Utilizando dados de Boston primeiro.

Procurar os principais drivers responsáveis pelo preço de um imóvel para aluguel.  
Procurar os principais drivers responsáveis pelo preço de um imóvel para alugar no AirBnB.

Sazonalidade, dia da semana, eventos especiais: isso tudo influencia no número de dias que alguém consegue alugar o seu imóvel.

The guests: entender qual o tipo de público que busca aquela cidade.

The property:

The Market:

The Fees:

The Industry:

Fontes:

<https://learnbnb.com/airbnb-supply-demand-dynamic-airbnb-pricing/>

* Depois de terminar, adicionar geolocalização de pontos de transporte público para entender como isso afeta o preço.
* Adicionar distância em relação ao centro financeiro e centro de turismo da cidade (caso haja mais que um, procurar).

**Prever Reviews Positivas e Negativas**

Utilização de um Sentiment Classifier\ opinion miner

* Bag of Words
* Categorical/Polarity
* Scalar/Degree

<https://github.com/AitorAl/Airbnb-Reviews-Sentiment-Analysis/blob/master/Quick%20Analysis/Boston%20Airbnb%20reviews.ipynb>

Nesse site ele indica como utilizar no pacote do nltk que já possui um sentiment analyzer pronto. Indica a polaridade do texto: quantos % positivo, negativo e neutro.

Sendo assim, posso utilizar esse modelo pronto. A questão é entender como ele funciona. Parece que o Bag of Words vem pronto e isso que possibilita entender a polaridade do texto.

Utilizar o sistema disponível para identificar reviews positivas e negativas. Explorar como é feito esse sistema já disponível para explicar no Medium, ou então indicar onde encontrar. A partir desse sistema dispoível, apresentar alguma análise que leve em conta as outras bases de dados disponíveis. --> uma análise possível é ver se a nota que é dada para cada listing\_id é parecida com o ranking de review encontrado no listings database.

- Discover if a raw text is positive or negative.

<https://towardsdatascience.com/detecting-bad-customer-reviews-with-nlp-d8b36134dc7e>

- Samira Munir: explica como criar o BOW, criar as features para cada review, calcular os scores (negativo e positivo) e treinar um modelo classificador além de testá-lo.

<https://towardsdatascience.com/building-a-sentiment-classifier-using-scikit-learn-54c8e7c5d2f0>

a) Tokenizing, c) Removing Stop words, b) Parts of Speech Tagging, c) Concordance

<https://towardsdatascience.com/testing-the-waters-with-nltk-3600574f891c>

-->> Unsuporvised Sentiment Analysis

<https://www.kaggle.com/pacogiu/sentiment-analysis-from-hotel-reviews/data>