**🔍 Página 4: Análise Exploratória dos Dados (EDA)**

**📊 Visão Geral da EDA**

A Análise Exploratória dos Dados (EDA) é uma etapa fundamental na ciência de dados que envolve a investigação inicial dos dados para identificar padrões, detectar anomalias e preparar os dados para análises mais avançadas.

Neste projeto, utilizamos Python, especialmente as bibliotecas Pandas, Matplotlib e Seaborn, para conduzir nossa análise exploratória.

**🔧 Tratamento Inicial dos Dados**

**1. Importação e Visualização Inicial dos Dados**

Usamos a biblioteca Pandas para carregar e visualizar as primeiras linhas das bases de dados:

import pandas as pd

orders\_df = pd.read\_csv('orders.csv')

drivers\_df = pd.read\_csv('drivers\_data.csv')

products\_df = pd.read\_csv('products\_data.csv')

customers\_df = pd.read\_csv('customers\_data.csv')

missing\_items\_df = pd.read\_csv('missing\_items\_data.csv')

# Visualizando primeiras linhas

display(orders\_df.head())

display(drivers\_df.head())

**2. Verificação e Tratamento de Dados Ausentes**

Analisamos valores ausentes usando Pandas:

# Identificando valores ausentes

missing\_data = orders\_df.isnull().sum()

print(missing\_data)

# Tratamento de valores ausentes (se necessário)

orders\_df.dropna(inplace=True)

**📉 Análise Estatística e Distribuições**

**3. Estatísticas Descritivas Básicas**

Calculamos as estatísticas descritivas básicas para variáveis-chave:

# Estatísticas descritivas

orders\_df.describe()

**4. Distribuição de Valores (Matplotlib e Seaborn)**

Visualizamos distribuições com histogramas e boxplots:

import matplotlib.pyplot as plt

import seaborn as sns

# Histograma dos valores dos pedidos

plt.figure(figsize=(10,6))

sns.histplot(orders\_df['order\_amount'], bins=30, kde=True)

plt.title('Distribuição dos Valores dos Pedidos')

plt.xlabel('Valor do Pedido ($)')

plt.ylabel('Frequência')

plt.show()

**🗺️ Análise Geográfica**

**5. Analisando Pedidos por Região**

Examinamos a frequência dos pedidos nas diferentes regiões:

# Pedidos por região

region\_counts = orders\_df['region'].value\_counts()

plt.figure(figsize=(8,5))

sns.barplot(x=region\_counts.index, y=region\_counts.values)

plt.title('Número de Pedidos por Região')

plt.xlabel('Região')

plt.ylabel('Número de Pedidos')

plt.xticks(rotation=45)

plt.show()

**🕒 Análise Temporal**

**6. Analisando Pedidos por Horário**

Examinamos as tendências horárias de entregas:

# Convertendo para formato datetime

orders\_df['delivery\_hour'] = pd.to\_datetime(orders\_df['delivery\_hour']).dt.hour

# Análise da frequência de entregas por horário

hour\_counts = orders\_df['delivery\_hour'].value\_counts().sort\_index()

plt.figure(figsize=(10,5))

sns.lineplot(x=hour\_counts.index, y=hour\_counts.values, marker='o')

plt.title('Distribuição de Pedidos por Horário')

plt.xlabel('Hora do Dia')

plt.ylabel('Número de Pedidos')

plt.xticks(range(0,24))

plt.grid(True)

plt.show()

**📌 Análise de Fraudes Potenciais**

**7. Frequência de Itens Faltantes**

Identificamos padrões em itens faltantes:

# Análise da frequência de itens faltantes

missing\_rate = orders\_df['items\_missing'].value\_counts(normalize=True) \* 100

plt.figure(figsize=(6,4))

sns.barplot(x=missing\_rate.index, y=missing\_rate.values)

plt.title('Taxa de Itens Faltantes por Pedido (%)')

plt.xlabel('Número de Itens Faltantes')

plt.ylabel('Taxa (%)')

plt.show()

**8. Análise de Motoristas Suspeitos**

Identificamos motoristas com altas taxas de itens faltantes:

# Calculando taxas de itens faltantes por motorista

fraud\_driver = orders\_df.groupby('driver\_id')['items\_missing'].mean().sort\_values(ascending=False).head(10)

plt.figure(figsize=(10,6))

sns.barplot(x=fraud\_driver.values, y=fraud\_driver.index)

plt.title('Motoristas com Maiores Taxas de Itens Faltantes')

plt.xlabel('Média de Itens Faltantes por Pedido')

plt.ylabel('ID do Motorista')

plt.show()

**🎯 Principais Insights da EDA**

* Identificação de regiões e horários com maior frequência de problemas.
* Motoristas e clientes suspeitos com taxas elevadas de reclamações.
* Variações significativas nos valores dos pedidos e frequência dos itens faltantes.