Capstone Project

THE BATTLE OF NEIGHBOURHOODS

By: Satyam Dutta

May 27th, 2020

Business Problem

- Problem: ‘What should be the recommended place to open a new shopping mall in a developed city like Surat?’

- Objective: To analyse and select the favourable locations for constructing or investing in a new shopping mall.

Data

Data Required:

- List of neighbourhoods in Surat
- GPS coordinates of the neighbourhoods
- Venue Data

Sources of Data:

- Wikipedia page 'https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Surat'
- Python Geocoder Package
- Foursquare API

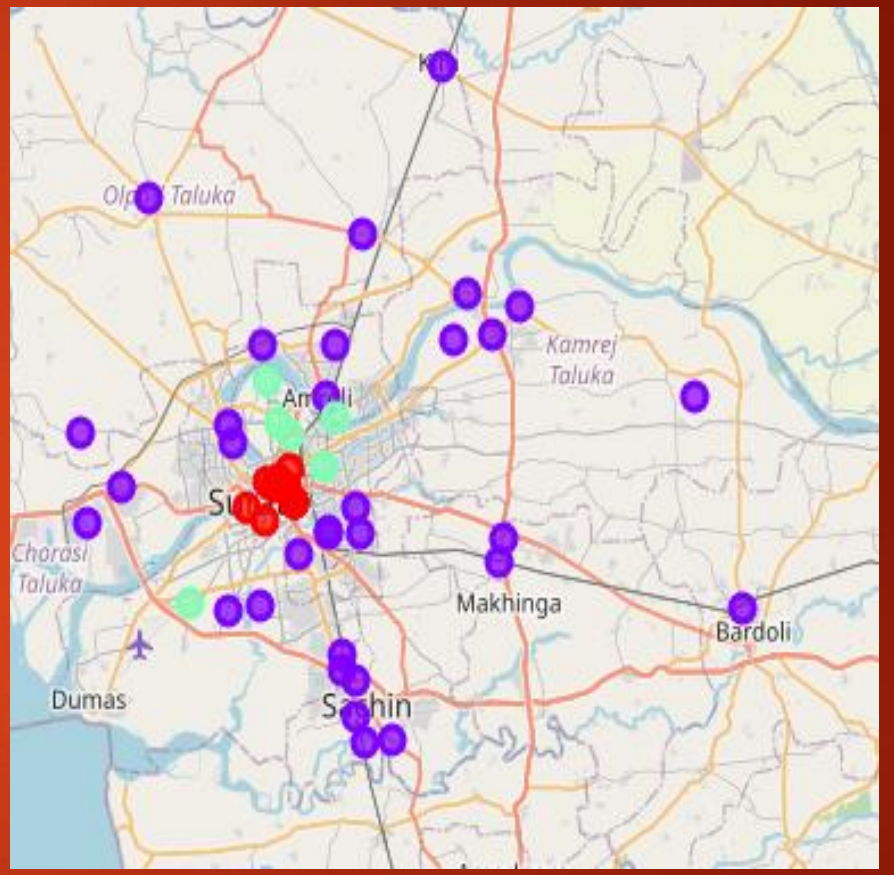
Methodology

- Web scraping the Wikipedia page to obtain neighbours list.
- Obtaining GPS coordinates of the neighbourhoods using Geocoder package.
- Using Foursquare API to obtain venue data.
- Group data by taking the mean of frequency of the occurrence of each venue category.
- Filter the grouped data by ‘Shopping Mall’.
- Performing K-means cluster on the filtered data.
- Visualize the clusters on map using Folium package.

Results

Neighbourhoods categorized into 3 clusters:

- Cluster 0: Neighbourhoods with moderate number of shopping malls.
- Cluster 1: Neighbourhoods with low number of shopping malls.
- Cluster 2: Neighbourhoods with large number of shopping malls.



Discussions

- Most of the shopping malls are concentrated in the central area of Surat.
- The highest number of malls is in cluster 2.
- Moderate number of shopping malls in cluster 0.
- Cluster 1 has very low number to totally no shopping malls in the neighbourhoods.

Conclusion

- Answer to the business problem:
  'The neighbourhoods in the Cluster 1 are the most preferred location for constructing or investing in new shopping malls.'

- Findings of this project will help the concerned stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded locations to open a new shopping mall.


Thank
You