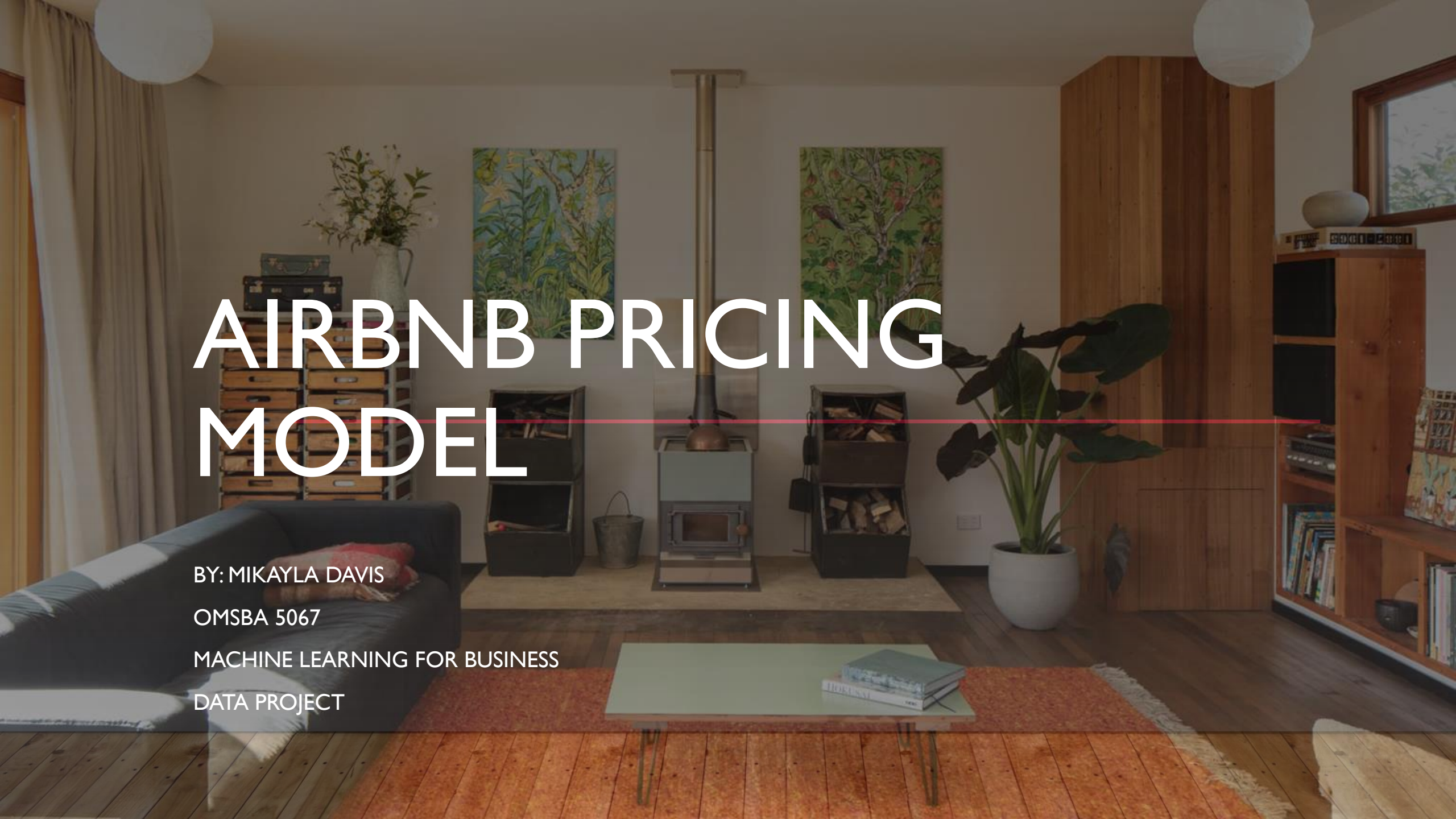


AIRBNB PRICING MODEL

BY: MIKAYLA DAVIS

OMSBA 5067

MACHINE LEARNING FOR BUSINESS

DATA PROJECT

GOAL AND FEATURES:

Goal:

- Create a predictive model that helps new users of AirBnB located in Seattle know where to price their listing.
- The data for this project was scraped from the following website:

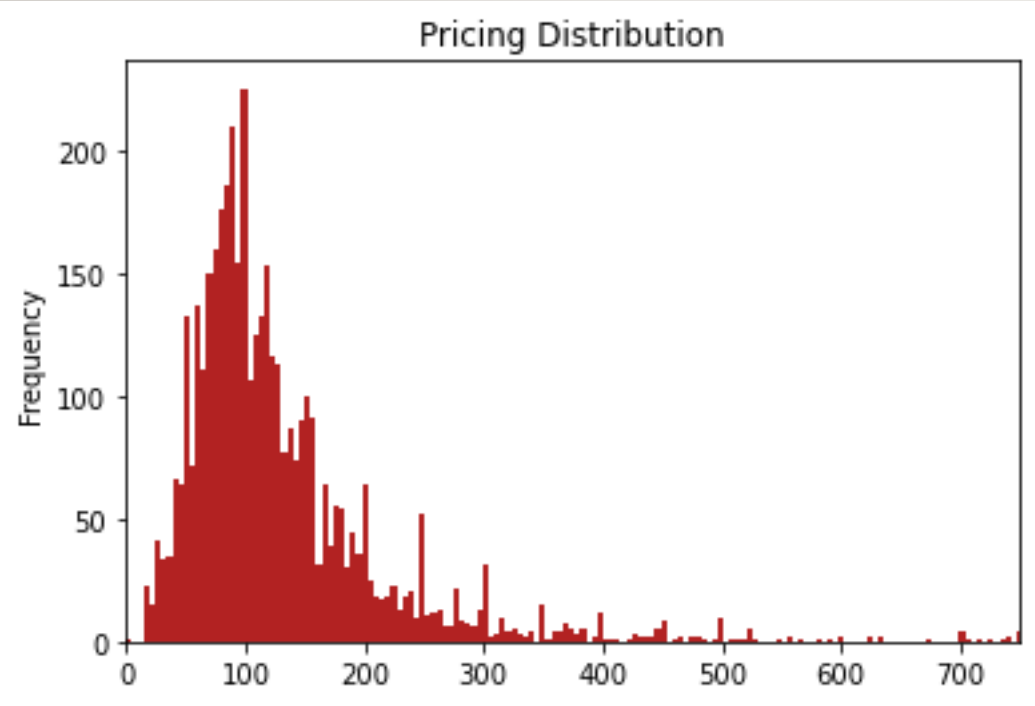
http://insideairbnb.com/get-the-data.html

- 4213 Airbnb Listings

Features:

- number_of_reviews
- host_response_rate
- neighbourhood
- availability_365
- minimum_nights
- accommodates
- room_type
- bedrooms

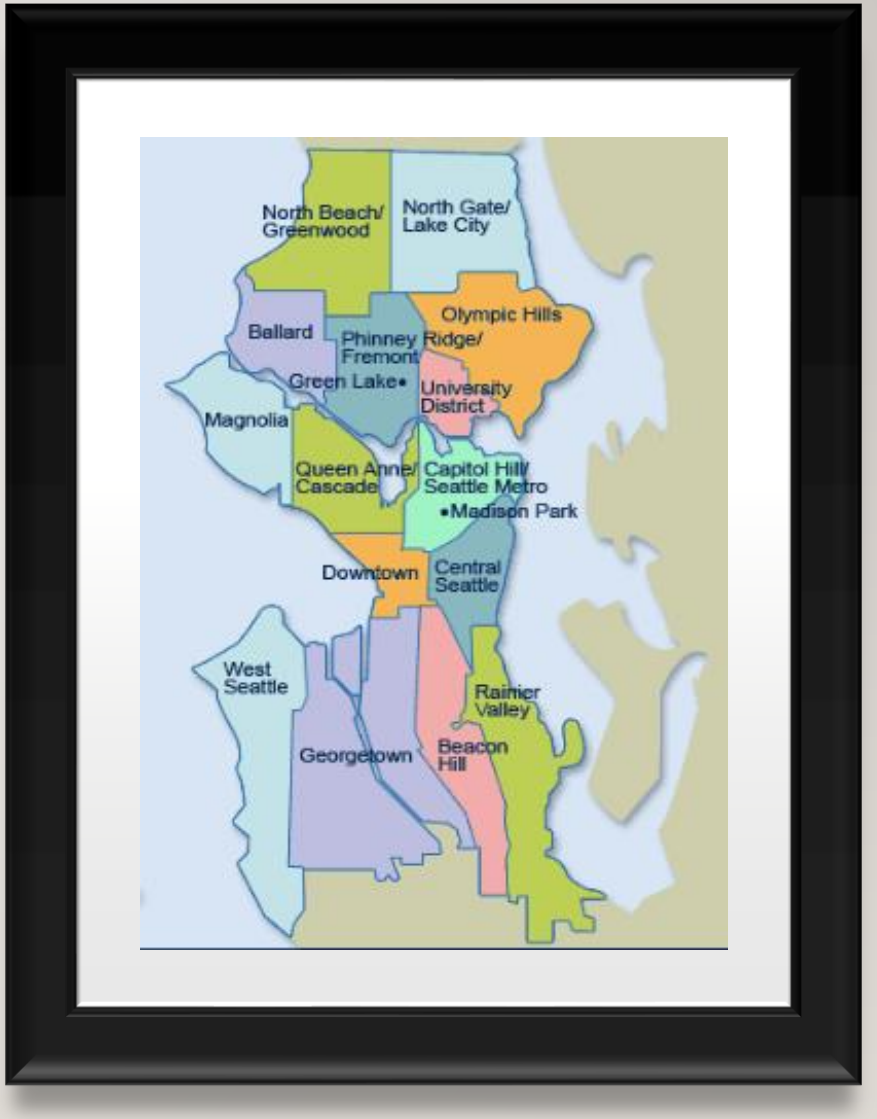
ANALYSIS OF DATA: PRICING



count	4213
mean	134.26822
std	119.99742
min	0
25%	78
50%	107
75%	154
max	3858

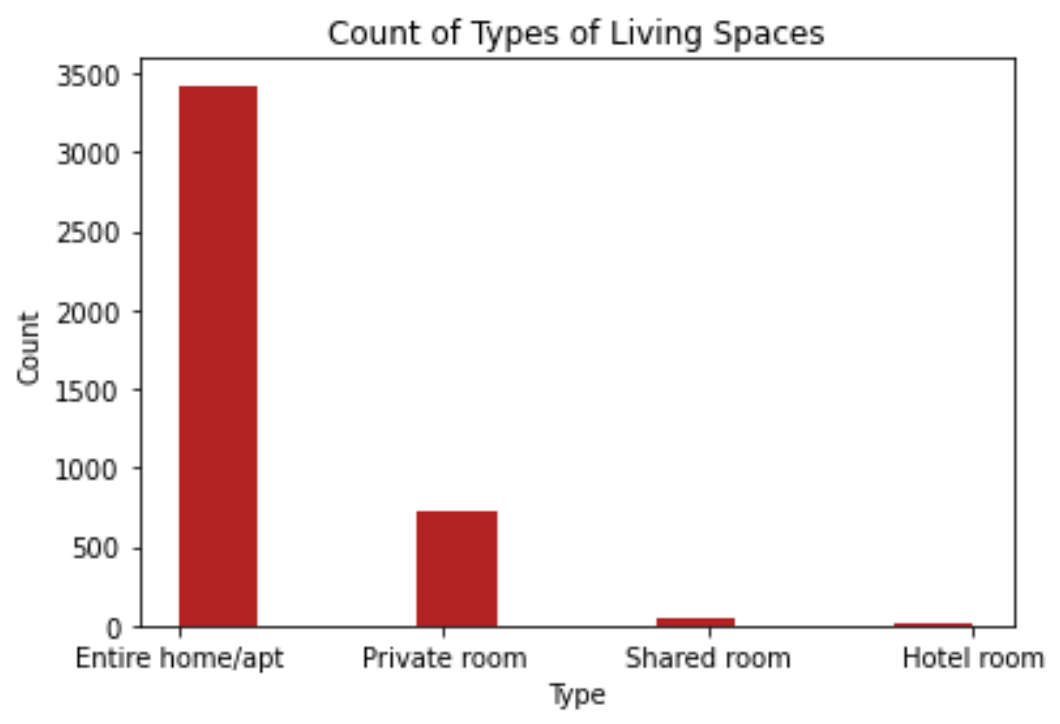
ANALYSIS OF DATA: NEIGHBORHOODS

Neighborhood	Count
Belltown	306
Broadway	230
University District	142
Wallingford	138
Fremont	129
South Lake Union	129
First Hill	118
Central Business District	117
Minor	108
North Beacon Hill	106

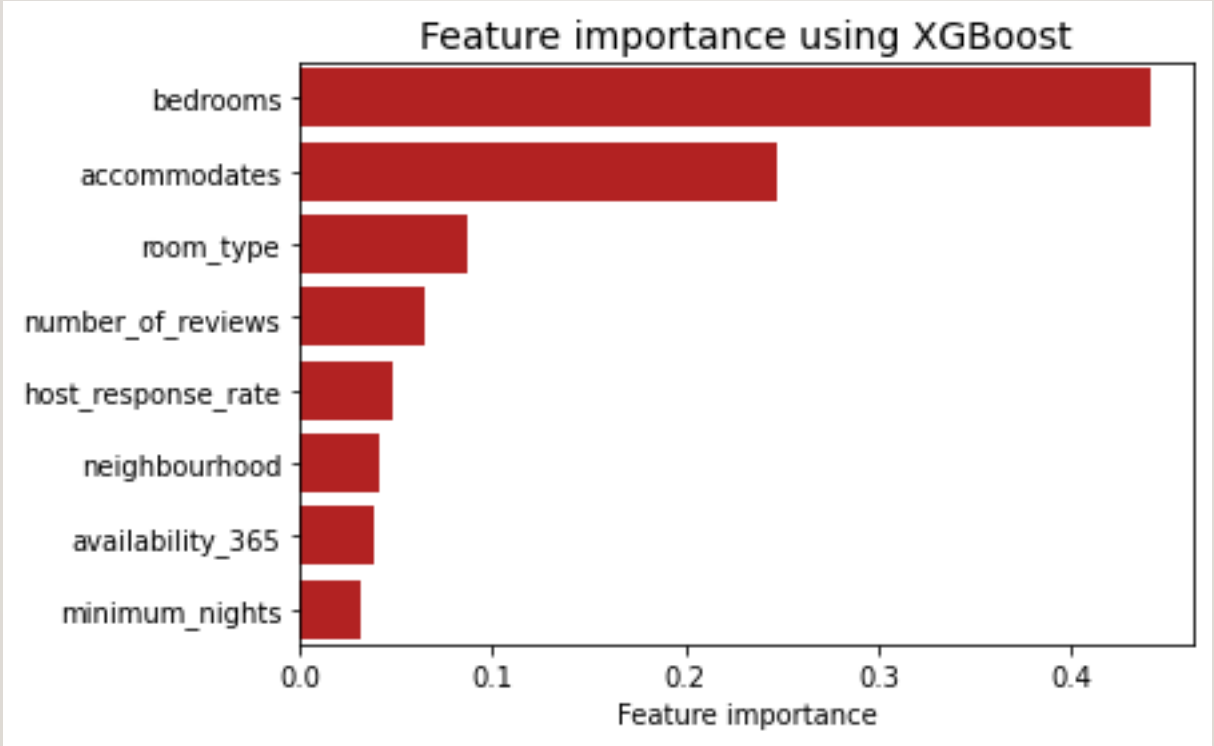


AVERAGE PRICE FOR THESE NEIGHBORHOODS

TYPES OF LIVING SPACES AVAILABLE



FEATURE IMPORTANCE USING XGBOOST



MODEL: NEURAL NETWORK

```
s1 = MLPRegressor(hidden_layer_sizes=(8,),random_state=1,max_iter=2000)
s2 = MLPRegressor(hidden_layer_sizes=(8,8),random_state=1,max_iter=2000)
s3 = MLPRegressor(hidden_layer_sizes=(8,4),random_state=1,max_iter=2000)
s4 = MLPRegressor(hidden_layer_sizes=(4,),random_state=1,max_iter=2000)
s5 = MLPRegressor(hidden_layer_sizes=(16,16),random_state=1,max_iter=2000)
```

CROSS VALIDATION TO DETERMINE THE BEST SETTING – LOOKING AT R2

S1	0.3539317
S2	0.3456551
S3	0.3525932
S4	0.3571191
S5	0.3466191

MODEL: LINEAR REGRESSION

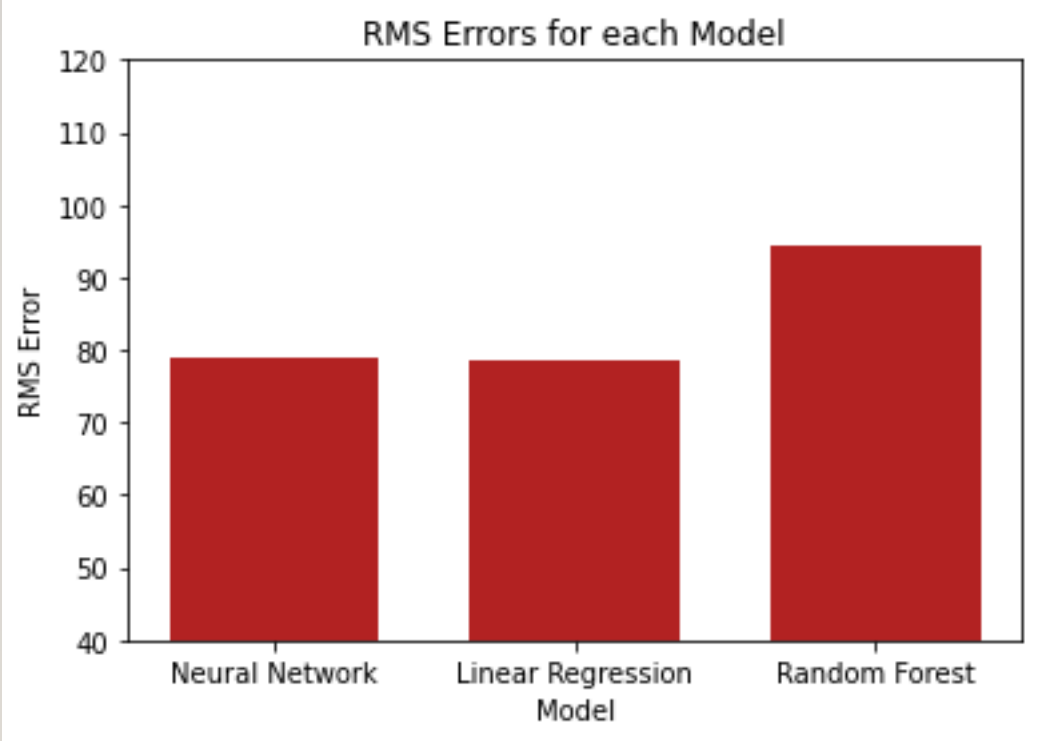
- Good ol' Classic regr

```
from sklearn import linear_model

regr = linear_model.LinearRegression()

regr = regr.fit(X_train,y_train)
#print("score = ", regr.score(X_test,y_test))

YestimateL = regr.predict(X_test)
err1L = (sum((YestimateL-y_test)**2)/len(y_test))**0.5
#print("RMS error = ",err1L)
```

MODEL: RANDOM FOREST CLASSIFIER

```
from sklearn.ensemble import RandomForestClassifier

clf = RandomForestClassifier(n_estimators=100, random_state=50)
clf = clf.fit(X_train, y_train)
```

BEST MODEL: LINEAR REGRESSION



RMS error NN = 78.89408565600093
RMS error Linear = 78.62348162461026
RMS error RF = 94.33136753757843

CONCLUSION

- The model to use for where to price Airbnb listings in Seattle is a linear regression model.
- The feature with the highest importance is bedrooms, the neighborhood had much less of an importance
- Future suggestions: Additional features – the Amenities were listed in the data as a list of strings. It would be good to look at some of these features to determine the importance and then keep those that have the most influence. (e.g., Lake Access, Hot Tub, BBQ Grill). Another feature that was not included in the raw data was a cleaning fee, which would influence the total price and impact interest in the listing, this would need to be explored further.