**IBM Data Science Capstone Project**

***Analysing***

***Neighbourhoods In***

***Delhi, India***

***To Start***

***A New***

***Restaurant***

***Suleman Salmani***

***11-09-2020***

**Table of Content**

Contents

Introduction

Business Problem

Data

*Neighbourhoods data*

*Geographical coordinates*

*Venues data from Foursquare API*

Methodology

*Feature Extraction*

*Plotting the Data on Map*

*Unsupervised Learning*

Results

Discussion

Conclusion

**Introduction**

Delhi is one of the biggest and most populated cities in India. Delhi has population of 3 crores that makes it most populated city.

The population in Delhi comprises of the people of various ethnicities from all over the country and world. The city is full of restaurant serving thousands of hungry people, every day. The diversity in population of city has brought in a vast diversity in food habits of people.

**Business Problem**

Our Client is interested to open a new restaurant in Delhi, India. They have approached us to study the market and suggest them a location in one of the neighbourhoods that would be a best business interest. Our main objective would be to study and analyse the right data using various data science techniques and suggest to our client a best place to start a restaurant.

**Data**

In order to achieve our final goal, we’ll have required the following data

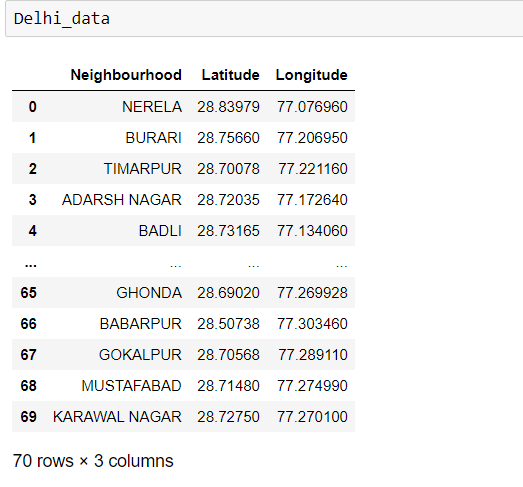
* Neighbourhoods of Delhi, India
* Geographical Coordinates of Neighbourhoods
* Venues Data from Foursquare API

**Neighbourhoods of Delhi, India**

This Data was Extracted from [here](https://ceodelhi.gov.in/Content/EntireDelhiLocalities.aspx). Here I have found all the Neighbourhoods of Delhi.

**Geographical Coordinates**

The Geographical coordinates are extracting using Geopy library in Python. These coordinates help to visualize neighbourhoods on map.



**Venue data from Foursquare API**

After that we have extracted venue data from the foursquare API. This data was used to study various venues in Neighbourhoods. This data helps us to understand the competition in the market and provide important details about the restaurants. This data also talks about the category of a venue.

**Methodology**

**Feature Extraction**

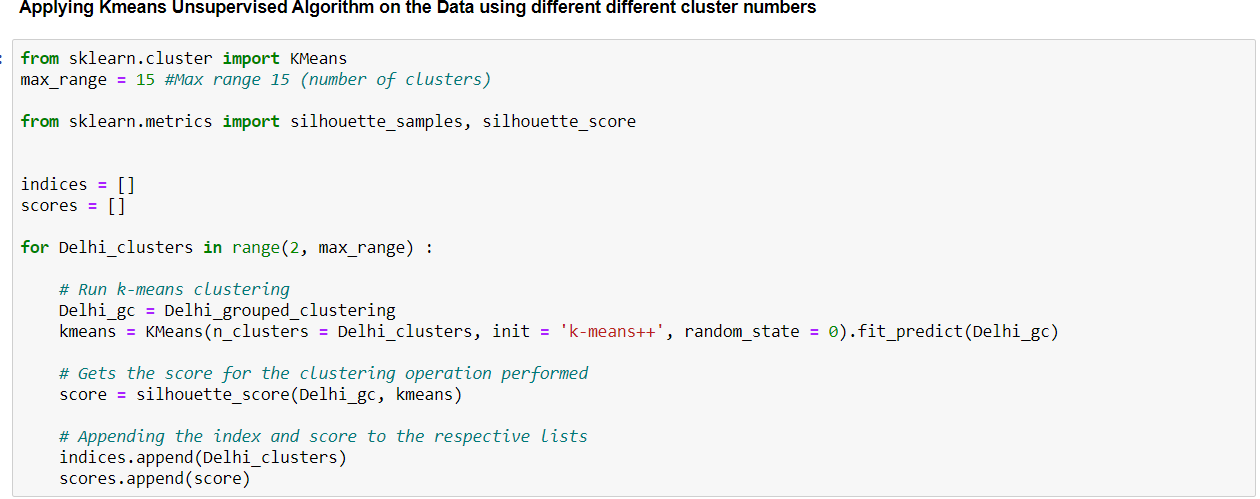
Here one hot encoding was performed and created a sparse matrix to make the category of venue to a feature of the dataframe and first column of that DataFrame was neighbourhood so that if any neighbourhood has that category of venue then the value of that category will be one and opposite.

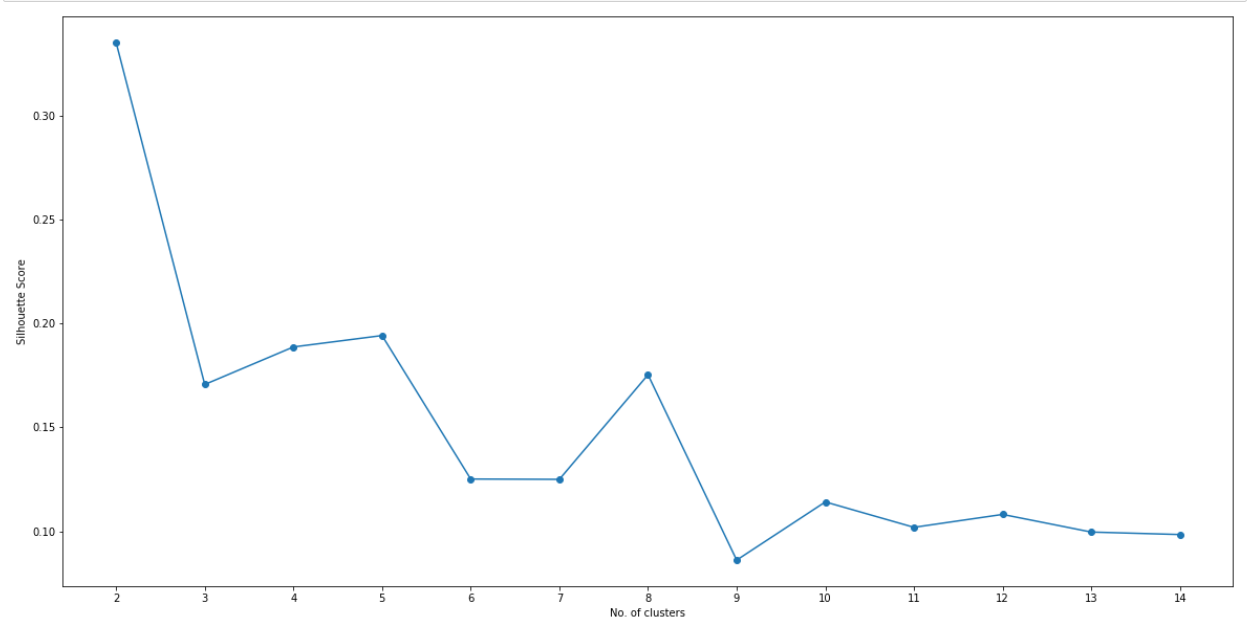


**Unsupervised Learning**

Unsupervised was used to find the similarities between the neighbourhoods. Kmeans a clustering algorithm is used to cluster the neighbourhoods that have similar category of venues.

**Kmeans** is a clustering algorithm that creates the cluster of data having similar type of characteristics by minimizing the data dispersion for each cluster. Each cluster represent a group of data having a pattern inside the multidimensional features. It is necessary for the algorithm to have the idea about the number of clusters since it is considered as the input of our algorithm. For this reason, elbow method was implemented.

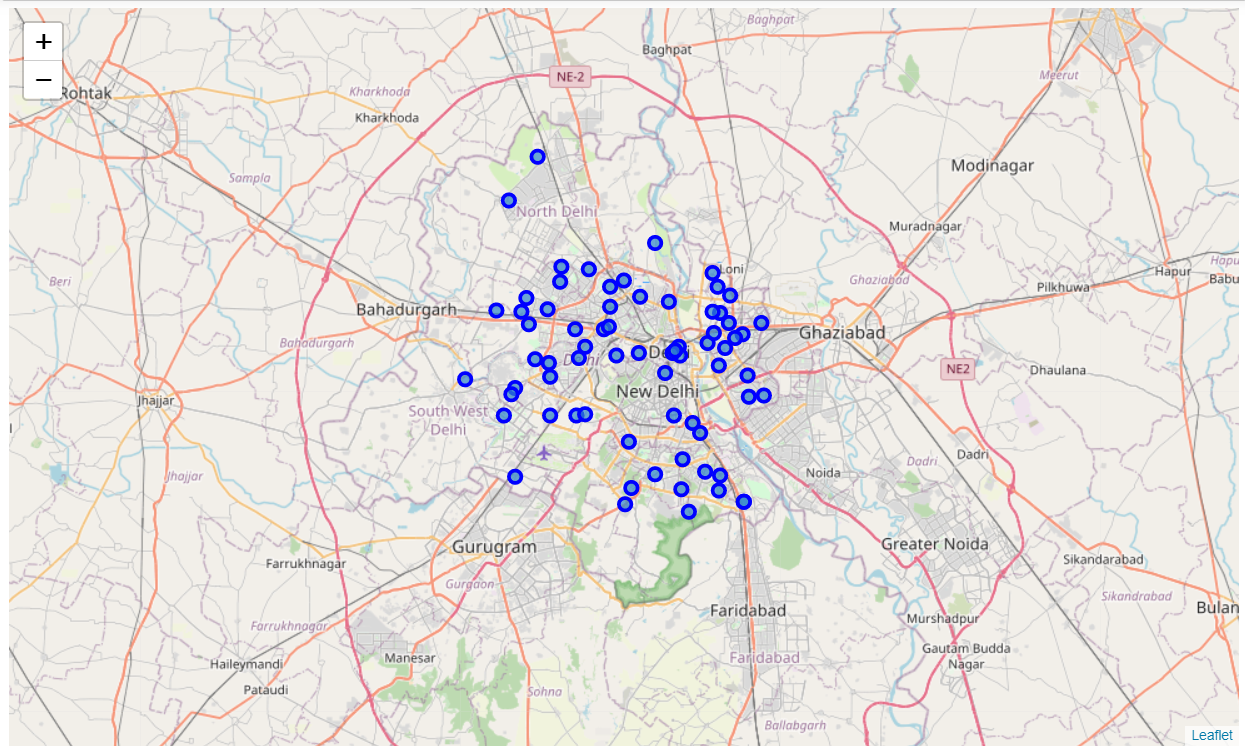




**Visualize the Data on MAP**

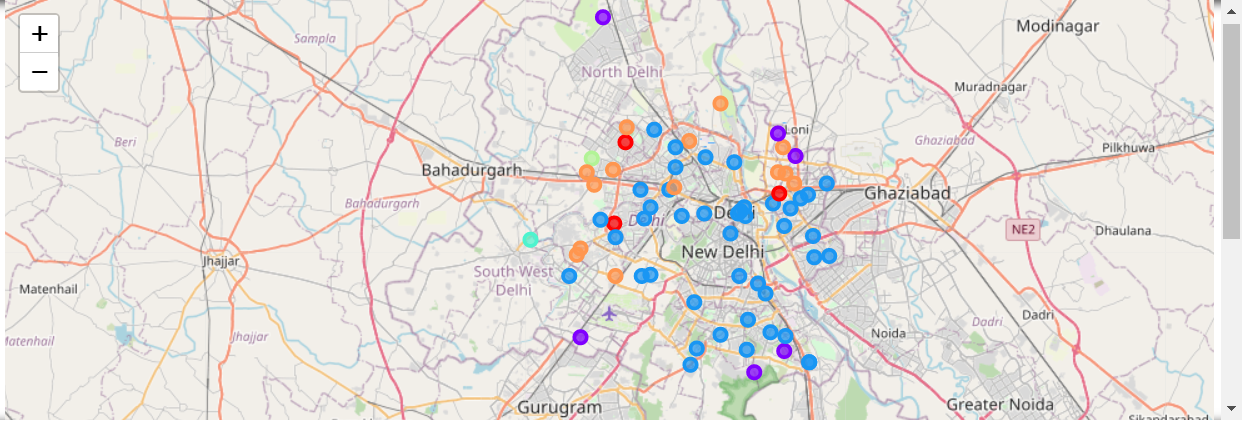
Visualization of data make it easy to understand the patterns behind the data and gives clear understanding about the data as compared to the quantitative data.

Folium Library was used to visualize the neighbourhoods on the map of Delhi. It was also used to show the clustered data.



**Result**

Kmeans algorithm was applied to the data of neighbourhoods of Delhi. Neighbourhoods clustered by Kmeans can be seen below.



After studying and visualizing all the clusters some important conclusions were made.

**Discussion**

As mentioned earlier the most suitable neighbourhoods to starting a restaurant are present in the cluster number 3.

Kmeans algorithm successfully clustered the neighbourhoods having similar type of venues.

After studying all the clusters, it is recommended to the client the **Rajouri Garden**, **Janak Puri** that fall in cluster 3 looks like good fit to start a restaurant. Client can go ahead and make decisions based on another factors.

**Conclusion**

Data Analysis and Machine Learning techniques used in this project can be very helpful in determine the solution of other business problems. Python’s inbuilt libraries such as Geopy, Folium makes it easy to visualize the geographical data.

In this project we have studied about the neighbourhoods in Delhi where our client can start a restaurant.