



Coursera Capstone 2021

City of Opportunity:
Toronto

May 9th, 2021

Introduction

Toronto is the capital city of the Canadian province of Ontario. It is the most populous city in Canada and the fourth most populous city in North America. Toronto is a culturally diverse and financially vibrant city and attracts many immigrants from all around the world.

Business Problem

Wendy, a Chinese artist and investor wants to open an art gallery in Toronto. However, she is not familiar with the local area developments and dynamics. Therefore, she hires a data scientist to conduct research and identify the ideal location for her art gallery. The two main questions are: 1) What are the different neighborhoods of Toronto? 2) Which neighborhood is the best location for the art gallery?

In order to answer the above two questions, the data scientist needs the following data:

1. List of the districts of Toronto

Data Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Description: We will get a list of postal codes in Toronto Canada.

2. Geo-coordinate of the districts in Toronto

Data Source: http://cocl.us/Geospatial_data

Description: From the file of toronto_coordinates.csv on this website, we will get the latitude and longitude information of Toronto.

3. Top venues of districts in Toronto

Data source: Foursquare API

Description: By using this API we will get all the venues in each district. We can group the areas into meaningful clusters.

Target Audience

The target audience of this project are business owners or entrepreneurs such as Wendy who are not familiar with Toronto around areas but want to start a new business in this big city.

Method

First of all, we downloaded the local districts and neighborhoods dataset from the Wikipedia web page, then cleaned and extracted the data into a Pandas DataFrame. Second, we created a table contains the list of districts in Toronto with the respective geo coordinates. Third, we collected venues for each district and see which venues are the most common. In this step, we used Foursquare API to collect venue data, organized into a Pandas DataFrame, and created a table of top venues in each district. Fourth, we used the K-means approach to group the Toronto neighborhoods into meaningful clusters. Finally, we visualized data with Folio.

Table 1. List of Districts in Toronto

| | | | |
|----|-----|-----------------------|--|
| 71 | M6A | North York | Lawrence Manor, Lawrence Heights |
| 72 | M6B | North York | Glencairn |
| 73 | M6C | York | Humewood-Cedarvale |
| 74 | M6E | York | Caledonia-Fairbanks |
| 75 | M6G | Downtown Toronto | Christie |
| 76 | M6H | West Toronto | Dufferin, Dovercourt Village |
| 77 | M6J | West Toronto | Little Portugal, Trinity |
| 78 | M6K | West Toronto | Brockton, Parkdale Village, Exhibition Place |
| 79 | M6L | North York | North Park, Maple Leaf Park, Upwood Park |
| 80 | M6M | York | Del Ray, Mount Dennis, Keelsdale and Silverthorn |
| 81 | M6N | York | Runnymede, The Junction North |
| 82 | M6P | West Toronto | High Park, The Junction South |
| 83 | M6R | West Toronto | Parkdale, Roncesvalles |
| 84 | M6S | West Toronto | Runnymede, Swansea |
| 85 | M7A | Queen's Park | Ontario Provincial Government |
| 86 | M7R | Mississauga | Enclave of L4W |
| 87 | M7Y | East Toronto Business | Enclave of M4L |
| 88 | M8V | Etobicoke | New Toronto, Mimico South, Humber Bay Shores |

Table 2. List of Districts in Toronto with Geo Coordinates

| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|------------|-------------|--|-----------|------------|
| 0 | M1B | Scarborough | Malvern, Rouge | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Rouge Hill, Port Union, Highland Creek | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |

Figure 1. Map of Toronto Neighborhoods

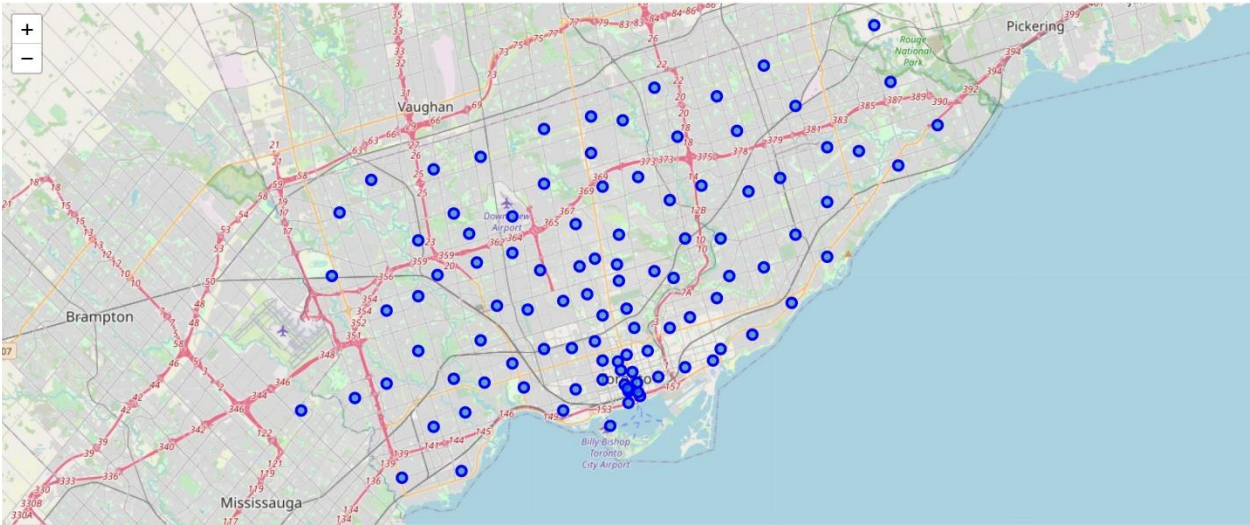


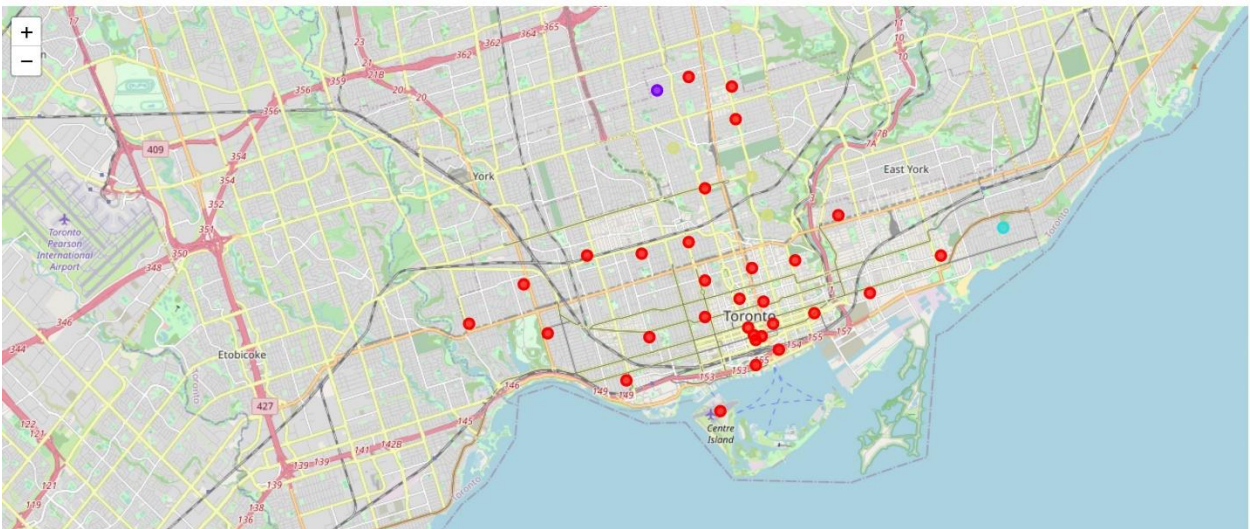
Table 3. 10 Most Occurrence Venue Types in Each Area

| | PostalCode | Borough | Neighborhoods | 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 |
|----|------------|------------------|---|-----------------------|-----------------------|-------------------------------|-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------------|------------------------|
| 27 | M5V | Downtown Toronto | CN Tower, King and Spadina, Railway Lands, Har... | Airport Service | Airport Lounge | Boutique | Harbor / Marina | Rental Car Location | Boat or Ferry | Sculpture Garden | Bar | Airport Terminal | Airport Gate |
| 31 | M6J | West Toronto | Little Portugal, Trinity | Bar | Vietnamese Restaurant | Café | Vegetarian / Vegan Restaurant | Restaurant | Coffee Shop | Men's Store | Asian Restaurant | Yoga Studio | Cuban Restaurant |
| 34 | M6R | West Toronto | Parkdale, Roncesvalles | Breakfast Spot | Gift Shop | Bookstore | Dog Run | Movie Theater | Bar | Restaurant | Dessert Shop | Eastern European Restaurant | Italian Restaurant |
| 26 | M5T | Downtown Toronto | Kensington Market, Chinatown, Grange Park | Café | Bar | Vegetarian / Vegan Restaurant | Coffee Shop | Vietnamese Restaurant | Mexican Restaurant | Gaming Cafe | Farmers Market | Comfort Food Restaurant | Grocery Store |
| 25 | M5S | Downtown Toronto | University of Toronto, Harbord | Café | Bookstore | Bar | Japanese Restaurant | Bakery | College Arts Building | Coffee Shop | Dessert Shop | Poutine Place | Pub |
| 32 | M6K | West Toronto | Brocton, Parkdale Village, Exhibition Place | Café | Coffee Shop | Breakfast Spot | Pet Store | Bakery | Performing Arts Venue | Nightclub | Climbing Gym | Restaurant | Burrito Place |

Table 4. Four Meaningful Clusters

| | PostalCode | Borough | Neighborhood | Latitude | Longitude | Cluster | 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 | |
|----|------------|------------------|---|-----------|------------|---------|-----------------------|-----------------------|-------------------------------|-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|---|
| 27 | M5V | Downtown Toronto | CN Tower, King and Spadina, Railway Lands, Har... | 43.628947 | -79.394420 | 0 | Airport Service | Airport Lounge | Boutique | Harbor / Marina | Rental Car Location | Boat or Ferry | Sculpture Garden | Bar | |
| 31 | M6J | West Toronto | Little Portugal, Trinity | 43.647927 | -79.419750 | 0 | Bar | Vietnamese Restaurant | Café | Vegetarian / Vegan Restaurant | Restaurant | Coffee Shop | Men's Store | Asian Restaurant | Y |
| 34 | M6R | West Toronto | Parkdale, Roncesvalles | 43.648960 | -79.456325 | 0 | Breakfast Spot | Gift Shop | Bookstore | Dog Run | Movie Theater | Bar | Restaurant | Dessert Shop | |
| 26 | M5T | Downtown Toronto | Kensington Market, Chinatown, Grange Park | 43.653206 | -79.400049 | 0 | Café | Bar | Vegetarian / Vegan Restaurant | Coffee Shop | Vietnamese Restaurant | Mexican Restaurant | Gaming Cafe | Farmers Market | |
| 25 | M5S | Downtown Toronto | University of Toronto, Harbord | 43.662696 | -79.400049 | 0 | Café | Bookstore | Bar | Japanese Restaurant | Bakery | College Arts Building | Coffee Shop | Dessert Shop | |
| 32 | M6K | West Toronto | Brockton, Parkdale Village, Exhibition Place | 43.636847 | -79.428191 | 0 | Café | Coffee Shop | Breakfast Spot | Pet Store | Bakery | Performing Arts Venue | Nightclub | Climbing Gym | |

Figure 2. Map of Clusters



Results

Upon observing the result, we can name the clusters as follow:

- Cluster 1 (Red): Living area. This area is with mostly park, trail, school, and some small businesses.
- Cluster 2 (Yellow): Roselawn. This area is in Central Toronto with nothing here except a garden.
- Cluster 3 (Purple): Business area. This is the busiest and centralized area with lots of business venues.
- Cluster 4 (Blue): Oldest neighborhood. This is the area with lot of nature parks.

Recommendation

Based on the result, the data scientist recommended the cluster 3 as the ideal location for Wendy to open her art gallery to attract more customers and generate maximum profit.

Conclusion

In this project, I worked as a data scientist to go through the process of identifying the business problems, specifying the data required, extracting and preparing the data, visualizing the results, performing machine learning by clustering the data into 4 clusters, generating insights, and making the recommendation to the client.