1 Introduction

Recent years, drinking bubble teas has slowly becoming a trend for Toronto citizens due to the cultural diversity of the city. Therefore opening up a bubble tea shop has been a huge market for Chinese investors not only because of the trend but also the low cost of making bubble teas. Hence opening a new bubble tea shop is a difficult task due to its competiveness. If the shop is opened in the area with high competition, the business will not be feasible. If the shop is opened in the area where bubble tea is not so common for the people who lived in that area, the business will not be successful. Therefore, choosing a right location is a huge first step for a successful bubble shop.

1.1 Business Problem

The objective of this capstone project is to find a right location for opening a new bubble tea shop in the city of Toronto, Canada. We will use data from Foursquare API and use Data Science methodologies to analysis the data. We will use clustering in machine learning to come up with a model that will predict that most appropriate location to open a new bubble tea shop.

2 Data acquisition and cleaning

2.1 Data Source

a. list of postal code of Canada from the wiki page:
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

b. geographical coordinates of the neighborhoods from https://cocl.us/Geospatial_data

c. explore various venues using Foursquare's explore API

2.2 Data Cleaning

Since the project only focus on downtown region of Toronto, I would need to come out with a dataframe that only consists information strictly from downtown region. Then I will combine the data from geographical coordinates of the neighborhoods to get the coordinates of each neighbourhood.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
5	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
6	M6G	Downtown Toronto	Christie	43.669542	-79.422564
7	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
8	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752
9	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576
10	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817
11	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049
12	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049
13	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420
14	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529
15	M5W	Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846
16	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-79.367675
17	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
18	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160

The next step is use the Foursquare API to explore various venues from each neighbourhood. There are two dataframes, one contains all the available bubble tea in each neighbourhood, the other one contains all Chinese related shops. Once I got all the venues from all the neighbourhoods, I combined all the venues to single dataframe and only contain the information I need for analysis.

```
In [53]: downtown_bubbletea_venues.head()
```

Out[53]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.654260	-79.360636	Palgong Tea	43.650090	-79.361374	Bubble Tea Shop
1	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Bubble Lee	43.665025	-79.384499	Bubble Tea Shop
2	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	琉璃鯨 The Whale Tea	43.663573	-79.384208	Bubble Tea Shop
3	Garden District, Ryerson	43.657162	-79.378937	Happy Fruit Bubble Tea	43.656339	-79.380783	Bubble Tea Shop
4	Garden District, Ryerson	43.657162	-79.378937	Real Fruit Bubble Tea	43.655994	-79.380577	Bubble Tea Shop

```
In [54]: downtown_chinese_venues.head()
```

Out[54]:

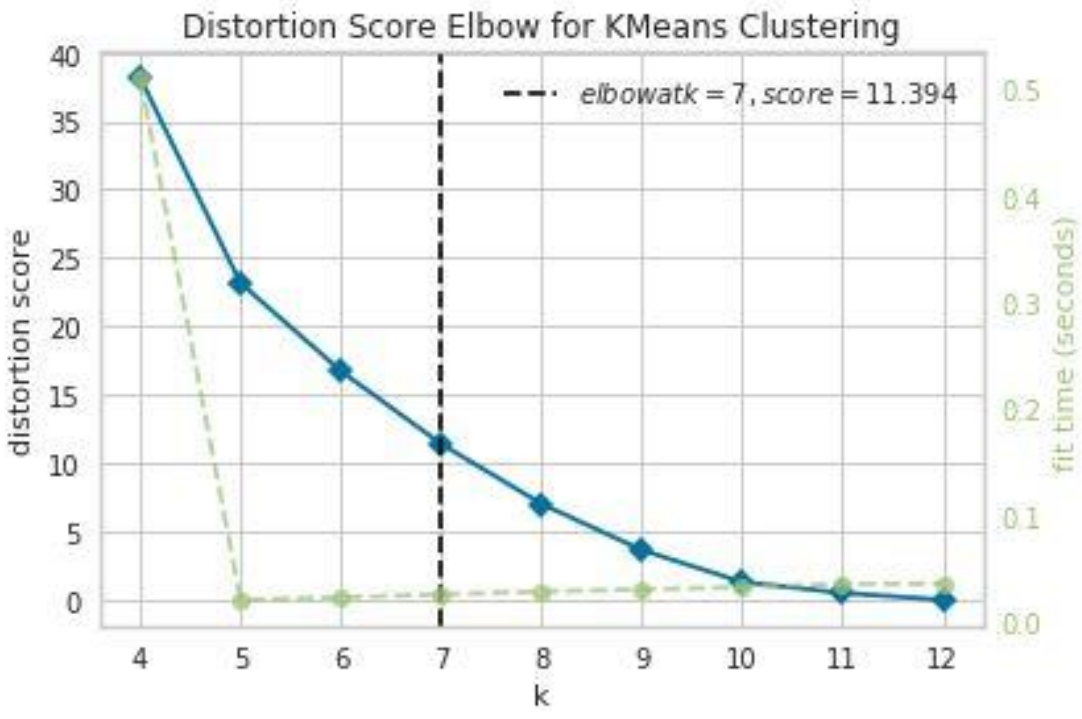
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Garden District, Ryerson	43.657162	-79.378937	Good View Chinese Restaurant	43.656791	-79.376139	Chinese Restaurant
1	Garden District, Ryerson	43.657162	-79.378937	Winner BBQ Chinese Cuisine DT 賊好吃的烤肉飯	43.659320	-79.382794	Chinese Restaurant
2	Garden District, Ryerson	43.657162	-79.378937	Yueh Tung Chinese Restaurant	43.655281	-79.385337	Chinese Restaurant
3	Garden District, Ryerson	43.657162	-79.378937	Ho Jan 好棧粉麵小廚	43.661483	-79.381111	Chinese Restaurant
4	St. James Town	43.651494	-79.375418	Good View Chinese Restaurant	43.656791	-79.376139	Chinese Restaurant

Out[76]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Central Bay Street	4	7	7	43.657952	-79.387383
Church and Wellesley	6	1	9	43.665860	-79.383160
Commerce Court, Victoria Hotel	0	1	0	43.648198	-79.379817
First Canadian Place, Underground city	5	2	2	43.648429	-79.382280
Garden District, Ryerson	2	4	8	43.657162	-79.378937
Harbourfront East, Union Station, Toronto Islands	0	1	0	43.640816	-79.381752
Kensington Market, Chinatown, Grange Park	1	18	12	43.653206	-79.400049
Richmond, Adelaide, King	3	6	3	43.650571	-79.384568
St. James Town	0	1	0	43.651494	-79.375418
St. James Town, Cabbagetown	0	1	0	43.667967	-79.367675
Toronto Dominion Centre, Design Exchange	5	2	0	43.647177	-79.381576
University of Toronto, Harbord	5	4	1	43.662696	-79.400049
Regent Park, Harbourfront	0	0	1	43.654260	-79.360636
Queen's Park, Ontario Provincial Government	0	0	2	43.662301	-79.389494
Berczy Park	0	0	0	43.644771	-79.373306
Christie	0	0	0	43.669542	-79.422564
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	0	0	0	43.628947	-79.394420
Rosedale	0	0	0	43.679563	-79.377529
Stn A PO Boxes	0	0	0	43.646435	-79.374846

3 Predictive Modelling

K-mean clustering was used to cluster the above datatframe. First the elbow method was used to determine the appropriate k value.



Then k = 7 was used to do the modeling.

Out[69]:

	Cluster Labels	Chinese Related Shops Count	Latitude_x	Longitude_x	Bubbletea Shops Count	Latitude_y	Longitude_y
Central Bay Street	4	7	43.657952	-79.387383	7	43.657952	-79.387383
Church and Wellesley	6	1	43.665860	-79.383160	9	43.665860	-79.383160
Commerce Court, Victoria Hotel	0	1	43.648198	-79.379817	0	43.648198	-79.379817
First Canadian Place, Underground city	5	2	43.648429	-79.382280	2	43.648429	-79.382280
Garden District, Ryerson	2	4	43.657162	-79.378937	8	43.657162	-79.378937
Harbourfront East, Union Station, Toronto Islands	0	1	43.640816	-79.381752	0	43.640816	-79.381752
Kensington Market, Chinatown, Grange Park	1	18	43.653206	-79.400049	12	43.653206	-79.400049
Richmond, Adelaide, King	3	6	43.650571	-79.384568	3	43.650571	-79.384568
St. James Town	0	1	43.651494	-79.375418	0	43.651494	-79.375418
St. James Town, Cabbagetown	0	1	43.667967	-79.367675	0	43.667967	-79.367675
Toronto Dominion Centre, Design Exchange	5	2	43.647177	-79.381576	0	43.647177	-79.381576
University of Toronto, Harbord	5	4	43.662696	-79.400049	1	43.662696	-79.400049
Regent Park, Harbourfront	0	0	43.654260	-79.360636	1	43.654260	-79.360636
Queen's Park, Ontario Provincial Government	0	0	43.662301	-79.389494	2	43.662301	-79.389494
Berczy Park	0	0	43.644771	-79.373306	0	43.644771	-79.373306
Christie	0	0	43.669542	-79.422564	0	43.669542	-79.422564
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	0	0	43.628947	-79.394420	0	43.628947	-79.394420
Rosedale	0	0	43.679563	-79.377529	0	43.679563	-79.377529
Stn A PO Boxes	0	0	43.646435	-79.374846	0	43.646435	-79.374846

4.2 Examination and Conclusion:

Let's examining cluster 0.

```
In [89]: final.loc[final['Cluster Labels']==0]
```

Out[89]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Commerce Court, Victoria Hotel	0	1	0	43.648198	-79.379817
Harbourfront East, Union Station, Toronto Islands	0	1	0	43.640816	-79.381752
St. James Town	0	1	0	43.651494	-79.375418
St. James Town, Cabbagetown	0	1	0	43.667967	-79.367675
Regent Park, Harbourfront	0	0	1	43.654260	-79.360636
Queen's Park, Ontario Provincial Government	0	0	2	43.662301	-79.389494
Berczy Park	0	0	0	43.644771	-79.373306
Christie	0	0	0	43.669542	-79.422564
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	0	0	0	43.628947	-79.394420
Rosedale	0	0	0	43.679563	-79.377529
Stn A PO Boxes	0	0	0	43.646435	-79.374846

This cluster contains the least amount of competition on bubble tea shops. There aren't many Chinese related shops opening in the areas, that means the neighbourhoods are not populated with Chinese people. So opening a bubble tea shop in these areas is not really a good ideal.

Let's examining Cluster 6

```
In [95]: final.loc[final['Cluster Labels']==6]
```

Out[95]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Church and Wellesley	6	1	9	43.66586	-79.38316

Cluster 6 contains neighbourhoods that have too much competition on bubble tea shops and also there aren't many Chinese related shops opening in the areas either. Therefore opening a bubble tea shop in this cluster is not ideal.

Cluster 1, 2, 3, 4 and 5 are good candidates for opening a bubble tea shop.

```
In [90]: final.loc[final['Cluster Labels']==1]
```

Out[90]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Kensington Market, Chinatown, Grange Park	1	18	12	43.653206	-79.400049

```
In [91]: final.loc[final['Cluster Labels']==2]
```

Out[91]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Garden District, Ryerson	2	4	8	43.657162	-79.378937

```
In [92]: final.loc[final['Cluster Labels']==3]
```

Out[92]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Richmond, Adelaide, King	3	6	3	43.650571	-79.384568

```
In [93]: final.loc[final['Cluster Labels']==4]
```

Out[93]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
Central Bay Street	4	7	7	43.657952	-79.387383

```
In [94]: final.loc[final['Cluster Labels']==5]
```

Out[94]:

	Cluster Labels	Chinese Related Shops Count	Bubbletea Shops Count	Latitude	Longitude
First Canadian Place, Underground city	5	2	2	43.648429	-79.382280
Toronto Dominion Centre, Design Exchange	5	2	0	43.647177	-79.381576
University of Toronto, Harbord	5	4	1	43.662696	-79.400049

Out of these clusters, cluster 3 would be the best candidate to open a bubble tea shop. The cluster contains good amount of Chinese related shops which means the neighbourhoods involve a lot of Chinese related activities. At the same time they are not packed with bubble tea shops which mean less competition. Since the cluster only contains a neighbourhood which is “Richmond, Adelaide, King”, this would be the neighbourhood to open a bubble tea shop!