Final Capstone Project – Week-4

IBM Applied Data Science Capstone

Introduction

This is a capstone project for Coursera's IBM Data Science Professional Certificate. In this project, a hypothetical scenario has been portraited for a concept that there might be better opportunity to open a new Yoga Studio in Toronto. Therefore, it might be a great opportunity for an entrepreneur who is from Canada. As the Yoga is very popular among Western as well as Asian communities. Identifying the best possible location to open such a Yoga Studio is one of the most important decisions for this entrepreneur and this project is to design to help and find the most suitable location.

Business Understanding

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Yoga Studio in the city of Toronto, Canada. By using data science methods and tools along with machine learning unsupervised algorithm i.e. Clustering, this project aims to provide solutions to answer the business requirement/question : ***If an entrepreneur wants to open a Yoga Studio in Toronto city in that case which location in Toronto city should he/she consider where his/her business get more benefit and opportunity.***

Target Audience

The entrepreneur who wants to find a best possible location to open a Yoga Studio in Toronto city.

Data Requirements

To analyse and solve this problem, there will be need for below data requirements :

- ✓ List of neighbourhoods of Toronto, Canada
- ✓ Latitude and Longitude of the neighbourhoods
- ✓ Venue data related to existing Yoga Studios in Toronto for final model input and comparison study.

Data Collection

- ✓ Scrapping of Toronto neighbourhoods from Wikipedia link i.e. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- ✓ Getting Latitude and Longitude data of these neighbourhoods via Geocoder package
- ✓ Using Foursquare API to get venue data related to these neighbourhoods