**The methodology used for the analysis of the Airnbnb NYC case study**

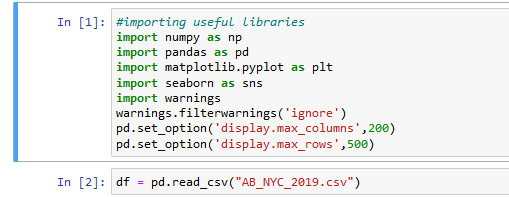
**Prepared by:**

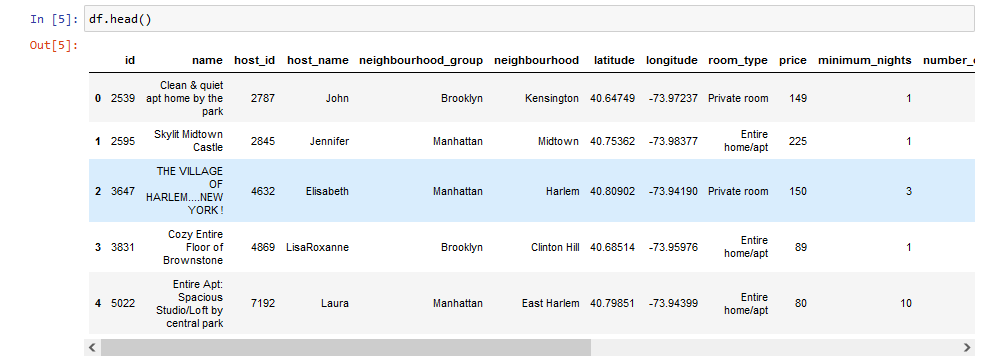
1. **Dharshak Chandra P**
2. **Digmakumari Tarunkumar Patel**
3. **Divya Darshani**

* The methodology used for data cleaning, and wrangling is in Jupyter Notebook using python.
* For visualization for data analysis, we used Tableau.

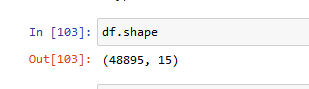
**Explanation for Presentation 1: for Data Analysis Manager, Lead Data Analyst:**

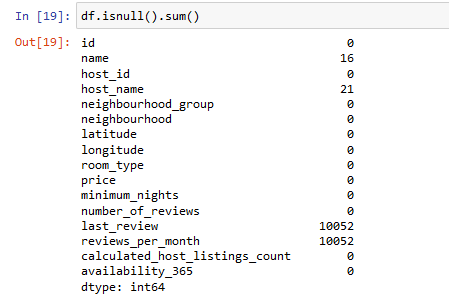
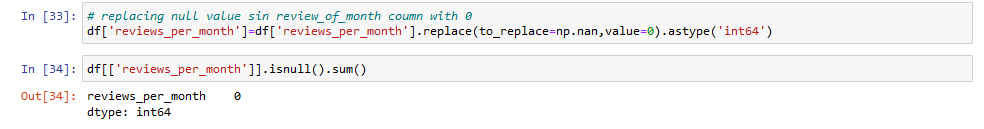
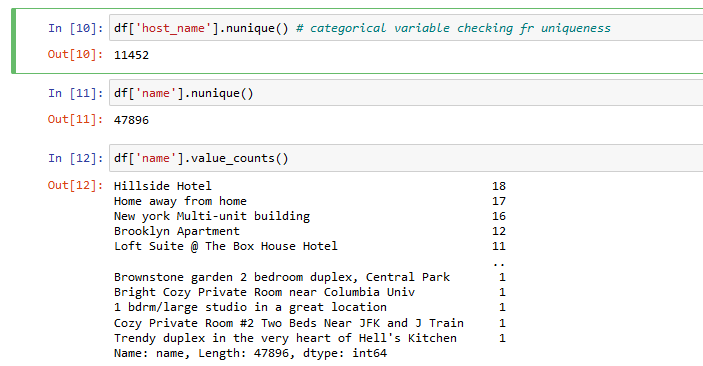
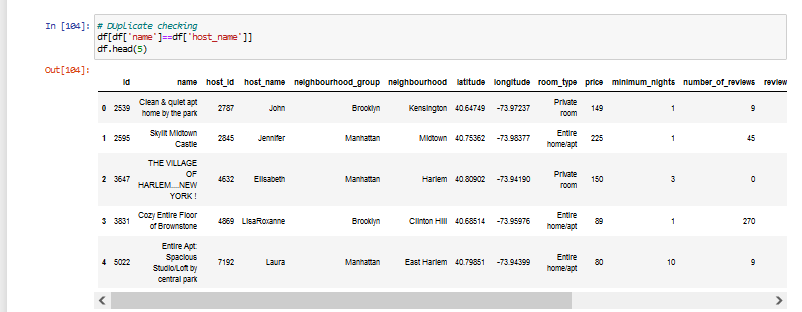
* Common Initial analysis included understanding data such as :
  + Using Jupyter notebook for data :AB\_NYC\_2019.csv

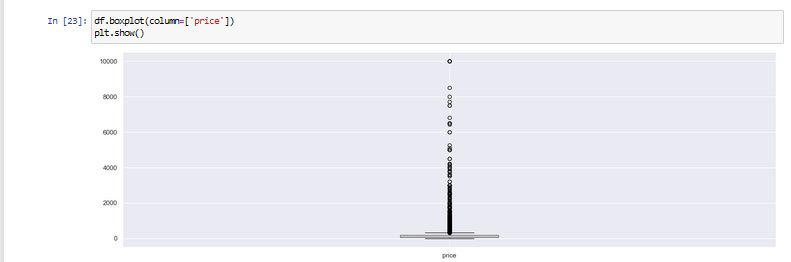




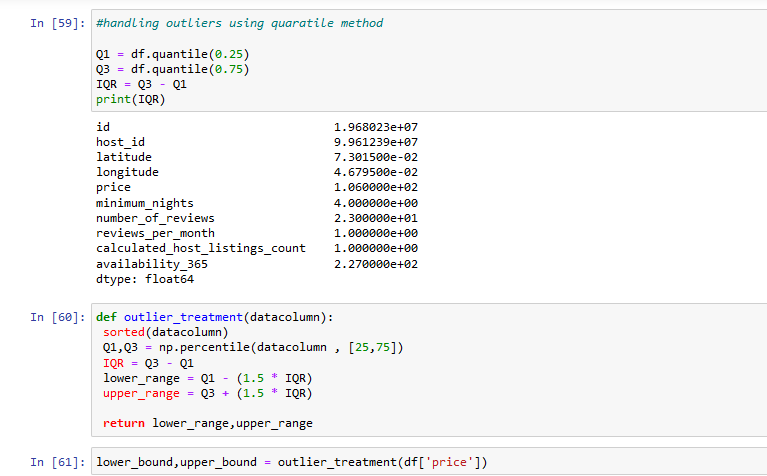
* + Number of rows & columns: 48895, 15

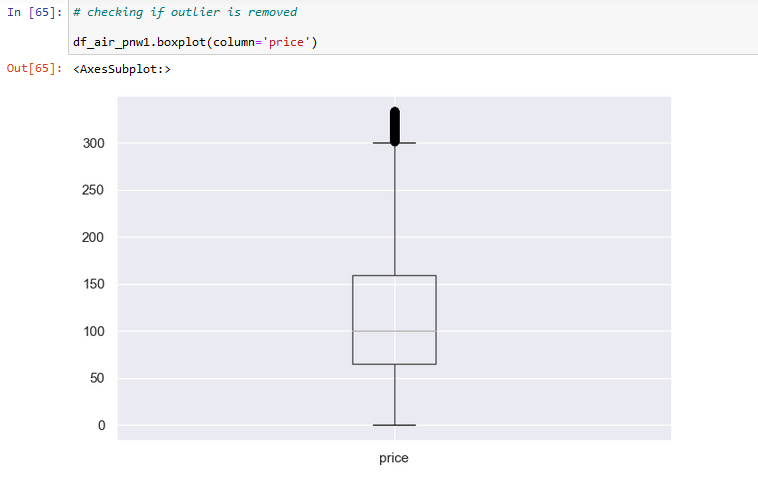


* Checking missing values & dealing with same ( by removing it if not required for further analysis)
  + 
  + 
  + 
  + 
  + 
  + Dealing with outlier columns (field)
    - Price column :



Handling outliers using quartile method:

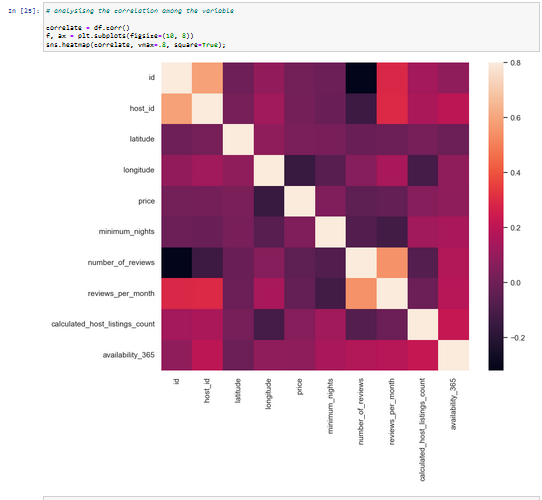




* Checking duplicate values in the table columns

Data wrangling & cleaning :

* Checked Missing values which are available in name, host name, last review, and review-per-month.
* Dropped that column having less percent of missing values as it won’t help much in the analysis
* Checked the uniqueness of the data in each column for values
* Handled the outlier using the Quartile method and made the column values ready for better analysis.
* **Identified the correlation among the present columns**

****

* From the heatmap, we can see the correlation between different features that can affect an Airbnb listing.
* There's a correlation among host\_id to reveiws\_per\_month & availability\_365. Also, there's a noticeable correlation between min\_nights to no\_of\_listings\_count & availability\_365. Price also shows some correlation with availability\_365 & host\_listings\_count.
* no\_of\_reviews and reviews\_per\_month gives almost the same information. so we can carry out an analysis with any of the two variables. Also, no\_of\_reviews is correlated to availability\_365

1. Top 10 host
   1. We identified the top 10 Host Ids, Host Name with count of Host Ids.
   2. Among the 10 lists of hosts for Airbnb seems Michael being the top most followed by David and to the end at 10 position its Maria.
2. Popular neighbour
   1. Manhattan & Brooklyn are the highest preferable neighbourhoods among all
   2. Whereas staten island is least preferred.
   3. Queens is third preferred airbnb spot followed with one third being preferred in Bronx.
   4. We took neighbourhoods in rows and the sum of reviews in column and took neighborhood groups in colour.
   5. We used a filter to show Top neighbours as per the sum of reviews.
3. Customer booking w.r.t minimum nights
   1. A very huge hike rate could be seen for Airbnb booking for 0-2 days listing out 2-4 days at just 50 % of previous one.
   2. 7 -14 & 20+ has a very ower bar followed with 4-7 & 14 - 20.
   3. We created the bin for Minimum nights.
   4. The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.
4. Airbnb availability in different Neighbourhood groups with respect to Minimum nights
   1. Most of the listings in all the different Neighbourhood groups have the Minimum nights requirement to be between 0 to 4 days.
5. Neighborhood vs Availability w.r .t minimum nights
   1. From the graph , the maximum availability for rooms is in Brooklyn & Manhattan.
   2. However, in both the neighborhood groups maximum price range preferred is 60 - 180

6. Median price with a neighborhood group

1. From the graph, it seems Manhattan has more price variance than the rest.
2. We changed the Price from a Sum Measure to a median measure

7. Preferred room type w.r.t neighborhood group

* 1. From the chart seems the private and Entire home occupied is more preferred by the customer than share room i.e 2.22%.
  2. We created a pie chart to understand the percentage of room type preferred by the neighbourhood group
  3. We added Room Type to the colours to highlight the different RoomType in different colours

**Explanation for Presentation 2: Head of Acquisitions and Operations, NYC, Head of User Experience:**

1. Popular neighbourhood
   1. Manhattan & Brooklyn are the highest preferable neighbourhood among all
   2. Whereas staten island is least preferred.
   3. Queens is third preferred airbnb spot followed with one third being preferred in Bronx.
2. Customer booking w.r.t minimum nights
   1. A very huge hike rate could be seen for Airbnb booking for 0-2 days listing out 2-4 days at just 50 % of previous one.
   2. 7 -14 & 20+ has a very ower bar followed with 4-7 & 14 - 20.
3. Airbnb availability in different Neighbourhood groups with respect to Minimum nights
   1. Most of the listings in all the different Neighbourhood groups have the Minimum nights requirement to be between 0 to 4 days.
4. Price range preferred by customers
   1. 35.78% of customer prefer price between 60 -180 ; followed by 0-60 & 120-180 on a range of 20.26% respectively.
   2. However, the preferences becomes lesser when going more further beyond 180 to 320+ which is nearly nil.
5. Neighbourhood vs availability w.r.t price
   1. From graph , maximum availability for rooms is in Brooklyn & Manhattan.
   2. However in both the neighbourhood groups maximum price range preferred is 60 - 180.
6. Understanding price variance w.r.t room type & neighbourhood
   1. From graph it seems Manhattan has more price variance then rest.
7. Room type w.r.t individual neighbourhood groups
   1. The preference for shared & private room is nearly same in Brooklyn where as in Bronx private room preference is 60% & that of shared room is 34 %.
   2. From chart seems the private and Entire home occupies is more preferred by customer than share room i.e 2.22% in Manhatten.
   3. In Queens, highly preferred are private rooms consisting of approximately 60 % where as only 37% is for Shared rooms
   4. From chart seems the private and Entire home occupies more preferred by customer than share room i.e 2.41% in Staten Island**.**