UNIVERSIDADE FEDERAL DE ALAGOAS
CENTRO DE TECNOLOGIA – CTEC
ENGENHARIA QUÍMICA

LEONARDO MOTA MARINHO LEMOS

OTIMIZAÇÃO DE UM SISTEMA DE DISTRIBUIÇÃO DE BEBIDAS

Maceió

2021

LEONARDO MOTA MARINHO LEMOS

OTIMIZAÇÃO DE UM SISTEMA DE DISTRIBUIÇÃO DE BEBIDAS

Trabalho de conclusão de curso apresentado ao curso de Engenharia Química da Universidade Federal de Alagoas, como requisito parcial para obtenção do título de Bacharel em Engenharia Química, sob orientação do Professor Dr. João Inácio Soletti.

Maceió

2021

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Valter dos Santos Andrade

L557o Lemos, Leonardo Mota Marinho.
Otimização de um sistema de distribuição de bebidas / Leonardo Mota Marinho Lemos, Maceió – 2021
35 f. : il.

Orientador: João Inácio Soletti.
Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química) – Universidade Federal de Alagoas, Centro de Tecnologia, Maceió, 2021.

Bibliografia: f. 35.

1. Bebidas - Distribuição. 2. Otimização de processo. 3. K-Means, Método. 4. Python (Linguagem de programação de computador). I. Título.

CDU: 66.012

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me permitido oportunidades de crescimento muito grandes durante minha graduação e vida. Agradeço aos meus pais pela confiança e apoio que me dão até hoje em relação às escolhas que fiz e faço e por estarem comigo em todos os momentos.

Também a todos amigos que fiz durante as principais experiências de aprendizado que tive na graduação no MEJ – Movimento Empresa Júnior, um agradecimento à PROTEQ – Empresa Júnior de Engenharia Química e Engenharia Ambiental, espaço que me possibilitou enxergar mais sobre o mercado de trabalho e à FEJEA – Federação das Empresas Juniores do Estado de Alagoas, organização que também me possibilitou ampliar minha visão de futuro.

No mais, agradeço aos meus gestores durante o estágio, Marco Brito e Diego Cândido, que me possibilitaram um troca de experiências e aprendizado também muito enriquecedor para minha jornada profissional.

Gostaria também de agradecer ao meu Orientador Professor João Inácio Soletti, por ter topado me auxiliar neste trabalho desde quando a demanda surgiu como um simples projeto de estágio e que se tornou meu trabalho de conclusão de curso.

Aos amigos que fiz durante a graduação e que irei levar para toda a vida, Alysson, Filipe, Daniel, Marcelo, Mariany, Andreza, Arthur Almeida, Arthur Montoro e tantos outros que de alguma forma contribuíram e foram de extrema importância para toda a trajetória acadêmica.

Por fim, deixo meu agradecimento final para minha companheira Fernanda, que sempre esteve ao meu lado me apoiando e dividindo nossos sonhos e conquistas.

RESUMO

Devido às crescentes necessidades do mercado de produção em se atingir simultaneamente um número cada vez maior de abrangência e escala na distribuição de seus produtos para a sociedade, a otimização de processos vem sempre criando alternativas para que se tenha uma maior redução nos custos da operação e aumento da eficiência de processos. Nesse sentido, o processo logístico nunca foi tão estratégico para o setor industrial. Após os anos 80s, a logística passa a ter realmente um desenvolvimento revolucionário, impulsionado pelas demandas ocasionadas pela Revolução Industrial, pela alteração da economia mundial e pelo grande crescimento de demanda, em que todo o gerenciamento da cadeia de abastecimento que planeja, implementa e controla o fluxo e armazenamento eficiente e econômico de matérias-primas, materiais semiacabados e produtos acabados, são controlados pela área. Nesse contexto, a engenharia se apresenta como peça fundamental para aplicação da otimização de processos envolvendo o desenvolvimento de algoritmos computacionais complexos que possam trazer a solução ótima para o problema. Um dos métodos amplamente utilizados para clusterização de dados envolvendo a distribuição de produtos na engenharia é o Método *K-Means*, modelo que objetiva particionar *n* observações de *k* grupos, sendo cada observação pertencente ao grupo mais próximo da média, que resulta na divisão do espaço de dados em um diagrama de Voronoi. Assim, o objetivo deste trabalho consiste na otimização computacional do sistema de distribuição de bebidas do Centro de Distribuição Direta de uma Cervejaria, utilizando algoritmos de resolução NP-Difícil (problema computacionalmente difícil) através da linguagem de programação *Python,* a fim de entender o melhor cenário de agrupamento de clientes e rotas de entrega para minimizar a distância percorrida diária durante a distribuição de bebidas da empresa.

Palavras-Chave: Otimização de Processos. *K-Means*. Python. Distribuição.

ABSTRACT

Given the growing need of the production market to simultaneously reach an increasing number of scope and scale in the distribution of its products to a society, the optimization of processes comes whenever creating alternatives so that there is a greater reduction in the costs of the operation. and increased process efficiency. In this sense, the logistical process has never been more strategic for the industrial sector. After the 1980s, logistics began to have a revolutionary development, driven by the demands caused by the Industrial Revolution, by the change in the world economy and by the great growth in demand, in which all the supply chain management that plans, implements and controls the efficient and economical flow and storage of raw materials, semi-finished materials and finished products are controlled by the area. In this context, engineering presents itself as a fundamental piece for the application of process optimization involving the development of computational computational algorithms that bring an optimal solution to the problem. One of the methods used for data clustering involving the distribution of products in engineering is the K-Means Method, a model that aims to partition n dissipation of groups, presenting each observation belonging to the group closest to the average, which results in the division of the data on a Voronoi diagram. Thus, the objective of this work consists in the computational optimization of the beverage distribution system of the Direct Distribution Center of a Brewery, using NP-Difficult resolution algorithms (computationally difficult problem) through the Python programming language, in order to understand the best scenario of grouping customers and delivery routes to minimize the daily distance traveled during the company's beverage distribution.

Key words: Process optimization. *K-Means*. Python. Distribution.

LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema geral de aprendizagem da máquina 16
Figura 2 - Exemplo do resultado do algoritmo K-means para diferentes valores de k .. 17
Figura 3 – Metodologia do trabalho para otimização do sistema de distribuição 19
Figura 4 - Processo de Vendas e Distribuição 20
Figura 5 - Posição geográfica dos clientes na cidade de Maceió-AL 22
Figura 6 - Esquema de aplicação do método de trabalho 25
Figura 7 - Clusterização de dados usando K-means 26
Figura 8 - Distribuição dos clientes por dia da semana (cenário antigo) 27
Figura 9 – Distribuição dos clientes por cada dia da semana (cenário antigo) 28
Figura 10 - Distribuição dos clientes em clusters (cenário novo) 28
Figura 11 - Distribuição de clientes por cluster (cenário novo) 29
Figura 12 - Clientes dos vendedores 24 e 9 (cenário antigo) 30
Figura 13 - Clientes dos vendedores 24 e 9 (cenário novo) 30
Figura 14 - Distribuição dos clientes por dia da semana (cenário novo) 31
Figura 15 - Distribuição de clientes por cluster (cenário novo) 31
Figura 16 – Comparação da economia de tempo de deslocamento por dia da semana.. 33
Figura 17 - Comparação do ganho em entregas/frota por dia da semana 33

LISTA DE TABELAS

Tabela 1 - Segmentos de clientes e % do total de nº de clientes 23
Tabela 2 - Nº de clientes visitados por dia da semana..... 23
Tabela 3 - N º de clientes por vendedores por sala de vendas..... 24
Tabela 4 - Comparação da dispersão de tempo de deslocamento entre os algoritmos aplicados..... 32

LISTA DE ABREVIATURAS E SIGLAS

K	*Número de clusters*
PDV	*Ponto de Venda*
TD	*Tempo de Deslocamento*
TME	*Tempo Médio de Entrega*
TR	*Tempo em Rota*
SKU	*Stock Keeping Unit (Unidade de Manutenção de Estoque)*

SUMÁRIO

1 INTRODUÇÃO 10
2 OBJETIVO 12
2.1 Objetivo Geral 12
2.2 Objetivos Específicos 12
3 REVISÃO BIBLIOGRÁFICA 13
3.1 Algoritmos Exatos 13
3.2 Algoritmos Heurísticos 15
3.2.1 Algoritmos Supervisionados 16
3.2.2 Algoritmos Não Supervisionados 17
4 METODOLOGIA 19
4.1 Avaliação do processo de distribuição atual 20
4.2 Levantamento e análise dos dados 22
4.3 Aplicação do método de clusterização *K-means* 25
5 RESULTADOS E DISCUSSÃO 27
5.1 Clusterização 1 27
5.2 Clusterização 2 30
5.3 Simulação de resultados 32
6 CONCLUSÃO 34
7 REFERÊNCIAS BIBLIOGRÁFICAS 35

1 INTRODUÇÃO

Os grandes avanços da vida moderna e da sociedade em geral têm provocado, de maneira acelerada, a necessidade de construir processos e métodos de cálculo eficientes para resolvê-los. Alguns desses problemas são de natureza complexa, com ordem de grandeza de milhares ou milhões de informações associadas, precisando de estudos adequados, técnicas específicas e grande velocidade dos cálculos para conseguir resultados em tempo viável. Ou seja, como produzir de maneira cada vez mais eficiente e como distribuir produtos e serviços em larga escala, em tempo hábil e com o menor custo, são questões muito relevantes atualmente e que merecem respostas efetivas.

O planejamento de operações de distribuição de bens e serviços é uma atividade de suma importância para muitas indústrias e empresas do estado de Alagoas. Uma das decisões centrais que envolvem o sistema de distribuição de produtos é a elaboração regular de um conjunto de rotas a serem realizadas por uma frota de veículos. Essas rotas definem um conjunto de clientes servidos por cada veículo e que são classificados e sequenciados de acordo com objetivos específicos. Entre os objetivos mais comuns, incluem-se a obtenção de rotas com menor custo associado (número de veículos utilizados da frota própria, número de veículos terceirizados ou fretados, distância percorrida, combustível utilizado, tempo em rota e a dispersão em km da rota), assim como rotas que ofereçam maior nível de serviço (maior quantidade de clientes atendidos, menor violação de prazos acordados).

Outro elemento importante que precisa ser considerado na elaboração das rotas são alguns fatores que as influenciam indiretamente, entre eles: o fluxo de visitas dos vendedores aos clientes, as limitações de tempo máximo de rota, a capacidade e peso da carga dos veículos, imposição de janelas de tempo para a coleta ou entrega em cada cliente, e tipos de veículos que podem atender cada cliente. É fácil perceber que o confronto entre os objetivos que se deseja alcançar e as restrições que limitam o seu alcance tornam o processo decisório de planejamento e programação de rotas bastante complexo e um grande desafio para os tomadores de decisão.

Dentro dessa perspectiva, no desenvolvimento deste trabalho será necessário desenvolver sobre as variações de algoritmos de otimização existentes que sejam capazes de realizar uma etapa crucial da execução de um sistema de distribuição de bebidas, que

é o agrupamento dos clientes que serão visitados, visto que as visitas dos vendedores antecedem em 1 dia a distribuição dos produtos.

2 OBJETIVO

2.1 Objetivo Geral

Desenvolver estratégias no formato de ferramenta computacional, utilizando-se da linguagem Python, para a modelagem e otimização de um sistema de distribuição de bebidas através do agrupamento de clientes da distribuição, envolvendo mais especificamente o algoritmo heurístico NP-Difícil *K-Means*.

2.2 Objetivos Específicos

1. Construir um modelo de agrupamento de clientes direcionados para visitas de vendedores da cervejaria em dias específicos da semana;
2. Avaliar a aplicabilidade da mudança dos dias de visita/vendas atuais dos vendedores da cervejaria;
3. Avaliar a aplicabilidade do modelo de clusterização *K-Means* para o agrupamento de clientes da cervejaria;
4. Verificar a evolução de produtividade ou redução de custos associados à otimização final do sistema de distribuição de bebidas da cervejaria.

3 REVISÃO BIBLIOGRÁFICA

Problemas de otimização ocorrem frequentemente em uma grande quantidade de aplicações industriais e científicas. Problemas desse tipo podem ser definidos essencialmente por um par $(X, f(.))$, em que X é um conjunto qualquer associado ao conjunto de restrições e $f:X \rightarrow R$ representa uma função de custo associada. Em particular, no caso de problemas de agrupamentos e minimização, deseja-se encontrar uma solução viável $x^* \in X$ em que $f(x^*) \leq f(x), \forall\ x \in X$. Se X é um conjunto discreto, a determinação de uma solução ótima x^* é teoricamente possível enumerando-se todos os elementos de X e computando-se, paralelamente, sua imagem associada. Na prática, entretanto, essa abordagem é geralmente inviável já que o número de soluções viáveis cresce exponencialmente com o tamanho do problema.

Devido ao elevado esforço computacional exigido na determinação de uma solução exata para uma grande quantidade de problemas combinatórios (denominados problemas NP-difíceis), muitos dos algoritmos propostos para a solução podem não buscar diretamente um mínimo global. Dessa forma, nesses casos normalmente é adotada uma estratégia que vise equilibrar a qualidade da solução obtida com o tempo total de processamento. Isso pode ser viabilizado, por exemplo, através de métodos heurísticos (aproximativos). A utilização de métodos exatos, por outro lado, se justifica principalmente em situações especiais, em que as possibilidades consideradas para um determinado problema são “suficientemente pequenas” para a aplicação considerada.

3.1 Algoritmos Exatos

Como já mencionado, são algoritmos mais aplicados a problemas de baixa complexidade ou menor número de variáveis. Na última década, a utilização de modelos de Programação Linear Inteira Mista (PLIM) tem sido muito recorrente e cada vez mais refinada. A disponibilidade de softwares robustos, máquinas mais rápidas e a possibilidade de se resolver problemas relativamente grandes em um menor tempo de processamento têm tornado a programação inteira muito mais atrativa a pesquisadores e analistas da iniciativa privada e do setor público.

O sucesso na solução de problemas de Programação Linear Inteira Mista de larga escala requer formulações cuja “relaxação linear” correspondente defina uma boa aproximação do conjunto de soluções viáveis. Na última década, um grande esforço tem sido direcionado a métodos do tipo Branch-and-Cut, Branch-and-Price e Relax-and-Cut

na resolução desses problemas. Hoffman e Padberg (1985) e Nemhauser & Wolsey (1988), Wolsey (1998) apresentam uma exposição geral dessas metodologias.

A ideia descrita no método Branch and Cut defende que conjuntos de dados associados à formulação de um problema são desprezadas na "relaxação linear". O grande número de restrições impede que o problema seja tratado convenientemente utilizando-se apenas ferramentas de programação linear. Dessa forma, se uma solução ótima associada à "relaxação linear" é inviável, um novo problema de separação deve ser "resolvido" buscando a identificação de uma ou mais restrições violadas pela relaxação corrente (*cutting procedure*). O novo problema obtido (com restrições adicionais) é novamente resolvido via programação linear e o processo é repetido até que novas classes de desigualdades violadas não sejam mais encontradas. Nesse momento, interrompe-se o processo e retorna-se à árvore de busca (Branch-and-Bound).

A filosofia do método Branch-and-Price é similar à do método Branch-and-Cut. A diferença básica é que, no primeiro, faz-se uma geração de colunas ao invés de geração de linhas (restrições violadas). Na verdade, trata-se de procedimentos complementares visando um incremento dos limites inferiores gerados pela relaxação linear associada.

Os procedimentos Relax-and-Cut visam combinar relaxação lagrangena (na geração de limites inferiores) com resultados de teoria poliédrica e Branch-and-Bound. Nesse caso, um subconjunto do conjunto de restrições é dualizado. Os multiplicadores de Lagrange associados a esse subconjunto são então atualizados utilizando-se, por exemplo, o método subgradiente de Reeves (1993), ou o algoritmo do volume, apresentado por Bararona e Ambil (1997). Neste caso, como o número de restrições dualizadas pode ser extremamente grande (exponencial), é necessário que se defina um subconjunto bem menor de restrições dualizadas permitindo sua atualização dinamicamente.

3.2 Algoritmos Heurísticos

Da mesma forma, os algoritmos heurísticos que também foram mencionados, são uma classe de métodos que levam a soluções não necessariamente exatas, mas que conseguem atingir resultados de boa qualidade, isto é, próximos do ótimo.

Uma das limitações dos métodos que visam à determinação de soluções ótimas para problemas combinatórios é que raramente eles resolvem grandes problemas em um tempo computacional aceitável. Em função disso, muitos pesquisadores se concentraram no desenvolvimento de heurísticas ou metaheurísticas inteligentes como *Simulated Annealing*, GRASP (*Greedy Randomized Adaptive Search Procedure*), VNS (*Variable Neighborhood Search*), Algoritmos Genéticos entre outros. O crescente avanço dos algoritmos aproximativos pode ser atribuído, basicamente, à dificuldade de resolução de uma grande variedade de importantes problemas combinatórios. Na verdade, pode-se afirmar que grande parte dos problemas combinatórios de interesse prático são NP-Difíceis.

Diante disso, a constatação que um determinado problema seja NP-Difícil nos coloca imediatamente diante de outra questão: qual a melhor estratégia de resolução a ser adotada? Essa dificuldade se agrava drasticamente à medida que grandes conjuntos de dados são considerados, como será o caso deste trabalho.

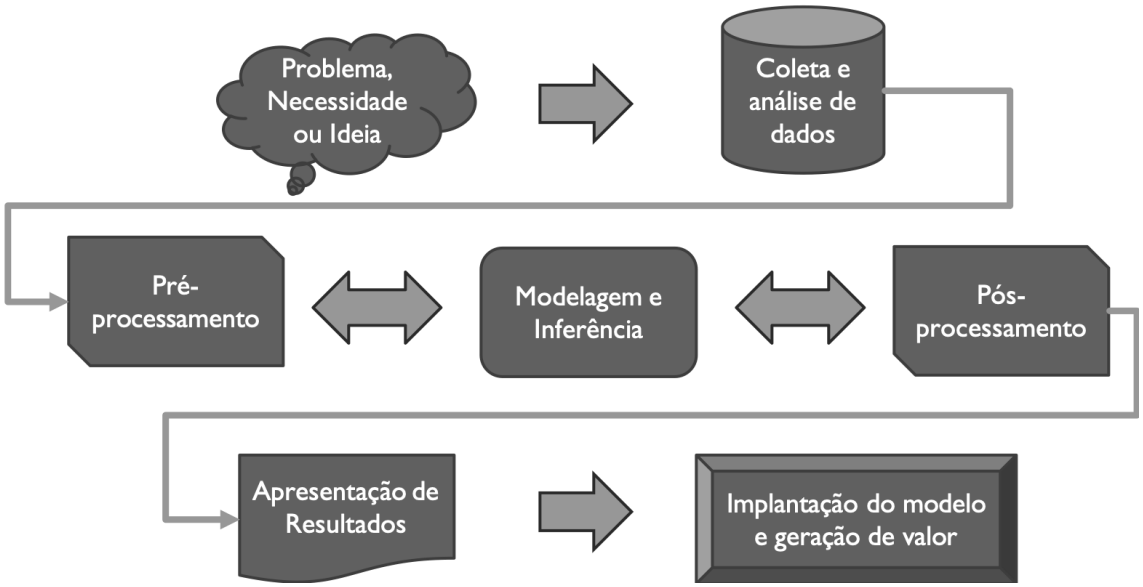
Os algoritmos aproximativos (especialmente os determinísticos aproximativos), foram introduzidos por Johnson (1974) e são algoritmos polinomiais que buscam sacrificar o mínimo possível da qualidade obtida nos métodos exatos, ganhando, simultaneamente, o máximo possível em eficiência (tempo polinomial). Como discutido em Hochbaum (1997), a busca do equilíbrio entre essas situações conflitantes é o grande paradigma dos algoritmos aproximativos.

Antes do surgimento dos algoritmos aproximativos a análise de desempenho dos métodos heurísticos se baseava, simplesmente, em sua execução para um conjunto finito de dados (*benchmark*). A performance da heurística era então comparada com a de outras heurísticas para o mesmo conjunto de dados considerado. Este tipo de comparação, ainda hoje bastante utilizado, retorna apenas uma medida parcial de desempenho, já que o conjunto de dados é normalmente pequeno, além de não representar satisfatoriamente o conjunto de todos os dados associadas ao problema. Em outras palavras, uma heurística

com bom desempenho para esse conjunto finito não mantém, necessariamente, a mesma performance quando aplicada a outro conjunto de dados com características distintas.

Dentro desse contexto, surge a aprendizagem computacional, um subcampo da Engenharia que evoluiu o estudo sobre reconhecimento de padrões e que pode ser definido como o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados (SIMON, 2013). Esse campo surge como um conjunto de métodos que podem detectar padrões em dados automaticamente para depois usá-los para predizer dados futuros.

Figura 1 - Esquema geral de aprendizagem da máquina.



Fonte: Escovedo (2020).

Ela geralmente é dividida em dois tipos principais: aprendizado supervisionado e não-supervisionado.

3.2.1 Algoritmos Supervisionados

O supervisionamento é utilizado quando, a partir de um conjunto predefinido de dados (*dataset*), se busca encontrar uma função que possa prever resultados futuros. Nesse tipo de aprendizado, é possível estimar números reais ou valores dentro de um conjunto finito. Essa estimativa normalmente se dá de duas formas:

- Classificação: estimada com base em um conjunto finito de dados, este modelo pode ser usado para classificar um grupo de dados com a busca de uma função matemática que permita estimar corretamente cada parâmetro do conjunto.

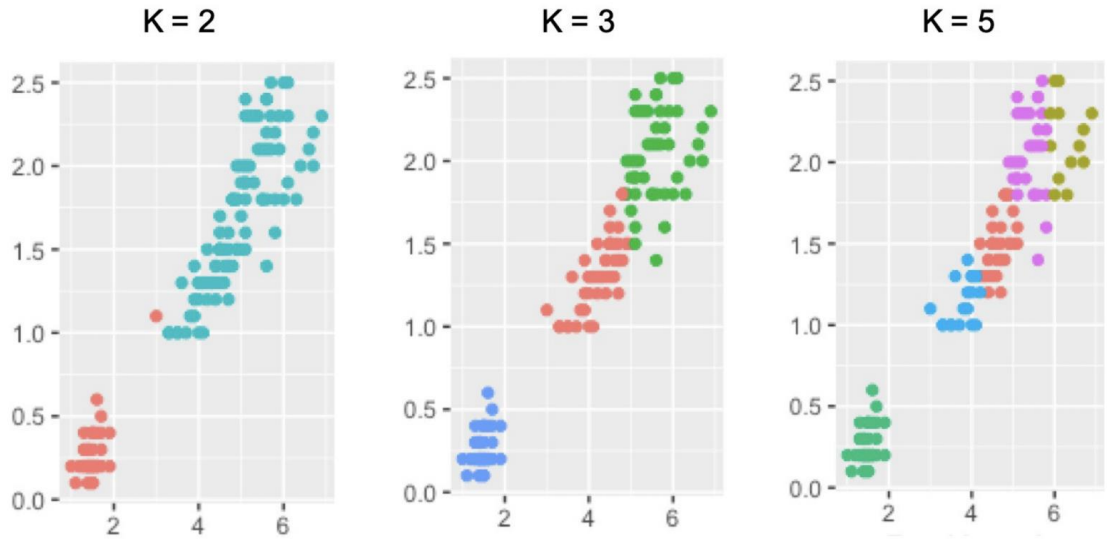
- Regressão: consiste em realizar aprendizado supervisionado a partir de dados históricos com uma saída de dados numérica (contínuo ou discreto). Neste modelo, verifica-se a distância ou o erro entre a saída do modelo e a saída desejada. ao estimar valores reais.

3.2.2 Algoritmos Não Supervisionados

A Aprendizagem não supervisionada, por outro lado, nos permite abordar problemas com pouca ou nenhuma ideia do que os resultados devem aparentar. Pode-se derivar de uma estrutura de dados onde não necessariamente se sabe o efeito ou histórico das variáveis. Assim, o processo de aprendizado busca identificar regularidades entre os dados a fim de agrupá-los ou organizá-los em função das similaridades que apresentam entre si.

Uma das categorias mais utilizadas para uso da aprendizagem não supervisionada é a clusterização, que visa agrupar os dados de interesse, ou separar os registros de um conjunto de dados em subconjuntos (*clusters*), com base em propriedades comuns entre os elementos de cada *cluster*. Um dos algoritmos mais utilizados e conhecidos de clusterização é o *K-Means*, que é baseado principalmente pela distância entre o conjunto dos dados.

Figura 2 - Exemplo do resultado do algoritmo *K-means* para diferentes valores de k.



Fonte: Escovedo (2020).

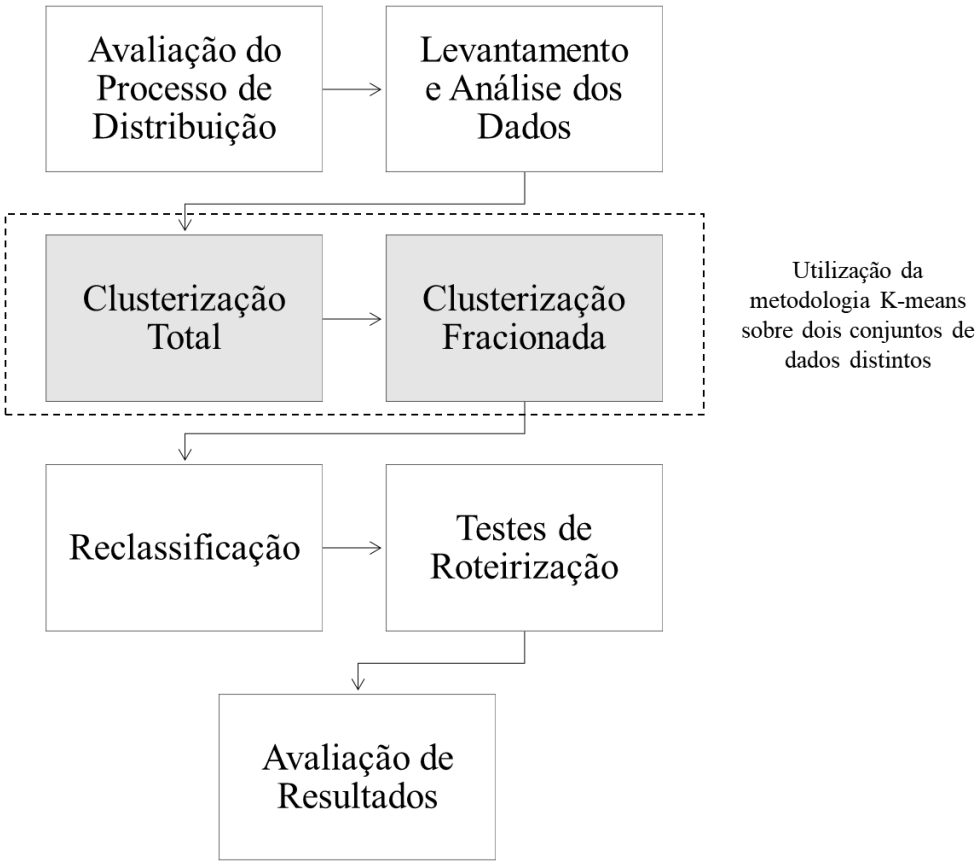
O algoritmo visa separar os dados em *K clusters* (um número predefinido), de acordo com a distância de cada ponto até o centroide. O centroide é um ponto que, no

início do algoritmo, é determinado de forma aleatória ou direta para cada um dos *K* clusters determinados. A seguir, é preciso associar cada dado do conjunto inicial de dados (*dataset*) ao centroide mais próximo. Depois, é recalculado *k* novos centroides como sendo os baricentros dos grupos resultantes dos dados. Em geral, o *K-means* apresenta bom desempenho quando os grupos de dados são densos, compactos e bem separados uns dos outros, além de ser computacionalmente rápido e de fácil entendimento e implementação.

4 METODOLOGIA

Neste trabalho, será utilizado como aplicação principal o algoritmo (não supervisionado) de clusterização *K-means* para a otimização do sistema de distribuição de uma cervejaria. Serão feitas duas aplicações sobre o sistema atual, a primeira consistirá em aplicar o método sobre todo o conjunto de dados de clientes da cervejaria, e a outra aplicação será somente no conjunto de dados de clientes que cada vendedor da empresa possui em sua agenda de visitas. O esquema a seguir ilustra como será a realização desse trabalho.

Figura 3 – Metodologia do trabalho para otimização do sistema de distribuição.

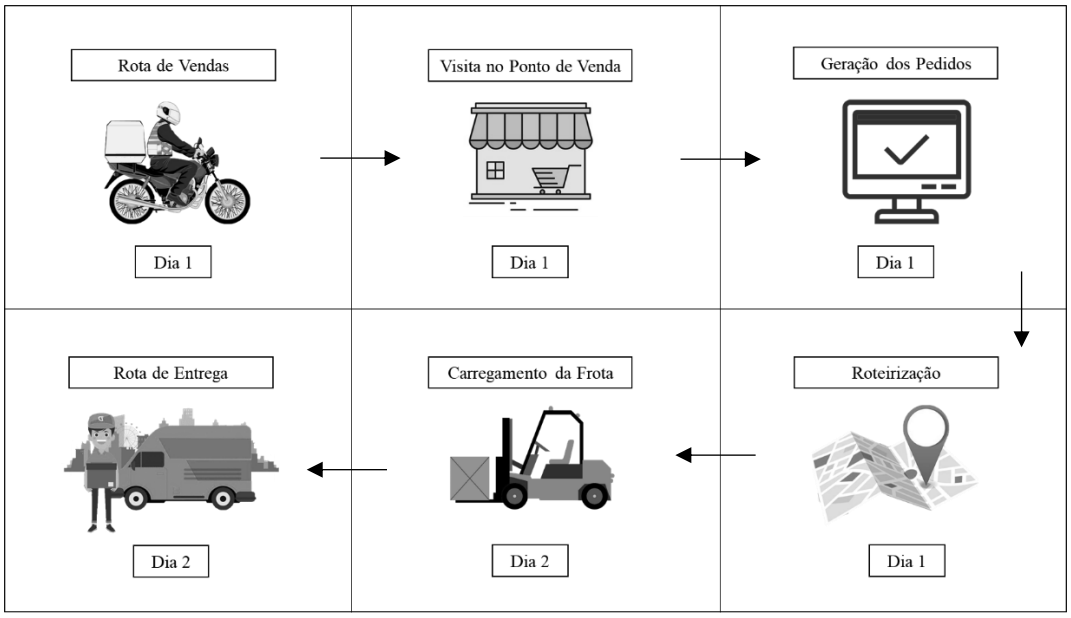


Fonte: O Autor (2021).

4.1 Avaliação do processo de distribuição atual

Para início deste trabalho, é necessário entender como funciona o processo atual da empresa em relação a venda e distribuição dos produtos. O esquema a seguir ilustra o processo atual de vendas e distribuição.

Figura 4 - Processo de vendas e distribuição atual.



Fonte: O Autor (2021).

- Rota de Vendas: Essa etapa inicial do processo consiste na visita de todos os pontos de vendas do dia. Eles são divididos entre todos os vendedores da empresa principalmente de acordo com a segmentação de cada cliente.
- Visita no Ponto de Venda: Nesta etapa o vendedor executa o processo de venda direcionando os produtos que o cliente necessita e alinhando sobre a entrega que será realizada no dia seguinte.
- Geração dos Pedidos: Às 18:00 horas do dia, todos os pedidos são direcionados para o sistema da empresa, são validados e enviados para o sistema de roteirização. Depois de 18:00 horas, não é mais possível fazer pedidos para entregar no dia seguinte.
- Roteirização: Nesta etapa, o setor de roteirização da empresa basicamente vai definir as rotas de entrega do dia seguinte, de acordo com o volume dos pedidos, o território de entrega do ponto de venda, o peso e volume que cada caminhão de entrega comporta, o tempo de entrega médio de cada ponto de venda (deslocamento + descarga) e a jornada de trabalho da equipe de entrega.

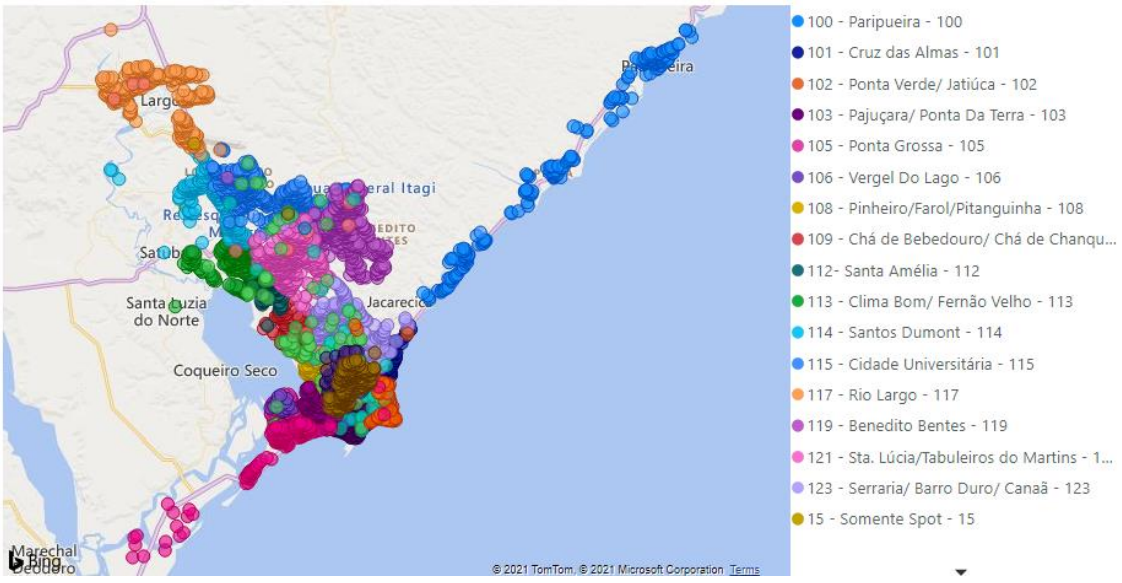
- Carregamento da Frota: De acordo com a roteirização realizada, os pedidos são faturados e carregados dentro da frota disponibilizada pela empresa. Essa etapa normalmente acontece durante a madrugada do dia seguinte.
- Rota de Entrega: Essa é a última etapa do processo, a equipe de entrega é dividida de acordo com o número de frotas que foram preparadas e então é realizada a sequência de entregas de todos os pedidos que foram realizados no dia anterior à visita dos vendedores.

4.2 Levantamento e análise dos dados

Os dados a serem utilizados para a aplicação do método foram coletados através da base de dados de clientes da cervejaria, para preservação da informação, os dados foram alterados de maneira que as informações confidenciais da empresa e de seus clientes fossem preservadas sem afetar nenhuma das partes.

Os principais dados coletados foram as informações de Latitude e Longitude dos clientes da empresa, além da frequência de visita dos vendedores a cada um desses clientes. A seguir apresentamos no mapa a posição geográfica de cada um dos clientes da empresa.

Figura 5 - Posição geográfica dos clientes na cidade de Maceió-AL



Fonte: O Autor (2021)

O conjunto de clientes estão concentrados na cidade de Maceió-AL e são distribuídos em territórios específicos, de acordo com o bairro de sua localização. Além disso, existe uma segmentação desses clientes nos seguintes perfis de mercado: depósito de bebidas, bares, minimercados, mercearias, padarias, conveniências, restaurantes etc. A tabela a seguir traz a quantidade de clientes de acordo com o seguimento pertencente.

Tabela 1 - Nº de clientes para cada segmento.

SEGMENTO	Nº CLIENTES
ARMAZÉM / MARCEARIA	2058
DEPÓSITO DE BEBIDAS	540
BAR / BARZINHO	521
MINIMERCADO	413
BOTECO / BOTEQUIM	191
RESTAURANTE / PIZZARIA	162
LANCHONETE / PASTELARIA	94
PADARIA / CONFEITARIA	49
POSTOS DE GASOLINA	48
LOJA DE CONVENIÊNCIA	40
SUPERMERCADO	29
TRAILLER / BARRACA	27
FOOD SERVICE	19
OUTROS	410

Fonte: O Autor (2021).

Esse conjunto de clientes é regularmente visitado pelos vendedores da empresa ao menos uma vez por semana. A frequência atual de visita é determinada basicamente pelo segmento de mercado de cada cliente e suas características de demanda. A seguir temos o número de cliente visitados atualmente por cada dia da semana pelos vendedores.

Tabela 2 - Nº de clientes visitados por dia da semana.

DIA DA SEMANA	Nº CLIENTES
SEGUNDA-FEIRA	858
TERÇA-FEIRA	820
QUARTA-FEIRA	907
QUINTA-FEIRA	820
SEXTA-FEIRA	787
SÁBADO	409

Fonte: O Autor (2021).

A partir dessas informações, é possível entender de maneira mais profunda os padrões gerais de visita e entrega dos produtos da cervejaria e como o sistema atual de distribuição é basicamente guiado pelas visitas de vendas. Por fim, é importante também visualizar a quantidade de clientes aberta por cada vendedor, visto que também esse fator influencia na divisão de clientes por dia de semana. Atualmente são 38 vendedores cada um com uma quantidade determinada de clientes para visitar e vender.

Tabela 3 - N º de clientes por vendedores por sala de vendas.

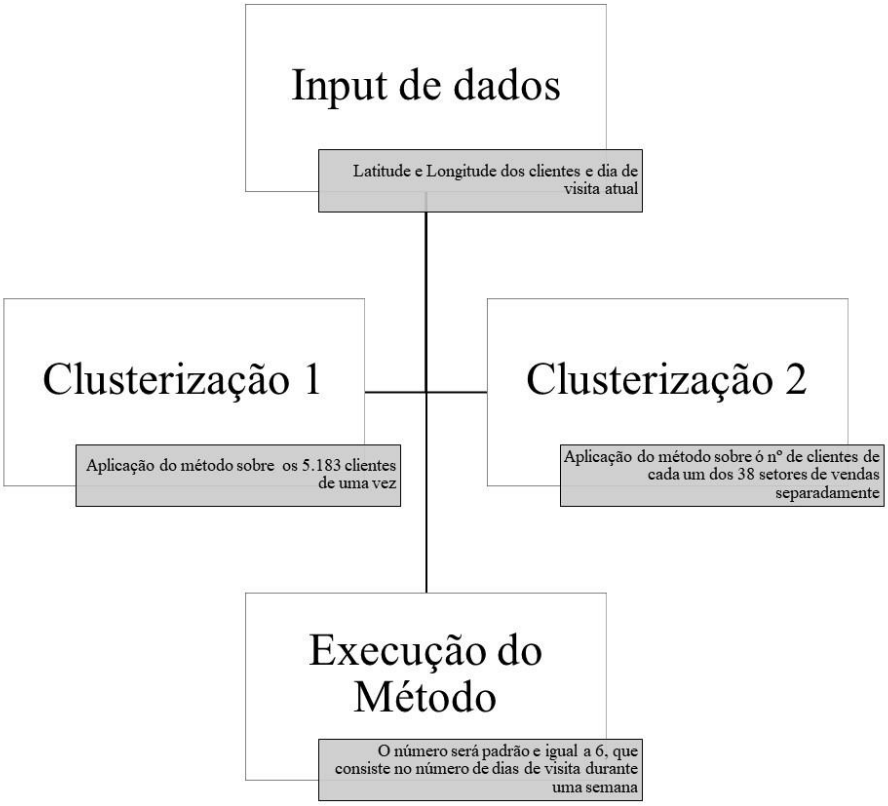
SALA 1	Nº CLIENTES	SALA 2	Nº CLIENTES
VENDEDOR 1	168	VENDEDOR 15	200
VENDEDOR 2	172	VENDEDOR 16	181
VENDEDOR 3	143	VENDEDOR 17	170
VENDEDOR 4	90	VENDEDOR 18	193
VENDEDOR 5	67	VENDEDOR 19	149
VENDEDOR 6	90	VENDEDOR 20	192
VENDEDOR 7	78	VENDEDOR 21	181
VENDEDOR 8	78	VENDEDOR 22	186
VENDEDOR 9	82	VENDEDOR 23	169
VENDEDOR 10	133	VENDEDOR 24	177
VENDEDOR 11	90	VENDEDOR 25	212
VENDEDOR 12	105	VENDEDOR 26	201
VENDEDOR 13	87	VENDEDOR 27	177
VENDEDOR 14	32	VENDEDOR 28	192
-	-	VENDEDOR 29	229
-	-	VENDEDOR 30	207
-	-	VENDEDOR 31	170

Fonte: O Autor (2021).

4.3 Aplicação do método de clusterização *K-means*

Nesta etapa do trabalho, como explicado anteriormente, será realizada a aplicação da metodologia de clusterização de duas formas diferentes sobre o mesmo conjunto de dados, o esquema a seguir ilustra as duas aplicações.

Figura 6 - Esquema de aplicação do método de trabalho.



Fonte: O Autor (2021).

Considerando então o conjunto de dados levantados $P = \{p_1, p_2, ..., p_n\}$, sendo n o número de clientes em análise, e p representando o conjunto de pontos $p_1 = (x_1, y_1)$ como sendo a latitude e a longitude, respectivamente, de cada um dos clientes. O algoritmo *K-means* inicia-se com a escolha de *K* centroides $\{C_1, C_2, ..., C_k\}$ para o agrupamento. Depois, cada ponto do conjunto *P* é associado ao seu centro mais próximo segundo o cálculo da distância euclidiana d_E, formando assim *k* grupos P_i. O próximo passo do algoritmo é atualizar os centroides. De acordo com Lloyd (1957), o centroide é escolhido como sendo o ponto que minimiza a soma do quadrado da distância d_E, entre ele mesmo e cada ponto do conjunto.

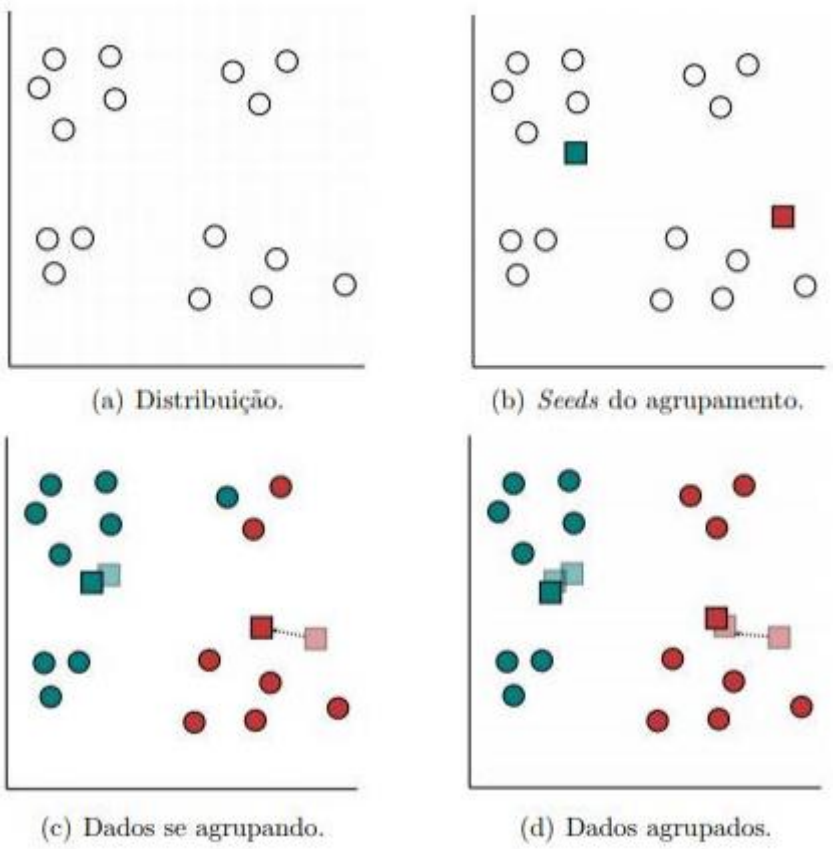
$$c_i = argmin \sum_{pj \in P_i} d_E^2(c_i, p_j),$$

Esse ponto é justamente o centro de massa do grupo P_i, e se dá por:

$$c_i = \frac{p_{i1} + p_{i2} + \cdots + p_{im}}{|P_i|}$$

em que $|P_i|$ corresponde a cardinalidade de P_i e $p_{ij} \in P_i$ com $j = 1, ..., m$ (PINELE, 2017). Como exemplo, a figura a seguir representa o funcionamento do método de maneira simplificada com um conjunto simples de dados.

Figura 7 - Clusterização de dados usando *K-means.*



Fonte: Prado (2008).

A aplicação dos algoritmos foi realizada utilizando as bibliotecas *numpy, pandas*, *matplotlib* e *seaborn*, ambas disponíveis na linguagem de programação *python.*

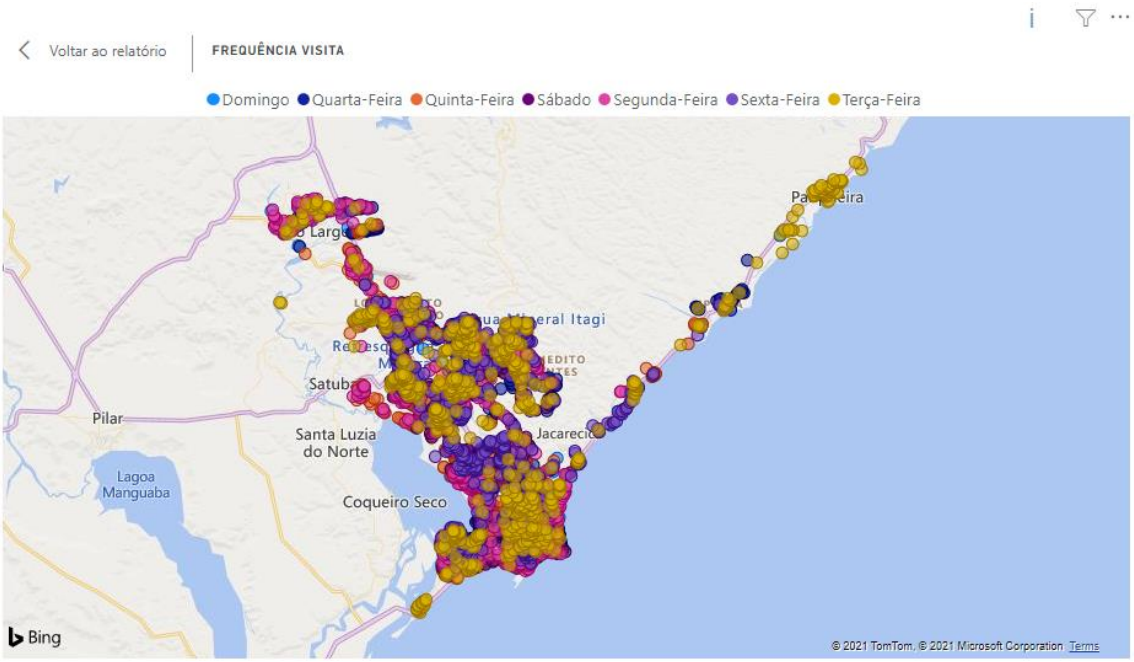
5 RESULTADOS E DISCUSSÃO

São apresentadas a seguir as duas aplicações do método adotado no trabalho. O número *K* de clusters utilizado pela aplicação foi igual a seis, considerando cada cluster como um dia da semana de segunda-feira a sábado. A sequência de resultados apresenta o cenário anterior e o posterior ao uso do método e a simulação de seus respectivos retornos para a empresa.

5.1 Clusterização 1

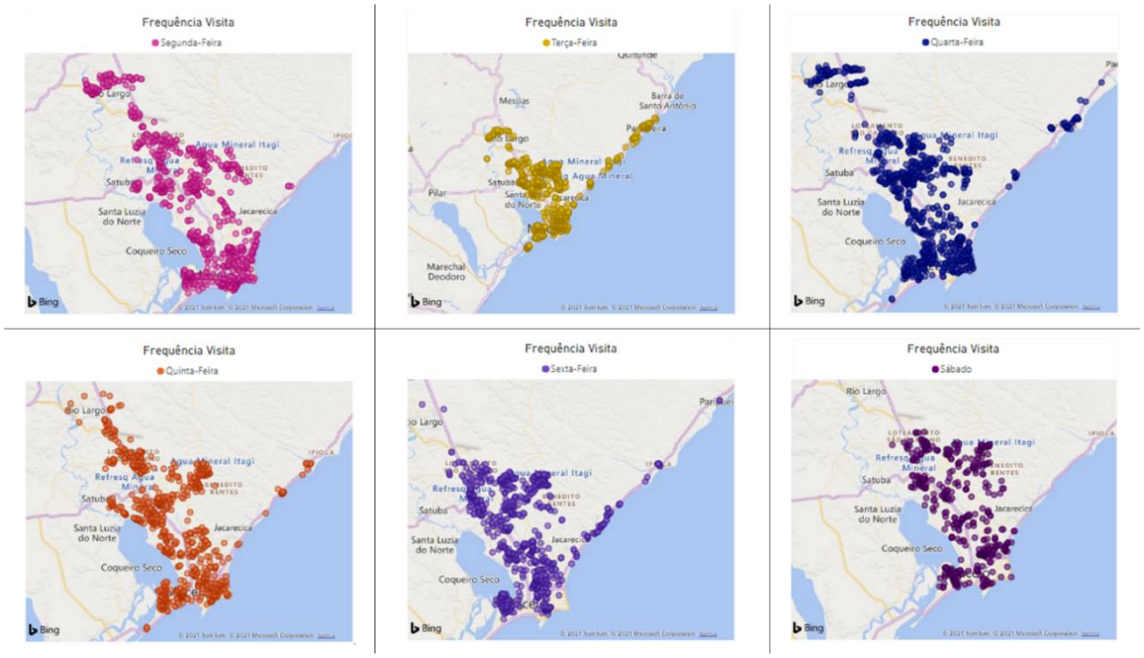
Nesta etapa de clusterização, o método foi realizado sobre todo o conjunto de clientes, sendo atribuído um cluster para dia de visita na semana (segunda-feira à sábado). A figura 8 apresenta no cenário antigo, sendo caracterizado através das cores, o dia de visita dos clientes.

Figura 8 - Distribuição dos clientes por dia da semana (cenário antigo).



Fonte: Bing (2021).

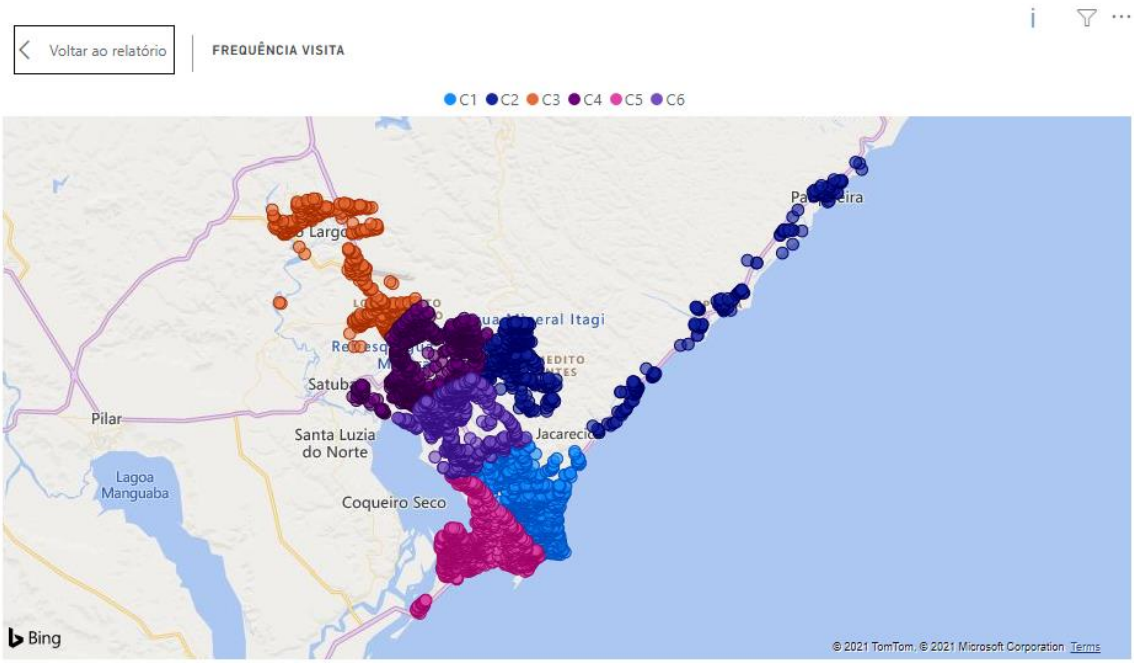
Figura 9 – Distribuição dos clientes por cada dia da semana (cenário antigo).



Fonte: Bing (2021).

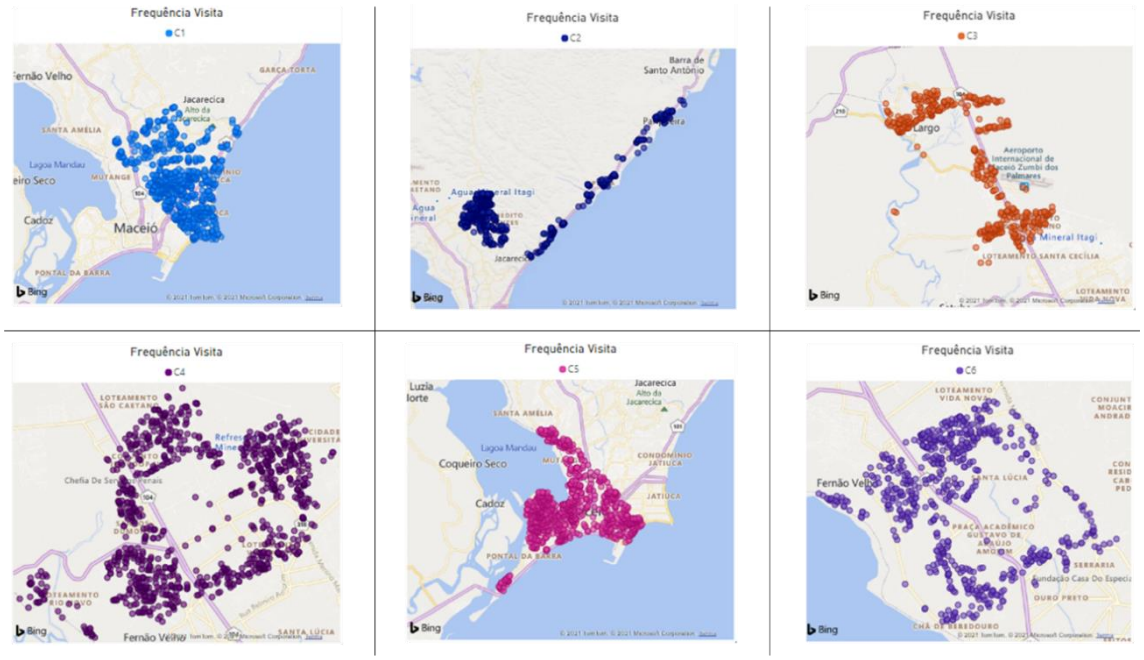
Após a aplicação do método, o resultado a seguir é alcançado. O método não fornece diretamente o novo dia de visita, mas sim uma classificação dentro dos novos clusters, conforme figura a seguir.

Figura 10 - Distribuição dos clientes em clusters (cenário novo).



Fonte: Bing (2021).

Figura 11 - Distribuição de clientes por cluster (cenário novo).

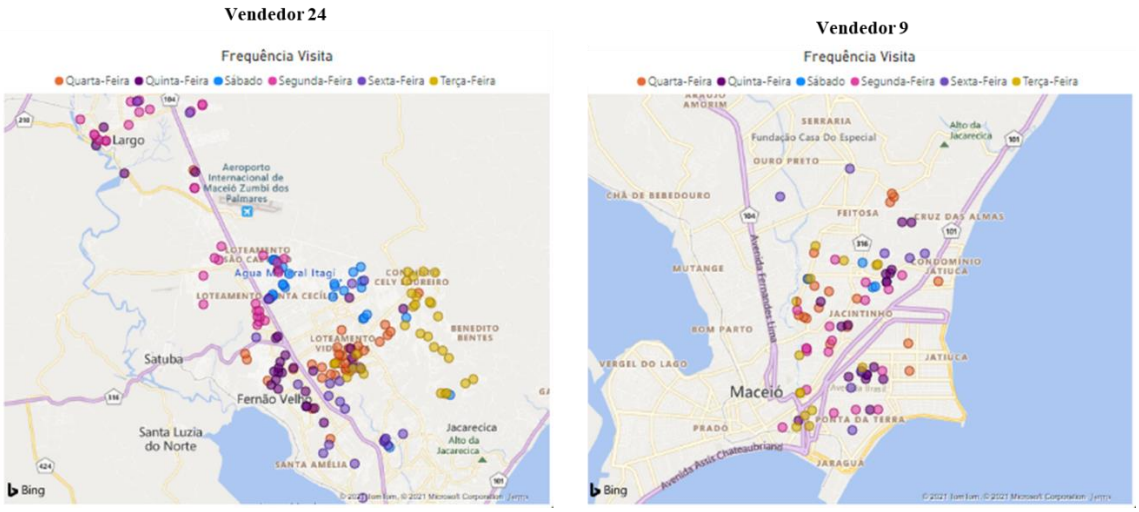


Fonte: Bing (2021).

5.2 Clusterização 2

Para a aplicação desta etapa, foi necessário executar o algoritmo 31 vezes, que é o número de vendedores da empresa. Basicamente, a aplicação foi realizada dentro do conjunto de clientes de cada vendedor de forma separada, gerando assim a melhor combinação para os clientes daquele vendedor. A seguir estão alguns vendedores específicos antes da aplicação da metodologia.

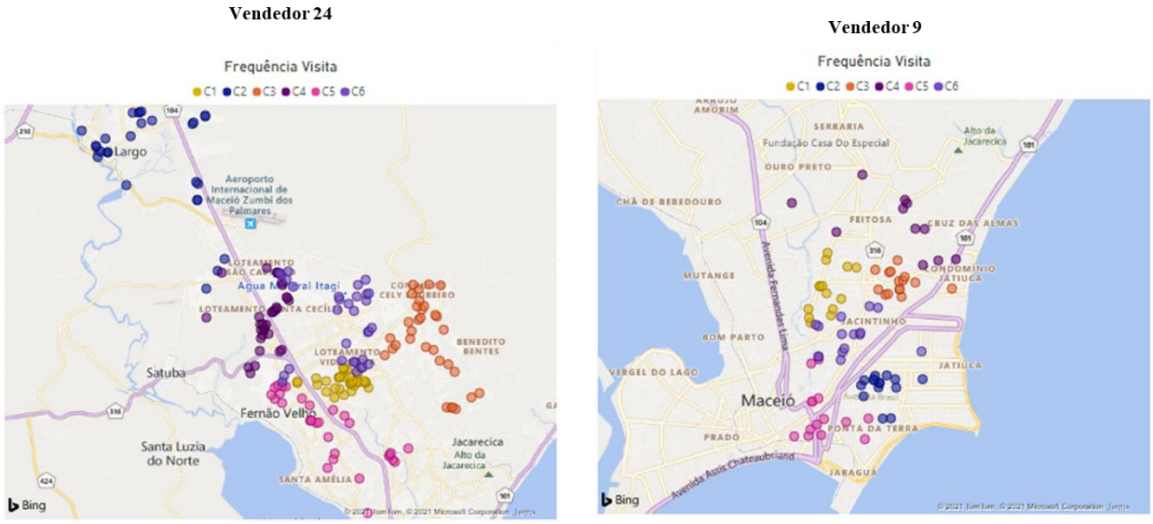
Figura 12 - Clientes dos vendedores 24 e 9 (cenário antigo).



Fonte: Bing (2021).

Aplicando o algoritmo, temos o seguinte resultado aberto pelos vendedores especificados.

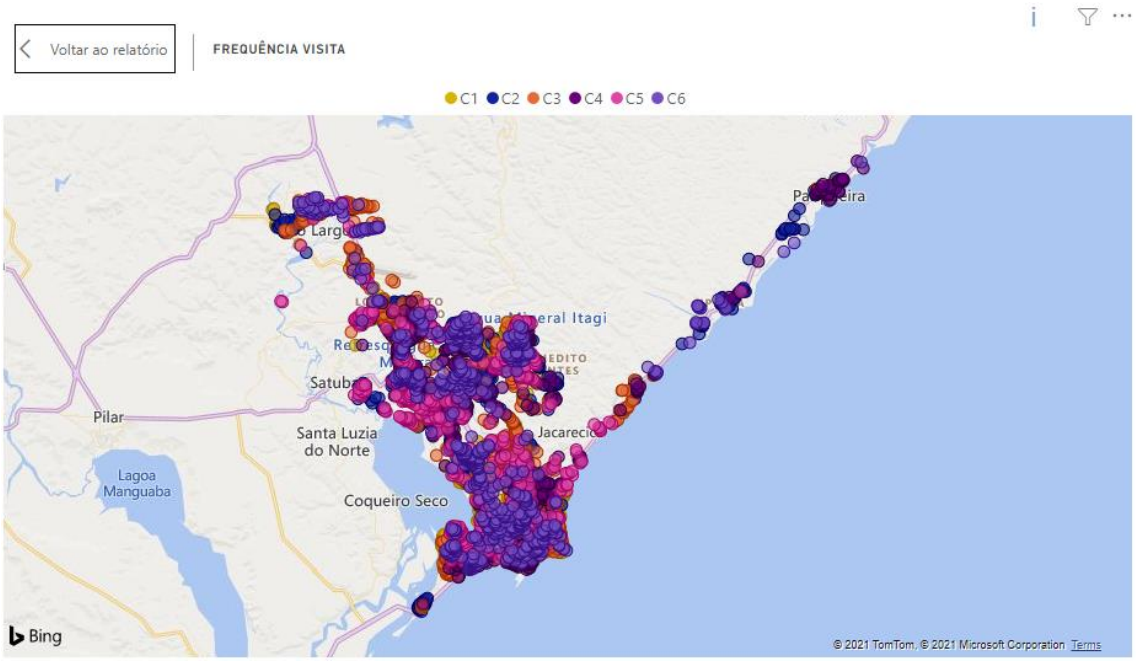
Figura 13 - Clientes dos vendedores 24 e 9 (cenário novo)



Fonte: Bing (2021).

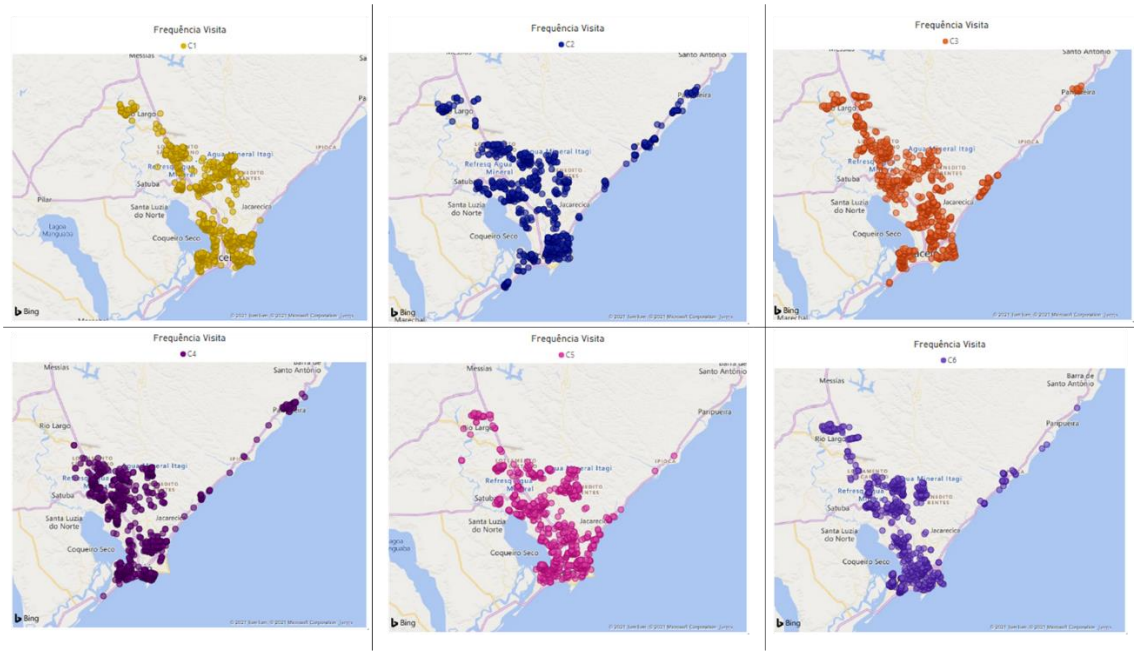
Já no cenário total de clientes, temos a seguinte visualização geral e aberta por cada cluster como dia da semana.

Figura 14 - Distribuição dos clientes por dia da semana (cenário novo).



Fonte: Bing (2021).

Figura 15 - Distribuição de clientes por cluster (cenário novo).



Fonte: Bing (2021).

Visualmente, consegue-se perceber que o resultado geral na visão de dia de semana não é tão afetado quando comparado com a visão de visita de cada vendedor.

5.3 Simulação de resultados

Para entender se de fato temos evolução em ambos os formatos de clusterização e o quanto se evolui ou não entre a aplicação de cada um dos modelos, será realizada uma simulação comparando o modelo atual com a clusterização 1 e 2. O cálculo de simulação será realizado com base nas seguintes equações:

$$Tempo\ em\ Rota = Tempo\ de\ Deslocamento + Tempo\ Médio\ de\ Entrega$$

(Equação 1)

$$Tempo\ de\ Deslocamento = \frac{Distância\ Planejada}{Velocidade\ da\ Via}$$

onde, *TR = Tempo em Rota, TD = Tempo de Deslocamento e TME = Tempo Médio de Entrega.* O *TR* é calculado em relação ao número de clientes que serão realizadas as entregas no dia da rota.

A partir disso, o cálculo de simulação será realizado utilizando-se os dados históricos registrados do *TME* de cada cliente da rota que foi avaliado. Esse valor é calculado com base na média do tempo das últimas 10 entregas realizadas no ponto de venda.

A seguir, na tabela 4 temos uma visão geral comparativa dentro de cada dia de entrega da semana com relação à dispersão do Tempo de Deslocamento (TD) comparando-se com o TD para a clusterização 1 e para a clusterização 2.

Tabela 4 - Comparação da dispersão de tempo de deslocamento entre os algoritmos aplicados.

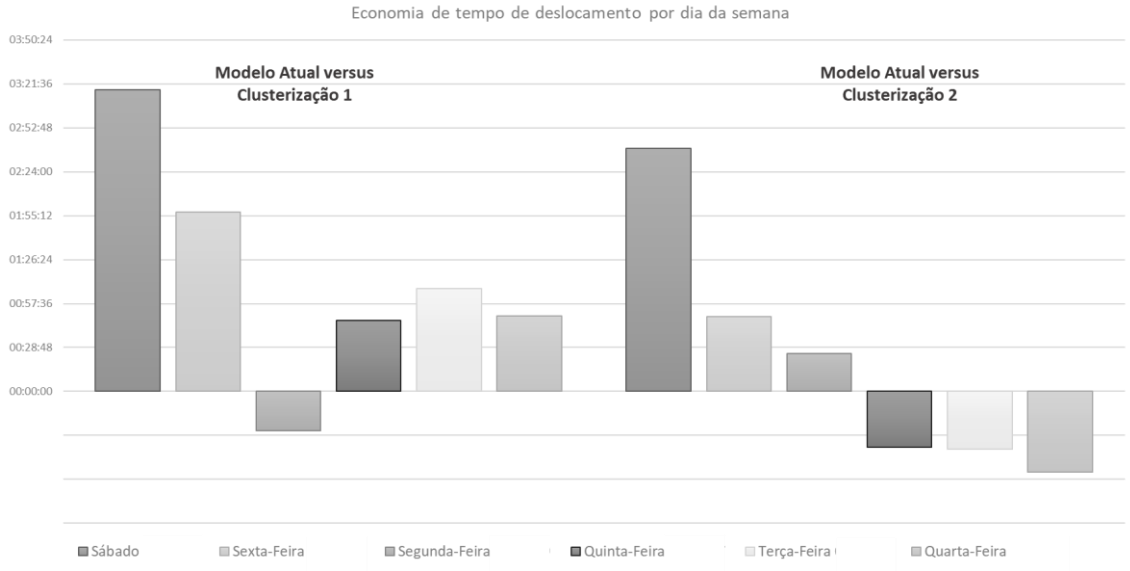
		COMPARAÇÃO					
Dia da Semana	**TD Atual (1)**	**Δ 12**	**Δ 13**	**TD C1 (2)**	**Δ 21**	**Δ 23**	**TD C2 (3)**
Sábado	06:35:55	-50%	-40%	03:18:11	100%	19%	03:56:40
Sexta-Feira	04:40:15	-42%	-18%	02:42:49	72%	42%	03:51:10
Segunda-Feira	04:13:49	10%	-10%	04:39:42	-9%	-18%	03:49:11
Quinta-Feira	03:09:37	-24%	19%	02:23:16	32%	58%	03:46:17
Terça-Feira	03:07:52	-36%	20%	02:00:37	56%	87%	03:45:34
Quarta-Feira	02:37:19	-32%	33%	01:47:41	46%	95%	03:30:00
Média	**04:04:08**	**-29%**	**1%**	**02:48:43**	**49%**	**47%**	**03:46:29**

Fonte: O Autor (2021).

É possível perceber com clareza que ambos os métodos trazem melhoras em geral em comparação ao modelo atual de visitas e entregas da empresa. O gráfico a seguir, traz

uma noção comparativa entre o tempo economizado no deslocamento comparando os dois modelos aplicados com a maneira atual de distribuição dos clientes da empresa.

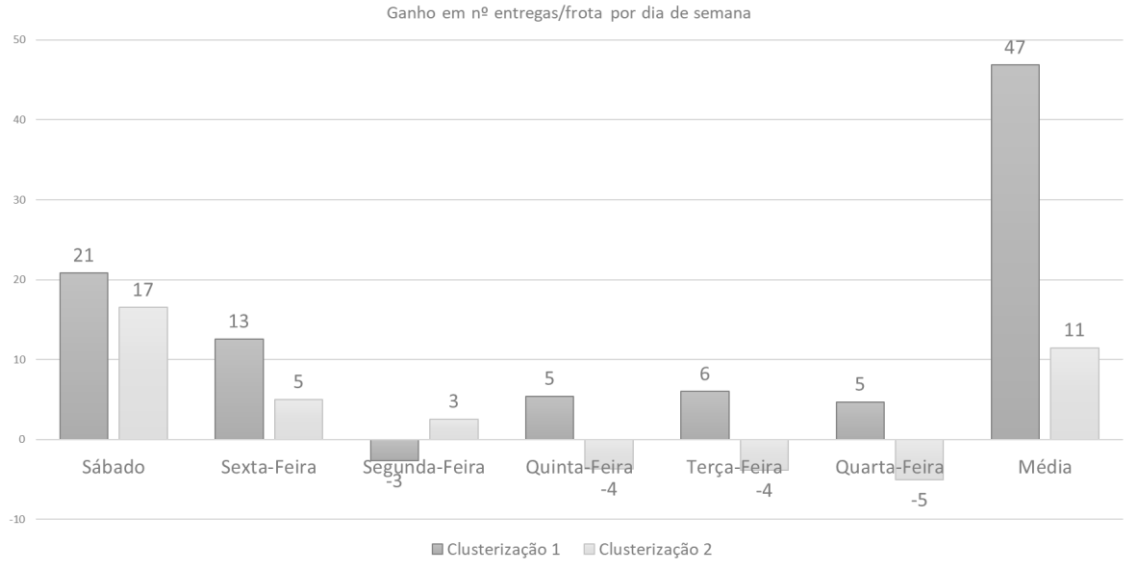
Figura 16 – Comparação da economia de tempo de deslocamento por dia da semana.



Fonte: O Autor (2021).

Com essa visualização, fica claro que o modelo de clusterização 1 consegue trazer mais retorno em relação à economia de tempo de deslocamento se comparado também com a clusterização 2. No geral, ambos métodos devem trazer retorno positivo para a empresa, a seguir, o gráfico traduz o quanto essa economia citada poderia ser transformada em mais outras entregas para a empresa.

Figura 17 - Comparação do ganho em entregas/frota por dia da semana.



Fonte: O Autor (2021).

6 CONCLUSÃO

Neste trabalho foram apresentadas duas maneiras alternativas de otimização do sistema de distribuição de bebidas de uma cervejaria. Essas alternativas visaram estritamente interesses logísticos que pudessem trazer retornos palpáveis com sua aplicação.

Através do estudo, foi possível realizar a clusterização e simular uma nova formação da programação de entrega por meio da utilização da linguagem de programação *Python* e testes diretamente realizados com a ferramenta de roteirização de entrega da cervejaria.

Foi possível avaliar os retornos diretos e indiretos que cada modelo de clusterização poderia trazer para o sistema de distribuição da cervejaria. Evidentemente, o modelo de clusterização 1 se destacou mais em seus retornos, devido ser direcionada a execução do algoritmo sobre todo o conjunto de dados de clientes de entrega, porém o modelo de clusterização também se mostrou uma alternativa de melhora para o sistema atual da empresa.

Um detalhe de relevância sobre a programação do sistema de distribuição da cervejaria é que os interesses logísticos e econômicos não são os únicos relevantes para a determinação da programação de entrega. Por se tratar de um mercado de varejo muito grande e segmentado, o perfil de clientes tem um peso grande sobre essa operação. Variáveis como o capital de giro do estabelecimento, o giro específico de cada SKU, e sua própria segmentação de mercado tendem a ser fatores de extrema relevância sobre a determinação ideal de entrega de mercadorias para um cliente.

Sendo assim, o estudo se mostrou de grande relevância como alternativa de otimização e aumento de produtividade/eficiência para o sistema de distribuição da cervejaria. Foi possível fornecer uma alternativa mais viável para a empresa em questão, trazendo clareza sobre uma oportunidade de retorno que deve ser avaliada pensando todos os pontos de peso que a otimização e mudança podem trazer.

7 REFERÊNCIAS BIBLIOGRÁFICAS

BARARONA, F.; ANBIL, R. **The Volume Algorithm: producing primal solutions with a subgradient metod**, Technical Report, IBM T. J. Watson Research Center NY 10589, 1997.

HOCHBAUM, D. **Approximation algorithms for NP-hard problems.** PWS Publishing Company, 1997.

HOFFMAN, K; PADBERG, M. **Pure zero-one linear programming problems: A computational study**, Tech. Rep., National Bureau of Standards, Gaithersburg, MD, 1985.

JOHNSON, D. **Approximation algorithms for combinatorial problems**. Journal of Computer and System Sciences 9, 256-278, 1974.

NEMHAUSER, G; WOLSEY, L. **Integer and Combinatorial Optimization**. John Wiley & Sons, Inc, 1988.

REEVES, C. **Modern Heuristic Techniques for Combinatorial Problems**; Blackwell Scientific Publications, 1993.

SIMON, P. **Too Big to Ignore: The Business Case for Big Data**; Wiley, 2013.

ESCOVEDO, Tatiana. **Introdução a Data Science: Algoritmos de Machine Learning e métodos de análise**. São Paulo: Casa do Código, 28 de fevereiro de 2020.

LLOYD, S. P. **Least square quantization in PCM**, Bell Telephone Laboratories Paper, 1957.

PINELE, J. **Geometria do Modelo Estatístico das Distribuições Normais Multivariadas**. Tese de doutorado. Campinas, 2017.

PRADO, T. C. **Segmentação de Imagens Coloridas Utilizando Técnicas de Agrupamento de Dados**. Florianópolis, Santa Catarina, 2008.

TOBERGTE, D. R.; CURTIS, S. **Python for Data Analaysis**. [s.l: s.n.]. v. 53