

# Análise e Predição do Valor de Aluguéis de Imóveis no Brasil

---

INTELIGÊNCIA ARTIFICIAL

**Discentes:**

Diego Metring Teixeira

Murilo Ribeiro da Costa



**CENTRO  
UNIVERSITÁRIO**

# PROBLEMA, DOMÍNIO E BASE DE DADOS

---

35.7K

2.4K

## PROBLEMA

Entender e prever o valor de aluguéis em um mercado complexo.

## DOMÍNIO

Mercado Imobiliário de Aluguéis no Brasil

## BASE DE DADOS

**Brazilian Real Estate to Rent (Kaggle)**

**Foco da análise:** Imóveis na faixa de R\$300 a R\$1300, visando uma média estimada do custo dos imóveis na época da pesquisa

**Volume inicial:** 35.773 entradas, 78 colunas

**Volume final:** 2399 entradas, 13 colunas

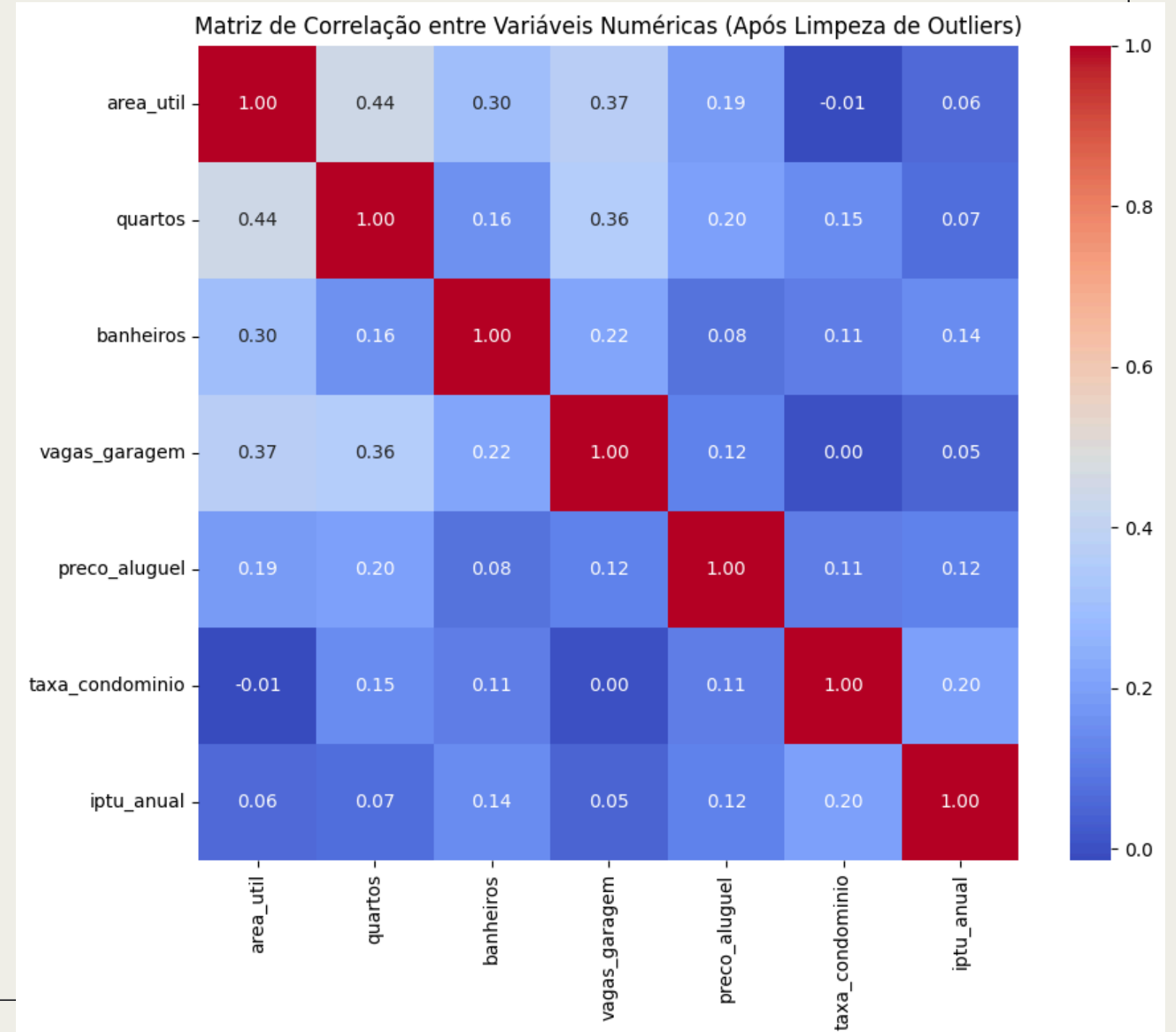
# PRÉ-PROCESSAMENTO E LIMPEZA DE DADOS

---

<p>●</p> <p><i>Seleção de Colunas Relevantes</i></p> <p>Como principais atributos do <i>dataset</i>, foram selecionadas as colunas cidade, estado, área útil, quartos, banheiros, suítes, vagas na garagem, se é mobiliado, se tem piscina, se tem academia, valor do aluguel, taxa de condomínio e IPTU.</p>	<p>●</p> <p><i>Renomeação e Conversão de Tipos</i></p> <p>Colunas renomeadas para facilitar o uso, conversão de dados textuais para numéricos onde aplicável, valores ausentes preenchidos.</p>	<p>●</p> <p><i>Preenchimento de Valores Ausentes</i></p> <p>Linhas com valores zero ou nulos nas colunas <b>preco_aluguel</b> e <b>area_util</b> removidas.</p>	<p>●</p> <p><i>Remoção de Outliers e Filtragem</i></p> <p><i>Outliers</i> identificados e removidos com base na coluna <b>area_util</b> utilizando o percentil 99. Foram incluídos somente os imóveis com faixa de preço de aluguel entre R\$300 e R\$1300.</p>	<p>●</p> <p><i>AED e Análise de Correlação</i></p> <p>Foi gerado um resumo estatístico das colunas numéricas após a última etapa para entender a distribuição e as principais medidas de tendência central e dispersão. As relações lineares entre os atributos foram identificadas por uma matriz de correlação.</p>
---	---	---	---	---

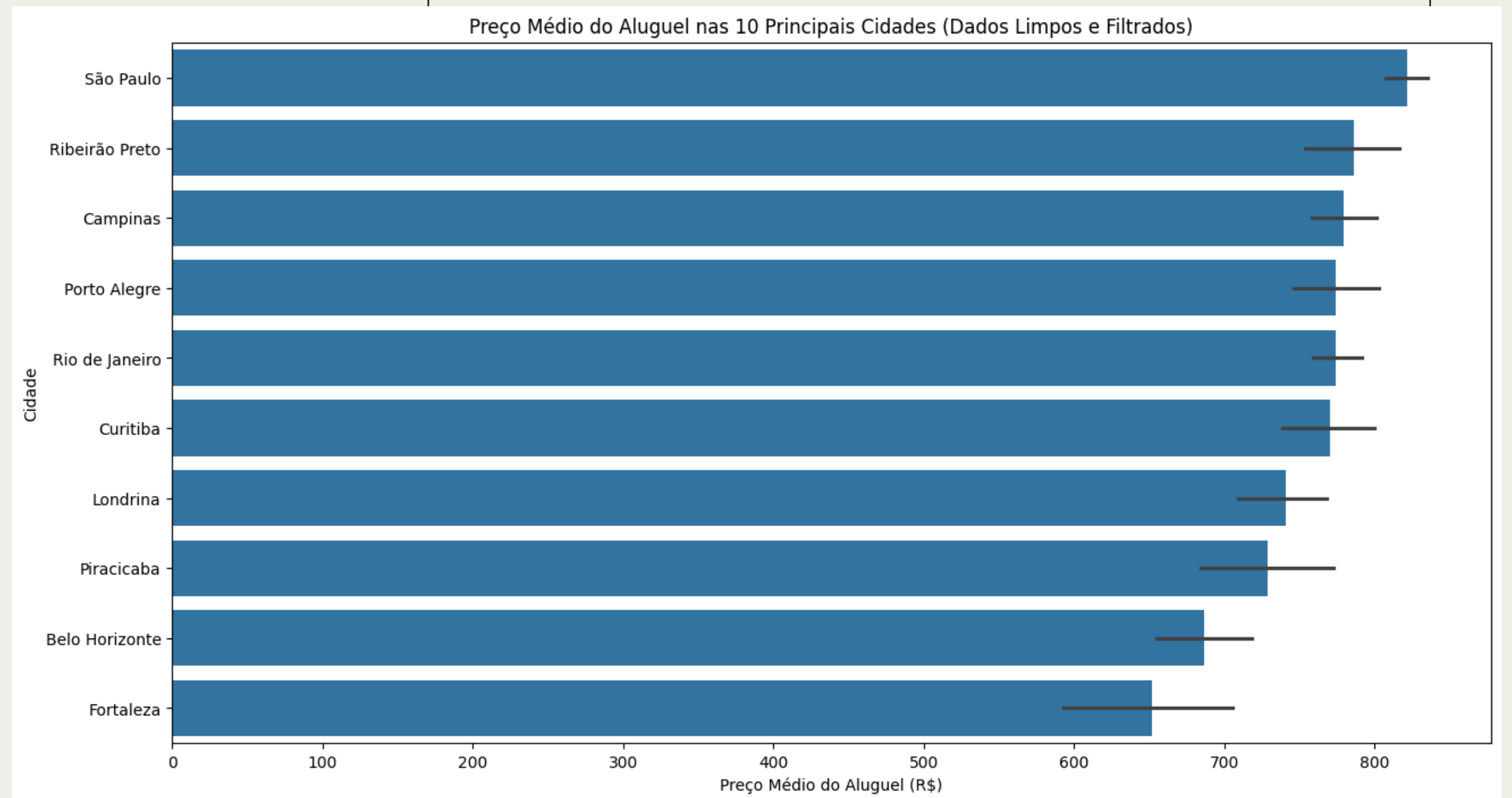
# RESULTADOS DA AED

- O preço do aluguel apresenta maior correlação positiva com os atributos:
  - Número de quartos (0.20)
  - Área útil (0.19)
  - Vagas na garagem e IPTU Anual (0.12)



# RESULTADOS DA AED

- A cidade também pode ser classificada como um dos principais fatores influentes no valor do aluguel, como pode ser visto na imagem



# PREPARAÇÃO PARA MODELAGEM

---

- **Variável Alvo (y):** preco\_aluguel.
- **Features (X):** Demais colunas após limpeza.
- **Engenharia de Atributos:**
  - One-Hot Encoding para variáveis categóricas (cidade, estado, mobiliado, piscina, academia).
  - Preenchimento de NaNs restantes em X com 0.
- **Divisão Treino/Teste:** 80% Treino, 20% Teste (random\_state=42)

# MACHINE LEARNING (ALGORITMO 1)

---

- **Algoritmo:** Random Forest Regressor.
- **Funcionamento:** Combina várias árvores de decisão para prever valores contínuos.
- **Treinamento:** Modelo inicial (`model.fit(X_train, y_train)`).
- **Otimização de Hiperparâmetros:** `RandomizedSearchCV` para buscar melhores `n_estimators`, `max_depth`, etc.
- **Métricas de Avaliação (Teste):**
  - **MAE (Erro Médio Absoluto):** R\$80.05 (inicial), R\$79.85 (otimizado).
  - **R<sup>2</sup> (Explicabilidade):** 0.42 (inicial), 0.43 (otimizado).

# MACHINE LEARNING (ALGORITMO 2)

---

- **Algoritmo:** Gradient Boosting Regressor.
- **Funcionamento:** Constrói modelos sequencialmente, corrigindo erros dos modelos anteriores.
- **Treinamento:** Modelo `gbr_model.fit(X_train, y_train)`.
- **Métricas de Avaliação (Teste):**
  - **MAE:** R\$ 107.11
  - **R<sup>2</sup>:** 0.36

# COMPARATIVO DE PERFORMANCE

---

<i>Random Forest Inicial</i>	<i>Random Forest Otimizado</i>	<i>Gradient Boosting Regressor</i>
<ul style="list-style-type: none"><li>• <math>R^2 = 0.42</math></li><li>• MAE = R\$80,05</li><li>• MSE = 13411.78</li></ul>	<ul style="list-style-type: none"><li>• <math>R^2 = 0.43</math></li><li>• MAE = R\$79.85</li><li>• MSE = 13227.05</li></ul>	<ul style="list-style-type: none"><li>• <math>R^2 = 0.36</math></li><li>• MAE = R\$107,11</li><li>• MSE = 17471.63</li></ul>

## INSIGHTS

---

<i>Modelo Vencedor</i>	<i>Fatores Não Capturados</i>	<i>Limitação</i>
Random Forest (Otimizado) com MAE de ~R\$80 e $R^2$ de 0.43.	Estado de conservação, vista, andar, barulho, proximidade a serviços, etc.	O modelo explica 43% da variação de preço; 57% ainda não é capturada

# REFERÊNCIAS

---

maverickjpa. (Ano do Upload ou Acesso, se disponível, caso contrário, omitir). Brazilian Real Estate to Rent. Kaggle. Disponível em: <https://www.kaggle.com/datasets/maverickjpa/brazilian-real-estate-to-rent>. Acesso em: 25 de novembro de 2025.

Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. Documentação oficial de *RandomForestRegressor* e *GradientBoostingRegressor*. Disponível em: <https://scikit-learn.org/stable/>. Acesso em: 25 de novembro de 2025.

# Obrigado!

---