**PROBLEM & BACKGROUND**

Mumbai, Pune and Nagpur are the three most famous cities of the state of Maharashtra. All the cities are diverse in many ways. There has been a war for supremacy in terms of quality of life, jobs, education, entertainment and recreational facilities that these cities have to offer. We want to explore how much they are similar or dissimilar in aspects from a resident point of view regarding food, accommodation, beautiful places, and many more. We will be exploring the following cities and tries to understand what is popular in them and what they have to offer to someone who is contemplating to make a choice on seeking a life in either of the metro cities

# The deciding factor for most would be on how lively, supportive, vibrant and unique each of the cities can be when compared to each other. The business problem in this study assumes that people who would be interested in this study are those who would like to create a projection of potential life and activities in these metro city neighbourhoods if the subject moves to live in one of them. The decision to choose one over the other would depend on popular venues in the neighbourhoods in each of these metro cities.

# DATA DESCRIPTION

### For any “data science project” data is of paramount importance. For this study, we needed data about neighborhoods in each of these metro cities. The data published by the government on postal codes for all India would serve us well for this study. We will specifically download the CSV provided under <https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude>.

For this problem, we will read the csv into a pandas Dataframe and curate it to remove the data related to all other cities, towns, and places which are not in our list. We shall then clean up the unnecessary columns in the CSV, which is not relevant or useful for our current problem. Post office names (*office name*) will be used as the neighborhood names in each of the regions. Duplicate Pincodes will be removed.

***Nominatim*** library will be used to find the longitude and latitude of each of the neighborhoods in both Mumbai ,Pune and Nagpur. This will form the dataset we will use for this study.

# The first few records of the dataset we now have after clean-up and curation appear as below.

# 

### 

### We now have the neighbourhood’s for both the metro cities

### The next step is to enhance the dataset with the required information. We would need the longitude and latitude values for the neighborhoods. We will use the *Nominatim* library from *geocoders.geopy* package to find the longitude and latitude for each of the neighborhoods and would eventually create a dataset having all the necessary columns for our analysis.

### 

### 

### 

### We now have the necessary information to visualize the neighborhoods for both the cities on a *folium* map.

### 

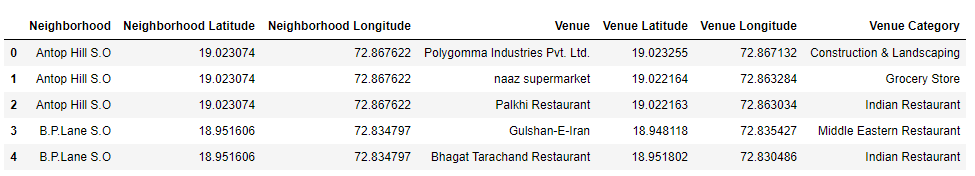
### 

### 

**METHODOLOGY**

**Finding top venues near neighbourhoods**

We will use the Foursquare API to find the top venues in the neighbourhoods of Mumbai. This will help us understand the nature of life Mumbai neighbourhoods have to offer. We will iteratively make Foursquare API calls for each of the Mumbai neighbourhoods in our dataset. Foursquare API returns the popular venues within 500m radius of this neighbourhood.



We will repeat the same step for both Pune and Nagpur and get popular venues within 500 m radius for their entire neighbourhood





Next, we will employ statistically and analytical methods to find the unique venues/venue categories in the Mumbai neighbourhoods and we will build a Dataframe that calibrates each of the neighbourhoods with the frequency of occurrence for each of the venue category.

From our analysis, we see that there are 184 unique venue categories in Mumbai neighbourhoods. We then create a dataset that lists the top 10 common venues against each of the neighbourhoods in Mumbai. We get a representation such as below for all the neighbourhoods in Mumbai.

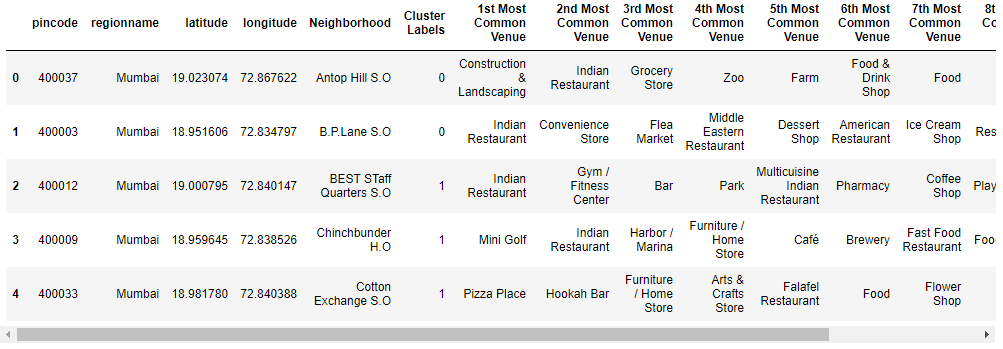


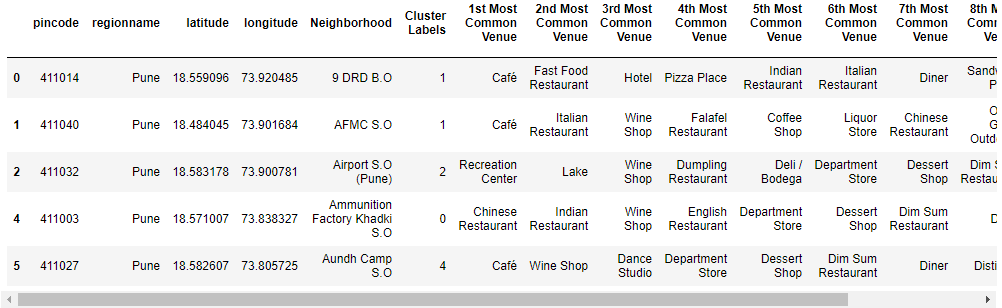
Similarly we will create a dataset that lists the top 10 common venues against each of the neighbourhoods in Pune & Nagpur. We get a representation such as below for all the neighbourhoods. Pune has 97 unique venues while Nagpur has 45 unique venues. 



**Cluster the neighbourhoods in cities based on the similarity of top common venues**

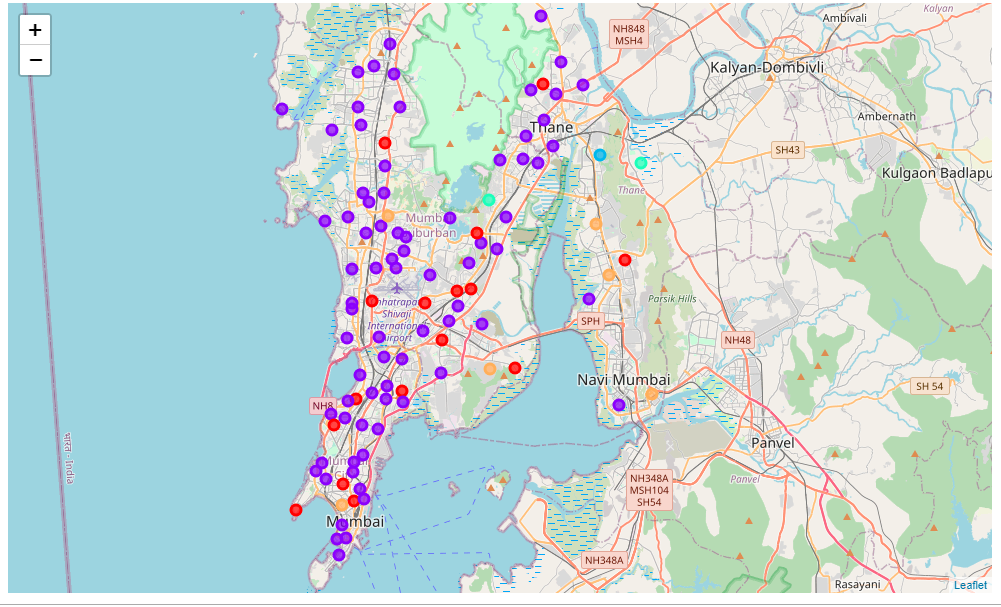
Given that we now have the required information regarding the top venues against each of the neighborhoods in all the cities, let us now apply a clustering algorithm to group the neighborhoods based on the similarity in types of venues they have. By clustering, we also provide information to users on a common type of neighbourhoods in each cities. We will use the k-Means clustering approach to cluster the neighbourhoods. k will be selected as 10. This means that we will group the neighborhoods into 10 clusters. Each of the neighborhoods gets a Cluster Label assigned.

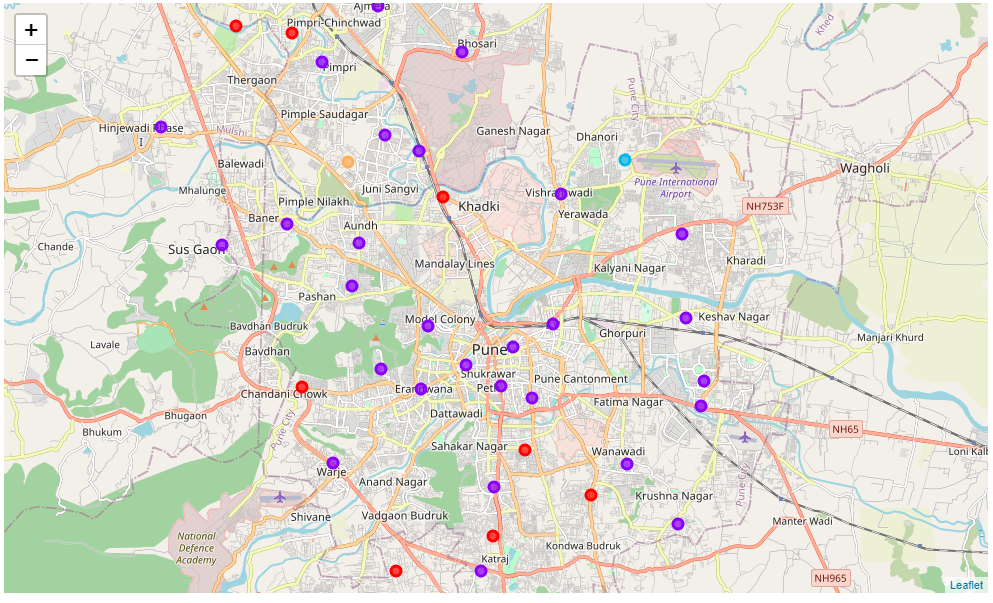


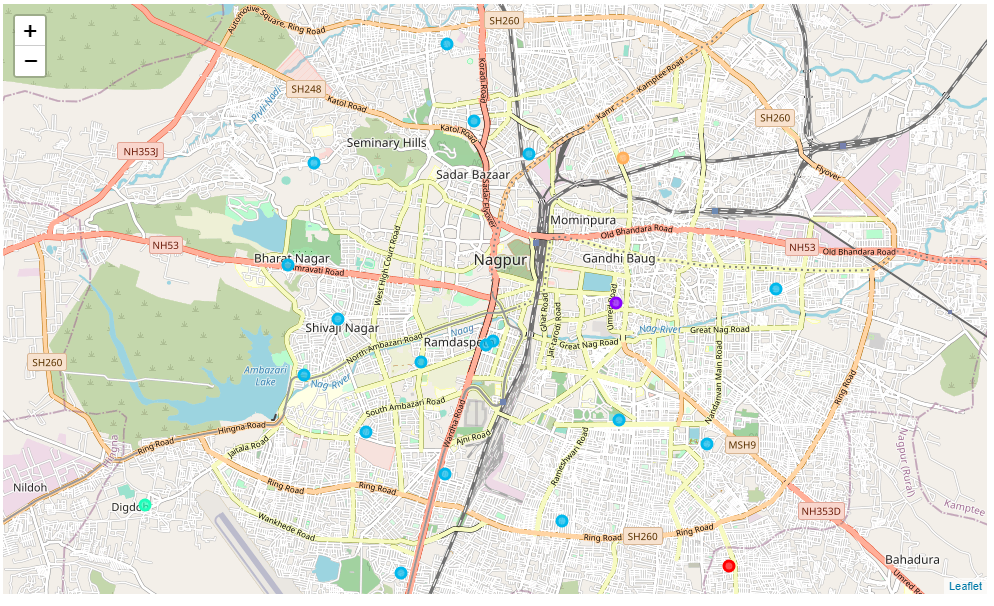




We will then use the dataset with cluster labels assigned to visualize the clusters in a *folium* map.







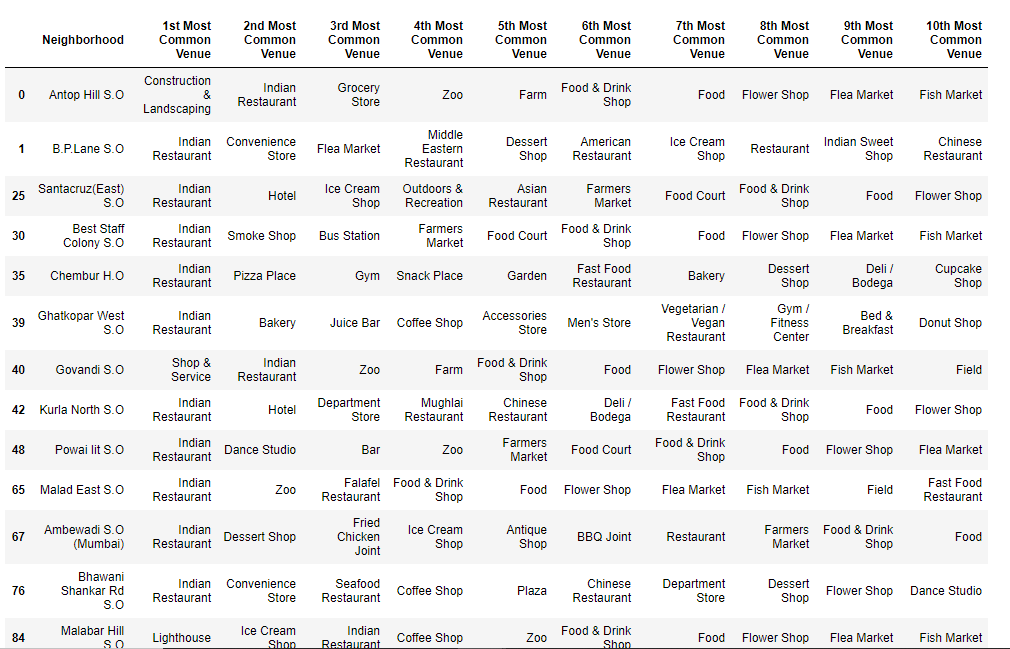
## EXAMINE CLUSTERS

### Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster.

### Mumbai

### A piece of important information this map provides is that many neighbourhoods in Mumbai are of similar nature concerning the venues they have around, indicated by the cluster marked in violet. Let us now dig a little deeper into how the neighbourhoods are clustered and what is the characteristic of the cluster that is very common across most neighbourhoods in Mumbai.

### Cluster 0



There are close to 16 neighbourhoods belonging to this cluster type. This cluster is popular for having Restaurant, Ice Cream Shop, Shop & Service, Bakery, Grocery Store. These neighborhoods are good for families, since the venues close to these neighborhoods are great for people with families.

**Cluster 1**

### 

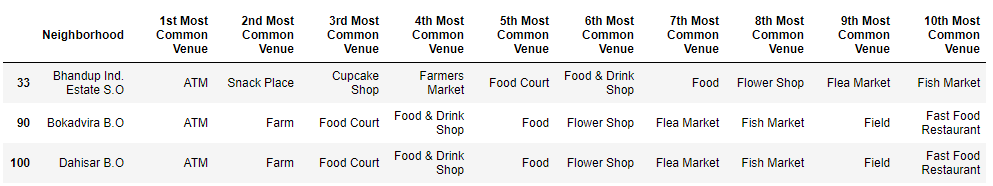
There are close to 76 neighbourhoods belonging to this cluster type. This cluster is popular for having Restaurant, Cafe Shop, Pizza Place, Dessert Shop, Bars etc. These neighborhoods are good for foodies.

**Cluster 2**

****

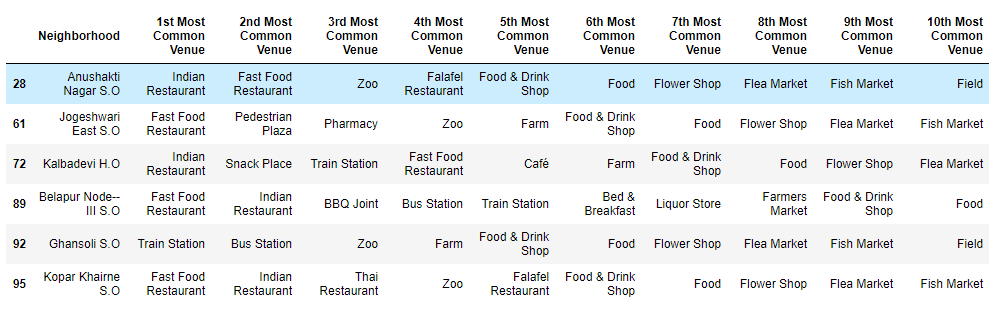
Not many neighborhoods belong to this cluster, Business Service, Zoos and Market seem to be popular venues close to the neighborhood in this cluster.

**Cluster 3**

****

Not many neighborhoods belong to this cluster, ATM, Market and Food Court seem to be popular venues close to the neighborhood in this cluster.

**Cluster 4**

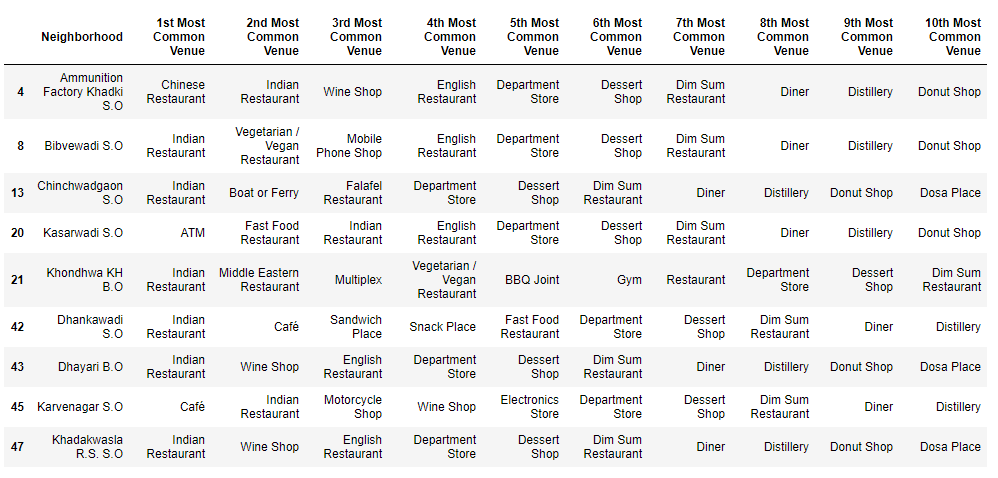
****

The neighborhoods belonging to this cluster is popular for having a mix of Indian and Thai restaurants, Train stations, Bus stations, Liquor Stores, Pharmacy, etc. We see that this neighborhood would be something that would be interesting to those who depend more on the public commute since these neighborhoods are closer to train and bus stations. Also, these neighborhoods may interest people who have diverse food choices starting from Indian, Asian to having Snacks, Sandwich, Ice-cream shops.

### Pune

### A piece of important information this map provides is that many neighbourhoods in Pune are of similar nature concerning the venues they have around, indicated by the cluster marked in violet. Let us now dig a little deeper into how the neighbourhoods are clustered and what is the characteristic of the cluster that is very common across most neighbourhoods in Pune.

### Cluster 0

****

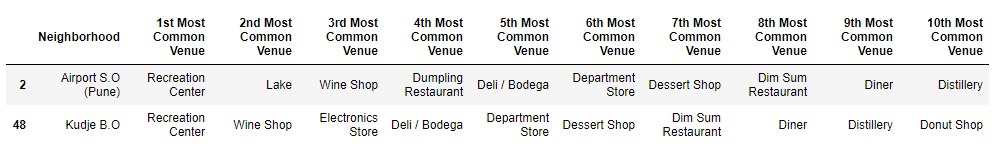
There are close to 9 neighbourhoods belonging to this cluster type. The neighbourhoods belonging to this cluster is popular for having a mix of Indian, Chinese and Middle Eastern restaurants, ATM, Mobile phone shop, Motorcycle shop etc. We see that this neighbourhood would be something that would be interesting to people who have diverse food choices starting from Indian, Arabic, Chinese, and Vegan to having Snacks, Sandwich and Fast Foods etc.

### Cluster 1

### There are close to 29 neighbourhoods belonging to this cluster type The neighbourhoods belonging to this cluster is popular for having Indian restaurants, Cafes, Seafood, and fast food joints. We see that this neighbourhood would be something that would be interesting to those who would like Seafood, fast food. Probably these neighbourhoods are also of interest to those who come from outside and would like to visit places serving different kind of food as there are large numbers of hotels here

### 

### Cluster 2

Not many neighborhoods belong to this cluster, This cluster is popular for having Recreation Centre, Lakes, Distillery and Deli/Bodega. These neighborhoods are not good for foodies. However, this should be good for those who have children, since the venues close to these neighborhoods are great to keep the children engaged.

### Cluster 3

### 

Not many neighborhoods belong to this cluster, Stores, Wine Shop and different type of restaurant seem to be popular venues close to the neighborhood in this cluster.

### Cluster 4

Not many neighborhoods belong to this cluster, Cafes, Wine Shop and different type of restaurant seem to be popular venues close to the neighborhood in this cluster.

### 

### Nagpur

### A piece of important information this map provides is that many neighbourhoods in Nagpur are of similar nature concerning the venues they have around, indicated by the cluster marked in blue. Let us now dig a little deeper into how the neighbourhoods are clustered and what is the characteristic of the cluster that is very common across most neighbourhoods in Nagpur.

### Cluster 0

### 

Not many neighborhoods belong to this cluster, Stores, IT Services, Garden and different type of restaurant seem to be popular venues close to the neighborhood in this cluster.

### Cluster 1

### 

Not many neighborhoods belong to this cluster, Different Type of Restaurants, Stores, Garden seem to be popular venues close to the neighborhood in this cluster.

### Cluster 2

There are close to 17 neighbourhoods belonging to this cluster type. The neighbourhoods belonging to this cluster is popular for having a different type of restaurants like cafes,Pizza,Indian,Hookah Bar etc. We see that this neighbourhood would be something that would be interesting to people who have diverse food choices.

### 

### Cluster 3

### 

Not many neighborhoods belong to this cluster, Dhabas, Gym, Garden seem to be popular venues close to the neighborhood in this cluster.

### Cluster 4

### 

Not many neighborhoods belong to this cluster, Pharmacy, Stores, Garden seem to be popular venues close to the neighborhood in this cluster.

**RESULT**

In this project, we have attempted to load the dataset for three of Maharashtra’s prime cities and have tried to analyse the neighbourhood regions in these cities based on the type of popular and top venues they have. We have clustered the neighbourhoods based on the most common top venues in each of the neighbourhood. Our intention with this project was to analyse and understand the difference in the type of life in these cities, which can offer decision points for anybody who is considering settling in either of the cities and can get a peek into what type of experience and facilities they will be provided with.

Given our cluster information for Mumbai, Pune and Nagpur, we see that all three cities and its neighbourhoods are a great place for a foodie. There are a lot of restaurants, cafes, bars, etc. in Mumbai and Pune neighbourhoods, while there are lots of street foods in Nagpur. Also due to the proximity of Mumbai to the seashore, Mumbai neighbourhoods offer for harbours, seafood, boat, and ferry rides. On the other hand, we see how dissimilar life in Nagpur neighbourhoods would be compared to Mumbai/Pune neighbourhoods. Nagpur neighbourhoods are good for those who like gardens and Street Food places. Pune neighbourhoods are more similar to Mumbai. There is very less in terms of foreign cuisine restaurants in Nagpur. Mumbai/Pune, on the other hand, is great for international visitors, expats, etc., because of the variety and types of food outlets it has.

**Discussion**

Thus with this project, we have analysed the kind of life each of these big cities has to offer based on the popular venues in their neighbourhood.

As one can notice, the number if venues we get from foursquare don’t reflect the reality, as these cities has far more venues. But, since Foursquare is not popular in India, not many venues are registered on foursquare.

Mumbai would be the choice if you want a fast life

Nagpur would be the choice if you want to live a relaxed life

Pune is somewhat between these two cities, having both a fast life and relaxed life.

**Conclusion**

The outcome of the project was impacted by the limited effectiveness of the Foursquare API for a city like Mumbai, Pune and Nagpur. The way the neighbourhoods were suggested was majorly due to Indian restaurants alone.

The results suggest that this approach would yield insights on common trends between areas. This analysis can be further improved by using more popular source like Google Maps

But with the data available the study reveals is that the categories of venues Mumbai/Pune offers are far too many compared to Nagpur. This means that Nagpur becomes restrictive in terms of variety and convenience. With the data, we have studied Pune wins this battle of cities!