Coursera Capstone

IBM Applied Data Science Capstone

Opening of a new Indian food restaurant in Pune city

Heramb Lonkar

November 2019

1. Introduction

“People who love to eat are always the best people”. [1]

Restaurants are where we always go to eat. These are the places where we eat peacefully. Try new dishes, cuisines. In India, people love to eat Indian food. Why? Because India is a diversified country. Each state has different dishes, different tastes. Hence you will find Indian restaurant all over the places. As expected, Pune city has so many Indian food restaurants. Hence to open a new restaurant, an analysis must be done.

The objective of this project is to analyse and get the best suitable location to open an Indian food restaurant in Pune city. For this analysis machine learning technique like clustering.

Problem Statement- “Best suitable location to open Indian food restaurant in Pune city.”

So, **the target audience** for this project is a businessman who is willing to invest in any location regardless of the property cost. Or for someone who is trying to extend his Indian food restaurant chain in Pune city. For small investors, this may not be a good idea.

2. Data and Methodology

Data from the following sources will be used:

- List of neighbourhoods from *Wikipedia*[2]
- Coordinates (Geocoder Python) using *OpenStreetMaps*[3]. But I have decided to shift to *HERE*[4] maps API for better accuracy.
- Restaurant venues using *Foursquare API*[5]

Now, the usage of this data will be explained:

i. From Wikipedia page, we will get the list of neighbourhoods. For this process, *BeautifulSoup4*[6] Python library will be used. By this, neighbourhoods will be added in the list and later into a data frame.

ii. Geocoder in Python will be used to get coordinates of the neighbourhoods. Coordinates will be joined with neighbourhoods into the data frame.

iii. Now using Foursquare API, we will get a list of venues and their category.

iv. Maximum 300 venues in 4000 meters radius will be fetched.

v. Data cleaning will be done at all stages whenever required.

vi. Also, data visualisation will be done using Folium maps, a Python library.

vii. Machine Learning technique K-means clustering will be used to get results.

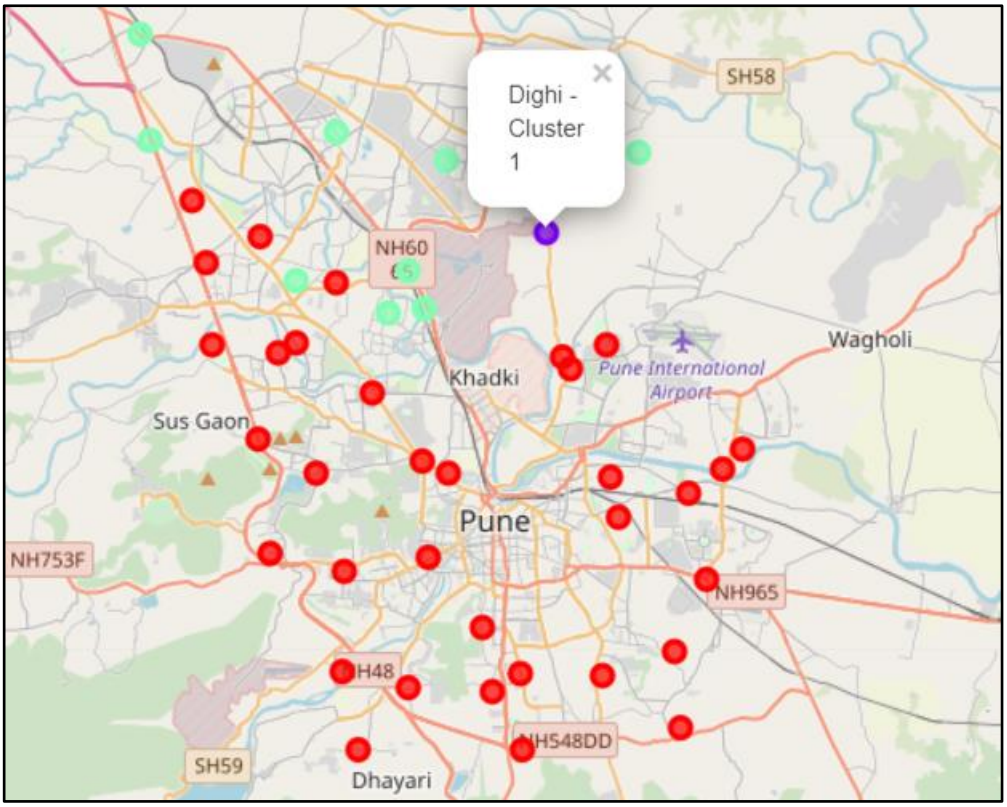
viii. Depending on the frequency occurrence of Indian Restaurants, neighbourhoods will be clustered.

ix. K-means will help us to understand the concertation of Indian Restaurants in neighbourhoods.

3.Results:

As you can see in the image below, the result is categorised into 3 clusters. These clusters are based on the frequency of occurrence for ‘Indian Restaurants’

- Cluster 0: Neighbourhoods with a moderate number of shopping malls in red colour
- Cluster 1: Neighbourhoods with a low number to no existence of shopping malls in purple colour
- Cluster 2: Neighbourhoods with a high concentration of shopping malls in light green colour



4. Discussion

As seen in results, Pune city has so many Indian food restaurants. Though the prediction pattern is very different, restaurants are located all over the places.

A moderate number of restaurants can be found all over the Pune city i.e. Cluster 0. A very high number of restaurants are in little outskirts. Probably, there was pretty much free or open space previously i.e. Cluster 2.

Hardly any restaurant can be seen in Cluster 1. So if anyone wants to open a new restaurant Cluster 1 is the best suitable location.

Also, those who are trying to extend their food chain restaurant, they can also invest in Cluster 0. But investor/businessman is highly suggested to avoid Cluster 2 since they will have intense computation already.

5. Conclusion

In this project, I 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.

Limitation of this project is that I have not constraint the money need to invest. If we consider that then scenario could be different.

Ultimately suggestion is given to open a new Indian food restaurant in Pune city. I conclude that the following areas as shown in the image are the best areas to open a new Indian food restaurant. Hence Dighi is the only recommended location for new businessman.

	Neighborhood	Indian Restaurant	Cluster Labels	Latitude	Longitude
13	Dighi	0.8	1	18.61522	73.87241

1. References:

[1] The All-Time Greatest Quotes About Food And Eating: https://www.delish.com/food/g25438962/food-quotes/

[2] List of neighbourhoods in Pune: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Pune

[3] OpenStreetMap API: https://wiki.openstreetmap.org/wiki/API

[4] HERE Maps API: https://developer.here.com/

[5] Foursquare Developers: https://developer.foursquare.com/

[6] BeautifulSoup4: https://pypi.org/project/beautifulsoup4/