IBM Capstone Project

The Battle of Neighborhoods

Business Consultancy for Property Developers to Open a Shopping Mall

By

Saad Muhammad

Introduction

Karachi is the biggest city of Pakistan with an area of 3780 square kilometers and the capital of the Province; Sindh. The population of Karachi is roughly 23 Million and is growing is a fast pace. The shoppers in Karachi are lively and go to different places around the town to have fun with friends and family. One of the places that shoppers in Karachi like the most is the shopping malls, one the examples in the Dolmen Mall located in the Clifton area which is always crowded with shoppers. Shopping malls in Karachi have recently become a very popular shopping destination for general population because of the advantages of having all the necessary shopping stuff under one roof, from grocery to clothes, from cosmetics to electronics and a proper food court with numerous eateries.

Business Problem/ Target Audience

For a city as big as Karachi there are not as much shopping malls in it and there is a huge room to property developers to invest in this area and exploit the huge potential of Karachi in terms of its population and importance as business hub of Pakistan. To open a shopping mall would require resources and a thorough plan. The first and foremost thing for the decision to open a shopping mall in Karachi would be the potential location. I will help the interested property developers in identifying the best strategic location for the mall using data science techniques which include ML models such as K-Means to cluster different neighborhoods with respect to concentration of shopping malls in them. The target audience for this project would be the property developers.

Data

To do our project as described, we would need the following data:

- List of all the neighborhoods in Karachi
- Geocoordinates of against all the neighborhoods
- And finally the venues in the neighborhoods

Tools/ Sources

- The list of all the areas in Karachi is available at Wikipedia, I have copied it into a CSV file and renamed a few areas manually after checking if the geocoder package gives their coordinates or not.

- Python Geocoder package is used to get the coordinates for most of the neighborhoods.
- Finally, FourSquare API is used to get details about the neighborhoods including list of venues.

https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Karachi

Methodology

The first and foremost thing was to retrieve the data for the neighborhoods in Karachi which was available on the above-mentioned address on Wikipedia. The data was copied to csv file and manually altered to rename some of the neighborhoods based on the acceptability of the names by the geocoder.

The coordinates were then retrieved and plotted against each neighborhood using geocoder package from Python. We used folium to visualize the neighborhoods on Karachi map to make sure the coordinates against the neighborhoods were correct.

Foursquare API was used to get top 10 venues within 2000 meters radius. The API call returned 1326 venues from Foursquare database against our neighborhood list. There are 117 unique venue categories but we had to choose only the shopping malls from the venue grouped data frame. The number of shopping malls in Karachi were found out to be only 26 which is a very small number considering the size and the population of the city which shows that there is a huge opportunity for property developers to build shopping malls in the city.

Finally, we used an ML model to cluster the malls. The ML model we used is K-Means which identifies centroids and allocates the data to the nearest cluster based on Euclidean distance. K-Means is an unsupervised ML algorithm and suits our problem well to cluster the shopping malls in the city of Karachi. We have used 3 clusters in our project. The results will allow us to make our recommendation based on the concentration of malls in different cluster and visualized on the map for easy understanding. Based on the concentration of shopping malls we can easily decide which neighborhood to choose to develop a new mall.

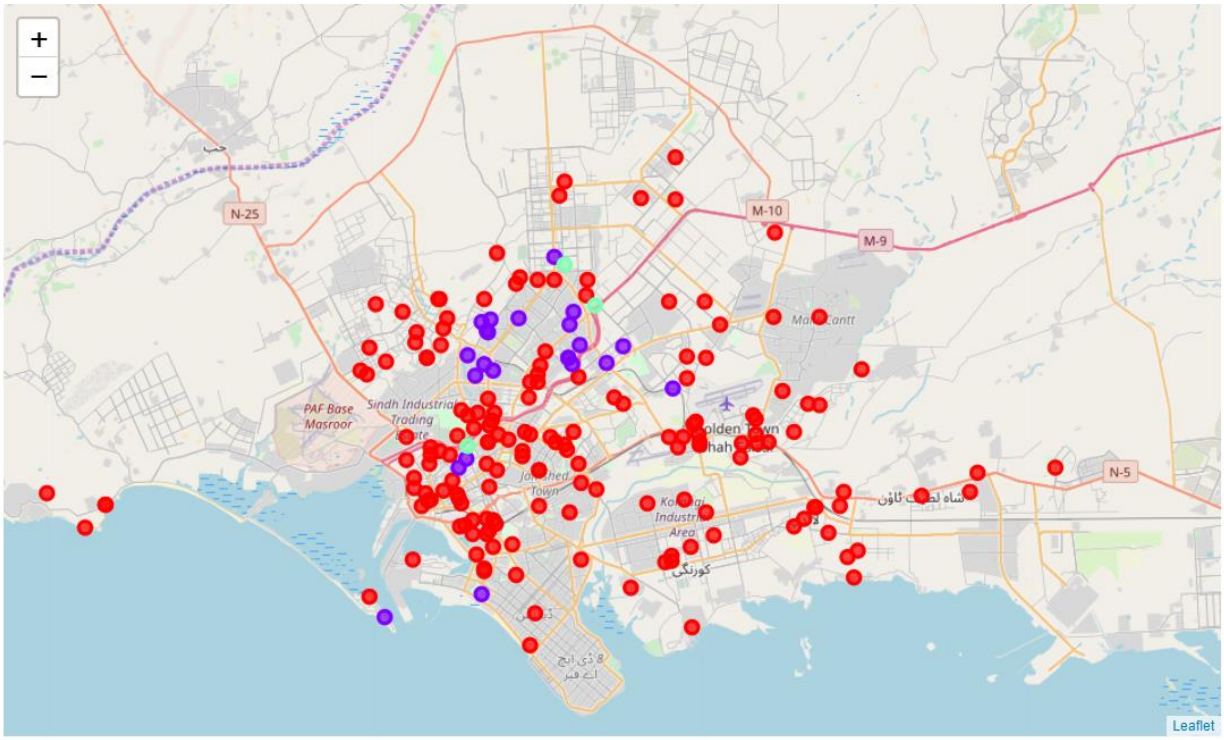
Results

We have used K-Means to cluster the shopping malls in the city

• Cluster 0: Neighborhoods with no shopping malls and is represented with red circles on the map

• Cluster 1: Neighborhoods with high concentration of shopping malls and is represented with purple circles on the map

• Cluster 2: Neighborhoods with low concentration of shopping malls and is represented with mint circles on the map



Discussion & Conclusion

Most of the shopping malls in the Karachi are in cluster 1 in the central Karachi or nearby neighborhoods as represented by the purple circles the map. There would be high competition in this area but most likely a greater customer footfall as well. There are only 3 malls in cluster 2 as represented by mint circles. There would be very little competition and good opportunity for interested developers to develop a shopping mall. Finally, the majority of neighborhoods fall in cluster 0 with no shopping malls as shown by the red circles, there would be little to no competition on these neighborhoods but it might be risky as well as no developer stepped into these neighborhoods it may be because of the wealth profile of these neighborhoods which may be clarified with more data and further analysis on the wealth profile of these neighborhoods. Cluster 1 is composed of middle to high income neighborhoods, considering these neighborhoods may be a little risky but it would surely be rewarding at the same time because of greater foot fall and awareness in the neighborhood. It would suit risk taking property developers to open Malls in cluster1. For the risk averse developers, cluster2 and its surrounding neighborhoods would be a good choice to open shopping malls.

The conclusion of this report is recommendation of cluster1 and cluster2 for the interested property developers based on their risk profile to develop shopping malls in these clusters. While there is little information available on cluster0 about the wealth profile of the neighborhoods, there may be slums and low income neighborhoods in this cluster and therefore not recommended for opening a shopping mall.