The Battle of Neighborhoods

Introduction / Business Problem

A resident of Clairville in Etobicoke of Toronto, Canada has been offered a job opportunity on the other side of Toronto at Highland Creek in Scarborough, and is currently attempting to find out more on whether the services and amenities available in Highland Creek or in neighborhoods in that region would be comparable or better than what he is currently having in Clairville. Thus, the problem at hand would require the comparison of the top 10 venues of each neighborhood based on the frequency of each venue in each neighborhood. Given that the whole research was undertaken to address the preferences on venues and services of the resident, the whole project is designed with the resident as the target audience.

Based on understanding of the resident, he has the following characteristics and interests:
- dog owner
- typically eats out with preference for Turkish cuisine
- enjoys social drinking with friends and colleagues on regular basis
- occasionally rents a car to drive out of town to have short vacations over the weekend or while on leave

Data

The capstone project would be relying primarily on the data provided during the assignment for Week 3. The main data sources would be as follows:
- the borough and neighborhood data from Wikipedia;
- the geospatial data from the CSV file; and
- the top 10 venues in each neighborhood pulled via FourSquare API to find out the the services and amenities for each neighborhood.

Thus, the fields that would be needed are as follows:
- Boroughs in Toronto extracted from Wikipedia
- Neighborhoods in Toronto extracted from Wikipedia
- Longitudes and latitudes of neighborhoods in Toronto extracted from the geospatial data
- Longitudes and latitudes of venues per neighborhood extracted via FourSquare API
- Type of each venue in each neighborhood extracted via FourSquare API

The data is sanitized by first using pandas to extract the boroughs and neighborhood data from Wikipedia into a data frame, and dropping all rows with boroughs 'Not assigned'. Then, the geospatial data is extracted from the CSV and merged together with the original data frame to form a new one containing columns for latitude and longitude for each unique postal code.

Methodology

After importing the relevant packages, an API request URL and the GET request to pull nearby venues of each neighborhood from FourSquare based on its name, latitude and longitude were both created.

After doing so, each neighborhood in Toronto was then processed via one-hot encoding (i.e. process by which categorical variables are converted into binary variables for further analysis via machine-learning subsequently) and rows were grouped by neighborhoods and by taking the mean of the frequency of occurrence of each category. A new data frame containing the top 10 most common venues for each neighborhood was then created, where an extract is shown in **Table 1**.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Latin American Restaurant	Breakfast Spot	Skating Rink	Escape Room	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Dim Sum Restaurant	Ethiopian Restaurant
1	Alderwood, Long Branch	Pizza Place	Pharmacy	Pub	Gym	Sandwich Place	Coffee Shop	Skating Rink	Dessert Shop	Dim Sum Restaurant	Diner
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Fried Chicken Joint	Frozen Yogurt Shop	Restaurant	Supermarket	Sushi Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Deli / Bodega
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Women's Store	Dim Sum Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Sandwich Place	Liquor Store	Juice Bar	Indian Restaurant	Restaurant	Sushi Restaurant	Café	Pub

Table 1: Extract of data frame containing top 10 most common venues for neighborhoods

The analysis continues with running k-means clustering of the neighborhoods of Toronto into 7 clusters via a reiterative process, where there were no significant changes in the clusters for additional clusters created. A map of the clusters can be seen in **Figure 1**.
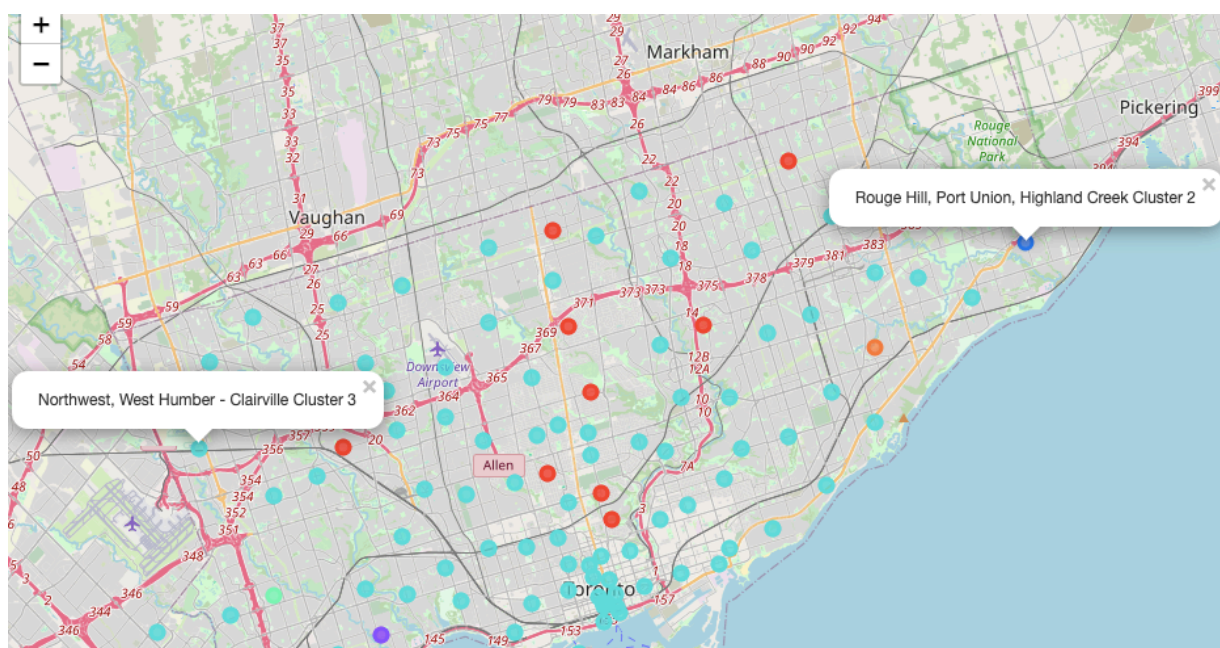

Figure 1: Map showing the clustering of the neighborhoods in Toronto

The usage of K-means clustering assists in zooming in on neighborhoods in the same cluster as Clairville, and allows analysis to be made on whether Highland Creek is similar to to Clairville in terms of the venues it has to offer, and if not, to determine what are the nearest neighborhoods to Highland Creek that provides similar venues as Clairville. This would allow a decision to be made quickly on whether the resident should take up the job opportunity knowing that his new neighborhood or nearby neighborhoods would be similar to that he has been familiar with, or to reject the job opportunity if he is not comfortable with a new living environment.

Results and Discussion

Per analysis of the results, it is noted that Clairville in Etobicoke (Cluster 3) and Highland Creek in Scarborough (Cluster 2) belongs to separate clusters. The top 10 most common venues for each neighborhood are as follows in **Table 2** and **Table 3** respectively:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
94	Etobicoke	3	Bar	Rental Car Location	Drugstore	Truck Stop	Women's Store	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaura

Table 2: Top 10 most common venues for Clairville

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	Scarborough	2	Bar	Women's Store	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Farm

Table 3: Top 10 most common venues for Highland Creek

It is noted that Highland Creek has venues which addresses most of the needs of the resident, such as the following:

- Dog runs are the 6th most common venue, allowing the resident to send his dog outside to play and expend energy while he is out at work
- Diners are the 3rd most common venue with doner restaurants being the 7th most common venue, providing the resident with options to eat out while satisfying any cravings for Turkish cuisine if necessary.
- Bars are the most common venue, thus the resident can continue to enjoy social drinking with friends and colleagues on a regular basis

However, Highland Creek does not have rental car locations, which may make it difficult for the resident to occasionally drive out of town to have short vacations over the weekend or while on leave.

An alternative solution may be to look at nearby neighborhoods. As seen in **Figure 2**, we further note that West Hill in Scarborough is the nearest neighborhood from Highland Creek that also belongs to the same cluster as Clairville in Etobicoke (i.e. Cluster 3). It is further noted that the two neighborhoods are only approximately 3-4km apart.
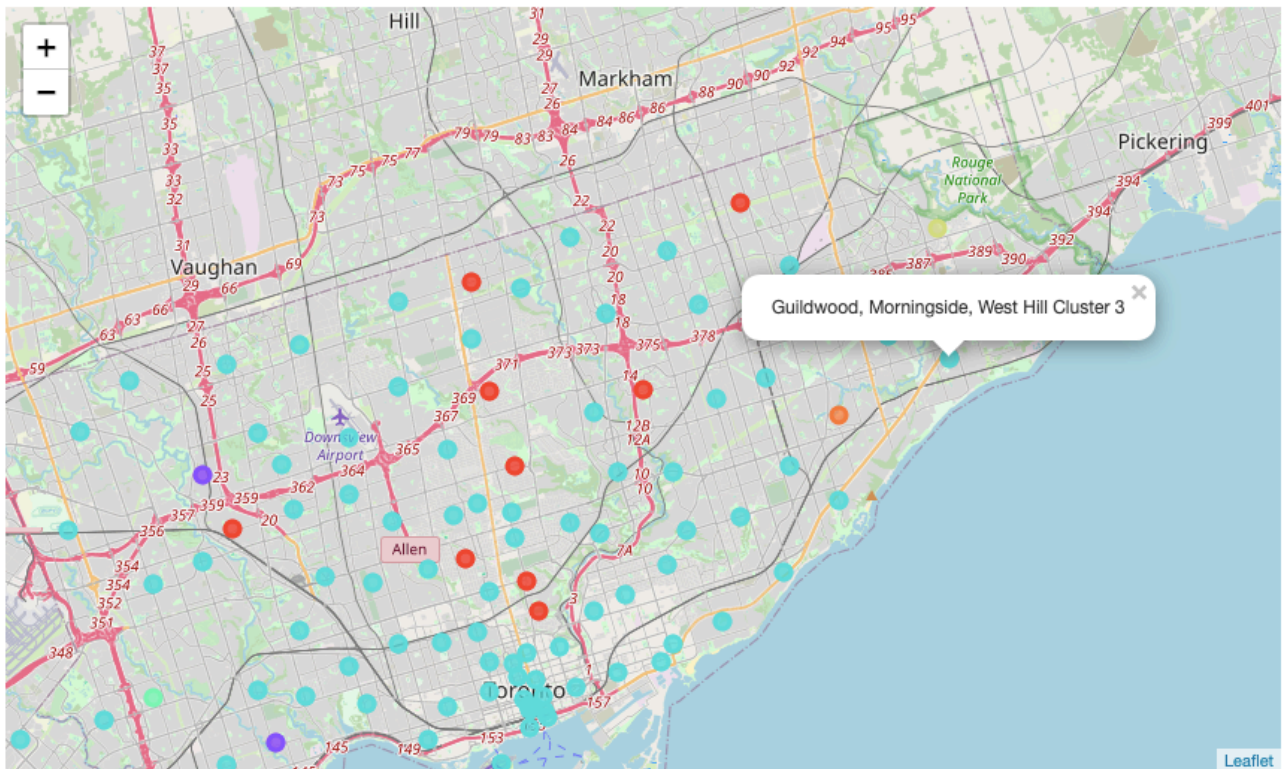
Figure 2: Location of West Hill in Scarborough

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
18	Scarborough	3	Bank	Electronics Store	Restaurant	Breakfast Spot	Medical Center	Rental Car Location	Mexican Restaurant	Women's Store	Discount Store	Distribution Center

Table 4: Top 10 most common venues for West Hill

However, through analysis of the top 10 most common venues of West Hill per **Table 4**, it was noted that while it belongs to the same cluster as Clairville, the top common venues only addresses one interest of the resident, which is for him to occasionally drive out of town to have short vacations over the weekend or while on leave through the presence of rental car locations as the 6th most common venue.

More analyses would need to be performed on whether the other venues in West Hill other than those in the top 10 are sufficiently common in addressing the needs of the resident.

Conclusion

Given the job opportunity is offered at Highland Creek in Scarborough, and that the top 10 most common venues there better satisfy the needs of the resident as compared to West Hill, it is recommended for the resident to take up the job opportunity and to move to Highland Creek.

While the resident may want to drive out of town from time to time for a short vacation, the short distance of 3-4km between Highland Creek and West Hill would mean that he can rely on the public transport between the two neighborhoods. This would only require a maximum traveling time of 30 minutes, and allows the resident to have a wide range of rental car locations at his disposal for rental of cars for the aforementioned purpose.