IBM Data Science Capstone Project Report

Introduction:

Pune city, which was previously known as an education hub is now also a growing IT hub. This attracts a lot of student and working-class emigrants. Naturally the go-to food of this demographic is fast food or snacks due to the fast-paced environment. This makes Pune a good market for fast food chains and snack bars.

Problem:

In this project we will analyze the neighborhoods in Pune to help a fast food chain, looking to enter Pune, formulate a strategy. We will be targeting mainly fast food restaurants and snack bars since they usually cater to a similar target audience.

This will help the fast food chain by giving them an idea about which locations are dense in population belonging to this demographic. This will ensure a high foot-traffic and help boost visibility and sales.

Data:

The data required for this project is:

1. List of neighborhoods in Pune
2. Geographical coordinates of neighborhoods
3. List of restaurants and snack bars in the neighborhoods

The following sources of Data will be used.

1. Wikipedia- To get the data regarding neighborhoods in Pune.
2. Python geolocation library: To get the coordinates of all the locations.
3. Foursquare API: To explore the neighborhoods and get a list of restaurants, snack bars and fast food outlets.

Methodology:

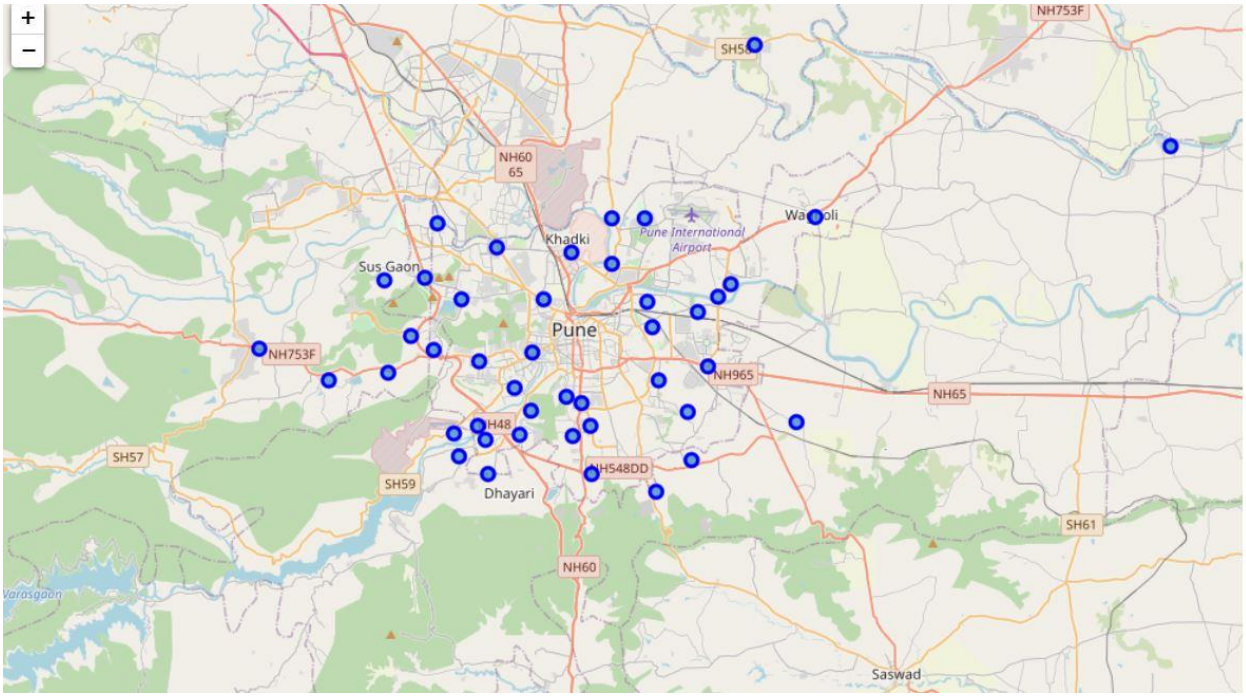
We begin by using the BeautifulSoup package to perform web scraping on the Wikipedia page to get the list of neighborhoods in Pune city. Next we use the geocoder package to get the coordinates of all the neighborhoods. We store these coordinates in a new dataframe.

Now we use the Foursquare API to explore all these neighborhoods to get the venue information. From this dataset we filter out all the venues to get a new dataframe which contains only fast food restaurants and snack bars since these are the only 2 venues we are interested in.

We perform a little exploratory data analysis on the new dataframe. Next we divide the areas into 3 clusters using the K Means algorithm. We now visualize these clusters using the folium library in Python. Based on these findings recommendations can be made to the fast food chain.

Results:

After web scraping and getting the list of neighborhoods in Pune City. We use the folium library to plot it on a map.

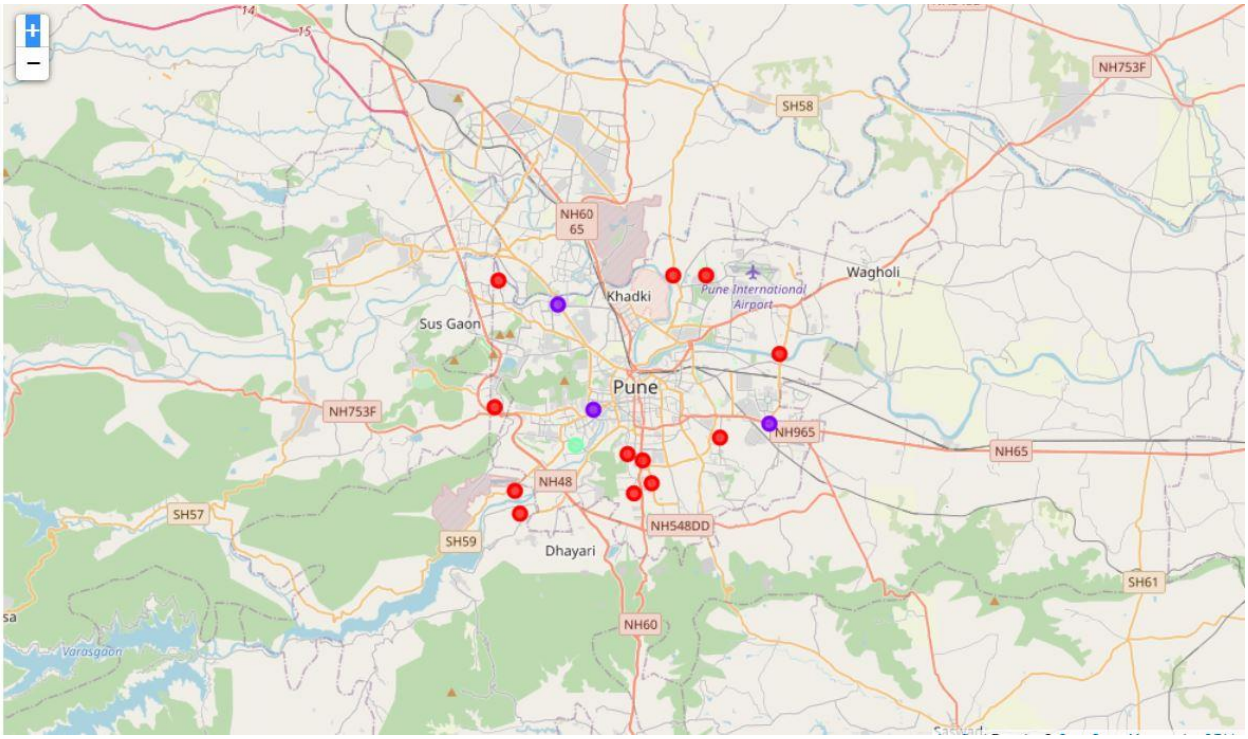


After this we used the K Means algorithm to divide the areas into three clusters.

Cluster0: The locations having only one fast food restaurant/snack bar.

Cluster1: The locations having three fast food restaurants/snack bars.

Cluster2: The locations having two fast food restaurants/snack bars.



Discussion:

From the observations it seems that the locations falling in the first cluster will be good locations to get started since the new outlets will face little or negligible resistance from competitors. Outlets opening in cluster 2 will face the highest amount of resistance from the already existing restaurants and snack bars since they will be closer to each other and will compete for the same customers.

Based on these findings it is recommended that the new outlets should first be opened in locations falling in Cluster 0. Depending on how they perform in Cluster 0 new investments must be made to open outlets in cluster 2 and then cluster 1.

Conclusion:

As mentioned previously locations in cluster 1 are the ideal locations to get started since they have only one fast food restaurant/snack bar. Once these outlets are able to capture some amount of market share they can start opening outlets in cluster 3 and then cluster 2 finally. Using the findings from this project the investors and the key decision makers in the fast food chain will be able to formulate a decent strategy to enter a brand new market and also optimize the spending.