Coursera Capstone Project Applied Data Science

By : Satyajeet Dhawale

satyajit.dhawale@gmail.com

https://github.com/d-satyajit

Introduction

- Nagpur is central India biggest city and has the number of interesting places within the boundaries of the city.
- There are many places, restaurant and malls in Nagpur, where many people visits on daily basis. Also, due to new style of food delivery many people here have opened several home based small food restaurants and people prefer ordering from them.
- Hence to succeed with retail there has to better way of selecting a nearby restaurant and provide a fast accessible experience.

Business Problem

- Now and then everyone wants to eat from outside and also people go for business lunch/dinner or required food for partying at home.
- The main idea is to find the ideal nearby, optimal and most density restaurant for the customer base in city.

Data Collection

- All the major locations in Nagpur city were taken from the Wikipedia page (https://en.wikipedia.org/wiki/List_of_localities_in_Nagpur) and scraped using BeautifulSoup library in Python.
- To get the latitude and longitude of each location I have used Geocoder library in python and stored in csv file for each location.
- The venue data is then found via the FourSquare API by passing coordinates of each location. And all the venue data is captured in another new DataFrame.

Methodology

- **Precision of the Geocoder library in Python :** It was noted that for some location the Geocoder library was giving the incorrect coordinates. And hence few of the locations has to be checked manually and changes because they were having same name as other locations in India. Also, some names were incorrectly mention in the Wikipedia that also needs to be corrected.

- **Folium :** Folium is used to display the location points on the virtual map and for the cluster visualization.

Methodology

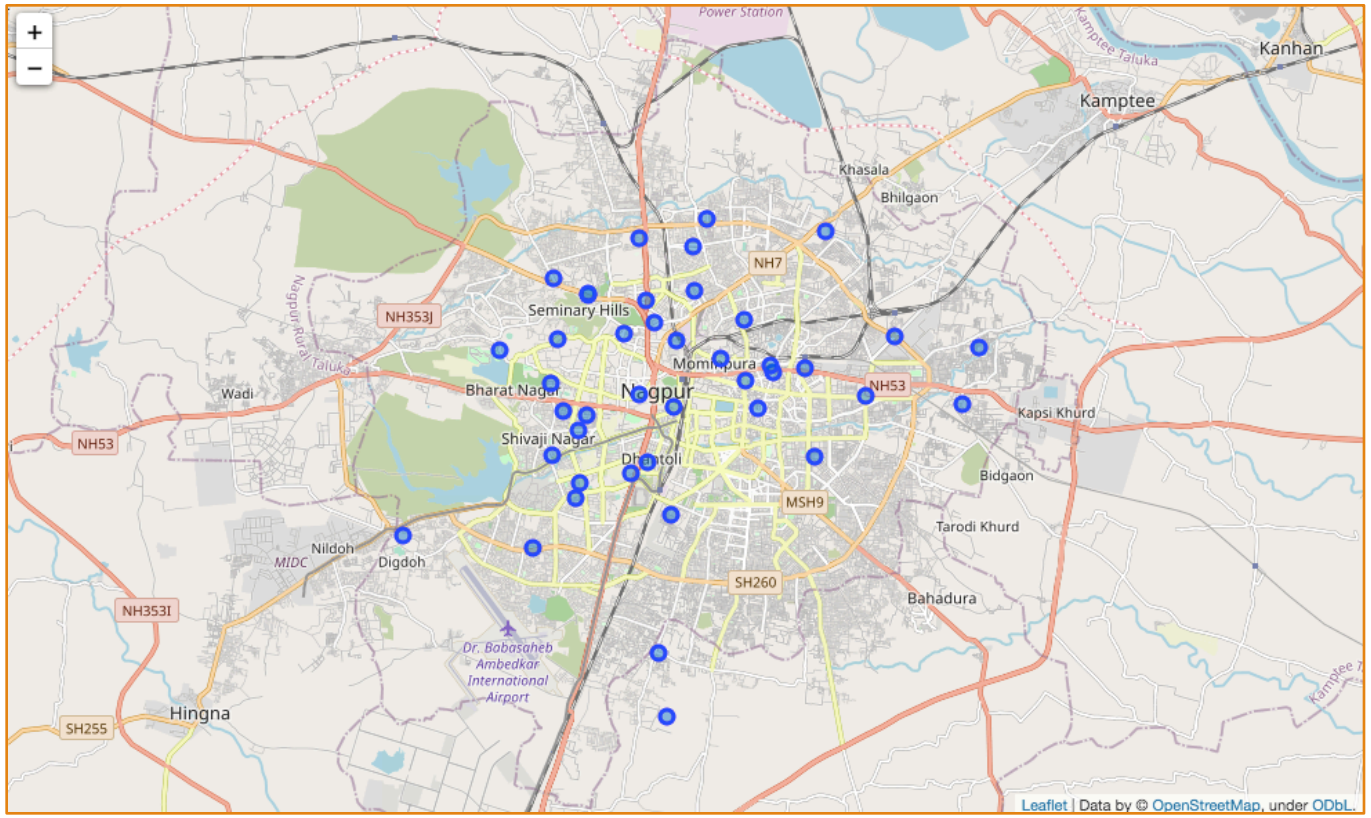

Figure : Neighborhood's of Nagpur.

Methodology

- **One hot encoding** : One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction. For the K-means Clustering Algorithm, all unique items under Venue Category are one-hot encoded.

- **Top 10 most common venues :** Due to high variety in the venues, only the top 10 common venues are selected and a new Data Frame is made, which is used to train the K-means Clustering Algorithm.

Methodology

- **Optimal number of clusters:** Silhouette Score is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from -1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. Based on the Silhouette Score of various clusters below 20, the optimal cluster size is determined.

- **K-means clustering:** The venue data is then trained using K-means Clustering Algorithm to get the desired clusters to base the analysis on. K-means was chosen as the variables (Venue Categories) are huge, and in such situations K-means will be computationally faster than other clustering algorithms.

Methodology

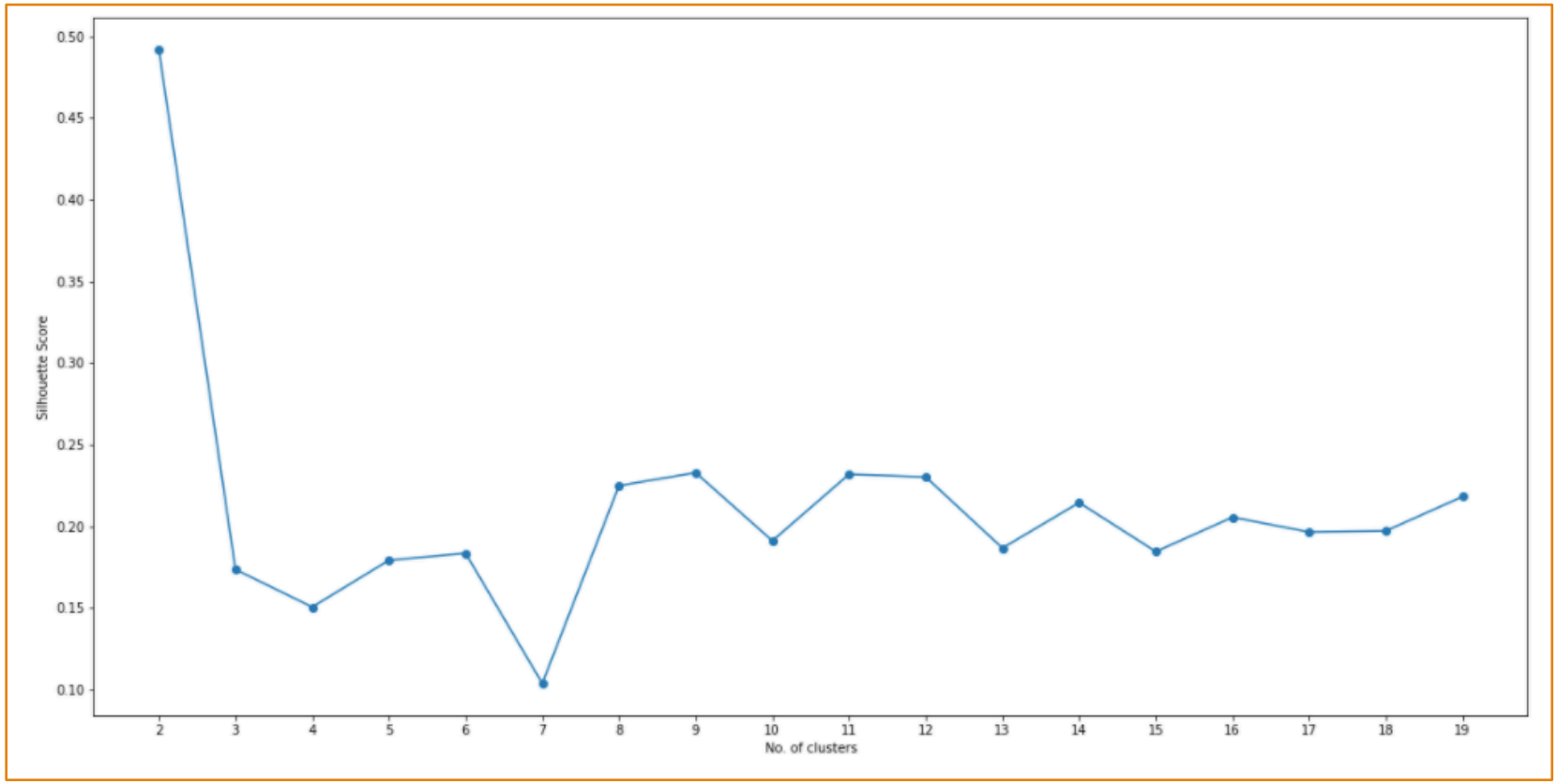

Figure : Silhouette score vs Number of clusters.

Results

- The neighborhoods are divided into ‘n’ clusters where ‘n’ is the number of clusters found using the optimal approach.
- The clustered neighborhoods are visualized using different colors so as to make them distinguishable.

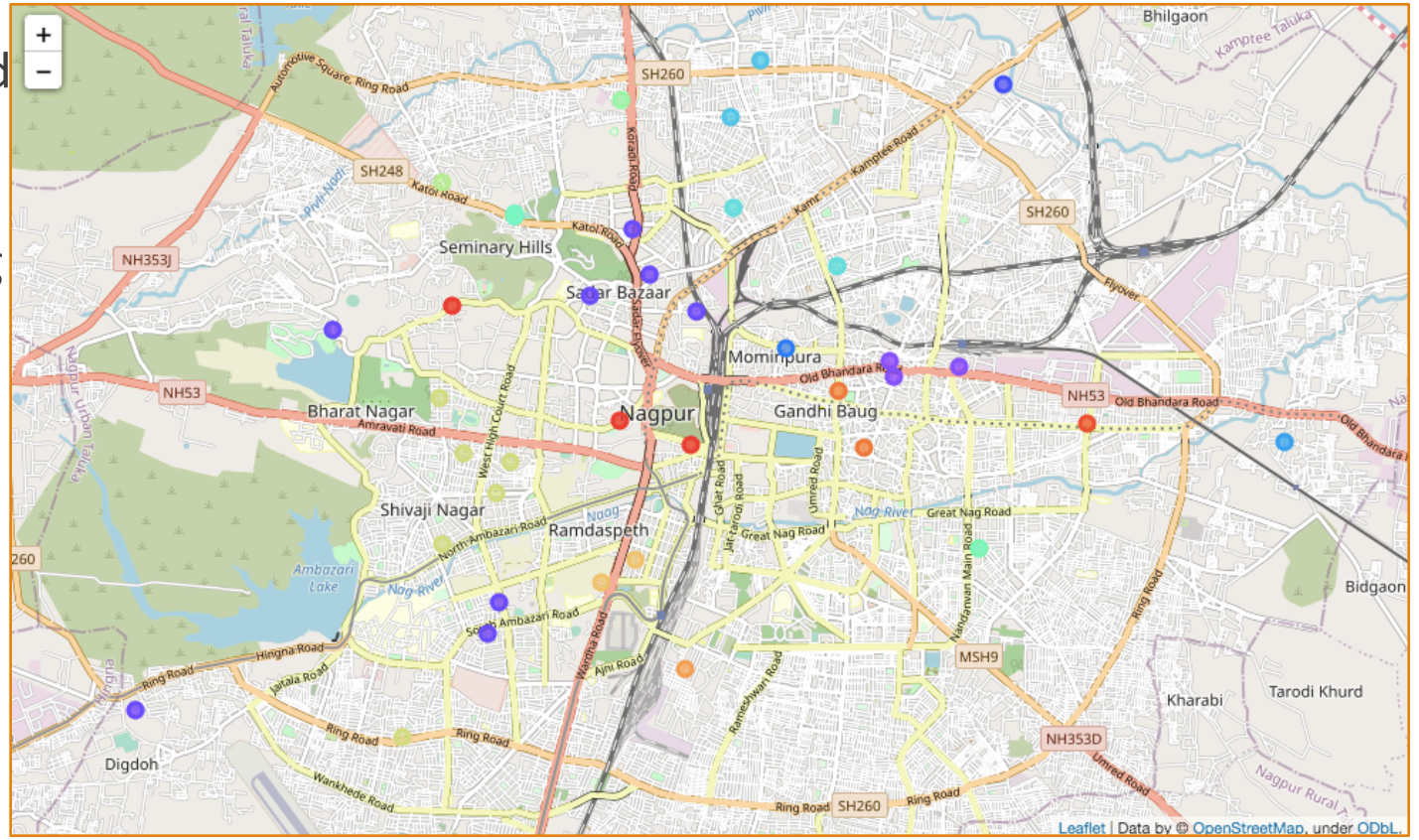

Figure : Clustered Neighborhoods of Nagpur.

Results

- The six places i.e. Dharampeth, Ravi Nagar, Pratap Nagar, Gokulpeth, Giripeth and Gandhinagar area fall in very dense areas of Nagpur and are mostly surrounded with restaurants, ice cream shops, clothing stores, coffee and snacks places.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Dharampeth	Ice Cream Shop	Coffee Shop	Indian Restaurant	Clothing Store	Fast Food Restaurant	Plaza	Snack Place	Market	Park	Nightclub
19	Ravi Nagar	Clothing Store	Indian Restaurant	Lounge	Coffee Shop	Restaurant	Park	Tea Room	Athletics & Sports	Trail	Hookah Bar
25	Pratap Nagar	Coffee Shop	Snack Place	Restaurant	Indian Restaurant	Sandwich Place	Ice Cream Shop	Lounge	Gym	Plaza	Furniture / Home Store
32	Gokulpeth	Indian Restaurant	Ice Cream Shop	Clothing Store	Park	Coffee Shop	Market	Nightclub	Fast Food Restaurant	Plaza	Café
33	Giripeth	Ice Cream Shop	Indian Restaurant	Coffee Shop	Clothing Store	Fast Food Restaurant	Restaurant	Market	Park	Nightclub	Plaza
37	Gandhinagar	Snack Place	Ice Cream Shop	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Clothing Store	Market	Diner	Park	Café

Discussion

- After analysing the various clusters produced by the Machine learning algorithm, cluster no.14, is a prime fit to that shows six different locations having different food corners that includes restaurants, ice cream shops, clothing stores, coffee and snacks places.

- These areas fall in very dense areas of Nagpur and are mostly surrounded with people, hence it is obvious that most of the people will go here only. But there is a lot of scope for food business owners to open their food corners in areas other than these six areas.

Conclusion

- As the high growing population of the Nagpur City and keeping in mind that India is a country of youngster's food corners like restaurants, ice cream shops, coffee and snacks places and clothing stores will have high demand.
- If the business owners try to focus on the areas where there is a high demand of customers and less food corners their business will definitely will grow. Also, the it will be very beneficial for the people if the food corners will be near to their home.