UNIVERSITY OF TORONTO
FACULTY OF INFORMATION

INF2178 Midterm Project

by the group55 of

Hung Wang
Rohit Pandit

Faculty of Information

February 26, 202

Table of Contents

Introduction 2
1. 1 Introduction 2
1.2 Literature Review 3
1.3 Research Questions and Objective 4
2 Exploratory Data Analysis 4
2.1 Descriptive Statistics 4
2.2 T-test 12
3 Methodology 16
3.1 Data Cleaning 16
3.3 ANOVA Tests 17
3.3.1 One-Way ANOVA with Sex (First IV of RQ1): 17
3.3.2 One-Way ANOVA with Race (Second IV of RQ1): 18
3.3.3 Two-Way ANOVA with Sex and Race 19
3.3.4 One-Way ANOVA with Sex (First IV of RQ2): 21
3.3.5 One-Way ANOVA with Perceived Race (Second IV of RQ2): 21
3.3.6 Two-Way ANOVA with Sex and Race 22
4 Results and Findings 23
4.1 Research Question 1 24
4.2 Research Question 2 24
5 Conclusion and Limitations: 24
6 References 25

Introduction

1. 1 Introduction

In 2013, a neighborhood watch volunteer who fatally shot an unarmed Black teenager named Trayvon Martin in Sanford, Florida, United States. However, following national protests, also known as the Black Lives Matter movement, Zimmerman was acquitted of all charges by a Florida jury. The case resulted in widespread debate and controversy around issues of racial profiling, gun usage and black over-presentation in the criminal justice system in the US. However, the controversy of racial profiling and over-representation of certain demographic groups in the criminal justice system is not unique to the United States.

In Canada, studies have shown that race and sex can have an impact on an individual's likelihood of being arrested. According to the Government of Canada's report, Black people are more likely to experience discrimination when dealing with the police (Government of Canada, 2022). Another report from the Government of Canada had shown that females accounted for 1 in 4 persons accused in police-reported crime incidents in Canada in 2017(Savage, 2019). Furthermore, the Ontario Human Rights Commission found that although Black males constitute only 4% of Toronto's population, they make up 26.7% of arrests that lead to street releases (Wortley & Jung, 2020). This means that Black males are 6.7 times more likely to be arrested and released on the street than the others, indicating a sign of racial profiling in Canada. This also suggests that the likelihood of being arrested can be influenced by the intersection of race and sex. As a result, although Canada is often seen as more culturally tolerant and diverse and progressive, its criminal justice system seems to not be immune to systematic biases and disparities.

This study aims to explore how personal attributes including sex and perceived race impact the likelihood of being arrested in Canada. Additionally, the study seeks to examine whether these personal attributes are associated with negative behavior during the time of arrest. 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, 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 is *Race/Ethnicity, Perceived Skin Color, and the Likelihood of Adult Arrest* by Finkeldey & Demuth, which examines the relationship between race/ethnicity, perceived skin color, and the likelihood of adult arrest in the United States. The study finds that individuals who are perceived to have darker skin color are more likely to be arrested than those with lighter skin color, regardless of their race or ethnicity. For example, self-identifying as Latinos or Native Americans and having darker skin color are more likely to be arrested than those who have lighter skin colour. The authors suggest that colorism does exist in the current arrest practices, resulting in implicit bias and stereotypes that associate darker skin color with criminality. Additionally, the study finds that African Americans and Hispanics are more likely to be arrested than Whites, even after controlling for perceived skin color. The authors suggest that this is due to systemic racism and discrimination in the criminal justice system.

The second 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 (RQ1) is centered on the impact of personal attributes on the number of arrests. Specifically, we will examine whether there is a relationship between sex, perceived race, and the number of arrests. We will also explore whether the number of arrests differs across different combinations of these attributes.

Our second research question (RQ2) focuses on the impact of personal attributes on negative behavior during arrest. We will investigate whether sex and perceived race have a relationship with negative behavior during arrest, and whether negative behavior differs across different combinations of these attributes.

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

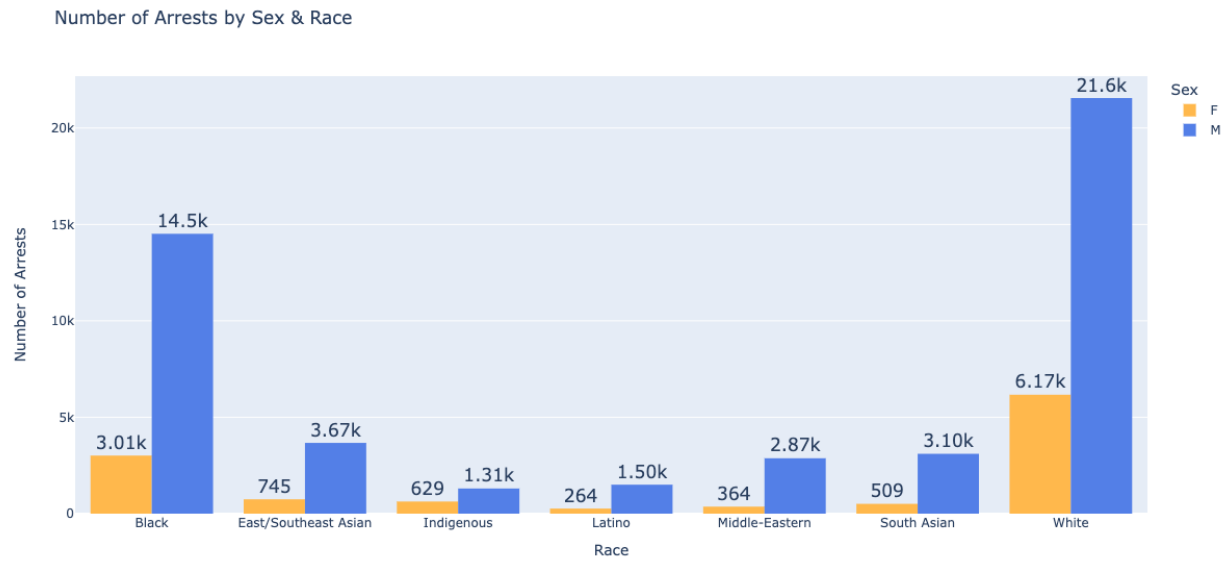
To first understand the dataset related to the number of arrests, we produced the following count chart showing the number of male and female arrestees in terms of their racial identities (Figure 1). Unsurprisingly, there were more male than female arrestees in the number of arrests across all

Table 1: Number of Arrests

Mean	Standard Deviation
80.83	148.92

the racial groups. 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.



We also constructed boxplots for both sex and racial comparison (Figure 2&3). In terms of age, we can see that males have a higher median than females, indicating that males have a higher number of arrests. Concerning race, we can see that the White group has the highest median, followed by the Black group, which suggests that the White group has the highest number of arrests overall, followed by the Black. The range for the White group is wider than the rest, suggesting that the White group may have more variation in the number of arrests than other racial groups. The boxplots for racial groups, except for the White and Black groups, are very left-skewed, meaning there may be a lower proportion of fewer arrests in these groups.

Figure 2

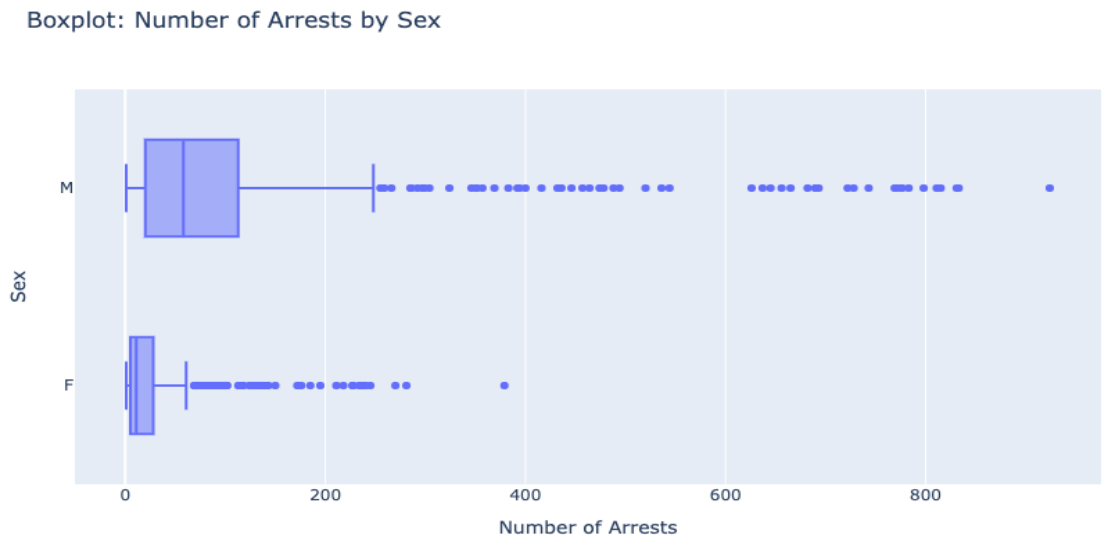
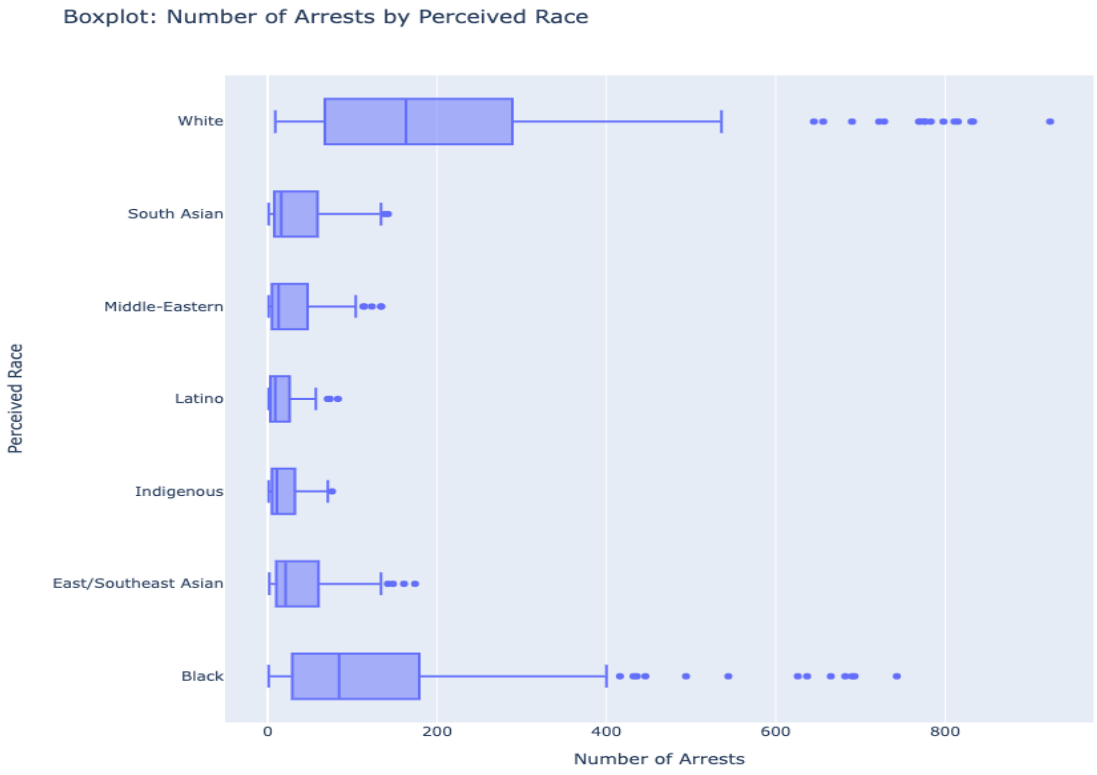


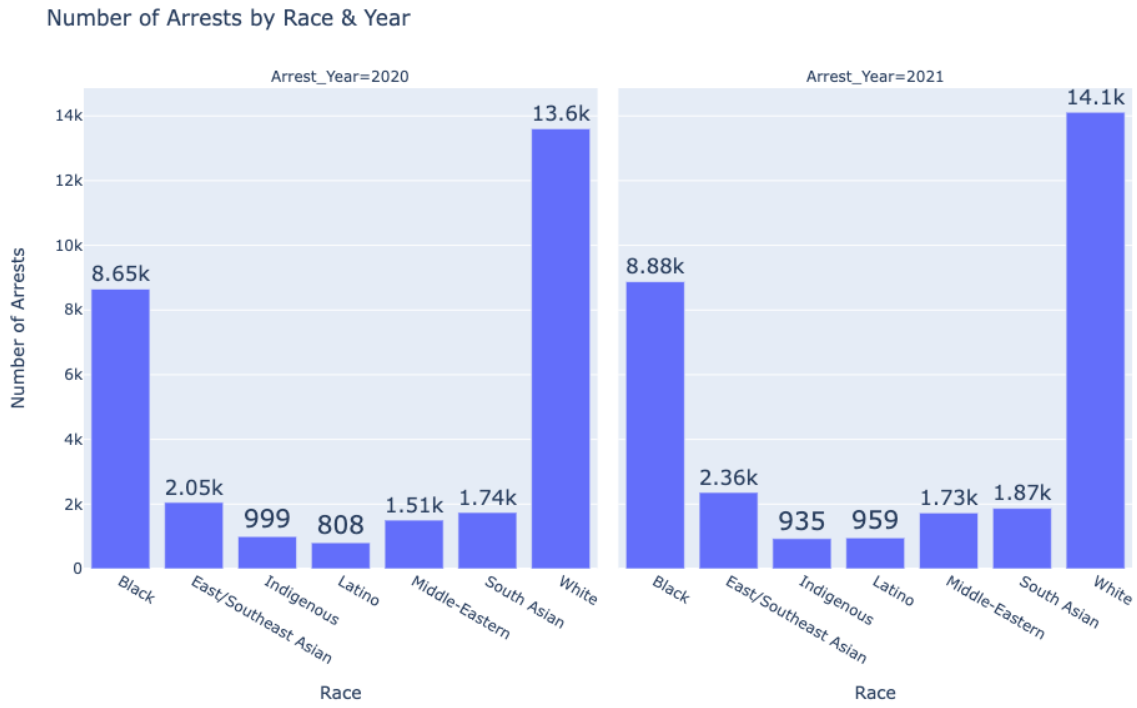
Figure 3



Given that 2020 was the year that the pandemic happened, resulting in an economic recession, we want to explore whether this economic downturn had an impact on the increase in criminal

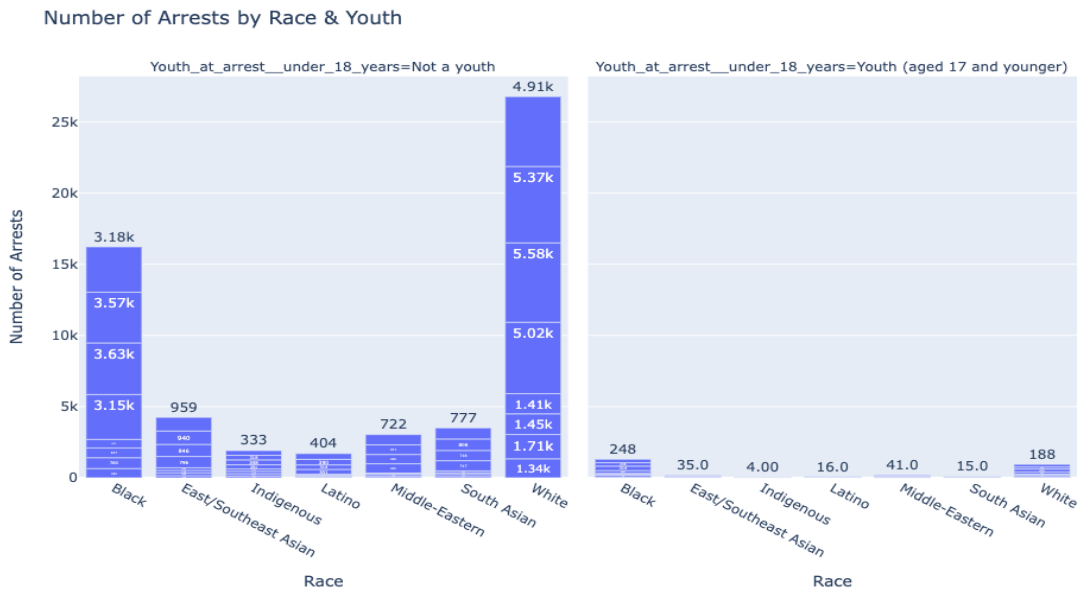
activities, which reflects on the number of arrests between 2020 and 2021. Figure 4 shows the difference between 2020 and 2021 with regard to race. However, we do not see any notable difference in the number of arrests between these two years.

Figure 4



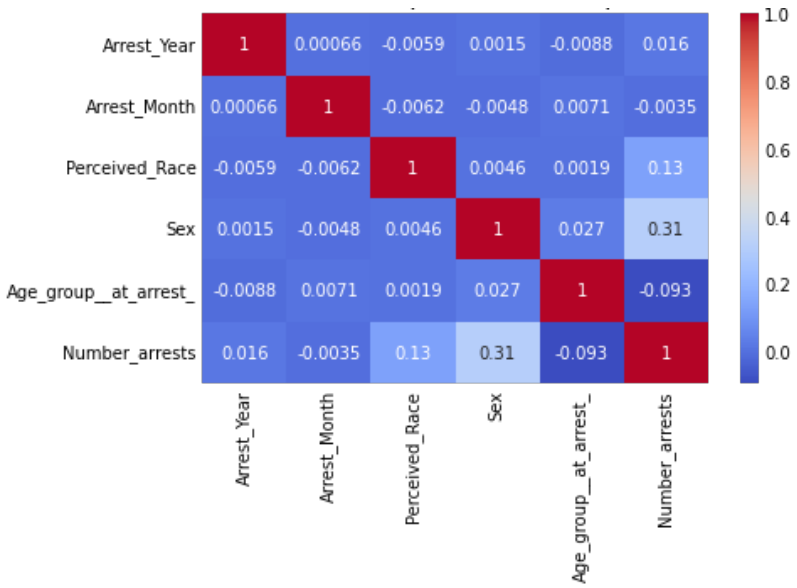
We also looked into whether being a youth or not can affect the likelihood of arrest (Figure 5). We see that the number of adult arrestees is dramatically higher than youth under 18. It is interesting to see that although youth has a much less number of arrests, Whites and Blacks are still the groups with the highest number of arrests. In this case, Black youths are more likely to be arrested than their white counterparts.

Figure 5



The correlation matrix informed the relationship between different variables with the imputed values (Figure 6).The correlation between Number of Arrest and Sex (0.31) and Number of Arrest and Race was (0.13) were the highest as compared to other variables. Whereas the correlation between Race and Sex was (0.0046) which would be of our interest in the Two-way ANOVA test.

Figure 6 Correlation Matrix of Independent and Dependent Variable



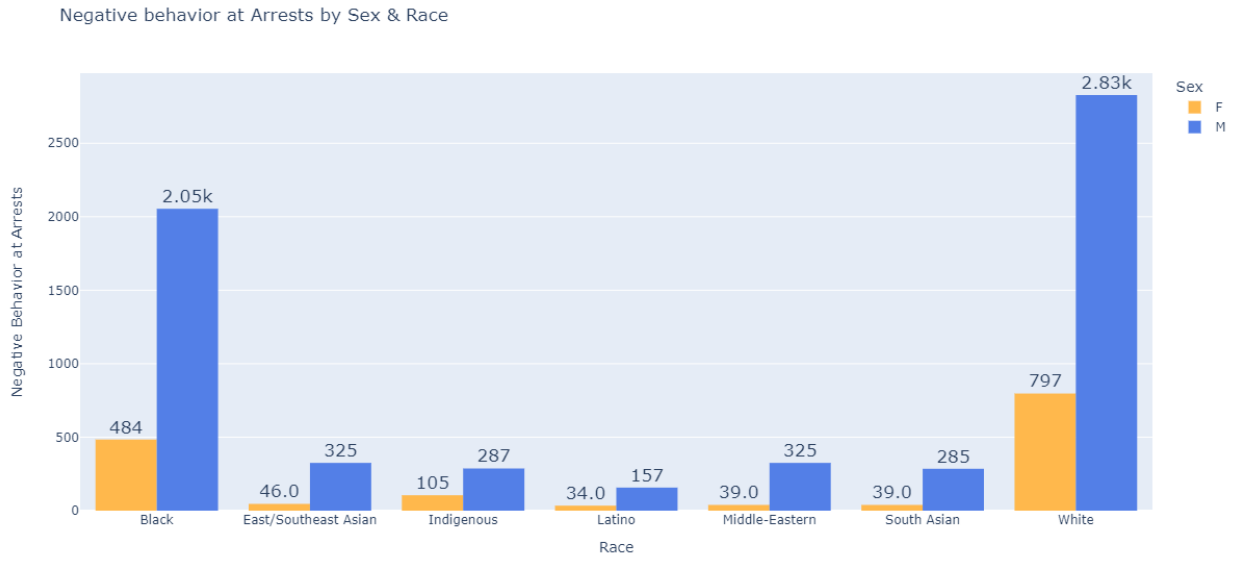
Regarding our second research question, here is some basic descriptive statistics for our second dependent variable Negative Behavior at arrest

Table 2: Negative Behavior at Arrest

Mean	Standard Deviation
139.38	312.53

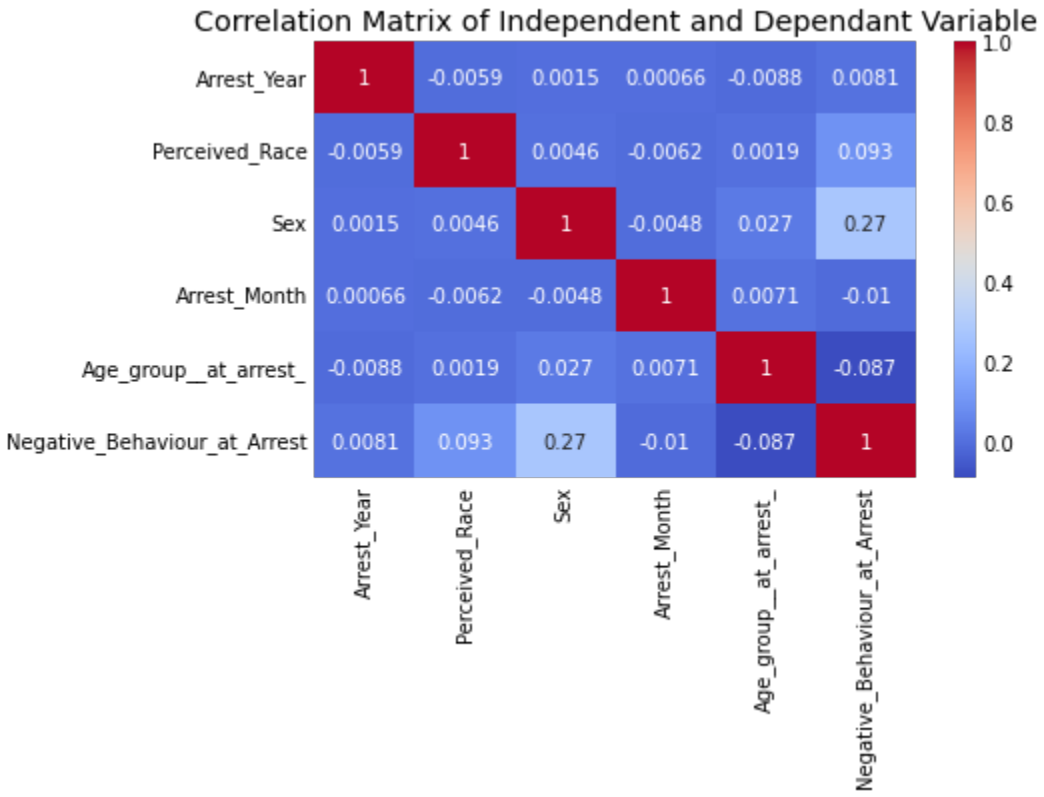
We plotted a bar chart to visualize the negative behavior of each gender of different perceived races at the times of arrests (Figure 7). It was observed that males in general had behaved negatively at the time of arrest as compared to females with White males having the highest number.

Figure 7



The Negative Behavior at Arrest variable had the highest correlation with Sex (0.27) and Race (0.093). Moreover, the correlation between Race and Sex was 0.0046. These factors would be of our interests while conducting the Two-way ANOVA test.

Figure 8. Correlation Matrix of Independent and Dependent Variable



To further show the relationship between sex, race and the number of arrests, we created an interaction plot (Figure 9). The minimal interception between almost all types of races indicate some sort of interaction between the number of arrests and sex. However, the parallel lines between the Black and White races indicate no interaction. We also created an interaction to explore the effect of sex and race on the negative behaviour at arrest (Figure 10). Most of the lines coincide and are parallel indicating absence of clear interaction, however it is difficult to get a clear interpretation. Hence, to further explore the relationship, we will perform various tests in the following sections.

Figure 9

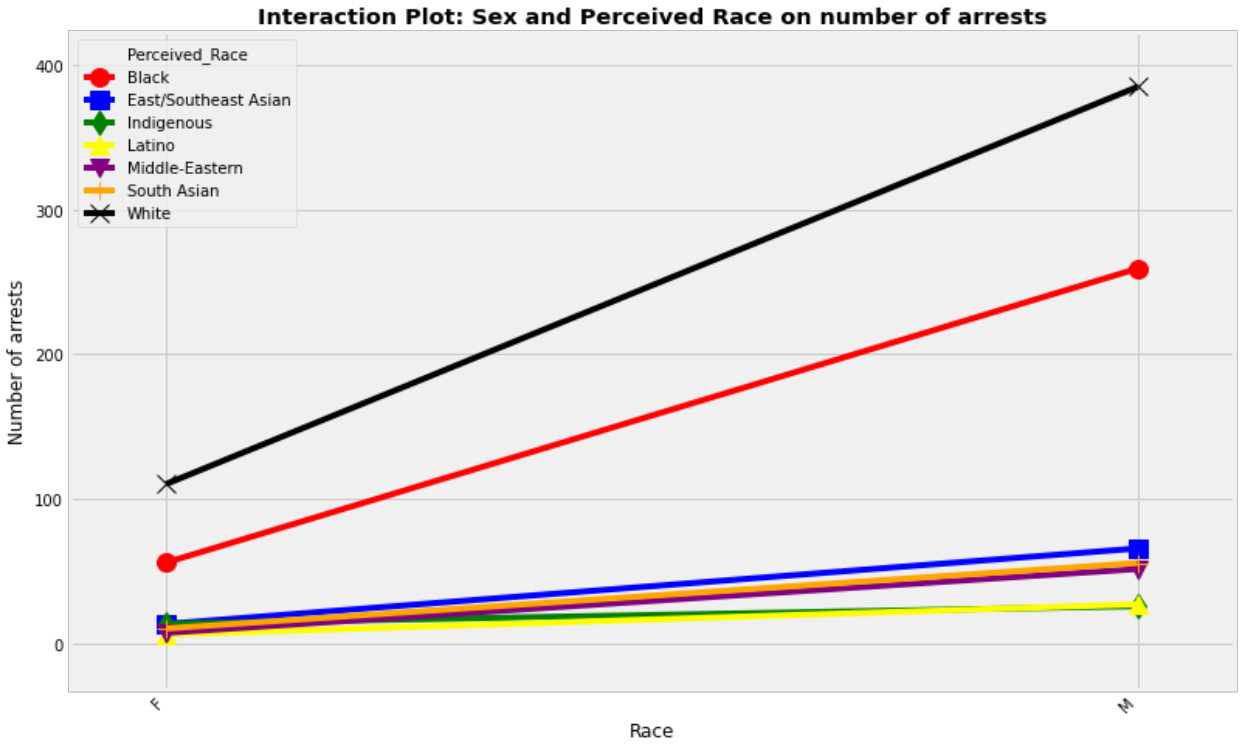
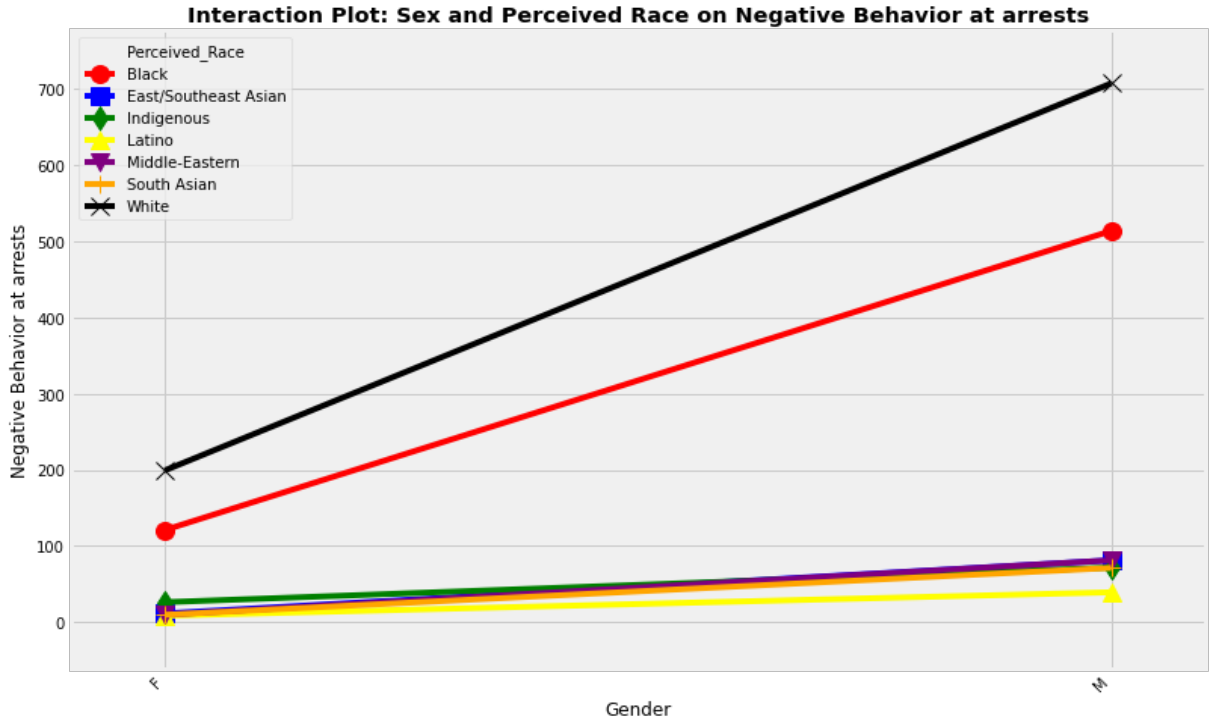


Figure 10



2.2 T-test

To explore the data further, we performed Welch's t-tests between different variables to check if there were any significant differences in the means. We use Welch's t-test rather than the one-sample or two-sample t-test for our analysis because the sample sizes of the two groups are different. In addition, by adjusting for unequal variances, Welch's t-test provides a more accurate test 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, and observed that 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 sexes. 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 results indicate that the mean number of arrests for males (M= 125.38, SD =189.37) is higher than for females (M= 32.66, SD = 54.42). In addition, with an alpha of 0.05 and a 95% CI [72.98, 112.47], we found that the p-value (1.04e-18) is lower than 0.05, indicating that it is of statistical significance, and we can reject the null hypothesis. In other words, if we randomly sample males and females from the dataset, the average number of arrests between males and females will most likely differ. Therefore, we will include the variable Sex to our ANOVA analysis.

Youth or Not and Number of Arrests

We calculated the mean number of arrests for people under 17 years old (Youth) and those over (Not Youth) respectively and observed that the Not Youth average was greater than Youth. Following that, we conducted a Welch's T-test to explore whether there is a significant difference in the mean number of arrests in terms of age. The following is our hypothesis:

- Null hypothesis H0: The population mean of the two independent groups, arrestee's age younger than 17 years old and those older than 17 years old, are equal in terms of the number of arrests.
- Alternative hypothesis H1: The population mean of the two independent groups, arrestee's age younger than 17 years old and those older than 17 years old, are not equal in terms of the number of arrests.

Our results indicate that the mean number of arrests for Not Youth (M= 89.11, SD = 157.79) is higher than for Youth (M= 27.97, SD = 40.33). In addition, with an alpha of 0.05 and a 95% CI [-75.68, -46.61], we found that the p-value (8.94e-16) is lower than 0.05, indicating that it is statistical significance, and we can reject the null hypothesis. In other words, if we randomly sample Youth and Not Youth from the dataset, the average number of arrests between them will most likely differ. However, since there is a significantly larger number of Not Youth than Youth in the dataset, we decided not to include Youth in our ANOVA analysis.

Year and Number of Arrests

We calculated the mean number of arrests between 2020 and 2021, respectively, and observed that the average in 2021 was slightly higher than in 2020. 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 years. The following is our hypothesis:

- Null hypothesis H0: The population mean of the two independent groups, 2020 and 2021, are equal in terms of the number of arrests.
- Alternative hypothesis H1: The population mean of the two independent groups, 2020 and 2021, are not equal in terms of the number of arrests.

Our results indicate that the mean number of arrests for 2021 (M= 83.15, SD = 151.82) is higher than for 2020 (M= 78.51, SD = 146.13). In addition, with an alpha of 0.05 and a 95% CI [-26.08, 16.80], we found that the p-value (0.67) is greater than 0.05, indicating that it is not statistical significance, and we cannot reject the null hypothesis. In other words, if we randomly sample 2020 and 2021 from the dataset, the average number of arrests between these years will most likely not differ. Hence, since there is statistical significance between these two years, we decided not to include Year in our ANOVA analysis.

Sex and Negative Behavior at the time of Arrests

Similarly we conducted another Welch's T-test to explore the significant difference in means of Negative Behavior for males and females respectively. The following was our hypothesis:

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

Our results indicate that the mean number of arrests for males (M= 223.61, SD =414.3) is higher than for females (M= 55.14, SD = 112.26). In addition, with an alpha of 0.05 and a 95% CI [3.01, 333.92], we found that the p-value (0.047) is lower than 0.05, Thus, average of random samples of negative behavior at times of arrest would differ for male and female. Thus, we included Sex variable to our ANOVA analysis.

Youth and Not a Youth and Negative Behavior at the time of Arrests

Next, We conducted another Welch's T-test to explore the significant difference in means of Negative Behavior for Youth and Not Youth respectively. The following was our hypothesis:

- Null hypothesis H0: The population mean of the two independent groups, arrestee's age younger than 17 years old and those older than 17 years old, are equal in terms of the negative behavior at times of arrests.
- Alternative hypothesis H1: The population mean of the two independent groups, arrestee's age younger than 17 years old and those older than 17 years old, are not equal in terms of the negative behavior at times of arrests.

The mean of negative behavior for Youth (M=7.36, SD =11.28) was very low as compared to the mean for negative behavior of Not Youth (M = 271.39, SD= 403.35). Even Though the p-value was less than 0.05, we did not include this variable in our analysis due to the larger number of data available for Not Youth category as compared to the Youth category.

Year and Negative Behavior at the time of Arrests

Lastly, we conducted another Welch's T-test to explore the significant difference in means of Negative Behavior for males and females respectively. The following was our hypothesis:

- Null hypothesis H0: The population mean of the two independent groups, 2020 and 2021 are equal in terms of the negative behavior at times of arrests.
- Alternative hypothesis H1: The population mean of the two independent groups, 2020 and 2021, are not equal in terms of the negative behavior at times of arrests.

The mean of negative behavior for people arrested in 2020 (M=137.57, SD =313.29) was lower than the mean for negative behavior of people arrested in 2021 (M = 141.18, SD= 317.5). However, the p-value 0.97 obtained for the t-test was higher than 0.05. Thus the difference was not statistically significant and we did not include this variable in our ANOVA test analysis.

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 6 columns: Arrest_Year, Arrest_Month, Perceived_Race, Sex, Youth_at_arrest_under_18years, Age group at arrest. We counted the number of times each of these combinations of variables were booked for any offenses and renamed to Number_arrests. This helped us in analyzing the patterns of the number of arrests and understanding the demographics of the arrested population 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 ANOVA Tests

For ANOVA analysis, we checked the following 4 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 observation.

However, using the Levene's test, we found that the population variances are not equal for the number of arrests but equal for negative behavior. We will discuss this limitation in the discussion section below.

3.3.1 One-Way ANOVA with Sex (First IV of RQ1):

We calculated the mean number of arrests for males and females, respectively, and observed that the male average was greater than the female average. Following that, we conducted a one-way ANOVA to explore whether there is a significant difference in the mean number of arrests between the sexes. 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.

The one-way ANOVA test yields a p-value of 3.31e-18. Since the p-value is less than the significance level of 0.05, we can reject the null hypothesis and conclude that there is a significant difference in the average number of arrests between males and females. In addition, the Tukey test also provides the same result.

3.3.2 One-Way ANOVA with Race (Second IV of RQ1):

We calculated the mean number of arrests for different racial groups and observed that the mean for White and Black groups was greater than that for other racial groups. Following that, we conducted a one-way ANOVA to explore whether there is a significant difference in the mean number of arrests between different racial groups. The following is our hypothesis:

- Null hypothesis H0: The population mean of the seven independent groups, White, Black, East/South East, Indigenous, South Asian, Middle Eastern, and Latino, are equal in terms of the number of arrests.
- Alternative hypothesis H1: The population mean of the seven independent groups, White, Black, East/South East, Indigenous, South Asian, Middle Eastern, and Latino, are not equal in terms of the number of arrests.

The one-way ANOVA test yields a p-value of 2.73e-58. Since the p-value is less than the significance level of 0.05, we can reject the null hypothesis and conclude that there is a significant difference in the average number of arrests between racial groups. In addition, To determine which groups differ significantly from one another, we conduct a Tukey test.

The Tukey test shows that the average number of arrests for Black is significantly different from the average number of the other racial group (p-value = 0.001) (Table 3). Additionally, the number of arrests of East/South East, Indigenous, South Asian, Middle Eastern and Latino are all significantly different from that of White (p-value = 0.001). Hence, we can conclude that there is a significant difference in the average number of arrests by different racial groups and that the average number of arrests of Whites and Blacks is significantly different from the other racial groups.

Table 3

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-119.9076	0.001	-168.8987	-70.9166	True
Black	Indigenous	-139.5926	0.001	-190.2879	-88.8973	True
Black	Latino	-141.6573	0.001	-192.0838	-91.2308	True
Black	Middle-Eastern	-129.0749	0.001	-178.6299	-79.52	True
Black	South Asian	-125.2424	0.001	-174.9157	-75.569	True
Black	White	88.1906	0.001	39.1995	137.1816	True
East/Southeast Asian	Indigenous	-19.6849	0.9	-70.1666	30.7967	False
East/Southeast Asian	Latino	-21.7496	0.8461	-71.9613	28.462	False
East/Southeast Asian	Middle-Eastern	-9.1673	0.9	-58.5036	40.169	False
East/Southeast Asian	South Asian	-5.3347	0.9	-54.7899	44.1205	False
East/Southeast Asian	White	208.0982	0.001	159.3283	256.8681	True
Indigenous	Latino	-2.0647	0.9	-53.9405	49.8112	False
Indigenous	Middle-Eastern	10.5176	0.9	-40.5114	61.5467	False
Indigenous	South Asian	14.3502	0.9	-36.7938	65.4942	False
Indigenous	White	227.7832	0.001	177.3015	278.2648	True
Latino	Middle-Eastern	12.5823	0.9	-38.1797	63.3444	False
Latino	South Asian	16.4149	0.9	-34.4627	67.2925	False
Latino	White	229.8479	0.001	179.6362	280.0595	True
Middle-Eastern	South Asian	3.8326	0.9	-46.1813	53.8465	False
Middle-Eastern	White	217.2655	0.001	167.9292	266.6018	True
South Asian	White	213.433	0.001	163.9778	262.8882	True

3.3.3 Two-Way ANOVA with Sex and Race

Following the one-way ANOVAs, we conducted a two-way ANOVA to explore the effects of two sex and race on the number of arrests. The following is our hypothesis:

- Null hypothesis H0:
  - There is no statistical significant difference in the mean of the number of arrests based on sex after controlling for race.
  - There is no statistical significant difference in the mean of the number of arrests based on race after controlling for sex.
  - There is no statistical significant interaction effect between sex and race on the mean number of arrests.
- Alternative hypothesis H1:
  - There is a statistical significant difference in the mean of the number of arrests based on sex after controlling for race.
  - There is a statistical significant difference in the mean of the number of arrests based on race after controlling for sex.
  - There is a statistical significant interaction effect between sex and race on the mean number of arrests.

The ANOVA table shows that both males and females have significant difference in mean numbers of arrests ($p < 0.05$), different racial groups have significant difference in mean numbers of arrests($p < 0.05$), and also had evidence about significant effect of sex on number of arrests across different racial groups ($p < 0.05$). Thus, we can reject the null hypotheses.

Table 4

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	5.384718e+06	6.0	84.125941	5.181618e-80
C(Sex)	1.731873e+06	1.0	162.343237	1.004823e-33
C(Perceived_Race):C(Sex)	1.716013e+06	6.0	26.809428	6.030188e-29
Residual	7.798286e+06	731.0	NaN	NaN

Subsequently, we conducted a Tukey's test identify which pairs of groups have significantly different means in this case. Only significant relationships are displayed in Table 5 out of the total 91 comparisons. We observed from the test results of the Tukey-HSD and found out that groups of both sex for Black and White racial groups had significant differences with groups of both sex for remaining races.

Table 5

group1	group2	meandiff	p-adj	lower	upper	reject
F,Black	M,Black	203.4008	0	137.1122	269.6894	TRUE
F,Black	M,White	329.1151	0	262.8265	395.4037	TRUE
F,East/Southeast Asian	F,White	96.8393	0.0001	31.1561	162.5225	TRUE
F,East/Southeast Asian	M,Black	245.875	0	180.1918	311.5582	TRUE
F,East/Southeast Asian	M,White	371.5893	0	305.9061	437.2725	TRUE
F,Indigenous	F,White	96.7599	0.0002	28.0041	165.5156	TRUE
F,Indigenous	M,Black	245.7956	0	177.0398	314.5513	TRUE
F,Indigenous	M,White	371.5099	0	302.7541	440.2656	TRUE
F,Latino	F,White	104.1429	0.0001	34.1243	174.1614	TRUE
F,Latino	M,Black	253.1786	0	183.16	323.1971	TRUE
F,Latino	M,White	378.8929	0	308.8743	448.9114	TRUE
F,Middle-Eastern	F,White	103.0056	0	35.7318	170.2794	TRUE
F,Middle-Eastern	M,Black	252.0413	0	184.7675	319.3151	TRUE
F,Middle-Eastern	M,White	377.7556	0	310.4818	445.0294	TRUE
F,South Asian	F,White	99.9629	0.0001	32.3379	167.5879	TRUE
F,South Asian	M,Black	248.9986	0	181.3736	316.6236	TRUE
F,South Asian	M,White	374.7129	0	307.0879	442.3379	TRUE
F,White	M,Black	149.0357	0	83.3525	214.7189	TRUE
F,White	M,Indigenous	-84.5546	0.0021	-151.829	-17.2808	TRUE
F,White	M,Latino	-83.3036	0.0018	-148.987	-17.6204	TRUE
F,White	M,White	274.75	0	209.0668	340.4332	TRUE
M,Black	M,East/Southeast Asian	-193.643	0	-259.326	-127.96	TRUE
M,Black	M,Indigenous	-233.59	0	-300.864	-166.317	TRUE
M,Black	M,Latino	-232.339	0	-298.023	-166.656	TRUE
M,Black	M,Middle-Eastern	-207.875	0	-273.558	-142.192	TRUE
M,Black	M,South Asian	-203.75	0	-269.433	-138.067	TRUE
M,Black	M,White	125.7143	0	60.0311	191.3975	TRUE
M,East/Southeast Asian	M,White	319.3571	0	253.6739	385.0404	TRUE
M,Indigenous	M,White	359.3046	0	292.0308	426.5785	TRUE
M,Latino	M,White	358.0536	0	292.3704	423.7368	TRUE
M,Middle-Eastern	M,White	333.5893	0	267.9061	399.2725	TRUE
M,South Asian	M,White	329.4643	0	263.7811	395.1475	TRUE

3.3.4 One-Way ANOVA with Sex (First IV of RQ2):

We observed that the mean value of negative behavior at the time of arrests was significantly different between sexes (p value < 0.05). The following is our hypothesis:

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

Therefore, we can reject the null hypothesis and conclude that there is a significant difference in negative behaviour at arrest between males and females. In addition, the Tukey test also provides the same result.

3.3.5 One-Way ANOVA with Perceived Race (Second IV of RQ2):

We observed that the mean value of negative behavior at the time of arrests was significantly different between different racial groups (p value < 0.05). The following is our hypothesis:

- Null hypothesis H0: The population mean of the seven independent groups, White, Black, East/South East, Indigenous, South Asian, Middle Eastern, and Latino, are equal in terms of negative behaviour at arrest.
- Alternative hypothesis H1: The population mean of the seven independent groups, White, Black, East/South East, Indigenous, South Asian, Middle Eastern, and Latino are not equal in terms of negative behaviour at arrest.

Therefore, we can reject the null hypothesis and conclude that there is a significant difference in negative behaviour at arrest between different races (Table 6). However, the Tukey test provides a different result. The post hoc test shows that all the groups are not statistically significant with each other (p value > 0.05).

Table 6

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-270.875	0.4843	-706.7938	165.0438	False
Black	Indigenous	-268.25	0.4953	-704.1688	167.6688	False
Black	Latino	-293.375	0.3884	-729.2938	142.5438	False
Black	Middle-Eastern	-271.75	0.4807	-707.6688	164.1688	False
Black	South Asian	-276.75	0.4594	-712.6688	159.1688	False
Black	White	135.875	0.9	-300.0438	571.7938	False
East/Southeast Asian	Indigenous	2.625	0.9	-433.2938	438.5438	False
East/Southeast Asian	Latino	-22.5	0.9	-458.4188	413.4188	False
East/Southeast Asian	Middle-Eastern	-0.875	0.9	-436.7938	435.0438	False
East/Southeast Asian	South Asian	-5.875	0.9	-441.7938	430.0438	False
East/Southeast Asian	White	406.75	0.0823	-29.1688	842.6688	False
Indigenous	Latino	-25.125	0.9	-461.0438	410.7938	False
Indigenous	Middle-Eastern	-3.5	0.9	-439.4188	432.4188	False
Indigenous	South Asian	-8.5	0.9	-444.4188	427.4188	False
Indigenous	White	404.125	0.0859	-31.7938	840.0438	False
Latino	Middle-Eastern	21.625	0.9	-414.2938	457.5438	False
Latino	South Asian	16.625	0.9	-419.2938	452.5438	False
Latino	White	429.25	0.0562	-6.6688	865.1688	False
Middle-Eastern	South Asian	-5.0	0.9	-440.9188	430.9188	False
Middle-Eastern	White	407.625	0.0812	-28.2938	843.5438	False
South Asian	White	412.625	0.0747	-23.2938	848.5438	False

3.3.6 Two-Way ANOVA with Sex and Race

We calculated the mean negative behavior at time of arrests by sex and race and observed that the male average was greater than the female average. The following is our hypothesis:

- Null hypothesis H0:
  - There is no statistical significant difference in the mean of the negative behavior at times of arrest based on sex after controlling for race.
  - There is no statistical significant difference in the mean of the behavior at times of arrest based on race after controlling for sex.
  - There is no statistical significant interaction effect between sex and race on the mean of the negative behavior at times of arrest .
- Alternative hypothesis H1:
  - There is a statistical significant difference in the mean of the negative behavior at times of arrest based on sex after controlling for race.
  - There is a statistical significant difference in the mean of the negative behavior at times of

arrest based on race after controlling for sex.

- There is a statistical significant interaction effect between sex and race on the mean of the negative behavior at times of arrest .

The Table 7 shows that both males and females have significant difference in the mean negative behaviour at arrests ($p < 0.05$), different racial groups have significant difference in the mean negative behaviour at arrests($p < 0.05$), but there was no evidence about significant effect of sex in mean negative behaviour at arrests across different racial groups ($p > 0.05$).

Table 7

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	1.430594e+06	6.0	3.246973	0.010322
C(Sex)	3.973230e+05	1.0	5.410748	0.024913
C(Perceived_Race):C(Sex)	4.599604e+05	6.0	1.043957	0.411086
Residual	3.084152e+06	42.0	NaN	NaN

Subsequently, we conducted a Tukey's test identify which pairs of groups have significantly different means in this case. Only significant relationships are displayed in Table 8 out of the total 91 comparisons. We observed from the test results of the Tukey-HSD and found out that groups of only Male sex for White race had significant differences with groups of female sex for minority races East/Southeast Asian, Latino, Middle-Eastern and South Asian.

Table 8

group 1	group 2	meandiff	p-adj	lower	upper	reject
F,East/Southeast Asian	M,White	695.5	0.0415	14.0784	1376.9216	TRUE
F,Latino	M,White	698.5	0.0399	17.0784	1379.9216	TRUE
F,Middle-Eastern	M,White	697.25	0.0405	15.8284	1378.6716	TRUE
F,South Asian	M,White	697.25	0.0405	15.8284	1378.6716	TRUE

4 Results and Findings

We derived some interesting insights about the underlying data considering several aspects and factors of the data. The two research questions were aimed to analyze how the number of arrests

and the negative behavior at the times of arrests differed on the basis of person's sex and the ethnicity. Overall, the results drew light on some interesting insights about the factors that affected the number of arrests and negative behavior of the person during the time of the arrests.

4.1 Research Question 1

According to the results of our analysis, sex and race (Independent Variables) of a person did have an effect on our dependent variable number of arrests. The interaction effect of each of these independent variables and their combined effect on our dependent variable were of statistical significance. While we observed that male sex had more influence on the number of arrests, when the race of the arrested person was considered, the mean number of arrests differed for males and females of different races. For instance, the mean number of arrests of a white male would differ from the mean number of arrests of a male or female of the other considered cases. It indicates that the race of the person would have an impact on the sex of the arrested person and vice versa. The results of this analysis encouraged us to explore which sex and race had negative behavior during the times of the arrests.

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 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).

Link: shorturl.at/jmN39