COURSERA

IBM APPLIED DATA SCIENCE CAPSTONE

Opening a new Café in Delhi, India

Made By:- Harshit Singhal

From:- India

INTRODUCTION

- Café these days have become one of the most common place to hang out with friends and family.
- In the city like Delhi having a population size of near about 1.9 crores (as of 2012) it provide a great opportunity to open a new café in such great city.
- The objective of this capstone project is to analyse and select the best locations that are available in the Delhi, India to open a new café.
- All the builders, businessman, café chain owners or some great future business tycoon or some entrepreneurs would be interested in knowing the location in the city like Delhi, India which is densely populated and have a great scope to open café and flourish them in near future.

DATA

- List of neighbourhoods in Delhi. This defines the scope of this project which is confined to the city of Delhi.
- https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi
- For the sake of geographical coordinates of the neighbourhoods we will be using the geopy library.
- For the the purpose of exploring the neighbourhoods we will be using the foursquare API.

METHODOLOGY

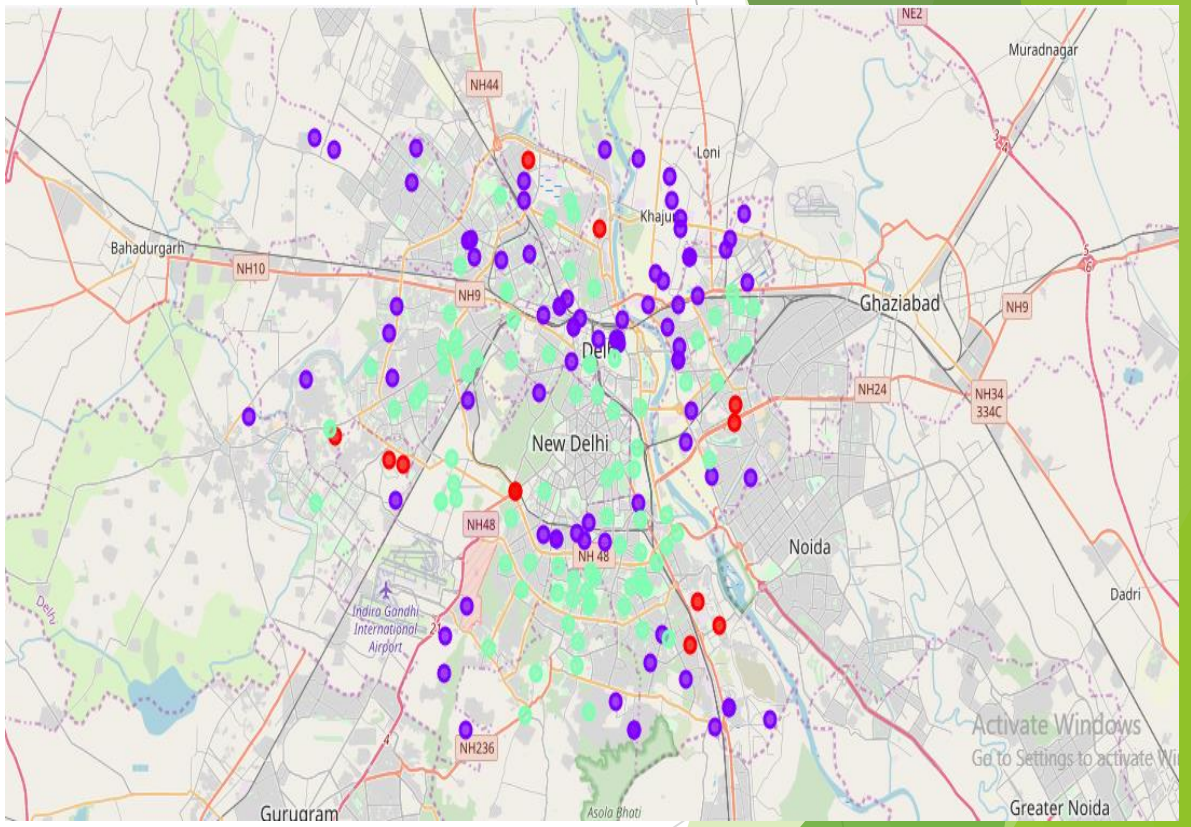
- Scrap the data of neighbourhood from Wikipedia
- Gathered geographical coordinate from Geopy
- Used folium to visualize the data on map.
- Used foursquare API to explore the locality
- Analysis of the venue data and selected the mean of occurrence of venue type grouped by neighbourhood as feature for clustering
- Used K-Means for clustering data in three clusters
- Presented the cluster on the map

RESULTS

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Café”:

- Cluster 0: Neighbourhoods with maximum number of café
- Cluster 1: Neighbourhoods with low number to no existence of café
- Cluster 2: Neighbourhoods with moderate concentration of café

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.



DISCUSSION

- From the above results we can see that the locality having high number of café are located at very faraway places.
- Cluster 2 have moderate number of café so opening a café in such location would be beneficial if you have a uniqueness in your business model or in the interior/decoration of café.
- Cluster 3 have low to negligible café so opening café in such location would lead to fast growth and neither you have to go through high competition of market.

CONCLUSION

For this project we have gone through the business problem and have solved it by going through the process of data collection, data cleaning, data analysis, data preparation, data visualization, choosing and training the model and finally giving the results by categorising each location in three different cluster based on the frequency of café in the locality. Now answering the question it is best to open the café in locations which fall under cluster 1 as these location have minimal or no café in their locality.

THANK YOU