Capstone Project - The Battle of Neighborhoods

1.Introduction

1.1 Scenario and Background:

I am planning to start a food truck in city of Toronto and I wanted to know in which areas the food is being supplied by different media and more venues are there other than food supplies. In that type of area, I can move my food truck with in the that area to make more business with less efforts to find a place to do food business.

1.2 Problem to be resolved:

How to find a neighborhood in Toronto with the following conditions to move food truck

- Food business is successful
- More venues along with restaurants
- Area with radius of 1km

1.3 Interested Audience:

I believe the methodology, tools and strategy used in this project is relevant for a person or entity considering moving to a major city in US, Europe or Asia. Europe, US or Asia, Likewise, it can be helpful approach to explore the opening of a new business. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case for a person developing Data Science skills.

2. Data

2.1 Data Requirements:

- Data with postal codes and neighborhoods in the city of Toronto with latitude and longitude
- Nearby venues of the neighborhoods

2.2 Data Sources, Data Processing and Tools used:

- Postal codes with neighborhoods will be web scrapped form wiki page using beautifulsoup
  'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
- Latitude and longitude of the neighborhood will be collected from the following geospatial data http://cocl.us/Geospatial_data
- Nearby venues of the neighborhoods collected by using foursquare api.

2.3 Data collection

2.3.1 getting Postal codes with neighborhoods from wikipedia

by doing web scraping ,using the BeautifulSoup package and creating a data frame with the data

Beautiful Soup *is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.*

2.3.2 collecting the geospatial data

now we will collect the geospatial data for lattitude and longitude of the collected postal codes in city of Toronto

2.3.3 preprocessing the data set

we remove the rows which contain not assigned as a value for borough or neighborhood usually this is refered as cleaning

2.3.4 nearby venues using foursquare api

now we have to find the nearby venues of our neighborhoods using the foursquare api with latitude and longitude

Foursquare *City Guide, commonly known as Foursquare, is a local search-and-discovery mobile app developed by Foursquare Labs Inc. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.*
now we have the all the data regarding venues nearby the neighborhoods and their locations

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Food truck

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

3. Methodology

In this project we will direct our efforts on detecting areas of Toronto that the food is being supplied by different media and more venues are there other than food supplies. We will limit our analysis to area ~1km around city center.

1. In first step we have collected the required data: location and venues with in 500m radius in toronto city . We have also identified venue category (according to Foursquare categorization).

2. Second step in our analysis will be calculation and exploration of 'venue freaqunecies' across different areas of Toronto - we group the venues by their neighborhoods and finding most common venue type.

3. In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established: we will take into consideration locations with more number of venues and food as a most common one. We will present map of all such locations but also create clusters (using k-means

clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final borough exploration and search for optimal venue locations by stakeholders.

4.Analysis

4.1 nearby venues

we found ,for each neighborhood how many venues are there
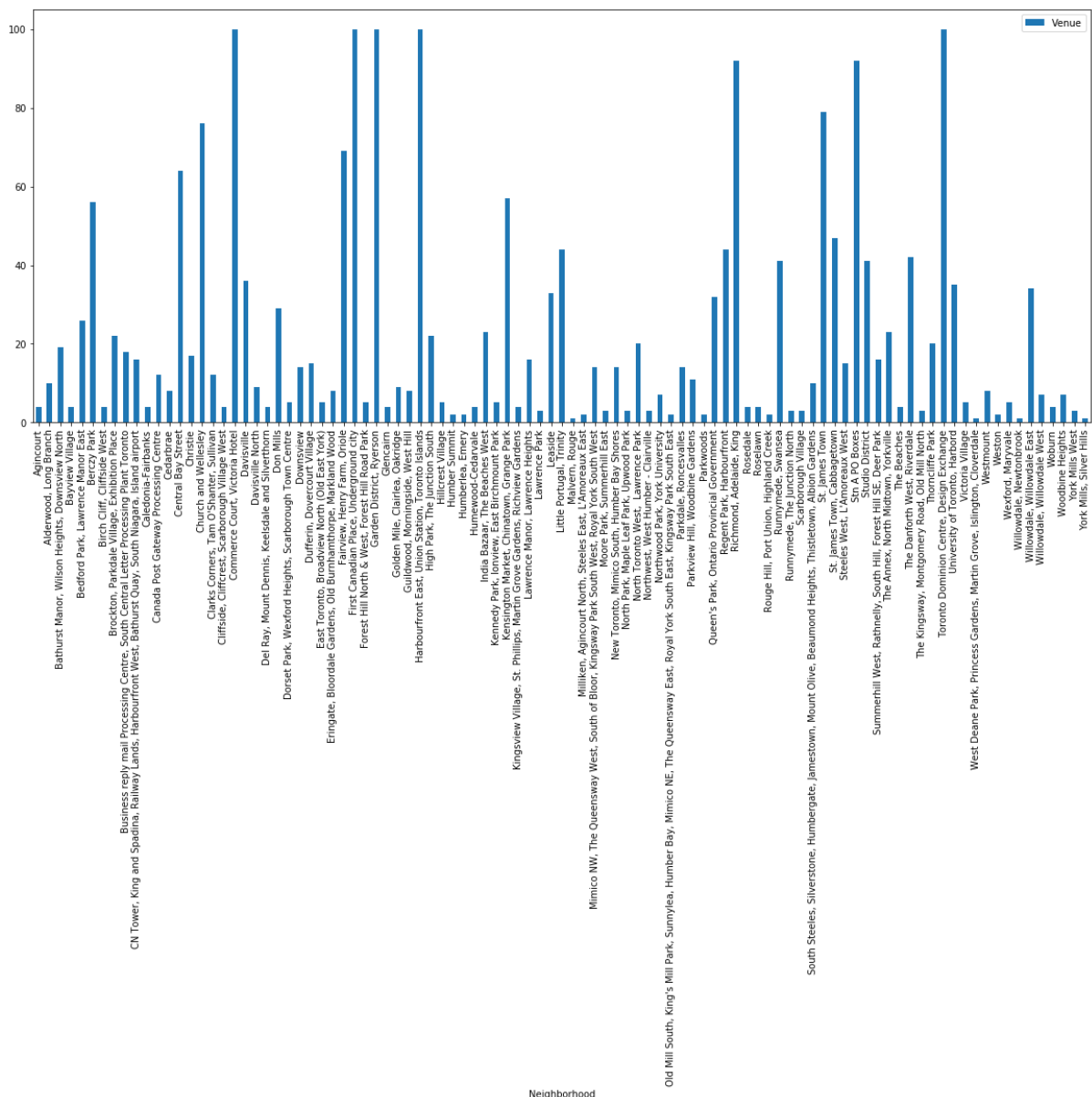
4.2 **start analysing each neighborhood**

we find the onehost encoding of the neighborhoods with venue category for clustering purpose after grouping all these with neighborhood,we can see the proportions of neighborhoods to venue categorywhich makes sense that which type of venue is famous in each neighborhood.

That frequency can be visualised uaing the wordcloud.
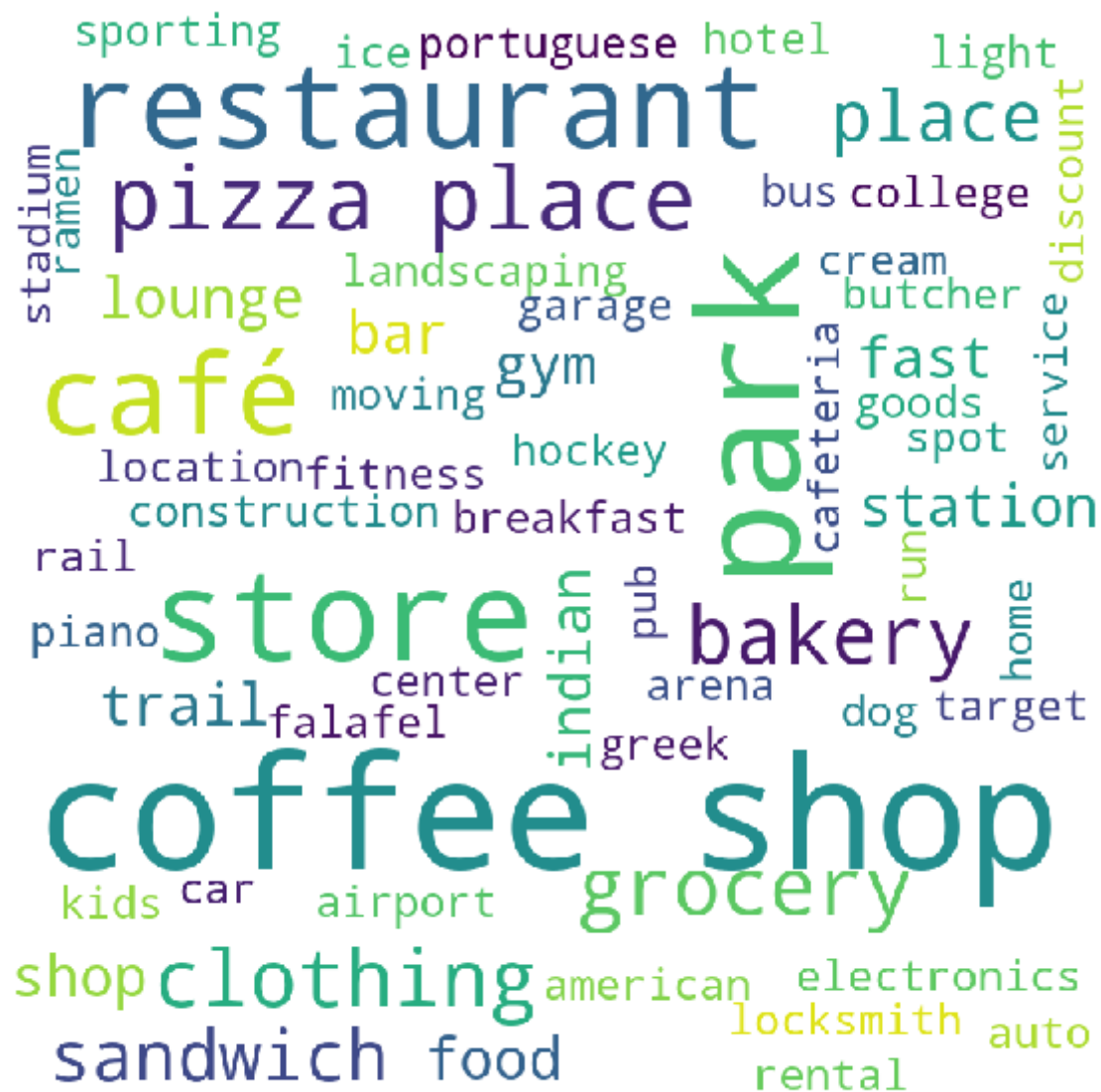
4.3 Clusetring the neighborhoods

4.3.1 kmeans

We cluster the neighborhood using Kmeans algorithm

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

We clustered all these neighborhoods as 4 clusters and we also found the most common type of venues in each neighborhood and that looks like

	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	
0	M3A	North York	Parkwoods	43.753259	-79.329656	2	Park	Food & Drink Shop	Yoga Studio	Dog Run	Dessert Shop	Dim Sum Restaurant	Diner	
1	M4A	North York	Victoria Village	43.725882	-79.315572	0	Portuguese Restaurant	Intersection	Coffee Shop	French Restaurant	Hockey Arena	Yoga Studio	Dessert Shop	R
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	Coffee Shop	Pub	Bakery	Park	Breakfast Spot	Café	Theater	
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	0	Clothing Store	Furniture / Home Store	Women's Store	Accessories Store	Boutique	Sporting Goods Shop	Miscellaneous Shop	
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Yoga Studio	Italian Restaurant	Diner	Bar	Bank	Beer Bar	

After visulaising the cluster with different color we can easily see the clusters.
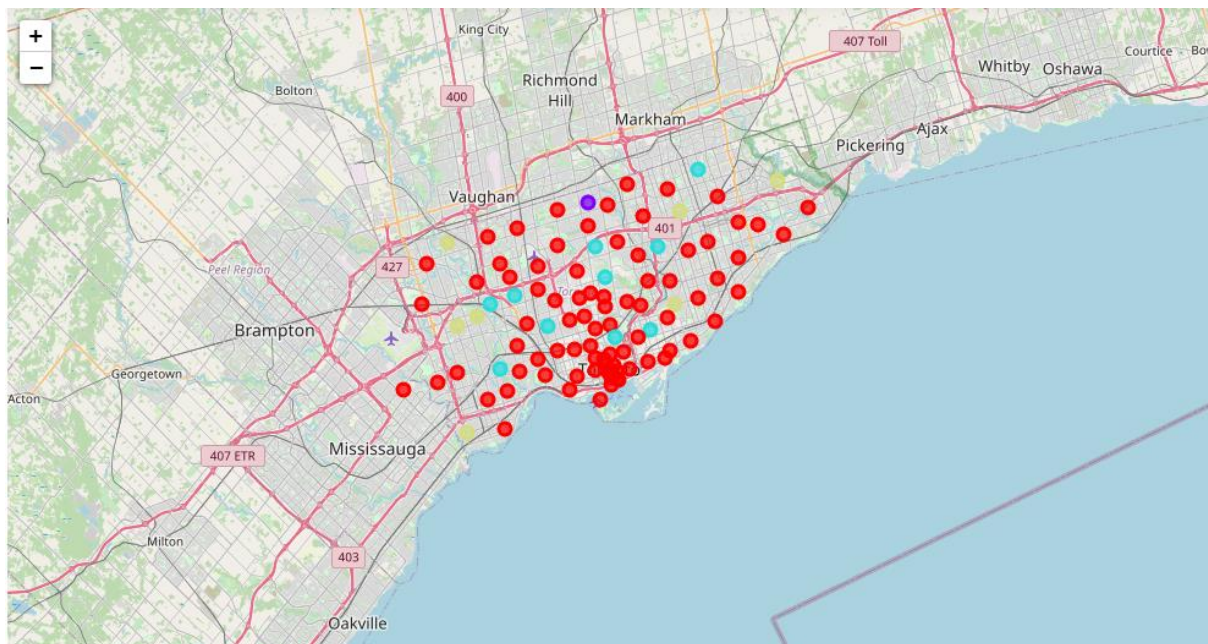
4.3.2 analysing clusters

lets find out how many venues are there in each cluster and what are the venue categories in each cluster

cluster	Number of eighborhoods	Most common venue type
Cluster 1	83	Portuguese Restaurant
Cluster 2	1	Piano Bar
Cluster 3	10	Park
Cluster 4	7	Fast Food Restaurant

5.Results and Discussion

From our analysis the summary of the clustering can show below
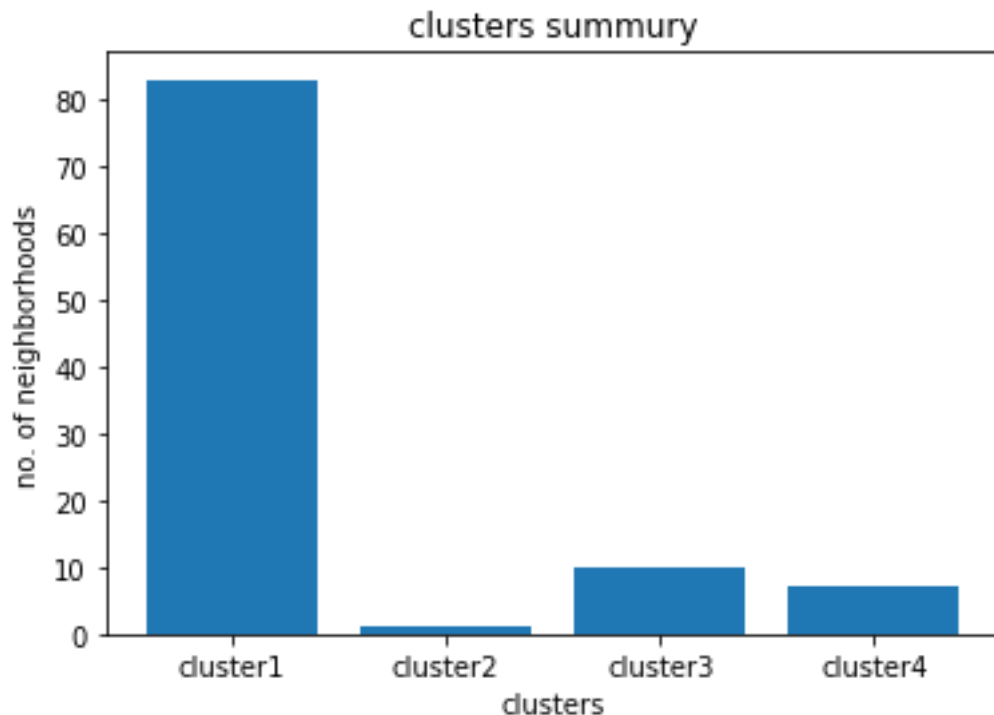
by seeing this graph we can easlily say that we have to select the cluster 1 and in cluster 1 the following are most common Portuguese Restaurant,Coffee Shop,Clothing Store etc.

Our analysis shows that although there is a great number of restaurants in Toronto, there are pockets of food truck in city center. Highest concentration of venues was detected in 10 places in a cluster which includes **North York,Downtown Toronto, Etobicoke,Scarborough,East York,York,East Toronto,West Toronto,Central Toronto,Mississauga** which satisifies our conditions.

firstly we concentrated on the nieghborhoods groups and calculated the frequent venues in each neighborhoods and got to know food is common so we decided that food has priority. after this we formed 4 clusters suing kmeans algorithm and found each cluster with its particulars.

among those 4 we have one cluster with more number of venues so people will roam around that places so we can have a food truck in that particluar places to make good money. we can do further by tracking the trending places and to move the truck accordingly.

6.Conclusion

our main goal is to find a neighborhood in Toronto where Food business is successful and More venues along with restaurants

now we got 10 places where there are more number of venues as you can see in maps with limited radius and food is being most common so we can choose those plcaes to move the food truck