Selection of location for Multiplex Establishment in Pune - India

A Project Report

Submitted in Partial Fulfillment of Requirements for IBM Data Science Professional Certificate Program on Coursera

Submitted By – Swaroop Todankar

Date – 20 August 2019

 [1]

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

1. Introduction

Indian Film Industry ranks 1st in number of films produced and admissions [1]. The number of screens for viewing films is 8 per million, which is very low as compared to western countries which are in the range of 120+ [2]. India has about only 2400 multiplex screens as compared to single screens which are 6700 [2].

Even though the multiplex are just 27 percent of the total screens, these contribute to about 45 percent of the total revenue of cinema [2]. This stresses the importance of such an establishment.

The main advantage of Multiplex is that multiple movies can be screened in at a single location and people are presented with a choice of entertainment. This also encourages the development of businesses like food court and merchandise industry which often compliment the movie industry. In this way, people not only find a getaway location for their weekends, but also all their other requirements are satisfied along with it.

With the advancements of film making and videography, more and more movies and plays are pushed to the audience every day. People spend most of their time engaged in work. Entertainment, whether it may be outdoors or indoors can be considered to be a major form of stress release. Outdoor entertainment can be concerts, fairs, meetups and many more things. Indoor entertainments include video games, watching movies, catching up on web series etc.

Considering movies and the whole market surrounded around it, Multiplex can serve the audience as well as develop a steady stream of income for the organization which runs such establishment.

The primary factors for the establishment of such facilities are the location of the establishments and the funding required. Most countries have private organizations interested to take up this business venture. The following question which arises is that – what are the factors that must be considered in order to select a location for such an establishment?

2. Business Problem

The main question is- If an organization is deciding to open a Multiplex establishment in Pune, which location would be the best considering all the factors?

This report aims to put forward an analysis in selection of location for a Multiplex establishment using location data obtained from foursquare API. The algorithms used can help to identify the perfect location considering proximity to nearby similar facilities and the demand for the services.

3. People interested in the Project – Target Audience

The target audience in this scenario are normal people who want to spend their leisure time for watching movies. The main factors contributing to the demand are time of travel and location.

The report also aims to make the business decision of construction of such an establishment by providing answers to exactly where can such a facility be constructed ?. This is meant to target the private organizations which are interested in running such a business venture.

4. Data required

The data required to build a model to suggest a location for the establishment is as follows:

4.1 Neighborhood Data: Data pertaining to neighborhoods of Pune city obtained from Wikipedia.

The data specifying the index and the neighborhoods of Pune city can be obtained through Wikipedia page: Neighborhoods of Pune city. This provides a table listing the neighborhoods.

Web scraping techniques such as using the Beautiful Soup library or the Wikipedia library can be used to convert this html data into a pandas data frame. This is helpful for analysis with python in Jupyter Notebook.

4.2 Location Data: Location data of these neighborhoods obtained from geocoder library of google or location data available online.

The latitudes and longitudes of neighborhoods of Pune city are required in order to access the third step of the model preparation. The latitude and longitude (co-ordinates) can be obtained using the geocoder library of google.

An alternative to the above step, if the geocoder library becomes unreliable, is obtaining a geospatial file detailing the location co-ordinates from web directly.

4.3 Venue Data: Data of venues in these neighborhoods, obtained from Foursquare API

Using the location data obtained in second step of model preparation, the Foursquare API is used to obtain the venue data of these neighborhoods. The data is then cleaned and the data pertaining to Multiplex establishments of each neighborhoods is obtained. The data is then used for machine learning algorithms to perform exploratory analysis in order to obtain results and make inferences.

5. Methodology

The following steps were employed to obtain the required results:

5.1 Importing Necessary Libraries

The first step is to import the necessary libraries and packages.

Numpy – For numerical calculations

Matplotlib – plotting and visualization

Pandas – Data manipulation

Geocoder – Obtaining location data

Folium – Creation of maps

Beautiful Soup – Web Scraping

Sklearn – Machine Learning

5.2 Web Scarping

Using the Beautiful Soup (bs4) package, the data from Wikipedia entry of 'Neighbourhoods of Pune' can be scrapped.

The data can be cleaned to obtain the required results and then stored in the form of a dataframe using Pandas library.

5.3 Visualization

Using the geocoder library the co-ordinates of Pune city are obtained and using the folium library, a map denoting the neighbourhoods as markers is visualized.

5.4 Obtaining Venues

By using the developer account of Foursquare API, venues data can be obtained. The details required to access the account are the credentials which are hidden in the code submitted.

Then using parameters such as setting search radius at 2000m and a limit of 100 venues, a venues list is obtained. This venue list is used to create a dataframe using pandas.

A quick check is carried out to verify if Multiplex as a category exists within it.

5.4 One Hot Encoding

The obtained data frame is one hot encoded using the get dummies method of pandas. Using the group by and mean, statistical information is obtained which is used to filter the dataframe to obtain only the values where “Multiplex” is positive.

5.5 Clustering

Using the sklearn library, K-Means clustering is applied on the data. K-Means is a clustering technique of Machine learning where depending upon the data (similarity or range), the data is divided within certain clusters.

The number of clusters is pre-defined in order to assign the centroids along with the cluster means can be calculated. The nearest data points to these centroids are grouped together in to similar clusters. Using the values of within cluster elements, a new centroid value is calculated and the process is iterated.

The number of clusters for this analysis was selected as 5. Obtaining the cluster labels from the results, the cluster data was merged with the dataframe and is presented as final dataframe for visualization.

5.6 Cluster Visualization

Using the folium library the visualization of clusters of venues in the Pune city are obtained.

6. Results

6.1 Clusters - After Cluster Visualization, the following clusters were obtained.

Cluster 0 – Red Colour

Table 1 – Cluster 0

	Neighbourhoods	Multiplex	Cluster Labels	Latitude	Longitude
0	Ambegaon Budruk	0.0	0	18.453520	73.838880
2	Balewadi	0.0	0	18.575980	73.779830
3	Baner	0.0	0	18.548200	73.773180
4	Bavdhan	0.0	0	18.517543	73.778533
8	Dhanori	0.0	0	18.578580	73.892670
9	Dhayari	0.0	0	18.447020	73.807570
10	Erandwane	0.0	0	18.509660	73.831240
11	Fursungi	0.0	0	18.473650	73.974730
15	Hingne Khurd	0.0	0	18.479800	73.830740
16	Kalas	0.0	0	18.578450	73.874880
17	Katraj	0.0	0	18.447320	73.864050
18	Khadki	0.0	0	18.561120	73.853010
19	Kharadi	0.0	0	18.544610	73.939250
20	Kondhwa	0.0	0	18.438260	73.898940
22	Kothrud	0.0	0	18.505170	73.802450
24	Markal	0.0	0	18.667570	73.952570
28	Pashan	0.0	0	18.536740	73.792900
29	Pirangut	0.0	0	18.511230	73.683170
30	Saswad	0.0	0	18.347370	74.029040
33	Vadgaon Budruk	0.0	0	18.467290	73.824730
34	Vishrantwadi	0.0	0	18.555330	73.874920
35	Wadgaon Sheri	0.0	0	18.537890	73.932670
36	Wagholi	0.0	0	18.579530	73.985290
38	Warje	0.0	0	18.472110	73.802130
39	Yerwada	0.0	0	18.544836	73.884677

Cluster 1 – Purple Colour

Table 2 – Cluster 1

	Neighbourhoods	Multiplex	Cluster Labels	Latitude	Longitude
14	Hadapsar	0.035088	1	18.502530	73.927060
23	Manjri	0.026316	1	18.481941	73.865628
26	Mundhwa	0.036585	1	18.530170	73.921250
27	Parvati	0.030303	1	18.486970	73.850070

Cluster 2 – Sky Blue Colour

Table 3 – Cluster 2

	Neighbourhoods	Multiplex	Cluster Labels	Latitude	Longitude
32	Undri	0.076923	2	18.4542	73.91719

Cluster 3 – Fluorescent Green Colour

Table 4 – Cluster 3

	Neighbourhoods	Multiplex	Cluster Labels	Latitude	Longitude
1	Aundh	0.017544	3	18.56345	73.81227
6	Dattawadi	0.020000	3	18.49994	73.83988
12	Ganesh khind	0.018182	3	18.54075	73.82923
13	Ghorpadi	0.012048	3	18.52233	73.89710
21	Koregaon Park	0.010000	3	18.53533	73.89382
31	Shivajinagar	0.010000	3	18.53724	73.83806
37	Wanowrie	0.021277	3	18.49538	73.90009

Cluster 4 – Yellow Colour

Table 5 – Cluster 4

	Neighbourhoods	Multiplex	Cluster Labels	Latitude	Longitude
5	Bibvewadi	0.054054	4	18.47187	73.86336
7	Dhankawadi	0.041667	4	18.46629	73.85324
25	Mohammedwadi	0.043478	4	18.47871	73.91591

6.2 Maps – The generated maps are as follows

The map of Pune with neighbourhoods superimposed on it

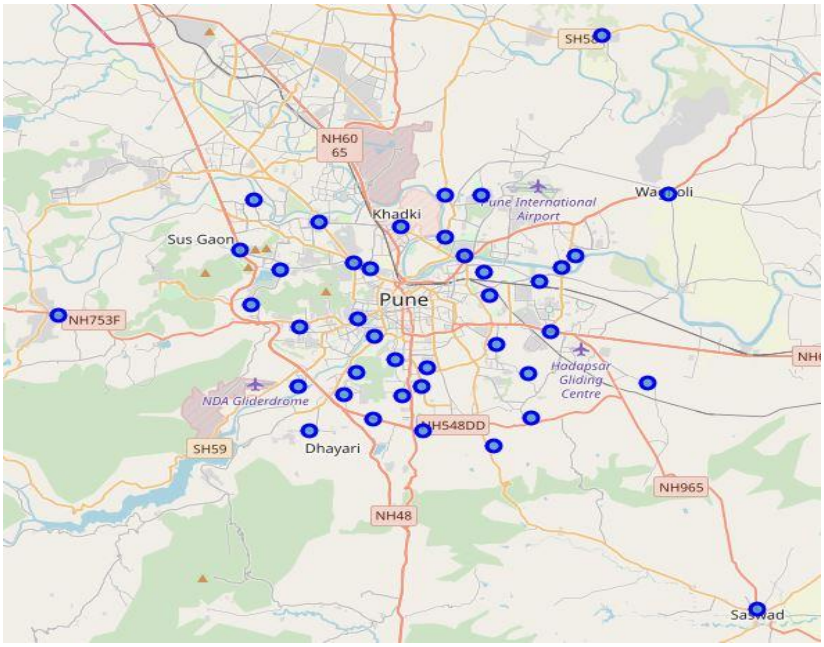

Figure 2 – Pune Neighbourhoods

The map of pune denoting the clusters of Multiplex:

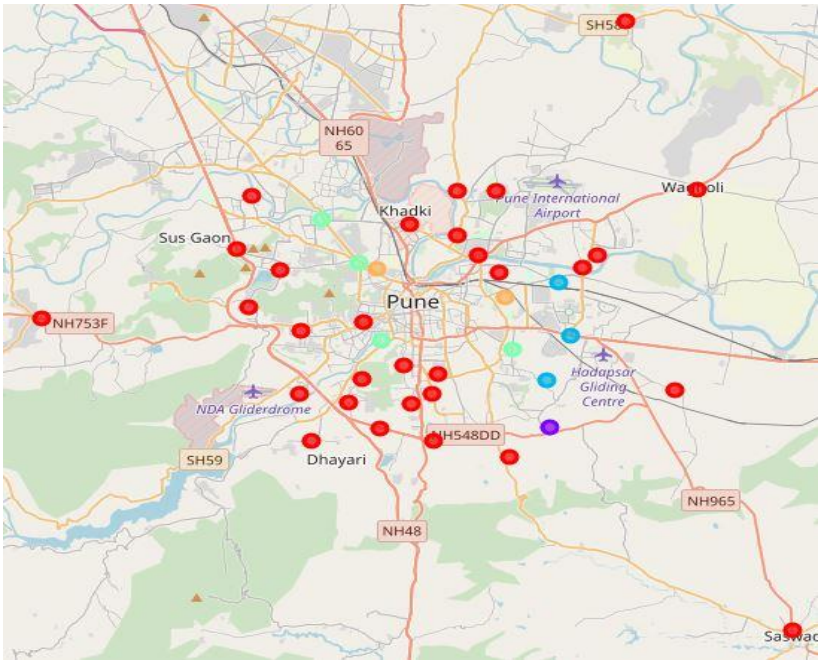

Figure 3 – Clusters of Neighbourhoods

7. Discussions – Observations

7.1 Observations: The results obtained lead to following observations:

1. Cluster label 0 represented by the red colour has the maximum number of elements as compared to other clusters

2. Cluster label 2 represented by sky blue colour has only one element

3. Cluster label 0 seems to be spread around the outskirts of Pune city

4. The number of multiplexes in Cluster label 0 is the least and in Cluster label 2 it is the maximum

7.2 Inference: This leads to the following inferences:

1. The Cluster label 0 neighbourhoods have the least amount of Multiplex venues around them.

2. The Cluster label 2 neighbourhoods have the most multiplex venues

3. Depending upon the location and proximity to nearby neighbourhoods, the following 3 spots can be considered for construction of a new Multiplex –

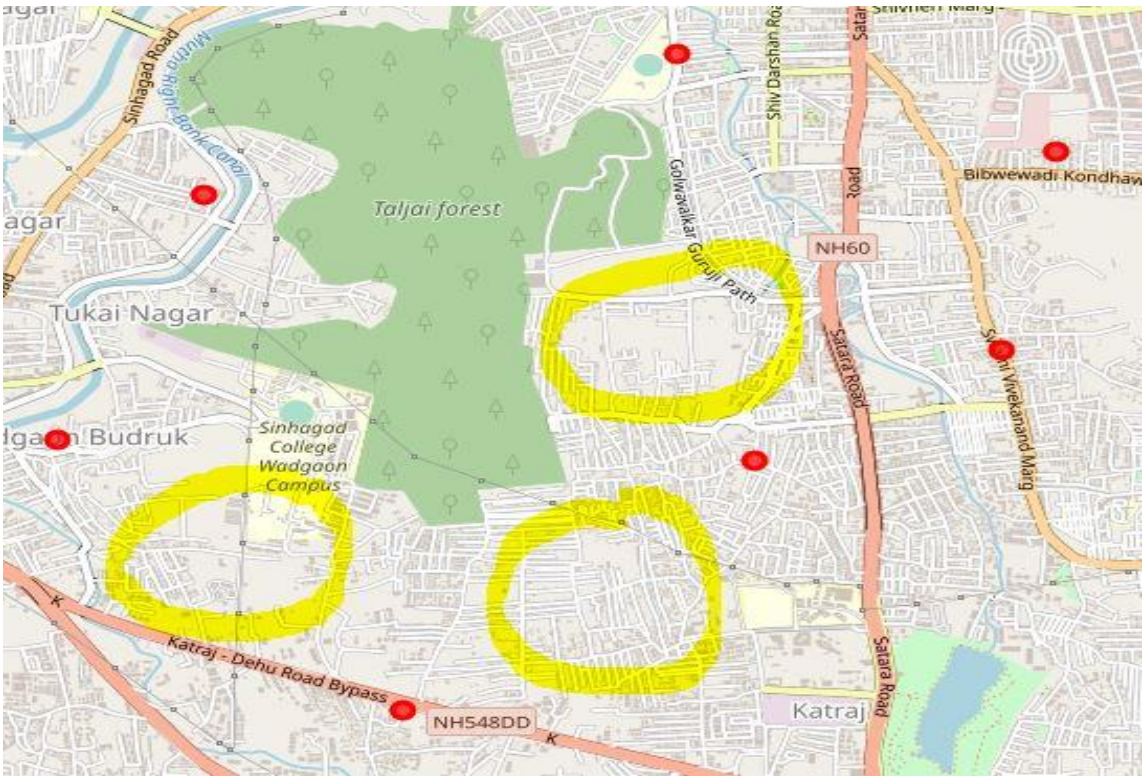

Figure 4 – Selection of Construction spots

The reasoning behind selection of these spots is that:

1. These areas are moderately to densely populated.

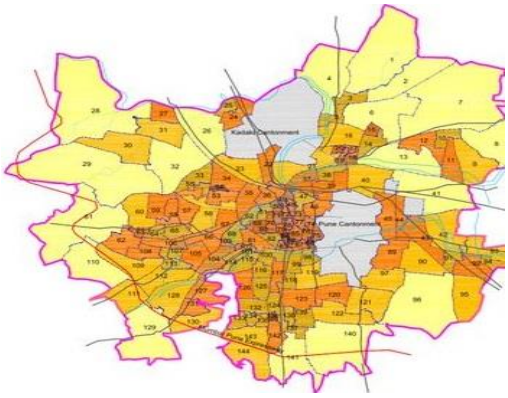

Figure 5 – Population Density of Pune [3]

2. Location is on the boundaries of 2 or more neighbourhoods

3. With less number of Multiplex in these areas, the competition would be less thus leading to more sales and profit.

8. Limitations

The following inferences can be more refined by the exact population distribution of Pune city. This will further help in discarding two of the three options provided above.

Furthermore, the information of the demographics and their spending habits does influence the decision making in this business decision. The data in hand at the moment fails to take into consideration of these criteria. Only a rough estimate can be made. But with the availability of these data elements the results can be refined and more focused answer to the business decision can be obtained.

9. Conclusion

Using web scraping technique, data was collected. It was manipulated to required format and machine learning was applied to reach the conclusion in order to answer the business question. The cluster label 0 was selected and further it was analyzed to reach to a conclusion that 3 locations are suitable for the construction of a Multiplex in Pune city.

10. References

1. Phdessay, “Multiplex Industry in India,” [Online]. Available: https://phdessay.com/multiplex-industry-in-india/
2. Business World, “The Lord of the Screens,” [Online]. Available: http://www.businessworld.in/article/The-Lord-Of-The-Screens/09-10-2018-161838/
3. NIUA.org, “Pune,” [Online]. Available: https://smartnet.niua.org/pune