**MOST VISITED COMMERCIAL SHOPS IN NEW YORK CITY**

**BY:**

**KIRUTHIKA ARUMUGAM VIJAYALAKSHMI**

**INTRODUCTION:**

In this project, I am interested in 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 neighbourhoods that are lacking the selected type of shop which could be potential business opportunity.

**BUSINESS PROBLEM:**

The problem is to find the most visited commercial shops in the New York city. The type of commercial shop could be anything like a restaurant or a movie theatre or a furniture showroom. The main idea is to find where the most number of people are going.

**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 two files 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

After extracting and reading the data, we will translate the above data into a Pandas data frame for processing which would look like this. These are the data elements that are needed when we call Foursquare web service call in order to get the venues available in that neighbourhood (Neighbourhoods are not included here)

Then we will create a dictionary in order to decide which category is the most popular (commercial type)

After all this, we will check the coordinates within given n number of kilometres and count how many ‘Bar’ are there (venues selected as 2000 as a trial)

Find the two neighbourhoods that are closest to the coordinate which has the most number of the specific shop type but lacking that within 4 kilometres