



FACULDADE SENAI DE TECNOLOGIA GASPAR RICARDO JÚNIOR

GRADUAÇÃO EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

**Giovani Miamoto
Luciano Júnior
Victor Geroto
Rafael Pecorari**

Trabalho Ciência de Dados

**SOROCABA - SP
2025**

**Giovani Miamoto
Luciano Júnior
Victor Geroto
Rafael Pecorari**

Trabalho Ciência de Dados

Documentação apresentanda à Faculdade SENAI de Tecnologia Gaspar Ricardo Júnior, para a entrega no dia 20/05, como requisito da matéria de ciência de dados na graduação em Análise e Desenvolvimento de Sistemas sob a orientação do professor:

Prof. André Cassulino Araujo Souza

**SOROCABA - SP
2025**

FACULDADE SENAI DE TECNOLOGIA GASPAR RICARDO JÚNIOR

Trabalho Ciência de Dados

SUMÁRIO

- 1. Introdução**
- 2. Referencial Teórico**
- 3. Metodologia**
- 4. Conclusão**
- 5. Referências**
- 6. Apêndices**

INTRODUÇÃO

Este projeto de Ciência de Dados tem como foco a análise estatística de um conjunto de dados simulados do mercado de aluguel de imóveis no Rio de Janeiro. O objetivo principal é aplicar técnicas de estatística descritiva, visualização de dados e modelagem preditiva para entender o comportamento dos preços de aluguel em diferentes bairros.

A escolha desse tema se justifica pela sua relevância social e econômica, além da oportunidade de aplicar conhecimentos práticos de Python e suas bibliotecas.

Referencial Teórico

Foram aplicados os seguintes conceitos estatísticos:

- **Medidas de Tendência Central:** média, mediana e moda.
- **Medidas de Dispersão:** amplitude, variância e desvio-padrão.
- **Visualizações Estatísticas:** histogramas, boxplots, gráficos de barras e dispersão.
- **Análise de Dados Categóricos:** frequências e proporções por categorias.
- **Correlação:** análise da relação entre múltiplas variáveis quantitativas.
- **Regressão Linear Simples:** análise preditiva entre número de quartos e o preço de aluguel.

Bibliotecas Python: **pandas**, **numpy**, **matplotlib**, **seaborn**, **scipy**, **scikit-learn**.

Metodologia

Base de Dados

Utilizamos uma base de dados simulada contendo **500 registros de imóveis**. As principais variáveis foram: bairro, tipo de imóvel, número de quartos, número de banheiros, número de avaliações, dias disponíveis e preço

Tratamento de Dados

Foram realizadas as seguintes etapas:

- Criação e limpeza de dados simulados
- Tratamento de outliers
- Cálculo de variáveis derivadas
- Normalização de nomes de bairros e tipos de imóvel

Ferramentas utilizadas:

Google Colab, Python, pandas, numpy, seaborn, matplotlib, scipy, scikit-learn.

Conclusão

Este trabalho permitiu identificar padrões relevantes no mercado de aluguel do Rio de Janeiro, mostrando que o número de quartos, banheiros e a localização são fatores que impactam significativamente os preços. O uso de técnicas estatísticas como medidas de tendência central, dispersão, análise categórica, correlação e regressão linear simples foi essencial para obter essas conclusões.

Entre as limitações encontradas, destaca-se o uso de dados simulados, que, apesar de bem construídos, não refletem todas as variáveis presentes no mercado real. Como sugestão para futuras análises, recomenda-se trabalhar com bases de dados reais, incluir variáveis adicionais (como idade do imóvel).

Referências

McKinney, W. (2018). Python for Data Analysis. O'Reilly Media.
Seaborn Documentation: <https://seaborn.pydata.org/>
Scikit-learn User Guide: <https://scikit-learn.org/>
Google Colab: <https://colab.research.google.com/>