Defence of the Master Thesis

Airbnb Copenhagen

An AI & Machine Learning approach to analyzing and modeling Copenhagen Airbnb prices

Agenda

1.The main insights from the thesis

2.Critical reflection on the feature selection

3. improvements master thesis’s sections

4.Neural Network Multi Layer Perceptron

5.Model comparison Test and Training set

6.Results

7..Ideas for further research

1.Main insights from the thesis

1. A field of machine learning and deep learning methods and techniques that provide an identifiable pattern and prediction price for Airbnb listing price

2.How well a panel of six machine learning methods can predict the Airbnb price in Copenhagen and establish a basis for a sustainable model approach for Copenhagen's Airbnb environment.

3. .Despite the amount of data in the Airbnb dataset, the models show insightful results of the features' impact on price, which can provide insight into which patterns and mechanisms are most associated.

2.Critical reflection on the cleaning process and feature selection

1.Data Cleaning and Exploration :

In short, the original dataset contained 28077 Airbnb listings and 106 features but **I dropped a bunch**. For example, some of those are free text variables, like the host description of the property and all the written reviews. To perform feature selection, it was very important to find an approach and process that demonstrated the relationship between price and data set functions. Heatmap shows an adaptive visualization for deeper insight into the correlation intensity of the variables.

2.Natural Language Processing(NLP) :

Natural Language Processing was not been used in the creation of this model. Therefore, text variables was and other variables which are not useful for predicting price (e.g. url, host name and other host-related features that are unrelated to the property).

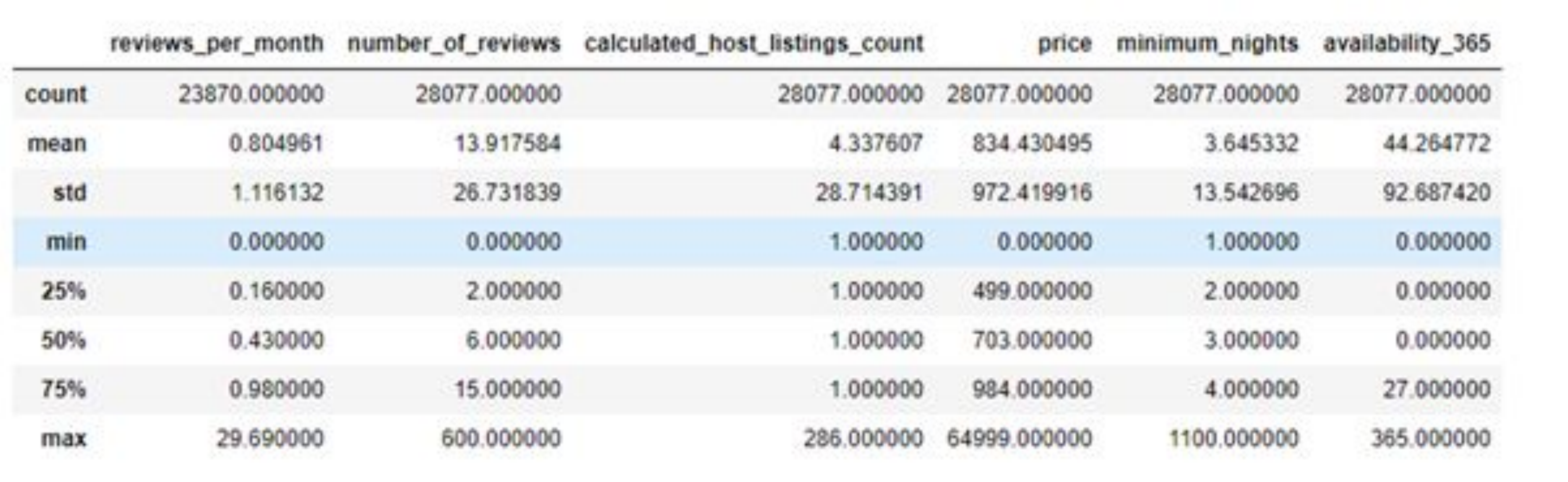
3.Descriptive Statistics : When we have a set of observations, it is useful to summarize features of our data into a single statement called a descriptive statistic. As their name suggests, descriptive statistics describe a particular quality of the data they summarize. These statistics fall into two general categories: the measures of central tendency and the measures of spread.

Listing.csv

28077 Airbnb Listings.

After Cleaning process : I.Training data sets with 15238 Airbnb listings and a testset with 3836 Airbnb listings

19074 Airbnb Listings



	reviews_per_month	number_of_reviews	calculated_host_listings_count	price	minimum_nights	availability_365
count	23870.000000	28077.000000	28077.000000	28077.000000	28077.000000	28077.000000
mean	0.804961	13.917584	4.337607	834.430495	3.645332	44.264772
std	1.116132	26.731839	28.714391	972.419916	13.542696	92.687420
min	0.000000	0.000000	1.000000	0.000000	1.000000	0.000000
25%	0.160000	2.000000	1.000000	499.000000	2.000000	0.000000
50%	0.430000	6.000000	1.000000	703.000000	3.000000	0.000000
75%	0.980000	15.000000	1.000000	984.000000	4.000000	27.000000
max	29.690000	600.000000	286.000000	64999.000000	1100.000000	365.000000

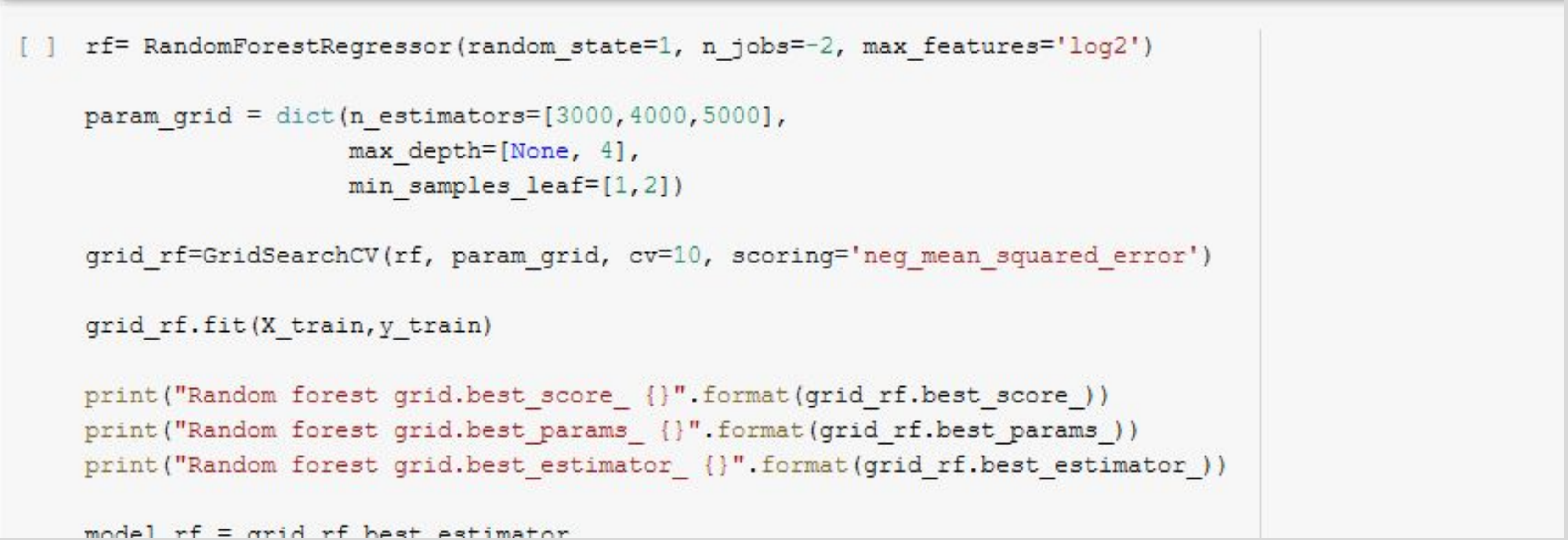
3. Improvements in the Master’s thesis sections ?

6.2 section :Exploratory Data Analysis of Airbnb Listings in Copenhagen

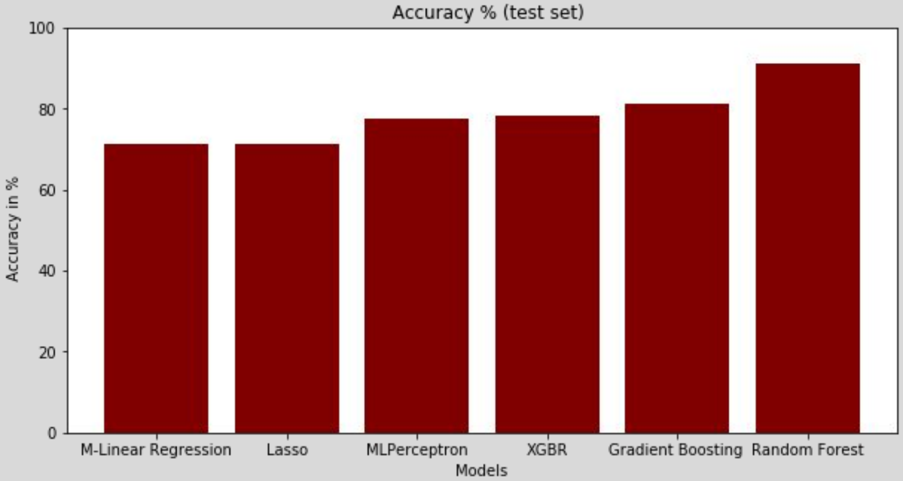
8.5 section : Feature Engineering

10.2.1 section : Multiple Linear Regression

10.2.2 section : Lasso



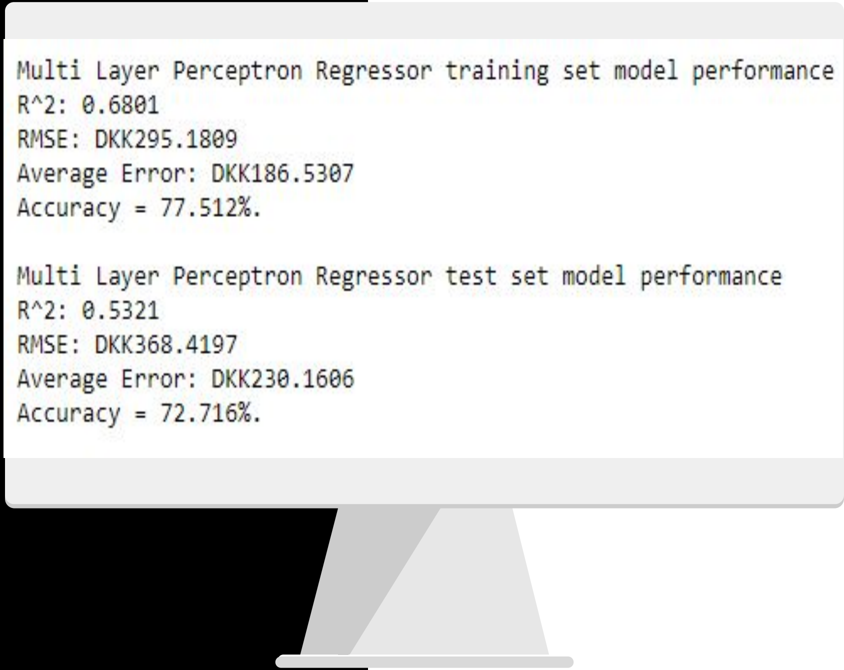
```
[ ]  rf= RandomForestRegressor(random_state=1, n_jobs=-2, max_features='log2')

     param_grid = dict(n_estimators=[3000,4000,5000],
                       max_depth=[None, 4],
                       min_samples_leaf=[1,2])

     grid_rf=GridSearchCV(rf, param_grid, cv=10, scoring='neg_mean_squared_error')

     grid_rf.fit(X_train,y_train)

     print("Random forest grid.best_score_ {}".format(grid_rf.best_score_))
     print("Random forest grid.best_params_ {}".format(grid_rf.best_params_))
     print("Random forest grid.best_estimator_ {}".format(grid_rf.best_estimator_))

     model_rf = grid_rf.best_estimator
```



4.Neural Network Multi Layer Perceptron

Multi layer perceptron Model's true scores and metrics

Wrong number results have been written. This applies to r2 score, accuracy and RMSE score.

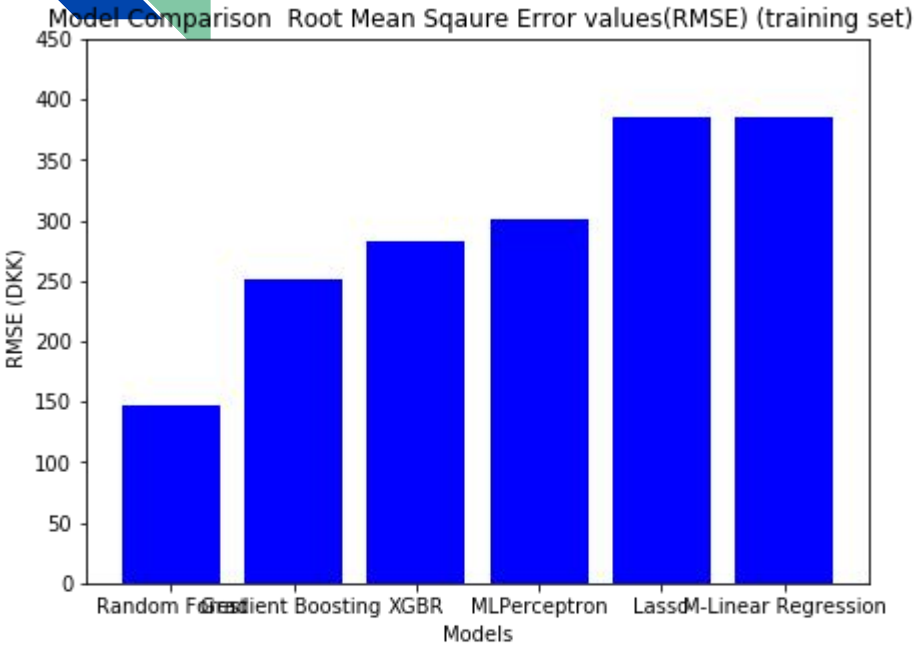


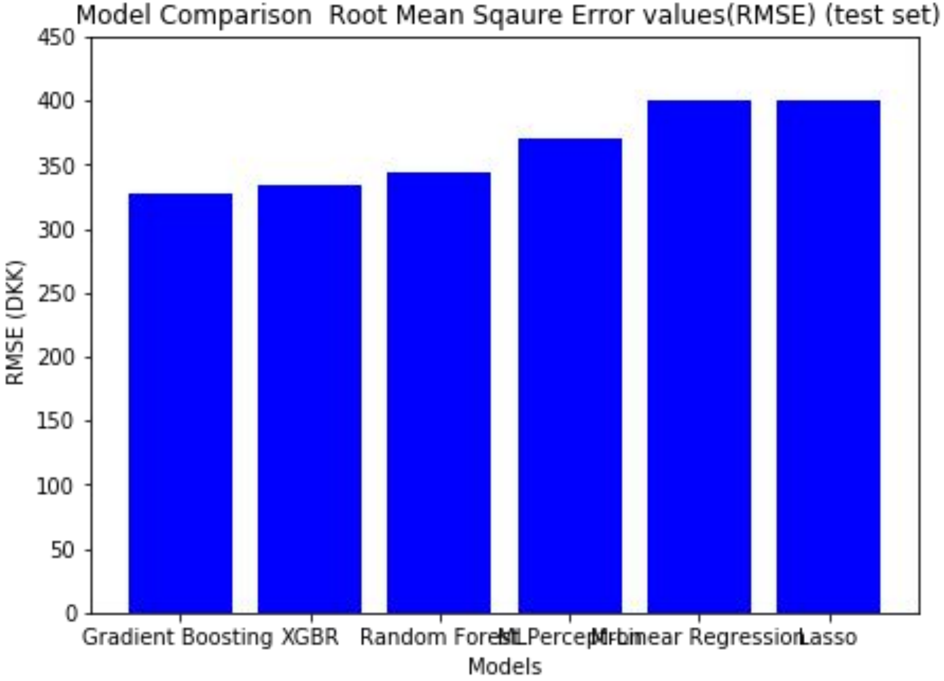
```
Multi Layer Perceptron Regressor training set model performance
R^2: 0.6801
RMSE: DKK295.1809
Average Error: DKK186.5307
Accuracy = 77.512%.

Multi Layer Perceptron Regressor test set model performance
R^2: 0.5321
RMSE: DKK368.4197
Average Error: DKK230.1606
Accuracy = 72.716%.
```

Old figures from one of the first results from the MLP model have been accidentally written down in the section.

5. Model comparison Test and Training set

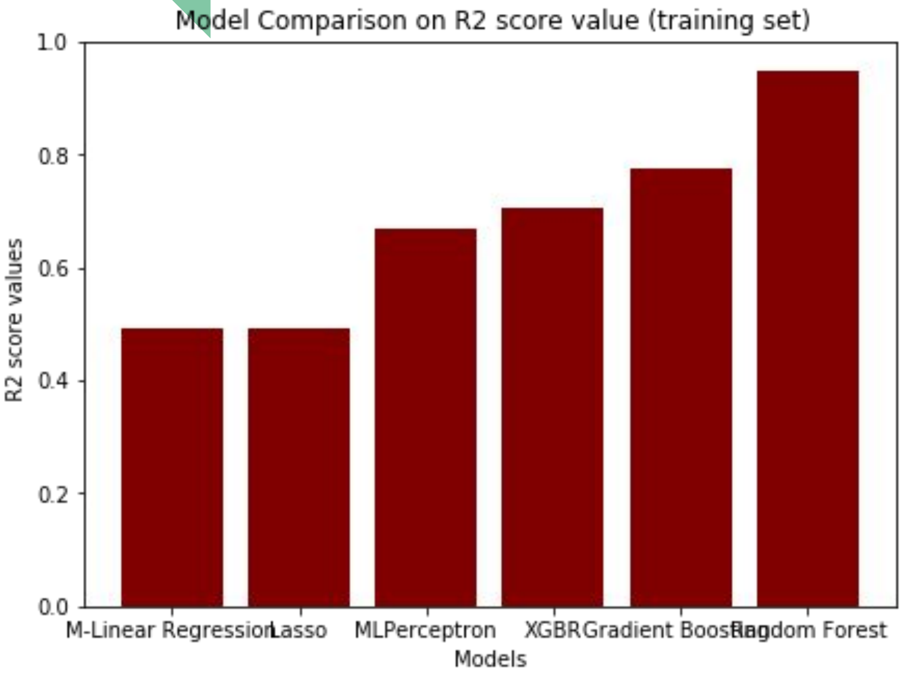
RMSE Value

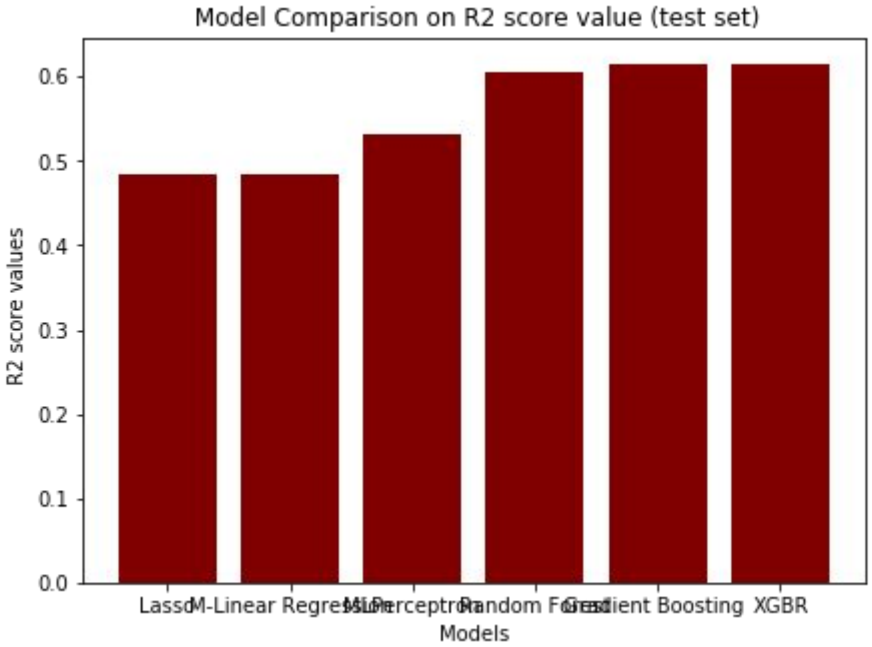




5. Model comparison Test and Training set

R2 Value



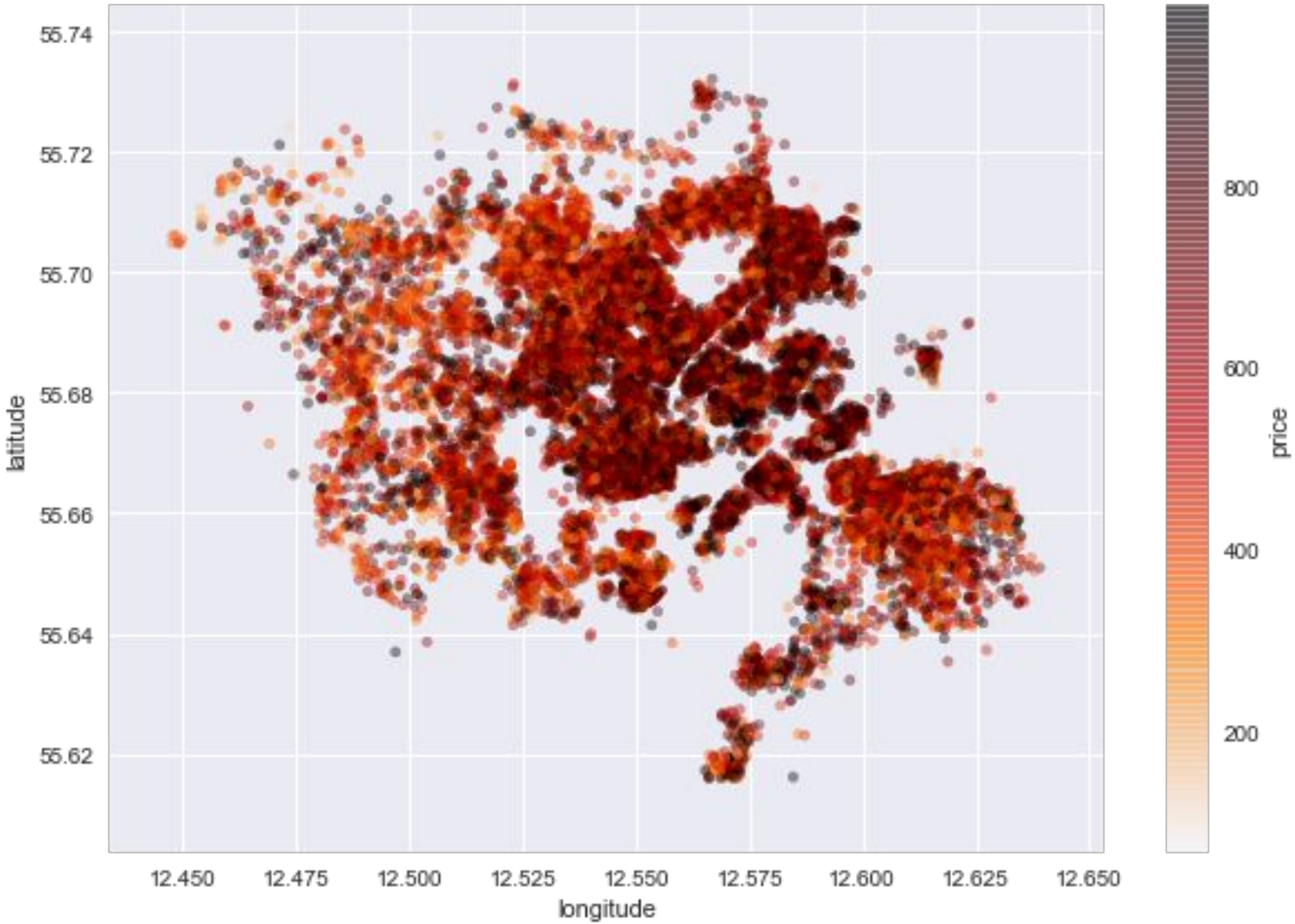


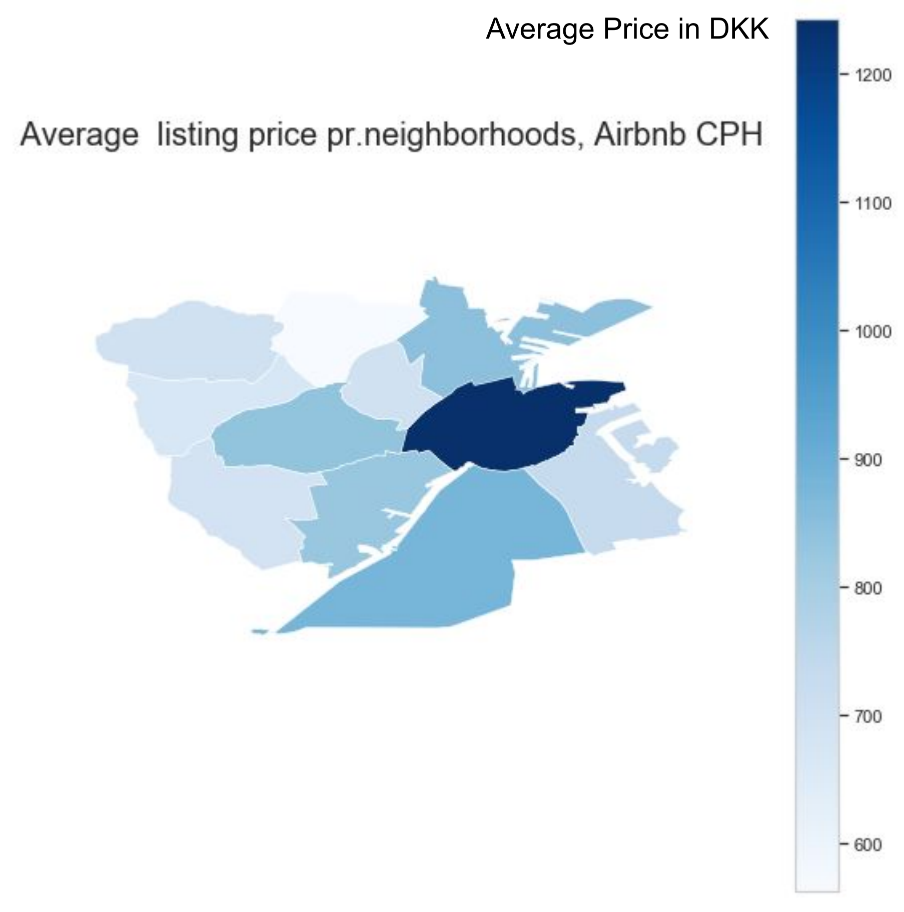
Results

Model	Training RMSE	Test RMSE	Training R2	Test R2	Training Average error	Test Average error	Training accuracy	Test accuracy
Linear	385DKK	401DKK	0,49	0.48	236DKK	242DKK	71,21 %	71,38 %
Lasso	385DKK	401DKK	0,49	0,48	236DKK	243DKK	71,211 %	71,38 %
GBR	252DKK	327DKK	0,77	0,61	155DKK	206DKK	81,24	75,1 %
XGBOOST	282DKK	333DKK	0,70	0,61	177DKK	206DKK	78,1 %	75,1 %
RF	146DKK	343DKK	0,94	0,60	79DKK	211DKK	91,1 %	74,9 %
MLP	295DKK	368DKK	0,68	0,53	186DKK	230DKK	77,5 %	72,7

Results

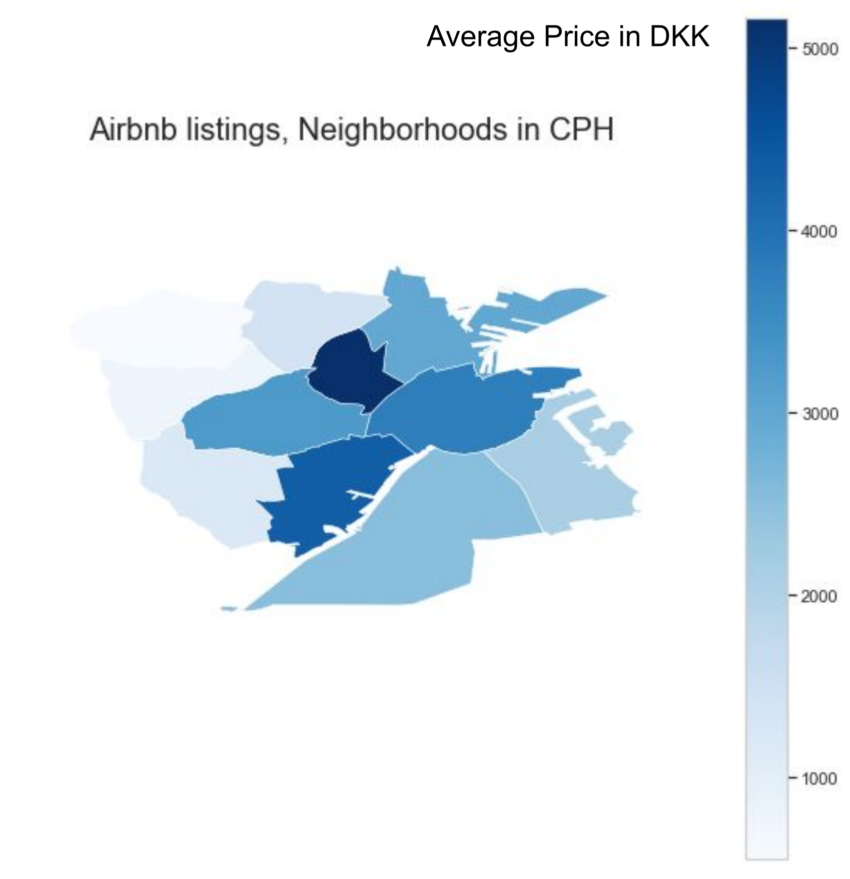
In this study, I modeled Airbnb listings data in Copenhagen from September 2018 to September 2019. About 80% of listings are apartments, with and the average nightly rate of 1114DKK. based on 24 features indicated, on tree-based models, namely gradient-boosting regression and extreme gradient-boosting regression, explains the price variation in the training dataset quite well (pictures at coefficient of variation R2). The R2 measurement for the test dataset was fairly strong (about 0.61) with a root mean square error of about $ 327. Room_type function was a function of the greatest importance.


55.74
55.72
55.70
55.68
55.66
55.64
55.62
latitude
12.450
12.475
12.500
12.525
12.550
12.575
12.600
12.625
12.650
longitude
800
600
400
200
price



Average Price in DKK

Average listing price pr.neighborhoods, Airbnb CPH



Average Price in DKK

Airbnb listings, Neighborhoods in CPH