INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Deteção de anomalias em dados de gestão de resíduos em Portugal

João Sérgio Vieira Zolá Ribeiro

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:
Doutor Luís Nunes, Professor Associado
Iscte - Instituto Universitário de Lisboa

Co-Orientador:
Doutor Sancho Moura Oliveira, Professor Associado
Iscte - Instituto Universitário de Lisboa

Setembro, 2024

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Deteção de anomalias em dados de gestão de resíduos em Portugal

João Sérgio Vieira Zolá Ribeiro

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:
Doutor Luís Nunes, Professor Associado
Iscte - Instituto Universitário de Lisboa

Co-Orientador:
Doutor Sancho Moura Oliveira, Professor Associado
Iscte - Instituto Universitário de Lisboa

Setembro, 2024

Aos meus avós.

Agradecimentos

Aos meus orientadores, Professor Doutor Luís Nunes e Professor Doutor Sancho Oliveira, o meu mais sincero obrigado por todo o acompanhamento e apoio prestados. Desde o momento em que confiaram em mim para embarcar neste tema tão desafiante, até ao último instante de entrega, estiveram sempre por perto, ajudando-me a superar cada obstáculo.

À minha família, especialmente aos meus pais, deixo um agradecimento especial por todo o apoio incondicional ao longo do meu percurso académico. A vossa presença constante, especialmente durante a realização desta dissertação, foi essencial para me manter focado e motivado, nunca permitindo que desistisse dos meus objetivos.

Aos meus amigos, agradeço por toda a motivação e apoio ao longo de todo este processo. Um agradecimento especial aos amigos que me acompanham desde a licenciatura e que partilharam comigo esta jornada universitária. Foi um privilégio poder contar com a vossa amizade, partilhar objetivos comuns, enfrentarmos desafios juntos e colecionar memórias que levo para a vida.

Gostaria também de agradecer à APA pela disponibilização dos dados que tornaram possível o desenvolvimento deste trabalho. Espero ter conseguido contribuir de forma positiva para uma instituição tão relevante em Portugal. Sem o apoio desta entidade, a concretização deste estudo não teria sido possível.

Por fim, um agradecimento à NTT DATA por me permitir conciliar a realização da dissertação de mestrado que coincidiu com o início da minha jornada no mundo do trabalho. Este ano foi, sem dúvida, repleto de desafios, mas a oportunidade de conseguir equilibrar estas duas vertentes foi essencial para o meu crescimento pessoal e profissional.

A realização desta dissertação foi parcialmente financiada por fundos nacionais através da FCT - Fundação para a Ciência e Tecnologia, I.P. no âmbito dos projetos UIDB/04466/2020 e UIDP/04466/2020.

Resumo

Esta dissertação tem como principal objetivo explorar e detetar anomalias na gestão de resíduos em Portugal, no período de 2019 a 2022, utilizando os dados disponibilizados pela APA (APA). A investigação centrou-se em três questões principais: a viabilidade de identificar anomalias no transporte e gestão de resíduos através dos dados da APA, a caracterização dos produtores e transformadores, e a relação entre os Códigos de Atividade Económica (CAE) e os tipos de resíduos gerados e transformados.

Através desta investigação, foi possível detetar discrepâncias significativas entre as quantidades de resíduos enviadas pelos produtores e as recebidas pelos transformadores. Ao longo dos quatro anos analisados, foram identificadas anomalias em 16,3% dos registos de transporte, considerados anulados.

Foram realizadas duas experiências para a caracterização dos produtores e transformadores através do *clustering*. A primeira revelou que os produtores de resíduos de misturas estão muitas vezes associados a grandes volumes de resíduos e a mercadorias perigosas, enquanto entre os transformadores, os resultados sugerem que as mercadorias perigosas estão associadas a grandes quantidades de resíduos. A segunda experiência mostrou que as embalagens com resíduos ou contaminadas estão distribuídas por vários setores económicos, ao passo que os resíduos de misturas e compósitos estão concentrados essencialmente no setor energético. Entre os transformadores, observou-se uma maior especialização, com cada grupo focado num setor económico específico.

As conclusões oferecem uma base para o desenvolvimento de trabalhos mais profundos que se espera que possam culminar em estratégias de melhoria na gestão de resíduos em Portugal.

Palavras-chave: Gestão de Resíduos; APA; Transporte de Resíduos; Produtores de Resíduos; Transformadores de Resíduos; Clustering.

Abstract

This dissertation aims to explore and detect anomalies in waste management in Portugal from 2019 to 2022, using data provided by the Portuguese Environment Agency (APA). The research focused on three main questions: the feasibility of identifying anomalies in waste transportation and management through APA data, the characterization of waste producers and processors, and the relationship between Economic Activity Codes (CAE) and the types of waste generated and processed.

Through this investigation, significant discrepancies were identified between the amounts of waste sent by producers and those received by processors. Over the four years analyzed, anomalies were found in 16.3% of transport records, which were considered canceled.

Two experiments were conducted to characterize producers and processors through clustering. The first experiment revealed that producers of mixed waste are often associated with large volumes of waste and hazardous goods, while among processors, the results suggest that hazardous goods are linked to large quantities of waste. The second experiment showed that packaging waste or contaminated packaging waste is distributed across various economic sectors, while mixed and composite waste is mainly concentrated in the energy sector. Among processors, a higher level of specialization was observed, with each group focused on a specific economic sector.

The conclusions provide a foundation for further research, which is expected to culminate in strategies for improving waste management in Portugal.

Keywords: Waste Management; APA; Waste Transportation; Waste Producers; Waste Processors; Clustering.

Índice

Agradecimentos iii

Resumo v

Abstract vii

CAPÍTULO 1 1

Introdução 1

1.1. Contextualização 1

1.2. Motivação 2

1.3. Questões de Investigação 3

1.4. Objetivos 4

1.5. Metodologia 4

CAPÍTULO 2 7

Revisão da Literatura 7

2.1. Processo de Revisão de Literatura 7

2.2. Gestão de Resíduos no Estrangeiro 9

2.3. Gestão de Resíduos em Portugal 11

2.4. Inteligência Artificial na Gestão de Resíduos 14

2.5. Resumo da Revisão de Literatura 16

2.6. Abordagem diferenciadora 17

CAPÍTULO 3 19

Análise de Dados 19

3.1. Business Understanding 19

3.2. Data Understanding 19

3.2.1. Descrição de variáveis de diferentes fontes de informação 19

3.2.2. Conceitos fundamentais 21

3.3. Data Preparation 22

3.3.1. Dataset Inicial 22

3.3.2. Tratamento e Junção de variáveis dos formulários 24

3.3.3. Preparação dos Dados para a Criação de Modelos 25

3.3.4. Junção dos dados referentes aos plásticos 27

3.3.5. Entrevistas a Produtores e Transformadores 27

CAPÍTULO 4 29

Modelação e Resultados 29
4.1. Análises Descritivas 29
4.1.1. Caracterização dos dados iniciais 29
..... 29
4.1.2. Análise Temporal 36
4.1.3. Análise por distrito 39
4.1.4. Análise da presença de plásticos 40
4.1.5. Código de Atividade Económica (CAE) 41
4.2. Clustering 44
4.3. Resultados 46
4.3.1. Experiência 1: Clustering excluindo CAEs 46
4.3.2. Experiência 2: Clustering incluindo CAEs 51
CAPÍTULO 5 59
Conclusões e Trabalho Futuro 59
5.1. Conclusões 59
5.2. Limitações 61
5.3. Trabalho Futuro 61
Referências Bibliográficas 63
Apêndices 67

Índice de Figuras

Figura 1 – CRISP - DM [11] 5
Figura 2 - Metodologia PRISMA, Adaptado de [12] 8
Figura 3 - Economia circular na Gestão de Resíduos em Portugal, adaptado de [25] 12
Figura 4 - Boxplots para a remoção de outliers 26
Figura 5 - Número de registos no dataset da variável ler_codigo (Tipo de Resíduo) 29
Figura 6 - Top 5 das operações definidas pelo Produtor 32
Figura 7- Top 5 das operações definidas pelo Transformador 32
Figura 8 - Tipo de Resíduo por quantidade enviada e recebida, respetivamente entre 2019-2022 (Toneladas) - Concluídas 32
Figura 9 – Boxplots das diferenças de quantidades enviadas e recebidas por tipo de resíduo - Concluídas 33
Figura 10 - Boxplots das diferenças de quantidades enviadas e recebidas por tipo de resíduo - Corrigidas 34
Figura 11 - Tempo médio de transporte para cada tipo de resíduo 35
Figura 12 - Distribuição do tempo de transporte por tipo de resíduo 35
Figura 13 - Quantidade enviada pelos produtores entre 2019 e 2022 (Toneladas) 36
Figura 14 - Quantidade recebida pelos transformadores entre 2019 e 2022 37
Figura 15 - Quantidade enviada por mês pelos produtores entre 2019 e 2022 38
Figura 16 - Quantidade recebida por mês pelos transformadores entre 2019 e 2022 39
Figura 17 - Mapa de calor com as quantidades enviadas pelos produtores nos distritos portugueses (2019-2022) 40
Figura 18 - Mapa de calor com a classificação do tipo de resíduos entre 2019-2022 41
Figura 19 - Número Ideal de clusters para produtores sem CAE 46
Figura 20 - Mapa de Calor da média das variáveis dos produtores por cluster exlcuindo o CAE 47
Figura 21 - Número Ideal de clusters para transformadores sem CAE 49
Figura 22 - Mapa de Calor da média das variáveis dos transformadores por cluster sem o CAE 50
Figura 23 - Número Ideal de clusters para Produtores com CAE 52
Figura 24 - Mapa de Calor da média das variáveis dos Produtores por cluster incluído o CAE 54
Figura 25 - Número Ideal de clusters para Transformadores com CAE 55
Figura 26 - Mapa de Calor da média das variáveis dos Transformadores por cluster com o CAE 56

Índice de Tabelas

Tabela 1 - Artigos principais por secção e fonte utilizados através do Prisma 17
Tabela 2 – Seleção das variáveis relevantes..... 20
Tabela 3 - Variáveis finais para análise..... 22
Tabela 4 - Números Omissos por variável..... 23
Tabela 5 - Caracterização da variável Origem 30
Tabela 6 - Relação entre o Tipo de Resíduo e o Estado do transporte 30
Tabela 7 - Tipo de resíduos face à perigosidade 31
Tabela 8 - Top de Operações realizadas pelos Produtores e Transformadores..... 31
Tabela 9 - Diferença entre a quantidade enviada e a quantidade recebida (em toneladas) - Concluídas 33
Tabela 10 - Crescimento (em %) das quantidades enviadas face ao ano anterior 36
Tabela 11 - Crescimento (em %) das quantidades recebidas face ao ano anterior..... 37
Tabela 12 - Quantidades enviadas entre 2019 e 2022 por Distrito 40
Tabela 13 - Top 5 de ocorrências de CAEs de produtores por tipo de resíduo..... 42
Tabela 14 - Top 5 de CAEs por produtores e transformadores..... 43
Tabela 15 - CAEs associados a mercadorias perigosas..... 43
Tabela 16 - Variáveis utilizadas no Clustering (Experiência 1) 44
Tabela 17 - Variáveis utilizadas no Clustering (Experiência 2) 45
Tabela 18 - Número de Produtores sem CAE incluído por cluster..... 47
Tabela 19 -Top 3 de CAEs associados aos clusters da 1ª Experiência (Produtores)..... 48
Tabela 20 - Número de transformadores sem CAE incluído por cluster..... 49
Tabela 21- Top 3 de CAEs associados aos clusters da 1ª Experiência (Transformadores)..... 51
Tabela 22 - Número de Produtores com CAE incluído por cluster..... 53
Tabela 23 - Número de Transformadores com CAE incluído por cluster..... 56

CAPÍTULO 1

Introdução

1.1. Contextualização

A crescente preocupação ambiental tem moldado a sociedade de um modo geral, destacando-se a gestão de resíduos como um dos desafios mais urgentes da sociedade atual em Portugal. O país enfrenta vários desafios, marcados por uma evolução notável na urbanização e nos padrões de consumo da população. A complexidade dessa transformação impõe barreiras cada vez maiores à gestão de resíduos, tornando-se crucial analisar aprofundadamente como é que estas podem ser ultrapassadas de forma inovadora e eficiente.

Neste contexto, a má gestão de resíduos conduz não só à degradação ambiental, mas também pode vir a afetar a qualidade de vida dos portugueses. Questões de saúde relacionadas com a contaminação do solo e da água, juntamente com a degradação de espaços públicos, destacam a urgência de uma abordagem mais eficaz neste tema.

Esta dissertação foca-se na deteção de anomalias no transporte de resíduos e na caracterização de produtores e transformadores em Portugal, utilizando dados disponibilizados pela Agência Portuguesa do Ambiente (APA). Num contexto em que a reciclagem é um procedimento ainda muito pouco adotado para o que era expectável, é crucial abordar este problema, uma vez que o país se encontra em risco de não cumprir as metas estabelecidas para a União Europeia [1], evidenciando uma necessidade cada vez maior na reformulação dos modelos económicos e operacionais.

Em Portugal, a APA, responsável pela gestão de resíduos estabelece determinadas obrigações para os produtores, numa ótica de possuir uma gestão adequada destes, com o objetivo de garantir uma utilização sustentável dos recursos naturais, reduzindo a pressão nos ecossistemas e promovendo uma maior saúde ambiental e humana.

Alinhado com a economia circular, esta gestão reduz a dependência de recursos importados, impulsionando oportunidades económicas e contribuindo para uma maior competitividade a longo prazo. A responsabilidade pela gestão recai no produtor inicial, podendo ser alargada por lei para outras entidades, dependendo da situação. A APA e as Comissões de Coordenação e Desenvolvimento Regional regulam e licenciam todas as operações, garantindo proteção ambiental e saúde pública [2].

Adicionalmente, o Regime Geral de Gestão de Resíduos (RGGR), em Portugal, estabelece definições e responsabilidades cruciais no âmbito da recolha e do tratamento de resíduos. Esta recolha inclui uma triagem e armazenamento dos resíduos para que estes possam ser transportados e instalados, designado por recolha seletiva. Desta forma, estes permanecem separados por tipo e por natureza facilitando todo o processo de tratamento [3].

Tanto a recolha como o tratamento têm obrigações específicas para cada uma destas atividades. O tratamento deve ser realizado por operadores autorizados, enquanto que, a recolha não requer qualquer tipo de licenciamento. No entanto, as pessoas responsáveis por estas tarefas devem estar registadas, e declarar toda a informação no Sistema Integrado de Registo Eletrónico de Resíduos, cumprindo as normas técnicas estabelecidas pela Autoridade Nacional de Resíduos [3]. No que diz respeito ao transporte, os produtores, os detentores, os transportadores e os destinatários devem registar todas as operações realizadas numa plataforma de acompanhamento deste processo nas Guias Eletrónicas de Acompanhamento de Resíduos (e-GAR) [4].

Ainda inserido no RGGR, desde 2007, Portugal tem em vigor uma Taxa de Gestão de Resíduos (TGR) com o intuito de incentivar, tantos os operadores como os consumidores, a adotarem comportamentos mais sustentáveis. Esta taxa tem como objetivo reduzir a produção de resíduos, assim como obter uma gestão mais eficiente dos mesmos [5].

1.2. Motivação

A preocupação ambiental tem vindo a ganhar um significado cada vez mais relevante na sociedade atual, especialmente no que diz respeito à gestão de resíduos. No contexto de Portugal, as atuais práticas na gestão de resíduos têm apresentado desafios cada vez maiores, exigindo uma maior dedicação a este tema. Neste sentido, ao longo do tempo, tem havido uma evolução notória no que diz respeito à crescente urbanização e padrões de consumo da população, sendo por isso necessário explorar cada vez mais as limitações nos sistemas de gestão de resíduos [6]. Este estudo, propõe-se analisar e, eventualmente, ajudar a encontrar soluções inovadoras para ultrapassar as barreiras na gestão de resíduos em Portugal, identificando e avaliando possíveis anomalias que possam comprometer a eficácia deste processo.

Uma vez que serão analisados dados entre 2019 e 2022 nesta investigação, é importante ressalvar que Portugal não alcançou os objetivos comunitários estipulados para a recolha de resíduos nos anos de 2019 e 2020, o que demonstra a existência de dificuldades na gestão de resíduos [7].

Através da identificação de anomalias na gestão de resíduos, e a partir dos dados disponibilizados pela APA, esta investigação poderá ter um impacto relevante, contribuindo para uma sociedade mais sustentável e saudável. Através destes, ter-se-á uma visão concreta do comportamento dos produtores e transformadores de resíduos e a forma como é feita a monitorização do transporte destes. Deste modo, será possível detetar erros e consequentemente otimizar todo o processo de gestão de resíduos.

Ao verificar alguns dados estatísticos, de acordo com a APA, em 2022 foram produzidas cerca de 5 323 mil toneladas de resíduos urbanos, representado por um aumento de 0,24 pontos percentuais

face ao ano anterior [8]. Este crescimento tem sido notório desde 2016, e apesar de lento, tem-se verificado constante ao longo dos últimos anos [3]. Para além disso, grande parte destes resíduos são depositados em aterro (mais de 50% desde 2019), sendo esta uma percentagem ainda significativa. O restante acaba por ter um destino mais útil, como para tratamentos mecânicos e biológicos, e até mesmo valorização energética, material ou orgânica. Desta forma, a previsão feita para 2035 é reduzir ao máximo os resíduos colocados em aterro, tendo a meta de atingir os 10% [8].

Atualmente vive-se numa sociedade que está virada cada vez mais para o mundo digital, no entanto, os resíduos elétricos e eletrónicos são ainda reciclados numa quantidade muito reduzida face ao que é expectável. Por essa razão deve-se adotar uma fiscalização mais rigorosa, prática esta que não tem acontecido nos últimos anos. Esta situação deve-se também ao facto de Portugal ter um sistema com falta de financiamento [9].

Por se tratar de um tema preocupante para o país, várias são as entidades que debatem este assunto, e de acordo com a APA, Portugal recicla apenas 21% dos resíduos urbanos, valor este, distante da meta estabelecida pela Comissão Europeia de 55%. A notícia publicada pelo Jornal de negócios em outubro de 2023, revela que tem havido uma estagnação na última década na evolução da gestão de resíduos [1], sendo que Portugal encontra-se entre os 10 países da União Europeia em situação de risco de não alcançar as metas estabelecidas, e esta situação deve-se a uma regressão nos resultados, comparativamente aos níveis alcançados antes do início da pandemia em 2020 [9].

O incumprimento contínuo das metas europeias evidencia a necessidade de reformulação nos modelos económicos e operacionais, impulsionando o país a adotar práticas sustentáveis para evitar penalizações financeiras e promover uma gestão eficaz dos resíduos. Este desafio exige uma abordagem integrada, envolvendo todos os intervenientes, a fim de alcançar os objetivos estabelecidos para a reciclagem até 2025 e 2030 [4].

Numa era em que a sustentabilidade ambiental é uma prioridade a nível global, pretende-se chegar a conclusões e soluções tangíveis para um problema real e urgente de colmatar em Portugal. Ao abordar um tema de cariz tão relevante, esta investigação será preponderante para a construção de um futuro mais sustentável e resiliente para as gerações futuras.

1.3. Questões de Investigação

Para a realização deste estudo, ter-se-á como ponto de partida as seguintes questões de investigação:

- É viável, através dos dados disponibilizados pela APA, detetar anomalias no registo do transporte / gestão de resíduos?
- Como se caracterizam os produtores e transformadores atuais em Portugal?

- Qual a relação que os Códigos de Atividade Económica têm na caracterização dos produtores e transformadores face aos tipos de resíduos analisados?

1.4. Objetivos

Para dar resposta às perguntas de investigação formuladas, esta dissertação tem como objetivo explorar as várias possibilidades de deteção de anomalias na gestão de resíduos, em Portugal, entre 2019 e 2022 baseada nos dados disponibilizados pela APA. Neste sentido, pretende-se desenvolver e otimizar diversas técnicas que permitam identificar irregularidades nos registos do processo de gestão de resíduos.

Adicionalmente, torna-se crucial perceber o papel dos produtores e transformadores que atuam nesta área e, assim, analisar a forma como estes funcionam. Deste modo, garante-se que cumprem com o seu papel neste processo.

Para além disso, é relevante analisar se as atividades económicas podem ajudar a caracterizar os produtores e transformadores de resíduos.

Estes objetivos visam identificar possíveis anomalias nos dados, mas também promover um melhor conhecimento dos produtores e transformadores de resíduos em Portugal. Para garantir que todos eles vão ao encontro do pretendido, utilizar-se-ão dados disponibilizados pela APA com informações relativas aos tipos de resíduos, quantidades associadas a cada tipologia, produtores, transformadores, códigos de atividade económica entre outras variáveis relevantes para este estudo.

1.5. Metodologia

A metodologia que será adotada nesta dissertação será o CRISP-DM, *Cross Industry Standard Process for Data Mining*, uma abordagem estruturada indicada para projetos na área de *Data Science*, mais concretamente para a realização de *Data Mining*.

Esta metodologia facilita as análises realizadas, na medida em que torna-se flexível e eficaz a forma como são tratados os dados. Composto por 6 fases, iterativas e independentes, permite que haja uma adaptação continua ao longo do desenvolvimento desta dissertação [10]. Na Figura 1, observa-se como é que deve ocorrer um projeto que segue esta metodologia [11].

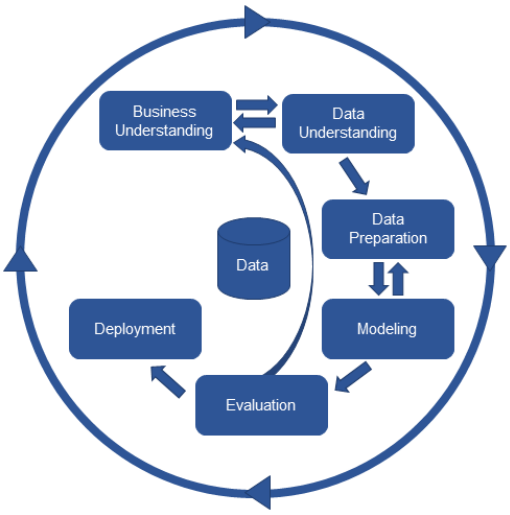


Figura 1 – CRISP - DM [11]

As seguintes fases dividem-se das seguinte forma:

- *Business Understanding*: Nesta fase inicial, o objetivo passa por perceber em conjunto com a APA os requisitos e os objetivos a alcançar ao longo desta dissertação. Assim, nesta fase, pretende-se identificar os problemas e os desafios que a APA apresenta e que pretende ver resolvidos com este estudo.
- Data Understanding: Após a compreensão dos objetivos da organização, a próxima etapa consiste na recolha e exploração dos dados disponíveis. Esta fase inclui a descrição dos dados, a garantia da qualidade dos mesmos e a identificação de quaisquer problemas ou anomalias que se detetem nos dados fornecidos. Nesta fase pretende-se ainda identificar as variáveis relevantes e obter uma visão geral sobre as informações contidas e que poderão ser analisadas resultando em conclusões pertinentes para a APA.
- Data Preparation: A preparação dos dados é uma fase crucial para organizar e transformar os dados, de modo a que estes estejam tratados corretamente e prontos a serem analisados. Esta etapa inclui a limpeza dos dados, integração de informação de diferentes ficheiros fornecidos pela APA, seleção de variáveis relevantes e a criação de novas, caso faça sentida para este estudo. Para além disso, irá proceder-se à divisão entre aquilo que são os registos de transporte dos resíduos, caracterização de produtores e transformadores, facilitando assim as etapas seguintes.
- Modeling: Nesta fase, são aplicadas técnicas de *data mining* para a criação de modelos preditivos ou descritivos que respondam às questões de investigação definidas inicialmente. Para tal, recorrer-se-á ao *clustering*, nomeadamente o *K-means*, para além de todas as medidas descritivas que permitirão retirar conclusões relevantes sobre os dados fornecidos.

Assim, será possível identificar padrões e/ou correlações sobre a forma como são geridos os resíduos por parte dos produtores e transformadores portugueses.

- Evaluation: A avaliação consiste na validação da qualidade dos modelos e análises descritivas criadas na fase anterior. Desta forma, pretende-se analisar os objetivos definidos inicialmente e perceber se está a ser dada a resposta necessária, e se vai ao encontro das necessidades da APA, garantindo um contributo que cause impacto na organização.
- Deployment: A última fase do CRISP-DM é a implementação dos modelos e das soluções desenvolvidas. Esta fase irá incluir a apresentação dos resultados, a criação de *dashboards* e a definição de ações práticas baseadas nos *insights* obtidos. No âmbito desta dissertação, as ações que possam resultar do trabalho desenvolvido serão recomendações para a APA, com vista a otimizar os processos internos e melhorar a sustentabilidade ambiental.

A aplicação desta metodologia permite uma abordagem sistemática e estruturada à análise de dados, garantindo que todas as etapas do processo são devidamente cumpridas e que os objetivos da APA são atingidos. Com o CRISP-DM, espera-se obter uma compreensão aprofundada e detalhada da gestão de resíduos em Portugal, identificando áreas críticas e propondo soluções inovadoras e eficazes.

CAPÍTULO 2

Revisão da Literatura

A deteção de anomalias na gestão de resíduos emerge como uma preocupação cada vez mais constante na sociedade, impulsionada pela crescente consciencialização ambiental que tem surgido ao longo dos últimos anos. Desta forma, surge a necessidade de perceber qual o impacto que a produção de resíduos e consequente distribuição dos mesmos tem causado em Portugal.

Neste contexto, a revisão de literatura visa fornecer uma análise abrangente das abordagens teóricas e práticas, que fundamentam a deteção de anomalias na gestão de resíduos. Assim, com o objetivo de perceber qual o impacto que a Inteligência Artificial pode ter neste tema, pretende-se perceber que técnicas podem ser aplicadas para otimizar o processo de gestão de resíduos, assim como as boas práticas já adotas noutros países e os moldes existentes atualmente em Portugal.

2.1. Processo de Revisão de Literatura

Para a realização da revisão de literatura torna-se necessário adotar uma metodologia capaz de assegurar a veracidade dos estudos analisados. Deste modo, recorrer-se-á ao PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) com vista a organizar todo o processo [12].

A partir desta metodologia, definiram-se 3 queries em 3 bases de dados distintas e fidedignas, sendo estas o Scopus, IEEE e o Web Of Science. Em cada uma destas aplicaram-se exatamente as mesmas queries, de forma a perceber quais é que ofereciam melhores resultados. É ainda de referir que em cada uma das bases de dados, aplicaram-se alguns filtros, nomeadamente para aceder a artigos realizados na última década (2014-2024), artigos de livre acesso e categorizados por waste management e/ou envirnomental sciences. Estas filtragens foram feitas antes de analisar cada um dos artigos, numa ótica de obter mais facilmente o conteúdo pretendido.

As queries utilizadas foram as seguintes:

- Waste AND Management AND Portugal;
- Waste AND Management AND Europe AND Union;
- Artificial AND Intelligence AND Waste AND Management.

De seguida, segue-se o esquema que demonstra todo o processo na seleção dos artigos utilizados para a realização da revisão de literatura (Figura 2).

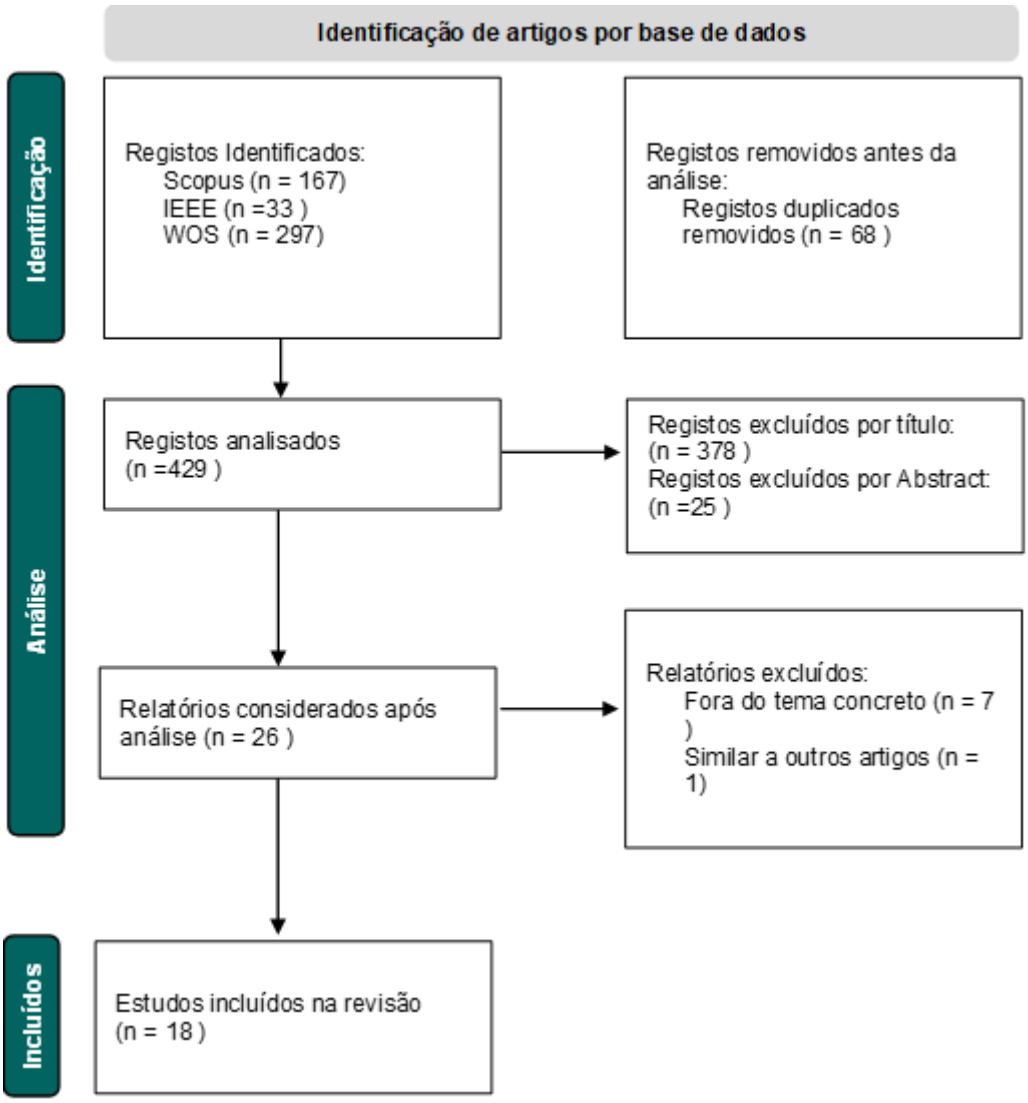


Figura 2 - Metodologia PRISMA, Adaptado de [12]

Através do esquema apresentado (Figura 2), verifica-se a existência de 68 artigos duplicados que foram removidos antes de iniciar a análise destes. De seguida, excluíram-se diversos artigos ao ler o titulo (378), tendo em consideração que não iam ao encontro do tema desta dissertação. Adicionalmente, excluíram-se 25 artigos após a leitura do *abstract* por não corresponder ao que se pretende explorar nesta investigação. Desta forma, leram-se, na totalidade 26 artigos, sendo que 8 não foram incorporados na revisão de literatura por abordarem questões não relacionadas com este tema, ou por já possuir na revisão outros artigos similares.

Para além dos artigos incluídos através da metodologia PRISMA, foram adicionados 5 artigos extra que permitiram adicionar informação relevante sobre alguns estudos realizados. Estes foram, em alguns casos, encontrados com critérios de seleção no Google Académico, e noutros casos referenciados pelos orientadores desta dissertação. Desta forma, perfez-se um total de 23 artigos.

2.2. Gestão de Resíduos no Estrangeiro

Para ter uma visão mais abrangente do que acontece sobre a gestão de resíduos em Portugal, torna-se crucial perceber o comportamento dos países pertencentes à União Europeia (UE), como medida de comparação dos resultados obtidos no estrangeiro [13]. Anualmente, a União Europeia produz cerca de 2 500 mil milhões de toneladas de resíduos [14], no entanto a forma como estes são geridos tem vindo a melhorar ao longo do tempo [13]. Este facto é corroborado pelo facto da taxa de reciclagem e compostagem da UE ter aumentado cerca de 10 pontos percentuais em 21 anos (de 2000 a 2021). Já a taxa de deposição em aterros da UE diminuiu para mais de metade no mesmo período. Estes resultados são a combinação de fatores, incluindo a aplicação de novas políticas da UE, investimentos em infraestruturas de reciclagem e compostagem e ainda uma maior sensibilização pública para a importância da gestão de resíduos [13].

A gestão de resíduos a nível mundial surge com um tema urgente e cada vez mais relevante na sociedade, uma vez que se prevê um aumento de 70% na criação anual de resíduos até ao ano de 2050 [15]. Fatores que implicam este aumento ao longo dos anos são a densidade populacional, o PIB per capita e o turismo. Na UE, um estudo mencionado pela Comissão Europeia, estima que com a aplicação dos princípios da economia circular à economia da UE ,o produto interno bruto (PIB) poderá aumentar em 0,5% até 2030 [15].

Em 2023, o Luxemburgo e a Irlanda, destacaram-se como os países da União Europeia que apresentaram métodos mais eficazes na gestão de resíduos [16]. Por outro lado, a Dinamarca, a Alemanha, a Noruega e a Suíça foram líderes na produção de resíduos em 2021 [17], havendo assim uma maior tendência, para que países com maior poder económico, produzam uma maior quantidade de resíduos por pessoa [13].

Para além disso, segundo um artigo publicado em 2023, a União Europeia tem vindo a tentar implementar cada vez mais práticas de prevenção, reutilização, reciclagem e compostagem, com vista a reduzir a quantidade de resíduos em aterros. É ainda de salientar que cerca de 49,6% dos resíduos urbanos na UE foram reciclados. Tendo em consideração que a meta estabelecida a nível europeu para 2030 será de 60%, verifica-se que ainda existe um longo caminho a percorrer, uma vez que se pretende transitar até 2050 para uma economia cada vez mais circular [13]. Esta economia surge como um pilar fundamental da implementação do Green Deal Europeu, com particular destaque para a gestão de resíduos de embalagens, que desempenham um papel crucial no cumprimento das metas de reciclagem estabelecidas. Contudo, a diversidade de modelos de resíduos existentes nos vários países da União Europeia representam um desafio, tendo em conta que as leis e objetivos impostos não são uniformes, dificultando o cumprimentos do mesmos. Desta forma, deverá partir da União Europeia a

necessidade de ser capaz de construir uma visão estratégica dos diferentes sistemas de gestão de resíduos que cada país da União Europeia poderá vir a adotar [18].

É ainda de realçar que, apesar da maior parte dos países terem vindo a aumentar a quantidade de resíduos urbanos per capita, alguns já têm apresentado uma diminuição face a anos anteriores, como é o caso, da Bulgária, Malta, Chipre, Espanha e Roménia [13]. No entanto, países que aumentaram a produção de resíduos apresentam também uma tendência superior na alocação de recursos financeiros para a gestão de resíduos, como é o caso da Eslováquia, demonstrando uma maior preocupação ambiental. Desta forma, algo a ter em consideração é a relação existente entre o investimento atribuído para o sistema que este país recorre para a gestão de resíduos com a quantidade de resíduos produzidos pelo mesmo. Tendo em consideração que a União Europeia apresenta uma tendência semelhante (em que os resíduos e os custos aumentam), devem ser analisados individualmente cada caso, sobre o que está a acontecer para a falta de melhores resultados, tendo em conta os gastos atribuídos para esta causa [19].

Segundo um estudo realizado em 2021, e tal como referido anteriormente, cada país adota práticas diferentes em relação à gestão dos seus resíduos, e por isso, ao efetuar-se uma comparação entre a Áustria, a Suécia e a Finlândia corrobora-se este facto. Vários são os fatores que contribuem para esta diferença tão significativa entre países da UE, nomeadamente, as políticas de cada país, as infraestruturas disponíveis para o efeito, ações de sensibilização dinamizadas em cada um deles e a qualidade de como são recolhidos, armazenados e tratados os dados sobre os resíduos. Neste caso, destaca-se a Finlândia como o país de entre estes 3, o que apresenta maiores barreiras para obter uma melhor gestão de resíduos. Estas barreiras incidem essencialmente sobre a capacidade de incineração, o mau funcionamento e a falta de conceção por parte das entidades de reciclagem [20].

Na União Europeia, a Responsabilidade Alargada do Produtor (EPR) tem sido implementada como um dos principais mecanismos de gestão de resíduos, focando-se na reciclagem e reutilização de materiais. Produtores e transformadores são incentivados a adotar práticas de *eco-design* que facilitam a reciclagem e reduzem o impacto ambiental. Países como a Alemanha, os produtores são responsáveis por cobrir os custos da recolha e tratamento de resíduos, incluindo embalagens e resíduos eletrónicos [21].

Ainda dentro do que acontece na Europa, na Holanda, mais concretamente em Delft, já se recorre a tecnologias de Internet das Coisas (IoT) e IA, tornando-a numa cidade mais inteligente. Atualmente, já existe um sistema inteligente que contém sensores em 15 contentores e que recolhem informação através de uma Rede Neuronal Convolucional (CNN) para prever a taxa de resíduos no contentor. Este é um sistema que aborda um problema recorrente e que soluciona facilmente algumas das barreiras em otimizar todo o processo de gestão de resíduos. Para além disto, permite que as rotas delineadas

para os camiões do lixo sejam definidas estrategicamente com base nos resultados obtidos. Estas rotas foram calculadas e desenhadas através de um algoritmo baseado em grafos, estando disponível num painel de gestão inteligente de resíduos, disponibilizados através de uma aplicação móvel para os recolhedores de resíduos. Tendo em consideração, que tem impactado de uma forma bastante significativa a gestão dos resíduos pretende-se expandir este método através da instalação de 50 novos sensores [22].

Extrapolando o que acontece para resto do mundo (fora da UE), um estudo realizado em 2022, em Istambul, teve como propósito avaliar a viabilidade financeira associada à construção de uma instalação de incineração de resíduos. Com foco em apostar fortemente em tecnologia cada vez mais avançada, acredita-se que esta seja uma boa forma de gerar valor económico a partir de resíduos [23]. Uma vez que a Turquia foi em 2021 o país que mais recebeu resíduos provenientes de outros países (cerca de 14,7 milhões de toneladas), faz todo o sentido que se aposte neste tipo de iniciativas para que possa evoluir neste ramo. No entanto, a UE tem feito um esforço para assegurar que todas estas exportações que são realizadas , para países como a Turquia, são realizadas de forma legal, garantido que os resíduos são corretamente tratados nos países de destino [13].

Com o intuito de ter uma perspetiva ainda mais abrangente, torna-se necessário perceber o que é feitos em países com realidades completamente distintas à de Portugal. Sendo a Austrália um país bastante evoluído, este tem recorrido a métodos mais avançados para conseguir otimizar todo o processo de gestão de resíduos sólidos, e simultaneamente, reduzir os custos associados a este tema. Desta forma, já são postas em prática técnicas de inteligência artificial que permitem às entidades competentes terem uma maior quantidade de informação, no que toca à recolha, classificação, reciclagem e monitorização dos resíduos (apenas aplicável a resíduos sólidos). Tendo em conta que estes dados são recolhidos em tempo real, é percetível que o sistema se torne cada vez mais robusto, consistente, e escalável para outras realidades. No entanto, a escalabilidade não deixa de ser uma barreira, na medida em que países menos desenvolvidos ainda não são capazes de implementar este tipo de abordagens, por falta de dados e recursos adequados. Atualmente, este projeto está a ser implementado apenas na Austrália uma vez que apresenta um âmbito mais curto [24].

Assim, ao longo dos vários estudos analisados no estrangeiro, é notório que já estão postas em prática diversas medidas, e por essa razão, Portugal é um país ainda com muita margem de evolução, uma vez que poderá ter por base alguns dos métodos mencionados anteriormente.

2.3. Gestão de Resíduos em Portugal

Com a entrada no século XXI, Portugal sofreu uma revolução nos sistemas de gestão de resíduos, em que ao longo dos últimos anos é notória uma evolução, na medida em que o foco está totalmente

direcionado para a redução do consumo e da produção de resíduos. Desta forma, Portugal é um potencial país que pode vir a tirar proveito de uma economia circular (Figura 3, adaptado de [25]) através da adoção de estratégias sustentáveis na gestão de resíduos, mais especificamente no que diz respeito à produção de plásticos. [25].

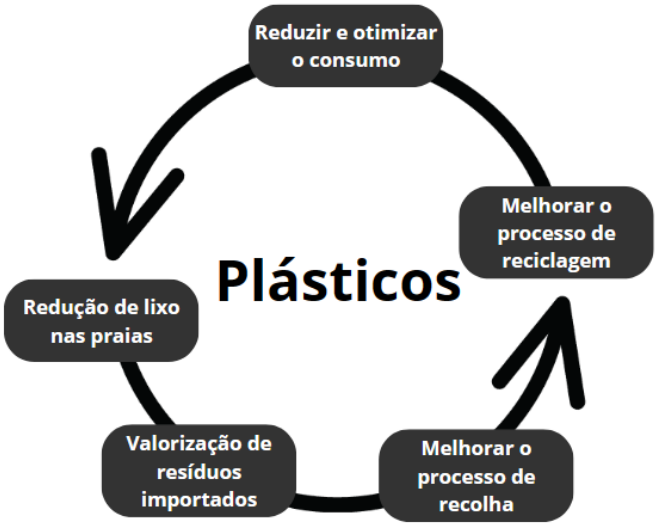


Figura 3 - Economia circular na Gestão de Resíduos em Portugal, adaptado de [25]

Em Portugal, uma das grandes preocupações está associada ao tratamento e destino inadequado de plásticos, nomeadamente em aterros. Por essa razão, um estudo realizado pela Universidade de Coimbra, revela que o país possui um índice de reciclagem de plástico reduzido, e por isso, deverão ser implementadas políticas públicas que vão ao encontro da resolução deste problema [26]. No entanto, com o intuito de acelerar o processo de transição para uma economia circular, foi criado em 2020, pela Associação Smart Waste Portugal, o Pacto Português para os Plásticos, que reúne diversos tipos de organizações (ONGs, associações, entre outras entidade) [27]. Este visa "eliminar os plásticos de uso único considerados problemáticos e/ou desnecessários; Garantir que 100% das embalagens são reutilizáveis, recicláveis ou compostáveis; Garantir que 70%, ou mais, das embalagens plásticas são efetivamente recicladas, através do aumento da recolha e da reciclagem; Incorporar, em média, 30% de plástico reciclado nas novas embalagens de plástico; Promover atividades de sensibilização e educação aos consumidores (atuais e futuros) para a utilização circular dos plásticos" [28]. O cumprimento deste pacto permitirá ao país avançar no âmbito da gestão de resíduos aproximando-se daqueles que são os objetivos a nível europeu [4].

Desde a década de 70, que é crucial conseguir tirar proveito dos resíduos, quer a nível económico, quer a nível social e ambiental. Contrariamente ao que é expectável, este reaproveitamento contribui para a redução de custos no processo de eliminação de resíduos. Através da União Europeia, existem metas de reciclagem, com o objetivo de otimizar as quantidades de resíduos reciclados. As empresas

portuguesas podem também ter um papel preponderante ao incentivar este tipo de práticas, e por essa razão analisou-se a eficiência destas no processo de reciclagem. Através da utilização de métodos de avaliação não paramétricos, verifica-se que a falta de incentivo é a principal razão para as falhas existentes nos sistemas de gestão de resíduos [29].

Nos dias de hoje, o método mais utilizado no país é a eliminação de resíduos em aterros licenciados. Contudo, há setores que podem optar por outras soluções, como por exemplo para a construção e destruição de estruturas (como madeira, metal, tijolos, entre outros), pode-se recorrer a uma maior reutilização deste tipo de resíduos. Estes continuam com uma baixa taxa de utilização (cerca de 6% em 2021), tendo margem para obter melhores resultados no futuro [30].

Desta forma, surge a necessidade de um maior apoio, por parte das entidades governamentais, e assim, tornar este um assunto de maior importância no meio político e social, e consequentemente possam ser tomadas ações que promovam uma melhor gestão de resíduos, tendo em consideração fatores económicos, culturais e tecnológicos [30] .

No entanto, a população portuguesa, maioritariamente, desconhece alguns dos novos métodos utilizados neste processo. Segundo um estudo realizado em 2022, pela Universidade de Aveiro a 626 inquiridos, 82% destes revelam este desconhecimento, apesar mostrarem abertura para a implementação de projetos inovadores neste ramo. Fatores demográficos, como a idade e o sexo, tem uma influência significativa em relação aos resultados obtidos, evidenciando a necessidade urgente de criar novas estratégias de sensibilização para as diferentes faixas etárias e géneros. Para além do desconhecimento neste tema, os portugueses revelam um descontentamento a nível económico e social que está diretamente relacionado com o impacto ambiental no âmbito da gestão de resíduos [31].

Numa ótica de compreender o desempenho da gestão de resíduos em Portugal, um estudo realizado pelo Centro de Estudos de Recursos Naturais, Ambiente e Sociedade, analisaram 42 municípios portugueses onde apresentam duas vertentes distintas: a qualidade da recolha feita dos resíduos e possíveis fatores sociodemográficos e económicos que possam influenciar os resultados das organizações responsáveis por esta matéria. Assim, verificou-se que o número de habitantes por zona de recolha, a acessibilidade a ecopontos, o grau de urbanização, o nível de escolaridade da população e a área geográfica podem ter um impacto significativo na forma como é feita a gestão de resíduos. Para tal, utilizaram-se modelos de regressão, de modo a quantificar o efeito de cada um destes fatores no processo de recolha de resíduos reciclados. Tal como referido anteriormente, Portugal, assim como todos os países da UE têm metas que devem ser cumpridas, e estudos como estes, poderão ser uma mais valia na medida em que as autoridades locais e regionais poderão definir estratégias que vão ao encontro das metas estabelecidas [32].

A nível mais tecnológico, Portugal já faz atualmente vários testes no âmbito da utilização de sensores que permitam estudar de uma forma mais eficaz a gestão de resíduos. Um estudo realizado em Bragança demonstra que já é possível, numa dimensão micro, medir o nível de resíduos existentes. Contudo verificam-se ainda várias limitações, nomeadamente a necessidade de sensores mais avançados que permitam abranger uma dimensão maior da sociedade e que estas operações estejam integradas nas cidades portuguesas, para além da necessidade de atualizar todas as infraestruturas portuguesas que acompanhem o uso deste tipo de tecnologia. Ao serem testados três sensores (S1, S2, S3) controlados pelo Arduino UNO, constatou-se que o que obtém melhores resultados é o S2, apesar do S3 ser o mais económico [33].

2.4. Inteligência Artificial na Gestão de Resíduos

A Inteligência Artificial tem cada vez mais um papel preponderante na gestão de resíduos, uma vez que é necessário perceber a melhor forma para otimizar todo o processo de gestão dos mesmos, desde o momento em que um indivíduo o coloca no contentor, até ao processo de recolha e de divisão de materiais nos centros de reciclagem.

É notória a aplicabilidade que a Inteligência Artificial (IA) pode ter neste âmbito, tal como na conversão de resíduos em energia, contentores inteligentes, redução de custos e consequente melhoria a nível de saúde pública. Assim, segundo um estudo realizado em 2023, a IA poder ser uma ferramenta com soluções automatizadas, que podem vir a reduzir a distância no processo de transporte até aproximadamente 37%, reduzir custos até 13% e economizar tempo até 28% [34].

Ao longo do tempo, têm surgido cada vez mais técnicas e modelos capazes de lidar com este tipo de problemas. Ao longo do processo de otimização da gestão de resíduos há que ter em consideração quais os modelos que mais se adequam à situação específica. Todos eles apresentam cada vez mais resultados positivos em abordagens relativas à gestão de resíduos, sendo as redes neuronais e o Support Vector Machine como aqueles que se destacam com maior número de vantagens para este tipo de aplicação [35].

Adicionalmente, os modelos de inteligência artificial são soluções eficazes, no que diz respeito à previsão de resíduos e roteiros de veículos. Contudo, a falta de dados e a falta de colaboração por parte das entidades competentes emergem como entraves significativos à implementação efetiva destas mesmas soluções. A opacidade inerente aos modelos de IA também se apresenta como um desafio, dificultando a sua adoção generalizada. Para além disso, surge ainda a necessidade de encontrar soluções mais personalizadas para os sistemas de gestão de resíduos [35], como é o caso da utilização de uma rede neuronal convolucional (CNN), das mais utilizadas para o efeito. Um estudo realizado em 2023, classificou imagens de resíduos em sete categorias (cartão, vidro, metal, orgânico,

papel, plástico e lixo), através da utilização de diversos modelos, destacando o VGG19 como o que obteve menor precisão e o InceptionV3 como o modelo que obteve melhores resultados na classificação do tipo de resíduos [36].

No que diz respeito, à gestão de resíduos sólidos urbanos baseados em técnicas de IA, existem já diversas investigações que destacam a potencial contribuição da comercialização de iniciativas fundamentadas em IA para atingir objetivos de desenvolvimento a longo prazo, de modo a promover a preservação ambiental. Desta forma, reconhece-se a necessidade de investigar e construir cada vez mais modelos específicos para a gestão de resíduos sólidos. Para tal, as *decision trees* consideram-se uma boa solução para a deteção de lixo ilegal e para a definição de estratégias capazes de alocar os resíduos de uma forma mais eficaz e económica [37], apresentado um erro de classificação de apenas 3,6%.

Para além dos modelos tradicionais de inteligência artificial referidos anteriormente, existem muitos outros que podem ser uma mais valia para a gestão de resíduos. O modelo AIEWO-WMC visa aprimorar a eficiência na gestão de resíduos urbanos, baseando-se em técnicas avançadas de IA, como o *deep learning* e *computer vision* para o reconhecimento e categorização exata de objetos residuais. Este é um dos modelos que mais se destaca pela elevada precisão que contém, atingindo uma taxa de classificação de 99,15% para os objetos residuais. A integração de técnicas de *deep learning*, a deteção de objetos baseada em RetinaNet, o otimizador Adagrad e a utilização do algoritmo de *autoencoder* empilhado (SAE) confere ao modelo uma base tecnológica mais robusta e inovadora. Modelos como este, permitem ter uma maior noção, de que forma, a inteligência artificial pode ter um papel fundamental na estratégia definida para a implementação de ambientes urbanos mais inteligentes [38].

Neste contexto, o desenvolvimento de contentores inteligentes, poderá promover uma gestão eficiente e inovadora dos resíduos através de IoT. Atualmente, já existem sensores nestes objetos que permitem monitorizar automaticamente o nível dos caixotes, e desta forma, garantir uma melhor limpeza das cidades e redução dos custos associados à recolha de resíduos. No entanto, para garantir que este método é sustentável, há que ter em consideração a quantidade de energia gasta e verificar se compensa efetivamente a sua utilização, tornando estas cidades mais inteligentes. O protótipo apresentado é autónomo, na medida em que ativa-se automaticamente quando deteta a proximidade de resíduos. A utilização de sensores ultrassónicos, de ESP8266 (microcontrolador que comunica via *wifi*), e a aplicação *Blynk* possibilitam a monitorização em tempo real, permitindo a transmissão eficiente de dados sobre o estado destes contentores. Este tipo de sistemas inteligentes representam uma evolução na gestão de resíduos, para além de oferecerem uma solução prática para diversos obstáculos impostos pelos sistemas tradicionais, como referido anteriormente [39].

Assim, é possível afirmar que no meio da inteligência artificial existem diversos caminhos a seguir, entre as escolhas mais tradicionais às mais inovadoras, existem diversos mecanismos que se podem adotar, numa ótica de mudar o processo de gestão de resíduos. Atualmente, existem técnicas que podem vir a substituir o sistema tradicional de gestão de resíduos, através de sensores inteligentes que comunicam diretamente com o LoRa e o *tensorflow* (modelo de *deep learning*). Este é um método identificado num estudo realizado em 2022, semelhante ao anteriormente descrito, com uma abordagem diferente em que o procedimento ocorre quando o LoRa envia os dados dos sensores, em que simultaneamente, o *tensorflow* deteta e classifica os resíduos em tempo real. O sensor colocado está incorporado dentro dos contentores, de modo a monitorizar a quantidade resíduos, permitindo ainda através de um sistema de GPS saber a localização exata do contentor [40].

A utilização deste tipo de tecnologia permite otimizar todo o processo de recolha e de separação de resíduos, na medida em que grande parte da população mundial não realiza ainda corretamente a separação dos resíduos, que permita obter resultados positivos. No entanto, a aplicabilidade deste tipo de tecnologia de forma massiva teria custos associados, algo a ter em conta por parte das organizações governamentais [40].

2.5. Resumo da Revisão de Literatura

A tabela abaixo representa um resumo geral dos artigos mais relevantes para cada uma das secções utilizadas ao longo da revisão de literatura, assim como a fonte utilizada para cada um deles (Tabela 1).

Estão representados os artigos que tiveram um maior impacto para este estudo, tendo em consideração cada uma das secções. Para o estrangeiro, foi possível perceber as práticas já utilizadas e casos de estudo para outros países sobre a forma como são geridos os resíduos. Em relação a Portugal, faz-se referência à gestão dos resíduos em Portugal, assim como a introdução da vertente mais tecnológica no país. Por fim, a última secção com foco no uso de modelos e algoritmos capazes de otimizar o processo de gestão de resíduos, para além da aplicabilidade de IoT neste ramo.

Tabela 1 - Artigos principais por secção e fonte utilizados através do Prisma

Secção	Títulos	Fonte
Gestão de Resíduos no Estrangeiro	[18] An Overview of Packaging Waste Models in Some European Countries	Scopus
	[19] Public Expenses and Investment in Environmental Protection and Its Impact on Waste Management	Scopus
	[17] Waste management as an element of sustainable development of the circular economy in the European Union	Scopus
	[20] Different pathways to a recycling society e Comparison of the transitions in Austria, Sweden and Finland	WOS
Gestão de Resíduos em Portugal	[32] Dataset of socio-economic and waste collection indicators for Portugal at municipal level	Scopus
	[33] Node Assembly for Waste Level Measurement: Embrace the Smart City	Scopus
	[30] Current state, comprehensive analysis and proposals on the practice of construction and demolition waste reuse and recycling in Portugal	WOS
	[31] A social receptiveness analysis on smart waste management – A case study for Portugal	WOS
Inteligência Artificial na Gestão de Resíduos	[36] Classification of Waste Materials using CNN Based on Transfer Learning	Scopus
	[40] An Internet of Things Based Smart Waste Management System Using LoRa and Tensorflow Deep Learning Model	IEEE
	[38] Artificial intelligence with earthworm optimization assisted waste management system for smart cities	WOS
	[35] Artificial intelligence applications in solid waste management: A systematic research review	WOS
	[39] IoT based Waste Management for Smart Cities	IEEE

2.6. Abordagem diferenciadora

Com base na investigação realizada para a revisão de literatura, torna-se pertinente perceber, de que forma esta dissertação se vai diferenciar face aos diversos estudos realizados nesta área. Assim, esta dissertação visa explorar os dados fornecidos pela APA, sendo já à partida algo que nunca foi realizado, na medida em que o foco do estudo está direcionado para as necessidades desta organização. Desta forma, realizar-se-á uma análise precisa e detalhada numa ótica de identificar anomalias que existam atualmente no sistema de gestão de resíduos em Portugal. Assim, ao identificar padrões pouco comuns ou discrepâncias nos dados, poderão ser retiradas conclusões relevantes e pertinentes para a APA.

Os principais fatores que diferenciam este estudo passam pela caracterização específica dos produtores e transformadores em Portugal e a relação que os códigos de atividade económica podem ter na classificação de resíduos. Deste modo, é possível perceber o impacto que a má classificação pode ter e, por consequência, a má gestão de resíduos pode impactar negativamente o país.

Adicionalmente, é ainda de referir que este estudo terá uma utilidade prática para a APA, na medida em que, através das anomalias detetadas poderá ganhar *insights* pertinentes e aplicáveis à organização, com vista a melhorar os métodos utilizados no registo do transporte de resíduos. Ao longo deste estudo, para além de se querer encontrar soluções, pretende-se ainda adotar técnicas inovadoras que permitam evoluir no método de gestão de resíduos, acompanhado da evolução tecnológica que se tem sentido nos últimos anos. Desta forma, acredita-se que este estudo terá ainda

um impacto ambiental, uma vez que é focado na tentativa de arranjar alternativas mais sustentáveis que promovam o bem estar social a nível ambiental e económico.

CAPÍTULO 3

Análise de Dados

3.1. Business Understanding

A APA tem um papel fulcral na supervisão e regulação da gestão de resíduos em Portugal, assegurando que as operações de recolha, tratamento e transporte de resíduos sejam realizadas de forma eficiente, e de acordo com as normas ambientais. Contudo, como detalhado no capítulo de contextualização desta dissertação, a APA enfrenta desafios importantes, nomeadamente no que diz respeito à garantia da qualidade dos dados recolhidos e ao acompanhamento eficaz das operações de gestão de resíduos. Este estudo visa contribuir para a resolução de alguns desses desafios, utilizando os dados fornecidos pela APA para identificar padrões, anomalias e possíveis melhorias nos seus processos internos.

O principal foco desta dissertação é dar resposta aos desafios através da análise dos dados fornecidos, com o intuito de fornecer soluções concretas que possam ser aplicadas pela APA. Através da deteção de irregularidades nos dados, espera-se desenvolver modelos que ajudem a organização a otimizar os seus processos, garantindo uma monitorização mais eficiente e uma gestão mais orientada para um país mais sustentável.

Este estudo visa ainda caracterizar os setores económicos que estão associados aos produtores e transformação de resíduos, fornecendo uma visão mais clara sobre as suas práticas e contribuindo para a adoção de soluções mais sustentáveis a longo prazo.

Para alcançar estes objetivos, serão utilizados Python e Power BI, ferramentas que permitirão criar modelos e relatórios que facilitem as análise referidas.

3.2. Data Understanding

A preparação dos dados é uma etapa fundamental para assegurar que os mesmos estão adequadamente preparados para o início do processo de análise. Esta fase permite perceber o que significam os dados através da descrição dos mesmos, identificar as variáveis relevantes e detetar anomalias ou possíveis problemas que estes possam vir a ter, para que sejam tratados na fase seguinte, seguindo a metodologia do CRISP-DM.

3.2.1. Descrição de variáveis de diferentes fontes de informação

Para este estudo foi fornecido pela APA um *dataset* com a informação principal designado por “view_ egar_guias_2019_2022”. Este contém 16 123 273 registos e 34 colunas. No apêndice a) encontra-se uma tabela com todas as variáveis que continha este csv.

Após verificar-se todas as colunas, selecionou-se apenas as variáveis para as quais se tinha conhecimento do que representavam no contexto da gestão de resíduos (por falta de resposta por parte da APA não foi possível esclarecer algumas dúvidas no que diz respeito ao significado de algumas das variáveis apresentadas no apêndice a)).

Excluindo algumas variáveis para as quais não foram recebidos esclarecimentos da APA atempadamente quanto ao seu significado, realizaram-se algumas medidas descritivas para perceber a distribuição das restantes variáveis e, desta forma, validar se seriam uma mais valia para este estudo. De seguida, na Tabela 2, encontram-se descritos os significados e o conteúdo de cada variável, tendo em consideração se vão ao encontro (ou não) das perguntas de investigação definidas inicialmente. Caso tenha um "x" na coluna utilizada, é porque foi relevante para o estudo.

Tabela 2 – Seleção das variáveis relevantes

Nome da Variável	Descrição da Variável	Utilizada
data_criacao	Data da criação do registo do transporte. Utilizada essencialmente para conseguir retirar o mês ou ano.	x
egar	Guia eletrónica de acompanhamento de resíduos. Contém apenas os códigos das guias, de cada transporte realizado.	
criada_por	Entidade responsável pelo transporte. Não foi utilizada, tendo em conta que o foco foi a caracterização dos produtores e transformadores e não a entidade que dá a entrada e/ou saída do transporte.	
produtor_origem	Tipo de Produtor. Contém categorias que não tem relevância para este estudo.	
produtor_nif	NIF do Produtor. Utilizado como chave primária para ligação a outras tabelas referidas mais abaixo.	x
produtor_nome	Nome do Produtor. Variável relevante para identificar os principais produtores e caracterização geral dos mesmos.	x
produtor_local	Local do Produtor. Nome da empresa a nível fiscal do produtor. Em alguns casos o mesmo nome que consta na variável produtor_nome.	
data_estado	Data de quando foi verificado o estado do transporte. A nível temporal utilizou-se variáveis associadas à data de início e fim de transporte	
estado	Estado do transporte. Pode assumir várias categorias.	x
data_inicio_transporte	Data de início de transporte, sendo relevante para conseguir estimar o tempo médio de um transporte.	x
data_fim_transporte	Data de fim de transporte, sendo relevante para conseguir estimar o tempo médio de um transporte.	x
dados_transportadores	Dados do transportador	
destinatario_nif	NIF do Transformador. Não tem relevância para extrair informação de outra fonte de informação.	
destinatario_nome	Nome do Transformador. Variável relevante para identificar os principais transformadores e caracterização geral dos mesmos.	x
destinatario_estabelecimento	Nome da empresa a nível fiscal do transformador. Em alguns casos o mesmo nome que consta na variável destinatario_nome.	
ler_codigo	Código do tipo de resíduo. Variável base para todas as análises realizadas ao longo da dissertação.	x
ler_perigosidade	Variável booleana que pode indicar a perigosidade do resíduo. Classificação feita com a variável mercadorias_perigosas.	
tipo_op_codigo	Código de utilização final dos resíduos.	x
quantidade	Quantidade enviada pelo Produtor (em Kilogramas).	x
tipo_op_recebido_codigo	Código de utilização final dos resíduos após transporte.	x
quantidade_recebida	Quantidade Recebida pelo Transformador (em Kilogramas).	x
ler_recebido_codigo	Código do tipo de resíduo recebido pelo transformador.	x
ler_recebido_perigosidade	Variável booleana que pode indicar a perigosidade do resíduo. Classificação feita com a variável mercadorias_perigosas.	
autorizada	Autorização do transporte (true/false).	x
origem	Origem dos dados (formulário ou WS_V2).	
mercadorias_perigosas	Classificação booleana sobre a perigosidade dos resíduos transportados.	x

Para além disso, foram disponibilizados 16 formulários (em formato xlsx) com informações sobre os produtores e transformadores. É de referir, que entre estes formulários, havia pouca uniformização de informação, tendo em conta que tinham um formato muito distinto entre todos eles (colunas com diferentes informações). Por essa razão, utilizou-se apenas a informação que fosse estritamente necessária, fazendo apenas *merge* ao *dataset* principal das colunas relevantes e que estavam em comum nos 16 formulários.

Destes extraiu-se informação adicional relativa ao Código de Atividade Económica e ao Distrito (para o distrito apenas está disponível a informação referente aos produtores). A ligação desta informação será feita através do NIF do produtor que consta também no *dataset* inicial (como chave primária).

Para dar início ao processo de tratamento dos dados, a verificação da qualidade de cada variável foi essencial. Ao validar cada uma das colunas a serem utilizadas (Tabela 2 e formulários adicionais), garante-se a consistência da informação, permitindo que todo o trabalho realizado a partir da próxima etapa do CRISP-DM é feito com a máxima veracidade e uniformização da informação disponibilizada.

3.2.2. Conceitos fundamentais

Para que se tenha um conhecimento fundamentado dos dados que serão analisados, torna-se fulcral perceber alguns conceitos. Neste contexto, de seguida, apresentam-se as definições dos tipos de resíduos principais que serão analisados ao longo deste estudo.

- Compósitos (código LER 150105): Referem-se a materiais compostos por dois ou mais constituintes com propriedades físicas ou químicas distintas, que permanecem separados e distintos na estrutura final. Este tipo de resíduo é caracterizado por ser difícil de reciclar devido à sua composição heterogénea, o que impede a separação fácil dos seus componentes para reutilização ou reciclagem [41].
- Embalagens com resíduos ou contaminadas (código LER 150110): Estes resíduos incluem embalagens que, após o uso, retêm restos do produto ou são contaminadas por substâncias perigosas. Estes resíduos são particularmente desafiantes para a gestão de resíduos, uma vez que as embalagens contaminadas requerem um tratamento especial para evitar a libertação de substâncias tóxicas no meio ambiente [42].
- Misturas (código LER 150106): Referem-se a resíduos compostos por uma combinação de diferentes materiais ou substâncias que não foram separadas ou classificadas em tipos específicos de resíduos. Estes resíduos podem incluir materiais heterogéneos que, devido à sua composição variada, requerem processos de tratamento para garantir a correta gestão e minimização de impactos ambientais [43].

3.3. Data Preparation

3.3.1. Dataset Inicial

Após serem analisadas as variáveis disponibilizadas, segue-se a fase do Data Preparation que é crucial para o desenrolar de todo o estudo, uma vez que a limpeza e o tratamento dos dados fornecidos pela APA podem influenciar significativamente os resultados finais. Numa fase inicial, ao proceder à leitura do CSV principal, foi realizada uma filtragem específica da variável ler_codigo para incluir apenas os tipos de resíduos solicitados pela APA para análise. Os códigos selecionados para essa filtragem foram os seguintes:

- 150105 – Compósitos;
- 150106 – Misturas;
- 150110 - Embalagens com resíduos ou contaminadas.

Após a aplicação desta filtragem, o *dataset* resultante ficou com um total de 363 733 observações, mantendo as 34 colunas originais. Após removeram-se as colunas referidas na secção anterior (Tabela 2), ficaram para análise as colunas presentes na Tabela 3, perfazendo um total de 16 colunas face às 34 iniciais.

Tabela 3 - Variáveis finais para análise

Variáveis para análise	
data_criacao	ler_codigo
produtor_nif	tipo_op_codigo
produtor_nome	quantidade
estado	tipo_op_codigo_recebido
data_inicio_transporte	quantidade_recebida
data_fim_transporte	ler_recebido_codigo
destinatario_nif	autorizada
destinatario_nome	mercadorias_perigosas

O passo seguinte, consiste na verificação de valores omissos, apresentando-se os seguintes valores na tabela representada abaixo (Tabela 4). Com base nos valores apresentados foi necessário realizar alguns procedimentos, nomeadamente eliminar algumas linhas que não devem conter valores NAs, e que não fazem sentido substituir por qualquer outro valor, como é o caso da variável produtor_nif, uma vez que serve de chave para outras tabelas, a variável tipo_op_codigo, para que se tenha conhecimento das operações realizadas com os resíduos e não enviesar os resultados aquando das análises efetuadas e, por último, a variável quantidade (referente à quantidade enviada pelos produtores). Nestes casos, considerando que são poucos registos procedeu-se à sua remoção.

Outras variáveis, como é o caso da quantidade_recebida (pelos transformadores), será feita a substituição dos seus NAs pelos valores que encontrar em cada linha pela variável quantidade. Isto porque são muitos os registos sem valor e em muitos casos (tal como é expectável), a quantidade enviada pelo produtor deverá ser exatamente a mesma que a quantidade recebida pelo transformador

(partindo do pressuposto que sempre que não foi registado o valor recebido, é porque assumem o mesmo valor da quantidade enviada). O mesmo acontece para o ler_recebido_codigo e tipo_op_recebido_codigo face às variáveis ler_codigo e tipo_op_codigo, respetivamente. Em relação aos restantes valores omissos, variáveis como data_incio_transporte, data_fim_transporte e autorizada, quando utilizadas serão analisadas de forma isolada, pelo que se optou por não remover do *dataframe* para não reduzir de forma tão drástica a dimensão dos dados, sendo apenas filtrado nessas mesmas análises para o efeito. Os restantes valores omissos para variáveis como o nome dos produtores, o nome e o NIF dos destinatários, uma vez que não têm qualquer impacto nas análises realizadas não se aplicou nenhum tipo de tratamento.

Tabela 4 - Números Omissos por variável

Nome da variável	Nº de NAs
data_criacao	0
produtor_nif	7
produtor_nome	25
estado	0
data_inicio_transporte	51
data_fim_transporte	19500
destinatario_nif	21
destinatario_nome	22
ler_codigo	0
tipo_op_codigo	48
quantidade	198
tipo_op_recebido_codigo	19502
quantidade_recebida	19369
ler_recebido_codigo	19502
autorizada	642
mercadorias_perigosas	0

Tendo em consideração que a filtragem dos tipos de resíduos feita inicialmente aplicou-se apenas pela variável ler_codigo, foi necessário verificar se existiam valores diferentes para além dos que estavam a ser analisados (150105, 150106,150110) na variável ler_recebido_codigo. Assim, verificou-se que apenas 1,6% (5794 linhas) apresentavam valores diferentes do tipo de resíduo classificado pelo produtor, e, por essa razão, decidiu-se filtrar a variável ler_recebido_codigo também pelos tipos de resíduos referidos anteriormente.

Adicionalmente, procedeu-se à substituição dos códigos presentes nas colunas ler_codigo e ler_recebido_codigo por descrições de texto que correspondem aos tipos de resíduos. Os códigos 150105, 150106 e 150110 foram substituídos por "Compósitos", "Misturas", e "Embalagens c/

resíduos ou contaminadas", respetivamente. Esta substituição foi realizada para facilitar a interpretação dos dados durante as análises subsequentes, permitindo uma compreensão mais intuitiva das categorias de resíduos analisadas.

É ainda de referir que desde o *dataset* incial que continha 16 123 273 registos, após a filtragem pelos códigos LER, ficou-se com 2,3% dos registos originais (363 733 observações). Após todo o tratamento mencionado totalizaram-se 357 702 observações, tendo sido descartadas 1,66% das observações que se tinha após a filtragem. Assim, do dataset orginal fornecido pela APA, estão para análise 2,22% dos dados originais.

A partir daqui, o foco foi identificar quais das colunas disponíveis poderiam ser utilizadas como identificadores de ligação (chave primária) para os formulários adicionais, de modo a extrair informações relevantes, como o Código de Atividade Económica e o Distrito. Constatou-se que o NIF do produtor (produtor_nif) seria a variável mais adequada para esta finalidade, por ser a única variável presente em todas as tabelas.

3.3.2. Tratamento e Junção de variáveis dos formulários

Para além do CSV principal, trataram-se também dos formulários (tal como descrito na secção de Data Understanding). Para tal, foi necessário combinar e integrar esses formulários ao *dataframe* principal, de modo a enriquecer as análises futuras. Inicialmente, carregaram-se os formulários de forma individual e de seguida, procedeu-se à concatenação destes, originando um único *dataframe*. Para garantir a consistência deste, mantiveram-se apenas as colunas comuns a todos os formulários durante o processo de concatenação.

Para efetuar esta extração, utilizou-se o NIF da organização correspondente ao produtor_nif do *dataset* principal e as colunas do CAE (com o descritivo da indústria pertencente), código_cae e distrito. Posteriormente, foi necessário garantir que apenas seriam extraídas combinações diferentes de um NIF com determinados CAEs associados para evitar duplicados. Após este procedimento, o *dataframe* ficou com um total de 64753 registos. De seguida exportou-se o excel CAES_Distritos para que fosse importado no *notebook* principal. A partir desta fase, efetuou-se a junção entre o dataframe principal com este novo excel, utilizando o NIF do produtor, presente tanto no CSV original (produtor_nif) como nos formulários (nif_organizacao), como chave primária. Desta forma, adicionaram-se as informações de CAE e distrito, ampliando a capacidade de análise com base em critérios económicos e geográficos. Aquando da utilização da variável Distrito removeram-se os NAs desta para não enviesar os resultados.

Depois de ter sido feita a junção, verificou-se que algumas entradas não tinham correspondência nas colunas adicionadas (cae e codigo_cae) uma vez que nem todos os NIFs constavam nos formulários disponibilizados. Para esses casos, os valores ausentes (78968 registos) foram substituídos por

'Inexistente', garantindo assim que as análises posteriores pudessem ser realizadas sem qualquer tipo de problemas em relação a valores omissos.

3.3.3. Preparação dos Dados para a Criação de Modelos

Para caracterizar de forma adequada os produtores e transformadores de resíduos, foi necessário realizar uma reorganização dos dados. Este processo envolveu a consolidação das informações de cada produtor e transformador, de modo a que cada linha representasse uma entidade única, com todos os atributos relevantes agregados.

Em primeiro lugar, agruparam-se os dados por produtor, com o objetivo de centralizar numa única linha todas as informações referentes a cada um deles. Para isso, foi criada uma função específica que processa as colunas categóricas, como os códigos CAE e o Código LER, associadas a cada produtor. Esta função permite converter os dados numa lista de valores únicos, removendo duplicados, ordenando todos os valores e, numa fase final, juntá-los numa única *string*. Além disso, foi extraído o ano através da coluna data_criacao para que as quantidades de resíduos pudessem ser analisadas anualmente, somando as quantidades para cada ano e adicionando-as como novas colunas. Para além disso, converteu-se a variável mercadorias perigosas em booleana para garantir que assumia os valores desejáveis (True or False). Após estes passos, obteve-se, para cada produtor, a quantidade total de resíduos gerados, os códigos CAE e os códigos LER únicos associados, e a indicação se lidava ou não com mercadorias perigosas. A abordagem das quantidades por ano foi também aplicada aos transformadores, permitindo uma análise detalhada ao longo dos anos.

O mesmo processo foi aplicado para os transformadores de resíduos. Agruparam-se os dados por transformador, juntando as informações de cada um. As quantidades de resíduos recebidas por cada transformador foram somadas, e os códigos CAE e os Código LER associados foram processados numa numa lista de valores únicos, e a variável mercadorias perigosas transformada. Esta tabela tem exatamente a mesma estrutura que a referida nos produtores, garantindo coerência aquando da criação dos modelos.

Após a consolidação dos dados em dois *dataframes* distintos para produtores e transformadores, realizou-se um tratamento adicional para garantir que a implementação dos modelos (*clustering*) fosse feita de forma consistente. Este tratamento envolveu a preparação e transformação das variáveis contidas em ambos, assegurando que cada coluna estaria formatada de maneira apropriada para o efeito.

Primeiramente, converteu-se a variável categórica codigos_ler_associados, em dummy gerando novas colunas para cada código LER (em formato binário). De seguida, procedeu-se à normalização da variável quantidade e quantidade recebida, para os produtores e transformadores, respetivamente.

Estas normalizações foram calculadas através da média e do desvio padrão destas colunas. Para tal, utilizou-se o Z-Score que facilitou a identificação e o tratamento de *outliers*. Após a normalização, geraram-se *boxplots* para identificar *outliers* (Figura 4). Com base nisso, filtraram-se os valores extremamente elevados, com o objetivo de garantir que os resultados dos modelos não ficariam distorcidos.

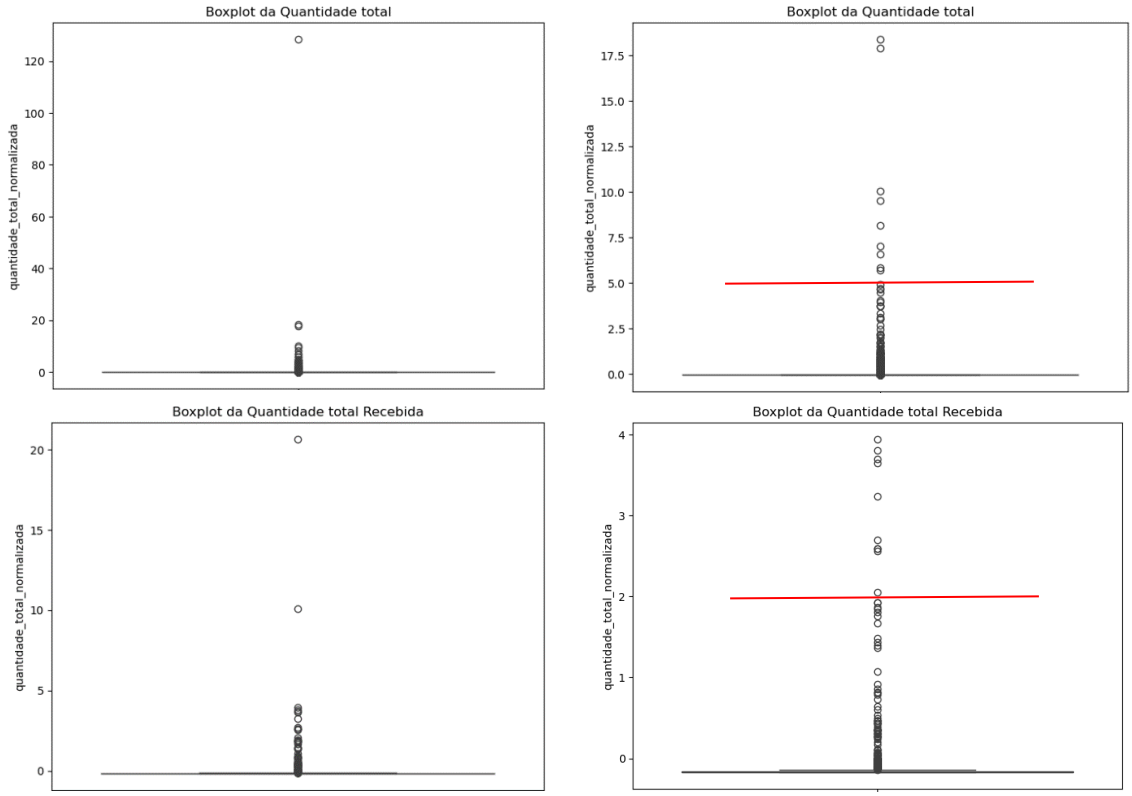


Figura 4 - Boxplots para a remoção de outliers

Os gráficos à esquerda representam a fase inicial, antes da remoção dos outliers, sendo visível a existência de valores muito afastados face à maioria. Após essa etapa inicial, efetuou-se um corte intermediário, resultando nos gráficos à direita. Os gráficos referentes à "Quantidade Total" correspondem aos dados dos produtores, enquanto os gráficos de "Quantidade Recebida" referem-se aos dados dos transformadores. Nos gráficos à direita, as linhas vermelhas indicam os pontos em que foram aplicados os cortes finais nos *dataframes*. Para a quantidade enviada pelos produtores, o corte foi feito no ponto 5, enquanto para a quantidade recebida pelos transformadores, o corte foi realizado no ponto 2. Esses valores foram escolhidos após várias tentativas, considerando os resultados apresentados no próximo capítulo de modelação, onde se observou que esses cortes proporcionavam melhores resultados aos modelos implementados. Numa fase posterior, uma vez que se incluiu a variável caes_associados, aplicou-se o mesmo processo de criação de variáveis dummy. Todo este tratamento foi também aplicado a ambos os *dataframes*.

3.3.4. Junção dos dados referentes aos plásticos

Com o intuito de compreender melhor a relação entre possíveis perdas de quantidades e anomalias associadas ao tratamento de resíduos plásticos, foi realizada uma filtragem específica dos dados na base de dados original. Nesta fase, foram selecionados os registos correspondentes aos códigos LER associados aos plásticos, em particular o código 150102, referente a embalagens de plástico. Adicionalmente, foram incluídos os três Códigos Ler como referência: 150105 (Compósitos), 150110 (Embalagens com resíduos ou contaminadas) e 150106 (Misturas). O mesmo filtro foi aplicado à variável ler_recebido_codigo. Após a exportação do CSV contendo toda esta informação, o processo de tratamento previamente aplicado ao *dataframe* inicial foi replicado neste *dataset* para garantir a consistência e evitar enviesamentos nos resultados (renomeação de atributos, remoção de *outliers* e substituição de valores omissos).

3.3.5. Entrevistas a Produtores e Transformadores

Com o intuito de recolher informações adicionais que pudessem complementar este estudo contactaram-se os produtores e transformadores mais frequentes na base de dados. A proposta de contacto com essas entidades surgiu principalmente da necessidade de entender onde poderá haver possíveis perdas de plásticos durante o transporte de resíduos. Por este motivo, foram selecionados os 10 principais produtores e os 10 principais transformadores que constam na base de dados.

Apesar das tentativas de contacto, este esforço foi em grande parte mal sucedido. Apenas duas empresas responderam, e, entre elas, apenas uma aceitou realizar uma reunião remota. No entanto, as respostas obtidas dessas duas empresas não forneceram informações relevantes ou que fossem ao encontro dos objetivos do estudo. Embora esta abordagem tenha sido limitada em termos de resultados práticos, foi percetível a dificuldade em aceder a dados adicionais que complementassem esta investigação.

CAPÍTULO 4

Modelação e Resultados

Neste capítulo, serão realizadas análises descritivas acerca dos dados previamente tratados, seguidas da aplicação de técnicas de *clustering*, com o intuito de identificar padrões dentro dos grupos de produtores e transformadores criados no capítulo anterior. As análises descritivas têm como finalidade oferecer uma visão abrangente sobre as principais características dos dados, permitindo uma melhor compreensão sobre estes, incluindo a sua distribuição e a correlação entre diferentes variáveis.

Numa fase posterior, será implementado o *clustering*, uma técnica de aprendizagem não supervisionada, que visa agrupar os dados em subconjuntos homogéneos, ou clusters, com base nas características selecionadas. O *clustering* é uma técnica muito utilizada neste âmbito, por ser capaz de segmentar e identificar padrões em grandes volumes de dados. A técnica de *clustering*, em particular, o *K-means*, será explorado, dado a sua eficiência e aplicabilidade no contexto dos dados da APA [44].

Através da modelação, espera-se dar resposta às perguntas de investigação definidas inicialmente, mas também descobrir novas perspetivas que possam contribuir significativamente para a APA. A integração das análises descritivas com o *clustering* permitirá uma exploração completa dos dados, proporcionando *insights* valiosos para a APA e contribuindo para o avanço das práticas de gestão de resíduos em Portugal.

4.1. Análises Descritivas

4.1.1. Caracterização dos dados iniciais

Dado que a variável principal é o ler_codigo que indica o tipo de resíduo, torna-se relevante perceber a distribuição desta variável pelo *dataset*. Assim, através da Figura 5, é notória uma maior presença de registos de resíduos de Misturas. Este facto pode ser explicado pela natureza abrangente desta categoria, que engloba múltiplos tipos de resíduos, resultando assim num maior número de ocorrências. Por outro lado, os compósitos são aqueles que aparecem menos vezes na classificação do código LER.

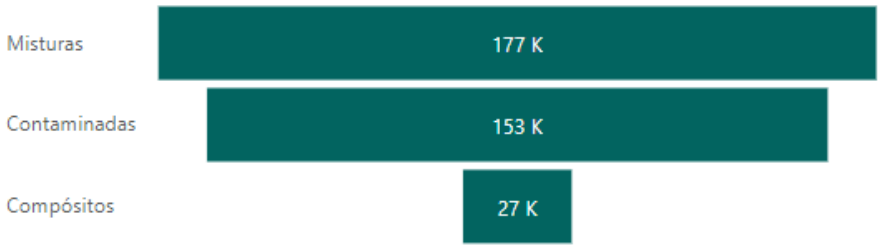


Figura 5 - Número de registos no dataset da variável ler_codigo (Tipo de Resíduo)

Com o objetivo de oferecer uma visão detalhada e contextualizada das informações previamente tratadas, uma das análises iniciais focou-se na caracterização dos estados dos registos, onde os dados foram classificados em diferentes categorias de estado. A Tabela 5 apresentada abaixo resume a distribuição dos estados em que os registos se encontram entre os anos 2019-2022.

Tabela 5 - Caracterização da variável Origem

Variável Estado	Número de Registos
Concluída (certificado de receção)	335198
Anulada	17678
Corrigida	2742
Rejeitada	782
Aceite	399
Emitida	363
Eliminada	327
Guardada	184
Correção Negada	29

A Tabela 5 revela que a grande maioria dos registos (335198 ocorrências) encontram-se no estado "Concluída (certificado de receção)", indicando que estes registos correspondem a processos que foram devidamente finalizados e validados pela APA. Este valor indica um elevado grau de sucesso na conclusão dos processos de transporte de resíduos.

Por outro lado, há um número significativo de registos que foram "Anulados" (17678 ocorrências), o que indica impedimentos na finalização dos processos ou mudanças nas condições iniciais do transporte dos resíduos. Registos "Corrigidos" indicam que alguns processos necessitaram de revisão antes de serem considerados válidos, enquanto as restantes categorias têm frequências muito menores, o que pode ser interpretado como exceções ou situações menos comuns no ciclo de vida do processo de transporte.

Tabela 6 - Relação entre o Tipo de Resíduo e o Estado do transporte

Tipo de Resíduo	Concluída (%)	Anulada (%)	Corrigida (%)	Rejeitada (%)	Outros (%)
Compósitos	92,3	**6,8**	0,3	0,3	0,3
Embalagens c/ resíduos ou contaminadas	95,5	3,6	0,3	0,2	0,4
Misturas	92,4	5,9	**1,2**	0,3	0,3

Nota: Outros (%) inclui percentagens menores de "Aceite", "Correção negada", "Eliminada", "Emitida" e "Guardada".

A partir da Tabela 6, é possível verificar de uma forma clara as diferenças entre os tipos de resíduos face ao estado dos transportes efetuado ao longo dos 4 anos em análise. É evidente que, para todos os tipos de resíduos, a maioria dos transportes foram realizados com sucesso, como era expectável, uma vez que esta categoria apresentou o maior número de ocorrências no geral. Destaca-se o valor

de anulação dos transportes para "Compósitos" que é notavelmente superior, atingindo os 6,8%, indicando que este tipo de transporte enfrentou mais dificuldades ou ajustes durante o processo. É ainda de realçar que para os resíduos de Misturas há uma percentagem relativamente superior aos restantes na categoria corrigida, sugerindo que pode haver uma maior dificuldade em classificar este tipo de resíduos.

No que diz respeito à variável 'mercadorias_perigosas', de um total de 357702 observações, apenas 20104 são classificadas como perigosas. Para averiguar qual o tipo de resíduos que apresenta uma maior taxa de mercadorias perigosas, procedeu-se ao cálculo das percentagens para cada tipo de resíduo, como se segue na Tabela 7.

Tabela 7 - Tipo de resíduos face à perigosidade

Tipo de Resíduo	Não Perigoso (%)	Perigoso (%)
Compósitos	99,9	0,01
Embalagens c/ resíduos ou contaminadas	86,9	13,09
Misturas	99,9	0,01

É notório que a grande maioria dos resíduos de Compósitos e Misturas são classificados como não perigosos, com valores de 99,9% na categoria "Falso" para ambos. No entanto, destaca-se as Embalagens com resíduos ou contaminadas que apresentam uma proporção significativamente maior de resíduos classificadas como perigosos, com 13,09% na categoria Verdadeiro. Este resultado era esperado, dado que as embalagens contaminadas têm uma maior probabilidade de conterem substâncias perigosas, justificando assim a maior proporção de resíduos perigosos nesta categoria.

Adicionalmente, no apêndice b) encontram-se as descrições para cada uma das operações que podem ser realizadas com os resíduos tanto do lado dos produtores como transformadores. No entanto, para que se consiga analisar a Figura 6 e a Figura 7 é importante recorrer à Tabela 8 que indica a nomenclatura atribuída a cada código sobre o Top de operações apresentadas.

Tabela 8 - Top de Operações realizadas pelos Produtores e Transformadores

Código	Operações
R12	Troca de resíduos com vista a submetê-los a uma das operações enumeradas de R1 a R11 – Apêndice b)
R13	Armazenamento de resíduos destinados a uma das operações enumeradas de R1 a R12 (com exclusão do armazenamento temporário, antes da recolha, no local onde os resíduos foram produzidos) - Apêndice b)
D15	Armazenamento antes de uma das operações enumeradas de D1 a D14 (com exclusão do armazenamento temporário, antes da recolha, no local onde os resíduos foram produzidos) - Apêndice b)
R3	Reciclagem/recuperação de substâncias orgânicas não utilizadas como solventes (incluindo digestão anaeróbia e ou compostagem e outros processos de transformação biológica)
R4	Reciclagem/recuperação de metais e compostos metálicos
D1	Depósito no solo, em profundidade ou à superfície (por exemplo, em aterros, etc.)

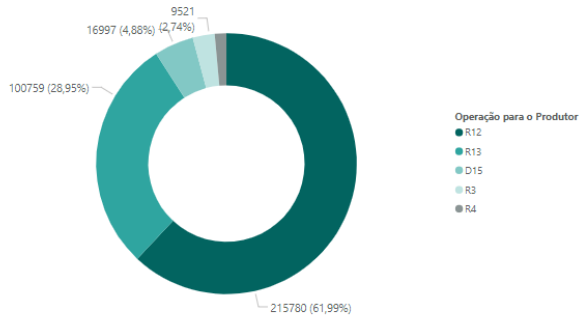


Figura 6 - Top 5 das operações definidas pelo Produtor

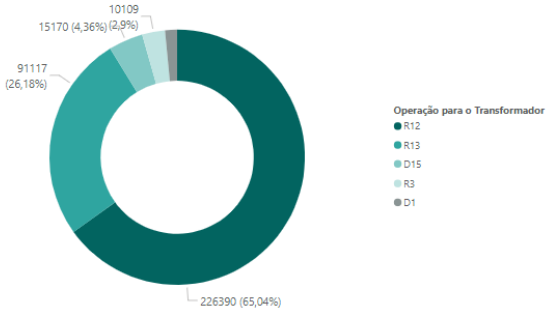


Figura 7- Top 5 das operações definidas pelo Transformador

Ao analisar simultaneamente a Figura 6 e a Figura 7 observa-se a obtenção de resultados muito semelhantes, entre produtores e transformadores. Estes diferenciam-se apenas no valor mais baixo de cada um. Maioritariamente, a operação R12 destaca-se e é referente à transferência de resíduos com o objetivo de serem processados por meio de uma das operações de recuperação ou reciclagem listadas no apêndice b) (R1-R11). A operação R13 e D15 são também referentes a outras operações, tal como referido na Tabela 8 e designadas no apêndice b). Já a R3 diz respeito à reciclagem/ recuperação de substâncias orgânicas não utilizadas como solventes, apresentando uma pequena percentagem face ao total, em ambos os casos. Para os produtores, diferencia-se o último lugar do top 5 com reciclagem/recuperação de metais e compostos metálicos e, no caso dos transformadores, o depósito em solo, em profundidade ou à superfície. Estes resultados não apresentam um resultado evidente das operações que são realizadas uma vez que a grande maioria não está alocada a uma ação específica.

Numa fase posterior, averiguou-se as quantidades enviadas e recebidas pelos respetivos produtores e transformadores (entre 2019-2022) para todas as operações dadas como concluídas. Apesar de serem idênticas apresentam diferenças, algo que não era expectável (Figura 8). Uma vez que se trata de um transporte, era esperado que a quantidade enviada fosse igual à quantidade recebida, indicando a perda de resíduos ou falhas na informação que é registada. Desta forma, procedeu-se ao cálculo das diferenças entre as quantidades enviadas e recebidas (Tabela 9).

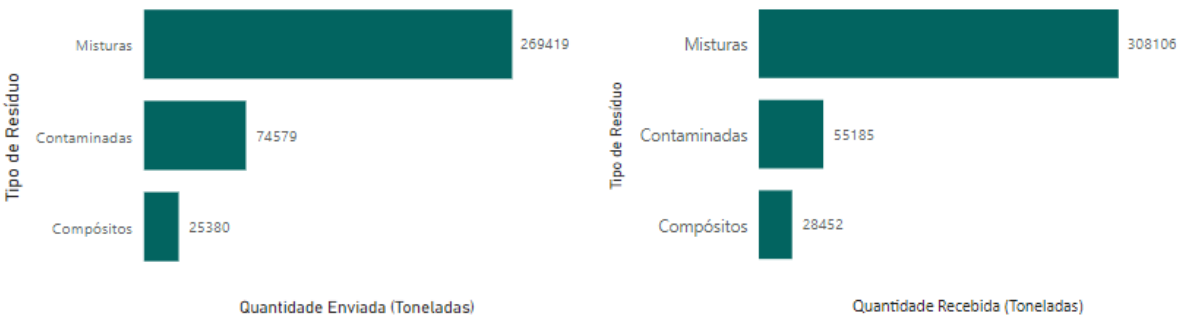


Figura 8 - Tipo de Resíduo por quantidade enviada e recebida, respetivamente entre 2019-2022 (Toneladas) - Concluídas

Tabela 9 - Diferença entre a quantidade enviada e a quantidade recebida (em toneladas) - Concluídas

Tipo de Resíduo	Diferença de Quantidade (T)	Nº de registos diferentes	% de registos diferentes face ao total
Misturas	-38687	146594	89,4%
Embalagens c/ resíduos ou contaminadas	19394	99613	68%
Compósitos	-3072	20477	82,6%

No caso das embalagens com resíduos ou contaminadas, verifica-se que a quantidade recebida pelos transformadores é inferior à quantidade enviada pelos produtores (valor positivo, Tabela 9), sugerindo possíveis perdas no transporte ou falhas no registo das quantidades transportadas.

Por outro lado, para os resíduos classificados como Misturas e Compósitos, a quantidade recebida é maior do que a enviada (valor negativo, Tabela 9), o que pode indicar uma eventual subestimação inicial ou adição de resíduos durante o transporte e que não tenha sido feito o registo. Este cálculo foi feito apenas para as operações dadas como concluídas. Grande parte dos registos têm valores diferentes entre as quantidades enviadas e recebidas dos resíduos, como se verifica nas duas últimas colunas da Tabela 9.

Destacam-se os compósitos por apresentarem uma percentagem menor face aos restantes, apesar desta ainda ser significativamente alta. Para validar se existiriam diferenças nos transportes que são considerados corrigidos, efetuou-se uma comparação que permita verificar a distribuição das diferenças de quantidades (enviadas e recebidas) face aos transportes concluídos.

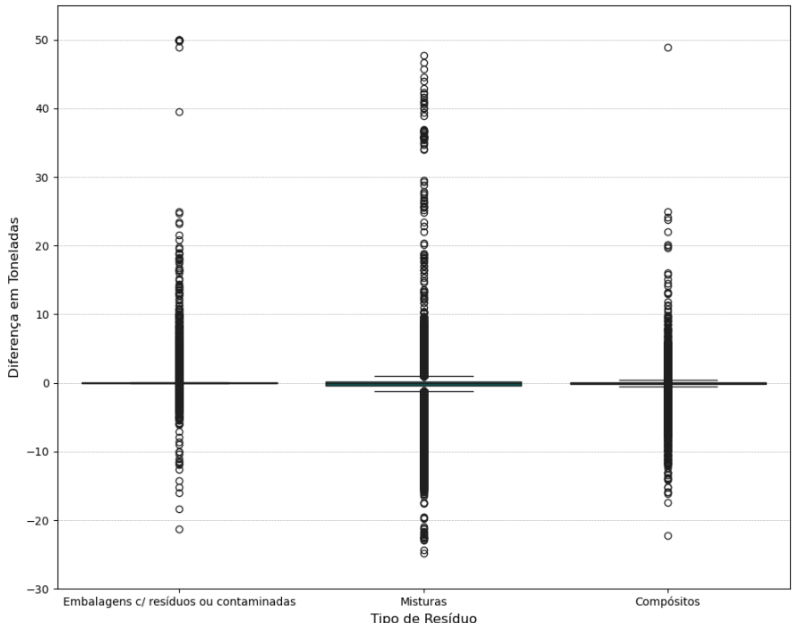


Figura 9 – Boxplots das diferenças de quantidades enviadas e recebidas por tipo de resíduo - Concluídas

Ao analisar a Figura 9, valida-se os resultados demonstrados na Tabela 9, uma vez que existe uma grande distribuição em relação à diferença de quantidades enviadas e recebidas (em toneladas) ao longo dos 4 anos para os três tipos de resíduos para as operações dadas como concluídas.

Através da Figura 10, que representa as diferenças de quantidades enviadas e recebidas dos transportes considerados corrigidos para cada tipo de resíduo, verifica-se que há uma menor distribuição destas diferenças de quantidades, revelando que as correções feitas têm um efeito significativo nestes resultados. Ainda assim, existem diferenças nas quantidades, mas numa escala muito menor. Os resíduos de Mistura destacam-se por serem os que, mesmo após serem corrigidos apresentam maiores discrepâncias nas diferenças de quantidades (Figura 10).

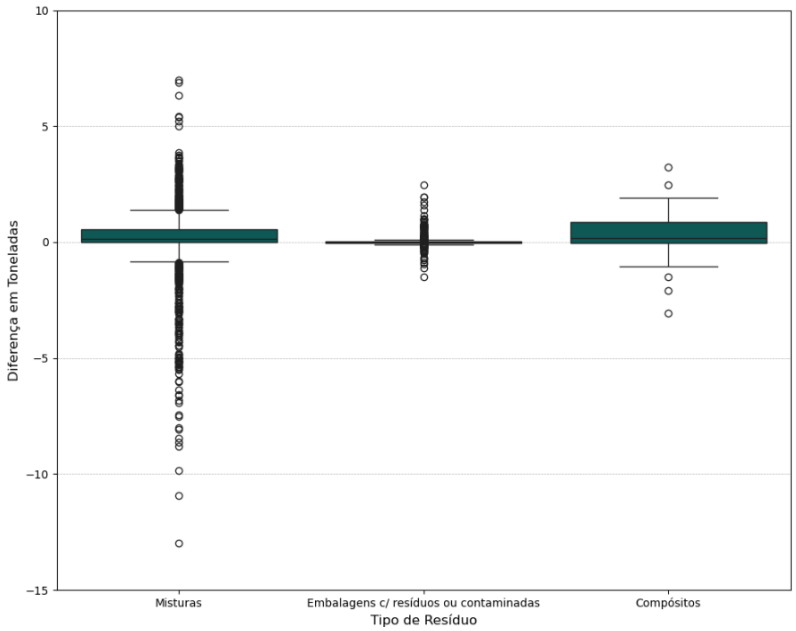


Figura 10 - Boxplots das diferenças de quantidades enviadas e recebidas por tipo de resíduo - Corrigidas

Para além disto, conforme mencionado na secção de Data Preparation (3.3.5), foram contactados os Top 10 de Produtores e Transformadores que mais aparecem na base de dados no período de 2019-2022, com o objetivo de obter mais informações sobre a entrada e saída desses tipos de resíduos e ainda averiguar a possível perda de plásticos durante os transportes. No entanto, essa tentativa foi, em grande parte infrutífera devido à falta de resposta por parte das empresas.

Adicionalmente, analisou-se se haveria empresas em comum por tipo de resíduo entre produtores e transformadores. Embora tenham sido identificados alguns casos em que certas empresas são destacadas como produtoras ou transformadoras para mais de um tipo de resíduo (em relação ao Top 5), essa não é uma ocorrência frequente. No entanto, verificou-se em toda a base de dados que, em 334 casos, as mesmas entidades aparecem tanto como produtoras quanto como destinatárias, o que indica que essas empresas operam em ambos os papéis. Para que haja um maior contexto, procuraram-se exemplos que verifiquem esta situação. A empresa responsável pela gestão do sistema

intermunicipal de resíduos de um distrito português, desempenha simultaneamente dois papéis: atua como produtora de resíduos, no âmbito dos serviços prestados aos municípios, e como transformadora, ao tratar esses resíduos através de processos de valorização e reciclagem [45]. Outro exemplo, são as empresas do setor automóvel, que geram resíduos ao longo do processo de produção. Essas mesmas empresas, em vez de descartarem esses resíduos, algumas delas têm operações de reciclagem internas para transformar esses materiais em matéria-prima para novos produtos [46].

A nível mais geral, no que diz respeito às variáveis que constam no *dataset*, analisou-se ainda o tempo médio dos transportes para cada um dos três tipos de resíduos em análise (Figura 11).

Para esta análise específica, filtrou-se a base de dados, removendo os 19500 registos que continham valores NAs na variável data_fim_transporte, conforme mencionado na secção de Data Preparation (3.3.1). Após este tratamento, observou-se que o tempo médio de transporte para os diferentes tipos de resíduos não apresenta discrepâncias significativas (Figura 11).

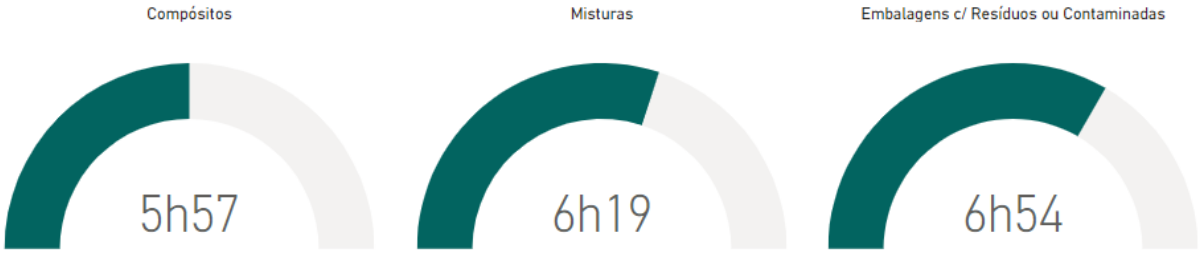


Figura 11 - Tempo médio de transporte para cada tipo de resíduo

Através da Figura 12, verificam-se oscilações grandes nos tempos de transporte. Isto pode indicar que algumas vezes o transporte é feito de forma quase imediata (em poucas horas), ou por outro lado, pode demorar algum tempo até começar ou ser registado o seu término. Para além disso, possivelmente transportes para outras zonas do país podem demorar mais tempo face aos que são produzidos e transformados no mesmo distrito justificando os tempos mais elevados, por exemplo.

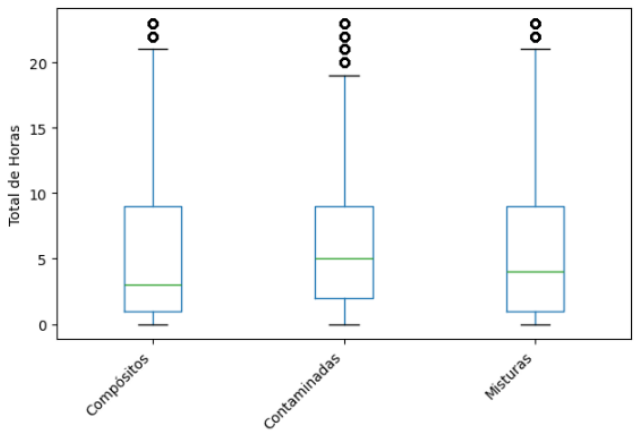


Figura 12 - Distribuição do tempo de transporte por tipo de resíduo

4.1.2. Análise Temporal

Através de uma análise temporal é possível entender as tendências ao longo dos anos e dos meses dentro do período estudado. Compreender essas variações ao longo do tempo pode fornecer informações sobre a evolução dos fluxos de resíduos, ajudando a identificar padrões, sazonalidades, e mudanças significativas no âmbito da gestão de resíduos.

Em relação à quantidade de resíduos enviada pelos produtores entre 2019 e 2022, a partir da Figura 13, é possível ter uma visão geral do que acontece para os três tipos de resíduos. Contudo, por haver valores muito diferentes entre estes três tipos de resíduos a Tabela 10 permite clarificar qual foi o comportamento de cada um para os diferentes anos.

Destacam-se as Misturas, que apresentaram a maior quantidade enviada, algo que já era expectável, considerando que é o tipo de resíduo com maior número de ocorrências. No entanto, registaram um decréscimo de 15,2% em 2020 face a 2019. As Embalagens com resíduos ou contaminadas mostraram um crescimento em 2021 de 42,4% face ao ano anterior (Tabela 10), fator que pode ser explicado devido à pandemia que surgiu neste intervalo de tempo, onde foram produzidos mais 30% dos resíduos face a 2020, considerados de risco biológico [47]. Já os compósitos, tiveram um ligeiro aumento em 2021 de 16,5% face a 2020 (Tabela 10).

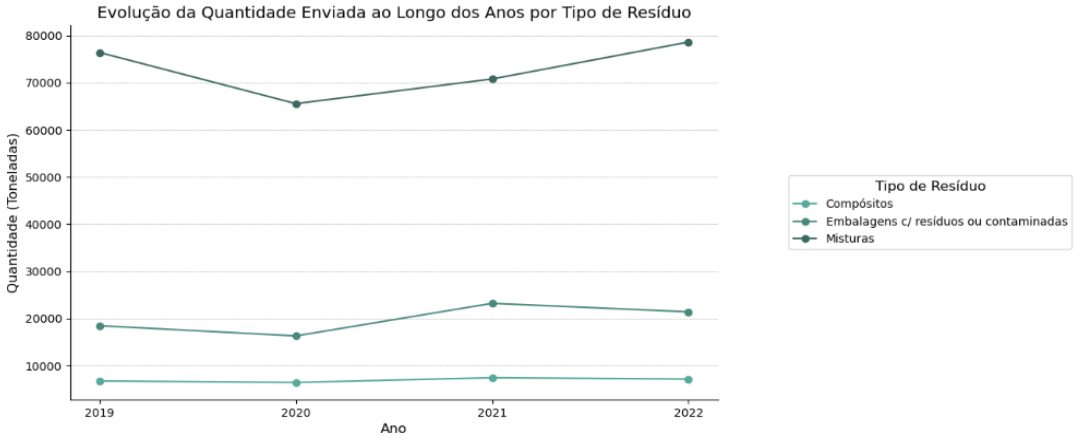


Figura 13 - Quantidade enviada pelos produtores entre 2019 e 2022 (Toneladas)

Tabela 10 - Crescimento (em %) das quantidades enviadas face ao ano anterior

Tipo de Resíduo	2020 (%)	2021 (%)	2022 (%)
Compósitos	-5,08	**16,48**	-4,71
Embalagens c/ resíduos ou contaminadas	-4,36	**42,43**	-8,63
Misturas	**-15,21**	8,54	4,14

Com o intuito de perceber qual o comportamento temporal dos resíduos recebidos pelos transformadores ao longo dos anos em análise, através da figura e da tabela abaixo (Figura 14 e Tabela 11), observa-se uma tendência de crescimento, para o ano de 2021. Assim como nas quantidades

enviadas, as Misturas demonstraram um aumento contínuo a partir desse ano, culminando em 2022 com a maior quantidade recebida registada (Figura 14). Apesar de não se verificar no gráfico, a partir da Tabela 11, destaca-se um aumento em 2021 de 18,8% face ao ano anterior para os resíduos Compósitos.

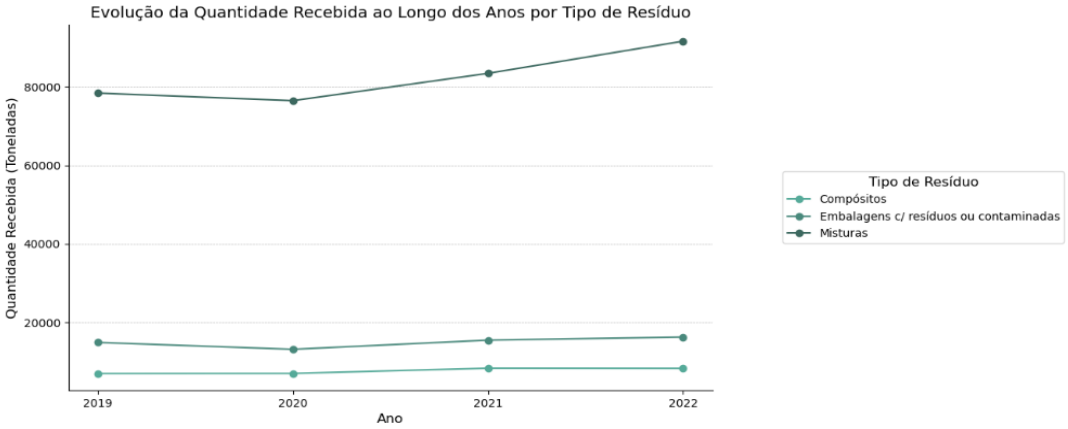


Figura 14 - Quantidade recebida pelos transformadores entre 2019 e 2022

Tabela 11 - Crescimento (em %) das quantidades recebidas face ao ano anterior

Tipo de Resíduo	2020 (%)	2021 (%)	2022 (%)
Compósitos	0,36	**18,79**	-0,44
Embalagens c/ resíduos ou contaminadas	-11,65	17,63	5,05
Misturas	-2,47	9,09	9,81

Tendo em consideração que as análises ao longo dos anos são genéricas, procedeu-se a uma análise mensal para verificar se existe alguma sazonalidade nas quantidades de resíduos recebidas e enviadas (Figura 15 e Figura 16).

No que diz respeito às quantidades enviadas pelos produtores (Figura 15), observa-se que os resíduos Compósitos mantêm um comportamento bastante consistente ao longo dos meses nos quatro anos analisados. Já os outros dois tipos de resíduos apresentam variações mais acentuadas ao longo do tempo. Relativamente às Embalagens com resíduos ou contaminadas, há uma tendência de aumento das quantidades enviadas antes ou após os meses de verão, como observado em 2019, 2021 e, de forma mais subtil, em 2022. Em 2020, considera-se a exceção onde se registou um aumento significativo durante o mês de agosto.

Por outro lado, os resíduos classificados como Misturas, que apresentam valores consistentemente superiores em comparação com os outros tipos de resíduos, como mencionado anteriormente, exibem várias oscilações ao longo dos meses, sem um padrão claro. No entanto,

destacam-se quedas ligeiras para todos os resíduos em abril de 2020, coincidindo com o início da pandemia de COVID-19.

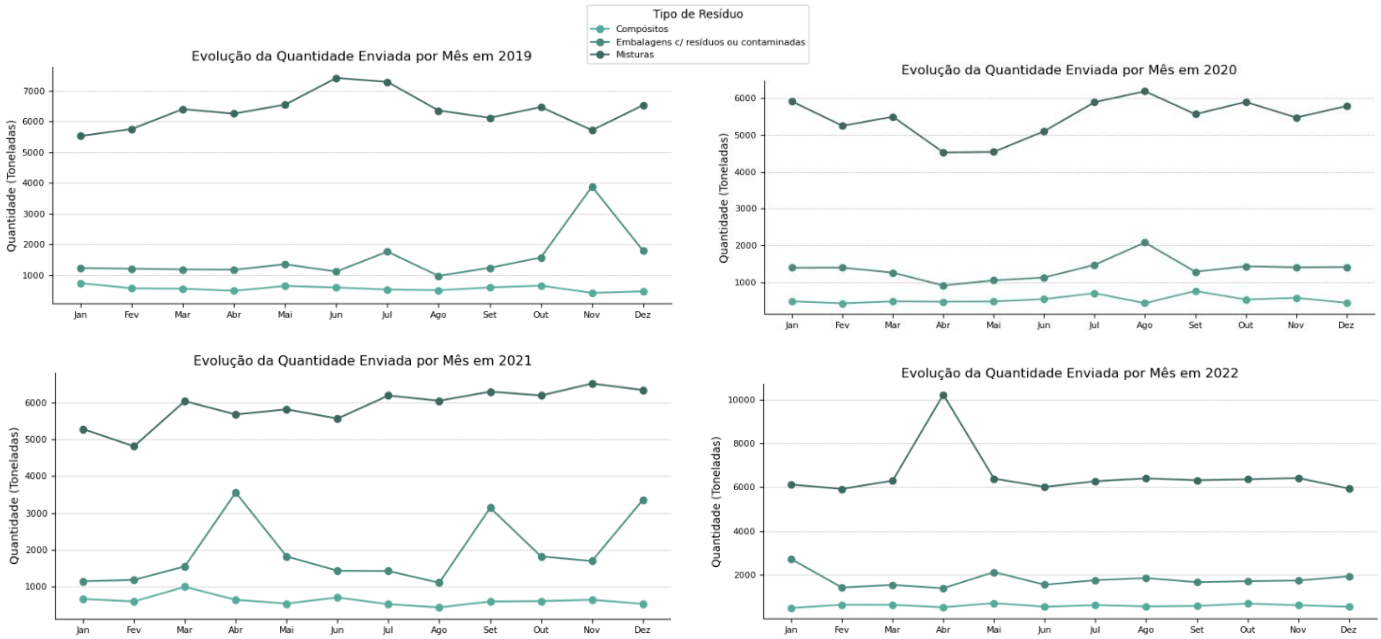


Figura 15 - Quantidade enviada por mês pelos produtores entre 2019 e 2022

Em relação às quantidades recebidas pelos transformadores (Figura 16), são notórias muitas semelhanças nos valores, o que faz sentido, tendo em conta que o expectável seria que tudo o que fosse enviado seria recebido também. Contudo, ao analisar os gráficos representados abaixo, é visível que algumas diferenças, vão ao encontro do descrito na análise das diferenças de quantidades referidas anteriormente. Por exemplo, em 2019, as Misturas têm alguns valores diferentes e com uma diferença significativa face ao que foi enviado pelos produtores (Figura 15), nomeadamente, no mês de junho, sendo um pico registado pelos produtores e uma queda para os transformadores. O mesmo acontece durante o ano de 2020 em que para o mesmo tipo de resíduo, a quantidade enviada é sempre menor do que quantidade recebida. Outro exemplo pertinente de analisar, é relativo às Embalagens com resíduos ou contaminadas que em agosto de 2020, apresentam uma subida ligeira na quantidade enviada (Figura 15), que em contrapartida, tem uma descida no valor referente à quantidade recebida (Figura 16). Apesar das diversas tentativas, não foi possível retomar o contacto com a APA, numa fase crítica deste estudo, que permitiria esclarecer a razão destas descontinuidades.

Comparando as duas figuras em 2021, os Compósitos são o único tipo de resíduo que mantém valores de envio e receção na mesma ordem. Destacam-se os picos de quantidades muito superiores enviadas em abril, setembro e dezembro de Embalagens com resíduos ou contaminadas face às quantidades recebidas, indicando falhas na receção destes resíduos ou no registo das quantidades. Em relação às Misturas, para o último ano em análise, em 2022, é claramente visível uma grande

discrepância no mês de setembro, tendo em conta que a quantidade total recebida atinge as 10 000 toneladas, enquanto a quantidade enviada mantém-se constante face ao mês anterior.

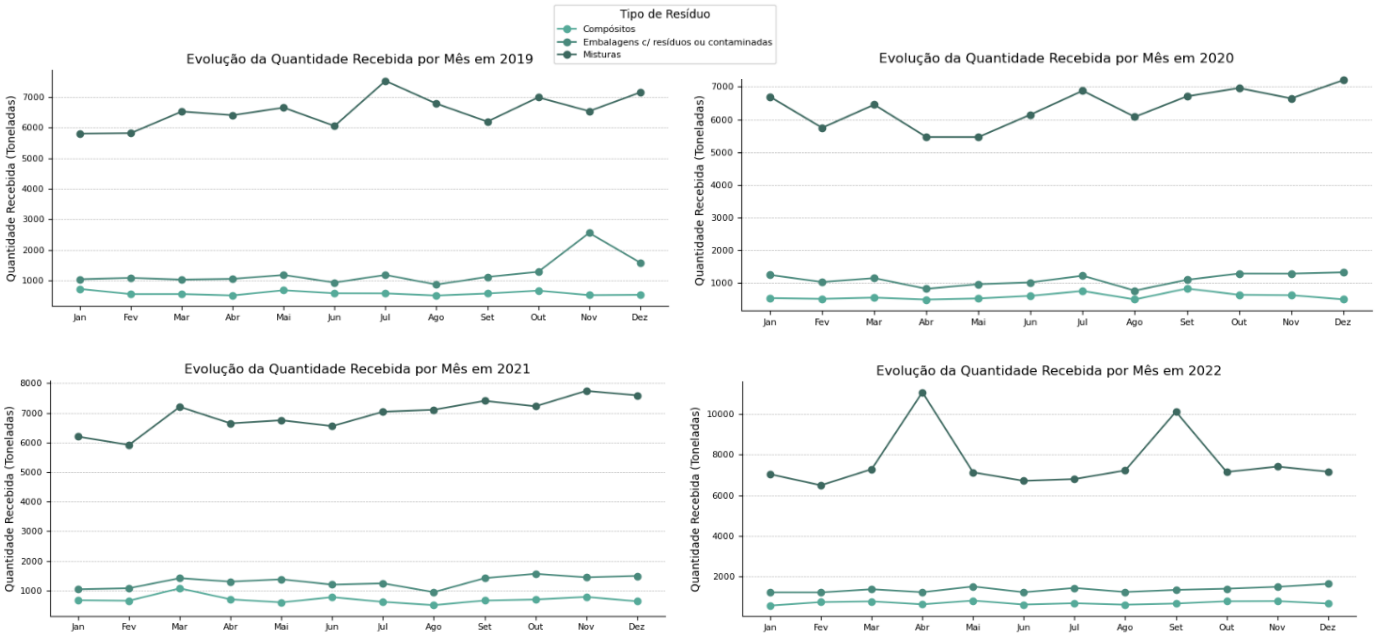


Figura 16 - Quantidade recebida por mês pelos transformadores entre 2019 e 2022

Com base nestas análises, é percetível a existência de falhas nos registos, na medida em que os valores deveriam coincidir. Para além desse facto, estas conclusões vão ao encontro das análises já mencionadas anteriormente (diferenças nas quantidades enviadas vs. quantidades recebidas), corroborando a sua veracidade.

4.1.3. Análise por distrito

Após a extração das informações referentes aos distritos dos produtores a partir dos formulários fornecidos pela APA, foi realizada uma análise detalhada para identificar os distritos que mais e menos quantidades de resíduos enviaram ao longo dos anos. Na Tabela 12, são apresentadas as quantidades totais de resíduos enviadas pelos diferentes distritos entre 2019 e 2022, organizadas de forma decrescente. Em apêndice c) pode-se consultar o detalhe da quantidade enviada por cada distrito em cada ano individualmente. Para a Figura 17, adquiriram-se as coordenadas de cada distrito [48], e através do Power BI procedeu-se à construção de um *heatmap*, representando graficamente os resultados obtidos na Tabela 12.

A Tabela 12 e a Figura 17 revelam claramente a disparidade nas quantidades de resíduos enviadas pelos distritos de Portugal. O Porto destaca-se com uma quantidade significativamente superior (116 858 toneladas), mais de 50% do que o distrito de Lisboa (75 731 toneladas), que ocupa o segundo

lugar. O mapa de calor reforça essa observação, evidenciando uma mancha mais intensa na região do Porto.

Tabela 12 - Quantidades enviadas entre 2019 e 2022 por Distrito

Distrito	Quantidade (Toneladas)
Porto	116858
Lisboa	75731,7
Setúbal	52593,3
Aveiro	26624,2
Leiria	24477,7
Santarém	23519,9
Braga	15683,7
Faro	8460,3
Coimbra	6355,1
Castelo Branco	6050,3
Viseu	6030,1
Ilha da Madeira (Madeira)	2960,9
Beja	2925,3
Viana do Castelo	2158,8
Portalegre	2088,7
Évora	1707,2
Vila Real	891,7
Bragança	732,7
Guarda	694
Ilha de Porto Santo (Madeira)	0,9

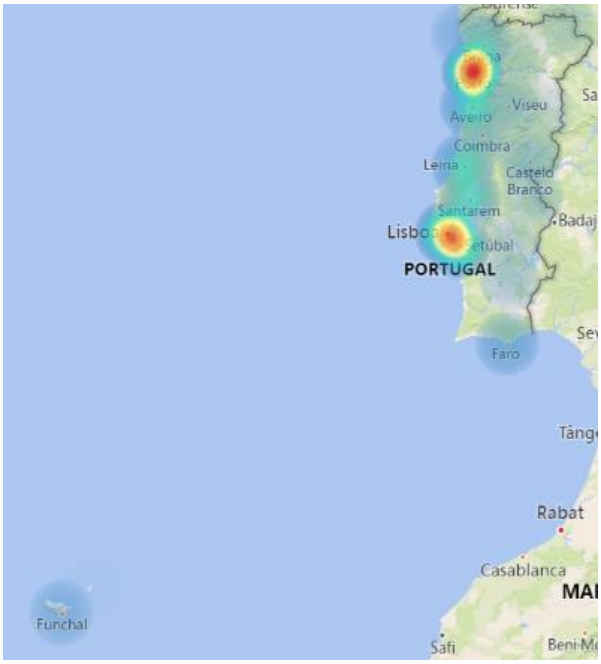


Figura 17 - Mapa de calor com as quantidades enviadas pelos produtores nos distritos portugueses (2019-2022)

Adicionalmente, é visível que os distritos localizados no litoral, como Setúbal, Aveiro, Leiria, Santarém, e Braga, também apresentam quantidades expressivas, refletidas tanto na tabela quanto no mapa. As regiões do interior, por outro lado, têm quantidades visivelmente menores, como por exemplo Guarda e Bragança, que apresentam valores muito baixos de resíduos enviados. A concentração maior de resíduos no litoral, visível pelas manchas intensas em torno do Porto, Lisboa, e Setúbal, é uma tendência clara que o mapa ajuda a corroborar.

Por falta de informação nos dados disponibilizados pela APA, não é possível fazer esta análise para os transformadores.

4.1.4. Análise da presença de plásticos

Tendo em consideração que a validação de presença de plásticos noutro tipos de resíduos é uma informação relevante para a APA, procedeu-se à análise deste tipo de resíduos face aos três que serviram de base para este estudo.

O mapa de calor representado abaixo (Figura 18) revela o peso dos resíduos classificados pelos produtores e os respetivos destinatários. É ainda de referir que os valores mais realçados a verde na diagonal representam os registos que foram corretamente classificados, isto é, enviados pelos produtores e recebidos e classificados da mesma forma pelos destinatários. A grande maioria dos

resíduos parece estar corretamente classificado, uma vez que os valores à volta da diagonal são quase todos zero.

No entanto, existem discrepâncias que devem ser tidas em conta, indicadas pelas áreas destacadas a vermelho na Figura 18. Esses dois pontos indicam que alguns resíduos foram classificados de maneira inconsistente entre produtores e destinatários. O quadrado com 0.1 mostra que alguns resíduos classificados como Misturas pelo produtor foram classificados à chegadas dos destinatários como plásticos. O mesmo acontece para produtores que enviam resíduos classificados como plásticos e são classificados pelos destinatários como Misturas. Essas inconsistências podem indicar problemas na categorização dos resíduos ou falhas durante o transporte, o que pode gerar erros na gestão de resíduos.

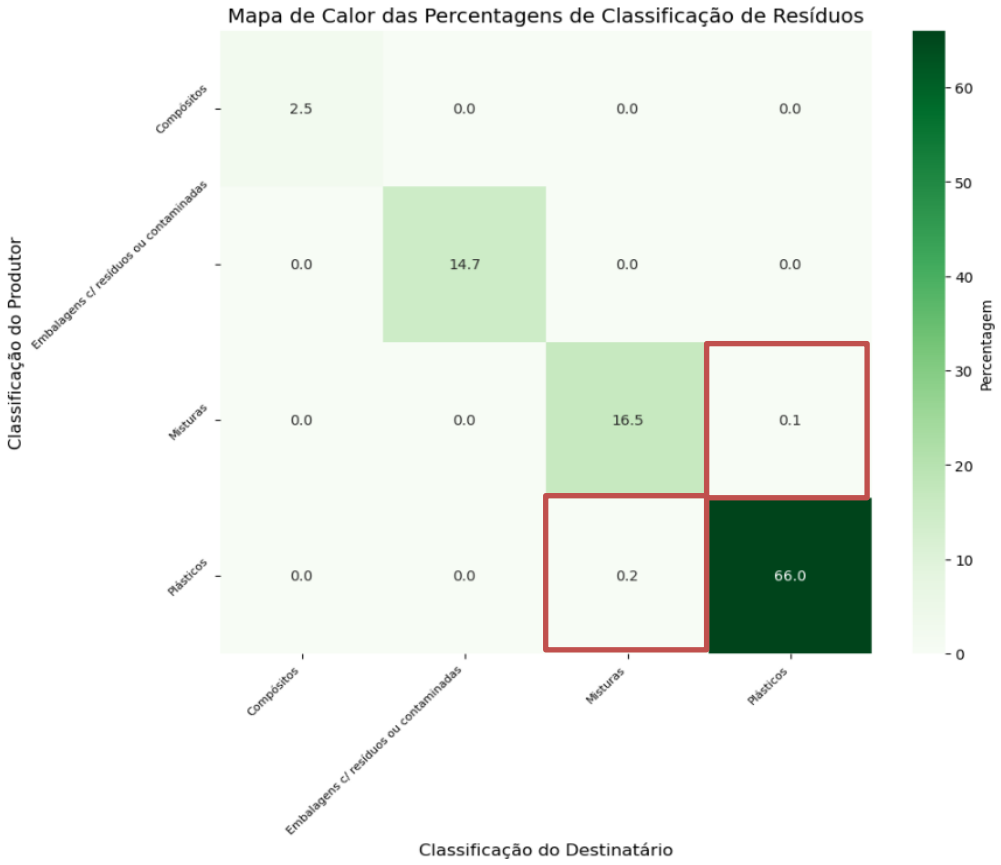


Figura 18 - Mapa de calor com a classificação do tipo de resíduos entre 2019-2022

4.1.5. Código de Atividade Económica (CAE)

Um Código de Atividade Económica é uma classificação oficial que serve para identificar e categorizar as atividades desempenhadas por empresas e organizações. Esta categorização permite não só a análise estatística do mercado empresarial, como também facilita a aplicação de políticas públicas,

fiscais e regulamentares. Cada entidade é obrigada a registar-se com um CAE, que reflete a sua principal área de atuação económica [49]. Desta forma, este código permitirá caracterizar os produtores e os transformadores (descrito na próxima secção) que constam na base de dados analisada e ainda identificar se existe alguma relação com o tipo de resíduo. Os compósitos continham 275 CAEs diferentes, sendo o tipo de resíduo que apresenta um menor número de CAEs associados. De seguida, encontram-se as Misturas com 508 CAEs diferentes, destacando-se as Embalagens com resíduos ou contaminadas com 592 CAEs distintos. Este valor indica que é o tipo de resíduo que atua mais em diferentes áreas de atuação e indústrias. Com o intuito de perceber as áreas dos principais Códigos de Atividade Económica, por tipo de resíduo, procedeu-se à criação da Tabela 13 que indica o Top 5 de CAEs mais frequentes, excluindo à priori, os Inexistentes (antigos NAs do dataset).

Tabela 13 - Top 5 de ocorrências de CAEs de produtores por tipo de resíduo

Tipo de Resíduo	Código de Atividade Económica	Nº de ocorrências
Compósitos	Construção de edifícios (residenciais e não residenciais)	6925
	Fabricação de medicamentos	2154
	Construção de outras obras de engenharia civil	2071
	Construção de estradas e pistas de aeroportos	1499
	Fabricação de misturas betuminosas	739
Embalagens c/ resíduos ou contaminadas	Manutenção e reparação de veículos automóveis	27577
	Comércio de veículos automóveis ligeiros	13049
	Tratamento e eliminação de resíduos perigosos	3908
	Comércio a retalho de peças e acessórios para veículos automóveis	2723
	Fabricação de tintas (exceto impressão), vernizes, mástiques e produtos similares	2650
Misturas	Comércio a retalho em supermercados e hipermercados	47972
	Hotéis com restaurante	5533
	Tratamento e eliminação de outros resíduos não perigosos	4331
	Recolha de outros resíduos não perigosos	2923
	Tratamento e eliminação de resíduos perigosos	2447

No caso dos Compósitos, destacam-se atividades na indústria da construção civil (edifícios, obras, estradas, etc..) e a indústria farmacêutica. As Embalagens com resíduos ou contaminadas são predominantes em setores do ramo automóvel, com um número elevado de ocorrências. Já as Misturas concentram-se em atividades na área do comércio, retalho e restauração. Desta forma, é evidente a divisão que há em relação às diferentes atividades praticadas para cada tipo de resíduo.

Após transformar os dados em tabelas específicas para produtores e transformadores, conforme descrito no capítulo de Data Preparation, foram identificados os principais Códigos de Atividade Económica associados a cada um deles e, assim poder efetuar uma caracterização destes quanto à sua atividade.

A partir da análise da Tabela 14, observa-se que, entre os CAEs listados, há apenas uma sobreposição entre produtores e transformadores que diz respeito à Construção de Edifícios (CAE

41200). Isto sugere que a construção é uma atividade que não só gera uma quantidade significativa de resíduos, mas também está envolvida no seu próprio processamento. Muitas vezes, o setor da construção pode estar associado à produção e transformação de resíduos na medida em que, reutilizam os materiais em matérias-primas para novos projetos, contribuindo para uma economia circular [50].

Tabela 14 - Top 5 de CAEs por produtores e transformadores

Código CAE	Descrição	Nº de Ocorrências
Produtores		
45200	Manutenção e reparação de veículos automóveis	2135
41200	Construção de edifícios (residenciais e não residenciais)	162
45110	Comércio de veículos automóveis ligeiros	140
45320	Comércio a retalho de peças e acessórios para veículos automóveis	130
18120	Outras atividades de impressão e de serviços gráficos	118
Transformadores		
41200	Construção de edifícios (residenciais e não residenciais)	19
42990	Construção de outras obras de engenharia civil	8
43992	Outras atividades especializadas de construção	5
23510	Fabricação de cimento	3
25110	Fabricação de estruturas e partes metálicas	3

No entanto, fora do setor da construção, os produtores estão mais concentrados em atividades de serviços e comércio, como a manutenção e reparação de veículos automóveis e o comércio de veículos e peças, setores que não têm um papel direto na transformação de resíduos, indicando uma clara separação de funções, em que os produtores de resíduos dependem de outros setores industriais (transformadores) especializados para o seu tratamento. Por essa razão, os transformadores estão mais ligados a setores industriais que envolvem o processamento de materiais, como a fabricação de cimento e estruturas metálicas. Estas atividades refletem a necessidade de tratar e reutilizar resíduos industriais, algo que não é tão comum nos setores de comércio e serviços.

Com o objetivo de identificar os CAEs associados a mercadorias perigosas, tanto para produtores como transformadores, efetuou-se uma filtragem pela variável mercadorias_perigosas=*true* (Tabela 15).

Tabela 15 - CAEs associados a mercadorias perigosas

Código CAE	Descrição	Nº de Ocorrências
Produtores		
45200	Manutenção e reparação de veículos automóveis	130
96010	Lavagem e limpeza a seco de têxteis e peles	69
18120	Outras atividades de impressão e de serviços gráficos	40
31091	Fabricação de mobiliário de madeira para outros fins	27
45110	Comércio de veículos automóveis ligeiros	20
Transformadores		
46750	Comércio por grosso de produtos químicos	1

Ao observar a Tabela 15, são notórias as diferenças significativas entre os setores associados à produção e transformação de mercadorias perigosas.

Nos produtores, há uma concentração em setores que estão frequentemente ligados à geração de resíduos perigosos devido ao uso de substâncias químicas, solventes e materiais contaminantes nas suas atividades operacionais, como é o caso do ramo automóvel.

No caso dos transformadores, aparece apenas uma área económica associada ao comércio por grosso de produtos químicos (CAE 46750), o que indica que, neste contexto, há uma menor diversidade de transformadores envolvidos no tratamento de mercadorias perigosas, ao contrário do que se observa nos produtores.

4.2. Clustering

Nesta secção, será aplicado o modelo *clustering* (*K-means*) a dois *dataframes* distintos: produtores e transformadores, conforme descrito anteriormente na secção do Data Preparation. Para cada um deles serão descritas as experiências realizadas e detalhados os resultados obtidos (na secção de Resultados) para cada um destes grupos, com foco na interpretação e identificação de padrões relevantes.

Para tal, efetuaram-se dois modelos para dois *dataframes*, com variáveis diferentes, de modo a testar qual o que obteria melhores resultados. Foi efetuado um tratamento prévio descrito no capítulo anterior que garante que todas as variáveis se encontram preparadas para fazerem parte do modelo implementado, nomeadamente a remoção de *outliers* e a normalização de valores (da variável referente à quantidade). Optou-se por não utilizar as variáveis referentes às quantidades por ano, visto que a sua inclusão não alteraria em nada os resultados, para além de permitir uma análise mais direta entre a quantidade total e as restantes variáveis consideradas.

A primeira experiência tem as variáveis presentes na Tabela 16:

Tabela 16 - Variáveis utilizadas no Clustering (Experiência 1)

Variáveis – Experiência 1
quantidade_total_normalizada
mercadoria_perigosa_transformada
código_ler_associados_150105
código_ler_associados_150106
código_ler_associados_150110

A única diferença nesta experiência para os dois *dataframes* consiste na variável quantidade_total_normalizada, uma vez que uma corresponde à quantidade enviada pelos produtores e noutra pela quantidade recebida pelos transformadores. O foco desta experiência baseia-se em

perceber se o modelo agrupa bem as variáveis referentes ao tipo de resíduo (código LER) com base na quantidade e na classificação de mercadoria perigosa. Para conseguir caracterizar de forma correta os produtores e transformadores, através do gráfico do cotovelo definiu-se o número de clusters ideal para cada modelo, onde numa fase posterior, foram representadas graficamente a relação entre a variável quantidade e as restantes para perceber a atribuição que foi efetuada pelo *K-means*. Adicionalmente, realizaram-se ainda testes ao utilizar a Análise de Componentes Principais, utilizada com o intuito de reduzir a dimensionalidade dos dados e preservar grande parte da variabilidade dos mesmos em menor número de dimensões, facilitando assim a sua visualização [51].

Para obter uma visão geral da construção desta experiência procedeu-se ainda à construção de uma tabela com as médias de cada variável em cada um dos clusters, representado graficamente também por um mapa de calor.

Já a segunda experiência, consiste no mesmo processo de construção do modelo, no entanto adicionaram-se nove novas variáveis criadas para o efeito. Estas dizem respeito ao código de atividade económica, e foram criadas após ser construída uma função a partir da coluna caes_associado onde se extraiu o primeiro dígito de CAEs mais frequentes num determinado produtor ou transformador.

Após esse procedimento, seguiu-se o tratamento já descrito na secção de Data Preparation, tornando-as em variáveis *dummy*. Esta experiência foi efetuada para ambos os *dataframes*, contudo, a nível gráfico contrui-se apenas um mapa geral da caracterização dos produtores e transformadores, tendo em consideração o elevado número de variáveis e de clusters. Desta forma, as variáveis utilizadas para a segunda experiência encontram-se representadas na Tabela 17.

Tabela 17 - Variáveis utilizadas no Clustering (Experiência 2)

Variáveis – Experiência 2	
quantidade_total_normalizada	caes_associados_3
mercadoria_perigosa_transformada	caes_associados_4
código_ler_associados_150105	caes_associados_5
código_ler_associados_150106	caes_associados_6
código_ler_associados_150110	caes_associados_7
caes_associados_1	caes_associados_8
caes_associados_2	caes_associados_9

As variáveis caes_asssociados_1 a caes_associados_9, nove variáveis, são referentes a um setor especifico de atividade económica. O objetivo desta segunda experiência consiste em perceber se o modelo agrupa bem as variáveis referentes ao tipo de resíduo (código LER) e/ou aos diferentes códigos de atividade económica.

Na próxima secção é possível encontrar todos os resultados obtidos para cada uma das 4 tentativas (2 para cada experiência, para produtores e transformadores, respetivamente).

4.3. Resultados

4.3.1. Experiência 1: Clustering excluindo CAEs

Para esta primeira experiência, utilizaram-se as variáveis referidas na Tabela 16 (presente na secção 4.2 do Clustering). O pré-processamento aplicado foi o mencionado na secção 3.3. do Data Preparation (3.3.3 Preparação dos Dados para a Criação de Modelos).

Produtores

Numa primeira instância, procedeu-se à construção do gráfico do cotovelo indicando o número ideal de clusters para este modelo.

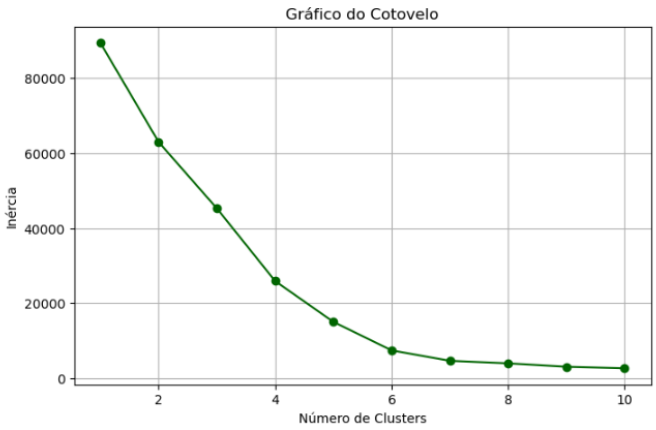


Figura 19 - Número Ideal de clusters para produtores sem CAE

A partir da Figura 19, representada pelo gráfico do cotovelo, foi possível identificar o número ideal de clusters para o modelo aplicado aos produtores. O gráfico do cotovelo é uma técnica amplamente utilizada para determinar o número ótimo de clusters em algoritmos de *clustering*, como o *K-means*. Este é feito com base na soma das distâncias quadráticas dentro dos clusters (inércia) em relação ao número de clusters. O ponto onde há uma diminuição abrupta seguida de uma estabilização gradual no gráfico é conhecido como o "cotovelo" e indica o número ideal de clusters, que neste caso foi identificado como cinco grupos. Este ponto de inflexão revela que, após esse número de clusters, a inércia continua a diminuir, mas de forma menos significativa, o que sugere que a adição de mais clusters não traria um ganho relevante para a segmentação dos dados [52] .

Para além disso, procedeu-se ao cálculo da silhueta (0,81), uma métrica utilizada para avaliar a qualidade dos agrupamentos feitos, em que mede a proximidade de cada produtor com o seu próprio cluster em comparação com os produtores de clusters vizinhos [53]. Sendo que este valor aproxima-se de 1, indica que as alocações feitas estão corretas e bem classificadas.

Assim, a distribuição destes produtores pelos respetivos clusters, foi feita da seguinte forma:

Tabela 18 - Número de Produtores sem CAE incluído por cluster

Cluster	Número de Produtores
0	3166
1	12725
2	1325
3	548
4	125

A distribuição dos clusters, como indicado na Tabela 18, não é muito uniforme, com uma grande concentração de produtores no cluster 1 e um número significativamente menor nos restantes clusters. O cluster 4, com apenas 125 produtores, é o menor. Essa disparidade pode indicar que o algoritmo identificou diferentes características nos dados que agruparam a maioria dos produtores num único grupo. De seguida, analisa-se de forma detalhada esta distribuição para entender a alocação feita pelo K-means. No apêndice d), encontram-se as medidas descritivas por cluster com a média de cada variável para este modelo. As análises detalhadas sobre a distribuição dos clusters para os resíduos Compósitos, Misturas, Embalagens contaminadas e mercadorias perigosas com base na quantidade enviada pelos produtores encontram-se em apêndice e), onde os gráficos apresentam a separação dos clusters com base nas variáveis analisadas. O mapa de calor (Figura 20) resume as médias das variáveis por cluster, que facilita a interpretação deste modelo. Para a variável quantidade, os valores estão normalizados através do Z-score, como mencionado na secção do Data Preparation.

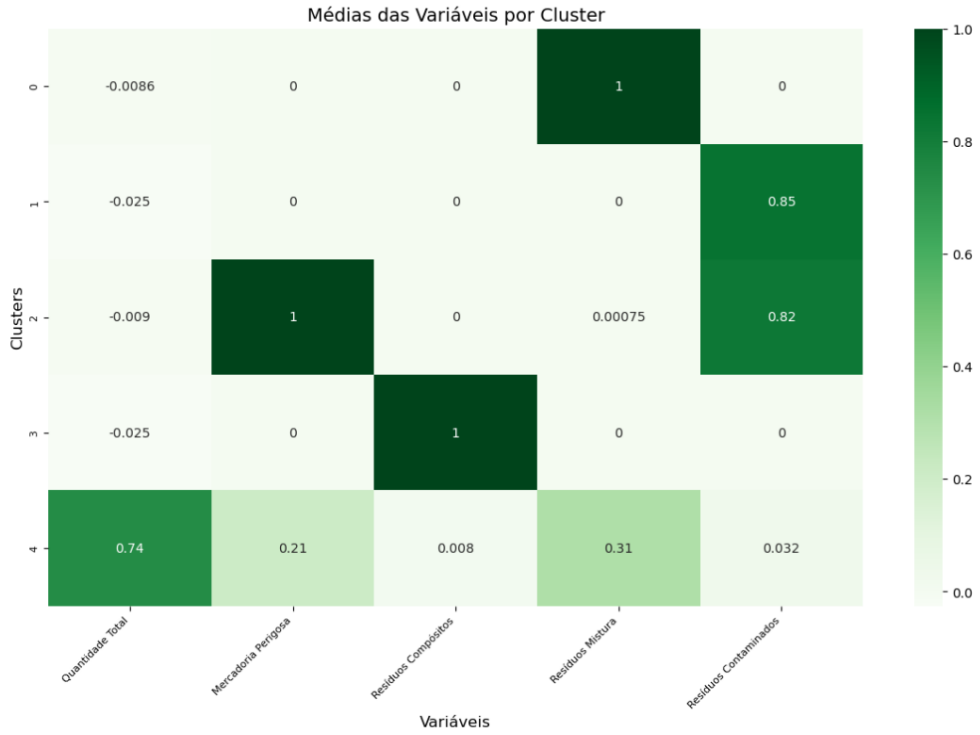


Figura 20 - Mapa de Calor da média das variáveis dos produtores por cluster exlcuindo o CAE

Através da Figura 20, compreende-se a relação que existe entre as variáveis e a atribuição dos produtores a cada um dos 5 clusters, sendo que a intensificação da cor verde, corresponde à representatividade de uma variável, num determinado cluster. De um modo geral, a alocação foi bem feita, tendo em consideração que as variáveis binárias (todas menos a quantidade total) tem uma grande representatividade num determinado cluster. No entanto, as Embalagens com resíduos ou contaminadas (Resíduos Contaminados) estão distribuídos por dois clusters distintos. O mais interessante de realçar nesta separação é a classificação enquanto mercadoria perigosa ou não, isto é, o algoritmo foi capaz de fazer a separação deste tipo de resíduo entre perigoso (cluster 2) e não perigoso (cluster 1). Outra análise relevante, diz respeito ao cluster 4, que atribui produtores que enviaram grandes quantidades de resíduos (verifica-se na primeira coluna do mapa de calor), e por essa razão, verifica-se também um peso adequado aos resíduos de Mistura, tendo em consideração, que na secção das Análises Descritivas era o tipo de resíduo com maiores quantidades enviadas. Por outro lado, os resíduos Compósitos são aqueles que apresentam menores quantidades elevadas.

Adicionalmente, recorreu-se à Análise das Componentes Principais, com o objetivo de testar se seria possível obter melhores resultados. No entanto, este não permitiu retirar nenhuma conclusão relevante, sendo por isso considerada uma tentativa de insucesso (em apêndice f)).

Após construir-se o modelo sem a inclusão das variáveis referentes ao CAE, torna-se pertinente perceber quais os Códigos de Atividade Económica mais associados a cada cluster (Tabela 19).

Tabela 19 -Top 3 de CAEs associados aos clusters da 1ª Experiência (Produtores)

Cluster	CAE	Nº Produtores
0	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	957
	Agricultura, Produção Animal, Caça, Florestas e Pesca	330
	Indústrias Extrativas	303
1	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	6560
	Indústrias Extrativas	1549
	Agricultura, Produção Animal, Caça, Florestas e Pesca	1340
2	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	349
	Indústrias Extrativas	316
	Agricultura, Produção Animal, Caça, Florestas e Pesca	250
3	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	194
	Indústrias Extrativas	53
	Agricultura, Produção Animal, Caça, Florestas e Pesca	49
4	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	38
	Agricultura, Produção Animal, Caça, Florestas e Pesca	25
	Indústrias Transformadoras	24

Desta forma, é percetível que para os 5 clusters, o setor energético, agrícola e as indústrias extrativas foram predominantes neste modelo. Apenas no último cluster, para o Top 3 de CAEs existe uma diferença sendo as Indústrias Extrativas substituídas pelas Indústrias Transformadoras. Este facto

poderá ser explicado pelo facto de ser o cluster associado também a grandes quantidades e a mercadorias perigosas, como verificado na Figura 20.

Transformadores

Para os transformadores, o processo de implementação do modelo foi feito de igual forma, mudando apenas a variável da quantidade (passando a ser utilizada a quantidade recebida). Após todo o processo de tratamento mencionado no Data Preparation, na Figura 21 é visível o gráfico do cotovelo que permite a identificação do número de clusters.

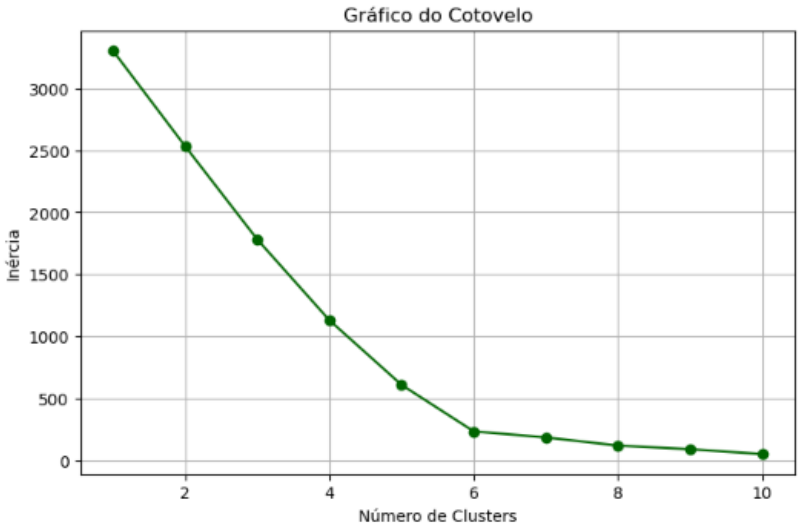


Figura 21 - Número Ideal de clusters para transformadores sem CAE

Como é possível observar na Figura 21, o número ideal de clusters identificado é 6, conforme indicado pela curva apresentada no gráfico. Comparativamente ao modelo anterior, este apresenta mais um agrupamento. Após a definição do número de clusters, foi realizado o cálculo do índice de silhueta, resultando num valor de 0,90. Este valor indica que o modelo apresenta uma boa qualidade de agrupamento, sugerindo que os clusters estão bem definidos e com uma separação clara entre eles. Para tal, segue-se abaixo a composição do mesmo (Tabela 20).

Tabela 20 - Número de transformadores sem CAE incluído por cluster

Cluster	Número de Transformadores
0	158
1	147
2	32
3	206
4	116
5	13

Ao analisar a Tabela 20, verifica-se uma distribuição heterogénea de transformadores pelos diferentes clusters, com alguns agrupamentos a concentrarem uma quantidade significativa de transformadores, enquanto outros apresentam um número consideravelmente menor. Esta variação

sugere que há padrões distintos entre os grupos, com alguns clusters a representarem transformadores com características mais comuns, enquanto outros agregam transformadores com características mais especificas.

Assim, procedeu-se à realização de uma análise aprofundada desta distribuição para compreender melhor como é que o *K-means* alocou os transformadores nos diferentes clusters. No Apêndice g), são apresentadas as estatísticas descritivas por cluster, mostrando a média de cada variável, à semelhança do que foi feito para o anterior.

No Apêndice h), encontram-se detalhadas as análises da distribuição dos clusters para as variáveis relacionadas aos resíduos Compósitos, Misturas, Embalagens contaminadas e mercadorias perigosas, com base nas quantidades recebidas pelos transformadores. Neste apêndice, os gráficos ilustram a segmentação dos clusters com base nessas variáveis específicas.

Também a Adicionalmente, foi realizada uma Análise de Componentes Principais, contudo, esta análise não permitiu tirar conclusões, tal como referido no modelo anterior, e por isso, foi considerada infrutífera (Gráfico no apêndice i)). *Tal como realizado no modelo dos produtores, averiguou-se o Top 3 de CAEs associados aos clusters deste modelo*, através de um mapa de calor, sintetiza as médias das variáveis por cluster, permitindo entender a relação entre as variáveis e a atribuição dos transformadores aos seis clusters definidos. Em particular, destaca-se que a quantidade recebida, mesmo após a normalização, tem uma representatividade significativa no Cluster 5. No geral, a alocação dos transformadores foi adequada, uma vez que as variáveis binárias (todas exceto a quantidade total) apresentam uma forte representatividade em clusters específicos. Esta distribuição evidencia a coerência do modelo em segmentar as entidades de uma forma mais clara e consistente.

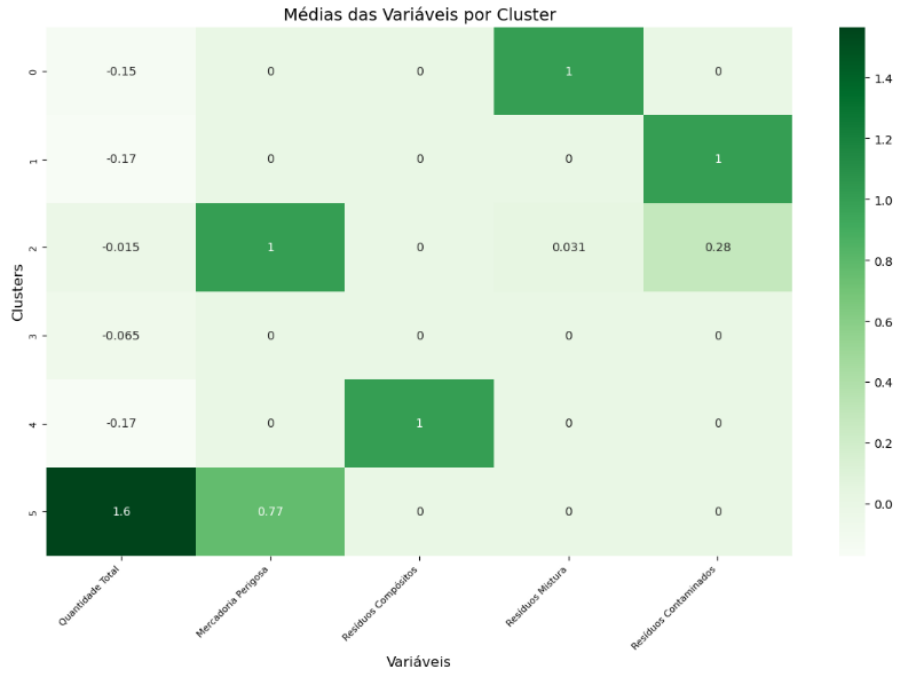


Figura 22 - Mapa de Calor da média das variáveis dos transformadores por cluster sem o CAE

À semelhança do que aconteceu no modelo anteriormente apresentado referente aos produtores, os resíduos classificados como Embalagens com resíduos ou contaminadas estão distribuídos em dois clusters distintos. Um destes clusters tem alguma representatividade em Mercadorias Perigosas (Cluster 2), enquanto o outro não apresenta essa característica (Cluster 1). Este comportamento do algoritmo demonstra sua capacidade de separar corretamente resíduos perigosos de não perigosos.

Ainda sobre as mercadorias perigosas, observa-se que estas estão divididas entre o já mencionado Cluster 2 e o Cluster 5, que se destaca por tratar grandes quantidades de resíduos. Isso sugere que, quando se trata de mercadorias perigosas, os transformadores tendem a receber essas substâncias em grandes volumes.

Adicionalmente, foi realizada uma Análise de Componentes Principais, contudo, esta análise não permitiu tirar conclusões, tal como referido no modelo anterior, e por isso, foi considerada infrutífera (Gráfico no apêndice i)). *Tal como realizado no modelo dos produtores, averiguou-se o Top 3 de CAEs associados aos clusters deste modelo (Tabela 21)*

Tabela 21- Top 3 de CAEs associados aos clusters da 1ª Experiência (Transformadores)

Cluster	CAE	Nº Transformadores
0	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	59
	Agricultura, Produção Animal, Caça, Florestas e Pesca	25
	Indústrias Extrativas	21
1	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	64
	Agricultura, Produção Animal, Caça, Florestas e Pesca	27
	Indústrias Extrativas	21
2	Indústrias Extrativas	16
	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	12
	Agricultura, Produção Animal, Caça, Florestas e Pesca	3
3	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	126
	Indústrias Extrativas	36
	Agricultura, Produção Animal, Caça, Florestas e Pesca	28
4	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	64
	Indústrias Extrativas	15
	Agricultura, Produção Animal, Caça, Florestas e Pesca	13
5	Indústrias Extrativas	7
	Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio	4
	Agricultura, Produção Animal, Caça, Florestas e Pesca	2

A partir da Tabela 21, observa-se que o Top 3 de CAEs mais frequentes neste cluster é o mesmo identificado para os produtores, abrangendo o setor energético, agrícola e indústrias extrativas. A única diferença entre os clusters está na variação da ordem desses três setores.

4.3.2. Experiência 2: Clustering incluindo CAEs

Para a realização desta segunda experiência, utilizou-se os primeiros dígitos dos códigos de atividade económica para perceber como se comportava o modelo K-means face ao anteriormente apresentado. Para tal, recorreu-se ao INE [54], para identificar a que setores se referem cada um dos primeiros dígitos:

1. Agricultura, Produção Animal, Caça, Florestas e Pesca
2. Indústrias Extrativas
3. Indústrias Transformadoras
4. Eletricidade, Gás, Vapor, Água Quente e Fria e Ar Frio
5. Captação, Tratamento e Distribuição de Água; Saneamento, Gestão de Resíduos e Despoluição
6. Construção
7. Comércio por Grosso e a Retalho; Reparação de Veículos Automóveis e Motociclos
8. Transportes e Armazenamento
9. Alojamento, Restauração e Similares

Com base nisto, utilizaram-se estas nove variáveis em formato binário, para além das que foram utilizadas na primeira experiência (conforme mencionado na Tabela 17 da secção 4.2 Clustering). Através desta segunda tentativa é também possível caracterizar os produtores e transformadores, não só quanto ao tipo de resíduo, mas também ao setor de atividade económica que estão mais diretamente relacionados.

Todo o tratamento realizado para esta experiência foi idêntico à experiência anteriormente apresentada e que se encontra também descrita na secção 3.3. do Data Preparation (3.3.3 Preparação dos Dados para a Criação de Modelos).

Produtores

Tal como na primeira experiência procedeu-se inicialmente à criação do gráfico do cotovelo para averiguar o número ideal de clusters.

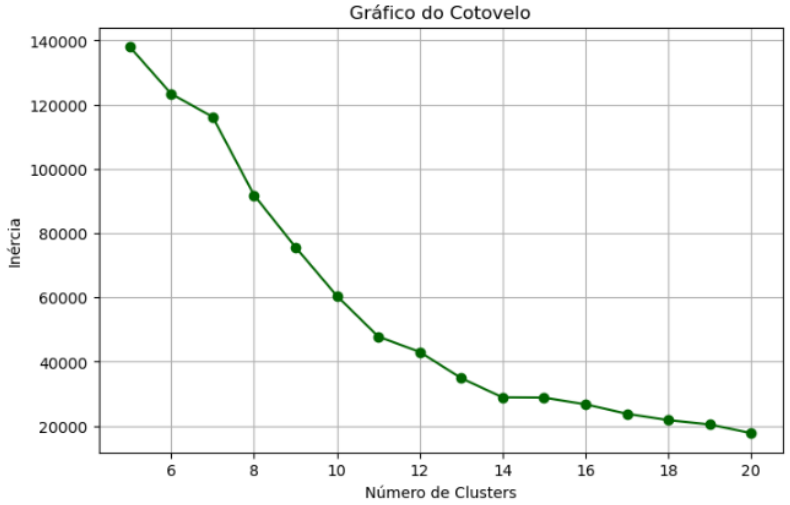


Figura 23 - Número Ideal de clusters para Produtores com CAE

Através da Figura 23, não se verifica uma curva clara que permita identificar de forma direta o número ideal de clusters. Dado que a diminuição da inércia começa diminuir e a estabilizar a partir do número 12, optou-se por utilizar esse valor para o número ideal de clusters . Esta escolha permite que haja uma segmentação mais detalhada com base nas variáveis selecionadas para esta experiência, tendo assim, uma melhor representatividade dos diferentes grupos. Para este número, obteve-se uma silhueta de 0,68 indicando que os grupos formados são coesos e fazem uma separação clara e entre os diferentes grupos.

Com base nisto, a Tabela 22 indica o número de produtores atribuídos a cada cluster.

Tabela 22 - Número de Produtores com CAE incluído por cluster

Cluster	Número de Produtores
0	964
1	5962
2	1614
3	1880
4	367
5	118
6	210
7	278
8	1020
9	251
10	314
11	1614

A partir da Tabela 22, é possível observar que alguns clusters concentram grande parte dos produtores, o que sugere a existência de grupos com características comuns entre um grande número de entidades. O cluster 1, por exemplo, destaca-se por conter o maior número de produtores, refletindo possivelmente a predominância de características semelhantes dentro deste grupo.

Por outro lado, há clusters com um número relativamente menor de produtores, como os clusters 5, 6, e 9, o que pode indicar a presença de grupos mais específicos, possivelmente associados a características menos comuns. Estes clusters com menor número de produtores podem ser uma mais valia na identificação de segmentos de produtores com perfis diferentes do que é habitual.

Em apêndice j), encontram-se as principais medidas descritivas para cada um dos clusters com todas as variáveis incluídas. Para uma melhor visualização, a Figura 24 permite retirar algumas conclusões sobre este modelo. É ainda de referir que as labels do eixo do x foram adaptadas face ao descrito acima devido à dimensão das categorias do CAE.

A Figura 24 permite observar a relação entre a atribuição dos produtores aos diferentes clusters, com base no tipo de resíduo e nos diversos códigos de atividade económica. O que mais se destaca, à primeira vista, é a predominância de resíduos de Embalagens contaminadas entre os vários clusters, com especial ênfase no cluster 1, que está fortemente associado ao CAE relacionado com Eletricidade,

Gás, Vapor, Água Quente e Fria, e Ar Frio. Esta concentração indica que os produtores deste setor estão envolvidos em atividades que geram este tipo específico de resíduo.

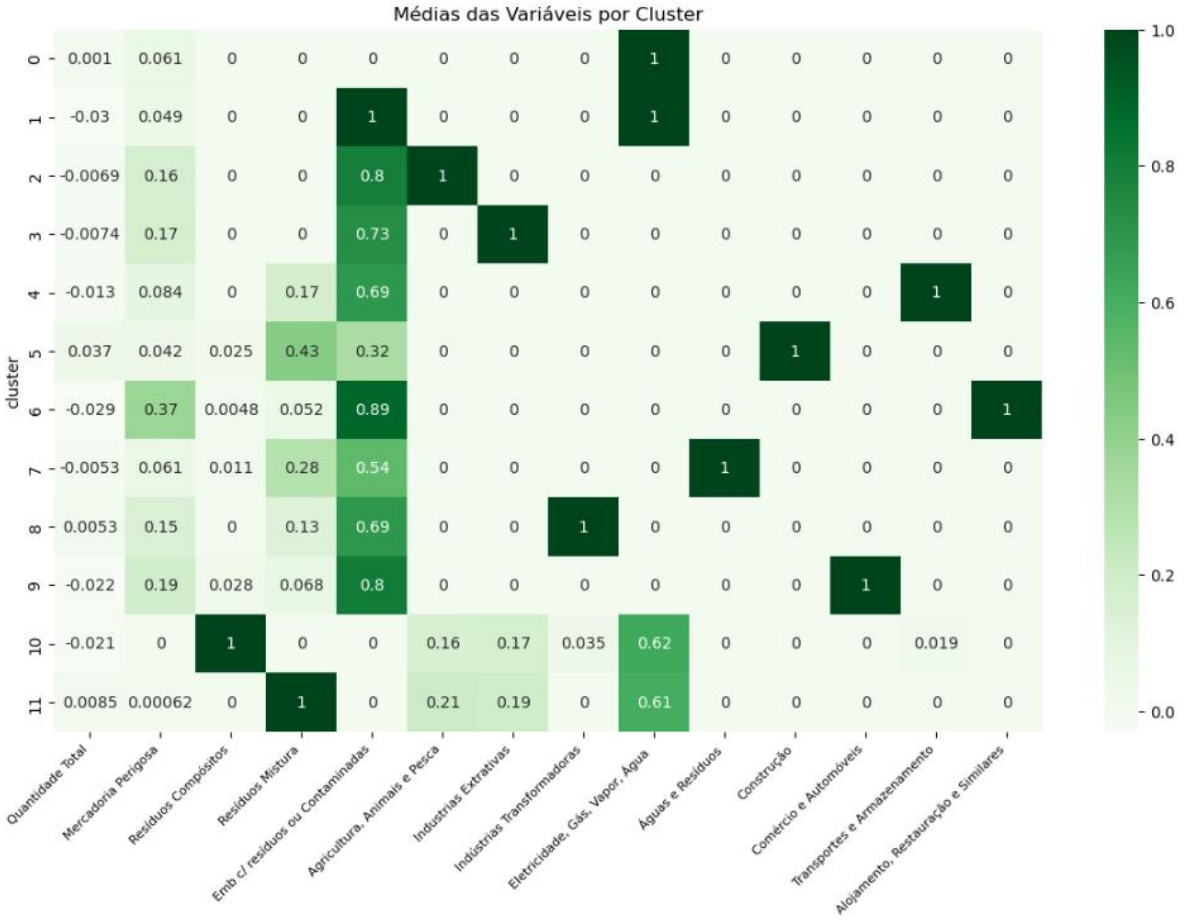


Figura 24 - Mapa de Calor da média das variáveis dos Produtores por cluster incluído o CAE

Para além disso, tanto os resíduos Compósitos como as Misturas (clusters 10 e 11, respetivamente) estão associados aos mesmos três códigos de atividade económica. Esta repetição revela que estes tipos de resíduos, diferem-se apenas pela sua natureza, atuando em setores semelhantes.

Outro ponto importante de referir é que o CAE relacionado com Eletricidade, Gás, Vapor, Água Quente e Fria, e Ar Frio tem uma presença significativa em dois clusters diferentes (clusters 0 e 1), o que não ocorre com os restantes CAEs. Este comportamento está relacionado com a divisão deste CAE entre o cluster 1 que contém embalagens Contaminadas e o cluster 0 que não possui nenhuma representação deste tipo de resíduo.

Os restantes CAEs, tal como era expectável, estão representados apenas num único cluster, o que demonstra que as atribuições dos produtores aos clusters parecem ter sido bem realizadas, com uma correspondência clara entre o tipo de resíduo e o setor económico responsável pela sua produção.

Contrariamente ao que foi observado na primeira experiência, onde os CAEs não foram incluídos, a quantidade de resíduos e a classificação de mercadorias perigosas não têm um peso significativo neste modelo, o que indica que a inclusão dos CAEs levou a uma segmentação mais orientada para os setores económicos e menos focada nas características dos resíduos em si.

Transformadores

O processo realizado para os transformadores seguiu um procedimento idêntico ao utilizado anteriormente para os produtores. Tal como na análise dos produtores, a Figura 25 não revela de forma clara um ponto de inflexão evidente que indique o número ideal de clusters para a construção deste modelo.

Figura 25 - Número Ideal de clusters para Transformadores com CAE

No entanto, decidiu-se criar também 12 clusters, tendo em consideração que a inércia começa a estabilizar e diminuir de forma mais lenta a partir desse valor. Embora o índice de silhueta obtido para este caso tenha sido inferior ao dos produtores (0,52), continua a representar uma boa separação entre os clusters. Este valor pode indicar apenas que, apesar dos clusters estarem relativamente bem definidos, há uma maior sobreposição entre grupos. Através das análises seguintes, será percetível o comportamento deste modelo.

Através da Tabela 23 que indica o número de transformadores para cada cluster, verifica-se uma grande disparidade na sua distribuição pelos diferentes clusters. Este desnível transmite a ideia que alguns grupos são significativamente mais representativos, enquanto outros contêm apenas um ou poucos transformadores. Para validar a eficácia do modelo e entender as razões para essa segmentação, é essencial realizar uma análise descritiva detalhada de cada cluster, com o intuito de identificar que tipo de transformadores estão concentrados em certos grupos e que características específicas os distinguem daqueles que estão mais isolados. Por outro lado, clusters mais povoados,

como o cluster 6 com 185 transformadores e o cluster 4 com 93 transformadores, é expectável que estejam a agrupar transformadores com características mais comuns e homogéneas.

Tabela 23 - Número de Transformadores com CAE incluído por cluster

Cluster	Número de Transformadores
0	68
1	91
2	49
3	42
4	93
5	3
6	185
7	1
8	7
9	3
10	64
11	1

A análise das médias das variáveis associadas a estes clusters encontra-se disponível no Apêndice k), e a Figura 26 ilustra essas mesmas medidas através de um mapa de calor.

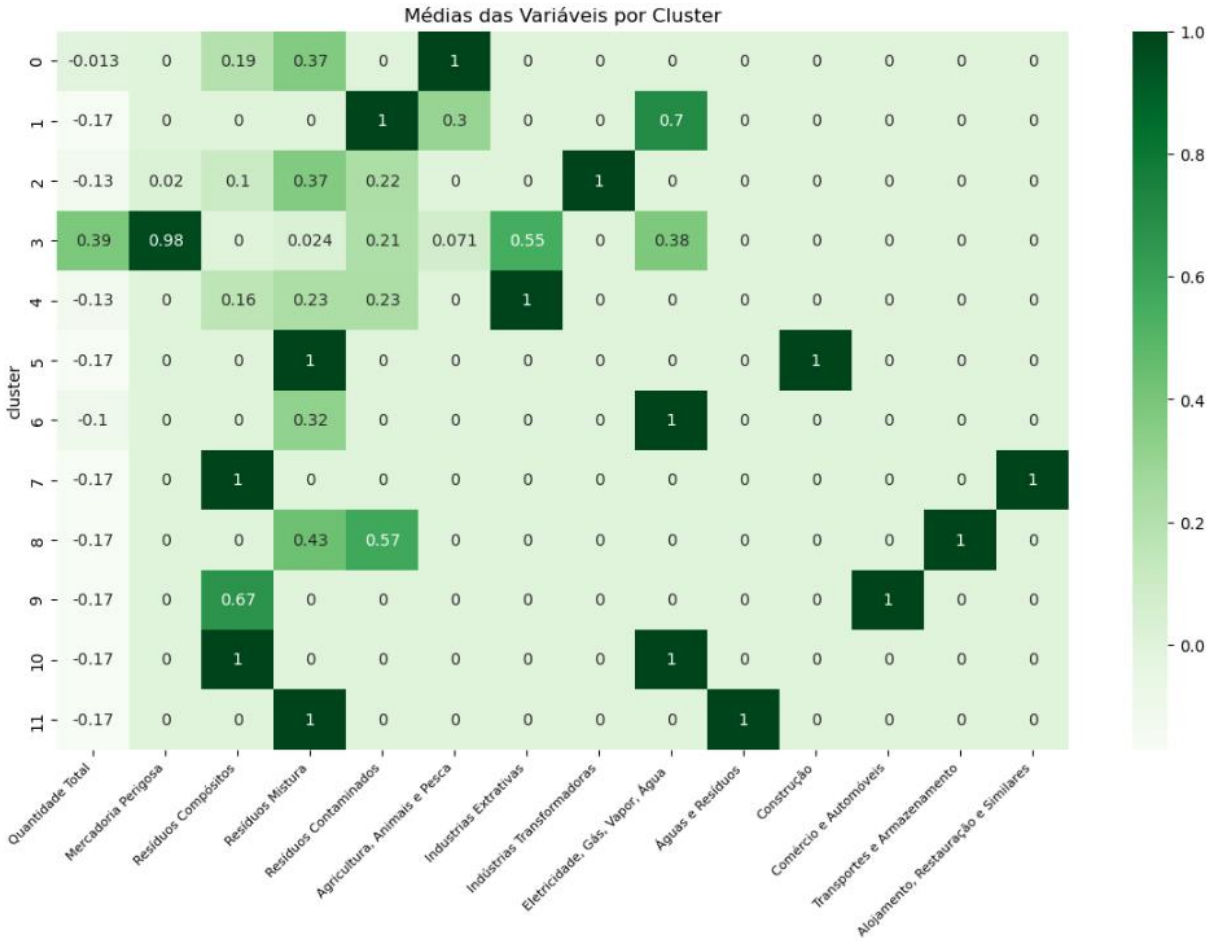


Figura 26 - Mapa de Calor da média das variáveis dos Transformadores por cluster com o CAE

A Figura 26 demonstra que todas as variáveis, com exceção da quantidade total, estão bem representadas nos diferentes clusters. Ao contrário do que foi observado nesta experiência com os

produtores, as mercadorias perigosas apresentam-se de forma significativa no cluster 3, onde há também uma representatividade expressiva do CAE relacionado com as Indústrias Extrativas e com o setor da Eletricidade, Gás, Vapor e Água. Isso pode indicar que este cluster agrupa transformadores com características particulares associadas a atividades de extração de recursos naturais e ao setor energético.

Os resíduos Compósitos estão distribuídos por três clusters distintos (7, 10 e, em menor grau, 9, com uma representatividade de 0.67), associados a três CAEs diferentes: Alojamento Restauração e Similares, Eletricidade, Gás, Vapor e Água, e Comércio e Automóveis. Esta separação reforça a ideia de que os resíduos Compósitos foram corretamente alocados, uma vez que foram atribuídos a atividades económicas bastante distintas entre si.

Por outro lado, os resíduos de Misturas encontram-se distribuídos por vários clusters, mas estão particularmente concentrados nos clusters 5 e 11, associados aos CAEs de Construção e Água e Resíduos, respetivamente. Esta distribuição sugere que os resíduos de Misturas têm uma ligação forte a setores relacionados com a construção civil e a gestão de recursos hídricos, refletindo a natureza dos resíduos transformados nesses setores.

O cluster 1 agrupa exclusivamente resíduos de Embalagens com resíduos ou contaminadas, sendo o único cluster onde este tipo de resíduo é predominante. No entanto, também é possível encontrar este tipo de resíduo noutros clusters, especialmente ligados ao CAE de Eletricidade, Gás, Vapor e Água, indicando que há subgrupos dentro deste setor que possuem diferentes características, justificando assim a sua separação em vários clusters.

Como mencionado na análise da Tabela 23, os clusters com poucos transformadores, como os clusters 5, 7, 9 e 11, onde há apenas 1 a 3 transformadores, representam entidades com características especificas, e por essa razão, é que todos eles estão distribuídos de forma tão uniforme (valor 1 para o tipo de resíduo e valor 1 no setor associado). Isto quer dizer que há transformadores específicos que atuam num setor em que os restantes não operam. Esta separação pode ser explicada pelo facto destes transformadores estarem envolvidos em atividades muito específicas ao operarem em nichos de mercado, justificando possivelmente esta distribuição. Esta diferenciação é reforçada pela análise apresentada na Figura 26, que revela a distribuição das variáveis associadas, permitindo compreender melhor as particularidades de cada grupo.

A escolha de um maior número de clusters para a Experiência 2 revelou ser a opção mais acertada, pois, após a realização de diversos testes, foi possível alcançar classificações mais precisas e com um melhor score.

CAPÍTULO 5

Conclusões e Trabalho Futuro

5.1. Conclusões

Esta dissertação teve como objetivo explorar a deteção de anomalias na gestão de resíduos em Portugal, entre 2019 e 2022, utilizando os dados disponibilizados pela APA. Através das análises detalhadas dos produtores e transformadores de resíduos, foi possível identificar alguns padrões e possíveis irregularidades no processo de gestão de resíduos, bem como avaliar a relação entre os Códigos de Atividade Económica (CAE) e os tipos de resíduos gerados e transformados.

As três questões de investigação formuladas inicialmente e apresentadas abaixo, foram parcialmente respondidas, contribuindo para uma visão um pouco mais clara sobre os desafios e as oportunidades de melhoria na gestão de resíduos em Portugal. As dificuldades no tratamento e a falta de contacto com a APA numa fase crítica do projeto impediram a obtenção de conclusões mais fundamentadas.

É viável, através dos dados disponibilizados pela APA, detetar anomalias no registo do transporte / gestão de resíduos?

O principal objetivo de detetar anomalias na gestão de resíduos foi parcialmente alcançado, na medida em que através das análises descritivas evidenciaram-se discrepâncias nos dados de transporte entre produtores e transformadores, apesar de não se ter conhecimento do motivo pelo qual acontece.

Ao longo dos quatro anos analisados, 16,3% dos registos de transporte foram anulados e 1,8% precisaram de correção, apontando possíveis falhas nos registos, algo que seria benéfico vir a ser validado pela APA. Uma das principais irregularidades detetadas foi a discrepância entre as quantidades enviadas e recebidas para os transportes dados como concluídos. Para as Misturas e Compósitos, a quantidade recebida foi menor do que a enviada, enquanto para as Embalagens Contaminadas, a quantidade recebida foi maior do que a enviada. Estas diferenças, que deveriam ser mínimas, totalizam milhares de toneladas, ao fim de quatro anos o que poderá indicar falhas no registo ou no próprio transporte. Detetou-se ainda, que para os transportes que foram corrigidos, as diferenças nas quantidades foram muito menores face aos transportes dados como concluídos, o que indica que estas correções têm impacto nestas diferenças significativas de quantidades.

Outra irregularidade identificada foi a presença de plásticos nos resíduos de Misturas, tanto no envio quanto na chegada aos transformadores, o que poderá indicar falhas na sua categorização. Adicionalmente, o tempo médio de transporte, apresenta valores anómalos que levam a crer que o seu registo pode nem sempre corresponder ao valor real.

Para complementar estas conclusões, no apêndice I) é apresentado o *dashboard* desenvolvido, que possibilita a filtragem e visualização detalhada de informações referentes aos resíduos dos produtores e transformadores.

Como se caracterizam os produtores e transformadores atuais em Portugal?

Com o intuito de caracterizar os produtores e os transformadores de resíduos, foram realizadas duas experiências, uma utilizando os Códigos de Atividade Económica e outra sem essas variáveis. Na experiência sem o CAE, observou-se que os produtores de resíduos de Misturas estão associados a grandes volumes de resíduos e a mercadorias perigosas. Ainda, no caso das Embalagens com resíduos ou contaminadas, foram identificados dois grupos distintos de produtores: um que lida com mercadorias perigosas e outro que não. Entre os transformadores, os resultados foram menos claros devido ao número reduzido destes agentes. Desta forma, não foi possível associar nenhum dos três tipos de resíduos a mercadorias perigosas ou fazer qualquer atribuição com base na quantidade recebida. No entanto, foi notório, que os transformadores que lidam com grandes quantidades de resíduos, são, em geral, os mesmos que estão associadas ao tratamento de mercadorias perigosas.

Qual a relação que os Códigos de Atividade Económica têm na caracterização dos produtores e transformadores face aos tipos de resíduos analisados?

Já na experiência com o CAE, foi possível observar que as Embalagens com resíduos ou contaminadas estão distribuídas por vários setores de atividade económica, enquanto os resíduos de Misturas e Compósitos estão concentrados principalmente no setor energético (eletricidade, gás, vapor e água). Entre os transformadores, verificou-se uma maior especialização, com cada grupo focado num setor económico específico, o que pode evidenciar que estes estão focados numa área económica para conseguirem transformar determinado tipo de resíduos.

A análise revelou ainda que, de um modo geral, os produtores destes resíduos (que potencialmente contêm plástico, mas não são resíduos totalmente plásticos) estão concentrados em atividades de serviços e comércio, como a manutenção e reparação de veículos automóveis e o comércio de veículos e peças. Estas atividades não estão diretamente envolvidas no tratamento de resíduos, o que evidencia uma clara divisão de funções, onde os produtores confiam em setores industriais especializados (transformadores) para realizar o tratamento adequado.

De um modo geral, esta dissertação foi ao encontro dos objetivos definidos inicialmente, respondendo às perguntas de investigação formuladas, fornecendo *insights* relevantes para o desenvolvimento e apoio interno da APA. Estes resultados oferecem uma base consistente para a melhoria contínua do sistema de gestão de resíduos português.

5.2. Limitações

Ao longo deste estudo, foram identificadas algumas limitações que poderiam ter vindo a impactar os resultados obtidos. Primeiramente, o contacto direto e contínuo com a APA teria sido uma mais-valia, permitindo uma compreensão mais profunda dos dados e possibilitando um diálogo direto sobre as irregularidades detetadas e o eventual despiste de informações dúbias. Tal envolvimento poderia ter facilitado a validação de algumas suposições sobre os resultados obtidos enriquecendo as conclusões retiradas.

Para além disso, a falta de respostas por parte dos produtores e transformadores presentes na base de dados da APA, e que foram contactados durante a realização deste estudo impediram a adição de informações complementares que poderiam ser relevantes para esta investigação.

Outro aspeto a considerar foi a limitação na diversidade das variáveis disponibilizadas. Embora os dados fornecessem uma boa base para o estudo, variáveis adicionais, como por exemplo os custos operacionais, poderiam ter proporcionado uma caracterização mais completa. Esta inclusão teria ainda permitido uma avaliação mais detalhada das práticas utilizadas no âmbito da gestão de resíduos em Portugal.

No que diz respeito aos modelos implementados, verificou-se que o número de clusters na experiência onde se incluiu os CAEs limitou a possibilidade de obter uma segmentação mais diferenciada, apesar dos esforços para ajustar os modelos construídos.

Outro desafio foi a realização da análise distrital, que foi possível apenas realizar para os produtores, uma vez que os dados dos transformadores não permitiram uma avaliação geográfica detalhada. Este facto limitou a realização de uma comparação dos locais principais onde atuavam os transformadores face ao que foi analisado em relação aos produtores.

Por fim, algumas variáveis presentes nos *datasets* disponibilizados não foram incluídas na análise devido à falta de clareza sobre o seu significado ou contexto, o que restringiu o seu uso para enriquecer os resultados. Uma maior compreensão dessas variáveis teria certamente contribuído para uma análise mais abrangente.

5.3. Trabalho Futuro

Tendo em conta as limitações identificadas e os resultados obtidos, há várias oportunidades para realizar uma investigação futura que possa complementar este trabalho e contribuir para o avanço no estudo da gestão de resíduos em Portugal em conjunto com a APA.

Uma das principais áreas que seria benéfico analisar aprofundadamente consistiria na análise de todos os resíduos associados a plásticos, compreendendo as suas diferenças e classificações. É importante referir, que há uma grande probabilidade de erros na classificação da tipologia de resíduos

que estejam associados ao plástico, tendo em consideração as falhas já verificadas para os três tipos de resíduos analisados nesta dissertação. Desta forma, seria interessante perceber as causas destas discrepâncias e de sugerir formas de mitigar o problema.

Outro aspeto relevante seria a análise detalhada das operações realizadas com os resíduos. Muitos registos não especificam de forma clara qual a operação realizada com os resíduos, o que limita a compreensão do que será feito com os mesmos. Um estudo mais focado nesta questão poderia proporcionar conclusões pertinentes para que haja um melhor rastreio no processo de tratamento de resíduos.

Adicionalmente, seria interessante testar novos algoritmos e técnicas de *clustering* que possam caracterizar de forma mais detalhada os produtores e transformadores. Tal como referido no Capítulo de Revisão de Literatura, também modelos de *deep learning* e *computer vision* podem gerar resultados muito positivos, como é o caso do Modelo AIEWO-WMC que permite o reconhecimento e categorização de objetos residuais com uma taxa de precisão muito elevada.

Por fim, futuros trabalhos neste âmbito poderiam beneficiar de uma colaboração mais direta com outras entidades reguladoras e industriais, como a APA, numa ótica de facilitar o acesso a informações complementares. A inclusão de variáveis adicionais, poderá ainda ajudar a criar um panorama mais completo e a oferecer soluções práticas e aplicáveis para melhorar a sustentabilidade da gestão de resíduos em Portugal.

Referências Bibliográficas

[1] «Portugal em vias de falhar (novamente) metas de reciclagem». Acedido: 25 de novembro de 2023. [Em linha]. Disponível em: https://www.jornaldenegocios.pt/sustentabilidade/detalhe/20231030-0933-portugal-em-vias-de-falhar-novamente-metas-de-reciclagem

[2] «Produção e Gestão de Resíduos | APA». Acedido: 11 de novembro de 2023. [Em linha]. Disponível em: https://apambiente.pt/residuos/producao-e-gestao-de-residuos

[3] «Recolha | APA». Acedido: 13 de janeiro de 2024. [Em linha]. Disponível em: https://apambiente.pt/index.php/residuos/recolha

[4] «Transporte | APA». Acedido: 14 de janeiro de 2024. [Em linha]. Disponível em: https://apambiente.pt/index.php/residuos/transporte

[5] «Taxa de Gestão de Resíduos | APA». Acedido: 18 de novembro de 2023. [Em linha]. Disponível em: https://apambiente.pt/residuos/taxa-de-gestao-de-residuos

[6] APA, «Estudo prévio sobre a implementação da recolha seletiva em Portugal Continental incindindo em especial sobre o fluxo dos biorresíduos», jul. 2019. Acedido: 22 de novembro de 2023. [Em linha]. Disponível em: https://apambiente.pt/sites/default/files/2023-09/Estudo%20Pr%C3%A9vio%20Biorresiduos%202019.pdf

[7] ECO, «Faltam verbas e fiscalização para Portugal cumprir meta de recolha de resíduos elétricos», ECO. Acedido: 26 de novembro de 2023. [Em linha]. Disponível em: https://eco.sapo.pt/especiais/faltam-verbas-e-fiscalizacao-para-portugal-cumprir-meta-de-recolha-de-residuos-eletricos/

[8] «Dados sobre resíduos urbanos | APA». Acedido: 21 de novembro de 2023. [Em linha]. Disponível em: https://apambiente.pt/residuos/dados-sobre-residuos-urbanos

[9] «Reciclagem: Portugal, o eterno incumpridor». Acedido: 25 de novembro de 2023. [Em linha]. Disponível em: https://www.jornaldenegocios.pt/sustentabilidade/ambiental/detalhe/reciclagem-portugal-o-eterno-incumpridor

[10] N. Hotz, «What is CRISP DM?», Data Science Process Alliance. Acedido: 12 de julho de 2024. [Em linha]. Disponível em: https://www.datascience-pm.com/crisp-dm-2/

[11] M. Gonzalez, «CRISP-DM na prática», Medium. Acedido: 13 de julho de 2024. [Em linha]. Disponível em: https://medium.com/@mgonzaleze/crisp-dm-na-pr%C3%A1tica-65be0ee92ada

[12] «PRISMA 2020 flow diagram», PRISMA statement. Acedido: 14 de setembro de 2024. [Em linha]. Disponível em: https://www.prisma-statement.org/prisma-2020-flow-diagram

[13] Parlamento Europeu, «Gestão de resíduos na UE: factos e números». novembro de 2023. [Em linha]. Disponível em: https://www.europarl.europa.eu/pdfs/news/expert/2020/12/story/20180328STO00751/20180328STO00751_pt.pdf

[14] «Como alcançar a economia circular na UE até 2050? | Atualidade | Parlamento Europeu». Acedido: 17 de janeiro de 2024. [Em linha]. Disponível em: https://www.europarl.europa.eu/news/pt/headlines/society/20210128STO96607/como-alcancar-a-economia-circular-na-ue-ate-2050

[15] *COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES Um novo Plano de Ação para a Economia Circular Para uma Europa mais limpa e competitiva*. 2020. Acedido: 22 de setembro de 2024. [Em linha]. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52020DC0098

[16] F. Gökgöz e E. Yalçın, «Investigating Waste Management Efficiencies and Dynamics of the EU Region», em *Pragmatic Engineering and Lifestyle*, D. S-K. Ting e J. A. Stagner, Eds., Emerald Publishing Limited, 2023, pp. 91–111. doi: 10.1108/978-1-80262-997-220231005.

[17] D. Sultanova, A. Maliashova, e S. Gadelshina, «Waste management as an element of sustainable development of the circular economy in the European Union», *E3S Web Conf.*, vol. 247, p. 01007, 2021, doi: 10.1051/e3sconf/202124701007.
[18] G. Di Foggia e M. Beccarello, «An Overview of Packaging Waste Models in Some European Countries», *Recycling*, vol. 7, n.º 3, p. 38, jun. 2022, doi: 10.3390/recycling7030038.
[19] E. Mihaliková, M. Taušová, e K. Čulková, «Public Expenses and Investment in Environmental Protection and Its Impact on Waste Management», *Sustainability*, vol. 14, n.º 9, p. 5270, abr. 2022, doi: 10.3390/su14095270.
[20] H. Salmenperä, «Different pathways to a recycling society – Comparison of the transitions in Austria, Sweden and Finland», *Journal of Cleaner Production*, vol. 292, p. 125986, abr. 2021, doi: 10.1016/j.jclepro.2021.125986.
[21] K. Pouikli, «Concretising the role of extended producer responsibility in European Union waste law and policy through the lens of the circular economy», *ERA Forum*, vol. 20, n.º 4, pp. 491–508, mar. 2020, doi: 10.1007/s12027-020-00596-9.
[22] M. D'Agostini, S. Venturi, e E. Vigo, «Smart Urban Waste Management System: The Case Study of Delft, Netherlands», em *Geomatics for Green and Digital Transition*, E. Borgogno-Mondino e P. Zamperlin, Eds., em Communications in Computer and Information Science. Cham: Springer International Publishing, 2022, pp. 178–192. doi: 10.1007/978-3-031-17439-1_13.
[23] M. Taştan e M. N. Uğural, «Applicability Of Incineration Technology In Waste Management: Istanbul Case Study», *International Scientific and Vocational Studies Journal*, vol. 6, n.º 2, pp. 121–137, dez. 2022, doi: 10.47897/bilmes.1212938.
[24] L. Andeobu, S. Wibowo, e S. Grandhi, «Artificial intelligence applications for sustainable solid waste management practices in Australia: A systematic review», *Science of The Total Environment*, vol. 834, p. 155389, ago. 2022, doi: 10.1016/j.scitotenv.2022.155389.
[25] J. C. Prata, A. L. P. Silva, A. C. Duarte, e T. Rocha-Santos, «The road to sustainable use and waste management of plastics in Portugal», *Front. Environ. Sci. Eng.*, vol. 16, n.º 1, p. 5, out. 2021, doi: 10.1007/s11783-021-1439-x.
[26] M. M. Conceição, M. D. F. P. Alves, J. T. P. Conceição, e P. Castro, «Resíduo Urbano em Portugal – Uma Análise dos Resíduos Plástico», *RSD*, vol. 7, n.º 8, p. e1278382, mai. 2018, doi: 10.17648/rsd-v7i8.382.
[27] «Pacto Português para os Plásticos | APA». Acedido: 2 de março de 2024. [Em linha]. Disponível em: https://apambiente.pt/apa/pacto-portugues-para-os-plasticos
[28] «Pacto Português para os Plásticos | Juntos, por uma economia circular para os plásticos, em Portugal.», Pacto Português para os Plásticos /> <meta property=. Acedido: 3 de março de 2024. [Em linha]. Disponível em: https://pactoplasticos.pt/
[29] R. C. Marques, N. F. Da Cruz, e P. Carvalho, «Assessing and exploring (in)efficiency in Portuguese recycling systems using non-parametric methods», *Resources, Conservation and Recycling*, vol. 67, pp. 34–43, out. 2012, doi: 10.1016/j.resconrec.2012.07.005.
[30] A. J. C. Marinho, J. Couto, e A. Camões, «Current state, comprehensive analysis and proposals on the practice of construction and demolition waste reuse and recycling in Portugal», *Journal of Civil Engineering and Management*, vol. 28, n.º 3, Art. n.º 3, mar. 2022, doi: 10.3846/jcem.2022.16447.
[31] M. Costa, R. C. Madureira, e C. Dias-Ferreira, «A social receptiveness analysis on smart waste management – A case study for Portugal», *Sustainable Chemistry and Pharmacy*, vol. 28, p. 100735, set. 2022, doi: 10.1016/j.scp.2022.100735.
[32] V. Oliveira, V. Sousa, e C. Dias-Ferreira, «Dataset of socio-economic and waste collection indicators for Portugal at municipal level», *Data in Brief*, vol. 22, pp. 658–661, fev. 2019, doi: 10.1016/j.dib.2018.12.069.
[33] A. S. Silva *et al.*, «Node Assembly for Waste Level Measurement: Embrace the Smart City», em *Optimization, Learning Algorithms and Applications*, vol. 1754, A. I. Pereira, A. Košir, F. P. Fernandes, M. F. Pacheco, J. P. Teixeira, e R. P. Lopes, Eds., em Communications in Computer and

Information Science, vol. 1754. , Cham: Springer International Publishing, 2022, pp. 604–619. doi: 10.1007/978-3-031-23236-7_42.
[34] B. Fang *et al.*, «Artificial intelligence for waste management in smart cities: a review», *Environ Chem Lett*, vol. 21, n.º 4, pp. 1959–1989, ago. 2023, doi: 10.1007/s10311-023-01604-3.
[35] M. Abdallah, M. Abu Talib, S. Feroz, Q. Nasir, H. Abdalla, e B. Mahfood, «Artificial intelligence applications in solid waste management: A systematic research review», *Waste Management*, vol. 109, pp. 231–246, mai. 2020, doi: 10.1016/j.wasman.2020.04.057.
[36] S. Poudel e P. Poudyal, «Classification of Waste Materials using CNN Based on Transfer Learning», em *Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation*, Kolkata India: ACM, dez. 2022, pp. 29–33. doi: 10.1145/3574318.3574345.
[37] Nilu Kumari, Sneha Pandey, Amit Kumar Pandey, e Mainak Banerjee, «Role of Artificial Intelligence in Municipal Solid Waste Management», doi: https://doi.org/10.37745/bjmas.2022.00180.
[38] Rajalakshmi J., Sumangali K., Jayanthi J., e Muthulakshmi K, «Artificial Intelligence with Earthworm Optimization Assisted Waste Management System for Smart Cities», jan. 2023, doi: 10.30955/gnj.004712.
[39] P. V. Rao *et al.*, «IoT based Waste Management for Smart Cities», *International Conference on Computer Communication and Informatics*, 2020.
[40] T. J. Sheng *et al.*, «An Internet of Things Based Smart Waste Management System Using LoRa and Tensorflow Deep Learning Model», *IEEE Access*, vol. 8, pp. 148793–148811, 2020, doi: 10.1109/ACCESS.2020.3016255.
[41] M. F. Ashby e D. R. H. Jones, *Engineering materials. 2: An introduction to microstructures, processing and design*, 2. ed., Repr. Oxford: Butterworth-Heinemann, 2005.
[42] «Directive - 2008/98 - EN - Waste framework directive - EUR-Lex». Acedido: 1 de setembro de 2024. [Em linha]. Disponível em: https://eur-lex.europa.eu/eli/dir/2008/98/oj
[43] P. T. Williams, *Waste treatment and disposal*, 2nd ed. Chichester, West Sussex, England ; Hoboken, NJ, USA: Wiley, 2005.
[44] A. K. Jain, «Data Clustering: 50 Years Beyond K-means», em *Machine Learning and Knowledge Discovery in Databases*, vol. 5211, W. Daelemans, B. Goethals, e K. Morik, Eds., em Lecture Notes in Computer Science, vol. 5211. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 3–4. doi: 10.1007/978-3-540-87479-9_3.
[45] «Gesamb – Gestão Ambiental e de Residuos». Acedido: 11 de setembro de 2024. [Em linha]. Disponível em: https://gesamb.pt/
[46] W. Wellbrock, D. Ludin, L. Röhrle, e W. Gerstlberger, «Sustainability in the automotive industry, importance of and impact on automobile interior – insights from an empirical survey», *Int J Corporate Soc Responsibility*, vol. 5, n.º 1, p. 10, dez. 2020, doi: 10.1186/s40991-020-00057-z.
[47] «Gestão de Resíduos na pandemia de COVID-19 | Intial PT». Acedido: 3 de setembro de 2024. [Em linha]. Disponível em: https://www.initial.com/pt/blog/gestao-de-residuos-hospitalares/gestao-de-residuos-na-pandemia-de-covid-19
[48] «Portugal Cities Database | Simplemaps.com». Acedido: 5 de setembro de 2024. [Em linha]. Disponível em: https://simplemaps.com/data/pt-cities
[49] «Atividade da empres e objeto social», justiça.gov.pt. Acedido: 5 de setembro de 2024. [Em linha]. Disponível em: https://justica.gov.pt/Como-criar-uma-empresa-online/Atividade-da-empresa-e-objeto-social
[50] N. Islam, M. Sandanayake, S. Muthukumaran, e D. Navaratna, «Review on Sustainable Construction and Demolition Waste Management—Challenges and Research Prospects», *Sustainability*, vol. 16, n.º 8, p. 3289, abr. 2024, doi: 10.3390/su16083289.
[51] «Principal component analysis: a review and recent developments». Acedido: 8 de setembro de 2024. [Em linha]. Disponível em: https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2015.0202
[52] P. Bholowalia, «EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN», *International Journal of Computer Applications*, vol. 105, n.º 9.

[53] «Mastering Cluster Validation with Silhouette Scores and Visualization in Python», CodeSignal Learn. Acedido: 8 de setembro de 2024. [Em linha]. Disponível em: https://learn.codesignal.com
[54] «Portal do INE». Acedido: 8 de setembro de 2024. [Em linha]. Disponível em: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=10376048&PUBLICACOESmodo=2

Apêndices

a) Nome das variáveis iniciais no dataset original

Nome das variáveis	
fk_guia_acompanhamento_residuos	destinatario_estabelecimento
data_criacao	ler_codigo
egar	ler_perigosidade
validacao	tipo_op_codigo
criada_por	quantidade
qualidade	tipo_op_recebido_codigo
produtor_origem	quantidade_recebida
produtor_nif	ler_recebido_codigo
produtor_nome	ler_recebido_perigosidade
produtor_local	autorizada
data_estado	utilizador_criacao
estado	utilizador_alteracao
data_inicio_transporte	origem
data_fim_transporte	mercadorias_perigosas
dados_transportadores	grupo_hospitalar
destinatario_nif	num_plano_gestao_lamas
destinatario_nome	num_plano_gestao_lamas_recebido

b) Descrição dos códigos da variável tipo_op_codigo e tipo_op_recebido_codigo

Código	Operações
R1	Utilização principal como combustível ou outro meio de produção de energia
R2	Recuperação/Regeneração de solventes
R3	Reciclagem/recuperação de substâncias orgânicas não utilizadas como solventes (incluindo digestão anaeróbia e ou compostagem e outros processos de transformação biológica)
R4	Reciclagem/recuperação de metais e compostos metálicos
R5	Reciclagem/recuperação de outros materiais inorgânicos
R6	Regeneração de ácidos ou bases
R7	Valorização de componentes utilizados na redução da poluição
R8	Valorização de componentes de catalisadores
R9	Refinação de óleos e outras reutilizações de óleos
R10	Tratamento do solo para benefício agrícola ou melhoramento ambiental
R11	Utilização de resíduos obtidos a partir de qualquer das operações enumeradas de R1 a R10
R12	Troca de resíduos com vista a submetê-los a uma das operações enumeradas de R1 a R11
R13	Armazenamento de resíduos destinados a uma das operações enumeradas de R1 a R12 (com exclusão do armazenamento temporário, antes da recolha, no local onde os resíduos foram produzidos)
D1	Depósito no solo, em profundidade ou à superfície (por exemplo, em aterros, etc.)
D2	Tratamento no solo (por exemplo, biodegradação de efluentes líquidos ou de lamas de depuração nos solos, etc.)
D3	Injeção em profundidade (por exemplo, injeção de resíduos por bombagem em poços, cúpulas salinas ou depósitos naturais, etc.)
D4	Lagunagem (por exemplo, descarga de resíduos líquidos ou de lamas de depuração em poços, lagos naturais ou artificiais, etc.)
D5	Depósitos subterrâneos especialmente concebidos (por exemplo, deposição em alinhamentos de células que são seladas e isoladas umas das outras e do ambiente, etc.)
D6	Descarga para massas de água, com exceção dos mares e dos oceanos
D7	Descargas para os mares e ou oceanos, incluindo inserção nos fundos marinhos
D8	Tratamento biológico não especificado em qualquer outra parte do presente apêndice que produza compostos ou misturas finais rejeitados por meio de qualquer das operações enumeradas de D1 a D12
D9	Tratamento físico-químico não especificado em qualquer outra parte do presente apêndice que produza compostos ou misturas finais rejeitados por meio de qualquer das operações enumeradas de D1 a D12 (por exemplo, evaporação, secagem, calcinação, etc.)
D10	Incineração em terra
D11	Incineração no mar
D12	Armazenamento permanente (por exemplo, armazenamento de contentores numa mina, etc.)
D13	Mistura anterior à execução de uma das operações enumeradas de D1 a D12
D14	Reembalagem anterior a uma das operações enumeradas de D1 a D13
D15	Armazenamento antes de uma das operações enumeradas de D1 a D14 (com exclusão do armazenamento temporário, antes da recolha, no local onde os resíduos foram produzidos)
AP	Armazenamento preliminar

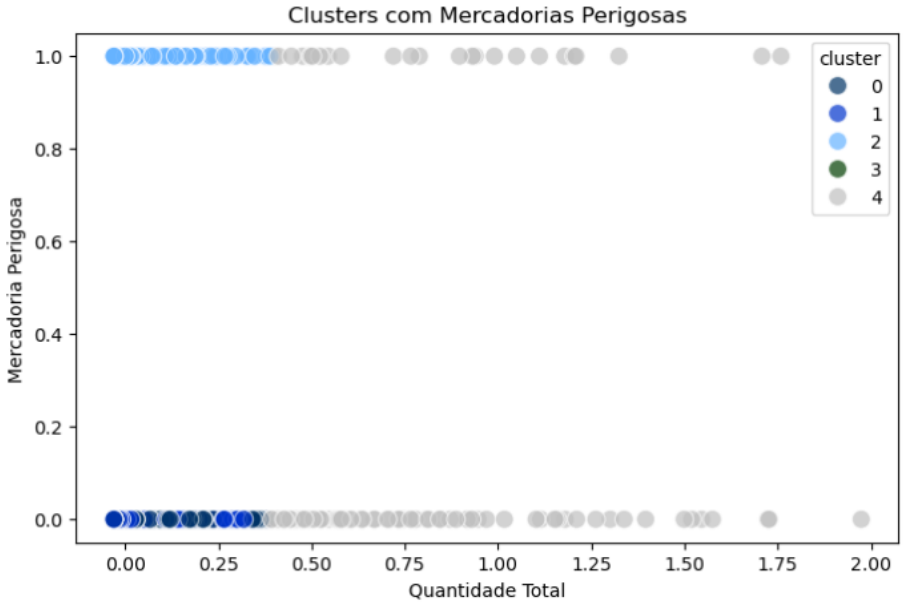
c) Quantidade de resíduos enviada por Distrito para cada ano em análise

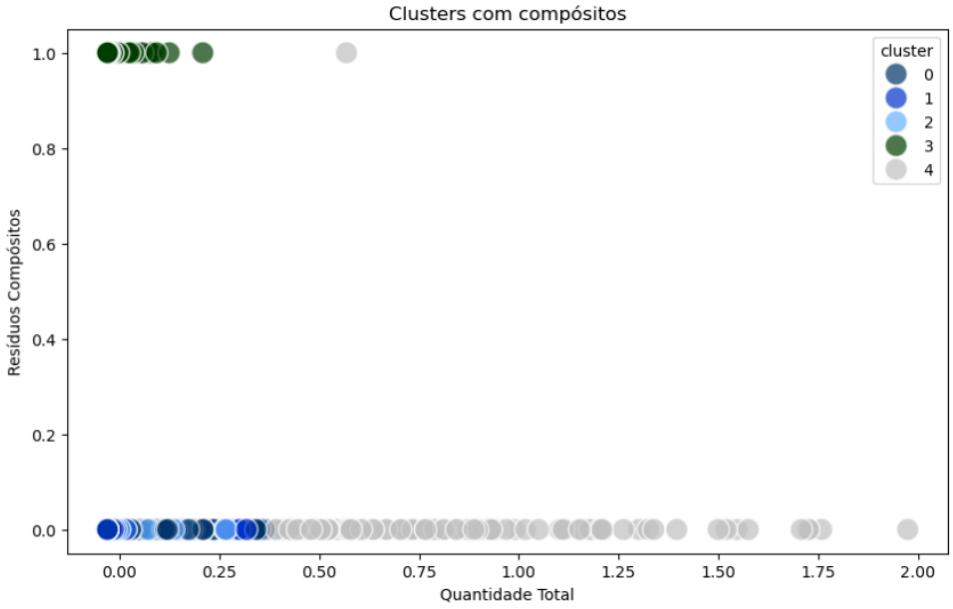
Distrito	2019	2020	2021	2022
Porto	28930,1	27130	29339	31458,9
Lisboa	25901,4	14647,7	17567,7	17615
Setúbal	10877,6	9293,7	17587,3	14834,6
Leiria	4937,3	5445,3	6726,7	7368,3
Aveiro	5841	6691,8	7526,4	4741,7
Santarém	4857	5819,9	6665,7	6177,3
Braga	3648,6	3241,6	4051,8	4741,7
Faro	2098,9	1897,1	1884,4	2018,5
Viseu	1039,9	952,1	2004	2040,2
Coimbra	1352	1333,1	1691,4	1980,9
Castelo Branco	1340,8	1444,4	1536,4	1728,7
Ilha da Madeira (Madeira)	995,5	450	553,6	1732,8
Beja	689,4	675,3	766,7	793,9
Portalegre	443,1	414,7	514,7	672,5
Évora	318,9	370,9	475,5	542
Viana do Castelo	708,5	434	489	527,3
Bragança	244,6	137,1	177,9	173,1
Vila Real	374,9	156,7	198,2	195,9
Guarda	108	105,2	328,4	132,5
Ilha de Porto Santo (Madeira)	0	0	0	0,9

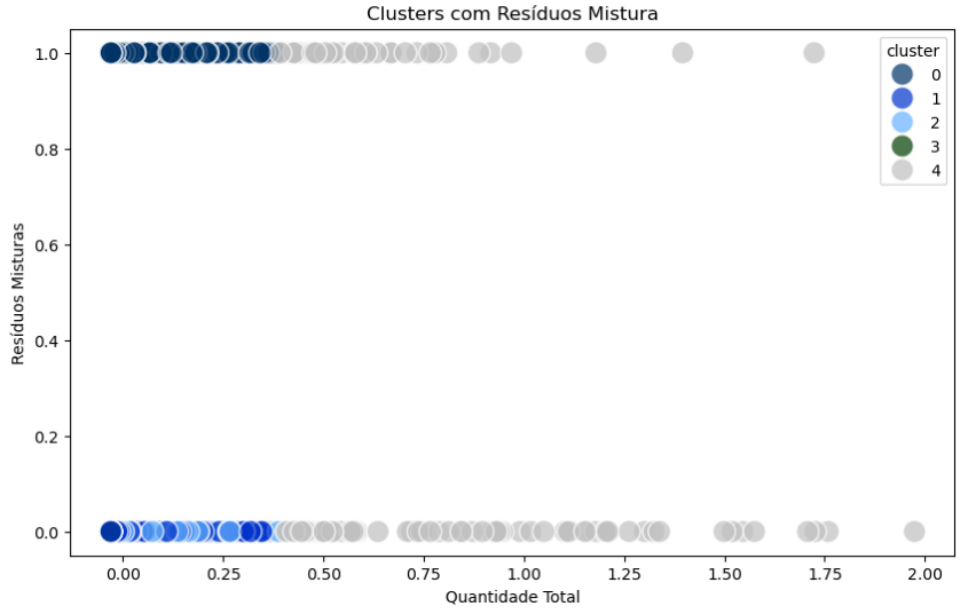
d) Estatísticas Descritivas por Cluster dos Produtores sem CAEs incluídos

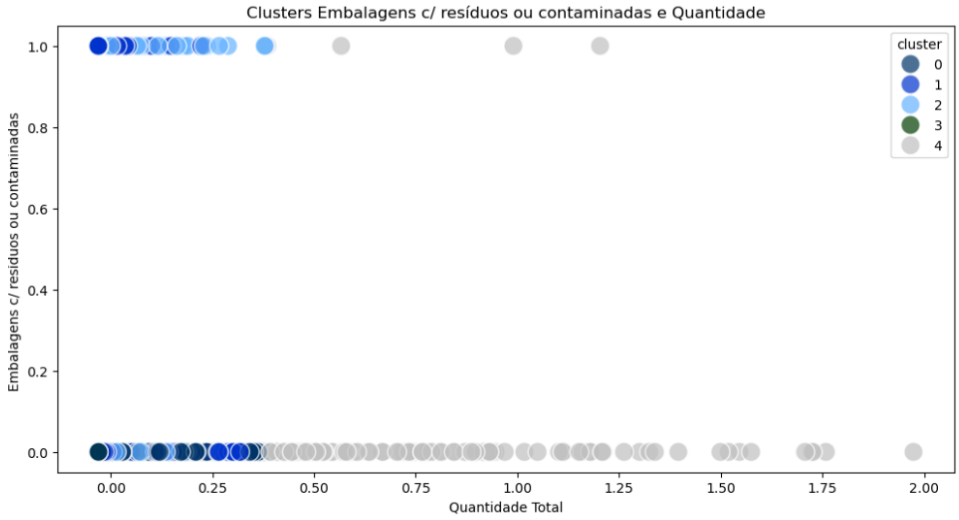
Cluster	Média Quantidade	Média Mercadoria Perigosa	Média Compósitos	Média Misturas	Média Embalagens Contaminadas
0	-0,008611	0	0	1	0
1	-0,025433	0	0	0	0,845422
2	-0,00897	1	0	0,000755	0,821887
3	-0,025105	0	0	1	0
4	0,738699	0,208	0,008	0,312	0,032

e) Distribuição dos clusters por variável com base na quantidade enviada pelo produtor (Clustering excluindo o CAE)

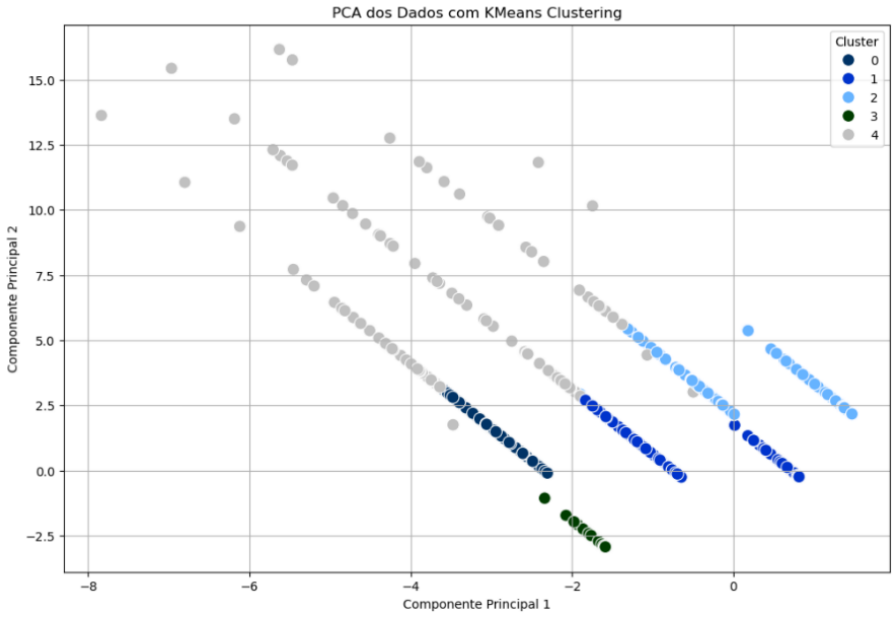








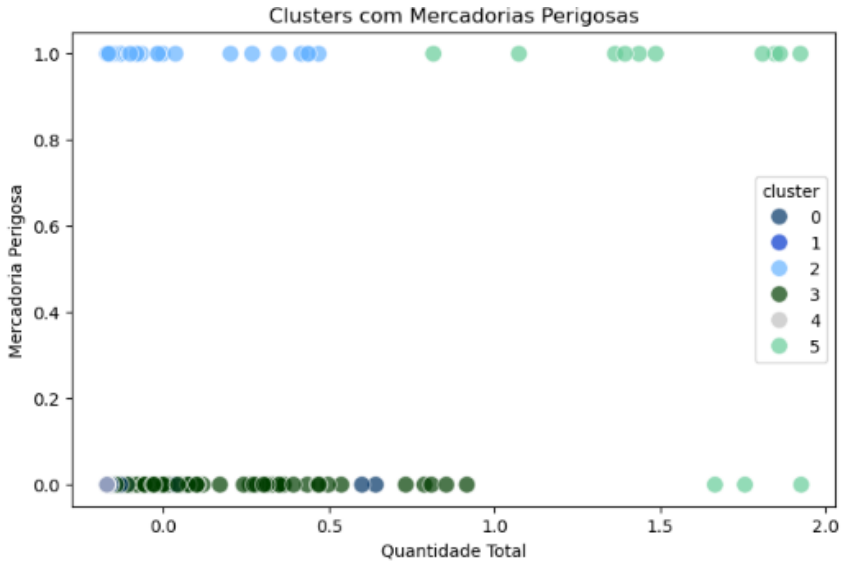
f) PCA referente ao clustering de produtores excluindo os CAEs

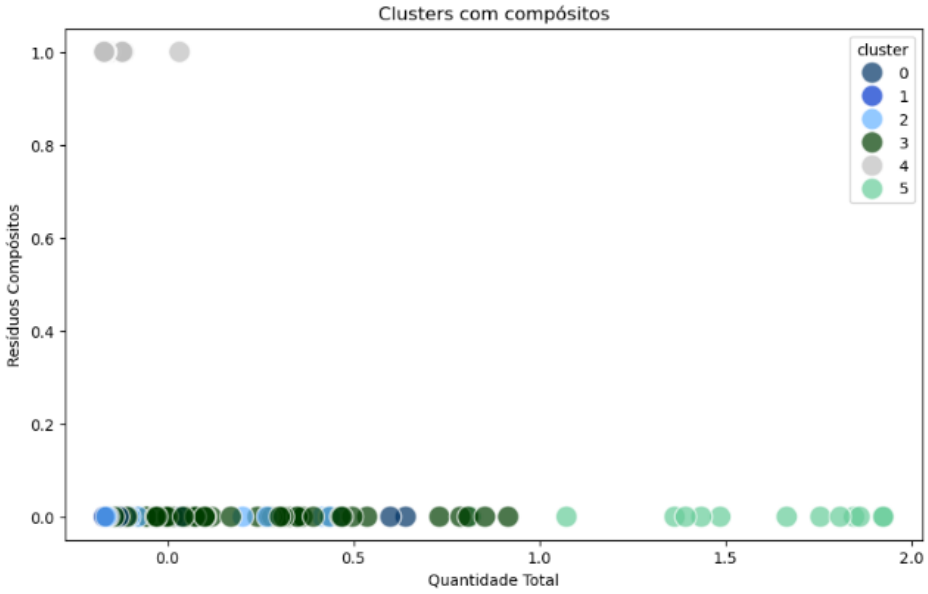


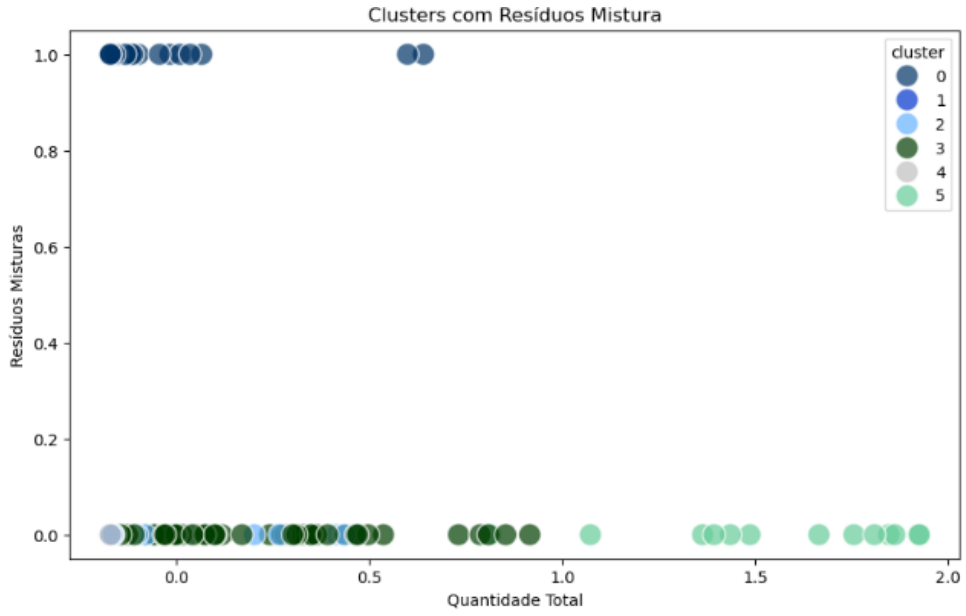
g) Estatísticas Descritivas por Cluster dos Transformadores sem CAEs incluídos

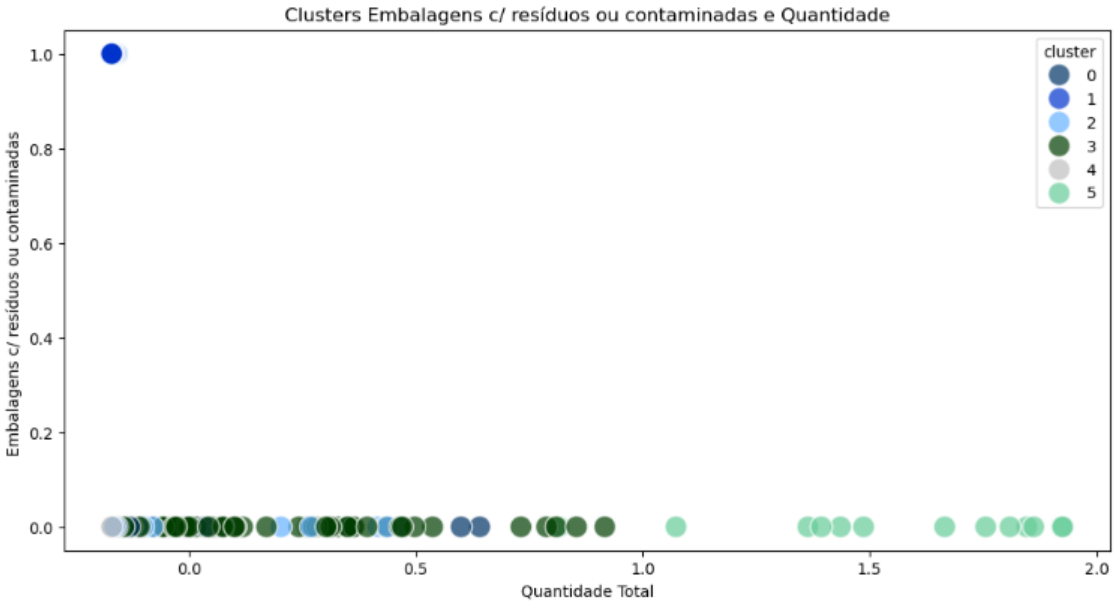
Cluster	Média Quantidade	Média Mercadoria Perigosa	Média Compósitos	Média Misturas	Média Embalagens Contaminadas
0	-0,149941	0	0	1	0
1	-0,170330	0	0	0	1
2	-0,014914	1	0	0,03125	0,28125
3	-0,064987	0	0	0	0
4	-0,166244	0	1	0	0
5	1,565909	0,769231	0	0	0

h) Distribuição dos clusters por variável com base na quantidade recebida pelo transformador (Clustering excluindo CAE)

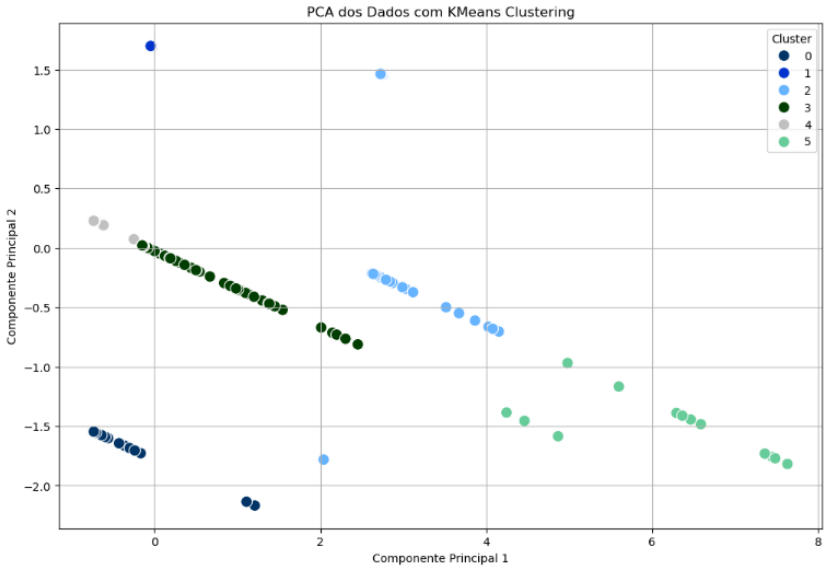








i) PCA referente ao clustering de transformadores excluindo os CAEs



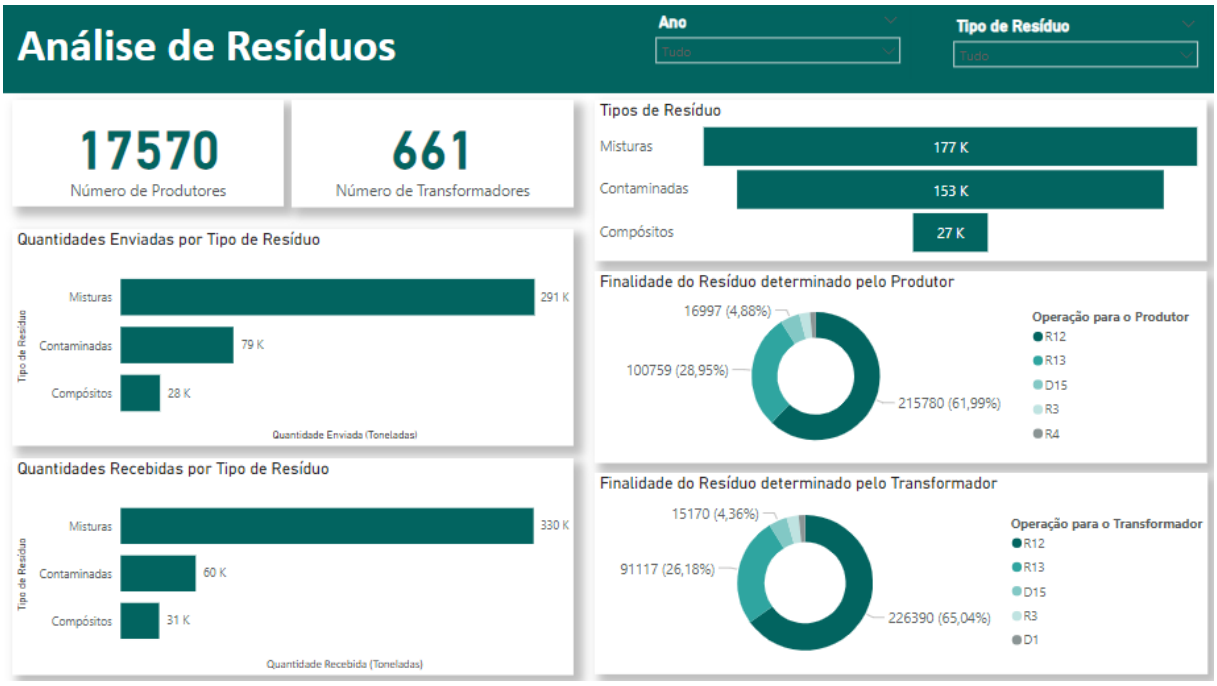
j) Estatísticas Descritivas por Cluster dos Produtores com CAEs incluídos

Cluster	Quantidade	Merc. Perigosa	Compósitos	Misturas	Contaminadas	CAE 1	CAE 2	CAE 3	CAE 4	CAE 5	CAE 6	CAE 7	CAE 8	CAE 9
0	0,001	0,061203	0	0	0	0	0	0	1	0	0	0	0	0
1	-0,02968	0,048809	0	0	1	0	0	0	1	0	0	0	0	0
2	-0,00691	0,159851	0	0	0,797398	1	0	0	0	0	0	0	0	0
3	-0,00742	0,171277	0	0	0,72766	0	1	0	0	0	0	0	0	0
4	-0,0134	0,084469	0	0,171662	0,689373	0	0	0	0	0	0	0	1	0
5	0,037183	0,042373	0,025424	0,432203	0,322034	0	0	0	0	0	1	0	0	0
6	-0,02916	0,371429	0,004762	0,052381	0,885714	0	0	0	0	0	0	0	0	1
7	-0,00532	0,061151	0,010791	0,284173	0,535971	0	0	0	0	1	0	0	0	0
8	0,005278	0,147059	0	0,128431	0,689216	0	0	1	0	0	0	0	0	0
9	-0,02205	0,187251	0,027888	0,067729	0,804781	0	0	0	0	0	0	1	0	0
10	-0,0213	0	1	0	0	0,156051	0,16879	0,035032	0,621019	0	0	0	0,019108	0
11	0,008546	0,00062	0	1	0	0,205081	0,189591	0	0,605328	0	0	0	0	0

k) Estatísticas Descritivas por Cluster dos Transformadores com CAEs incluídos

Cluster	Quantidade	Merc. Perigosa	Compósitos	Misturas	Contaminadas	CAE 1	CAE 2	CAE 3	CAE 4	CAE 5	CAE 6	CAE 7	CAE 8	CAE 9
0	-0,012871	0	0,191176	0,367647	0	1	0	0	0	0	0	0	0	0
1	-0,170318	0	0	0	1	0,296703	0	0	0,703297	0	0	0	0	0
2	-0,131265	0,020408	0,102041	0,367347	0,22449	0	0	1	0	0	0	0	0	0
3	0,39155	0,97619	0	0,02381	0,214286	0,071429	0,547619	0	0,380952	0	0	0	0	0
4	-0,12934	0	0,16129	0,225806	0,225806	0	1	0	0	0	0	0	0	0
5	-0,170513	0	0	1	0	0	0	0	0	0	1	0	0	0
6	-0,101081	0	0	0,318919	0	0	0	0	1	0	0	0	0	0
7	-0,170552	0	1	0	0	0	0	0	0	0	0	0	0	1
8	-0,166339	0	0	0,428571	0,571429	0	0	0	0	0	0	0	1	0
9	-0,170694	0	0,666667	0	0	0	0	0	0	0	0	1	0	0
10	-0,169598	0	1	0	0	0	0	0	1	0	0	0	0	0
11	-0,170655	0	0	1	0	0	0	0	0	1	0	0	0	0

l) Dashboard



m) Repositório do Código desenvolvido: [joaozr/APA: Tese João Ribeiro (github.com)]