Jobs offered in two world capital city's

Introduction:

What would you choose? You just got your IBM Data science specialist certificate in your pocket and now you get jobs offered far from your home.

This is an extremely important decisions because these are two capital cities where you have never been and you only know from movies and the Internet.

Since you don't want to travel for hours (you already did that in your current job and don't want to keep this up)

These are your options:

- New York – Manhattan
- Toronto – Down Town

You might want to visit these city's but since this is very expensive (you live in Europe) there must be a better way to figure out which city suits you best.

You will use your new skillset in this case to choose the right neighborhood for you. What would you choose?

Data section

We will use the foursquare API to get data from the venues in these neighborhoods. To get this data we need the coordinates from New York and Toronto. We get there from the following locations:

-New York : https://geo.nyu.edu/catalog/nyu_2451_34572
-Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (we need to extract this list with BS4)

We will clean this data and remove all unnecessary points (what is importa nt for living there). We also need to remove characters and get all the data into the same columns. We will need our data cleaning on Python skills for this one.

We need to get this all into one Pandas data frame so we can cluster the da ta.

We will see in these neighborhoods what are the top 5 most popular venue s and then we can see what suits us best. The top venues tell a lot about a neighborhood. If there are a lot of restaurants and cinema's this we be a ve ry crowded neighborhood. Do you want that ? or would you like some plac e a little more quite.

Methodology

A way to compare these two neighborhoods is by comparing the venues ba sed similarity. This will tell a lot about a neighborhood and this will help us decide where we want to live.

In order to know we will use the venues most visited in there neighborhoo ds. For example, if people are going to parks a lot, then parks must be a str ong attribute of the neighborhood. Conversely, if there is a coffee shop in t he neighborhood but it is not visited as much, then coffee shops would not be an attribute of the neighborhood.

In our case we will use the venue information from both city's an put them together in a Pandas data frame. Then we will build 8 clusters using the 10 most visited venues in these neighborhoods. We will use the mean frequen cy of visits to determine the top 10. Then we will run a K-Means to clusters these into 8 clusters. So this will let use see which neighborhoods are simila r and which don't . then we use it's characteristics to determine where we would wont to move to.

Results

Cluster 1 This was all New York neighborhoods that was heavily Italian Rest aurants and then fitness such as gyms and yoga
Cluster 2 This was a single Toronto neighborhood that consisted of outdoor activities such as parks, trails and playgrounds
Cluster 3 This was all Toronto neighborhoods that probably surround the ai rport since Airport terminal and airport lounge was all that was included
Cluster 4 Mostly a New York cluster with one Toronto neighborhood. This was mainly cafe and hotels

Cluster 5 Single Toronto neighborhood consisting of Cafes, Grocery Stores a nd parks
Cluster 6 Single New York neighborhood consisting of bars, parks and ferry service
Cluster 7 A split of New York and Toronto neighborhoods with a variety of r estaurants as their main venues
Cluster 8 A split of New York and Toronto neighborhoods with a heavy focu s on cafes

Discussion

After reviewing my results, I am very surprised that there are some neighb orhoods that are unique to the city except for the Toronto neighborhoods t hat has or borders an airport.
We know that Manhattan does not have any airports since they are across the river in Queens.
You can see in Cluster 2 that there seems to be a single neighborhood mor e outdoor activities. It must be on the outer part of the city that borders a s uburban type neighborhood. This makes sense since Manhattan is bordere d by other large boroughs such as Queens and the Bronx.
Cluster 5 seems more like residential area since it has grocery stores and pa rks that is unique to Toronto
Cluster 6 is a unique neighborhood that is focused heavily on the ferry with bars. This must be the neighborhood where commuters are coming into, Cl uster 7 & 8 are similar to both City's. They do contains a lot of neighborhoo ds in both city's.

Conclusion

As we see when clustering. These boroughs look a lot like each other. They have a lot in common.
I think my personal conclusion would be that I would choose a neighborho od from New York. Battery Park would be a great option.
I love the American people and Battery park has some great fitness options what I think is really important to stay in shape and healthy plus I love a go od walk at the park. Battery park is a very nice park since I have been there in the beginning of this year.