How to price your Harlem Airbnb?

Jan-Willem Reijnen July - 1- 2020

SUGAR HILL
UPPER MANHATTAN
WEST HARLEM
HARLEM
MORNINGSIDE HEIGHTS
EAST HARLEM
Holcombe Rucker Park
Riverbank State Park
The City College of New York
Harlem YMCA
Third Avenue Bridge
Marcus Garvey Park
The Cathedral Church of St. John the Divine
General Grant National Memorial
Henry Hudson Pkwy
Broadway
Amsterdam Ave
Harlem River Dr
Harlem River
NEW JERSEY
NEW YORK
W 110th St
Park Ave
Lexington Ave
3rd Ave
2nd Ave
1st Avenue
Google

Source: Google Maps 2020

Overview

- Introduction
- Data description
- Methodology
- Results
- Discussion & Conclusion

Introduction

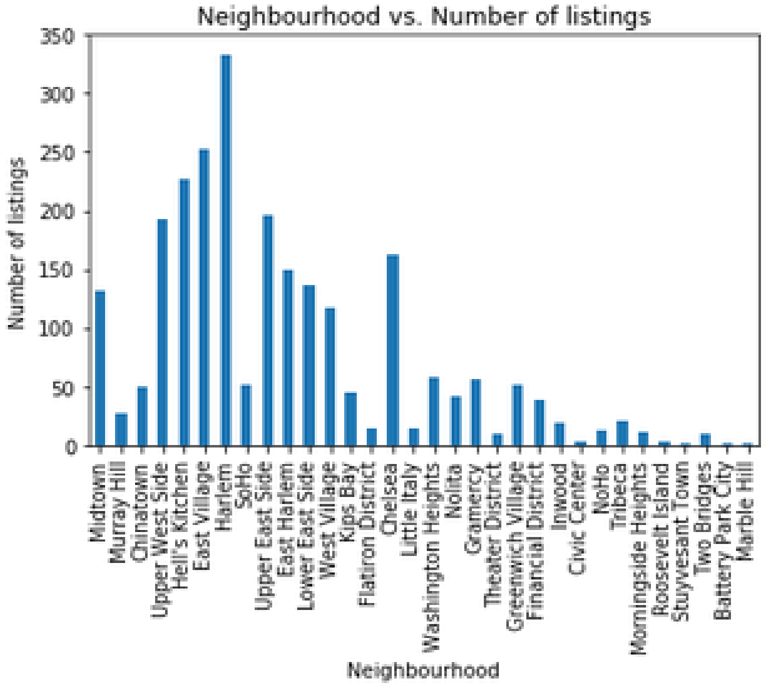
- **“Airbnb's** growth is alarming and threatening hoteliers. Having recorded more than **4 million spaces** for rent **across the world** in 65,000 cities and 191 countries, the company is waxing strong in the **United States** with approximately **600,000** listings.”
- What can be learnt from current listings for future listings

- Goal: Can the venues nearby an Airbnb listing be used for future price consulting?

Amsterdam Avenue
New York
Harlem River Drive
Ogden Avenue
Yankees-East 153rd Street
Harlem-148th Street
137th Street-City College
125th Street
Riverside Drive
Broadway
Morningside
Harlem River
River Avenue
138th Street-Grand Concourse
Harlem-125th Street
116th Street
Cathedral Parkway-110th Street
Central Park North-110th Street
103rd
5th Avenue
Madison Avenue
Park Avenue
Hudson Line MT4
3rd Avenue
2nd Avenue
1st Avenue
116th Street

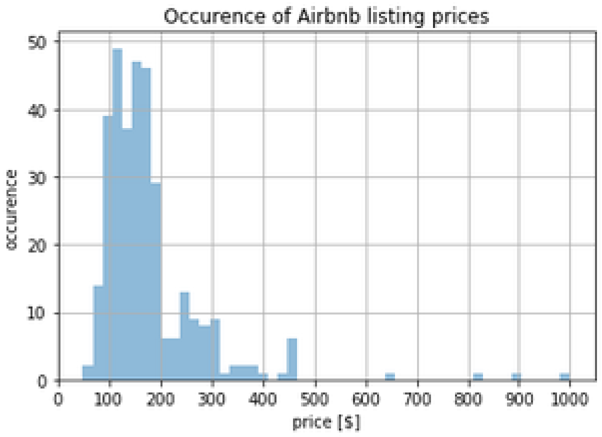
Introduction

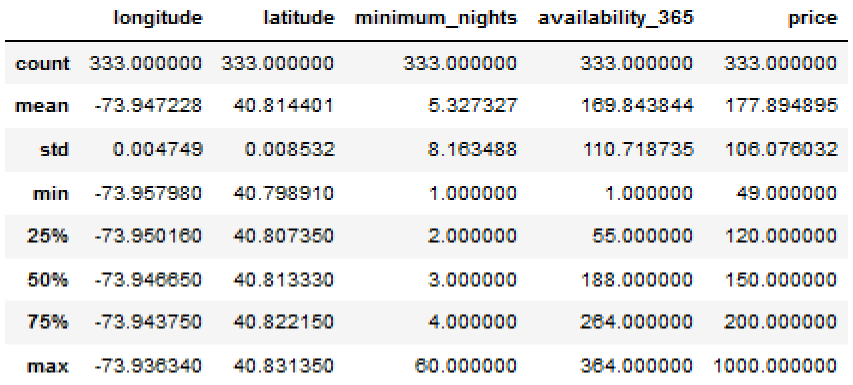
- NYC has 47900 listings in 2019
- Limited to Harlem due to Foursquare request restrictions



Data description

- Filtered set contains 263 listings in Harlem
- All listings are full apartments or houses

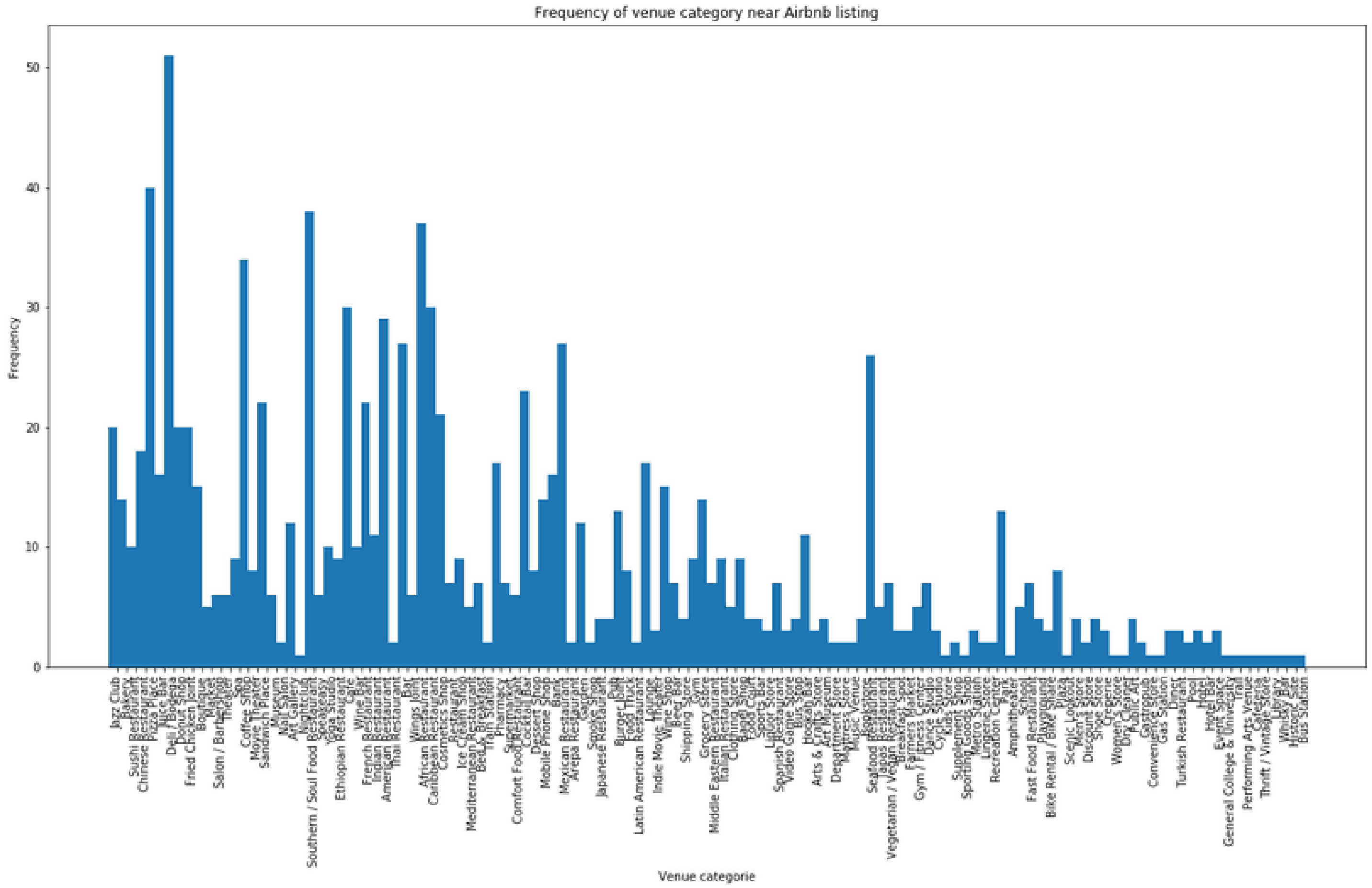




	longitude	latitude	minimum_nights	availability_365	price
count	333.000000	333.000000	333.000000	333.000000	333.000000
mean	-73.947228	40.814401	5.327327	169.843844	177.894895
std	0.004749	0.008532	8.163488	110.718735	106.076032
min	-73.957980	40.798910	1.000000	1.000000	49.000000
25%	-73.950160	40.807350	2.000000	55.000000	120.000000
50%	-73.946650	40.813330	3.000000	188.000000	150.000000
75%	-73.943750	40.822150	4.000000	264.000000	200.000000
max	-73.936340	40.831350	60.000000	364.000000	1000.000000

Data description

- Venue occurences
- 127 venue types

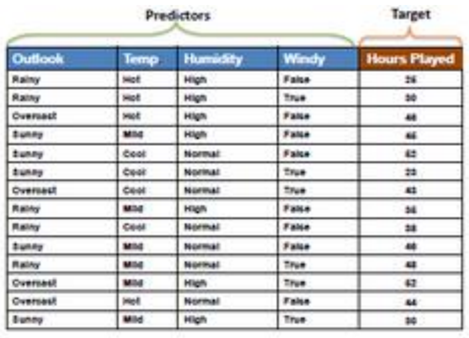


Methodology

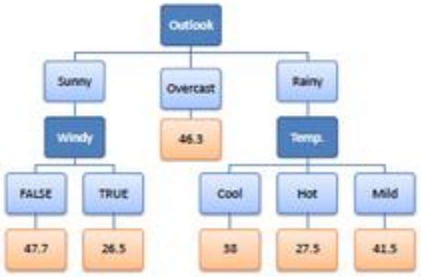
- 2 types of regression: Decision tree regression and Multiple linear regression
- Fitted on 80% of data set, tested on 20%
- 131 features to consider in regression
- Decision tree regression:

Multiple linear regression:

$y = a\,x_1 + b\,x_2 + c\,x_3 \ldots$

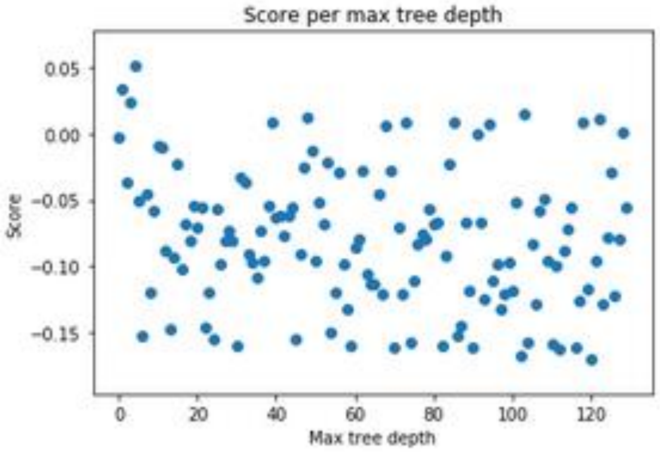


Predictors				Target
Outlook	Temp	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	46
Sunny	Mild	High	False	45
Sunny	Cool	Normal	False	52
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	35
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	46
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

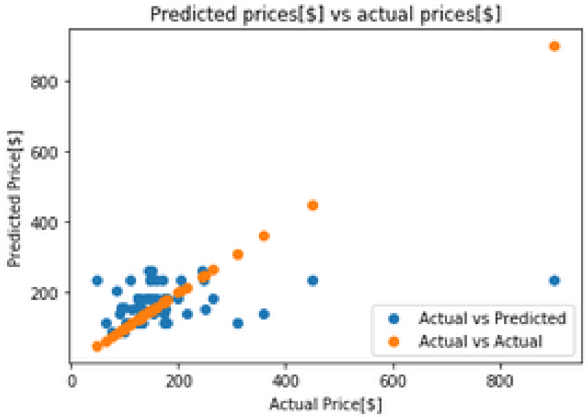


Decision tree regression

- Optimal depth for decision tree:

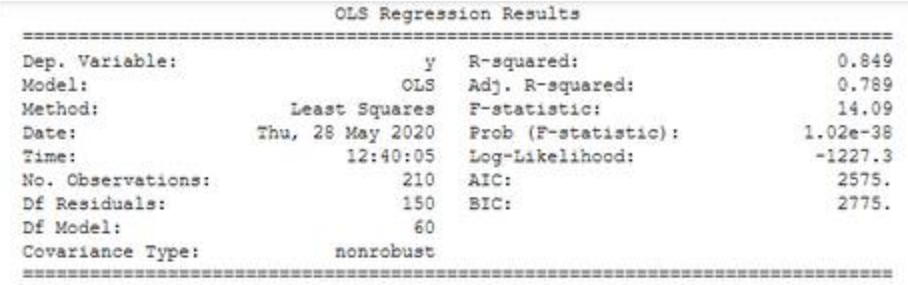


- Best results from optimal decision tree:
  - Not very close, low accuracy and high variance



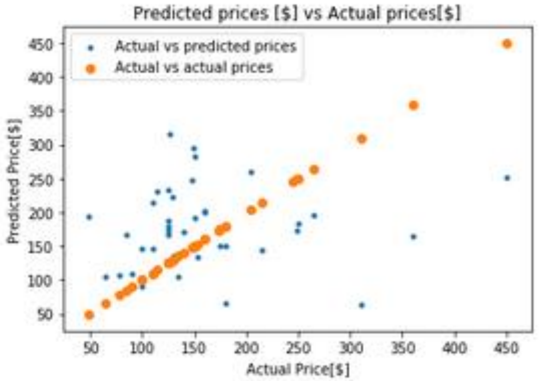
Multiple Linear Regression

- Best fitted model has R-squared of .849 (very good)



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.849
Model:                            OLS   Adj. R-squared:                  0.789
Method:                 Least Squares   F-statistic:                     14.09
Date:                Thu, 28 May 2020   Prob (F-statistic):           1.02e-38
Time:                        12:40:05   Log-Likelihood:                -1227.3
No. Observations:                 210   AIC:                             2575.
Df Residuals:                     150   BIC:                             2775.
Df Model:                          60
Covariance Type:            nonrobust
==============================================================================
```

- However, prediction accuracy still quite off:



Results

- From multiple regression the best venues to have nearby are:
  - Gastropub
  - Food truck
  - Comfort food restaurant
- The worst venues to have nearby:
  - Chinese restaurant
  - Smoke shop
  - Hookah bar

Discussion & conclusion

- Model should be tested with more data, as 131 features are to many for 263 observations

- However, Multiple regression performs way better and could definitely be used