



AIRBNB: RIO DE JANEIRO

Find your next Airbnb with us...

Problem Statement:

1. Accurately predict what a property in Rio De Janeiro would list for on Airbnb.
2. Build Streamlit app allowing people to input the features of their property and generate a price recommendation.

A close-up photograph of a person's hand wearing a dark long-sleeved shirt. The hand is gripping a black spray bottle and spraying a fine mist of water onto a dark, textured surface. The lighting is dramatic, coming from the side to highlight the spray and the contours of the hand.

DATA CLEANING

- Using Functions & Regex to remove special characters and spaces from key features such as bedroom, baths, & price
- Reformatting features from objects to categorical, integers, or floats
- Replacing missing values: imputing with linear regression, then later median, mode(bed / baths)
- Dropping unneeded columns

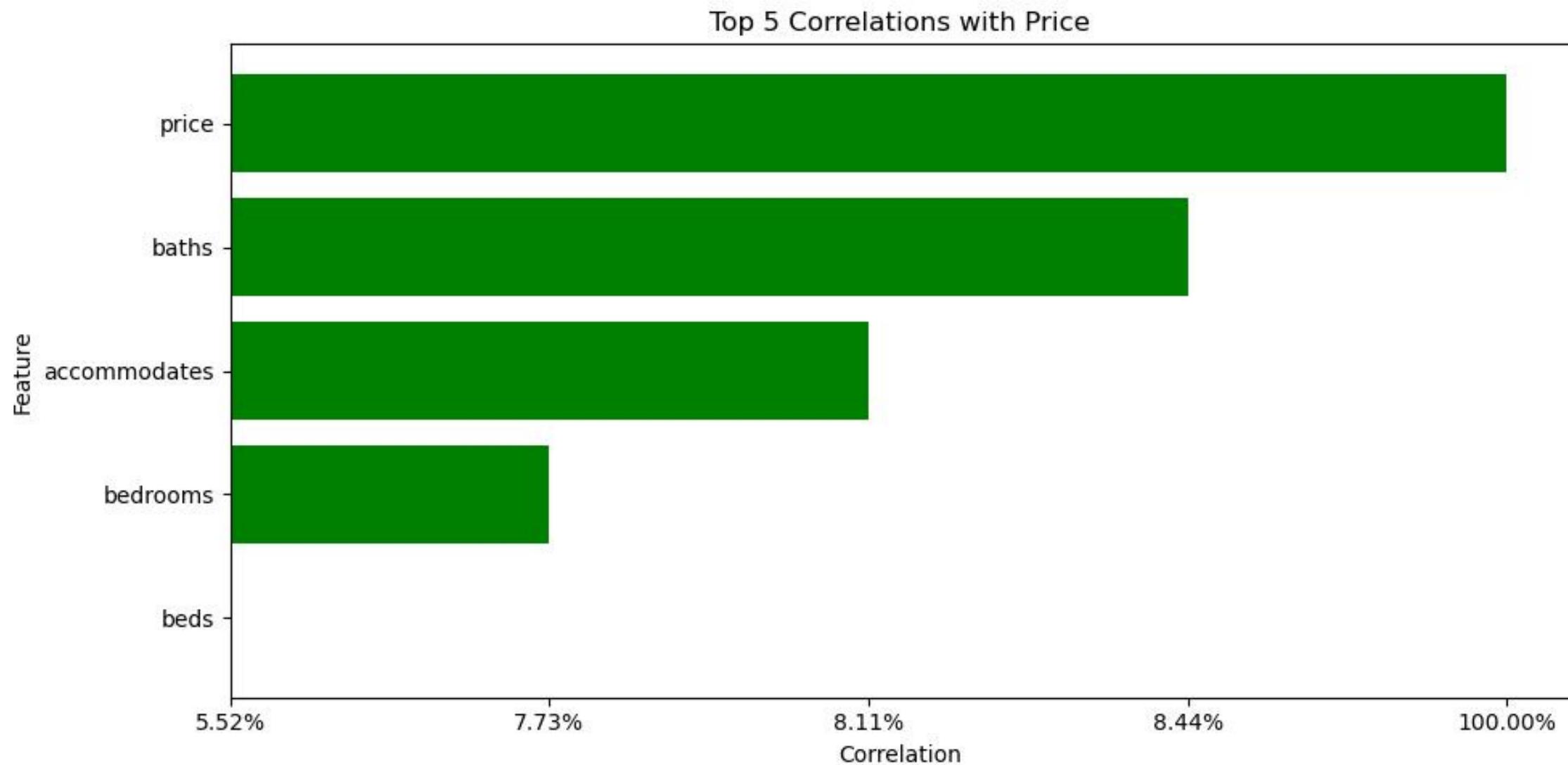
EXPLORATORY DATA ANALYSIS

Early issues....

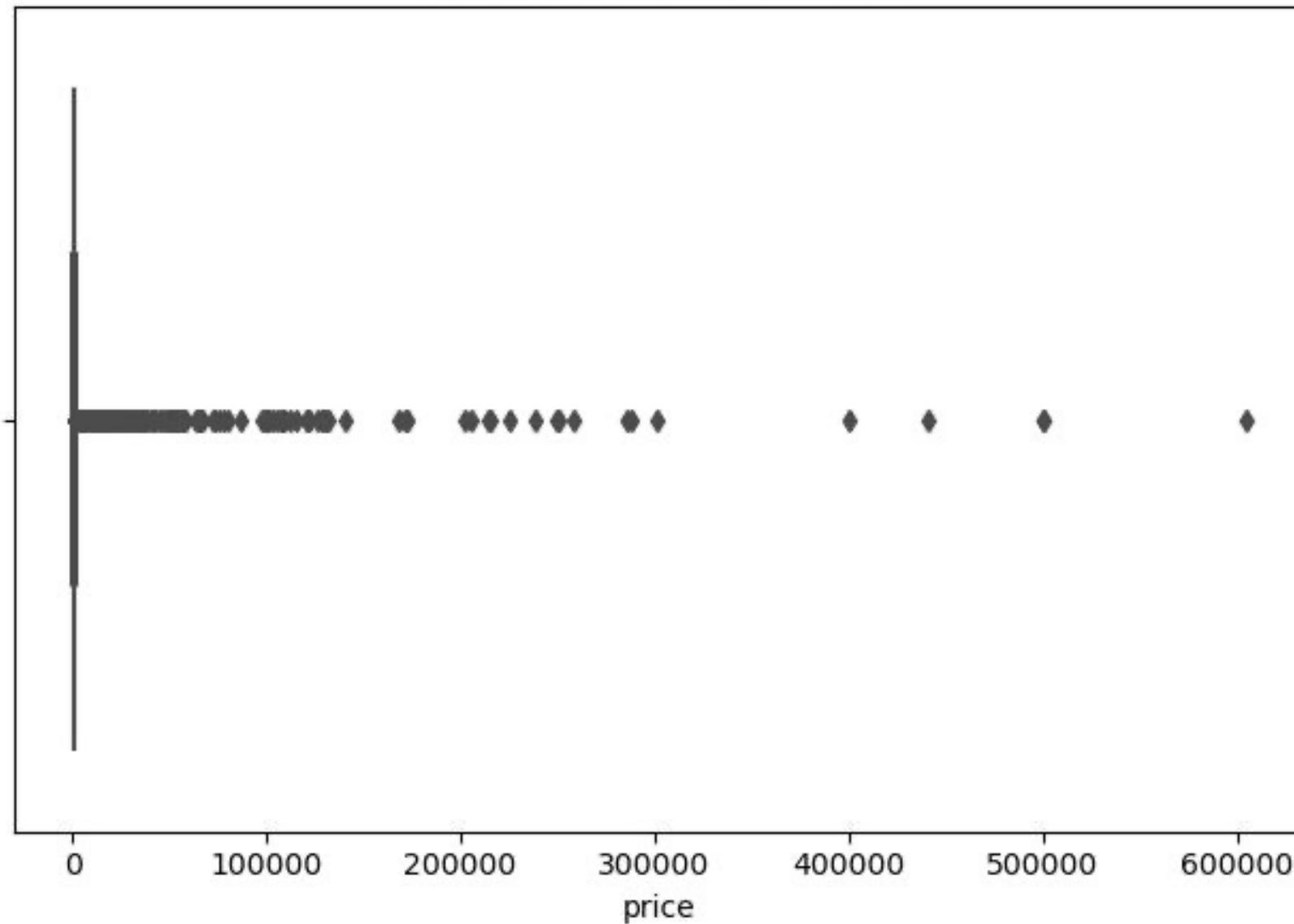
1. Huge range in listing price (few hundred dollars to a few hundred thousand dollars).
2. Reason for range difficult to decipher
3. Very little correlation between features and target variable price



Price as it correlates to features...

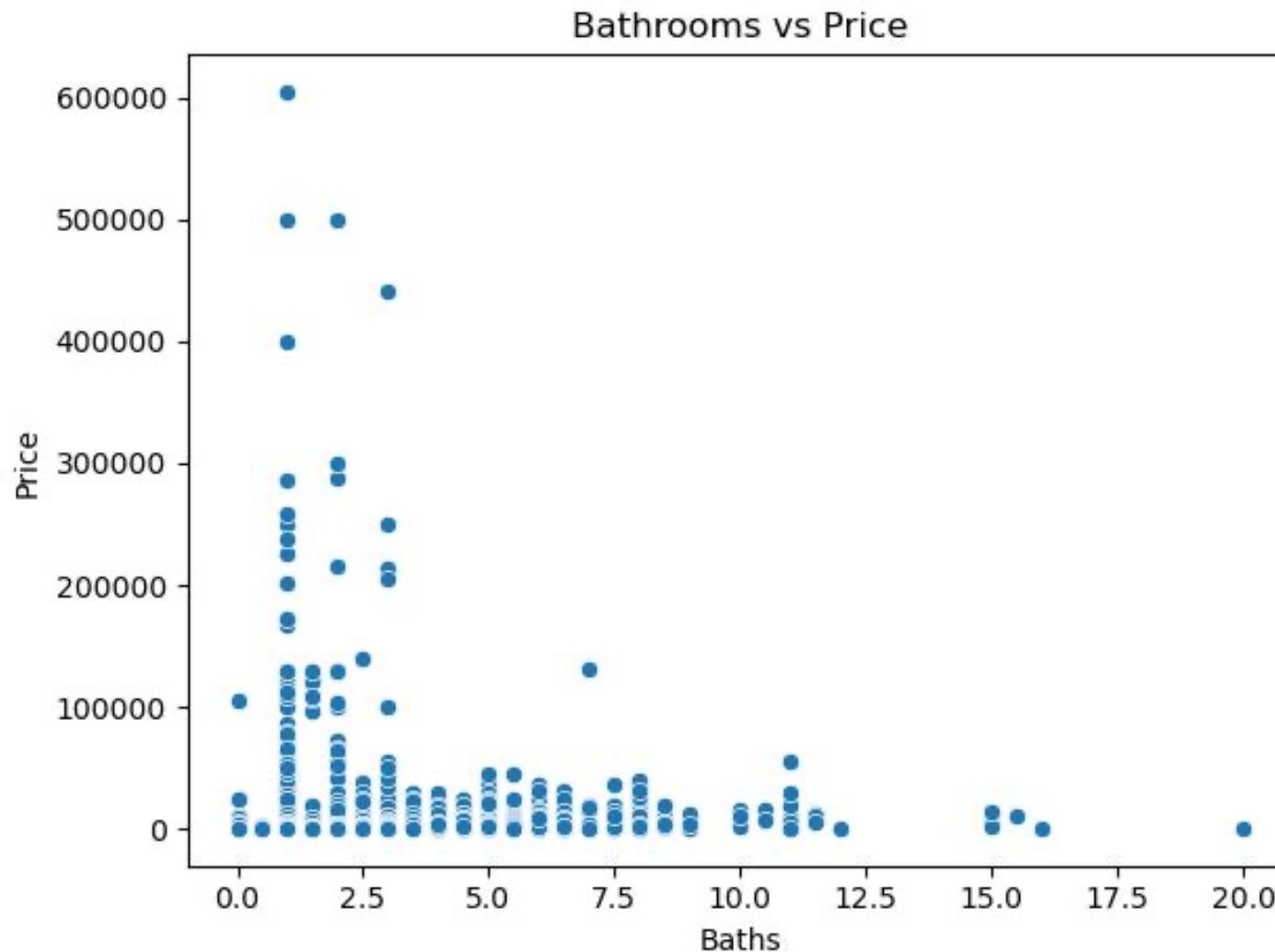


Rio Airbnb : Price - Boxplot



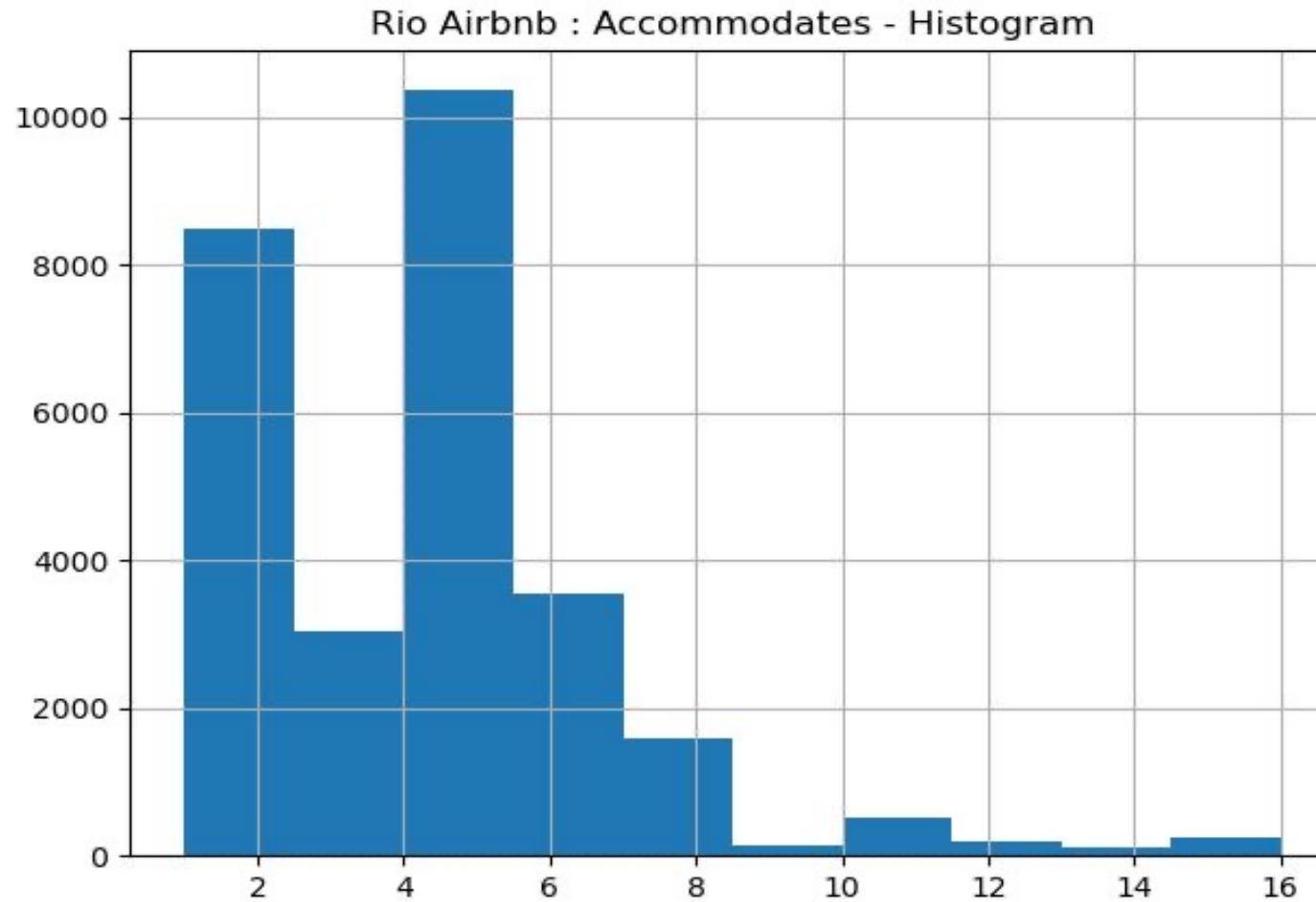
While most listings are clustered at a few hundred or a few thousand dollars, there are several that are in the tens to hundreds of thousands of dollars.

Bathrooms as they correlate to price...



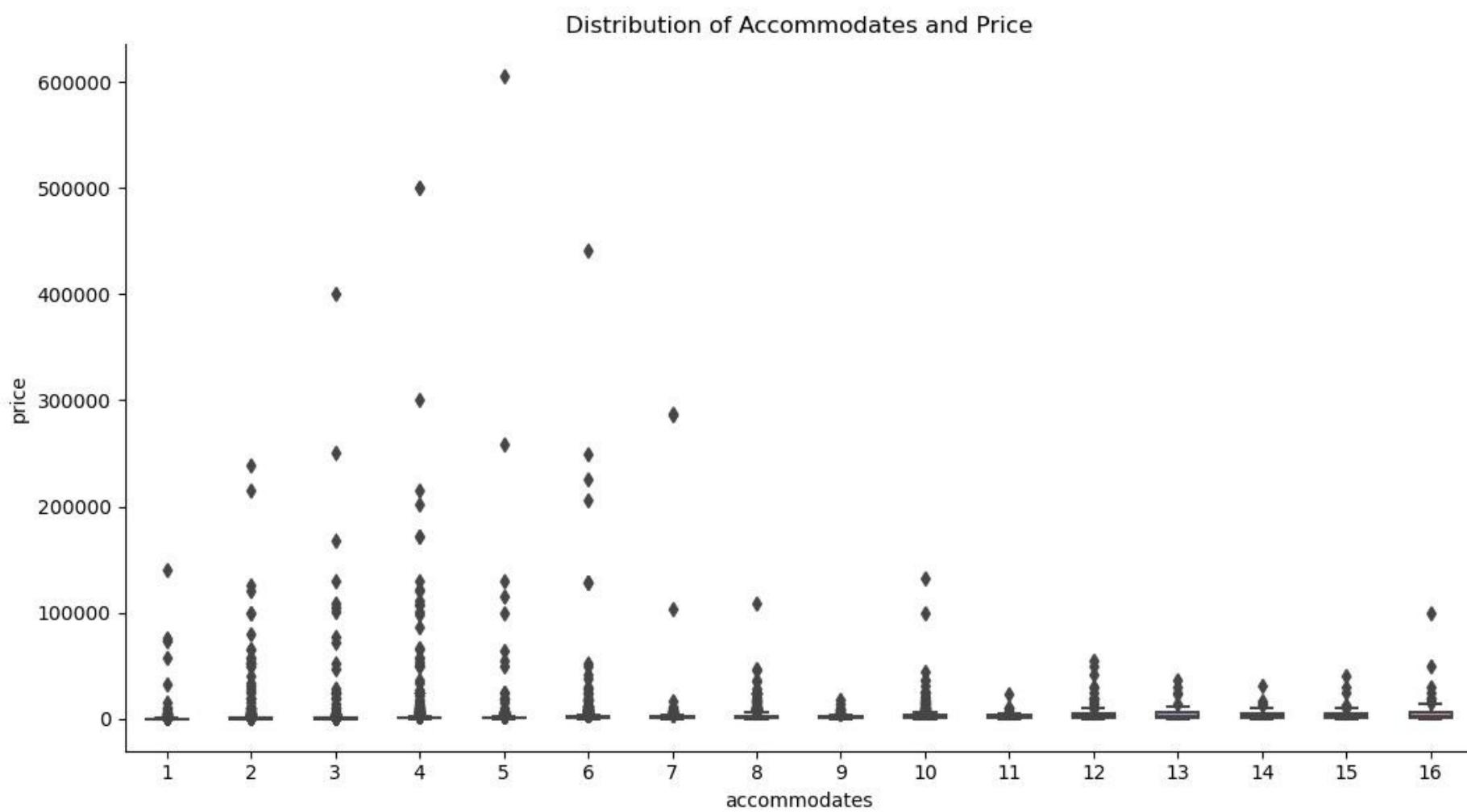
★ Several listings in the tens of thousands of dollars per night range, and many even higher

Accommodates as a feature...



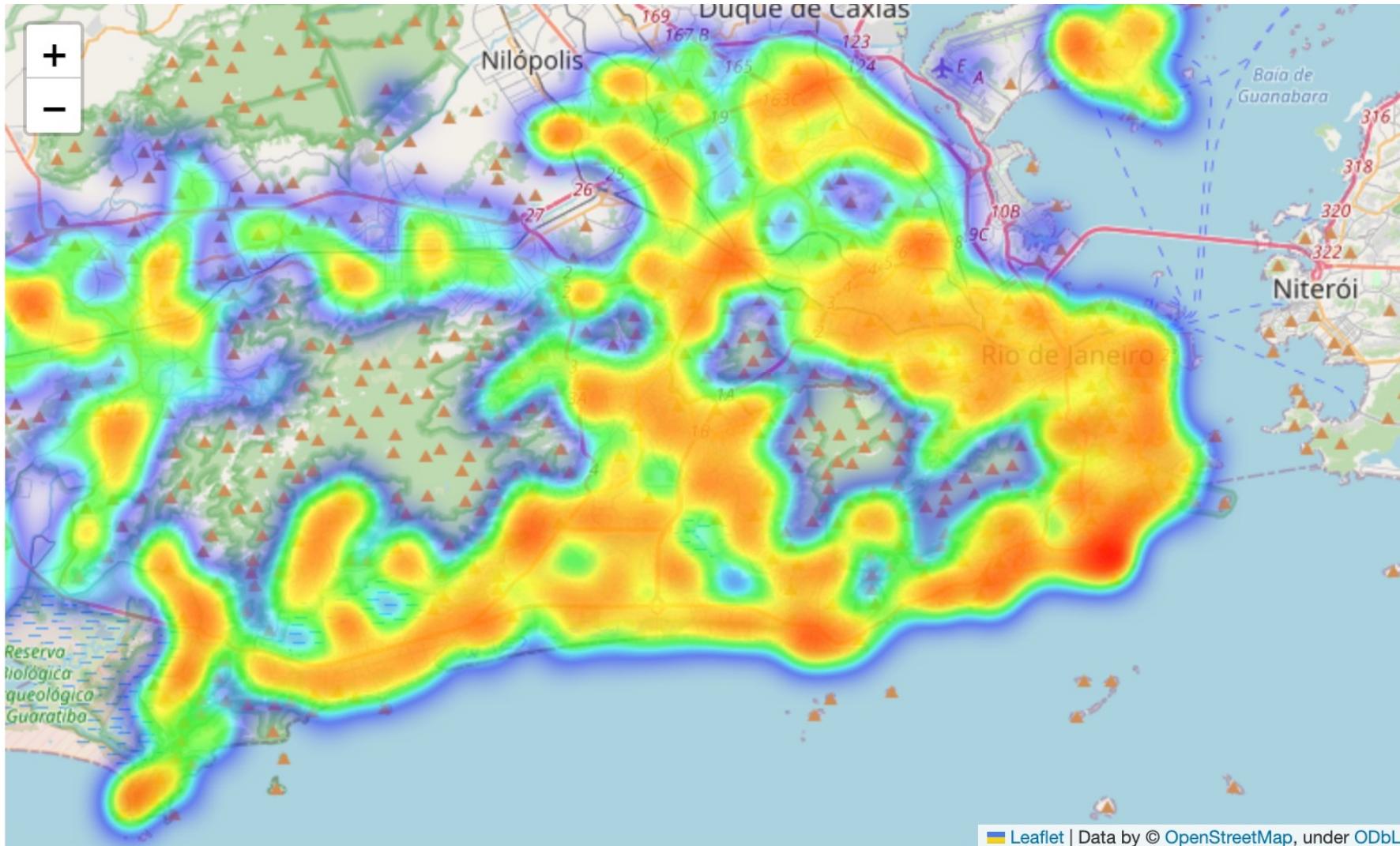
- ★ Most listings will accommodate four to six people

Plotting Accommodates and price...



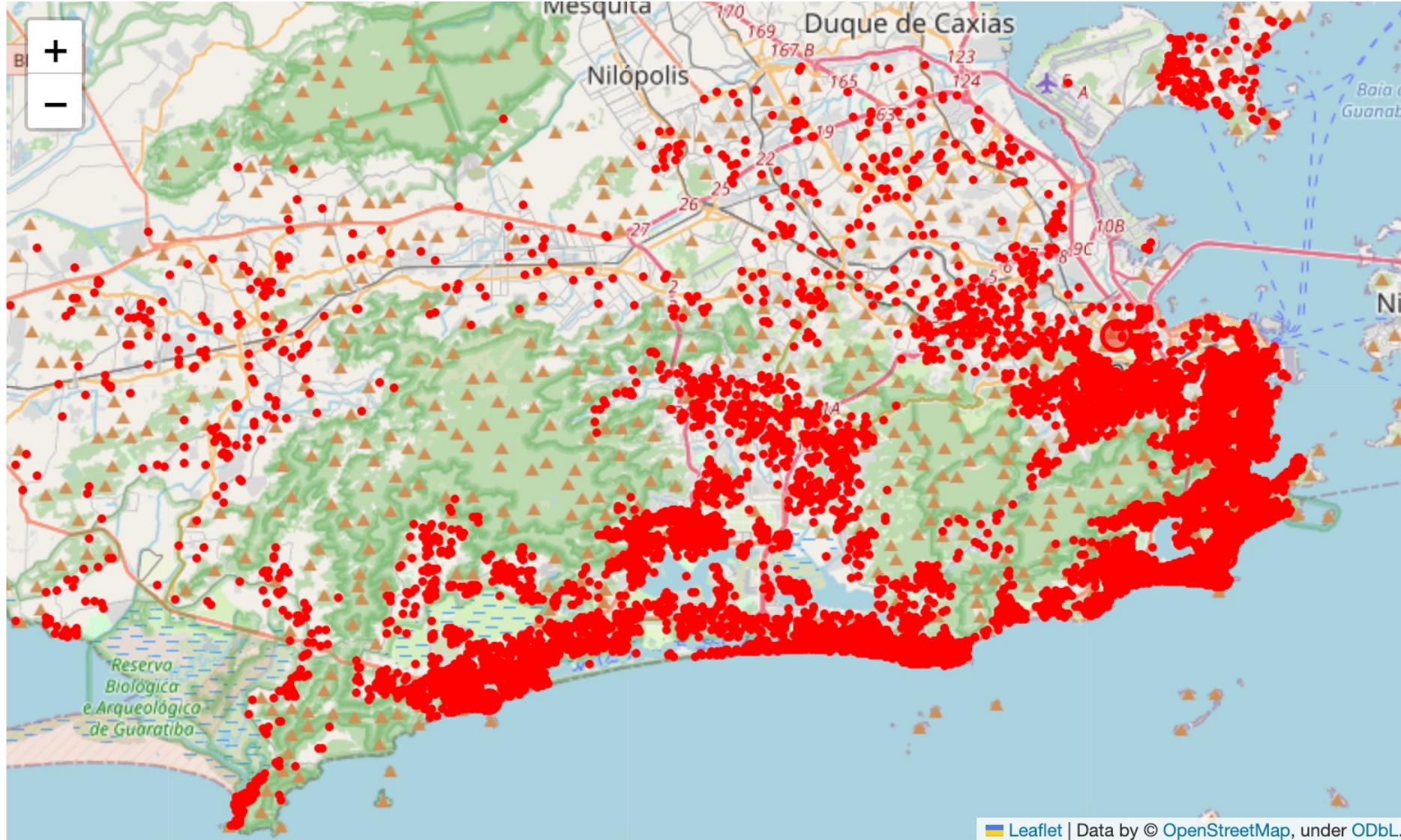
★ The number of people a listing accommodates doesn't seem to be a clear indicator for predicting price

Location, Location, Location?

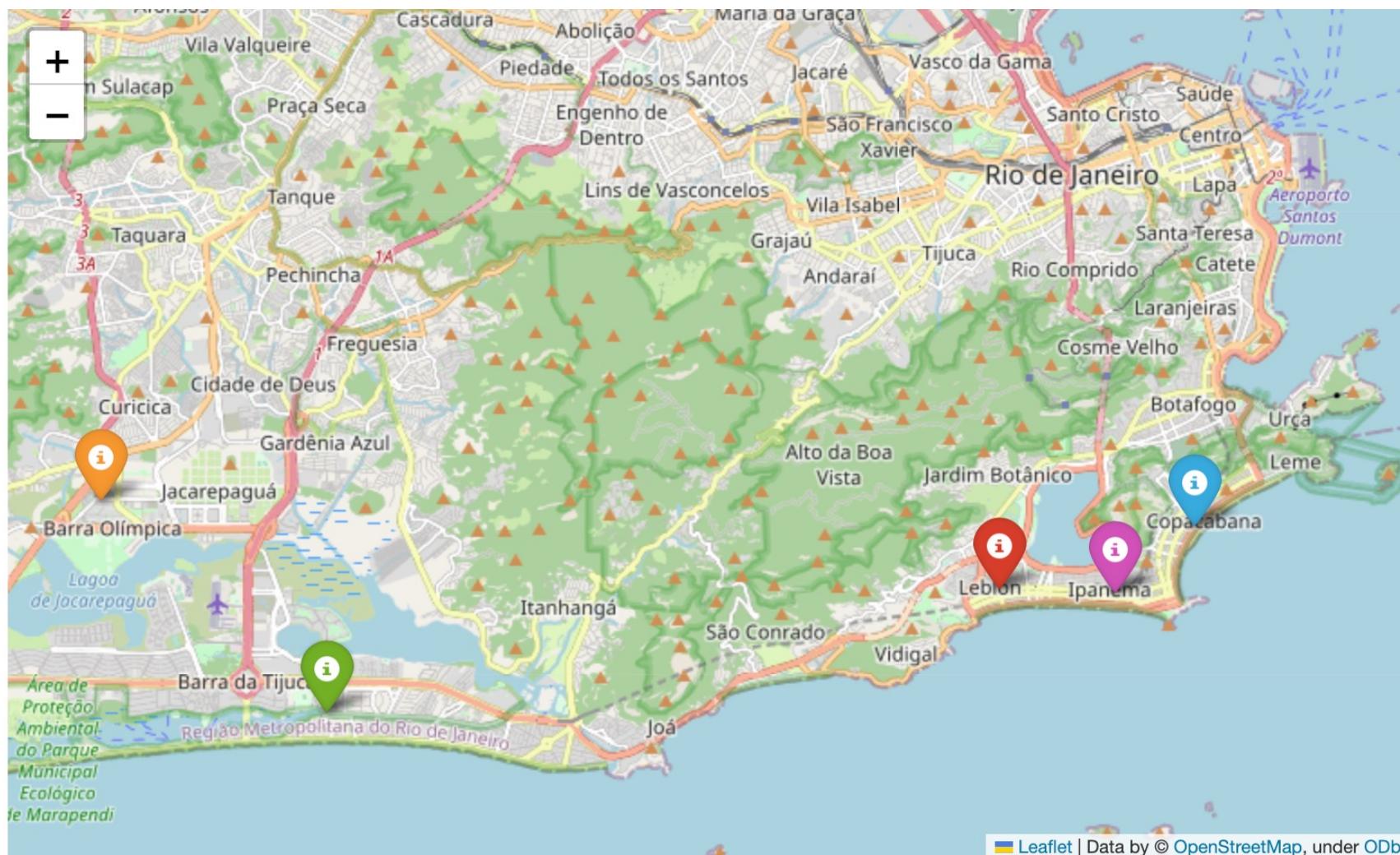


★ The topography of Brazil and popularity of the beach lead to most Airbnb listings being highly concentrated along the coast

Location seen yet another way...



Most expensive neighborhoods in Rio...



1. Leblon
2. Ipanema
3. Copacabana
4. Barra de Tijuca
5. Barra de Olimpica

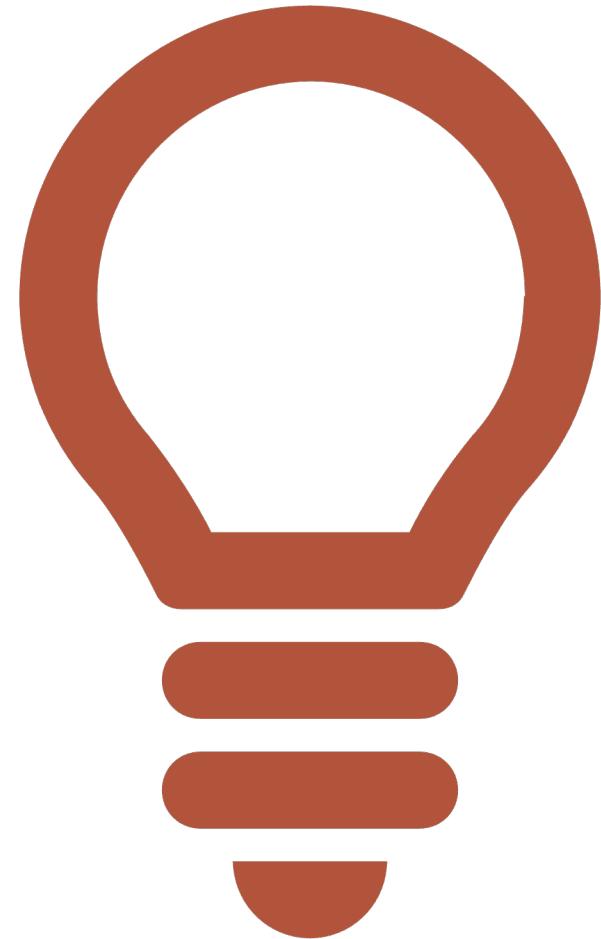


Attempts to Model

Revisited Cleaning Data for modeling...

- Polynomial Features
- Feature Selection
 - PCA
- One Hot Encoding Neighborhoods
- Highest score reached on a test score was .02%





RECOMMENDER SYSTEM



- Interactive recommender system on Streamlit that suggests 10 additional Airbnb options to clients based on a selected listing
- Uses TfidVectorizer to analyze the description, accounting for Portuguese and English
- Runs through Cosine Similarity Matrix

Conclusions & Next Steps:



Regression Model & Price Recommendation:

- Further exploring correlation between features and pricing in Rio.
- Explore feature extraction, engineering, PCA, and predictive pricing.

Listing Recommender:

- Concatenating additional columns to evaluate performance
- Better organization and naming of listings both for searching and recommending
- Adding ability to input features on app based on features (price, accommodates, bed, baths, etc.)

