IBM Data Science on Coursera: Capstone
The Battle of Neighborhoods

1 INTRODUCTION

Finding places to visit when we are in a new place is something that could take too long, by indecisions or a lot of information to get in. Because of that, Big Data and Data Science can help us to make decision based on a data set of information about these locations. As an example, data can be collected from websites that allow users to publish their reviews and opinions about the experience found in this place. Then, it is possible, with Data Science to automate the ‘decision-making’ with powerful algorithms which take just few seconds to show possible places to go based in data collected.

2 PROBLEM SITUATION

The algorithm has to compare different places in Toronto based in Data Collected on the internet using APIs and another database to help people choosing the best place to go. It includes parks, gym, restaurants, movie theatres etc. To do that, anyone who has not been in Toronto before could use it to provide the best decision based in a less time possible.

3 DATA

It will be used data based on internet resources such as Wiki, websites. Also, API settings will be used to return Data based in postal codes and GEO location.

3.1 REQUIRED DATA

A list with postal codes in Toronto, a dataset that contains different types of venues located in Toronto with their respective postal code and GEO Location. Therefore, data should contain Postal Code, Borough, Neighbourhood, Latitude,

Longitude. Based in the avenue, a list of places required will be returned by the API used in the algorithm.

3.2 DATA SOURCE

To apply these skills, a List of Postal Codes of Canada: Toronto used is available at Wikipedia. Foursquare API can provide information and Data about places inside avenue and next from the avenues and localization. Those data is returned as JSON and will be handled in Data Science Python.

3.3 DATA WRANGLING

Using Python Data Science, it is possible to make decisions based in an algorithm that uses clusters and k-means to show the best result needed. For this process, all the learning skills gotten around the whole specialization will be used, therefore the principal is: Python Data Science Analysis and Machine Learning.