Similarities of the neighborhood structure of Capital Cities

1. The Problem

Due to the more and more growing Population in this world combined with the rural exodus, cities are growing faster than ever before. Managing such big cities is not an easy task. To achieve a high standard of living for the inhabitants it is necessary to provide a lot of public infrastructure for instance continuous electricity, tap water and sewage as well as public transportation. So, lets focus on public transportation, because it is the most flexible one and therefore easiest to improve. Public transportation providers are facing different issues. In some cities this working fine, in others it is not. So, there is the question, if there are similarities between the neighborhoods of big cities. If there any similarities, good working solutions for public transportation of some cities can be applied to other ones. The same principle might be practical for other infrastructure like water and electricity. Thus, the target audience of this analysis might public transportation provider as well as the city administration. While the city administration might improve their processes and the living standards for the inhabitants, the public transportation provider might increase their revenue with optimized offers.

2. Data

Based on the neighborhood data as well as the venues in the neighborhoods both cities Toronto and New York City will be compared. With the coordinates of each neighborhood the venues will be retrieved from the Foursquare API. This Data will be clustered and drawn out on a map to show similarities or dissimilarities between the countries.

The used Dataset looks like this with 12248 entries:

	City	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	NewYork	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	NewYork	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	NewYork	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	NewYork	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
4	NewYork	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

3. Methodology

3.1. Statistical Methods

Firstly, the structure of both cities based on the occurrence has been compared. Due to the big difference in the number of venues for each city, the relative occurrence of a venue category where chosen. The top 15 categories of each city were compared, and differences and common top categories was demonstrated.

3.2. Machine Learning Method

In the following the neighborhoods of both cities were clustered by their relative occurrence of venue categories in a neighborhood. For this, each proportion of a category in a neighborhood where calculated. Then they were clustered only by the category proportion into three different clusters with the K-Means algorithm.

4. Results

4.1. Comparison of top categories

So, there are some similarities in the occurrence of various venue categories in both cities. Nine of the Top 15 categories of both cities are equal and 6 categories per city are only in the top 15 in one city. At the chart we see, that the proportion of most categories are quite equal, but we also see, that for instance Coffee Shop and Cafe are more present in Toronto, whereas Donut Shops and Pizza Places are mor e present in New York. This shows that it might be of interest to explore the date further for similarities of both cities.

Common Categories of both cities:

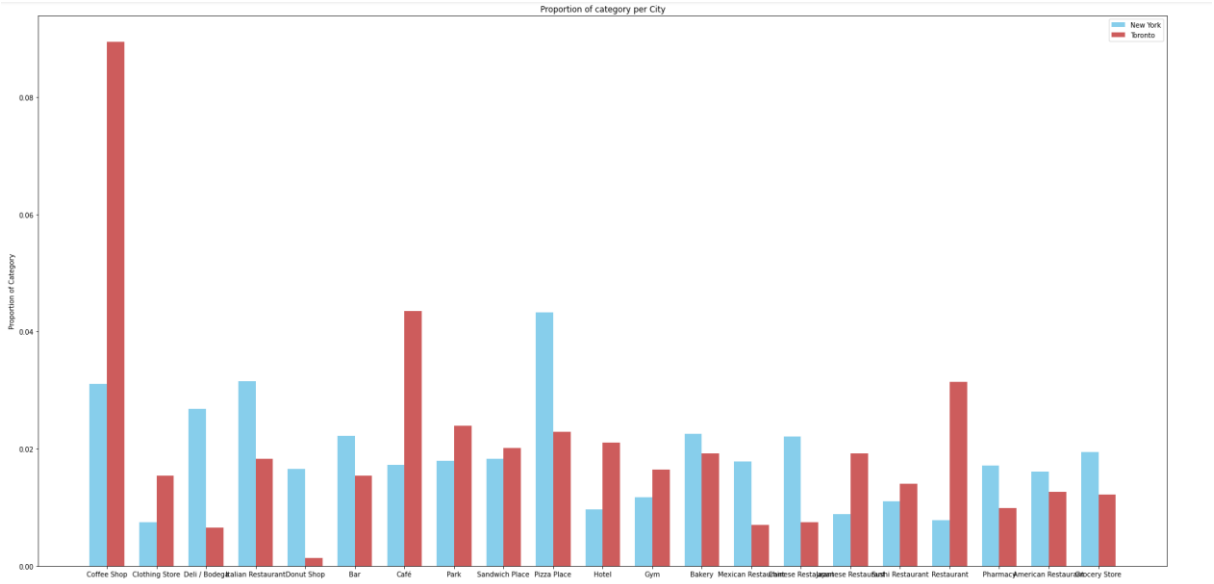
American Restaurant, Bakery, Bar, Café, Coffee Shop, Italian Restaurant, Park, Pizza Place, Sandwich Place

Top Categories of New York only:

Chinese Restaurant, Deli / Bodega, Donut Shop, Grocery Store, Mexican Restaurant, Pharmacy

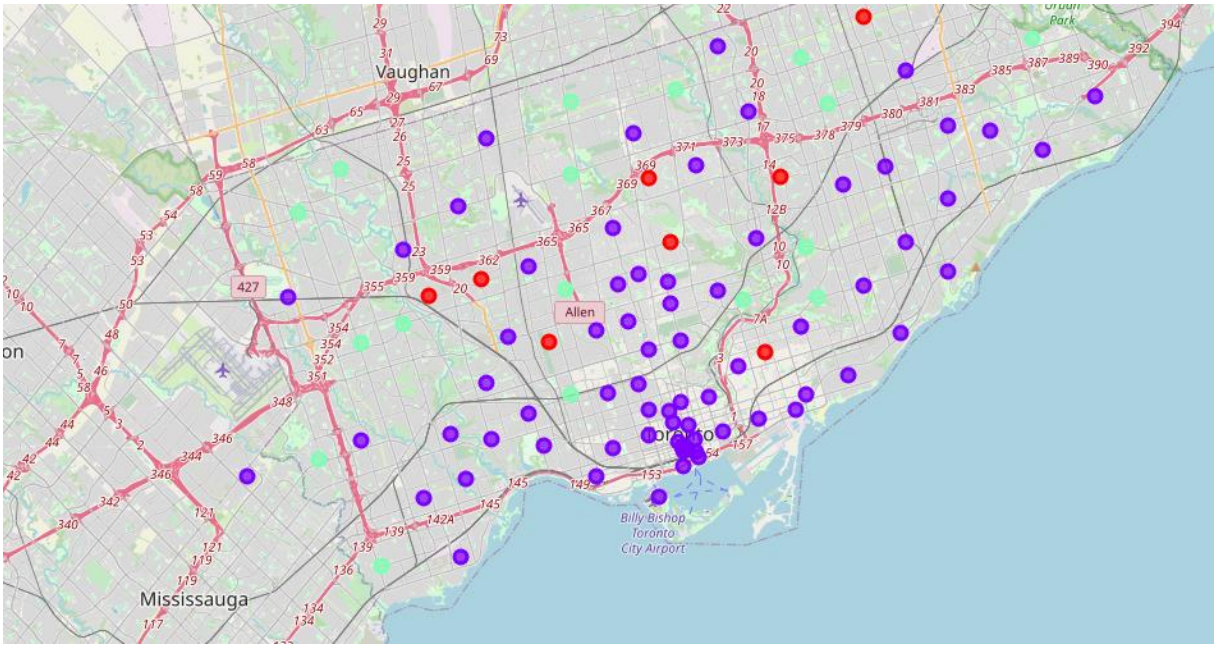
Top Categories of Toronto only:

Clothing Store, Gym, Hotel, Japanese Restaurant, Restaurant, Sushi Restaurant

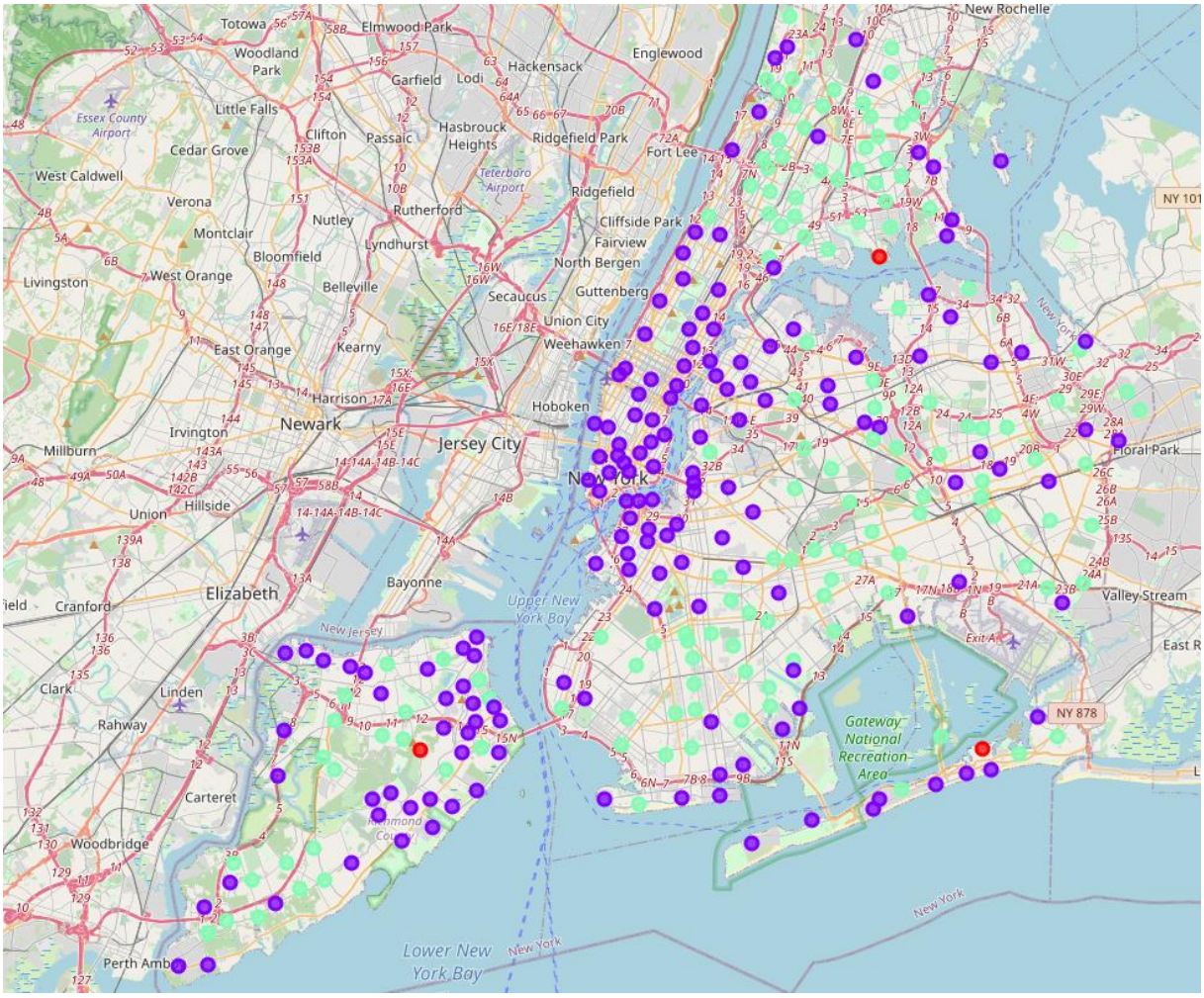


4.2. Clustering Results

Map of Toronto:



Map of New York:



The clustering shows, that in both cities, Category purple is most represented, whereas category blue is more represented in New York than in Toronto. The third red category is not represented well at all.

5. Discussion

The Results show that the structure of the neighborhood, has similarities but are not completely equal at all. There are different types of categories and in Toronto it is mostly one category. Regarding the problem explained at the beginning, it might be said, that may be infrastructure problems which are not solved in category purple neighborhoods in New York can be solved by solutions in category Purple neighborhoods in Toronto or vice versa. Unfortunately, this analysis lacks data for other cities in other countries. Maybe there will be better results with more Information

6. Conclusion

The Problem in this analysis is, that the obtained information might not be generalized, since there are only 2 cities which has been compared. Maybe the categories were a bit to fine-grained and it will bring better results with combined categories for e. g. restaurants as combination of Italian Restaurant, American Restaurant, etc. Also the density, population as well es companies could be of interest for comparing the city structure.