

Allocation of shopping mall, Pune

✓ Anant Sakhare



IBM – APPILED DATA SCIENCE CAPSTONE

CAPSTONE PROJECT

MAY – 2020

Table of Content

Sr. No	Topic	Page
1	Introduction	3
2	Target	4
3	Data Collection	5
4	Methods	6
5	Finding	7
6	Conclusion	8
7	Limitation	9
8	Reference	10

Introduction

Pune, formerly known as Poona until 1978, is the second largest city in the Indian State of Maharashtra, after Mumbai, and the eighth most populous city in India, with an estimated population of 7.4 million as of 2020. It has been ranked as "the most liveable city in India" several times.

Along with the municipal corporation limits of Pimpri Chinchwad and the three cantonment towns of Camp, Khadki and Dehu Road, Pune forms the urban core of the eponymous Pune Metropolitan Region (PMR). According to the 2011 census, the urban area has a combined population of 5.05 million while the population of the metropolitan region is estimated at 7.4 million. Situated 560 metres (1,837 feet) above sea level on the Deccan plateau on the right bank of the Mutha river, Pune is also the administrative headquarters of its namesake district. In the 18th century, the city was the seat of the Peshwas, the prime ministers of the Maratha Empire and one of the most important political centres on the Indian subcontinent.

Pune is widely regarded as the second major "IT hub of India" and the top "automobile and manufacturing hub of India". It is known as the "Oxford of the East" with the presence of a wide range of educational institutions. India's first indigenously-run girls' school was started in Pune by Savitribai Phule along with her associate Fatima Sheikh. The city has emerged as a major educational hub in recent decades, with nearly half of the total international students in the country studying in Pune. Research institutes of information technology, education, management and training attract students and professionals from India and overseas.ⁱ

Pune has a connection with many World organisation such WTC (World trade centre). With this pune has 12 IT parks in different area of pune .The second biggest city in the state of Maharashtra with around 4.5 million people living happily in an area of 450 sq. km. I have lived in many cities over the years. However, PUNE holds a special place in my heart. No other place has given the level of thrill, fun, memories, holidays & opportunities as Pune. The city emerged drastically over the span of last 10 years and was awarded as the 2nd best city to live in India. Everyone knows that India produces the largest number of Software Experts in the world & IT Industry has become the backbone of the economy. With over 110 MNCs, Pune is home to some of the biggest and stunning IT parks of India. There are more than 5 official IT zones in Pune with Rajiv Gandhi Infotech Park in Hinjewadi being the biggest of all. This IT park is spread over an area of 13 KMs with the presence of all major companies like TCS, Infosys, Wipro, Persistent, Cognizant, Accenture, etc. Such large landscape of companies gives great exposure to Job seekers in the Silicon Valley of India

- Comfortable Lifestyle

Although Pune is next big city in Maharashtra after Mumbai, it is still pretty much cheaper when it comes to lifestyle. Property rates are too much lower in comparison to Mumbai, and this really helps the budget of a salaried person. The city is very much safe for anyone even at night. Local people of the city easily get along with outsiders. Food & Transportation is fairly cheap. One can roam around the city in an own vehicle without traffic problems. Pune has many army cantonments which keep the city safer, greener and cleaner. Law & Order in the city is also good. Altogether a family can live happily forever with an amazing lifestyle.

- Current Situation

As discussed above many new villages are added in new in Pune continent Area. New malls are under the construction phase and they are fulfilling customer demand.

So we are going to answer where should we locate a mall ?

This is the business decision we need to take.

Target

The main target of this project is that find a optimum location to build a mall.
We are going to find them a optimistic location in pune from no other mall is near.

With this our indirect target is customers or people in that area.

Data Collection

Types of data is needed

- 1 . Latitude and longitude of locations
2. Neighborhoods location
3. Lat- long point for neighborhood location point

From getting work done, we did scarped several website.

1. https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Pune
2. https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Pimpri-Chinchwad
3. <https://foursquare.com/explore?mode=url&near=Pune%2C%20Mah%C4%81r%C4%81shtra&nearGeold=72057594039187165>

- ✓ Tools/ packages are used are used to solve the problem
 - Python packages
 - GeoCoder
 - BeautifulSoup
 - Wikipedia

By using above packages we collected following data:

- ✓ Latitude
- ✓ Longitide
- ✓ Venue Name
- ✓ Neighborhood
- ✓ Venue Category

Method

For project I have analyzed locations and selected the location for creating new malls in area of Pune city. This being a Data Science Project the methodology followed as follow.

1. Understand the question (Problem)

As we know that since last decade pune city has changed into a IT hub of india. In 2010, There are 2 Million people were living but now this count raised to several billion. So malls are help to get various item in a single roof. So current allocation of malls are not fulfill requirement of people.

2. Data Collection

3. Data Validation and cleaning

4. Modeling

5. Creating Visualization

6. Business decision's

7. Data driven decision

8. Final closure

For solving the problem we did API calls on foursquare API. From this api calls we get neighborhood's and venue data. We used python loops to scroll all over the point.

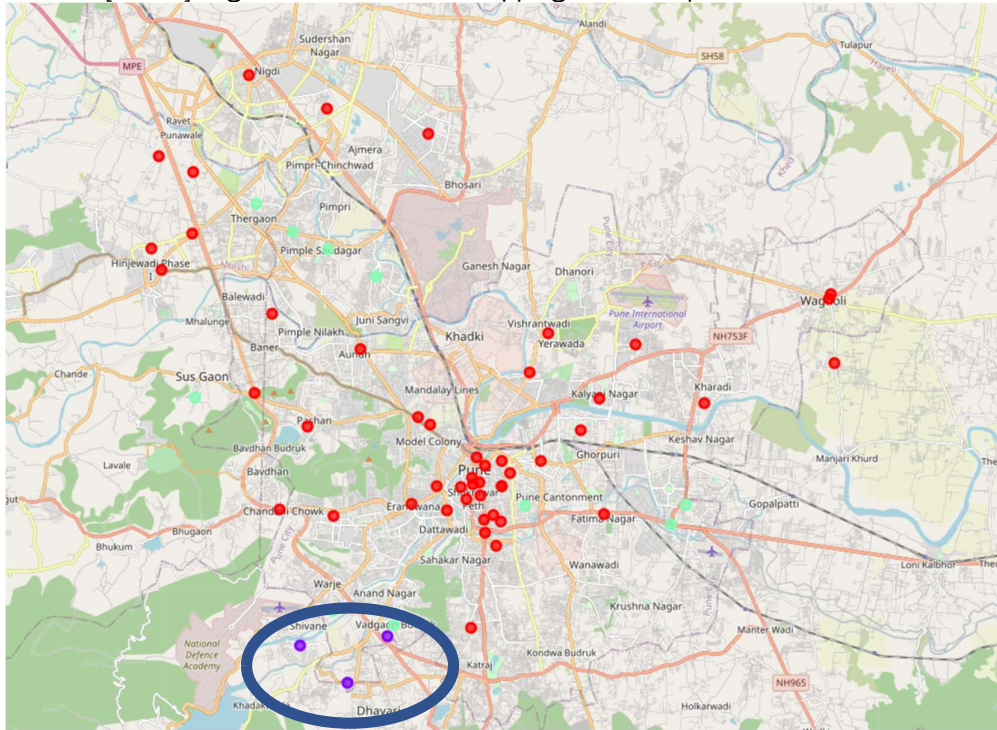
Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Shopping Mall" data, we will filter the "Shopping Mall" as venue category for the neighbourhoods.

Finally we applied K mean Clustering, to identify centroid. We chose 3 number of clusters to get data. Then, we will analyses each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analyzing the "Shopping Mall" data, we will filter the "Shopping Mall" as venue category for the neighborhoods.

Findings

The results from the k-means clustering show that we categorize the neighbourhoods into 3 Clusters are based on the frequency of occurrence for “Shopping Mall”:

- Cluster 0 [Red]: moderate number of shopping malls are present
- Cluster 1 [Purple]: low number to no existence of shopping malls are present
- Cluster 2 [Green]: high concentration of shopping malls are present



AS per data dhayari, shivane, and Vadgaon (B) is newly developed area has less number of malls are available.

Conclusion

- As we can see in the map maximum portion has the limited number of malls in nearby location.
- But in South-west region is recently added in Municipal Corporation so this locations are not having any shopping mall,
- Optimal area
 - Shivane
 - Wadgoan
 - Dhayari

This represents a great opportunity and high potential areas to open new shopping malls as there is no competition

LIMITATIONS AND FUTURE SCOPE

- data collection is not accurate means not fully updated so model is not 100% accurate.
Many malls are under construction in various area is not captured in foursquare api .
 - In future along with the geographical distributions we can also add the population attribute to the model.
-

Reference

ⁱ <https://en.wikipedia.org/wiki/Pune>