DATA SCIENCE CAPSTONE PROJECT: BATTLE OF NEIGHBORHOODS IN TORONTO

Business Problem:

Toronto is the heart of Ontario, Canada with a population of 2.93 Million people. It is the most densely populated and urban place of Canada, and known best for the CN tower, Scotiabank Arena, Toronto Islands, and much more. Aside from being the most popular cities to visit in Canada, it is also a crucial business hub, contains many universities, attractions, and residential areas.

Due to that reason, the city continuously expands with new venues creating a very competitive atmosphere for entrepreneurs looking to open a location based business. The location of the venue tremendously impacts the strength of the business.

This project aims to explore patterns of multiple Neighborhoods situated in Toronto by categorizing them into clusters in order to explore the strength and weakness of similar types of locations. Doing so can benefit entrepreneurs looking to open business by choosing the cluster supporting the type of business.

Data Description

The data used to analyze in this project has been collecting using the scraping method from the wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M along with using Four Square API's to get information of venues, and geocoder to obtain coordinates. A file is generated beforehand containing the coordinates of each postal code, named 'Geospatial Coordinates.csv'.

The data will be used to generate clusters to group similar neighborhoods together based on their characteristic, namely venue types and their concentration. This will help deduce the pattern between neighborhoods.