

CURSO: Ciência de Dados

POLO DE APOIO PRESENCIAL: Santana

SEMESTRE: 2º

COMPONENTE CURRICULAR / TEMA:

NOME COMPLETO DO ALUNO: Enzo Vemado

TIA: 23020911

NOME DO PROFESSOR:

## **Análise Exploratória de Dados Olist Store**

### **1. Apresentação do Grupo**

Enzo Vemado, líder do projeto, que ao assumir os múltiplos papéis necessários para levar este projeto adiante, fui responsável por cada etapa do processo: desde a formulação da questão de pesquisa, passando pela coleta e limpeza dos dados, até a análise exploratória e interpretação dos resultados. Este esforço solo foi tanto um desafio quanto uma oportunidade de demonstrar minha capacidade de conduzir uma pesquisa independente e produzir insights significativos para o campo do e-commerce.

### **2. Nome do Projeto**

Desvendando Padrões de Comportamento de Consumo: Olist Store

### **3. Empresa/Organização de Estudo**

A empresa foco deste estudo é a Olist Store, uma destacada plataforma de vendas no mercado brasileiro que permite a comerciantes de diversos tamanhos e segmentos anunciar seus produtos para venda. A Olist se posiciona como uma facilitadora no e-

commerce, combinando a variedade e a conveniência de um marketplace com a qualidade de serviço de uma loja de varejo boutique.

Fundada em 2015, a Olist tem desempenhado um papel vital na transformação do cenário do e-commerce no Brasil. Com uma proposta de valor única, ela ajuda pequenos e médios varejistas a competir em um ambiente dominado por grandes players, proporcionando-lhes acesso a uma base de clientes mais ampla e serviços logísticos e de marketing de qualidade superior.

No entanto, apesar de seu sucesso e crescimento contínuo, a Olist enfrenta desafios intrínsecos à dinâmica do mercado online. Um desses desafios é entender profundamente o comportamento dos consumidores. Este estudo se concentra em desvendar os padrões e tendências subjacentes nas interações dos clientes com a plataforma, visando otimizar as estratégias de vendas e marketing da empresa. Através de uma análise rigorosa dos dados disponíveis, este projeto busca não só realçar oportunidades para a Olist mas também contribuir com insights valiosos que poderão ser aplicados ao setor de e-commerce como um todo.

#### **4. Área do Problema**

O e-commerce no Brasil tem visto um crescimento sem precedentes. No entanto, com o aumento do volume de vendas, surgiram desafios significativos relacionados à compreensão do comportamento do consumidor, otimização da jornada do cliente, e retenção de clientes.

#### **5. Descrição do Problema / Gap**

Apesar dos avanços na infraestrutura de e-commerce, há uma lacuna evidente no entendimento dos padrões de compra dos consumidores. A falta de insights acionáveis resulta em campanhas de marketing mal direcionadas, baixa retenção de clientes e, por fim, perda de receita.

#### **6. Proposta Analítica**



Este projeto propõe uma abordagem analítica para desvendar os padrões de comportamento de consumo através de técnicas de análise de dados avançadas. Alavancando os dados disponíveis, pretendemos extrair insights que poderão orientar estratégias de marketing e tomadas de decisão baseadas em dados.

## 7. Dados Disponíveis

Os dados para esta análise foram obtidos de [www.kaggle.com](http://www.kaggle.com). Eles abrangem:

- Informações dos clientes e localização
- CEPs e coordenadas
- Itens de cada transação
- Informações do pagamento de cada transação
- Reviews realizados pelos clientes
- Dados de cada transação
- Informações sobre cada produto
- Dados sobre os vendedores
- Nomes de categoria

## 8. Análise Exploratória

Utilizando scripts Python, conduzimos uma análise exploratória abrangente. Os scripts, que são parte integral deste estudo, estão disponíveis em nosso repositório GitHub e podem ser acessados através do seguinte link: [https://github.com/vemado/PROJE\\_APLIC\\_I/blob/main/LTV\\_brazilian\\_ecommerce.ipynb](https://github.com/vemado/PROJE_APLIC_I/blob/main/LTV_brazilian_ecommerce.ipynb)

Esta análise incluiu:

### Variáveis que foram consideradas:

- order\_id (O indicador único do pedido);
- customer\_id (O indicador de qual cliente);
- order\_status (O indicador se o pedido tinha sido finalizado ou não);
- order\_purchase\_timestamp;
- order\_approved\_at (Quanto o pedido foi aprovado);

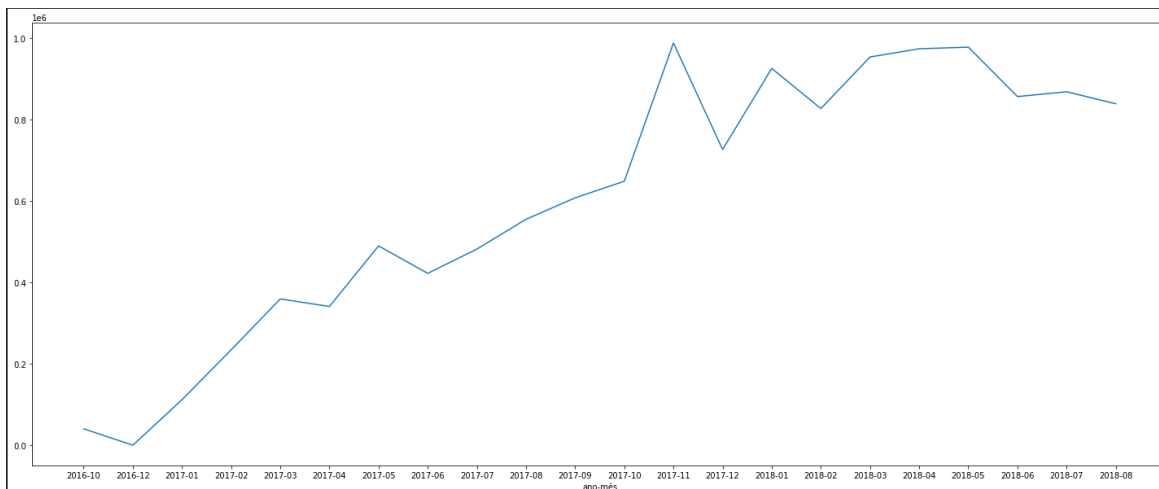


## Universidade Presbiteriana Mackenzie

- order\_delivered\_carrier\_date (Quando o pedido foi despachado);
- order\_delivered\_customer\_date (Quando o pedido foi entregue);
- order\_estimated\_delivery\_date (Previsão de entrega);
- payment\_sequential (Quantidade de tipo de pagamento);
- payment\_type (Tipo de pagamento);
- payment\_installments (Quantidade de parcelas);
- payment\_value (Pagamento total por transação);
- price (Valor da venda);
- freight\_value (Valor do frete).

**Visualizações de Dados:** Gráficos elucidativos foram utilizados para detalhar 9. Resultados Pretendidos

Receita líquida mensal:

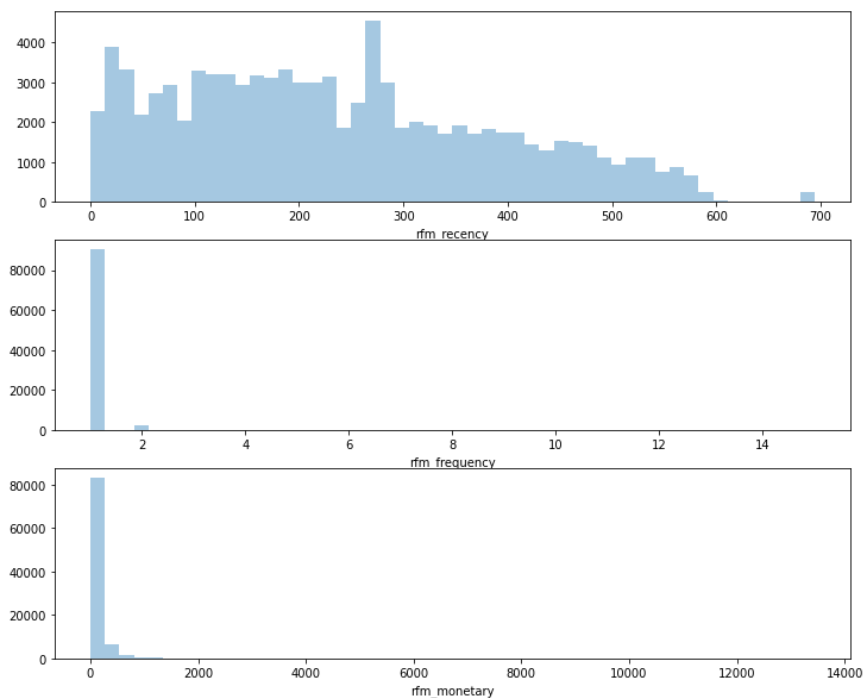


RFM, inicialmente é realizado a análise RFM dos clientes, que se refere a:

Recência: Tempo entre última transação e fim da base

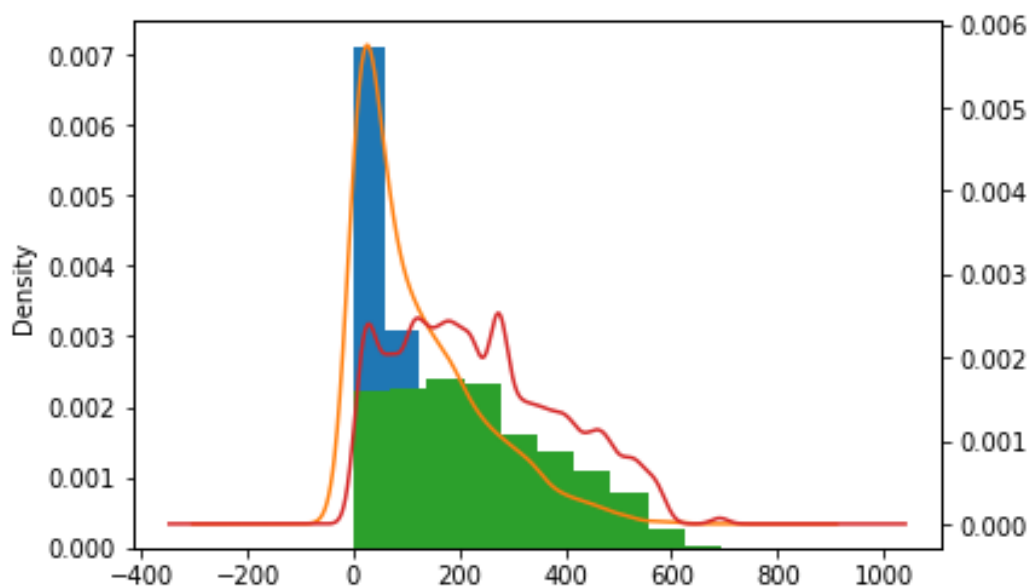
Frequência: Quantidade de transações realizadas

Valor Monetário: Valor gerado pelo cliente



Como não existe um critério para determinação da "morte" dos clientes, é realizado uma análise para encontrar a quantidade de dias sem transacionar onde o cliente tem a maior probabilidade de abandonar o ecommerce (Churn).

O Churn foi o método usado para determinar se o cliente ainda está vivo na base ou não, baseado no cruzamento da recência máxima de clientes recorrentes e da recência histórica da base completa.



Com esta análise, pretendemos alcançar uma compreensão clara dos padrões de comportamento de consumo, que poderá ser utilizada para aprimorar estratégias de marketing, otimizar a jornada do cliente e, em última análise, aumentar a retenção de clientes. Insights específicos e recomendações serão fornecidos como resultado final deste projeto.