

Apresentação Airbnb

Pedro Beaumord
Diniz



○ problema





Contexto

O Airbnb possui uma base de dados com inúmeros imóveis, onde temos:

- Informações sobre as características do imóvel, como quantidade de quartos, camas e banheiros;
- Informações sobre a região onde o imóvel está localizado;
- Latitude e Longitude do imóvel, o que permite verificar a proximidade com pontos de interesse;
- Quantidade de locações realizadas no imóvel.

Objetivo

O que queremos resolver com base nos dados fornecidos?

Em posse das informações descritas desejamos melhorar a rentabilidade e a taxa de ocupação dos imóveis, para isso, é necessário aumentar a quantidade de locações desses imóveis.

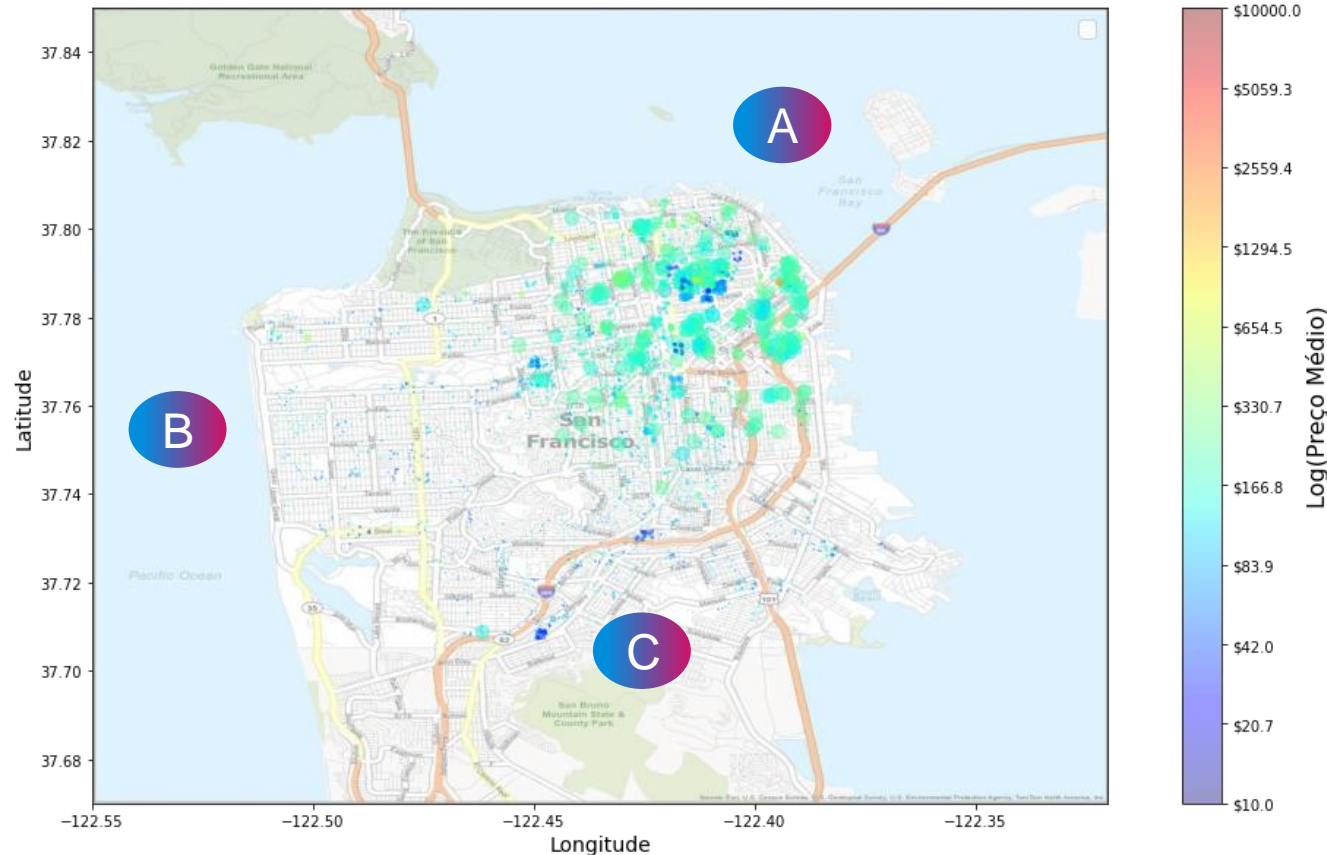


Solução



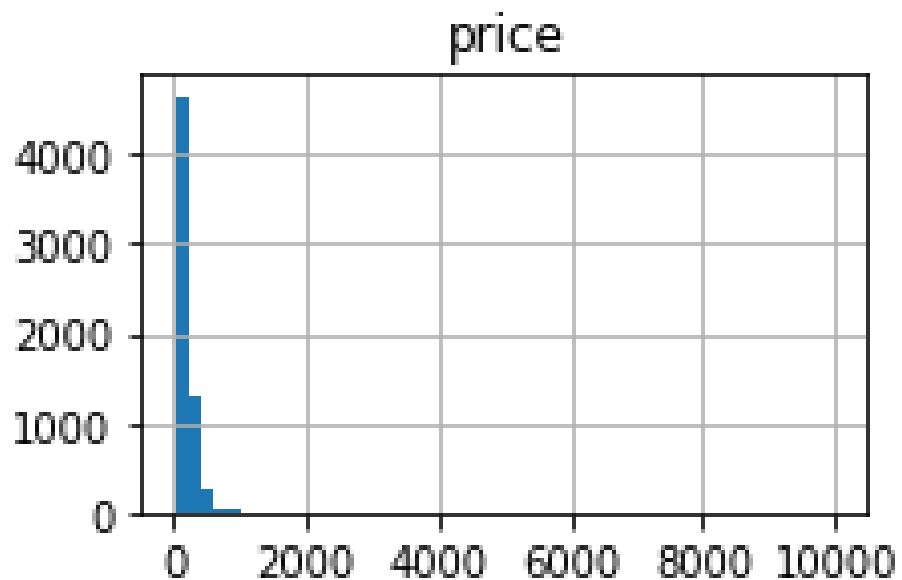
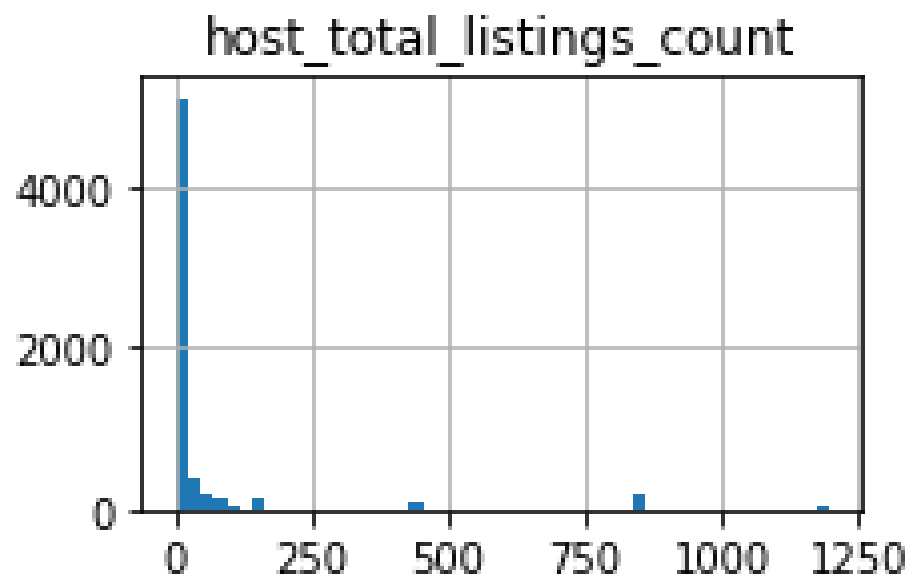
Insights Obtidos

Os gráficos a seguir, foram gerados nas análises exploratórias, aqui serão discutidas as conclusões mais pertinentes tiradas a partir dessas visualizações



Localização dos imóveis

- Os imóveis que apresentaram maior quantidade de locações (representado pelo tamanho da bolha), estão localizados na região da *San Francisco Bay* (ponto A);
- Existe uma procura por imóveis próximos ao Oceano Pacífico (ponto B) e a rodovia 280 (ponto C);
- Os imóveis próximos aos pontos B e C apresentaram valores menores de locação (cor da bolha);
- Próximos ao ponto A, observamos que alguns imóveis são locados mesmo com valor elevado, entretanto na região central vários imóveis apresentam baixa quantidade de locações.



Distribuição dos dados

Comparando as duas variáveis de maior importância, que indicam o valor de locação do imóvel e a quantidade de locações feitas, observamos que:

- A quantidade de locações chega a valores elevados em alguns casos, acima de 1000 locações, mas a grande concentração está em baixas quantidades, inferiores a 20 locações;
- A maior parte dos imóveis apresenta baixos valores para sua locação, entretanto, alguns chegam a 10000 dólares para locação;
- A comparação entre essas duas variáveis levanta a questão se existe uma relação entre os preços das locações e a quantidade que elas são reservadas, nos dando a pergunta norteadora para a modelagem do problema:

Os preços de locação estão adequados para os imóveis?



Modelagem

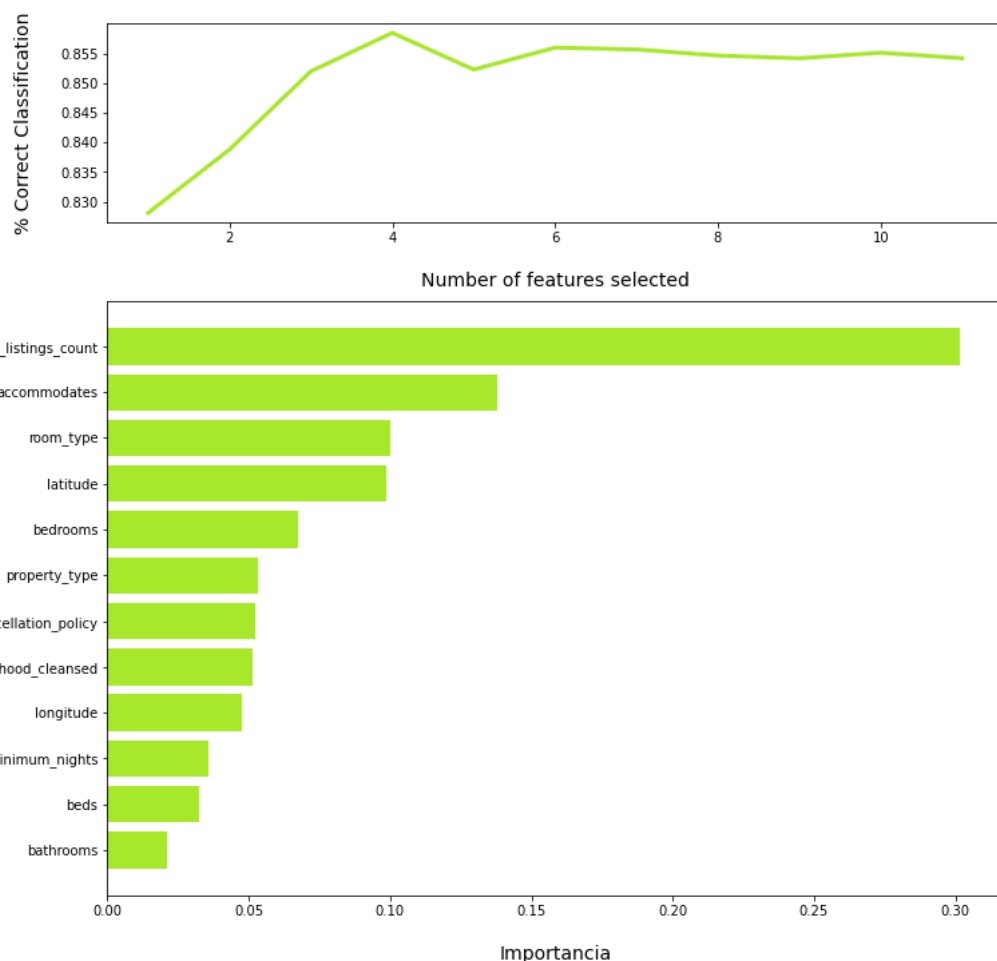
Baseado na pergunta norteadora, iniciamos a etapa de modelagem com as seguintes premissas:

- Imóveis com grande quantidade de locações apresentam preços adequados;
- Os imóveis com baixas locações precisam passar por um processo de precificação em busca do valor ideal de locação;
- Divisão da base em duas:
 - **Base de modelagem:** Imóveis com quantidade elevada de locações e preço adequado
 - **Base de validação:** Imóveis com baixa quantidade de locações, onde será utilizado o modelo criado para encontrar a precificação ideal

Para isso, foi utilizado o modelo de Random Forest, pois este consegue identificar com precisão comportamentos não lineares em busca de relações entre as features.

Recursive Feature Elimination (RFE)

Recursive Feature Elimination with Cross-Validation

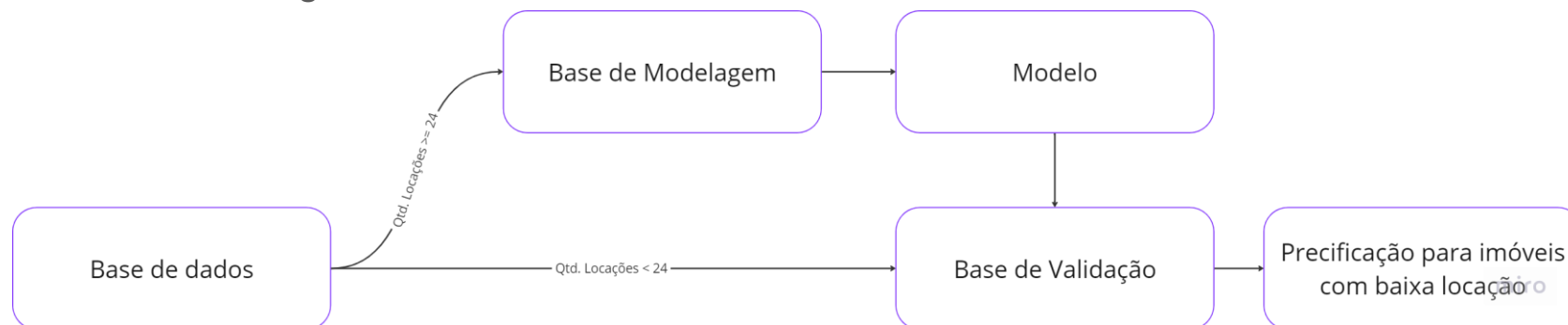


- O RFE é uma técnica utilizada para encontrar a quantidade e quais as features ideais para o modelo, consiste no uso do cross validation para ir testando feature a feature em busca daquelas que apresentam maior importância para o modelo;
- As features não selecionadas pelo RFE são aquelas que não são significativas para impactos nos resultados;
- Para essa análise são utilizadas as variáveis “NOISE”, consistem em variáveis com valores totalmente aleatórios, essas variáveis permitem selecionar com exatidão as features mais pertinentes, uma vez que, uma feature com importância menor do que uma totalmente aleatória não possui representatividade.

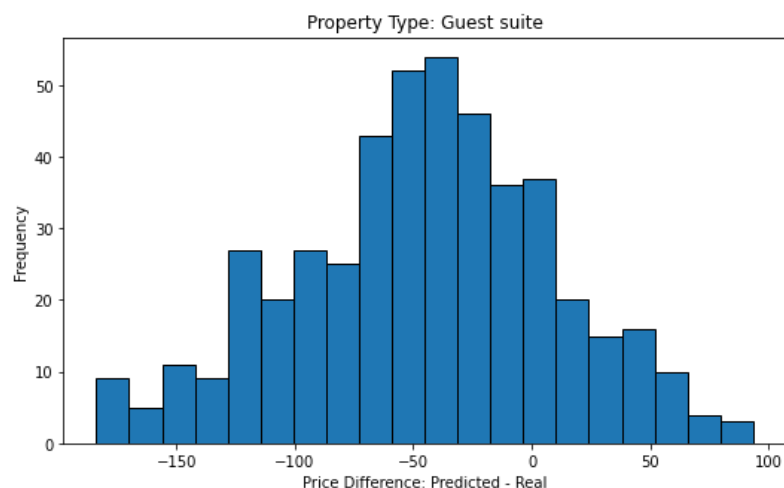
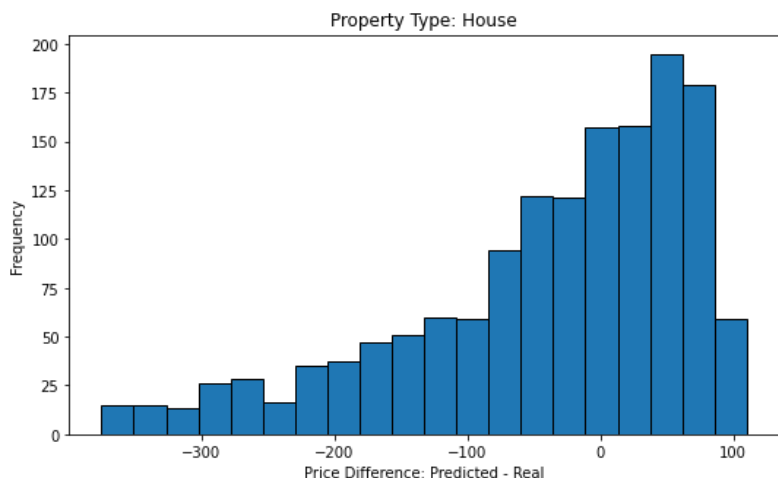
Modelo

- Para separar as bases de dados, foi utilizado como quantidade de locações o número 24, esse valor indicaria, em média, duas locações por mês em um horizonte de um ano;
- O modelo foi submetido ao uso do GridSearch CV para busca dos melhores hiperparâmetros para melhora do resultado do modelo;
- Os resultados para R^2 estão apresentados na tabela ao lado;
- O modelo foi executado nos dados com imóveis com menos de 24 locações em busca da precificação ideal;
- Pipeline da modelagem:

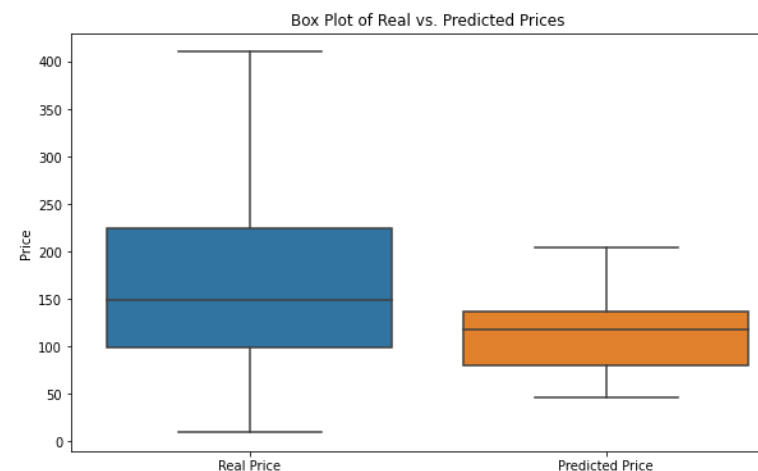
	R^2
Base de Treino	95%
Base de Teste	83%
Base de Modelagem (Completa)	92%



Resultados da Validação



- Como era esperado, o R^2 (-1,9%) para a base de validação apresentou um valor bem baixo, indicando que realmente os valores praticados não estão sendo definidos de forma similar aos dos imóveis com alta quantidade de locação;
- Comparando o preço predito com o real, em alguns tipos de imóveis como “House”, foi observado que os valores praticados poderiam ser maiores;
- Para outros tipos de imóveis esse valor estaria bem abaixo do real;
- A distribuição dos valores preditos foi bem menor do que os valores reais, com seu ponto máximo próximo ao terceiro quartil dos valores reais praticados.



Conclusões e Próximos Passos



Conclusões

A abordagem escolhida demonstrou que os preços praticados poderiam ser alterados nos imóveis com baixa locação, apesar dos valores apontados em alguns casos serem inferiores, indicando uma queda no faturamento por locação individual, o objetivo era encontrar um valor de precificação que aumentasse a procura e a quantidade de locações, resultando em um aumento da taxa de ocupação desses imóveis e assim sua rentabilidade.



Próximos Passos

É notável que os imóveis dispostos na base possuem características bem diferentes entre si, sendo então a abordagem escolhida enviesada por comparar imóveis de características distintas.

Uma evolução da solução proposta que resultaria em resultados melhores, seria o uso de análises conjuntas, criando inicialmente um modelo de clusterização e dentro de cada cluster executar a solução proposta, garantindo assim que vamos comparar imóveis similares.



OBRIGADO!

Pedro Beaumord Diniz

Telefone:

(31) 986472403

Email:

pedrobeaumord@outlook.com

