

Estudo de Caso 3 – Parte 1:

Quanto vale alugar um apê?

A startup AlugaFácil deseja automatizar suas recomendações de preços de aluguel com base em características dos imóveis. Eles também querem prever se um imóvel é caro ou barato com base na média da região.

Você e sua equipe foram contratados como cientistas de dados para criar e avaliar esse modelo.

Dataset fornecido:

O arquivo dados_apartamentos.csv contém aproximadamente 1000 registros com as seguintes colunas:

- area_m2
- quartos
- banheiros
- garagem
- distancia_centro_km
- andar
- preco_aluguel

Objetivos:

1. Criar um modelo de regressão linear para prever o valor do aluguel.
2. Classificar os imóveis em "caro" ou "barato" com base na média geral.
3. Avaliar a performance do modelo de classificação.
4. Propor ajustes para melhorar a performance.

Tarefas por grupo:

Parte 1: Análise exploratória

- Carregue os dados com pandas
- Verifique valores ausentes, tipos e estatísticas básicas
- Faça gráficos de dispersão entre as variáveis independentes e o preço
- Há alguma correlação aparente?

Parte 2: Regressão Linear

- Crie um modelo de regressão linear com scikit-learn
- Use as variáveis: area_m2, quartos, banheiros, garagem e distancia_centro_km
- Separe os dados em treino e teste (80/20)
- Faça previsões no conjunto de teste

- Meça a performance com `mean_absolute_error` e `r2_score`
- Interprete os coeficientes do modelo

Parte 3: Classificação

- Crie uma nova coluna "caro":
`df["caro"] = df["preco_aluguel"] > df["preco_aluguel"].mean()`
- Treine um modelo de regressão logística
- Meça a performance com:
 - Acurácia
 - Precisão
 - Recall
 - Matriz de confusão

Parte 4: Discussão em grupo

- O modelo previu bem o valor do aluguel?
- O modelo classificou bem os imóveis?
- Que ajustes poderiam melhorar o desempenho?

Parte 5: Entrega

- Notebook do Google Colab com:
 - Código executado
 - Gráficos
 - Respostas escritas em formato de relatório

Dataset: dados_apartamentos.csv