Investigating the Relationship Between Demographic Factors, Strip Searches, and Severity of Crime

Group 53:

Teresa Lau

1001782314

teresacm.lau@mail.utoronto.ca

Maidah Shah

999859014

maidah.shah@mail.utoronto.ca

INF2178: Experimental Design for Data Science

Final Project

Dr. Shion Guha

Faculty of Information, University of Toronto

Master of Information

April 16, 2023

TABLE OF CONTENTS

1. INTRODUCTION **4**

1.1 Research Context and Motivation **4**

1.2 Literature Review **4**

1.3 Research Objectives **8**

2. EXPLORATORY DATA ANALYSIS (EDA) **9**

2.1 Selection of Variables **9**

2.2 EDA Through Data Visualization **9**

2.3 EDA Through Descriptive Statistics **15**

2.4 EDA Through T-tests **17**

3. METHODS **21**

3.1 Dataset Description **21**

3.2 Exploratory Data Analysis **21**

3.2.1 Preliminary Phase: Selection of Variables and Data Visualization **21**

3.2.2 Transformation of Variables **21**

3.2.3 T-tests **24**

3.3 Power Analysis **24**

3.4 ANCOVA **25**

3.5 Logistic Regression **26**

4. RESULTS **28**

4.1 Power Analysis **28**

4.2 ANCOVA **29**

4.1.1 Overview **29**

4.1.2 ANCOVA Test 1 **30**

4.2.3 ANCOVA Test 2 **31**

4.3 Logistic Regression **32**

4.3.1 Overview **32**

4.3.2 Logistic Regression Results **32**

4.3.3 Evaluate Logistic Regression Model Performance **33**

5. DISCUSSION **35**

6. CONCLUSION **38**

6.1 Overview of Findings **38**

6.2 Limitations **39**

7. REFERENCES **41**

1. INTRODUCTION

1.1 Research Context And Motivation

Racial, sex, and age based disparities in crime statistics have become a growing area of concern, and in turn, a growing area of interest for academic studies. Understanding how the race, sex, and age of an offender is correlated to whether they are strip searched, can help to determine if there are biases present in law enforcement decision making, which can lead to inequitable treatment and outcomes for certain groups of people, and result in their overrepresentation in the criminal justice system. It is also important to understand how these demographic factors impact the severity of crime, which can help to assess whether certain racial, sex, or age groups are at a higher risk of committing certain crimes.

Knowledge of factors that put individuals at a higher risk of being strip searched can be leveraged to develop training for law enforcement agencies to reduce biases in decision making. These insights can also be used to develop targeted mitigation measures for use in crime prevention work for at-risk groups. This is important work to advance and ensure a more equitable approach to public safety, while acknowledging and addressing the disadvantages faced by at-risk groups.

1.2 Literature Review

Demographic Factors and Strip Searches

Strip searches remain a highly controversial and debated topic amongst law enforcement agencies and advocates for fairer policing. It can be argued that the decision to strip search an individual is highly subjective on the part of a police officer. As such, it is critically important that officers are equipped with the right knowledge and tools to enable fair, and evidence-based decision making.

Monika Lemke, a researcher in socio-legal studies at York University who focuses on police-led strip searches in Canada, has argued that the Toronto Police's excessive use of strip searches has unfairly targeted Black people since the 1990s, despite their smaller representation amongst the population (2022). Similarly, a report by the University College London looking at police strip search data between 2018 to 2022 in England and Wales found that Black children were eleven times more likely to be strip searched than White children

(Dodd, 2023). The report also found that 38% of children strip searched were Black, despite only making up less than 6% of the population. On the other hand, 42% of children strip searched were White, despite making up more than 70% of the population (Dodd, 2023).

Additional data released by the Metropolitan Police in the United Kingdom looking at the race, sex, and age of detainees who were strip searched between 2019 to 2021, found that across all three years, the greatest number of strip searches occurred amongst White offenders, followed closely in number by Black offenders (*Age, gender and ethnicity*, n.d.). Additionally, males accounted for 90% of all strip searches, while adults (18 years or older) accounted for 94% of all strip searches. In particular, the most strip searches occurred amongst the age groups of 21-30, followed by 31-40. And amongst youth (17 years or younger), those aged 16-17 accounted for 73% of all strip searches (*Age, gender and ethnicity*, n.d.).

Overall, research has shown that an individual is at higher risk of being strip searched if they are Black, male, or a young adult.

Demographic Factors and Severity of Crime

Motivated by communal safety and well-being, research on predictors of crime have taken an interdisciplinary approach by applying a sociological, biological, law enforcement, and social justice lens, to ensure a holistic approach to uncovering and addressing the root cause of issues that lead offenders to commit crimes (Ulmer & Steffensmeier, 2015). A 2018 study by the U.S. Department of Justice, which took a look at the race of violent crime offenders in proportion to the U.S. population, found that Black people were overrepresented amongst offenders in non-fatal violent crimes (Beck, 2021). Of all crimes, almost half of all offenders in robberies (51%) and one third of all offenders for aggravated assault (34%) were Black. It was also found that White people were underrepresented amongst offenders in non-fatal violent crimes. Of all crimes, almost half of all offenders for aggravated assault (45%) and one third of all offenders in robberies (31%) were White. Hispanic people were represented proportionally to their representation in the population in serious, non-fatal, violent crimes, but underrepresented amongst offenders for simple assault. Asians were also underrepresented except for crimes involving rape or sexual assault (Beck, 2021). Overall, it can be seen that both Black and White people were involved in more non-fatal violent crimes than other racial groups.

Additionally, studies by both the University of New Brunswick and Public Safety Canada have found that one of the most consistent results across research on crime and sex, is that females commit significantly less crime than males across all categories of crime (Meares, n.d.; *Incorporating Gender*, 2022). As well, when females do commit crimes, they are typically less severe in nature, such as theft, fraud, and drug violations, and usually against known victims (*Incorporating Gender*, 2022). Another argument made by studies is that higher levels of testosterone in men than women lead to aggression and violence, which in turn, results in males committing more crimes (Ulmer & Steffensmeier, 2015).

In the book, *The Nurture Versus Biosocial Debate in Criminology: On the Origins of Criminal Behavior and Criminality*, Jeffery Ulmer and Darrell Steffensmeier (2015), Professors of Sociology and Criminology at Pennsylvania State University, authored a chapter on the relationship between age and crime. From their scan of major literature on the topic, they highlighted age as being one of the strongest predictors of crime, a trend that is consistent with earlier studies on criminology dating as far back as the nineteenth-century. At a high-level, they found that a majority of studies proposed that crime peaks in early adolescence or early adulthood, and declines thereafter. Several studies attributed this trend to biological factors, such as developments in the prefrontal cortex during the early twenties, which can improve functioning, reasoning, and impulse control, which in turn, would reduce the likelihood of committing crimes. However, sociologists argue that while biological factors do play a role, they must also be considered in context of the social structure and culture of the environment . Studies by social scientists have found that trends that show a decline in criminality into adulthood can be attributed to a setting where youth have greater access to legitimate opportunities as they transition into adulthood. However, this is not the case for all populations, as minority groups tend to face more barriers as a result of systematic inequalities which hinder their transition into adulthood (Ulmer & Steffensmeier, 2015).

In one study highlighted in the chapter, it was found that adult offending levels were higher in Black versus White adults (Ulmer & Steffensmeier, 2015). Applying a social science lens, it was argued that this trend could be attributed to Black youth having limited access to legitimate adult jobs, leaving them more likely to partake in criminal activity. It was also found that age is correlated with the severity of crime. Typically, younger offenders are more likely to commit crimes with less severe consequences, such as vandalism, petty theft, robbery, and drug violations. More severe, public, and sophisticated crimes like homicide, aggravated assault, drinking and driving, and fraud, have age distributions closer to the late twenties to thirties. Trends amongst older populations show a shift towards less visible and less physical crimes, such as bookie, fence, or involvement in criminal enterprises. As mentioned earlier, studies also argue that testosterone levels, which are typically higher in young and middle aged males, can explain the physical and severe nature of crimes that they commit, which subdue as men get older and testosterone levels begin to decrease (Ulmer & Steffensmeier, 2015). As such, the intersection of race and age can also be considered as a predictor of criminality.

Despite the trends being observed in these studies, there are always exceptions to the norm. Some researchers have argued that crime peaks across the ages of fifteen to thirty four, whereas others have proposed fifteen to fifty as the lifespan of criminal activity for an offender (Ulmer & Steffensmeier, 2015). Another exception to the norm is major shifts in social and cultural environments. More recently, the COVID-19 pandemic and resulting lockdowns which started in 2020, presented such an instance. Statistics Canada found that police-reported crime decreased eight percent in the first year of the pandemic (*After Five Years*, 2021). At the same time, the combined severity and volume of non-violent crime decreased ten percent. However, this was not consistent across all categories of crime, as hate crime increased thirty-seven percent in the same year (*After Five Years*, 2021).

Overall, there is abundant research on the correlation between the occurrence and severity of crime, race, sex, and age. However, these should be considered alongside confounding factors as presented by biological and sociological changes when being considered as predictors of criminality.

Strip Searches and Severity of Crime

While there was no research found showing an explicit correlation between strip searches and severity of crime, it should be considered that most offenders who are strip searched are done so given a police officer’s belief that they are carrying drugs or weapons. Should these items be found, they would most likely be charged with a contraband related crime, which is one of the less severe forms of crime. However, it would be interesting to investigate whether the number of strip searches is indicative of severity of crime. For example, looking at whether offenders who commit severe crimes have ever been strip searched, or whether those who are strip searched are repeat offenders. However, in the absence of this research and data, this report will focus primarily on the effects of demographic factors on strip searches and the severity of crime.

1.3 Research Objectives

Leveraging the findings from existing literature, this study will aim to further investigate the relationship between demographic factors, strip searches, and severity of crime (rated by the maximum sentence one can receive if convicted). A dataset published by the Toronto Police Service was used to investigate the following research questions:

Research Question 1: Does the severity of crime vary by the occurrence of strip searches, while controlling for the demographics of an offender?

Research Question 2: Do the demographics of an offender predict the occurrence of strip searches?

2. EXPLORATORY DATA ANALYSIS

This section of the report will outline the steps undertaken for the exploratory data analysis (EDA). It will provide justification for the variables that were selected for the study, and include the various data visualizations that were developed, along with details around what data was included in each figure and why, and provide interpretations to highlight key information and trends from each plot. This section will also outline the descriptive statistics, and the hypotheses and results of the t-tests that were undertaken in the study.

2.1 Selection of Variables

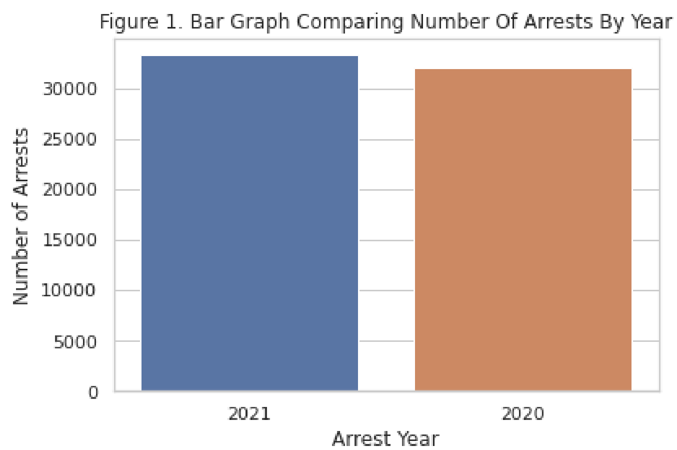
For the first step of the exploratory data analysis, it was decided to streamline the variables of initial interest to: Arrest Year, Arrest Month, Perceived Race, Sex, Age Group At Arrest, Youth At Arrest Under 18 Years, Arrest Local Division, Occurrence Category, and Strip Search. Year and month were selected to enable observation of trends over time, such as seasonality, as related to the number of arrests. Demographic factors, such as race, sex, age, and youth, were selected to enable observation of whether certain groups of people were at higher risk of arrest. Similarly, local division was selected to enable observation of whether people from certain communities or neighbourhoods were at higher risk of arrest. Both occurrence category and strip search in combination with the mentioned predictors were selected to enable observation of whether certain groups (based on race, sex, age, youth, and location) were at higher risk of being strip searched or committing certain crimes.

Variables including identification numbers were not selected as they did not provide meaningful information. However, while variables such as, Booked, Actions At Arrest, Search Reason, and Items Found, could be used for analysis, they were not included, as it was decided to narrow the scope of the study to the correlation between demographic factors (race, sex, and age), strip searches, and severity of crime.

2.2 EDA Through Data Visualization

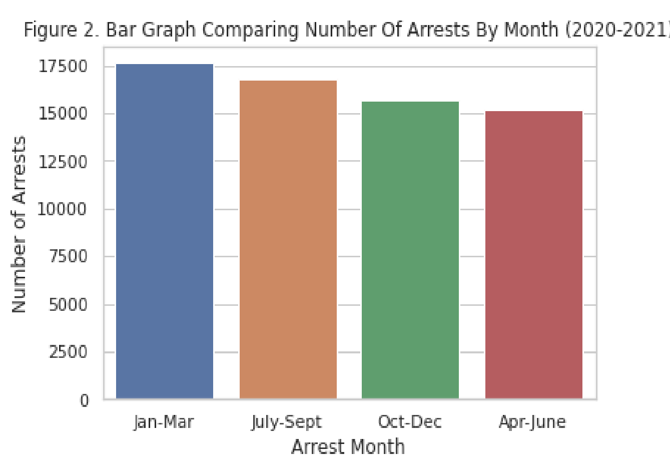
The purpose of the preliminary EDA in this study was to determine if the number of initial variables of interest could be narrowed before conducting t-tests. This was accomplished by developing bar plots for each variable of interest to observe if there were any trends that would be meaningful to explore further. If not, the variables would not be considered for further analysis.

Figure 1. Bar Graph Comparing Number Of Arrests By Year



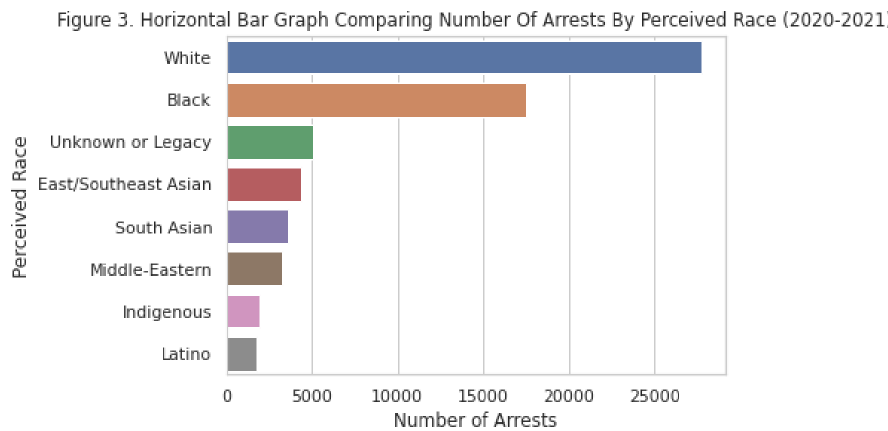
This graph shows that there was not a significant difference between the number of arrests from 2020-2021. However, there was a slight uptick in the number of arrests in 2021. Additionally, as the data in this study was limited to a time period of two years, this limits the analysis. As the COVID-19 pandemic and resulting lockdowns were occurring at this time, as a next step, it would be interesting to analyze data prior to and after this time, to observe if the trends noted by Statistics Canada (as mentioned in the above literature review) hold true and show a decrease in crime in 2020, and observe whether crime increased after lockdown restrictions were lifted. However, these inferences could not be made with this dataset, so arrest year was not explored further as a variable.

Figure 2. Bar Graph Comparing Number of Arrests By Month (2020-2021)



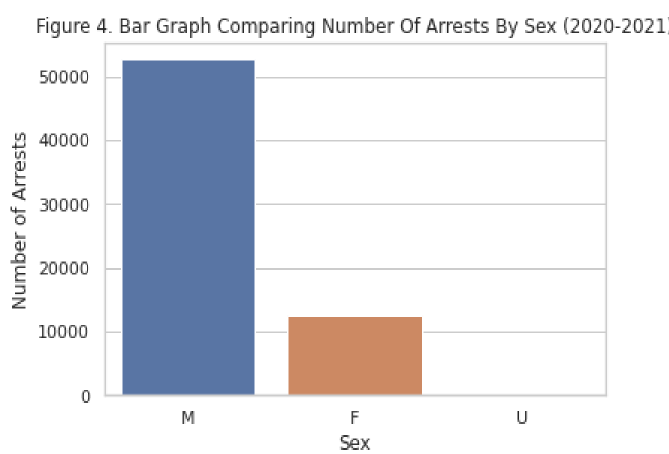
This graph provides an additional layer to the year data presented in Figure 1, by breaking the time down further into months. It appears that there were no significant differences between each quarter. While there were slightly more arrests in the Winter and Summer quarters, it does not appear that the number of arrests followed seasonal trends. As such, arrest month was not explored further as a variable.

Figure 3. Horizontal Bar Graph Comparing Number Of Arrests By Perceived Race (2020-2021)



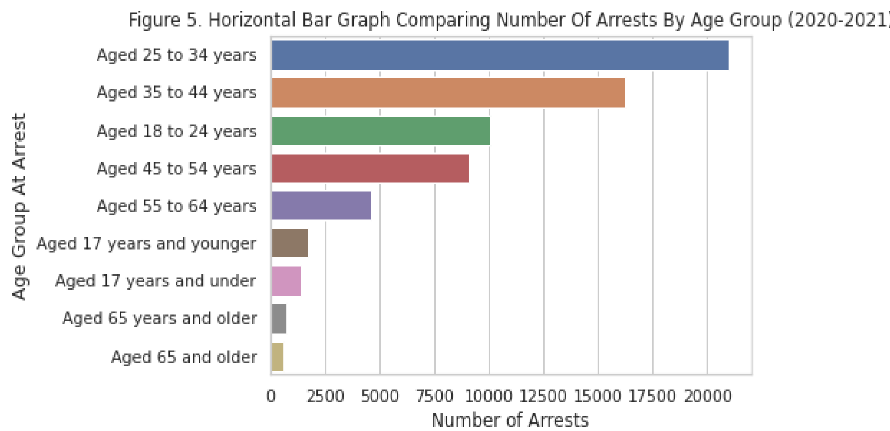
This graph shows that White offenders accounted for the greatest number of arrests, followed by Black offenders. As well, these two races accounted for significantly more arrests than any other race, even if the number of arrests amongst all other races were combined. As there was a major discrepancy between race and number of arrests, perceived race was selected to be explored further as a variable.

Figure 4. Bar Graph Comparing Number Of Arrests By Sex (2020-2021)



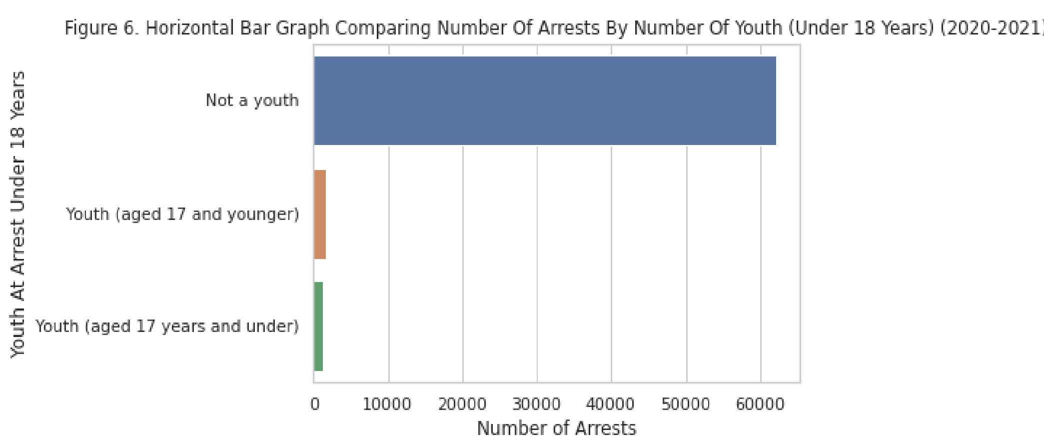
This graph shows that male offenders accounted for significantly more arrests than females. As well, the number of arrests amongst those who did not identify with "male" or "female" were extremely small and negligible. As there was a major discrepancy between sex and number of arrests, sex was selected to be explored further as a variable.

Figure 5. Horizontal Bar Graph Comparing Number Of Arrests By Age Group (2020-2021)



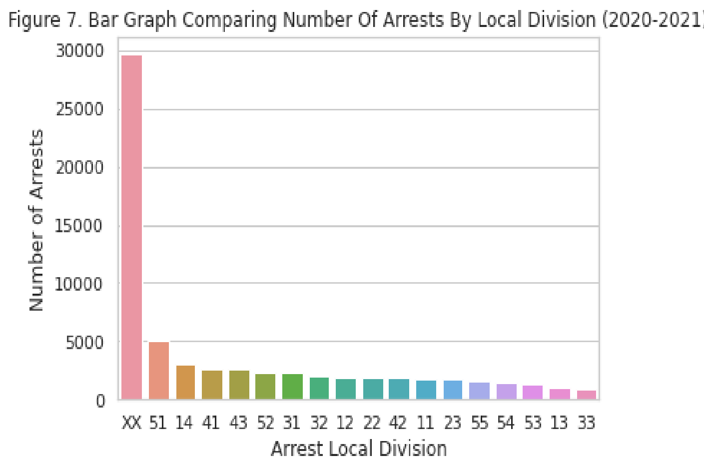
This graph shows that most arrests occur amongst those aged 25-34 years, followed by those aged 35-44 years. There also appears to be a trend that more arrests occur amongst adults (18+ years) than youth. As well, it can be seen that those 17 years and under are grouped in different categories between the two years, which will need to be considered and consolidated into one group for further analysis. Overall, as there was a major discrepancy between age and number of arrests, age group was selected to be explored further as a variable.

Figure 6. Horizontal Bar Graph Comparing Number Of Arrests By Number Of Youth (Under 18 Years) (2020-2021)



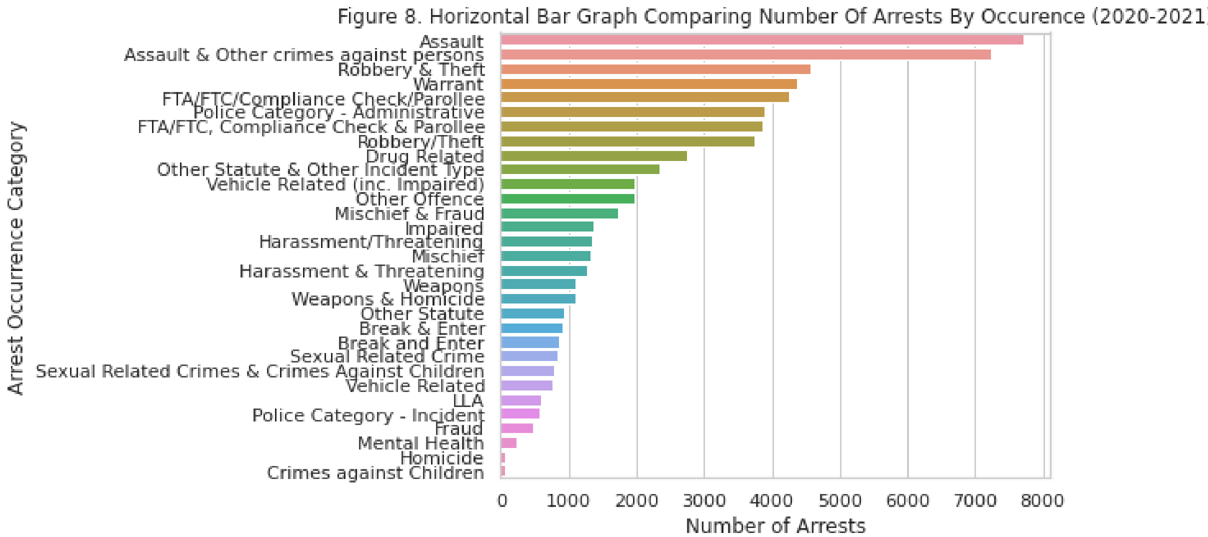
This graph provides an additional layer to the data presented in Figure 5 by consolidating the age groups into two broader categories. Similar to the findings from that plot, there appears to be a trend that more arrests occur amongst adults (18+ years) than youth. However, as both figures 5 and 6 are looking at the same data, further analysis will focus on the age group, rather than whether an offender was a youth or not, as the larger number of groupings enables more insight into trends by age.

Figure 7. Bar Graph Comparing Number Of Arrests By Local Division (2020-2021)



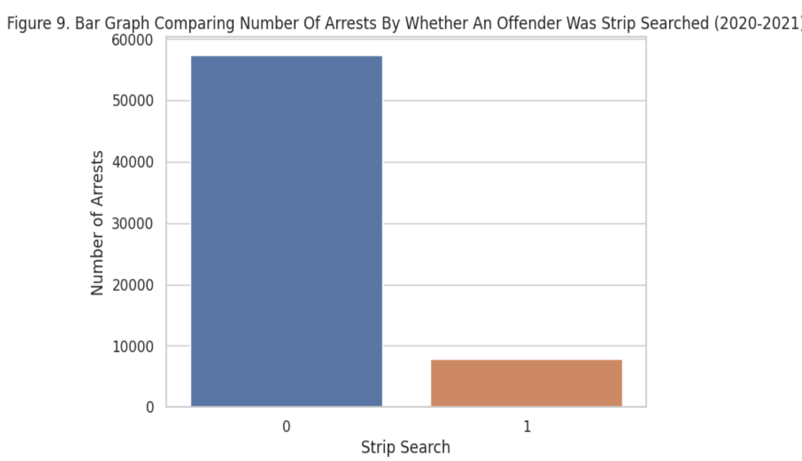
This graph shows that most arrests do not specify the local division where they occur, which does not provide much insight. As well, the major discrepancy between unidentified divisions versus all other divisions skews the graph. If removed, it would appear that most arrests occur in local division 51. Overall, there does not appear to be much meaningful information that can be extracted from this plot, at least for the purpose of this study, which is looking at more prominent trends. As such, local division was not explored further as a variable.

Figure 8. Horizontal Bar Graph Comparing Number Of Arrests By Occurence (2020-2021)



This graph shows that most arrests occurred for assaults, and overall, for less severe and non-fatal crimes, such as robbery, theft, and drug related incidents. As well, it can be seen that several occurrences are duplicated and in categories with slightly different names across the two years, which will need to be considered and consolidated accordingly for further analysis. Overall, as there is a major discrepancy between occurrence and number of arrests, occurrence will be explored further as a variable.

Figure 9. Horizontal Bar Graph Comparing Number Of Arrests By Whether an Offender Was Strip Searched (2020-2021)



This graph shows that significantly more arrests occurred without strip searches than with them (0=no, 1=yes). Overall, as there is a major discrepancy between strip searches and number of arrests, strip search will be explored further as a variable**.**

2.3 EDA Through Descriptive Statistics

Severity of Crime

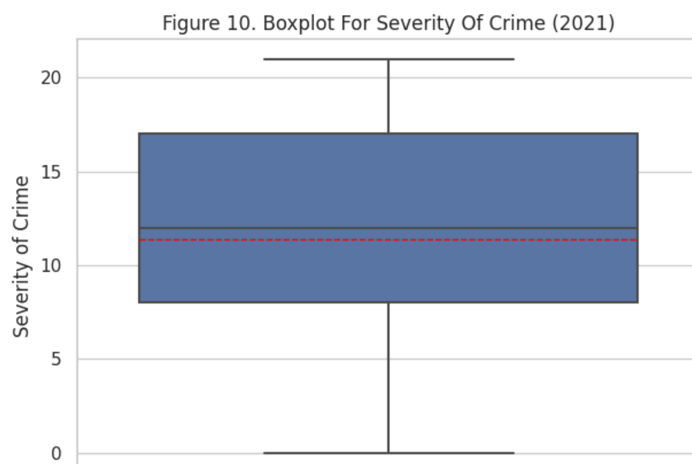
For the next step of the exploratory data analysis, it was decided to investigate the effects of demographic factors, including race, sex, and age, on the severity of crime. In order to do so, the discrete variable, occurrence category, was transformed into a continuous outcome variable, which measured the severity of crime. This process is further detailed in the methods section of this report. Descriptive statistics and a boxplot were then produced for the outcome variable, severity of crime, to extract key information on the variable, as outlined in both *Table 1* and *Figure 10*.

Table 1. Descriptive Statistics For Severity Of Crime (2021)*

count	33294
mean	11.37
std	5.40
min	0
25%	8
50%	12
75%	17
max	21

*Rounded to the nearest hundredth

Figure 10. Box plot for Severity of Crime Score (2021)

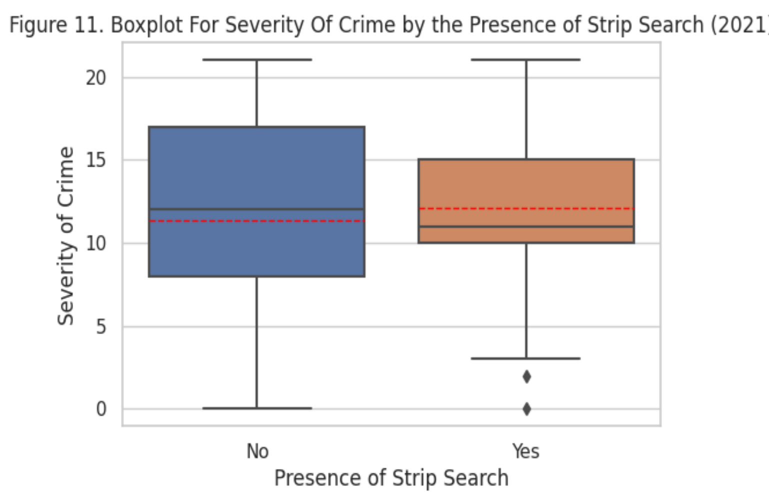


From both *Table 1* and *Figure 10*, it can be seen that in 2021, there were 33,294 crimes that led to an arrest. On average, these crimes ranked 11.37 on the severity scale, where 11 denotes a crime as severe as a drug related offense, and 12 denotes a crime as severe as a vehicle related offense. Of all arrests, the most severe crime committed was homicide, as denoted by a severity level of 21. The mean, as represented by the dotted red line in the box plot, appears lower than the median. The bottom whisker of the boxplot is also longer than the top. As such, it can be interpreted that the distribution of arrests is negatively skewed. In other words, this means that there were more crimes of lower severity that led to arrests in 2021. Also, as the boxplot is relatively long, this means that over all arrests, there was an extensive range of crime severity. 75% of all arrests in 2021 were a result of crimes below a severity level of 17, where 17 denotes a crime as severe as assault. 50% of all arrests were a result of crimes below a severity level of 12, where 12 denotes a crime as severe as a vehicle related offense. And 25% of all arrests were a result of crimes below a severity level of 8, where 8 denotes a crime as severe as an FTA, FTC, compliance check, or parolee related offense.

Strip Search and Severity of Crime

Next, it was decided to investigate the effects of whether an individual is strip searched on the severity of crime. In order to do so, the numerical dichotomous variable, strip search, was transformed into a categorical variable, where 0=no and 1=yes to indicate the presence of a strip search. This process is further detailed in the methods section of this report. A boxplot was then produced for the interaction of strip search and the outcome variable, severity of crime, to extract key information on the relationship, as shown in *Figure 11*.

Figure 11. Box plot for Severity of Crime by the Presence and Absence Of A Strip Search At The Time Of Arrest (2021)



From *Figure 11*, it can be seen that the mean, as represented by the dotted red line in the box, appears lower than the median for the “no” plot. The bottom whisker of the plot is also longer than the top. As such, it can be interpreted that the distribution of arrests is negatively skewed. In other words, this means that there were more crimes of lower severity where no strip search was conducted that led to arrests in 2021. Also, as the box is relatively long, this means that over all arrests where no strip search was conducted, there was an extensive range of crime severity. For the “yes” plot, it can be seen that the mean is higher than the median. As such, it can be interpreted that the distribution of arrests is positively skewed. In other words, this means that there were more crimes of higher severity where a strip search was conducted that led to arrests in 2021. Also, as the box is shorter, this means that over all arrests where a strip search was conducted, there was less variability in crime severity. There also appear to be two outliers in the plot at the lower end of severity, which makes sense as the distribution of arrests is positively skewed.

2.4 EDA Through T-tests

For the final step of the exploratory data analysis, three t-tests were conducted between the predictor variables, race, sex, and age, and the outcome variable, severity of crime. This step was undertaken to determine which of the variables were statistically significant predictors of severity of crime, which would be used to inform what variables to include in the ANCOVA tests as “blocking” factors. An additional t-test was conducted after these tests between the predictor variable, strip search, and the outcome variable, severity of crime to determine if strip search was a statistically significant predictor of crime severity for inclusion in the ANCOVA tests.

T-test 1: Race And Severity Of Crime

The first t-test was conducted between offenders who were perceived as White versus racialized. To arrive at this dichotomy, all races other than White, including Black, East/Southeast Asian, Indigenous, Latino, Middle Eastern, South Asian, and Unknown or Legacy, were re-coded into one new category, BIPOC (Black, Indigenous, People of Colour). A t-test was then conducted to infer the difference in means in the severity of crime between White offenders and BIPOC offenders. The hypotheses of the test were:

- *Null Hypothesis:* There is no difference in the mean severity of crime between White and BIPOC offenders.
- *Alternative Hypothesis:* There is a difference in the mean severity of crime between White and BIPOC offenders.

A Welch’s t-test was performed. Results of the test are outlined in *Table 2*.

Table 2: Welch’s T-test On The Mean Severity Of Crime Between White And BIPOC Offenders

T-statistic (Welch’s T-test)	p-value	Degrees of Freedom
–9.3528	0.0000	29688.5151

Results show that the difference in means between White offenders (Mean = 11.05; SD = 5.51) and BIPOC offenders (Mean = 11.61; SD = 5.29) is statistically significant, with the p-value (0) being smaller than 0.05 at a 95% confidence interval [-0.68, -0.44]. As such, the null hypothesis can be rejected. This means that there is a difference in the mean severity of crime between White offenders and BIPOC offenders.

T-test 2: Sex And Severity Of Crime

The second t-test was conducted to determine whether the mean severity of crime differed between male and female offenders. Rows where the sex of the offender was unidentified were removed, as they accounted for a very small and negligible proportion of the sample. As a result, the remaining data consisted of only offenders who identified as "male" or "female". The hypotheses of the test were:

- *Null Hypothesis:* There is no difference in the mean severity of crime between male and female offenders.
- *Alternative Hypothesis:* There is a difference in the mean severity of crime between male and female offenders.

A Welch's t-test was performed. Results of the test are outlined in *Table 3*.

Table 3: Welch's T-test On The Mean Severity Of Crime Between Male And Female Offenders

T-statistic (Welch's T-test)	p-value	Degrees of Freedom
-1.3577	0.1746	9985.6940

Results show that the difference in the mean severity of crime between male (Mean = 11.35; SD = 5.41) and female (M = 11.45; SD = 5.31) offenders is not statistically significant, given that the p-value (0.1746) being larger than 0.05 at a 95% confidence interval [-0.24, 0.04]. As such, the null hypothesis cannot be rejected. This means that there is not sufficient evidence that there is a difference in the mean severity of crime between male and female offenders.

T-test 3: Youth vs. Adult And Severity Of Crime

The third t-test was conducted to determine if the mean severity of crime differed between youth (people aged 17 and below) and adult (people aged 18 or above) offenders. In the dataset, the column, "Youth_at_arrest__under_18_years" was a binary variable with inputs of either, “Youth” (aged 17 and below) or “Not a Youth” (aged 18 and above). The hypotheses of the test were:

- *Null Hypothesis:* There is no difference in the mean severity of crime between youth and adult offenders.
- *Alternative Hypothesis:* There is a difference in the mean severity of crime between youth and adult offenders.

A Welch’s t-test was performed. Results of the test are outlined in *Table 4.*

Table 4: Welch’s T-test On The Mean Severity Of Crime Between Youth and Adult Offenders

T-statistic (Welch’s T-test)	p-value	Degrees of Freedom
12.0205	0.0000	1514.2090

Results show that the difference in the mean severity of crime between youth (Mean = 12.89, SD = 4.75) and adult (Mean = 11.30; SD = 5.41) offenders is statistically significant, with the p-value (0) being smaller than 0.05 at a 95% confidence interval [1.33, 1.85]. As such, the null hypothesis can be rejected. This means that there is a difference in the mean severity of crime between youth and adult offenders.

T-test 4: Presence of Strip Search And Severity Of Crime

The final t-test was conducted to determine if the mean severity of crime differed between offenders who did or did not undergo a strip search at the time of arrest. The recoding of this variable from numerical to categorical is further detailed in the methods section of this report. The hypotheses of the test were:

- *Null Hypothesis:* There is no difference in the mean severity of crime between offenders where a strip search was performed or not performed at the time of arrest.
- *Alternative Hypothesis:* There is a difference in the mean severity of crime between offenders where a strip search was performed or not performed at the time of arrest.

A Welch’s t-test was performed. Results of the test are outlined in *Table 5.*

Table 5: Welch’s T-test On The Mean Severity Of Crime Between Offenders Where A Strip Search Was Performed Or Not Performed At The Time of Arrest

T-statistic (Welch’s T-test)	p-value	Degrees of Freedom
4.6357	0.0000	732.1052

Results show that the difference in the mean severity of crime between the group where a strip search was performed at the time of arrest (Mean = 12.12, SD = 4.27) and the group where a strip search was not performed at the time of arrest (Mean = 11.35; SD =5.41) is statistically significant, with the p-value (0) being smaller than 0.05 at a 95% confidence interval [0.44, 1.09]. As such, the null hypothesis can be rejected. This means that there is a difference in the mean severity of crime between offenders who underwent a strip search and those who did not at the time of arrest.

Additionally, a power analysis was conducted prior to carrying out this t-test, as strip search was a new variable of interest. For the results of the power analysis, please refer to *Section 4.1* of this report.

3. METHODS

3.1 Dataset Description

The Arrests and Strip Searches dataset that was analyzed in this study was sourced from The Toronto Police Service's public safety data portal (*Arrests and Strip Searches,* 2022). It contained 65,276 observations across 24 attributes. The dataset included information related to all arrests and strip searches conducted in the City of Toronto and other jurisdictions across a time period of 2020-2021. While the original dataset spanned two years, after the initial exploratory data analysis, the study narrowed its scope to focus on information from only 2021. Therefore, rows that contained 2020 data were removed in the process, resulting in 33,294 observations across 26 attributes remaining in the dataset. This was decided as there were inconsistencies in grouping and naming conventions between the two years of data. 2021 was selected as it contained more data and groupings, and was more recent.

3.2 Exploratory Data Analysis

3.2.1 Preliminary Phase: Selection of Variables and Data Visualization

Exploratory data analysis was undertaken as the first step in the study. The initial phase of the EDA involved streamlining the 24 attributes in the dataset to 9 attributes based on areas of interest. The next phase involved producing bar plots for each of the 9 attributes to uncover high-level trends. If major discrepancies or other interesting trends were noted in the data visualizations, the variables were selected for further exploration. A literature review was also conducted to learn more about the existing research on the relationships amongst the attributes of interest. This further streamlined the dataset to 5 attributes of interest. It was then decided to focus the study on the correlation between demographic factors, including race, sex, and age, strip searches, and the severity of crime. For more details on the process and findings of the EDA, please refer to the exploratory data analysis section at the beginning of this report.

3.2.2 Transformation of Variables

Severity of Crime

The next step in the study involved transforming the outcome variable. As the study looked to investigate factors that impact the severity of crime, the concept of "severity" had to be operationalized. In this analysis, severity of crime was measured by the maximum sentence an individual could receive, should they be convicted with an indictable offence. Indictable offences refer to the more severe types of criminal activities that warrant an immediate arrest. All 21 categories of the offence were ranked by the maximum receivable sentence. An arrest made without a description of the offence was assigned a level of 0. A new column titled, “Severity”, was added to the data frame to reflect these levels. Severity was a continuous variable with values ranging from 0 to 21, where higher ratings were assigned to increasingly severe crimes.

Information on the maximum sentence for each category of offence was extracted from the Canadian and Ontario Criminal Codes, which are the statute legislations that govern criminal offences in Canada and the Province of Ontario (where the City of Toronto is located) (*Consolidated Federal Laws*, 2023; *E-laws*, 2019). In cases where the maximum sentence was the same for two or more categories (e.g., homicide and sex related crimes result in a maximum sentence of life in prison without parole), the category with the higher minimum sentence was noted as more severe. It was decided to apply this operationalized definition to assess the severity of an offence, as maximum sentence was codified and measurable.

There were several considerations with this method. While some offence categories do not vary in type (e.g., there is only one kind of homicide), other categories, such as theft, can be of various types. Theft of items or services that are more valuable are deemed as more serious in the Canadian Criminal Code, hence, the maximum sentence would be longer than that of a theft of a lower value item or service (*Consolidated Federal Laws*, 2023). However, the dataset did not include a detailed description of the offence, and therefore, it was not possible to determine the specific type of offence that an individual committed. As such, the longest sentence that an offender could receive within that broader category of crime was taken to represent the severity of the crime. It is acknowledged that this may have exaggerated the severity of an offence in broader categories and resulted in a wide range of sentence lengths. However, this was mitigated by applying the same method consistently across all categories of crime, aiming to achieve a fairer and more objective measure.

Additionally, some of the criminal categories in the dataset did not include a detailed description of the individual crimes. As an example, the category of, “Police Category - Administrative”, did not correspond to any particular offence in the Canadian or Ontario Criminal Codes. In this case, additional reports from Statistics Canada were referenced to estimate the severity of the offences (*Section 1*, 2015). However, these offences were typically of a lower severity in relation to those that were recognized in the criminal codes.

Strip Search

In the dataset, the column “StripSearch'' was a binary variable with “1” indicating the presence of a strip search at the time of arrest, and “0” indicating the absence of a strip search. This column was re-coded to a categorical variable for the purposes of the t-tests, ANCOVA, and logistic regression. This process also provided a more descriptive label, where rows that took the value of “1” were re-coded as “Yes”, and rows that took the value of “0” were re-coded as “No”.

Demographics: Race and Age

As a result of the t-tests, the statistically significant demographic variables, “Age Group” and “Perceived Race”, were selected to be transformed from categorical to continuous variables in preparation for the ANCOVA analysis. For “Age Group”, the seven groups were assigned a value from 1 to 7, with the youngest age group, “Aged 17 years and under” being 1 and the oldest age group, “Aged 65 and older” being 7. Records without any age group were assigned a value of 0. For “Perceived Race”, as there is no logical order to organize the categories unlike age groups, the 7 perceived races were randomly assigned a value from 1-7, while records with no specification of race or unknown race were assigned a value of 0.

It is recognized that these data transformation strategies could not produce fully continuous variables as there were only seven levels for each of the explanatory variables. As such, it is important to acknowledge that this may limit the results of the analysis from being fully representative of the actual trends in the dataset.

3.2.3 T-tests

As the final step of the EDA, methods of statistical inference were used to compare and infer relationships between perceived race, sex, age, strip search, and the severity of crime. A t-test is designed to test if there is a statistically significant difference between the means of two groups, where the explanatory variable is a nominal, two-level factor while the outcome variable is quantitative. A traditional Student's T-test assumes that the sample is of a normal distribution and has the same variance as the population. Instead of a Student's T-test, this report made use of the Welch's T-test, which assumes that the populations of two groups have a difference in variance.

The dataset largely follows the requirement for the t-test, with a quantitative outcome variable (severity of crime), and categorical predictor variables that have exactly two levels: race (White vs. BIPOC); sex (Male vs. Female); age (Youth vs. Not a youth), and strip search (Yes vs. No). Perceived race, which had more than two levels, was re-coded to "White" and "BIPOC" (Black, Indigenous, People of Colour) to satisfy the requirement for the t-test. Four t-tests were then conducted on each of selected explanatory variables and the outcome variable to determine which of the variables were statistically significant predictors of severity of crime, which would be used to inform further statistical tests.

It is important to acknowledge that while the explanatory variables of race, sex, and age were treated as categorical in the t-tests, they were treated as continuous in ANCOVA. However, the results of the t-tests were still used to inform the inclusion of statistically significant variables in the ANCOVA as it was assumed that the results would stand regardless of the form that the variable took (i.e., categorical or continuous). This can be considered as a limitation in this study.

3.3 Power Analysis

Statistical power refers to the probability of correctly detecting an effect if it is present. It can also be interpreted as the probability of correctly rejecting the null hypothesis if the alternative hypothesis is true. As such, a higher power means that an effect is more likely to be found in an experiment. There are several factors that can impact an experiment's power: (1) sample size; (2) effect size; and (3) confidence interval. The confidence interval is usually set at

95% by convention and the sample size is usually observed or determined by the researchers. The effect size of an experiment refers to the strength of the statistical findings, which are calculated using *Cohen's D* in this report.

Statistical power can be improved by: (1) reducing common variance, (2) increasing sample size; and (3) increasing the mean difference between two groups. In this report, power analysis was performed on the variable, "Strip Search", before carrying out a t-test to assess if the group means for severity of crime differed amongst those who had been strip searched at the time of arrest versus those who had not. The t-test was conducted after the power analysis as strip search was a new variable added to the research and of great importance to help address the research questions. The power analysis would help to inform if the sample size was adequate to attain a high statistical power, given the different effect size intervals. The results of the power analysis are documented in *Section 4.1* of this report.

3.4 ANCOVA

Analysis of Covariance, also known as ANCOVA, is a statistical method that allows researchers to assess group effects of at least one continuous and one categorical explanatory variable on a continuous outcome variable. Usually, the main focus is placed on assessing the effect of the categorical explanatory variable with the continuous explanatory variable utilized as a control or "blocking" variable.

There are two types of ANCOVA: ANCOVA without interaction and ANCOVA with interaction. The former is used when the categorical explanatory variable and the continuous explanatory variable are independent, which enables the continuous explanatory variable to function as a control. In this case, researchers are able to assess the effect of the categorical explanatory variable on the outcome variable, adjusting for the continuous explanatory variable.

In contrast, ANCOVA with interaction refers to a case where the categorical explanatory variable has an effect on the continuous explanatory variable. This results in researchers no longer being able to use the continuous variable as a control to test the sole effect of the explanatory variable on the outcome variable, as there are external factors (i.e., the interaction between two explanatory variables) that influence the observed effect. One way to assess whether the additive model (i.e., no interaction) or the interaction model should be used by looking at the adjusted R^2 value, which provides information on what extent the data can be explained by each model.

Additionally, when conducting an ANCOVA analysis, it is imperative for researchers to consider the following assumptions: (1) normality of errors; (2) linearity; (3) equal variance; and (4) independence of errors. In this case, a quantile-quantile plot (QQ plot) was created prior to conducting the ANCOVA to test for the linearity of the severity of crime at both levels of strip search (present and absent at the time of arrest). In this report, ANCOVA was used to assess the effect of strip search on the severity of crime, after adjusting for the covariates of perceived race and age group.

3.5 Logistic Regression

Logistic regression is a technique that models the relationships between one or more categorical or continuous explanatory variable(s) and a binary, categorical outcome variable. Logistic regression is distinctive from other regression models as it takes on a categorical, binary outcome that is often conventionally labelled as, “success” or “failure”. It is important to note that the interpretation of the outcome variable in a logistic regression is via the lens of the *odds ratio*. As the results of a logistic regression are either “success” or “failure”, the interpretation of the odds ratio should be the ratio of the probability of success to the probability of failure.

This report took a machine learning approach to conducting logistic regression, meaning that the model was developed through an algorithm “learning” the dataset, which improved prediction accuracy after learning additional unseen data. The first step in this process included splitting the entire dataset into a “train” set and a “test” set in an 80-20 ratio. Using Python libraries like *Scikitlearn*, a logistic regression model was generated, as the algorithm learns through the training dataset. Afterwards, the same model was applied to the test dataset to assess how well the model reacted to unseen data. The effectiveness of the model was measured by a confusion matrix, which shows the number of correct and false positive/negative results in a table format. The lower the number of false positives and negatives in the confusion matrix, the fewer errors the model is making, thus, resulting in a more accurate model. In this report, logistic regression was used to assess whether demographic factors like race, sex, and age of an offender could predict the occurrence of a strip search at the time of arrest.

4. RESULTS

4.1 Power Analysis

This section details the power analysis conducted for the fourth t-test looking at the difference in mean severity scores between those who underwent a strip search and those who did not. Further details on this t-test can be found in *Section 2.4* of the report. This power analysis was conducted to help inform whether the sample size included in this dataset was sufficient to generate a high statistical power (i.e., the probability of correctly rejecting the null hypothesis). To do so, the effect size of the experiment was first calculated using the *Cohen's D* formula. The result returned an effect size of 0.142, indicating a small effect size that points towards a small difference in the mean severity score between those who had been strip searched and those who had not.

Afterwards, a power analysis was conducted to test for the statistical power of this t-test, given the previously calculated effect size and the sample size in the dataset, with the confidence interval set at 95%. Results showed that the statistical power for this t-test was 1.0, suggesting a very high probability of correctly rejecting the null hypothesis. However, it is important to note that the number of records where a strip search was not present (N = 32,608) was significantly higher than those where a strip search was present (N = 686). Therefore, a further analysis was conducted to assess if the sample size for each of the levels was substantial enough to establish a statistical power of 0.8. As a result, it was found that the required sample size for the “strip search = yes” group was 395 and for the “strip search = no” group was 18,766. Thus, the sample sizes for both levels met the respective required sample sizes to establish a statistical power of 0.8, given the effect size of 0.142.

To understand why the statistical power turned out to be that substantial, a power curve, as seen in *Figure 12*, was plotted to illustrate the relationship amongst statistical power, effect size, and sample size at a 95% confidence interval. With an effect size of 0.14 (teal line), the required sample size is around 3,000 observations. Since the sample sizes at both levels of the strip search variable met the required sample sizes, it can be concluded that the experiment had enough observations to establish a reliable result.

Figure 12: Power Plot of T-test #4: Severity Score By The Presence And Absence Of Strip Search

4.2 ANCOVA

4.2.1 Overview

This section details the ANCOVA results from testing whether the severity of crime is influenced by the presence or absence of a strip search, given race and age as “blocking” factors. The explanatory variable of sex was excluded from the ANCOVA following the result of a t-test that yielded no statistically significant difference in the mean severity of crime between male and female offenders. The severity of crime in both ANCOVA tests was measured as a continuous outcome variable, and strip search was measured as a categorical explanatory variable. The re-coded race and age group served as continuous outcome variables in their respective tests.

Before conducting the analysis, the linearity of the data was checked to assess if it met the assumptions of ANCOVA. A quantile-quantile plot was used to visualize the data against a perfect normal distribution. As shown in *Figure 13* and *Figure 14*, at both levels of strip search, the severity of crime was not normally distributed. It is recognized that such a phenomenon could be attributed to the fact that the severity of crime was re-coded from categorical data, and as a result, had an upper boundary at a severity score of 21. However, it was decided to proceed with the ANCOVA analysis with this limitation in mind, after considering there to be no suitable alternatives to answer the research questions with the data that was available.

Figure 13: QQ Plot Of Severity Of Crime When A Strip Search Was Present At Arrest

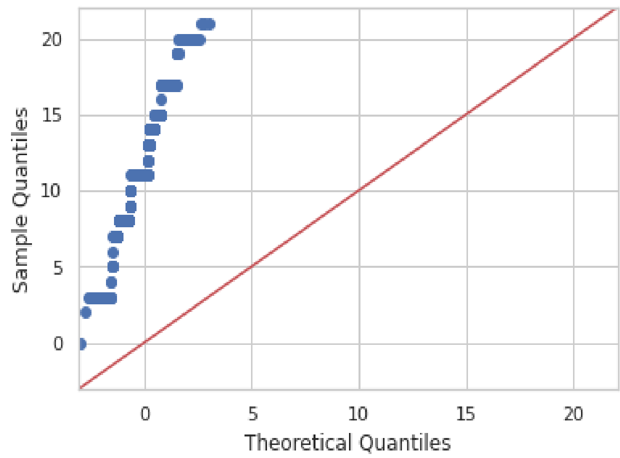
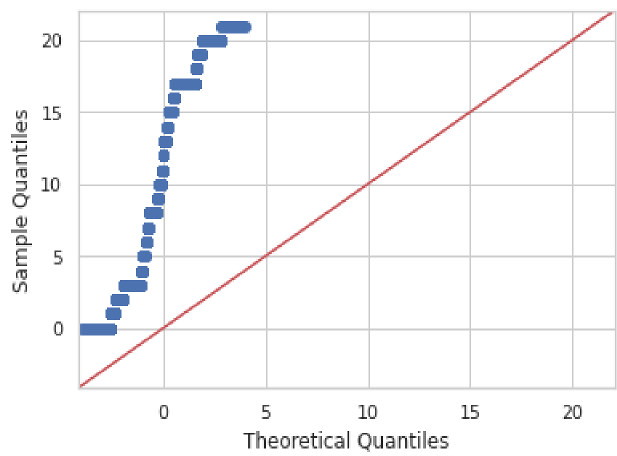


Figure 14: QQ Plot Of Severity Of Crime When A Strip Search Was Absent At Arrest



4.2.3 ANCOVA Test 1: Severity Of Crime And Presence Of Strip Search While Controlling For Race

The first ANCOVA was conducted to assess whether the presence or absence of a strip search had any effect on the severity of crime carried out by an offender, while controlling for the race of an offender. The hypotheses of the test were:

- *Null Hypothesis:* There is no difference in the mean severity of crime between offenders where a strip search was performed or not performed at the time of arrest, while controlling for race.
- *Alternative Hypothesis:* There is a difference in the mean severity of crime between offenders where a strip search was performed or not performed at the time of arrest, while controlling for race.

An ANCOVA was performed. Results of the test are outlined in *Table 6*.

Table 6: Results Of ANCOVA Analysis On Severity Of Crime And Presence Of Strip Search While Controlling For Race

Source	SS	DF	F-Statistic	p-value	np2
Strip Search	395.5983	1	13.62861	0.000223	0.000409
Race	568.0864	1	19.57093	0.00001	0.000588
Residual	966339.5	33391			

Results show that the p-value for strip search is smaller than 0.05. Thus, the null hypothesis can be rejected. This means that there is a statistically significant difference in the mean severity of crime between offenders where a strip search was performed at the time of arrest and offenders where a strip search was not performed, while controlling for race.

4.2.2 ANCOVA Test 2: Severity Of Crime And Presence Of Strip Search While Controlling For Age

The second ANCOVA was conducted to assess whether the presence or absence of a strip search had any effect on the severity of crime carried out by an offender, while controlling for the age of an offender. The hypotheses of the test were:

- *Null Hypothesis:* There is no difference in the mean severity of crime between offenders where a strip search was performed or not performed at the time of arrest, while controlling for age.
- *Alternative Hypothesis:* There is a difference in the mean severity of crime between offenders where a strip search was performed or not performed at the time of arrest, while controlling for age.

An ANCOVA was performed. Results of the test are outlined in *Table 7*.

Table 7: Results Of ANCOVA Analysis On Severity Of Crime And Presence Of Strip Search While Controlling For Age

Source	SS	DF	F-Statistic	p-value	np2
Strip Search	390.010798	1	13.42969	0.000248	0.000403
Age	105.758376	1	3.6417	0.056358	0.000109
Residual	966801.859	33391			

Results show that the p-value for strip search is smaller than 0.05. Thus, the null hypothesis can be rejected. This means that there is a statistically significant difference in the mean severity of crime between offenders where a strip search was performed at the time of arrest and offenders where a strip search was not performed, while controlling for age.

4.3 Logistic Regression

4.3.1 Overview

Following the ANCOVA tests, it was decided to run a logistic regression with all three demographics, race, sex, and age, as categorical explanatory variables, and strip search as a dichotomous, outcome variable. These variables would be used to test whether demographics could predict the occurrence of a strip search. However, as a machine learning approach was used for the logistic regression in this report, and most machine learning algorithms work with numerical values, the explanatory variables needed to be in numerical form. As race and age were already re-coded to numerical categories (outlined in *Section 3.2.2)*, sex was then re-coded to: Male = 0 and Female = 1.

4.3.2 Logistic Regression Results

A single logistic regression was conducted following the model outlined above. The results of the test are outlined in *Table 8*.

Table 8: Results Of Logistic Regression On Strip Search And Demographics (Race, Sex, Age)

Source	Coefficient	SE	Z	p-value	Odds Ratio
Intercept	-3.6040	0.136	-26.456	0.000	0.027214
Age	-0.0725	0.033	-2.174	0.030	0.930041
Race	0.0163	0.021	0.792	0.428	1.016384
Sex	-0.2108	0.015	-1.836	0.066	0.809914

Results show that the p-value for age is less than 0.05. This means that age is the only statistically significant predictor for the occurrence of strip search in the model. In other words, increasing age was associated with a decreased likelihood of an offender being strip searched at the time of arrest. As the odds ratio is approximately 0.93, this means that for each additional age increase for an offender (i.e., each increase in age group), the odds that they are strip searched at the time of arrest decreases by about 0.93 times, holding all other variables constant.

4.3.3 Evaluate Logistic Regression Model Performance

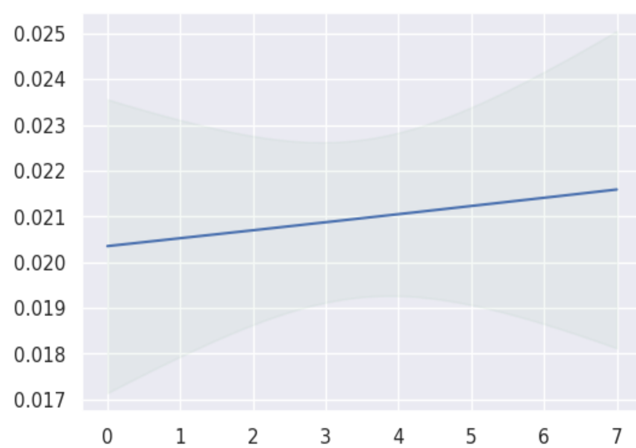
Following the logistic regression, steps were taken to evaluate the model performance. This was accomplished through testing the accuracy of the model, confusion matrices, and prediction intervals. Both a test of accuracy and the confusion matrix in *Figure 15* showed an approximately 98% accuracy. This means that the logistic regression model in this report makes correct predictions 98% of the time.

Figure 15. Confusion Matrix For Logistic Regression Model

```
Confusion Matrix :
 [[6531    0]
 [ 128    0]]
```

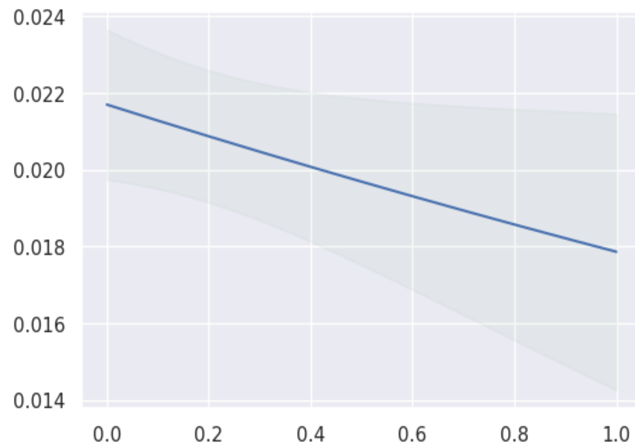
Below are the three prediction intervals that were plotted for the logistic regression model and the variables, race, sex, and age.

Figure 16. 95% Prediction Interval For Logistic Regression Model For Race



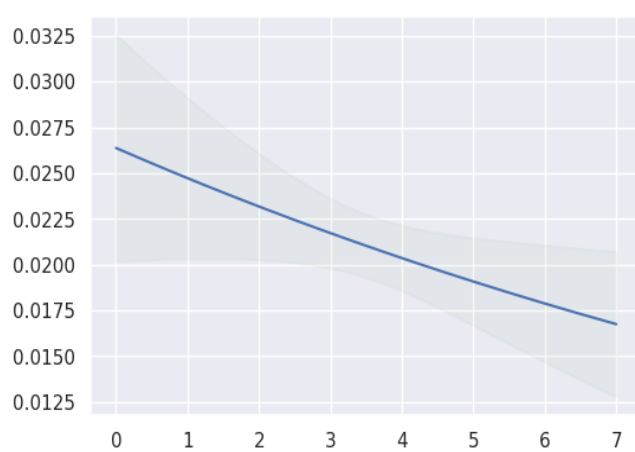
This graph shows the probability that an offender is strip searched based on their race, as predicted by the logistic regression model in the report (blue line). The 95% confidence interval depicted by the shaded green area represents the range of values for new observations. In other words, it shows that 95% of future observations will fall within the shaded range of values. In this case, the graph shows higher probabilities that an offender is strip searched if their race is unknown (value of 0), or if they are Black (value of 1), White (value of 5), East/Southeast Asian (value of 6), or Latino (value of 7). However, the larger shaded range of values for these racial groups also means that the predictions are less reliable than predictions for offenders who are Indigenous (value of 3).

Figure 17. 95% Prediction Interval For Logistic Regression Model For Sex



This graph shows much higher probabilities that an offender is strip searched if they are male rather than female. However, the larger shaded range of values for females (value of 1) also means that the predictions are less reliable than predictions for male offenders (value of 0).

Figure 18. 95% Prediction Interval For Logistic Regression Model For Age



This graph shows the highest probability that an offender is strip searched is if their age is unknown (value of 0) or if they are a youth (17 and under) (value of 1), after which the probability sharply declines, and is lowest for those aged 65 plus (value of 7). However, the larger shaded ranges for those aged 17-24 and 45 plus means that the predictions are less reliable than predictions for offenders who are aged 25-44.

5. DISCUSSION

Key Finding 1: Strip Search And Severity Of Crime Have A Statistically Significant Relationship, While Controlling For Race And Age

The first research question in this study looked to answer whether the severity of crime varies by the occurrence of strip searches, while controlling for the demographics of an offender. The two ANCOVA tests showed a statistically significant relationship between the severity of crime and age and race of the individuals being arrested. This means that the severity of crime does vary by the presence or absence of a strip search at the time of arrest, regardless of an offender's race or age. This result raises interesting insights, as it goes against the hypothesis of the research team based on our own perceived understanding of the variables of interest and the findings from the literature review. The research team hypothesized that the severity of crime would have been related to strip search at the time of arrest, and further impacted by the race or age of an offender. For example, it was discussed that police officers may be more inclined to strip search people of a certain age group or race. However, after isolating the effect of race and age, the presence or absence of a strip search still relates to the severity of crime committed by an offender.

As such, the results of the ANCOVA are contrary to some of the findings in the literature review, where other researchers found that offenders who were Black, youth, or male were more likely to be strip searched at the point of arrest. Nonetheless, similar to what was discussed in the literature review, these findings suggest that the presence of a strip search has a relationship with the severity of crime committed by an offender. Even though this report does not seek to investigate if people who committed a particular type of crime would be more likely to get strip searched, the literature review shows that individuals who committed a crime related to drugs and were in possession of weapons were more likely to be strip searched. Further research in this direction is required to investigate if this dataset follows the same patterns as other research findings in the literature.

Overall, the ANCOVA tests helped to address the research questions which sought to understand the relationship between strip search and severity of crime after controlling for an offender's socio-demographic characteristics, like race and age. The result challenges the research team's conventional understandings of the topic, and sparks further discussions and analysis to unveil the intricacy of the relationship between strip searches and severity of crime in future research.

Key Finding 2: Age Is The Only Statistically Significant Demographic Predictor Of The Occurrence Of Strip Search

The second research question in this study looked to understand whether the demographics of an offender, including race, sex, and age, could predict whether an offender would be strip searched at the time of arrest. The results of the logistic regression that was conducted to answer this question showed that of the aforementioned demographic predictors, age was the only statistically significant predictor of the occurrence of strip search. Furthermore, the odds ratio showed that for each additional age increase for an offender (i.e., each increase in age group), the odds that they are strip searched at the time of arrest decreases by about 0.93 times, holding all other variables constant. Furthermore, the plot for the 95% prediction interval for age showed that the offenders whose ages were unknown, or youth offenders (under 17 years of age) had the highest probability of being strip searched, while those aged 65 plus had the lowest probability. However, it should be considered that larger shaded ranges for those aged 17-24 and 45 plus means that the predictions are less reliable than predictions for offenders who are aged 25-44.

Existing literature, as highlighted in the literature review at the beginning of this report, appears to support these findings. Research coming out of the United Kingdom has shown that the most strip searches occurred amongst the age groups of 21-30, followed by 31-40. In *Figure 18*, it can be seen that while these age groups may have had lower probabilities of being strip searched than youth, the tighter ranges of probabilities (i.e., smaller shaded region) for those aged 25-44 means that the predictions are more reliable than those of other age groups. However, it should also be noted that the highest probability of strip searches occurred amongst groups of offenders with unknown ages which could skew results. Should their ages be known and updated in the dataset and analysis, these trends could shift.

Research from both Toronto and the United Kingdom have also shown that Black offenders are the highest targets for strip searches (when considering their representation amongst the population). This can also be seen in *Figure 16*, even though the regression model did not show race as a statistically significant predictor for strip search. However, just because the model did not note race as significant, does not mean that these trends are not observed in the population, or that the findings of existing literature are insignificant. Rather, it should be considered that similar to age, the high probability of offenders being strip searched with unknown or unrecorded race could skew results and shift trends should their race be known, possibly resulting in the model finding race to be a statistically significant predictor of strip search.

A similar argument can also be made for the sex of an offender. Research has shown that males have a significantly higher chance of being strip searched than females, a trend consistent with that observed in *Figure 17*. However, just because the model in this report did not find sex to be a statistically significant predictor of strip search does not mean that the trend is not observed in the population.

Overall, the takeaway should be that both data collection and analysis methods can impact statistically significant results. Factors like unknown data values, where and when the sample was collected, how the variables were re-coded in this study to be numerical, or how the model was developed should all be considered as potential impacts. As well, just because a variable is not statistically significant in a particular dataset does not mean that the trend cannot be observed in a larger population. It may just come down to a specific trend not being observed in a particular sample. Thus, it is critically important to not take the results of a single study or one model as the truth or generalizable to a larger population. Rather, limitations of the study should be considered, as well as the results of existing research before generalizing results.

6. CONCLUSION

6.1 Overview of Findings

The research questions in this study sought to investigate the relationships amongst demographic factors, strip searches, and severity of crime, as strip searches are one of the more controversial practices in the police force that have sparked discussions within academia and the general public. Specifically, this study looked to understand the relationship between the severity of crime and strip search, after adjusting for the demographic characteristics of an offender. In addition, the study looked to infer if any of the demographic factors of interest could predict if an offender would be strip searched at the time of arrest.

To answer the aforementioned research questions, variables in the dataset were first examined through a series of exploratory data analysis including data visualization and t-tests, to help inform the selection of variables to proceed with further statistical tests. A power analysis was also performed prior to conducting the t-test on strip search, as this was the key variable for examination in the following ANCOVA tests and logistic regression.

The ANCOVA tests have shown that the presence or absence of a strip search is related to severity of crime, after adjusting for an offender's race and age. This was contrary to existing literature which has found that strip searches tend to be performed more often to Black, youth, or male offenders. However, the findings from this report align with existing literature which have found that strip searches are related to the type of crime committed by offenders, as measured by severity level in this study.

As for the second research question, the logistic regression has shown that amongst the three demographic indicators being examined, age is the only statistically significant predictor of strip search at the time of arrest. Other indicators, such as sex and race, did not have a statistically significant relationship with strip search. Specifically, younger individuals were more likely to experience a strip search compared to older individuals. From the results of the logistic regression, it is observed that as age increases, the odds of the individual being strip searched decreases by 0.93 times, holding other factors constant. As well, the prediction intervals varied by the age group in their accuracy, with a wider shading area for the "17 and under" age group and other groups aged 45 and above. This means that the predictions for these age groups would be less reliable.

This finding is similar to what was observed in the literature, which has shown that younger individuals are more likely to experience a strip search by the police. However, the research being used for comparison in this study seems to suggest that other demographic factors, such as race and sex, are also related to the odds of an offender being strip searched. Specifically, the literature has shown that Black, youth, and male offenders are more likely to be strip searched at the time of the arrest, which are supported by the prediction interval plots, but not the results of the logistic regression in this report. Nonetheless, it is important to acknowledge that even though the findings from this report did not suggest a statistically significant relationship between strip search and race or sex, this does not mean that such relationships do not exist at all. Further research is required to examine these relationships in more detail to establish more reliable and valid conclusions.

6.2 Limitations

There were several limitations with this study. As a secondary dataset was used, there were restrictions on the explanatory variables that could be assessed in relation to severity of crime, as all potential variables of interest may not have been included in the data collection process. As well, there were instances where the data lacked detail. As an example, when details of a type of crime were not specified, a subjective opinion had to be used in the decision making process to assign levels of severity to a type of crime.

In addition, multiple data transformations were conducted to re-code categorical variables to continuous variables by assigning a number to each of the categories. Even though this is one of the most common methods of data transformation in statistical research, the re-coded quantitative variables often are limited to very few levels with a small range. For example, the age group and perceived race variables in this dataset only had a range of 1 to 7. This limited range deviates from the natural properties and characteristics of a true continuous variable, which has a wider range and more granular intervals. Therefore, it is recognized that this limitation may have skewed the statistical results of the tests, undermining the internal validity of this research.

Lastly, as the data reflected arrests over a time period of 2020-2021, it should be considered that this was the same time that the COVID pandemic and resulting lockdowns were occurring. Such a major event can result in noticeable shifts in the social and cultural environment, which in turn, can impact crime trends. As such, this may limit the external validity and specifically, the generalizability of the findings of this study to other time periods.

Overall, further research with more detailed metadata included in the dataset to provide context, revised strategies to factor in the limitation in the recoding process of the categorical data, and additional data from other periods would be needed for the future direction of this research.

7. REFERENCES

Allen J. Beck. 2021. Race And Ethnicity Of Violent Crime Offenders And Arrestees. (January 2021). Retrieved February 22, 2023 from https://bjs.ojp.gov/content/pub/pdf/revcoa18.pdf

After Five Years of Increases, Police-reported Crime In Canada Was Down In 2020, But Incidents Of Hate Crime Increased Sharply. (July 2021). Retrieved February 22, 2023 from https://www150.statcan.gc.ca/n1/daily-quotidien/210727/dq210727a-eng.htm

Age, gender and ethnicity of detainees who were strip searched from 2019 to 2021. Metropolitan Police. (n.d.). Retrieved April 14, 2023, from https://www.met.police.uk/foi-ai/metropolitan-police/d/march-2022/stop-and-search-receipt-and-record/

Arrests and Strip Searches (RBDC-arr-TBL-001). (November 2022). Retrieved February 26, 2023 from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Consolidated Federal Laws of Canada, Criminal code. (February 2023). Retrieved February 26, 2023 from https://laws-lois.justice.gc.ca/eng/acts/C-46/

Dodd, V. (2023, March 26). *Black children 11 times more likely to be strip-searched in England and Wales than White Peers*. The Guardian. Retrieved April 14, 2023, from https://www.theguardian.com/uk-news/2023/mar/26/race-disparity-police-strip-searches-of-children-england-and-wales

E-laws. (October 2019). Retrieved February 26, 2023 from https://www.ontario.ca/laws

Incorporating Gender-based Analysis Into Crime Prevention. (August 2022). Retrieved February 22, 2023 from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2020-s002/index-en.aspx

Jeffery T. Ulmer and Darrell Steffensmeier. 2015. The Age And Crime Relationship: Social Variation, Social Explanations. In *The Nurture Versus Biosocial Debate In Criminology: On The Origins Of Criminal Behavior And Criminality*. SAGE Publications, 377–395.

Lemke, M. (2022, July 18). *Policing toronto: Strip searching in a divided city - the bullet*. Socialist Project. Retrieved April 14, 2023, from https://socialistproject.ca/2022/07/policing-toronto-strip-searching-in-a-divided-city/

Section 1: The Crime Severity Index. (November 2015). Retrieved February 26, 2023 from https://www150.statcan.gc.ca/n1/pub/85-004-x/2009001/part-partie1-eng.htm

T.L. Meares. Crime and Gender - UNB. Retrieved February 23, 2023 from https://web.lib.unb.ca/instruction/bcull/CrimeGenderIESBSArticle.pdf