**Battle of Cities**

**Sastry Akundy**

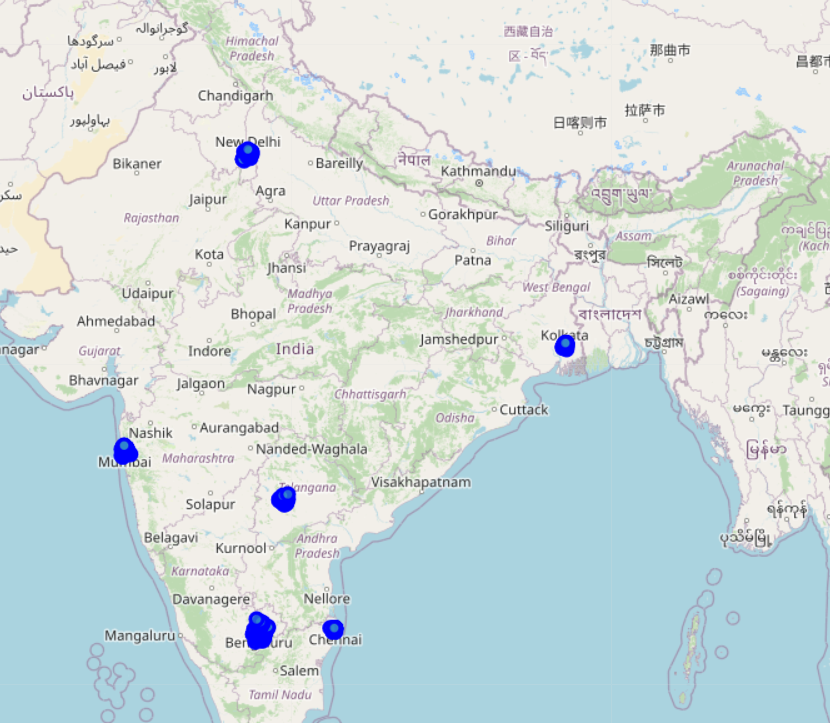
19th July 2020

**Problem Statement:**

The culture, language and cuisine significantly varies across India. Employees in India hop jobs and move across cities very frequently. The major problem people face in doing so is to find a neighborhood which similar to what they were used to. If someone has an option to chose among cities to move, it is important to identify which city would provide more favourable locality.

**Introduction:**

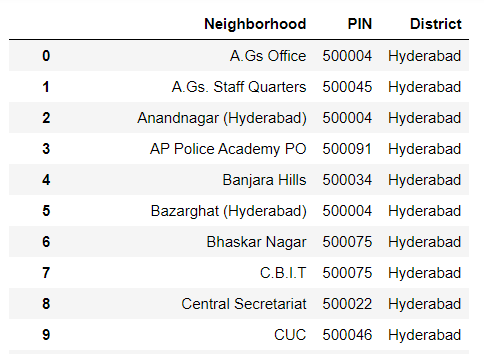
The major cities of India are spread across and are major source of jobs. Employees who are in their early careers frequently shift jobs for quick career growth. But a major concern for people while moving is to find a neighborhood which is similar to their existing ones. If the new company provides an option to pick among cities other that he currently resides in, he/she would need the information that provides which city is favourable. Six major cities of India are shown in Figure 1.



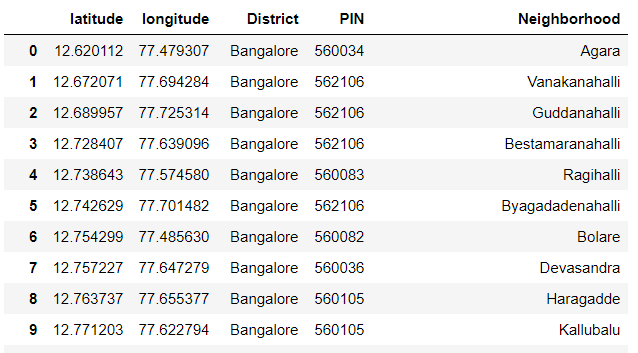
*Figure 1: Major Indian Cities*

**Data:**

Geographical location data is collected from All India post office list. After dropping few unnecessary columns the data for six major cities namely Delhi, Mumbai, Kolkata, Hyderabad, Bangalore and Chennai is as shown in Figure 2.



*Figure 2: Data from All India Post Office list*

The latitude and longitude data for each location is obtained from geopy and the data is grouped by latitude and longitude data for overlapping locations. Cleaned up data is as shown in Figure 3.



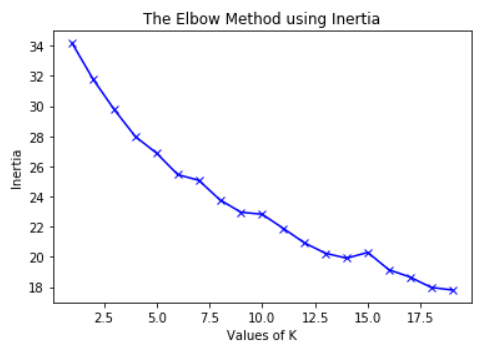
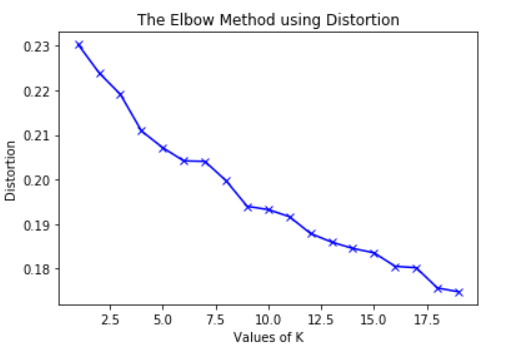
*Figure 3: Cleaned grouped data and Sample Foursquare data*

**Clustering:**

Data of Delhi, Mumbai and Bangalore is considered to evaluate if a person has to move from one of those cities to any of the other two, which one he can chose based one his current location.

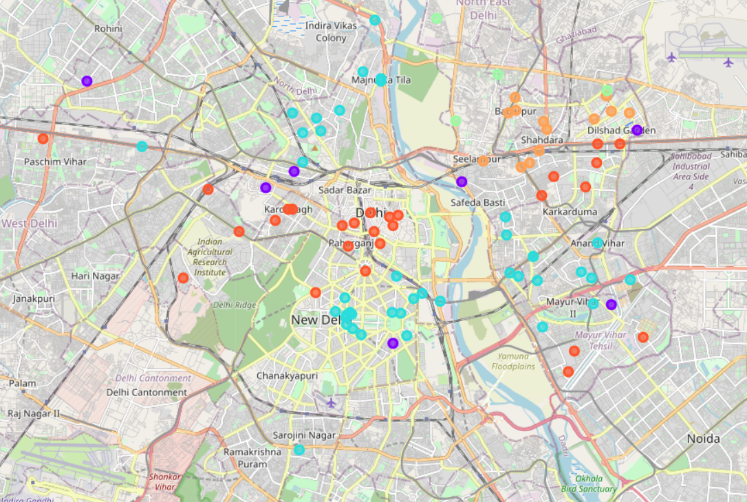
The neighborhoods are converted to clusters using the k-means algorithm. The Elbow method is used to estimate the optimum number of clusters into which the data is clustered. There are two methods to estimate the optimum number of clusters using the Elbow method.

1. Distortion:  It is calculated as the average of the squared distances from the cluster centers of the respective clusters. Figure 4 shows the result of the distortion method. Though not clearly indicative, K appears to have slight elbow at 10
2. Inertia: It is the sum of squared distances of samples to their closest cluster center. Figure 5 shows the result of the inertia method. Both 10 and 15 appears to show change in direction. So, 10 is chosen as optimal K.

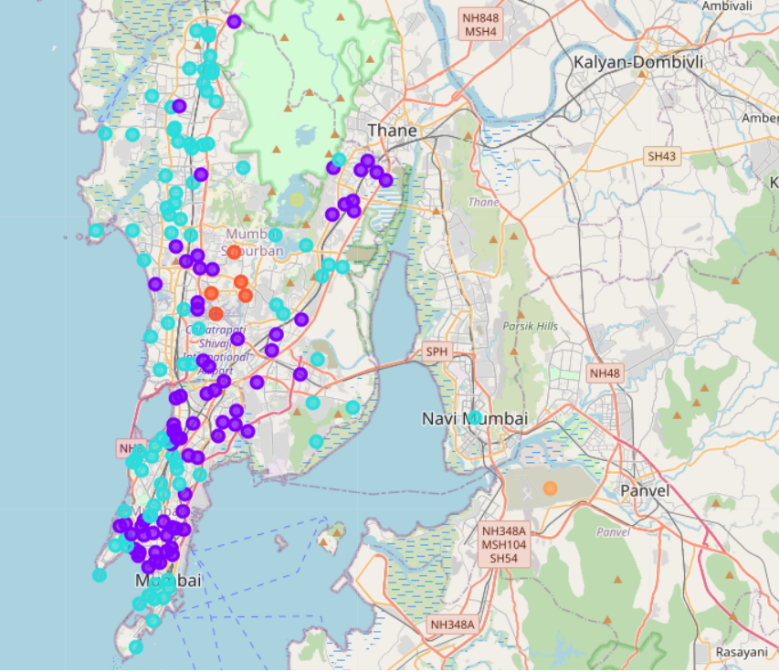


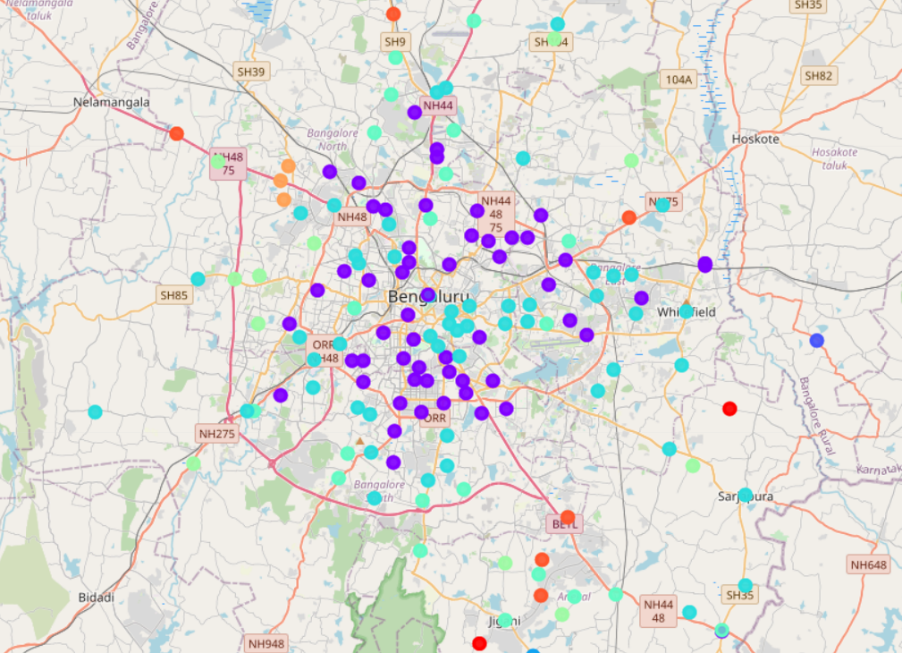
*Figure 4: Distortion method Figure 5: Inertia method*

**Results & Discussion:**

The clustered maps for Delhi, Mumbai and Bangalore are in figures 6, 7 and 8 respectively.

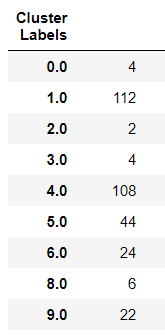
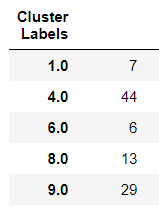
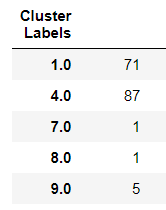
*Figure 6: Delhi Cluster Map*

 *Figure 7: Mumbai Cluster Map*



*Figure 8: Bangalore Cluster Map*

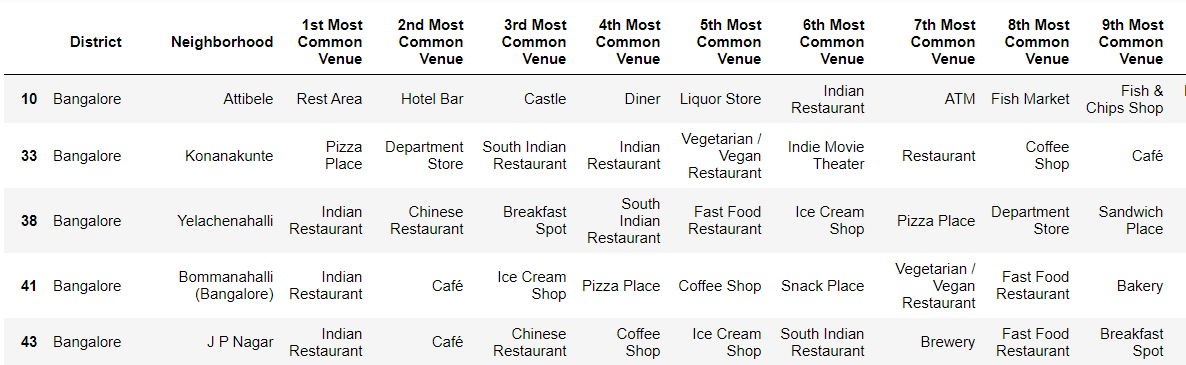
The density of clusters for Delhi, Mumbai and Bangalore are in figures 9, 10 and 11 respectively.

From the cluster density charts, it is evident that Bangalore has more unique clusters compared to Mumbai and Delhi. Cluster 4 is a major common selection among all the cities. Cluster 1, which has high density in Mumbai and Bangalore is sparse in Delhi. Sample Cluster 1 data obtained from Foursquare is shown in Figure 12.

*Figure 10: Mumbai Cluster density*

*Figure 11: Bangalore Cluster density*

*Figure 9: Delhi Cluster density*



*Figure 12: Cluster 1 Foursquare data*

**Conclusion:**

To conclude the results from this study, consider a case where a person from Bangalore has to move to either Delhi or Mumbai. He/She can select a favourable location based on the cluster that he/she is currently located. If the current cluster is 4, either both Delhi and Mumbai seems favourable, but, if current cluster is 1, Mumbai would be the preferred option.

But, if the current cluster is one of the unique clusters of Bangalore, it would be difficult to identify a suitable place in either Mumbai or Delhi and further detailed study is required to select a suitable location.