airbnb
New York

Project 4, Group 15

Lina Blanco

Gunel Garayeva

Fatina Hamadi

Kade Rivers

An analysis of data from Airbnb New York listings over several years and insight into Airbnb price predictions for consumers and hosts

What Is Airbnb

Airbnb is an online marketplace that connects people who want to rent out their property with people who are looking for accommodations, typically for short stays. Airbnb offers hosts a relatively easy way to earn some income from their property. Guests often find that Airbnb rentals are cheaper and homier than hotels.

Project Overview and Inspiration 🔍

This Airbnb Open Dataset is based in New York, USA.
This project provides insight for both travel and business entrepreneurs who may want to get utilize or work with Airbnb or other similar companies.

We were able to add in open source data between the 2010's and 2024 to allow for additional insight and and predictive capabilities.

Fields of interest for this project include the location information for mapping, prices on rentals, minimum number of nights required to rent as well as availability throughout the year, and information on the host's listings and reviews of the locations.

Minimal cleaning was needed after combining for this dataset and additional datasets could be added in the future for re-use and increased predictive qualities over time.

Project Goals

1 Map out New York Airbnb Locations based on Neighborhood
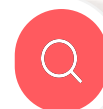
2 Discover Insights into Location, Property Features, and Price Ranges

3 Pricing Predictions and Recommended Similar Properties

Local Law 18

A law that changed how Airbnb Rentals in New York were regulated

2022-2023

A law that went into effect in 2022 (but wasn't fully enforced until 2023) to ensure safer rental arrangements and further verify hosts with fines for breaking this law and the regulations listed with short-term rental units.

Reasoning

Prior to this law going into effect, there were over 47,000 listings in New York, there were no limitations on guests staying in the rental, hosts were not always verified as living in or near the property, and there were no safety regulations being enforced for renters to ensure the property was up to code and actually safe for rental. This law saw a drop of around 7,500 rentals that were no longer able to lawfully operate after New York declared to start enforcing the law September 5, 2023
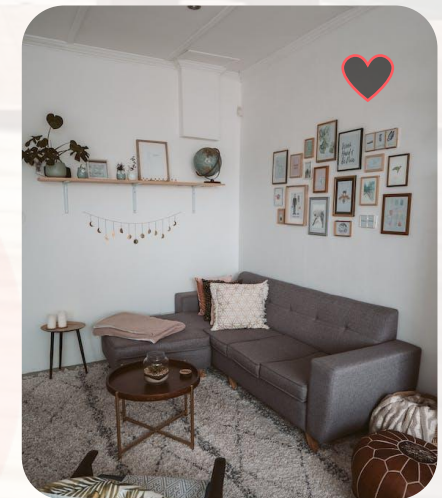
Design Concepts

Color Palette

The slidedeck and theme is the color palette for Airbnb and is located here.

Web Theme

The bootswatch theme chosen is titled "Solar" and is located here.

Slidedeck

The slidedeck was chosen from Slidechef and is credited on the final slide and is located here.

Tableau and Machine Learning

The tableau stories and dashboards were a collaborative effort to avoid repetition in design and insight into the project. The machine learning and recommender are intended to be used together or separate for insights into pricing.

Python Anywhere

This project was deployed using Python Anywere and is located here: https://project4group15.pythonanywhere.com/

GitHub

This project was shared and created using a GitHub repository located here:

https://github.com/musicmaj/project-4-group-15
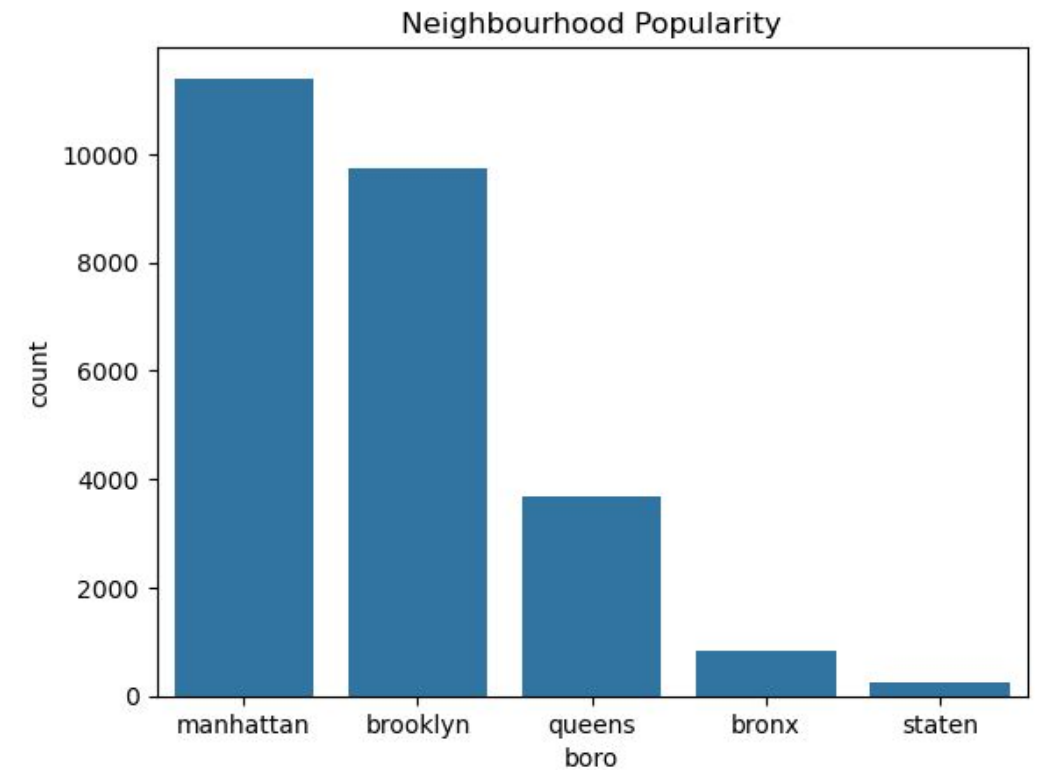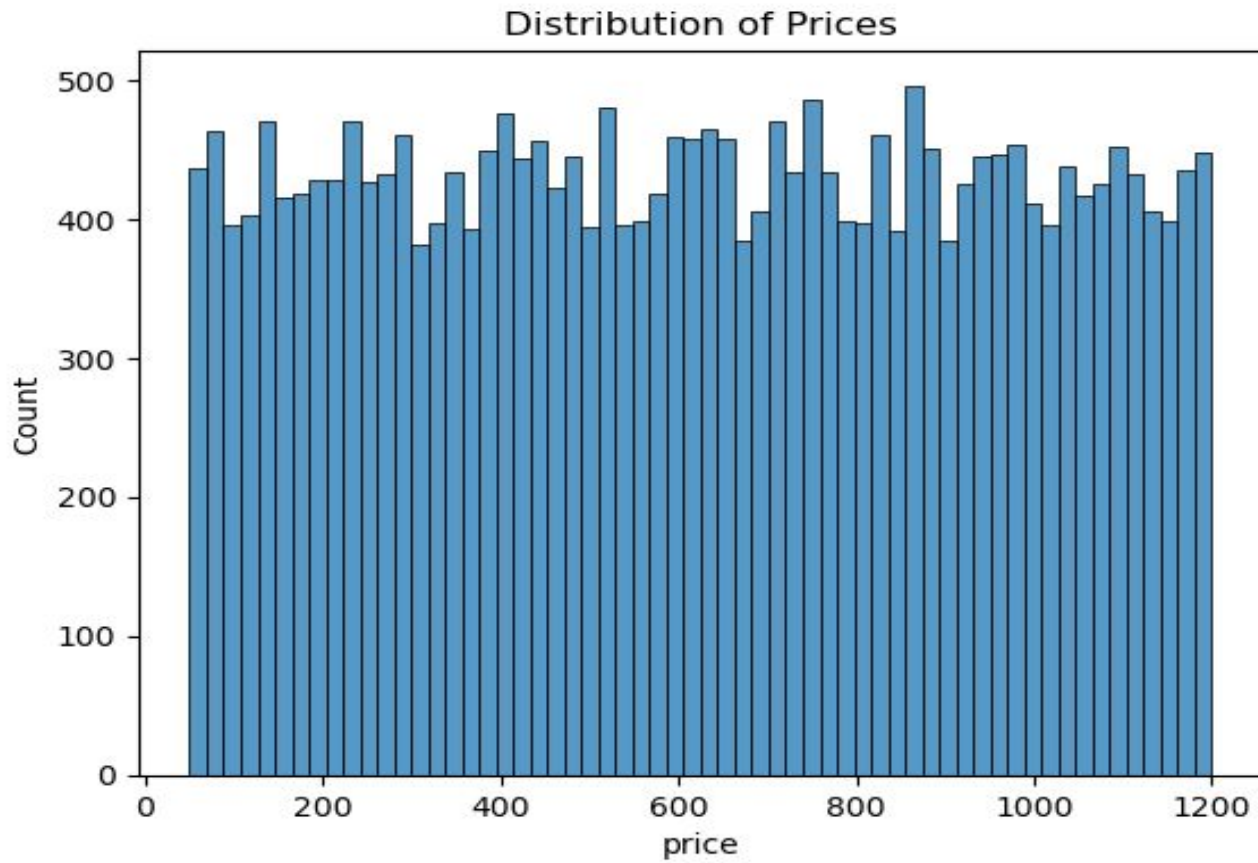
Data Selection

The original dataset was from Kaggle and is located here. The additional datasets to add in additional years came from here.

Notebook Visualizations

Jupyter Notebook visualizations created during data cleaning and preparations for machine learning
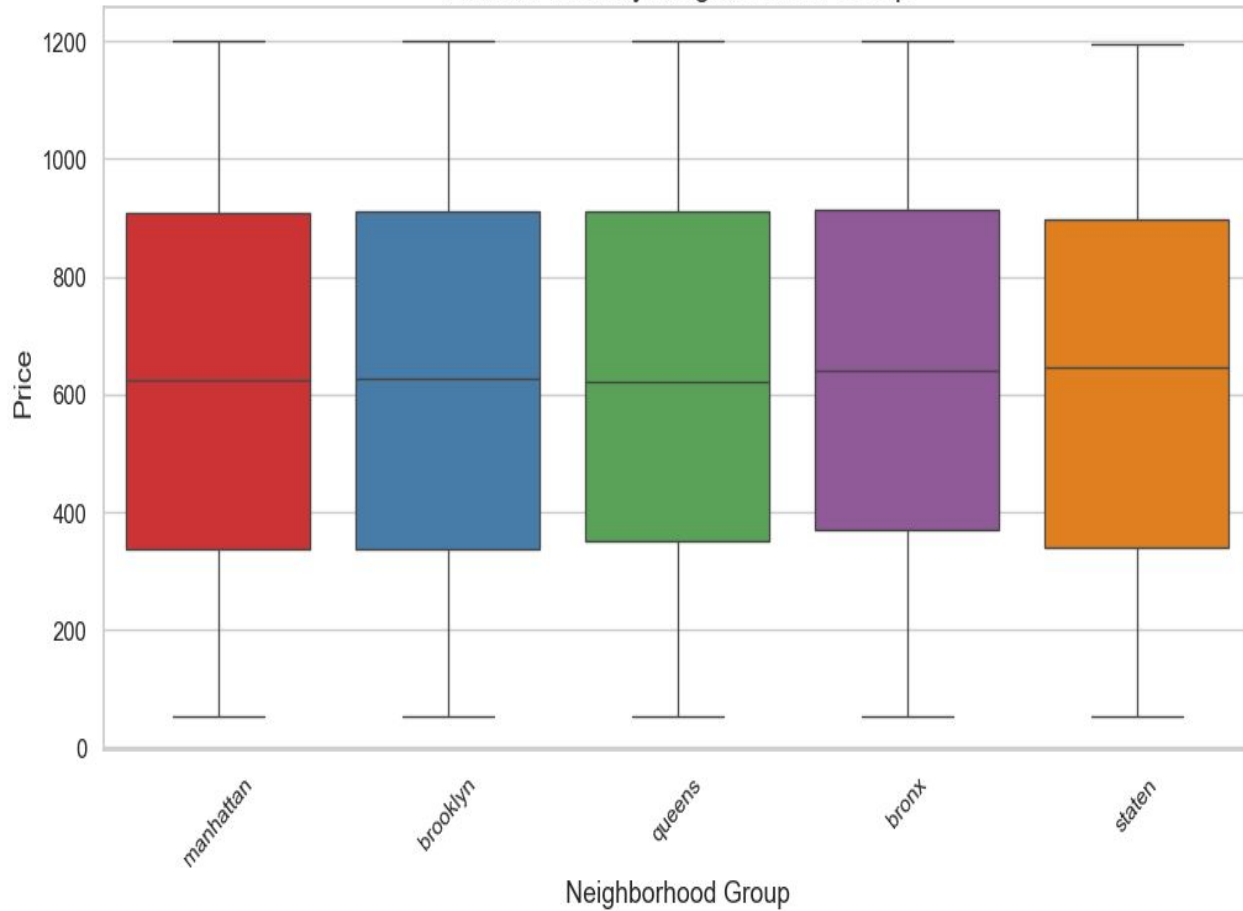
Data Cleaning for Machine Learning

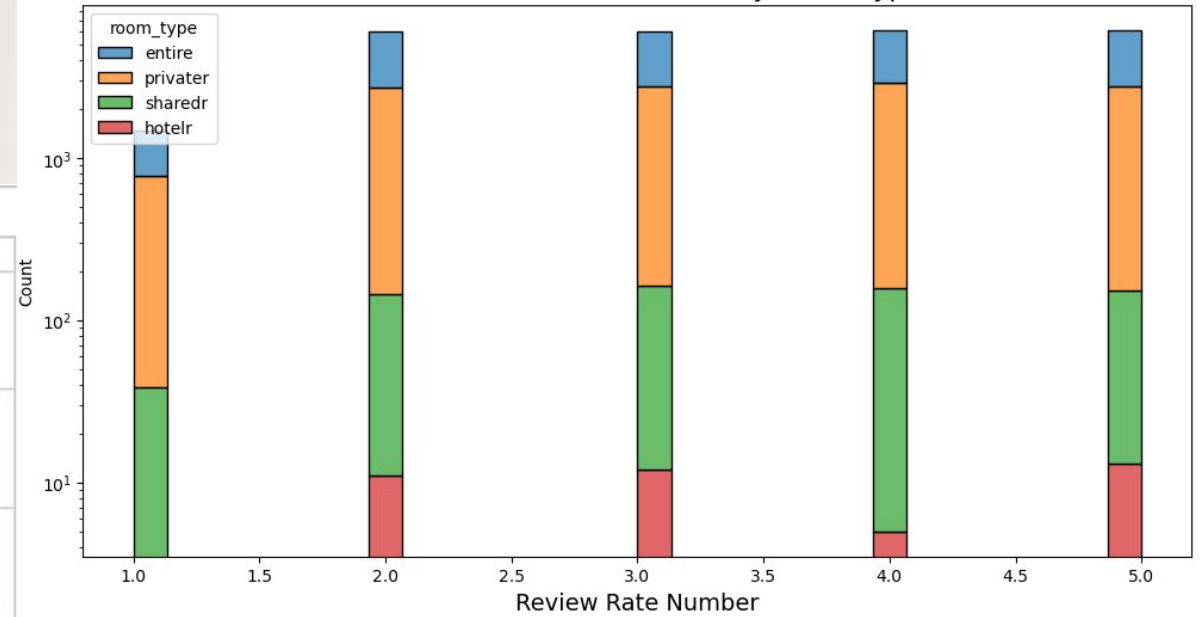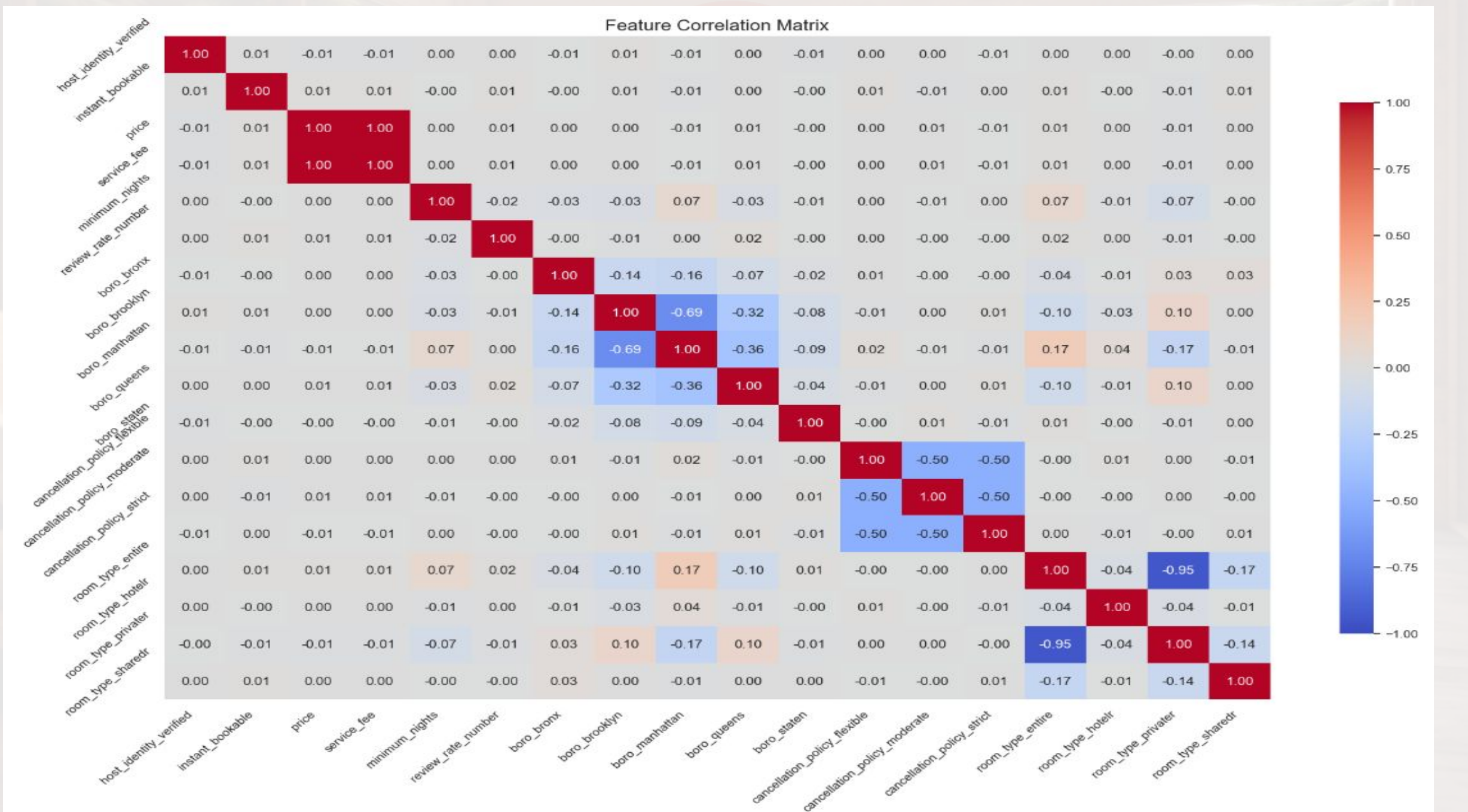Distribution of Prices

Neighbourhood Popularity

Data Cleaning for Machine Learning

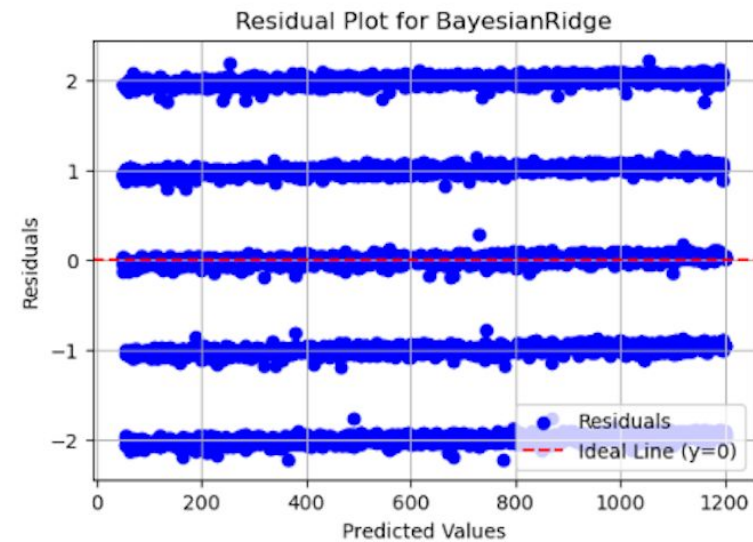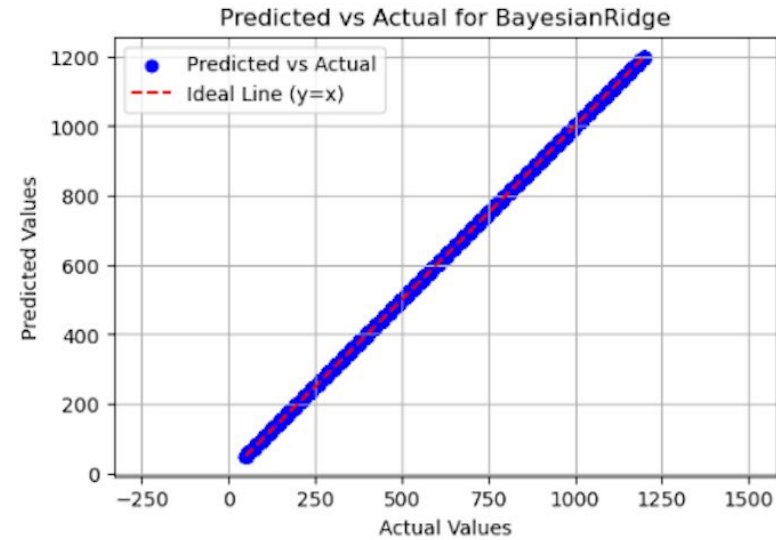Airbnb Prices by Neighborhood Group

Distribution of Review Rates by Room Type

Feature Correlation Matrix

Machine Learning Results

The best model is: BayesianRidge with parameters: {}
Final Evaluation Metrics for BayesianRidge:
MSE: 2.0128061814385974, MAE: 1.2108597179785323, RMSE: 1.4187340065842495, R²: 0.9999817971928254

Machine Learning - Model Evaluation

```
Training Set Evaluation:
Mean Absolute Error: 1.2154024057603618
Mean Squared Error: 2.023922885400747
Root Mean Squared Error: 1.422646437243192
R-squared: 0.9999814933075519

Test Set Evaluation:
Mean Absolute Error: 1.2108597179785323
Mean Squared Error: 2.0128061814385974
Root Mean Squared Error: 1.4187340065842495
R-squared: 0.9999817971928254
```


Model Evaluation: Trained vs. Tested Set
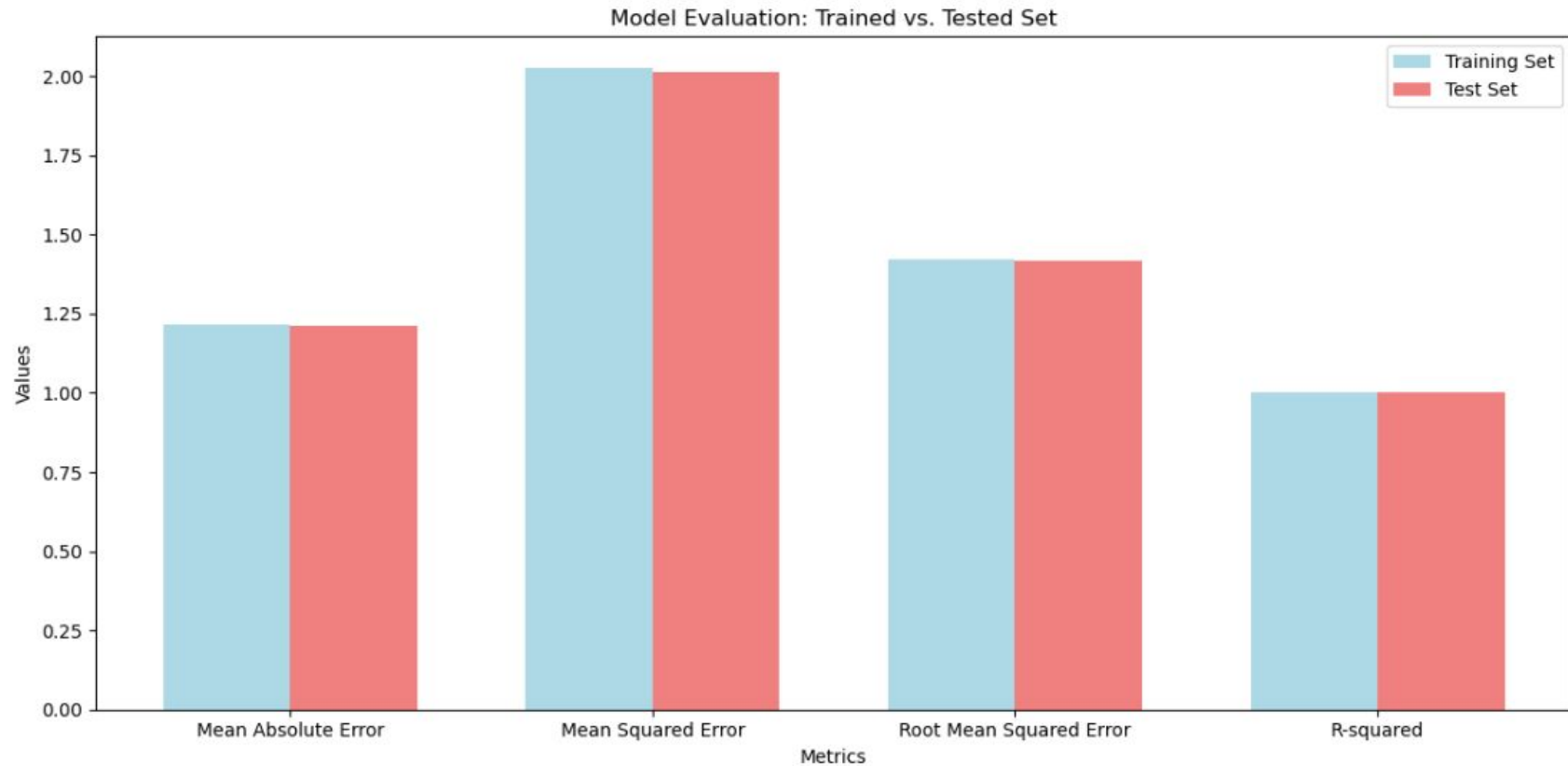
Machine Learning - Model & Metrix Performance

Live Demo

https://project4group15.pythonanywhere.com/
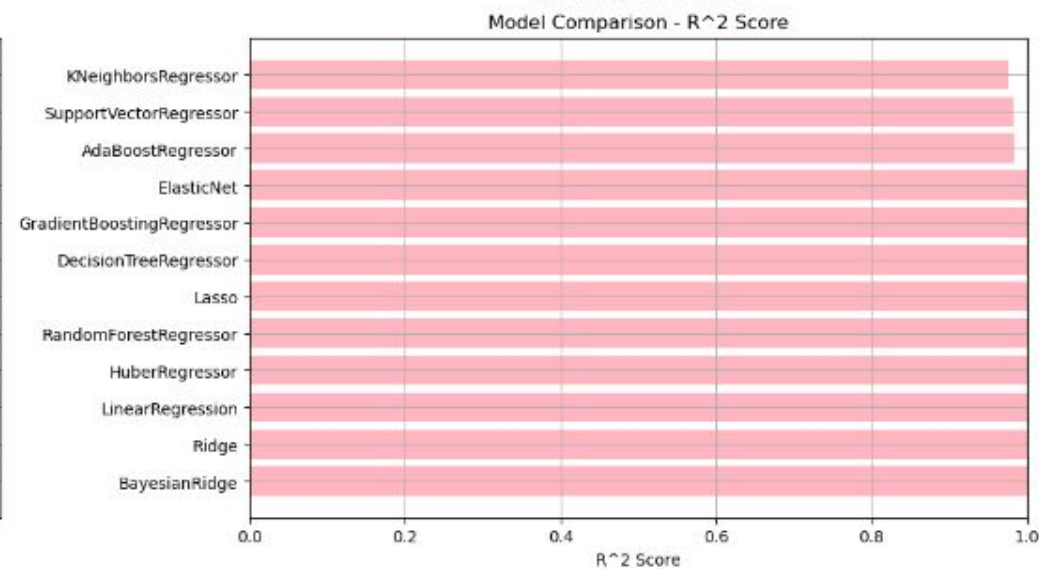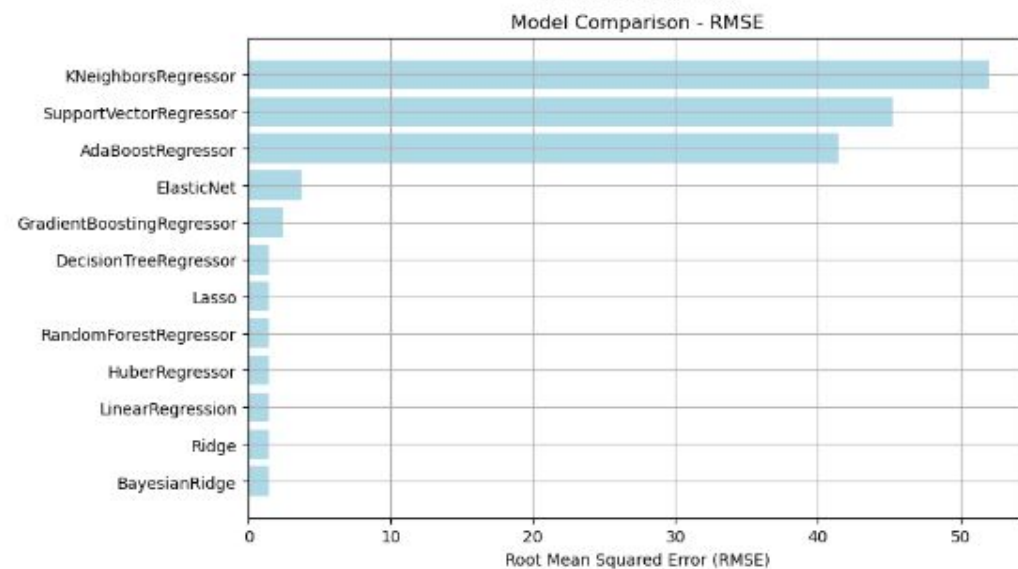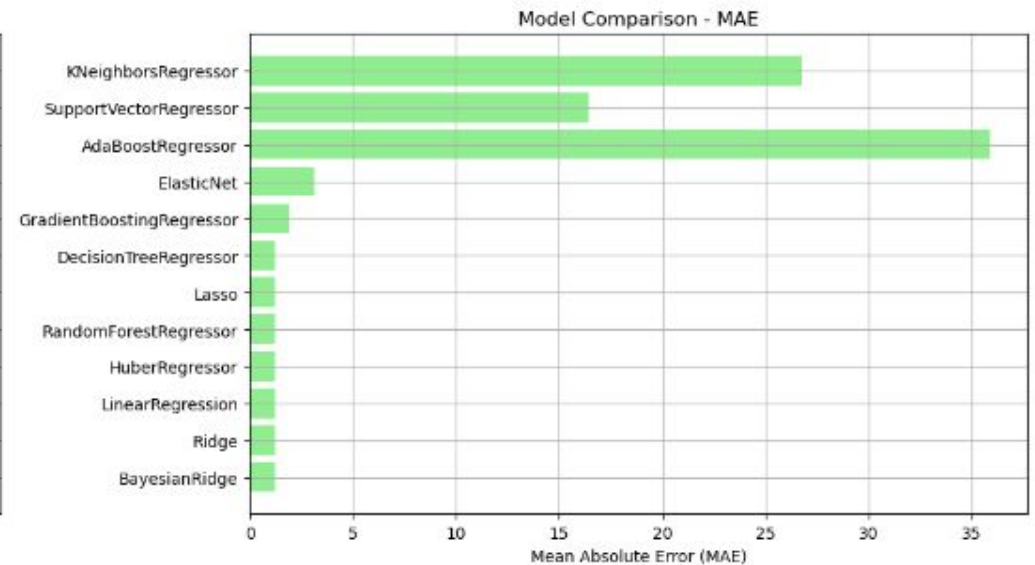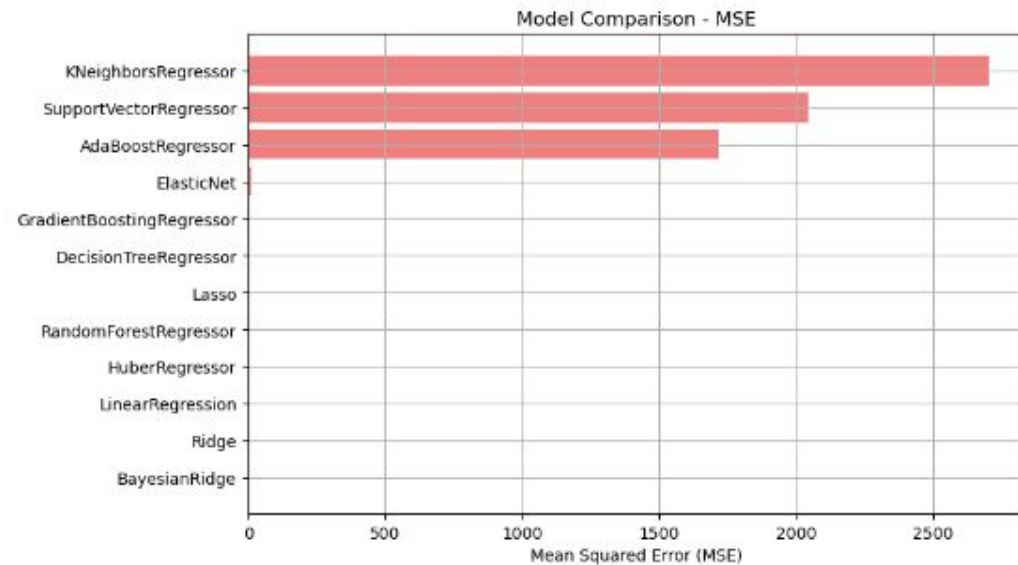
Recommender

A tool created to input information about potential rentals and generating a list of the closest related rentals in the 2024 dataset.

2024 Cleaned DataFrame

	rental_id	neighbourhood_group	neighbourhood	room_type	price	minimum_nights	number_of_reviews
0	1189243425411300671	Manhattan	Murray Hill	Entire home/apt	58.0	30	1
1	651593916026998398	Brooklyn	Flatlands	Private room	80.0	30	0
2	310325	Manhattan	Harlem	Private room	75.0	30	31
3	57261212561500056	Brooklyn	Sunset Park	Private room	45.0	30	6
4	1020282701018874374	Brooklyn	Bedford-Stuyvesant	Private room	47.0	30	0

Function Name and User Input to generate recommendations

```python
def recommendation(closest_rentals_length, neighbourhood_group, room_type, price):
```

```python
# User input closest_rentals_length, neighbourhood_group, room_type, price
closest_rentals_length = 10
neighbourhood_group = "Brooklyn"
room_type = "Private room"
price = 80


response = recommendation(closest_rentals_length, neighbourhood_group, room_type, price) # used for the flask app, returns JSON


# to test
pd.DataFrame(response)
```

Recommender Output

	rental_id	neighbourhood_group	neighbourhood	room_type	price	minimum_nights	number_of_reviews	distance
0	17160286	Brooklyn	Bedford-Stuyvesant	Private room	100000.0	30	29	0.000000
1	605115521796576121	Brooklyn	Bedford-Stuyvesant	Private room	100000.0	30	9	0.000004
2	100441620024086150	Manhattan	Upper East Side	Entire home/apt	15618.0	31	0	0.002764
3	904524100106225881	Manhattan	SoHo	Private room	20000.0	1	0	0.003593
4	904523946643334652	Manhattan	SoHo	Private room	20000.0	1	0	0.003593
5	904524223667771043	Manhattan	SoHo	Private room	20000.0	1	0	0.003593
6	52862058	Manhattan	Financial District	Entire home/apt	10000.0	30	1	0.006624
7	830682282357157632	Manhattan	Midtown	Entire home/apt	10000.0	30	0	0.006692
8	692813905111173363	Manhattan	Battery Park City	Entire home/apt	10000.0	30	0	0.006692
9	830656153550799267	Manhattan	Midtown	Entire home/apt	10000.0	30	0	0.006692

Final Notes

Conclusions

While Airbnb has changed the way we travel for business or pleasure, it's also called for updated regulations and laws to ensure it's continuation and the verification of hosts and safety for renters. With the open source data available, we were able to visualize the geographic locations of rentals, discover important and impactful features of top performing rentals and neighborhoods as well as use machine learning to make predictions on the prices of rentals which can be used by consumers or future hosts. This information was able to be uploaded and usable in Python Anywhere.

Bias/Limitations

- The dataset didn't include how many nights the rentals were rented throughout the year
- The dataset was focused in New York City only
- The limitations of versions of PythonAnywhere instead of a container prevented the use of the Recommender portion of the website without modifying or creating a new dev environment with older libraries

Future Work

The options are limitless when it comes to adding onto this project and continuing on with additional datasets, other short-term rental companies like Vrbo, further insight into how often each rental is rented out throughout the year, etc… This was a fantastic project to show the skills we have learned and how we can transform, visualize, and predict using data.

[Slide Chef](https://slidechef.net)