EDA & Predicting Prices for Airbnb

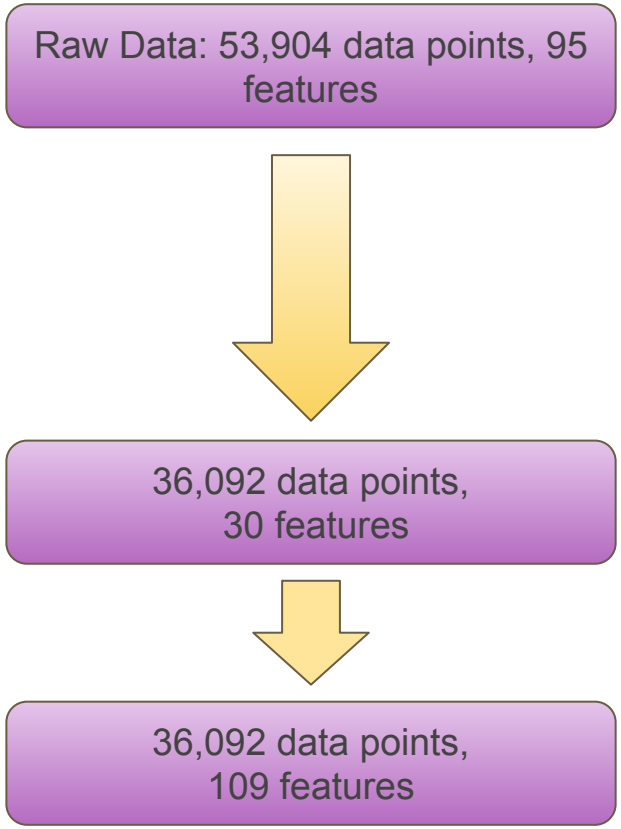
Sinem Y. Polat
31 Nov 2017

Overview

- Dataset overview & Problem Definition
- Exploratory Visualization
- The ML algorithms & Evaluation
- Conclusions

(The most used keywords in house listings names)

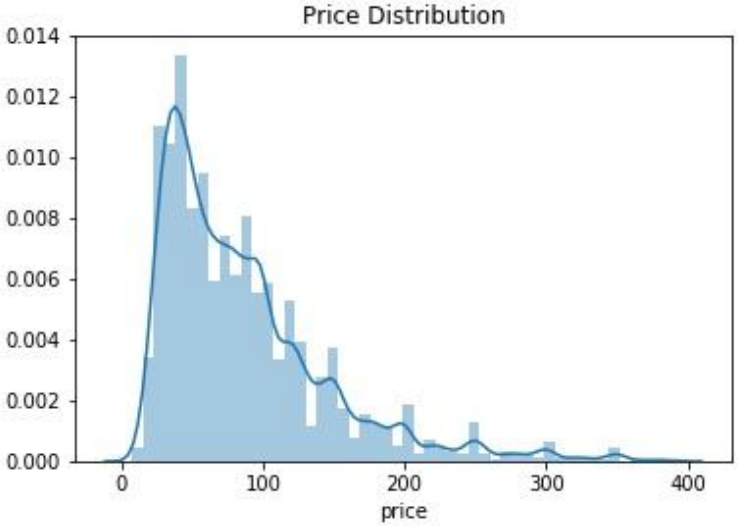
Dataset Overview & Problem Definition

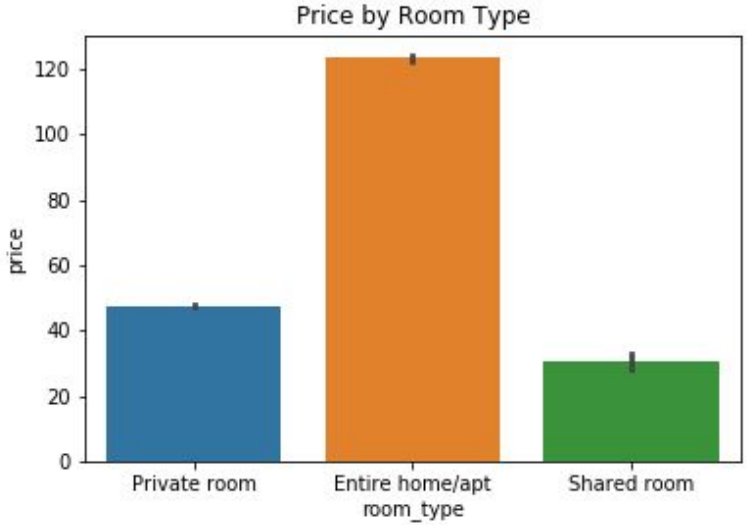


- Manually eliminating some redundant features (links, city, street, pictures etc.)
- Controlling for null or 0 values for the output column (price)
- Converting ‘f’ and ‘t’ values to 0 and 1
- Replacing $ signs for price columns
- Converting null values to 0. (security deposit, cleaning fee...)
- Eliminating rows including null. (bedrooms, rating...)
- Filter out outliers (1% of dataset)

- Converting text columns (amenities, locations) to individual columns to get the data ready for modeling

Features Affecting Price

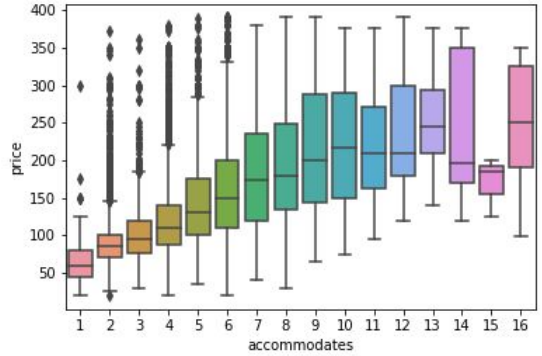


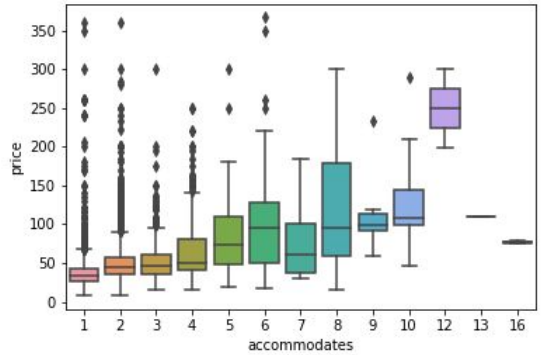


- Most house prices are between $0 and $200 per day in London.
- Being entire home is one the most important features.

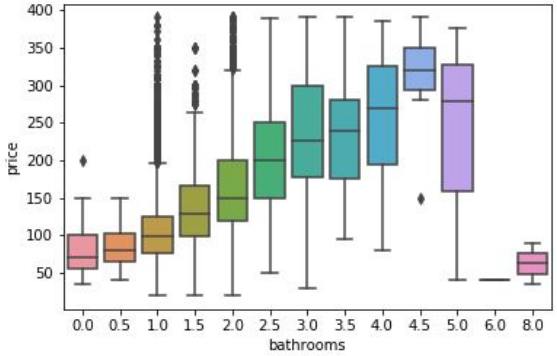
Differences between Entire Place and Private Room

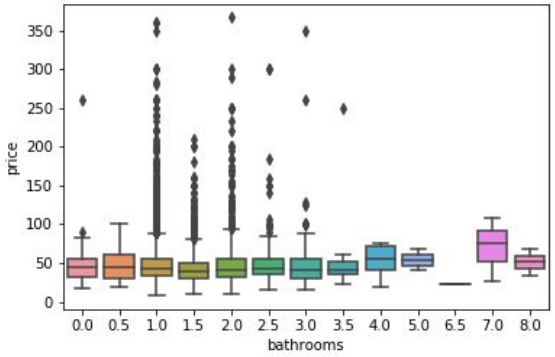
Number of people accommodating



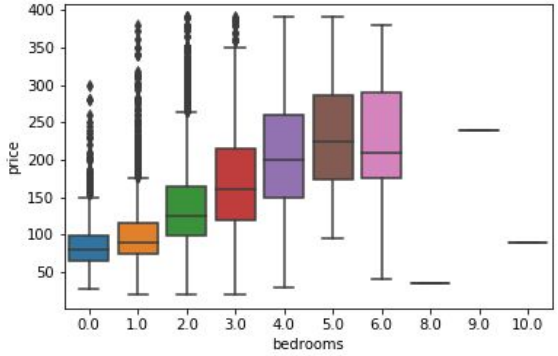


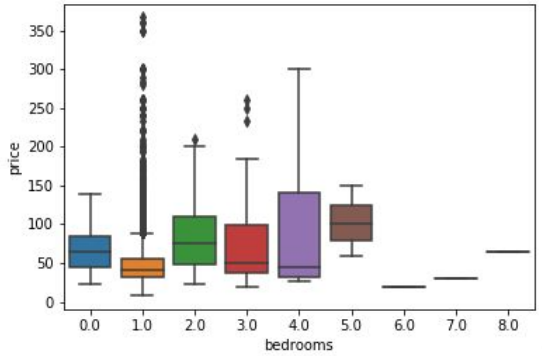
Number of bathrooms



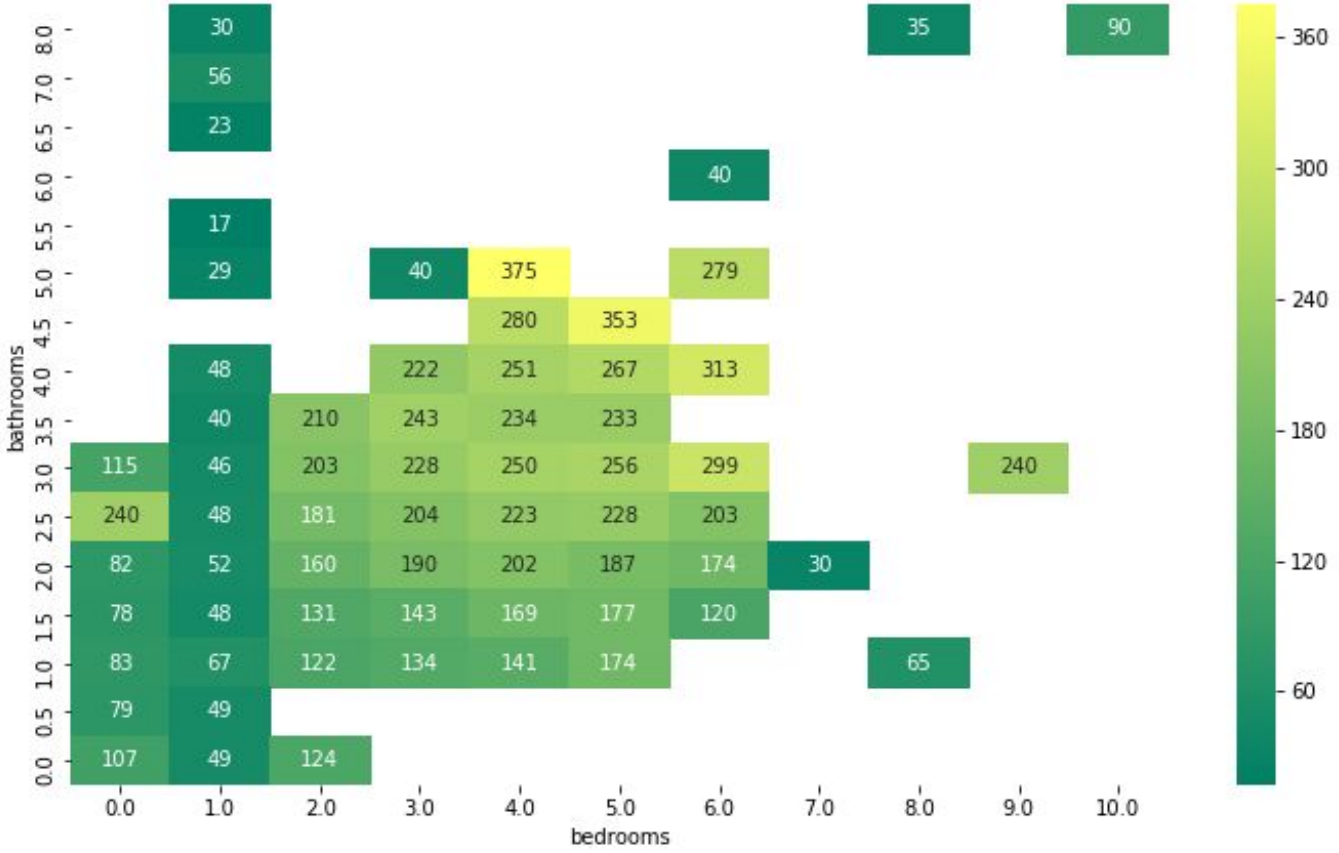


Number of bedrooms

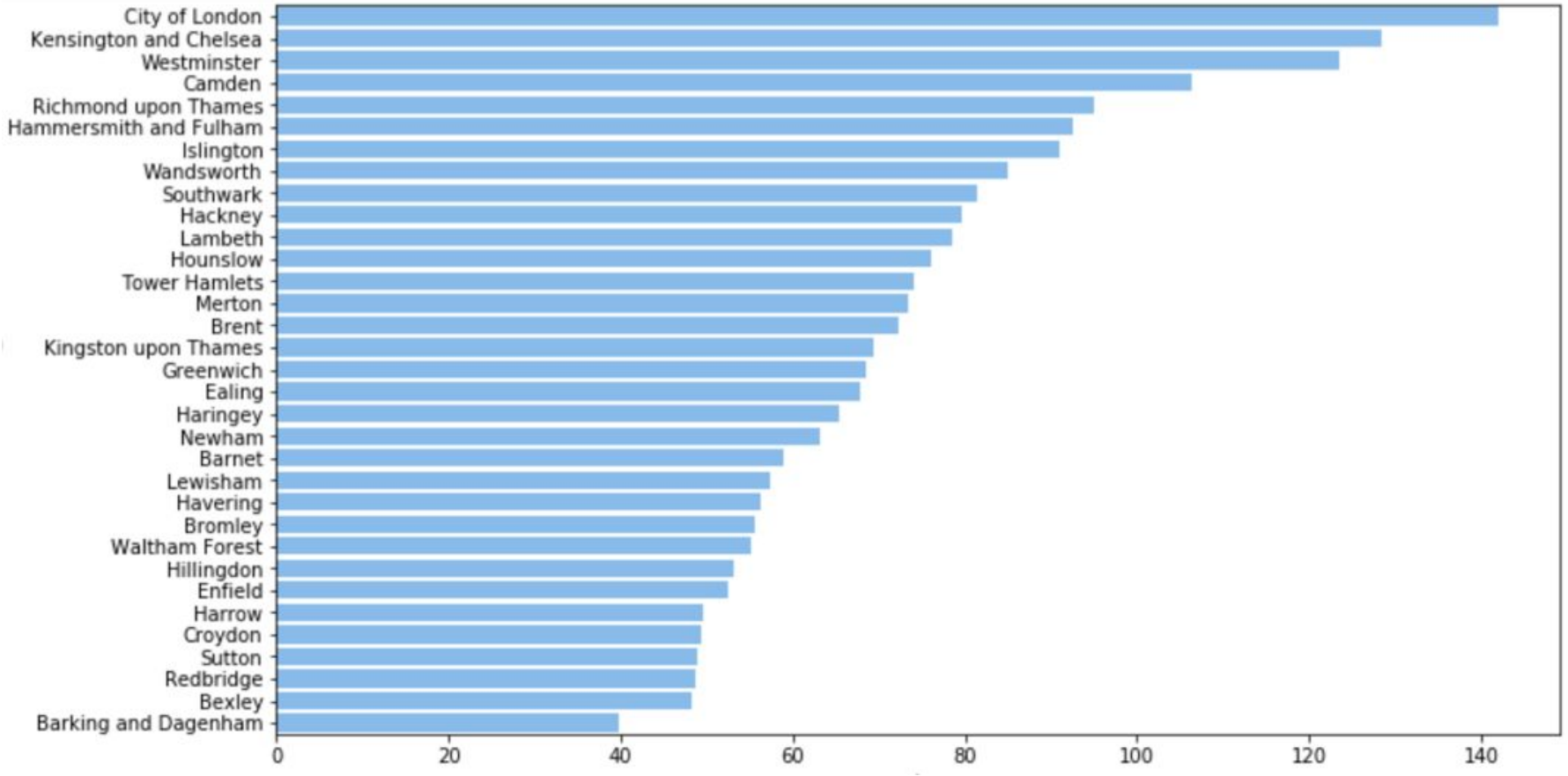




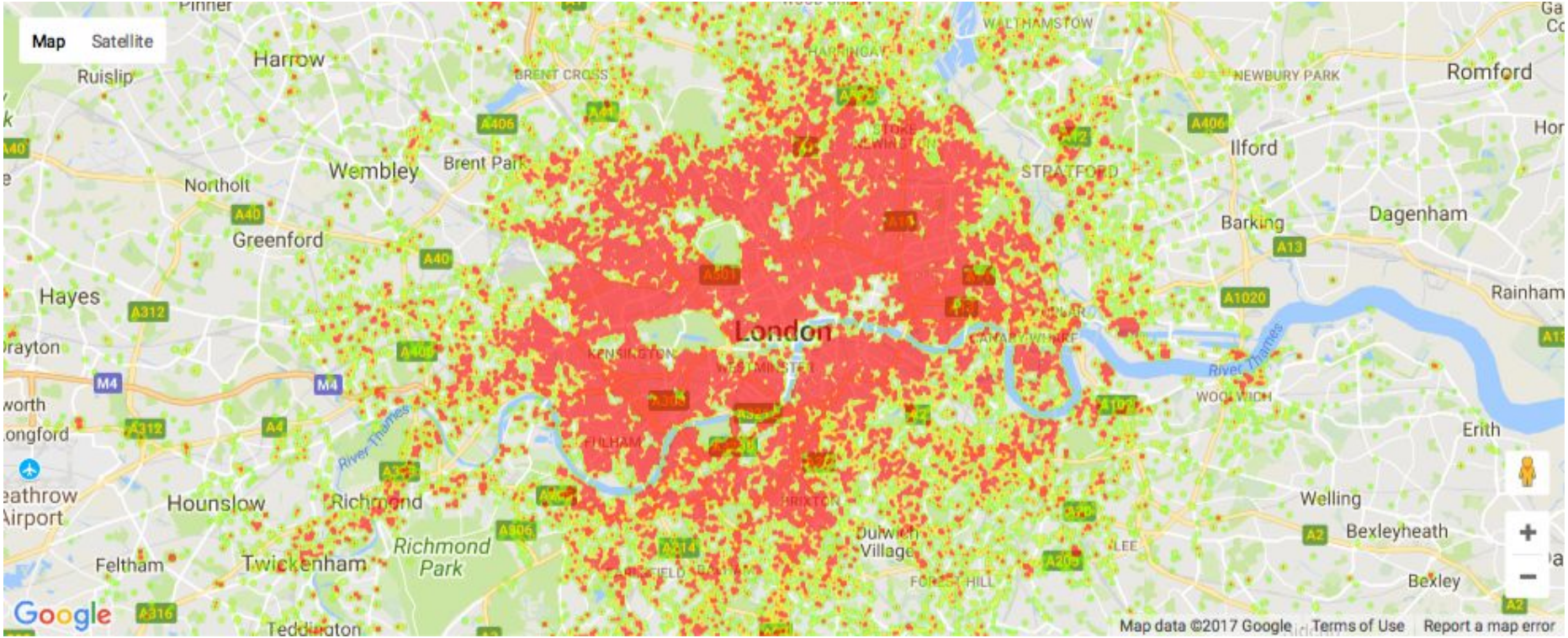
Bathroom & Bedroom Effects



Average House Prices ($ per day)

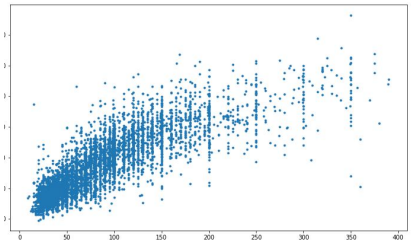
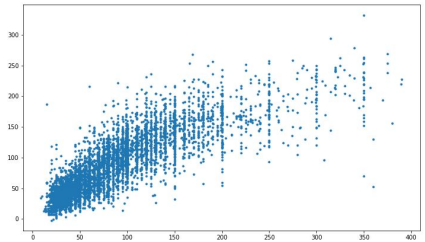
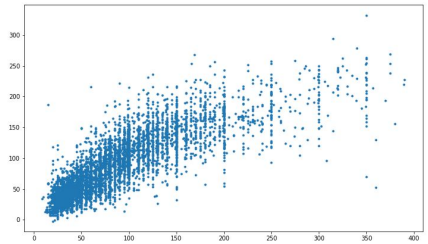


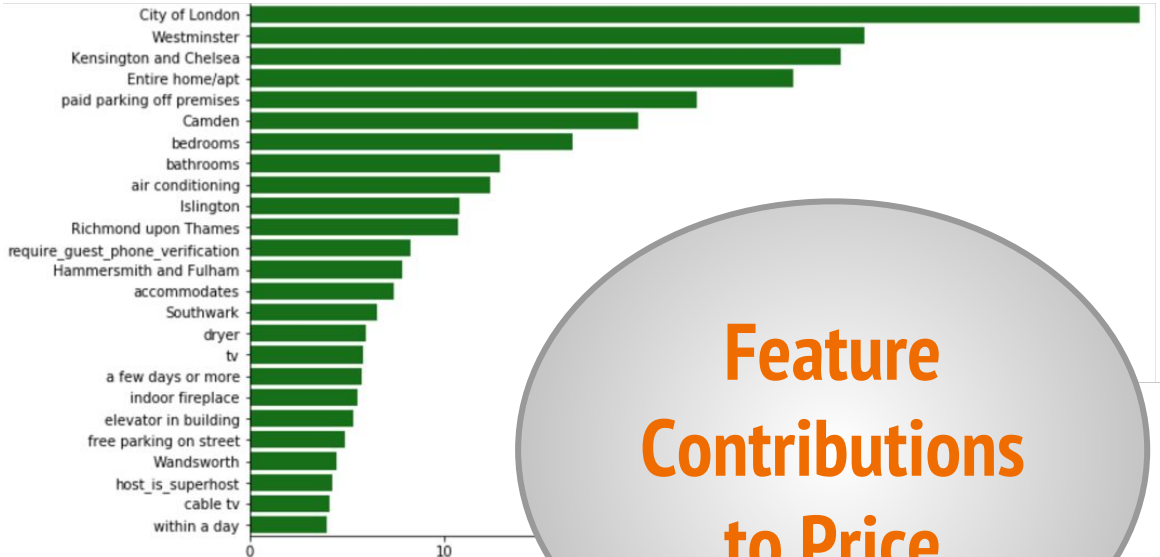
Location Effects



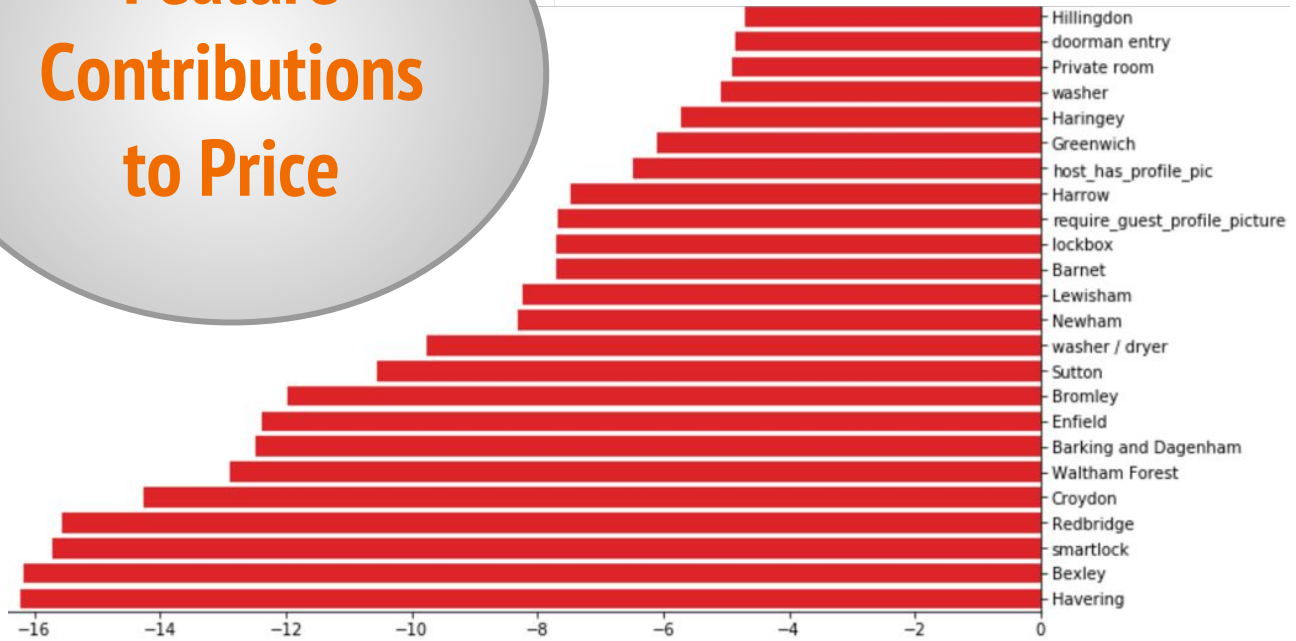
(See the dynamic map on Jupyter)

Algorithms

	Decision Tree	K-Nearest Neighbours	Linear Regression
Mean Squared Error (MSE)	1478	1775	1170
Root of MSE	38	42	34
R^2 Score	0.61	0.53	0.69
After optimizing modeling parameters (GridSearchCV)			
MSE	1387	1709	1170
Root of MSE	37	41	34
R^2 Score	0.63	0.55	0.69
Comparison Plots			



Feature Contributions to Price



Conclusions

- Regression models are used to predict house price:
  - 70% Training
  - 15% Validation
  - 15% Test
- Model can be trained to predict house price with a RMSE $34.
- Most important features are :
  - Location
  - Entire home
  - Some amenities (air conditioning, tv/cable tv, elevator, free parking on street)
  - Bathroom and bedroom numbers
  - Replying requests within a day or few days
  - Number of people accommodating