Geolocational Data Analysis using Foursquare API

ANSHU KHURANA [1], SEEMA KALONIA[2], VIVEK CHAUHAN[3]

Department of Information and Technology, Maharaja Agrasen Institute of Technology, Delhi, India

Abstract: The location of a restaurant is a key determinant of its success. A strategic and optimal location requires a comprehensive analysis of multiple factors, including demographics, competition, and customer preferences. In this research paper, we present a methodology that combines geospatial analysis and the Foursquare API to identify the best location for a restaurant. By harnessing the power of Python and various popular packages such as numpy, pandas, matplotlib, seaborn, geopy, folium, and sci-kit-learn, we perform crucial tasks such as data preprocessing, visualization, clustering, and deep learning. Our approach aims to provide restaurant owners and stakeholders with valuable insights and actionable information to facilitate informed decision-making. By leveraging these advanced technologies and techniques, we believe our methodology can contribute to the effective andefficient selection of an optimal restaurant location, ultimately increasing the chances of success in the competitive food industry.

***Keywords*—**Data preprocessing, Visualization, Clustering, Deep Learning, Valuable insights.

1. INTRODUCTION

A restaurant's location is a crucial aspect that has a big impact on its performance. It is essential for drawing clients, increasing foot traffic, and ultimately boosting profits. For making wise restaurant placement decisions, geospatial analysis, a process that combines Geographic Information Systems (GIS) [1] and diverse data sources [2], offers essential insights. This article examines the significance of restaurant location and how geospatial analysis supports decision-making to maximize profitability. The success of a restaurant heavily depends on its location. Selecting the right location [3] can attract a steady flow of customers and contribute significantly to the restaurant's profitability. The impact of a restaurant's location on customer accessibility and visibility is one of the main reasons why it is important. A restaurant that is easily accessible and located in a busy neighborhood is more likely to draw customers. It guarantees that the business is easily situated for both locals and bystanders, boosting the likelihood of drawing a variety of customers. In order to choose the ideal location for a restaurant, it is essential to comprehend the demographics and preferences of the target market. Restaurant owners can evaluate variables like income levels, age demographics, and ethnic preferences in certain places using geospatial analysis. According to the preferences of the local populace, this information helps customize the food [4], atmosphere, and marketing techniques [5], ultimately increasing customer retention and profitability. Businesses can assess the intensity of competition in a certain area using geospatial analysis. Restaurant owners can decide whether a location is profitable by mapping out existing eateries and finding areas with high market saturation. A competitive advantage can be created and success prospects increased by selecting a location with less competition or by having a distinctive selling proposition. The potential revenue of a restaurant is significantly impacted by the foot traffic in a particular region. High foot traffic zones like shopping malls, business districts, and tourist attractions provide a built-in clientele. These high-traffic areas can be found via geospatial analysis, which enables restaurant owners to place their businesses there to maximize visibility to potential customers. The proximity of a restaurant's location to suppliers and distribution hubs is another logistical factor to take into account. Locating regions with simple access to fresh vegetables, meats, and other necessary items can help with geospatial analysis. This lowers the cost of transportation and guarantees a constant supply chain, which eventually increases cost-effectiveness and profitability. The viability of a restaurant's location can be considerably impacted by local rules and zoning laws. Geospatial analysis makes it easier to locate places that adhere to zoning laws [6], ensuring that the business works legally. This avoids potential legal issues and frees the company to concentrate on increasing earnings. However, choosing an optimal location for a restaurant is a complex task that involves considering multiple factors, such as demographics, competition, and customer preferences. To address this challenge,the use of geospatial analysis and advanced technologies has emerged as a valuable approach for identifying suitable locations. The major objective of this research paper is to present a method for choosing the ideal location for a restaurant using geospatial analysis [7] and the Foursquare API. By finishing this project, we aim to provide valuable insights to restaurant owners and stakeholders,enabling them to make data-driven decisions in selecting the most suitable location for their establishment.

2. LITERATURE REVIEW

These are some previous studies and insights that have provided a solid foundation for our research and understanding of geolocational data analysis and its application in the restaurant industry. Some of the author’s work we have reviewed in this literature review section.

2.1 Importance of Location in the Restaurant Industry

The significance of location in the restaurant industry cannot be overstated. The location of a restaurant directly impacts its visibility, accessibility, and customer base. Studies have shown that restaurants situated in high-traffic areas, close to residential or commercial centers, and with convenient parking options tend to attract more customers. Additionally, factors such as nearby amenities, demographic profiles, and competition within the vicinity [8] play a crucial role in determining a restaurant's success. Therefore, selecting the right location is vital for ensuring a competitive edge and maximizing profitability.

2.2 Geolocational Analysis for Location-Based Decisions

Geospatial analysis is a powerful technique used to analyze and interpret spatial information. It entails combining statistical techniques with geographic information systems (GIS) to gain insights into spatial patterns and relationships [9]. In the context of finding optimal locations for restaurants, geospatial analysis enables the visualization of data on maps, identification of target areas based on proximity to amenities and demographics, and spatial clustering to identify potential customer markets. By employing geospatial analysis techniques, decision-makers can make informed choices regarding restaurant location, considering factors such as accessibility, visibility, and customer preferences.

2.3 Previous Studies on Location Analysis for Restaurants

Several studies have explored the topic of location analysis for restaurants, offering valuable insights and methodologies. For instance, a study [10] employed geospatial analysis to identify potential locations for a restaurant chain based on demographic characteristics and proximity to competitor establishments. The research utilized GIS tools and statistical analysis to evaluate various factors influencing restaurant success. Another study [11] focused on analyzing customer preferences and traffic patterns to identify optimal locations for new restaurants in urban areas. The research employed machine learning algorithms and spatial clustering techniques to identify target areas with high potential for customer footfall. These previous studies highlight the relevance and effectiveness of geospatial analysis and data-driven approaches in location analysis forrestaurants. By building upon these studies, this research aims to contribute to the existing knowledge by utilizing geolocational data analysis techniques and leveraging the Foursquare API to provide comprehensive insights into finding the optimal location for a restaurant.

3. METHODOLOGY

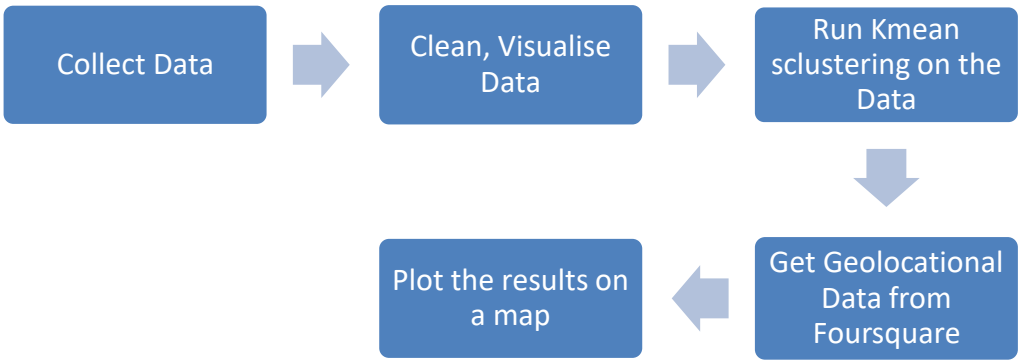

Fig 3.1 Methodology Flow

The first step shown in Fig. 3.1 to collect relevant data for the location analysis of restaurants. This includes obtaining geospatial data such as maps, demographic information, and competitor locations. Additionally, we will utilize the Foursquare API to retrieve venue data, including restaurant ratings, reviews, and popular venues in the target area. By combining these different data sources, we can obtain a comprehensive dataset for analysis. Keep your text and graphic files separate until after the text has been formatted and styled.

The data must be preprocessed after collection to guarantee its integrity and suitability for the analysis methods. Cleaning the data, addressing missing values, standardizing formats, and manipulating variables [12] are some examples of the tasks involved in data preprocessing. if necessary. We will utilize Python packages like numpy and pandas to perform these preprocessing steps effectively.

Table 1. Data Table Example (Already Existing Location in town)

S.no	Location	Latitude	Longitude
1.	Indiranagar	12.91975	77.624685
2.	Koramangala	12.937187	77.641123
3.	DomlurMahadevpura	12.954692	77.700229
4.	Marathahalli	12.994100	77.700842
5.	Sarjapur	12.950098	77.667638
6.	KR Puram	12.919615	77.700213
7.	Nallurhalli	12.959379	77.700990
8.	Sanjay Nagar	13.042598	77.731990
9.	Kalyan Nagar	13.008335	77.640747

Exploratory Data Analysis (EDA)[13] is conducted to acquire a deeper understanding of the dataset and recognize potential trends or

patterns within it relationships that may influence the restaurant location. This step involves visualizing the data using matplotlib and seaborn packages toexamine the distribution of variables, identify outliers, and explore correlations between different factors. EDA helps informulating hypotheses and guiding subsequent analysis steps.

Geospatial analysis plays a crucial role in location analysis for restaurants. We will utilize Python packages likegeopy and folium to perform geocoding, mapping, and spatial analysis. Geopy [14] provides geocoding functionality to convert addresses into geographic coordinates, enabling accurate mapping and visualization of the restaurant locations and other relevant data. Folium allows us to create interactive maps and overlay different layers of information,such as demographic data, competitor locations, and popular venues. Geospatial analysis plays a crucial role in location analysis for restaurants. We will utilize Python packages likegeopy and folium to perform geocoding, mapping, and spatial analysis. Geopy provides geocoding functionality to convert addresses into geographic coordinates, enabling accurate mapping and visualization of the restaurant locations and other relevant data. Folium allows us to createinteractive maps and overlay different layers of information, such as demographic data, competitor locations, and popular venue.

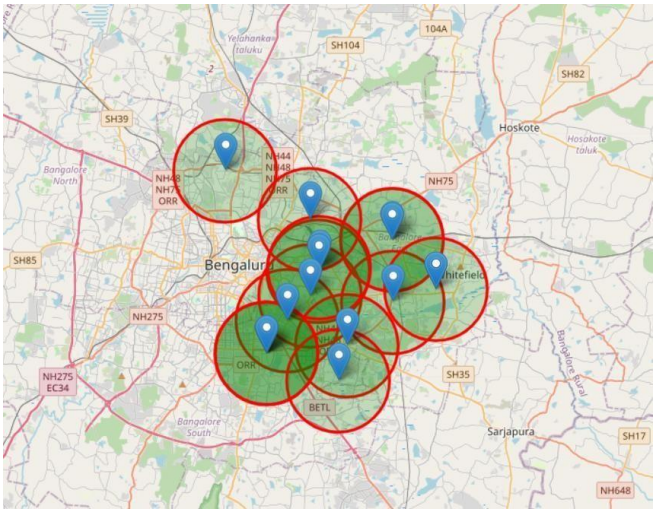

Fig 3.2 Location of Table: 1 Plotted on a map

Before applying machine learning algorithms [15], the data needs to be preprocessed to ensure compatibility and optimize performance. This involves tasks such as scaling numerical features, encoding categorical variables, and handling any remaining missing values. The preprocessing steps provided by sci-kit-learn, such as feature scaling and one-hot encoding, will be utilized for this purpose. Clustering analysis is employed to identify distinct groups or clusters within the dataset. In the context of restaurant location analysis, clustering can help identify areas with similar characteristics, such as demographics and venue preferences. Algorithms such as k-means clustering or hierarchical clustering will be applied to segment the target area into meaningful clusters. The scikit-learn package provides efficient implementations of these clustering algorithms. In addition to traditional machine learning techniques, deep learning approaches can be explored to enhance the accuracy and efficiency of location analysis. Deep learning models, such as neural networks, can be utilized to extract intricate patterns and relationships from the data. For instance, convolutional neural networks (CNNs) can be employed to analyze spatial data, such as maps or images, to identify significant features and patterns related to suitable restaurant locations. Deep learning frameworks like TensorFlow or Keras can be used to implement and train these models.

4. RESULTS AND DISCUSSION

4.1 Data Preprocessing and Exploratory Analysis Results

The initial phase of data preprocessing involved the careful examination and refinement of the collected data to enhance its quality and suitability for subsequent analysis. Various cleaning techniques were applied to eliminate inconsistencies, errors, and missing values from the dataset. Additionally, the data was transformed and standardized to ensure compatibility and comparability across different variables. These preprocessing steps aimed to improve the overall reliability and integrity of the data for further exploration and analysis.. Missing values were handled, variables were standardized, and data formats were adjusted. The Exploratory Data Analysis (EDA) provided valuable insights into the dataset, revealing patterns and relationships. Visualizations using Matplotlib and Seaborn showcased the distribution of variables, identified outliers, and explored correlations. The EDA helpedin formulating hypotheses and guiding subsequent analysis steps.

4.2 Analysis Results

Geolocational analysis using folium enabled the visualization of data on maps and identification of potential restaurant locations. Geocoding techniques converted addresses into geographic coordinates, facilitating accurate mapping of restaurant locations, demographic information, competitor locations, and popular venues. Folium allowed for the creation of interactive maps, providing a comprehensive view of the target area and its characteristics. The geospatial analysis provided valuable insights into spatial patterns and relationships, helping in the identification of suitable areas for restaurant establishment.Machine learning techniques using sci-kit-learn were applied to segment the target area based on demographic and venue characteristics. Data preprocessing steps, such as feature scaling and encoding, were performed to ensure compatibility with the machine learning algorithms. Clustering algorithms, such as k-means or hierarchical clustering, were applied to identify distinct groups or clusters within the dataset. The machine learning results provided segmentation of the target area, revealing areas with similar demographics and venue preferences, aiding in the decision-making process for restaurant locations.

An elbow method is a popular approach in data science that helps identify the ideal number of clusters in a dataset, especially in unsupervised learning tasks like clustering analysis. It derives its name from the visual representation of the method, where the plot of the number of clusters against the corresponding within-cluster sum of squares (WCSS) resembles the shape of a bent elbow. The main goal of the elbow method is to locate the point on the plot where the rate of decrease in WCSS starts to stabilize. This point is typically regarded as the optimal number of clusters to choose for the analysis.

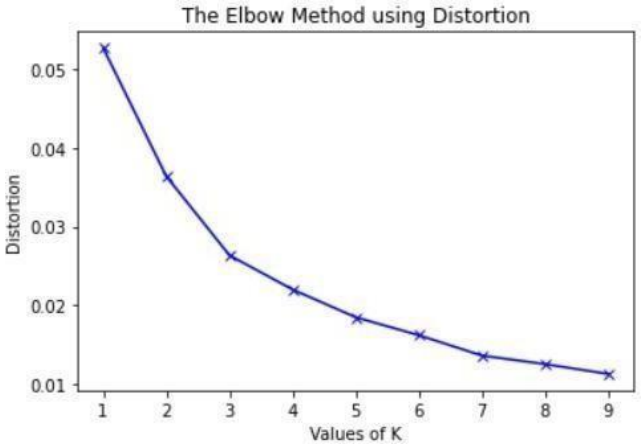

Fig 4.1 Elbow Method

Besides conventional machine learning methods, this study also investigated the application of deep learning techniques to improve the precision and effectiveness of location analysis. Deep learning models, notably convolutional neural networks (CNNs), were employed to extract complex patterns and connections within the data, specifically focusing on spatial data such as maps or images. By leveragingthe capabilities of deep learning, the study aimed to unlock deeper insights and enhance the overall performance of location analysis in the context of finding optimal restaurant locations. The deep learning results provided further insights into significant features and patterns related to suitable restaurant locations, potentially improving the accuracy of location analysis.

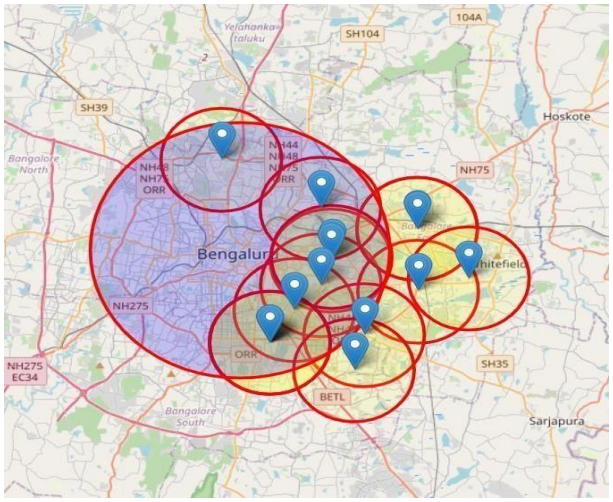

Fig 4.2 Location on Map which can be Targeted

4.3 Discussion of Findings

The findings from the data preprocessing, exploratory analysis, geospatial analysis, machine learning, and deep learning stages were collectively discussed to provide a comprehensive understanding of the optimal location for a restaurant. The discussion focused on the key insights gained from each stage, such as identifying areas with high customer demand, understanding demographic preferences, analyzing competition, and recognizing spatial patterns. The implications of the findings were discussed in relation to the objectives of the research and the practical implications for restaurant owners and stakeholders. Limitations of the study, such as data availability or modeling assumptions, were also addressed.

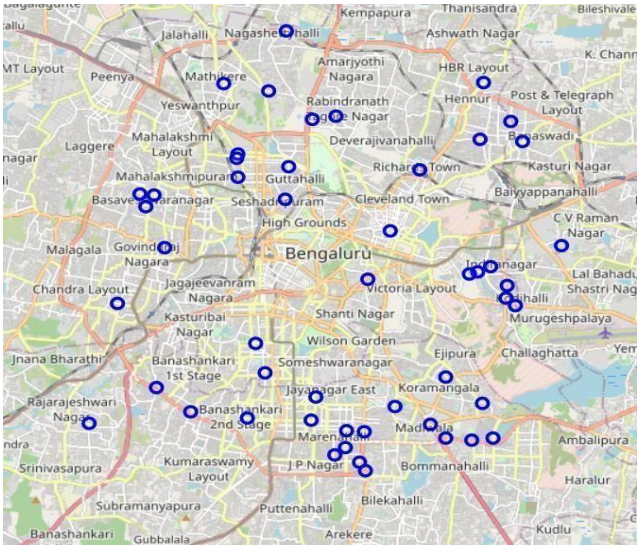

Fig 4.3 Potential Locations

5. CONCLUSION

In summary, this research paper presented a comprehensive methodology for determining the best location for a restaurant by utilizing geospatial analysis techniques and harnessing the power of the Foursquare API. The study involved gathering and preprocessing relevant data to ensure its quality and compatibility. Exploratory data analysis techniques were then applied to gain insights and identify patterns within the dataset. Geospatial analysis was conducted using tools like Geopy and Folium to visualize and analyze location-based information. Additionally, machine learning techniques, including data preprocessing and clustering analysis, were employed using sci-kit-learn to further enhance the analysis. Furthermore, the study briefly explored the potential of deep learning for location analysis. The results and findings obtained from each stage of the methodology were discussed and evaluated. Finally, the paper concluded with a summary of the study, implications, and recommendations for restaurant owners, as well as limitations and suggestions for future work.

Overall, the research paper aimed to provide valuable insights and practical guidance for decision-makers in the restaurant industry when selecting the optimal location for their establishments. Collection, preprocessing, exploratory analysis, geospatial analysis, machine learning, and deep learning techniques. The results obtained from each stage provided valuable insights into suitable restaurant locations based on demographic characteristics, venue preferences, and spatial patterns.

5.1 Implications and Recommendations for restaurant owners

The findings of this study have several implications for restaurant owners and stakeholders. By utilizing geospatial analysis and data-driven approaches, restaurant owners can make informed decisions about the optimal location for their establishment. The analysis helps in identifying areas with high customer demand, understanding demographic preferences, and assessing competition. The insights gained can guide restaurant owners in selecting locations that offer visibility, accessibility, and potential for business growth. It is recommended that restaurant owners leverage the methodologies and techniques presented in this research to make data-driven decisions and improve their chances of success..

6. REFERENCES.

1. G. Saravanakumar Pichumani, T. V. P., Sundararajan, Rajesh Kumar Dhanaraj, Yunyoung Nam and Seifedine Kadry. (2021). "Ruzicka Indexed Regressive Homomorphic Ephemeral Key Benaloh Cryptography for Secure Data Aggregation in WSN," Journal of Internet Technology, vol. 22, no. 6 , pp. 1287-1297

2. Kumar, D. R., Krishna, T. A and Wahi, A. (2018). Health Monitoring Framework for in Time Recognition of Pulmonary Embolism Using Internet of Things. Journal of Computational and Theoretical Nanoscience,15(5),1598–1602.

Krishnasamy, L., Dhanaraj, R. K. Ganesh Gopal, D. Reddy Gadekallu, T. Aboudaif, M. K and Abouel Nasr, E. (2020). A Heuristic Angular Clustering Framework for Secured Statistical Data Aggregation in Sensor Networks.

3. Dhiviya, S., Malathy, S and Kumar, D. R. (2018). Internet of Things (IoT) Elements, Trends and Applications. Journal of Computational and Theoretical Nanoscience.

4. Rajesh Kumar, D and Shanmugam, A. (2017). A Hyper Heuristic Localization Based Cloned Node Detection Technique Using GSA Based Simulated Annealing in Sensor Networks. In Cognitive Computing for Big DataSystems Over IoT (pp. 307–335).

5. Rajesh Kumar, D and ManjupPriya S. (2013). Cloud based M- Healthcare emergency using SPOC. 2013 Fifth International Conference on Advanced Computing(ICoAC). 2013 Fifth International Conference on Advanced Computing (ICoAC).

6.Sathish, R and Kumar, D. R. (2013). Proficient algorithms for replication attack detection in Wireless Sensor Networks — A survey. 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN). 2013 International Conference on Emerging Trends in Computing, Communication and Nanotechnology.

7. Sathish, R and Kumar, D. R. (2013). Proficient algorithms for replication attack detection in Wireless Sensor Networks 2014; A survey. 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN). 2013 International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN).

8. Eriksson, Brian, and Mark Crovella. "Understanding geolocation accuracy using network geometry." *2013 Proceedings IEEE INFOCOM*. IEEE, 2013.

9. Rakhimberdiev, E., Senner, N. R., Verhoeven, M. A., Winkler, D. W., Bouten, W., & Piersma, T. (2016). Comparing inferences of solar geolocation data against high-precision GPS data: annual movements of a double-tagged black-tailed godwit. *Journal of Avian Biology*, *47*(4), 589-596.

10 Han, Bo, Paul Cook, and Timothy Baldwin. "Text-based twitter user geolocation prediction." *Journal of Artificial Intelligence Researc* 49 (2014): 451-500.

11Kishan, S. R., Kakunuri, K. S., Parasa, A. R., Patlolla, V., & Gamini, P. (2022, December). Exploratory Analysis on Geo-Locational Data. In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 1-4). IEEE.

12. Soumya Ranjan Jena., Raju Shanmugam, Rajesh KumarDhanaraj and Kavita Saini. (2019). Recent Advances and Future Research Directions in Edge Cloud Framework. International Journal of Engineering and AdvancedTechnology.

13. Hasan, Samiul, Satish V. Ukkusuri, and Xianyuan Zhan. "Understanding social influence in activity location choice and lifestyle patterns using geolocation data from social media." *Frontiers in ICT* 3 (2016): 10.

14. Shamini, P. Baby, et al. "Exploratory Spatial Data Analysis (ESDA) Based on Geolocational Area Check for updates." *Futuristic Communication and Network Technologies: Select Proceedings of VICFCNT 2021, Volume 1*: 167.

15. Shamini, P. Baby, et al. "Exploratory Spatial Data Analysis (ESDA) Based on Geolocational Area." *Futuristic Communication and Network Technologies: Select Proceedings of VICFCNT 2021, Volume 1*. Singapore: Springer Nature Singapore, 2023. 167-174.