Analyzing the neighbourhoods of Coimbatore for starting a new restaurant

~ Jayasankar S

IBM Applied Data Science Capstone

19-08-2020

INTRODUCTION

- This project deals with the major venue categories in the neighborhoods of **Coimbatore, The Manchester of South India**.
- This project would specifically help business personal plan to start new Restaurants in Coimbatore, Tamil Nadu, India.
- Coimbatore is a metropolitan city that has around 36 neighbourhoods

BUSINESS PROBLEM

- To start a restaurant where there are very less competition.
- To find a neighbourhood that is most likely to give a good business.

DATA

- Neighbourhoods of Coimbatore – Obtained through Data scraping from the Neighbourhoods of Coimbatore Wikipedia page.
- Geographical coordinates of the neighbourhoods – Obtained using GeoPy library.
- Venue data from Foursquare – Using FourSquare API

METHODOLOGY

- Feature Extraction – One Hot Encoding

```
cov_1hot = pd.get_dummies(explore_cov[['Venue Category']], prefix="", prefix_sep="")


# Add neighbourhood column back to dataframe
cov_1hot['Neighbourhood'] = explore_cov['Neighbourhood']


# Move neighbourhood column to the first column
fixed_columns = [cov_1hot.columns[-1]] + cov_1hot.columns[:-1].values.tolist()
cov_1hot = cov_1hot[fixed_columns]

cov_1hot.head()
```

METHDOLOGY

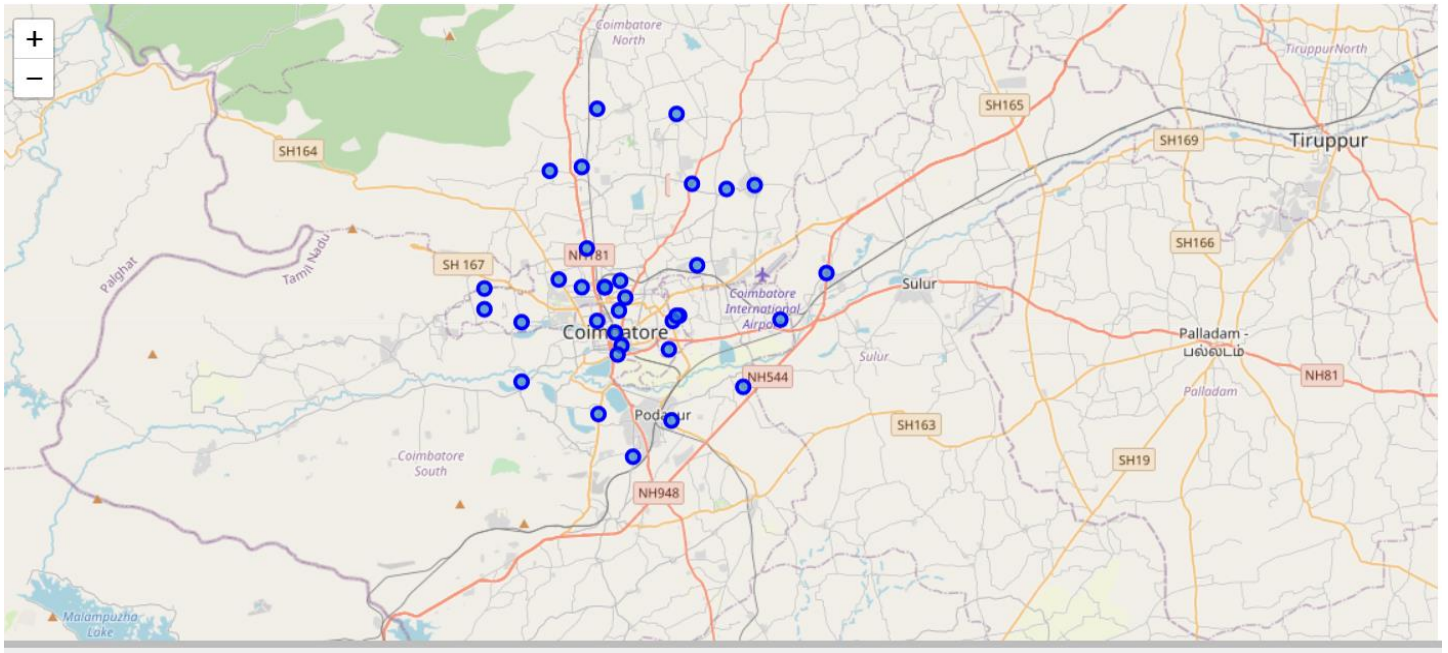
- Unsupervised Learning – K means clustering

```
max_range = 15 #Max range 15 (number of clusters)

from sklearn.metrics import silhouette_samples, silhouette_score


indices = []
scores = []

for cov_clusters in range(2, max_range) :

    # Run k-means clustering
    cov_gc = cov_grouped_clustering
    kmeans = KMeans(n_clusters = cov_clusters, init = 'k-means++', random_state = 0).fit_predict(cov_gc)

    # Gets the score for the clustering operation performed
    score = silhouette_score(cov_gc, kmeans)

    # Appending the index and score to the respective lists
    indices.append(cov_clusters)
    scores.append(score)
```

```
plot(max_range, scores, "No. of clusters", "Silhouette Score")
```

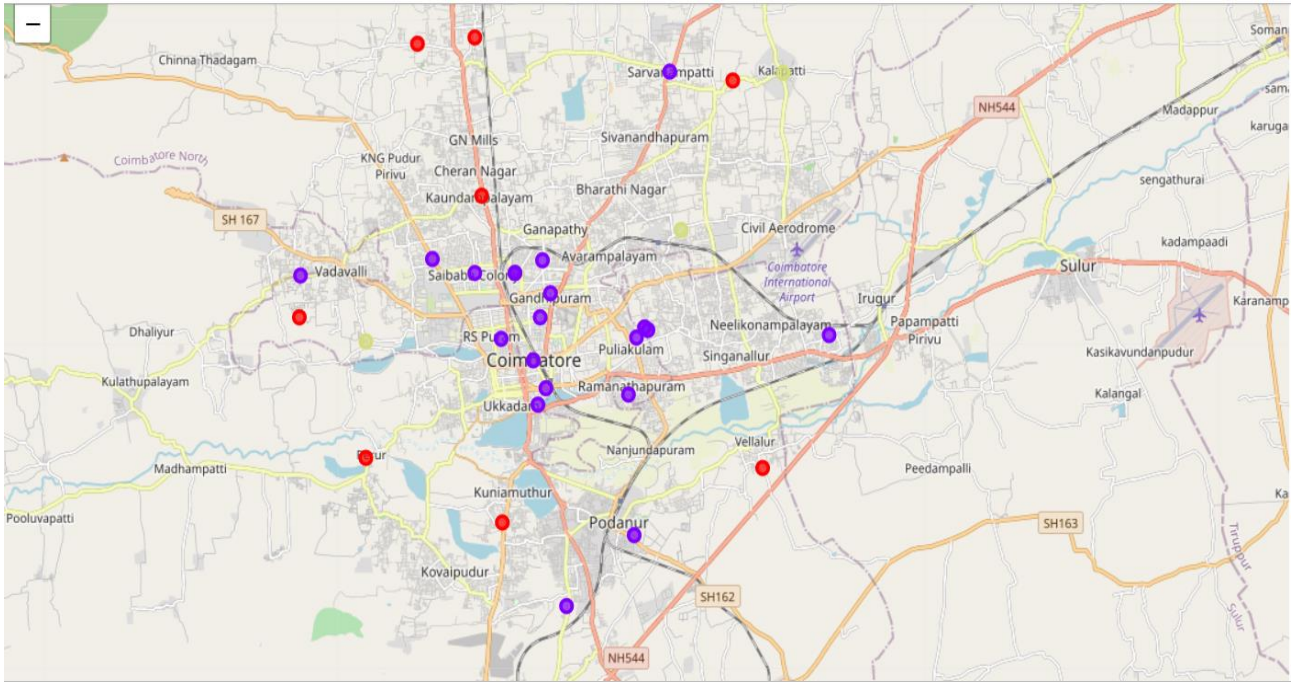
METHODOLOGY

- Plotting – Using Folium



RESULT

- Visualization of clusters



- After visualization of clusters, it was concluded that cluster 1 would be suitable for starting a restaurant.

DISCUSSION

- The most suitable neighbourhood for starting a new restaurant are in the cluster 1, as there are very less competition in those areas.
- The K-means model worked well in clustering similar neighbourhoods together.
- The neighbourhoods such as Ganapathy, Perur, Vilankurichi can be chosen for starting new restautrants.