

Cousera Capstone Project Introduction + Data

Report by: Hani Raza

Business Problem Statement

This project will analyse the New York City data. To start with, we will find the most visited commercial shop in New York according to the number of check-ins to that shop, then we will try to find the neighbourhoods that are lacking the selected type of shop which could be potential business opportunity and help the people to open shop in that area since these types of shops are not available.

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 shop should he open and in which neighbourhood.

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 .txt 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