

**Universidade Presbiteriana Mackenzie**  
**Faculdade de Computação e Informática - Ciência de Dados**

Projeto Aplicado III – Sistema de recomendação em site de compras

Membros do grupo: Carlos Oliveira, Felipe Ferraz, Erick Isidoro

São Paulo, 30 de maio de 2025

## Resumo

Este projeto aplicado concentrou-se no desenvolvimento de sistemas de recomendação para sites de vendas, visando contribuir significativamente para o aumento das vendas e a otimização da alocação de produtos. O objetivo principal foi compreender os hábitos de compra dos clientes em sites de comércio eletrônico e sugerir produtos que são frequentemente adquiridos em conjunto.

O projeto também se alinha a objetivos extensionistas, atendendo aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, particularmente o ODS 12 (consumo e produção responsáveis), o ODS 8 (crescimento econômico inclusivo) e o ODS 9 (inovação na gestão de vendas e estoques). A integração desses objetivos busca impulsionar práticas comerciais responsáveis e socialmente conscientes.

Foram utilizados dados reais da plataforma Kaggle, Análise do e-Commerce no Brasil - Olist Dataset, abrangendo o período de 2016 a 2018 e mais de 100 mil pedidos. Os dados, disponibilizados pela Olist, foram organizados inicialmente em oito conjuntos distintos e tiveram o anonimato preservado.

A metodologia envolveu a importação, análise, cruzamento e tratamento dos dados, consolidando-os em uma única tabela. Foram realizadas análises exploratórias, como a evolução do volume de compras, gasto médio com frete por estado, volume de compras por horário, volume e quantidade de itens por consumidor, e análise da distância entre consumidores.

Para o sistema de recomendação, foi implementado o algoritmo K-Vizinhos Mais Próximos (KNN), utilizando variáveis como `order_item_id`, `product_id`, `price` e `product_category_name`. O modelo foi testado com os 10 produtos mais vendidos, e o melhor desempenho para o KNN foi obtido com  $k=15$ . Os resultados da recomendação final consistiram em uma lista dos 5 itens mais próximos em similaridade para cada um dos 10 produtos mais vendidos, acompanhada da probabilidade de venda.

Os resultados obtidos a partir do dataset da Olist foram organizados e analisados, proporcionando insights relevantes sobre o comportamento dos clientes, logística e satisfação do consumidor. Visualizações de dados foram amplamente utilizadas, mostrando distribuição geográfica, categorias mais vendidas e tempo médio de entrega.

A análise crítica destacou pontos positivos, como a utilização de técnicas robustas de Machine Learning e visualizações claras, mas também pontos negativos, pois o banco não continha a descrição dos produtos para análise previa e visual do usuário. As conclusões indicam a possibilidade de prever a indicação de produtos aos clientes e identificam oportunidades de melhorias nos resultados de vendas. Trabalhos futuros sugerem a implementação de técnicas de balanceamento de dados, teste de novos algoritmos e a ampliação da base de dados externos para análise relacionadas.

## Sumário

Resumo.....	2
Sumário .....	4
Introdução .....	5
Metodologia .....	8
1 Importação do Dataset .....	8
2 Importação das Bibliotecas.....	8
3 Análise e Cruzamento de Tabelas .....	8
4 Tratamento de Dados.....	9
5 Análises Exploratórias .....	9
6 Sistema de Recomendação com KNN .....	9
Resultado.....	11
1 Organização e Análise dos Resultados.....	11
2 Análise Crítica dos Resultados.....	11
Conclusões e trabalhos Futuros .....	13
GitHub e Youtube.....	15
Link para acesso ao projeto no GitHub e Youtube .....	15
Tabela de imagens (anexos) .....	16
Bibliografia .....	19

## Introdução

Neste projeto, trabalhamos com sistemas de recomendação para sites de vendas, uma ferramenta que contribui significativamente para o aumento das vendas, podendo, em muitos casos, representar a sobrevivência da empresa. Esses sistemas evitam o acúmulo de produtos, otimizam a alocação de espaço e direcionam esforços para produtos que, de fato, trarão resultados relevantes.

O objetivo principal do sistema é compreender os hábitos de compra dos clientes em sites de comércio eletrônico e sugerir produtos frequentemente adquiridos em conjunto. Por exemplo, ao adicionar farinha ao carrinho, o sistema pode sugerir automaticamente ingredientes como açúcar e ovos, que são comumente comprados juntos para o preparo de uma receita. Além disso, o sistema também pode sugerir receitas completas com base nos ingredientes já selecionados ou recomendar ingredientes adicionais de acordo com a preferência culinária do usuário, como pratos típicos italianos, japoneses ou brasileiros.

Este projeto aplicado considera um objetivo extensionista, atendendo às necessidades dos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, especialmente no que diz respeito ao consumo e produção responsáveis (ODS 12). Através da recomendação de produtos que otimizam o uso de recursos e reduzem o desperdício, o sistema contribui para práticas comerciais mais sustentáveis. Além disso, ao priorizar produtos de pequenos comerciantes e negócios locais, o projeto também apoia o crescimento econômico inclusivo (ODS 8) e promove a inovação na gestão de vendas e estoques (ODS 9).

A integração desses objetivos no sistema de recomendação possibilita não apenas aumentar as vendas e a satisfação do cliente, mas também impulsionar práticas comerciais responsáveis e socialmente conscientes, fortalecendo o impacto positivo do projeto tanto no âmbito econômico quanto social.

Os dados utilizados foram obtidos na plataforma Kaggle, [Análise do e-Commerce no Brasil - Olist Dataset](#), abrangendo o período de 2016 a 2018. Os dados foram disponibilizados pela Olist, uma loja de

departamentos com mais de 100 mil pedidos registrados. Por se tratar de dados reais, foram preservados o anonimato dos indivíduos, parceiros e empresas envolvidas.

Os dados foram organizados em oito conjuntos distintos: Consumidores, Vendedores, Produtos, Pedidos, Artigos dos Pedidos, Pagamento e Geolocalização.

## Referencial Teórico

Os sistemas de recomendação têm se tornado ferramentas essenciais em plataformas de comércio eletrônico, redes sociais e serviços de streaming, pois auxiliam os usuários na descoberta de produtos, conteúdos ou informações de interesse. Dentre os algoritmos utilizados nesse contexto, destaca-se o K-Vizinhos Mais Próximos (KNN), um método baseado em instâncias que avalia a similaridade entre itens ou usuários com base em características pré-definidas (RICCI; ROKACH; SHAPIRA, 2015).

O algoritmo KNN foi originalmente proposto por Cover e Hart (1967) e se caracteriza por ser não-paramétrico, realizando classificações ou previsões com base nos exemplos mais próximos em um espaço métrico. Em sistemas de recomendação, o KNN pode ser aplicado tanto na filtragem colaborativa quanto no conteúdo, identificando itens similares com base em atributos ou no comportamento de usuários (SARWAR et al., 2001).

Segundo Hastie, Tibshirani e Friedman (2009), o desempenho do KNN está diretamente relacionado à escolha da métrica de distância e à seleção de características relevantes, sendo comuns o uso das distâncias Euclidiana, Manhattan e Minkowski. Apesar de sua simplicidade, o algoritmo apresenta bons resultados em bases de dados menores ou de média escala, com o benefício adicional de não requerer um treinamento prévio complexo.

Além disso, McKinney (2010) destaca o papel de bibliotecas como o Pandas na manipulação eficiente de dados estruturados, enquanto Géron (2019) explora a aplicação do KNN em projetos práticos utilizando a biblioteca scikit-learn, amplamente adotada na comunidade científica e profissional para implementação de algoritmos de aprendizado de máquina em Python.

Estudos recentes continuam explorando variações e aplicações do KNN em sistemas de recomendação, muitas vezes combinando-o com técnicas de pré-processamento, redução de dimensionalidade e análise de agrupamentos para melhorar a performance do sistema (ZHANG et al., 2020). Isso demonstra a relevância contínua desse algoritmo como base para soluções práticas e acessíveis no campo da recomendação de produtos.

## Metodologia

O projeto envolveu uma série de decisões metodológicas com o objetivo de garantir a eficácia dos modelos testados. Abaixo, descrevemos as etapas realizadas no pipeline:

Fluxo das etapas.



### 1 Importação do Dataset

Foi realizada a importação do dataframe disponibilizado na plataforma Kaggle, no repositório `olistbr/brazilian-ecommerce`.

### 2 Importação das Bibliotecas

Utilizamos as bibliotecas `pandas`, `numpy`, `seaborn`, `matplotlib`, `warnings`, `folium` e `sklearn.cluster` para análise, visualização e modelagem dos dados.

### 3 Análise e Cruzamento de Tabelas

As tabelas separadas inicialmente — `customers`, `geolocation`, `order_items`, `orders`, `products`, `sellers` e `reviews` — foram gradualmente cruzadas. Em cada etapa, selecionamos



apenas as colunas de interesse, resultando em uma tabela consolidada com 113.007 linhas e 28 colunas.

#### 4 Tratamento de Dados

No tratamento da tabela consolidada:

- a. Removemos os registros de 2016 (apenas 375 registros, não cobrindo o ano completo).
- b. Também excluimos setembro de 2018, que continha apenas um registro.
- c. Convertemos as colunas de data para o tipo `datetime`.
- d. Criamos duas colunas: uma para o tempo de entrega e outra para a codificação do `product_id` (com Label Encoding).

Ao final do tratamento, o *dataframe* ficou com 112.632 linhas e 30 colunas.

#### 5 Análises Exploratórias

A seguir, realizamos diversas análises sobre os dados tratados:

- e. Evolução do volume de compras ao longo do tempo (**Anexo - Gráfico I**);
- f. Gasto médio com frete por estado e volume de compras por estado (**Anexo - Gráfico II**);
- g. Volume de compras por horário, destacando os períodos com maior concentração de pedidos (**Anexo - Gráfico III**);
- h. Volume e quantidade de itens por consumidor;
- i. Análise da distância entre consumidores, com base em latitude e longitude de 100 clientes selecionados (**Anexo - Imagem II**).

#### 6 Sistema de Recomendação com KNN

Para o sistema de recomendações, utilizamos as seguintes variáveis:

- j. `order_item_id` (identificador único do item no pedido),
- k. `product_id` (identificador único do produto),
- l. `price` (preço),

m. `product_category_name` (categoria do produto).

Selecionamos os 10 produtos mais vendidos para testar o modelo. Utilizamos *Label Encoding* para transformar os `product_id` em valores numéricos. Em seguida:

- n. Dividimos os dados entre treino e teste;
- o. Definimos como variáveis preditoras (X) o `order_item_id` e o `price`, e como variável-alvo (y) o `product_id` codificado;
- p. Testamos o modelo KNN com  $k = 5$  e  $k = 10$ , mas os melhores resultados foram obtidos com  $k = 15$ ;
- q. Também testamos outras composições de variáveis preditoras e alvo, mas todas apresentaram desempenho inferior.

Para a recomendação final, aplicamos o modelo aos 10 produtos mais vendidos. Geramos, para cada um deles, uma lista com os 5 itens mais próximos em termos de similaridade, acompanhados da probabilidade de venda. Os resultados foram consolidados em uma planilha Excel para facilitar a visualização (**Anexo - Imagem III**).

## **Resultado**

### **1 Organização e Análise dos Resultados**

Com a construção do modelo finalizada, foram organizados e analisados os resultados obtidos a partir do dataset da Olist. As análises descritivas e preditivas proporcionaram insights relevantes sobre o comportamento de compra dos clientes, logística e satisfação do consumidor.

Técnicas de visualização de dados foram amplamente utilizadas, incluindo gráficos que mostraram:

- A distribuição geográfica dos clientes.
- As categorias de produtos mais vendidos.
- O tempo médio de entrega em diferentes regiões.
- O desempenho dos modelos preditivos, através de matrizes de confusão e métricas como Acurácia e F1-Score.

Quando comparados com outras abordagens presentes no referencial teórico, os modelos utilizados apresentaram desempenho satisfatório, especialmente o KNN , que demonstrou maior capacidade de recomendação.

### **2 Análise Crítica dos Resultados**

#### **Pontos Positivos:**

Utilização de técnicas robustas de Machine Learning que possibilitaram boas taxas de acurácia.

Visualizações claras, facilitando a compreensão dos dados e dos resultados, obtendo o processo completo de coleta, tratamento, modelagem e avaliação.

#### **Pontos Negativos:**

O banco de dados não continha a descrição dos produtos, principalmente para manter a confidencialidade dos dados disponibilizados para estudos, impossibilitando uma análise mais criteriosa no entendimento de cotidiano para os departamentos de vendas em seguimento relacionados ao estudo, podendo ser validado com um conhecimento prévio dos itens que podem ser vendidos juntos.

## Conclusões e trabalhos Futuros

### Recapitulação:

O trabalho teve como foco o desenvolvimento de um pipeline completo para análise e indicação de vendas a um cliente em um ambiente de e-commerce, utilizando o dataset da Olist. Foram realizadas etapas de integração e tratamento de dados, análise exploratória e modelagem com algoritmos recomendado para sistemas de indicação.

### Conclusão:

Os resultados confirmaram que é possível prever, com boa acurácia, a indicação de produtos correlatos aos clientes com base em características de perfil de aquisição do consumidor. O Modelo KNN destacou-se como o modelo mais eficiente trazendo resultados com forte probabilidade de indicação.

Segue tabela demonstrando os resultados obtidos com o modelo e a chance de venda dos itens indicados, a tabela contém os top 10 itens mais vendidos com a indicação de 5 produtos para cada um deles, contendo a categoria de vendas de ambos os produtos.

Top Produto	Top 10 Produto	Categoria do Produto	Produto Recomendado	Categoria do Recomendado	Chance (%)
1	aca2eb7d00ea1a7b8ebd4e68314663af	moveis_decoracao	30d1842be9d62546d3fb853119602ac2	eletrodomesticos_2	87,5
1	aca2eb7d00ea1a7b8ebd4e68314663af	moveis_decoracao	fe5827d110a9ba59d4ab407e86ccd6eb	construcao_ferramentas_jardim	65,52
1	aca2eb7d00ea1a7b8ebd4e68314663af	moveis_decoracao	ee8b16196604ed47b34b094351ce68e2	bebes	73,45
1	aca2eb7d00ea1a7b8ebd4e68314663af	moveis_decoracao	cc9c93a7d06ba4b5913d8f3fba62612e	esporte_lazer	72,78
1	aca2eb7d00ea1a7b8ebd4e68314663af	moveis_decoracao	43423cdfde7fda63d0414ed38c11a73	relogios_presentes	55,4
2	99a4788cb24856965c36a24e339b6058	cama_mesa_banho	8ac47b3ab13d68f49f10dde899674149	dvds_blu_ray	58,48
2	99a4788cb24856965c36a24e339b6058	cama_mesa_banho	437d05a395e9e47f9762e677a7068ce7	beleza_saude	78,04
2	99a4788cb24856965c36a24e339b6058	cama_mesa_banho	f30de5fdde000c5deb03cc49d782249	casa_conforto	94,07
2	99a4788cb24856965c36a24e339b6058	cama_mesa_banho	7c1e2b3fa0233e46fb3bcdcb9919a72f	papelaria	51,91
2	99a4788cb24856965c36a24e339b6058	cama_mesa_banho	a6ad77b15e566298a4e8ee2011ab1255	moveis_decoracao	70,13
3	422879e10f46682990de24d770e7f83d	ferramentas_jardim	b9142260cefbdb5688748061179bb7fe	cama_mesa_banho	88,46
3	422879e10f46682990de24d770e7f83d	ferramentas_jardim	167b19e93bacb17916b9a6dd03264e7	eletronicos	57,93
3	422879e10f46682990de24d770e7f83d	ferramentas_jardim	6413f7a28e149a324c4a914000399fb2	cool_stuff	62,25
3	422879e10f46682990de24d770e7f83d	ferramentas_jardim	17606c7d7254ed1f0351fd48a28be932	cama_mesa_banho	82,92
3	422879e10f46682990de24d770e7f83d	ferramentas_jardim	54d9ac713e253fa1fae9c8003b011c2a	cool_stuff	62,06
4	389d119b48cf3043d311335e499d9c6b	ferramentas_jardim	b7605b5b483063d12bd90a772bfff9d21	cama_mesa_banho	87,84
4	389d119b48cf3043d311335e499d9c6b	ferramentas_jardim	362b773250263786dd58670d2df42c3b	esporte_lazer	57,39
4	389d119b48cf3043d311335e499d9c6b	ferramentas_jardim	ea546947e2412f88a20ac74103592b42	beleza_saude	67,77
4	389d119b48cf3043d311335e499d9c6b	ferramentas_jardim	777d2e438a1b645f3aec9bd57e92672c	cama_mesa_banho	61,59
4	389d119b48cf3043d311335e499d9c6b	ferramentas_jardim	9007d9a8a0d332d61d9dd611fa341f4b	papelaria	56,97
5	368d6c730842d78016ad823897a372db	ferramentas_jardim	b9142260cefbdb5688748061179bb7fe	cama_mesa_banho	82,49
5	368d6c730842d78016ad823897a372db	ferramentas_jardim	167b19e93bacb17916b9a6dd03264e7	eletronicos	81,91
5	368d6c730842d78016ad823897a372db	ferramentas_jardim	6413f7a28e149a324c4a914000399fb2	cool_stuff	84,62
5	368d6c730842d78016ad823897a372db	ferramentas_jardim	17606c7d7254ed1f0351fd48a28be932	cama_mesa_banho	88,68
5	368d6c730842d78016ad823897a372db	ferramentas_jardim	54d9ac713e253fa1fae9c8003b011c2a	cool_stuff	51,16

6	53759a2ecddad2bb87a079a1f1519f73	ferramentas_jardim	b2fb8ccacff645e701cd692064ad09c9	cama_mesa_banho	67,64
6	53759a2ecddad2bb87a079a1f1519f73	ferramentas_jardim	e0d64dcfaa3b6db5c54ca298ae101d05	relogios_presentes	59,95
6	53759a2ecddad2bb87a079a1f1519f73	ferramentas_jardim	34d6306314a47fe15d1efdb23302020	informatica_acessorios	63,41
6	53759a2ecddad2bb87a079a1f1519f73	ferramentas_jardim	51250f90d798d377a1928e8a4e2e9ae1	perfumaria	71,83
6	53759a2ecddad2bb87a079a1f1519f73	ferramentas_jardim	1b9c9d11fda157197ea45527ffabae2	esporte_lazer	51,09
7	d1c427060a0f73f6b889a5c7c61f2ac4	informatica_acessorios	df892547f0d5bb0c35cf26ab0e93d8d	fashion_bolsas_e_acessorios	84,41
7	d1c427060a0f73f6b889a5c7c61f2ac4	informatica_acessorios	2b25ef4ad1788d699dff628cbbfbfd0	utilidades_domesticas	64,05
7	d1c427060a0f73f6b889a5c7c61f2ac4	informatica_acessorios	0f9df08ca6757f6d5158b71f5d307724	casa_construcao	82,34
7	d1c427060a0f73f6b889a5c7c61f2ac4	informatica_acessorios	84f5c4f480ad6c999d6a6860f1a2e41	alimentos_bebidas	70,92
7	d1c427060a0f73f6b889a5c7c61f2ac4	informatica_acessorios	b9ee7519d0187d2389af62ba6c612963	esporte_lazer	84,87
8	53b36df67ebb7c41585e8d54d6772e08	relogios_presentes	e7e04378fa044f94470c5486c896ed5	esporte_lazer	51,02
8	53b36df67ebb7c41585e8d54d6772e08	relogios_presentes	21fb5057dd6a737df6851a7ab7a130da	moveis_decoracao	57,82
8	53b36df67ebb7c41585e8d54d6772e08	relogios_presentes	de72ea95a1cd05b1a597d294a425be57	automotivo	61,08
8	53b36df67ebb7c41585e8d54d6772e08	relogios_presentes	35b0c77029697004675d3a2ce505916	informatica_acessorios	54,57
8	53b36df67ebb7c41585e8d54d6772e08	relogios_presentes	23365beed316535b4105bd800c46670e	telefonica	60,87
9	154e7e31ebfa092203795c972e5804a6	beleza_saude	e0cf79767c5b016251fe139915c59a26	beleza_saude	94,04
9	154e7e31ebfa092203795c972e5804a6	beleza_saude	0347f9daa0a33f82b4c9d821450cd8a	malas_acessorios	93,97
9	154e7e31ebfa092203795c972e5804a6	beleza_saude	82e4ad16521ca131d95e198d507db370	eletronicos	55,26
9	154e7e31ebfa092203795c972e5804a6	beleza_saude	1acb18f0869c5489d5f76abcd681165e	esporte_lazer	91,05
9	154e7e31ebfa092203795c972e5804a6	beleza_saude	cb92637d91f8267856d2143b6fc7bd9e	esporte_lazer	90,13
10	3dd2a17168ec895c781a9191c1e95ad7	informatica_acessorios	e44f675b60b3a3a2453ec36421e06f0f	esporte_lazer	57,53
10	3dd2a17168ec895c781a9191c1e95ad7	informatica_acessorios	73fbb5c7d95425e37d003af6ee943997	cama_mesa_banho	59,4
10	3dd2a17168ec895c781a9191c1e95ad7	informatica_acessorios	d017a2151d543a9885604dc62a3d9dcc	fashion_bolsas_e_acessorios	93,96
10	3dd2a17168ec895c781a9191c1e95ad7	informatica_acessorios	fb6782985a98aa8a59238f58239f6f1e	informatica_acessorios	62,44
10	3dd2a17168ec895c781a9191c1e95ad7	informatica_acessorios	edb58f769af494909a8851152c3a2d1d	cama_mesa_banho	74,54

### Melhorias no Projeto:

- Implementação de técnicas de balanceamento de dados.
- Teste de novos algoritmos, como Gradient Boosting e XGBoost.
- Refinamento dos hiperparâmetros utilizando Grid Search.

### Trabalhos Futuros:

- Ampliação da base de dados com variáveis externas, como indicadores econômicos.
- Desenvolvimento de um sistema automatizado de recomendação de produtos utilizando outros algoritmos.
- Aplicação do modelo preditivo em tempo real, integrando-o a sistemas de CRM.

## GitHub e Youtube

### Link para acesso ao projeto no GitHub e Youtube

Abaixo tem-se o link do GitHub onde estão compartilhados, dataset, cronograma, script outros documentos atualizados do projeto:

Demonstração da imagem do cronograma em tabelas de imagem (Imagem 1 – Cronograma de Atividade), arquivo disponibilizado em link Github.

1 – Link GithHub

[https://github.com/Ferraz0Felipe/Projeto\\_Aplicado\\_III.git](https://github.com/Ferraz0Felipe/Projeto_Aplicado_III.git)

2 – Link Youtube

<https://www.youtube.com/watch?v=vsJ3CSjwoLM>

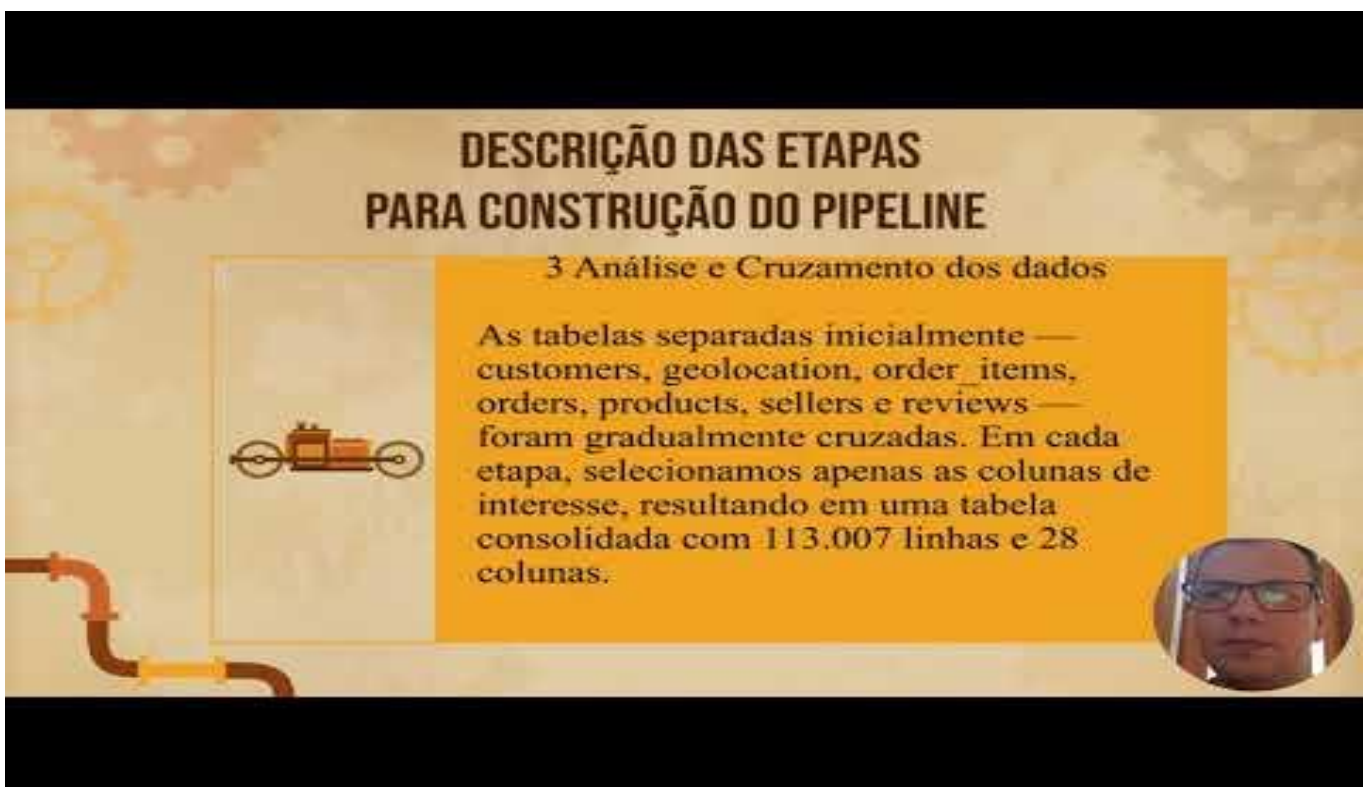


Tabela de imagens (anexos)

Imagem I – Cronograma de Atividade (disponível na integra link Github)

									Concluído	Programado	Atrasado																					
									Fevereiro																							
Etapa	Título	Atividades	Responsável	Início	Término	Duração (dias)	Milestones	Status	16	17	18	19	20	21	22	23	24	25	26	27	28	29	01	02	03	04	05	06	07	08	09	10
1	Concepção do Produto	Organizar o grupo	Carlos	2-fev-2025	4-mar-2025	15	6-mar-2024	Concluído																								
		Escolher Tema	Erick	2-fev-2025	4-mar-2025	3	6-mar-2024	Concluído																								
		Escolher a base de dados	Felipe	2-fev-2025	4-mar-2025	5	6-mar-2024	Concluído																								
		definir o cronograma do projeto	Erick	2-fev-2025	4-mar-2025	4	6-mar-2024	Concluído																								
		Elabora o documento inicial	Carlos	2-fev-2025	4-mar-2025	4	6-mar-2024	Concluído																								
2	Definição do produto	Analisar a base de dados	Erick	11-mar-2025	10-abr-2025	24	3-abr-2024	Concluído																								
		limpar e preparar a base de dados	Felipe	11-mar-2025	10-abr-2025	16	3-abr-2024	Concluído																								
		Escolher a técnica para treinamento do modelo	Carlos	11-mar-2025	10-abr-2025	4	3-abr-2024	Concluído																								
		construir uma prova de conceito	Carlos	11-mar-2025	10-abr-2025	10	28-abr-2024	Concluído																								
		Definir métricas de avaliação de desempenho	Erick	11-mar-2025	10-abr-2025	12	28-abr-2024	Concluído																								
3	Metodologia	Implementar a técnica proposta	Felipe	3-abr-2025	3-mai-2025	8	28-abr-2024	Concluído																								
		Ajustar o pipeline de dados	Carlos	3-abr-2025	3-mai-2025	5	28-abr-2024	Concluído																								
		Documentar os passos implementados	Erick	3-abr-2025	3-mai-2025																											
4	Resultados e Conclusão	Organizar e Documentar os resultados	Carlos	2-mai-2025	1-jun-2025	11	17-mai-2024	Concluído																								
		Finalizar Documento do projeto	Erick	2-mai-2025	1-jun-2025	10	17-mai-2024	Concluído																								
		Entregar Materiais produzidos e apresentação	Felipe	2-mai-2025	1-jun-2025	8	17-mai-2024	Concluído																								

Gráfico I - evolução do volume de compras realizadas no decorrer do tempo

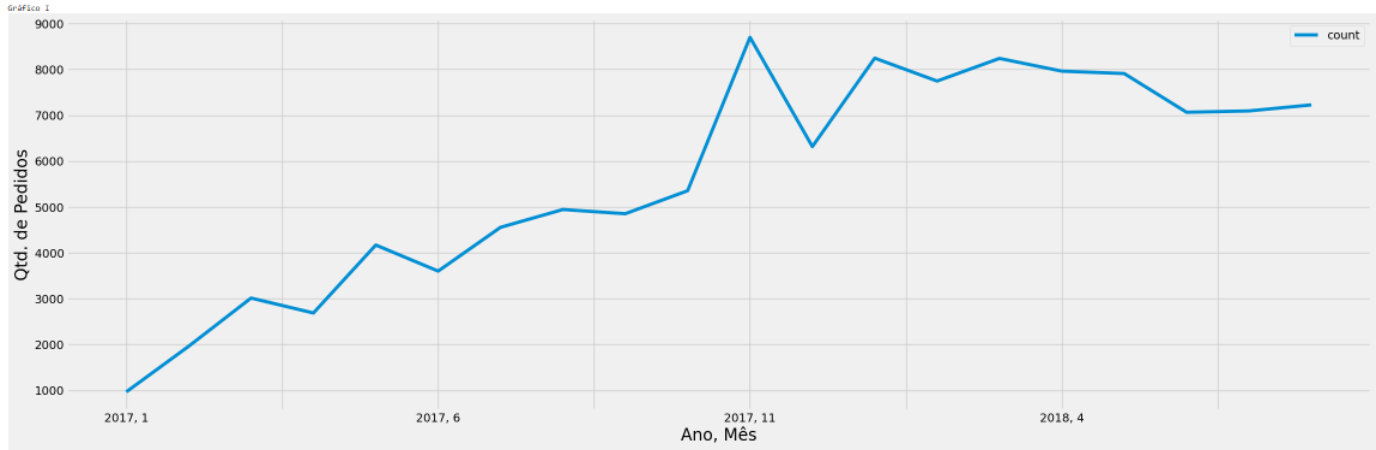


Gráfico II – Volume de vendas por estado e valor médio de gastos com frete por estado.



Gráfico II

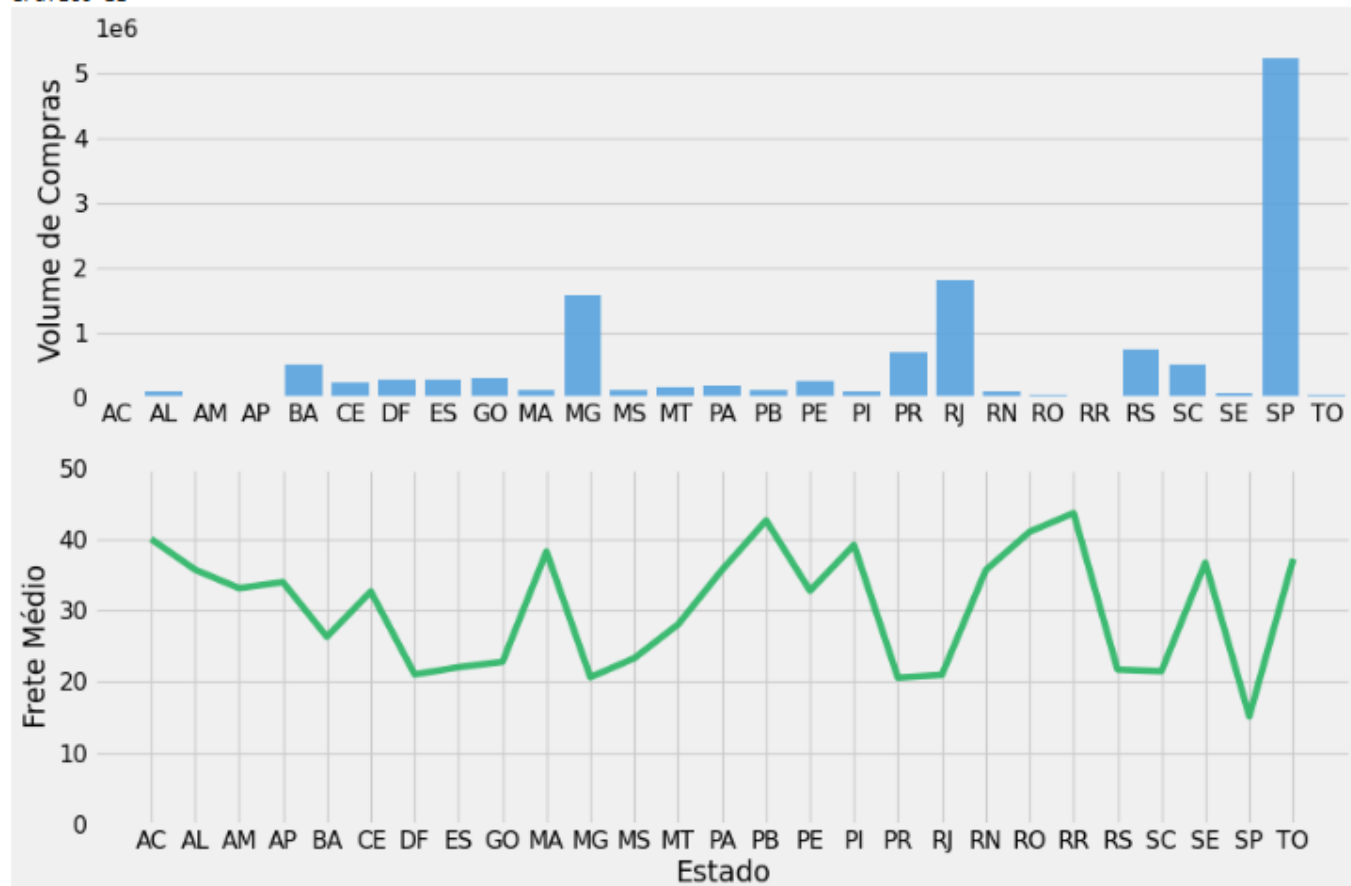


Gráfico III – Volume de vendas por horas.

Gráfico III

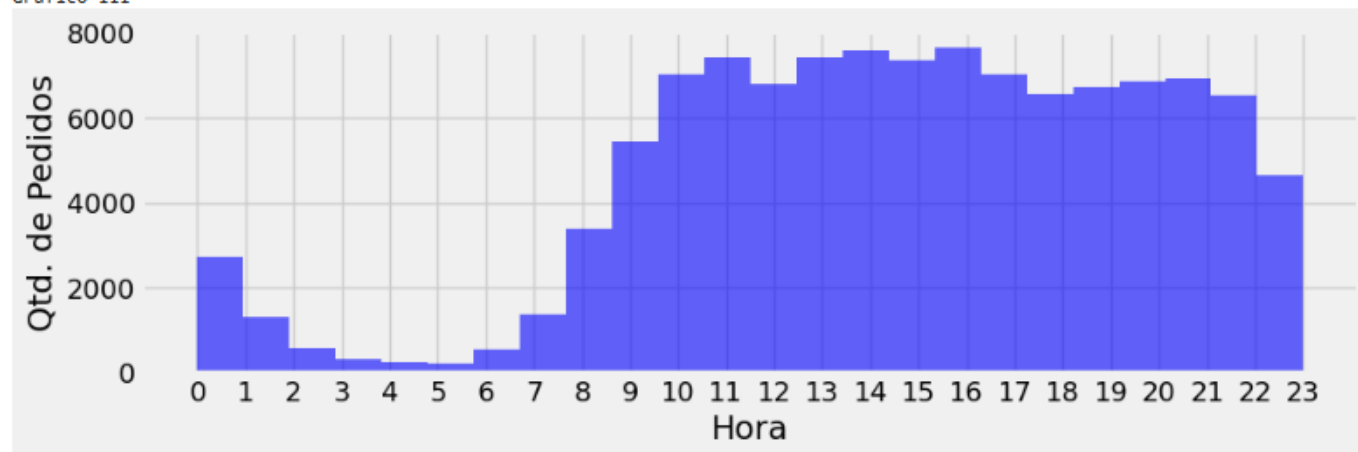


Imagem II - Análise de distância entre consumidores.

Cluster -1: Ciano  
Cluster 0: Rosa  
Cluster 1: Vermelho  
Cluster 2: Azul  
Cluster 3: Laranja  
Cluster 4: Marrom

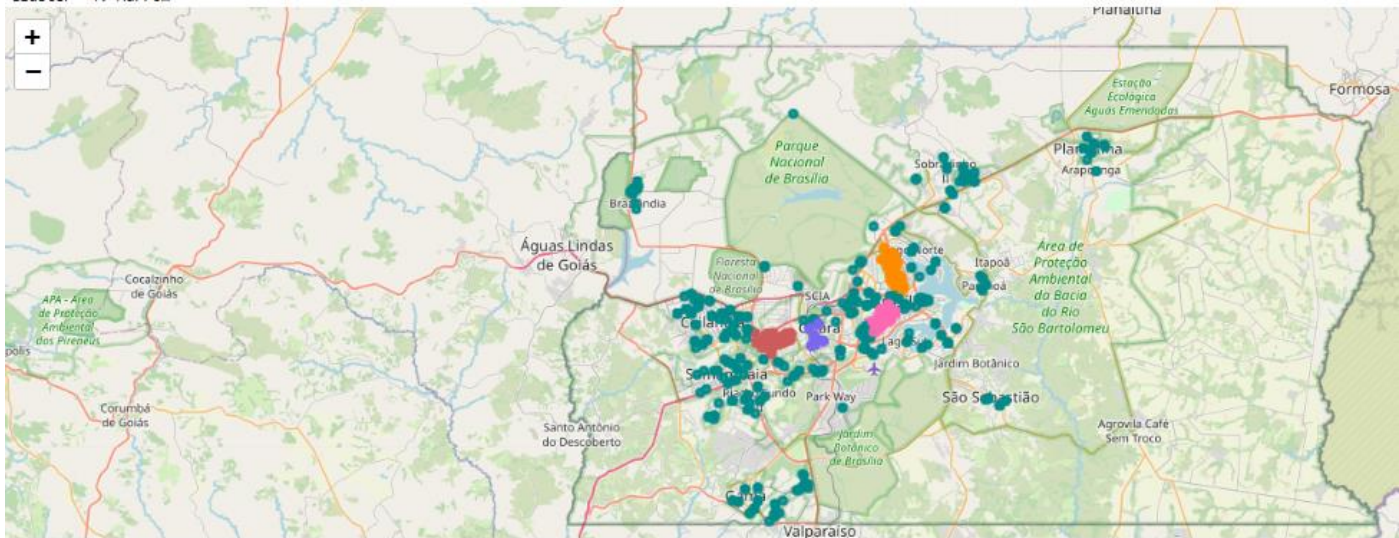


Imagem III - demonstração de Excel (disponível na integra link Github).

D1	Categoria do Recomendado				
A	B	C	D	E	F
Top 10 Produto	Categoria do Produto	Produto Recomendado	Categoria do Recomendado	Chance (%)	
aca2eb7d00ea1a7b8e	moveis_decoracao	30d1842be9c62546d3fb853119602ac2	eletrodomesticos_2	87,5	
aca2eb7d00ea1a7b8e	moveis_decoracao	fe5827d110a9ba59d4ab407e86ccd6eb	construcao_ferramentas_jardim	65,52	
aca2eb7d00ea1a7b8e	moveis_decoracao	ee8b16196604ed47b34b094351ce68e2	bebes	73,45	
aca2eb7d00ea1a7b8e	moveis_decoracao	cc9c93a7dc6ba4b5913c8f3fba62612e	esporte_lazer	72,78	
aca2eb7d00ea1a7b8e	moveis_decoracao	43423cdfde7fda63d0414ed38c11a73	relogios_presentes	55,4	
99a4788cb24856965c3f	cama_mesa_banho	8ac47b3ab13c68f49f10dde899674149	dvds_blu_ray	58,48	
99a4788cb24856965c3f	cama_mesa_banho	437c05a395e9e47f9762e677a7068ce7	beleza_saude	78,04	
99a4788cb24856965c3f	cama_mesa_banho	f30de5fdde000c5debf03cc49d782249	casa_conforto	94,07	
99a4788cb24856965c3f	cama_mesa_banho	7c1e2b3fa0233e46fb3bcdcb9919a72f	papelaria	51,91	
99a4788cb24856965c3f	cama_mesa_banho	a6ad77b15e566298a4e8ee2011ab1255	moveis_decoracao	70,13	
422879e10f46682990d	ferramentas_jardim	b9142260ceffb5688748061179bb7fe	cama_mesa_banho	88,46	
422879e10f46682990d	ferramentas_jardim	167b19e93baccb17916b9a6dd03264e7	eletronicos	57,93	
422879e10f46682990d	ferramentas_jardim	6413f7a28e149a324c4a914000399fb2	cool_stuff	62,25	
422879e10f46682990d	ferramentas_jardim	17606c7d7254ed1f0351fd48a28be932	cama_mesa_banho	82,92	
422879e10f46682990d	ferramentas_jardim	54d9ac713e253fa1fae9c8003b011c2a	cool_stuff	62,06	
389d119b48cf3043d31	ferramentas_jardim	b7605b5b483063d12bd90a772bfff9d21	cama_mesa_banho	87,84	
389d119b48cf3043d31	ferramentas_jardim	362b773250263786dd58670d2df42c3b	esporte_lazer	57,39	
389d119b48cf3043d31	ferramentas_jardim	ea546947e2412f88a20ac74103592b42	beleza_saude	67,77	
389d119b48cf3043d31	ferramentas_jardim	777d2e438a1b645f3aec9bd57e92672c	cama_mesa_banho	61,59	
389d119b48cf3043d31	ferramentas_jardim	9007d9a8a0d332c61d9dd611fa341f4b	papelaria	56,97	
368c6c730842d78016ac	ferramentas_jardim	b9142260ceffb5688748061179bb7fe	cama_mesa_banho	82,49	
368c6c730842d78016ac	ferramentas_jardim	167b19e93baccb17916b9a6dd03264e7	eletronicos	81,91	
368c6c730842d78016ac	ferramentas_jardim	6413f7a28e149a324c4a914000399fb2	cool_stuff	84,62	
368c6c730842d78016ac	ferramentas_jardim	17606c7d7254ed1f0351fd48a28be932	cama_mesa_banho	88,68	

## Bibliografia

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21-27, 1967.

GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2. ed. Sebastopol: O'Reilly Media, 2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009.

MCKINNEY, W. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference*. 2010.

RICCI, F.; ROKACH, L.; SHAPIRA, B. *Recommender Systems Handbook*. 2. ed. Springer, 2015.

MATIAS-PEREIRA, J. *Manual de metodologia da pesquisa científica*. 4. ed. São Paulo: Atlas, 2019.

WAZLAWICK, R. S. *Metodologia de pesquisa para ciência da computação*. 3. ed. Rio de Janeiro: LTC, 2021.