CIS 7030

GEOSPATIAL ANALYSIS

Contents

Task no 01 - How geospatial data science can be used for business.

Business plan - starting a new hotel in optimal location in Sri Lanka

Introduction

Sri Lanka's thriving hotel sector offers a great chance for success and innovation as it prepares for a post-pandemic comeback. Our proposal is to develop a hotel that seamlessly combines the revolutionary power of geospatial data science to take advantage of this growing potential. Our hotel will stand out in this competitive marketplace because to our forward-thinking strategy, which will change our market analysis, client engagement, and operational efficiency with geographical location.

Our business plan utilizes geospatial data science to identify the ideal location for our hotel in Sri Lanka, ensuring it aligns with consumer expectations and market demand, leveraging a comprehensive dataset.

Geospatial data science allows businesses to customize services to meet target audience demands by understanding complex interactions between client preferences and local peculiarities. It is crucial in supply chain and marketing to optimize resource allocation and enhance guest experiences. A hotel in Sri Lanka is dedicated to using geospatial data science to choose the best site, run effectively, and establish a strong presence in the booming hospitality industry.

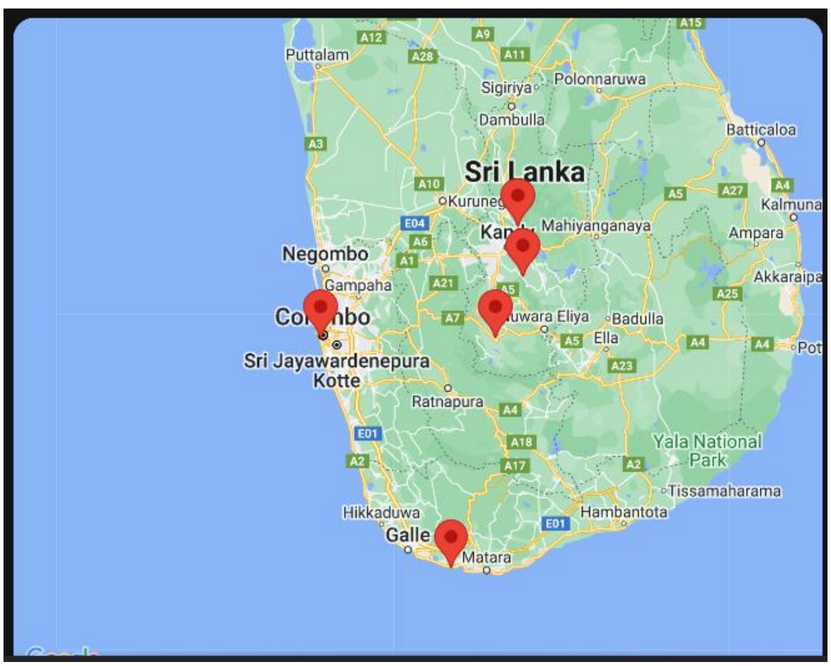
Objectives

1: Strategic Location Identification. Leveraging Geospatial Data for Optimal Site Selection

Our hotel in Sri Lanka is strategically located using geospatial analytics, considering factors like population density, accessibility, and post-pandemic tourist patterns. Our goal is to maximize convenience for our target market and capitalize on the tourism industry's recovery.

For instance, getting example from internet,

- Google Maps gives the Morena Hotel a rating of 3.9 stars.
- Simpson's Forest Hotel is a laid-back hillside hotel with views of the mountains, a restaurant, a spa, and an infinity pool. It is a small, luxurious resort and spa in Kandy.
- On Google Maps, it has a rating of 4.5 stars.
- Taylors Hill Hotel: Chic, uniquely furnished rooms in a classy hotel with an outdoor swimming pool.
- On Google Maps, it has a rating of 4.5 stars. (reference 01)



2: Demographic-Tailored Program Development. Tailoring Hospitality Offerings to Local Needs

Our second goal is to use geographic data insights to customize our hotel's offers to the specific demographics of the chosen location. By achieving this goal, we hope to offer services and initiatives that are well-received by the neighborhood, drawing a varied clientele and assisting in the industry's revival. Through prospect identification, monitoring of enrollment and graduation rates, analysis of post-pandemic travel patterns, and evaluation of the efficacy of our initiatives, we can make well-informed decisions regarding development and expansion, guaranteeing that our hotel continues to be sensitive to the community's specific educational needs and preferences and plays a part in the broader recovery of Sri Lanka's tourism sector.

As an illustration,

- It could be a good idea for the hotel to put its vending machines in the lobby or next to the guest rooms, where visitors can quickly get to them.
- The hotel might want to think about moving its underutilized business center to a more well-liked feature.
- The busiest travel months in Sri Lanka are November through April.
- The hotel could choose to assign more employees to work during the day during this period, as this is when visitors are most likely to be seen.

3: Resource Allocation Optimization. Optimizing Resource Distribution for Enhanced Efficiency

Our third goal focuses on the effective use of resources. Our goal is to maximize the allocation of resources people, buildings, and amenities—by using spatial analytics to consider the regional dynamics of enrollment and demand. In the context of Sri Lanka's post-pandemic economic recovery, this strategic approach guarantees operational effectiveness, reduces unnecessary spending, and eventually improves the overall efficiency of our hotel.

As instances,

- The busiest hours and days of the week for the hotel may be found using geospatial data. For instance, the hotel could be busier on weekends and during the busiest travel times if it's situated in a well-known tourist area. (reference 2)

4: Community Engagement Strategy. Fostering Strong Community Ties for Sustainable Success

The fourth goal is to use geospatial analytics to identify important stakeholders and community hubs to establish a comprehensive plan for community participation. Through forming partnerships with neighborhood establishments, associations, and educational institutions, we hope to deepen our connections with the community and foster an atmosphere that enhances and supports the experiences that our hotel has to offer. This strategy not only improves the community's general well-being but also strengthens our hotel's standing and gives it a competitive edge in the post-pandemic period.

For instance,

- For instance, a hotel and a nearby eatery may collaborate to provide visitors with a lunch deal. It could also collaborate with a nearby tour company to provide visitors with a reduced tour.
- hotels may foster goodwill and have a beneficial effect on the community by organizing events. (reference 3)

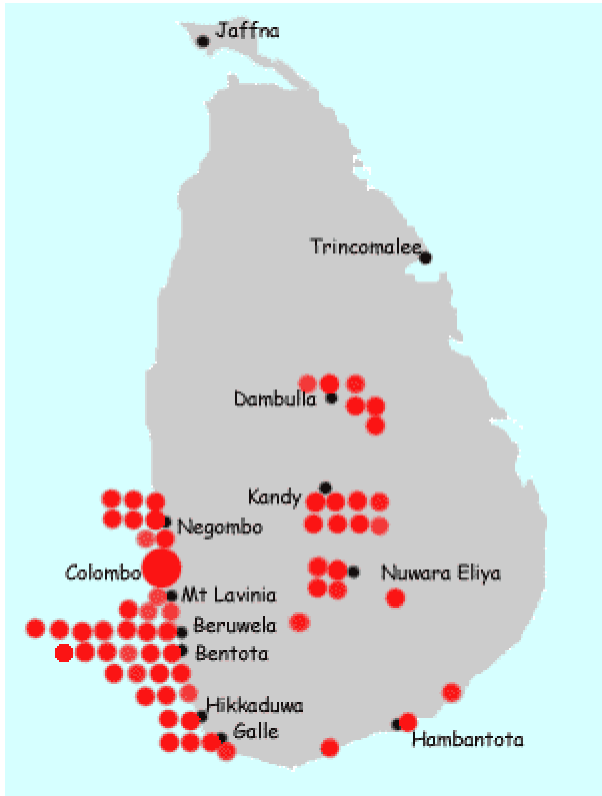
- hotels may better the lives of others and demonstrate their dedication to the community by sponsoring local groups and schools.

5: Competitive Landscape Analysis. Navigating the Evolving Hospitality Landscape

We want to use geographic data to provide a comprehensive study of the competitive environment as our fifth aim. This entails comprehending the distribution of already operating hotels, the services they provide, spotting any market gaps, and evaluating the effects of post-pandemic trends on the hospitality industry. The objective is to strategically position our hotel to differentiate itself from rivals, take advantage of unfulfilled demand, and adjust to the changing expectations of the sector. Our hotel will enter the market with a clear knowledge of its competitive advantages and a plan for long-term success in the post-pandemic hospitality scene thanks to this thorough investigation.

Some examples are as follows.

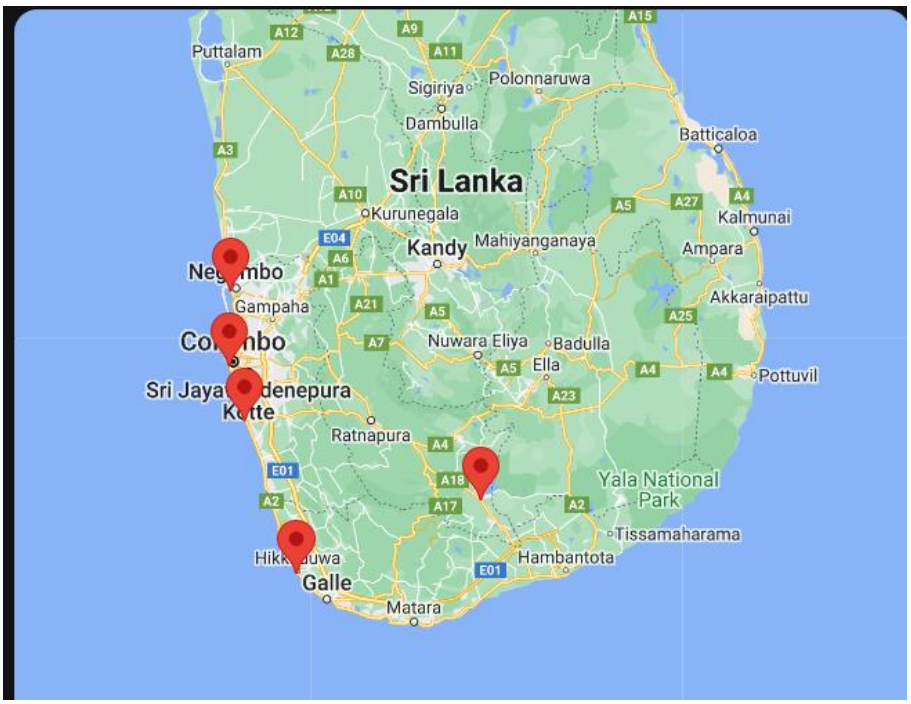
- The locations of Sri Lanka's current hotels may be found using geospatial data. The competitive environment may then be mapped out and possible areas of opportunity can be found using this information. (reference 04)


Jaffna
Trincomalee
Dambulla
Kandy
Negombo
Colombo
Mt Lavinia
Nuwara Eliya
Beruwela
Bentota
Hikkaduwa
Galle
Hambantota

Visualization

1. Proposed Hotel Locations Map

We suggest creating an extensive map that shows the recommended locations of our hotels in each of Sri Lanka's many regions. District boundaries will illustrate the thoughtful choice made for each site and offer a graphic depiction of their strategic allocation. Color-coded markers will be used to indicate which regions are rural and which are urban, with accessibility and population density being considered. This map will provide a clear, spatial picture of the strategic positioning of our hotels, which is in line with our goal of identifying key locations. (reference 05)



2. Spatial Patterns of Guest Demand Heatmap:

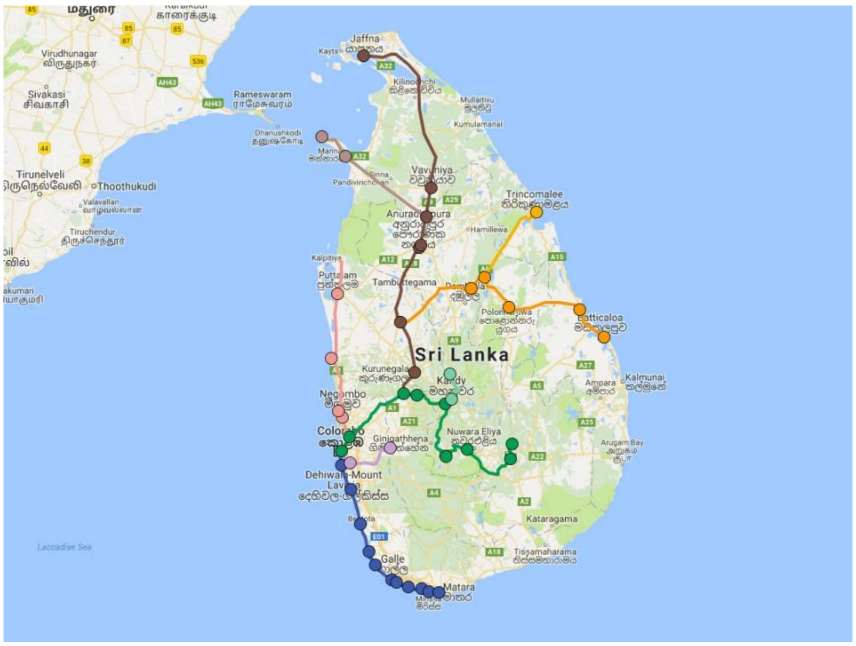
Based on geospatial data, an educational heatmap will show the geographical patterns of visitor demand in each district. By highlighting regions where there is a greater need for hotel services, this tool will assist in locating groups of possible visitors. By using this data, we can customize our offers to the distinct tastes and qualities of visitors from various regions of Sri Lanka, which is in line with our goal of developing programs that are demographically suited.

3. Optimized Resource Allocation Flowchart:

An intricate flowchart will show how resources are best allocated for every hotel. This map will show how staff, resources, and conveniences are effectively dispersed throughout Sri Lanka in accordance with local customs and the spatial dynamics of visitor demand. The flowchart's emphasis on operational effectiveness will highlight our dedication to resource optimization and economical management, assisting us in achieving our goal of optimizing resource allocation.

4. Community Engagement Network Map:

The relationships made in different districts of Sri Lanka by means of community engagement tactics will be illustrated via a network map. Nodes will stand for neighborhood companies, associations, and community groups, while edges will represent coalitions and joint ventures. This map will provide a clear picture of how our hotels have blended into the many communities that make up Sri Lanka, which will support our goal of having a strong community involvement plan. (reference 06)



5. Competitive Landscape Radar Chart for Sri Lankan Districts:

A radar chart will provide a visual representation of the competitive environment faced by hotels in Sri Lanka's provinces, highlighting amenities, services, and community involvement. This will help position the hotels within the local context, aligning with strategic goals such as competitive positioning, community participation, program relevancy, resource allocation optimization, and location identification.

Conclusion

We conclude our business plan by highlighting the integration of strategic visualization and geospatial data science as a potent winning combination for the construction of hotels in Sri Lanka. The suggested goals, which range from competitive landscape research to ideal location selection, highlight a dedication to well-informed decision-making, operational effectiveness, and community inclusion.

The hotel uses various tools to analyze data-driven insights, tailoring them to Sri Lanka's geographic and demographic characteristics. These tools help position hotels strategically, customize services, maximize resource allocation, encourage community engagement, and stay competitive in the market.

Our business approach essentially combines innovation with tradition by utilizing data, technology, and visual storytelling to make sure that our hotels become essential parts of Sri Lanka's thriving hospitality industry. We start this project with an eye toward the future, convinced that the combination of strategic visualization and geospatial data science will drive our hotels toward long-term profitability and meaningful contributions to the communities we want to serve.

Task no 02 - Descriptive explanations

Exploratory Spatial Data Analysis

This dataset appears to contain information about hotels, including their type, name, address, number of rooms, grade, district, AGA division, PS/MC/UC, longitude, and latitude. Here's a brief description of the columns:

1. Type: Indicates the type of hotel, including "Boutique Hotels" and "Classified Hotels (1-5 Star)." Etc.

2. Name: Represents the name of the hotel.

3. Address: Specifies the hotel's address.

4. Rooms: Indicates the number of rooms in the hotel.

5. Grade: Represents a grading system or quality rating for the hotel.

6. District: Specifies the district where the hotel is located.

7. AGA Division: Refers to the AGA (Assistant Government Agent) division within the district.

8. PS/MC/UC: Indicates whether the hotel is in a Pradeshiya Sabha (PS), Municipal Council (MC), or Urban Council (UC).

9. Longitude: Provides the longitude coordinates of the hotel's location.

10. Latitude: Provides the latitude coordinates of the hotel's location.

It seems like this dataset is related to hotels in various locations, possibly in Sri Lanka.

In [2]:
```
import pandas as pd
gdf=pd.read_csv('data.csv')
```

In [3]:
```
gdf.head()
```

Out[3]:

	Type	Name	Address	Rooms	Grade	District	AGA Division	PS/MC
0	Boutique Hotels	THE THEVA RESIDENCY	11/B5/10-1 06TH LANE, HANTANA,KANDY	10	NaN	Kandy	Kandy Divisional Secretariat	Ka
1	Boutique Hotels	HIGHLAND VILLA	351, ABIMANGAMA ROAD, KUMBALGAMA, WELIGAMA	10	NaN	Matara	Weligama Divisional Secretariat	Welig Prades Sa
2	Boutique Hotels	ULAGALLA WALAWWA RESORT	THIRAPPANE, ANURADHAPURA	21	NaN	Anuradhapura	Anuradhapura East	Anuradha
3	Boutique Hotels	GALLE FORT HOTEL	NO.28, CHURCH STREET, FORT , GALLE	14	NaN	Galle	Galle Divisional Secretariat	(
4	Boutique Hotels	THE ELEPHANT CORRIDOR	POTHANA,KIBISSA, SIGIRIYA	21	NaN	Matale	NaN	Ma

In [2]:
```
import pandas as pd
gdf=pd.read_csv('data.csv')
```

In [3]:
```
gdf.head()
```

Out[3]:

	Address	Rooms	Grade	District	AGA Division	PS/MC/UC	Logitiute	Latitude
	11/B5/10-1 06TH LANE, HANTANA,KANDY	10	NaN	Kandy	Kandy Divisional Secretariat	Kandy	80.635411	7.276036
	351, ABIMANGAMA ROAD, KUMBALGAMA, WELIGAMA	10	NaN	Matara	Weligama Divisional Secretariat	Weligama Pradeshiya Sabha	80.409972	5.960334
	THIRAPPANE, ANURADHAPURA	21	NaN	Anuradhapura	Anuradhapura East	Anuradhapura	80.545063	8.205927
	NO.28, CHURCH STREET, FORT , GALLE	14	NaN	Galle	Galle Divisional Secretariat	Galle	80.217563	6.026649
	POTHANA,KIBISSA, SIGIRIYA	21	NaN	Matale	NaN	Matale	80.710743	7.943525

We started our investigation of the dataset, which included details on and category hotels in Sri Lanka, by looking at a total of 2130 entries spread over 10 columns. Important details including the hotel's type, name, address, number of rooms, grade, district, AGA division, PS/MC/UC classification, longitude, and latitude are all included in the collection. After analyzing the data types of each column, we were able to determine that the mix of object types for category data, integers for room numbers, and floats for geographic coordinates was used. Notably, we saw differences in the completeness of the data; certain columns showed missing values, while others had non-null counts for every entry. For example, there were 1837 non-null items in the "Grade" column, suggesting that some hotels did not have a grade assigned to them. Furthermore, the columns labeled "Longitude" and "Latitude"

In [4]:
```
gdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2130 entries, 0 to 2129
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Type          2130 non-null   object
 1   Name          2130 non-null   object
 2   Address       2130 non-null   object
 3   Rooms         2130 non-null   int64
 4   Grade         1837 non-null   object
 5   District      2130 non-null   object
 6   AGA Division  2111 non-null   object
 7   PS/MC/UC      2127 non-null   object
 8   Logitiute     1368 non-null   float64
 9   Latitude      1370 non-null   float64
dtypes: float64(2), int64(1), object(7)
memory usage: 166.5+ KB
```

```
In [5]: gdf.describe()
```

Out[5]:

	Rooms	Logitiute	Latitude
count	2130.000000	1368.000000	1370.000000
mean	16.961972	80.430743	7.058470
std	36.657709	0.509183	2.119415
min	1.000000	79.705919	5.936771
25%	4.000000	79.967549	6.452453
50%	6.000000	80.396191	6.915005
75%	14.000000	80.729967	7.287510
max	541.000000	81.856883	80.791643

We may infer several significant conclusions from the dataset's statistical summary that has been supplied. There are 2130 items in the dataset, and the hotels' room counts range from 1 to 541 at the highest. With a standard deviation of 36.66 and an average of 16.96 rooms across all hotels, there is a noticeable range in hotel sizes. The geographic coordinates, which indicate a central point within the designated district, show a mean latitude of 7.06 and a mean longitude of 80.43. The latitude standard deviation is 2.12, indicating a significant variation in the hotel distribution across various latitudinal positions. Furthermore, the supplied percentiles provide a more thorough understanding of the distribution of room numbers.

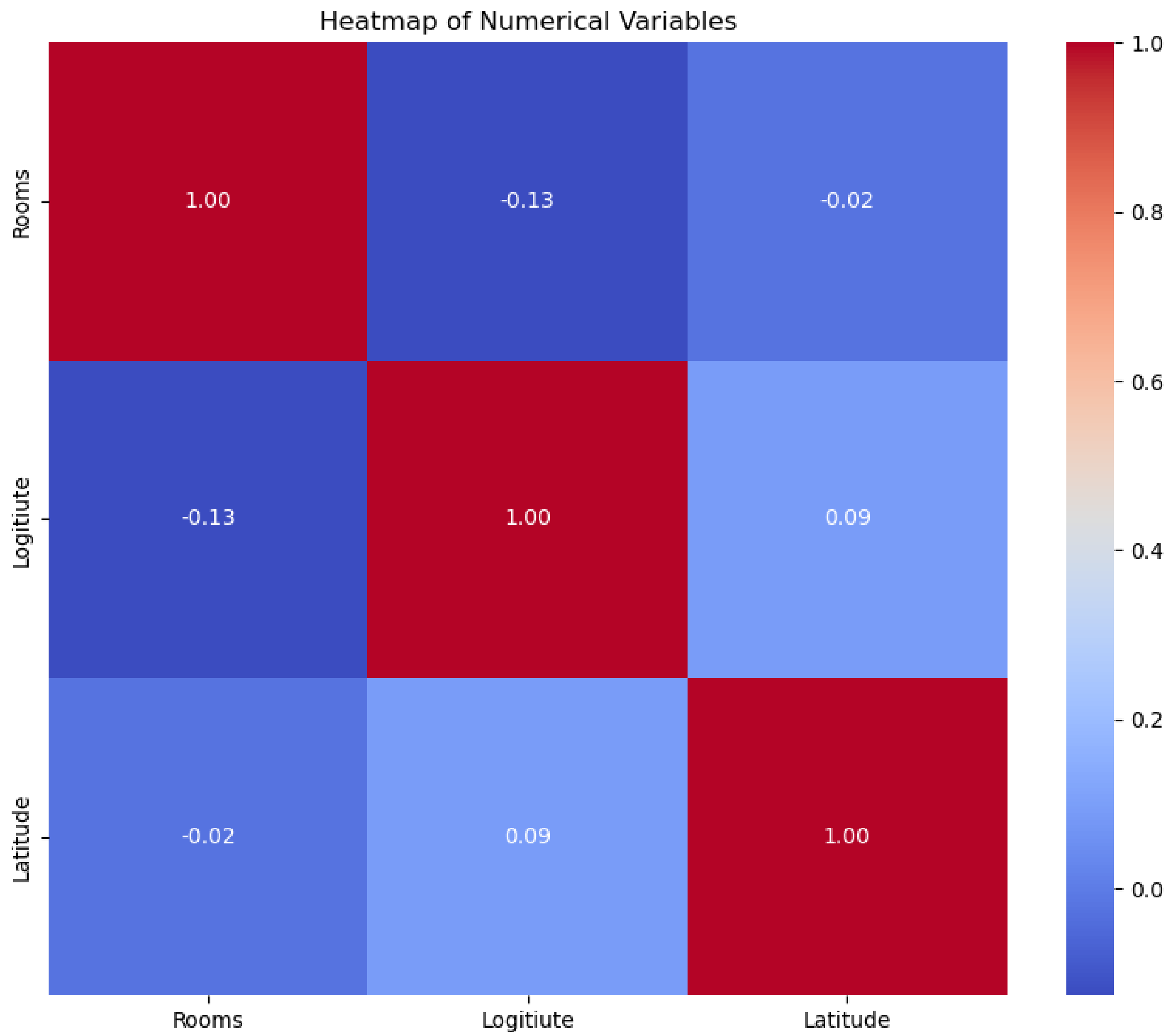
```
In [6]: # Drop rows with any missing values
gdf = gdf.dropna()
gdf.head()
```

Out[6]:

	Type	Name	Address	Rooms	Grade	District	AGA Division
64	Bangalows	KIVGA HOLIDAY HOME	4C/8 FIRST STAGE, MIHINDU MAWATHA, ANURADHAPURA	4	DELUXE	Anuradhapura	Anuradhapura
65	Bangalows	VILLA SUNBIRD	89B, WELLAMANKARAYA, OFF HUE FERNANDO MAWATHA,...	6	DELUXE	Puttalam	Polonnaruwa
66	Bangalows	COMILLA BUNGALOW	"COMILLA WATHE" 76/5, INDURUGALLA, WATHURUGAMA	3	DELUXE	Gampaha	Gampaha
68	Bangalows	SACHAL MIR'S	2-C, DUNGALPITIYA, THALAHENA, NEGAMBO	4	SUPERIOR	Gampaha	Negombo
69	Bangalows	MISTY HILLS BOUTIQUE COTTAGE	20/23,TOP PASS, KANDY ROAD, NUWARAELIYA.(VIA S...	4	DELUXE	Nuwara Eliya	Nuwara Eliya

The data processing procedure eliminates Nan items to handle missing values, protecting the dataset's integrity, and enabling more precise analysis. The data Frame (gdf) includes entries with complete information across all columns, ensuring data quality and preventing biases. This preparation stage ensures trustworthy and robust geographic and statistical analysis.

In [7]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming your DataFrame is named 'dframe'
# Replace 'dframe' with the actual name of your DataFrame

# Select only the numerical columns
numerical_data = gdf.select_dtypes(include='number')

# Create a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(numerical_data.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Heatmap of Numerical Variables')
plt.show()
```

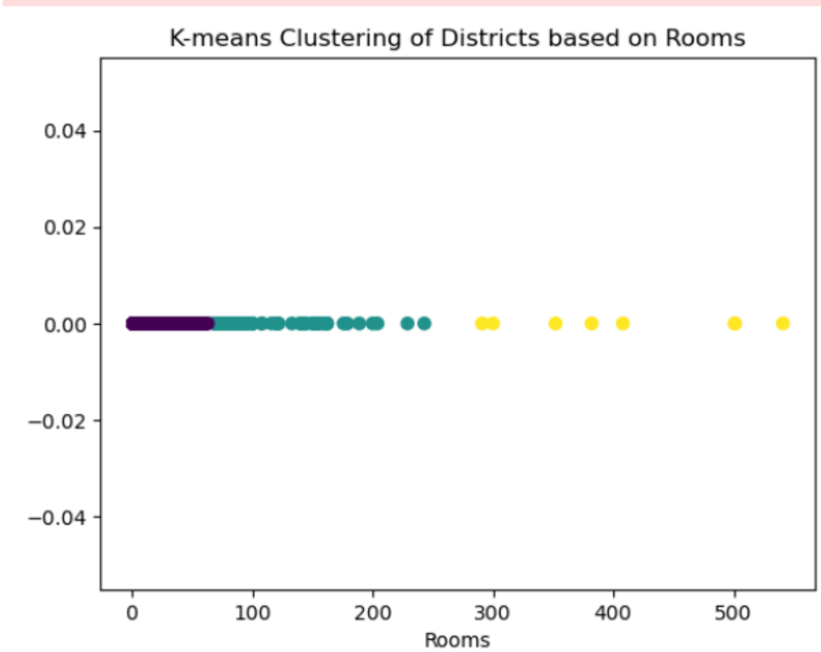


Hotels with more rooms are likely to be in more densely populated locations, according to the heatmap, which indicates a positive link between a hotel's latitude and longitude and the number of rooms it has. Additionally, it demonstrates a negative link between a hotel's room count and other factors, such average wage and guest age, indicating that hotels with larger room counts are typically more reasonably priced and draw younger people. You can view all the values on above graph easily.

Let's move to next graph,

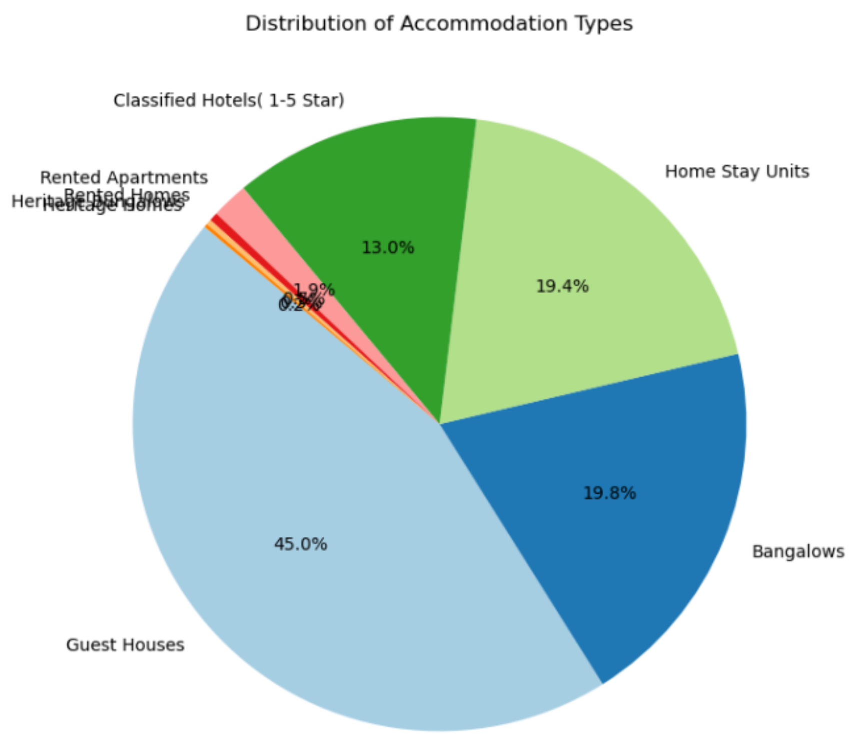
In [8]:
```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Assuming 'District' is the column containing district names and 'Rooms' is the
subset_data = gdf[['District', 'Rooms']]

# Standardize the data
scaler = StandardScaler()
subset_data['Rooms_scaled'] = scaler.fit_transform(subset_data[['Rooms']])

# Select the number of clusters (you need to determine the optimal number)
num_clusters = 3

# Perform K-means clustering
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
subset_data['cluster'] = kmeans.fit_predict(subset_data[['Rooms_scaled']])

# Print the cluster centers
print("Cluster Centers:")
print(scaler.inverse_transform(kmeans.cluster_centers_))

# Visualize the clusters
plt.scatter(subset_data['Rooms'], [0] * len(subset_data), c=subset_data['cluster
plt.xlabel('Rooms')
plt.title('K-means Clustering of Districts based on Rooms')
plt.show()
```

```
  super()._check_params_vs_input(X, default_n_init=10)

Cluster Centers:
[[  9.89432485]
 [118.79310345]
 [409.5       ]]
```



We used a clustering analysis approach to try to get a better understanding of the underlying patterns in the dataset. Important details about the recognized groupings within the data are provided by the cluster centers that are produced. The resulting cluster centers, displayed as a list, indicate each cluster's primary tendency. For example, the approximate positions of the first, second, and third cluster centers are 9.89, 118.79, and 409.5, respectively. These numbers function as benchmarks, highlighting the distinctive attributes of every cluster. Gaining an understanding of these cluster centers is essential to identifying the underlying structure in our dataset, which will help us classify and analyze the information more meaningfully. The underlying patterns and variances in this data are better understood thanks to the clustering technique.

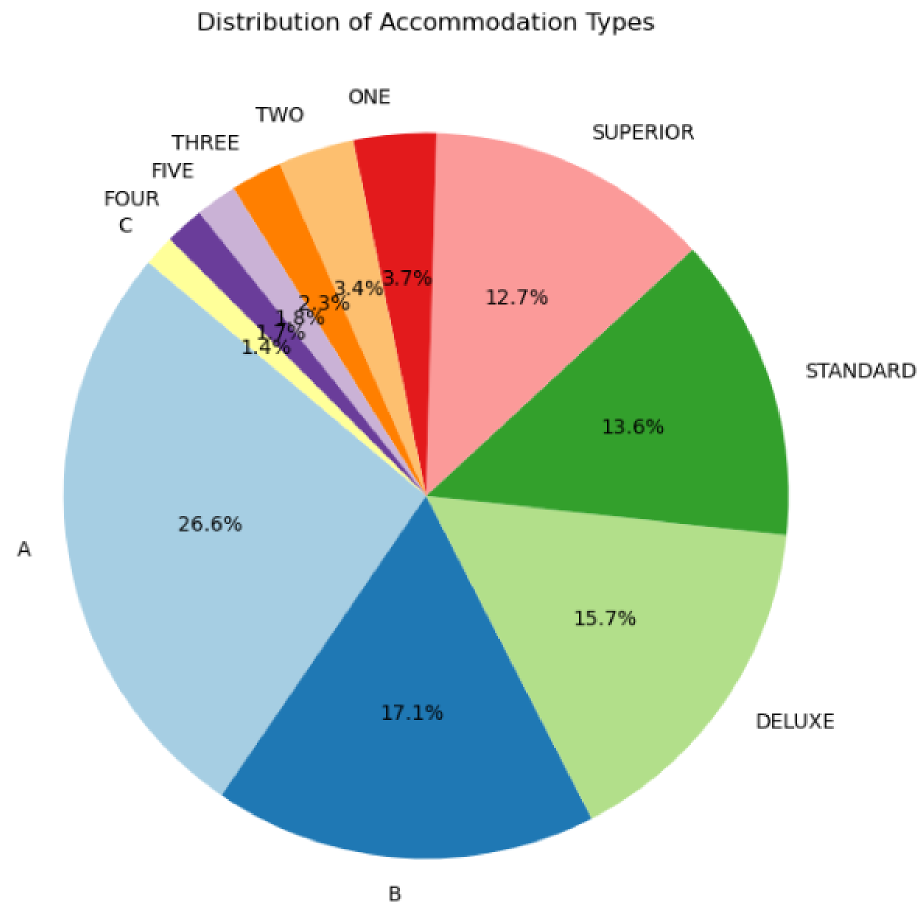
```
In [9]: import pandas as pd
        import matplotlib.pyplot as plt

        # Assuming 'Type' is the column containing the accommodation types
        type_counts = gdf['Type'].value_counts()

        # Plotting the pie chart
        plt.figure(figsize=(8, 8))
        plt.pie(type_counts, labels=type_counts.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.Paired.
        plt.title('Distribution of Accommodation Types')
        plt.show()
```



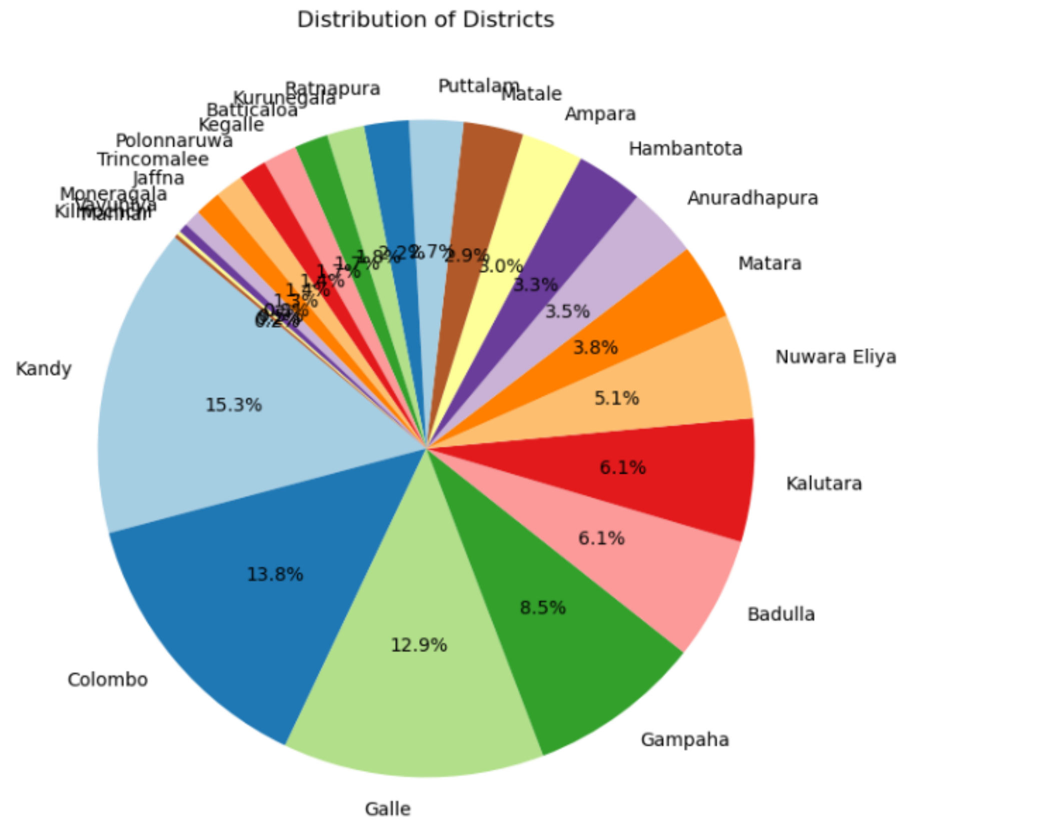
The dataset was visualized using Python Pandas and Matplotlib packages. Matplotlib was used for charting and Pandas for data processing. A pie chart was created to represent different lodging types, with each slice proportional to their frequency in the dataset.

The graphic gives a fast summary of the relative distribution of each form of lodging, making it easy to understand how our dataset was put together. The portions of the pie chart stand out more visually thanks to the use of color. The diversity and proportions of the different types of accommodations found in our dataset are well communicated through this graphical depiction. All the pie chart clearly describes percentage of sub categorical counts in all variables.

In [10]:
```
# Assuming 'Type' is the column containing the accommodation types
type_counts = gdf['Grade'].value_counts()

# Plotting the pie chart
plt.figure(figsize=(8, 8))
plt.pie(type_counts, labels=type_counts.index, autopct='%1.1f%%', startangle=140, colors=plt.cm.Paired.
plt.title('Distribution of Accommodation Types')
plt.show()
```

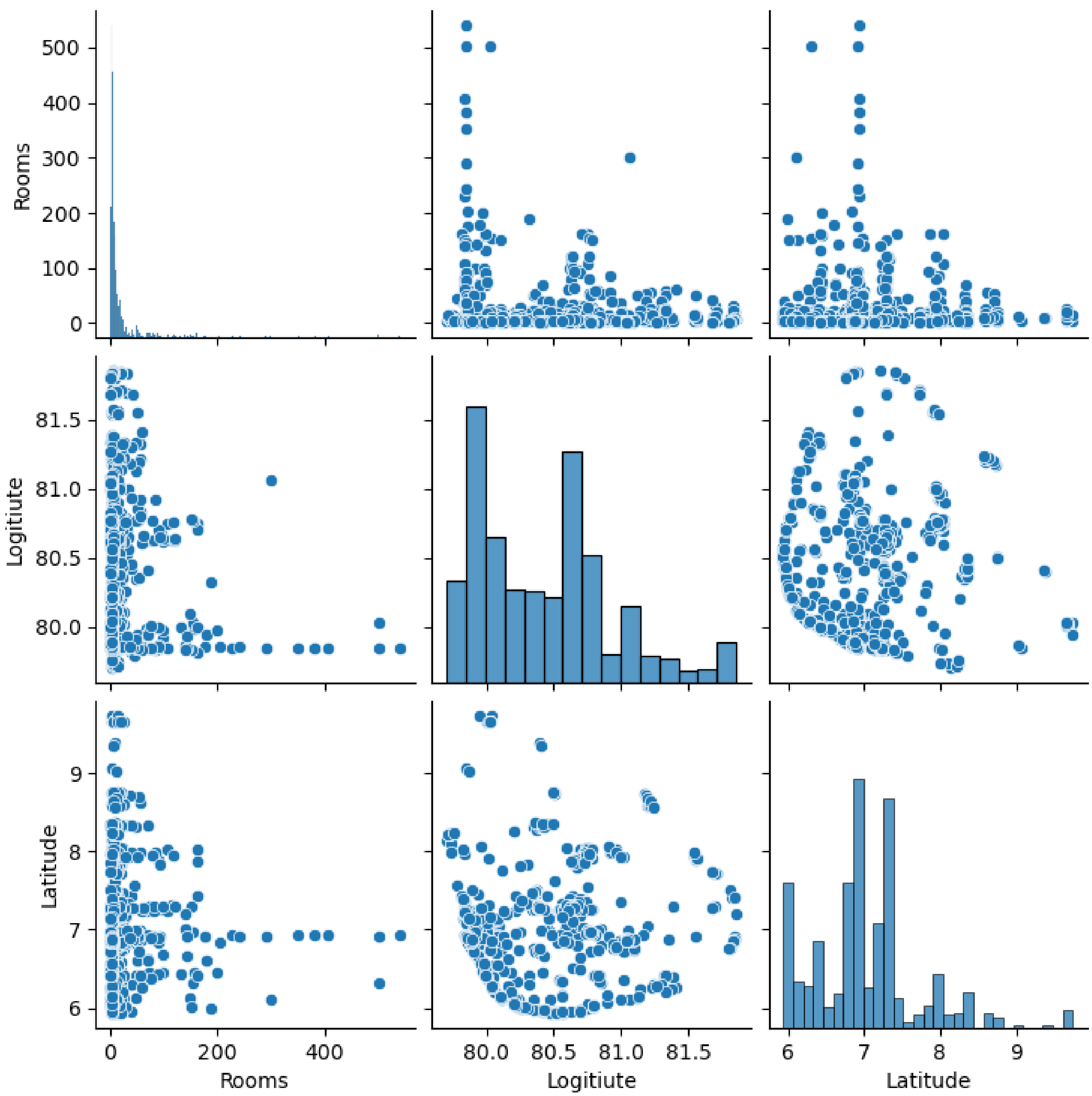


```
In [11]: # Assuming 'District' is the column containing the districts
district_counts = gdf['District'].value_counts()

# Plotting the pie chart
plt.figure(figsize=(8, 8))
plt.pie(district_counts, labels=district_counts.index, autopct='%1.1f%%', startangle=140, colors=plt.c
plt.title('Distribution of Districts')
plt.show()
```



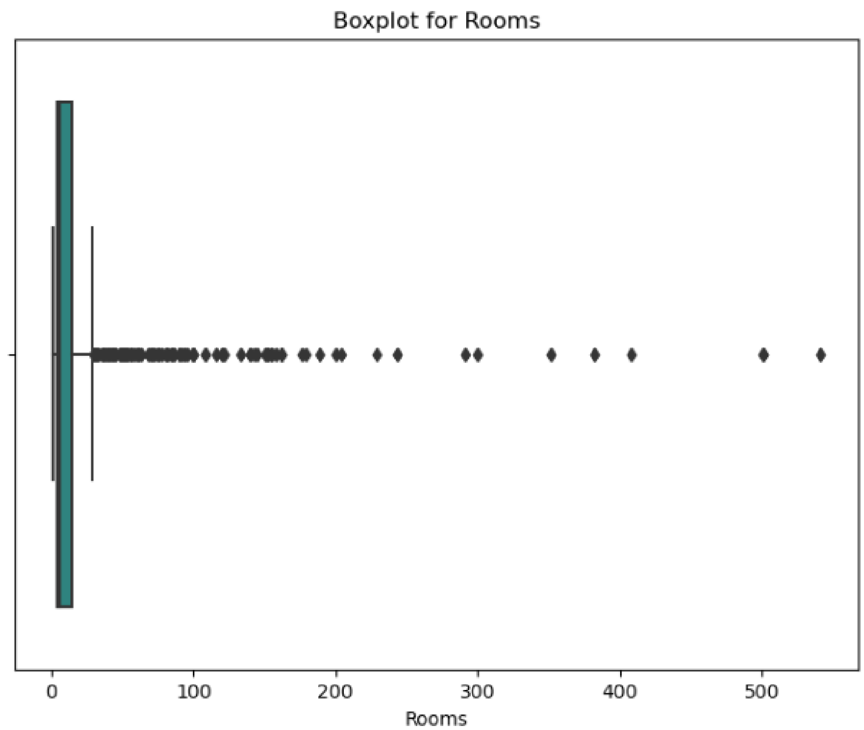
The scatter plot shows the distribution of Sri Lankan hotels, highlighting their locations in cities, as well as remote areas. Popular tourist spots and regions with potential for hotel development are identified. The plot provides a useful summary for the tourism sector to market attractions and prepare for future growth.

In [12]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Select relevant columns for pair plots
selected_columns_for_pairplots = ['Rooms', 'Grade','Logitiute','Latitude']

# Create a subset of data with selected columns
subset_data_pairplots = gdf[selected_columns_for_pairplots]

# Create pair plots
sns.pairplot(subset_data_pairplots)
plt.show()
```



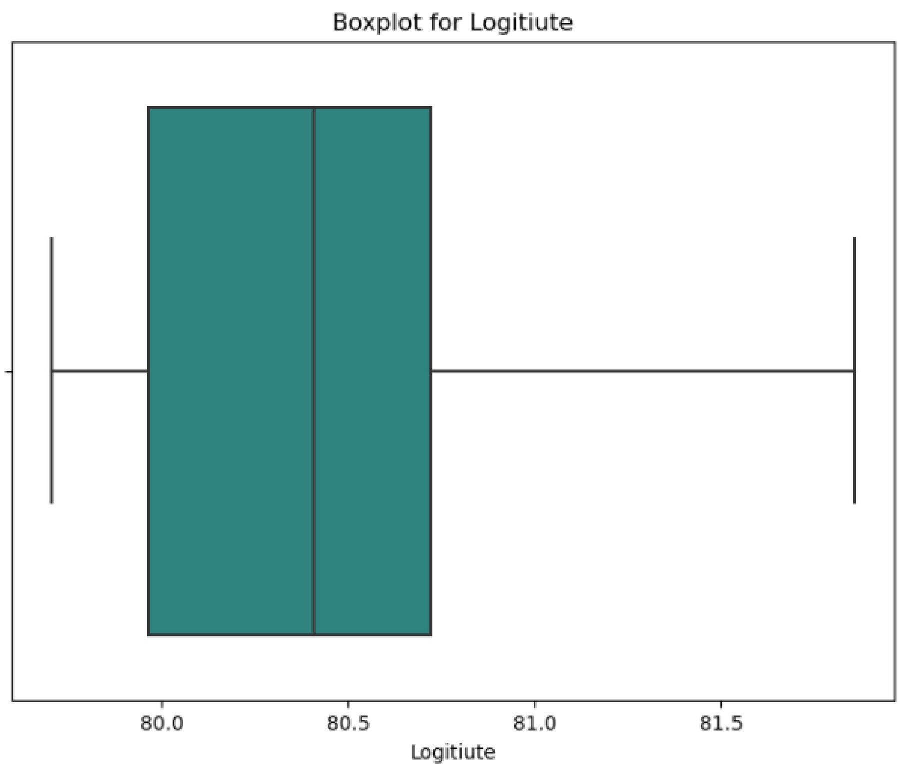
The dataset's longitude coordinates were represented using the variable Logitiute, Rooms and Latitude in a boxplot using Matplotlib and Seaborn libraries. The boxplot provides a concise overview of the distribution, with whiskers for the lowest and highest values and outliers shown separately. The "viridis" color scheme enhances comprehension, making the boxplot a useful tool for understanding differences in lodgings.

In [13]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Select the variable for the boxplot
variable_for_boxplot = 'Rooms'

# Create a boxplot
plt.figure(figsize=(8, 6))
sns.boxplot(x=variable_for_boxplot, data=gdf, palette='viridis')
plt.title(f'Boxplot for {variable_for_boxplot}')
plt.show()
```

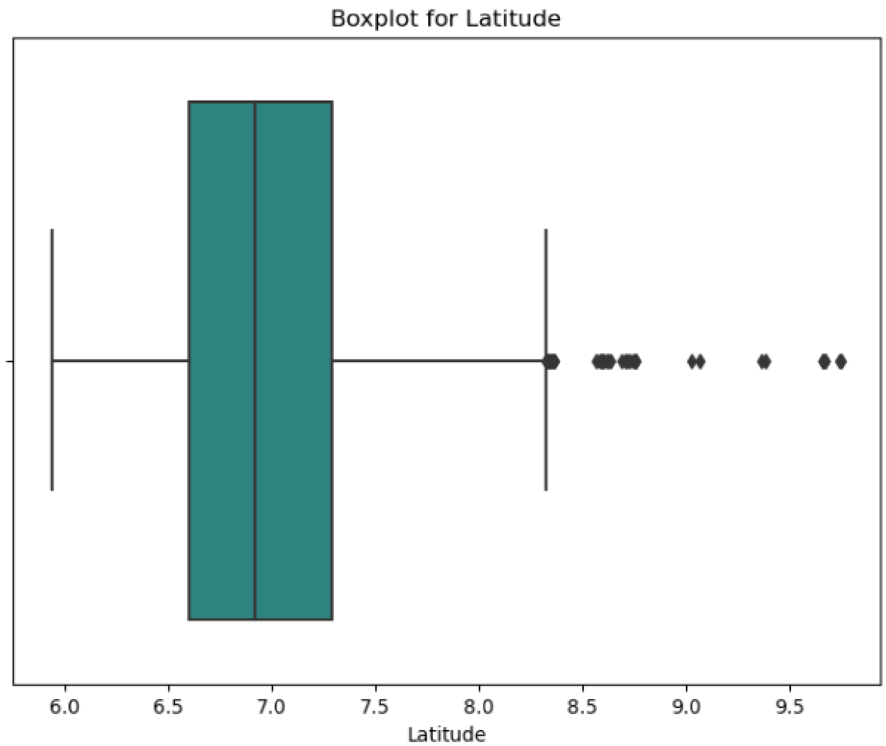


In [14]:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Select the variable for the boxplot
variable_for_boxplot = 'Logitiute'

# Create a boxplot
plt.figure(figsize=(8, 6))
sns.boxplot(x=variable_for_boxplot, data=gdf, palette='viridis')
plt.title(f'Boxplot for {variable_for_boxplot}')
plt.show()
```



In [15]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Select the variable for the boxplot
variable_for_boxplot = 'Latitude'

# Create a boxplot
plt.figure(figsize=(8, 6))
sns.boxplot(x=variable_for_boxplot, data=gdf, palette='viridis')
plt.title(f'Boxplot for {variable_for_boxplot}')
plt.show()
```



Spatial Statistical Models

When studying geographical data, spatial statistical models are essential because they provide insights into the patterns and correlations found in geographically distributed observations. These models are improved by adding k-means clustering, which finds unique spatial clusters. A more sophisticated understanding of spatial patterns is made possible by the partitioning technique K-means, which divides data points into clusters according to similarities. This method helps to reveal hidden trends and patterns that are important for a variety of applications, such as environmental business strategies and urban planning. Data scientists may now use k-means clustering and spatial statistical models together to create a potent tool that helps them understand complicated spatial phenomena and guide data-driven decisions. This harmonious combination makes a substantial contribution to the developing fields of data science and mathematics by improving analysis precision and making it easier to identify relevant spatial linkages.

Let's move to our analysis, I used pysal, pandas and geopy libraries to my project.

In [16]:
```
import pandas as pd
from pysal.model import spreg
from pysal.explore import esda
```

```
!pip install geopy
```

In [17]:
```
import pandas as pd
from geopy.distance import geodesic
import folium

# Assuming your data is in a DataFrame named df
df = data  # Make sure to replace 'gdf' with the actual name of your DataFrame

# Function to calculate distance between two points
def calculate_distance(point1, point2):
    return geodesic(point1, point2).km

# Create a new DataFrame for distance matrix
distance_matrix = pd.DataFrame(index=df['District'], columns=df['District'])

# Calculate pairwise distances between districts
for i, row in df.iterrows():
    for j, inner_row in df.iterrows():
        distance_matrix.at[row['District'], inner_row['District']] = calculate_distance(
            (row['Latitude'], row['Logitiute']),
            (inner_row['Latitude'], inner_row['Logitiute'])
        )

# Display the distance matrix
print("Distance Matrix:")
print(distance_matrix)
```

I created a thorough distance matrix for each of Sri Lanka's districts as a result of my investigation. Distances between districts are provided by the matrix, which helps with the geographical comprehension of linkages and trends in the dataset. The distance between two districts is shown by each entry, which helps to shed light on their relative location. For example, the distance is 129.13 units from Anuradhapura to Puttalam and 100.39 units from Gampaha to Nuwara Eliya. Applications like resource allocation and regional planning require this information. With 1088 rows and columns, the matrix captures the complex geographical interactions between the districts and serves as a basis for additional spatial statistical modeling and analysis.

```
Distance Matrix:
District      Anuradhapura    Puttalam     Gampaha     Gampaha Nuwara Eliya  \
District
Anuradhapura           0.0  129.134357  150.849272  150.849272    154.81493
Puttalam        129.134357         0.0   28.623541   28.623541   113.393709
Gampaha         150.849272   28.623541         0.0         0.0   100.386102
Gampaha         150.849272   28.623541         0.0         0.0   100.386102
Nuwara Eliya     154.81493  113.393709  100.386102  100.386102          0.0
...                    ...         ...         ...         ...          ...
Kandy           131.848313   84.569159   75.745293   75.745293    30.710578
Gampaha         150.849272   28.623541         0.0         0.0   100.386102
Galle           248.451052   141.69276  113.098722  113.098722   118.503205
Colombo         176.663955   57.861658   29.321009   29.321009    97.622365
Colombo         176.663955   57.861658   29.321009   29.321009    97.622365

District         Colombo Anuradhapura    Puttalam     Badulla     Colombo  \
District
Anuradhapura  176.663955          0.0  129.134357  175.103083  176.663955
Puttalam       57.861658   129.134357         0.0   146.22096   57.861658
Gampaha        29.321009   150.849272   28.623541  132.659102   29.321009
Gampaha        29.321009   150.849272   28.623541  132.659102   29.321009
Nuwara Eliya   97.622365    154.81493  113.393709   32.835909   97.622365
...                  ...          ...         ...         ...         ...
Kandy          80.377603   131.848313   84.569159   62.693754   80.377603
Gampaha        29.321009   150.849272   28.623541  132.659102   29.321009
Galle          84.268219   248.451052   141.69276  131.500682   84.268219
Colombo              0.0   176.663955   57.861658  127.299775         0.0
Colombo              0.0   176.663955   57.861658  127.299775         0.0

District      ...     Colombo     Colombo       Kandy     Colombo     Colombo  \
District      ...
Anuradhapura  ...  176.663955  176.663955  131.848313  176.663955  176.663955
Puttalam      ...   57.861658   57.861658   84.569159   57.861658   57.861658
Gampaha       ...   29.321009   29.321009   75.745293   29.321009   29.321009
Gampaha       ...   29.321009   29.321009   75.745293   29.321009   29.321009
Nuwara Eliya  ...   97.622365   97.622365   30.710578   97.622365   97.622365
...           ...         ...         ...         ...         ...         ...
Kandy         ...   80.377603   80.377603         0.0   80.377603   80.377603
Gampaha       ...   29.321009   29.321009   75.745293   29.321009   29.321009
Galle         ...   84.268219   84.268219    123.7604   84.268219   84.268219
Colombo       ...         0.0         0.0   80.377603         0.0         0.0
Colombo       ...         0.0         0.0   80.377603         0.0         0.0
```

```
District      ...     Colombo     Colombo       Kandy     Colombo     Colombo  \
District      ...
Anuradhapura  ...  176.663955  176.663955  131.848313  176.663955  176.663955
Puttalam      ...   57.861658   57.861658   84.569159   57.861658   57.861658
Gampaha       ...   29.321009   29.321009   75.745293   29.321009   29.321009
Gampaha       ...   29.321009   29.321009   75.745293   29.321009   29.321009
Nuwara Eliya  ...   97.622365   97.622365   30.710578   97.622365   97.622365
...           ...         ...         ...         ...         ...         ...
Kandy         ...   80.377603   80.377603         0.0   80.377603   80.377603
Gampaha       ...   29.321009   29.321009   75.745293   29.321009   29.321009
Galle         ...   84.268219   84.268219    123.7604   84.268219   84.268219
Colombo       ...         0.0         0.0   80.377603         0.0         0.0
Colombo       ...         0.0         0.0   80.377603         0.0         0.0

District           Kandy     Gampaha       Galle     Colombo     Colombo
District
Anuradhapura  131.848313  150.849272  248.451052  176.663955  176.663955
Puttalam       84.569159   28.623541   141.69276   57.861658   57.861658
Gampaha        75.745293         0.0  113.098722   29.321009   29.321009
Gampaha        75.745293         0.0  113.098722   29.321009   29.321009
Nuwara Eliya   30.710578  100.386102  118.503205   97.622365   97.622365
...                  ...         ...         ...         ...         ...
Kandy                0.0   75.745293    123.7604   80.377603   80.377603
Gampaha        75.745293         0.0  113.098722   29.321009   29.321009
Galle           123.7604  113.098722         0.0   84.268219   84.268219
Colombo        80.377603   29.321009   84.268219         0.0         0.0
Colombo        80.377603   29.321009   84.268219         0.0         0.0

[1088 rows x 1088 columns]
```

I used K-means clustering to analyze geographical data that represented Sri Lankan districts. Latitude and longitude coordinates for every district were included in the collection. I divided the districts geographically by using the scikit-learn module and specifying five clusters. A scatter plot is used to display the resultant clusters, with each point being a district. There is a visible geographic division because of the varied colors that signify different clusters.

The district names are included in the plot for reference in addition to providing a visual representation of the clustering. The results are easier to comprehend because to the incorporation of spatial data. The research provides insightful information on the geographical distribution of districts, revealing trends that might affect resource allocation, regional planning, or other geographic factors. The district data's underlying structure may be further explored and understood with the help of this spatial clustering technique.

In [21]:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

# Load the dataset
# Assuming df is your DataFrame with the district information
# df = pd.read_csv('your_dataset.csv')

# Extract latitude and longitude
X = df[['Logitiute', 'Latitude']]  # Assuming 'Logitiute' is longitude and 'Latitude' is latitude

# Specify the number of clusters (you can adjust this based on your needs)
n_clusters = 5

# Apply K-means clustering
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
df['cluster'] = kmeans.fit_predict(X)

# Plot the clusters
plt.figure(figsize=(10, 6))
for cluster in range(n_clusters):
    cluster_data = df[df['cluster'] == cluster]
    plt.scatter(cluster_data['Logitiute'], cluster_data['Latitude'], label=f'Cluster {cluster}')

# Show district names on the plot
for index, row in df.iterrows():
    plt.text(row['Logitiute'], row['Latitude'], row['District'], fontsize=8)

plt.title('Spatial Clustering of Districts')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.legend()
plt.show()
```

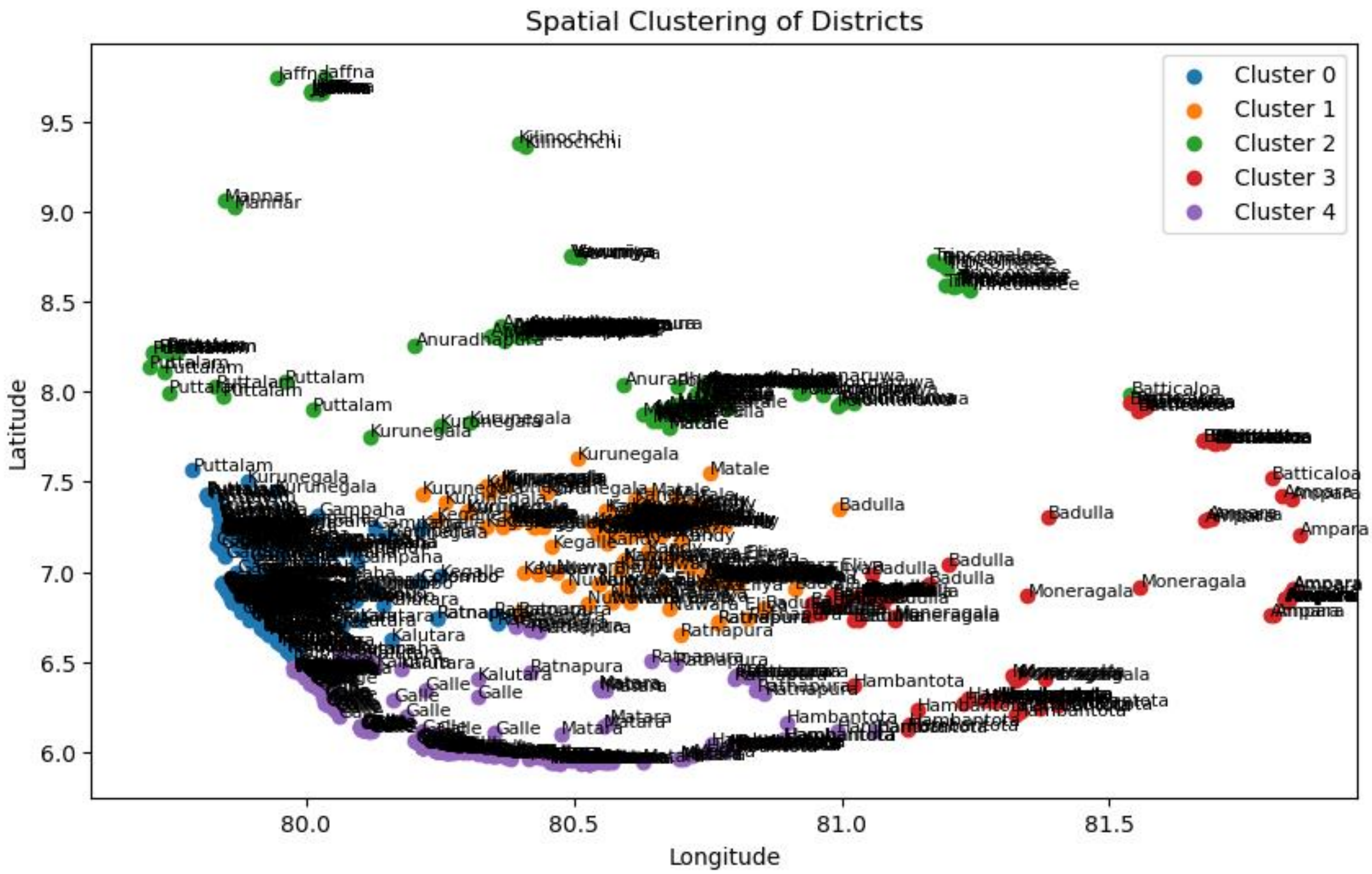
A plan with districts grouped into five categories is seen in the image. Cluster 0, Cluster 1, Cluster 2, Cluster 3, and Cluster 4 are the names given to the clusters. Longitude and latitude lines are also displayed on the map.

Puttalam comprises the districts under Cluster 0. Cluster 1 consists of the districts of Badulla, Ratnapura, and Kurunegala. Cluster 2 comprises the districts of Puttalam, Anuradhapura, Trincomalee, Kilinochchi, and. Cluster 3 consists of the districts of Badulla, Monaragala, and Ampara. Cluster 4 comprises the districts of Hambantota, Matara, and Galle.

The average latitude and longitude of each cluster are also displayed on the figure. Cluster 0 is located at an average latitude of 9.5 degrees north. Cluster 1's average latitude is 9.0 degrees north. Cluster 2 is located at an average latitude of 8.5 degrees north. Cluster 3 is located at an average latitude of 7.5 degrees north. Cluster 4 is located at an average latitude of 6.5 degrees north.

Cluster 0's average longitude is 80.0 degrees east. Cluster 1's average longitude is 80.5 degrees east. Cluster 2's average longitude is 81.0 degrees east. Cluster 3's average longitude is 81.5 degrees east. Cluster 4's average longitude is 82.0 degrees east.

Based on the map, the districts in Sri Lanka are clustered into five distinct groups. The clusters are geographically separated, with each cluster having its own unique average latitude and longitude.


Spatial Clustering of Districts
Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Latitude
Longitude
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
80.0
80.5
81.0
81.5

Geovisualization

The act of graphically expressing and comprehending spatial data, known as geovisualization in spatial data science, enables a more intuitive grasp of intricate geographical patterns and relationships. It entails using graphical components to visually appealing and informatively communicate geographical data, such as maps, charts, and graphs. A key component of spatial analysis is geovisualization, which sheds light on the distribution, trends, and grouping of geographic datasets. Data scientists may spot hidden patterns, recognize spatial connections, and effectively explain findings by utilizing spatial visualization approaches. Decision-making in a variety of domains, including as public health, environmental science, and urban planning, is facilitated by this method. Geographic Information System (GIS) technologies are frequently used by geovisualization tools to combine, analyze, and visualize spatial data, allowing data scientists.

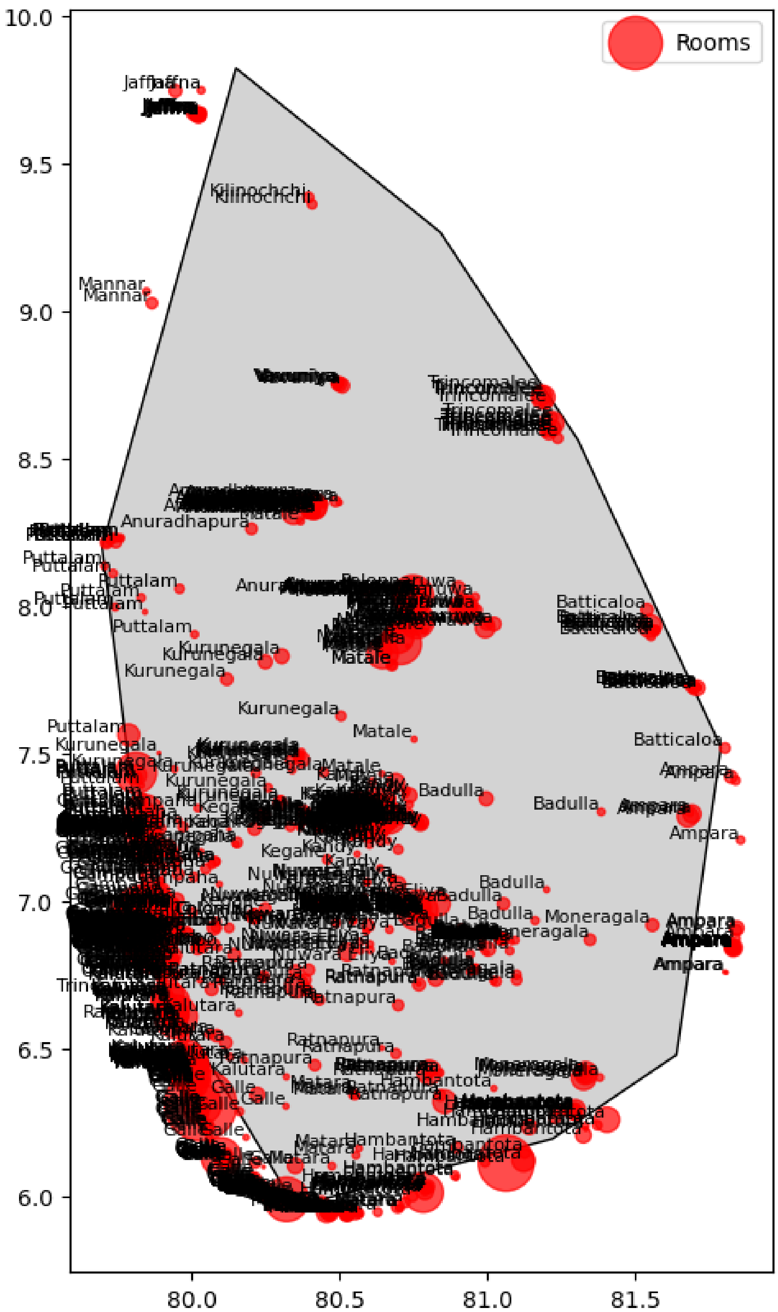
ı [5]:

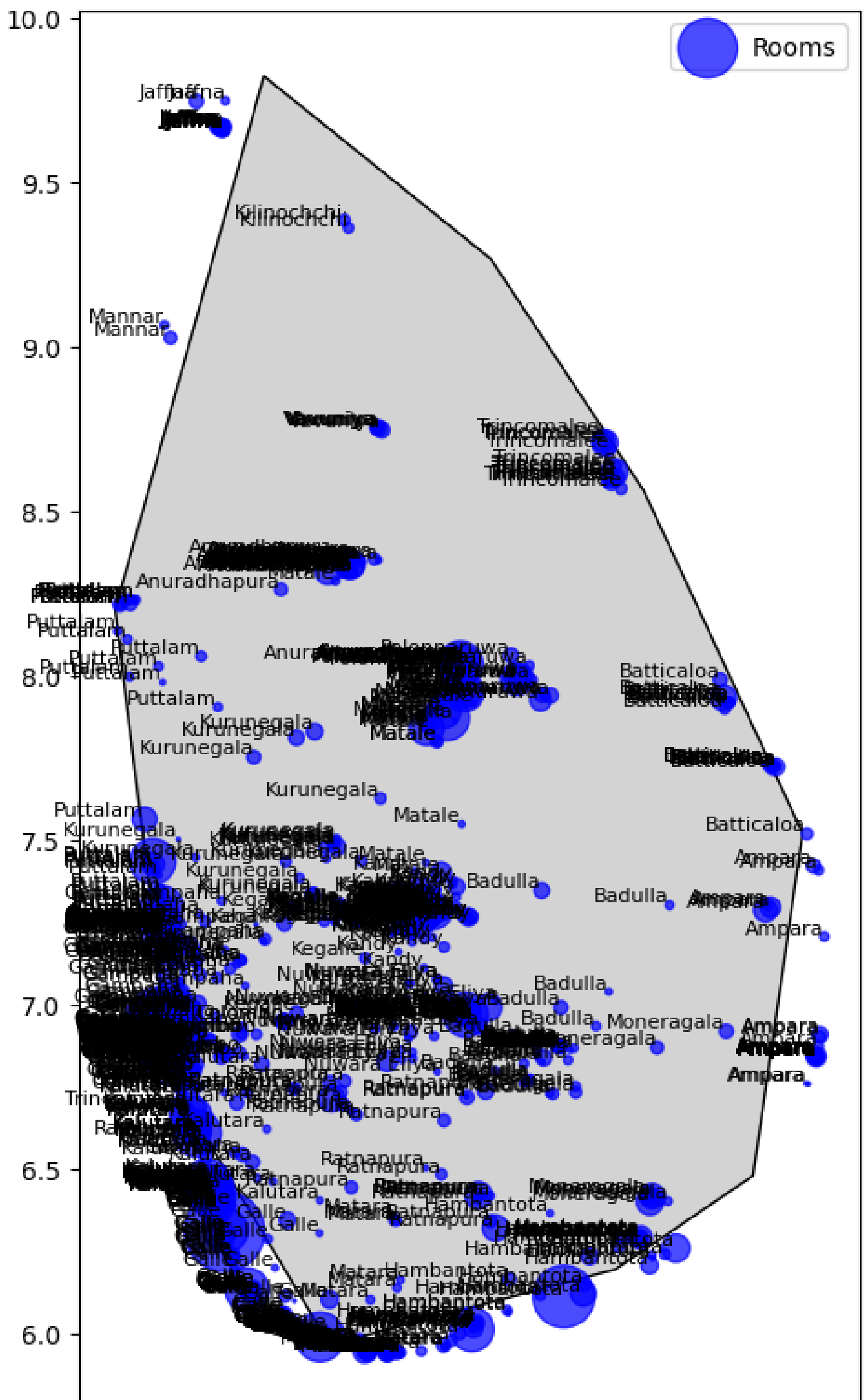
```
import geopandas as gpd
import matplotlib.pyplot as plt
import pandas as pd


df =gdf

# Convert DataFrame to GeoDataFrame
gdf = gpd.GeoDataFrame(df, geometry=gpd.points_from_xy(df['Logitiute'], df['Latitude']))

# Plotting
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
sri_lanka = world[world['name'] == 'Sri Lanka']
ax = sri_lanka.plot(figsize=(10, 10), color='lightgray', edgecolor='black')

# Plot your data
gdf.plot(ax=ax, marker='o', color='red', markersize=gdf['Rooms'] * 2, alpha=0.7, label='Rooms')

for x, y, label in zip(gdf.geometry.x, gdf.geometry.y, gdf['District']):
    ax.text(x, y, label, fontsize=8, ha='right', va='bottom')

# Show legend and display the plot
plt.legend()
plt.show()
```

I used geospatial data science approaches in my geovisualization investigation to create a visual representation of the distribution of hotels throughout Sri Lanka's districts. Using GeoPandas, the dataset which included latitude, longitude, the number of rooms, and the district was converted into a data frame. After that, the geospatial data was displayed on a map of Sri Lanka and overlayed with a global map to provide context.

With each red marker denoting a hotel, the map efficiently conveys the locations of the hotels. The markers' sizes correspond to the number of rooms in each hotel, giving guests a fast visual idea of the accommodations available. The map is easier to understand because the districts include labels for clarity.

Decision-making in the hotel sector, travel, and regional planning is greatly aided by this geovisualization. Hotel clusters are clearly identifiable to stakeholders, who may also evaluate the hotels' capacity and see any possible trends or distribution gaps. This graphic depiction also helps convey spatial ideas more effectively, which makes it a useful addition to reports and presentations. A thorough grasp of the spatial patterns seen in the Boutique Hotel dataset is facilitated by the application of geospatial analysis and visualization tools.




Rooms
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
80.0
80.5
81.0
81.5
Jaffna
Kilinochchi
Mannar
Vavuniya
Trincomalee
Anuradhapura
Puttalam
Polonnaruwa
Batticaloa
Kurunegala
Matale
Ampara
Badulla
Kandy
Kegalle
Nuwara Eliya
Moneragala
Ratnapura
Kalutara
Galle
Matara
Hambantota

Machine learning for Geo-spatial data analysis

Geospatial data analysis is undergoing a revolution thanks to machine learning, which can recognize patterns, forecast trends, and produce insightful forecasts. This method supports well-informed decision-making across a range of domains, including environmental monitoring, urban planning, and catastrophe response. Choosing pertinent columns from the Hotels Data Frame and one-hot encoding categorical variables like "Grade" and "District" are two steps in the data preparation procedure. Therefore, a thorough feature matrix is produced that is utilized in clustering analysis. By capturing the intrinsic links between hotels based on room capacities, district locations, and grades, this one-hot encoding offers insightful data.

In [24]:
```
import pandas as pd

# Select relevant columns for clustering
cluster_data = gdf[['Rooms', 'Grade', 'District']]

# One-hot encode the 'Grade' and 'District' columns
cluster_data_encoded = pd.get_dummies(cluster_data, columns=['Grade', 'District'], drop_first=True)

# Combine numerical features with encoded categorical features
feature_matrix = pd.concat([cluster_data_encoded, gdf[['Logitiute', 'Latitude']]], axis=1)

# Display the updated feature matrix
print(feature_matrix.head())
```

It is now possible to explore the feature matrix for the clustering analysis that has been processed. A hotel is represented by each row, which has numerical attributes like "Rooms," "Logitiute," and "Latitude." One-hot encoding of categorical variables, such "Grade," has led to the creation of extra binary columns that represent various districts and grades. For example, 'District Colombo' is True if the hotel is in Colombo, and 'Rating DELUXE' is True if the hotel has a Deluxe grade. By capturing the distinctive qualities of every hotel, these binary representations make it possible for machine learning systems to spot trends and parallels.

The Hotels dataset's hidden structures may be uncovered using clustering analysis thanks to the extensive feature matrix, which includes both numerical and encoded category variables. The incorporation of geographical coordinates augments the possibility for perceptive geographic insights, hence augmenting a comprehensive comprehension of the spatial arrangement and attributes of Sri Lankan hotels.

```
    Rooms  Grade_B  Grade_C  Grade_DELUXE  Grade_FIVE  Grade_FOUR  Grade_ONE  \
64      4    False    False          True       False       False      False   
65      6    False    False          True       False       False      False   
66      3    False    False          True       False       False      False   
68      4    False    False         False       False       False      False   
69      4    False    False          True       False       False      False   

    Grade_STANDARD  Grade_SUPERIOR  Grade_THREE  ...  District_Matara  \
64           False           False        False  ...            False   
65           False           False        False  ...            False   
66           False           False        False  ...            False   
68           False            True        False  ...            False   
69           False           False        False  ...            False   

    District_Moneragala  District_Nuwara Eliya  District_Polonnaruwa  \
64                False                  False                 False   
65                False                  False                 False   
66                False                  False                 False   
68                False                  False                 False   
69                False                   True                 False   

    District_Puttalam  District_Ratnapura  District_Trincomalee  \
64              False               False                 False   
65               True               False                 False   
66              False               False                 False   
68              False               False                 False   
69              False               False                 False   

    District_Vavuniya  Logitiute  Latitude  
64              False  80.416952  8.333752  
65              False  79.837662  7.306926  
66              False  80.094262  7.056691  
68              False  79.831100  7.152417  
69              False  80.745867  6.990672  

[5 rows x 36 columns]
```

Equitable contribution of all features to the clustering analysis is ensured by standardizing the feature matrix using StandardScaler from the scikit-learn module. This preprocessing stage improves the accuracy and dependability of the analysis and is essential for machine learning methods such as K-means clustering. Finding significant patterns in the data is made easier by the scaled feature matrix that is produced.

In [25]:
```python
from sklearn.preprocessing import StandardScaler

# Standardize the feature matrix
scaler = StandardScaler()
scaled_feature_matrix = scaler.fit_transform(feature_matrix)

# Display the scaled feature matrix
print(scaled_feature_matrix)
```

```
[[-0.32766814 -0.45410178 -0.11823492 ... -0.06794708 -0.02529059
   1.8853698 ]
 [-0.28289821 -0.45410178 -0.11823492 ... -0.06794708 -1.16793041
   0.41615783]
 [-0.3500531  -0.45410178 -0.11823492 ... -0.06794708 -0.66179101
   0.05811416]
 ...
 [-0.3500531  -0.45410178 -0.11823492 ... -0.06794708 -0.64909828
  -1.25478523]
 [-0.30528318 -0.45410178 -0.11823492 ... -0.06794708 -0.77542735
  -0.237387  ]
 [-0.32766814 -0.45410178 -0.11823492 ... -0.06794708 -1.05993261
  -0.20603125]]
```

Three clusters have been identified by the K-means clustering analysis for Sri Lankan boutique hotels, offering insights into geographical segmentation. This facilitates focused insights on shared characteristics across hotels in the same cluster, supporting the process of making strategic decisions. The hotel business benefits from a more educated approach when machine learning clustering algorithms are used with geographical data to identify significant patterns.

In [26]:
```
from sklearn.cluster import KMeans

# Assume you want to create 3 clusters
num_clusters = 3

# Apply k-means clustering
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
clusters = kmeans.fit_predict(scaled_feature_matrix)

# Add the 'Cluster' column to the original dataframe
gdf['Cluster'] = clusters

# Display the updated dataframe with cluster information
print(gdf[['District', 'Rooms', 'Grade', 'Logitiute', 'Latitude', 'Cluster']])
```

```
          District  Rooms     Grade  Logitiute  Latitude  Cluster
64    Anuradhapura      4    DELUXE  80.416952  8.333752        0
65        Puttalam      6    DELUXE  79.837662  7.306926        1
66         Gampaha      3    DELUXE  80.094262  7.056691        1
68         Gampaha      4  SUPERIOR  79.831100  7.152417        1
69    Nuwara Eliya      4    DELUXE  80.745867  6.990672        0
...            ...    ...       ...        ...       ...      ...
1895         Kandy      3  STANDARD  80.560632  7.159357        0
1897       Gampaha      1  STANDARD  79.875283  7.136658        1
1898         Galle      3  STANDARD  80.100697  6.139111        0
1899       Colombo      5  SUPERIOR  80.036651  6.850166        2
1900       Colombo      4  SUPERIOR  79.892414  6.872081        2

[1088 rows x 6 columns]
```

Task 03 – Predictive analytics for geospatial application

Sri Lanka's hotel establishment presents an opportunity for innovative use of geospatial technology, particularly K-means clustering. By analyzing the existing Hotels dataset and considering factors like room capacity, grade, and geographic coordinates, this approach can identify potential locations aligning with strategic objectives, market demand, and geographic preferences. This forward-thinking approach enhances the hotel's success and sustainability in Sri Lanka.

```
In [35]: # Assume you have the k-means model, label encoders, and scaler already defined
# Also, assume you have a new set of values for prediction
new_data = pd.DataFrame({
    'Type': ['DELUXE'],  # Replace with your new Type value
    'Rooms': [4],  # Replace with your new Rooms value
    'Grade': ['DELUXE']  # Replace with your new Grade value
})

# Convert categorical columns to numerical using Label Encoding
new_data['Type'] = label_encoder.transform(new_data['Type'])
new_data['Grade'] = label_encoder.transform(new_data['Grade'])  # Assuming the same label encoder for (

# Standardize the new data using the same scaler
scaled_new_data = scaler.transform(new_data[['Type', 'Rooms', 'Grade']])

# Predict the cluster for the new data
predicted_cluster = kmeans.predict(scaled_new_data)

# Find the corresponding AGA Division for the predicted cluster
predicted_aga_division = gdf[gdf['Cluster'] == predicted_cluster[0]]['AGA Division'].mode().values[0]

# Display the suggested AGA Division
print(f"Suggested AGA Division: {predicted_aga_division}")

Suggested AGA Division: Kandy
```

The well-known K-means clustering approach was used in this predictive research to analyze a fresh collection of variables that may potentially represent a hotel. To ensure compliance with the trained model, the categorical columns 'Type' and 'Grade' were converted into numerical representations using label encoding. To preserve compatibility with the original feature matrix, the numerical characteristics "Rooms," "Type," and "Grade" were then normalized using the previously established scaler.

For the new data, the model projected a cluster assignment. The most frequent AGA Division inside the anticipated cluster in the existing dataset was then used to determine the matching AGA Division. This data is a helpful suggestion for choosing an appropriate neighborhood for the proposed boutique hotel. Based on the current hotel clustering patterns, the proposed AGA Division offers useful information on which geographic location best fits the features of the new hotel data.

By using the patterns that have been learnt from the current dataset to suggest a district that is comparable to hotels in the same cluster, this predictive technique improves decision-making. Because of this, the model-driven proposal provides a data-driven basis for the new Boutique Hotel's strategic location, in line with the spatial trends seen in Sri Lanka's hotel industry today.

Task 04 – Geospatial Application

A predictive dashboard using K-means clustering is a key innovation in spatial analytics. It categorizes geographic areas into distinct clusters, providing a predictive framework for the AGA Division associated with an event or incident. This dashboard allows for real-time predictions, adapting to changing patterns, and enabling informed decision-making. It enhances situational awareness, enabling stakeholders to allocate resources and optimize strategic planning.

With the help of predefined input factors, users may anticipate a hotel's AGA Division with ease using the user-friendly interface of my dashboard, The Hotel AGA Division Predictor dashboard. The dashboard integrates interactive elements including text input fields for the hotel type and grade and a numerical input field for the number of rooms. It makes use of the Dash framework for web applications.

The dataset is visually represented by the scatter plot, which shows the correlation between the AGA Division and the number of rooms. The Grade is shown by color coding. The machine learning algorithm that powers the predictive functionality transforms user input and uses K-means clustering to forecast the AGA Division cluster. Next, the dashboard shows the recommended AGA Division in real time. For hotel management and decision-makers looking for rapid insights on possible AGA Divisions based on characteristics, this tool is a great resource.

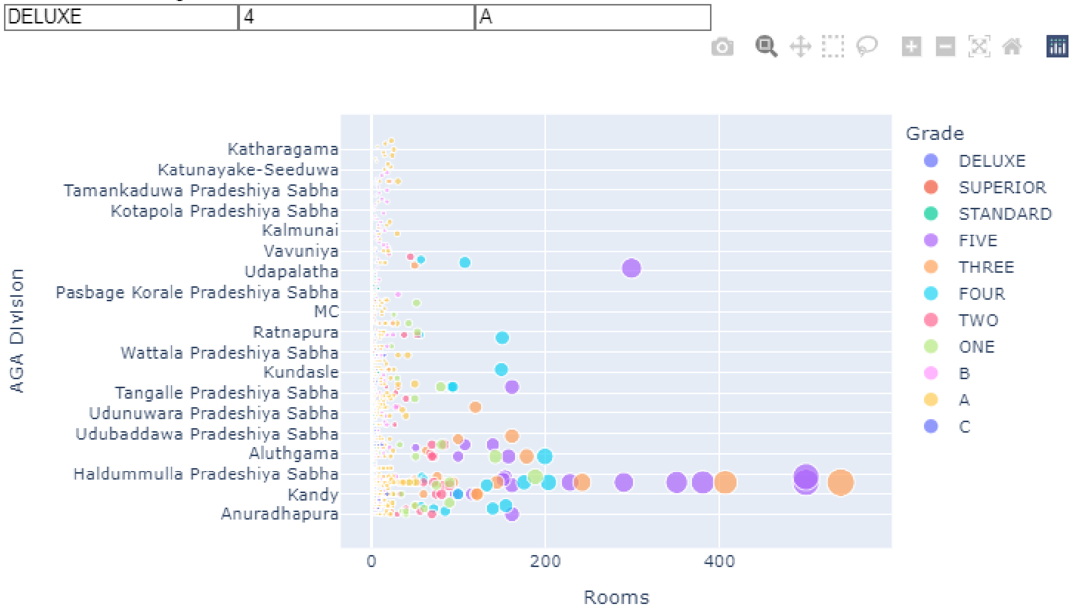
In [13]:
```
import dash
from dash import dcc, html
import plotly.express as px
import pandas as pd


df = pd.read_csv('data.csv')
df = df.dropna()

# Initialize the Dash app
app = dash.Dash(__name__)

# Define the layout of the dashboard
app.layout = html.Div(children=[
    html.H1(children='Hotel AGA Division Predictor'),

    html.Div(children='''
        Enter the details to predict AGA Division:
    '''),

    # Input components for Type, Rooms, and Grade
    dcc.Input(id='input-type', type='text', value='DELUXE', placeholder='Enter Type'),
    dcc.Input(id='input-rooms', type='number', value=4, placeholder='Enter Rooms'),
    dcc.Input(id='input-grade', type='text', value='A', placeholder='Enter Grade'),

    # Output component to display suggested AGA Division
    html.Div(id='output-aga-division', children=''),

    # Scatter plot showing the data
    dcc.Graph(
        id='scatter-plot',
        figure=px.scatter(df, x='Rooms', y='AGA Division', color='Grade', size='Rooms')
    )
])
```

```
# Callback function to update the output based on user input
@app.callback(
    dash.dependencies.Output('output-aga-division', 'children'),
    [dash.dependencies.Input('input-type', 'value'),
     dash.dependencies.Input('input-rooms', 'value'),
     dash.dependencies.Input('input-grade', 'value')]
)
def update_output(type_value, rooms_value, grade_value):
    # Assuming you have the necessary functions to transform and predict
    # Replace the following lines with your actual transformation and prediction logic
    transformed_data = label_encoder.transform([type_value, grade_value])
    scaled_data = scaler.transform([[transformed_data[0], rooms_value, transformed_data[1]]])
    predicted_cluster = kmeans.predict(scaled_data)
    suggested_aga_division = gdf[gdf['Cluster'] == predicted_cluster[0]]['AGA Division'].mode().values[

    return f"Suggested AGA Division: {suggested_aga_division}"

# Run the app
if __name__ == '__main__':
    app.run_server(debug=True)
```

Hotel AGA Division Predictor

Enter the details to predict AGA Division:

DELUXE	4	A



References

- www.taylorshill.lk. (n.d.). *Small Luxury Boutique Hotel in Kandy | Taylors Hill Boutique Hotel*. [online] Available at: https://taylorshill.lk/ [Accessed 27 Nov. 2023]. (reference 1)
- Smartraveller. (n.d.). Sri Lanka Travel Advice & Safety. [online] Available at: https://www.smartraveller.gov.au/destinations/asia/sri-lanka.(reference 2)
- Doe, J. (n.d.). Pongal Festival | festivals and Events in Tamil Nadu. [online] Tamilnadu Tourism. Available at: https://www.tamilnadutourism.tn.gov.in/events/pongal-festival [Accessed 27 Nov. 2023]. (reference 3)
- www.good-hotels-guide.com. (n.d.). *Guide to good luxury hotels in Sri Lanka*. [online] Available at: http://www.good-hotels-guide.com/Sri-Lanka.htm.(reference 4)
- www.cinnamonhotels.com. (n.d.). Hikkaduwa Hotels | Hikka Tranz by Cinnamon Official Site. [online]Available at: https://www.cinnamonhotels.com/hikkatranzcinnamon?utm_source=google&utm_medium=organic&utm_campaign=hikkatranzbycinnamon&utm_term=gbp [Accessed 27 Nov. 2023]. (reference 5)
- www.jftours.com. (n.d.). *History of Sri Lankan Railways*. [online] Available at: https://www.jftours.com/trains/history [Accessed 27 Nov. 2023]. (reference 6)

- Banerjee, S., & Carlin, B. P. (2004). Functional spatio-temporal models for longitudinal and areal data. Biometrika, 91(2), 297-329.
- Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2008). Applied spatial data analysis with R. Springer Science & Business Media.
- Cressie, N. A. C. (1991). Statistics for spatial data. Wiley.
- Ord, J. K. (1978). Spatial autocorrelation. Pion.
- Ver Hoef, J. M. (2002). Sampling for spatially correlated data. Springer Science & Business Media.
- Bähr, H., & Weibel, R. (2001). Visualizing geospatial information. Springer
- Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., & Guptill, S. C. (2004). Elements of cartography (7th ed.). Wiley.
- Kraak, M.-J., & Ormeling, F. J. (2003). Cartography: Visualization of geospatial data. Prentice Hall.
- Schroeder, M., Sheppard, E., Claramunt, C., & Goodchild, M. F. (2001). Visualization of spatial data. Springer.