

Análise Exploratória dos Dados (EDA)

Iniciamos o processo tratando os valores faltantes ou zeros nas variáveis numéricas. Removemos também as colunas desnecessárias para nosso modelo como 'id', 'host_name', 'ultima_review' e 'host_id'

Também foi feita uma limpeza dos outliers na coluna de 'price' e 'bairro' para visualização de valores mais consistentes. Com isso, realizamos diversas análises com as colunas 'bairro_group', 'bairro' e 'room_type'



Figura 1 - Comparação entre as variáveis com a coluna bairro_group

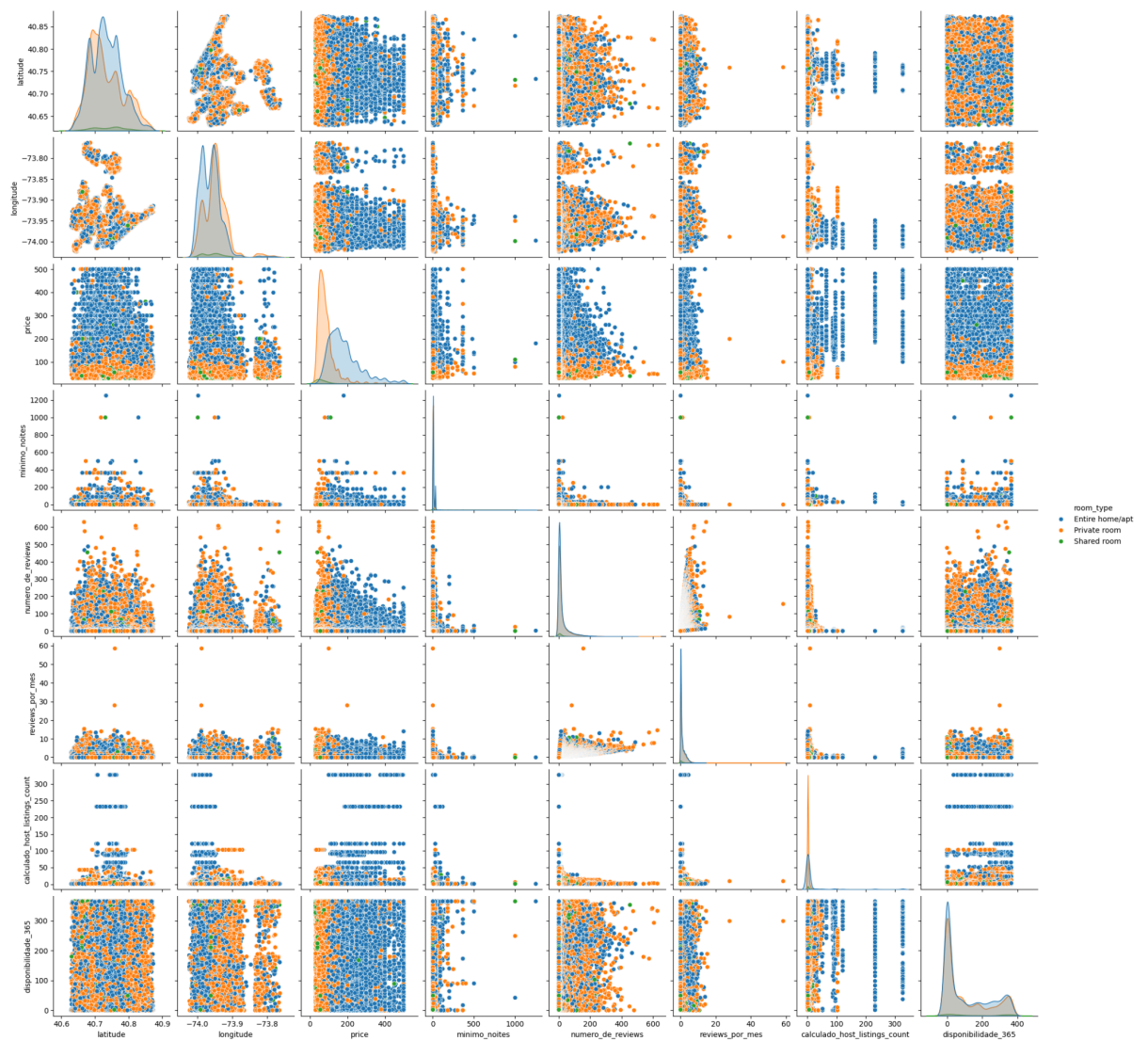


Figura 2 - Comparação entre as variáveis com a coluna room_type

Conforme mostrado na Figura 3, analisando as variáveis '**bairro_group**', e '**room_type**', observamos que elas influenciam no preço. Após o devido tratamento, utilizaremos essas variáveis no modelo de treinamento.

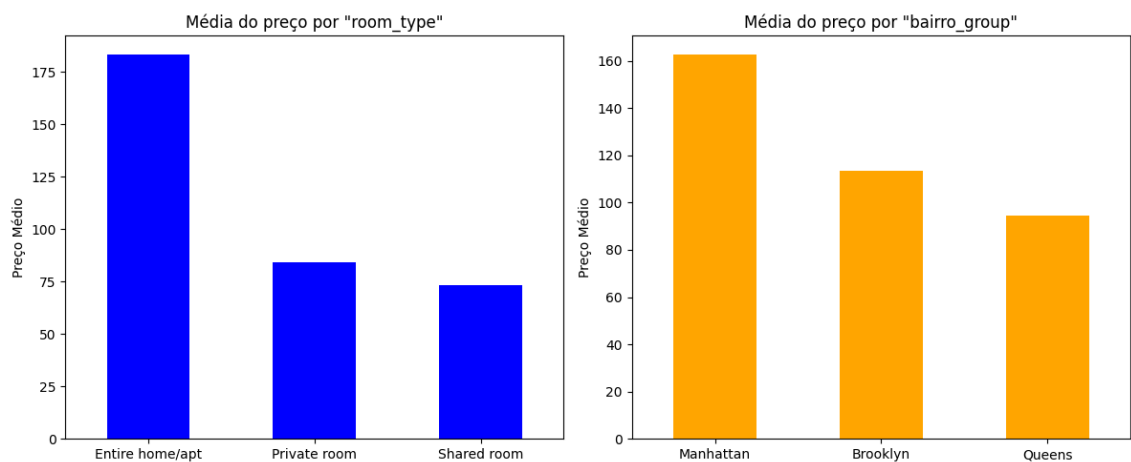


Figura 1 - Média do preço por "room_type" e média do preço por "bairro_group"

Podemos concluir que o bairro Manhattan e o tipo de quarto “Entire home” apresentam o maior preço médio de imóveis. Isso sugere que a demanda nessa localidade é alta.

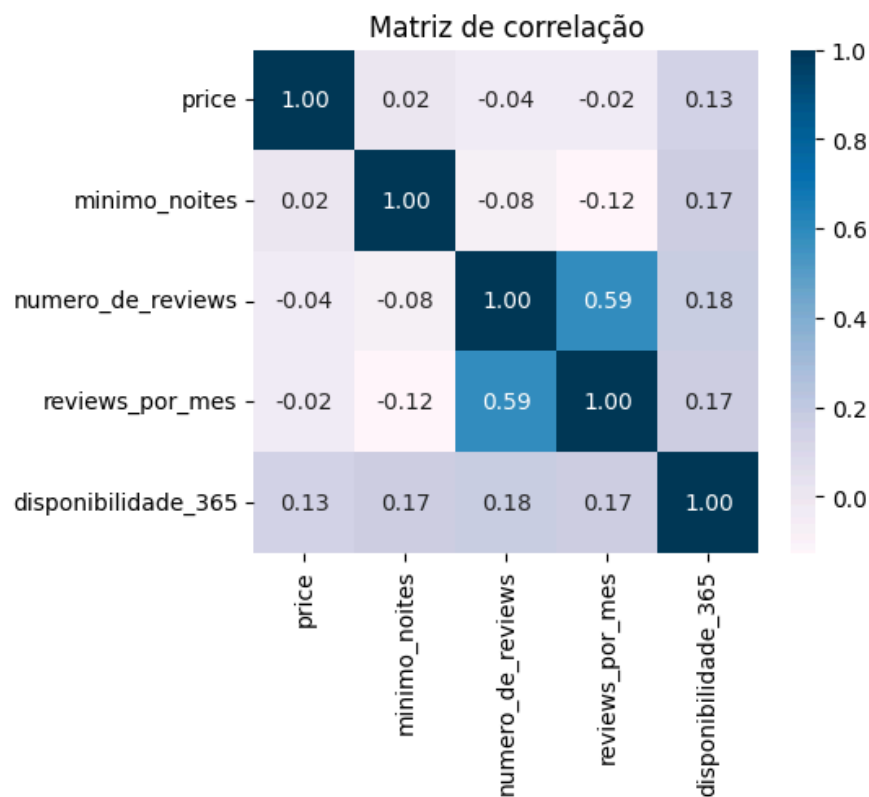


Figura 4 - Matriz de Correlação

Conforme a figura acima não há nenhuma correlação linear significativa entre o número mínimo de noites e a disponibilidade com o preço. Entretanto, é crucial ressaltar que a falta de correlação linear não elimina a possibilidade de haver alguma relação não linear entre as variáveis.

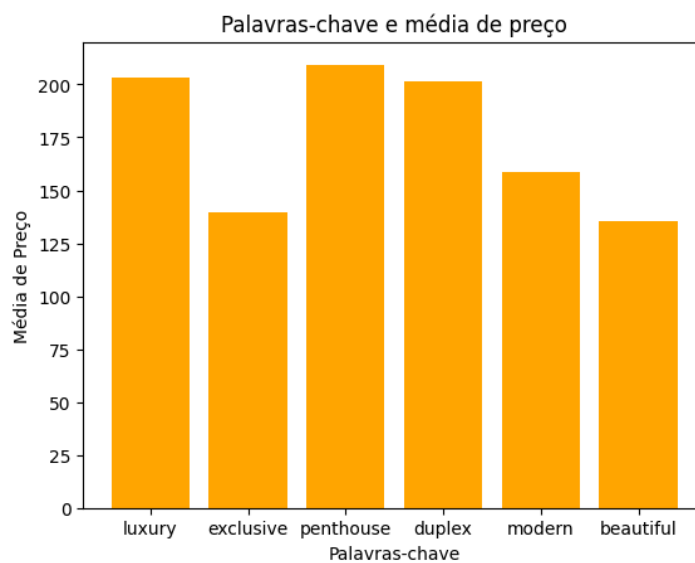


Figura 5 - Correlação entre palavra-chave e média de preço

Perguntas do Desafio

1. **Supondo que uma pessoa esteja pensando em investir em um apartamento para alugar na plataforma, onde seria mais indicada a compra?** Com base na análise conduzida, parece ser uma estratégia promissora concentrar os investimentos em apartamentos para aluguel na área de Manhattan.
2. **O número mínimo de noites e a disponibilidade ao longo do ano interferem no preço?** Conforme a figura 4, não foi identificada uma correlação clara entre o número mínimo de noites e a disponibilidade ao longo do ano com o preço dos imóveis.
3. **Existe algum padrão no texto do nome do local para lugares de mais alto valor?** Quando analisamos determinadas palavras-chave em conjunto com os preços, observamos que termos como "luxo", "cobertura" ou "duplex" geralmente estão associados a locais de maior valor. No entanto, é importante notar que essas palavras foram selecionadas previamente e não podemos garantir com total certeza que elas influenciam os preços da mesma forma que outras variáveis.