**Comparing two cities Windsor and Detroit**

Sogra Bilal Memon

April 1st 2020

1. Introduction:
   1. Background

This project aims to compare the neighborhoods of two cities: The Canadian city Windsor and the American city Detroit. Windsor and Detroit are located right across the Canadian – US boarder separated by the Detroit River that runs in between them. These two cities have shared multiple partnerships and agreements such as the Detroit and Windsor Tunnel Corporation which is jointly owned by both the cities. The two cities are connected to each other through the Michigan railway tunnel, Ambassador bridge and the Detroit – Windsor tunnel. More than 40,000 people travel between these two cities per day.

* 1. Problem

As discussed earlier more than 40,000 people cross the US – Canada border between these two cities on a daily basis. This could be for several reasons including tourism, people living in one city while working in the other city or the availability of certain venues in one city that are absent in the other. In April 2013 the development of a new bridge The Gordie Howe International Bridge was approved between these two cities to reduce the traffic on the existing bridge. The Gordie Howe International Bridge is expected to be constructed by 2024 and is predicted to cost a total of $5.7billion. It has also been predicted that this bridge will earn $70.5 million in toll revenue. Understanding the differences between the neighborhoods and the available venues in each of these cities could help us understand why residence one city would commute to the other city.

* 1. Interest

Many stakeholders can benefit from understanding this data. The government can make important decisions related to providing important services such as public transport or required infrastructure based on its citizen’s needs. The analysis of this data can also be helpful to new comers, tourists and business owners. New comers can compare neighborhoods to help them select a suitable location for residence that is close enough to their workplace and meets their other requirements. On the other hand, tourists may be interested in planning a trip that includes different types of venues from both cities. Business owners can understand the missing venues in their city to set up a new business venture that gives users the same facility while giving an added advantage of being closer to the user.

1. Data Acquisition and Cleaning
   1. Data Source

For this project I will be extracting the data from two Wikipedia pages using the web scraping python library Beautiful Soup. I selected two Wikipedia pages that have the names of all the neighborhoods in each of the two cities. You can find the links to these pages here: [Windsor](https://en.wikipedia.org/wiki/Neighbourhoods_of_Windsor,_Ontario) and [Detroit](https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Detroit). From these two pages all the neighborhoods and the boroughs they belong to can be extracted. The geolocation for each of these neighborhoods can be found using the Nominatim python module that converts addresses into latitude and longitude. Using the latitude and longitude of each neighborhood we can call the Foursquare API to get the nearby venues. We can manipulate the data received from the Foursquare API to find the most popular venues in each neighborhood. The neighborhoods can then be cluster using the top 5 venues for each neighborhood as the features. The k-means clustering algorithm will group similar neighborhoods in the same cluster and the dissimilar neighborhoods in a different cluster. The results can then be plotted on a folium map using the geolocation of each neighborhood to position the marker correctly and using a different color for each cluster created.