PROBLEM & BACKGROUND

Toronto is one of the most famous places in the world. We want to explore how to establish a new business from a market point of view regarding food, accommodation, beautiful places, and many more.When it comes to a new business, it is important to explore the competition in the area, especially for a busy area. As we all known, establishing a good business idea is one of the pillars of the success and the people most often fails if there is not enough researches from a market prospective, like friendly environment. Every city is unique in their own way and give something new. And now the information is so common regarding location of every place around the world on your fingertips which make it easier to explore. Therefore, this project designed to explore the existed business around the Toronto on the basis of available information, which will generate the new business ideas based on the research.

DATA DESCRIPTION

For this problem, we will get the services of Foursquare API to explore the data, in terms of their neighborhoods. The data also include the information about the places around each neighborhood like restaurants, hotels, coffee shops, parks. We selected one Borough from each city to analyze their neighborhoods. Data source: 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'

METHODOLOGY

We will use machine learning technique, "Clustering", K-means to segment the neighborhoods with similar objects on the basis of each neighborhood data and explore more about these clusters based on the model. This will help to locate the target business's areas and hubs, and then we can judge the best location to start a new business.
k-means clustering is rather easy to apply to even large data sets, particularly when using heuristics such as Lloyd's algorithm. It has been successfully used in market segmentation, computer vision, and astronomy among many other domains. It often is used as a preprocessing step for other algorithms, for example to find a starting configuration.

EXPLORATION

For our target country: Toronto, we first have to clean and tidy our dirty data, including extracting table from Wikipedia page by beautifulsoup method. Then we plan to explore the data according to our requirements. In the processing phase, which

applied multiple steps, eliminating "Not assigned" values, combine neighborhoods. For data verification and further exploration, we use Foursquare API to get the coordinates of Toronto and explore its neighborhoods. The neighborhoods are further characterized as venues and venue categories.We cluster our locations based on their characters and explore each cluster after our kmeans methodology. This appoarch will help us to generate the whole picture of the business existed in Toronto place and help us to decide our ultimate business location and type.

1. First, Let's get a whole picture of our clean dataset:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Port Union,Rouge Hill	43.784535	-79.160497
2	M1E	Scarborough	West Hill,Morningside,Guildwood	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Scarborough Village West,Cliffcrest,Cliffside	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848

As we can see that the Borough neighbourhood are listed above with latitude and longitude.

2. Then let get the Toronto latitude lontitude coordinate:

The geograpical coordinate of Toronto are 43.653963, -79.387207.

3. Create map of New York using latitude and longitude values and add markers to map:

We can find that the business distributed evenly, except a point: the place near the airport. We can figure out that most of the business located near the airport, let get another zoom in graph to better identify the locations:
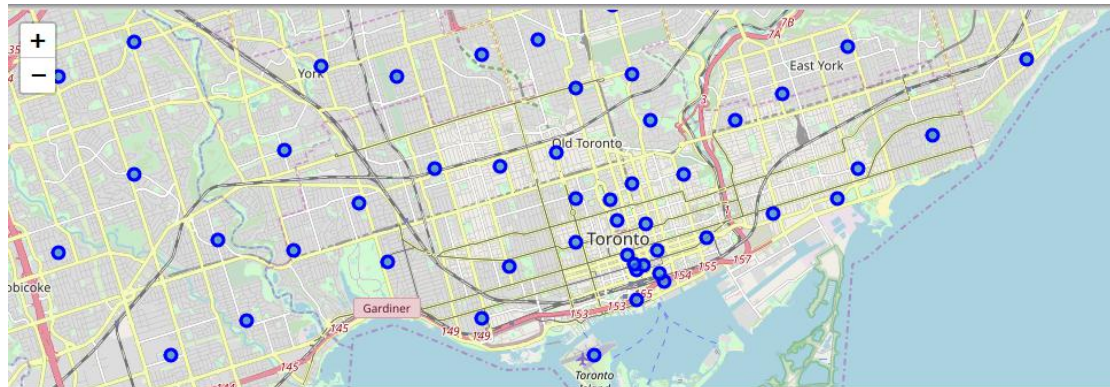
We can also figure out that the more roads with more traffic the more store will be.

4. Log in Foursquare and get the Foursquare ID and secret number and print the Studio district name, latitude and longitude:

Latitude and longitude values of Studio District are 43.6595255, -79.340923.

5. Getting the top 50 venues that are in Studio District within a radius of 200 meters:

{'meta': {'code': 200, 'requestId': '5d74e725446ea6002c70265a'},
 'response': {'headerLocation': 'Leslieville',
   'headerFullLocation': 'Leslieville, Toronto',
   'headerLocationGranularity': 'neighborhood',
   'totalResults': 15,
   'suggestedBounds': {'ne': {'lat': 43.6613255018, 'lng': -79.33843958229481},
    'sw': {'lat': 43.6577254982, 'lng': -79.3434064177052}},
   'groups': [{'type': 'Recommended Places',
     'name': 'recommended',
     'items': [{'reasons': {'count': 0,
        'items': [{'summary': 'This spot is popular',
          'type': 'general',
          'reasonName': 'globalInteractionReason'}]},
       'venue': {'id': '4ad7e958f964a520001021e3',

      'name': "Ed's Real Scoop",
       'location': {'address': '920 Queen St. E',
         'crossStreet': 'btwn Logan Ave. & Morse St.',
         'lat': 43.660655832455014,
   'lng': -79.3420187548006,
......
      'referralId': 'e-0-4bda1bd03904a593c42b459e-13'},
{'reasons': {'count': 0,
      'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
    'venue': {'id': '54318dcb498e7edd4e17bceb',
      'name': 'Good Neighbour',
      'location': {'address': '935 Queen St East',
         'lat': 43.6610004575794,
         'lng': -79.34029809683763,
         'labeledLatLngs': [{'label': 'display',
            'lat': 43.6610004575794,
            'lng': -79.34029809683763}],
         'distance': 171,
         'cc': 'CA',
         'city': 'Toronto',
         'state': 'ON',
         'country': 'Canada',
         'formattedAddress': ['935 Queen St East', 'Toronto ON', 'Canada']},
      'categories': [{'id': '4bf58dd8d48988d103951735',
         'name': 'Clothing Store',
         'pluralName': 'Clothing Stores',
         'shortName': 'Apparel',
         'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/apparel_',
            'suffix': '.png'},
         'primary': True}],
      'photos': {'count': 0, 'groups': []}},
    'referralId': 'e-0-54318dcb498e7edd4e17bceb-14'}]}]}}

6. Extracts the category of the venue:

	name	categories	lat	lng
0	Ed's Real Scoop	Ice Cream Shop	43.660656	-79.342019
1	Leslieville Pumps	Sandwich Place	43.660892	-79.340626
2	Queen Books	Bookstore	43.660651	-79.342267
3	Purple Penguin Cafe	Café	43.660501	-79.342565
4	Leslieville Cheese Market	Cheese Shop	43.660546	-79.342302

We can identify that the categories are: Ice Cream Shop, Sandwich Place, Bookstore, Cafe and Cheese shop. Four of the five are related to the food and one of them is related to the education area. Three of the food area are related to the fast food, including ice cream, cafe and sandwich. We can generate that most people will tend to get fast food in the revenue. We considered that fast food industry is one of our top choice when it come to a new business in Toronoto.

7. 15 venues were returned by Foursquare. Then let generating venues within Studio District neighborhood:
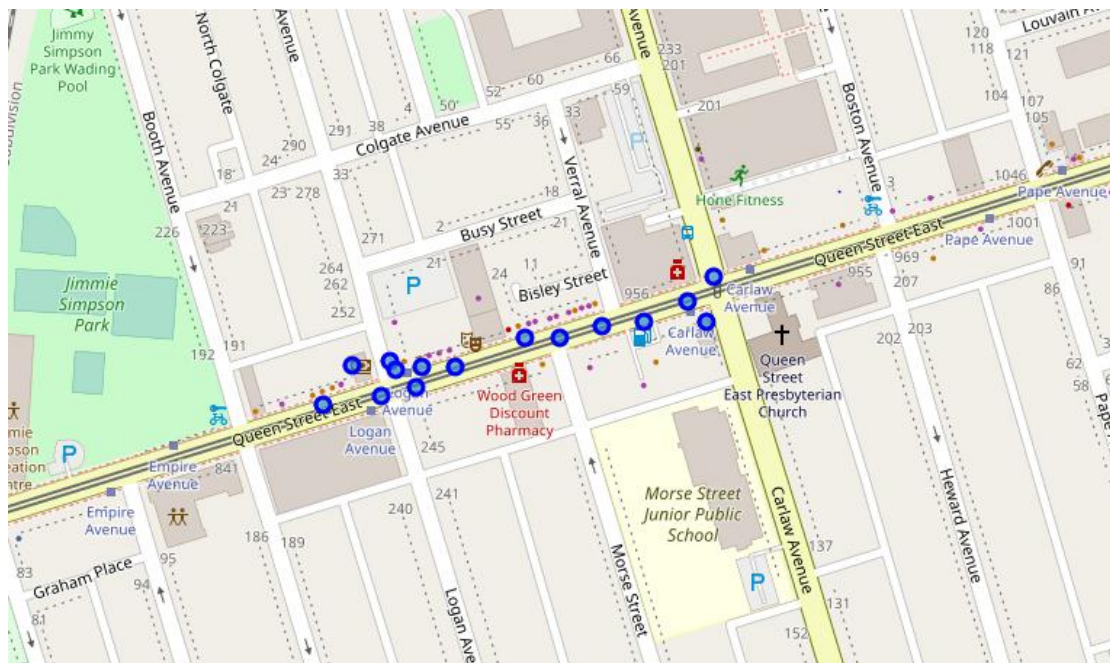
Exploring Neighbourhoods in Toronto:

| Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bathurst Quay,South Niagara,Island airport,CN Tower,Railway Lands,King and Spadina,Harbourfront West	15	15	15	15	15	15
Berczy Park	50	50	50	50	50	50
Business Reply Mail Processing Centre 969 Eastern	19	19	19	19	19	19
Cabbagetown,St. James Town	44	44	44	44	44	44
Central Bay Street	50	50	50	50	50	50
Christie	16	16	16	16	16	16
Church and Wellesley	50	50	50	50	50	50
Davisville	38	38	38	38	38	38
Davisville North	9	9	9	9	9	9
Deer Park,Summerhill West,South Hill,Rathnelly,Forest Hill SE	16	16	16	16	16	16

There are 216 uniques categories.

	Neighbourhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop
0	Bathurst Quay,South Niagara,Island airport,CN ...	0.066667	0.066667	0.066667	0.066667	0.2	0.133333	0.000000	0.00	0.00	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
1	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.02	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.020000
2	Business Reply Mail Processing Centre 969 Eastern	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.00	0.00	0.000000	0.00	0.0000	0.052632	0.000000	0.0000	0.000000
3	Cabbagetown,St. James Town	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.022727	0.00	0.00	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
4	Central Bay Street	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.020000	0.00	0.00	0.02	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
5	Christie	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.00	0.00	0.000000	0.00	0.0625	0.000000	0.000000	0.0625	0.000000
6	Church and Wellesley	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.00	0.00	0.020000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
7	Davisville	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.00	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
8	Davisville North	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.00	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
9	Deer Park,Summerhill West,South Hill,Rathnelly...	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.062500	0.00	0.00	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.062500
10	Design Exchange,Toronto Dominion Centre	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.020000	0.00	0.02	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000
11	Dovercourt Village,Dufferin	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.00	0.00	0.000000	0.00	0.0000	0.000000	0.000000	0.0000	0.000000

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bathurst Quay,South Niagara,Island airport,CN ...	Airport Service	Airport Terminal	Harbor / Marina	Sculpture Garden	Airport Food Court
1	Berczy Park	Coffee Shop	Cocktail Bar	Steakhouse	Farmers Market	Bakery
2	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Auto Workshop	Pizza Place	Restaurant
3	Cabbagetown,St. James Town	Coffee Shop	Café	Restaurant	Italian Restaurant	Pub
4	Central Bay Street	Coffee Shop	Café	Bubble Tea Shop	Ice Cream Shop	Italian Restaurant

We can identify that the most common renue are as picture above. We can consider cafe or other fast food around airport or other street listed above.

8. Based on what we have, we cluster our locations into 5 clusters:
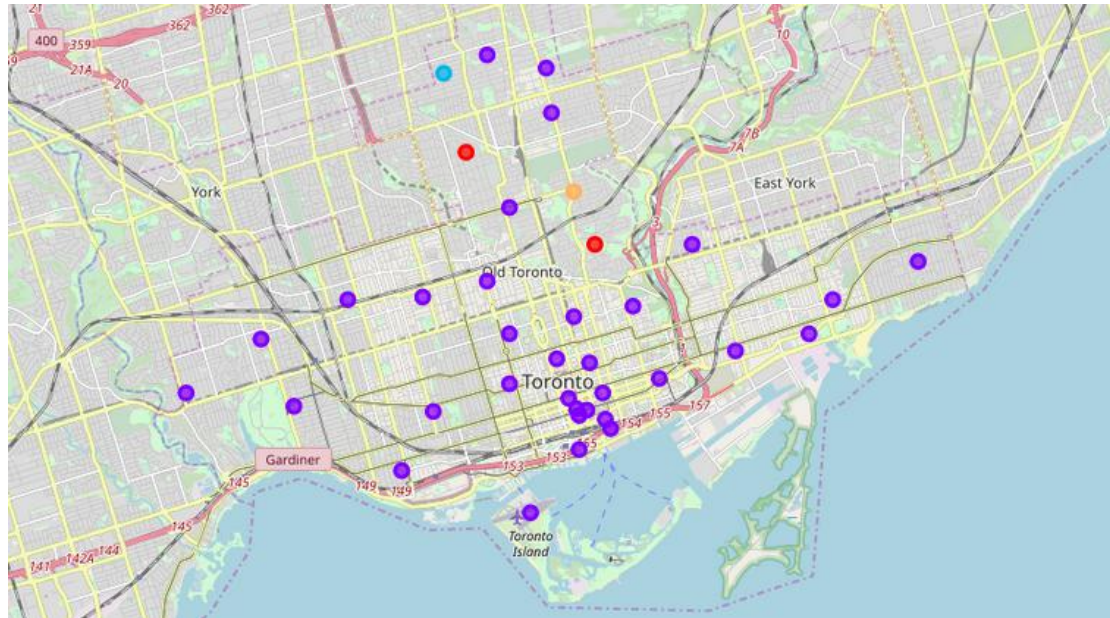
The purple one represent the most of our locations. For this cluster, let make a more detail analysis:

	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Downtown Toronto	43.679563	-79.377529	0	Park	Playground	Trail	Building	Department Store
23	Central Toronto	43.696948	-79.411307	0	Park	Trail	Sushi Restaurant	Jewelry Store	Yoga Studio

We can see that downtown Toronto and central Toronto are in Borough for cluster1. Park is the most common revenue, then is trail, playground and restaurant. We consider that for this cluster, entertainment is the main function. If we have located this area, we can consider to build an entertainment business.

	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	East Toronto	43.676357	-79.293031	1	Neighborhood	Health Food Store	Trail	Pub	Other Great Outdoors
1	East Toronto	43.679557	-79.352188	1	Greek Restaurant	Coffee Shop	Italian Restaurant	Bookstore	Furniture / Home Store
2	East Toronto	43.668999	-79.315572	1	Sandwich Place	Park	Pizza Place	Ice Cream Shop	Burger Joint
3	East Toronto	43.659526	-79.340923	1	Café	Coffee Shop	Italian Restaurant	American Restaurant	Bakery
5	Central Toronto	43.712751	-79.390197	1	Dance Studio	Clothing Store	Restaurant	Food & Drink Shop	Hotel
6	Central Toronto	43.715383	-79.405678	1	Coffee Shop	Clothing Store	Sporting Goods Shop	Burger Joint	Salon / Barbershop
7	Central Toronto	43.704324	-79.388790	1	Pizza Place	Sandwich Place	Dessert Shop	Sushi Restaurant	Gym
9	Central Toronto	43.686412	-79.400049	1	Coffee Shop	Pub	Liquor Store	Supermarket	Sushi Restaurant
11	Downtown Toronto	43.667967	-79.367675	1	Coffee Shop	Café	Restaurant	Italian Restaurant	Pub
12	Downtown Toronto	43.665860	-79.383160	1	Coffee Shop	Burger Joint	Gastropub	Gay Bar	Restaurant
13	Downtown Toronto	43.654260	-79.360636	1	Coffee Shop	Park	Bakery	Pub	Café
14	Downtown Toronto	43.657162	-79.378937	1	Coffee Shop	Café	Fast Food Restaurant	Ramen Restaurant	Clothing Store
15	Downtown Toronto	43.651494	-79.375418	1	Coffee Shop	Gastropub	Hotel	Café	Restaurant
16	Downtown Toronto	43.644771	-79.373306	1	Coffee Shop	Cocktail Bar	Steakhouse	Farmers Market	Bakery
17	Downtown Toronto	43.657952	-79.387383	1	Coffee Shop	Café	Bubble Tea Shop	Ice Cream Shop	Italian Restaurant
18	Downtown Toronto	43.650571	-79.384568	1	Coffee Shop	Café	Steakhouse	Bar	American Restaurant
19	Downtown Toronto	43.640816	-79.381752	1	Coffee Shop	Café	Hotel	Plaza	Aquarium
20	Downtown Toronto	43.647177	-79.381576	1	Coffee Shop	Café	Restaurant	Gym	Deli / Bodega
21	Downtown Toronto	43.648198	-79.379817	1	Coffee Shop	Hotel	Café	Restaurant	Gym
24	Central Toronto	43.672710	-79.405678	1	Sandwich Place	Café	Coffee Shop	Pizza Place	Liquor Store

For our second cluster listed above, we can consider to run a coffee shop. While at the same time we should brand ourselves, because too many cafe here will lead to a intensive competition. If we want less competition, then we can try restuarant or gym nearby.

```
#cluster 3
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
22	Central Toronto	2	Pool	Garden	Yoga Studio	Department Store	Event Space

```
#cluster 4
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Central Toronto	3	Park	Swim School	Bus Line	Yoga Studio	Diner

```
#cluster 5
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Central Toronto	4	Playground	Tennis Court	Department Store	Falafel Restaurant	Event Space

For the rest of clusters, we can find that the entertainment is the focus in these area. Therefore, we can develop some better entertainment activities in this area, or we can start some fast food business around which lead to less competiton in this area. To decide which locations and business type to choose, we should futher discuss the business strategy and type. Then combine the specific business type and mission of our new business, we can choose a suitable location.

Conclusion

We decide to establish our business in a location near the busy street, around the airport. We divided our choosed locations into 5 clusters and figure out the top 5 most common venue in the locations. We find most of them are related to entertainment. We can try to establish a fastfood resturant based on the current situation, which will reduce the competition and people tend to get food during the activity. While the entertainment business can also be established only if we have a blue ocean strategy which can attract more people for our the business.