



Statistical Analysis of Airbnb Rates in New York City

Group 29

Alexa Schultz
Amber Chiotasso
Cristonique Basden
Duyen Nguyen
Stephen McFarlane
Taylor Pitts

CONTENTS

- 1. Introduction**
- 2. Problem Statement**
- 3. Modeling Type**
- 4. Dataset**
- 5. Visualization**
- 6. Results**
- 7. Observations**
- 8. Conclusion**
- 9. References**

INTRODUCTION

- **What is the problem?**
 - **Finding appropriate rates for Airbnb hosts to charge for their listings**
- **Why you chose this problem?**
 - **Help hosts figure out realistic and appropriate rates**
 - **Help consumers ensure they are getting a fair price**
- **Why it is a pressing problem now?**
 - **Lost sales from unhappy renters**
 - **Many airbnb hosts have little guidance on how to price their listings**
 - **Oversaturated markets**
 - **Unreasonably priced units**
 - **Lost profits**

PROBLEM STATEMENT

- **Problem statement**
 - An analysis of the NYC Airbnb unit pricing can determine quickly and appropriately how to price Airbnb units according to current market demands for their particular accommodations.
- **Hypothesis**
 - What are the factors that have significant influences on NYC Airbnb rates?
- **Solution Approach**
 - Building a forecast model that is able to predict Airbnb rates based on the historical data in terms of locations, number of reviews, room types, and availability.

■ Modeling types used: linear regression and random forests.

Linear regression

$$Y_i = f(X_i, \beta) + e_i$$

Y_i = dependent variable

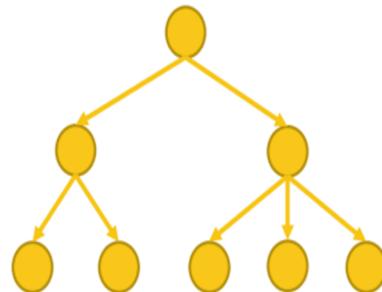
f = function

X_i = independent variable

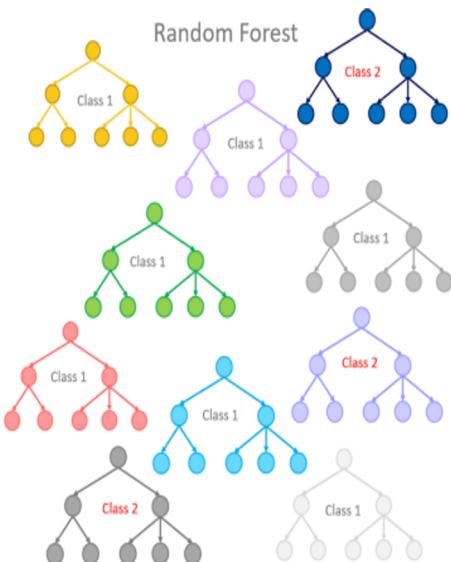
β = unknown parameters

e_i = error terms

Single Decision Tree



Random forest



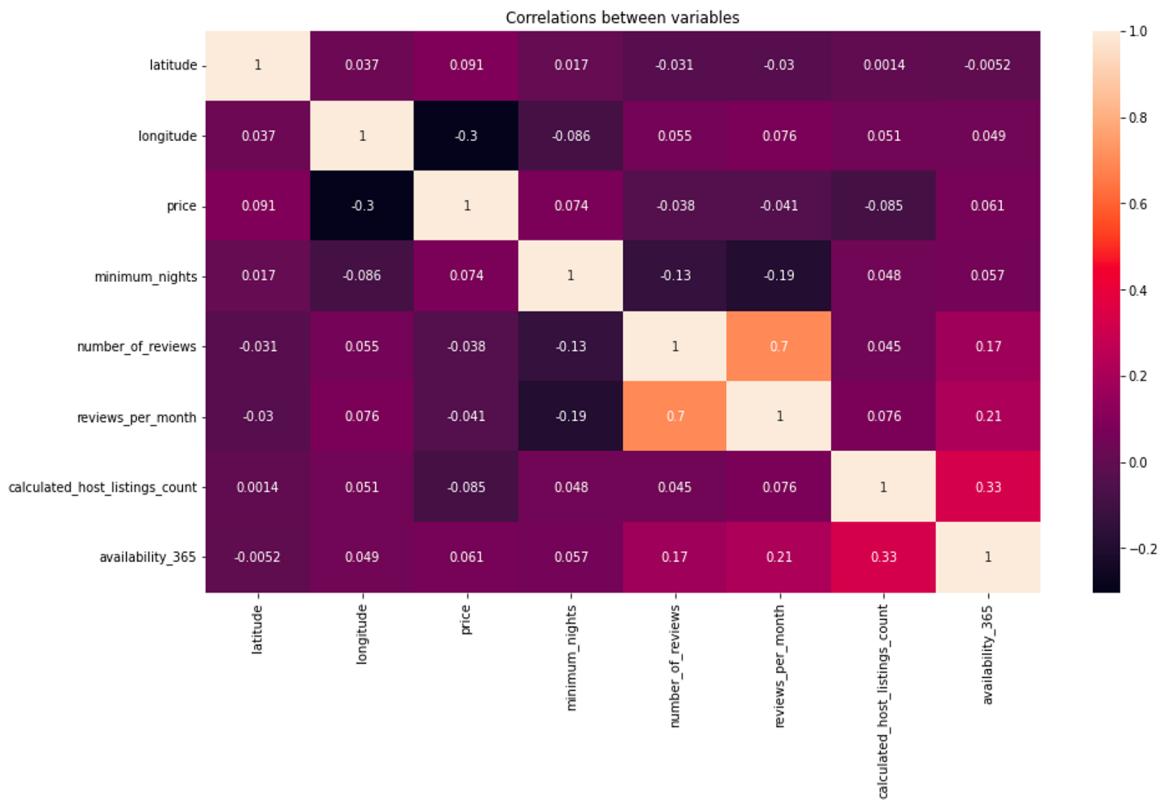
DATASET

- What are the variables in your dataset?
 - Independent variables

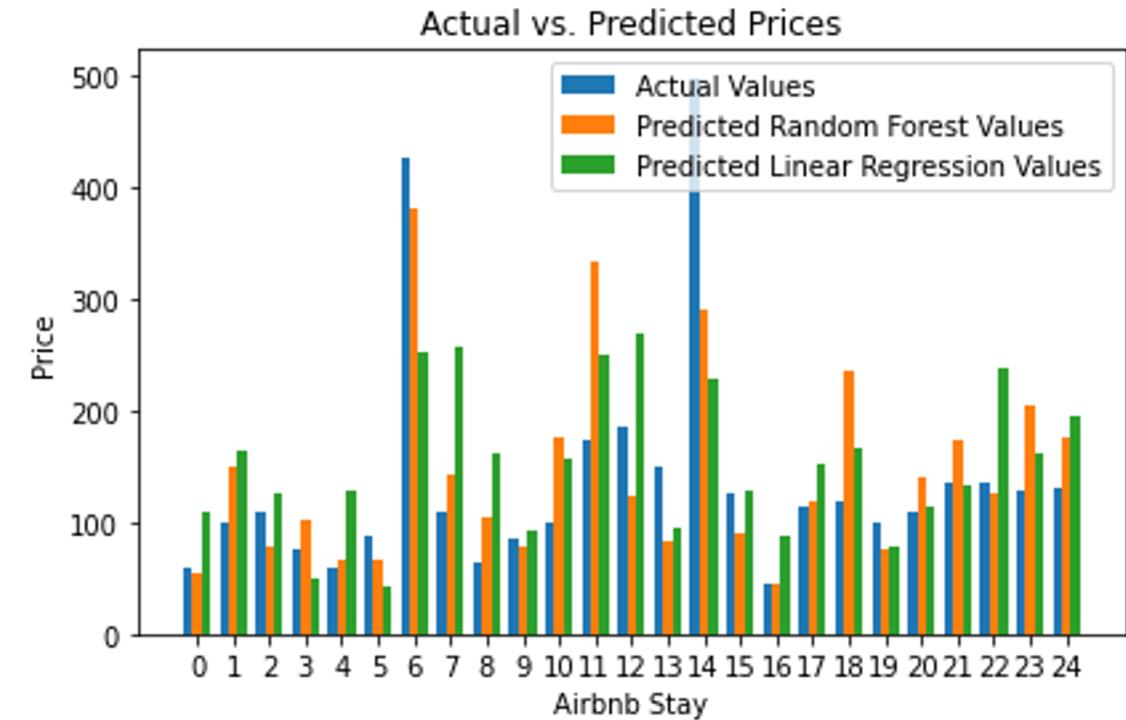
neighbourhood group	neighbourhood	latitude	longitude	room type	minimum nights	number of reviews	reviews per month	calculated host listings count	availability 365
---------------------	---------------	----------	-----------	-----------	----------------	-------------------	-------------------	--------------------------------	------------------

- Dependent variable: price
- Why do you think this dataset is appropriate for your project?
 - Free, open-source
 - Easy to use (csv formatted)
 - Provided enough qualitative and quantitative details on NYC Airbnb units
- Where did you get the data from? Mention the legitimate reference for your dataset
 - Kaggle Database, New York City Airbnb Open Data
https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/version/3?select=AB_NYC_2019.csv

VISUALIZATION

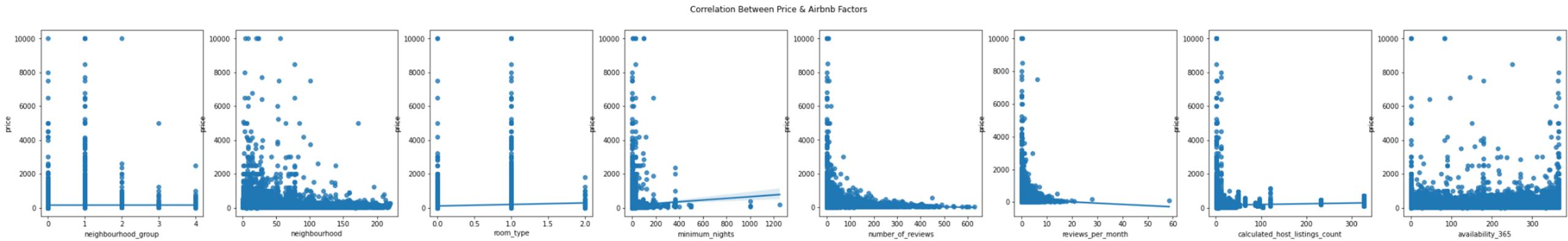


The matrix above reveals the correlation of each variable. This matrix can help identify relevant features for Airbnbs



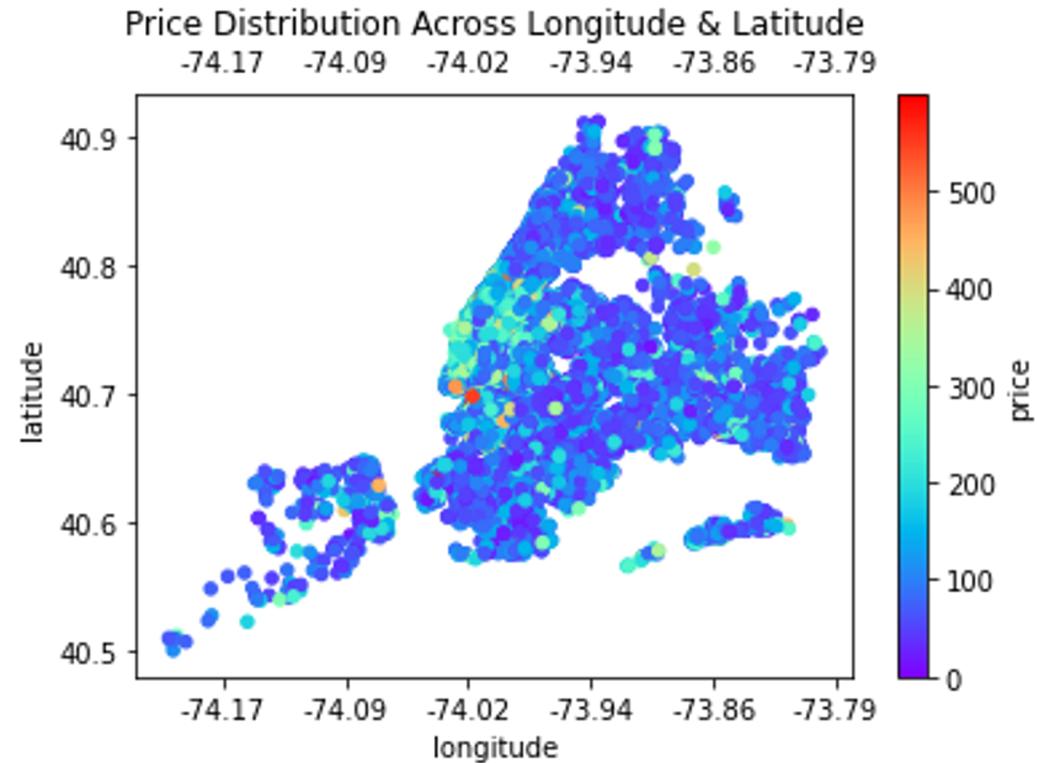
The bar chart reveals the differences between the actual price values of 25 random airbnbs and the two types of price values from the predicted models of individual airbnbs. This chart can help determine the accuracy of the models.

VISUALIZATION

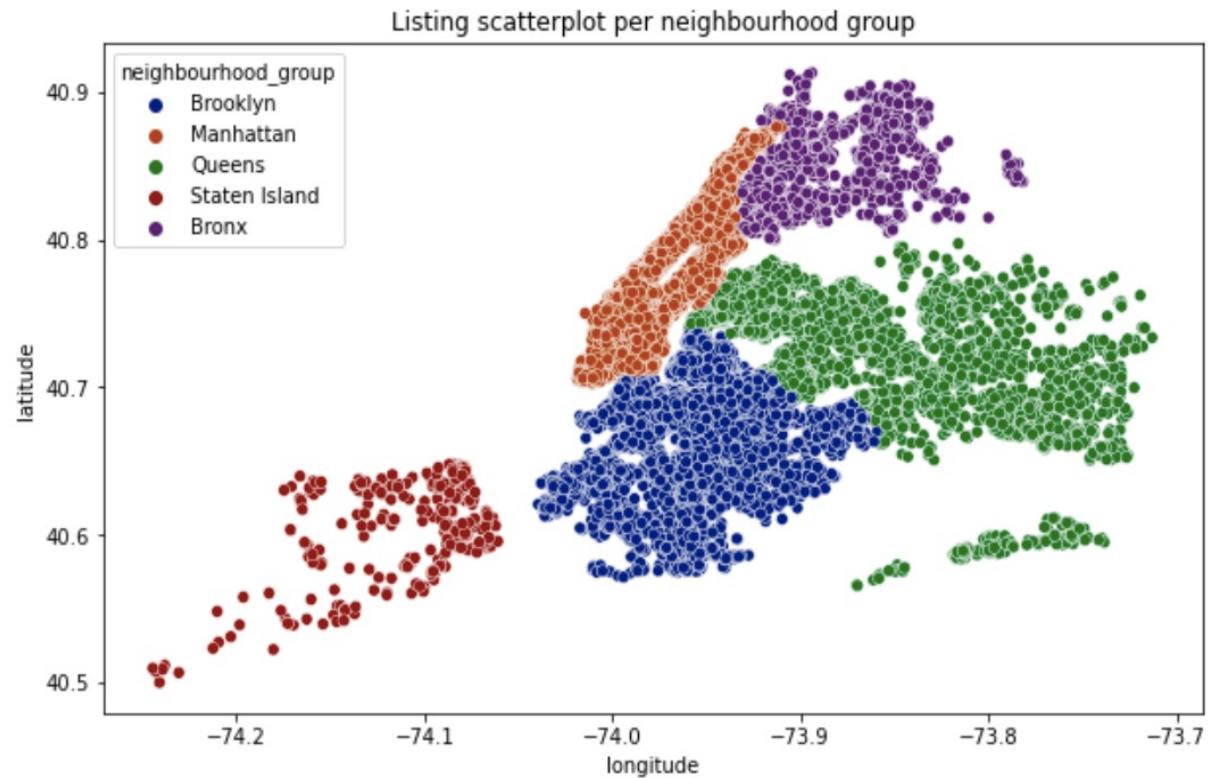


- Scatter plots illustrate and allow for observation of the relationship of some features & price.
- Understanding the relationships between the features and pricing can help airbnb host to improve certain features so as to make more profit.

VISUALIZATION



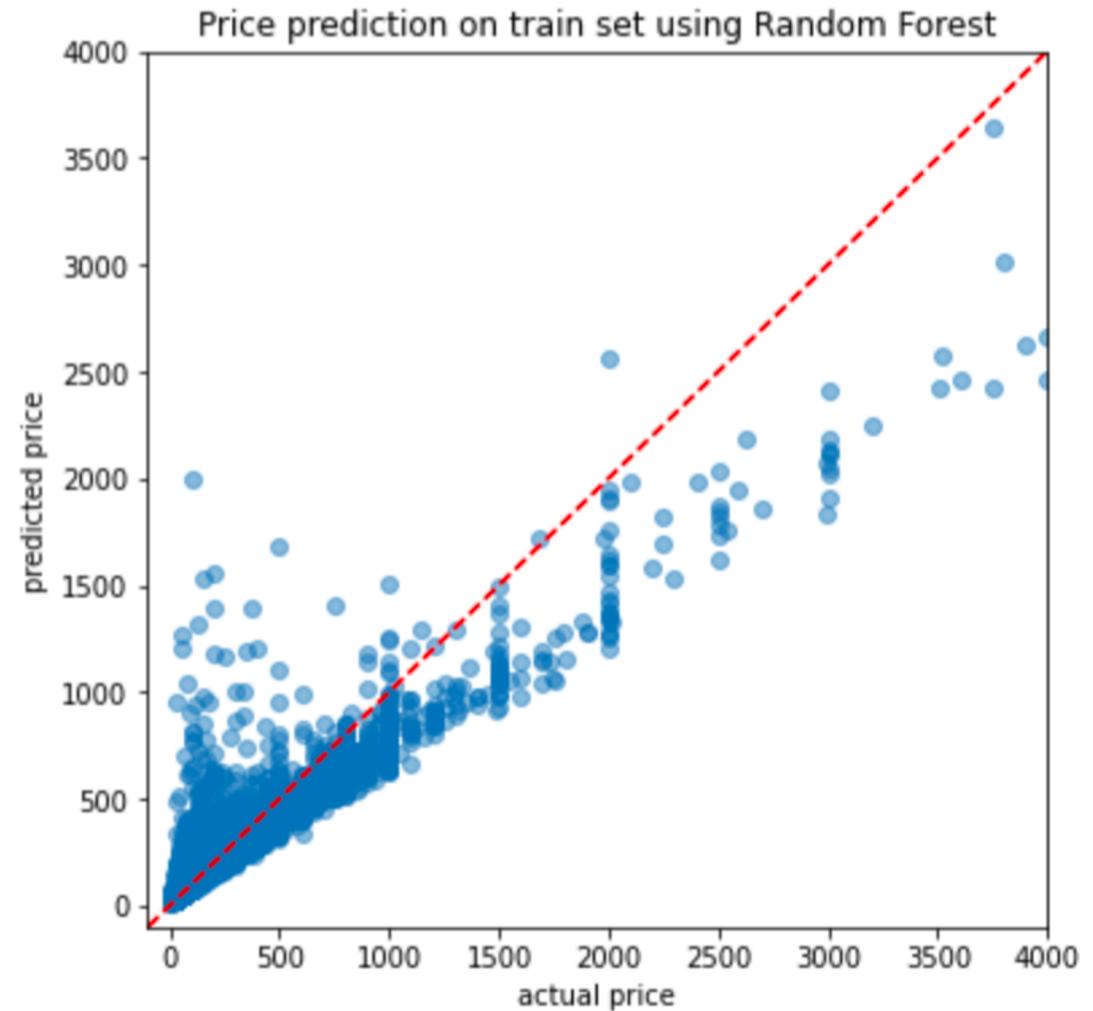
The heat map reveals the magnitude of the effects of longitude and latitude on the pricing of Airbnb. From the map, one can see that the location can influence the pricing of Airbnbs.



The scatter plot shows the listing density of 5 neighbourhood groups or cities.

RESULTS

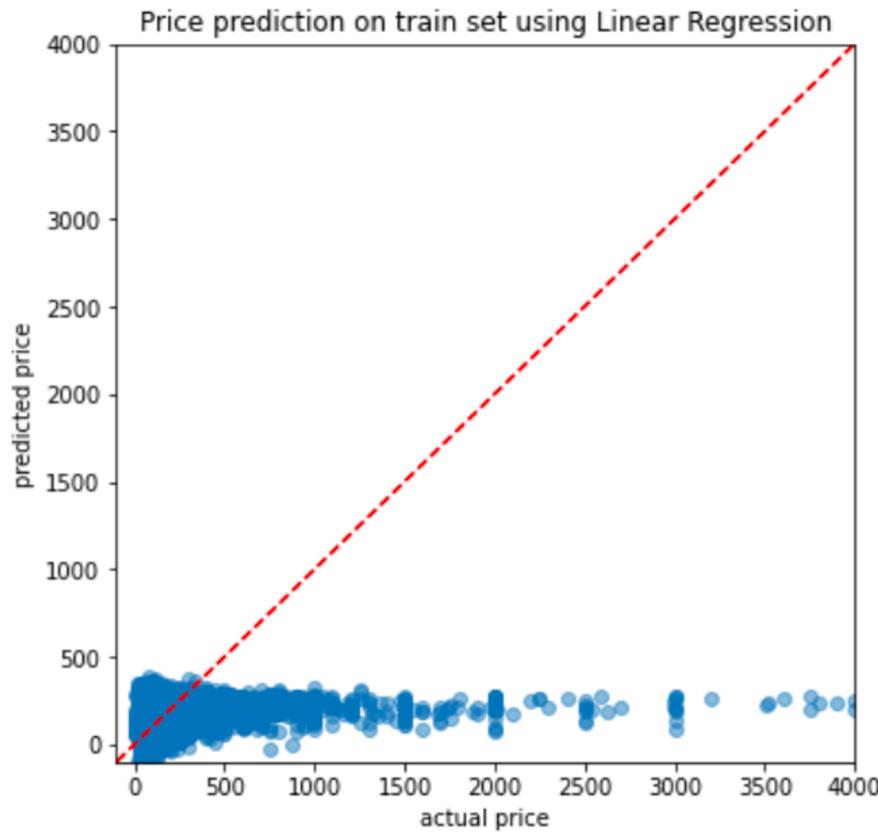
The random forest model was able to predict NYC Airbnb prices with the accuracy of 87%.



$R^2: 0.8763909986780196$

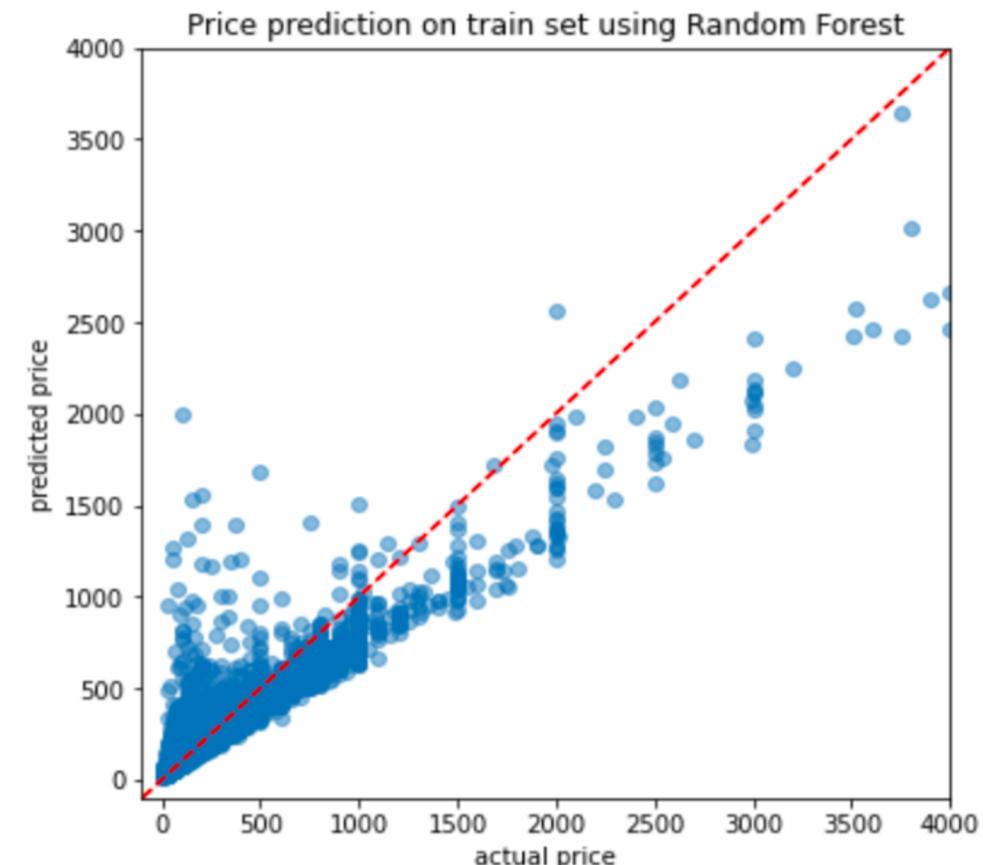
OBSERVATIONS

Linear Regression



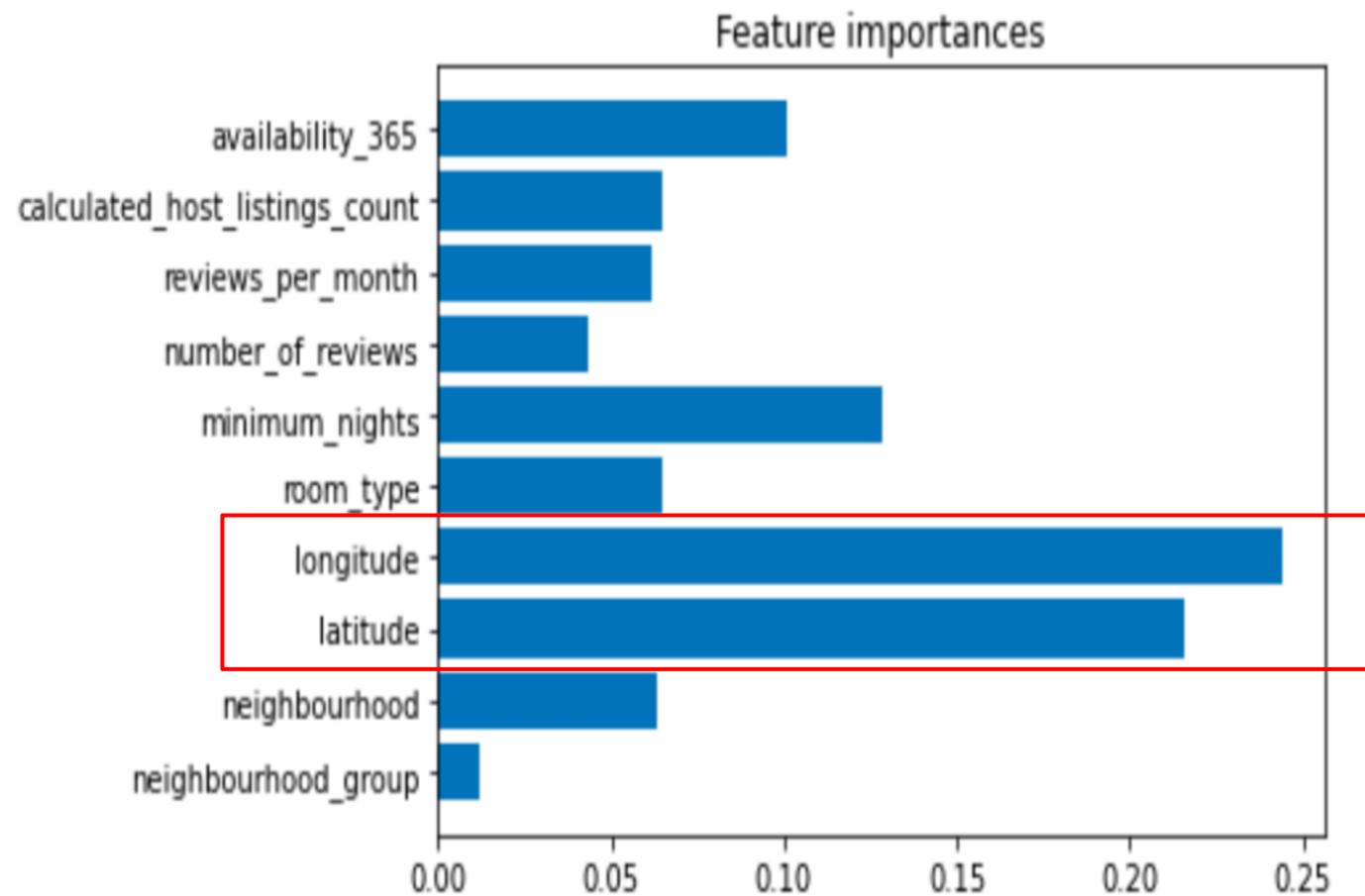
R2: 0.06734472901120103

Random forest



R2: 0.8763909986780196

CONCLUSION



- **Longitudes and latitudes have heavy influences on Airbnb pricing.**
- **Probably because of:**
 - **Downtown tourist attractions of each borough**
 - **Shorter commute to places**

REFERENCES

- Guttentag, D. (2019). Progress on Airbnb: A literature review. *Journal of Hospitality and Tourism Technology*. <https://doi.org/10.1108/jhtt-08-2018-0075>.
- Adamiak, C. (n.d.). Current state and development of Airbnb accommodation offer in 167 countries. Taylor & Francis. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/13683500.2019.1696758>.
- Nurdialit, D. G. (2020, October 6). Python exploratory data Analysis (EDA) on NYC Airbnb. Medium. Retrieved September 16, 2021, from <https://medium.com/analytics-vidhya/python-exploratory-data-analysis-eda-on-nyc-airbnb-cbeabd622e30>.
- Arfiano, A. (2020, June 26). Data exploration on Airbnb singapore: 01. Medium. Retrieved September 16, 2021, from <https://towardsdatascience.com/data-exploration-on-airbnb-singapore-01-40698c54cac3.k>
- Dgomonov. (2019, August 12). New York City airbnb open data. Kaggle. Retrieved from https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/version/3?select=AB_NYC_2019.csv.



Thank you for watching!