EXPLORATORY ANALYSIS OF GEOLOCATIONAL DATA

Likhitha D, Pooja K, Ashwini, Bhagyashree Shankar, Gauri Sameer Rapate
Dept. of Computer Science
PES UNIVERSITY
Bengaluru, India

Abstract—**This paper is about development of a website to find best accommodation for people or wishing to relocate based on the budget and proximity to the desired location. The user can log in and search for the houses based on the desired location and budget and user can also contact owner and can view the place with 360-degree view of image. For this we use K-means Clustering algorithm to refine the search.. Grocery shopping is the second objective of the proposed work.**

Index Terms—**K-means Clustering Algorithm, User profile, latitude and longitude.**

I. INTRODUCTION

In this paper, the goal is to create a website that will assist residents and the people intending to relocate in finding the best housing options based on the needs, preferences, and proximity to the new location. The customer logs in and looks for properties in the area and price range they choose. Using pictures and a 360-degree virtual tour, user can also get in touch with the owner and see the property. To accomplish this, K-Means clustering is being used to find the best lodging for Bangalore-based or relocating customers. By classifying dwellings for users based on the "Amenities," "Budget," and "Proximity" preferences. The idea saves clients time and aids in helping the user to choose the best lodging options based on the Amenities. With regard to this project, The objective is to give consumers a user-friendly interface that enables user to look for local homes, obtain housing evaluations, prices, and ratings, as well as purchase groceries, dairy products, snacks, and drinks.

II. LITERATURE REVIEW

A. Development of online based smart house renting web application.

The research's objective is to provide a standard web-based online platform for both tenants and property owners so that everyone may take advantage of it. This essay explains the development of a web-based application for the citizens of Identify applicable funding agency here. If none, delete this.

Houses for rent with intelligent communication with the owner are available for registration by both tenants and home owners. Benefits: The web application is very effective and user- friendly. Limitations: Data analysis will be challenging when working with large datasets. I need a quick internet connection.

B. Exploratory Spatial analysis of housing prices obtained from web scraping technique

The purpose of the study is to verify the spatial auto-correlation between the mean house prices in Salvador that were obtained using web scraping methods on online platforms. Web scraping technology is used in the case study as a reference for current data on the formal real estate market that operates in the urban space, which favor's greater updating of online platforms. This is due to the speed, low costs of access, adhesion, and popularity among advertisers. 1.Advantages We selected it due to its effectiveness, versatility, cost, and speed in extracting data at scale from websites.
2.Limitations: It takes time to perform data analysis of information obtained through web scraping. This frequently takes a lot of time.

C. Development and evaluation of mobile application for room rental information with chat and push notification.

The proposed method develops a mobile application with chat and push notification capabilities to provide details on the amenities offered, the cost, the number of rooms that are available, and even the owner's contact details. 1.Techniques: The server is accessible via a REST web service, and XMPP protocol is used to convey information to all listening in- stances, mostly to let people know whether they are available to chat or busy. 2.Benefits: This app has special conversation and broadcast features (push notification). Examples of tech- nologies that can be used to exchange housing information and communicate between room seekers and building admin- istrators include REST web services and XMPP. 3.Limitations: The database is quickly updated with the application's data (no backend application).

D. Exploratory data analysis, visualizations, interactive plots,animations insights into the Airbnb data.

Data quality assessment Examples of key feature engineering include user review mining, comment analysis using a word cloud, building word vectors from reviews, and cat- egorical variable analysis.

1.ADVANTAGES: With the help of this study, which involved exploratory data analysis and visualization, we discovered some useful information on the Airbnb rental industry. Rentals nearer to popular city attractions cost more. Prices are higher for rentals that the host rates highly for the area. The annual growth in rental average prices c c is correlated with demand. By borough, the overall number of listings by type varies. The likelihood of a host getting ”promoted” to the level of the super host tends to be directly correlated with ratings and response rates. The importance of the environment, setting, and cleanliness is indicated by words like ”peaceful,” ”walkable,” ”clean,” and ”spotless” that are associated with the word ”comfortable”.

III. PROPOSED METHODOLOGY

The project's approach would start by grouping Bangalore's neighbourhoods according to their latitude and longitude co-ordinates. The least error ideal number of clusters is identified and clustered in accordance with the Unsupervised Machine Learning approach K-Means Clustering and Elbow Curve method.. The K-Means technique finds k centroids and then assigns each data point to the closest cluster, minimising the size of the centroids. Pre-processed data is stored in this clustered area for later use. Using the user's latitude and longitude as a starting point, the distance is used to determine which cluster the user belongs to in order to make the best classification of the cluster for its representation. The Bangalore venue data is handled for missing, duplicate, and inconsistent values before being cleaned, pre-processed , and stored. The same data is used for exploratory data analysis and visualization
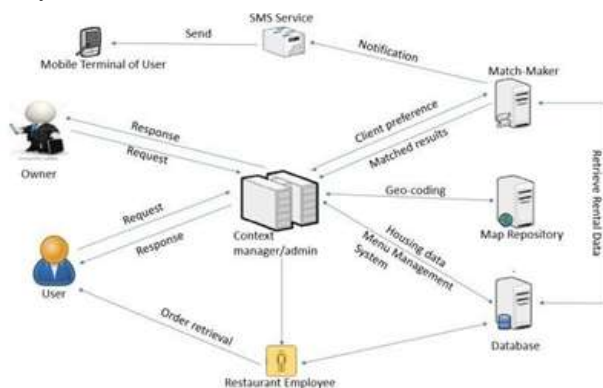

Fig. 1. Architecture Diagram.

A. K-Means Clustering

K-means clustering is an unsupervised machine learning algorithm that finds a rule to group similar data together.

Non-hierarchical clustering requires that the starting partition/number of clusters is known a priori. We want to partition the data points into k number of clusters so that the within- cluster variation for all clusters is as small as possible. Steps of k-means clustering • Select k number of clusters

• Initialization by randomly assigning each observation to one of the k clusters. • Iterate the following two steps until the cluster assignment stops changing: • Compute the centroid for each cluster • Assign/re-assign each observation to the cluster with the closest centroid

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2 \qquad (1)$$

Where, $[x_i - y_j]$ is Euclidean distance between x_i and y_j, c_i is the number of data points in ith cluster, 'c' is the number of cluster canters. Here, it's used to organize large location coordinates of unlabelled data (latitude and longitude) to generate clusters of Bangalore areas. The Elbow Method is one of the most widely used techniques for figuring out the optimal value of k, In the data-set k is turned out to be 5.

IV. RESULTS

A. Exploratory data analysis

The data consists of 200 rows and four columns, The columns include, address, metro, local market, bus stop.

1) Elbow plot: Plotting the within-sum-squares of each cluster for the different number of clusters. The within-sum- squares will decrease with each additional cluster, but we will look for the elbow, where the within-sum-squares decrease the most from the previous number of clusters and then flattens out for additional clusters.The method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use. x-axis represents number of clusters and y-axis representsthe wcss (within cluster sum of square). :
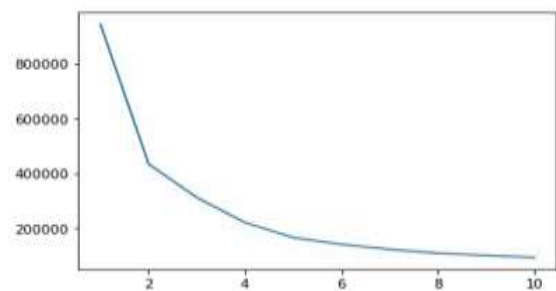

Fig. 2. Elbow Curve Method

2) The Data points are scattered randomly before and

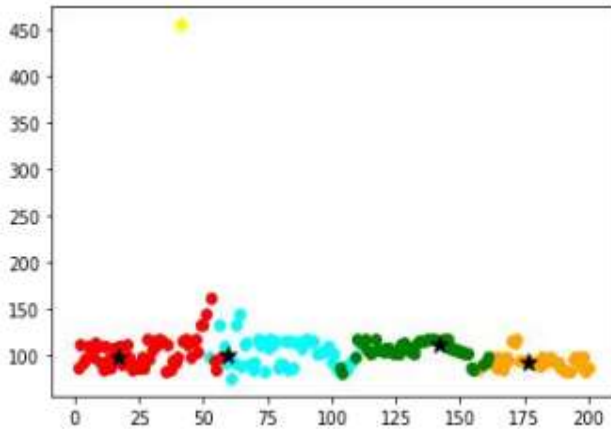afterapplying the k-means clustering algorithm. :
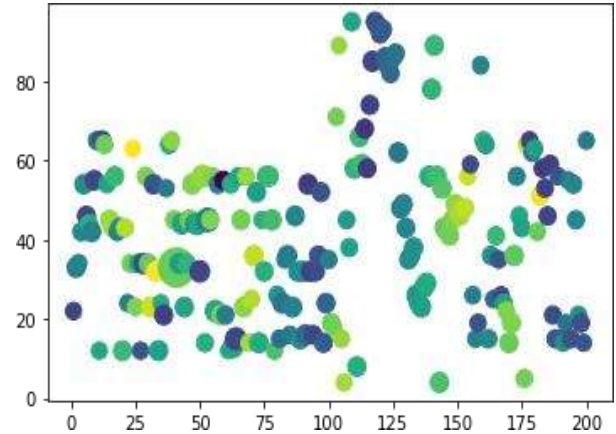

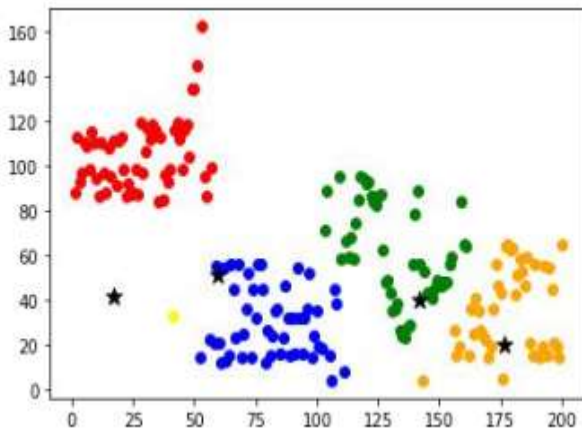Fig. 3. K-means Clustering Visualized.

Fig. 4. K-means Clustering Visualized.

Fig. 5. K-means Clustering Visualized.

Fig. 6. Bangalore Venue Before Clustering.
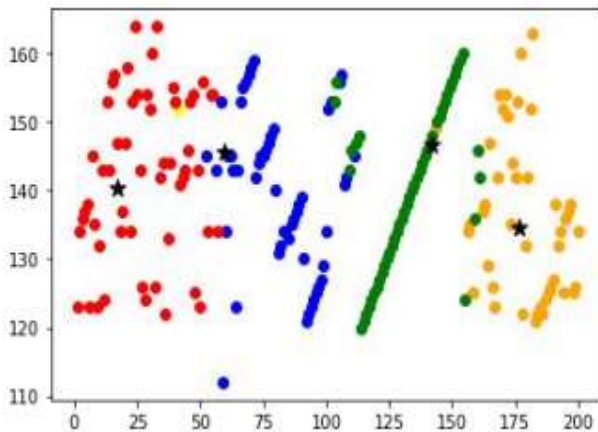
V. DISCUSSIONS

The evaluation of the user's satisfaction is the most vital part of the assessment of the system as it is the clear indicationof the degree to which a user's likes and interests are catered. The data is also limited as the range taken into consideration is just Bangalore alone. The system searches the house based on user venue. The user searches the rooms based on the price and the locations which is near to his/her college or work placeaccording to their needs. The user has to subscribe to get the contact details of the owner so that the owner details will be secured. And the user can also order the grocery things such as vegetables and the dairy products, snacks and beverages. K-Means clustering to find the best lodging for Bangalore-based or relocating customers

VI. CONCLUSION

In this study, a website to find best accommodation for people in Bangalore or wishing to relocate based on their budget and proximity to the desired location has been developed. In this study, it was proposed to use machine learning technique i.e, K-means clustering algorithm to identify best places for accommodation. The developed solution consists of an easy- to-use interface with usability being an important factor.

ACKNOWLEDGMENT

REFERENCES

[1] Dipta Voumick, Prince Deb, Sourav Sutradhar, Mohammad Monirujjaman Khan "Development of Online Based Smart House Renting Web Application", published by Journal of Software Engineering and Applications, Vol.14 No.7, 2021

[2] S.R. Manalu, A. Wibisurya, N. Chandra and A. P. Oedijanto, "Development and evaluation of mobile application for room rental information with chat and push notifica- tion," 2016 International Conference on Information Man- agement and Technology (ICIMTech), 2016, pp. 7-11, doi: 10.1109/ICIMTech.2016.7930293

[3] Erguden, S, Low cost housing policies and constraints in developing countries, International conference on spatial development for sustainable development, Nairobi 2001.

[4] Priya Gupta, 2 Surendra Sutar, "Multiple Targets Detection And Tracking System For Location Prediction", Inter- national Journal of Innovations in Scientific and Engineering Research (IJISER), Vol.1, No.3, pp.127-130,2014.

[5] Jia Sheng , Ying Zhou, Shuqun Li "Analysis of rental housing paper Market Localization"2nd International Confer- ence on Education, Management and Social Science (ICEMSS2014).

[6] Shriram, R.B., Nandhakumar, P., Revathy, N. and Kavitha, V. House (Individual House/Apartment) Rental Man- agement System. International Journal for Computer Science and Mobile Computing, 19, 1-43, 2019.

[7] Gommans, H.P., Njiru, G.M. and Owange, A.N. (2014) Rental House Management System. International Journal of Scientific and Research Publications, 4, 1-24.

[8] Nandhini, R., Mounika, K., Muthu Subhashini, S. and Suganthi, S. (2018) Rental Home System for Nearest Place. International Journal of Pure and Applied Mathematics, 19, 1681.

[9] Bristi, W.R., Chowdhury, F. and Sharmin, S. (2019) Stable Matching between House Owner and Tenant for De- veloping Countries. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, 6-8 July 2019, 1-6.

[10] Benjamin, D.J.The Environment and Performance of Real Estate. Journal of Real Estate Literature, 11, 279-324, 2003.

[11] Cooper, M, Ideas to develop a literature review, vol. 3, page, 39, 1998.

[12] Golland , A, Housing supply, profit and housing pro- duction: The case of the United Kingdom, Netherlands and Germany, Journal of Housing and the Built Environment, vol.11, no1, 1996.

[13] Priya Gupta, 2 Surendra Sutar, "Multiple Targets Detection And Tracking System For Location Prediction", Inter- national Journal of Innovations in Scientific and Engineering Research (IJISER), Vol.1, No.3, pp.127-130,2014.

[14] Macoloo, G, The changing nature of financing low in- come urban housing development in Kenya, Housing Studies, vol. 9, Issue 2, pages 18-92, 1994.

[15] Kim H, Shin G. On predictive routing of security contexts in an all-IP network. Secur Commun Netw, 3: 4–15, 2010.