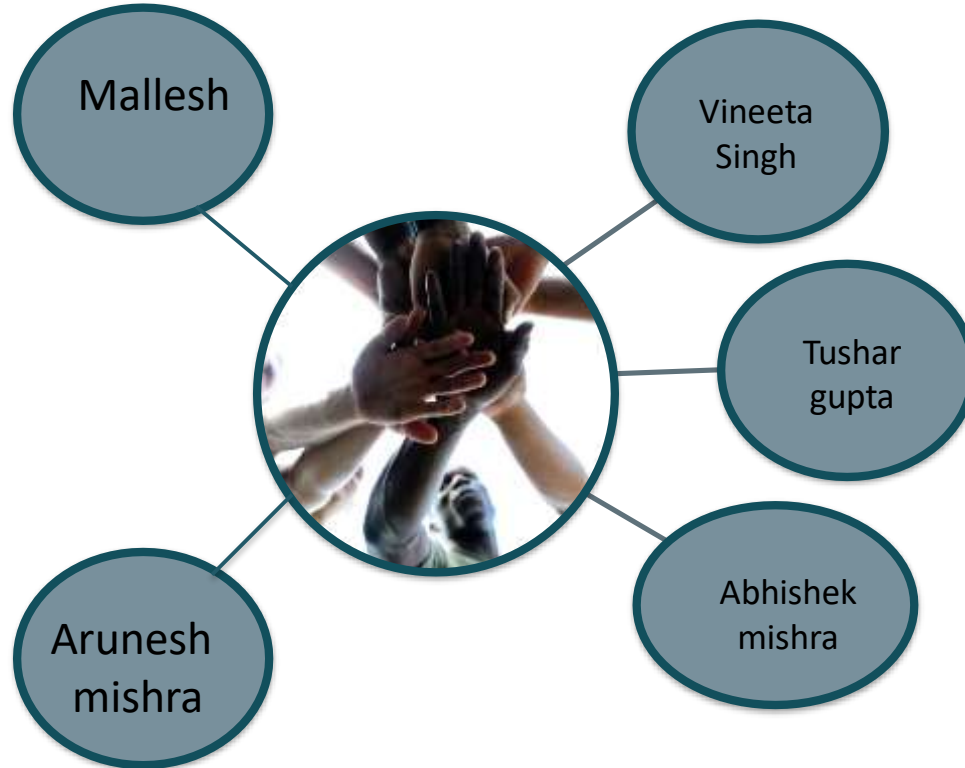


Capstone Project

Airbnb Booking Analysis

TEAM MEMBERS



WHAT IS AIRBNB



Airbnb began in 2008 when two designers who had space to share hosted three travellers looking for a place to stay. Now, millions of hosts and travellers choose to create a free Airbnb account so they can list their space and book unique accommodation anywhere in the world. And Airbnb experience hosts share their passions and interests with both travellers and locals.



Objective of The Project

The Main Objective Of This Project is to find the Relationship Between Property Type ,Neighbourhood or Price or any Correlation Observation Between Review and Price. Also giving meaningful insights from the data to make business decisions

Business Problems

What can we learn about different hosts and areas?

Airbnb need to know the Booking factors and predictions? (ex: locations, prices, reviews, etc)

Its mandatory for Airbnb to find the hosts which are the most busiest and why?

Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Road Map for Data Analysis

HOST

**Finding
Top Host**

Top three
busiest host
and also
location
wise top
busiest host

NEIGHBOUR -HOOD

**Neighbourhood
-group and
Neighbourhood
max bookings**

Relationship
Between price
and neighbour
and its listing

Location getting
maximum number
of bookings on
basis of reviews

REVIEWS

**Review and
price
correlation**

Relationship
between
review and
Price

Relationship
between
review and
location

Price

**Relation
between
price vs
different
factor**

Relation
between
Price vs
location

Data Cleaning

- Dropped all that column which was not useful for doing data analysis like Id, review from last month
- Also we have column name review per month which have 10052 null value. We need to make it zero for our analysis
- Column which is mandatory to take for the observation we Removed all the null values from the column.
- Data with zero null value gives a accurate measures and help data analysts to give right suggestion to airbnb

DATA ANALYSIS

During analysis we found out that most of the rooms are available in price range below 1000. Also there are rooms which are quite expensive ranging above 4000 in Manhattan and Brooklyn as people needs luxury vacations

Using the count function we found that “Entire home/apt” was the maximum number of selected room type. We explore the observation on the basis of the room type connected to every factor like reviews,min night stay ,price etc

People are not price conscious.They give privacy more importance than money
Which divergences the price and review per month relation
Manhattan is the most listed neighbourhood

Data Visualization

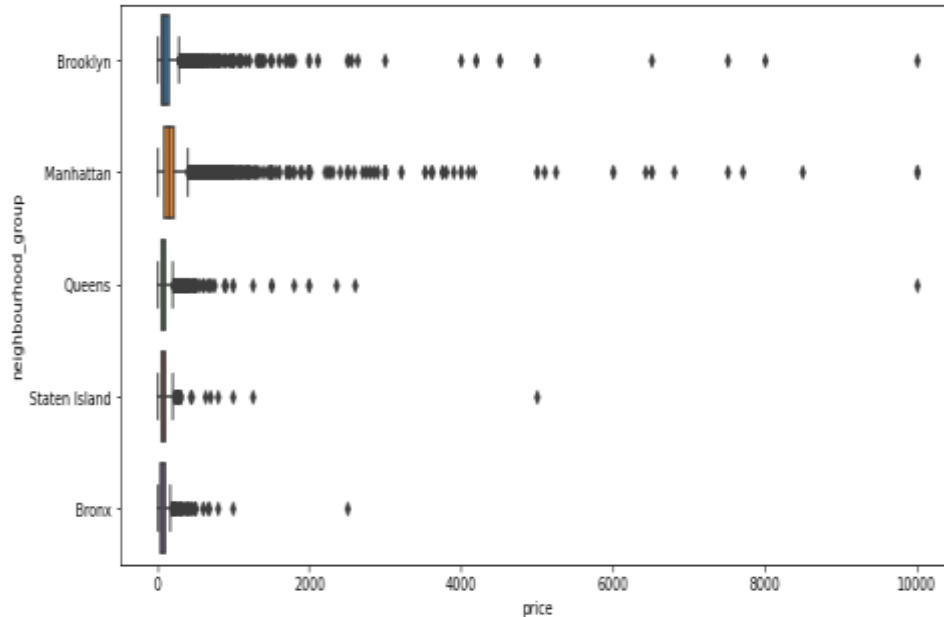
Correlation Plot

	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
host_id	1.000000	0.020199	0.127022	0.015294	-0.017391	-0.140181	0.209752	0.154949	0.203434
latitude	0.020199	1.000000	0.084830	0.033899	0.024893	-0.015357	-0.018719	0.019518	-0.010942
longitude	0.127022	0.084830	1.000000	-0.149954	-0.062772	0.059015	0.138435	-0.114715	0.082669
price	0.015294	0.033899	-0.149954	1.000000	0.042805	-0.047926	-0.050531	0.057462	0.081847
minimum_nights	-0.017391	0.024893	-0.062772	0.042805	1.000000	-0.080080	-0.124870	0.127962	0.144275
number_of_reviews	-0.140181	-0.015357	0.059015	-0.047926	-0.080080	1.000000	0.589311	-0.072385	0.171975
reviews_per_month	0.209752	-0.018719	0.138435	-0.050531	-0.124870	0.589311	1.000000	-0.047322	0.163664
calculated_host_listings_count	0.154949	0.019518	-0.114715	0.057462	0.127962	-0.072385	-0.047322	1.000000	0.225712
availability_365	0.203434	-0.010942	0.082669	0.081847	0.144275	0.171975	0.163664	0.225712	1.000000

Conclusion

Plot shows there is no strong correlation between variables except number reviews and review per month

Data Visualization

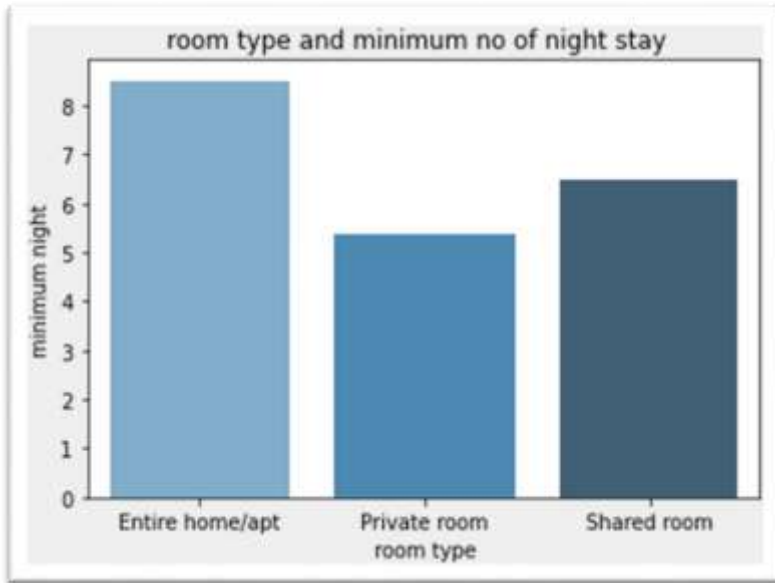


Conclusion

Box plot shows that there has been outliers in the price range. Most of the booking prices are below 1000

Data Visualization

Plot For Room Type and Minimum No of Night Stays

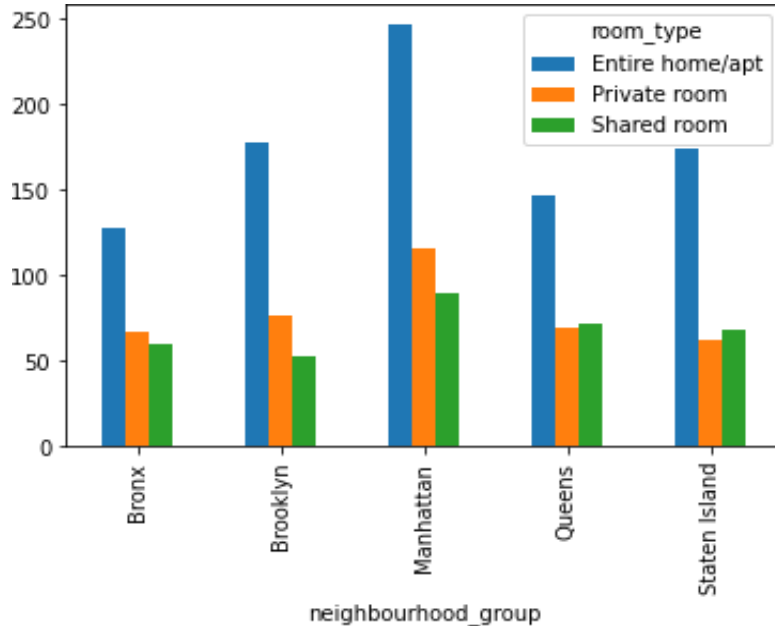


Conclusion

“Entire home/apt was the maximum number of selected room type with 8.51 minimum night stays. that means people like to stay in the privacy in the U.S”

Data Visualization

Plot For Finding The Costly Neighbourhood Group

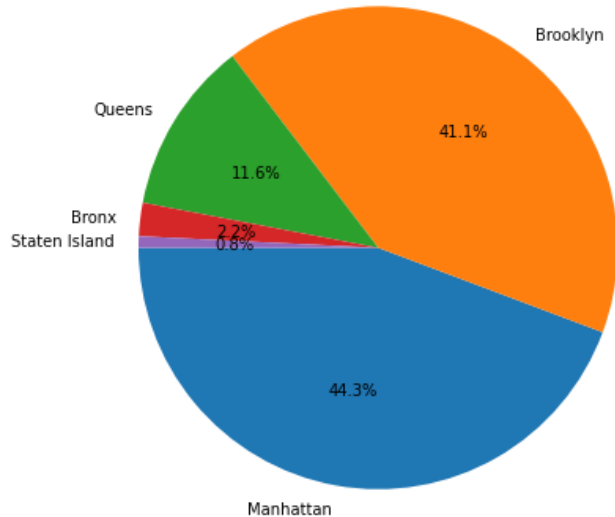


Conclusion

Manhattan is the most expensive for the all the given room type. After checking its location on map we can find it's situated at mid of the country so more denser and also have important financial status, so movement of financial act is happened here.

Data Visualization

Plot For Finding The Costly Neighbourhood Group



Conclusion

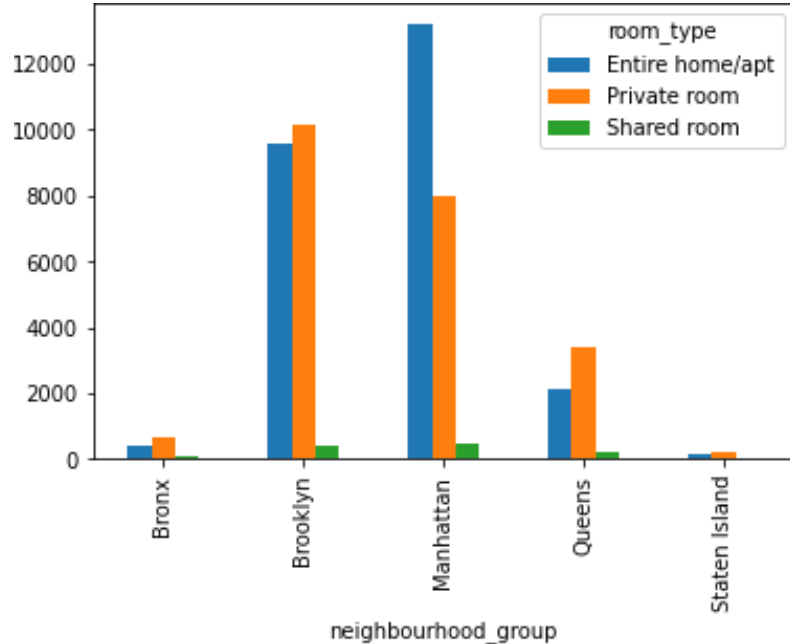
Graph shows the highest 21642 (approx 44%) hosts available in Manhattan

Graph shows the second highest 20080 (approx 41%) hosts available in Brooklyn

It can be clearly seen that majority of hosts are from these two areas alone

Data Visualization

Plot Between Room Type And Neighbourhood Group

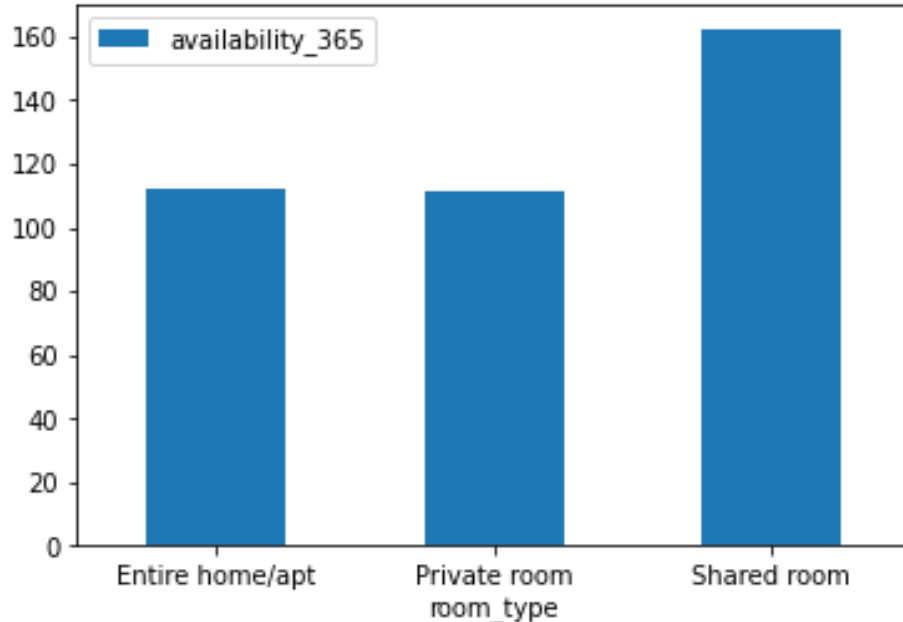


Conclusion

Plot shows that private rooms and entire home/apt are more preferred in Manhattan, Brooklyn and Queens. Shared rooms are least populated, which means most of the people select privacy and safety over the less price.

Data Visualization

Plot Between the room type and it's availability throughout the year

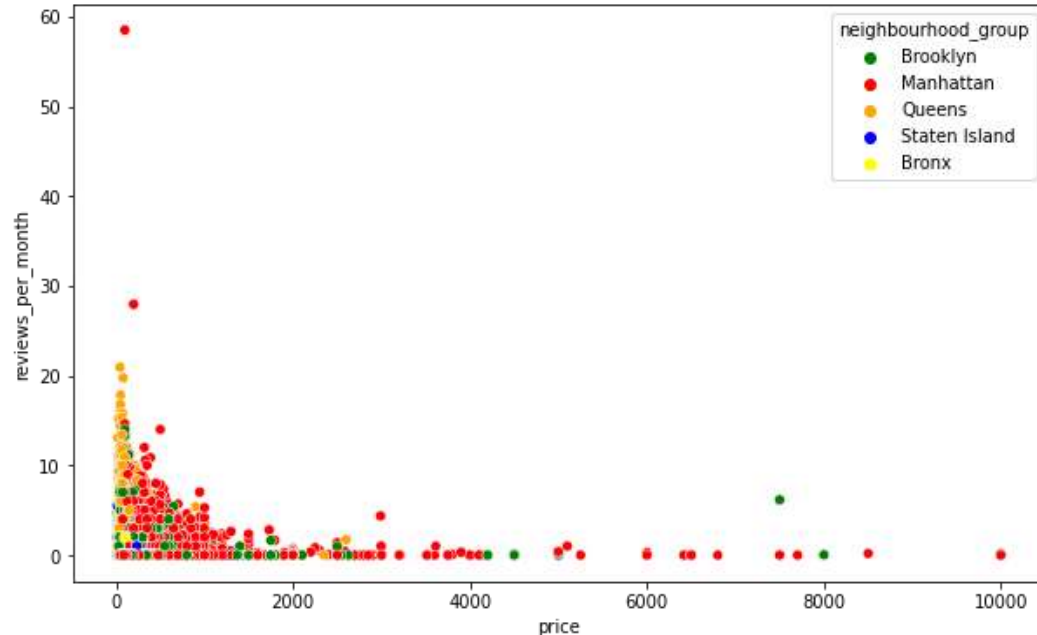


Conclusion

From the given bar plot we can easily understand that shared room type on an average is available for more no of days throughout the year

Data Visualization

Plot For Finding The Correlation Between Price And Review Per Month

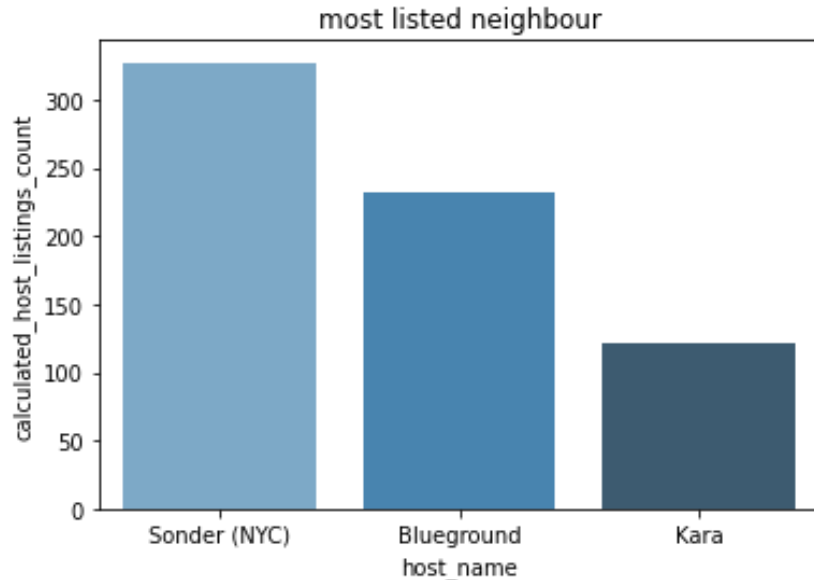


Conclusion

Plot shows that review and price is in a negative correlation in neighbourhood group with each other which means people prefer affordable property and hence given more reviews

Data Visualization

Top Three Busiest Host on BNB

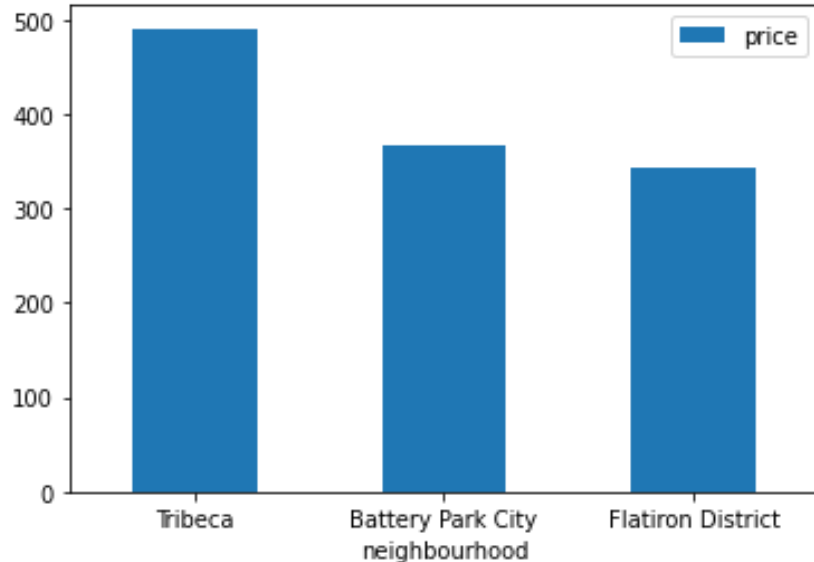


Conclusion

The bar plot shows the busiest host in term of traffic . Sonder is a biggest chain of hotel with the total listing of 327 , all present in Manhattan.

Data Visualization

Plot For Finding The top neighbourhood inside Manhattan

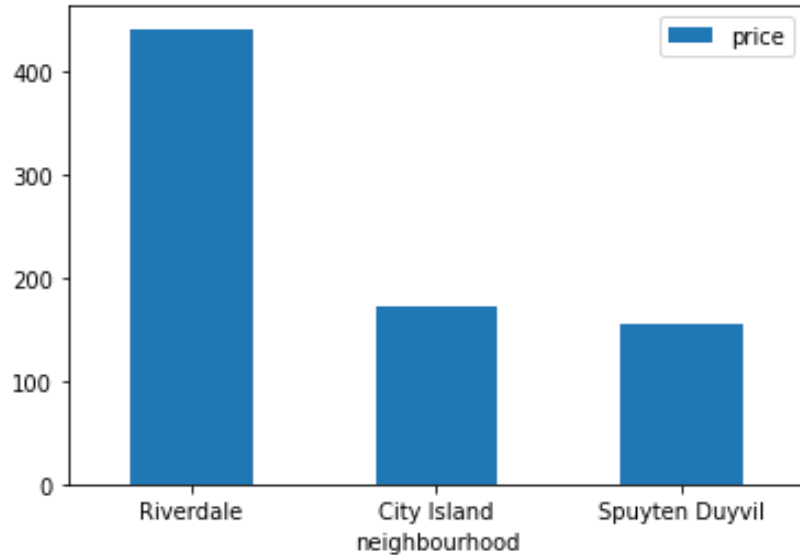


Conclusion

The bar plot between different neighbourhood of the given neighbourhood group based on the price range . which clear us about top three neighbourhood from Manhattan

Data Visualization

Plot For Finding The top neighbourhood inside Bronx

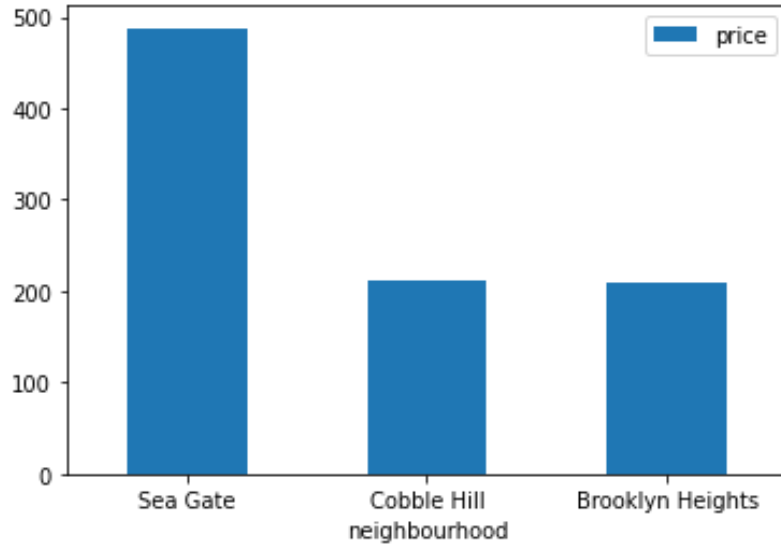


Conclusion

The bar plot between different neighbourhood of the given neighbourhood group base on the price range . which clear us about top three neighbourhood from Bronx

Data Visualization

- Plot For Finding The top neighbourhood inside Brooklyn

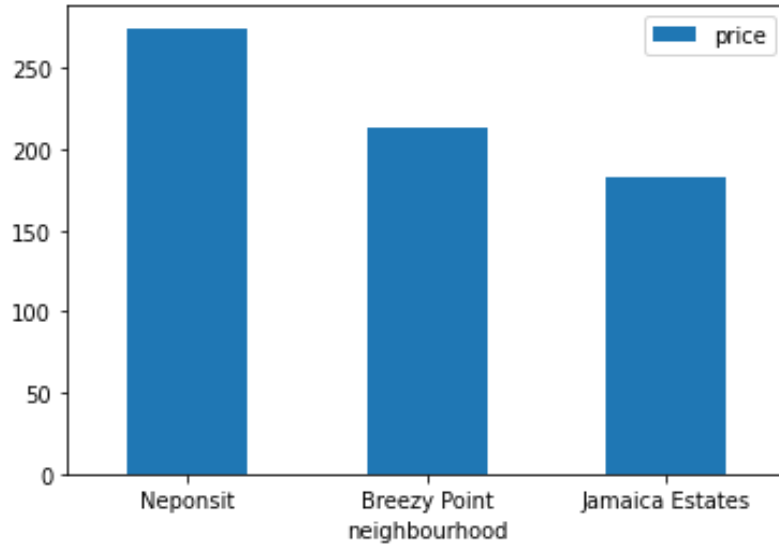


Conclusion

The bar plot between different neighbourhood of the given neighbourhood group base on the price range . which clear us about top three neighbourhood from Brooklyn

Data Visualization

- Plot For Finding The top neighbourhood inside Queens

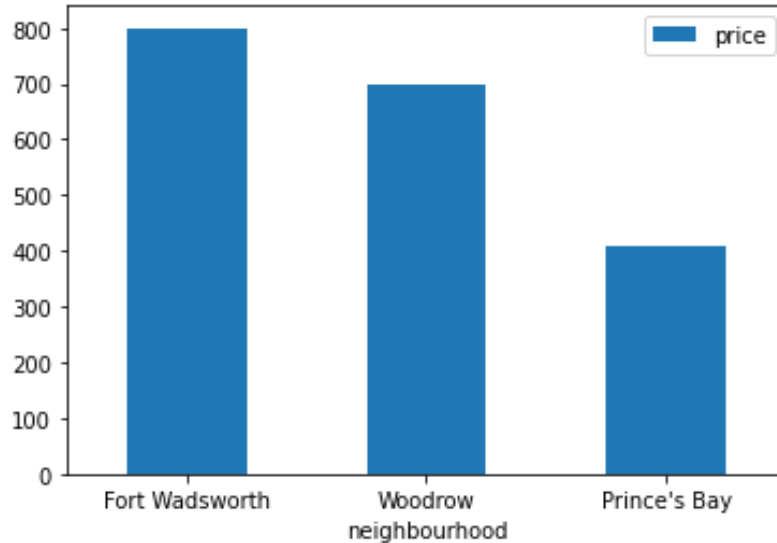


Conclusion

The bar plot between different neighbourhood of the given neighbourhood group base on the price range . which clear us about top three neighbourhood from Queens

Data Visualization

- Plot For Finding The top neighbourhood inside Staten-Island

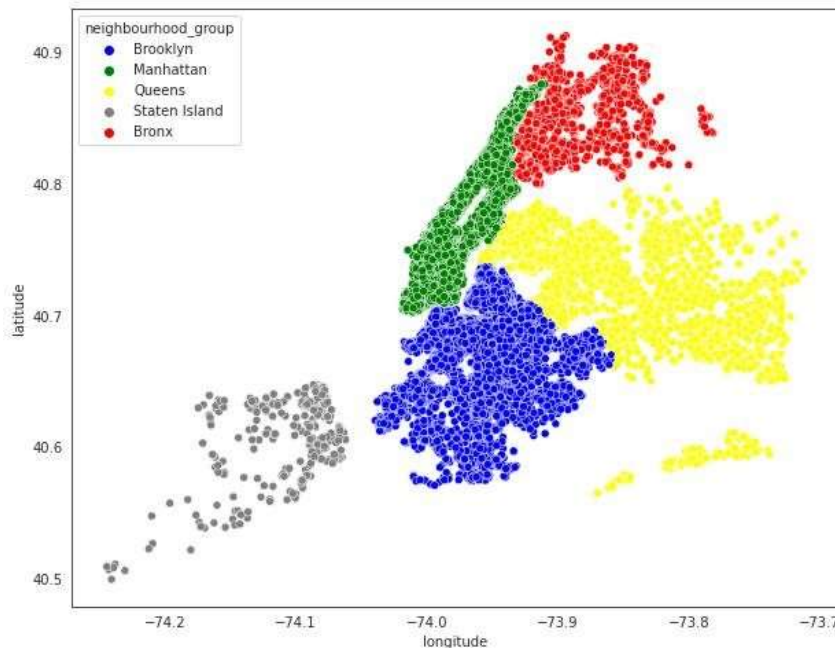


Conclusion

The bar plot between different neighbourhood of the given neighbourhood group base on the price range . which clear us about top three neighbourhood from Staten-Island

Data Visualization

Insights from the longitude and latitude for the no of count.

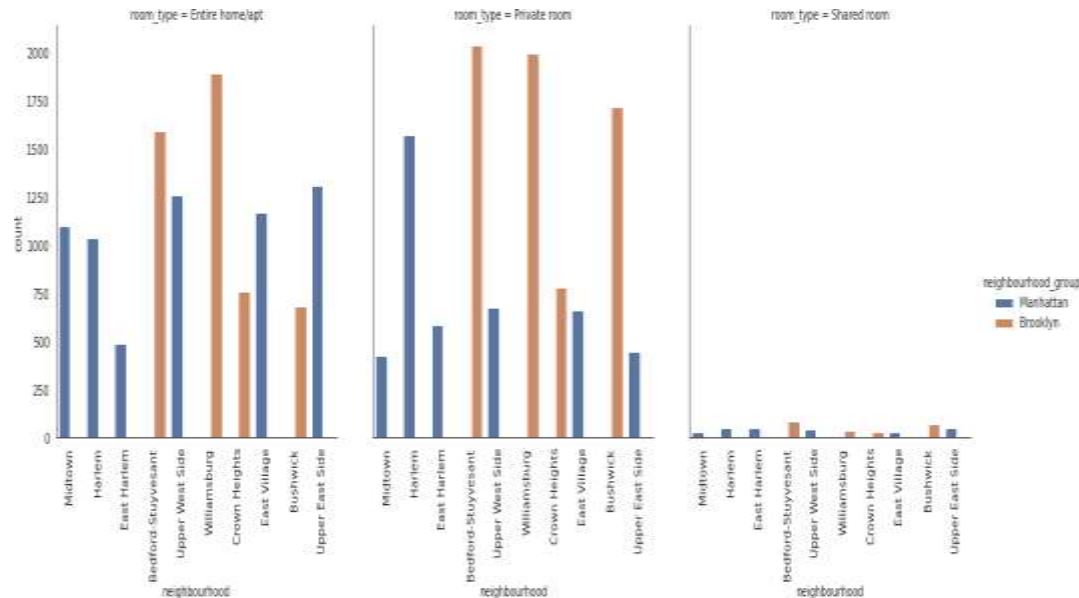


Conclusion

This is area wise Map-plot for the all the airbnb with respect to most no of counting from the different Neighborhood-group and their Neighbourhood. This plot clear us that Manhattan have most no of counts. Because it is located in a centre as per longitude and latitude .its a financial state of U.S so that why it is most demanding.

Data Visualization

Categorical Plot for top ten neighbourhood showing the relation between price and the review per for the top temonth



Conclusion

This plot is for the top Neighborhood from the different Neighborhood-group having the room type selection .plot clear us that most of the customer from the U.S looking for Entire room/apptt

Challenges Faced

- It was very challenging to completely understand the data and to comprehend the relevance of each CSV file
- Since it was our first capstone project so it was difficult to establish proper coordination between team members
- Filtering out the best indicators from indicators to keep for analysis as data contains lot of missing values



Thank You