New York park venues that should be avoided

Author: XYZ

Introduction

New York is one of the most popular cities in the world. Due to its reputation, each year the city is visited by approximately 40 million tourists from all over the world. The city is divided into 5 Boroughs: Manhattan, Bronx, Brooklyn, Queens, and Staten Island. Each of them is characterized by unique places to visit. All districts have a large number of parks of various sizes. These parks, due to their characteristics, are visited by people who want to rest from concrete buildings.

A very important part of the tour is ensuring yourself and your loved one's safety while watching places that attract attention. We can achieve such security in several ways, the most effective one is to avoid places with a higher risk of a criminal incident.

The report presents a list of venues in parks with a higher probability of an incident in which we may be a victim. These places are grouped into two groups: the highly probable and medium probability of a criminal event.

The report has been prepared for groups of people. The first is people who just want to explore the venues in New York's parks (tourists or NY residents). The second group is the NY police, which, after the publication of this report, should increase patrols at the shown locations to ensure public safety.

Data

To expose venues that are more risky, we use data taken from Foursquare. This method allows us to exhibit all venues in specific radius from given place of interest.

In order to show the selected objects, the Folium package will be used, which allows us to display maps of the relevant area. Below is shown a map with marked parks and playgrounds in New York city

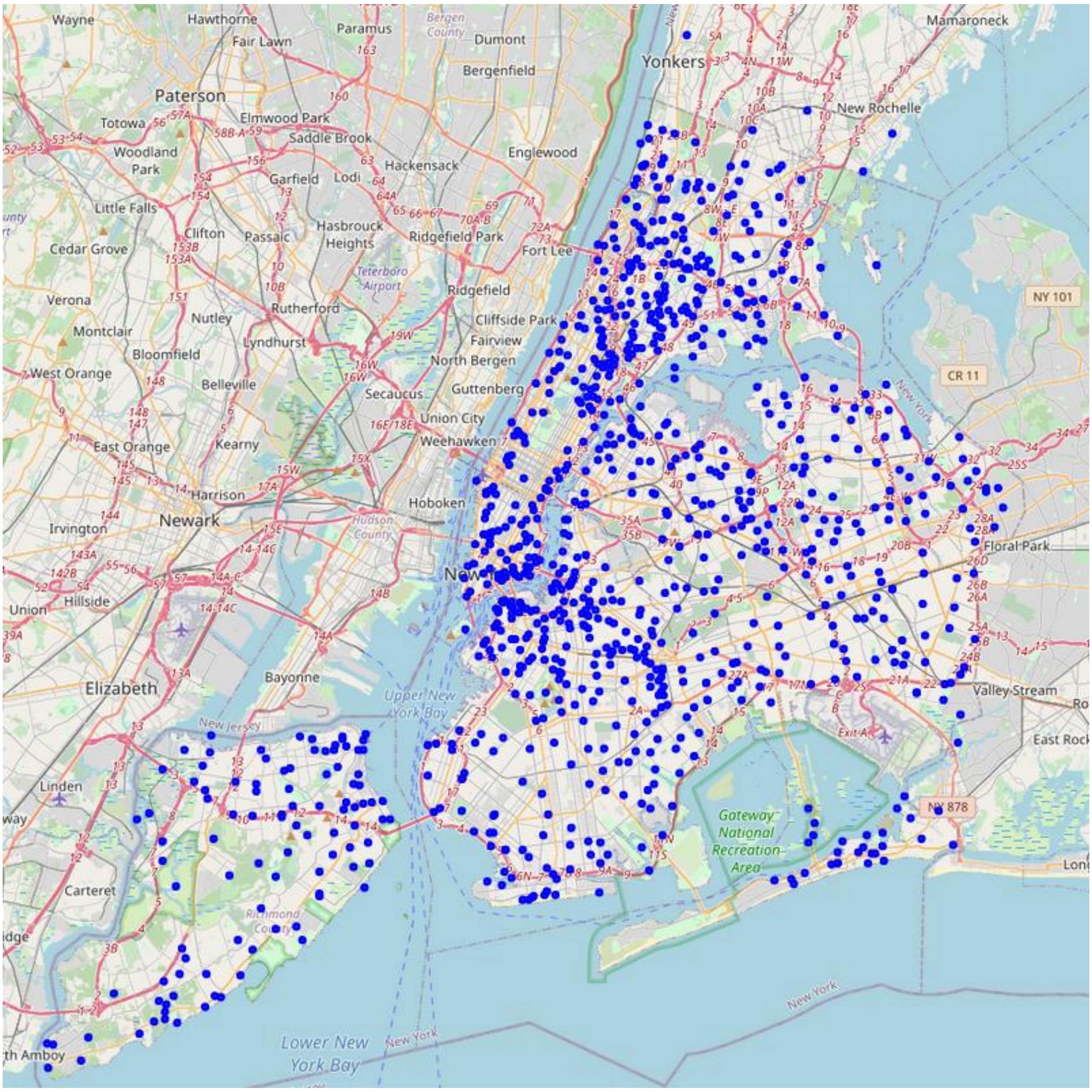

FIG. 1 Map of New York with marked parks and playgrounds as blue dots

The last and most important type of data is data taken from police statistics from 2015-2020 from https://www1.nyc.gov/site/nypd/stats/crime-statistics/park-crime-stats.page. These data show the number of crimes committed in parks and playgrounds throughout the city. These data are published in the .xlsx format and are divided into the years from 2015 to 2020 (data for 2021 are still collected), and additionally, each year is divided into 4 quarters. A sample table (excerpt) is shown below

Table 1 Small fragment of table of crime data obtained from police statistic record

N1130

4th QTRPARK CRIME REPORT

SEVEN MAJOR COMPLAINTS

Report covering the period Between Oct 1, 2020 and Dec 31, 2020

PARK	BOROUGH	SIZE (ACRES)	CATEGORY	MURDER	RAPE	ROBBERY	FELONY ASSAULT	BURGLARY	GRAND LARCENY	GRAND LARCENY OF MOTOR VEHICLE	TOTAL
PELHAM BAY PARK	BRONX	2771.747	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
VAN CORTLANDT PARK	BRONX	1146.430	ONE ACRE OR LARGER	0	0	0	0	0	1	0	1
ROCKAWAY BEACH AND BOARDWALK	QUEENS	1072.564	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
FRESHKILLS PARK	STATEN ISLAND	913.320	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
FLUSHING MEADOWS CORONA PARK	QUEENS	897.690	ONE ACRE OR LARGER	0	0	1	1	1	1	0	4
LATOURETTE PARK & GOLF COURSE	STATEN ISLAND	843.970	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
MARINE PARK	BROOKLYN	798.000	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
BELT PARKWAY/SHORE PARKWAY	BROOKLYN/QUEENS	760.430	ONE ACRE OR LARGER	0	0	0	0	0	1	0	1
BRONX PARK	BRONX	718.373	ONE ACRE OR LARGER	0	0	2	1	0	1	0	4
FRANKLIN D. ROOSEVELT BOARDWALK AND BEACH	STATEN ISLAND	644.350	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
ALLEY POND PARK	QUEENS	635.514	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
PROSPECT PARK	BROOKLYN	526.250	ONE ACRE OR LARGER	0	0	1	2	0	2	0	5
FOREST PARK	QUEENS	506.860	ONE ACRE OR LARGER	0	1	0	0	0	0	0	1
GRAND CENTRAL PARKWAY	QUEENS	460.160	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
FERRY POINT PARK	BRONX	413.800	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0
CONEY ISLAND BEACH & BOARDWALK	BROOKLYN	399.203	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0

The presented data in table is processed as follows:

- Combining all quarters into one year
- Using the Nominatim package to determine the geographic coordinates of all parks listed in the table
- The use of machine learning to cluster the parks into three groups that are characterized by: high, medium and negligible number of crimes committed on parks area.
- With the help of Foursquare, designate venues in parks from the high and medium risk cluster group
- Plotting data on a NY map using the Folium package

Methodology

Data analysis was performed using mainly pandas package with DataFrame objects.

For obtaining geographical location was used **geopy** package, code below:

```python
from geopy.geocoders import Nominatim
```

```python
from geopy.exc import GeocoderTimedOut

def do_geocode(address, attempt=1, max_attempts=10):
    try:
        return geolocator.geocode(address, timeout=None)
    except GeocoderTimedOut:
        if attempt <= max_attempts:
            return do_geocode(address, attempt=attempt+1)
        raise
```

FIG. 2 Code lines used to obtain geographical location information

Due to many TimeOut Errors this recursive definition was used to help prevent this kind of errors from receiving information from server.

To render map we used **Folium** package. In some analysis markers with different markers size was used.

```python
# add markers to map
for lat, lng, label,total in zip(crime2020['LATITUDE'], crime2020['LONGITUDE'], crime2020['PARK'],crime2020['TOTAL'
    label_1 = folium.Popup(str(label), parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=2*math.log(float(total)),
        popup=label_1,
        color='red',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.0,
        parse_html=False).add_to(crime_NY)

crime_NY
```

FIG. 3 Part of code lines to render NY maps with marked crimes commited in 2020. Please note that size of marker denotes higher number of crimes.

This changing in size was achieve by connecting number of crimes with marker radius size.

All bar graphs were plotted using **matplotlib** package.

To group parks into high, medium and low risk area unsupervised machine learnig method was used. Using SciKit-learn package we introduce to analysis a KMeans clustering model with cluster number = 3. As an output we received:

Cluster 2 – high-risk area

Cluster 1 – medium-risk area

Cluster 0 – low-risk area

	PARK
Cluster Labels	
High (red)	1
Medium (purple)	34
Low (green)	932

FIG. 4 Pandas DataFrame showing total number of parks and playgrounds divided into 3 cluster groups

For cluster 2 and 1 we used Foursquare server to obtain nearby venues names and their location.

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

```
venues1=getNearbyVenues(df1.PARK,df1.LATITUDE, df1.LONGITUDE, radius=150)
```

FIG. 5 Code lines of receiving venues information from Foursquare server

Results

The location of parks and playgrounds are included as a blue dots on NY map was showed on FIG.1. To compare that map with total numbers of crimes committed in area of parks and playgrounds render map showed below.

FIG. 6 NY map with marked crimes commited in from 2015 to 2019. Diffrenet colors repres diffrent year: 2015-black; 2016-purple; 2017-green; 2018- yellow; 2019-red

Size of mark represents total value of committed crimes. The bigger mark is, the higher number of crimes was registered on given area. The different colors represent different years from 2015 up 2019 (2020 here was not included, please see FIG.3).

Before we use machine learning to group parks lets we check commited crimes depends on different parameters. Figure below show number of crimes from 2018 as a function of park/playground size. As we can see the number of crimes registered for area larger the one acre is around 10-times higher than for smaller parks/playgrounds.

Statistics

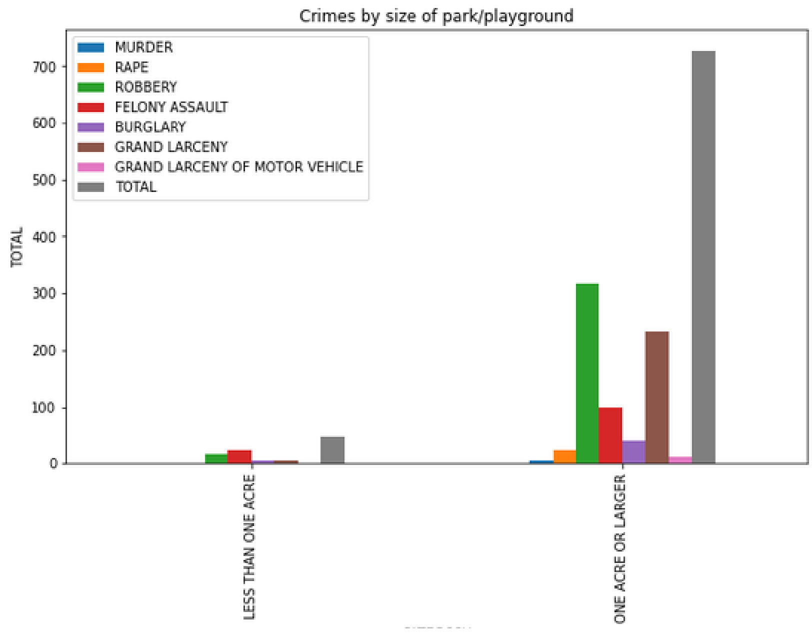

FIG. 7 Numbers of crimes by size of park/playground in 2018

Lets compare now, how total number of crimes was changing past 6 years. As we can see, up to 2019 the total number was almost the same, however in 2020 we can observed drastically fall in numbers

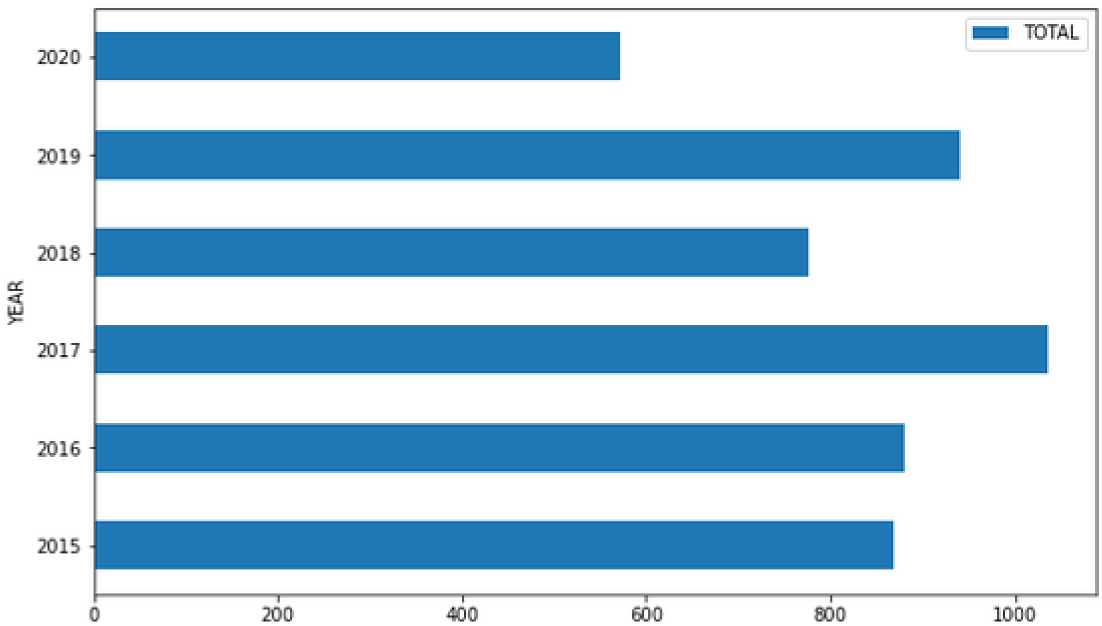

FIG. 8 Total number of crimes committed from 2015 to 2020

When we look at specific crime type we can observe that murder, rape, grand larceny of motor vehicle are constant past 6 years. However, when we look at burglary we can observe monotonic increase past this years.

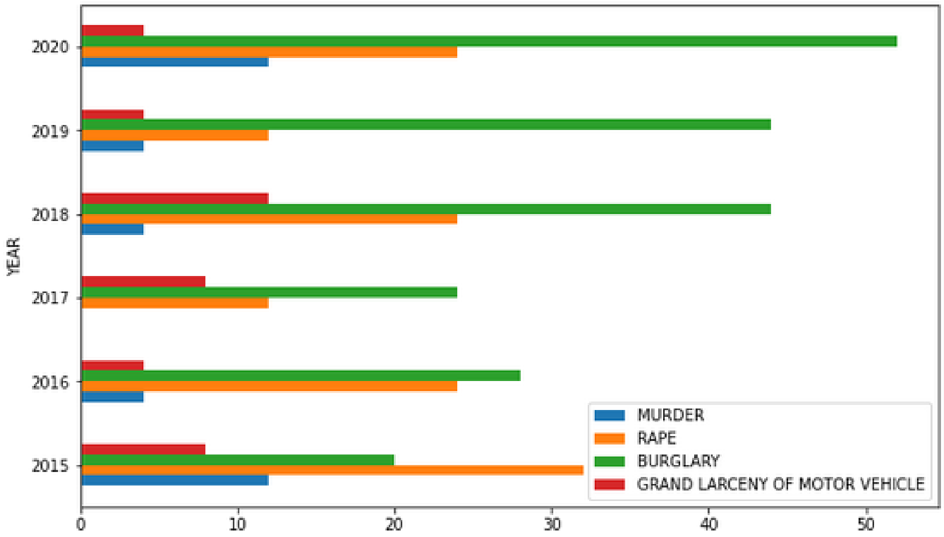

FIG. 9 Number of murder, rape, burglary, and grand larceny of motor vehicle from 2015 o 2020

For robbery, felony assault and grand larceny tendency looks diffrent that for crimes presented at FIG. 9. We can observed for robbery and grand larceny around 50% decrease in number of committed these crimes.

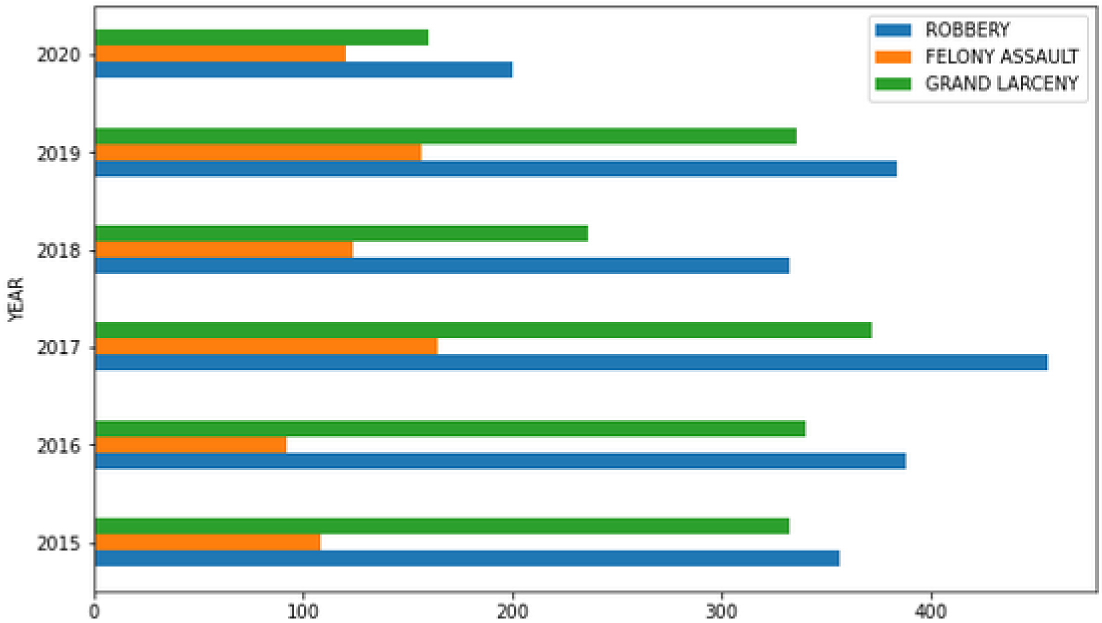

FIG. 10 Number of robbery, felony assault, and grand larceny from 2015 o 2020

Clustering

As was presented in methodology, machine learning was used to group parks in 3 diffrents clusters. The best way to do that is to use KMeans model. This model was analyzing 7 types of crimes and their number in each park over past 6 years. The output of that is shown below

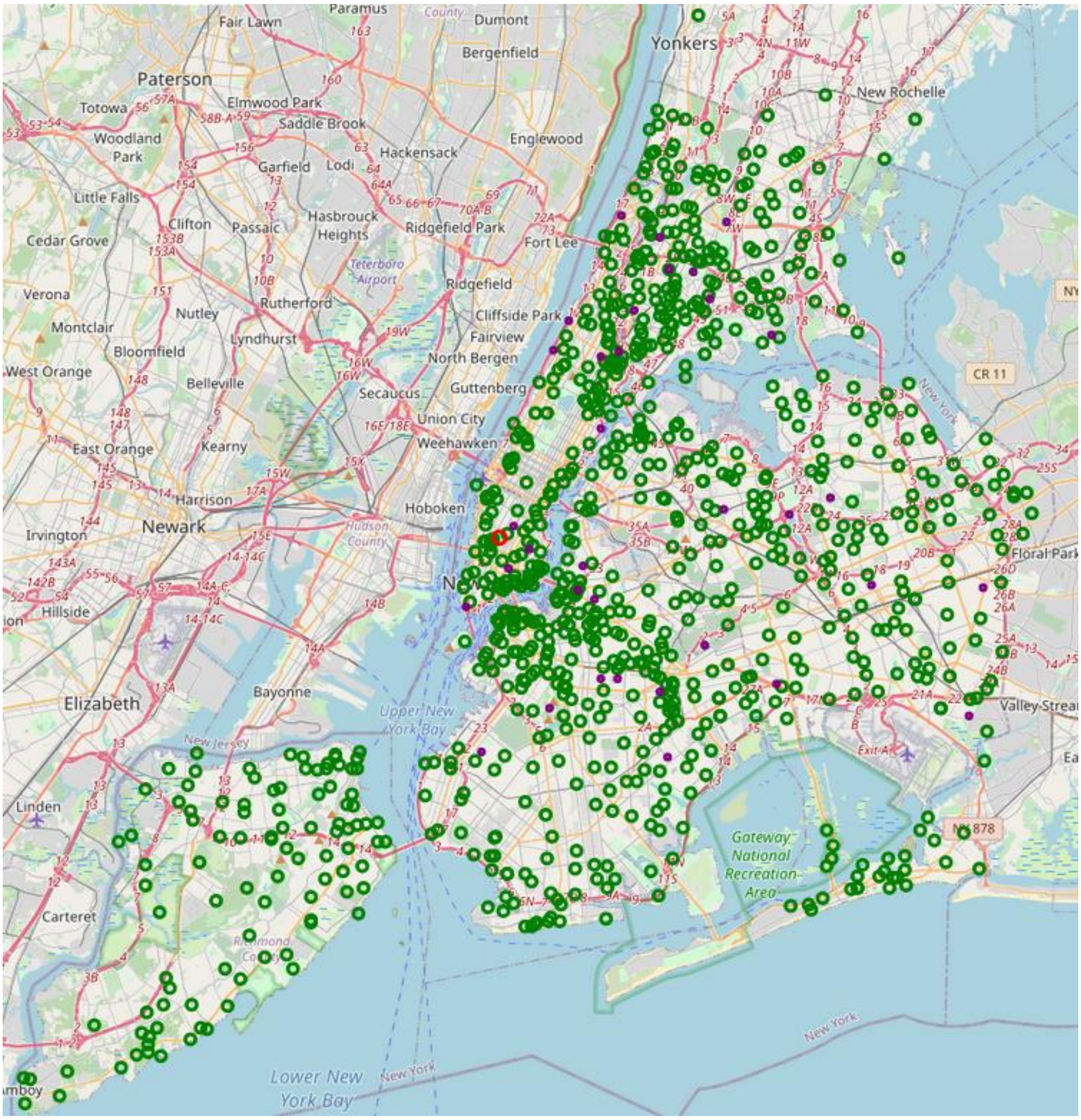

FIG. 11 NY map with marked all parks. Diffrent colors represent safety of that park/playground. Green denote park/playground is safety; purple denote as medium risk in park/playground area; red denote as high risk area

The last step is retrieve from Foursquare server information about venues that are located on medium and high risk park area. For high risk we have venues that are located at Washington Square Park (FIG.12)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	WASHINGTON SQUARE PARK	40.730891	-73.997585	Washington Square Park	40.730816	-73.997458	Park
1	WASHINGTON SQUARE PARK	40.730891	-73.997585	Washington Square Dog Run	40.730767	-73.998490	Dog Run
2	WASHINGTON SQUARE PARK	40.730891	-73.997585	N.Y. Dosas	40.731017	-73.998070	Food Truck
3	WASHINGTON SQUARE PARK	40.730891	-73.997585	Washington Square Fountain	40.730818	-73.997459	Fountain
4	WASHINGTON SQUARE PARK	40.730891	-73.997585	Judson Memorial Church	40.730267	-73.998272	Performing Arts Venue
5	WASHINGTON SQUARE PARK	40.730891	-73.997585	Robin Kovary Run for Small Dogs	40.730117	-73.997526	Dog Run
6	WASHINGTON SQUARE PARK	40.730891	-73.997585	Former Location of Edward Hopper's Studio (191...	40.731017	-73.995846	Historic Site
7	WASHINGTON SQUARE PARK	40.730891	-73.997585	NYU Skirball Center for Performing Arts	40.729848	-73.997716	Performing Arts Venue
8	WASHINGTON SQUARE PARK	40.730891	-73.997585	Washington Square Playground	40.730759	-73.996588	Playground
9	WASHINGTON SQUARE PARK	40.730891	-73.997585	Kimmel Marketplace	40.730090	-73.997656	College Cafeteria

FIG. 12 Washington Square Park venues

However for medium risk we have 151 venues

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BRONX PARK	40.858847	-73.875904	Stone Mill	40.859716	-73.876196	Historic Site
1	BRONX PARK	40.858847	-73.875904	Clay Picnic Pavilions	40.858271	-73.875393	Picnic Shelter
2	PROSPECT PARK	40.661774	-73.971089	Prospect Park	40.661938	-73.969617	Park
3	PROSPECT PARK	40.661774	-73.971089	Prospect Park Dog Beach	40.662484	-73.971967	Dog Run
4	PROSPECT PARK	40.661774	-73.971089	Fallkill Falls	40.661706	-73.971505	Waterfall
...	...	...	...	...	...	...	...
146	VETERANS GROVE	40.742500	-73.877800	Sky Cafe	40.741999	-73.878626	Indonesian Restaurant
147	VETERANS GROVE	40.742500	-73.877800	Yaya Tea Garden	40.741886	-73.878630	Bubble Tea Shop
148	VETERANS GROVE	40.742500	-73.877800	Asian Taste 86	40.741879	-73.878709	Indonesian Restaurant
149	RODNEY PLAYGROUND SOUTH	40.709741	-73.955933	Zeff's Pizzeria	40.709737	-73.954793	Pizza Place
150	RODNEY PLAYGROUND SOUTH	40.709741	-73.955933	Shalom Japan	40.709219	-73.955839	Japanese Restaurant

151 rows × 7 columns

FIG. 13 Venues that are in the parks with medium risk

Map below shows locations of venues that should be avoided

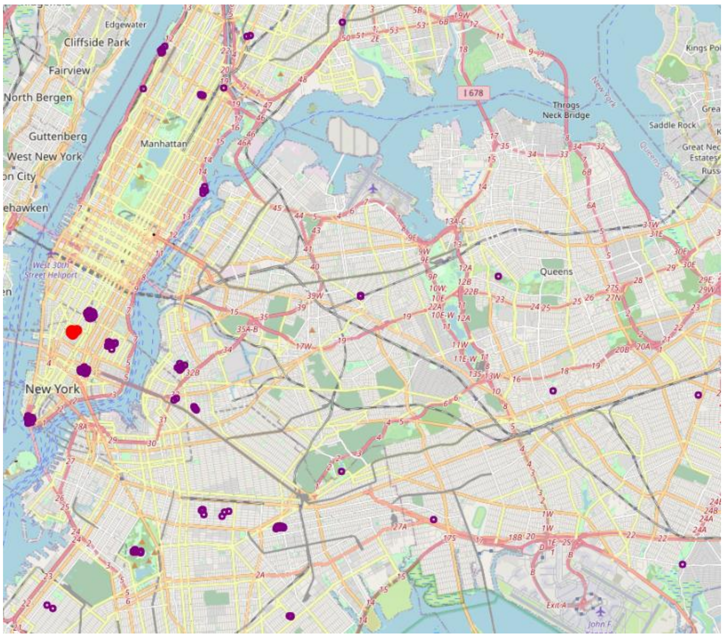

FIG. 14 NY map with marked venues that should be avoided

Discussion

Fig. 7 present differents in crimes number for size area. As we could observed smaller area are almost 10-times safer that larger ones. This is caused by the fact that smaller area is more likely to be playgrounds where the density of people is much more higher that for larger area. More people density makes for criminal much more difficult for to get away from crime incident, than crime is committed in large area of empty space.

FIG. 8 to FIG. 10 shows that total number of crimes besides burglary drastically fall down in 2020. To explain this is quite easy. From the end of 2019 and begging of 2020 all people suffer from COVID-19 pandemic. By this fact huge part of society do not come out from their apartments. This worldwide lock down explain also increase in burglary. Due to low number of people that are outside, the only source of income for criminals are only now from burglary to the apartments and houses.

FIG.11 and FIG.14 shows that the most high-risk area is located at Washington Square Park.

Conclusions

To sum up analysis of parks and venues, we observe that huge part of parks and venues near these parks in NY city are completely safe, besides few places that I would recommended to avoid it. As we can see machine learning are a great way to group huge number of data points into cluster of similar properties, as was shown here to divide around one thousand parks/playground are by their safety.