



# Data Science Capstone Project

A Comparison of two top cities if India

# The Problem

- The problem stands like comparing the two cities in terms of important venues i:e which city has more number of a particular category of venues. Also, I will try to compare the neighbourhoods of individual cities and try to infer some similarities and differences between them. Lastly, I will try to extract certain locations from both the cities where establishing a business will be of interest. I use here the concept of absence of a particular venue category and establish the fact that it will be beneficial to establish that particular business unit in that Neighbourhood.



# Why Delhi and Mumbai

- Delhi is the capital city of India and is regarded as the heart of the nation. The city is popular for its enriched culture and heritage. The city hosts some famous historical monuments and is developing with the passing of time. The capital city is divided into two sections popularly known as Purani Dilli or Old Delhi and Nayi Dilli or New Delhi.
- Mumbai, formerly known as Bombay, is among the top five metropolitans of the country, and one of the most rapidly developing regions in India as well. It is also called as the financial capital of India. Mumbai stands tall as the hub of economic development, contributing to as much as 5% of India's GDP and it also finds its name amongst the top 10 commercial centres across the globe.

# Data Description

I am supposed to use Wikipedia to get the neighbourhoods list of both the cities. This data is available. Later I will manually add the coordinates of each neighbourhood to the list. I will compile two .csv files for both the cities.

I will then use Four Square API to generate a list of venues of each neighbourhood. The populated list will be used to infer important facts about the two cities. We will do a brief comparison based on this data.

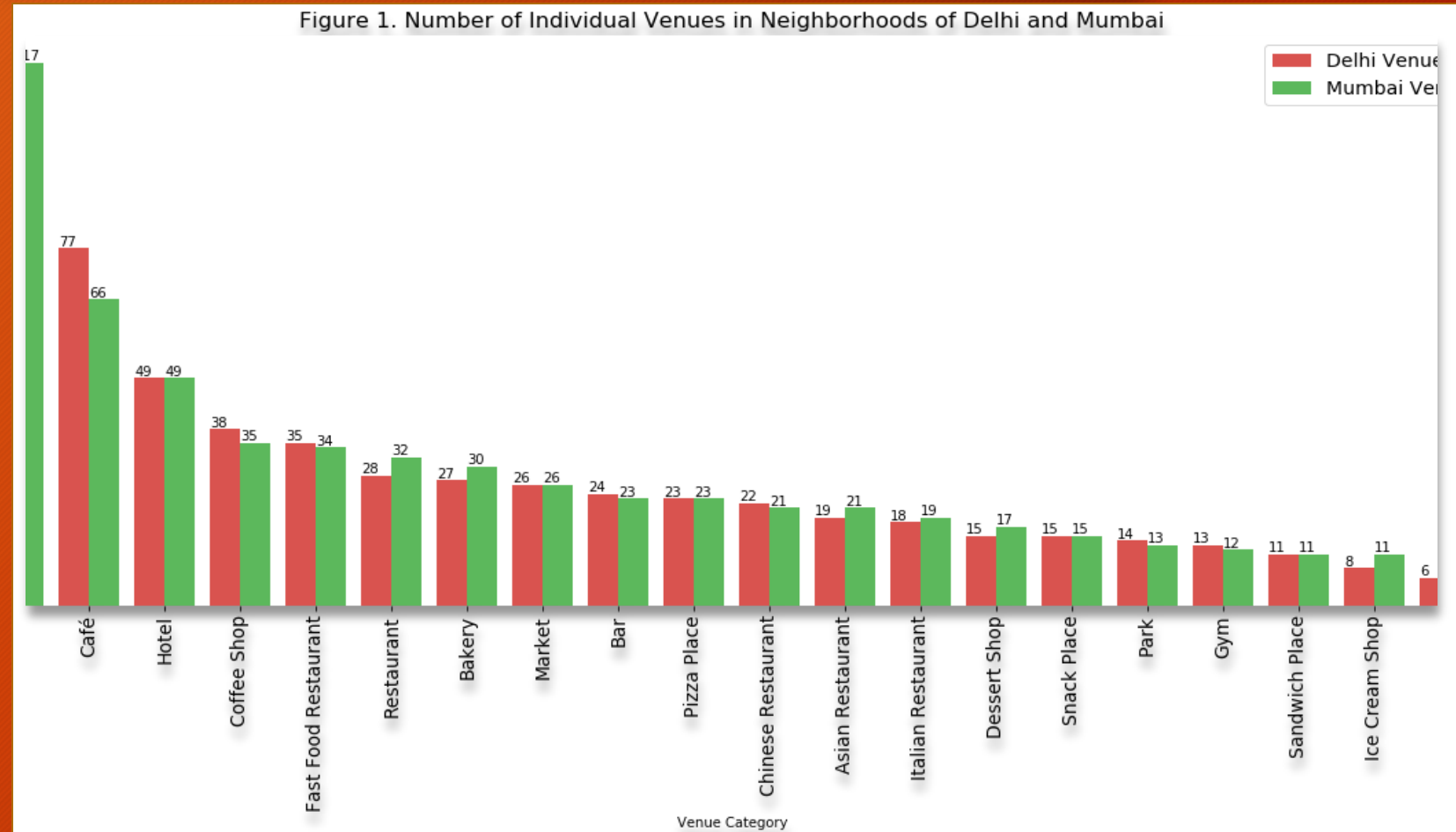
# Methodology

## B.1 Finding the relation between Delhi and Mumbai

As a database, I used GitHub repository in my study. My master data which has the main components *Neighbourhood*, *City*, *Latitude* and *Longitude* information of the city.

I utilized the Foursquare API to explore the Neighbourhoods and segment them. I designed the limit as **100 venue** and the radius **500 meter** for each Neighbourhood from their given latitude and longitude information.

Plotting these values with a bar chart for visualization.





# Methodology

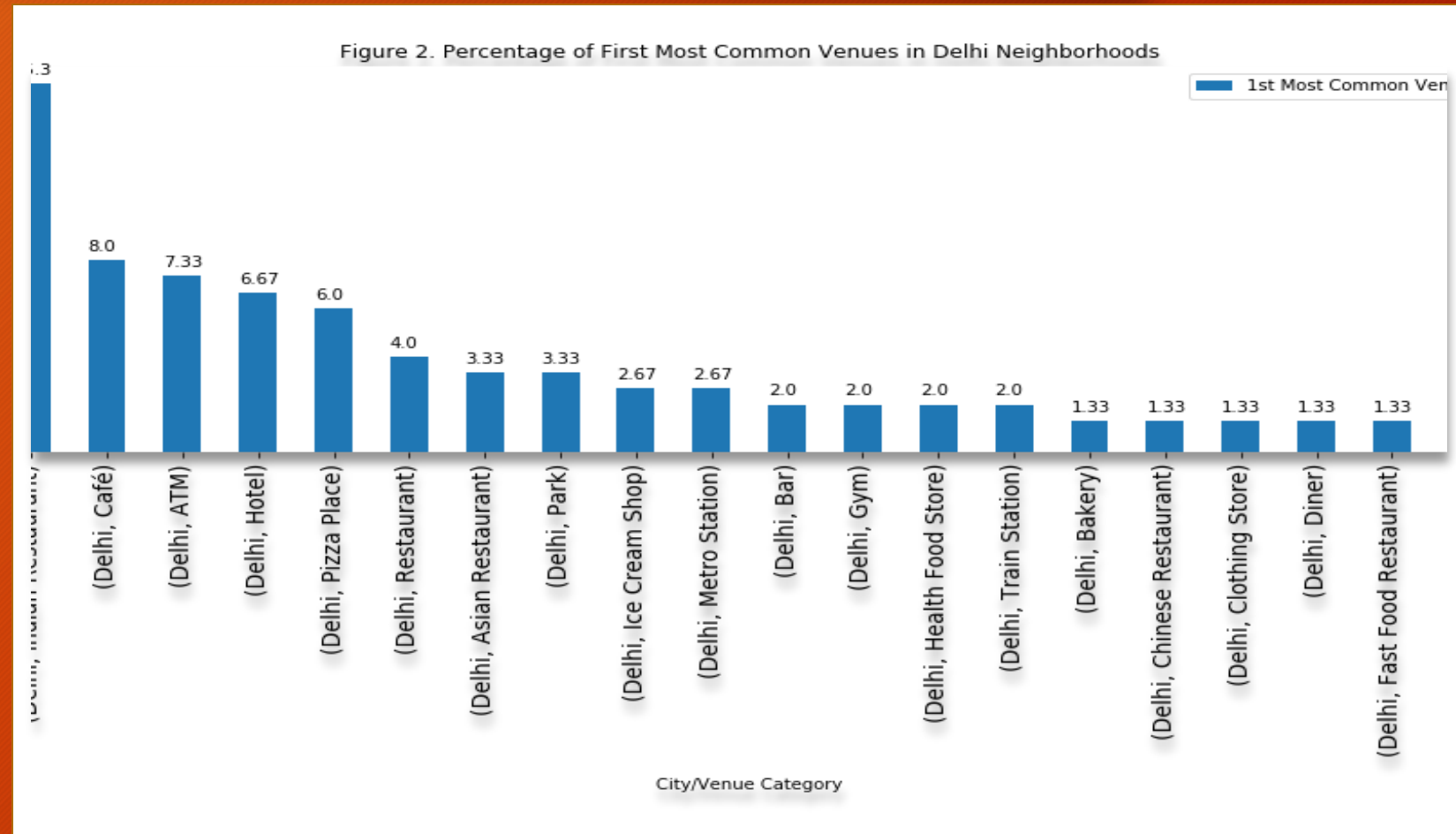
## B.2 Finding Relations between various individual neighbourhoods of Delhi

Next to find relation between various individual neighbourhoods of Delhi I, first extracted top five venues of each neighbourhood. I, did so by applying one hot encoding to the dataset returned by the Four Square API.

Then I extracted from this dataframe two dataframes one each for Delhi and Mumbai.

Next I calculated the percentage values of 1st most Common Venues and selected the top 20 for comparison.

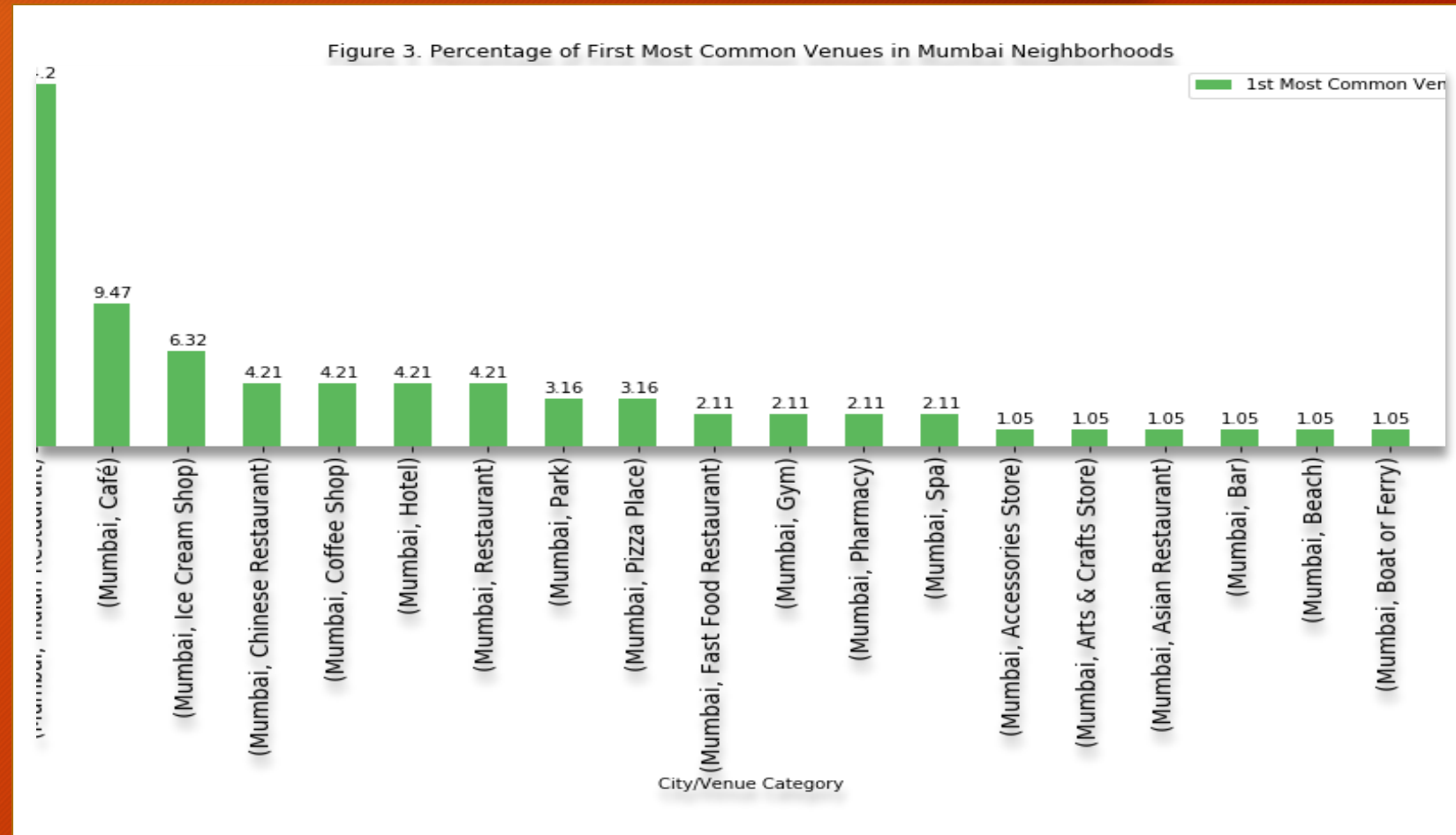
I used matplotlib's bar chart to visualise the above dataframe.



# Methodology

## B.2 Finding Relations between various individual neighbourhoods of Mumbai

The same process is repeated for Mumbai Neighbour hoods and the final visualization for Mumbai looks like this



# Methodology

## B.3 Locations where business can be setup in Delhi and Mumbai

Now I will try to find out locations where certain business can be set up. We define the criteria of selection based on the absence of that particular venue category in a particular neighborhood. We take locations where a particular venue category is not common. We undertake this process for 'Indian Restaurants' and 'Cafes ' as these are the Venue Categories with the highest number of Venues. I repeat the same process for Mumbai also. The same process can be repeated for any venue category.



# Results

For B.1 i:e comparison between two cities viz Delhi and Mumbai see (Figure 1) we can infer the following results:

1. There are more Indian Restaurants in both Delhi and Mumbai followed by Cafe as expected.
2. The highest among venue categories are restaurants which include Indian Restaurants, Restaurants, Fast Food Restaurant and Chinese Restaurant.
3. There are more Indian Restaurants in Mumbai than in Delhi.
4. There are more Cafe in Delhi than in Mumbai.
5. The hotels are equal in number in both cities.
6. Delhi being the capital has a greater number of Chinese Restaurants than Mumbai.
7. There are more Ice Cream Shops in Mumbai than in Delhi and that may be due to the Harsh Winters of Delhi.
8. There are more centres of Gym in Delhi than in Mumbai.
9. One most important assumption that we could make is that 'based on Venue Categories Delhi and Mumbai are a lot similar'.

# Results

For B.2 i.e comparison between Neighbourhoods of Delhi (see Figure 2) we can infer the following results:

1. There are about 15% Neighbourhoods in Delhi which have '*Indian Restaurants*' as their 1st Most common place.
2. Similarly about 8% of Neighbourhoods have '*Cafe*' as their 1st most common venue.
3. Other most important interpretation is that even though in **Figure 1** the third ranked venue category is '*Hotel*' but as we see in **Figure 2** the third ranked venue category is '*ATM*' with a percentage value of **7.33** which indicates that *ATM's* in Delhi Neighborhoods are located close to each other than that of *Hotels* being close to each other.
4. Also **pizza places** are close to each other than *Coffee Shops, Restaurants, Bakery, Market and Bars* being close to each other.
5. **Fast Food Restaurants** are widely distributed in *Delhi Neighborhoods*.



# Results

For B.2 i.e comparison between Neighbourhoods of Mumbai (see Figure 3) we can infer the following results:

1. There are about 24% Neighborhoods in Delhi which have 'Indian Restaurants' as their 1st Most common place.
2. Similarly about 9% of Neighborhoods have 'Cafe' as their 1st most common venue.
3. Other most important interpretation is that even though in Figure 1 the third ranked venue category is 'Hotel' but as we see in Figure 3 the third ranked venue category is *Ice Cream Shops* with a percentage value of 6.32 which indicates that Ice Cream Shops in Mumbai Neighborhoods are located too close to each other than that of hotels being close to each other.
4. **Coffee Shops** are almost evenly distributed in Mumbai Neighborhoods.
5. Also **Chinese Restaurants** are close to each other than *Hotels, Coffee Shops, Restaurants, Bakery, Market, Pizza Places and Bars*.
6. **Bars** are widely distributed in Mumbai Neighborhoods.
7. Also **Chinese Restaurants** in *Mumbai Neighborhoods* are in close proximity than that in *Delhi Neighborhoods*.



## D. Discussions

The results drawn are just assumptions that I'm making about the sample set of data. I made sure that the Four Square API returned almost same number of venues for both Delhi and Mumbai. Later I made my assumptions about the categories of venues in the neighbourhoods.

The results may vary based on the time you try to access the API as it constantly updates the venues of the areas. One other option that is worth a try is to analyse by increasing the radius passed to the API.

I will try to make the data available on Wikipedia in the form of a table which can be scrapped directly to access the data. As of now the dataset is available on my Github Repository available for downloading.

## E. Conclusions

One thing that is lacking about the said cities is a proper dataset about the population. The government is supposed to make it available in the future. With more data about the population we can make good assumptions regarding the businesses.

I conclude with the findings that I made and just hope that the information provided is of some help to the people or business.





# Thank You!

Compiled by Zubair Khaliq