IBM DATA SCIENCE CAPSTONE PROJECT

PROBLEM DEFINITION AND DATA

PART1

Background

This Project is going to involve a theoretical business expansion from an early entrepreneur, whose vision is to expand his little pizzeria to another neighborhood in the Bronx or Brooklyn.

Problem Description

Mr. Smith a successful pizzeria owner, has the intention to further expand his little business located in Bronxdale-Bronx, into the borough of Brooklyn, since it is the nearest located borough.

Ideally Mr. Smith desires to open a new pizzeria in Brooklyn but is afraid that neighborhoods in Brooklyn will not share the same characteristics as the ones in the Bronx, and specially his neighborhood. Also, Mr. Smith is worried that competition for pizzeria places will crushed his little venture. In a worst-case scenario, the entrepreneur is willing to only expand withing Bronx's neighborhoods.

Data

The data used for this problem will be the one provided from the Lab of Battle of Neighborhoods Manhatthan in this course, containing data of all the boroughs, neighborhoods in New York.

https://cocl.us/new_york_dataset

Furthermore, using Foursquare API to search for top venues around each neighborhood in the two Boroughs (Brooklyn and Bronx).

Usage of Data

Primary, handling the data for each neighborhood and searching for the top venues in each of the neighborhoods will provide useful insights of the characteristics for each area of Boroughs, whether is a highly commercial, residential or business sector. Then applying K-means machine learning for creating clusters of neighborhoods, without taking into consideration their respective Borough