'Pubs & Breweries' near Top 10 Software (IT) Parks in Bangalore, India.

1. Introduction

1.1 Background

Bangalore, officially known as Bengaluru, is the capital city of the Indian state of Karnataka. It has a population of over ten million, making it a megacity & is third most populous city in India. Bangalore is sometimes referred to as the “Silicon Valley of India” (or “IT capital of India”) because of its role as the nation’s leading information technology (IT) exporter.

Bangalore has been high on the list of young talented IT professionals from across India as it offers a decent standard of living and good job opportunities. Young millennials & recent graduates often move to Bangalore to find employment. Due to this reason, the ‘Nightlife’ culture, of this otherwise traditional southern Indian city, is booming. After a day of hard work, Bangalorians may prefer grabbing a drink at a nearby Pub/Bar. We try to explore the nearby locations of top 10 Bangalore IT Hubs and see where they may head to, for relaxing & having a good time.

2. Data acquisition and cleaning

2.1 Data sources

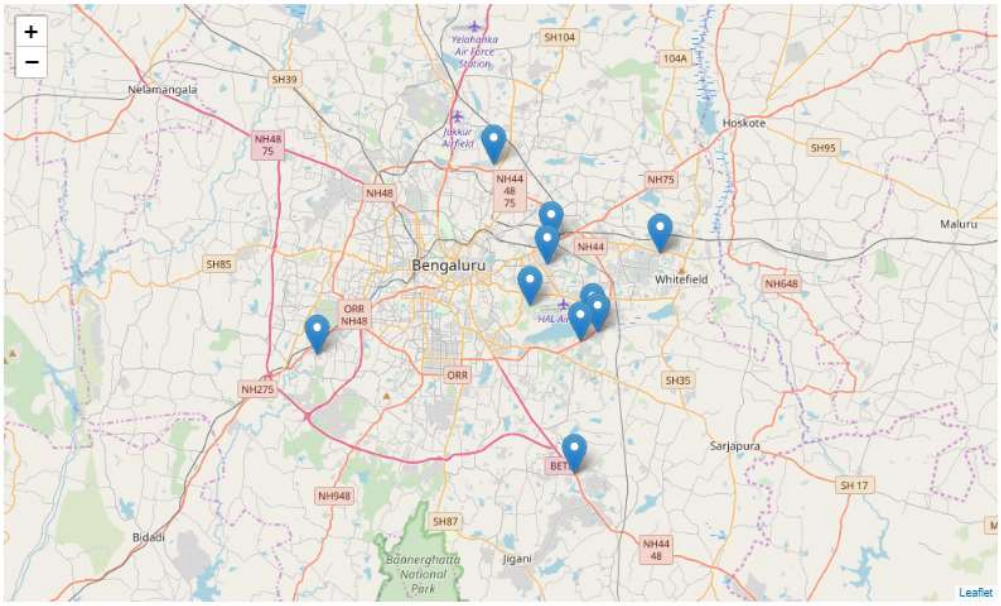
Data obtained from the following links:
http://www.walkthroughindia.com/walkthroughs/10-amazing-software-technological-parks-in-bengaluru/
https://www.tripadvisor.in/Tourism-g297628-Bengaluru_Bangalore_District_Karnataka-Vacations.html#photos;aggregationId=&albumid=101&filter=7
Used FourSquare to obtain the nearby venues near the Tech Parks.

2.2 Data Cleaning

Performed Web-scraping using Beautiful Soup python library. Filtered the required venues obtained from FourSquare.

3. Exploratory Data Analysis

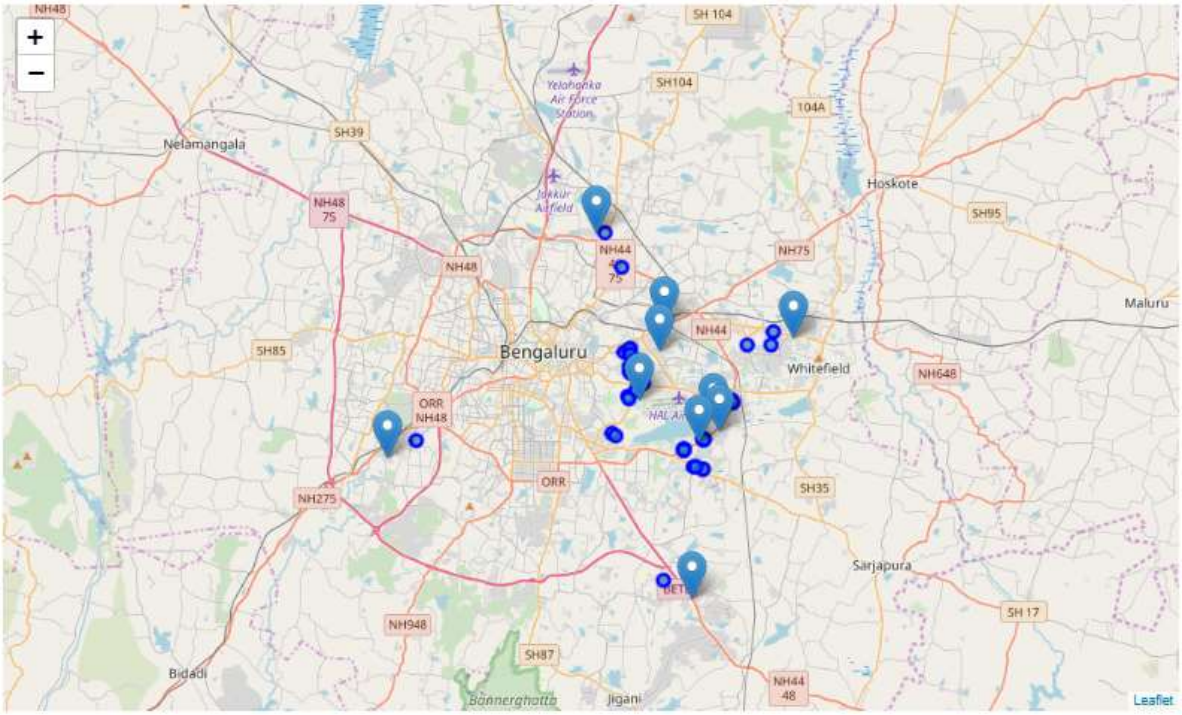
Once the coordinates of these 10 IT Hubs were obtained, attempted to plot these Hubs on a map to understand there locations of the Map of Bangalore.



Then used FourSquare using the Client ID and Client Secret created on the FourSquare website using Developer account to obtain the nearby venues to these 10 IT Parks. I was interested to know the nearby venues within walking distance. Assumed that 2 - 3 kms walking distance should be ok hence took the radius as 3kms. Found there were various venues some showing label as Bar some showing cocktail bar etc so added all of them to the venues list.
Filtered the data and created a table with the Number of Pubs / Breweries etc within 3 km of the IT Hubs.

Name of Tech Park	Longitude	Latitude
Electronic City	77.6770	12.8399
Bagmane Tech Park	77.6581	12.9785
Manyata Tech Park	77.6221	13.0448
International Tech Park	77.7354	12.9858
Prestige Tech Park	77.6889	12.9398
RMZ Infinity	77.6611	12.9941
Ecospace Business Park	77.6810	12.9275
Global Village Tech Park	77.5014	12.9188
Embassy GolfLinks Business Park	77.6465	12.9513
Embassy TechVillage	77.6925	12.9339

Plotted the Venue data on the Map along with the location of the IT Hubs to get a visual representation of the data.



4. ML practice model used – K Means Clustering.

Practicing Machine Learning. Considered that we have a set of data (List of 'All Venues' from above) and we need to Cluster them in groups.

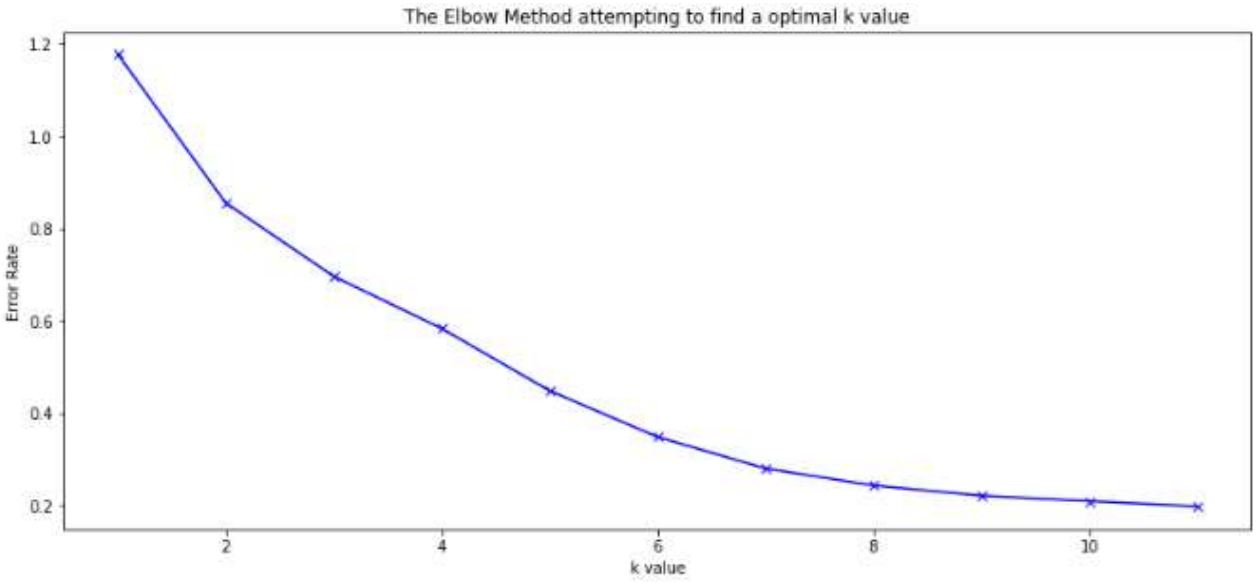
Assumption: Lets assume we have a list of Venues (Venue Coordinates) and we are not sure where they are located or which venues are located Nearby. Plotted the data for all the venues to get an understanding and visually see the data spread.

I will attempt to use K Means Clustering to cluster these Venues together in Groups and attempt to figure out the best k value or the number of clusters it can be divided into.

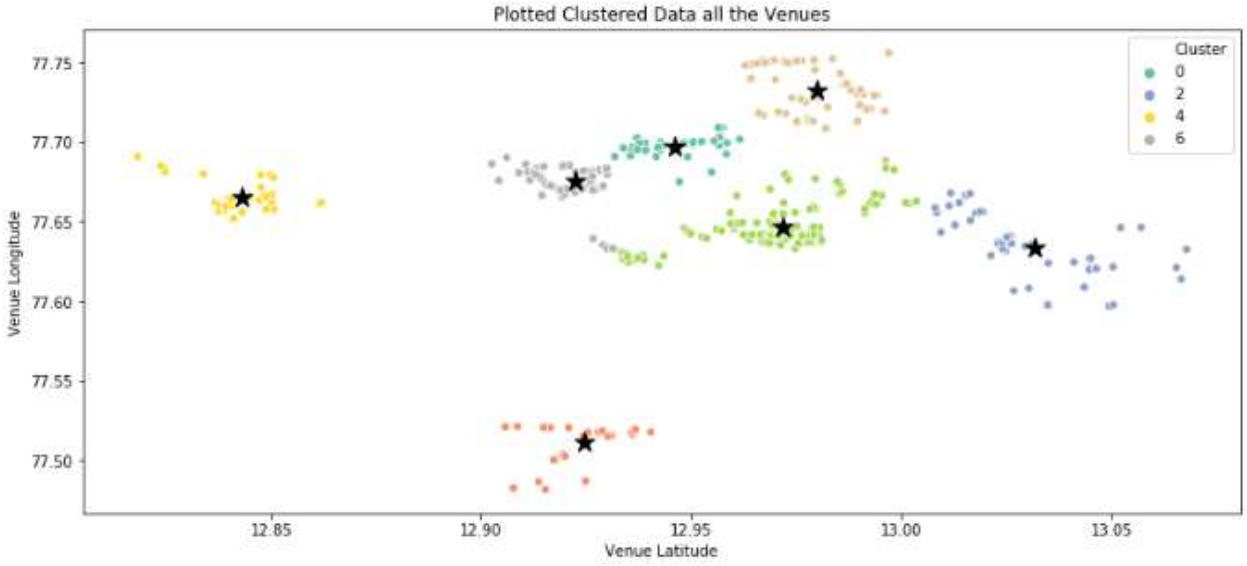
Calculated the Elbow point to check what values are obtained for k in range 1 to 12.

seems k = 7 could be an acceptable value for k.

Then attempting to cluster the data and plotting the Venue data with Tech Hub location data on map of Bangalore.



Venue data & Tech Hub location data plotted on map of Bangalore.



5. Conclusion

I noticed that within 3km Radius of these 10 IT Parks in Bangalore, 'Embassy GolfLinks Business Park' has most options available to hang out with friends post work (for grabing a beer etc) while 'Global Village Tech Park' & 'Electronic City' has least no of options available within the 3km Radius.
Additionally, if we are given a set of venues, then we can use K Means Clustering to cluster these venues together.
