Capstone Project

The Battle of Neighborhoods-Presantation

Mustafa YAZICI

1.Introduction

- This report can help those who are planning to start a new business such as an international cuisine restaurant in Toronto city and need to find the best location for their shop.

1.1 Background

- The number of population increases due to newborns and migration. Therefore demands are increasing in every areas.Such as demands of restaurants.
- In addition, people's preferences and needs change day by day. For this reason, new shopping malls, shops, restaurants etc. are set up every day all over the world.
- The population of Toronto is increasing. New restaurants are needed in this city. There are also different restaurants, which include different cuisine in Toronto.

1.1 Background

- Especially the central districts of the Toronto are becoming more and more become a shining star.
- Since Toronto is a very important place, it attracts attention from very famous brands. Many famous brands want to be in Toronto. The entrepreneurs want to take advantage by having restaurants in the central of Toronto.

1.2 Business Problem

- Unfortunately, it is not possible to access all kinds of data. Even if it is reached, it may take years to make sense of them with today's technology. therefore, it is necessary to collect the important data and make realistic assumptions when necessary.

- Acceptances:

  1. Restaurant customers are located around the available restaurants. And it is appropriate to specify these clients in the make_blobs method with cluster_std = 0.003.

  2. Since it covers the world cuisines, it aspires to the customers of every available restaurant.

  3. Since the company wanted to open a place in the center, the information of the restaurants close to the center was used.

  4. It is accepted that every restaurant has registered to the site where the data is received.

1.2 Business Problem

- Since the restaurant is a world brand, it covers the world food and is a competitive company, it will increase the probability of success, rather than escaping from the center. Therefore, I need to find where the customer cluster is most concentrated.
- The problem is :If the client want to open an international luxury cuisine restaurant in Toronto city, which point is the best option to open the restaurant?

1.3 Interest

- Selecting the right location for a business is one of the first and very important decisions in running a business.
- This report can help those who are planning to start a new business such as an international luxury cuisine restaurant in Toronto city and need to find the best location for their shop.
- A world brand restaurant chain, which can better understand people's new wishes, is innovative, contains many world cuisines and can attract customers from existing restaurants, will want to make a place in Toronto. And they will want to know where the most suitable place is for this place. The customer will look for a location with the highest potential.
- The company concerned is a company that wants to appeal to a large group of customers in the center, not to a small group of customers.

2. Data Acquisition And Cleaning

- I extracted the name of the neighborhoods of Toronto from the following link in Wikipedia: ' https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M '

```
url = 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
html = urlopen(url)
soup = BeautifulSoup(html, 'html.parser')
```

```
row = soup.find('tbody') # Extract and return first occurrence of tbody
#Let's fix the data in a few commands below.
list_=row.get_text().split("\n")  # we try to get the writings. we do data separation from spaces.
list_[:20]
```

```
['',
 '',
 'M1ANot assigned',
 '',
 '',
 '',
 'M2ANot assigned',
 '',
 '',
 '',
 'M3ANorth York(Parkwoods)',
 '',
 '',
 '',
 'M4ANorth York(Victoria Village)',
 '',
 '',
 '',
 'M5ADowntown Toronto(Regent Park / Harbourfront)',
 '']
```

```
list_1=[]

for i in list_:
    if i!="":
        if (i[3:])!='Not assigned': #eliminating 'Not assigned'
            list_1.append(i[3:])
print(list_1[:20])
```

```
['North York(Parkwoods)', 'North York(Victoria Village)', 'Downtown Toronto(Regent Park / Harbourfront)', 'North
York(Lawrence Manor / Lawrence Heights)', "Queen's Park / Ontario Provincial Government", 'Etobicoke(Islington Av
enue)', 'Scarborough(Malvern / Rouge)', 'North York(Don Mills)North', 'East York(Parkview Hill / Woodbine Garden
s)', 'Downtown Toronto(Garden District, Ryerson)', 'North York(Glencairn)', 'Etobicoke(West Deane Park / Princess
Gardens / Martin Grove / Islington / Cloverdale)', 'Scarborough(Rouge Hill / Port Union / Highland Creek)', 'Nort
h York(Don Mills)South(Flemingdon Park)', 'East York(Woodbine Heights)', 'Downtown Toronto(St. James Town)', 'Yor
k(Humewood-Cedarvale)', 'Etobicoke(Eringate / Bloordale Gardens / Old Burnhamthorpe / Markland Wood)', 'Scarborou
gh(Guildwood / Morningside / West Hill)', 'East Toronto(The Beaches)']
```

2. Data Acquisition And Cleaning

```
print(len(list_1),len(borough),sep="\n")
```

```
103
104
```

There are a problem. The numbers is not aqual.

I found the missing data. Let's put it in its place.

```
borough.insert(4,"Queen's Park")
borough[:10]
```

```
['North York',
 'North York',
 'Downtown Toronto',
 'North York',
 "Queen's Park",
 'Etobicoke',
 'Scarborough',
 'North York',
 'East York',
 'Downtown Toronto']
```

```
df = pd.DataFrame()
df['Borough']=borough
df.head()
```

	Borough
0	North York
1	North York
2	Downtown Toronto
3	North York
4	Queen's Park

	Borough
0	North York
1	Downtown Toronto
2	Queen's Park
3	Etobicoke
4	Scarborough
5	East York
6	York
7	East Toronto
8	West Toronto
9	Central Toronto
10	Mississauga

2. Data Acquisition And Cleaning

```
borough_coordinate_lat=[]
borough_coordinate_long=[]
for i in borough :
    address = [ (i+',Toronto, CA')]
    geolocator = Nominatim(user_agent="foursquare_agent")
    location = geolocator.geocode(address)
    lat = location.latitude
    lng = location.longitude
    borough_coordinate_lat.append(lat)
    borough_coordinate_long.append(lng)
print(borough_coordinate_lat)
print(borough_coordinate_long)
```

```
[43.7543263, 43.6563221, 43.659659, 43.671459150000004, 43.773077, 43.69
3.653963, 43.653963, 43.6668555]
[-79.44911696639593, -79.3809161, -79.3903399, -79.55249206611668, -79.2
7154, -79.3934918, -79.387207, -79.387207, -79.5879563]
```

In [208]:

```
toronto_data = pd.DataFrame([borough,borough_coordinate_lat,borough_coo
toronto_data.columns=['Borough', 'Latitude','Longitude']
toronto_data
```

Out[208]:

	Borough	Latitude	Longitude
0	North York	43.7543	-79.4491
1	Downtown Toronto	43.6563	-79.3809
2	Queen's Park	43.6597	-79.3903
3	Etobicoke	43.6715	-79.5525
4	Scarborough	43.7731	-79.2578
5	East York	43.7	-79.3325
6	York	43.6791	-79.4912
7	East Toronto	43.6248	-79.3935
8	West Toronto	43.654	-79.3872
9	Central Toronto	43.654	-79.3872
10	Mississauga	43.6669	-79.588

2. Data Acquisition And Cleaning

- After that, I used Foursquare API to get the information about the restaurants.The informations get extracted from the following link:'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(

```
In [265]: toronto_venues = getNearbyVenues(names=toronto_data['Borough'],
                                            latitudes=toronto_data['Latitude'],
                                            longitudes=toronto_data['Longitude']
                                           )
          print("done")
```

```
North York
Downtown Toronto
Queen's Park
Etobicoke
Scarborough
East York
York
East Toronto
West Toronto
Central Toronto
Mississauga
https://api.foursquare.com/v2/venues/explore?&client_id=3DY3IKVXYZDQFTBN3XEPQW2RWHSPCCFM0DRSLPS3YK1Z513O&client_secret=APLJQGA24KCQVXBBRG0ZNJ5EK2TWR21SCTL0VG2WLAXJ5PNY&v=20180605&ll=43.6668555,-79.5879563&radius=30000&limit=10
[{'reasons': {'count': 0, 'items': [{'summary': 'This spot is popular', 'type': 'general', 'reasonName': 'globalInteractionReason'}]}, 'venue': {'id': '4af4709bf964a52079f221e3', 'name': 'Porta Via', 'location': {'address': '5399 Eglinton Avenue West, Unit 104', 'crossStreet': 'btw Renforth & Commerce', 'lat': 43.66344890472386, 'lng': -79.58963815332794, 'labeledLatLngs': [{'label': 'display', 'lat': 43.66344890472386, 'lng': -79.58963815332794}], 'distance': 402, 'cc': 'CA', 'city': 'Toronto', 'state': 'ON', 'country': 'Canada', 'formattedAddress
```

```
In [267]: toronto_venues.head()
```

Out[267]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	43.754326	-79.449117	Earl Bales Park	43.753043	-79.436228	Park
1	North York	43.754326	-79.449117	Nordstrom	43.726054	-79.449443	Clothing Store
2	North York	43.754326	-79.449117	Maryam Hotel	43.766961	-79.401199	Hotel
3	North York	43.754326	-79.449117	Kinka Izakaya	43.760161	-79.409827	Japanese Restaurant
4	North York	43.754326	-79.449117	Crate & Barrel	43.726584	-79.452661	Furniture / Home Store

2. Data Acquisition And Cleaning

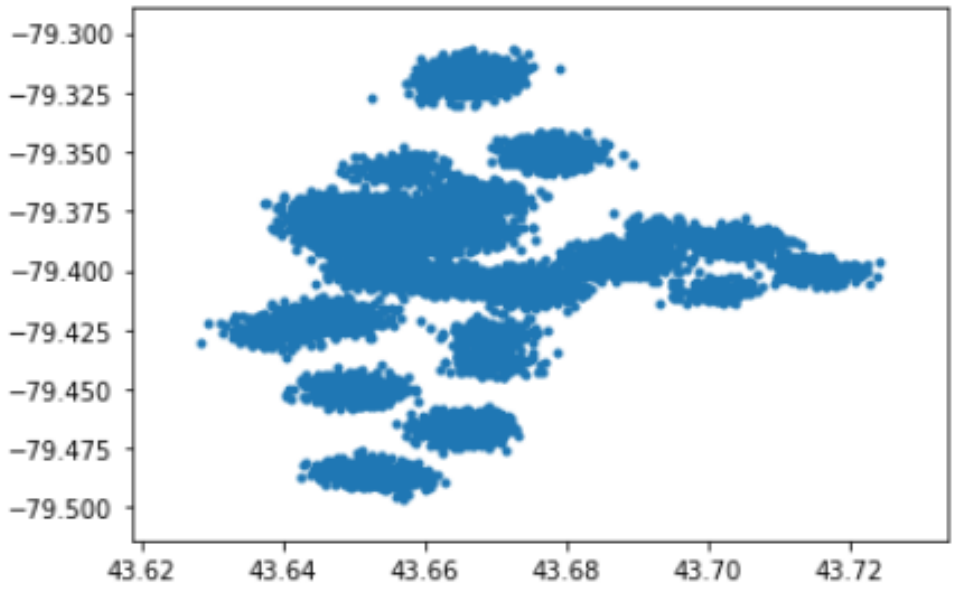
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
22	Central Toronto	43.653963	-79.387207	Byblos Toronto	43.647615	-79.388381	Mediterranean Restaurant
23	Central Toronto	43.653963	-79.387207	Richmond Station	43.651569	-79.379266	American Restaurant
24	Mississauga	43.666855	-79.587956	Mrakovic	43.666041	-79.578850	Eastern European Restaurant
25	Mississauga	43.666855	-79.587956	Bravo Bistro	43.659420	-79.603604	Eastern European Restaurant
26	Mississauga	43.666855	-79.587956	The Keg Steakhouse + Bar - Dixon Road	43.686329	-79.597353	Restaurant

2. Data Acquisition And Cleaning

- In accordance with the acceptance_1 using the restaurant coordinates, I found the places where the customers are most intense. I used cluster_std = 0.004 in the make_blobs method to find them.

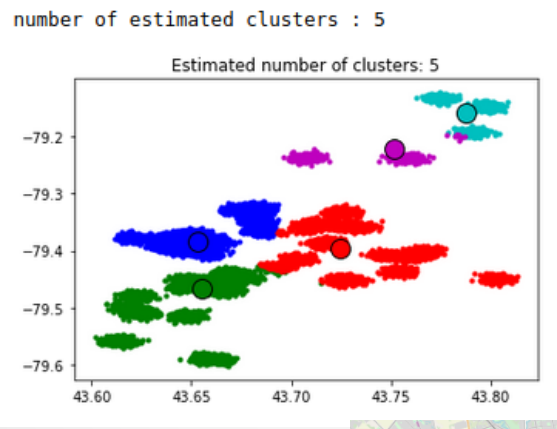
```
X, y = make_blobs(n_samples=15900, centers=list_centers, cluster_std=0.003)
plt.scatter(X[:, 0], X[:, 1], marker='.')
```

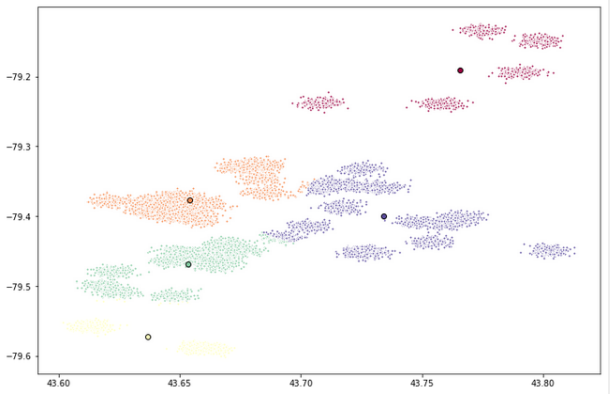
```
<matplotlib.collections.PathCollection at 0x7f08fe186e10>
```

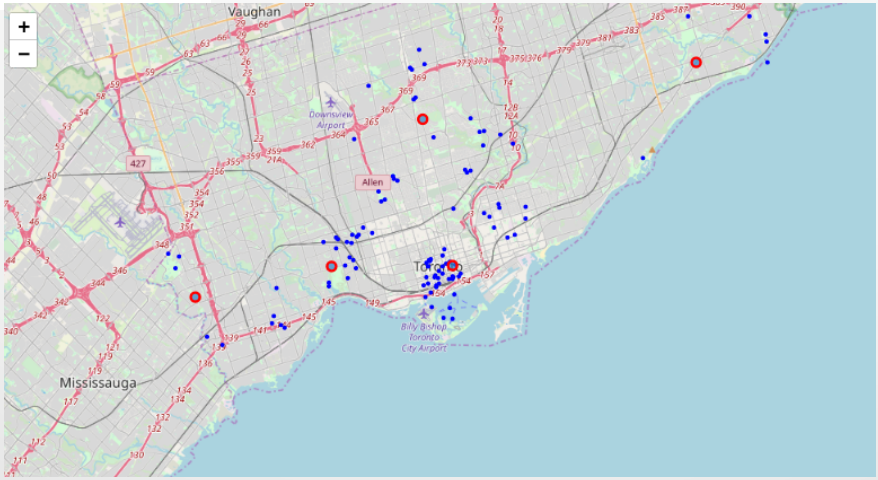


2. Data Acquisition And Cleaning

- I found the most suitable region with the K-means-clustering method by using the places where the customers are the most crowded. I found k in the K-means-clustering method that I used here with estimate_bandwidth.

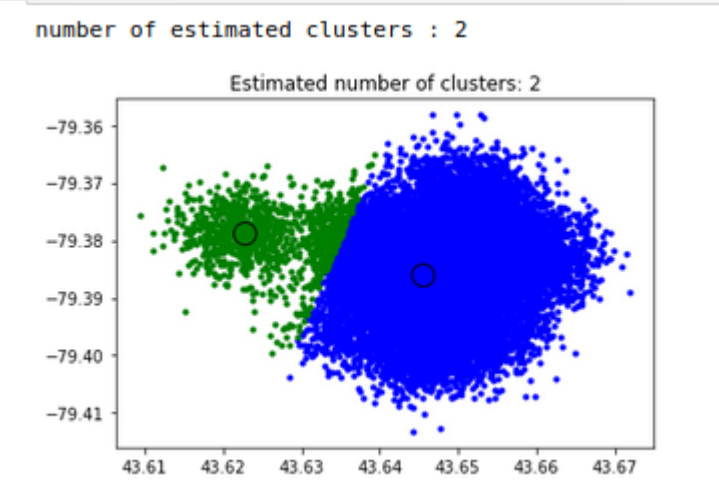


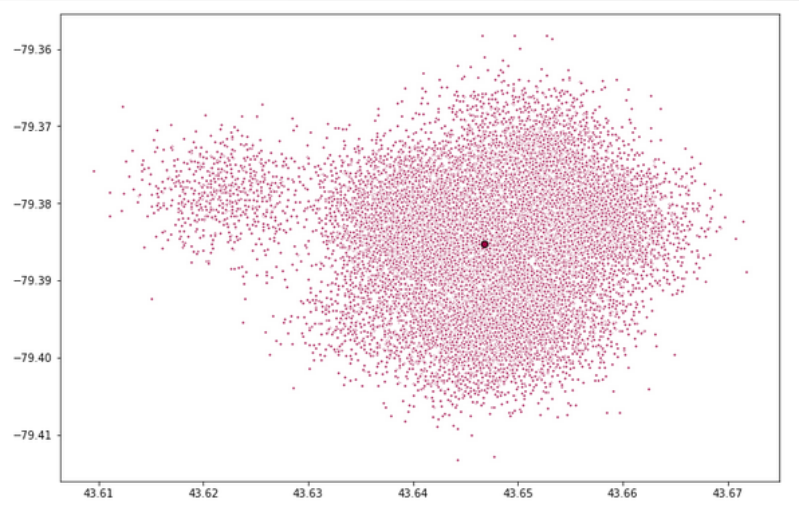




2. Data Acquisition And Cleaning

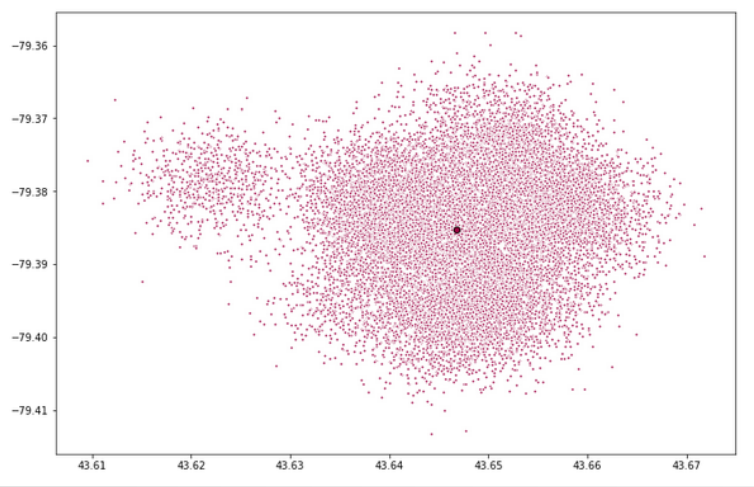
- I applied the K-Means Clustering method again in this region and find the center of this place. so I found the best location of the best region.

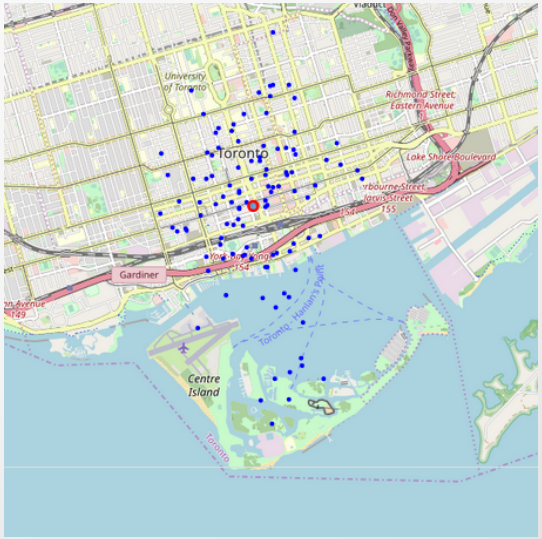




Methodology and ML

- Modelling : To help clients(who want to open an international luxury cuisine restaurant in a Toronto city) find best point for opening an international luxury cuisine restaurant in a Toronto city. I was clustering resourant customers using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.





Results and Discussion

- From the result of first clustering algorithm, it was determined 5 centers were the best choise for opening an international luxury cuisine restaurant.

- Looking at the graphics, we are sure that it is more convenient to open the restaurant in the center of Toronto. Since our company will create a place in the center, so I eliminated it in small and remote places. And I selected the best area according to the chart.

- From the result of second clustering algorithm, it was determined 1 centers were the best choise for opening an international luxury cuisine restaurant.

Results and Discussion

- My advice to the customer is to determine the places that are suitable to open a restaurant near the point I found, and then choose the most suitable place for these points by calculating.

- In addition, customer experience should be used and acceptance should be made more accurately, so that the correct result can be approached more. (unfortunately, I could not benefit from the customer experience in this project.)

- The presence of restaurants that are available in the area and not registered with foursquare will increase the probability of success.

- I also showed the types of restaurants in the area. Finding successful cooks especially in these types will increase the success of the business.

5. Conclusion

- Opening a new international luxury cuisine restaurant is a complex task that can lead to a large monetary loss if not done properly. Thus, extensive research about the area would greatly get bigger the likelihood of the restourant succeding.
- From the project above, I revealed the workflow neccessary for the client to determine what area the restourant should open.for specially, I determined that the optimal location to open an international luxury cuisine restaurant in Toronto shoul be that (k_means3.cluster_centers_) point.