Applied Data Science Capstone Project

Introduction :

Coffee shops are trending in this time and are becoming a high commodity. They can provide multiple services such as serving coffee, drinks, pastry and providing a place to employees and students to work from distance. Therefore, it might be a great opportunity for an entrepreneur to open a coffee shop in Toronto, Canada. This project is for entrepreneurs who are looking for a good and stable investment in Toronto by helping them where to open their Coffee shop. This report aims to create an analysis of features to search best neighborhoods by comparing between neighborhoods.

Buisness Problem :

The major purpose of this project, is to suggest a good neighborhood in a Toronto for the investor to open a Coffee Shop. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question : In Toronto, What is the best place to open a Coffee Shop ?

Approach:

This project would us Scrapping of Toronto neighborhoods via Wikipedia. Extract Latitude and Longitude data of these neighborhoods via Geocoder package. Use Four-square API as its prime data gathering source. And to compare the similarities between multiple locations, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Libraries Used to Develope the Project :

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Geocoder: To retrieve Location Data.

- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.