**Exploring the street trees for Waterloo to see environmental care.**

**1. Introduction**

**a. Background**

**This report is for those who want to see how much environmental friendly the city of Waterloo is. It provides a suggestion on what would be the best venue to start planting trees or at least try to maintain the existing ones.**

**Waterloo may not be the most populated city in Canada right now but it the influx of new students will change that. Waterloo is becoming the Technology hub of Canada with many Tech companies and start-ups opened especially influx of new 100,000 students enrolling every year in universities.**

**b. Business Problem**

**This report focusses on the issue of where to plant trees.**

**c. Interest**

**On what basis can the government decide the plant location?**

**While selecting the place there are key points to consider like they need to check out like where the most well-visited venues of the city are. If the place is more populated, the more trees should be planted.**

**2. Data Preparation**

1. **Scrapping Waterloo Table from City of Waterloo homepage**

<http://data.waterloo.ca/datasets/cityofcambridge::street-trees/data>

**I first make use of waterloo page from City of waterloo to scrap the table to create a data-frame. For this, I used requests and Beautifulsoup4 library to create a data-frame containing name of the districts of Tokoyo, Area, population and 1st Major District. We start as below —**

**2. Approach**

**Foursquare API:**

**This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.**

**Custering Approach:**

**To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a city like Waterloo. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm**

**Libraries Which are 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.**