

PROJECT REPORT

Applied Data Science Capstone Project – Battle of the Neighborhoods

Introduction: Business Problem

This project will analyze neighborhoods between Toronto and Montreal in Canada. A multi-national company is planning to set up shop in Canada. They have narrowed their shortlist to two of the biggest cities on the Canadian East coast. The company wants insight into the neighborhoods and local economy in each of the cities, so that its employees may have the optimum living standards and which is commercially viable as well. This project will explore the similarities and dissimilarities between certain neighborhoods in the two cities, and determine which neighborhood best fits the MNC's cultural values and vision.

Data Section

There will be primarily two sets of data to work with. The first set will consist of Montreal's different neighborhoods - [https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: H](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H)

The second set will have Toronto's neighborhoods:

[https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

The geometric coordinates will be retrieved from a CSV file downloaded for –

Montreal:



PostalCodes_Canada.csv

<https://www.serviceobjects.com/blog/free-zip-code-and-postal-code-database-with-geocoordinates/>

Toronto:



Geospatial_Coordinates (1).csv

https://cocl.us/Geospatial_data

The data is in CSV format. Data cleaning for the Montreal coordinates file will be done in the CSV itself, where we don't need the sub-postal codes, we will have just the borough's codes for the first location and delete the rest. For example, there are multiple coordinates under H1A XXX. We will just keep the first row's coordinates.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood.

The data retrieved from Foursquare will contain information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g., the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

In addition to this, various python packages and machine learning functionalities will be used to provide insights on these neighborhoods for comparison.