

Projeto Final: Análise exploratória de dados

- **Objetivo**
 - Aplicar os conhecimentos aprendidos ao longo do curso para resolver um problema real de negócios, utilizando análise exploratória de dados (EDA), estatística descritiva e visualizações.
- **Escolha do Dataset**
 - Aqui estão algumas sugestões:
 - Sales Data – comportamento de vendas em e-commerce: <https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store>
 - Airbnb Listings – análise de preço, localização e disponibilidade: <https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data>
 - Bank Marketing Dataset – segmentação de clientes, estatísticas descritivas e visualizações de resposta por perfil. Também permite testes de hipótese para comparar grupos: <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset>
 - USA Bank Financial Data – análises temporais, estatísticas descritivas de indicadores e comparação entre instituições: <https://www.kaggle.com/datasets/vishalsinghsangral/usa-bank-financial-data>
- **Roteiro do Projeto**
 - **Introdução**
 - Apresente o dataset e justifique sua escolha.
 - Defina o problema de negócio a ser investigado (Ex: “Qual a categoria de produto mais lucrativa?”, “Como a pandemia afetou as vendas?”, “Qual perfil de cliente aluga mais no Airbnb?”).
 - **Pré-processamento e Limpeza**
 - Verifique e trate:
 - Valores faltantes
 - Dados duplicados
 - Colunas irrelevantes ou com muitos valores constantes
 - Aplique `.drop()`, `.fillna()`, `.apply()` quando necessário.
 - **Exploração Inicial com Estatísticas Descritivas**
 - Use `.describe()` e `.value_counts()` para entender os dados.
 - Calcule médias, desvios, máximos, mínimos, etc.
 - Interprete os resultados: O que os dados dizem?
 - **Tratamento de Outliers**
 - Use IQR e desvio padrão para identificar e tratar outliers.
 - Compare gráficos antes/depois do tratamento.
 - **Visualizações com Seaborn**
 - Gráficos obrigatórios:
 - histplot ou boxplot para análise univariada.
 - scatterplot ou pairplot para relações entre variáveis.
 - lineplot ou barplot para evolução no tempo ou comparação de categorias.

- **Group By e Agregações**
 - Aplique groupby para responder perguntas específicas, como:
 - Qual cidade vende mais?
 - Quais categorias de produto geram maior receita?
 - Qual dia da semana tem mais movimento?
- **Correlação e Teste de Hipóteses (Avançado)**
 - Mostre o heatmap de correlações.
 - Opcional: realize um teste de hipótese simples (Ex: comparar médias antes e depois de uma campanha).
- **Conclusão e Recomendações**
 - Apresente os principais achados.
 - Dê **duas a três recomendações práticas** para o negócio com base na análise feita.
- **Entregáveis**
 - **Notebook Jupyter (ou Google Colab) com:**
 - Título e subtítulos claros
 - Código comentado
 - Gráficos bem apresentados
 - Conclusões em texto
 - **Apresentação (pdf):**
 - Slide com problema de negócio
 - Slide com visualizações principais
 - Slide com insights e recomendações
- **Prazo**
 - **Entrega do notebook e Apresentação:** até início da aula 9 (última aula)
- **Dicas**
 - Escreva sempre **insights** junto com seus gráficos.
 - Mantenha o foco: não analise tudo, escolha **uma ou duas perguntas de negócio bem definidas**.
 - Capriche no storytelling visual: seus gráficos são o seu argumento.