Price Optimization Analysis for AirBnB

About Dataset

Context

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019.

Content

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

Acknowledgements

This public dataset is part of Airbnb, and the original source can be found on this website.

Inspiration

- What can we learn about different hosts and areas?
- What can we learn from predictions? (ex: locations, prices, reviews, etc)
- Which hosts are the busiest and why?
- Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Introduction & Objective

Business Objective - To create a pricing strategy for Airbnb, which helps Airbnb hosts set the right price for their Airbnb listing and provides customers the benefit of cost.

Type of Problem - Regression Analysis (Dependent Variable is 'Airbnb listing price per night (in USD)' which is regressed against a bunch of independent variables (listing attributes - accommodates, bedrooms, bathrooms, beds, amenities provided, neighborhood and the room type).

Metrics - RMSE (Root Mean Squared Error) to check which ML model provides accurate predictions, Cross Validation Score for hyperparameter tuning in certain ML models, R^2 and Adjusted R^2 for explainability power and to check model fit.

Data Science Lifecycle:

Business Understanding

Data Acquisition and Understanding

2.1 Raw to Relevant Data

2.2 Data Type Inspection and Conversion

2.3 Dirty Data due to Constraints

2.4 Outlier Detection and Treatment

2.5 Handling Missing Values

2.6 Exploratory Data Analysis (Correlation Analysis, Visualizations & Statistical Analysis)

2.7 Handling Categorical Variables for ML Modeling (Text Encoding)

2.8 Train and Test Split and Data Scaling/Feature Scaling

Modeling

Deployment

Customer Acceptance