

|  |
| --- |
| Battle of neighbourhood  Report |
|  |
| **September 12**  For Couresra IBM Data Science Certification  Authored by: Sanjeev Gupta |

1. **Introduction**
   1. Background

Restaurants are classified or distinguished in many different ways. The primary factors are usually the food itself (e.g. [vegetarian](https://en.wikipedia.org/wiki/Vegetarianism), [seafood](https://en.wikipedia.org/wiki/Seafood), [steak](https://en.wikipedia.org/wiki/Steak)); the [cuisine](https://en.wikipedia.org/wiki/Cuisine) (e.g. Italian, Chinese, Japanese, Indian, French, Mexican, Thai) or the style of offering (e.g. [tapas](https://en.wikipedia.org/wiki/Tapas) bar, a [sushi](https://en.wikipedia.org/wiki/Sushi) train, a [tastet](https://en.wikipedia.org/wiki/Tastet) restaurant, a [buffet](https://en.wikipedia.org/wiki/Buffet) restaurant or a [yum cha](https://en.wikipedia.org/wiki/Yum_cha) restaurant). Beyond this, restaurants may differentiate themselves on factors including speed (see [fast food](https://en.wikipedia.org/wiki/Fast_food)), formality, location, cost, service, or [novelty themes](https://en.wikipedia.org/wiki/Theme_restaurant) (such as [automated restaurants](https://en.wikipedia.org/wiki/Automated_restaurant)).

1.2 Problem

Urban conglomerations follow a pattern of development and this pattern varies from country to country to country and city to city. However, the need of a restaurant at a specific place is closely linked to the economic development of the area, paying capacity of the public, surrounding business and type of tourists and vistors. To open a new specialty restaurant in a city, the pattern of existing restaurant has to be analyzed as they are successfully running there venture and hence the needed criteria are met.

1. **Data acquisition and preprocessing**
   1. Area-wise latitude and longitude data

The data for pin code ( postal code of India) alongwith the location is available at Geonames site. They have provided the IN.txt file that contains all post codes of India. The following is the structure of the file The data format is tab-delimited text in utf8 encoding, with the following fields :

* **country code** : iso country code, 2 characters
* **postal code** : varchar(20)
* **place name** : varchar(180)
* **admin name1** : 1. order subdivision (state) varchar(100)
* **admin code1** : 1. order subdivision (state) varchar(20)
* **admin name2** : 2. order subdivision (county/province) varchar(100)
* **admin code2** : 2. order subdivision (county/province) varchar(20)
* **admin name3** : 3. order subdivision (community) varchar(100)
* **admin code3** : 3. order subdivision (community) varchar(20)
* **latitude** : estimated latitude (wgs84)
* **longitude** : estimated longitude (wgs84)
* **accuracy** : accuracy of lat/lng from 1=estimated, 4=geonameid, 6=centroid of addresses or shape
  1. Foursquare api

Foursquare is a social networking service available for common PC as well as for smartphones. The app's purpose is to help you discover and share information about businesses and attractions around you. The api requires the location parameters (longitude and latitude). Based on location parameters, it provides details of different venues within a periphery of configured distance.

In case of this project, the input of location for each of the post codes of the two cities were provided to the api and it received the details of the surrounding venues.

* 1. Preprocessing

Postal code file contained the details of all postal codes in India. Two dataframes were created from it. One for Mumbai and another for Delhi. The following cleaning was done.

* File was tab delimited and there was no header row
* Headings were added.
* The columns not required were removed.
* Duplicate rows were removed

After these changes, the data was looking like below

|  | **PIN** | **Place** | **area** | **region** | **area2** | **latitude** | **longitude** | **accuracy** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 14927 | 110001 | New Delhi G.P.O. | Delhi | New Delhi | New Delhi | 28.6369 | 77.2183 | 3 |
| 14928 | 110002 | Civic Centre | Delhi | New Delhi | New Delhi Central | 28.6453 | 77.2456 | 3 |
| 14929 | 110020 | F F C Okhla | Delhi | New Delhi | New Delhi | 28.5345 | 77.2779 | 3 |
| 14930 | 110029 | Nauroji Nagar | Delhi | New Delhi | New Delhi | 28.5650 | 77.1972 | 3 |
| 14931 | 110043 | Gopal Nagar | Delhi | New Delhi | Delhi | 28.6109 | 77.2140 | 4 |

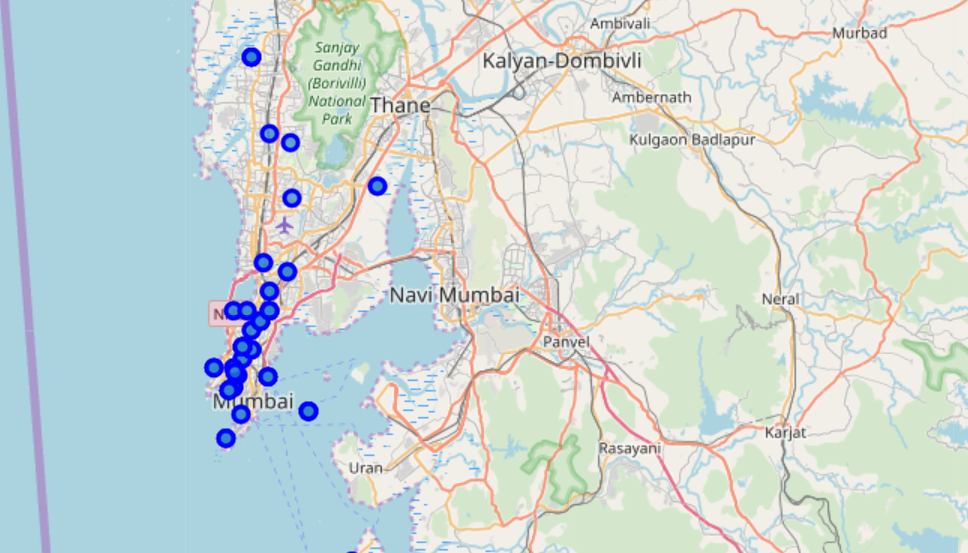
# Methodology

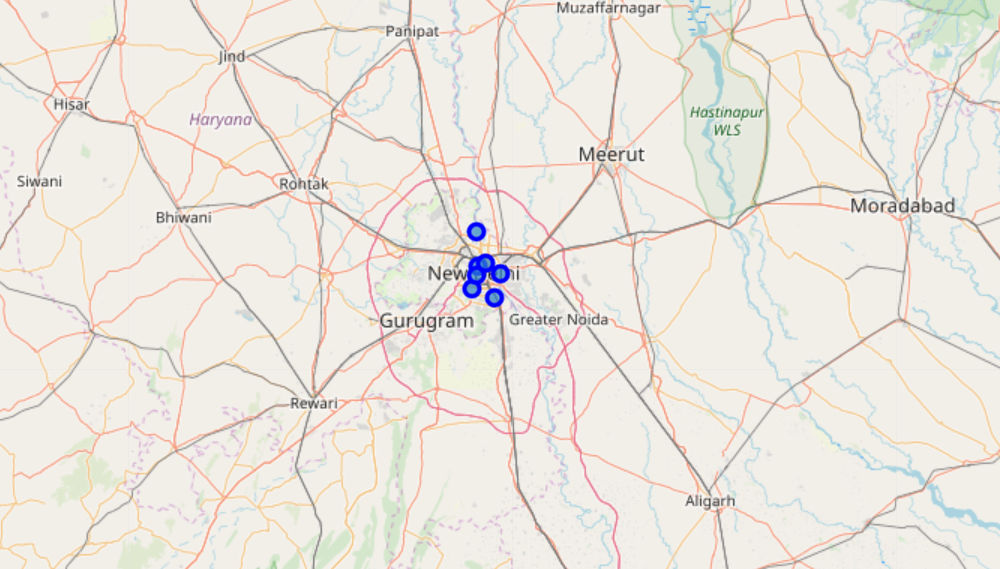
* 1. Process Steps and strategy

1. Check the major venue category based on the business run in the two cities using foursquare api
2. Check for the businesses with the lower numbers for the specialty
3. Check different venue category by filtering data only for the food businesses
4. Look for the lower number outlets in the food business
5. Create bar chart and Folium maps for visual interpretation
6. Filter out the specialty restaurants from the data
7. Map again to see their concentration
8. Propose the recommendation

* 1. Exploratory Analysis

Different postal areas were drawn on the map using Folium library. The following is the view of the two cities





Next, the data was checked for most popular venue groups in the cities. The following are the top ones from Mumbai

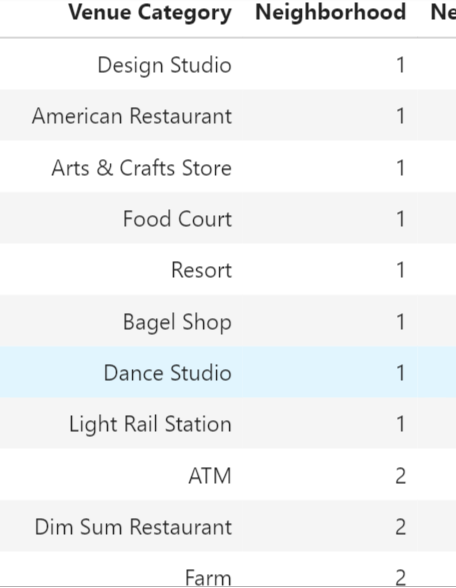


The following is the list of top venue category in New Delhi.

One thing is common in both the lists is that both are predominantly food serving categories



The next thing is to check the list of businesses having less venues/outlets.



List for New Delhi

# 

The lists on previous page are the sample from the lists from Mumbai and New Delhi respectively. These have non restaurant businesses as well. If we check for eateries, the following is the list :

1, Portugese restaurants

2. American restaurants

3. South American Restaurants

4. Korean Restaurants

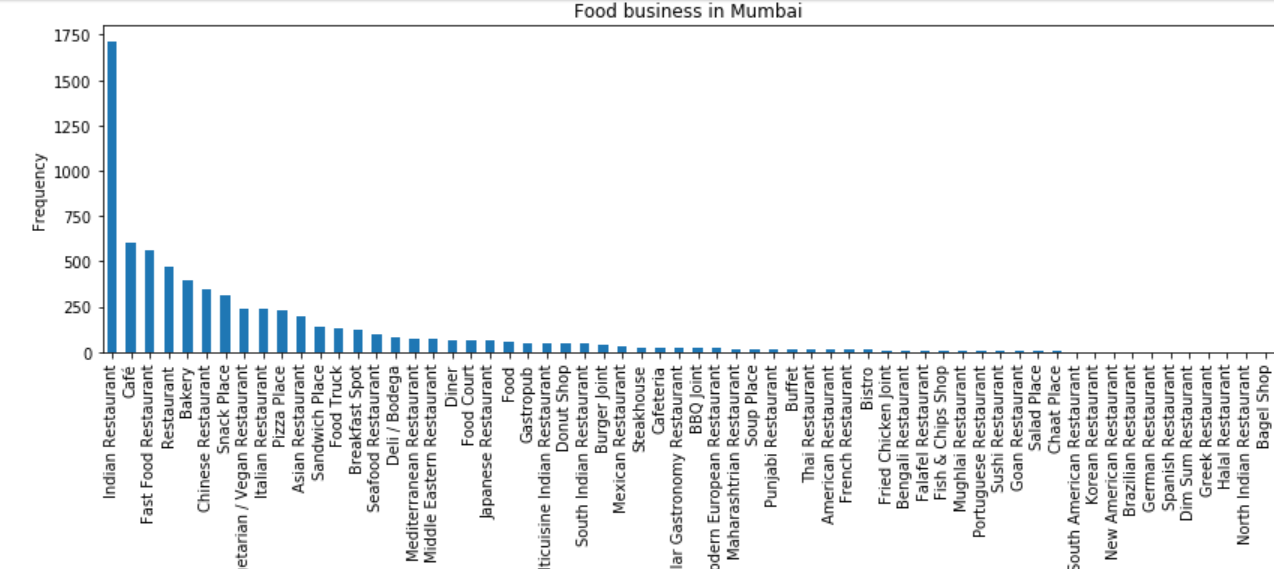
5. Spanish Restaurants

6. Scandinavian

7. Modern European

8. Turkish

Other view of the same data in bar form is depicted below

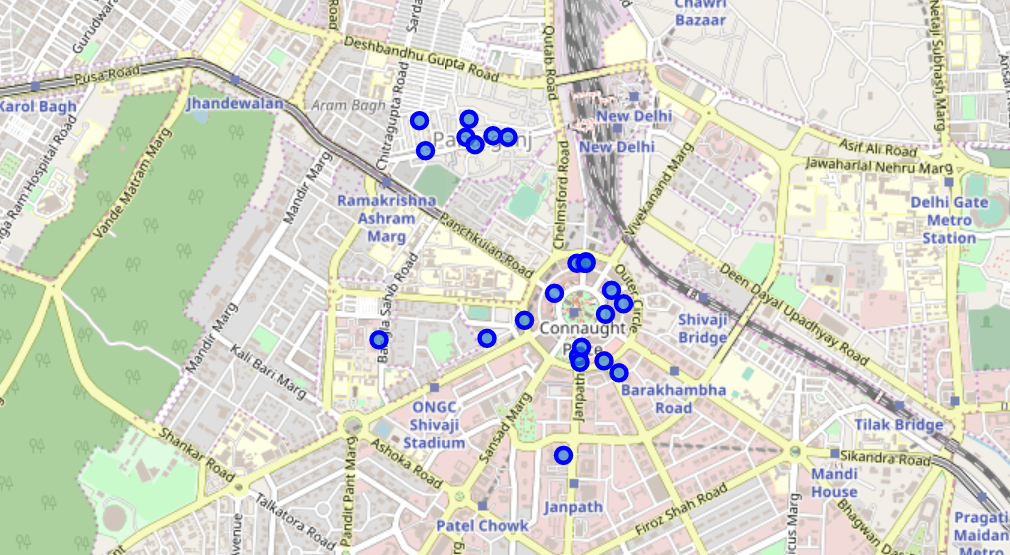


# 

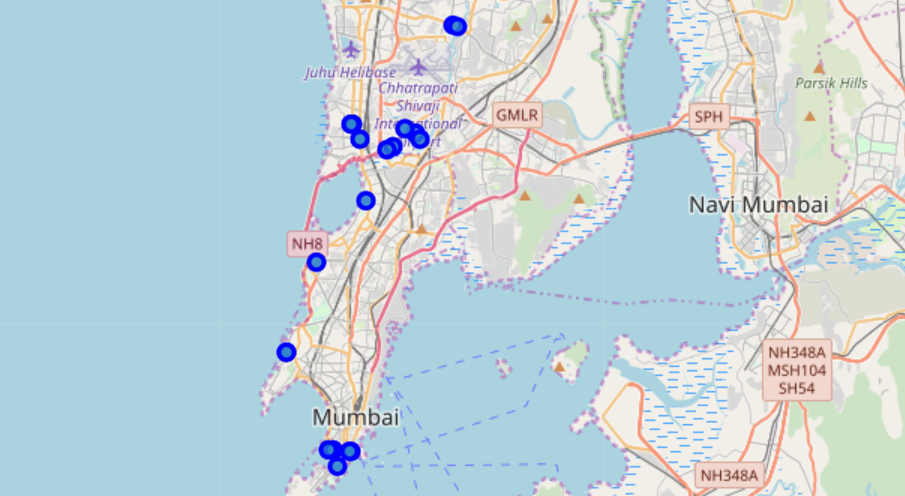
The above is the bar chart of Food business in Delhi.

Let us see the concentration of the Specialty Food business in two cities

In the following map, we can see that the specialty restaurants are concentrated in Paharganj and Cannaught Place areas ( as seen in map below)



In case of Mumbai, the specialty restaurants are at Nariman point, beaches or in Bandra



1. Result and Conclusion

1. The data shows that the restaurants, Eateries, bars are the top businesses in Mumbai. The same pattern is seen in New Delhi. Indian restaurants, Fast Food, Chinese restaurants, Coffee, ice cream shops, juice bar are at the top.

2. Portugese restaurants, American restaurants are few in numbers. There is oppotunity for such restaurants as speciality ones. The same is true for South American, Korean and Spanish restaurants.

3. In New Delhi, Museum also appeared in top 5 as there are many museums in New Delhi. While in Mumbai, it is very low in the venue category.

5. Recommendation

The following places are good for specialty restaurants

**Delhi** : Cannaught Place and Pahar Ganj

**Mumbai** : Nariman Point and Bandra

6. Observation

The categorization available at the Foursquare database is not refined as there are separate categories for 'New American Restaurant' and 'American Restaurant', it is difficult to grasp the difference. Similarly, there is one item falling under category 'Road'.