**Coursera and IBM**

**Week 1 Project Overview**

**Michael Ding**

**Applied Data Science Capstone**

**June 14, 2020**

Week 1 Project Overview

**Problem Definition**

It is no doubt that cultural diversity serves an integral role in the success of many Western nations in an increasingly globalized world. Multiculturalism helps countries build long-lasting products that capture a broad global audience. But when immigrants come to countries like Canada, it is critical that they find a supportive community for their shift and integration as part of Canadian society. Elsewise, immigrants feel secluded and become unwilling to contribute the cultural melting pot of diverse cities like Toronto, NYC, and Sydney.

Our problem today will focus on determining the best communities to open a restaurant for families from Greece that are immigrating to Toronto. To solve this problem, we will focus on finding clusters of neighbourhoods with high frequencies of “Greek restaurant” places as classified by the Foursquare API.

**Solution Method**

As mentioned, the method we will use to solve our problem is to examine places data from the Foursquare API. We will find the mean frequency of “Greek restaurants” places within neighbourhoods and cluster high frequency neighbourhoods together using the K-means clustering method.

We will split our data into a 90% training set and 10% test set to find error values for different values of K. Using the elbow method, we find the optimum value of K to use for our neighbourhood clustering. Afterwards, we should be able to visualize on our Toronto map different clusters of Greek communities and the cluster of non-Greek communities.