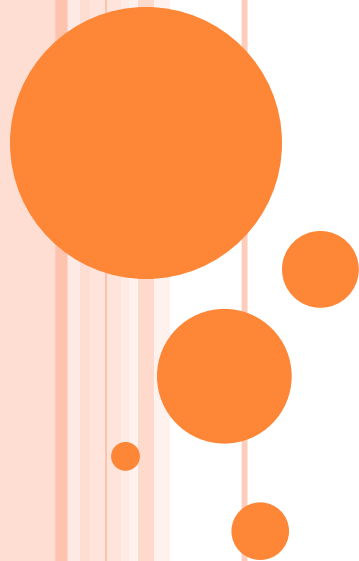


IBM APPLIED DATA SCIENCE CAPSTONE

OPENING OF NEW RESTAURANT IN THE NEIGHBORHOOD.

By: Sneha Videkar

Date: 15 May 2020



INTRODUCTION

- Yelp [1][2] is a company which publishes crowd-sourced reviews about businesses using the company website.
- Yelp dataset captures business data from all around the world.
- Mainly this site captures reviews related to business of various categories like Dentist, Restaurant, etc.
- Along with reviews, it lists location details of every business in Yelp database.



BUSINESS PROBLEM

- Success of new business depends on many factors. Location is the key factor which should be considered while opening new business.
- For this purpose, location attributes present in the Yelp dataset[4] will help find a place to start a new business.
- This project aim to help investors/entrepreneur to find an appropriate location for their prospective restaurant business.
- With the help of Yelp Business dataset and machine learning techniques like clustering, this project will locate top 10 most common restaurant businesses in a neighborhood.
- This will guide investors while opening a new restaurant business at particular location.



TARGET AUDIENCE

- This project is mainly helpful for the investors/entrepreneur who are looking for an appropriate location for their prospective restaurant business.
- Also, with the help of exploratory analysis from this project can help restaurant business owners to identify potential competitors.



DATA ACQUISITION

- For this problem below dataset will be used:1. Business.csv will be used from Yelp Dataset.
 - This dataset is taken from Kaggle website[3] and Yelp dataset website[4]. Figure 1 gives an idea about few records from the Business.csv dataset.

business_id	name	neighborhood	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	categories
kCoE3jvEtg6UVz5SOD3GVw	"BDJ Realty"	Summerlin	"2620 Regatta Dr, Ste 102"	Las Vegas	NV	89128	36.207430	-115.268460	4.0	5	1	Real Estate Services;Real Estate;Home Services...
xcgFnd-MwkZeO5G2HQ0gAQ	"T & T Bakery and Cafe"	Markham Village	"35 Main Street N"	Markham	ON	L3P 1X3	43.875177	-79.260153	4.0	38	1	Bakeries;Bagels;Food
fNMVV_ZX7CJSDWQGdOM8Nw	"Showmars Government Center"	Uptown	"600 E 4th St"	Charlotte	NC	28202	35.221647	-80.839345	3.5	7	1	Restaurants;American (Traditional)
l09JfMeQ6ynYs5MCJtrcmQ	"Alike Catering"	Yonge and Eglinton	"2459 Yonge St"	Toronto	ON	M4P 2H6	43.711399	-79.399339	3.0	12	0	Italian;French;Restaurants
lHYICS-y8AFJuitv6MGpxg	"Starbucks"	Liberty Village	"85 Hanna Avenue"	Toronto	ON	M6K 3S3	43.639863	-79.419533	4.0	21	1	Food;Coffee & Tea

- Forsquare API [5] will be used to get the most common venues for abovementioned business and in Toronto area. Only near by venues of typerestaurants will be considered for further analysis to achieve the projectgoal



DATA PREPROCESSING

- Yelp Business data was loaded in the IBM Watson Studio and accordingly csv file was read in dataframe called businessDF.
- Yelp business data frame has captured various categories of the businessdata.
- In this project, business with “Restaurant” as a category are considered for modelling



METHODOLOGY

- Exploratory analysis was performed to get better understanding of the data.
- Yelp dataset tries to captures details about different businesses. This is highlighted in the figure.

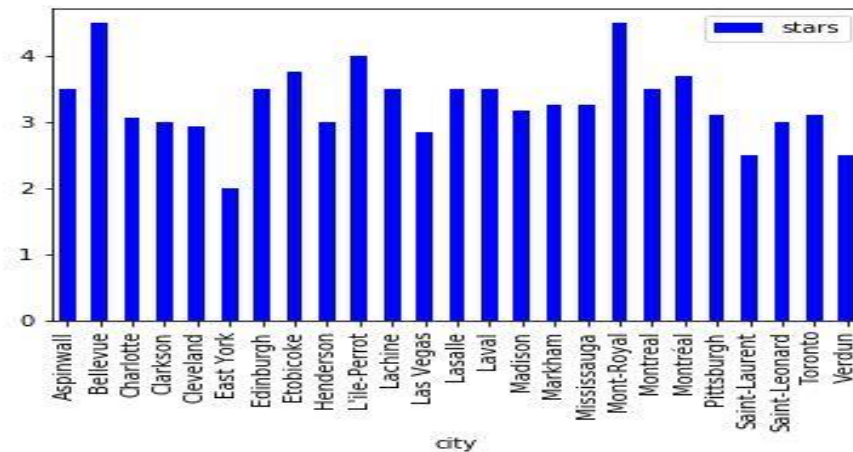


- Here, Worcloud library was used as data visualization technique.



METHODOLOGY

- This project has focused on business with category as Restaurants.
- Average ratings of all the restaurants were compared using bar plot.



- From above bar graph, we can identify city namely Mont-Royal and Bellevue has achieved highest ratings for Restaurant business.



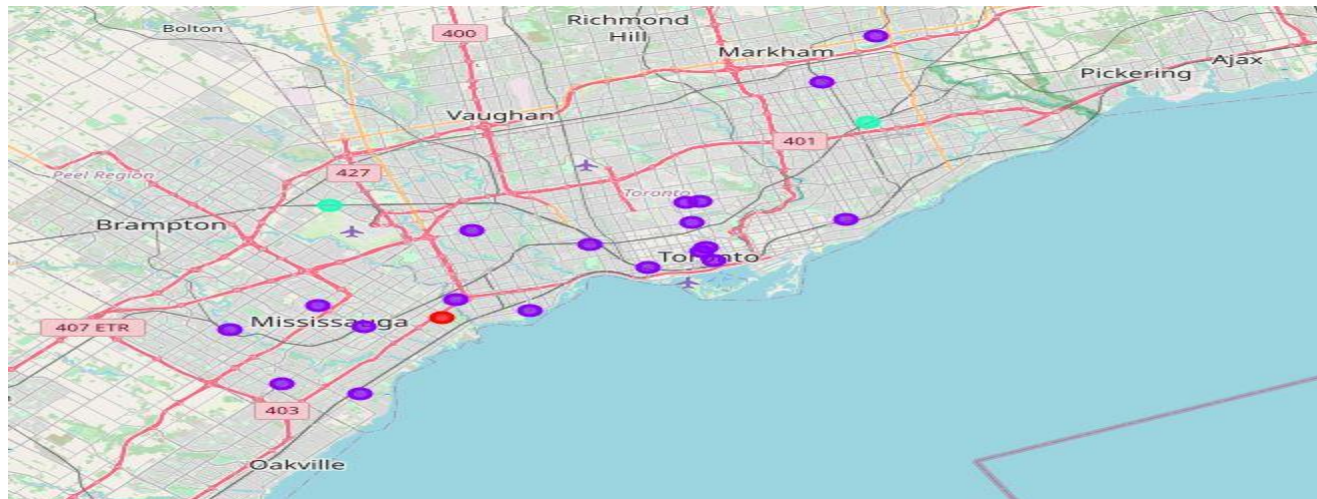
METHODOLOGY

- With help of Forsquare API[5] and Forsquare developer account, neighborhood area of restaurant businesses derived from the Yelp Business data was explored.
- Area with radius of 500 meters was explored to get nearby venues.
- Machine learning technique like clustering was used in this project.
- K-means clustering was used to cluster the neighborhood in 3 clusters.
- Map for visualization were used.



RESULTS

- The result of clustering is shown in figure below for the Toronto area.

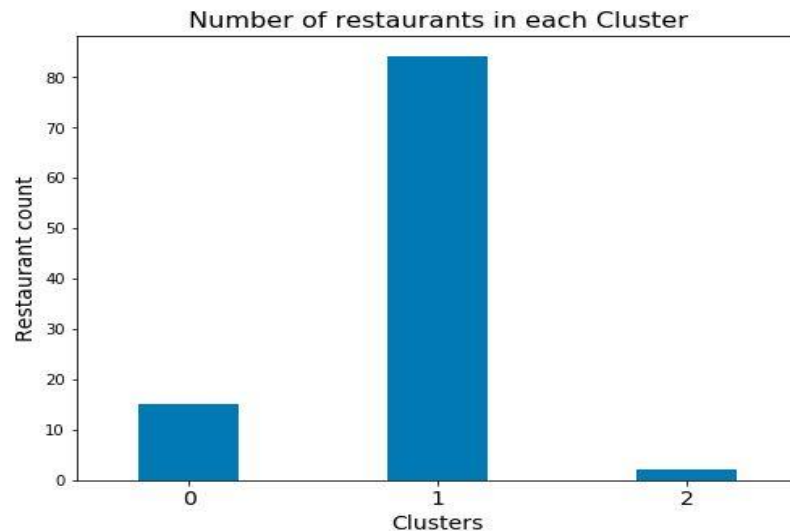


- From the map it is clear that, neighborhood is divided in 3 different cluster.
- Cluster 0 : is represented with red color and has moderate number of restaurants.
- Cluster 1 : is represented with purple color and has highest number of restaurants.
- Cluster 2 : is represented with green color and has few restaurants.



DISCUSSION

- We can say most of the restaurants are in the neighborhood which belong to cluster 1.
- In total, 84 restaurants are in cluster 1.



- With this information, we can advise new business owners to avoid neighborhood area which belongs to cluster 1 while opening their prospective restaurant's business.
- On the other hand Cluster 0 and 2 has a handful of restaurants and it gives an excellent opportunity to start a new business.



DISCUSSION

- Further cluster analysis provides detailed information on type of top 10 restaurant in each cluster along with neighborhood area in which they live.
- With this information, the business owner is notified with potential competitors in the neighborhood.
- Cluster 1 has restaurants of type Italian, Indian, Korean etc.
- Upon comparison of cluster 0 with cluster 1, we can say that more fast-food restaurants are listed in a neighborhood which belongs to cluster 0.
- So if an investor is looking to open Fast-food restaurant then neighborhood which belongs to cluster 1 or 2 will be of good choice rather than cluster 0.



CONCLUSION AND FUTURE WORK

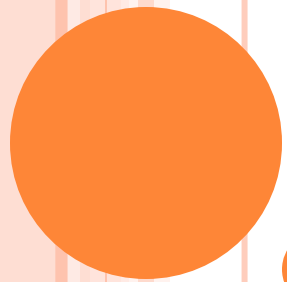
- With this project, we can guide investors to find an appropriate location for their prospective restaurant business.
- Machine learning and visualization techniques gives better information to the investors.
- Interactive GUI could be future work for this project.



REFERENCES

- [1] <https://en.wikipedia.org/wiki/Yelp>
- [2] <https://www.yelp.com/about>
- [3] <https://www.kaggle.com>
- [4] <https://www.yelp.com/dataset/documentation/main>
- [5] <https://foursquare.com/>





THANK YOU!