

Predicting Airbnb Prices in New York

PRESENTATION



Overview - Business Problem

- Airbnb is an open platform, where hosts set the price.
- There are many variables that influence price.
- In this project, I create a predictive model for pricing, taking into account features like location and amenities.

Business Problem



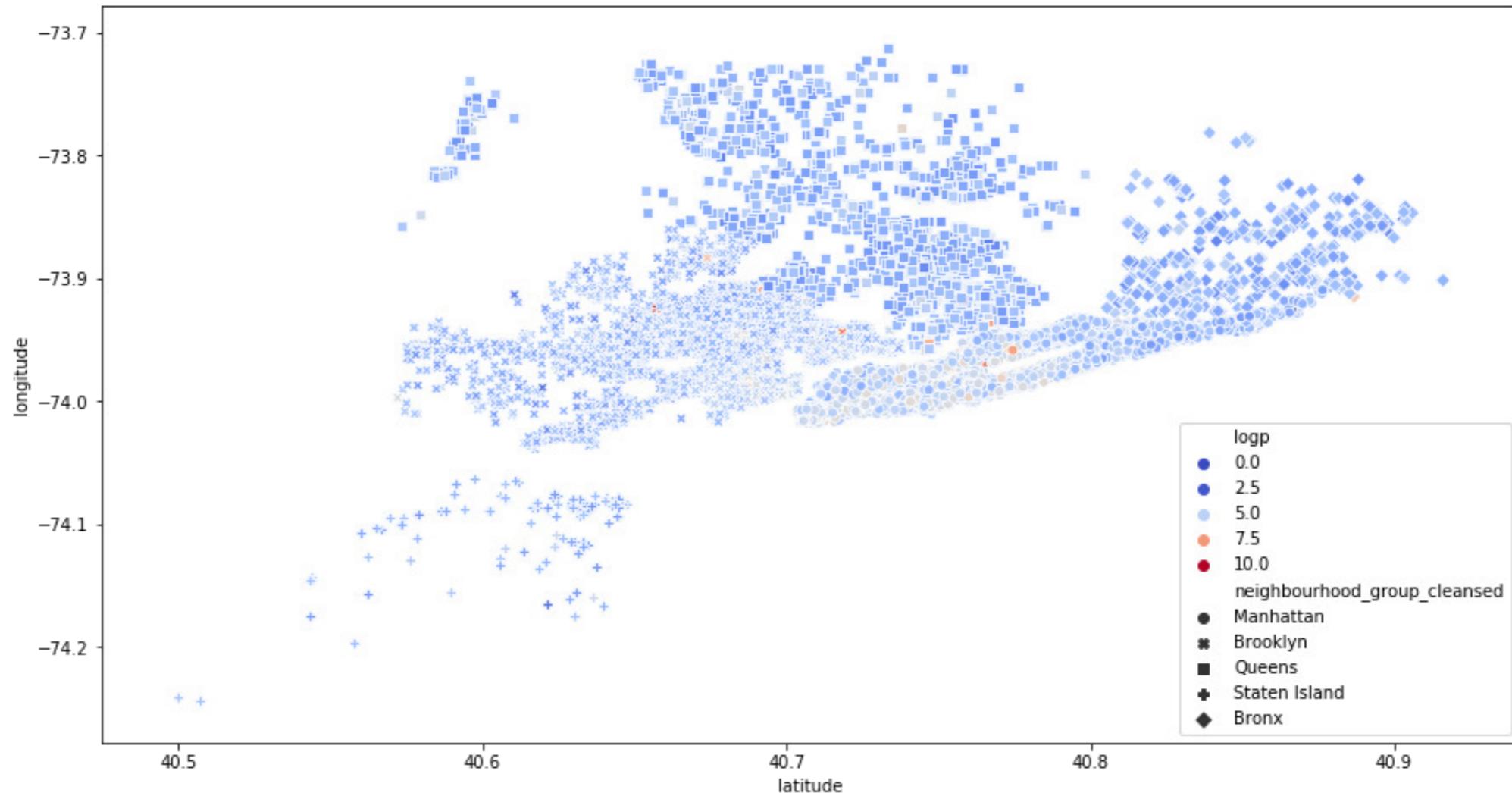
Understanding price is important for guest, host and people interested in knowing how these market work (hotels, public policy makers, Airbnb and other businesses).



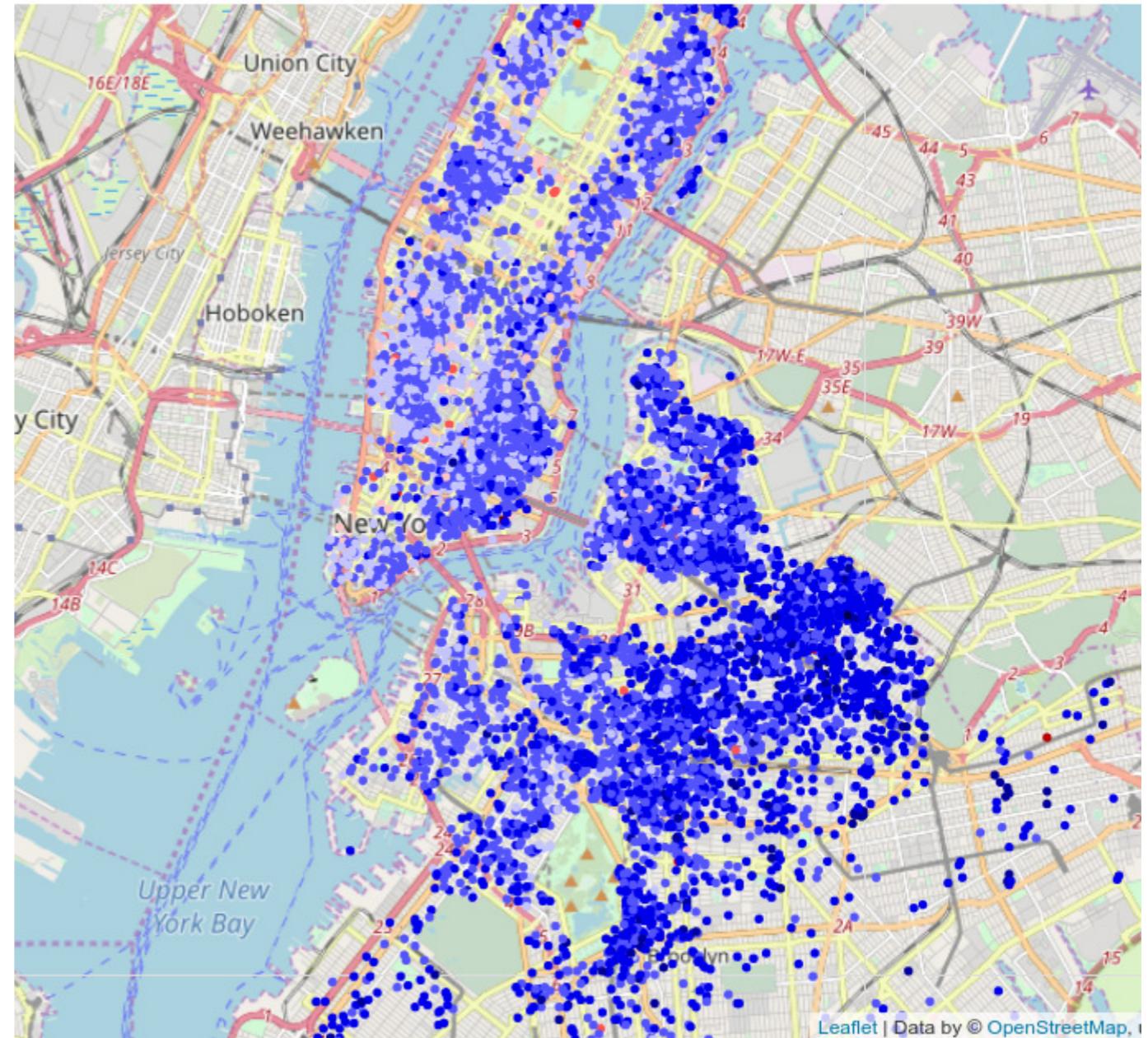
Data acquisition and cleaning

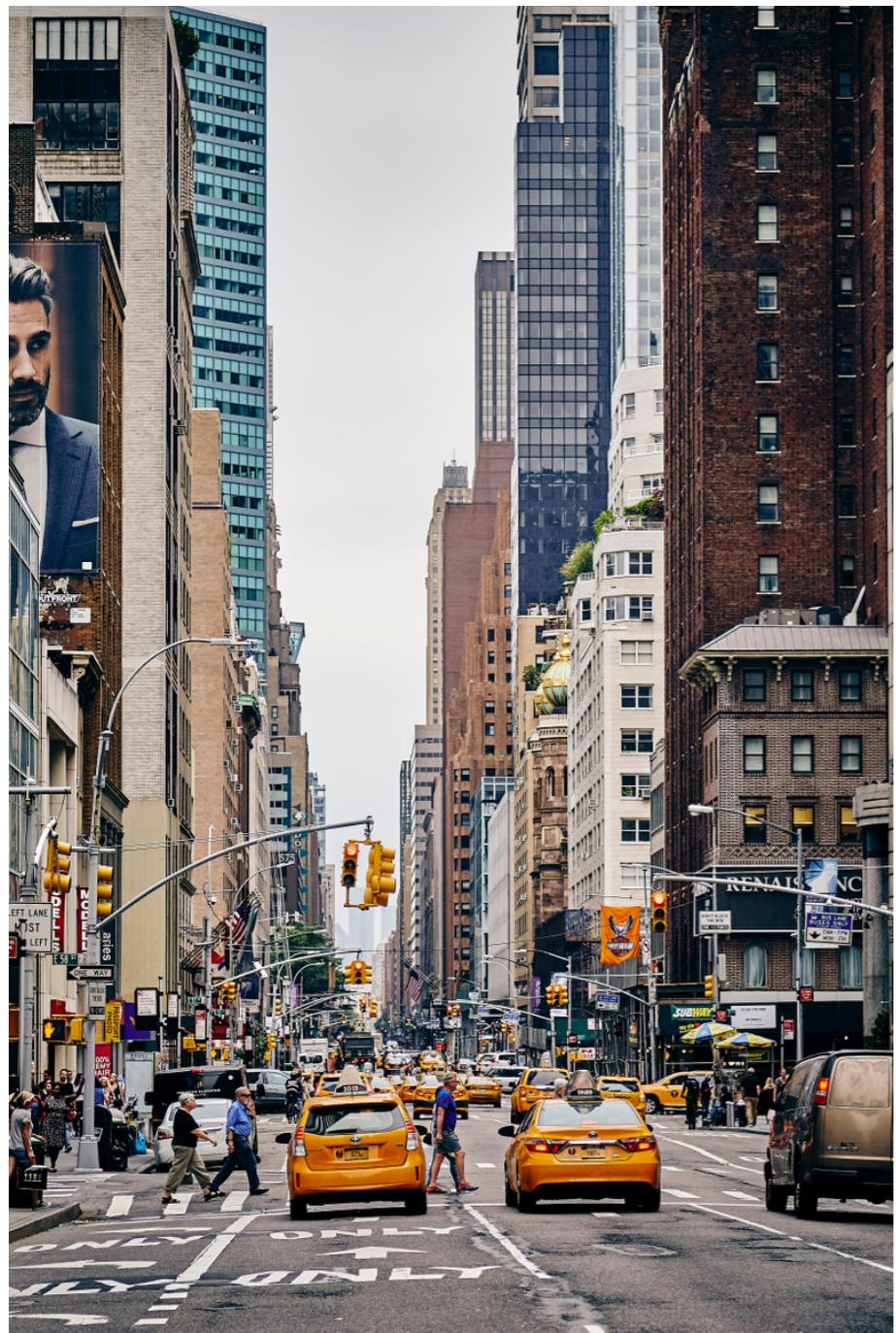
- Airbnb listing data from
<http://insideairbnb.com> - 04 December, 2019
- Venues data from Foursquare API
<https://developer.foursquare.com/>
February 2020

Airbnb Price Distribution over Latitude and Longitude



Airbnb Price Distribution over Latitude and Longitude

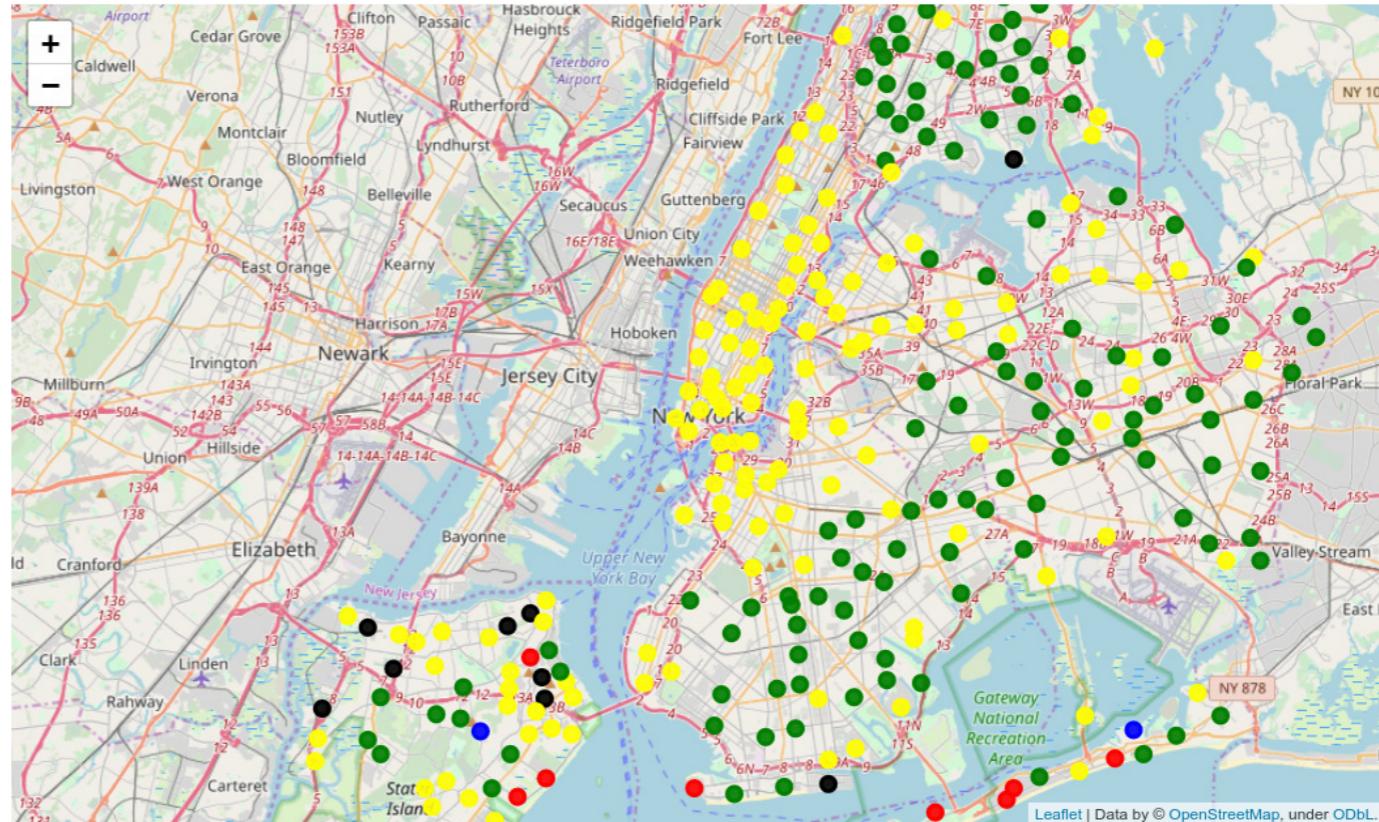




Venues Clustering

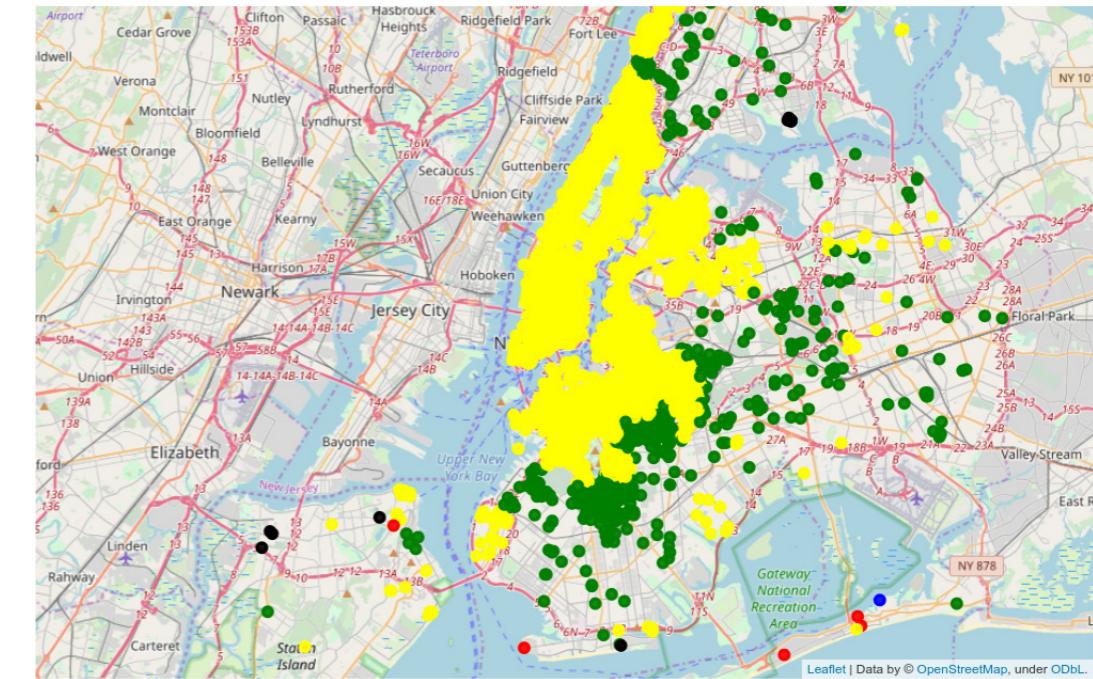
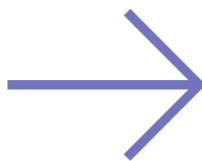
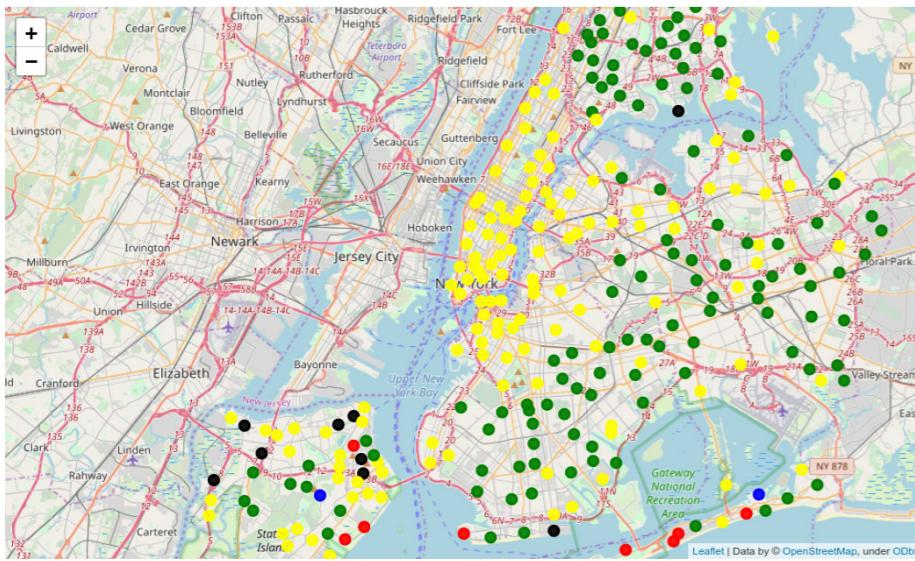
NYC neighborhoods were grouped into 5 clusters

using Foursquare API data and an KNN model.

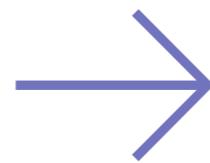
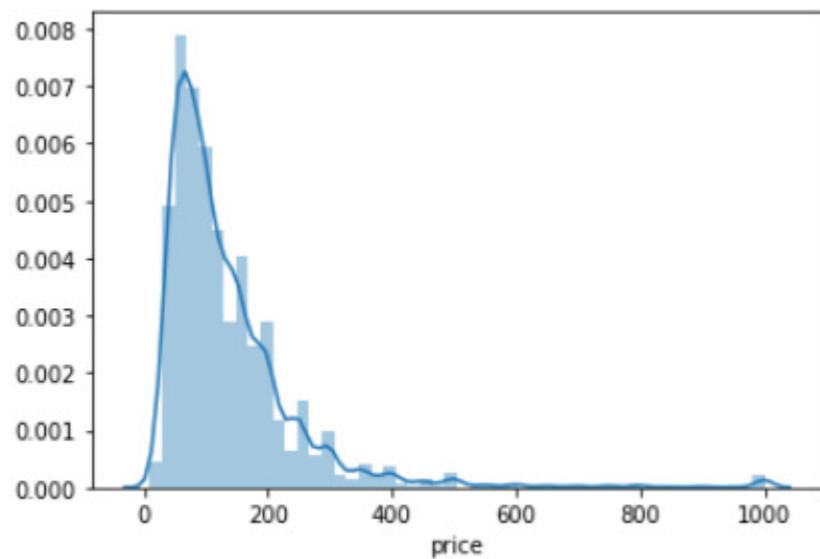


Airbnb Locations to Cluster

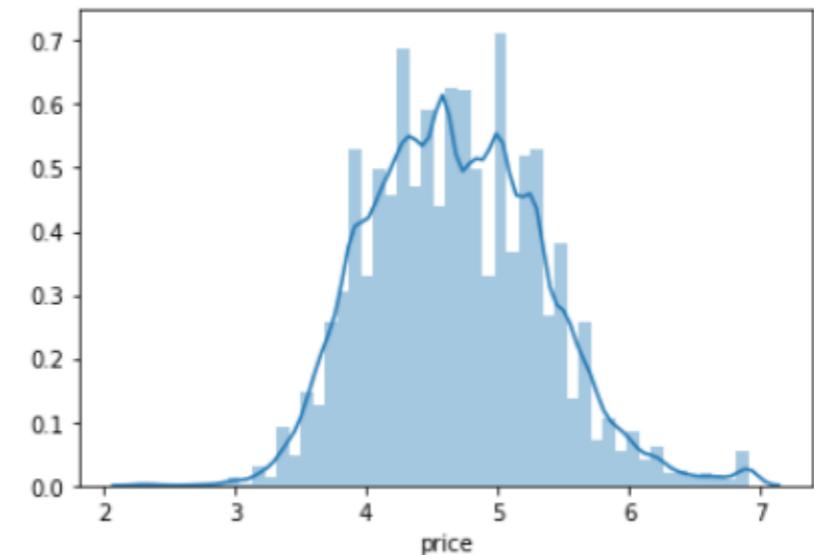
Cluster is a new model feature



Data Preparation



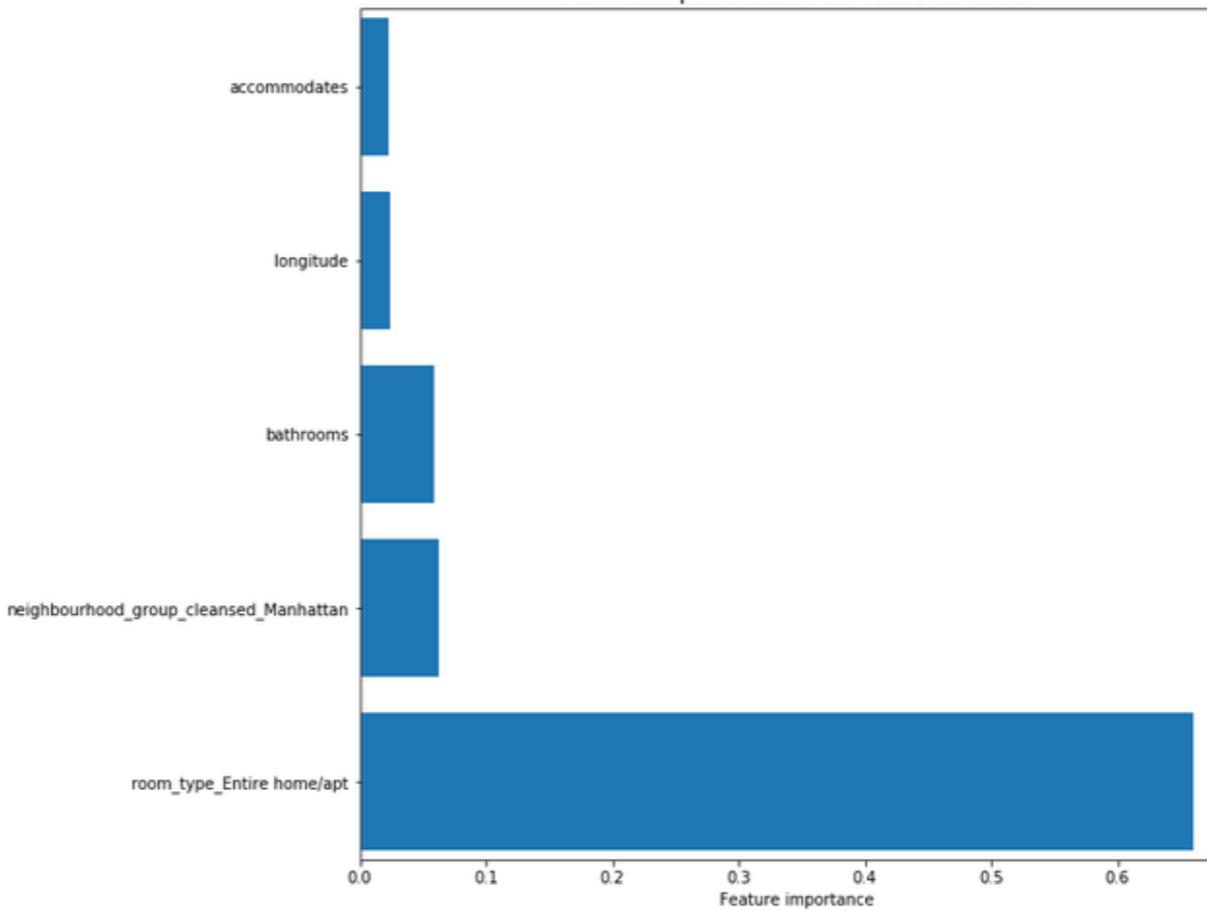
Log Transformation



Modelling: Xgboost 1



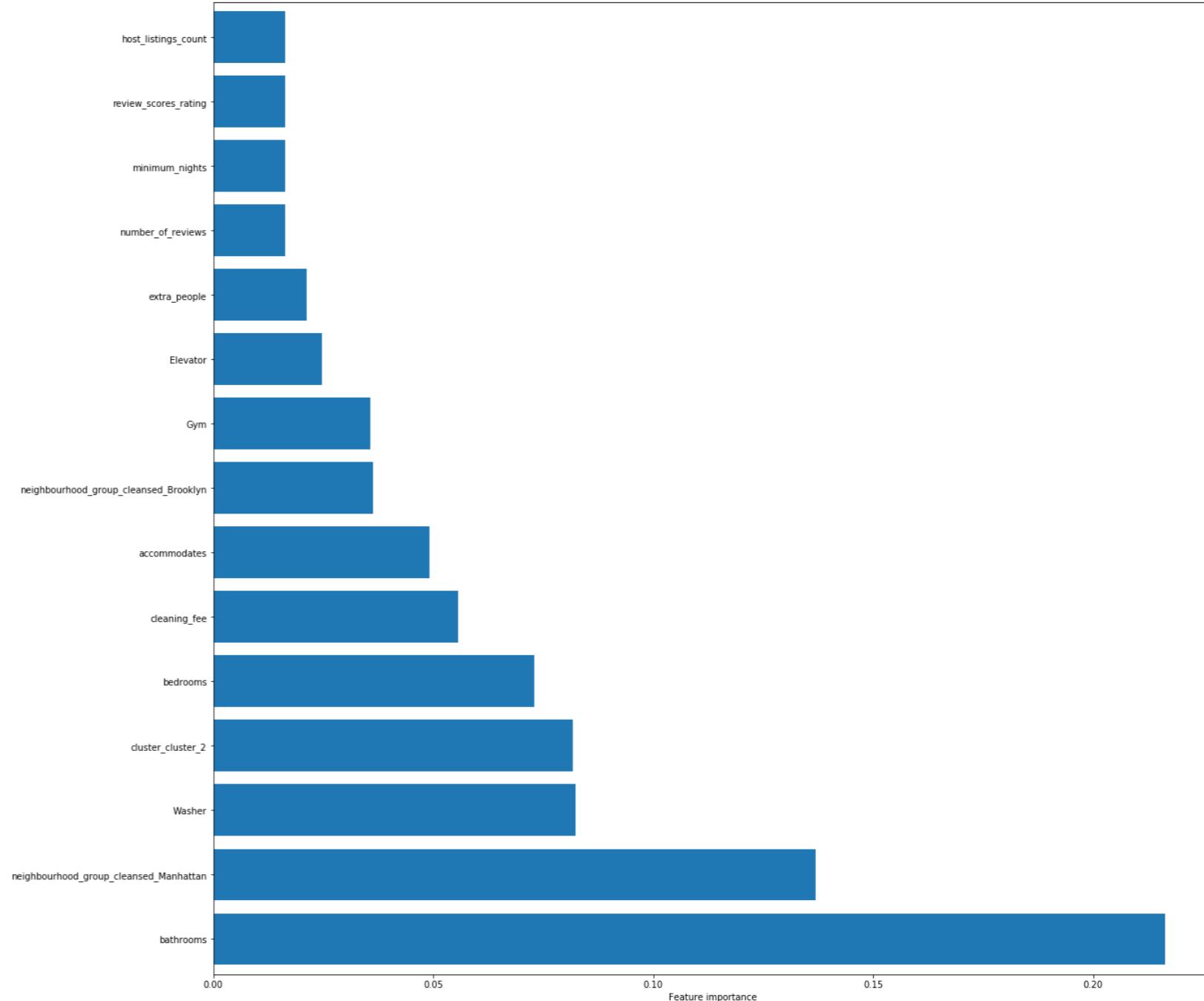
Room Type is the most important feature



Feature importances in the XGBoost model

Modelling: Xgboost 2

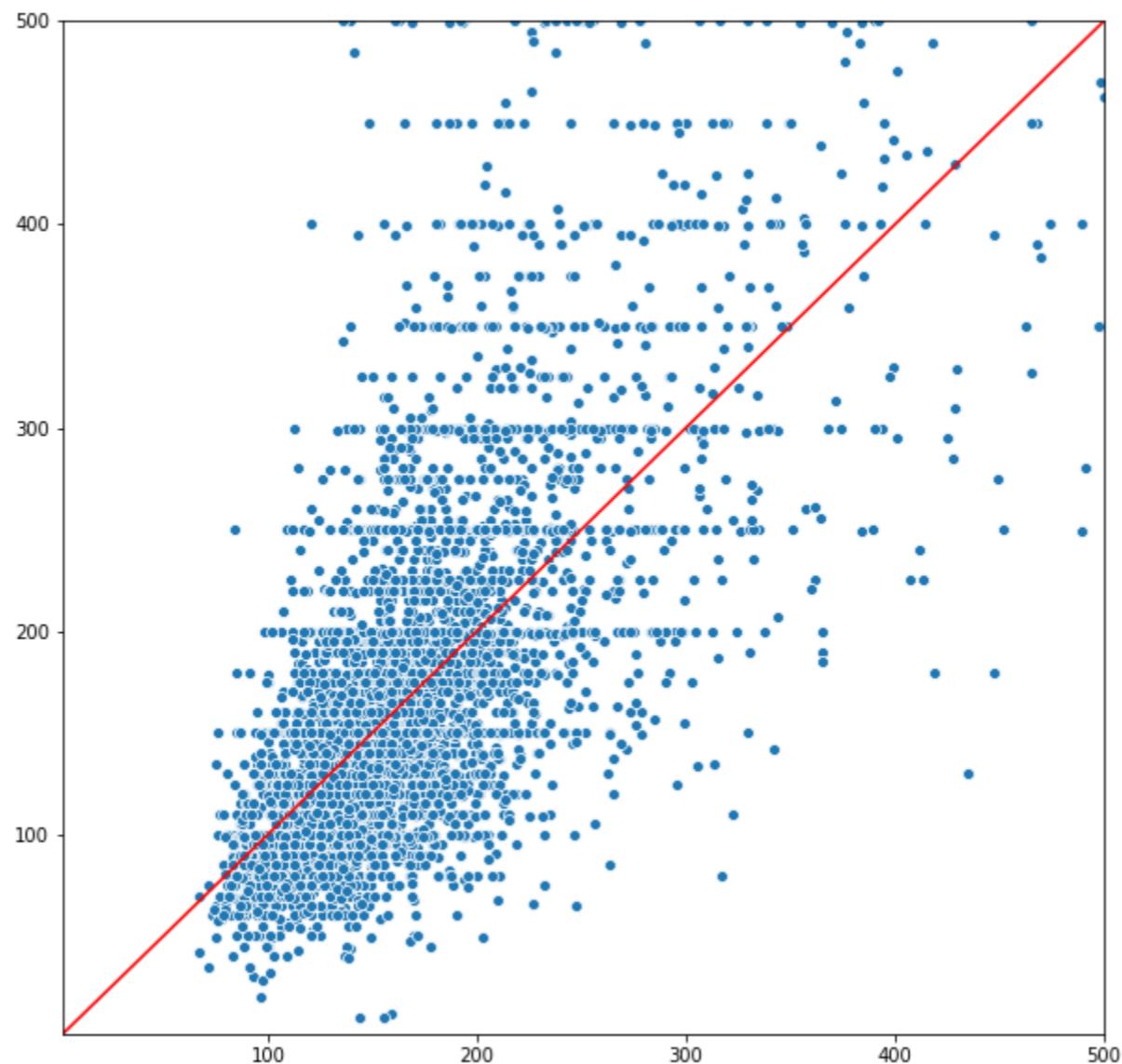
Focusing only in entire properties



Tree



Model Accuracy





Conclusion and future directions

- With a limited set of features, built an useful model for price prediction.
- Location and amenities are important features.
- Review score didn't play an important role.
- Future research could include distance to important venues.