

Finding the most common venues in Toronto



Saritha Aseen
7th July 2020

This report is for the capstone project of IBM Data Science Professional Certificate. In this project, the neighborhoods of Toronto are clustered into groups of similar neighborhoods. Foursquare API was used to retrieve data on the venues of the neighborhoods and their categories.

Business problem

ABC delivery service is an international delivery business. Their primary domain is courier and food delivery. They want to explore the possibility of striking partnership with health systems, grocery stores to expand their delivery business in these fields. Recently, the key business stakeholders decided to expand the delivery base to Toronto, the financial capital of Canada. ABC believes that convenience store and pharmacy delivery may leverage the drop in revenues faced from lesser restaurant food deliveries due to the COVID pandemic. They would like to analyze if convenience stores and pharmacies are among the major venue categories that exist in Toronto. This analysis will help them confirm if convenient store and pharmacies will be the right option to choose from to start their delivery diversifying project.

Data description

The Foursquare API is used to fetch and explore neighborhoods in Toronto. Then, I use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. Then, I use the *k*-means clustering algorithm to complete this task.

Finally, I use the Folium library to visualize the neighborhoods in Toronto and their emerging clusters.

The information I am focussing on are restaurants, convenience stores and pharmacies. We have neighborhood details about Downtown Toronto. We need to apply Neighborhood Segmentation and Clustering to analyze the neighborhood data and prioritize restaurants, convenience stores and pharmacies located in Toronto. Lastly, we can decide if Toronto is a good option to start with convenience store and drug store delivery.

Data Acquisition and Preparation

Firstly, we will take the processed Toronto data from week 3 but only include Downtown Toronto for this project, then explore the city by using Foursquare API and visualize the results separately.

Since clustering will be based on the categories of venues in the neighborhoods, we need data that specifies the venues in the neighborhoods and their categories. Of course we, in the first place, need a list of the neighborhoods of Toronto.

The figure below shows a map of Toronto with its neighborhoods represented as circles.



To acquire data on venues and their categories, Foursquare API is used. Foursquare is one of the world largest sources of location and venue data. To retrieve the venues and their categories in a given neighborhood, the coordinates—the latitude and the longitude—of the neighborhood are sent in the API request.

The API-request URL looks like the following:

```
https://api.foursquare.com/v2/venues/search?
&client_id=1234&client_secret=1234&v=20180605&
ll=40.89470517661,-73.84720052054902&radius=500&limit=100
```

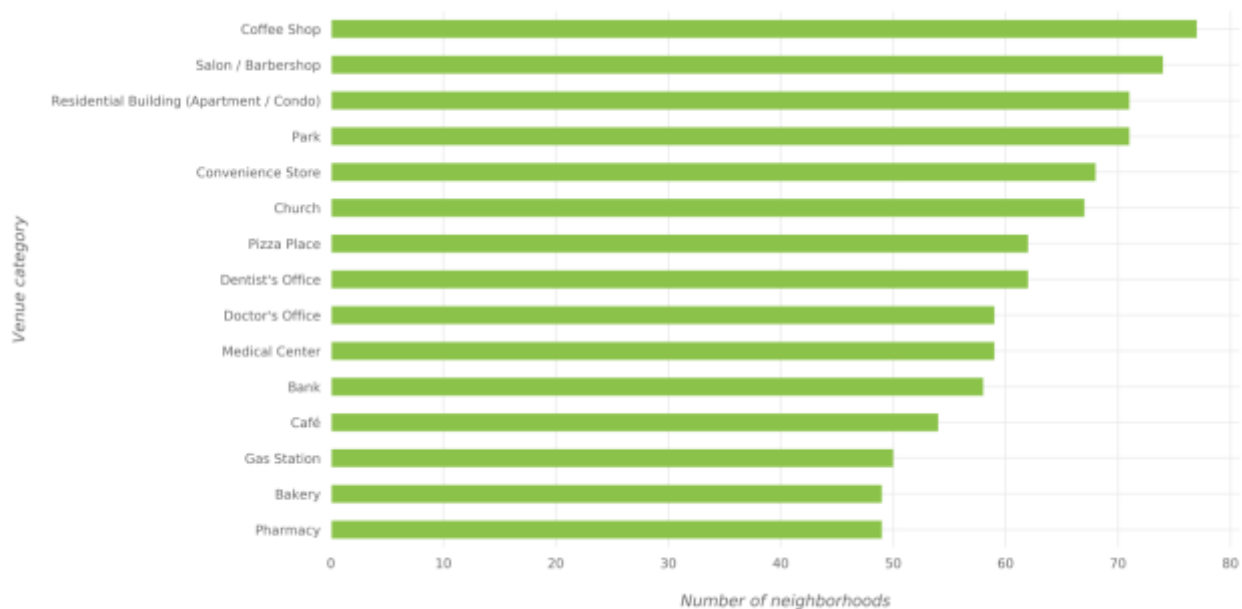
Client_id and client_secret are credentials used to access the API service and are obtained when registering a Foursquare developer account, v indicates the API version to use, ll indicates the latitude and longitude of the desired location, radius is the maximum distance in meters between the specified location and the retrieved venues, and limit is used to limit the number of returned results if necessary.

Exploratory Data Analysis

To get a better understanding of the venues data, we performed some exploratory analysis. In this analysis, we found the most common venues in Toronto. Before proceeding with exploratory

data analysis and the subsequent steps, a data-preparation operation was performed: venues whose category is “Building”, “Office”, “Bus Line”, “Bus Station”, “Bus Stop”, or “Road” were excluded because they are not expected to add analytical value in this project.

The figure below shows the most common venue categories in Toronto.



Clustering of Neighborhoods

To perform clustering, we need to feed the clustering algorithm with features in appropriate format. The data format that we

saw above—show also below—is not suitable for the clustering algorithm.

To perform clustering, we need to feed the clustering algorithm with features in appropriate format. The data format shown below is not suitable for the clustering algorithm. Hence one-hot encoding will be performed on the “Venue Category” field in the dataframe (table) shown below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station
1	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
2	Wakefield	40.894705	-73.847201	Pitman Deli	40.894149	-73.845748	Food
3	Wakefield	40.894705	-73.847201	Julio C Barber Shop 2	40.894165	-73.845748	Salon / Barbershop
4	Wakefield	40.894705	-73.847201	Public School 87	40.895331	-73.845918	School

The next step is aggregating the values for each neighborhood so that each neighborhood becomes represented by only one row. The aggregation will be done by grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

	Neighborhood	ATM	Accessories Store	Acupuncturist	Adult Boutique	Advertising Agency	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	...	Watch Shop	Water Park	Wine Bar
1	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0
2	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0
3	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0
4	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0
5	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0

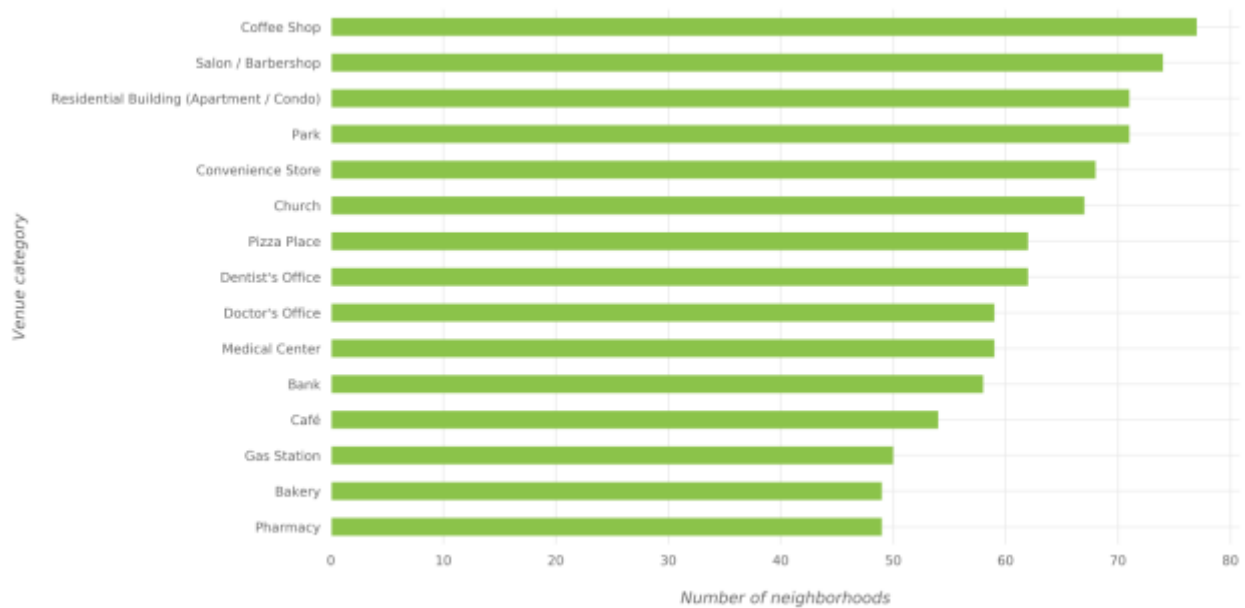
The clustering algorithm used is the K-means algorithm of Scikit-learn package and number of clusters is chosen to be 5 clusters. The output of clustering is a label for each neighborhood indicating to which cluster this neighborhood

belongs. The figure below shows a sample of a dataframe created with the cluster labels.

	Neighborhood_	ATM	Accessories Store	Acupuncturist	Adult Boutique	Advertising Agency	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	...	Watch Shop	Water Park	Wine Bar	Wi Sh
0	Agincourt	0.0	0.0	0.0000	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
1	Alderwood, Long Branch	0.0	0.0	0.0000	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0000	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
3	Bayview Village	0.0	0.0	0.0125	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0000	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0

I used the Foursquare API to explore the venues in neighborhoods with setting the limit as 100 venues and the radius 500 meter based on latitude and longitude of each Borough. I sorted the top 10 venues for each Borough by using KMeans- Clustering. The value of kcluster was set as 5.

Most common venues in Toronto



Recommendation & Conclusion

As we can see from clustering results:

Downtown Toronto has convenience stores and pharmacies among its most common venues. As a result, it could attract more delivery services apart from restaurants. This analysis will help lower the risk for ABC delivery services expanding their delivery service to include deliveries from convenience stores and pharmacies in Canada starting from downtown Toronto.