



**Politécnico
de Viseu**

Escola Superior
de Tecnologia
e Gestão de Lamego

P/Viseu - Centro de Lamego

Simão Costa Santos

ESTGL | 2025

Otimização de rotas de recolha de lixo – Caso de estudo: Município de Cascais

Otimização de rotas de recolha de lixo - Caso de Estudo: Município de Cascais



Simão Costa Santos

Estudante N.º 25766

Lamego, 22 de junho de 2025



Simão Costa Santos
25766

**Otimização de rotas de recolha de lixo - Caso de Estudo:
Município de Cascais**

Licenciatura em Engenharia Informática e Telecomunicações

Orientador: Prof. Ricardo Gama

Co-orientador: Prof. José Lousado

Lamego, 22 de junho de 2025

Termo de Responsabilidade / Declaração de Autoria / Declaração Anti Plágio

Eu, Simão Costa Santos, n.º de aluno 25766 declaro que sou o único autor(a) deste relatório de projeto/estágio apresentado para obtenção de aprovação na unidade curricular de Projeto Final da Licenciatura em Engenharia Informática e Telecomunicações ministrada na Escola Superior de Tecnologia e Gestão de Lamego.

Declaro por minha honra que o presente relatório é original, que não contém qualquer plágio, uma vez identifiquei de forma clara e citei corretamente trabalhos de outros autores que foram utilizados; no caso de ter utilizado frases retiradas de trabalhos de outros autores referenciei-as devidamente ou, se as redigi com palavras diferentes, indiquei o original de onde foram adaptadas, bem como os contributos significativos de outras obras publicadas da minha autoria.

Nesta base, declaro que tenho consciência que a prática de plágio pode gerar responsabilidade civil, criminal e disciplinar, bem como a reprovação a esta unidade curricular ou, no limite, à retirada do grau académico e que constitui uma grave violação da ética académica.

Por fim, declaro que este trabalho, em parte ou no todo, não foi previamente submetido para outra finalidade.

Lamego, 22 de junho de 2025

Assinatura

Simão Costa Santos

AGRADECIMENTOS

Gostaria de expressar o meu sincero agradecimento ao Sr. Professor Ricardo Gama por ter aceitado ser o meu orientador neste projeto final. A sua constante disponibilidade, interesse e dedicação foram essenciais ao longo de todo o processo. Sempre pronto a esclarecer dúvidas, a sua análise minuciosa do conteúdo do trabalho e as sugestões apresentadas contribuíram significativamente para a qualidade deste trabalho. A sua orientação foi fundamental não só na execução técnica do projeto, mas também no planeamento e estruturação das várias etapas. Sem o seu apoio, dificilmente teria alcançado os resultados que obtive.

Agradeço também ao Sr. Professor José Lousado pela colaboração prestada ao longo do desenvolvimento do projeto.

Agradeço igualmente a todos os professores que fizeram parte do meu percurso académico, pelo conhecimento e experiência que partilharam ao longo destes anos, e que foram determinantes para a minha formação pessoal e profissional.

Não poderia deixar de agradecer aos colegas que me acompanharam durante este trajeto. O apoio mútuo, a partilha de ideias e os momentos vividos em conjunto marcaram de forma positiva esta etapa da minha vida.

RESUMO

Este trabalho incide sobre a otimização de rotas de recolha de lixo, com o objetivo de melhorar a eficiência operacional deste tipo de serviço. Para tal, foi desenvolvido um estudo composto por três componentes principais: análise exploratória dos dados fornecidos pelo município de Cascais, modelo preditivo e análise dos resultados. A análise dos dados permitiu extrair informações relevantes sobre os circuitos, padrões de recolha e quantidades de lixo recolhido. O modelo preditivo prevê os níveis de enchimento e constrói rotas otimizadas e por fim, analisou-se os resultados obtidos. A integração destes módulos, visa oferecer uma ferramenta mais eficaz para recolhas mais sustentáveis.

Palavras-chave: *Otimização de rotas, Análise de Dados, Modelo preditivo*

ÍNDICE GERAL

INTRODUÇÃO.....	2
1.ESTADO DA ARTE	4
1.1 Custos e impacto ambiental da operação logística	4
1.2 O papel da otimização na gestão operacional	5
1.3 Modelos matemáticos aplicados à otimização da recolha de resíduos	6
1.3.1 Formulação Geral	7
1.3.2 Complexidade do VRP	7
1.3.3 Variantes do VRP.....	8
1.4 Aplicação prática ao contexto do projeto.....	8
2.DEFINIÇÃO DO PROBLEMA	9
3.ANÁLISE DE DADOS	12
3.1 Preparação e Tratamento dos Dados.....	12
3.2 Cruzamento e verificação de IDs.....	15
3.3 Normalização dos nomes dos circuitos.....	16
3.4 Estruturação dos dados por Circuito de Recolha	17
3.5 Tabela de registo dos níveis por tipo e recolha.....	19
3.6 Tabela com as Observações e Recolhas.....	20
3.7 Taxas de enchimento e estatísticas	22
3.8 Cálculo dos níveis estimados no início do dia e no final do dia	23
3.9 Ficheiro com os circuitos realizados em Cascais.....	25
4.OTIMIZAÇÃO DAS ROTAS.....	27
4.1 Previsão de níveis de enchimento usando <i>Machine Learning</i>	29
4.2 Resultados Obtidos Do Modelo Preditivo	30

4.2.1 Quantidade de lixo não recolhido.....	30
4.2.2 Distância	33
4.3 Conclusão do resultado obtido.....	36
5. DESENVOLVIMENTO DE UM DASHBOARD.....	37
5.1 Estrutura do Dashboard	37
5.2 Módulo Análise de Dados.....	38
5.2.1 Frequência semanal e médias por enchimento	38
5.2.2 Distribuição de pontos por número de circuito	40
5.2.3 Quantidade de pontos por circuito.....	41
5.2.4 Visualização geográfica dos pontos por circuito.....	42
5.2.5 Gráficos de um ponto aleatório por circuito.....	43
5.2.6 Evolução temporal dos níveis de enchimento por circuito.....	46
5.2.7 Quantidade de lixo recolhido por tipo e circuito vs total	48
5.2.8 Evolução temporal do lixo recolhido por tipo	49
5.3 Módulo das Rotas Reais	52
5.4 Módulo Otimização de Rotas (Solver)	55
6. OUTROS TESTES REALIZADOS.....	61
6.1 Época de Treino	61
6.2 Época de Validação	63
6.3 Conclusão do resultado obtido.....	66
CONCLUSÃO E TRABALHOS FUTUROS	67
REFERÊNCIAS	68

ÍNDICE DE FIGURAS

Figura 1 - Esquema VRP	6
Figura 2 - Exemplificação de uma rota falhada devido a uma previsão incorreta	10
Figura 3 - Registo de níveis de enchimentos sucessivos	11
Figura 4 - Ficheiro com as coordenadas dos pontos.....	13
Figura 5 - Estrutura do ficheiro das operações de recolha	14
Figura 6 - Nomenclatura dos circuitos antes da normalização	16
Figura 7 - Nomenclatura dos circuitos após a normalização.....	17
Figura 8 - Ficheiro de agregação por Data, Circuito, Tipo e Pontos.....	18
Figura 9 - Tabela de registo dos níveis por tipo e recolha.....	19
Figura 10 - Tabela com as observações e recolhas.....	21
Figura 11 - Distribuição dos níveis de enchimento usando a taxa de enchimento média ...	24
Figura 12 - Distribuição dos níveis de enchimento usando a taxa de enchimento mediana	24
Figura 13 - Ficheiro fornecido para ajudar a identificar as rotas.....	26
Figura 14 - Tabela com as soluções do solver para o ano todo	28
Figura 15 - Distribuição do lixo não recolhido para o tipo Embalagens	31
Figura 16 - Distribuição da percentagem do lixo não recolhido para o tipo Papel	32
Figura 17 - Distribuição da percentagem do lixo não recolhido para o tipo Vidro	32
Figura 18 - Distribuição da distância para o tipo Embalagens	34
Figura 19 - Distribuição da distância para o tipo Vidro	35
Figura 20 - Distribuição da distância para o tipo Papel.....	35
Figura 21 - Frequência semanal e média dos enchimentos para o Circuito 02	39
Figura 22 - Distribuição de pontos por número de circuitos	40
Figura 23 - Quantidade de pontos por circuito	41

Figura 24 - Representação da caixa de seleção	42
Figura 25 - Mapa com os pontos do Circuito 10	43
Figura 26 - Frequência de papel para o ponto 3712	44
Figura 27 - Série temporal do nível de enchimento de papel do ponto 3712	45
Figura 28 - Intervalo de Recolhas de papel para o ponto 3712	45
Figura 29 - Evolução temporal (papel) do Circuito 12.....	46
Figura 30 - Evolução temporal (embalagens) do Circuito 12	47
Figura 31 - Evolução temporal (vidro) do Circuito 12	47
Figura 32 - Quantidade total de lixo recolhido por tipo e circuito	48
Figura 33 - Proporção de lixo recolhido por tipo de resíduo.....	49
Figura 34 - Gráfico da evolução temporal do lixo recolhido para Papel.....	50
Figura 35 - Gráfico da evolução temporal do lixo recolhido para Embalagens	50
Figura 36 - Gráfico da evolução temporal do lixo recolhido para Vidro	51
Figura 37 - Filtro para seleção da rota a visualizar.....	53
Figura 38 - Rota efetuada no dia 02/01/2024, no Circuito 03 para a recolha de Embalagens	54
Figura 39 - Estatísticas da rota	55
Figura 40 - Representação das caixas de seleção do solver para o dia, circuito e tipo	56
Figura 41 - Rota de Início efetuada no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens.....	57
Figura 42 - Rota de Fim efetuada no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens.....	58
Figura 43 - Sub-Rota efetuada no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens.....	59

Figura 44 - Estatísticas das rotas efetuadas no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens.....	60
Figura 45 – Média do retorno da rede neuronal na época de treino	61
Figura 46 - Média de pontos não visitados pela rede neuronal na época de treino	62
Figura 47 - Média de lixo não recolhido pela rede neuronal na época de treino.....	63
Figura 48 - Média do retorno da rede neuronal para a época de validação	64
Figura 49 - Média de pontos não visitados pela rede neuronal na época de validação	65
Figura 50 - Média da quantidade de lixo não recolhido pela rede neuronal durante a fase de validação	66

ÍNDICE DE TABELAS

Tabela 1 - Média geral da percentagem da quantidade de lixo não recolhido	30
Tabela 2 – Média da percentagem da quantidade de lixo não recolhido por tipo e rota	31
Tabela 3 - Média geral da distância.....	33
Tabela 4 – Média da distância por tipo e rota.....	34

Lista de abreviaturas, acrónimos e siglas

ID	<i>Identity</i> Identidade
NaN	<i>Not a Number</i> Não é Número
NP	<i>Nondeterministic Polynomial</i> Polinômio não determinístico
PCVRP	<i>Prize Collecting Capacitated Vehicle Routing Problem</i> Problema de Roteamento com Capacidade e Prêmios
VRP	<i>Vehicle Routing Problem</i> Problema de Roteamento de Veículos

INTRODUÇÃO

A gestão de resíduos urbanos enfrenta desafios crescentes devido ao aumento contínuo da produção de resíduos, à urbanização acelerada e à pressão por soluções ambientalmente sustentáveis. A recolha de resíduos sólidos representa uma das atividades mais dispendiosas da gestão municipal, podendo consumir mais de 50% dos custos totais do sistema (de Moraes e Soares, 2024). Quando mal planeada, essa operação pode resultar em recolhas ineficazes, trajetos redundantes e um desperdício significativo de recursos logísticos e energéticos.

No município de Cascais, a recolha de resíduos é monitorada através do registo dos níveis de enchimento dos contentores (Papel, Embalagens e Vidro) quando a recolha é efetuada. Embora esta abordagem forneça dados históricos valiosos, apresenta uma limitação crítica: entre duas recolhas consecutivas, não existe informação sobre a evolução do nível de enchimento, ou seja, até os operadores chegarem fisicamente ao ponto de recolha, desconhece-se o estado real do contentor. Esta incerteza representa um desafio significativo para o planeamento eficiente das rotas. Se, ao chegar ao local, o contentor estiver mais cheio do que o previsto, pode não haver capacidade suficiente no veículo para completar o percurso planeado ou ter de deixar de recolher resíduos dos contentores na totalidade para concluir a rota. Isto torna todo o processo ineficiente e pouco sustentável, uma vez que pode resultar em contentores a transbordar, rotas mal aproveitadas e um uso desnecessário de recursos logísticos e energéticos.

Com o intuito de resolver este problema, desenvolveu-se um modelo preditivo baseado em técnicas de *machine learning*. Este modelo tem como objetivo estimar, para cada ponto de recolha, o nível de enchimento esperado no início do dia, com base nos níveis registados nas últimas recolhas e no intervalo de tempo decorrido entre elas até ao dia em questão. Adicionalmente, estima-se também o nível final esperado para esse dia, considerando as recolhas realizadas nesse dia. Estas previsões permitem um planeamento mais eficiente das rotas de recolha, ajustado ao comportamento esperado dos contentores, contribuindo para a redução de custos operacionais e para uma maior sustentabilidade do sistema. A integração do modelo com o solver PyVRP possibilitou ainda a criação de rotas otimizadas em relação à distância e à quantidade de lixo não recolhido, adaptadas às previsões, promovendo uma operação mais inteligente e eficaz.

Nos próximos capítulos será feito o estado da arte, uma análise exploratória dos dados fornecidos pela Câmara Municipal de Cascais, a descrição do desenvolvimento do modelo preditivo e a integração com a otimização de rotas e a análise dos resultados obtidos.

1. ESTADO DA ARTE

A gestão de resíduos é um dos principais desafios enfrentados pelas cidades modernas, sendo diretamente influenciada pelo crescimento populacional, aumento do consumo e crescente exigência por soluções ambientalmente sustentáveis. Um dos componentes mais críticos deste processo é a operação logística de recolha de resíduos, que representa uma parcela significativa dos custos totais do sistema de gestão de resíduos.

Tradicionalmente, a recolha é realizada segundo rotas fixas, que não consideram as variações diárias na produção de resíduos, podendo resultar em ineficiências e custos elevados. Assim, a incorporação de dados em tempo real e a aplicação de técnicas preditivas surgem como soluções promissoras para aumentar a eficácia do planeamento e operação das rotas. Recentemente a inteligência artificial, Internet das Coisas (IoT) e análise de dados em tempo real, tem revolucionado o setor da gestão de resíduos, possibilitando abordagens mais inteligentes e baseadas em dados. Por exemplo (Prata et al., 2025) desenvolveram uma abordagem de otimização baseada em dados para planear operações de recolha de resíduos inteligentes, utilizando informações em tempo real fornecidas por sensores volumétricos instalados nos contentores.

1.1 Custos e impacto ambiental da operação logística

A operação logística associada às recolhas de resíduos constitui uma das componentes mais dispendiosas na gestão urbana de resíduos sólidos, representando frequentemente entre 50% a 70% do custo total do sistema. Estes custos estão relacionados principalmente com o consumo de combustível, manutenção dos veículos e tempo de operação. Por essa razão a otimização das rotas de recolha assume um papel de extrema importância para a redução de despesas operacionais e aumento da sustentabilidade económica do sistema.

Além dos custos financeiros, as rotas mal planeadas contribuem significativamente para o impacto ambiental negativo das atividades de recolha. O consumo excessivo de combustível por percursos desnecessariamente longos ou ineficientes resulta na emissão elevada de gases de efeito estufa, poluentes atmosféricos e ruído, afetando a qualidade do ar e o bem-estar das populações urbanas (de Moraes e Soares, 2024).

1.2 O papel da otimização na gestão operacional

A otimização é um elemento central na modernização dos sistemas de recolha de resíduos, que permite melhorar a eficiência operacional através da definição de rotas adaptadas às necessidades reais de serviço. Este processo visa encontrar soluções que minimizem custos, distâncias percorridas, lixo não recolhido e tempos de recolha, cumprindo as restrições existentes, como a capacidade dos veículos, tempos de trabalho e prioridades de recolha (Herrera-Granda et al., 2024).

O problema de otimização mais comum neste contexto é o *Vehicle Routing Problem* (VRP), que consiste em determinar o rotas em que se deve visitar um conjunto de localizações com a menor distância ou custo possível. Este problema, tem múltiplas variantes que se ajustam a diferentes cenários operacionais, incluindo janelas temporais, múltiplos depósitos, capacidades limitadas, entre outros (Leal & Pereira, 2022).

A otimização tem vários benefícios como:

- Melhoria da Eficiência e Redução de Custos: A escolha de uma frota e veículos adequados, bem como a otimização das rotas, proporcionam uma redução de custos estimada em 45%.
- Redução de Impactos Ambientais: Com a redução dos consumos de combustível e das emissões libertadas pelas viaturas, pode-se ter a uma diminuição de cerca de 25% nas emissões.
- Otimização de Tempo de Trabalho: Pequenas melhorias podem evitar a necessidade de horas extra de trabalho. Otimizar estas rotas poderia resultar num abatimento anual de 353.70 horas de trabalho.

1.3 Modelos matemáticos aplicados à otimização da recolha de resíduos

O VRP é um problema clássico da investigação operacional que tem como objetivo encontrar o conjunto ótimo de rotas para uma frota de veículos que deve servir um conjunto de clientes ou pontos de recolha, minimizando o custo total, por exemplo, distância percorrida, tempo ou número de veículos utilizados onde se respeita simultaneamente um conjunto de restrições operacionais (Liu et al., 2023). A Figura 1 demonstra uma representação esquemática do problema de roteamento de veículos.

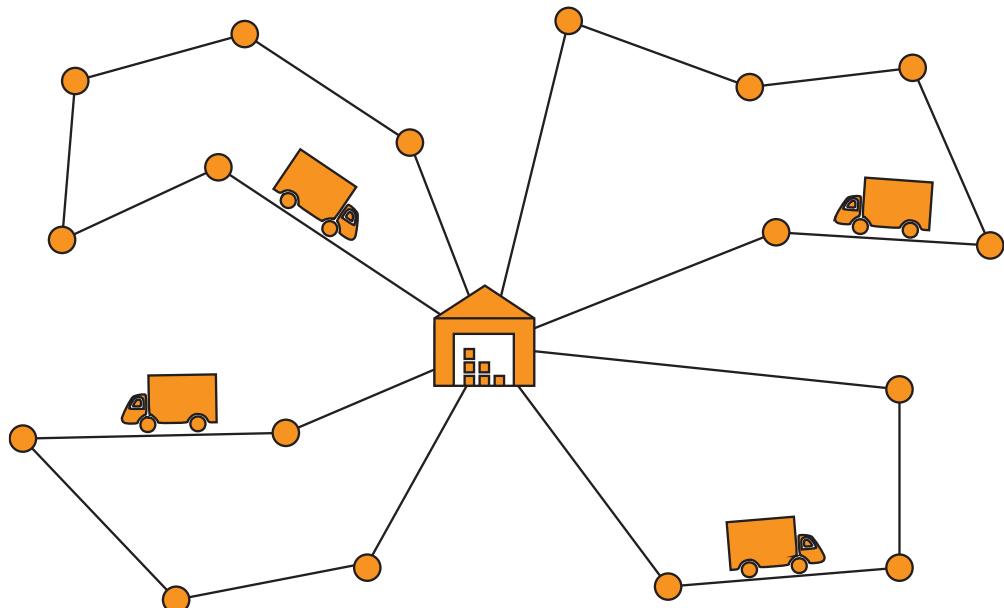


Figura 1 - Esquema VRP

Fonte: *A brief introduction to VRP - PyVRP 0.12.0a0 documentation*

1.3.1 Formulação Geral

De uma forma básica, pode-se descrever assim o VRP:

- Existe um armazém central (ou vários) de onde partem e regressam os veículos.
- Um conjunto de clientes/pontos com uma procura associada.
- Um conjunto de veículos (ou apenas um) com capacidade limitada.
- Um grafo com os custos associados às arestas (por exemplo, distâncias entre pontos ou tempo)
- Objetivo final: Encontrar o conjunto de rotas que satisfaçam as exigências, com o menor custo possível.

A complexidade computacional do VRP aumenta exponencialmente com o número de pontos a visitar, sendo considerado NP-difícil, o que justifica o desenvolvimento de abordagens heurísticas e métodos exatos. No ponto abaixo encontra-se uma breve explicação do que é um problema NP-difícil.

1.3.2 Complexidade do VRP

A complexidade computacional associada a este problema é um dos principais desafios na sua resolução prática. O VRP é classificado como um problema NP-difícil, o que significa que o número de combinações de rotas possíveis cresce exponencialmente com o aumento do número de clientes. Assim, torna-se computacionalmente inviável resolver instâncias de grande dimensão através de métodos exatos num tempo aceitável.

De forma geral, um problema NP-difícil caracteriza-se por não possuir, até ao momento, um algoritmo conhecido que o resolva em tempo polinomial para todos os casos. Além disso, a verificação de soluções pode ser também complexa. Por essa razão, a resolução de problemas como o VRP recorre frequentemente a técnicas heurísticas ou metaheurísticas, que embora não garantam a solução ótima, permitem obter soluções satisfatórias num tempo computacionalmente eficiente.

1.3.3 Variantes do VRP

Ao longo do tempo, diversas variantes do VRP para a recolha de resíduos, foram desenvolvidas para refletir cenários reais, mas neste contexto a mais relevante é o Waste Collection Vehicle Routing Problem (WCVRP) (Li et al., 2025). Dentro desta existem ainda algumas subvariantes, tais como:

- CVRP (*Capacitated VRP*): inclui veículos com capacidade limitada.
- VRPTW (*VRP with Time Windows*): com janelas temporais para atendimento.
- MDVRP (*Multi-Depot VRP*): com múltiplos depósitos.
- PCVRP (*Prize Collecting VRP*): inclui prémios para cada cliente visitado.

Neste caso, a variante escolhida foi a PCVRP. Nesta variante, não é obrigatório visitar todos os clientes, em vez disso, cada cliente tem um "prémio" ou uma recompensa associada e o objetivo é maximizar o total de prémios recolhidos, enquanto se minimiza o custo total das rotas, respeitando as capacidades dos veículos. No caso deste projeto, algumas características do PCVRP são particularmente relevantes, uma vez que nem todos os contentores precisam ser recolhidos diariamente, sendo necessário encontrar um equilíbrio entre o custo da operação e o benefício de atender determinados pontos.

1.4 Aplicação prática ao contexto do projeto

Após a apresentação das principais variantes do problema de roteamento de veículos aplicadas à recolha de resíduos, com especial destaque para a PCVRP, importa agora estabelecer a ligação entre o modelo teórico e o contexto de estudo. É necessário compreender os dados disponíveis para se definir corretamente o problema, uma vez que são estes que permitem parametrizar adequadamente o modelo de otimização. Neste sentido, torna-se necessária uma análise exploratória dos dados relativos, nomeadamente no que respeita à localização e características dos pontos de recolha, capacidades dos veículos e frequências de serviço. Esta análise constitui a base para a formulação do modelo, sendo fundamental para garantir que as soluções propostas são viáveis, eficazes e alinhadas com a realidade do sistema em estudo. As secções seguintes apresentam, portanto, a definição do problema e análise aos dados fornecidos pelo município de Cascais.

2. DEFINIÇÃO DO PROBLEMA

A recolha de resíduos urbanos no município de Cascais é um processo logístico complexo, que envolve a monitorização e o esvaziamento de contentores com diferentes tipos de resíduo (Papel, Embalagens e Vidro). No passado, alguns contentores estavam equipados com sensores o que permitia obter informação dos níveis de enchimento e planear rotas de recolha mais eficientes (Smart Waste | Cascais Data, 2025). Atualmente, os níveis de enchimento destes contentores são registados apenas quando ocorre a recolha, o que significa que entre duas observações consecutivas não existe qualquer visibilidade sobre a evolução real do enchimento.

Este cenário levanta um desafio central: o planeamento das rotas de recolha é feito com base em informação incompleta e desatualizada, o que pode levar a decisões não eficientes, como:

- Visitar contentores ainda com pouca carga.
- Chegar a contentores que já atingiram a capacidade máxima e estão a transbordar.
- Comprometer a capacidade do veículo ao recolher pontos mais cheios do que o esperado, o que pode obrigar a alterações de rota ou deixar contentores por recolher.

Na figura seguinte, apresenta-se um exemplo ilustrativo de uma situação problemática comum: um camião com capacidade de 10 unidades inicia a sua rota com base no planeamento feito a partir dos dados disponíveis, no entanto, o volume de resíduos é maior que a capacidade do camião. Como consequência, o veículo já não dispõe de capacidade suficiente para recolher todos os resíduos, comprometendo assim o cumprimento da rota planeada.

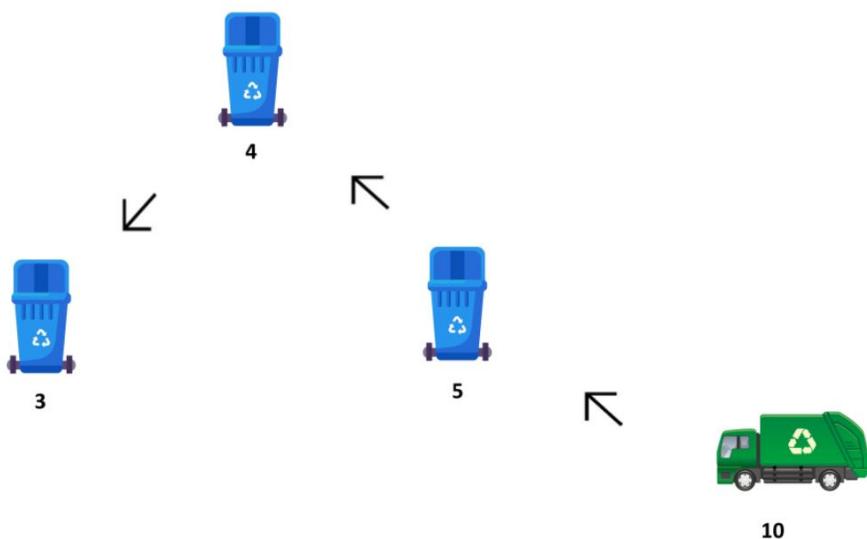


Figura 2 - Exemplificação de uma rota falhada devido a uma previsão incorreta

Fonte: Freepik | All-in-One AI Creative Suite; Elaboração própria

Para lidar com este problema, pretende-se desenvolver um modelo preditivo que faça a previsão do nível de enchimento dos contentores com a otimização das rotas de recolha. Para isso, é necessário ter a informação para cada ponto, registos das últimas recolhas com os respetivos níveis de enchimento observados e o tempo decorrido entre cada recolha.

Esta informação está disponível graças à forma como o processo é atualmente conduzido no município de Cascais: em cada operação, é selecionado um subconjunto de pontos pertencentes a um circuito e a um tipo específico de resíduo (Papel, Embalagens ou Vidro), sendo depois realizada a rota de recolha. Durante a rota, os operadores registam para cada contentor, o nível de enchimento no momento da visita, a quantidade recolhida, bem como a data e hora da recolha.

Assim, com esta informação é possível estimar a quantidade que cada ecoponto enche entre recolhas sucessivas, bem como o tempo necessário para atingir esse volume. Na figura seguinte, apresenta-se um exemplo de um ponto em que o nível foi registado no dia 2 de fevereiro às 08:00, e cujo nível de enchimento foi novamente registado no dia seguinte, à mesma hora.

Registo no dia 2 de fevereiro às 8:00



Nível - 2

Registo no dia 3 de fevereiro às 8:00



Nível 3

Figura 3 - Registo de níveis de enchimentos sucessivos

Fonte: Freepik | All-in-One AI Creative Suite; Elaboração própria

Como é possível observar na informação representada acima, o nível de enchimento do ponto aumentou um valor de um dia para outro. Assim, tendo em conta os registos dos níveis de enchimento, torna-se possível estimar a taxa de enchimento de cada ponto ao longo do tempo, o que é útil para desenvolver o modelo preditivo.

Na secção seguinte irá ser descrita a análise dos dados que foi realizada com o objetivo de extrair a informação necessária para construir o modelo e fazer a otimização das rotas.

3. ANÁLISE DE DADOS

A análise de dados foi uma etapa fundamental para compreender o comportamento dos níveis de enchimento dos contentores ao longo do tempo e para suportar as fases de previsão e otimização de rotas. Os dados utilizados neste projeto foram fornecidos pela Câmara Municipal de Cascais e incluem registo históricos de recolhas, níveis de enchimento dos ecopontos, localização destes contentores e informações associadas aos circuitos de recolha.

Inicialmente, procedeu-se à limpeza e preparação dos dados, com o objetivo de garantir a sua consistência e adequação às análises subsequentes. Entre os passos realizados destacam-se a identificação e tratamento de valores em falta, a uniformização de formatos e a segmentação dos dados por circuito, tipo de resíduo e dia.

Nos subcapítulos seguintes, são apresentados os principais indicadores extraídos dos dados, a caracterização estatística dos níveis de enchimento e a forma como esses dados foram utilizados para alimentar o sistema de previsão e o solver de otimização.

3.1 Preparação e Tratamento dos Dados

Nesta fase inicial, foi realizada uma análise exploratória dos dados disponibilizados, com o objetivo de compreender a sua estrutura, identificar possíveis inconsistências e prepará-los para posterior modelação. Foram disponibilizados dois ficheiros no formato .csv, com a seguinte composição:

- Ficheiro 1 – Localização dos pontos de recolha: contém os identificadores únicos de cada ponto de recolha e as respetivas coordenadas geográficas (latitude e longitude), que permitem a sua georreferenciação. Na figura seguinte, encontra-se a estrutura desse ficheiro.

id	time	info_adicional	tipo	nivel	local_id
5392957	2024-03-08 06:22:58+00	Círculo 11	Papel	4	3666
5392958	2024-03-08 06:22:58+00	Círculo 11	Embalager	5	3666
5392959	2024-03-08 06:22:58+00	Círculo 11	Vidro	2	3666
5392960	2024-03-08 06:23:05+00	Círculo 11	Embalager	0	3666
5392961	2024-03-08 06:24:19+00	Círculo 07	Embalager	2	6558
5392962	2024-03-08 06:24:19+00	Círculo 07	Papel	4	6558
5392963	2024-03-08 06:24:19+00	Círculo 07	Vidro	1	6558
5392964	2024-03-08 06:25:17+00	GP 2	Embalager	1	6549
5392965	2024-03-08 06:25:17+00	Círculo 03	Papel	3	6549
5392966	2024-03-08 06:25:38+00	Círculo 08	Vidro	0	3647
5392967	2024-03-08 06:25:43+00	Circ. Ilhas	Vidro	0	2924
5392968	2024-03-08 06:27:16+00	Círculo 03	Embalager	2	3927
5392969	2024-03-08 06:27:16+00	Círculo 03	Vidro	1	3927
5392970	2024-03-08 06:28:01+00	Círculo 11	Embalager	3	3712
5392971	2024-03-08 06:28:01+00	Círculo 11	Vidro	2	3712
5392972	2024-03-08 06:28:06+00	Círculo 11	Embalager	0	3712
5392973	2024-03-08 06:31:14+00	Círculo 11	Embalager	1	3719
5392974	2024-03-08 06:31:14+00	Círculo 11	Vidro	2	3719
5392975	2024-03-08 06:33:08+00	Circ. Ilhas	Vidro	0	3863
5392976	2024-03-08 06:33:35+00	Círculo 05	Embalager	1	1128
5392977	2024-03-08 06:33:35+00	Círculo 05	Vidro	2	1128
5392978	2024-03-08 06:33:50+00	Círculo 05	Embalager	1	6574
5392979	2024-03-08 06:33:50+00	Círculo 05	Vidro	2	6574
5392980	2024-03-08 06:33:56+00	Círculo 05	Embalager	1	3830
5392981	2024-03-08 06:33:56+00	Círculo 05	Vidro	2	3830
5392982	2024-03-08 06:34:10+00	Círculo 05	Embalager	1	3817
5392983	2024-03-08 06:34:10+00	Círculo 05	Vidro	2	3817

< >
Result 1
Result 2
Query
+

Figura 5 - Estrutura do ficheiro das operações de recolha

Fonte: Câmara Municipal de Cascais

Durante o processo de preparação dos dados, foram detetadas algumas inconsistências nos nomes dos circuitos, que por vezes se encontravam mal escritos ou correspondiam a circuitos apenas com um ponto associado. Estes casos foram revistos e normalizados de modo a garantir uniformidade nas análises subsequentes.

É importante referir que apenas os tipos Embalagens, Papel e Vidro foram utilizados no desenvolvimento do projeto, por isso, todas as linhas associadas ao tipo Indiferenciado foram descartadas.

Adicionalmente, os níveis de enchimento foram interpretados como uma escala ordinal, sendo possível inferir a evolução do volume de resíduos entre recolhas consecutivas. Este aspeto é particularmente relevante para a identificação de padrões de acumulação.

A etapa de preparação incluiu ainda a verificação da correspondência entre os identificadores dos pontos nos dois ficheiros. Após o tratamento, os dados encontraram-se em condições para serem utilizados nas etapas seguintes do estudo.

3.2 Cruzamento e verificação de IDs

Após a preparação inicial dos dados, procedeu-se à verificação da consistência entre os identificadores dos pontos presentes nos dois ficheiros disponibilizados. Para tal, foi realizada uma análise comparativa com o objetivo de:

- Determinar o número total de identificadores únicos existentes em cada ficheiro;
- Identificar quais os identificadores que estão presentes em ambos os conjuntos de dados;
- Verificar inconsistências, como identificadores existentes num ficheiro, mas ausentes no outro.

Depois de se ter realizado esta etapa verificou-se que no Ficheiro 1 existiam mil e quinze pontos únicos e que no Ficheiro 2 existia cinco mil setecentos e trinta. O número de pontos utilizados foram os pontos comuns aos dois ficheiros, que foram novecentos e quinze. Este cruzamento permitiu garantir que apenas os pontos com dados completos, isto é, com coordenadas e registos de operações, fossem considerados nas etapas posteriores do estudo.

3.3 Normalização dos nomes dos circuitos

Durante a exploração dos dados, foi identificada a presença de inconsistências na nomenclatura dos circuitos de recolha. Estas inconsistências incluíam erros ortográficos, variações na capitalização e divergências na formatação, como por exemplo: “Circuito 08”, “circuito 05” ou “Circuito 03 32”. Tais variações dificultavam a análise e por isso foi necessário normalizar os nomes dos circuitos. Na figura seguinte, encontra-se os nomes dos circuitos antes da normalização.

```
Total de circuitos: 22

Circ. Ilhas: 41 pontos únicos
Círculo 03: 65 pontos únicos
Círculo 03 32: 8 pontos únicos
Círculo 05: 73 pontos únicos
Círculo 06: 73 pontos únicos
Círculo 07: 67 pontos únicos
Círculo 08: 74 pontos únicos
Círculo 09: 48 pontos únicos
Círculo 10: 76 pontos únicos
Círculo 11: 68 pontos únicos
Círculo 12: 142 pontos únicos
Círculo 08: 1 pontos únicos
Circ. Ilhas 5: 1 pontos únicos
Círculo 01: 64 pontos únicos
Círculo 02: 70 pontos únicos
Círculo 04: 70 pontos únicos
círculo 05: 1 pontos únicos
via recolha: 991 pontos únicos
Circ. Ilhas 04: 1 pontos únicos
GP 91: 1 pontos únicos
Círculo 09 92: 1 pontos únicos
Círculo Ilhas: 2 pontos únicos
```

Figura 6 - Nomenclatura dos circuitos antes da normalização

Fonte: Elaboração própria

Para resolver esta situação, foi realizada uma etapa de normalização, na qual todos os nomes foram padronizados segundo uma convenção uniforme, garantindo a consistência na identificação de cada circuito. Como o “via recolha” continha apenas pontos em que o tipo de resíduo era indiferenciado, foi descartado da análise. Após este processo de limpeza e harmonização dos dados, foi possível determinar que existiam treze circuitos distintos, como se verifica na Figura 7.

Total de circuitos: 13

Circ. Ilhas: 42 pontos únicos
Círculo 03: 65 pontos únicos
Círculo 05: 73 pontos únicos
Círculo 06: 73 pontos únicos
Círculo 07: 67 pontos únicos
Círculo 08: 74 pontos únicos
Círculo 09: 48 pontos únicos
Círculo 10: 76 pontos únicos
Círculo 11: 68 pontos únicos
Círculo 12: 142 pontos únicos
Círculo 01: 64 pontos únicos
Círculo 02: 70 pontos únicos
Círculo 04: 70 pontos únicos

Figura 7 - Nomenclatura dos circuitos após a normalização

Fonte: Elaboração própria

Como é possível observar, os circuitos tem uma média de setenta pontos à exceção do Circ. Ilhas, Círculo 09 e Círculo 12.

3.4 Estruturação dos dados por Circuito de Recolha

Para facilitar a análise, foi necessário construir um ficheiro que sintetizasse a informação dispersa nos registos individuais. Este ficheiro teve como objetivo organizar os dados relativos a cada percurso de recolha, o que permitiu uma visão consolidada da atividade por circuito.

O ficheiro final incluiu os seguintes campos:

- Data do circuito – correspondente ao dia em que o circuito foi feito.
- Nome do circuito – já normalizado conforme descrito na secção anterior.
- Tipo de circuito – indica o tipo de resíduo recolhido (Embalagens, Papel ou Vidro).
- Lista de pontos associados – identificadores dos pontos de recolha servidos nesse circuito e nessa data.

Na Figura 8, apresentada a seguir, está uma parte do ficheiro que foi criado com esta informação.

3.5 Tabela de registo dos níveis por tipo e recolha

Para se aprofundar mais a análise, foi construído um ficheiro detalhado por rota, a partir da estrutura agregada previamente elaborada. Este novo ficheiro permite observar, para cada instância de execução de um circuito (definido por data, nome e tipo de resíduo), os pontos efetivamente visitados com o nível de enchimento registado por tipo, organizados por ordem crescente da hora de registo. Para cada ponto incluído na rota, foram registadas as seguintes informações:

- Data
- Hora da operação
- ID do ponto
- Nível de enchimento registado por tipo
- Indicação de recolha efetuada ou não para cada tipo

A figura seguinte, mostra uma parte da tabela criada.

A	B	C	D	E	F	G	H	I	J	K
Data do circuito	Hora	Círculo	local_id	Papel	Embalagem Vidro		Recolhido	Recolhido	Recolhido	Vidro
02/01/2024	04:20:43	Circ. Ilhas	3863	NaN	1.0	1.0	0	0	0	0
02/01/2024	04:20:49	Circ. Ilhas	3864	NaN	1.0	1.0	0	0	0	0
02/01/2024	04:32:14	Circ. Ilhas	3844	2.0	2.0	2.0	0	0	0	0
02/01/2024	04:32:30	Circ. Ilhas	3865	4.0	3.0	3.0	1	0	0	0
02/01/2024	04:53:56	Circ. Ilhas	3732	NaN	1.0	1.0	0	0	0	0
02/01/2024	04:54:02	Circ. Ilhas	3733	NaN	1.0	1.0	0	0	0	0
02/01/2024	04:54:33	Circ. Ilhas	3734	4.0	3.0	1.0	1	0	0	0
02/01/2024	06:15:55	Circ. Ilhas	3682	NaN	1.0	1.0	0	0	0	0
02/01/2024	06:16:09	Circ. Ilhas	3681	NaN	1.0	2.0	0	0	0	0
02/01/2024	06:16:18	Circ. Ilhas	3685	3.0	1.0	1.0	1	0	0	0
02/01/2024	06:19:29	Circ. Ilhas	3702	NaN	4.0	1.0	1	0	0	0
02/01/2024	07:41:09	Circ. Ilhas	3929	NaN	4.0		0	0	0	0
02/01/2024	07:42:20	Circ. Ilhas	3909	NaN	5.0		0	0	0	0
02/01/2024	08:55:44	Circ. Ilhas	6568	3.0	1.0	1.0	1	0	0	0
02/01/2024	09:06:41	Circ. Ilhas	3664	NaN	2.0		0	0	0	0
02/01/2024	09:08:01	Circ. Ilhas	6565	3.0	1.0	1.0	1	0	0	0
02/01/2024	09:08:53	Circ. Ilhas	6594	NaN	1.0	1.0	0	0	0	0
02/01/2024	06:36:34	Círculo 03	3205	2.0	5.0	2.0	0	1	0	0
02/01/2024	06:38:54	Círculo 03	3206	3.0	5.0	2.0	0	1	0	0
02/01/2024	06:41:23	Círculo 03	3207	4.0	5.0	2.0	0	1	0	0
02/01/2024	06:45:17	Círculo 03	3208	NaN	5.0	2.0	0	1	0	0
02/01/2024	06:48:27	Círculo 03	3920	5.0	5.0	4.0	0	1	0	0
02/01/2024	06:51:23	Círculo 03	3927	5.0	5.0	4.0	0	1	0	0
02/01/2024	06:55:07	Círculo 03	3921	5.0	5.0	4.0	0	1	0	0
02/01/2024	06:58:27	Círculo 03	6508	2.0	5.0	2.0	0	1	0	0

Figura 9 - Tabela de registo dos níveis por tipo e recolha

Fonte: Elaboração própria

Depois de analisar a tabela criada, tirou-se as seguintes conclusões:

- Falta de informação em determinados tipos de resíduo: Verificou-se a presença de valores NaN (*Not a Number*) nas colunas relativas aos níveis de Papel, Embalagens e Vidro, o que indica ausência de registo por parte do operador. Nestes casos, não foi possível determinar o estado de enchimento do contentor aquando da passagem.
- Indicação explícita de recolha: As colunas Recolhido Papel, Recolhido Embalagens e Recolhido Vidro permitiram identificar quando ocorreu efetivamente uma operação de recolha. Por exemplo, na quarta linha do ficheiro, é registada uma recolha de papel com um nível de enchimento de 4, realizada às 04:32:30 do dia 02-01-2024.
- Recolhas sem indicação de nível: Em alguns casos, constatou-se que uma recolha foi registada sem que estivesse disponível o nível correspondente NaN, sugerindo falhas no registo do contentor no momento da operação.
- Recolhas simultâneas de diferentes tipos: Foi possível observar que, em certos casos, foram efetuadas recolhas de Papel e Embalagens em simultâneo num mesmo ponto.
- Correspondência com os circuitos identificados anteriormente: As rotas presentes neste ficheiro coincidem com os circuitos normalizados no ficheiro agregado descrito anteriormente.

3.6 Tabela com as Observações e Recolhas

A partir da tabela construída na secção anterior, foi desenvolvida uma nova estrutura de dados com o objetivo de consolidar, para cada ponto de recolha, a informação essencial agregada por data, circuito e tipo de resíduo.

Esta tabela sintetiza o histórico de observações e operações de recolha e contém os seguintes campos-chave para cada combinação de data, ponto e tipo:

- ult_valor_obs: último nível de enchimento registado, independentemente da existência de recolha.
- ult_valor_rec: nível registado no momento da última recolha.
- ult_obs_foi_rec: booleano que indica que se na última vez que se observou o ponto também se recolheu.

- informação_diaria: nível de enchimento caso tenha sido observado/recolhido no próprio dia.
- recolhido: indica se ocorreu recolha no ponto na data indicada.
- observacao: indica se esse ponto foi observado na data em questão.
- ultima_observacao e ultima_recolha: *Timestamp* da última observação e *timestamp* da última recolha.

Na Figura 10, apresentada a seguir, está representada uma parte da tabela descrita nesta secção.

A	B	C	D	E	F	G	H	I	J	K	L	M
Data	Círculo	Tipo	local_id	ult_valor_obs	ult_valor_rec	ult_obs_foi_rec	informacao_diaria	recolhido	observacao	ultima_recolha	ultima_observacao	
02/01/2024	Circ_Ilhás	Papel	3863	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3864	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3864	NaN	NaN	0 2.0	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3865	NaN	NaN	0 4.0	1	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3732	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3733	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3734	NaN	NaN	0 4.0	1	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3682	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3681	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3685	NaN	NaN	0 3.0	1	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3702	NaN	NaN	0 NaN	1	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3929	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3909	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	6568	NaN	NaN	0 3.0	1	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3664	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	6565	NaN	NaN	0 3.0	1	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	6594	NaN	NaN	0 NaN	0	1	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3722	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3799	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3922	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3193	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3195	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3915	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	4127	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3826	NaN	NaN	0 NaN	0	0	NaN	NaN		
02/01/2024	Circ_Ilhás	Papel	3926	NaN	NaN	0 NaN	0	0	NaN	NaN		

Figura 10 - Tabela com as observações e recolhas

Fonte: Elaboração própria

Esta tabela é de grande relevância para as fases seguintes do projeto, pois a partir dela serão determinados dois parâmetros fundamentais para cada ponto e para cada dia: o valor de início (nível mais recente conhecido para o dia) e o valor de fim (nível previsto no final do dia, após acumulação e eventual recolha).

Estes dois valores constituem os principais inputs dos métodos de apoio à decisão que serão explorados nas fases seguintes, nomeadamente:

- O modelo de otimização (solver), que utiliza os níveis previstos para decidir que pontos devem ser incluídos nas rotas diárias.
- O modelo preditivo, que irá prever os níveis de enchimento, com base nos históricos de observações, padrões de acumulação e características do circuito.

Assim, esta tabela serve como uma ponte entre a análise feita aos dados e a aplicação de métodos de previsão.

3.7 Taxas de enchimento e estatísticas

Foi necessário criar estatísticas específicas por ponto com o objetivo de estimar os níveis de enchimento no início e no final do dia e para alimentar a rede neuronal e o modelo preditivo com variáveis que capturam o comportamento histórico de cada ponto.

Para cada registo foram calculadas as seguintes métricas:

- Delta de nível: diferença entre o nível de enchimento antes e depois de cada registo.
- Delta de tempo: número de horas entre dois regístros consecutivos do mesmo ponto e tipo.
- Taxa de enchimento: taxa média de enchimento por hora (delta de nível / delta de tempo).
- Nível aquando da recolha: valor registado no momento da recolha.

Para cada uma destas métricas, foram obtidas estatísticas descritivas como média, mediana, mínimo, máximo, primeiro quartil (Q1) e terceiro quartil (Q3). Esta análise foi feita em duas escalas temporais: mensal e trimestral. A taxa de enchimento vai ser especialmente importante pois vai ser utilizada para calcular os níveis de enxamentos estimados no início do dia e no fim do dia.

3.8 Cálculo dos níveis estimados no início do dia e no final do dia

Para estimar os níveis de enchimento dos pontos no início do dia, foi utilizada uma abordagem baseada numa equação linear simples, que relaciona o tempo decorrido desde a última observação do ponto com a taxa de enchimento. A fórmula foi a seguinte:

Nível estimado no início do dia

$$\begin{aligned} &= (\text{última observação em horas}) * (\text{taxa de enchimento mediana}) \\ &+ \text{último valor observado} \end{aligned}$$

Tanto o tempo da última observação, expresso em horas, quanto o último valor observado, foram extraídos da tabela apresentada no ponto 3.6.

A escolha de utilizar a taxa de enchimento mediana, em vez da média, deve-se ao facto de a mediana ser menos sensível a valores extremos, o que proporciona resultados mais robustos e com menos incoerências. Para mostrar a diferença entre a utilização da taxa de enchimento mediana e a taxa de enchimento média, apresenta-se a seguir dois gráficos com a distribuição dos níveis estimados no início do dia calculados com base na taxa de enchimento mediana e na taxa de enchimento média, respetivamente.

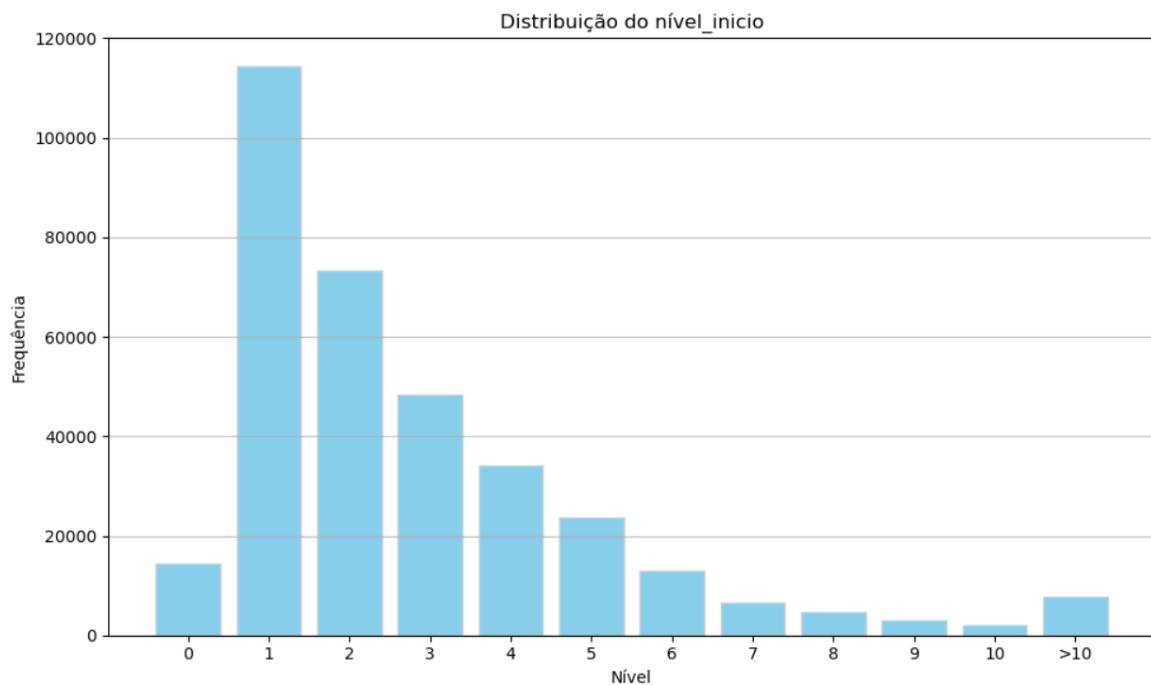


Figura 12 - Distribuição dos níveis de enchimento usando a taxa de enchimento mediana

Fonte: Elaboração própria

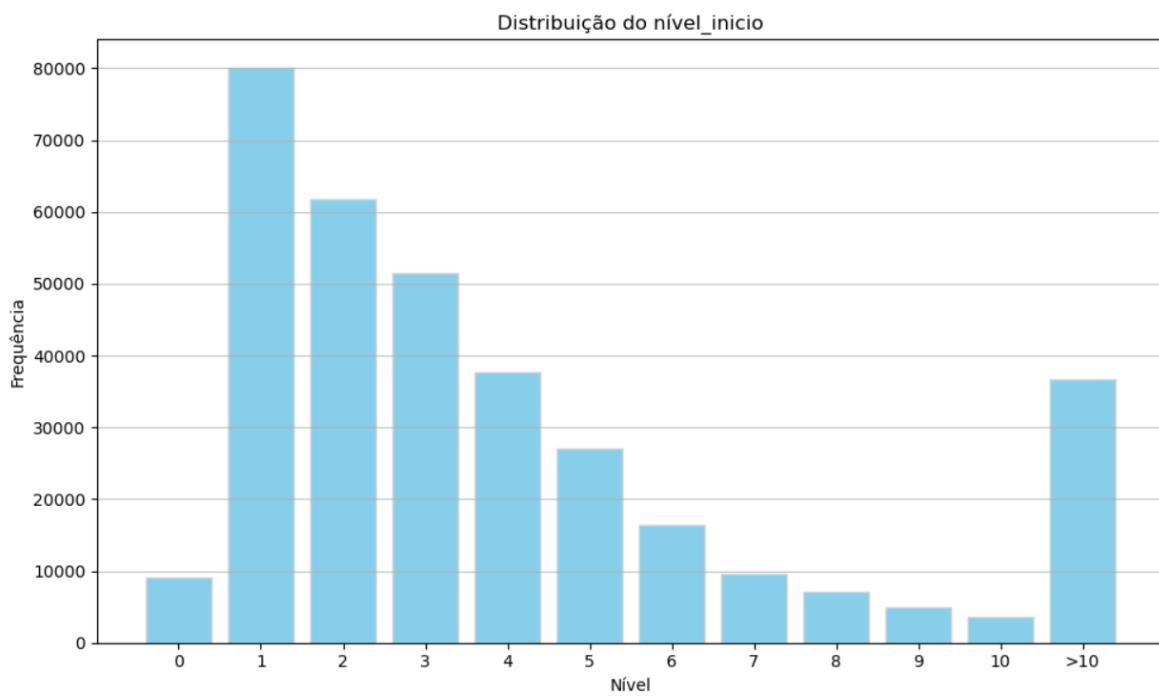


Figura 11 - Distribuição dos níveis de enchimento usando a taxa de enchimento média

Fonte: Elaboração própria

Os valores acima do nível 5 representam inconsistências e como é possível observar, ao usar a taxa de enchimento mediana, essas inconsistências são significativamente menores. Idealmente, os níveis de enchimento deveriam situar-se entre 0 e 5. No entanto, ao utilizar a taxa média, observou-se uma maior dispersão dos valores, incluindo níveis fora desse intervalo, logo, utilizou-se a taxa mediana pois reduziu significativamente essa variabilidade, produzindo estimativas mais realistas e coerentes com os limites operacionais dos contentores. Ainda assim, para corrigir eventuais inconsistências, como valores NaN, decidiu-se limitar os níveis de enchimento a um máximo de 5 e a um mínimo de 1.

Os valores do nível de enchimento de fim de dia foram obtidos através da coluna informação_diaria, presente no ficheiro demonstrado no ponto 4.6. Quando o valor era NaN, decidiu-se igualar ao valor estimado do nível de enchimento de início do dia.

3.9 Ficheiro com os circuitos realizados em Cascais

Numa fase posterior da análise, solicitámos à Câmara Municipal de Cascais um apoio adicional para melhorar a identificação das rotas realizadas por cada camião. Em resposta, foi-nos fornecido um ficheiro que continha informações detalhadas sobre as operações de recolha, incluindo a data de início e de fim, a distância percorrida, os locais por onde passou o camião, o ID da viatura, a identificação do circuito e o tipo de resíduo recolhido.

Contudo, após uma análise cuidadosa deste ficheiro, verificámos inconsistências significativas. Os pontos indicados como pertencentes a determinados circuitos não correspondiam à realidade, segundo a informação que nos foi fornecida. Observou-se que os circuitos tinham muitos mais pontos do que aqueles registados no ficheiro, e vários pontos que supostamente pertenciam a um circuito estavam, na verdade, associados a outro.

Deste modo, o ficheiro não contribuiu para uma melhoria significativa na análise das rotas, não sendo possível utilizá-lo para clarificar ou validar a associação dos pontos aos circuitos de forma precisa. Na Figura 13, encontra-se uma amostra do ficheiro fornecido.

inicio	fim	distancia_percorrida	locais	viatura_id	slug	slug
2024-09-04 05:06:20+00	2024-09-04 10:27:05+00	43643	{3795,2581,3840,	95	circuito-vidro3-sel1	vid
2024-12-13 07:55:56+00	2024-12-13 11:25:03+00	35497	{3861,3886,3178,	95	circuito-vidro1-sel1	vid
2024-12-17 07:50:13+00	2024-12-17 11:36:01+00	37188	{3802,3871,3694,	95	circuito-vidro1-sel1	vid
2024-12-17 13:18:21+00	2024-12-17 18:41:04+00	51767	{2971,2975,2977,	95	circuito-vidro1-sel2	vid
2024-12-18 06:14:00+00	2024-12-18 11:15:05+00	41329	{3863,3864,3865,	95	circuito-vidro1-sel1	vid
2024-12-18 13:25:26+00	2024-12-18 18:39:17+00	63860	{2967,2968,4329,	95	circuito-vidro1-sel2	vid
2024-12-19 06:40:43+00	2024-12-19 10:21:50+00	31512	{3822,2574,3823,	95	circuito-vidro1-sel1	vid
2024-12-19 13:23:52+00	2024-12-19 18:11:53+00	67219	{4398,3084,3059,	95	circuito-vidro1-sel2	vid
2024-12-20 06:28:26+00	2024-12-20 10:40:30+00	60915	{2606,2924,2608,	95	circuito-vidro1-sel1	vid
2024-12-21 06:41:15+00	2024-12-21 10:53:44+00	57504	{2571,2836,3571,	95	circuito-vidro1-sel1	vid
2024-12-21 13:24:07+00	2024-12-21 18:37:38+00	56449	{3110,4290,4292,	95	circuito-vidro1-sel2	vid
2024-12-23 06:25:54+00	2024-12-23 11:15:49+00	53776	{3577,2604,3696,	95	circuito-vidro1-sel1	vid
2024-12-23 13:17:56+00	2024-12-23 18:46:24+00	52720	{3109,2995,3057,	95	circuito-vidro1-sel2	vid
2024-12-24 05:11:49+00	2024-12-24 09:57:19+00	51621	{3571,3669,3870,	95	circuito-vidro1-sel1	vid
2024-12-26 06:47:18+00	2024-12-26 11:12:41+00	41068	{3692,3569,3679,	95	circuito-vidro1-sel1	vid
2024-12-27 06:53:26+00	2024-12-27 11:11:16+00	29510	{2572,3865,3844,	95	circuito-vidro1-sel1	vid
2024-12-27 13:04:50+00	2024-12-27 18:41:02+00	49372	{3000,3020,4307,	95	circuito-vidro1-sel2	vid
2024-12-28 06:19:39+00	2024-12-28 11:09:53+00	67036	{2782,3154,3611,	95	circuito-vidro1-sel1	vid
2024-01-02 06:07:34+00	2024-01-02 11:33:52+00	55033	{2747,3716,3711,	111	circuito-09-plastico-sel1	pla
2024-01-03 06:05:48+00	2024-01-03 11:04:32+00	86175	{344,2631,2633,2	111	gpm-plastico-sel1	pla
2024-01-03 13:16:08+00	2024-01-03 17:48:40+00	49694	{4403,4324,4242,	111	circuito-sabado-4f-papel	pap
2024-01-04 06:05:00+00	2024-01-04 11:01:21+00	38407	{6549,3927,4181,	111	circuito-03-papel-sel1	pap
2024-01-04 13:00:40+00	2024-01-04 19:06:33+00	76253	{3058,3057,38,36	111	circuito-08-plastico-sel2	pla
2024-01-05 05:41:37+00	2024-01-05 11:01:50+00	44581	{3927,4181,3205,	111	circuito-03-plastico-sel1	pla
2024-01-05 13:01:59+00	2024-01-05 18:58:34+00	69244	{3058,6419,3057,	111	circuito-08-papel-sel2	pap
2024-01-06 13:11:58+00	2024-01-06 18:20:36+00	54345	{4323,4300,2899,	111	circuito-sabado-4f-papel	pap
2024-01-08 00:05:24+00	2024-01-08 12:21:00+00	60855	{2820,2807,2806,	111	-----	---

Figura 13 - Ficheiro fornecido para ajudar a identificar as rotas

Fonte: Câmara Municipal de Cascais

4. OTIMIZAÇÃO DAS ROTAS

Com base nos níveis obtidos no ponto 3.8, foi criada um ficheiro com as soluções obtidas pelo solver. Esta tem para cada combinação de dia, circuito e tipo de resíduo, as diferentes rotas obtidas. Foram incluídas três versões distintas de rota:

- Rota de início: criada com base nos níveis de enchimento estimados no início do dia, ou seja, antes de qualquer operação de recolha. Esta rota representa o planeamento feito com dados disponíveis à partida e reflete um cenário comum em ambientes reais, onde não existe conhecimento completo do estado dos ecopontos ao longo do dia.
- Sub-rota: esta rota não é calculada pelo solver. É derivada posteriormente, utilizando os mesmos pontos da rota de início, mas considerando os níveis de enchimento estimados para o final do dia. O objetivo é simular o que teria acontecido caso a rota de início tivesse sido executada, mas com os valores reais (ou previstos com mais precisão) no final do dia. Assim, é possível perceber que, mesmo com um planeamento prévio, podem ocorrer sobrecargas ou falhas na recolha, caso os dados não reflitam corretamente a realidade.
- Rota de fim: construída com base nos níveis de enchimento estimados para o final do dia, ou seja, assume que se dispõe de informação completa e precisa. Representa o cenário ideal em que todas as decisões são tomadas com base em dados mais atualizados permitindo obter a solução teoricamente mais eficiente.

O solver utilizado neste projeto é o PyVRP, uma biblioteca open-source desenvolvida especificamente para resolver variantes do problema de roteamento de veículos. Este solver recebe como entrada os pontos de recolha e os armazéns com as respetivas coordenadas, a quantidade de resíduos a recolher em cada ponto, a capacidade do veículo, o tempo de serviço para cada ponto que neste caso foi de 5 minutos e matrizes de tempo e distância.

A função objetivo do *solver* foi definida com o propósito de minimizar dois aspectos fundamentais:

- A distância total percorrida pelas viaturas.
- A quantidade de resíduos não recolhidos, considerando os limites de capacidade dos veículos e os níveis dos contentores.

Para cada rota foram ainda registados os seguintes dados:

- Distância total percorrida.
- Duração estimada da rota.
- Lista dos pontos visitados.
- Quantidade de resíduos não recolhidos.
- Quantidade total de resíduos recolhidos.
- Ocupação do camião.

Na figura seguinte, encontra-se uma amostra da tabela obtida com as soluções do solver para cada rota.

A	B	C	D	E	F	G	H	I	J
data	circuito	tipo	rota1	distancia_rota1	duracao_rota1	uncollected_prizes_rota1	prizes_rota1	num_clientes_rota1	ocupacao_camião_rota1
02/01/2024	Circ_Ilhas	Papel	[0, 27, 34525.32	219.469823		48	120	30	120
02/01/2024	Circuito_03	Embalagens	[0, 40, 32973.6	381.136311		10	120	60	120
02/01/2024	Circuito_05	Embalagens	[0, 2, 1119918.48	248.599815		99	120	40	120
02/01/2024	Circuito_06	Papel	[0, 28, 20349.620000000003	251.627152		99	120	40	120
02/01/2024	Circuito_07	Papel	[0, 3, 4, 24161.46	257.565986		81	120	40	120
02/01/2024	Circuito_08	Papel	[0, 3, 6, 31189.94	212.293821		172	120	30	120
02/01/2024	Circuito_09	Papel	[0, 38, 25974.98	207.900653		69	120	30	120
02/01/2024	Circuito_09	Embalagens	[0, 23, 35829.38	323.646646		0	99	48	99
02/01/2024	Circuito_10	Papel	[0, 12, 20945.239999999998	200.827489		184	120	30	120
02/01/2024	Circuito_11	Papel	[0, 42, 26997.78	203.294156		152	120	30	120
02/01/2024	Circuito_11	Embalagens	[0, 42, 29902.280000000002	264.409653		88	120	40	120
02/01/2024	Circuito_12	Papel	[0, 15, 123464.66	260.395821		306	120	40	120
03/01/2024	Circ_Ilhas	Papel	[0, 28, 36195.4	240.382486		30	120	33	120
03/01/2024	Circ_Ilhas	Vidro	[0, 28, 51857.78	321.563982		0	56	42	56
03/01/2024	Circuito_01	Embalagens	[0, 42, 23284.120000000003	255.927484		72	120	40	120
03/01/2024	Circuito_02	Embalagens	[0, 2, 4128548.52	378.227974		20	120	60	120
03/01/2024	Circuito_03	Vidro	[0, 39, 30243.3	332.85998		28	120	52	120
03/01/2024	Circuito_04	Papel	[0, 38, 24249.239999999998	264.232152		90	120	40	120
03/01/2024	Circuito_05	Embalagens	[0, 42, 29329.58	313.532144		38	120	48	120
03/01/2024	Circuito_07	Papel	[0, 3, 4, 24498.02	258.716819		79	120	40	120
03/01/2024	Circuito_07	Embalagens	[0, 3, 4, 25221.44	259.732323		87	120	40	120
03/01/2024	Circuito_09	Papel	[0, 38, 27706.12	232.062318		45	120	34	120
03/01/2024	Circuito_09	Vidro	[0, 23, 35829.38	323.646646		0	83	48	83
03/01/2024	Circuito_11	Vidro	[0, 44, 35077.42	355.395815		29	120	54	120
03/01/2024	Circuito_12	Papel	[0, 8, 7, 21625.5	241.473987		331	119	37	119
03/01/2024	Circuito_12	Embalagens	[0, 8, 1129681.8	351.548815		180	119	54	119
04/01/2024	Circuito_11	Vidro	[0, 28, 51857.78	321.563982		0	100	40	100

Figura 14 - Tabela com as soluções do solver para o ano todo

Fonte: Elaboração própria

4.1 Previsão de níveis de enchimento usando *Machine Learning*

Após ter sido calculada a estimativa para o nível de enchimento no início do dia e no final, decidiu-se usar um modelo preditivo para prever os níveis de enchimento. Para isso usou-se a scikit-learn, uma das bibliotecas em Python mais usadas para *machine learning*. Esta biblioteca disponibiliza uma ampla variedade de algoritmos, entre os quais foram testados o *LinearRegression*, o *RandomForestRegressor* e o *HistGradientBoostingRegressor*.

Para verificar se o modelo iria apresentar resultados melhores aos previamente calculados, foi necessário calcular o erro absoluto médio para dezembro para os níveis obtidos usando a taxa de enchimento mediana e para os níveis obtidos usando os modelos preditivos. Verificou-se que enquanto o erro absoluto médio para os níveis usando a taxa de enchimento mediana era de 1 unidade, para os níveis obtidos usando o modelo *HistGradientBoostingRegressor* era de 0.6 unidades, sendo assim o melhor dos modelos testados.

Este modelo constrói várias árvores de decisão de forma interativa e depois ajusta-se progressivamente aos erros do modelo anterior, o que o torna eficaz na captura de padrões.

Para treinar este modelo é necessário definir um conjunto de treino e um conjunto de validação. O conjunto de treino foi constituído pelos dados compreendidos entre janeiro e novembro e o conjunto de validação de desempenho do modelo, foi com os dados do mês de dezembro.

Em seguida, usando o modelo, calculou-se os níveis de enchimento estimados para o início do dia e para o final do dia no mês de dezembro. Posteriormente, o solver foi novamente utilizado, desta vez com os dados previstos pelo modelo, com o objetivo de comparar as rotas criadas com aquelas obtidas utilizando os níveis estimados a partir da taxa de enchimento mediana.

4.2 Resultados Obtidos Do Modelo Preditivo

Para verificar o desempenho do modelo preditivo, decidiu-se calcular o lixo não recolhido e distância, pois são os parâmetros otimizados pelo solver, das seguintes rotas:

- Sub-Rota Base: Rota que passa pelos mesmos pontos da rota criada usando os níveis de início calculados com a taxa mediana, mas usa os níveis de enchimento para o final do dia calculados com a taxa mediana.
- Sub-Rota HGB: Rota que passa pelos mesmos pontos da rota criada usando os níveis de início previstos pelo modelo preditivo, mas usa os níveis de enchimento para o final do dia previstos do modelo preditivo.
- Rota Fim: Rota com os níveis de enchimento para o final do dia previstos pelo modelo preditivo.

4.2.1 Quantidade de lixo não recolhido

Na tabela seguinte é possível observar a média geral da percentagem de lixo não recolhido.

Tabela 1 - Média geral da percentagem da quantidade de lixo não recolhido

Tipo de Rota	Média
Sub-Rota Base	37,91
Sub-Rota HGB	36,70
Rota Fim	35,60

Na tabela seguinte encontra-se a média da percentagem de lixo não recolhido por tipo (Embalagens, Vidro e Papel) e rota acima mencionada.

Tabela 2 – Média da percentagem da quantidade de lixo não recolhido por tipo e rota

Tipo de Rota	Embalagens (%)	Papel (%)	Vidro (%)
Sub-Rota Base	46,37	57,73	10,79
Sub-Rota HGB	45,55	56,97	8,77
Rota Fim	44,09	55,61	8,28

Nas figuras seguintes, encontram-se gráficos com a distribuição da percentagem do lixo não recolhido para cada tipo de lixo e para a Sub-Rota Base, Sub-Rota HGB e Rota Fim.

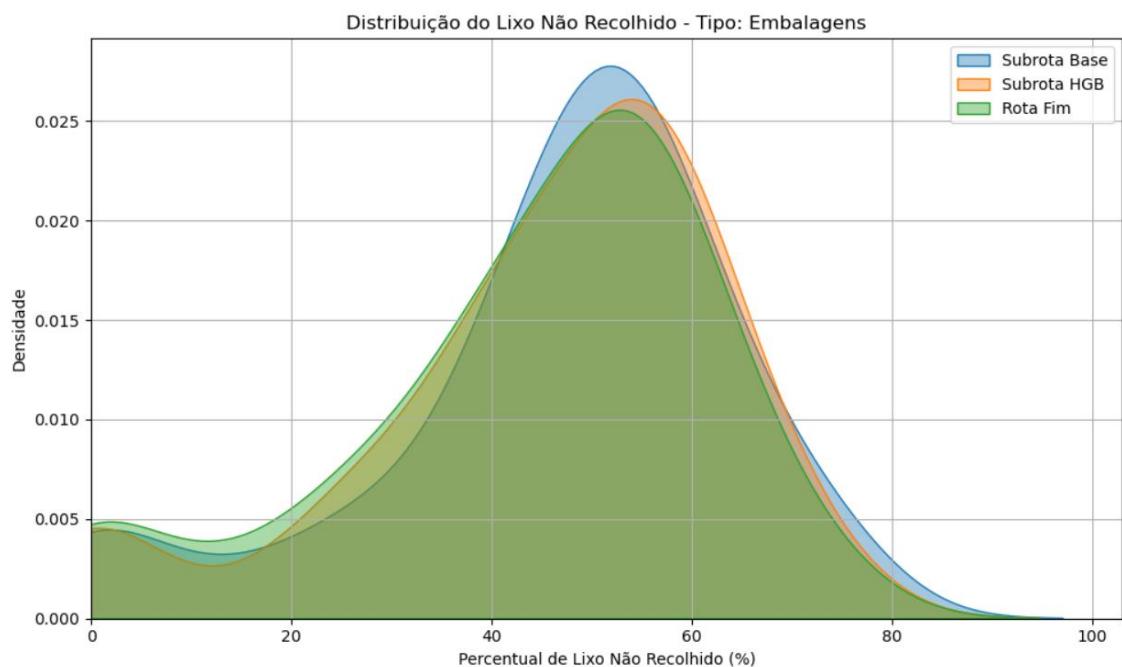


Figura 15 - Distribuição do lixo não recolhido para o tipo Embalagens

Fonte: Elaboração própria

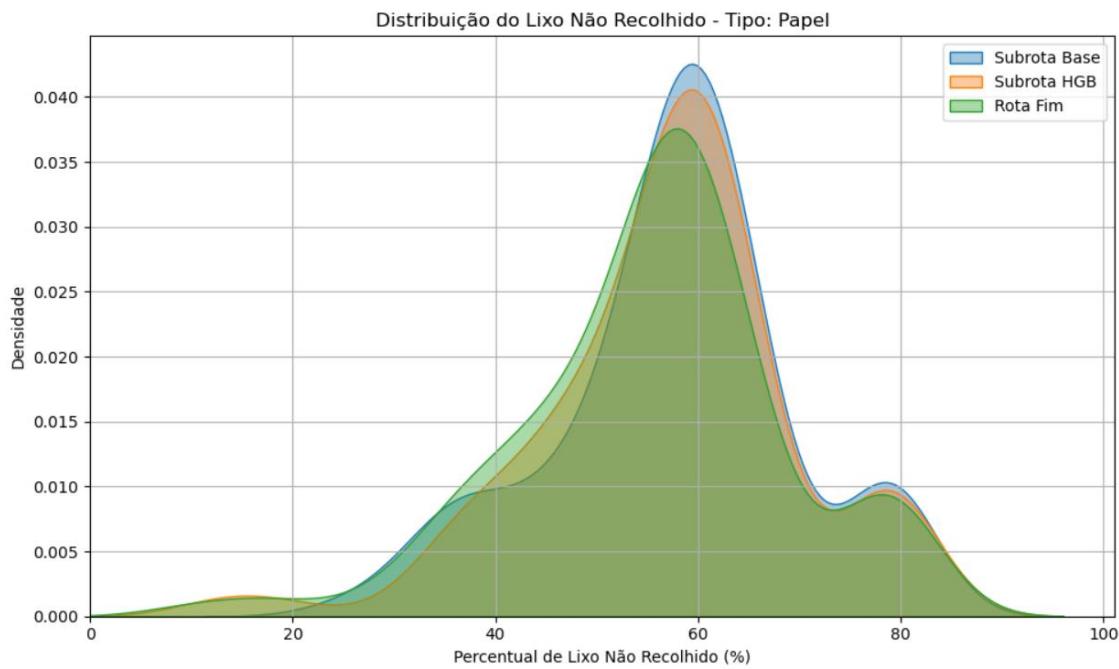


Figura 16 - Distribuição da percentagem do lixo não recolhido para o tipo Papel

Fonte: Elaboração própria

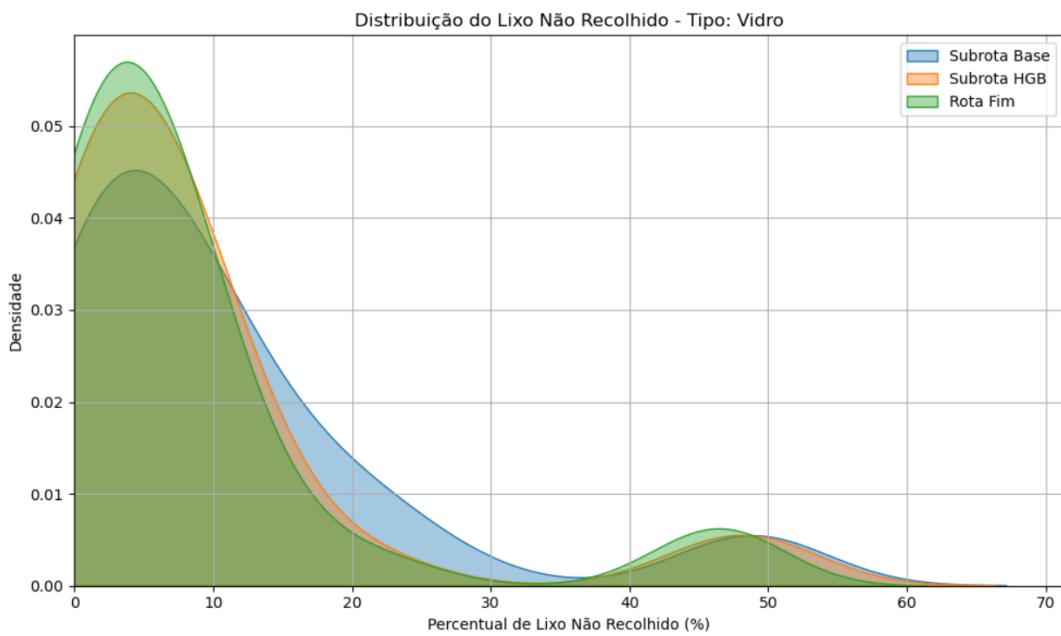


Figura 17 - Distribuição da percentagem do lixo não recolhido para o tipo Vidro

Fonte: Elaboração própria

Depois de obtidos os resultados, é importante destacar que a Rota Fim representa a rota criada com os dados mais atualizados, correspondendo assim ao cenário ideal e, naturalmente, ao melhor desempenho possível. Por essa razão, a comparação mais relevante para avaliar a qualidade do modelo preditivo é entre a Sub-Rota Base e a Sub-Rota HGB. A melhoria da Sub-Rota HGB face à Sub-Rota Base nesta métrica é a seguinte:

- Embalagens – 1,77%
- Papel – 1,32%
- Vidro – 18,72%
- Geral – 3,19%

Logo, Sub-Rota HGB apresentou um desempenho superior nesta métrica, o que indica que o modelo preditivo conseguiu fornecer estimativas mais eficazes, o que resultou em rotas mais eficientes.

4.2.2 Distância

Na tabela apresentada a seguir, é possível observar a média geral em metros para a distância das rotas.

Tabela 3 - Média geral da distância

Tipo de Rota	Média
Sub-Rota Base	28962,37
Sub-Rota HGB	28577,48
Rota Fim	28283,19

Na tabela a seguir, é possível observar a média da distância em metros para cada tipo de lixo e rota.

Tabela 4 – Média da distância por tipo e rota

Tipo de Rota	Embalagens	Papel	Vidro
Sub-Rota Base	27028,11	23694,69	35870,70
Sub-Rota HGB	26037,62	22616,34	36735,76
Rota Fim	25672,69	22501,98	36338,12

Nas figuras seguintes encontram-se gráficos com a distribuição distância para cada tipo de lixo e para a Sub-Rota Base, Sub-Rota HGB e Rota Fim.

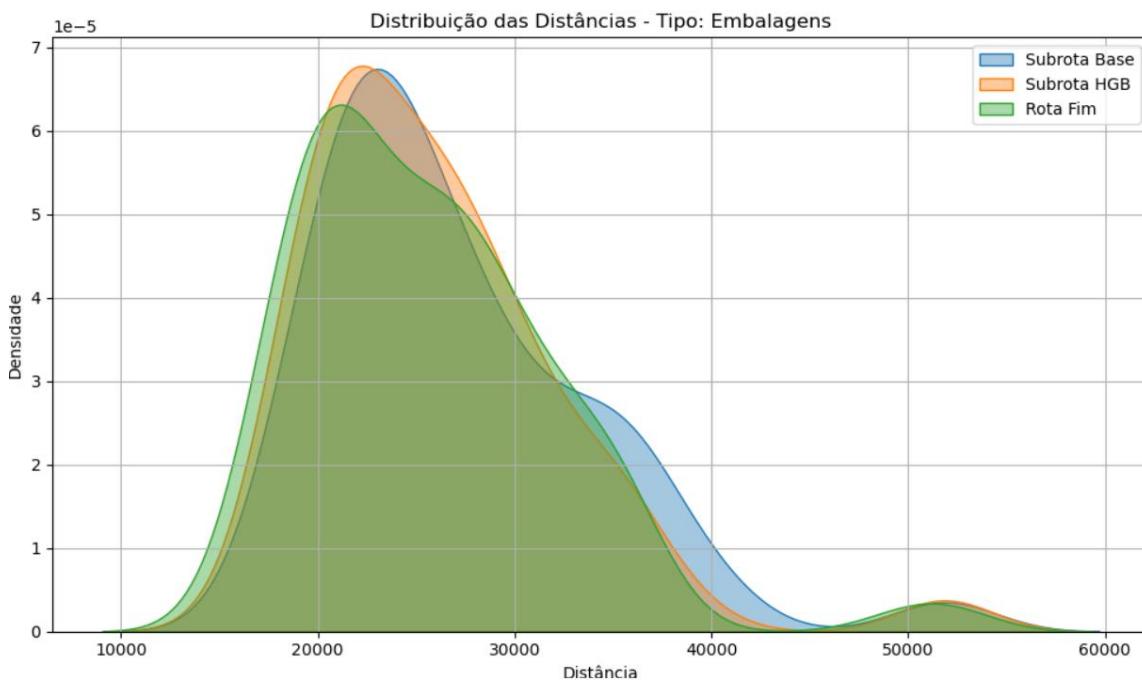


Figura 18 - Distribuição da distância para o tipo Embalagens

Fonte: Elaboração própria

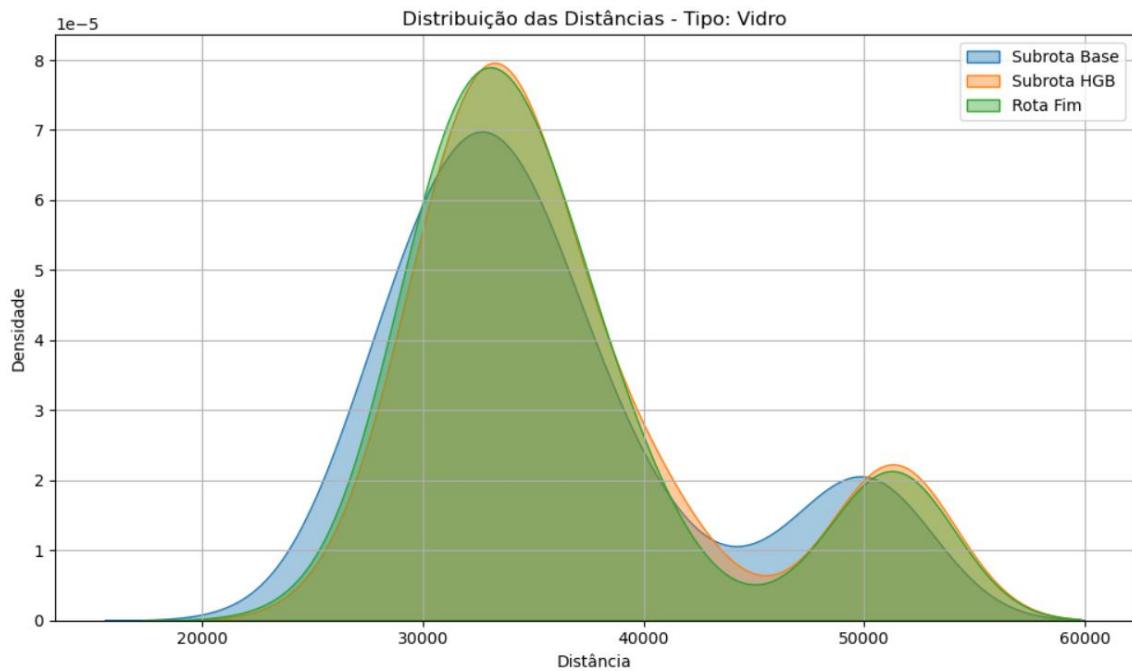


Figura 19 - Distribuição da distância para o tipo Vidro

Fonte: Elaboração própria

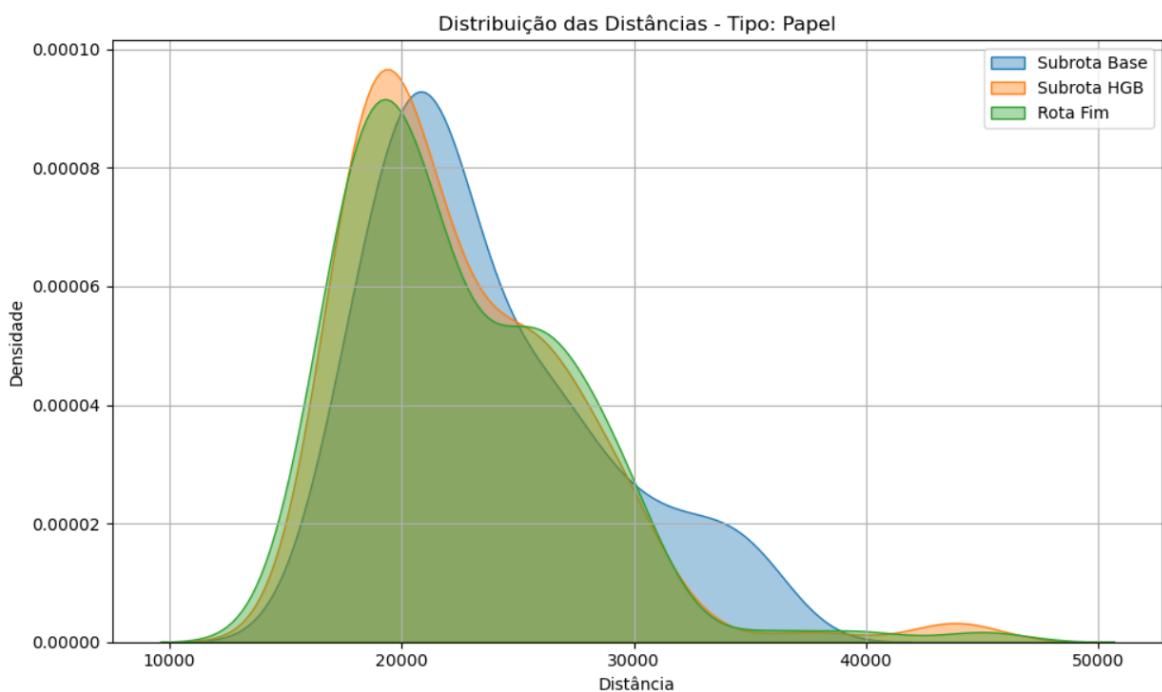


Figura 20 - Distribuição da distância para o tipo Papel

Fonte: Elaboração própria

Depois de obtidos os resultados, fez-se a mesma análise da quantidade de lixo não recolhido para a distância e verificou-se os seguintes valores da Sub-Rota HGB em relação à Sub-Rota Base:

- Embalagens: 3,66%
- Papel: 4,55%
- Vidro: + 2,41%
- Geral: 1,33%

Apesar da Sub-Rota HGB ter registado, em média, um aumento de 2,41% na distância percorrida nas rotas de Vidro, os resultados obtidos para os restantes tipos de resíduos foram positivos. No geral, isto evidencia a eficácia da abordagem preditiva utilizada na criação da Sub-Rota HGB.

4.3 Conclusão do resultado obtido

De forma geral, o modelo preditivo demonstrou-se eficaz na criação de rotas otimizadas, quando comparado com a abordagem baseada na taxa de enchimento mediana para a estimativa dos níveis de enchimento. Nas métricas analisadas, nomeadamente a distância percorrida e a quantidade de lixo não recolhido, observou-se uma melhoria consistente nos diferentes tipos de resíduo. Estes resultados evidenciam o potencial da abordagem preditiva como uma ferramenta de apoio ao planeamento logístico da recolha de resíduos urbanos, o que pode contribuir para operações mais eficientes e sustentáveis.

5. DESENVOLVIMENTO DE UM DASHBOARD

No âmbito deste projeto, foi desenvolvido um *dashboard* interativo com recurso à biblioteca Streamlit, com o objetivo de integrar, visualizar e comparar diferentes abordagens de otimização de rotas de recolha de resíduos. A estrutura da aplicação permitir uma fácil navegação entre os módulos, desde a exploração dos dados até à visualização das rotas otimizadas de forma clara e funcional.

5.1 Estrutura do Dashboard

A aplicação foi estruturada em vários módulos:

- Página Inicial: Apresenta de forma breve o conteúdo da aplicação.
- Análise de Dados: Este módulo permite uma exploração visual e interativa dos dados. É possível analisar estatísticas, padrões de enchimento dos contentores ao longo do tempo, distribuições por circuito, tipo de contentor, quantidade de pontos por circuito e frequência de recolha. Para isso, foram utilizados gráficos para uma melhor compreensão.
- Rotas Reais: Esta secção permite visualizar as rotas de recolha efetivamente realizadas no concelho de Cascais e as informações obtidas de cada rota. É possível selecionar o dia, circuito e tipo de resíduo.
- Otimização de Rotas (Solver): Nesta parte da aplicação são apresentadas as rotas criadas pelo solver, bem como a suas características, com base nos dados de níveis de enchimento estimados no início e final do dia. Estas rotas foram construídas de modo a minimizar custos como a distância percorrida e a quantidade de lixo não recolhido.

A aplicação tem como objetivo não só integrar as diferentes componentes do projeto, mas também permitir a comparação prática entre as rotas reais e as soluções geradas computacionalmente. Esta ferramenta pode, assim, servir de base para apoiar decisões estratégicas no planeamento de operações urbanas de recolha de resíduos.

5.2 Módulo Análise de Dados

Uma das componentes fundamentais da aplicação é a funcionalidade dedicada à análise exploratória de dados. Esta secção permite ao utilizador compreender o comportamento e os padrões dos dados que foram usados tanto para fundamentar o algoritmo de otimização, como o modelo preditivo.

Aqui, pode-se explorar os seguintes elementos:

- Frequência semanal e médias de enchimento.
- Distribuição de pontos por número de circuito.
- Quantidade de pontos por circuito.
- Visualização geográfica dos pontos por circuito.
- Gráficos diversos de um ponto aleatório por circuito.
- Evolução temporal dos níveis de enchimento por circuito.
- Quantidade de lixo recolhido por circuito e tipo vs. lixo total.

Os gráficos foram construídos com bibliotecas como Plotly e Matplotlib que permite uma visualização mais apelativa. Além disso, em algumas secções, é possível selecionar intervalos de datas, tipos de contentores e circuitos específicos, que torna possível fazer análises mais direcionadas. Nos subpontos abaixo irem explicar melhor cada subsecção do módulo.

5.2.1 Frequência semanal e médias por enchimento

Nesta subsecção é possível ver dois gráficos para todos os circuitos: um com a frequência semanal, onde mostra o dia da semana e a quantidade de vezes que nesse dia foi realizado o circuito e no outro é possível ver para cada tipo, uma média dos níveis de enchimento. A Figura 21 mostra um exemplo dos gráficos para o Circuito 02.

Círcuito 02

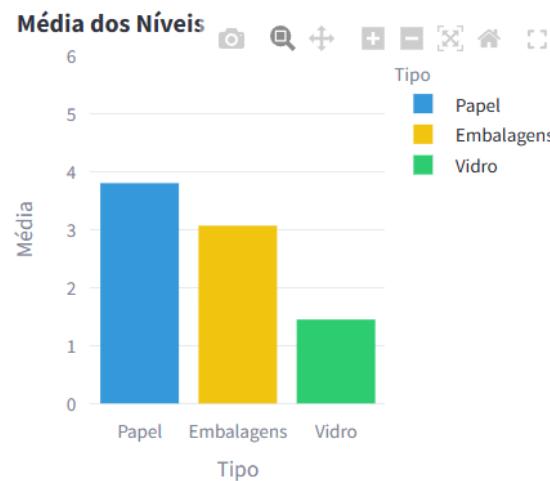
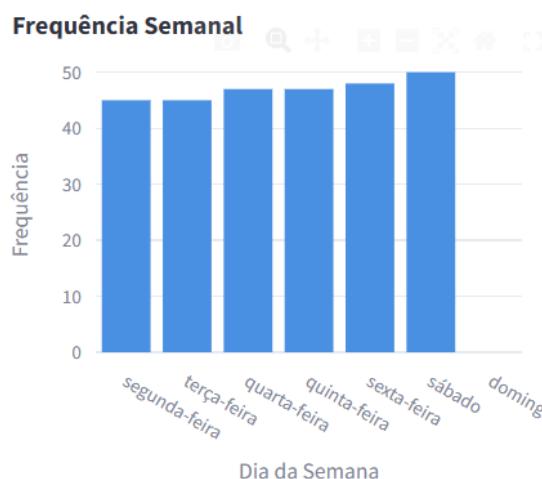


Figura 21 - Frequência semanal e média dos enchimentos para o Círculo 02

Fonte: Elaboração própria

A partir dos gráficos da frequência semanal para todos os circuitos, foi possível analisar o seguinte:

- Segunda-feira até sábado apresentam frequências altas e estáveis para a todos os circuitos, que indica que os circuitos ocorrem regularmente nesses dias.
- Em todos os circuitos, domingo é o dia com menor frequência sendo nula em alguns casos.

Relativamente aos gráficos da média de enchimento, depois de serem analisados, concluiu-se o seguinte:

- O nível de papel é o que apresenta as médias mais altas em todos os circuitos, em seguida as embalagens e por último, o vidro.
- Todos os circuitos seguem este padrão.

5.2.2 Distribuição de pontos por número de circuito

Nesta subsecção, é apresentada um gráfico que mostra os ecopontos e o número de circuitos a que pertencem. O objetivo desta visualização é perceber se os pontos estão distribuídos exclusivamente por circuitos distintos ou se existem sobreposições, o que pode ter implicações tanto na operação como na otimização das rotas. É possível observar este gráfico na Figura 22.

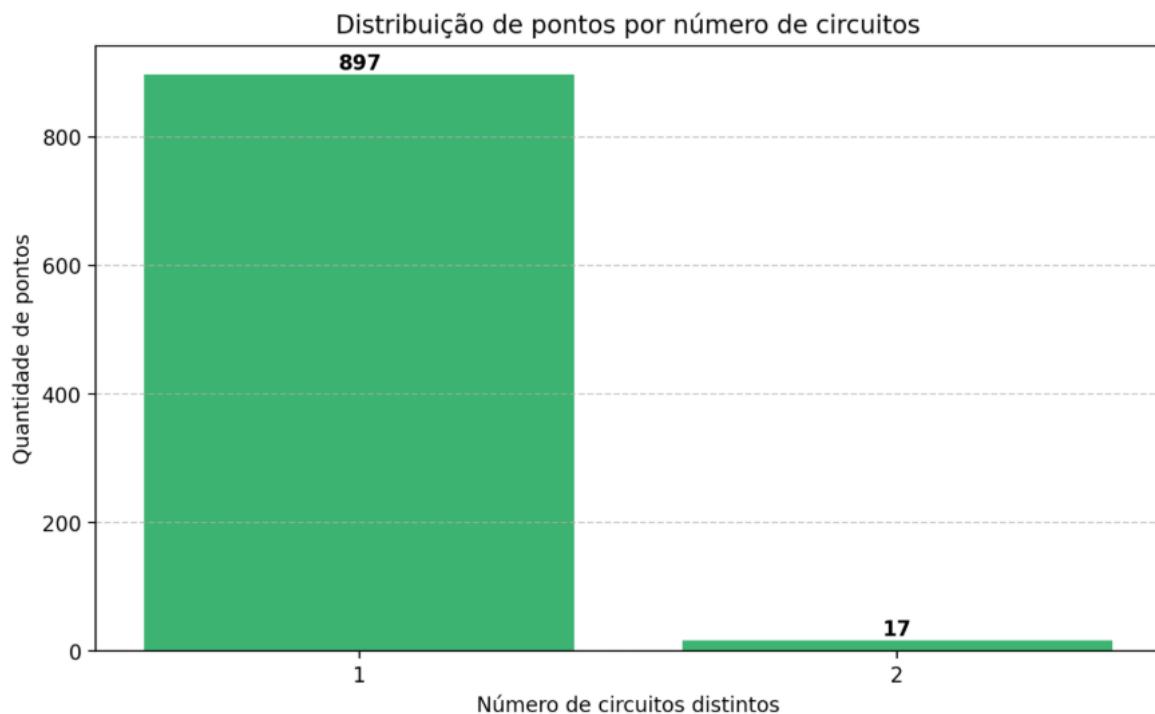


Figura 22 - Distribuição de pontos por número de circuitos

Fonte: Elaboração própria

Como é possível observar no gráfico anterior, existem um total de novecentos e quinze pontos sendo oitocentos e noventa e sete pertencem a exclusivamente a um circuito e dezassete deles estão associados a dois circuitos. Esta informação é importante pois permite perceber que a grande maioria dos ecopontos é gerida de forma única e exclusiva por um circuito, o que simplifica o planeamento e execução das rotas.

5.2.3 Quantidade de pontos por circuito

O gráfico construído nesta subsecção apresenta a distribuição do número de pontos por circuito. Na Figura 23 é possível visualizar claramente a quantidade de pontos atribuída a cada circuito.

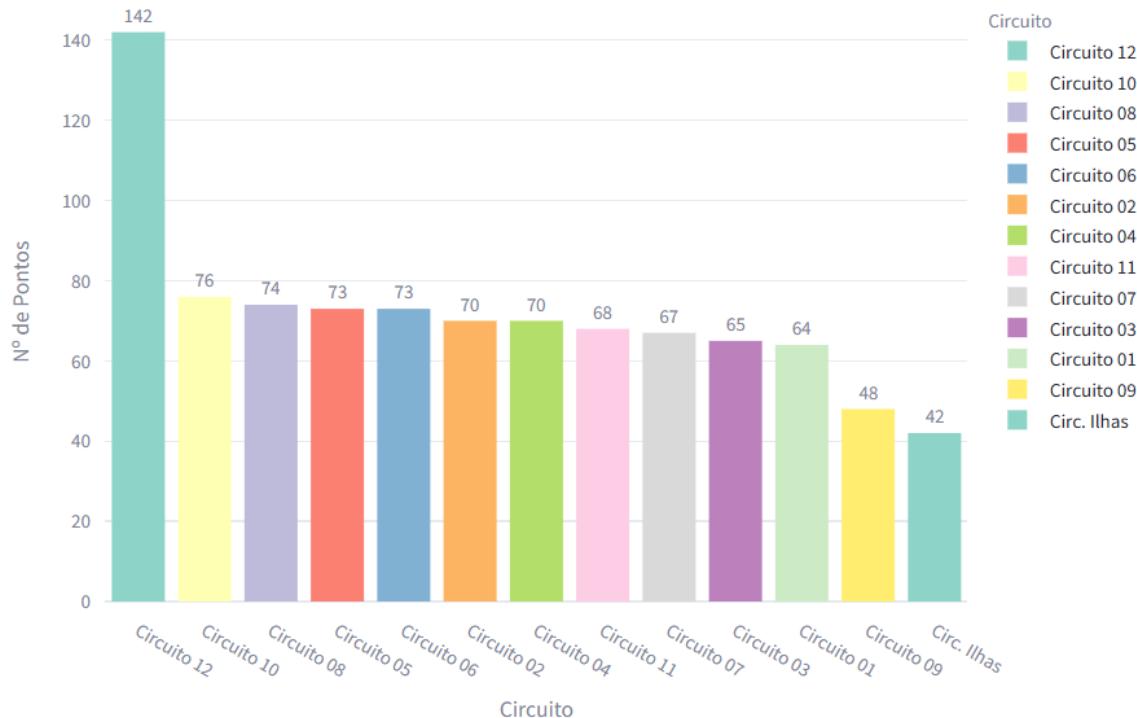


Figura 23 - Quantidade de pontos por circuito

Fonte: Elaboração própria

Com base nos dados analisados, observou-se que:

- O número médio de pontos por circuito é de aproximadamente setenta.
- No entanto, existem exceções relevantes, nomeadamente:
 - O Circuito 12, com cento e quarenta e dois ecopontos que apresenta uma quantidade do dobro de pontos médio.
 - O Circuito 09, com quarenta e oito ecopontos, e o Circuito Ilhas, com quarenta e dois ecopontos, apresentam valores menores de pontos.

Estas discrepâncias podem estar relacionadas com características geográficas, acessibilidade ou frequências de recolha que são fatores que influenciam diretamente o planeamento das rotas.

5.2.4 Visualização geográfica dos pontos por circuito

Para complementar a análise descritiva dos dados, foi desenvolvida uma visualização interativa em mapa, que visualizar geograficamente os ecopontos atribuídos a cada circuito. Esta abordagem facilita a compreensão espacial da operação e fornece uma perspetiva mais intuitiva da distribuição territorial dos pontos.

A funcionalidade implementada permite ao utilizador:

- Selecionar um ou mais circuitos através de uma caixa de seleção.
- Visualizar no mapa os ecopontos correspondentes aos circuitos escolhidos, com cores diferentes para cada circuito.
- Observar concentrações geográficas e dispersão dos ecopontos, o que pode ajudar a identificar zonas de maior densidade ou áreas isoladas.

Abaixo, na Figura 24, encontra-se uma representação da caixa de seleção e na Figura 25, um exemplo do mapa com os pontos do Circuito 10 representados a vermelho.

Mapa interativo com pontos dos circuitos

Selecione os circuitos a visualizar no mapa:

Choose an option 

Selecione ao menos um circuito para visualizar os pontos no mapa.

Figura 24 - Representação da caixa de seleção

Fonte: Streamlit; Elaboração própria

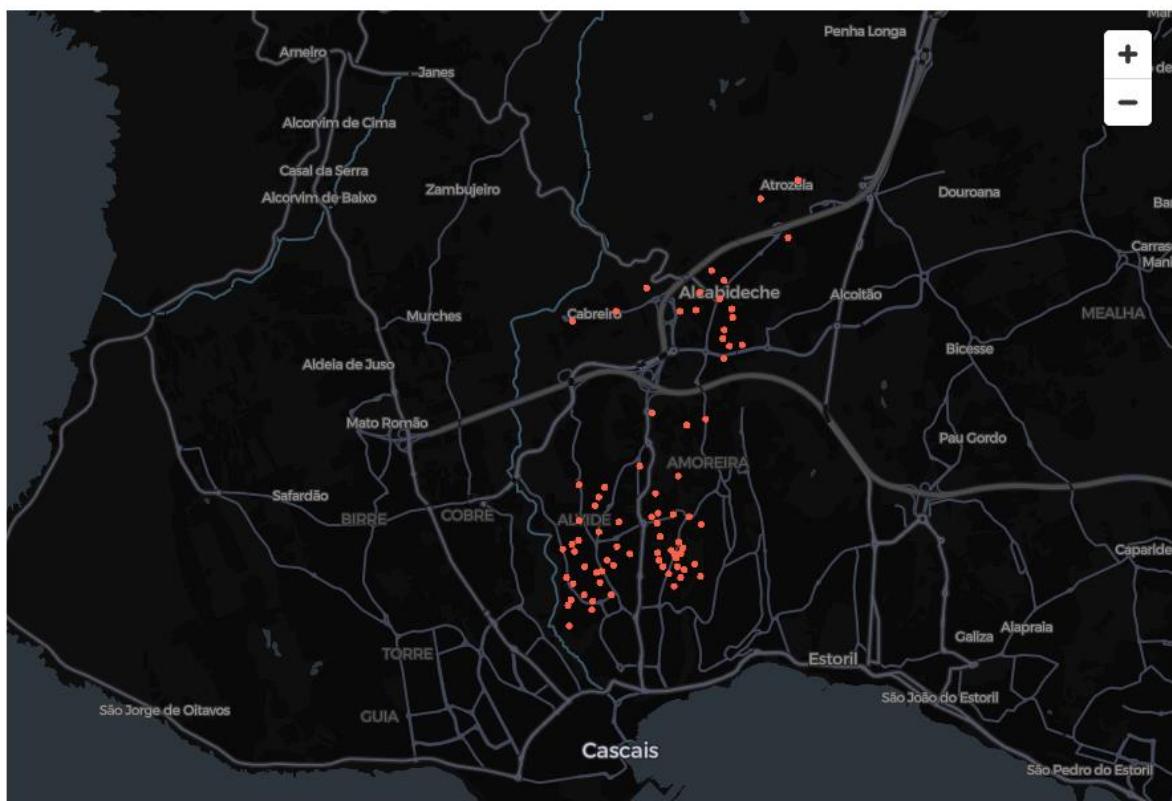


Figura 25 - Mapa com os pontos do Circuito 10

Fonte: Elaboração própria

5.2.5 Gráficos de um ponto aleatório por circuito

Com o objetivo de analisar o comportamento dos ecopontos ao longo do tempo de forma mais aprofundada, foi selecionado aleatoriamente um ponto por circuito e analisado separadamente para cada tipo de resíduo: papel, embalagens e vidro.

Para cada um destes resíduos, foram criados três tipos de visualizações gráficas:

- Frequência de recolhas: permite observar quantas vezes o ponto foi recolhido ao longo do tempo, o que permite detetar padrões regulares, irregularidades ou sazonalidade.

- Série temporal dos níveis de enchimento: mostra a evolução contínua dos níveis de enchimento ao longo dos dias, o que ajuda a perceber o comportamento típico de acumulação de resíduos no ponto em questão.
- Intervalo entre recolhas: apresenta o número de dias entre recolhas consecutivas.

As figuras abaixo apresentam os três tipos de gráficos criados para o ponto 3712 do Circuito 12. Neste caso, foi exemplificado apenas para papel, sendo apresentados: a frequência de recolhas, a série temporal dos níveis de enchimento e o intervalo entre recolhas.

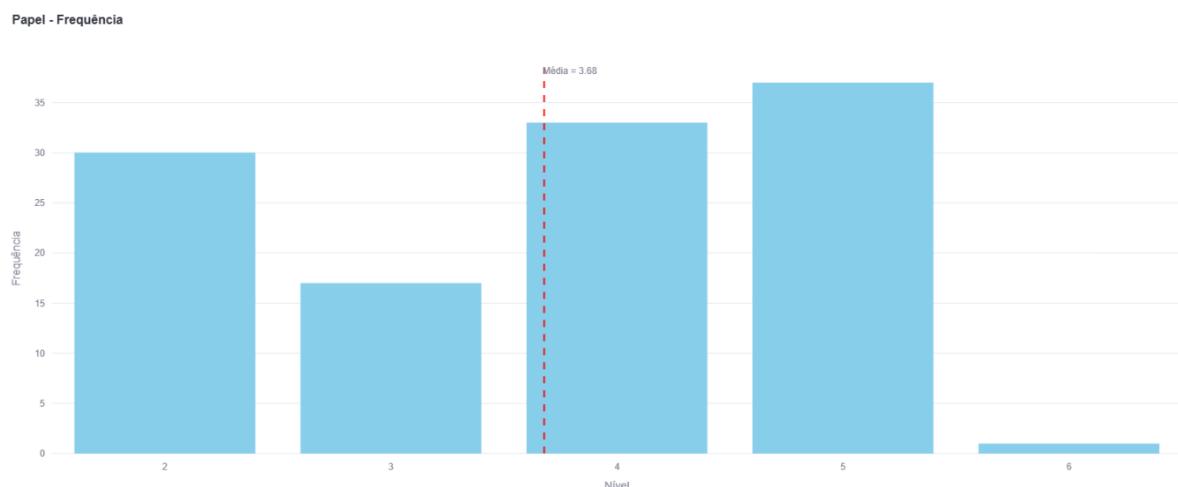


Figura 26 - Frequência de papel para o ponto 3712

Fonte: Elaboração própria

Papel - Série Temporal



Figura 27 - Série temporal do nível de enchimento de papel do ponto 3712

Fonte: Elaboração própria

Papel - Intervalo Recolhas

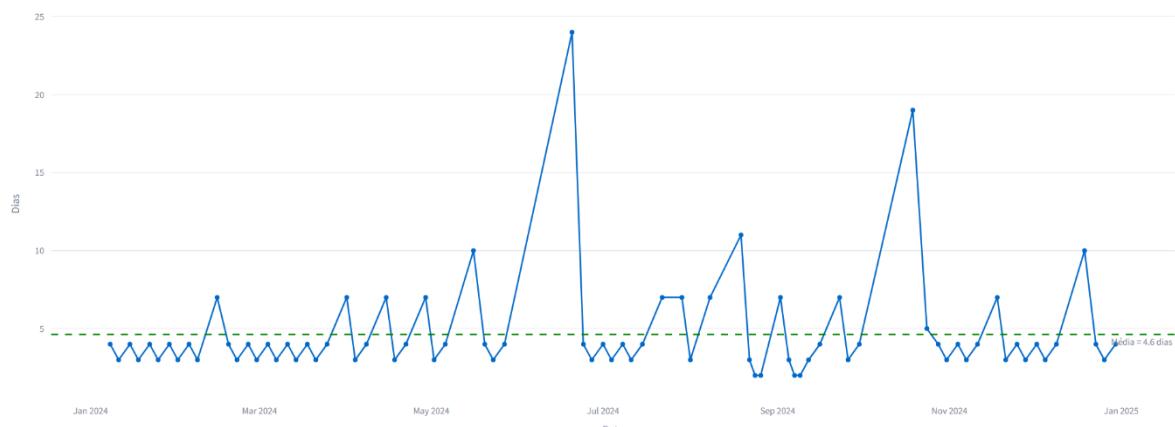


Figura 28 - Intervalo de Recolhas de papel para o ponto 3712

Fonte: Elaboração própria

Depois de uma análise geral para todos os gráficos de cada circuito, chegou-se à seguinte conclusão:

- O nível 1 em Papel não foi registado.
- Em relação às recolhas, Embalagens e Papel apresentam intervalos de recolha mais regulares.
- O Vidro apresenta um espaçamento maior entre recolhas.
- A quantidade de informação disponível varia por tipo, sendo o Vidro o que tem menos dados registados.

5.2.6 Evolução temporal dos níveis de enchimento por circuito

Nesta subsecção, são apresentados gráficos que ilustram a evolução temporal dos níveis de enchimento dos ecopontos, agrupados por circuito e organizados segundo o tipo de resíduo: papel, embalagens e vidro. Para cada tipo de resíduo, foi criado um gráfico onde se visualiza a evolução dos níveis de enchimento ao longo do tempo para todos os circuitos existentes. Nas figuras abaixo, estão representados estes gráficos para o Circuito 04.



Figura 29 - Evolução temporal (papel) do Círculo 12

Fonte: Elaboração própria

Círculo 04 – Evolução Temporal (Embalagens)

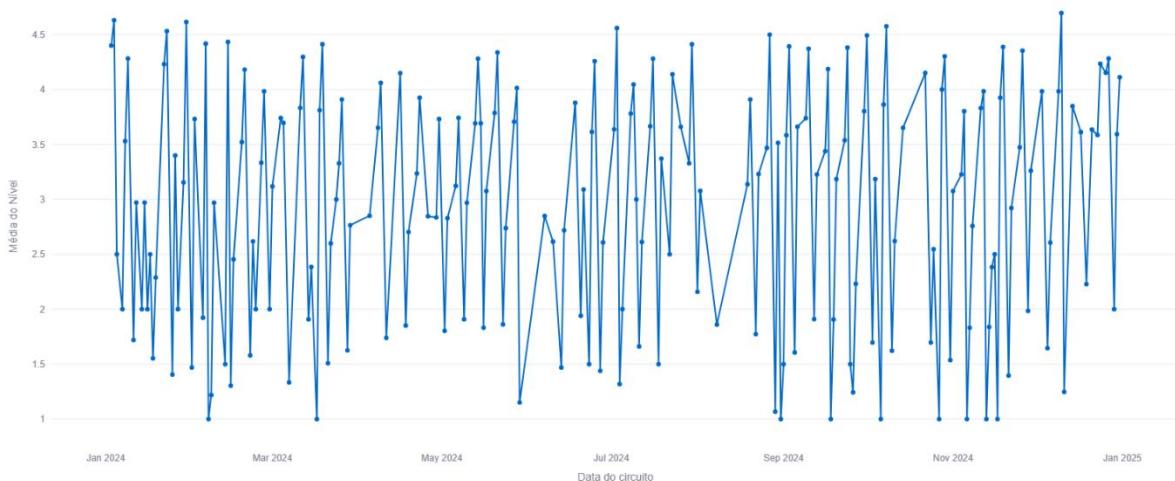


Figura 30 - Evolução temporal (embalagens) do Círculo 12

Fonte: Elaboração própria

Círculo 04 – Evolução Temporal (Vidro)

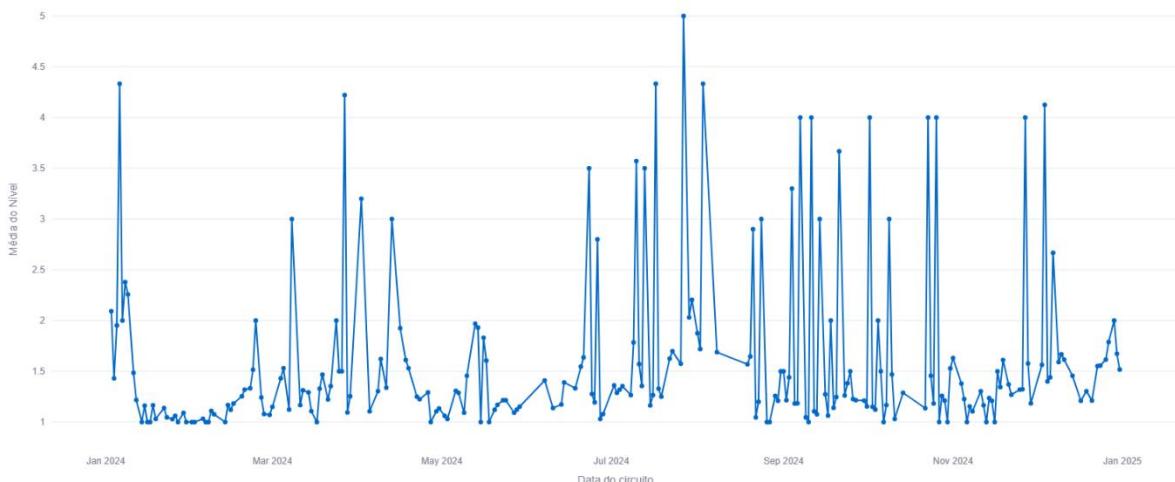


Figura 31 - Evolução temporal (vidro) do Círculo 12

Fonte: Elaboração própria

5.2.7 Quantidade de lixo recolhido por tipo e circuito vs total

Nesta subsecção é possível ver duas visualizações distintas que permitem compreender melhor a distribuição da quantidade de resíduos recolhidos, tanto a nível individual por circuito como de forma agregada por tipo de resíduo. A Figura 32 mostra a quantidade total de lixo recolhido por tipo de resíduo (papel, embalagens e vidro) em cada circuito. Esta visualização permite identificar quais os circuitos que geram maior volume de resíduos, bem como o tipo predominante em cada um. A Figura 33 apresenta a proporção total de lixo recolhido no município, agregando todos os circuitos, dividido por tipo de resíduo. Estes gráficos são úteis para perceber a representatividade relativa de cada tipo de resíduo na recolha global.

Quantidade de lixo recolhido por tipo e circuito

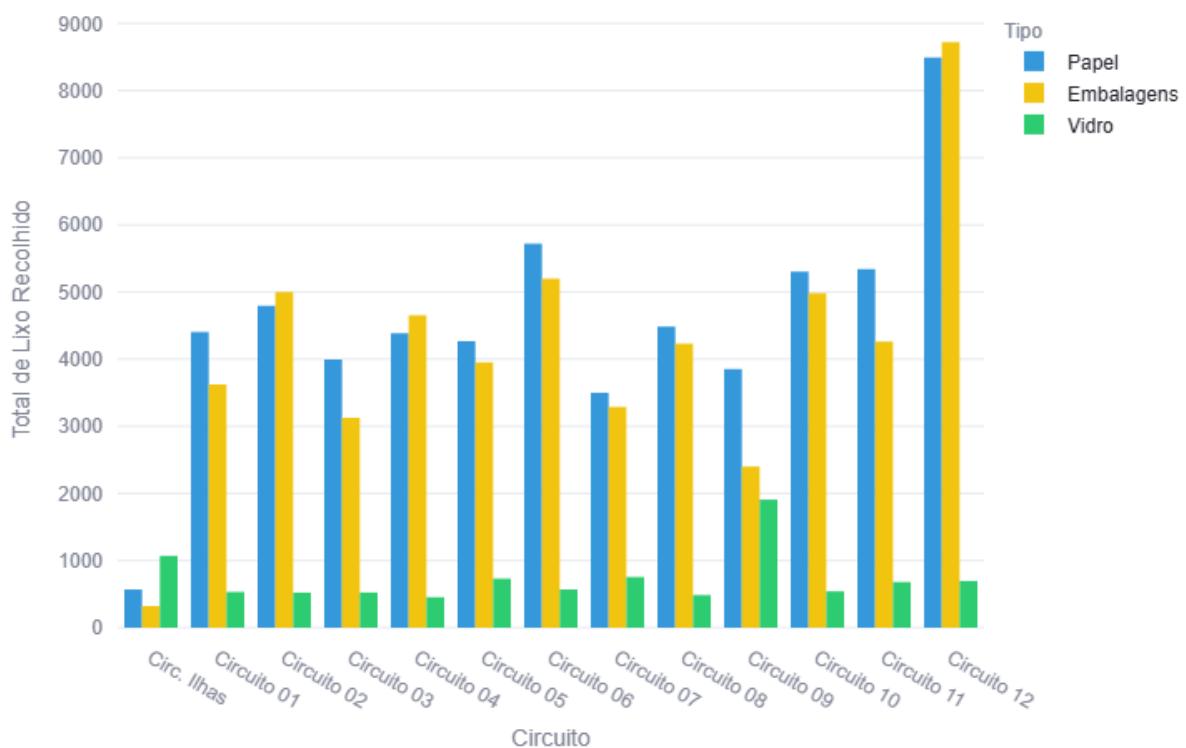


Figura 32 - Quantidade total de lixo recolhido por tipo e circuito

Fonte: Elaboração própria

Proporção total de lixo recolhido por tipo

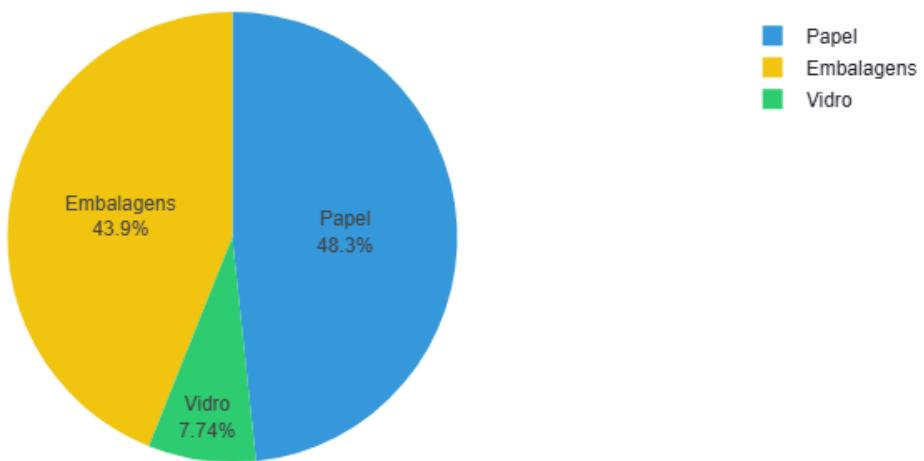


Figura 33 - Proporção de lixo recolhido por tipo de resíduo

Fonte: Elaboração própria

Como é possível observar, o tipo de resíduo com maior volume recolhido é o papel, seguido pelas embalagens, e por fim o vidro, que representa a menor fração.

5.2.8 Evolução temporal do lixo recolhido por tipo

Nesta subsecção é possível ver a evolução diária da quantidade de lixo recolhida ao longo do ano de 2024 para cada tipo de resíduo: Papel, Embalagens e Vidro. Para isso, foram construídos três gráficos separados, um por tipo, onde se representa a quantidade total recolhida por dia, considerando todos os circuitos. Na Figura 34, 35 e 36 estão representados respetivamente, os gráficos para Papel, Embalagens e Vidro.

Evolução Temporal do lixo recolhido de Papel

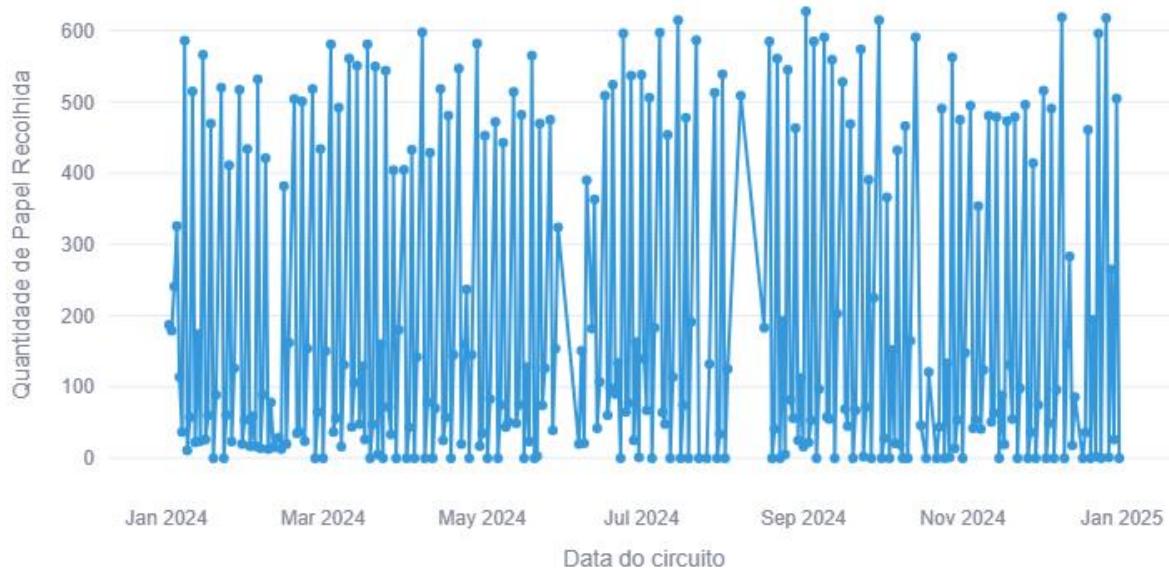


Figura 34 - Gráfico da evolução temporal do lixo recolhido para Papel

Fonte: Streamlit; Elaboração própria

Evolução Temporal do lixo recolhido de Embalagens

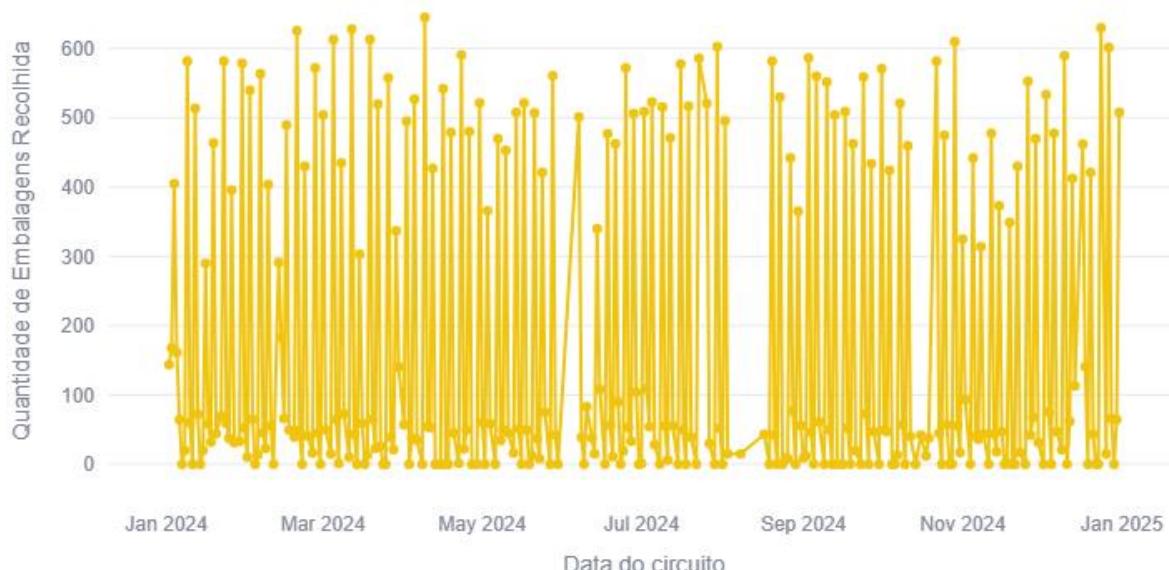


Figura 35 - Gráfico da evolução temporal do lixo recolhido para Embalagens

Fonte: Streamlit; Elaboração própria

Evolução Temporal do lixo recolhido de Vidro

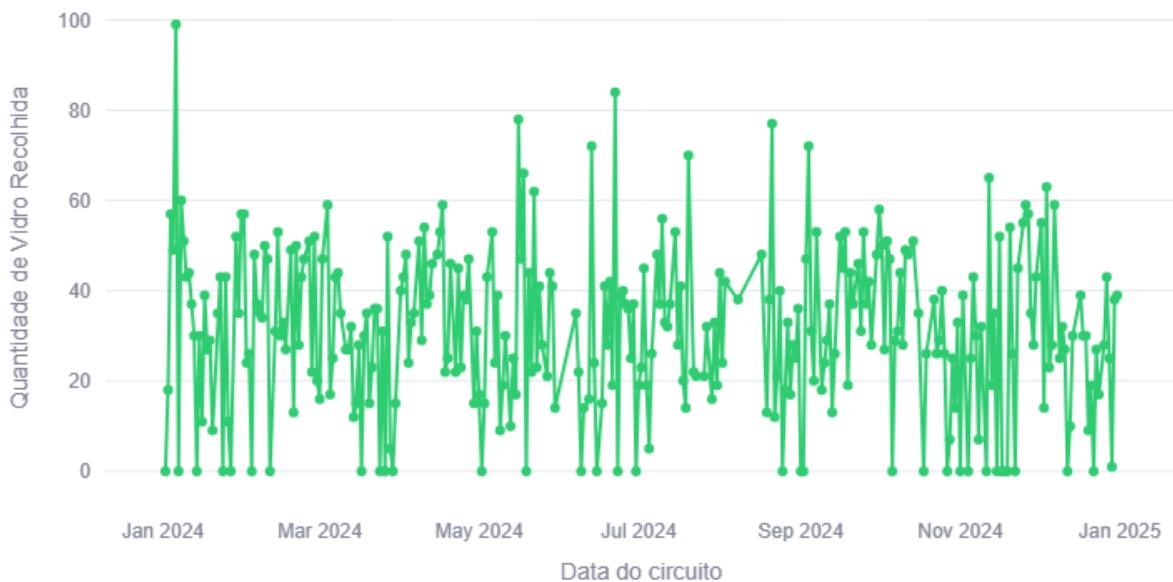


Figura 36 - Gráfico da evolução temporal do lixo recolhido para Vidro

Fonte: Streamlit; Elaboração própria

A análise visual destes gráficos permitiu identificar alguns padrões relevantes:

- A evolução ao longo do tempo das recolhas de Papel e Embalagens mostra um comportamento bastante semelhante, com valores próximos e variações parecidas ao longo dos dias. Este padrão pode indicar que as pessoas costumam deitar fora estes dois tipos de resíduos de forma parecida, talvez porque estão relacionados com os mesmos hábitos de consumo, como o uso doméstico e comercial.
- Por outro lado, o Vidro apresenta valores de recolha consistentemente mais baixos, tanto em quantidade como na frequência com que é recolhido. Isto pode estar relacionado com o menor uso de embalagens de vidro no dia a dia, em comparação com papel e cartão, que são mais comuns em embalagens de produtos e encomendas, por exemplo.

5.3 Módulo das Rotas Reais

Este módulo da aplicação permite ao utilizador visualizar as rotas reais de recolha de resíduos realizadas no município, com base nos dados históricos disponibilizados.

A interface permite filtrar a informação através da seleção de três parâmetros principais:

- Data da operação.
- Circuito de recolha.
- Tipo de resíduo (papel, embalagens ou vidro).

Após a escolha da combinação desejada, é apresentado um mapa onde é desenhada a rota efetuada nessa data, circuito e tipo de resíduo. Este mapa proporciona uma visualização geográfica clara do trajeto seguido pelo veículo. Adicionalmente, são disponibilizadas estatísticas da rota, que incluem:

- Distância total percorrida.
- Duração estimada da rota.
- Quantidade de lixo não recolhido.
- Número de pontos visitados.
- Ocupação do camião (máximo é 120).

A rota visualizada tem sempre início no Estaleiro e termina no TratoLixo, representados no mapa, respetivamente, por um ponto verde (origem) e um ponto vermelho (destino). Estes dois pontos foram também fornecidos pela Câmara Municipal de Cascais e refletem o modelo real de operação adotado pelo município, onde os veículos partem do Estaleiro no início do turno e descarregam os resíduos no TratoLixo no final da recolha.

A título de exemplo vai se usar o dia 02/01/2024, o Circuito 03 e o tipo de resíduos Embalagens. O filtro para selecionar a rota a ser visualizada pode ser observado na Figura 37, que esta representada abaixo.

Selecione a data

2024/01/02

Selecione o circuito:

Círculo 03

Selecione o tipo de lixo

Embalagens

Figura 37 - Filtro para seleção da rota a visualizar

Fonte: Streamlit; Elaboração própria

Na figura abaixo, está representada a rota efetuada com os parâmetros da Figura 37, acima mencionada.

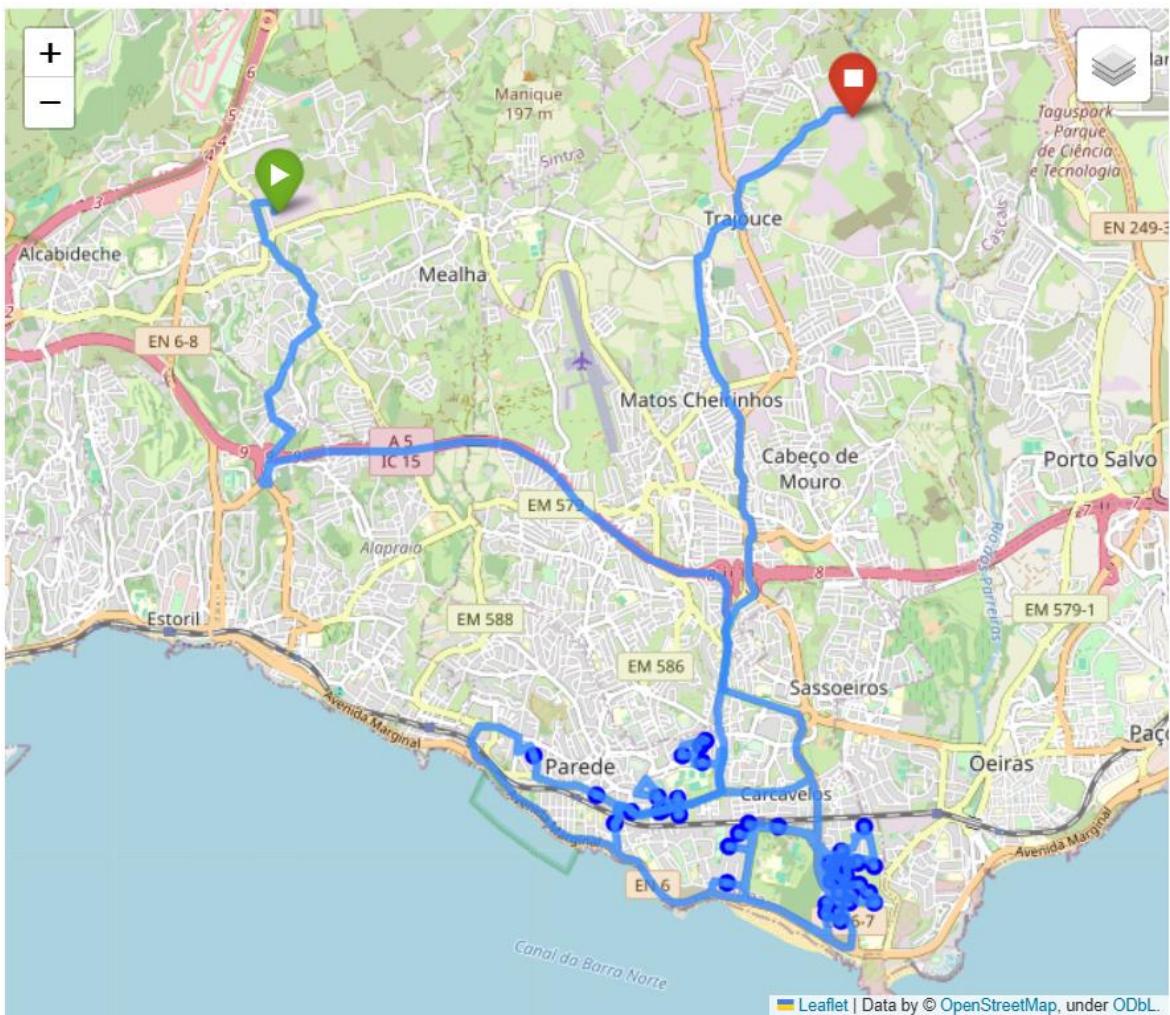


Figura 38 - Rota efetuada no dia 02/01/2024, no Circuito 03 para a recolha de Embalagens

Fonte: OpenStreetMap; Elaboração própria

Na figura acima, a linha azul representa o trajeto realizado pelo camião da recolha e os pontos azuis, representam os ecopontos visitados.

Na Figura 39, que se encontra em baixo, é possível ver as estatísticas desta rota.



Figura 39 - Estatísticas da rota

Fonte: Elaboração própria

É importante salientar que os valores apresentados para a distância e a duração da rota são estimativas, obtidas com através de matrizes de tempo e distância. No caso da duração, além do tempo estimado de deslocação entre dois pontos consecutivos, é também considerado um tempo fixo de serviço de cinco minutos por ponto visitado, refletindo a média estimada para a operação de recolha em cada ecoponto.

5.4 Módulo Otimização de Rotas (Solver)

Este módulo da aplicação permite visualizar as rotas otimizadas geradas por um solver com base no problema de roteamento de veículos. À semelhança do módulo anterior, o utilizador pode selecionar a data, o circuito e o tipo de resíduo. A aplicação processa essa informação e apresenta no ecrã:

- Um mapa com camadas com as rotas calculadas
- O mesmo conjunto de estatísticas do módulo anterior:
 - Distância total da rota.
 - Duração estimada da operação.
 - Quantidade de lixo não recolhido.
 - Número de pontos visitados.
 - Ocupação do camião.

Este módulo permite comparar diferentes abordagens ao problema da recolha de resíduos, fornecendo uma perspetiva quantitativa sobre a eficiência das rotas geradas com base em diferentes dados de entrada. As rotas são criadas através do solver PyVRP, uma biblioteca especializada na resolução de problemas do tipo VRP. Neste caso foi escolhido uma variante do CVRP, que é o PCVRP, explicado no ponto 1.3.3.

Na Figura 40, estão representadas as caixas de seleção onde se escolhe o dia, circuito e tipo de resíduo para depois o solver filtrar a informação pretendida. Neste exemplo escolheu-se o dia 2 de fevereiro, o Circuito 02 e Embalagens para o tipo de resíduo a recolher.

Solver - PyVRP

Selecione a data:

2024/02/16

Selecione o circuito:

Círculo_02

Selecione o tipo:

Embalagens

Figura 40 - Representação das caixas de seleção do solver para o dia, circuito e tipo

Fonte: Streamlit; Elaboração própria

Na Figura 41, 42 e 43 encontram-se os mapas para cada tipo de rota que foram criados, que são as rotas mencionadas no ponto 3.10: Rota de Início, Rota de Fim e Sub-rota.

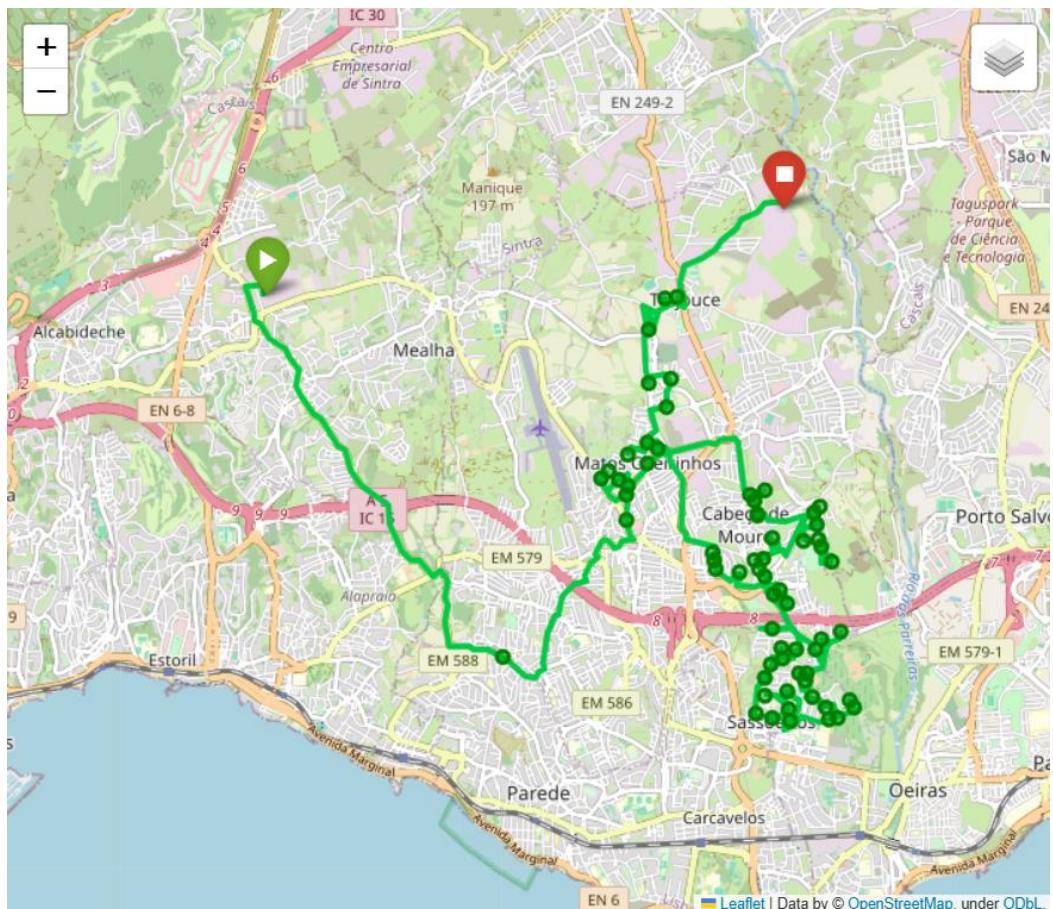


Figura 41 - Rota de Início efetuada no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens

Fonte: OpenStreetMap; Elaboração própria

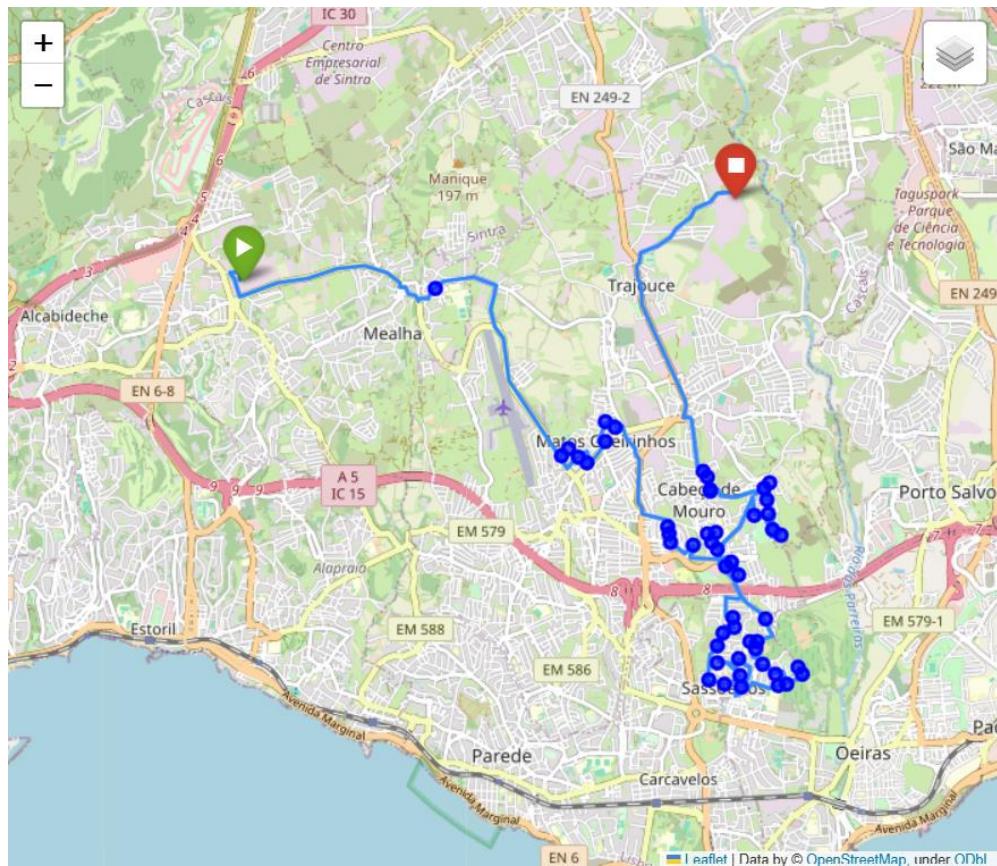


Figura 42 - Rota de Fim efetuada no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens

Fonte: OpenStreetMap; Elaboração própria

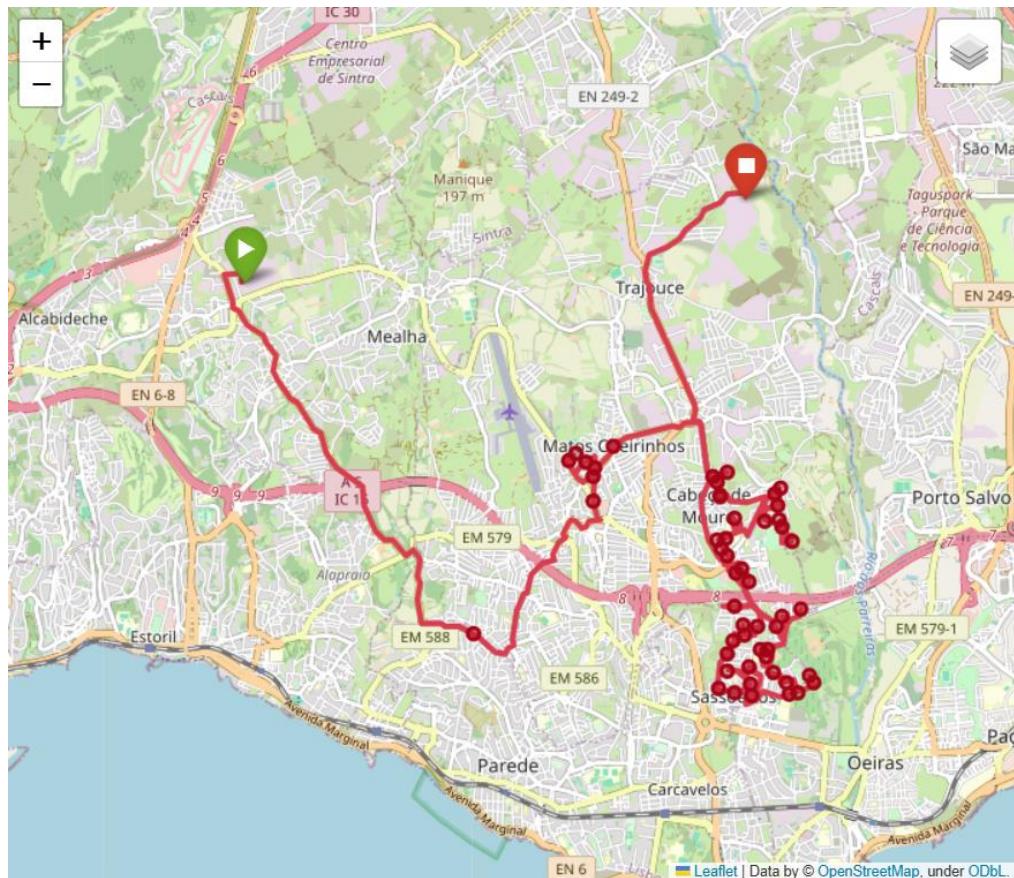


Figura 43 - Sub-Rota efetuada no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens

Fonte: OpenStreetMap; Elaboração própria

Também como disponível no módulo das rotas reais, são apresentadas as estatísticas para cada rota, como se pode observar na figura seguinte.

Rota de Início

- Distance: 34.57 km
- Duração: 420 minutos
- Pontos visitados: 65
- Quantidade de lixo não recolhido: 5
- Ocupação total do camião: 74

Rota de Fim

- Distance: 25.54 km
- Duração: 319 minutos
- Pontos visitados: 50
- Quantidade de lixo não recolhido: 36
- Ocupação total do camião: 120

Sub Rota

- Distance: 26.87 km
- Duração: 333 minutos
- Pontos visitados: 52
- Quantidade de lixo não recolhido: 36.0
- Ocupação total do camião: 120.0

Figura 44 - Estatísticas das rotas efetuadas no dia 16/02/2024, no Circuito 02 para a recolha de Embalagens

Fonte: Elaboração própria

6. OUTROS TESTES REALIZADOS

Posteriormente, foi desenvolvido um simulador baseado numa rede neuronal para prever os níveis de enchimento e criar rotas de recolha otimizadas. As previsões foram baseadas nas estatísticas previamente calculadas no ponto 3.7.

A rede foi treinada com os dados históricos recolhidos entre o mês de janeiro e novembro, sendo o mês de dezembro reservado para validação. Esta divisão permite que a avaliação é feita sobre dados nunca vistos pelo modelo durante a fase de treino, o que permite aferir a sua capacidade de generalização.

6.1 Época de Treino

Os resultados obtidos durante a fase de treino permitiam concluir que em média, o modelo apresentava resultados inferiores aos obtidos pelo solver. Nas figuras seguintes é possível ver alguns gráficos obtidos para a época de treino, nomeadamente o retorno, quantidade de pontos não visitados e quantidade de lixo não recolhida (em decimal). Na Figura 38 encontra-se um gráfico com o retorno da rede neuronal para a época de treino.

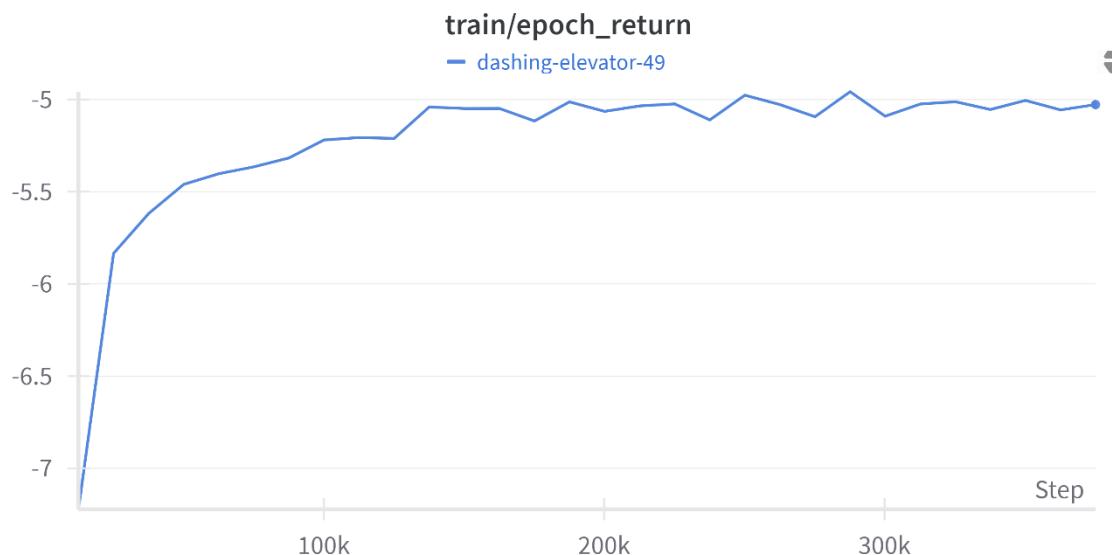


Figura 45 – Média do retorno da rede neuronal na época de treino

Fonte: Weights & Biases: The AI Developer Platform

Neste gráfico é possível observar o retorno médio para a época de treinamento. Em aprendizagem por reforço, o retorno é tipicamente a soma dos prémios recebidos durante um episódio ou época. É possível observar que o retorno aumentou de -7 para -5. Isso indica que o agente aprendeu comportamentos iniciais que melhoraram o seu desempenho e depois a curva fica estável nesse valor. Assim, o modelo parece estável, mas não ótimo pois os retornos ainda são negativos e deveriam ser positivos.

Na Figura 39, que se encontra a seguir, é possível ver um gráfico com a média de quantidade de pontos não visitados na época de treino.

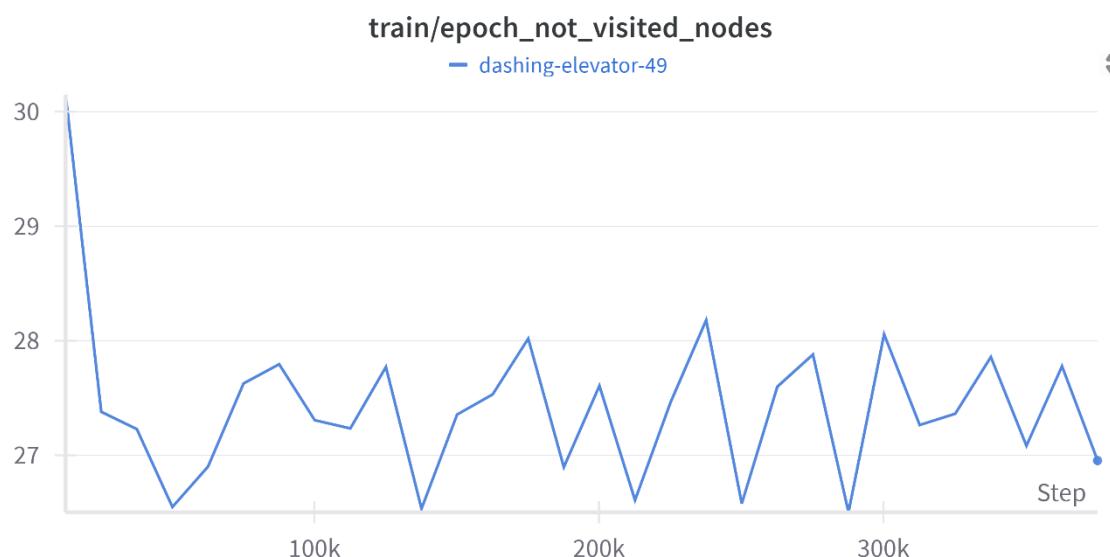


Figura 46 - Média de pontos não visitados pela rede neuronal na época de treino

Fonte: Weights & Biases: The AI Developer Platform

Como é possível observar, o agente começa por deixar cerca de trinta pontos por visitar, mas rapidamente passa a deixar cerca de vinte e sete, o que reflete uma melhoria na exploração. Idealmente a curva no gráfico teria de ficar mais suave e constante o que não aconteceu.

Na Figura 40, que se encontra a seguir, está representado o gráfico com a média de lixo não recolhido na época de treino.

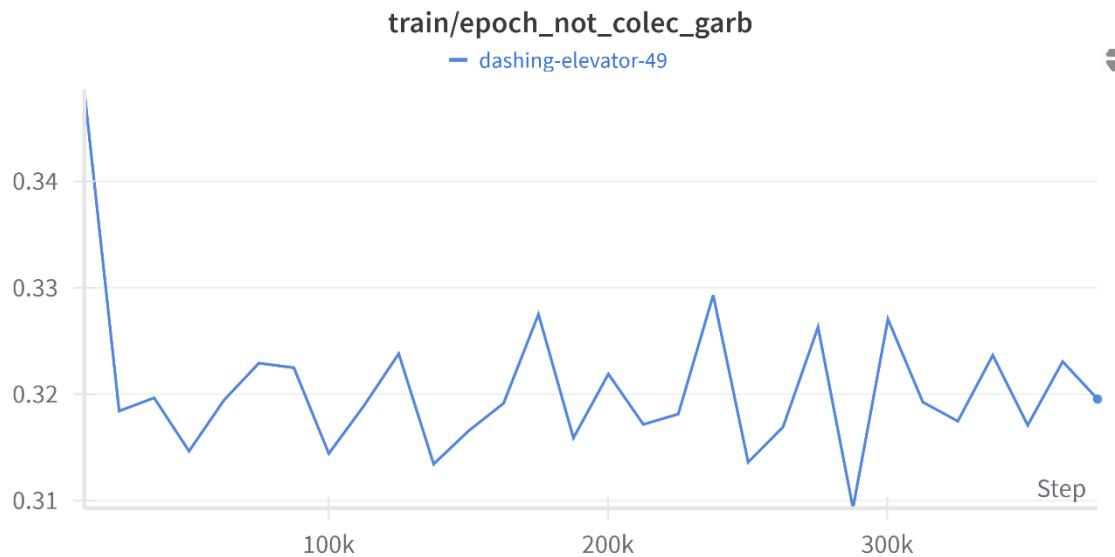


Figura 47 - Média de lixo não recolhido pela rede neuronal na época de treino

Fonte: (Weights & Biases: The AI Developer Platform, sem data)

Neste gráfico, observa-se que a média da quantidade de lixo não recolhido pela rede neuronal diminuiu inicialmente, mas depois oscilou ao longo do tempo, o que indica a ausência de uma tendência clara de melhoria contínua. Idealmente, esse valor deveria permanecer abaixo de 0,30 para que a rede pudesse superar as rotas geradas pelo solver.

6.2 Época de Validação

Os resultados obtidos para a época de validação foram bastante insatisfatórios, o que indica que provavelmente o modelo sofreu de *overfitting*, ou seja, ajustou-se demasiado aos dados de treino e não conseguiu generalizar bem para dados novos. Na figura seguinte, é possível ver o retorno médio da rede neuronal para a época de validação.

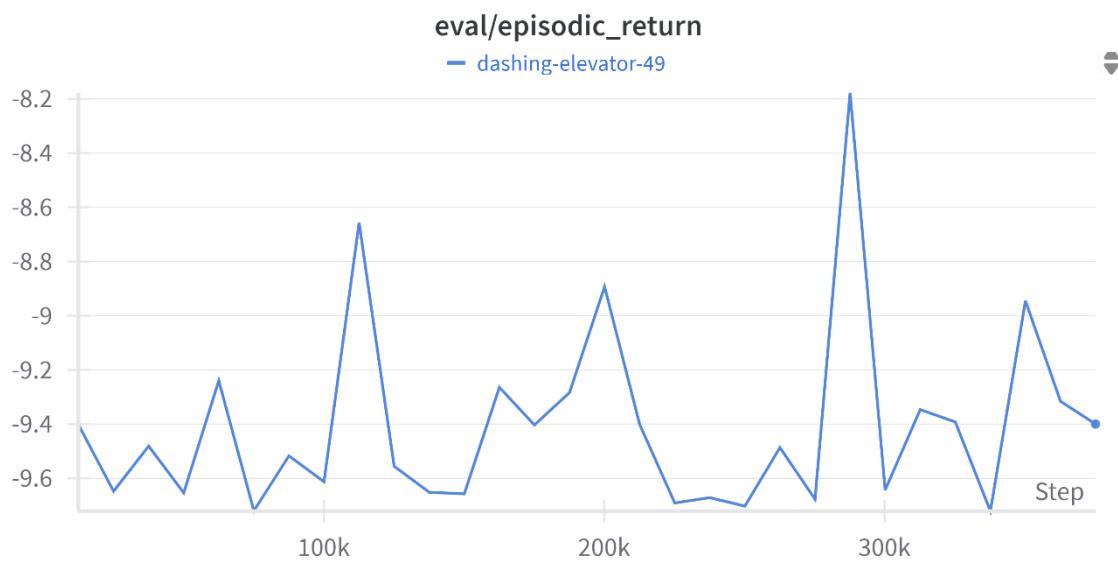


Figura 48 - Média do retorno da rede neuronal para a época de validação

Fonte: Weights & Biases: The AI Developer Platform

Neste gráfico é possível ver que o retorno também é negativo. Não há uma tendência clara de melhoria consistente ao longo do tempo. Apesar de alguns picos de melhoria, o retorno volta a cair depois.

Na Figura 42, que se encontra a seguir, é possível ver o gráfico para a quantidade de pontos não visitados pelo modelo durante a época de validação.

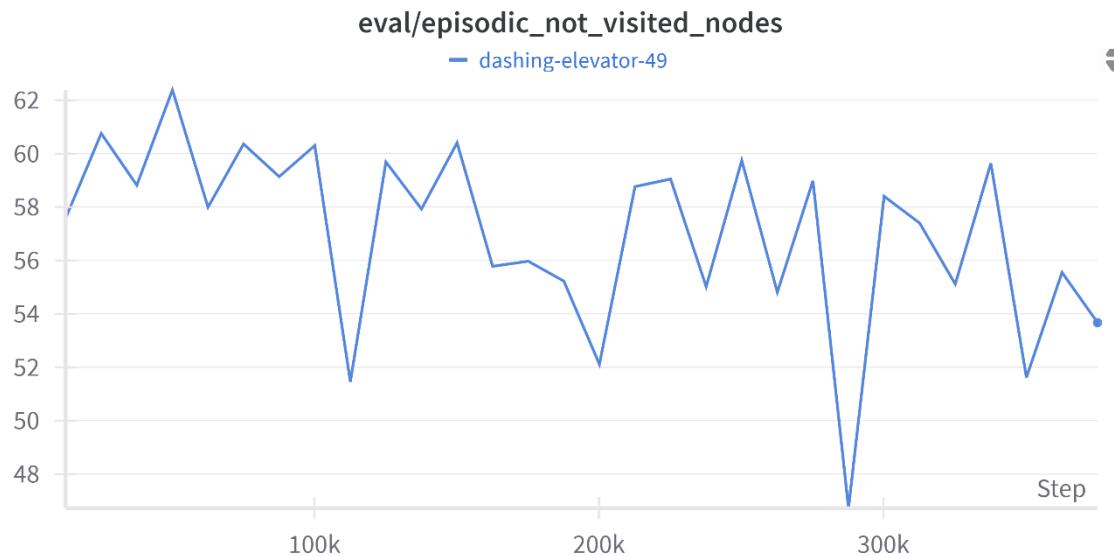


Figura 49 - Média de pontos não visitados pela rede neuronal na época de validação

Fonte: Weights & Biases: The AI Developer Platform

Nesta figura, é possível ver que o número de pontos não visitados oscila significativamente sem uma tendência clara e com uma tendência continua de queda. Há momentos em que o número de nós não visitados cai, mas rapidamente volta a subir o que mostra falta de consistência. A meta ideal seria ver a curva decrescendo consistentemente.

Por fim, temos o gráfico do lixo não recolhido pelo modelo durante a época de validação, que se encontra na Figura 43.

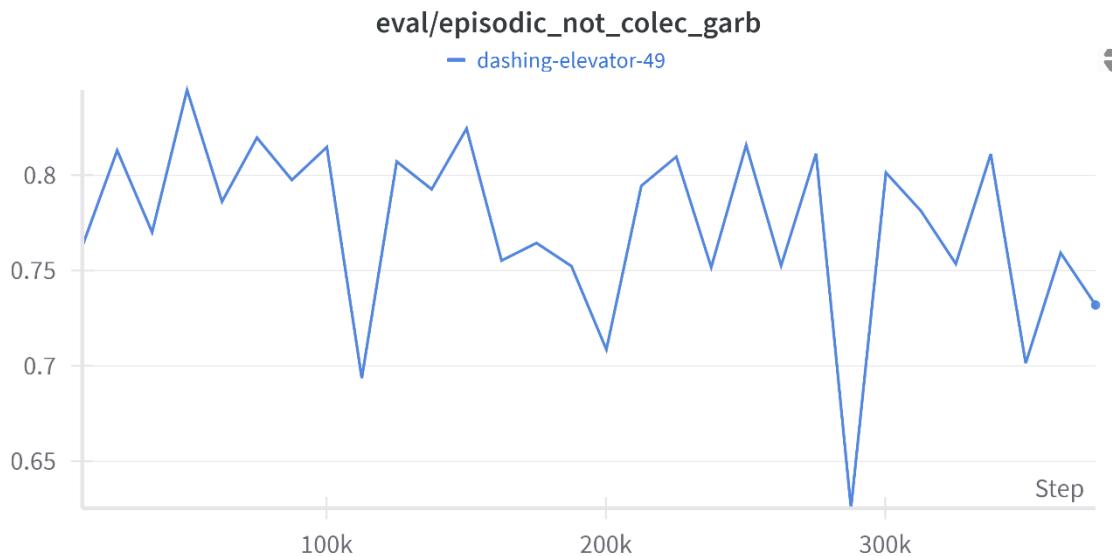


Figura 50 - Média da quantidade de lixo não recolhido pela rede neuronal durante a fase de validação

Fonte: (Weights & Biases: The AI Developer Platform, sem data)

É possível observar que o agente não está a recolher o lixo de forma consistente, dado o elevado grau de variação nos valores ao longo do tempo. Embora haja momentos pontuais de melhoria, como o valor mínimo em torno de 0.63, este ainda se encontra muito aquém do desempenho esperado, que seria idealmente próximo de 0.30 para esta métrica.

6.3 Conclusão do resultado obtido

Os resultados obtidos durante a época de validação não foram totalmente satisfatórios, apresentando valores abaixo do esperado nas variáveis avaliadas. Verificou-se uma elevada oscilação nos desempenhos ao longo do treino e na validação, com oscilações frequentes tanto no retorno acumulado como nas tarefas específicas, como a recolha de lixo e a visita de pontos.

Uma possível explicação para este comportamento instável poderá ser o *overfitting*, previamente explicado no ponto 6.2. Para além disso, a possível falta de dados consistentes longo do processo pode ter dificultado a aprendizagem.

CONCLUSÃO E TRABALHOS FUTUROS

O presente trabalho teve como principal objetivo realizar um estudo para a otimização da recolha de resíduos urbanos no município de Cascais, combinando a previsão do nível de enchimento dos contentores com a criação de rotas otimizadas. Através da análise dos dados históricos recolhidos pela Câmara Municipal, foi possível estimar o comportamento dos ecopontos ao longo do tempo e construir modelos preditivos com base nessa informação.

Os resultados obtidos com o modelo preditivo, nomeadamente o *HistGradientBoostingRegressor* revelou-se positivo, conseguindo prever os níveis de enchimento dos contentores de forma mais precisa do que com a taxa de enchimento mediana. No entanto, o simulador criado para uma abordagem com redes neurais não apresentou resultados satisfatórios, possivelmente devido à complexidade do problema, eventuais inconsistências ou como referido, problemas de *overfitting*.

A integração dos resultados preditivos com o solver PyVRP permitiu criar rotas de recolha mais eficientes, adaptadas ao comportamento esperado dos contentores, promovendo uma operação mais sustentável e com menor desperdício de recursos.

Como continuação deste trabalho, existem várias oportunidades de melhoria e aprofundamento. Em primeiro lugar, pode-se refinar o modelo preditivo, quer através da incorporação de novas variáveis ou através da utilização de outros algoritmos *de machine learning*.

Adicionalmente, será importante melhorar o simulador desenvolvido para a rede neuronal, explorando arquiteturas mais robustas no sentido de obter resultados positivos.

Em suma, este estudo demonstrou o potencial da combinação entre previsão e otimização para melhorar significativamente a gestão da recolha de resíduos, constituindo uma base sólida para futuras soluções mais inteligentes, sustentáveis e adaptáveis.

REFERÊNCIAS

- A brief introduction to VRP - PyVRP 0.12.0a0 documentation.* (sem data). Obtido 25 de Maio de 2025, de https://pyvrp.org/setup/introduction_to_vrp.html
- de Moraes e Soares, R. (2024). Evaluating the efficiency of public expenditure in municipal waste collection: A comparative study of Portuguese municipalities. *Environmental Economics*, 15(1), 203–216. [https://doi.org/10.21511/ee.15\(1\).2024.15](https://doi.org/10.21511/ee.15(1).2024.15)
- Freepik | All-in-One AI Creative Suite.* (sem data). Obtido 21 de Junho de 2025, de <https://www.freepik.com/>
- Herrera-Granda, I. D., Cadena-Echeverría, J., León-Jácome, J. C., Herrera-Granda, E. P., Chavez Garcia, D., & Rosales, A. (2024). A Heuristic Procedure for Improving the Routing of Urban Waste Collection Vehicles Using ArcGIS. *Sustainability (Switzerland)*, 16(13). <https://doi.org/10.3390/su16135660>
- Leal, C. H., & Pereira, S. (sem data). *Universidade de Aveiro 2022*.
- Li, W., Theam Foo, N., Haidi, I., & and Wang, S. L. (2025). A literature review of the state of the art of sustainable waste collection and vehicle routing problem. *Journal of the Air & Waste Management Association*, 75(1), 3–26. <https://doi.org/10.1080/10962247.2024.2415298>
- Liu, F., Lu, C., Gui, L., Zhang, Q., Tong, X., & Yuan, M. (2023). *Heuristics for Vehicle Routing Problem: A Survey and Recent Advances*. <http://arxiv.org/abs/2303.04147>
- OpenStreetMap.* (sem data). Obtido 27 de Maio de 2025, de <https://www.openstreetmap.org/about>
- Prata, J., Simões, C. L., & Simoes, R. (2025). Improvements to Municipal Solid Waste Collection Systems Using Real-Time Monitoring. *Sustainability (Switzerland)*, 17(4). <https://doi.org/10.3390/su17041405>
- Smart Waste | Cascais Data.* (sem data). Obtido 21 de Junho de 2025, de <https://data.cascais.pt/ambiente-energia/smart-waste>

Streamlit • A faster way to build and share data apps. (sem data). Obtido 27 de Maio de 2025, de <https://streamlit.io/>

Weights & Biases: The AI Developer Platform. (sem data). Obtido 19 de Junho de 2025, de <https://wandb.ai/site/>