

Exercícios de Manipulação de Dados com Pandas

Exercício 1: Feature Engineering

Crie uma nova coluna chamada **price_category** que categoriza as propriedades em "High", "Medium" e "Low" com base nos quartis do preço total.

Exercício 2: Exploração de Propriedades

Calcule a proporção de propriedades mobiliadas (*furnished*) em comparação com as não mobiliadas (*unfurnished*) ou não especificadas (*not specified*).

Exercício 3: Análise por Tipo de Transação

Calcule a média do preço por metro quadrado para cada tipo de transação (*resale*, *new booking*) e determine qual tipo de transação tem o maior preço médio por metro quadrado.

Exercício 4: Estatísticas por Status da Propriedade

Analise as estatísticas (média, mediana, desvio padrão) do preço total e da área útil (*super area*) agrupadas por status da propriedade (*Ready to Move*, *Under Construction*).

Exercício 5: Tratamento de Valores Nulos

Preencha os valores nulos na coluna "*balcony_count*" com a mediana dos valores não nulos. Em seguida, verifique novamente se ainda existem valores nulos no conjunto de dados.

Exercício 6: Análise Temporal

Divida o dataset em anos com base na coluna de "*data de transação*" e crie um gráfico de linha mostrando a tendência anual do preço médio por metro quadrado ao longo do tempo.

Exercício 7: Agregação Multi-Característica

Calcule a média, mediana e desvio padrão da área útil (*super area*) para cada combinação única de tipo de propriedade (*property type*) e número de quartos (*bedroom count*).

Exercício 8: Tratamento de Categorias Raras

Identifique as categorias na coluna "*society name*" que aparecem menos de 10 vezes no conjunto de dados e substitua-as por "*Other Society*".

Exercício 9: Correlação e Visualização

Calcule a matriz de correlação entre as colunas numéricas e crie um mapa de calor (*heatmap*) para visualizar as correlações. Destaque as correlações fortes com cores diferentes.

Exercício 10: Transformação de Dados

Crie uma nova coluna chamada "*log_price*" que contenha o logaritmo natural (*ln*) do preço total. Em seguida, normalize a coluna "*super area*" utilizando a técnica Min-Max, ou seja, ajuste os valores para um intervalo entre 0 e 1.