Applied Data Science 1

**Clustring and Fitting**

# Submitted by: Bilal Ahmad #23035733

# Link: [Github Repository](https://github.com/bilal147-st/Fitting-n-clustring)

**Exploratory Analysis Report: Insights from [Airbnb Dataset](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)**

# Introduction:

In this report, I have presented an exploratory analysis of the New York Airbnb dataset 2019, aiming to uncover insights into pricing trends, property types, neighbourhood distribution, and their relationships. Utilizing various visualizations, we delve into the data to reveal patterns and associations that can inform strategic decisions for hosts and travelers alike.

# Methodology:

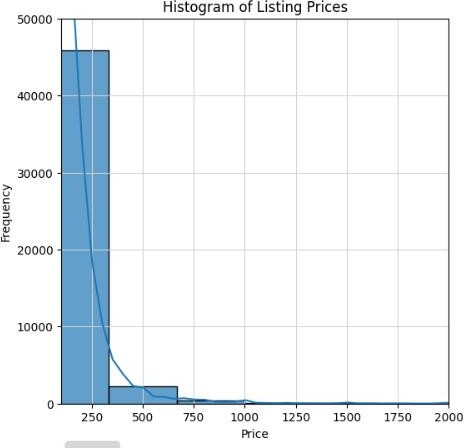
The analysis is based on a comprehensive dataset obtained from kaggle, encompassing listings from various neighbourhoods. We utilize Python programming language along with popular data analysis libraries such as Pandas, Matplotlib, and Seaborn for visualization and exploration as we learn in classroom sessions.

# Histogram of Price Distribution:

The histogram illustrates the distribution of listing prices across the dataset. We generate a histogram of the 'price' column in the Airbnb dataset using matplotlib.pyplot.hist(). The number of bins in the histogram is determined by the bins parameter. The colour of the edges are determined by the edgecolor parameters, respectively. Lastly, we label the x and y axes as Price and Frequency, respectively and give the histogram a title.

# Insights:

Prices are skewed to the right, indicating that the majority of listings are concentrated at lower price points. A long tail suggests the presence of high-priced outliers.



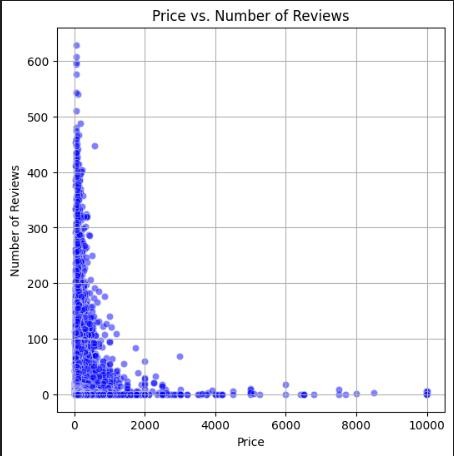
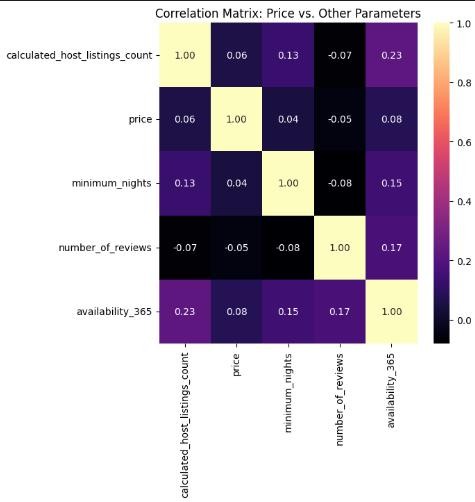
# Scatter Plot of Price vs. Number of Views:

The scatter plot examines the relationship between listing price and the number of views. For the purpose of visualising overlapping data points, the alpha value is set to 0.5, which semi-transparently displays the points. To alter the transparency of the points, modify the alpha value.

# Insights:

No clear linear relationship is observed between price and views, indicating that factors beyond price influence listing popularity.

Clusters of points at various price levels suggest different pricing strategies among hosts.

# Heatmap of Numerical Parameters:

The heatmap illustrating the correlations between numerical parameters such as price, number of reviews, and availability of Airbnb listings in a New York was one of the main visualizations we produced. We were able to determine which Airbnb listings had the highest and lowest number of reviews thanks to this heatmap.

# Insights:

Price correlates negatively with the no\_of\_reviews which indicates that expensive places have fewer customers thus fewer reviews. Availability shows a positive trend with the listing which indicates that places which are available for a longer period of time will have greater listings irrespective of their price.

# Conclusion:

Through exploratory analysis of the Airbnb dataset, we have gained valuable insights into pricing dynamics, property types, and neighbourhood distributions. These insights can inform strategic decisions for both hosts and travellers, aiding in pricing optimization, property selection, and neighbourhood exploration. Further analysis and modelling could delve deeper into predictive analytics and market trends, providing even greater value to stakeholders in the Airbnb ecosystem.