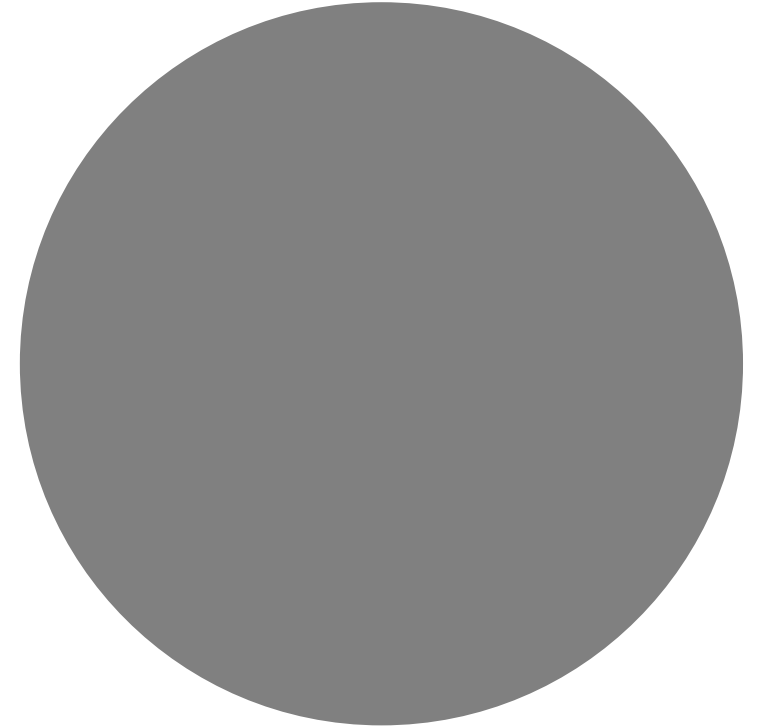


Analyzing New York City Airbnb Data Using Data Visualizations and Machine Learning Techniques

10/4/2019



Predicting Airbnb Listing Price

- Airbnb has become a popular marketplace for arranging or offering lodging services around the globe.
- Analysis of the listing data is important for business decisions.
- This study is intended to study whether listed room price on Airbnb can be predicted based on other listed information and how good the prediction is.

Data

- New York City Airbnb data on Kaggle
- New York City neighborhoods and their latitude and longitude coordinates
- Foursquare location data obtained through API calls

Data Preparation

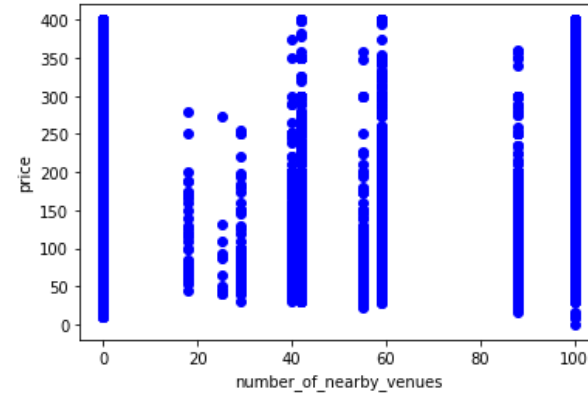
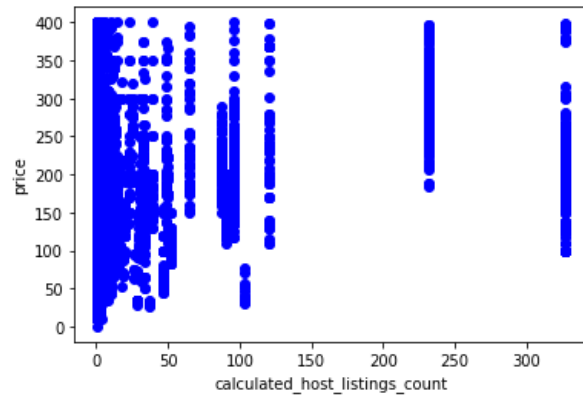
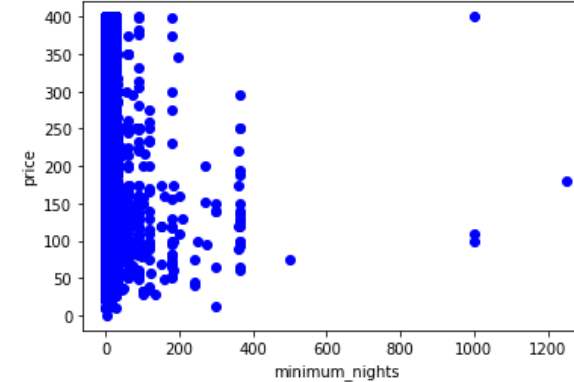
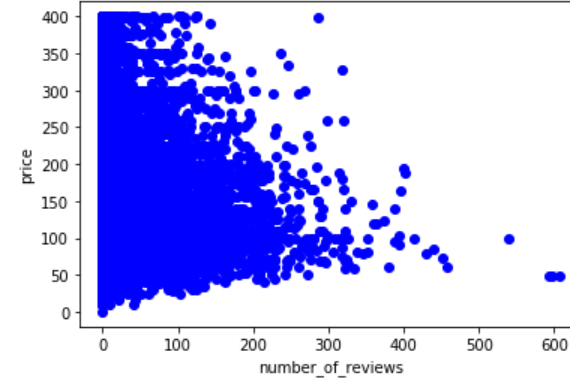
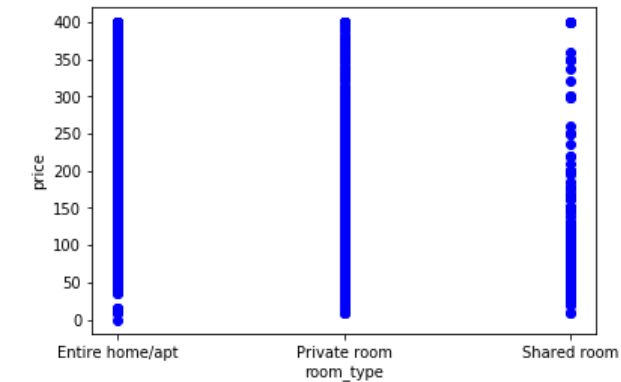
- Variables with a lot of missing values were excluded from analysis.
- Only listings in Manhattan were retained.
- Outliers were excluded.
- Airbnb listing data were merged with results from the Foursquare API calls.
- Final data frame looks like the following:

	room_type	minimum_nights	number_of_reviews	calculated_host_listings_count	number_of_nearby_venues	price
0	Entire home/apt	1	45	2	100.0	225
1	Private room	3	0	1	0.0	150
2	Entire home/apt	10	9	1	41.0	80
3	Entire home/apt	3	74	1	100.0	200
4	Private room	2	430	1	0.0	79

Multiple Linear Regression Model

- Price is the outcome variable.
- The predictor variables include the following:
 - room type
 - number of minimum nights
 - calculated host listing count
 - number of reviews
 - number of nearby venues
- The library “scikit learn” was imported to train and evaluate the multiple linear regression model.

Scatterplots between Price and its Predictors



Correlation Matrix between Price and its Predictors

	minimum_nights	number_of_reviews	calculated_host_listings_count	number_of_nearby_venues	price
minimum_nights	1.000000	-0.087959	0.141754	0.044290	0.024943
number_of_reviews	-0.087959	1.000000	-0.100310	-0.074107	-0.078694
calculated_host_listings_count	0.141754	-0.100310	1.000000	0.076138	0.188121
number_of_nearby_venues	0.044290	-0.074107	0.076138	1.000000	0.161208
price	0.024943	-0.078694	0.188121	0.161208	1.000000

Multiple Linear Regression Model Summary

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.327			
Model:	OLS	Adj. R-squared:	0.327			
Method:	Least Squares	F-statistic:	1322.			
Date:	Fri, 04 Oct 2019	Prob (F-statistic):	0.00			
Time:	15:01:10	Log-Likelihood:	-91939.			
No. Observations:	16308	AIC:	1.839e+05			
Df Residuals:	16301	BIC:	1.839e+05			
Df Model:	6					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	90.0604	1.096	82.159	0.000	87.912	92.209
x1	-0.1564	0.021	-7.494	0.000	-0.197	-0.116
x2	-0.0341	0.013	-2.718	0.007	-0.059	-0.010
x3	0.1798	0.011	15.897	0.000	0.158	0.202
x4	0.1124	0.012	9.154	0.000	0.088	0.137
x5	94.9697	1.084	87.624	0.000	92.845	97.094
x6	6.0977	1.106	5.514	0.000	3.930	8.265
x7	-11.0070	2.607	-4.222	0.000	-16.117	-5.897
=====						

Conclusions

- Although the correlations between price and the predictors are weak at best, the predictors, as a whole, explained a relatively big proportion of the variance in price.
- Price of a listing on Airbnb can be influenced by many factors other than those included in this study, such as square feet, amenities, distance to popular tourist attractions, and accessibility to public transportation. Data on these factors, when obtained, can be used as predictors in future studies.