

## Cousera Capstone Project Report

By: Mohammad Taha Wahab

### Introduction

This project will analyze the New York City data. First, we will find the most visited commercial shop according to the number of check-ins, then we will try to find the neighborhoods that are lacking the selected type of shop which could be potential business opportunity.

### Target Audience

The target audience of this report is any one that is interested in opening a shop but have no idea what kind of and in which neighborhood.

### Data Section

The data comes from **Dingqi Yang** from the following link <https://sites.google.com/site/yangdingqi/home/foursquare-dataset>. It contains 227,428 check-ins in New York city. The data contains a file in tsv format. Each file contains 8 columns, which are:

1. User ID (anonymized)
2. Venue ID (Foursquare)
3. Venue category ID (Foursquare)
4. Venue category name (Foursquare)
5. Latitude
6. Longitude
7. Time zone offset in minutes (The offset in minutes between when this check-in occurred and the same time in UTC)
8. UTC time

## **Application**

We will find the most visited type of shop (commercial) according to the number of check-ins given in the data, then we will try to find neighborhoods that has none of this type of shop.

Examples are for 2000 venues, and the red dot is the center neighborhood which has the most number of Bars between selected coordinates. We did find two neighborhoods that are closest to it having none Bars within 4 kilometers.