Geospatial Analysis for choosing suitable location to start hotel in Sri Lanka Using Machine Learning

Sachith Nimesh Yamannage\textsuperscript{1} and W A R De Mel\textsuperscript{2}

\textsuperscript{1}B.SC (hon) Financial Mathematics and Industrial Statistics, University of Ruhuna
\textsuperscript{2}Affiliation not available

February 27, 2024

Abstract

The post-pandemic recapture of Sri Lanka’s tourist industry is anticipated, and this paper suggests a hotel that employs geospatial data science for strategic placement and operational excellence. The hotel will use geospatial analytics to choose the best site, customize its hospitality offerings to meet local demands, assign resources as efficiently as possible, build strong community relationships, and navigate the competitive marketplace. The hotel will succeed in the resurgent Sri Lankan hospitality sector cheers to its data-driven strategy.
Geospatial Analysis for choosing suitable location to start hotel in Sri Lanka Using Machine Learning

Sachith Nimesh Yamannage  
https://orcid.org/0000-0002-9320-1890  
Email: sachithnimesh999@gmail.com

Dr. W. A. R. De Mel  
https://orcid.org/0000-0002-3958-060X  
Email: ajith@maths.ruh.ac.lk

Faculty of Science, University of Ruhuna, Matara, Sri Lanka

Abstract

The post-pandemic recapture of Sri Lanka’s tourist industry is anticipated, and this paper suggests a hotel that employs geospatial data science for strategic placement and operational excellence. The hotel will use geospatial analytics to choose the best site, customize its hospitality offerings to meet local demands, assign resources as efficiently as possible, build strong community relationships, and navigate the competitive marketplace. The hotel will succeed in the resurgent Sri Lankan hospitality sector cheers to its data-driven strategy.

Keywords: geospatial data science, hotel, Sri Lanka, Machine Learning

1. Introduction

As Sri Lanka gets ready for a post-pandemic shrinking, the country's booming hotel industry grants a huge opportunity for success and innovation. To exploit on this expanding potential, we propose to build a hotel that logically integrates the revolutionary power of geospatial data science. Owing to our pioneering approach, which will modify our market analysis, customer interface, and operational efficacy with geographic location, our hotel will be palatable in this crowded market.

Using a huge dataset, our company put on geospatial data science to determine the top spot for our hotel in Sri Lanka, making sure it meets market demand and customer prospects.

Through an alertness of the intricate relationships between local characteristics and customer likings, geospatial data science enables establishments to tailor services to match the wants of their target audience. Optimizing resource provision and improving visitor experiences are critical in supply chain and marketing. One Sri Lankan hotel is devoted to leveraging geospatial data science to select the ideal location, operate efficiently, and make a lasting brand in the rapidly expanding hospitality sector.

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

After the sweeping, Sri Lanka's hospitality sector is expected to grow significantly, and to stay viable, lodging establishments are increasingly utilizing geospatial data science. The best place for a hotel may be found using geospatial data science, which can also be used to achieve competition, improve resource apportionment, customize hospitality products to meet local requirements, and build strong communal links.

One study by Mariani et al. (2021) found that geospatial data science can be used to identify areas with high tourist petition and low hotel supply. This information can then be used to develop targeted marketing operations and attract more guests.\(^1\)

Another study by Centobelli and Ndou (2019) found that geospatial data science can be used to optimize hotel operations. For example, geospatial data can be used to realize the movement of guests and staff, which can then be used to rally staffing levels and reduce wait times.\(^2\)

Geospatial data science is a powerful tool that can be used to expand the performance of hotels in Sri Lanka. By leveraging geospatial data, hotels can gain treasured insights into their target market, optimize their operations, and build a sturdy brand.

3. Methodology:

The approach adopted in the study of the info on hotels in Sri Lanka incorporates several stages to discover, process, and interpret the data. The main techniques employed are charted in the following. The machine learning language secondhand in this work was Python.

Data Processing:

In order to preserve the integrity of the dataset, missing values were handled appropriately by means of the drop NA method. After that, a GeoDataFrame (gdf) was built to assure accuracy for statistical and geographic analysis.

Visualization Techniques:

A heatmap was used in the research to find trends in the latitude, longitude, and number of rooms. Hidden patterns were revealed by exploiting k-means clustering technique. The distribution of hotel kinds and locations in Sri Lanka was revealed graphically using Pie Charts and Scatter Plots.

Spatial Statistical Models:

Sri Lankan districts were divided into five clusters consistent with latitude and longitude using k-means clustering. A thorough distance matrix was erected to clarify the geographical relationships and patterns crosswise districts.

Geovisualization:

Utilizing GeoPandas, the dataset was altered into a GeoDataFrame, facilitating the visualization of hotel distribution on a Sri Lankan map. Map overlaps incorporated hotel locations onto a
global map, as long as additional context. Spatial analysis was hired to visualize trends, patterns, and potential gaps in hotel distribution clusters.

**Machine Learning for Geo-spatial Data Analysis:**

By choosing related columns, one-hot encoding categorical variables, and building an extensive representation, the feature matrix was ready. StandardScaler allowed standardization, which guaranteed the feature matrix's consistency for precise machine learning analysis. Next, K-means clustering was used to identify regional hotel segmentation trends.

4. Findings

The collection 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, most likely in Sri Lanka.

**Figure 1**

Heatmap shows clearly about correlations between numerical variables.

**Figure 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, greatly 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, tightening the distances between them. With 1088 rows and columns, this matrix is crucial for spatial statistical modeling and offers information for regional planning and reserve allocation, among other uses.
The dataset was separated into five clusters using K-means clustering on Sri Lankan districts using latitude and longitude coordinates. These clusters were graphically shown 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.

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 an 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.
4.5 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 props 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 systematic examination of Sri Lanka's hotel dataset has provided important new information about the geographical subtleties of the sector. There are evident relationships between numerical variables and different types within the dataset, as demonstrated by the heatmap and clustering analysis. The several percentages of lodging categories and the geographic distribution of hotels are clearly interconnected by the visualizations, which include pie charts and scatter plots. The spatial statistical models express the one with k-means clustering offer important insights into the patterns seen in data that is dispersed spatially. Creating a detailed distance matrix for each district in Sri Lanka enlarges the possible customs for regional planning and resource distribution.

A real technique that gives stakeholders an innate considerate of hotel distribution, clusters, and possible trends is geovisualization. A prophetic element is added by the machine learning method, notably K-means clustering, which acclaims potential sites for new hotel based on patterns discovered from the offered information.

6. Conclusion

Of conclusion, a inclusive knowledge of the spatial undercurrents in Sri Lanka’s hotel business is made possible by the grouping of exploratory data analysis, machine learning tactics, and spatial statistical models. The results afford industry stakeholders with practical insights to help them style well-informed conclusions on resource provision, regional planning, and calculated hotel placement, among other topics. Unconventional analytics combined with the compliance of geospatial technology equips the segment for innovation and long-term success. This research adds to our thoughtful of the hotel environment and arranges the groundwork for additional studies and claims in the rapidly developing subject of spatial data science in the hospitality business.

7. References


8. Appendix

- Metadata - Accommodation Information for Tourists | Open Data Portal - Sri Lanka

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boutique Hotels</td>
<td>THE THEVA RESIDENCY</td>
<td>11/BS/10-1 06TH LANE, KANDY</td>
</tr>
<tr>
<td>Boutique Hotels</td>
<td>HIGHLAND VILLA</td>
<td>31, ABIMANGAMA ROAD, KANDY</td>
</tr>
<tr>
<td>Boutique Hotels</td>
<td>ULAGALLA WALAWWA RESORT</td>
<td>THIRAPPANE, ANURADHA</td>
</tr>
<tr>
<td>Boutique Hotels</td>
<td>GALLE FORT HOTEL</td>
<td>NO.28, CHURCH STREET, KANDY</td>
</tr>
<tr>
<td>Boutique Hotels</td>
<td>THE ELEPHANT CORRIDOR</td>
<td>POTHANA, KIBISSA, SIGIRIYA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rooms</th>
<th>Grade</th>
<th>District</th>
<th>AGA Division</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Kandy</td>
<td>Kandy Divisional Secretariat</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Matara</td>
<td>Weligama Divisional Secretariat</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Anuradhapura</td>
<td>Anuradhapura East</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Galle</td>
<td>Galle Divisional Secretariat</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Matale</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>District</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kandy</td>
<td>80.63541</td>
<td>7.276036</td>
</tr>
<tr>
<td>Weligama Pradeshiya Sabha</td>
<td>80.40997</td>
<td>5.960334</td>
</tr>
<tr>
<td>Anuradhapura</td>
<td>80.54506</td>
<td>8.205927</td>
</tr>
<tr>
<td>Galle</td>
<td>80.21756</td>
<td>6.026649</td>
</tr>
<tr>
<td>Matale</td>
<td>80.71074</td>
<td>7.943525</td>
</tr>
</tbody>
</table>