The Battle of Neighborhood: The Best Location for a sub-Saharan

African restaurant in Toronto

Stephane Mviena September 20, 2020

I. Introduction

a. Background

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016, It is the most populous city in Canada and the fourth most populous city in North America. According to official statistics, Toronto is the city with the largest number of blacks in Canada (approximately 36%) and this proportion represents approximately 7.5% of the city's population. According to official statistics, Toronto is the city with the largest number of blacks in Canada (approximately 36%) and this proportion represents approximately 7.5% of the city's population.

Because of this multiculturalism, Toronto is full of many culinary treasures and therefore many restaurants. There are Portuguese, Japanese, Chinese, Vietnamese, Indian restaurants and more. by doing research on the internet, we realize that there are also African restaurants that exist in the city but most of them are Ethiopian or Somali

b. Problem

The objective of this project is to be able to find the best location for establish the sub-Saharan African restaurant in the city of Toronto using Data analysis and Machine Learning.

c. Interest

Potential African investors would be interested in this project which will make African cuisine known in the city. Also, all hotel or catering professionals could be very interested by this project.

II. Data acquisition and cleaning

a. Data Source

The Foursquare API will be used to explore neighborhoods in Toronto, more specifically, we will be using the explore function to get the most common venue categories in each neighborhood.

Neighborhood names, alongside their corresponding boroughs and postal codes, scraped from Wikipedia: Toronto Wikipedia

The Toronto's census data for its social demographic characteristics will be distilled from Toronto's Neighborhood Profile : Census Data.

b. Data Description

Most common venue categories in each neighborhood: these data will give us the ability to analyses the number of the occurrences for each venue category in the neighborhood. But further the main category which is important for our project is ‘’Restaurant”. It will be added in our final data set

Neighborhood names data scrapped from Wikipedia will allow us to have the names of the neighborhoods of the city of Toronto according to the local postal codes

The census data which is our main source of data allows us to have demographic data on the population of neighborhoods in the city of Toronto and their income.

The combination of all these reprocessed data sources will give us our final data set to apply our machine learning model

III. Methodology

The first step of this data project consists to cleaning all the datasets which we are use.

After download a census data of Toronto I am processing by drop 2 features which are **Neighbourhood Id, Combined Indicators** and rename **After-Tax Household Income** to **Income.** The result is as follows:

	Neighbourhood	Income	Africa	Total Population
0	West Humber-Clairville	59703.0	260.0	33312.0
1	Mount Olive-Silverstone-Jamestown	46986.0	830.0	32954.0
2	Thistletown-Beaumond Heights	57522.0	130.0	10360.0
3	Rexdale-Kipling	51194.0	0.0	10529.0
4	Elms-Old Rexdale	49425.0	90.0	9456.0

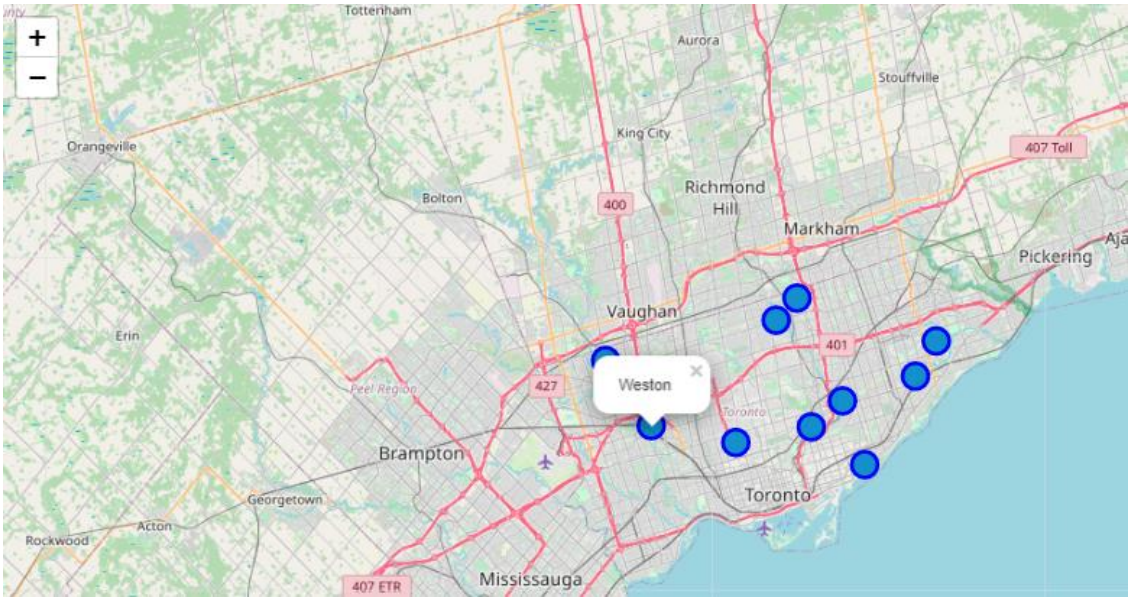
The other datasets for the city of Toronto have been downloaded from Wikipedia and the internet. By combining these 3 datasets we have the following result:

[6]:

	Latitude	Longitude	Neighbourhood	Income	Africa	Total Population	Percentage of Africans
0	43.756303	-79.565963	Humber Summit	53272.0	40.0	12416.0	0.322165
1	43.725882	-79.315572	Victoria Village	43743.0	215.0	17510.0	1.227870
2	43.803762	-79.363452	Hillcrest Village	57682.0	30.0	16934.0	0.177158
3	43.786947	-79.385975	Bayview Village	58028.0	25.0	21396.0	0.116844
4	43.705369	-79.349372	Thorncliffe Park	38645.0	230.0	21108.0	1.089634

In this new dataset we can clearly see the proportion of Africans where they are present, the average income and the coordinates of each neighborhood of Toronto.

With this result, we can visualize the neighborhood using the folium library. I created a map of Toronto with neighborhood superimposed on top.



To explore the venues of Neighborhoods where the Africans are present, I use Foursquare API. I designed to limit as 100 venues and the radius of 1500 for each borough from their given latitude and longitude information's. Here is a head of the list neighborhood, Venues name, category, latitude and longitude information's from Foursquare API

```
Humber Summit
Victoria Village
Hillcrest Village
Bayview Village
Thorncliffe Park
The Beaches
Humewood-Cedarvale
Weston
Woburn
Scarborough Village
```

14]:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Humber Summit	43.756303	-79.565963	Pizza Monza	43.755043	-79.567195	Pizza Place
1	Humber Summit	43.756303	-79.565963	HNS ARARAT	43.757519	-79.563653	Furniture / Home Store
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

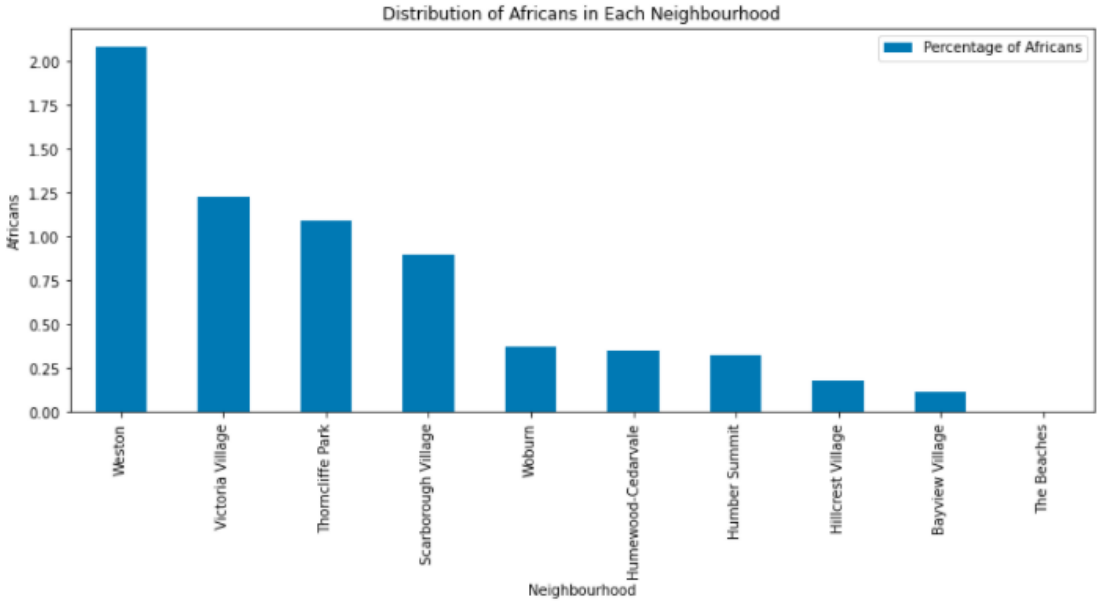
We can anlyzed the Neighborhood as follow :

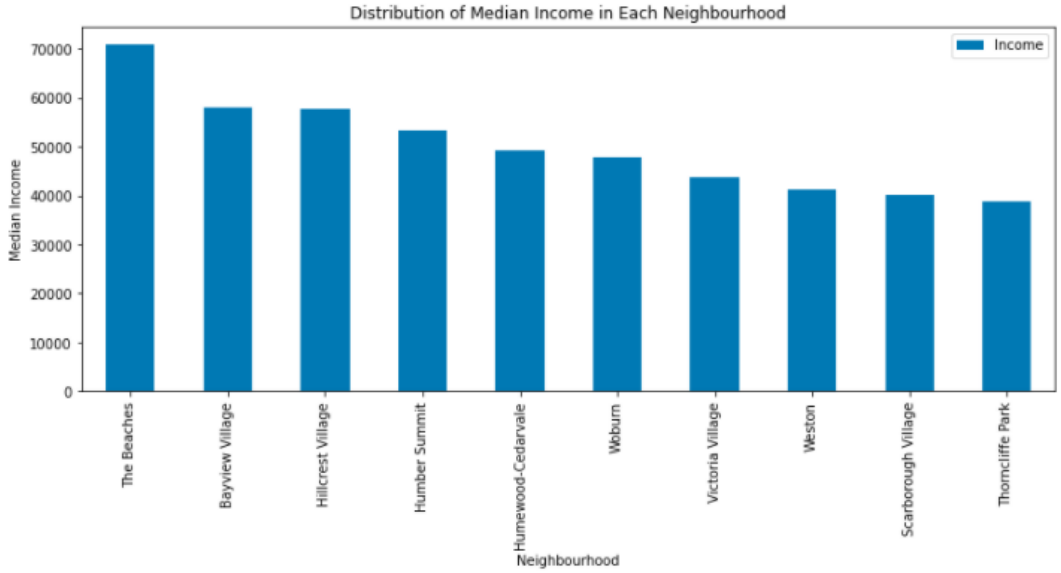
The number of Venues show that Thorncliffe Park is the Neighborhood with many places.

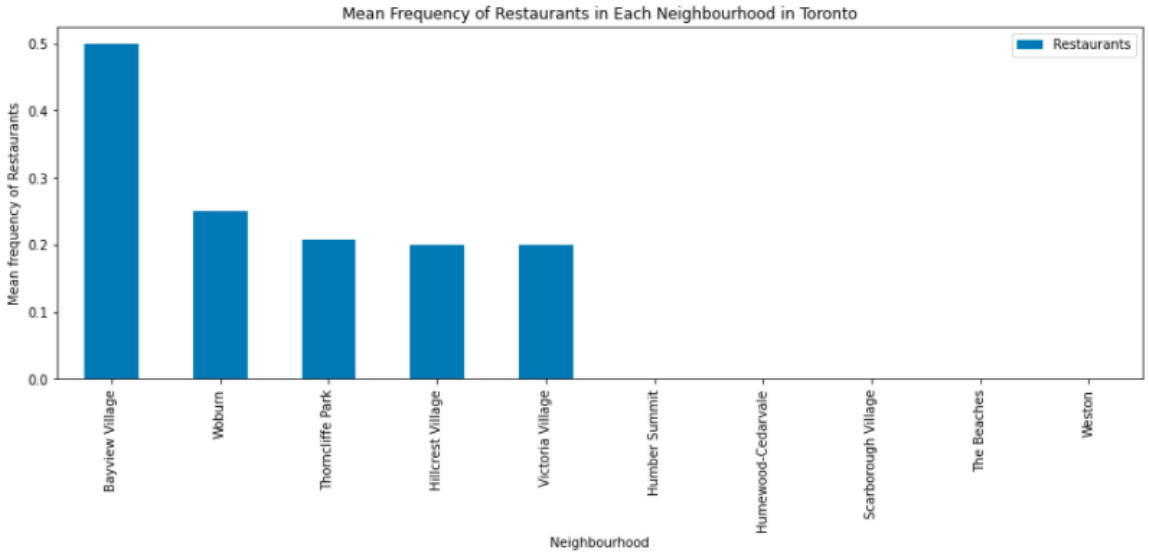
Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bayview Village	4	4	4	4	4	4
Hillcrest Village	5	5	5	5	5	5
Humber Summit	2	2	2	2	2	2
Humewood-Cedarvale	4	4	4	4	4	4
Scarborough Village	3	3	3	3	3	3
The Beaches	4	4	4	4	4	4
Thorncliffe Park	24	24	24	24	24	24
Victoria Village	5	5	5	5	5	5
Weston	2	2	2	2	2	2
Woburn	4	4	4	4	4	4

At this stage, the following three graphs already allow us to draw some information:

- **Weston** is the neighborhood that has a lot more Africans in the city of Toronto
- Incomes are roughly average in all neighborhoods
- In terms of competition, there are five neighborhoods where finding a restaurant is quite difficult. these are **Weston, Humber Summit, The Beaches, Humewood-Cedarvale, Scarborough Village.**


Distribution of Africans in Each Neighbourhood
Percentage of Africans
Africans
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Weston
Victoria Village
Thorncliffe Park
Scarborough Village
Woburn
Humewood-Cedarvale
Humber Summit
Hillcrest Village
Bayview Village
The Beaches
Neighbourhood


Distribution of Median Income in Each Neighbourhood
Income
Median Income
70000
60000
50000
40000
30000
20000
10000
0
The Beaches
Bayview Village
Hillcrest Village
Humber Summit
Humewood-Cedarvale
Woburn
Victoria Village
Weston
Scarborough Village
Thorncliffe Park
Neighbourhood


Mean Frequency of Restaurants in Each Neighbourhood in Toronto
Restaurants
Mean frequency of Restaurants
0.5
0.4
0.3
0.2
0.1
0.0
Bayview Village
Woburn
Thorncliffe Park
Hillcrest Village
Victoria Village
Humber Summit
Humewood-Cedarvale
Scarborough Village
The Beaches
Weston
Neighbourhood

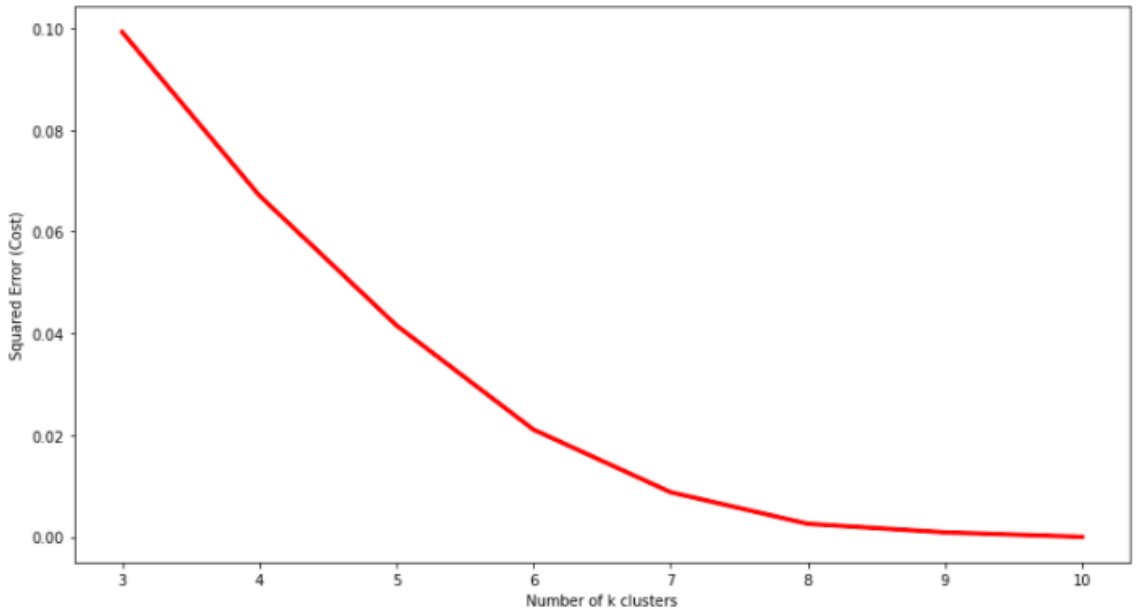
In order to go further in our analysis, we will apply a Machine Learning model called K-Means. it is a model which allows to classify data in homogeneous groups called **clusters**. These clusters have roughly the same characteristics. We will apply this model to the new data set that we will get by merging the dataset from the venues found in the neighborhoods and our first dataset.

[21]:

	Latitude	Longitude	Neighbourhood	Income	Percentage of Africans	Restaurants
0	43.756303	-79.565963	Humber Summit	53272.0	0.322165	0.000000
1	43.725882	-79.315572	Victoria Village	43743.0	1.227870	0.200000
2	43.803762	-79.363452	Hillcrest Village	57682.0	0.177158	0.200000
3	43.786947	-79.385975	Bayview Village	58028.0	0.116844	0.500000
4	43.705369	-79.349372	Thorncliffe Park	38645.0	1.089634	0.208333

We will first determine the optimal number of clusters. The most common way to choose the number of clusters is to run K-Means with different values of K and calculate the variance of the different clusters. The variance is the sum of the distances between each centroid in a cluster and the different observations included in the same cluster. Thus, we try to find several clusters K such that the selected clusters minimize the distance between their centers (centroids) and the observations in the same cluster

Generally, by putting in a graph the different numbers of clusters K according to the variance, we find a graph like this one:



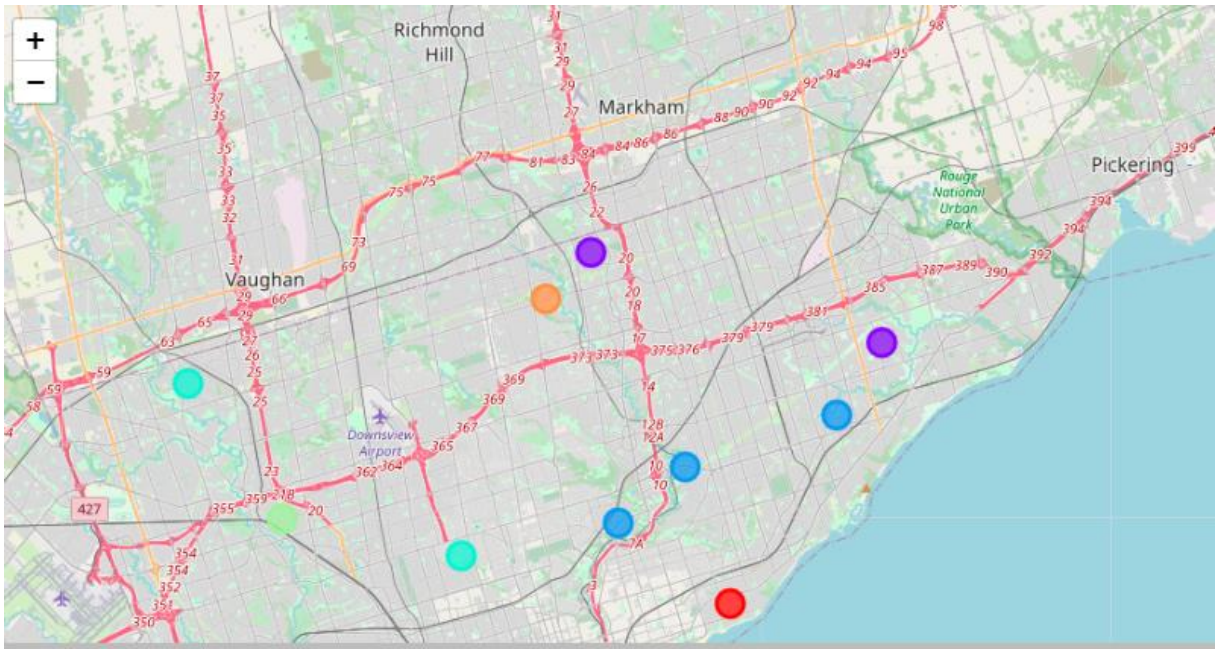
Note on this graph, the shape of an arm where the highest point represents the shoulder and the point where K is 10 represents the other end: the hand. The optimal number of clusters is the point representing the elbow. Here the elbow can be represented by K being **6**. This is the optimal number of clusters. Generally, the knee point is that of the number of clusters from which the variance no longer decreases significantly. The fact of looking for the point representing the elbow, gave name to this method: **The Elbow method**

IV. Results

After determining the number of clusters, we can easily proceed to our classification of the neighborhood of the city of Toronto according to the features that we have retained. We then get this new data frame that we can put on a map like this

[30]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
0	3	43.756303	-79.565963	Humber Summit	53272.0	40.0	12416.0	0.322165	-0.856697
1	2	43.725882	-79.315572	Victoria Village	43743.0	215.0	17510.0	1.227870	0.404697
2	1	43.803762	-79.363452	Hillcrest Village	57682.0	30.0	16934.0	0.177158	0.404697
3	5	43.786947	-79.385975	Bayview Village	58028.0	25.0	21396.0	0.116844	2.296789
4	2	43.705369	-79.349372	Thorncliffe Park	38645.0	230.0	21108.0	1.089634	0.457256



The detail of all the clusters is as follows:

Cluster 0

```
[33]: df_cluster.loc[df_cluster['Cluster Label'] == 0]
```

[33]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
5	0	43.676357	-79.293031	The Beaches	70957.0	0.0	21567.0	0.0	-0.856697

Cluster 1

```
[34]: df_cluster.loc[df_cluster['Cluster Label'] == 1]
```

[34]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
2	1	43.803762	-79.363452	Hillcrest Village	57682.0	30.0	16934.0	0.177158	0.404697
8	1	43.770992	-79.216917	Woburn	47908.0	200.0	53485.0	0.373937	0.720046

Cluster 2

```
[35]: df_cluster.loc[df_cluster['Cluster Label'] == 2]
```

[35]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
1	2	43.725882	-79.315572	Victoria Village	43743.0	215.0	17510.0	1.227870	0.404697
4	2	43.705369	-79.349372	Thorncliffe Park	38645.0	230.0	21108.0	1.089634	0.457256
9	2	43.744734	-79.239476	Scarborough Village	40181.0	150.0	16724.0	0.896915	-0.856697

Cluster 3

```
[36]: df_cluster.loc[df_cluster['Cluster Label'] == 3]
```

[36]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
0	3	43.756303	-79.565963	Humber Summit	53272.0	40.0	12416.0	0.322165	-0.856697
6	3	43.693781	-79.428191	Humewood-Cedarvale	49252.0	50.0	14365.0	0.348068	-0.856697

Cluster 4

```
[38]: df_cluster.loc[df_cluster['Cluster Label'] == 4]
```

[38]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
7	4	43.706876	-79.518188	Weston	41356.0	375.0	17992.0	2.08426	-0.856697

Cluster 5

```
[39]: df_cluster.loc[df_cluster['Cluster Label'] == 5]
```

[39]:

	Cluster Label	Latitude	Longitude	Neighbourhood	Income_x	Africa	Total Population	Percentage of Africans	Restaurants
3	5	43.786947	-79.385975	Bayview Village	58028.0	25.0	21396.0	0.116844	2.296789

The results of our analysis show 6 clusters whose important characteristics are as follows:

Cluster 0: in this group the African population is not present and therefore cannot bring any interest in our project

Cluster 1: this group is characterized by an African population of less than 1% but incomes in these neighborhoods are high

Cluster 2: this group is characterized by an African population of 1% on average with average incomes and the presence of restaurants in the districts of Victoria Village and Thorncliffe Park.
Cluster 3: this is characterized by high incomes but a low African population and almost no restaurants.
Cluster 4: the highest African population in the city for an average income and almost no restaurant.
Cluster 5: for this group, smaller African population but with a high density of restaurants.

V. Discussion

Toronto is a big city in Canada and this project is interesting for any investor wishing to settle in Toronto. the methods and model used can be replicated for other areas such as banking, parks and even startups.

To complete the analysis, it would be interesting to have the origin of Africans in each Neighborhood, which would allow us to have the type of African cuisine appropriate for our restaurant.

In addition, data on local food markets could also help us to find locations near these markets where we could get supplies.

VI. Conclusion

From this analysis, it appears the best location for a sub-Saharan African restaurant is the neighborhood of **Weston** which therefore wins **The Battle of Neighborhood in Toronto.** It must be said that Weston district where there are many more Africans with an average income of 40K$. Moreover, the analysis showed us that this is one of the neighborhood where restaurants are rare, so we don't really have to worry about the competition**.**