NYProject

Bhuvana

11/15/2021

This report is being prepared based on tha data that is retrieved from <https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD> on the NY Shooting Incident.

There are mainly information on five boroughs, namely “Queens, Bronx, Manhattan, Staten Island and Brooklyn”, of New York that is collected and an analysis is being done to understand on the Perpetrators who were involved in them, in particular on the females who were involved in each borough

## Importing the Data Set

Data for this Project is being retrieved from here <https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD>

**Initial information that is got is as below.**

## Rows: 23585 Columns: 19

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (10): OCCUR\_DATE, BORO, LOCATION\_DESC, PERP\_AGE\_GROUP, PERP\_SEX, PERP\_R...  
## dbl (7): INCIDENT\_KEY, PRECINCT, JURISDICTION\_CODE, X\_COORD\_CD, Y\_COORD\_CD...  
## lgl (1): STATISTICAL\_MURDER\_FLAG  
## time (1): OCCUR\_TIME

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## # A tibble: 23,585 x 19  
## INCIDENT\_KEY OCCUR\_DATE OCCUR\_TIME BORO PRECINCT JURISDICTION\_CODE  
## <dbl> <chr> <time> <chr> <dbl> <dbl>  
## 1 24050482 08/27/2006 05:35 BRONX 52 0  
## 2 77673979 03/11/2011 12:03 QUEENS 106 0  
## 3 203350417 10/06/2019 01:09 BROOKLYN 77 0  
## 4 80584527 09/04/2011 03:35 BRONX 40 0  
## 5 90843766 05/27/2013 21:16 QUEENS 100 0  
## 6 92393427 09/01/2013 04:17 BROOKLYN 67 0  
## 7 73057167 06/05/2010 21:16 BROOKLYN 77 0  
## 8 211362213 03/20/2020 21:27 BROOKLYN 81 0  
## 9 137564752 07/04/2014 00:25 QUEENS 101 0  
## 10 147024011 10/18/2015 01:33 QUEENS 106 0  
## # ... with 23,575 more rows, and 13 more variables: LOCATION\_DESC <chr>,  
## # STATISTICAL\_MURDER\_FLAG <lgl>, PERP\_AGE\_GROUP <chr>, PERP\_SEX <chr>,  
## # PERP\_RACE <chr>, VIC\_AGE\_GROUP <chr>, VIC\_SEX <chr>, VIC\_RACE <chr>,  
## # X\_COORD\_CD <dbl>, Y\_COORD\_CD <dbl>, Latitude <dbl>, Longitude <dbl>,  
## # Lon\_Lat <chr>

**A snapshot of the data is as below.**

## INCIDENT\_KEY OCCUR\_DATE OCCUR\_TIME BORO   
## Min. : 9953245 Length:23585 Length:23585 Length:23585   
## 1st Qu.: 55322804 Class :character Class1:hms Class :character   
## Median : 83435362 Mode :character Class2:difftime Mode :character   
## Mean :102280741 Mode :numeric   
## 3rd Qu.:150911774   
## Max. :230611229   
##   
## PRECINCT JURISDICTION\_CODE LOCATION\_DESC STATISTICAL\_MURDER\_FLAG  
## Min. : 1.00 Min. :0.000 Length:23585 Mode :logical   
## 1st Qu.: 44.00 1st Qu.:0.000 Class :character FALSE:19085   
## Median : 69.00 Median :0.000 Mode :character TRUE :4500   
## Mean : 66.21 Mean :0.333   
## 3rd Qu.: 81.00 3rd Qu.:0.000   
## Max. :123.00 Max. :2.000   
## NA's :2   
## PERP\_AGE\_GROUP PERP\_SEX PERP\_RACE VIC\_AGE\_GROUP   
## Length:23585 Length:23585 Length:23585 Length:23585   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## VIC\_SEX VIC\_RACE X\_COORD\_CD Y\_COORD\_CD   
## Length:23585 Length:23585 Min. : 914928 Min. :125757   
## Class :character Class :character 1st Qu.: 999925 1st Qu.:182539   
## Mode :character Mode :character Median :1007654 Median :193470   
## Mean :1009379 Mean :207300   
## 3rd Qu.:1016782 3rd Qu.:239163   
## Max. :1066815 Max. :271128   
##   
## Latitude Longitude Lon\_Lat   
## Min. :40.51 Min. :-74.25 Length:23585   
## 1st Qu.:40.67 1st Qu.:-73.94 Class :character   
## Median :40.70 Median :-73.92 Mode :character   
## Mean :40.74 Mean :-73.91   
## 3rd Qu.:40.82 3rd Qu.:-73.88   
## Max. :40.91 Max. :-73.70   
##

## Tidying the data

Data is being tidyed by renaming the Column **BORO** to **BOROUGH** and removing the columns OCCUR\_TIME, X\_COORD\_CD, Y\_COORD\_CD, Lon\_Lat, Latitude, Longitude, JURISDICTION\_CODE

## # A tibble: 23,585 x 12  
## INCIDENT\_KEY OCCUR\_DATE BOROUGH PRECINCT LOCATION\_DESC STATISTICAL\_MURDER\_F~  
## <dbl> <chr> <chr> <dbl> <chr> <lgl>   
## 1 24050482 08/27/2006 BRONX 52 <NA> TRUE   
## 2 77673979 03/11/2011 QUEENS 106 <NA> FALSE   
## 3 203350417 10/06/2019 BROOKLYN 77 <NA> FALSE   
## 4 80584527 09/04/2011 BRONX 40 <NA> FALSE   
## 5 90843766 05/27/2013 QUEENS 100 <NA> FALSE   
## 6 92393427 09/01/2013 BROOKLYN 67 <NA> FALSE   
## 7 73057167 06/05/2010 BROOKLYN 77 <NA> FALSE   
## 8 211362213 03/20/2020 BROOKLYN 81 <NA> FALSE   
## 9 137564752 07/04/2014 QUEENS 101 <NA> FALSE   
## 10 147024011 10/18/2015 QUEENS 106 <NA> FALSE   
## # ... with 23,575 more rows, and 6 more variables: PERP\_AGE\_GROUP <chr>,  
## # PERP\_SEX <chr>, PERP\_RACE <chr>, VIC\_AGE\_GROUP <chr>, VIC\_SEX <chr>,  
## # VIC\_RACE <chr>

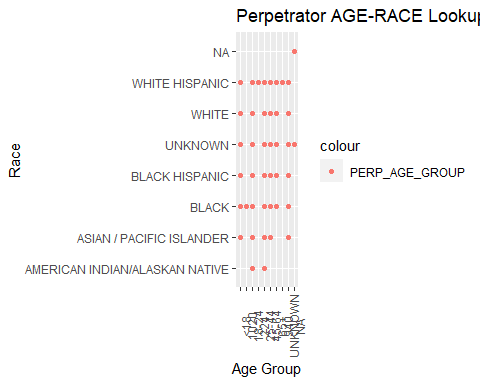
## Transforming the Data

Since we are more interested in knowing about the Perpetrator’s involved, data is being transformed to include only the City , PERP details and VIC details of Age, Sex and Race and STATISTICAL\_MURDER\_FLAG.

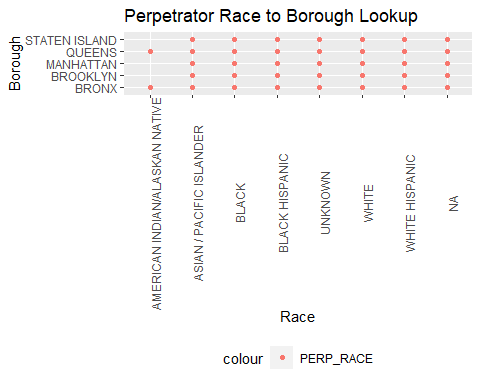
## # A tibble: 23,585 x 8  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_AGE\_GROUP PERP\_SEX PERP\_RACE STATISTICAL\_MURDER\_~ VIC\_AGE\_GROUP  
## <chr> <chr> <chr> <chr> <lgl> <chr>   
## 1 BRONX <NA> <NA> <NA> TRUE 25-44   
## 2 QUEENS <NA> <NA> <NA> FALSE 65+   
## 3 BROOKLYN <NA> <NA> <NA> FALSE 18-24   
## 4 BRONX <NA> <NA> <NA> FALSE <18   
## 5 QUEENS <NA> <NA> <NA> FALSE 18-24   
## 6 BROOKLYN <NA> <NA> <NA> FALSE <18   
## 7 BROOKLYN <NA> <NA> <NA> FALSE <18   
## 8 BROOKLYN <NA> <NA> <NA> FALSE 25-44   
## 9 QUEENS <NA> <NA> <NA> FALSE 18-24   
## 10 QUEENS <NA> <NA> <NA> FALSE 18-24   
## # ... with 23,575 more rows, and 2 more variables: VIC\_SEX <chr>,  
## # VIC\_RACE <chr>

## Visualizing the Data

1. In this first visualization, information about the Perpetrator’s race and the Age Group’s who were involved in is being looked into.



1. In this second visualization, the information with regard to Perpetrator’s race and the Borough, where they are involved in is being looked into.



## Analysing the Data

On analyzing the first visual data, it can be observed that the data of the age group **1020** and **940** could be invalid inputs or they may be Typo while entering the data, of which we may not be sure. Also there are values with **UNKNOWN** and **NA**, where the age group of the people is not being specified. The Race also has data of **NA** and **UNKNOWN**.

From the second visual data, it can be seen that every Borough has reported data of all Race types of people being involved except for “AMERICAN INDIAN/ALASKAN NATIVE” race type, who were involved in only in Bronx and Queens Borough.

A look into the data shows that the data of **NA** and **UNKNOWN** could be valid records. So is for the **1020** and **940** age group data. Therefore, they are retained as received.

## # A tibble: 19,085 x 8  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_AGE\_GROUP PERP\_SEX PERP\_RACE STATISTICAL\_MURDER\_~ VIC\_AGE\_GROUP  
## <chr> <chr> <chr> <chr> <lgl> <chr>   
## 1 QUEENS <NA> <NA> <NA> FALSE 65+   
## 2 BROOKLYN <NA> <NA> <NA> FALSE 18-24   
## 3 BRONX <NA> <NA> <NA> FALSE <18   
## 4 QUEENS <NA> <NA> <NA> FALSE 18-24   
## 5 BROOKLYN <NA> <NA> <NA> FALSE <18   
## 6 BROOKLYN <NA> <NA> <NA> FALSE <18   
## 7 BROOKLYN <NA> <NA> <NA> FALSE 25-44   
## 8 QUEENS <NA> <NA> <NA> FALSE 18-24   
## 9 QUEENS <NA> <NA> <NA> FALSE 18-24   
## 10 BROOKLYN <NA> <NA> <NA> FALSE <18   
## # ... with 19,075 more rows, and 2 more variables: VIC\_SEX <chr>,  
## # VIC\_RACE <chr>

## # A tibble: 4,500 x 8  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_AGE\_GROUP PERP\_SEX PERP\_RACE STATISTICAL\_MURD~ VIC\_AGE\_GROUP  
## <chr> <chr> <chr> <chr> <lgl> <chr>   
## 1 BRONX <NA> <NA> <NA> TRUE 25-44   
## 2 BRONX <NA> <NA> <NA> TRUE 45-64   
## 3 QUEENS <NA> <NA> <NA> TRUE 25-44   
## 4 BRONX 18-24 M BLACK TRUE 25-44   
## 5 QUEENS <NA> <NA> <NA> TRUE 18-24   
## 6 QUEENS <NA> <NA> <NA> TRUE 25-44   
## 7 BRONX <NA> <NA> <NA> TRUE 25-44   
## 8 BROOKLYN 25-44 M BLACK TRUE 25-44   
## 9 MANHATTAN 25-44 M WHITE HISP~ TRUE 45-64   
## 10 BROOKLYN 18-24 M BLACK TRUE 45-64   
## # ... with 4,490 more rows, and 2 more variables: VIC\_SEX <chr>, VIC\_RACE <chr>

Following analysis is being done further down.

* Total number of people who are involved from each Borough
* Number of people of Particular Gender type is checked
* Number of Females involved from each Borough and the Race and Age Group they belong to is looked into.

Henceforth a model is prepared on the Prediction of number of females involved on Total Count

## Total Count from each Borough

## # A tibble: 5 x 2  
## BOROUGH Total\_Count  
## <chr> <int>  
## 1 BRONX 6701  
## 2 BROOKLYN 9734  
## 3 MANHATTAN 2922  
## 4 QUEENS 3532  
## 5 STATEN ISLAND 696

## Count of Particular Gender of people from each Borough

## # A tibble: 20 x 3  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_SEX Gender  
## <chr> <chr> <int>  
## 1 BRONX F 95  
## 2 BRONX M 4047  
## 3 BRONX U 396  
## 4 BRONX <NA> 2163  
## 5 BROOKLYN F 111  
## 6 BROOKLYN M 5079  
## 7 BROOKLYN U 652  
## 8 BROOKLYN <NA> 3892  
## 9 MANHATTAN F 51  
## 10 MANHATTAN M 1792  
## 11 MANHATTAN U 191  
## 12 MANHATTAN <NA> 888  
## 13 QUEENS F 66  
## 14 QUEENS M 2043  
## 15 QUEENS U 231  
## 16 QUEENS <NA> 1192  
## 17 STATEN ISLAND F 12  
## 18 STATEN ISLAND M 529  
## 19 STATEN ISLAND U 29  
## 20 STATEN ISLAND <NA> 126

## Count of Particular Race of people from each Borough

## # A tibble: 37 x 3  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 BRONX AMERICAN INDIAN/ALASKAN NATIVE 1  
## 2 BRONX ASIAN / PACIFIC ISLANDER 28  
## 3 BRONX BLACK 2480  
## 4 BRONX BLACK HISPANIC 531  
## 5 BRONX UNKNOWN 525  
## 6 BRONX WHITE 47  
## 7 BRONX WHITE HISPANIC 926  
## 8 BRONX <NA> 2163  
## 9 BROOKLYN ASIAN / PACIFIC ISLANDER 33  
## 10 BROOKLYN BLACK 4362  
## # ... with 27 more rows

## Count of Females Involved from each Borough

## # A tibble: 5 x 3  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_SEX Gender  
## <chr> <chr> <int>  
## 1 BRONX F 95  
## 2 BROOKLYN F 111  
## 3 MANHATTAN F 51  
## 4 QUEENS F 66  
## 5 STATEN ISLAND F 12

## Joining, by = "BOROUGH"

## # A tibble: 5 x 4  
## BOROUGH Total\_Count PERP\_SEX Gender  
## <chr> <int> <chr> <int>  
## 1 BRONX 6701 F 95  
## 2 BROOKLYN 9734 F 111  
## 3 MANHATTAN 2922 F 51  
## 4 QUEENS 3532 F 66  
## 5 STATEN ISLAND 696 F 12

## Count on Females involved with regard to their Race and Age Group

## # A tibble: 23 x 3  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 BRONX AMERICAN INDIAN/ALASKAN NATIVE 1  
## 2 BRONX BLACK 54  
## 3 BRONX BLACK HISPANIC 8  
## 4 BRONX UNKNOWN 3  
## 5 BRONX WHITE 2  
## 6 BRONX WHITE HISPANIC 27  
## 7 BROOKLYN ASIAN / PACIFIC ISLANDER 1  
## 8 BROOKLYN BLACK 87  
## 9 BROOKLYN BLACK HISPANIC 5  
## 10 BROOKLYN UNKNOWN 2  
## # ... with 13 more rows

## # A tibble: 23 x 3  
## # Groups: BOROUGH [5]  
## BOROUGH PERP\_AGE\_GROUP Group  
## <chr> <chr> <int>  
## 1 BRONX <18 7  
## 2 BRONX 18-24 32  
## 3 BRONX 25-44 45  
## 4 BRONX 45-64 4  
## 5 BRONX UNKNOWN 7  
## 6 BROOKLYN <18 14  
## 7 BROOKLYN 18-24 40  
## 8 BROOKLYN 25-44 44  
## 9 BROOKLYN 45-64 7  
## 10 BROOKLYN UNKNOWN 6  
## # ... with 13 more rows

## Count on Females involved specific to each Borough with regard to Race and Age Group

## # A tibble: 6 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 BRONX AMERICAN INDIAN/ALASKAN NATIVE 1  
## 2 BRONX BLACK 54  
## 3 BRONX BLACK HISPANIC 8  
## 4 BRONX UNKNOWN 3  
## 5 BRONX WHITE 2  
## 6 BRONX WHITE HISPANIC 27

## # A tibble: 4 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 QUEENS BLACK 36  
## 2 QUEENS BLACK HISPANIC 3  
## 3 QUEENS WHITE 4  
## 4 QUEENS WHITE HISPANIC 23

## # A tibble: 3 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 MANHATTAN BLACK 41  
## 2 MANHATTAN BLACK HISPANIC 6  
## 3 MANHATTAN WHITE HISPANIC 4

## # A tibble: 4 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 STATEN ISLAND BLACK 4  
## 2 STATEN ISLAND BLACK HISPANIC 1  
## 3 STATEN ISLAND WHITE 6  
## 4 STATEN ISLAND WHITE HISPANIC 1

## # A tibble: 6 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_RACE Race  
## <chr> <chr> <int>  
## 1 BROOKLYN ASIAN / PACIFIC ISLANDER 1  
## 2 BROOKLYN BLACK 87  
## 3 BROOKLYN BLACK HISPANIC 5  
## 4 BROOKLYN UNKNOWN 2  
## 5 BROOKLYN WHITE 3  
## 6 BROOKLYN WHITE HISPANIC 13

## # A tibble: 5 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_AGE\_GROUP Group  
## <chr> <chr> <int>  
## 1 BRONX <18 7  
## 2 BRONX 18-24 32  
## 3 BRONX 25-44 45  
## 4 BRONX 45-64 4  
## 5 BRONX UNKNOWN 7

## # A tibble: 6 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_AGE\_GROUP Group  
## <chr> <chr> <int>  
## 1 QUEENS <18 6  
## 2 QUEENS 18-24 36  
## 3 QUEENS 25-44 17  
## 4 QUEENS 45-64 3  
## 5 QUEENS 65+ 1  
## 6 QUEENS UNKNOWN 3

## # A tibble: 5 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_AGE\_GROUP Group  
## <chr> <chr> <int>  
## 1 MANHATTAN <18 7  
## 2 MANHATTAN 18-24 14  
## 3 MANHATTAN 25-44 25  
## 4 MANHATTAN 45-64 3  
## 5 MANHATTAN UNKNOWN 2

## # A tibble: 2 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_AGE\_GROUP Group  
## <chr> <chr> <int>  
## 1 STATEN ISLAND 18-24 4  
## 2 STATEN ISLAND 25-44 8

## # A tibble: 5 x 3  
## # Groups: BOROUGH [1]  
## BOROUGH PERP\_AGE\_GROUP Group  
## <chr> <chr> <int>  
## 1 BROOKLYN <18 14  
## 2 BROOKLYN 18-24 40  
## 3 BROOKLYN 25-44 44  
## 4 BROOKLYN 45-64 7  
## 5 BROOKLYN UNKNOWN 6

It can be interpreted that the Female Perpetrators belonging to the race of Black and White Hispanic are mostly involved and people of Age Group’s 18-24 and 25-44, in short, 18-45 age group people are responsible for the acts.

Although other race types of people too account in the Borough’s, their numbers are comparatively less. The issue on race and their impact is a sensitive topic, but still it raises a few questions.

* Why are particular type of people responsible
* What measures could be taken to mitigate their adverse impact
* Is there any specific reasons for their act

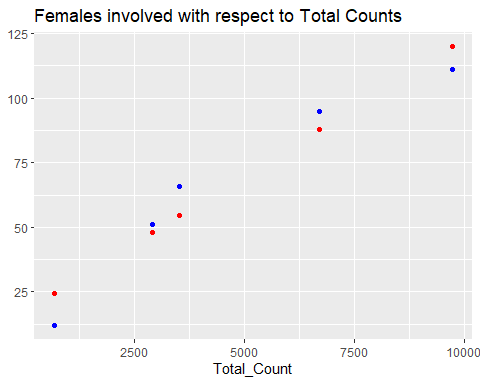
Age group’s between 18-24 and 24-45 also would be raising the same questions.

## Modelling the Data

A model is being predicted on females involved with regard to total counts.

##   
## Call:  
## lm(formula = Gender ~ Total\_Count, data = femaleGenderJoin)  
##   
## Residuals:  
## 1 2 3 4 5   
## 7.033 -9.020 2.970 11.523 -12.506   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.150391 9.608457 1.785 0.17226   
## Total\_Count 0.010568 0.001692 6.244 0.00829 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.96 on 3 degrees of freedom  
## Multiple R-squared: 0.9286, Adjusted R-squared: 0.9047   
## F-statistic: 38.99 on 1 and 3 DF, p-value: 0.008286

## # A tibble: 5 x 5  
## BOROUGH Total\_Count PERP\_SEX Gender pred  
## <chr> <int> <chr> <int> <dbl>  
## 1 BRONX 6701 F 95 88.0  
## 2 BROOKLYN 9734 F 111 120.   
## 3 MANHATTAN 2922 F 51 48.0  
## 4 QUEENS 3532 F 66 54.5  
## 5 STATEN ISLAND 696 F 12 24.5



It can be observed that the predictive values in red is in line with the actual values in blue that we have for each borough of the perpetrators impact.

## Conclusion

From the analysis that is being done above on the Female Perpetrators, it can be seen that the Borough of Brooklyn is being affected highly. Staten Island looks to be far less in number compared to the other Borough’s. Though there are different race types of people who were associated and people of different age groups, it is particular race types who are engaged in them, so is a specific age group people.

There were a couple of factors that went into in the Initial Analysis,

While preparing the report, possible source of bias that was there was to identify

* the group of people who were affected
* the borough that is least affected overall

This was thought to understand on the borough of New york, if moving over there.

Then opted to look out for other options that provided a little general specific insight

* on a particular gender who have committed

The questions that have come out from this analysis are :

* would the data on other gender types also be similar
* are the same age group people related, to account for in particular borough’s

This report generation has helped me to understand the details on female perpetrators involved in NY.