### Capstone Project – Exploring Bengaluru

### Introduction / Business Problem

### Bengaluru - the capital city of Karnataka, India. This bustling city is known as the Silicon Valley of India. It’s also the start-up capital of India. Hundreds of people from all over the world come to Bengaluru to make their careers, to build their life. Hence this city hosts all kinds of people be it unskilled or skilled like Software professionals, Government officials, Diplomats, Administrators, Students and so on. It has a population over 12.34 million and spread across 709 square kilometres. As a resident of this city, I have decided to use Bengaluru in my project. Through this project I will explore this beautiful city so that the new comers can decide which area/neighbourhood they would want to stay in or figure out the best places to eat/shop.

### The Data Section

* I made use of a python library called ***pgeocode* (**[**https://github.com/symerio/pgeocode**](https://github.com/symerio/pgeocode)**)** which is a high-performance off-line querying of GPS coordinates, region name and municipality name from postal codes. Distances between postal codes as well as general distance queries are also supported. The used GeoNames database includes postal codes for 83 countries.

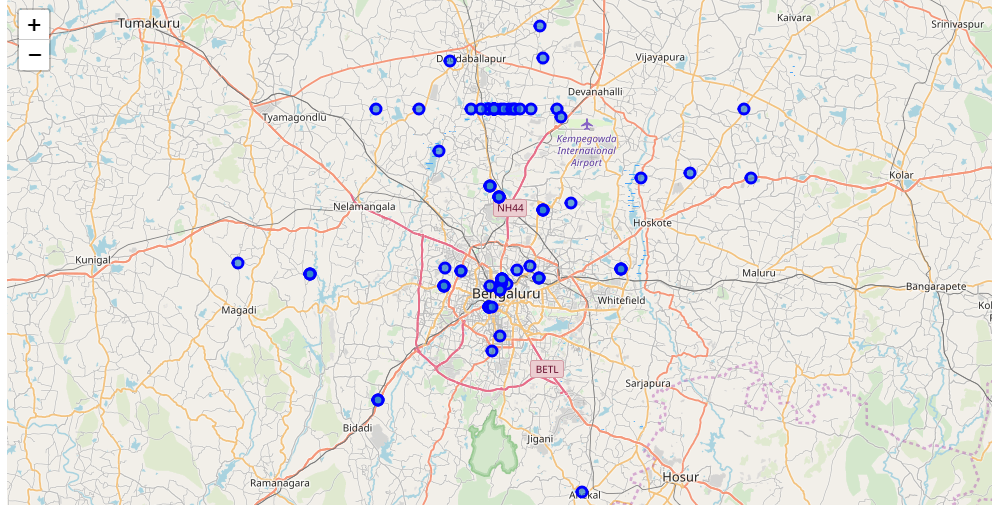
I used '*index* postal\_codes' function which creates a data frame of unique postal codes of a given country. The data frame consists of following columns:

* + country code: iso country code, 2 characters
  + postal code: postal code
  + place name: place name (e.g. town, city etc)
  + state\_name: 1. order subdivision (state)
  + state\_code: 1. order subdivision (state)
  + county\_name: 2. order subdivision (county/province)
  + county\_code: 2. order subdivision (county/province)
  + community\_name: 3. order subdivision (community)
  + community\_code: 3. order subdivision (community)
  + latitude: estimated latitude (wgs84)
  + longitude: estimated longitude (wgs84)
  + accuracy: accuracy of lat/lng from 1=estimated to 6=centroid
* Using this function, I first I obtained the postal code details for the entire country India and then queried the data to retain the postal codes of Bengaluru. The sample data is shown below:

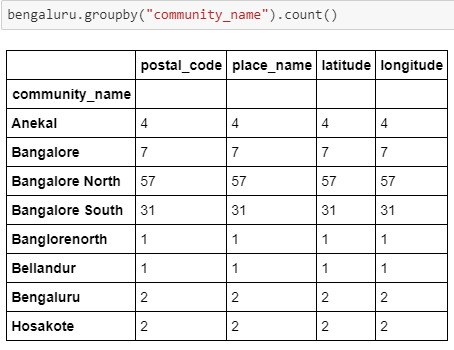


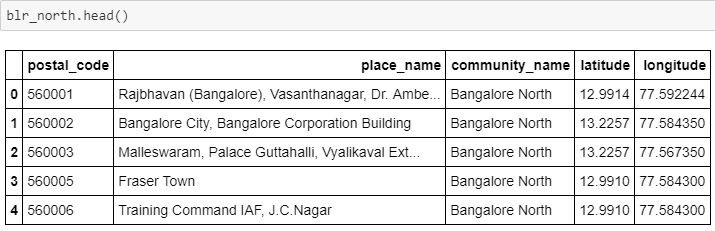
### Methodology

* I used **Forsquare API** to get the most common venues of given community of Bengaluru.
* I used python **folium** library to visualize geographic details of Bengaluru and its places and I created a map of Bengaluru as shown below:



* As Bengaluru North had highest number of places, I decided to just focus on Bengaluru North alone





* Retrieved all the venue details from Foursquare API and

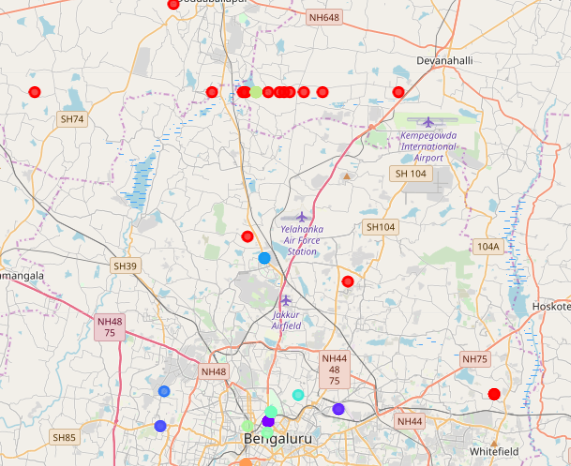


### Results

* In summary, **50** unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each neighbourhood:



* We have some common venue categories in neighbourhoods. For that reason I used unsupervised learning **K-means algorithm** to cluster the neighbourhoods. K-Means algorithm is one of the most common cluster methods of unsupervised learning.



* I ended the study by visualizing the data and clustering information on the Bengaluru map.

### Conclusion

People can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.