Capstone Project-

The Battle of Neighborhoods

PRESENTATION BY:- ANKIT AGRAWAL

1. 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 neighborhoods that are lacking the selected type of shop which could be potential business opportunity.

Data Section

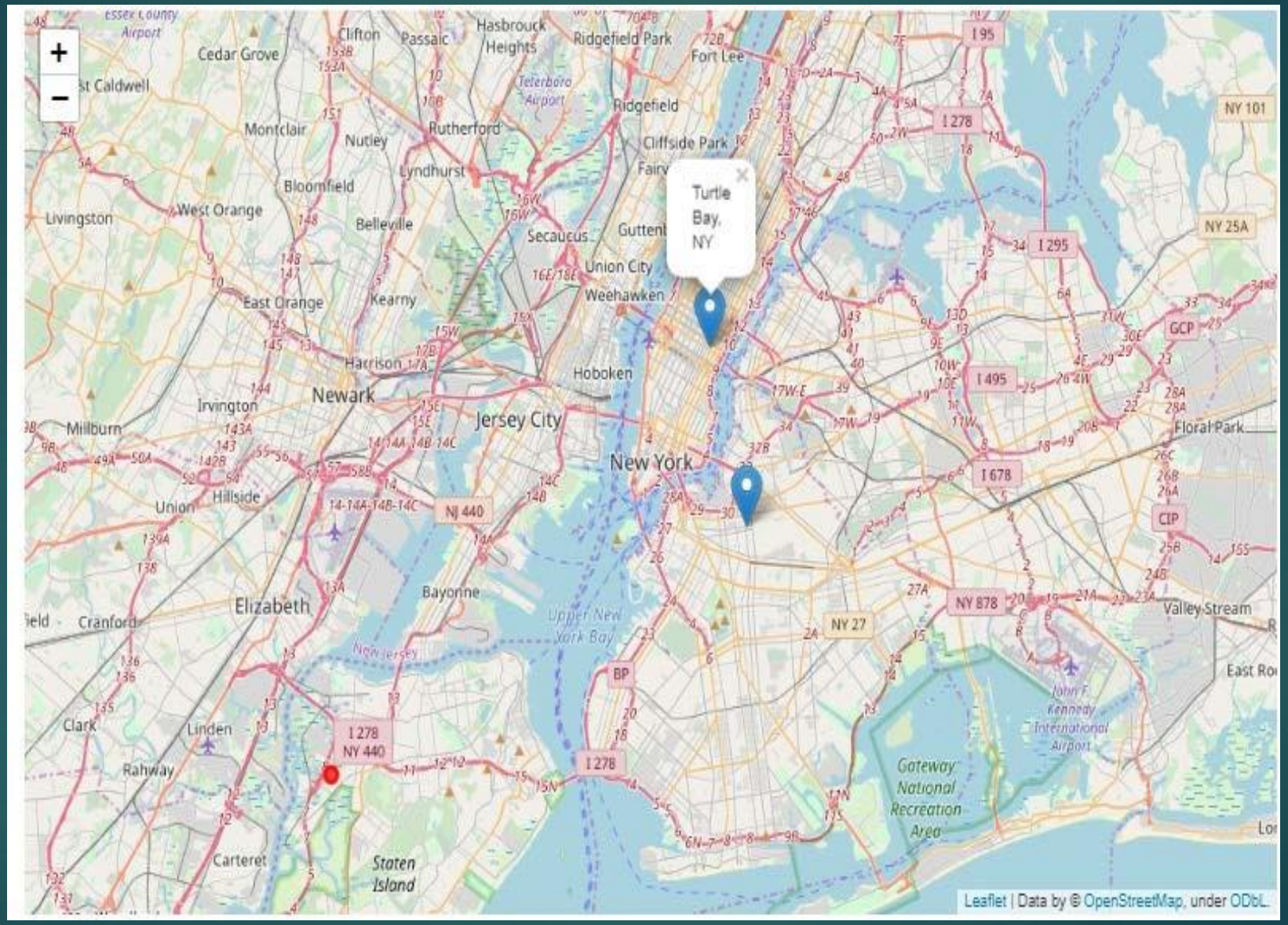
It contains 227,428 check-ins in New York city. The data contains two files in tsv format. Each file contains 8 columns, which are: 1. User ID (anonymized) 2. Venue ID (Foursquare) 3. Venue category ID (Foursquare) 4. Venue category name (Foursquare) 5. Latitude 6. Longitude Salim T. | 2 7. Time zone 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 neighborhood (Neighborhoods are not included here)

	VenueID	CategoryName	Visitor Count	Latitude	Longitude
0	49bbd6c0f964a520f4531fe3	Arts & Crafts Store	7	40.719810375488535	-74.00258103213994
1	4a43c0aef964a520c6a61fe3	Bridge	37	40.60679958140643	-74.04416981025437
2	4c5cc7b485a1e21e00d35711	Home (private)	1	40.716161684843215	-73.88307005845945
3	4bc7086715a7ef3bef9878da	Medical Center	1	40.7451638	-73.982518775
4	4cf2c5321d18a143951b5cec	Food Truck	4	40.74010382743943	-73.98965835571289

Bar is the most visited commercial category according to given data.

After all this, we will check the coordinates within given n number of kilometers and count how many ‘Bar’ are there (venues selected as 2000 as a trial) Find the two neighborhoods that are closest to the coordinate which has the most number of the specific shop type but lacking that within 4 kilometers


Turtle Bay, NY
Cedar Grove
Clifton
Passaic
Hasbrouck Heights
Ridgefield Park
Fort Lee
Ridgefield
Cliffside Park
Montclair
Nutley
Rutherford
Lyndhurst
Bloomfield
Belleville
Livingston
West Orange
Secaucus
Union City
Weehawken
East Orange
Kearny
Harrison
Hoboken
Newark
Irvington
Millburn
Jersey City
New York
Union
Hillside
Elizabeth
Bayonne
Cranford
Clark
Linden
Rahway
Carteret
Staten Island
Upper New York Bay
New Jersey
Gateway National Recreation Area
John F. Kennedy International Airport
Valley Stream
Floral Park
I 95
I 278
I 295
I 495
I 678
NY 101
NY 25A
NY 878
NY 27
NJ 440
GCP
CIP
BP
Leaflet | Data by © OpenStreetMap, under ODbL.

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 neighborhoods that has none of this type of shop.
- Examples are for 2000 venues, and the red dot is the center neighborhood which has the most number of Bars between selected coordinates. We did find two neighborhoods that are closest to it having none Bars within 4 kilometers.

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-kilometer sphere. And we did manage to get the neighborhoods' names from foursquare database and pin down the two closest neighborhoods, '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 above can be of use for future entrepreneurs.