

**Description:**

**●** Airbnb is an online marketplace that connects people who want to rent out their homes with people looking for accommodations in specific locales.

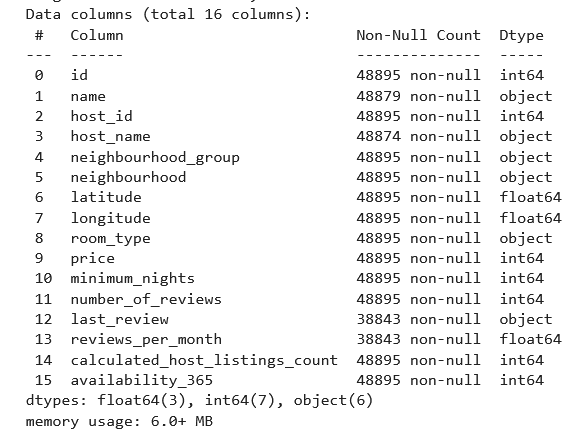
**●**I have performed Exploratory Data Analysis (EDA) on the dataset to communicate insights from data and answer the question about the bussiness growth and provide solutions for other challenges as well.

**●**For conveying insights and for easy and clear understanding I have used multiple visualisation.

**Why Airbnb Dataset for EDA project:**

**●**Airbnb dataset has huge amount of data ,to handle and perform various operation on such data is challenge in itself.

**●**The data contains 48895 rows and 16 columns containg data in various data type like object,integer,float,many nulls and duplicate values too.



**About dataset:**

Our original dataset i.e airbnb\_df has:

1) 0.03% and 0.02% null values in names and host names column.

2) 20% of null values in last review and reviews per month columns.

3) Also the dataset has no duplicated values.

#### ****Business Objective :****

"To create a world where anyone can belong anywhere.”

**Problem Statement :**

Here are some statements of which we will try to communicate insights:

1)The type of rooms and neighbourhood people prefers.

2) Distribution of preferred room over the price

3)Host listing distribution in neighbourhood groups and in each neighbourhood group.

4)Average booking price and normalization.

5)Visualization on assumed revenue and growth over the years and visualization on assumed revenue and growth over the months of last three years.

6)Visualization on Co-relation between the variables in data.

**Python libraries imported:**

**●**Pandas

**●**Numpy

**●**Seaborn

**●**Matplotlib

**Operations performed:**

**●**Data Cleaning

**●**Data Wrangling,Manipulation,normalization where needed.

**●**Data Visualization

### What all manipulations are performed:

Manipulations done in the original dataset i.e airbnb\_df are mentiond below:

**●**Droped corresponding subset of the null values of columns named "names" and "host\_names" which will make no impact in the originality of data.

**●**Removed whole columns named "last\_review" and "reviews\_per\_month" because droping the corresponding values will dilute the orignality of data and the insights will be mis-leading as a big chunk of data will be removed.

After doing required manipulations,I have assigned a variable "man\_airbnb\_df" to the manipulated dataset that stands for manipulated airbnb dataframe.

**Insights,problem statments and solutions:**

1)The type of rooms and neighbourhood people prefers.

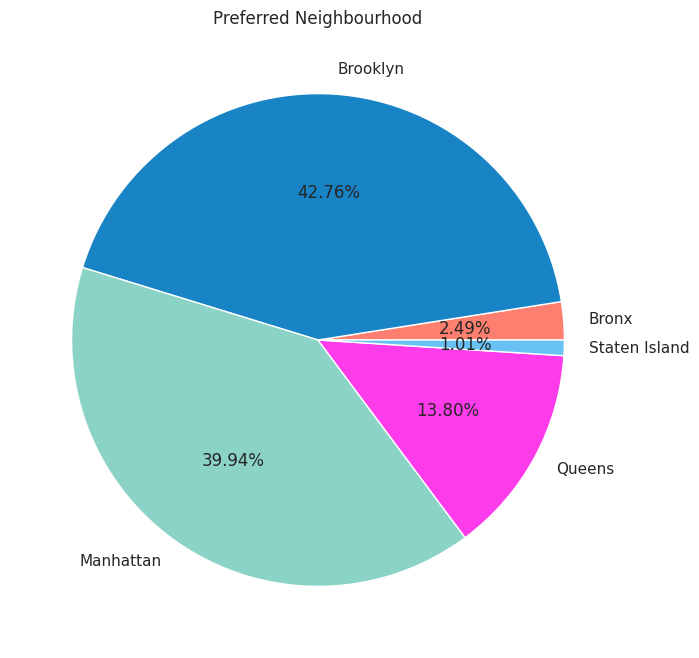
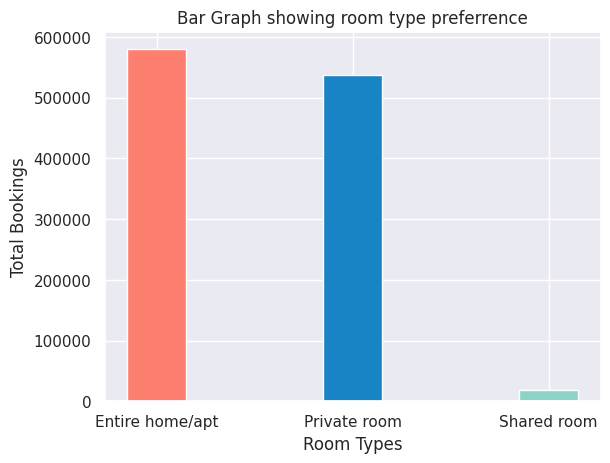


Chart 1.2

Chart 1.1

Entire home/apt are most preferred droom type followed by private rooms and shared rooms are least preferred.

The reason behind the stunted growth of shared room should be known and addressed properly.

Privacy could be the reason.

2)Distribution of preferred room over the price

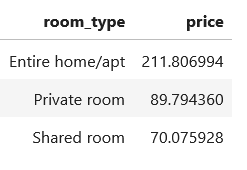
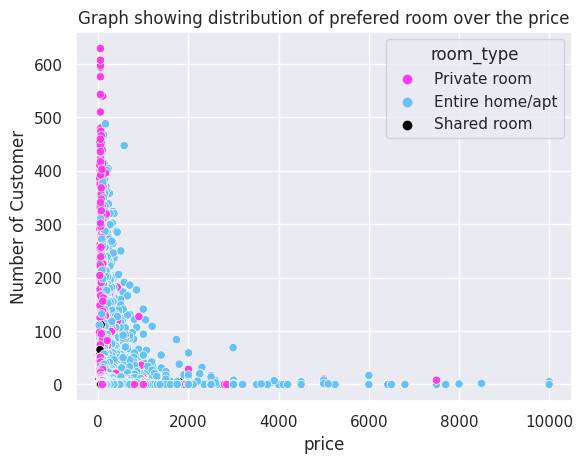


Chart 2.1

**Mean Price of Rooms**

Majority are finding the price range 0 to 2000 comfertably affordable, but it should be noted that other range are also keeping the booking register ringing.

Price range 0 to 2000 is the most preferred price range irrespective of room type,even the least prefered i.e shared room also lies in the same price range. So adding more room(private and entire home) in this range can be benificial form bussiness point of view.And also try to find the reason why shared room are performing consistently low.

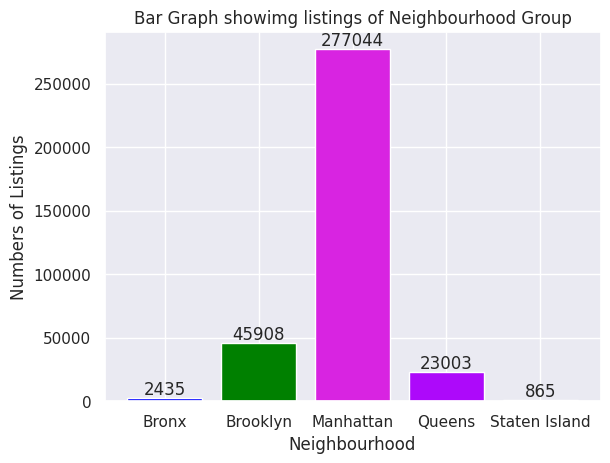
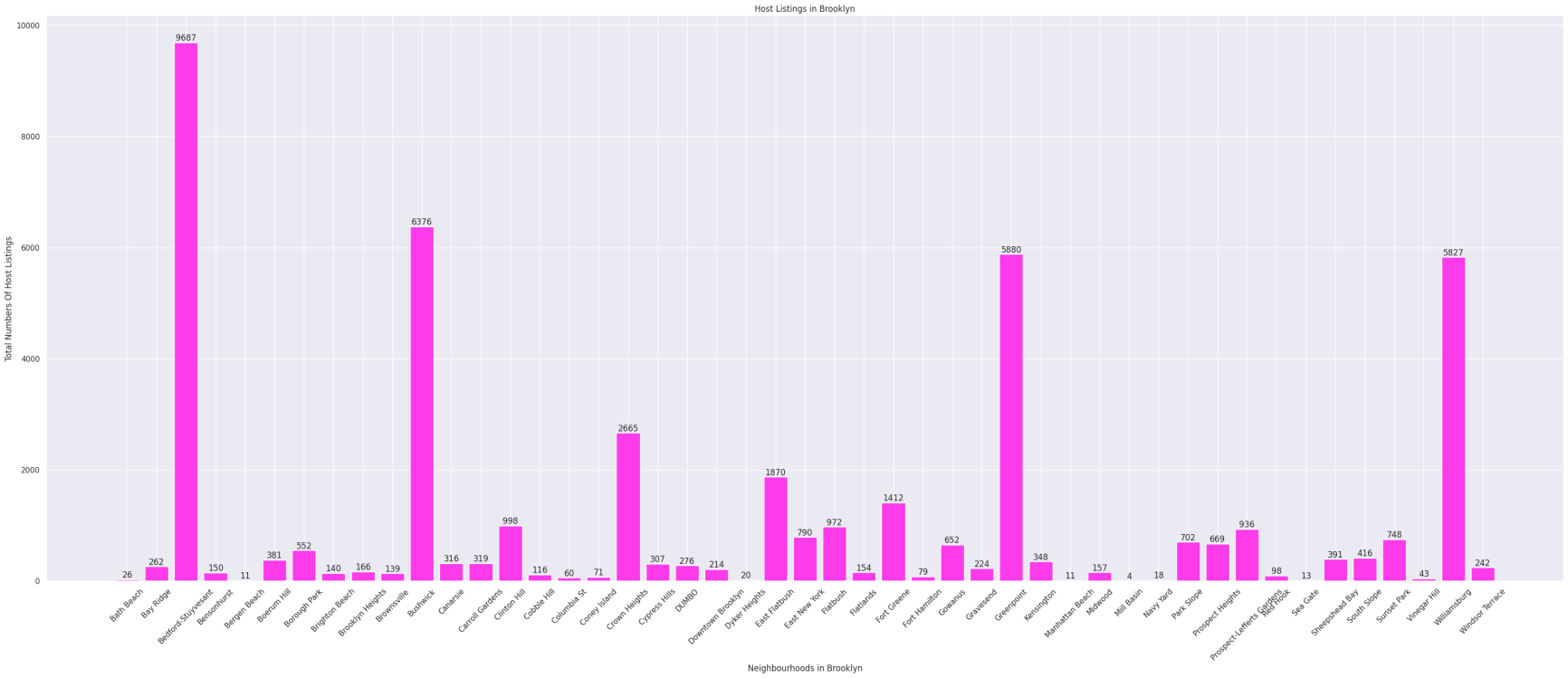
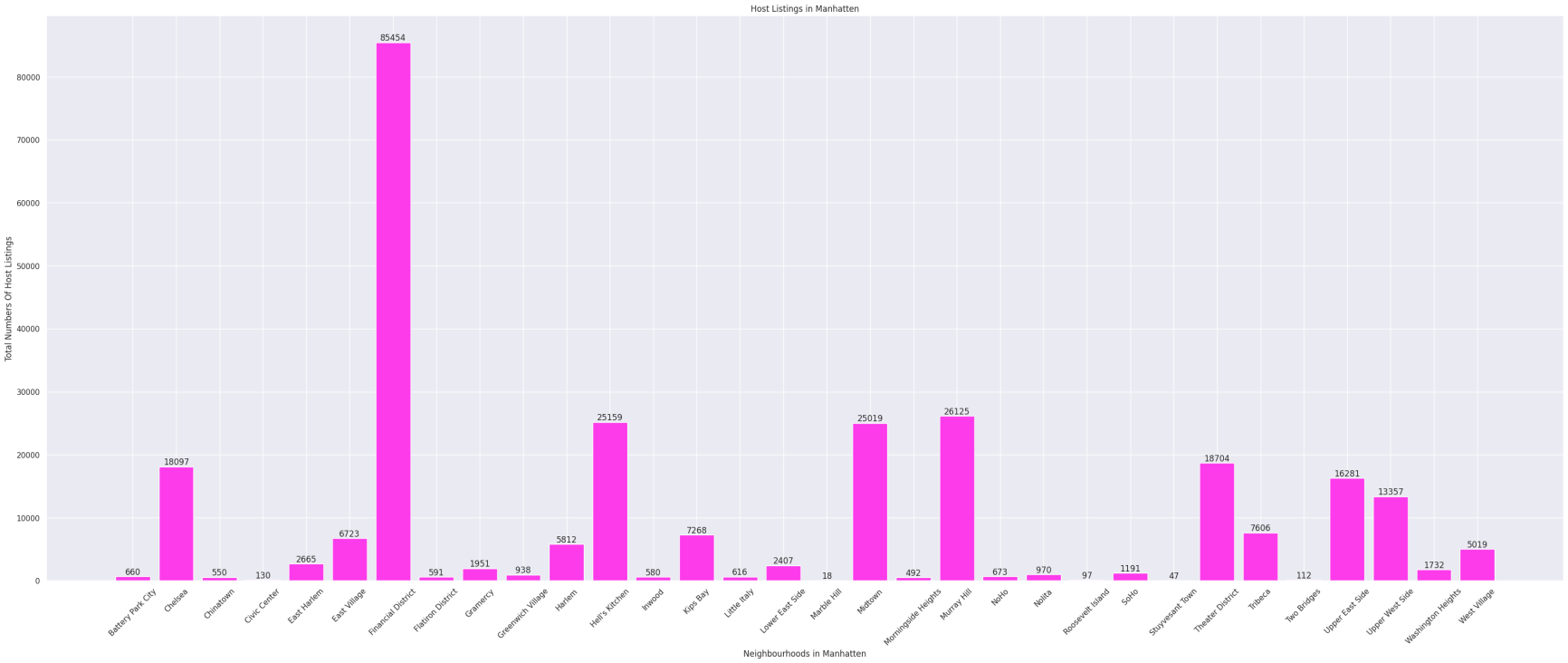
3)Listings in Neighbourhoods and in each group: 

Chart 3.1

It can be seen that Manhatten has much higher listings than other four group combine. Staten island and bronx are last names again. 

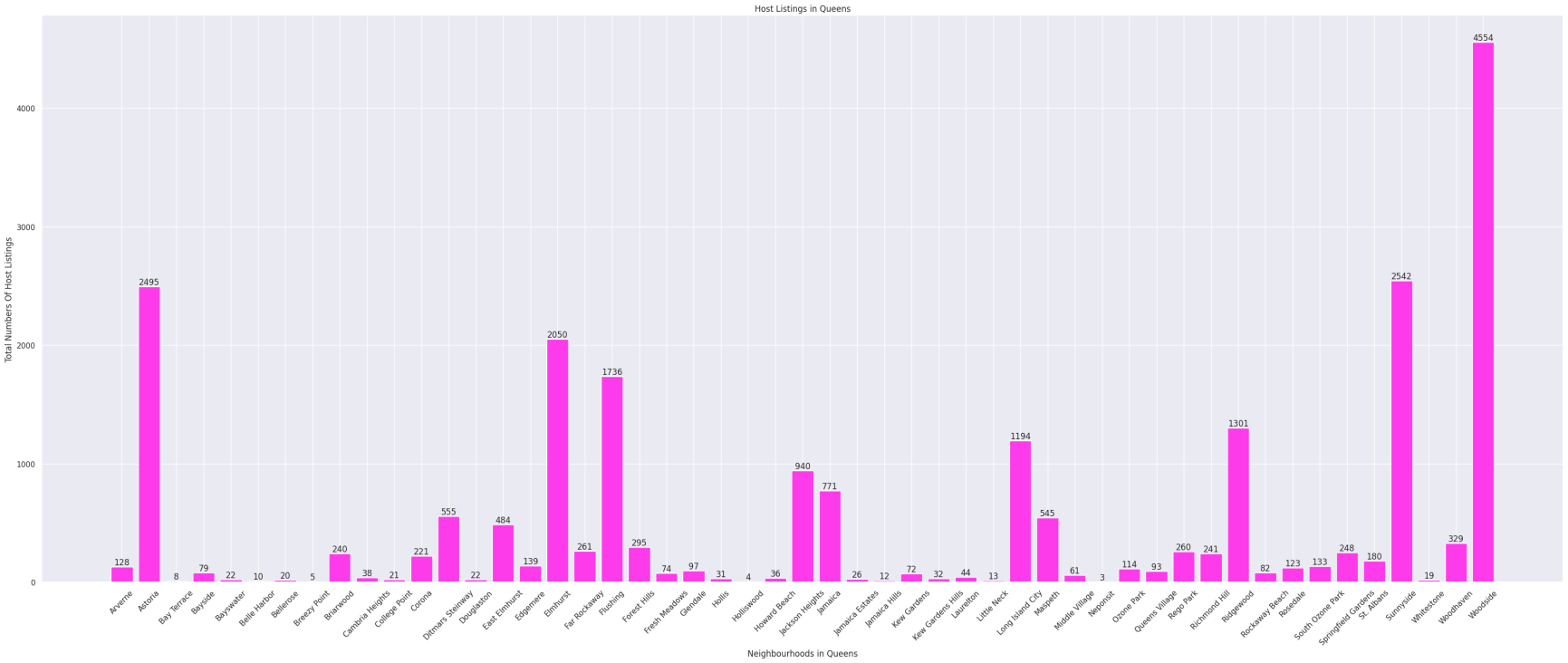


Chart3.4

Chart3.3

Chart3.2

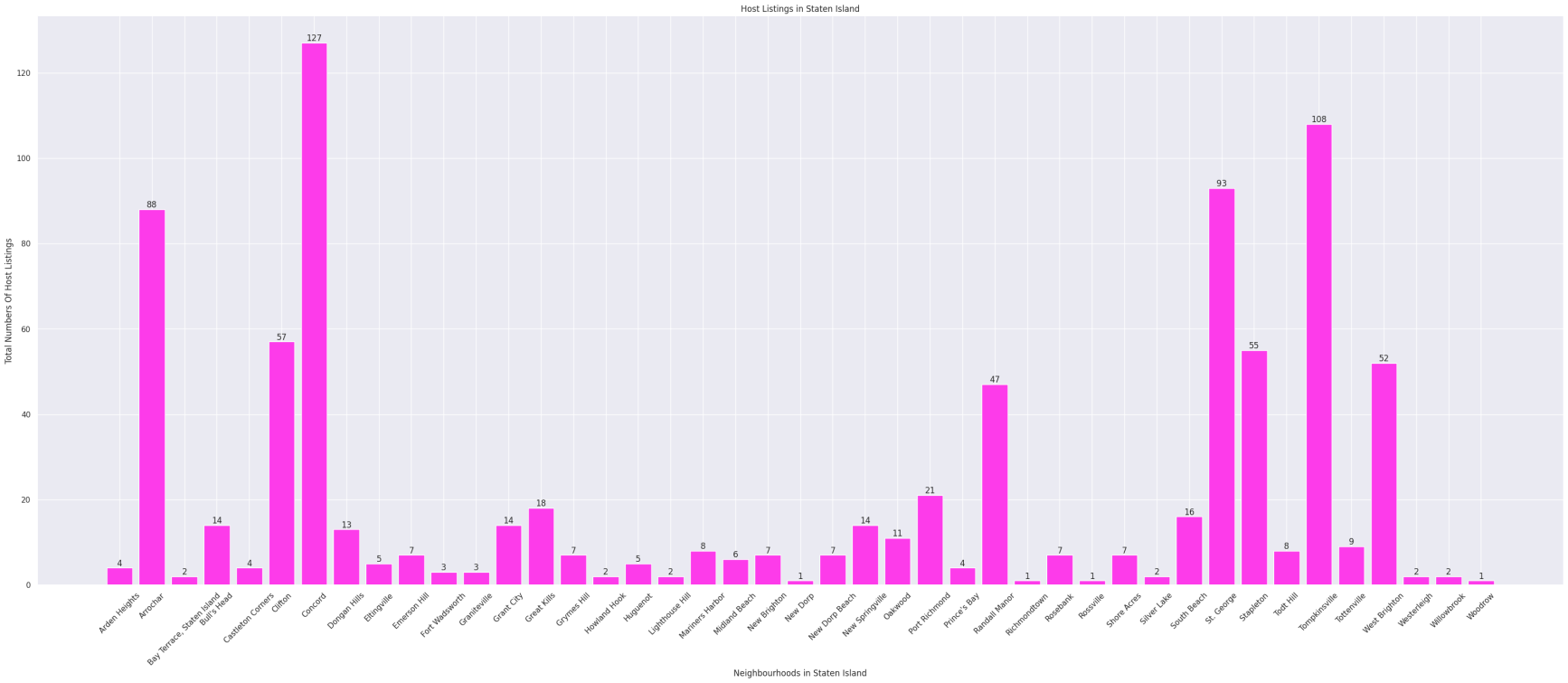
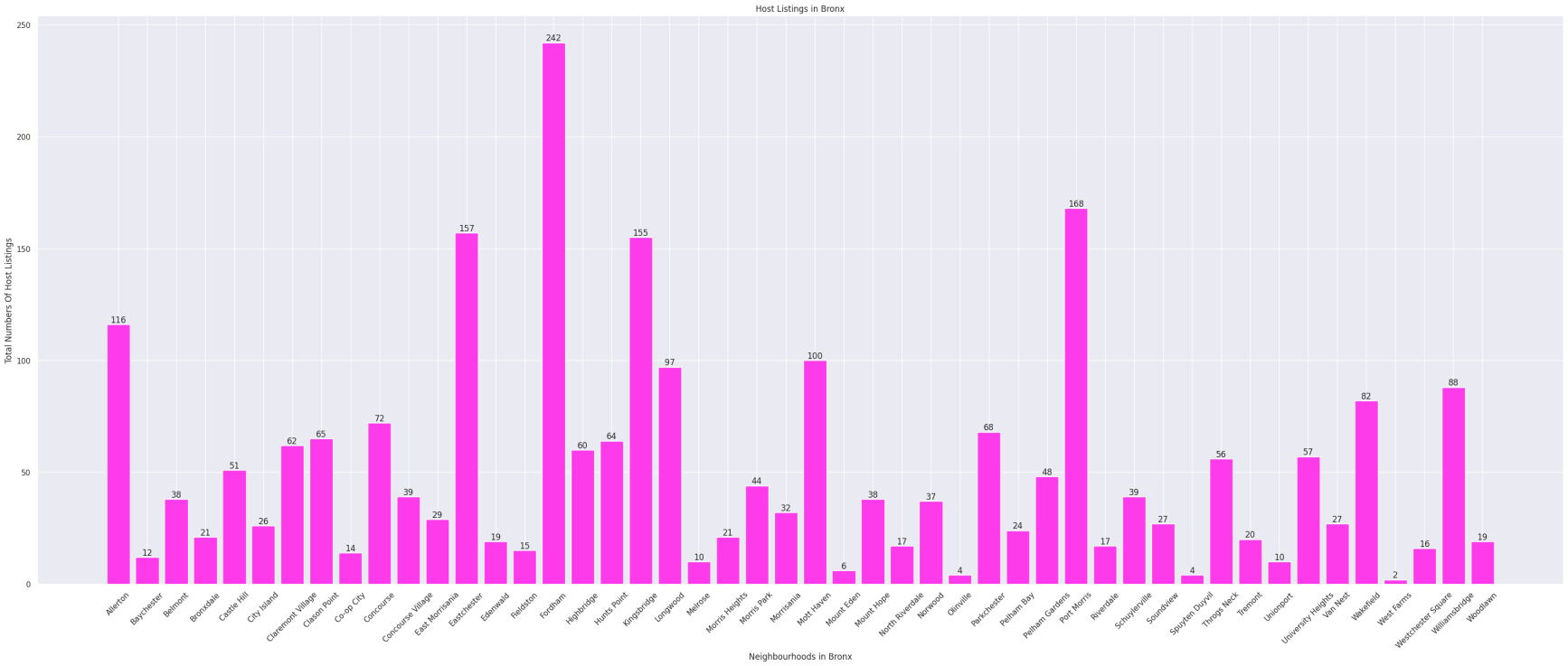


Chart3.6

Chart3.5

I found that the host listings in Manhatten is mainly concentrated in it's neighbourhood named as "financial District". Though the State Island and Queens has low listings they are showing healthy distribution. Brooklyn has 4 or 5 neighbourhood showing listings near and above 6000 and rest are having average 1000 approximetly.

4)Average booking price over the years**:**

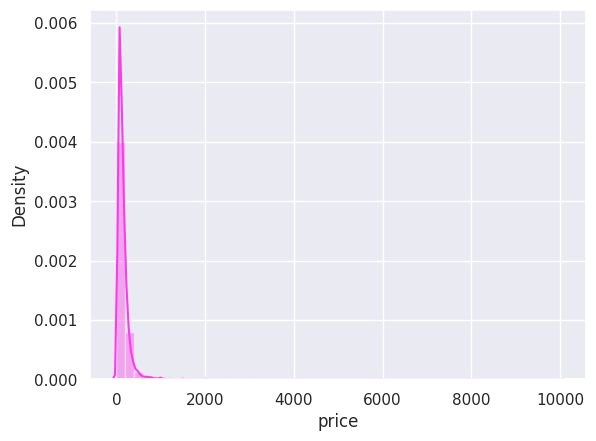
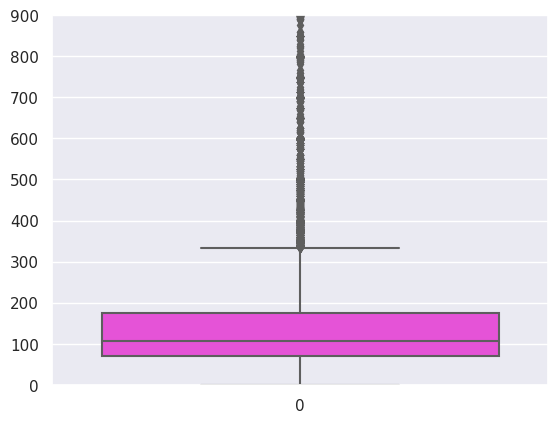
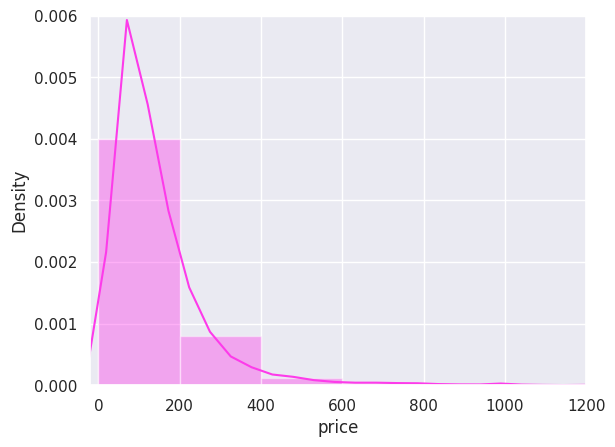


Chart 4.1

Before Normalization

****

Box Plot showing outliers

Chart 4.3

Chart 4.2

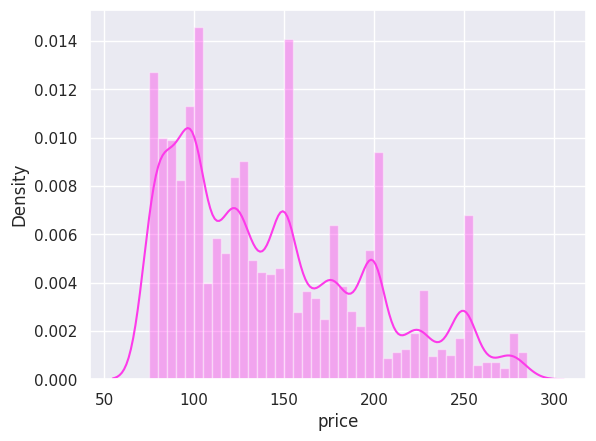
****

Chart 4.4

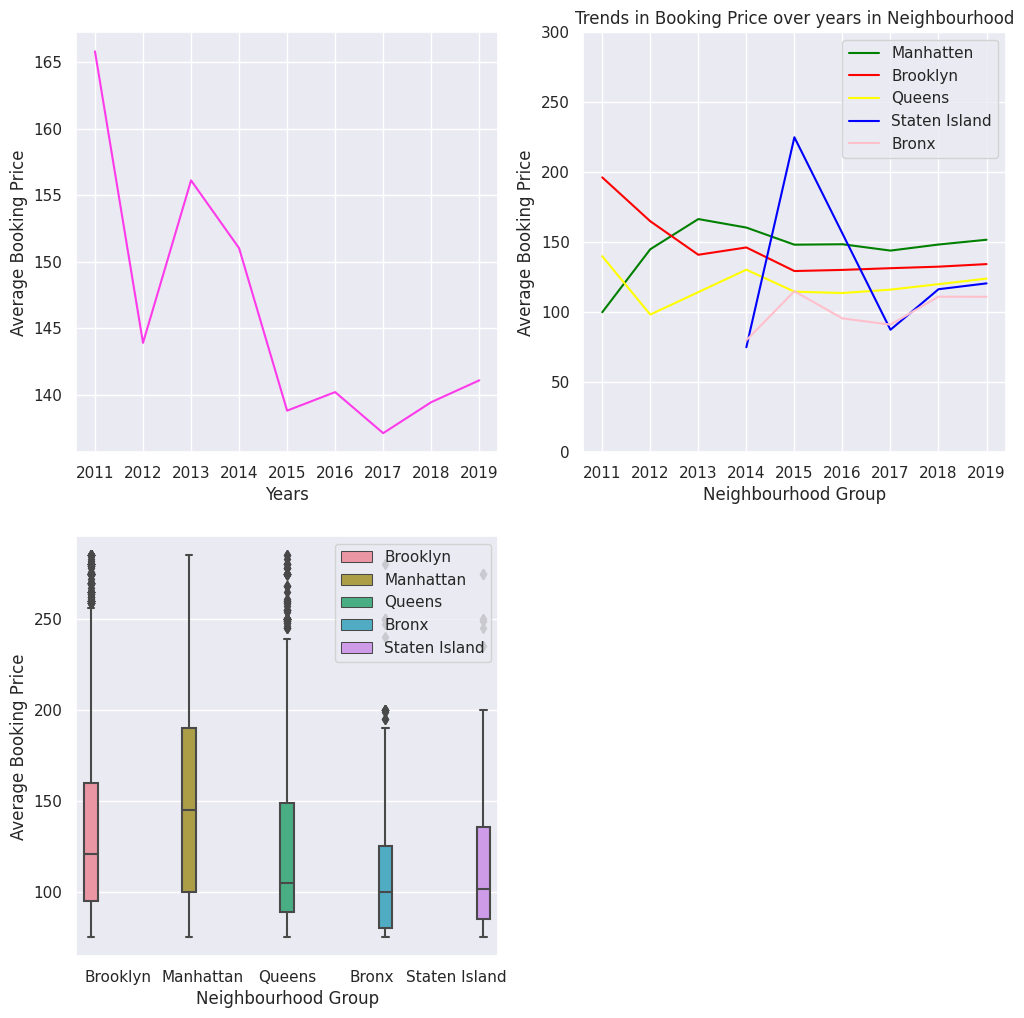


Chart 4.5

The average booking price has declined from 2011 to 2019,in which Bronx has minimum average booking price,Staten Island is having huge decline in average booking price,where as neighbourhood group like Manhatten,Queens,Brooklyn are stable or rock steady.

Chart 4.1

5)Visualization on assumed revenue and growth over the years

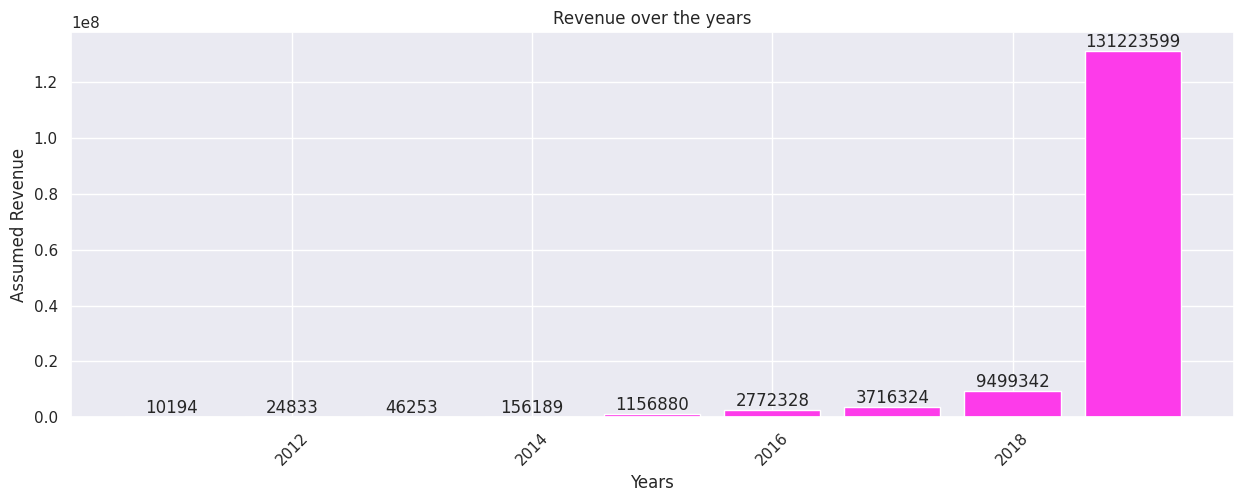


Chart 5.1

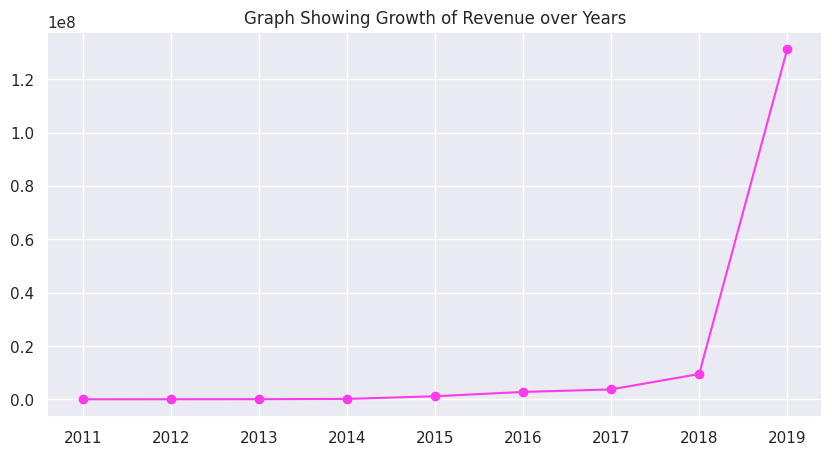


Chart 5.2

The only insight from the plot is that the revenue(assumed)is growing,and also showing the exponential growth in last year.This may be because the people habit of reviewing has been increasing.

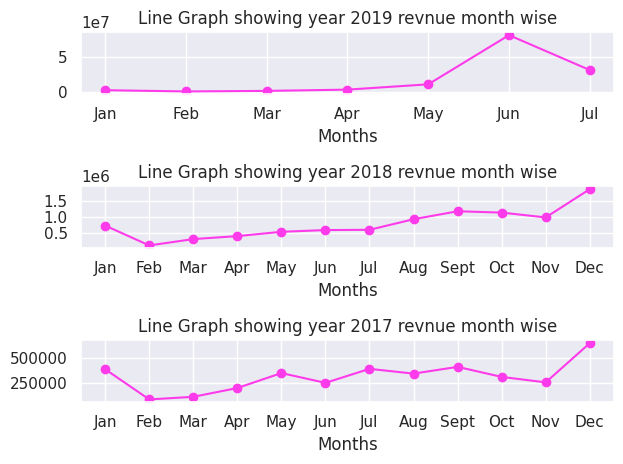


Chart 5.3

1)For year 2017,18,19:

The month of April-May are showing growth in all three years,summer vacation might be the reason.

Jan and Februray are showing the slight fall,this might not be the fall but because there was higher number of bookings in last month the graph is showing the fall.(Base Effect)

2)For years 2017,18(for 2019 we don't have complete data):

Months of November and December are also showing the growth ,major festivals like Christmas,Winter Solistice,New Year Celebration

3)For years 2017,18: August and september are also showing growth reason may be a major festivel in New York like Herlim

6)Visualization on Co-relation between the variables in data.

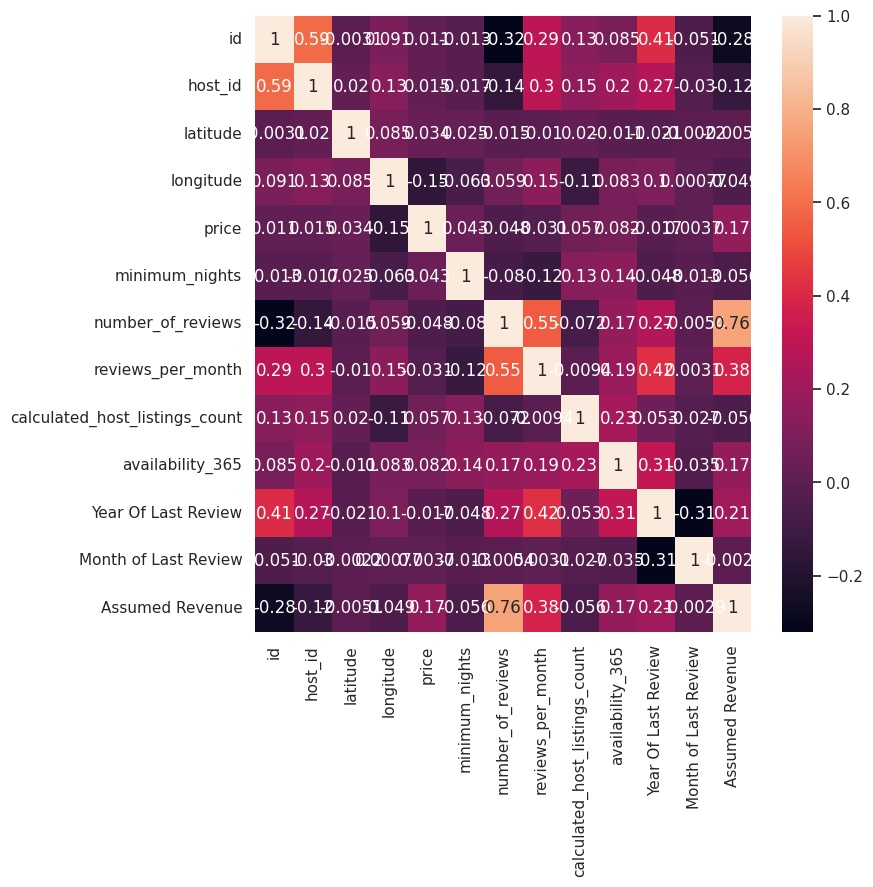


Chart 6.1

Positive co-relation between availability,reviews,minimum night means when rooms/homes are having high number of reviews or offering high minimum nights then the price is also increasing

Negative Co-relation:

1)When lattitudes are increasing the listings are reducing

2)Longtitude and price are showing a strong negative co relation.

If the issue regarding the latitudes and listings having negative co-relation is addressed (means reasons like competition from local agency,poor penitration,etc) the this can create a positive impact on bussiness.

**Solution to achieve Business Objective:**

## Airbnb is on it's mission of creating a world where anyone can belong anywhere,for achieving this goal Airbnb needs to expand its reach i.e penitration across the world keeping this in mind i have done data analysis and mentioned some points below:

1)In 2011 Manhatten,Brooklyn,Queens was having Airbnb's listings and 2014 Staten Island and Bronx was also added.In this way it can expand the business across the world. (**for visualisation refer chart 4.5**)

2)For expanding the business ,revenue is very essential,which is also in favour because revenue is growng over the years and 2018 and 2019 saw a exponential growth in terms of revenue. (**for visualisation refer chart 5.1 & 5.2**)

3)It should catter all type of audiance like family,indivisual,rich,poor,middle class, chart 9(heatmap) shows that price and number of reviews are propertional.

**Below mention are some point which need a special attention:**

1)Listings concentration:Manhatten is 2nd most prefered neighbourhood group and has the highest number are host listings and in manhatten "Financial District" is having most listings.That means huge portion of Airbnb's revenue is from Manhatten.If competitors like Verbo,flipkey,etc takes over in Mnhatten then it would be very harsh gor business. (**for visualisation refer chart 1.2,3.1,3.2,3.3,3.4,3.4,3.6**)

2)Privacy concern:People are prefering private room over the shared despite so privacy might be the reason.Another possible reason is the booking price of shared room because private room and shared rooms are not having a much diffrence. (**for visualisation refer chart 2.1**)

# ****Conclusion****

Analysing the given data it is certain that Airbnb is growing and it has something for everyone(customer) which is very important in such business.But there are some faultlines which I have mentioned in previous sections,if they are not addressed or considred a problem the business might get affected in a negative way.