

2)

a)

Depende. Se a pessoa estiver procurando menores preços, eu recomendaria o Bronx.

Porém, se estiver procurando um maior conforto, com todo o imóvel disponível durante o tempo de estadia, eu recomendaria Manhattan. Embora o preço médio seja mais caro, há maior disponibilidade de imóveis com "Entire home/apt"

b) Há uma pequena interferência, quase nula. Na matriz de correlação, disponibilidade ao longo do ano tem uma relação de 0.082 enquanto número mínimo de noites possui 0.043. Esses valores são muito baixos, indicando uma fraca interferência.

c) Sim. Algumas fileiras que têm na coluna "nome" referências a eventos como SuperBowl, Film, Event, Photo apresentam uma média de preços muito maior comparado a média do conjunto de dados

3)

Primeiro, na análise dos dados, fica claro que as colunas 'id', 'host_id' e 'host_name' não são necessárias para o treinamento. Essas variáveis são para o controle no aplicativo, e não possuem uma relação com o preço.

Ao analisar o dataset, escolhi duas abordagens: Uma previsão utilizando a coluna 'name' e uma sem. Existe sim uma relação no 'nome' com o preço, porém essa variável não é categórica (o usuário pode dar qualquer nome para seu imóvel que estará disponível no aplicativo).

Então, como transformar os dados? Minha escolha foi: Verificando se no nome há a presença de uma das referências citadas acima (item 2c). Além dessa nova variável, também utilizei a proximidade com o Central Park.

Estamos resolvendo um problema de regressão (Tentaremos prever um número entre uma infinidade de números). Dois modelos me vêm em mente: LinearRegression e DecisionTreeRegression. Um modelo de Regressão com Árvore de Decisão normalmente consegue performar mais em relações não lineares, enquanto o de Regressão Linear é o contrário.

A medida de erro utilizada foi o RMSE (Root Mean Squared Error). Basicamente, captura a distância euclidiana entre o valor real e a previsão. Como estamos querendo prever um número (preço), essa métrica me parece a melhor.

4) O apartamento está localizado nas proximidades de Manhattan. Logo, minha sugestão de preço seria a média de preços de aluguel dos imóveis na área (aproximadamente 196.9 dólares por noite)