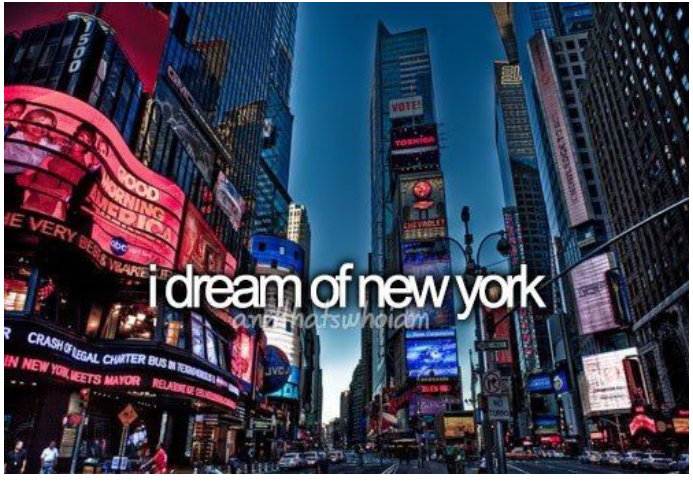
New York City - Safest Borough

Gun violence in the 21st century

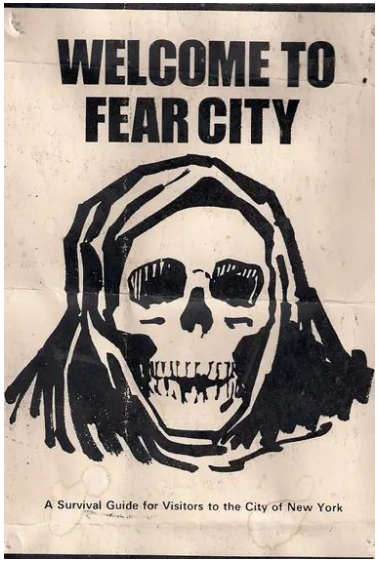
New York City - The City of Dreams



New York City is seen as the city of dreams where anything is possible. Success stories are born in the city. Movies and TV shows often depict a struggling artist who moves to NYC and becomes extremely successful. But NYC wasn't always seen as this.

How “Fear City” became “The City of Dreams”

In the 1970’s crime and violence in new york city was rampant. NYC was coined “Fear City” Where police corruption and crime ruled.



Violence peaked in the early 1990’s.

There are several theories why violence went down: *Broken Window Policing, Stronger repercussions for law breaking, etc*

Our study will explore whether or not this downward trend is continuing.

Datasets - NYPD Historical Dataset + Population

1

Original Dataset contains 19 features (NYPD)

```
## # A tibble: 6 x 19
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     PRECINCT JURISDICTION_CODE
##          <int> <fct>      <fct>      <fct>       <int>              <int>
## 1     24050482 08/27/2006 05:35:00   BRONX          52                  0
## 2     77673979 03/11/2011 12:03:00   QUEENS        106                  0
## 3    203350417 10/06/2019 01:09:00   BROOKLYN       77                  0
## 4     80584527 09/04/2011 03:35:00   BRONX          40                  0
## 5     90843766 05/27/2013 21:16:00   QUEENS        100                  0
## 6     92393427 09/01/2013 04:17:00   BROOKLYN       67                  0
```

*Transforming and Cleaning Data reduced our data to 7 features

The dataset breaks down each time a shooting incident has occured in NYC since 2006. The dataset also includes information about the victim and perpetrator. We also know whether or not an incident was fatal.

Since 2006:

23,585 Shooting Incidents

4,500 Shooting related deaths

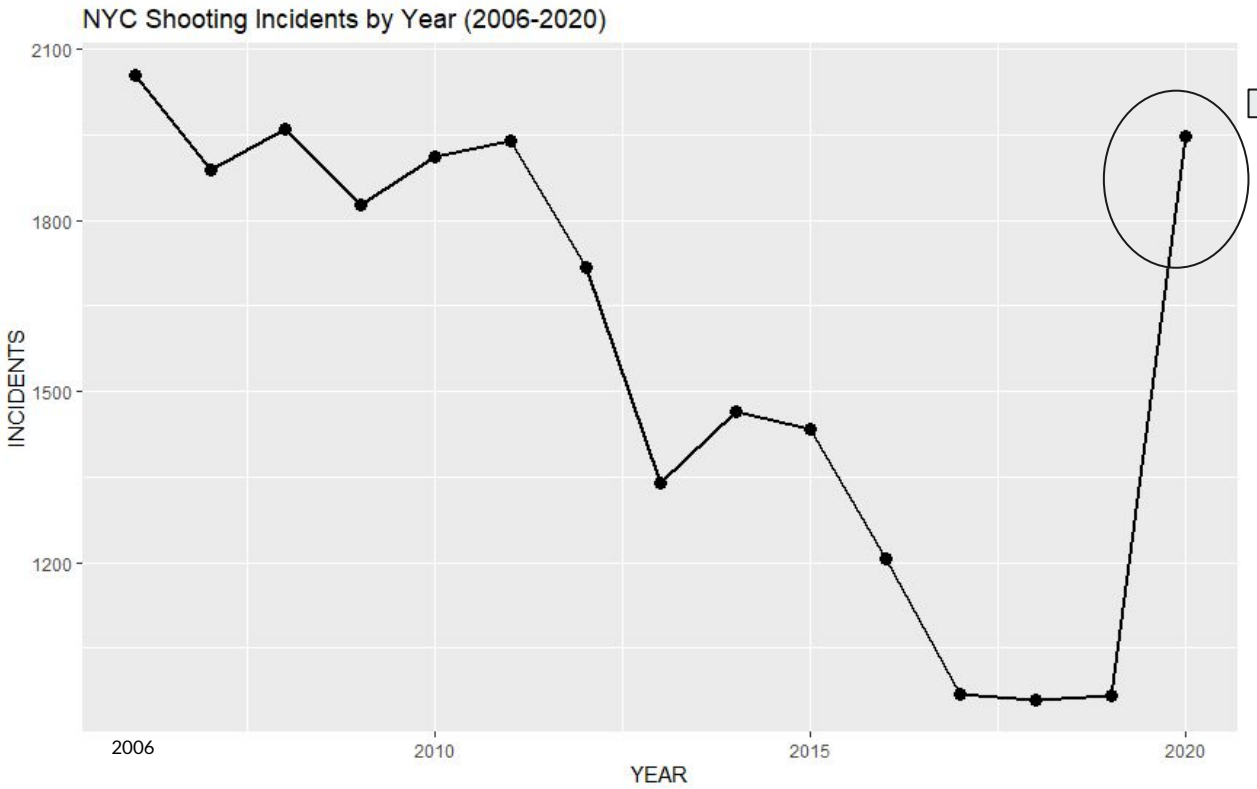
2

Population Dataset (Census)

```
##            BORO POPULATION
## 1        QUEENS    2287000
## 2     MANHATTAN    1632000
## 3         BRONX    1435000
## 4      BROOKLYN    2590000
## 5 STATEN ISLAND     474893
```

The population for each Borough was taken from the census data. This information will be to properly compare between different Boroughs.

Number of shooting incidents has been decreasing yearly



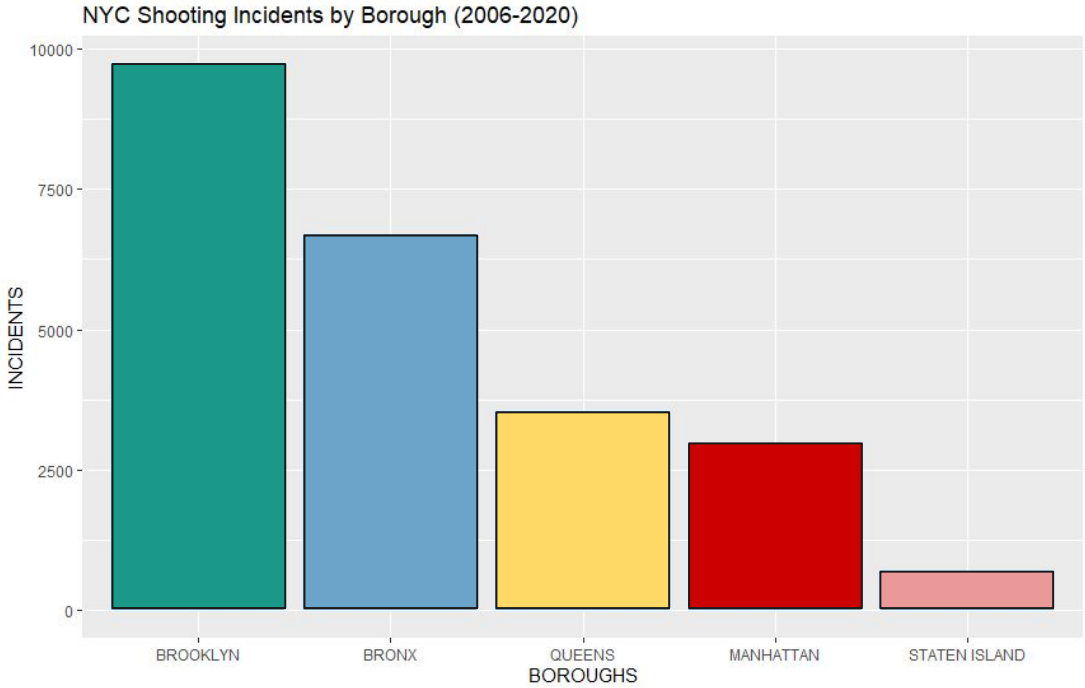
Until Covid...

Since 2006, shooting incidents in NYC were decreasing every year. Before starting the analysis, I thought 2020 would have much fewer incidents because people who remain at home because of COVID. That assumption was proven to be false.

Covid and potentially other factors (2020 was also a hotly contested election year) actually increased the number of shooting incidents.

(see bias section - slide 10 for more info).

Which Boroughs have the highest number of incidents?



NYC is broken into 5 unique boroughs:

From a pure number of incidents perspective, Brooklyn has the highest number of shooting incidents.

Revisiting Shooting Incidents with Population

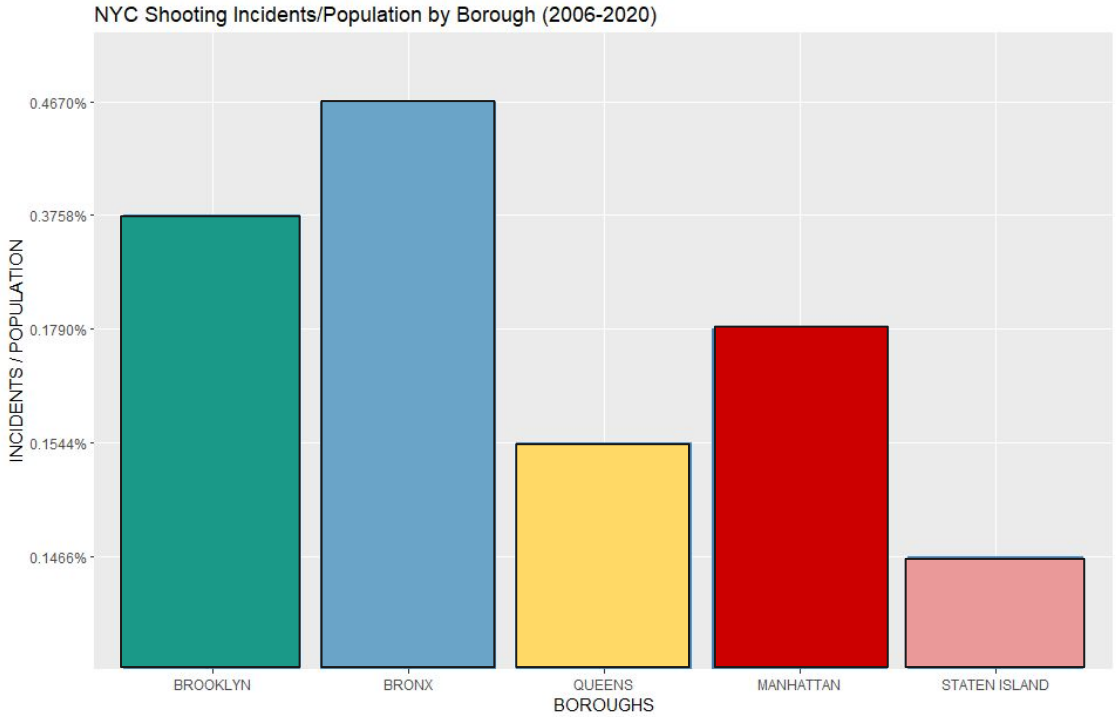
Population data

```
##           BORO POPULATION
## 1       QUEENS    2287000
## 2    MANHATTAN    1632000
## 3        BRONX    1435000
## 4     BROOKLYN    2590000
## 5 STATEN ISLAND     474893
```

Using this information we can create a new metric:

# of Shooting incidents / Population

This metric will allow us to more compare the 5 boroughs since the population discrepancies between each borough is very large

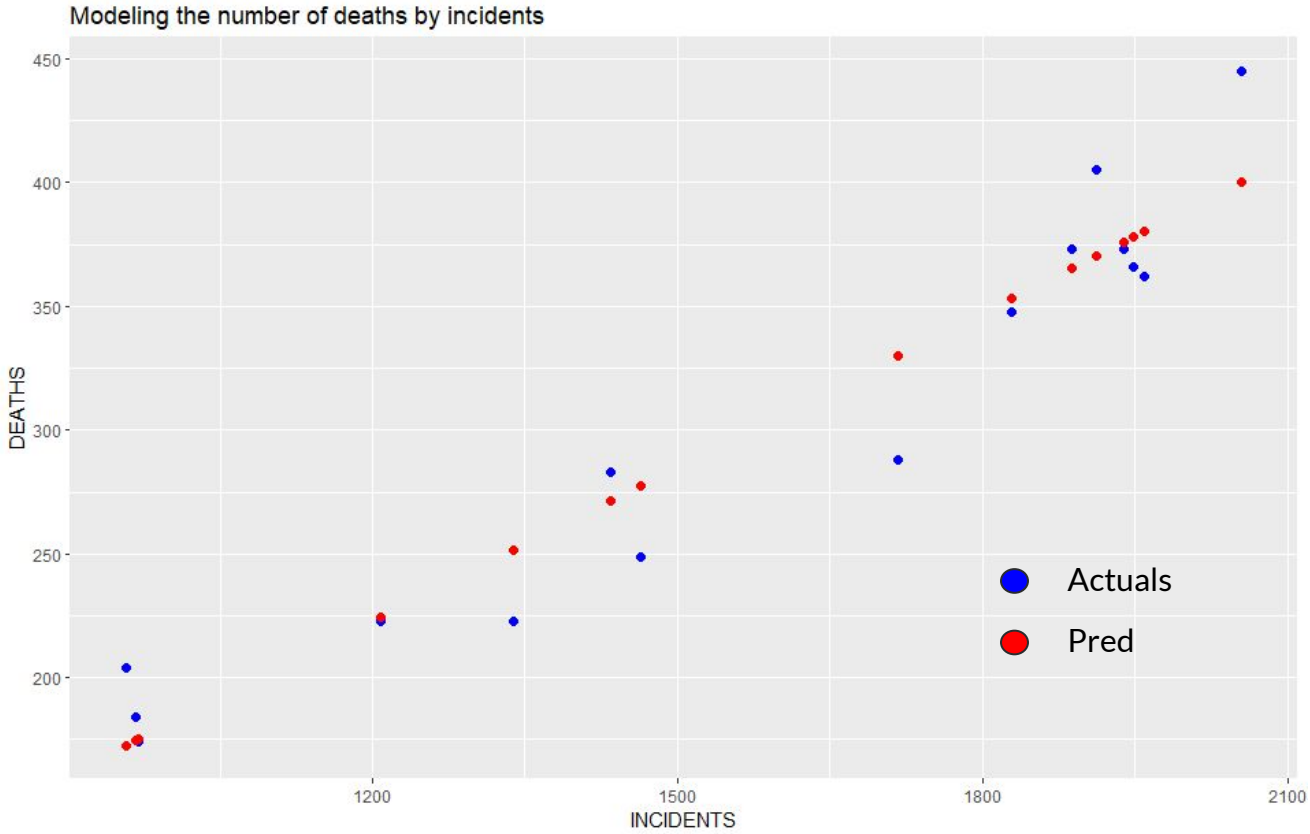


BRONX is the borough with the most incidents/population

Modeling the number of deaths using Linear Regression

```
# A tibble: 15 x 3
    YEAR INCIDENTS DEATHS
   <dbl>     <dbl>  <int>
 1  2006      2055    445
 2  2007      1887    373
 3  2008      1959    362
 4  2009      1828    348
 5  2010      1912    405
 6  2011      1939    373
 7  2012      1717    288
 8  2013      1339    223
 9  2014      1464    249
10  2015      1434    283
11  2016      1208    223
12  2017       970    174
13  2018       958    204
14  2019       967    184
15  2020      1948    366
```

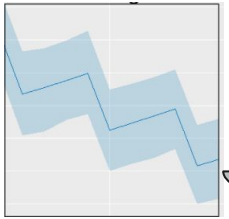
Using our cleaned and aggregated data we will use # of incidents to predict the number of deaths



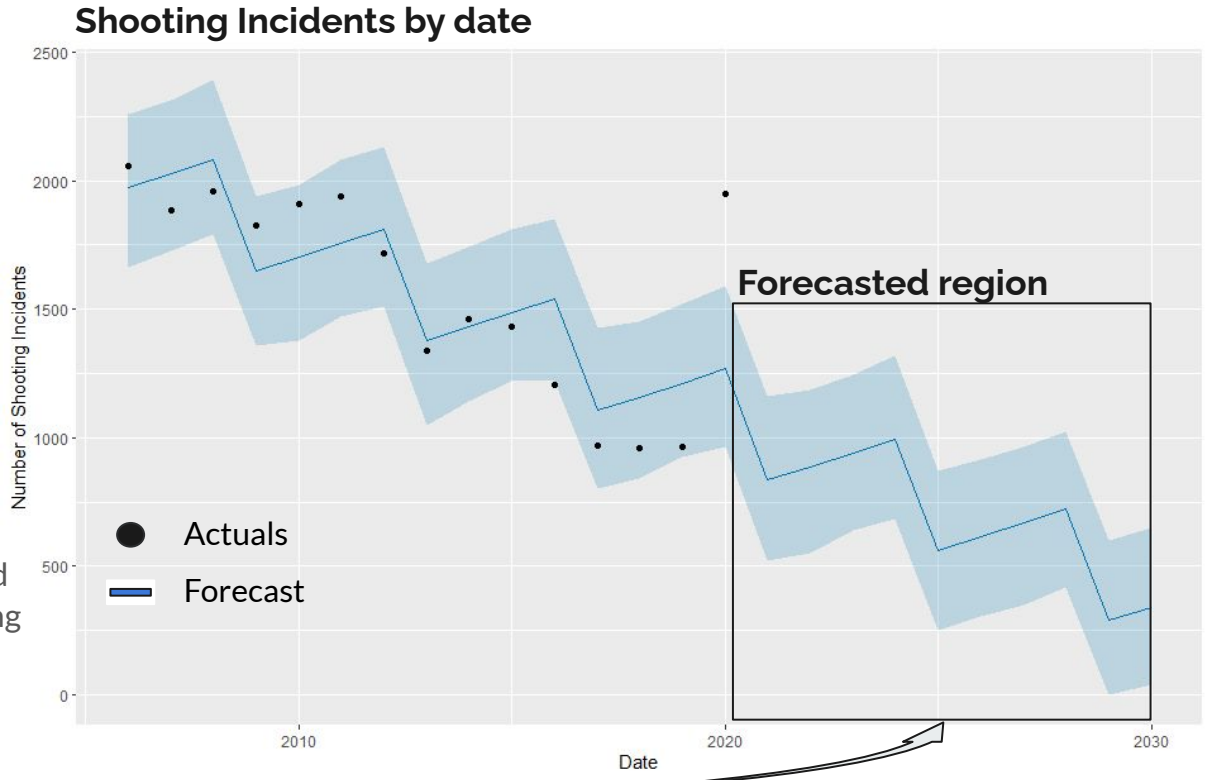
Using Facebook Prophet to forecast # of incidents

Facebook created a specialized package for R that is extremely useful for time-series data.

This package was used to forecast the number of shooting incidents into the future.



Model predicts a continued downward trend in shooting incidents *(it considers 2020 an outlier year)*



Bias Identification

1 Personal Bias

As previously mentioned, one of my personal biases was the belief that there would be less shooting related incidents during 2020 because of COVID.

I assumed that people would be locked down at home. From examining the data (slide 5) that assumption was proven to be false. COVID and potentially other factors *(2020 was also a hotly contested election year)* actually increased the number of shooting incidents.

2 Information Bias

Another bias that I noticed would potentially be the time the shooting incident was recorded. Since these observations are human dependent, I wanted to remove it entirely from my dataset. I also wanted to remove any factors involving race.

There have been several algorithms which ended up being inherently racist. See this MIT article describing police specific algorithms:

Conclusion

We cleaned, examined, visualized, analyzed and modeled the data from the NYPD. It looks like the number of shooting incidents has been steadily decreasing year of year, until we hit 2020. Where it seems that COVID (and potentially other factors) seem to have had a negative impact and has lead to an increase in the number of shooting incidents.

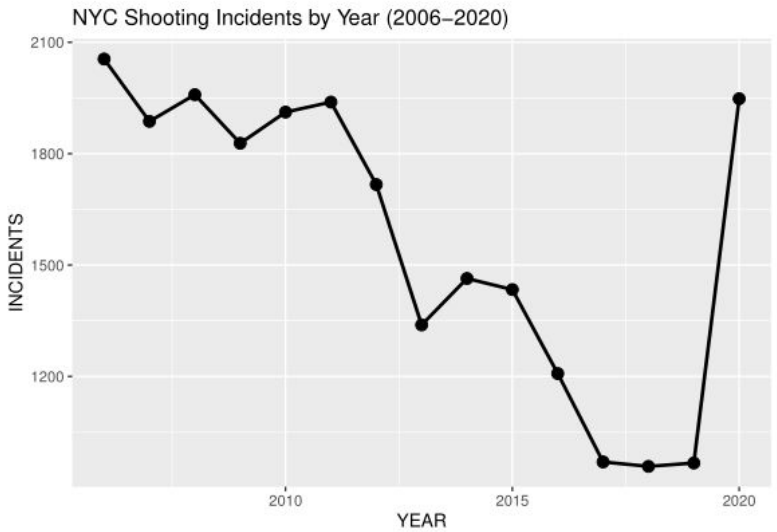
The Bronx seems to be the most violent borough in NYC with the highest number of shooting incidents population.

Finally, we showed that we are able to model the number of shooting related deaths by using the # of shooting incidents and using Facebook Prophet we forecasted a downward trend for the 5 next years.

Backup Slides

Full R Markdown file can be found on Github. The document contains the full analytical breakdown and the code for the visualizations:

```
ggplot(shooting_data_year, aes(x = YEAR, y = INCIDENTS)) + geom_line(size = 1) +
    geom_point(size = 3) + ggtitle("NYC Shooting Incidents by Year (2006-2020)")
```



It looks like the number of shooting incidents were decreasing year over year, until we hit 2020. Before starting the analysis I thought 2020 would have much fewer incidents because of COVID (see bias section further down for more elaboration).

Now that we looked at the incidents over the years, let's create a new data set to see the incidents by NYC borough:

```
# Group by borough
shooting_data_borough <- shooting_data_clean %>%
    group_by(BORO) %>%
    summarize(INCIDENTS = sum(INCIDENTS), DEATHS = sum(STATISTICAL_MURDER_FLAG))
shooting_data_borough
```

```
## # A tibble: 5 x 3
##   BORO          INCIDENTS DEATHS
##   <fct>             <dbl>  <int>
## 1 BRONX              6701   1247
## 2 BROOKLYN           9734   1898
## 3 MANHATTAN          2922    515
## 4 QUEENS             3532    697
## 5 STATEN ISLAND       696    143
```

It looks like some boroughs have more incidents than others, let's take a look: