https://colab.research.google.com/drive/1-B7JiwdcJ9vtuzxjVMh71lAV3_o9Uz73#scrollTo=qmqHgOuImbdf

shorturl.at/jmN39

INF2178 Final Project

by the group40 of

Hung Wang
Rohit Pandit

Faculty of Information

April 15, 2023

Table of Contents

Introduction

1. 1 Introduction

In 2019, a Sunday school teacher from Oregon was falsely accused of being a drug smuggler and got strip-searched at Vancouver International Airport. After a body cavity search and an X-ray, no drugs were found, and the woman stated she was traumatized. This controversial incident hit the news headline and caused a heated debate on the legality and legitimacy of strip search by the police on suspects, especially on females. Additionally, one study by Liberty Investigates has revealed that black girls are three times more likely to be subjected to invasive strip searches by the Police than white girls in the UK. The report also discovered that the police use strip searches more frequently on women than they do on men. Consequently, such searches have traumatizing effects on the victims and violate their human rights. This also implies that sex and race could have an impact on the individual likelihood of being strip searched.

Besides the topic of strip search, in Canada, studies have shown that sex can have an impact on an individual's likelihood of being arrested. According to the Government of Canada, one report had shown that females only accounted for 1 in 4 persons accused in police-reported crime incidents in Canada in 2017(Savage, 2019). Chesney-Lind and Pasko's (2013) research has shown that women are typically apprehended for criminal offenses such as theft, drug offenses, fraud, and impaired driving. In cases where women participate in violent behavior, it is usually of a less severe intensity, taking place in private settings and directed towards individuals known to them, as per Schwartz's (2013) findings. Conversely, male aggression tends to result in more harm, happening more frequently in public spaces, and directed towards strangers, as per Schwartz's (2013) observations. As a result, this suggests that the likelihood of being arrested can be influenced by sex.

This study aims to explore how a personal attribute of sex impacts the likelihood of being arrested in Canada. Additionally, the study seeks to examine whether personal attributes including sex and perceived race are associated with individual likelihood of strip search. For this study, we will use a dataset on arrests and strip searches published by Toronto Police

Service. By investigating the influence of personal attributes on arrest rates and strip search rates, the study hopes to highlight the potential for discriminatory practices within law enforcement. The analysis of the dataset and the subsequent evaluation of the evidence aims to provide valuable insights for policymakers, criminal justice professionals, and scholars in the field to better understand and address issues of bias and discrimination within the criminal justice system.

1.2 Literature Review

The study will reference two studies on the criminal justice system and outcomes in the United States for the comparison in the following sections of this paper. The first study *Sex Differences in the Likelihood of Arrest* by Stolzenberg & D'Alessio, examined data from the National Incident-Based Reporting System (NIBRS) from FBI to determine whether sex differences exist in the likelihood of arrest in nineteen states and the District of Columbia during 2000. Seven offenses were analyzed including kidnapping, forcible rape, forcible fondling, robbery, aggravated assault, simple assault, and intimidation. The results found that males are significantly more likely to be arrested for kidnapping, forcible fondling and intimidation than females. Similarly, tying back to the first study, this study also showed that Black females are more likely to be arrested for aggravated and simple assault compared to White females. However, in some cases, the offender's sex on the probability of arrest was more pronounced when an individual was black. By analyzing the data, the analyses suggest that the sex of a criminal offender has an influence on the police decision-making process. Overall, the results of the study indicate that the reason behind the lower arrest rate for females is partially due to the fact that law enforcement officials show more leniency towards women. The authors suggest that these findings have important implications for law enforcement policies and practices, including the need for more gender-specific approaches to policing.

1.3 Research Questions and Objective

The purpose of our research is to investigate how personal attributes such as sex and perceived race interact with the number of arrests and negative behavior during arrest. We will explore two research questions based on our literature review and preliminary analysis of the dataset.

Our first research question seeks to shed light on the potential influence of personal attribute of sex on the number of arrests. In order to achieve this goal, we aim to closely scrutinize whether a discernible relationship exists between sex and the number of arrests. To ensure minimize potential confounds, we will utilize a collective covariate of negative behavior at arrest, which combines concealment of items, combative behavior, violence, spitting or biting, resistance, defensive or escape tendencies, mental instability, suicidal tendencies, and assault of law enforcement personnel. By controlling for these additional variables, we can more accurately isolate the effect of sex on arrest frequency, and arrive at a more nuanced understanding of the underlying factors at play.

Through these research questions, we aim to provide a deeper understanding of the dataset, as we believe that individual factors might influence the number of arrests and negative behavior during arrest. By exploring the interaction of sex and perceived race with these outcomes, we can better identify potential patterns or biases that may exist within the criminal justice system. To guide our investigation, we conducted a literature review and analyzed descriptive statistics ,t-tests, one way ANOVA and two way ANOVA and post-hoc tests on the dataset. Our preliminary analysis has provided us with insight into the nature of the data, which we will use to inform our further exploration.

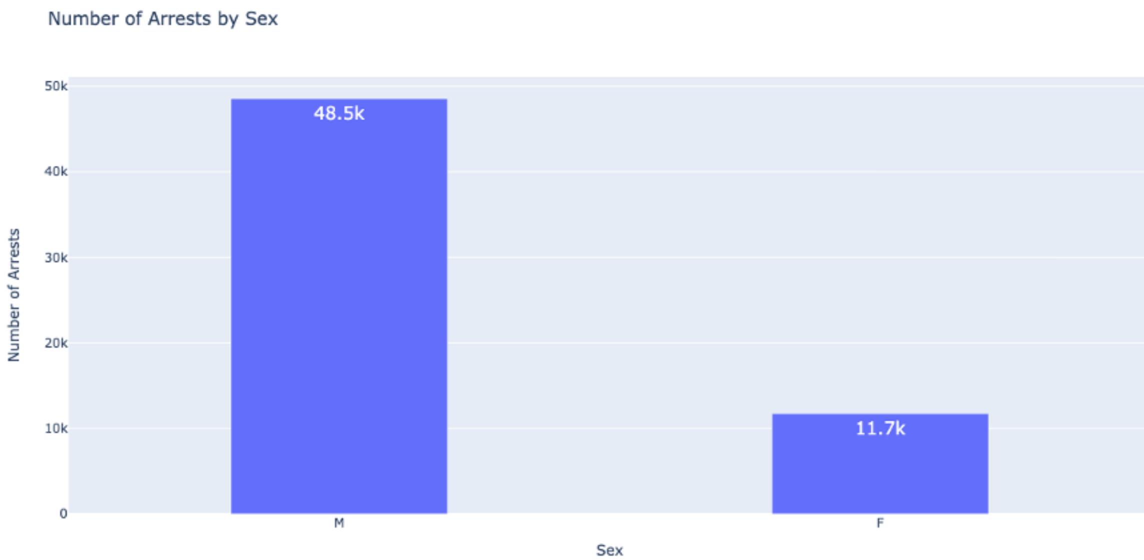
2 Exploratory Data Analysis

2.1 Descriptive Statistics

As a preliminary step towards obtaining a comprehensive understanding of the dataset related to the number of arrests, we generated a count chart that depicts the number of male and female arrestees (see Figure 1). As expected, the result revealed a notable difference in the number of arrests between males and females, with the former having significantly greater proportion of the total arrests recorded. This finding underscores the importance of examining potential sex-based differences in the context of criminal justice system situation.

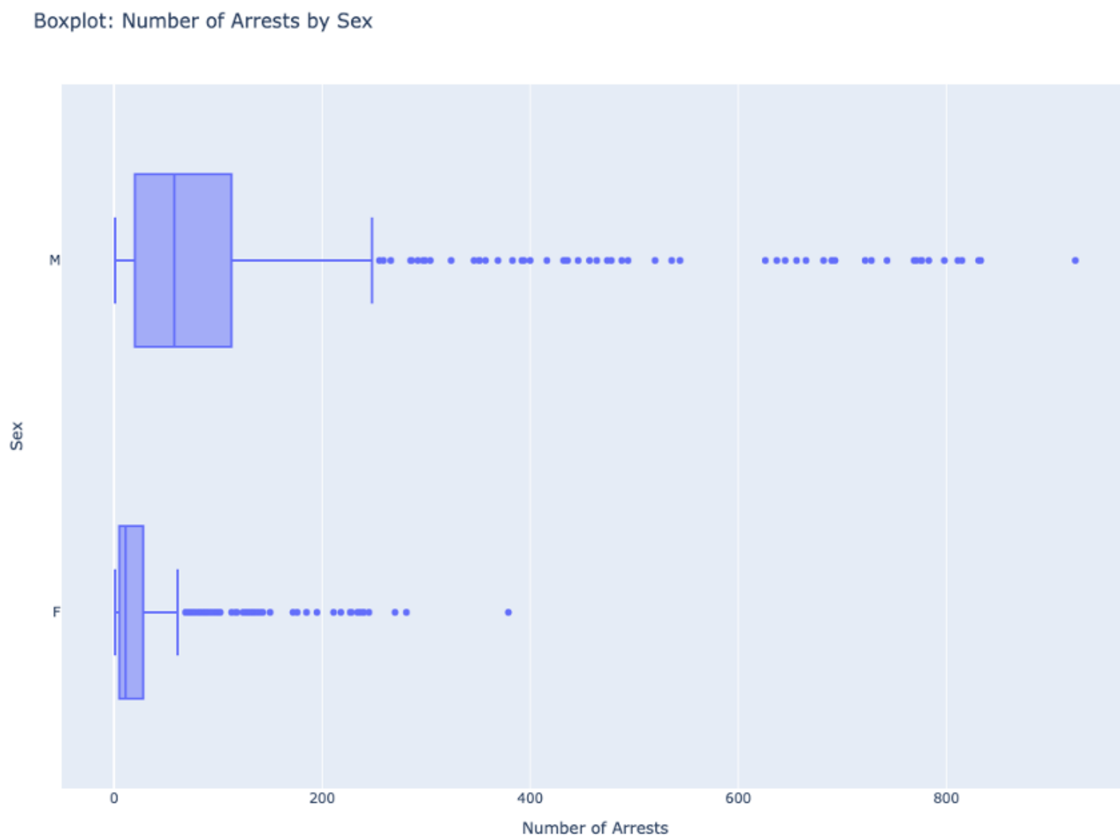
In particular, this sex difference is evident in the White group. However, what surprises us is that White arrestees have the highest numbers while Black arrestees are the second. Other racial groups have a much less number of arrests, with Latinos being the least ones

Figure 1.



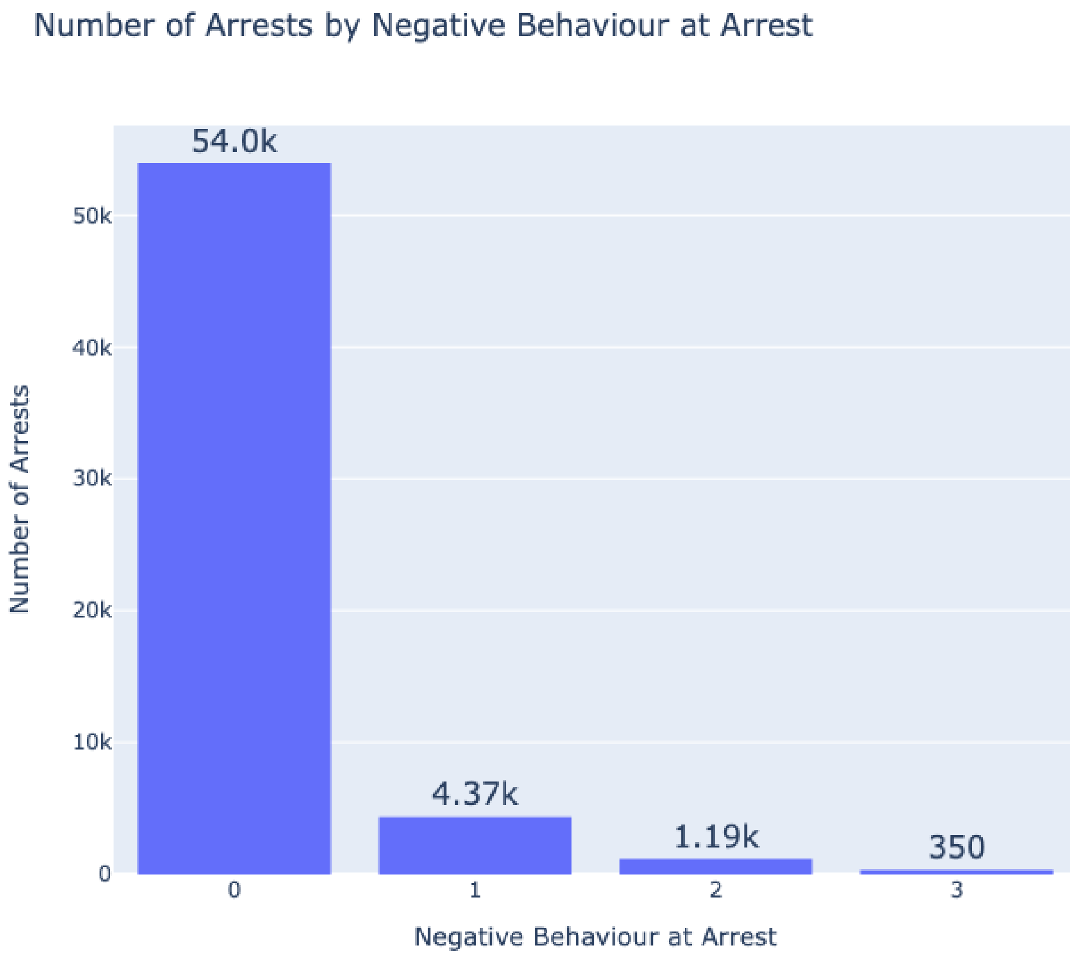
In addition to the count chart analysis, we further explored potential differences in the number of arrests between males and females by generating boxplots (Figure 2). Our examination of the boxplot revealed a higher median value for males, indicating that on average, males have a greater number of arrests than females. Additionally, the range of values for males was found to be wider than that of females, meaning a greater variation in the number of arrests recorded among males. Conversely, the boxplot for females displayed a notably left-skewed distribution, indicating a higher proportion of female arrestees with a lower number of arrests compared to males. Again, these findings provide additional insights into the potential disparities in the number of arrests between males and females, and can serve as a valuable reference for further analysis.

Figure 2



Regarding the covariate of negative behavior at arrest, we produced a count chart to display the frequency of the different frequency of levels of negative behavior observed (Figure 3). The bar chart reveals that the majority of arrestees scored 0 on the negative behavior, indicating that a considerable proportion of arrestees did not exhibit any negative behavior during their arrest. Furthermore, the second most common score was 1, which was less frequent than 0. This suggests that some arrestees displayed minor negative behavior during their arrest. In contrast, scores of 2 and 3 had lower frequencies than 0 and 1, indicating that fewer arrestees exhibited more severe negative behavior. Overall, the distribution of scores for negative behavior suggests that most arrestees did not display significant negative behavior during their arrest. These findings could be valuable for understanding the prevalence and severity of negative behavior during arrests.

Figure 3



Additionally, we produce a stacked bar chart, a bar chart and an interaction plot showing the distribution of negative behaviour at arrest broken down by male and female (Figure 4&5). Based on our analysis of the negative behavior at arrest dataset, it is clear that males exhibit more negative behaviour than females across all scores from 0 to 3. The score of 0 had the highest proportion for both males and females, but the proportion for males was still higher. The data suggests that there is a notable difference in the negative behaviour exhibited by males and females during arrest. However, we are also aware that more male arrestees are recorded in this dataset.

Figure 4

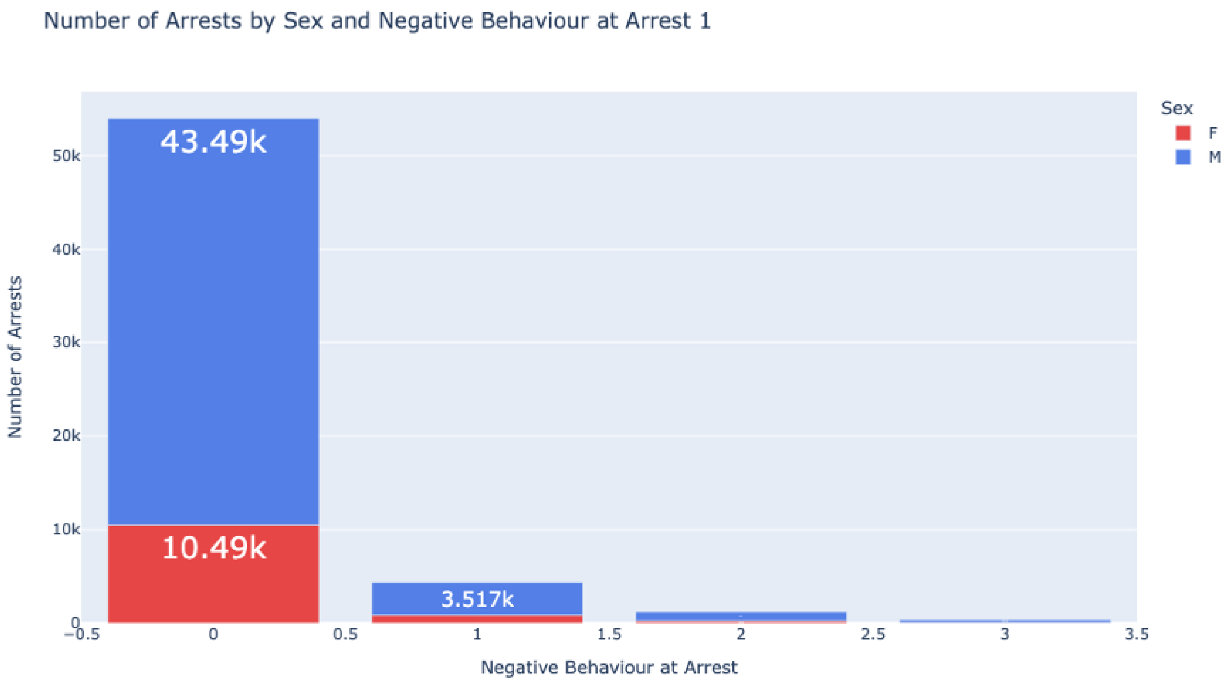


Figure 5

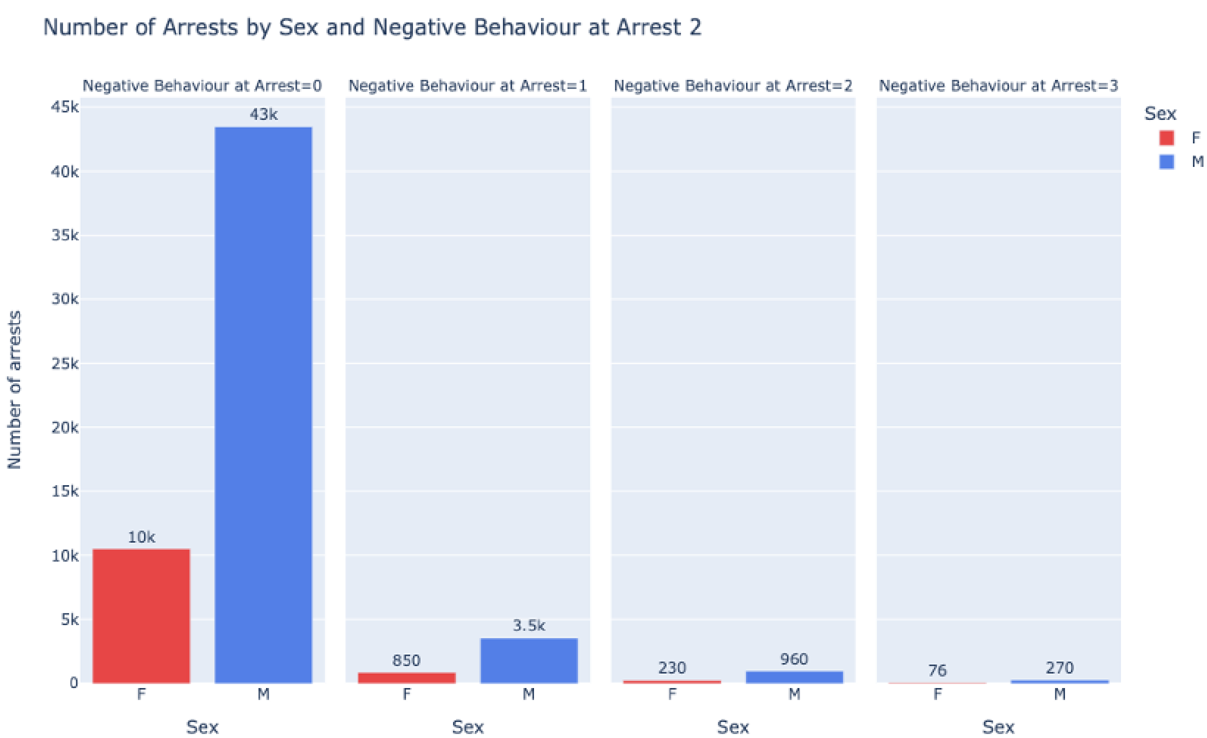
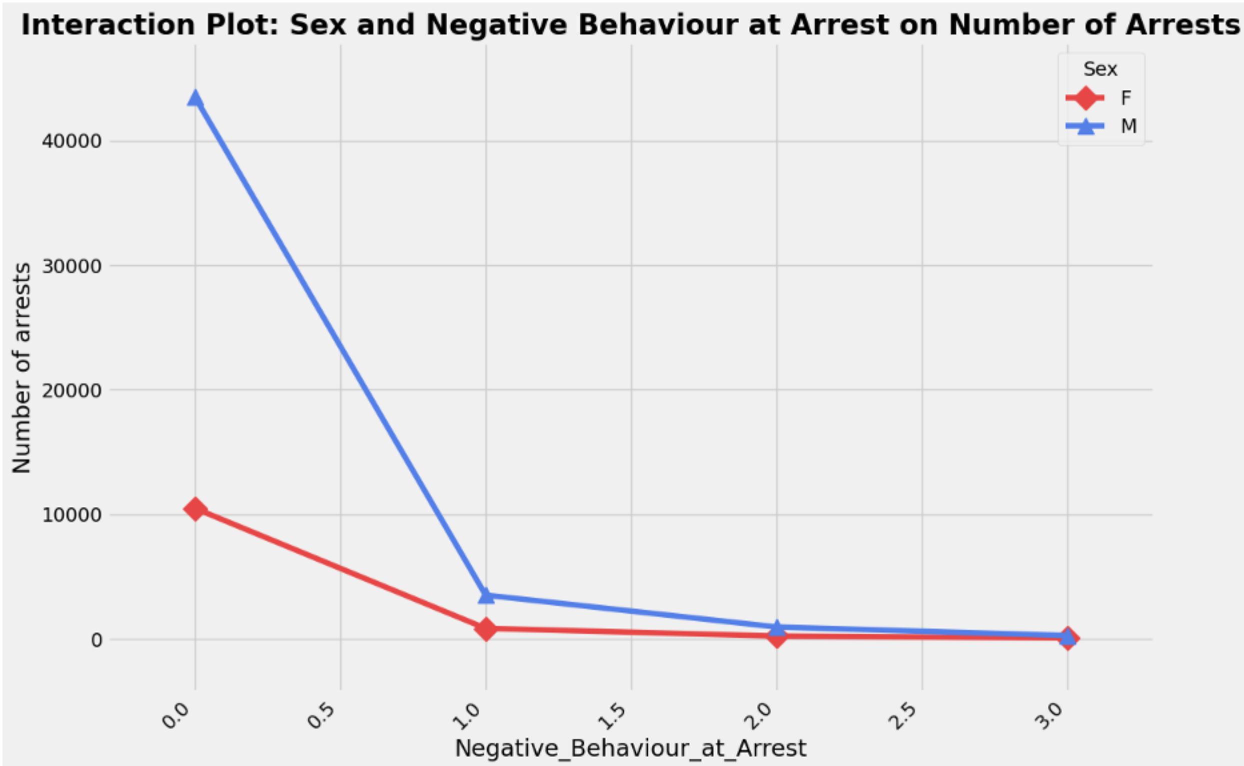


Figure 6



2.2 T-test

In order to further investigate the data, we conducted Welch's t-tests between sex and the number of arrest to determine whether there were significant differences in the means. Welch's t-test was selected over one-sample or two-sample t-tests due to the unequal sample sizes of the two groups since the sample sizes of the two groups are different. In addition, by adjusting for unequal variances, Welch's t-test provides a more precise measure of differences between the two groups. These tests were conducted separately on the 2 different subsets that we had created addressing the two research questions respectively.

Before running the T-test, we ensured the following assumptions were fulfilled.

1. A nominal explanatory independent variable with two levels.
2. A dependent variable is measured on a continuous scale.
3. Normality assumption: The Shapiro-Wilk test shows that the dataset provided is not normally distributed, meaning the normality assumption is violated. However, when the

sample size is larger than 50, the Central Limit Theorem suggests that the sample mean will be normally distributed regardless of the distribution of the population. In this case, we argue that the t-test assumption of normality is met and proceed with the t-test.

4. Independence of observations: The observations used in the following t-tests are independent of each other, meaning that the values in one sample do not affect the values in the other sample.

Sex and Number of Arrests

We calculated the mean number of arrests for males and females respectively. Upon observation, we noticed the male average was greater than the female average. Following that, we conducted a Welch's T-test to explore whether there is a significant difference in the mean number of arrests between the male and female arrestees. The following is our hypothesis:

- Null hypothesis H0: The population mean of the two independent groups, male and female are equal in terms of the number of arrests.
- Alternative hypothesis H1: The population mean of the two independent groups, male and female, are not equal in terms of the number of arrests.

Our statistical analysis, with a significance level (alpha) of 0.05 and a 95% confidence interval, yielded a p-value of 1.56e-09, which is less than the predetermined significance level. This result allows us to reject the null hypothesis and conclude that there is a significant difference in the average number of arrests between males and females in the population. Therefore, if we were to randomly select individuals from the dataset, it is likely that we would observe a difference in the average number of arrests between males and females. As a result, we will include the variable 'Sex' in our analysis ANCOVA to determine the impact of sex on the number of arrests.

3 Methodology

3.1 Data Cleaning

We checked for NULL, NA or NaN values in all the columns of the dataset. Since the number for missing value rows are too many, instead of deleting them, we replaced the missing values in numerical columns including Arrest ID, Age, Occurrence, various Search reasons and Items Found with zeros. All the blank rows of categorical variables were replaced by the modal values of those respective columns. After ensuring that the dataset did not have any missing or blank value, we performed the describe and info function to view the summary statistic and data type of each column.

3.2 Dataset Description

In our project, we used a dataset that gives information about arrests and strip searches in Toronto. The dataset comprises information about 65,275 entries for 37347 unique people including 25 different attributes. Each column attribute name represents a variable and each row is an observation. The cell represents the value.

We created two subsets of the main dataset. The first subset, Number of Arrests, comprised 8 columns: Arrest_Year , Arrest_Month, Perceived_Race, Sex, Youth_at_arrest__under_18_years, Age_group__at_arrest_, Number_arrests, and Negative_Behaviour_at_Arrest. We counted the number of times each of these combinations of variables were booked for any offenses and renamed to Number_arrests. For Negative Behavior at time of Arrests, we summed up the values of all action attributes including Actions at arrest Concealed i, Actions at arrest combative, Actions at arrest Resisted d, Actions at arrest Mental inst and Actions at arrest Assaulted o under one column named Negative Behavior at Arrest. This helped us in analyzing the patterns of the number of arrests and understanding the demographics of the arrested population and control their negative behaviour at arrest for our first research question.

For the second subset, Negative Behavior at time of Arrests, we summed up the values of all action attributes including Actions at arrest Concealed i, Actions at arrest combative, Actions at arrest Resisted d, Actions at arrest Mental inst and Actions at arrest Assaulted o under one column named Negative Behavior at Arrest. We then grouped by the attributes Arrest Year, Perceived Race, Sex and Youth at arrest under 17 years and measured the count of negative behavior of each combination at the time of arrests. This helped us in analyzing the pattern of negative behavior of each demographic group during the time of arrest based on our second research question.

3.3 ANCOVA

For ANCOVA analysis, we checked the following 5 test requirements:

1. Data is independently and randomly sampled.
2. The level of measurement is interval/ratio.
3. Normality assumption: Populations are normally distributed. However, this is not very strict, especially since our groups have a large sample size.
4. Independence of observations: The observations in each group should be independent of one another. This means that the value of one observation should not affect the value of another
5. Homogeneity of regression slopes: The relationship between the covariate and the dependent variable should be linear, and the slopes of the regression lines should be equal across groups of the independent variable.

3.3.1 Power analysis

Power is a metric that measures the probability of correctly identifying a positive result. It is calculated as the complement of the probability of failing to identify a true effect, also known as the Type 2 error rate. In practice, a power value of 0.8 is commonly employed. The potential statistical power of an experiment can be determined by taking into account the significance level, sample size and estimated effect size. Therefore, one can also calculate the desired sample size required to achieve a desired statical power, for example, 0.8, for their statistical experiment, whi ch is the objective of this section.

To understand what ideal sample size for our variable Sex in terms of number of arrests to achieve desired statistical power, we first define a function called pooled_standard_deviation to calculate the pooled standard deviation of two samples, in this case, males and females. The function will calculate the sample sizes of the two samples, and then calculate the sample variances using np.var function from the NumPy library with the degree of freedom equal to 1.

Next, the function calculate the pooled standard deviation. Additionally, a second function called Cohens_d is created to calculates Cohen's d, which is a measure of effect size for the difference between two means, in our case, male and female. The function calculates the mean values of the two samples using np.mean, and then calls the pooled_standard_deviation function to calculate the pooled standard deviation.

Next, we want to calculate the effect size for the number of arrests. First, we import the TTestIndPower class from the statsmodels.stats.power module, which is to calculate the statistical power of a two-sample t-test. The previously-defined Cohens_d function is used here to calculates the effect size between males and females. The following lines set the significance level, which is the standard of 0.05, and the desired statistical power, which is 0.8, and calculate the ratio of the sample sizes of males to females. The result will give us the effect size for number of arrests.

In order to calculate the ideal sample size to achieve a desired statistical power for males and females respectively, we import the the TTestIndPower class from the statsmodels library. We use the solve_power method of the analysis object to calculate the sample size required to achieve the desired statistical power for a given effect size, significance level, and ratio of sample sizes. The result gives us the required sample size for both males and females to achieve statistical power. We also create a power curve graph for two samples to illustrate the relationship between sample size and effect size, holding power at 0.8 and alpha at 0.05.

3.3.2 ANCOVA

In Our ANCOVA analysis, the dependent variable (Number of Arrest) was measured on a Continuous scale, and the independent variable (Sex) and a covariate (Negative_Behaviour_at_Arrest) were considered. To perform ANCOVA analysis, we first created a new dataframe called “number_arrests__sex_1way” that includes only the variables “Sex”, “Negative Behaviour at Arrest’, and “Number of Arrests” from the original dataframe. It is also important to ensure that the data is correctly stored in the new dataframe. Next, we use the ANCOVA function from the pingouin library on the Number_arrests variable, with

Negative_Behaviour_at_Arrest as a covariate and Sex as the independent variable. The output of this function will provide conditional probability of predicting Number_arrests in terms of the sex of the individual, holding Negative_Behaviour_at_Arrest constant.

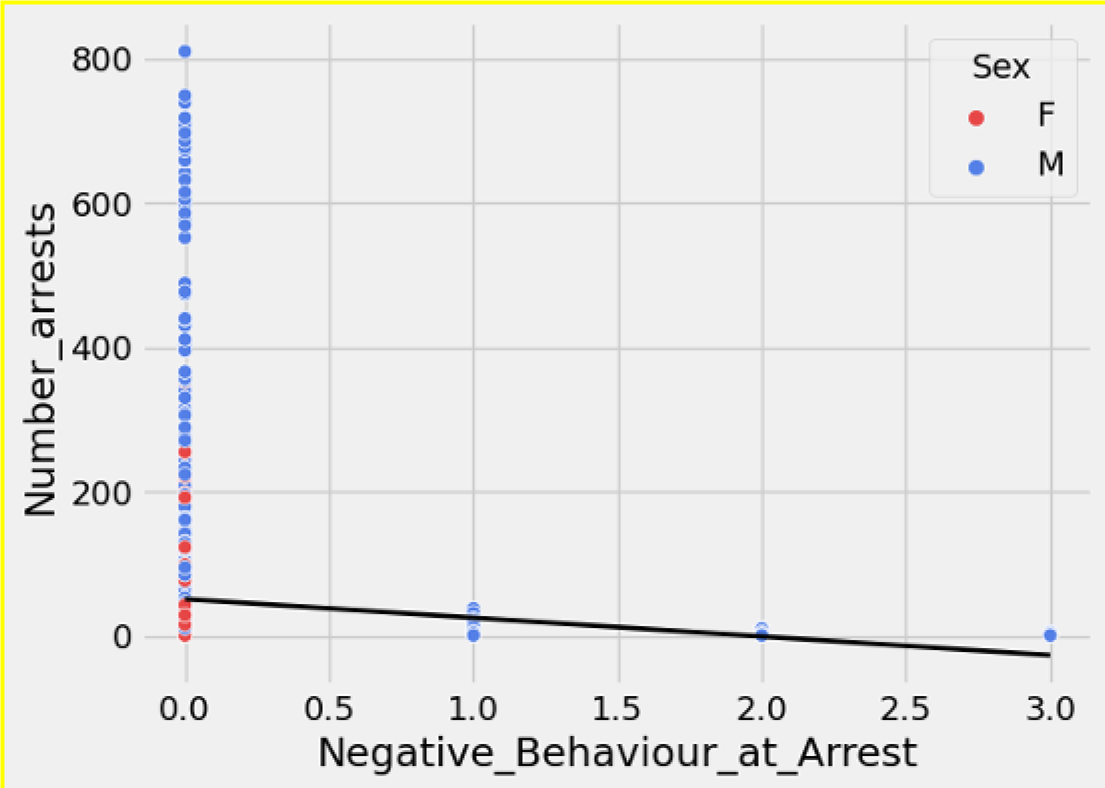
4 Results and Findings

We derived some interesting insights about the underlying data considering several aspects and factors of the data. The first research question was aimed to analyze how the number of arrests and differed on the basis of person's sex after controlling their negative behaviour at arrest. [RQ2]. Overall, the results drew light on some interesting insights about the factors that affected the number of arrests and strip search.

4.1 Research Question 1 with ANCOVA

For ANOCA assumption checks, we found that the data is independently and randomly sampled. The level of measurement is also interval/ratio. The Shapiro-Wilk test shows that the dataset provided is not normally distributed, meaning the normality assumption is violated. However, when the sample size is larger than 50, the Central Limit Theorem suggests that the sample mean will be normally distributed regardless of the distribution of the population. In this case, we argue that the t-test assumption of normality is met and proceed with the t-test. Furthermore, using the Levene's test, we found that the population variances are not equal for the number of arrests. As demonstrated by the scatterplot (Figure7), the relationship between the covariate and the dependent variable is non-linear, and that the regression line slopes differ across the independent variable groups. For the sake of this project, we will continue our analysis but we will also address these limitations in the discussion section below.

Figure 7



4.1.1 Power Analysis

The power curve graph (Figure) illustrates how statistical power varies across different sample sizes for each effect size and each sex level. For instrance, at a small effect size of 0.20, the power curve graph shows that a sample size of at least 250 participants is needed to achieve a statistical power of 0.80 or higher. At a medium effect size of 0.5, a sample size of 80 observations is needed to achieve a statistical power of 0.80 or higher. At a large effect size of 1.3, a sample size of 20 participants is needed to achieve a statistical power of 0.80 or higher. In other words, as the effect size increase, the number of samples needed to find significance reduces.

Figure 7

In our experiment, we calculate that our effect size for the difference in the mean number of arrests between males and females is 0.18. According to Cohen's guidelines, an effect size of 0.2 is small, 0.5 is medium, and 0.8 is large. As a result, the effect size of 0.18 in our case indicates that there is a small-to-medium difference in the number of arrests between males and females. With the effect size of 0.18 at alpha of 0.05 and power of 0.8, we derive that the required sample size for males and females are 335.11 and 643.63 while the actual sample size we have for males are females are 1840 and 958. This indicate that to conduct the ANCOVA statistical experiment to test our hypothesis using our current dataset can yield a strong statical power of more than 0.8.

Table

Sex	Required Sample Size	Actual Sample Size
Male	335.11	1840
Female	643.63	958

4.1.2 ANCOVA

According to the ANCOVA results of our analysis, sex (Independent Variable) of a person did have an significant effect on our dependent variable number of arrests after controlling the individual's negative behaviour at arrest as a covariate. In more details, "uncorrected p-value" for Sex is less than 0.05. Therefore, we can reject the null hypothesis that each of the sex, including males and female respectively, results in the same number of arrests, even after controlling for their negative behaviour at arrests. In terms of practical interpretation, we hypothesized that individual sex would be able to predict the one's likelihood of being arrest. Drawing the insight from our analysis, we see that there is statistically significant relationship between individual sex and the likelihood of being arrest when controlling for their negative behaviour at arrest.

Table

	Source	p-unc
0	Sex	5.357803e-15
1	Negative_Behaviour_at_Arrest	2.306641e-74
2	Residual	NaN

4.2 Research Question 2

We explored the interaction effect of sex and race of the arrested person with the person's negative behavior at the time of his arrest. The results of our analysis are not clear cut as the combined interaction effect of the independent variables sex and race on the dependent variable negative behavior at the time of arrest was not of statistical significance. However, we observed that the independent variables had individual effects on the dependent variable which were of statistical significance. Further analysis needs to be performed to draw insights from it.

5 Discussion:

5 Conclusion and Limitations:

With respect to the recent surge in violence, it is essential to understand the patterns and trends of the arrested populations. We attempted to study the impact of different demographics of the arrested person with the number of arrests and behavioral patterns of the arrested person. We tried to study the two research questions resorting to the use of t-tests, 1-way ANOVA , 2-way ANOVA followed by post-hoc tests respectively.

We faced several limitations while performing the analysis. We performed the Shapiro-Wilk test to verify the normality distribution assumption. The results of the Shapiro-Wilk test indicated that our normality assumptions about the subsets of the dataset that we had created were not satisfied. Also, we conducted the Levene’s test to check the assumption of the homogeneity of the variances. The results suggested that these assumptions were not satisfied by the independent variables (Sex and Race) for the number of arrests for the Research Question 1. We also observed a few outliers in the box plot distributions in some groups. Future research should take into consideration of the aforementioned limitations before conducting a follow-up research.

6 References

Department of Justice Government of Canada. 2022. Overrepresentation of black people in the Canadian Criminal Justice System. (December 2022). Retrieved February 27, 2023 from https://www.justice.gc.ca/eng/rp-pr/jr/obpccjs-spnsjpc/index.html

Jessica G. Finkeldey and Stephen Demuth. 2019. Race/ethnicity, perceived skin color, and the likelihood of adult arrest. Race and Justice 11, 4 (2019), 567–591. DOI:http://dx.doi.org/10.1177/2153368719826269

Laura Savage. 2019. Female offenders in Canada, 2017. (January 2019). Retrieved February 27, 2023 from https://www150.statcan.gc.ca/n1/pub/85-002-x/2019001/article/00001-eng.htm

Lisa Stolzenberg and Stewart J. D'Alessio. 2004. Sex differences in the likelihood of arrest. Journal of Criminal Justice 32, 5 (2004), 443–454. DOI:http://dx.doi.org/10.1016/j.jcrimjus.2004.06.006

US Customs and Border Protection. 2022. What is racial profiling? (May 2022). Retrieved February 27, 2023 from https://www.cbp.gov/faqs/what-racial-profiling

Scot Wortley and Maria Jung. 2020. Racial Disparity in Arrests and Charges An analysis of arrest and charge data from the Toronto Police Service. (2020).

Chesney-Lind, M., & Pasko, L. (2013). *The female offender: Girls, women, and crime (*3rd ed.)*.* Thousand Oaks, CA: SAGE Publications, Inc.

Schwartz, J. (2013). A "new" female offender or increasing social control of women's behaviour? Cross-national evidence. *Feminist Studies, 39*(3), 790-821.