Chicago’s Airbnb price predictor

Profesor: Shafqat Shad
Students: Rabab Mohamed, Gabriel Palacios

Objective

As people try to get into new businesses to increase their income after a hard year, we found that there is no way to predict how much someone should charge for an Airbnb listing. To make that transition easier we want to create a model that can predict the price of the market for a new place taking into account its characteristics. This model then could be used to create web applications or mobile apps and help people to get into the Airbnb business.

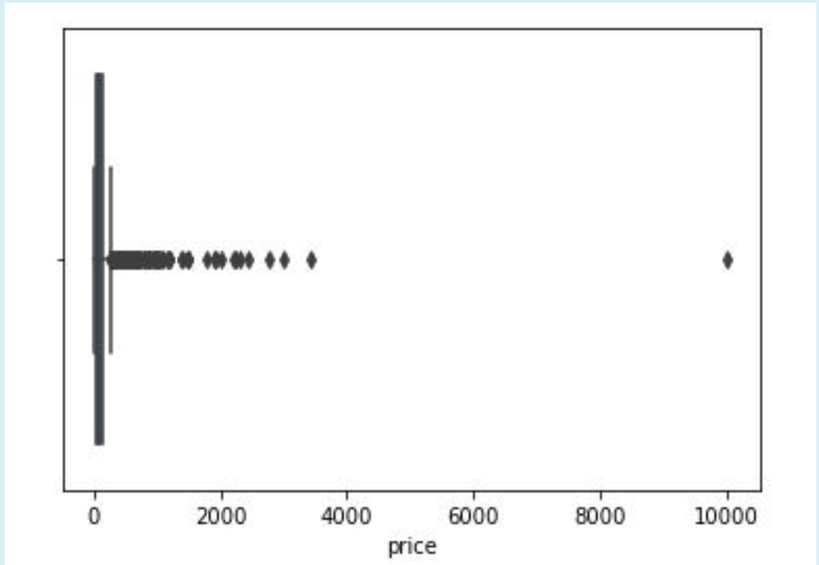
Data

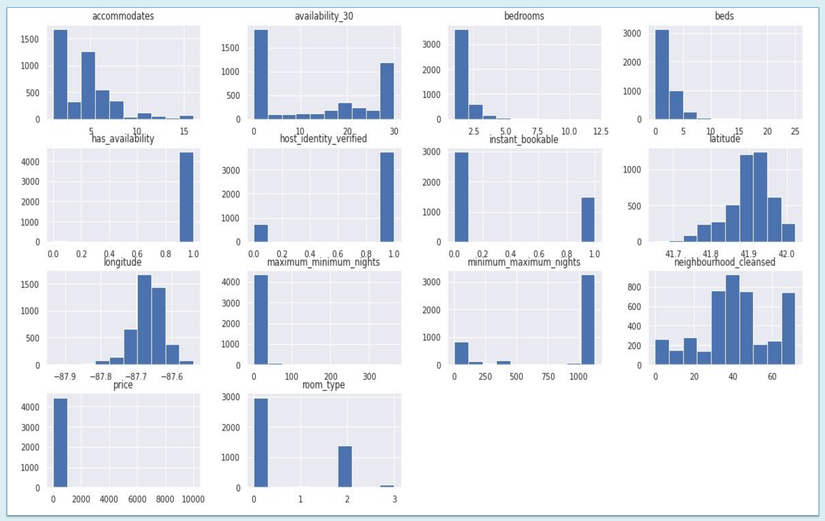
Source: This dataset belongs to *Inside Airbnb*. Specifically, we will use Chicago's dataset posted on December 20th, 2020. The dataset consists of 6523 rows and 74 columns.

The dataset after cleaning consists of 4179 rows and 14 columns. The data types are integers and floats.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4477 entries, 0 to 6522
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   host_identity_verified  4477 non-null   int64
 1   neighbourhood_cleansed  4477 non-null   int64
 2   latitude                4477 non-null   float64
 3   longitude               4477 non-null   float64
 4   room_type               4477 non-null   int64
 5   accommodates            4477 non-null   int64
 6   bedrooms                4477 non-null   float64
 7   beds                    4477 non-null   float64
 8   price                   4477 non-null   float64
 9   maximum_minimum_nights  4477 non-null   int64
 10  minimum_maximum_nights  4477 non-null   int64
 11  has_availability        4477 non-null   int64
 12  availability_30         4477 non-null   int64
 13  instant_bookable        4477 non-null   int64
dtypes: float64(5), int64(9)
memory usage: 524.6 KB
```

Data Visualization





Results

R-squared	Training Set	Testing Set
Multilinear regression	0.35	0.29
Polynomial Regression degree= 4	0.61	0.99
Polynomial Regression degree= 6	0.92	0.99

Conclusion

- The model with the highest R-squared value was polynomial regression with degree = 6.
- Airbnb's characteristics explain 92% of the total variation in the Airbnb price.