



PEDRO MIGUEL ALMEIDA CARVALHO BRANCO ALVES

Licenciado em Ciências de Engenharia e Gestão Industrial

ALGORITMO DE CRIAÇÃO DE ROTAS PARA ENTREGA DE MERCADORIAS: ESTUDO DE CASO

MESTRADO INTEGRADO EM ENGENHARIA E GESTÃO INDUSTRIAL

Universidade NOVA de Lisboa

setembro, 2024

ALGORITMO DE CRIAÇÃO DE ROTAS PARA ENTREGA DE MERCADORIAS: ESTUDO DE CASO

PEDRO MIGUEL ALMEIDA CARVALHO BRANCO ALVES

Licenciado em Ciências de Engenharia e Gestão Industrial

Orientadora: Ana Paula Barroso,
Professora Auxiliar, Universidade NOVA de Lisboa

Algoritmo de criação de rotas para entrega de mercadorias: Estudo de caso

Copyright © Pedro Miguel Almeida Carvalho Branco Alves, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

À minha família e à Maria, por toda a ajuda e apoio.

AGRADECIMENTOS

Quero começar por expressar a minha gratidão à professora Ana Barroso pelo acompanhamento e orientação. Sem o seu apoio e dedicação, a conclusão deste trabalho não teria sido possível. Agradeço pela ajuda, pelos conselhos valiosos, correções cuidadosas e sugestões que me permitiram crescer academicamente.

Também quero agradecer ao Engenheiro Luís Mascarenhas, ao Pramod Maugi e ao Ricardo Ferro da TNB. As suas colaborações foram essenciais para a obtenção dos dados e informações necessárias para a realização desta dissertação. Além disso, as suas orientações internas ajudaram-me a compreender o funcionamento da empresa e a esclarecer diversas dúvidas.

Por fim, gostaria de dedicar um especial agradecimento à minha família e amigos e à Maria, que sempre me apoiaram e motivaram a continuar. O vosso apoio incondicional foi fundamental para superar os desafios desta etapa, que se mostrou exigente, mas transformadora.

RESUMO

A melhoria do planeamento das rotas de entrega é de grande importância na atividade de empresas de logística, como a TNB, para aumentar a eficiência operacional e a satisfação dos clientes. Atualmente, o processo de criação de rotas é ineficiente, pois consome tempo excessivo e está sujeito a erros que impactam negativamente a produtividade e os custos operacionais.

Este estudo propõe o desenvolvimento de um algoritmo para a criação de rotas de entrega, e para implementar numa base diária. Utiliza dados como o peso, o volume, as coordenadas geográficas e as janelas horárias das entregas, respeitando as restrições de capacidade das carrinhas e os horários pré-estabelecidos pelos clientes.

A abordagem adotada envolve a análise de dados históricos de entregas realizadas para modelar os tempos de entrega através de regressão linear múltipla, incluindo variáveis como o tempo de serviço, o peso, o volume e o número de caixas. Com base nesses dados, o algoritmo desenvolvido organiza as rotas de forma a reduzir a distância total percorrida, garantindo o cumprimento das janelas horárias e resultando numa utilização da frota mais eficiente, sejam carrinhas elétricas ou a gasolina.

Os resultados obtidos na aplicação do algoritmo a um mês de planeamento de rotas mostram uma redução no tempo consumido na criação das rotas e um aumento na eficiência das entregas, refletindo-se em menores custos operacionais e maior satisfação dos clientes. A implementação do algoritmo permite que a empresa realize uma melhor utilização da sua frota, alinhando-se com os objetivos de sustentabilidade da empresa. Além disso, contribui para a diminuição do consumo de combustível, promovendo benefícios económicos e ambientais. Este estudo evidencia a importância da automatização e da análise de dados na gestão de operações logísticas, propondo uma solução eficaz para a criação de rotas de entrega que pode ser adaptada a diferentes contextos empresariais. As implicações deste trabalho sugerem que a adoção de algoritmos pode transformar a eficiência operacional de empresas de transporte, proporcionando vantagens competitivas no mercado.

Palavras chave: Planeamento de rotas, Algoritmo de criação de rotas, Logística, Eficiência operacional, Análise de dados

ABSTRACT

The improvement of delivery routes is of great importance for logistics companies, such as TNB, to improve operational efficiency and customer satisfaction. Currently, the process of route planning is inefficient, as it is very time-consuming, and prone to errors, which negatively impact productivity and operational costs.

This study proposes the development of an algorithm for creating delivery routes and implementing daily. Utilizing data such as weight, volume, geographic coordinates, and delivery time windows, while respecting the trucks' capacity constraints and customer pre-specified delivery times.

The approach involves analyzing historical delivery data to model delivery times through multiple linear regression, including variables such as service time, weight, volume, and number of packages. Based on this data, the developed algorithm organizes routes to reduce total distance travelled, ensuring compliance with delivery time windows, and a better use of the companies' fleet, whether they are electric or gasoline-powered trucks.

The results obtained after applying the algorithm to a full month of route planning show a reduction in the time spent on route planning and an increase in delivery efficiency, resulting in lower operational costs and higher customer satisfaction. The implementation of the algorithm enables the company to improve the use of its fleet, aligning with the company's sustainability goals. Furthermore, it contributes to reduce fuel consumption, promoting both economic and environmental benefits. This study highlights the importance of automation and data analysis in managing logistics operations, proposing an effective solution for delivery route planning that can be adapted to different business contexts. The implications of this work suggest that the adoption of algorithms can transform the operational efficiency of transportation companies, providing competitive advantages in the market.

Keywords: Route planning, Route creation algorithm, Logistics, Operational efficiency, Data analysis

ÍNDICE

1	Introdução	1
1.1	<i>Justificação do tema</i>	1
1.2	<i>Enquadramento</i>	2
1.3	<i>Objetivo</i>	3
1.4	<i>Estrutura do documento</i>	3
1.5	<i>Métodos de Investigação</i>	4
2	Logística e análise de dados no planeamento de rotas de entrega	5
2.1	<i>Logística de armazém no contexto de entregas a cliente</i>	5
2.2	<i>Problemas de planeamento de rotas de veículos e do Caixeiro Viajante</i>	7
2.3	<i>Técnicas de análise de dados</i>	11
3	Caracterização do estudo de caso	19
3.1	<i>Empresa transportadora</i>	19
3.2	<i>Caracterização das mercadorias a transportar</i>	21
3.3	<i>Caracterização da frota e equipas de transporte</i>	21
3.3.1	<i>Caracterização da frota</i>	22
3.3.2	<i>Caracterização das equipas de transporte</i>	22
3.4	<i>Descrição do processo de criação de rotas</i>	23
3.5	<i>Tempo de Serviço</i>	26
3.5.1	<i>Recolha de dados</i>	26
3.5.2	<i>Caracterização dos dados</i>	27
4	Algoritmo de criação de rotas	31
4.1	<i>Objetivos do algoritmo</i>	31
4.2	<i>Variáveis de input</i>	31
4.3	<i>Restrições</i>	32
4.4	<i>Tempo de serviço</i>	33
4.4.1	<i>Análise de regressão linear múltipla</i>	33
4.5	<i>Proposta de algoritmo</i>	37
4.5.1	<i>Algoritmo de criação de rotas</i>	38
4.5.2	<i>Funções desenvolvidas</i>	40
4.5.3	<i>Pré-processamento dos dados de input</i>	43

4.5.4	Análise de Resultados	44
4.6	<i>Comparação de resultados</i>	45
4.6.1	Comparação das rotas de entrega	45
4.6.2	Análise de custos	47
5	Conclusão	49
5.1	<i>Conclusões sobre o trabalho desenvolvido</i>	49
5.2	<i>Limitações</i>	50
5.3	<i>Proposta para trabalhos futuros</i>	50
	Referências bibliográficas	53
	Anexo A	59
	Anexo B	63
	Anexo C	67
	Anexo D	69

ÍNDICE DE FIGURAS

Figura 3.1 - Excerto de visualização da lista de contrato no sistema <i>iMoovit</i>	24
Figura 3.2 - Visualização de uma rota no menu de criação de rotas	25
Figura 3.3 - Relação entre o peso e o tempo de entrega	27
Figura 3.4 - Relação entre o volume e o tempo de entrega.....	28
Figura 3.5 - Relação entre o número de caixas e o tempo de entrega	28
Figura 4.1 - Diagrama de <i>inputs</i> e <i>outputs</i>	32
Figura 4.2 - Estrutura genérica do algoritmo	38
Figura 4.3 - Exemplo da exportação de um contrato.....	43
Figura 4.4 - Exemplo de uma rota criada pelo algoritmo.....	45
Figura C.1 - Inicialização do algoritmo desenvolvido	67
Figura C.2 - Funções desenvolvidas.....	68
Figura C.3 - Últimas linhas do código desenvolvido.....	68

ÍNDICE DE TABELAS

Tabela 4.1 - Estatísticas da regressão inicial	33
Tabela 4.2 - ANOVA inicial.....	34
Tabela 4.3 - Coeficientes da regressão inicial	34
Tabela 4.4 - Estatísticas da regressão do modelo final	36
Tabela 4.5 - ANOVA do modelo final.....	37
Tabela 4.6 - Coeficientes do modelo final.....	37
Tabela 4.7 - Dados para <i>input</i> no algoritmo, resultantes da concatenação das colunas.....	44
Tabela 4.8 - Estatísticas das rotas criadas pelo algoritmo.....	46
Tabela 4.9 - Estatísticas das rotas criadas pelo gestor de tráfego	46
Tabela A.1 - Dados recolhidos, no acompanhamento das rotas, sobre os contratos entregues	59
Tabela B.1 - Resíduos do modelo de regressão	63
Tabela D.1 - Resumo das rotas criadas pelo algoritmo	69
Tabela D.2 - Resumo das rotas criadas pelo gestor de tráfego.....	70

GLOSSÁRIO

Algoritmo - Conjunto de instruções ou regras definidas passo a passo para resolver um problema/re-alizar uma tarefa. Usados em várias áreas, desde programação até otimização de processos (Dantzig & Ramser, 1959).

Análise de regressão linear múltipla - Técnica estatística que modela a relação linear entre uma variável dependente e várias variáveis independentes. É utilizada para prever valores e entender como diferentes fatores influenciam um determinado resultado (Cohen, 2013; Durbin & Watson, 1992).

Armazém - Local usado para armazenamento e movimentação de mercadorias, onde operações como receção, inspeção, organização do espaço, controle de *stock*, *picking*, e *packing* são fundamentais para a distribuição mais eficiente (Gu et al., 2007; Rouwenhorst et al., 2000).

Arrasto - No contexto de logística, refere-se ao transporte de mercadorias para um local intermediário, como um armazém, antes da entrega final.

Capacitated vehicle routing problem (CVRP) - Variante do problema de planeamento de rotas de veículos, em que se deve planear rotas, para uma frota de veículos com capacidade limitada, para visitar um conjunto de clientes, minimizando custos e respeitando restrições de capacidade (Toth & Vigo, 2002).

Cluster - Agrupamento de objetos ou pontos de dados com características similares. Em diversas áreas, como análise de dados e *machine learning*, o termo é utilizado para identificar padrões e segmentar conjuntos de dados (Tan et al., 2014).

Contrato - Conjunto de serviços a serem realizados na casa do cliente, podendo incluir entrega, montagem e/ou recolha de mercadorias.

Distância de Cook - Métrica usada na análise de regressão linear para identificar influências de observações individuais, permitindo avaliar quais os pontos de dados que têm um impacto significativo no modelo (Cook, 1977).

DFITS (Difference in fits) - Métrica utilizada para medir a influência de uma observação no valor ajustado do modelo de regressão (Belsley et al., 1980).

Google maps API - Interface de programação de aplicações fornecida pela *Google*, que permite a integração de mapas e funcionalidades de geolocalização em aplicações web ou móveis.

Janela horária - Intervalo de tempo selecionado pelo cliente durante o qual a entrega deve ser realizada.

Leverage (Alavancagem) - Métrica na análise de regressão linear que avalia a distância de uma observação em relação ao centro de massa das variáveis independentes. Pode indicar a presença de *outliers* (Chatterjee & Hadi, 2012).

Logística de armazém - Gestão das atividades relacionadas com o armazenamento e a movimentação de mercadorias dentro de um armazém, incluindo o controlo de *stock*, *picking* e *packing* (Gu et al., 2007).

Outliers - Observações que se desviam significativamente do padrão dos dados, podendo afetar a modelação e/ou a análise estatística (Rousseeuw & Leroy, 1987).

Planeamento de rotas - Processo de determinar o melhor caminho ou sequência de passos para alcançar um destino ou realizar uma série de atividades. Utilizado em logística, navegação e planeamento urbano (Dantzig & Ramser, 1959).

Problema do caixeiro viajante (TSP) - Problema de otimização que tem o objetivo de encontrar o percurso mais curto para visitar um conjunto de cidades e retornar ao ponto de partida. Este problema serve como base para a resolução de problemas de planeamento de rotas de veículos (Chvátal et al., 2010).

Problema de planeamento de rotas de veículos (VRP) - Problema de otimização que visa determinar as rotas mais eficientes para uma frota de veículos visitar um conjunto de clientes dispersos geograficamente (Dantzig & Ramser, 1959; Laporte, 2009).

Processamento de dados - Atividade de recolher, organizar e analisar dados para extrair informações relevantes.

Rotas - Percursos planeados para os veículos de entrega, que incluem a sequência de clientes a serem visitados e a ordem dessas visitas.

Serviço de entrega - Atividade de transportar mercadorias desde o armazém até ao cliente final, incluindo a descarga e posicionamento da mercadoria no local indicado pelo cliente.

Tempo de entrega - Período medido desde o momento em que a equipa de entrega estaciona o veículo na morada do cliente até quando retoma a viagem. Esse tempo varia conforme o número e tipo de mercadoria a ser entregue.

Warehouse management system (WMS) - Sistema de gestão de armazéns que melhora processos de logística, como o controlo de *stock*, *picking* e *packing* (Rouwenhorst et al., 2000).

SIGLAS

API - *Application programming interface*

CVRP - *Capacitated vehicle routing problem*

DBSCAN - *Density-based spatial clustering of applications with noise*

DFFITS - *Difference in fits*

EVRP - *Electric vehicle routing problem*

MS - *Mean square*

NP - *Nondeterministic polynomial time*

PCA - *Principal component analysis*

PV - *Prior Velho*

RFID - *Radio frequency identification*

TSP - *Problema do caixeiro viajante (traveling salesman problem)*

VRP - *Problemas de planeamento de rotas de veículos (Vehicle routing problem)*

VRPTW - *Vehicle routing problem with time windows*

VIF - *Variance inflation factor*

WMS - *Warehouse management system*

INTRODUÇÃO

Neste primeiro capítulo, é apresentada uma visão geral do contexto e da motivação para a realização deste trabalho. Inicialmente, justifica-se a escolha do tema, destacando sua relevância no âmbito da criação de rotas de entrega em empresas de logística. Em seguida, descreve-se o enquadramento do estudo, incluindo informações sobre a empresa onde o trabalho foi desenvolvido e os principais desafios enfrentados. São também definidos os objetivos da dissertação e detalhada a estrutura do documento, fornecendo um guia para a compreensão do conteúdo abordado nos capítulos subsequentes. Por fim, descrevem-se os métodos de investigação adotados, evidenciando as abordagens utilizadas na elaboração e validação do algoritmo proposto.

1.1 Justificação do tema

A ideia de realizar um trabalho sobre um algoritmo de criação de rotas surgiu após uma análise inicial aos procedimentos da empresa. Foi observado que a empresa despendia uma quantidade elevada de recursos na realização das tarefas relacionadas com a criação de rotas. Um colaborador da empresa gasta grande parte do seu tempo de trabalho diário, a definir quais os contratos que devem ser colocados numa mesma rota e qual é a rota de entrega a ser realizada por cada carrinha. Este é um processo realizado com o apoio do sistema utilizado pela empresa para todas as suas operações, o *Imoovit*.

Imediatamente foi possível perceber que este processo não é o mais indicado por ser bastante ineficiente. Os clientes são associados a uma rota tendo por base a zona de entrega correspondente à morada do cliente. As rotas são então definidas com contratos com a mesma zona de entrega, ou zonas de entrega com localizações próximas. Esta análise é feita pelo conhecimento prévio do colaborador que, pela sua experiência de trabalho, sabe prever as rotas que terão de ser feitas diariamente, de acordo com as rotas dos dias anteriores.

Espera-se então obter um algoritmo que, ao receber a informação dos contratos a realizar para o dia seguinte, consiga desenvolver um conjunto de rotas, definindo também a ordem pela qual cada

cliente deve ser visitado. A criação deste algoritmo resultaria na redução do tempo despendido, pelos gestores de tráfego da TNB, na criação de rotas, bem como, numa maior eficiência das mesmas, o que pode representar um aumento do número de contratos concluídos dentro da janela horária, o que por sua vez aumenta os níveis de satisfação dos clientes, pois eles podem não estar dispostos a aceitar receber a mercadoria fora da janela horária, o que levaria a reagendamentos.

Percebe-se então que este tema tem bastante interesse para a empresa pois permitirá definir medidas concretas para que as rotas percorram o mínimo de quilómetros possível, tornando-as mais eficientes.

1.2 Enquadramento

Este trabalho foi desenvolvido no âmbito de um estágio profissional, realizado na TNB, empresa pertencente ao Grupo Totalmédia. O estágio iniciou a 17 de março de 2022, teve a duração de nove meses, e teve o objetivo de dar apoio a nível da gestão da informação e análise de dados.

A TNB é uma empresa que realiza a entrega de última milha, o que significa que realiza o transporte de mercadoria desde um terminal de distribuição, até à morada do cliente final. A empresa realiza vários serviços diferentes como transporte, montagem de mobiliário, montagem de cozinhas e mudanças. Cada entrega de mercadoria para um cliente final implica a criação de um contrato, e cada contrato pode conter vários tipos diferentes de mercadoria.

A TNB realiza então um serviço a um cliente, que é na maior parte das vezes uma loja de venda de mobiliário, na casa de um cliente final, que é a pessoa que adquire o produto na loja. O cliente principal da empresa é a Ikea, que representa cerca de 90% da faturação, por esta razão, existe um departamento completamente dedicado à gestão dos contratos desta loja. Todos os outros clientes são geridos pela mesma equipa pois não requerem o mesmo nível de gestão.

Depois da entrega ser realizada, o cliente final dispõe de 48 horas para apresentar uma queixa de mercadoria danificada ou em falta. Quando isto acontece a TNB também realiza o serviço de recolha e/ou substituição das peças e é realizada uma análise em conjunto com a loja, para as falhas serem imputadas a quem as realizou. Por ser uma empresa que concretiza serviços para outra empresa, torna-se muito importante a satisfação do cliente final, ou seja, é necessário garantir que o cliente recebe a mercadoria correta e em bom estado. Para isto, é necessário realizar uma excelente gestão logística do armazém e ter uma boa comunicação entre os gestores de tráfego e as equipas de transporte, garantindo assim que as entregas são realizadas dentro da janela horária agendada com o cliente.

1.3 Objetivo

O objetivo desta dissertação é desenvolver um algoritmo que crie rotas para um dia de entregas, de uma forma automática, a partir de dados que caracterizam as entregas a realizar aos clientes. Os *inputs* fornecidos pelo utilizador no dia anterior ao dia da entrega são o peso, o volume, as coordenadas geográficas da morada do cliente e a janela horária escolhida no momento da compra da mercadoria, indicando também a ordem por que cada cliente deve ser visitado. O *output* do algoritmo deve ser um conjunto de rotas organizadas por ordem de visita. Para cada rota, cujo percurso se inicia no armazém de distribuição e termina na localização do último cliente a ser visitado, deve ser identificada a hora a que cada cliente foi visitado.

Neste trabalho pretende-se também comparar as rotas criadas pelo algoritmo desenvolvido com as rotas efetivamente criadas para um certo dia de entrega utilizando algumas medidas de desempenho como a distância total percorrida pelas carrinhas em cada rota.

1.4 Estrutura do documento

Este documento encontra-se dividido em cinco capítulos, com o propósito de tornar a leitura e a compreensão do trabalho desenvolvido mais acessíveis.

No primeiro capítulo é contextualizado o estudo ao apresentar de forma clara o problema que motiva este trabalho. São explorados os objetivos que se pretendem alcançar com o estudo desenvolvido, destacando a relevância do trabalho no contexto da criação de rotas. Neste capítulo é ainda apresentada a estrutura geral do documento para orientar o leitor.

De seguida no segundo capítulo é estabelecida a base teórica necessária para compreender o algoritmo de criação de rotas automáticas. São abordados conceitos, teorias, modelos e abordagens que sustentam a pesquisa. É, portanto, apresentada a pesquisa literária que foi feita, com o objetivo de delinear a estratégia a utilizar para dar resposta ao problema.

No capítulo 3 é apresentada uma caracterização detalhada do sistema em que o algoritmo é aplicado. É realizada uma exploração aprofundada do ambiente, incluindo dados de entrada, restrições, objetivos e requisitos. Pretende-se que o leitor compreenda o modo de funcionamento da empresa em questão.

O quarto capítulo é dedicado à explicação minuciosa do funcionamento do algoritmo e o seu desenvolvimento, bem como os princípios subjacentes e técnicas aplicadas à sua criação. O capítulo termina com a aplicação do algoritmo a um exemplo de pequena dimensão, para demonstrar o seu funcionamento, e com a comparação dos resultados com as rotas criadas pelo colaborador da TNB.

O quinto e último capítulo é respetivo às conclusões. São expostas as conclusões finais sobre o trabalho realizado, sendo também discutidas as limitações e possíveis direções para trabalhos futuros.

1.5 Métodos de Investigação

Neste trabalho foi aplicada a metodologia do estudo de caso, pois esta metodologia, permite realizar uma análise minuciosa, com o objetivo de compreender as suas particularidades e características. O estudo de caso é uma ferramenta muito utilizada em áreas como a psicologia, sociologia e engenharia, entre outras.

O objetivo do estudo de caso é compreender o caso selecionado em toda a sua complexidade e profundidade, identificando os fatores que o influenciam e todas as interações entre eles. Desta forma, é possível compreender como o objetivo do estudo se desenvolve e quais são as suas características essenciais.

Os resultados obtidos a partir da utilização deste método podem ser utilizados para o desenvolvimento de novas teorias, bem como para a tomada de decisões práticas que estão assim fundamentadas com o estudo realizado.

O estudo de caso é assim uma ferramenta que exige um trabalho minucioso e detalhado, que envolve uma recolha de dados extensa e uma análise profunda sobre esses mesmos dados.

Neste trabalho, o estudo de caso dividiu-se em 4 fases. Na primeira fase, foi efetuada uma análise do trabalho conduzido pelo gestor de tráfego nas suas tarefas de criação de rotas. Esta fase tem o intuito de estudar o processo e conhecer todos os seus detalhes pois, desta forma, é possível identificar padrões e tendências nas rotas que são criadas diariamente, como também áreas que necessitam de ações de melhoria ou pontos fortes do processo que possam ser replicados na criação de rotas no futuro.

A segunda fase foi composta pela recolha e tratamento dos dados necessários para o desenvolvimento do trabalho. Nesta fase foram recolhidos os tempos que uma equipa leva a realizar a entrega de cada contrato. Utilizando algumas características de cada contrato como o seu peso, volume e número de caixas a entregar foi possível obter uma fórmula para simular o tempo de entrega de um contrato, realizando uma análise de regressão linear múltipla. Nesta fase foi também importante definir todas as restrições relevantes para o problema em questão.

Na terceira fase foi desenvolvido o algoritmo de criação de rotas.

A última fase consistiu na aplicação do algoritmo aos contratos de vários dias de entrega para determinar a validade das rotas criadas. Foram também criados alguns critérios de desempenho de forma que seja possível comparar as rotas criadas pelo algoritmo com as rotas criadas efetivamente pelo gestor de tráfego.

LOGÍSTICA E ANÁLISE DE DADOS NO PLANEAMENTO DE ROTAS DE ENTREGA

De forma a criar um algoritmo para resolver o problema de criação de rotas, é essencial compreender os princípios fundamentais envolvidos, bem como os trabalhos e pesquisas pré-existentes nesse campo. Assim, neste capítulo, serão explorados os conceitos logísticos relacionados com o planeamento de rotas para entregas ao cliente.

Além disso, serão abordados temas mais amplos, como a logística de armazém no contexto de entregas ao cliente, os problemas de planeamento de rotas de veículos (VRP) e o problema do caixeiro viajante (TSP). Também serão abordadas técnicas de análise de dados, com foco na análise de regressão linear múltipla, para aprofundar a compreensão dos critérios utilizados, a modelação dos problemas, e as diferentes metodologias aplicadas na criação de rotas.

2.1 Logística de armazém no contexto de entregas a cliente

Uma gestão eficiente de um armazém tem um impacto direto no planeamento e na execução das rotas de entrega, que é uma fase crítica na distribuição de mercadorias aos clientes finais.

A logística de armazém envolve a gestão de todas as atividades relacionadas com o armazenamento e movimentação de mercadorias dentro de um armazém. Isto inclui operações como a receção e inspeção de mercadorias, organização do espaço de armazenamento, controlo de *stock* e, até, o processo de *picking* e *packing* (seleção e embalagem) para a distribuição (Gu et al., 2007).

A logística de armazém tem os objetivos principais de:

1. Maximizar a utilização do espaço disponível;
2. Otimizar a movimentação interna de produtos;
3. Reduzir os tempos de ciclo de processamento de pedidos;
4. Assegurar a precisão do *stock*;

5. Facilitar a saída eficiente das mercadorias para distribuição.

O planeamento das rotas de entrega é um processo que visa determinar o percurso mais eficiente para transportar mercadorias do armazém até aos destinos finais. Este planeamento está ligado com a eficiência das operações de armazém por várias razões.

Um controlo de *stock* preciso é essencial para garantir que os produtos corretos estejam disponíveis para o *picking* e sejam despachados na hora certa. Erros no *stock* podem levar a atrasos na preparação dos pedidos, o que, por sua vez, afeta o planeamento das rotas de entrega (Rouwenhorst et al., 2000).

O tempo necessário para selecionar e embalar os produtos no armazém afeta diretamente o horário de saída das mercadorias para entrega. Um processo de *picking* ineficiente pode resultar em atrasos, o que compromete o planeamento das rotas e pode levar ao incumprimento das janelas horárias das entregas (De Koster et al., 2007).

A logística de armazém permite o agrupamento de pedidos que têm destinos próximos ou que seguem rotas semelhantes o que facilita a consolidação de cargas, fator essencial para aumentar a eficiência das entregas, minimizando a distância total percorrida e o número de viagens necessárias (Berg & Zijm, 1999).

Armazéns que operam de forma eficiente conseguem sincronizar as suas operações com os horários de entrega exigidos pelos clientes. Essa capacidade de sincronização é vital para o planeamento das rotas, especialmente em casos de entrega *just-in-time* ou em horários específicos (Rushton et al., 2022). Um aumento da eficiência da gestão do armazém resulta, ainda, na redução dos custos operacionais, pois aumenta a precisão e rapidez com que as mercadorias são preparadas para despacho, evitando deslocações desnecessárias e garantindo uma utilização eficiente dos veículos (Bartholdi & Hackman, 2008).

Apesar dos benefícios mencionados, integrar de forma eficaz a logística de armazém com o planeamento de rotas de entrega apresenta vários desafios.

Flutuações na procura podem resultar em períodos de alta e baixa atividade no armazém, o que complica a previsão e o planeamento de rotas. Em períodos de pico, os armazéns podem enfrentar dificuldades para processar pedidos rapidamente, afetando as rotas de entrega (Carvalho et al., 2012).

Consolidar pedidos de forma eficiente requer um equilíbrio entre o tempo de espera para agrupar pedidos e a necessidade de despachar rapidamente. Falhas neste equilíbrio podem resultar na criação de rotas ineficientes ou em entregas parciais, aumentando os custos de transporte (Goetschalckx & Ashayeri, 1989).

A determinação do número ideal de zonas envolve a análise de diversos fatores que afetam o desempenho do sistema de *picking*. Mercadorias com alta rotatividade podem ser agrupadas em zonas específicas para reduzir o tempo de *picking*. A análise do perfil da procura ajuda a entender qual o tipo de mercadoria que deve ser armazenado em zonas próximas para facilitar o acesso rápido. O *layout*

do armazém deve ser projetado para facilitar a movimentação eficiente dentro de cada zona, evitando rotas de *picking* que se cruzam ou áreas de congestionamento. Com o objetivo de minimizar o tempo de *picking* e o tempo de movimentação em mente, foram desenvolvidas metodologias baseadas em modelos matemáticos e técnicas de simulação. Através destes métodos é possível identificar a configuração que proporciona a melhor eficiência operacional. Estudos mostram que a concepção eficiente do sistema de *picking*, incluindo a escolha adequada do número de zonas, pode reduzir consideravelmente o tempo de ciclo e melhorar a eficiência na consolidação de pedidos, impactando positivamente a logística de distribuição (De Koster et al., 2012).

A necessidade de monitorizar o *stock* em tempo real para garantir que as mercadorias estejam disponíveis quando necessário é muito importante para evitar a ocorrência de erros na preparação dos pedidos (Gunasekaran & Ngai, 2004).

A adoção de tecnologias avançadas, como sistemas de gestão de armazém (WMS), RFID e automação, pode melhorar a eficiência do armazém e, por extensão, o planeamento das rotas (Rouwenhorst et al., 2000). No entanto, a implementação e a integração dessas tecnologias apresentam desafios em termos de custo, complexidade e necessidade de formação dos recursos humanos (Faber et al., 2013).

Garantir que os funcionários do armazém estão bem treinados nas melhores práticas de operação pode reduzir erros e aumentar a eficiência, impactando positivamente a rapidez e a precisão das rotas de entrega (Richards, 2017).

A implementação de técnicas avançadas de previsão de procura ajuda a alinhar a capacidade do armazém com os requisitos de entrega, minimizando o impacto das flutuações no planeamento de rotas (Christopher, 2016).

A automatização parcial ou total do *picking* pode aumentar consideravelmente a velocidade e a precisão deste processo, reduzindo o tempo de ciclo e permitindo um planeamento de rotas mais ágil (De Koster et al., 2007).

2.2 Problemas de planeamento de rotas de veículos e do Caixeiro Viajante

Os problemas de planeamento de rotas de veículos (*Vehicle routing problems* - VRP) e o problema do caixeiro viajante (*Traveling salesman problem* - TSP) são problemas muito estudados na área de otimização combinatória e investigação operacional. Ambos têm vasta aplicabilidade prática em áreas como logística, transporte, planeamento de rotas e distribuição, sendo essenciais para a eficiência operacional de muitas empresas.

O VRP, introduzido por Dantzig e Ramser em 1959, envolve a determinação de rotas para um conjunto de veículos que devem visitar uma série de clientes dispersos geograficamente. O objetivo é

reduzir os custos associados ao transporte, o que pode incluir distâncias percorridas, tempo de viagem e outros fatores (Dantzig & Ramser, 1959).

O TSP pode ser visto como um caso particular do VRP onde, dado um conjunto de pontos geográficos a visitar e as distâncias entre cada ponto, o objetivo é encontrar o percurso mais curto que visite cada cidade uma vez e retorne ao ponto de partida. Este problema foi formulado originalmente por Dantzig, Fulkerson e Johnson em 1954 (Chvátal et al., 2010), que propuseram métodos iniciais para a sua solução. O TSP é regularmente classificado como NP-difícil, ou seja, não há conhecimento da existência de nenhum algoritmo conhecido que possa resolvê-lo em tempo polinomial para todas as instâncias possíveis (Applegate, 2006).

Existem várias variantes do VRP que refletem diferentes restrições e aplicações práticas:

- VRP com Janelas de Tempo (VRPTW): Introduce restrições de tempo para a visita a cada cliente. Bräysy & Gendreau (2005) identificaram que, devido à complexidade computacional do VRPTW, as abordagens heurísticas e meta-heurísticas (como algoritmos de busca tabu e algoritmos genéticos) mostraram-se mais adequadas para a solução de problemas em larga escala, oferecendo um equilíbrio entre qualidade da solução e tempo de processamento. Estes autores também realçaram que as restrições temporais são críticas em muitas aplicações práticas, como na logística de entregas urbanas e na distribuição de produtos frescos, onde atrasos podem comprometer seriamente o serviço ou a qualidade do produto.
- VRP Capacitado (CVRP): No caso do CVRP, os veículos têm uma capacidade limitada, o que exige que a solução do problema leve em consideração tanto a rota mais curta quanto a carga máxima que cada veículo pode transportar. Toth & Vigo (2002) discutiram diversas formulações matemáticas do CVRP e apresentaram algoritmos exatos, como o *Branch-and-Cut*, que se mostraram eficazes para resolver instâncias de tamanho médio do problema. Foi também investigada a eficiência de diferentes métodos de relaxamento, como os relaxamentos linear e de Lagrange, e avaliado como esses métodos podem ser combinados para melhorar a resolução de problemas de grande escala. Um dos principais avanços apresentados por Toth & Vigo (2002) foi o desenvolvimento de estratégias híbridas que combinam técnicas exatas com heurísticas, permitindo resolver problemas mais complexos.
- VRP Dinâmico e Estocástico: Envolve incertezas nas distâncias ou na disponibilidade dos clientes. Bertsimas & Simchi-Levi (1996) desenvolveram modelos matemáticos que incorporam incertezas diretamente na formulação do problema, utilizando técnicas como a programação estocástica e a programação robusta. No estudo desenvolvido pelos autores ficou demonstrado que estas abordagens são fundamentais para aumentar a robustez das soluções em cenários onde a informação é incompleta ou sujeita a

mudanças, como na logística em tempo real e no transporte de emergência. A pesquisa mostrou que ao considerar incertezas, é possível obter soluções que, embora não sejam as mais eficientes em condições ideais, apresentam um desempenho significativamente melhor em situações adversas, oferecendo um equilíbrio entre custo e confiabilidade.

Para resolver o VRP foram desenvolvidos métodos como algoritmos exatos, heurísticas e meta-heurísticas. Métodos exatos, como a programação linear inteira e o *branch and bound*, são utilizados para instâncias menores devido à sua complexidade computacional (Laporte, 2009). Heurísticas, como o algoritmo de Clarke e Wright, fornecem boas soluções em tempo razoável (Golden et al., 2008). Meta-heurísticas, como algoritmos genéticos e busca tabu, são frequentemente utilizadas para instâncias maiores e mais complexas do VRP (Gendreau et al., 1994).

Os avanços em métodos exatos e heurísticos têm impulsionado a capacidade de resolver grandes instâncias de TSP e VRP. Applegate (2006) forneceu um estudo detalhado sobre o TSP, destacando avanços em técnicas de *branch and bound* e métodos de corte. Baldacci et al. (2012) reviram algoritmos exatos para o VRP com restrições de capacidade e janelas de tempo, mostrando como esses avanços têm aumentado a eficiência das soluções.

Adicionalmente, a utilização de algoritmos híbridos e a combinação de diferentes abordagens têm-se revelado promissoras. Por exemplo, a integração de algoritmos genéticos com busca tabu ou outras heurísticas locais permitem explorar o espaço de soluções de forma mais eficiente, combinando a exploração global dos genéticos com a intensificação local da busca tabu (Blum & Roli, 2003).

A solução destes problemas envolve frequentemente técnicas avançadas de programação que podem ser implementadas usando diversas linguagens e ferramentas. Entre essas ferramentas, o *Python* tem-se destacado como uma das linguagens de programação preferidas para a solução de problemas de otimização, devido às suas características únicas.

A sintaxe clara e concisa do *Python* facilita a implementação e a experimentação com diferentes algoritmos o que é muito importante para a prototipagem rápida e o desenvolvimento de soluções eficientes. O *Python* oferece uma vasta gama de bibliotecas dedicadas à otimização e à matemática como a PuLP, uma biblioteca para programação linear e inteira (Mitchell et al., 2011), *Google OR-Tools*, uma biblioteca desenvolvida pela *Google* para resolver problemas de planejamento de rotas e otimização e *SciPy* e *NumPy* que oferecem suporte para operações matemáticas e algoritmos necessários para implementar heurísticas e meta-heurísticas (Oliphant, 2007; Harris et al., 2020). Esta linguagem de programação tem ainda uma comunidade ativa que contribui regularmente com novos pacotes, documentação e suporte, o que facilita a implementação de algoritmos complexos e a solução de problemas específicos. O *Python* é ainda facilmente integrável com outras ferramentas e linguagens, como C ou C++ para otimização de desempenho, e com ambientes de visualização, como *Matplotlib* e *Seaborn* para realizar a análise de resultados (Hunter, 2007).

As pesquisas mais recentes sobre o TSP e o VRP têm se concentrado em melhorar a eficiência dos algoritmos existentes e em integrar novas técnicas de *machine learning* e inteligência artificial. Avanços como o uso de redes neurais, *reinforcement learning*, e algoritmos híbridos estão a abrir as fronteiras do que é possível na solução destes problemas complexos. A evolução contínua destas técnicas promete aumentar a qualidade e a eficiência das soluções para o TSP e o VRP, impactando positivamente várias aplicações práticas em logística e planejamento de rotas.

A investigação contínua não só aprofunda o entendimento teórico, como também impulsiona avanços tecnológicos em diversas indústrias, melhorando a eficiência das operações logísticas. A aplicação de novos paradigmas de computação, como a computação em nuvem e a inteligência artificial, está a abrir novas fronteiras para a solução destes problemas, permitindo tratar instâncias cada vez maiores e mais complexas em tempos reduzidos (Pillac et al., 2013).

Bengio et al. (2021) realizaram uma revisão abrangente sobre como técnicas de *machine learning* estão a ser aplicadas à otimização combinatória. Entre as descobertas, destacam-se as aplicações de redes neurais para melhorar heurísticas e meta-heurísticas para o TSP. Os autores discutem como as técnicas de *deep learning* têm o potencial de aumentar a eficácia das abordagens tradicionais, como algoritmos genéticos e busca local, ao aprender padrões a partir de soluções anteriores e adaptar os algoritmos em tempo real para aumentar a qualidade das soluções e reduzir o tempo de computação.

Yu et al. (2019) introduziram uma abordagem baseada em *deep reinforcement learning* para problemas de planejamento de rotas online de veículos. Utilizando um modelo de otimização combinatória neural, a abordagem foi capaz de se adaptar rapidamente a mudanças em tempo real, como condições de tráfego e cancelamentos de pedidos. Os resultados mostraram que a abordagem proposta superou algoritmos heurísticos tradicionais em termos de eficiência de rota e tempo de resposta. O modelo neural treinado demonstrou uma melhoria significativa na otimização de rotas em cenários de entrega urbana, especialmente em situações onde a flexibilidade e a adaptação rápida são muito importantes.

Tan & Yeh (2021) ofereceram uma análise detalhada da literatura recente sobre o VRP, classificando as soluções em três grandes categorias: métodos exatos, heurísticos e meta-heurísticos. Os autores destacaram que, nos últimos anos, tem existido uma preferência pela utilização de métodos baseados em inteligência artificial e *machine learning* para resolver variantes do VRP. Os autores também apontaram que a integração de técnicas como algoritmos genéticos, busca tabu, e *reinforcement learning*, trouxe melhorias na qualidade das soluções e na escalabilidade dos algoritmos. Esta revisão forneceu uma visão abrangente das tendências atuais, destacando a crescente importância de algoritmos adaptativos e a aplicação de modelos que incorporam complexidades do mundo real, como incertezas no tráfego e procuras dinâmicas dos clientes.

Lin et al. (2022) focaram-se no problema de planejamento de rotas de veículos elétricos (EVRP) com janelas de tempo, uma variante complexa do VRP devido às limitações de autonomia e tempos de recarga dos veículos. Os autores aplicaram *deep reinforcement learning* para otimizar a rota dos

veículos, considerando tanto as janelas de tempo quanto a necessidade de recarregar as baterias. O modelo foi capaz de aprender políticas de planejamento de rotas que equilibram a minimização da distância percorrida e o cumprimento das restrições de tempo. Os resultados mostraram que a abordagem proposta reduziu consideravelmente os custos operacionais e melhorou a eficiência do uso da frota de veículos elétricos, fornecendo soluções robustas em cenários urbanos complexos.

O TSP e o VRP são problemas fundamentais na otimização combinatória, com aplicações práticas significativas na logística, planejamento de rotas e distribuição. A vasta gama de métodos e abordagens desenvolvidos para resolver esses problemas reflete a sua importância e a diversidade dos desafios que apresentam. O progresso contínuo nesta área é muito importante para o desenvolvimento de soluções mais eficientes e adaptáveis, respondendo às crescentes exigências dos sistemas logísticos modernos.

2.3 Técnicas de análise de dados

Algumas técnicas de análise de dados têm sido amplamente utilizadas em diversas disciplinas, desde as ciências sociais até a engenharia, e revelam-se especialmente úteis em cenários onde se procura prever o comportamento de uma variável de interesse, como o tempo de entrega de mercadorias, a partir de vários fatores. Nestas técnicas, destaca-se a análise de regressão linear múltipla.

A análise de regressão linear múltipla, é uma técnica estatística utilizada para modelar a relação entre uma variável dependente e múltiplas variáveis independentes.

A regressão linear múltipla estende a regressão linear simples onde, em vez de uma única variável, existe um conjunto de variáveis independentes ($X_1, X_2, \dots, X_p$) que explicam a variabilidade na variável dependente (Y). O modelo de regressão linear múltipla é apresentado na equação (1):

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_p X_p + \epsilon \quad (1)$$

Onde:

- Y é a variável dependente que se deseja prever;
- θ_0 é o intercepto do modelo e representa o valor esperado para Y quando as variáveis independentes são iguais a zero;
- $\theta_1, \theta_2, \dots, \theta_p$ são os coeficientes de regressão que quantificam o impacto de cada variável independente $X_1, X_2, \dots, X_p$ sobre Y ;
- ϵ é o termo de erro, que engloba a variabilidade em Y não explicada pelas variáveis independentes.

Para que os resultados da regressão linear múltipla sejam válidos, é necessário que alguns pressupostos básicos sejam atendidos:

1. Linearidade: Este pressuposto implica que uma alteração numa variável independente deve resultar numa alteração proporcional na variável dependente. A linearidade pode ser verificada através da análise gráfica dos erros e utilizando testes estatísticos específicos (Cohen, 2013).
2. Independência dos erros: Os erros (ϵ) devem ser independentes, ou seja, não deve haver correlação entre os erros dos diferentes pontos de dados. Este aspeto é particularmente importante na presença de séries temporais ou dados sequenciais. O teste de *Durbin-Watson* é regularmente utilizado para testar este pressuposto (Durbin & Watson, 1992).
3. Homoscedasticidade: A variância dos erros deve ser constante ao longo dos valores previstos. Quando esta suposição não é atendida, ocorre heteroscedasticidade, o que pode tornar as estimativas dos coeficientes menos eficientes e os intervalos de confiança imprecisos (Breusch & Pagan, 1979). A homoscedasticidade pode ser verificada através da análise gráficos de resíduos.
4. Normalidade dos erros: Os erros devem ser distribuídos normalmente para que os testes de significância dos coeficientes de regressão sejam válidos. A normalidade dos erros pode ser avaliada através do teste de *Shapiro-Wilk* ou através da inspeção visual de um gráfico *Q-Q plot* (Shapiro & Wilk, 1965).
5. Ausência de multicolinearidade: As variáveis independentes não devem ser muito correlacionadas entre si. A multicolinearidade inflaciona as variâncias dos coeficientes de regressão, tornando-os instáveis e dificultando a interpretação do modelo. A multicolinearidade pode ser diagnosticada pelo Fator de Inflação da Variância (VIF) e pelos valores de tolerância (Neter et al., 1983).

Os coeficientes, $\theta_1, \theta_2, \dots, \theta_p$, representam a magnitude e a direção da influência de cada variável independente sobre a variável dependente. A significância de cada coeficiente pode ser testada utilizando o *P-value*, que indica a probabilidade de que o coeficiente seja diferente de zero. Um *P-value* inferior a um nível de significância escolhido (geralmente 0,05) sugere que o coeficiente é estatisticamente significativo (Mundfrom et al., 2006).

A tabela ANOVA (Análise de Variância) é uma ferramenta fundamental na regressão linear múltipla, pois permite avaliar a qualidade do modelo. A tabela ANOVA divide a variabilidade total em variabilidade explicada pelo modelo (variabilidade entre grupos) e variabilidade não explicada (variabilidade dentro dos grupos). A significância do modelo como um todo pode ser avaliada pelo *F-statistic*, comparando a variabilidade explicada com a não explicada (Montgomery et al., 2007).

A análise de resíduos é fundamental para validar os pressupostos do modelo. Os resíduos são calculados através das diferenças entre o valor observado e o valor previsto pelo modelo. Para verificar a linearidade e a homoscedasticidade são frequentemente utilizados gráficos de resíduos *versus* valores ajustados (Chatterjee & Hadi, 2012).

Além disso, os gráficos dos resíduos padronizados são também úteis para identificar *outliers*, que são observações com resíduos muito grandes em magnitude. A remoção de *outliers* pode melhorar o ajuste do modelo, mas deve ser feita com precaução, de forma a não enviesar os dados do modelo (Rousseeuw & Leroy, 1987).

A análise de resíduos é uma técnica regularmente utilizada para identificar *outliers*, uma vez que estes tendem a produzir valores absolutos de resíduos elevados em comparação com as outras observações. Existem vários métodos para identificar *outliers* com base nos resíduos, nomeadamente:

1. Resíduos padronizados: São calculados dividindo-se os resíduos pelo seu desvio-padrão estimado. Como regra geral, resíduos padronizados superiores a 2 ou 3, em valor absoluto, são considerados indicativos de *outliers* (Montgomery et al., 2007). No entanto, em grandes conjuntos de dados, é comum encontrar alguns resíduos padronizados além deste limite, por isso, o contexto e a distribuição geral dos resíduos também devem ser considerados.
2. *Leverage* (Alavancagem): A alavancagem mede a distância de uma observação em relação ao centro de massa das variáveis independentes. Centro de massa refere-se ao ponto médio de todos os dados das variáveis independentes no espaço multidimensional. Essencialmente, representa as médias de todas as variáveis independentes e serve como uma referência para medir o quão distante cada observação está desse ponto central. Observações com alta alavancagem estão longe desse centro de massa, indicando que são diferentes da maioria das observações e, por essa razão, têm o potencial de ser *outliers*. Um valor de alavancagem superior a $2(p+1)/x$, onde p representa a quantidade de variáveis independentes e x é o número total de observações, pode indicar a presença de *outliers* (Chatterjee & Hadi, 2012).
3. Resíduos *standard*: Estes são semelhantes aos resíduos padronizados, mas ajustados para refletir a importância de cada observação no cálculo do modelo. Para calcular os resíduos *standard*, primeiro, é calculado o resíduo da observação i , de seguida, este valor é ajustado dividindo-o pelo erro padrão dos resíduos que exclui a influência da observação i . A fórmula de cálculo dos resíduos *standard* é apresentada na equação (2).

$$t_i = \frac{e_i}{\hat{\sigma}(i)\sqrt{1-h_{ii}}} \quad (2)$$

onde $\hat{\sigma}(i)$ é a estimativa do erro padrão dos resíduos sem a i -ésima observação, e h_{ii} é a alavancagem da observação i , que indica a sua influência no ajuste do modelo. Os resíduos *standard* são uma ferramenta mais precisa para detetar *outliers*. Resíduos *standard* com

valores absolutos superiores a 3 são frequentemente considerados potenciais *outliers* (Neter et al., 1983).

4. Distância de Cook: A Distância de Cook é um método que combina informações sobre a magnitude dos resíduos e a influência de cada observação no ajuste do modelo. Observações com uma Distância de Cook superior a 1 podem ser consideradas influentes e potencialmente *outliers* (Cook, 1977). Este critério é particularmente útil para identificar observações que não só têm valores de resíduos grandes, mas que também alteram significativamente os coeficientes do modelo.
5. Resíduos de pressão DFFITS (*Difference in fits*): O DFFITS mede também a influência de cada observação no valor ajustado do modelo. A fórmula de cálculo dos DFFITS é apresentada na equação (3).

$$DFFITS_i = \frac{e_i}{\hat{\sigma}(i)\sqrt{h_{ii}}} \quad (3)$$

onde $\hat{\sigma}(i)$ é a estimativa do erro padrão dos resíduos sem a i -ésima observação, e h_{ii} é a alavancagem da observação i . Valores absolutos de DFFITS superiores a $2\sqrt{p/n}$ são considerados indicativos de observações que podem ser *outliers* (Belsley et al., 1980).

A remoção de *outliers* é uma prática importante no desenvolvimento de modelos com base na análise de regressão, pois estes podem distorcer as estimativas realizadas pelo modelo para os coeficientes de regressão e afetar a validade dos seus pressupostos. No entanto, a remoção de *outliers* deve ser justificada com base em análises robustas e, idealmente, confirmada através de replicações do modelo com e sem os *outliers* para verificar a estabilidade dos resultados (Barnett & Lewis, 1994).

A multicolinearidade é um problema encontrado regularmente em modelos de regressão múltipla. A correlação entre variáveis independentes pode dificultar a estimação precisa dos coeficientes de regressão, tornando os resultados do modelo menos. Em essência, a multicolinearidade cria redundância na informação fornecida pelas variáveis independentes e faz com que seja difícil isolar o efeito individual de cada variável sobre a variável dependente, já que as variáveis estão correlacionadas. O diagnóstico e a correção da multicolinearidade são muito importantes para garantir que os resultados do modelo são interpretáveis e que os coeficientes estimados são adequados ao problema.

Uma métrica regularmente utilizada para detetar a presença de multicolinearidade é o Fator de Inflação da Variância (VIF). Valores de VIF superiores a 10 indicam problemas severos de multicolinearidade, sugerindo a necessidade de remover ou combinar variáveis independentes (O'Brien, 2007).

Geralmente, combinar variáveis independentes para lidar com a multicolinearidade, envolve técnicas para reduzir a redundância entre variáveis altamente correlacionadas. Isto pode ser

concretizado, criando variáveis compostas ou índices que sintetizam a informação de várias variáveis em uma única métrica, como combinar "altura" e "peso" num Índice de Massa Corporal.

A análise de regressão linear múltipla é uma técnica estatística versátil que, quando corretamente aplicada, pode fornecer informações valiosas sobre as relações entre múltiplas variáveis. A sua aplicação no contexto da previsão de tempo de entrega, por exemplo, permite identificar os fatores mais críticos que afetam a eficiência logística e, consequentemente, melhorar a performance operacional. Contudo, a eficácia deste método depende da verificação rigorosa dos pressupostos subjacentes e da interpretação criteriosa dos resultados obtidos.

Outra técnica de análise de dados utilizada no contexto de previsão de uma variável de interesse é a análise de *clusters*. A análise de *clusters* é uma técnica de aprendizagem não supervisionada usada para agrupar dados em subconjuntos (*clusters*) de forma que as observações dentro de cada *cluster* sejam mais semelhantes entre si do que com as observações de outros *clusters*. Esta similaridade é baseada na distância Euclidiana entre os dados. A análise de *clusters* é amplamente aplicada em várias áreas, como segmentação de mercado, reconhecimento de padrões, biologia e, no contexto da logística e transporte, para agrupar clientes ou rotas de entrega de forma (Tan et al., 2014).

Existem diversas abordagens para realizar a análise de *clusters*, sendo as mais comuns o *K-Means*, a análise hierárquica e o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). O *K-Means* é um dos métodos mais utilizados, que divide os dados em *K clusters*, onde *K* é um número pré-definido. O algoritmo procura minimizar a variância dentro dos *clusters*, atribuindo cada ponto ao *cluster* com o centro de massa mais próximo. Este método é simples e eficiente para grandes conjuntos de dados, mas requer a indicação prévia do número de *clusters*, o que pode ser uma limitação (Macqueen, 1967). A análise hierárquica cria uma hierarquia de *clusters*, começando com cada observação no seu próprio *cluster* e, em seguida, unindo iterativamente os *clusters* mais semelhantes até que se forme um único *cluster* geral ou até atingir um determinado número de *clusters*. Esta técnica pode ser representada num dendrograma, o que ajuda a visualizar as relações de similaridade entre as observações. Ao contrário do *K-Means*, não é necessário definir o número de clusters antes da sua utilização (Ward, 1963). Já o DBSCAN agrupa pontos que estão em regiões densas e identifica *outliers* como pontos que estão em regiões esparsas. É especialmente útil para identificar *clusters* de formas arbitrárias e para dados com ruído ou *outliers* (Ester et al., 1996).

Na análise de *clusters*, os resíduos são valores que ajudam a avaliar o grau de ajuste do modelo aos dados. Para métodos como o *K-Means*, os resíduos são calculados pela distância entre cada ponto e o centro de massa do *cluster* ao qual pertence. Pontos com resíduos altos indicam que estão longe do centro de massa e podem ser considerados candidatos a *outliers*. A presença de outliers pode distorcer a formação dos *clusters*, especialmente em métodos como o *K-Means*, que dependem da média. Uma abordagem comum é detetar e remover os *outliers* antes da análise de *clusters*. Isto pode ser feito usando medidas de resíduos, como a distância ao centro de massa, ou aplicando um método

robusto como o DBSCAN, que pode identificar e ignorar automaticamente *outliers* como pontos de ruído (Kriegel et al., 2011).

Na análise de *clusters*, a seleção e normalização das variáveis são críticas, pois as variáveis determinam a métrica de similaridade usada para formar os *clusters*. A análise da significância das variáveis, ajuda a identificar quais são as variáveis mais importantes para a formação dos *clusters*. Métodos como a análise de componentes principais (PCA) podem ser utilizados para reduzir a dimensionalidade dos dados, selecionando as variáveis que mais contribuem para a variação nos dados e, consequentemente, para a formação de *clusters* (Joliffe, 2002).

Para avaliar a qualidade do modelo desenvolvido, várias métricas podem ser utilizadas, como o coeficiente de silhueta, que mede a similaridade de uma observação com o seu próprio *cluster* em relação a outros *clusters*, indicando quão bem a observação foi classificada (Rousseeuw, 1987). O Índice de Dunn avalia a compactação e a separação dos *clusters*, sendo que valores mais altos indicam melhor qualidade (Dunn, 1974). Além disso, métodos de validação cruzada podem ser aplicados para avaliar a estabilidade e a robustez dos *clusters* quando o modelo é aplicado a diferentes subconjuntos dos dados (Tibshirani et al., 2001).

Devido à sua capacidade de lidar com dados complexos, interpretação fácil dos resultados e fornecer previsões precisas as árvores de decisão e as florestas aleatórias são também muito úteis no âmbito da previsão de variáveis, especialmente em situações onde existe uma relação complexa e não linear entre as variáveis (Breiman et al., 2017).

As Árvores de Decisão são estruturas hierárquicas que dividem o conjunto de dados em subconjuntos mais pequenos, baseando-se em algumas características dos dados. Em cada nó da árvore, é selecionada a variável que melhor separa os dados em diferentes classes ou que melhor prevê o valor da variável dependente. Esta divisão é realizada até que cada nó terminal contenha uma observação predominante de uma classe específica ou um valor previsto (Quinlan, 1986).

Os critérios para dividir os nós podem variar, sendo os mais comuns a entropia e o índice Gini para problemas de classificação, e o erro quadrático médio para problemas de regressão. A entropia mede a impureza ou desordem de um conjunto de dados, e a divisão é feita para reduzir a entropia. Já o Índice Gini avalia a frequência de uma determinada classe no nó, e a árvore procura minimizar este índice a cada divisão (Breiman et al., 2017).

Uma das vantagens das árvores de decisão é a sua interpretabilidade. Este método permite visualizar facilmente o processo de tomada de decisão, tornando-o valioso para aplicações onde a explicabilidade do modelo é importante. No entanto, as Árvores de Decisão têm algumas limitações, como a tendência para o *overfitting*, especialmente quando a árvore é muito profunda e se ajusta demasiado aos dados de treino (Hastie et al., 2009).

Para superar as limitações das árvores de decisão, como a alta variância e o *overfitting*, foram desenvolvidas as florestas aleatórias. Este método é um tipo de *ensemble learning* que combina

múltiplas árvores de decisão para melhorar a precisão das previsões e a robustez do modelo (Breiman, 2001).

As florestas aleatórias constroem um conjunto de árvores de decisão a partir de diferentes subconjuntos aleatórios dos dados, com substituição (*bootstrap sampling*). Cada árvore na floresta é treinada com um subconjunto diferente dos dados e, em cada nó da árvore, é considerada apenas uma amostra aleatória de características para a divisão. Esta aleatoriedade ajuda a reduzir a correlação entre as árvores, resultando num modelo mais robusto (Breiman, 2001).

Para fazer previsões, as florestas aleatórias utilizam um processo de votação ou média das previsões de todas as árvores na floresta. Este método reduz a variância do modelo, levando a previsões mais estáveis e precisas. Além disso, as florestas aleatórias fornecem uma medida de importância das características, indicando quais variáveis são mais relevantes para a previsão (Liaw & Wiener, 2002).

Na análise de resíduos em árvores de decisão e florestas aleatórias, os resíduos representam a diferença entre o valor observado e o valor previsto pelo modelo. Analisar os resíduos ajuda a identificar padrões que podem indicar problemas no modelo, como tendência ou variabilidade não explicada. Como as florestas aleatórias são menos suscetíveis a *overfitting*, os *outliers* têm um efeito reduzido em comparação com árvores de decisão isoladas. No entanto, é essencial detetar e analisar *outliers*, pois podem indicar a necessidade de transformações nos dados (Geurts et al., 2006).

As árvores de decisão e florestas aleatórias têm uma vasta gama de aplicações, desde a previsão de risco financeiro e deteção de fraudes até à análise de dados médicos e previsão de comportamentos do cliente. No contexto da logística e do transporte, podem ser utilizadas para prever o tempo de entrega, otimizar rotas e identificar os principais fatores que afetam a eficiência das operações.

A partir da revisão dos conceitos e técnicas apresentados neste capítulo, conclui-se que a eficiência no planeamento de rotas de entrega depende diretamente da integração entre a logística de armazém e os métodos de otimização. A análise de problemas de planeamento de rotas, como o VRP e o TSP, demonstrou que soluções tradicionais, embora eficazes, necessitam de ser adaptadas a realidades operacionais, como restrições de capacidade e janelas temporais. Técnicas de análise de dados, como a análise de regressão linear múltipla, revelam-se fundamentais para a estimação de tempos de serviço. Ademais, a aplicação de heurísticas e meta-heurísticas foi identificada como essencial para lidar com a complexidade desses problemas em grande escala.

Em suma, o estudo destes temas evidenciou que o uso de algoritmos e modelos matemáticos avançados é muito importante para o aumento da eficiência das operações logísticas, reduzindo os custos das empresas de transporte.

CARACTERIZAÇÃO DO ESTUDO DE CASO

Neste capítulo é apresentado o caso para qual o algoritmo é desenvolvido e aplicado. É feita uma contextualização da realidade da empresa, bem como as suas necessidades de acordo com a sua operação. É também descrito todo o processo levado a cabo, atualmente, pelo gestor de tráfego para a criação das rotas.

3.1 Empresa transportadora

Um contrato pode ser definido por um conjunto de serviços a realizar na casa do cliente. Os principais serviços realizados pela TNB são os de entrega, montagem e recolha sendo que estes serviços podem também ter algumas variantes dependendo do tipo de mercadoria neles contidos. Existem também outros tipos de serviços que não serão aprofundados por não necessitarem de entrar numa rota para serem efetuados, ou por terem uma amostra residual. Um contrato pode conter serviços de Entrega e Recolha de mercadoria diferente e o mesmo pode ser dito para os serviços de montagem. O foco deste trabalho são os contratos que contenham serviços de Entrega pois estes compõem cerca de 90% dos contratos realizados. A TNB realiza cerca de 20 mil contratos por mês sendo este valor, habitualmente, mais elevado nos meses de Verão.

As entregas são realizadas maioritariamente nas zonas da área metropolitana de Lisboa e do Algarve, mas a empresa assegura também a entrega de mercadoria para outros pontos do país. Para realizar as entregas para zonas como Leiria, Castelo Branco, Évora, Porto e ilhas a empresa utiliza armazéns externos para armazenar a mercadoria que terá de ser entregue. Assim é apenas necessário enviar camiões periodicamente para entregar a mercadoria nesses armazéns, e a gestão das entregas dos contratos aos clientes fica a cargo de outras empresas subcontratadas pela TNB para essas áreas. Este tipo de entrega realizado pelo camião é designado por "arrasto".

Para garantir que as rotas cumprem as janelas horárias de todos os contratos é necessário estimar o tempo que uma equipa demora a realizar a entrega, ou seja, o tempo que passa desde que a equipa estaciona a carrinha na morada do cliente, até que retomam a marcha. Este tempo terá o nome de Tempo de Entrega. O valor desta medida será bastante diferente em cada entrega pois o número de artigos a entregar pode ser apenas um ou mais de uma centena. Também mercadoria mais pesada ou de maior volume terá um maior tempo de entrega pois as equipas têm maior dificuldade em transportá-la até casa do cliente. Por outro lado, também será importante ter em consideração as condições do prédio onde está a ser realizada a entrega, tais como o piso do cliente e se é possível utilizar elevador ou não. Para isto serão acompanhadas várias rotas e serão medidos os tempos despendidos em cada entrega.

Tendo em conta o objetivo final deste trabalho foi necessário ter uma extensa compreensão do processo de criação de rotas levado a cabo pela empresa. Foi muito importante perceber como é realizado este processo e assimilar todos os fatores que têm relevância nas decisões tomadas pelos trabalhadores. A TNB é uma empresa que realiza a entrega de mercadoria e outros serviços em casa do cliente, por isso, foi considerado muito importante realizar o acompanhamento de uma equipa, no desenvolvimento das suas atividades. A realização destas duas fases permite a identificação de várias situações que podem influenciar a criação das rotas, pois tanto os gestores de tráfego como as equipas de entrega desempenham papéis importante neste processo e podem apresentar algumas dificuldades e oportunidades de melhoria no processo.

A mercadoria transportada pela TNB pode variar muito nas suas características, a entrega de artigos com maiores dimensões como um colchão ou um roupeiro representa condições de entrega diferentes de artigos como cadeiras e mesas que vêm divididos em peças de menores dimensões e são mais fáceis de transportar.

Depois de os clientes adquirirem os serviços de entrega de mercadoria nas lojas, estas criam os contratos a serem realizados no sistema *iMoovit*, para que a TNB saiba qual a mercadoria a receber nos seus armazéns e comece a criar as rotas a serem realizadas. Ao adquirir a mercadoria, o cliente escolhe o dia e a janela horária na qual pretende receber o contrato na sua casa. Existem várias janelas horárias disponíveis para os clientes escolherem para as suas entregas, mas a maior parte dos contratos podem ser divididos em dois grupos, manhã (contratos a serem realizados antes das 13h) e tarde (contratos a serem realizados depois das 14h). Existe também a janela das 09:00-18:00 que é considerada como de dia inteiro e contratos com esta janela horária tanto podem ser agrupados com contratos da manhã como da tarde.

3.2 Caracterização das mercadorias a transportar

Para realizar os serviços de entrega na casa do cliente é necessário ter uma boa compreensão do tipo de mercadoria a transportar. Como existe uma loja que representa cerca de 90% da faturação da TNB, a empresa dedicou os seus recursos a aperfeiçoar o processo de entrega, baseando-se na mercadoria vendida por esta loja.

A Ikea é uma loja de venda de mobiliário doméstico e outros produtos complementares ao mobiliário, que oferece uma ampla gama de produtos aos seus clientes, divididos em várias categorias diferentes, que incluem:

- Móveis: Vários tipos de mobiliário para espaços da casa como sala de estar, quartos, cozinhas e casas de banho. Inclui produtos como sofás, cadeiras, camas, secretárias, armários e guarda-roupas. Nesta categoria está inserida a grande maioria da mercadoria a ser transportada pela TNB pois é a mercadoria que os clientes, geralmente, não conseguem transportar sozinhos;
- Decoração da casa: Engloba produtos que adicionam estilo e personalidade à casa do cliente, como cortinados, tapetes, fronhas de almofada e produtos para colocar nas paredes;
- Cozinha e jantar: Inclui produtos como talheres, utensílios e têxteis de cozinha;
- Iluminação: Envolve diferentes opções para iluminar os espaços como candeeiros de teto, de mesa e de pé;
- Têxteis: Inclui roupas de cama, toalhas e outros produtos à base de tecido;
- Organização da casa: São produtos que ajudam na organização da casa como caixas, cestos e cabides;
- Exterior: Inclui mobiliário para o exterior da casa como mesas e cadeiras de jardim;

Todos os seus produtos são desenhados com foco em acessibilidade e praticidade sendo muitos deles modulares, ou seja, permitem que os clientes os personalizem e configurem de modo a satisfazer as suas necessidades.

3.3 Caracterização da frota e equipas de transporte

Nesta secção, descreve-se a frota de carrinhas e as equipas de transporte utilizadas pela empresa para a realização das entregas. Inicialmente, apresenta-se a composição da frota, destacando as características dos diferentes tipos de carrinhas, como a capacidade de carga e volume. Em seguida, discute-se a estrutura e organização das equipas de transporte, elucidando as funções de cada membro e a sua importância para o cumprimento dos prazos e janelas horárias estabelecidas com os clientes. Esta caracterização é fundamental para compreender as restrições e possibilidades inerentes ao

processo de criação de rotas, servindo como base para o desenvolvimento do algoritmo proposto neste trabalho.

3.3.1 Caracterização da frota

A frota da TNB é composta por dois tipos de carrinhas: elétricas e a gasolina. Todas as carrinhas são alugadas. Neste momento, a empresa tem a obrigação (imposta pelos seus clientes, correndo o risco de penalizações pelo incumprimento) de realizar pelo menos metade das suas rotas diárias em carrinhas elétricas, mas, num futuro próximo, tem o objetivo de atingir os 100% de utilização de carrinhas deste tipo.

O número de carrinhas alugadas pela empresa, neste momento, é de 50, sendo 22 elétricas e 28 a gasolina. Pelo facto de poderem ocorrer imprevistos na operação diária das carrinhas, como avarias ou acidentes, que resultem na impossibilidade da sua utilização, de forma temporária ou permanente, a empresa mantém na sua frota um número de carrinhas mais elevado do que seria a sua necessidade diária.

As carrinhas elétricas utilizadas na frota da empresa são da marca *Maxus*, modelo *eDeliver 9* e apresentam as seguintes características:

- Volume útil até 9,7 metros cúbicos;
- Peso máximo da carga de 1040 Kg;
- Autonomia de cerca de 240 Km.

As carrinhas a gasolina são da marca *Renault*, modelo *Master* e apresentam as seguintes características:

- Volume útil até 10,8 metros cúbicos;
- Peso máximo da carga de 1177 Kg.

De forma a salvaguardar as alterações no peso dos elementos das equipas de transporte e evitar que o peso total bruto das carrinhas, aquando da sua circulação, ultrapasse o limite legal de 3500 Kg, a empresa definiu um limite de 1000 Kg, na criação das rotas de entrega. Foi também definido o limite de 6 metros cúbicos, para reduzir o risco de causar danos na mercadoria, durante o transporte nas carrinhas.

3.3.2 Caracterização das equipas de transporte

Os trabalhadores responsáveis pela realização das entregas são subcontratados pela TNB a outras empresas. Estas equipas de transporte são compostas por dois elementos, o chefe de equipa e um ajudante. O papel do chefe de equipa é definir a ordem pela qual os clientes devem ser visitados e

conduzir a carrinha até às moradas dos clientes. O ajudante é responsável por comunicar com os gestores de tráfego para os manter informados do estado da rota e se existiram imprevistos na sua realização. O ajudante é também responsável por telefonar a cada cliente 15 a 30 minutos antes da chegada à sua morada. Ambos os elementos são responsáveis por carregar a mercadoria na carrinha (no armazém de distribuição) e de a entregar ao cliente na sua morada.

Geralmente é atribuída uma carrinha a cada equipa, sendo que os dois elementos da equipa ficam responsáveis pela sua utilização até que os gestores de tráfego decidam que é necessário trocar de carrinha.

3.4 Descrição do processo de criação de rotas

O processo de criação de rotas de entrega de mercadoria na TNB é levado a cabo pela equipa de três gestores de tráfego da empresa. Geralmente, cada gestor de tráfego está encarregue de criar as rotas para apenas um terminal de distribuição. Na zona de Lisboa existem três terminais de distribuição Alfragide, Loures e Prior Velho (PV), a partir dos quais a mercadoria pode sair em rota para casa do cliente. Este processo de criação de rotas é efetuado no dia anterior ao da realização da rota.

A mercadoria chega ao armazém entre as 07:00 e as 10:00 da manhã de cada dia. Só quando toda a mercadoria tiver dado entrada no armazém é que os gestores de tráfego iniciam a criação das rotas. Para impedir que um contrato seja colocado em duas rotas diferentes, os gestores avisam os restantes quando estão a realizar este processo.

O processo de criação de rotas é realizado no sistema *iMoovit*. Neste sistema, no menu de gestão de contratos é possível obter uma listagem dos contratos que têm agendamento para o dia seguinte. Esta listagem apresenta várias características do contrato como:

- Zona de entrega - A empresa dividiu o país em "zonas", utilizando o nome da localidade e, quando o volume de entregas é elevado, números para a sua diferenciação. Em cada zona de entrega estão agrupados vários códigos postais pela sua proximidade geográfica;
- Peso - Apresenta o somatório do peso da mercadoria inserida no contrato em quilogramas;
- Volume - Apresenta o somatório do volume da mercadoria inserida no contrato em metros cúbicos;
- Tipo de contrato - Esta informação serve como referência aos serviços contidos no contrato;
- Estado do contrato - Indica o estado do contrato no momento da procura. Um contrato pode-se nos estados "Concluído", "A agendar", "Agendado", "Em rota", "Não executado" e "Serviço Cancelado" dependendo da situação em que se encontra;

- Janela horária - Apresenta a janela horária escolhida pelo cliente para a entrega da mercadoria.

Todas estas características dos contratos são muito relevantes para definir a rota em que serão inseridos. Na figura 3.1 observa-se a listagem que se obteve ao procurar por contratos com o tipo “a agendar” e criados no dia 7 de dezembro de 2023. Pode-se observar algumas das características mencionadas, bem como outras não tão relevantes para o problema em questão. Note-se que na figura 3.1 a procura é realizada para contratos criados naquele dia, e não para contratos agendados para esse dia. Alguns contratos não apresentam a informação da data e janela horária da entrega e teriam ainda de ser agendados com o cliente.

ID	Data	Tipo Contrato	Estado	Data Auto	Localidade	Terminal Entrega	Zona Entrega	Data Entrega	Janela Entrega	Ten
CDPL02/017346	07-12-2023 10:23	Transporte com Montagem	A agendar		Alcoentre	PV	Nova NE - Alenquer/Vila F de Xira/Arruda Vinhos	29/12/2023	09:00 - 18:00	
CDPL02/017900	07-12-2023 10:23	Transporte	A agendar		Torres Vedras	PV	Nova NE - Mafra/TorreVedras/Sobral	12/12/2023	09:00 - 18:00	
CDPL02/018293	07-12-2023 10:23	Transporte com Montagem	A agendar		Alvalade	Palmela	Nova NE - Alentejo Litoral - TA12	21/12/2023	09:00 - 18:00	
CDPL02/016957	07-12-2023 10:23	Transporte com Montagem	A agendar		Entroncamento	PV	Nova NE - Santarém - TA08	26/12/2023	09:00 - 18:00	
CDPL02/017682	07-12-2023 10:23	Transporte	A agendar		Silvares	Castelo Branco	Nova NE - Castelo Branco - TA09	13/12/2023	09:00 - 18:00	
CDPL02/017819	07-12-2023 10:23	Transporte	A agendar		Fundão	Castelo Branco	Nova NE - Castelo Branco - TA09	15/12/2023	09:00 - 18:00	
CDPL02/017979	07-12-2023 10:23	Transporte com Montagem	A agendar		Longueira	Algarve	Nova NE - Baixo Alentejo Litoral	12/12/2023	09:00 - 18:00	
	07-12-2023 12:36	SAV	A agendar		Casais Novos	Santarém	Nova NE - Santarém - TA08	12/12/2023	09:00 - 13:00	
	07-12-2023 12:44	Transporte com Montagem	A agendar		Porto	Porto	Nova NE - Porto			
	07-12-2023 13:21	Transporte	A agendar		SETÚBAL	Palmela	Nova NE - Setúbal - TA07			
	07-12-2023 13:21	Transporte com Montagem	A agendar		CAPARICA	Palmela	Nova NE - Margem Sul - TA06			
	07-12-2023 13:21	Transporte	A agendar		OEIRAS	PV	Nova NE - Oeiras			
	07-12-2023 13:21	Transporte com Montagem	A agendar		SESIMBRA	Palmela	Nova NE - Setúbal - TA07			
	07-12-2023 13:21	Transporte com Montagem	A agendar		CARNAXIDE	PV	Nova NE - Oeiras			
	07-12-2023 13:22	Transporte	A agendar		Rio Tinto	Porto	Nova NE - Porto			
	07-12-2023 13:22	Transporte com Montagem	A agendar		CABANAS DE TAVIRA	Algarve	Nova NE - Tavira			
	07-12-2023 13:22	Transporte	A agendar		ALMADA	Palmela	Nova NE - Margem Sul - TA06			

Figura 3.1 - Excerto de visualização da lista de contrato no sistema iMoovit

De seguida, o gestor de tráfego seleciona todos os contratos correspondentes ao terminal de distribuição que lhe foi alocado, que tenham a zona de entrega em comum e, copia-os para o menu de criação de rotas. Este é o menu onde efetivamente são criadas as rotas. Neste menu é possível selecionar algumas rotas “pré-definidas” que foram criadas em sistema, pois é sabido que haverá sempre um número elevado de contratos para entregar por dia para certas zonas de entrega. Assim estas rotas quando são selecionadas, agregam os contratos com a zona de entrega definida e, apresentam alguns indicadores que permitem ao gestor de tráfego perceber o estado da rota em questão. São apresentados valores como o número de contratos, o peso, volume total e valor dos contratos selecionados em euros. Na figura 3.2 observa-se uma rota criada para a zona da Margem Sul e Setúbal.

14ctls 79% 539,94€ 3434mi 200,60€				Nº 1612082 - PV Transporte - Margem Sul										Manhã: 6 Tarde: 8 Noite: 0 Dia: 0			
2627,20kg 131%		13,80m³ 106%		13835,74€		Nº Contrato	Ref. Externa	Estado	T. Distribuição	Trab. Cond.	Operação	Tipo Contrato	Data Entrega	Janela Horária	Zona Entrega	Localidade	Cliente
Trans.	Estrada Infinita - 613 - Paulo Marques v			101-1196244		Em rota	PV	100 %	Entrega	Transporte	29/10/2022	09:00 - 13:00	Nova NE - Margem Sul - TA06	2860-457 Moita			Nuno Miguel Dias Ja...
				101-1196941		Em rota	PV	100 %	Entrega	CDC Transporte	29/10/2022	09:00 - 13:00	Nova NE - Margem Sul - TA06	2860-472 Moita			Walter Monteiro
				110-1152239		Em rota	PV	100 %	Entrega	CDC Transporte	29/10/2022	09:00 - 13:00	Nova NE - Setúbal - TA07	2975-146 Quinta do Conde			Miguel Beito
				168-14098	675382_3	A agendar	PV	100 %	Entrega	Transporte	01/11/2022	09:00 - 18:00	Nova NE - Setúbal - TA07	2975-146 Quinta do Conde			JOSE ATAIDE QUINTA...
Aux.				101-1196672		Em rota	PV	100 %	Entrega	Transporte	29/10/2022	09:00 - 13:00	Nova NE - Margem Sul - TA06	2845-589 Amora			Sara Patricia da Cruz...
				101-1194932		Em rota	PV	100 %	Entrega	CDC Transporte	29/10/2022	09:00 - 13:00	Nova NE - Margem Sul - TA06	2805-239 Almada			sonia versosimo
Aux.2				101-1196420		Em rota	PV	100 %	Entrega	Transporte	29/10/2022	13:00 - 17:00	Nova NE - Margem Sul - TA06	2830-591 Barreiro			Margarida Lobato
				101-1198331		Em rota	PV	100 %	Entrega	Transporte	29/10/2022	13:00 - 17:00	Nova NE - Margem Sul - TA06	2830-283 Barreiro			Tiago Corona
Viat.	Sub. Estrada Infinita - 613 v			110-1152121		Em rota	PV	100 %	Entrega	CDC Transporte	29/10/2022	13:00 - 17:00	Nova NE - Margem Sul - TA06	2830-529 Santo André			Bruno Monteiro
				110-1152044		Em rota	PV	100 %	Entrega	CDC Transporte	29/10/2022	13:00 - 17:00	Nova NE - Margem Sul - TA06	2835-410 Lavradio			Solanch Milanis Delg...
Term.				154-7985		Em rota	PV	75 %	Entrega	Transporte	29/10/2022	14:00 - 18:00	Nova NE - Margem Sul - TA06	2860-316 Brejos			Luis Seco
Arm. Armazém Prió Velho																	

Figura 3.2 - Visualização de uma rota no menu de criação de rotas

Na parte esquerda da figura 3.2 é possível observar que o sistema apresenta uma breve análise das características dos contratos que estão inseridos na rota 1612082 - PV Transporte - Margem Sul, aos quais associa um código de cores. Estas cores representam o nível de ocupação da carrinha. O verde indica que o valor desta característica se encontra abaixo do limite definido. A verde encontra-se o número de contratos, que se encontra a 70% da ocupação. A vermelho encontra-se o peso total dos contratos que se encontra a 131% do limite, ou seja, neste caso a rota está com peso superior à carga permitida. Quando a rota está apenas um pouco acima do limite num dos valores, aparece o amarelo, neste caso isso acontece com o volume total dos contratos. Estas cores são apenas indicadores para os gestores de tráfego, pois existem outros fatores que podem permitir que uma carrinha transporte mercadoria que ultrapasse os limites definidos. Pela experiência que têm sobre a destreza das diferentes equipas de transporte, os gestores de tráfego podem decidir definir uma rota com um peso total da mercadoria superior ao limite definido e alocá-la a uma equipa de transporte mais experiente, por exemplo.

É com esse conhecimento sobre as equipas que os gestores de tráfego analisam a mercadoria dos contratos colocados nas rotas e decidem se algum contrato deve ser retirado da rota por conter mercadoria com volume ou peso elevado. Se uma rota já contiver contratos com peso ou volume mais elevado que o limite definido, um contrato com mercadoria difícil de transportar será passado para outra rota. Existe também o cuidado de tentar equilibrar o número de contratos realizados durante a manhã e a tarde, tentando garantir sempre o cumprimento das janelas horárias.

Quando o gestor de tráfego está satisfeito com a rota criada tem de alocá-la a uma equipa. Normalmente esta escolha parece ser fácil, pois algumas equipas são atribuídas todos os dias às mesmas áreas geográficas. O tipo de mercadoria a ser transportada também pode influenciar a escolha da equipa, pois pode ser necessário uma equipa mais experiente.

No dia seguinte, a informação sobre a rota é impressa e entregue às respetivas equipas. As equipas são também informadas sobre algumas características peculiares dos contratos a entregar, bem como sobre as localidades que terão de visitar para realizar as entregas. Quando as equipas concluem o carregamento da mercadoria nas carrinhas, escolhem a ordem pela qual visitam cada cliente, bem como o caminho que vão ter de realizar para lá chegar, tendo em consideração as janelas horárias das entregas.

3.5 Tempo de Serviço

Neste contexto de entrega de mercadorias, tempo de serviço, refere-se ao período que decorre desde a chegada da equipa de entrega à casa do cliente até ao momento em que retomam a marcha para visitar o cliente seguinte. Este intervalo de tempo inclui todas as atividades relacionadas com a entrega, como a descarga das mercadorias, a interação com o cliente, a verificação da documentação e outros processos operacionais. No capítulo 4 é apresentada uma proposta de estimação do tempo de serviço de um contrato.

3.5.1 Recolha de dados

De modo a estimar o tempo de serviço necessário em cada contrato realizado, foram recolhidos alguns dados. Para o efeito, durante o mês de março de 2024, foi feito o acompanhamento de várias equipas de transporte enquanto efetuam as entregas. Foram consideradas rotas efetuadas em diferentes zonas como Lisboa, Sintra, Cascais, Oeiras, Amadora, Almada, Seixal, Barreiro e Setúbal. Deste modo foi possível medir o tempo de serviço de cada contrato, bem como recolher outros dados eventualmente relevantes sobre a mercadoria transportada de cada um, tais como o seu peso, volume e o número de caixas. Foi ainda assinalada a ordem pelos quais os contratos de cada rota foram entregues, pois foi considerada a hipótese do desgaste da equipa de entrega poder influenciar o tempo de entrega.

O tempo de entrega foi medido a partir do momento que a equipa estaciona a carrinha no local da entrega, até ao momento em que retoma a marcha para visitar o cliente seguinte.

No total foram recolhidos os dados de 136 contratos realizados por 3 equipas diferentes. Esta informação pode ser consultada na tabela A.1 do anexo A.

No processo de acompanhamento das equipas no ato da entrega da mercadoria aos clientes foi possível perceber, por observação e comunicação com as equipas, que o peso da mercadoria a transportar é o fator que mais influencia o tempo de entrega, para além de outros fatores como o volume, o número de caixas e a ordem de entrega na rota. Fatores como o andar da casa do cliente, a existência de elevador no prédio ou os acessos a casa do cliente não serão considerados para a estimação do tempo de entrega porque, na maior parte das vezes, não fazem parte da informação solicitada ao cliente, ou por serem difíceis de quantificar.

3.5.2 Caracterização dos dados

Os dados recolhidos estão divididos nas seguintes características:

- Tempo de entrega: medido em minutos, valores compreendidos entre 1 e 37 minutos com um valor médio de 6,9 minutos;
- Peso: medido em quilogramas, valores compreendidos entre 0,03 e 745,17 Kg com um valor médio de 101,61 Kg;
- Volume: medido em metros cúbicos, valores compreendidos entre 0,003 e 3,343 m^3 com um valor médio de 0,439 m^3 ;
- Número de caixas: valores compreendidos entre 1 e 113 caixas com um valor médio de 12,8 caixas e uma mediana de 6 caixas;
- Ordem na rota: atribuído o valor de 1 ao primeiro contrato entregue da rota o valor de 2 ao segundo e assim em diante.

Na figura 3.3 é apresentada a relação entre os valores do peso de cada contrato, em quilogramas e o tempo da entrega, em minutos.

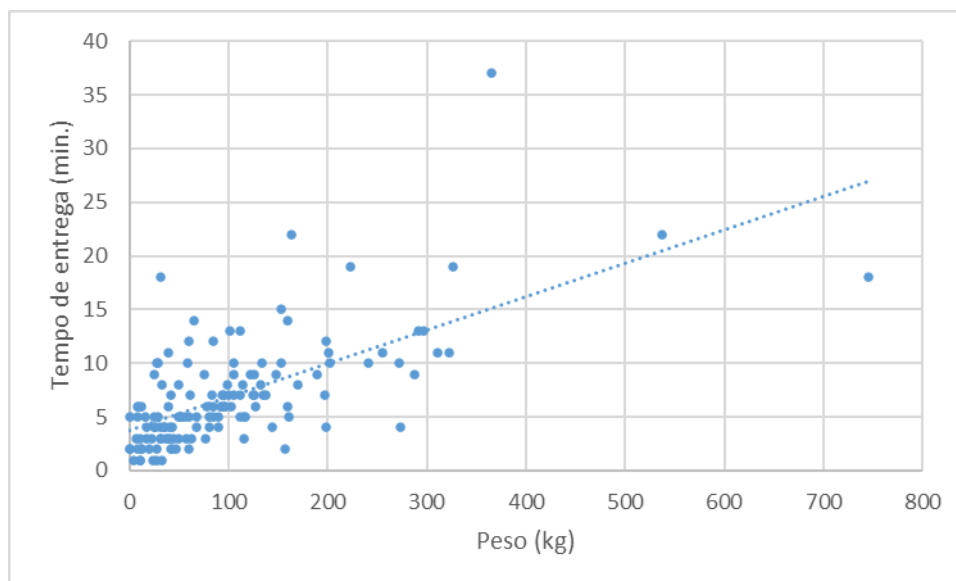


Figura 3.3 - Relação entre o peso e o tempo de entrega

Verifica-se que o tempo de entrega é maior quando o peso da mercadoria é mais elevado representando assim uma correlação positiva entre os dados das duas variáveis. Dada a dispersão observada em alguns valores pode haver a possibilidade da existência de *outliers* nos dados, pelo que este facto deverá ser considerado de forma a alcançar uma estimacão mais fidedigna aos dados.

De seguida analisou-se a possibilidade de existência de uma relação entre o volume e o tempo de entrega dos contratos, figura 3.4.

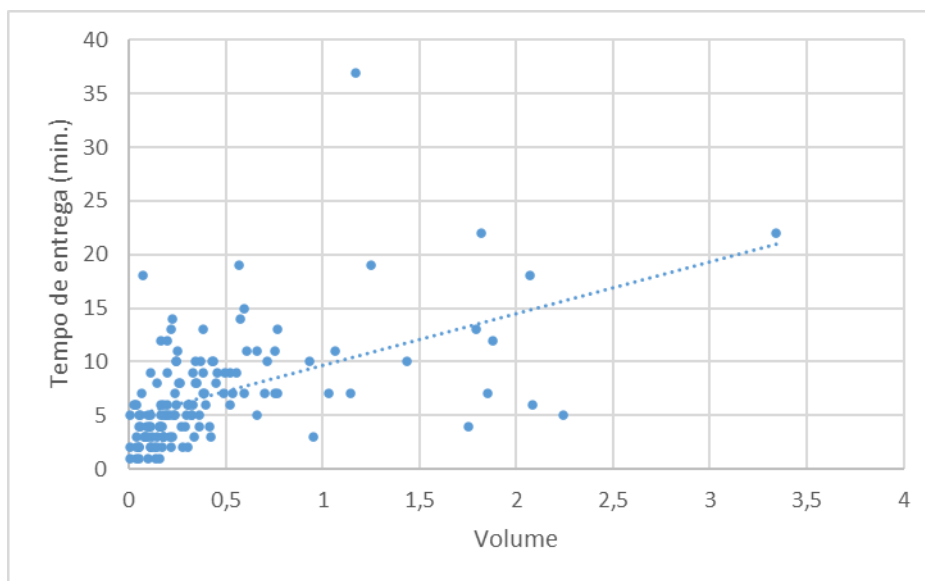


Figura 3.4 - Relação entre o volume e o tempo de entrega

Na figura 3.4, apresenta-se a relação entre o volume, em metros cúbicos, e o tempo de entrega, em minutos. Nesta figura, é possível observar que existe uma correlação positiva entre os dados, embora não tão clara como na relação entre o tempo de entrega e o peso. Mais uma vez é notada a possibilidade de existência de *outliers*.

Foi também analisada a relação entre o número de caixas e o tempo de entrega, figura 3.5.

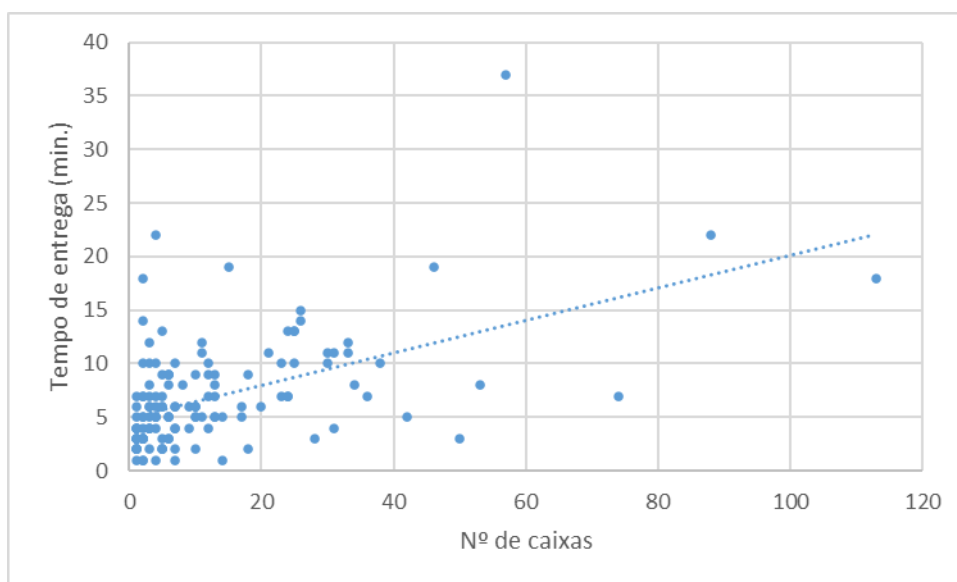


Figura 3.5 - Relação entre o número de caixas e o tempo de entrega

Pela análise da figura 3.5, parece existir uma grande dispersão dos dados, sendo assim possível que o número de caixas não tenha uma influência relevante no tempo de entrega, pelo que poderá não ser significativo na sua estimação. Esta conclusão terá de ser comprovada por uma análise dos dados, realizada no próximo capítulo (capítulo 4).

ALGORITMO DE CRIAÇÃO DE ROTAS

Neste capítulo apresenta-se o desenvolvimento do algoritmo de criação de rotas. Inicia-se com a apresentação dos objetivos, a identificação das variáveis utilizadas e as restrições a cumprir. Seguidamente é apresentada a proposta do algoritmo em pseudocódigo, bem como uma explicação da lógica do código desenvolvido.

Mostra-se também o tratamento que é necessário realizar aos dados para servirem de input ao algoritmo desenvolvido e aplica-se o algoritmo para definição das rotas a seguir, para fazer as entregas de mercadoria aos clientes de diferentes dias. Por fim, realiza-se uma análise comparativa das rotas obtidas pelo algoritmo proposto e das executadas atualmente.

4.1 Objetivos do algoritmo

O objetivo do algoritmo é criar uma série de rotas a partir dos dados de *input* relativos aos contratos de entrega de mercadoria a clientes. As rotas devem ser criadas, realizando um agrupamento dos contratos pela distância entre as moradas dos clientes em questão, com o propósito de reduzir as distâncias percorridas pelas carrinhas de entrega. Devem ainda garantir o cumprimento das janelas horárias dos contratos, sem ultrapassar os limites de peso e volume das carrinhas.

O *output* do algoritmo será então uma lista de rotas para a "Manhã" e para a parte da "Tarde", sendo que a parte relativa à manhã apenas deverá conter contratos com janelas horárias que terminem até às 13 horas ou janelas horárias de dia inteiro. A parte da tarde deverá conter os contratos cujas janelas horárias se iniciem às 14 horas e os contratos com janelas horárias de dia inteiro.

4.2 Variáveis de input

Na figura 4.1 é apresentado um diagrama que apresenta as variáveis deste problema.

As variáveis de *input* serão fornecidas ao algoritmo pelo utilizador. Para cada contrato, é necessário indicar o peso, o volume, bem como as coordenadas geográficas da morada de entrega e as características da janela horária. Estes dados são cruciais para permitir que o algoritmo crie as rotas e defina a ordem de entrega.

Estes dados permitem ao algoritmo também estimar os tempos de serviço e de entrega para cada contrato. O algoritmo inicia então o processo de criação de rotas tendo em conta as restrições de peso e de volume máximos das carrinhas de entrega.

O *output* do algoritmo será uma lista de rotas, nas quais cada rota é constituída por uma série de contratos organizados por ordem de visita. Serão também apresentadas algumas informações sobre cada rota, como o peso e volume totais da mercadoria e o tempo estimado para a chegada à morada de cada cliente.

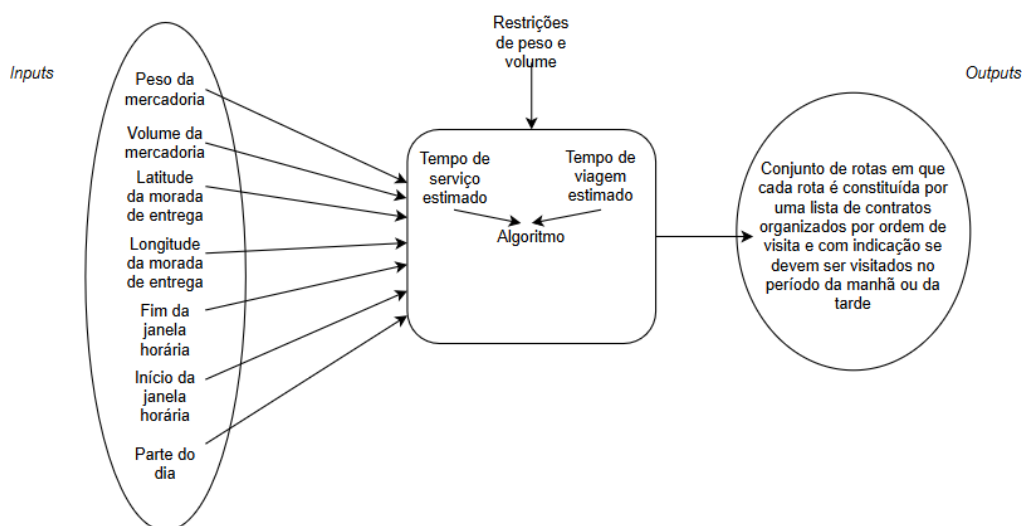


Figura 4.1 - Diagrama de *inputs* e *outputs*

4.3 Restrições

O algoritmo de criação de rotas a ser desenvolvido tem o objetivo de resolver um problema complexo associado à entrega de mercadoria na casa dos clientes uma vez que, cada rota, tem de obedecer a uma série de restrições temporais e de capacidade das carrinhas usadas nas entregas.

Existem restrições temporais porque cada cliente possui uma janela horária associada que deve ser cumprida. Estas restrições revelam que é de extrema importância o algoritmo estimar o tempo despendido nas viagens entre os clientes, bem como o tempo de serviço.

As restrições na carga, no que diz respeito ao peso e ao volume das carrinhas de transporte significam que o algoritmo tem de controlar estes valores e garantir que qualquer rota não tenha

valores superiores aos limites impostos. Para as carrinhas usadas no transporte, o limite para o peso da mercadoria a transportar é de 1000 quilogramas e o limite de volume é de 6 metros cúbicos.

4.4 Tempo de serviço

4.4.1 Análise de regressão linear múltipla

Para conseguir estimar o tempo de serviço de um contrato foi realizada uma análise de regressão linear múltipla dos dados recolhidos. Nesta análise o tempo de entrega é a variável dependente e o peso, o volume e o número de caixas são as variáveis independentes. Esta análise permite perceber os efeitos, que as variáveis independentes têm no tempo de serviço.

Esta análise foi realizada utilizando a ferramenta de análise de dados do *Microsoft Excel*.

Os resultados da utilização desta ferramenta podem ser observados nas tabelas 4.1, 4.2 e 4.3.

Na tabela 4.1 podemos observar as estatísticas da regressão. Estas estatísticas providenciam informações sobre a qualidade do ajuste e o poder explicativo do modelo de regressão.

Tabela 4.1 - Estatísticas da regressão inicial

R Múltiplo	0,66
R Quadrado	0,44
R Quadrado Ajustado	0,42
Erro Padrão	3,84
Observações	136

O valor de R múltiplo representa o coeficiente de correlação entre os valores observados para a variável dependente, neste caso o tempo de entrega, e os valores previstos pelo modelo. Um R múltiplo de 0,66 indica uma correlação positiva moderada entre as variáveis independentes e a variável dependente.

O coeficiente R quadrado (*R Square*) representa a percentagem da variância na variável dependente que é explicada pelas variáveis independentes. Este modelo explica 44% da variabilidade dos tempos de entrega.

O coeficiente R quadrado ajustado (*Adjusted R Square*), foi desenvolvido para ter em conta o número de variáveis independentes no modelo e penaliza a inclusão de variáveis desnecessárias. Neste modelo, existe uma ligeira diferença entre estes dois coeficientes, revelando uma ligeira penalização pela inclusão de três variáveis independentes. Isto significa que uma ou mais variáveis podem não contribuir significativamente para o ajuste do modelo.

O Erro Padrão é uma medida do desvio médio dos valores observados, em relação à linha de regressão. Neste modelo este valor é, aproximadamente 3,84.

A Tabela ANOVA apresentada na tabela 4.2, é um método estatístico utilizado para dividir a variabilidade total dos dados em diferentes componentes, com o propósito de avaliar a significância do modelo de regressão.

Tabela 4.2 - ANOVA inicial

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	F Significante
Regressão	4	1490,65	372,66	25,22	1,68E-15
Resíduos	131	1935,46	14,77		
Total	135	3426,11			

A primeira linha da tabela 4.2 (Regressão) representa a variabilidade na variável dependente, neste caso, o tempo de entrega, que é explicada pelo modelo de regressão. Este modelo apresenta 4 graus de liberdade e a soma dos quadrados (SS) é a variabilidade total explicada pelo modelo. A média dos quadrados (MS) é obtida através do quociente entre o valor da soma dos quadrados e o número de graus de liberdade. O valor de F é uma razão da variância explicada pelo modelo para a variância não explicada. Quanto maior for este valor mais significativo é o modelo de regressão. O baixo *P-value* (valor muito próximo de zero), para esta estatística, significa que o modelo é estatisticamente significativo.

Por fim, a tabela dos coeficientes da regressão, tabela 4.3 apresenta informação sobre os coeficientes estimados para as variáveis independentes no modelo de regressão.

Tabela 4.3 - Coeficientes da regressão inicial

	Coeficientes	Erro Padrão	<i>t Stat</i>	<i>P-value</i>
Interceção	3,49	0,75	4,66	0,00
Ordem	0,01	0,07	0,16	0,87
Peso	0,02	0,01	3,75	0,00
Volume	1,58	0,84	1,88	0,06
Nº caixas	0,03	0,03	1,08	0,28

A primeira coluna apresenta os coeficientes estimados para cada variável independente no modelo. Os coeficientes representam a alteração na variável dependente (tempo de entrega) resultante pela alteração de uma unidade na variável independente a que cada um está associado, mantendo todas as outras variáveis constantes. Por exemplo, se o coeficiente para "Peso" for 0,02, sugere-se que para cada aumento de uma unidade no peso, o tempo de entrega é esperado aumentar em 0,02 unidades de tempo, mantendo tudo o resto constante.

Na segunda coluna é apresentado o valor para o erro padrão, que mede a variabilidade ou precisão da estimativa do coeficiente. Um erro padrão menor indica uma estimativa mais precisa do coeficiente. Representa a quantidade média pela qual a estimativa do coeficiente difere do valor real do parâmetro em amostragens repetidas. Por exemplo, um erro padrão de 0,01 para o coeficiente da variável "Peso" sugere que o efeito estimado do peso no tempo de entrega é relativamente preciso.

A estatística *t* avalia a significância de cada estimativa de coeficiente. Compara o coeficiente estimado com zero e indica se o coeficiente é significativamente diferente de zero. Quanto maior o valor absoluto da estatística *t*, mais evidência há contra a hipótese de que o coeficiente é zero. Uma estatística *t* com magnitude superior a 2 ou inferior a -2 é frequentemente considerada estatisticamente significativa para um nível de significância de 95%. Por exemplo, se a estatística *t* para "Peso" for 3,75, o coeficiente para o peso é significativamente diferente de zero.

O *P-value* associado a cada coeficiente testa a hipótese nula de que o coeficiente é igual a zero (ou seja, que a variável independente não tem efeito sobre a variável dependente). Um *P-value* baixo (tipicamente abaixo de um nível de significância escolhido, como 0,05) indica que o coeficiente é estatisticamente significativo, significando que há forte evidência contra a hipótese nula. Por outro lado, um *P-value* alto sugere que o coeficiente não é estatisticamente significativo e que não existe evidência suficiente para rejeitar a hipótese nula. Por exemplo, o valor do *P-value* para "Peso" ser inferior a 0,05 sugere que o efeito do peso no tempo de entrega é estatisticamente significativo (para um nível de significância de 0,05).

O termo *intercept* representa o valor estimado para o tempo de entrega quando todas as variáveis independentes estão a 0. Um *P-value* próximo de 0 significa que o *intercept* é estatisticamente significativo.

O *P-value* da variável "Peso" indica que é estatisticamente significativa. O coeficiente da variável "Volume" é 1,58, mas, neste caso, o seu *P-value* de 0,06 (superior a 0,05) sugere que a variável pode não ser significativa para prever a variável dependente. As variáveis "Número de caixas" e "Ordem" apresentam valores muito elevados de *P-value* o que indica que estas variáveis não são significativas para prever a variável dependente.

O facto das variáveis "Número de caixas" e "Ordem" não serem estatisticamente significativas significa que para este modelo, alterações nestas variáveis não estão associadas a alterações significativas na variável dependente. Como, em teoria, estas variáveis seriam importantes na previsão do tempo de entrega, foi realizada uma nova análise de regressão apenas com as variáveis "Peso" e "Volume".

Desta nova análise resultaram valores muito semelhantes para as estatísticas de regressão, o que comprova que a remoção das variáveis não significativas, não reduziu o poder preditivo do modelo. Foi também obtido um *P-value* de 0,08 para a variável "Volume". Com o objetivo de concluir se esta variável deve ser considerada no modelo foi analisada a tabela de resíduos resultante da análise de regressão, que se encontra no Anexo B (tabela B.1).

A tabela B.1 apresenta o valor da variável dependente estimado pelo modelo e apresenta também os valores do resíduo e do resíduo padronizado para cada observação. O valor dos resíduos é obtido a partir da diferença entre o valor estimado e o valor obtido efetivamente para cada observação. O valor dos resíduos padronizados é obtido pelo quociente entre o valor dos resíduos e o valor do seu desvio padrão.

Se uma observação apresenta um valor extremo (superior a 2 ou inferior a -2) para os resíduos padronizados, isto pode significar que essa observação é um *outlier* e a sua remoção do modelo de dados pode resultar num modelo mais forte. Foram então selecionadas todas as observações que apresentam valores extremos para os resíduos padronizados.

De seguida foi realizada novamente uma análise de regressão linear múltipla com o conjunto de dados, sem os referentes às observações consideradas como *outliers* e foram analisadas as tabelas, respetivas.

Analisando as tabelas é possível concluir que a remoção dos *outliers* resultou num modelo mais competente a prever os tempos de entrega. Este novo modelo explica 51,5% da variabilidade da variável dependente, revelando assim uma subida de 7 pontos percentuais em relação ao modelo anterior. Também o valor do Erro Padrão desceu dos 3,84 para aproximadamente 2,62.

Na tabela ANOVA o valor de F também sofreu um aumento, comprovando assim que a remoção de *outliers* resultou num modelo de regressão mais significativo.

Em relação às variáveis, obteve-se um valor de *P-value* associado à variável "Volume" de 0,66 o que indica que esta variável é estatisticamente insignificante e deverá ser removida do modelo.

O processo de remoção de *outliers* pela análise da tabela de resíduos do modelo de regressão, foi repetido até que não existissem observações com valores extremos de resíduos padronizados ou que a sua remoção não beneficiasse o poder preditivo do modelo.

Finalmente obteve-se um modelo de regressão com as seguintes características (tabelas 4.4, 4.5 e 4.6):

Tabela 4.4 - Estatísticas da regressão do modelo final

R Múltiplo	0,88
R Quadrado	0,77
R Quadrado Ajustado	0,77
Erro Padrão	1,56
Observações	109

Tabela 4.5 - ANOVA do modelo final

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	F Significante
Regressão	1	891,19	891,19	366,22	0,00
Resíduos	107	260,39	2,43		
Total	108	1151,58			

Tabela 4.6 - Coeficientes do modelo final

	Coeficientes	Erro Padrão	<i>t Stat</i>	<i>P-value</i>
Interceção	2,63	0,22	12,18	0,00
Peso	0,03	0,00	19,14	0,00

O modelo apresentado nas tabelas 4.4, 4.5 e 4.6 será utilizado para prever os tempos de entrega baseado nos dados recolhidos do peso da mercadoria a ser entregue.

Este modelo explica cerca de 77% da variabilidade dos tempos de entrega. O resto da variabilidade estará relacionado com alguns dos fatores já mencionados anteriormente que não foram considerados neste modelo, como o andar da casa onde é realizada a entrega, acessos a casa do cliente e a existência de elevador ou não, entre outros.

Assim, o modelo indica que o tempo de serviço de um contrato (TS), é estimado com base no peso (P) da mercadoria a ser entregue, utilizando a equação (4).

$$TS = 2,63 + 0,03 * P \quad (4)$$

Sendo "TS" o tempo de serviço em minutos e "P" o peso do contrato em quilogramas.

4.5 Proposta de algoritmo

Nesta secção será apresentada a proposta de solução para o problema. Será apresentado um diagrama que demonstra a lógica do funcionamento do algoritmo desenvolvido. Será também feita uma explicação sobre todos os passos necessários para cumprir os objetivos propostos para o algoritmo.

A escolha da linguagem de programação desempenha um papel crucial na eficácia e eficiência do desenvolvimento de algoritmos, especialmente em contextos logísticos complexos como os problemas de criação de rotas de entregas. No caso da implementação deste algoritmo, optou-se por utilizar *Python*, devido a várias razões fundamentais.

4.5.1 Algoritmo de criação de rotas

O algoritmo de criação de rotas a ser desenvolvido, necessita de cumprir uma série de objetivos e de restrições, como visto na secção 4.1. Na figura 4.2 observa-se um diagrama que demonstra o fluxo de funcionamento do algoritmo. Nas sub-subsecções 4.5.1.1 até 4.5.1.5 são explicadas, em detalhe, as diferentes fases de funcionamento do algoritmo. No anexo C, nas figuras C.1, C.2 e C.3, é possível observar partes do código desenvolvido.

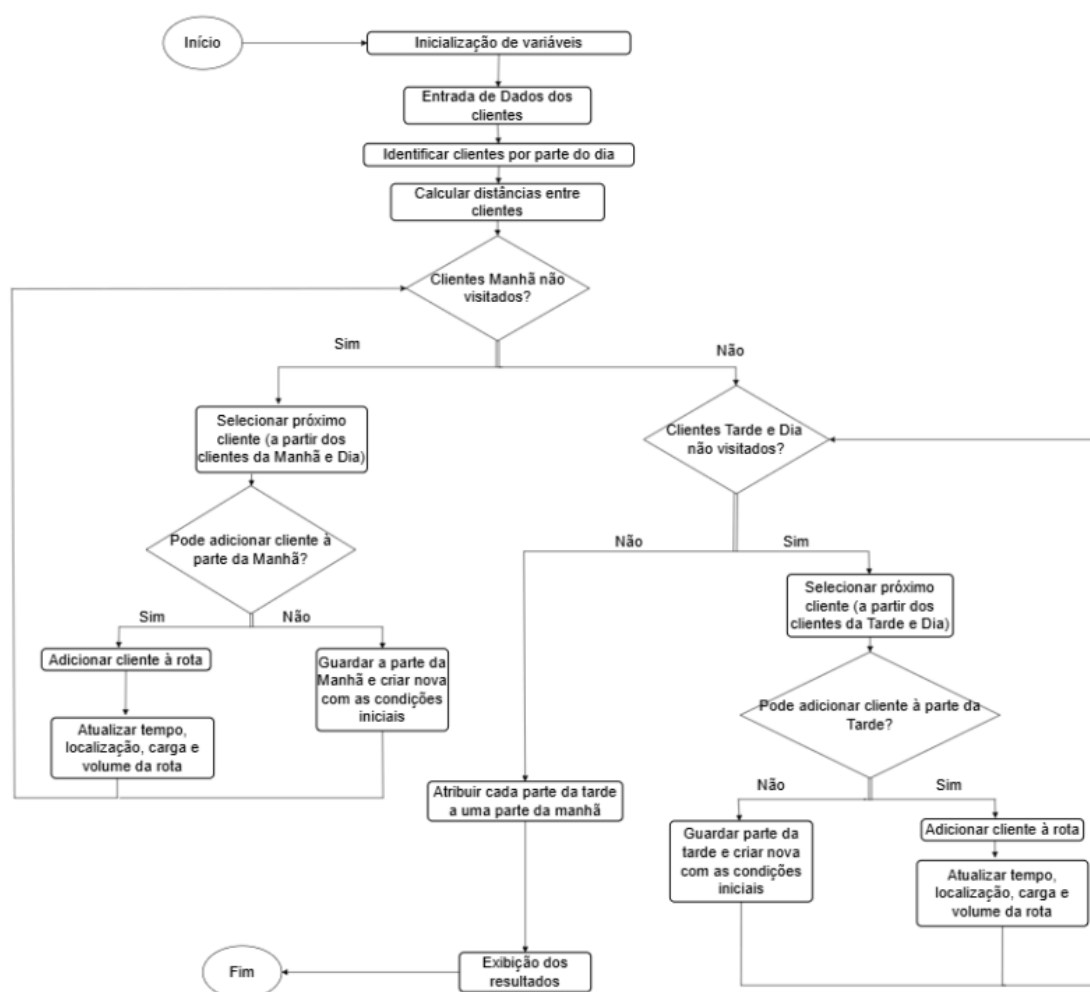


Figura 4.2 - Estrutura genérica do algoritmo

4.5.1.1 Configuração inicial

Numa fase inicial, é realizada a importação das bibliotecas *googlemaps*, *datetime* e *numpy* para fornecer funcionalidades relacionadas com mapas, manipulação de data/hora e operações numéricas, respetivamente.

Duas classes são definidas: i) *TimeWindow* para representar janelas de tempo e ii) *Client* para representar os contratos. Valores constantes como o peso máximo (Kg), volume máximo (m^3) e o ponto de partida de qualquer rota (coordenadas geográficas do armazém de distribuição) também são definidas.

É ainda efetuada a configuração do cliente da API do *Google Maps* com uma chave de API que não é fornecida neste documento por razões de segurança. Isto permitirá a obtenção de informações de distância e tempo de viagem entre locais.

4.5.1.2 Entrada de Dados

Para realizar a entrada dos dados no algoritmo, a função *collect_client_data* permite que o utilizador insira dados dos contratos, incluindo número (identificado como ID), coordenadas, peso, volume, janela horária e *slot* de tempo (Manhã, Tarde e Dia).

4.5.1.3 Lógica de seleção do próximo cliente

A função *can_add_to_route* verifica se um contrato pode ser adicionado à rota com base nas restrições de peso, volume e janela horária. A função *select_next_client* escolhe o próximo contrato a ser adicionado à rota com base na menor distância em relação ao último contrato adicionado à rota ou ao ponto de partida (se a rota estiver vazia). A função *add_candidate_to_route* adiciona o contrato candidato à rota e retorna as coordenadas do cliente do contrato adicionado. Quando um cliente é adicionado a uma rota, é necessário atualizar o tempo atual com o tempo de viagem estimado entre o cliente do último contrato adicionado e o cliente do contrato candidato e atualizar a carga e o volume com os dados do contrato adicionado.

4.5.1.4 Construção de rotas

Na construção de rotas, são utilizadas todas as funções definidas anteriormente para criar as rotas.

A função *construct_routes* divide os contratos em grupos de acordo com os seus *slots* de tempo de entrega e constrói rotas considerando esses grupos. Esta função itera sobre os contratos de cada *slot* de tempo (Manhã, Tarde, Dia) e constrói rotas separadamente para cada parte do dia, Manhã e Tarde, usando a lógica de seleção do próximo contrato e atualiza a informação da rota.

Quando todos os contratos fornecidos como input, estão inseridos numa rota, todas as rotas da parte da "Tarde" são atribuídas a uma rota da parte da "Manhã", quando possível. Este processo permite criar a rota final, que é atribuída, mais tarde, à equipa de transporte, que obtém assim a informação de qual a mercadoria a carregar para entrega de manhã, e à tarde. Quando existem mais rotas criadas da parte da "Tarde" do que da parte da "Manhã" ou vice-versa, é criada uma rota apenas com a parte disponível, significando assim que, para essas rotas, a equipa de transporte irá apenas realizar entregas na parte do dia correspondente.

4.5.1.5 Apresentação dos resultados

Após as rotas serem criadas, as quais, incluem informações como a janela horária do contrato e a parte do dia em que cada cliente vai ser visitado, bem como o peso total e volume total da rota, são apresentados os resultados. As informações são apresentadas de forma clara, para análise e interpretação do utilizador.

4.5.2 Funções desenvolvidas

Para escolher quais os contratos que devem ser adicionados a cada rota foi necessário criar funções que estimam a distância e o tempo de viagem entre as coordenadas geográficas das moradas de todos os clientes aos quais é necessário proceder à entrega no dia para o qual se está a fazer o planeamento. Essas funções utilizam os serviços de API da *Google* para estimar estes valores acedendo assim aos dados do *Google Maps*.

Com base no conjunto de contratos selecionados, foi definido que o primeiro contrato a ser adicionado a uma rota é aquele cuja morada de entrega seja a mais próxima do armazém de distribuição. A partir daqui, foi criada uma função que avalia qual o contrato com a morada mais próxima do último cliente inserido na rota, sendo assim esse o candidato a ser considerado. Foi também criada uma função que avalia se, o contrato pode ser realizado dentro da sua janela horária, considerando o tempo de viagem e o tempo de entrega estimados.

Quando um cliente é adicionado a uma rota é necessário atualizar as informações sobre a localização geográfica da carrinha e o tempo atual, bem como o peso e o volume total envolvidos na rota até ao momento.

Finalmente foi necessário criar a função de criação de rotas. Para as rotas a criar terem em conta as partes do dia (Manhã e Tarde), a função cria 3 listas diferentes dos contratos pela parte do dia relativa à sua janela horária. De seguida, esta função percorre iterativamente as listas de clientes e constrói rotas para a parte da "Manhã" e parte da "Tarde" sendo que os contratos correspondentes à lista "Dia" podem ser inseridos nas rotas de ambas as partes do dia.

Em seguida, apresenta-se as funções desenvolvidas (com uma breve explicação do seu funcionamento) para que o algoritmo funcione da forma devida.

calcular_distancia (localizacao1, localizacao2):

Função que utiliza a API do *Google Maps* para calcular a distância entre duas localizações geográficas, definidas por *localizacao1* e *localizacao2*, argumentos da função.

calcular_tempo_de_viagem (localizacao1, localizacao2):

Função que utiliza a API do *Google Maps* para calcular o tempo de viagem entre duas localizações geográficas, definidas por *localizacao1* e *localizacao2*, argumentos da função.

calcular_tempo_de_entrega (cliente):

Função que calcula o tempo de entrega de um cliente (único argumento) com base no peso da mercadoria a ser entregue utilizando a equação obtida na subsecção 4.4.1.

pode_adicionar_ao_percurso (candidato, percurso_atual, tempo_atual):

Função que verifica se a rota está vazia. Se estiver vazia, permite a adição sem restrições, do candidato selecionado, que está identificado no primeiro argumento da função.

Se a rota não estiver vazia, verifica se a adição do candidato à rota, não ultrapassa as capacidades máximas de peso e volume da carrinha.

Caso não ultrapasse, calcula o tempo de viagem até ao candidato a partir da localização atual, que está identificada no segundo argumento da função, e soma o tempo de entrega estimado. Em seguida, verifica se esse tempo, adicionado ao tempo que já tem a rota, que está identificado no terceiro argumento da função por tempo atual, está dentro da janela horária do cliente.

selecionar_proximo_cliente (percurso_atual, clientes_nao_visitados, tempo_atual):

Função que seleciona o próximo cliente que potencialmente poderá ser inserido na rota, com base na localização atual da rota, no conjunto de clientes que ainda não foi inserido em rotas e no tempo já consumido pelos clientes atribuídos à rota, identificados pelos argumentos 1, 2 e 3 respetivamente.

Se a rota está vazia, define a localização atual como a posição inicial, se não está, define a localização atual como as coordenadas geográficas do último cliente inserido.

Para cada cliente não visitado, calcula a distância até ao cliente candidato a ser inserido na rota, a partir da localização atual.

Seleciona o cliente mais próximo que atende às restrições de tempo e capacidade.

adicionar_cliente_ao_percurso (cliente, percurso_atual, tempo_atual):

Função que adiciona um cliente (primeiro argumento) à rota que está a ser construída no momento (segundo argumento) e calcula a hora em que o cliente irá ser visitado com base no tempo atual da rota (terceiro argumento)

Adiciona o cliente à rota.

Utiliza a função *calcular_tempo_de_viagem* para calcular o tempo de viagem entre a localização atual da rota e as coordenadas geográficas do cliente inserido

Utiliza a função *calcular_tempo_de_entrega* para calcular o tempo de entrega do cliente inserido.

Define o tempo de visita do cliente como a soma do tempo atual da rota com o tempo de viagem e o tempo de entrega do cliente.

Atualiza a localização atual para as coordenadas do cliente.

Atualiza a carga e o volume da carrinha com base nas características do cliente.

construir_rotas (clientes):

Inicialização do algoritmo e Divisão de Clientes:

Cria conjuntos de clientes para a parte da manhã, tarde, e dia inteiro.

Inicializa uma lista vazia de rotas.

Iteração Principal:

Enquanto houver clientes não visitados para a parte da manhã ou tarde:

Define a localização atual como o ponto inicial (coordenadas geográficas do armazém de distribuição).

Constrói a parte da Manhã:

Enquanto houver clientes não visitados para a parte da Manhã.

Cria uma rota vazia com as condições iniciais.

Enquanto a rota não estiver cheia.

Selecione o próximo cliente a ser visitado (na lista da manhã e dia inteiro).

Se o cliente puder ser adicionado à rota, adiciona-o e atualiza as condições da rota.

Senão é selecionado o cliente seguinte.

Se nenhum cliente puder ser adicionado, a rota está cheia.

A rota é guardada.

Constrói a parte da Tarde:

Enquanto houver clientes não visitados para a tarde ou dia inteiro.

Cria uma rota vazia com as condições iniciais.

Enquanto a rota não estiver cheia.

Selecione o próximo cliente a ser visitado (na lista da tarde e dia inteiro).

Se o cliente puder ser adicionado à rota, adiciona-o e atualiza as condições da rota.

Senão é selecionado o cliente seguinte.

Se nenhum cliente puder ser adicionado, a rota está cheia.

A rota é guardada.

Atribui cada parte da tarde a uma parte da manhã, quando possível

Retorna a lista de rotas construídas.

4.5.3 Pré-processamento dos dados de input

Os dados sobre cada contrato têm de ser inseridos no algoritmo. No entanto, como não é possível exportar os dados a partir do *iMoovit* no formato desejado, é necessário efetuar um pré-processamento dos mesmos.

O algoritmo necessita de um input que corresponde à concatenação do Nº contrato, Coordenadas Geográficas, Peso, Volume, Início Janela Horária, Fim Janela Horária e Parte do dia, separados por ",". No número de contrato não é necessário realizar nenhuma alteração em relação ao número que é exportado do sistema. As coordenadas geográficas, o peso e o volume devem ser apresentados com um ponto final a separar as unidades das casas decimais. O início e fim de janela horária devem ser apresentados com o seguinte formato: "09:00" (nove horas da manhã). A parte do dia será uma de três hipóteses: Manhã, Tarde e Dia inteiro. Todas estas informações devem estar separadas por vírgulas e sem espaços entre elas.

A partir do *iMoovit*, no menu Exportação de dados de Contratos, é possível escolher quais os campos de dados dos contratos a exportar. Para obter a lista de contratos a utilizar no algoritmo é necessário escolher os seguintes campos para realizar a exportação:

- Nº Contrato;
- Entrega - Morada;
- Entrega - Código Postal;
- Entrega - Localidade;
- Entrega - Janela Horária;
- Peso;
- Volume.

Na figura 4.3 observa-se o exemplo de um contrato exportado com os campos referenciados selecionados.

A	B	C	D	E	F	G
Nº Contrato	Entrega - Morada	Entrega - Código Postal	Entrega - Localidade	Entrega - Janela Horária	Peso	Volume
102-1356115	Rua Passos Manuel Nº78 2º Esq	1150-260	Lisbon	09:00 - 13:00	84,71	0,502

Figura 4.3 - Exemplo da exportação de um contrato

Depois de realizar a exportação foram feitas alterações nos dados para os transformar e colocar no formato requerido pelo algoritmo.

Para obter as coordenadas geográficas relativas à morada do cliente, foi criada uma função no *Excel*, utilizando as ferramentas do *Microsoft Visual Basic*, que utiliza a chave API da *Google* para obter as coordenadas geográficas de uma morada a partir do *Google Maps*. De forma a garantir que as coordenadas geográficas correspondem à morada indicada pelo cliente, foram concatenadas as colunas

correspondentes à morada, código postal e localidade numa outra coluna, designada por "Morada completa". A morada obtida nesta coluna é então utilizada na função criada para obter as suas coordenadas geográficas.

De seguida foi necessário separar a janela horária, resultando numa coluna com a hora de início da janela horária e outra coluna com a hora de fim. Foi, ainda, necessário classificar as janelas horárias pela parte do dia em que se inserem. Para isso foi criada uma fórmula que avalia a hora de início e fim da janela horária e atribui uma classificação de "Manhã", "Tarde" ou "Dia".

Para os valores do peso e do volume foi necessário substituir a vírgula por um ponto final para o algoritmo ler corretamente os números. Os dados de input de cada contrato resultam da concatenação de todos estes dados, separados por uma vírgula, como se observa na tabela 4.7

Tabela 4.7 - Dados para *input* no algoritmo, resultantes da concatenação das colunas

Nº Contrato	Coordenadas geográficas	Peso	Volume	Início JH	Fim JH	Parte do dia	Dados Input
102-1420945	38.8234616, -9.0894936	193.49	0.369	9	13	Manhã	102-1420945,38.8234616, -9.0894936,193.49,0.369,9,13,Manhã
101-1322887	38.8166301, -9.0918143	32.98	0.025	9	13	Manhã	101-1322887,38.8166301, -9.0918143,32.98,0.025,9,13,Manhã
110-1176608	38.8952713, -9.0343573	84.1	0.163	9	13	Manhã	110-1176608,38.8952713, -9.0343573,84.1,0.163,9,13,Manhã
168-37113	38.9532046, -8.9925509	20	0.018	9	13	Manhã	168-37113,38.9532046, -8.9925509,20,0.018,9,13,Manhã
101-1323019	38.9549676, -8.9890667	268.82	1.142	9	13	Manhã	101-1323019,38.9549676, -8.9890667,268.82,1.142,9,13,Manhã
101-1322098	38.9371013, -8.8638596	132.59	1.638	9	13	Manhã	101-1322098,38.9371013, -8.8638596,132.59,1.638,9,13,Manhã
101-1319854	38.9252702, -8.5786010	137.19	1.325	13	17	Tarde	101-1319854,38.9252702, -8.5786010,137.19,1.325,13,17,Tarde
101-1321764	38.9252702, -8.5786010	2.8	0.025	13	17	Tarde	101-1321764,38.9252702, -8.5786010,2.8,0.025,13,17,Tarde
101-1322318	38.9637309, -8.5225249	29.33	0.106	13	17	Tarde	101-1322318,38.9637309, -8.5225249,29.33,0.106,13,17,Tarde
101-1322082	38.9018056, -9.0366186	375.17	1.071	13	17	Tarde	101-1322082,38.9018056, -9.0366186,375.17,1.071,13,17,Tarde
110-1176580	39.1322015, -8.9239630	197.91	0.716	13	17	Tarde	110-1176580,39.1322015, -8.9239630,197.91,0.716,13,17,Tarde
110-1176613	39.1618166, -8.7888254	220.87	0.591	13	17	Tarde	110-1176613,39.1618166, -8.7888254,220.87,0.591,13,17,Tarde
110-1176531	38.8675296, -9.0583167	224.9	0.938	13	17	Tarde	110-1176531,38.8675296, -9.0583167,224.9,0.938,13,17,Tarde
168-37249	38.9967022, -8.9727951	35.23	0.023	9	18	Dia	168-37249,38.9967022, -8.9727951,35.23,0.023,9,18,Dia

4.5.4 Análise de Resultados

Na figura 4.4 é apresentada uma rota criada pelo algoritmo, para um determinado dia.

Pela análise da figura 4.4 percebe-se que o algoritmo apresenta a rota criada, com uma separação entre os contratos a realizar durante a manhã e durante a tarde. Cada contrato é apresentado com o seu número de identificação, seguido das coordenadas da morada do cliente, seguido do seu peso, volume e janela horária. É também apresentada a hora prevista para a sua entrega. É também possível observar que existe uma diferença substancial no número de contratos a realizar na parte da manhã e na parte da tarde da rota evidenciada. Isto deve-se ao facto de existir um contrato, na parte da manhã, com um peso muito elevado (667,77 kg), o que resultou em que o peso da mercadoria atingisse o limite máximo mais facilmente. Este facto fez também com que a parte da manhã fosse concluída bastante mais cedo do que a hora normal de término das janelas horárias da manhã (13h).

No fim da lista de contratos respetivos a cada parte do dia, são apresentados o peso e volume totais. É ainda apresentado o cálculo da distância total percorrida, incluindo as viagens desde e até o armazém.

Percebe-se que o formato do output facilita a análise da informação da rota por parte do utilizador do algoritmo. Os valores do peso, volume e distância totais permitem também realizar uma análise comparativa com as rotas criadas pelo gestor de tráfego. Esta análise será realizada na secção 4.6.

Manhã

Cliente 101-1328378 (38.7913138, -9.1223052),	Peso: 13.99, Volume: 0.1, Janela horária: 9-13, Hora de visita:	09:07:49
Cliente 102-1431470 (38.7913138, -9.1223052),	Peso: 667.77, Volume: 1.927, Janela horária: 9-13, Hora de visita:	09:30:29
Cliente 102-1433495 (38.7916763, -9.1200594),	Peso: 212.42, Volume: 0.716, Janela horária: 9-13, Hora de visita:	09:41:26
Cliente 101-1329963 (38.7805269, -9.117779),	Peso: 38.14, Volume: 0.128, Janela horária: 8-12, Hora de visita:	09:53:03
Cliente 110-1177669 (38.776658, -9.1028226),	Peso: 60.69, Volume: 0.129, Janela horária: 9-13, Hora de visita:	10:04:33
Cliente 102-1430522 (38.7392859, -9.14986),	Peso: 4.71, Volume: 0.154, Janela horária: 9-13, Hora de visita:	10:28:34
Peso Total: 997.72, Volume Total: 3.154, Distância Total: 26.44		

Tarde

Cliente 102-1433482 (38.7910428, -9.1198988),	Peso: 198.08, Volume: 0.467, Janela horária: 13-17, Hora de visita:	14:14:19
Cliente 102-1433756 (38.7792142, -9.1186511),	Peso: 39.88, Volume: 0.098, Janela horária: 13-17, Hora de visita:	14:24:52
Cliente 102-1433730 (38.7706795, -9.104893),	Peso: 73.51, Volume: 0.516, Janela horária: 13-17, Hora de visita:	14:34:47
Cliente 101-1330014 (38.7456657, -9.1129369),	Peso: 77.21, Volume: 0.559, Janela horária: 12-16, Hora de visita:	14:48:41
Cliente 101-1329937 (38.7396565, -9.1064546),	Peso: 117.84, Volume: 1.441, Janela horária: 12-16, Hora de visita:	15:00:21
Cliente 101-1330068 (38.7257821, -9.11719),	Peso: 33.09, Volume: 0.147, Janela horária: 12-16, Hora de visita:	15:12:11
Cliente 101-1329790 (38.7269991, -9.1213397),	Peso: 107.77, Volume: 1.048, Janela horária: 12-16, Hora de visita:	15:21:47
Cliente 102-1433978 (38.7272121, -9.12121),	Peso: 44.2, Volume: 0.201, Janela horária: 13-17, Hora de visita:	15:27:37
Cliente 101-1329818 (38.7300365, -9.127831),	Peso: 9.83, Volume: 0.064, Janela horária: 12-16, Hora de visita:	15:34:47
Cliente 102-1433703 (38.7300365, -9.127831),	Peso: 98.81, Volume: 0.697, Janela horária: 13-17, Hora de visita:	15:40:23
Cliente 101-1328754 (38.7287999, -9.1299423),	Peso: 10.85, Volume: 0.033, Janela horária: 13-17, Hora de visita:	15:45:16
Cliente 102-1433542 (38.7340132, -9.1312587),	Peso: 50.58, Volume: 0.158, Janela horária: 13-17, Hora de visita:	15:53:51
Cliente 110-1177652 (38.7336276, -9.1335738),	Peso: 59.38, Volume: 0.234, Janela horária: 13-17, Hora de visita:	16:02:27
Cliente 102-1433945 (38.7306553, -9.1328638),	Peso: 21.77, Volume: 0.033, Janela horária: 13-17, Hora de visita:	16:10:55
Cliente 102-1434100 (38.7394089, -9.1309069),	Peso: 1.66, Volume: 0.065, Janela horária: 13-17, Hora de visita:	16:20:17
Cliente 102-1433664 (38.718979, -9.1273235),	Peso: 38.97, Volume: 0.195, Janela horária: 13-17, Hora de visita:	16:33:26
Cliente 102-1433900 (38.7239019, -9.1501704),	Peso: 9.56, Volume: 0.031, Janela horária: 13-17, Hora de visita:	16:51:24
Peso Total: 992.99, Volume Total: 5.987, Distância Total: 40.95		

Figura 4.4 - Exemplo de uma rota criada pelo algoritmo

4.6 Comparação de resultados

4.6.1 Comparação das rotas de entrega

Para realizar uma comparação relevante, o algoritmo foi aplicado a um mês de planeamento de rotas de entrega (26 dias de trabalho) e apenas aos contratos referentes ao armazém de distribuição do Prior Velho (PV). Na tabela D.1 do anexo D apresenta-se um resumo das rotas, para cada dia em que o algoritmo foi aplicado. É possível observar o número de contratos a entregar no dia, o número de rotas criadas pelo algoritmo, a distância total percorrida pelas carrinhas e a média do número de contratos, peso, volume e distância por rota. Também no anexo D, na tabela D.2 é apresentado um resumo semelhante, para as rotas criadas pelo gestor de tráfego para o mesmo mês de trabalho.

De modo a comparar as rotas criadas pelo algoritmo, com as rotas criadas pelo gestor de tráfego, foram utilizadas algumas métricas, nomeadamente:

- Número de rotas criadas;
- Distância total percorrida;
- Percentagem de ocupação das carrinhas (Peso e Volume).

Nas tabelas 4.8 e 4.9 são apresentados os valores relevantes para as métricas consideradas, sobre as rotas criadas pelo algoritmo e pelo gestor de tráfego, respetivamente.

Tabela 4.8 - Estatísticas das rotas criadas pelo algoritmo

	Nº de contratos	Nº de rotas	Distância percorrida (KM)	% Peso	% Volume
Total	10097	704	127115,09	-	-
Média	388,3	27,1	4889,04	86,68%	68,16%

Tabela 4.9 - Estatísticas das rotas criadas pelo gestor de tráfego

	Nº de contratos	Nº de rotas	Distância percorrida (KM)	% Peso	% Volume
Total	10097	768	130757,69		
Média	388,3	29,5	5029,14	79,22%	62,32%

Pela análise realizada às tabelas 4.8, constata-se que a utilização do algoritmo resultou num número de rotas criadas mais reduzido, sendo que em apenas um dos dias, o número de rotas criadas foi igual (Anexo D). Em todos os outros dias analisados, o algoritmo criou pelo menos uma rota a menos, comparado com o gestor de tráfego. Este facto revela-se como um fator positivo para o algoritmo, pois a redução do número de rotas diárias a realizar, permite diminuir o número de viaturas a utilizar, bem como os seus custos associados, resultando assim numa utilização mais eficiente dos recursos.

Em relação à distância percorrida, as rotas criadas pelo algoritmo percorrem uma distância média de 4889 quilómetros e para as rotas criadas pelo gestor de tráfego este valor é de 5029 quilómetros. Nesta métrica, o algoritmo resultou numa menor distância percorrida em 21 dos 26 dias analisados (80%). Apesar de os resultados serem positivos, isto revela que as rotas criadas pelo algoritmo poderão ser melhoradas de modo a produzirem uma distância percorrida total mais reduzida.

Sobre a ocupação das carrinhas, a sua utilização resultou no aumento de cerca de sete pontos percentuais para o peso e seis pontos percentuais para o volume. Isto confirma a utilização mais eficiente dos recursos nas rotas criadas pelo algoritmo.

Sobre esta métrica, é ainda importante avaliar o número de rotas em que o peso da mercadoria transportada, excede o limite legal. Cerca de 29% das rotas criadas pelo gestor de tráfego resultam em as carrinhas com um peso superior a 1000 quilogramas. Este facto representa um risco para empresa, pois pode ser multada, se ocorrer uma fiscalização a carrinhas com peso superior ao permitido. Em relação às rotas criadas pelo algoritmo, este risco é limitado por ser imposto o limite de peso. Nestas,

apenas 4% estão com peso superior ao permitido. Este valor não é nulo pois existem alguns contratos que individualmente, apresentam um peso superior a 1000 quilogramas. O algoritmo não foi programado para dividir estes contratos por mais que uma carrinha de entrega, pois teria de ser feita uma análise à mercadoria correspondente a esses contratos e feita uma divisão das embalagens de acordo com o seu peso, o que aumentaria a sua complexidade.

4.6.2 Análise de custos

Com as métricas analisadas na subsecção 4.6.1, procedeu-se a uma análise dos ganhos monetários com a utilização do algoritmo.

O aluguer de uma carrinha de entrega representa um custo mensal para a empresa de 1300 €. Como na subsecção anterior a aplicação do algoritmo resultou, consistentemente, na redução do número de rotas criadas para os vários dias (em média o número de rotas criadas pelo algoritmo foi inferior em 2,4 rotas), a empresa poderia reduzir o número de carrinhas alugadas em 2, significando uma redução mensal de custos de 2600 €. Considerando a frota da empresa esta redução do número de carrinhas representa uma redução de 4% nos custos associados ao seu aluguer.

O facto de necessitar de menos carrinhas para realizar as entregas, significa também que a empresa reduziria os custos associados à subcontratação das equipas de entrega.

As rotas propostas pelo algoritmo percorrem por dia, em média, cerca de 140 quilómetros menos do que as criadas pelo gestor de tráfego. Considerando o consumo médio de 6,1 litros, por cada 100 quilómetros percorridos, o algoritmo permitiria uma redução diária de cerca de 8,5 litros de gasolina. Esta redução no consumo de gasolina representaria uma redução de custos mensais de cerca de 375 €.

Considerando que a empresa tem o objetivo de substituir a sua frota por carrinhas elétricas, num futuro próximo, é também importante avaliar os ganhos relativos a este tipo de carrinhas. Para carrinhas elétricas o consumo médio situa-se nos 31,6 kWh por cada 100 quilómetros percorridos. Considerando o preço médio do kWh na rede pública de 0,38 €, estima-se uma redução do custo mensal para a empresa em cerca de 437 €.

Por cada carrinha que transporta um peso de mercadoria superior ao limite de 1000 quilogramas, a empresa corre o risco de ser multada numa coima entre 500 € e 1500 €. Como referido anteriormente em 4.6.1, este risco é reduzido pelo facto de o algoritmo criar, maioritariamente, rotas com peso total inferior ao limite.

A utilização do algoritmo permite ainda anular o tempo despendido pelo gestor de tráfego na criação das rotas de entrega. Este facto permitiria à empresa realizar uma gestão dos seus recursos humanos mais eficiente.

Pelos resultados apresentados, podemos afirmar que a utilização do algoritmo desenvolvido seria positiva para a empresa, pois cumpre todos os objetivos e requisitos definidos e permite também reduzir custos.

CONCLUSÃO

Neste capítulo final, são expostas as conclusões, as limitações e as recomendações para trabalhos futuros no âmbito da utilização de algoritmos de criação de rotas. Inicialmente, são apresentadas as conclusões finais, abordando o processo seguido na realização do trabalho e os resultados alcançados. Em seguida, surgem as limitações encontradas durante o desenvolvimento do trabalho.

Finalmente são propostas recomendações para desenvolvimento futuro de trabalhos sobre o tema abordado, focando-se em aspectos considerados relevantes e identificados ao longo do desenvolvimento desta dissertação.

5.1 Conclusões sobre o trabalho desenvolvido

Esta dissertação foi motivada por um problema concreto enfrentado pela empresa TNB no planejamento das rotas de entrega de mercadoria. A motivação para o estudo surgiu da observação dos procedimentos ineficientes da empresa em relação à criação de rotas de entrega. O processo atual, realizado pelos gestores de tráfego, consome uma quantidade significativa de recursos e tempo. Definiu-se então, o objetivo de criar um algoritmo capaz de automatizar a criação de rotas, reduzindo, assim, o tempo despendido e aumentando a eficiência operacional.

A primeira etapa do estudo envolveu uma análise detalhada do processo de criação de rotas, que atualmente é realizado pelos gestores de tráfego. Essa análise permitiu a identificação de padrões, tendências e áreas que necessitavam de melhorias.

Através do acompanhamento das rotas de entrega, foi possível identificar vários fatores que influenciam o valor do tempo de entrega, como o peso e o volume das mercadorias. Foi ainda possível recolher dados sobre os tempos de entrega, o peso, o volume, e o número de caixas dos contratos entregues. Utilizando a análise de regressão linear múltipla aplicada aos dados recolhidos, foi desenvolvido um modelo que estima o tempo de entrega de cada contrato, proporcionando uma base sólida para a automatização do processo.

O algoritmo desenvolvido demonstrou ser eficaz na criação de rotas, considerando as restrições impostas tais como janelas horárias e limites de peso e volume das carrinhas. A aplicação do algoritmo em cenários reais revelou uma melhoria apreciável em termos da redução da distância percorrida pelas carrinhas. Comparando-se as rotas criadas pelo algoritmo com as rotas criadas pelo gestor de tráfego, observou-se que a aplicação do algoritmo não diminuiu apenas a carga de trabalho dos gestores de tráfego, como também resultou em rotas mais eficientes.

Além disso, o algoritmo contribuiu para um aumento no número de entregas realizadas dentro das janelas horárias acordadas com os clientes o que permite um menor número de reagendamentos.

Este estudo reafirmou a importância da combinação entre tecnologia e conhecimento humano no aumento da eficiência de processos logísticos complexos. Embora a automatização traga claras vantagens em termos de eficiência, a experiência e o conhecimento dos trabalhadores continuam a ser um fator diferencial importante. A integração de ambas as abordagens pode resultar em soluções que não apenas tornam o processo mais eficiente, mas também garantam flexibilidade e adaptabilidade às variáveis dinâmicas do ambiente de operação. Assim, considera-se que este trabalho alcançou os seus objetivos principais e também gerou informações valiosas sobre as possibilidades e os desafios da automatização na logística, oferecendo uma base sólida para futuros desenvolvimentos na área.

5.2 Limitações

Uma das limitações identificadas neste estudo é o facto de o algoritmo não ter sido efetivamente testado em contexto real. A sua aplicação efetiva a um dia de rota poderia revelar problemas inesperados, que afetariam a qualidade das rotas realizadas. Isto permitiria identificar questões a acrescentar ao algoritmo de forma a torná-lo mais adequado ao contexto real das rotas de entrega.

O facto de o algoritmo não considerar a possibilidade de uma entrega não ser realizada (por ausência do cliente ou as características da morada não o permitirem), pode resultar em dificuldades em cumprir a rota definida porque o tempo de entrega, nestes casos, teria de ser substituído.

5.3 Proposta para trabalhos futuros

Na realização deste estudo foram identificadas diversas limitações e oportunidades de melhoria. Apesar de ser uma área já muito estudada, cada caso estudado apresenta características muito específicas que representam oportunidades de desenvolvimentos futuros.

Começando pela estimação do tempo de entrega, a abordagem levada a cabo neste estudo poderá ser aperfeiçoada em trabalhos futuros. Como observado no acompanhamento das rotas, o tempo de entrega depende de vários fatores que foram desconsiderados na análise de regressão linear múltipla realizada nomeadamente o andar da casa do cliente, a experiência dos elementos da equipa de

transporte, a existência de elevador no prédio e os acessos a casa do cliente. A inclusão destes fatores pode resultar num tempo de entrega previsto mais preciso para o problema em questão.

Outra área promissora para pesquisas futuras, envolve a incorporação de variáveis externas ao algoritmo desenvolvido, tais como condições de tráfego em tempo real, características específicas dos locais de entrega (por exemplo, acessibilidade e presença de elevadores) e fatores meteorológicos. A inclusão dessas variáveis tem o potencial de aumentar significativamente a precisão das estimativas do tempo de viagem, resultando numa melhor adaptação do algoritmo ao contexto real das rotas de entrega. Estudos futuros podem explorar a aplicação de modelos estocásticos que considerem estas variáveis, proporcionando soluções mais robustas e adaptáveis às condições operacionais.

Em relação às rotas criadas, faz sentido explorar a possibilidade de utilizar algoritmos de busca heurística ou programação linear, para as tornar mais eficientes. Isto permitiria tornar o algoritmo ainda mais robusto e adequado às necessidades da empresa.

Este estudo foi realizado para solucionar um problema concreto da TNB e representa uma abordagem inicial a um problema que tem inúmeras componentes que se podem alterar no futuro de acordo com as necessidades da empresa. A continuação deste trabalho poderá validar os resultados apresentados e a sua adaptação a outros problemas poderá complementar o estudo nesta área.

REFERÊNCIAS BIBLIOGRÁFICAS

- Applegate, D. L. (Ed.). (2006). *The traveling salesman problem: A computational study*. Princeton University Press.
- Baldacci, R., Mingozzi, A., & Roberti, R. (2012). Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints. *European journal of operational research*, 218(1), 1–6. <https://doi.org/10.1016/j.ejor.2011.07.037>
- Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed). Wiley.
- Bartholdi, J. J., & Hackman, S. T. (2008). *Warehouse & Distribution Science: Release 0.89* (p. 13). Atlanta: Supply Chain and Logistics Institute. Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley.
- Bengio, Y., Lodi, A., & Prouvost, A. (2021). Machine learning for combinatorial optimization: A methodological tour d’horizon. *European journal of operational research*, 290(2), 405–421. <https://doi.org/10.1016/j.ejor.2020.07.063>
- Berg, J. P. van D., & Zijm, W. H. M. (1999). Models for warehouse management: Classification and examples. *International journal of production economics*, 59(1–3), 519–528. [https://doi.org/10.1016/S0925-5273\(98\)00114-5](https://doi.org/10.1016/S0925-5273(98)00114-5)
- Bertsimas, D. J., & Simchi-Levi, D. (1996). A new generation of vehicle routing research: Robust algorithms, addressing uncertainty. *Operations research*, 44(2), 286–304. <https://doi.org/10.1287/opre.44.2.286>
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM computing surveys*, 35(3), 268–308. <https://doi.org/10.1145/937503.937505>
- Bräysy, O., & Gendreau, M. (2005). Vehicle routing problem with time windows, Part I: Route construction and local search algorithms. *Transportation science*, 39(1), 104–118. <https://doi.org/10.1287/trsc.1030.0056>
- Breiman, L. (2001). Random Forests. *Machine learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees* (1.^a ed.). Routledge. <https://doi.org/10.1201/9781315139470>
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287. <https://doi.org/10.2307/1911963>
- Carvalho, H., Azevedo, S. G., & Cruz-Machado, V. (2012). Agile and resilient approaches to supply chain management: Influence on performance and competitiveness. *Logistics Research*, 4(1–2), 49–62. <https://doi.org/10.1007/s12159-012-0064-2>
- Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example* (5. ed). Wiley.

- Christopher, M. (2016). *Logistics and supply chain management: Logistics & supply chain management*. Pearson UK.
- Chvátal, V., Cook, W., Dantzig, G. B., Fulkerson, D. R., & Johnson, S. M. (2010). Solution of a large-scale traveling-salesman problem. Em M. Jünger, T. M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, & L. A. Wolsey (Eds.), *50 years of integer programming 1958-2008* (pp. 7–28). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68279-0_1
- Cohen. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences* (3.^a ed.). Routledge. <https://doi.org/10.4324/9780203774441>
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18. <https://doi.org/10.1080/00401706.1977.10489493>
- Dantzig, G. B., & Ramser, J. H. (1959). The truck dispatching problem. *Management science*, 6(1), 80–91. <https://doi.org/10.1287/mnsc.6.1.80>
- De Koster, R. B. M., Le-Duc, T., & Zaerpour, N. (2012). Determining the number of zones in a pick-and-sort order picking system. *International journal of production research*, 50(3), 757–771. <https://doi.org/10.1080/00207543.2010.543941>
- De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European journal of operational research*, 182(2), 481–501. <https://doi.org/10.1016/j.ejor.2006.07.009>
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1), 95–104. <https://doi.org/10.1080/01969727408546059>
- Durbin, J., & Watson, G. S. (1992). Testing for serial correlation in least squares regression. II. Em S. Kotz & N. L. Johnson (Eds.), *breakthroughs in statistics* (pp. 260–266). Springer New York. https://doi.org/10.1007/978-1-4612-4380-9_21
- Ester, M., Kriegel, H.-P., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD*, 96(34), 226–331.
- Faber, N., de Koster, M. B. M., & Smidts, A. (2013). Organizing warehouse management. *International journal of operations & production management*, 33(9), 1230–1256. <https://doi.org/10.1108/IJOPM-12-2011-0471>
- Gendreau, M., Hertz, A., & Laporte, G. (1994). A tabu search heuristic for the vehicle routing problem. *Management science*, 40(10), 1276–1290. <https://doi.org/10.1287/mnsc.40.10.1276>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Goetschalckx, M., & Ashayeri, J. (1989). Classification and design of order picking. *Logistics world*, 2(2), 99–106. <https://doi.org/10.1108/eb007469>
- Golden, B., Raghavan, S., & Wasil, E. (Eds.). (2008). *The vehicle routing problem: Latest advances and new challenges* (Vol. 43). Springer US. <https://doi.org/10.1007/978-0-387-77778-8>
- Gu, J., Goetschalckx, M., & McGinnis, L. F. (2007). Research on warehouse operation: A comprehensive review. *European journal of operational research*, 177(1), 1–21. <https://doi.org/10.1016/j.ejor.2006.02.025>
- Gunasekaran, A., & Ngai, E. W. T. (2004). Information systems in supply chain integration and management. *European journal of operational research*, 159(2), 269–295. <https://doi.org/10.1016/j.ejor.2003.08.016>

- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed). Springer.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Joliffe, I. (2002). Principal component analysis for special types of data. In *principal component analysis* (pp. 338–372). Springer-Verlag. https://doi.org/10.1007/0-387-22440-8_13
- Kriegel, H., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *WIREs data mining and knowledge discovery*, 1(3), 231–240. <https://doi.org/10.1002/widm.30>
- Laporte, G. (2009). Fifty years of vehicle routing. *Transportation science*, 43(4), 408–416. <https://doi.org/10.1287/trsc.1090.0301>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22.
- Lin, B., Ghaddar, B., & Nathwani, J. (2022). Deep reinforcement learning for the electric vehicle routing problem with time windows. *IEEE transactions on intelligent transportation systems*, 23(8), 11528–11538. <https://doi.org/10.1109/tits.2021.3105232>
- Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281–297.
- Mitchell, S., O’Sullivan, M., & Dunning, I. (2011). *PuLP: A linear programming toolkit for Python*.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (com Krueger, D.). (2007). *Student solutions manual to accompany Introduction to linear regression analysis, fourth edition, Douglas C. Montgomery; Elizabeth A. Peck; G. Geoffrey Vining* (Fourth edition). Wiley-interscience.
- Mundfrom, D. J., Perrett, J. J., Schaffer, J., Piccone, A., & Roozeboom, M. (2006). Bonferroni adjustments in tests for regression coefficients. *Multiple linear regression viewpoints*, 32(1), 1–6.
- Neter, J., Wasserman, W., & Kutner, M. H. (1983). *Applied linear regression models*. R.D. Irwin.
- O’Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Oliphant, T. E. (2007). Python for scientific computing. *Computing in science & engineering*, 9(3), 10–20. <https://doi.org/10.1109/MCSE.2007.58>
- Pillac, V., Gendreau, M., Guéret, C., & Medaglia, A. L. (2013). A review of dynamic vehicle routing problems. *European journal of operational research*, 225(1), 1–11. <https://doi.org/10.1016/j.ejor.2012.08.015>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81–106. <https://doi.org/10.1007/BF00116251>
- Richards, G. (2017). *Warehouse management: A complete guide to improving efficiency and minimizing costs in the modern warehouse*. Kogan Page Publishers.

- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. Wiley.
- Rouwenhorst, B., Reuter, B., Stockrahm, V., Van Houtum, G. J., Mantel, R. J., & Zijm, W. H. M. (2000). Warehouse design and control: Framework and literature review. *European journal of operational Research*, 122(3), 515–533. [https://doi.org/10.1016/S0377-2217\(99\)00020-X](https://doi.org/10.1016/S0377-2217(99)00020-X)
- Rushton, A., Croucher, P., & Baker, P. (2022). *The handbook of logistics and distribution management: Understanding the supply chain*. Kogan Page Publishers.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2014). *Introduction to data mining* (New internat. edition). Pearson.
- Tan, S.-Y., & Yeh, W.-C. (2021). The vehicle routing problem: State-of-the-art classification and review. *Applied sciences*, 11(21), 10295. <https://doi.org/10.3390/app112110295>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the royal statistical society series B: Statistical methodology*, 63(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>
- Toth, P., & Vigo, D. (2002). Models, relaxations and exact approaches for the capacitated vehicle routing problem. *Discrete applied mathematics*, 123(1–3), 487–512. [https://doi.org/10.1016/S0166-218X\(01\)00351-1](https://doi.org/10.1016/S0166-218X(01)00351-1)
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Yu, J. J. Q., Yu, W., & Gu, J. (2019). Online vehicle routing with neural combinatorial optimization and deep reinforcement learning. *IEEE transactions on intelligent transportation systems*, 20(10), 3806–3817. <https://doi.org/10.1109/tits.2019.2909109>

ANEXOS

ANEXO A

Tabela A.1 - Dados recolhidos, no acompanhamento das rotas, sobre os contratos entregues

Nº contrato	Ordem	Peso (Kg)	Volume (m3)	Nº caixas	T. Entrega (min)
154-10909	6	0,03	0,17	1	2
101-1322245	5	0,46	0,003	1	2
110-1176838	16	6,73	0,036	1	3
102-1424877	12	8,6	0,038	1	2
102-1422850	10	11,33	0,119	1	3
101-1323327	3	12,66	0,024	1	6
168-37382	13	27,13	0,219	1	2
101-1323308	3	27,98	0,052	1	1
168-37798	10	31	0,361	1	4
414-152	9	32	0,18	1	3
414-148	12	32	0,18	1	3
101-1321367	12	35	0,102	1	4
168-37339	10	35,5	0,093	1	4
168-37459	5	43	0,115	1	2
168-37714	6	43	0,115	1	4
168-37193	8	49	0,088	1	3
168-37016	10	49	0,167	1	5
168-37084	8	105	1,852	1	7
102-1424082	4	4,1	0,008	2	1
101-1322069	7	10,94	0,041	2	1
102-1423055	6	15,86	0,1	2	5
101-1321402	9	22,84	0,088	2	3
102-1422525	5	39,88	0,098	2	3
102-1422655	3	41,06	0,079	2	3
168-37369	8	41,96	0,068	2	7
110-1177120	8	58	0,424	2	3
102-1427431	11	59,98	0,213	2	5
102-1422141	8	61,6	0,698	2	7
101-1323087	18	62,65	0,955	2	3

Tabela A.1 - Dados recolhidos, no acompanhamento das rotas, sobre os contratos entregues (cont.)

Nº contrato	Ordem	Peso (Kg)	Volume (m3)	Nº caixas	T. Entrega (min)
168-37262	9	68,3	0,158	2	4
102-1424092	1	9,1	0,031	3	6
102-1421376	12	11,52	0,05	3	2
101-1324174	15	24,52	0,057	3	4
110-1176912	8	80,71	0,169	3	4
101-1323527	9	84,1	0,163	3	6
102-1423495	15	84,21	0,201	3	5
102-1424379	17	89,94	0,165	3	4
102-1424684	15	100,14	1,033	3	7
101-1324109	6	25,93	0,159	4	4
102-1422268	3	32,72	0,156	4	1
102-1424741	19	52,73	0,294	4	5
102-1424923	18	56,25	0,189	4	5
101-1323346	6	111,33	1,147	4	7
102-1422155	9	116,6	0,229	4	5
401-3709	12	159	2,086	4	6
102-1422317	1	8,56	0,039	5	6
102-1424060	2	11,91	0,051	5	2
102-1421229	2	20,05	0,115	5	2
102-1424592	1	47,23	0,134	5	2
102-1424376	17	78,32	0,332	5	6
101-1323830	7	93,2	0,487	5	7
101-1321588	5	0,52	0,003	6	5
102-1421395	11	17,19	0,148	6	3
102-1421755	4	44,8	0,179	6	3
102-1424022	6	111,82	0,235	6	5
101-1322162	13	121,78	0,381	6	9
101-1325559	4	125,3	0,327	6	9
102-1425047	1	161,3	2,241	6	5
102-1423554	2	10,2	0,101	7	1
102-1423084	2	17,67	0,052	7	4
101-1323823	1	92,63	0,181	7	6
102-1424606	11	102,04	0,303	7	6
102-1423592	4	153,67	0,428	7	10
102-1422696	3	98,63	0,342	8	8
102-1425435	7	93,45	0,313	9	6
102-1422209	8	8,47	0,111	10	5

Tabela A.1 - Dados recolhidos, no acompanhamento das rotas, sobre os contratos entregues (cont.)

Nº contrato	Ordem	Peso (Kg)	Volume (m3)	Nº caixas	T. Entrega (min)
102-1423460	13	60,42	0,146	10	2
101-1324675	1	95,59	0,243	10	6
102-1422010	12	115,43	0,323	10	5
101-1322971	14	127,07	0,399	10	6
102-1423093	5	189,53	0,521	10	9
102-1422018	4	25,42	0,054	11	5
110-1176656	1	41,06	0,27	12	4
102-1425193	14	124,39	0,235	12	7
102-1419790	7	287,4	0,558	12	9
101-1323981	14	28,87	0,061	13	5
102-1422128	10	80,11	0,18	13	5
101-1323233	19	113,96	0,264	13	8
102-1424583	10	126,51	0,755	13	7
102-1424109	7	148,24	0,496	13	9
101-1322647	10	24,25	0,137	14	1
102-1424041	9	67,81	0,324	14	5
101-1323210	18	38,75	0,2	17	6
101-1324169	10	89,32	0,658	17	5
110-1176940	5	41,56	0,301	18	2
101-1322454	1	105,55	0,454	18	9
102-1421418	3	96,24	0,525	20	6
101-1322447	7	254,64	0,662	21	11
102-1423288	8	197,14	0,767	23	7
102-1424761	2	241,21	1,435	23	10
102-1424206	5	94,69	0,597	24	7
102-1421651	3	135,42	0,38	24	7
102-1414938	13	296,23	0,769	24	13
101-1323035	12	202,35	0,436	25	10
101-1322270	6	290,86	1,792	25	13
101-1324084	11	36,32	0,21	28	3
102-1426713	2	272,07	0,716	30	10
101-1325541	5	322,63	0,756	30	11
102-1417702	7	310,32	1,062	31	11
101-1322879	10	198,3	1,88	33	12
102-1420802	2	200,7	0,611	33	11
102-1422769	11	169,43	0,348	34	8
102-1421752	2	83,4	0,389	36	7

Tabela A.1 - Dados recolhidos, no acompanhamento das rotas, sobre os contratos entregues (cont.)

Nº contrato	Ordem	Peso (Kg)	Volume (m3)	Nº caixas	T. Entrega (min)
102-1424045	4	133,89	0,368	38	10
102-1423581	9	50,98	0,366	42	5
102-1424242	4	76,58	0,223	50	3
101-1322006	4	132,71	0,447	53	8
102-1420972	17	137,6	0,537	74	7
101-1316824	7	537,58	1,817	88	22
101-1321842	6	745,17	2,07	113	18
101-1323409	5	365,58	1,171	57	37
101-1325338	3	222,42	0,57	46	19
102-1421575	19	273,73	1,751	31	4
442-1084	11	164	3,343	4	22
102-1424822	13	31,08	0,073	2	18
168-37155	9	65,5	0,226	2	14
102-1421864	11	160,01	0,573	26	14
110-1176733	4	152,55	0,592	26	15
102-1420659	16	111,86	0,217	25	13
102-1419620	6	326,78	1,247	15	19
110-1176887	11	39,85	0,248	11	11
101-1322380	7	59,66	0,195	11	12
102-1423306	11	197,83	0,417	9	4
102-1424809	16	156,4	0,275	7	2
102-1419627	14	100,88	0,38	5	13
102-1421600	2	28,2	0,243	4	10
168-37572	16	28,71	0,242	3	10
110-1177154	9	84,1	0,163	3	12
102-1425275	12	58,34	0,345	12	10
102-1423273	15	33,09	0,147	6	8
102-1423819	14	25,31	0,112	5	9
102-1424388	13	105,7	0,93	2	10
102-1421862	13	75,81	0,198	6	9
102-1414799	3	115,54	0,335	5	3
101-1321323	1	49,3	0,26	3	8
101-1323236	8	144,6	0,293	7	4

ANEXO B

Tabela B.1 - Resíduos do modelo de regressão

Observação	Y previsto	Resíduos	Resíduos padronizados
1	21,04	0,96	0,25
2	9,92	-2,92	-0,77
3	8,85	-0,85	-0,22
4	7,17	-4,17	-1,10
5	6,63	-1,63	-0,43
6	8,27	1,73	0,46
7	7,11	-0,11	-0,03
8	8,97	-0,97	-0,26
9	11,98	0,02	0,00
10	9,94	1,06	0,28
11	13,05	-2,05	-0,54
12	12,78	-1,78	-0,47
13	11,57	-1,57	-0,41
14	5,64	-2,64	-0,70
15	9,55	0,45	0,12
16	13,57	-0,57	-0,15
17	12,12	0,88	0,23
18	7,34	-0,34	-0,09
19	7,86	-0,86	-0,23
20	9,85	-2,85	-0,75
21	11,81	-1,81	-0,48
22	10,87	0,13	0,03
23	7,11	-1,11	-0,29
24	5,51	-3,51	-0,93
25	7,11	1,89	0,50
26	5,40	0,60	0,16
27	7,15	-2,15	-0,57
28	4,80	-3,80	-1,00
29	6,04	-1,04	-0,27
30	7,04	0,96	0,25
31	4,79	0,21	0,05
32	7,99	-0,99	-0,26
33	6,06	-1,06	-0,28

Tabela B.1 - Resíduos do modelo de regressão (cont.)

Observação	Y previsto	Resíduos	Resíduos padronizados
34	8,02	0,98	0,26
35	7,13	-0,13	-0,03
36	11,14	-2,14	-0,56
37	5,21	-1,21	-0,32
38	4,53	0,47	0,12
39	7,39	-1,39	-0,37
40	6,99	-1,99	-0,53
41	5,51	-3,51	-0,93
42	4,26	0,74	0,19
43	8,85	0,15	0,04
44	6,30	-0,30	-0,08
45	6,40	-0,40	-0,11
46	6,49	1,51	0,40
47	6,56	-0,56	-0,15
48	7,44	-3,44	-0,91
49	7,81	2,19	0,58
50	4,12	-3,12	-0,82
51	4,21	-0,21	-0,06
52	6,04	-0,04	-0,01
53	7,10	1,90	0,50
54	5,81	3,19	0,84
55	4,42	-1,42	-0,37
56	6,58	-1,58	-0,42
57	3,76	1,24	0,33
58	6,99	2,01	0,53
59	4,99	-1,99	-0,53
60	10,78	-5,78	-1,53
61	6,09	-0,09	-0,02
62	6,54	0,46	0,12
63	6,75	-3,75	-0,99
64	4,30	-2,30	-0,61
65	4,02	-2,02	-0,53
66	4,91	-2,91	-0,77
67	3,91	2,09	0,55
68	5,46	-0,46	-0,12
69	5,36	-0,36	-0,09
70	10,54	-4,54	-1,20
71	6,64	-1,64	-0,43
72	7,94	-0,94	-0,25
73	4,51	-0,51	-0,13
74	4,62	-3,62	-0,96
75	6,01	-2,01	-0,53
76	7,59	-0,59	-0,16

Tabela B.1 - Resíduos do modelo de regressão (cont.)

Observação	Y previsto	Resíduos	Resíduos padronizados
77	5,92	-0,92	-0,24
78	4,38	-0,38	-0,10
79	4,06	-2,06	-0,54
80	5,79	0,21	0,06
81	5,72	-1,72	-0,45
82	5,09	2,91	0,77
83	3,85	2,15	0,57
84	6,64	-3,64	-0,96
85	5,33	-0,33	-0,09
86	5,41	-1,41	-0,37
87	6,10	0,90	0,24
88	5,59	-2,59	-0,68
89	4,67	2,33	0,61
90	4,30	-1,30	-0,34
91	3,94	-2,94	-0,78
92	4,13	0,87	0,23
93	3,70	-2,70	-0,71
94	4,64	-1,64	-0,43
95	4,62	-1,62	-0,43
96	3,91	-0,91	-0,24
97	4,61	-2,61	-0,69
98	4,65	-1,65	-0,43
99	3,91	-1,91	-0,50
100	4,89	-0,89	-0,23
101	4,98	0,02	0,01
102	4,59	-0,59	-0,16
103	4,56	-0,56	-0,15
104	4,61	-1,61	-0,43
105	8,85	-1,85	-0,49
106	4,07	-1,07	-0,28
107	4,83	-1,83	-0,48
108	3,86	-1,86	-0,49
109	4,72	-0,72	-0,19
110	4,71	-2,71	-0,71
111	3,60	-1,60	-0,42
112	4,25	-3,25	-0,86
113	3,87	2,13	0,56
114	26,78	-8,78	-2,32
115	15,24	21,76	5,75
116	10,77	8,23	2,17
117	13,47	-9,47	-2,50
118	12,63	9,37	2,47
119	4,50	13,50	3,57

Tabela B.1 - Resíduos do modelo de regressão (cont.)

Observação	Y previsto	Resíduos	Resíduos padronizados
120	5,45	8,55	2,26
121	8,86	5,14	1,36
122	8,65	6,35	1,68
123	7,27	5,73	1,51
124	13,18	5,82	1,54
125	5,23	5,77	1,52
126	5,54	6,46	1,71
127	8,90	-4,90	-1,29
128	7,76	-5,76	-1,52
129	6,62	6,38	1,68
130	4,65	5,35	1,41
131	4,78	5,22	1,38
132	5,79	6,21	1,64
133	5,84	4,16	1,10
134	4,81	3,19	0,84
135	4,54	4,46	1,18
136	7,49	2,51	0,66

ANEXO C

```
1 import googlemaps
2 from datetime import datetime
3
4 class TimeWindow:
5     def __init__(self, start, end):
6         self.start = start
7         self.end = end
8
9     def __str__(self):
10         return f"{self.start} - {self.end}"
11
12 class Client:
13     def __init__(self, client_id, coordinates, weight, volume, time_window, time_slot):
14         self.client_id = client_id
15         self.coordinates = coordinates
16         self.weight = weight
17         self.volume = volume
18         self.time_window = time_window
19         self.time_slot = time_slot
20         self.visit_time = None
21
22     def __str__(self):
23         return f"Client {self.client_id} at {self.coordinates}, Weight: {self.weight}, Volume: {self.volume}, Time window: {self.time_window}"
24
25 max_weight = 1000
26 max_volume = 6
27 starting_point = (38.788093, -9.122211)
28
29 gmaps = googlemaps.Client(key='API KEY')
30
31 distance_cache = {}
32 travel_time_cache = {}
33
34 def get_distance_matrix(client1, client2):
35     if (client1, client2) in distance_cache:
36         return distance_cache[(client1, client2)]
37     lat1, lng1 = client1
38     lat2, lng2 = client2
39     result = gmaps.distance_matrix((lat1, lng1), (lat2, lng2), mode="driving")
40     if result['status'] == 'OK':
41         distance_value = result['rows'][0]['elements'][0].get('distance', {}).get('value')
42         if distance_value is not None:
43             distance_value_km = distance_value / 1000 # Convert to kilometers
44             distance_cache[(client1, client2)] = distance_value_km
45             return distance_value_km
46     return float('inf')
47
```

Figura C.1 - Inicialização do algoritmo desenvolvido

```

48 def get_travel_time(client1, client2):
49     if (client1, client2) in travel_time_cache:
50         return travel_time_cache[(client1, client2)]
51
52     lat1, lng1 = client1
53     lat2, lng2 = client2
54     now = datetime.now()
55     result = gmaps.directions((lat1, lng1), (lat2, lng2), mode="driving", departure_time=now)
56     if result:
57         # Extract the duration in traffic value from the first route
58         duration_value = result[0]['legs'][0]['duration']['value']
59         if duration_value is not None:
60             travel_time_seconds = duration_value # Travel time in seconds
61             travel_time_cache[(client1, client2)] = travel_time_seconds
62             return travel_time_seconds
63
64     return float('inf')
65
66
67 def calculate_distance(client1, client2):
68     return get_distance_matrix(client1, client2)
69
70 def calculate_travel_time(client1, client2):
71     return get_travel_time(client1, client2)
72

```

Figura C.2 - Funções desenvolvidas

```

252 total_client_to_client_distance = 0
253 for idx, (manha_route, tarde_route) in enumerate(combined_routes):
254     print(f"Route {idx + 1}:\n Manhã:")
255     if manha_route[0]:
256         for client in manha_route[0]:
257             visit_time_str = datetime.datetime.fromtimestamp(client.visit_time).strftime('%Y-%m-%d %H:%M:%S')
258             print(f" {client} - Visited at: {visit_time_str}")
259         print(f" Total Weight: {manha_route[1]} KG, Total Volume: {manha_route[2]} M3, Total Distance: {manha_route[3]:.2f} km")
260         print(f" Client-to-Client Distance: {manha_route[4]:.2f} km")
261         total_client_to_client_distance += manha_route[4]
262     else:
263         print(" No Manhã part for this route.")
264
265     print(" Tarde:")
266     if tarde_route[0]:
267         for client in tarde_route[0]:
268             visit_time_str = datetime.datetime.fromtimestamp(client.visit_time).strftime('%Y-%m-%d %H:%M:%S')
269             print(f" {client} - Visited at: {visit_time_str}")
270         print(f" Total Weight: {tarde_route[1]} KG, Total Volume: {tarde_route[2]} M3, Total Distance: {tarde_route[3]:.2f} km")
271         print(f" Client-to-Client Distance: {tarde_route[4]:.2f} km")
272         total_client_to_client_distance += tarde_route[4]
273     else:
274         print(" No Tarde part for this route.")
275
276     if manha_route[0]:
277         return_to_start_distance = calculate_distance(starting_point, manha_route[0][-1].coordinates)
278         print(f" ('Return to starting point', {return_to_start_distance:.2f} km)")
279         if tarde_route[0]:
280             transition_to_tarde_distance = calculate_distance(manha_route[0][-1].coordinates, tarde_route[0][0].coordinates)
281             print(f" ('Transition to Tarde', {transition_to_tarde_distance:.2f} km)")
282     elif tarde_route[0]:
283         return_to_start_distance = calculate_distance(starting_point, tarde_route[0][-1].coordinates)
284         print(f" ('Return to starting point', {return_to_start_distance:.2f} km)")
285
286 print(f"Total Client-to-Client Distance for all routes: {total_client_to_client_distance:.2f} km")

```

Figura C.3 - Últimas linhas do código desenvolvido

ANEXO D

Tabela D.1 - Resumo das rotas criadas pelo algoritmo

Dia	Nº de rotas	Contratos/rota	Peso(Kg)/rota	% (peso)	V(m3)/rota	% (volume)	Distância (Km)	Distância média/rota
1	28	13,5	1747,64	87,38%	7,34	61,19%	5438,99	194,25
2	23	15,0	1770,61	88,53%	8,64	71,99%	4719,13	205,18
3	27	14,0	1815,52	90,78%	7,09	59,10%	4448,89	164,77
4	32	14,5	1730,47	86,52%	7,61	63,45%	5104,66	159,52
5	29	16,0	1811,05	90,55%	9,33	77,72%	6250,03	215,52
6	30	13,5	1597,69	79,88%	7,74	64,47%	5359,95	178,67
7	30	12,9	1658,62	82,93%	9,01	75,08%	5274,38	175,81
8	25	13,4	1678,90	83,95%	7,26	60,49%	3880,72	155,23
9	30	14,3	1759,74	87,99%	7,71	64,25%	5459,72	181,99
10	30	14,4	1772,58	88,63%	9,29	77,43%	5266,90	175,56
11	29	14,6	1684,14	84,21%	8,35	69,58%	4574,62	157,75
12	31	12,3	1512,46	75,62%	7,32	61,00%	5050,00	162,90
13	25	13,6	1680,35	84,02%	9,07	75,58%	4408,85	176,35
14	22	14,7	1736,31	86,82%	9,21	76,72%	4114,89	187,04
15	27	12,6	1728,06	86,40%	8,24	68,68%	5937,04	219,89
16	26	15,8	1874,45	93,72%	8,83	73,58%	4764,17	183,24
17	29	14,9	1689,97	84,50%	8,30	69,13%	5169,62	178,26
18	28	14,1	1695,33	84,77%	7,59	63,24%	5084,80	181,60
19	24	15,3	1827,71	91,39%	8,47	70,57%	4740,56	197,52
20	20	15,2	1710,68	85,53%	7,63	63,55%	3464,07	173,20
21	28	14,1	1844,82	92,24%	7,54	62,82%	6122,04	218,64
22	30	15,3	1781,67	89,08%	9,46	78,84%	5143,87	171,46
23	29	15,1	1764,21	88,21%	8,29	69,10%	5318,67	183,40
24	27	15,8	1797,86	89,89%	8,47	70,62%	4542,67	168,25
25	26	13,2	1609,92	80,50%	7,15	59,54%	4302,00	165,46
26	19	15,7	1793,32	89,67%	7,74	64,51%	3173,85	167,04

Tabela D.2 - Resumo das rotas criadas pelo gestor de tráfego

Dia	Nº de rotas	Contratos/rota	Peso(Kg)/rota	% (peso)	V(m3)/rota	% (volume)	Distância (Km)	Distância média/rota
1	28	13,5	1747,64	87,38%	7,34	61,19%	5534,77	197,67
2	28	12,4	1454,43	72,72%	7,10	59,14%	4609,95	164,64
3	30	12,6	1633,97	81,70%	6,38	53,19%	4421,08	147,37
4	33	14,1	1678,03	83,90%	7,38	61,53%	5689,34	172,40
5	31	14,9	1694,21	84,71%	8,72	72,70%	6340,44	204,53
6	31	13,1	1547,44	77,37%	7,50	62,47%	5383,27	173,65
7	32	12,1	1554,96	77,75%	8,45	70,39%	5605,27	175,16
8	26	12,9	1614,33	80,72%	6,98	58,16%	4190,87	161,19
9	31	13,8	1702,97	85,15%	7,46	62,18%	5473,41	176,56
10	33	13,1	1611,43	80,57%	8,45	70,39%	5507,08	166,88
11	32	13,3	1526,25	76,31%	7,57	63,05%	4985,94	155,81
12	32	11,9	1465,19	73,26%	7,09	59,10%	4994,55	156,08
13	26	13,1	1615,72	80,79%	8,72	72,67%	4390,26	168,86
14	27	12,0	1414,77	70,74%	7,50	62,51%	4397,92	162,89
15	29	11,7	1608,89	80,44%	7,67	63,94%	6113,26	210,80
16	31	13,3	1572,12	78,61%	7,41	61,71%	5222,79	168,48
17	30	14,4	1633,64	81,68%	8,02	66,83%	5429,86	181,00
18	29	13,6	1636,87	81,84%	7,33	61,06%	4952,95	170,79
19	29	12,7	1512,59	75,63%	7,01	58,41%	4694,12	161,87
20	24	12,6	1425,56	71,28%	6,35	52,95%	3863,08	160,96
21	29	13,6	1781,20	89,06%	7,28	60,65%	6132,25	211,46
22	31	14,8	1724,20	86,21%	9,16	76,30%	5297,42	170,88
23	32	13,7	1598,82	79,94%	7,51	62,62%	5152,53	161,02
24	29	14,8	1684,13	84,21%	7,98	66,48%	4766,80	164,37
25	28	12,2	1494,92	74,75%	6,63	55,29%	4356,23	155,58
26	27	11,1	1261,96	63,10%	5,45	45,40%	3252,26	120,45



