Airbnb Listings in New York City

**Tasks each one of us completed:**

Akhila –

Shivani –

**Why did we pick Airbnb Listings Dataset?**

Airbnb has become a popular service since it started in 2008. We wanted to analyze the listings in New York City as it is one of the most populated cities with Airbnb listings. We wanted to learn which hosts are the most popular and the reason behind it to validate some of our assumptions based on the dataset. Also, we analyzed which neighborhoods have the highest prices and what room types are popular in which neighborhoods.

**What were some of the challenges we faced?**

While we were cleaning the dataset, there were many null values needed to be removed. We decided to drop the host\_name column as it is unnecessary personal information which does not add to our analysis. Also, reviews\_per\_month had many NA’s which we didn’t know how to remove without affecting the dataset. Later we realized, if there were no guests, reviews for that month could be NA.

Understanding which visualizations would have a strong impact was challenging. After plotting several different visualizations, we came to a conclusion. Moreover, while plotting the graphs, the challenging task was to understand the aesthetics of the graphs. Overtime, we got used to it.

**Which were the topics we had to learn more about?**

While cleaning the data, we learned about many functions like drop, fill.na and isnull(). This helped us clean more efficiently.

We had to learn about various plots in the seaborn library which would be efficient for our purpose. Also learned to make each visualization using the same color palette.

Even though we didn’t use the function ‘. describe ()’, we learned that it gives a summary of all the statistical analysis such as mean, max, min, standard deviation, etc.

**How else could we expand the project?**

This is a rich dataset which could be used to predict the average price point by season. Since New York City has seasonal high traffic. Airbnb hosts could charge the optimal high price by neighborhood which would help attract the most amount of traffic to their listing.

As far as analyzing the project even further, a number of correlation graphs can be plotted between each column (such as mimumum\_graphs, number\_of\_reviews) and price. Further regression analysis & predictions can also be done.

**Citations:**

<https://datavizpyr.com/sort-bars-in-barplot-using-seaborn-in-python/>

<https://seaborn.pydata.org/tutorial/aesthetics.html>

<https://datavizpyr.com/countplot-or-barplot-with-seaborn-catplot/>

<https://seaborn.pydata.org/tutorial/color_palettes.html>

<https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html#visualization-barplot>