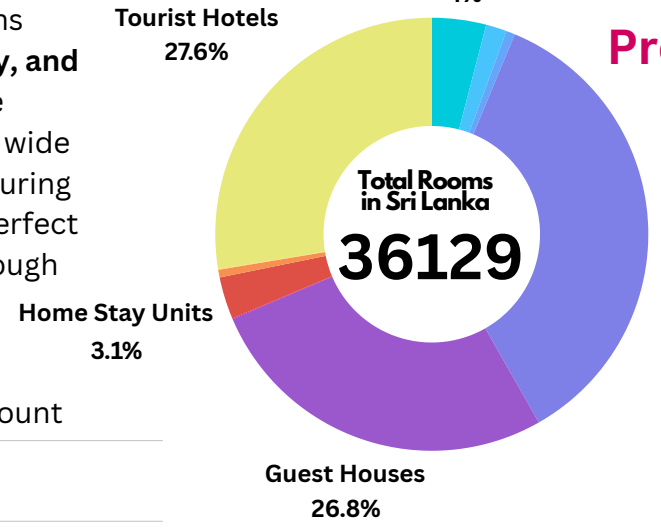
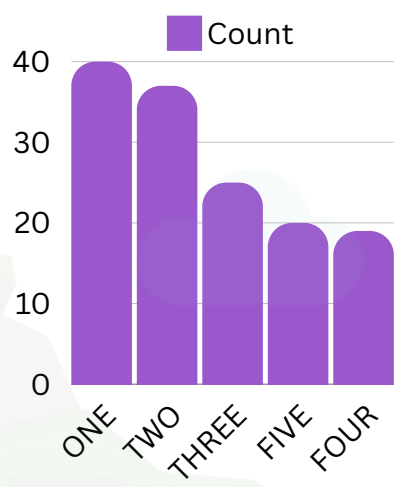
SRI LANKA TOURIST ACCOMODATION

Sri Lanka posses an impressive array of accommodation options, with a total of **2,130** establishments offering **36,129** rooms, according to **data.gov.lk.** The highest concentration of these accommodations can be found in the **Western and Southern Provinces**, with popular destinations including the **South-West Beach, Kandy, and the Central Province**. This diverse accommodation landscape caters to a wide range of preferences and budgets, ensuring that visitors to Sri Lanka can find the perfect place to stay during their journey through this captivating island nation.

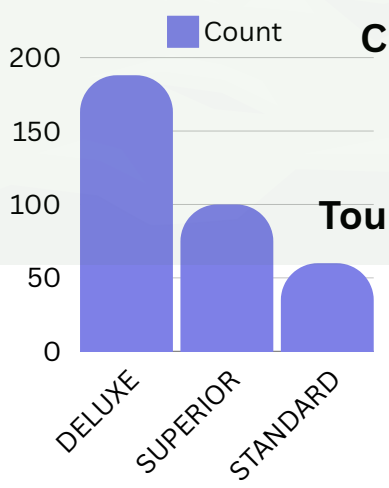
39.71%
of Accomondation is in Western Province



70.41%
of Accomondation is in Coastal Area

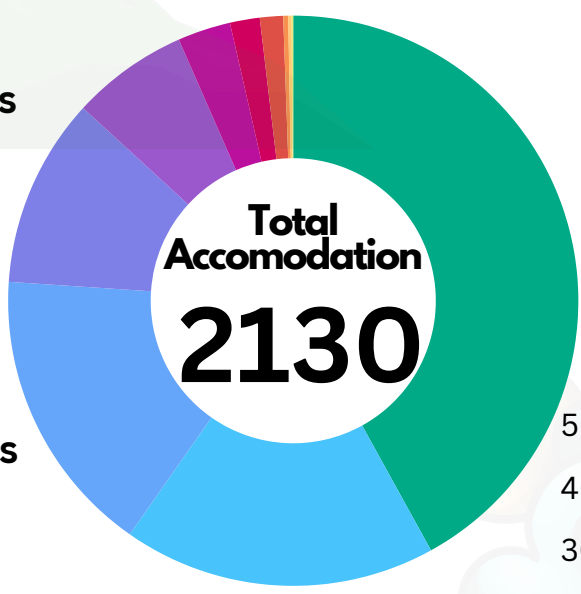


Highest Percentage of Total Accomodation is catered by **Classified Star Hotels** even though total number of Guest Houses, Home Stay Units and Bunglows are numerically higher.

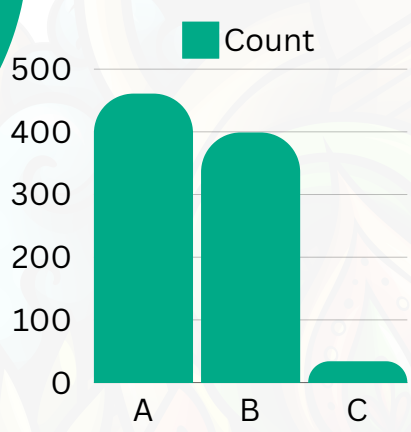


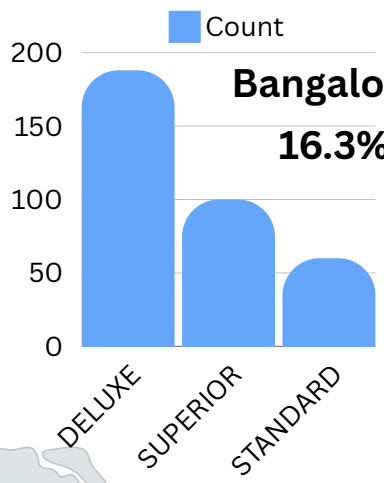
Classified Hotels(1-5 Star)
6.6%

Tourist Hotels
10.8%



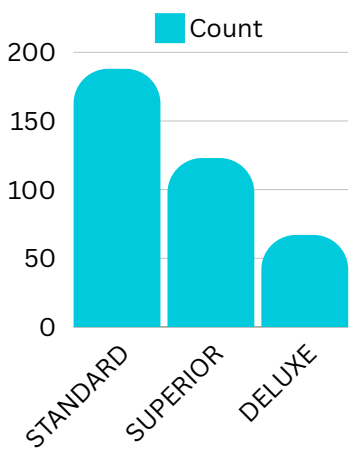
Guest Houses
42%

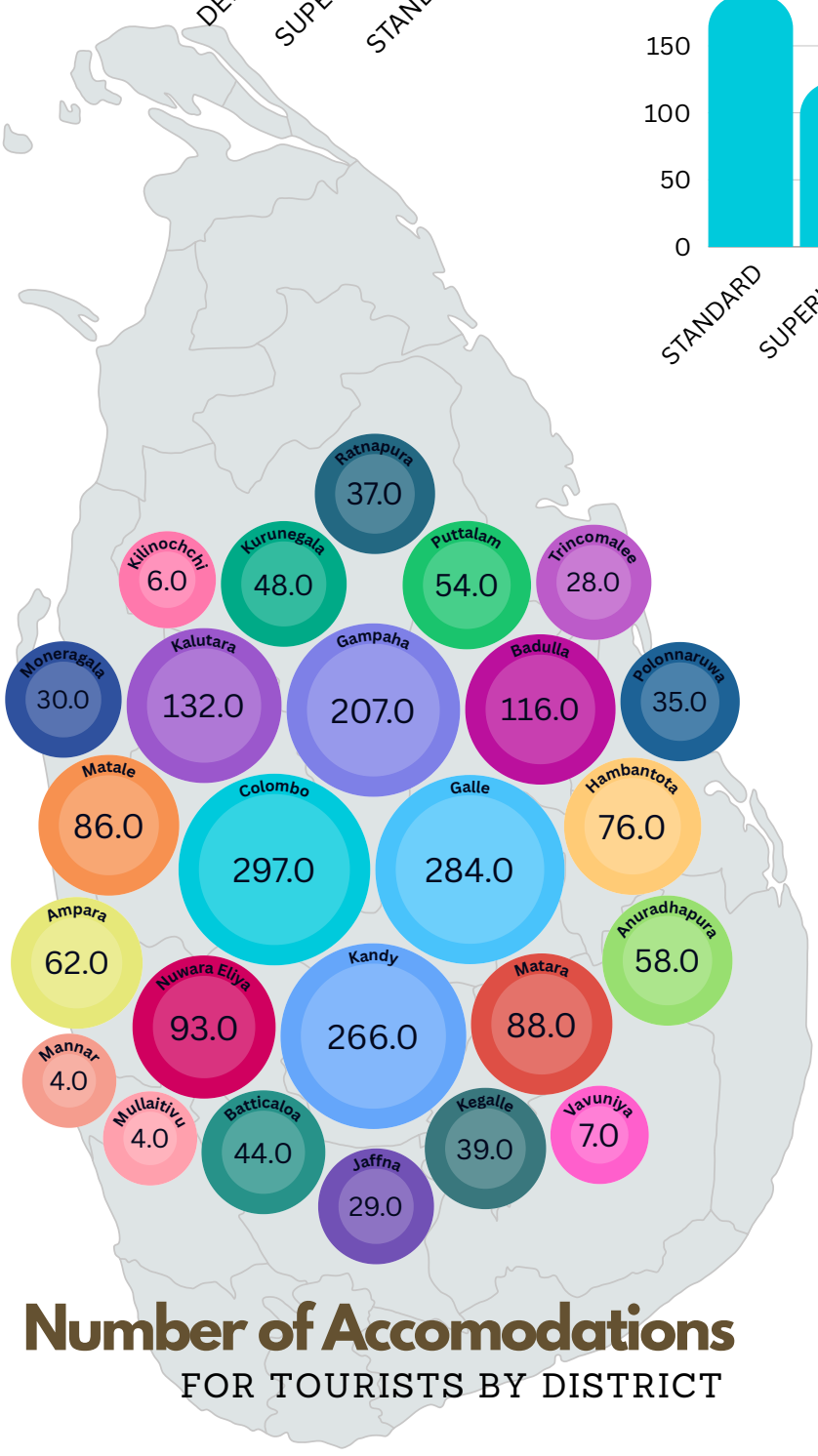




Bangalows
16.3%

Home Stay Units
17.7%





Number of Accomodations
FOR TOURISTS BY DISTRICT

7519
Rooms in Colombo

5287
Rooms in Galle

3543
Rooms in Gampaha

3285
Rooms in Kalutara

2789
Rooms in Kandy

Dataset used: https://data.gov.lk/dataset/accommodation-information-tourists/
Information for accommodation.csv file.
Used Libraries: Pandas, Matplotlib, Seaborn, Keplergl, Geopandas, Scikitlearn, Plotly, Numpy, contextily
Used Tools: VScode Notebook, Canva to build the infographic poster
Reason of Selection of Dataset: Most of the Datasets on data.gov.lk were smaller csv files with limited number of records and very small number of dimesions. This dataset had significant number of records even though the dimensions were quite limited for an effective predictive and prescriptive analysis.
Main Challenges: Missing Values, Limited Number of Dimensions of the dataset, Limited Number of information, Difficulty in making useful prediction model building.

1. At the beginning to get an idea about the dataset summary statistics like mean, median, mode, range, and standard deviation for numerical columns like 'Rooms'. This gave an idea of the distribution and central tendency of data. Moreover, the missing values were viewed using seaborn heatmap.
2. There were multiple missing values, Such as the location information Latitude and Longitude information and Accommodation Grade related information.
3. However, for the analysis primary dimensions were Type, District, Number of Rooms that mattered the most for the visualizations.
4. Accommodation count, room count by district were analyzed and did considerable manipulations and views on the data.
5. There were challenges in the analysis when visualizing the locations information on the map since there were numerous records with location information. However, made the decision to not visualize that on the infographic poster itself since considering the whole dataset there were many data points with location information missing and imputation will not make much sense. However, number of accommodations was visualized using bubble diagram.
6. However, visualizations were done using Kepler, geopandas some libraries of the location information.
7. Prediction modelling was difficult and not making much sense due the nature of the dataset.
8. However clustering analysis was done using K-means clustering to check the possibility of clustering the geospatial data and separate the information to tourism clusters. It was somewhat successful. However, it was not included in the infographic since this dataset didn't have more dimensions to visualize the clustering information with more valuable insights and recommendations.
9. In addition, Classification of hotel room types and condition was presented in the infographic per Type.

Github Repo:

https://github.com/amilawijesooriya/CS5617-Poster-Assignment

https://github.com/amilawijesooriya/CS5617-Poster-Assignment/blob/main/poster_analysis.ipynb

Student Name: Wijesooriya D.A.P
Student ID: 248385K