

# **DOCUMENTAÇÃO DO PROJETO FINAL EBAC**

**Passo a Passo da Análise do E-Commerce da Olist Store**

# COLETA DE DADOS

- Projeto criado utilizando a fonte de dados do E-Commerce da Olist Store, disponível no [www.kaggle.com](https://www.kaggle.com)
- Download direto de 5 planilhas no formato CSV; atendendo os requisitos mínimos de informação de data, linhas e colunas para o projeto.
- Todas as tabelas foram selecionadas cuidadosamente para fazer o cruzamento dos dados.

# TRATAMENTO DE DADOS

- Os campos de data e preços foram corretamente formatados no notebook usando PySpark. Data para timestamp e Preço para double.
- Novas colunas foram criadas a partir dos dados, como: Mês, Ano, Hora, Dia da Semana e Dias de Entrega. Tudo para preparar os dados e extrair o máximo de informações.
- Valores ausentes foram verificados no Excel e no Google Colab.

# EXPLORAÇÃO DOS DADOS (EDA)

- A exploração foi feita por inteiro usando o PySpark no Colab.
- Os dados no geral foram extraídos a partir das funções: count, groupBy, orderBy e etc.
- Uso de funções de agregação, como: soma , média, desvio padrão, mínimo, máximo e etc.
- Testes estatísticos como regressão linear simples e correlação foram utilizados para extrair informações.

# INSIGHTS DE NEGÓCIO

- Houve um volume de R\$ 16 milhões em vendas e 99,4 mil pedidos, um resultado aproximadamente 24 % acima da expectativa (meta fictícia).
- A categoria de maior valor pago e maior quantidade de pedidos foi “Cama, Mesa e Banho”; demonstrando grande interesse do público por produtos de algodão e materiais similares.
- A taxa de clientes recorrentes (12,5 mil) em relação aos clientes únicos (86,2 mil) é de aproximadamente 14%, que é considerada baixa para um E-Commerce. A maior concentração está presente na região Sudeste, em especial SP e RJ.
- A média de atrasos é -12,04, o que sugere que os pedidos em geral chegam antes do prazo. Mês de março tem o maior pico de atraso.

# INSIGHTS ESTATÍSTICOS

- O ticket médio por pedido (R\$ 160) apresentado se mostra bem distante do seu desvio padrão (R\$ 620), mostrando grande diferença entre os perfis de compra que se distanciam muito da média.
- A média de parcelamento foi maior para “Computadores”, o que pode demonstrar ser a categoria menos acessível financeiramente.
- A regressão linear simples feita com o preço do produto e frete e seu impacto no pagamento, mostra um coeficiente de 1 ponto de influência em cada um e um intercepto de 21.78 (os pedidos sem itens ou frete, ainda custariam R\$ 21,78).
- A correlação entre o peso do produto e o frete é de 0.55, considerada correlação positiva moderada.

# CONCLUSÕES / SUGESTÕES

1. Dar atenção às categorias que vendem menos, como “Ferramentas de Jardim” e “Automotivo”.
2. Fazer programas de incentivo como “Cashback pra quem volta” e outros similares para aumentar a taxa de clientes que compram mais de uma vez.
3. Ficar atento ao motivo dos atrasos nos pedidos em março, se pode estar relacionado a períodos festivos, como o Carnaval.
4. Entender os diferentes perfis de compra dos clientes e como aumentar as compras dos perfis menores.
5. Alternativas cabíveis para diminuir parcelamento ou aumentar as vendas de “Computadores.”

# **OBRIGADO PELA ATENÇÃO!**

O projeto demonstrou como a análise de dados, combinada com ferramentas como PySpark e Looker Studio, pode transformar dados brutos em decisões concretas e insights acionáveis para e-commerce.