

# COURSERA CAPSTONE

## IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

SETTING UP A CAFFE IN AHMEDABAD, INDIA

BY: Nandan Kakadiya

August 13, 2021



## INTRODUCTION

The city of Ahmedabad is endowed with a rich architectural heritage that is vital to the local identity and continuity of the place. Along with the foremost heritage Indo-Islamic monuments of the 15th to 17th centuries, there are potential heritage precincts in the form of the Pols, the traditional residential clusters of the medieval period, which makes Ahmedabad exceptional. Combining these all, the historic walled city of Ahmedabad has it all to be the first and only city in India to be inscribed in UNESCO's World Heritage City list of 2017. As a result, Ahmedabad city is hotspot for foreigners as well as local tourists. Opening cafe allow owner to earn good revenue. Opening a new Cafe, like any other business decision, needs careful analysis and is far more complicated than it appears. As with any business, the cafe's location is one of the most critical considerations that will determine whether the business succeeds or fails.

## BUSINESS PROBLEM

There are already many Cafes in the city and many more are being built. The objective of this project is to analyze and select the best locations in the city of Ahmedabad, India, to open a new Cafe. This project is mainly focused on geospatial analysis of the Ahmedabad City to understand which would be the best place to open a new Cafe. Using data science methodology and machine learning Algorithms like clustering, this project aims to provide solutions to answer the business question: In the city of Ahmedabad, if someone is looking to open a cafe, where would you recommend that they open it?

## TARGET AUDIENCE

Food franchises owner and investors who are planning to open or invest in new restaurants or Caffe in Gujarat's Economic capital city, Ahmedabad, would find this initiative particularly valuable.

## DATA

We'll need the following data to solve the problem:

- List of neighborhoods in Ahmedabad. This defines the scope of this project which is confined to the city of Ahmedabad.
- In order to plot the neighborhood we will need latitude and longitude of the neighborhood.
- Venue data, particularly data related to Cafe. We will use this data to perform clustering on the neighborhoods.

- ### SOURCES

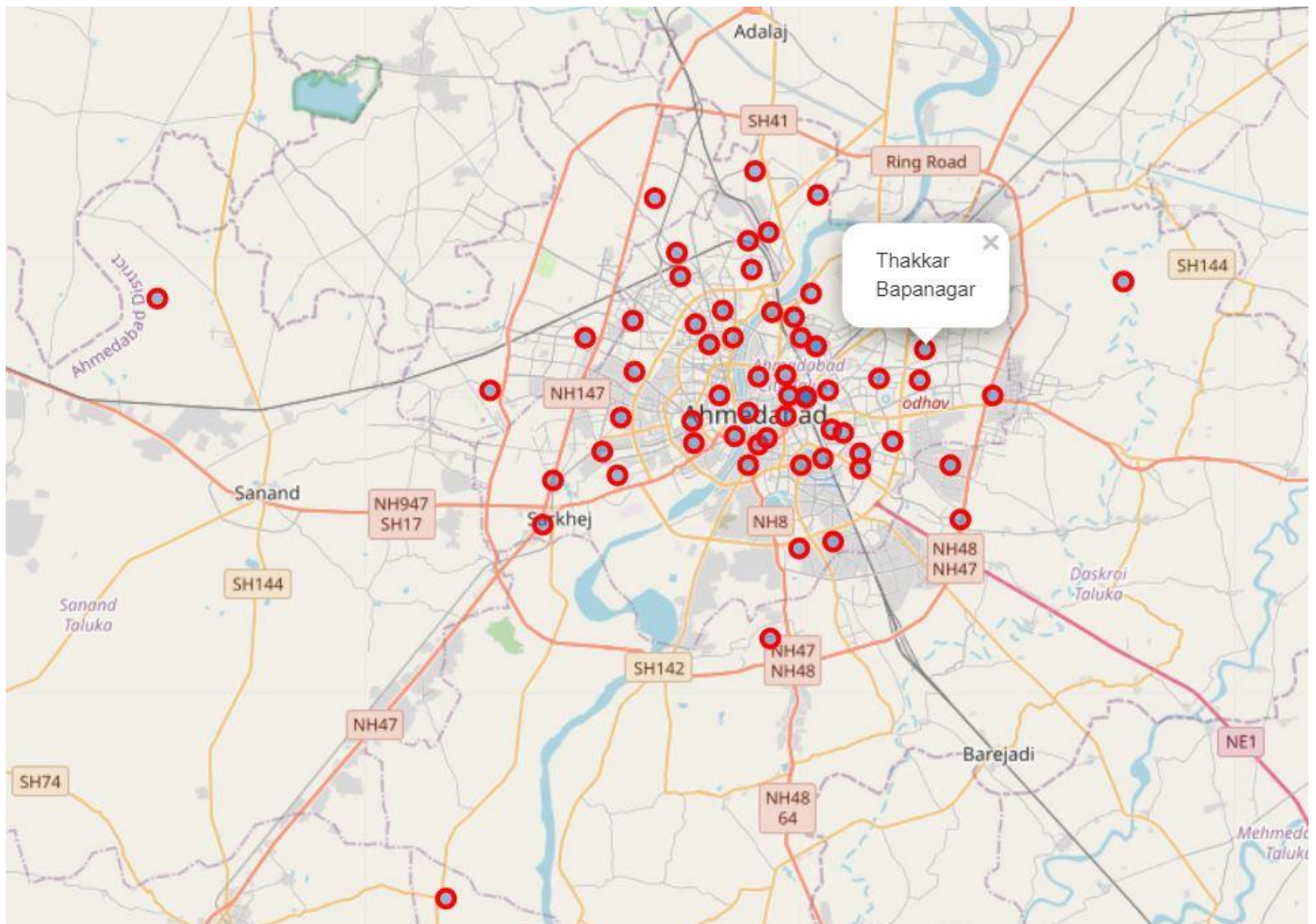
This page ([https://en.wikipedia.org/wiki/Category:Neighbourhoods\\_in\\_Ahmedabad](https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Ahmedabad)) is a list of neighbourhoods in Ahmedabad. I have extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup library.

Then, using the Python Geocoder library, we can get the latitude and longitude coordinates of the neighbourhoods. After that, to access the venue data for those neighbourhoods, I used the Foursquare API.

Foursquare API will provide many categories of the venue data, and we are particularly interested in the Cafe category in order to help us solve the business problem. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium)

## AHMEDABAD MAP

This map represents all the 69 neighbourhoods of the city Hyderabad. We then try to make clusters by using the Machine learning technique(K-means clustering) and group all the neighbourhoods into different clusters.





## METHODOLOGY

Firstly, we need to get the list of neighborhoods in the city of Ahmedabad.

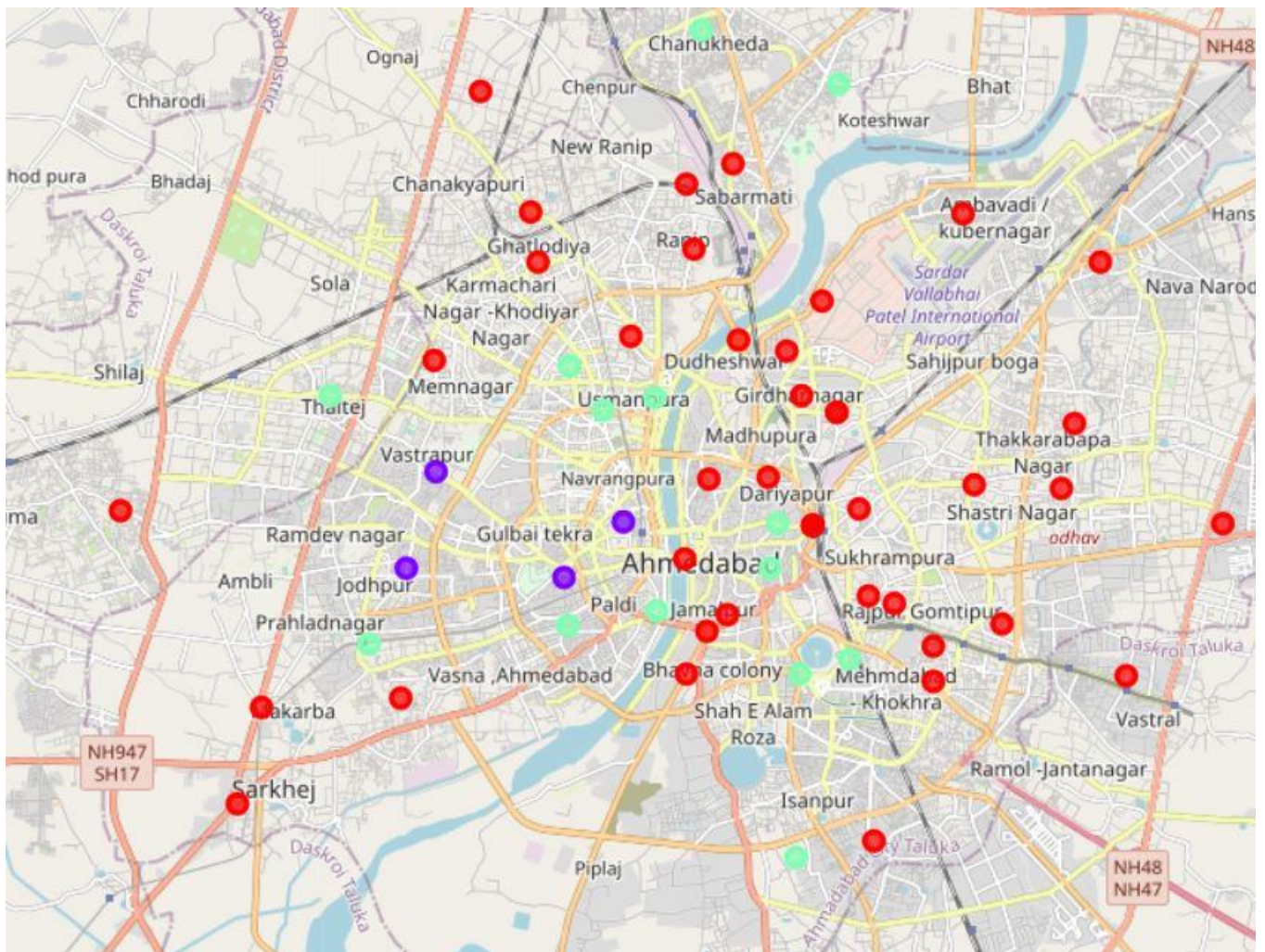
To extract the list of neighborhoods data, we'll use web scraping using Python requests and BeautifulSoup packages. This is, however, only a list of names. To use the Foursquare API, we need to obtain geographical coordinates in the form of latitude and longitude. To do so, we'll utilize the fantastic Geocoder library, which will translate the address into latitude and longitude coordinates. We'll collect the data, load it into a pandas DataFrame, and then use the Folium package to show the neighbourhoods on a map. This allows us to run a sanity check to ensure that the geographical coordinates data given by Geocoder is plotted appropriately in Ahmedabad. Then, we'll use the Foursquare API to get the top 50 venues within a 1000-meter radius. We need to register a Foursquare Developer Account in order to obtain the Foursquare CLIENT ID and Foursquare secret key. In a Python loop, we then make API requests to Foursquare, passing in the geographical coordinates of the neighborhoods. Foursquare will supply the venue data in JSON format, from which we will extract the name, category, latitude, and longitude of the venue.

We can use the data to see how many venues were returned for each neighbourhood and how many unique categories can be selected from all of the venues that were returned. Then, by grouping the rows by neighbourhood and calculating the sum of the frequency of occurrence of each venue category, we will analyse each neighbourhood. By doing so, we are also preparing the data for use in clustering. Since we are analysing the “Café and fast food stalls” data, we will filter out the this as venue allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms called KMEANS Clustering and is particularly suited to solve the problem for this project.

We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for Cafes. The findings will allow us to determine which neighbourhoods have a larger concentration and which neighbourhoods have fewer concentration of Cafes. It will help us answer the question of which neighbourhoods are most suited to open new shops based on the presence of different shops in different neighbourhoods.

## RESULTS

A good number of Caffe are concentrated near Sabarmati river in Ahmedabad. Cluster 2 shows the neighborhood which has highest number of cafe in surroundings which are marked with color blue. Cluster 1 shows the moderate amount of Snacks shop in the area. This represents a great opportunity and high potential areas to open new cafe as there is very little to no competition area marked with red color. Meanwhile, Shops in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of Cafes. Therefore, this project recommends Cafe frenchies owners to capitalize on these findings to open new shoppes in neighbourhoods in cluster 0 with little to no competition. Owner with unique idea can also try in Cluster 2. Lastly, They are advised to avoid neighbourhoods in cluster 1 which already have a high concentration of caffes and restaurants and suffering from intense competition.



## DISCUSSIONS

As observations noted from the map in the Results section, most of the Cafes are concentrated in the central city, with the highest datapoint in cluster 0 which represents less or zero number of Caffes nearby and moderate number in cluster 2. On the other hand, cluster 1 has a very low datapoints it denotes neighborhood which has more than caffes nearby. This represents a great opportunity and high potential areas to open new venture as there is very little to no competition from existing restaurants in Cluster 0. Meanwhile, cafes in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of other cafes. From another perspective, the results also show that the oversupply of Cafe mostly happened in the central area of the city, with the some area still have very few Caffes. Therefore, this project recommends franchise owners to capitalize on these findings to open new stall in neighbourhoods in cluster 0 with little to no competition.

## FUTURE WORK

The purpose of this project was to explore the Ahmedabad city and see what are potential spots to open new Café. In this project, we only consider one factor i.e. frequency of occurrence of shoppingmalls, there are other factors such as population and income of residents that could influence the location decision and rating of other restaurents. However, to the best of this researcher's knowledge, such data are not available at the neighbourhood level that this project requires. Future study could develop a mechanism for estimating such data, which could then be used in the clustering process to select the best sites for a new Cafe to operate. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

## CONCLUSION

In this project, we 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, and lastly providing recommendations to the relevant stakeholders. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a new cafe. The findings of this project will help the relevant stakeholders to capitalize on the opportunities.

## REFERENCES

[https://en.wikipedia.org/wiki/Category:Neighbourhoods in Hyderabad, India](https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India)

<https://developer.foursquare.com/docs/>



