SETTING UP A SHOPPING MALL IN HYDERABAD, INDIA

Hiral

31.12.2020

OUTLINE

- Introduction
  - Business Problem
  - Target Audience
- Data
  - Sources of data
  - Methods to extract the data
- Methodology
- Results
- Discussions
  - Limitations and Suggestions for future research
- Conclusion
- References

IBM **Developer**

SKILLS NETWORK

INTRODUCTION

- For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities.

- Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provide a great distribution channel to market their products and services.

- Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand.

- As a result, there are many shopping malls in the city of Hyderabad and many more are being built. Opening shopping malls allow property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

BUSINESS PROBLEM

- The objective of this capstone project is to analyze and select the best locations in the city of Hyderabad, India to open a new shopping mall.

- Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question:

- In the city of Hyderabad, India if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

TARGET AUDIENCE OF THIS PROJECT

- This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Telangana i.e. Hyderabad.

DATA

- To solve the problem, we will need the following data:

- List of neighborhood's in Hyderabad. This defines the scope of this project which is confined to the city of Hyderabad, the capital city of Telangana which is in South India

- Latitude and longitude coordinates of those neighborhood's. This is required in order to plot the map and also to get the venue data.

- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhood's.

SOURCES OF DATA AND METHODS TO EXTRACT THE DATA

- This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India) contains a list of neighborhood's in Hyderabad, with a total of 200 neighborhood's.

- We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup packages. Then we will get the geographical coordinates of the neighborhood's using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhood's.

- After that, we will use Foursquare API to get the venue data for those neighborhood's.

- Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward.

- This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

HYDERABAD MAP
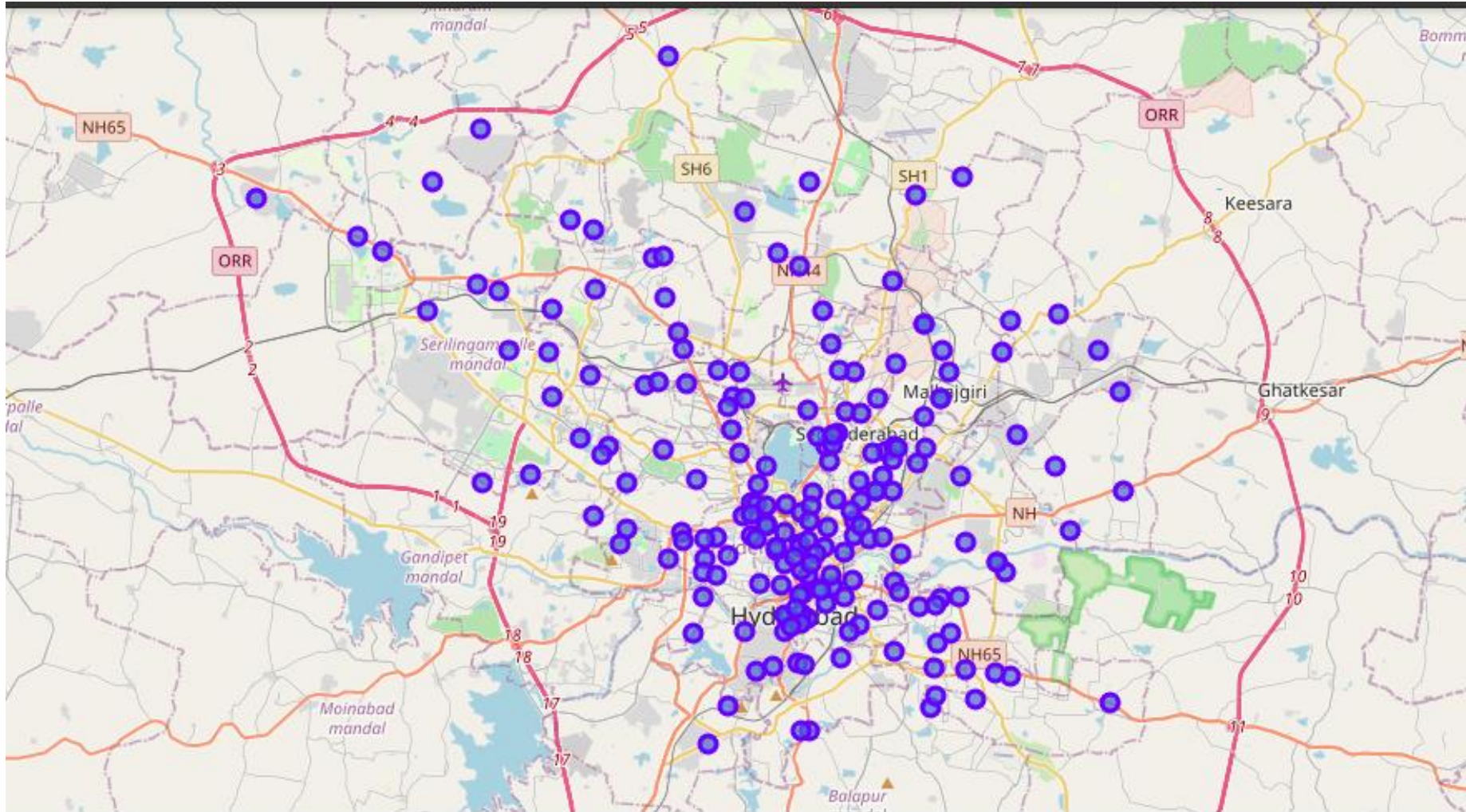
- This map represents all the 200 neighborhood's of the city Hyderabad. We then try to make clusters by using the Machine learning technique(K-means clustering) and group all the neighborhood's into different clusters.

METHODOLOGY

- Firstly, we need to get the list of neighborhood's in the city of Hyderabad. Fortunately, the list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India) .

- We will do web scraping using Python requests and beautiful soup packages to extract the list of neighborhood's data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API.

- To do so, we will use the wonderful Geocoder package that will allow us to convert the address into geographical coordinates in the form of latitude and longitude.

- After gathering the data, we will populate the data into a pandas Data Frame and then visualize the neighborhood's in a map using Folium package.

- This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Hyderabad.

- Next, we will use the Foursquare API to get the top 100 venues that are within a radius of 2000 meters.

IBM Developer

SKILLS NETWORK

METHODOLOGY

- We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhood's in a Python loop.

- Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude.

- With the data, we can check how many venues were returned for each neighborhoods and examine how many unique categories can be curated from all the returned venues .

- Then, we will analyze each neighborhood by grouping the rows by neighborhoods 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 allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

- It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project.

METHODOLOGY

- We will cluster the neighborhood's into 3 clusters based on their frequency of occurrence for "Shopping Mall".

- The results will allow us to identify which neighborhood's have a higher concentration of shopping malls while which neighborhood's have a fewer number of shopping malls.

- Based on the occurrence of shopping malls in different neighborhood's, it will help us to answer the question as to which neighborhood's are most suitable to open new shopping malls.

- Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhood's in cluster 2 with little to no competition.

IBM Developer

SKILLS NETWORK

RESULTS

- The results from the k-means clustering show that we can categorize the neighborhood's into 3 clusters based on the frequency of occurrence for "Shopping Mall":

- Cluster 0: Neighborhood's with a very less number of shopping malls

- Cluster 1: Neighborhood's with a moderate concentration of shopping malls

- Cluster 2: Neighborhood's with a high concentration of shopping malls

- The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in mint green color.

RESULTS
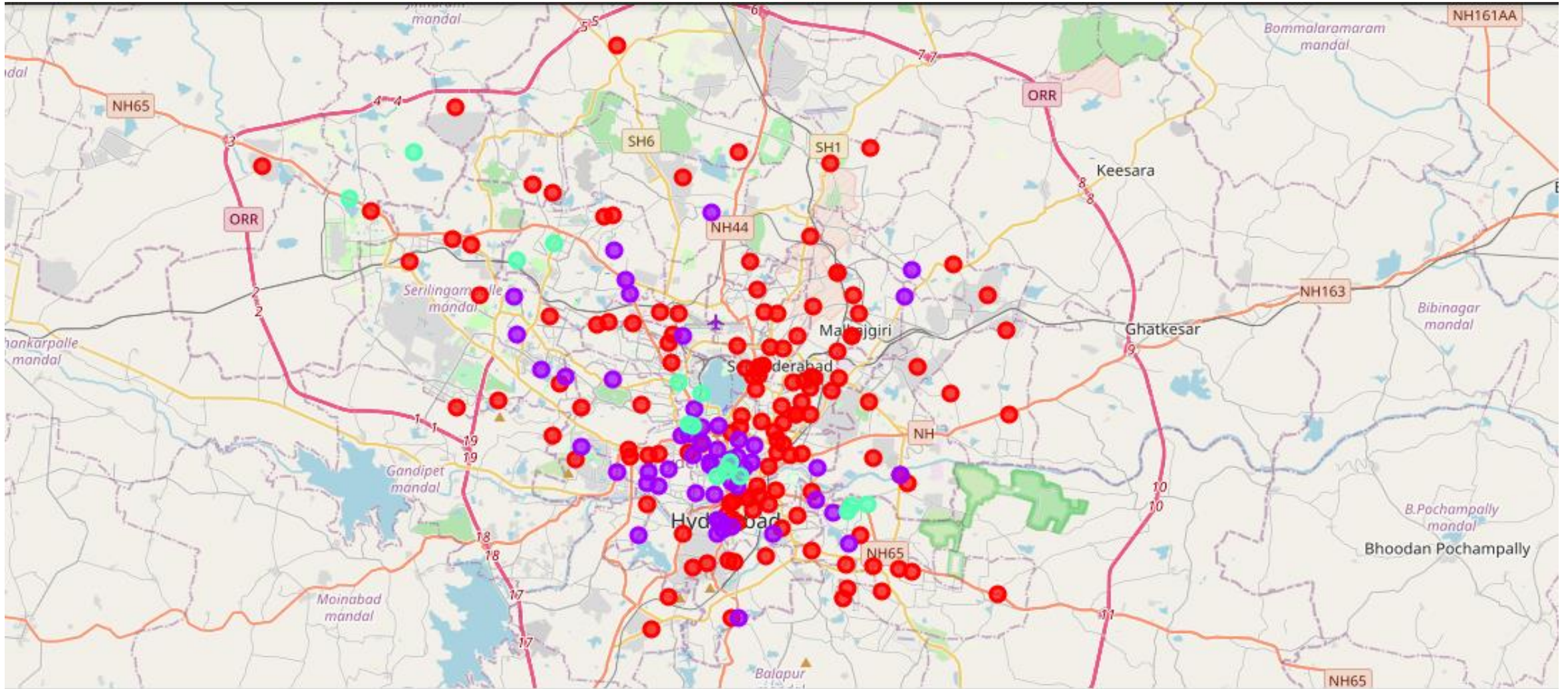
DISCUSSION

- As observations noted from the map in the Results section, most of the shopping malls are concentrated in the central area of Hyderabad city, with the highest number in cluster 0 and moderate number in cluster 1.

- On the other hand, cluster 2 has a very low number(15) in the neighborhood's. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls.

- Meanwhile, shopping malls in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of shopping malls.

- From another perspective, the results also show that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls.

- Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhood's in cluster 0 with little to no competition.

- Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhood's in cluster 1 with moderate competition.

- Lastly, property developers are advised to avoid neighborhood's in cluster 2 which already have a high concentration of shopping malls and suffering from intense competition.

LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

- In this project, we only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall.

- However, to the best knowledge of this researcher, such data are not available to the neighborhood level required by this project.

- Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall.

- In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned.

- Future research could make use of paid account to bypass these limitations and obtain more results.

CONCLUSION

- In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new shopping mall.

- To answer the business question that was raised in the introduction section, the answer proposed by this project is:

- The neighborhood's in cluster 1 are the most preferred locations to open a new shopping mall.

- The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.

IBM Developer

SKILLS NETWORK

REFERENCES

- Category: Neighborhood's in Hyderabad, India.
- Wikipedia Retrieved from :
- https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India
- Foursquare Developers Documentation.
- Foursquare. Retrieved from :
- https://developer.foursquare.com/docs/

IBM Developer

SKILLS NETWORK