Data

For this project, we have used the Foursquare API. Two CVS files about New York and Toronto are downloaded from the links attached below so that their longitude and latitude coordinates are obtained.

New York neighbourhoods: https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json

Toronto neighbourhoods: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We have structured the data and focus on neighbourhoods in Downtown Toronto and Manhattan. Then we merged the data. A Foursquare API GET request is sent in order to acquire the surrounds venues that are within a radius of 500m. The data is formatted using one hot encoding with the categories of each venue. Then, the venues are grouped by neighbourhoods computing the mean of each feature. The similarities will be determined based on the frequency of the categories found in the neighbourhoods. These clusters help my customer to decide which areas are good for her taste.