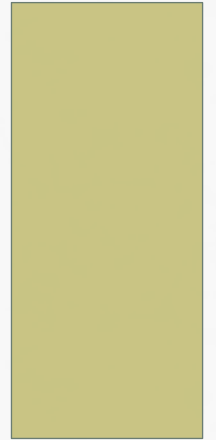# FINDING THE MOST COMMON VENUES IN TORONTO

FINDING THE MOST COMMON VENUES IN TORONTO

- 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 focusing 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 CLEANING

To explore, cluster the neighborhoods in Toronto and to scrape data, the following Wikipedia page was used:

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Neighbourhoods with the same borough were merged. If a cell has a borough but a **Not assigned** neighbourhood, then the neighbourhood will be the same as the borough. BeautifulSoup was used to parse html data.

DATA ACQUISITION AND PREPARATION

- We explore the city by using Foursquare API and visualize the results. 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.

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.

CLUSTERING OF NEIGHBORHOODS

To perform clustering, we need to feed the clustering algorithm with features in appropriate format. Onehot encoding was performed on the "Venue Category" of the data frame.

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

The clustering algorithm used is the K-means algorithm of the Scikit-learn package.

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
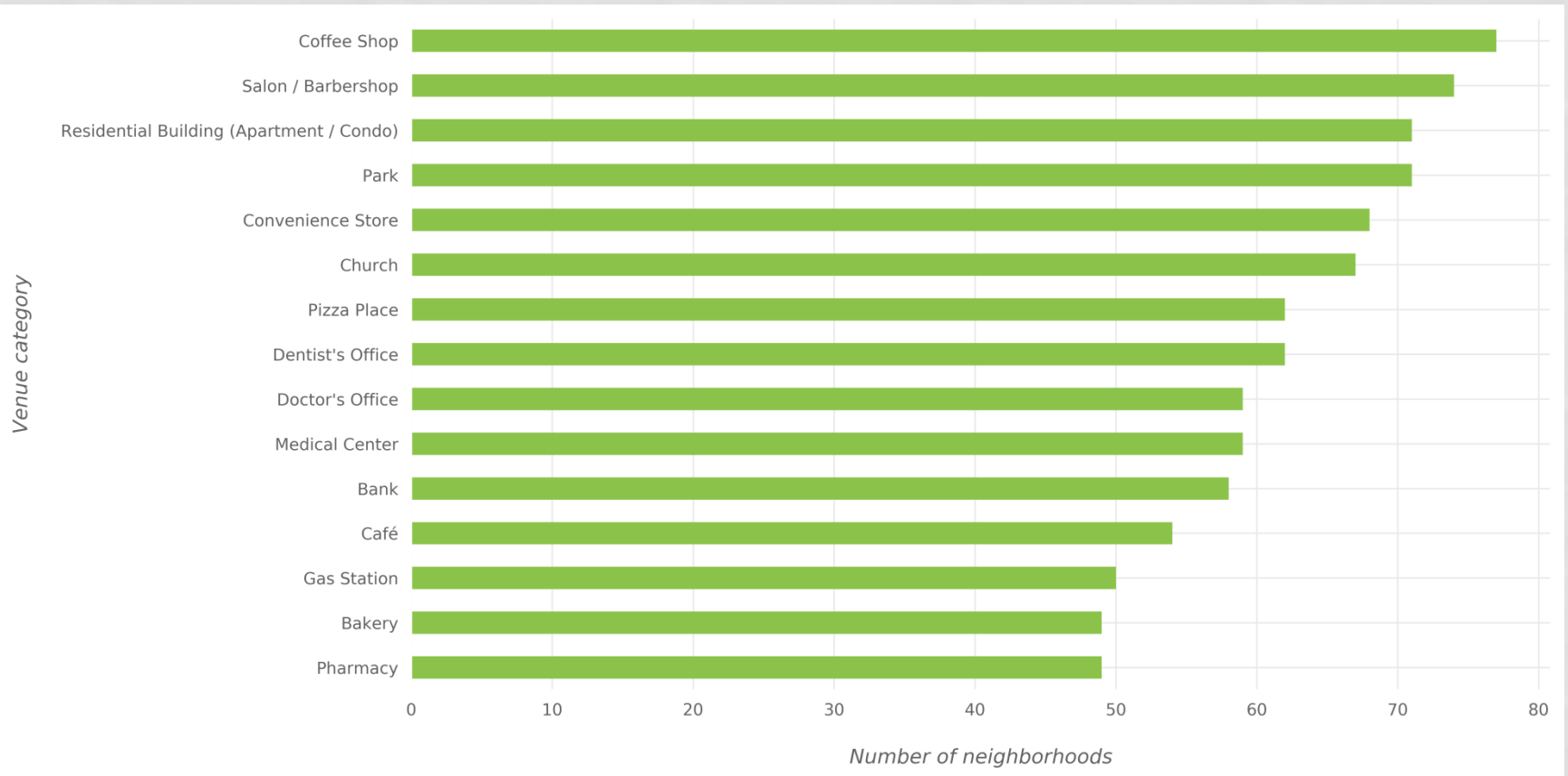
RECOMMENDATION & CONCLUSION

- 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