**Air BnB Case Study Methodology Document**

**By Madhura Pathak & Prachi Faye**

* **Storyboarding**
* First, we got familiarized with dataset and noted down the important fields.
* We mapped down the various slides of presentation to answer the questions asked then made a rough template on basis of this.
* Then visualized a rough skeleton and flow of the presentation to tell the data story insights which will cover an important aspect of our story and that will make insightful presentation to the stakeholders.
* **Data Reading and understanding**

Data Reading and cleaning was done in Jupiter notebook and visualizations majorly done in tableau.

Importing libraries and Data

Text

Description automatically generated



Checking Data info

Table

Description automatically generated

Data didn’t have any duplicate values

Graphical user interface, text, application, chat or text message

Description automatically generated

Missing values were detected

Table

Description automatically generated

Tried to understand data in more depth by finding mean & percentiles

Table

Description automatically generated

* **Data preparation and Pre-Processing**

Null values for reviews\_per\_month were replaced with “0” and column the last\_review was split into year, month and day columns (in tableau dashboard) to get important insights.

Text

Description automatically generated

Changing the data type of ID and Host\_Id

Graphical user interface, text, application

Description automatically generated

* **Dropping the variables which are not useful for analysis**

Graphical user interface

Description automatically generated with medium confidence

* We checked for outliers, but we did not deal with it as when we tried dealing with it, It smoothened out the data and hence all insights were getting lost.
* There was small number of Null values which was not affecting analysis hence didn’t remove them.
* We analyzed the data set in python as well as in tableau to draw meaningful insights by conducting univariate and bivariate analysis and tried to draw a conclusion from the correlation analysis.
* **Data Analysis :**

In order to measure our analysis, we created a Matrix to provide us a direction while creating graphs using different Dimensions and Measures.

Text

Description automatically generated

* **Top 25 words used in naming the place**

Graphical user interface, text, application, email

Description automatically generated

Chart

Description automatically generated

* We observed after going over top 25 used listings' name words.
* Simple words were used to describe Space and area where the listing is such as 'room', 'bedroom', 'private', 'apartment', 'studio'.
* No catchphrases or 'popular/trending' terms that are used.
* This might have helped easier search by a potential traveler which is needed in dealing with multilingual customers.
* **In Tableau, we applied binning in few variables like price, availability\_365, No of review, Min nights.**

Text

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

* **Count of listing & Total revenue w.r.t. Room type and neighborhood group**

Chart

Description automatically generated

* Host listing count & revenue is maximum for Entire home /Apt and that too for Manhatten neighborhood group.
* Brooklyn leading with highest revenue in private rooms.
* Listing and revenue of shared rooms is very negligible in all neighborhood groups.
* **Neighborhood group Vs Reviews range & Neighborhood group Vs Average Price**

Chart, waterfall chart

Description automatically generated

Chart

Description automatically generated

* The majority listing are from Manhattan followed by Brooklyn and Queens.
* **Room Type Vs Total Reviews vs Total Revenue**

Line chart

Description automatically generated

* Entire home Room type leads in Revenue and no of reviews too.
* However comparatively revenue from Private room is less.
* The Revenue and Reviews for shared room are lowest.
* **Top 10 Hostedplace (property) names vs Avg Price vs Avg Review**

Chart, histogram

Description automatically generated

* Max review Average per month for any AirBnb property is **58.5**.
* Property” Enjoy great views of city in our Delux room” has max reviews and the avg price is 100 USD per night.
* This is excellent example to keep Price moderate and have greater business and more satisfied customer.
* **Room Type Availability basis Price range**

Chart, bar chart

Description automatically generated

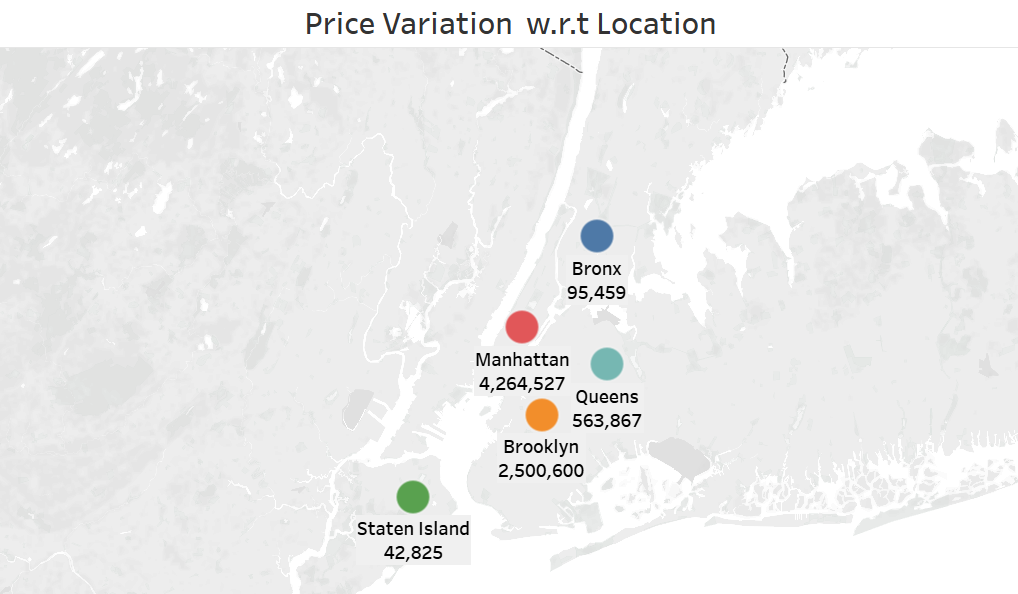
* We can see majorly number of properties having availability upto **100 nights** per year are more.
* In that leading is Price range 71-150 USD per night then 0-70 USD.
* Properties in availability range 201-300 are majorly from price range 501-800USD that means we need to reduce the price to increase consumption of those rooms.
* **Number of Listing vs Review range**

Chart

Description automatically generated

We have highest reviews for properties in price range 0-50USD. Manhatten, Brooklyn and Queens are among top 3.

* **Price Variation w.r.t Geography**

****

* This shows us the dataset distribution in New York city with respect to latitude and longitude.
* From the above graph Price of accommodation in Manhattan is pretty higher and affordable for high class people.
* **Last Review Monthly Vs Number of Reviews**

Graphical user interface, application

Description automatically generated

* We observed that highest monthly reviews were received in 2019 and even the listings are highest in 2019 (both from properties in Brooklyn& Manhattan).
* Above Visualization shows the marking for highest and lowest monthly reviews.
* **Average Minimum Nights vs Room type**

Chart, bar chart

Description automatically generated

* Average room night for entire home/Apt are highest which is 11 in Manhatten.
* In comparison Private room has lower min nights.
* Shared room also have very high min nights and may be that’s one of the reason their demand is low.
* **Average Price Vs Review Range**

Chart, bar chart

Description automatically generated

* **Average Minimum Nights vs Price Range & Roomtype**

Chart, bar chart

Description automatically generated

* We can see for price range above 1500USD the min room nights is very high.
* We can certainly reduce the min room nights to increase consumption of those room nights.
* **Top 10 Host Name on basis of Highest Reviews with Price insight**

Chart, bar chart

Description automatically generated

* Host John from Williamsburg, Brooklyn have highest 719 reviews.
* However he has not given name to the property correctly so its appearing it as #Name?
* Similarly, there are two more host which has done the same. Hence the Property with highest review appears to be #NAME?
* As the field name is such we cannot impute missing value with mode as the insight will go we kept it as is.
* We saw majority of the Top host properties are from price range 0-70 USD.
* **Top Locations on basis of Avg.Price Vs Avg.Reviews**

Bar chart

Description automatically generated with medium confidence

* Manhatten has highest avg.price for property but Staten island has highest avg reviews.
* The Avg. price of Staten Island properties is lower.
* Also from earlier visualization we saw Staten Island have high availability too.
* We need to think of other ways to increase bookings in Staten Island along with reducing rate little bit.
* **Top 20 Hosted Place Name on basis of Price**

Chart, bar chart

Description automatically generated

From above graph are we can see majority of hosted places are from Manhattan neighborhood.

**Presentations :**

* Made presentations adhering to best practices and pyramid principle.
* Added recommendations for respective departments.