

HealthZone Restaurant:
New York City Location Search Report

Report By David Skoff
July 12th, 2020

Table of Contents

Introduction	Page 3
Data	Page 4
Methodology	Page 5
Results	Page 6
Discussion	Page 8
Conclusion	Page 9

1. Introduction

HealthZone restaurant, privately founded in 2014, has been successfully operating near downtown Toronto for over six years. During this time, loyal customers have praised the restaurant for incredible food and atmosphere which in turn has established HealthZone as a cornerstone of their neighborhood.

HealthZone has a single location in the Rosedale neighborhood on the outskirts of downtown Toronto. This neighborhood is filled with parks, trails and playgrounds. The neighborhood is occupied by residents that value health, exercise and healthy food. It is just west of the Lower Don River which provides great paths for running and biking, so everyone from casual exercisers to athletes funnel into this location from multiple directions. Again, many of these passersby frequently visit the HealthZone restaurant.

The private owners of HealthZone are now looking to expand to a second location in the New York City area. They are targeting another neighborhood that mimics the Rosedale (Toronto) neighborhood in its venues (parks, trails, playgrounds) because of the type of people that these venues attract. Finding people that value healthy food options is critical to the success of the upcoming New York City location.

To this end, the owners of HealthZone have commissioned our data analytics firm (DataBeGood) to source and analyze data to help determine similar neighborhoods in New York City that they can use to narrow the search for their second location. This report outlines the data, methods, results, discussion and conclusion of our analysis.

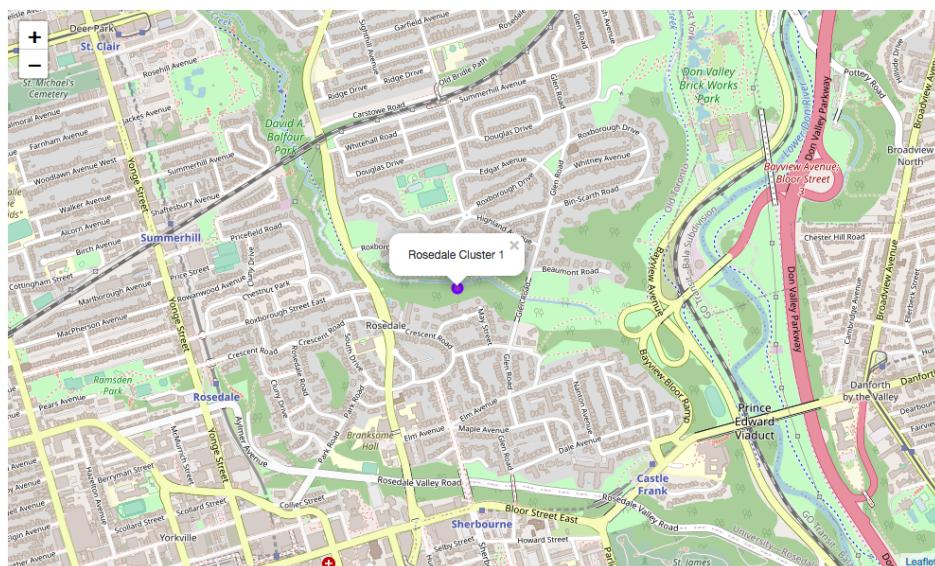


Figure 1: Map of area surrounding Rosedale neighborhood in downtown Toronto

2. Data

The aim of the DataBeGood analysis firm was to collect data describing the popular venue types (parks, restaurants, retail stores, etc.) for each neighborhood near downtown Toronto and for each neighborhood in all five boroughs of New York City which could be potential locations for the second HealthZone restaurant.

To do this, the first step was sourcing the latitude and longitude data for each neighborhood. For Toronto, the postal codes, boroughs and neighborhoods were scraped from Wikipedia. An online repository (IBM/Coursera) was used to source a mapping of postal codes to latitude and longitude data. Afterwards, the data was joined via the postal codes in each data set so that the borough and neighborhood each had latitude and longitude points. Lastly, this analysis focused on the boroughs near downtown Toronto where HealthZone is currently located. Therefore, the central, west, east and downtown Toronto neighborhoods were considered in the analysis. Combined the four boroughs have 39 neighborhoods that were part of the analysis.

For New York City, a dataset of the boroughs, neighborhoods, latitudes and longitudes was sourced as a GeoJSON file from NYU (https://geo.nyu.edu/catalog/nyu_2451_34572). The dataset included all five boroughs and 306 neighborhoods.

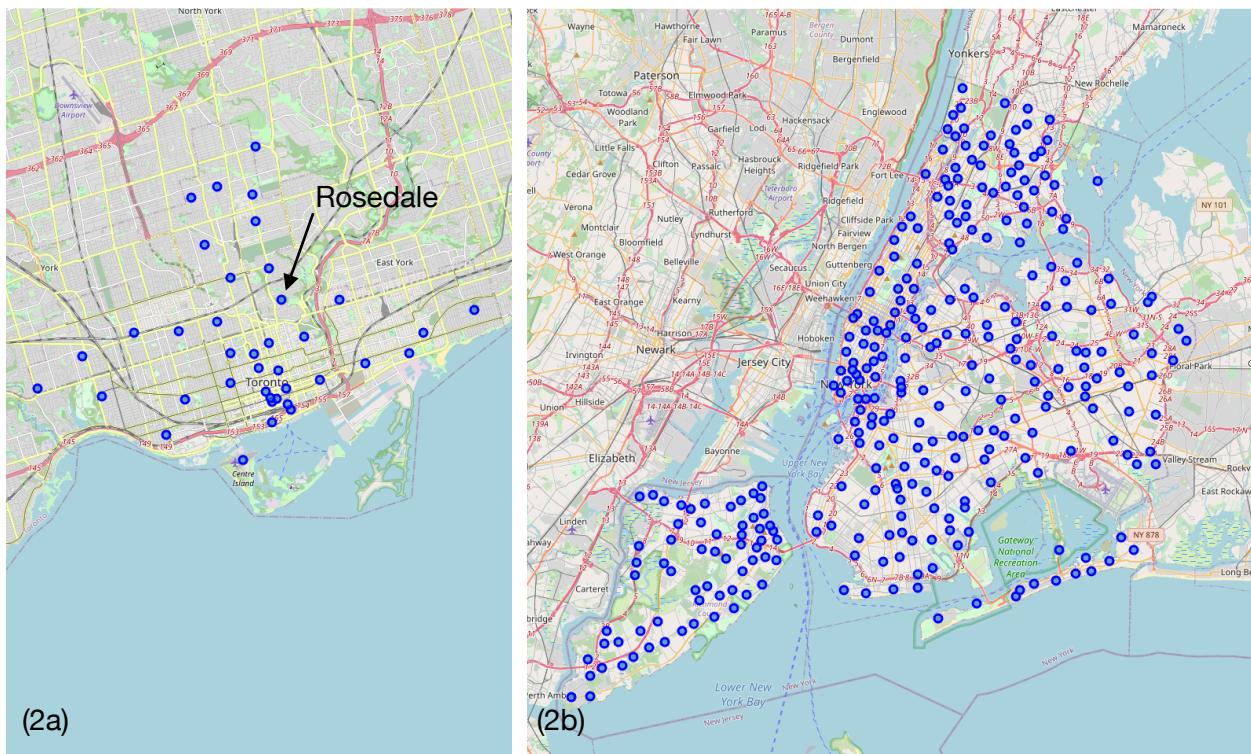


Figure 2: Map of neighborhoods in (a) downtown Toronto and (b) New York City considered in analysis

Lastly, the Foursquare API (<https://developer.foursquare.com/>) was used to source data relating to the venues in each neighborhood. The inputs into the API are the neighborhood's latitude and longitude. The Foursquare API most importantly returns the venue category (along with venue name, latitude and longitude) for each venue within a set radius from the input location. For this analysis, a radius of 500 meters (~0.3 miles) was used. For the combined 345 neighborhoods between Toronto and NYC, the Foursquare API returned 11,715 venues which were used in the analysis (an average of ~34 venues per neighborhood).

3. Methodology

After collecting the venue category for 11,715 venues across the 345 neighborhoods of downtown Toronto and NYC, each neighborhood was analyzed to determine the highest frequency venue types in those neighborhoods. This was done by simply calculating the frequency of venue type in the neighborhood. For instance, if a neighborhood had 20 venues total and five were of type 'coffee shop', then the frequency of 'coffee shop' would be 0.25. Finding the most populated venue types/categories in each neighborhood gives a description of the characteristics of the neighborhood.

Once the frequency of venue categories were determined, machine learning was applied to the dataset to find similar neighborhood clusters between Toronto and NYC. The type of machine learning algorithm used was *k*-means clustering. The purpose of *k*-means clustering is to partition the entire dataset (neighborhoods in this case) into *k* groups where within a group there is similarity and outside a group there is dissimilarity. No datapoint (i.e. neighborhood) can exist in more than one cluster using this method. Therefore each neighborhood is exclusive to a single cluster type which contains other similar neighborhoods. This machine learning algorithm was chosen because it does not have any datapoint overlapping into multiple clusters and because it can be used relatively efficiently on medium to large databases.

Exploratory data analysis in the manner described above was done first on just downtown Toronto to understand how the Rosedale neighborhood was similar and dissimilar within its own city. This exploratory analysis showed that Rosedale is unique within its own city and is only very similar to the adjacent neighborhood of Moore Park. Both neighborhoods are characterized by parks, playgrounds and trails. Compared to neighborhoods in near downtown, the Rosedale neighborhood does not have the plethora of coffee shops, restaurants and bars to compete with. This will be an important point for the second location in NYC. It will be important to find another neighborhood of people with healthy values and lower competition size.

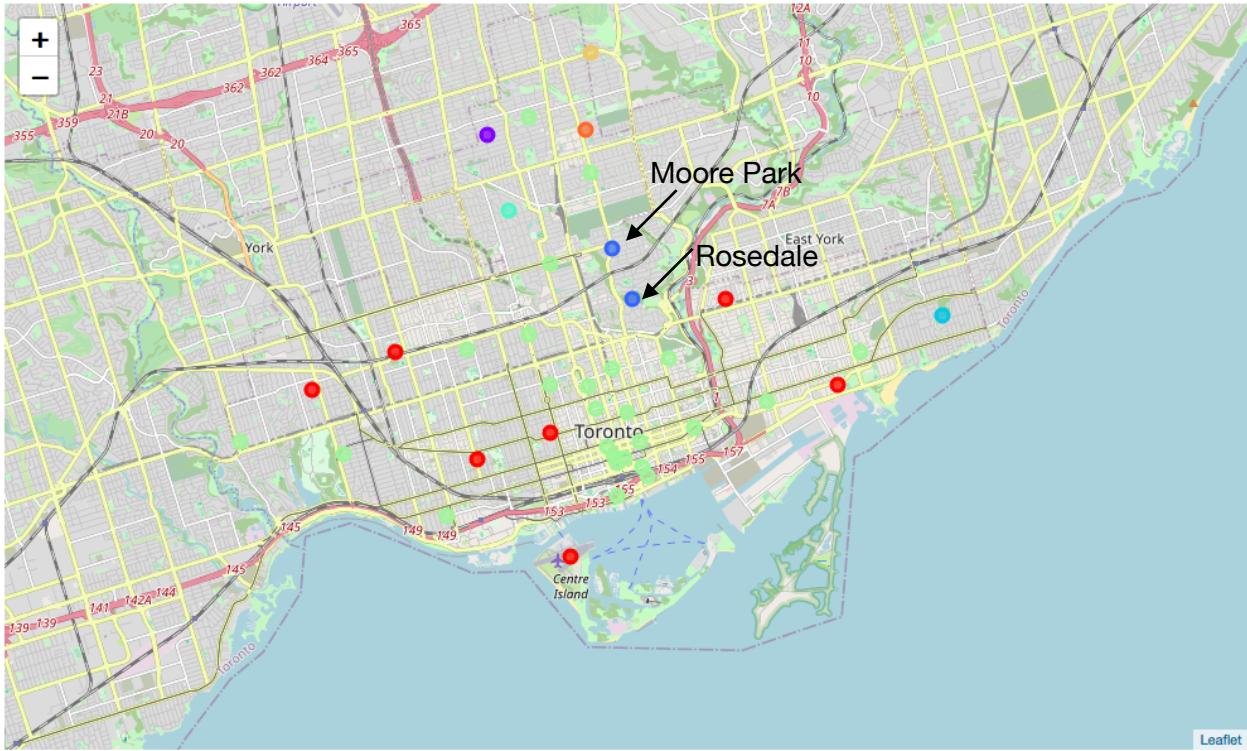


Figure 3: Map showing exploratory clustering analysis of Toronto using k -means clustering ($k = 8$). Analysis showed Rosedale (blue dot) is unique in parks and trails while having much lower coffee shop, restaurant and bar competition like many neighborhoods nearer downtown (green dots).

4. Results

The 11,715 venues across the 345 neighborhoods of downtown Toronto and NYC were analyzed using k -means clustering to produce eight clusters which show similar neighborhoods within any cluster.

Two clusters were found to contain both neighborhoods in Toronto and in NYC. The other six clusters contained neighborhoods in NYC only. The Rosedale neighborhood in Toronto was found to be in a cluster with two other neighborhoods in Toronto (Moore Park and Lawrence Park) and six other neighborhoods in NYC. This cluster and the three most common venue types in the neighborhoods are shown in Table 1.

Table 1 shows that this cluster heavily features parks with 8/9 neighborhoods having this venue as the most common. It also shows a pattern of similarity with trails, swimming pools and yoga studios. These are all venues where people can carry out healthy activities.

Borough	Neighborhood	1st Most Common Venue Type	2nd Most Common Venue Type	3rd Most Common Venue Type	
a	Downtown Toronto	Rosedale	Park	Trail	Playground
b	Central Toronto	Lawrence Park	Park	Swim School	Bus Line
c	Central Toronto	Moore Park	Park	Lawyer	Trail
d	Bronx	Clason Point	Park	Pool	Grocery Store
e	Queens	South Ozone Park	Park	Food Truck	Deli/Bodega
f	Queens	Somerville	Park	Yoga Studio	Egyptian Restaurant
g	Staten Island	Todt Hill	Park	Trail	Yoga Studio
h	Staten Island	Chelsea	Park	Steakhouse	Spanish Restaurant
i	Staten Island	Randall Manor	Home Service	Park	Bus Stop

Table 1: Cluster that contains the Rosedale neighborhood of Toronto.
This cluster is made up of three neighborhoods in Toronto and six in New York City.

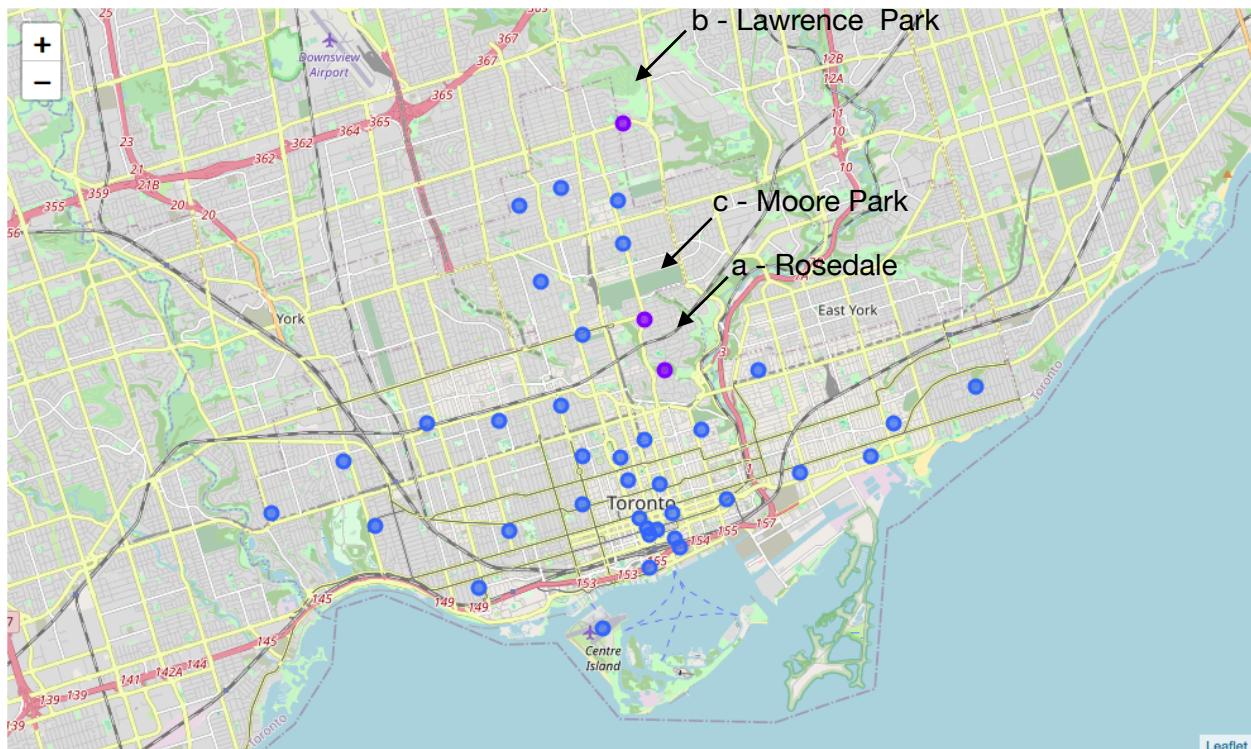


Figure 4: Map of downtown Toronto with neighborhoods colored according to cluster.
Rosedale is included in the small purple cluster which is labeled according to Table 1.

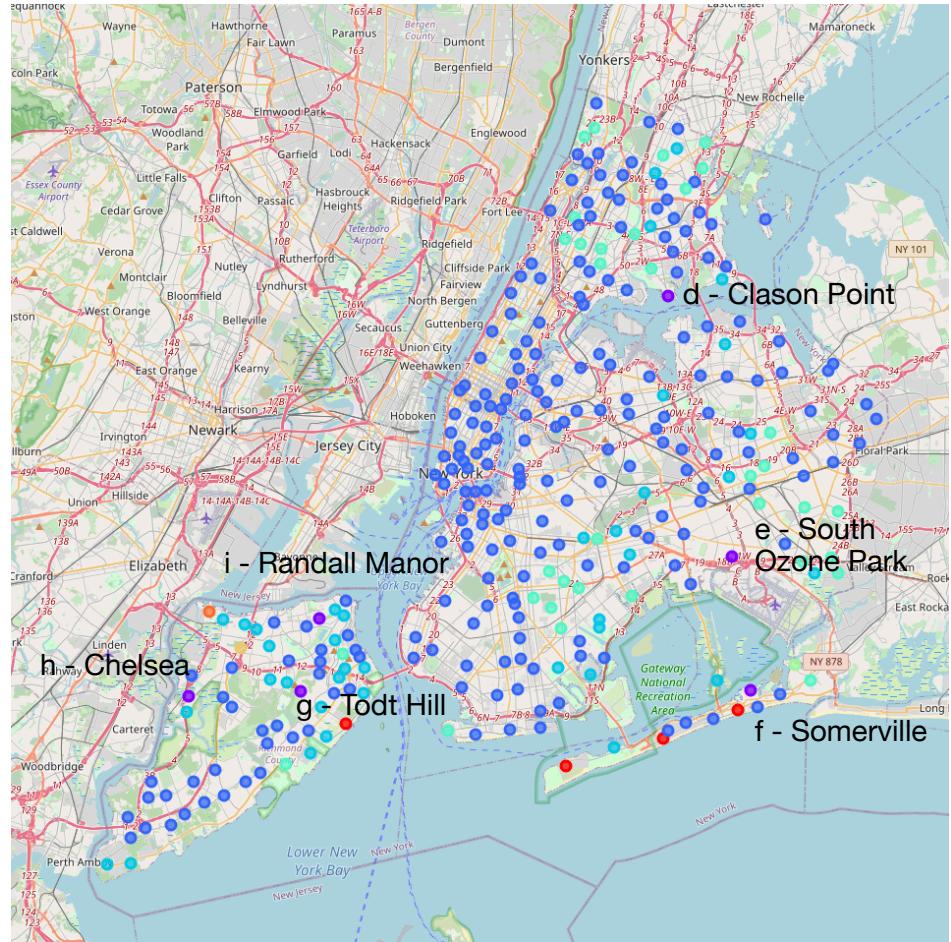


Figure 5: Map of New York City with neighborhoods colored according to cluster. Rosedale is included in the small purple cluster which is labeled according to Table 1.

5. Discussion

The results show that the 306 neighborhoods in New York City have been narrowed down substantially to just 6 neighborhoods that are similar in venue characteristics to the Rosedale neighborhood where HealthZone's first successful business resides. This is a key discovery that will make the next steps of finalizing the second location much easier for HealthZone's owners.

To recap, the six similar neighborhoods discovered in this analysis as potential choices for the second location are shown in Table 2.

Borough	Neighborhood	1st Most Common Venue Type	2nd Most Common Venue Type	3rd Most Common Venue Type
d	Bronx	Clason Point	Park	Pool
e	Queens	South Ozone Park	Park	Food Truck
f	Queens	Somerville	Park	Yoga Studio
g	Staten Island	Todt Hill	Park	Trail
h	Staten Island	Chelsea	Park	Steakhouse
i	Staten Island	Randall Manor	Home Service	Park
				Bus Stop

Table 2: Six neighborhoods recommended for HealthZone's NYC location based on k -means clustering

Additionally, we observe that the South Ozone Park location is very close the JFK airport and large highways, which may not be as desirable compared to the others.

Clason Point is completely isolated geographically from the other similar points in this cluster. It is the farthest north location and the only one in the Bronx. It could provide a unique atmosphere and competitive advantage since it would stand apart from anything else close by to people in this neighborhood.

Lastly, three of the neighborhoods are located on Staten Island. Staten Island shows many parks and green areas that should attract the type of customer HealthZone is aiming to have. For this reason, there may be more potential customers for HealthZone located on Staten Island.

6. Conclusion

The DataBeGood data analytics firm collected location and venue information from both Toronto and New York City to determine similar neighborhoods between the two cities and help narrow the choice of a second location spot.

After analyzing 345 neighborhoods and 11,715 venues with a machine learning algorithm called k -means clustering, the analysis found six neighborhoods in New York City that are similar to the Rosedale neighborhood in Toronto (see Table 2).

Three neighborhoods are located on Staten Island which has ample parks, trails, yoga and healthy activities. This looks to offer the type of customer that HealthZone desires.

From here, it is recommended that the final choice of the second location is found by examining the real estate markets and foot traffic in the six neighborhoods of Table 2.

The analysis presented here filtered 306 neighborhoods down to 6, which makes a final research project on the local real estate and foot traffic of each six neighborhoods much more feasible and quicker to accomplish.