THE BATTLE OF NEIGHBORHOOD;

A Study of Toronto, Chicago.

OLADITI SAMUEL OLATIDE

IBM APPLIED DATA SCIECE CAPSTONE

IN COMPLETION OF

PROFESSIONAL CERTIFICATE IN DATA SCIENCE BY IBM ON COURSERA

The Battle of Neighborhoods

1. Introduction

An average man changes his location about twelve times in his lifetime. This brings us to some questions which are itemized below:

1. Why do people move from place to place until they finally find a place?
2. Do they move to settle down where they truly feel happy?
3. Does what we need change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction?
4. Do we often move to a new area without knowing the exact environment we moved into, which might result into forcing us to run at any sign of discomfort?
5. Does social life affect people of all age group?

To reduce these frequent change of location, a proper research should be done before deciding our next move in life. Considering the following factors when picking a new place to live, you won't end up wasting your valuable time and money making moves you'll end up regretting or get tired of easily. One of the first thing to consider is "Safety" which is important before moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there. Does the area have a high crime rate? Sometimes, an area may seem perfect - low taxes, moderate climate, affordable housing. However, it's important to check the prevalence of violent crimes in the area before making your final decision. Exposure to a high crime rate leaves your home and your family susceptible to theft, assault and personal harm. Crime and property value are entwined, so find out if that's why houses are cheap. You shouldn't sacrifice safety to save money.

Does the area have a high cost of living? The cost of goods and services varies from city to city and tends to be highest in the most densely-populated cities. The essential expenses to consider when relocating are: Groceries, Healthcare, Gas, Utilities, Entertainment, etc. Cost of living is the amount of money spent to maintain the average standard of day-to-day life.

How will the climate affect energy costs? You may think that leaving a high-cost city such as San Diego in favor of the more affordable Dallas is a no-brainer, but the price of homes or rent is not the only factor to consider. The moderate climate of San Diego will keep your energy costs low. Your air-conditioning bills will hit the roof when attempting to keep your home comfortable in humid Texas weather. If the climate is extreme, your cooling (or heating) costs will increase dramatically.

1.2 Problem

The statistical dataset from BeautifulSoup found on IBM-COURSERA Professional Data Science Capstone course which is a web scraping package, has information of each neighborhoods in Toronto and a Wikipedia page which has a list of postal codes. This project aims to select the safest borough based on the total crimes, exploring the neighborhoods of that borough to find the 10 most

common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

1.3 Interest

Anyone considering to relocate to Toronto will be interested to identify the best borough to live, exploring its neighborhoods and common venues around each neighborhood. This project will be using the Foursquare API to explore neighborhoods in some selected cities of Toronto. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The k-means clustering algorithm will also be used for the analysis. Finally, using the Folium library to visualize the neighborhoods in Toronto and their emerging clusters. The target audience for this project is twofold. Firstly, any planning to visit Toronto, Canada can use the plots and maps from this project to quickly select places that suit their budget and rating preferences. Secondly, a company can use this information to create a website or a mobile application, which is updated on a regular basis, to allow individuals to the city or even expand same functionality to other places.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data from two sources. The first data source of the project uses a BeautifulSoup Package that shows the Borough in Toronto.
The dataset contains the following columns:

1. Postcode
2. Borough: common names for Boroughs in Toronto
3. Neighbourhood

The second source of data is scraped from a wikipedia page that contains the list of Canada postal codes. This page contains additional information about the boroughs.
And lastly, the Foursquare API gave the project the ability to locate each borough and provide an overview of the common venues with a view on the world map.
Using the Foursquare’s explore API (which gives venues recommendations), I fetched venues from the centre of Toronto and collected their names, categories and locations (latitude and longitude).

2.2 Data Cleaning

Drop any 'None' row in the DataFrame.
Drop any row which contains a 'Not assigned' value. And all "Not assigned" will be replace to 'NaN' using numpy for convenience.

I checked that the latitude and longitude values of each corresponding venue match. After the careful analysis.

	name	categories	lat	lng
0	Downtown Toronto	[{'id': '4f2a25ac4b909258e854f55f', 'name': 'N...	43.653232	-79.385296
1	Nathan Phillips Square	[{'id': '4bf58dd8d48988d164941735', 'name': 'P...	43.652270	-79.383516
2	Indigo	[{'id': '4bf58dd8d48988d114951735', 'name': 'B...	43.653515	-79.380696
3	LUSH	[{'id': '4bf58dd8d48988d10c951735', 'name': 'C...	43.653557	-79.380400
4	Eggspectation Bell Trinity Square	[{'id': '4bf58dd8d48988d143941735', 'name': 'B...	43.653144	-79.381980

Now let check to be sure that Foursquare returned 100 venues.

They can be categorized as follows:

1. There are venues that have specific restaurants/cafes inside them as provided by Foursquare API
2. Two locations are so close that they have practically same latitude and longitude values
3. Some venues have been replaced with new venues. Venues belonging to category 1 and 3 are perfect to keep. However, the venues that belong to category 2 should be dropped.
4. After careful inspection and removal, the final dataset had a total of 49 venues with which we can work. As a final dataset, we're left with 65 venues.

	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 Mos Commo Venue
0	Adelaide,King,Richmond	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot
1	Berczy Park	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot
2	Brockton,Exhibition Place,Parkdale Village	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot
3	Business Reply Mail Processing Centre 969 Eastern	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot
5	Cabbagetown,St. James Town	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot
6	Central Bay Street	Clothing Store	Restaurant	Hotel	Plaza	Seafood Restaurant	Coffee Shop	Italian Restaurant	Burger Joint	Breakfas Spot

Using data cleaning, the dataset from the API will be used based on their venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the dataset. The final data will include the venue name, category, address, latitude, longitude and rated based on the neighborhood.

3. Methodology and Exploratory Data Analysis

This project strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two places for visit. The information will be consolidated into a single map where you can see the details of the various locations, the cluster of venues in the neighborhood and their relative locations. The pop-ups on the map items will display rent various location and the cluster of venues applicable.

The Tools:

Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. 'Geodata' was obtained by coding a program to use 'Nominatim' to get latitude and longitude of different locations.

Geopy_distance and Nominatim were used to establish relative distances. Seaborn graphic was used for general statistics on rental data.

Maps with popups labels allow quick identification of location, rating and features, thus making the selection very easy to observe by a layman.

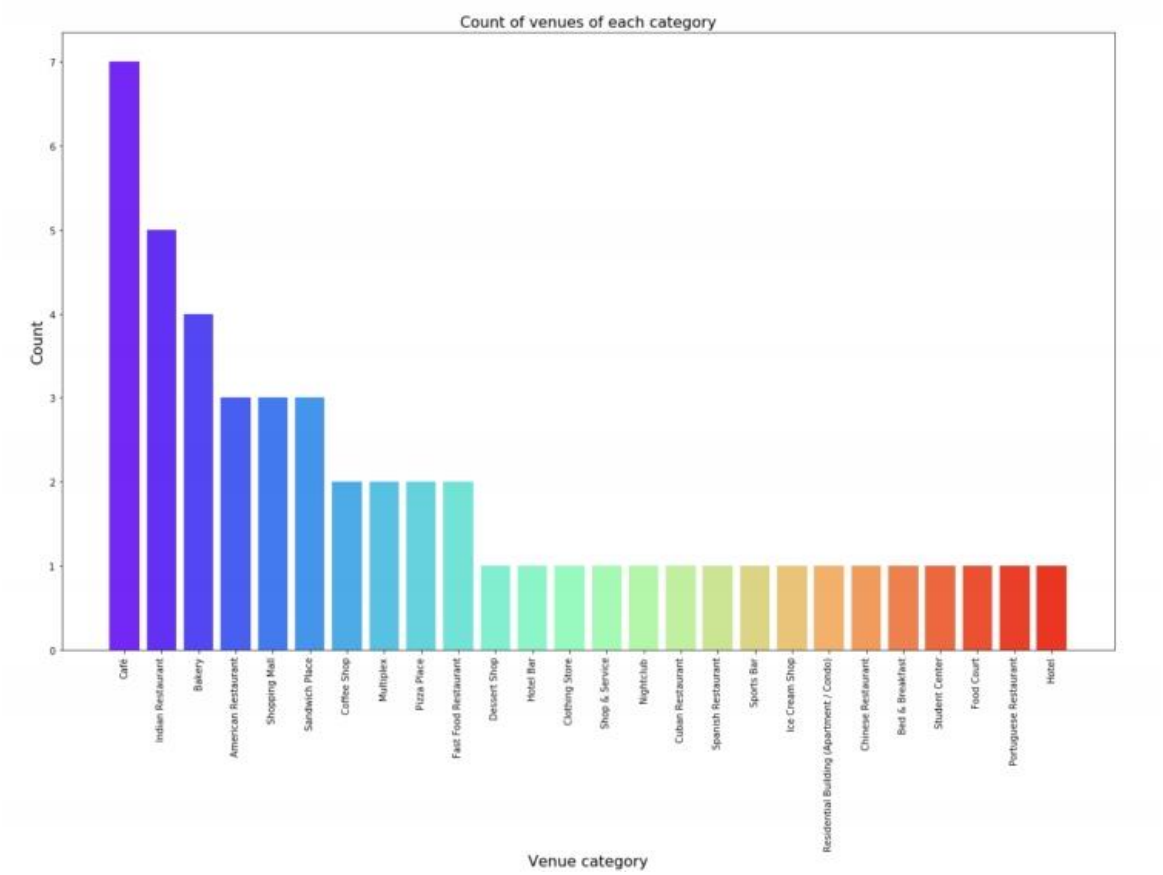
Retrieving various venues in Toronto from Foursquare API, I extract the location data from the API for all venues within the city of Toronto. Using this to fetch the venue information including the rating as common venues within a grade range of 1 to 10.

Using data cleaning, the dataset from the API and postal codes from wikipedia will be combined based on the venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the two datasets. The final data will include the venue name, category, address, latitude, longitude,

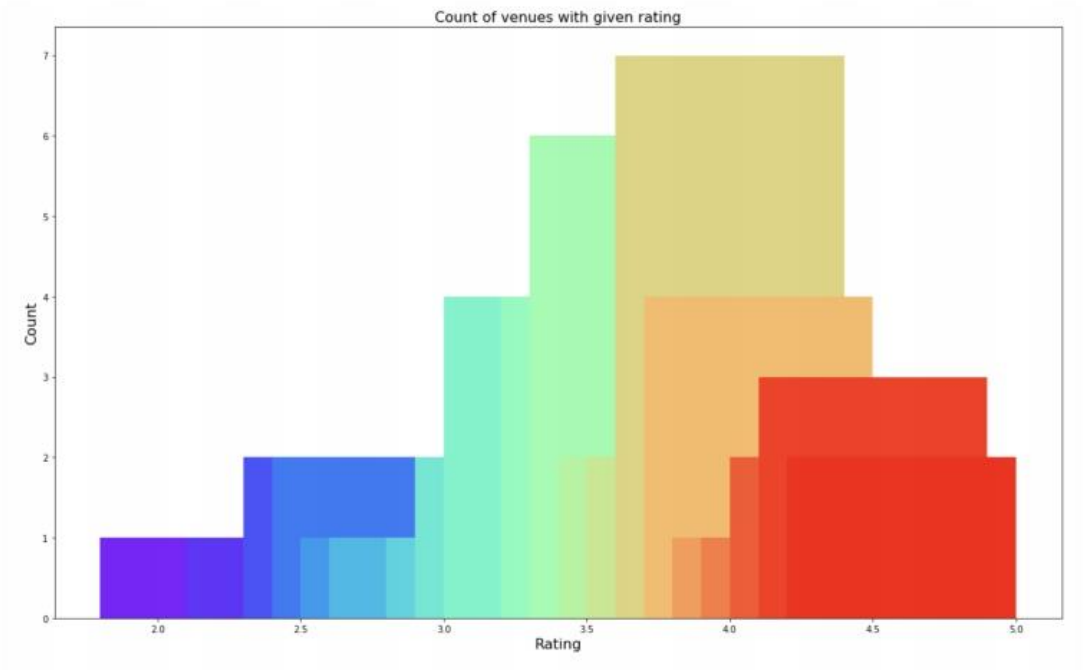
rating, price range, and average cost per person.

Using this dataset, I analyzed the top venue types that exist in Toronto. I then explore the venues on maps. This will allow us to better understand the location of various venues and the places where many venues co-exist and create place worth visiting. I'll also explore the venues based on the ratings and price range of various venues. The venues will be plot using proper color coding such that a simple glance at the map would reveal the location of the venues as well as give information about them. I aim to identify places which can be recommended to visitors based on their price and rating preferences. I'll also cluster the venues and see if we can draw meaningful information out of what kind of venues exist in Toronto.

I began my analysis by taking a critical look at the various categories of venues that exist in Toronto. As there are many restaurants, I believe that the majority venues shall include restaurants. The same thjng was done to for other venues like stores.



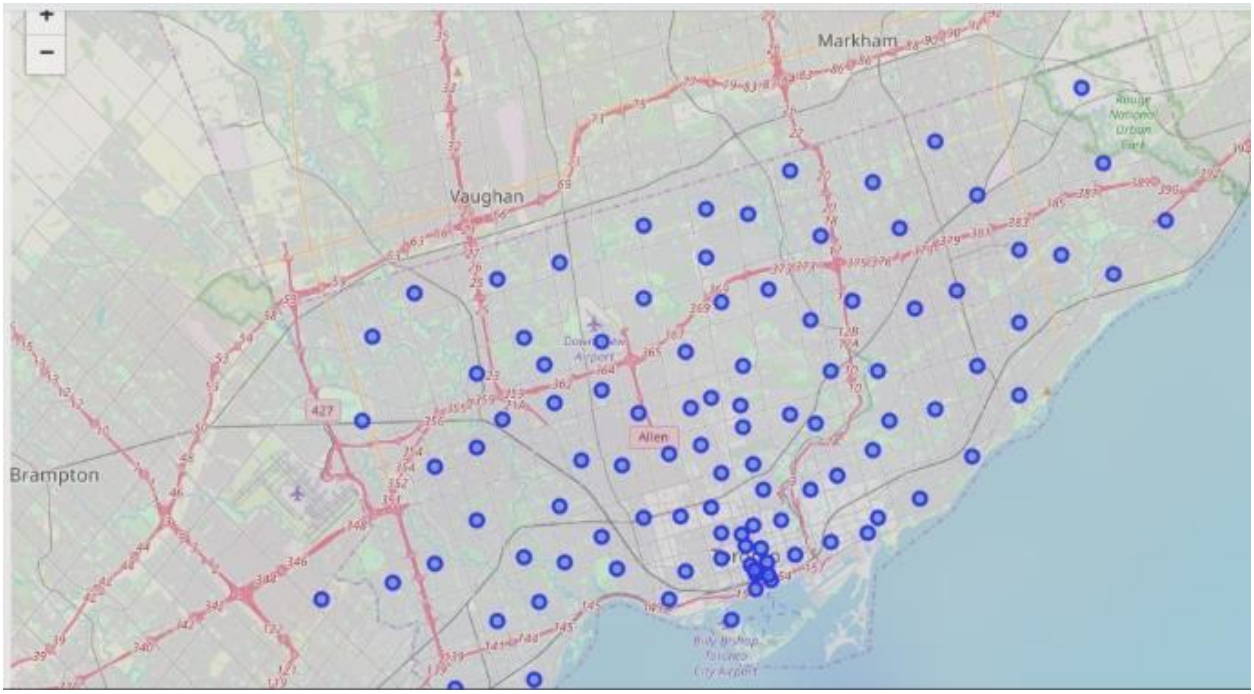
Count of various types of venues in Toronto


Count of venues with given rating
Count
Rating
0
1
2
3
4
5
6
7
2.0
2.5
3.0
3.5
4.0
4.5
5.0

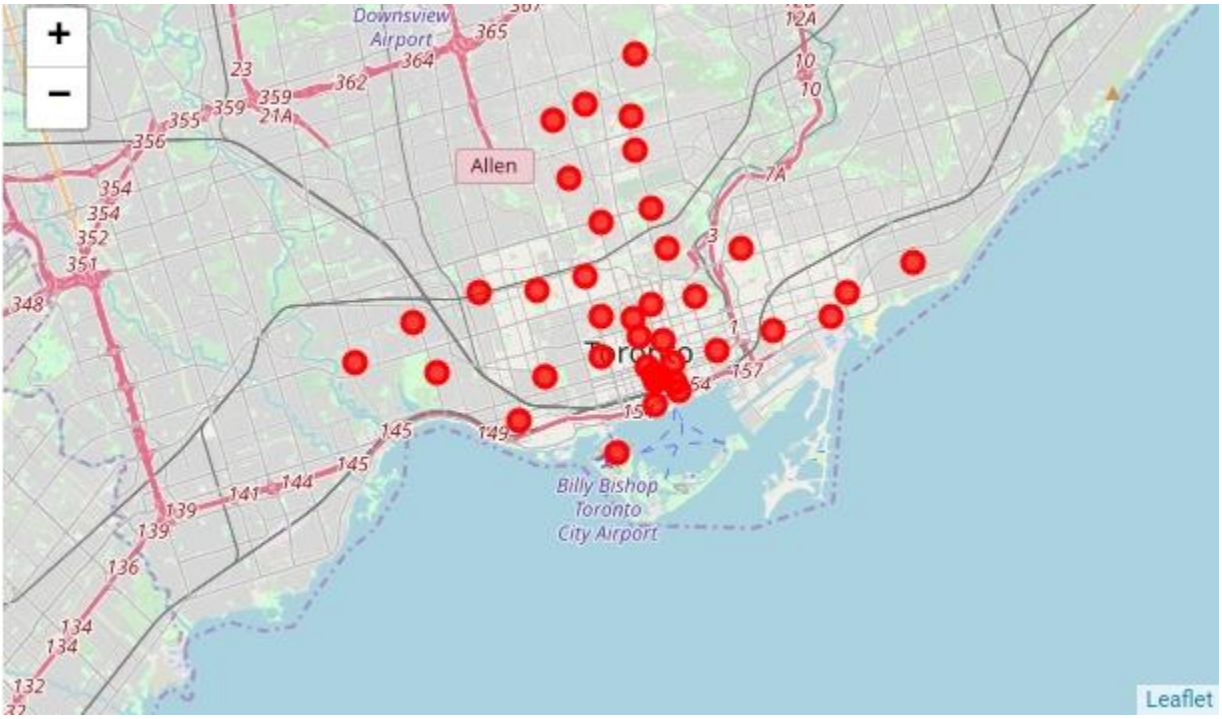
4. Clustering

Finally, all the venues are clustered based on their rating range, location and more to identify similar venues and the relationship amongst them. Using 'KMeans' clustering and decided to cluster the venues into two separate groups.

1. The first cluster (blue) is spread across the whole city and includes the majority venues. These venues have mean price range of 1.71 and rating spread around 3.57.



2. The second cluster (red) is very sparsely spread and has very limited venues.



5. Results and Discussion

After collecting data from the Foursquare API, we got a list of 163 different venues. Hence, we had to inspect their latitude and longitude values as well as their names to combine them and remove all the outliers. This resulted in a total venue count of 49.

We identified that from the total set of venues, majority of them were Cafes and Indian Restaurants. A visitor who loves Cafes/Indian Restaurants would surely benefit from coming to Toronto.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Mo Comm Venue
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Clothing Store	Restaurant	Hotel	Plaza	Seafoo Restau
1	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	0	Clothing Store	Restaurant	Hotel	Plaza	Seafoo Restau
2	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	0	Clothing Store	Restaurant	Hotel	Plaza	Seafoo Restau
3	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Clothing Store	Restaurant	Hotel	Plaza	Seafoo Restau
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0	Clothing Store	Restaurant	Hotel	Plaza	Seafoo Restau

While the ratings range from 1 to 10, majority venues have ratings close to 4. This means that most restaurants provide good quality food which is liked by the people of the city, thus indicating the high rating. When we plot these venues on the map, we discover that there are clusters of venues around The Beaches, Central Toronto and Downtown Toronto. These clusters also have very high ratings (more than 3).

When we take a look at the rating of each venue, we explore that many venues along the East Toronto are suspected to be clothing stores. Generally considering the whole datasets, we can deduce that there are various spots you could stop to eat such as the Italian Restaurant, American Restaurants, Burger joints, coffee shops, seafood Restaurants, hotels, and commercial plazas.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.57. On the other hand, there are few venues which are high priced and have average rating of 4.03.

1. The most common neighborhood in East Toronto is The Beaches with the clothing store rated higher than any other place, so if you're looking for a place to shop for your wears both male and female, you should plan to visit The Beaches.

2. If you're looking for the best places for your breakfast, with the highest rating but might also carry a high price tag, you should visit Eggspectation Bell Trinity Square ate The Beaches.

3. If you're looking to explore the city and have no specific criteria to decide upon the places you want to visit, you should try The Beaches.

6. Conclusion

A company can use this information to build an online website/mobile application, to provide users with up to date information about various venues in the city based on the search criteria (name, rating and price).

The purpose of this project was to explore the places that a person visiting Toronto could explore. The venues have been identified using Foursquare API and have been plotted on the map. The map reveals the major areas of the city. Based on the visitor's venue rating and price preferences, he/she can choose amongst the three places.