Finding the right neighborhood to open a new bank in Etobicoke

By José G. Pérez

Business understanding

- An investor is interested in opening a new Bank in the city of Etobicoke, Toronto.
- The investor asked his new business department through his team of Data Scientists an analysis that allows generating recommendations of which is the ideal neighborhood to open to the new Bank based on three considerations:
    1) Less competition.
    2) Greater amount of new clients.
    3) Potential clients with better income.
- From these preliminary variables, the neighborhood where the new bank agency will be opened must be identified.

Analytic approach

- To select the ideal neighborhood to open the new Bank, we will do an analysis of the demographic data of the Etobicoke neighborhoods, highlighting the number of population, per capita income and population growth.
- This will allow us to classify the most attractive neighborhoods considering these variables. From these variables we will use the clustering machine learning algorithm (k-means)
- Prior to the application of the algorithm, the standardization of the variables will be carried out, so that they have the same scale. In this way we will group neighborhoods with common demographic characteristics.
- In this analysis we will incorporate the information from the most common offices around (2000m) of the Eobicoke neighborhoods, with emphasis on the Bank's offices.
- In this way, we will build a data set that will provide us with demographic data, information on bank offices and the results of the clustering analysis.
- We will support these analyzes with bank office location maps around neighborhoods acquired from the Foursquare API service.

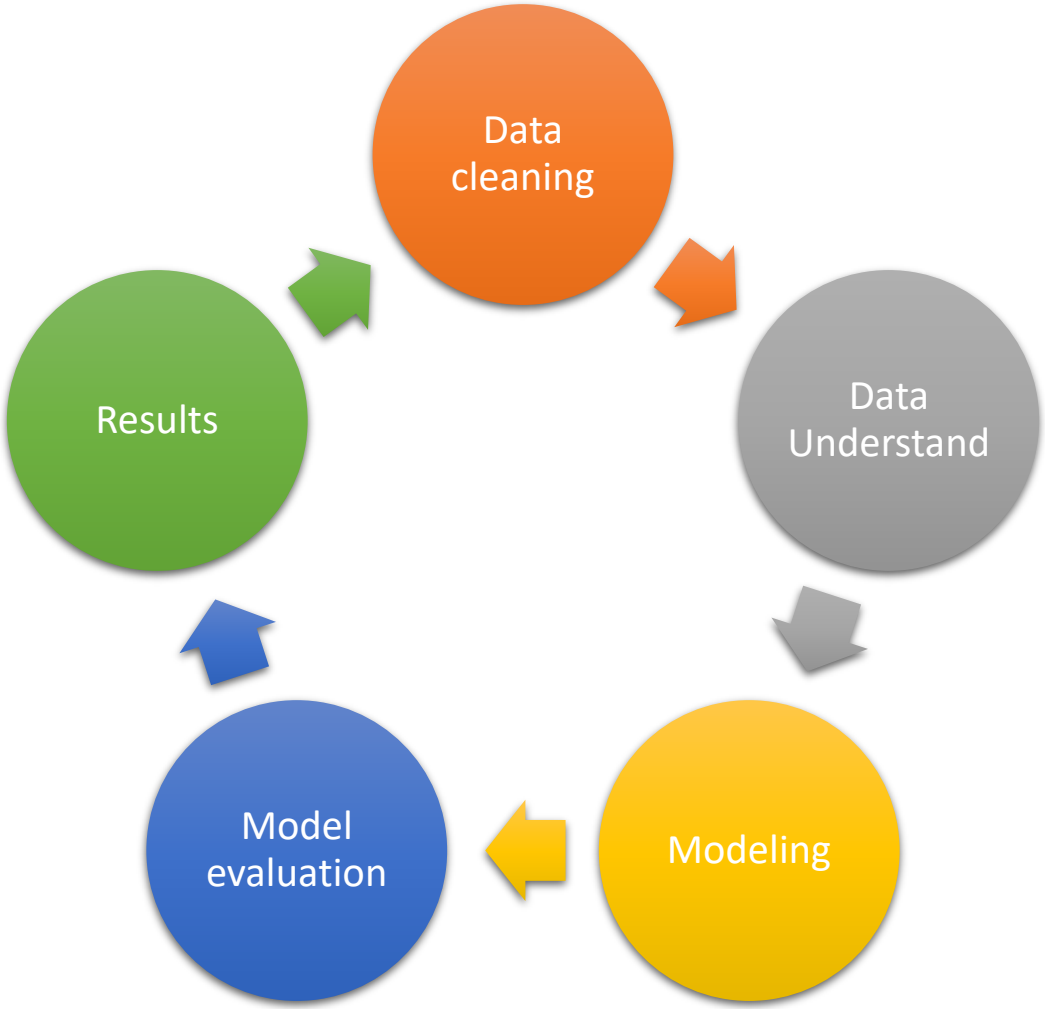
Data description

In order to make the necessary recommendations to the investor who is interested in opening a new Bank in the city of Etobicoke, the following data sources were used:

- Toronto demographics features [1].

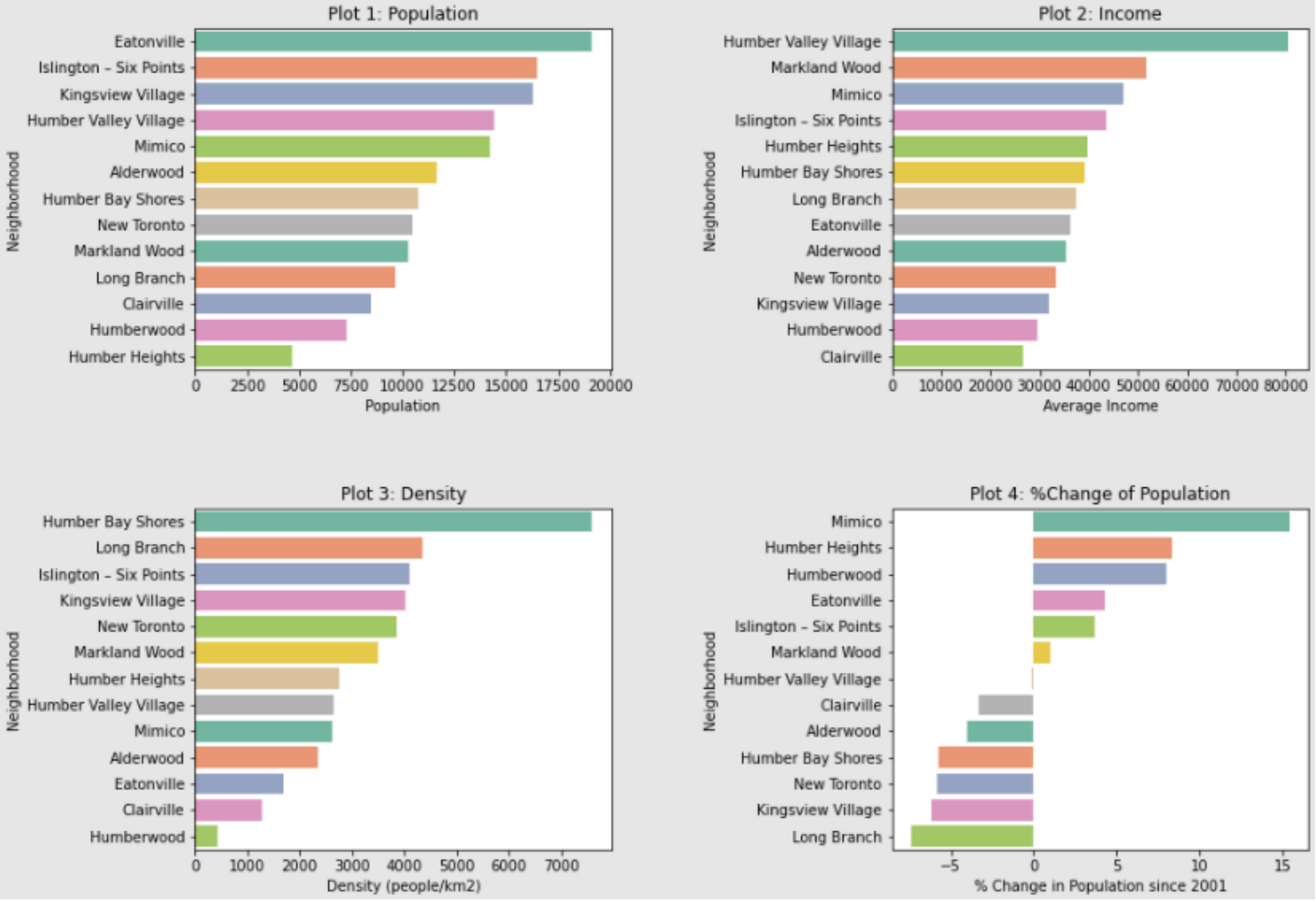
- Principal 100 venues around 2000m of the Etobicoke neighborhoods using the Foursquare API services [2].

Methodology



Data understanding

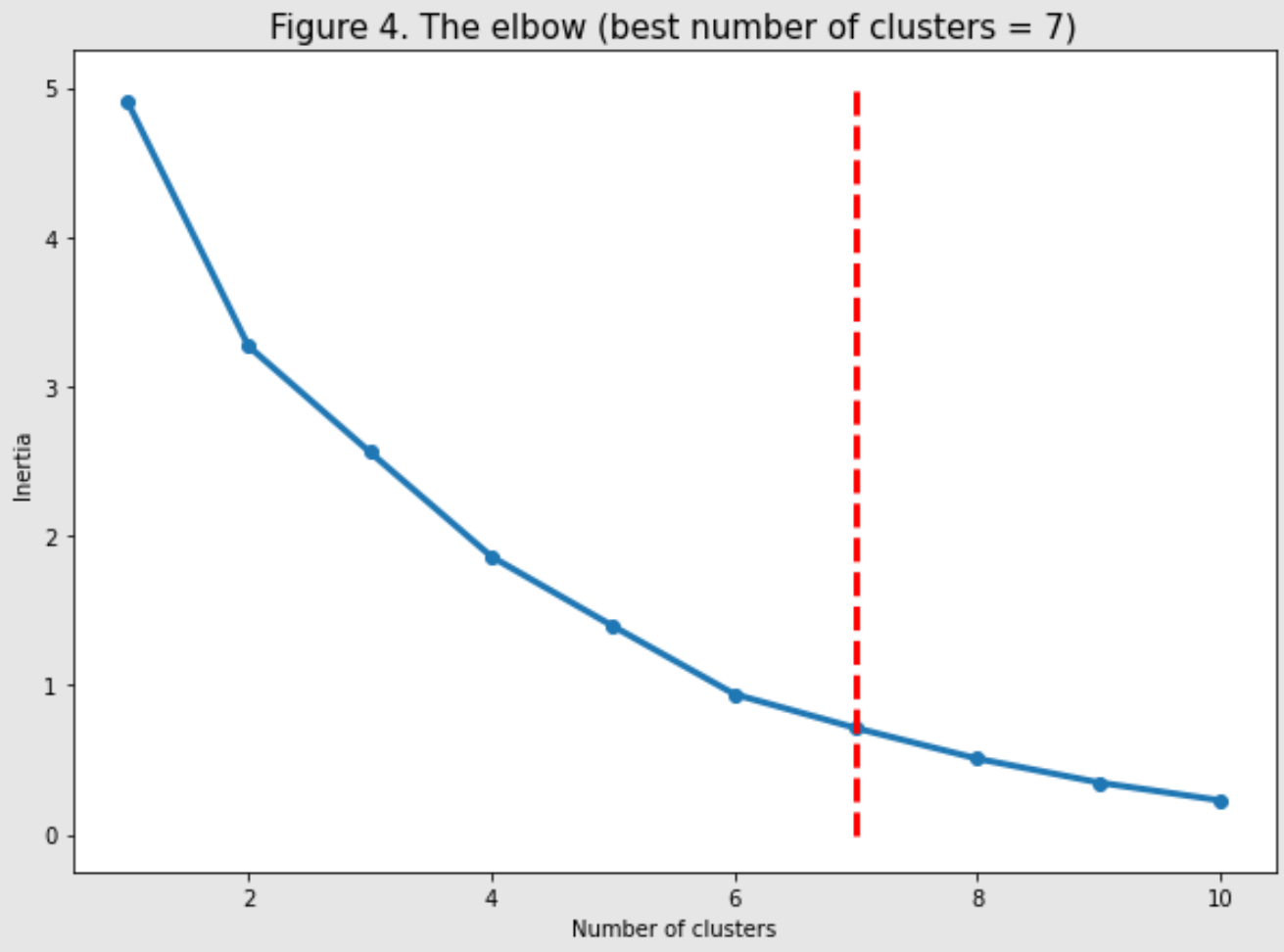
Figure 3. Etobicoke demographics features



From demographics data (Figure 3):

- Eatonville has the highest population.
- Humber valley has the highest average income.
- Humber Bay has the highest population density.
- Mimico has the highest % change of population.

Modelling



For data modeling we use:

- Clustering analysis machine learning algorithm (k-means).
- Demographics data was standardized using MinMax Scaler.
- To evaluate the cluster model we use The elbow plot.
- The Elbow plot shows that the ideal number of clusters is 7.

Results

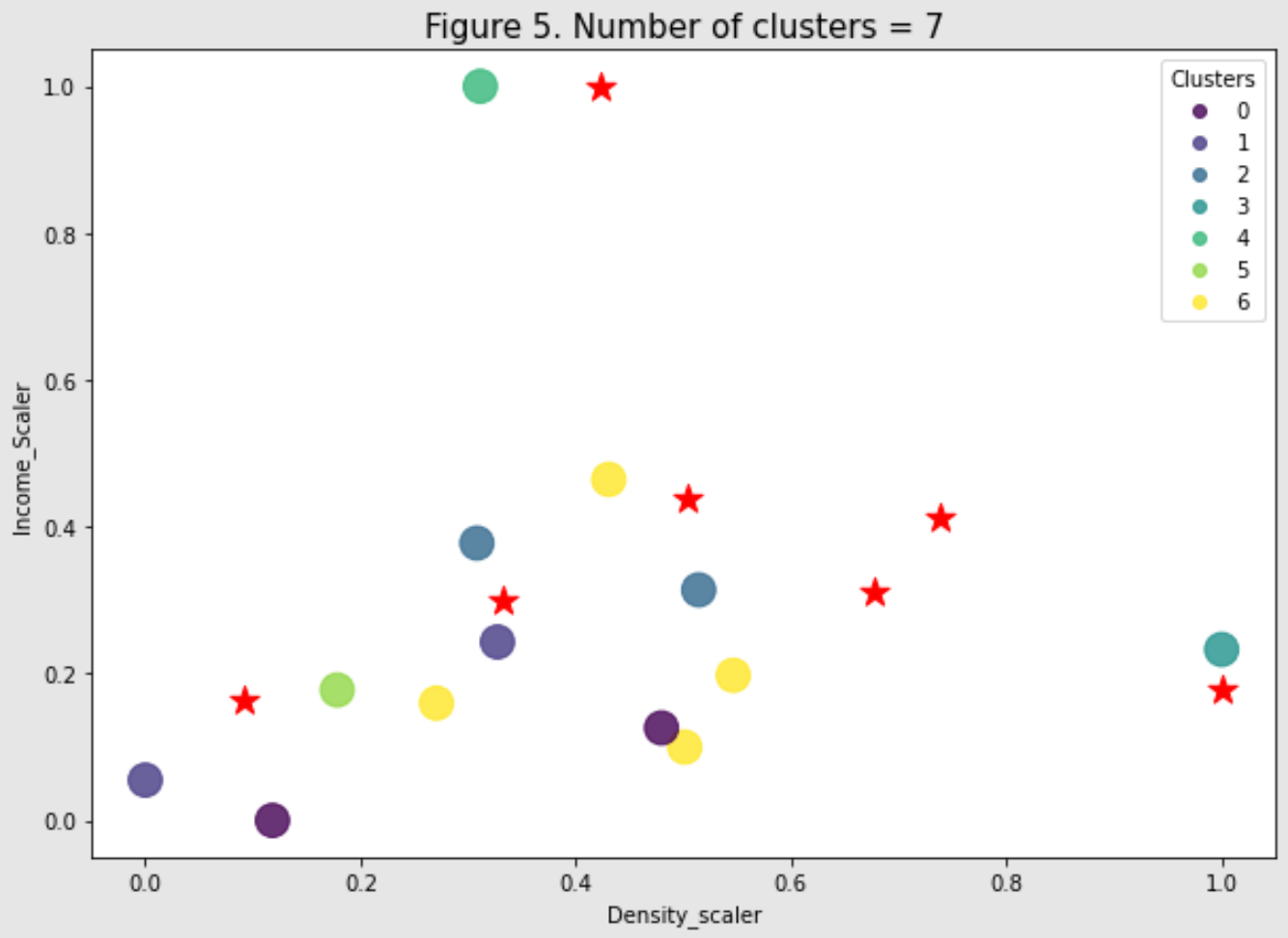


Figure 5 shows:

- Number of clusters after the analysis.
- The centroids of each clusters.

Results

	Neighborhood	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	Latitude	Longitude	ClusterLabels	...	1st Most Common Venue
0	Alderwood	11656	4.94	2360	-4.0	35239	8.8	43.604960	-79.541160	6	...	Coffee Shop
1	Clairville	8506	6.71	1268	-3.3	26610	13.2	43.748030	-79.631220	0	...	Intersection
2	Eatonville	19131	11.26	1699	4.3	36206	12.6	43.689581	-79.494751	5	...	Coffee Shop
3	Humber Bay Shores	10775	1.42	7588	-5.7	39186	15.7	43.626860	-79.476710	3	...	Park
4	Humber Heights	4674	1.69	2766	8.3	39738	10.1	43.652247	-79.486697	1	...	Bakery
5	Humberwood	7319	17.40	421	8.0	29576	7.9	43.725165	-79.621555	1	...	Coffee Shop
6	Humber Valley Village	14453	5.45	2652	-0.1	80618	12.0	43.641466	-79.492537	4	...	Coffee Shop
7	Islington – Six Points	16508	4.02	4106	3.7	43570	17.1	43.634870	-79.530520	2	...	Coffee Shop
8	Kingsview Village	16254	4.05	4013	-6.2	32004	11.8	43.702510	-79.572090	6	...	Coffee Shop
9	Long Branch	9625	2.22	4336	-7.4	37288	14.2	43.593540	-79.532750	6	...	Coffee Shop
10	Markland Wood	10240	2.92	3507	1.0	51695	9.2	43.633910	-79.569480	6	...	Coffee Shop
11	Mimico	14198	5.40	2629	15.4	47011	11.6	43.617290	-79.498850	2	...	Restaurant
12	New Toronto	10455	2.71	3858	-5.8	33415	13.0	43.601430	-79.509250	0	...	Park

- This is the resulting dataset with the cluster labels added.
- Then we can group the demographics features by cluster to select the cluster with better demographics features and minus Banks around it.

Results

Figure 6. Etobicoke Neighborhood Cluster Map

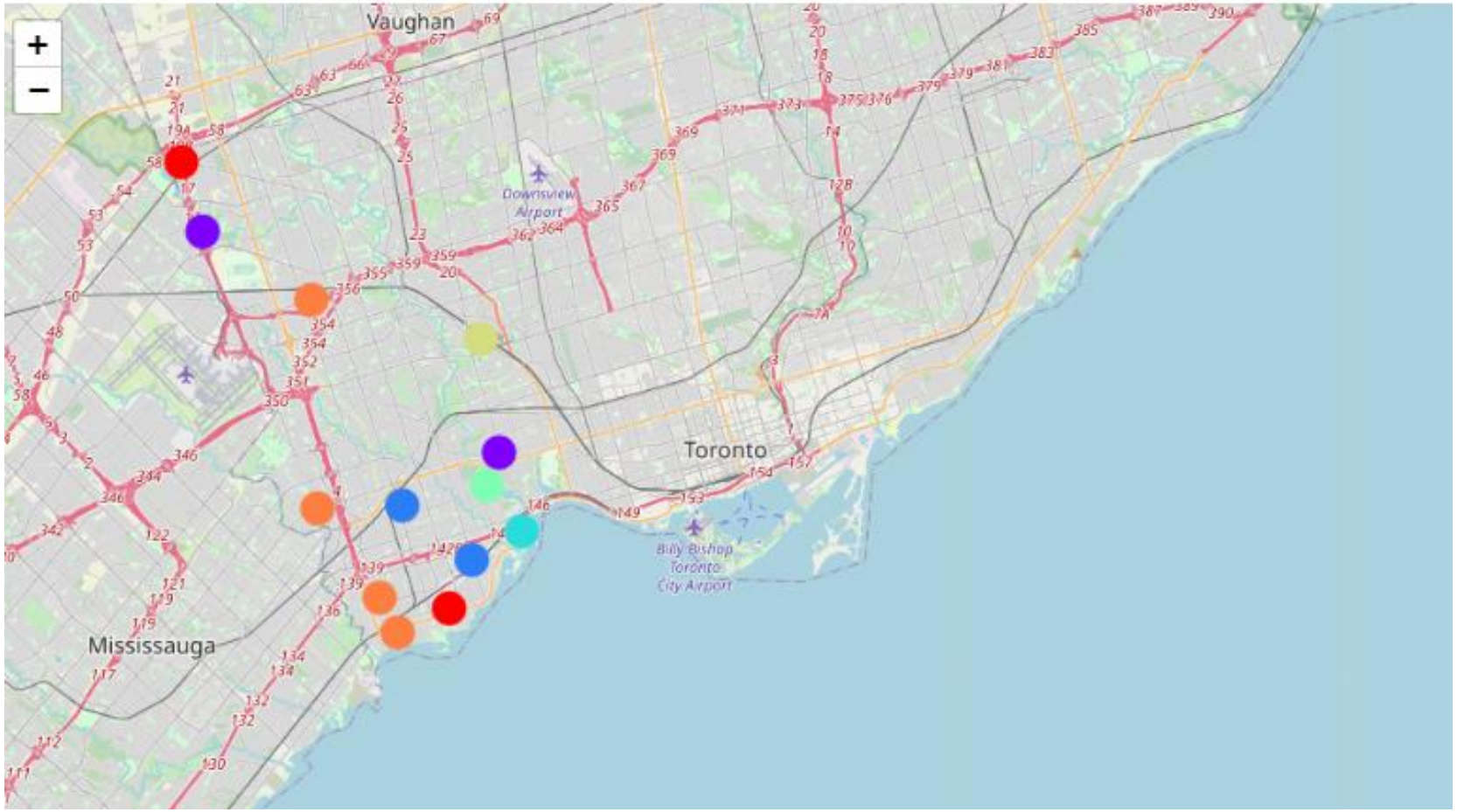


Figure 6 shows:

- The Etobicoke map with the cluster analysis superimposed on each neighborhood.

Results

Figure 7. Banks location map on the Etobicoke neighborhoods (2000 m radius)

- Eatonville and Clairville neighborhoods don't have Banks around.2000 meters radius.
- We must compare the demographics features for both neighborhoods to select the best to open a new bank.

Results

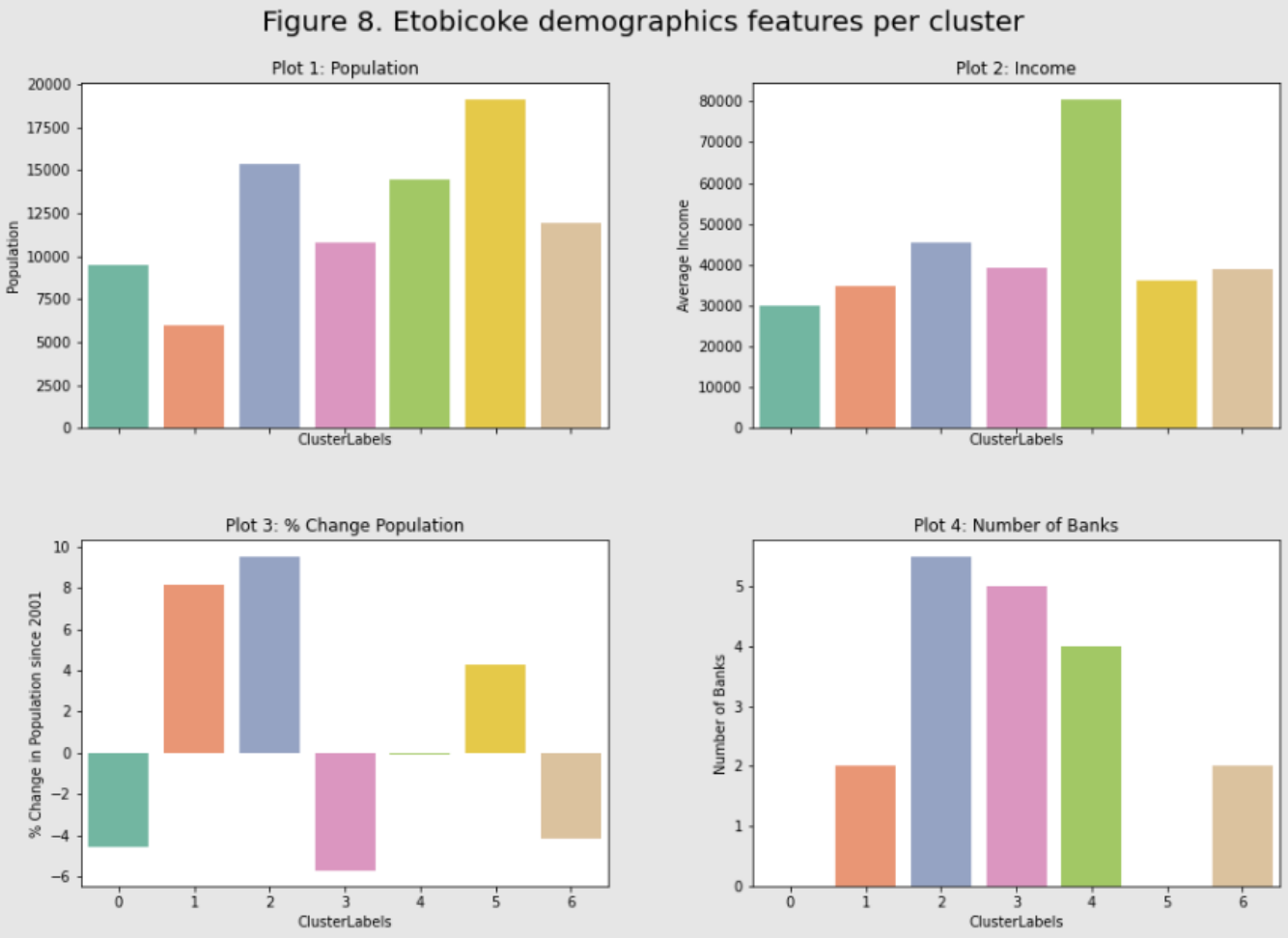
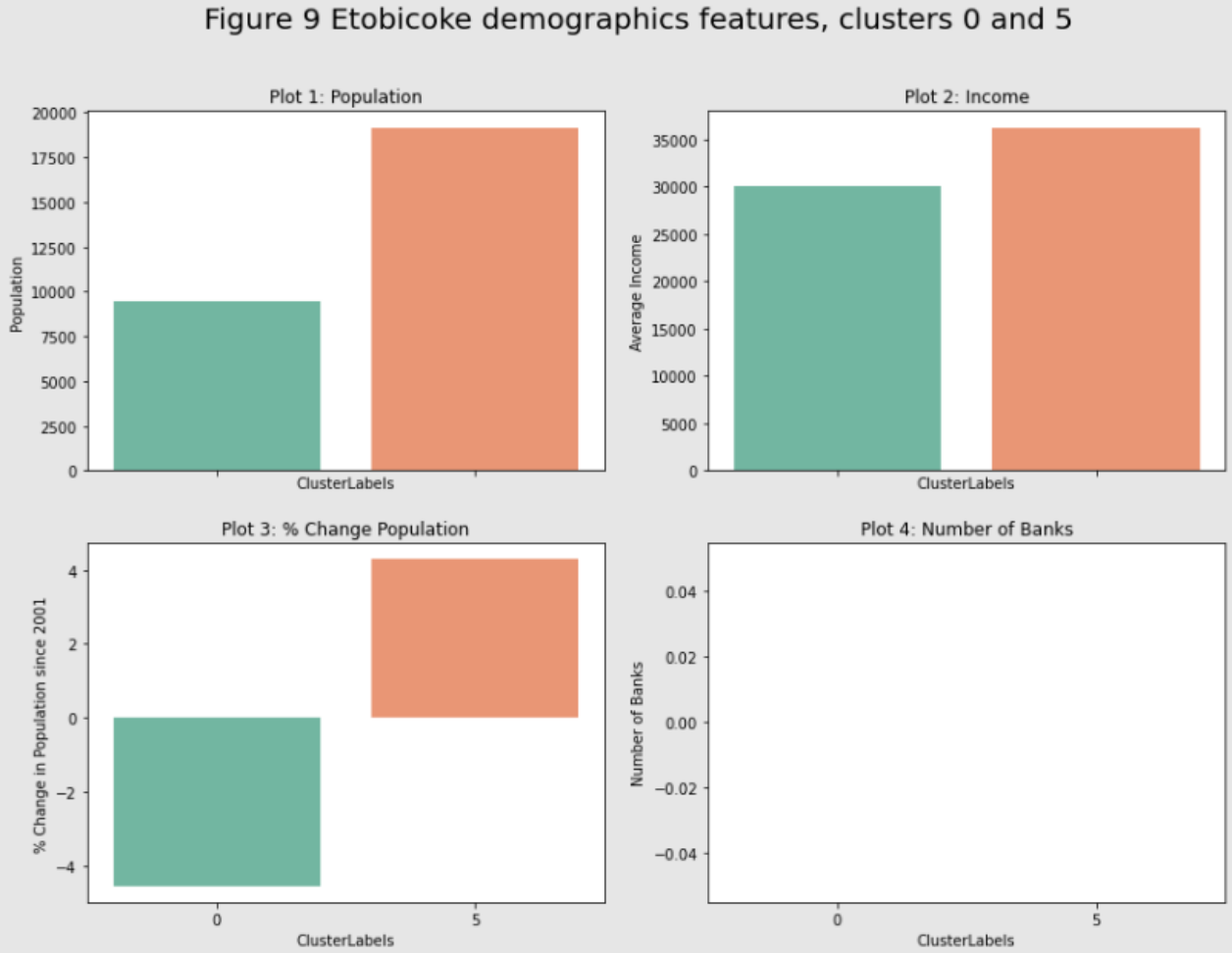


Figure 8 shows:

- Cluster 5 has the highest population.
- Cluster 4 has the highest income.
- Clusters 0, 3, and 6 decreased population.
- **Clusters 0 and 5 don't have Banks around it.**

Results



When we compare clusters 0 and 5

- Cluster 5 has the highest population.
- Cluster 5 has the highest income.
- Clusters 0 have decreased population since 2001.
- **Both clusters don't have Banks around it.**

Results

	Neighborhood	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	Latitude	Longitude	ClusterLabels
2	Eatonville	19131	11.26	1699	4.3	36206	12.6	43.689581	-79.494751	5

	Neighborhood	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	Latitude	Longitude	ClusterLabels
1	Clairville	8506	6.71	1268	-3.3	26610	13.2	43.74803	-79.63122	0
12	New Toronto	10455	2.71	3858	-5.8	33415	13.0	43.60143	-79.50925	0

Filtering clusters 0 and 5:

- Group 5 has only one neighborhood (Eatonville)
- Group 0 has two neighborhoods, Clairville and New Toronto, geographically distant.
- Eatonville, compared to neighborhoods in cluster 0, has better demographics.

Conclusion

- Cluster 5 there is only the neighborhood of Eatonville with a population of 19,131 inhabitants and an average income of 36,206, it has also had a population growth of 4.3% since 2001.
- Cluster 0 groups the neighborhoods of Clairville and New Toronto that are geographically distant (see Figure 7). In addition, both do not have a larger population or higher income than the Eatonville neighborhood, and additionally, since 2001 their population has had a decrease of 4.55%.
- According to these analyzed variables, the new Bank office should be opened in the neighborhood of Eatonville, since there are no other banks around and it is the neighborhood with the best demographic indexes of the neighborhoods where there are no banks.