Opening new Italian Restaurant in Vadodara, India

**1. Introduction**

**1.1 Background**

Vadodara District is a district in the eastern part of the state of Gujarat in western India. The city of Vadodara (Baroda), in the western part of the district, is the administrative headquarters. Vadodara District covers an area of 7,794 km². As of 2011, the district had a population of 4,165,626 of which 49.6% were urban, 50.4% were rural, 5.3% were scheduled castes and 27.6% were scheduled tribes.[6] As of 2011 it is the third most populous district of Gujarat (out of 33), after Ahmadabad and Surat.[1]

The district is bounded by Panchmahal district to the north, Anand and Kheda districts to the west, Bharuch and Narmada districts to the south, and Chhota Udaipur to the east. The Mahi River passes through the district.

The historical city of Baroda was the capital of Baroda Residency, and one of the princely states of India under Bombay Presidency.

**1.2. Problem**

In this project, I will investigate that if there is any good location in Vadodara for opening Italian

restaurant and it really values to open one.

**1.3. Interest**

Anyone wants to open a new venue in any geographic location may be interested in this project by modifying search criteria.

**2. Data**

**2.1 Data sources**

To consider the problem we can list the data as below:

I get the neighborhood data of vadodara from Wikipedia [3]

<https://en.wikipedia.org/wiki/Vadodara>

<https://vmc.gov.in/development_plant.aspx>

<https://www.proptiger.com/vadodara/all-localities>

* I use python geocoder library to get geographical coordinates of neighborhoods[4]
* I use **Foursquare API** **venues explore** method to get the venues of given neighborhoods of vadodara [5].
* I use **Foursquare API venues method** to get ranks and likes of restaurants by given venue id [5].

**A. References:**

[3]<https://en.wikipedia.org/wiki/Vadodara>

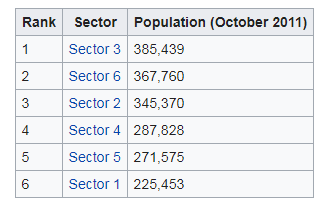
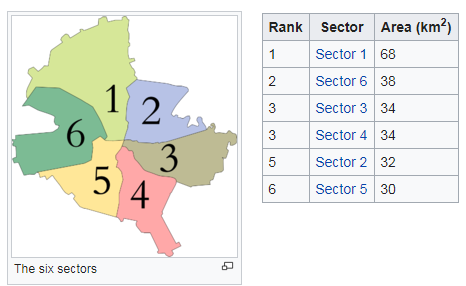
[4] [Python Geocoder Library](https://geocoder.readthedocs.io/)

[5] [Foursquare API](https://developer.foursquare.com/)

**2.2 Feature Selection and Data Usage**

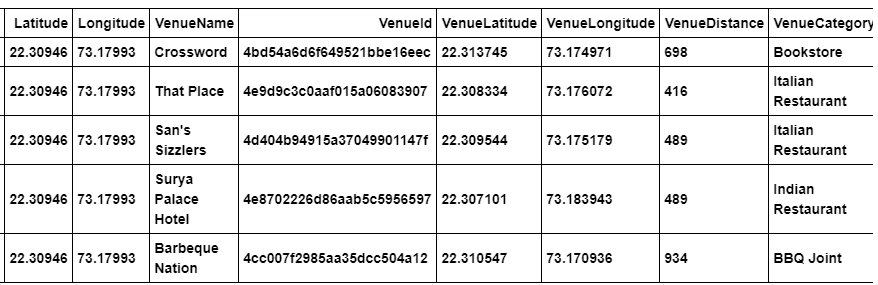
I will use neighborhood location values to analyze vadodara geographical structure. I will use folium library of python to draw maps by using given latitude and longitudes of neighborhoods. Selected features will be as below

I will use populations of neighborhoods who has big population may need a new restaurant.

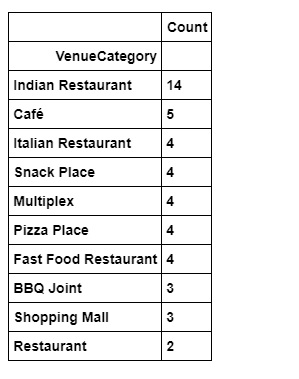
 

I will also classify neighborhoods by using venues distribution and counts. In this way, I will find similarities of neighborhoods which will help me to choose location for opening a new restaurant. Similar structured neighborhoods may handle same type of venue.

I will use venue list category to find distribution of restaurants and Italian restaurants in neighborhoods.



I will use also ratings and likes of restaurants in the areas. It may help me to find location with bad rating restaurants which may need a new restaurant.

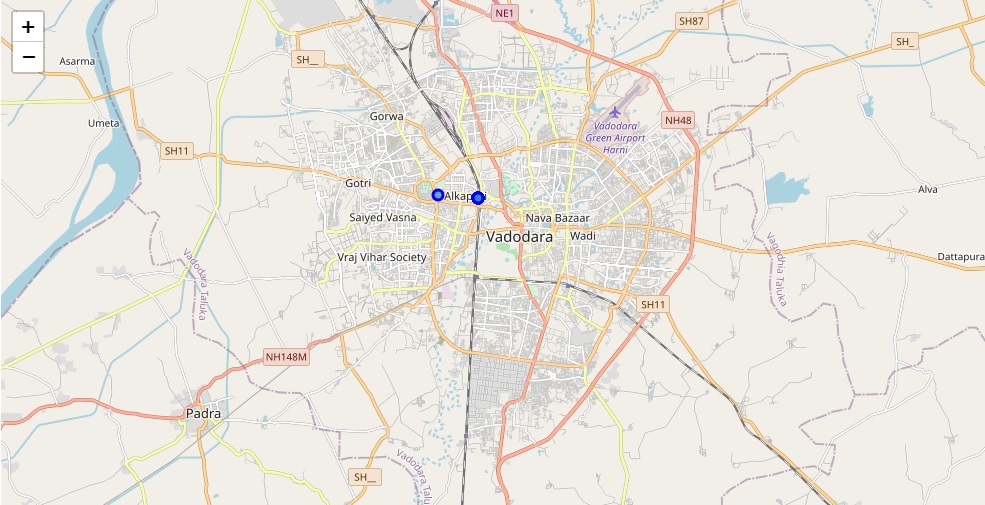


After that I will use clustering methods to produce sector wise distribution of venues.As shown in the figure below.

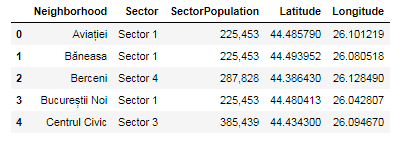


# 3. Methodology

* I used GitHub repository for code versioning.
* The Vadodara data is available with the neighborhood name in wikipedia. Location information (latitude and longitude) of neighborhoods are taken from geocoder library.
* I took neighborhood from wikipedia and put blue dots on Vadodara map to see centers of neighborhoods. There Geographical coordinates of five neighborhoods are as below

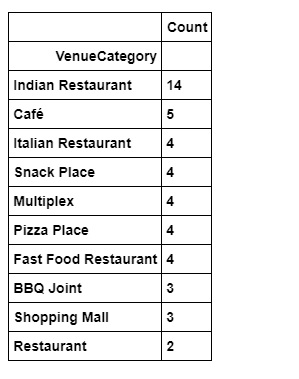


* I also got sector names of neighborhoods and populations of sectors. There are 6 sectors in Vadodara. I also merged Sector and Neighborhood data to see in which sector neighborhood places.



By using Foursquare API I got venues **1 km** around center of each neighborhood with limit **100 venue**. I merged data with Neighborhood data. Head of the merged data is as below

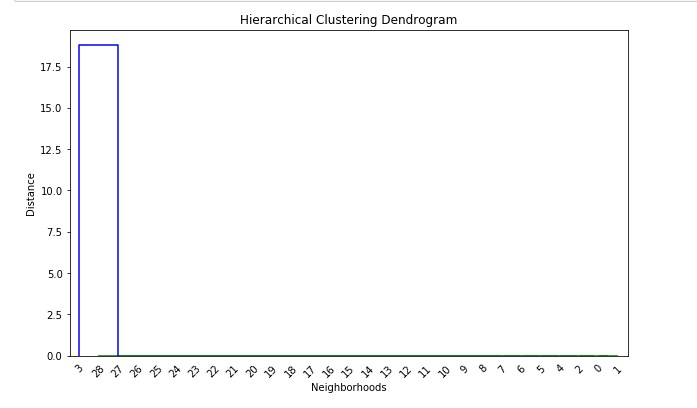
* **There are 39 unique venue categories. Some of them are as below:**



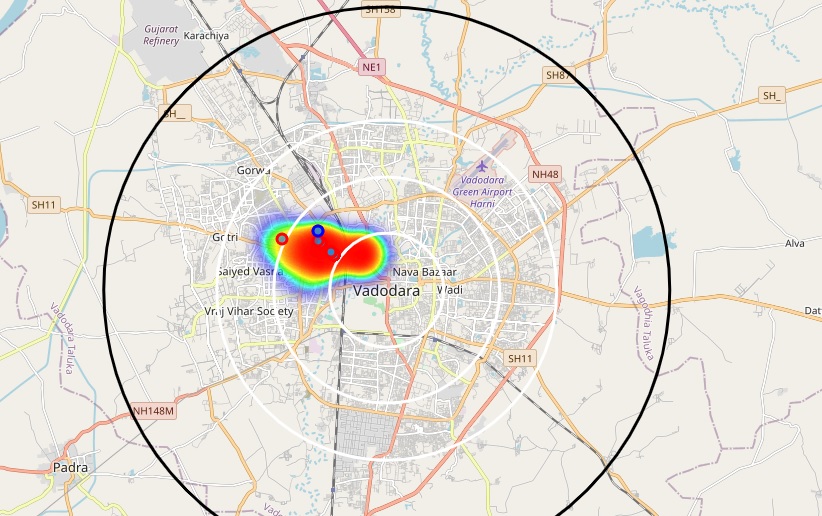
**Total number of restaurants: 46**

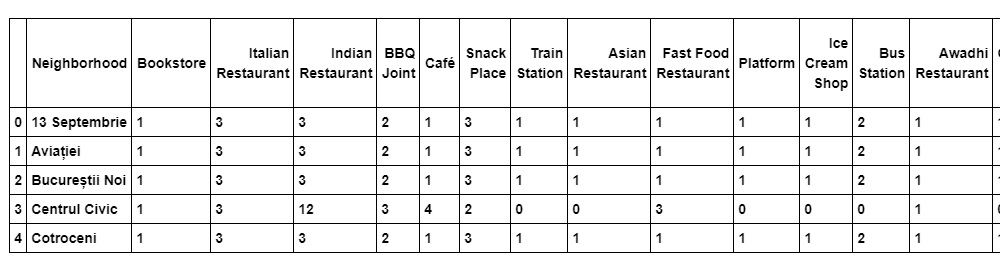
**Total number of Italian restaurants: 6**

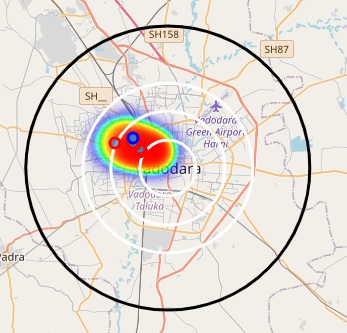
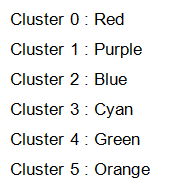
**Percentage of Italian restaurants: 13.04%**



* I used foursquare to get restaurant ratings by using venuid and merged with venue data
* I drawed heatmap for counts of restaurants in neighborhoods by using Folium library. On heatmap, I also drawed 4 few circles indicating distance of 2km, 4km, 6km and 10km from Bucharest center.
* **Red markers are Italian restaurants**
* **Blue markers are Italian restaurants with rating less than 7 restaurants**



* I used Agglomerative Clustering to cluster neighborhoods according to count of venue categories in each neighborhood. According to the below graph, I decided to separate our neighborhoods into 6 clusters (cut at distance of 20, horizontal black line).



Cluster 0 : Italian Restaurant , Pizza, Café

Cluster 1 : Restaurant , Bar

Cluster 2 : Park , Plaza, Clothing stores , Museum, Gym

Cluster 3 : Coffee Shop, Hotel, Pub

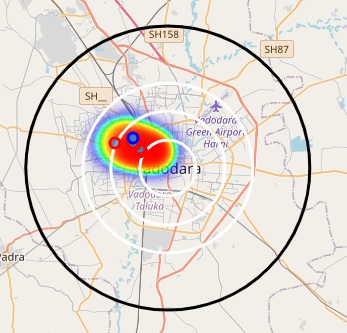
Cluster 4 : Café, Suprmarket

Cluster 5 : Café

# 4. Results

* As from the above experiments regarding heat map, folium map and clustering methods we can say that, Cluster 0 has most restaurants.
* Especially there is no Italian restaurant in west part including Cluster 3.
* We can think about Cluster 2, west part Cluster 0 especially likes Italian tastes.
* So, may be this part will not like any other tastes.

# 5. Conclusion



This way we have analyzed the neighborhood of Vadodara city and we have found the details of restaurants, mall, and multiplex. From restaurants we got insides about Indian restaurants. Italian restaurants, fast food, snacks, and cafes categories. After using clustering techniques we got more ideas about Italian restaurants. We found clusters, define heat map on it and showed practical results. Cluster 0 has most restaurants. Especially there is no Italian restaurant in west part including Cluster 3. We can think about Cluster 2, west part Cluster 0 especially likes Italian tastes, may be this part will not like any other tastes.

So we can set up Italian restaurant in cluster 1,2,3,4. Cluster 3 we be most beneficial as it has multiplex area.