CAPSTONE PROJECT-BATTLE OF NEIGHBORHOODS

Samira Amirli
IBM Professional Data Science Course

Background Information

- Various cities in every part of the world contain various kinds of venues and their categories in itself which inform about the specific local cultures of those cities to the mankind. Let me introduce the most important cities which attract the people's attention from all over the world, such as New York City which is the main capital, international and most populated city of the United States, and Toronto which is the main financial capital, and also most populated city of Canada.

Business Problem

- I will choose the main boroughs from abovementioned two countries, such that I will choose Brooklyn from New York, and York from Toronto, which are one of most important and famous places of their corresponding cities. After that, according to customers' answers to the survey questions, I will recommend them which one of the places to visit.

- **Target Audience:** older retired people above age of 50-60 years.

Survey Questions

- 1. How old are you and what is the purpose of your travelling?
- 2. Are you travelling with your family/friends or alone?
- 3. Are you interested in entertainment, such as going to pubs, clubs, cinema or others?
- 4. Are you interested in doing various sport types and do you want to go for some winter sports?
- 5. Would you like mainly to visit the old places, such as not investigated remote places and historical museums?

Objective

- The purpose of this project is to categorically segment the neighborhoods of New York City and Toronto (Brooklyn, York respectively) into major clusters and examine those clusters to find the appropriate travel places considering the preferences and tastes of our customers who are in the category of above 50- 60 years.

Data Source

- The first dataset of New York City consists of 5 boroughs and neighborhoods in each borough, and also geometric coordinates, such as latitude and longitude coordinates of each neighborhood. The link to this dataset can be found easily on the web and is the following: https://geo.nyu.edu/catalog/nyu_2451_34572
- The second dataset of Toronto city consists of different boroughs, neighborhood in each borough and their respective postal codes. The link to this dataset is taken from Wikipedia page and is the following: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Also the data for the geographical coordinates for each neighborhood in Toronto are collected, thus, the following csv file contains those data: https://cocl.us/Geospatial_data
- The Foursquare API will be utilized to obtain the geographical location data, such as for Brooklyn in New York, and for York in Toronto. These datas will be used to explore the venues in the neighbourhoods of Brooklyn and York,respectively.The venues will provide the categories needed for my dataset analysis : https://developer.foursquare.com/

Methodology

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M6C	York	Humewood-Cedarvale	43.693781	-79.428191
1	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512
2	M6M	York	Del Ray , Mount Dennis , Keelsdale and Silvert...	43.691116	-79.476013
3	M6N	York	Runnymede , The Junction North	43.673185	-79.487262
4	M9N	York	Weston	43.706876	-79.518188

Brooklyn before and after clustering

York before and after clustering

Exploratory Data Analysis

- 'getNearbyVenues' function is created, which functions loop through all the neighborhoods of Brooklyn and sets an API request URL by defining radius = 500 and LIMIT = 50.

```
In [18]: print(brooklyn_venues.shape)
         brooklyn_venues.head()
```

```
(2091, 7)
```

Out[18]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	The Bookmark Shoppe	40.624577	-74.030562	Bookstore

```
In [46]: print(york_venues.shape)
         york_venues.head()
```

```
(17, 7)
```

Out[46]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Humewood-Cedarvale	43.693781	-79.428191	Cedarvale Park	43.692535	-79.428705	Field
1	Humewood-Cedarvale	43.693781	-79.428191	Cedarvale Ravine	43.690188	-79.426106	Trail
2	Humewood-Cedarvale	43.693781	-79.428191	Phil White Arena	43.691303	-79.431761	Hockey Arena
3	Humewood-Cedarvale	43.693781	-79.428191	Prince's Parkette	43.697385	-79.424704	Park
4	Caledonia-Fairbanks	43.689026	-79.453512	Nairn Park	43.690654	-79.456300	Park

Feature Engineering

- 'Onehot encoding' function is applied for the purpose of converting the categorical variables.

```
print(brooklyn_onehot.shape)
brooklyn_onehot.head()
```

```
(2091, 263)
```

Out[47]:

	Yoga Studio	Accessories Store	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field
0	0	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	1	0	0	0	0
2	0	0	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	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
print(york_onehot.shape)
york_onehot.head()
```

```
(17, 14)
```

Out[50]:

	Neighborhood	Bus Line	Caribbean Restaurant	Convenience Store	Fast Food Restaurant	Field	Grocery Store	Hockey Arena	Park	Pool	Restaurant	Sandwich Place	Trail	Women's Store
0	Humewood-Cedarvale	0	0	0	0	1	0	0	0	0	0	0	0	0
1	Humewood-Cedarvale	0	0	0	0	0	0	0	0	0	0	0	1	0
2	Humewood-Cedarvale	0	0	0	0	0	0	1	0	0	0	0	0	0
3	Humewood-Cedarvale	0	0	0	0	0	0	0	1	0	0	0	0	0
4	Caledonia-Fairbanks	0	0	0	0	0	0	0	1	0	0	0	0	0

Top 10 most common venues in both Brooklyn and York respectively.

```
for ind in np.arange(brooklyn_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(brooklyn_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[52]:

	Neighborhood	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	Bath Beach	Chinese Restaurant	Pharmacy	Gas Station	Bubble Tea Shop	Italian Restaurant	Fast Food Restaurant	Sushi Restaurant	Check Cashing Service	Kebab Restaurant	Sports Bar
1	Bay Ridge	Spa	Italian Restaurant	Pizza Place	Chinese Restaurant	Greek Restaurant	Grocery Store	Ice Cream Shop	Hookah Bar	American Restaurant	Lounge
2	Bedford Stuyvesant	Pizza Place	Coffee Shop	Café	Bar	Fried Chicken Joint	Bagel Shop	Cocktail Bar	Gift Shop	Gourmet Shop	Boutique
3	Bensonhurst	Sushi Restaurant	Bakery	Ice Cream Shop	Chinese Restaurant	Italian Restaurant	Donut Shop	Factory	Road	Noodle House	Liquor Store
4	Bergen Beach	Harbor / Marina	Baseball Field	Athletics & Sports	Playground	Donut Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field

```
for ind in np.arange(york_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(york_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[56]:

	Neighborhood	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	Caledonia-Fairbanks	Women's Store	Pool	Park	Trail	Sandwich Place	Restaurant	Hockey Arena	Grocery Store	Field	Fast Food Restaurant
1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	Sandwich Place	Restaurant	Fast Food Restaurant	Convenience Store	Women's Store	Trail	Pool	Park	Hockey Arena	Grocery Store
2	Humewood-Cedarvale	Trail	Park	Hockey Arena	Field	Women's Store	Sandwich Place	Restaurant	Pool	Grocery Store	Fast Food Restaurant
3	Runnymede, The Junction North	Grocery Store	Convenience Store	Caribbean Restaurant	Bus Line	Women's Store	Trail	Sandwich Place	Restaurant	Pool	Park
4	Weston	Park	Convenience Store	Women's Store	Trail	Sandwich Place	Restaurant	Pool	Hockey Arena	Grocery Store	Field

Machine Learning: K-Means

- K-means is an unsupervised machine learning algorithm which groups certain data points into certain user-specified number of k clusters based on their similar characteristics or features.

- Set the number of clusters to 3 and run the k-means to cluster each neighborhood into 3 clusters, in both Brooklyn and York respectively. My result will segment each neighborhood in both Brooklyn and York based upon the most common venues in its near surrounding.

Machine Learning: k= 3 clusters

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	0	Spa	Italian Restaurant	Pizza Place	Chinese Restaurant	Greek Restaurant	Grocery Store	Ice Cream Shop	Hookah Bar	American Restaurant	Lounge
1	Brooklyn	Bensonhurst	40.611009	-73.995180	0	Sushi Restaurant	Bakery	Ice Cream Shop	Chinese Restaurant	Italian Restaurant	Donut Shop	Factory	Road	Noodle House	Liquor Store
2	Brooklyn	Sunset Park	40.645103	-74.010316	2	Bank	Latin American Restaurant	Pizza Place	Bakery	Mexican Restaurant	Gym	Fried Chicken Joint	Pharmacy	Deli / Bodega	Mobile Phone Shop
3	Brooklyn	Greenpoint	40.730201	-73.954241	0	Coffee Shop	Bar	Cocktail Bar	Yoga Studio	Spa	Café	Bakery	Record Shop	Pizza Place	French Restaurant
4	Brooklyn	Gravesend	40.595260	-73.973471	2	Pizza Place	Bakery	Bus Station	Lounge	Diner	Chinese Restaurant	Breakfast Spot	Furniture / Home Store	Bar	Electronics Store

Postal Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
M6C	York	Humewood-Cedarvale	43.693781	-79.428191	1.0	Trail	Tennis Court	Hockey Arena	Field	Women's Store	Turkish Restaurant	Skating Rink	Sandwich Place	Restaurant	Pool
M6E	York	Caledonia-Fairbanks	43.689026	-79.453512	0.0	Park	Women's Store	Pool	Turkish Restaurant	Trail	Tennis Court	Skating Rink	Sandwich Place	Restaurant	Pizza Place
M6M	York	Del Ray , Mount Dennis , Keelsdale and Silvert...	43.691116	-79.476013	1.0	Turkish Restaurant	Skating Rink	Sandwich Place	Restaurant	Bar	Women's Store	Trail	Tennis Court	Pool	Pizza Place
M6N	York	Runnymede , The Junction North	43.673185	-79.487262	2.0	Pizza Place	Caribbean Restaurant	Bus Line	Breakfast Spot	Women's Store	Turkish Restaurant	Trail	Tennis Court	Skating Rink	Sandwich Place
M9N	York	Weston	43.706876	-79.518188	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Results

- **Cluster 1 in Brookyln**

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bay Ridge	-74.030621	0	Spa	Italian Restaurant	Pizza Place	Chinese Restaurant	Greek Restaurant	Grocery Store	Ice Cream Shop	Hookah Bar	American Restaurant	Lounge
1	Bensonhurst	-73.995180	0	Sushi Restaurant	Bakery	Ice Cream Shop	Chinese Restaurant	Italian Restaurant	Donut Shop	Factory	Road	Noodle House	Liquor Store
3	Greenpoint	-73.954241	0	Coffee Shop	Bar	Cocktail Bar	Yoga Studio	Spa	Café	Bakery	Record Shop	Pizza Place	French Restaurant
6	Sheepshead Bay	-73.943186	0	Dessert Shop	Turkish Restaurant	Sandwich Place	Harbor / Marina	Yoga Studio	Grocery Store	Hotel	Creperie	Restaurant	Outlet Store
12	Windsor Terrace	-73.980073	0	Deli / Bodega	Café	Diner	Park	Plaza	Bakery	Coffee Shop	Sushi Restaurant	Chinese Restaurant	Beer Store
13	Prospect Heights	-73.964859	0	Bar	Mexican Restaurant	Cocktail Bar	Beer Bar	Coffee Shop	Wine Bar	Café	Thai Restaurant	Bakery	Ice Cream Shop
15	Williamsburg	-73.958115	0	Coffee Shop	Bar	Bagel Shop	Yoga Studio	Taco Place	Breakfast Spot	Burger Joint	Café	Clothing Store	Playground

Results

- **Cluster 2 in Brookyln**

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
30	Mill Island	-73.908186	1	Pool	Locksmith	Women's Store	Filipino Restaurant	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
100	Mill Island	-73.908186	1	Pool	Locksmith	Women's Store	Filipino Restaurant	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop

Results

- **Cluster 3 in Brookyln**

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Sunset Park	-74.010316	2	Bank	Latin American Restaurant	Pizza Place	Bakery	Mexican Restaurant	Gym	Fried Chicken Joint	Pharmacy	Deli / Bodega	Mobile Phone Shop
4	Gravesend	-73.973471	2	Pizza Place	Bakery	Bus Station	Lounge	Diner	Chinese Restaurant	Breakfast Spot	Furniture / Home Store	Bar	Electronics Store
5	Brighton Beach	-73.965094	2	Restaurant	Russian Restaurant	Beach	Eastern European Restaurant	Gourmet Shop	Sushi Restaurant	Bank	Mobile Phone Shop	Taco Place	Korean Restaurant
7	Manhattan Terrace	-73.957438	2	Pizza Place	Donut Shop	Ice Cream Shop	Steakhouse	Coffee Shop	Grocery Store	Convenience Store	Restaurant	Bank	Cosmetics Shop
8	Flatbush	-73.958401	2	Mexican Restaurant	Coffee Shop	Juice Bar	Bank	Caribbean Restaurant	Pizza Place	Lounge	Chinese Restaurant	Bagel Shop	Sandwich Place
9	Crown Heights	-73.943291	2	Pizza Place	Café	Museum	Bakery	Pharmacy	Coffee Shop	Salon / Barbershop	Candy Store	Burger Joint	Playground

Results

- **Cluster 1, 2 and 3 in York**

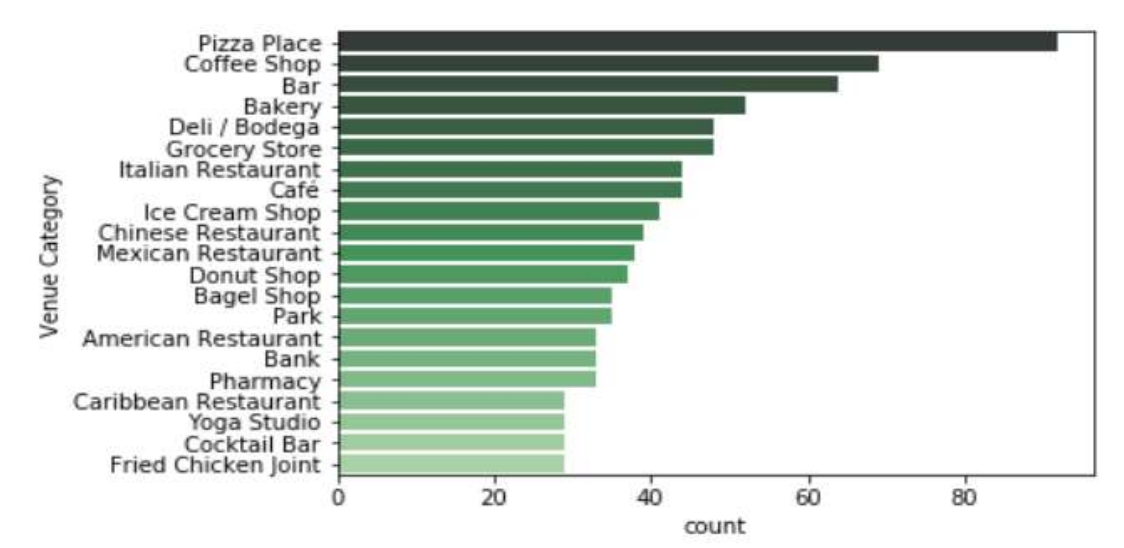
	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	York	43.706876	-79.518188	0	Park	Convenience Store	Women's Store	Trail	Sandwich Place	Restaurant	Pool	Hockey Arena	Grocery Store	Field

	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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	York	43.691116	-79.476013	1	Sandwich Place	Restaurant	Fast Food Restaurant	Convenience Store	Women's Store	Trail	Pool	Park	Hockey Arena	Grocery Store
3	York	43.673185	-79.487262	1	Grocery Store	Convenience Store	Caribbean Restaurant	Bus Line	Women's Store	Trail	Sandwich Place	Restaurant	Pool	Park

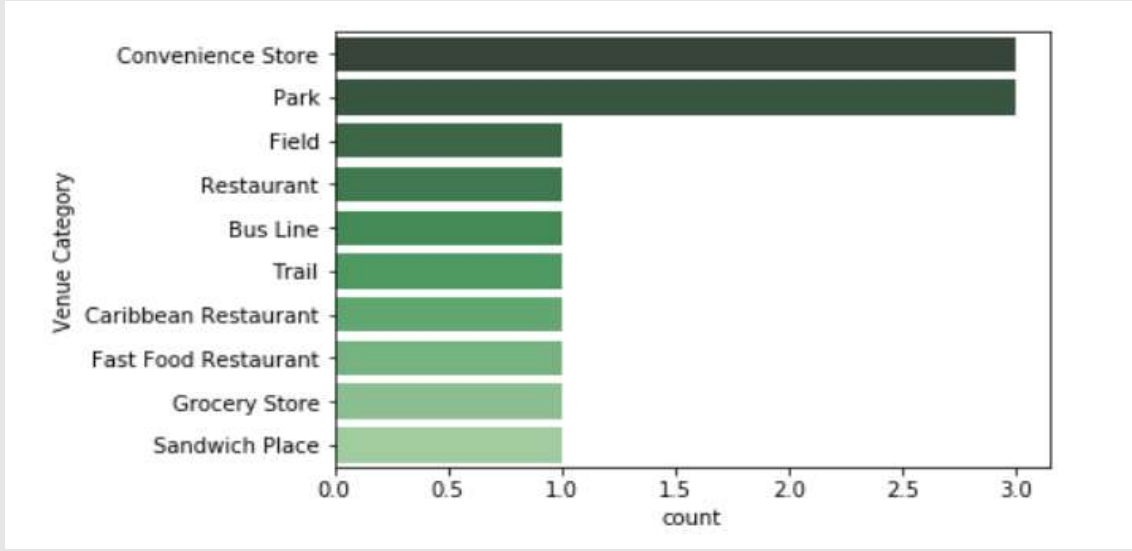
	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	York	43.693781	-79.428191	2	Trail	Park	Hockey Arena	Field	Women's Store	Sandwich Place	Restaurant	Pool	Grocery Store	Fast Food Restaurant
1	York	43.689026	-79.453512	2	Women's Store	Pool	Park	Trail	Sandwich Place	Restaurant	Hockey Arena	Grocery Store	Field	Fast Food Restaurant

Data Visualization: Bar Plots

- **Brookyln**



York



Discussion

- Brooklyn is much more diverse than York in terms of entertainment places, such as restaurants, cafes, bars, museums and so on.
- On the other side, York is mainly diverse for various sport activities, also diverse for its parks, calm rest places and small stores.
- Recommend customers travelling to Brooklyn instead of York, taking into account their answers to previously mentioned survey questions.
- As a further research, various machine learning algorithm methods as Elbow Method, Silhouette Method, or even Density Based Clustering Method can be applied for the purpose of finding the optimal k for better prediction

Conclusion

- To recapitulate, these datasets can be applied to many various problems for the further research along with new datasets and API platforms.

References

- 1. New York City data: https://geo.nyu.edu/catalog/nyu_2451_34572
- 2. Wikipedia page for Toronto data: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- 3. Geographic location data for Toronto : https://cocl.us/Geospatial_data
- 4. Foursquare API: https://foursquare.com/