# Predicting Prices for Airbnb Accommodations

Capstone Project

# Problem Statement

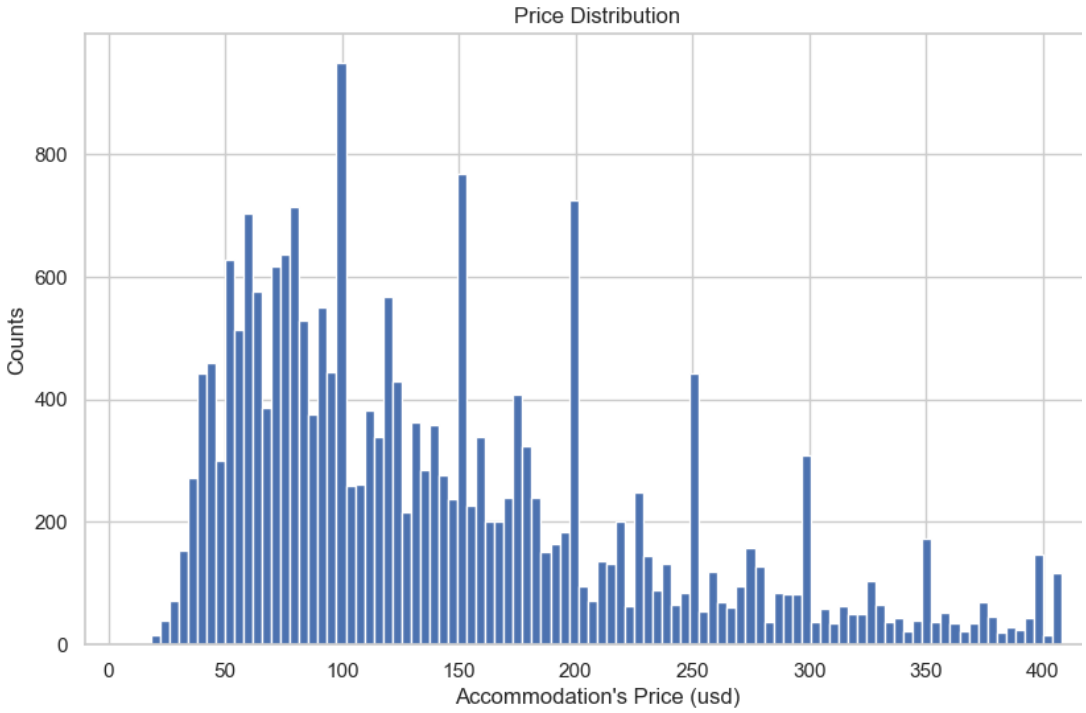
- About: Renting accommodation's site
- Audience: primary and secondary
- Metrics: Regression Problem - RMSE

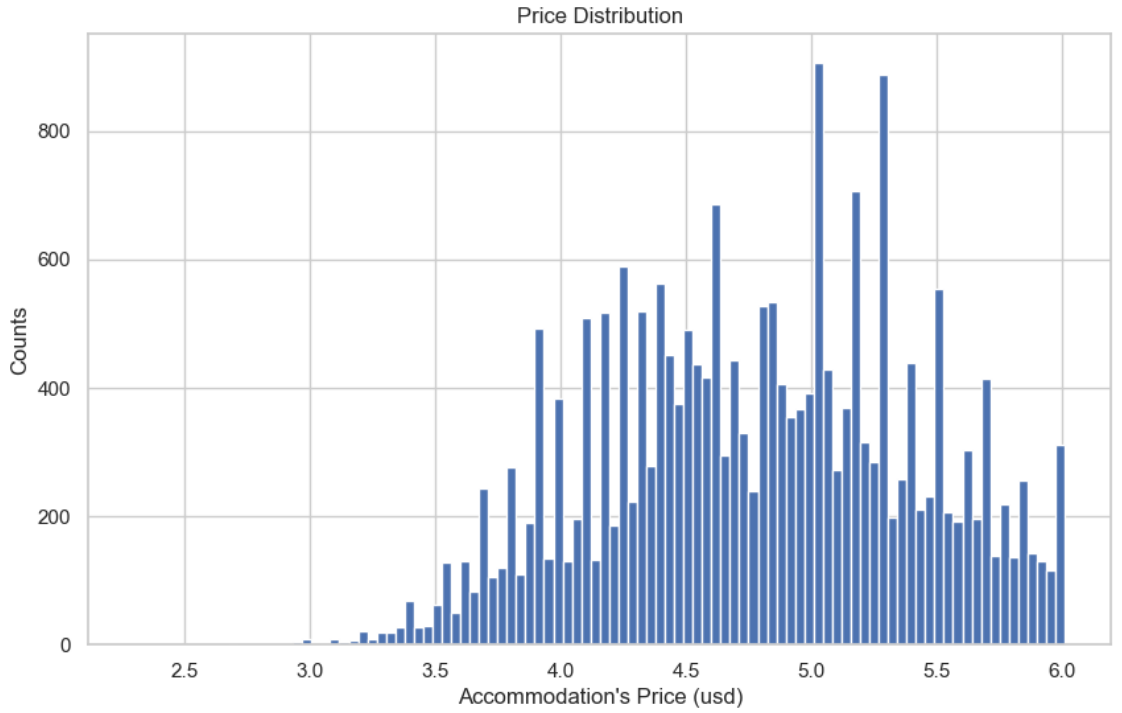
# About Data

- **Datasets:**
  - Listing accommodations New York City;
    - 38,000 x 75
  - Neighborhoods price;
    - ~180 neighborhoods
- **Cleaning Process:**
  - Null values, Input Strategies, Regex, Feature Engineering, handle outliers;
    - 25,000 x 23

# Exploratory Data Analysis

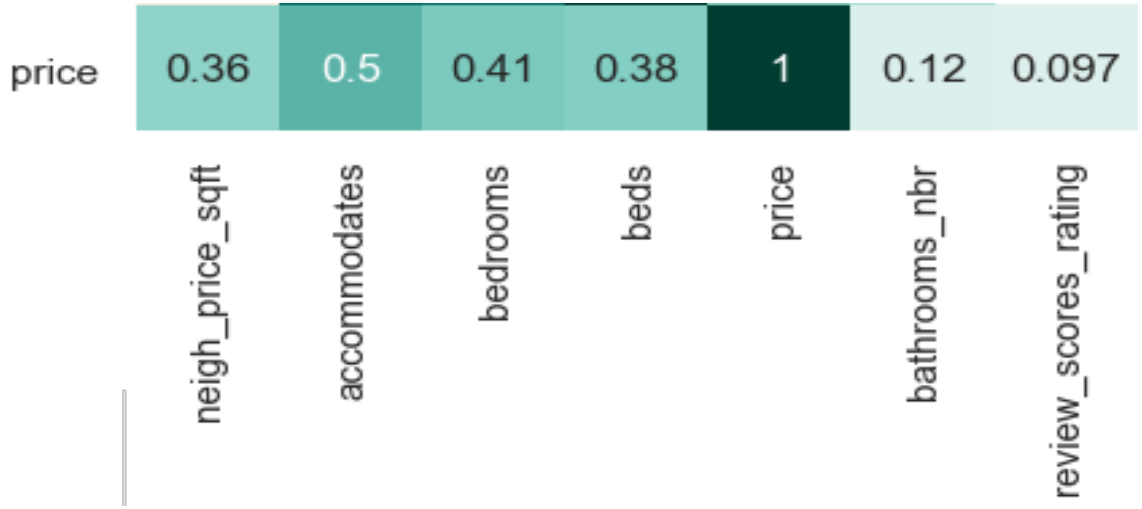
- **Distributions:**
  - skewness /log transformation



# Exploratory Data Analysis

- Correlations with the target;
  - latitude/longitude x neighborhood;

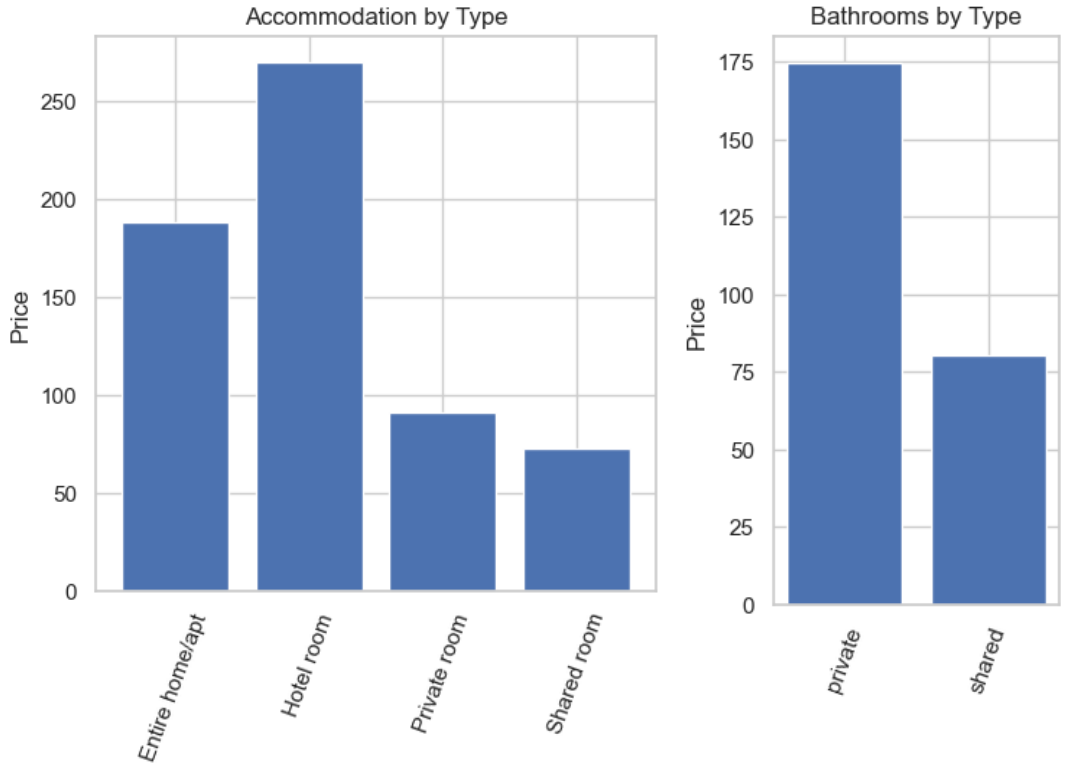



# Exploratory Data Analysis

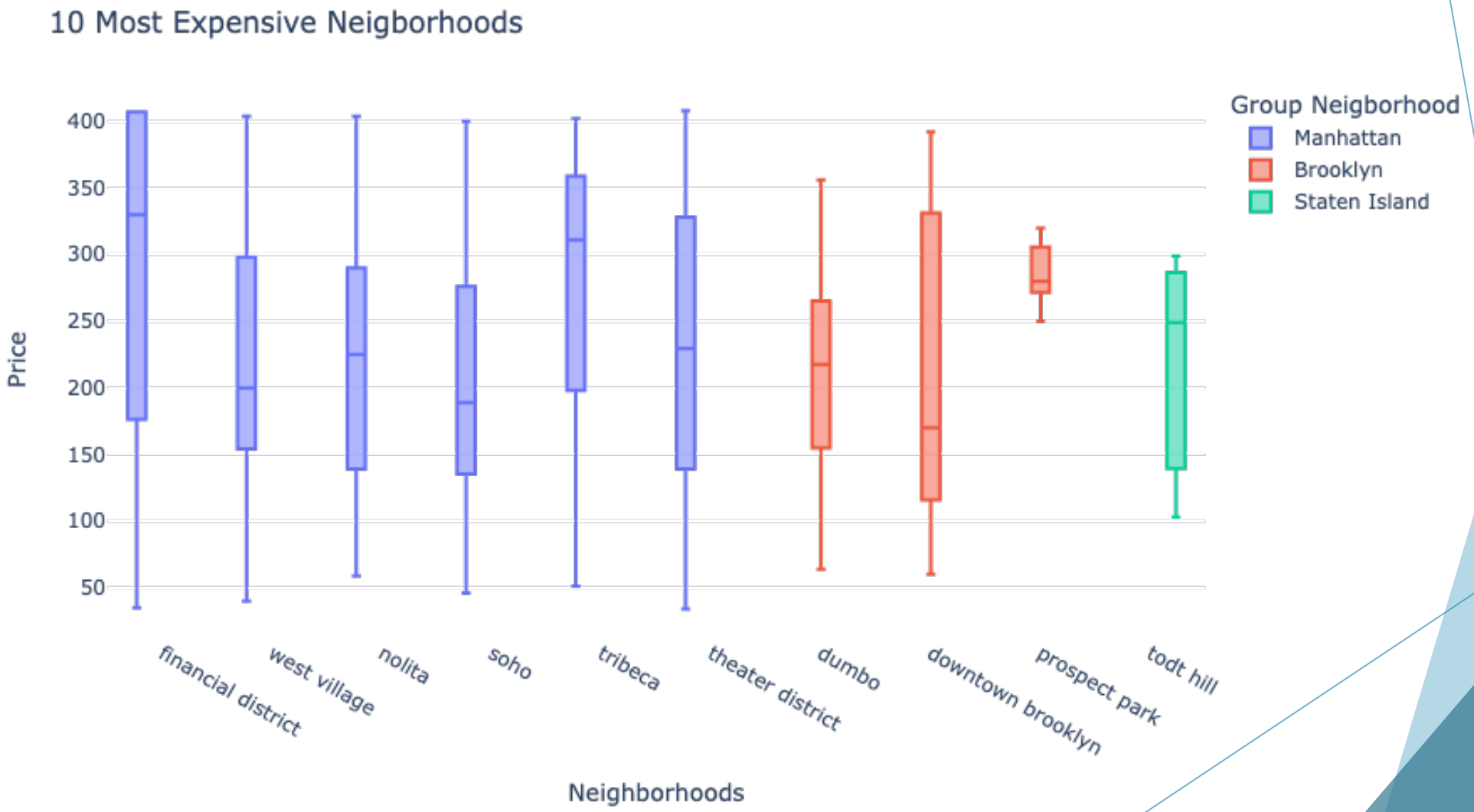
- latitude/longitude variables;


# Exploratory Data Analysis

- Boxplot for categorical variables




# Exploratory Data Analysis




# Models Evaluation

- **Tranformers:**
  - One Hot Encoded;
  - Scalling;
- **Models:**
  - Supervised: Linear Regression / KNN
  - Unsupervised: Decision Trees / RainForest / Neural Networks
- **Techniques:**
  - Regularization
  - Gridsearch

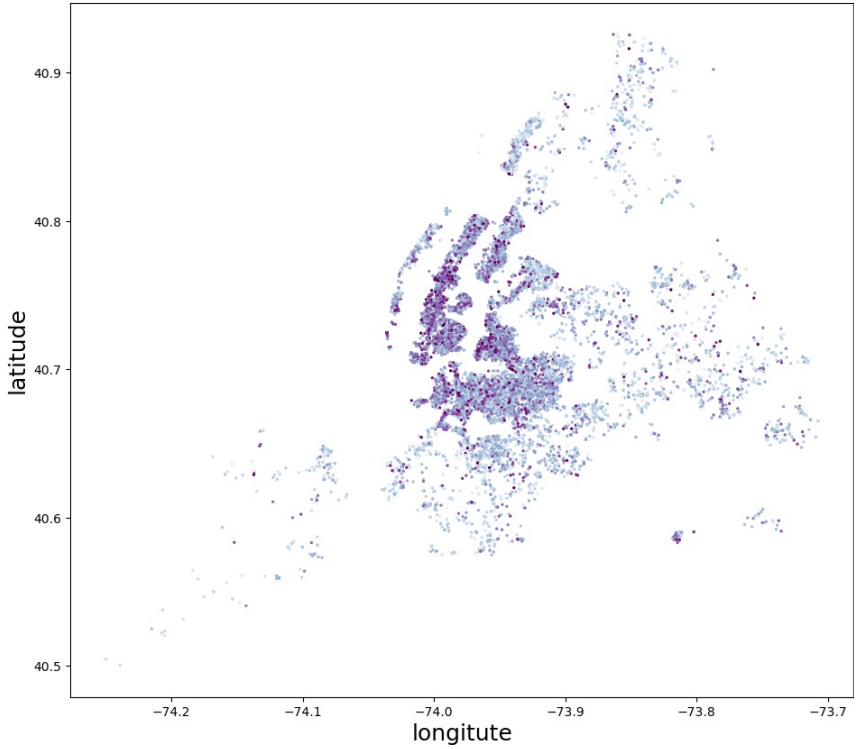
# Models Evaluation

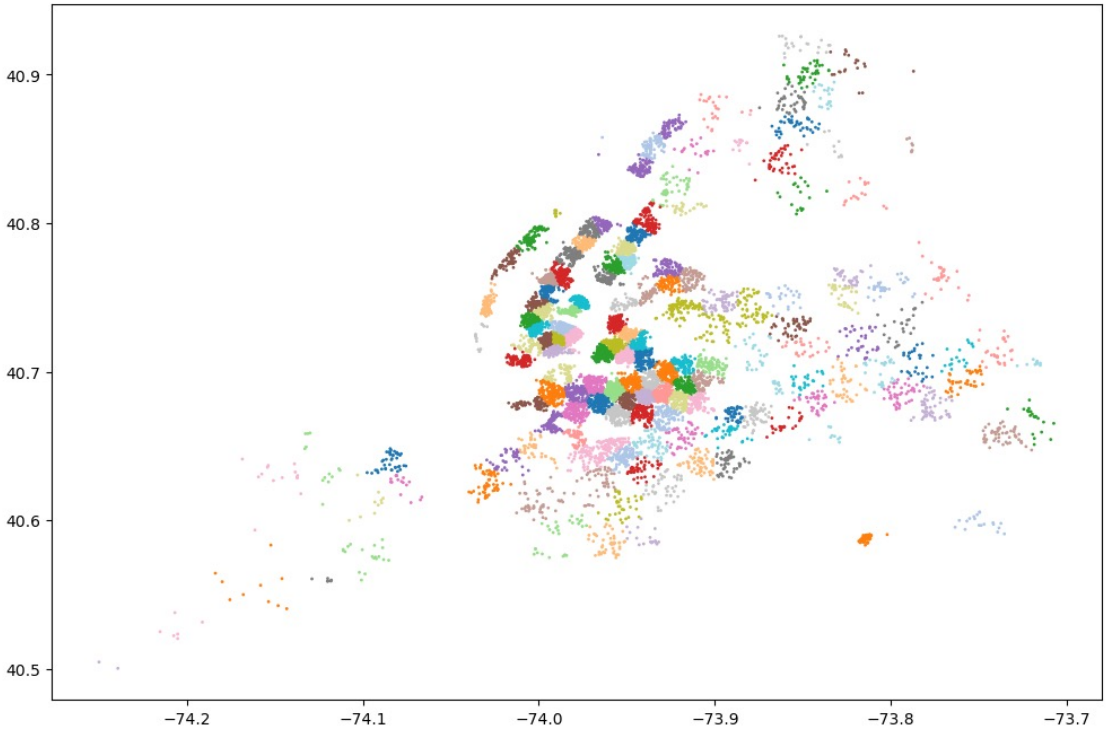
- Benchmark's Model:

| Model | Train Score | Test Score | Diff. | RMSE |
|---|---|---|---|---|
| Random Forest | 0.8866 | 0.6818 | 20.48% | 150.465 |
| K-NNeighbor | 0.7198 | 0.6619 | 5.79% | 150.487 |
| Stacked Model ElasticNet | 0.6526 | 0.5790 | 7.36% | 126.150 |

# Transfer Learning

- **Transfer Learning using KMeans**
  - k = 150, using silhouette score;



# Conclusions and Recommendations

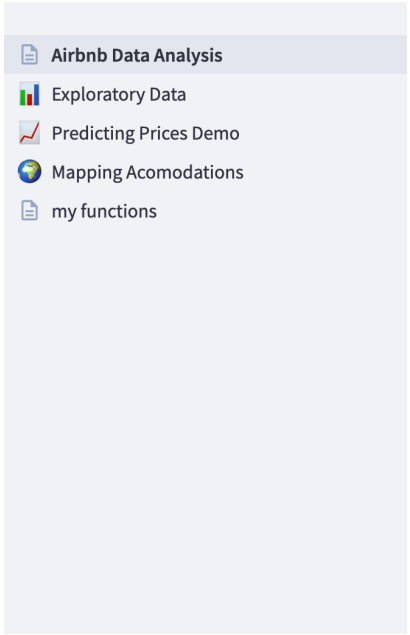
- **Conclusions:**
  - feature engineering 'amenities_count' and 'description_listing_count' ;
  - latitude/longitude or cluster with transfer learning to replace the neighborhood;
  - some variables are more important than others in determining the accommodation' prices
    - (the neighborhood feature carries more weight to the target than the number of beds or baths)

# Conclusions and Recommendations

- **Recommendations:**
  - Between 5 groups of neighborhood Manthan has so far the higher If you living in Manhattan
  - more efficient increase the capacity of accommodate people than necessarily adding a room;
- **Future works:**
  - Try to work in a different way with the ‘amenities’ variable and try to get some information from it.

# Airbnb App

- App to explorer Airbnb Listing Data and Predicting Prices:
  - Airbnb Explorer App




**_ Data Analysis on Airbnb Listings _**

Sunrise by the mountains


# Thank You!