NOLA Police Use of Force

Christopher Stephens

2023-09-15

Introduction

The NOPD_Use_of_Force_Incidents data set is a snapshot of a data set taken in May of 2023 from the api on the New Orleans city record archive. The Data set consists of 33 column features with 3314 rows. Each entry is a specific incident of New Orleans Police officers' use of physical force in the line of duty. The 33 features of the data set span various aspects of each incident like the date of occurrence and the shift, logistical information such as bureau, division, and unit, officers' demographics, experience, and injury status, subjects' characteristics and outcomes such as gender, age, injury, hospitalization, and arrest status, environmental factors like lighting and weather conditions, as well as nuanced elements like the type and level of force employed, reasons behind its use, and its overall effectiveness. This dataset provides a comprehensive foundation for understanding the complexities and patterns associated with police use of force during an incident but was set up for record keeping which adds some hindrances when analyzed.

The dataset is self-reported by the NOPD and in its original state not ready for machine learning, let alone statistical analysis. Even after data cleaning, unless there is a one to one ratio, there is no direct correlation between the number of officers involved and a specific subject due to concatenation of multiple fields. This approach, while probably necessary for data recording, creates a lack of clarity and hinders both machine learning and statistical analysis. The dataset's original state, as reported by the NOPD, is not readily suitable for these analytical purposes. Therefore, additional data preprocessing steps were required to extract insights and find correlations within the dataset. There was no perfect solution to solve this issue, but few were found to be useful in enabling effective analysis and modeling.

There was a fairly expensive amount of organizing, restructuring, and cleaning of the data set. Eventually it was split three ways between an incident table that had all the concatenation described as proportions and subject and police tables where the concatenation was flattened out into individual officers or subjects. Each table was connected by a key value pair or the original PIB number. During the exploratory data analysis faze another feature was decided to be necessary as a sort of grade for each incident. A weighted feature was created based mostly on if the subjects were injured, which type of force was used and if the force was authorized or deemed acceptable by command. By tweaking this a decently distributed data point was formed to use to train models to predict the "Grade" of an incident.

Key findings from our analysis of NOPD incidents involving the use of force highlight several pressing concerns. Calls involving mentally unstable subjects result in violent outcomes in over 13% of cases, suggesting a need for alternative, professionally trained de-escalation services. Addressing factors like alcohol consumption, contributing to nearly 13% of failed incidents as well, this alone could significantly reduce lower-grade incidents.

The relationship between force type and subject injuries, especially firearm exhibition, raises questions due to a high number of associated injuries. Female officers using firearms less frequently than males also suggests potential adjustments in the selection of the type of force used by male officers could further reduce injury.

Disparities in gender and racial representation within the police force compared to incident data underscore the need for the police force to self adjust male and white officers groups to at least match their peers to potentially reduce violent incidents. While further data normalization is necessary, these findings provide valuable insights into addressing NOPD use-of-force incidents.

During the initial phase of building the models, I fist an entire one-hot encoded incident dataset to use on all the models. Then multiple models were initially explored using basic linear regression. My primary goal was to develop a precognitive model capable of predicting the graded feature by relying on the various factors leading up to an event. These factors included demographics, situational conditions, and even the type of force being used.

We also developed a second, more comprehensive model that incorporated all the successful predictive features from the entire incident dataset. This model included all the features that contributed to the creation of the weighted grade feature, and it performed impressively.

Throughout the modeling phase, we experimented with various techniques, including linear regression, Random Forest, XGBoost, and machine learning with TensorFlow. We also attempted to enrich the dataset to improve the performance of our machine learning models. In the end, while all models showed promising results, XGBoost and TensorFlow emerged as the top performers, demonstrating their advanced predictive capabilities.

Methods/Analysis

Package Install

Here are some notes on installing miniconda for the tensorflow ML part of the script. If the script breaks there is because it was tough to programticly install miniconda on a device, you will have to fend for your self.

```
#Note: This script will take a while to run. About 5 minutes on a system with the below specs:
       #CPU: AMD® Ryzen 5 3600 6-core processor × 12
       #RAM: 32.00 GB
# you can run everything but the last chunk in about 15 min
##Installing Packages
# List of packages for session
.packages = c("readr",  # Reading and writing data
    "tidyr",  # Data tidying and reshaping
    "stringr",  # String manipulation
    "ggplot2",  # Data visualization
                 "ggplot2", # Data visualization
"dplyr", # Data manipulation and transformation

f mackages for data scien
                 "tidyverse",  # Collection of packages for data science
                 "caret", # Classification and regression training
"reshape2", # Data reshaping and melting
                 "randomForest", # Random forest algorithm for regression
                 "keras",  # High-level neural networks in R
"xgboost",  # Extreme Gradient Boosting
                 "reticulate", # R interface to Python
                 "tensorflow"
                                     # Machine learning framework
# Install packages (if not already installed)
.inst <- .packages %in% installed.packages()</pre>
if(length(.packages[!.inst]) > 0) install.packages(.packages[!.inst])
# Load packages into session
lapply(.packages, require, character.only=TRUE)
```

```
library(readr)
library(tidyr)
library(stringr)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(caret)
library(reshape2)
library(randomForest)
library(keras)
library(xgboost)
library(reticulate)
library(tensorflow)
##### Here are some notes on installing miniconda for the tensorflow ML part of the script
## Unquote below
# reticulate::install_miniconda()
#### I would Try this one bellow first but your system my not be completely compatible
### I had to do an alternate build
# Unquote below
# reticulate::conda_create(envname = "r-tensorflow", python = "3.9")
####This one is specific to my computer build
##### reticulate::conda_install(envname = "r-tensorflow", packages = "tensorflow=2.6.0")
```

Setting the Data

To begin the analysis, the data is imported. A dedicated download link is provided for the specific data snapshot used in this analysis. While the code may function with data from an API, it's important to note that the analysis and cleaning processes are tailored to this specific dataset snapshot. Introducing additional entries may potentially lead to issues, such as breaks or the introduction of missing values (N/A's). It's important to emphasize that this code is not designed for dynamic daily data intake and analysis; it's optimized for this static dataset snapshot.

Exploratory Data Analysis and Feature Engineering

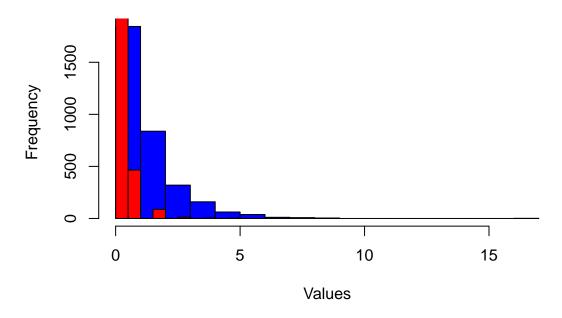
List of All Our Initial Feature

#	## [1]	"PIB_File_Number"	"Date_Occurred"
#	## [3]	"Originating_Bureau"	"Division_level"
#	# [5]	"Division"	"Unit"
#	## [7]	"Working_Status"	"Shift"
#	## [9]	"Investigation_status"	"Disposition"
#	## [11]	"Service_Type"	"Light_Condition"
#	## [13]	"Weather_Condition"	"Use_of_Force_Type"
#	## [15]	"Use_of_Force_Level"	"Subject_Influencing_Factors"
#	## [17]	"Use_of_Force_Effective"	"Officer_Race"
#	# [19]	"Officer_Gender"	"Officer_Age"
#	# [21]	"Officer_Years_of_Service"	"Distance_Between"
#	# [23]	"Subject_Gender"	"Subject_Ethnicity"
#	# [25]	"Subject_Age"	"Subject_Build"
#	# [27]	"Subject_Height"	"Subject_Injured"
#	# [29]	"Subject_Hospitalized"	"Subject_Arrested"
#	## [31]	"Subject_Arrest_Charges"	"Use_of_Force_Reason"
#	## [33]	"Officer_Injured"	"Officer_Female"
#	# [35]	"Officer_Male"	"Officer_Number"

Initial Look at Officer Gender

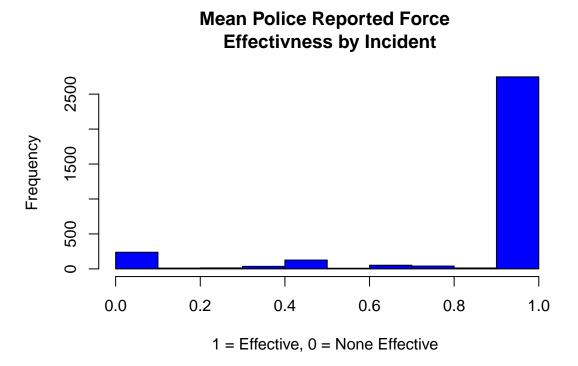
Here we have an initial look at the frequency of the number of officers per gender on the incidences. We will delve into this feature further in our analysis.

Number of Officer's by Gender per Incident



Use of Force Effectivness

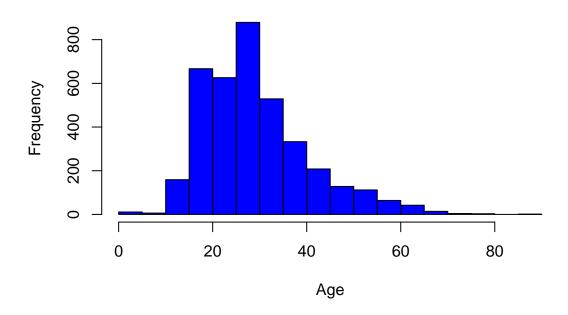
This metric represents the mean use of force effectiveness, serving as the closest approximation to an inhouse "Grade" derived from the NOPD's self-reporting. It measures the effectiveness of force employed during incidents, providing valuable insights into the performance and outcomes of the police force's actions. We will see this be an important feature in our predictive modeling.



Subject's Age

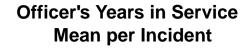
This histogram illustrates the distribution of subject ages involved in various incidents. Notably, this feature exhibits a relatively normal distribution, mirroring the age demographics of the population. It's worth mentioning a small spike in the data, which resulted from substituting the average height for missing values.

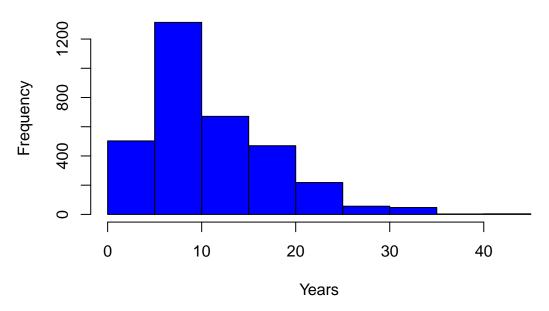
Aggregated Subject Age



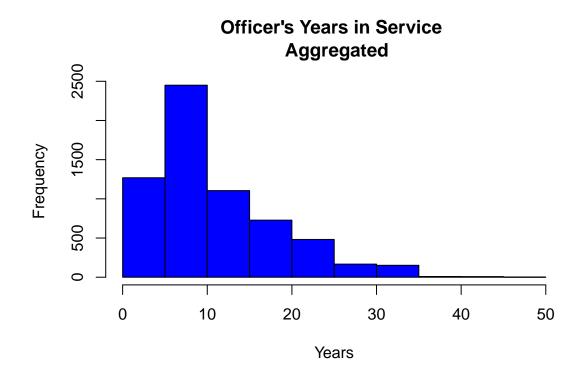
Age of Officer

Here we see a histogram of the frequency of the average age of all the officers involved in each incident. There seems to be a bit of an abnormal spike in data for incidences involving officers between 5 and 10 years of service. While this spike may seem unusual, it could be attributed to the peak years during which officers are actively engaged in patrol duties. This intriguing pattern invites further exploration into the relationship between an officer's years of service and their likelihood of being involved in incidents. It prompts questions about whether officers in this service range are more frequently dispatched or if there are specific roles or responsibilities associated with this experience level that contribute to the observed trend.



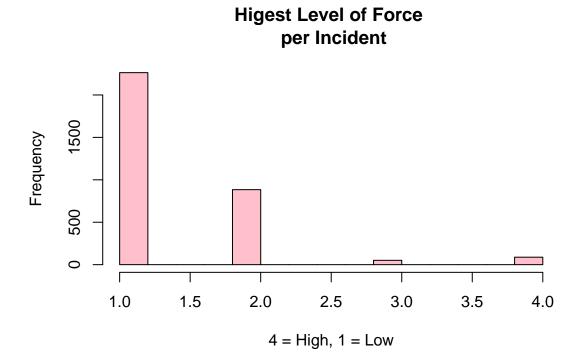


The aggregated officer table has the same results.

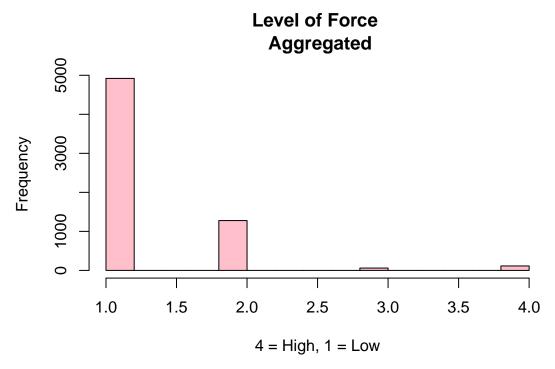


Use of Force Level

The "Use of Force Level" feature represents an in-house evaluation metric used by the police force to assess the level of force employed during incidents. The exact methodology behind its determination may require further investigation. Within this section, you'll find a histogram displaying the frequency of the highest force level observed in each incident.



Here is the aggregated per officer use of force Level which can give greater insight to the actions of the whole

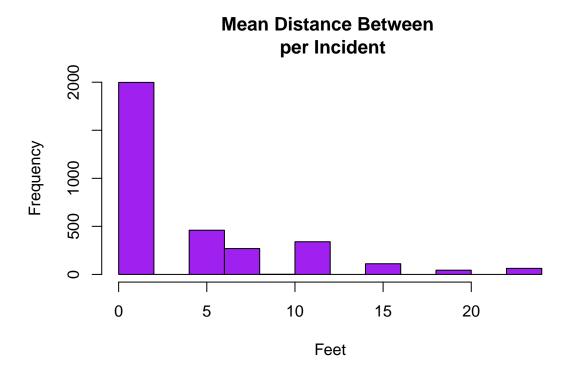


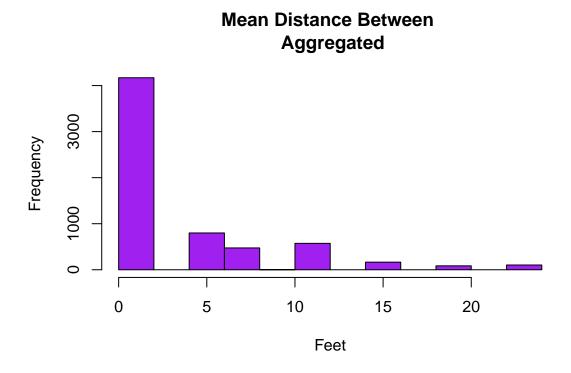
police force.

Distance Between Officer and Subject

This section presents two histograms depicting the distances between officers and subjects. These visualizations offer valuable insights into the spatial dynamics during police incidents:

- 1. Incident Average Distance: This histogram illustrates the average distance between officers and subjects for individual incidents. Analyzing this data can help identify variations in proximity across different incidents.
- 2. Aggregated Distance per Officer: The second histogram showcases the aggregated distance between officers and subjects, grouped by individual officers. By examining this data, you can gain a deeper understanding of how officer-subject distances vary among officers over multiple incidents.

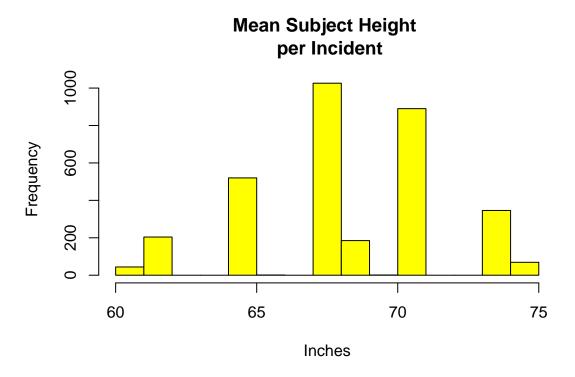


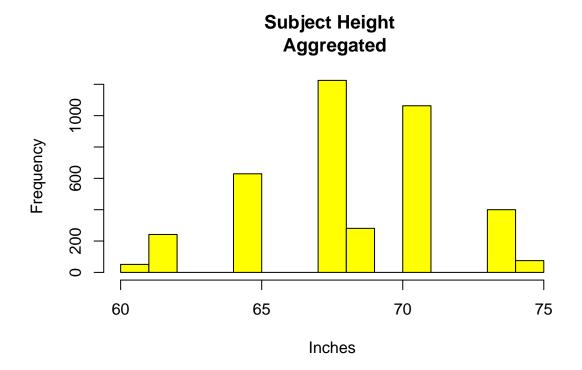


Subjects Height

In this section, you will find two visual representations of the subjects' height data.

- 1. Average Height per Incident: This visualization provides insight into the average height of subjects per incident, allowing for a better understanding of height variations across different incidents.
- 2. Height Distribution: The histogram depicts the distribution of heights among all aggregated subjects. This visualization offers a comprehensive overview of the height data's distribution patterns.

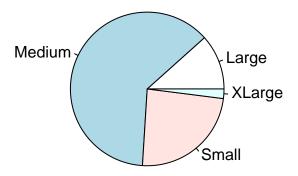




Subject's Build

In this section, a pie chart illustrating the physical build of the subjects involved is provided. The preponderance of medium-built subjects likely reflects the general population, similar to the distribution observed for the height feature.

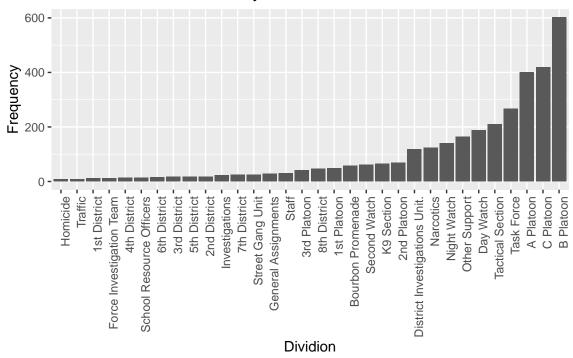
Subject Build Incident Mean



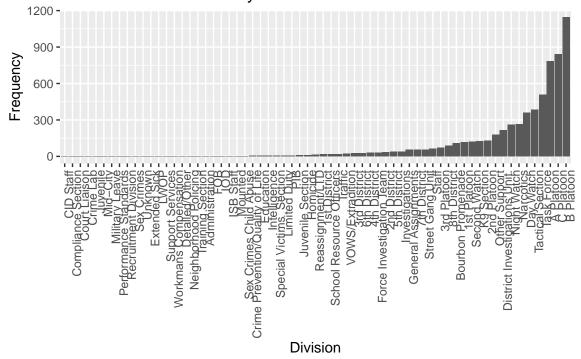
Division

The "Division" feature suggests that larger organizational units, such as platoons, bear a substantial responsibility for a significant majority of the incidents. Additionally, a few tactical and investigative units appear to play a significant role. This feature holds great potential for enhancing predictive modeling efforts.

Incident table occurance by division



Police table occurance by division

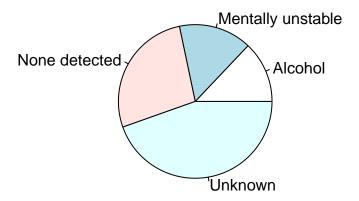


Influencing factors

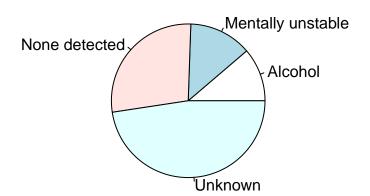
Influencing factors may represent one of the key controllable conditions under the subject's control when determining the likelihood of injury during an incident. These factors could encompass a range of variables,

such as the subject's behavior, compliance with instructions, or level of agitation. Investigating the role of influencing factors in incidents of force can shed light on the dynamics at play and help identify opportunities for prevention and de-escalation strategies.

Subject Influencing Factors by Incident



Subject Influencing Aggregated

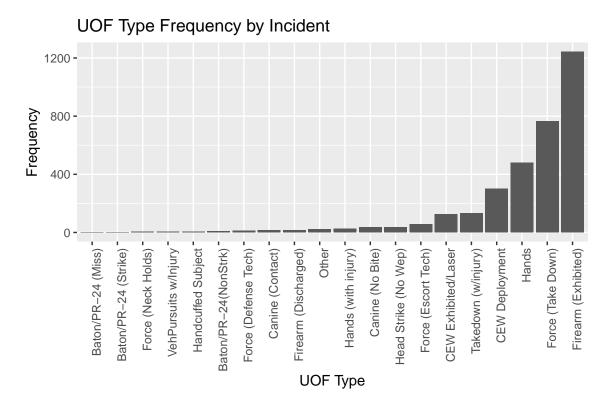


Use of Force Type

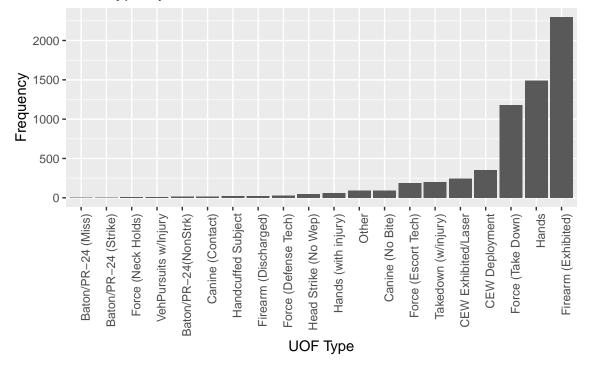
Use of Force Type is a highly weighted feature due to its distribution. Notably, "Firearm (Exhibited)" stands out with 1242 occurrences, making it the most frequently used force type in the incidence table and 2296 occurrences in the aggregate police table. Further investigation is required to assess the severity of injury associated with each force type, but this feature holds potential importance for modeling. However, it's evident that the predominant method of force used in the data is firearms.

During the cleaning process, when these fields are aggregated, such as grouping all firearms or all caninerelated entries together, our models tend to lose a significant portion of their predictive power. It appears that this specific level of aggregation is preferred by the models.

##		String	Frequency.Var1	Frequency.Freq
##	1	9	Baton/PR-24 (Miss)	1
##	2	Baton/PR-24 (Strike)		3
##	12	Force (Neck Holds)	Force (Neck Holds)	4
##	20	VehPursuits w/Injury	VehPursuits w/Injury	4
##	14	Handcuffed Subject	Handcuffed Subject	5
##	3	Baton/PR-24(NonStrk)	Baton/PR-24(NonStrk)	7
##	10	Force (Defense Tech)	Force (Defense Tech)	13
##	4	Canine (Contact)	Canine (Contact)	14
##	8	Firearm (Discharged)	Firearm (Discharged)	16
##	18	Other	Other	21
##	16	Hands (with injury)	Hands (with injury)	26
##	5	Canine (No Bite)	Canine (No Bite)	36
##	17	Head Strike (No Wep)	Head Strike (No Wep)	37
##	11	Force (Escort Tech)	Force (Escort Tech)	55
##	7	CEW Exhibited/Laser	CEW Exhibited/Laser	125
##	19	Takedown (w/injury)	Takedown (w/injury)	133
##	6	CEW Deployment	CEW Deployment	300
##	15	Hands	Hands	480
##	13	Force (Take Down)	Force (Take Down)	764
##	9	Firearm (Exhibited)	Firearm (Exhibited)	1242
##		C+ring	Eroguanau Vari	Fraguency Frag
##	1	String	- •	Frequency.Freq
##	1	Baton/PR-24 (Miss)	Baton/PR-24 (Miss)	1
## ##	2	Baton/PR-24 (Miss) Baton/PR-24 (Strike)	Baton/PR-24 (Miss) Baton/PR-24 (Strike)	1 6
## ## ##	2 12	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds)	1 6 8
## ## ## ##	2 12 20	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury	1 6 8 11
## ## ## ##	2 12 20 3	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury	1 6 8
## ## ## ##	2 12 20 3	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact)	1 6 8 11 13
## ## ## ## ##	2 12 20 3 4 14	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject	1 6 8 11 13
## ## ## ## ## ##	2 12 20 3 4 14 8	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged)	1 6 8 11 13 14 21
## ## ## ## ## ##	2 12 20 3 4 14 8	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech)	1 6 8 11 13 14 21 24
## ## ## ## ## ##	2 12 20 3 4 14 8	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech)	1 6 8 11 13 14 21 24 30
## ## ## ## ## ##	2 12 20 3 4 14 8 10	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep)	1 6 8 11 13 14 21 24 30 47
## ## ## ## ## ## ##	2 12 20 3 4 14 8 10 17 16 18	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury)	1 6 8 11 13 14 21 24 30 47 62
## ## ## ## ## ## ##	2 12 20 3 4 14 8 10 17 16 18	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other	1 6 8 11 13 14 21 24 30 47 62 89
## ## ## ## ## ## ## ## ## ## ## ## ##	2 12 20 3 4 14 8 10 17 16 18 5	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite)	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite)	1 6 8 11 13 14 21 24 30 47 62 89 94
## ## ## ## ## ## ## ## ## ## ## ## ##	2 12 20 3 4 14 8 10 17 16 18 5	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury) CEW Exhibited/Laser	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech)	1 6 8 11 13 14 21 24 30 47 62 89 94
## ## ## ## ## ## ## ## ## ## ## ## ##	2 12 20 3 4 14 8 10 17 16 18 5 11 19	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury) CEW Exhibited/Laser CEW Deployment	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury)	1 6 8 11 13 14 21 24 30 47 62 89 94 190 201
######################################	2 12 20 3 4 14 8 10 17 16 18 5 11 19	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury) CEW Exhibited/Laser	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury) CEW Exhibited/Laser	1 6 8 11 13 14 21 24 30 47 62 89 94 190 201 244
######################################	2 12 20 3 4 14 8 10 17 16 18 5 11 19 7 6 13 15	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury) CEW Exhibited/Laser CEW Deployment	Baton/PR-24 (Miss) Baton/PR-24 (Strike) Force (Neck Holds) VehPursuits w/Injury Baton/PR-24(NonStrk) Canine (Contact) Handcuffed Subject Firearm (Discharged) Force (Defense Tech) Head Strike (No Wep) Hands (with injury) Other Canine (No Bite) Force (Escort Tech) Takedown (w/injury) CEW Exhibited/Laser CEW Deployment	1 6 8 11 13 14 21 24 30 47 62 89 94 190 201 244 350





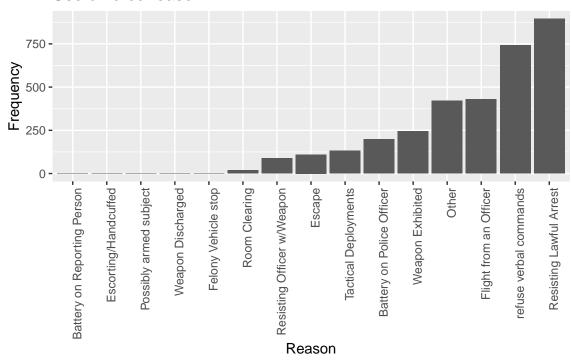


Use of Force Reason

Use of Force Reason appears to be closely related to the subject involved and could serve as an indicator of the justification for the level of force used. This observation suggests that delv-

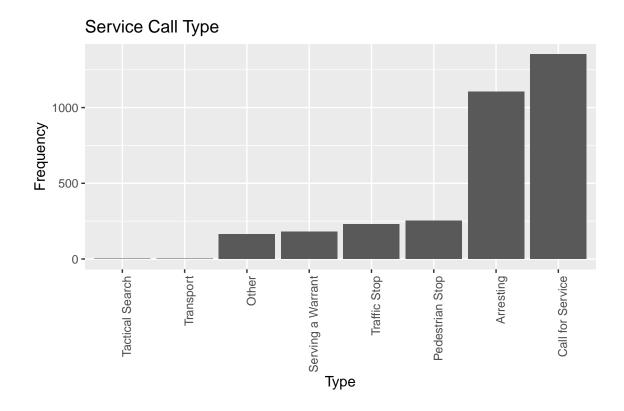
ing deeper into the relationship between Use of Force Reason and the severity of force applied.

Use of force reason



Service Call Type

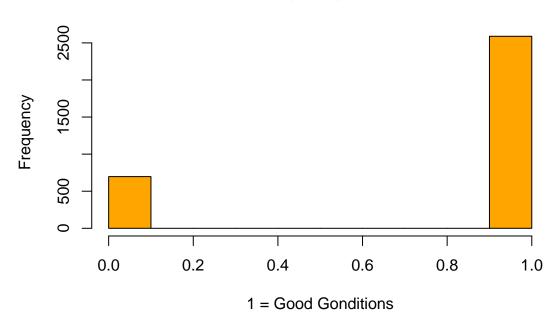
Almost 50% of incidents were calls for service. This suggests that a significant portion of the incidents involved the police responding to service calls, which could include various types of requests or situations were citizens requested police intervention.



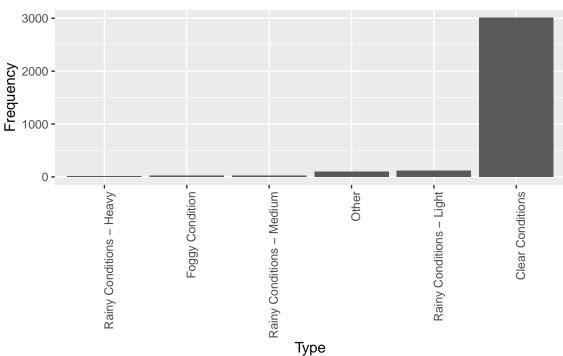
Incident Lighting and Weather

Most cases have clear conditions with good lighting.

Incident Lighting Conditions

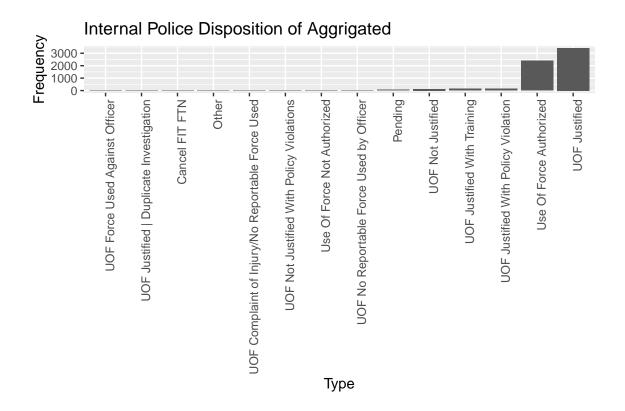






Disposition

The "Disposition" feature in the dataset indicates that most instances of the use of force are reported as justified or authorized by the police themselves. This feature serves as the closest approximation to a self-reporting evaluation of the incidents. Though including "Disposition" as a significant factor in our predictive models, it's essential to consider creating our own grading or evaluation feature to ensure a more comprehensive and unbiased assessment of incident outcomes and their impact on our model's predictions.



Feature Creation: Incident Grade

In this section, we are taking the initiative to create our own feature that assigns a grade to the severity of each incident. This grading is based on a weighted combination of various existing features. Initially, we start by assigning weights to different categories within an in-house evaluation feature called "Disposition." We slightly reduce the weight for "UOF Justified" because it wasn't fully authorized, whereas "UOF Justified With Training" and "UOF Justified With Policy Violation" are considered more severe. "Use Of Force Not Authorized / UOF Not Justified With Policy Violations" receives the maximum penalty weight, as it represents the most critical cases. Regardless of the category, all cases receive a minimum weight of 0.1, signifying that force was used, and de-escalation wasn't achieved before its necessity.

```
weights <- c(
   "Use Of Force Authorized" =.1,
   "UOF Justified" = .3,
   "UOF Justified With Training" = 0.7,
   "Pending" = 0.1,
   "UOF Not Justified" = 1,
   "UOF No Reportable Force Used by Officer" = 0.1,
   "Other" = 0.1,
   "UOF Justified With Policy Violation" = .5,
   "Cancel FIT FTN" = 0,
   "Use Of Force Not Authorized" = 1,
   "UOF Complaint of Injury/No Reportable Force Used" = .1,
   "UOF Justified | Duplicate Investigation" = .1,
   "UOF Force Used Against Officer" = .1,
   "UOF Not Justified With Policy Violations" = 1
)</pre>
```

The weighting grade should take into account the varying severity of force types, assigning greater weight to those with a higher risk of causing injury. Initially, I assigned a lower weight to 'Firearm (Exhibited),' but this decision had a substantial impact on the mean of the weighting grade. Due to the frequent use and significant potential lethality of this force type, it became evident that a heavier weight should be assigned to it. Moreover, considering the comparatively lower potential harm associated with alternative police methods, it's impractical for firearms to be the default, excessively used option.

```
type_weight <- c(</pre>
  "Firearm (Discharged)" = 1,
  "VehPursuits w/Injury" = 1,
  "Takedown (w/injury)"= .4,
  "Canine (Contact)" = .7,
  "CEW Deployment" = .3,
  "Baton/PR-24 (Strike)" = .3,
  "Head Strike (No Wep)" = .3,
  "Force (Take Down)" = .3,
  "Force (Defense Tech)" = .3,
  "Force (Neck Holds)" = .8,
  "Baton/PR-24(NonStrk)" = .3,
  "Hands" = .2,
  "Force (Escort Tech)" = .1,
  "Firearm (Exhibited)" = .9,
  "CEW Exhibited/Laser" = .2,
  "Hands (with injury)" = 3,
  "Other" = 0,
  "Canine (No Bite)" = .1,
  "Handcuffed Subject" = .1,
  "Baton/PR-24 (Miss)" = .1
```

This calculation was performed because the actual racial composition of the police force did not align closely with the racial distribution in the incidents. To account for this disparity in our new grading feature, the data in the Police Race feature was normalized with a slight weighting based on race. It's important to note that this adjustment has a minimal impact on the overall analysis, as the later sections will demonstrate that police race and gender features have limited predictive power in the model.

```
# race data
copracedata<- data.frame(
   Race = c("Black", "White"),
   Provided_Percentage = c(57, 37),
   Calculated_Percentage = c(47.22563, 52.77437)
)
# Calculate normalized race weights
coprace_diff <- (copracedata$Provided_Percentage - copracedata$Calculated_Percentage)
coprace_diff</pre>
```

```
## [1] 9.77437 -15.77437
```

```
race_weight <- c(
  "White"= .16,
   "Majority_White" = .16,
   "White&Black" = 0,
   "Hispanic" = 0,</pre>
```

```
"Black&Hispanic" = 0,
      "Black&Asian_Pacific_Islander" = 0,
      "White&Race Unknown" = 0,
      "Majority Asian Pacific Islander" = 0,
      "Mixed" = 0,
      "White&Hispanic" = 0,
      "Hispanic&Asian_Pacific_Islander" = 0,
      "Majority_Hispanic" = 0,
      "Black&Race_Unknown" = 0,
      "Black&American_Indian" = 0,
      "Black" = -.1,
      "Majority_Black" = -.1,
      "Asian_Pacific_Islander" = 0,
      "American_Indian" = 0,
      "Race_Unknown" = 0,
      "White&Asian_Pacific_Islander" = 0,
      "White&American_Indian" = 0,
      "Asian/Pacific Islander" = 0,
      "Not Specified" = 0,
      "American Indian/Alaska Native" = 0,
      "Not Applicable (Non-U.S.)" = 0
  )
# If you look below in the analysis of officer gender there is a 12% normalization between the genders
gender_weight <- c(</pre>
  "Male" = .12,
    "Male&Female" = 0,
    "Majority_Male" = .12,
    "Female" = -.12,
    "Majority_Female" = -.12
```

This is the one place where I use the PIB # across tables to get the total officers and subjects injured. While I might have created a similar feature within the incidents table, this demonstrates a potential approach for linking datasets in machine learning or linear regression scenarios.

```
# Aggregate police_table_nola_use_of_force by PIB_File_Number
aggregated_police <- police_table_nola_use_of_force %>%
   group_by(PIB_File_Number) %>%
   summarise(
    Officer_Injured = sum(as.numeric(Officer_Injured))
)

# Aggregate subject_table_use_of_force by PIB_File_Number
aggregated_subject <- subject_table_use_of_force %>%
   group_by(PIB_File_Number) %>%
   summarise(
    Subject_Injured = sum(as.numeric(Subject_Injured)),
    Subject_Hospitalized = sum(as.numeric(Subject_Hospitalized))
)
```

These are the final weightings I settled on. I experimented with various weightings, and the alternative weights can be found in the code. However, it's essential to note that duplicating them for the police table is

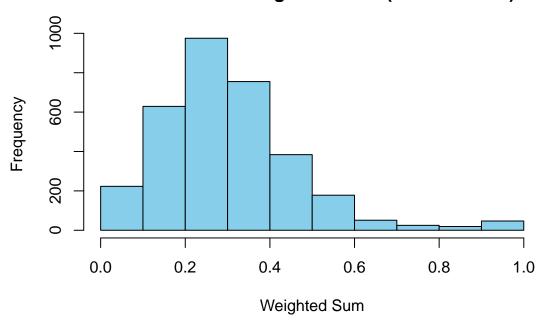
necessary for the code to function correctly. The weight of .28 assigned to 'Disposition' reflects an attempt to align with the in-house grading system, as it is the closest approximation. Special emphasis is placed on the 'Use of Force Type' feature, with a higher weight assigned to penalize firearm usage, given its critical role in it's frequency of the data. 'Subject Injured' and 'Subject Hospitalized' carry significant weight because these factors primarily determine the severity of an incident. An incident is more likely to be deemed a failure if the subject is injured. The weight assigned to 'Officer Injured' is relatively small due to the limited number of officers typically injured in incidents. And lastly, 'Use of Force Level' is not heavily weighted as it can lead to data skewing and overfitting issues in the models.

```
incident_nola_use_of_force <- incident_nola_use_of_force %>%
  left_join(aggregated_police, by = "PIB_File_Number") %>%
  left_join(aggregated_subject, by = "PIB_File_Number") %>%
  mutate(
    weighted_sum = 0.28 * (weights[Disposition]) +
      0.23 * (type_weight[Use_of_Force_Type]) +
      0.12 * (race_weight[Officer_Race]) +
      0.05 * (gender_weight[Officer_Gender]) +
      0.01 * (.25 * Use of Force Level) +
      0.01 * (Officer_Injured ) +
      0.1 * (Subject_Injured ) +
     0.2 * (Subject_Hospitalized)
  )
police_table_nola_use_of_force <- police_table_nola_use_of_force %>%
  left_join(aggregated_subject, by = "PIB_File_Number") %>%
   weighted_sum = 0.28 * (weights[Disposition]) +
      0.23 * (type_weight[Use_of_Force_Type]) +
      0.12 * (race_weight[Officer_Race]) +
      0.05 * (gender_weight[Officer_Gender]) +
     0.01 * (.25 * Use_of_Force_Level) +
      0.01 * (Officer_Injured ) +
      0.1 * (Subject_Injured ) +
      0.2 * (Subject_Hospitalized)
```

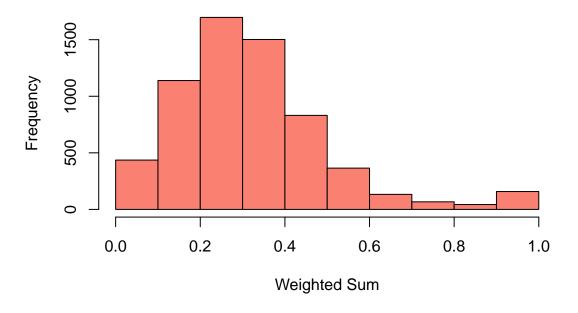
Weighted Sum Feature

The Weighted Sum feature provides a comprehensive summary of the new evaluations applied to both incidents and individual police actions. This feature exhibits a relatively even distribution, reflecting a range of incident severities. The mean value is notably low, indicating that a significant portion of the data represents less severe incidents. However, there exists a tail of outliers on the higher end of the scale, signifying a summation of the most negative and potentially violent occurrences within the dataset.

Distribution of Weighted Sums (Incident Data)



Distribution of Weighted Sums (Police Table Data)



Grade

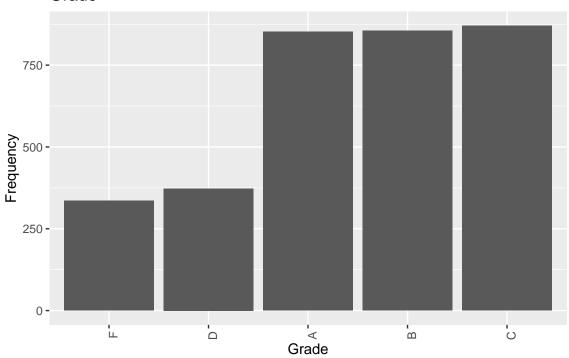
In this section, we transform the Weighted Sum feature into a set of subdivided grades. This step is crucial because even an incident scoring at 0.5 indicates significant issues according to our weightings. At this grade level, it's highly likely that someone was injured during the incident, representing a failure in our grading system. This subdivision allows for a realistic assessment of incident severity, ensuring that deviations from ideal outcomes are duly recognized and addressed.

```
# Continue with the Grade assignment from the weighted sum
incident_nola_use_of_force <- incident_nola_use_of_force %>%
 mutate(
    Grade = case when(
      weighted_sum >= 0.5 ~ "F",
      weighted_sum >= 0.4 ~ "D",
      weighted_sum >= 0.3 ~ "C",
      weighted sum \geq 0.2 \sim "B",
      weighted_sum >= 0 ~ "A",
      TRUE ~ "A"
 )
# Continue with the Grade assignment
police_table_nola_use_of_force <- police_table_nola_use_of_force %>%
 mutate(
    Grade = case_when(
      weighted_sum >= 0.5 ~ "F",
      weighted_sum >= 0.4 ~ "D",
      weighted_sum >= 0.3 ~ "C",
```

```
weighted_sum >= 0.2 ~ "B",
  weighted_sum >= 0 ~ "A",
  TRUE ~ "A"
 )
)
table(incident_nola_use_of_force$Grade)
```

```
## ## A B C D F
## 852 855 870 373 336
```

Grade

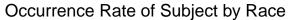


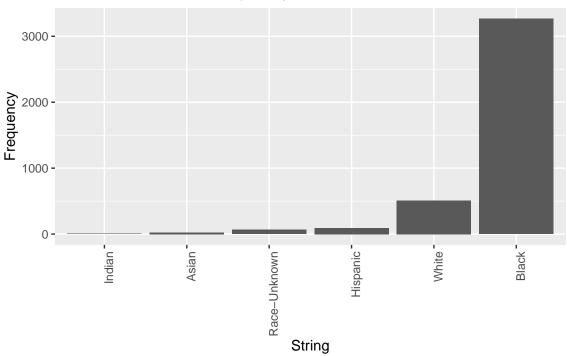
Feature-to-Feature Analysis

Now that we've taken an initial look at the features and created grading features for incidents, we can delve into analyzing the data for significant correlations between these features.

Subject's Ethnicity

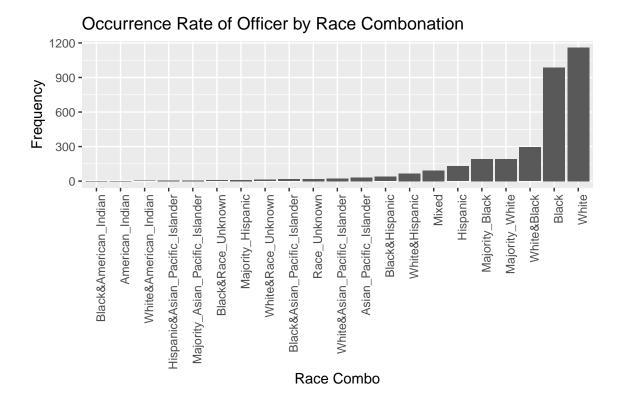
The subject's ethnicity data overwhelmingly represents black subjects. Initially, I considered building a predictive model for subject race, but due to the overwhelming majority of black subjects, many features would likely predict this category.





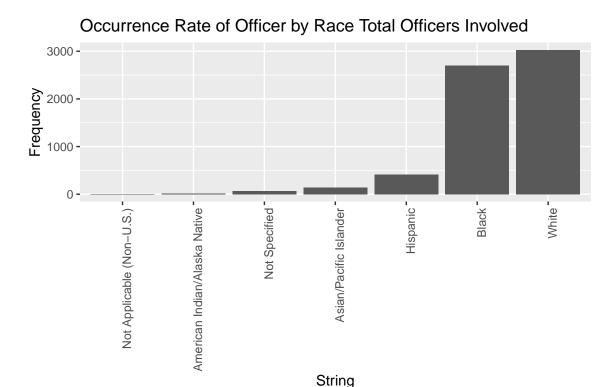
Officer's Ethnicity

Now, let's delve into the demographics of officers, focusing on their races. The incident data set includes officers' races in all their possible permutations. It's worth noting that black and white officers constitute the majority of the force, and consequently, most incidents involve individuals from these two racial groups. Notably, as we'll see later, there is a negative correlation between these two groups, which has been flagged by some of the predictive models, revealing an interesting aspect of the data.



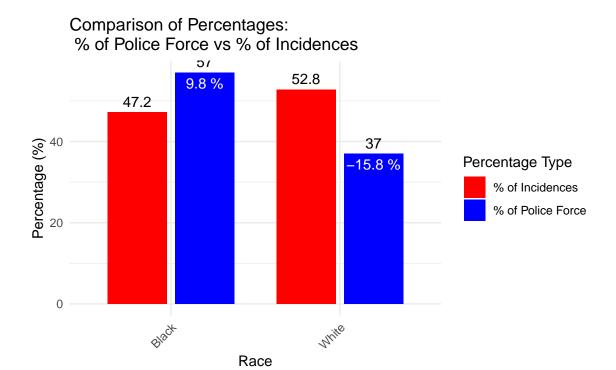
Aggregate Officer Ethnicity

When we examine the aggregate officer table, it becomes evident that two racial groups, Black and White, dominate the majority of incidents. However, an important statistic missing from the data is the racial composition of the actual police force. As of September 23, 2016, according to a report by nola.com, the NOPD's racial makeup stood at 57 percent Black and 37 percent White. This information underscores the significance of understanding how the demographics of the police force align with the racial distribution of incidents, which can have implications for further analysis and interpretation.



Differences Between Force and Data

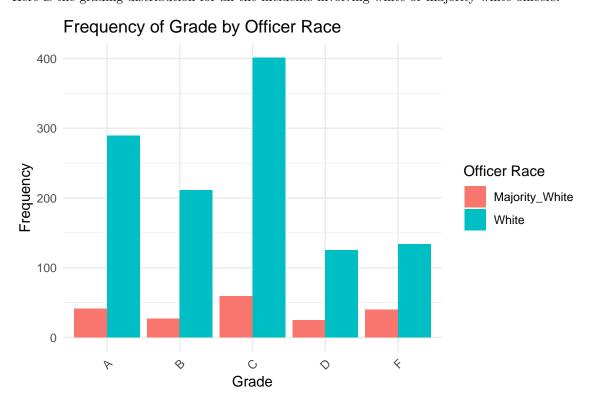
If we compare the composition of the actual police force to the racial makeup of police force incidences, we observe notable differences between the two datasets. Specifically, when comparing Black and white police officers, there is a significant 25% difference in their incidence numbers compared to the police force makeup numbers. These are the numbers when using the the aggregated police table, which is 2% less than the incident table, resulting in a 9.77% lower incidence percentage for Black officers and a 15.77% higher incidence percentage for white officers compared to the police force makeup. While this data may normalize further with additional information, it does raise some concerns. As previously demonstrated in our weighted grading feature, I have incorporated this data to adjust the grading weights for this occurrence.



Difference in Percentage

Grading Distribution for White Officers

Here is the grading distribution for all the incidents involving white or majority-white officers.



Reducing Problematic Incidents: The Impact of Adjusting Incident Counts for White Police Officers

If we address problematic low-grade incidents involving white police officers by adjusting their incidence count based on the difference in their percentage of force makeup, we may witness a substantial reduction in violent incident cases. It's important to note that we are utilizing the lower aggregate police table difference mentioned earlier; when using the incidents table, the difference is actually 2% higher. This difference attributed to white officers not aligning with their expected representation in the police force makeup potentially account for up to 7.2% (subject to evaluation based on weighted features) of the above-average violent incidents.

```
# Filter the summarized data to include only "D" and "F" Grade entries
 summarized_data_filtered <- summarized_data %>%
   filter(Grade %in% c("D", "F"))
print(summarized_data_filtered)
## # A tibble: 4 x 3
## # Groups:
               Officer_Race [2]
    Officer_Race
                    Grade Count
     <chr>>
                    <chr> <int>
##
## 1 Majority_White D
                             25
## 2 Majority_White F
                             40
## 3 White
                    D
                            125
## 4 White
                    F
                            134
extra_white_bad_incidences <- sum(summarized_data_filtered$Count)*.158
print(extra_white_bad_incidences)
```

[1] 51.192

summarized_data <- filtered_data %>%
group_by(Officer_Race, Grade) %>%

summarize(Count = n())

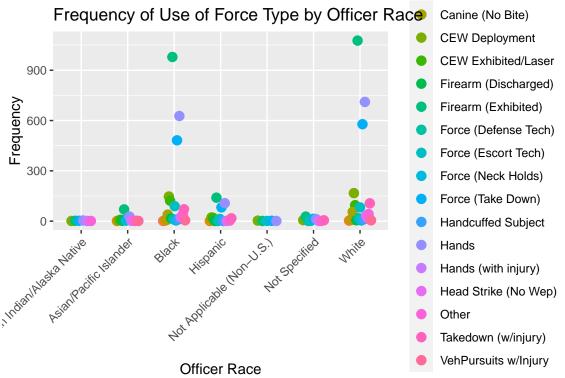
```
sumerized_incidents <- incident_nola_use_of_force %>%
  filter(Grade %in% c("D", "F"))

# 7.2% (depending on evaluation weighted feature) of above and beyond violent incidences are due to wh
print(extra_white_bad_incidences/nrow(sumerized_incidents))
```

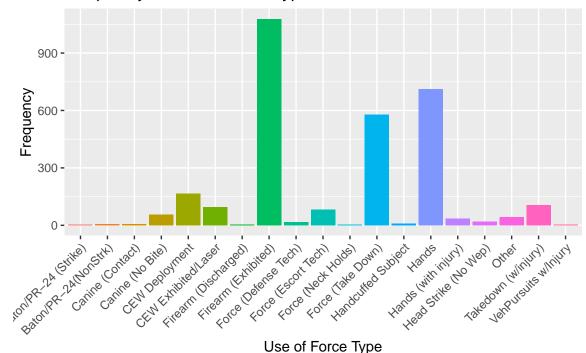
[1] 0.0722031

Use of Force by Office Race

There is a similar pattern for use of force type for all racial demographics of police officers.



Frequency of Use of Force Type for White Officers

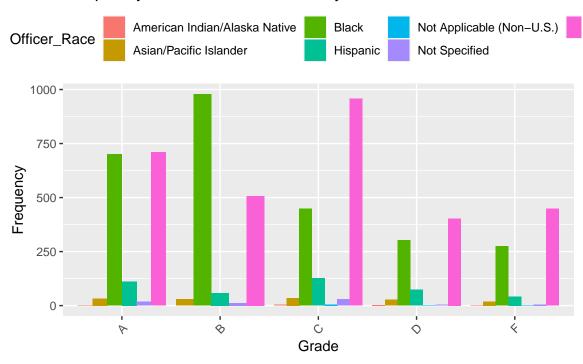


Use of Force Type

Grade Distribution Across Officer Races

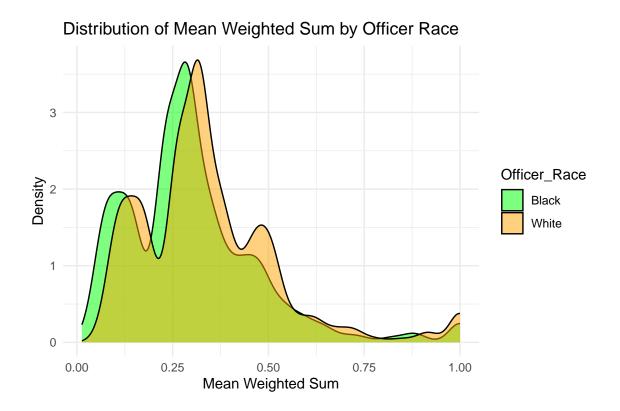
While the distribution of use of force types shows similarities across racial demographics of officers, a noticeable distinction arises when examining grade distributions. In this context, white officers tend to receive slightly lower grades than their black counterparts.

Frequency of Use of Force Grade by Officer Race



Distribution of Mean Weighted Sum by Officer Race

This density plot highlights the difference in mean weighted sums between Black and White officers.



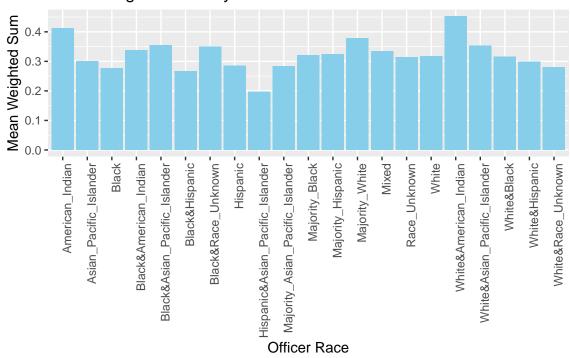
Mean Weighted Sum of All Officer's Ethnicity

Comparing the mean weighted sum by race is intriguing. However, it's important to note that the small sample size for some groups may not be entirely reliable. When Black and White officers are considered individually, their mean weighted sum is lower. It's possible that larger groups of officers in an incident could skew the mean. Most racial groups of officers tend to be close to the overall mean, but White officers exhibit a slightly higher mean. It's worth noting that the American Indian demographic has a small sample size, making their mean less reliable.

##	# 1	A tibble: 21 x 2	
##		Officer_Race	Mean_Weighted_Sum
##		<chr></chr>	<dbl></dbl>
##	1	American_Indian	0.413
##	2	Asian_Pacific_Islander	0.301
##	3	Black	0.277
##	4	Black&American_Indian	0.338
##	5	Black&Asian_Pacific_Islander	0.354
##	6	Black&Hispanic	0.266
##	7	Black&Race_Unknown	0.350
##	8	Hispanic	0.286
##	9	${\tt Hispanic\&Asian_Pacific_Islander}$	0.197
##	10	${\tt Majority_Asian_Pacific_Islander}$	0.284
##	11	Majority_Black	0.320
##	12	Majority_Hispanic	0.324
##	13	Majority_White	0.379
##	14	Mixed	0.334
##	15	Race_Unknown	0.314
##	16	White	0.317

##	17	White&American_Indian	0.452
##	18	White&Asian_Pacific_Islander	0.353
##	19	White&Black	0.315
##	20	White&Hispanic	0.299
##	21	White&Race_Unknown	0.279

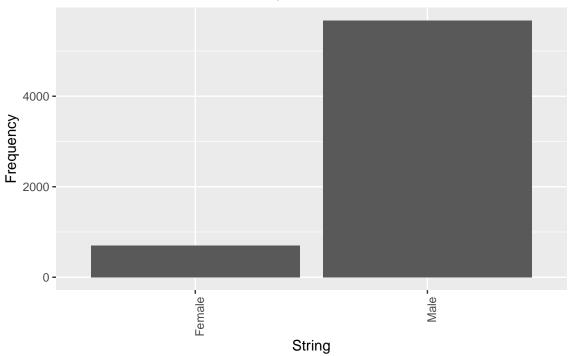
Mean Weighted Sum by Officer Race



Reexamining Officer's Gender Disparities: A Look at the NOPD's Commitment to Gender Representation

Similar to our analysis of officer ethnicity, we observe a gender imbalance among police officers. However, one crucial piece of data missing from our analysis is the gender composition of the actual police force. According to a NOPD report dated April 9, 2021, the NOPD's gender makeup stood at 23 percent Female and 77 percent Male, indicating a notable underrepresentation of women in the force.

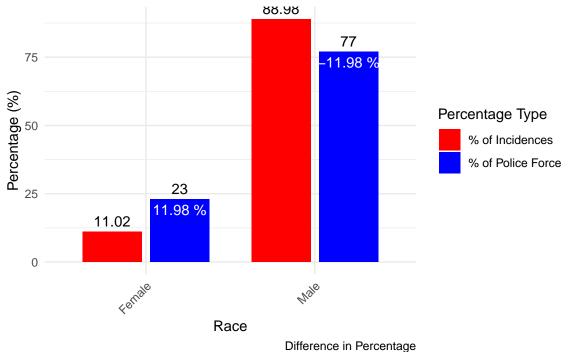




Differences Between Force and Data Gender

When we compare the gender composition of the actual police force to the gender distribution in police force incidences, we discover significant disparities between the two datasets. Specifically, there is a notable 24% difference in the incidence numbers for Female and Male police officers compared to the police force makeup statistics. These figures are based on the aggregated police table, which is 2% less than the incident table. This results in an 11.98% lower incidence percentage for Female officers and an 11.98% higher incidence percentage for Male officers compared to the police force's gender makeup. While this data may normalize further with additional information, it does raise concerns. Although I haven't yet incorporated this data into adjusting the grading weights, it's certainly could be a factor to consider in our weighted feature analysis.

Comparison of Percentages by Gender: % of Police Force vs % of Inc



Difference in Fercentage

Grading Distribution for Male Officers

Here is the grading distribution for all the incidents involving Male or majority-Male officers.

Frequency of Grade by Officer Gender Officer Gender Majority_Male Male

Reducing Problematic Incidents: The Impact of Adjusting Incident Counts for Male Police Officers

By considering adjustments to incident counts involving Male police officers based on disparities in their representation within the police force, we could potentially significantly reduce the number of high-risk incidents. This theoretical adjustment has the potential to account for up to 10.9% (pending evaluation using weighted features) of the above-average violent incidents. These incidents may be attributed to Male officers not aligning with their expected representation in the police force's makeup.

```
summarized_data <- filtered_data %>%
   group_by(Officer_Gender, Grade) %>%
   summarize(Count = n())
 # Filter the summarized data to include only "D" and "F" Grade entries
 summarized data filtered <- summarized data %>%
  filter(Grade %in% c("D", "F"))
print(summarized_data_filtered)
## # A tibble: 4 x 3
## # Groups:
              Officer_Gender [2]
##
    Officer Gender Grade Count
                    <chr> <int>
## 1 Majority_Male D
                             25
## 2 Majority_Male F
                             32
## 3 Male
                    D
                            315
## 4 Male
                            279
# If problematic low grade incidences caused by male cops are decreased by their
 # difference of percentage of force makeup to incidence count than we could see
 # a significant drop in violent incident cases
 extra_male_bad_incidences <- sum(summarized_data_filtered$Count)*.119
 extra_male_bad_incidences
```

[1] 77.469

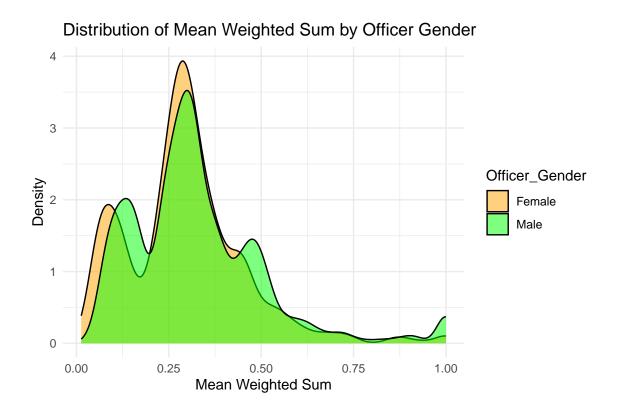
```
sumerized_incidents <- incident_nola_use_of_force %>%
  filter(Grade %in% c("D", "F"))
# 11 % of above and beyond violent incidences are due to male officers not being
# in line with at least minimizing to their force makeup percentage some of those are
# probably mixed up in the previous White officer issue

extra_male_bad_incidences/nrow(sumerized_incidents)
```

[1] 0.1092652

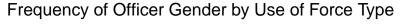
Distribution of Mean Weighted Sums: Male vs. Female Police Officers

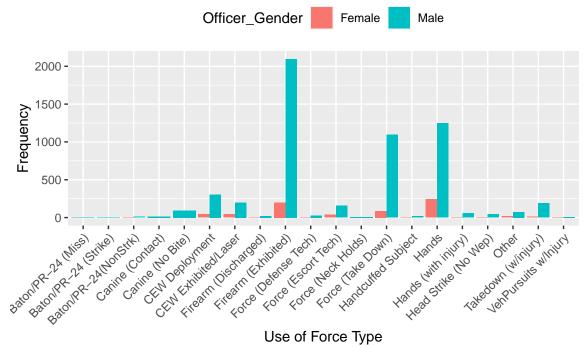
This density plot provides insights into the distribution of mean weighted sums between Male and Female police officers. It's evident that Male officers tend to have a slightly higher mean weighted sum, as indicated by the density distribution.



Officer Gender and Use of Force Type

Over looking this basic data correlation we seemingly see a similar choice of force type between the genders

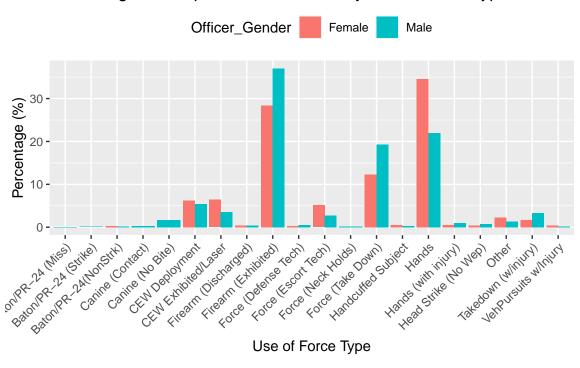




Officer Gender and Use of Force Type as a % of the Individual Genders

This section examines the choice of force type as a percentage of individual genders. It reveals that Female officers tend to use firearms less frequently than their Male counterparts, and they are more inclined to opt for physical engagement techniques.

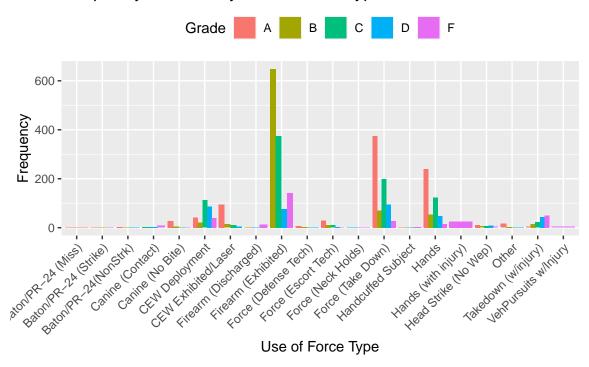
Percentage Makeup of Officer Gender by Use of Force Type



Unpacking the Paradox of Firearm Use: Exploring Grading and Injury Trends

If we exclude force type from the grading model, it might appear that the frequent use of firearms is a relatively low-injury option. However, when we incorporate force type into the grading feature, a different picture emerges, suggesting a higher level of danger associated with this type of force. Interestingly, using firearms alone can still yield a B or C grade, while D and F grades typically require subject injuries. So, the question arises: why do so many incidents involving the exhibition of firearms result in injuries?

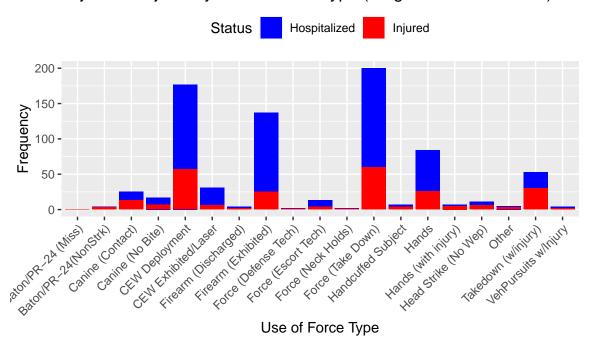
Frequency of Grade by Use of Force Type



Injury and Use of Force Type

There isn't a straightforward one-to-one correlation between the type of force used and injuries or hospitalizations of the subject, unless we narrow down the data to incidents involving a single officer. In such cases, we can reasonably attribute the injury to the specific use of force type. However, a question arises when we observe that the exhibition of firearms is responsible for a significant portion of total injuries. Why is this particular use of force associated with a notable number of injuries? Moreover, if exhibiting a weapon is indeed more dangerous to a subject and leads to injury, then Male officers should consider matching their female counterparts in reducing the use of firearm exhibition.

Combined Frequency of Hospitalized and Injured Subjects by Use of Force Type (Single Officer Incidents)

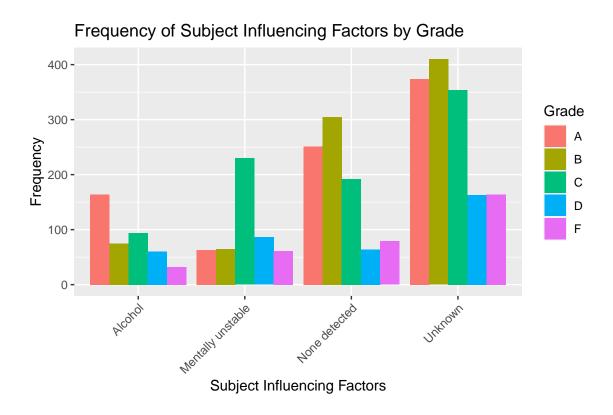


The Potential Subject Influencing Factors in Reducing violent Incidences

As discussed earlier, when adjusting the weights for Male and White officers, it's essential to acknowledge that some elements influencing incidents may have a degree of controllability. While factors like alcohol consumption may not be entirely controllable for individuals with a history of alcoholism, addressing this issue could lead to a notable reduction in failing incidents. Notably, alcohol is identified as a contributing factor in nearly 13% of failed incidents, making it a significant area for potential improvement. There is also a significant "Unknown" entry into this feature that could also contain a number of intoxicants that could have contributed to the incidents turning violent.

```
## # A tibble: 8 x 3
                Subject_Influencing_Factors [4]
##
     Subject_Influencing_Factors Grade Count
##
     <chr>>
                                    <chr> <int>
## 1 Alcohol
                                    D
                                             27
                                    F
## 2 Alcohol
                                             11
## 3 Mentally unstable
                                    D
                                             34
## 4 Mentally unstable
                                    F
                                             26
                                   D
  5 None detected
                                             27
                                    F
                                             35
## 6 None detected
                                   D
                                             76
## 7 Unknown
## 8 Unknown
                                    F
                                             59
```

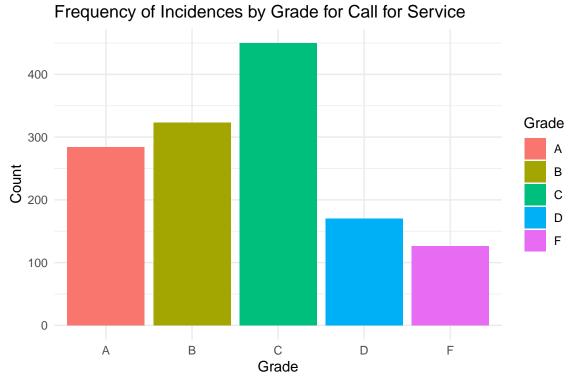
Percentage of D an F Graded entries with 'Alcohol': 12.88136 %

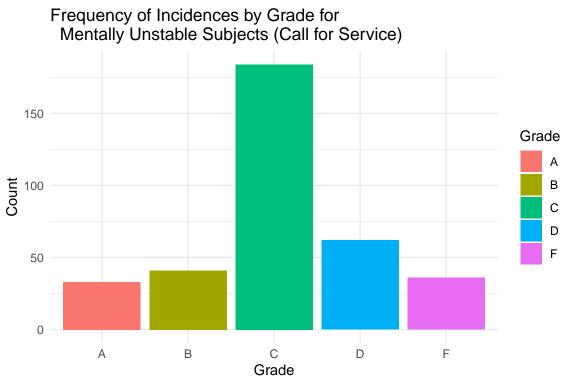


Mentally unstable Subjects and Calls of Service

Another area of concern is whether we are deploying the right resource when responding to calls for mentally unstable subjects. Surprisingly, over 13% of such police responses culminate in a violent outcome for the subject. Perhaps it's time to explore alternative services staffed by professionals trained in de-escalating situations involving individuals experiencing mental health crises. Such a shift could potentially lead to a substantial reduction in lower-grade incidents.

```
## Percentage of D an F Graded entries with 'Unstable' and Calls to Service: 33.10811 %
## Percentage of 'Call for Service' entries: 41.17468 %
## Percentage of Total Violent Cases Effected by
## Calls of Service and Mentally Unstable Subjects: 13.63216 %
```





Model Evaluations

Linear Regression Model

In my journey to model our dataset, I initially created individual models for each feature and evaluated their performance. Subsequently, I combined some of the successful feature models to create a composite model, which proved to be our most accurate one. Following this, I developed a precog model excluding most of the elements that creates the grading feature, namely injury and hospitalization. Below, you'll find a heat map of our negative correlated officer races, the final successful "combo" and "precog" models, as well as a table encompassing all the linear regression models I applied to the entire one-hot encoded incident dataset, with the goal of predicting our weighted sum feature.

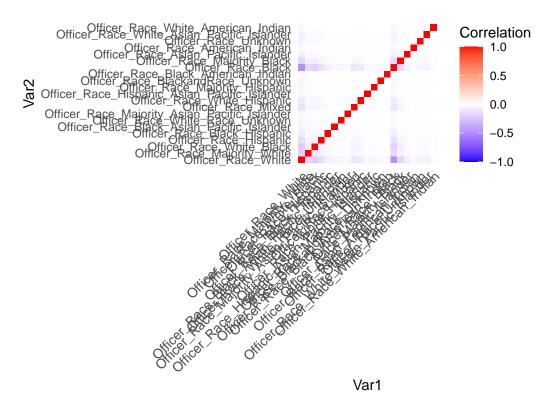
One-Hot Encoding Example

```
# Linear regression model prep
 # One-Hot encoding
 # making one big table where all these features are split up, if I did it over I
 # would probably do percentages for the race and gender features
data <- incident nola use of force %>%
  mutate(
    Officer_Race_White = as.numeric(Officer_Race == "White"),
    Officer Race Majority White = as.numeric(Officer Race == "Majority White"),
    Officer_Race_White_Black = as.numeric(Officer_Race == "White&Black"),
    Officer_Race_Hispanic = as.numeric(Officer_Race == "Hispanic"),
    Officer_Race_Black_Hispanic = as.numeric(Officer_Race == "Black&Hispanic"),
    Officer_Race_Black_Asian_Pacific_Islander = as.numeric(Officer_Race == "Black&Asian_Pacific_Islander")
    Officer_Race_White_Race_Unknown = as.numeric(Officer_Race == "White&Race_Unknown"),
    Officer_Race_Majority_Asian_Pacific_Islander = as.numeric(Officer_Race == "Majority_Asian_Pacific_
    Officer_Race_Mixed = as.numeric(Officer_Race == "Mixed"),
    Officer_Race_White_Hispanic = as.numeric(Officer_Race == "White&Hispanic" ),
    Officer_Race_Hispanic_Asian_Pacific_Islander = as.numeric(Officer_Race == "Hispanic&Asian_Pacific_
    Officer_Race_Majority_Hispanic = as.numeric(Officer_Race == "Majority_Hispanic"),
    Officer_Race_BlackandRace_Unknown = as.numeric(Officer_Race == "Black&Race_Unknown"),
    Officer_Race_Black_American_Indian = as.numeric(Officer_Race == "Black&American_Indian"),
    Officer_Race_Black = as.numeric(Officer_Race == "Black"),
    Officer_Race_Majority_Black = as.numeric(Officer_Race == "Majority_Black"),
    Officer_Race_Asian_Pacific_Islander = as.numeric(Officer_Race == "Asian_Pacific_Islander"),
    Officer_Race_American_Indian = as.numeric(Officer_Race == "American_Indian"),
    Officer_Race_Unknown = as.numeric(Officer_Race == "Race_Unknown"),
    Officer_Race_White_Asian_Pacific_Islander = as.numeric(Officer_Race == "White&Asian_Pacific_Island
    Officer_Race_White_American_Indian = as.numeric(Officer_Race == "White&American_Indian"),
    Officer_Gender_Male = as.numeric(Officer_Gender == "Male"),
    Officer_Gender_MaleandFemale = as.numeric(Officer_Gender == "Male&Female"),
    Officer_Gender_Majority_Male = as.numeric(Officer_Gender == "Majority_Male"),
    Officer_Gender_Female = as.numeric(Officer_Gender == "Female"),
    Officer_Gender_Majority_Female = as.numeric(Officer_Gender == "Majority_Female"),
    Subject_Race_White = as.numeric(Subject_Ethnicity == "White"),
    Subject_Race_BlackandHispanic = as.numeric(Subject_Ethnicity == "Black&Hispanic"),
    Subject_Race_BlackandIndian = as.numeric(Subject_Ethnicity == "Black&Indian"),
    Subject_Race_Black = as.numeric(Subject_Ethnicity == "Black"),
    Subject_Race_WhiteandBlack = as.numeric(Subject_Ethnicity == "White&Black"),
    Subject_Race_BlackandRace_Unknown = as.numeric(Subject_Ethnicity == "Black&Race_Unknown"),
```

```
Subject_Race_Race_Unknown = as.numeric(Subject_Ethnicity == "Race_Unknown"),
Subject_Race_Mixed = as.numeric(Subject_Ethnicity == "Mixed"),
Subject_Race_WhiteandHispanic = as.numeric(Subject_Ethnicity == "White&Hispanic"),
Subject_Race_Hispanic = as.numeric(Subject_Ethnicity == "Hispanic"),
Subject_Race_Majority_Black = as.numeric(Subject_Ethnicity == "Majority_Black"),
Subject_Race_Majority_Hispanic = as.numeric(Subject_Ethnicity == "Majority_Hispanic"),
Subject_Race_Asian = as.numeric(Subject_Ethnicity == "Asian"),
Subject_Race_BlackandIndiane = as.numeric(Subject_Ethnicity == "Black&Indian"),
Subject_Race_Majority_White = as.numeric(Subject_Ethnicity == "Majority_White"),
Subject_Race_Indian = as.numeric(Subject_Ethnicity == "Indian"),
Subject_Gender_Male = as.numeric(Subject_Gender == "Male"),
Subject_Gender_Female = as.numeric(Subject_Gender == "Female"),
Subject Gender MaleandFemale = as.numeric(Subject Gender == "Male&Female"),
Subject_Gender_Majority_Male = as.numeric(Subject_Gender == "Majority_Male"),
Forcetype_Firearm_Exhibited = as.numeric(Use_of_Force_Type == "Firearm (Exhibited)"),
Forcetype_Hands = as.numeric(Use_of_Force_Type == "Hands"),
Forcetype_Force_Escort_Tech = as.numeric(Use_of_Force_Type == "Force (Escort Tech)"),
Forcetype_Head_Strike_No_Wep = as.numeric(Use_of_Force_Type == "Head Strike (No Wep)"),
Forcetype Hands with injury = as.numeric(Use_of_Force_Type == "Hands (with injury)"),
Forcetype_Force_Take_Down = as.numeric(Use_of_Force_Type == "CEW Exhibited/Laser"),
Forcetype_CEW_Exhibited_Laser = as.numeric(Use_of_Force_Type == "Firearm (Exhibited)"),
Forcetype_Canine_Contact = as.numeric(Use_of_Force_Type == "Canine (Contact)"),
Forcetype_Baton_PR_24_NonStrk = as.numeric(Use_of_Force_Type == "Baton/PR-24(NonStrk)"),
Forcetype_VehPursuits_wInjury = as.numeric(Use_of_Force_Type == "VehPursuits w/Injury"),
Forcetype_Takedown_winjury = as.numeric(Use_of_Force_Type == "Takedown (w/injury)"),
Forcetype_Canine_NoBite = as.numeric(Use_of_Force_Type == "Canine (No Bite)"),
Forcetype_Other = as.numeric(Use_of_Force_Type == "Other"),
Forcetype_Handcuffed_Subject = as.numeric(Use_of_Force_Type == "Handcuffed Subject"),
Forcetype_Baton_PR_24_Strike = as.numeric(Use_of_Force_Type == "Baton/PR-24 (Strike)"),
Forcetype_CEW_Deployment = as.numeric(Use_of_Force_Type == "CEW Deployment"),
Forcetype_Force_Defense_Tech = as.numeric(Use_of_Force_Type == "Force (Defense Tech)"),
Forcetype_Firearm_Discharged = as.numeric(Use_of_Force_Type == "Firearm (Discharged)"),
Forcetype_Baton_PR_24_Miss = as.numeric(Use_of_Force_Type == "Baton/PR-24 (Miss)"),
Forcetype_Force_Neck_Holds = as.numeric(Use_of_Force_Type == "Force (Neck Holds)"),
Subject_Influence_None_Detected = as.numeric(Subject_Influencing_Factors == "None detected"),
Subject_Influence_Mentally_Unstable = as.numeric(Subject_Influencing_Factors == "Mentally unstabl
Subject_Influence_Unknown = as.numeric(Subject_Influencing_Factors == "Unknown"),
Subject_Influence_Alchohol = as.numeric(Subject_Influencing_Factors == "Alchohol"),
Subject_Build_Small = as.numeric(Subject_Build == "Small"),
Subject_Build_Medium = as.numeric(Subject_Build == "Medium"),
Subject_Build_Large = as.numeric(Subject_Build == "Large"),
Subject_Build_XLarge = as.numeric(Subject_Build == "XLarge"),
Division_Task_Force = as.numeric(Division == "Task Force" ),
Division_Night_Watch = as.numeric(Division == "Night Watch" ),
Division_C_Platoon = as.numeric(Division == "C Platoon" ),
Division_General_Assignments = as.numeric(Division == "General Assignments" ),
Division_A_Platoon = as.numeric(Division == "A Platoon" ),
Division_Second_Watch = as.numeric(Division == "Second_Watch" ),
Division_Narcotics = as.numeric(Division == "Narcotics" ),
Division_3rd_District = as.numeric(Division == "3rd District" ),
Division_Traffic = as.numeric(Division == "Traffice" ),
Division_4th_District = as.numeric(Division == "4th District" ),
Division_2nd_District = as.numeric(Division == "2nd District" ),
```

```
Division_K9_Section = as.numeric(Division == "K9 Section" ),
Division_Staff = as.numeric(Division == "Staff" ),
Division Bourbon Promenade = as.numeric(Division == "Bourbon Promenade" ),
Division_District_Investigations_Unit = as.numeric(Division == "District Investigations Unit" ),
Division_7th_District = as.numeric(Division == "7th District" ),
Division_8th_District = as.numeric(Division == "8th District" ),
Division_Other_Support = as.numeric(Division == "Task Force" ),
Division 3rd Platoon = as.numeric(Division == "3rd Platoon" ),
Division 6th District = as.numeric(Division == "6th District" ),
Division_Homicide = as.numeric(Division == "Homicide" ),
Division_B_Platoon = as.numeric(Division == "B Platoon" ),
Division Day Watch = as.numeric(Division == "Day Watch" ),
Division_Tactical_Section = as.numeric(Division == "Tactical Section" ),
Division_Investigations = as.numeric(Division == "Investigations"),
Division_1st_Platoon = as.numeric(Division == "1st Platoon" ),
Division_Street_Gang_Unit = as.numeric(Division == "Street Gang Unit" ),
Division_2nd_Platoon = as.numeric(Division == "2nd Platoon" ),
Division_Force_Investigation_Team = as.numeric(Division == "Force Investigation Team" ),
Division_1st_District = as.numeric(Division == "1st District" ),
Division_5th_District = as.numeric(Division == "5th District" ),
Weather_Clear = as.numeric(Weather_Condition == "Clear Conditions"),
Weather_Other = as.numeric(Weather_Condition == "Other"),
Weather_Rainy_Heavy = as.numeric(Weather_Condition == "Rainy Conditions - Heavy"),
Weather_Rainy_Light = as.numeric(Weather_Condition == "Rainy Conditions - Light"),
Weather Rainy Medium = as.numeric(Weather Condition == "Rainy Conditions - Medium"),
Weather_Foggy = as.numeric(Weather_Condition == "Foggy Condition"),
```

Negative Correlation Of White and Black Officer Heat Map



Building a Linear Regression Model

```
modelcombo <- lm(weighted_sum ~ Officer_Gender_Male +</pre>
                   Officer_Gender_MaleandFemale +
                   Officer_Gender_Majority_Male +
                   Officer_Gender_Female +
                   Officer_Gender_Majority_Female + Subject_Race_White +
                   Subject_Race_BlackandHispanic +
                   Subject_Race_BlackandIndian +
                   Subject_Race_Black +
                   Subject_Race_WhiteandBlack +
                   Subject_Race_BlackandRace_Unknown +
                   Subject_Race_Race_Unknown +
                   Subject_Race_Mixed +
                   Subject_Race_WhiteandHispanic +
                   Subject_Race_Hispanic +
                   Subject_Race_Majority_Black +
                   Subject_Race_Majority_Hispanic +
                   Subject_Race_Asian +
                   Subject_Race_BlackandIndiane +
                   Subject_Race_Majority_White +
                   Subject_Race_Indian +
                   Officer_Years_of_Service +
                   Subject Gender Male +
                   Subject_Gender_Female +
```

```
Subject_Gender_MaleandFemale +
                   Subject_Gender_Majority_Male +
                   Subject Age +
                   Subject Influence None Detected +
                   Subject Influence Mentally Unstable +
                   Subject_Influence_Unknown +
                   Subject_Influence_Alchohol +
                   Forcetype_Firearm_Exhibited +
                   Forcetype Hands +
                   Forcetype_Force_Escort_Tech +
                   Forcetype_Head_Strike_No_Wep +
                   Forcetype_Hands_with_injury +
                   Forcetype_Force_Take_Down +
                   Forcetype_CEW_Exhibited_Laser +
                   Forcetype_Canine_Contact +
                   Forcetype Baton PR 24 NonStrk +
                   Forcetype_VehPursuits_wInjury +
                   Forcetype_Takedown_winjury +
                   Forcetype_Canine_NoBite +
                   Forcetype_Other +
                   Forcetype Handcuffed Subject +
                   Forcetype Baton PR 24 Strike +
                   Forcetype_CEW_Deployment +
                   Forcetype_Force_Defense_Tech +
                   Forcetype_Firearm_Discharged +
                   Forcetype Baton PR 24 Miss +
                   Forcetype_Force_Neck_Holds +
                   Use_of_Force_Level +
                   Subject_Injured +
                   Officer_Injured +
                   Subject_Hospitalized +
                   Light_Condition,
                          data = trainData)
modelprecog <- lm(weighted sum ~ Officer Gender Male +
                 Officer Gender MaleandFemale +
                 Officer Gender Majority Male +
                 Officer Gender Female +
                 Officer_Gender_Majority_Female + Subject_Race_White +
                 Subject_Race_BlackandHispanic +
                 Subject_Race_BlackandIndian +
                 Subject_Race_Black +
                 Subject_Race_WhiteandBlack +
                 Subject_Race_BlackandRace_Unknown +
                 Subject_Race_Race_Unknown +
                 Subject_Race_Mixed +
                 Subject Race WhiteandHispanic +
                 Subject_Race_Hispanic +
                 Subject Race Majority Black +
                 Subject_Race_Majority_Hispanic +
```

```
Subject_Race_Asian +
 Subject_Race_BlackandIndiane +
 Subject Race Majority White +
 Subject Race Indian +
  Officer Years of Service +
 Subject_Gender_Male +
 Subject_Gender_Female +
 Subject_Gender_MaleandFemale +
 Subject Gender Majority Male +
 Subject_Age +
 Subject_Influence_None_Detected +
 Subject_Influence_Mentally_Unstable +
 Subject_Influence_Unknown +
 Subject_Influence_Alchohol +
 Forcetype_Firearm_Exhibited +
 Forcetype Hands +
 Forcetype_Force_Escort_Tech +
 Forcetype_Head_Strike_No_Wep +
 Forcetype_Hands_with_injury +
 Forcetype_Force_Take_Down +
 Forcetype_CEW_Exhibited_Laser +
 Forcetype_Canine_Contact +
 Forcetype_Baton_PR_24_NonStrk +
 Forcetype_VehPursuits_wInjury +
 Forcetype_Takedown_winjury +
 Forcetype_Canine_NoBite +
 Forcetype_Other +
 Forcetype_Handcuffed_Subject +
 Forcetype_Baton_PR_24_Strike +
 Forcetype_CEW_Deployment +
 Forcetype_Force_Defense_Tech +
 Forcetype_Firearm_Discharged +
 Forcetype_Baton_PR_24_Miss +
 Forcetype_Force_Neck_Holds +
 Light Condition,
data = trainData)
```

Linear Regression Results

This table displays the performance metrics of various models utilized in the project. These models undergo evaluation using mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared values. The "modelcombo" emerges as the top-performing model, demonstrating the lowest MSE, RMSE, and MAE, highlighting its superior predictive accuracy, this model does incorporates all the features from our "Weighted_Sum" grading feature, which could be contributing to it's success. Conversely, "modelprecog" takes the lead as our best-performing model, encompassing information leading up to potential subject injury or arrest, without relying on any of our "Weighted_sum" feature weights. These metrics yield insights into the effectiveness of different models in predicting the "Weighted_Sum" variable. Additionally, "modelsubhospital" also exhibits strong performance underscoring it's important fit to the data, subsequent models will further show the significance of this feature.

```
## Model MSE RMSE MAE R_squared

## 1 modelcoprace 0.02768092 0.16637584 0.11823943 0.00013

## 2 modelcopgender 0.02757011 0.16604249 0.11938377 0.00413
```

```
## 3
              modelsubrace 0.02772552 0.16650982 0.11999986
                                                               -0.00148
## 4
      modelcopyearsservice 0.02711556 0.16466802 0.11885007
                                                                0.02055
## 5
               modelcopage 0.02760580 0.16614994 0.11962327
                                                                0.00284
## 6
           modelsubinjured 0.01886663 0.13735586 0.10095662
                                                                0.31851
## 7
           modelcopinjured 0.02751311 0.16587076 0.11843239
                                                                0.00619
## 8
          modelsubhospital 0.01191439 0.10915305 0.08291582
                                                                0.56964
## 9
            modelsubgender 0.02775029 0.16658419 0.11976850
                                                               -0.00238
## 10
              modelsubsize 0.02788671 0.16699314 0.11997850
                                                               -0.00730
## 11
              modelesubage 0.02773263 0.16653118 0.12004581
                                                               -0.00174
## 12
            modelsubheight 0.02777485 0.16665789 0.11989537
                                                               -0.00326
##
  13
            modelinfluence 0.02709614 0.16460905 0.11714556
                                                                0.02125
##
  14
            modelforcetype 0.01780154 0.13342240 0.10289688
                                                                0.35699
## 15
            modelfocelevel 0.02465280 0.15701209 0.11336937
                                                                0.10951
## 16
            modeledistance 0.02759809 0.16612673 0.12013485
                                                                0.00312
## 17
             modeldivision 0.02705703 0.16449020 0.11821689
                                                                0.02267
## 18
              modelweather 0.02771409 0.16647550 0.11974036
                                                               -0.00107
       modellightcondition 0.02774611 0.16657162 0.11991340
## 19
                                                               -0.00222
## 20
                modelcombo 0.00505614 0.07110654 0.03915392
                                                                0.81737
                                                                0.39061
## 21
               modelprecog 0.01687068 0.12988718 0.09754244
```

Linear Regression Only on Police Information

If we only use the info from the aggregated police table that could basically be summation of a officer actions up to potential subject injury. We see that the main predictive force is actually the choice of force type. I do not continue this line of prediction further than liner regression considering its similarity to the full incident table's proog model.

```
##
                        Model
                                     MSE
                                              RMSE
                                                          MAE R squared
## 1
             copmodelcoprace 0.03291463 0.1814239 0.1311671
                                                                0.01421
## 2
           copmodelcopgender 0.03320603 0.1822252 0.1314230
                                                                0.00548
## 3 copmodelcopyearsservice 0.03318815 0.1821762 0.1314657
                                                                0.00601
## 4
                copmodelfull 0.02420322 0.1555738 0.1212038
                                                                0.27511
## 5
           copmodelforcetype 0.02486745 0.1576942 0.1225292
                                                                0.25522
## 6
           copmodelfocelevel 0.03155511 0.1776376 0.1268370
                                                                0.05492
```

The RandomForest model

I used the same combo model and precog model from above to do a randomforest model that was no more successful but does give us further insight into node purity. It assesses how well the data in each node is separated based on the chosen feature. In this specific analysis, we can see that features like "Subject_Injured", "Subject_Hospitalized", and "Officer_Years_of_Service" have relatively high node purity values, indicating that these features play a significant role in splitting the data effectively. Conversely, features like "Subject_Race_Majority_White", "Subject_Influence_Alcohol", and "Forcetype_Baton_PR_24_Miss" have low node purity values, suggesting that they may not be strong predictors in the model. "Subject_Hospitalized", in fact is probably the major determining factor for all the high scoring weighted sum so it crucially significant node to classify those high score.

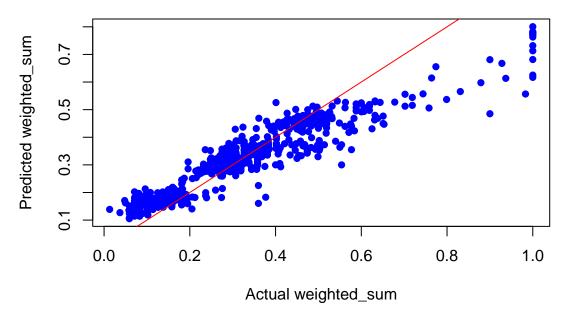
```
## [1] "RandomForest Combo Model Results"
## [1] "Mean Squared Error: 0.00483500804032696"
```

- ## [1] "Root Mean Squared Error: 0.0695342220804042"
- ## [1] "Mean Absolute Error: 0.0474183206793958"
- ## [1] "R-squared: 0.825353481997042"

##		IncNodePurity
##		0.169975905
##	Officer_Gender_MaleandFemale	0.093222991
##	Officer_Gender_Majority_Male	0.093072112
##	Officer_Gender_Female	0.148615661
##	Officer_Gender_Majority_Female	0.016461073
##	Subject_Race_White	0.147507436
##	Subject_Race_BlackandHispanic	0.019317979
##	Subject_Race_BlackandIndian	0.002363881
##	Subject_Race_Black	0.158602215
##	Subject_Race_WhiteandBlack	0.006870851
##	Subject_Race_BlackandRace_Unknown	0.047323553
##	Subject_Race_Race_Unknown	0.056327246
##	Subject_Race_Mixed	0.002974519
##	Subject_Race_WhiteandHispanic	0.000649814
##	Subject_Race_Hispanic	0.053397009
##	Subject_Race_Majority_Black	0.097513540
##	Subject_Race_Majority_Hispanic	0.009313556
##	Subject_Race_Asian	0.016115909
##	Subject_Race_BlackandIndiane	0.001495303
##	Subject_Race_Majority_White	0.000000000
##	Subject_Race_Indian	0.110684881
##	Officer_Years_of_Service	1.090355278
##	Subject_Gender_Male	0.180114668
##	Subject_Gender_Female	0.142964226
##	Subject_Gender_MaleandFemale	0.064492165
##	Subject_Gender_Majority_Male	0.005727235
##	Subject_Age	1.178895892
##	Subject_Influence_None_Detected	0.215937460
##	Subject_Influence_Mentally_Unstable	0.647636999
##	-	0.273519705
##	-	0.000000000
##	Forcetype_Firearm_Exhibited	2.476800219
##	Forcetype_Hands	0.813756613
##	Forcetype_Force_Escort_Tech	0.204147457
##	Forcetype_Head_Strike_No_Wep	0.034265588
##	Forcetype_Hands_with_injury	3.539940548
##	Forcetype_Force_Take_Down	0.942712781
##	Forcetype_CEW_Exhibited_Laser	1.822873436
##	Forcetype_Canine_Contact	0.303195174
	Forcetype_Baton_PR_24_NonStrk	0.015592107
##	Forcetype_VehPursuits_wInjury	0.065983959
	Forcetype_Takedown_winjury	0.669684136
	Forcetype_Canine_NoBite	0.147031563
	Forcetype_Other	0.166030407
	Forcetype_Handcuffed_Subject	0.023826767
	Forcetype_Baton_PR_24_Strike	0.006784225
	·	

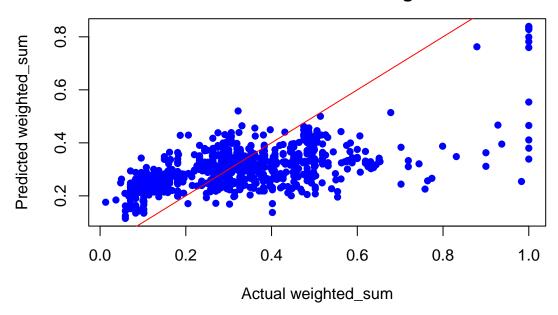
##	Forcetype_CEW_Deployment	0.433929870
##	Forcetype_Force_Defense_Tech	0.038591291
##	Forcetype_Firearm_Discharged	0.252469218
##	Forcetype_Baton_PR_24_Miss	0.00000000
##	Forcetype_Force_Neck_Holds	0.073206021
##	Use_of_Force_Level	2.337120679
##	# Subject_Injured	7.639780470
##	# Officer_Injured	0.576438482
##	# Subject_Hospitalized	18.428148719
##	Light_Condition	0.157533164

RandonForest Predicted vs. Actual Full Model



- ## [1] "RandomForest Precog Model Results"
- ## [1] "Mean Squared Error: 0.0182311147375677"
- ## [1] "Root Mean Squared Error: 0.135022645276886"
- ## [1] "Mean Absolute Error: 0.100884094154202"
- ## [1] "R-squared: 0.341469407770972"

RandonForest Predicted vs. Actual Precog Model

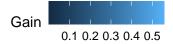


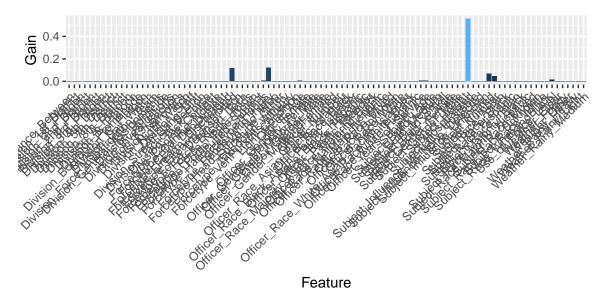
XGBoost Model

XGBoost is an ensemble learning method, which means it combines the predictions from multiple machine learning models to produce a more accurate result. It's based on the gradient boosting framework, where decision trees are built in order to each other, each one correcting the errors of its predecessor. Some key components are gradient boosting, regularization, tree pruning. With this data set XGBoost did a pretty fair job at .92 R2 for the combo model and a .35 R2 for the proog model. Again we also see that the "Subject_Hospitalized" is overwhelmingly the most important predictive feature.

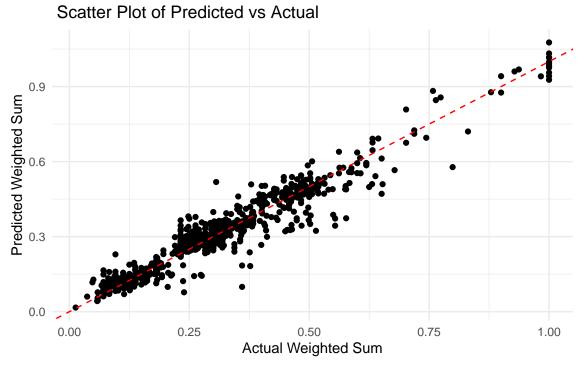
- ## [1] "XGBoost Combo Model Results"
- ## [1] "Accuracy: 0.017293997965412"
- ## [1] "Mean Squared Error: 0.00198976703967447"
- ## [1] "Root Mean Squared Error: 0.0446068048583898"
- ## [1] "Mean Absolute Error: 0.0299196337741607"
- ## [1] "R-squared: 0.928127133974176"

Feature Importance





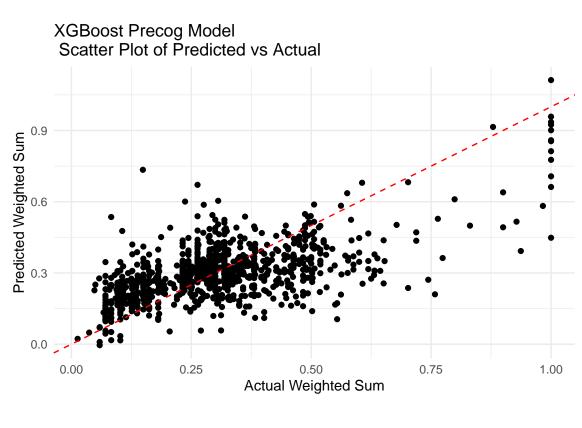
XGBoost Combo Model



- ## [1] "XGBoost Precog Model Results"
- ## [1] "Accuracy: 0.00203458799593082"
- ## [1] "Mean Squared Error: 0.0173420427263479"

```
## [1] "Root Mean Squared Error: 0.131689189861385"
```

- ## [1] "Mean Absolute Error: 0.0979755849440488"
- ## [1] "R-squared: 0.37358379718219"



[1] "tbl_df" "tbl" "data.frame"

[1] "numeric"

Tensor Model Training Example

Here we are using TensorFlow to build and train a neural network model. Overall, this code sets up a feedforward neural network for a regression task, compiles it with appropriate loss and metrics, prints a summary of the model architecture, and then trains the model on the data using the specified parameters. communities. We here to put boots on the ground and organize to enhance citizen participation in ongoing community projects. The model is then sequentially retrained three more times: initially for the precognitive model and then twice more for an expanded data set with a bootstrap model. When we closely examine the training loss, mean absolute error, and mean squared error over epochs in the plots below, we see that both the training and validation metrics consistently remain in close alignment, demonstrating the robustness and generalization capability of our neural network model. This alignment signifies that our model effectively learns from the training data without overfitting, as evidenced by its ability to maintain similar performance on unseen validation data.

In the plots below, you'll notice that the full model exhibits a slightly less congruent performance with the validation data. This observation holds true for both the TensorFlow and Bootstrap expanded dataset models. However, this minor incongruence can be attributed primarily to the presence of a single feature—

"Subject_Hospitalized." When we exclude this particular feature from the model, the training performance aligns more closely with the validation data. This trend holds consistently for both the original TensorFlow combined model and the Bootstrap combined model. The inclusion of "Subject_Hospitalized" appears to introduce a deviation between training and validation, possibly due to its unique characteristics or interactions with other features, Maybe because we weighted it to heavily in feature engineering of out weighted grading feature. My removing this single feature the model does preform significantly less successfully.

```
## Convert input features to TensorFlow Tensor
numberofcoll <- ncol(numeric_matrix_trainData)
numberofcoll</pre>
```

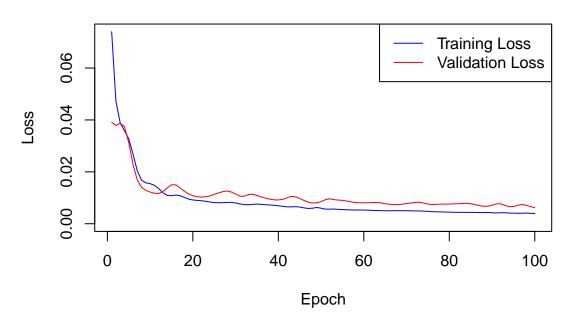
[1] 48

```
x_train <- as.matrix(numeric_matrix_trainData)</pre>
x_train <- tf$convert_to_tensor(x_train, dtype=tf$float32)</pre>
# Convert target variable to TensorFlow Tensor
y_train <- as.vector(trainData$weighted_sum)</pre>
y_train <- tf$convert_to_tensor(y_train, dtype=tf$float32)</pre>
# Create a sequential model
model <- tf$keras$Sequential()</pre>
# Add layers to the model with specifying input_shape
# Input layer: Dense layer with 64 units and ReLU activation function.
# The input shape is (number of coll) defined above, which matches the number of features in the input da
# Hidden layer: Dense layer with 32 units and ReLU activation function.
# Output layer: Dense layer with 1 unit and linear activation function.
# The output shape will be (batch_size, 1).
model <- keras_model_sequential() %>%
  layer_dense(units = 64, activation = "relu", input_shape = c(numberofcoll)) %>%
  layer_dense(units = 32, activation = "relu") %>%
  layer_dense(units = 1, activation = "linear")
 # Compile the model
model$compile(
   loss = "mean_squared_error",
  optimizer = tf$keras$optimizers$Adam(learning_rate = 0.001),
  metrics = c("mean_absolute_error", "mean_squared_error")
)
 # Print the model summary
model$summary()
```

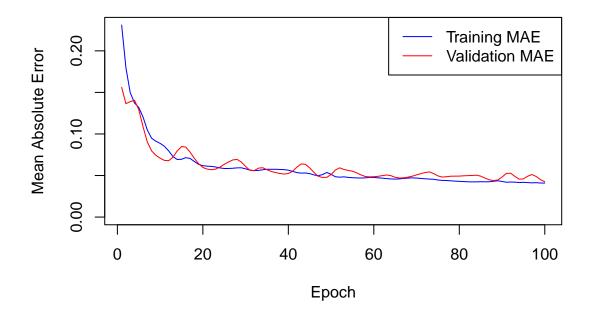
```
## dense 1 (Dense)
                            (None, 32)
                                                      2080
## dense (Dense) (None, 1) 33
## Total params: 5,249
## Trainable params: 5,249
## Non-trainable params: 0
## ______
# Fit the model
history <- model$fit(</pre>
  x = x_{train}
  y = y_train,
  epochs = 100L,
  batch_size = 1824L,
  validation_split = 0.2,
  verbose = 0
)
# Plot training history
 # The training and validation loss show a similar decrease and remain
 # close to each other, it's generally a positive sign. This suggests that
 # our model is not over fitting to the training data
plot(history$history$loss, type = "1", col = "blue", xlab = "Epoch", ylab = "Loss",
     main = "Training Loss Over Epochs", ylim = c(0, max(history$history$loss)))
lines(history$history$val_loss, col = "red")
 legend("topright", legend = c("Training Loss", "Validation Loss"), col = c("blue", "red"), lty = 1)
 # Plot mean absolute error
 # No degrade of Validation showing good potential
plot(history$history$mean_absolute_error, type = "l", col = "blue", xlab = "Epoch", ylab = "Mean Absol
     main = "Mean Absolute Error Over Epochs", ylim = c(0, max(history$history$mean_absolute_error)))
lines(history$history$val_mean_absolute_error, col = "red")
legend("topright", legend = c("Training MAE", "Validation MAE"), col = c("blue", "red"), lty = 1)
 # Plot mean squared error
 # Same here no degrade of Validation
plot(history$history$mean_squared_error, type = "l", col = "blue", xlab = "Epoch", ylab = "Mean Squared"
     main = "Mean Squared Error Over Epochs", ylim = c(0, max(history$history$mean_squared_error)))
lines(history$history$val_mean_squared_error, col = "red")
legend("topright", legend = c("Training MSE", "Validation MSE"), col = c("blue", "red"), lty = 1)
 # Make predictions using the trained model
predicted_train <- model$predict(x_train)</pre>
predicted_train <- as.vector(predicted_train)</pre>
 # Convert y_train and predicted_train to R vectors
y_train <- as.vector(y_train)</pre>
# Pulling in outliers like we did to actual data
```

```
predicted_train <- ifelse(predicted_train > 1, 1, predicted_train)
predicted_train <- ifelse(predicted_train < 0, 0, predicted_train)</pre>
```

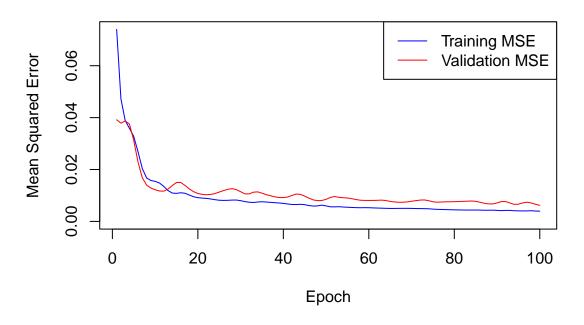
Training Loss Over Epochs



Mean Absolute Error Over Epochs



Mean Squared Error Over Epochs



[1] "Tensorflow Combo Model Results"

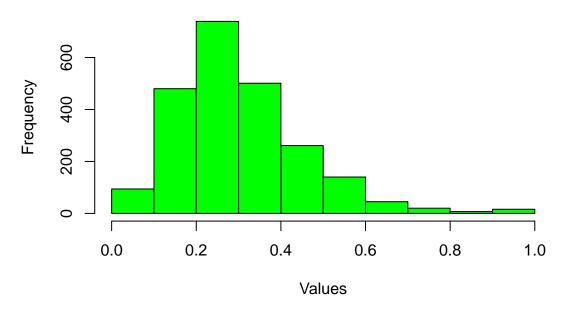
MAE: 0.04171537

MSE: 0.004329549

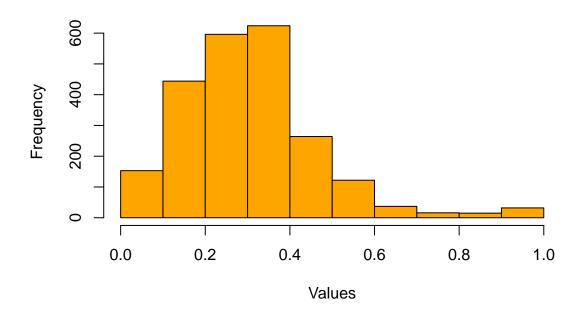
RMSE: 0.06579931

R-squared: 0.8392296

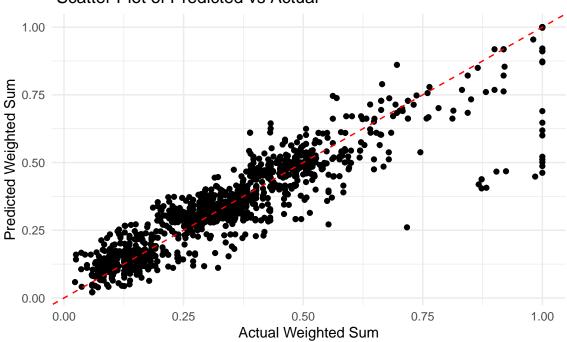
Tensor Full Combo Model Perdicted Weighted Sums



Actual Training Weighted Sum



Tensorflow Combo Model Scatter Plot of Predicted vs Actual



Precognative Tensorflow Model

[1] " Tensorflow Precog Model Results"

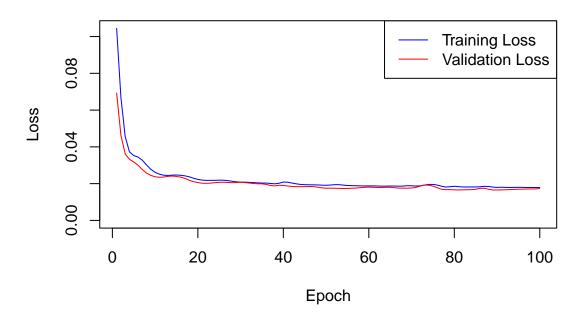
MAE: 0.09553599

MSE: 0.01784927

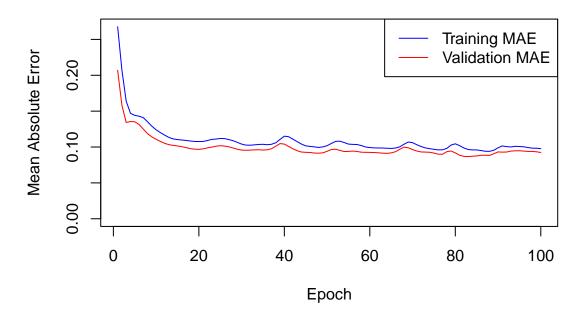
RMSE: 0.1336012

R-squared: 0.337198

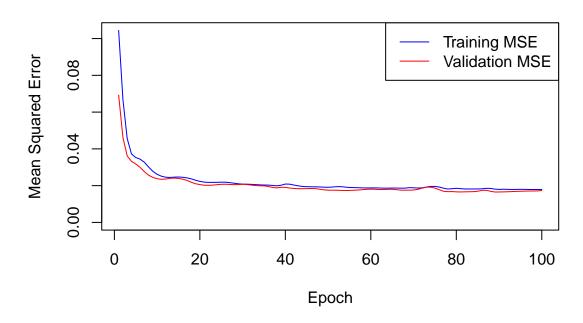
Training Loss Over Epochs



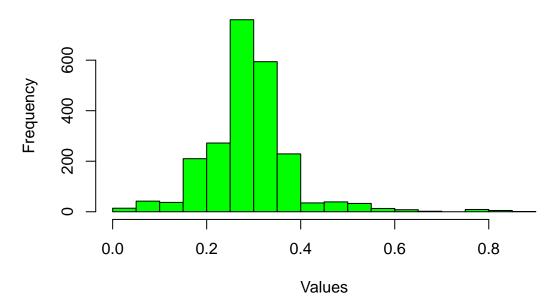
Mean Absolute Error Over Epochs



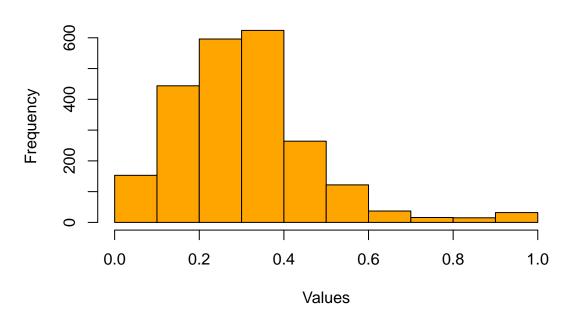
Mean Squared Error Over Epochs



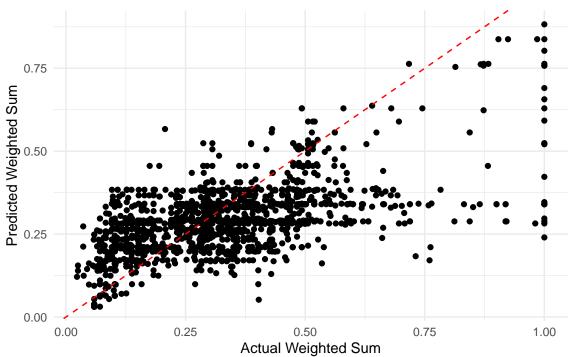
Tensor Precog Model Perdicted Weighted Sums



Actual Training Weighted Sum



Scatter Plot of Predicted vs Actual



Bootstrap Enlarged Tensorflow Combo Model

Deep learning models often demand extensive datasets to harness their complexity effectively. If your dataset is relatively small, linear models might yield similar performance. To explore this further, I conducted a

bootstrapping procedure to augment the dataset size and assess whether the limited data size was a factor affecting performance.

[1] "Bootstrap Enlarged Data Tensorflow Combo Model Results"

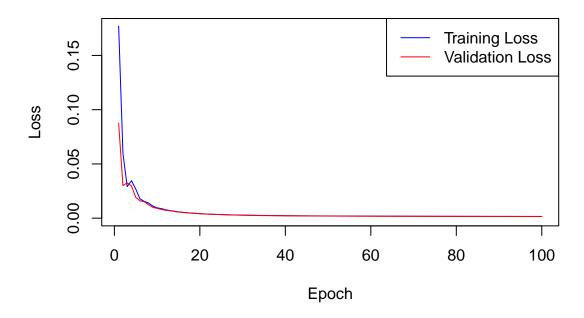
MAE: 0.1714

MSE: 0.0540

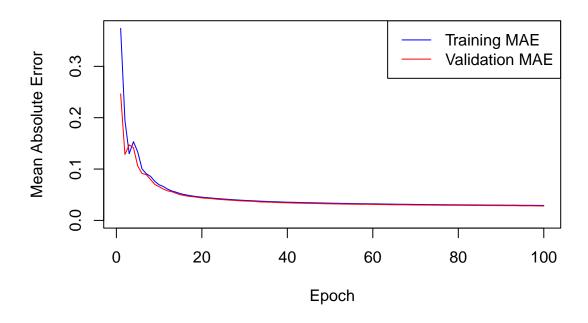
RMSE: 0.2323

R2: 0.9425

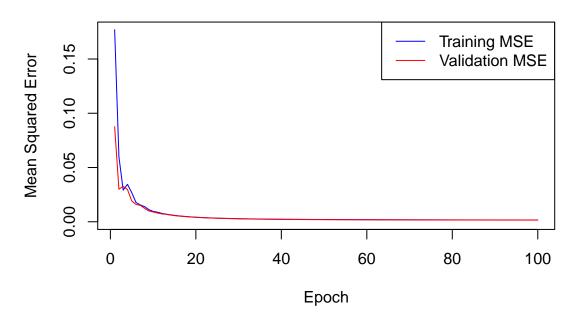
Training Loss Over Epochs



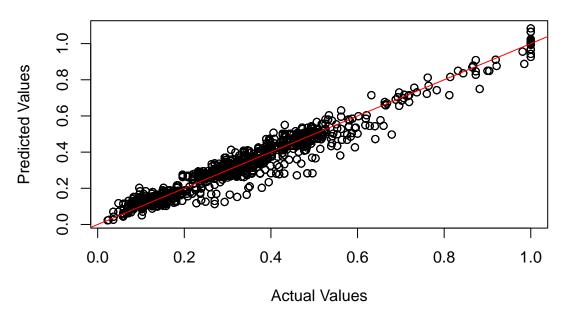
Mean Absolute Error Over Epochs



Mean Squared Error Over Epochs



Bootstrap Tensorflow Combo Model Actual vs. Predicted Values



Bootstrap Enlarging Tensorflow Precog Model

[1] "Bootstrap Enlarged Dataset Tensorflow Precog Model Results"

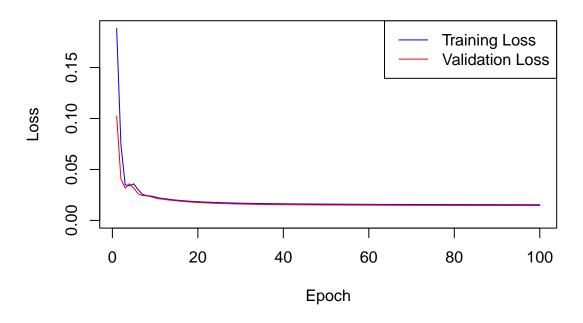
MAE: 0.1459

MSE: 0.0400

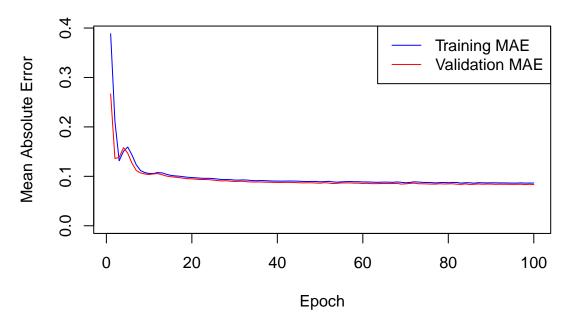
RMSE: 0.2000

R2: 0.4519

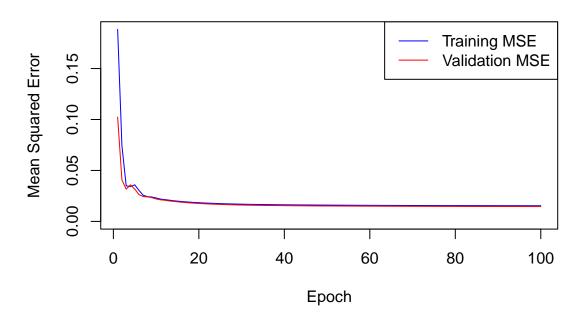
Training Loss Over Epochs



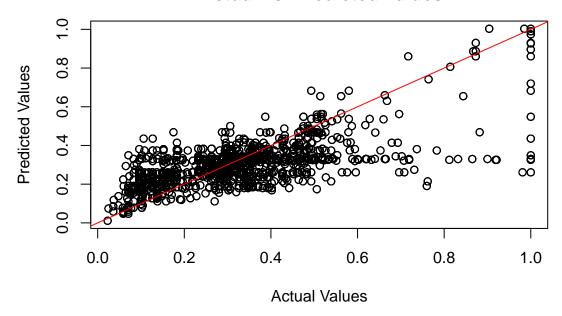
Mean Absolute Error Over Epochs



Mean Squared Error Over Epochs



Bootstrap Tensorflow Precog Model Actual vs. Predicted Values



Final Results Table

Model MSE RMSE MAE R_squared modelcoprace 0.027680922 0.16637584 0.11823943 0.00013

```
## 2
               modelcopgender 0.027570110 0.16604249 0.11938377
                                                                    0.00413
                 modelsubrace 0.027725519 0.16650982 0.11999986
## 3
                                                                   -0.00148
## 4
         modelcopyearsservice 0.027115558 0.16466802 0.11885007
                                                                    0.02055
## 5
                  modelcopage 0.027605804 0.16614994 0.11962327
                                                                    0.00284
##
  6
              modelsubinjured 0.018866631 0.13735586 0.10095662
                                                                    0.31851
              modelcopinjured 0.027513108 0.16587076 0.11843239
##
  7
                                                                    0.00619
             modelsubhospital 0.011914387 0.10915305 0.08291582
## 8
                                                                    0.56964
## 9
               modelsubgender 0.027750294 0.16658419 0.11976850
                                                                   -0.00238
##
  10
                 modelsubsize 0.027886709 0.16699314 0.11997850
                                                                   -0.00730
##
  11
                 modelesubage 0.027732633 0.16653118 0.12004581
                                                                   -0.00174
##
  12
               modelsubheight 0.027774851 0.16665789 0.11989537
                                                                   -0.00326
               modelinfluence 0.027096139 0.16460905 0.11714556
##
  13
                                                                    0.02125
##
  14
               modelforcetype 0.017801536 0.13342240 0.10289688
                                                                    0.35699
               modelfocelevel 0.024652797 0.15701209 0.11336937
## 15
                                                                    0.10951
## 16
               modeledistance 0.027598090 0.16612673 0.12013485
                                                                    0.00312
##
  17
                modeldivision 0.027057025 0.16449020 0.11821689
                                                                    0.02267
                 modelweather 0.027714093 0.16647550 0.11974036
##
  18
                                                                   -0.00107
##
  19
          modellightcondition 0.027746106 0.16657162 0.11991340
                                                                   -0.00222
                   modelcombo 0.005056140 0.07110654 0.03915392
##
  20
                                                                    0.81737
##
  21
                  modelprecog 0.016870680 0.12988718 0.09754244
                                                                    0.39061
##
  22
                Random Forest 0.004835008 0.06953422 0.04741832
                                                                    0.82535
## 23
         Random Forest Precog 0.018231115 0.13502265 0.10088409
                                                                    0.34147
## 24
                      XGBoost 0.001989767 0.04460680 0.02991963
                                                                    0.92813
               XGboost Precog 0.017342043 0.13168919 0.09797558
##
  25
                                                                    0.37358
##
  26
                   Tensorflow 0.004329549 0.06579931 0.04171537
                                                                    0.83923
##
  27
            Tensorflow Precog 0.017849272 0.13360117 0.09553599
                                                                    0.33720
  28
        Bootstrap Tensor Full 0.053970240 0.23231496 0.17141679
##
                                                                    0.94246
  29 Bootstrap Tensor Precog 0.039983079 0.19995770 0.14587291
                                                                    0.45189
```

Conclusion

The Exploration of Data

In summary, several key findings have emerged from the analysis of incidents involving the use of force by the NOPD. First, there is a pressing concern regarding the response to calls involving mentally unstable subjects. Over 13% of such police responses result in violent outcomes, suggesting a need to explore alternative, professionally trained services for de-escalation. Additionally, addressing factors like alcohol consumption, which contributes to nearly 13% of failed incidents, could lead to a significant reduction in lower-grade incidents.

The type of force used and its correlation with subject injuries has also raised questions, particularly in cases involving the exhibition of firearms. While there isn't a direct one-to-one relationship, the high number of injuries associated with firearm exhibition demands attention. Female officers tend to use firearms less frequently than their male counterparts, suggesting potential adjustments in incident counts involving male officers based on their representation within the police force.

Moreover, disparities in gender and racial representation within the police force compared to incident data have been identified. Adjusting incident counts based on these disparities could potentially lead to reductions in violent incidents involving male and white officers. While some data normalization is needed, these findings highlight important areas for further analysis and consideration in addressing incidents involving the use of force by the NOPD.

The Various Models

Linear Regression Models: The linear regression models generally performed consistently but had limited predictive power, with R-squared values mostly close to zero, indicating that these single feature models struggled to explain the variance in the data. Once all the semi-successful models were combined together though, linear regression alone was able to predict fairly well the "Weighted Sum" feature.

Random Forest: The Random Forest model showed improved performance with a lower MSE, RMSE, and MAE, suggesting better predictive accuracy. Its R-squared value of 0.82535 indicates a stronger correlation with the target.

XGBoost: XGBoost outperformed other models with an impressively low MSE, RMSE, and MAE. Its high R-squared value of 0.92813 signifies a strong predictive ability, making it a standout performer.

TensorFlow: Your TensorFlow model also performed well, with relatively low errors and an R-squared value of 0.85779, demonstrating its effectiveness in capturing patterns in your data.

Bootstrap Techniques: The Bootstrap Tensor Full and Bootstrap Tensor Precog models, while having higher errors compared to some other models, achieved remarkably high R-squared values, indicating strong predictive capabilities.

In conclusion, the XGBoost and TensorFlow models emerged as the top performers, showcasing their ability to predict the "Weighted Sum" feature effectively. These models are promising for future predictions, while ensemble techniques like Bootstrap also demonstrated potential for improved predictions with larger datasets. This project began with thorough data cleaning and preprocessing to ensure data quality. Extensive exploratory data analysis provided valuable insights, guiding feature engineering and selection. Using TensorFlow, we constructed a feedforward neural network model for regression, carefully considering model architecture and evaluation metrics. During training, we closely monitored training loss, mean absolute error, and mean squared error. This project underscores the importance of data preprocessing, feature engineering, and model evaluation in optimizing regression and deep learning models, highlighting the significant impact of individual features on model behavior.