# CAPSTONE PROJECT

**INTRODUCTION/BUSINESS PROBLEM**

New York city data will analyse by this project. According to the number of check-ins, we will find the most visited commercial shop, there could be potential business opportunity but some neighbourhoods that are lacking the selected type of shop we try to find it.

**TARGET AUDIENCE**

people who want to do business in commercial shops who have no idea

**DATA SECTION**

The data comes from Ding qi 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 csv format. Each file contains 8 columns, which are:

User ID (anonymized)

Venue ID (Foursquare)

Venue category ID (Foursquare)

Venue category name (Foursquare)

Latitude

Longitude

Time zone offset in minutes (The offset in minutes between when this check-in occurred and the same time in UTC)

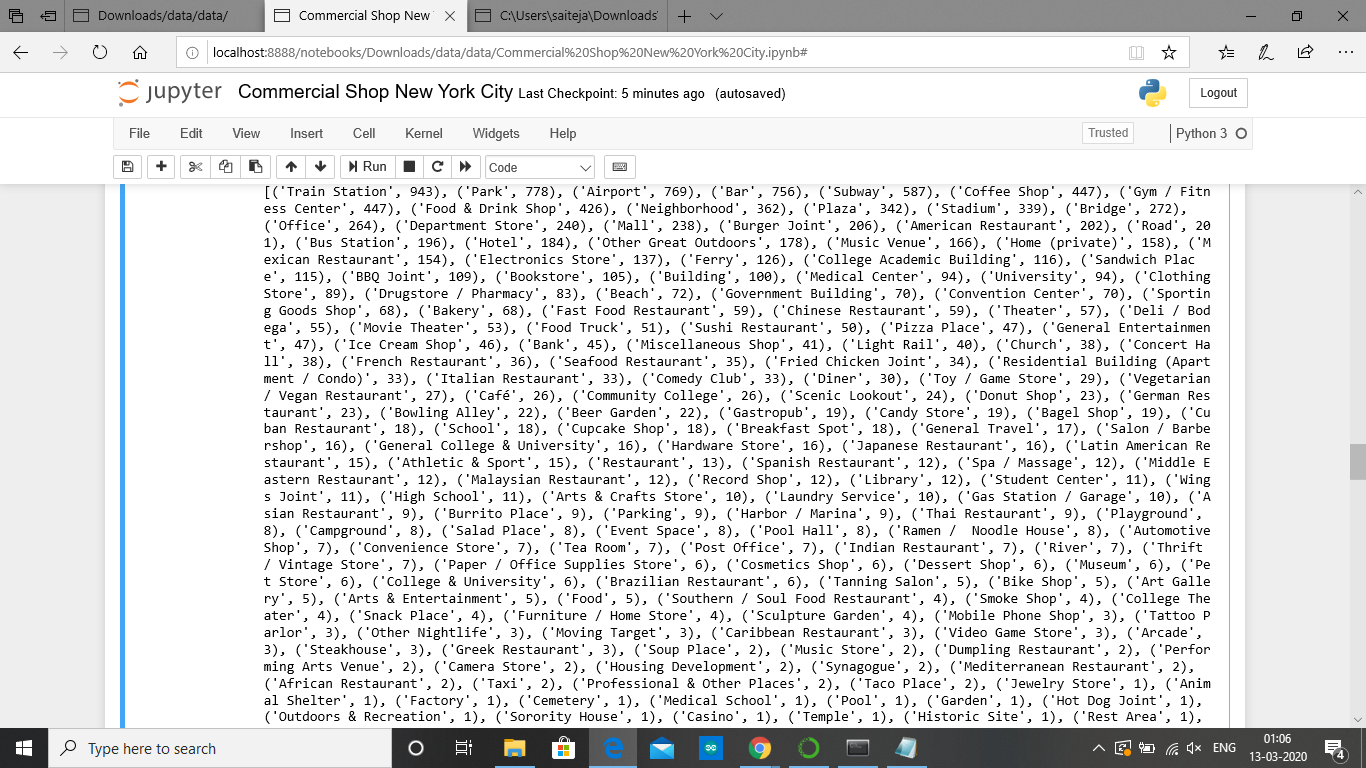
UTC time

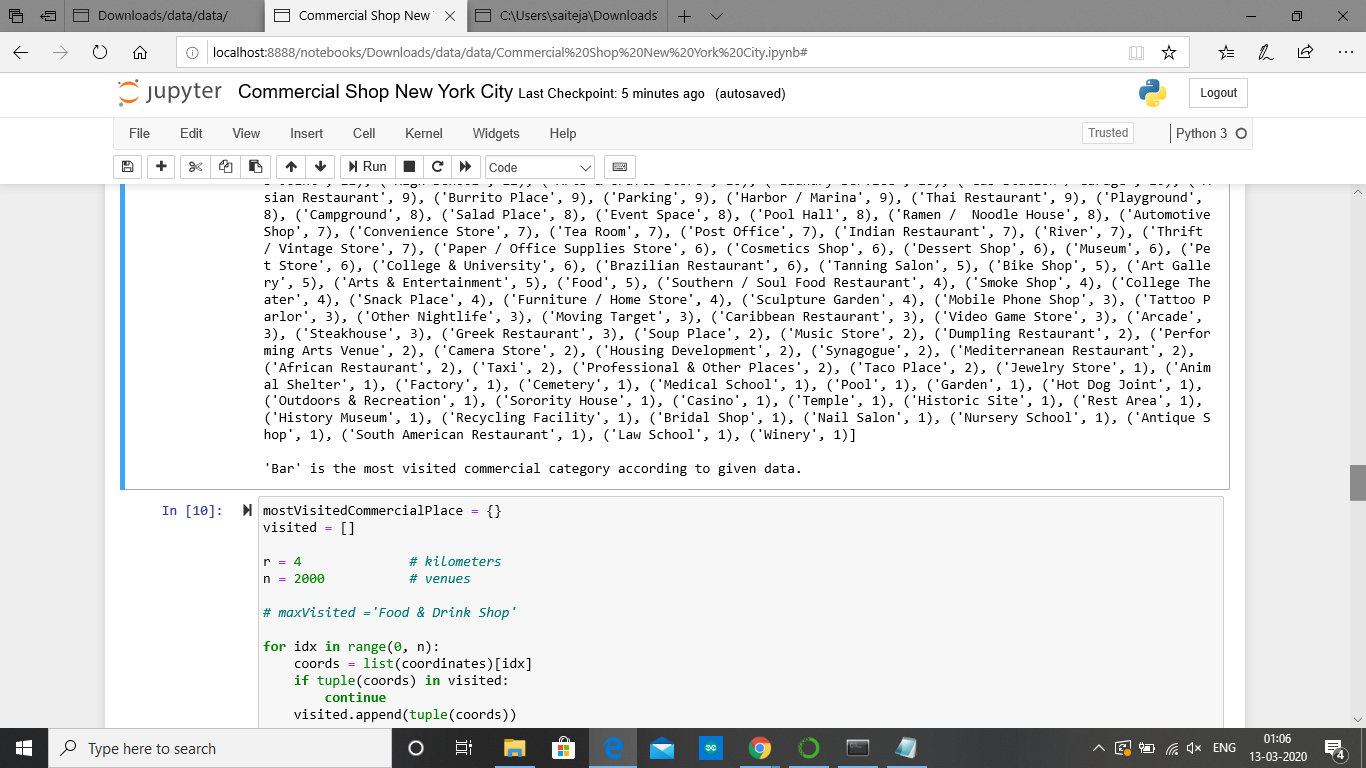
**APPLICATION**

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 neighbourhoods that has none of this type of shop.

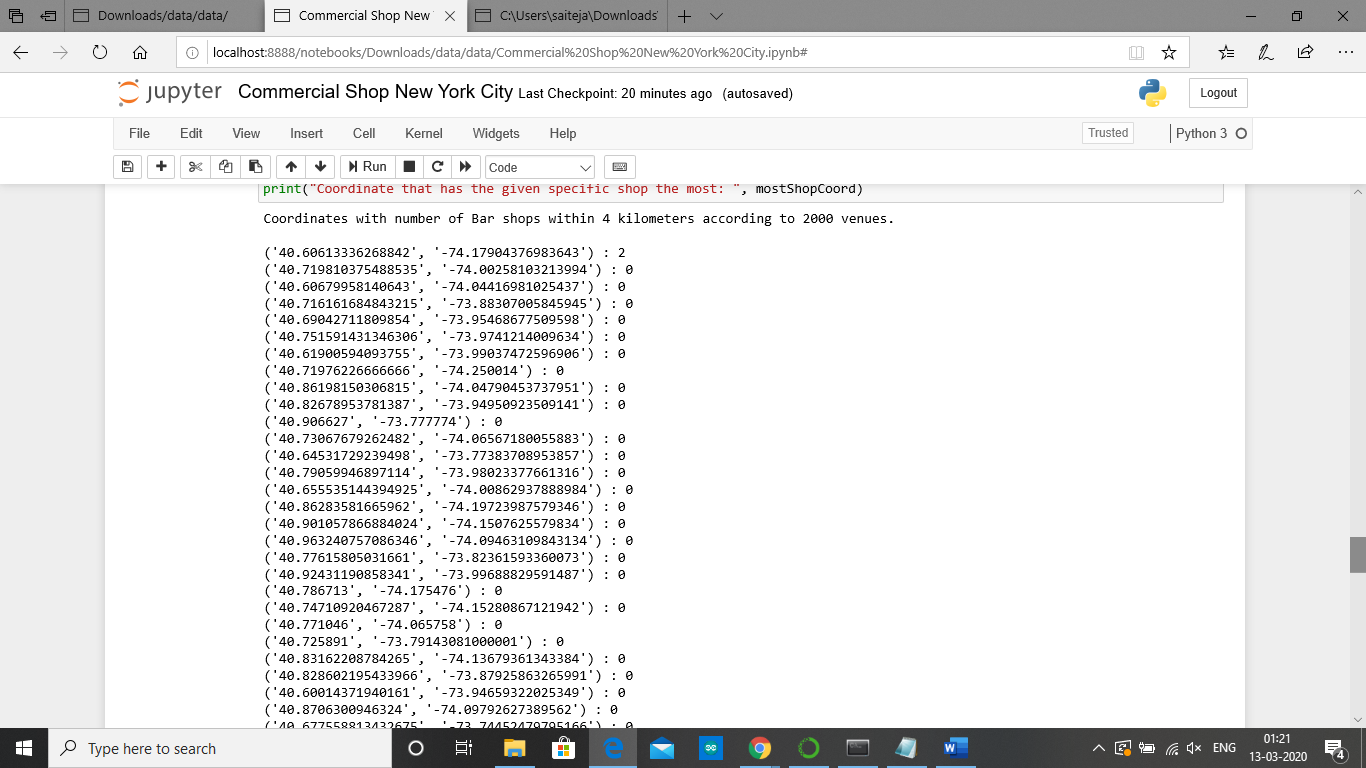
Examples are for 2000 venues, and the red dot is the centre neighbourhood which has the greatest number of Bars between selected coordinates. We did find two neighbourhoods that are closest to it having none Bars within 4 kilometres.







Then we check coordinates and count how many numbers of bars are there (2000 trials)



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

* Bensonhurst
* Bedford-Stuyvesant

