**7PAM2000 APPLIED DATA SCIENCE 1**

**ASSIGNMENT 1: VISUALIZATION**

**STUDENT NAME :PRASAD KIRANAMPALLI**

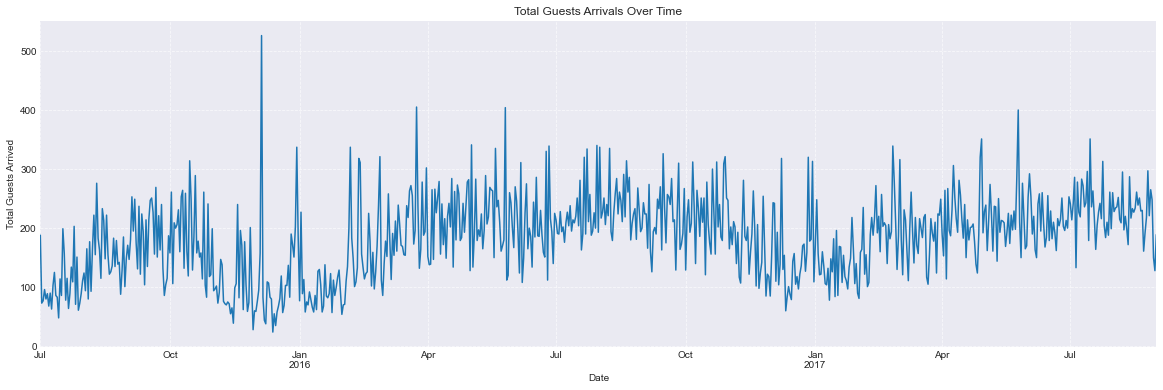
**STUDENT ID:22073170**

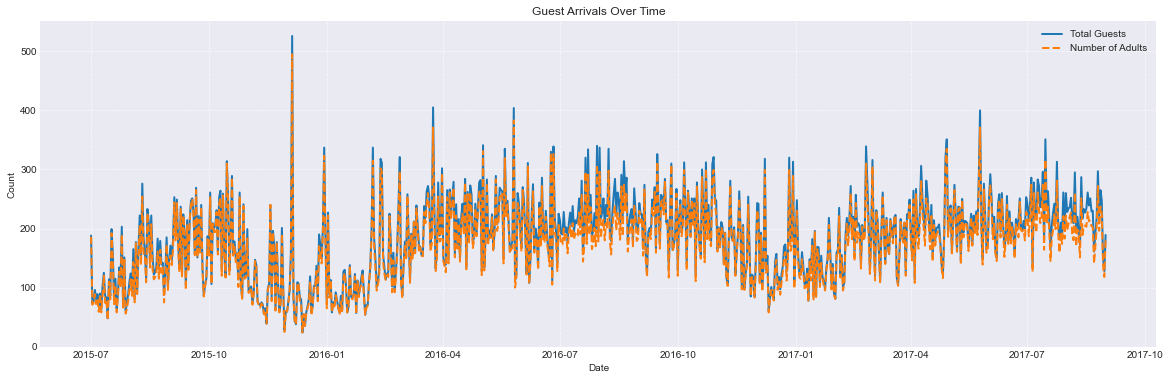
**GITHUB LINK:[ https://github.com/kp23aav/ADS1-ASSIGNMENT.git ]**

**Dataset:**

The dataset, ‘Hotel Booking Deman’ is available on Kaggle. It contains information related to the hotel bookings and contains different features which could be used for analysis and research purpose. (Hotel booking demand, n.d.):

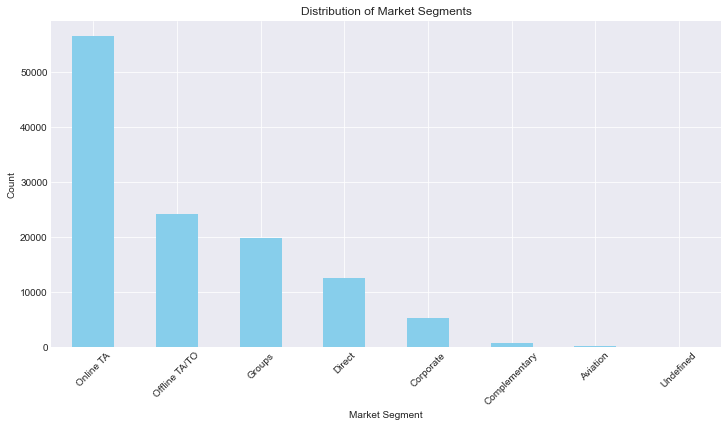
**Line Plot:**





In this data the line plot is used for plotting the guests who are arriving verses the time. So, the first line plot shows the arrival of total guests over a time period while the second line plot has multiple line plot on it which shows the total number of guests arriving and the total number of adults who are arriving in the hotels. The blue solid line shows the total guests and the dashed orange line shows the total number of adults who are arriving. The line plots are suitable for such data because these data are plotted against time and also helps in understanding the data points over time ultimately helping in understanding the pattern of data.

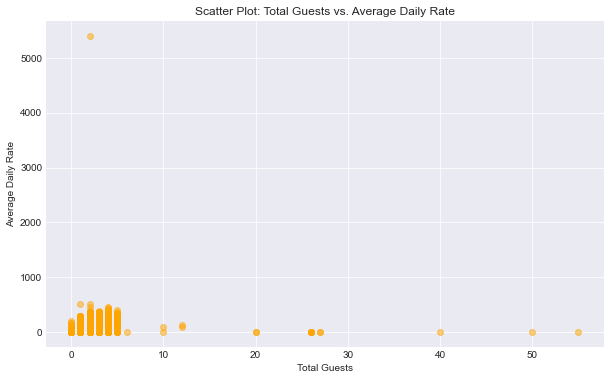
**Bar Plot:**



Bar plots are used for plotting the categorical data points or the frequencies of the data. The bar height is decided by the number of frequencies in the data points. As the frequency count increases the bar height also increases.

A bar plot is used for displaying the distribution of market segments in the dataset. Each different market segment is shown by a new bar on the bar chart. Bar plot is clearly depicting the proper distribution of the market segment and also helps in understanding which market segment is high on demand. Thus, this bar plot s selected for plotting the charts.

**Scatter plot:**



Scatter plot is used to find out the relationship between two variables and the correlation between the variables.

Here the variables ‘Total guests’ verses the ‘Average daily rates’ are being plotted using the scatter plot which helps in understanding the relation between these two variables. It could be observed that there isn’t any relevant relationship between the variables.

# References

*Hotel booking demand*. (n.d.). Retrieved from Kaggle: https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand/data