“Shopping Mall Site Selection using Machine Learning”

September 6 , 2022

- **By Suraj Maurya**

ACKNOWLEDGEMENT

It gives us a great pleasure to present our project on Machine Learning. This is our milestone in b.tech in computer science and engineering.

We would like to express our sincere thanks to all the mentors who helped us throughout the project. I would like to acknowledge the help and guidance provided by our project guide Feynn labs services in all places during the presentation of this project.

I am thankful to our honourable Feynn labs services . We are also thankful to the staff member of the Faculty of Science for their moral supports towards the project.

Contents

Introduction

Domain - Retail

The Domain of this Project is Retail. In this project we are going to suggest the best locality to open a shopping mall in the Pune city to a business man. With increase in the population, demands are also increasing. As demands are unlimited, and demands become habits and customs, this is the reason a huge number of opportunities in the market for businesses are being created on daily basis. This leads to an intense competition. So, businessmen need to consider many different factors to compete the competitors surrounding them. The factors such as locality, demography, population and many more are needed to be bought into consideration while starting a new business as they highly affect to the growth of a particular business.

Among all other factors mentioned above, location plays a very crucial role within the business. Selection of a site should be based on a systematic approach. As these are the building blocks of a business on basis which the future of that business totally depends. According to our survey many people did research on this topic of appropriate site selection for a store, and different people considered many different factors as mentioned above and used various regression techniques, algorithms and neural networks to get optimal results. This paper presents a method to solve the issue of appropriate site selection for a store, considering neighbourhood as a major factor as location decision relates to the entire physical structure of that particular outlet. As we get an idea from the neighbourhoods that what are the existing businesses running in that particular locality.

Motivation

Location selection is one of the key success factors of Shopping Mall businesses. Location determines the number of contacts with customers, business volume and total income of business. To prove the hypothesis of location as the decisive factor in Shopping Mall businesses, the following project applied scientific analysis and mathematical methods. The obtained findings are based on the analysis of Shopping Malls locations in Pune city. The purpose of analysis of these practical examples is to examine to what extent the businesses use location as a means of competitive advantage.

Problem Statement

One of the most important strategic decisions made by Shopping Mall businesses is where to locate their operations, because location is such a significant cost driver. The location decision often depends on the type of business. For Shopping Mall businesses, the strategy focuses on maximising revenue. When these businesses make decisions about location of its facilities they must consider the geographical area and the number of shopping malls already present near the localities. A fundamental decision Shopping Mall businesses make is whether to locate their facilities close to competitors or far from them. How these businesses compete and whether external factors force them to locate close to each other influence this decision.

Literature Survey

In this project our main aim was to study, what are the key factors behind start of any new business, that is how store locators choose locations to open a store in order to maintain their existence in this competitive world as well as maximise profit at the same time. It is well known that the location selection is vital for successful operation of businesses in stores. Especially in case of retail stores the location plays major role because for retailers availability of resources at their store location is mandatory. This is the main reason a good site selection strategy is thought to be an effective means to reduce cost and obtain high benefits in business. There are various studies on deriving the best parameters for location selection for opening new business in store at selected location. Regarding location analysis, researches are also done for a store in global context.

In a work by Divaries, Cosmas, Jaravaza in 2013["The Role of Store Location in Influencing Customers' Store Choice", Published in Journal of Emerging Trends in Economics and Management Sciences (JETEMS) 4(3):302-307 (ISSN: 2141-7016)] using trade area analysis proved that different locations have different trade area characteristics .

In 2018 another study was completed on site selection of retail shop by Luyao Wang, Hong Fan and Yankun Wang["Site Selection of Retail Shops Based on Spatial Accessibility and Hybrid BP Neural Network", Published by International Journal of Geo-Information, 29 May 2018], this study had provided a new method that selects site for opening a new business, which fills the gap in the site selection for small retail shops. The two-step model, including the spatial accessibility estimation process with gravity model and the market potential evaluation process with BP (backpropagation network) and PCA (principal component analysis) model makes the site selection convincing and near reality.

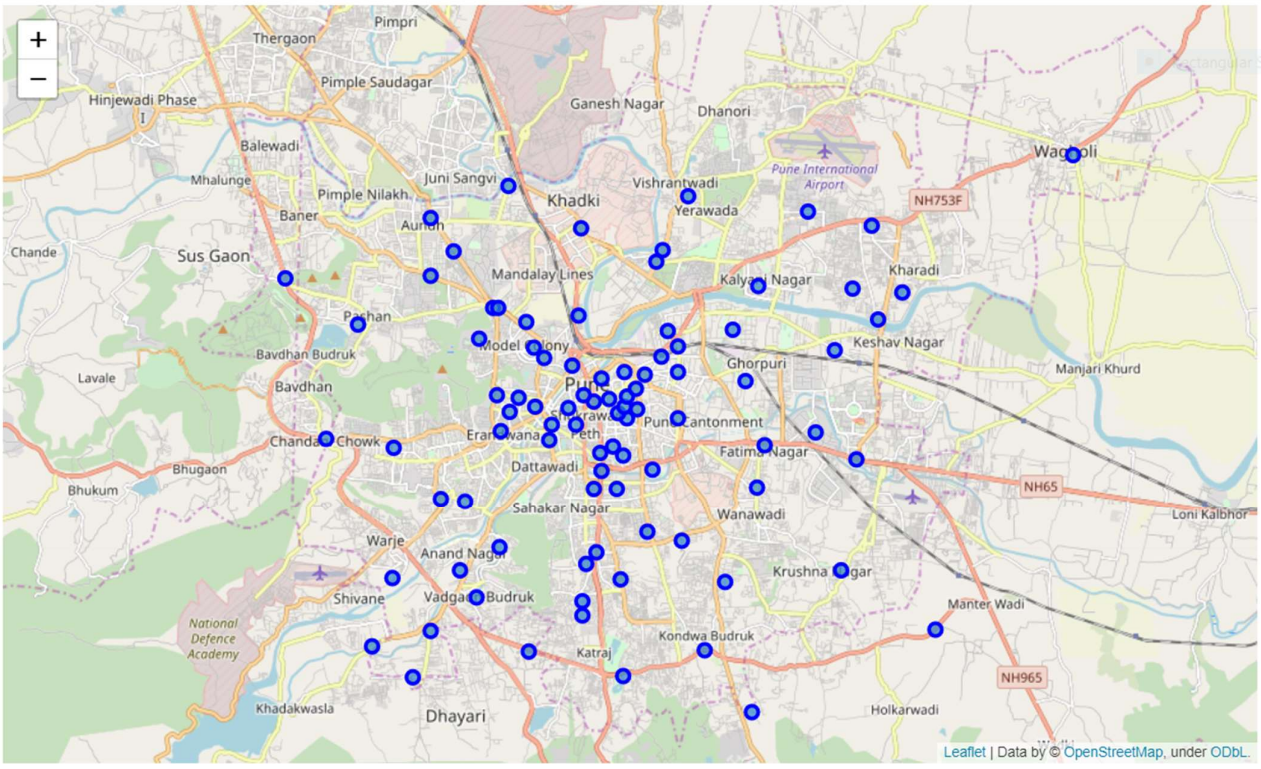
In 2019 a study was conducted by Mansi Karna["A study on selection of Location by retail chain: Big Mart", Published in International Journal of Research - *Granthaalayah*, 7(1), 383-395, 2019. https://doi.org/10.5281/zenodo.2561093] to understand multi criteria problem like how chain stores select most convenient locations, here Analytical Hierarchy Process was used which considered qualitative as well as quantitative approach in decision making.

After going through many different research papers and according to the researches done previously, we came to know that for different location selection many criteria comes into picture and all of them vary from each other. Each of those criteria have unique quality which is somehow required for selecting location of a store. Through this, we came up with the fact that a single model cannot explain all such criteria, as each of them is unique in their own contexts. So, it is very necessary to understand that factors such as geography, population, and total number of Shopping Malls present in a locality for opening a new shopping mall.

Solution Design

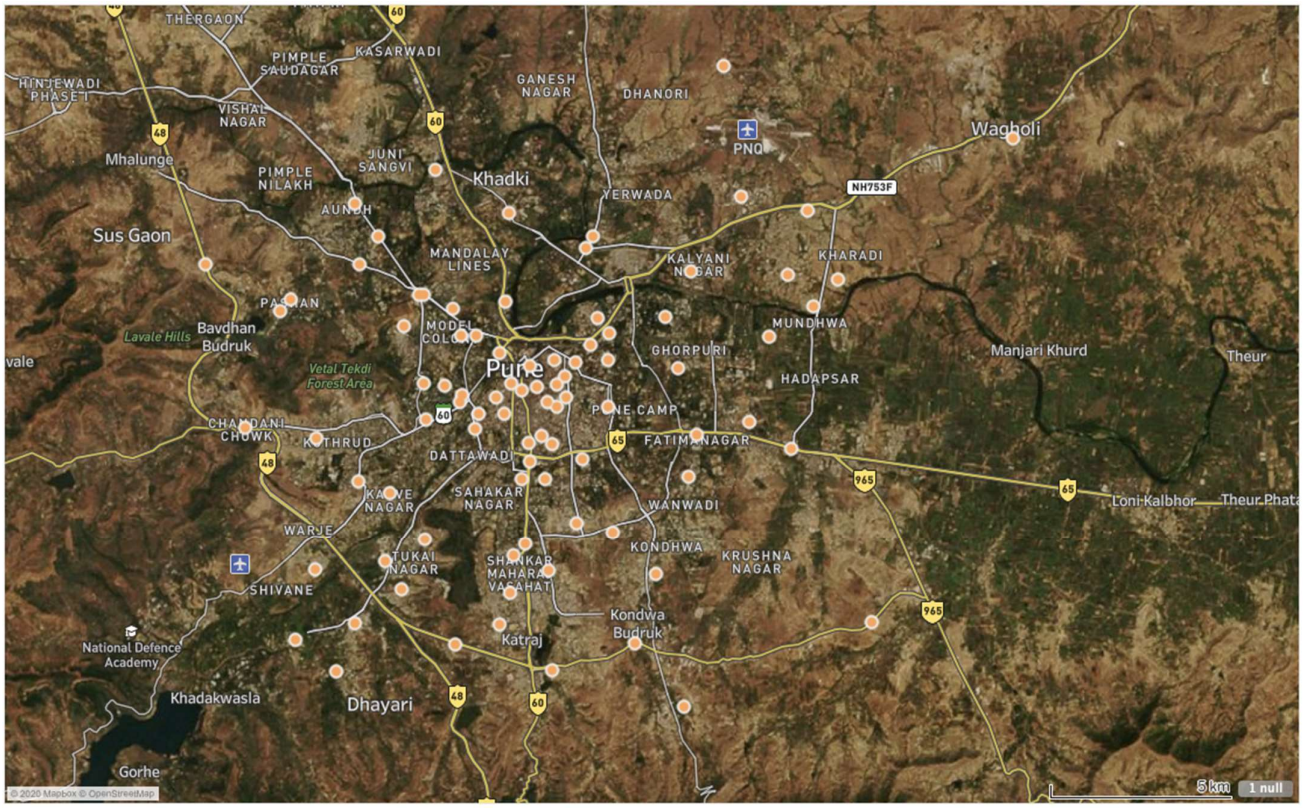
Solution Approach

To build this project we needed the data of all the localities of Pune city. As the appropriate data was not available on any medium, we came up with an idea to scrape the data from a website i.e. to use 'Web Scraping'. From the website https://www.mapsofindia.com/pune/localities/, where a list of all the 96 different localities of Pune were listed. So the data was scraped using an online platform namely 'Parse Hub' https://www.parsehub.com/ which is a powerful & open source web scraping tool that scrapes the data from any given URL. The data collected was then imported in a 'CSV' i.e. comma separated value format. We used 'Python' a robust open source programming language which supports multiple libraries and is very efficient to use. Dataset was imported into Python using 'Pandas' library for further processing. Our next aim was to get the geographical co-ordinates of all the localities of the Pune city which was done using 'Geopy'. Geopy is a library which helps to convert an address into a Latitude & Longitude values. Therefore, we retrieved the Latitudes and Longitudes of the localities and also found the Geographical Co-ordinates of the Pune city. Further 'Folium' library was used which is a map rendering library in python which provides a leaflet of a map using geographical co-ordinates and latitude,



longitude values.

Figure 1:Localities in Pune City



This above image is map of all the localities of Pune city. We have highlighted them using blue coloured dots.

Figure 2: Localities in Pune City Visualized by Tableau

We used Foursquare application programming interface which explores the neighbourhoods of a particular locality using latitude and longitude based on given radius and given limitations by using GIS (Geographical Information System) tool. GIS tools presents geographical data as this tool is designed to locate all the given points on the earth surface and map all of them according to their positions on the earth surface itself. By using this tool, we were able to get the most accurate datapoints. Further by providing a 5000meter radius we got datapoints i.e. all neighbourhoods in the specific radius. To generate this information, we needed to create an API request URL to Foursquare.

We proceeded with Data Wrangling and by using 'One Hot Encoding' we labelled the data into binary categories resulting into whether a venue exists or not. Then we proceeded with creating a new data frame specifically for the number of shopping malls and the population in the locality. The next step was to create clusters of multiple shopping malls existing in all localities of Pune city, for this approach we applied an Unsupervised Machine Learning Algorithm namely K-Means, this algorithm creates clusters for every unique data point based on centroids. Further we used Folium which is a library that supports python. With the help of this library we generated maps showing clusters.

Technology Stack

The different technologies used in this project are:-

1. Foursquare API (https://developer.foursquare.com/)

The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag. Additionally, Foursquare allows developers to build audience segments for analysis and measurement. JSON is the preferred response format.

2. ParseHub (https://www.parsehub.com/)

ParseHub is a visual data extraction tool that anyone can use to get data from the web. You'll never have to write a web scraper again and can easily create APIs from websites that don't have them. ParseHub can handle interactive maps, calendars, search, forums, nested comments, infinite scrolling, authentication, dropdowns, forms, Javascript, Ajax and much more with ease. ParseHub offer both a free plan for everyone and custom enterprise plans for massive data extraction.

3. Python

Python is the language that is stable, flexible, and provides various tools to developers. All these properties of Python make it the first choice for Machine learning. From development to implementation and maintenance, Python is helping developers to be productive and confident about the software they are developing. There are many benefits of using Python in Machine learning. These benefits are increasing the popularity of Python. That is why you see Python mostly in Machine learning Software.

4. Tableau

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into the very easily understandable format. Data analysis is very fast with Tableau and the visualizations created are in the form of dashboards and worksheets. The data that is created using Tableau can be understood by professional at any level in an organization. It even allows a non-technical user to create a customized dashboard.

Design Model

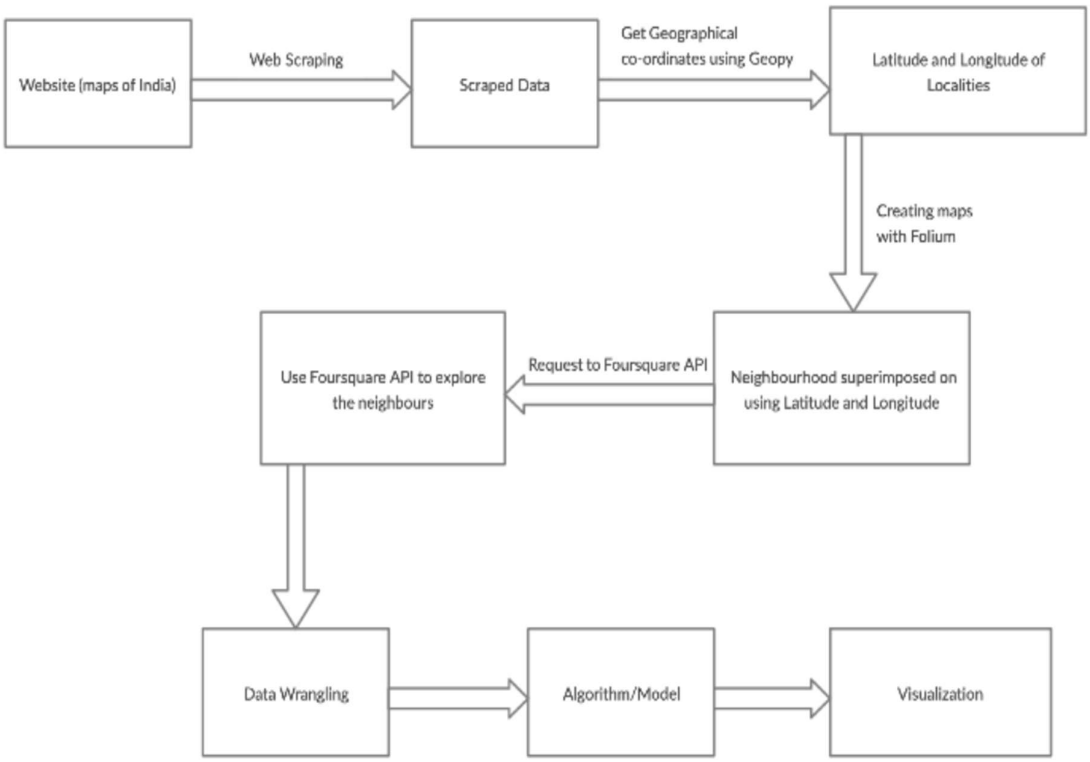


Figure 3: Architecture

Solution Implementation and Results

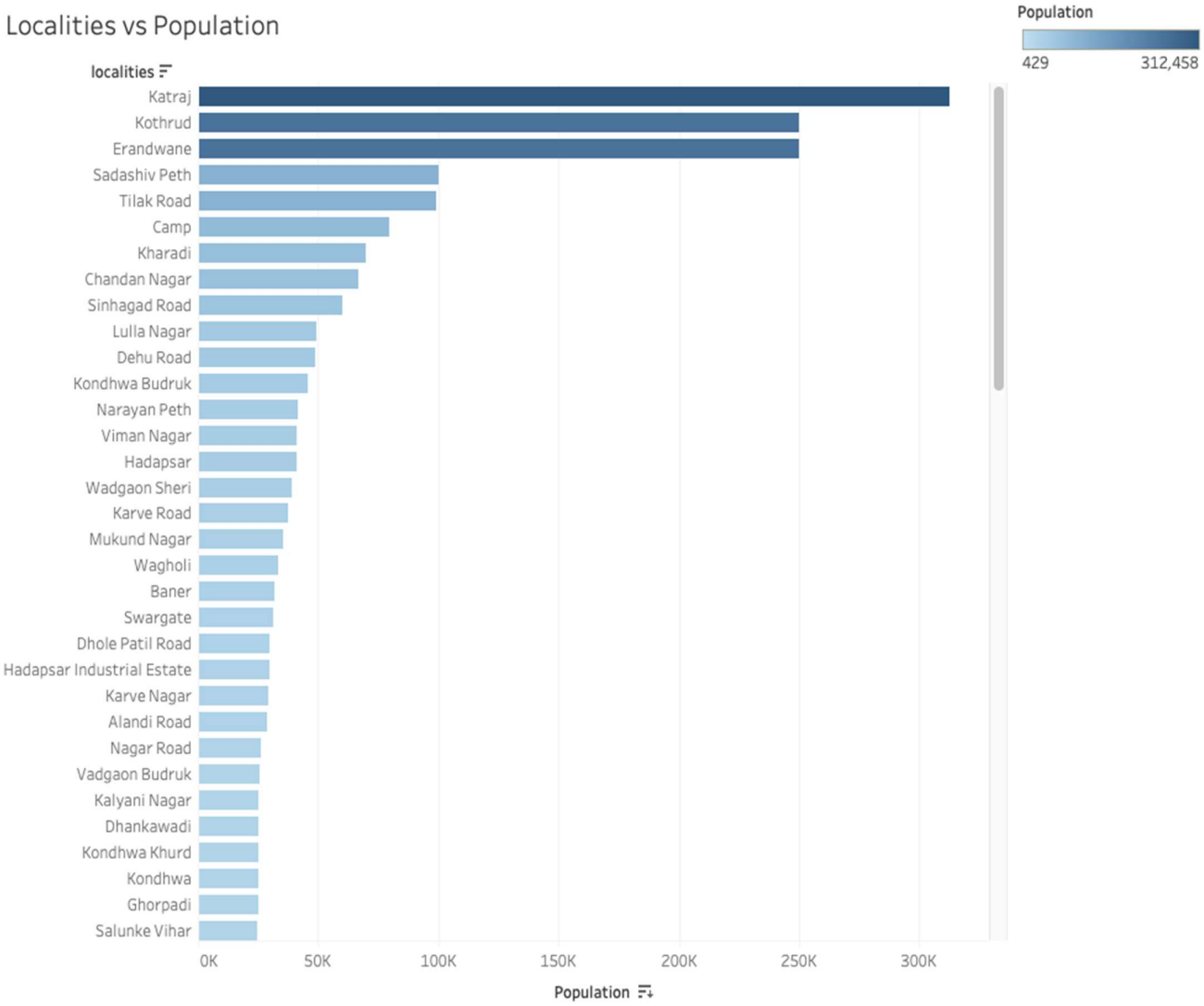
Obtaining Data

As the dataset was not readily available, we had to scrape the data. We created our first dataset from the website called Maps of India, URL - https://www.mapsofindia.com/. From this website we got the Localities name. There were 96 distinct localities present. Then by using Geopy(a library in python, used for generating the latitude and longitude of a location) we extracted the Latitude and Longitude of the localities from which, we then merged with the dataset which we got from Maps of India. So now the dataset had 4 attributes namely Localities_name, Localities_url, Latitude and Longitude. With the help of Fouresquare API we explored the neighbourhoods of the localities. We gave two parameters for exploring the neighbourhoods, radius = 5000m and LIMIT = 500. The radius indicates the total area to be explored from the give Latitude and Longitude, whereas LIMIT indicates the maximum venues to explore from the given point. Exploring the neighbourhoods gave us more 4 attributes namely VenueName eg: Restaurant name, Cafe name, Shopping Malls name, Bakery name, Pub name etc, VenueLocation in terms of Latitude and Longitude and VenueCategory eg: Gym, Super market, Coffee shop, Chocolate shop, Hotel etc. After this, we needed population data and number of malls present in the respective localities. We got the Population data from Pune municipal corporation website, URL - https://pmc.gov.in/mr and Number of Malls from the website named HOUSING.com, URL - https://housing.com. We didn't get enough information about Shopping malls in a specific Locality from HOUSING.com so we found the rest from Google by manually searching for number of malls in a locality.

Maps of India	https://www.mapsofindia.com/
	https://pmc.gov.in/mr
Housing.com	https://housing.com

EDA

We superimposed the localities on the Folium map with the help of their Latitude and Longitude. This gave us an idea of how spread the points are across the map. We then, plotted a Bar Graph of Localities vs Population and found that Katraj, Kothrud and Erandwane has



the highest population.

Figure 4: Localities vs Population

We also plotted a Bar Graph of Localities vs Shopping Malls and found that Yerawada, Shivaji Nagar and Hadapsar had highest number of Shopping Malls.

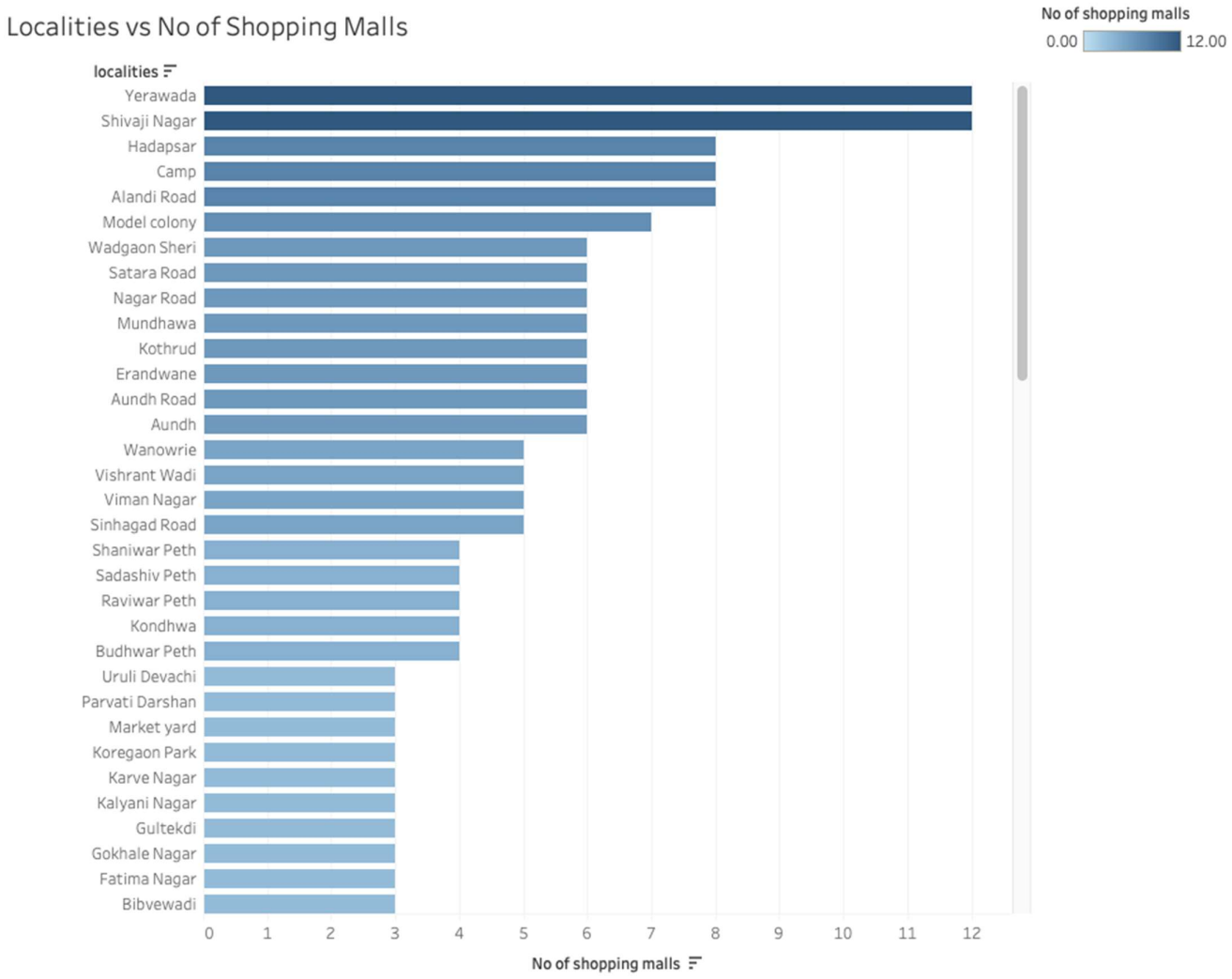


Figure 5: Localities vs No. Of Shopping Malls

Pre-Processing

As our dataset had numeric values except for Localities_name and VenueCategory, we decided to apply One Hot Encoding to the VenueCategory attribute. As we had 134 distinct VenueCategory, so we got 134 distinct attribute as well. Next we grouped rows by neighbourhood and took the mean of the frequency of occurrence of each category. This gave

us an idea about how many distinct neighbourhoods would be present at any locality. But our main target was Shopping Malls. So we dropped all columns except for VenueCategory with Shopping Malls. This way we got rid of unnecessary dimensions in our dataset.

Machine Learning Algorithms Used

K Means Clustering

The algorithm which we used for this problem statement is K - Means Clustering. With this algorithm we got clusters 3 clusters. The clusters are formed on the basis of number of shopping malls in a locality. The clusters with high number of shopping malls indicate that there is already a lot of competition in that locality and the idea of building a shopping mall should be moved to some other locality which has low number of shopping malls and high population.

K - Means Algorithm asks for the value of parameter n_clusters. Giving this value on the fly is not a good idea as it may form wrong clusters. So for that we used the Elbow Method. We gave a range from 1 to 10 for n_clusters, and found that for K = 3 was the optimal value of K.

We also did Silhouette Analysis on the different number of clusters and found K = 3 the optimal one because the Silhouette Coefficient Values for K = 3 were closer to 1. This states that the points in clusters were similar to each other and has dissimilarity with points which belongs to different clusters.

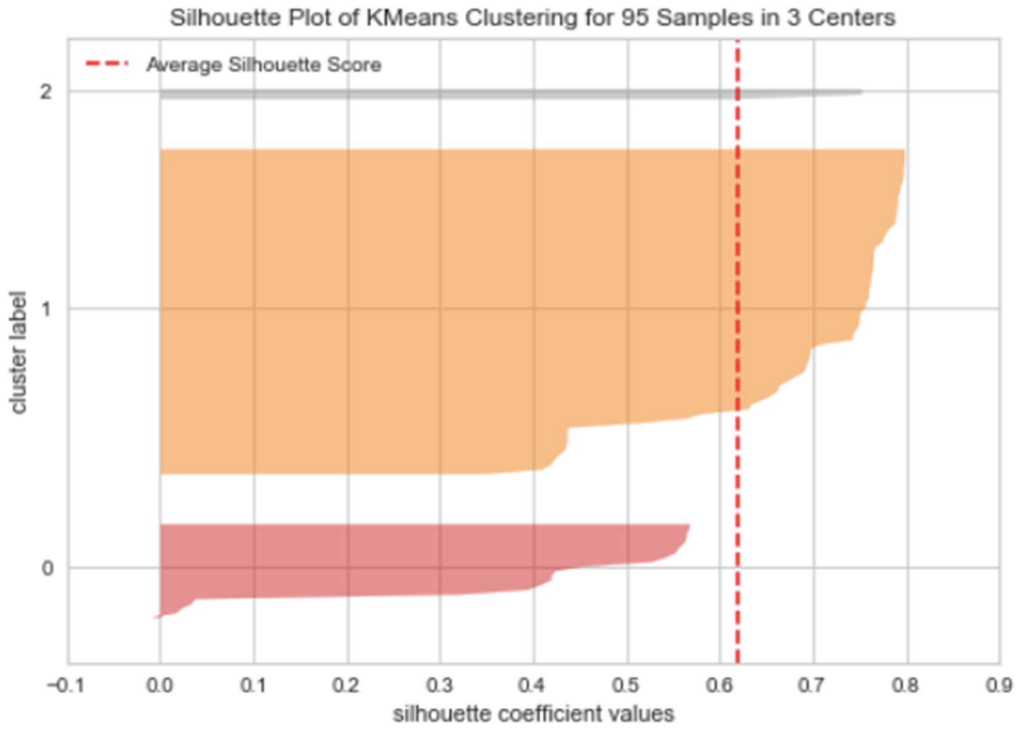


Figure 6: Silhouette plot for K = 3

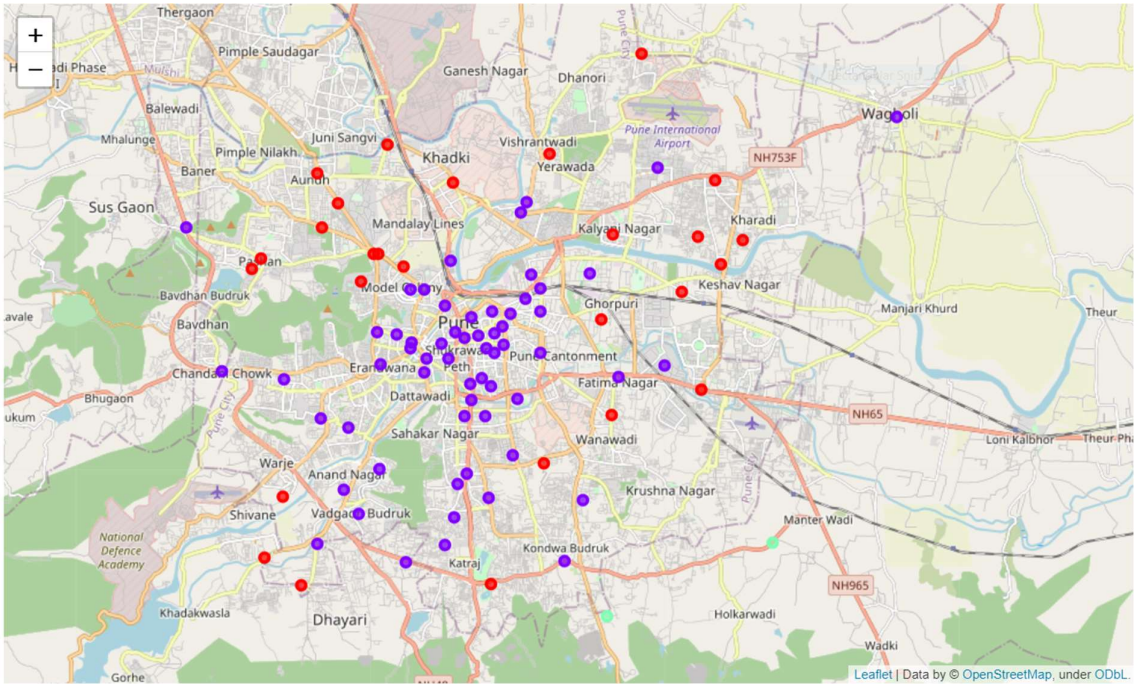


Figure 7: K-means Clustering on Map

The output of the above map is explained in the next section.

Results

The key insights that we got by K - Means Clustering :-

1. In Cluster 1 there is high Population and low number of Shopping Malls. So we can suggest a Businessman to open Shopping Malls in the localities belonging to cluster 1 as the scope of business would be high.

2. In Cluster 2 there is low Population and low number of Shopping Malls. So we can assume that the scope of business in the localities belonging to cluster 2 might be moderate.

3. In Cluster 3 there is low Population and high number of Shopping Malls. So definitely the competitions is more in this localities and it is not advisable to suggest business of Shopping Malls in the localities of cluster 3.

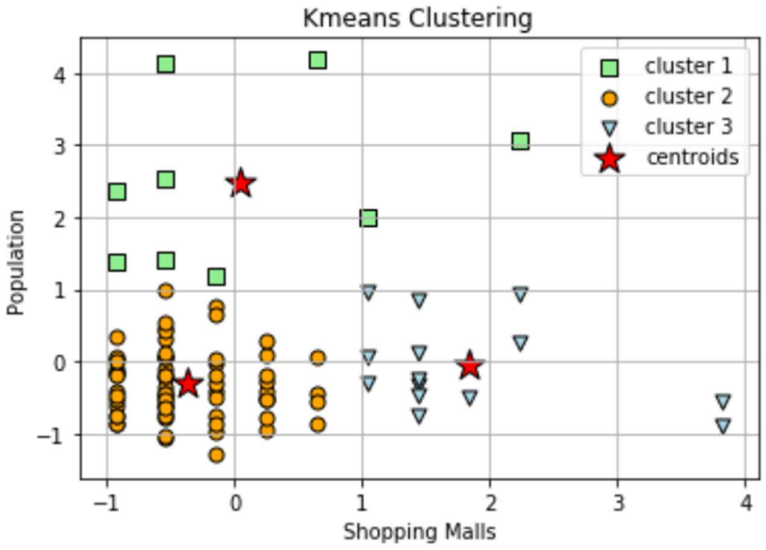


Figure 8: K-means Clustering

Conclusion and Future Work

Conclusion

In this particular project we tried to understand the major role which neighbourhood plays in order to select a particular location to open a new Shopping Mall at any site. As mentioned above we explored neighbourhoods of all localities of Pune city listed in the created dataset. With this we came to know that by exploring neighbourhoods it was much easier to do market analysis of each and every locality of the city. As this gave us accurate details of what all are the types of businesses currently running in all localities. Through which we got an idea of all the existing competitors of all types of businesses, this made decision making easier with respect to the formation of any new business. This project can help shopping mall businesses to find the right locality so that the business can sustain and grow.
In the above proposed model, we formed clusters by taking a specific venue category of Shopping Malls into consideration. In total 3 clusters are formed.

Limitations

The limitation of the project is that the data is a bit biased thus affecting the analysis.

Future Work

In the future, we can extend the scope of the project by suggesting almost any type of businesses which is listed in the Venue Category in our project. The only thing we would require is the data about the different Venue Categories and we would be able to suggest the best locality for that business. Also we would include data about the demographics which would support such scope of the project.

References

[1]. Maps of India https://www.mapsofindia.com/

[2]. Paresehub https://www.parsehub.com/

[3]. Foursquare https://foursquare.com/

[4]. Google Maps https://www.google.com/maps/

[5]. Hikmit Erbiyik, "Retail store location selection problem with multiple analytical hierarchy process of decision making an application in Turkey", Published by Elsevier Ltd., The 8th International Strategic Management Conference, 2012

[6]. Dongdong Ge and Luhui Hu, "Intelligent site selection for bricks-and-mortar stores", Published in Modern Supply Chain Research and Applications. Published by Emerald Publishing Limited, 2019. http://creativecommons.org/licences/by/4.0/legalcode

[7]. Mansi Karna and Anusha Rai, "A study on selection of Location by retail chain: Big Mart", Published in International Journal of Research - *Granthaalayah*, 7(1), 383-395, 2019. https://doi.org/10.5281/zenodo.2561093.

[8]. Drago Pupavac, "Choice of location for retail businesses" Submitted to Special Issue on Profit-Driven Analytics January 17, 2018

[9]. Luyao Wang, "Site Selection of Retail Shops Based on Spatial Accessibility and Hybrid BP Neural Network", Published by International Journal of Geo-Information, 29 May 2018.

[10]. Joshua K, "Retail Site Selection: A New, Innovative Model for Retail Development". Cornell Real Estate Review, 7(1), 1-26. Retrieved from http://scholarship.sha.cornell.edu/crer/vol7/iss1/17.

[11]. Divaries Cosmas, "The Role of Store Location in Influencing Customers' Store Choice", Published in Journal of Emerging Trends in Economics and Management Sciences (JETEMS) 4(3):302-307 (ISSN: 2141-7016).

[12]. Rogers, D.S., "Retail location analysis in practice", Research Review, Vol. 14 No. 2, pp. 73-78, 2017.

[13]. Shmoys and Tardos, "Approximation algorithms for facility location problems", Proceedings of the 29th Annual ACM Symposium on Theory of Computing, pp. 265-274.

[14]. Evaluating the Neighbourhood When Choosing a Business Facility

https://www.bizfilings.com/toolkit/research-topics/office-hr/evaluating-the-neighborhood-when-choosing-a-business-facility

[15]. Location needs of various business types https://www.inc.com/encyclopedia/site-selection.html

[16]. Site Selection Wikipedia https://en.wikipedia.org/wiki/Site_selection

[17]. Choosing a retail store location https://www.thebalancesmb.com/choosing-a-retail-store-location-2890245