PROJECT: THE BATTLE OF NEIGHBORHOODS
COURSE: APPLIED DATA SCIENCE
PROGRAM: IBM DATA SCIENCE PROFESSIONAL

Living it up in Hollywood North

SEGMENTING AND CLUSTERING NEIGHBORHOODS IN TORONTO

<Extra Sections Removed for Ease of Review>

II. Data

In this notebook, we will first scrape data from a Wikipedia page (List of postal codes of Canada: M) to get a list of Canadian neighbourhoods. Next we will pull in data from Foursquare's API to obtain further perspective about the neighborhoods in Toronto. We will continue to layer more data by pulling in data from the city of Toronto's Open Data Portal. This is a free online data collection that is regularly maintained and updated. From this collection we will use the following sources;

- Apartment Building Evaluation
    - compliance records for maintenance standards for residential buildings that are three or more storeys and ten or more units
- Apartment Building Registration
    - details about all residential buildings that are three or more storeys and ten or more units
- Ward Profiles, 2014-2018 Wards
    - census information for each Ward in the City of Toronto
    - demographic, social and economic information
- Neighbourhoods
    - boundaries of City of Toronto Neighbourhoods
- City Wards
    - boundaries for the City of Toronto's 47 wards
- Bikeways
    - Toronto bicycle lanes, signed bicycle routes and pathways
- Pedestrian Network
    - Toronto walkability measures, routes and pathways

We will wrangle the data, clean it, and then read it into a pandas dataframe, so that it is in a structured format. Then we will use the explore function to get the most common venue categories in each neighborhood. Finally, we will use this feature to group the neighborhoods into clusters.