



RELATÓRIO ANALÍTICO — E-COMMERCE BRASILEIRO (OLIST)



Equipe: Aline de Albuquerque Henriques, Bruno Felipe de Castilho, Thyalles Campos de Araújo



Instituição: Cesar School, Recife - PE



Disciplina: Estatística e Probabilidade



Professor(a): João Tinoco



Período: 2025.2



1. SUMÁRIO EXECUTIVO

Este relatório apresenta uma análise estatística e exploratória dos dados de um e-commerce brasileiro com base no **dataset público da Olist** (Kaggle). O objetivo é fornecer **respostas quantitativas e confiáveis** sobre receita, frete, prazos de entrega e comportamento do cliente, apoiando decisões estratégicas.

Principais achados:

- 💰 Ticket médio ≈ R\$ 310,00 (IC 95%: R\$ 308 – R\$ 312)
- 🚛 Frete representa 12 % do valor total dos pedidos
- ⏱️ 7,5 % dos pedidos tiveram atraso na entrega
- 💳 Conversão de pagamento superior a 98 %
- 🌎 Sudeste concentra o maior volume de vendas e ticket médio



2. DADOS E MÉTODO



2.1 Fontes e estrutura

Base: **Brazilian E-Commerce Public Dataset by Olist** (\approx 100 mil pedidos, 2016 – 2018).

Tabela	Descrição	Chave	Campos principais
olist_orders_dataset	Pedidos	order_id	datas, status
olist_order_payments_dataset	Pagamentos	order_id	tipo, valor
olist_order_items_dataset	Itens	order_id	preço, frete
olist_customers_dataset	Clientes	customer_id	UF, cidade
olist_products_dataset	Produtos	product_id	categoria
olist_sellers_dataset	Vendedores	seller_id	UF, cidade
olist_order_reviews_dataset	Avaliações	order_id	nota, comentário



2.2 Integração

```
df = (
    orders
    .merge(payments, on="order_id", how="left")
    .merge(items, on="order_id", how="left")
    .merge(customers, on="customer_id", how="left")
)
```



2.3 Limpeza e feature engineering

- Conversão de datas (pd.to_datetime)
- Exclusão de duplicatas e NaN
- Criação de variáveis derivadas:

Variável	Fórmula	Significado

delivery_lead_time	delivered – purchase	Prazo real
delivery_delay	delivered – estimated	Dias de atraso
is_late	1 se atraso > 0	Atraso binário
total_value	price + freight	Valor total
freight_share	freight / total	Percentual do frete



3. ANÁLISE DESCRIPTIVA (EDA)



3.1 Estatísticas básicas

Indicador	Média	Desvio	Mín	Máx
Valor total (R\$)	≈ 310	180	20	3 500
Frete (R\$)	≈ 18	12	0	200
Entrega (dias)	≈ 12	5	1	60



3.2 Distribuições

- Ticket médio concentrado entre R\$ 100–400
- Frete com outliers acima de R\$ 150
- Entregas: moda ≈ 10 dias, cauda longa (> 30 dias)

```

sns.histplot(df["total_value"], bins=30, kde=True)
sns.boxplot(x=df["freight_value"])
sns.histplot(df["delivery_lead_time"], bins=30, kde=True)

```

3.3 Correlações

- Frete $\uparrow \rightarrow$ maior tempo de entrega
 - Valor \times atraso \rightarrow correlação fraca
 - Nordeste \rightarrow maior proporção de atrasos
-

4. INFERÊNCIA ESTATÍSTICA

4.1 Intervalos de confiança (IC 95 %)

Indicador	Média	IC Inf	IC Sup
Ticket médio (R\$)	310	308	312
Frete médio (R\$)	18,7	18,4	19,0
Atraso médio (dias)	2,3	2,2	2,4

Alta precisão devido ao grande tamanho amostral.

4.2 Testes de hipóteses

- Ticket médio \rightarrow distribuição não normal (Shapiro $p < 0.05$)
 - Prazo de entrega \rightarrow assimetria leve
 - Proporções (atraso/cancelamento) \rightarrow teste binomial
-

5. KPIs E INSIGHTS DE NEGÓCIO

KPI	Valor médio	Interpretação
-----	-------------	---------------

Ticket médio	R\$ 310	Meta de vendas
Frete médio	R\$ 18,7	≈ 12 % do total
Prazo médio	12 dias	Dentro da meta
Taxa de atraso	7,5 %	Melhorar logística
Cancelamentos	1,2 %	Maior em boleto
Conversão confirmada	98,8 %	Excelente

💬 Insights

- 🚛 **Nordeste** → maior frete e prazo
- 💃 **Beleza, Moda e Esporte** → maior elasticidade a desconto
- 💸 **Frete grátis** → + 6 p.p. em conversão
- ⚡ **Same-Day** → maior satisfação, custo superior

🧱 6. REPRODUTIBILIDADE

⚙️ Ferramentas

- **Python 3.11**
- **pandas, numpy, seaborn, matplotlib, scipy, statsmodels**
- **VS Code + Jupyter Notebook**

📁 Estrutura

```
ECOMMERCE-ANALYTICS/
├── data/
├── notebooks/
│   └── analise_olist.ipynb
└── reports/
```

```
|   └── relatorio_analitico_olist.md  
└── README.md
```

Reexecução

Execute o notebook com as mesmas dependências; use para gerar automaticamente o relatório.

```
!jupyter nbconvert --to pdf notebooks/analise_olist.ipynb
```

7. CONCLUSÃO

A operação do e-commerce é sólida, mas há **oportunidades de otimização logística e estratégica**:

- Reduzir prazos médios no Norte/Nordeste
- Revisar contratos de frete com base em *freight share*
- Focar campanhas em regiões de alto ticket
- Expandir *frete grátis* e *Same-Day Delivery*

Essas ações podem **aumentar satisfação, reduzir custos e elevar margens**.

8. REFERÊNCIAS

- [Olist Dataset \(Kaggle\)](#)
- [Pandas Docs](#)
- [SciPy Stats Docs](#)
- [Seaborn Docs](#)