**Airbnb Case Study**

**Problem background**

Suppose that you are working as a data analyst at Airbnb. For the past few months, Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.

The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue such as -

* Which type of hosts to acquire more and where?
* The categorisation of customers based on their preferences.
* What are the neighbourhoods they need to target?
* What is the pricing ranges preferred by customers?
* The various kinds of properties that exist w.r.t. customer preferences.
* Adjustments in the existing properties to make it more customer-oriented.
* What are the most popular localities and properties in New York currently?
* How to get unpopular properties more traction? and so on...

**End Objective**

To prepare for the next best steps that Airbnb needs to take as a business, you have been asked to analyse a dataset consisting of various Airbnb listings in New York. Based on this analysis, you need to give two presentations to the following groups.

Presentation - I

* Data Analysis Managers: These people manage the data analysts directly for processes and their technical expertise is basic.
* Lead Data Analyst: The lead data analyst looks after the entire team of data and business analysts and is technically sound.

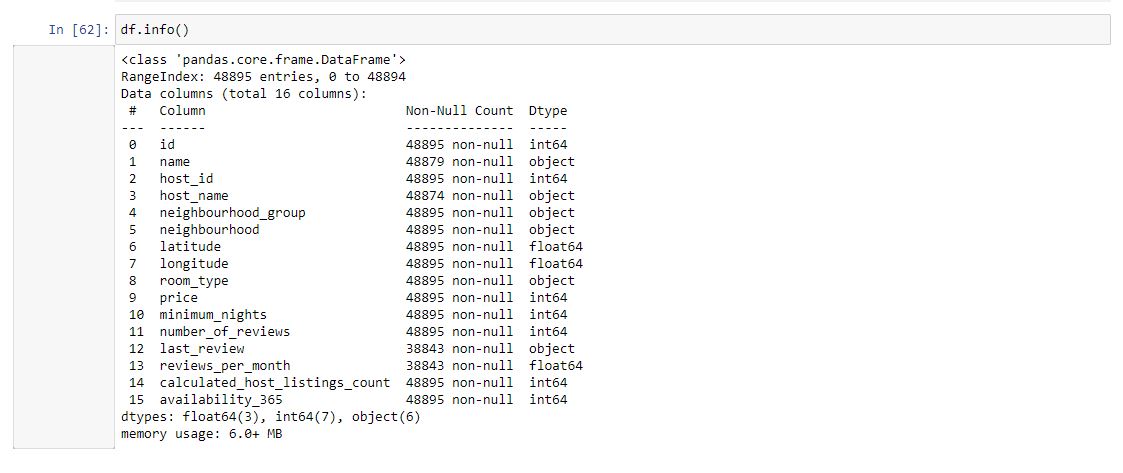
Presentation - II

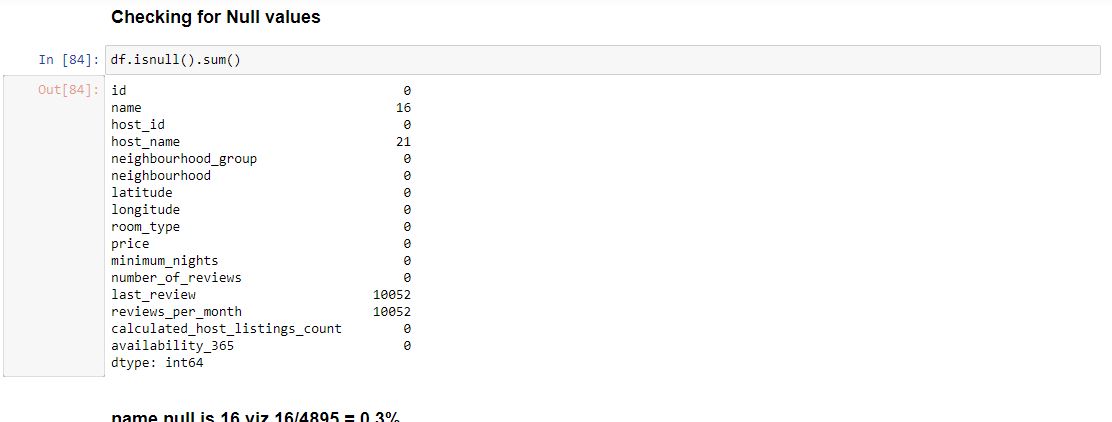
* Head of Acquisitions and Operations, NYC: This head looks after all the property and host acquisitions and operations. Acquisition of the best properties, price negotiation, and negotiating the services the properties offer falls under the purview of this role.
* Head of User Experience, NYC: The head of user experience looks after the customer preferences and also handles the properties listed on the website and the Airbnb app. Basically, the head of user experience tries to optimise the order of property listing in certain neighbourhoods and cities in order to get every property the optimal amount of traction.

**Methodology**

Data Cleaning & Preparation for visualization and analysis purpose

Tool Used - Python





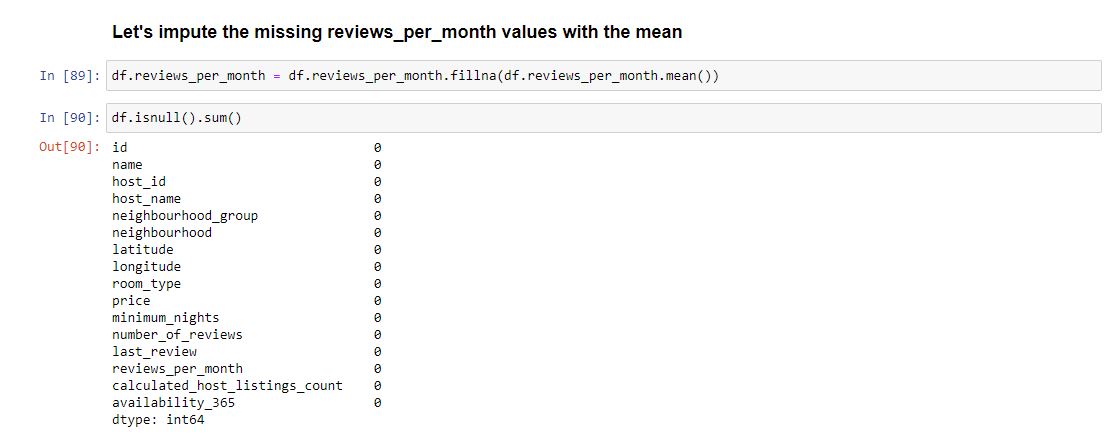
name null is 16 viz 16/4895 = 0.3%

host\_name null is 21 viz = 0.4%

PS these are unique values so we cannot impute with any other value so we unfortunately have to drop these records

For the below columns null value % is 20.5 % viz slightly above the threshold but these columns are important so we cannot drop these, we have to impute with the median value

* last\_review
* reviews\_per\_month





* Checked the Duplicate rows in our dataset and no duplicate data was found.
* Checked the Null Values in our dataset.
* Checked the formatting in our dataset.
* Identified and review outliers.

**Data Analysis and Visualizations using Tableau**

**PPT 1**

We have used tableau to visualize the data for the assignment. Below are the detailed steps used for each

visualization.

* Top 10 Host: We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree Map
* Preferred Room type with respect to Neighbourhood group: We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
* For Variance of price with Neighbourhood Groups: We used a box and whisker’s plot with Neighbourhood Groups in rows and Price in columns.
* Customer Booking w r t minimum nights:
  + We created the bin.
  + The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.
* Popular Neighborhoods: We took neighbourhood in columns and sum of reviews in rows and took neighbourhood groups in colour.
* Neighbourhood vs Availability: We created a dual axis chart using bar chart for availability 365 and line chart for price for top neighbourhood group sorted by price.

**PPT 2**

* Preferred Room type with respect to Neighbourhood group: We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
* Customer Booking w r t minimum nights
  + We created the bin.
  + The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.
* Neighbourhood vs Availability: We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.
* We showed famous top10 host id according to the number of reviews. We showed it geographically with the help of map and also by bar chart.