

Capstone Report

Introduction

This project will analyze 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 a potential business opportunity.

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