Report on Tourist Attractions Dataset

Author: Raghuvaran Boddupally
Student ID: 23013721
GitHub Repository: https://github.com/BRaghuvaran/Applied-Datascience-Assignment/tree/main

Introduction:

The tourism sector is a vital contributor to regional economies, drawing visitors and supporting local businesses. Analyzing data on tourist attractions, including visit duration, Google review ratings, entrance fees, and other details, is crucial for strategic planning and development. This report delves into such data to glean insights into the characteristics and factors that shape tourist attractions.

Overview:

The dataset contains information on 325 tourist attractions spanning various zones, states, and cities. It encompasses details like establishment year, visit duration, Google review ratings, entrance fees, and more. Our objective through exploratory data analysis is to reveal insights into these attractions' attributes and their interrelations.

Statistical Summary:

A statistical analysis of the data reveals the following insights:

The time needed to visit in hrs: The mean time required for a visit is 1.81 hours, with a standard deviation of 0.97 hours, indicating moderate variability. The duration ranges from a minimum of 0.5 hours to a maximum of 7 hours, with the 25th, 50th, and 75th percentiles falling at 1 hour, 1.5 hours, and 2 hours, respectively.

Google review ratings exhibit a mean of 4.49 with a standard deviation of 0.27, showcasing generally high ratings. Ratings range from a minimum of 1.4 to a maximum of 4.9, with the median rating at 4.5 and the 25th and 75th percentiles at 4.4 and 4.6, respectively.

The entrance fee distribution is skewed, with a mean of 115.81 INR and a wide standard deviation of 530.86 INR. Fees vary greatly, ranging from 0 INR to 7500 INR, with 25th, 50th, and 75th percentiles all at 0 INR, reflecting a significant proportion of venues with no entrance fee.

Lastly, the number of Google reviews has a mean of 0.41 lakhs and a standard deviation of 0.65 lakhs, indicating a substantial range in review counts. Reviews span from a minimum of 0.01 lakhs to a maximum of 7.4 lakhs, with the median review count at 0.17 lakhs and the 25th and 75th percentiles at 0.059 lakhs and 0.5 lakhs, respectively.

Exploratory Data Analysis Report:

Pie Chart: Type Distribution

This report seeks to visualize the distribution of different types of tourist attractions using a pie chart. The aim is to understand the proportion of each attraction type within the dataset, providing insights into the diversity of tourist offerings.

Observations:

Temples constitute the largest portion of tourist attractions in the dataset, followed by Parks and Museums.
Observatories, Theme Parks, and Tombs represent a moderate portion of attractions.
Monuments, Markets, Forts, Stepwells, and Zoos constitute smaller proportions within the dataset.

Scatter Plot: Establishment Year vs. Number of Google Reviews

The goal of this report is to examine the relationship between the establishment year of tourist attractions and the number of Google reviews they receive. By plotting the establishment year against the number of Google reviews for each attraction, we aim to identify any trends or patterns in attraction popularity over time.

Observations:

There is no clear linear trend between the establishment year and the number of Google reviews.
Attractions established in various years exhibit a wide range of review counts, indicating diverse levels of popularity irrespective of establishment year.
Certain outliers may represent attractions that have experienced exceptionally high or low levels of engagement compared to others.

Heatmap: Zone vs. Google Review Rating

The objective of this report is to analyze the relationship between different zones and the corresponding Google review ratings of tourist attractions. By visualizing this relationship using a heatmap, we aim to identify any patterns or discrepancies in visitor satisfaction levels across different zones.

Observations:

The heatmap illustrates varying average Google review ratings across different zones.
Attractions in the Northern and Southern zones generally receive higher ratings compared to those in the Eastern, Western, and Central zones. The Central zone exhibits the widest range of ratings, indicating diverse visitor experiences within this zone.

Correlation Matrix:

A correlation matrix reveals the relationships between different variables in the dataset. For example, we can examine if there's any correlation between visit duration, Google review ratings, entrance fees, and the number of Google reviews.

Insights and Observations:

The Google review ratings generally reflect high levels of visitor satisfaction across the attractions. There is a wide variation in entrance fees, implying diverse pricing strategies among different attractions. The visit durations at attractions vary considerably, indicating a range of visitor experiences and levels of engagement. Certain correlations may exist between attributes; for example, attractions with longer visit durations may tend to have higher entrance fees.

Conclusion:

Analyzing the dataset on tourist attractions provides valuable insights into their features and influencing factors. By understanding visit durations, Google review ratings, entrance fees, and other attributes, tourism officials and businesses can make informed decisions, refine marketing strategies, and enhance visitor experiences. Further exploration and analysis could offer deeper insights into the dynamics of tourist attractions and their contributions to local economies and tourism industries.