Exploratory Data Analysis
On
AIRBNB Bookings Analysis

TEAM NAME : Pirates

Team Members

1. Aniket Raj
2. Kushal Kishor
3. Aswathaman R
4. A Balaji
5. Akhilesh Bhardwaj

ABOUT AIRBNB

- AIRBNB is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app.
- Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.
- Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking
- The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia.

CONTENTS

- Data summary
- Dataset cleaning
- EDA on dataset
- conclusion

PROBLEM STATEMENT

To Explore and analyze the data to discover ,

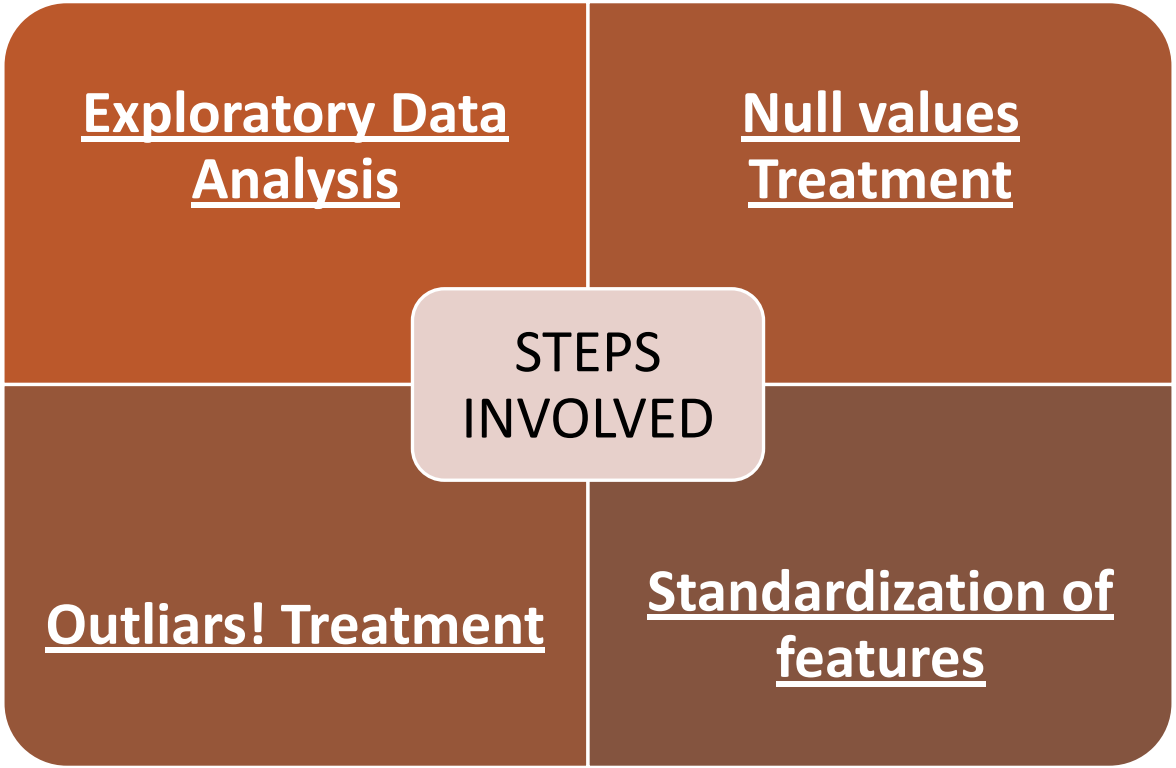
1. Check how many host ids belongs to which neighbourhood group.
2. Which hosts are the busiest and why?
3. Overall descriptive summary of the main variable('Price')
4. Analyze 'Price' variable with 'latitude' variable and 'longitude' variable of the data
5. How many types of room and number of different types of room in each neighborhood group?

PROBLEM STATEMENT

AI

6. Overall descriptive summary of the main variable ('Price') along different types of rooms.
7. What is the average price of different types of rooms in each neighborhood group?
8. Which neighbourhood of neighbourhood group(Bronx,Brooklyn, Manhattan, Queens,Staten Island) got maximum visits?

STEPS INVOLVED IN EDA



DATA SUMMARY

Name of the Dataset	AIRBNB Booking Analysis		
Number of variables	16		
Number of observations	48895		
Duplicate rows	0 (0.0%)		
Total size in memory	6.0 MB		
Missing Data (Columns)	**last_review**	10052	20.55 %
	reviews_per_month	10052	20.55 %
	host_name	21	0.04 %
	name	16	0.03 %

VARIABLE DATA TYPE

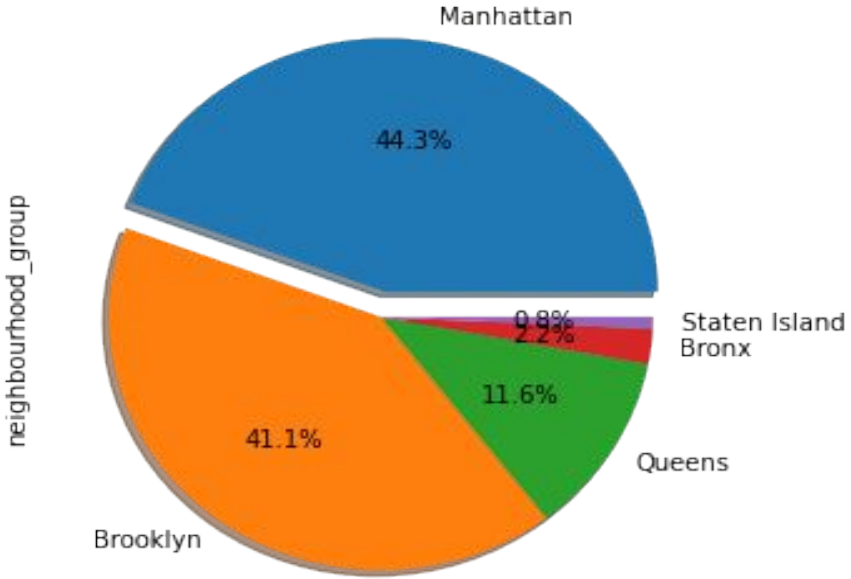
Date Type	Column
Numeric - int64	0 id 2 host_id 9 price 10 minimum_nights 11 number_of_reviews 12 calculated_host_listings_count 13 availability_365
Numeric – float64	6 latitude 7 longitude
String - object	1 name 3 host_name 4 neighbourhood_group 5 neighbourhood 8 room_type

CLEANING DATASET

Columns	Total	Percent
last_review	10052	0.205583
reviews_per_month	10052	0.205583
host_name	21	0.000429
name	16	0.000327

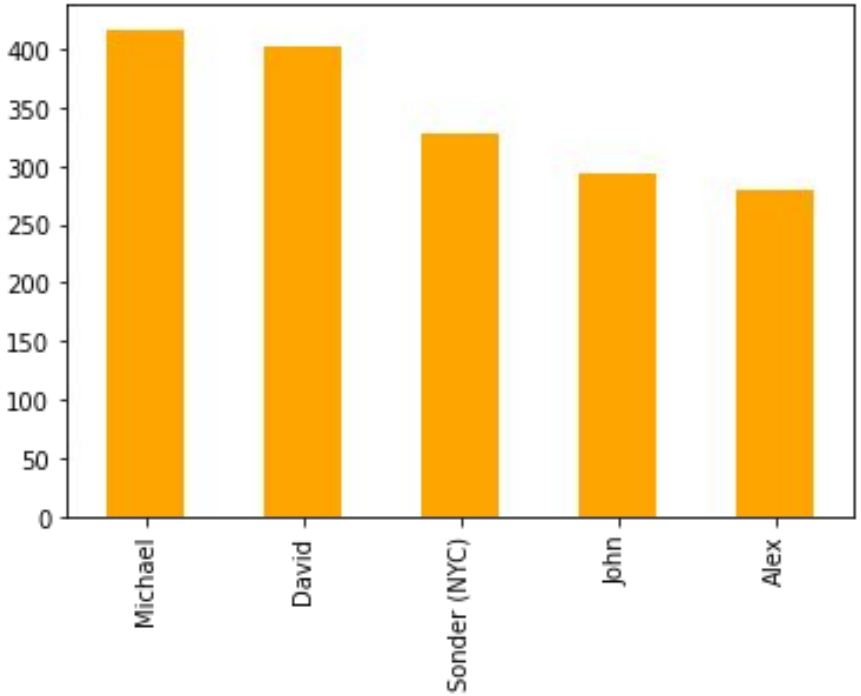
To handle missing data, variables with missing data has been removed, except the variable 'host_name' and 'name'.

Check how many host ids belongs to which neighbourhood group.



This data shows the neighborhood group of the NYC and also the host id belongs to which neighborhood group

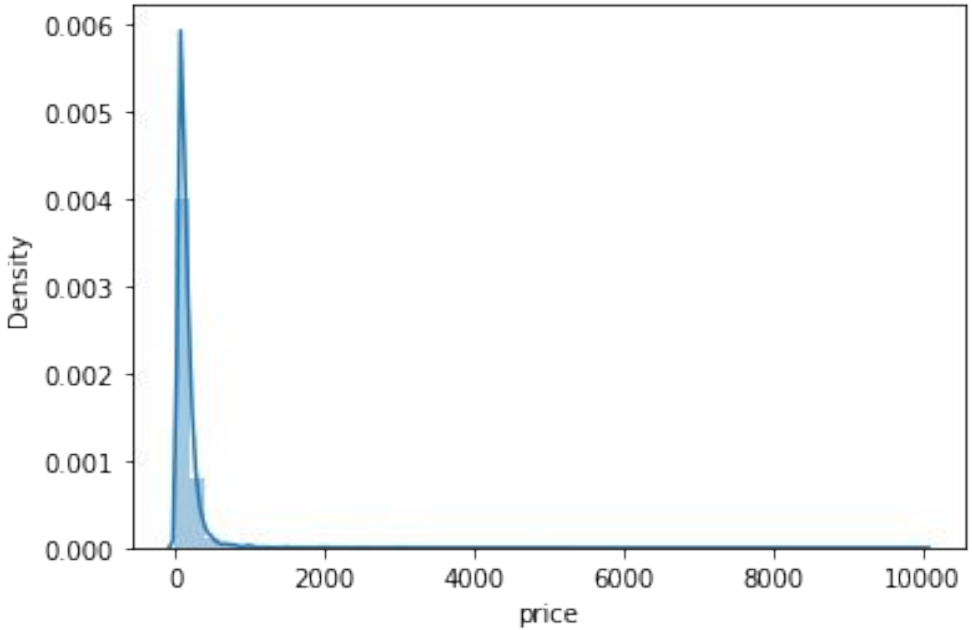
Which hosts are the busiest and why?



This data shows that Michael, David etc. are the busiest host

summary of the main variable('Price')

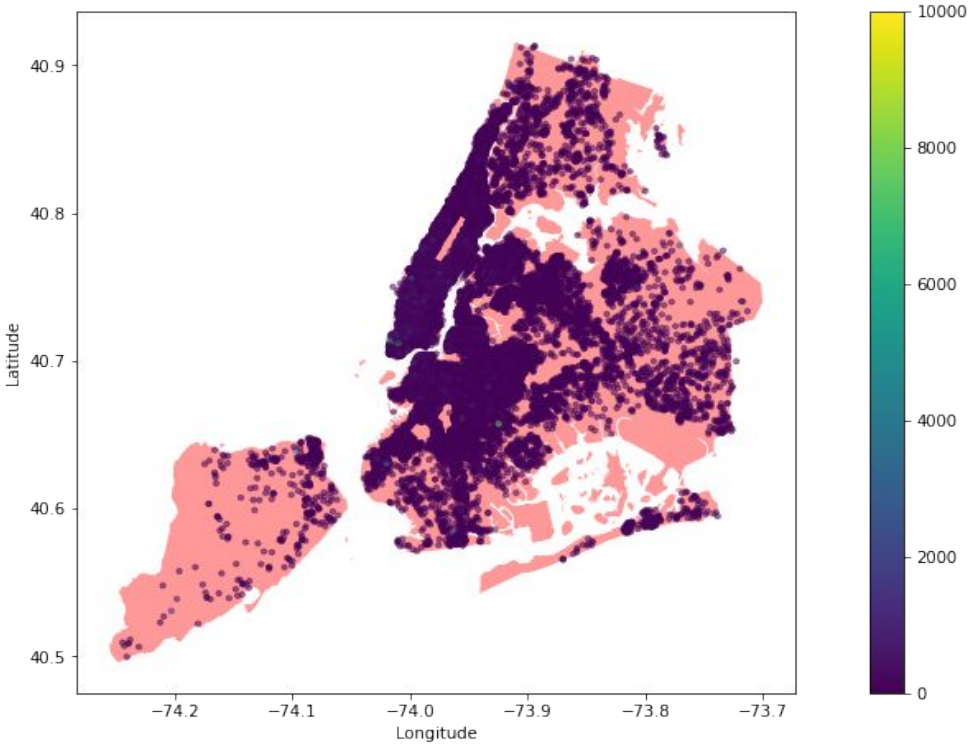
count	48895.000000
mean	152.720687
std	240.154170
min	10.000000
25%	69.000000
50%	106.000000
75%	175.000000
max	10000.000000



This data shows that more bookings are done at the price range of 10 - 500

Analyze 'Price' variable with 'latitude' variable and 'longitude' variable of the data

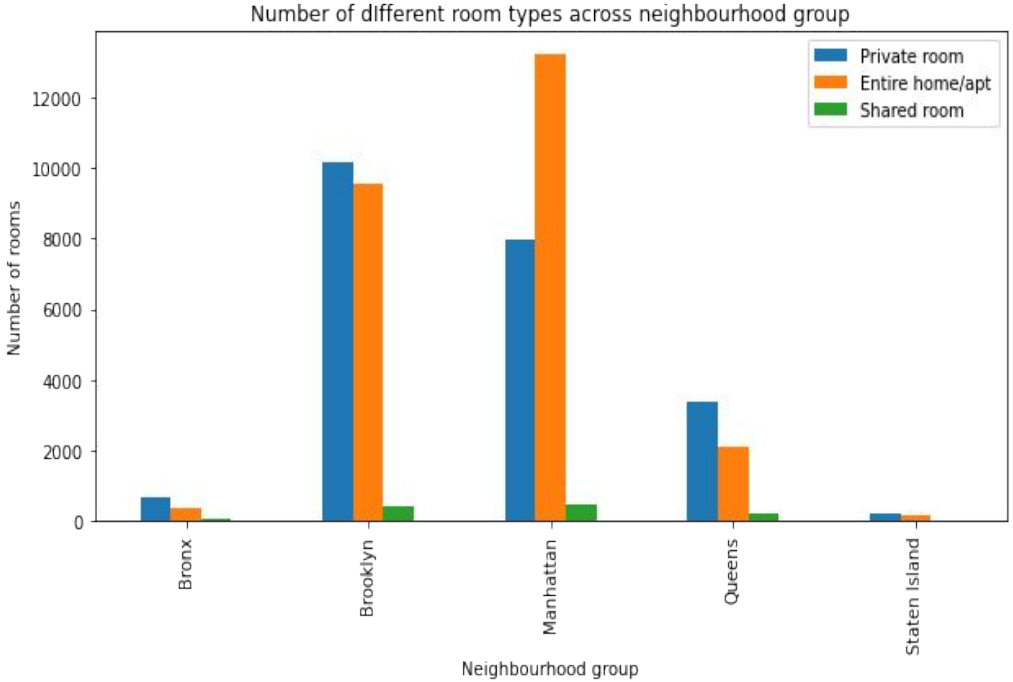
This data shows the geographic location of price variable.



How many types of room and number of different types of room in each neighborhood group?

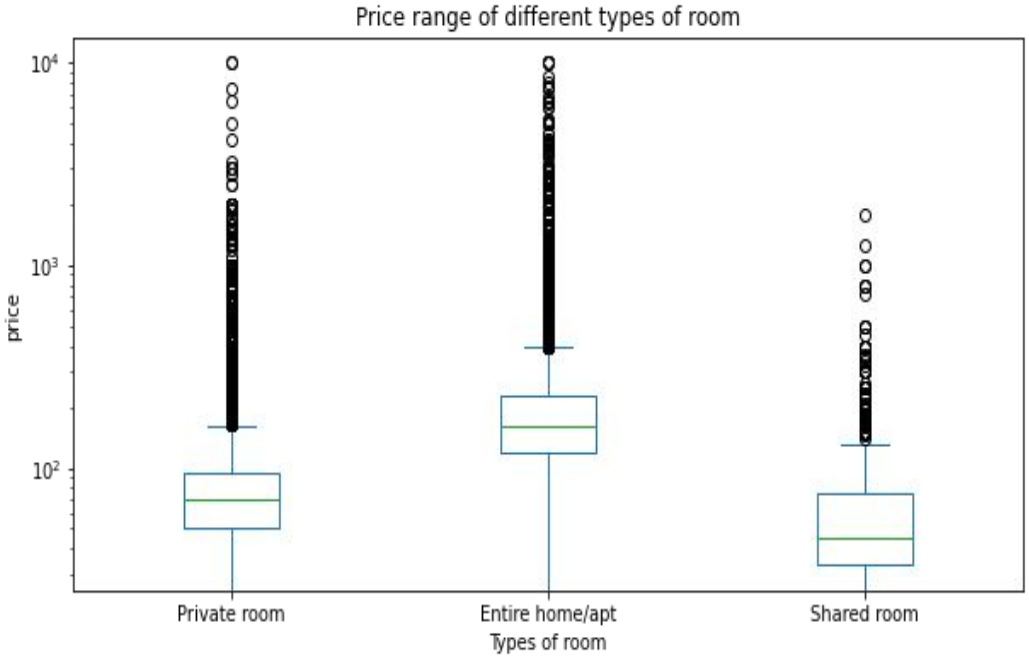
From the data , we can able to find that

- Brooklyn has more private room as compared to others
- Manhattan has more number of entire home/apartments
- Shared room is less in all the neighbourhood group.

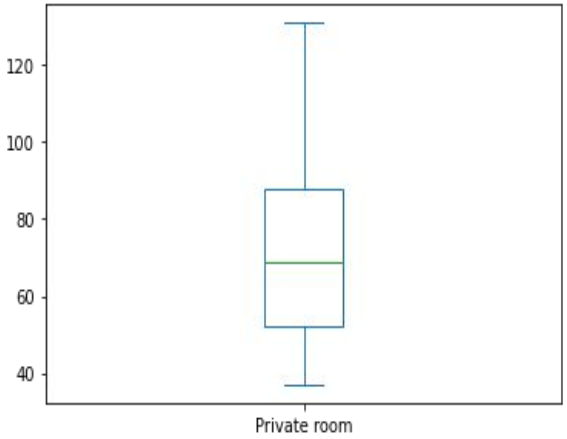


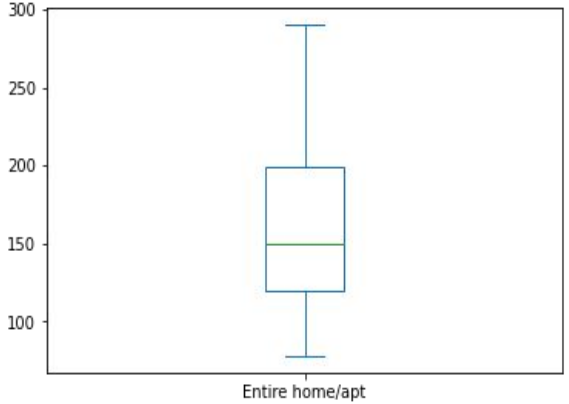
Summary of the main variable('Price') along different types of rooms

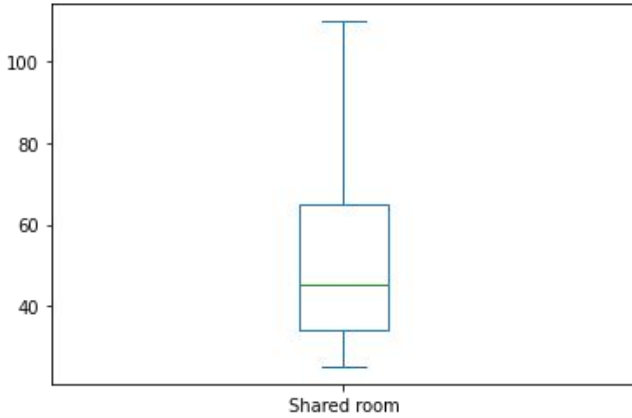
This data shows the overall descriptive summary of the main variable('Price') along different types of rooms.



OUTLIERS





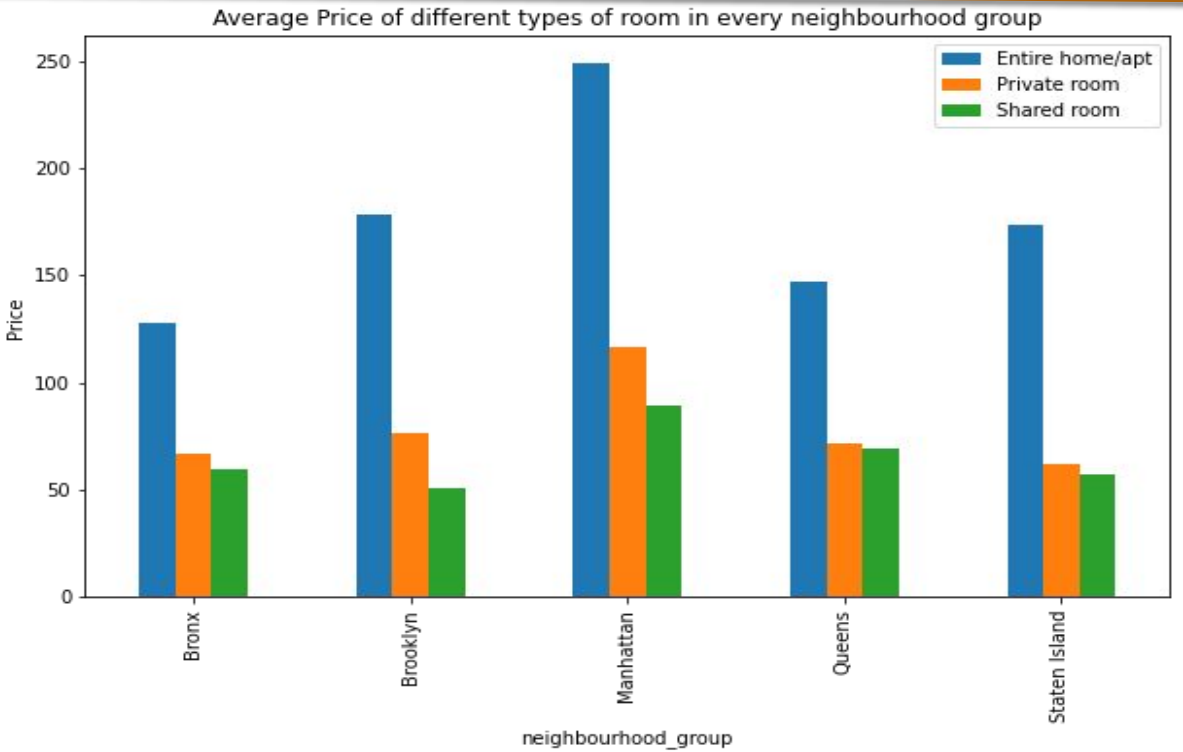


This analysis shows the standard deviation of 'Price' and a set of

Box plots for different room types.

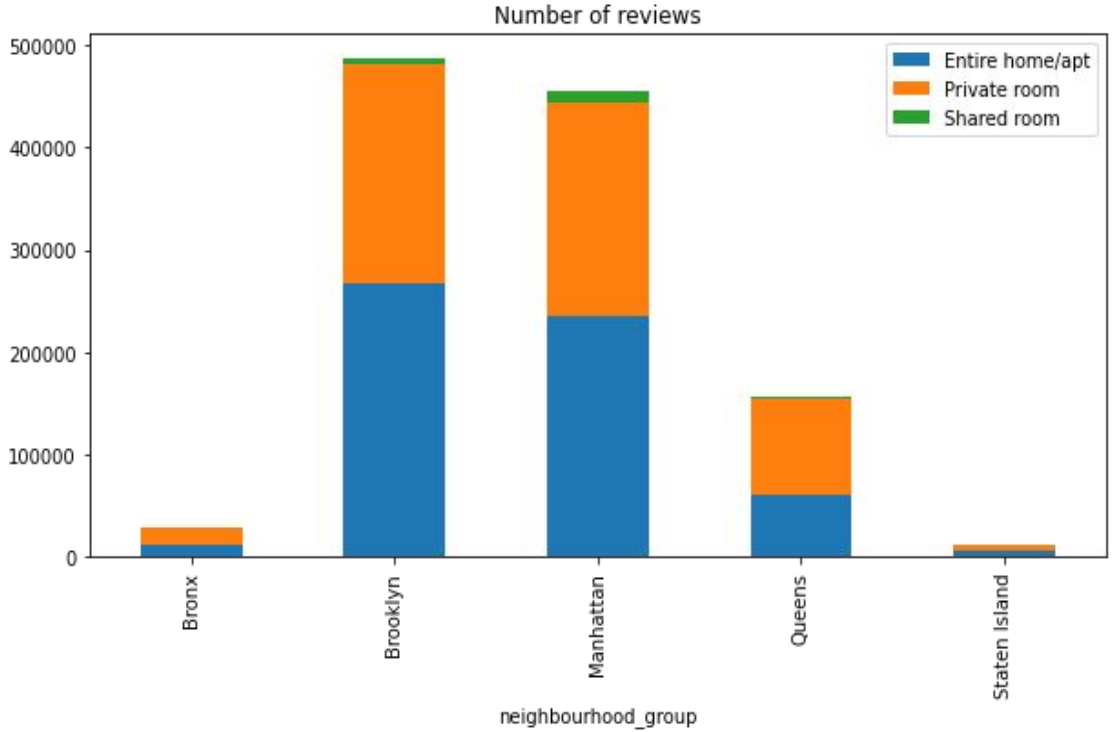
What are the average price of different types of rooms in each neighborhood group?

This data shows that average price of different types of room in every neighbourhood group



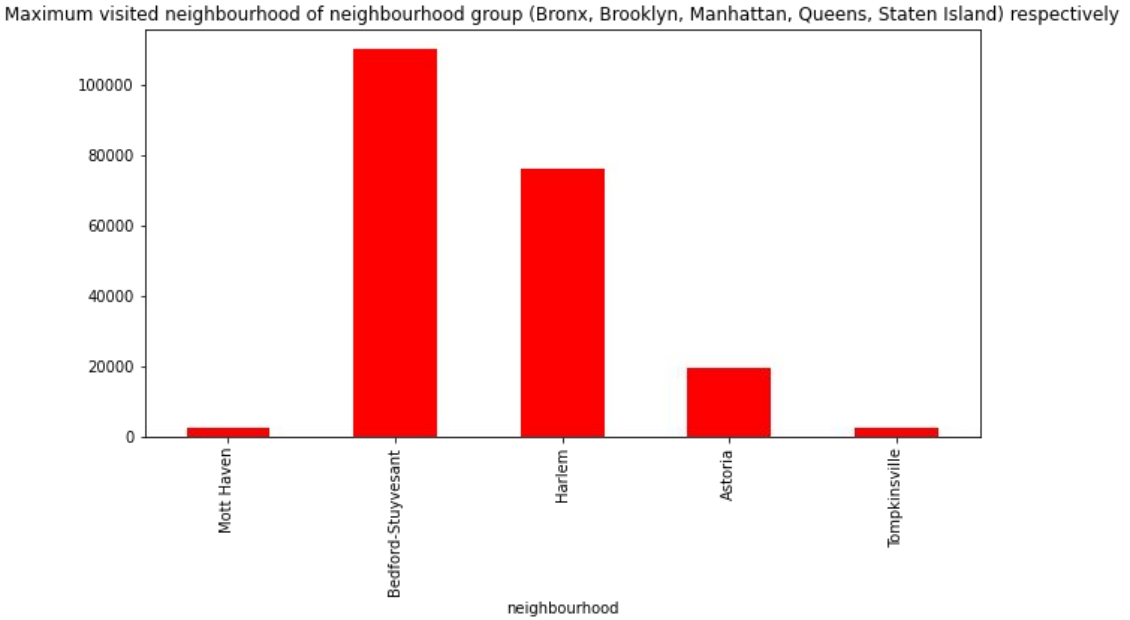
What are the number of visits in different types of rooms in each neighbourhood group?

This data shows the number of visits in different types of rooms in each neighborhood group.



which neighbourhood of neighbourhood group(Bronx,Brooklyn, Manhattan, Queens,Staten Island) got maximum visits?

This data shows that neighborhood which got maximum visits in each neighbourhood_group based on the number of reviews.



CONCLUSION

Throughout this dataset we put in practice many of the strategies and We philosophized about the variables, we analyzed 'Price' alone and with the most correlated variables, we dealt with missing data and outliers, we tested some of the fundamental statistical assumptions and we even transformed categorical variables into dummy variables.