Opening a new supermarket in Pune, Maharashtra

IBM data science capstone project

By - Vivek Mankar

Business Problem

- The location of a supermarket plays an important role in its success.
- Especially considering the number of other supermarkets, grocery stores, Markets, etc in the neighborhood.
- The objective of the project is to analyze and select the best location in the city to open a new supermarket. Using data science and machine learning techniques like clustering, this project aims to solve the following business Question:
- **Que: In the city of Pune if a company wants to open a new supermarket where would you recommend that they open it?**

Data

To solve this problem we will need the following data :

- List of neighborhoods in Pune.
- Latitude and Longitudinal coordinates of those neighborhoods.
- Venue data for each neighborhood.

Sources

- For the list of neighborhoods, I used (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Pune)
- For Latitude and Longitudinal coordinates: Python Geocoder Package (https://geocoder.readthedocs.io/)
- For Venue data: Foursquare API (https://foursquare.com/)

Methodology

- Web scraping Wikipedia page for neighborhoods list.

- Get latitude and longitude coordinates using geocoder.

- Use Foursquare API to get venue data.

- Group data by neighbourhood and taking the mean of the frequency of occurrence of each venue category.

- Decide the value of k by using elbow method.

- Perform clustering on the data by using k-means clustering.

- Visualise the clusters in the map using folium.

- Visualise the clusters using 3D plot of matplotlib.

K-Means Clustering Results

The Elbow Method using Distortion
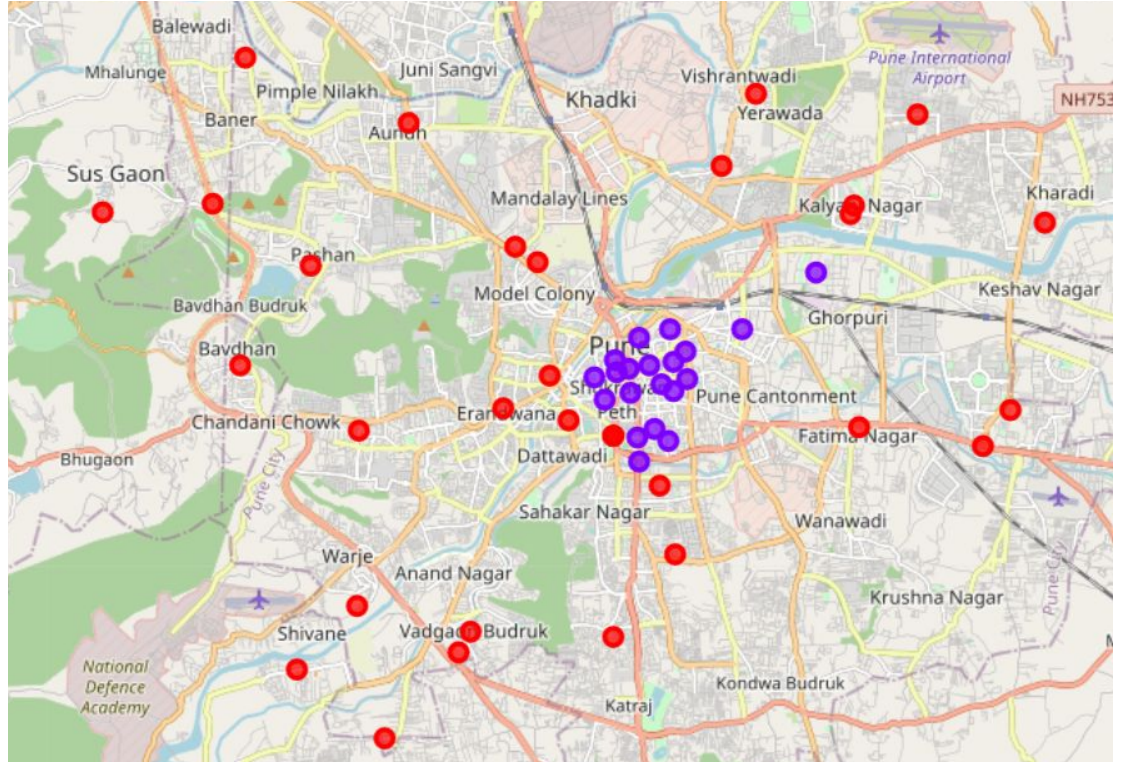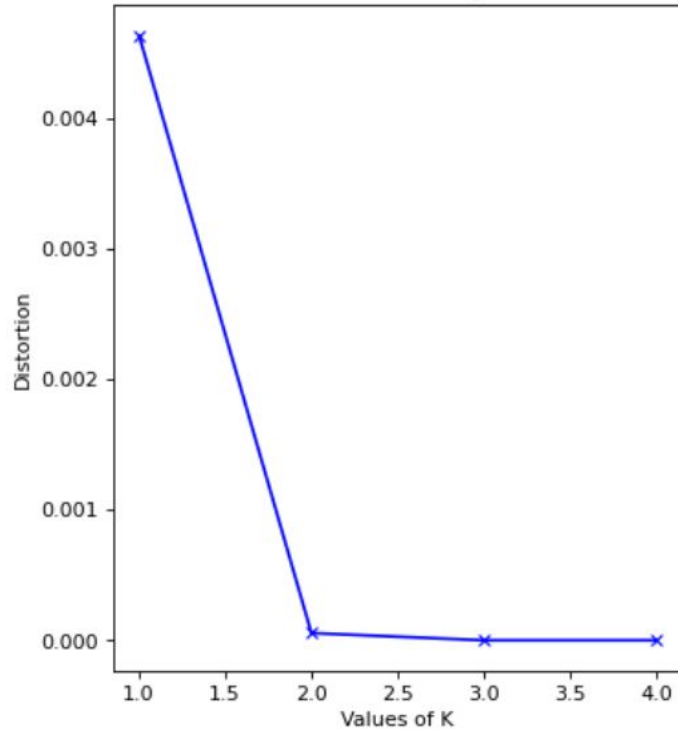
3D Plot of clusters
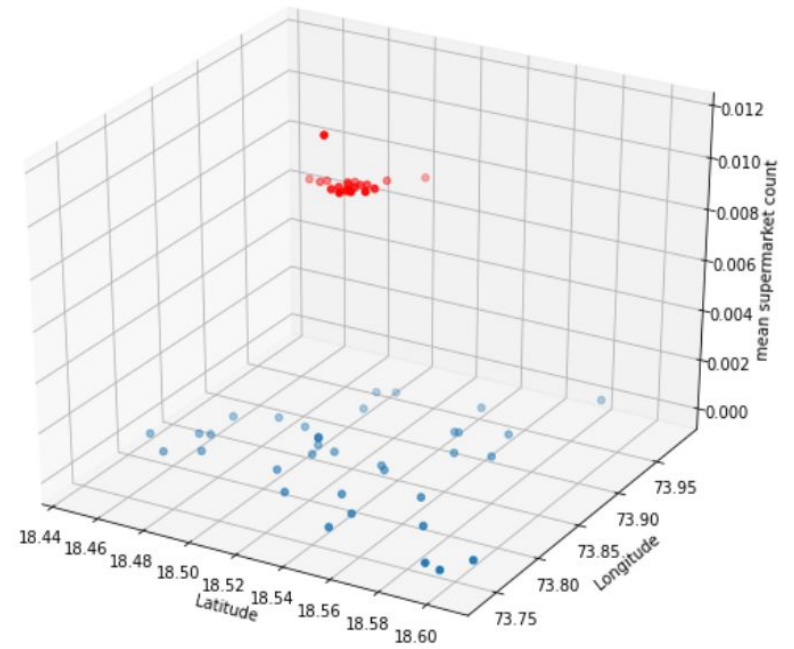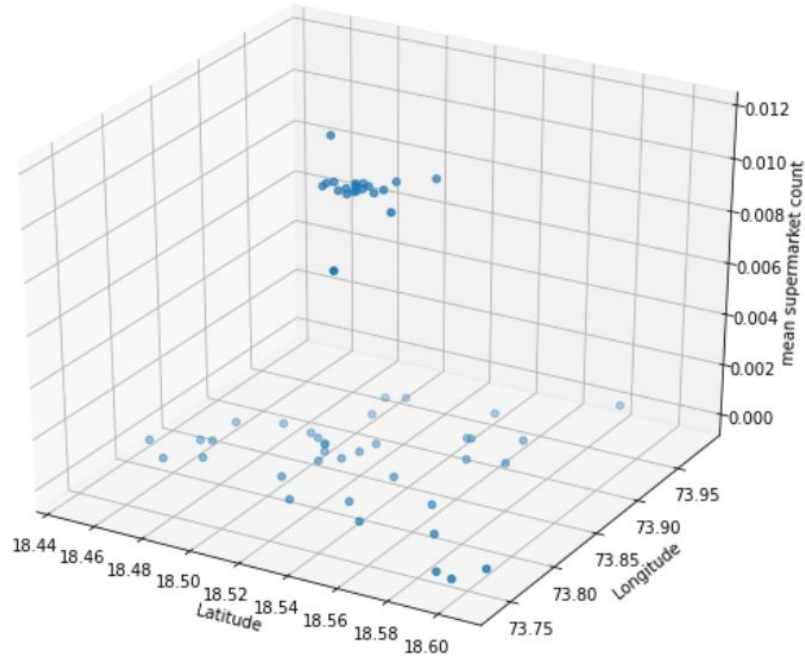
Discussion

Most of the supermarkets are concentrated in the central area of the city

From the result it is clear that there are only two kinds of neighborhoods i.e. neighborhoods with supermarkets and without supermarkets.

Oversupply of supermarkets mostly happened in the central area of the city while the suburb area still have very few or no supermarkets.

Recommendations

Open new supermarkets in neighborhoods in cluster zero with little or no competition

Avoid neighbourhoods in cluster one already high concentration of Supermarkets and intense competition

The number of neighborhoods in cluster 0 is greater than the number of neighborhoods in cluster 1 this means that there are more opportunities in Pune for such stores.

Conclusions

The neighborhoods in the cluster zero are most preferred locations to open a new supermarket.

The findings of this project will help the relevant stakeholders to capitalise on the opportunities on high potential locations for avoiding crowded areas in their discussions to open New supermarket

Thank you !...