REPORT

Battle Of Neghborhoods

1.Introduction

A group of businessmen are going to open a restaurant of Ukrainian cuisine. They have in mind two major cities: New York and Toronto. The intended group of visitors is people living or working nearby, wanting a tasty and inexpensive lunch or dinner, possibly tourists. Important information when choosing a place will be remoteness from the city center, the availability of tourist places, train stations, as well as other restaurants.

We need to choose a place for a new restaurant. It should be noted that the description of the client is not entirely specific, it can be both tourists and couples living nearby. The problem is the lack of an indicator that has borders that can be checked for compliance. Rather, we are talking about the optimal combination of a number of factors, some of which have already been taken into account when dividing the city neighborhoods into clusters, so we have to focus on studying them. Interest in this problem may arise among people who are going to start their own business (not necessarily restaurant) and choose the best place for these purposes.

2.Data

We explored New York City and the city of Toronto and segmented and clustered their neighborhoods. We also have data on the location of the neighborhoods of these two cities (table 1 and table 2 respectively).

Table 1. New York City Dataframe

Dataframe's columns	Data Types
Borough	object
Neighborhood	object
Latitude	float64
Longitude	float64
Cluster Labels	float64
1st Most Common Venue	object
2nd Most Common Venue	object
3rd Most Common Venue	object
4th Most Common Venue	object
5th Most Common Venue	object

6th Most Common Venue	object
7th Most Common Venue	object
8th Most Common Venue	object
9th Most Common Venue	object
10th Most Common Venue	object

Table 2. Toronto Dataframe

Dataframe's columns	Data Types
Borough	object
Neighborhood	object
Latitude	float64
Longitude	float64
Cluster Labels	float64
1st Most Common Venue	object
2nd Most Common Venue	object
3rd Most Common Venue	object
4th Most Common Venue	object
5th Most Common Venue	object
6th Most Common Venue	object
7th Most Common Venue	object
8th Most Common Venue	object
9th Most Common Venue	object
10th Most Common Venue	object

Thus, it is necessary to study these clusters, paying attention to the availability of tourist places, other restaurants, train stations (venue category). It also seems possible to calculate the average remoteness of the neighborhoods included in each cluster from the city center. Thus, the answer will be preferred one or more neighborhoods that satisfy the client's request.

3.Methodology

Let's start exploring NY clusters with counting the number of neighborhoods in them.

Table 3. NY clusters' division when the number of clusters is 5.

Cluster	Number of Neighborhoods
0	280
2	19
4	4

1	1
3	1

We can immediately see that the division into clusters is not optimal, since the vast majority of neighborhoods belong to one cluster. Therefore, we change the number of clusters. It should be noted that the optimal number of clusters in this case was found by enumerating their number from 5 to 10 based on the criterion of uniform distribution of neighborhoods across clusters. The number of clusters in the New York area has been changed to 8.

Table 4. NY clusters' division when the number of clusters is 8.

Cluster	Number of Neighborhoods
2	134
6	72
4	72
1	9
0	9
5	5
3	2
7	1

It looks better. So let's count the number of neighborhoods in Toronto's clusters.

Table 5. Toronto clusters' division when the number of clusters is 5.

Cluster	Number of Neighborhoods
1	162
0	27
3	10
4	5
2	3

Let's select the data of the cluster which has the largest number of neighborhoods in NY and see it's 1st Most Common Venue category.

Table 6. The 1st Most Common Venue category of the first cluster in NY.

Venue category	1st Most Common Venue
Italian Restaurant	23

Bar	15
Coffee Shop	11
Chinese Restaurant	6
Café	5
Grocery Store	5
Clothing Store	4
Pizza Place	3
Park	3
Burger Joint	3

So we see here 10 caribbean restaurant's, several other restaurants, supermarkets, banks. I suppose it could be a good place to relax, may be situated near the see or some park zones.

Let's see the 2nd most common venue category of the first cluster in NY.

Table 7. The 2nd Most Common Venue category of the first cluster in NY.

Venue category	2nd Most Common Venue
Coffee Shop	15
Italian Restaurant	14
Pizza Place	11
Park	7
Deli / Bodega	5
American Restaurant	5
Bakery	4
Cosmetics Shop	4
Café	3
Bar	3

And here we really see 7 parks.

Let's select the data of the cluster which has 72 neighborhoods in NY. Let's see it's first & second most common places.

Table 8. The 1st Most Common Venue category of the Second Cluster in NY.

Venue category	1st Most Common Venue
Pizza Place	26
Pharmacy	8

Grocery Store	5
Bank	4
Italian Restaurant	3
Bus Station	3
Chinese Restaurant	2
Donut Shop	2
Buffet	1
Deli / Bodega	1

Table 9. The 2nd Most Common Venue category of the Second Cluster in NY.

Venue category	2nd Most Common Venue
Pizza Place	17
Bank	4
Italian Restaurant	4
Donut Shop	3
Deli / Bodega	3
Bagel Shop	3
Mexican Restaurant	3
Food Truck	2
Food	2
Park	2

We see a lot of pizza places, several banks, pharmacies, grocery stores. I suppose this cluster composes usual places for living middle-class people. Let's check it out later and go to next cluster which combines either 72 neighborhoods.

Table 10. The 1st Most Common Venue category of the Third Cluster in NY.

Venue category	1st Most Common Venue
Deli / Bodega	10
Caribbean Restaurant	10
Donut Shop	6
Bank	5
Chinese Restaurant	4
Fried Chicken Joint	3
Pharmacy	3
Mobile Phone Shop	3
Bus Station	2

Supermarket	2

Table 11. The 2nd Most Common Venue category of the Third Cluster in NY.

Venue category	2nd Most Common Venue
Donut Shop	5
Fried Chicken Joint	4
Grocery Store	4
Deli / Bodega	4
Chinese Restaurant	4
Fast Food Restaurant	3
Bank	3
Cosmetics Shop	2
Gym	2
Bus Station	2

The third cluster at first glance is something alike to the first one, but instead of parks - two baseball fields.

So we studied the most numerous clusters of New York. Now let's get started on exploring Toronto clusters.

Table 12. The 1-st Most Common Venue category of the First Cluster in Toronto.

Venue category	1st Most Common Venue
Coffee Shop	26
Café	14
Skating Rink	11
Grocery Store	11
Pizza Place	10
Airport Lounge	7
Pub	6
Convenience Store	6
Clothing Store	5
Bakery	5

Table 13. The 2-nd Most Common Venue category of the First Cluster in Toronto.

Venue category	2nd Most Common Venue
Coffee Shop	23
Pizza Place	17
Café	13
Sandwich Place	10
Restaurant	9
Airport Service	7
Bakery	6
Women's Store	4
Chinese Restaurant	4
Fast Food Restaurant	4

So it looks like a place near the airport, as there are Aiport Lounges & Airport Services, respectively, a lot of coffee houses. There are also 11 Skating Rink and a lot of places to have a snack.

Table 13. The 1-st Most Common Venue category of the Second Cluster in Toronto.

Venue category	1st Most Common Venue
Park	11
Playground	6
Construction & Landscaping	3
River	3
Trail	2
Swim School	1
Pub	1

Table 14. The 2-nd Most Common Venue category of the Second Cluster in Toronto.

Venue category	2nd Most Common Venue
Park	7
Coffee Shop	4
Pizza Place	4
Jewelry Store	2
Gym	2
Airport	2
Women's Store	1

Metro Station	1
Food & Drink Shop	1
Trail	1

So cluster 2 in Toronto is some kind of a park's place, a place for relax.

Table 15. The 1-st Most Common Venue category of the Second Cluster in Toronto.

Venue category	1st Most Common Venue
Baseball Field	10

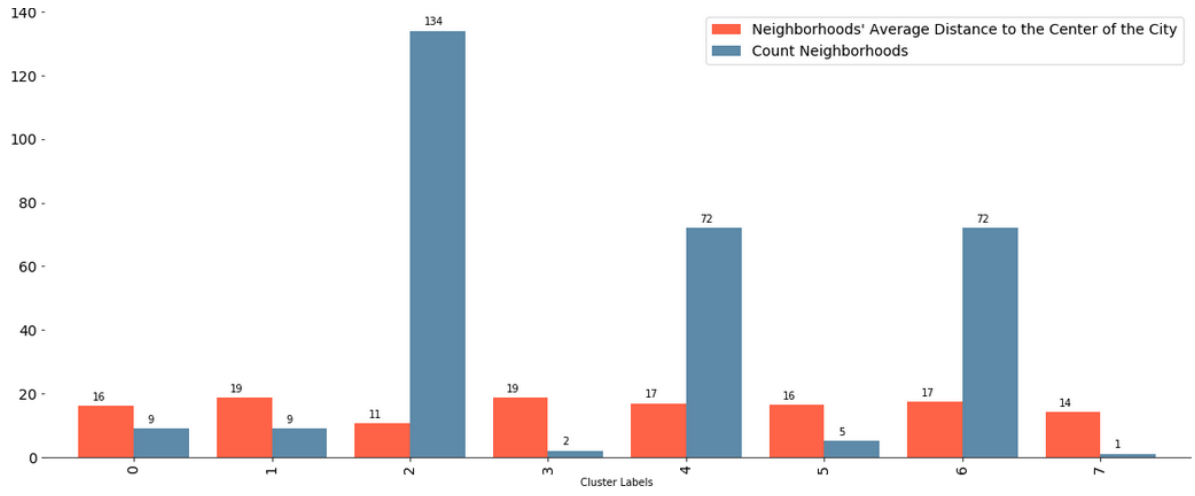
So it looks like we studied the most numerous clusters of Toronto.

Let's now estimate the average distance to the city center of the neighborhoods included in each cluster of New York and Toronto, respectively.

4.Result

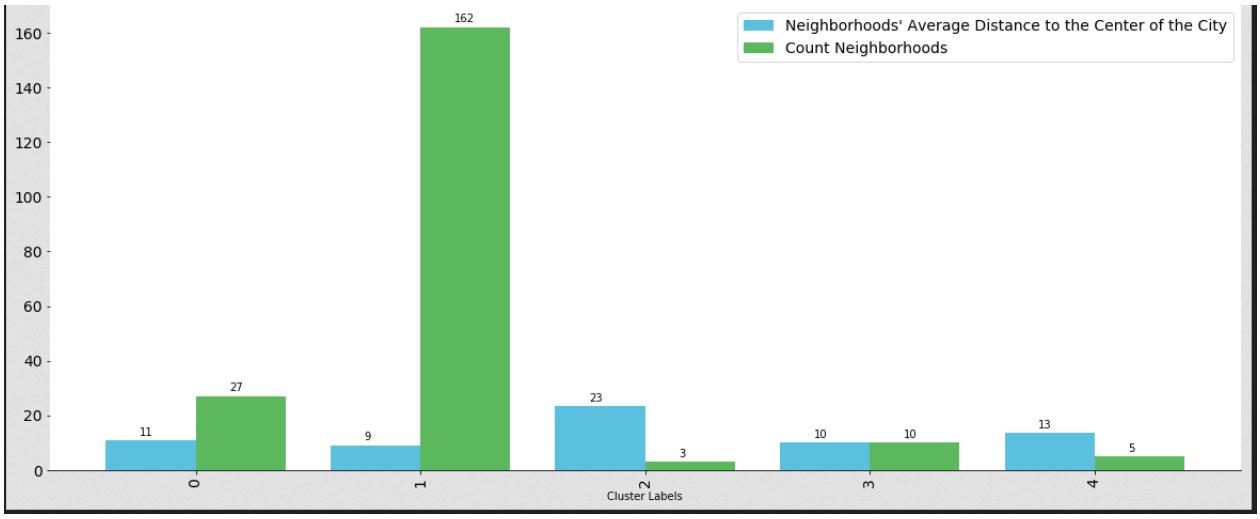
In each of the databases (New York and Toronto), we have latitude and longitude corresponding to each neighborhood. With their help, the distance of each neighborhood to the city center was first calculated. Then, the average distance of the regions included in each cluster to the city center was calculated.

Let us summarize the information we have about the number of neighborhoods included in each cluster and their average distance to the city center using a bar chart.



Char 1. NY Clusters' Neighborhoods Quantity And Average Distance to the Center of the City

As we see from the char 1 the most numerous cluster of NY with the number of neighborhoods 134 has the shortest average neighborhoods distance to the center of the city is among all the clusters – 11 kilometers.



Char 2. Toronto Clusters' Neighborhoods Quantity And Average Distance to the Center of the City

As we see from the char 2 the most numerous cluster of Toronto with the number of neighborhoods 162 has the shortest average neighborhoods distance to the center of the city is among all the clusters – 9 kilometers.

5.Discussion

Of the 8 clusters in New York, the most numerous were 3 with the number of neighborhoods 134, 72 and 72, respectively.

1. In the first cluster of NY we see a lot of restaurants, several supermarkets, banks. I suppose it could be a good place to relax. It is situated near the see or park zones. Good place for tourists and entertainment of any kind. The average neighborhoods distance to the center of the city is the shortest among all the clusters – 11 kilometers.
2. In the second cluster of NY we see a lot of pizza places, several banks, pharmacies, grocery stores. It consumes neighborhoods areas partly situated near the beach zones. I suppose this cluster composes usual places for living middle-class people. The average neighborhoods distance to the center of the city is 17 kilometers.
3. The third cluster at first glance is something alike to the first one, but instead of parks - two baseball fields and the average neighborhoods distance to the center of the city is 17 kilometers.

Of the 5 clusters in Toronto, the most numerous were 3 with the number of neighborhoods 162, 27 and 10, respectively.

1. The first cluster of Toronto looks like a place near the airport, as there are Airport Lounges & Airport Services, respectively, a lot of coffee houses. There are also 11 Skating Rink and a lot of places to have a snack. The average neighborhoods distance to the center of the city is the shortest among all the clusters - 9 kilometers.
2. The second cluster in Toronto has 7 parks and not very much places to have a snack. It is some kind of a park's place, a place for relax. The average neighborhoods distance to the center of the city is 11 kilometers.
3. The third cluster in Toronto has 10 baseball fields and several shops. The average neighborhoods distance to the center of the city is 10 kilometers.

6.Conclusion

I would recommend opening a restaurant in the first cluster of New York. Yes, there will be high competition, but this place clearly attracts a large number of people. Regarding Toronto, there are clearly fewer restaurants, but this is most likely due to less demand for such places. If the owner is still inclined to choose Toronto for starting a business, I would recommend additional research focused on Toronto.