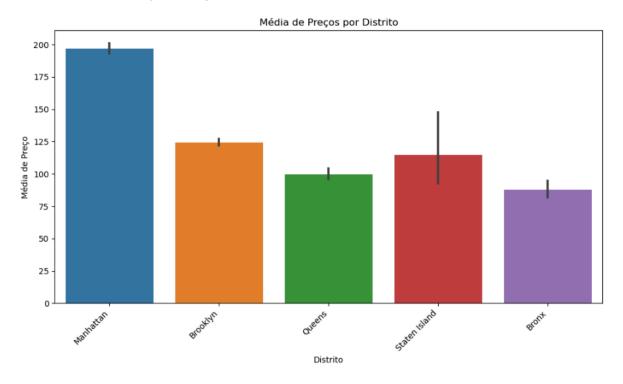
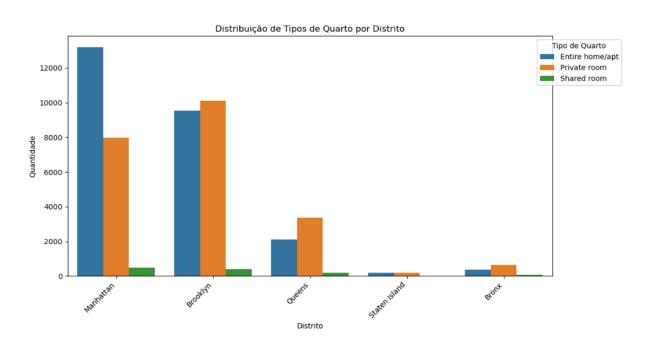
### Respondendo as perguntas do Desafio

# a. Supondo que uma pessoa esteja pensando em investir em um apartamento para alugar na plataforma, onde seria mais indicada a compra?

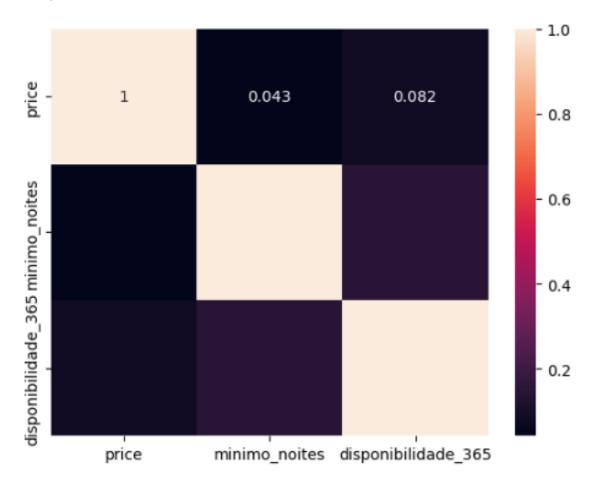
Apartamento/Casa e quartos privativos têm maior demanda, a média de preços é maior e quase não tem disponibilidade pois tendem a estar sempre alugados. Manhattan tem maior média de preços e maior quantidade de imóveis para alugar. Manhattan e Brooklyn são bairros com melhor reputação, localização e proximidade. Manhattan é ideal para negócios e turismo.





# b. O número mínimo de noites e a disponibilidade ao longo do ano interferem no preço?

Sim, embora a correlação entre esses fatores seja moderada, a quantidade mínima de noites e a disponibilidade podem impactar diretamente nos preços. Além disso, observamos que os preços geralmente aumentam durante períodos de alta temporada, como férias, feriados e eventos. Outros elementos, como a localização em áreas populares ou turísticas, também exercem uma influência significativa nos valores praticados.



## c. Existe algum padrão no texto do nome do local para lugares de mais alto valor?

A análise dos nomes revelou padrões associados a quartos com preço mais alto. Termos frequentes identificados - Apartamento, Manhattan, espaçoso, sugerem características mais procuradas pelos usuários.

#### Nuvem de Palavras para Locais com Preços Mais Altos



2. Explique como você faria a previsão do preço a partir dos dados. Quais variáveis e/ou suas transformações você utilizou e por quê? Qual tipo de problema estamos resolvendo (regressão, classificação)? Qual modelo melhor se aproxima dos dados e quais seus prós e contras? Qual medida de performance do modelo foi escolhida e por quê?

Estamos resolvendo um problema de regressão, já que a variável alvo a ser prevista (preço) é numérica e contínua. Algoritmos de regressão são apropriados para prever variáveis contínuas. Utilizamos transformações, como o Label Encoder, para converter variáveis categóricas (strings) em numéricas, pois alguns algoritmos de machine learning exigem representações numéricas. Dividimos os previsores e a classe em X e y, respectivamente, removendo o preço e a última\_revisão de X.

Padronizamos todas as variáveis usando o Standard Scaler para colocar todos os valores na mesma escala, evitando que o algoritmo atribua uma importância desproporcional a algumas variáveis. Em seguida, realizamos a divisão entre conjuntos de treino e teste.

Testamos os algoritmos de regressão linear, SVR, Random Forest e XGBoost. O escolhido para prever os preços em geral foi o Random Forest, pois apresentou o menor Mean Squared Error. As vantagens do uso do Random Forest incluem sua versatilidade, alta precisão e robustez em relação ao overfitting. No entanto, é importante destacar que requer ajuste de hiperparâmetros.

#### 3. Supondo um apartamento com as seguintes características:

```
{'id': 2595,
 'nome': 'Skylit Midtown Castle',
'host_id': 2845,
'host_name': 'Jennifer',
'bairro_group': 'Manhattan',
 'bairro': 'Midtown',
 'latitude': 40.75362,
 'longitude': -73.98377,
 'room type': 'Entire home/apt',
 'price': 225,
 'minimo noites': 1,
 'numero de reviews': 45,
 'ultima review': '2019-05-21',
'reviews_por_mes': 0.38,
 'calculado host listings count': 2,
 'disponibilidade 365': 355}
```

#### Qual seria a sua sugestão de preço?

Com base nos dados sugeridos, a previsão do preço médio de aluguel para a região de Manhattan, especificamente no bairro Midtown, seria de \$268.28.