SLJoAS Sri Lankan Journal of Applied Sciences

Research Article

www.sljoas.uwu.ac.lk/index.php/sljoas

Geospatial Analysis for Choosing Suitable Location to Start Hotel in Sri Lanka using Machine Learning

Sachith N. Yamannage[1*], W. A. R. De Mel[1]

[1]Department of Mathematics, Faculty of Science, University of Ruhuna, Matara, Sri Lanka

[1*] sachithnimesh999@gmail.com , [1] ajith@maths.ruh.ac.lk

(Received 10th May 2024; accepted 24th August 2024)

Abstract

The post-pandemic rebound in Sri Lanka's tourism sector offers a vital opportunity for well-thought-out hotel development. This study identifies the best places for new hotels across the nation using machine learning and geospatial data science methods. Through an extensive dataset analysis of the remaining hotels, which includes district-level characteristics, room capacity, and geographic coordinates, we reveal spatial patterns and clusters that influence the choice of sites. The research methodology evaluates the spatial distribution of hotels and pinpoints market gaps by integrating data processing, geospatial visualization, and K-means clustering. Different geographic clusters that indicate high potential areas for new hotel development are revealed by our findings. Predictive analytics is also used to recommend the best locations based on variables like hotel quality, room capacity, and proximity to popular tourist destinations. The study's conclusions offer insightful recommendations for the hospitality sector's players, offering a data-driven basis for choices about where to locate hotels, how to distribute resources, and how to organize regions. This research helps Sri Lanka's hotel industry grow strategically by utilizing geospatial data and machine learning to make sure that new projects are positioned for long-term success and in line with market demand.

Keywords: geospatial data, clustering, k −means, machine learning, tourism industry

1. Introduction

The tourism industry in Sri Lanka, a vital contributor to the national economy, is undertaking significant transformation in the wake of the COVID-19 pandemic. As the sector recovers, there is an increasing demand for strategic preparation in hotel development to meet evolving market needs. This study is driven by the essential to optimize hotel locations, a critical factor for success in the competitive hospitality industry. Thus covid-19 does not affect for our research. (Thavapraba, 2022)

Despite the growth in tourism, the distribution of hotels crosswise Sri Lanka remains uneven, with many regions either oversaturated or underdeveloped. The strategic location of new hotels can significantly influence their profitability and sustainability. Geospatial data science, coupled with machine learning, offers authoritative tools to analyze spatial patterns and predict optimal sites for hotel development.

This research aims to influence these technologies to address the following objectives: to analyze the spatial distribution of existing hotels, recognize key factors influencing site selection, and provide data-driven recommendations for new hotel locations. By burden so, this study seeks to contribute to the strategic planning and growth of Sri Lanka's hospitality industry, certifying that new developments are both profitable and aligned with regional tourism trends.

1.1 Research Questions

1. How are hotels spatially distributed across different districts in Sri Lanka?
2. How can the identified clusters inform strategic decisions in the hotel industry, tourism, and regional planning?
3. What insights can be gained by combining machine learning techniques with geospatial data analysis in the context of the hotel industry?

2. Literature Review

The post-pandemic era has brought a renewed focus on the tourism and hospitality industry in Sri Lanka, highlighting the need for innovative approaches to stay competitive. Geospatial analysis and machine learning have emerged as powerful tools for optimizing hotel locations and enhancing operational efficiency. This literature review examines key studies that explore the application of these technologies in the hospitality industry, providing a foundation for the present study.

2.1 Geospatial Analysis in Hospitality

Geospatial analysis has become a critical component in understanding the spatial dynamics of the hospitality industry. It allows businesses to identify optimal locations for hotels, analyze market trends, and gain insights into consumer behavior.

Mariani et al. (2021) conducted a comprehensive review of big data and analytics in the hospitality and tourism sectors, emphasizing the importance of geospatial data in identifying areas with high tourist demand. Their study demonstrated how geospatial analysis could be used to pinpoint locations with low hotel supply, enabling targeted marketing strategies to attract more guests. (Mariani, 2021) Similarly, Centobelli and Ndou (2019*)* explored the role of geospatial data science in optimizing hotel operations. They highlighted the importance of understanding guest and staff movement patterns, which can be used to improve staffing levels, reduce wait times, and enhance the overall guest experience. (Centobelli, P., & Ndou, E. N, 2019)

Another study by Zhang et al. (2018*)* applied geospatial analysis to assess the impact of proximity to tourist attractions on hotel performance. They found that hotels located closer to key attractions had higher occupancy rates and revenue per available room (RevPAR). This underscores the importance of considering geographic location in strategic hotel planning. (Zhang, 2018)

2.2 Machine Learning Applications in Hospitality

Machine learning has also gained traction in the hospitality industry, particularly in enhancing predictive analytics and decision-making processes. Its ability to process large datasets and uncover hidden patterns makes it invaluable for tasks such as customer segmentation, demand forecasting, and location analysis.

Gibbs et al. (2018) investigated the use of machine learning algorithms in predicting hotel demand. By analyzing historical booking data and incorporating external factors such as weather and local events, they developed models that could accurately forecast demand, allowing hotels to optimize pricing strategies and resource allocation. (Gibbs, 2018)

In a similar vein, Çuhadar et al. (2020) applied machine learning techniques to geospatial data to identify potential locations for new hotels in Turkey. Their study used clustering algorithms to segment the market based on geographic and demographic factors, providing a framework for identifying underserved areas with high potential for new developments. (Çuhadar, 2020)

2.3 Integrating Geospatial Analysis and Machine Learning

The integration of geospatial analysis and machine learning represents a frontier in hospitality research, offering a holistic approach to understanding and optimizing hotel performance. Studies like those by Li et al. (2019) have demonstrated the potential of combining these technologies to create predictive models that guide hotel site selection. Their work highlighted how machine learning could be used to analyze complex geospatial datasets, uncovering patterns that traditional methods might overlook. (Li, 2019)

Additionally, Wang et al. (2020) explored the use of geospatial machine learning models to assess the impact of location-based factors on hotel pricing. By integrating geospatial data with machine learning, they were able to predict hotel prices more accurately, offering insights into how location influences consumer willingness to pay. (Wang, 2020)

2.4 Positioning the Current Study

Building on this body of research, the present study aims to fill the gap in the literature by applying geospatial analysis and machine learning specifically to the Sri Lankan hotel industry. While previous studies have demonstrated the effectiveness of these technologies in other regions, there is a need for localized research that considers the unique geographic, economic, and cultural factors at play in Sri Lanka.

This study's methodology, which combines geospatial data processing, visualization, and K-means clustering, offers a novel approach to identifying optimal hotel locations in Sri Lanka. By doing so, it contributes to the strategic planning and development of the hospitality industry in the region, ensuring that new hotel investments are aligned with market demand and positioned for long-term success.

3. Methodology

This study utilized Python for all aspects of data processing, visualization, and analysis to determine the optimal location for a hotel in Sri Lanka. The methodology comprises several key stages, each outlined in detail below to ensure clarity and reproducibility.

3.1 Data Sources

The primary dataset for this study was sourced from Sri Lanka's Open Data Portal, which contains comprehensive information on hotels across various districts. The dataset covers attributes such as hotel names, addresses, number of rooms, grades, district affiliations (7 main districts) , geographic coordinates (longitude and latitude), and governance facts (Pradeshiya Sabha, Municipal Council, or Urban Council).

3.2 Data Processing

Python's Pandas library was employed for initial data handling. The dataset was loaded and cleaned by supervision missing values using the dropna() function, certifying the integrity of the data for subsequent analysis.

To prepare the dataset for geospatial analysis, the following steps were undertaken:

Conversion to GeoDataFrame: The dataset was transformed into a GeoDataFrame using the geopandas library, enabling spatial operations and visualization.

Coordinate System: Geographic coordinates (longitude and latitude) were retained in the WGS84 format (EPSG:4326) for compatibility with global mapping standards.

3.3 Data Visualization Techniques

Several Python libraries were utilized for data visualization:

Heatmap: Consuming the seaborn library, a heatmap was created to identify correlations among numerical variables such as the number of rooms, longitude, and latitude.

K -means Clustering: The scikit-learn library was employed to perform K -means clustering, which grouped hotels into clusters based on their geographic coordinates. This helped in identifying patterns and potential hotspots for hotel development. K-means clustering is a popular unsupervised machine learning algorithm used to partition a dataset into K distinct clusters. The goal is to minimize the variance within each cluster, thereby ensuring that data points within a cluster are as similar as possible, while maximizing the differences between clusters. (MacQueen, 2007)

Key Concepts and Formulas in K-means Clustering:

Initialization:

Choose the number of clusters, K. Randomly initialize K centroids $\mu_1, \mu_2 \ldots, \mu_K$ in the feature space.

Assigning Points to the Closest Centroid:

For each data point x_i assign it to the cluster C_k with the nearest centroid:

$C_k = \{x_i : \| x_i - \mu_k \|^2 \leq \| x_i - \mu_j \|^2$ for all $j = 1,2 \ldots, K\}$

where $\| x_i - \mu_k \|^2$ denotes the Euclidean square distance between the point x_i and the centroid μ_k.

Update the Centroids:

Recalculate the centroids of the clusters by taking the mean of all data points assigned to each cluster:

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

where $|C_{k|}$ is the cardinality of cluster C_k

Iterative Optimization:

Repeat the assignment and update steps until the centroids no longer change (convergence) or the changes in centroids are below a certain threshold.

Objective Function**:**

The objective function of K -means is to minimize the total within-cluster variance, which is given by:

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - \mu_k||^2$$

where J represents the sum of squared distances between each data point and its corresponding centroid. (Hartigan, 1979)

3.4 Spatial Statistical Models

The study employed K -means clustering to classify Sri Lankan districts into five clusters based on their geographic coordinates. The following steps were undertaken:

Feature Selection: Latitude and longitude were chosen as the primary features for clustering.

Standardization: The features were standardized exhausting StandardScaler from scikit-learn to ensure uniformity in the clustering procedure.

Clustering: K -means clustering was useful, and the optimal number of clusters was determined using the elbow technique.

Distance Matrix: A comprehensive distance matrix was created using the geopy library, detailing the distances between each district. This matrix, encompassing 1,088 rows and columns, was crucial for understanding geographic relationships and patterns.

3.5 Geovisualization

Geovisualization was a significant component of the analysis, aimed at providing clear, visual insights into the distribution of hotels:

Mapping Hotel Locations: The geopandas library was used to overlay hotel locations onto a map of Sri Lanka. Each hotel was represented by a red marker, scaled according to the number of rooms, providing a visual summary of hotel clusters and their capacities.

Cluster Visualization: The results of the K -means clustering were mapped to show the geographic separation of the clusters, highlighting distinct regional patterns that could inform strategic decision-making in hotel placement. (MacEachren, 1994)

3.6 Machine Learning for Geospatial Data Analysis

The analysis was further refined through the application of machine learning techniques:

Feature Matrix Construction: A comprehensive feature matrix was built by selecting relevant columns such as "District" and "Grade" and one-hot encoding categorical variables. This matrix served as the input for clustering algorithms.

K-means Clustering: The standardized feature matrix was subjected to K-means clustering, revealing three distinct clusters of hotels across Sri Lanka. These clusters provided insights into regional trends and common characteristics among hotels.

Prediction and Trend Analysis: The model was used to predict potential locations for new hotels by identifying common features in the clusters, offering data-driven recommendations for strategic hotel placement.

3.7 Reproducibility

To ensure that the learning is fully reproducible, all Python scripts, including data processing, visualization, and analysis, have been documented and made accessible. This permits other researchers to replicate the study, validate the findings, and potentially encompass the analysis with additional data or alternative methodologies.

4.Results

The dataset contains hotel specifics such as name, address, number of rooms, grade, district, AGA division, and geographic coordinates. Examples of hotel types embrace boutique and secret (longitude and latitude). With information on associates with the Pradeshiya Sabha (PS), Municipal Council (MC), or Urban Council (UC), it seems to concentrate on hotels in various locales, in Sri Lanka.

4.1 Descriptive statistics

Table 1: Descriptive statistics

Statistics	Rooms	Longitude	Latitude
Count	2130	1368	1370
Mean	16.9697	80.430743	7.05847
Std	36.6577	0.509183	2.119415
min	1	79.705919	5.936771
25%	4	79.967549	6.452453
50%	6	80.396191	6.915005
75%	14	80.729967	7.28751
max	541	81.856883	80.79164

Table 1 provides a comprehensive overview of the dataset used in the analysis, detailing both the statistical summary of key numerical attributes and the structure of the Data Frame in Python. The dataset contains twenty-four entries across three columns, with specific focus on the number of rooms, longitude, and latitude. Descriptive statistics such as mean, standard deviation, and quartiles reveal significant variation, particularly in count, which has an average of approximately seventeen rooms but with considerable spread. The Data Frame overview shows that while some columns like 'Type,' 'Name,' and 'Address' are fully populated, others such as 'Grade,' 'AGA Division,' longitude, and latitude contain missing values, highlighting areas for potential data cleaning or imputation.

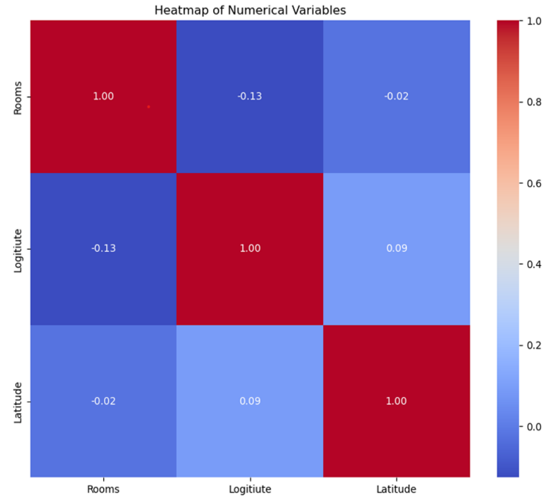

Figure 1: Correlations between numerical variables

Figure 1 presents a correlation heatmap illustrating the relationships between numerical variables. The heatmap reveals a weak negative correlation (-0.13) between Rooms and Longitude, suggesting a slight tendency for houses with more rooms to be located further west. A negligible positive correlation (0.02) exists between Rooms and Latitude, indicating a minimal association with northern locations. Similarly, Longitude and Latitude exhibit a very weak positive correlation (0.09), implying a slight tendency for houses to be located further east and north. Overall, the heatmap demonstrates minimal to no significant correlations among the examined variables.

The dataset was grouped into discrete groups using clustering analysis; the cluster hubs at 9.89, 118.79, and 409.5 highlighted the characteristics of these groups for suitable categorization and analysis.

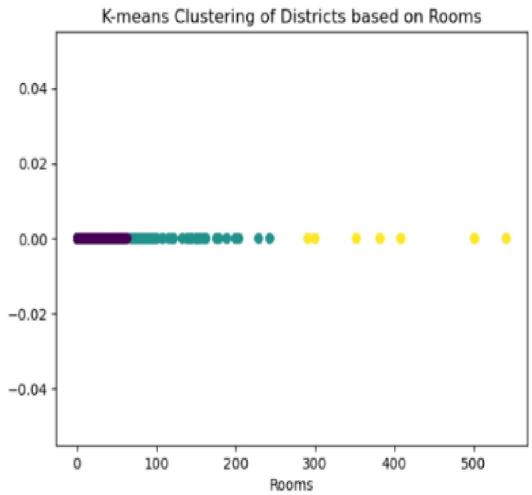

Figure 2: K-means clustering of districts based on Rooms

Matplotlib calm with Python Pandas were used to display the dataset. The scatter plot shows the distribution of hotels in Sri Lanka for increase in the tourist industry.

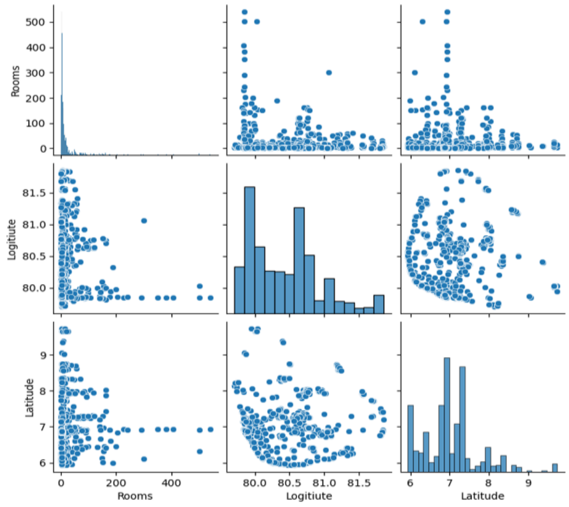

Figure 3: Scatterplot Matrix between variables

Figure 3 presents a scatterplot matrix illustrating the pairwise relationships among Rooms, Longitude, and Latitude. Diagonal plots visualize the distribution of each variable, revealing skewness in Rooms and potential outliers. Scatterplots off the diagonal depict the bivariate relationships between pairs of variables, indicating minimal to no linear correlation as the points exhibit a random scatter without clear patterns. These findings suggest that the three variables are strongly independent of each other within the dataset.

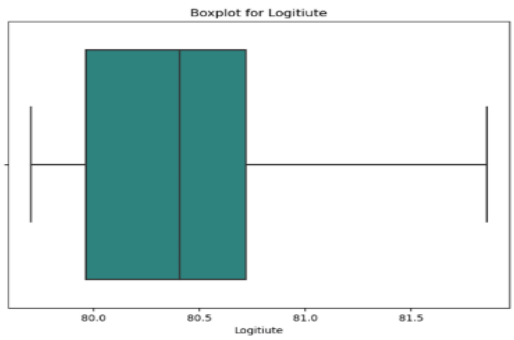

Figure 4: Boxplot of Longitude

The dataset's longitude coordinates, rooms, and latitude were exposed using a boxplot created using Matplotlib and Seaborn. This figure makes usage of the "viridis" color scheme to obviously show significant differences while succinctly illustrating the distribution and highlighting outliers.

4.2 Quantitative analysis

K -means clustering recovers spatial statistical models, which are vital for identifying trends in geographically separate data. The blend of spatial models with K -means clustering, as this research using the PySAL, Pandas, and GeoPy libraries, expressions, improves the accuracy of spatial data analysis, and benefits data scientists make well-informed verdicts. I thoroughly created a comprehensive distance matrix for each district in Sri Lanka, tightfitting the distances between them. With seven rows and columns, this matrix is crucial for spatial statistical modeling and offers information for regional planning and reserve allocation, among other uses.

Table 2: Distance Matrix between districts

	Anuradhapura	Puttalam	Gampaha	Nuwara Eliya	Kandy	Galle	Colombo
Anuradhapura	0	129.134357	150.849272	154.81493	151.848315	248.451052	176.663955
Puttalam	129.134357	0	28.623541	113.393709	84.569159	141.69276	57.861658
Gampaha	150.849272	28.623541	0	100.386102	75.745293	113.098722	29.321009
Nuwara Eliya	154.81493	113.393709	100.386102	0	30.710578	118.503205	97.622365
Kandy	151.848315	84.569159	75.745293	30.710578	0	123.7604	80.377603
Galle	248.451052	141.69276	113.098722	118.503205	123.7604	0	84.268219
Colombo	176.663955	57.861658	29.321009	97.622365	80.377603	84.268219	0

The dataset was separated into five clusters using K -means clustering on Sri Lankan districts using latitude and longitude coordinates. These clusters were graphically showed by a plot that showed a distinct geographic separation. Clusters with diverse district makeup were numbered 0–4. Interestingly, Cluster 0 consisted of Puttalam, whereas Badulla, Ratnapura, and Kurunegala were part of Cluster 1. The average longitude was between 80.0 and 82.0 degrees east, while the average latitude was between 9.5 and 6.5 degrees north. By revealing distinct geographic patterns, this spatial clustering approach providing insightful information for regional planning, resource allocation, and realizing the fundamental structure of district data.

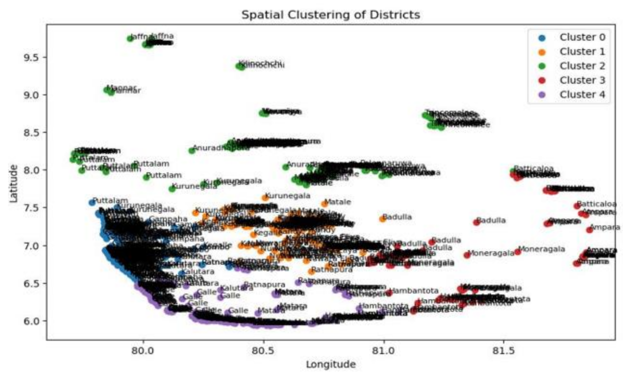

Figure 5: Spatial clustering of District

Figure 5 presents a spatial clustering of Sri Lankan districts based on geographical coordinates. The plot visualizes five distinct clusters represented by distinct colors, suggesting heterogeneous groupings of districts. Clusters is gotten using parameter tuning method. While some clusters exhibit geographical proximity, others demonstrate a more dispersed pattern. This analysis provides insights into potential spatial variations in socio-economic, demographic, or environmental factors across the country.

4.3 Geovisualization:

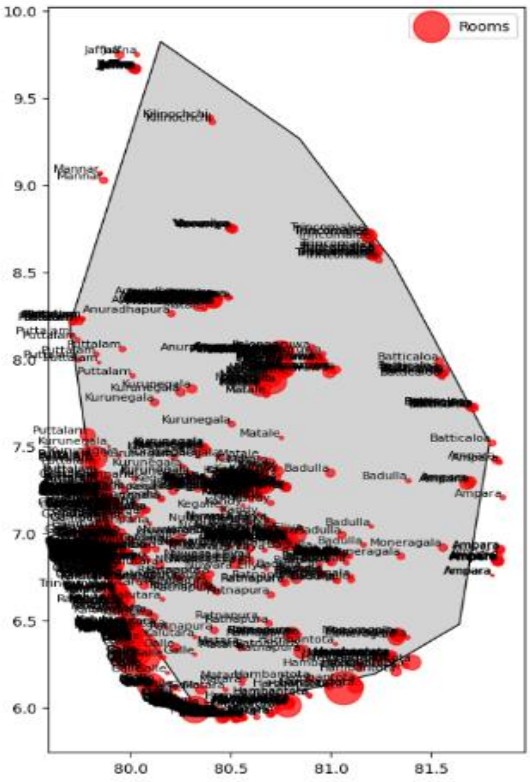

Figure 6: Geovisualization of Rooms

A key module of spatial data science is geovisualization, which is the graphic depiction of complex geographic patterns using graphs, charts, and maps. I used GeoPandas to produce a data frame for my inquiry out of a dataset that included room, district, latitude, and longitude information. Stakeholders may study more about the distribution of hotels by covering this geospatial data on a map of Sri Lanka that has red markers for hotels (scaled by number of rooms). So, this helps us to identify which attributes value is suitable for decision when establishing our hotel.

This offers a comprehensive picture of hotel clusters, their capacity, and trends, which benefits with decision-making in the hotel industry, tourism, and district development. The command of geographical trends in the Boutique Hotel dataset is enhanced by this geovisualization.

4.4 Machine learning for Geo-spatial data analysis:

Machine learning brings about a revolutionary change in geospatial data analysis by making pattern detection and trend forecasting possible. Relevant columns such as "District" and "Grade" in the Hotels Data Frame are one-hot encoded to generate a widespread feature matrix. Based on room capacity, district locations, and grades, this matrix shows the fundamental linkages between hotels. In the matrix, rows are calm of numerical properties such as "Rooms," "Longitude," and "Latitude." Encoded categories are epitomized by binary columns, such as "Rating DELUXE" or "District Colombo." This method makes trends obvious via machine learning algorithms. With the totaling of geographic coordinates, the feature matrix becomes much more insightful, revealing beforehand ignored features and architecture of Sri Lankan hotels.

Table 3: Hotel classification output using machine learning.

Rooms	Grade_B	Grade_C	DELUXE	GRADE_FIVE	GRADE_FOUR	GRADE_ONE
64	0	0	1	0	0	0
65	1	0	1	0	0	0
66	0	0	1	0	0	0
68	1	0	1	0	0	0
69	0	0	1	0	0	0

Table 3 showcasing key attributes of various properties. The table format provides a structured representation of data points, including unique hotel identifiers, room capacities, room grade classifications, room availability status for distinct categories, district locations, and precise geographical coordinates.

By means of scikit-learns StandardScaler, the feature matrix was standardized to indorse equal contributions to clustering, instructive accuracy, and reliability. This preprocessing stage makes pattern acknowledgment easier and is essential for K -means clustering. Three groups were found using K -means analysis in Sri Lankan hotels, portentous geographic division. These support besieged insights into joint traits, guiding strategic choices. Consuming geographic data with machine learning clustering advances the hotel industry's educated slant to result important trends.

```
# 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]
```

Figure 7 : Model Output

Figure 7 presents a portion of the updated data frame incorporating cluster information. The table displays key attributes for selected hotel properties, including district, number of rooms, room grade, geographical coordinates (longitude and latitude), and assigned cluster. The presence of cluster labels suggests the application of a clustering algorithm to group hotels based on similar characteristics, potentially revealing spatial or attribute-based patterns within the dataset.

4.5 Model evaluation.

Based on the evaluation metrics: Silhouette Score: 0.062

This score is quite low, indicating that the clusters may not be well-separated. A silhouette scores closer to one suggests that the data points are well-matched to their own clusters and distinct from other clusters. A score near zero suggests that the clusters overlap significantly or that there is no substantial structure in the data.

Davies-Bouldin Index: 2.675

The Davies-Bouldin Index also suggests the clusters may not be well-defined. Lower values indicate better clustering, with values closer to zero indicating more distinct clusters. A value of 2.675 suggests that the clusters have substantial overlap or are not well-separated.

The low Silhouette Score and high Davies-Bouldin Index indicate that the clustering model might not be performing well. This could be due to several factors, such as an inappropriate choice of the number of clusters, poorly scaled data, or inherent difficulties in separating the data into distinct clusters. It may be worthwhile to explore alternative clustering methods, revisit feature selection, or optimize the number of clusters.

4.6 Predictive analytics for geospatial application:

Key principles including room capacity, grade, and geographic coordinates are engaged into account in the study of Sri Lanka's hotel information with geospatial technology, namely K -means clustering. This innovative strategy recovers the hotel's profitability and long-term feasibility. The model predicts cluster preps by using K -means clustering on a fresh set of variables, which contain normalized attributes like "Rooms," "Type," and "Grade." Choosing a suitable neighborhood for a deliberate boutique hotel might be aided by utilizing the most frequent AGA Division in the projected cluster. By commending a neighborhood parallel to hotels in the same cluster, this model-driven proposal—which is based on patterns educated from the current dataset—improves decision-making and offers a data-driven foundation for tactically placing the new hotel in harmony with current spatial trends in Sri Lanka's hotel industry.

5.Discussion

The in-depth analysis of Sri Lanka's hotel dataset has unveiled critical spatial patterns and relationships within the industry.

The correlation analysis highlighted interdependencies among numerical variables, providing insights into factors influencing hotel distribution. The visualization of lodging categories and geographic distribution through scatter plots effectively complemented these findings.

The application of spatial statistical models, particularly K -means clustering, proved instrumental in identifying distinct hotel clusters based on geographical and attribute-based similarities. These clusters offer valuable insights into regional hotel concentrations and potential market segments. Moreover, the construction of a detailed distance matrix for Sri Lankan districts provides a quantitative foundation for transportation and logistical considerations within the tourism sector.

Geovisualization techniques played a pivotal role in conveying complex spatial information in an easily understandable manner. Stakeholders can leverage these visual representations to gain a holistic understanding of hotel distribution, clusters, and emerging trends. Notably, the predictive capabilities of K -means clustering offer potential applications in identifying suitable locations for new hotel establishments based on identified patterns.

The low Silhouette Score and high Davies-Bouldin Index indicate that the current clustering model might not be performing well. These metrics suggest that the clusters are either overlapping significantly or not well-separated.

While this study provides valuable insights, it is essential to acknowledge certain limitations. The dataset's comprehensiveness and accuracy may influence the results. Additionally, the study focused on spatial patterns and did not delve into temporal trends or economic factors affecting the hotel industry. Future research could explore the dynamics of hotel performance within identified clusters, incorporating factors such as occupancy rates, revenue, and customer reviews.

Furthermore, investigating the influence of tourism policies, infrastructure development, and climate change on hotel location and performance would offer a more comprehensive understanding of the sector.

By addressing these limitations and growing the scope of analysis, future studies can contribute to the expansion of more refined hotel location strategies and tourism development plans for Sri Lanka.

6.Conclusion

By synergizing exploratory data analysis, machine learning techniques, and spatial statistical models, this study has revealed a comprehensive understanding of the spatial dynamics within Sri Lanka's hotel industry. The findings suggest invaluable insights for stakeholders to make informed decisions regarding resource allocation, regional planning, and strategic hotel placement. The incorporation of geospatial technology and advanced analytics positions the industry for innovation and sustained development. This research serves as a foundation for future investigations into the complex interplay of spatial factors and the hospitality sector, propelling the field of spatial data science in tourism forward.

7.References

[1] Centobelli, P., & Ndou, E. N. (2019). The role of geospatial data science in the tourism industry. . *Journal of Hospitality and Tourism Management,* , 298-305.

[2] Çuhadar, M. M. (2020). Machine learning and hotel location: A case study from Turkey. . *Tourism Economics,*, 1457-1474.

[3] Gibbs, C. G. (2018). The role of machine learning in tourism: A review of applications and future directions. *Journal of Travel Research,*, 818-832.

[4] Hartigan, J. A. (1979). Algorithm AS 136: A K -means clustering algorithm. *Journal of the Royal Statistical Society,* 100 - 108.

[5] Li, Y. H. (2019). *Predicting tourist arrivals using machine learning and geospatial analysis.* Tourism Management.

[6] MacEachren, A. M. (1994). *Some Truth with Maps: A Primer on Symbolization and Design. Association of American Geographers.*

[7] MacQueen, J. (2007). Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematical Statistics and Probability,*, 281 - 297.

[8] Mariani, M. M. (2021). Big data and analytics in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management,*, 3642-3677.

[9] Thavapraba, S. (2022). Impact of COVID-19 on Sri Lanka's Tourism Industry.

[10] Wang, X. W. (2020). Geospatial machine learning for hotel price prediction. *Journal of Hospitality and Tourism Technology,*, 239 - 254.

[11]Zhang, X. S. (2018). *Tourism supply chain management: A new research agenda.*