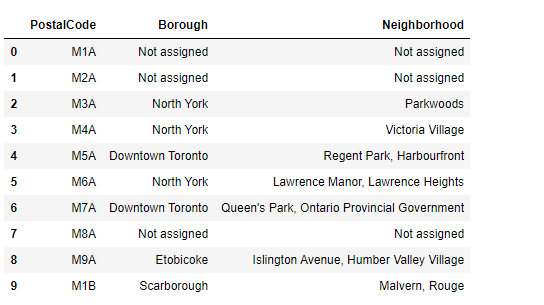
**COURSERA CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOODS**

**INTRODUCTION**

The aim of this project is to select an ideal location for opening a shopping mall in the city of Toronto, Canada. This report is targeted towards stakeholders who wants to invest in a shopping mall in the city. Using Data Science methodologies, we will find the perfect location where there are very few malls in the vicinity, to avoid competitions. We will use Clustering methodology to cluster the neighborhoods to find a location that is highly optimal for opening the mall.

**DATA DESCRIPTION**

The Geographical data to analyze the data is scrapped from the internet as the structured format of the data was not available. It is taken from the Wikipedia page using BeautifulSoup Package in Python. It consists of columns like Postal Code, Boroughs and Neighborhood which would be ideal for our analysis.

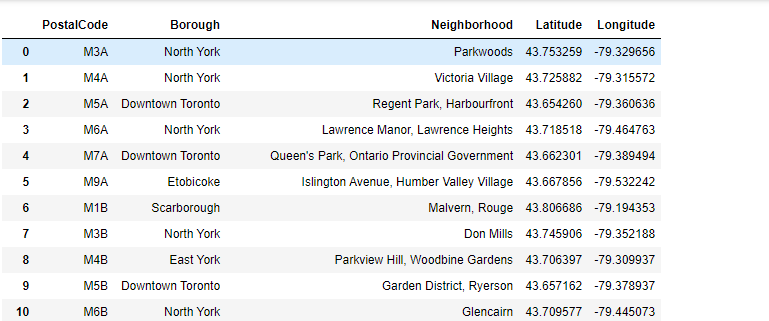


*Fig 1. Scrapped data from Wikipedia*

Further, the latitude and longitude information is added to get a clear picture of the location of neighborhoods and the venues in the neighborhoods are got through Foursquare API.

**DATA CLEANING AND TRANSFORMATION**

There were quite a few places in which the boroughs were unassigned, those rows were dropped as there are no useful information that can be obtained. The Neighborhoods which were unassigned were changed to the respective borough names. The data was joined with the latitude and longitude data and a sample of the cleaned data frames is show in figure 2.



*Fig 2. Data after cleaning*

After the data is cleaned, the venues in the neighborhoods is got through the Foursquare API for analysis. We then filter the data to get only the count of the shopping malls in the respective neighborhoods. Once this is done, we will have Neighborhood and the list of venues in them. We further transform the data such that each neighborhood has the mean of count of all of its venues to make it easier for analysis.



*Fig 3. Neighborhoods and Shopping malls(mean)*