**Project report**

**Introduction**

In This project will analyse the New York City data. Along with it, we will find the most visited commercial shop according to the number of check-ins, then we will try to find the neighbourhood’s that are lacking the given type of shop which Could be potential business opportunity.

**Target Audience**

The target audience of this project is any one that is interested in opening a shop but also have no idea of what kind of shop they want to open and in which neighbourhood should the shop will be situated.

**Data Section**

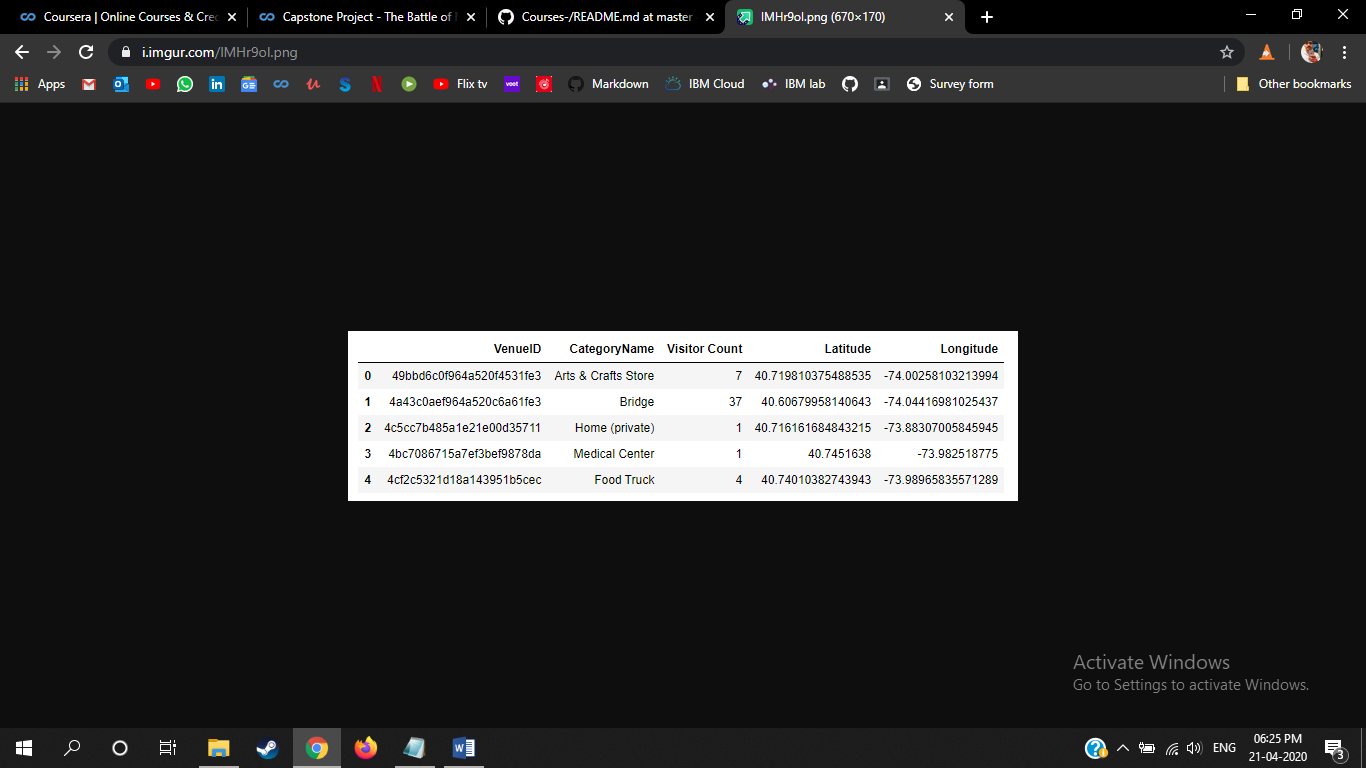
The data comes from **Dingqi Yang** from the following ink [https://sites.google.com/site/yangdingqi/home/foursquare-dataset](#_top)

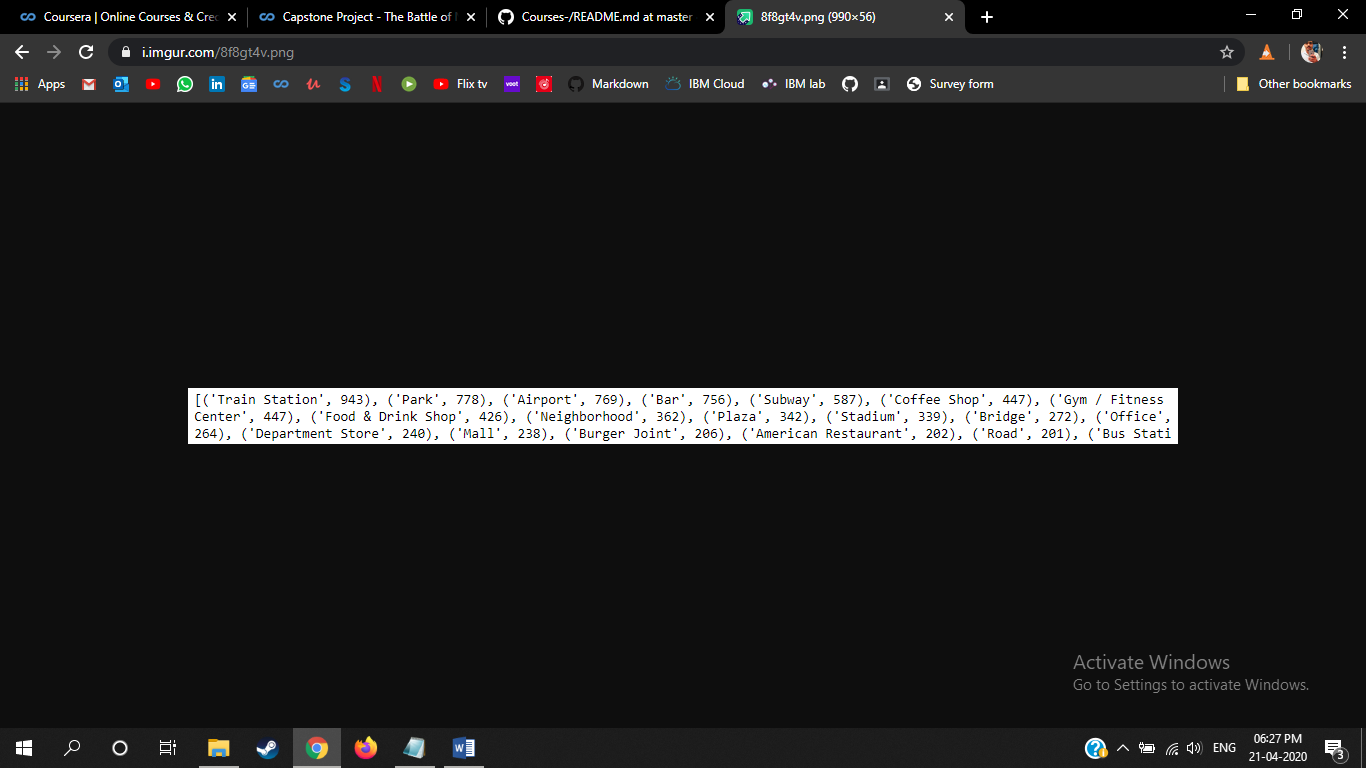
It contains 227,428 check-ins in New York City. The data contains a file in tsv format. Each file contains 8 columns, which are:

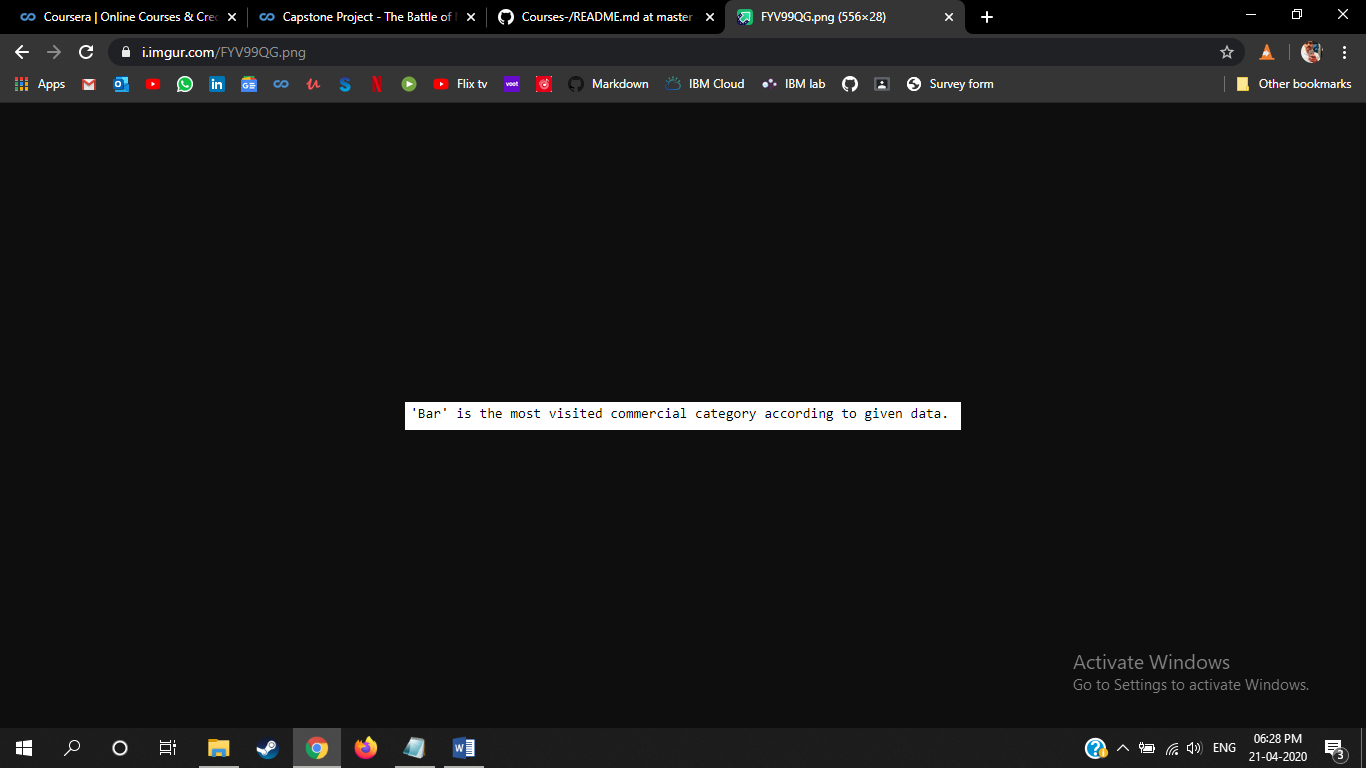
* User ID (anonymized)
* Venue ID (Foursquare)
* Venue category ID (Foursquare)
* Venue category name (Foursquare)
* Latitude
* Longitude
* Time zone offset in minutes (The offset in minutes between when this check-inoccurred and the same time in UTC)
* UTC time

We will find the most visited type of commercial shops according to the number of check-ins given in the data which came from the internet, then we will try to find neighbourhood’s that has none of this type of shop. Examples are for 2000 venues, and the red dot is the centre neighbourhood which has the most number of Bars between selected coordinates. We did find two neighbourhood’s that are closest to it having none Bars within 4 kilometres.

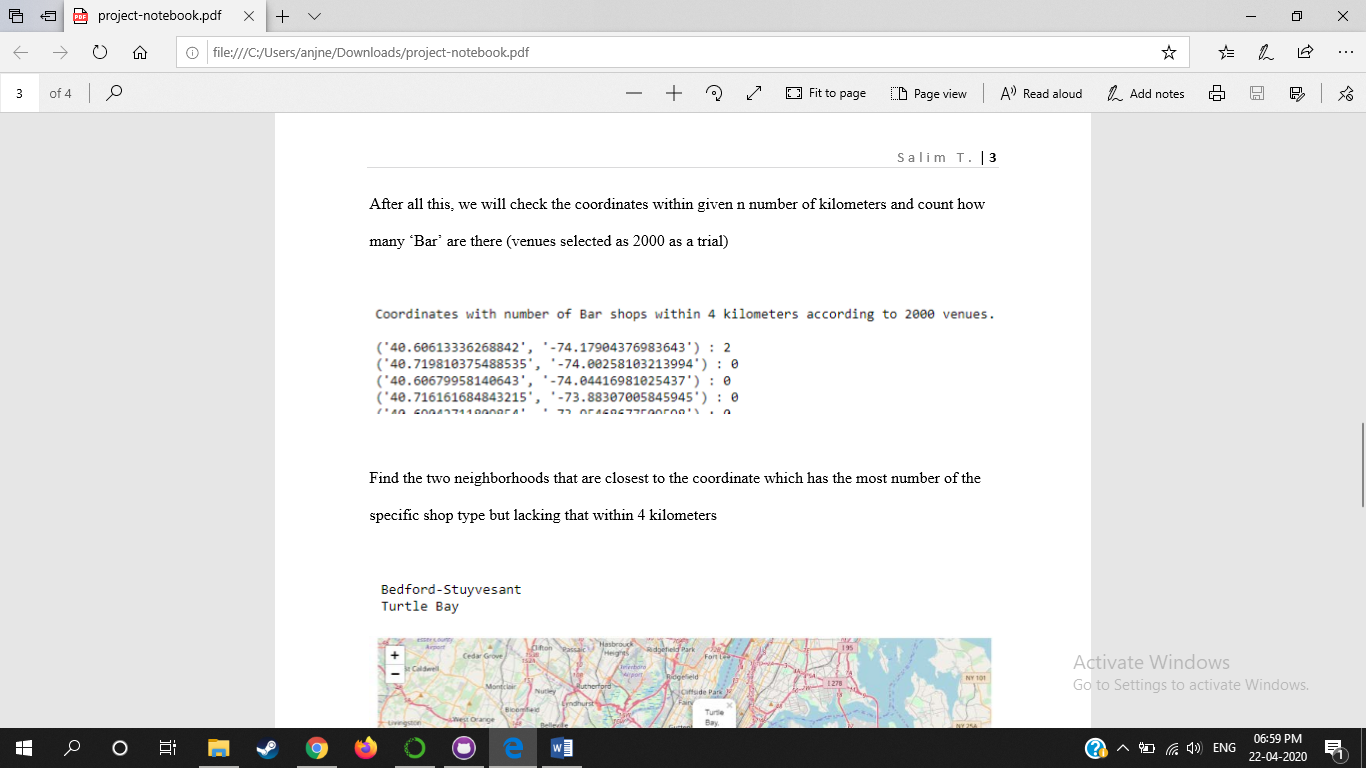
After extracting and reading the data, we will translate the above data into a Pandas data frame for processing which would look like this. These are the data elements that are needed when we call Foursquare web service call in order to get the venues available in that neighbourhood



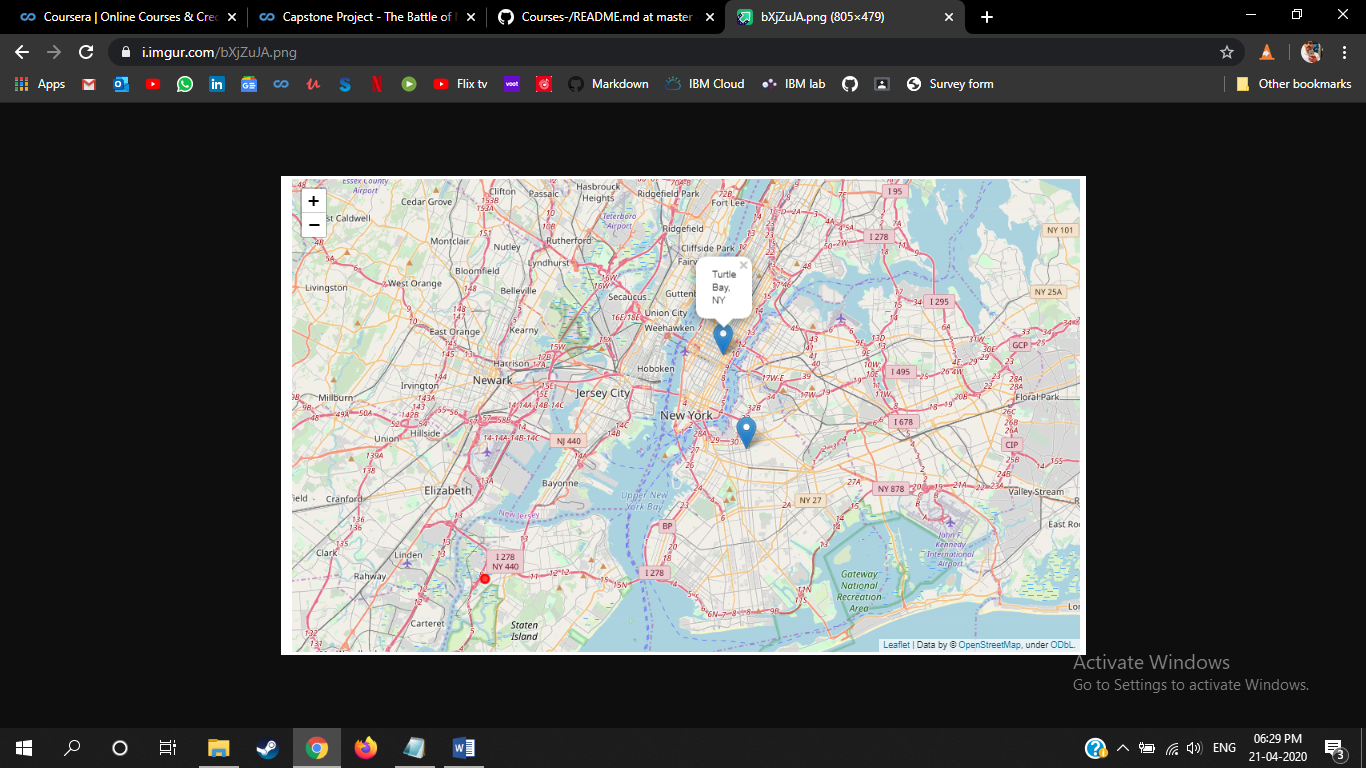




After this we will check the co-ordinates within a given number of Kilometres and count how many bar are there



Then find the two neighbourhood that are closet to the co-ordinate which has the most number of the specific shop type but lacking that within 4-5 km



**Results & Conclusion**

In our sample of 2000 venues, we did find more than 10 coordinates that has no Bar (the most visited shop type according to sample) within four-kilometre sphere. And we did manage to get the neighbourhoods’ names from foursquare database and pin down the two closest neighbourhoods, ‘Bedford-Stuyvesant’, and ‘Turtle Bay’, into the map. Of course, it should not be forgotten that the data used above is almost 6-year old so further research might be needed. Anyways, the results according to the data in hand can be checked from the map and analysis.