Python Capstone Course Final Report

25 July 2021

Comparison of New York and Toronto:

Which city is best to attract 25-35 year old millennials?

Introduction

Business Problem: **Large technology corporations need a data driven method to choose new R&D HQs that will attract highly educated millennials.** These large technology corporations are having trouble attracting able-minded millennials with significant coding and technology experience. Millennials are demanding better work-life balance and to work in interesting cities.

The corporations cannot grow in Silicon Valley and Seattle areas any longer. High taxes, access to housing concerns, and overall decrease in quality of life since 2000, are driving knowledge workers to other regions.

> _Hypothesis:_
>
> _Analysis of Quality of Life of two metro-urban centers can provide these corporations with alternatives._
>
> _One factor in Quality of Life for millennials is how "liveable" a city is_

Data and Sources

Method of Analysis: Determine "liveable" factors between New York and Toronto by determining how neighbourhoods are grouped, and the amenities these neighbourhoods offer

Data Preparation: Import Neighbourhood data from New York and Toronto sourced from course data file for New York: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

and for Toronto using webscraping of Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Methodology

Data Analysis: Use K-means clustering to find how many neighbourhoods are in each urban center. This will lead to the comparison: How are neighbourhoods structured, their distance from each other, and their typical amenities.

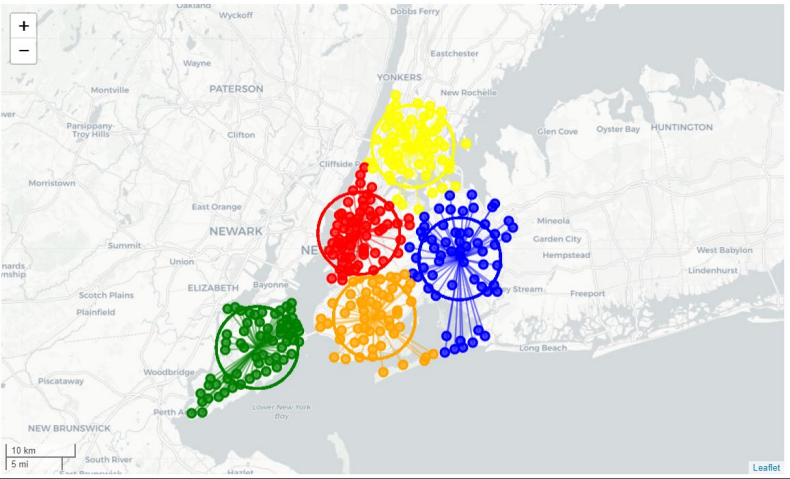k-means clustering done using sklearn KMeans with n = 5

Results

K-means clustering of the two cities: K-means clusters into 5 clusters for New York, and add centroid and color:

New York City:

Toronto, metro area:

```
Toronto_data.head()
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

|	PostalCode	Borough	Neighborhood	Latitude	Longitude	cluster	cen_x	cen_y	color	x_line	y_line
0	M3A	North York	Parkwoods	43.753259	-79.329656	4	43.700213	-79.324393	yellow	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572	4	43.700213	-79.324393	yellow	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2	43.662429	-79.397089	red	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	3	43.750577	-79.414192	orange	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494	2	43.662429	-79.397089	red	43.662301	-79.389494
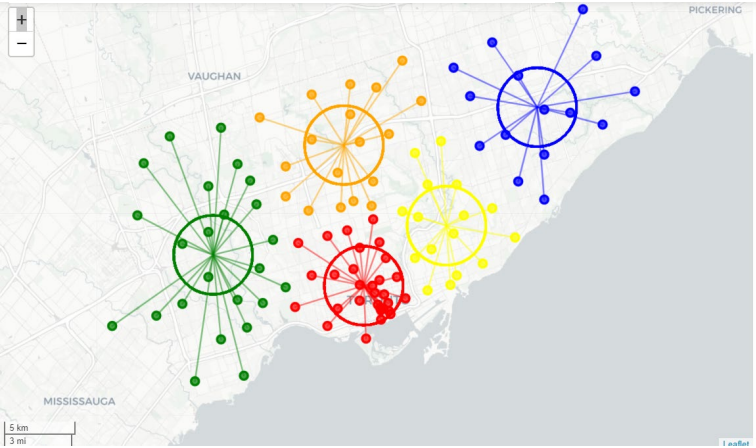
Further refinement was done in order to concentrate only on Manhattan and Toronto core area:

Manhattan:

```
manhattan_data.head()
```

|	Borough	Neighborhood	Latitude	Longitude	cluster	cen_x	cen_y	color
0	Manhattan	Marble Hill	40.876551	-73.910660	4	40.843482	-73.876488	yellow
1	Manhattan	Chinatown	40.715618	-73.994279	2	40.734036	-73.966889	red
2	Manhattan	Washington Heights	40.851903	-73.936900	4	40.843482	-73.876488	yellow
3	Manhattan	Inwood	40.867684	-73.921210	4	40.843482	-73.876488	yellow
4	Manhattan	Hamilton Heights	40.823604	-73.949688	4	40.843482	-73.876488	yellow
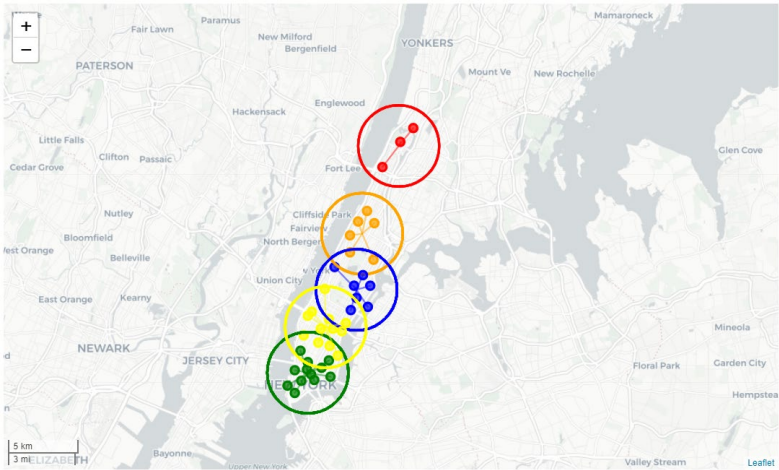
Toronto Core:

By selecting the main neighbourhoods in the city center.

```
central_list = ["Queen's Park", "East York", "East Toronto Business", "Downtown Toronto Stn A",
                "Downtown Toronto","Central Toronto", "Etobicoke", "West Toronto", "East York/East Toronto","West Toronto"]
Toronto_downtown = Toronto_data.loc[Toronto_data['Borough'].isin(central_list)]
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude	cluster	cen_x	cen_y	color	x_line	y_line
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1	43.654949	-79.390928	red	43.654260	-79.360636
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494	1	43.654949	-79.390928	orange	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	2	43.681159	-79.556595	orange	43.667856	-79.532242
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	3	43.691040	-79.327472	blue	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	1	43.654949	-79.390928	blue	43.657162	-79.378937
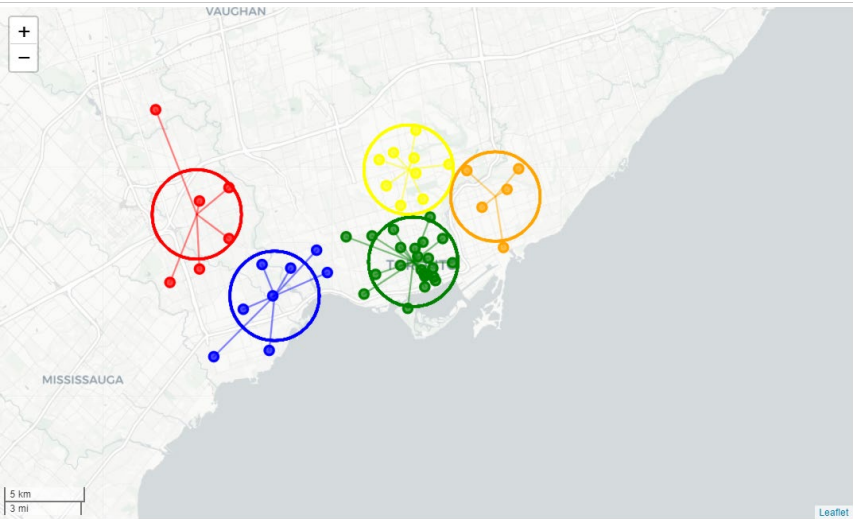
Merging of both NYC and Greater Toronto:
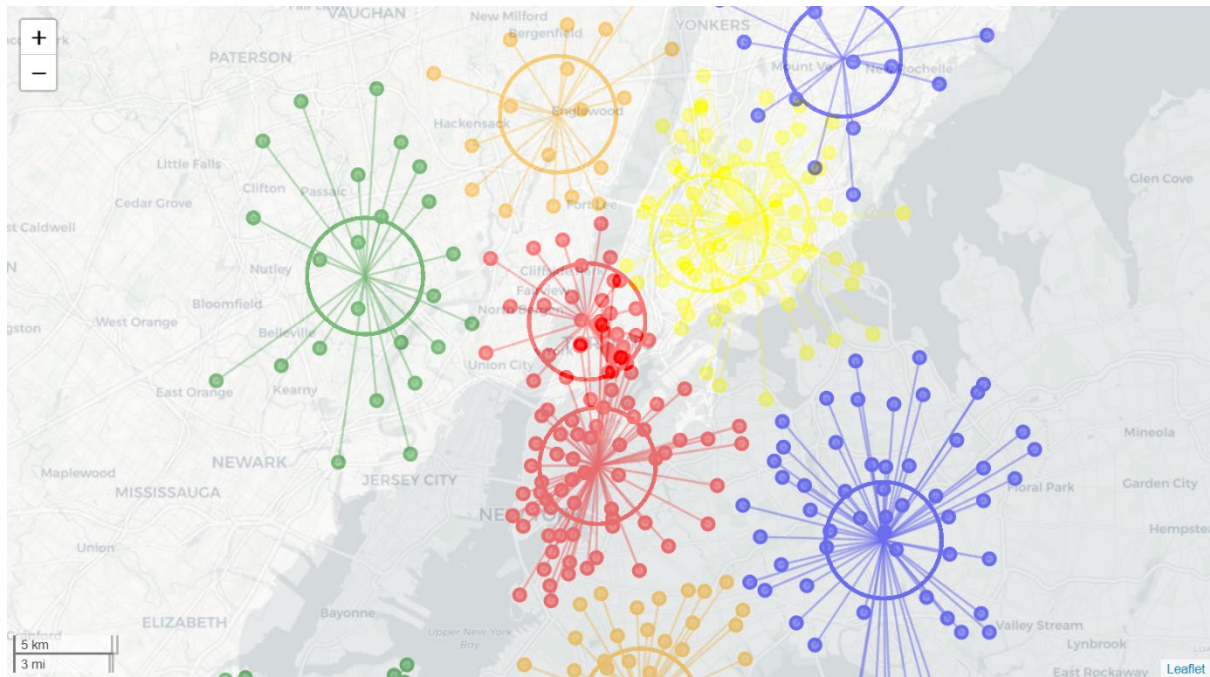
Merging of just downtown cores:
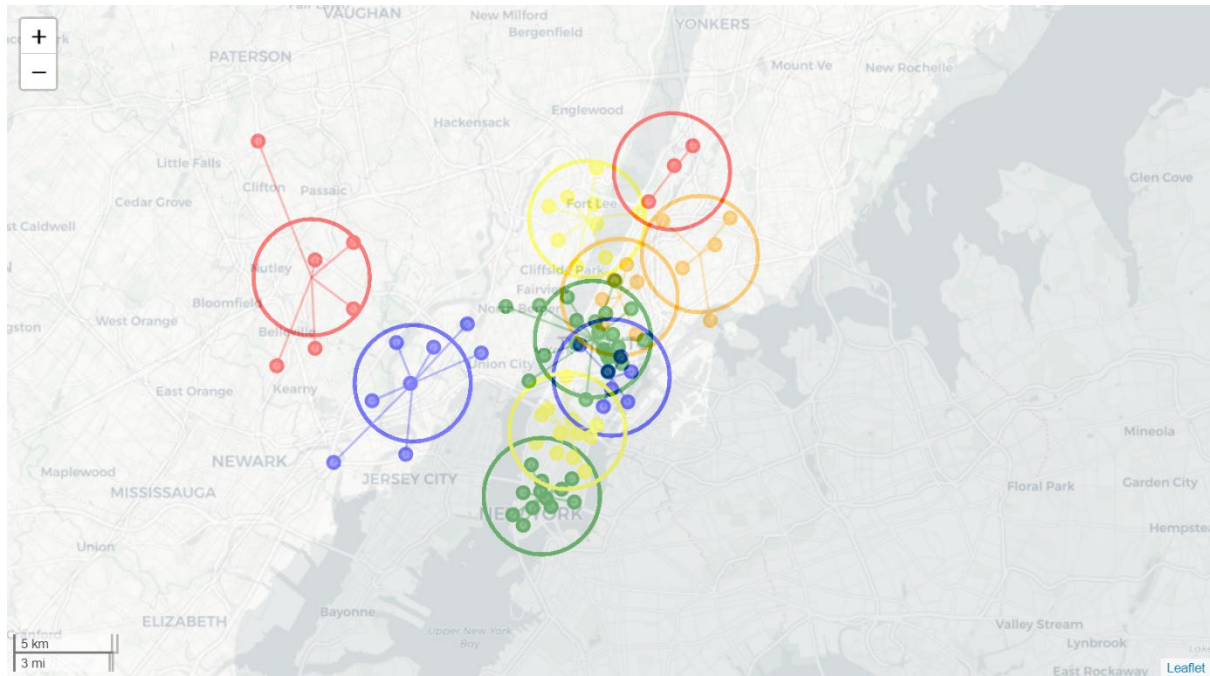
Discussion

Superimposing the two downtown cores, it is evident that NYC has a more concentred group of interesting neighborhoods.

Conclusion: NYC may provide a better Quality of Life than Toronto when determined by neighbourhood clustering

Conclusion

Solution to Business Problem: For Large technology corporations to attract highly education Millennials, they must place their technology headquarters in interesting urban areas with a high Quality of Life. The decision on which urban area to choose can be made by analysing aspects of Quality of Life that are important to Millennials.

Demonstration of hypothesis:

The Analysis demonstrates that Quality of Life can be explained by k-means clustering of neighbourhoods, to show that interesting amenities are grouped which makes the city more liveable.

Future Analysis can be expanded to include different cities, including European metro-urban areas. Further, it would be useful to use different metrics, such as access to medical care, biking routes, art and leisure.