Dessert Shop business in Toronto, Ontario.

Khouloud Jaouani

I. *Background* :

- Often times, when we make that leap of faith and move to a whole new country, we find ourselves facing the struggles of settling in and finding our way around the new environment.

- As far as a business owner is concerned, opening a new business in a new country or area could be challenging since there might be more than one competitor in the same neighborhood. That is why a good analysis must be conducted in order to figure out the best area to start ones business for it to succeed.

II. *Business Problem :*

- With that being said, the goal of this project aims to make the process of choosing the best area to open a Dessert Shop easier on the owner, where formalities of looking for the best neighborhood area will be taken care of.

- We will indeed extract the list of neighborhoods in Toronto, and analyze the area around each neighborhood to see if it suits the prospective Dessert Shop owner or not.

- We will use k-means clustering machine learning clustering algorithm, to cluster the venues based on their category. At the end of this analysis, we will know which area will be the best suited to start the Dessert Shop business.

III. *Datasets and APIs used :*

○ Datasets : For the Paris areas, we will scrape the data from this page. Here is the URL : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

○ Foursquare API : We will be using the Foursquare API to extract the data related to the venues around the neighborhood areas, and merge them to our initial dataset in order to make our analysis.

○ **NB :** The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag.

IV. Data Scraping :

- After scraping the data and cleaning it we obtain the following dataframe.

|	Postcode	Borough	Neighbourhood
0	None	None	None
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village

We indeed followed these steps : Dropping the cells with a borough that is "Not assigned".

Group the data by Borough and Postcode.

If a cell has a borough but a "Not assigned" neighborhood, then the neighborhood will be the same as the borough.

Downloading the Geospatial data :

○ Here is the map of Toronto.
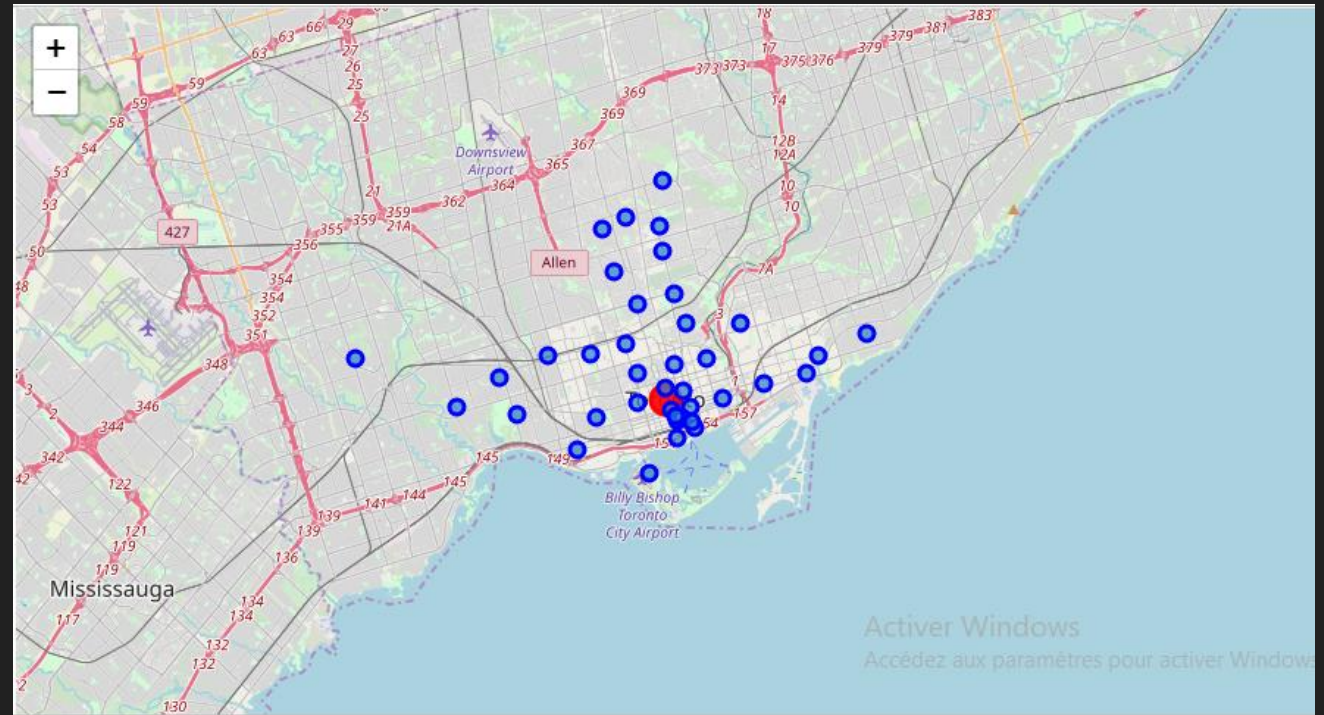
Top 10 venues in Toronto :

	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
0	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Bar	Salad Place	Thai Restaurant	Bakery	Restaurant	Sushi Restaurant
1	Agincourt	Latin American Restaurant	Lounge	Breakfast Spot	Sandwich Place	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Discount Store	Dance Studio	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Fast Food Restaurant	Beer Store	Pharmacy	Pizza Place	Fried Chicken Joint	Sandwich Place	Discount Store	Department Store
4	Alderwood, Long Branch	Pizza Place	Coffee Shop	Skating Rink	Pub	Gym	Pharmacy	Sandwich Place	Dim Sum Restaurant	Dance Studio

Let's check if there are any dessert shops as top 1 venue.

As we can see, we have 20 neighborhoods with a bakery as one of the top 10 venues which is quite a lot !

So for a better analysis, let's see how many bakeries are the top 1 venue in those neighborhoods.

```
b1 = neighborhoods_venues_sorted[neighborhoods_venues_sorted['1st Most Common Venue'] == 'Bakery']
b1
```

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

What about Dessert Shops.

```
ds1 = neighborhoods_venues_sorted[neighborhoods_venues_sorted['1st Most Common Venue'] == 'Dessert Shop']
ds1
```

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

Okay Great ! So it seems like there aren't any Dessert Shops as the number one venue. This is our chance.

Before making a decision; let's first cluster the neighborhoods to have a better understanding.

K-means clustering :

We extracted 5 different clusters in the Toronto area :

Cluster 01

```
toronto_merged.loc[toronto_merged['Cluster_Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

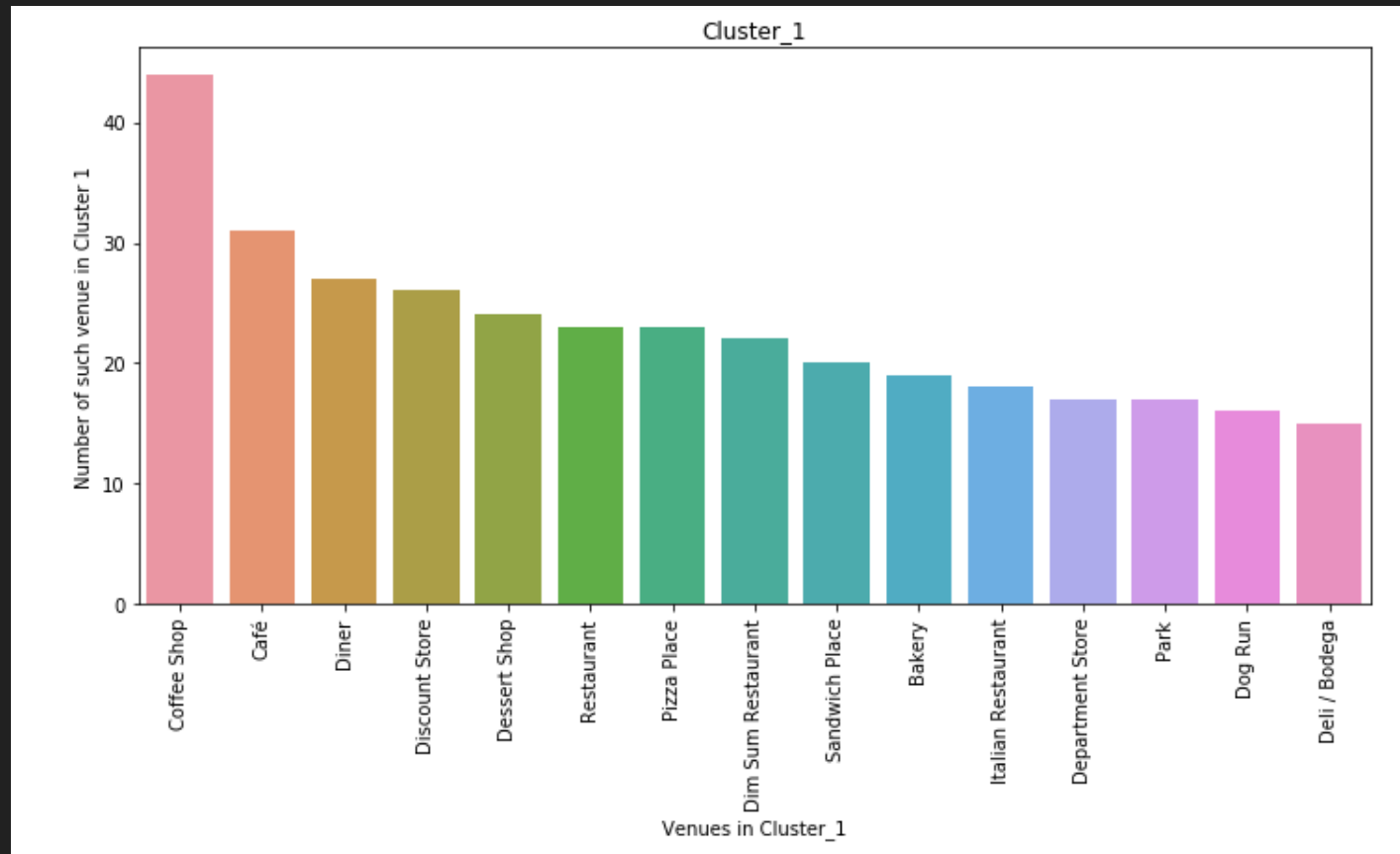|	Borough	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Co Ve
2	Scarborough	0	Electronics Store	Pizza Place	Breakfast Spot	Intersection	Medical Center	Mexican Restaurant	Re Lo
3	Scarborough	0	Coffee Shop	Korean Restaurant	Gift Shop	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Du Re
4	Scarborough	0	Caribbean Restaurant	Bakery	Bank	Athletics & Sports	Lounge	Hakka Restaurant	Ga
6	Scarborough	0	Discount Store	Coffee Shop	Department Store	Dance Studio	Empanada Restaurant	Electronics Store	Ea Eu Re
7	Scarborough	0	Bus Line	Fast Food Restaurant	Bakery	Metro Station	Intersection	Park	So
8	Scarborough	0	Motel	American Restaurant	Women's Store	Curling Ice	Deli / Bodega	Department Store	De

Cluster 03

```
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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Mo Comm Venue
20	North York	2	Cafeteria	Women's Store	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Doner Resta

Cluster 04

```
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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9 C V
0	Scarborough	3	Fast Food Restaurant	Women's Store	Doner Restaurant	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	D

Cluster 02

```
toronto_merged.loc[toronto_merged['Cluster_Labels'] == 1, 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	6th Most Common Venue	7th Most Common Venue
5	Scarborough	1	Playground	Convenience Store	Women's Store	Discount Store	Deli / Bodega	Department Store	Dessert Shop
14	Scarborough	1	Park	Playground	Discount Store	Dance Studio	Deli / Bodega	Department Store	Dessert Shop
23	North York	1	Park	Convenience Store	Bank	Discount Store	Department Store	Dessert Shop	Dim Sum Restauran
25	North York	1	Park	Food & Drink Shop	Construction & Landscaping	Discount Store	Department Store	Dessert Shop	Dim Sum Restauran
30	North York	1	Park	Airport	Dog Run	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restauran
40	East York	1	Park	Metro Station	Convenience Store	Discount Store	Department Store	Dessert Shop	Dim Sum Restauran
44	Central Toronto	1	Park	Swim School	Bus Line	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Donut Sho

Cluster 05

```
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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Con Ven
1	Scarborough	4	Golf Course	Bar	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Don Res

We then extracted all the venues in a cluster

Deciding on the location : North York

Let's direct our focus on cluster 01 and 02 since they have
the most human traffic within them. The top venues in those
clusters are mostly coffee shops. So opening a dessert shop
within those areas could be beneficial for the business since
there might be a chance of developing business ties with those
coffee shops, as well as profiting of the movement in these areas.
Let's see what the recurrent neighborhoods in those two clusters are.

```
cluster["Cluster_1"]["Borough"].value_counts().argmax()
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:1: FutureWarning:
The current behaviour of 'Series.argmax' is deprecated, use 'idxmax'
instead.
The behavior of 'argmax' will be corrected to return the positional
maximum in the future. For now, use 'series.values.argmax' or
'np.argmax(np.array(values))' to get the position of the maximum
row.
  """Entry point for launching an IPython kernel.
```

```
'North York'
```

```
cluster["Cluster_2"]["Borough"].value_counts().argmax()
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:1: FutureWarning:
The current behaviour of 'Series.argmax' is deprecated, use 'idxmax'
instead.
The behavior of 'argmax' will be corrected to return the positional
maximum in the future. For now, use 'series.values.argmax' or
'np.argmax(np.array(values))' to get the position of the maximum
row.
  """Entry point for launching an IPython kernel.
```

```
'North York'
```

As we can see here, North York seems to be a good location to open a Dessert Shop.

Activer Wind
Accédez aux pa

Conclusion :

Finally, we drew the conclusion that opening a Dessert Shop around the North York area would be perfect. We indeed analyzed our results according to the five clusters we produced. While Clusters 3, 4 and 5 may not be suitable since there is mainly parks and gardens in those neighborhoods, clusters 1 and 2 would be ideal since they had quite the human traffic with many Coffee shops, bars and restaurants around the area.