DATA SCIENCE CAPSTONE PROJECT

Battle of Neighbourhoods

Introduction

A start up based on providing renewable energy services has recently made autonomous vehicles for delivery of food in a smaller region of Toronto .The vehicles are solar powered .The company wants to begin it's services from the area where there are lots of restaurants so that it can pick up food from those restaurants and deliver in nearby areas.

The vehicles available are less and moreover they can't travel for longer distances .The startup has taken loan for building these vehicles now it's time to repay that loan as per agreement. For that they have to employ vehicles in such a way that it maximizes there profit . So they decided to employ these vehicles in areas where there are more restaurants. By doing so they will be able to collect food from larger restaurants spending less time and resources. So, as a pilot project as well as to maximize the profit they decided to employ these vehicles in restaurants rich area. So, they need help of a data scientist to help them in this cause.

Who would be interested?

Anyone who needs to start food delivery service or any other delivery services to and from restaurants.

Data Used in problem solving

1. **Wikipedia Data scraped from web (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)**

2. **Geospatial Data**

   (https://cocl.us/Geospatial_data)

Wikipedia Data

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website,[1] via its applications for such smartphones as the iPhone and BlackBerry,[2] and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Geospatial Data

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917

This data consists of longitude and latitude location corresponding to postal codes given in the first column.

Merged Data

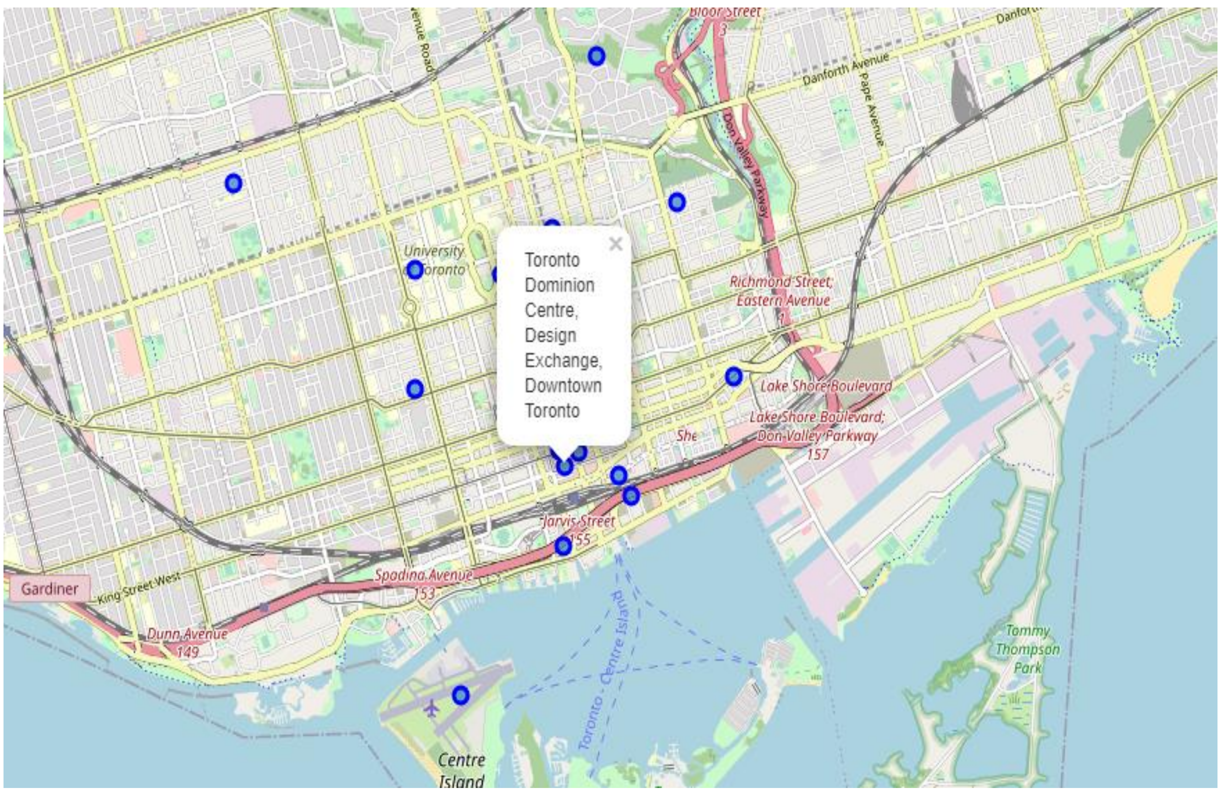
Now we require data that will contain that will contain neighbourhoods ,boroughs ,latitude and longitude corresponding to postal codes in the first column.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Methodology

1. Data Visualization

At the very onset we will visualize the neighbourhoods in the toronto city of Canada .Here we will use geocoder library to obtain location as longitude and latitude. Then using that data we will plot map using folium .



2. Using Foursquare Api to obtain nearby venues

Now we will use Foursquare Api to obtain nearby venues in city of Toronto within radius of 500 meters

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
Regent Park, Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

3. Obtaining Neighbourhood Count

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Berczy Park	56	56	56	56	56	56
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	15	15	15	15	15	15
Central Bay Street	63	63	63	63	63	63
Christie	17	17	17	17	17	17
Church and Wellesley	77	77	77	77	77	77
Commerce Court, Victoria Hotel	100	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100	100
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
Kensington Market, Chinatown, Grange Park	56	56	56	56	56	56
Queen's Park, Ontario Provincial Government	33	33	33	33	33	33
Regent Park, Harbourfront	48	48	48	48	48	48
Richmond, Adelaide, King	93	93	93	93	93	93

4. Converting Data into onehot vector

	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

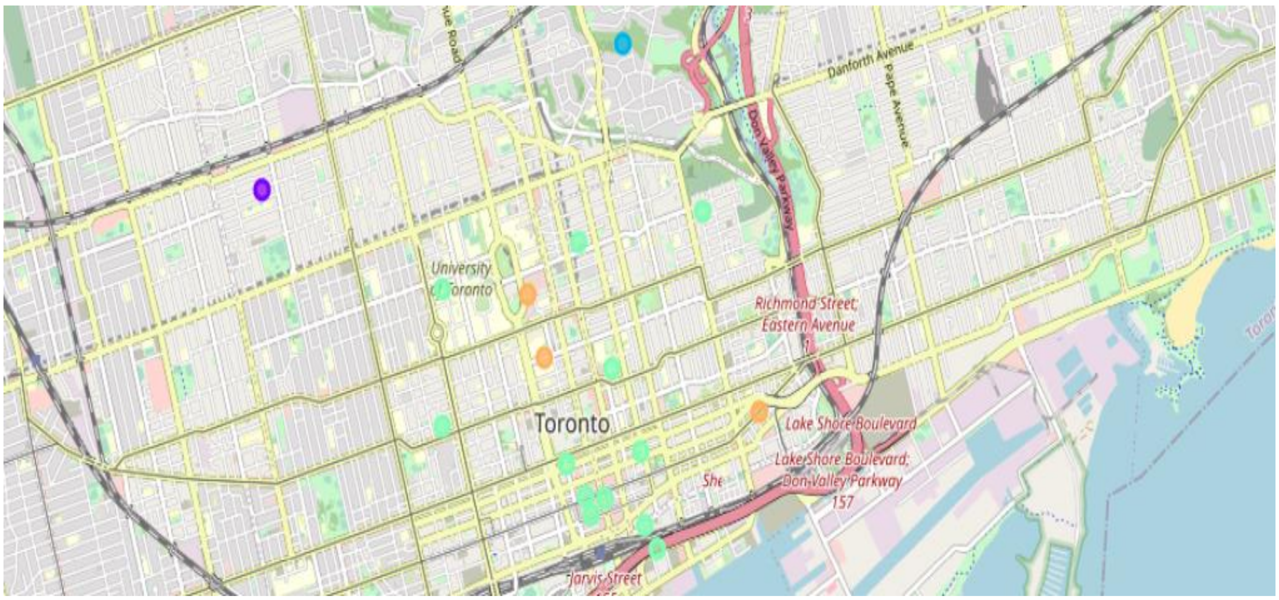
5. Groupby Data

Now we will group by data based on the mean of the frequency of occurrence of a particular venue.

	Neighborhood	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	A Muse
0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.017857	0.0000
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000	0.066667	0.066667	0.066667	0.066667	0.2	0.133333	0.000000	0.000000	0.00	0.000000	0.0000
2	Central Bay Street	0.015873	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.0158
3	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.0000
4	Church and Wellesley	0.025974	0.012987	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.012987	0.000000	0.00	0.000000	0.0000
5	Commerce Court, Victoria Hotel	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.040000	0.000000	0.00	0.010000	0.0000

6. <u>Finding Cluster</u>

Now we will use unsupervised machine learning algorithm known as clustering and find clusters of similar venues.
We will take value of k=5 to find five clusters.



This map shows the clusters obtained through clustering algorithm.

Cluster 1:

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	Downtown Toronto	0	Airport Service	Airport Terminal	Sculpture Garden	Plane	Boutique	Rental Car Location	Harbor / Marina	Airport Lounge	Airport Gate	Airport Food Court

Cluster 2:

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Downtown Toronto	1	Grocery Store	Café	Park	Candy Store	Nightclub	Coffee Shop	Italian Restaurant	Restaurant	Diner	Athletics & Sports

Cluster 3:

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	Downtown Toronto	2	Park	Trail	Playground	Creperie	Doner Restaurant	Dog Run	Distribution Center	Discount Store	Diner	Dessert Shop

Cluster 4:

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Downtown Toronto	3	Clothing Store	Coffee Shop	Bubble Tea Shop	Italian Restaurant	Middle Eastern Restaurant	Cosmetics Shop	Japanese Restaurant	Restaurant	Café	Dine
3	Downtown Toronto	3	Coffee Shop	Café	Cocktail Bar	Gastropub	American Restaurant	Seafood Restaurant	Italian Restaurant	Department Store	Lingerie Store	Creperie
4	Downtown Toronto	3	Coffee Shop	Cocktail Bar	Seafood Restaurant	Bakery	Beer Bar	Restaurant	Cheese Shop	Café	Hotel	Japanese Restauran
7	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Gym	Thai Restaurant	Hotel	Clothing Store	Deli / Bodega	Bookstore	Sush Restauran
9	Downtown Toronto	3	Coffee Shop	Café	Hotel	Restaurant	Seafood Restaurant	American Restaurant	Salad Place	Japanese Restaurant	Italian Restaurant	Sush Restauran
10	Downtown Toronto	3	Coffee Shop	Café	Restaurant	Hotel	American Restaurant	Gym	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Deli / Bodega
11	Downtown Toronto	3	Café	Restaurant	Bar	Italian Restaurant	Japanese Restaurant	Bookstore	Bakery	Yoga Studio	Beer Bar	Beer Store
12	Downtown Toronto	3	Café	Bakery	Mexican Restaurant	Vietnamese Restaurant	Coffee Shop	Gaming Cafe	Vegetarian / Vegan Restaurant	Dessert Shop	Bar	Breakfas Spo
15	Downtown Toronto	3	Coffee Shop	Seafood Restaurant	Café	Italian Restaurant	Restaurant	Cocktail Bar	Beer Bar	Japanese Restaurant	Park	Hote

Cluster 5:

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Downtown Toronto	4	Coffee Shop	Pub	Park	Bakery	Restaurant	Breakfast Spot	Café	Theater	Spa	Ice Cream Shop
1	Downtown Toronto	4	Coffee Shop	Sushi Restaurant	Gym	Diner	Park	Mexican Restaurant	Japanese Restaurant	Italian Restaurant	Hobby Shop	Wings Joint
5	Downtown Toronto	4	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Japanese Restaurant	Bar	Ice Cream Shop	Burger Joint	Thai Restaurant	Salad Place
8	Downtown Toronto	4	Coffee Shop	Aquarium	Café	Hotel	Scenic Lookout	Sporting Goods Shop	Italian Restaurant	Brewery	Restaurant	Fried Chicken Joint

Observations

From above observations it is clear that cluster 4 will be most appropriate choice for our purpose as it has lots of restaurants.

Conclusion

The startup wanted to find the area where they can deploy there solar powered autonomous vehicles and we have used clustering algorithm to find the most appropriate areas that they can target. This will not only boost up their revenue but also lead to optimum utilization of their resources.

Future prospects

As renewable resources are future of humanity in future there will many solar powered vehicles that may run on the roads. But the limitation of those vehicles is that they are not efficient for long distances. So ,we can use them for short distances for specific services like delivery services, where clustering algorithm will come in handy.