IBM Capstone Project - The Battle of Neighborhoods

Exploring neighborhoods in Pune, Maharashtra

By Nandhitha .V

Business Problem

- I have recently got transferred to Pune for work. Which neighborhoods had got vegetarian restaurants? Which neighborhoods have shopping malls and theatres? Which neighborhoods are great for a coffee? Which neighborhoods are famous for its markets? Where are the ATMs? were the questions running on my mind as a new resident.

- So, in this capstone project, I am going to find a good neighborhood location in Pune to stay in by using data science methods and algorithms like clustering.

Data

- The required data to do this project is as follows:
- Latitude and Longitude of neighborhood
- List of neighborhoods in Pune, Maharashtra
- Venue data(restaurants, ATM, theatres) of neighborhoods

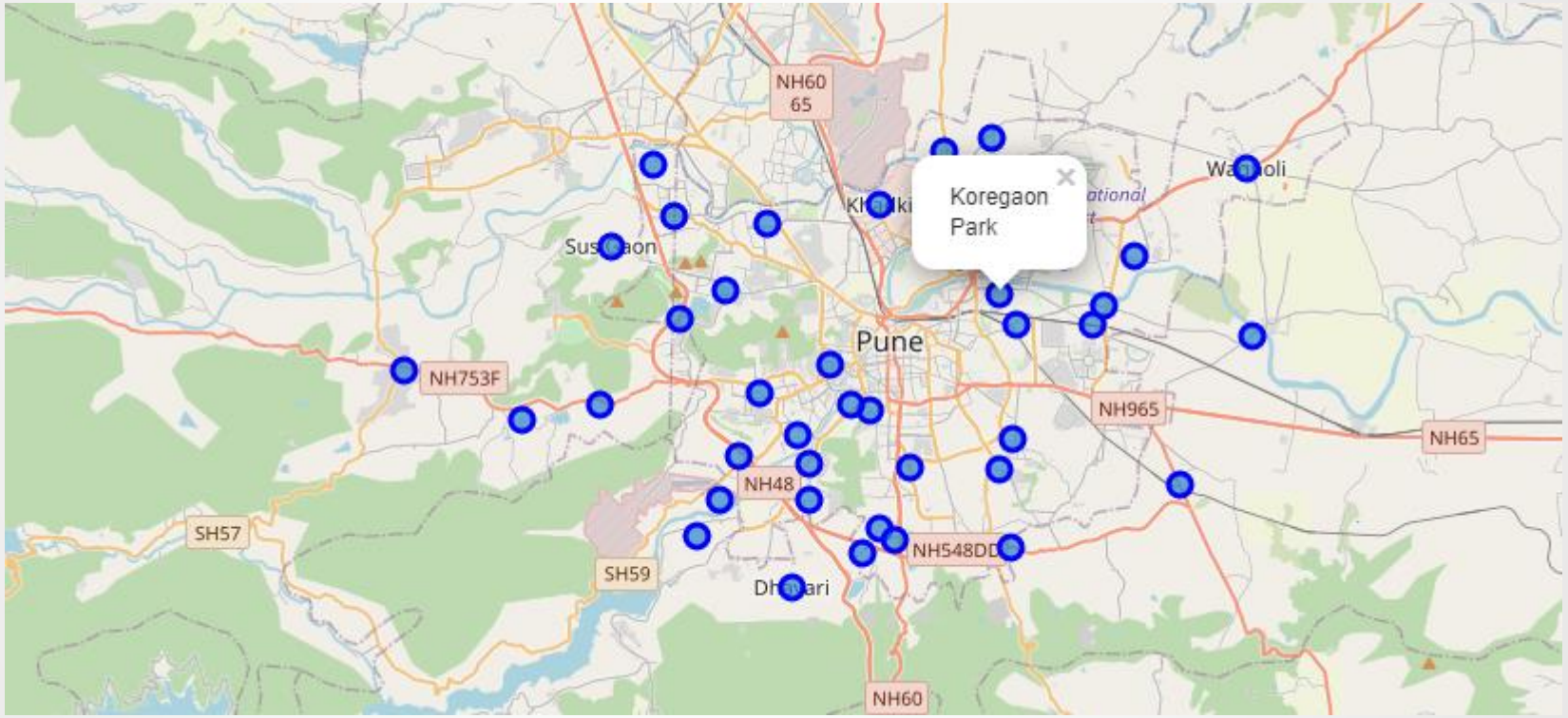
- **Data sources**
- Following data sources will be needed to extract the required information:
- Web scraping data from Wikipedia to get information on Pune neighborhoods using Beautiful soup.

  The link to the data is : 'https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Pune'
- Generating Longitude and Longitude coordinates of Pune as well the neighborhoods via geocoder, geopy package.
- Using Foursquare API to create a map of Pune and also to generate the venue data related to the neighborhood.

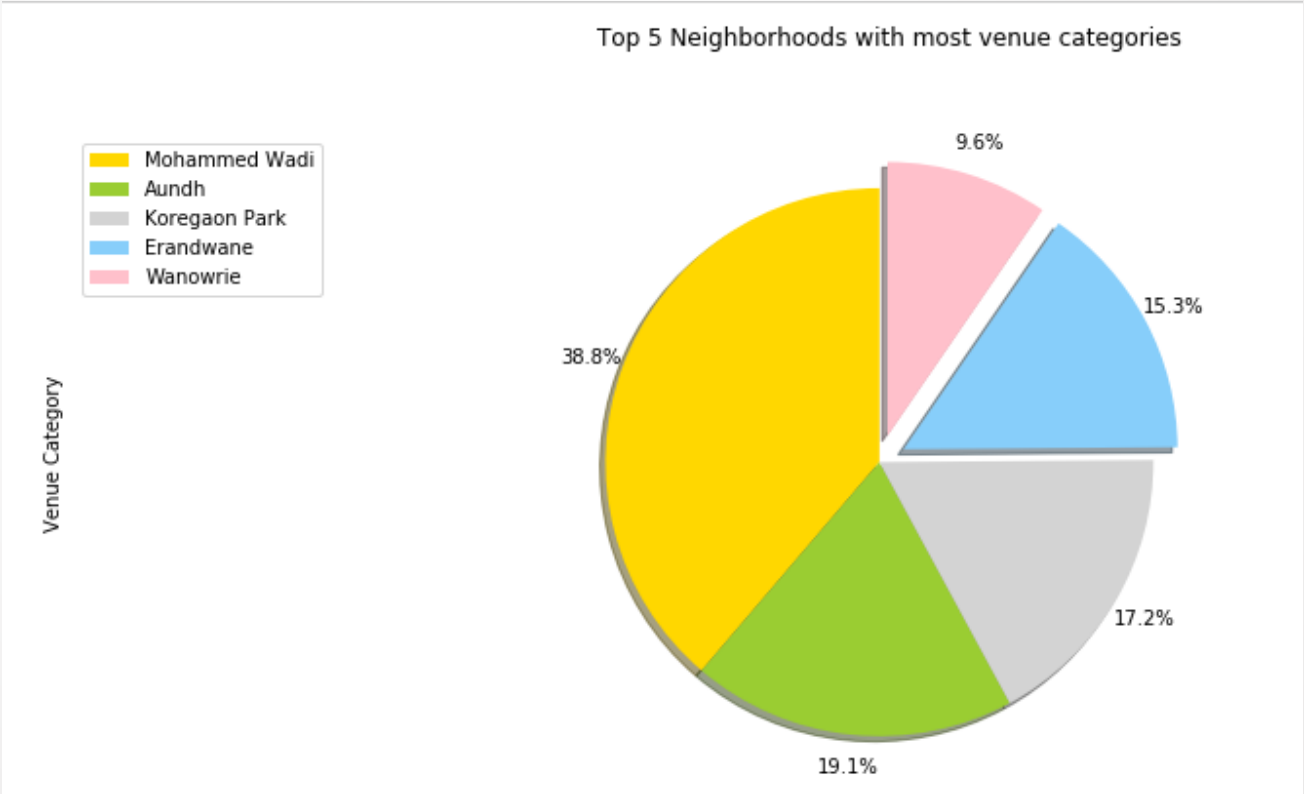
Methodology

- Importing packages
- Extracting data from Wikipedia by Web scraping
- Data cleaning
- Finding latitude and longitude
- Finding nearby venues using Foursquare API
- Generating the neighborhoods most common places
- Clustering on the data using K-Means algorithm
- Examining the Clusters

Map of Pune before Clustering

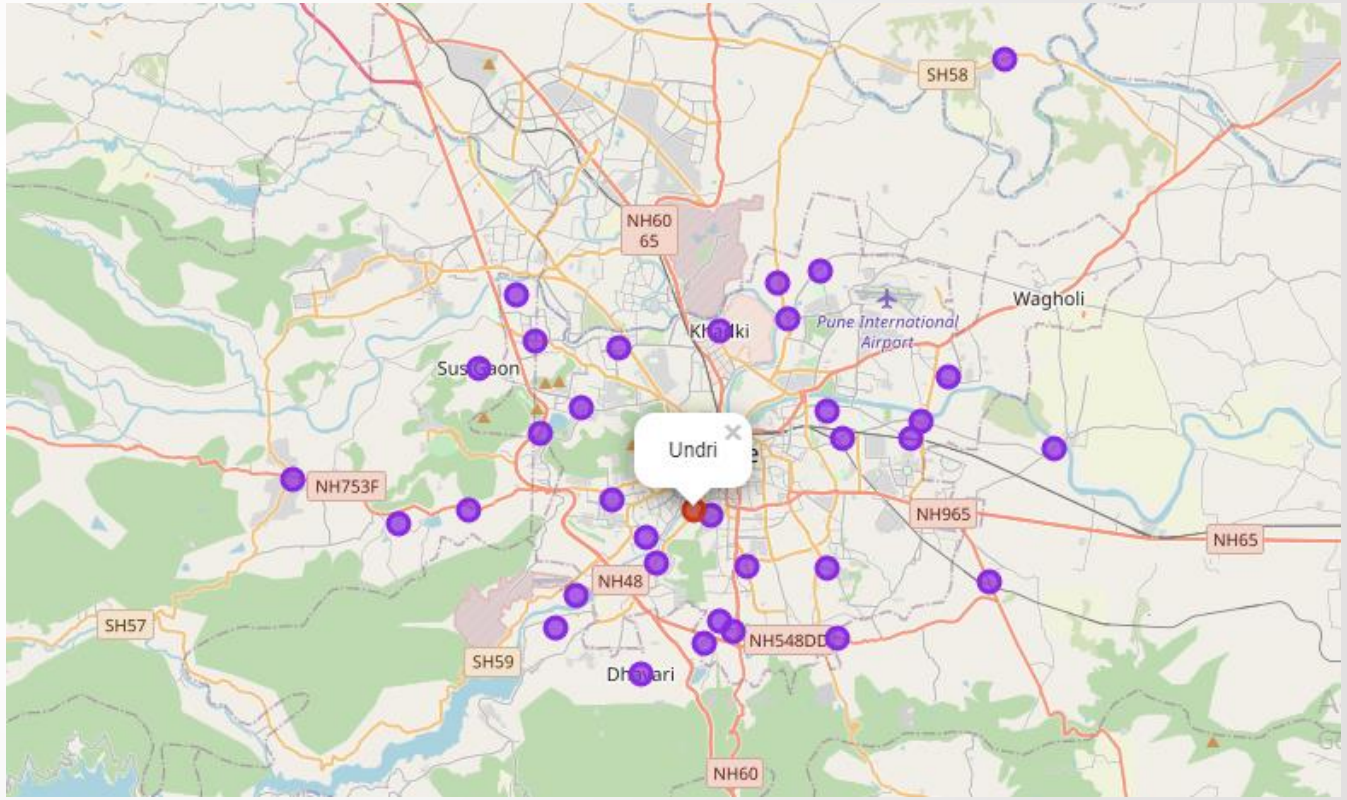


Top 5 Neighborhoods with most venue categories is depicted using a pie chart.



Map of Pune after Clustering

```
kclusters = 2
clus = gp.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(clus)
labels=kmeans.labels_
kmeans.labels_[0:10]
pune_clus['Label']=labels
#pune_clus.insert(1,"Label",labels)
```



Results

- It is found that there are
  - 41 unique neighborhoods
  - 393 unique venues
  - 111 unique venue categories
- The k-means clustering was performed with k=2.The data was divided into two clusters with label-0 and label-1.The total number of neighborhoods were 41.
- Cluster 0 - All the neighborhoods in the cluster have diverse venues mostly restaurants, coffee shops, malls, department stores etc. as the first two common venues.
- Cluster 1 - The neighborhood in this cluster have other venues like furniture store, zoo etc.
- On examining the clusters, it is clear that 40 neighborhoods are similar and thus it falls under the same cluster with the label-0.
- The only neighborhoods which is different from the other are ' Undri ' and thus it is falls into another cluster with label-1.

Discussion

- In this project, Analysis of Pune neighborhoods recommendations based on venue categories like restaurants, mall, coffee shops etc. has been presented. This will be a great recommendation for visitors like me who are new to the place to find out nearby venues of interest and in deciding a place to stay in.

- The generated results shows the 2 clusters associated with the neighborhoods. It is evident that Cluster 0 is the most representative of the Pune city. Just looking at this cluster, it shows there are 80% of restaurants, grocery stores, coffee shops, electronic stores which will make life easier if we stay in this neighborhood.

- On the other hand, as we observe Cluster 1, the percentage is less than 40% .There are not many restaurants in these area. But this place is has got other spots like Furniture store, Zoo etc.

Conclusion

- Using the combination of data from the URL and the foursquare API, we were able to collect, clean, analyse, discover and examine the venues of the neighborhoods in Pune to find the best neighborhood to live in.

- Since all the neighborhoods were similar with respect to the most common venues, and also likely to live , I would find a good place to stay in any of it other than the neighborhood 'Undri' which didn't seem to be a residential and accessible place.

Thank You