

# The Battle of Neighbourhoods of Ahmedabad

## Introduction

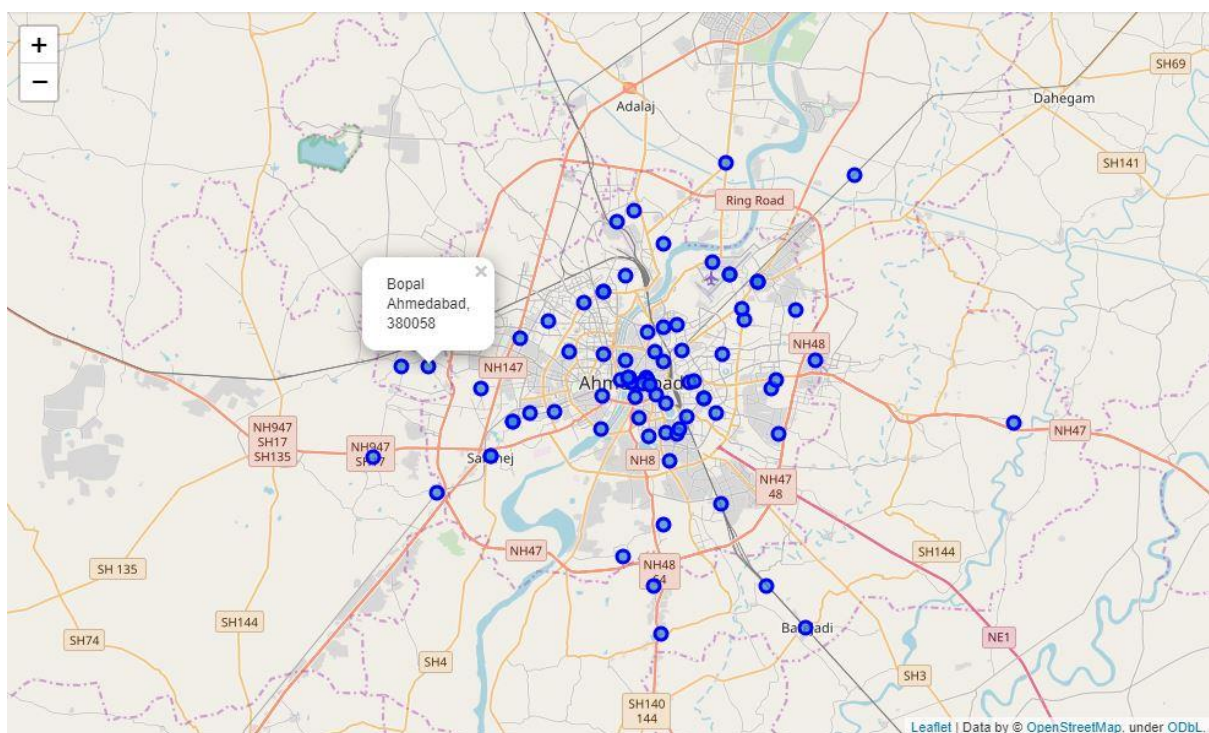
Ahmedabad is a UNESCO World Heritage city situated in the western region of India. One of the most common establishments of this city is paan parlour. These are small shops in the nature of mom and pop stores, which sells all kinds of stuff ranging from snacks to perfumes. One of the most popular products sold in these shops is paan – which is a recipe of arecanut and betel leaf garnished with flavouring agents like rose petals, fennel seeds, cherries, cardamom powder, coconut flakes and chopped dates. Local populace actively consumes this recipe after a sumptuous meal for its digestive properties. In this exercise, I try to identify suitable locations where a paan shop can be a thriving business. My goal is to find out places where Indian restaurants are in abundance. The rationale is that people visiting restaurants will be interested to savour a paan after their meal.

## Data

The areas of Ahmedabad, with their respective pincode, have been obtained from [this link](#). As the file format was a pdf, I used the online converter [Smallpdf](#) to convert my file to a docx file. Once my docx file was ready, I used MS Excel to convert text to columns and finally saved the file as csv. This method was extremely convenient for this exercise but I realised that this shall not be a suitable method for a database that runs into millions of rows.

## Analysis

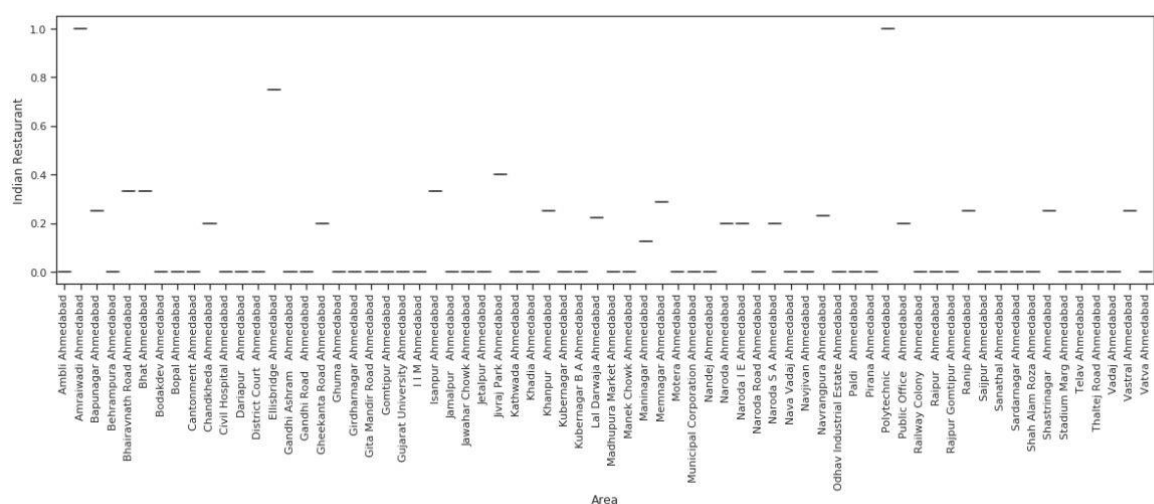
The notebook is prepared in IBM Watson Studio. The csv file prepared earlier was imported as data asset and inserted to code. Since my file had only areas and columns, I used geopy to add coordinates, latitude and longitude. Then, I used Folium to display the map based on the dataframe.



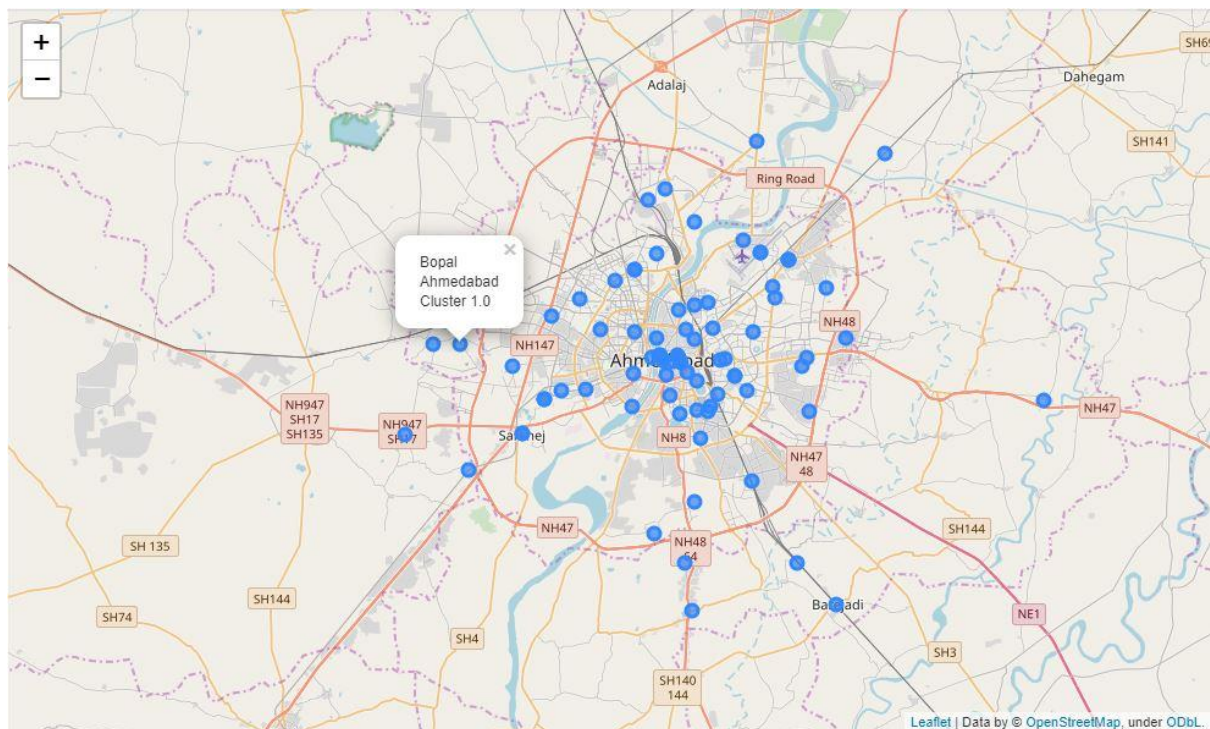
By this process, I was able to visualize the areas of the city along with their pincode.

Foursquare API was used to retrieve information about the areas based on geocoordinates. From the json file I obtained from Foursquare, I obtained the nearby venues and venue categories. Then I found out the most common venues for all areas.

I used seaborn boxplot to have a basic overview of the Indian restaurants' presence across the city. Here is the plot:



To group the areas based on their similarity (in terms of most common venue), I used scikit-learn. The clusters were thus obtained. The visualization of cluster marker on map was achieved through Matplotlib (along with Folium).



From the analysis, I identified five clusters with a certain amount of distinction. The first cluster is based on public transport systems like bus station. The second cluster is

recreational. It consists of parks, lakes, cafes, historical sites, zoo etc. The third cluster is geared towards water transportation like boat ferry (mostly for the riverfront location). The fourth cluster represents the business hub – electronic stores, farmers market etc. The fifth cluster is to my interest. It consists of umpteen Indian restaurants, diners and snack places.

Thus from the classification based on clusters, I found these places most suitable to establish a paan parlour:

1. Amraiwadi
2. Bapunagar
3. Bhat
4. Bhairavnath Road
5. Ellisbridge
6. Gheekanta Road
7. Isanpur
8. Jivraj Park
9. Lal Darwaja
10. Polytechnic
11. Shastrinagar

These areas are recommended to establish a paan parlour as the first most common venue is an Indian restaurant.

## **Conclusion**

By the end of this exercise I realised how powerful a tool is data science for informed decision making. Our life throws at us numerous data points which can be harnessed for information processing and insightful analysis. Given the paucity of resources, decisive allocation becomes utmost important. Through a mechanism like machine learning we can identify synergies for transformation of data points and visualize them for efficient and quick comprehension.

Hereby I would also like to mention the shortcomings, which I would like to improve in due course. This exercise is a preliminary effort towards use of tools like scikit-learn. My clustering was restricted to a limit for the sake of simplicity. “K-means clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm.” (Fränti & Sieranoja, 2019) For further discourse, model refinement can be done for better representation and accuracy of the outcome.

## **References**

Fränti, P., & Sieranoja, S. (2019, September). *ScienceDirect*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0031320319301608>