airbnb

NEW YORK AIRBNB PRICE PREDICTION

BACKGROUND

Airbnb is a rental platform which offers rental properties

It provide a variety of shared and independent short-term need based options to tourists, students and business professionals.

Airbnb is trying to introduce and price its services based on the features listed for the lodging so that they can provide customized solutions on the basis of the customer's budget and demand.

DATA ACQUISITION AND CLEANING

- A WELL-DEVELOPED DATASET COULD BE FOUND FROM KAGGLE DATASET "NEW YORK CITY AIRBNB OPEN DATA" HTTPS://WWW.KAGGLE.COM/DGOMONOV/NEW-YORK-CITY-AIRBNB-OPEN-DATA

- THERE ARE 48895 ROWS AND 16 COLUMNS

- IRRELEVANT FEATURES ARE DROPPED. 12 COLUMNS LEFT AFTER DROPPING.

- MISSING VALUES IN REVIEW_PER_MONTH ARE REPLACED BY 0.

MISSING VALUES BEFORE CLEANING

```
id                                0
name                             16
host_id                           0
host_name                        21
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
last_review                   10052
reviews_per_month             10052
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

MISSING VALUES AFTER CLEANING

```
host_id                           0
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
reviews_per_month                 0
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```
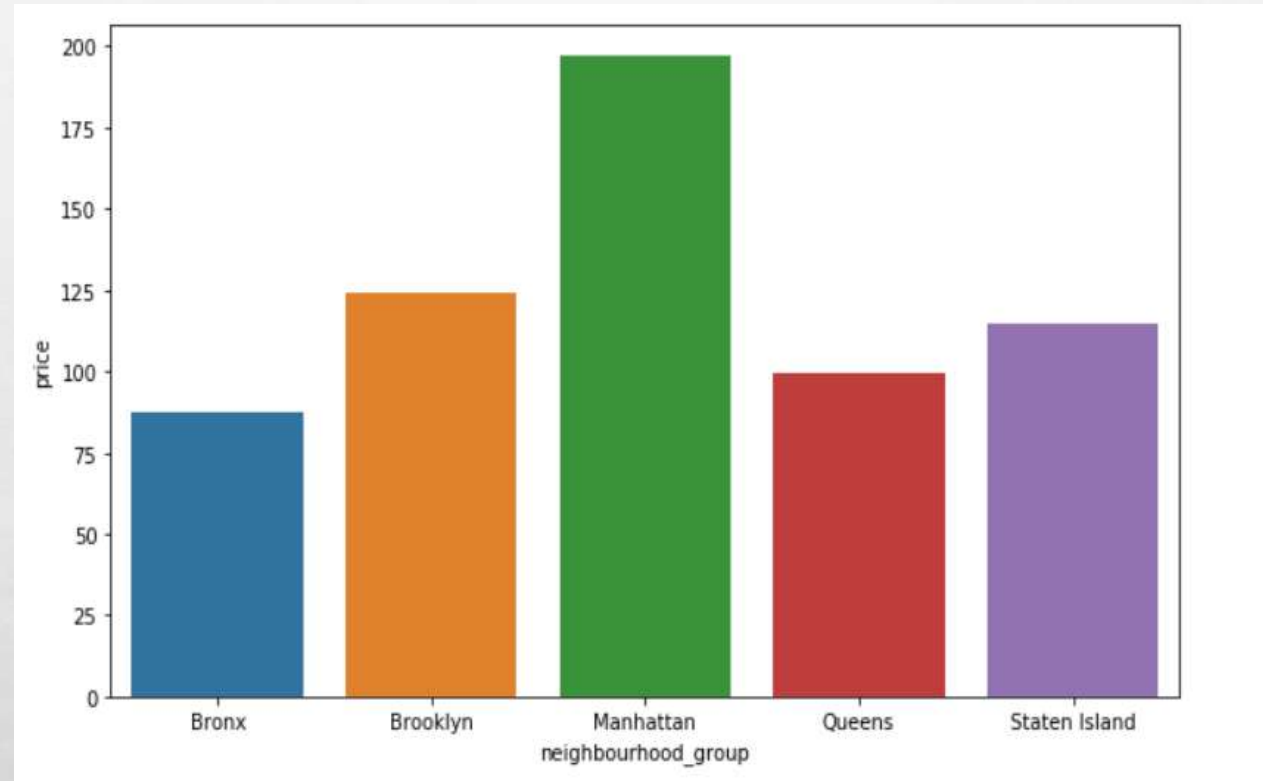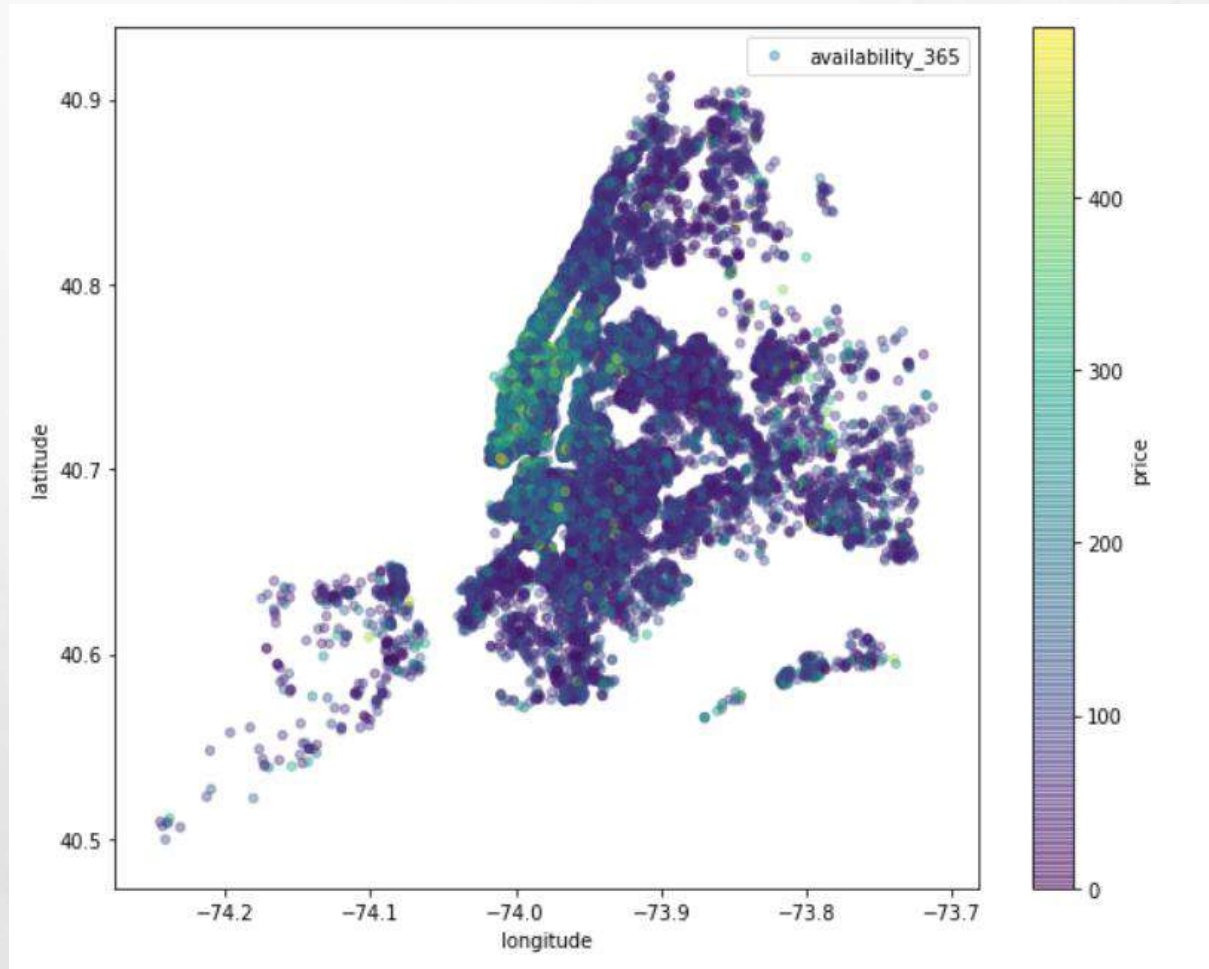
CORRELATION HEAT MAP
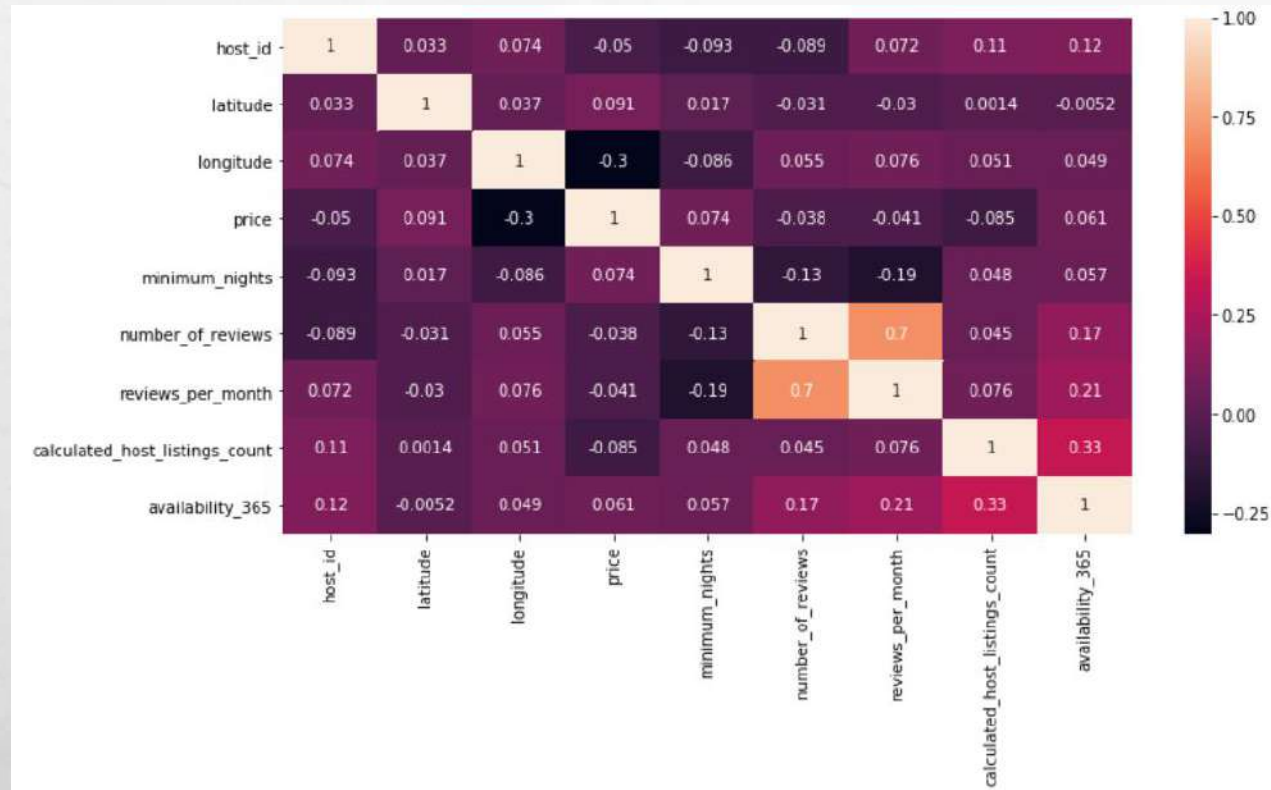
- **Correlation matrix as heatmap is used to check the correlation among different numeric variables as there are multiple columns in the dataset**

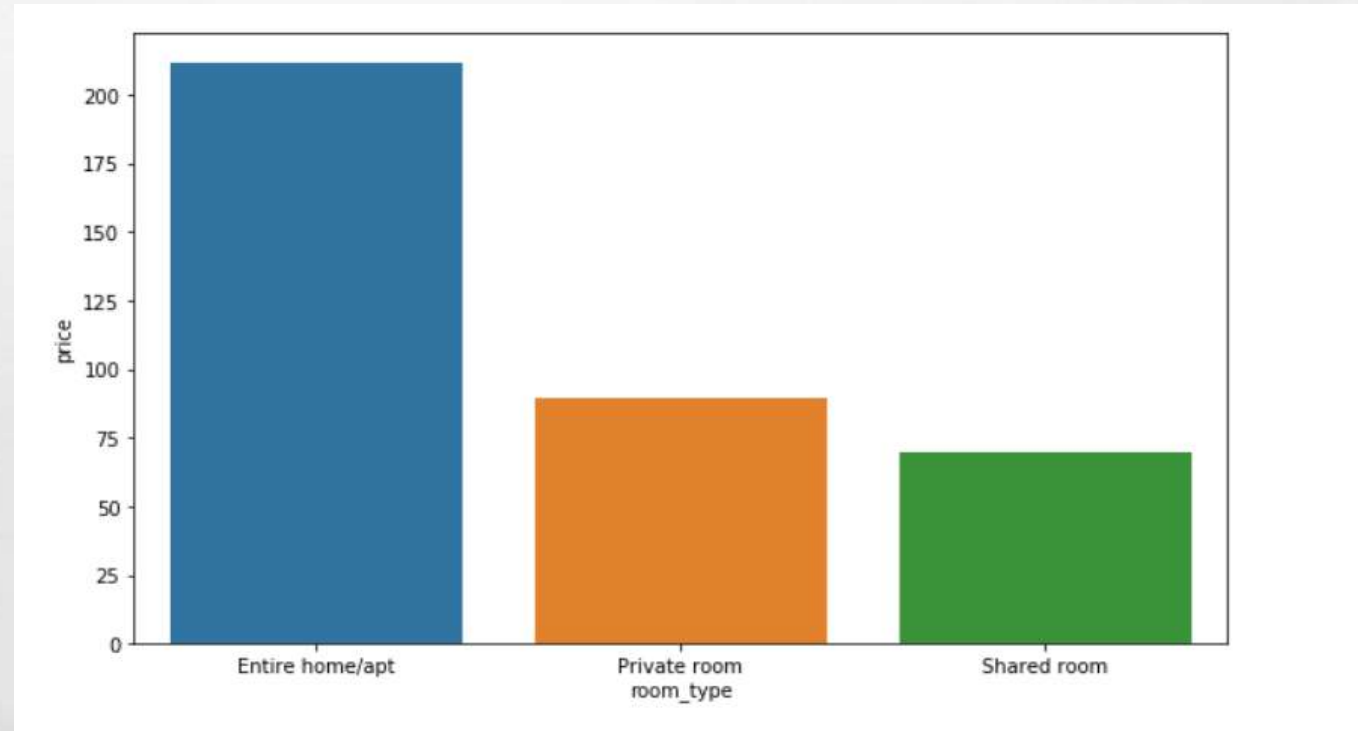- **Higher correlation means there are higher chances of multicollinearity**
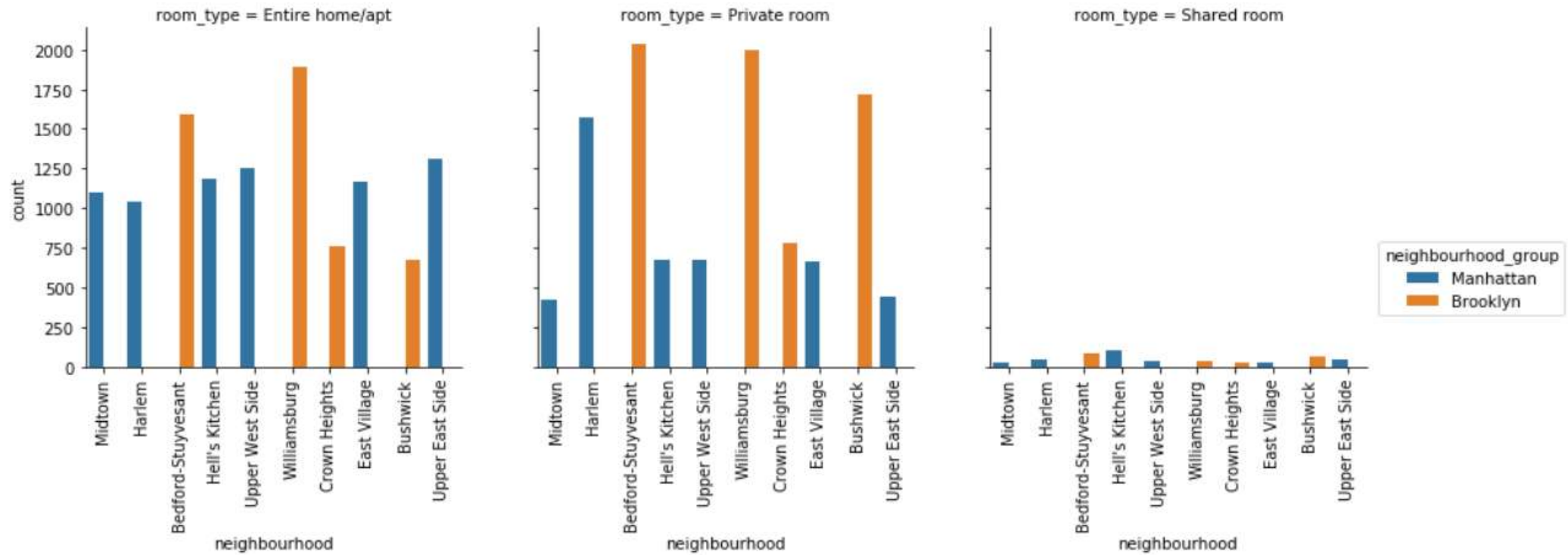
RELATIONSHIPS BETWEEN NEIGHBOURHOOD AND PRICE

PRICE DISTRIBUTION

RELATIONSHIP BETWEEN ROOM TYPE AND PRICE

POPULAR ROOM TYPES IN TOP 10 NEIGHBORHOODS

METHODOLOGY

- MULTIPLE LINEAR REGRESSION
- LASSO
- RIDGE
- ELASTIC NET
- BAGGING
- RANDOM FOREST

RESULT

Method	RMSE	R2 score
MLR	176.58337	0.07480
Lasso	176.58618	0.07477
Ridge	176.57157	0.07493
Elastic Net	176.57958	0.074843
Bagging	178.10512	0.058788
Random Forest	161.56591	0.22547

Random Forest produces the lowest RMSE, indicating the best model is this analysis.

DISCUSSION AND CONCLUSION

- THERE ARE STILL MANY LIMITATIONS THAT WILL AFFECT THE ACCURACY.

- FIRSTLY, THERE MIGHT BE OTHER MODELS THAT COULD PRODUCE MORE ACCURATE PREDICTIONS.

- SECONDLY, THE PRICE OF AN AIRBNB PROPERTY MIGHT BE AFFECTED BY OTHER FEATURES, FOR EXAMPLE, THE INTERIOR DESIGN.

- THIS PREDICTION CAN BE HELPFUL FOR PEOPLE WHO WOULD LIKE TO START A AIRBNB BUSINESS OR TOURISTS WHO ARE LOOKING FOR AN AIRBNB.

- THANK YOU