Comparing different cities using Foursquare API

Héctor Aristizábal
April 29th, 2020

The problem

- Jhon lives in Toronto and got a job offer in Bogota Colombia and Berlin Germany

- He wants to know which city he would feel more familiar

- He would also like to know which neigborhoods of these cities would have similar characteristics as the one he is living in right now in Toronto Canada

?

BOGOTA

?

?

BERLIN

TORONTO

?

Data sources

- A Neighborhood list of the cities of Toronto, Bogota and Berlin is available online in Wikipedia and can be found in the following links:

- **Toronto**

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- **Bogota**

- https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin#Localities

- **Berlin**

- https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin#Localities

- FOURSQUARE API

Cleansing

- Data scraped from Wikipedia website for to get the list of neigborhood for the 3 cities

- Only Borough and neighborhood information was kept and the following dataframe was built

	City	Borough	Neighborhood	Latitude	Longitude
0	Toronto	North York	Parkwoods	43.758872	-79.320292
1	Toronto	North York	Victoria Village	43.731540	-79.314280
2	Toronto	Downtown Toronto	Regent Park / Harbourfront	43.660690	-79.360310
3	Toronto	North York	Lawrence Manor / Lawrence Heights	43.723570	-79.437110
4	Toronto	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.666630	-79.393268
...	...	...	...	...	...
306	Berlin	Reinickendorf	Waidmannslust	52.575450	13.349700
307	Berlin	Reinickendorf	Lübars	52.575450	13.349700
308	Berlin	Reinickendorf	Wittenau	52.575450	13.349700
309	Berlin	Reinickendorf	Märkisches Viertel	52.596800	13.358310
310	Berlin	Reinickendorf	Borsigwalde	52.575450	13.349700

311 rows × 5 columns

Methodology

K means clustering

- To explore the cities, we will use the *Venue Recommendation* which returns a list of recommended venues near a certain location.

- The category of the places and businesses in each area is used to characterize each Neighborhood and compare them(from Foursquare)

- K means clustering to the most common venues for each Neighborhood

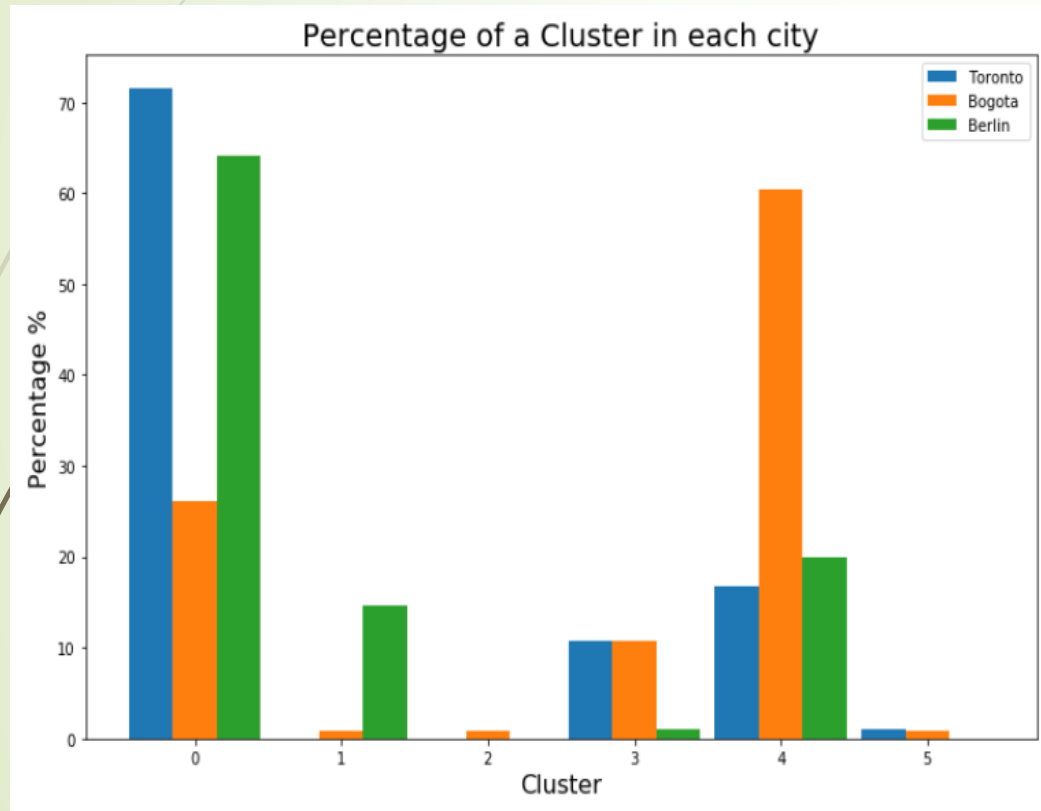- Explore different values of K

K-means clustering
Results

K=3
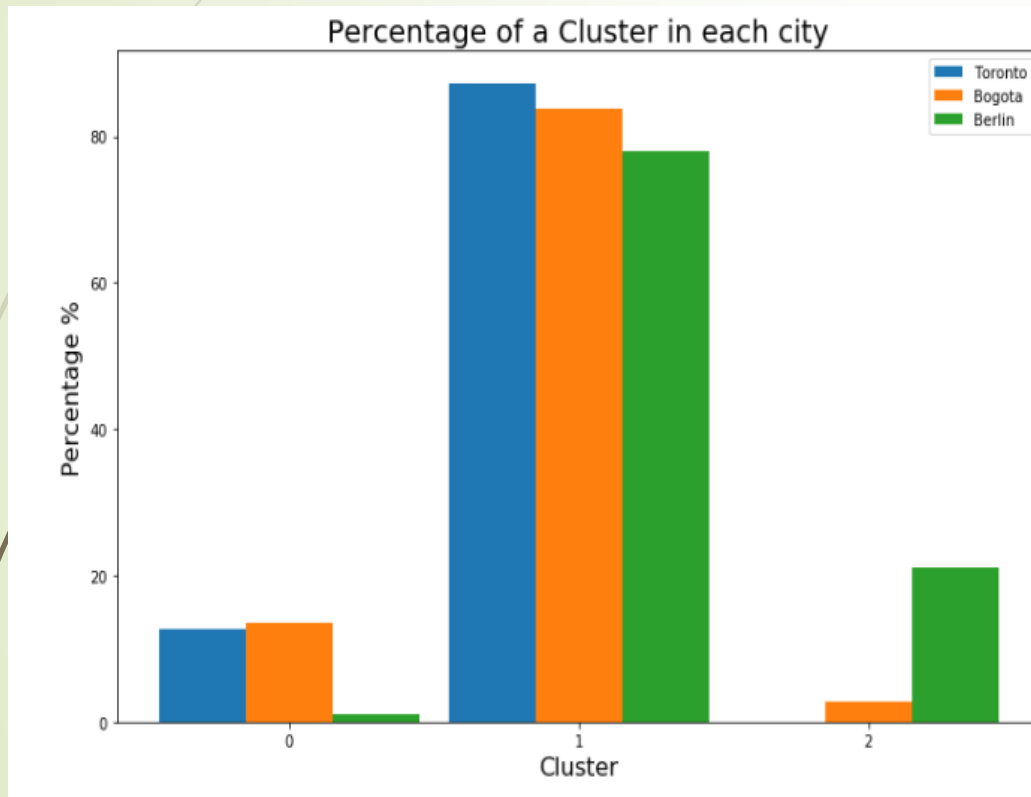
K=4
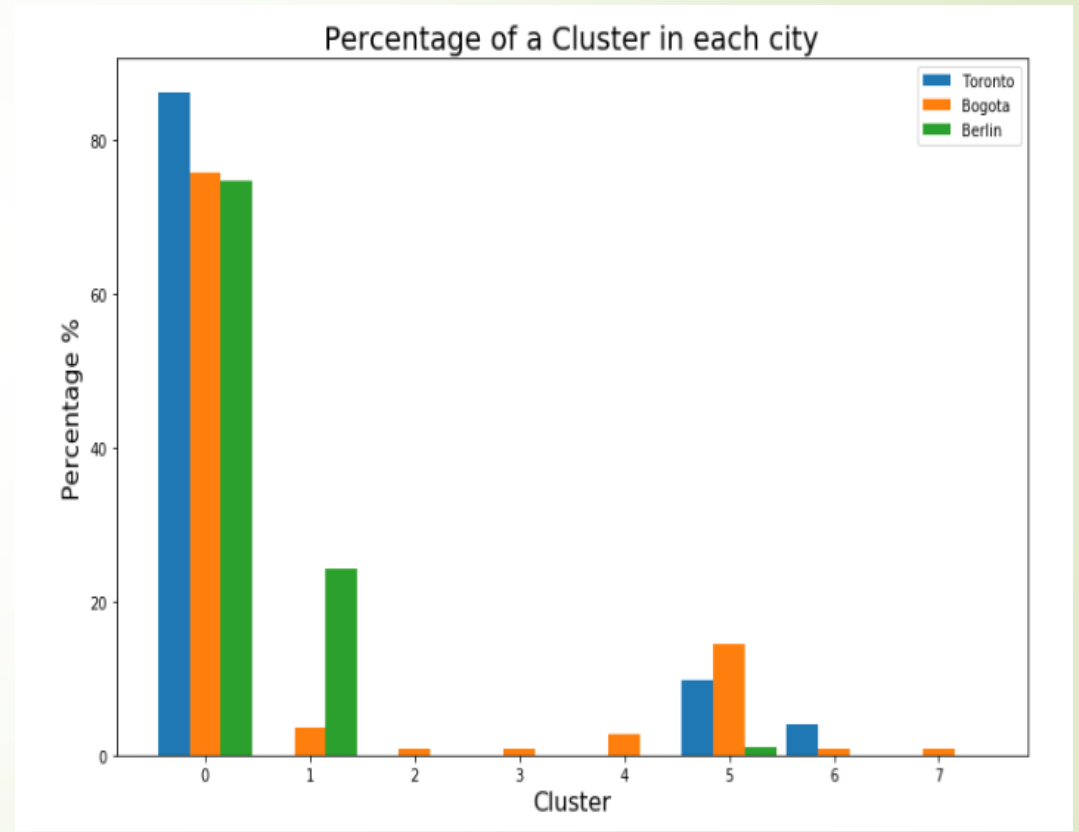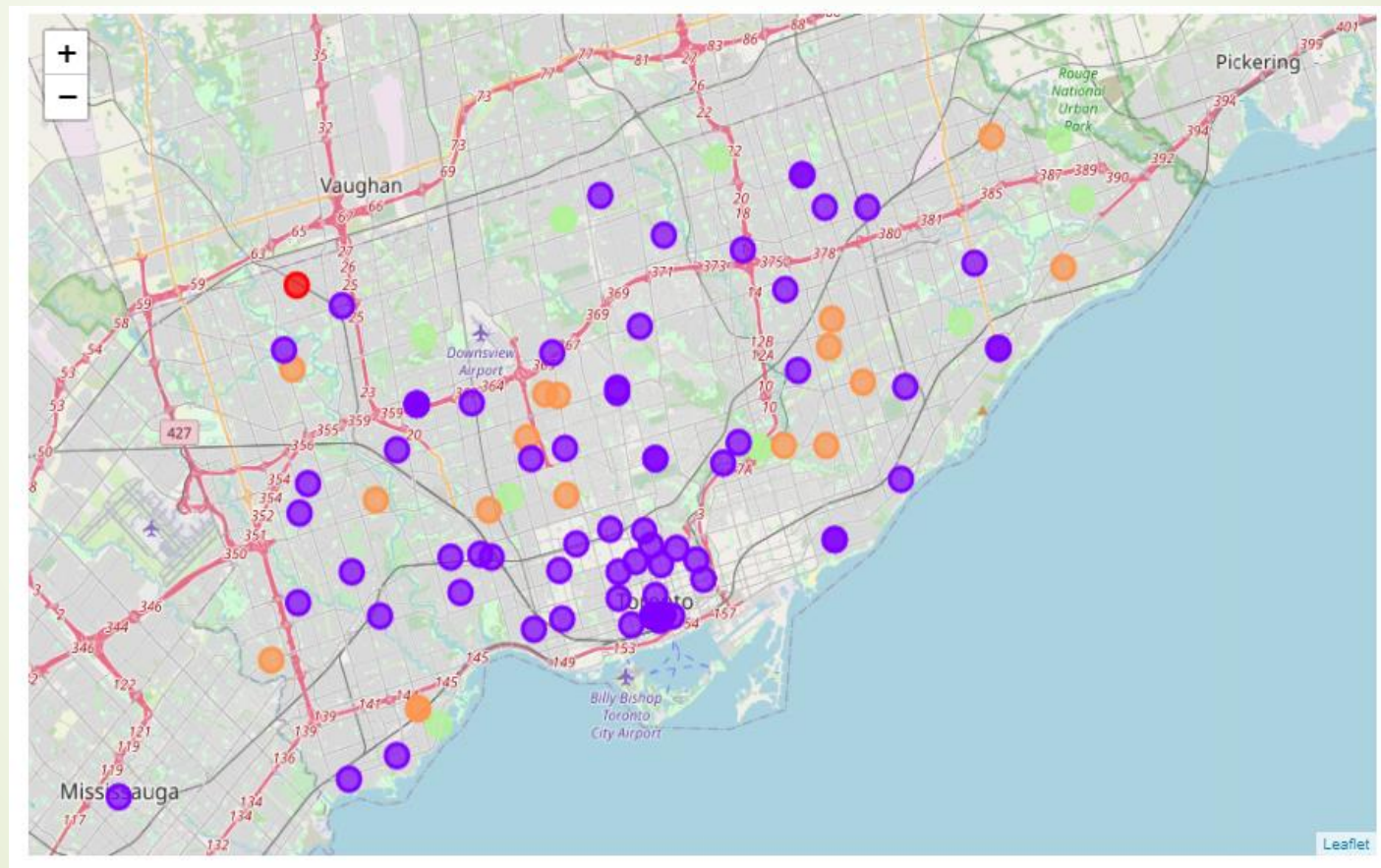
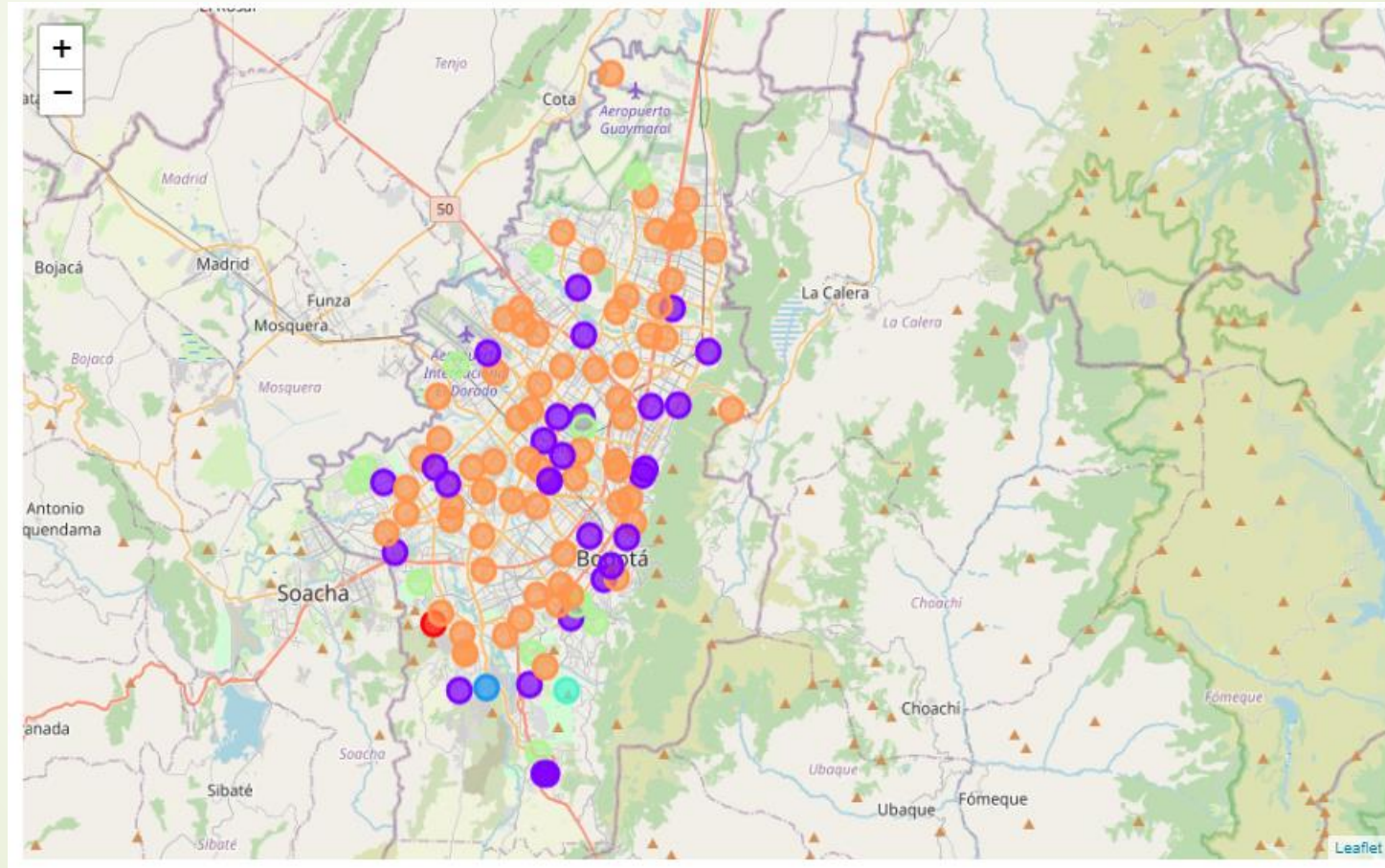# K-means clustering
Results

K=6

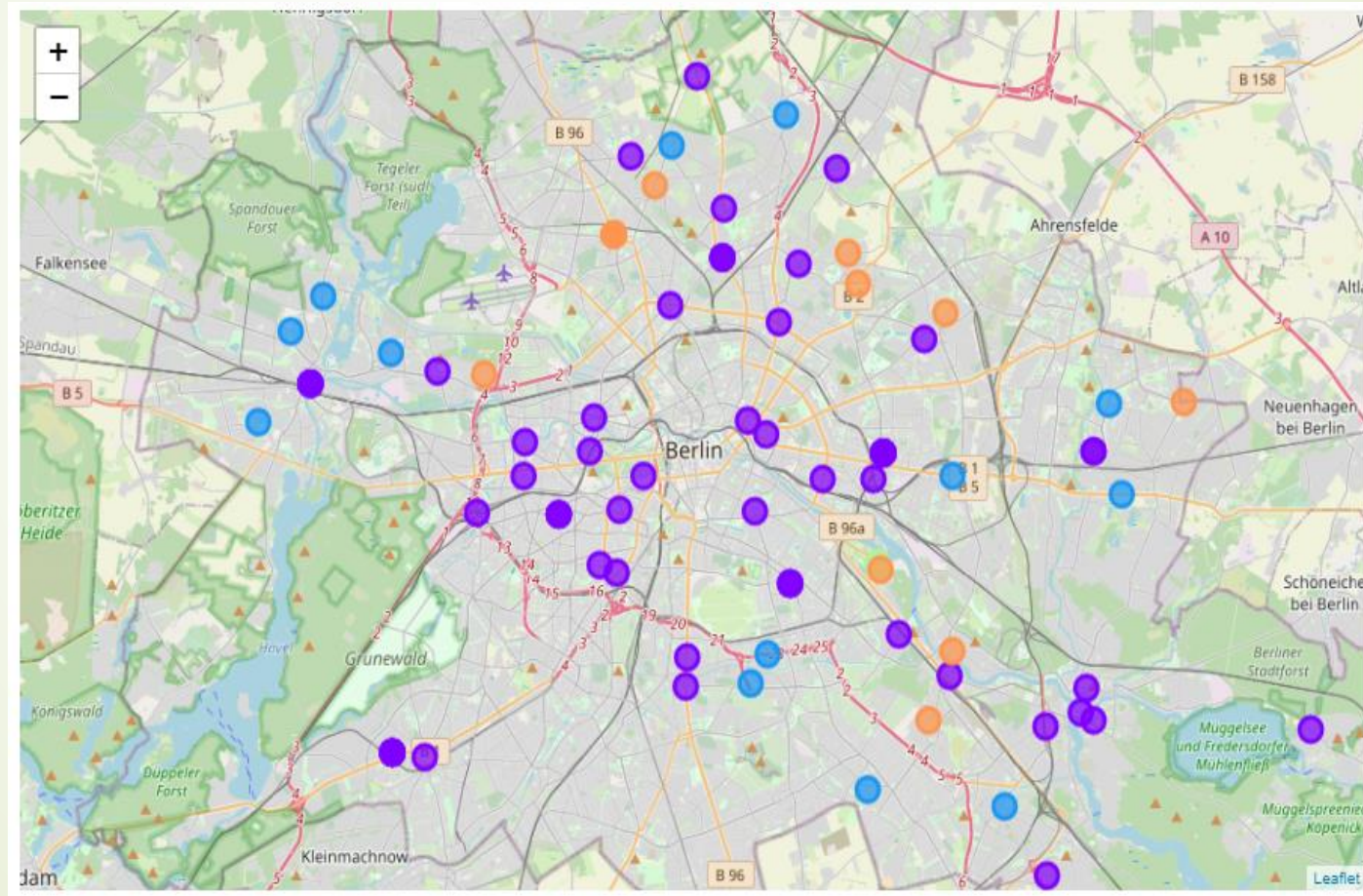K=8

Maps
Toronto K-means clustering k=6

Maps
Bogota K-means clustering k=6

Maps
Berlin K-means clustering k=6

Conclusions

- Knowing the socioeconomical context of the different countries maybe it is worthy to take the results of the analysis using k=6 more into detail. It is probably more likely that this clustering tell us more about the neighborhoods of the 3 cities, taking into account that the Foursquare API delivers a lot less results in the case of the city of Bogota. It would be interesting to add some socioeconomic data to the analysis like average income or access to healthcare and other variables in order to make the clustering much more reliable.

- After analyzing the clustering results obtained there is no much left to say that this is not an easy decision for John to make, giving that the results observed showed a lot of similarities between the cities. Based on what we have seen in the K-means clustering results John will probably find a good neighborhood to live in any of the cities analyzed.

References

- [1] Toronto neighborhoods – Wikipedia
- [2] Bogota neighborhoods – Wikipedia
- [3] Berlin neighborhoods – Wikipedia
- [4] Foursquare API