

# Random Forest Analysis on Felony Conviction from the New York Police Department Crime Statistics

Lincy Chen

5/16/2022

## Introduction

Discussions concerning the treatment of law enforcement towards protected classes have been a frequent subject of contention in recent years. One notable example includes the May 2020 murder of George Floyd, an unarmed 46-year-old African-American male, by former police officer Derek Chauvin in Minneapolis. Movements such as Black Lives Matter have been driven by similar inequitable displays of power by police against black people. While their core relies on race being a factor in determining the severity of arrests, many testify against this notion, claiming that All Lives Matter, and that race is a negligible contributor to arrests.

Previous literature on the topic have explored survey opinions on the criminal justice system. A Pew Research Center article, published in June 2020, visualized the political standings of white views on the unethical treatment of blacks. More right-leaning whites argued that black people were not treated unfairly in comparison to more left-leaning whites<sup>3</sup>. Another 2016 survey by the Pew can be linked to this study; it concluded that the majority of officers believed fatal-encounters with blacks to be isolated incidents<sup>4</sup>.

In this paper, I will employ data science methods to achieve a deeper understanding of whether or not a pattern exists among the demographics of arrests made by NYPD. My data set<sup>1</sup> of interest was initially found on Kaggle and compiled to provoke thoughtful and fact-based analysis of police violence and racial equity in the United States. Its source<sup>2</sup> belongs to the New York Police Department's effort to increase transparency and accountability within its police division. The original data set contains almost 5,012,956 observations and 19 features. At first glance, the appearance of missing data seems sparse. However, past revisions of this project have deemed it necessary to be wary of the presence of 'dirty data'. My approach to addressing these issues are expanded upon in the Data Wrangling section of this paper.

I selected to analyze the following qualitative features:

- description of the offense: responses varied in the interest of adequately detailing specific crimes
- level the offense was classified as: felony (F), violation (V), misdemeanor (M), infraction (I)
- the date of the arrest: converted to an object of class `Date` in the form of YYYY-MM-DD
- borough of the arrest: Manhattan (M), Staten Island (S), Bronx (B), Queens (Q), Brooklyn (K)
- age of the perpetrator: sorted into the following groups (<18, 18-24, 25-44, 45-64, 65+)
- race of the perpetrator: categorized by the following (BLACK, WHITE HISPANIC, WHITE, BLACK HISPANIC, ASIAN / PACIFIC ISLANDER, AMERICAN INDIAN/ALASKAN NATIVE, UNKNOWN)
- sex of the perpetrator: Male (M), Female (F)
- jurisdiction code: grounds under which the officer made the arrest - Patrol (0), Transit (1), Housing (2), 3+ (outside jurisdiction)

These variables will help me answer the following research question:

## What are the main factors associated with a felony being committed in the New York City area?

The following sections will detail my reasoning behind methods I chose for analysis, my analysis of results, a discussion delineating potential explorations of related data sets and methods, and an appendix containing the code that led to my findings.

## Methods

### Data Wrangling

To facilitate the deployment of a random forest, I dropped any rows containing NAs as shown in Section III of the Appendix. I was not concerned with the possibility of increasing variance and error as a result of this change since only 0.8987911% of the data were NAs. Additionally, I dropped rows containing anomalies in my data. For example, there was one instance of an individual with a recorded age of 1042.

As referenced above, the original data set contained 19 features. I regarded features describing the longitude, latitude, x-coordinate, y-coordinate, to be irrelevant and redundant towards this project's goals, so I chose not to include them in preparing my data set for model building. I also deemed it necessary to rename certain features from the original data set in the interest of following a standardized convention.

Considering the limitations of R and my 2020 MacBook Air, processing and debugging a random forest with a data set of almost 5 million observations proved to be computationally taxing. The data set I used to apply the model was limited to a calculation of the top 50 offenses committed based on the original cleaned data set's counts. The result was still a reasonably large data set of 99,608 observations.

In designing the Shiny application, I converted the string date variable to data that R could recognize as calendar dates. Using the processed data frames from previous cleaning tasks, I created new ones containing counts of offense levels based on attributes such as `rac` and `date`. I also replaced category labels within `sex`, `offenselevel`, and `location` to enhance interpretability of the application.

Lastly, I created a new column for my model's predictor variable—whether a felony was committed or not. I distinguished `crime_dummies` to be a factor with 2 levels. 0 indicative of no felony committed and 1 indicative of a felony committed. This is in part due to the fact that a felony is regarded as a crime of high seriousness whereas the other categories of crimes are of less serious level.

*Referenced code can be viewed in Section III of the Appendix.*

### Model Building

To help answer the proposed research question, I employed a random forest classifier as I wanted to see how important certain features would be in relation to others through prediction. For example, if race emerged at the top of the Gini coefficient output, one could reasonably infer that race was a distinguishing factor in determining whether or not a felony was committed. This would imply a racial bias towards the classification of a crime being a felony by police. I favored a random forest algorithm because I wanted to prioritize accuracy considering the sensitive context of this investigation. I also deemed the large nature of my data set to be appropriate for random forest classification. Given my curiosity in analyzing whether or not an association is present, I prioritized using a predictive model over an explanatory model because of the categorical and descriptive nature of my features.

I used a 80/20 training/test set split. Within implementation of model, I tried 2 variables at each split (`mtry = 2`) since Hastie, Tibshirani, and Friedman (2009) recommend using  $\sqrt{p}$  variables when dealing with classification trees. 201 was arbitrarily chosen to satisfy using a reasonably large odd number, reflective of the needs of this large data set (`trees = 201`).

My methodology was primarily derived from Chapter 11 of *Modern Data Science with R*.

*Referenced code can be viewed in Section IV of the Appendix.*

## Exploratory Data Analysis

I made preliminary charts to illustrate the selected predictors I used against the classification of whether the associated crime was a felony or not. They are primarily in bar chart format since the data is of qualitative behavior. I also found it visually appealing in ways where taller bars would easily imply a more disproportionate nature against other characteristics. I calculated proportions instead of numbers for ease of comparability between my dummy variable and ethical data reasons. To elaborate, many far-right outlets will manipulate graphics to incriminate a certain demographic when in actuality, these groups faced inequity in the form of racial profiling. *Referenced code can be viewed in Section V of the Appendix.*

To assist in exploratory data analysis, I developed an interactive Shiny application based on content found in *Modern Data Science for R*'s Chapter 14.

*Referenced code can be viewed in Appendix, V, A*

My program displays time series graphs of the number of crimes committed by day. The user is able to choose specific time frames and analyze the levels of arrests made over time. I considered this display to be useful for exploratory data analysis as it could spot whether or not there is an imbalance in data based on different years, which could possibly impact results in case there was a time when the NYPD was not as well-funded and did not make as many arrests.

*Referenced code can be viewed in Appendix, V, B*

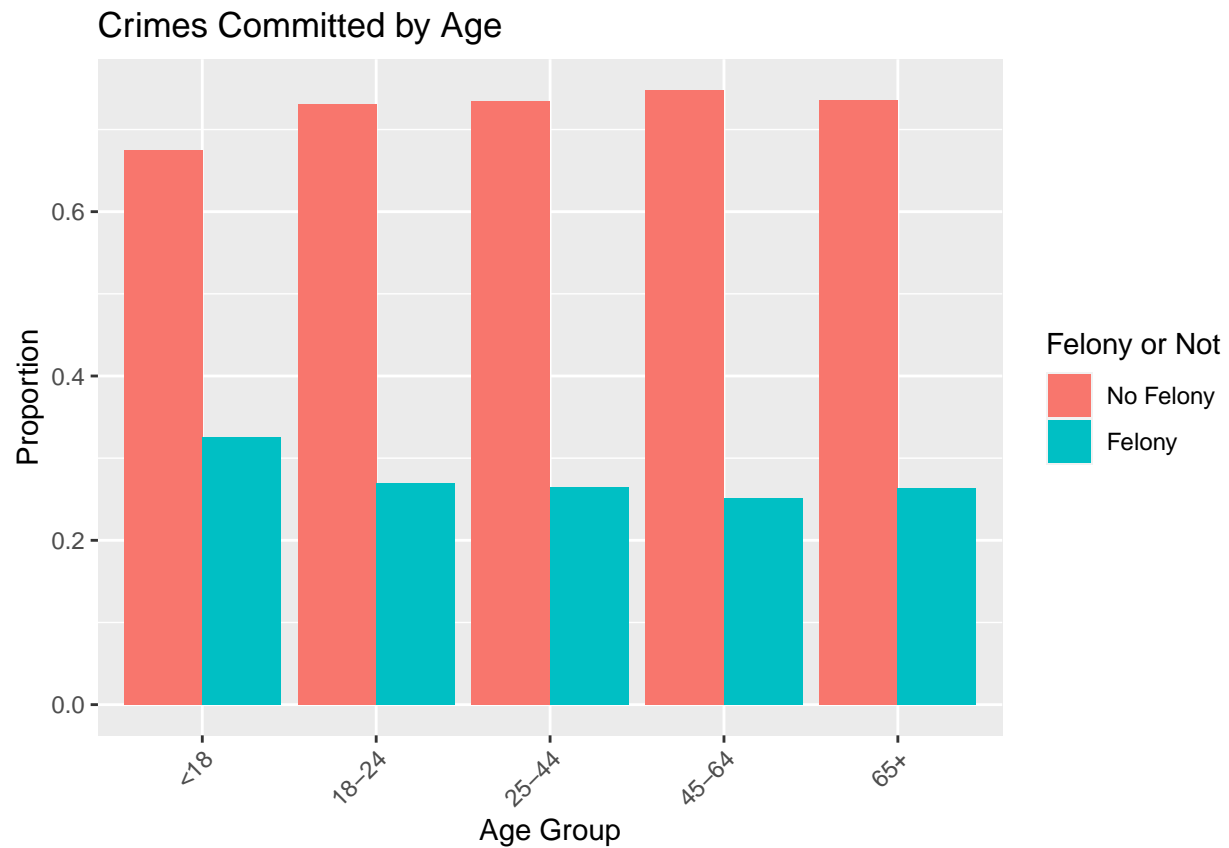
I was appealed by the intuitive nature of the 'Search' option supplied by the DataTables package. I thought that its capabilities would be particularly useful given the immense size and unique offense descriptions provided by the data set. Having a queryable tool partitioned by demographic features such as location and levels of offense also made sense to me as I was curious about the balance of data between certain characteristics.

*Referenced code can be viewed in Appendix, V, C*

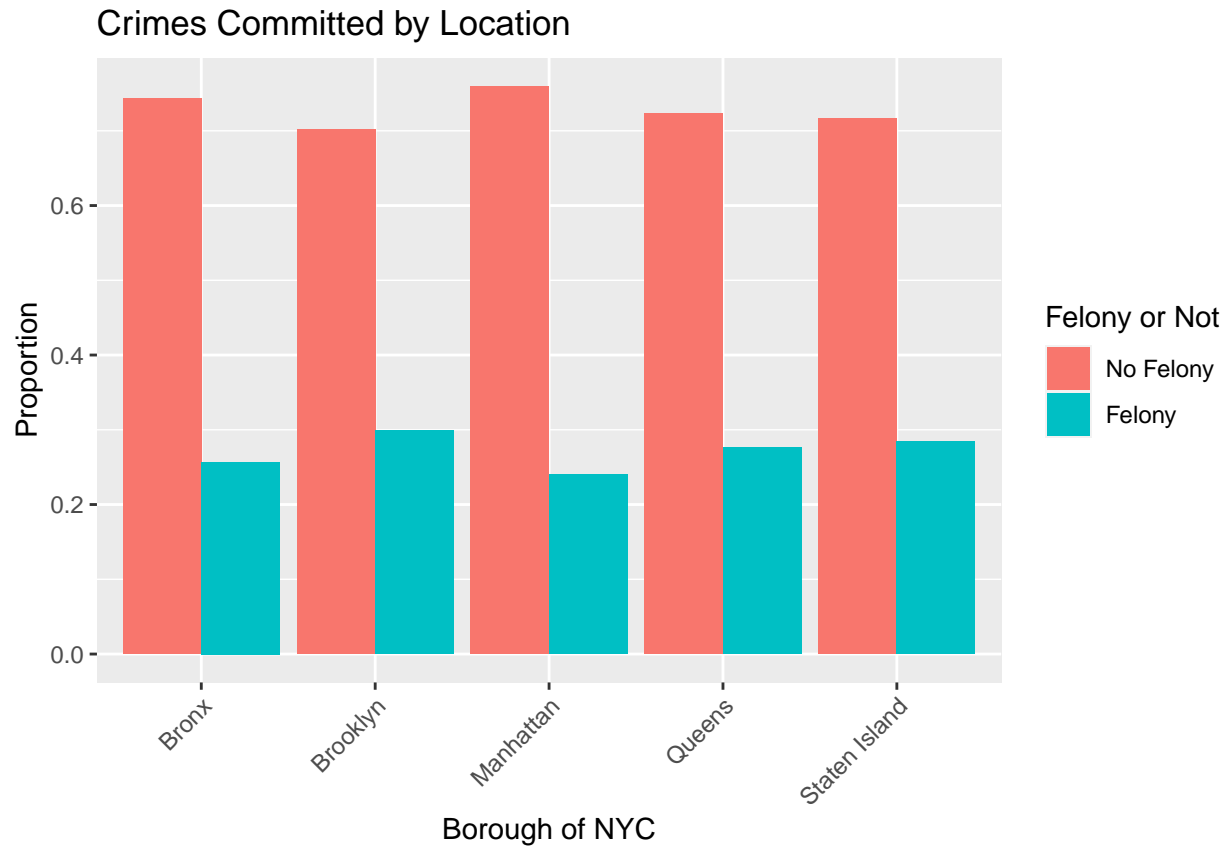
For my third application, I wanted to create an easy viewing tool for the distribution of crimes offenses grouped by race. The user has the option to select the level of offense they'd wish to display and the race. The sum of these counts are displayed on the y-axis with the preferred racial groupings on the x. This way, it would be easy to visualize whether or not one group is disproportionately targeted for higher offense levels than others.

## Results

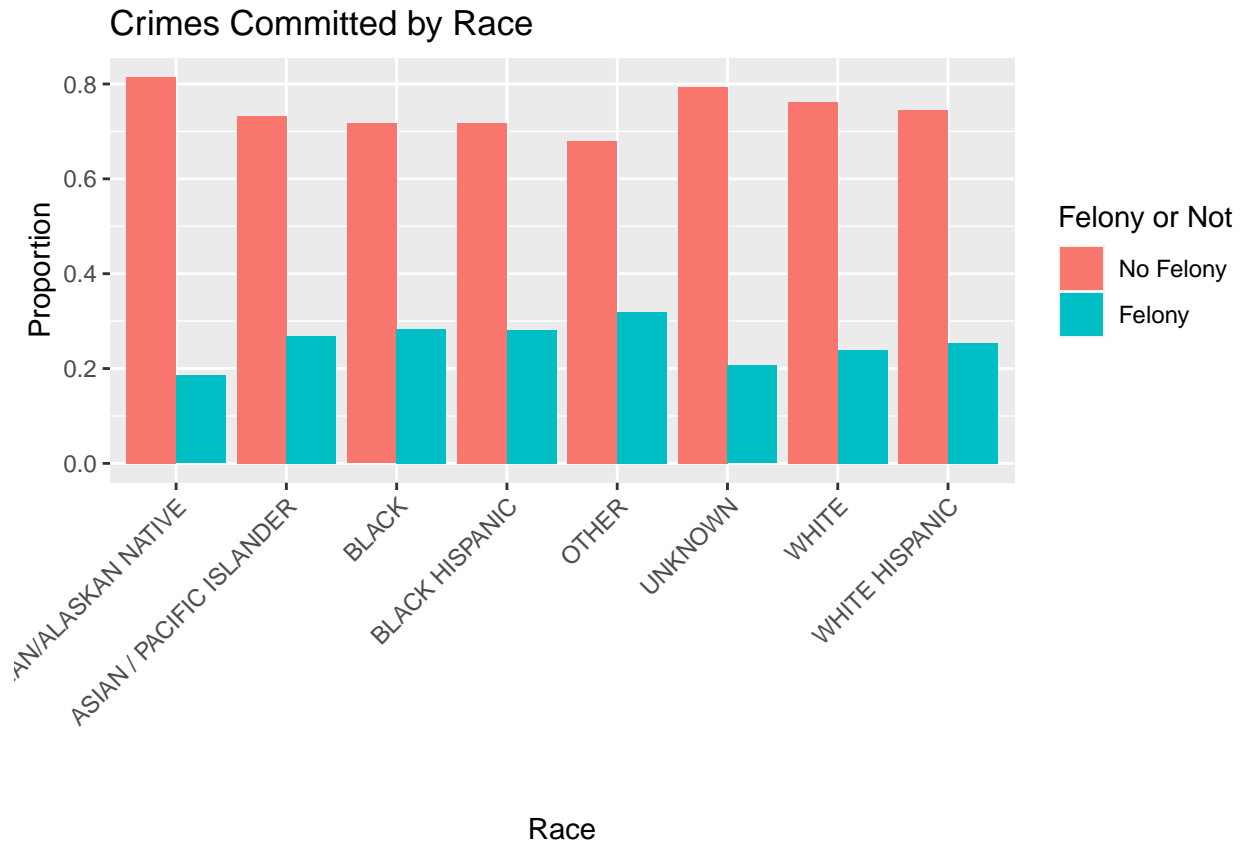
## Exploratory Data Analysis



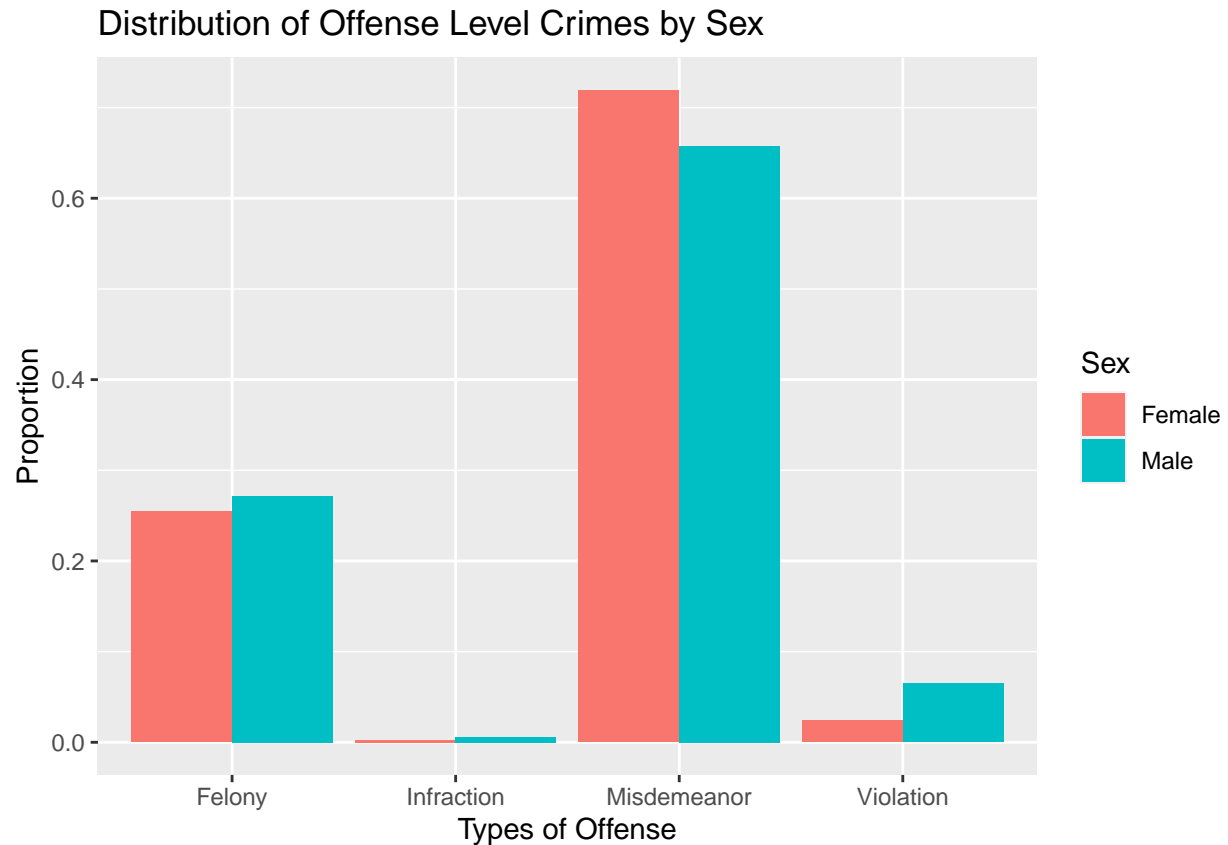
This visualization compares proportion of felony crimes between each age group. On the x-axis, age groups are listed from <18 to 65+. On the y-axis, proportion of crimes that are considered felonies are measured. It is noteworthy to see that the proportion of felonies committed is highest among perpetrators less than 18 years of age as evidenced by the tallest bar for the **Felony** label. On the other hand, 45-64 year olds are less likely to commit a felony as it has the lowest bar.



This visualization compares proportion of felony crimes between each borough of NYC. On the x-axis, locations recorded of arrests are listed from the Bronx, Brooklyn, Manhattan, Queens, and Staten Islands. On the y-axis, proportion of crimes that are considered felonies are measured. It is noteworthy to see that the proportion of felonies committed is highest among perpetrators in Brooklyn as evidenced by the tallest bar for the **Felony** label. On the other hand, Manhattan crimes are less likely to be of a felony as it has the lowest bar.



This visualization compares proportion of felony crimes between each race group. On the x-axis, races are listed from the following: Black, Black Hispanic, White, White Hispanic, Asian, American Indian, other, and unknown. On the y-axis, the proportion of crimes that are considered felonies are measured out of each group. It is noteworthy to see that the proportion of felonies committed is highest among perpetrators in 'Other' races as evidenced by the tallest bar for the **Felony** label. On the other hand, American Indians/Alaskan Natives are less likely to commit a felony as it has the lowest bar from this data set.



This visualization compares proportion of males vs females between each level of offense grouping. On the x-axis, offense levels are listed from the following: Felony, Infraction, Misdemeanor, and Violation. On the y-axis, the proportion of crimes from males and females are measured out of each group. It is noteworthy to see that the proportion of felonies committed is about the same for males and females as evidenced by their similar heights for the **Felony** label. On the other hand, women are more likely to commit a misdemeanor as witnessed by their somewhat higher height in relation to the proportion for male misdemeanors.

## Random Forest Results

```
mod_forest
```

```
## parsnip model object
##
##
## Call:
## randomForest(x = maybe_data_frame(x), y = y, ntree = ~201, mtry = min_cols(~2,      x))
##           Type of random forest: classification
##           Number of trees: 201
## No. of variables tried at each split: 2
##
##           OOB estimate of  error rate: 26.84%
## Confusion matrix:
##           0    1 class.error
## 0 58179  70 0.001201737
```

```
## 1 21314 120 0.994401418
```

The confusion matrix generated by the above random forest model can be interpreted in the following manner:

- There are 58179 true negatives, where the model correctly predicted no felony when there really was no felony committed by the perpetrator
- There were 120 true positives, where the model correctly identified a felony occurrence in perpetrators that did commit a felony
- There were 21314 false negatives, where the model predicted that no felony in a crime when in actuality there was a felony committed
- There were 70 false positives, where the model incorrectly predicted that an arrest made was a felony when in actuality there was no felony committed.

The model was able to correctly classify felony arrests more than it incorrectly classified felony arrests.

```
randomForest::importance(mod_forest$fit) %>%  
  as_tibble(rownames = "variables") %>%  
  arrange(desc(MeanDecreaseGini))
```

```
## # A tibble: 5 x 2  
##   variables MeanDecreaseGini  
##   <chr>          <dbl>  
## 1 code           833.  
## 2 race           160.  
## 3 age            155.  
## 4 location       133.  
## 5 sex            65.8
```

The resulting matrix represents the level of importance each predictor has in determining whether or not a felony was committed. As evidenced by the highest Gini-values, jurisdiction code followed by race, age, and location all play an important role in the random forest algorithm.

## Discussion

At face value, the matrix reports jurisdiction code to be a significantly more important predictor than race, age, and location as evidenced by its high value of 832 in relation to the others' near 150 levels. Though, it would be ill-advised to believe that race, age, and location do not play a significant role in determining whether a felony was committed due to their lower Gini-coefficient values. Rather, considering the meaning of jurisdiction codes in real life, it can be implied that a felony would be more likely to be committed when an officer is within jurisdiction as they are defined to be crimes of more serious level. Race and location may be related with jurisdiction code in the lens where police are more concentrated in certain boroughs more than others. This confounding factor should be taken into consideration since it can be intuitively said that people of similar races tend to live in same areas. Furthermore, previous revisions of the random forest model without jurisdiction code have highlighted the importance of race out of the 4. Therefore, results imply inequality in police conviction of felonies among demographics such as race, location, and age. On the other hand, sex cannot be said to be a useful factor in distinguishing between felonies and regular crimes.

From analyzing the Frequency of Crimes by Race Shiny Application, some of the meaning from the random forest results can be explained. A lot of felonies committed are disproportionately represented by black racial groups. Combined with the model's results, the race's influential role in classification between whether or not a felony is committed comes into view. Future analysis of this topic might delve into studying other demographics such as income level of the perpetrator. A variable like location, could then be used to identify areas of disinvestment and gentrification inside New York City.



# Appendix

## I. Library imports

```
library(tidyverse)
library(devtools)
library(ggthemr)
library(randomForest)
library(yardstick)
library(tidymodels)
library(extrafont)
```

## II. Data collection

```
popo <- read_csv("crimnypd.csv")
glimpse(popo)
```

## III. Data wrangling

```
# original data set overview
summary(popo)
```

```
# missing data removal
na_free_popo <- na.omit(popo)
# subsetting of data for variables used in this project
less_popo_data <- na_free_popo %>%
  select(
    OFNS_DESC,
    ARREST_DATE,
    ARREST_BORO,
    AGE_GROUP,
    PERP_SEX,
    PERP_RACE,
    JURISDICTION_CODE,
    LAW_CAT_CD
  ) %>%
  rename(
    offense = OFNS_DESC,
    crimedate = ARREST_DATE,
    age = AGE_GROUP,
    sex = PERP_SEX,
    race = PERP_RACE,
    code = JURISDICTION_CODE,
    offenselevel = LAW_CAT_CD,
    location = ARREST_BORO
  )

# calculation of top 50 offenses committed
```

```

evenlesstry <- less_popo_data %>%
  group_by(offense) %>%
  summarise(offensefreq = length(offense)) %>%
  arrange(desc(offensefreq)) %>%
  head(50)

top50offense <- unique(evenlesstry$offense)

# subset of the top 50 offenses committed
evenlesspopo <- less_popo_data %>%
  select(
    offense,
    crimdate,
    location,
    age,
    sex,
    race,
    code,
    offenselevel
  ) %>%
  filter(
    offense == top50offense
  )

# conversion of data for shiny app
evenlesspopo$crimdate <- as.Date(evenlesspopo$crimdate, format = "%m/%d/%Y")

# preparation for random forest model
modeldf <- evenlesspopo %>%
  mutate(
    crime_dummies = ifelse(offenselevel == "F", 1, 0),
    crime_dummies = factor(crime_dummies)
  )

# removal of dirty data
modeldf = filter(modeldf, age != "309" & age != "1042" & age != "709" & age != "932")

# cleaning for shiny app
shinydf <- modeldf
# for a more user-friendly experience
shinydf[shinydf$location == "B", "location"] <- "Bronx"
shinydf[shinydf$location == "M", "location"] <- "Manhattan"
shinydf[shinydf$location == "K", "location"] <- "Brooklyn"
shinydf[shinydf$location == "Q", "location"] <- "Queens"
shinydf[shinydf$location == "S", "location"] <- "Staten Island"

shinydf[shinydf$sex == "M", "sex"] <- "Male"
shinydf[shinydf$sex == "F", "sex"] <- "Female"

shinydf[shinydf$offenselevel == "F", "offenselevel"] <- "Felony"
shinydf[shinydf$offenselevel == "M", "offenselevel"] <- "Misdemeanor"
shinydf[shinydf$offenselevel == "I", "offenselevel"] <- "Infraction"
shinydf[shinydf$offenselevel == "V", "offenselevel"] <- "Violation"

# data frame for time series creation

```

```
test <- shinydf %>%
  group_by(crimedate) %>%
  count(offenselevel)
# data frame for racial distribution
bardf <- shinydf %>%
  group_by(race) %>%
  count(offenselevel)
```

## IV. Model building

```
# for a consistent generation of results
set.seed(2000)
# training/test split
crime_parts <- modeldf %>%
  initial_split(prop = 0.8)
train <- crime_parts %>% training()
test <- crime_parts %>% testing()
# variables supplied to the decision tree
form <- as.formula(
  "crime_dummies ~ age + sex + race + location + code"
)
# assembly of one tree
mod_tree <- decision_tree(mode = "classification") %>%
  set_engine("rpart") %>%
  fit(form, data = train)
# evaluation of the model
pred <- train %>%
  select(offenselevel) %>%
  bind_cols(
    predict(mod_tree, new_data = train, type = "class")
  ) %>%
  rename(offenselevel_dtree = .pred_class)
# assembly of the forest with 5 predictors for classification of felony crime or not
mod_forest <- rand_forest(
  mode = "classification",
  mtry = 2,
  trees = 201
) %>%
  set_engine("randomForest") %>%
  fit(form, data = train)
```

Random forest implementation

## V. Exploratory Data Analysis

Preliminary Graphs

```

# code for crime counts vs age graph
modeldf %>%
  count(crime_dummies, age) %>%
  group_by(age) %>%
  mutate(freq = n/sum(n)) %>%
  ggplot(aes(x = age, y = freq, fill = crime_dummies)) +
  geom_bar(stat="identity", position = 'dodge') +
  labs(
    title = "Crimes Committed by Age",
    x = "Age Group",
    y = "Proportion"
  ) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  scale_fill_discrete(name="Felony or Not",
    breaks=c("0", "1"),
    labels=c("No Felony", "Felony"))

# code for crime counts vs race graph
modeldf %>%
  count(crime_dummies, race) %>%
  group_by(race) %>%
  mutate(freq = n/sum(n)) %>%
  ggplot(aes(x = race, y = freq, fill = crime_dummies)) +
  geom_bar(stat="identity", position = 'dodge') +
  labs(
    title = "Crimes Committed by Race",
    x = "Race",
    y = "Proportion"
  ) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  scale_fill_discrete(name="Felony or Not",
    breaks=c("0", "1"),
    labels=c("No Felony", "Felony"))

# code for crime counts vs oocation graph
modeldf %>%
  count(crime_dummies, location) %>%
  group_by(location) %>%
  mutate(freq = n/sum(n)) %>%
  ggplot(aes(x = location, y = freq, fill = crime_dummies)) +
  geom_bar(stat="identity", position = 'dodge') +
  labs(
    title = "Crimes Committed by Location",
    x = "Borough of NYC",
    y = "Proportion"
  ) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  scale_fill_discrete(name="Felony or Not",
    breaks=c("0", "1"),
    labels=c("No Felony", "Felony")) +
  scale_x_discrete(labels = c("B" = "Bronx", "K" = "Brooklyn", "M" = "Manhattan", "Q" = "Queens", "S" = "Staten Island"))

# code for crime counts vs sex graph
evenlesspopo %>%
  count(sex, offenselevel) %>%

```

```

group_by(sex) %>%
mutate(freq = n/sum(n)) %>%
ggplot(aes(x = offenselevel, y = freq, fill = sex)) +
geom_bar(stat="identity", position = 'dodge') +
labs(
  title = "Distribution of Offense Level Crimes by Sex",
  x = "Types of Offense",
  y = "Proportion"
) +
scale_x_discrete(labels = c("F" = "Felony", "I" = "Infraction", "M" = "Misdemeanor", "V" = "Violation"),
scale_fill_discrete(name="Sex",
                     breaks=c("F", "M"),
                     labels=c("Female", "Male"))

```

## A. Interactive Time Series Program to Showcase Crime over Time

```

library(shiny)
library(tidyverse)

# ui.R
ui <- shinyUI(
  bootstrapPage(
    h3("Frequency of Crimes Over Time"),
    dateInput(
      "startdate",
      "Enter starting date:",
      value = "2010-02-02",
      min = "2006-01-01",
      max = "2019-12-30"
    ),
    dateInput(
      "enddate",
      "Enter ending date:",
      value = "2010-09-02",
      min = "2006-01-02",
      max = "2019-12-31"
    ),
    checkboxGroupInput(
      "crimelevel",
      "Level(s) of Offense to display:",
      choices = sort(unique(test$offenselevel)),
      selected = c("Misdemeanor", "Infraction"),
    ),
    plotOutput("plot")
  )
)

# server.R
server <- shinyServer(
  function(input, output) {
    output$plot <- renderPlot({

```

```

ds <- test %>%
  filter(
    crimedate >= input$startdate, crimedate <= input$enddate,
    offenselevel %in% input$crimelevel
  )
ggplot(
  data = ds,
  aes_string(x = "crimedate", y = "n", color = "offenselevel")
) +
  geom_line(size = 2) +
  labs(title = paste("Types of Crimes Committed from ", input$startdate,
    "to", input$enddate), y="Frequency", x = "Date",
    color = "Type of Offense") +
  theme_classic() +
  theme(plot.title = element_text(size = 22))
})
}
)

shinyApp(ui=ui, server=server)

```

## B. Interactive Database Queried by Location, Offense Description, and Offense Level

```

library(tidyverse)
library(shiny)
library(shinybusy)

ui <- fluidPage(
  titlePanel("Crime Explorer"),
  fluidRow(
    # some things take time: this lets users know
    add_busy_spinner(spin = "fading-circle"),
    column(
      4,
      selectInput(inputId = "location",
        label = "Borough:",
        choices = c(
          "ALL",
          unique(as.character(shinydf$location))
        )
      )
    ),
    # display dynamic list of offenses
    column(4, uiOutput("offensecontrols")),
    column(4, uiOutput("offenselevelcontrols"))
  ),
  # Create a new row for the table.
  fluidRow(
    DT::dataTableOutput("table")
  )
)

```

```

)

server <- function(input, output) {

  datasetlocation <- reactive({ # Filter data based on selections
    data <- shinydf %>%
      select(
        offense, crimdate, location,
        age, sex, race, offenselevel
      ) %>%
      distinct()
    req(input$location) # wait until there's a selection
    if (input$location != "ALL") {
      data <- data %>%
        filter(location == input$location)
    }
    data
  })

  datasetoffense <- reactive({ # dynamic list of offenses
    req(input$offense) # wait until list is available
    data <- datasetlocation() %>%
      unique()
    if (input$offense != "ALL") {
      data <- data %>%
        filter(offense == input$offense)
    }
    data
  })

  datasetoffenselevel <- reactive({ # dynamic list of offenses
    req(input$offenselevel) # wait until list is available
    data <- datasetoffense() %>%
      unique()
    if (input$offenselevel != "ALL") {
      data <- data %>%
        filter(offenselevel == input$offenselevel)
    }
    data
  })

  output$table <- DT::renderDataTable(DT::datatable(datasetoffenselevel()))

  output$offensecontrols <- renderUI({
    availablelevels <-
      unique(sort(as.character(datasetlocation())$offense))
    selectInput(
      inputId = "offense",
      label = "Offense Description:",
      choices = c("ALL", availablelevels)
    )
  })
}

```

```

output$offenselevelcontrols <- renderUI({
  availablelevels <-
    unique(sort(as.character(datasetlocation()$offenselevel)))
  selectInput(
    inputId = "offenselevel",
    label = "Offense Level:",
    choices = c("ALL", availablelevels)
  )
})
}

shinyApp(ui = ui, server = server)

```

### C. Interactive Bar Chart Builder by Race and Crime Level

```

library(shiny)
library(tidyverse)

# ui.R
ui <- shinyUI(
  bootstrapPage(
    h3("Frequency of Crimes by Race")
    ,
    checkboxGroupInput(
      "crimelevel",
      "Level(s) of Offense to display:",
      choices = sort(unique(bardf$offenselevel)),
      selected = c("Misdemeanor", "Infraction"),
    ),
    checkboxGroupInput(
      "race",
      "Race to display:",
      choices = sort(unique(bardf$race)),
      selected = c("BLACK", "WHITE", "ASIAN / PACIFIC ISLANDER", "WHITE HISPANIC", "BLACK HISPANIC"),
    ),
    plotOutput("plot")
  )
)

# server.R
server <- shinyServer(
  function(input, output) {
    output$plot <- renderPlot({
      ds <- bardf %>%
        filter(
          offenselevel %in% input$crimelevel,
          race %in% input$race
        )
      ggplot(
        data = ds,

```



```

    aes_string(x = "race", y = "n", fill = "offenselevel")
  ) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Racial Distribution of Crimes in the NYC area",
       y="Frequency", x = "Race", color = "Type of Offense") +
  theme_classic() +
  theme(plot.title = element_text(size = 22))
})
}
)

shinyApp(ui=ui, server=server)

```

## Bibliography

1. <https://www.kaggle.com/datasets/jpmiller/police-violence-racial-equity>
2. <https://www1.nyc.gov/site/nypd/stats/crime-statistics/crime-statistics-landing.page>
3. <https://www.pewresearch.org/fact-tank/2020/06/03/10-things-we-know-about-race-and-policing-in-the-u-s/>
4. <https://www.pewresearch.org/social-trends/2017/01/11/behind-the-badge/>

Chapter 11 of Modern Data Science with R

Chapter 14 of Modern Data Science with R

## Frequency of Crimes Over Time

Enter starting date:

2010-07-01

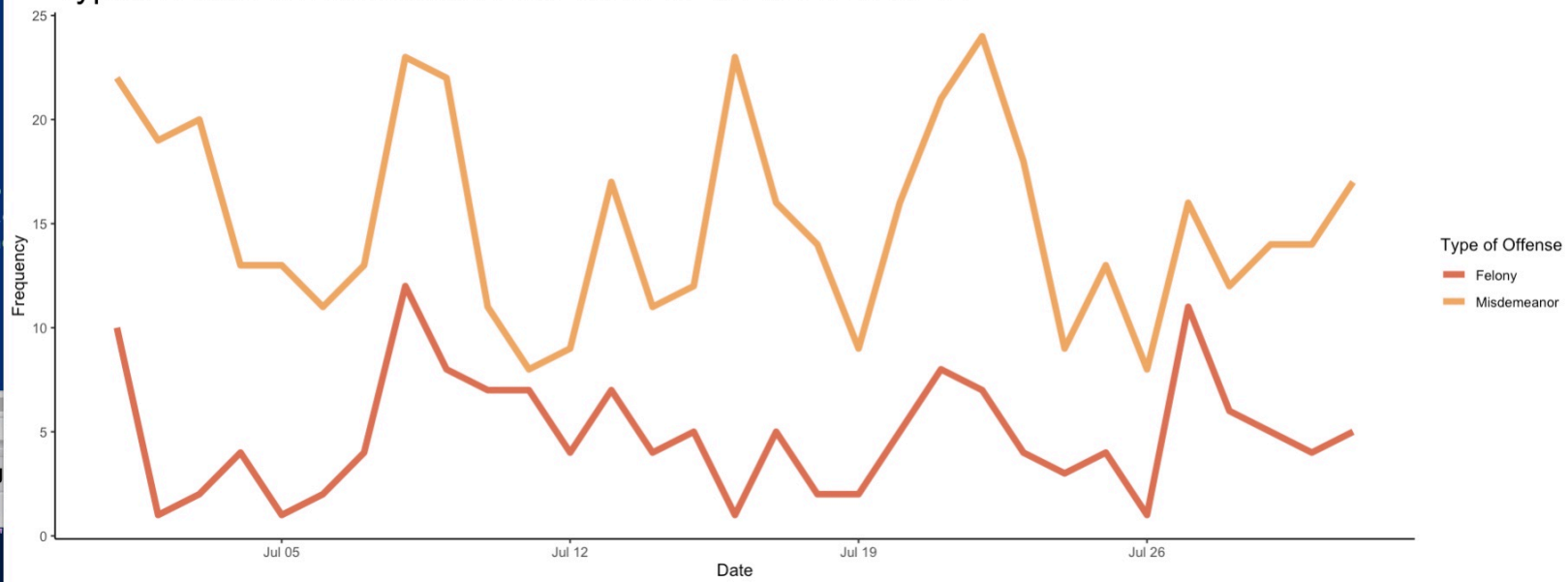
Enter ending date:

2010-07-31

Level(s) of Offense to display:

- ☒ Felony
- ☐ Infraction
- ☒ Misdemeanor
- ☐ Violation

### Types of Crimes Committed from 2010-07-01 to 2010-07-31



# Crime Explorer

**Borough:**  

Queens

**Offense Description:**  

PROSTITUTION & RELATED OFFENSES

**Offense Level:**  

Misdemeanor

Show 

10

 entries

Search:

	offense	crimedate	location	age	sex	race	offenselevel
1	PROSTITUTION & RELATED OFFENSES	2018-10-24	Queens	25-44	Male	ASIAN / PACIFIC ISLANDER	Misdemeanor
2	PROSTITUTION & RELATED OFFENSES	2018-07-14	Queens	25-44	Female	BLACK	Misdemeanor
3	PROSTITUTION & RELATED OFFENSES	2018-08-03	Queens	25-44	Female	WHITE HISPANIC	Misdemeanor
4	PROSTITUTION & RELATED OFFENSES	2018-06-27	Queens	25-44	Female	BLACK	Misdemeanor
5	PROSTITUTION & RELATED OFFENSES	2018-05-25	Queens	45-64	Male	WHITE HISPANIC	Misdemeanor
6	PROSTITUTION & RELATED OFFENSES	2018-01-11	Queens	25-44	Female	ASIAN / PACIFIC ISLANDER	Misdemeanor
7	PROSTITUTION & RELATED OFFENSES	2019-11-16	Queens	25-44	Male	WHITE HISPANIC	Misdemeanor
8	PROSTITUTION & RELATED OFFENSES	2018-05-05	Queens	25-44	Male	WHITE HISPANIC	Misdemeanor
9	PROSTITUTION & RELATED OFFENSES	2018-04-26	Queens	25-44	Male	ASIAN / PACIFIC ISLANDER	Misdemeanor
10	PROSTITUTION & RELATED OFFENSES	2018-04-21	Queens	18-24	Male	BLACK HISPANIC	Misdemeanor

# Frequency of Crimes by Race

Level(s) of Offense to display:

- ☒ Felony
- ☐ Infraction
- ☒ Misdemeanor
- ☒ Violation

Race to display:

- ☐ AMERICAN INDIAN/ALASKAN NATIVE
- ☒ ASIAN / PACIFIC ISLANDER
- ☒ BLACK
- ☒ BLACK HISPANIC
- ☐ OTHER
- ☐ UNKNOWN
- ☒ WHITE
- ☒ WHITE HISPANIC

Racial Distribution of Crimes in the NYC area

