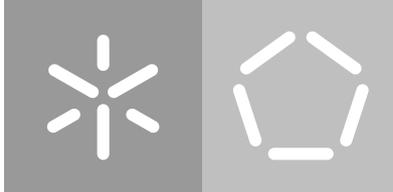


Universidade do Minho

Escola de Engenharia

Carlos Daniel Coutinho Faria

**Data Science para o Controlo Analítico
e Eficiência Energética numa ETAR**



Universidade do Minho

Escola de Engenharia

Carlos Daniel Coutinho Faria

**Data Science para o Controlo Analítico
e Eficiência Energética numa ETAR**

Dissertação de Mestrado

Mestrado em Engenharia Informática

Trabalho efetuado sob a orientação do(a)

Paulo Jorge Freitas de Oliveira Novais

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositoriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



**Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
CC BY-NC-SA 4.0**

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en>

Agradecimentos

Em primeiro lugar, gostaria de agradecer ao meu orientador, professor Paulo Jorge Freitas de Oliveira Novais, por me ter dado a oportunidade de realizar esta dissertação, sob sua orientação. Quero destacar todo o seu empenho, compromisso e auxílio prestado ao longo deste projeto. Sem dúvida, foi uma pessoa fundamental para o resultado final desta dissertação. Para o Pedro Oliveira, quero agradecer por toda a sua disponibilidade e apoio no decorrer deste trabalho, onde sempre prestou esclarecimento das dúvidas que surgiram com o decorrer do mesmo. Gostaria de agradecer também aos engenheiros Francisco Aguiar e Frederico Barros Lopes das Águas do Norte pela disponibilização dos dados utilizados ao longo da dissertação e por toda a amabilidade prestada.

A todos os meus amigos gostaria de agradecer o apoio constante e amizade, ajudando-me a manter um espírito positivo ao longo desta temporada.

Por fim, e não menos importante, um agradecimento à minha família, especialmente aos meus pais e irmão, que sempre me apoiaram em todos os momentos, dando-me sempre a força necessária e motivação.

Este trabalho foi parcialmente apoiado por Fundos Nacionais através da agência de financiamento portuguesa, FCT - Fundação para a Ciência e a Tecnologia, dentro do projeto DSAIPA/AI/0099/2019.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

(Carlos Daniel Coutinho Faria)

Resumo

Atualmente, principalmente nos países mais desenvolvidos, as águas residuais são tratadas através das Estação de Tratamento de Águas Residuais (ETARs) de forma a tentar atenuar os efeitos da atividade humana na poluição da água potável. As ETARs são infraestruturas que desempenham um papel fundamental e imprescindível para a sociedade, pois estas permitem que a água potável já utilizada em diversas atividades e no uso doméstico possa retornar ao seu habitat natural nas melhores condições, podendo ser reaproveitada.

Contudo, uma ETAR para realizar a sua função requer elevados consumos energéticos, devido a todo o processo de tratamento que é auxiliado com um número elevado de equipamentos, desde das águas residuais afluentes até ao seu destino final. Esses consumos oscilam consoante o tipo de afluente, sendo importante analisar através de um controlo analítico, quais as substâncias que, quando presentes no afluente, requerem maior tratamento, e conseqüentemente maior consumo de energia.

Com isto, o objetivo desta dissertação é, através do uso de *Data Science*, elaborar um forte controlo analítico às substâncias do afluente e relacionar o mesmo com a oscilação do consumo energético, tornando as ETARs mais eficientes e sustentáveis.

Palavras-chave: Controlo Analítico, Data Science, Eficiência Energética, Estação de Tratamentos de Águas Residuais

Abstract

Currently, mainly in more developed countries, wastewater is treated through Wastewater Treatment Plants (WWTPs) to mitigate the effects of human activity on the source of drinking water. WWTPs are infrastructures that play a fundamental and indispensable role for society, as these allow the water supplied in various activities and for domestic use to return to its natural habitat in the best conditions, to be able to reuse it.

However, a WWTPs to perform its function requires high energy consumption due to the entire treatment process assisted with an increased number of equipment, from the inflow of the affluent wastewater to its final destination. These consumptions vary depending on the affluent, being essential the analysis through an analytical control, which substances, when present in the affluent, require more effective treatment and consequently greater energy consumption.

With this, this dissertation's objective is, through the use of Data Science, to elaborate a robust analytical control to the affluent substances and relate it with the oscillation of energy consumption, making the WWTPs more efficient and sustainable.

Keywords: Analytical Control, Data Science, Energy Efficiency, Wastewater Treatment Plant

Índice

Lista de Figuras	viii
Lista de Tabelas	xi
Siglas	xiii
1 Introdução	1
1.1 Enquadramento e Motivação	1
1.1.1 ETAR	3
1.1.2 Controlo Analítico	5
1.1.3 Eficiência Energética	5
1.2 Objetivos	6
1.3 Estrutura do Documento	7
2 Estado da Arte	8
2.1 Data Science	8
2.1.1 Machine Learning	10
2.2 Controlo Analítico numa ETAR	14
2.2.1 Data Science para Controlo Analítico numa ETAR	15
2.2.2 Substâncias Relevantes	16
2.2.3 Valores Padrão	21
2.3 Eficiência Energética numa ETAR	21
2.3.1 Data Science para Eficiência Energética numa ETAR	23
2.3.2 Fatores que Afetam o Consumo Energético numa ETAR	24
2.4 Revisão da Literatura	25
2.4.1 Controlo Analítico	25
2.4.2 Eficiência Energética	28
3 Materiais e Métodos	32
3.1 Recolha e Armazenamento dos Dados	32
3.2 Exploração Inicial aos Dados	34

3.2.1	Vila Real	34
3.2.2	Guimarães	41
4	Manipulação dos Dados	45
4.1	Tratamento dos Dados	45
4.1.1	Vila Real	45
4.1.2	Guimarães	48
4.2	Análise Exploratória dos Dados	50
4.2.1	Vila Real	51
4.2.2	Guimarães	75
4.3	Plataforma de Visualização	88
4.4	Tecnologias Utilizadas	91
5	Casos de Estudo	93
5.1	Enquadramento	93
5.2	Análise e tratamento dos dados	94
5.3	Experiências	95
5.4	Resultados e discussão	96
6	Conclusão e Trabalho Futuro	99
	Bibliografia	103
	Anexos	109
I	Dados Electricidade	109
I.1	Vila Real	109
I.2	Guimarães	112
II	Dados Controlo Analítico	117

Lista de Figuras

1.1	Esquema do ciclo urbano da água.	3
1.2	Esquema de funcionamento de uma ETAR.	4
2.1	Fases aplicação <i>Data Science</i>	9
2.2	Exemplo de uma Decision Tree (DT).	12
2.3	Exemplo de uma Random Forest (RF).	13
2.4	Variação do fluxo diário do caudal de águas residuais	24
3.1	Modelo Conceptual Base de Dados.	33
3.2	Esquema Base de Dados.	34
3.3	Variação média da Electricidade de Vila Real ao longo dos meses de todos os anos.	36
3.4	Variação média do Funcionamento ao longo dos meses de todos os anos.	40
3.5	Variação média da Electricidade de Guimarães ao longo dos meses de todos os anos.	42
3.6	Variação do setor Lamas Flotadas ao longo dos meses de todos os anos.	43
4.1	Variação média da Electricidade de Vila Real ao longo dos meses de todos os anos após tratamento.	47
4.2	Variação média do Funcionamento ao longo dos meses de todos os anos após tratamento.	48
4.3	Variação média da Electricidade ao longo dos meses de todos os anos após tratamento.	49
4.4	Variação média do setor Lamas Flotadas ao longo dos meses de todos os anos após tratamento.	51
4.5	Variação Média da Electricidade de Vila Real ao longos dos meses em cada ano.	52
4.6	Variação Média da Electricidade de Vila Real ao longos dos meses em cada ano por indicadores.	53
4.7	Valores médios da Electricidade de Vila Real em cada ano.	53
4.8	Valores médios da Electricidade de Vila Real em cada ano por indicadores.	54
4.9	Valores médios da Electricidade de Vila Real em cada trimestre.	55
4.10	Valores médios da Electricidade de Vila Real em cada trimestre por indicador.	56
4.11	Valores Médios da Electricidade de Vila Real em cada estação do ano.	56
4.12	Valores Médios da Electricidade de Vila Real em cada estação do ano por indicador.	57
4.13	Correlações entre os indicadores da Electricidade de Vila Real.	58

4.14	Varição Média de alguns indicadores do Controlo Analítico em cada mês do ano.	60
4.15	Varição Média de alguns indicadores do Controlo Analítico por mês em cada ano.	61
4.16	Varição Média de alguns indicadores do Controlo Analítico por trimestres.	62
4.18	Correlações entre os indicadores e respetivos setores do Controlo Analítico.	63
4.17	Varição Média de alguns indicadores do Controlo Analítico por estações do ano.	64
4.19	Valores médios do Funcionamento por mês em cada ano.	65
4.20	Varição Média geral dos processos do funcionamento por mês em cada ano.	67
4.21	Valores médios do Funcionamento em cada ano.	68
4.22	Varição Média dos processos do Funcionamento em cada ano.	69
4.23	Valores médios do Funcionamento em cada trimestre.	70
4.24	Varição Média dos processos do funcionamento por trimestres.	71
4.25	Valores médios do Funcionamento em cada estação do ano.	72
4.26	Varição Média dos processos do funcionamento por estações do ano.	73
4.27	Varição Média da Electricidade de Guimarães ao longos dos meses em cada ano.	76
4.28	Varição Média da Electricidade de Guimarães ao longos dos meses em cada ano por indicadores.	77
4.29	Valores médios da Electricidade de Guimarães em cada ano.	77
4.30	Valores médios da Electricidade de Guimarães em cada ano por indicadores.	78
4.31	Valores médios da Electricidade de Guimarães em cada trimestre.	79
4.32	Valores médios da Electricidade de Guimarães em cada trimestre por indicador.	79
4.33	Valores Médios da Electricidade de Guimarães em cada estação do ano.	80
4.34	Valores Médios da Electricidade de Guimarães em cada estação do ano por indicador.	81
4.35	Correlações entre os indicadores da Electricidade de Guimarães.	81
4.36	Varição Média dos Caudais ao longos dos meses em cada ano.	83
4.37	Varição Média de alguns indicadores dos Caudais ao longos dos meses em cada ano.	83
4.38	Valores médios dos Caudais em cada ano.	84
4.39	Valores médios de alguns indicadores dos Caudais em cada ano.	85
4.40	Valores médios dos Caudais em cada trimestre.	85
4.41	Valores médios de alguns indicadores dos Caudais em cada trimestre.	86
4.42	Valores Médios dos Caudais em cada estação do ano.	87
4.43	Valores Médios de alguns indicadores dos Caudais em cada estação do ano.	87
4.44	Arquitetura da plataforma desenvolvida.	89
4.45	Ilustração na plataforma dos dados da electricidade da Estação de Tratamento de Águas Residuais (ETAR) de Vila Real.	89
4.46	Ilustração na plataforma dos dados da electricidade da ETAR de Vila Real.	90
4.47	Ilustração na plataforma dos dados da electricidade da ETAR de Vila Real.	90
4.48	Ilustração na plataforma dos dados do Controlo Analítico da ETAR de Vila Real.	91
4.49	Ilustração na plataforma dos dados do Controlo Analítico da ETAR de Vila Real.	91

5.1	Previsão do azoto total utilizando os melhores modelos de DT e RF.	97
5.2	Ilustração na plataforma das previsões efetuadas.	98
I.1	Soma dos consumos de electricidade por ano em cada indicador.	109
I.2	Soma dos consumos de electricidade por trimestre em cada indicador.	110
I.3	Soma dos consumos de electricidade por estação do ano em cada indicador.	110
I.4	Distribuição dos consumos totais da electricidade por indicador.	111
I.5	Distribuição dos consumos totais da electricidade por anos.	111
I.6	Distribuição dos consumos totais da electricidade por trimestre.	112
I.7	Distribuição dos consumos totais da electricidade por estação do ano.	112
I.8	Soma dos consumos de electricidade por ano em cada indicador.	113
I.9	Soma dos consumos de electricidade por trimestre em cada indicador.	113
I.10	Soma dos consumos de electricidade por estação do ano em cada indicador.	114
I.11	Distribuição dos consumos totais da electricidade por indicador.	114
I.12	Distribuição dos consumos totais da electricidade por ano.	115
I.13	Distribuição dos consumos totais da electricidade por trimestre.	115
I.14	Distribuição dos consumos totais da electricidade por estação do ano.	116
II.1	Variação Média do indicador Fósforo Total em cada mês do ano.	117
II.2	Variação Média do indicador Amónia em cada mês do ano.	118
II.3	Variação Média do indicador Nitratos em cada mês do ano.	118
II.4	Variação Média do indicador Oxigénio em cada mês do ano.	119
II.5	Variação Média do indicador Sólidos Totais (ST) em cada mês do ano.	119
II.6	Variação Média do indicador Sólidos Voláteis (SV) em cada mês do ano.	120
II.7	Variação Média do indicador Sólidos Suspensos Totais (SST) em cada mês do ano.	120
II.8	Variação Média do indicador Sólidos Suspensos Voláteis (SSV) em cada mês do ano.	121

Lista de Tabelas

2.1	Tabela dos Valores Limite de Emissão em descargas de Águas Residuais	21
3.1	Períodos Diários da Electricidade.	35
3.2	Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Vila Real.	35
3.3	Dados Controlo Analítico.	37
3.4	Dados Funcionamento.	39
3.5	Dados Tempo Vila Real.	40
3.6	Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Guimarães.	41
3.7	Dados Caudais.	43
3.8	Dados Tempo Guimarães.	44
4.1	Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Vila Real, após tratamento.	51
4.2	Interpretação dos valores das correlações.	58
4.3	Resultado da Estatística Descritiva aplicada aos dados do Controlo Analítico, após tratamento.	59
4.4	Resultado da Estatística Descritiva aplicada aos dados do Funcionamento, após tratamento.	66
4.5	Maiores correlações encontradas entre Electricidade e Controlo Analítico.	73
4.6	Maiores correlações encontradas entre Electricidade e Funcionamento.	74
4.7	Algumas das maiores correlações encontradas entre Controlo Analítico e Funcionamento.	74
4.8	Maiores correlações encontradas entre o Tempo e os restantes tipo de dados.	75
4.9	Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Guimarães, após tratamento.	75
4.10	Dados Caudais após tratamento.	82
4.11	Maiores correlações encontradas entre Electricidade e Caudais.	88
4.12	Maiores correlações encontradas entre o Tempo e os restantes tipo de dados.	88
5.1	Maiores correlações encontradas entre o <i>target</i> de estudo e os restantes indicadores.	95
5.2	Variáveis presentes no <i>dataset</i> final.	95
5.3	Hiperparâmetros testados	96
5.4	Top-3 modelos DT e RT. Legenda: a - max_depth; b - max_samples_split; c - min_samples_leaf; d - max_features; e - n_estimators; f - splitter; g - RMSE; h - MAE; i - time (segundos).	97

5.5 Algumas regras definidas pelo melhor modelo candidato. 98

Siglas

<i>CBO</i> ₅	Carência Bioquímica de Oxigênio após 5 dias
API	Application Programming Interface
AR	Águas Residuais
ART	Águas Residuais Tratadas
CART	Classification And Regression Tree
CBO	Carência Bioquímica de Oxigênio
COT	Carbono Orgânico Total
CQO	Carência Química de Oxigênio
CRISP-DM	Cross Industry Standard Process for Data Mining
DEA	Data Envelopment Analysis
DEMOCON	Decentralized Monitoring and Control
DL	Deep Learning
DS	Decision Support
DT	Decision Tree
EDA	Exploratory Data Analysis
ETA	Estação de Tratamento de Água
ETAR	Estação de Tratamento de Águas Residuais
IA	Inteligência Artificial
ID3	Iterative Dichotomiser 3
KPI	Key Performance Indicator
kWh	Quilowatt-Hora

MAE	Mean Absolute Error
ML	Machine Learning
MLR	Multiple Linear Regression
RA	Rapid Audit
RBF	Radical Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
SEMMA	Sample, Explore, Modify, Model, and Assess
SFT	Sólidos Fixos Totais
SSF	Sólidos Suspensos Fixos
SSsed	Sólidos Sedimentáveis
SST	Sólidos Suspensos Totais
SSV	Sólidos Suspensos Voláteis
ST	Sólidos Totais
SV	Sólidos Voláteis
SVT	Sólidos Voláteis Totais
UM	Universidade do Minho
URE	Utilização Racional da Energia
VLE	Valores Limite de Emissão

Introdução

O presente trabalho surge na unidade curricular de Dissertação em Engenharia Informática, na [Universidade do Minho \(UM\)](#), com o título “Data Science para o Controlo Analítico e Eficiência Energética numa ETAR”. Este projeto aborda o estudo na área de ciências de computação, mais concretamente *Data Science* para auxiliar no apoio de tomada de decisão, no contexto das [ETAR](#). Neste capítulo, na secção 1.1 é feito um enquadramento do tema assim como a motivação para a realização desta dissertação. Na secção 1.2 são referenciados os objetivos inerentes ao desenvolvimento da dissertação. Por fim, na secção 1.3 é descrita a respetiva organização deste documento e os tópicos que cada capítulo abrange.

1.1 Enquadramento e Motivação

Com o crescimento de forma exponencial da espécie humana e com a necessidade de satisfazer os bens essenciais, tem-se vindo a registar um aumento dos problemas ambientais do nosso planeta, resultantes entre outros das atividades industriais [1]. A sociedade atual produz elevadas quantidades de resíduos (sólidos, líquidos e gasosos), que podem ter como consequência perturbações no meio ambiente, provocando mudanças na qualidade dos solos, água e ar [2].

O crescimento populacional resultante do desenvolvimento das sociedades tem elevado a um aumento das preocupações relativamente à preservação da água. Tendo em consideração essa preocupação, juntamente com a mudança nos estilos de vida da espécie humana e com o aumento dos padrões de consumo de água, é estimado que nos próximos anos o aumento da procura pela água, energia e outros bens e serviços (que necessitam de água para o seu desenvolvimento), possam tornar o recurso natural mais escasso.

Com a qualidade e quantidade de água potável a ser um tópico cada vez mais em foco mundialmente, torna-se necessário que haja uma melhor gestão deste recurso natural [3]. Com base nesta necessidade, tornou-se imperativo tratar as [Águas Residuais \(AR\)](#) para atenuar efeitos ambientais prejudiciais, além

da possibilidade da reutilização destas noutras atividades, como por exemplo, na agricultura. As **AR**, normalmente conhecidas como esgotos, é o termo utilizado para definir as águas, que após a utilização humana, apresentam as suas características alteradas [4]. Portanto, o tratamento das **AR** é de extrema importância pois, para além do supracitado, pode evitar risco para a saúde pública, poluição dos recursos hídricos para os quais são descarregados os efluentes e poluição do ambiente num modo geral, sob a forma de odores desagradáveis ou contaminação das águas subterrâneas [5].

Com o intuito de colmatar essa necessidade, surgiram então as **ETAR**. Atualmente, as **ETAR** são infraestruturas de extrema importância para o meio ambiente, uma vez que, realizam a despoluição de inúmeros cursos de água onde, diariamente, são canalizadas elevadas cargas de afluentes poluentes, através das redes de esgotos [6]. Porém, as **ETAR** com a execução das suas funções têm elevados consumos energéticos, refletindo cerca de 7% de toda a energia consumida à escala mundial [7]. A nível nacional, cerca de 4% da energia elétrica utilizada é da responsabilidade do ciclo urbano da água, sendo aproximadamente 25% dessa energia utilizada nas **ETAR** [8]. Por isso, alcançar um nível de eficiência energética numa **ETAR** tornou-se num tópico imprescindível na gestão das mesmas, pois prevê-se que o consumo energético destas infraestruturas tende a aumentar com o envelhecimento dos equipamentos. Além do mais, a necessidade de implementar novas tecnologias para, salvaguardando a qualidade das águas residuais tratadas, ir ao encontro de padrões de qualidade mais exigentes, tornando-se mais eficaz o tratamento ao afluente. Consequentemente, isso acarreta um acréscimo nos custos associados a este setor, sendo crucial alcançar uma melhor eficiência energética nas **ETAR**, garantindo que o tratamento seja igualmente cumprido ou melhorado.

Aqui entra a *Data Science*, que é uma área interdisciplinar ligada ao estudo e análise de dados provenientes de diferentes setores, sendo os dados estruturados ou não estruturados, combinando diferentes campos de estatística e computação para a interpretação dos dados, visando a extração do conhecimento para obtenção de variáveis relevantes para possíveis tomadas de decisão. A aplicação desta área no contexto das **ETAR** pode trazer inúmeras vantagens, uma vez que com uma análise profunda a todos os dados inerentes a estas infraestruturas utilizando-se diversas técnicas de estatística e computação, pode levar à extração de diversos *insights* o que ajuda na tomada de decisão.

Assim sendo, a motivação para a realização desta dissertação foi o facto de ser evidente que as **ETAR** necessitam de uma gestão apoiada na sustentabilidade e na estabilidade nos processos de tratamento. É importante procurar obter, através do uso de *Data Science*, o alicerce necessário para auxiliar as tomadas de decisão, baseado nos fenómenos verificados na exploração dos dados relativos aos sistemas, de modo a expor as razões desses mesmos comportamentos ou resultados. O uso de *Data Science* na gestão de uma **ETAR** pode ajudar na rentabilização dos custos operacionais, tornando as **ETAR** mais eficientes, estáveis e sustentáveis.

1.1.1 ETAR

Uma **ETAR** é uma instalação de tratamento de **AR**, que recorre a vários processos físicos, químicos e biológicos, através de vários órgãos de tratamento (equipamentos e instalações) para tratar as **AR**. Após efetuado este tratamento, as água são devolvidas ao meio ambiente em condições ambientalmente seguras [9]. As **AR** devem ser devidamente tratadas numa **ETAR**, onde as **Águas Residuais Tratadas (ART)** antes de serem devolvidas ao meio ambiente, têm de cumprir com a legislação e a licença de descarga, garantindo assim que os efluentes não poluem o meio ambiente e estão em condições de ser reutilizados, assegurando assim o ciclo urbano da água. A Figura 1.1 ilustra esse mesmo ciclo.

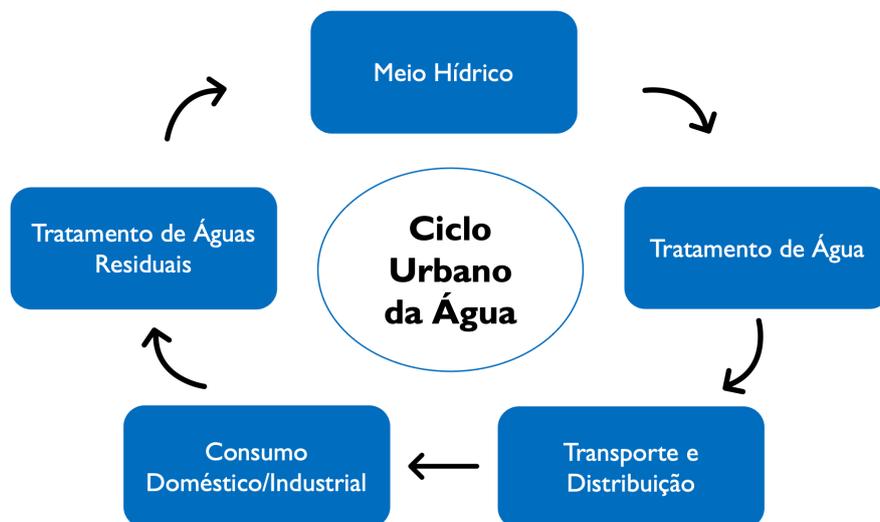


Figura 1.1: Esquema do ciclo urbano da água.

O ciclo urbano da água corresponde a todas as etapas ou fases de utilização da água desde o momento da sua captação até ser restituída à natureza. A água que chega às residências, inicia-se na captação da água em meios hídricos, seguindo para as **Estação de Tratamento de Água (ETA)** onde é tratada para ser possível proceder-se à sua utilização. Após isso, é armazenada em reservatórios e daí são distribuídas à população [10]. Depois de utilizada, é tratada nas **ETAR**, sendo novamente devolvidas ao meio ambiente, em condições ambientais seguras.

Uma **ETA**, a par das **ETAR**, têm um papel fundamental para a sociedade. Garantem que a água, antes de ser consumida pela população, cumpre todos os requisitos necessários para que seja possível proceder-se ao seu consumo de uma forma segura. Após ser captada, a água é encaminhada para a **ETA** onde é feita a correção das características físicas, químicas e bacteriológicas, garantindo assim que a água é adequada para consumo [10]. Após o consumo da água, como supracitado, a mesma é encaminhada para a **ETAR**.

Para se proceder ao tratamento das águas residuais nas **ETAR** são utilizadas técnicas avançadas de purificação e processamento, como se pode verificar na Figura 1.2:

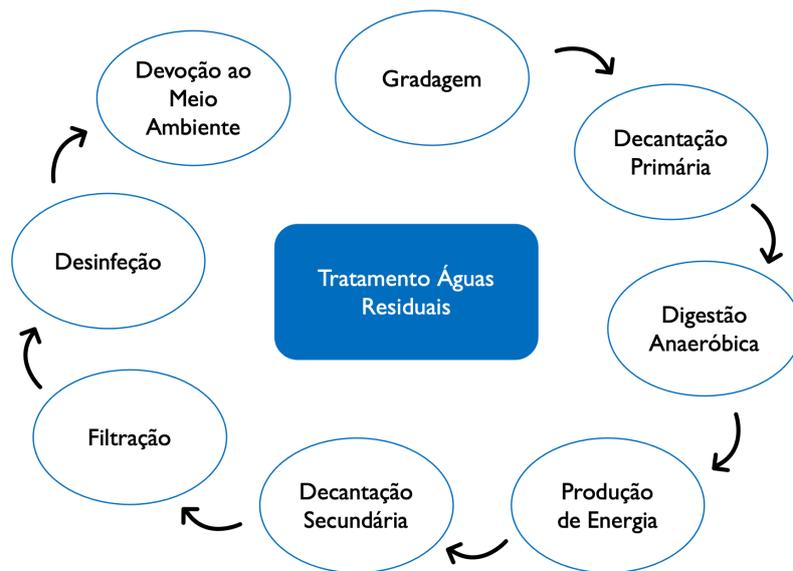


Figura 1.2: Esquema de funcionamento de uma ETAR.

Além disso, uma **ETAR** pode integrar cinco fases de tratamento:

- **Tratamento Preliminar:** Nesta primeira fase de tratamento são aplicadas três técnicas ao afluente para efetuar o tratamento inicial, sendo elas a gradagem, a desarenação e o desengorduramento. A gradagem consiste na separação dos sólidos que contêm maior dimensão de forma automática ou manual. A desarenação consta na realização da decantação às areias, sólidos e outros detritos imóveis numa caixa apropriada. Por fim, no desengorduramento é efetuada a remoção das gorduras em emulsão no afluente [11].
- **Tratamento Primário:** Nesta fase, é efetuado o tratamento físico-químico ao afluente. É executado a regularização do pH do afluente, uma vez que nas próximas etapas de tratamento necessitam que o afluente apresente uma escala de pH ideal. Depois é feita a coagulação/floculação, onde é feita a separação da matéria presente no afluente, mediante a densidade de flóculos, para mais tarde permitir que seja realizada a sua decantação. Por fim, este tratamento é realizado por vezes em determinados processos industriais onde é necessária uma depuração maior, recorrendo-se à osmose inversa para se conseguir obter um efluente de maior qualidade [12].
- **Tratamento Secundário:** Conhecido também como tratamento biológico, o tratamento secundário tem como objetivo reduzir a carga orgânica dos afluentes, recorrendo-se à digestão anaeróbia ou à aeróbia por microrganismos. Após isso, o afluente passa pela decantação secundária, o que permite o depósito das lamas resultantes do tratamento. Esta fase é responsável, em norma, por 60% a 70% de todo o consumo energético numa **ETAR** [11].
- **Tratamento Terciário:** Nesta etapa, é realizado o polimento final do afluente. É feita a remoção de excesso de fósforo ou azoto, caso a descarga final seja realizada para um meio sensível, assim

como um qualquer tipo de tratamento adicional para ser possível reaproveitar o efluente. O tratamento aplicado nesta fase pode ser muito específico, consoante o tipo de reaproveitamento que será aplicado ao afluente. Os destinos finais do afluente podem ser para rega de espaços verdes, combate a incêndios, agricultura, lavagens de pavimentos e ruas, entre outras opções [9].

- **Tratamento das Lamas:** Este tratamento é específico às lamas que provêm das várias fases de tratamento realizadas numa ETAR. As lamas contêm um teor de microrganismos patogénicos elevados, portanto devem ser tratados de forma a transformar as lamas num resíduo inerte. Para isso, é realizada a desidratação e desinfecção das lamas de forma mecânica ou química. Após ser efetuado este tratamento, as lamas podem ser utilizadas no setor dos combustíveis ou então como fertilizantes na agricultura [11].

1.1.2 Controlo Analítico

O controlo analítico refere-se a todos os procedimentos e processos que pretendam garantir que os resultados obtidos das análises laboratoriais sejam precisos, consistentes e sobretudo que estejam dentro dos limites dos padrões estabelecidos, referentes a uma determinada norma, dependendo da área em questão. Com isso, um controlo analítico apertado assume um papel cada vez mais preponderante na manutenção da garantia da qualidade e segurança, com o objetivo de fornecer informações para a tomada de decisão com base em determinadas amostras de análise [13].

A realização de um bom controlo analítico em diversas infraestruturas é peça fundamental para que seja possível agir em conformidade, quando situações anómalas ocorrem e é necessários corrigi-las [14].

Portanto, o controlo analítico é a melhor forma de garantir que as características de determinado produto, água, etc, estão dentro dos limites aceitáveis assegurando assim que não existe risco para os seres vivos ou perturbações negativas no meio ambiente.

1.1.3 Eficiência Energética

Com a clara ameaça da extinção das reservas de combustíveis fósseis, as preocupações sobre o ambiente e a pressão dos resultados económicos, levou que a eficiência energética seja encarada como uma das melhores soluções para equilibrar o modelo de consumo existente, assim como para combater as alterações climáticas [15]. A eficiência energética, de uma forma simplificada, significa utilizar a menor energia possível para a execução da mesma tarefa, isto é, ao ser realizada uma determinada tarefa tentar sempre que o desperdício energético seja o menor possível [16].

Toda a energia percorre um processo de transformação, ao qual resulta a luz, frio, calor, entre outros. Durante toda a transformação, uma parte dessa energia é desperdiçada, e a outra que chega ao consumidor final, nem sempre é aproveitada da melhor forma. Por isso, a eficiência energética implementa medidas para combater o desperdício de energia ao longo de todo o processo de transformação.

A eficiência energética segue todo um processo de produção, distribuição e utilização da energia, onde pode ser dividido em duas fases, a transformação e a utilização [15]:

- **Transformação:** A energia está presente na natureza em várias formas, e para ser possível a sua utilização, é necessário proceder-se à sua transformação. Os processos de transformação, transporte e utilização final da energia acarretam impactos negativos no meio ambiente. Parte das perdas são inevitáveis e outras resultam de questões físicas, mas a restante parte é perdida devido ao mau aproveitamento e à inexistência de otimização dos sistemas de transformação. Este desperdício, ao longo dos anos, tem merecido especial atenção por parte das empresas que processam e vendem a energia, que pretendem reduzir ao máximo essas perdas, pois esse desperdício energético traduz-se em dinheiro perdido. Por parte dos governos de vários países, são implementadas medidas que visam uma melhor utilização da energia de forma a refletir um consumo energético otimizado, permitindo assim uma melhor racionalização deste recurso [17].
- **Utilização:** O desperdício de energia não está unicamente presente no processo de transformação, mas também no processo de utilização. Relativamente a este processo, a eficiência energética está constantemente ligada ao termo [Utilização Racional da Energia \(URE\)](#), que implica a adoção de várias medidas que possibilite uma melhor utilização e aproveitamento da energia em diversos setores, como setor doméstico, serviços e industrial, assim como para o meio ambiente [17].

Utilizar a energia que dispomos de forma otimizada e responsável, poderá atenuar de forma significativa os problemas ambientais existentes, permitindo às gerações futuras um futuro mais risonho.

1.2 Objetivos

Para a resolução deste projeto é necessário superar várias etapas, onde os principais objetivos são:

- Levantamento da literatura existente no âmbito da ciência dos dados no campo das [ETAR](#);
- Utilização de uma metodologia de extração de conhecimento ([Cross Industry Standard Process for Data Mining \(CRISP-DM\)](#), [Sample, Explore, Modify, Model, and Assess \(SEMMA\)](#), etc);
- Compreensão dos dados fornecidos de uma [ETAR](#);
- Análise aos dados fornecidos e deteção de padrões e correlações nos mesmos;
- Tratamento dos dados, como deteção e resolução de *missing values*;
- Elaboração de *dashboards* de visualização da análise de dados realizada.

1.3 Estrutura do Documento

O presente documento está estruturado em vários capítulos. No capítulo 1 é descrito os conceitos teóricos que estão na base deste documento, bem como a motivação para a elaboração deste tema. São referidos também os objetivos inerentes à realização desta dissertação.

No capítulo 2 está presente o estado da arte, onde são apresentados os fundamentos teóricos que sustentam os conceitos que serão aplicados no desenvolvimento do projeto, bem como a revisão da literatura existente.

No capítulo 3 ilustra-se os matérias e métodos utilizados na resolução desta dissertação, desde o armazenamento dos dados a uma análise inicial aos mesmos.

Já no capítulo 4 são exibidas as manipulações realizadas nos dados para se obter o máximo de informação possível. É realizada também uma análise profunda aos dados utilizados com o intuito de expor as suas principais características.

No capítulo 5 é descrito o caso de estudo realizado neste projeto de dissertação, onde exhibe com pormenor todas as etapas realizadas no mesmo.

Por fim, no capítulo 6 é apresentada uma conclusão do trabalho realizado, assim como o trabalho futuro a executar.

Estado da Arte

Neste capítulo é apresentada uma análise do estado da arte com o principal objetivo de identificar os diferentes trabalhos disponíveis na literatura, assim como clarificar em que consiste a área de *Data Science* e direcionar a mesma nos contextos de controlo analítico e eficiência energética. A metodologia utilizada para a elaboração da pesquisa sobre os temas abordados neste capítulo consistiu, principalmente, na análise de artigos disponíveis nas bases de dados do Google Scholar, ResearchGate, Science Direct, entre outros. A secção 2.1 fornece uma análise à área de *Data Science*, descrevendo em que consiste a mesma, a sua finalidade, assim como as suas características e etapas de aplicação. Na secção 2.2 é descrito em que consiste um controlo analítico numa ETAR, como se pode e quais as vantagens em aplicar *Data Science* para controlo analítico, as principais substâncias e respetivas características presentes nas AR, os valores padrão e os valores de emissão de descarga que os efluentes têm de apresentar. Já na secção 2.3 é retratada a eficiência energética numa ETAR, como se consegue e quais as mais valias em aplicar *Data Science* para a eficiência energética, assim como os fatores que influenciam os consumos energéticos numa ETAR. Por fim, na secção 2.4 é feita uma análise aos trabalhos já existentes na literatura no contexto do tema desta dissertação, onde se dá foco aos métodos de resolução usados, tecnologias e abordagens utilizadas, resultados e respetivas conclusões.

2.1 Data Science

A *Data Science* é uma ampla área de estudo no domínio dos sistemas e processamento de dados, com o principal objetivo de, combinando diferentes campos da estatística e da computação para a interpretação dos dados, aplicar uma série de técnicas para obter *insights* para auxiliar no processo de tomada de decisão [18]. A estatística e o uso de modelos estatísticos estão profundamente ligados à *Data Science*, onde inicialmente se utilizava apenas estatística, sendo que ao longo dos anos a evolução levou a que se comesçassem a incluir conceitos/práticas de *Inteligência Artificial (IA)* e *Machine Learning (ML)* [19].

Cada vez mais se aplica *Data Science* em diversas áreas, como medicina, engenharia, ciências sociais e empresas, tornando-se cada vez mais necessária e imprescindível para a evolução destas áreas que geram enormes quantidades de dados [20] [21]. O processo de *Data Science* tem diversos componentes chave que permitem chegar ao resultado final, como pode ser verificado na Figura 2.1, sendo eles:



Figura 2.1: Fases aplicação *Data Science*

- **Percepção do tema/área de estudo:** A primeira etapa a ser efetuada em qualquer projeto de *Data Science* e a de maior relevância é o estudo do tema/área em questão. É de extrema importância perceber a área de estudo para se obter uma compreensão do domínio em causa. Esse conhecimento permite ter uma melhor ideia de como aplicar os recursos de engenharia ao problema nas etapas seguintes [22].
- **Obtenção dos dados:** Os dados são a base da *Data Science* pois permitem, após todos os processos de manipulação e desenvolvimento, chegar a conclusões e respostas que auxiliam a tomada de decisão. Em alguns casos a quantidade de dados disponível para o estudo é baixa, por isso é importante recolher o maior número de dados possível. Para isso existem diversas bibliotecas e métodos que permitem obter a informação desejada. Quanto mais dados se tiver acesso, mais certo e eficaz será o resultado final [22].
- **Análise e Tratamento dos dados:** Esta etapa é considerada fundamental em *Data Science*, tendo como objetivo um estudo minucioso aos dados, identificando dados incompletos, imprecisos, irrelevantes e ausentes. De seguida, esses dados identificados são modificados, substituídos ou excluídos conforme a necessidade. Caso não fosse realizada esta etapa, os resultados finais podiam ficar muito distantes da realidade, o que podia levar a tomadas de decisão erradas [23].
- **Análise exploratória dos dados:** Nesta fase é feita uma análise metódica aos dados, onde se aplicam diversas técnicas estatísticas e de visualização para se identificar mais detalhadamente os padrões e correlações que os dados apresentam [24].

- **Criação de modelos de ML e Deep Learning (DL):** Nesta fase os dados são modelados de modo a que seja possível prever resultados e valores futuros de certas variáveis. São usados sobretudo fórmulas e algoritmos de regressão ou classificação para gerar modelos capazes de classificar ou prever valores e/ou padrões futuros. Isso permite generalizar as ocorrências assim como melhorar a eficiência da tomada de decisão. Dependendo do projeto em questão, esta etapa pode nem sempre ser realizada [25].
- **Interpretação dos dados** A última fase a efetuar é a interpretação dos dados e a respetiva representação dos mesmos. Nesta etapa são ilustradas todas as conclusões finais, os *insights*, correlações, assim como os padrões que se verificaram durante todo o estudo. Essa ilustração é feita normalmente utilizando *dashboards*, que permitem uma fácil interpretação dos resultados obtidos e com isso tirar conclusões para auxiliar na tomada de decisão [26].

2.1.1 Machine Learning

Uma enorme vantagem que os humanos possuem em relação aos computadores é a capacidade de conseguirem alterar automaticamente o seu tipo de comportamento baseado nos erros cometidos em episódios passados. Os humanos tentam resolver os seus problemas tentando-os corrigir ou encontrar uma nova abordagem em lidar com eles, sempre baseado no conhecimento adquirido em experiências antecedentes, assim como o uso do bom senso para se obter conclusões [27, 28]. Olhando para os algoritmos tradicionais, estes são criados para desempenhar uma determinada tarefa, mas no final não olham para o resultado obtido, e conseqüentemente, não conseguem melhorar o seu desempenho. Para colmatar esta limitação, foi criado o conceito de ML. A sua essência remete-se à criação de modelos e ferramentas que tem a capacidade de aprender e conseqüentemente melhor o desempenho obtido, através de uma recolha contínua de dados onde é adquirida a experiência e o conhecimento [29]. ML é um subcampo da IA que, com base num determinado conjunto de dados, executa algoritmos que cria de forma automática modelos de representação de conhecimento. Esta aprendizagem assenta na ideia de que se deve treinar as máquinas, fornecendo-lhes como *input* o máximo de dados históricos, uma ou mais medidas de desempenho, e deixar o algoritmo aprender, ou seja, ajusta-lo de modo iterativo a fim que este melhore o seu desempenho [30]. Todo este procedimento corresponde ao processo de treino do modelo, onde quanto melhor o mesmo for realizado, mais qualidade terão as previsões ou classificações. Essas previsões ou classificações correspondem ao *output* obtido dos modelos treinados que podem representar, por exemplo, classificação de determinadas classes ou previsões de qualidade em situações futuras e que estejam relacionadas com padrões históricos[31]

2.1.1.1 Paradigmas de Aprendizagem

Em ML existem três tipos de aprendizagens principais, sendo elas Aprendizagem Supervisionada, Aprendizagem não Supervisionada e Aprendizagem por Reforço [32].

- **Aprendizagem Supervisionada:** Este tipo de aprendizagem requer que o conjunto de dados utilizado contenha informações sobre o resultado. Esta aprendizagem usa exemplos rotulados, utilizando *inputs* com os respectivos *outputs* conhecidos para treinar os modelos. Os algoritmos que seguem este tipo de aprendizagem aprendem ajustando um conjunto de parâmetros do modelo que operam, a fim de o mesmo se ajustar melhor ao conjunto de *outputs*. O objetivo de Aprendizagem Supervisionada é treinar um modelo na forma $y = f(x)$, para prever os *outputs* y , com base nos *inputs* x . Existem muitos tipos de modelos que seguem esta aprendizagem, sendo alguns deles: *DT*, *RF*, *Neural Networks*, entre outros [33].
- **Aprendizagem não Supervisionada:** Em contraste com Aprendizagem Supervisionada, os modelos que usam Aprendizagem não Supervisionada utilizam dados não rotulados no seus processos de treino. Nestes casos, os modelos operam sem saberem os resultados ou observações conhecidas, ou seja, esta técnica não tem como finalidade prever um valor específico. Em vez disso, o objetivo é descobrir padrões no *dataset* utilizado. Os modelos *K-means* e *K-Nearest neighbors* são alguns que utilizam este tipo de aprendizagem [34].
- **Aprendizagem por Reforço:** Na Aprendizagem por Reforço os modelos são treinados para tomarem uma sequência de decisões. O agente aprende com o intuito de atingir uma determinada meta num ambiente incerto e complexo, onde as soluções encontradas são obtidas utilizando uma abordagem de tentativa erro. O modelo sabe quais são os seus objetivos e recebe recompensas ou penalidades pelas ações que realiza. Portanto, o objetivo neste tipo de aprendizagem é maximizar ao máximo as recompensas [35].

2.1.1.2 Modelos Machine Learning

Como já mencionado, atualmente existem vários modelos de *ML*. Neste tópico o foco serão as *DT* e as *RF*, uma vez que foram os modelos utilizados no caso de estudo realizado, devido sobretudo ao reduzido tamanho do *dataset* utilizado para o efeito.

- **DT:** As *DT* são um método, não paramétrico, que utiliza aprendizagem supervisionada para problemas de classificação e regressão. Este modelo aprende com os dados a aproximarem-se de uma curva senoidal com um conjunto de regras *if-then-else*. Quanto mais profunda for a árvore, mais complexas serão as regras de decisão. Uma *DT* constrói os modelos de classificação ou regressão em forma de árvore, dividindo o conjunto dos dados em subconjuntos cada vez menores, e ao mesmo tempo, é desenvolvida incrementalmente uma árvore de decisão associada [36]. Na Figura 2.2 é representada um exemplo da estrutura deste tipo de modelo.
 - **Nó Raiz:** Representa a variável preditora, sendo o primeiro nó da árvore;
 - **Ramo:** Retrata os resultados de cada teste realizado num nó, onde este decidirá qual será a seguinte ramificação;

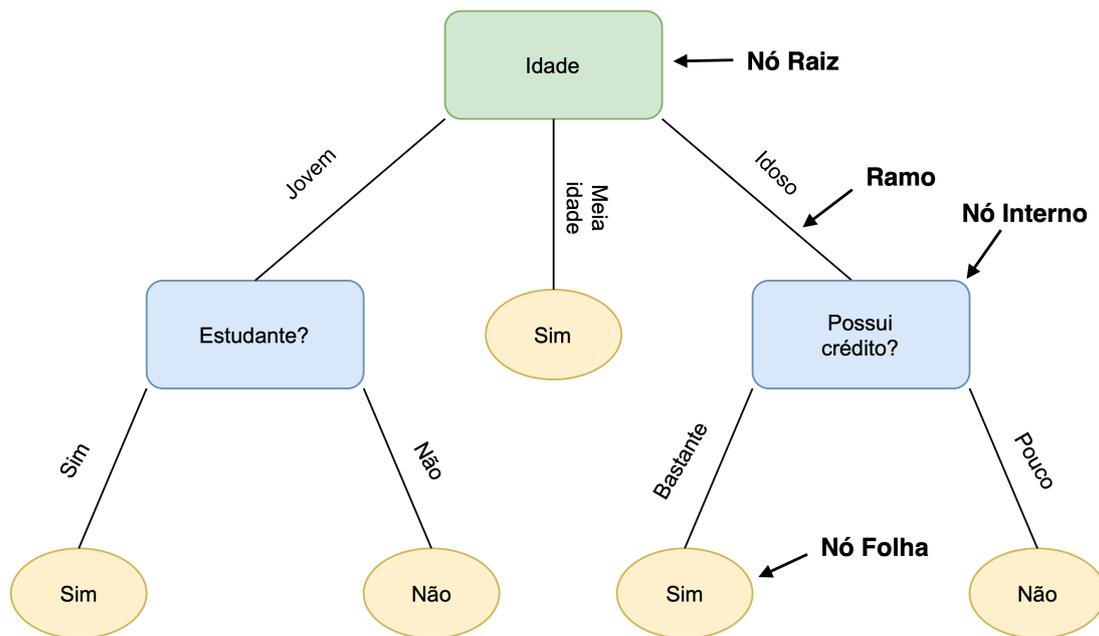


Figura 2.2: Exemplo de uma DT.

- **Nó Interno:** Representa também uma variável envolvendo todos os outros nós, com exceção do nó raiz e os terminais;
- **Nó Folha:** Corresponde aos nós terminais da árvore. Quando é atingido o nó folha, é atribuído o valor final para aquela variável, resultando uma classe ou um valor, dependendo do problema em questão.

Para a construção de uma DT existem vários algoritmos disponíveis, sendo os mais populares o *Iterative Dichotomiser 3 (ID3)*, *C4.5* e *Classification And Regression Tree (CART)*. O ID3, que foi criado por *Ross Quinlan*, onde cria árvores simples construídas de cima para baixo, e em cada iteração, é selecionada a melhor característica no momento atual para criar o nó. Contudo, torna-se um algoritmo limitado, uma vez que só é capaz de lidar com variáveis nominais, e por isso só pode ser utilizado em problemas de classificação. Já o algoritmo *C4.5*, também criado pelo *Ross Quinlan*, surgiu pela necessidade de melhorar o ID3, sendo que este algoritmo tem a capacidade de lidar com variáveis contínuas numéricas, permitindo assim ser utilizado em problemas de regressão. Outra vantagem que apresenta, é a possibilidade de ser realizada a "poda" da árvore após a sua criação, permitindo que sejam removidos os ramos que causam "ruído" nos dados, e conseqüentemente, melhorar a precisão das previsões ou classificações. Por fim, o algoritmo CART é muito similar ao C4.5, onde pode ser utilizado em problemas de classificação ou regressão [37].

Por fim, a utilização das DT tem algumas vantagens, mas também algumas limitações. Quando comparado com outros modelos, as DT são considerados modelos simples, onde ao mesmo tempo conseguem obter bons resultados com esforços reduzidos. Conseguem ser bastante rápidas e intuitivas, onde são capazes de fornecer uma boa visão de quais os atributos que são mais significativos

para a classificação ou previsão, podendo lidar com dados categóricos ou numéricos. Contudo, estes modelos não produzem bons resultados se existirem muitas variáveis não correlacionadas nos dados utilizados. Além disso, em *datasets* que sejam muito complexos seriam geradas árvores muito complexas, com inúmeras divisões e múltiplas subárvores, o que levaria a muito tempo no processo de treino e, conseqüentemente, tornar-se caro do ponto de vista computacional [38].

- **RF:** Outro modelo de ML considerado foi as RF, que tal como as DT, são um algoritmo de aprendizagem supervisionada. A "floresta" que este modelo constrói é um conjunto de DT que geralmente são treinadas com o método de *bagging*. A ideia geral desse método é criar uma combinação de modelos de aprendizagem que aumente os resultados gerais. As DT criadas numa RF são executadas em paralelo, sem nunca interagirem umas com as outras. Além disso, uma RF adiciona aleatoriedade adicional ao modelo ao "cultivar" as árvores, onde em vez de procurar a característica mais importante ao dividir o nó, procura a melhor característica entre um subconjunto aleatório de características. Isto dá resultado a uma grande diversidade que geralmente resulta num modelo mais preciso [39]. A Figura 2.3 ilustra um exemplo de uma RF.

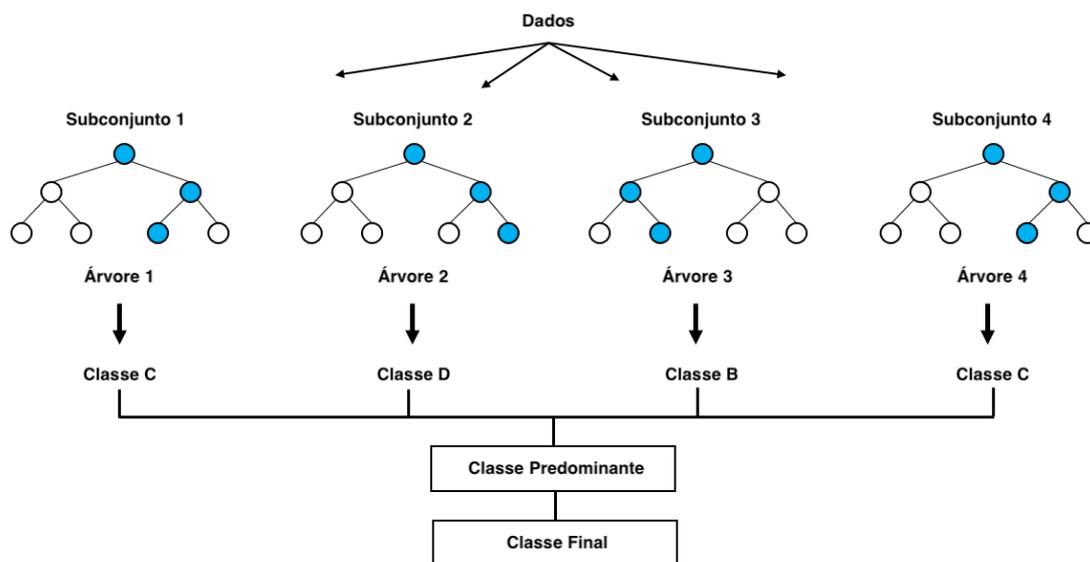


Figura 2.3: Exemplo de uma RF.

As RF também apresentam as suas vantagens e desvantagens. São capazes de lidar com um grande *dataset*, onde a presença de *outliers* e "ruído" nos dados não influencia o resultado final. Tal como as DT, a utilização deste tipo de modelos permite entender quais as variáveis que são mais importantes para o estudo, podendo também ser utilizado em problemas de classificação e regressão. No entanto, apesar de normalmente obter melhores resultados que as DT, devido à sua maior complexidade, demoram mais tempo no processo de treino [40].

2.2 Controlo Analítico numa ETAR

Gerir uma ETAR é um processo exaustivo e complexo, pois é dependente de fatores incontrolláveis como as condições climáticas ou descargas ilegais (ligações ilegais à rede de saneamento). Todos estes fatores causam variações no caudal e características do afluente na ETAR, sendo necessário um tratamento mais robusto ao afluente [41]. Pode-se equiparar o funcionamento de uma ETAR com a indústria, cujo o principal objetivo é controlar todos os processos de forma a minimizar ao máximo os impactos ambientais, assim como os custos de exploração. Neste sentido, torna-se necessário, diariamente, analisar regularmente os vários parâmetros que caracterizam o afluente de modo a controlar o processo e eventualmente proceder-se a ajustes no equipamento.

Para o funcionamento de uma ETAR ser monitorizado é necessário que seja realizada a análise de vários parâmetros em pontos estratégicos, desde a entrada do afluente até à saída do efluente final. Isso é obtido através da realização do controlo analítico, que visa a avaliação do cumprimento das exigências de qualidade à descarga de efluentes tratados. O controlo analítico permite que se possam comparar os valores do afluente na obra de entrada com o efluente final de modo a verificar a percentagem de remoção de cada parâmetro presente no afluente e apurar se as concentrações no efluente final cumprem a legislação em vigor, assim como determinar a eficiência do tratamento na sua globalidade, tal como em cada unidade individual [42].

Praticar ensaios de controlo analítico no campo (controlo expedito), levados ao efeito na própria ETAR, possibilita obter uma ideia bastante aceitável e rápida das características do afluente, durante o processo de tratamento. O controlo expedito na ETAR é crucial pois possibilita a deteção das alterações nas características do afluente, em pleno terreno [43]. Além disso, permite que sejam tomadas as devidas ações para garantir que, no fim do tratamento, as características do efluente cumprem com os requisitos de qualidade impostos na licença de descarga, sem ser necessário aguardar pelos resultados obtidos em laboratório, que são sempre mais lentos.

As AR que afluem às ETAR apresentam variadas características, dependendo da sua proveniência:

- **Águas Residuais Domésticas:** provêm de instalações sanitárias, cozinhas e zonas de lavagem de roupas, caracterizando-se por conterem quantidades elevadas de matéria orgânica, serem facilmente biodegradáveis e manterem uma relativa constância das suas características no tempo [44].
- **Águas Residuais Industriais:** derivam da atividade industrial e caracterizam-se pela diversidade dos compostos físicos e químicos que contêm, dependendo do tipo de processamento industrial. Apresentam, em geral, grande variedade nas suas características ao longo do tempo [44].
- **Águas Residuais Pluviais:** resultam da precipitação atmosférica caída diretamente no local ou em bacias limítrofes contribuintes. Apresentam geralmente menores quantidades de matéria poluente, particularmente de origem orgânica[45].

O afluente que entra na ETAR pode conter diversas variações das suas características, o que pode influenciar a eficácia do tratamento. As variações de carga poluente podem ser sazonais, consoante as variações da população e as variações climáticas, podendo também ser diárias [46]. Para garantir a eficácia do tratamento, é aconselhado nestes casos, que seja efetuado um controlo expedito na ETAR de modo que a monitorização do tratamento do afluente seja mais frequente, salvaguardando a qualidade final exigida do efluente [43].

Por vezes, quando não são respeitadas por parte do setor industrial os limites máximos de descargas diárias, o funcionamento normal da ETAR é inevitavelmente alterado. São várias as consequências, desde da paragem dos equipamentos, devido à elevada afluência, até à ineficácia do tratamento ao afluente. Isto implica elevadas consequências ambientais, pois uma vez que o tratamento do afluente não é realizado nas condições normais, iram ser poluídos os recursos hídricos onde a descarga é efetuada [47]. De notar que, as AR industriais não são semelhantes às AR domésticas e devem ser pré-tratadas no próprio estabelecimento industrial antes de serem lançadas na rede de drenagem urbana, para assim ser assegurado o tratamento eficaz na ETAR [48].

Nem todas as ETAR têm os mesmos requisitos de qualidade do efluente na licença de descarga, onde consoante a zona em que seja efetuada a descarga, os requisitos podem ser de maior ou menor exigência. Para se classificar as zonas de descarga, estas estão divididas em duas categorias, como descrito no Decreto-Lei n.º 236/98 [48]:

- **Zonas Sensíveis:** Lagos naturais e outras extensões de água doce, estuários e águas costeiras que se revelem eutróficos ou suscetíveis de se tornarem eutróficos num futuro próximo, se não forem tomadas medidas de proteção. São também consideradas como zonas sensíveis as águas doces de superfície, destinadas à captação de água potável.
- **Zonas menos Sensíveis:** Uma extensão ou zona de água marinha pode ser identificada como uma zona menos sensível se a descarga de AR não deteriorar o ambiente devido à morfologia, à hidrologia ou às condições hidráulicas específicas existentes nessa zona. Nestas zonas tem-se sempre em consideração o risco de a descarga poder ser transferidas para zonas adjacentes, onde pode desempenhar efeitos nocivos ao meio ambiente.

2.2.1 Data Science para Controlo Analítico numa ETAR

Nas últimas duas décadas, as tecnologias da computação tornaram-se imprescindíveis em todo o mundo e em praticamente todas as ferramentas usadas. Na indústria das ETAR, resultou num aumento exponencial na quantidade de dados que podem ser armazenados, tanto dos equipamentos usados, como das características dos processos físicos. No entanto, a inexistência de um meio eficaz de processar e visualizar os dados, levou a que melhorias significativas não surtam grande efeito, pois a sua exploração não está a ser utilizada e todo o potencial poderá descartado [49].

Para todos os gestores, engenheiros e operadores das ETAR, a capacidade de filtrar dados de uma forma rápida e inteligente, de forma a armazenar informações adicionais, é fundamental para eficácia e otimização dos processos de tratamento realizados nestas instalações. Na maioria dos casos, o problema não se remete ao acesso aos dados, mas sim à incapacidade de entender as suas origens, pontos fortes e limitações [49]. Em muitos casos, os dados não são inspecionados e interpretados de uma forma mais profunda até que surjam problemas, mas essa abordagem pode levar a análises mais desleixadas e conclusões impróprias. Para tirar o melhor proveito dos dados, as ETAR devem avaliar regularmente a qualidade e as tendências destes durante as operações normais de tratamento. Dados esses, que são, sobretudo provenientes das análises laboratoriais e dos diversos sensores presentes em cada etapa de tratamento [50].

É nesta fase que a aplicação de *Data Science* no controlo analítico é uma mais valia. Permite que seja realizado um estudo extremamente completo a todos os dados relativos ao controlo analítico do afluente em todos os processos de tratamento. Com esse estudo, ficam expostas todas as características de cada tratamento, sendo depois perceptível aos especialistas químicos e operadores da ETAR perceber onde se podem efetuar as devidas alterações para uma melhoria na qualidade da água. O uso de técnicas avançadas de visualização de dados, permite que as ETAR se concentrem em aspectos críticos relacionados com as condições de entrada do afluente e saída do efluente, ajustando-se os processos de tratamento em resposta à natureza em constante mudança do fluxo de resíduo [51].

2.2.2 Substâncias Relevantes

As AR consistem numa complexa mistura de substâncias dissolvidas e em suspensão, contendo inúmeros microrganismos de diversos tipos. Definir a caracterização qualitativa das AR é essencial para se conceptualizar a linha de tratamento das ETAR, consoante as características que as AR afluentes apresentam. Muitas das características da água são avaliadas somente por um parâmetro, como por exemplo, a temperatura que representa uma característica de elevada importância ambiental, onde é avaliada por um único parâmetro, denominado de Temperatura. Já outras características podem ser avaliadas por mais que um parâmetro, como é o caso da matéria orgânica na água que pode ser avaliada por *Carência Bioquímica de Oxigénio (CBO)*, *Carbono Orgânico Total (COT)*, entre outros [46].

As características das AR podem ser classificadas em três grupos:

- **Físicas:** Como principais características físicas, as AR contém:
 - **Sólidos:** Os ST que estão presentes nas AR são das principais e mais importantes características físicas, compreendem as substâncias orgânicas e inorgânicas presentes. Os ST calculam-se pela pesagem do resíduo seco, obtido após a evaporação da água na amostra realizada. Os sólidos são uma parte essencial para a realização do controlo de processos de tratamento físicos e biológicos nas AR, assim como para avaliar a conformidade com a

legislação. Por norma, parte dos **ST**, com natureza orgânica, são designados por **Sólidos Voláteis Totais (SVT)**, devido ao facto de se volatilizarem após a ignição ou calcinação, quando submetidos a temperaturas elevadas. O resíduo seco é relativo à parte inorgânica dos sólidos presentes nas águas, sendo designados por **Sólidos Fixos Totais (SFT)**. A determinação destes dois compostos (**SVT** e **SFT**) não permite que seja distinguida a matéria orgânica da inorgânica com precisão, devido ao facto da calcinação conduzir à perda de alguns compostos inorgânicos. Os **Sólidos Sedimentáveis (SS_{sed})** são uma característica muito relevante no tratamento das **AR**, pois permitem estimar a quantidade de lamas que foram produzidas. Existem também os **SST** que são obtidos através da filtração das **AR** através de um filtro ou por centrifugação. Os sólidos suspensos podem ser classificados em **Sólidos Suspensos Fixos (SSF)**, que reflete a porção de sólidos (totais, suspensos ou dissolvidos) que resta após a ignição ou calcinação, e em **SSV**, que representa a porção de sólidos (totais, suspensos ou dissolvidos) que se perdem após a ignição ou calcinação [46].

- **Turvação:** É uma característica física da água que expressa a dificuldade da penetração da luz na água. A turvação acontece devido à presença de partículas em suspensão de uma dimensão muito reduzida, que absorvem os raios incidentes e efetuam a dispersão dos mesmos. As partículas, por norma, são provenientes de diferentes naturezas, desde microrganismos até partículas minerais, como composto de ferro e fibras [46].
 - **Cor:** Esta característica é definida como a coloração com que a água fica após a remoção, por filtração, das partículas que estão em suspensão. A cor apresentada pela água antes do processo de filtração é designado por cor aparente [46].
 - **Odor:** O odor que as **AR** apresentam é resultante de algumas substâncias dissolvidas na água, provenientes da utilização da água em múltiplas atividades, assim como da decomposição da matéria orgânica que se desenvolve nas próprias **AR** [46].
 - **Temperatura:** É uma característica que é influenciada pela sua localização e também pelas descargas de águas quentes na **ETAR**. A temperatura das **AR** não é coincidente com a temperatura ambiente, estando no inverno geralmente, acima da média 2°C. A temperatura é uma característica muito importante, pois além de afetar outras características da água, influencia a velocidade dos processos bioquímicos utilizados no tratamento das **AR** [46].
 - **Densidade:** A densidade de uma determinada substância caracteriza-se como a razão entre a massa de um determinado volume da substância e a massa de igual volume de água pura a uma temperatura de 4°C [46].
 - **Viscosidade:** É uma característica física dos fluidos que reflete a facilidade com que o fluido escoar. A viscosidade é diretamente influenciada com a temperatura, sendo que quanto maior for o valor da temperatura menor será a viscosidade [46].
- **Químicas:** As **AR** têm elevadas concentrações de diversas substâncias químicas. Anualmente,

com a introdução de novas substâncias no mercado, esse número acaba por aumentar pois essas novas substâncias quando utilizadas acabam sempre por atingir as AR. Contudo, as principais características químicas das AR são:

- **pH:** É o parâmetro que caracteriza o nível de acidez na água em, ácido, neutro ou básico. Os valores de pH variam entre 0 e 14, onde 7 representa o valor da neutralidade. As soluções básicas contêm valores de pH superiores a 7, enquanto as soluções ácidas têm valores abaixo de 7 [46].
- **Potencial Redox:** Este parâmetro avalia a capacidade de uma certa substância perder ou ganhar elétrons, isto é, de ser reduzida ou oxidada. O potencial Redox é inteiramente dependente do valor de pH da água, da atividade oxidada e reduzida e da temperatura [46].
- **Alcalinidade:** É a característica capaz de refletir a capacidade da água neutralizar os ácidos. É um parâmetro de elevada importância no tratamento das AR, pois constitui um fator condicionante em alguns processos de tratamento, como nos tratamentos químicos, remoção biológica de nutrientes e na remoção de amoníaco por extração na fase gasosa [46].
- **Cloretos:** Os cloretos presentes nas AR são provenientes das atividades humanas (excretos humanos), industriais, agrícolas e em águas de infiltração, onde ocorra intrusão salina. É um parâmetro que, quando está presente em elevadas quantidades nas AR, pode comprometer a reutilização das AR tratadas [46].
- **Oxigênio dissolvido:** Esta característica representa o valor do teor de moléculas de oxigênio dissolvidas na água. É um parâmetro muito importante na gestão ambiental da água, pois trata-se de um indicador relativo ao estado de poluição da água. Água não poluída apresenta valores de oxigênio dissolvido em elevada quantidade, ao invés da água poluída, como as AR, que praticamente não contêm valores de oxigênio dissolvido [46].
- **Matéria orgânica:** A matéria orgânica presente nas AR podem ser biodegradáveis ou não biodegradáveis. A biodegradação é o processo que leva a transformação dos compostos orgânicos em composto minerais ou inorgânicos, por ação de seres vivos, mais concretamente de microorganismos. São diversos os compostos orgânicos presentes em AR, como por exemplo: a borracha, plásticos, antibióticos, fertilizantes agrícolas, pesticidas, açúcares, corantes, vitaminas, entre outros. Calcular a matéria orgânica presente nas AR é muito importante, e como a indústria está constantemente a introduzir novas substâncias no mercado, torna-se impossível analisar todos os compostos orgânicos. Para isso, a avaliação da matéria orgânica faz-se de uma forma indireta e agrupada, utilizando-se os seguintes parâmetros: CBO, Carência Química de Oxigênio (CQO) e COT. A CBO é definida como a quantidade de oxigênio utilizado na oxidação bioquímica da matéria orgânica na água. Este parâmetro representa uma avaliação indireta da matéria orgânica biodegradável na água (proteínas, hidratos de carbono) em que o que é medido experimentalmente é o consumo de oxigênio,

sendo proporcional à quantidade de matéria orgânica que foi bioquimicamente oxidada. A **CQO** é o parâmetro que avalia a matéria orgânica presente na água, onde traduz o oxigênio necessário para a oxidação química dessa mesma matéria orgânica. O **COT** permite medir a concentração total de matéria orgânica presente na água [46].

- **Nutrientes:** As **AR** contêm nutrientes necessários para o desenvolvimento das plantas, que ao não ser efetuada a sua remoção, irão potenciar a eutrofização da água, através das descargas efetuadas pela **ETAR** nos meios recetores aquáticos. Este fenómeno ocorre quando é feita a excessiva proliferação de plantas macroscópicas e microscópicas, levando ao desequilíbrio do meio recetor. O Azoto e o Fósforo são os principais nutrientes, pois estão em quantidades superiores em relação aos outros, fazendo parte dos 16 elementos indispensáveis ao desenvolvimento das plantas. Atualmente, a preservação da qualidade das massas de água tem como um dos principais objetivos a prevenção ou remediação da sua eutrofização. Essa prevenção é conseguida através da remoção do Azoto e Fosforo dos afluentes, conseguida através do tratamento terciário realizado nas **ETAR**, onde a descarga do efluente de situa numa zona classificada como zona sensível. O Azoto é um nutriente vital devido a ser um componente fundamental nas estruturas tecidulares proteicas. A presença do Azoto na água, normalmente é indicador de poluição. O Fosforo é outro elemento importante, pois também pode originar a eutrofização das águas, devendo ser controlada a sua presença, tal como a do Azoto. A sua presença em elevadas concentrações nas **AR** provém de detergentes e dos compostos orgânicos proteicos [46].
- **Enxofre e compostos sulfurados:** O Enxofre pertence aos aminoácidos das proteínas, sendo portanto um dos constituintes do protoplasma dos seres vivos. A sua presença na água provem dos microrganismos, animais e plantas. O Enxofre pode também se encontrar na água sob diferentes formas inorgânicas: sulfatos, tiosulfatos, sulfitos e sulfuretos. A presença destes compostos sulfurados nas **AR** são muito prejudiciais, pois atribuem às **AR** características muito particulares, que se não forem removidas, irão ser altamente prejudiciais à vida aquática e até para os próprios equipamentos de tratamento nas **ETAR**. Por exemplo, os sulfuretos conferem à água uma cor bastante escura e promove a formação de maus cheiros, além de tornar a água tóxica e corrosiva, podendo danificar assim os equipamentos presentes na **ETAR** [46].
- **Óleos e gorduras:** Esta característica quando se encontra em elevadas quantidades é suscetível de provocar elevados problemas nos órgãos de tratamento das **ETAR** e nas fossas sépticas. Podem provar o entupimento das canalizações e a ineficácia do tratamento biológica ao interferir com a atividade microbiana, devido a provocarem o revestimento de paredes e superfícies das águas, com a obstrução ao arejamento. Os óleos e as gorduras têm uma biodegradabilidade bastante reduzida, portanto a sua remoção das **AR** é bastante importante. Provém sobretudo das oficinas mecânicas, e das habitações, consequência das lavagens da

louça que contém restos como, manteigas, óleos alimentares, gorduras, etc [46].

- **Substâncias tensoativas:** As substâncias tensoativas são sobretudo detergentes usados no meio doméstico e industrial. Estes quando presentes nas AR, provocam presença de espumas superficiais, nas ETAR e nos meios recetores. Provocam consequências negativas pois dificultam a solução do oxigénio na água, contrariando a sedimentação dos flocos biológicos nos reatores de lamas ativadas [46].
- **Metais pesados:** Os metais pesados, também podem incluir metais e semi-metais, representam um potencial de toxicidade ecológica e um risco para a saúde, sendo provenientes sobretudo de origem industrial. Os metais que representam maior risco para a saúde pública são: chumbo, bário, mercúrio, selénio, prata, berílio, cádmio, crómio e o alumínio quando presente em grandes quantidades. Já o cobre é muito prejudicial para a vida aquática. Por isso, é de extrema importância remover estes metais das AR antes de ser efetuada a descarga por parte da ETAR [46].
- **Compostos orgânicos perigosos:** Esta característica é das que requer maior atenção e atenção por parte das autoridades. Muitos destes compostos não são biodegradáveis e muito tóxicos, podendo causar inúmeras complicações quando ingeridos. São provenientes das pesticidas utilizadas na agricultura, exterminação de roedores, tratamento de madeiras, entre outros. Torna-se imprescindível que seja feita a sua remoção nas fases de tratamento das ETAR, quando presente nas AR [46].
- **Biológicas ou Microbiológicas:** As características biológicas ou microbiológicas das AR tem como objetivo a determinação do grau de contaminação por microorganismos patogénicos, assim como a possibilidade de transmissão de doenças. Os microrganismos estão em elevadas quantidades nas AR, sendo alguns deles patogénicos (que geram doenças). Estes microrganismos desempenham um papel fulcral nos processos de depuração biológica das AR, pois são agentes da oxidação bioquímica dos poluentes orgânicos biodegradáveis [46].
 - **Microrganismos patogénicos:** Estes microrganismos tem origem nas fezes e urina de hospedeiros humanos, dos animais provenientes das indústrias agropecuárias, dos matadouros e AR pluviais em sistemas de drenagem unitários. A tipologia e quantidade de microrganismos patogénicos nas AR varia de população para população, alterando-se ao longo dos meses e até do próprio dia no mesmo aglomerado. Estes microrganismos patogénicos podem ser classificados nos seguintes grupos: bactérias, protozoários, vírus e helmintas [46].
 - **Toxicidade:** Define-se como a capacidade de uma substância ou mistura causar danos no sistema biológico dos seres vivos, alterando o normal funcionamento do seu metabolismo. A toxicidade varia consoante a localização da fonte de contaminação, quantidade de substâncias tóxicas e duração de exposição ao tóxico. Esta característica pode ser classificada como

imediate ou retardada, conforme o tempo decorrido entre a exposição ao tóxico e o primeiro sintoma consequente dessa mesma exposição [46].

2.2.3 Valores Padrão

Tal como supracitado, é necessário que o tratamento das AR cumpra com vários requisitos de qualidade, antes de se proceder à descarga para o meio ambiente. De modo a salvaguardar alterações negativas que as mesmas podem impor caso não seja realizado o devido tratamento, o efluente tem de cumprir com os requisitos de qualidade impostos. O decreto lei nº236/98, anexo XVIII [52], define os Valores Limite de Emissão (VLE) na descarga de AR, como mostra a Tabela 2.1:

Tabela 2.1: Tabela dos Valores Limite de Emissão em descargas de Águas Residuais

Parâmetros	Expressão dos resultados	VLE
pH	Escala de Sorensen	6,0 - 9,0
Temperatura	°C	Aumento de 3°C
Carência Bioquímica de Oxigénio após 5 dias (CBO ₅)	mg/l O ₂	40
CQO	mg/l O ₂	150
SST	mg/l	60
Alumínio	mg/l Al	10
Ferro total	mg/l Fe	2,0
Manganés total	mg/l Mn	2,0
Cheiro	—	Não detetável na diluição
Cor	—	Não detetável na diluição
Cloro residual disponível Livre	mg/l Cl ₂	0,5
Cloro residual disponível Total	mg/l Cl ₂	1,0
Fenóis	mg/l C ₆ H ₅ OH	0,5
Óleo e gorduras	mg/l	15
Sulfuretos	mg/l S	1,0
Sulfitos	mg/l SO ₃	1,0
Sulfatos	mg/l SO ₄	2000
Fósforo total	mg/l P	10 (valor normal) 3 em águas que alimentem lagoas ou albufeiras 0,5 em lagoas ou albufeiras
Azoto amoniacal	mg/l NH ₄	10
Azoto total	mg/l N	15
Nitratos	mg/l NO ₃	50
Aldeídos	mg/l	1,0
Arsénio total	mg/l As	1,0
Chumbo total	mg/l Pb	1,0
Cádmio total	mg/l Cd	0,2
Crómio total	mg/l Cr	2,0
Crómio hexavalente	mg/l Cr	0,1
Cobre total	mg/l Cu	1,0
Níquel total	mg/l Ni	2,0
Merúrio total	mg/l Hg	0,05
Cianetos totais	mg/l CN	0,5
Óleos minerais	mg/l	15
Detergentes	mg/l	2,0

2.3 Eficiência Energética numa ETAR

Na indústria das ETAR, alcançar um elevado nível de eficiência energética tornou-se um tópico cada vez mais importante na gestão da mesma. Reduzir os consumos energéticos, emissão dos gases com efeito de

estufa e custos operacionais tem sido das principais preocupações dos gestores das **ETAR**, que têm vindo a adotar a escolha de equipamentos e tecnologias mais eficientes, aliado a sistemas de recuperação de energia e a uma gestão de custos mais eficiente. É importante que uma **ETAR** não se foque exclusivamente no único objetivo de tratar a água residual, mas também incluir a gestão eficiente de todos os recursos que utiliza [53]. Isso é conseguido através da eficiência energética, onde um dos principais fatores é a realização de um controlo efetivo dos custos, resultando num acompanhamento e análise contínua de todos os sistemas. A eficiência energética retorna inúmeros benefícios a vários níveis, ambiental (com a redução da pressão atmosférica e das emissões de gases com efeito de estufa), económico (crescimento económico através da criação de novos postos de emprego e desenvolvimento do mercado que envolve a área de eficiência energética), financeiro (redução dos custos com a energia), e por fim, a nível operacional (aumento da vida útil de equipamentos e infraestruturas e melhoria da qualidade da água) [54].

Para se conseguir obter uma eficiência energética e consequentemente usufruir dos seus benefícios, é necessário identificar onde podem ser efetuadas melhorias de desempenho em todos os processos existentes na **ETAR**. Para ser possível identificar essas lacunas, é fundamental que seja executada uma auditoria energética [55], onde a utilização de *Data Science* neste processo pode ter uma enorme relevância, podendo ser ilustrados aspetos fulcrais que permitam a redução dos consumos energéticos, que uma auditoria normal não seria capaz. As auditorias energéticas, resumidamente, são um exame detalhado às condições da utilização da energia nas infraestruturas, com o objetivo de implementar uma utilização mais económica e eficiente da energia, sem afetar os resultados de produção, sendo no caso das **ETAR** a eficácia do tratamento às **AR**. A realização de uma auditoria é dependente do seu âmbito e da dimensão da infraestrutura a auditar, mas, por norma, uma auditoria energética engloba as seguintes etapas [56]:

- Realização de um levantamento em termos energéticos de todas as formas de energias consumidas, assim como as necessidades energéticas específicas de utilização e de cada divisão da empresa.
- Análise de todos os registos históricos dos consumos das diversas formas de energia envolvidas, com discriminação dos fluxos energéticos.
- Elaboração de balanços energéticos e avaliação da desagregação dos consumos.
- Identificação das principais áreas de intervenção, com vista na racionalização dos consumos energéticos e diminuição dos custos.

Estas auditorias não desempenham um custo acrescido sobre os custos relativos ao tratamento das **AR**, nem a implementação de medidas para se obter eficiência energética requer forçosamente investimentos elevados. É possível atingir bons resultados elaborando apenas alguns ajustes operacionais ou reduzidos investimentos para um maior controlo e automação em todos os processos de tratamento [55].

Em alguns casos, as **ETAR** são dimensionadas para receberem um determinado caudal de **AR**, o que acontece é que esse caudal, por vezes, nunca chega a ser como o dimensionado. Com isso, a **ETAR** fica

a funcionar pela metade da sua capacidade, com equipamentos sobredimensionados resultando num exagerado consumo de energia, e com isso, valores muito longínquos para se atingir níveis de eficiência adequados [57].

Contudo, apesar das medidas da eficiência energética terem imenso potencial, é necessário que haja um acompanhamento das mesmas, caso a caso. Existem muitos fatores que podem influenciar a eficiência energética como, por exemplo, o tamanho da ETAR, as características do aflente (quando a carga orgânica é muito elevada, requer um tratamento mais robusto), as soluções tecnológicas utilizadas e a sazonalidade das descargas para a ETAR [55].

2.3.1 Data Science para Eficiência Energética numa ETAR

Atualmente com o aumento crescente do consumo energético por parte da população, principalmente na forma de eletricidade em residências e indústria, as preocupações no setor da energia têm gerado cada vez mais preocupação. Apesar de, nos últimos anos, haver uma maior consciencialização sobre o impacto das fontes de energia não renováveis no planeta e a elevada emissão de gases com efeito de estufa, se não forem aplicadas medidas concretas e imperativas, este problema só tende em piorar [58].

Dessa forma, como supramencionado, o termo eficiência energética tem vindo, ao longo dos anos, a ganhar cada vez mais importância tornando-se imprescindível na indústria e em novas habitações. A eficiência energética, pode ajudar na redução da produção de energia e conseqüentemente, reduzir as emissões de gases com efeito de estufa e na preservação dos recursos de combustíveis fósseis, assegurando um contributo notório para a redução dos problemas ambientais no planeta.

Com o avanço da tecnologia, os equipamentos industriais já vêm integrados com diversos sensores que permitem medir os consumos energéticos de uma forma simples e eficaz, sendo que no caso de equipamentos mais antigos pode-se utilizar sensores específicos nos equipamentos para obter os mesmos dados. Isto permite que possam ser armazenados dados de diferentes variáveis relativas ao consumo de energia por parte de cada equipamento.

Quanto maior a complexidade do processo, maior será a dificuldade na tomada de decisão por parte dos responsáveis da instituição, no que diz respeito a medidas que visam a redução dos consumos energéticos [59]. Portanto, quando todos os dados que se podem obter de uma instituição não são devidamente analisados ou usados, as mesmas ficam privadas de conhecer detalhadamente as despesas energéticas mensais, verificar a evolução ao longo do tempo e de identificar as ações que podem ser adotadas para minimizar os custos energéticos.

Aqui entra a *Data Science* que permite obter o máximo partido dos dados que se dispõem. Como supracitado, é realizado um estudo absoluto a todos os dados relativos aos consumos energéticos, para fornecer *insights* dos dados que possam auxiliar no processo de tomada de decisão, para alcançar uma melhor eficiência energética nestas instalações [60]. Este procedimento tem como principal objetivo considerar todas as possíveis variáveis relacionadas com o consumo de recursos energéticos durante uma operação e ilustrar, de forma objetiva, os indicadores que visem explicar quais os equipamentos ou

fatores externos que estão mais relacionados com o gasto de recursos durante a operação [61]. No caso das ETAR, o uso de *Data Science* iria permitir, aos gestores e operadores, verificar toda a informação sobre os gastos de energia por parte de todos os equipamentos e quais as variáveis que influenciam mais a oscilação do consumo energético no funcionamento da ETAR. Assim, será possível aos gestores das ETAR identificarem onde podem ser feitas alterações em todo o esquema de funcionamento da ETAR, de modo a garantir uma eficiência energética cada vez maior, salvaguardando sempre os requisitos de tratamento das AR.

2.3.2 Fatores que Afetam o Consumo Energético numa ETAR

O consumo energético das ETAR varia substancialmente, dependendo do tipo de tratamento que se tem de aplicar ao afluente, assim como as técnicas e tecnologias de tratamento. Quanto maior for a carga orgânica, maior serão os consumos energéticos, pois são necessários tratamentos mais exigentes para que, no final, o efluente cumpra com as normas de descargas exigidas para o meio ambiente. Outro fator que pode influenciar os consumos energéticos é o fluxo dos caudais, onde quanto maior for esse fluxo mais esforço requererá aos equipamentos e, conseqüentemente, mais energia eléctrica será utilizada [53]. De modo a ter-se uma melhor visão de como é a variação normal do fluxo caudal das AR nas ETAR, na Figura 2.4 é ilustrada essa mesma variação:

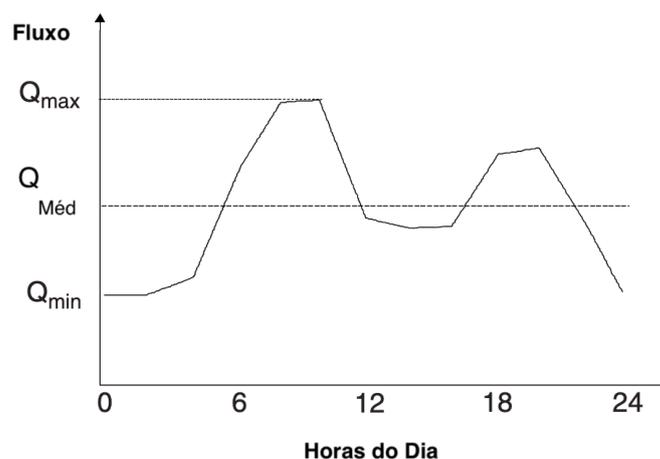


Figura 2.4: Variação do fluxo diário do caudal de águas residuais

Como se pode verificar, o fluxo mínimo do caudal acontece nas primeiras horas do dia, correspondente ao baixo consumo de água da população. O primeiro pico ocorre ao final da manhã, onde condiz com a chegada da água à ETAR do pico matinal de água utilizada. Por norma, o segundo pico acontece ao final na tarde, entre as 18h e as 20h [62].

Atualmente, encontra-se na maioria das ETAR baixos níveis de desempenho a nível de eficiência energética. São diversos os fatores que influenciam os consumos energéticos numa ETAR, dependendo muito das características da mesma e respetivamente dos tipos de tratamento utilizados, mas num modo geral um baixo nível de eficiência energética deve-se sobretudo a [63]:

- Crescente necessidade de ser realizada a reciclagem da água, graças à crescente escassez deste recurso;
- Tipos de motores e bombas que são utilizados;
- Exigência crescente dos parâmetros de descarga dos efluentes tratados.
- Etapas de desinfecção e infiltração são energeticamente veementes.
- Processos de bombagem das águas requerem elevados consumos energéticos, devido às características do terreno ou configurações da ETAR.
- Ausência de mecanismos de recuperação energética.
- Eficiência baixa nas operações, mistura e sistemas de arejamento.

De forma a corrigir alguns destes problemas, existem diversas tecnologias e oportunidades a nível de controlo, monitorização, eficiência, entre outros, como forma de aumentar a eficiência energética e consequentemente diminuir os consumos energéticos das ETAR. Algumas destas medidas, passam por exemplo, pela substituição de equipamentos por outros mais eficientes, garantindo que os mesmos realizam o mesmo trabalho mas utilizem menos energia na sua operação [64].

2.4 Revisão da Literatura

Nesta secção é feita a análise de várias soluções e trabalhos de investigação existentes na literatura, de modo a examinar as diferentes abordagens e tecnologias que foram estudadas neste contexto. O objetivo desta análise não é apenas identificar alguns dos pontos mais relevantes nas diferentes soluções existentes na literatura, mas também refletir sobre os aspetos mais importantes no domínio desta dissertação. Para cada artigo é realizado um pequeno resumo, exposto os objetivos a que se propõem, os dados em que trabalha, tipos de algoritmos/modelos utilizados e os resultados obtidos. Na subsecção 2.4.1 são revistos artigos direcionados ao controlo analítico e na subsecção 2.4.2 direcionados à eficiência energética.

2.4.1 Controlo Analítico

Newhart et al. [65] realizaram uma análise de desempenho baseado nos dados de algumas ETAR dos Estados Unidos, com o objetivo de reduzir os custos, assim como otimizar as operações. Começa por ser explicado as características dos dados relativos às ETAR, descrevendo a sua estrutura, frequência e variação temporal e as características das variáveis de estudo. Seguidamente, os autores referenciam a abordagem de controlo orientado por dados (controlo avançado ou automatizado) que, usando dados históricos e conhecimento do sistema da ETAR, uma função pode ser desenvolvida para minimizar custos ou energia enquanto mantém a qualidade do efluente. Com isso, pode-se identificar o melhor conjunto de pontos de ajuste e decisões de controlo. Segundo os autores, esta é uma abordagem fundamentalmente diferente de

ajustar heurísticamente os pontos de ajuste das variáveis e observar a resposta do sistema. Neste estudo, foram utilizados diferentes modelos como, modelo de controlo preditivo (*Model predictive control*), redes neuronais (*neural networks*), modelos de função de transferência (*Transfer function models*) e lógica difusa (fuzzy logic). O modelo de controlo preditivo foi implementado em processos simplistas nas ETAR, que podem ser controlados com uma única variável. Os resultados foram, no geral, melhores quando comparados com um controlador tradicional. Já as redes neuronais foram utilizadas para modelar a concentração de oxigénio dissolvido e nitratos, onde os autores não compararam os resultados com o desempenho de um controlador convencional, devido ao facto de ser necessário requisitos computacionais para controlo em tempo real num sistema maior. Os modelos de função de transferência apenas foi descrita a possibilidade dos mesmos serem utilizados nos modelos de controlo preditivo e otimização, obtendo-se melhorias significativas. Por fim, para a lógica difusa são descritos vários exemplos da sua implementação no contexto real, ilustrando os diferentes resultados que podem ser obtidos com a sua utilização, assim como vantagens da sua utilização quando comparado com o uso de outras abordagens.

Noutro estudo levado a cabo por Gonçalves et al. [66] foi realizada uma análise estatística na gestão de nove ETAR. O estudo focou-se na análise de quatro variáveis físico-químicas (SSV, SST, CQO e CBO₅), duas variáveis hidro-meteorológicas (precipitação pluviométrica e número de dias, por mês, de precipitação pluviométrica) e também teve em consideração o valor do volume de afluentes de AR nas ETAR. Inicialmente os autores efetuaram uma análise exploratória aos dados, com o principal objetivo de caracterizar e identificar a gestão de processos das ETAR, em particular processos biológicos usados no tratamento de afluentes. Para isso, foram desenvolvidos métodos estatísticos com o intuito de quantificar e caracterizar a variabilidade das variáveis de qualidade do afluente que entra nas ETAR, onde foi possível caracterizar e quantificar os padrões e correlações entre o volume dos afluente às ETAR, as variáveis físico-químicas, as variáveis hidro-meteorológicas e as variáveis associadas ao custo dos tratamentos. De salientar que a correlação entre as variáveis físico-químicas e as variáveis meteorológicas foi essencial para se conseguir obter conhecimento que suporte a minimização dos recursos utilizados nos sistemas de tratamento nas ETAR. Foram estabelecidos modelos de calibração e modelos lineares que visam prever e estimar os procedimentos de monitorização dinâmicos abrangidos nesses processos. Estes modelos adaptam-se aos comportamentos sazonais ao longo do ano, onde os processos de modelação foram realizados em dois períodos hidrológicos: estação seca e estação chuvosa. Como resultados, foi possível obter e perceber relações de dependência entre variáveis e efeitos que os técnicos das ETAR desconheciam, estando agora esses dados ao dispor dos mesmos para auxiliar no processo de tomada de decisão, permitindo-lhes avaliar e prever comportamentos em diferentes subsistemas das ETAR.

No estudo levado a cabo por Ribeiro et al. [67] é descrito uma metodologia intitulada de *Decentralized Monitoring and Control (DEMOCON)*, que consiste num sistema simples de apoio à tomada de decisão, especialmente direcionado a ETAR descentralizadas, onde é utilizada modelação matemática para controlo de gestão operacional nos processos de tratamento numa ETAR. Este estudo, teve como objetivo apoiar a tomada de decisão sobre as ações de controlo, estando organizada em três fases sequenciais associadas a etapas específicas do processo de aquisição e análise de informação, completas por uma

fase de validação das ações de controlo. Este sistema trabalha sobre diversos tipos de dados relativos às **ETAR** sendo eles classificados em três categorias: dados básicos (informações relativas às características do sistema), dados complementares (valores obtidos através de inspeções, reclamações e observações), e dados operacionais (dados relativos ao desempenho do sistema de tratamento, afluência de **AR** e registo de falhas). Este sistema monitoriza essencialmente os seguintes parâmetros em afluentes e efluentes: caudal, condutividade, oxigénio dissolvido, pH, turvação, azoto amoniacal, sólidos suspensos, carência química em oxigénio, nitratos, carbono orgânico e compostos de fósforo. Com esta monitorização é fornecido, aos responsáveis das **ETAR**, um aumento do conhecimento sobre as condições de funcionamento de cada fase de tratamento, assim como uma possibilidade de ser alertados para uma possível anomalia em qualquer sistema da instalação, devido à monitorização ser efetuada em modo contínuo. Com este estudo, os autores conseguiram concluir que com a implementação deste sistema, é obtido um aumento significativo da robustez e eficiência dos sistemas de tratamento das **ETAR**.

2.4.1.1 Análise Crítica sobre os Estudos do Controlo Analítico

Após ser realizada a leitura dos artigos acima descritos, verifica-se a existência de diversas abordagens que podem auxiliar nas tomadas de decisão sobre controlo analítico numa **ETAR**. No artigo [65] apesar de ser descrito a realização da análise e tratamento aos dados, não foi especificado como foi feito esses processos, que variáveis foram descartadas, as suas características e métodos utilizados. Também é exposto a utilização de diferentes modelos para a resolução do problema mas a informação sobre as variáveis utilizadas e os resultados é muito escassa, não ficando totalmente esclarecido a mais valia na utilização destes modelos, apesar de ser narrado a obtenção de resultados positivos. No segundo artigo estudado [66], tal como [65] foi realizada uma análise e tratamento aos dados, mas a informação sobre essa etapa é também bastante escassa, apesar dos autores referirem as variáveis que foram alvo de estudo. Nesse mesmo artigo, é descrita a utilização de métodos estatísticos para a resolução do problema, mas não são especificados quais os modelos utilizados, assim como, as suas características. No que toca aos resultados obtidos, não foram muito explícitos, sendo descrito apenas que a utilização destes métodos estatísticos provaram ser uma mais valia na gestão de uma **ETAR**, onde auxiliam no processo de tomada de decisão. Por fim, no artigo [67], tal como acontece em [65] e [66], o processo de análise e tratamento dos dados não é muito clarificado, faltando informação dos procedimentos e opções tomadas neste processo tão importante. Contudo, relativamente aos dados utilizados já é bastante esclarecedor, classificando os dados em diferentes grupos, caracterizando e descrevendo as suas principais características. Também é feita uma explicação muito pormenorizada do funcionamento do sistema criado. Quanto aos resultados obtidos, apenas é descrito, resumidamente, que a utilização deste sistema acarreta um aumento significativo da robustez e eficiência dos sistemas de tratamento das **ETAR**, faltando gráficos ou tabelas que ilustrem essas melhorias.

Analisa-se também que, em todos estes estudos, as variáveis utilizadas como *input* são de um modo geral reduzidas, podendo ter sido descartadas variáveis que podiam retornar, quando conjugadas com as

restantes, informações e posteriormente conclusões importantes para a tomada de decisão. É de notar que, nos estudos analisados, apenas um faz a análise dos padrões e correlações entre as variáveis do afluente, sendo um aspeto crucial para ilustrar que características provocam determinados comportamentos no tratamento do afluente. Também é importante referir que em nenhum dos estudos é descrito uma análise sazonal às características do afluente, o que permitia a perceção de comportamentos específicos, relacionados com a sazonalidade ou com os trimestres/semestres ao longo do ano.

2.4.2 Eficiência Energética

No estudo levado a cabo por Li et al. [68] foi realizada a análise e previsão dos consumos energéticos numa ETAR. É concluído que neste estudo, para se efetuar uma análise completa dos consumos energéticos numa ETAR, a melhor solução é utilizar a análise de agrupamento (clustering analysis), sendo usado então para esta abordagem o método Fuzzy C-Means clustering. São consideradas três variáveis principais como indicadores de desempenho, sendo elas: as cargas afluentes, CQO, e o azoto. Para a previsão dos consumos energéticos foi utilizado o método Radical Basis Function (RBF) Neural Network, com os autores a justificarem esta opção com a presença de muito fatores capazes de influenciar os consumos energéticos, e com isso os sistemas de energia sofrem alterações não lineares, tornando difícil um modelo convencional ter bom desempenho. Este modelo foi introduzido para prever os valores da variável que foram consideradas para o estudo. Como resultados, os autores conseguiram expor diversos padrões e correlações entre as variáveis de estudo, percebendo fatores que podem influenciar os seus valores e com isso aumentar o consumo energético, ficando mais explícito aos gestores das ETAR onde é consumida mais energia e o que provoca os picos de consumo, oferecendo-lhes referências para melhorar a eficiência energética. Para a previsão dos consumos energéticos, os autores implementaram o modelo Multiple Linear Regression (MLR) para comparar os resultados com o modelo RBF. Como conclusão, os autores verificaram que a abordagem utilizada com o modelo RBF retorna melhores resultados, quando comparados com o modelo MLR.

Longo et al. [69] realizaram um estudo onde é utilizada uma metodologia sistemática para a quantificação robusta da eficiência energética numa ETAR, utilizando Data Envelopment Analysis (DEA). É justificado o uso da metodologia DEA devido ao facto desta reunir enorme capacidade em lidar com múltiplas variáveis de entrada e saída, identificar relações eficientes de entrada e saída, identificar fontes e quantificar a ineficiência em cada uma das unidades comparadas. Primeiramente, os autores realizaram uma análise superficial das características dos dados e concluíram que a abordagem DEA é uma mais valia quando o estudo se remete a uma única ETAR ou em várias que partilhem o mesmo tipo de tratamentos, os fatores de entrada e saída que as caracterizam são idênticos e que funcionam sob o mesmo conjunto de condições ambientais, caso contrario, as conclusões retornadas ao utilizar esta metodologia podem não ser as mais corretas. Como este estudo se concentrou em várias ETAR de diferentes características é utilizado a metodologia Robust Energy Efficiency DEA, que envolve quatro fases de desenvolvimento: obtenção e preparação dos dados, seleção do modelo, estimativa da eficiência e por fim, refinamento

e validação do modelo. São consideradas como potenciais variáveis de *input* a eletricidade e produtos químicos e como *outputs* são consideradas a remoção de CQO, remoção de nitrogénio e remoção de fósforo. Foi feita uma elevada relevância às variáveis que representam os fatores exógenos (qualquer fator que não esteja sob o controlo direto da gestão da ETAR é exógeno ao sistema), onde foi descrito que estes fatores contém uma importância bastante significativa para a obtenção de resultados, pois sem o seu controlo, as estimativas da eficiências geradas pelo DEA serão potencialmente enviesadas. Como resultados, foi conseguida uma quantificação robusta da eficiência energética da ETAR, aumentando a qualidade das estimativas para futuros valores de consumos energéticos e exposto os fatores que mais influenciam os consumos energéticos nestas infraestruturas. Os autores concluíram também que, o não uso das variáveis que representam fatores exógenos pode levar a alterações das estimativas de eficiência energética na ordem dos 50%, sendo portanto extremamente importante a inclusão destes fatores.

Em outro estudo realizado por Iglesias et al. [70] foi descrita uma metodologia denominada de ENERWATER, que consiste num método padrão com o objetivo de avaliar e melhorar a eficiência energética nas ETAR. Esta metodologia pode ser aplicada de duas maneiras diferentes utilizando diferentes métodos: Rapid Audit (RA) e Decision Support (DS). O método RA leva a uma rápida estimativa dos índices de energia, utilizando dados históricos do uso de energia juntamente com os valores de qualidade de afluentes e efluentes obtidos por análises de rotina. O método DS requer uma monitorização intensiva do uso de energia e parâmetros de qualidade da água para fornecer um cálculo preciso e detalhado dos índices de energia. As variáveis utilizadas pelos autores foram: valores dos consumos das bombas e motores, remoção de SST, remoção de CQO, remoção de Nitrogénio, remoção de Fósforo e remoção de Sólidos. Como resultados, os autores concluíram que as estimativas obtidas com esta metodologia foram validadas com sucesso, sugerindo um alto nível de robustez das estimativas de eficiência energética, pois os autores utilizaram uma forma inovadora de mensurar essa eficiência energética ao desenvolver uma ferramenta de *benchmarking* e diagnosticar o uso de energia, formulando ações de melhoria com base em análises anteriores. O ENERWATER pode ser facilmente aplicada por operadores das ETAR, uma vez que requer a medição de parâmetros geralmente medidos nestas infraestruturas, e portanto é expectável que a aplicação desta metodologia facilite o processo de diagnóstico energético, pelo menos na fase inicial de identificação da ineficiência existente. É assim fornecido aos operadores das ETAR sugestões baseadas em casos para a se obter a eficiência energética na instalação. As principais contribuições da ENERWATER como metodologia padrão para a eficiência energética são: definição precisa das funções das ETAR por meio da identificação dos Key Performance Indicator (KPI) que refletem a eficiência operacional de cada função, capacidade para se adaptar a diferentes ETAR e consideração da possibilidade de ser produzida energia na própria ETAR.

No estudo levado a cabo por Sancho et al. [71] foi descrita uma abordagem *non-radial* DEA para a eficiência energética em 177 ETAR espanholas. A aplicação desta metodologia teve como objetivo numa primeira etapa calcular os índices de eficiência energética nas ETAR de estudo, sendo na segunda etapa analisadas as variáveis operacionais que contribuem para as diferenças de eficiência energética entre as ETAR estudadas. A utilização desta abordagem foi justificada pelos autores por ser capaz de fornecer

um indicador de eficiência para cada uma das variáveis do processo. A grande vantagem deste tipo de indicadores é que permite identificar em quais das variáveis de estudo é preciso atuar para aumentar a eficiência energética da instalação. Na segunda etapa de análise são utilizados dois métodos: *Regression Analysis*, utilizado *Tobit Regression* e *Ordinary Least Squares*, e também o método *Kruskal-Wallis non-parametric test*. Para a realização deste estudo foram utilizadas com *inputs* as variáveis: energia utilizada, reagentes, custos com pessoal e manutenção, e como *output* a remoção de SST e CQO. Foi também considerado o gasto de energia por quilo de remoção de CQO e o volume total de AR tratada num ano. Como resultados, os autores conseguiram verificar que foi possível obter o índice de eficiência global de cada instalação estudada, assim como os índices de consumo de cada uma. Concluíram que os índices de eficiência (neste estudo considera-se uma instalação eficiente quando a pontuação é igual a 1) nas ETAR estudadas são baixos, estando em média o valor desse índice em 0,288. Das 177 ETAR estudadas, apenas 10% são energeticamente eficientes, 69% apresentam um índice de eficiência entre 0 e 0,3, enquanto 21% apresentam pontuação entre 0,3 e 0,9. Os autores ilustraram também que, para cada ETAR, os valores de redução de energia e a percentagem de redução de emissão dos gases sob efeito de estufa que estas infraestruturas podiam obter caso operassem com um índice de eficiência 1. É concluído também pelos autores que a eficiência energética pode ser afetada pelo seguintes fatores: tamanho da instalação, idade da instalação, características do afluente, parâmetros de descarga exigidos e tipos de tratamentos utilizados.

2.4.2.1 Análise Crítica sobre os Estudos da Eficiência Energética

Concluída a revisão dos artigos acima descritos, demonstra-se a existência de diversas abordagens e metodologias que podem colaborar na detecção de fatores que estejam a afetar os consumos energéticos nas ETAR, refletindo baixos níveis de eficiência energética nas infraestruturas. No artigo [68] é clara a falta de informação relativa à fase de análise e tratamento dos dados que foi efetuado neste estudo, onde não consta qualquer informação sobre as opções e abordagens tomadas nesta fase que é considerada uma das mais importantes num projeto deste tipo. Apenas é descrito a proveniência dos dados utilizados neste estudo. Contudo é feita uma explicação completa do funcionamento dos algoritmos utilizados, descrevendo como os mesmos funcionam e como foram implementados. Relativamente aos resultados obtidos, são ilustrados de uma forma muito completa, utilizando tabelas e gráficos para a respetiva representação dos mesmos. No artigo [69] é realizada uma boa análise às características dos dados e uma forte explicação dos métodos utilizados, demonstrando o que foi efetuado em cada fase de desenvolvimento da metodologia. Contudo, tal como no artigo [68] não é feita qualquer referência à fase de tratamento aos dados, onde não existe qualquer tipo de demonstração das abordagens tomadas nessa fase de estudo. No que toca à ilustração dos resultados obtidos, foi demonstrando diferentes comparações dos resultados obtidos entre os modelos utilizados, onde é realizada uma análise muito completa. No artigo [70] é explicada de uma forma bastante completa as características da metodologia criada, expondo todas as suas características. Relativamente à análise e tratamento dos dados essa informação é descrita, mas

as opções tomadas nessa fase não são totalmente esclarecedoras, faltando informações relativamente às mesmas. Os resultados conseguidos são descritos com muito rigor e pormenor, expondo todas as informações e mais valias conseguidas com a utilização desta metodologia. Por fim, no artigo [71] não é descrita qualquer informação sobre a fase de análise e tratamento dos dados, sendo essa a sua principal lacuna. Os métodos utilizados são descritos com algum pormenor, explicando o funcionamento de cada um, assim como as suas características. A demonstração dos resultados obtidos é completa, expondo-os em gráficos e tabelas para ser mais perceptível a sua interpretação. Em suma, todos os artigos revistos relatam imensa informação sobre os métodos e abordagens tomadas assim como os resultados obtidos, no entanto pecam todos é na ausência de informação disposta relativa à fase de análise e tratamento dos dados.

Verifica-se que, em todos estes estudos, nenhum utilizou como variáveis de *input* dados climatológicos, podendo estes desempenhar um papel muito relevante nas conclusões finais, como por exemplo, para a justificação de determinadas características dos afluentes numa ETAR, num determinado período, ou então para se detetar a possibilidade da existência de descargas ilegais. Relativamente a este segundo ponto pode ser detetado caso, para um determinado período de tempo, o clima seja seco e o caudal do afluente à ETAR seja acima da média. Também o facto destes estudos serem focados exclusivamente na eficiência energética podem retornar conclusões, que posteriormente, podem comprometer a qualidade do efluente, pois como utilizam apenas duas ou três variáveis como *output*, pode-se focar em aumentar a eficiência do tratamento na remoção destas variáveis mas podem prejudicar a remoção das restantes, comprometendo assim a qualidade exigida do efluente.

Materiais e Métodos

Este capítulo descreve os materiais e métodos considerados nesta dissertação. A secção 3.1 faculto o processo de recolha e armazenamento dos dados utilizados neste desígnio, descrevendo todas as etapas executadas nesta fase, desde de como foi cumprida a recolha dos dados até as implementações e abordagens efetuadas para o armazenamento dos mesmos. Na secção 3.2 é elaborada uma análise inicial a todos os dados a que se tem acesso. É produzida uma análise singular para cada tipo de dados, onde é feita uma análise geral das suas características aplicando-se estatística descritiva com o intuito de se adquirir um conhecimento mais amplo sobre os mesmos. Dessa análise também é investigado se os dados contém lacunas entre os seus registos.

3.1 Recolha e Armazenamento dos Dados

Num projeto deste género, recolher o máximo de dados possível é um dos principais objetivos, pois quantos mais dados se tiver ao dispor, maiores e melhores serão os resultados obtidos. Por isso, neste processo tentou-se recolher o maior número de dados possível.

Os dados recolhidos para este estudo foram fornecidos por uma empresa multi-municipal de [AR](#), que facultaram dados de duas [ETAR](#). Uma delas situa-se na cidade de Vila Real, onde foram disponibilizadas informações relativas à electricidade, controlo analítico e funcionamento desta mesma infraestrutura. A outra [ETAR](#) localiza-se em Serzedo, na cidade de Guimarães e, relativamente a esta instalação, foram fornecidos dados alusivos à electricidade e caudais.

Todos os dados cedidos vinham como dados brutos, ou seja, para cada tipo (electricidade, controlo analítico, etc) os dados estavam todos juntos num único ficheiro, no formato `.x/sx`. Posto isso, decidiu-se que seria melhor separar os dados em ficheiros diferentes, para que a sua inserção na base de dados fosse mais eficiente. Para isso, foi realizada uma *script* em *python* que, para cada tipo, separa então os dados em indicadores, como por exemplo, a electricidade passou de um único ficheiro para cinco

ficheiros diferentes, correspondentes aos seus indicadores. Dentro de cada ficheiro gerado, estão todos os registos relativos ao indicador em questão, contendo toda a informação alusiva ao mesmo.

Posteriormente, foram adicionados ao estudo os dados climatológicos das cidades onde se encontram as ETAR, nomeadamente Vila Real e Guimarães. Esta inclusão deve-se ao facto de, nesta fase, se acreditar que estes dados podem retornar inúmeras vantagens ao estudo. Desde logo, perceber até que ponto a meteorologia influencia a qualidade das AR, se relaciona com os altos consumos energéticos, entre outros fenómenos. Estes dados foram adquiridos através da utilização da *API Open Weather Map*, que devolve informações meteorológicas diárias muito completas, como temperatura, chuva, pressão atmosférica, entre outros indicadores meteorológicos.

Para armazenar todos estes dados, foi criada uma base de dados *MySQL*, com o objetivo de que os dados sejam guardados de forma eficaz e simplista. A Figura 3.1 ilustra o modelo conceptual da mesma:

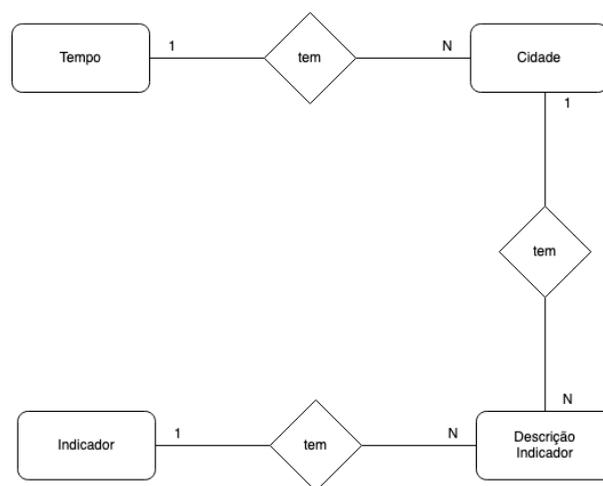


Figura 3.1: Modelo Conceptual Base de Dados.

Como pode ser observado, a base de dados é composta por quatro tabelas diferentes, sendo elas a tabela Cidade que é relativa aos dados referentes às duas cidades onde estão localizadas as infraestruturas, a tabela Indicador que armazena informação base relativa aos diferentes indicadores presentes em todos os dados, a tabela Descrição Indicador que deposita informações adicionais de cada indicador e por fim, a tabela Tempo que conserva informações relacionadas com a meteorologia. Na Figura 3.2 pode ser examinado com mais pormenor a estrutura e as características da base de dados criada.



Figura 3.2: Esquema Base de Dados.

Tendo a estrutura da base de dados criada, foi elaborada outra *script* em *python*, que de forma otimizada, faz a inserção dos dados na base de dados. Nessa *script* apenas se tem de definir o caminho onde se encontram as pastas dos dados e automaticamente são inseridos na base de dados, respeitando a estrutura da mesma.

Com a base de dados operacional, e conseqüentemente com o armazenamento dos dados de forma organizada, a resolução das etapas seguintes torna-se mais facilitada, uma vez que a consulta aos dados é bastante mais intuitiva.

3.2 Exploração Inicial aos Dados

De forma a identificar as características iniciais dos dados a que se tinha ao dispor, nesta secção é feita uma análise inicial aos dados fornecidos. Desta análise pode-se verificar as lacunas existentes nos dados, expondo alguns dos tratamentos que posteriormente serão necessários aplicar, de modo a colocar os dados na melhor forma possível, para se poder extrair dos mesmos o máximo de informação real.

3.2.1 Vila Real

Como supramencionado, o primeiro *dataset* é relativo à *ETAR* de Vila Real, onde foram fornecidos dados sobre a electricidade, controlo analítico e funcionamento, desta mesma infraestrutura.

3.2.1.1 Electricidade

Os dados referentes à electricidade correspondem aos gastos de energia elétrica efetuados pela infraestrutura, durante o seu funcionamento. Analisado os dados, verifica-se desde logo que se dispõem de 4 indicadores diferentes, sendo eles *cheia*, *ponta*, *super vazio* e *vazio*. Estes indicadores correspondem aos períodos de consumo da electricidade durante o dia, ou seja, representam o consumo eléctrico nos diferentes períodos horários, onde as taxas a pagar variam consoante o período em questão. De modo a ser mais perceptível, a Tabela 3.1 descreve os horários que definem esses períodos:

Tabela 3.1: Períodos Diários da Electricidade.

Indicador	Hora Legal de Inverno	Hora Legal de Verão
Ponta	09h00 às 10h30	10h30 às 13h00
	18h00 às 20h30	19h30 às 21h00
Cheia	08h00 às 09h00	08h00 às 10h30
	10h30 às 18h00	13h00 às 19h30
	20h30 às 22h00	21h00 às 22h00
Vazio	00h00 às 02h00	00h00 às 02h00
	06h00 às 08h00	06h00 às 08h00
	22h00 às 00h00	22h00 às 00h00
Super Vazio	02h00 às 06h00	02h00 às 06h00

Todos os dados da electricidade equivalem a um total de 5392 registos, onde cada indicador abarca 1348 registos entre 01/01/2016 e 27/05/2020, com uma unidade de medida em *Quilowatt-Hora (kWh)*. Em todos os indicadores não existe a presença de valores em falta mas, em contrapartida, verifica-se a existência de *timesteps* em falta, num total de 263 registos, em cada um deles. Em cada indicador, 95% dos registos contém periodicidade diária, apresentando depois sobretudo registos de dois em dois dias e de sete em sete dias. De modo a ter uma primeira perceção dos valores que cada indicador apresenta, foi elaborada uma estatística descritiva a cada um, onde foi calculada a média, moda, mediana, desvio padrão e valor mínimo e máximo. A Tabela 3.2 ilustra os resultados obtidos.

Tabela 3.2: Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Vila Real.

Métricas	Cheia	Ponta	Super Vazio	Vazio
Média	1951437.288	539147.179	703131.769	1358468.661
Moda	681480	187062	472522	498975
Mediana	1250587	347913.5	472522	922215
Desvio Padrão	1465015.578	398554.091	599736.465	987887.925
Mínimo	0	0	0	0
Máximo	4313751	1175533.0	11011102	2949111

Examinada a tabela 3.2 pode-se verificar prontamente que todos os indicadores apresentam valores elevadíssimos, expondo uma clara existência de uma lacuna nos dados, uma vez que uma infraestrutura equivalente nunca chega perto destes valores. Outra conclusão que se pode tirar, é a homogeneidade presente em todos os indicadores no que toca ao valor mínimo de cada um. Como supramencionado,

uma ETAR funciona 24h/24h, portanto o valor mínimo nunca pode ser igual a zero, logo todos os zeros existentes terão de se converter em valores nulos.

Para se ter uma ideia de como é a variação média de cada indicador da electricidade ao longo dos meses de todos os anos, foi concebido um gráfico para ilustrar essa mesma variação. A Figura 3.3 exibe essa mesma informação.

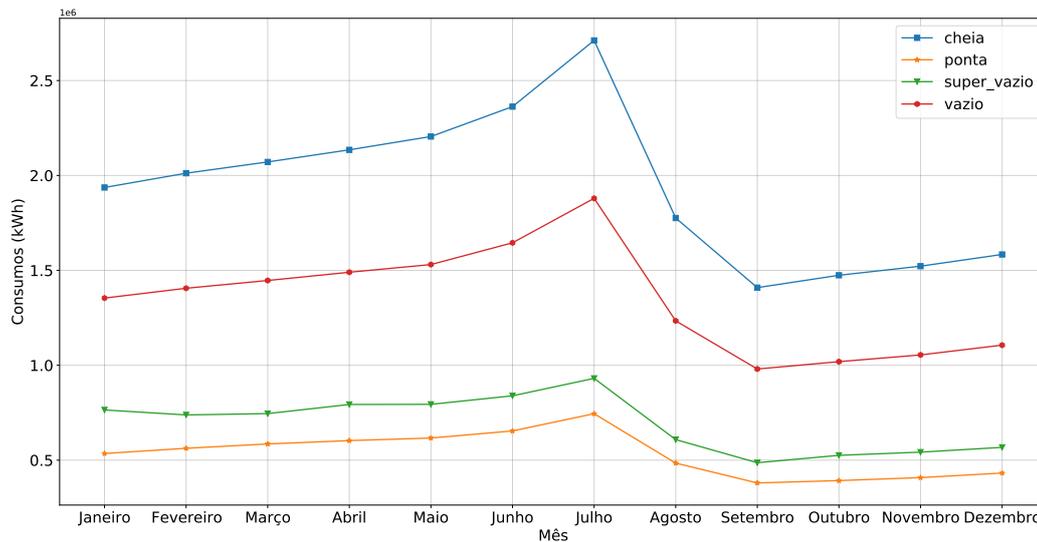


Figura 3.3: Variação média da Electricidade de Vila Real ao longo dos meses de todos os anos.

Como se pode observar na Figura 3.3, a electricidade tem um crescimento linear de Janeiro a Julho, que depois baixa ligeiramente até Setembro voltando a crescer até Dezembro. Isto reflete que os dados contém outra lacuna, uma vez que os consumos energéticos nestas infraestruturas não varia assim tanto entre os meses do ano. Posto isso, foi elaborada uma análise registo a registo com o intuito de identificar essa mesma deficiência. Verificou-se que todos os indicadores contém dados acumulados, ou seja, o consumo do dia é igual ao consumo do dia anterior mais o consumo do próprio dia, sendo que a meio do ano existe uma espécie de *reset*, onde depois os dados voltam a ficar acumulados. Identificou-se também a presença de erros de inserção nos dados, onde se encontrava diferenças diárias na ordem dos milhares de kWh, sendo que era perfeitamente perceptível que se tratava de erros de inserção.

3.2.1.2 Controlo Analítico

Relativamente aos dados do controlo analítico, como o próprio nome indica, correspondem aos valores apresentados pelas substâncias presentes nas AR, em diversas etapas/setores da instalação, adquiridos através do controlo analítico efetuado às mesmas. Tem-se então acesso a informação referente a distintos indicadores, presentes em diferentes setores da infraestrutura. Com o intuito de ter uma compreensão inicial dos dados, foi aplicada estatística descritiva aos mesmos, calculando-se a média, moda, mediana, desvio padrão, valor mínimo e máximo. A Tabela 3.3 ilustra o resultado da estatística descritiva efetuada,

e também os setores e indicadores existentes, assim como o número de registos, periodicidade e unidade de medida que cada um apresenta.

Tabela 3.3: Dados Controlo Analítico.

Setor	Indicador	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo	Periodicidade
Afluente Bruto	Azoto Total	mg/l	95	66.045	0	71.3	28.050	0	145	Semanal
	CBO	mg/l	48	438.979	0	405.5	270.389	0	1050	Quinzenal
	CQO	mg/l	95	699.547	0	719	340.661	0	1568	Semanal
	Fósforo Total	mg/l	97	7.118	0	7.64	3.333	0	14.5	Semanal
	Ph	Escala de Sorensen	205	7.337	7.6	7.6	1.471	0	8.94	Semanal
Câmara de Degaseificação	SST	mg/l	95	264.916	0	245	172.686	0	820	Semanal
	Amónia	mg/l	109	7.415	0.5	2.6	11.994	0	58.5	Semanal
	Nitratos (NO3)	mg/l	90	4.729	0	4.905	3.403	0	19.8	Semanal
	SST	mg/l	95	317.842	0	265	264.051	0	1380	Semanal
Efluente Tratado	Amónia	mg/l	214	11.379	0	7.4	10.977	0	47.6	Dois em Dois dias
	Azoto Total	mg/l	111	15.859	11.4	13.4	9.504	0	54	Semanal
	CBO	mg/l	49	4.327	2	3	3.023	0	12	Quinzenal
	CQO	mg/l	98	28.311	0	28.4	10.724	0	53.9	Semanal
	Fósforo Total	mg/l	104	1.315	0.5	0.825	1.183	0	6.130	Semanal
	Nitratos (NO3)	mg/l	209	4.284	0	4	3.199	0	26.3	Dois em Dois dias
	Ortofosfatos	mg/l	110	1.538	0.5	0.886	1.701	0	7.99	Semanal
	Ph	Escala de Sorensen	273	6.832	6.9	6.99	1.2001	0	8.490	Semanal
Entrada Reator Biológico	SST	mg/l	98	5.541	3	4	5.501	0	33	Semanal
	Azoto Total	mg/l	96	58.879	0	63.7	25.878	0	126	Semanal
	CBO	mg/l	47	274.894	240	240	183.586	0	900	Quinzenal
	Fósforo Total	mg/l	96	6.844	0	7.375	3.141	0	14.1	Semanal
	Ph	Escala de Sorensen	201	7.404	7.5	7.52	1.162	0	9.560	Semanal
Lamas Biológicas 1	SST	mg/l	96	96.343	0	78	73.688	0	480	Semanal
	Amónia	mg/l	192	4.607	0	0.5	8.835	0	49.9	Dois em Dois dias
	IVL	mg/l	96	129.169	0	145.15	80.969	0	258.3	Semanal
	Nitratos (NO3)	mg/l	192	3.429	0	3.38	3.118	0	21.7	Dois em Dois dias
	Oxigénio	mg/l	167	1.502	0	0.9	1.459	0	6.650	Dois em Dois dias
	Ph	Escala de Sorensen	298	6.554	6.7	6.655	1.045	0	8.730	Semanal
	SST	mg/l	74	4914.851	0	5030	1678.577	0	8100	Semanal
Lamas Biológicas 2	SSV	mg/l	41	3691.244	0	4060	1756.244	0	6570	Quinzenal
	Amónia	mg/l	196	4.629	0	0.5	8.129	0	35.4	Dois em Dois dias
	IVL	mg/l	96	129.44	0	142	90.789	0	312.4	Semanal
	Nitratos (NO3)	mg/l	193	3.517	0	2.78	3.749	0	23.1	Dois em Dois dias
	Oxigénio	mg/l	172	2.062	0	1.32	1.819	0	8.7	Dois em Dois dias
	Ph	Escala de Sorensen	296	6.588	6.4	6.64	0.896	0	9.09	Semanal
	SST	mg/l	78	3565.586	0	4235	2464.846	0	8210	Semanal
Lamas Biológicas Espessadas	SSV	mg/l	40	2681.825	0	3250	2131.129	0	6710	Quinzenal
	ST	mg/l	53	25.720	0	26.634	10.853	0	48.134	Quinzenal
Lamas Biológicas Recirculadas	ST	mg/l	94	8.671	0	8.848	3.781	0	28.954	Semanal
Lamas Digeridas 1	Ácidos Gordos Voláteis	mg/l	94	36613.404	0	200	14213.697	0	67900	Semanal
	Alcalinidade	mg/l	94	1928.617	0	2000	2046.617	0	13850	Semanal
	Amónia	mg/l	94	520.535	0	660	347.627	0	1055	Semanal
	Fósforo Total	mg/l	94	287.824	0	355.5	196.218	0	765	Semanal
	Ph	Escala de Sorensen	124	6.764	7.15	7.15	1.318	0	8.4	Semanal
	ST	mg/l	94	10.774	0	13.947	7.432	0	26.710	Semanal
	SV	mg/l	94	7.511	0	9.506	5.225	0	17.424	Semanal
Lamas Digeridas 2	Ácidos Gordos Voláteis	mg/l	95	3593.158	240	240	13892.099	0	66800	Semanal
	Alcalinidade	mg/l	95	2275.579	2080	2080	1456.578	0	8250	Semanal
	Amónia	mg/l	95	631.546	0	705	283.885	0	1100	Semanal
	Fósforo Total	mg/l	95	348.292	0	370	149.826	0	630	Semanal
	Ph	Escala de Sorensen	130	6.756	7.18	7.05	1.169	0	8.5	Semanal
	ST	mg/l	95	15.673	0	16.12	4.766	0	36.684	Semanal
	SV	mg/l	95	10.486	0	11.148	4.239	0	25.862	Semanal
Lamas Mistas	ST	mg/l	95	25.695	0	25.523	12.051	0	52.792	Semanal
	SV	mg/l	73	19.921	0	19.794	10.348	0	41.658	Semanal
Lamas Primárias 1	ST	mg/l	96	31.633	0	31.776	17.715	0	121.428	Semanal
	SV	mg/l	74	26.186	0	27.297	16.050	0	103.810	Semanal
Lamas Primárias 2	ST	mg/l	96	28.412	0	28.891	17.904	0	111.592	Semanal
	SV	mg/l	74	23.125	0	24.29	16.483	0	99.821	Semanal
Poço Escorrências	Azoto Total	mg/l	48	170.896	0	172.5	86.940	0	315	Quinzenal
	CQO	mg/l	48	1625.833	0	1340.5	1336.944	0	5710	Quinzenal
	Fósforo Total	mg/l	48	29.469	0	28.25	20.573	0	138	Quinzenal
	SST	mg/l	48	1040.452	0	750.5	974.804	0	3550	Quinzenal
Tanque Anóxico Tanque 1	Nitratos (NO3)	mg/l	95	0.707	0	0.443	1.618	0	13	Semanal
	Oxigénio	mg/l	52	0.310	0.2	0.21	0.484	0	3.560	Semanal
Tanque Anóxico Tanque 2	Nitratos (NO3)	mg/l	95	0.663	0	0.483	1.492	0	12.9	Semanal
	Oxigénio	mg/l	61	0.349	0.38	0.29	0.413	0	3.270	Semanal

Os dados alusivos ao controlo analítico correspondem a um total de 7400 registos, compreendidos entre Agosto de 2018 e Maio de 2020, com exceção do indicador pH , onde concentra registos entre Janeiro de 2016 e Maio de 2020. Conta-se então com 16 setores e 16 indicadores diferentes, onde em média cada indicador contém 110 registos, por setor. Verifica-se que em todos os indicadores não existe valores em falta, ao invés da presença de *timesteps* em falta, onde se verifica que se inclui em

praticamente todos os indicadores. Na generalidade, praticamente todos os indicadores apresentam um periodicidade semanal, sendo que em alguns casos, existem indicadores que apresentam registos de 2 em 2 dias, como é o caso, por exemplo, da *amónia* no *efluente tratado*. Também existem casos onde os indicadores apresentam uma periodicidade quinzenal. De ressaltar que, todos os indicadores, em todos os setores, contém valor mínimo igual a zero, sendo que a maior parte deles contém também a moda igual a zero, ou seja, a maioria dos registos nestes casos contém valor igual a zero. Como nestas infraestruturas é praticamente impossível estes indicadores apresentarem valores iguais a zero, assume-se com a devida ajuda de pessoas da área, que estes valores estão incorretos sendo que, posteriormente, terão de corresponder a valores nulos.

3.2.1.3 Funcionamento

Os dados referentes ao funcionamento correspondem às horas de funcionamento de diversos equipamentos presentes nos diferentes setores da instalação. Estes dados abrangem um total de 8436 registos, onde se contém informações sobre dois setores (linha líquida e linha sólida) e 89 indicadores distintos, 36 correspondentes à *linha líquida* e os restantes 53 à *linha sólida*. Entende-se como indicadores os pontos/equipamentos de consumo presentes na infraestrutura. De forma a ter uma perceção mais ampla dos dados, aplicou-se estatística descritiva aos mesmos, onde se calculou a média, moda, mediana, desvio padrão valor mínimo e máximo. A Tabela 3.4 ilustra então os resultados obtidos da estatística descritiva aplicada e, adicionalmente, expõem os setores e indicadores existentes, assim como o número de registos, periodicidade e unidade de medida que cada um apresenta.

Verifica-se desde logo a escassez de dados em praticamente todos os indicadores, uma vez que possuem periodicidade mensal. Os dados na esmagadora maioria estão compreendidos entre Setembro de 2017 e Maio de 2020, sendo que em alguns casos então compreendidos entre Janeiro de 2016 e Maio de 2020, como é o caso, por exemplo, do indicador *Sistema de Apoio Geral Caldeira*, na *linha líquida*. Apurou-se que não existe valores em falta, mas identificou-se a ausência de *timesteps* em falta, nos indicadores que apresentam periodicidade diária.

De modo a possuir uma visão geral dos valores que estes dados apresentam, foi criado um gráfico que ilustra a variação média de cada linha de funcionamento ao longo dos meses, de todos os anos. A Figura 3.4 elucida essa mesma variação.

Tabela 3.4: Dados Funcionamento.

Setor	Indicador	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo	Periodicidade	
Linha Líquida	Equalização Efluente Biológico Bomba 1	Horas	21	1212.762	997	1010	438.389	808	2116	Mensal	
	Equalização Efluente Biológico Bomba 2	Horas	21	451.238	284	416	193.860	284	918	Mensal	
	Tanque Equalização B Agitador 1	Horas	21	3021.714	2974	2545	1521.582	1845	7082	Mensal	
	Tanque Equalização B Agitador 2	Horas	21	3402.238	1362	2009	3818.271	1309	12027	Mensal	
	Tanque Equalização C Agitador 1	Horas	21	3878.905	1892	2539	3553.135	1839	12842	Mensal	
	Tanque Equalização C Agitador 2	Horas	21	3942.238	1892	2538	3715.543	1839	13296	Mensal	
	Obra de Entrada 15V101	Horas	33	38740.091	3687	15717	29405.358	3687	75365	Mensal	
	Obra de Entrada 15V102	Horas	33	38816.121	3687	15800	29407.358	3687	75454	Mensal	
	Obra de Entrada Agitador Bacial Fossas 1	Horas	33	4274.333	343	646	4330.112	343	9945	Mensal	
	Obra de Entrada Agitador Bacial Fossas 2	Horas	33	3153.303	343	646	3064.471	343	8550	Mensal	
	Obra de Entrada Barras Automáticas 1	Horas	52	578.308	277	625.5	153.186	277	750	Diária	
	Obra de Entrada Barras Automáticas 2	Horas	45	573.578	443	577	71.721	443	687	Diária	
	Obra de Entrada Barras Automáticas 3	Horas	45	459.578	529	466	63.241	347	552	Diária	
	Obra de Entrada Bomba Bacica Fossica 1	Horas	33	113.394	13	33	103.669	13	247	Mensal	
	Obra de Entrada Bomba Bacica Fossica 2	Horas	33	196.727	73	80	134.240	73	353	Mensal	
	Obra de Entrada Bomba Bacica Fossica 3	Horas	33	120.939	7	14	123.409	7	270	Mensal	
	Obra de Entrada Grelha Mecânica	Horas	33	194	252	157	53.899	141	252	Mensal	
	Obra de Entrada Tamisador 1	Horas	33	14419.788	3478	6185	10794.833	3478	27339	Mensal	
	Obra de Entrada Tamisador 2	Horas	33	25871.485	6794	10548	19096.368	6794	49553	Mensal	
	Obra de Entrada Tamisador 3	Horas	33	20772.394	5275	8093	15580.120	5275	40127	Mensal	
	Sistema de Apoio Geral Caldeira	Horas	1602	2206.911	2171	2171	864.651	0	5225	Diária	
	Sistema de Apoio Geral Gerador	Horas	1602	168.047	755	42	283.517	1	755	Diária	
	Tratamento Primário 04-CR-01	Horas	33	15662.576	22453	11215	7598.504	6656	26249	Mensal	
	Tratamento Primário 04-CR-02	Horas	33	15398.727	2225	6319	13595.083	2220	32189	Mensal	
	Tratamento Primário 04-CR-03	Horas	33	22591.576	9582	13496	13647.398	8695	40268	Mensal	
	Tratamento Primário Agitador Poço Gorduras	Horas	33	9032.545	13458	5632	4298.348	4880	13778	Mensal	
	Tratamento Primário Bomba 1 Gorduras	Horas	33	517.030	290	312	244.631	282	793	Mensal	
	Tratamento Primário Bomba 1 Lamas Primárias	Horas	33	3496.278	1037	1935	2266.660	1037	6652	Mensal	
	Tratamento Primário Bomba 2 Gorduras	Horas	33	379.303	553	247	161.316	221	550	Mensal	
	Tratamento Primário Bomba 2 Lamas Primárias	Horas	33	3229.576	5185	1729	2192.149	853	6263	Mensal	
	Tratamento Primário Bomba Areias Desarenador 1	Horas	33	837.667	225	423	589.239	225	1714	Mensal	
	Tratamento Primário Bomba Areias Desarenador 2	Horas	33	789.818	237	389	542.603	237	1549	Mensal	
	Tratamento Primário Ponte Raspadora Desarenador 1	Horas	33	22344.273	8322	12074	13521.188	8322	40145	Mensal	
	Tratamento Primário Ponte Raspadora Desarenador 2	Horas	33	21187.152	33505	11486	12797.824	8139	37310	Mensal	
	Tratamento Primário Ponte Raspadora Sedipac 1	Horas	33	25848.667	5961	11746	19041.025	5961	51281	Mensal	
	Tratamento Primário Ponte Raspadora Sedipac 2	Horas	33	22414.030	37323	10161	16799.066	4642	45013	Mensal	
	Linha Sólida	Desidratação Bomba Elevação Lamas Desidratadas	Horas	33	17149.303	4281	6894	12823.808	4281	33763	Mensal
		Desidratação Bomba Poli Centrifuga 1	Horas	33	13107	23308	5513	9465.957	4079	23309	Mensal
		Desidratação Bomba Poli Centrifuga 2	Horas	33	12248.909	3612	5663	8565.459	3612	22723	Mensal
		Desidratação Bomba Poli Reserva Centrifuga	Horas	33	3383.121	622	1722	2799.261	548	7857	Mensal
		Desidratação Centrifuga 1	Horas	33	14749.394	26351	6067	10764.547	4516	26351	Mensal
		Desidratação Centrifuga 2	Horas	33	16159.515	4539	7212	11429.121	4539	31361	Mensal
		Desidratação Parafuso Lamas Saída Centrifuga	Horas	33	17416.303	4349	6966	13039.107	4349	34271	Mensal
		Desodorização 15-VE-01	Horas	33	21228.455	4335	4371	18836.293	4333	44889	Mensal
		Desodorização Bomba Ácido Sulfúrico	Horas	33	59.818	21	21	43.189	21	107	Mensal
		Desodorização Bomba Hipoclorito de Sódio	Horas	33	324.697	79	79	273.468	79	625	Mensal
		Desodorização Bomba Recirculação Torre Ácida	Horas	33	23902.969	6532	6576	19361.274	6530	48124	Mensal
Desodorização Bomba Recirculação Torre Básica		Horas	33	22834.545	5912	5956	18858.868	5910	46316	Mensal	
Desodorização Bomba Soda		Horas	33	490.151	126	126	405.263	126	932	Mensal	
Desodorização Ventilador 1		Horas	33	10561.545	3677	3701	7682.174	3672	20414	Mensal	
Desodorização Ventilador 2		Horas	33	14763.181	2856	2856	13269.571	2846	30655	Mensal	
Esspessamento Agitador Tanque Lamas Mistas		Horas	33	37840.909	68197	15149	28190.496	9974	71781	Mensal	
Esspessamento Bomba 1 Lamas Espessadas		Horas	452	2864.752	2882	2703	1131.742	2018	10061	Diária	
Esspessamento Bomba 1 Lamas de Saída das Mesas		Horas	33	10815.303	3970	2925	8222.142	2925	20751	Mensal	
Esspessamento Bomba 1 Saída TLM		Horas	450	3991.131	3382	3632	1772.246	3137	13400	Diária	
Esspessamento Bomba 2 Lamas Espessadas		Horas	450	1466.48	6876	1317	1006.042	718	6881	Diária	
Esspessamento Bomba 2 Lamas Saída das Mesas		Horas	33	10512.367	2950	3908	7919.357	2950	20005	Mensal	
Esspessamento Bomba 2 Saída TLM		Horas	450	4336.524	4303	4073.5	2067.729	1020	17335	Diária	
Esspessamento Bomba Poli Mesa 1		Horas	33	5443.879	10193	2070	4410.125	1185	10216	Mensal	
Esspessamento Bomba Poli Mesa 2		Horas	33	9154.030	2705	3502	6948.652	2705	18067	Mensal	
Esspessamento Bomba Poli Reserva		Horas	33	3840.576	5949	2580	2990.674	1411	6113	Mensal	
Esspessamento Bomba de Água de Lavagem da Mesa 1		Horas	33	8304.091	14387	3970	5650.873	2818	14429	Mensal	
Esspessamento Bomba de Água de Lavagem da Mesa 2		Horas	33	12899.152	4047	5147	10006.743	2876	26207	Mensal	
Esspessamento Compressor de Ar de Serviço 1		Horas	33	176.848	178	178	226.474	47	1335	Mensal	
Esspessamento Compressor de Ar de Serviço 2		Horas	33	35.455	41	41	20.537	1	78	Mensal	
Esspessamento Mesa 1		Horas	33	8327.969	14422	3986	5661.361	2827	14463	Mensal	
Esspessamento Mesa 2		Horas	33	12897.424	4056	5160	10001.491	2883	26269	Mensal	
Esspessamento Ponte Raspadora Espessador Gravítico		Horas	33	28234.393	7535	13365	19910.203	7535	54559	Mensal	
Esspessamento Preparador Polímetro 1		Horas	33	604.333	344	344	401.844	167	1209	Mensal	
Esspessamento Preparador Polímetro 2		Horas	33	990.545	246	297	818.078	246	1885	Mensal	
Estabilização Biológica Agitador Cuba Dig 1		Horas	30	31047.833	8405	10892	24596.714	6499	59852	Mensal	
Estabilização Biológica Agitador Cuba Dig 2		Horas	30	30019.333	6332	10825	23287.029	6332	59163	Mensal	
Estabilização Biológica Bomba Lamas Digeridas Dig 1		Horas	30	3790.633	5383	3246.5	1460.318	1858	5410	Mensal	
Estabilização Biológica Bomba Lamas Digeridas Dig 2		Horas	30	3754.733	166	3605.5	1522.915	166	5858	Mensal	
Estabilização Biológica Bomba Poço Condensados Dig 1		Horas	30	143.533	54	66	95.568	51	249	Mensal	
Estabilização Biológica Bomba Poço Condensados Dig 2		Horas	30	150.233	87	130.5	56.288	83	223	Mensal	
Estabilização Biológica Bomba Recirculação Água Quente Dig 1		Horas	30	23833.733	5751	8775.5	18006.130	5751	45636	Mensal	
Estabilização Biológica Bomba Recirculação Água Quente Dig 2		Horas	30	23921.533	5721	8982.5	18019.829	5721	45790	Mensal	
Estabilização Biológica Bomba Recirculação Água Quente Reserva		Horas	30	23591.267	4001	8456.5	19844.784	4001	70635	Mensal	
Estabilização Biológica Bomba Recirculação Lamas Dig 1		Horas	30	27774.333	9949	12763	18921.006	9026	50849	Mensal	
Estabilização Biológica Bomba Recirculação Lamas Dig 2		Horas	30	26569.433	49363	8165	21989.467	4249	52797	Mensal	
Estabilização Biológica Bomba Recirculação Lamas Reserva		Horas	30	20923.467	4742	7543.5	16095.949	4742	40294	Mensal	
Estabilização Biológica Bomba Recirculação Caldeira		Horas	30	2194.167	2171	2171	1079.720	33	4712	Mensal	
Estabilização Biológica Caldeira		Horas	30	604.067	1198	148.5	543.591	116	1240	Mensal	
Estabilização Biológica Compressor Biogas Digestor 1		Horas	30	35883.1	8478	11059.5	28790.565	8478	71047	Mensal	
Estabilização Biológica Compressor Biogas Digestor 2		Horas	30	35548.3	62864	14021	26408.289	8539	67191	Mensal	
Estabilização Biológica Ventilador Gasometro		Horas	30	39798.5	10039	15525.5	29376.767	10039	75580	Mensal	
Produção de Energia Máquina de Cogeração 1135		Horas	417	5676.866	6424	6222	936.929	547	6434	Diária	
Produção de Energia Máquina de Cogeração 1136		Horas	417	7012.386	8309	7345	1159.753	4933	8310	Diária	

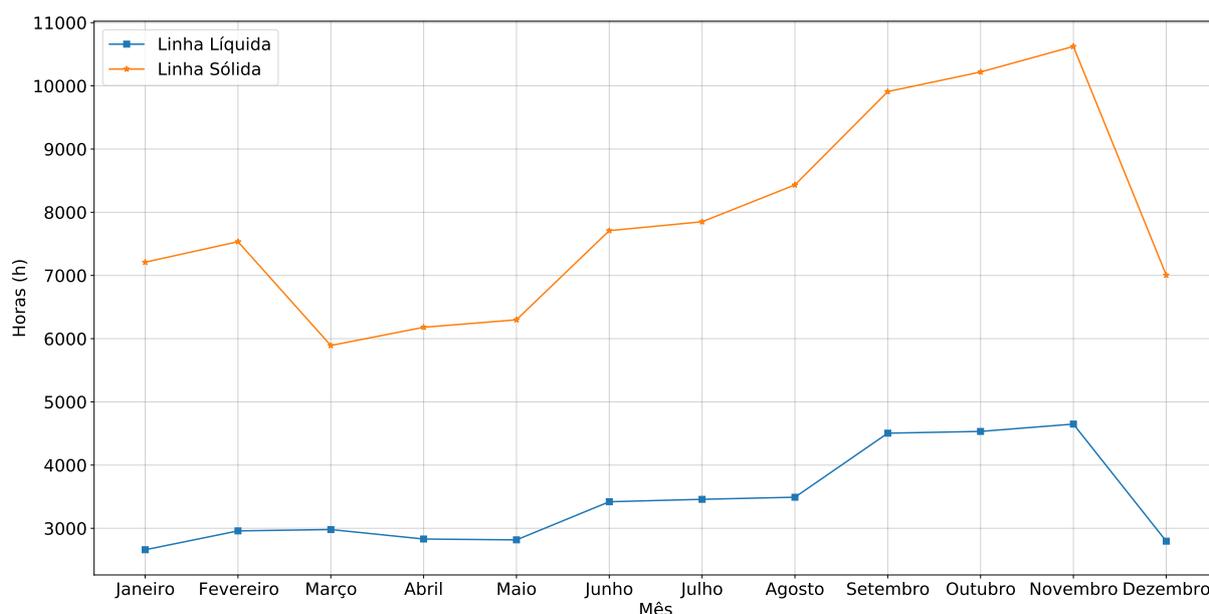


Figura 3.4: Variação média do Funcionamento ao longo dos meses de todos os anos.

Como se pode examinar na Figura 3.4, em praticamente todos os meses as horas de funcionamento tem tendência crescente, salvo três exceções, de Fevereiro a Março na linha sólida, de Março a Abril na *linha líquida* e de Novembro a Dezembro em ambas as linhas. Esse crescimento contínuo em quase todos os meses do ano raramente ocorre nestas infraestruturas, por isso está-se perante uma lacuna nos dados. Para averiguar o que provoca este crescimento, foi feita uma análise registo a registo aos dados. Conclui-se então que esse crescimento sucede devido aos dados estarem acumulados, ou seja, o total de horas de um certo dia é igual à soma das horas do dia anterior com o próprio dia, o que acontece com os dados relativos à electricidade visto anteriormente. Foi identificado também erros de inserção, onde a diferença entre registos era na ordem dos milhares, o que não está certamente correto.

3.2.1.4 Tempo

Relativamente aos dados atinentes ao tempo, correspondem aos valores do estado do tempo na cidade onde se localiza a ETAR, neste caso em Vila Real. No total, tem-se acesso a 1363 registos compreendidos entre Janeiro de 2016 e Maio de 2020, por cada indicador, com periodicidade diária. A Tabela 3.5 clarifica mais informações relativas a estes dados, expondo também o resultado da estatística descritiva aplicada aos mesmos.

Tabela 3.5: Dados Tempo Vila Real.

Indicador	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo
Temperatura	°C	1363	14.247	14.856	13.729	6.141	-0.099	30.373
Pressão Atmosférica	°C	1363	1018.803	1016.458	1018.375	6.678	984.583	1038.708
Humidade	%	1363	63.713	66.417	67.25	21.265	6.542	97.917
Velocidade do Vento	m/s	1363	2.035	2.375	1.878	0.809	0.755	7.765
Direção do Vento	Graus	1363	240.643	259.125	252.333	53.026	48.958	360
Nuvens	%	1363	47.106	0	46.75	32.394	0	100
Chuva	mm	450	0.638	0.12	0.498	0.519	0.11	4.062

Em todos os indicadores não existe valores nulos, exceto no indicador *chuva* que inclui 913 valores nulos. Relativamente aos *timesteps* em falta, não se identificou a sua presença.

3.2.2 Guimarães

O segundo *dataset* fornecido é relativo à ETAR de Serzedo, localizada na cidade de Guimarães, no qual foram facultados dados alusivo à Electricidade e Caudais desta mesma instalação.

3.2.2.1 Electricidade

Os dados pertencentes à electricidade, tal como em 3.2.1.1, são relativos aos gastos de energia elétrica realizado pela instalação, durante o seu funcionamento. Após uma análise aos mesmos, identificou-se que também se dispõem de 4 indicadores diferentes, nomeadamente *cheia*, *ponta*, *super vazio* e *vazio*. Esses indicadores equivalem aos diferentes períodos de consumo de electricidade durante o dia, exatamente o mesmo que em 3.2.1.1. Todos os estes dados correspondem a um total de 6088 registos, onde cada indicador abrange 1522 registos, compreendidos entre 01/01/2016 e 28/05/2020, com uma unidade de medida em kWh. Verifica-se que, em todos os indicadores, não existe valores em falta mas, identificou-se a existência de *timesteps* em falta, mais concretamente 91 registos em cada indicador. Relativamente à periodicidade, todos os indicadores apresentam uma periodicidade diária. Com o designio de obter uma perceção mais concreta sobre os dados, foi aplicada estatística descritiva aos mesmos, calculando a média, moda, mediana, desvio padrão, valor mínimo e máximo. A Tabela 3.6 aclara os resultados adquiridos.

Tabela 3.6: Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Guimarães.

Métricas	Cheia	Ponta	Super Vazio	Vazio
Média	2493261.499	631069.336	846335.405	1561107.892
Moda	2761603	699830	559339.0	498975
Mediana	2754295.5	699976	939203.5	1762589
Desvio Padrão	1396078.312	419849.825	392741.506	735588.486
Mínimo	6575	1021	1294	1172
Máximo	32336770	9111615	1398611	2659268

Tal como acontece em 3.2.1.1, os dados apresentam valores bastante elevados, sendo logo visível que os mesmos apresentam uma deficiência, uma vez que uma infraestrutura deste tipo nunca atinge valores de consumo equivalentes a estes, como supramencionado. Com isto, para ser mais perceptível a compreensão da variação dos consumos de um modo superficial, foi elaborado um gráfico que ilustra a variação média de cada indicador ao longo dos meses, em todos os anos de registos (2016 - 2020). A Figura 3.5 expõem essa mesma informação.

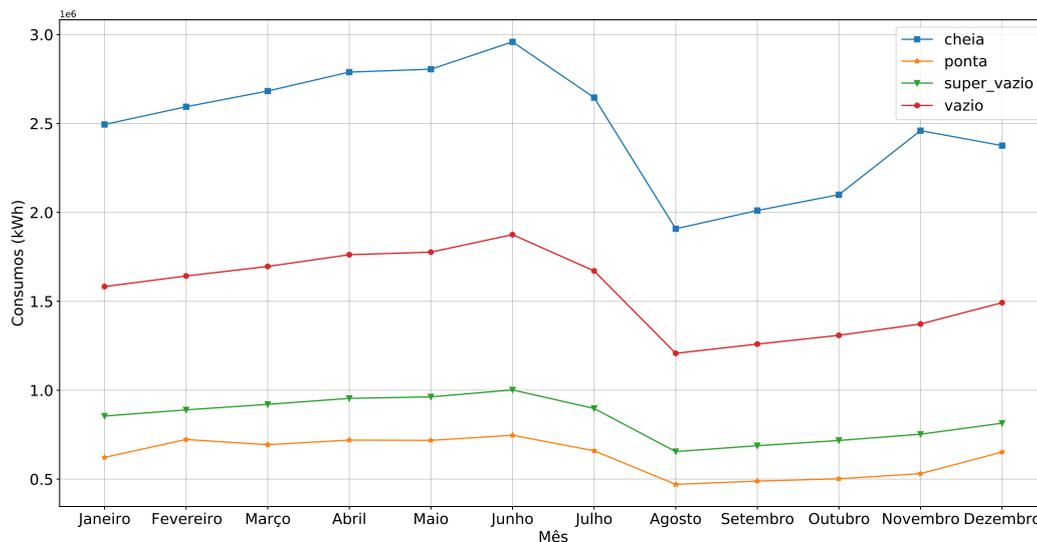


Figura 3.5: Variação média da Electricidade de Guimarães ao longo dos meses de todos os anos.

Como se pode observar, os dados tem uma tendência crescente em praticamente em todos os meses. Verifica-se que de Janeiro a Junho existe um crescimento contínuo nos consumos energéticos, que desce entre os meses de Junho e Agosto, que posteriormente volta a aumentar até Dezembro. Este comportamento é irrealista, uma vez que uma ETAR, por norma, não apresenta um crescimento contínuo de consumos em praticamente todos os meses do ano, portanto só prova que os dados estão acumulados. Para comprovar isso foi feita uma análise registo a registo e identificou-se que realmente os dados estavam acumulados, ou seja, o consumo de um determinado dia corresponde à soma dos consumos do dia anterior com o próprio dia. Também se verificou que existem erros de inserção nos dados, onde existem diferenças na ordem dos milhares de kWh entre dois dias, o que decididamente está errado.

3.2.2.2 Caudais

Em relação aos dados dos caudais, estes referem-se a informação relativa ao fluxo de AR presente em diversos setores na ETAR. No total, estes dados correspondem a 46638 registos, com informações sobre 15 setores presentes na infraestrutura. A fim de se ter um conhecimento mais amplo dos dados, foi aplicada estatística descritiva aos mesmos, onde foi calculada a média, moda, mediana, desvio padrão, valor mínimo e máximo. Na Tabela 3.7 é possível visualizar os resultados obtidos, assim como informação adicional sobre os dados, como o número de registos e periodicidades de cada setor.

Tabela 3.7: Dados Caudais.

Setor	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo	Periodicidade
Bugio	mm ³	87	70347.552	79610	70835	9492.054	52254	86147	Dois em Dois dias
Calvelos	mm ³	73	765906	765906	765906	0	765906.0	765906.0	Diária
Tanque Neutralização Caudal Faturação	mm ³	1518	5760218.692	5931753	5775097.5	3418351.301	10230	97766923	Diária
Caudal PqIndSocorro	mm ³	293	51776.884	0	31501	58038.112	0	172204	Diária
Central Preparação Polieletrólito Desidratação Centrífuga Bomba 1	mm ³	1531	9737.268	5458	9209	4863.210	5038	137782	Diária
Central Preparação Polieletrólito Desidratação Centrífuga Bomba 2	mm ³	1531	11321.385	12350	11575	4673.174	5646	155198	Diária
Escavanca	mm ³	87	9621.655	0	7244	9686.047	0	51488.0	Dois em Dois dias
Obra de Entrada Escorrências	mm ³	1535	2226082.153	2475909.0	2192354	784278.352	1408801	25830418	Diária
Fossas Sépticas	mm ³	1530	4952.614	0	7128	4504.769	0	78713.0	Diária
Lamas Flotadas	mm ³	1534	50709.546	53654	53654	11221.199	19176	466086	Diária
Central Preparação Polieletrólito Espessamento Tambor Bomba 1	mm ³	1529	1060.908	0	0	2574.739	0	52623	Diária
Central Preparação Polieletrólito Espessamento Tambor Bomba 2	mm ³	1531	4754.723	3614	5143	971.649	3066	13628	Diária
Tanque Lamas 1	mm ³	1533	304034.298	422244	295931	164754.702	155438	3522508	Diária
Tanque Lamas 2	mm ³	1533	429831.221	565670	422126	124320.875	286089	3701266	Diária
Vima Água	mm ³	607	32195.761	0	25384	25173.511	0	75042	Diária

Como se pode observar, tirando os setores *Bugio* e *Escavanca* que apresentam periodicidade de dois em dois dias, todos os setores apresentam uma periodicidade diária. Maioritariamente, os setores apresentam os registos compreendidos entre Janeiro de 2016 e Maio de 2020, sendo que os setores *Bugio*, *Escavanca* e *PqIndSocorro* apresentam os seus registos compreendidos entre Outubro de 2018 a Maio de 2020 e o setor *Vima Água* exhibe registos de Setembro de 2017 a Maio de 2020. No que respeito a *timesteps* em falta, não se verifica a existência dos mesmos, em todos os setores. Em contrapartida, constatou-se a existência, em todos os setores, de valores nulos (DISCRIMINAR??). De ressaltar que, todos os setores apresentam valores bastante elevados de fluxo dos caudais, onde dificilmente estas infraestruturas atingem estes valores. Para averiguar a possibilidade da presença de lacunas nos dados, foi elaborado um gráfico da variação dos valores durante os meses ao longo de todos os anos de registo, por exemplo, do setor *Lamas Flotadas*, uma vez que coloca-los todos juntos não se iria ter uma boa compreensão. A Figura 3.6 ilustra essa mesma variação.

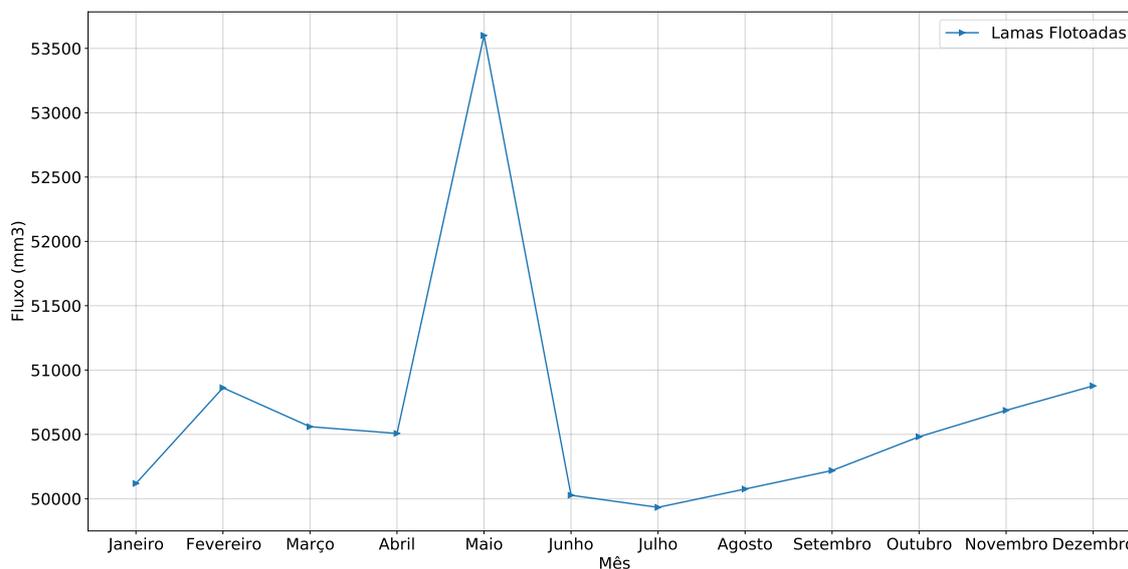


Figura 3.6: Variação do setor Lamas Flotadas ao longo dos meses de todos os anos.

Como pode ser examinado na Figura 3.6, os dados apresentam duas lacunas. A primeira, e que salta mais à vista, é a existência de erros de inserção. Como se pode observar, durante os meses de Abril e Maio os valores do fluxo apresenta uma diferença enorme, na ordem dos milhares, comparativamente

com os outros meses. Essa diferença é muito improvável de ocorrer, por isso foi feita uma análise registo a registo, com o intuito de confirmar essa lacuna. Feita a análise, confirma-se então que não só este setor, mas todos eles, apresentam erros de inserção nos seus registos. Relativamente à segunda lacuna existente, trata-se da acumulação dos dados. Como se pode ver, entre os meses de Julho e Dezembro, o fluxo tem um crescimento contínuo em todos os meses, onde mais uma vez, é um fenómeno muito incomum, o que prova a possibilidade de mais uma lacuna nos dados. Para se confirmar definitivamente essa lacuna, na análise registo a registo elaborada anteriormente, verificou-se que os dados, em todos os setores, apresentam valores acumulados, onde o valor do dia corresponde à soma do dia anterior com o próprio dia.

3.2.2.3 Tempo

No que se refere aos dados do tempo, estes correspondem aos dados climatológicos relativos à cidade de Guimarães, onde se localiza a [ETAR](#). Tem-se, no total, acesso a 1250 registos, entre Janeiro de 2016 e Maio de 2020, com uma periodicidade diária, para cada indicador. De modo a clarificar melhor estes dados, foi elaborada estatística descritiva nos dados, onde foi calculada a média, moda, mediana, desvio padrão, valor mínimo e máximo. Na Tabela 3.8 é possível examinar os resultados obtidos.

Tabela 3.8: Dados Tempo Guimarães.

Indicador	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo
Temperatura	°C	1250	14.241	11.703750	14.013	4.6197	2.081	27.233750
Pressão Atmosférica	°C	1250	1018.497	1019.583	1018.292	6.865	983.577	1037.667
Humidade	%	1250	80.300	84.333	83.583	13.055	31.250	100
Velocidade do Vento	m/s	1250	3.506	2.908	3.233	1.414	0.938	11.388
Direção do Vento	Graus	1250	192.711	176.667	189.465	58.029	47.917	337.083
Nuvens	%	1250	42.087	0	45.100	25.661	0	99.917
Chuva	mm	258	0.919	0.25	0.574	1.113	0.11	13.970

Tirando o indicador *chuva*, que contém 992 valores nulos, todos os restantes indicadores não apresentam valores nulos. No que toca a *timesteps* em falta, em todos os indicadores não se verificou que o mesmo ocorre.

Manipulação dos Dados

Neste capítulo são expostas as etapas realizadas em todo o processo de manipulação dos dados, onde o principal objetivo é extrair o máximo de informação possível dos dados e tentar garantir que a mesma representa a realidade, e não guie para conclusões irrealistas. Portanto, na secção 4.1 são exibidos todos os passos realizados no processo de tratamento nos diferentes tipos de dados, nas respetivas infraestruturas. Esse tratamento visa corrigir todas as lacunas previamente identificadas em cada um. Já a secção 4.2 disponibiliza uma análise exploratória para cada tipo de dados. Nessa análise são exploradas meticolosamente todas as características e padrões existentes nos diversos tipo de dados, usando-se diferentes *dashboards* para permitir uma melhor interpretação dos dados. Por fim, na secção 4.4 são descritas as tecnologias e *softwares* utilizados na realização deste projeto, nas diferentes fases que o mesmo apresenta.

4.1 Tratamento dos Dados

Antes de se efetuar uma análise mais profunda aos dados e evitar que se obtenha informações enganosas que podem levar a conclusões incorretas, é necessário corrigir e tratar algumas lacunas que os mesmos apresentam. Por isso, nesta secção serão ilustradas todas as etapas e abordagens realizadas no tratamento dos dados com o intuito de corrigir essas falhas, para que posteriormente, se possa realizar uma análise exploratória nos mesmos e se obtenha informação real.

4.1.1 Vila Real

Para a [ETAR](#) que se localiza na cidade de Vila Real, como visto anteriormente, tem-se acesso a dados relativos à electricidade, controlo analítico, funcionamento e tempo. Em relação aos dados do tempo, não será necessário proceder-se a qualquer tratamento nos mesmos. Estes apenas apresentaram valores

nulos no indicador *Chuva*, mas uma vez que esses valores nulos correspondem a 67% dos dados do indicador, não se torna muito confiável preencher esses valores. Portanto, abaixo serão então expostas todas as ações realizadas no processo de tratamento nos dados da electricidade, controlo analítico e funcionamento. De salientar que, os dados do funcionamento foram tratados e analisados com o intuito de se poder encontrar possíveis relações que possam influenciar tanto os valores da electricidade, como do controlo analítico.

4.1.1.1 Electricidade

Relativamente aos dados da electricidade, como se pode verificar em 3.2.1.1, estes dados contém algumas falhas, como por exemplo dados acumulados e erros de inserção. Começou-se então por agrupar os dados com periodicidade diária, uma vez que praticamente todos os registos eram diários. A utilização deste procedimento provocou a existência de valores nulos, que correspondem aos *timesteps* em falta. De forma a colmatar a presença desses *timesteps* em falta, foram inseridas essas datas nos dados, ficando as *features* associadas a esses registos com o valor zero, em vez de um valor nulo. Atribuiu-se valor zero, ao invés de valor nulo, devido ao facto de se ter verificado que todos os indicadores contém alguns registos com valor zero, o que está incorreto, e com isso bastou fazer uma única pesquisa nos dados que continham o valor zero e transforma-los em valores nulos. Após isso, calculou-se a diferença entre o registo i com o registo $i - 1$. Essa diferença corresponde ao valor do registo i , uma vez que os dados estão acumulados. Devido a essa ação, apareceram alguns registos com valores negativos (um número muito reduzido), que correspondem ao *reset* que é realizado nos dados, como se identificou anteriormente. Para corrigir essas situações, esses registos foram colocados como valores nulos. Para corrigir os erros de inserção claramente identificados, criou-se uma "regra" onde, se o registo i fosse muito menor que registo $i + 1$, o registo $i + 1$ passa a ser nulo também. Isto não abrange todas as situações, uma vez que não se podia colocar um valor da diferença relativamente baixo e correr o risco de meter valores nulos em situações onde houve realmente picos de corrente elétrica. Por isso, para corrigir os restantes erros de inserção, nos registos onde se tinha diferenças com um número bastante elevado, confirmou-se se realmente houve erro de inserção. Por fim, os valores nulos foram preenchidos com o resultado da média dos sete dias anteriores a cada nulo [72]. Para se visualizar como ficou a variação média da electricidade ao longo dos meses de todos os anos, após o tratamento, a Figura 4.1 ilustra essa mesma variação.

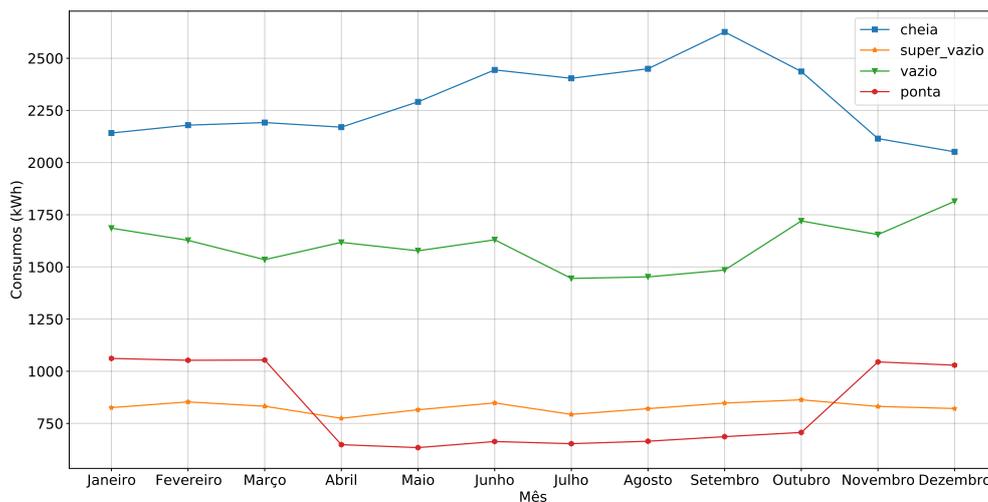


Figura 4.1: Variação média da Electricidade de Vila Real ao longo dos meses de todos os anos após tratamento.

Como se pode verificar, os valores da electricidade baixaram consideravelmente pois já não aparecem acumulados, como em 3.3, estando agora mais de acordo com o expectável. Observa-se também que, todos os indicadores da electricidade, apresentam pouca variação dos seus valores ao longo dos meses do ano, mantendo-se estáveis.

4.1.1.2 Controlo Analítico

Os dados do controlo analítico apresentam algumas lacunas, como por exemplo *timesteps* em falta, como pode ser examinado em 3.2.1.2. A primeira etapa do tratamento deste tipo dados remeteu-se a colocar os indicadores com a periodicidade mais comum, em cada um. Este procedimento levou à presença de valores nulos, que retratam os *timesteps* em falta. Por isso, foram então inseridos esses registos em falta, com valor nulo. A tarefa seguinte foi converter todos os valores iguais a zero em valores nulos, pelas razões já descritas em 3.2.1.2. Com a utilização destes mecanismos, o tratamento destes dados encontrou-se concluído, uma vez que não se identificou erros de inserção, nem valores acumulados.

4.1.1.3 Funcionamento

Como se pode apurar em 3.2.1.3, este conjunto de dados contém algumas falhas, como por exemplo *timesteps* em falta e dados acumulados. Iniciou-se por agrupar os dados com a periodicidade mais comum, em cada indicador. Através dessa ação provocou-se a existência de valores nulos, o que era esperado pois correspondem aos *timesteps* em falta, já detetados. Com isso, foram então inseridos esses *timesteps* em falta, com valor nulo. Feito isso, foi feito o calculo da diferença entre o registo i com o registo $i - 1$. Uma vez que foi identificado que os dados estão acumulados, este calculo é efetuado para colmatar essa lacuna, onde o resultado corresponde ao valor final do registo i . Apesar de os dados estarem

acumulados, existe a presença de alguns *resets* entre os registos, que depois voltam a ficar acumulados. O que acontece é que ao desacumular os dados, nessas situações a diferença calculada resulta um valor negativo, o que está incorreto. Para corrigir isso, colocou-se todos os valores negativos com valor nulo. Foi também reconhecido anteriormente que os dados apresentavam erros de inserção. Portanto, para reparar essas situações, foi criada uma "regra", onde se o registo i fosse menor que o registo $i + 1$, com uma elevada diferença, o registo $i + 1$ passa a ser nulo. Esta "regra" não abrange todas as situações que apresentam erros de inserção, uma vez que ao se colocar um valor da diferença mais baixo, podia-se estar a colocar a nulo registo onde ilustram que houve um pico nas horas de funcionamento. Por isso, foi-se verificar onde existiam as maiores diferenças entre registos e verificar manualmente se realmente existe um erro de inserção. Por último, todos os valores nulos foram preenchidos com o resultado da média dos últimos três meses (para indicadores com periodicidade mensal) ou dos últimos sete dias (para indicadores com periodicidade diária).

A Figura 4.2 ilustra a variação média do funcionamento ao longo dos meses de todos os anos, após o tratamento efetuado.

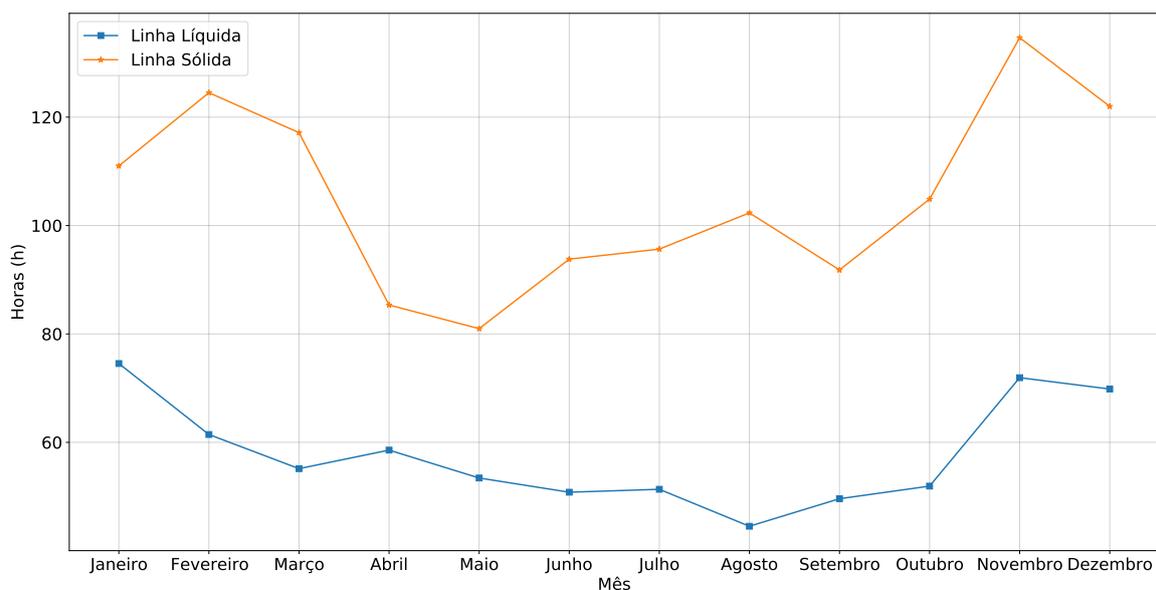


Figura 4.2: Variação média do Funcionamento ao longo dos meses de todos os anos após tratamento.

Como se pode visualizar, os valores do funcionamento baixaram significativamente, uma vez que já não estão acumulados, como em 3.4. Também é perfeitamente visível o maior número de horas utilizadas, em todos os meses do ano, por parte da *Linha Sólida*.

4.1.2 Guimarães

Relativamente à ETAR que se localiza em Serzedo, na cidade de Guimarães, teve-se acesso a dados relativos à electricidade, caudais e ao tempo. Como se pode verificar em 3.2.2.3, os dados do tempo não necessitam de tratamento. O único reparo a fazer à cerca destes dados é apenas a presença de muitos valores nulos no indicador *chuva*, mais precisamente 80% dos dados do indicador, o que não

torna fiável preencher os valores nulos, tal como em 4.1.1. Por isso, serão efetuados tratamentos nos dados de electricidade e caudais, onde abaixo serão ilustradas todas as ações e abordagens utilizadas neste processo. É importante referir que a utilização dos dados relativos aos caudais foi considerada com o propósito de identificar possíveis relações com os dados da electricidade.

4.1.2.1 Electricidade

Como se pode conferir em 3.2.2.1, estes dados apresentam algumas lacunas. A primeira tarefa efetuada foi colocar os dados com periodicidade diária, visto que praticamente todos os registos, em todos os indicadores, eram diários. Isto provocou a aparecimento de valores nulos, o que já era expectável, uma vez que se tinha identificado a existência de *timesteps* em falta. Por isso, foram então inseridos esses *timesteps* em falta nos dados, com valor nulo. Para desacumular os dados, efetuou-se a diferença entre o registo i e o registo $i - 1$, onde o resultado deste calculo irá corresponder ao valor do registo i . Como visto anteriormente, pontualmente ocorrem uns *resets* entre registos, isto provoca que da diferença calculada anteriormente resulte num valor negativo, o que está errado. Portanto, todos os valores negativos foram colocados como valores nulos. Tal como nos outros dados, previamente descritos, foi criada uma "regra" que tem como finalidade corrigir os erros de inserção existentes. Essa "regra" é aplicada quando o registo i é menor, com uma elevada diferença, que o registo $i + 1$, o que é sinonimo que esta-se perante um erro de inserção e portanto esse registo $i + 1$ passa a ter valor nulo. Como esta abordagem não abrange a totalidade dos erros de inserção, teve-se que ir manualmente às diferenças maiores e verificar se realmente houve um erro de inserção. Finalmente, todos os valores nulos foram preenchidos através da média dos registos dos últimos sete dias ao valores nulo. Para se visualizar como ficou a variação média da electricidade ao longo dos meses de todos os anos após o tratamento, foi criado um gráfico que pode ser examinado na Figura 4.3.

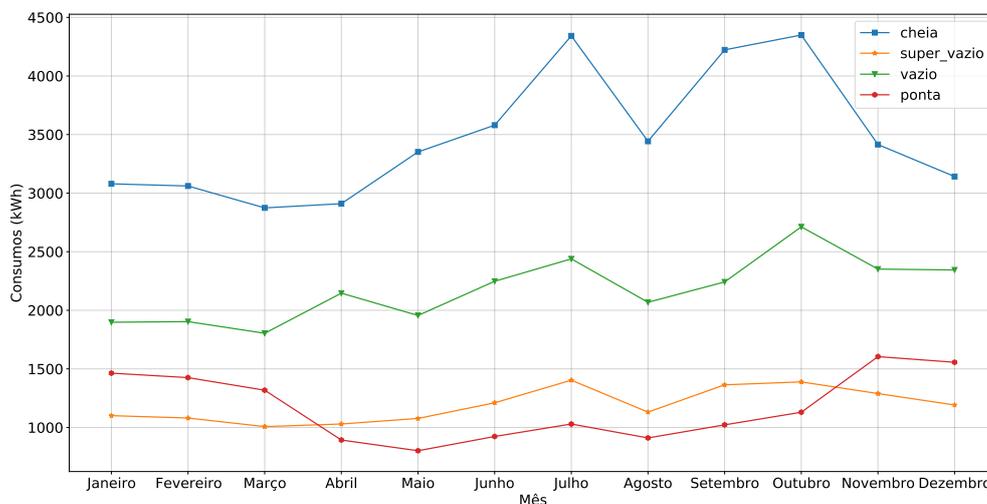


Figura 4.3: Variação média da Electricidade ao longo dos meses de todos os anos após tratamento.

Como se pode observar, os valores da electricidade baixaram drasticamente quando comparados com a análise inicial dos dados, consequência do desacumulado dos dados. Além disso, a variação já corresponde a um contexto real, o que não se sucedia antes do tratamento efetuado. Também se pode verificar que, nos meses referentes à Primavera e Outono, os consumos em todos os indicadores apresentam uma tendência crescente com o passar dos meses.

4.1.2.2 Caudais

No que se refere aos dados dos caudais, estes também apresentam falhas nos dados. O tratamento iniciou-se com o agrupar dos dados na periodicidade mais comum, em cada setor. Como aconteceu anteriormente, essa ação causou também a presença de valores em falta, correspondentes aos *timesteps* em falta, já identificados. Por isso, foram então inseridos esses registos com valor nulo. Após isso, e como se tinha visto que estes dados estavam acumulados, procedeu-se ao desacumulado dos mesmos. Para isso, utilizou-se a mesma abordagem utilizada nos outros dados, que se remete a calcular a diferença entre o registo i e o registo $i - 1$, onde o resultado passa a corresponder ao valor do registo i . Essa abordagem causa o aparecimento de valores negativos, pois também nos dados dos caudais são realizados *resets*. Portanto, todos os valores negativos passaram a valores nulos. Além disto, outra lacuna presente nestes dados é a presença de erros de inserção, onde para resolver isso foi utilizada a "regra" criada e utilizada para os outros dados, como supramencionado. Essa "regra" faz com que se sempre que o registo i seja muito menor que o registo $i + 1$, assume-se que se trata de um erro de inserção, e por isso o registo $i + 1$ passa a ter valor nulo. Ainda assim, esta "regra" não consegue alcançar todos os casos de erros de inserção, então verificou-se as maiores diferenças existentes e foi-se caso a caso verificar se era realmente um erro de inserção. Por fim, os valores nulos foram preenchidos calculando-se a média dos últimos sete dias (para setores com periodicidade diária) e dez dias (para setores com periodicidade de dois em dois dias).

A Figura 4.4 ilustra como ficou a variação média ao longo dos meses de todos os anos do setor Lamas Flotoadas, utilizado como exemplo em 3.6.

Visualizando a Figura 4.4, é perfeitamente visível que o valor médio deste setor ao longo dos meses, em todos os anos de registos, baixou radicalmente, uma vez que os dados deixaram de estar acumulados. Também se observa que, os valores apresentados por este setor sobem consideravelmente a partir do mês de Junho, obtendo uma quebra significativa em Outubro, mas que voltam a disparar até o mês de Dezembro.

4.2 Análise Exploratória dos Dados

Uma vez concluída a fase de tratamento dos dados, procedeu-se então a uma análise mais profunda aos mesmos, com a finalidade de se encontrar padrões, relações ou fenómenos que possam explicar certos comportamentos/valores apresentados pelos dados. Desse modo, nesta secção será realizada a

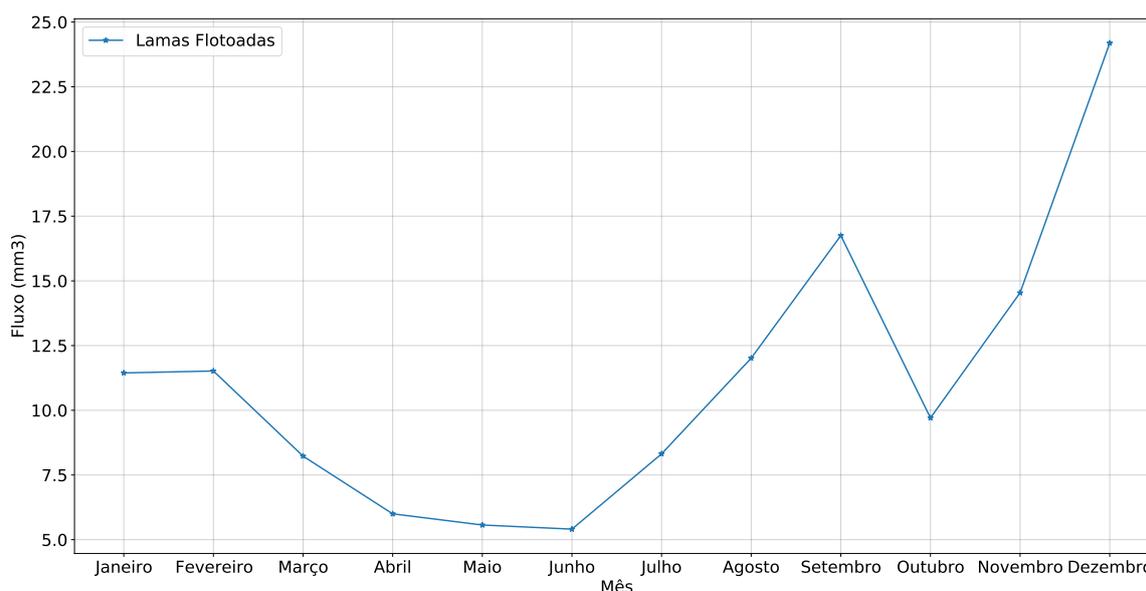


Figura 4.4: Variação média do setor Lamas Flotoadas ao longo dos meses de todos os anos após tratamento.

análise exploratória a cada tipo de dados, onde no final se pretende obter os *insights*, em cada um, para posteriormente ajudar no processo de tomada de decisão nas instalações em questão.

4.2.1 Vila Real

Para a ETAR de Vila Real foi efetuada então uma análise exploratória a cada tipo de dados, nomeadamente, dados relativos à electricidade, controlo analítico e funcionamento. Esta análise não se remeterá apenas a cada tipo de dados individualmente, mas também no relacionamento entre si com a intenção de se encontrar possíveis relações que possam ser interessantes para o estudo.

4.2.1.1 Electricidade

Relativamente aos dados da electricidade, começou-se por se aplicar novamente estatística descritiva aos dados, uma vez que nesta fase já se encontravam tratados e por isso são correspondentes à realidade. A Tabela 4.1 ilustra os resultados obtidos.

Tabela 4.1: Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Vila Real, após tratamento.

Métricas	Cheia	Ponta	Super Vazio	Vazio
Registos	1609	1607	1609	1610
Média	2283.321	829.779	826.556	1604.208
Moda	1993.179	638.387	773.107	1812.429
Mediana	2473	744.646	831	1029.5
Desvio Padrão	981.509	291.811	134.409	1238.239
Mínimo	123	4	346	49
Máximo	5298	1805	1429	6498

Como se pode examinar, os dados já apresentam valores mais realistas para este tipo de infraestruturas, obtendo-se diferenças em praticamente todas as métricas na ordem dos milhares, quando comparado com os valores apresentados antes da realização do tratamento. Como se pode analisar, os indicadores *cheia* e *vazio* apresentam valores médios bem mais elevados que os indicadores *ponta* e *super vazio*, o que reflete que esta instalação tem maiores consumos energéticos nos períodos correspondentes a estes indicadores.

Para se ter uma visão geral dos dados foi verificada a variação destes ao longo dos meses, em cada ano. Na Figura 4.5 pode ser examinada essa variação. Como se pode observar, os valores gastos de electricidade tem vindo a decrescer com o passar dos anos em praticamente todos os meses do ano, salvo algumas exceções. Entre os meses de Janeiro e Março verificou-se que, o ano de 2016, que o valor da electricidade apresentou consumos inferiores aos anos de 2017 e 2019, onde apenas o ano de 2018 se mostrou mais baixo. De ressaltar que, o ano de 2020 apresenta sempre o mesmo consumo médio, em todos os meses, onde entre os meses de Janeiro e Maio supera os anos de 2018 e 2019 ao nível de consumos de electricidade. A partir do mês de Maio, a cada ano que passa, os consumos são então cada vez menores nestes meses.

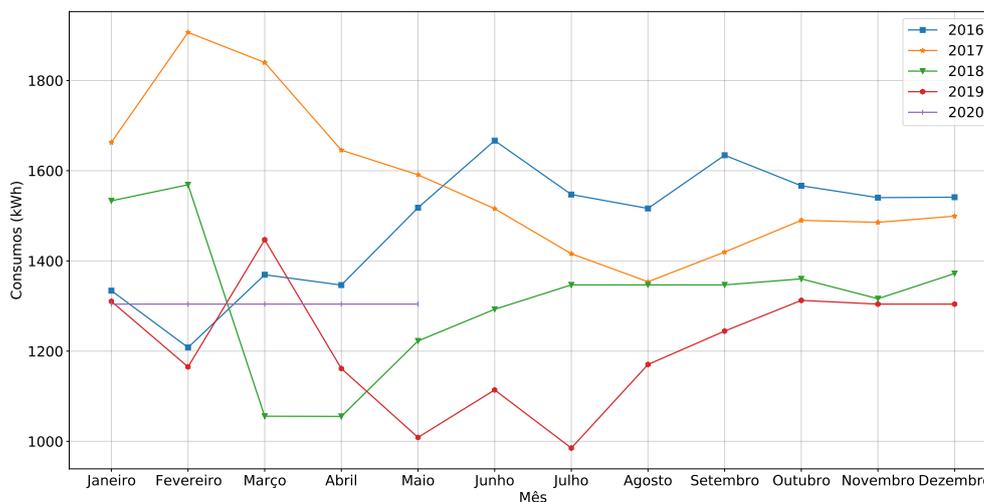


Figura 4.5: Variação Média da Electricidade de Vila Real ao longo dos meses em cada ano.

Na Figura 4.6 é possível efetuar a mesma análise, só que neste caso, por indicador. Olhando individualmente para os diferentes indicadores, verifica-se prontamente que o indicador *ponta*, em todos os anos com exceção de 2020, apresenta os seus valores bastante mais elevados entre Novembro e Março, comparativamente aos restantes meses do ano onde os valores baixam consideravelmente. Já no indicador *cheia* parece acontecer um pouco o oposto, onde os seus valores se apresentam ligeiramente mais baixos entre Novembro e Março. Nos restantes indicadores não se verifica a presença de grandes oscilações dos consumos, durante os meses de todos os anos. De ressaltar que em todos os indicadores, os consumos, a cada ano que passa tendem a baixar em praticamente todos os meses do ano, com exceção

de 2020, que principalmente em *super vazio* e *vazio* existe uma tendência onde os consumos voltam a subir neste ano.

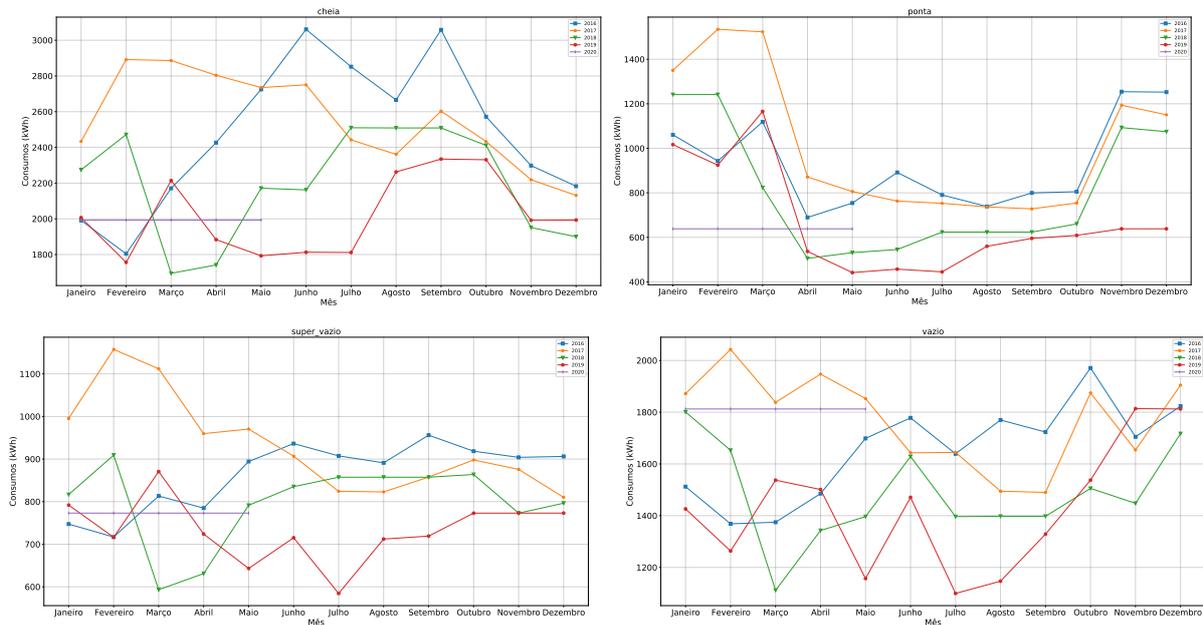


Figura 4.6: Variação Média da Electricidade de Vila Real ao longos dos meses em cada ano por indicadores.

De modo a perceber se com o decorrer dos anos os consumos gerais diminuem, foi calculada a média dos consumos por ano. A Figura 4.7 expõem essa mesma informação. Como se pode observar, 2017 é o ano que contem os consumos mais elevados seguido do ano de 2018. Verifica-se também que, a partir de 2017, os consumos apresentam uma tendência negativa até ao ano de 2019, onde em 2020 voltam a subir.

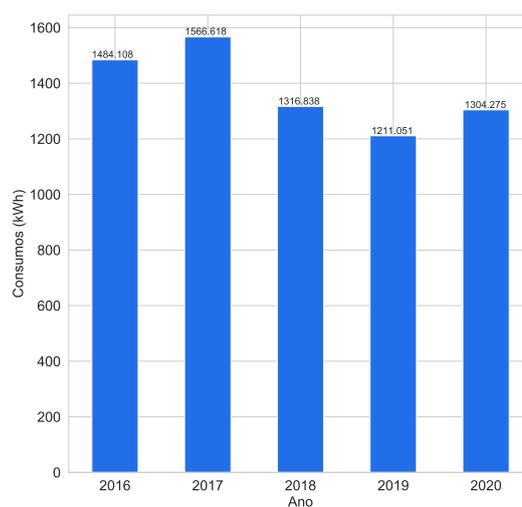


Figura 4.7: Valores médios da Electricidade de Vila Real em cada ano.

Na Figura 4.8 é elaborada a mesma abordagem, só que aplicada a cada indicador individualmente. Olhando então para os indicadores, o ano de 2017 apresenta-se como o ano que contém a média de

consumos mais elevados em todos os indicadores, com exceção em *vazio*, onde o ano de 2020 supera o ano de 2017. Também se verifica que nos indicadores *cheia* e *ponta* a partir do ano de 2017 a média de consumos anuais tem tendência decrescente, já em *super vazio* e *vazio* essa tendência também ocorre mas em 2020 a média volta a subir.

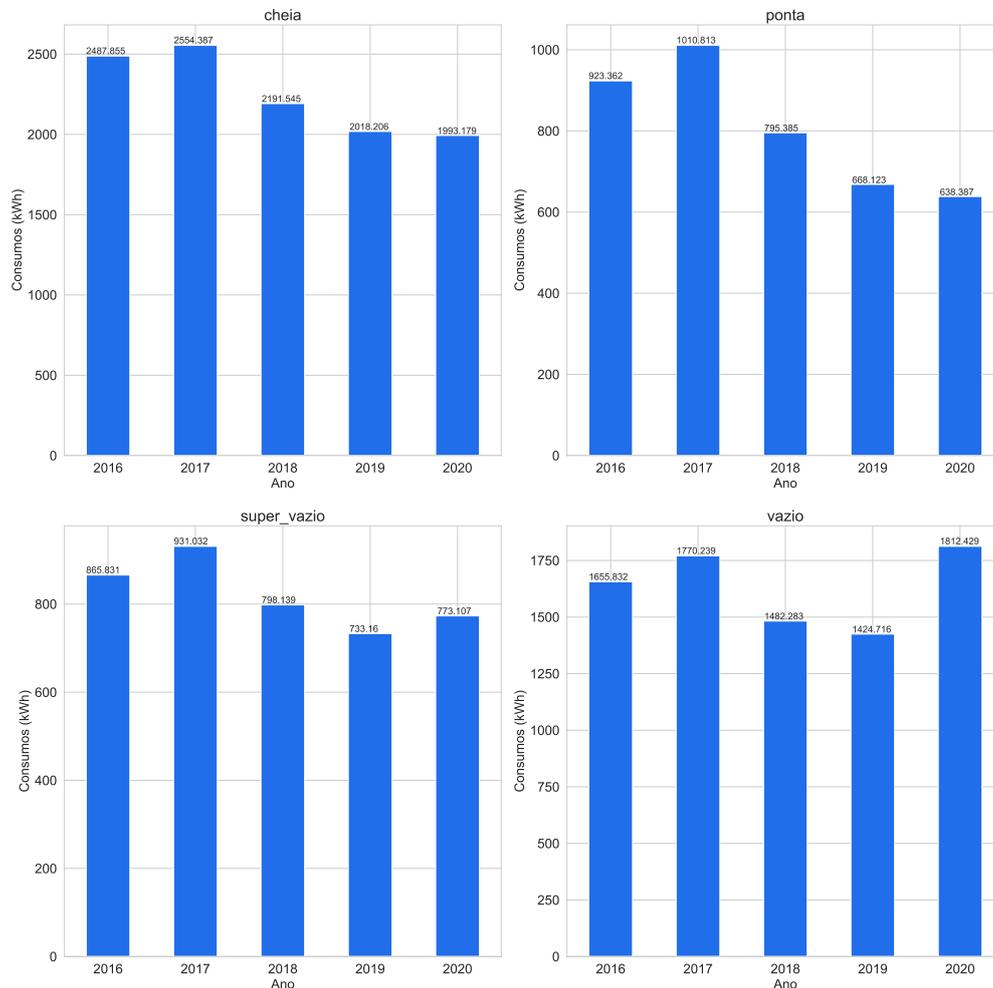


Figura 4.8: Valores médios da Electricidade de Vila Real em cada ano por indicadores.

Com o intuito de se obter mais informações sobre os dados, foram criadas duas variáveis, sendo elas trimestre e estação do ano, que posteriormente serão utilizadas também nos outros tipos de dados. Relativamente à variável trimestre, corresponde aos quatro trimestres anuais, já a variável estação do ano é relativa às diferentes estações do ano.

Para se compreender como são os consumos em cada trimestre, de todos os anos, foi calculada a média dos mesmos. A Figura 4.9 expõem então os valores obtidos. Como se pode visualizar, em praticamente todos os anos o 1º trimestre assume os valores mais altos de consumos, com exceção do ano de 2016 onde é o trimestre com menor consumo. De resto, vê-se a existência de um equilíbrio nos consumos durante os trimestres, ou seja, apresentam valores idênticos não existindo grandes divergências entre os mesmos.

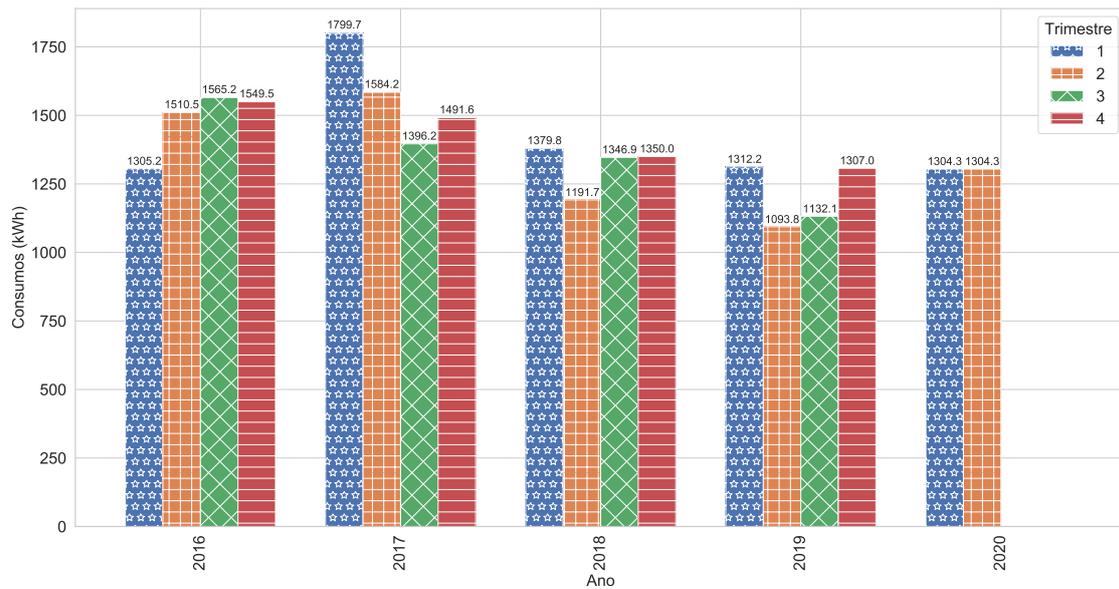


Figura 4.9: Valores médios da Electricidade de Vila Real em cada trimestre.

Já a Figura 4.10 mostra a mesma informação mas de forma individual. Analisada a figura, observa-se que *cheia* assume os seus maiores valores, em praticamente todos os anos, no 3º trimestre. De salientar também, um certo equilíbrio existente entre os valores dos trimestres durante os anos. No que toca ao indicador *ponta* ressalta à vista a grande discrepância nos valores relativos aos 1º trimestres comparando com os restantes, principalmente no ano de 2017. O 4º trimestre também assume valores significativamente elevados nos anos em que apresenta registos. Relativamente ao indicador *super vazio* salientar os 1º e 2º trimestres do ano de 2017 que exibem os valores mais elevados em todos os anos deste indicador. Os restantes indicadores demonstram um equilíbrio nos seus valores, em todos os trimestres. Por fim, no indicador *vazio* evidencia-se os valores exibidos pelo 1º e 2º trimestres no ano de 2017 e 2020, que estão entre os valores mais elevados apresentados pelos trimestres neste indicador. De ressaltar também os valores assumidos pelo 3º trimestre em todos os anos em que tem registo, sendo inclusive o trimestre com os valores mais elevados em 2016, 2018 e 2019. Do ponto de vista geral, no ano de 2020 os dois trimestres correspondentes assumem um total equilíbrio nos seus valores, nos diferentes indicadores existentes.

No que diz respeito à variável estação do ano, tal como para a variável trimestre, foi calculada a média dos valores nas estações do ano, em todos os anos, onde a Figura 4.11 ilustra essa mesma informação. Como se pode observar, em todos os anos, com a exceção do ano de 2016, o Inverno é responsável pelos maiores consumos energéticos da infraestrutura. Na Primavera e Verão, nos anos de 2017 e 2019, nota-se um claro decréscimo nos consumos, que posteriormente voltam a subir no Outono. Já em 2016 e 2018 repara-se num aumento nos consumos a partir da Primavera até ao Outono.

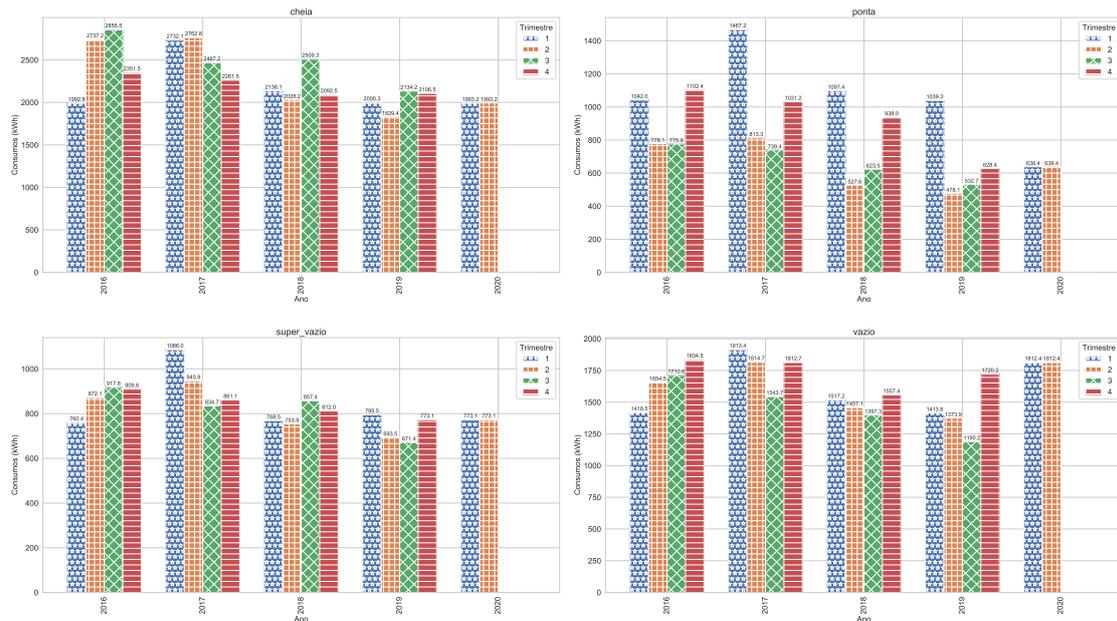


Figura 4.10: Valores médios da Electricidade de Vila Real em cada trimestre por indicador.

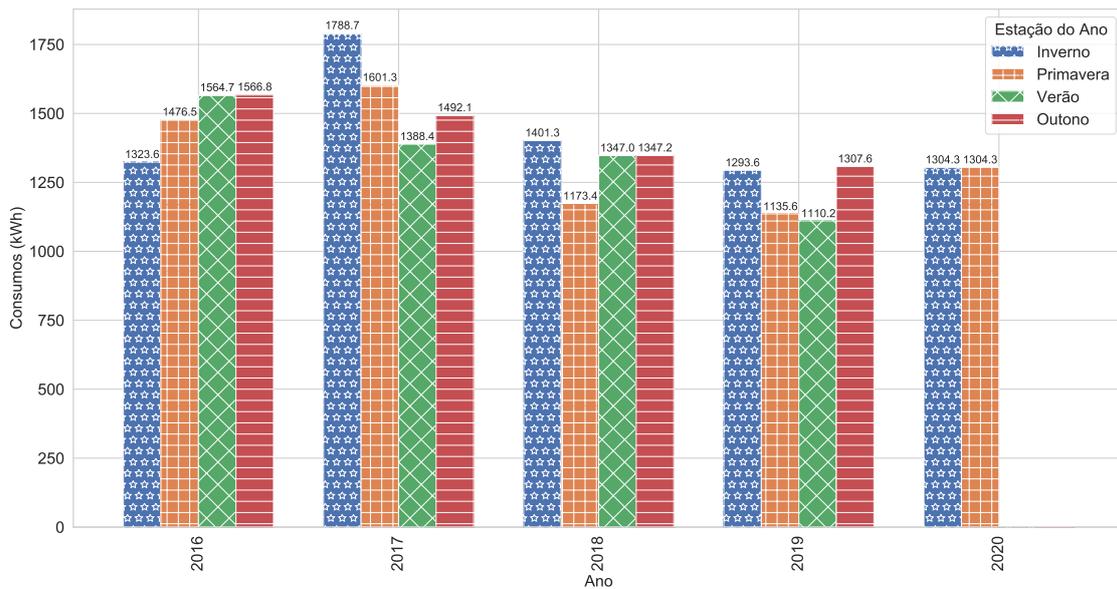


Figura 4.11: Valores Médios da Electricidade de Vila Real em cada estação do ano.

Na Figura 4.12 pode-se analisar, individualmente, os valores médios da electricidade nas diferentes estações do ano, em todos os anos, por cada indicador. Em relação ao indicador *cheia* o Verão exibe-se como a estação do ano com os valores mais elevados em 2016, 2018 e 2019. A Primavera assume também valores elevados nos diferentes anos, sobretudo em 2017, onde é a estação do ano com os maiores valores. No que diz respeito ao indicador *ponta*, o Inverno é claramente a estação do ano que apresenta os valores mais elevados em todos os anos. O Outono também assume igualmente valores elevados em comparação com a Primavera e o Verão, onde estes assumem valores muito idênticos em todos os anos. Já no indicador *super vazio* salienta-se o Inverno e Primavera no ano de 2017, pois assumem os valores mais elevados em todos os anos neste indicador. De resto, existe uma estabilidade

nos valores durante os anos, em todos as estações do ano. Por último, no indicador *vazio* o Inverno volta a apresentar os valores mais elevados, sobretudo em 2017 e 2020. O Outono assume igualmente valores bastante altos, especialmente em 2016 e 2019, sendo inclusive a estação do ano com maiores valores nesses dois anos. No geral, no ano de 2020 os valores do Inverno e Primavera apresentam-se com alto equilíbrio em todos os indicadores.

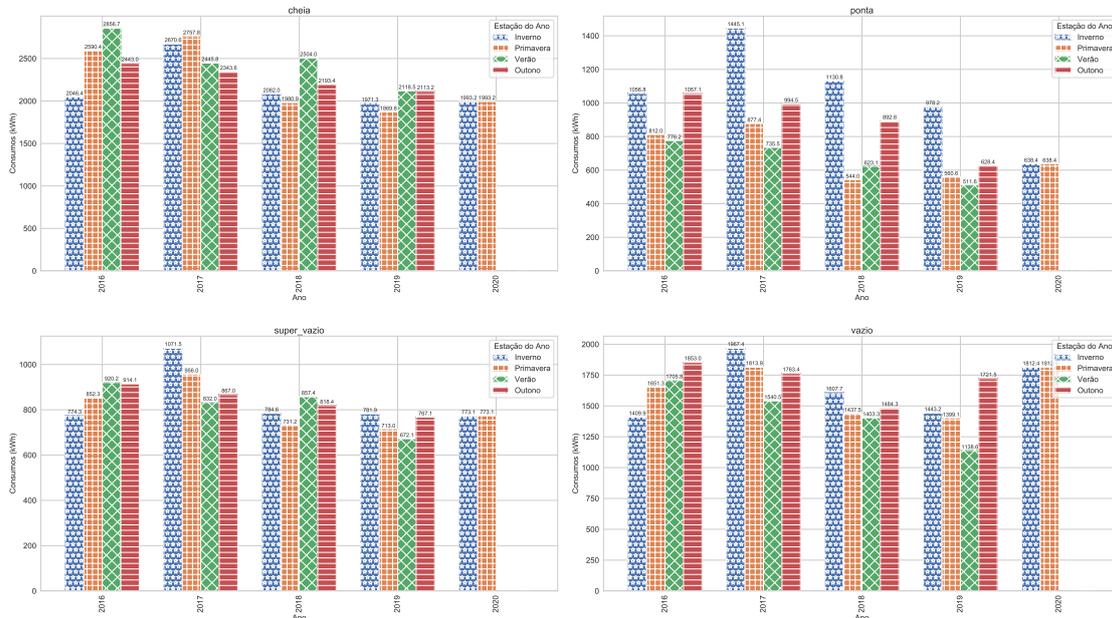


Figura 4.12: Valores Médios da Electricidade de Vila Real em cada estação do ano por indicador.

Por fim, de forma a identificar possíveis relações entre os indicadores e perceber de que forma se relacionam, foi calculada a correlação entre os mesmos. Para esse efeito, foi aplicado o teste de *Shapiro-Wilk* para verificar se os dados seguiam uma distribuição normal. Como se obteve $p < 0.05$, assume-se que os dados não seguem uma distribuição normal. Por isso, para o cálculo das correlações foi utilizado o método de Coeficiente de Correlação de Spearman, que avalia relações monótonas, sejam elas lineares ou não.

A Figura 4.13 exibe então os resultados obtidos das correlações calculadas entre os indicadores da electricidade. Como se pode observar, existem relações um pouco interessantes entre os indicadores. Começando pela única correlação moderada existente, entre o indicador *cheia* com o *vazio*, trata-se de uma dependência negativa, ou seja, quando um cresce o outro diminui. Depois tem-se um conjunto de correlações fracas, que significa que os indicadores em questão tem pouca interferência um no outro. Portanto, conclui-se que os indicadores da electricidade, salvo a exceção dos indicadores *cheia* com *vazio*, influenciam-se pouco entre eles. Para se compreender melhor os intervalos de valores dos diferentes tipos de graus de correlações, na Tabela 4.2 é possível verificar os graus existentes na interpretação dos resultados do cálculo de correlações.

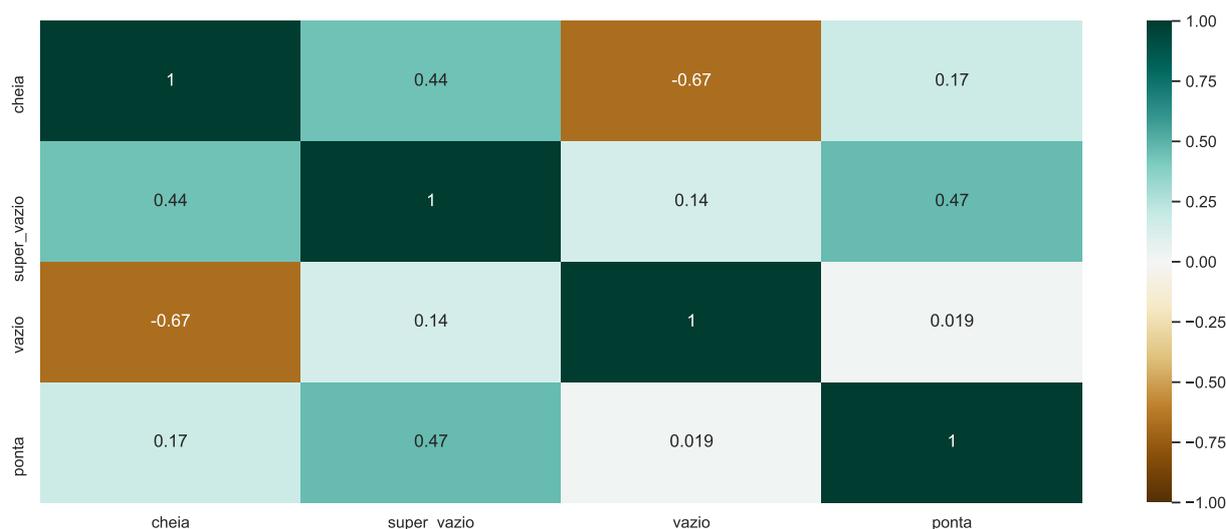


Figura 4.13: Correlações entre os indicadores da Electricidade de Vila Real.

Tabela 4.2: Interpretação dos valores das correlações.

Intervalos	Descrição
0 a 0.3 ou 0 a -0.3	Correlação desprezível
0.3 a 0.5 ou -0.3 a -0.5	Correlação fraca
0.5 a 0.7 ou -0.5 a -0.7	Correlação moderada
0.7 a 0.9 ou -0.7 a -0.9	Correlação forte
0.9 a 1 ou -0.9 a -1	Correlação muito forte

4.2.1.2 Controlo Analítico

No que diz respeito aos dados do controlo analítico, foi novamente aplicada estatística descritiva aos dados uma vez que nesta fase os mesmos se encontram tratados, o que significa que nesta ponto, representam a realidade. Na Tabela 4.3 é possível então averiguar os resultados obtidos.

Como se pode examinar, todos os indicadores, em todos os setores, deixaram de ter como valor mínimo o zero, o que acontecia inicialmente antes do tratamento. Outra mudança significativa é o decréscimo, em praticamente todos os indicadores, do valor do desvio padrão, o que significa que os valores passaram a estar menos dispersos em torno da média de cada um. Também se registaram diferenças consideráveis nas outras métricas calculadas, sobretudo no valor da média e moda, onde se identificam diferenças mais significativas quando comparado com os dados iniciais.

Para se proceder a uma análise mais específica de cada indicador, foi calculada inicialmente a variação média de cada um ao longo dos meses, em todos os anos de registo, nos respetivos setores. A Figura 4.14 expõem essa mesma informação, mas como se contém 16 indicadores diferentes a colocação de todas as imagens tornaria este tópico massivo, portanto decidiu-se selecionar 3 indicadores e alguns dos restantes encontram-se nos anexos II. Assim sendo, os indicadores a serem abordados neste ponto serão o *azoto total*, *CBO* e *CQO*. Como se pode observar na Figura 4.14, os indicadores seguem o mesmo padrão, apresentando os menores valores, em todos os meses, no setor efluente tratado, o que tem toda

Tabela 4.3: Resultado da Estatística Descritiva aplicada aos dados do Controlo Analítico, após tratamento.

Setor	Indicador	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo	Periodicidade
Afluente Bruto	Azoto Total	mg/l	95	70.521	68	71.3	20.964	20.4	145	Semanal
	CBO	mg/l	47	471.914	520	411	239.089	80	1050	Quinzenal
	CQO	mg/l	95	740.609	400	719	276.478	179	1568	Semanal
	Fósforo Total	mg/l	95	7.771	6.2	7.67	2.481	2.927	14.5	Semanal
	Ph	Escala de Sorensen	222	7.561	7.6	7.573	0.461	5.48	8.94	Semanal
	SST	mg/l	95	274.173	515	245	153.878	65	820	Semanal
Câmara de Degaseificação	Amónia	mg/l	96	7.298	0.015	3	11.748	0.013	58.5	Semanal
	Nitratos (NO3)	mg/l	83	5.042	9.1	4.93	0.020	0.020	19.8	Semanal
	SST	mg/l	95	348.087	195	276.38	243.255	30	1380	Semanal
Efluente Tratado	Amónia	mg/l	331	10.811	0.015	7.685	9.471	0.015	45.80	Dois em Dois dias
	Azoto Total	mg/l	96	16.668	11.4	13.425	9.504	3.140	54	Semanal
	CBO	mg/l	47	4.455	2	3	3.006	1	12	Quinzenal
	CQO	mg/l	96	29.499	22	28.9	9.046	13.100	53.2	Semanal
	Fósforo Total	mg/l	96	1.446	0.5	1.017	1.158	0.211	6.13	Semanal
	Nitratos (NO3)	mg/l	331	4.499	0.5	4.19	2.699	0.443	26.3	Dois em Dois dias
	Ortofósforos	mg/l	95	1.470	0.5	0.9	1.631	0.015	7.99	Semanal
	Ph	mg/l	228	6.981	6.9	6.984	0.408	5.050	8.4	Semanal
	SST	mg/l	96	5.670	3	4	5.135	1	33	Semanal
Entrada Reator Biológico	Azoto Total	mg/l	96	61.989	21.3	64.7	21.401	19.2	126	Semanal
	CBO	mg/l	47	278.383	240	240	179.766	20	900	Quinzenal
	Fósforo Total	mg/l	96	7.315	9.1	7.575	179.766	2.32	14.1	Semanal
	Ph	Escala de Sorensen	225	7.532	7.3	7.534	0.415	5.72	9.560	Semanal
	SST	mg/l	96	100.444	60	80	69.938	16	480	Semanal
Lamas Biológicas 1	Amónia	mg/l	292	5.409	0.015	2.005	8.427	0.005	49.9	Dois em Dois dias
	IVL	mg/l	84	172.817	139	170.7	38.510	98.8	258.3	Semanal
	Nitratos (NO3)	mg/l	292	4.176	3.5	2.383	3.118	0.078	21.7	Dois em Dois dias
	Oxigénio	mg/l	780	1.930	3.19	1.489	1.334	0.07	6.65	Dois em Dois dias
	Ph	Escala de Sorensen	228	6.724	6.9	0.366	1.045	5.67	8.3	Semanal
	SST	mg/l	74	5248.240	3880	5050	1025.367	3600	8100	Semanal
	SSV	mg/l	37	4345.585	3120	4178	849.819	2950	6570	Quinzenal
Lamas Biológicas 2	Amónia	mg/l	331	4.863	0.5	2.03	7.063	0.015	35.4	Dois em Dois dias
	IVL	mg/l	96	174.907	103.4	165.568	45.634	103.4	312.4	Semanal
	Nitratos (NO3)	mg/l	331	4.753	0.5	4.85	2.969	0.014	23.1	Dois em Dois dias
	Oxigénio	mg/l	780	2.259	0.53	2.041	1.386	0.1	8.7	Dois em Dois dias
	Ph	Escala de Sorensen	228	6.662	6.7	6.652	0.365	5.65	8.3	Semanal
	SST	mg/l	85	5114.191	5103.333	5156	1157.033	6.68	7980	Semanal
	SSV	mg/l	43	3905.685	710	4360	1533.086	710	6710	Quinzenal
Lamas Biológicas Espessadas	ST	mg/l	47	25.821	12.318	24.252	8.509	12.318	46.582	Quinzenal
Lamas Biológicas Recirculadas	ST	mg/l	95	9.602	4.234	9.406	2.945	4.234	28.954	Semanal
Lamas Digeridas 1	Ácidos Gordos Voláteis	mg/l	95	2958.222	200	200	12716.189	40	65250	Semanal
	Alcalinidade	mg/l	95	2231.772	1960	2000	1440.634	80	10765	Semanal
	Amónia	mg/l	95	644.023	515	655	180.482	39.25	1055	Semanal
	Fósforo Total	mg/l	95	366.805	490	353.5	95.406	125.5	765	Semanal
	Ph	Escala de Sorensen	204	7.321	8.4	7.33	0.506	5.215	8.4	Semanal
	ST	mg/l	95	14.708	17.71	14.881	3.552	2.690	26.710	Semanal
	SV	mg/l	95	10.176	12.534	10.076	2.567	1.680	17.424	Semanal
Lamas Digeridas 2	Ácidos Gordos Voláteis	mg/l	95	3013.234	240	280	12643.560	12643.560	66150	Semanal
	Alcalinidade	mg/l	95	2315.410	1800	2080	1222.789	1120	8050	Semanal
	Amónia	mg/l	95	709.525	820	721.236	178.754	33.350	1100	Semanal
	Fósforo Total	mg/l	95	379.014	370	370	94.180	123.250	630	Semanal
	Ph	Escala de Sorensen	220	7.167	6.18	7.135	0.546	4.670	8.5	Semanal
	ST	mg/l	95	16.530	18.058	16.672	2.149	10.351	26.224	Semanal
	SV	mg/l	95	11.821	12.674	12.172	1.658	6.994	17.981	Semanal
Lamas Mistas	ST	mg/l	95	25.735	3.238	26.622	9.521	3.238	52.792	Semanal
	SV	mg/l	73	22.558	17.056	21.126	7.125	7.970	41.658	Semanal
Lamas Primárias 1	ST	mg/l	95	34.045	0.408	31.818	14.785	0.452	121.428	Semanal
	SV	mg/l	73	28.831	35.414	27.464	13.089	0.206	103.81	Semanal
Lamas Primárias 2	ST	mg/l	95	31.059	0.408	29.288	15.071	0.408	111.592	Semanal
	SV	mg/l	73	26.000	0.236	24.638	13.723	0.236	99.821	Semanal
Poço Escorrências	Azoto Total	mg/l	48	190.543	138	190.5	63.434	44	315	Quinzenal
	CQO	mg/l	48	1797.077	3147.5	1580.174	1094.574	294	4200	Quinzenal
	Fósforo Total	mg/l	48	32.651	33	30.044	17.909	17.909	138	Quinzenal
	SST	mg/l	48	1149.703	1285	981.171	875.340	19.7	3550	Quinzenal
Tanque Anóxico Tanque 1	Nitratos (NO3)	mg/l	83	0.928	0.5	0.526	1.705	0.153	13	Semanal
	Oxigénio	mg/l	131	0.357	0.2	0.368	0.331	0.03	3.56	Semanal
Tanque Anóxico Tanque 2	Nitratos (NO3)	mg/l	95	0.822	0.5	0.5	1.446	0.137	12.9	Semanal
	Oxigénio	mg/l	113	0.333	0.4	0.28	0.320	0.02	3.27	Semanal

a lógica, uma vez que neste setor as *AR* terminam todo o processo de tratamento. Outra característica visível é relativa ao setor poço escorrências, onde os indicadores *azoto total* e *CQO* apresentam os seus valores mais elevados durante todo o ano. Também se verifica que nos indicadores *azoto total* e *CBO* os seus valores, com exceção no setor *efluente tratado*, exibem um crescimento acentuado sobretudo entre os meses de Maio e Julho.

De forma a perceber como o valor destes indicadores varia ao longo dos anos, foi calculada a variação média de cada um, em todos os setores, durante os meses nos diferentes anos de registo. Essa variação pode ser visualizada na Figura 4.15.

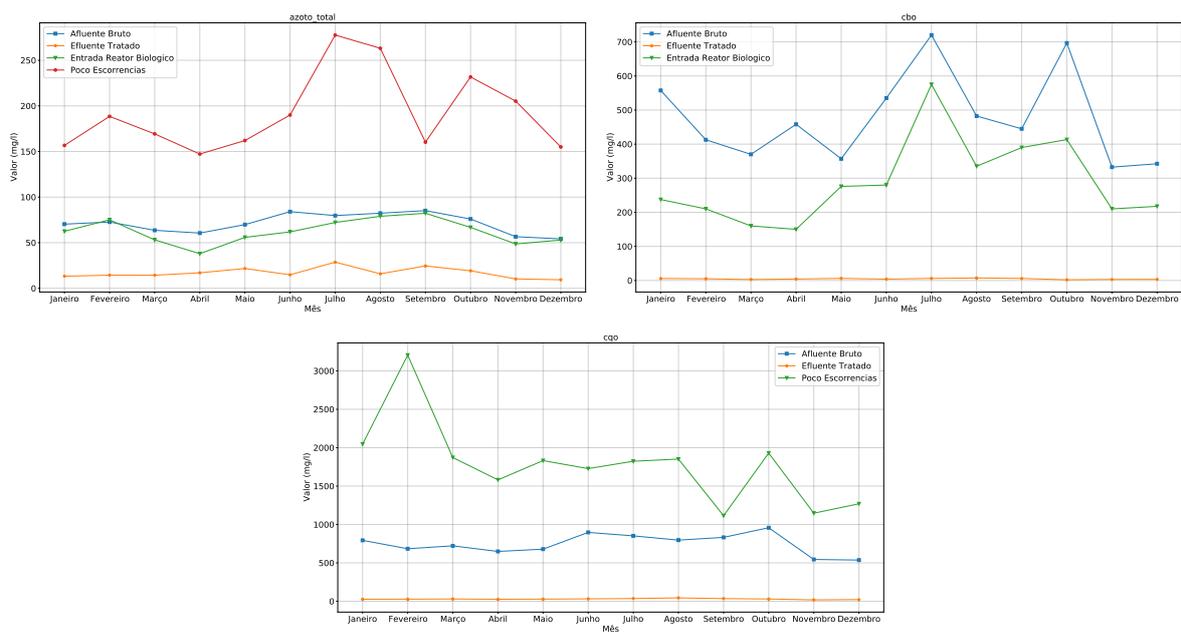


Figura 4.14: Variação Média de alguns indicadores do Controlo Analítico em cada mês do ano.

Examinada a figura, verifica-se desde logo que no mês de Julho no ano de 2019 coincidiu em todos os indicadores, e em praticamente todos os setores, com um pico nos valores, apresentando nesse mês valores bastante altos quando comparado com os restantes. Na mesma linha, e sabendo que no ano de 2018 apenas se contém dados a partir do mês de Agosto, verifica-se que nos meses de Agosto, Setembro e Outubro os valores são mais elevados, mas que com o aproximar do fim do ano os mesmos tendem a baixar consideravelmente. Relativamente ao ano de 2020, onde se abrange dados de Janeiro a maio, Fevereiro revela-se como o mês com os valores mais altos, com exceção do indicador *CBO* no setor efluente tratado, onde Janeiro se destaca como o mês que abarca os maiores valores. No ponto de vista geral, em todos os indicadores e respetivos setores, os valores tendem a ser menores no fim do ano, registando-se descidas acentuadas em alguns deles.

De modo a compreender como se apresentam os valores durante os trimestres, foi calculada a variação média dos indicadores por trimestres em todos os anos. Essa variação pode ser consultada na Figura 4.16.

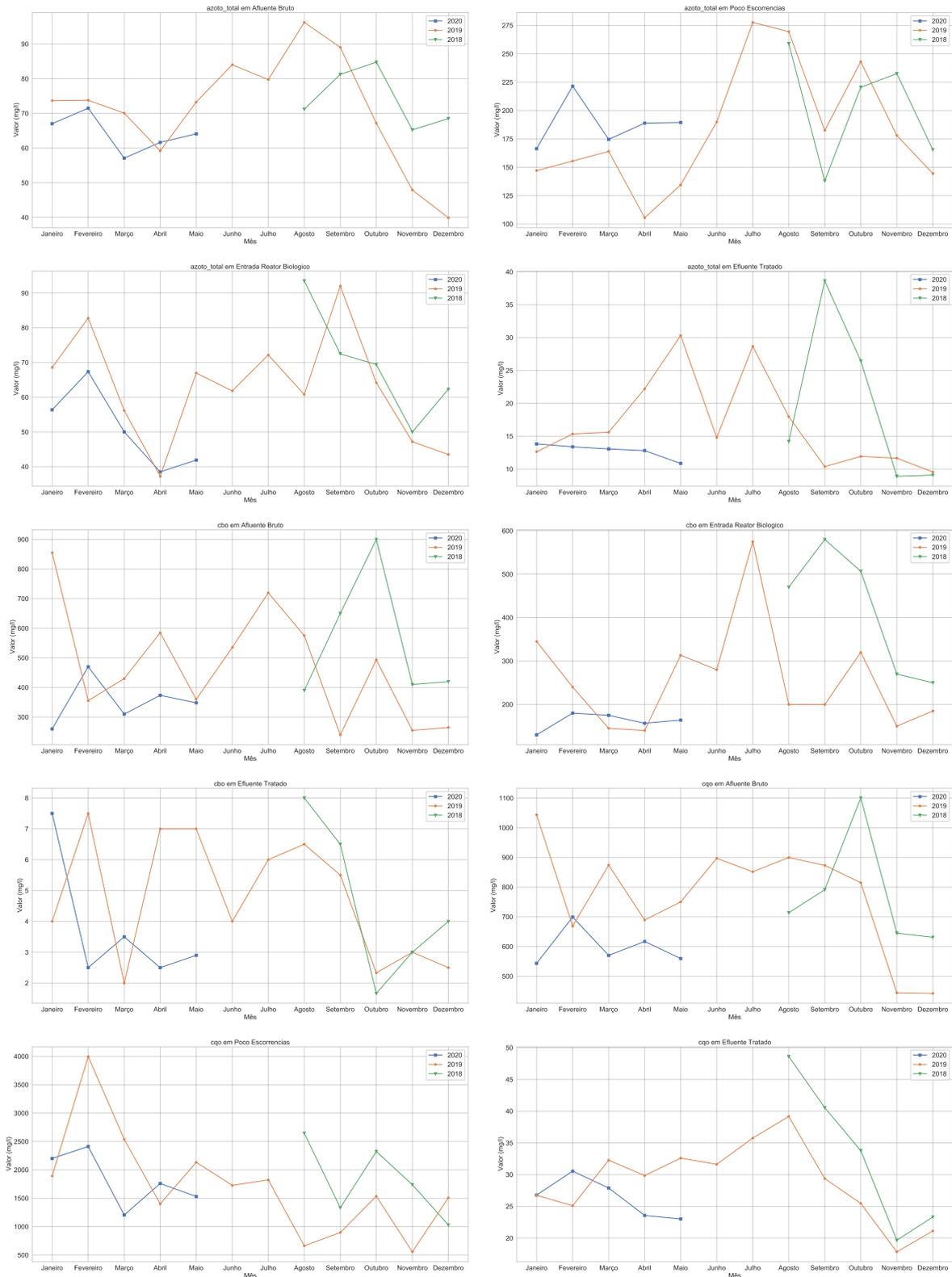


Figura 4.15: Variação Média de alguns indicadores do Controlo Analítico por mês em cada ano.

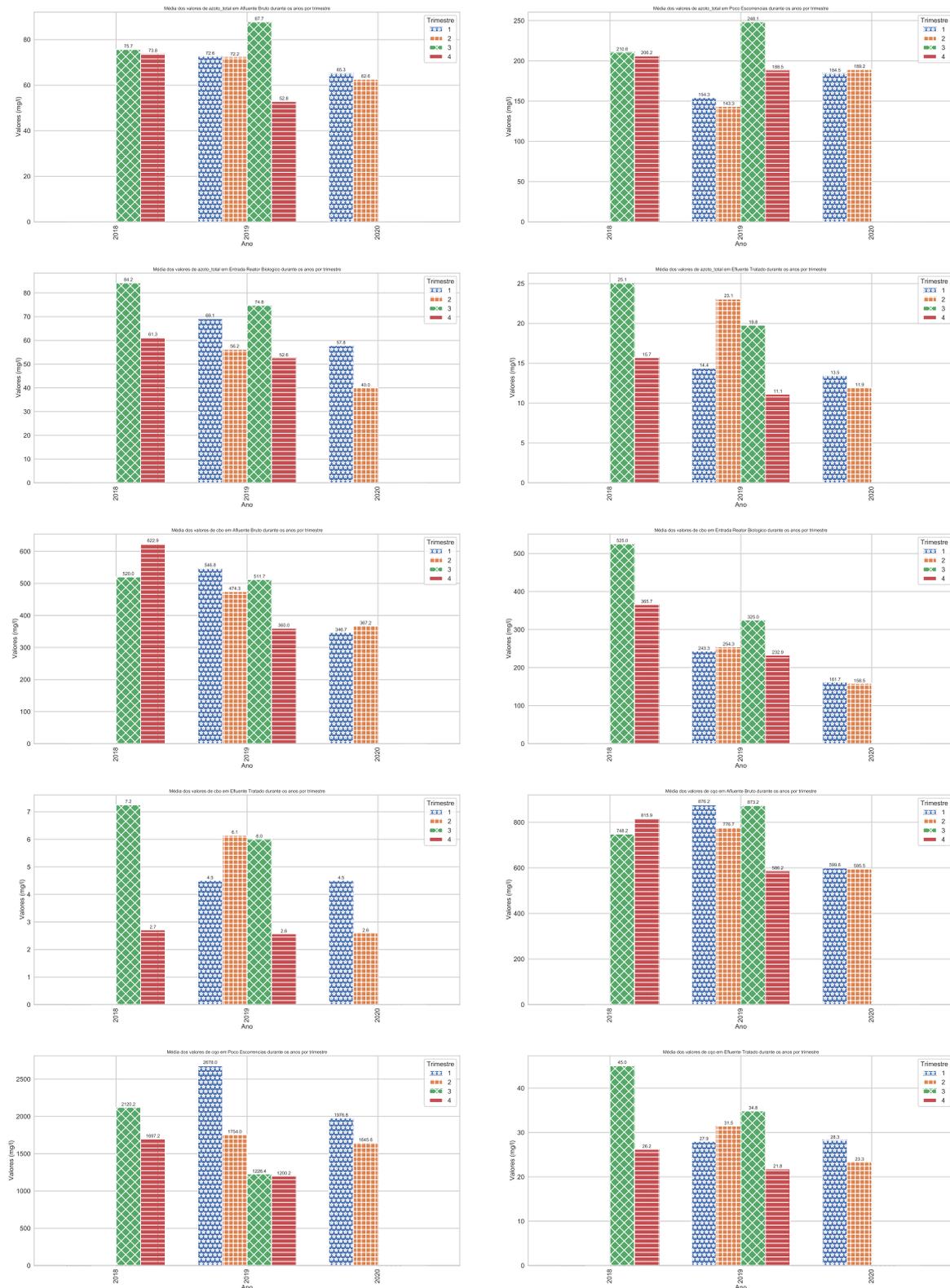


Figura 4.16: Variação Média de alguns indicadores do Controlo Analítico por trimestres.

Perscrutada a figura, o 3º trimestre assume-se, nos anos de 2018 e 2019, como o trimestre com os valores mais elevados, em quase todos os setores. Em contrapartida, o 4º trimestre apresenta os valores mais baixos nesses anos, com exceção no indicador *azoto total* no *poço escorrências*, onde o 1º e 2º trimestres contêm valores inferiores. De ressaltar que o 1º trimestre, no ano de 2020, exibe os valores

mais elevados em quase todos os setores analisados.

Já na Figura 4.17 é ilustrada a variação média dos indicadores por estação do ano, em todos os anos de registo.

Como se pode observar na figura, nos anos de 2018 e 2019, o Verão é a estação do ano que assume os maiores valores em quase todos os indicadores e respetivos setores. Relativamente a 2019, em todos os indicadores e nos seus diferentes setores, o Outono é a estação do ano que apresenta os valores mais baixos, apenas com exceção do indicador *azoto total* no *poço escorrências*, onde a Primavera e o Inverno exibem valores inferiores. Em 2020, os valores apresentados pelo meses de Inverno revelam valores mais elevados quando comparado com os meses da Primavera, exceto no indicador *CBO* em *afluente bruto*, onde a Primavera apresenta valores superiores ao Inverno.

Por último, calculou-se as correlações entre os diferentes indicadores e respetivos setores com o intuito de verificar possíveis relações entre os mesmos. Tal como em 4.2.1.1, foi também aplicado o teste de *Shapiro-Wilk* para identificar se os dados seguiam uma distribuição normal. O resultado obtido desse teste foi um $p < 0.05$, logo os dados não seguem essa mesma distribuição. Com isso, para se conseguir extrair ao máximo as informações sobre as correlações existentes, foi aplicado o método Coeficiente de Correlação de Spearman. A Figura 4.18 expõem então os resultados obtidos.



Figura 4.18: Correlações entre os indicadores e respetivos setores do Controlo Analítico.

Analisada a figura, verificou-se a existência de boas relações entre os indicadores analisados nos respetivos setores, sendo todas elas positivas, ou seja, sempre que um aumenta o outro também vai aumentar e vice-versa. Olhando para as correlações salienta-se a existência de 3 correlações fortes, sendo todas elas no setor *afluente bruto*, entre o indicador *azoto total* com o *CBO*, *CBO* com *CQO* e entre *azoto total* com *CQO*. Verificou-se também a existência de um conjunto elevado de correlações moderadas nos dados. De realçar que em 4.2 é possível examinar os diferentes graus existentes na interpretação dos resultados do cálculo de correlações.

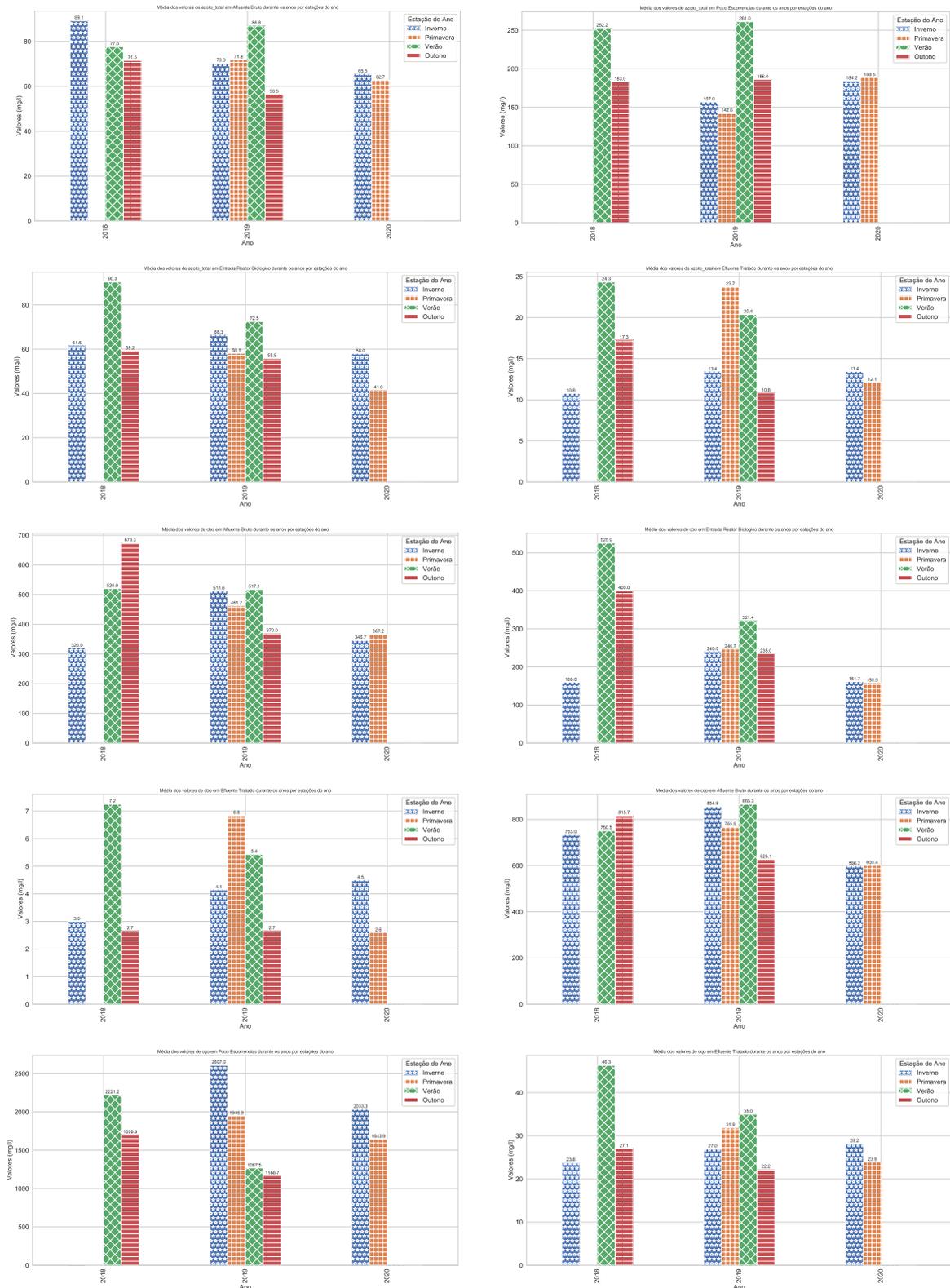


Figura 4.17: Variação Média de alguns indicadores do Controle Analítico por estações do ano.

4.2.1.3 Funcionamento

Em relação aos dados do funcionamento, uma vez que nesta fase os mesmos se encontram tratados, foi novamente aplicada estatística descritiva aos dados para se identificar as primeiras características dos mesmos, onde a Tabela 4.4 expõem então os resultados obtidos.

Analisada a tabela, verifica-se desde logo diferenças enormes, em todas as métricas calculadas, comparativamente à Tabela 3.4 onde os dados ainda não tinham sido tratados. Os valores apresentados após o tratamento já são mais realistas, uma vez que era impossível um indicador conter, por exemplo, como média de horas um valor na ordem dos milhares, como acontecia anteriormente. De salientar que no setor linha líquida o indicador *Tratamento Primário Ponte Raspadora Sedipac 1* é o que contém o número médio de horas de funcionamento mais elevado, já na linha sólida é o indicador *Estabilização Biológica Ventilador Gasometro*.

Para se ter uma melhor visão dos dados do funcionamento, inicialmente calculou-se a variação média geral dos mesmos durante os meses, em todos os anos observado no *dataset*. A Figura 4.19 expõem essa mesma variação média.

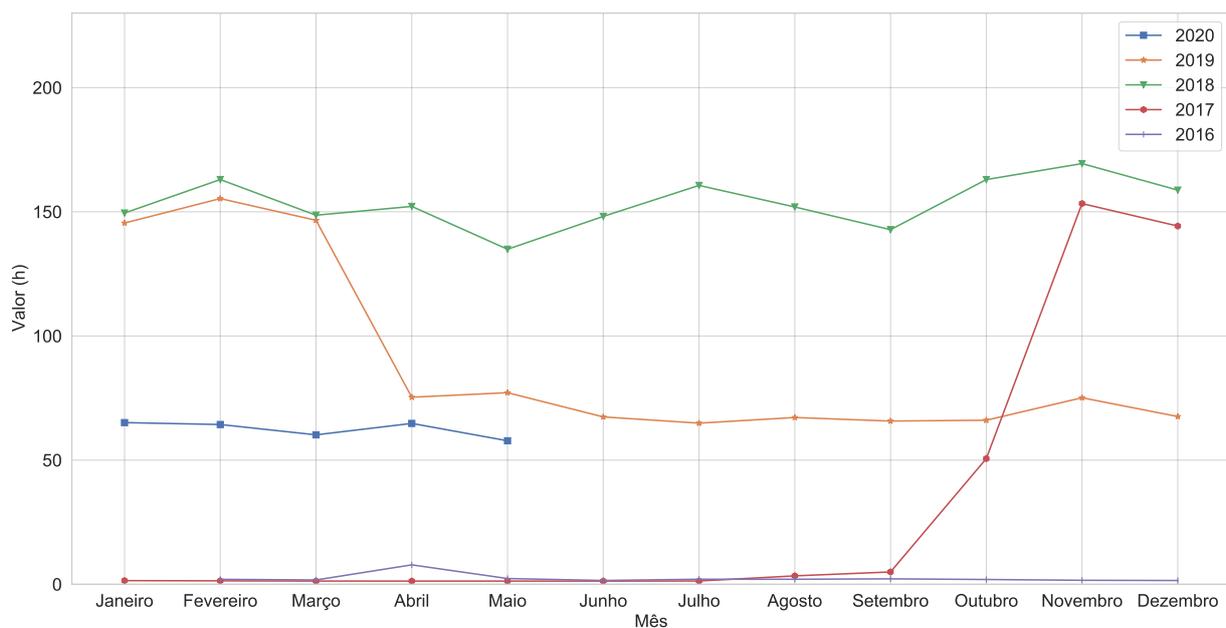


Figura 4.19: Valores médios do Funcionamento por mês em cada ano.

Como se pode examinar na figura, 2018 é, em todos meses, o ano que contém os valores mais elevados, seguindo-se o ano de 2019. Do lado oposto, o ano de 2016 foi claramente o ano em que se utilizou menos horas de funcionamento. Verificou-se também que, em todos os anos, existe uma certa estabilidade nas horas de funcionamento utilizadas durante os meses dos anos, salvo algumas exceções. Uma delas é relativa ao ano de 2017 que se mantém estável de Janeiro a Setembro, mas posteriormente tem um crescimento abrupto durante os meses de Outubro e Novembro mantendo-se nesses valores até ao fim do ano. A outra exceção é alusiva ao ano de 2019, onde de Janeiro a Março estão refletidos os

Tabela 4.4: Resultado da Estatística Descritiva aplicada aos dados do Funcionamento, após tratamento.

Setor	Processo	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo	Periodicidade	
Linha Líquida	Equalização Efluente Biológico Bomba 1	Horas	20	77.058	23	63.25	61.424	7	222	Mensal	
	Equalização Efluente Biológico Bomba 2	Horas	20	36.956	5	35.5	27.634	1	92	Mensal	
	Tanque Equalização B Agitador 1	Horas	20	188.486	10	135.5	169.653	10	631	Mensal	
	Tanque Equalização B Agitador 2	Horas	20	190.092	353	135	169.197	10	636	Mensal	
	Tanque Equalização C Agitador 1	Horas	20	196.917	10	135.5	171.668	10	632	Mensal	
	Tanque Equalização C Agitador 2	Horas	20	196.719	10	124.5	173.156	10	631	Mensal	
	Obra de Entrada 15VI01	Horas	32	702.967	742	719.5	87.212	235	743	Mensal	
	Obra de Entrada 15VI02	Horas	32	705.658	671	724.093	88.149	234	744	Mensal	
	Obra de Entrada Agitador Bacial Fossas 1	Horas	30	150.079	92.5	91.75	121.034	30	546	Mensal	
	Obra de Entrada Agitador Bacial Fossas 2	Horas	30	124.545	170	107.6	55.085	30	258	Mensal	
	Obra de Entrada Barras Automáticas 1	Horas	51	9.275	5	5	28.569	4	209	Diária	
	Obra de Entrada Barras Automáticas 2	Horas	44	5.545	5	5	1.810	4	13	Diária	
	Obra de Entrada Barras Automáticas 3	Horas	44	5.447	5	5	1.933	3	14	Diária	
	Obra de Entrada Bomba Bacia Fossica 1	Horas	32	4.945	1	3.898	4.718	1	19	Mensal	
	Obra de Entrada Bomba Bacica Fossica 2	Horas	32	2.368	1	2	1.503	1	9	Mensal	
	Obra de Entrada Bomba Bacica Fossica 3	Horas	32	3.065	2	2.28	2.394	1	11	Mensal	
	Obra de Entrada Grelha Mecânica	Horas	31	3.742	1	1	3.827	1	16	Mensal	
	Obra de Entrada Tamisador 1	Horas	32	267.594	281	269.333	177.753	22	634	Mensal	
	Obra de Entrada Tamisador 2	Horas	32	459.677	126	467.5	106.608	126	647	Mensal	
	Obra de Entrada Tamisador 3	Horas	32	366.219	276	387.5	117.836	98	628	Mensal	
	Sistema de Apoio Geral Caldeira	Horas	1577	9.217	24	7	7.679	1	43	Diária	
	Sistema de Apoio Geral Gerador	Horas	1571	1.350	1	1	0.640	0.12	10	Diária	
	Tratamento Primário 04-CR-01	Horas	30	479.785	12	577	234.705	12	743	Mensal	
	Tratamento Primário 04-CR-02	Horas	32	335.809	5	288.333	209.493	5	743	Mensal	
	Tratamento Primário 04-CR-03	Horas	32	431.283	718	426.265	194.254	33	740	Mensal	
	Tratamento Primário Agitador Poço Gorduras	Horas	30	37.881	11	12	69.349	1	370	Mensal	
	Tratamento Primário Bomba 1 Gorduras	Horas	31	2.817	2	2.667	1.580	1	8	Mensal	
	Tratamento Primário Bomba 1 Lamas Primárias	Horas	32	104.927	65	111.5	38.541	19	172	Mensal	
	Tratamento Primário Bomba 2 Gorduras	Horas	19	5.111	2	2	8.286	1	37	Mensal	
	Tratamento Primário Bomba 2 Lamas Primárias	Horas	32	96.833	4	109.5	43.710	4	160	Mensal	
	Tratamento Primário Bomba Areias Desarenador 1	Horas	32	26.750	21	24	11.170	5	44	Mensal	
	Tratamento Primário Bomba Areias Desarenador 2	Horas	30	23.070	22	22.056	9.879	3	44	Mensal	
	Tratamento Primário Ponte Raspadora Desarenador 1	Horas	32	456.094	516	503.5	120.515	97	614	Mensal	
	Tratamento Primário Ponte Raspadora Desarenador 2	Horas	30	407.341	402	473.5	155.098	8	618	Mensal	
	Tratamento Primário Ponte Raspadora Sedipac 1	Horas	32	707.853	671	724.5	87.804	238	744	Mensal	
	Tratamento Primário Ponte Raspadora Sedipac 2	Horas	32	626.814	42	713.293	196.376	42	743	Mensal	
	Linha Sólida	Desidratação Bomba Elevação Lamas Desidratadas	Horas	32	361.417	292	361	90.101	92	511	Mensal
		Desidratação Bomba Poli Centrifuga 1	Horas	19	115.895	1	64	124.094	1	322	Mensal
		Desidratação Bomba Poli Centrifuga 2	Horas	32	211.083	219	217.5	81.423	20	347	Mensal
		Desidratação Bomba Poli Reserva Centrifuga	Horas	32	152.542	2	122.5	107.773	2	352	Mensal
		Desidratação Centrifuga 1	Horas	12	202.944	52	204.167	101.483	52	337	Mensal
		Desidratação Centrifuga 2	Horas	32	370.469	98	369.5	94.313	98	531	Mensal
		Desidratação Parafuso Lamas Saída Centrifuga	Horas	32	362.406	294	361.5	90.056	92	512	Mensal
		Desodorização 15-VE-01	Horas	32	307.339	2	245.735	218.346	2	742	Mensal
		Desodorização Bomba Ácido Sulfúrico	Horas	32	1.276	1	1.25	0.161	1	2	Mensal
Desodorização Bomba Hipoclorito de Sódio		Horas	32	13.042	22	12	4.227	2	28	Mensal	
Desodorização Bomba Recirculação Torre Ácida		Horas	32	306.761	2	245.438	218.058	2	742	Mensal	
Desodorização Bomba Recirculação Torre Básica		Horas	32	297.824	2	244.938	214.525	2	742	Mensal	
Desodorização Bomba Soda		Horas	32	37.990	68	34.5	13.209	1	68	Mensal	
Desodorização Ventilador 1		Horas	32	142.405	249	113.5	110.847	2	364	Mensal	
Desodorização Ventilador 2		Horas	32	162.182	1	135.843	91.740	1	363	Mensal	
Espessamento Agitador Tanque Lamas Mistas		Horas	32	642.831	718	717.611	159.346	167	744	Mensal	
Espessamento Bomba 1 Lamas Espessadas		Horas	427	6.458	4	4	159.346	0.5	271	Diária	
Espessamento Bomba 1 Lamas de Saída das Mesas		Horas	32	130.562	87	125	55.228	42	238	Mensal	
Espessamento Bomba 1 Saída TLM		Horas	370	5.794	1	6	3.260	0.5	15	Diária	
Espessamento Bomba 2 Lamas Espessadas		Horas	427	3.913	3	3.5	2.199	0.5	12	Diária	
Espessamento Bomba 2 Lamas Saída das Mesas		Horas	32	119.260	46	124	43.424	44	271	Mensal	
Espessamento Bomba 2 Saída TLM		Horas	427	5.973	1	6	3.237	1	17	Diária	
Espessamento Bomba Poli Mesa 1		Horas	20	94.625	11.5	104.216	73.607	2	238	Mensal	
Espessamento Bomba Poli Mesa 2		Horas	32	167.453	39	142.778	74.033	39	304	Mensal	
Espessamento Bomba Poli Reserva		Horas	21	149.212	205	158.111	77.984	12	327	Mensal	
Espessamento Bomba de Água de Lavagem da Mesa 1		Horas	31	104.552	1	42	113.598	1	340	Mensal	
Espessamento Bomba de Água de Lavagem da Mesa 2		Horas	32	305.878	380	307	106.889	91	475	Mensal	
Espessamento Compressor de Ar de Serviço 1		Horas	17	6.900	3	3	7.422	1	30	Mensal	
Espessamento Compressor de Ar de Serviço 2		Horas	15	5.333	6	5	7.422	1	13	Mensal	
Espessamento Mesa 1		Horas	31	104.409	1	50	112.068	1	342	Mensal	
Espessamento Mesa 2		Horas	32	293.924	197	291.5	106.822	92	476	Mensal	
Espessamento Ponte Raspadora Espessador Gravitico		Horas	32	698.456	670	719.21	91.433	218	743	Mensal	
Espessamento Preparador Polímero 1		Horas	32	26.931	28	28.667	11.420	6	62	Mensal	
Espessamento Preparador Polímero 2		Horas	32	20.214	1	20.726	6.958	1	36	Mensal	
Estabilização Biológica Agitador Cuba Dig 1		Horas	32	522.440	725	549.667	195.518	102	740	Mensal	
Estabilização Biológica Agitador Cuba Dig 2		Horas	32	610.698	736	688.5	153.638	187	741	Mensal	
Estabilização Biológica Bomba Lamas Digeridas Dig 1		Horas	32	80.932	3	16.108	103.229	3	322	Mensal	
Estabilização Biológica Bomba Lamas Digeridas Dig 2		Horas	32	181.334	288	188.506	112.773	4	351	Mensal	
Estabilização Biológica Bomba Poço Condensados Dig 1		Horas	31	1.471	1	1	1.077	1	6	Mensal	
Estabilização Biológica Bomba Poço Condensados Dig 2		Horas	31	3.649	1	30	3.023	1	11	Mensal	
Estabilização Biológica Bomba Recirculação Água Quente Dig 1		Horas	32	418.038	469	471.222	115.662	132	499	Mensal	
Estabilização Biológica Bomba Recirculação Água Quente Dig 2		Horas	32	413.316	148	478.722	117.494	148	516	Mensal	
Estabilização Biológica Bomba Recirculação Água Quente Reserva		Horas	29	434.393	460	486	111.586	123	590	Mensal	
Estabilização Biológica Bomba Recirculação Lamas Dig 1		Horas	32	420.537	560	498.667	204.230	10	741	Mensal	
Estabilização Biológica Bomba Recirculação Lamas Dig 2		Horas	32	529.214	404	550.167	123.305	148	722	Mensal	
Estabilização Biológica Bomba Recirculação Lamas Reserva		Horas	32	355.274	221	355.167	101.557	70	580	Mensal	
Estabilização Biológica Bomba Recirculação Caldeira		Horas	28	340.409	1	358.5	227.350	1	709	Mensal	
Estabilização Biológica Caldeira		Horas	28	6.368	1	6	4.790	1	17	Mensal	
Estabilização Biológica Compressor Biogás Digestor 1		Horas	32	701.470	725	725.06	90.919	218	741	Mensal	
Estabilização Biológica Compressor Biogás Digestor 2		Horas	32	704.112	744	728	92.968	218	744	Mensal	
Estabilização Biológica Ventilador Gasometro		Horas	32	713.407	744	734.167	91.384	218	744	Mensal	
Produção de Energia Máquina de Cogeração 1135		Horas	419	17.483	10	9	22.024	2	205	Diária	
Produção de Energia Máquina de Cogeração 1136		Horas	422	11.265	11	10.327	5.768	1	25	Diária	

maiores consumos desse ano, que depois descem consideravelmente em Abril e mantém-se estáveis até ao final de 2019.

Além disso, foi também calculada a variação média durante os meses, de todos os anos, das horas de funcionamento dos diferentes processos realizados na instalação. Como dentro de cada setor se podem identificar diversos processos, e para não tornar este tópico muito massivo, fez-se foco nos processos como uma visão geral. Na Figura 4.20 ilustra então a variação média dos valores do funcionamento dos processos realizados na infraestrutura ao longo dos meses, de todos os anos.

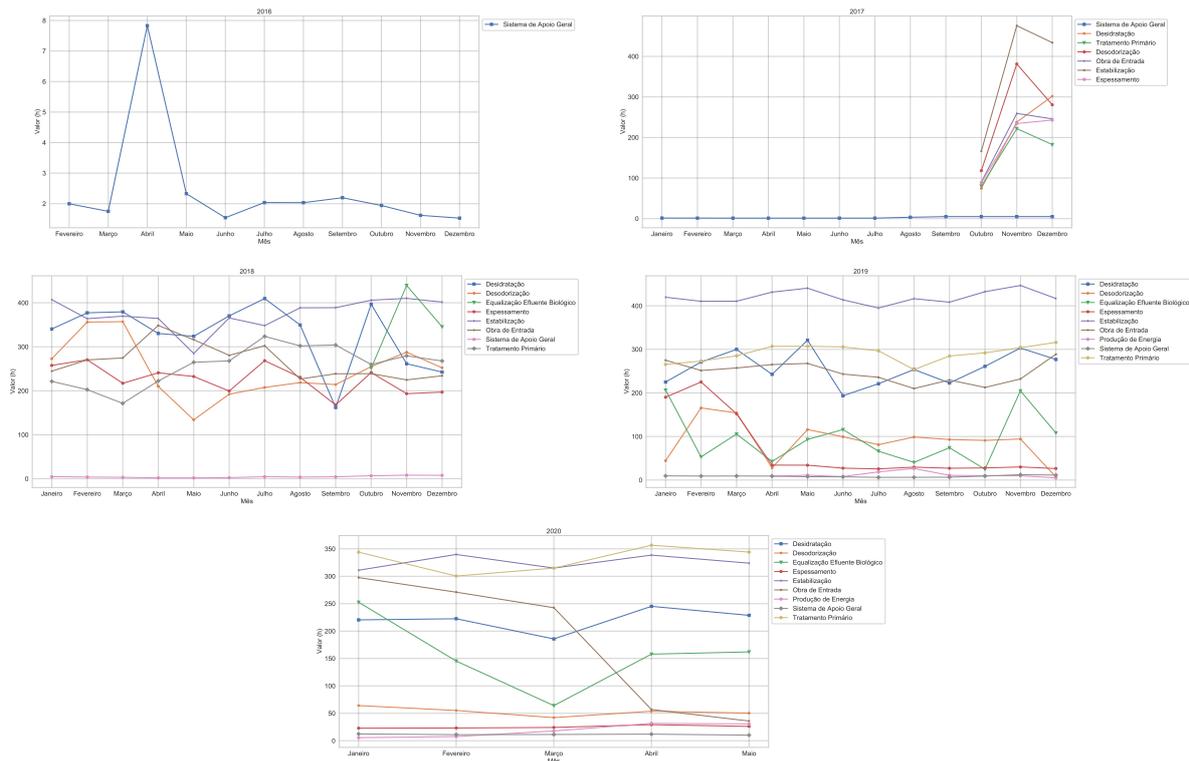


Figura 4.20: Variação Média geral dos processos do funcionamento por mês em cada ano.

Como se pode observar na Figura 4.20, relativamente ao ano de 2016 apenas se tem a acesso a dados relativos ao processo *Sistema de Apoio Geral*. É visível então que, no mês de Abril ocorre um pico de horas utilizadas comparativamente com os restantes meses do ano, onde este processo se mantém bastante homogéneo nos valores apresentados, tendo uma tendência decrescente a partir do mês de Setembro. Já em relação ao ano de 2017, fica claro a escassez de dados existentes, onde para todos os processos, com exceção do *Sistema de Apoio Geral*, apenas se tem dados entre os meses de Outubro e Dezembro. No que concerne ao processo *Sistema de Apoio Geral*, durante todo o ano, apresenta valores sempre muito próximos de zero, sendo um pouco maiores a partir do mês de Julho. Relativamente aos outros processos, todos tem um crescimento acentuado do mês de Outubro para Novembro, onde na maioria deles volta a decrescer no mês de Dezembro, sendo o processo *Estabilização* que apresenta os valores mais elevados.

No que tange ao ano de 2018, o processo *Sistema de Apoio Geral* volta a apresentar os seus valores próximos de zero, como no ano anterior. Os restantes processos concentram os seus valores

médios de funcionamentos entre 200 a 400 horas por mês, onde o processo *Estabilização* é quem assume os valores médios deste ano mais elevados. De realçar um decréscimo assinalável no processo *Desidratação* no mês de Setembro e em *Desodorização* no mês de Maio. Em relação ao ano de 2019, é perfeitamente claro que o processo *Estabilização* assume os valores mais elevados em todos os meses deste ano. Depois, pode resultar numa perceção em que os processos se dividem em dois grupos distintos ao nível dos valores apresentados, estando num deles os processos *Tratamento Primário*, *Desidratação* e *Obra de Entrada* e no outro os restantes processos. Nesse último grupo, que exhibe os valores mais baixos, vê-se que no início do ano quase todos os processos assumem valores mais elevados quando comparado com os restantes meses do ano, onde os valores estabilizam, salvo em *Equalização Efluente Biológico* que volta a subir em Novembro. Por fim, no ano de 2020 os processos *Tratamento Primário* e *Estabilização* são os que assumem os valores mais elevados durante os meses do ano. De ressaltar o comportamento do processo *Obra de Entrada* que em Janeiro apresenta dos valores mais elevados, valor esse que decresce a cada mês que passa, tendo inclusive um decréscimo bastante pronunciado em Abril. Já *Equalização Efluente Biológico* de Janeiro a Março os seus valores tem clara tendência decrescente, onde em Abril voltam a subir consideravelmente mantendo-se nesses valores até maio.

De modo a verificar o número de horas de funcionamento anuais em cada ano, foi elaborado o cálculo, por ano, do valor médio de horas de funcionamento, onde a Figura 4.21 representa o resultado obtido.

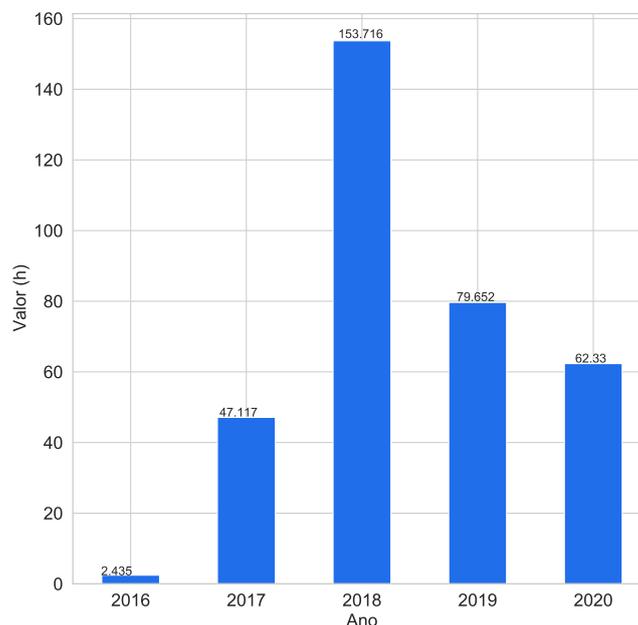


Figura 4.21: Valores médios do Funcionamento em cada ano.

Analisada a figura, verifica-se que o ano de 2018 se destaca largamente a nível de horas de funcionamento utilizadas no mesmo, quase igualando o dobro do valor do ano que obteve a segunda maior média de horas utilizadas, o ano de 2019. O ano de 2016 obteve a média anual mais baixa, com elevada diferença para os restantes anos, onde mesmo pertencendo apenas a um processo, os seus valores são realmente baixos.

Olhando individualmente para os processos, que pode ser examinado na Figura 4.22, observa-se que 2018, em muitos dos processos, é o ano que apresenta os valores mais elevados, sendo que nos casos em que não é o maior, os seus valores estão próximos do mais elevado, o que vem comprovar o descrito anteriormente em 4.21. Verifica-se também que os valores apresentados pelo processo *Estabilização*, exibe os valores mais elevados de todos os anos, com exceção do ano de 2020 onde é ligeiramente ultrapassado pelo processo *Tratamento Primário*. Percebe-se também que os processos de *Desidratação*, *Desodorização*, *Espessamento* e *Obra de Entrada*, a partir do ano de 2018, os seus valores tem uma tendência decrescente com o passar dos anos. No sentido inverso, o *Sistema de Apoio Geral* e *Tratamento Primário*, com o passar dos anos, os seus valores tendem a subir.

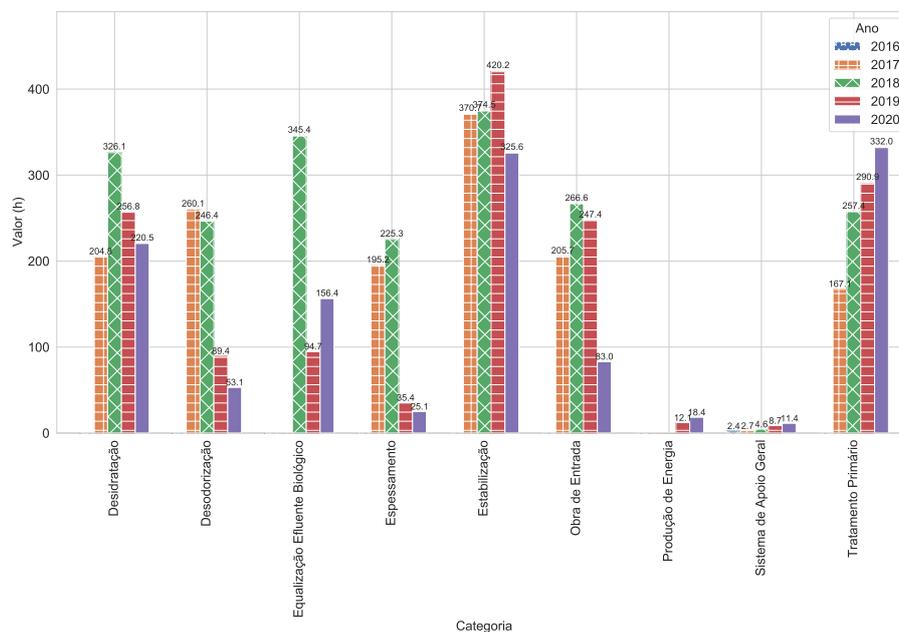


Figura 4.22: Variação Média dos processos do Funcionamento em cada ano.

De modo a investigar como se comportam os dados ao longo dos trimestres, foi realizado o cálculo dos valores médios em cada um deles. A Figura 4.23 ilustra então os resultados obtidos.

Como se pode observar, é no 1º trimestre que, nos anos de 2019 e 2020, se verifica mais horas de funcionamento, destacando-se mais a diferença para os outros trimestres no ano de 2019. Já relativamente aos anos de 2017 e 2018 o 4º trimestre é o que assume os valores mais elevados, onde a diferença é mais marcada no ano de 2017, justificada pelo facto destes anos conterem a esmagadora maioria dos dados nos últimos três meses do ano, como se verificou em 4.20. No ano de 2016, como já se tinha visto previamente, os valores são muito baixos, portanto por trimestres seguem a mesma linha. De destacar também que em cada ano tirando os trimestres com maior valor, os restantes apresentam valores bastante próximos uns dos outros.

Já na Figura 4.24 é possível analisar os valores dos processos por trimestres, individualmente.

Analisada a figura, em relação ao ano de 2016 apesar dos trimestres conterem valores baixos e próximos entre si, é no 2º trimestre que os valores são mais elevados, tendo os restantes praticamente o mesmo valor. No que toca ao ano de 2017, é o tratamento *Estabilização* que exibe os valores mais

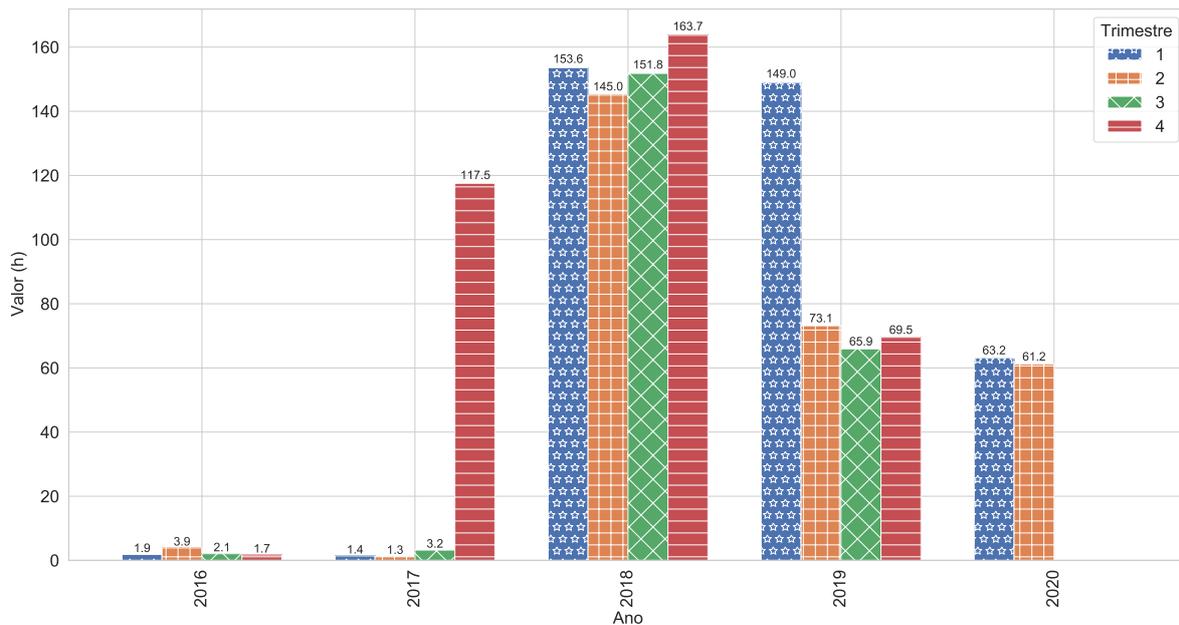


Figura 4.23: Valores médios do Funcionamento em cada trimestre.

elevados, seguindo-se da *Desodorização*. Os restantes processos apresentam os seus valores próximos uns dos outros, com exceção em *Sistema de Apoio Geral* onde os seus valores são muito inferiores a todos. A respeito do ano de 2018, são os processos *Estabilização* e *Desidratação* que assumem os valores mais elevados na sua generalidade. Verifica-se também que *Desidratação* e *Espessamento* ao longo dos trimestres manifestam uma tendência decrescente nos seus valores. Já com um comportamento quase inverso, encontram-se os processos *Desodorização* e *Estabilização*, que tirando o 1º trimestre, assume uma propensão crescente com o passar dos trimestres. Em relação ao ano de 2019 *Estabilização* destaca-se claramente como o processo que contém os valores mais elevados, apresentando uma distância considerável para os restantes. Já *Desidratação*, *Obra de Entrada* e *Tratamento Primário* apresentam os seus valores muito idênticos entre si. De salientar que, em todos os processos, os valores entre os trimestres não sofrem grandes variações, com a exceção de *Espessamento* onde o 1º trimestre se realça com valores mais elevados quando comparado com os restantes trimestres desse processo. Por último, em 2020 *Estabilização* e *Tratamento Primário* são os processos que exibem os valores mais elevados, apresentando inclusive valores muito próximos nos respetivos trimestres. Verifica-se também que o 2º trimestre assume valores mais altos, em quase todos os processos, salvo poucas exceções como é o caso de *Obra de Entrada* onde o 1º trimestre apresenta claramente os valores mais elevados.

Em relação à disposição dos dados pelas diferentes estações do ano, foi calculado os valores médios em cada uma delas, onde a Figura 4.25 expõem essa mesma informação.

Como se pode observar na figura, o ano de 2018 é quem assume os valores mais elevados, estando todas as estações do ano nesse ano com valores próximos entre si. No ano de 2019 destaca-se o Inverno que tem os valores bem mais altos comparativamente com as restantes estações do ano. Já em 2017 observa-se uma notória diferença entre as estações do ano Inverno e Outono com a Primavera e Verão,

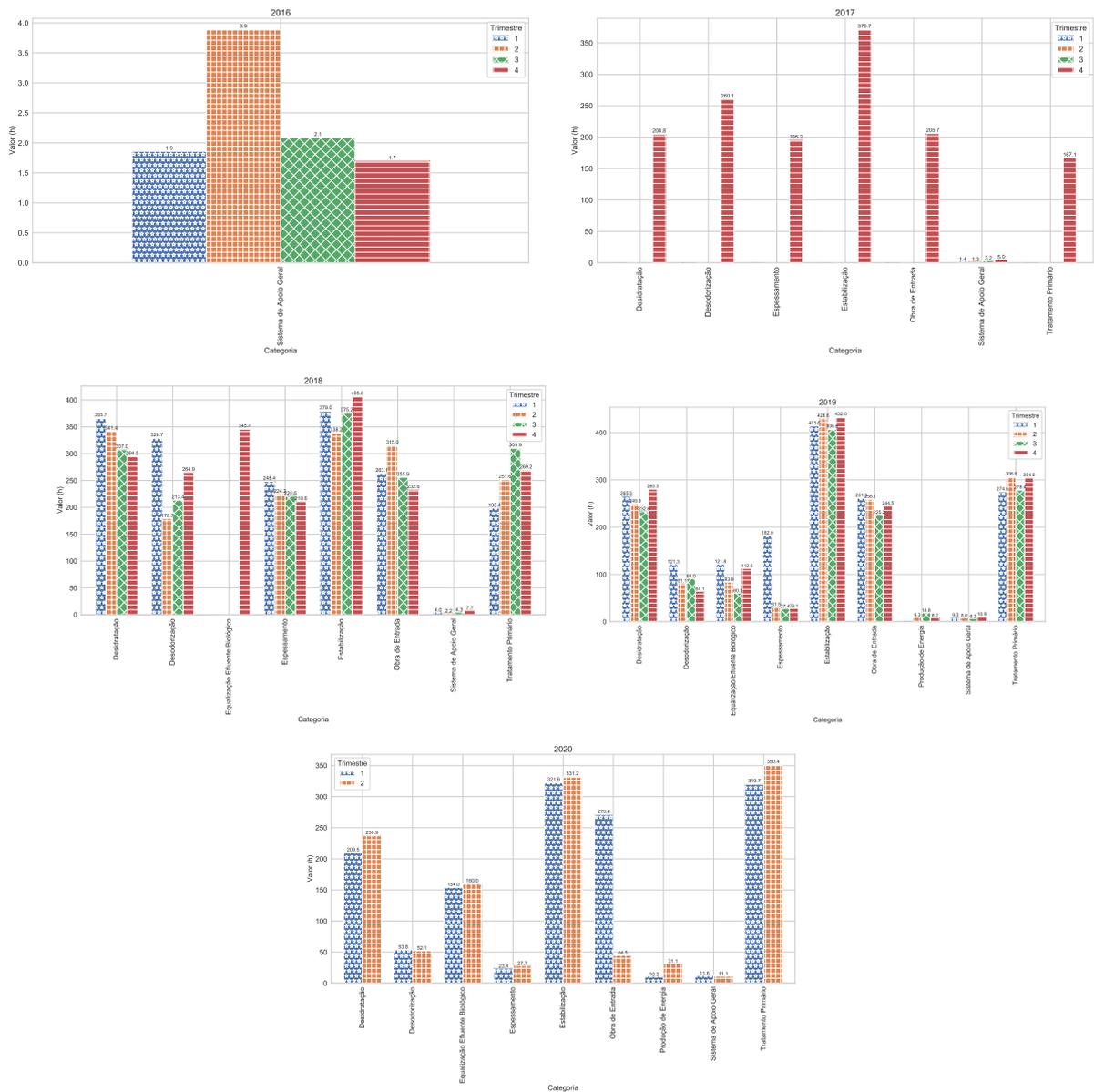


Figura 4.24: Variação Média dos processos do funcionamento por trimestres.

onde ocorre uma discrepância muito acentuada nos valores.

Observando a disposições dos valores médios apresentados pelos processos nas estações do ano, que pode ser consultado na Figura 4.26, verifica-se que em 2016 os valores foram realmente baixos, com destaque para a Primavera que apresentou os valores mais elevados, estando as restantes estações do ano com valores muito equivalentes. Em 2017, em todos os processos, o Inverno foi a estação do ano com os maiores valores, destacando-se sobretudo em *Estabilização* onde os valores são consideravelmente maiores. A única exceção é relativa a *Sistema de Apoio Geral* onde os valores são baixíssimos, e apresentam uma pequena tendência crescente com o passar das estações do ano. Em relação a 2018 *Estabilização*, *Desidratação* e *Equalização Efluente Biológico* possuem os valores mais elevados, do modo geral, nas estações do ano correspondentes. Já *Desodorização*, *Espessamento*, *Obra de Entrada* e *Tratamento Primário* acabam por, na sua generalidade, apresentar valores idênticos,

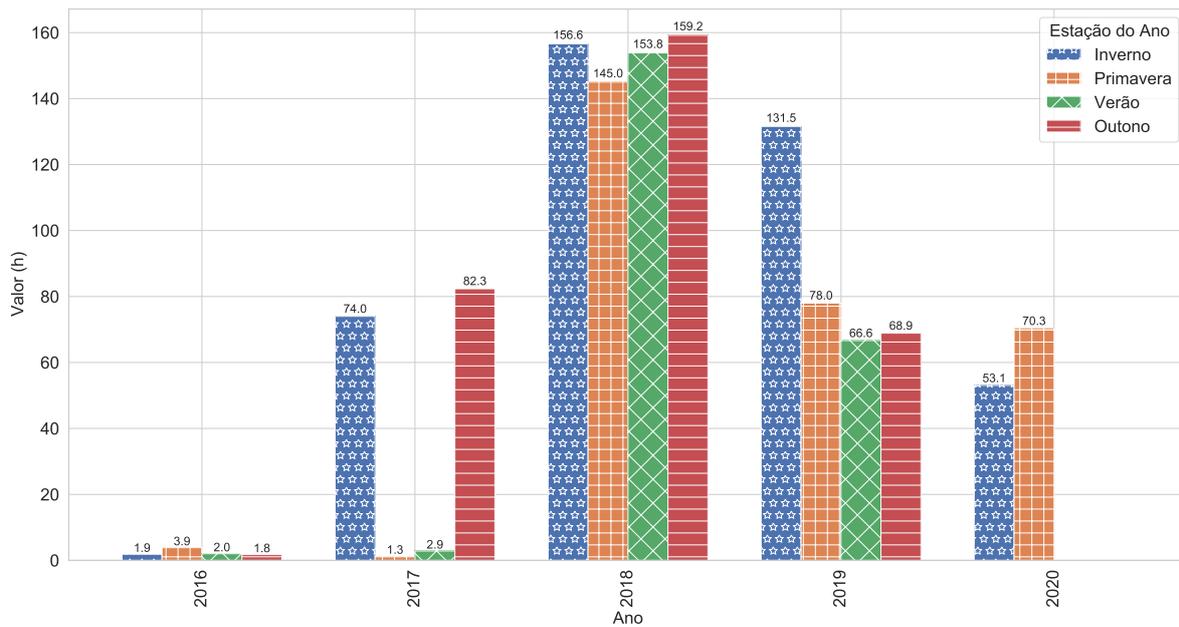


Figura 4.25: Valores médios do Funcionamento em cada estação do ano.

mas apresentam por exemplo os seus maiores valores em estações de ano distintas. No que toca ao ano de 2019, *Estabilização* é com toda a clareza o processo que utilizou mais horas de funcionamento em todas as estações do ano. Já *Desodorização*, *Obra de Entrada* e *Tratamento Primário* assumem valores próximos entre si, em todas as estações do ano. De realçar a existência de um certo equilíbrio de valores, em todos os processos, entre as estações do ano. Por fim, em 2020 o *Tratamento Primário* e *Estabilização* abarcam os valores mais elevados relativos a este ano, nas respetivas estações do ano. De evidenciar a existência de uma enorme diferença entre o Inverno e a Primavera em *Obra de Entrada*, onde o valor do Inverno está próximo dos máximos ocorridos neste ano.

4.2.1.4 Relação entre os diferentes tipos de dados

De forma a identificar possíveis relações entre os diferentes tipos de dados e perceber-se até que ponto se influenciam entre si, foram calculadas as correlações entre eles, utilizando mais uma vez o método Coeficiente de Correlação de Spearman, de modo a extrair essas informações. Começando pelos dados da electricidade e controlo analítico, onde a Tabela 4.5 ilustra as maiores correlações encontradas entre estes dois tipos de dados, conseguiu-se verificar que pouco se relacionam, existindo apenas cinco correlações significativas, sendo todas elas moderadas. Verifica-se então que a maior relação entre estes dois dados é entre o indicador *Ponta* da electricidade com o indicador *Sólidos Totais* no setor *Lamas Digeridas 1* no controlo analítico, com um valor de -0.576 , significando que se relacionam negativamente.

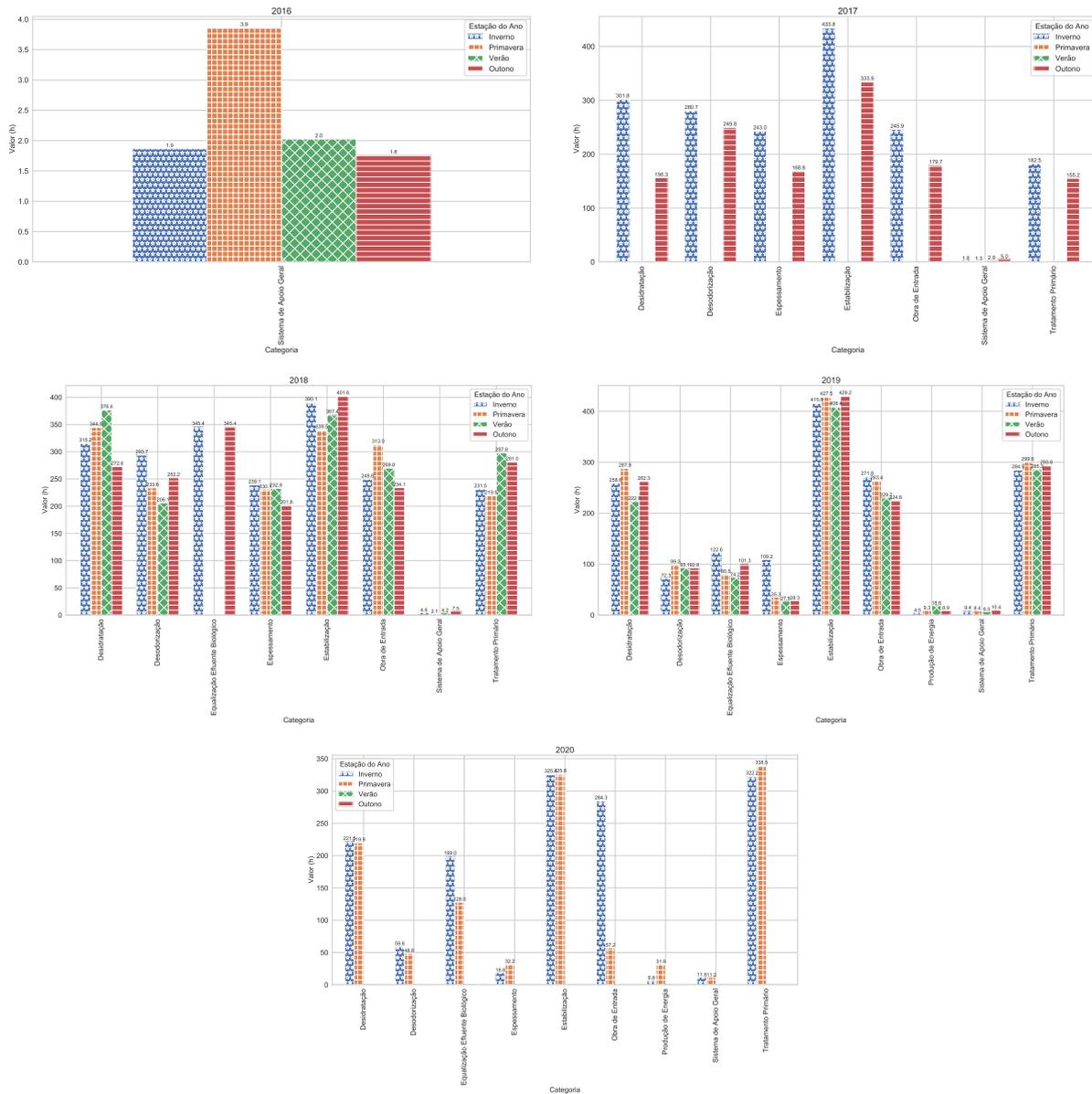


Figura 4.26: Variação Média dos processos do funcionamento por estações do ano.

Tabela 4.5: Maiores correlações encontradas entre Electricidade e Controlo Analítico.

Electricidade	Controlo Analítico	Valor Correlação
Ponta	Sólidos Totais em Lamas Digeridas 1	-0.576
Ponta	Sólidos Voláteis Lamas Digeridas 1	-0.561
Ponta	Sólidos Suspensos Voláteis em Lamas Biológicas 1	0.54
Ponta	Sólidos Suspensos Totais em Lamas Biológicas 1	0.533

Já entre os dados do Electricidade e Funcionamento, que pode ser consultado na Tabela 4.6, já se identificam duas correlações fortes entre *Ponta* com *Tratamento Primário Bomba 1 Lamas Primárias* na *Linha Líquida* e entre o indicador *Ponta* com *Tratamento Primário Bomba 2 Lamas Primárias* na *Linha Líquida*, ambas com influência negativa. As restantes relações apresentadas tratam-se de correlações moderadas, o que indica que existe uma relação considerável entre os casos em questão.

Contudo, foram ainda detetadas um conjunto de correlações fracas entre os dados da Electricidade e Funcionamento.

Tabela 4.6: Maiores correlações encontradas entre Electricidade e Funcionamento.

Electricidade	Funcionamento	Valor Correlação
Ponta	Tratamento Primário Bomba 1 Lamas Primárias em Linha Líquida	-0.773
Ponta	Tratamento Primário Bomba 2 Lamas Primárias em Linha Líquida	-0.763
Ponta	Obra de Entrada Grelha Mecânica em Linha Líquida	-0.582
Ponta	Obra de Entrada Tamisador 1 em Linha Líquida	0.562
Vazio	Espessamento Preparador Polímero 2 em Linha Sólida	0.555
Ponta	Espessamento Compressor de Ar de Serviço 1 em Linha Sólida	0.55
Ponta	Estabilização Biológica Bomba Poço Condensados Dig 2 em Linha Sólida	-0.547
Ponta	Desodorização Bomba Recirculação Torre Básica em Linha Sólida	0.537
Vazio	Espessamento Compressor de Ar de Serviço 2 em Linha Sólida	-0.525

Em relação aos dados do Controlo Analítico e Funcionamento, foram encontradas 210 correlações muito fortes, 236 fortes e 273 moderadas. Olhando para um total de 1424 relações possíveis, e visto que se identificou 446 relações com significado forte e muito forte e contabilizando as 273 com relevância moderada, dá um total de 719 relações relevantes, demonstrando que aproximadamente 51% de todas as relações possíveis tem uma importância relevante. Com isto, conclui-se que os dados do Controlo Analítico e Funcionamento estão evidentemente relacionados entre si, uma vez que apresentam uma clara influência um no outro. Na Tabela 4.7 podem ser consultados algumas das correlações muito fortes identificadas.

Tabela 4.7: Algumas das maiores correlações encontradas entre Controlo Analítico e Funcionamento.

Controlo Analítico	Funcionamento	Valor Correlação
Ph em Lamas Digeridas 1	Espessamento Bomba Poli Mesa 1 em Linha Sólida	-1
SV em Lamas Digeridas 1	Obra de Entrada Bomba Bacil Fossica 3 em Linha Líquida	-1
SST em Efluente Tratado	Desodorização Bomba Hipoclorito de Sódio em Linha Sólida	-1
Ortofosfatos em Efluente Tratado	Tratamento Primário Ponte Raspadora Sedipac em Linha Líquida	-1
SV em Lamas Digeridas 1	Espessamento Preparador Polímero 1 em Linha Sólida	-1
SST em Entrada Reator Biológico	Obra de Entrada Bomba Basica Fossica 2 em Linha Líquida	1
Fosforo Total em Afluente Bruto	Estabilização Biológica Compressor Biogás Digestor 1 em Linha Sólida	1
ST em Lamas Biológicas Recirculadas	Desodorização Bomba Ácido Sulfúrico	1
CQO em Efluente Tratado	Desodorização Bomba Ácido Sulfúrico	1
IVL em Lamas Biológicas 2	Estabilização Biológica Caldeira	1

Por fim, na Tabela 4.8 pode ser consultado a relação entre os dados do tempo com os restantes tipos de dados. Começando pela Electricidade com o Tempo, vê-se nitidamente que não existe relações entre os dois tipos de dados, uma vez que as maiores correlações encontradas foram apenas duas correlações fracas. Já entre o Tempo e Controlo Analítico identificam-se 3 relações moderadas, todas elas entre o indicador *Temperatura* nos dados do tempo, divergindo apenas os indicadores nos dados do Controlo Analítico. Contudo, é um número bastante reduzido de relações, o que demonstra uma baixa influência entre estes dois tipos de dados.

Tabela 4.8: Maiores correlações encontradas entre o Tempo e os restantes tipo de dados.

Tempo	Dados	Valor Correlação
Electricidade		
Temperatura	Ponta	-0.433
Pressão atmosférica	Ponta	0.317
Controlo Analítico		
Temperatura	SV em Lamas Digeridas 1	0.543
Temperatura	ST em Lamas Digeridas 1	0.509
Temperatura	CQO em Efluente Tratado	0.5

4.2.2 Guimarães

Já para a ETAR de Guimarães, será também realizada uma análise exploratória a cada tipo de dados desta infraestrutura, especificamente aos dados relacionados com a electricidade e caudais. Tal como em 4.2.1, esta análise não se irá focar apenas nos dois tipos de dados pertencentes à ETAR em questão, mas também na relação entre eles, com o intuito de se encontrar possíveis relações que possam ser relevantes para o estudo.

4.2.2.1 Electricidade

Em relação aos dados da electricidade, iniciou-se por aplicar de novo estatística descritiva nos dados, dado que neste ponto se encontram já com o tratamento concluído. O resultado obtido pode ser consultado na Tabela 4.9.

Tabela 4.9: Resultado da Estatística Descritiva aplicada aos Indicadores da Electricidade de Guimarães, após tratamento.

Métricas	Cheia	Ponta	Super Vazio	Vazio
Registos	1612	1610	1612	1612
Média	3444.308	1172.549	1177.910	2155.570
Moda	348	991	823	687
Mediana	3514	1075.147	1154.929	1436
Desvio Padrão	1811.966	520.614	405.240	1614.917
Mínimo	76	3	282	203
Máximo	11064	2964	3024	8733

Como se pode explorar na tabela, os valores apresentados pelos indicadores já são mais realísticos para uma infraestrutura como uma ETAR, onde se verificam diferenças enormes, na ordem dos milhares, comparativamente com os dados exibidos antes do processo de tratamento. Verifica-se também que o indicador *cheia* é o que apresenta os valores médios mais elevados, com alguma distância para os restantes. No sentido inverso, os indicadores *ponta* e *super vazio* apresentam os menores valores médios, tendo inclusive praticamente o mesmo valor. Portanto nesta infraestrutura os maiores consumos energéticos, em média, ocorrem nos períodos correspondentes ao indicador *cheia*, seguido dos equivalentes ao indicador *vazio*.

Com o intuito de se ter uma visão mais geral dos dados, foi realizada a variação dos mesmos em cada ano de registo. A Figura 4.27 ilustra essa mesma informação. Como pode ser observado na figura, os consumos, em praticamente todos os meses, tem tendência a aumentar com o passar dos anos, exceto no ano de 2020 onde os valores chegam a estar abaixo do ano de 2017, a partir do mês de Fevereiro. Também é de ressaltar que, em todos os anos, os consumos têm uma quebra no mês de Agosto, mas que posteriormente voltam a subir.

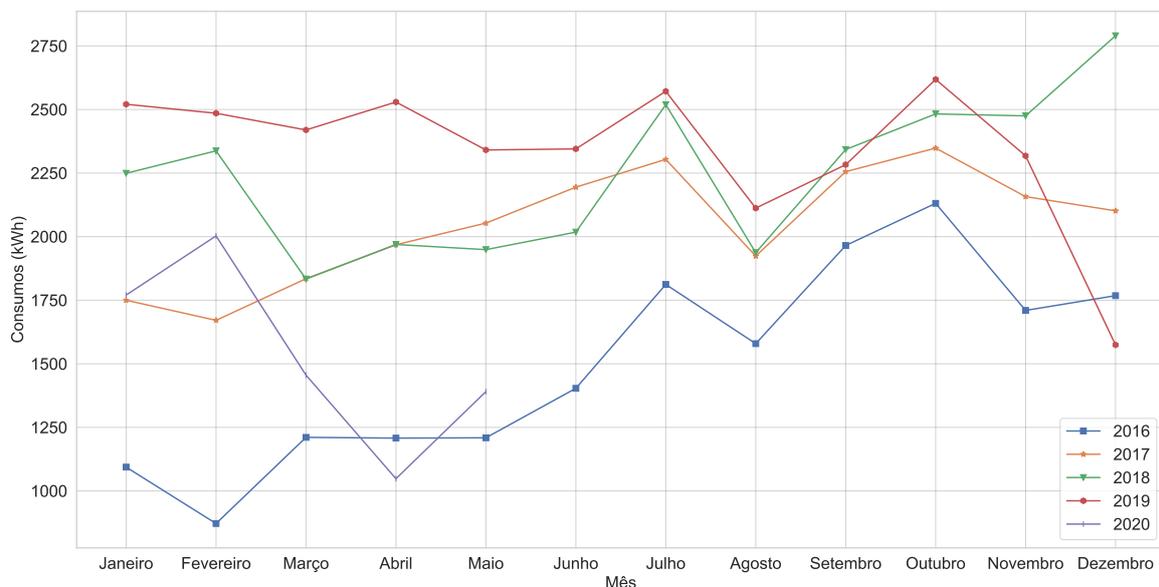


Figura 4.27: Variação Média da Electricidade de Guimarães ao longos dos meses em cada ano.

Observando essa variação individualmente, ilustrada na Figura 4.28, nota-se que no indicador *cheia* em todos os anos, exceto 2020, os consumos tem um crescimento a partir do mês de Março e que se prolonga até ao mês de Julho, existindo uma quebra em Agosto, voltando a subir em Setembro, até decrescerem novamente mais próximo do final do ano. O ano de 2020 tem um pequeno aumento no mês de Fevereiro mas já a partir de Março vai de encontro aos consumos registados no ano de 2016. Em relação ao indicador *ponta* os consumos desde Janeiro a maio têm uma tendência decrescente nos seus valores, mantendo-se equilibrados até Outubro e voltam a subir até ao fim do ano. No que toca ao indicador *super vazio* vê-se uma clara diferença dos consumos entre os anos de 2017, 2018 e 2019 com o ano de 2016 na primeira metade do ano. Também se verifica que no mês de Agosto existe uma queda dos consumos, em todos os anos. De salientar também uma queda bastante acentuada dos consumos no ano de 2020 a partir do mês de Fevereiro. Por fim, relativamente ao indicador *vazio*, com exceção do ano de 2020, em Abril regista-se um aumento significativo dos consumos energéticos em todos os restantes anos, sendo que em Agosto voltam a sofrer uma queda acentuada. De um modo geral, vê-se que em todos os indicadores com o passar dos anos, os consumos tendem a subir em quase todos os meses, onde 2020 é o ano que reverte essa inclinação, aproximando os valores de consumo de 2016, ano esse onde se registou menor consumo energético na instalação.

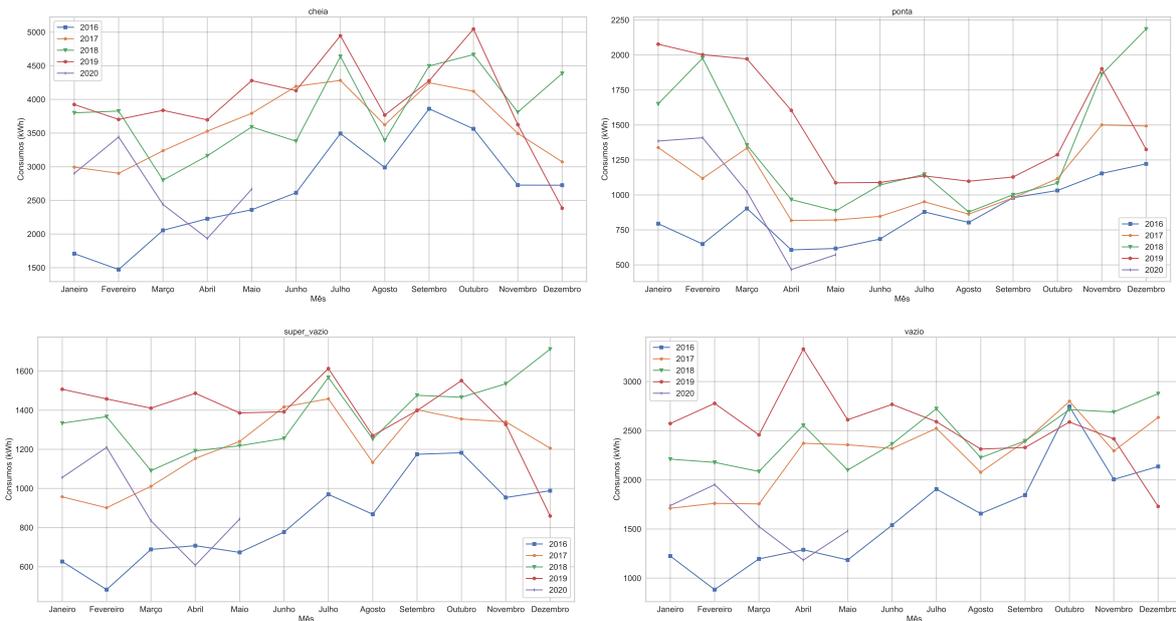


Figura 4.28: Variação Média da Electricidade de Guimarães ao longos dos meses em cada ano por indicadores.

De modo a verificar como quais os consumos médios anuais da infraestrutura, foi calculada a média de consumos em cada ano de registo, onde a Figura 4.29 elucida os resultados obtidos. Como pode ser visto, os consumos anuais tem vindo a aumentar a cada ano que passa, entre 2016 e 2019. Já em 2020 regista-se um decréscimo significativo nos consumos anuais.

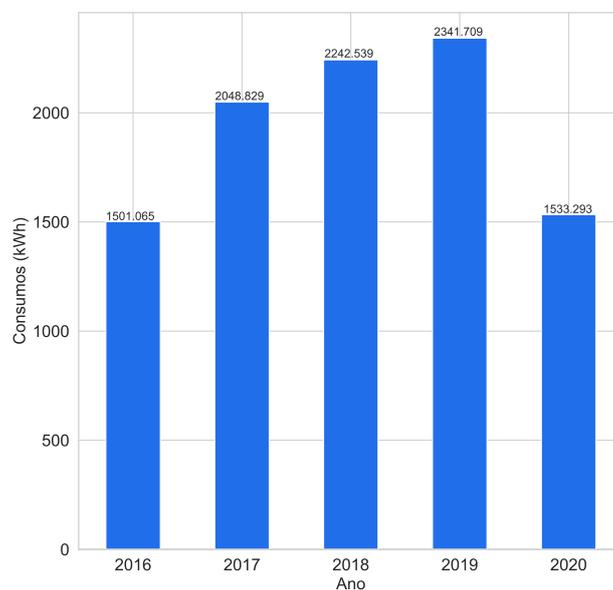


Figura 4.29: Valores médios da Electricidade de Guimarães em cada ano.

Relativamente aos consumos anuais de cada indicador, que pode ser consultado na Figura 4.30, verifica-se também que em todos eles os consumos anuais tem vindo a aumentar com o passar dos anos, mais concretamente entre 2016 e 2019, onde em 2020 observa-se uma redução considerável nos consumos.

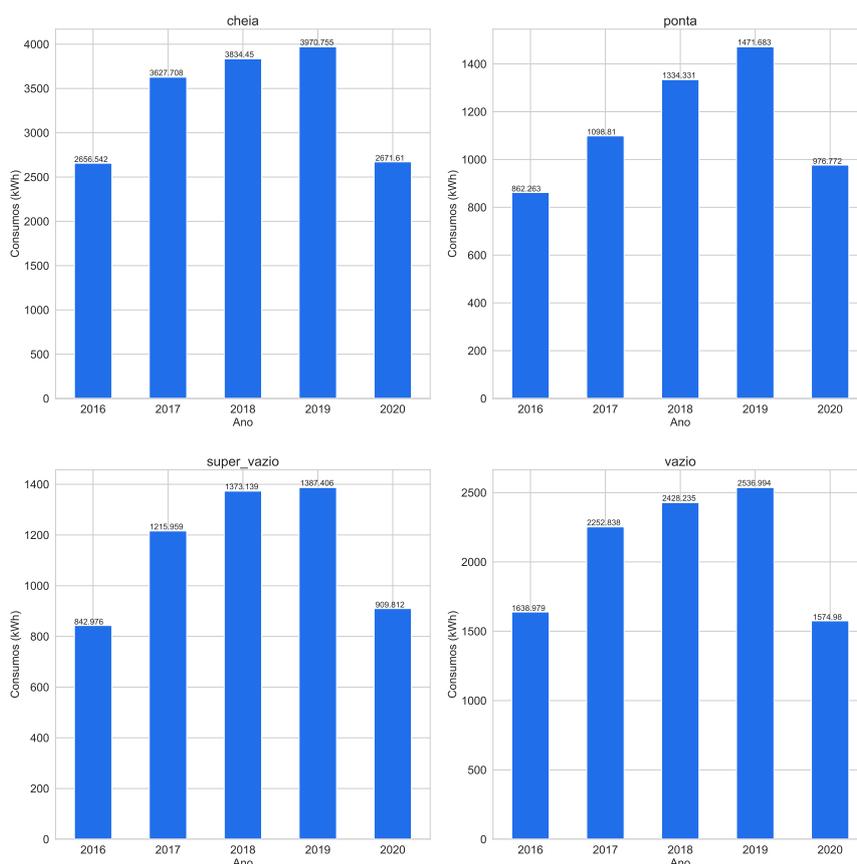


Figura 4.30: Valores médios da Electricidade de Guimarães em cada ano por indicadores.

De forma a possuir-se uma visão dos consumos por trimestres, foi feito o cálculo dos consumos energéticos por esse fator, em cada ano, onde na Figura 4.31 é possível examinar o resultado obtido. Como se pode observar, entre 2016 e 2018 nota-se uma tendência clara para o aumento dos consumos a cada trimestre que passa, exceto em 2018 onde do 1º para o 2º trimestres os valores baixam parcialmente, sendo sempre o 4º trimestre nestes anos o que apresenta os valores mais elevados. Em 2019 e 2020 verifica-se o oposto, onde o 1º trimestre assume os valores mais elevados e está-se perante uma inclinação negativa dos valores a cada trimestre que passa, ou seja, com o passar dos trimestres os consumos energéticos baixam.

Já na Figura 4.32, é possível analisar os consumos por trimestre em cada ano, por indicador. Portanto, no indicador *cheia* é o 3º trimestre que assume os valores mais elevados nos anos de 2016, 2017 e 2019. O 2º trimestre também apresenta valores relativamente altos, principalmente em 2017 e 2019. Observa-se também que os valores, em 2016, 2017 e 2019, tendem a crescer durante o 1º e 3º trimestres, onde depois se verifica uma queda considerável dos valores no 4º trimestre. Em relação ao indicador *ponta* entre 2016 e 2018 nota-se a existência de uma descida dos consumos do 1º para o 2º trimestre, onde posteriormente sobem consecutivamente até ao 4º trimestre, sendo o último com os valores mais elevados nestes anos. Em 2019 o 1º trimestre claramente assume os maiores valores, onde os mesmos descem até ao 3º trimestre, mas voltam a subir no 4º trimestre. Já em 2020, é nítido que o 1º trimestre exhibe valores bem mais elevados que o 2º trimestre. No que toca ao *super vazio* verifica-se que entre 2016 e

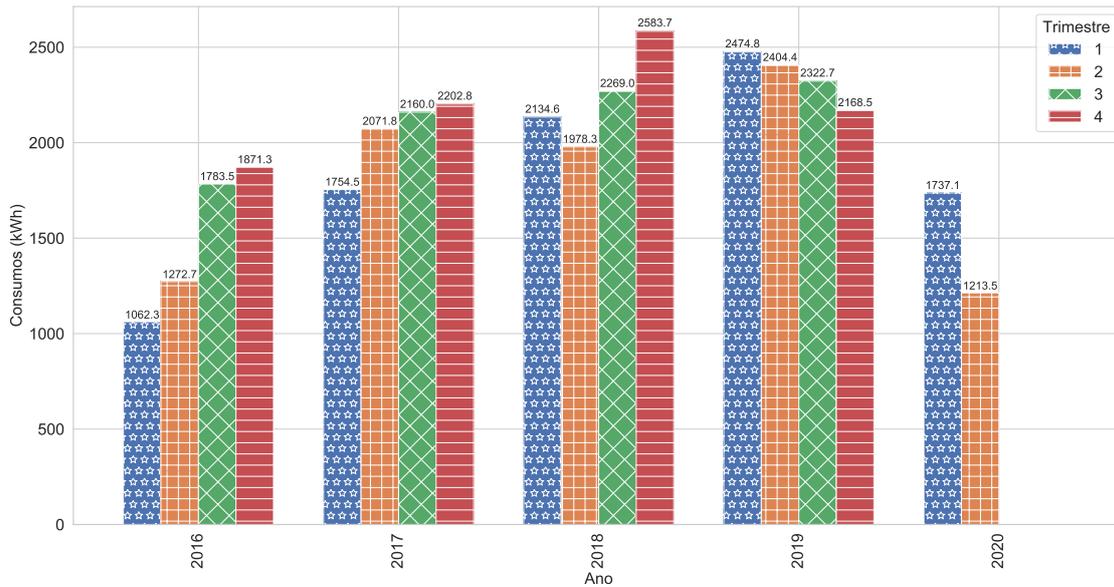


Figura 4.31: Valores médios da Electricidade de Guimarães em cada trimestre.

2018 os maiores valores foram verificados nos 3º e 4º trimestres, onde existe uma maior diferença desses dois trimestres para os restantes no ano de 2018. Já em 2019 e 2020, foi no 1º trimestre que se utilizou mais electricidade. Por fim, no indicador *vazio* é possível verificar um crescimento nos consumos com o passar dos trimestres entre 2016 e 2018, sendo o 4º trimestre aquele que exhibe os valores mais elevados. Já em 2019, foi no 2º trimestre que se registaram os maiores consumos, onde no 3º e 3º trimestres o mesmo baixa consideravelmente.

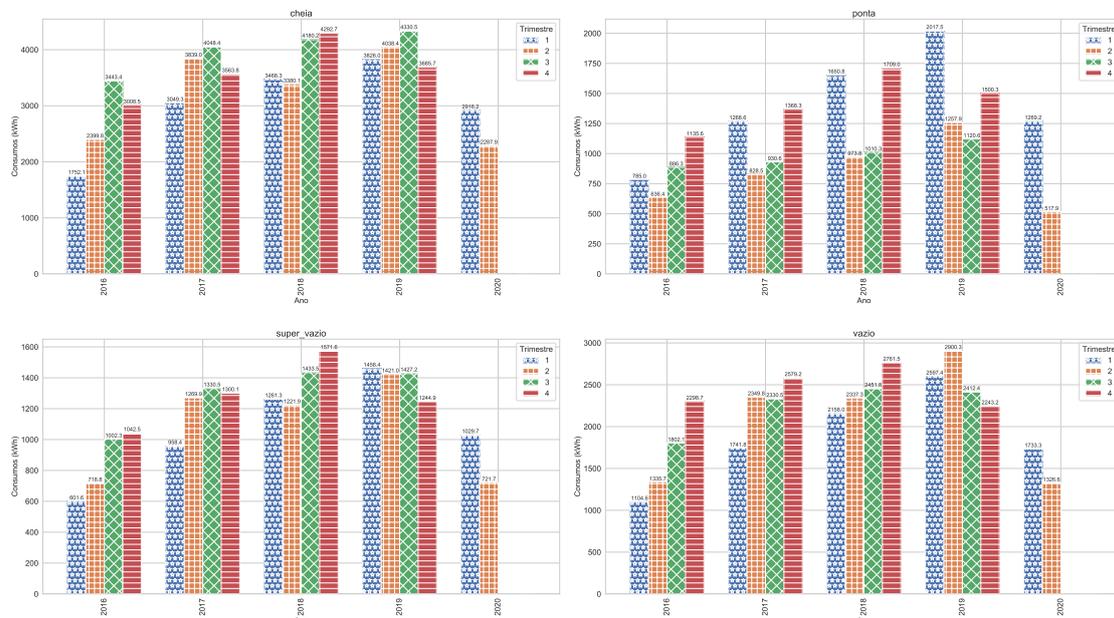


Figura 4.32: Valores médios da Electricidade de Guimarães em cada trimestre por indicador.

Para se perceber a disposição dos consumos pelas estações do ano, foi calculado os consumos energéticos médios em cada estação do ano, para todos os anos, com os resultados ilustrados na Figura 4.33. É possível verificar claramente que o Outono entre 2016 e 2018 foi a estação de ano que apresentou

os maiores consumos energéticos, sendo que neste mesmo período nota-se uma tendência crescente nos consumos entre o Inverno e o Outono, exceto em 2018, em que na Primavera os consumos baixam em relação ao Inverno. Em 2019 observa-se um nítido equilíbrio dos consumos nas estações do ano. Já em 2020, o Inverno apresenta notoriamente os consumos mais elevados.

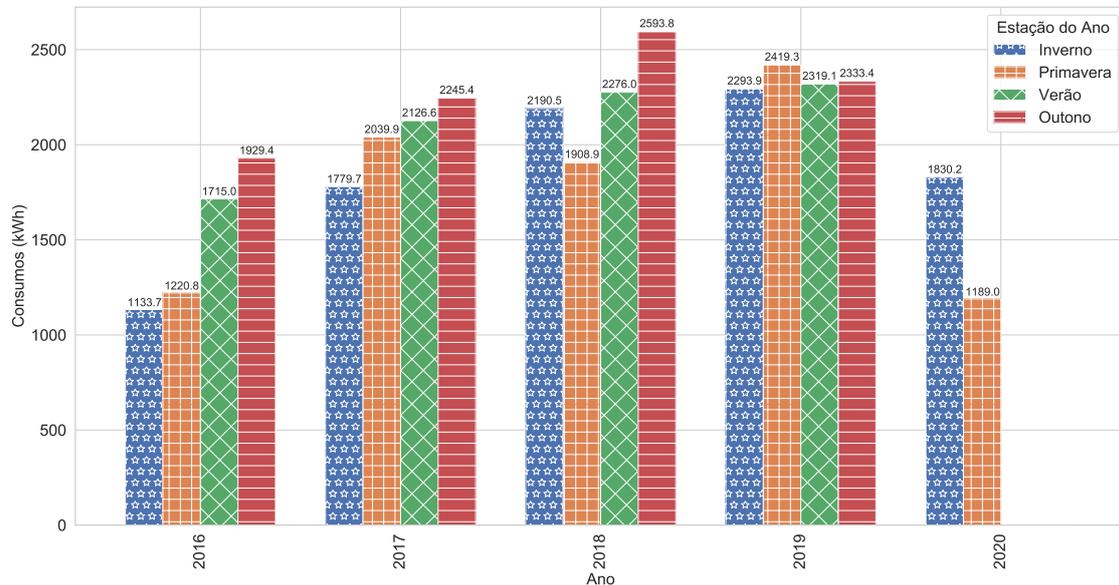


Figura 4.33: Valores Médios da Electricidade de Guimarães em cada estação do ano.

Olhando em particular para os distintos indicadores, que pode ser consultado na Figura 4.34, identifica-se que no indicador *cheia* o Verão, com exceção em 2018 e 2020, foi a estação do ano em que os consumos foram mais elevados, seguindo-se de muito perto o Outono. É possível observar também que em 2016, 2017 e 2019 os consumos aumentam entre o Inverno e Verão, onde no Outono apresentam um ligeiro decréscimo. Em relação a *ponta*, tirando 2016, o Inverno assume-se em todos os anos como a estação do ano com os maiores consumos energéticos, seguindo-se o Outono que, inclusive, em 2017 ultrapassa por pouco os consumos atingidos no Inverno. De ressaltar a existência de uma acentuada diferença entre as estações do ano Inverno e Outono com Primavera e Verão em todos os anos de registo. No que toca ao indicador *super vazio*, o Outono em 2016, 2017 e 2018 foi a estação do ano que utilizou mais energia, seguido do Verão. Nestes anos os consumos aumentam com o passar das estações do ano, onde se vê uma tendência crescente entre o Inverno e o Outono, menos em 2018 no qual a Primavera atinge valores menos elevados que o Inverno. De realçar um equilíbrio de valores entre todas as estações do ano em 2019. Por último, no indicador *vazio* em 2016, 2017 e 2018, o Outono volta a ser a estação do ano com os maiores consumos, acontecendo nestes mesmos anos praticamente o mesmo que no indicador *super vazio*, referido anteriormente. Em 2019 a Primavera destaca-se como a estação do ano com os maiores valores, estando as restantes em total equilíbrio.

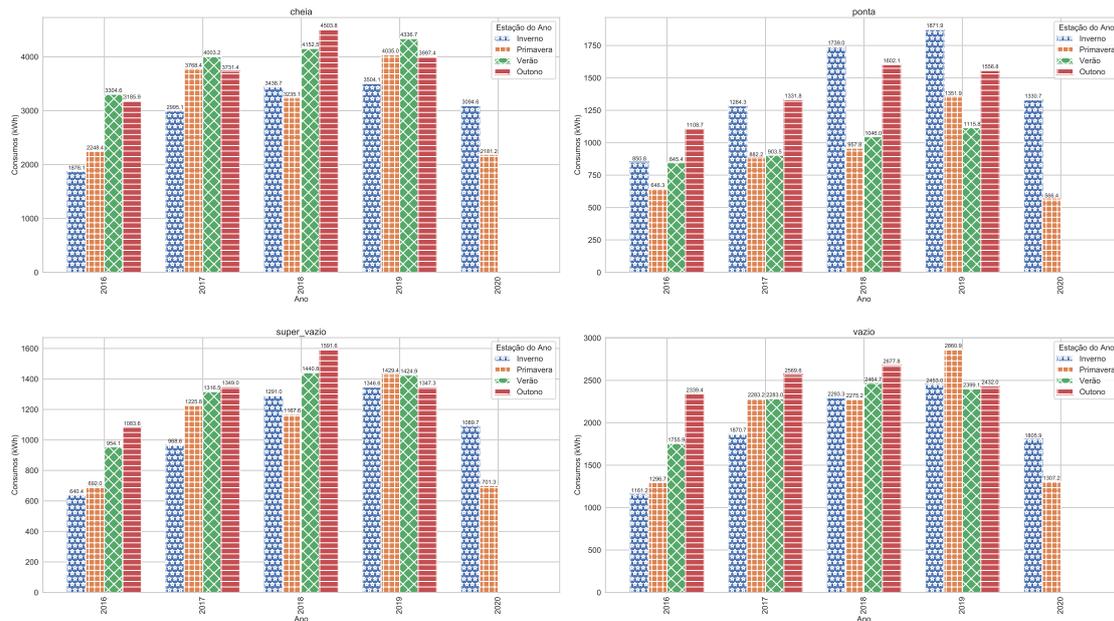


Figura 4.34: Valores Médios da Electricidade de Guimarães em cada estação do ano por indicador.

Por último, de modo a detetar as relações entre os indicadores da electricidade e, posteriormente, perceber até que ponto se relacionam entre si, foi realizado o cálculo das correlações entre eles. Para isso, foi novamente aplicado o teste de *Shapiro-Wilk* para identificar se os dados seguem uma distribuição normal. O resultado foi $p < 0.05$, concluindo-se que os dados não seguem uma distribuição Gaussiana. Com isso, para calcular as correlações existentes, foi utilizado o método de Coeficiente de Correlação Spearman, pelas razões já justificadas anteriormente, noutros tipos de dados. Na Figura 4.35 pode-se então observar os resultados obtidos. Como se pode examinar apenas se obteve uma correlação forte entre os indicadores *cheia* e *super vazio*, sendo a mesma positiva o que significa que se influenciam positivamente. Quanto a correlações moderadas contabilizou-se apenas uma entre os indicadores *super vazio* e *ponta*. As restantes são correlações fracas ou desprezíveis.

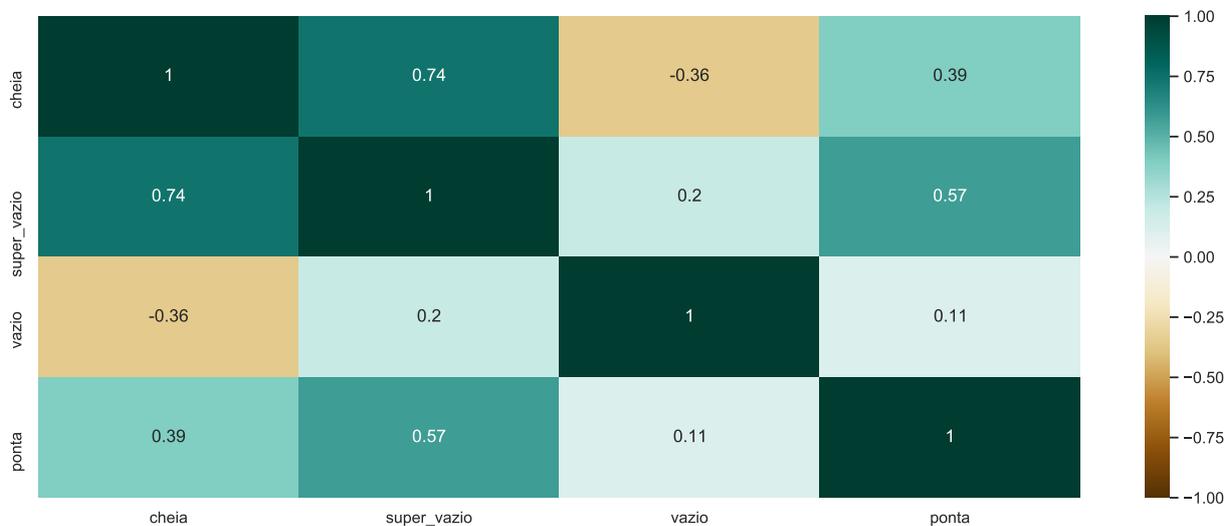


Figura 4.35: Correlações entre os indicadores da Electricidade de Guimarães.

4.2.2.2 Caudais

Relativamente aos dados dos caudais, inicialmente começou-se por realizar estatística descritiva nos dados, visto que nesta fase os mesmos já se encontram com o processo de tratamento concluído. Na Tabela 4.10 é possível verificar os resultados obtidos. Como se pode examinar na tabela, todos os setores apresentam valores significativamente mais baixos, quando comparado com os dados iniciais, onde se registam diferenças na ordem dos milhares, em todas as métricas calculadas. De realçar que o *Tanque Neutralização Caudal Faturação* destaca-se largamente como o setor com os valores médios mais elevados, seguido de *Obra de Entrada Escorrências* que também assume valores elevados em comparação com os restantes setores. Também convém salientar que o setor *Calvelos* que surgia nos dados iniciais foi descartado devido a não acrescentarem nada ao estudo, uma vez que todos os registos continham o mesmo valor.

Tabela 4.10: Dados Caudais após tratamento.

Setor	Unidade Medida	Registos	Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo	Periodicidade
Bugio	mm ³	246	109.320	115	113.763	50.475	2	343	Dois em Dois dias
Tanque Neutralização Caudal Faturação	mm ³	1609	9798.457	4716	9393	5182.877	2598	118159	Diária
Caudal PqIndSocorro	mm ³	610	347.763	140.5	279.634	358.987	26	5607	Diária
Central Preparação Polielectrólito Desidratação Centrífuga Bomba 1	mm ³	1609	7.672	11	7	4.267	1	99	Diária
Central Preparação Polielectrólito Desidratação Centrífuga Bomba 2	mm ³	1609	7.194	6	6.557	3.633	1	55	Diária
Escavanca	mm ³	224	115.467	125	114.081	25.463	5	243	Dois em Dois dias
Obra de Entrada Escorrências	mm ³	1609	1011.266	975	1005	299.983	52	3254	Diária
Fossas Sépticas	mm ³	1512	43.477	148.378	23.833	57.316	1	528	Diária
Lamas Flotadas	mm ³	1607	10.905	5	5.27	20.317	1	436	Diária
Central Preparação Polielectrólito Espessamento Tambor Bomba 1	mm ³	1592	3.048	2.861	2.861	1.352	1	11	Diária
Central Preparação Polielectrólito Espessamento Tambor Bomba 2	mm ³	1608	3.034	3	2.749	1.916	1	30	Diária
Tanque Lamas 1	mm ³	1609	193.726	184	175.918	118.040	2	1137	Diária
Tanque Lamas 2	mm ³	1609	207.723	177	182	128.699	1	1330	Diária
Vima Água	mm ³	990	118.298	5	80.108	125.949	1	1382	Diária

De modo a perceber o comportamento dos setores ao longo dos meses dos anos, foi calculada a variação média dos mesmos, ilustrado na Figura 4.36. Como se pode examinar, na primeira metade do ano de 2018 os caudais apresentados foram consideravelmente superiores em relação aos outros anos, no mesmo período. Verifica-se também que os valores, em todos os anos salvo algumas variações, tendem a baixar até ao mês de agosto atingindo neste mês os valores mais baixos, voltando a subir até ao final do ano.

Considerando os setores individualmente, como se possuem cerca de 14 setores distintos, e para não tornar este tópico muito massivo, optou-se por se efetuar a análise a três deles. Assim sendo, os setores analisados neste tópico serão *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 1*, *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 2* e *Obra de Entrada Escorrências*, que podem ser observados na Figura 4.37. Em relação à *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 1* os valores seguem uma certa homogeneidade durante os meses, em todos os anos, exceto em algumas situações. Uma delas ocorre no ano de 2017 onde, entre os meses de Abril e Julho, ocorre uma descida acentuada do caudal, no qual em 2018, no mesmo período, o mesmo tem uma subida considerável. Também em 2018 regista-se um aumento evidente entre os meses de Agosto e Setembro. Relativamente à *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 2* também se verifica um certo equilíbrio dos valores ao longo dos anos, excetuando sobretudo dois casos. O primeiro no ano de 2018 onde de Julho a Outubro o caudal tem uma tendência

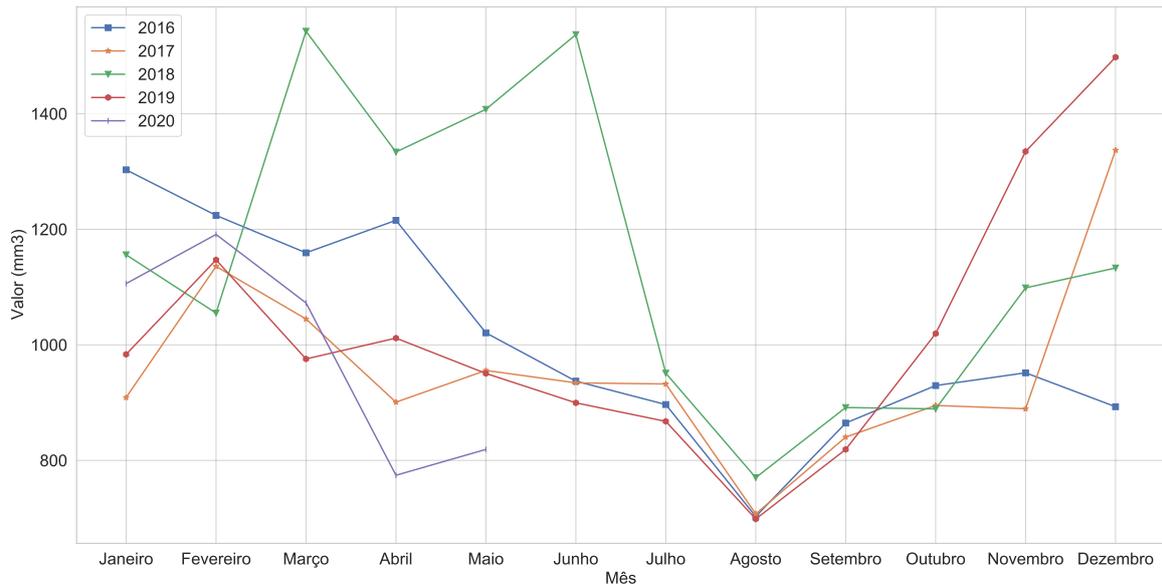


Figura 4.36: Variação Média dos Caudais ao longos dos meses em cada ano.

decrecente, e entre Outubro e Novembro onde é apresentado um crescimento abrupto. O outro caso remete-se ao ano de 2019 onde entre maio e Julho os valores descem a cada mês que passa, mas entre Julho e Setembro voltam a subir consideravelmente. Por fim, em *Obra de Entrada Escorrências* entre 2016 e 2019, o caudal assume uma tendência crescente nos seus valores especialmente entre Abril e Outubro, baixando posteriormente até ao fim do ano. Já em 2018 e 2019 os valores dos caudais exibem um equilíbrio nos seus valores em praticamente todos os meses, acentuando-se uma diferença apenas entre Novembro e Dezembro. Em relação a 2020 o fluxo assume um crescimento contínuo entre os meses de Março e Maio, tendo inclusive um crescimento acentuado entre Março e Abril.

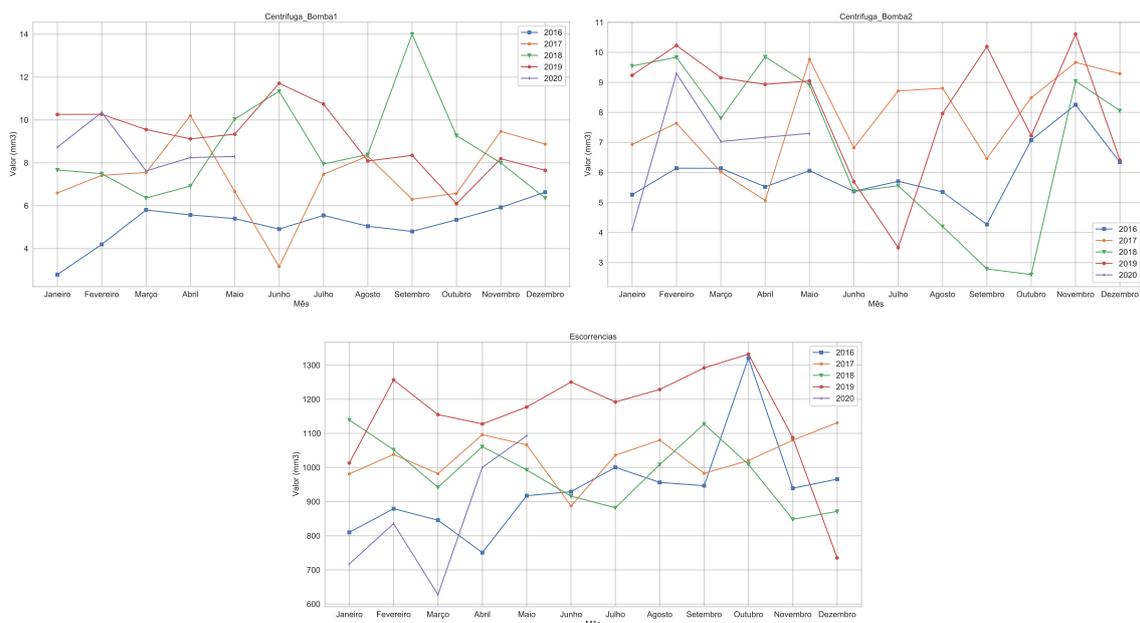


Figura 4.37: Variação Média de alguns indicadores dos Caudais ao longos dos meses em cada ano.

Visto isso, foi também calculado o caudal médio geral nos diferentes anos, no qual o resultado pode

ser consultado na Figura 4.38. Como se pode observar, foi no ano de 2018 que se registou uma maior caudal na ETAR. Já os restantes anos apresentam um equilíbrio nos seus valores, não divergindo muito uns dos outros.

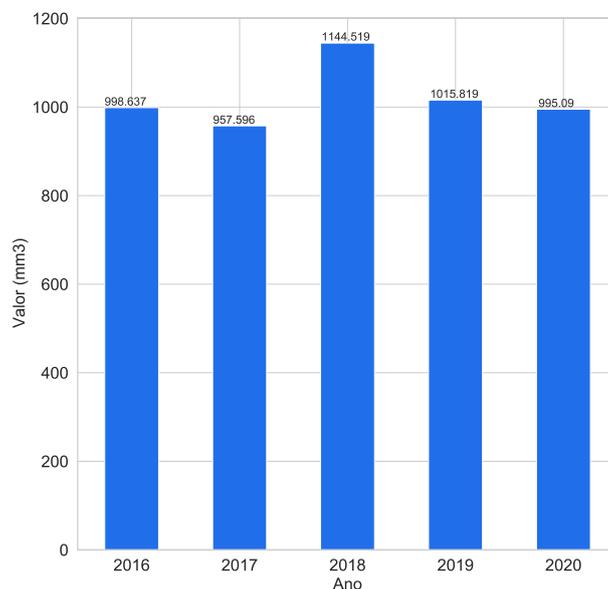


Figura 4.38: Valores médios dos Caudais em cada ano.

Olhando para os setores individualmente, que pode ser examinado na Figura 4.39, vê-se que na *Central Preparação Polieletrólito Desidratação Centrífuga Bomba 1* os valores com o passar dos anos tende a subir, neste caso entre 2016 e 2019. Já em 2020 nota-se um decréscimo nos valores. Relativamente à *Central Preparação Polieletrólito Desidratação Centrífuga Bomba 2* verifica-se uma pequena variação dos valores com o passar dos anos, mas de modo geral encontra-se com valores muito próximos entre si. Por último, na *Obra de Entrada Escorrências*, 2019 destaca-se como o ano em que o caudal médio foi mais elevado neste setor, seguindo-se do ano de 2017. O ano de 2020 por sua vez foi onde se registou um menor valor de caudal na ETAR.

Em relação à disposição dos dados pelos diferentes trimestres, foi realizado o cálculo do caudal médio em cada trimestre de todos os setores existentes, ilustrado na Figura 4.40. Através da análise da figura, é possível verificar que no 2º trimestre de 2018 foi onde se registaram os maiores valores do fluxo da infraestrutura. Também se observa a existência de um equilíbrio de valores no 3º e 4º trimestres entre 2016 e 2018, e entre os 1º trimestres em todos os anos. De realçar a diferença acentuada nos valores do 4º trimestre para os restantes, no ano de 2019.

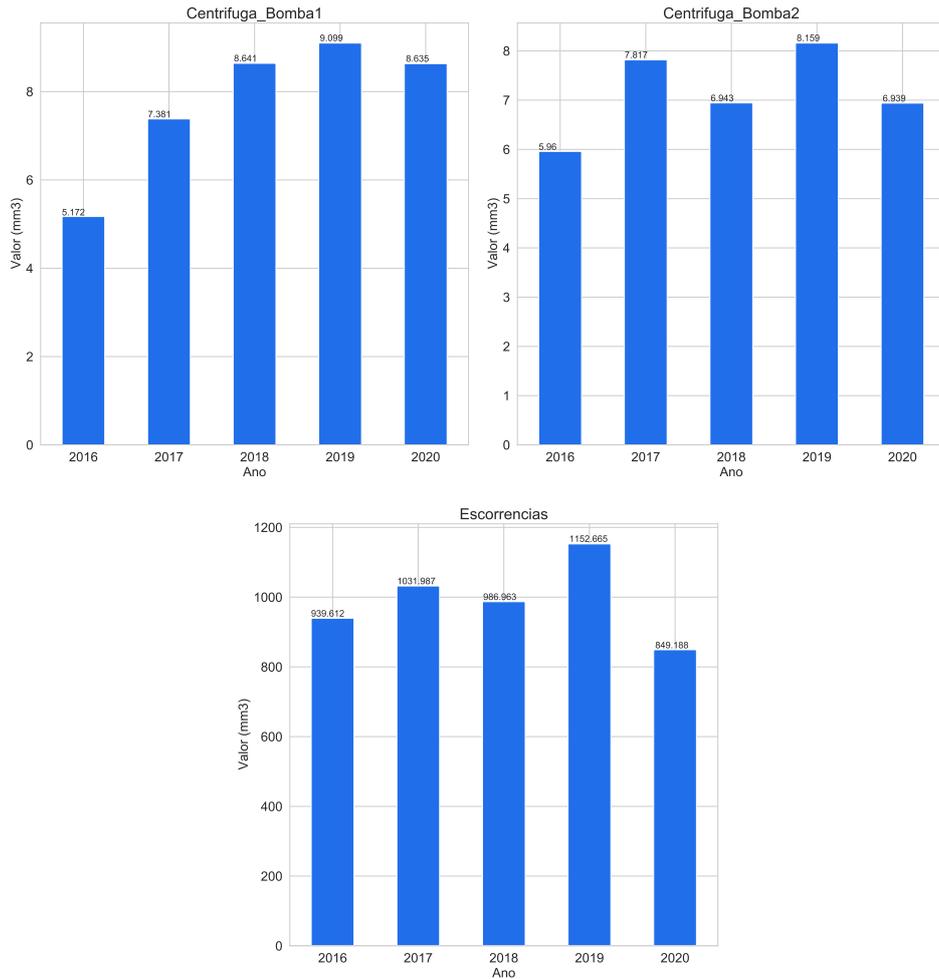


Figura 4.39: Valores médios de alguns indicadores dos Caudais em cada ano.

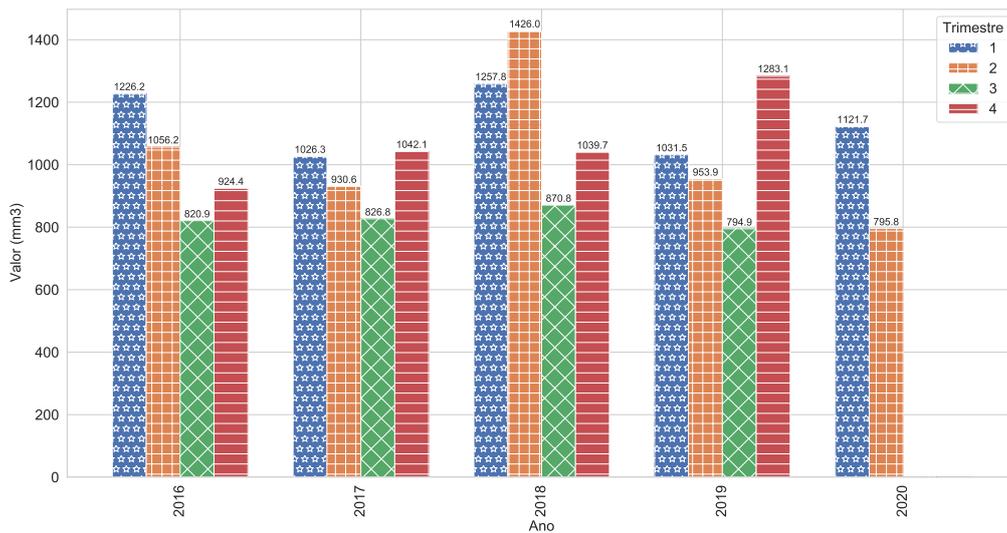


Figura 4.40: Valores médios dos Caudais em cada trimestre.

Considerando os setores individualmente, onde pode ser analisado na Figura 4.41, verifica-se que na *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 1*, o 2º e 3º trimestres, em 2018 e 2019, exibem os valores mais elevados neste setor. Já no ano de 2019, o 1º trimestre iguala o

2º trimestre. O ano de 2020 também exibe valores elevados nos trimestres, comparativamente com os restantes anos. Em relação à *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 2* o 3º trimestre foi o que registou menos caudal nos anos de 2016, 2018 e 2019. Do lado oposto, o 4º trimestre assume os valores mais elevados em 2016 e 2017 e o 1º trimestre em 2018 e 2019. Por fim, na *Obra de Entrada Escorrências* salta logo à vista um certo equilíbrio dos valores apresentados pelos trimestres em todos os anos. De salientar que em 2019, excluindo o 4º trimestre, foi onde os trimestres apresentaram os valores mais elevados, comparativamente com os restantes anos.

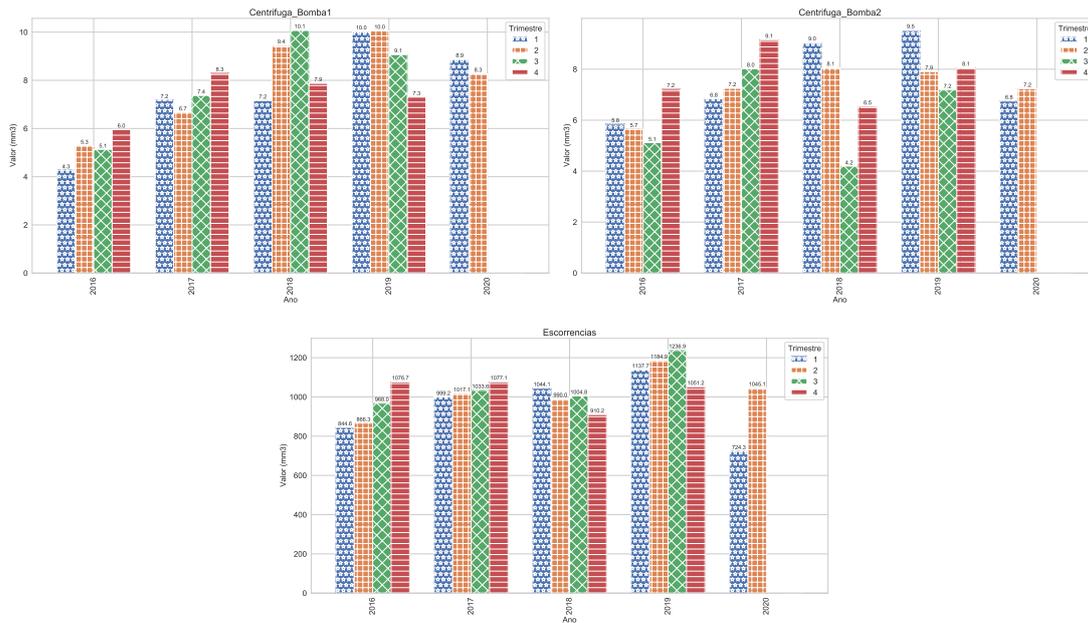


Figura 4.41: Valores médios de alguns indicadores dos Caudais em cada trimestre.

Olhando para a distribuição dos dados pelas estações do ano, foi calculado o valor médio de todos os setores em cada estação do ano, ilustrado na Figura 4.42. Com a análise da figura é possível observar que foi na Primavera, no ano de 2018, que o caudal médio foi maior. Verifica-se também que o Outono abarca os maiores valores nos anos de 2017 e 2019. Já o Inverno em 2016 e 2020 acolhe os valores mais elevados nestes anos. De salientar que a Primavera assume sempre, em todos os anos de registo, maior caudal médio que o Verão, apresentado uma diferença considerável em 2018.

Olhando para os setores de modo individual, identifica-se que na *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 1* existe uma estabilidade dos valores em todas as estações do ano, durante os vários anos representados. A Primavera é responsável pelos maiores valores médios de caudais em 2018 e 2019, sendo superada pelo Outono em 2016 e 2017 e pelo Inverno em 2020. No que toca à *Central Preparação Polielectrólito Desidratação Centrífuga Bomba 2* os maiores valores atingidos são no Inverno, em 2018 e 2019, e no Outono, em 2017. Em 2020 o Inverno e a Primavera acabam por apresentar praticamente o mesmo valor médio. De realçar que o Verão, com exceção em 2017, é a estação do ano com os menores valores nos anos em que tem registo. Por último, na *Obra de Entrada Escorrências*, sobretudo entre 2017 e 2019, existe um equilíbrio dos valores apresentados pelas estações do ano. Também se observa que o Verão em 2016, 2018 e 2019 supera o Inverno e a

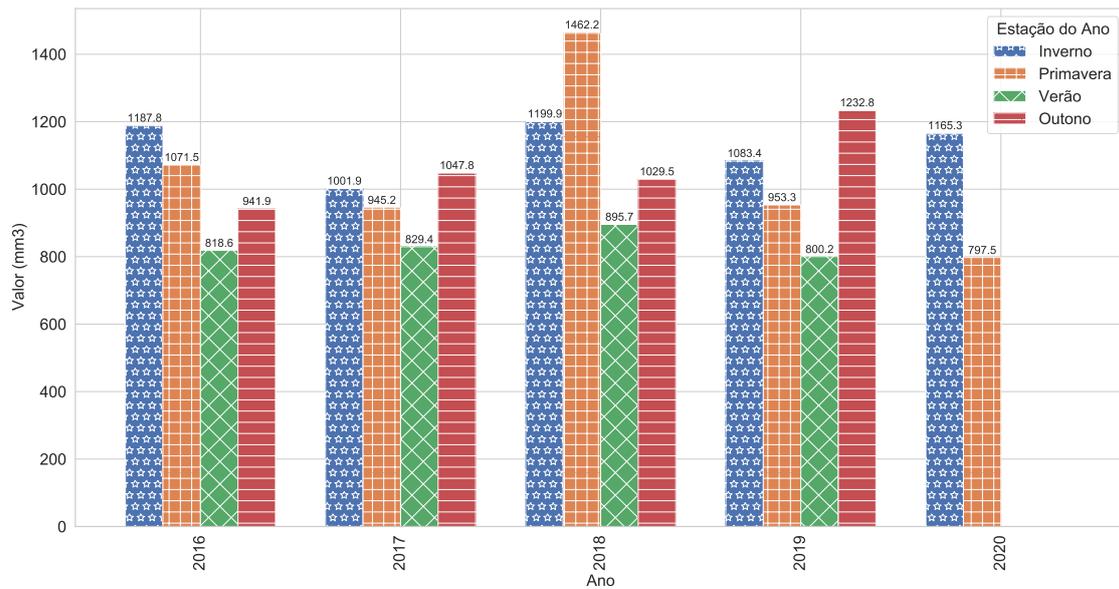


Figura 4.42: Valores Médios dos Caudais em cada estação do ano.

Primavera, sendo inclusive a estação do ano com os valores mais elevados em 2019. Já em 2020, a Primavera atinge claramente valores mais elevados que o Inverno.

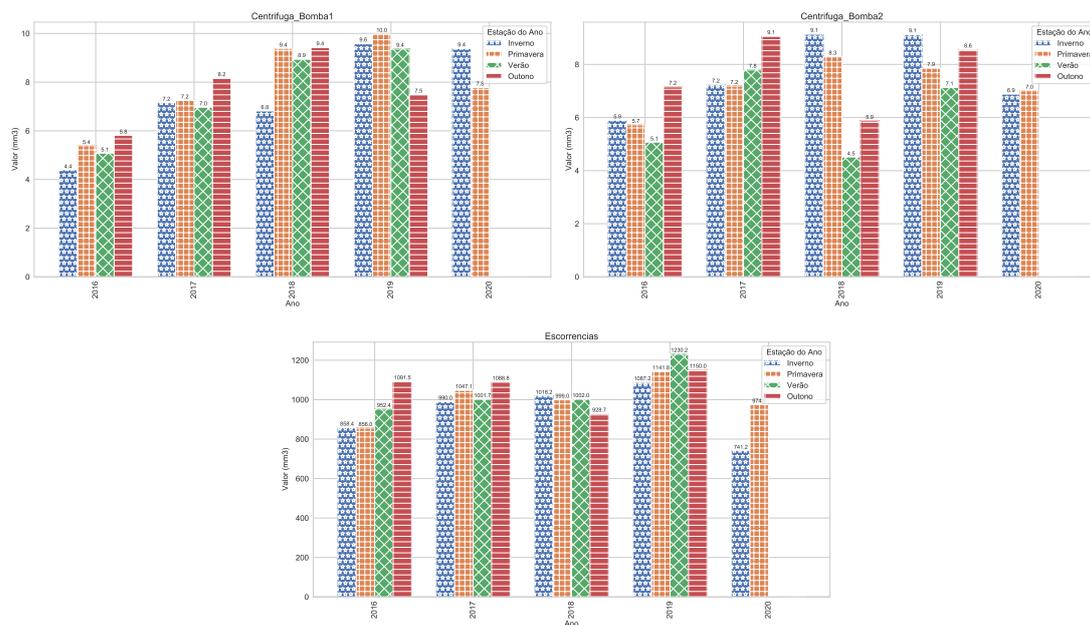


Figura 4.43: Valores Médios de alguns indicadores dos Caudais em cada estação do ano.

4.2.2.3 Relação entre os diferentes tipos de dados

Com o intuito de se identificar possíveis relações entre os dados da Electricidade e Caudais desta ETAR e perceber até que ponto têm influência um no outro, foram calculadas as correlações entre os mesmos. Na Tabela 4.11 podem ser consultados os resultados obtidos. Como se pode analisar, as maiores correlações detetadas foram duas correlações fracas, o que nos diz que os dados da Electricidade e Caudais não se

relacionam, onde apenas existe uma pequena influência entre *Ponta* com *Vima Água* e *Super Vazio* com *Obra de Entrada Escorrências*.

Tabela 4.11: Maiores correlações encontradas entre Electricidade e Caudais.

Electricidade	Caudais	Valor Correlação
Ponta	Vima Água	0.369
Super Vazio	Obra de Entrada Escorrências	0.342

Por último, na Tabela 4.12 podem ser visualizados os resultados obtidos das correlações calculadas entre os dados do Tempo e da electricidade. Como pode ser analisado, os dados do tempo com a Electricidade claramente não se relacionam, uma vez que apenas existe uma correlação fraca entre eles, mais concretamente entre o indicador meteorológico *temperatura* com o indicador *ponta*. Portanto não há qualquer tipo de relação relevante.

Tabela 4.12: Maiores correlações encontradas entre o Tempo e os restantes tipo de dados.

Tempo	Electricidade	Valor Correlação
Temperatura	Ponta	-0.32

4.3 Plataforma de Visualização

De maneira a apresentar toda a informação e permitir aos gestores das *ETAR* visualizá-la de forma organizada, intuitiva e eficaz, foi desenvolvida uma plataforma para o efeito, tratando-se de um *website* onde está integrado todo o trabalho concebido até este ponto. Nele estão presentes múltiplos *dashboards* para cada tipo de dados, onde é exibida toda a informação relevante de cada um com a possibilidade de aplicar diversos filtros de informação. Além disso, a plataforma permite também que sejam inseridos novos dados, que automaticamente serão processados, tratados e inseridos na base de dados, sendo posteriormente a informação atualizada nos respetivos *dashboards*.

Para se perceber melhor de que forma funciona a plataforma, na Figura 4.44 está então representada a arquitetura da mesma. Como pode ser visualizado na figura, todos os dados coletados da *ETAR* estão armazenados num servidor *Cloud*, onde o seu armazenamento segue já uma estrutura específica para que a consulta aos mesmos seja realizada de forma mais simples e clara possível. Posteriormente, caso os responsáveis da infraestrutura pretendam inserir novos dados, a plataforma permite que essa inserção seja realizada de forma muito simples, como já referido. Basta que se dê *upload* do ficheiro de dados e submeter, uma vez que no servidor esses dados são devidamente tratados de forma automática, passando por um conjunto de validações que deteta os tratamentos necessários a aplicar. Todo este processo é realizado utilizando a linguagem de programação *Python* e as respetivas bibliotecas, como por exemplo *Pandas* para o tratamento dos dados. Por essa razão, a comunicação entre o servidor e a plataforma é conseguida utilizando a *framework Flask*, onde se consegue efetuar uma transição rápida e simplista da informação. Relativamente ao *website*, foi desenvolvido utilizando diferentes linguagens de programação

e frameworks como *HTML*, *Bootstrap*, *JavaScript* e *CSS*, tendo sempre em vista transmitir a melhor experiência de navegação e percepção dos conteúdos ao utilizador.

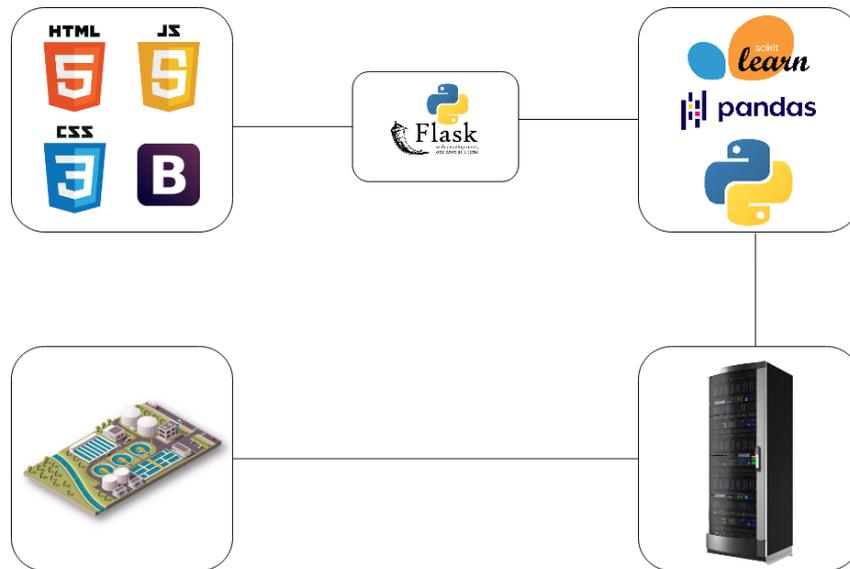


Figura 4.44: Arquitetura da plataforma desenvolvida.

De modo a elucidar melhor a plataforma que foi desenvolvida, as seguintes figuras demonstram o resultado final. Nas Figuras 4.45, 4.46 e 4.47 é ilustrado como está disposta toda a informação relevante dos dados, neste caso em particular, dos dados da electricidade da ETAR de Vila Real. Na Figura 4.45 verifica-se que a plataforma disponibiliza um conjunto de filtros para se aplicar nos dados, para um análise mais específica. Apresenta também diversas informações, como por exemplo, média, mediana e soma dos valores. Além disso, exibe alguns *dashboards* onde os dados estão representados, como por exemplo, a variação média da electricidade ao longo dos meses.

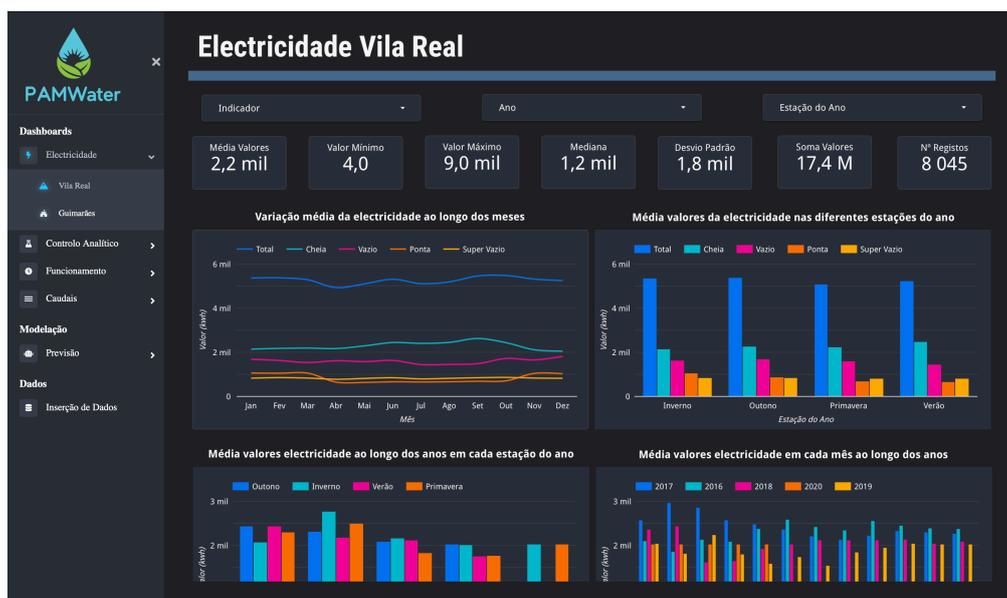


Figura 4.45: Ilustração na plataforma dos dados da electricidade da ETAR de Vila Real.

Já na Figura 4.46 identifica-se um conjunto de *dashboards* que exibem informações sobre os dados. Vê-se, por exemplo, a média dos valores da electricidade ao longo dos anos em cada estação do ano ou a média dos valores da electricidade em cada dia da semana ao longo dos anos.



Figura 4.46: Ilustração na plataforma dos dados da electricidade da ETAR de Vila Real.

Por fim, na Figura 4.47 observam-se *dashboards* que representam a distribuição da electricidade em diferentes cenários, tais como, por indicador, ano, estação do ano ou trimestre. Além disso, a plataforma fornece também uma tabela que representa os valores dos dados, em todos os registos existentes.



Figura 4.47: Ilustração na plataforma dos dados da electricidade da ETAR de Vila Real.

Relativamente à ilustração dos dados do controlo analítico da ETAR de Vila Real na plataforma, é possível verificar nas Figuras 4.48 e 4.49 parte do resultado final. Como se pode observar na Figura

4.48, estão aplicados filtros para a visualização dos dados, nomeadamente o indicador *Azoto Total* no *Efluente Tratado*, onde mostra diversas informações sobre os seus valores e variação.



Figura 4.48: Ilustração na plataforma dos dados do Controlo Analítico da ETAR de Vila Real.

Na Figura 4.49 identificam-se diversos *dashboards* que ilustram distintas informações dos dados, como por exemplo, a média dos valores do indicador ao longo dos anos, em cada semestre.



Figura 4.49: Ilustração na plataforma dos dados do Controlo Analítico da ETAR de Vila Real.

4.4 Tecnologias Utilizadas

Para o desenvolvimento de praticamente todo o projeto, foi utilizada a linguagem de programação *Python*. Todos os processos desde a análise dos dados brutos até a análise exploratória dos dados foi elaborada

utilizando o *software Jupyter Notebook*. Em relação à parte da modelação, todas as etapas até a construção dos modelos foram feitas no *software Spyder*. Já a parte de treino e *tuning* dos modelos foi concebido no *Google Colaboratory*, que se trata de um serviço *cloud* gratuito disponibilizado pela *Google*, que oferece grandes recursos computacionais, o que facilita este desenvolvimento. Além disso, utilizou-se também bibliotecas importantes, como por exemplo, *Pandas*, *Matplotlib*, *Numpy* para a parte de análise e tratamento dos dados, e *Scikit-learn* para a concepção dos modelos. Em relação à plataforma desenvolvida, utilizou-se as linguagens de programação *HTML*, *CSS* e *JavaScript*.

Casos de Estudo

Este capítulo aduz o caso de estudo que foi realizado neste projeto, que tem como finalidade fornecer mais um auxílio aos gestores das **ETAR** na difícil gestão destas mesmas infraestruturas. Portanto na secção 5.1 é descrito o que motivou a realização deste caso de estudo e do que se trata o mesmo, sendo demonstrados todos os pormenores. Na secção 5.2 é detalhado todo o processo de análise e tratamento dos dados, ilustrando as abordagens que foram seguidas e as técnicas aplicadas para a construção do *dataset* final. Já na secção 5.3 são relatadas as experiências realizadas para a construção e otimização dos modelos utilizados, onde é especificado quais as técnicas e ferramentas aplicadas neste processo. Por fim, na secção 5.4 são ilustrados os resultados adquiridos, mostrando as diferenças entre os modelos usados, assim como as respetivas conclusões.

5.1 Enquadramento

A gestão de uma infraestrutura como uma **ETAR** é uma tarefa muito complexa, pois as características das **AR** são sempre uma incógnita, dado que existem muitos fatores que podem alterar completamente as características das mesmas, e fugir do que é expectável. Com isto, a inteligência artificial (mais concretamente o uso de modelos de **ML**) pode exercer um papel fundamental na gestão destas infraestruturas, podendo prever valores futuros de certos indicadores presentes nas **AR** e garantir que, caso haja a previsão de valores indesejados, certas medidas sejam executadas, a fim de assegurar que as descargas de **AR** cumprem com os limites estabelecidos por lei.

Portanto, este caso de estudo surgiu pela clara necessidade de auxiliar, noutra patamar, os responsáveis das **ETAR** na gestão destas instalações. Por isso, foca-se então nos dados relativos ao controlo analítico uma vez que os mesmos representam os valores das **AR** em vários setores da **ETAR**, como por exemplo, no *afluente bruto*, *lamas mistas*, entre outros. Desse modo, o objetivo deste caso de estudo é possibilitar que os gestores das **ETAR** tenham acesso a previsões dos valores dos indicadores presentes

no último setor (*efluente tratado*), antes de se proceder à descarga de AR para o meio ambiente, e saberem antecipadamente, se os mesmos vão assumir valores indesejados.

Dado que existem vários indicadores distintos, para este caso de estudo decidiu-se focar no indicador *azoto total*, que corresponde à soma de todas as espécies de nitrogénio encontradas numa amostra de água, como por exemplo, o nitrato-nitrogénio ($NO_3 - N$) ou nitrito-nitrogénio ($NO_2 - N$) [73]. Esta escolha justifica-se também pelo facto do tratamento deste indicador assumir uma elevada importância uma vez que, existindo valores acima do permitido desta substância nas descargas das AR, pode trazer consequências indesejadas para o meio aquático. Elevadas quantidades deste indicador nas descargas de AR pode por exemplo, aumentar o processo de eutrofização provocando uma degradação acentuada da qualidade da água nos ambientes aquáticos [74].

Além disto, dado que a meteorologia pode ter influência nos valores apresentados pelos indicadores das AR, foram também utilizados os dados climatológicos da cidade de Vila Real pois é onde se localiza a ETAR em questão para este caso de estudo. Esta inclusão teve como finalidade tentar obter mais variáveis que possam influenciar os valores das AR, mais precisamente, do *azoto total*.

5.2 Análise e tratamento dos dados

Como dito anteriormente, os dados utilizados foram os relativos ao controlo analítico na ETAR de Vila Real, no qual a análise inicial aos mesmos pode ser consultada em 3.2.1.2 e a análise exploratória em 4.2.1.2, onde também inclui o indicador *azoto total* no *efluente tratado*, que corresponde ao *target* deste caso de estudo. Como visto em 4.1.1.2, os dados neste ponto já se encontram tratados, ou seja, já não existe a presença de valores nem *timesteps* em falta. Contudo, para a resolução deste caso de estudo foram realizadas algumas manipulações adicionais nos dados. Iniciou-se com a aplicação de *feature engineering*, onde foram criadas três novas variáveis, o dia, mês e ano, com o propósito de encontrar novas variáveis que possam influenciar o valor do *target* em estudo. De seguida, para colocar todos os indicadores com os mesmos *timesteps* e se identificar quais dos indicadores apresentavam uma relação forte com o *target*, foi realizada uma espécie de "one-hot encoding" onde os indicadores passaram a ser representados por colunas, com o valor correspondente em cada data. Por exemplo, foi criada uma coluna chamada "*azoto_total_afluente_bruto*" com os respetivos valores do indicador *azoto total* no *afluente bruto* nas correspondentes datas. Na Tabela 5.1 podem ser consultados os valores obtidos do cálculo das correlações, onde foi aplicada a mesma abordagem que em 4.2.1.2. Como se pode verificar, identificaram-se 2 correlações fortes, todas com o indicador *amónia* nos setores *efluente tratado* e *câmara de degaseificação*. Registaram-se também 2 correlações moderadas com o indicador CQO no *efluente tratado* e *amónia* nas *lamas Biológicas 2*.

A disposição dos dados desta forma provocou o aparecimento de valores em falta, uma vez que os indicadores tinham periodicidades distintas e registos em datas diferentes. Como a maioria dos indicadores, inclusive o *target*, continham periodicidade semanal, todos eles foram colocados com essa periodicidade.

Tabela 5.1: Maiores correlações encontradas entre o *target* de estudo e os restantes indicadores.

Indicador	Valor Correlação
Amónia em Câmara de Degaseificação	0.842
Amónia em Efluente Tratado	0.813690
CQO em Efluente Tratado	0.551
Amónia em Lamas Biologicas 2	0.511

Posto isso, para se preencher os valores em falta foi calculada a média das 3 semanas anteriores, onde o resultado corresponde ao registo em falta.

De realçar que os dados climatológicos assim como as 3 novas variáveis criadas não mostraram correlações consideráveis com o *target* em estudo. Com isto, todas as variáveis que não apresentaram uma relação relevante com o *target* foram descartadas. Por fim, o *dataset* final continha cerca de 218 registos e 6 variáveis, onde a Tabela 5.2 exemplifica um registo do mesmo.

Tabela 5.2: Variáveis presentes no *dataset* final.

Variável	Exemplo observação
data	2019-10-14
azoto_total_em_Efluente_Tratado	13.7
amonia_em_Lamas_Biologicas_2	1.161
cqo_em_Efluente_Tratado	29.1
amonia_em_Efluente_Tratado	4.969
amonia_em_Camara_de_Degaseificacao	4.8

5.3 Experiências

Uma vez concluída a análise e tratamento dos dados, como descrito em 5.2, o próximo passo centrou-se na conceção de modelos candidatos, mais concretamente DT e RF. Foram realizadas diversas experiências, concebendo diversos modelos candidatos de forma a encontrar a melhor combinação de hiperparâmetros para cada um dos modelos supracitados. Tendo isso em conta, para cada experiência desenvolvida foi efetuada a avaliação da performance da mesma, tendo em conta as diferentes métricas de avaliação. As métricas utilizadas para avaliar a qualidade das previsões dos modelos candidatos foram:

- **Root Mean Square Error (RMSE):** Esta métrica corresponde ao desvio padrão dos erros de previsão do modelo. Por outras palavras, é uma medida de precisão onde mede a diferença entre os valores previstos pelo modelo e os valores reais. Em 5.1 está representada a sua equação.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5.1)$$

- **Mean Absolute Error (MAE):** Outra métrica utilizada foi a MAE, que calcula a média das diferenças entre os valores previstos e os reais. A utilização desta métrica visa reforçar a confiança dos valores obtidos por meio do modelo utilizado. A sua equação pode ser consultada em 5.2.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5.2)$$

Portanto, utilizou-se a ferramenta *GridSearchCV*, que efetua a procura pelos melhores hiperparâmetros do modelo, aplicando ao mesmo tempo a técnica de *cross-validation* em cada experiência, garantindo assim que o melhor modelo candidato é o mais generalista possível. Neste caso, foi utilizado na técnica de *cross-validation* um CV igual a 5, o que significa que o *dataset* foi "partido" em cinco partes distintas durante o processo de treino do modelo. O valor alcançado em ambas as métricas de erro corresponde à média dos valores obtidos, em cada uma, nas três divisões efetuadas, resultando o valor final que avalia o desempenho do modelo. Na Tabela 5.3 é descrita a pesquisa dos hiperparâmetros considerados para cada um dos modelos utilizados.

Tabela 5.3: Hiperparâmetros testados

Parâmetro	DT	RF	Descrição
max_depth	[5,10]	[5,12]	profundidade máxima
min_samples_split	[2,6]	[2,8]	amostras mínimas necessárias para dividir
min_samples_leaf	[2,4]	[2,6]	amostras mínimas necessárias para estar numa folha
max_features	[auto,sqrt,log2]	[auto,sqrt,log2]	número de <i>features</i> para a melhor divisão
n_estimators	-	[20,100]	número de árvores na floresta
splitter	[best,random]	-	estratégia usada para escolher a divisão em cada nó

Como se pode verificar, os hiperparâmetros procurados em cada modelo são muito similares, com exceção de *n_estimators* que apenas é utilizado pelas *RF* e o *splitter* onde só é usado nas *DT*. Contudo, os valores dos hiperparâmetros entre os modelos são um pouco distintos, devido ao facto dos modelos de *RF* serem mais complexos que as *DT*, e por isso, pode exigir valores mais altos nos hiperparâmetros para obter resultados superiores.

5.4 Resultados e discussão

Com a conclusão de todo o processo de experiências, foi necessário efetuar uma análise dos resultados obtidos. A Tabela 5.4 exhibe esses mesmos resultados, mais concretamente dos três melhores modelos candidatos de cada tipo.

Como pode ser observado, o melhor modelo candidato obtido foi uma *DT* com um *RMSE* de 2.384 e *MAE* de 1.416. Este modelo utiliza uma *max_depth* de 10, *max_samples_split* de 2 e *min_samples_leaf* de 2. Já o melhor modelo candidato de *RF* obteve um *RMSE* de 2.640 e *MAE* de 1.561. De ressaltar que, os valores dos hiperparâmetros são um pouco homogêneos nos três melhores modelos de *DT*, como o *min_samples_leaf* que é igual em todos eles. Por outro lado, nos três melhores modelos de *RF* existe um baixo nível de homogeneidade relativos aos hiperparâmetros. Também é importante referir que os modelos de *RF* utilizam consideravelmente mais tempo nos processos de treino quando comparados com os modelos de *DT*, o que já era expectável devido à maior complexidade destes modelos.

Tabela 5.4: Top-3 modelos DT e RT. Legenda: a - max_depth; b - max_samples_split; c - min_samples_leaf; d - max_features; e - n_estimators; f - splitter; g - RMSE; h - MAE; i - time (segundos).

a	b	c	d	e	f	g	h	i
<i>Modelos candidatos DT</i>								
10	2	2	auto	-	best	2.384	1.416	0.0168
9	3	2	auto	-	best	2.704	1.571	0.0165
9	5	2	auto	-	best	2.791	1.613	0.0158
<i>Modelos candidatos RF</i>								
12	2	2	auto	50	-	2.640	1.561	0.1883
7	3	3	auto	60	-	2.827	1.734	0.1773
10	2	3	auto	80	-	2.911	1.785	0.1603

Além disso, em relação ao hiperparâmetro *max_samples_split* nos três melhores modelos de DT, verifica-se que este valor tende a aumentar com o crescer do erro, enquanto que nos modelos de RF isso não acontece, onde este valor tende a ficar sempre muito próximo. De realçar que, nos três melhores modelos de DT e RF, os hiperparâmetros *max_samples_split* e *min_samples_leaf* assumem praticamente os mesmos valores em ambos os modelos.

Na figura 5.1 é ilustrada a previsão efetuada pelo melhor modelo candidato de DT e RF. Como se pode verificar, as previsões de ambos os modelos são bastante próximas, contendo algumas diferenças superficiais, como por exemplo nas observações 29 e 61. Portanto, ambos os modelos são bastante válidos, uma vez que obtêm resultados muito satisfatórios. Contudo, na generalidade o modelo de DT consegue obter previsões mais precisas que o modelo de RF.

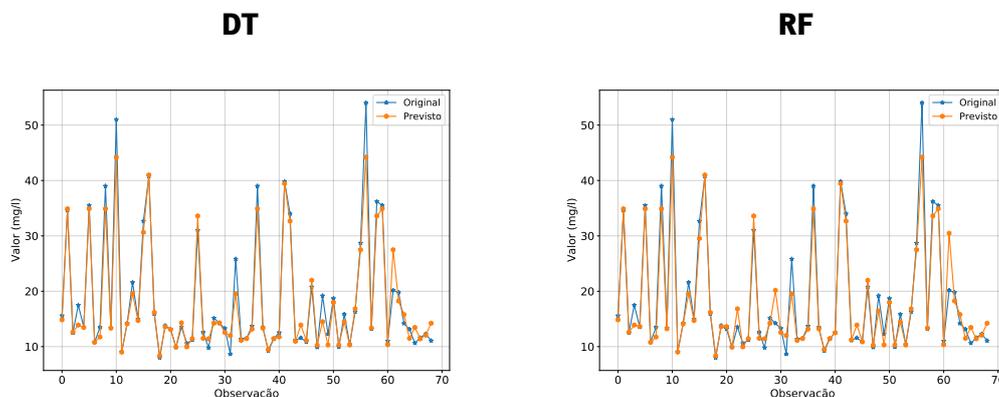


Figura 5.1: Previsão do azoto total utilizando os melhores modelos de DT e RF.

Com o intuito de fornecer ainda mais informações aos gestores das ETAR, foram também verificadas as regras que representam a declaração condicional, ou seja, que justificam os valores das previsões obtidas do melhor modelo candidato. Essas regras podem permitir aos responsáveis das ETAR saber os padrões existentes e o que contribui para o aumento indesejado dos valores do *target* em estudo. Na Tabela 5.5 é possível então verificar algumas das regras obtidas, onde são demonstrados alguns casos onde o *azoto total* no *efluente tratado* ultrapassa os valores permitidos por lei, valor esse que não pode ser superior a 15 mg/l nas descargas das AR. Por exemplo, pode-se constatar que o *azoto total*

no *efluente tratado* alcança o valor de 21.98 mg/l quando a *amónia* na *câmara de degaseificação* tem valor menor ou igual que 25.84 mg/l, a *amónia* no *efluente tratado* menor ou igual a 16.42 mg/l e a *amónia* nas *lamas biológicas 2* menor ou igual que 3.01 mg/l.

Tabela 5.5: Algumas regras definidas pelo melhor modelo candidato.

Regras	1	2	3	4	5
Amónia em Câmara de Degaseificação	≤ 25.84	> 25.84	≤ 27.43	≤ 10.60	≤ 10.60
Amónia em Efluente	≤ 16.42	≤ 6.24	≤ 6.24	> 13.15	> 6.24
CQO em Efluente	-	< 33.17	≤ 47.69	≤ 47.69	≤ 21.80
Amónia em Lamas Biológicas 2	≤ 3.01	≤ 2.03	> 2.34	≤ 1.30	≤ 1.25
Azoto Total em Efluente	21.98	41	35.83	18.50	29

De ressaltar que todo o trabalho realizado no caso de estudo foi implementado na plataforma desenvolvida, onde podem ser consultadas as previsões efetuadas através do uso de gráficos dinâmicos e tabela de resultados. Na Figura 5.2 é possível verificar como as previsões efetuadas são apresentadas, na plataforma, ao utilizador.

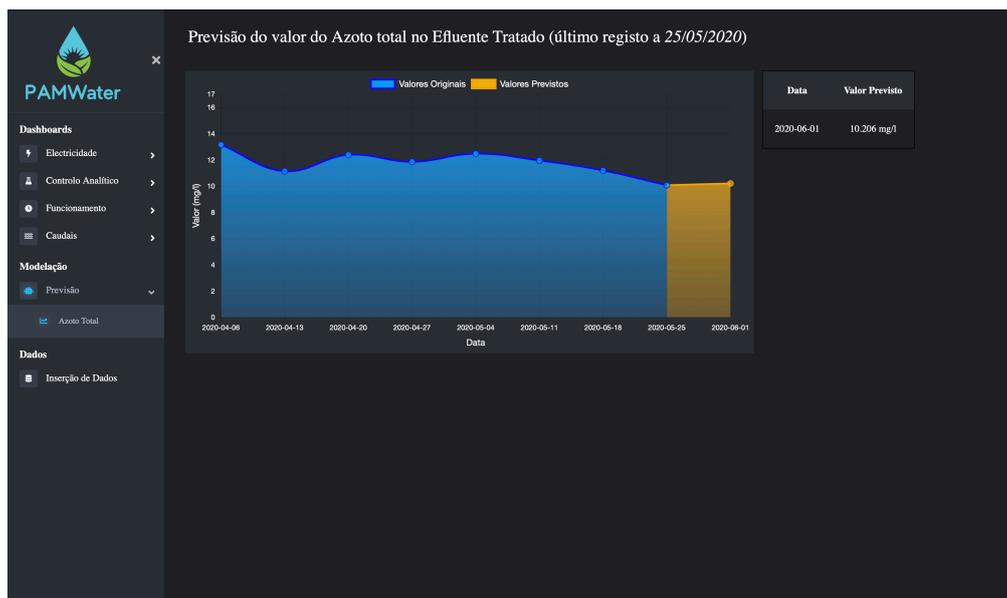


Figura 5.2: Ilustração na plataforma das previsões efetuadas.

Conclusão e Trabalho Futuro

Atualmente a sociedade enfrenta um enorme desafio, conhecido como sustentabilidade ambiental derivado, entre outros fatores, ao crescimento exponencial da população mundial. Esse crescimento acarreta diversos problemas com impactos muito negativos para o planeta, nomeadamente problemas relacionados com a água potável. Portanto, torna-se um tema bastante sensível, uma vez que se nada for feito no sentido inverso, as gerações futuras podem ficar comprometidas e enfrentar problemas bastante adversos, como a escassez de água consumível. Com isso, esta dissertação pode ser um auxílio na melhoria da gestão das ETAR (responsáveis pela despoluição das AR), onde tenta que haja um melhor controlo das substâncias presentes nas AR e, conseqüentemente, salvaguardar que valores indesejados na hora da descarga das AR sejam atingidos. Além disso, esta dissertação teve também o seu foco na melhoria da eficiência energética nestas infraestruturas, no qual analisa todos os tratamentos existentes e as respetivas substâncias, onde é exposta toda a informação de modo a poder clarificar onde podem ser realizados ajustes para a redução da energia consumida, sem nunca interferir na eficiência dos tratamentos.

Inicialmente foi efetuada uma investigação ao tema, onde foram expostos os conceitos mais cruciais sobre o mesmo. Começou-se por evidenciar os atuais problemas existentes com os recursos hídricos e a necessidade crescente de os corrigir ou atenuar, verificando-se que as ETAR desempenham um dos principais papéis para combater esta carência. Por isso, foi então feito um estudo minucioso a estas infraestruturas, onde se verificou quais as principais características e que tratamentos são realizados nas mesmas para tratar as AR. Além disso, muitos estudos relacionados ao tema foram analisados, onde foram identificadas algumas carências que foram tidas em conta para este trabalho, de modo a tentar corrigi-las. Outra ação que teve enorme relevância para a realização desta dissertação, foi a identificação e estudo de diversas substâncias presentes nas AR, o que permitiu adquirir uma maior sensibilidade na análise dos resultados obtidos.

Em relação aos dados da Electricidade, foram identificados inúmeros padrões e relações, em ambas as ETAR de estudo. Começando por Vila Real, verificou-se a ausência de relação entre os indicadores (fuso

horário energético), ou seja, não interferem entre eles, dizendo então que, por exemplo, o facto de um ter valores muito elevados não irá interferir com os restantes, exceto uma situação relativa aos indicadores *cheia* e *vazio*. Entre estes indicadores foi identificada uma relação no sentido negativo interessante, apesar de moderada, onde quando um cresce o outro diminui. Verificou-se também a existência de inúmeras variações características em cada indicador. Por exemplo, o indicador *ponta* assume os valores mais elevados nos meses de outono e inverno ao invés do indicador *cheia*, onde os seus maiores valores correspondem aos meses de primavera e verão. Identificou-se também que, na generalidade, os consumos médios a partir do ano de 2017 apresentam uma tendência decrescente em todos os indicadores, mas que em 2020 começam a ter o sentido inverso, sendo mais notório no indicador *vazio*.

Já na cidade de Guimarães, identificou-se uma maior relação entre os indicadores, existindo já uma relação forte e outra moderada, sendo o indicador *super vazio* o mais preponderante, uma vez que se encontra nestas maiores relações identificadas. Nesta [ETAR](#) verifica-se que, em todos os indicadores, os consumos tem as suas variações mas no mês de Agosto apresentam uma quebra. Observa-se também que, tal como na [ETAR](#) de Vila Real, o indicador *ponta* nos meses de primavera e verão assume claramente os seus valores mais baixos, disparando os consumos nos meses de outono e inverno. Em relação aos consumos médios por ano em cada indicador, tem-se uma clara tendência crescente com o passar dos anos, onde de 2016 a 2019 os consumos aumentam a cada ano que passa mas em 2020 parece ocorrer o contrário, registando-se uma clara decadência nos mesmos. É importante realçar que esta quebra pode estar relacionada ao facto do ano de 2020 apenas conter dados compreendidos entre os meses de Janeiro e Maio, portanto é provável que com os registos até ao final deste ano os consumos médios possam subir.

Relativamente aos dados do Controlo Analítico, apurou-se que os diferentes indicadores estudados, nos seus respetivos setores, apresentam uma relação bastante próxima. Essas relações ocorrem todas no setor *afluente bruto*, onde os indicadores *azoto total*, *CBO* e *CQO* se influenciam fortemente entre si. Também foi possível ver que noutros setores, estes indicadores também se relacionam, mas de forma moderada. Foi verificado também que no setor *poço escorrências* é onde os indicadores *azoto total* e *CQO* assumem, de modo claro, os seus valores mais elevados com uma diferença considerável comparativamente aos restantes setores. De ressaltar também que, no indicador *azoto total*, é nos meses de primavera e verão que no setor *efluente tratado* os valores são maiores, estando inclusive acima do que é desejável, uma vez que o valor em média ultrapassa o máximo permitido por lei (15 mg/l).

Olhando como os dados climatológicos de Vila Real se relaciona com os restantes dados da [ETAR](#) desta cidade, foram obtidas algumas conclusões. Pode ser visto que o indicador *temperatura* é o que tem mais influência com os restantes dados. Por exemplo, em relação aos dados do controlo analítico, as três maiores relações encontradas foram todas com o indicador *temperatura*. Já na cidade de Guimarães, os dados do tempo desta cidade não apresentam relação com os dados da electricidade, onde apenas se identificou uma correlação fraca.

Observando como todos os dados da [ETAR](#) de Vila Real se relacionam, verifica-se a existência de relações interessantes. Relativamente aos dados da Electricidade com Controlo Analítico encontrou-se quatro

correlações moderadas, todas elas com indicador *ponta* da Electricidade, mostrando que aquelas substâncias em questão nos respetivo setores, influenciam o valor da electricidade. Em relação à Electricidade com Funcionamento, como já era expectável, foram detetadas mais relações relevantes. São um total de nove relações (duas fortes e sete moderadas), sendo sete delas com o indicador *ponta* e as restantes com *vazio*. Portanto, conclui-se que os processos influenciam sobretudo os valores do indicador *ponta* da electricidade. Por último, entre os dados do Controlo Analítico e Funcionamento, são inúmeras as relações existentes, sendo mais de metade, de todas as relações possíveis, com valores relevantes. Isto prova que os valores do controlo analítico tem efeitos diretos no funcionamento, e vice-versa.

Em relação aos dados da ETAR de Guimarães, mais concretamente entre a Electricidade e Caudais, conclui-se que não se influenciam, uma vez que as maiores correlações encontradas foram fracas. Logo os valores de um não possui praticamente efeito nenhum no outro.

Para além da aplicação de *Data Science* aos dados onde se conseguiu extrair imensos *insights*, foi realizado também o caso de estudo que visa auxiliar ainda mais na gestão das ETAR. O caso de estudo foca-se na previsão do azoto total no *efluente tratado*, onde foram utilizados dois modelos de ML para o efeito. Trata-se de dois modelos baseados em árvore (DT e RF), onde esta opção teve por base dois motivos cruciais. Um deles é o reduzido tamanho do *dataset* utilizado onde, por exemplo, o uso de DL não seria muito eficaz, devido a necessitar de mais dados para demonstrar a sua *performance*. A outra razão remete-se à intenção de extrair do melhor modelo as "regras" que esclarecem os valores que o mesmo prevê, onde isso é obtido facilmente nos modelos baseados em árvore. Portanto, o melhor modelo obtido foi uma DT com um MAE de 1.612 e um RMSE de 2.207, provando que é um modelo que consegue ser bastante preciso nas suas previsões. Contudo, as RF também mostraram resultados promissores, sendo muito próximos das DT, revelando serem também um modelo altamente válido para o efeito. Um ponto interessante de ressaltar é que o modelo "mais simples" conseguiu obter resultados ligeiramente superiores, do que uma maior complexidade por parte das RF. Isto pode ser justificado pela pequeno tamanho do *dataset*, onde uma DT pode superar uma RF, com significativamente menos tempo de processamento. Este caso de estudo pode então ser uma ferramenta muito útil aos gestores das ETAR, uma vez que para além de poder prever valores indesejados da substância, indica também o que pode estar a influenciar os seus valores, sendo assim mais fácil proceder-se aos devidos ajustes na infraestrutura. De realçar também que este caso de estudo fez parte da *International Conference on Distributed Computing and Artificial Intelligence* de 2021, onde foi apresentado a todos os membros desta conferência [75].

Por fim, com a inclusão de todo este trabalho numa plataforma dinâmica e intuitiva, é possível concluir que podem ser antecipadas situação problemáticas nas ETAR, não apenas através da análise completa aos dados, mas também pelo processo de previsão criado. Isto permite aos responsáveis das infraestruturas estarem continuamente a par dos valores praticados nas diversas secções, nomeadamente energia e controlo analítico, assim como agir ou ficar em alerta quando sejam previstos valores indesejados da substância estudada, precavendo as consequências negativas. De evidenciar que o processo de previsão pode facilmente ser replicado a outros tipos de dados ou substâncias.

Por conseguinte, como trabalho futuro propõem-se:

- Recolher mais dados para enriquecer os *datasets* utilizados e com isso, através da aplicação de *Data Science*, extrair ainda mais *insights* que poderão fornecer novas conclusões e, consequentemente, ajudar no processo de tomada de decisão.
- Considerar adicionar mais substâncias que estejam presentes nas *AR*, mas que ainda não estão nos dados. Essa inclusão poderia permitir mais correlações com o *target* em estudo, e consequentemente, melhorar o processo de previsão.
- Aplicar a mesma abordagem só que em outra substância que desempenhe uma preocupação acrescida na gestão da infraestrutura, ou então a outro tipo de dados, como por exemplo, aos dados da electricidade para prever o consumo energético.
- Com um *dataset* maior usar *DL* para o processo de previsão. Utilizar, por exemplo, *Recurrent Neural Networks (RNN)* que já demonstraram obter um ótimo desempenho em problemas de séries temporais.

Bibliografia

- [1] D. d. S. Vaz. “Alterações climáticas, riscos ambientais e problemas de saúde: breves Considerações”. Em: *VI Seminário Latino Americano de Geografia Física II, Seminário Ibero Americano de Geografia Física, Universidade de Coimbra. Disponível em: <http://www.uc.pt/fluc/cegot/VISLAGF/actas/tema4/dirley>* (2010).
- [2] A. Kunz, P. Peralta-Zamora, M. de e N. Duran. “Novas Tendências no tratamento de efluentes têxteis”. Em: *Química Nova* 25 (fev. de 2002). doi: [10.1590/S0100-40422002000100014](https://doi.org/10.1590/S0100-40422002000100014).
- [3] W. H. Organization, UNICEF et al. “Progress on drinking water and sanitation: special focus on sanitation”. Em: *Progress on drinking water and sanitation: special focus on sanitation*. 2008, pp. 58–58.
- [4] J. Mateo-Sagasta, L. Raschid-Sally e A. Thebo. “Global wastewater and sludge production, treatment and use”. Em: *Wastewater*. Springer, 2015, pp. 15–38.
- [5] USGS. *Wastewater Treatment Water Use*. 2020.
- [6] P. Editora. *ETAR*. 2020. url: [https://www.infopedia.pt/\\\$etar](https://www.infopedia.pt/\$etar).
- [7] F. Liu, A. Ouedraogo, S. Manghee e A. Danilenko. *A Primer on Energy Efficiency for Municipal Water and Wastewater Utilities*. World Bank Group, 2012.
- [8] G. do Secretário de Estado do Ambiente. *Pensar 2020 - Uma Estratégia ao Serviço da População: Serviços de Qualidade a um Preço Sustentável*. Vol. 2. 2020.
- [9] Águas do Norte. *Tratamento de Águas Residuais*. 2020. url: <http://www.adnorte.pt/pt/educacao-ambiental/ciclo-urbano-da-agua/tratamento-de-aguas-residuais/?id=116>.
- [10] Águas de Portugal. *O ciclo urbano da água*. 2020. url: <https://www.adp.pt/pt/atividade/o-que-fazemos/o-ciclo-urbano-da-agua/?id=28>.
- [11] Geonatura. *O que é uma ETAR?* 2020.
- [12] P. Editora. *Funcionamento de uma ETAR*. 2020. url: https://www.portoeditora.pt/conteudos/emanuais/emanuais2014/32918/recursos/vter8_fl_etar.pdf.
- [13] J Obiols. “The importance of analytical control”. Em: *Options Méditerranéennes* 26 (1997), pp. 83–86.

- [14] E. Kovacs, J. Ermer, P. McGregor, P. Nethercote, R. Lobrutto, G. Martin e H. Pappa. “Analytical control strategy”. Em: *Pharmacoepial Forum* 42 (jan. de 2016).
- [15] EDP. *Guia Prático da Eficiência Energética*. 2020. url: http://www.louleadapta.pt/uploads/document/13_EDP_Guia_Eficiencia_Energetica.pdf.
- [16] I. N. de Eficiência Energética. *O que é Eficiência Energética*. 2020. url: http://www.inee.org.br/eficiencia_o_que_eh.asp?Cat=eficiencia.
- [17] P. da Eficiência Energética. *Eficiência Energética*. 2010. url: <http://www.portal-eficienciaenergetica.com.pt/conceitos.html>.
- [18] F. Provost e T. Fawcett. “Data Science and Its Relationship to Big Data and Data-Driven Decision Making”. Em: *Big Data* 1 (mar. de 2013). doi: [10.1089/big.2013.1508](https://doi.org/10.1089/big.2013.1508).
- [19] K. D. Foote. *A Brief History of Data Science*. 2016.
- [20] U. D. S. TEAM. *A Modern History of Data Science*. 2017.
- [21] L. Eriksson. *Understanding the Relationship Between Data Science, Artificial Intelligence and Machine Learning*. 2020.
- [22] D. Donoho. “50 Years of Data Science”. Em: *Journal of Computational and Graphical Statistics* 26.4 (2017), pp. 745–766. doi: [10.1080/10618600.2017.1384734](https://doi.org/10.1080/10618600.2017.1384734). eprint: <https://doi.org/10.1080/10618600.2017.1384734>. url: <https://doi.org/10.1080/10618600.2017.1384734>.
- [23] Sanat. *Data Science Life cycle*”. 2019.
- [24] G. Peake. *The Life Cycle of Data Science*. 2018. url: <https://www.decisivedata.com/blog/the-lifecycle-of-data-science>.
- [25] M. Boehm, I. Antonov, M. Dokter, R. Ginhör, K. Innerebner, F. Klezin, S. N. Lindstaedt, A. Phani e B. Rath. “SystemDS: A Declarative Machine Learning System for the End-to-End Data Science Lifecycle”. Em: *CoRR abs/1909.02976* (2019). arXiv: [1909.02976](https://arxiv.org/abs/1909.02976). url: <http://arxiv.org/abs/1909.02976>.
- [26] D. Banerjee. *Data Science : Brief understanding of Typical Project Life-cycle, Tools, Techniques and skills*. 2019.
- [27] J. Carneiro, P. Saraiva, D. Martinho, G. Marreiros e P. Novais. “Representing decision-makers using styles of behavior: An approach designed for group decision support systems”. Em: *Cognitive Systems Research* 47 (2018), pp. 109–132. issn: 1389-0417. doi: <https://doi.org/10.1016/j.cogsys.2017.09.002>. url: <https://www.sciencedirect.com/science/article/pii/S1389041717300906>.

- [28] J. P. Monteiro, D. Ramos, D. Carneiro, F. Duarte, J. M. Fernandes e P. Novais. “Meta-learning and the new challenges of machine learning”. Em: *International Journal of Intelligent Systems* 36.11 (2021), pp. 6240–6272. doi: <https://doi.org/10.1002/int.22549>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/int.22549>. url: <https://onlinelibrary.wiley.com/doi/abs/10.1002/int.22549>.
- [29] X. Yao e Y. Liu. “Machine Learning”. Em: *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. Ed. por E. K. Burke e G. Kendall. Boston, MA: Springer US, 2014, pp. 477–517. isbn: 978-1-4614-6940-7. doi: [10.1007/978-1-4614-6940-7_17](https://doi.org/10.1007/978-1-4614-6940-7_17). url: https://doi.org/10.1007/978-1-4614-6940-7_17.
- [30] I. El Naqa e M. J. Murphy. “What Is Machine Learning?” Em: *Machine Learning in Radiation Oncology: Theory and Applications*. Ed. por I. El Naqa, R. Li e M. J. Murphy. Cham: Springer International Publishing, 2015, pp. 3–11. isbn: 978-3-319-18305-3. doi: [10.1007/978-3-319-18305-3_1](https://doi.org/10.1007/978-3-319-18305-3_1). url: https://doi.org/10.1007/978-3-319-18305-3_1.
- [31] C. de Computação Gráfica. *Machine learning: o que é e para que serve?* 2017. url: <https://www.ccg.pt/machine-learning-o-que-e/>.
- [32] K. Jayanthi e C. Mahesh. “A Study on machine learning methods and applications in genetics and genomics”. Em: *International Journal of Engineering and Technology(UAE)* 7 (fev. de 2018), pp. 201–204. doi: [10.14419/ijet.v7i1.7.10653](https://doi.org/10.14419/ijet.v7i1.7.10653).
- [33] B. Liu. “Supervised Learning”. Em: *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 63–132. isbn: 978-3-642-19460-3. doi: [10.1007/978-3-642-19460-3_3](https://doi.org/10.1007/978-3-642-19460-3_3). url: https://doi.org/10.1007/978-3-642-19460-3_3.
- [34] L. Castro, E. A. Wasserman e M. Lauffer. “Unsupervised learning of complex associations in an animal model”. Em: *Cognition* 173 (2018), pp. 28–33. issn: 0010-0277. doi: <https://doi.org/10.1016/j.cognition.2017.12.015>. url: <https://www.sciencedirect.com/science/article/pii/S0010027717303347>.
- [35] R. S. Sutton e A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [36] H. Patel e P. Prajapati. “Study and Analysis of Decision Tree Based Classification Algorithms”. Em: *International Journal of Computer Sciences and Engineering* 6 (out. de 2018), pp. 74–78. doi: [10.26438/ijcse/v6i10.7478](https://doi.org/10.26438/ijcse/v6i10.7478).
- [37] J. Han, M. Kamber e J. Pei. *Data Mining: Concepts and Techniques*. English (US). Elsevier Inc., jan. de 2012. isbn: 9780123814791. doi: [10.1016/C2009-0-61819-5](https://doi.org/10.1016/C2009-0-61819-5).
- [38] P. Kashyap. *Machine learning for decision makers: Cognitive computing fundamentals for better decision making*. Springer, 2017.

- [39] M. Schonlau e R. Y. Zou. “The random forest algorithm for statistical learning”. Em: *The Stata Journal* 20.1 (2020), pp. 3–29. doi: [10.1177/1536867X20909688](https://doi.org/10.1177/1536867X20909688).
- [40] U. D. Kumar. *Business analytics: The science of data-driven decision making*. Wiley, 2017.
- [41] C. Silva, J. S. Matos e M. J. Rosa. *Estratégia para o desenvolvimento da 3.ª geração do sistema de avaliação de desempenho de ETAR*. APESB, 2016. doi: DOI : [10.22181/aer.2016.0107](https://doi.org/10.22181/aer.2016.0107).
- [42] S. Lakshmana Prabu, S. Timmakondu e A. Kumar. “Wastewater treatment technologies: A review”. Em: 43 (mai. de 2011), pp. 9–13.
- [43] T. S. Frazão. *Wastewater treatment plant analytic control through expeditious analysis methods*. 2015.
- [44] Águas de Santo André. *Regulamento Geral de Recolha, Tratamento e Rejeição de Efluentes do Sistema de Santo André*. 2020. url: <http://www.adsa.pt/contents/documents/rgesa.pdf>.
- [45] M. do Ambiente. *Regulamento n.º 253/2018 de 3 de Maio do Ministério do Ambiente*. 2018. url: <https://dre.pt/application/conteudo/115214285>.
- [46] H. Marecos do Monte, M. Santos, A. Barreiros e A. Albuquerque. *TRATAMENTO DE ÁGUAS RESIDUAIS - Operações e Processos de Tratamento Físico e Químico*. Dez. de 2016. isbn: 978-989-8360-32-8.
- [47] d. M. d. A. e. d. O. d. T. Inspeção-Geral da Agricultura. “20 Anos de Inspeção Ambiental em Portugal”. Em: *Républica Portuguesa* (2017).
- [48] M. do Ambiente. *Decreto-Lei n.º 152/97 de 19 de Junho do Ministério do Ambiente*. 1997. url: <https://data.dre.pt/eli/dec-lei/152/1997/06/19/p/dre/pt/html>.
- [49] J. Irving e K. Bilyk. *So Much Data, So Little Information: Advances in Wastewater Data Analytics and Visualization*. 2020. url: <https://www.hazenandsawyer.com/publications/so-much-data-so-little-information-advances-in-wastewater-data-analytics-an/>.
- [50] Opseyes. *How Wastewater Treatment Plants Are Embracing the Digital Age*. 2020. url: <https://opseyes.com/how-wastewater-treatment-plants-are-embracing-the-digital-age/>.
- [51] S. Eerikäinen, H. Haimi, A. Mikola e R. Vahala. “Data analytics in control and operation of municipal wastewater treatment plants: qualitative analysis of needs and barriers”. Em: *Water Science and Technology* 82 (jul. de 2020). doi: [10.2166/wst.2020.311](https://doi.org/10.2166/wst.2020.311).
- [52] M. do Ambiente. *Decreto-Lei n.º 236/98 de 1 de Agosto do Ministério do Ambiente*. 1998. url: <https://data.dre.pt/eli/dec-lei/236/1998/08/01/p/dre/pt/html>.

- [53] A. Santos, F. Oliveira, G. Soares, J. Carneiro, L. Lima, M. Viegas, P. Corigo, P. C. Paulo Nogueira e P. Quaresma. *Uso Eficiente de Energia nos Serviços de Águas*. ERSAR, 2018.
- [54] U. E. P. AGENCY. *Energy Efficiency in Water and Wastewater Facilities*. EPA, 2013.
- [55] A. Guerrini, G. Romano e A. Indipendenza. *Energy Efficiency Drivers in Wastewater Treatment Plants: A Double Bootstrap DEA Analysis*. Sustainability, 2017.
- [56] P. Reis. *A importância e as fases de uma auditoria energética*. 2012.
- [57] P. Foladori, M. Vaccari e F. Vitali. “Energy audit in small wastewater treatment plants: Methodology, energy consumption indicators, and lessons learned”. Em: *Water science and technology : a journal of the International Association on Water Pollution Research* 72 (set. de 2015), pp. 1007–1015. doi: [10.2166/wst.2015.306](https://doi.org/10.2166/wst.2015.306).
- [58] A. E. do Ambiente. *Energia*. 2019. url: <https://www.eea.europa.eu/pt/themes/energy/intro>.
- [59] E. Guillard. *How Big Data Analytics builds Energy Efficiency*. 2020.
- [60] B. S. Pimentel. *O poder da análise de dados na gestão de energia*. 2017. url: <https://viridis.energy/pt/blog/o-poder-da-analise-de-dados-na-gestao-de-energia>.
- [61] Tenbu. *Eficiência Energética: IOT E Modelos de Machine Learning*. 2020.
- [62] M. Sperling. “Wastewater Characteristics, Treatment and Disposal”. Em: *Water Intelligence Online* 6 (jul. de 2007). doi: [10.2166/9781780402086](https://doi.org/10.2166/9781780402086).
- [63] D. De Haas, J. Foley, B Marshall, M. Dancy, S Vierboom e J Bartle-Smith. “Benchmarking Wastewater Treatment Plant Energy Use in Australia”. Em: mai. de 2015.
- [64] S. Long e E. Cudney. “Strategic Process Integration of Energy and Environmental Systems in Wastewater Treatment Plants”. Em: *International Journal of Energy and Environment* 3 (jan. de 2012), pp. 521–530.
- [65] K. B. Newhart, R. W. Holloway, A. S. Hering e T. Y. Cath. “Data-driven performance analyses of wastewater treatment plants: A review”. Em: *Water Research* 157 (2019), pp. 498 –513. issn: 0043-1354. doi: <https://doi.org/10.1016/j.watres.2019.03.030>. url: <http://www.sciencedirect.com/science/article/pii/S0043135419302490>.
- [66] A. M. Gonçalves, M. Amorim e M. Costa. “UM CONTRIBUTO DA ANÁLISE ESTATÍSTICA NA GESTÃO DE UMA ESTAÇÃO DE TRATAMENTO DE ÁGUAS RESIDUAIS (ETAR)”. Em: jun. de 2018.
- [67] R. Ribeiro, H. Pinheiro, C. Pinheiro e M. Almeida. “Metodologia DEMOCON - Sistema expedito para supervisão e controlo avançados de ETAR descentralizadas”. Em: (abr. de 2017). doi: [10.13140/RG.2.2.19817.90722](https://doi.org/10.13140/RG.2.2.19817.90722).
- [68] Z. Li, Z. Zou e L. Wang. “Analysis and Forecasting of the Energy Consumption in Wastewater Treatment Plant”. Em: *Hindawi* (2019). url: <https://doi.org/10.1155/2019/8690898>.

- [69] S. Longo, A. Hospido, J. Lema e M. Mauricio-Iglesias. “A systematic methodology for the robust quantification of energy efficiency at wastewater treatment plants featuring Data Envelopment Analysis”. Em: *Water Research* 141 (2018), pp. 317 –328. issn: 0043-1354. doi: <https://doi.org/10.1016/j.watres.2018.04.067>. url: <http://www.sciencedirect.com/science/article/pii/S0043135418303610>.
- [70] S. Longo, M. Mauricio-Iglesias, A. Soares, P. Campo, F. Fatone, A. Eusebi, E. Akkersdijk, L. Stefani e A. Hospido. “ENERWATER – A standard method for assessing and improving the energy efficiency of wastewater treatment plants”. Em: *Applied Energy* 242 (2019), pp. 897 –910. issn: 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2019.03.130>. url: <http://www.sciencedirect.com/science/article/pii/S0306261919305367>.
- [71] F. Hernández-Sancho, M. Molinos-Senante e R. Sala-Garrido. “Energy efficiency in Spanish wastewater treatment plants: A non-radial DEA approach”. Em: *Science of The Total Environment* 409.14 (2011), pp. 2693 –2699. issn: 0048-9697. doi: <https://doi.org/10.1016/j.scitotenv.2011.04.018>. url: <http://www.sciencedirect.com/science/article/pii/S0048969711003755>.
- [72] P. Oliveira, B. Fernandes, C. Analide e P. Novais. “Forecasting Energy Consumption of Wastewater Treatment Plants with a Transfer Learning Approach for Sustainable Cities”. Em: *Electronics* 10.10 (2021). issn: 2079-9292. doi: [10.3390/electronics10101149](https://doi.org/10.3390/electronics10101149). url: <https://www.mdpi.com/2079-9292/10/10/1149>.
- [73] LAWA. *What is Nitrogen*. 2021. url: <https://www.lawa.org.nz/learn/factsheets/nitrogen/>.
- [74] A. Wood, M. Blackhurst, T. Hawkins, X. Xue, N. Ashbolt e J. Garland. “Cost-effectiveness of nitrogen mitigation by alternative household wastewater management technologies”. Em: *Journal of Environmental Management* 150 (2015), pp. 344–354. issn: 0301-4797. doi: <https://doi.org/10.1016/j.jenvman.2014.10.002>. url: <https://www.sciencedirect.com/science/article/pii/S0301479714004927>.
- [75] C. Faria, P. Oliveira, B. Fernandes, F. Aguiar, M. A. Pereira e P. Novais. “A Tree-Based Approach to Forecast the Total Nitrogen in Wastewater Treatment Plants”. Em: *Distributed Computing and Artificial Intelligence, Volume 1: 18th International Conference*. Ed. por K. Matsui, S. Omatu, T. Yigitcanlar e S. R. González. Cham: Springer International Publishing, 2022, pp. 137–147.

Dados Electricidade

I.1 Vila Real

Nesta secção estão ilustrados diversos gráficos relativos aos dados da electricidade na ETAR de Vila Real. Nestes gráficos está exposta mais informação relevante sobre estes dados, como por exemplo, a distribuição dos consumos totais da electricidade pelos diferentes indicadores.

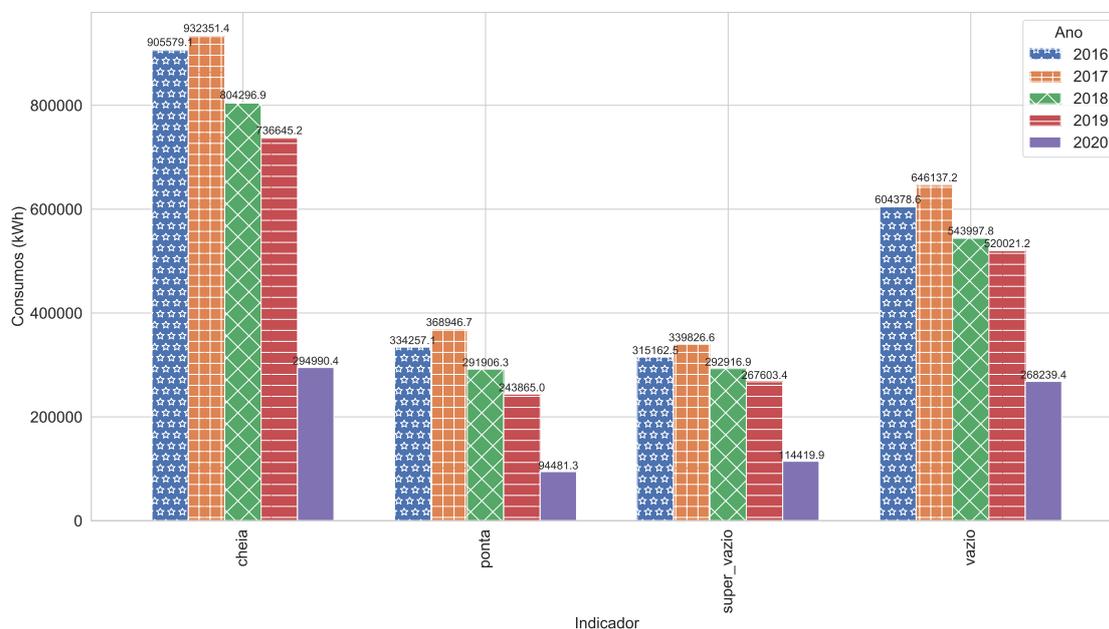


Figura I.1: Soma dos consumos de electricidade por ano em cada indicador.

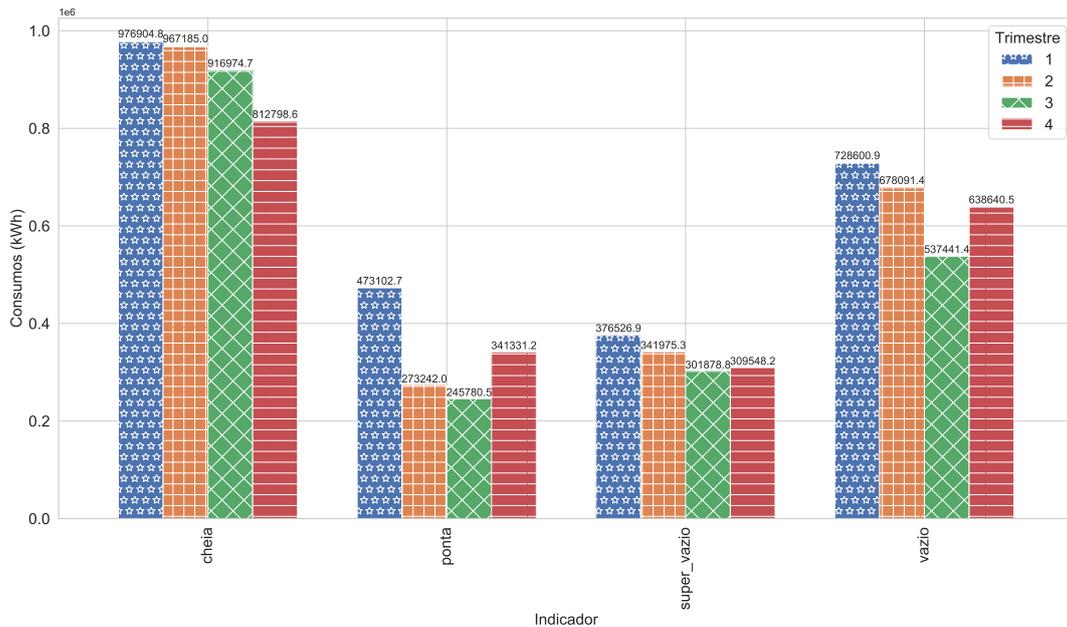


Figura I.2: Soma dos consumos de electricidade por trimestre em cada indicador.

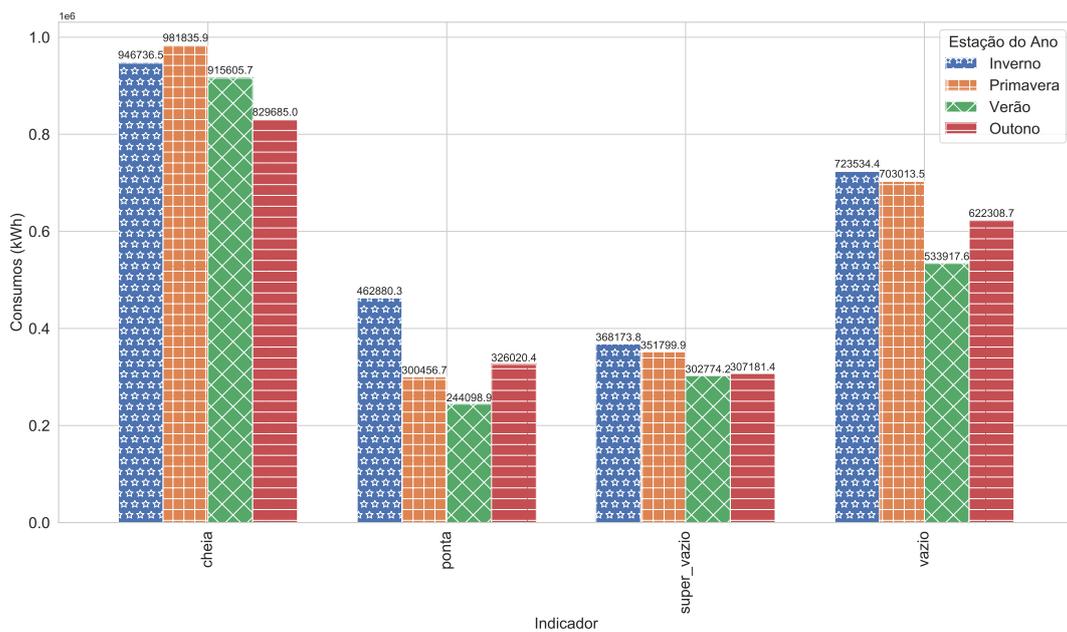


Figura I.3: Soma dos consumos de electricidade por estação do ano em cada indicador.

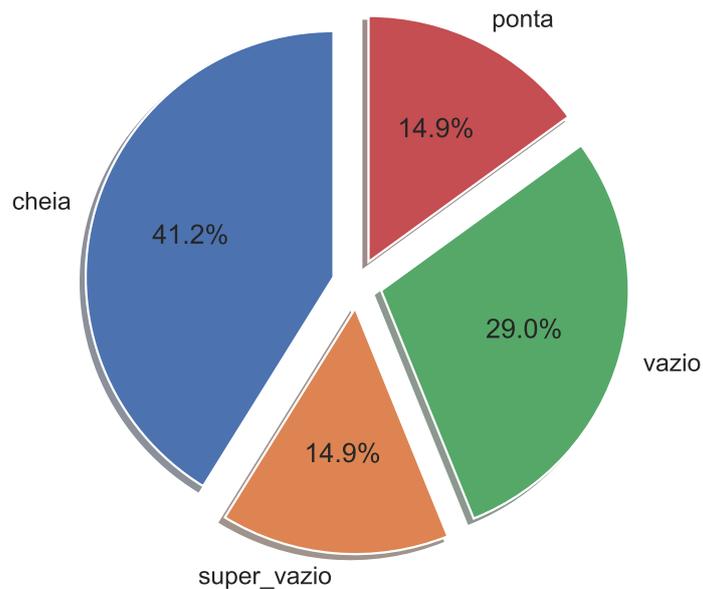


Figura I.4: Distribuição dos consumos totais da electricidade por indicador.

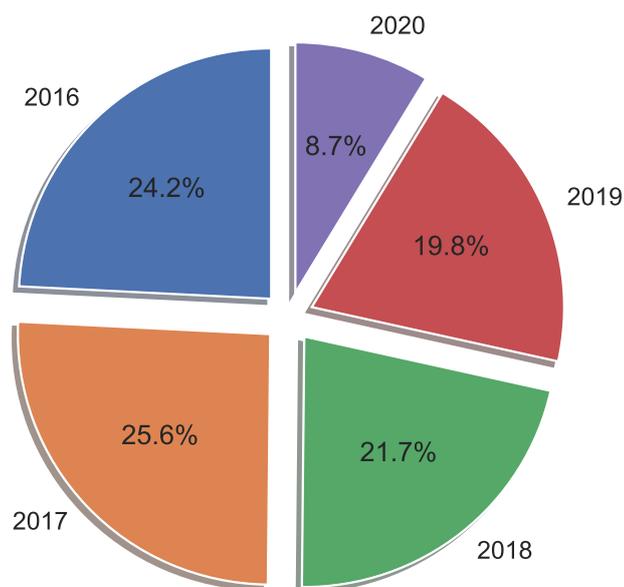


Figura I.5: Distribuição dos consumos totais da electricidade por anos.

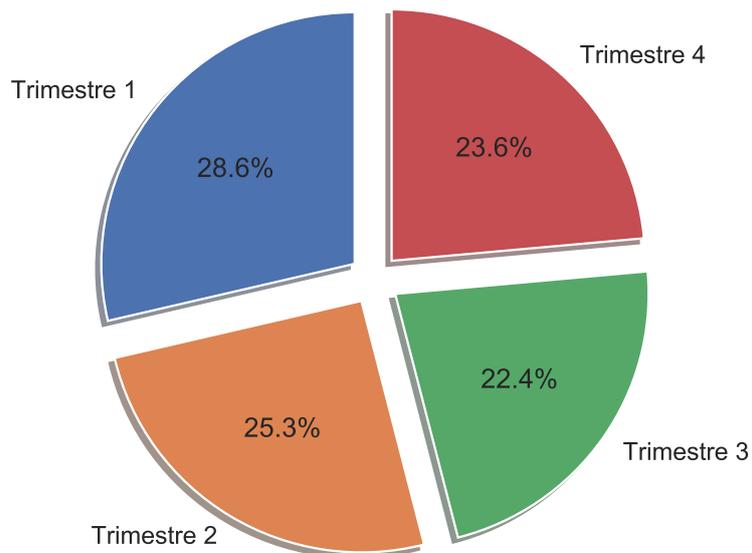


Figura I.6: Distribuição dos consumos totais da electricidade por trimestre.

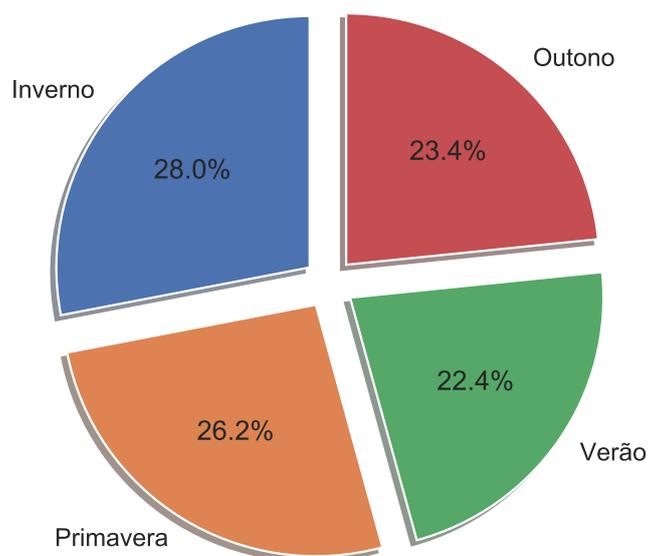


Figura I.7: Distribuição dos consumos totais da electricidade por estação do ano.

I.2 Guimarães

Já nesta secção é exibido múltiplos gráficos em relação aos dados da electricidade na [ETAR](#) de Guimarães. Neles são ilustradas mais informações sobre estes mesmos dados, como por exemplo, a soma dos consumos de electricidade por ano em cada indicador.

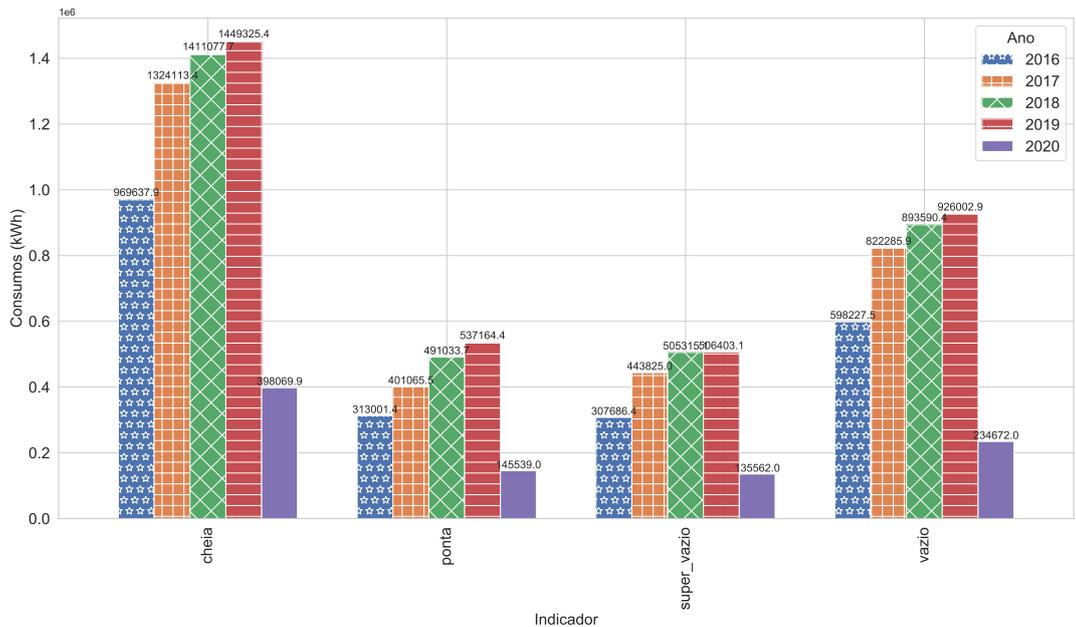


Figura I.8: Soma dos consumos de electricidade por ano em cada indicador.

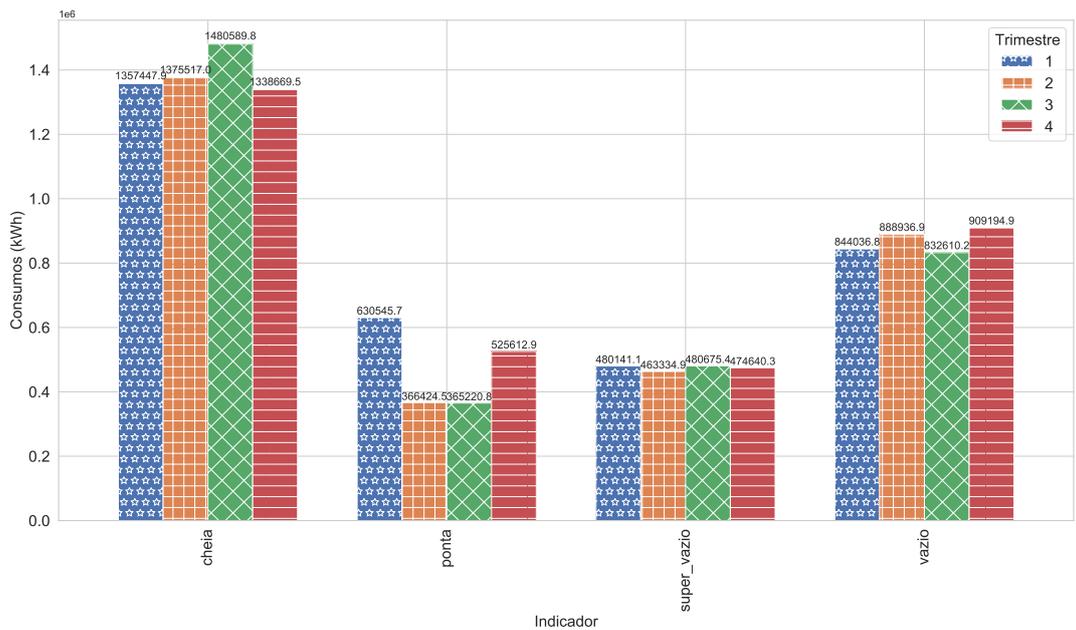


Figura I.9: Soma dos consumos de electricidade por trimestre em cada indicador.

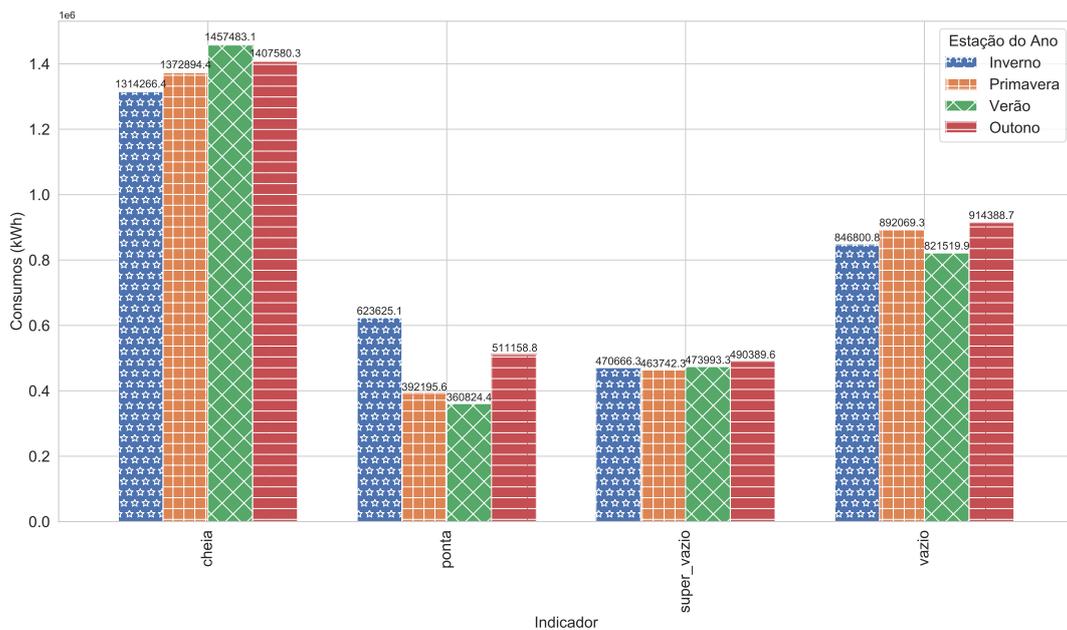


Figura I.10: Soma dos consumos de electricidade por estação do ano em cada indicador.

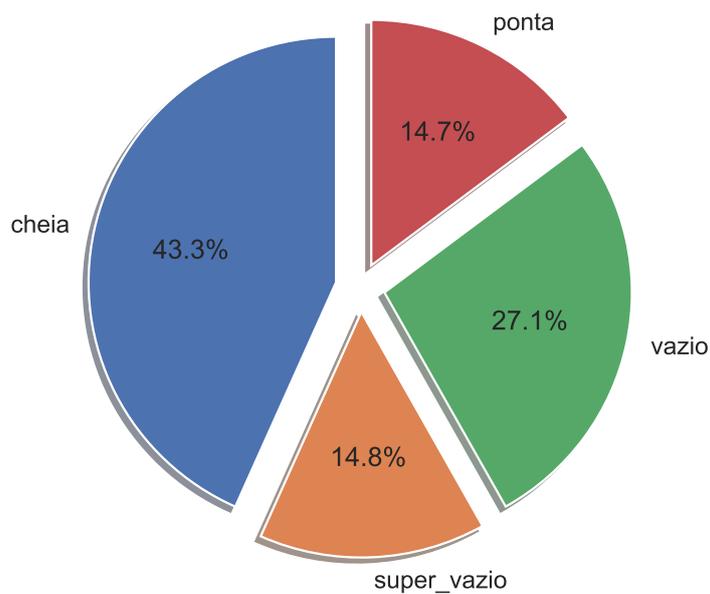


Figura I.11: Distribuição dos consumos totais da electricidade por indicador.

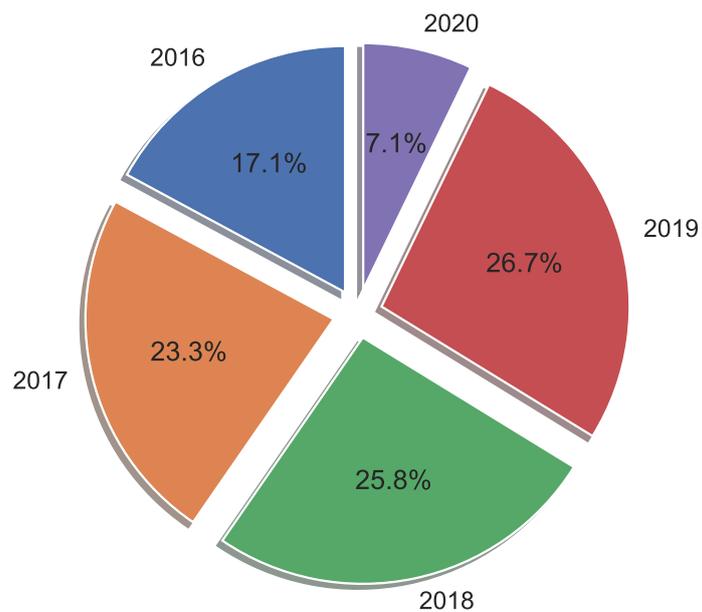


Figura I.12: Distribuição dos consumos totais da electricidade por ano.

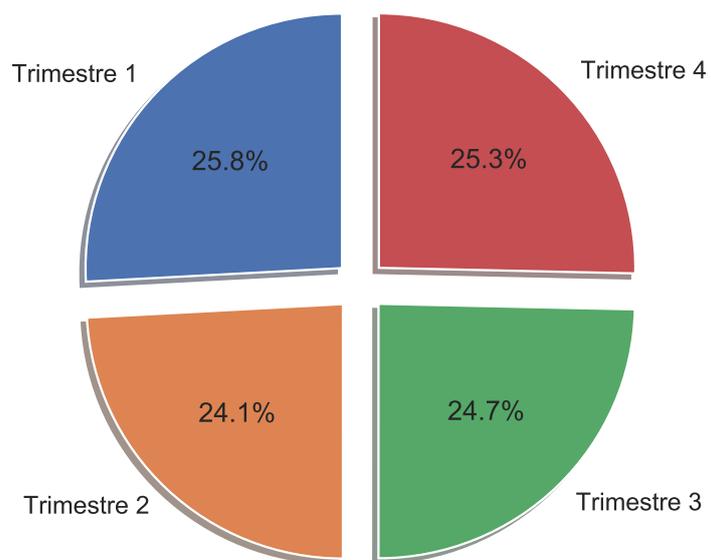


Figura I.13: Distribuição dos consumos totais da electricidade por trimestre.

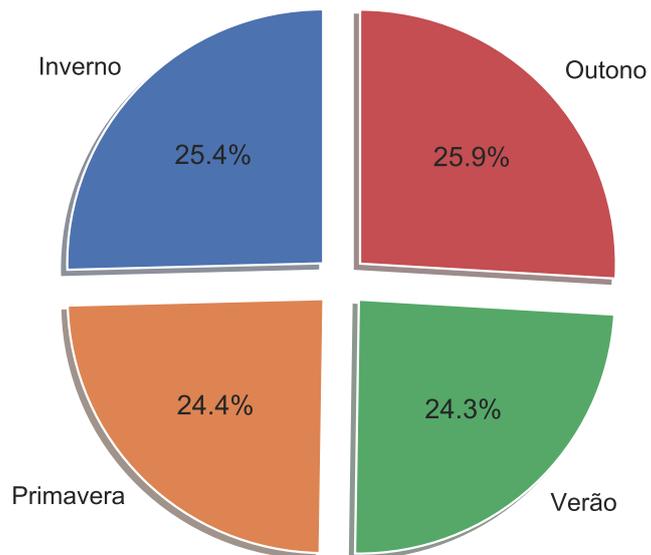
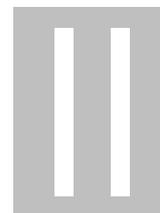


Figura I.14: Distribuição dos consumos totais da electricidade por estação do ano.



Dados Controlo Analítico

Neste capítulo estão ilustrados diversos gráficos relativos aos dados do controlo analítico da ETAR de Vila Real. É apresentada mais informação sobre a variação média em cada mês do ano de outros indicadores, nos seus respetivos setores.

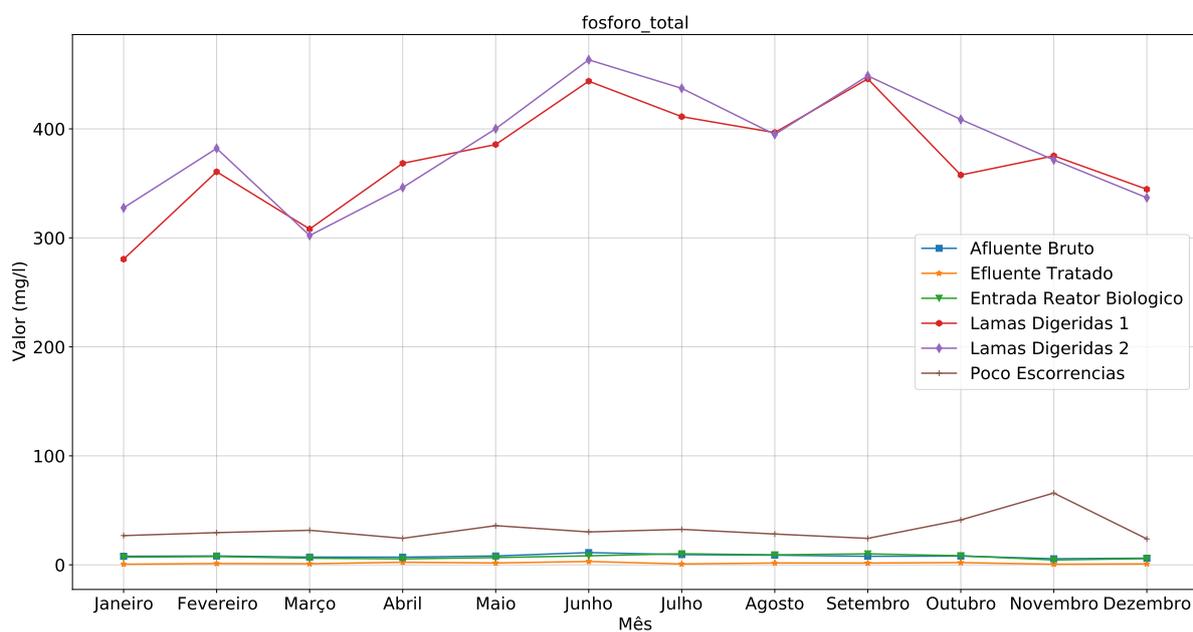


Figura II.1: Variação Média do indicador Fósforo Total em cada mês do ano.

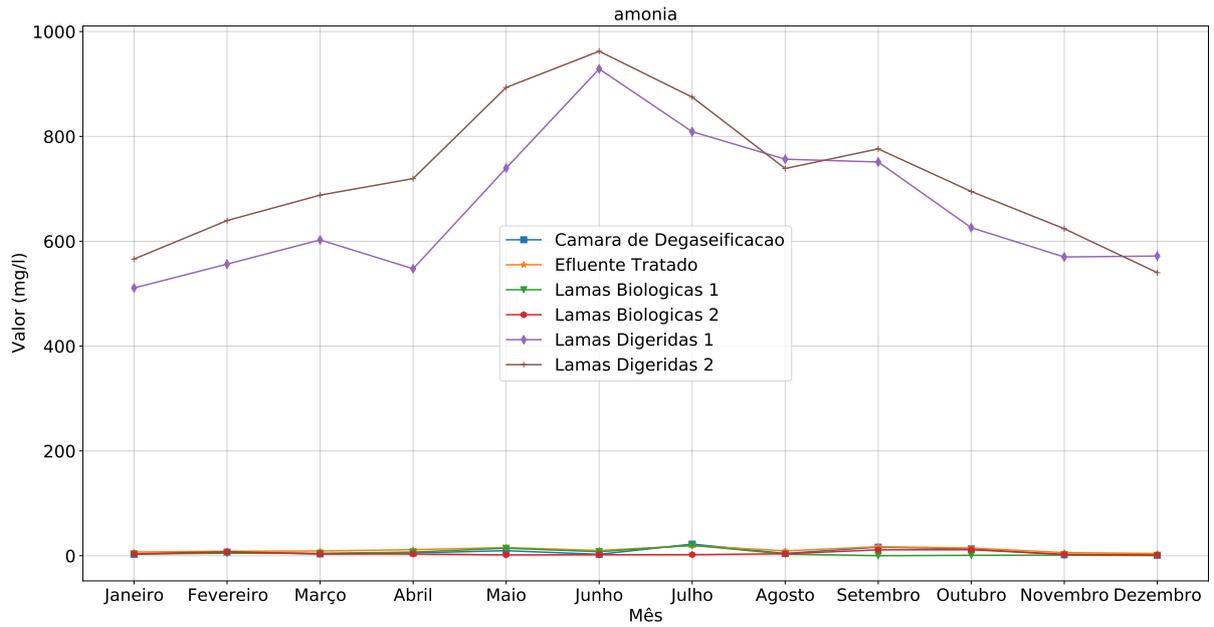


Figura II.2: Variação Média do indicador Amónia em cada mês do ano.

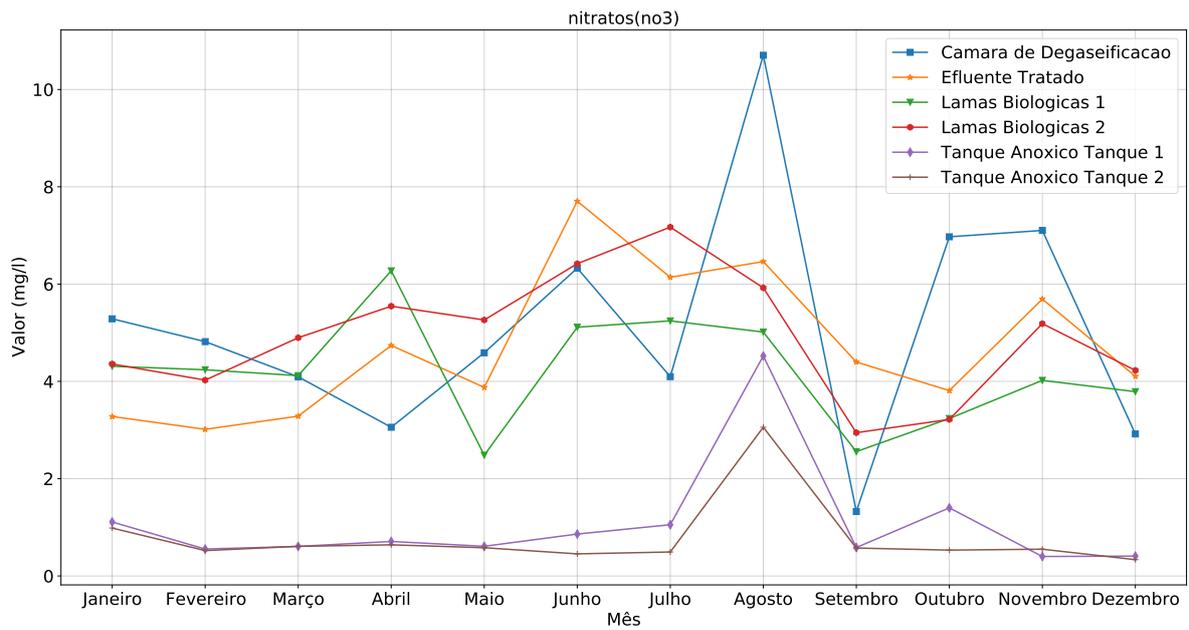


Figura II.3: Variação Média do indicador Nitratos em cada mês do ano.

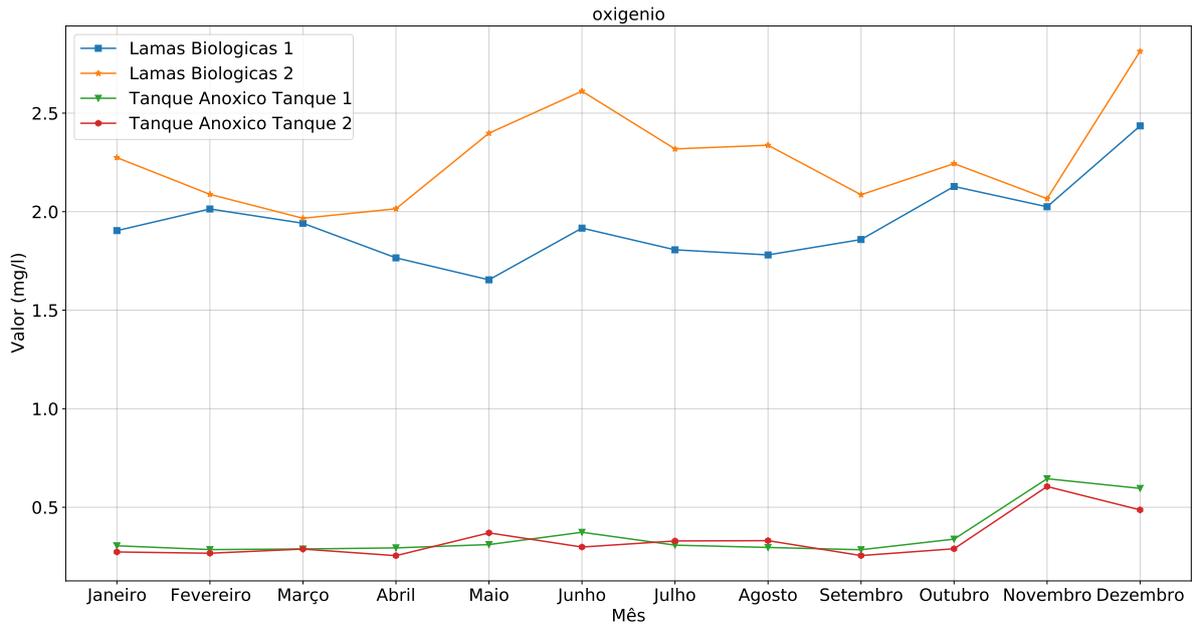


Figura II.4: Variação Média do indicador Oxigénio em cada mês do ano.

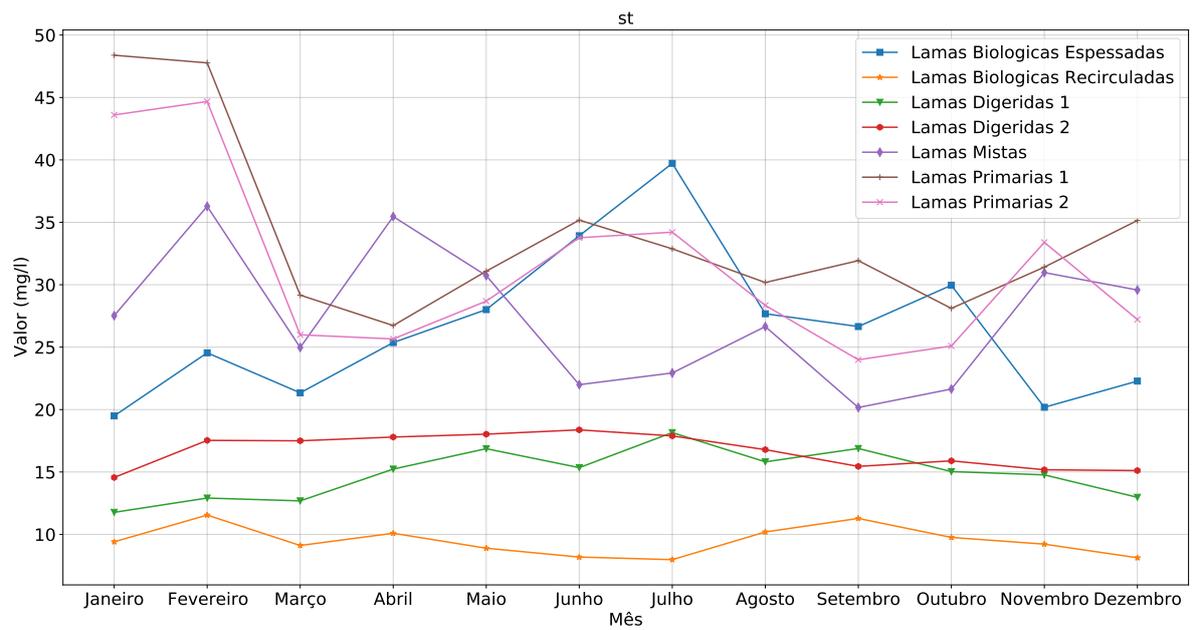


Figura II.5: Variação Média do indicador ST em cada mês do ano.

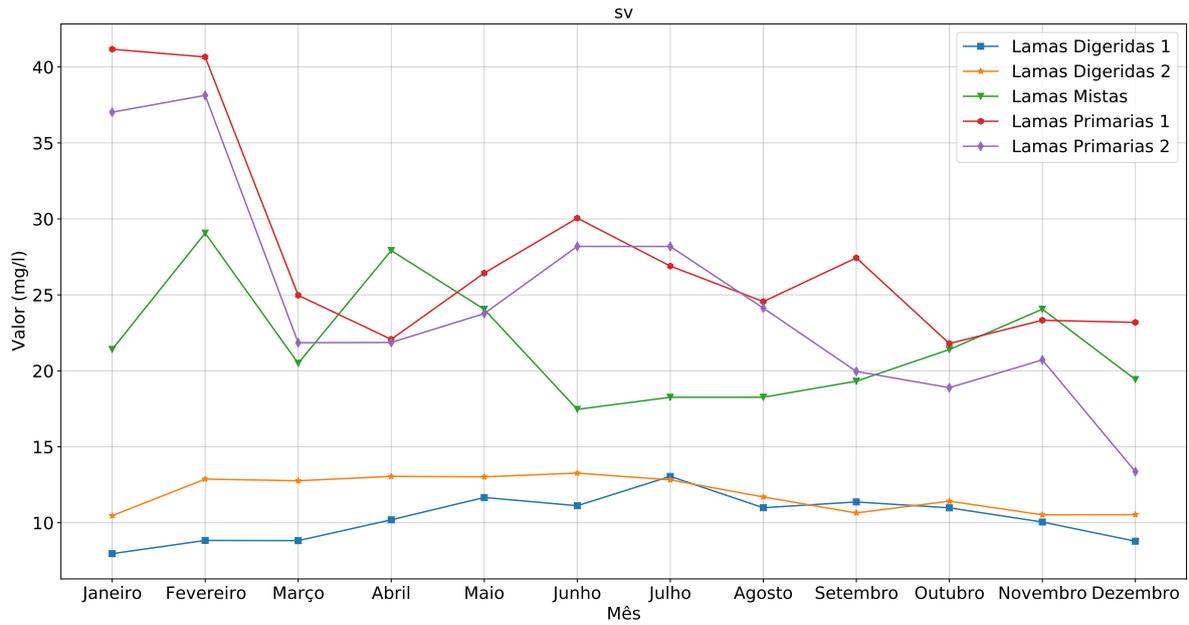


Figura II.6: Variação Média do indicador SV em cada mês do ano.

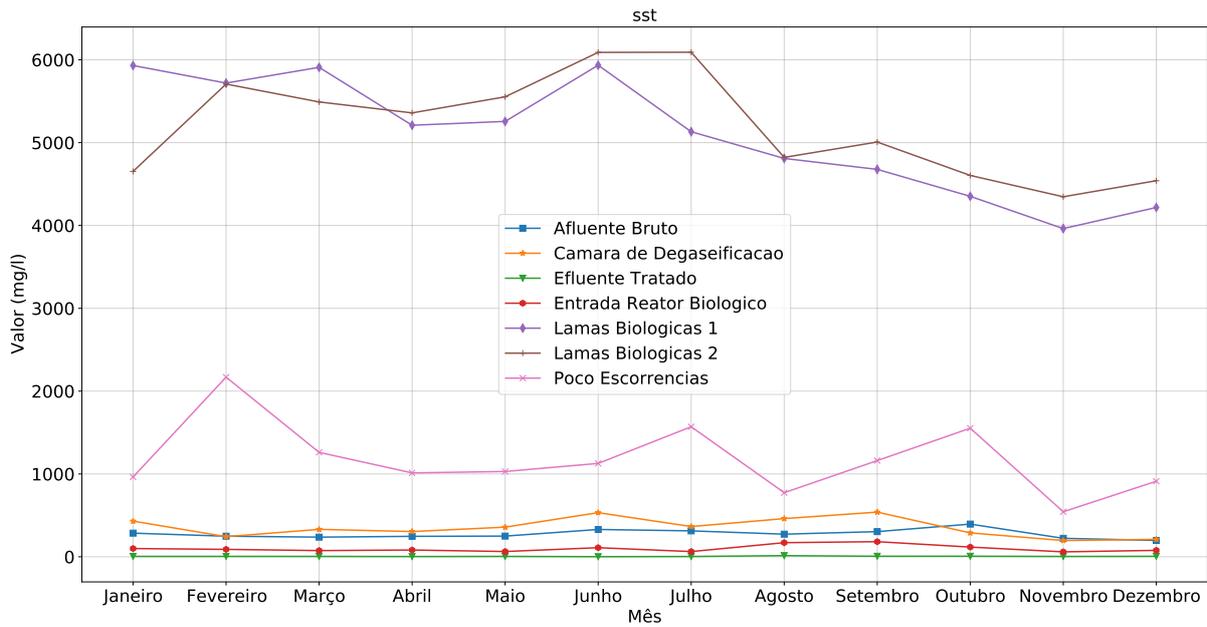


Figura II.7: Variação Média do indicador SST em cada mês do ano.

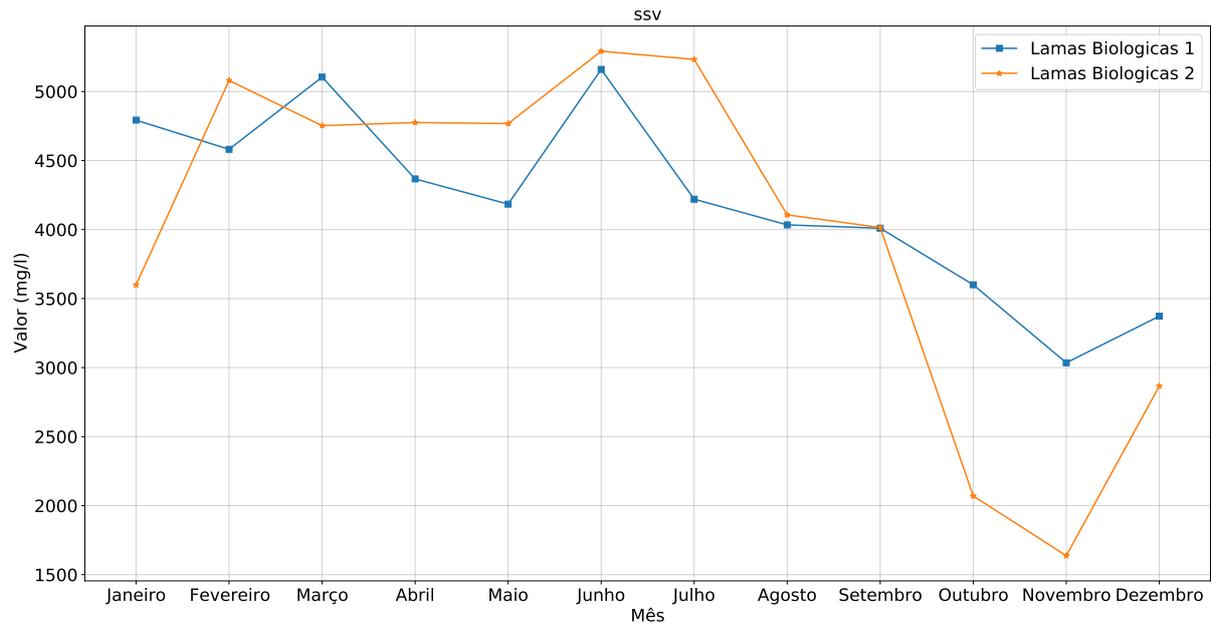


Figura II.8: Variação Média do indicador SSV em cada mês do ano.