Introduction/Business Problem

- **Introduction**

  In this hypothetical scenario, the client is a food delivery service similar to Zomato or Uber Eats. They are interested in gaining an understanding of the restaurant 'landscape' in Toronto so that they can effectively target the most suitable restaurants to sign up for their service. The restaurants will be categorised by postal code, meaning there may be multiple neighbourhoods per postal code. However, using geographical distancing form one venue to another, it may also be possible to group restaurants in a way other than by geographical or postal code region. By putting restaurants into catchment areas such as these, it may be possible to enable more efficient (i.e. quicker) deliveries than if restaurants are simply categorised by catchment area. There may also be scope for analysing socioeconomic data of residents in each neighbourhood. The data from this analysis will be used to build up an understanding of the restaurant business in Toronto to allow our client to make an informed choice.
- ⬚ Find out the number of restaurants by postal code. Postal code will be used instead of neighbourhood as it will be used to categorise delivery areas more efficiently.
- ⬚ The data will be sourced as follows:
- o Postal code regions will be web scraped from Wikipedia (https://en.wikipedia.org/iki/List_of_postal_codes_of_Canada:_M)
- o Restaurant data, including data conering ratings, food time and respective geographical distances will be acquired from FourSquare, using their API developer functionality.
- o Geographical data will be accessed from Geospatial Cognitive Class (https://cocl.us/Geospatial_data), including postal codes, neighbourhoods and boroughs as well as longitudinal and latitudinal data.

- **Data**

The data required are the locations of hotels, coffee shops, bars and restaurants in Toronto. To gather the data, we will use the locations of all

Introduction/Business Problem

neighborhoods in Toronto gathered from Wikipedia. Based on these locations we will gather the locations of all venues in these neighborhoods from Foursquare. We will filter the data to acquire the locations of the targeted venues. In order to inspect the data, we will use the folium library to extract the map of Toronto and visualize the locations of the venues on the map. A DBSCAN algorithm will be applied to the locations features to define the 5 clusters of venue. The locations of the warehouse will be defined as the centroids of the clusters. Again, to visualize the map of Toronto, the 5 clusters and the locations of the warehouses we will use the folium library.

•DATA EDA AND VISUALIZATION

Folium Library and Leaflet Map: Folium is a python library that can create interactive leaflet map using coordinate data. Since I am interested in restaurants as popular spots first I create a data-frame where the 'Venue_Category' column in previous data-frame contains the word 'Restaurant'.