INF 2178 FINAL WRITE UP

"Investigating the relationship between Age, Gender, Race, Number of arrests, Arrestee behavior at the time of arrest"

Group 52: Cunrong Liu, Huayu Wu, Ke Wu

Instructor: Shion Guha

Table of Contents

Introduction

As the main force for investigating and preventing crimes, the police are often faced with various critical situations. In order to stop harmful behaviors promptly, and protect the personal safety of citizens and policemen as well as the property safety of citizens and the country, the law grants the police the power to use force. Detailed data on police activity, including the identities and backgrounds of suspects or arrestees, is kept in U.S. federal and state records. Statistics show that 90 percent of those killed by police in the past five years were radicals (Published by Statista Research Department & 2, 2023), with the highest rate of radicals being black compared to other races. But an individual's behavior during an arrest may be influenced by different factors, including their personal history and experiences, their mental health, the circumstances under which they were arrested, and their interactions with law enforcement officers. Through this dataset, our motivation is to examine some behaviors of different ages, genders, and skin colors when they are arrested, and to study whether certain groups have a greater probability of showing negative arrest actions when they are arrested and whether different ages, genders and skin colors are encountered when they are arrested. Differential treatment, which is mainly manifested in strip searches. This facilitates discussions around issues related to policing and community policing. The studies are important because they raise issues related to fairness and justice in the criminal justice system and stereotypes about certain groups. By studying the ways that race, gender, and age intersect with criminal justice practice, we can work towards a more just and equitable system that treats all individuals with fairness and dignity. In the first half of report, our main test methods include: T-test, One-way ANOVA and Two-way ANOVA. We mainly studied the relationship with negative arrest actions based on race, age, and gender. Our research questions are two:

1. Do negative arrest actions have anything to do with race, and what influences strip searches?
2. Do negative arrest actions have a relationship with sex and age?

In the second half of the report, our main test methods include: T-test, power analysis, ANCOVA and logistic regression. Our research question are following:

1. Whether there is a significant difference in mean of total adverse actions between male and female suspects, after controlling for the effect of the number of arrests.
2. How do demographic factors (age, race), total adverse actions during arrest, and the number of arrests predict the likelihood of being male suspects versus female suspects in the criminal justice system?

Literature review

There has always been a dispute between race and arrest. The research from D'Alessio and Stolzenberg in 2003 assessed the effect of race on the probability of arrest for nearly 340,000 crimes and found that white and black criminals were arrested at roughly similar rates (D'Alessio & Stolzenberg, 2003). The rise in black arrests might simply reflect their greater involvement in crime, rather than racism by law enforcement officers. The analysis showed that whites were more likely than blacks to be arrested for robbery, aggravated assault, and common assault (D'Alessio & Stolzenberg, 2003). There was no significant difference in arrest rates between blacks and whites for rape. And this analysis refuted the point that police bias affected the arrest rate of blacks. D'Alessio also stated that black citizens' distrust of the police is also one of the reasons for this problem because black citizens were unwilling to report or testify about crimes committed by other black citizens, which led to an increase in the arrest rate of white criminals (D'Alessio & Stolzenberg, 2003). In addition, whites generally had higher expectations for arrests. In the second article： who commits crime, different aspects of the causes and associations between crime rates are described. Firstly, it mentioned that the crime rate of men was much higher than that of women and this significant difference could be attributed to the socialization of gender roles and the chances of boys spending nights out alone more than girls (Newburn, 2018). Secondly, age also affected crime rates as well especially at a young age around teens to twenties. Teenagers and older adults who lack full-time employment were also more likely than other age groups to commit crimes to obtain money when they need it (Newburn, 2018). Thirdly, the crime rate influenced by social class is not as clear as the first two (Newburn, 2018). It was interesting that in the arrest data, poor people were more likely than rich people to commit a street crime but rich people were more likely than poor people to commit a white-collar crime which was more harmful than street crime. Therefore, from this perspective, the relationship between social class and crime was not significant because of different emphases. The fourth point was urban vs rural residential and urban had higher crime rates than rural areas. The key factor was that urban had a high population density. The last point was race. The article mentioned race was bound to be associated with racial discrimination. But over time, the race differences in crime were also related to the previous reasons. For example, whites commit more white-collar crimes. What's more, the average level of African-Americans and Hispanics was poorer than whites, which may lead to high street crime rates. Experienced racism

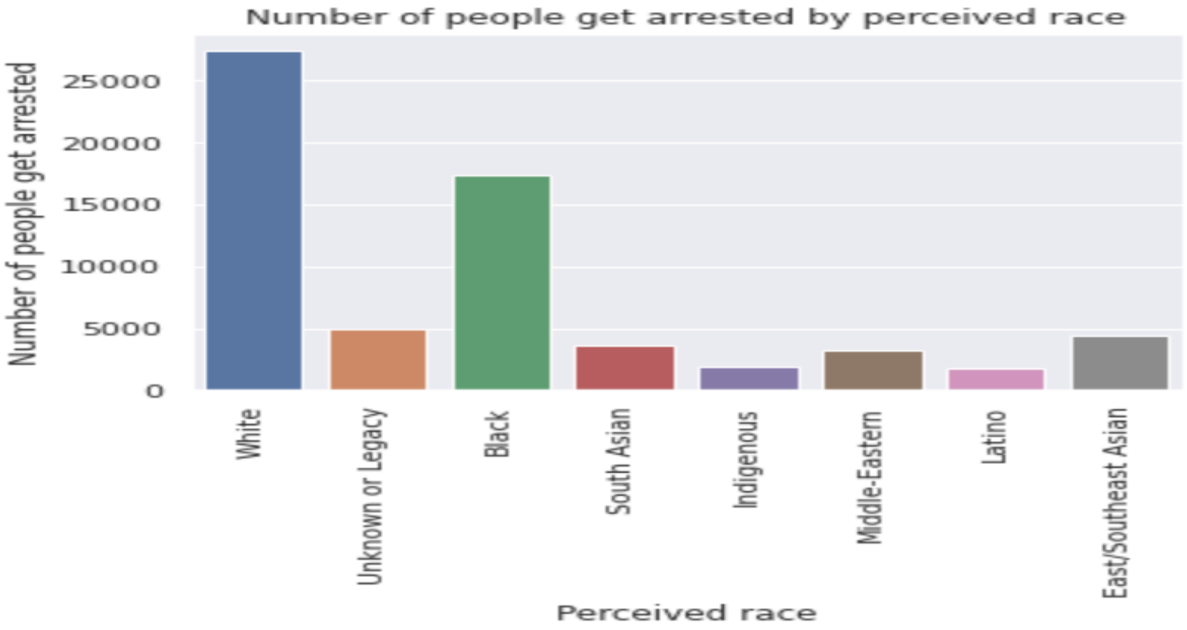
also led to anger and frustration that fuel criminal behavior. In the third article, Piquero and Brame used official records and self-report data from samples of serious juvenile offenders in Philadelphia and Phoenix to better understand the relationship between race and criminal activity in 2008. (Piquero & Brame, 2008) stated that in the research, white men in Philadelphia self-reported crime more frequently than Hispanics and blacks, while the results for men and women in Phoenix were similar to those for women in Philadelphia. Median self-reported crime frequencies did not appear to differ between racial groups. In the analysis of crime diversity scores, scores among Philadelphia male racial groups were greater than self-reported crime frequency (Piquero&Brame, 2008). The black crime rate was only a small part of Phoenix males. Overall, prior-year arrest brigades did not differ significantly between racial and ethnic groups and self-reported criminal diversity was positively associated with expected official arrests (Piquero&Brame, 2008). The article published by Statistics Canada titled "Police-reported crime statistics in Canada, 2020" provides an overview of crime rates and trends in Canada during the year 2020. According to the article, there was an increase in the rate of certain types of violent crimes in Canada in 2020, including homicides and firearm-related crimes. This increase in violent crime may be reflective of a range of factors, including social and economic inequality, substance use disorders, and mental health issues, which can impact the behavior of offenders upon arrest (Statistics Canada 2022). Research has shown that the behavior of offenders upon arrest can vary widely depending on a range of factors, including the severity of the crime, the individual's prior criminal history, and the presence of substance use or mental health issues. For example, individuals with a history of violence or substance use disorders may be more likely to exhibit aggressive or uncooperative behavior when arrested. Efforts to reduce adverse outcomes during arrest, such as excessive use of force, can benefit from a better understanding of the underlying factors that contribute to criminal behavior (Statistics Canada 2022). By addressing the root causes of criminal behavior, such as poverty, lack of access to education and job opportunities, and social inequality, it may be possible to reduce the incidence of criminal behavior and improve outcomes for both offenders and the broader community (Statistics Canada 2022).

EDA

1. Descriptive Statistics

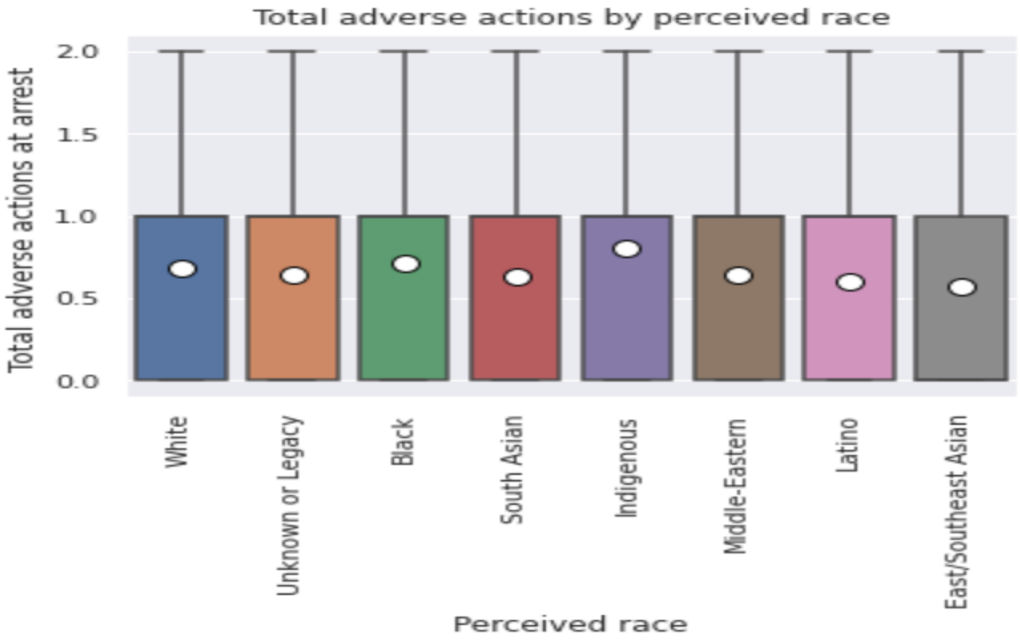
	Race	Sex	Age	Occurrence category	Action at arrest: concealed items	Action at arrest:Combative, violent or spitter/biter	Action at arrest: resistd, defensive or escape risk	Action at arrest: mental instability	Action at arrest: assaulted officer	Action at arrest: cooperative	Total adverse actions	Strip search
Count	64615	64615	64615	64615	64615	64615	64615	64615	64615	64615	64615	64615
Mean	NaN	NaN	NaN	NaN	0.0041	0.0444	0.0385	0.0336	0.0064	0.5506	1.9233	0.113
Std	NaN	NaN	NaN	NaN	0.0639	0.2060	0.1923	0.1802	0.0797	0.4974	1.4537	0.317
Min	NaN	NaN	NaN	NaN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
Max	NaN	NaN	NaN	NaN	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	11.0000	1.0000

2. Data Visualization



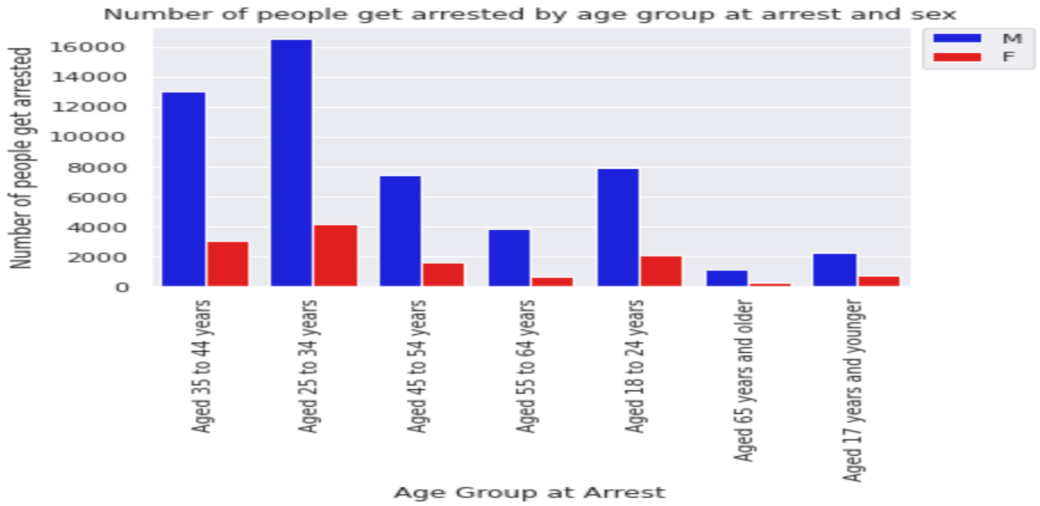
Plot 1: Number of people get arrested by perceived race

In this plot, the x-axis represents the different races of the arrested individuals, while the y-axis shows the total number of people arrested by the police. It can be seen that white people account for the largest proportion, followed by black and brown people.



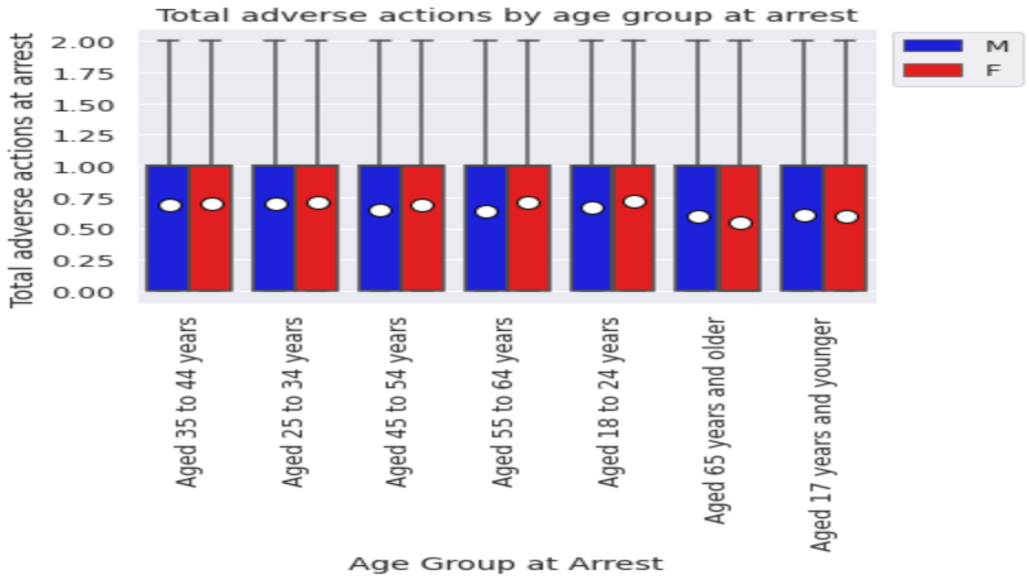
plot 2: Total adverse action by perceived race

This plot is a box chart that displays the distribution of total adverse actions at arrest by perceived race. The x-axis represents the different perceived races, and the y-axis represents the total adverse actions at arrest. Median values for black and indigenous individuals were slightly higher than those for other races, suggesting that these groups tend to have more negative behavior when arrested.



Plot 3: Number of people get arrested by age group at arrest and sex

This boxplot shows the number of arrests by age group and gender at the time of the arrest. The x-axis represents the age group at arrest, while the y-axis represents the number of people who get arrested. The bars are grouped by gender, with blue representing males and red representing females. The plot shows that the 25-34 age group made the most arrests, followed by the 35-44 age group.The number of arrests is the lowest in the two age groups：Aged 65 and older & aged 17 years and younger.

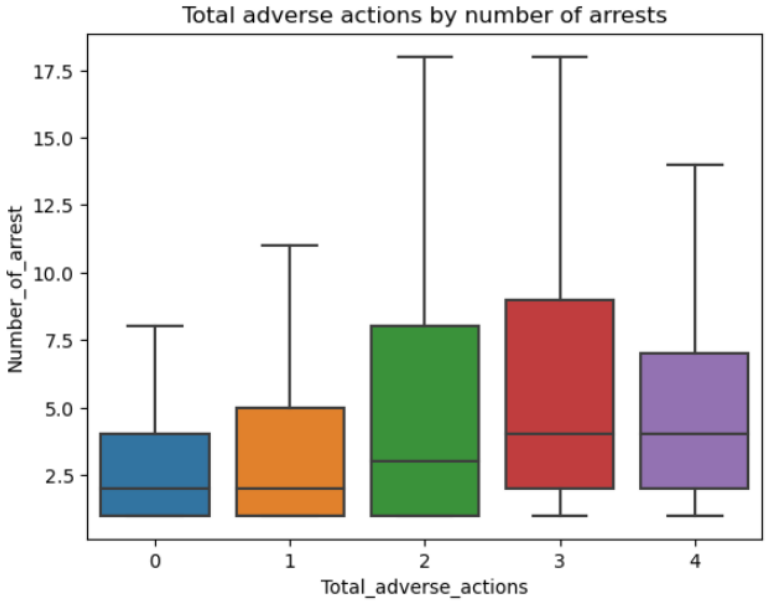


Plot 4: Total adverse actions by age group at arrest

This is a box plot that shows the distribution of the total adverse actions at arrest for different age groups at arrest. The x-axis represents the age group at the time of arrest, and the y-axis represents the total adverse actions at arrest. The plot is split into two groups based on sex, with males shown in blue and females shown in red. Median values for individuals aged 25 to 34 years and aged in 18 to 24 were slightly higher than those for other age groups, suggesting that these groups tend to have more negative behavior when arrested. And the median is lowest as age goes beyond 65, or when they are below 18.

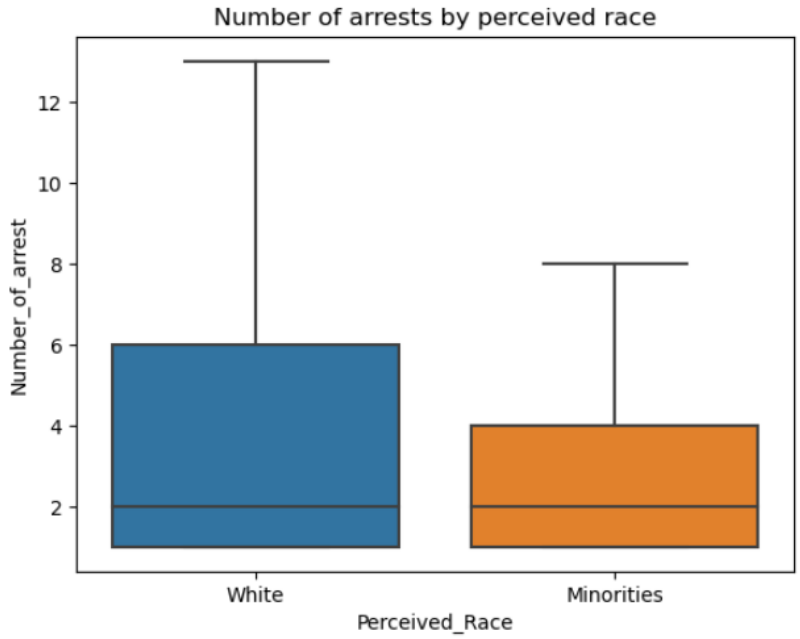
Plot 5: Adverse actions by youth at arrest under 18 years

This plot shows the distribution of the total adverse actions at arrest taken by police against youth at the time of their arrest, based on whether they were youths or not. From this graph, we can observe that people who are not youth have a slightly higher median rate of misconduct than adults when they are arrested.



Plot 6: Total adverse actions by number of arrests

This plot shows the distribution of total adverse actions at arrest for different numbers of arrests. From this graph, we can observe that people who have been arrested more times (higher median for the number of arrests) tend to exhibit more adverse actions when they are arrested. Since both variables are continuous, we could have used a scatter plot, but instead, we used a box plot because total adverse actions behave like a categorical variable with a limited number of values.



Plot 7: Number of arrests by perceived race

The box plot displays the distribution of the number of arrests for two racial groups, white and minorities. It indicates that the median number of arrests for the white group is slightly higher than that for the minorities group. Moreover, the spread of data for the white group is more extensive than that for the minorities group.

3. T-test (T)

Because t-tests are able to provide statistical measures of the significance of the difference between two means while taking into account the variability in the data, therefore, we ran Welch's t-tests with the categorical attributes in the dataset. We chose Welch's t-test instead of the student's t-test since equal variance among the residuals was not assumed. We checked the assumptions before implementing Welch's t-test and confirmed that the observations are independent, the data for each group are normally distributed, and there are no significant outliers. The following paragraphs demonstrate the way we ran the t-test as well as the results of the t-tests.

T-test 1:Sex and adverse actions at arrest

We conducted the first t-test to analyze whether the suspects' adverse actions(outcome variables) at arrest differ between males and females (Two-level explanatory variables). The hypothesis being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, female suspects and male suspects are equal.

Ha (Alternative Hypothesis): The population means of the two independent groups, female suspects and male suspects are different.

	Statistic	P-value	Mean F	Mean M	Sd F	Sd M
0	3.147228	0.001651	0.69584	0.673128	0.725651	0.720153
				CI	DOF	
0	(0.008567046580899382, 0.03685724390388165)				18850.712659	

Table 1: T-test Sex and adverse actions at arrest result

The results indicate that the mean adverse actions at arrest for female suspects (M=0.6958, SD=0.7257) are higher than the mean adverse actions at arrest for male suspects (M=0.6731, SD=0.7202). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.00165) is less than 0.05, 95% CI [0.0086, 0.0369]. Therefore, we can reject the null hypothesis that there is no difference in adverse actions at arrest for female suspects and male suspects.

T-test 2:Age(Youth) and adverse actions at arrest

We conducted the Second t-test to analyze whether the suspects' adverse actions(outcome variables) at arrest differ between youth and adults(Two-level explanatory variables). The hypothesis being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, youth suspects and adult suspects are equal.

Ha (Alternative Hypothesis): The population means of the two independent groups, youth suspects and adult suspects are different.

	Statistic	P-value	Mean Y	Mean A	Sd Y	Sd A
0	-6.646632	3.477880e-11	0.602258	0.681249	0.63222	0.725195
				CI		
0	(-0.10229304318502694, -0.05569026418524628)					

Table 2: T-test Age(Youth) and adverse actions at arrest result

The results indicate that the mean adverse actions at arrest for youth suspects (M=0.6023, SD=0.6812) are higher than the mean adverse actions at arrest for adult suspects (M=0.6322, SD=0.7252). With alpha established at 0.05, this is a statistically significant difference as the p-value (3.47e^-11) is less than 0.05, 95% CI [-0.1023, -0.0557]. Therefore, we can reject the null hypothesis that there is no difference in adverse actions at arrest for youth suspects and adult suspects.

T-test 3:Asian&other race and adverse actions at arrest

We conducted the third t-test to analyze whether the suspects' adverse actions(outcome variables) at arrest differ between Asian and other races. The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, Asian suspects and other races suspects are equal.

Ha (Alternative Hypothesis): The population means of the two independent groups, Asian suspects and other races suspects are different.

	Statistic	P-value	Mean A	Mean O	Sd A	Sd O	CIlower	CIhigher
0	-7.236391	4.850913e-13	0.593961	0.677241	0.650295	0.738973	-0.105838	-0.060721

Table 3: T-test Asian&other race and adverse actions at arrest result

The results indicate that the mean adverse actions at arrest for Asian suspects (M=0.5940, SD=0.6503) are lower than the mean adverse actions at arrest for other races suspects (M=0.6772, SD=0.7390). With alpha established at 0.05, this is a statistically significant difference as the p-value (4.85e^-13) is less than 0.05, 95% CI [-0.1058, -0.0607]. Therefore, we can reject the null hypothesis that there is no difference in adverse actions at arrest for Asian suspects and other races suspects.

T-test 4:Black&White race and adverse actions at arrest

We conducted the fourth t-test to analyze whether the suspects' adverse actions(outcome variables) at arrest differ between Black and White. The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, Black suspects and White suspects are equal.

Ha (Alternative Hypothesis): The population means of the two independent groups, Black suspects and White suspects are different.

	Statistic	P-value	Mean W	Mean B	Sd W	Sd B	CIlower	CIhigher
0	-4.582654	0.000005	0.683439	0.716344	0.726564	0.748612	-0.046979	-0.018832

Table 4: T-test Black&White race and adverse actions at arrest result

The results indicate that the mean adverse actions at arrest for White suspects (M=0.6834, SD=0.7266) are lower than the mean adverse actions at arrest for Black suspects (M=0.7163, SD=0.7486). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.000005) is less than 0.05, 95% CI [-0.0470, -0.0188]. Therefore, we can reject the null hypothesis that there is no difference in adverse actions at arrest for Black suspects and White suspects.

T-test 5:Asian&Black race and adverse actions at arrest

We conducted the fifth t-test to analyze whether the suspects' adverse actions(outcome variables) at arrest differ between Asian and Black (Two-level explanatory variables). The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, Asian suspects and Black suspects are equal.

Ha (Alternative Hypothesis): The population means of the two independent groups, Asian suspects and Black suspects are different.

	Statistic	P-value	Mean A	Mean B	Sd A	Sd B	CIlower	CIhigher
0	-13.252695	6.750427e-40	0.593961	0.716344	0.650295	0.748612	-0.140483	-0.104282

Table 5: T-test Asian&Black race and adverse actions at arrest result

The results indicate that the mean adverse actions at arrest for Asian suspects (M=0.5940, SD=0.6503) are lower than the mean adverse actions at arrest for Black suspects (M=0.7163, SD=0.7486). With alpha established at 0.05, this is a statistically significant difference as the p-value (6.75e^-40) is less than 0.05, 95% CI [-0.1405, -0.1043]. Therefore, we can reject the null hypothesis that there is no difference in adverse actions at arrest for Asian suspects and Black suspects.

T-test 6: Asian&other race and number of arrest

We conducted the sixth t-test to analyze whether the number of arrest(outcome variables) differ between Asian and other race (Two-level explanatory variables). The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, Asian suspects and other race suspects were arrested equally.

Ha (Alternative Hypothesis): The population means of the two independent groups, Asian suspects and other race suspects were arrested differently.

Ttest results for number of arrest: Ttest_indResult (statistic = -19.8322868452296, p-value = 4.683876588017642e-86)
Mean number of arrest, Asian, Other: 2.768292682926829 4.466744799296806

Standard deviation for number of arrest, Asian, Other: 3.859925803911002 6.1055784385087115								
Confidence interval for number of arrest: (-1.8663230938045714, -1.5305811389353823)								
	Statistic	P-value	Mean A	Mean O	Sd A	Sd O	CIlower	CIhigher
0	-19.832287	4.683877e-86	2.768293	4.466745	3.859926	6.105578	-1.866323	-1.530581

Table 6: T-test Asian & other race and number of arrest result

The results indicate that the mean number of arrests for Asian suspects (M= 2.768293, SD=3.859926) are lower than the mean number of arrests for Black suspects (M=4.466745, SD=6.105578). With alpha established at 0.05, this is a statistically significant difference as the p-value (4.683877e-86) is less than 0.05, 95% CI [-1.8663230938045714, -1.5305811389353823]. Therefore, we can reject the null hypothesis that there is no difference in the number of arrests for Asian suspects and other race suspects.

T-test 7: Black&White and number of arrest

We conducted the seventh t-test to analyze whether the number of arrest(outcome variables) differ between Black and White race (Two-level explanatory variables). The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, Black suspects and White race suspects were arrested equally.

Ha (Alternative Hypothesis): The population means of the two independent groups, Black suspects and White race suspects were arrested differently.

Confidence interval for number of arrest: (0.5387118880845995, 0.7431017899754385)								
	Statistic	P-value	Mean W	Mean B	Sd W	Sd B	CIlower	CIhigher
0	12.292133	1.160054e-34	4.557535	3.916628	5.573331	5.212871	0.538712	0.743102

Table 7: T-test Black&White race and number of arrest result

The results indicate that the mean number of arrests for Black suspects (M=3.916628, SD=5.212871) are lower than the mean number of arrests for White suspects (M= 4.557535, SD=5.573331). With alpha established at 0.05, this is a statistically significant difference as the p-value (1.160054e-34) is less than 0.05, 95% CI [0.5387118880845995, 0.7431017899754385].

Therefore, we can reject the null hypothesis that there is no difference in the number of arrests for Black suspects and White race suspects.

T-test 8: Asian&Black and number of arrest

We conducted the eighth t-test to analyze whether the number of arrest(outcome variables) differ between Asian and Black (Two-level explanatory variables). The hypotheses being tested are the following:

H0 (Null Hypothesis): The population means of the two independent groups, Asian suspects and Black suspects were arrested equally.

Ha (Alternative Hypothesis): The population means of the two independent groups, Asian suspects and Black suspects were arrested differently.

Confidence interval for number of arrest: (-1.2634778461834628, -1.0331927099322238)								
	Statistic	P-value	Mean A	Mean B	Sd A	Sd B	CIlower	CIhigher
0	-19.548199	2.540819e-84	2.768293	3.916628	3.859926	5.212871	-1.263478	-1.033193

Table 8: T-test Asian&Black race and number of arrest result

The results indicate that the mean number of arrests for Asian suspects (M=2.768293, SD=3.859926) are lower than the mean number of arrests for Black suspects (M= 3.916628, SD=5.212871). With alpha established at 0.05, this is a statistically significant difference as the p-value (2.540819e-84) is less than 0.05, 95% CI [-1.2634778461834628, -1.0331927099322238]. Therefore, we can reject the null hypothesis that there is no difference in number of arrest for Asian suspects and Black race suspects.

T-test concluding remarks

From the above T-tests, we found that the suspects' mean value of total adverse actions at arrest was statistically significant depending on their gender, age and race. Also, the suspects' mean value of the number of arrests was no big difference between their race group.

Research Design and Methods

Dataset Description

The Arrests and Strip Searches (RBDC-ARR-TBL-001) dataset contains information about all arrests and strip searches conducted by the Toronto Police Service between 2020 and 2021. This data includes the arrestee's age, race, gender, behavior at the time of arrest, and the reason for the search, as well as the location, time, and case code of the arrest.

Dependent variables: The dependent variable (DV) was not clearly defined. However, By performing data cleaning, We combine several variables, including Actions at arrest - Concealed items, Actions at arrest - Combative, violent or spitter/biter, Actions at arrest - Resisted, defensive or escape risk, Actions at arrest - Mental instability or possibly suicidal, Actions at arrest - Assaulted officer and Actions at arrest - Cooperative. we can consider the number of adverse arrest actions (Total adverse actions) as a potential DV reflecting negative arrest actions during the arrest.

Independent variables: Age group at arrest, Sex, Race, Number of arrest.

In addition, we also created a new variable named Number of Arrests, which represents how many times that a person is arrested in this dataset.

Based on our findings from the descriptive analysis and T-tests, we will use inferential statistical tests to explore our research questions.

Research Questions

Research question 1: From the EDA, the boxplot of total adverse actions at arrest vs perceived race illustrates that there is no big difference between each race group. We would like to use one-way ANOVA to explore if race has a relationship with total adverse actions at arrest, and how it influences total adverse actions at arrest. We checked the assumption before conducting the one-way ANOVA and confirmed that the data are independent, the distributions have the same variance and the responses for each factor level have a normal population distribution.

- H0 (Null Hypothesis): The population means of total adverse actions at arrest are equal for each perceived race.
- Ha (Alternative Hypothesis): The population means of total adverse actions at arrest of at least one perceived race are different from other perceived races.

Research question 2: From the T-test, we found that the suspects' mean value of total adverse actions at arrest was statistically significant depending on their gender as well as their age. Therefore, we would like to use two-way ANOVA to explore how the combination of gender and age changes the suspects' mean value of total adverse actions at arrest. We checked the assumption before conducting the one-way ANOVA and confirmed that the data are independent, the distributions have the same variance and the responses for each factor level have a normal population distribution.

- H0 (Null Hypothesis): The population means of total adverse actions at arrest are equal for each gender/age combination.
- Ha (Alternative Hypothesis): The population means of total adverse actions at arrest of at least one gender/age combination are different from other combinations.

New Research Questions

Research Question 3: From the T-test, we found that the suspects' mean value of the number of arrests was no big difference between each race group. Therefore, we would like to use ANCOVA to explore how the race changes the suspects' mean value of the number of arrests. We checked the assumption before conducting the ANCOVA and confirmed that the data are independent, the distributions have the same variance and the responses for each factor level have a normal population distribution.

$$\hat{Y}\ (\text{Perceived\_Race, Number\_of\_arrest}) = \beta 0 + \beta 1 \textit{Perceived\_Race[T.White]} + \textit{β2}\text{Number\_of\_arrest}$$

- H0 (Null Hypothesis): The population means of total adverse actions at arrest are equal for each perceived race, after controlling for the effect of number of arrest
- Ha (Alternative Hypothesis): The population means of total adverse actions at arrest of at least one perceived race are different from other perceived races, after controlling for the effect of number of arrest

Research Question 4: We want to investigate how demographic factors (age and race), total adverse actions during arrest, and the number of arrests predict the likelihood of being male suspects versus female suspects in the criminal justice system.

$$\hat{Y}\ (\text{Sex}) = \beta 0 + \beta 1 \text{Age[T.Youth]} + \beta 2 \text{Race[T.White]} + \beta 3 \text{Total\_Adverse\_Actions} + \beta 4 \text{Number\_of\_arrest}$$

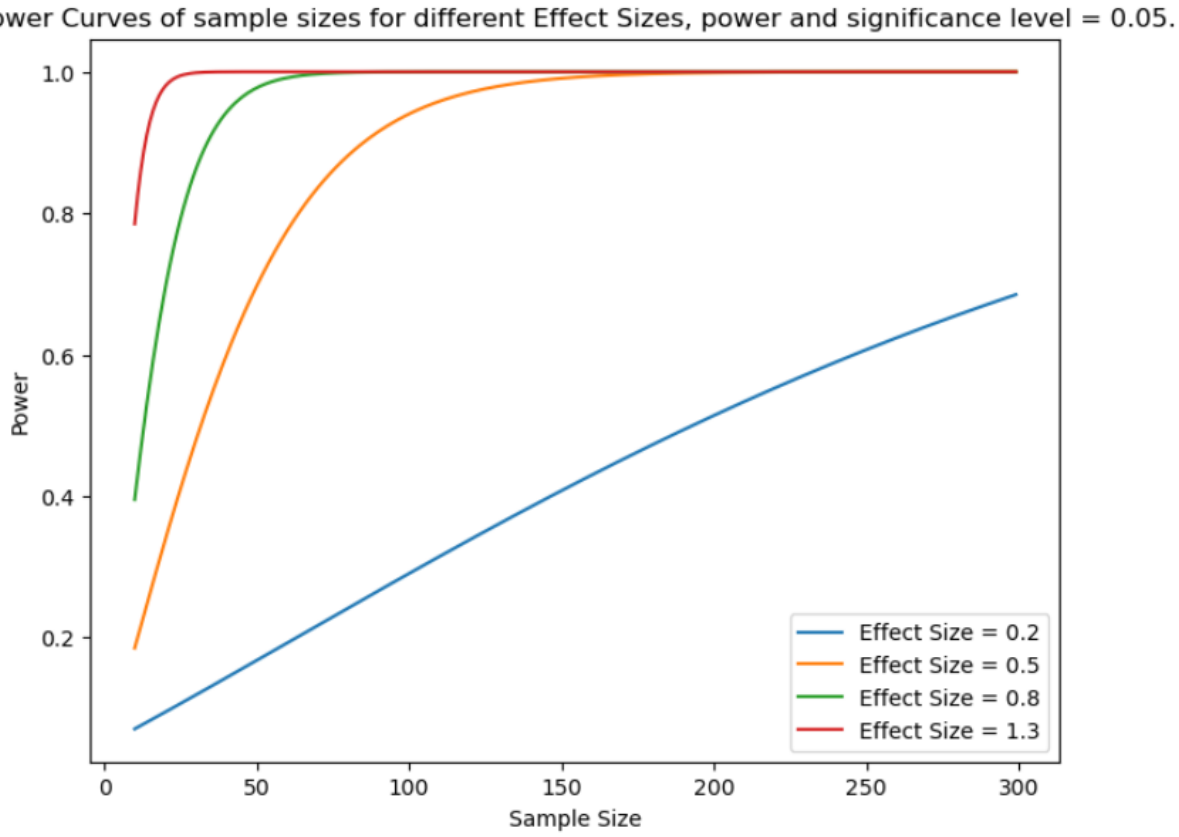
- H0 (Null Hypothesis): Demographic factors, total adverse actions during arrest, and the number of arrests do not significantly predict the likelihood of being male suspects versus female suspects in the criminal justice system.
- Ha (Alternative Hypothesis): Demographic factors, total adverse actions during arrest, and the number of arrests significantly predict the likelihood of being male suspects versus female suspects in the criminal justice system.

Power Analysis

In this experiment, we decided to perform two types of power analysis: prior power analysis and retrospective power analysis.

1. **A priori power analysis:**
   For any given experiments, We desired to perform this analysis before data collection, which involved determining the appropriate sample size based on the desired power level, effect size, and significance level.

Initially, we plotted a power curve to observe the general patterns for how different power and effect size determine the sample sizes.



Plot 8: Power curve of sample sizes for different effect sizes

We subsequently computed the required sample size, and the ideal sample size is 394 with the predetermined significance level, power, and effect size. The precise rationale underlying our choices are detailed below:

- **Significance level:** We adopted the usual Significance level in the statistical test, which is often set to 0.05 (5%).
- **Power:** We desired a high statistical power.Power refers.the probability of accepting the alternative hypothesis if it's true. A high statistical power suggests a small risk of committing TypeII errors. (e.g. False Negatives). We used a power of 0.8, since it is considered to be a reasonable standard in many fields of research. It means that there is an 80% chance of detecting a true effect if it is present.
- **Effect size:** We are unsure whether we could detect any differences between any two groups in our experiment given limited prior information. Therefore, we

decided to choose a small effect size to ensure adequate sample sizes are adopted to detect even small effects, which is 0.2

Effect Size	Power	Significance Level	Required Sample Size	Full Dataset Size	ANCOVA Dataset Size
0.2	0.8	0.05	394	64190	Minorities 36972 White 27227

Table 9: Required Sample size with predetermined significance level

Based on the table presented, we can determine that a sample size of 394 is required for our dataset, which would provide adequate statistical power to detect the effect size and achieve the desired level of significance. Our dataset for each group that are needed way exceeded this number. Therefore, we can confidently state that our dataset meets the required sample size despite the number of predictors we have for the analyses conducted and the results reported above.

2. **Retrospective power analysis:**
   We have also conducted retrospective power analysis, and the outcome is included in the Results and Findings section. We calculated the effect sizes for both proposed models. We adopted Cohen's d for the ANCOVA model, and odds ratio for the Logistic Regression.

Results and Findings

RQ1: A one-way ANOVA was performed to compare the effect of race on total adverse actions at arrest. It revealed that there was a statistically significant difference in the mean of total adverse actions at arrest at least two groups (F(3, 59609) = [52.26], p = 9.9925e-34). Tukey's HSD Test for multiple comparisons found that the mean value of total adverse actions at arrest was significantly different between Asian and Black(p = 0.001, 95% C.I. = [0.0972, 0.1476]),

Asian and other races(p = 0.001, 95% C.I. = [0.0526, 0.1139]), Asian and White(p = 0.001, 95% C.I. = [0.0658, 0.1132]), Black and other races(p = 0.001, 95% C.I. = [-0.0656, -0.0126]), black and white(p = 0.001, 95% C.I. = [-0.0510, -0.0148]). There was no statistically significant difference in the mean value of total adverse actions at arrest between other races and White (p=0.9).

	sum_sq	df	F	PR(>F)
C(Perceived_ Race)	82.378001	3.0	52.263776	9.992512e-34
Residual	31318.507041	59609.0	NaN	NaN

Table 10: One-way ANOVA, race and total adverse actions at arrest result

Multiple Comparison of Means - Tukey HSD, FWER = 0.05						
Group 1	Group 2	meandiff	p-adj	lower	uper	reject
Asian	Black	0.1224	0.001	0.0972	0.1476	Ture
Asian	Other	0.0833	0.001	0.0526	0.1139	Ture
Asian	White	0.0895	0.001	0.0658	0.1132	Ture
Black	Other	-0.0391	0.001	-0.0656	-0.0126	Ture
Black	White	-0.0329	0.001	-0.051	-0.0148	Ture
Other	White	0.0062	0.9	-0.0189	0.0313	False

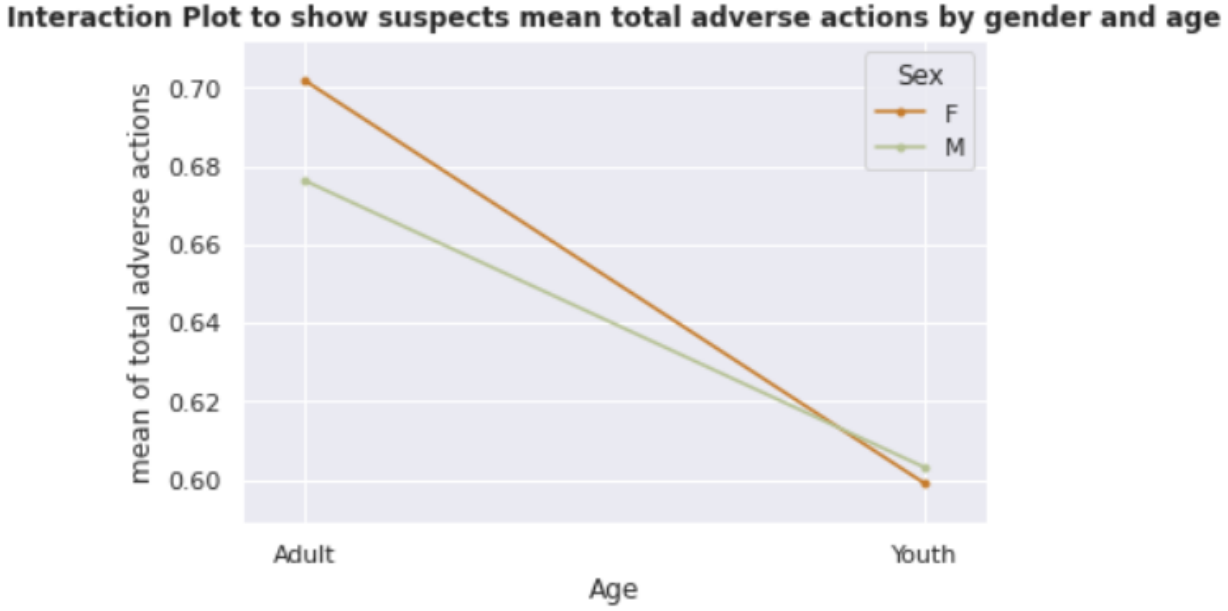
Table 11: Tukey HSD for one-way ANOVA

RQ2: A two-way ANOVA was performed to analyze the effect of gender and age on total adverse actions at arrest. It revealed that there was not a statistically significant interaction between the effects of gender and age (F(1, 64602) = 0.88, p = 0.348). Therefore, we fail to reject the null hypothesis that the mean total adverse actions at arrest are equal among gender and age combinations, and there is no need to perform the Tukey test. Additionally, simple main effects analysis showed that gender has a statistically significant effect on total adverse actions at arrest. (F(1, 64602) = 11.0602, p = 8.825e^-04). Another Simple main effects analysis showed

that age has a statistically significant effect on total adverse actions at arrest. (F(1, 64602) = 35.4814, p = 2.588e^-09). While the interaction plot does not provide any information on the statistically significant differences, the plot showed that the adult suspects have a higher mean of total adverse actions than the youth suspects. Furthermore, the graph also showed that the female adult suspects have a higher mean of total adverse actions than the male adult suspects.

	sum_sq	df	F	PR(>F)
C(Sex)	5.749983	1.0	11.060177	8.825084e-04
C(Youth_at_arrest_under_18_years)	18.446134	1.0	35.481411	2.588235e-09
C(sex): C(Youth_at_arrest_under_18_years)	0.457811	1.0	0.880606	3.480392e-01
Residual	33585.393357	64602.0	NaN	NaN

Table 12: two-way ANOVA, gender, age and total adverse actions at arrest result



Plot 9: Interaction Plot to show suspects' mean total adverse actions by gender and age

RQ3: The goal of ANCOVA is to detect the difference of total number of adverse actions that in White or Minority groups exhibited while controlling the number of arrests

In order to ensure the validity of model, we checked the model assumption, and the results are presented below:

ANCOVA Model Assumptions

1. Independence of observations: Observations within each group must be independent of each other. This assumption is met in this case since duplicated suspects were removed, assuming each suspect can only have one perceived race, either white or minority.
2. Homogeneity of regression slopes: ANCOVA assumes that the relationship between the covariate(s) and the dependent variable is the same for each group. The regression slopes should be similar across all groups. This assumption is met in this case as the interaction plot between the covariate and the dependent variable for each group shows that the lines are approximately parallel to each other, indicating that the regression slopes are similar across all groups.
3. Linearity and constant variance of residuals: The variance of the dependent variable should be equal across all groups, and the relationship between the covariate and dependent variable should be linear. This assumption is not met based on the residual vs fitted plot.
4. Normality: ANCOVA assumes that the dependent variable is normally distributed within each group, and a normal distribution of residuals is also assumed. The assumption of normality of the dependent variable is met through checking the Shapiro-Wilk test, but the normality of residuals is not met based on the QQ-plots.
5. Homogeneity of variance: The variance of the dependent variable should be equal across all groups. This assumption is met based on the Levene's test, which shows that the null hypothesis of equal variances is rejected, indicating that the assumption of equal variance is met.

ANCOVA Model Results and Analysis

The overall model is statistically significant as indicated by the F-statistics 461.1 and the p-value(1.45e-199). However, this model only explains a minimal portion(1.4%) of the dependent variable(total number of adverse actions).

More specifically, from this table of summary statistics, we obtained the following equations:

$$\hat{Y}\ (\text{Race}\ ,\ \text{Number of Arrests}\ ) = 0.6142 - 0.0042\ \text{Race} + 0.0161\ \text{NumArrests}$$

The intercept (0.6142) represents the estimated mean value of the dependent variable (total adverse actions) when the independent variable (perceived race) and the covariate (number of arrests) are both equal to zero.

The coefficient for the perceived race variable (white) is not statistically significant, as indicated by the high p-value (0.462). This suggests that there is no significant difference in the mean total adverse actions score between white individuals and non-white(minority)individuals, after controlling for the number of arrests. In addition, this intercept (- 0. 0.0042) suggests that compared with minority individuals, white individuals have slightly lower average total adverse action. This could also be showed in the following two equations:

$$\hat{Y} \text{ White: } Y(\text{Race} = 1, \text{NumArrests}) = 0.61\ \text{Race} + 0.0161\ \text{NumArrests}$$

$$\hat{Y} \text{ Minority: } Y(\text{Race} = 0, \text{NumArrests}) = 0.6142 + 0.0161\ \text{NumArrests}$$

The coefficient for the number of arrests variable (0.0161) is statistically significant, as indicated by the low p-value. This suggests that for each additional arrest, the mean total adverse actions score increases by approximately 0.016, after controlling for perceived race.

OLS Regression Results

Dep. variable:	Total_adverse_actions	R-squared:	0.014
Model:	OLS	Adj. R-squared:	0.014
Method:	Least Squares	F-statistic:	452.7
Date:	Fri, 14 Apr 2023	Prob (F-statistic):	6.12e-196
Time:	20:23:29	Log-Likelihood:	-69747.
No. Observations:	64199	AIC:	1.395e+05
Df Residuals:	64199	BIC:	1.395e+05
Df Model:	2		
Covariance Type:	nonrobust		

	Coef	Std err	z	P > |z|	[0.025	0.975]
Intercept	0.6142	0.004	145.848	0.000	0.606	0.622
perceived_Race[T.White]	-0.0042	0.006	-0.736	0.462	-0.015	0.007
Number_of_arrest	0.0161	0.001	30.034	0.000	0.015	0.017

Omnibus:	13877.280	Durbin-Watson:	1.902
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33365.455
Skew:	1.207	Prob(JB):	0.00
Kurtosis:	5.577	Cond. No.	15.1

Table 13: Statistical results for ANCOVA without interaction terms

ANCOVA With Interaction

In addition, we decided to test whether there is a difference in the total adverse actions between perceived races when the number of arrests is different, therefore, a model with interaction is performed.

We observed that the overall model does not improve based on adjusted R-squared value. However, the interaction term is statistically significant, indicating that the relationship between perceived race and total adverse actions depends on the number of arrests. More specifically, the coefficient is -0.0042, which suggests that the effect of the number of arrests on the total adverse actions at arrest is different for the white and non-white groups. Specifically, for the white group, an additional arrest is associated with a decrease in the mean value of the total adverse actions at arrest by 0.0042, holding all other variables constant.

OLS Regression Results

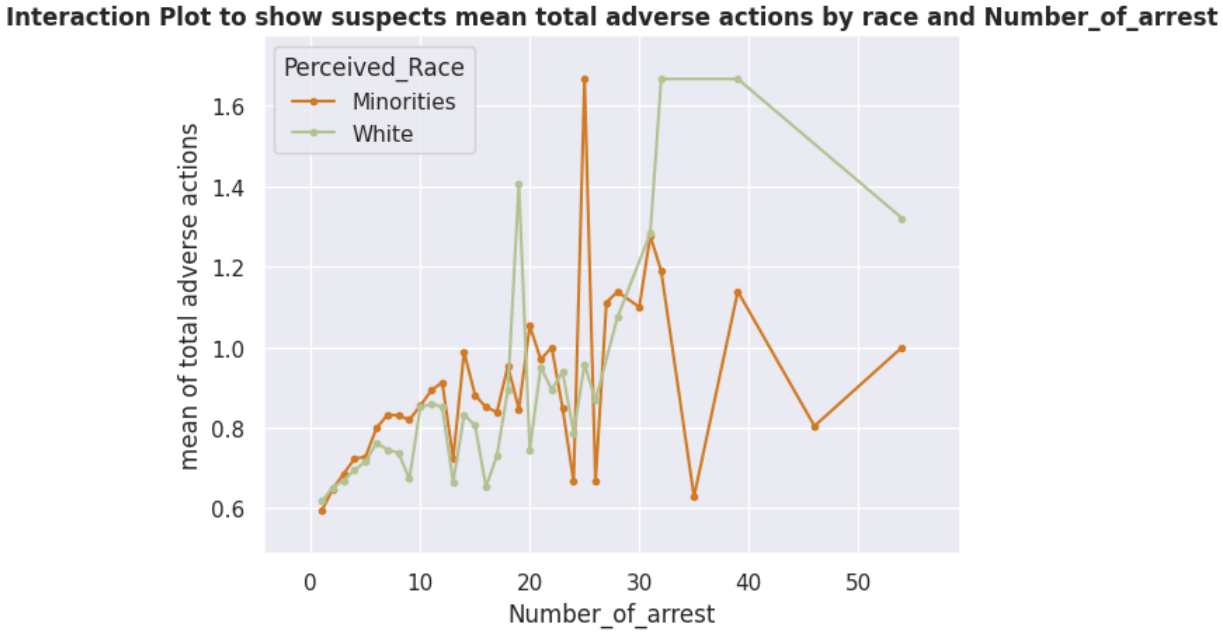
Dep. variable:	Total_adverse_actions	R-squared:	0.014
Model:	OLS	Adj. R-squared:	0.014
Method:	Least Squares	F-statistic:	306.9
Date:	Fri, 14 Apr 2023	Prob (F-statistic):	7.42e-198
Time:	20:24:41	Log-Likelihood:	-69739.
No. Observations:	64199	AIC:	1.395e+05
Df Residuals:	64195	BIC:	1.395e+05
Df Model:	3		
Covariance Type:	nonrobust		

	Coef	Std err	t	P > |t|	[0.025	0.975]
Intercept	0.6070	0.005	132.009	0.000	0.598	0.616
perceived_Race[T.White]	0.0130	0.007	1.797	0.072	-0.001	0.027
Number_of_arrest	0.0181	0.001	24.512	0.000	0.017	0.020
Perceived_Race[T.White]: Number_of_arrest	-0.0042	0.001	-3.899	0.000	-0.006	-0.002

Omnibus:	13853.288	Durbin-Watson:	1.903
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33273.383
Skew:	1.206	Prob(JB):	0.00
Kurtosis:	5.573	Cond. No.	21.6

Table 14: Statistical results for ANCOVA with interaction terms

Furthermore, we noticed that this phenomenon does not hold when the suspects are arrested many times (more than 30 times). Now, the white individuals would have a higher mean of total adverse actions than minority individuals, holding all other variables constant.



Plot 10: Suspects mean total adverse actions by race and number of arrest

ANCOVA Effect size

In general, a Cohen's d value of 0.2 is considered a small effect size, 0.5 is considered a medium effect size, and 0.8 is considered a large effect size. Therefore, Cohen's d value of 0.014 suggests that the difference between the white and minority groups is not practically significant and is likely due to chance. This outcome is consistent with the results obtained from our p-value analysis.

RQ4: The goal of our logistic regression is to analyze the relationship between an individual's sex (binary outcome: 0 for male and 1 for female) and a set of independent variables, including age (Youth & Not Youth), race (White & Minorities), number of total adverse actions and number of arrests. We aim to determine the influence of these factors on the likelihood of a suspect being male or female, and assess the strength and direction of these associations.

In order to ensure the validity of model, we checked the model assumption, and the results are presented below:

Logistic Regression Assumption Check:

1. Binary outcome: Logistic regression assumes that the outcome variable is binary. In this case, the outcome variable is sex (Female or Male), which meets this assumption.
2. Independence of observations: Observations within each group must be independent of each other. Duplicated observation rows were removed, which meets this assumption.
3. Linearity of the logit: Logistic regression assumes that the relationship between the logit of the outcome variable and all continuous variables is linear. The scatterplot of log-odds and total adverse action shows a clear non-linear pattern, which violates this assumption. However, the scatterplot of log-odds and number of arrests presents a linear relationship, which satisfies this assumption.
4. No multicollinearity: Logistic regression assumes that there is no high correlation between predictor variables. This assumption is met by computing the variance inflation factor (VIF) for each independent variable. All VIF values are below 1.5, indicating low correlation between predictor variables.
5. Large sample size: Logistic regression assumes a large sample size. In this dataset, we have a sufficiently large sample size according to the general rule of thumb.
6. No strongly influential outliers: Logistic regression assumes that there are no highly influential outlier data points. Cook's distance and boxplots were used to determine the influence of data points, and it was found that the variable of number of arrests has potential influential outliers, while the variable of total adverse action does not.

Logistic Regression Results and Analysis:

Firstly, we generated a logistic regression model. However, after getting the confusion matrix and calculating the F1 score(test data) of it, it appears that the model is not predicting any true positive cases (class 1: females) at all, which indicates that the model may be biased toward class 0 (males) and not performing well in identifying class 1 (females). This is due to an imbalanced dataset, there is 52271 male data but only 12535 female data. Therefore, we used Synthetic

Minority Over-sampling Technique (SMOTE) to resample it and generated another logistic regression model.

Optimization terminated successfully.
Current function value: 0.688559
Iterations 4

Logit Regression Results

Dep.variable	Sex	No. Observation:	83000
Model	Logit	Df Residuals:	82995
Method	MLE	Df Model:	4
Date:	Fri, 14 Apr 2023	Pseudo R-squ. :	0.006619
Time:	19:44:18	Log-Likelihood:	-57150.
Converged:	Ture	LL-Null:	-57531.
Covariance Type:	nonrobust	LLR p-value:	1.608e-163

	Coef	Std err	z	P > |z|	[0.025	0.975]
Intercept	-0.1539	0.012	-12.401	0.000	-0.178	-0.130
Youth_at_arrest_ under_ 18_years_youth	0.3216	0.032	10.105	0.000	0.259	0.384
Perceived_Race_White	0.3530	0.014	25.001	0.000	0.325	0.381
Total_adverse_actions	0.0390	0.010	3.995	0.000	0.020	0.058
Number_of_arrest	-0.0118	0.001	-8.581	0.000	-0.014	-0.009

Table 15: Logistic regression results

Table 16. The odds ratio of the Logistic Regression Coefficient

Intercept	0.857371
Youth_at_arrest_ under_ 18_years_youth	1.379377
Perceived_Race_White	1.423288
Total_adverse_actions	1.039788
Number_of_arrest	0.988276

Dtupe: float 64

Table 16: The odds ratio of the Logistic Regression Coefficient

All the features are statistically significant, since the p-value of each independent variable is lower than 0.05. For perceived_Race (odds ratio = 1.423288, 95%CI : [1.38445, 1.463221]): Since the odds ratio is greater than 1, this indicates that white have 1.42 times higher odds of being female suspects compared to minorities, all other variables being held constant. Youth_at_arrest__under_18_years (odds ratio = 1.379377, 95%CI: [1.295959, 1.468164]): With an odds ratio greater than 1, this means that youth (aged 17 and younger) have 1.38 times higher odds of being female suspects compared to those who are not youth, while holding other variables constant. Increasing number of total adverse actions were associated with an increased likelihood of a suspect being a female. The odds ratio is approximately 1.039788, the 95%CI of it is between 1.020074 and 1.059883. For each additional number of total adverse actions increase for the suspect, the odds that the suspect is a female increase by about 1.04 times. What's more, the increasing number of arrests was associated with a decreased likelihood of a suspect being a female. The odds ratio is approximately 0.988276, the 95%CI of it is between 0985617 and 0.990942. For each additional number of arrests increase for the suspect, the odds that the suspect is a female decrease by (1-0.988276) 0.011724 times.

	Lower CI	Upper CI	OR
Intercept	0.836769	0.878479	0.857371
Youth_at_arrest_ under_18_years_youth	1.295959	1.468164	1.379377
Perceived_Race_White	1.384445	1.463221	1.423288
Total_adverse_actions	1.020074	1.059883	1.039788
Number_of_arrest	0.985617	0.990942	0.988276

Table 17: Lower CI, Upper CI OR in logistic regression

The first confusion matrix is calculated on the training dataset, which is the data used to train the logistic regression model. This confusion matrix tells us how well the model has learned the patterns in the training data and can help us diagnose potential issues with overfitting or underfitting. The detailed values are stored in the following table:

Confusion Matrix: [[23960 17540] [20433 21067]]

Test accuracy is 0.5424939759036145
The F1 score is 0.525971513101227

Table18. Confusion Matrix for the training dataset

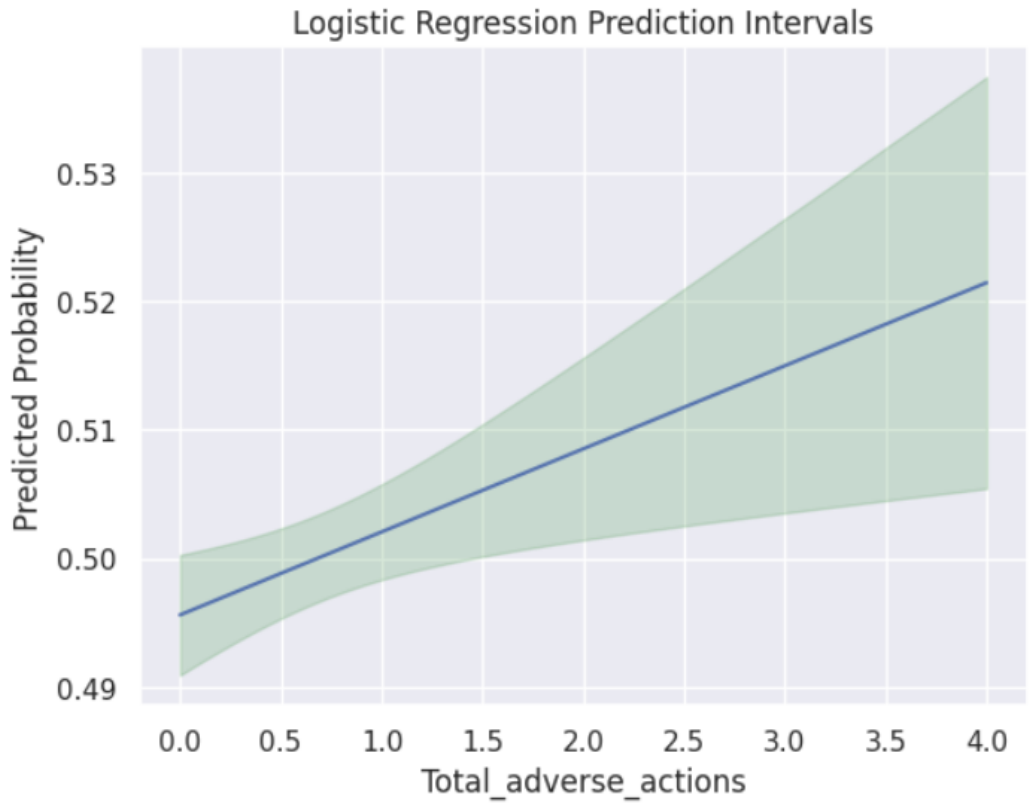
The second confusion matrix is calculated on the testing dataset, which is data that the model has never seen before. This confusion matrix tells us that our model lacks the ability to generalize well to new, unseen data (F1 scores dropped from 0.53 to 0.31). More specifically, on this new testing dataset, we see True Negatives (TN): 5724 - These are the instances where the model correctly predicted the person to be male (0). False Positives (FP): 4542 - These are the instances where the model incorrectly predicted the person to be female (1), while they were actually male (0). False Negatives (FN): 1278 - These are the instances where the model incorrectly predicted the person to be male (0), while they were actually female (1). True Positives (TP): 1294 - These are the instances where the model correctly predicted the person to be female (1). Based on these results, we calculated the F1 score to evaluate the performance of the model, since it is a balanced measure of accuracy giving equal importance to recall and precision. The F1 score is 0.3078, which is really low.

Confusion Matrix:
[[5724 4542]
[1278 1294]]

Test accuracy is 0.5466583579996884
The F1 score is 0.307802093244529

Table 19. Confusion Matrix for the testing dataset

The prediction interval plot is the showcase of possible prediction intervals, it represents the predicted probability of an outcome based on the logistic regression analysis with the independent variable 'Total_adverse_actions' and the dependent variable (DV) 'Sex'. The blue line in the plot is the predicted probability curve, which shows the relationship between 'Total_adverse_actions' and the predicted probability of the suspect is female occurring. The green shaded area represents the 95% prediction intervals for the logistic regression model. These intervals provide a range within which the probability of a female suspect occurring is likely to fall 95% of the time, given the value of 'Total_adverse_actions'. The prediction intervals provide an idea of the uncertainty associated with the model's predictions.



Plot 11: Logistic Regression Prediction Intervals

Discussion

RQ1 explored the effect of race on the total adverse actions at arrest. The one-way ANOVA test showed that there was a statistically significant difference in the mean of total adverse actions at arrest between at least two racial groups. The Tukey HSD test was conducted to perform multiple comparisons, and it revealed that the mean value of total adverse actions at arrest was significantly different between Asian and Black, Asian and other races, Asian and White, Black and other races, and Black and White. However, if we group all other non-white racial groups together and name them as the minority group.We find out that there was no statistically significant difference between the minority group(all other non-white racial groups) and White. These findings suggest that overall adverse behavior at the time of arrest is racially different, despite whether they are in minority groups or white The fact that some racial groups had higher averages for overall bad behavior at the time of arrest suggests that they were disproportionately affected by police actions during the arrest process. This may be due to a variety of reasons, including conscious or unconscious bias, prejudice or racial profiling. It is important to

acknowledge and address this issue to ensure fair and equitable treatment for all, regardless of race.

The two-way ANOVA examines the impact of age and gender on total adverse actions at arrest. The results suggest that there is no interaction between age and gender, indicating that the effects of gender and age are independent. However, gender and age individually have a statistically significant effect on the total adverse actions at arrest. Our graphs show that adult suspects have a higher mean of total adverse actions than youth suspects, which may be due to the perception that adult suspects pose a greater threat. Moreover, our graphs show that female adult suspects have a higher mean of total adverse actions than male adult suspects, indicating that gender bias may also exist in law enforcement. Likewise, these findings highlight the existence of implicit biases in law enforcement that may affect the treatment of individuals based on race, age, and gender. These biases can lead to discrepancies in the total number of adverse actions taken by law enforcement, with serious consequences for suspects.

Thirdly, power analysis is a statistical tool used to determine the appropriate sample size needed to detect a statistically significant effect in experimental research. In this report, the two types of power analysis are a priori and retrospective power analysis. The choice of significance level and effect size are important in power analysis, and a small effect size was chosen in the given experiment to ensure even small effects could be detected. The required sample size was determined to be 394, which was considered adequate to detect the effect size and achieve the desired level of significance. But the shortcoming of this power analysis is that the power analysis should be performed before the experiment, but we are given the dataset.

Fourth, in this study, ANCOVA was used to examine the differences in adverse actions taken against suspects of different races. The results showed no significant difference in the number of adverse actions when controlling for the covariate. However, other factors, such as socioeconomic status, education level, and prior criminal history, may contribute to the observed differences between racial groups. Future research could explore the influence of these potential predictors to promote equity in the criminal justice system.

Lastly, logistic regression was used in this study to predict the likelihood of a suspect being female based on several independent variables. Results showed that all independent variables were statistically significant with p-values less than 0.05. Perceived race and youth at arrest (under 18 years) were found to be significant predictors of a suspect being female. Additionally, each additional adverse action increased the odds of a female suspect by approximately 1.04 times, while each additional arrest decreased the odds of a female suspect by 0.011724 times. These findings have important implications for promoting equity and reducing gender disparities

in the criminal justice system. The model has a low F1 score and several factors could be contributing to this. Insufficient or irrelevant features, the model's simplicity, SMOTE limitations, and data quality issues could all be affecting the model's performance. To improve the model's accuracy, more relevant features could be added, advanced models could be used, SMOTE limitations could be considered, and data preprocessing and cleaning could be performed.

Conclusion

In this report, We used a dataset from the Toronto Police Service to explore the link between arrestee age, gender, race, number of arrests and arrestee behavior at the time of arrest. Through our analysis, we found that the adverse actions by arrestees at the time of arrest varied by race, gender, and age. First, the results of the analysis showed that female suspects were more likely to engage in negative arrest actions during arrest than male suspects. Second, younger suspects tend to take fewer negative adverse actions at the time of arrest than adult suspects. Third, our results show that blacks tend to be arrested at a higher rate than whites and Asians, and they also tend to engage in more negative adverse behavior during their arrests. Forth, the logistic regression analysis in this study revealed significant predictors of a suspect being female, including perceived race, youth at arrest, adverse actions, and arrests. These findings highlight the need for promoting equity and reducing gender disparities in the criminal justice system. Additionally, the ANCOVA analysis showed no significant difference in adverse actions between racial groups when controlling for the covariate. But it's not accurate to generalize an individual's behavior at the time of arrest simply by race, age or gender. Because, an individual's behavior during arrest may also be influenced by other factors, including their personal history, their mental health, and the circumstances of their arrest. Or the behavior of individuals may also be influenced by their living environment and cultural background, including poverty, discrimination, etc. Therefore, while studying such problems, these problems should also be considered in the test, or some variables should be controlled to improve the accuracy of the results. Therefore, we should abandon the stereotypes of some social groups, treat them with tolerance, and strive to create a fair and peaceful society.

Reference

Tim Newburn. 2018. 3. who commits crime? *Criminology: A Very Short Introduction* (2018), 24–38. DOI:http://dx.doi.org/10.1093/actrade/9780199643257.003.0003

S.J. D'Alessio and L. Stolzenberg. 2003. Race and the probability of arrest. *Social Forces* 81, 4 (2003), 1381–1397. DOI:http://dx.doi.org/10.1353/sof.2003.0051

Alex R. Piquero and Robert W. Brame. 2008. Assessing the race–crime and ethnicity–crime relationship in a sample of serious adolescent delinquents. *Crime & Delinquency* 54, 3 (2008), 390–422. DOI:http://dx.doi.org/10.1177/0011128707307219

Anon. Arrests and strip searches (RBDC-arr-TBL-001). Retrieved February 21, 2023 from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Published by Statista Research Department and Feb 2. 2023. People shot to death by U.S. police, by race 2023. (February 2023). Retrieved February 28, 2023 from https://www.statista.com/statistics/585152/people-shot-to-death-by-us-police-by-race/

Statistics Canada. (2022). Police-reported crime statistics in Canada, 2020. Retrieved from https://www150.statcan.gc.ca/n1/pub/85-002-x/2022001/article/00013-eng.htm