Studying impact of COVID-19 in USA

Unfolding the reason of why USA is most affected?

1. Introduction

a. Background

We are facing a global health crisis unlike any in the 75-year history of the United Nations — one that is killing people, spreading human suffering, and upending people's lives. But this is much more than a health crisis. It is a human, economic and social crisis. The coronavirus disease (COVID-19), which has been characterized as a pandemic by the World Health Organization (WHO), is attacking societies at their core.

The COVID-19 outbreak affects all segments of the population and is particularly detrimental to members of those social groups in the most vulnerable situations, continues to affect populations, including people living in poverty situations, older persons, persons with disabilities, youth, and indigenous peoples.

Life will be different, even after COVID-19 is brought under control. It will be very challenging for many individuals and businesses. But there will be new opportunities, new endeavors, and we will adapt.

b. Problem Statement

Covid-19 has brought whole world to a standstill which has never been witnessed in last 100 years. With countries worldwide enforcing lockdown, people had to adjust to this new normal of social distancing. Society and Economy thrives on people being social and we are just realizing its importance right now. Last few months have been hard on everyone but, things are looking to improve. A vaccination seems to be coming in near future.

With this ray of hope countries and economies are starting to open. Through this analysis, I wish to analyze the impact of COVID-19 on the world's largest economy, the USA. USA has reported over 2 million COVID-19 cases positive and has the largest death toll in the world. USA government has projected over 150k deaths and they say if they are able to limit the death toll to this number, it will be win for them. In this analysis, we will see

trends of COVID-19 cases and deaths at different levels, first as a whole country, then we dive into states and then we deep dive in to different counties (districts. We will also try to find out reason why USA is most effected.

Finally, we will recommend safe outdoor activities using Foursquare APIs that we can indulge ourselves in and try to get back to our normal lives, like in pre-COVID times.

c. Interest

This pandemic has affected each one of us. Hence, this analysis is for everyone who is interested in finding following questions:

i. How has COVID-19 affected USA?
ii. Which states/counties are the most affected?
iii. How your age plays role in fighting COVID-19?
iv. Is it safe to venture out right now?
v. What all places are safe to go out?

2. Data Acquisition & Cleaning

a. Data Sources

To tackle this problem, we need COVID-19 data for USA. The data should have Key Performance Metrics like number of cases, number of deaths, total population, state name, county name and most importantly this data should be at daily level (for each day).

To get this data we will leverage, https://covidtracking.com/data/us-daily . This organization has developed COVID-19 data set for USA. They have created APIs that we can include in our codes or download a CSV file. For this analysis I have downloaded CSV file. It contains data from January to May. The dataset has following fields:

#	Field Name	Description
1	Date	Date
2	County	Name of County
3	State	Name of State
4	FIPS	Federal Unique number, like ZIP code
5	County_FIPS	Federal Unique number for county
6	State_FIPS	Federal Unique number for state
7	Cases	Number of confirmed cases
8	Deaths	Number of deaths due to COVID-19
9	New_day_cases	Number of new cases on that date
10	New_day_deaths	Number of new deaths on that date

11	Cases_per_capita_100k	Number of cases per 100k population
12	Deaths_per_capita_100k	Number of deaths per 100k population
13	New_day_cases_per_capita_100k	Number of new cases per 100k population
14	New_day_deaths_per_capita_100k	Number of new deaths per 100k population
15	County_pop_2019_est	Estimated Population of County
16	Pop_per_sq_mile_2010	Population per square mile as of 2010

Another requirement would be to get demographic values for these counties. I have download from Kaggle. CSV file contains following fields:

#	Field Name	Description
1	State_FIPS	Federal unique number for State
2	County_FIPS	Federal unique number for County
3	County	Name of County
4	Tot_pop	Total population in that county
5	Male__perc	% of Male population in that county
6	Female_perc	% of Female population in that county
7	Age_0to4	Population between age 0-4
8	Age_5to14	Population between age 5-14
9	Age_15to24	Population between age 15-24
10	Age_25to34	Population between age 25-34
11	Age_35to44	Population between age 35-44
12	Age_45to54	Population between age 45-54
13	Age_55to64	Population between age 55-64
14	Age_65to74	Population between age 65-74
15	Age_75to84	Population between age 75-84
16	Age_85Plus	Population age >= 85

Last but not the least we will use FourSquare API to explore these counties and recommend outdoor activities and places for people to go out. We are using Venue Recommendations API. It returns a list of recommended venues near the current location.

b. Data Cleaning

Foursquare API response parsing was required. I build a custom function to do the trick. It takes in the response JSON and returns a Pandas DataFrame. The structured data facilitates data insights.

The DataFrame contains following fields:

#	Field Name	Description
1	ID	Unique ID of the Venue
2	Name	Name of the Venue
3	Latitude	Latitude of the Venue
4	Longitude	Longitude of the Venue

5	State	Name of State when Venue is located
6	Category_ID	ID of Type of Venue
7	Category_name	Name of Type of Category

```
def build_df(results):

    trend = pd.DataFrame(columns=['id', 'name', 'latitude', 'longitude', 'state', 'category_ids', 'category_names'])
    for itm in results:
        item = itm['venue']
        _id = item['id']
        _name = item['name']
        _lat = item['location']['lat']
        _long = item['location']['lng']
        _state = item['location']['state']
        _category_ids = []
        _category_names = []
        for category in item['categories']:
            _category_ids.append(category['id'])
            _category_names.append(category['name'])
        temp_df = pd.DataFrame(data={'id':[_id], 'name':[_name], 'latitude':[_lat], 'longitude':[_long], 'state':[_state],
        trend = trend.append(temp_df).reset_index(drop=True)

    return trend
```

3. Exploratory Data Analysis

a. Defining Target KPIs

There are two key metrics that we are tracking in this analysis. One is number of cases and second is number of deaths. This KPI is available to us at different levels like county level, state level and national level.

b. Visualizing target KPIs at National level

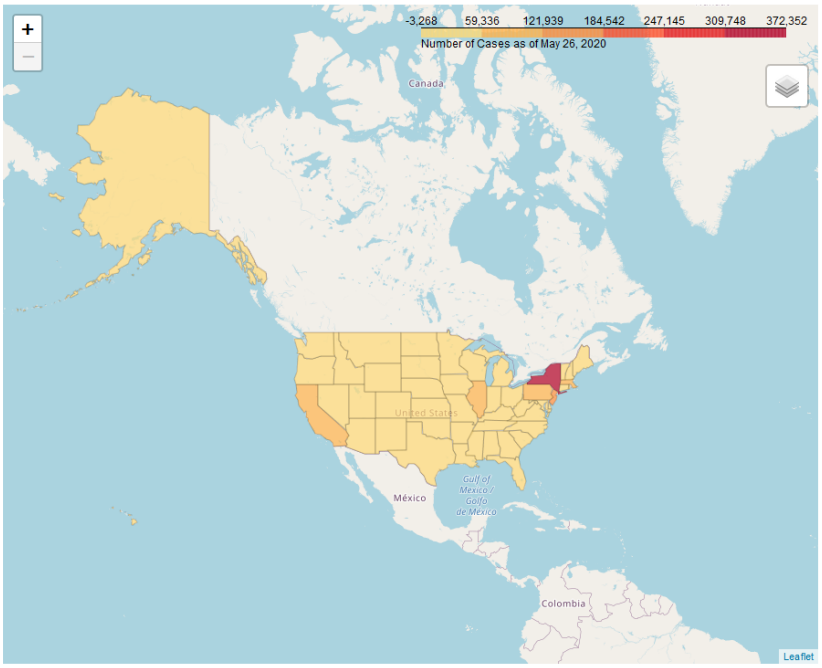


Figure 1 Number of Cases across USA

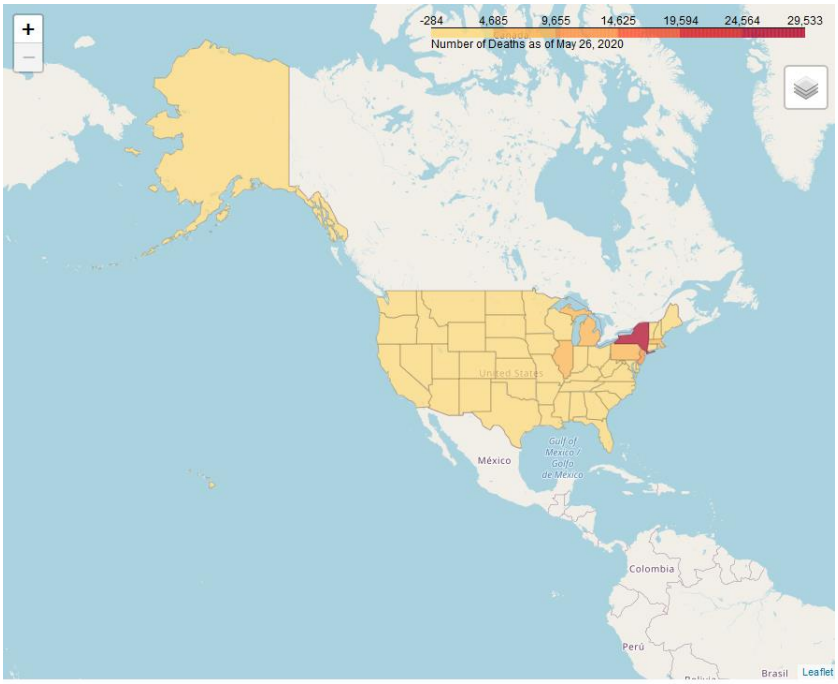


Figure 2 Number of Deaths across USA

c. Visualizing target KPIs at State Level

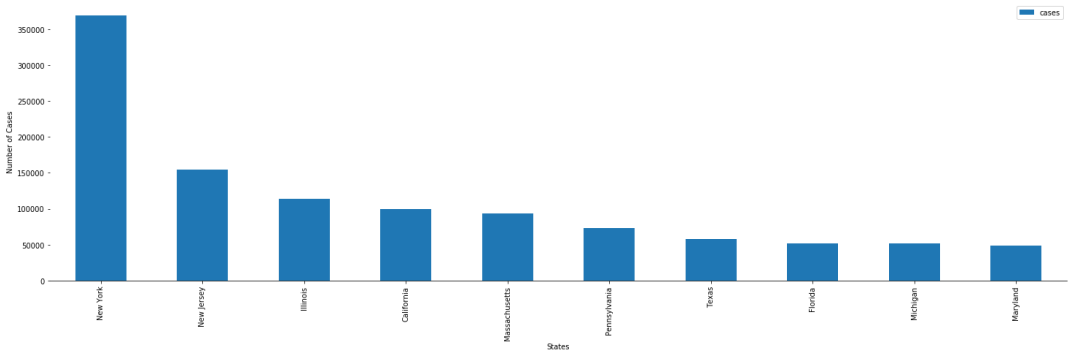


Figure 3 Top 10 states with highest number of Cases

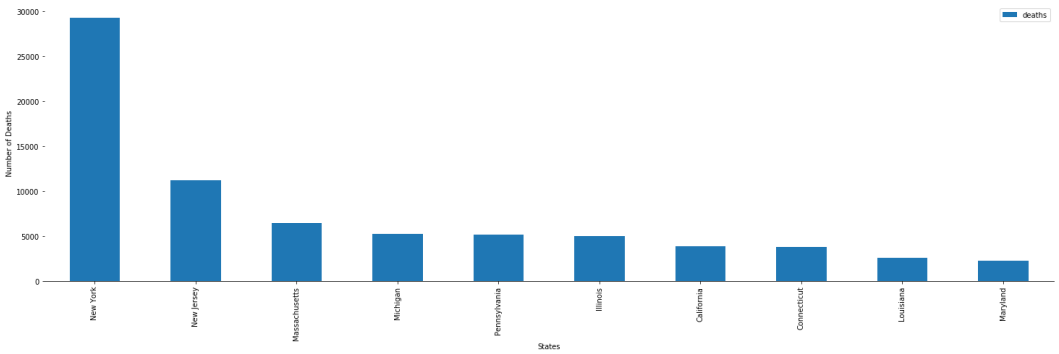


Figure 4 Top 10 states with highest number of Deaths

4. Results based on statistical techniques

a. Does COVID-19 effect males more than females or vice-versa?

```
cases        -0.10
deaths       -0.11
MALE_PERC     1.00
Name: MALE_PERC, dtype: float64
```

```
cases          0.10
deaths         0.11
FEMALE_PERC    1.00
Name: FEMALE_PERC, dtype: float64
```

Correlation values are near to 0 which suggests there is no correlation between our target KPIs and number of cases or number of deaths.

b. Does COVID-19 effect differ based on your age?

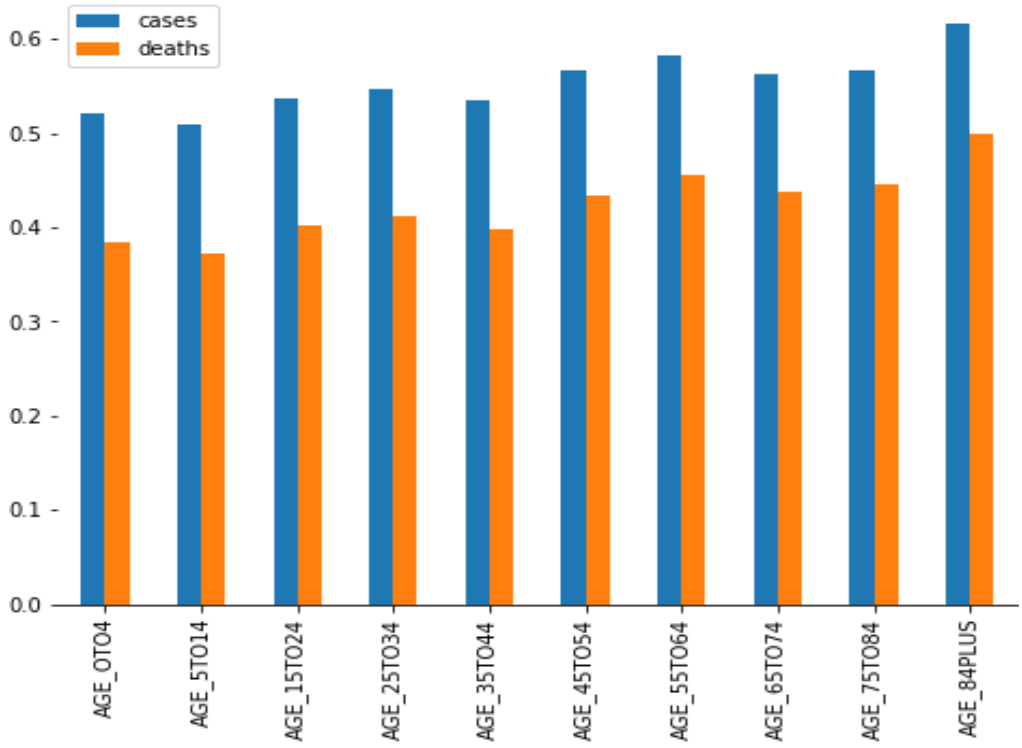


Figure 5 Correlation values for different age groups w.r.t to number of cases and deaths

From graph above it is very evident that as your age increases, you're more vulnerable to get infected with COVID-19 and survival rate decreases.

c. How is New York affected?

i. Trend of COVID-19 in state of New York

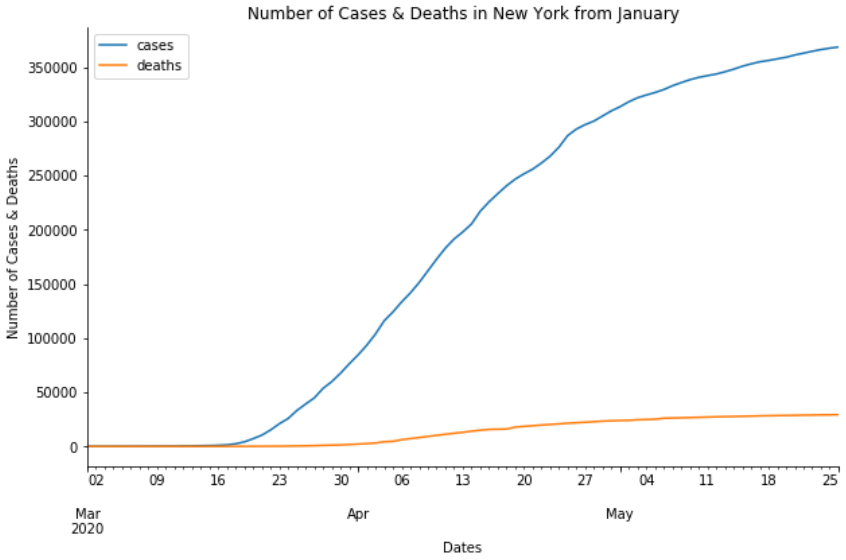


Figure 6 Trend line for number of cases & deaths in state of New York

ii. What is population distribution in state of New York?

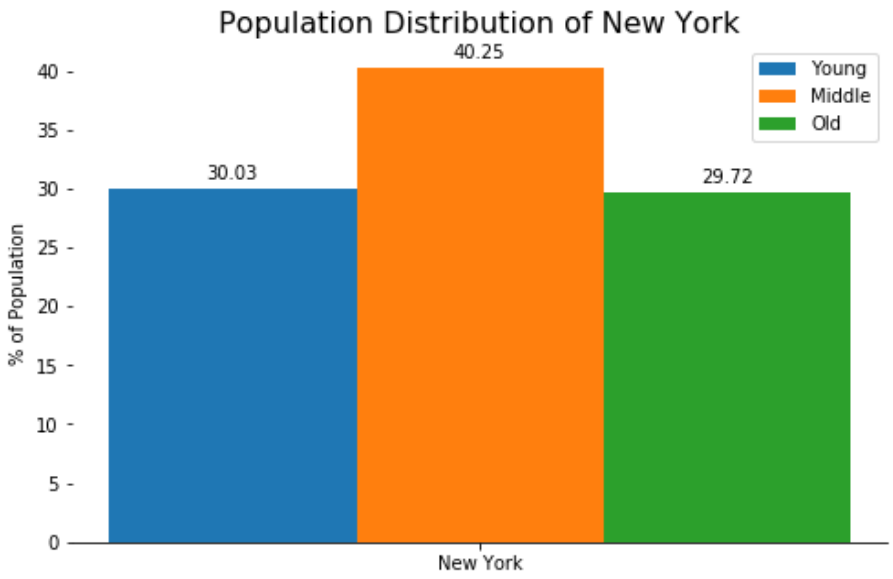


Figure 7 Population distribution in the state of New York

The chart above again presents the fact that older you are, more vulnerable you are to get infected with COVID-19. New York population is 70% + Middle age & Old age people.

iii. Which counties are most affected in New York?

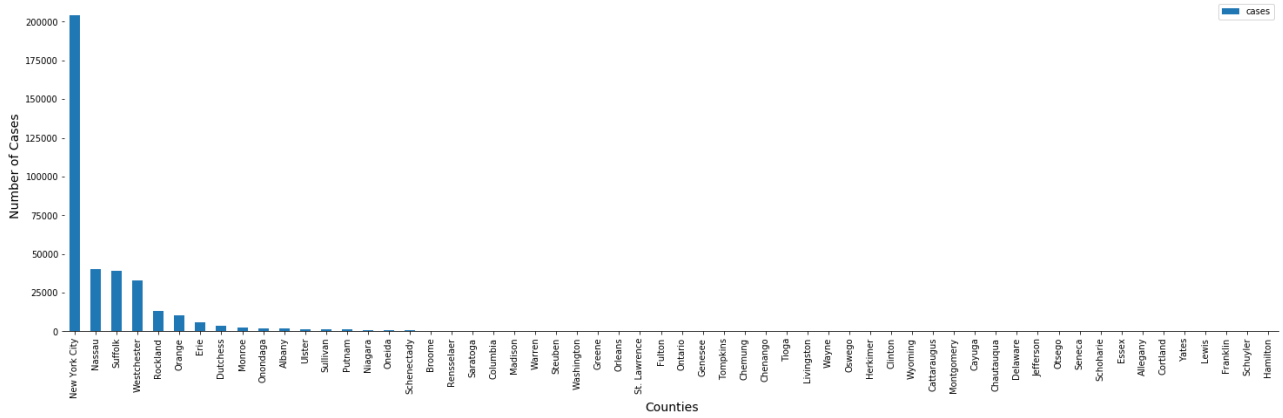


Figure 8 Number of Cases for each county in state of New York

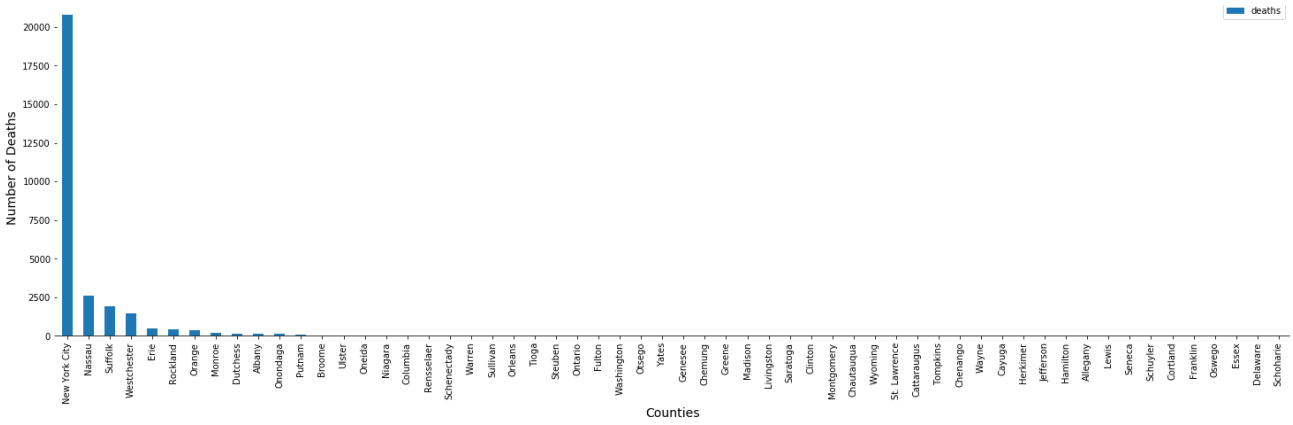


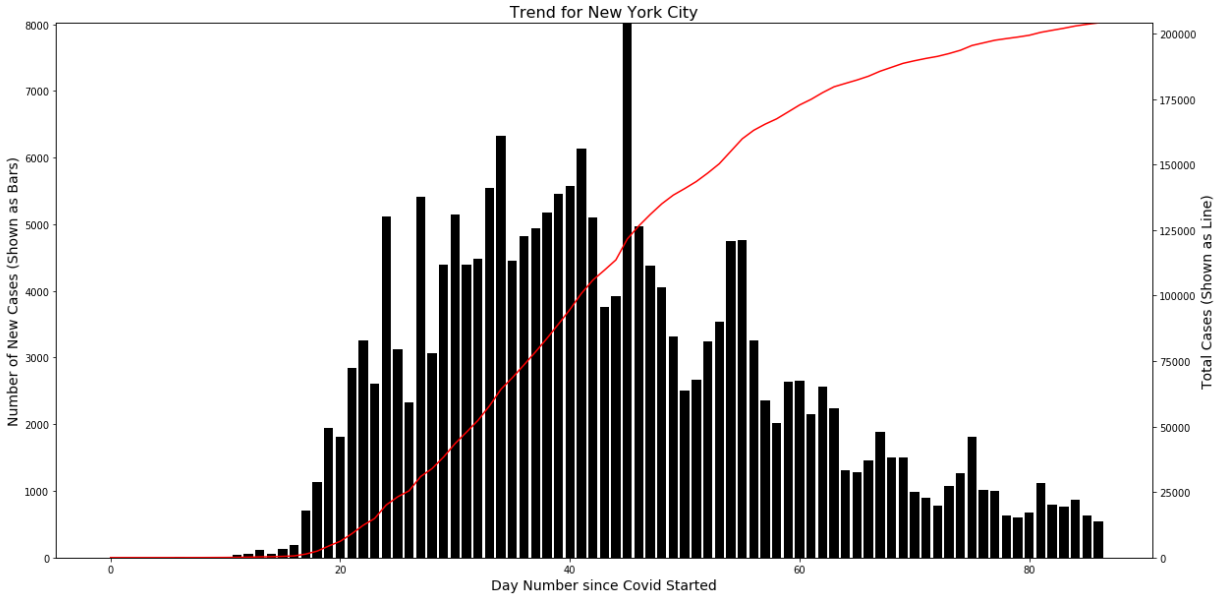
Figure 9 Number of Deaths for each county in state of New York

Results show that Top 5 infected counties contribute 98% of the total cases & deaths in the state of New York.

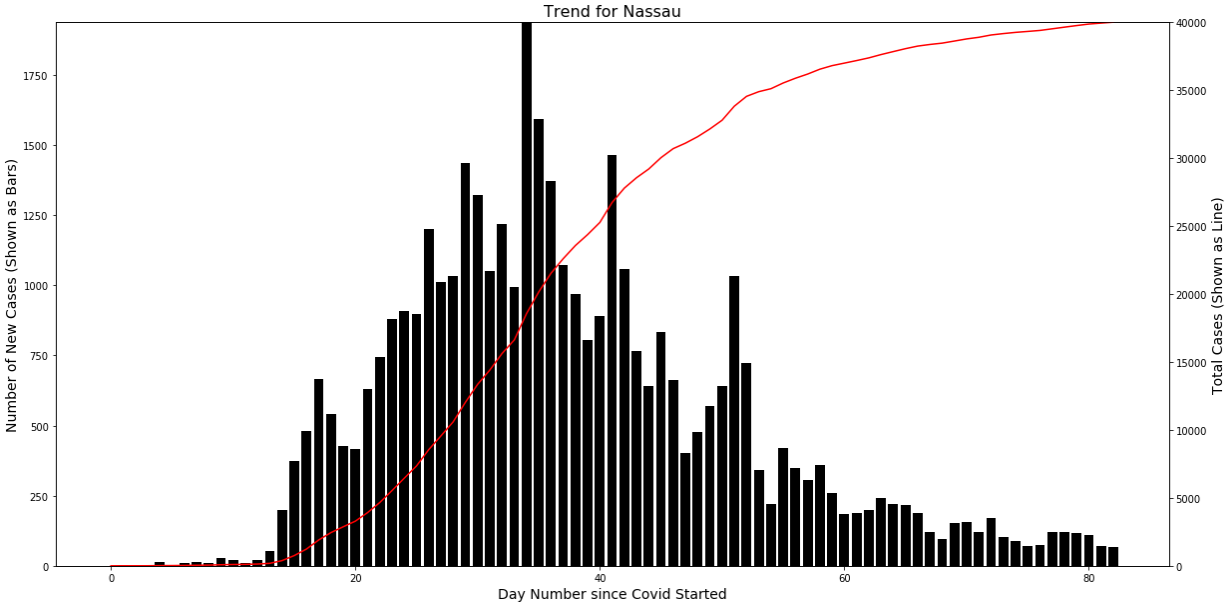
iv. Is the situation getting better in New York?

To answer this, we investigate trends for top 5 infected counties in New York since they constitute 98% of total cases & deaths in the state.

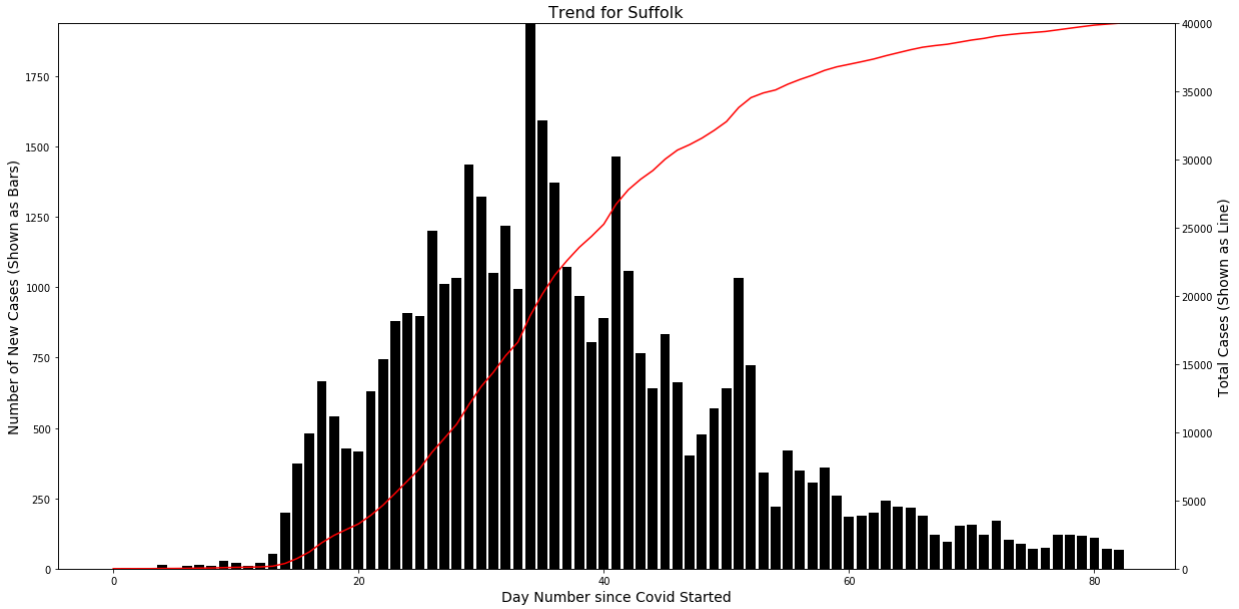
1. County 1: New York City



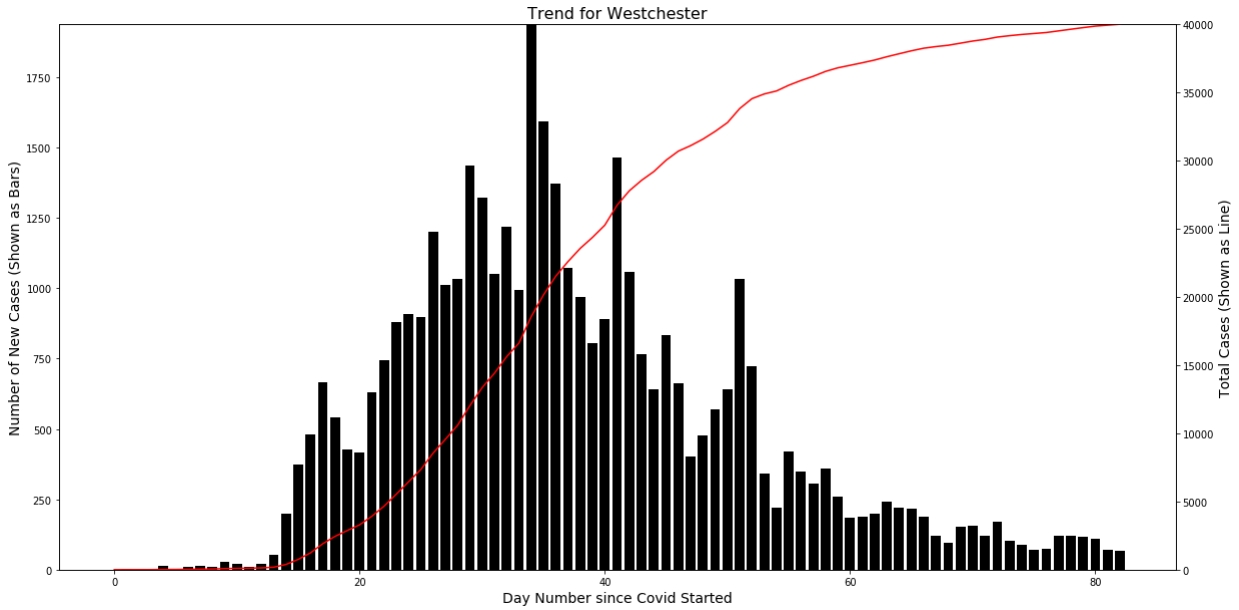
2. County 2: Nassau



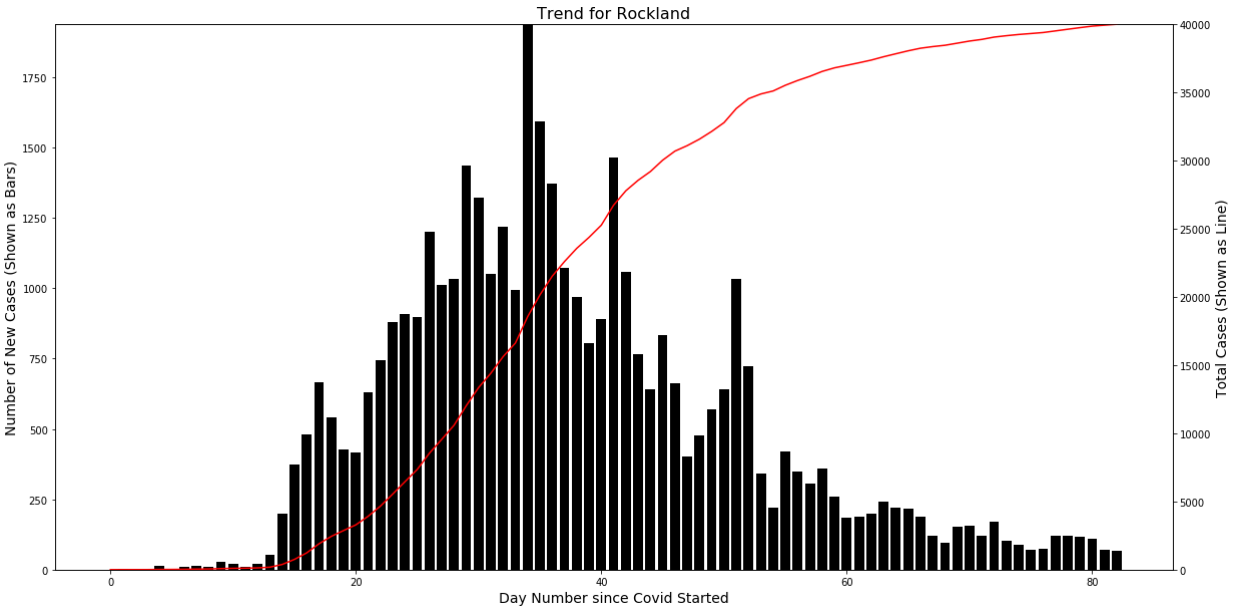
3. County 3: Suffolk



4. County 4: Westchester



5. County 5: Rockland



All graphs above suggest that though the number of cases is high, the number of new cases being reported daily have decreased significantly over the last 30 days. Hence, it can be said that it is safer than before to venture out.

v. Recommended places to venture out in the state of New York

Using Foursquare API, I have scraped recommended venues. The list of contain following categories:

#	Category
1	Park
2	Pizza Place
3	Ice Cream Shop
4	Gym
5	Brewery
6	Bakery
7	Grocery Store
8	Trail
9	American Restaurant
10	Beach
11	Italian Restaurant
12	Scenic Lookout
13	State / Provincial Park

14	Sandwich Place
15	Gym / Fitness Center
16	Boxing Gym
17	Mexican Restaurant
18	Museum
19	Bar
20	Restaurant
21	Supermarket
22	Theater
23	Café
24	Wine Shop
25	Seafood Restaurant
26	Fountain
27	Plaza
28	New American Restaurant
29	Surf Spot
30	Japanese Restaurant
31	Art Museum
32	Indie Movie Theater
33	Beer Bar
34	Yoga Studio
35	Deli / Bodega
36	Concert Hall
37	Cheese Shop
38	Farmers Market
39	Farm
40	Coffee Shop
41	Chinese Restaurant
42	Cycle Studio
43	Reservoir
44	Convenience Store
45	Community Center
46	Sculpture Garden
47	Cocktail Bar
48	Church
49	Tennis Stadium
50	Sushi Restaurant
51	Performing Arts Venue
52	Thai Restaurant
53	Track
54	Caribbean Restaurant
55	Vineyard
56	Warehouse Store
57	Waterfront
58	Butcher
59	BBQ Joint

60	Food Truck
61	Flea Market
62	Outdoor Sculpture
63	French Restaurant
64	Botanical Garden
65	Greek Restaurant
66	Field
67	Arepa Restaurant
68	Fast Food Restaurant
69	Hot Dog Joint
70	Falafel Restaurant
71	Exhibit
72	Jazz Club
73	Library
74	Lighthouse
75	Event Space
76	Movie Theater
77	Dog Run
78	Nature Preserve
79	Diner
80	Opera House
81	Breakfast Spot

Are you looking to go to Park?

```
get_places('Park')
```

```
: 'Riverside Park, Central Park, Central Park West - W 86th St, Carl Schurz Park, Astoria Park, Riverside Park South, Central Park - Wien Walk, Manor Park, Kensico Dam Plaza, Piers
on Park, Congers Lake Memorial Park, Waveny Park, Harbor Island Park, Untermyer Park, Greenwich Point Park, Van Cortlandt Park, Inwood Hill Park, Fort Tryon Park'
```

Are you looking to go to eat good Sushi?

```
get_places('Sushi Restaurant')
```

```
: 'Onsen Sushi and Restaurant'
```

5. Future Scope

Analysis around COVID-19 have caught eyes for all data scientists around the world and many questions are still left unanswered. We can explore how forecasting models can be leveraged to have be better prepared for all businesses that have come to standstill.