University of Toronto

INF 2178 LEC0101: EXPERIMENTAL DESIGN FOR DATA SCIENCE

End-term Project

Final Version

Winter 2023

Instructor: Prof. Shion Guha

Group Number: Group 61

Group Members:
Qiqi He
Yuqing Wu

April 16, 2023

Table of Contents

List of figures

List of tables

Abstract

This paper explores the relationship between “Cooperative”, “Booked”, “Race”, “Age” and “Sex” using the “Arrests and Strip Searches (RBDC-ARR-TBL-001)” (Arrests and Strip Searches (RBDC-ARR-TBL-001), 2022) dataset from the Toronto Police Service. Our paper provides a detailed description of the dataset, including information on the variables, data limitations or bias, or any other caveats that should be taken into account when analyzing and reporting the data. It provides exploratory data analysis on this dataset, including information on the number of arrests, number of booking, demographic information on the individuals involved, and the relationship between variables. The paper aims to find out if there is a relationship between race, sex and the booked event. It concludes by summarizing the main findings, stating the limitations, and suggesting avenues for future research.

Introduction

The arrest and search of individuals are common practices implemented by police departments worldwide to ensure public safety by identifying and apprehending individuals suspected of criminal activity. However, the implementation of these procedures has been widely criticized for their potential to create inequalities and discrimination in police practices. This study aims to investigate the relationship between an individual's demographic characteristics and the likelihood of being booked by the Toronto Police Service.

Literature Review

Police searches have been the subject of much research, with a focus on the impact of race, age, and gender on police practices. A growing body of literature has highlighted the overrepresentation of certain groups in police arrest and search statistics, suggesting that police practices may be influenced by implicit biases and stereotypes.

Research on the impact of race on police practices has been extensive. A study by Jain (2021) found that African Americans were more likely to be arrested and subjected to searches than any other racial group. The study also found that the race of the officer did not significantly affect the likelihood of arrest, indicating that implicit biases may be at play in police practices.

While the research on police practices and race has primarily focused on the overrepresentation of African Americans in arrest and search statistics, there is also evidence of disparities affecting other racial groups. One study by Mitchell & MacKenzie, 2004) found that Native Americans were more likely to be arrested and searched than whites or African Americans, even after controlling for factors such as offence type and criminal history. The study suggests that Native Americans may be subject to bias and discrimination in the criminal justice system, which could be attributed to historical and ongoing patterns of discrimination.

Research has also shown disparities affecting Hispanic and Latino individuals. A study by Ghandnoosh (2015) found that Hispanics were more likely to be stopped and searched by the police than whites or African Americans, and was also more likely to experience the use of force during encounters with police. The study suggests that the high rates of immigration enforcement in Hispanic communities may contribute to the disparities observed.

Asian Americans have also been subject to disparities in police practices. A study by Lee, Yoo, and Park (2019) found that Asian Americans were less likely to be arrested and searched than whites or African Americans, but were more likely to experience non-lethal use of force during encounters with police. The study suggests that Asian Americans may be subject to stereotypes that portray them as non-threatening or passive, which may lead police to use force to assert their authority.

Age is another factor that has been shown to influence police practices. A study by Brunson (2007) found that young men were more likely to be arrested and searched than older men, indicating that age-based stereotypes may be at play in police practices. Similarly, a study by Piquero, Piquero, and MacDonald (2003) found that young adults were more likely to be arrested than juveniles or older adults, suggesting that police may be targeting this age group for surveillance and enforcement.

Gender is another factor that has been shown to influence police practices. A study by Miller and Schwartz (2018) found that women were less likely to be arrested and searched than men. The study also found that the gender of the officer did not significantly affect the likelihood of arrest, indicating that implicit biases may be at play in police practices.

In addition to individual-level factors, institutional and contextual factors have been shown to influence police practices. A study by Weitzer and Tuch (2006) found that the organizational culture of police departments influenced arrest and search practices, with more aggressive departments having higher rates of arrests and searches. Another study by Goff, Lloyd, Geller, Raphael, and Glaser (2016) found that racial bias was more likely present in departments with more negative public perceptions and lower levels of community trust.

Overall, the literature suggests that police practices are influenced by a complex interplay of individual-level, institutional, and contextual factors. The analysis of the Arrests and Strip Searches dataset will provide important insights into the extent of disparities in police practices and help identify potential areas for intervention and reform.

Dataset Description

The data set used in this project is from the website of the Toronto Police Department, which provides information about all arrests and strip searches every three months from 2020 to 2021.

You can find it through the following link: https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about. The data set collected 65276 lines of crime data, with a total of 25 attributes, including object ID, the year of arrest, the month of arrest, the event ID, the arrest ID, the person ID, and the demographic information of the arrested (such as race, sex, and age group). The place of arrest, the type of crime, whether to register at the police station within 24 hours and the strip search is also included. At the same time, with regard to the action and search reasons at the time of arrest, the data set has some decision variables, such as whether to find Concealed items, mental ability or possibly feasible, AssistEscape, PossessWeapons, etc. However, there are many null values for the four attributes related to the search reason. ObjectID is unique for each line of the record. Among them, 12 are numeric variables and 12 are text variables. Numeric variables include binary values and numeric integer values.

Research Questions

This study seeks to explore one research question, which is: ***How will the frequency of actions at arrest cooperatives affect the number of booked, considering controlling factors for people's perceived race, sex and age groups? And if there any other external factors that will affect people's attitudes towards cooperation?*** The research question aims to investigate how "Booked" will affect "Cooperative" based on controlling "Race", "Sex" and "Age". The study intends to determine whether the frequency of actions at arrest that are cooperative will have an impact on the likelihood of being booked while controlling for other factors such as race, sex, and age groups. Prior research has suggested that gender biases may play a role in police practices, with men being more likely to be arrested than women. However, studies have also found that this relationship may vary depending on the context, such as the nature of the crime and the race of the individual involved. Therefore, this study aims to examine this relationship in the context of the Toronto Police Service and contribute to the broader literature on the topic.

Significance of Study

This study's findings have important implications for understanding potential biases and discrimination in police practices in Toronto, Canada. The research question aims to investigate how the frequency of cooperative actions at arrest will affect the number of booked individuals while controlling for perceived race, sex, and age. By examining the relationship between an individual's demographic characteristics and the likelihood of being booked, this study aims to contribute to the broader literature on policing practices and help inform policy decisions.

Outline of the Paper

The paper is structured as follows: The introduction provides a comprehensive overview of the existing literature on police practices and the impact of race, age, and gender on policing and the basic information of the dataset. Then the data cleaning gives a general explanation of how to

choose the appropriate dataset. The exploratory data analysis section presents an overview of the dataset used in this study and presents key descriptive statistics. The methods section provides a detailed description of the statistical methods used in this study, including power analysis, ANCOVA tests and logistic regression. The results and findings section presents the results of the analysis, including the relationships between the variables of interest. The discussion section provides a critical evaluation of the results and discusses the implications for policy and future research. The paper concludes with a summary of the key findings, limitations and recommendations for future research.

Data Cleaning

Considering that the dataset is huge and has 24 attributes, for the convenience of subsequent analysis, we need to preprocess the dataset. We first use the following histogram to show the counting of null and not null values for all 24 attributes in the dataset (Figure 1). By looking at Figure 1, blue represents a null value while orange represents not null value, we can see that there are 5 attributes 'SearchReason_CauseInjury', 'SearchReason_AssistEscape', 'SearchReason_PossessWeapons', 'SearchReason_PossessEvidence' and 'ItemsFound' that have more than 50000 null values, for the accuracy of the follow-up analysis results, these four columns of data will not be used in the later analysis. For a small amount of null value data of other attributes, we choose to drop the row corresponding to the null value.

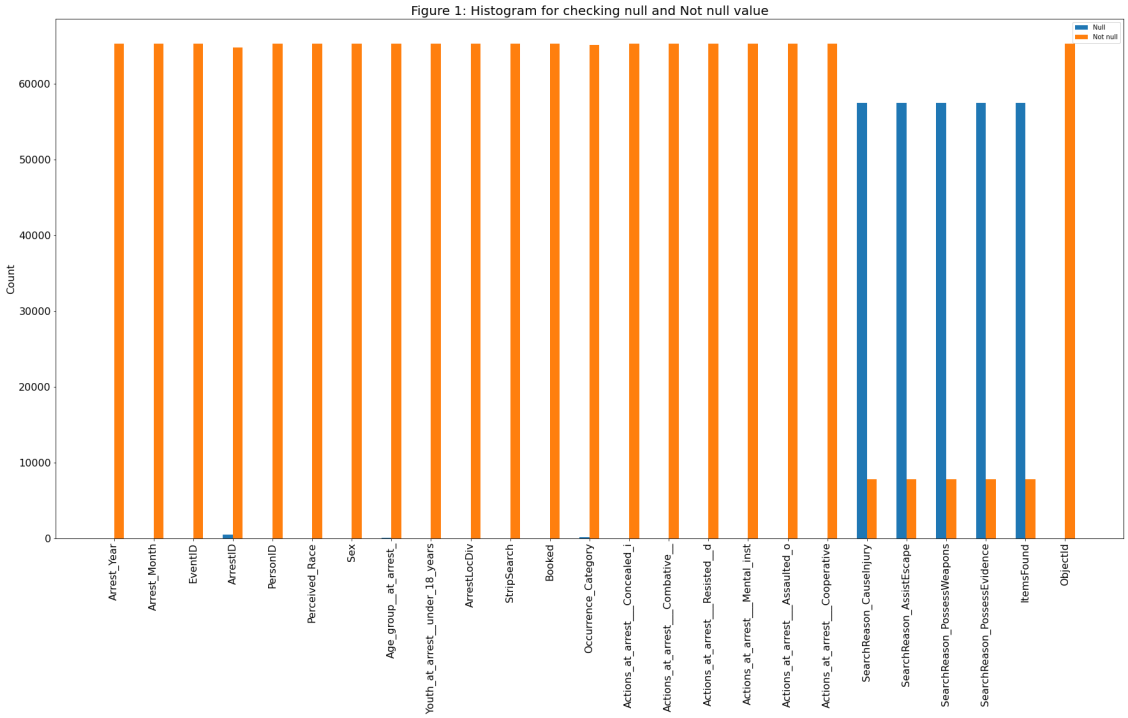

Figure 1: Histogram for checking null and not null values

After dropping columns with too many null values, considering that we want to discuss sex, age group and race for further analysis, the following plots show the distribution of these on the

original dataset. Through Figure 2 and Figure 3, we can conclude that in this dataset, men account for the majority, while the union sex accounts for only 9, which is too small. Therefore, we removed union sex in the subsequent analysis. Not only that but also the bias of the data itself may cause certain limitations to our research.

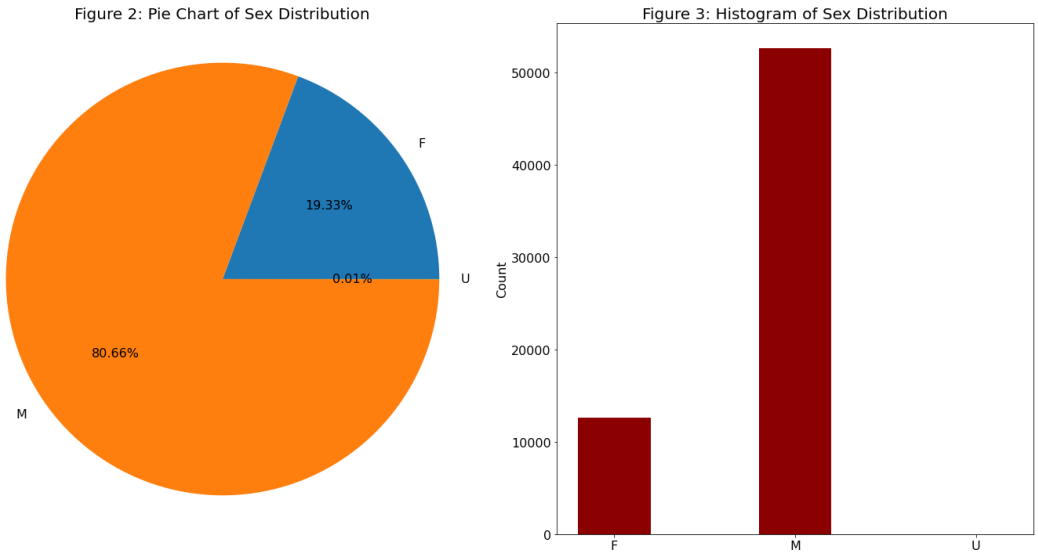

Figure 2 & Figure 3: Pie Chart and Histogram of sex distribution

Similarly, for the age group, we use the histogram to show its distribution. We can find that there are a total of 9 groups in Figure 4, but when we see the x-axis carefully, there exists overlapping in the categories. For example, “Aged 17 years and under” and “Aged 17 years and younger” are the same meaning, thus we combined these two groups into “Aged 17 years and under”. Similarly “Aged 65 and older” and “Aged 65 years and older” are merged into “Aged 65 years and older”. Meanwhile, we can find that among the criminals, those aged 25 to 34 are the most, followed by those aged 35 to 44.

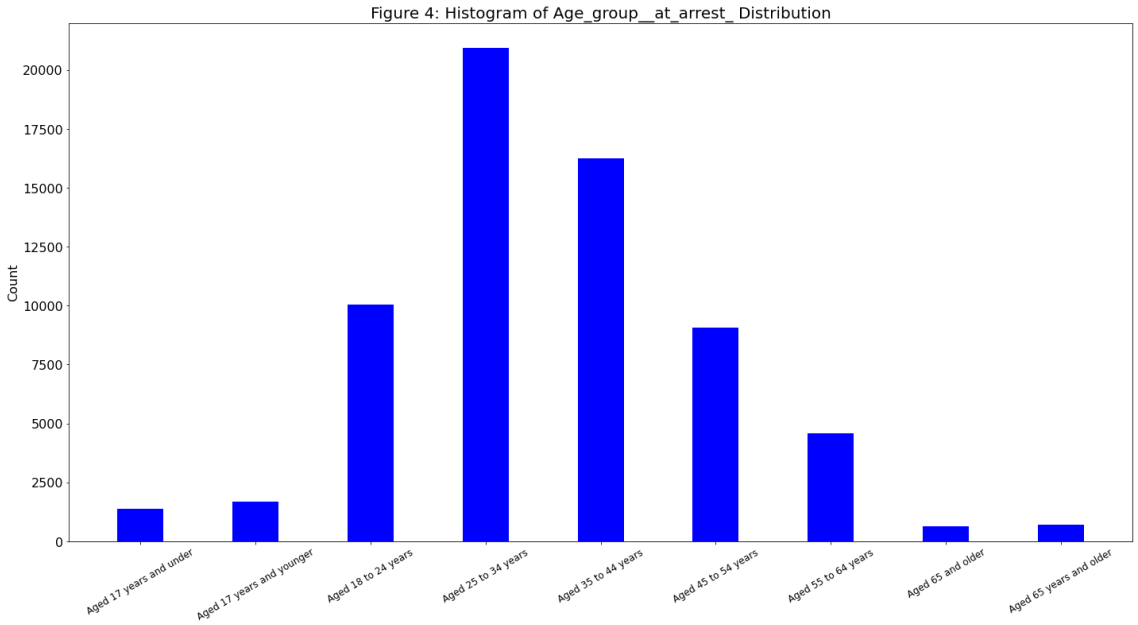

Figure 4: Histogram of Age_group_at_arrest_ distribution

From Figure 5, it can be seen that there are eight racial types, of which white people account for the largest number, and white and black people account for the majority of the race. This does not mean that the data has bias because the most commonly reported races in Toronto were White (50.2%), East Asian (12.7%), South Asian (12.3%), and Black (8.5%) according to the racial demographic data of Toronto in 2021. (Toronto Population 2021/2022, n.d.) Therefore, we think that the data on perceived race is logical and not very biased in analysis.

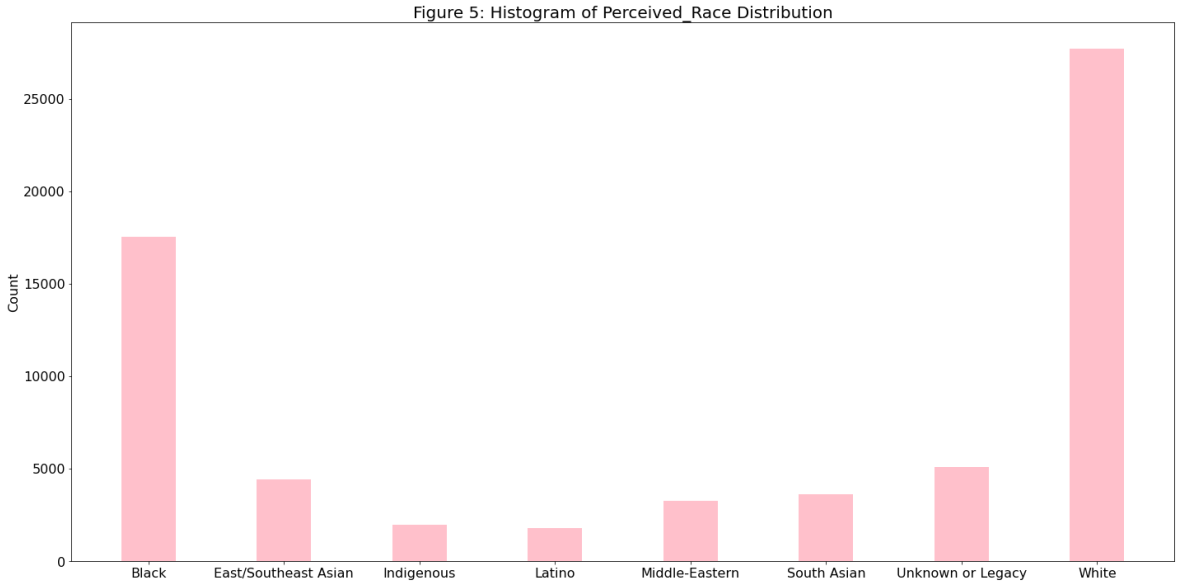

Figure 5: Histogram of Perceived_Race distribution

Exploratory Data Analysis

Descriptive Statistics

Based on our general thought and research question, we need to have a general understanding of age group, sex, race and booked count based on the cooperative count. Firstly, we want to learn about whether the cooperative count will associate with the booked count, age group and race, thus we provide a correlation heatmap with the cooperative count (Figure 6). Through the heatmap, it is clear that the booked count and race are the two variables with the highest correlation with the booked count. And sex is also highly related to the cooperative count. However, the age group may have no such strong relationship with the cooperative count, which is only about 5%. But this can't mean that age group has no relationship with the effect of booked count and cooperative count. As a result, we believe that it is reasonable for us to study whether the theme of cooperative count is different due to the booked count, age groups, sex and race.

Figure 6: Cooperative Count Correlation Heatmap

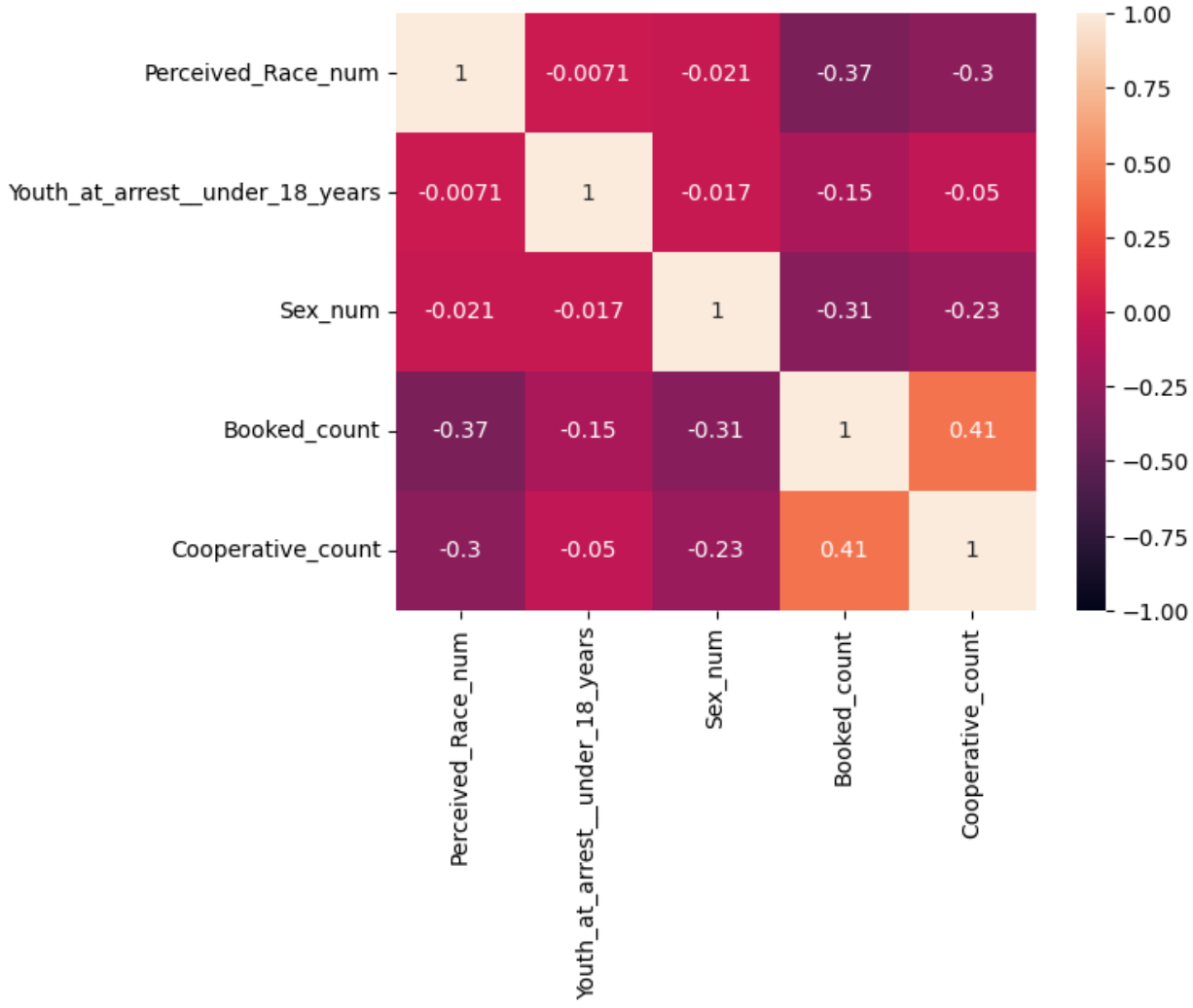

Figure 6: Cooperative Count Correlation Heatmap

The following table shows the mean booked count and cooperative count based on race and sex (Table 1). We can find that the mean cooperative count is different in different race groups and sex groups, the average number of cooperative times of black men and white men is significantly higher than that of other races. Same for the booked count. And the frequency of booking in men is higher than that in women. However, this may be due to the fact that the number of males is more than that of females, and the number of blacks and whites is higher than that of other races in this data set. So, we need to do more statistical analysis to prove this phenomenon.

Table 1: Race and Sex based on Booked Count and Cooperative Count

Perceived_Race	Sex	Booked_count	Cooperative_count
Black	**F**	96.142857	107.500000
	M	594.571429	433.785714
East/Southeast Asian	**F**	16.428571	41.357143
	M	133.142857	132.714286
Indigenous	**F**	26.500000	24.250000
	M	53.928571	33.071429
Latino	**F**	11.500000	12.300000
	M	60.714286	30.142857
Middle-Eastern	**F**	11.181818	43.454545
	M	107.785714	48.857143
South Asian	**F**	15.750000	57.000000
	M	113.428571	71.285714
Unknown or Legacy	**F**	30.833333	49.583333
	M	147.571429	95.071429
White	**F**	200.000000	127.642857
	M	818.285714	409.357143

To better conduct our research question, we also proceed with some t-tests before we use the ANCOVA as the final test. Considering that although we have learned from the above heatmap and statistical table that the cooperative count varies for different races and genders, due to the bias in some data packets, the number of black and white people is much higher than that of other races, and the number of men is higher than that of women. Therefore, we need more accurate statistical methods for detection.

T-tests

Considering that the above heatmap could only reveal the correlations between different attributes, we decided to conduct independent sample t-tests on some of the attributes. The t-test is a statistical test used to determine whether there is a significant difference between the means of two groups or populations. Because t-tests have some assumptions to run before we do the test, we have already checked the following assumptions: (1) Normality: The distribution of the population should be normal or approximately normal. (2) Homogeneity of variance: The variances of the two groups being compared should be equal. (3) Independence: The observations in each group should be independent of each other. This means that there should be no relationship between the observations in each group. Before we conduct ANCOVA tests and power analysis, we first do the t-tests for the dataset on the cooperative count. There is a total of two t-tests we made in this project.

Sex and Cooperative Count

The first one is for sex based on the cooperative count. We compare the mean of cooperative count between males and females and want to know whether they are different. The hypothesis of this test is as follows:

H0: The population mean of Males and Female's cooperative count are equal

Ha: The population mean of Males and Female's cooperative count are not equal

After calculation, the p-value is 0.00021054649011604 83, which is statistically significant at a *p=0.05* level. We can conclude that there is evidence that male and female cooperative count means are unequal, which is the same as the previous heatmap and table conclusions.

Race and Cooperative Count

The second one is for race based on the cooperative count. Considering that this dataset has the most white race data, and the second is black race. Other races' data is too small. Thus, we combine the other race data and the black race data as the "Not-White" data and compare this with the white race data. In conclusion, we compare the mean of cooperative count between the white race and the not-white race and want to know whether there exists a difference. And the hypothesis of this test is as follows:

H0: The population mean of white race and not-white races' cooperative count are equal

Ha: The population mean of white race and not-white races' cooperative count are not equal

After calculation, the p-value is 0.006315337824927366, which is statistically significant at a *p=0.05* level. We can conclude that there is evidence that white race and not-white races' cooperative count are not equal, which is the same as the previous heatmap and table conclusions.

From these two t-tests result, we hope to further explore how these variables affect the cooperative count. And we choose to do the power analysis and ANCOVA as the further methods to help us understand and solve the research question better.

Methods

Power Analysis

Statistical power refers to the probability of accepting the correct alternative hypothesis when the null hypothesis is wrong in the hypothesis test. It is related to type II error, also known as false negative rate. The greater the power, the lower the probability of making a type II error. Therefore, in the study, we had better reduce the type II error as much as possible to improve the accuracy of the study. Normally, a power of 0.8 is preferred and accepted in statistical studies. Based on Seltman's book, there are several ways to improve statistical power: (1) increase sample size, (2) reduce variance, and (3) increase population mean differences. To ensure that the power we use is effective and reasonable, our group choose to calculate the sample size based on the effective size for each independent categorical variable. This could help us optimize sample size to achieve sufficient research objectives while avoiding unnecessary costs and resources associated with collecting larger sample sizes. As for some attributes with more than two levels such as the race, we merged them into two groups. Additionally, a plot of power analysis could help us better study and understand the relationships among sample size, effect size and statistical power, and help improve the presentation, reliability and validity of research findings.

ANCOVA

Analysis of covariance (ANCOVA) is a statistical analysis method that combines linear regression analysis and analysis of variance. It is used to control the confounding factors that affect the dependent variables when comparing the differences between two or more means groups. To conclude the ANCOVA tests successfully, there are some assumptions we need to fulfill, and we have already checked the following assumptions: (1) Normality: The distribution of the population should be normal or approximately normal. (2) Homogeneity of variance: The variances of the two groups being compared should be equal. (3) Independence: The observations in each group should be independent of each other. This means there should be no relationship between dependent variables in each group. (4) Linearity. (5) Homoscedasticity: For dependent variables, it should have a consistent change for all levels of covariation.

Our group conducts three ANCOVA tests in this project. The hypothesis is as follows:

Age Group and Booked Count Based on Cooperative Count

We choose the age group, booked count and cooperative count to do the ANCOVA test to examine the relationship between different people's age groups and the cooperative count based on controlling the booked count as the covariance.

H0: There is no significant relationship between age groups and cooperative count based on controlling the booked count. This means each age group results in the same cooperative count by controlling the booked count.

Ha: There is a significant relationship between age groups and cooperative count based on controlling the booked count. This means each age group results in a different cooperative count by controlling the booked count.

Race and Booked Count Based on Cooperative Count

We choose the original perceived race, booked count and cooperative count to do the ANCOVA test to examine the relationship between different people's race groups and the cooperative count based on controlling the booked count as the covariance.

H0: There is no significant relationship between perceived race groups and cooperative count based on controlling the booked count. This means each race group results in the same cooperative count by controlling the booked count.

Ha: There is a significant relationship between perceived race groups and cooperative count based on controlling the booked count. This means each race group results in a different cooperative count by controlling the booked count.

Sex and Booked Count Based on Cooperative Count

We choose the original perceived race, booked count and cooperative count to do the ANCOVA test to examine the relationship between different people's race groups and the cooperative count based on controlling the booked count as the covariance.

H0: There is no significant relationship between perceived race groups and cooperative count based on controlling the booked count. This means each sex group results in the same cooperative count by controlling the booked count.

Ha: There is a significant relationship between perceived race groups and cooperative count based on controlling the booked count. This means each sex group results in a different cooperative count by controlling the booked count.

Logistic Regression

Logistic Regression is a generalized linear model that is a machine learning method used to solve binary classification (0 or 1) problems and estimate the likelihood of something. For example, the possibility of a user purchasing a certain product, the possibility of a patient suffering from a certain disease, and the possibility of an advertisement being clicked on by the user. As for our group, the detailed research question for the logistic regression will be: Will some personal factors such as age, race or sex and some external factors such as strip search not affect people's attitudes towards cooperation? As for the whole dataset, all attributes exclude some ID information such as Event_ID, Person_ID, Arrest_ID and the last five attributes which with too many null values are considered in this logistic regression. We do not want to affect the final training results by artificially deleting some available attributes.

Similarly, before we conduct the logistic regression, there are some assumptions we need to fulfill, and we have already checked the following assumptions: (1) Linearity of the Logit: The

relationship between the independent variable and the logit of the outcome is linear. (2) Binary outcome: The outcome variable is binary. (3) Homoscedasticity: The variance of errors is constant for different levels of independent variables. (4) Large sample size: The sample size is large enough for doing the logistic regression. (5) Random sample: The data used for analysis is randomized and is representative of the population.

As for our project, the purpose of doing this logistic regression model is that we want to figure out how different factors contribute to affect the change in the cooperative count, and the logistic regression model is such a strong tool to fulfill this. There are some steps to conduct this model. Firstly, after selecting the attributes we want to put into the model, we encode all categorical data into the numeric. Then, we split the dataset into the training set and the testing set with a ratio of 4:1. The training set is used to build an appropriate model, and the test set is used to evaluate the accuracy of the training model. By looking at the summary of the logistic regression, the confidence interval could tell use the estimate of the range within which the true population parameter is likely to lie. And the prediction intervals could provide the estimate of the range within which the predicted probability of the binary outcome is likely to fall. Finally, the accuracy value and the confusion matrix could give us an intuitive attitude toward the advantages and disadvantages of the final logistic regression model.

Results and Findings

Power Analysis

Cooperative Count and Age Groups

The cooperative's study investigated the effect of age groups on a specific variable and found a large effect size of 6.28 (Cohen's D) for age. To ensure the study's statistical power, a sample size of 11.226 was required for the not youth group, while a sample size of 70.453 was needed for the youth group. However, the actual sample size for the not youth group was 29, which exceeded the required number. The actual sample size for the youth group was 182, which was also larger than the required number. These findings suggest that the study had sufficient statistical power to accurately examine the effect of the age group on the variable in question, and the larger sample sizes for both age groups likely enhance the study's generalizability.

Cooperative Count and Perceived Race

The cooperative study investigated the effect of race on a specific variable and found a small effect size of 0.15 (Cohen's D) for race. To ensure the study's statistical power, a sample size of 72.917 was needed for the non-white race group, while a sample size of 11.157 was required for the white race group. The actual sample size for the white race group was 183, which exceeded the required number. However, the actual sample size for the non-white group was only 28, which fell short of the required sample size. These findings suggest that the study had limited

statistical power to accurately examine the effect of race on the variable in question for the non-white group. The larger sample size for the white group could enhance the study's generalizability for this group, but caution is needed when interpreting the results for the non-white group due to the smaller sample size.

Cooperative Count and Sex

The study examined the effect of sex on a specific variable and found a large effect size of 1.13 (Cohen's D) for sex. To achieve sufficient statistical power, a sample size of 66.644 was required for the female group, while a sample size of 75.396 was needed for the male group. The actual sample size for the male group was 99, which exceeded the required number. Similarly, the actual sample size for the female group was 112, which also exceeded the required number. These results suggest that the study had the sufficient statistical power to accurately examine the effect of sex on the variable in question for both male and female groups. The larger sample sizes for both groups could enhance the study's generalizability and provide more reliable estimates of the effect of sex on the variable of interest.

Power Analysis Plot

The plot (Figure 7) shows the relationship between sample size (x-axis) and statistical power (y-axis) for different effect sizes (lines with different colors). The effect sizes shown in the plot are 0.2 (blue line), 0.5 (orange line), and 0.8 (green line).

It indicates that as the sample size increases, the statistical power of the test also increases. This is expected, as a larger sample size allows for a more precise estimation of population parameters and reduces the variability of the estimate. It also shows that for a given effect size, a larger sample size results in higher statistical power. For example, for an effect size of 0.5, a sample size of 50 provides approximately 0.5 power, whereas a sample size of 150 provides approximately 0.95 power.

Overall, the plot highlights the importance of selecting an appropriate sample size to achieve the desired statistical power for a given effect size.

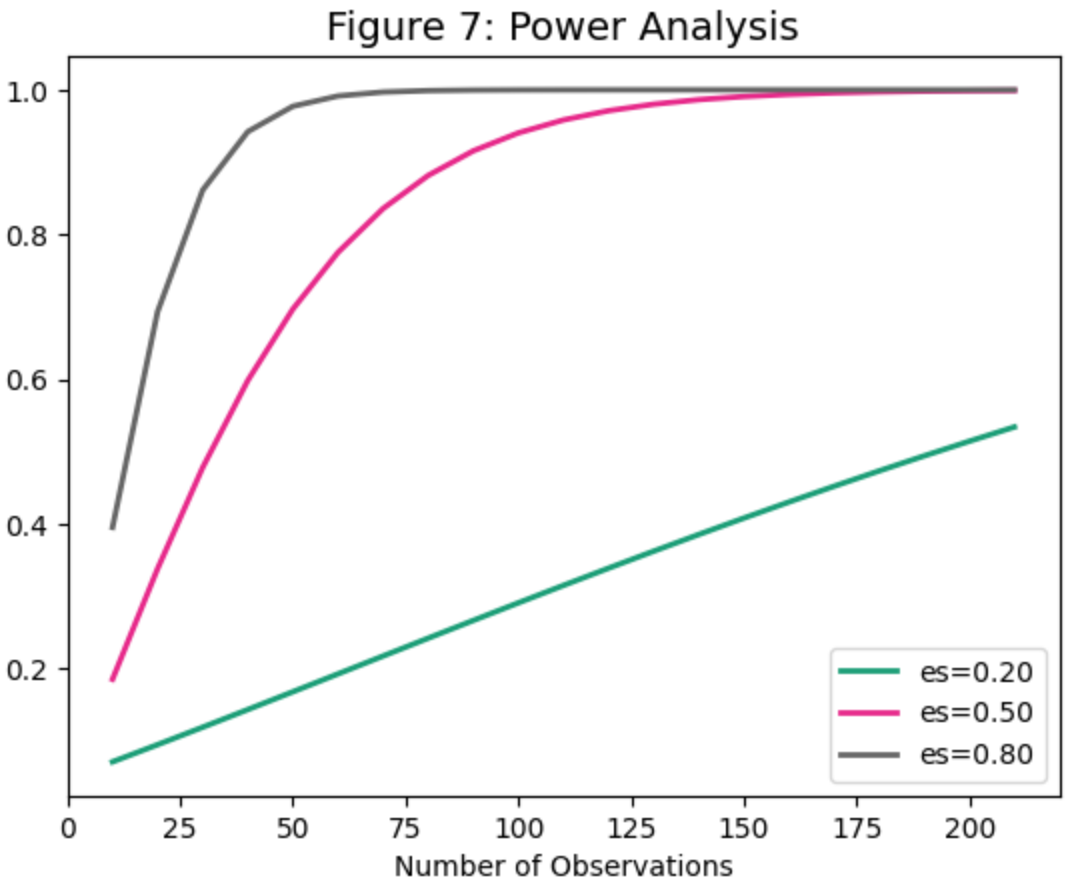

Figure 7: Power Analysis

ANCOVA

Age Group and Booked Count Based on Cooperative Count

Statistical Interpretation The ANCOVA test examined the effect of age_group_at_arrest on the cooperative count variable. The p-value of 7.647881e-02 indicates that there is some evidence against the null hypothesis that there is no difference in the means of the groups after adjusting for age ($p>0.05$). The p-value is greater than the conventional alpha level of 0.05, suggesting that the evidence is not strong enough to reject the null hypothesis. The F-value of 1.936807 is relatively low, suggesting that the age variable has only a small effect on the outcome variable.

Practical Interpretation A p-value greater than 0.05 generally means that no effect was observed. Therefore, we can conclude that age may have some effect on the outcome variable, but the evidence is not strong enough to make a definitive conclusion. Additional analyses or larger sample sizes may be needed. We cannot reject the null hypothesis meaning there is no significant relationship between age groups and cooperative count based on controlling the

booked count. The age of the arrested person may not affect the cooperative count, but might still impact the policers' attitude during attestation.

Table 2: Age Group and Booked Count Based on Cooperative Count

	Source	SS	DF	F	p-unc	np2
0	Age_group__at_arrest_	4.059594e+05	6	1.936807	7.647881e-02	0.054146
1	Booked_count	1.714817e+06	1	49.087719	3.531313e-11	0.194725
2	Residual	7.091548e+06	203	NaN	NaN	NaN

Race and Booked Count Based on Cooperative Count

Statistical Interpretation The results of the ANCOVA test indicate that the perceived race variable has a p-value of 0.034455 ($p < 0.05$) and an F-value of 4.531379, while the booked count variable has a p-value of 0.00001 and an F-value of 24.682181. The p-value for perceived race is less than 0.05, thus we can reject the null hypothesis, meaning there is a significant relationship between age groups and cooperative count based on controlling the booked count. Each race group results in a different cooperative count by controlling the booked count.

Table 3: Race and Booked Count Based on Cooperative Count

	Source	SS	DF	F	p-unc	np2
0	Perceived_Race	1.598543e+05	1	4.531379	0.034455	0.021321
1	Booked_count	8.707177e+05	1	24.682181	0.000001	0.106077
2	Residual	7.337653e+06	208	NaN	NaN	NaN

Sex and Booked Count Based on Cooperative Count

Statistical Interpretation The perceived sex variable has a p-value of 0.08584875, and an F-value of 2.99791, while the booked count variable has a p-value of 0.05689867 and an F-value of 31.73. The p-value for perceived sex is greater than 0.05, indicating that there is not enough evidence to reject the null hypothesis that there is no significant relationship between perceived sex and cooperative count when controlling for the booked count. This suggests that there may not be a significant difference in the cooperative count between different perceived sex groups after controlling for the effect of the booked count. We cannot reject the null hypothesis and conclude that there is no significant relationship between perceived sex and cooperative count when controlling for the booked count.

Table 4: Sex and Booked Count Based on Cooperative Count

	Source	SS	DF	F	p-unc	np2
0	Sex	1.065264e+05	1	2.997910	8.485475e-02	0.014208
1	Booked_count	1.127705e+06	1	31.736329	5.689867e-08	0.132380
2	Residual	7.390981e+06	208	NaN	NaN	NaN

Logistic Regression

Compared with females, the log odds of men getting booked are 0.038 higher, controlling for other features. However, the feature sex is not statistically significant (see Table 5 for the summary of logistic regression). We get the odds ratio to make the features more interpretable (see Table 6).

Here below are the interpretation for all 26 independent variables with intercept terms with the dependent variable cooperative at arrest:

1. Intercept: If all the independent variables are with the value of 0, the log odds ratio for cooperative at arrest is 234.3669. By comparing with the significant level, it is statistically significant.
2. Arrest_Year: If one year/unit increases in arrest year, the log odds ratio for cooperative at arrest will decrease by 0.1164. By comparing with the significant level, it is statistically significant.
3. Youth_at_arrest_under_18_years_: If people are under 18 years when arrested, the log odds ratio for cooperative at arrest will increase by 0.5405. By comparing with the significant level, it is not statistically significant.
4. Booked: If people are being booked, the log odds ratio for cooperative at arrest will increase by 0.4146. By comparing with the significant level, it is statistically significant.
5. Strip_Search: If people are being strip-searched, the log odds ratio for cooperative at arrest will decrease by 0.2331. By comparing with the significant level, it is statistically significant.
6. Actions_at_arrest___Concealed_i: If people are having concealed item action at arrest, the log odds ratio for cooperative at arrest will decrease by 0.2725. By comparing with the significant level, it is not statistically significant.
7. Actions_at_arrest___Combative__: If people are having combative action at arrest, the log odds ratio for cooperative at arrest will decrease by 3.1272. By comparing with the significant level, it is statistically significant.

8. Actions_at_arrest___Resisted__d: If people have resisted and defensive action at arrest, the log odds ratio for cooperative at arrest will decrease by 2.7640. By comparing with the significant level, it is statistically significant.
9. Actions_at_arrest___Mental_inst: If people are having mental instability action at arrest, the log odds ratio for cooperative at arrest will decrease by 0.4012. By comparing with the significant level, it is statistically significant.
10. Actions_at_arrest___Assaulted_o: If people are having assaulted officer action at arrest, the log odds ratio for cooperative at arrest will decrease by 4.5351. By comparing with the significant level, it is statistically significant.
11. Jan-Mar: By comparing with arrest months Apr-June, if people are arrested in Jan-Mar, the log odds ratio for cooperative at arrest will increase by 0.0620. By comparing with the significant level, it is statistically significant.
12. July-Sept: By comparing with arrest months Apr-June, if people are arrested in July-Sept, the log odds ratio for cooperative at arrest will decrease by 0.0179. By comparing with the significant level, it is not statistically significant.
13. Oct-Dec: By comparing with arrest months Apr-June, if people are arrested in Oct-Dec, the log odds ratio for cooperative at arrest will increase by 0.0217. By comparing with the significant level, it is not statistically significant.
14. East/Southeast Asian: By comparing with the black race group, if people belong to East/Southeast Asian, the log odds ratio for cooperative at arrest will increase by 0.3177. By comparing with the significant level, it is statistically significant.
15. Indigenous: By comparing with the black race group, if people belong to Indigenous, the log odds ratio for cooperative at arrest will decrease by 0.0077. By comparing with the significant level, it is not statistically significant.
16. Latino: By comparing with the black race group, if people belong to Latino, the log odds ratio for cooperative at arrest will increase by 0.3020. By comparing with the significant level, it is statistically significant.
17. Middle Eastern: By comparing with the black race group, if people belong to Middle Eastern, the log odds ratio for cooperative at arrest will increase by 0.1077. By comparing with the significant level, it is statistically significant.
18. South Asian: By comparing with the black race group, if people belong to South Asian, the log odds ratio for cooperative at arrest will increase by 0.0868. By comparing with the significant level, it is statistically significant.
19. Unkown Legacy: By comparing with the black race group, if people belong to Unkown Legacy, the log odds ratio for cooperative at arrest will decrease by 0.0031. By comparing with the significant level, it is not statistically significant.
20. White: By comparing with the black race group, if people belong to White, the log odds ratio for cooperative at arrest will increase by 0.0643. By comparing with the significant level, it is statistically significant.

21. M: By comparing with the female, if people are male, the log odds ratio for cooperative at arrest will increase by 0.0380. By comparing with the significant level, it is not statistically significant.
22. Aged 18 to 24 years: By comparing with the group Aged 17 years and younger, if people are Aged 18 to 24 years, the log odds ratio for cooperative at arrest will increase by 0.4348. By comparing with the significant level, it is not statistically significant.
23. Aged 25 to 34 years: By comparing with the group Aged 17 years and younger, if people are Aged 25 to 34 years, the log odds ratio for cooperative at arrest will increase by 0.4411. By comparing with the significant level, it is not statistically significant.
24. Aged 35 to 45 years: By comparing with the group Aged 17 years and younger, if people are Aged 35 to 45 years, the log odds ratio for cooperative at arrest will increase by 0.4704. By comparing with the significant level, it is not statistically significant.
25. Aged 45 to 54 years: By comparing with the group Aged 17 years and younger, if people are Aged 45 to 54 years, the log odds ratio for cooperative at arrest will increase by 0.5113. By comparing with the significant level, it is not statistically significant.
26. Aged 55 to 64 years: By comparing with the group Aged 17 years and younger, if people are Aged 55 to 64 years, the log odds ratio for cooperative at arrest will increase by 0.4775. By comparing with the significant level, it is not statistically significant.
27. Aged 65 years and older: By comparing with the group Aged 17 years and younger, if people are Aged 55 to 64 years, the log odds ratio for cooperative at arrest will increase by 0.6139. By comparing with the significant level, it is not statistically significant.

Table 5: Summary Statistic Results of Logistic Regression

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	234.3669	38.851	6.033	0.000	158.221	310.513
ArrestYear	-0.1164	0.019	-6.055	0.000	-0.154	-0.079
Youthatarrestunder18years	0.5405	0.557	0.971	0.331	-0.550	1.631
Booked	0.4146	0.019	21.467	0.000	0.377	0.453
StripSearch	-0.2331	0.032	-7.244	0.000	-0.296	-0.170
ActionsatarrestConcealedi	-0.2725	0.165	-1.647	0.100	-0.597	0.052
ActionsatarrestCombative	-3.1272	0.127	-24.549	0.000	-3.377	-2.878
ActionsatarrestResistedd	-2.7640	0.117	-23.605	0.000	-2.994	-2.535
ActionsatarrestMentalinst	-0.4012	0.062	-6.483	0.000	-0.523	-0.280

ActionsatarrestAssaultedo	-4.5351	1.005	-4.512	0.000	-6.505	-2.565
JanMar	0.0620	0.026	2.395	0.017	0.011	0.113
JulySept	-0.0179	0.026	-0.685	0.493	-0.069	0.033
OctDec	-0.0217	0.027	-0.811	0.417	-0.074	0.031
M	0.0380	0.023	1.630	0.103	-0.008	0.084
EastSoutheastAsian	0.3177	0.039	8.134	0.000	0.241	0.394
Indigenous	-0.0077	0.058	-0.134	0.894	-0.120	0.105
Latino	0.3020	0.058	5.177	0.000	0.188	0.416
MiddleEastern	0.1077	0.044	2.434	0.015	0.021	0.194
SouthAsian	0.0868	0.042	2.056	0.040	0.004	0.170
UnknownorLegacy	-0.0031	0.037	-0.084	0.933	-0.076	0.070
White	0.0643	0.023	2.785	0.005	0.019	0.110
Aged18to24years	0.4348	0.555	0.783	0.434	-0.654	1.523
Aged25to34years	0.4411	0.555	0.794	0.427	-0.647	1.529
Aged35to44years	0.4704	0.555	0.847	0.397	-0.618	1.559
Aged45to54years	0.5113	0.556	0.920	0.357	-0.577	1.600
Aged55to64years	0.4775	0.556	0.859	0.390	-0.612	1.567
Aged65yearsandolder	0.6139	0.559	1.099	0.272	-0.481	1.709

The odds ratio estimates for the different age groups (18-24, 25-34, 35-44, 45-54, 55-64, and 65+) all suggest that being in a higher age group increases the odds of a higher cooperative count. For example, the odds ratio estimate for "Aged 18 to 24 years" is 0.520 to 4.588, suggesting that being in this age group may have a large effect on the cooperative count; the estimates for the other age groups are similar. Therefore the difference is not considered significant.

The odds ratio estimates for the different race groups (East/Southeast Asian, Indigenous, Latino, Middle Eastern, South Asian, Unknown or Legacy, and White) suggest that being in certain racial groups may increase or decrease the odds of a higher cooperative count. For example, the odds ratio estimate for "Latino" is 1.206 to 1.516, suggesting that being Latino increases the odds

of a higher cooperative count. The estimates for the other racial groups also suggest some effect on the cooperative count.

The odds ratio estimate for "Sex" is 0.992 to 1.087, suggesting that being female or male does not have a significant effect on the cooperative count, holding all other variables constant. The coefficient is very close to 1, and the 95% confidence interval for the coefficient ranges from 0.992 to 1.087, suggesting that the effect of sex on the cooperative count is not statistically significant. Therefore, we would interpret this coefficient as indicating that there is little to no effect of sex on the cooperative count in this logistic regression model.

Table 6: Summary Statistic Results of CI & Odds Ratio

	Lower CI	Upper CI	OR
Intercept	5.183830e+68	7.143122e+134	6.085124e+101
ArrestYear	8.571934e-01	9.242926e-01	8.901109e-01
Sex	9.923257e-01	1.087400e+00	1.038776e+00
Youthatarrestunder18years	5.767810e-01	5.109980e+00	1.716782e+00
Booked	1.457593e+00	1.572240e+00	1.513832e+00
StripSearch	7.437074e-01	8.436664e-01	7.921117e-01
ActionsatarrestConcealedi	5.505908e-01	1.053223e+00	7.615084e-01
ActionsatarrestCombative	3.415458e-02	5.627419e-02	4.384086e-02
ActionsatarrestResistedd	5.010905e-02	7.929800e-02	6.303608e-02
ActionsatarrestMentalinst	5.930060e-01	7.558346e-01	6.694882e-01
ActionsatarrestAssaultedo	1.495949e-03	7.690230e-02	1.072576e-02
JanMar	1.011324e+00	1.119346e+00	1.063965e+00
JulySept	9.330723e-01	1.033959e+00	9.822210e-01
OctDec	9.286802e-01	1.031149e+00	9.785744e-01
EastSoutheastAsian	1.272664e+00	1.483188e+00	1.373900e+00
Indigenous	8.864940e-01	1.110806e+00	9.923321e-01
Latino	1.206435e+00	1.516399e+00	1.352567e+00
MiddleEastern	1.021220e+00	1.214678e+00	1.113757e+00
SouthAsian	1.004072e+00	1.184839e+00	1.090717e+00
UnknownorLegacy	9.267085e-01	1.072368e+00	9.968816e-01
White	1.019231e+00	1.115795e+00	1.066421e+00
Aged18to24years	5.199987e-01	4.587955e+00	1.544581e+00
Aged25to34years	5.235684e-01	4.614709e+00	1.554386e+00
Aged35to44years	5.390844e-01	4.752843e+00	1.600682e+00
Aged45to54years	5.613018e-01	4.953172e+00	1.667401e+00
Aged55to64years	5.420809e-01	4.793717e+00	1.612012e+00
Aged65yearsandolder	6.182611e-01	5.521778e+00	1.847674e+00

Tables 7 and 8 below demonstrate that the model correctly predicted the outcome about 58% of the time for training data, and 57% for test data. Actually, the accuracy score is not high enough which means this logistic regression model may not be the best model to interpret this result.

There are a few reasons why this happens:

(1) The relationship between the cooperative count and these independent variables is not linear.

(2) Because of excluding some attributes such as the occurrence type, insufficient or inappropriate features make this happen.

(3) There exists the overfitting problem because of the overly complex or has too many features in this logistic regression model.
(4) Considering that our dataset is imbalanced on some attributes, for example, for the perceived race attributed, white and black data are more than other race groups, and for sex data, the male population is more than the female one. And the logistic regression models can be affected by this class imbalance, where one class has significantly fewer examples than the other.
(5) The sample size is also a factor that may affect the accuracy score. A smaller sample size will make the model have not enough data to learn meaningful patterns, resulting in lower accuracy.

Table 7: Confusion Matrix for Train Data

	Prediction True	Prediction False
Actual True	20224	8780
Actual False	13197	10012

Test accuracy (Train) = 0.5790894987838278

Table 8: Confusion Matrix for Test Data

	Prediction True	Prediction False
Actual True	4826	2231
Actual False	3358	2539

Test accuracy (Test) = 0.5718553700015321

Prediction Interval

The prediction intervals are used to estimate a range of possible values for an individual data point or observation that is yet to be observed, based on a model or a set of data. Based on our result, the age groups suggest that being in any of the age groups from 18 to 64 years old does not have a significant effect on the number of individuals booked, as the intervals for all age groups include 1, which means that there is no significant difference in booking rates between these age groups and the reference group (those aged 17 years and under).

The prediction interval for sex indicates that the model is confident that the effect of sex on the number of booked individuals falls between 0.992 and 1.087, which means that being male or female does not have a significant effect on the number of individuals booked, all else being equal.

The prediction intervals for race/ethnicity suggest that individuals who identify as East/Southeast Asian, Latino, or Middle Eastern may be more likely to be booked compared to the reference group (individuals who identify as Black), as their prediction intervals do not include 1. However, individuals who identify as Indigenous, South Asian, Unknown/Legacy, or White do not show significant differences in booking rates compared to the reference group, as their prediction intervals include 1.

All these are consistent with our previous analysis, so we can conclude that age and sex do not have a significant effect on the number of booked, and race has significant differences among groups. We can see the results from the plots as well (Figure 8 and Figure 9).

Table 9: Summary Statistic Results of PI

	Lower PI	Upper PI
Intercept	5.176582e+68	7.153124e+134
ArrestYear	8.571928e-01	9.242932e-01
Sex	9.923249e-01	1.087401e+00
Youthatarrestunder18years	5.767694e-01	5.110083e+00
Booked	1.457592e+00	1.572241e+00
StripSearch	7.437065e-01	8.436673e-01
ActionsatarrestConcealedi	5.505875e-01	1.053229e+00
ActionsatarrestCombative	3.415442e-02	5.627445e-02
ActionsatarrestResistedd	5.010884e-02	7.929834e-02
ActionsatarrestMentalinst	5.930047e-01	7.558363e-01
ActionsatarrestAssaultedo	1.495895e-03	7.690509e-02
JanMar	1.011323e+00	1.119347e+00
JulySept	9.330714e-01	1.033960e+00
OctDec	9.286793e-01	1.031150e+00
EastSoutheastAsian	1.272662e+00	1.483190e+00
Indigenous	8.864921e-01	1.110808e+00
Latino	1.206433e+00	1.516403e+00
MiddleEastern	1.021219e+00	1.214680e+00
SouthAsian	1.004070e+00	1.184841e+00
UnknownorLegacy	9.267072e-01	1.072370e+00
White	1.019230e+00	1.115796e+00
Aged18to24years	5.199883e-01	4.588046e+00
Aged25to34years	5.235579e-01	4.614801e+00
Aged35to44years	5.390736e-01	4.752939e+00
Aged45to54years	5.612906e-01	4.953271e+00
Aged55to64years	5.420701e-01	4.793813e+00
Aged65yearsandolder	6.182486e-01	5.521889e+00

As for our two prediction interval plots, we have the following explanations for them. Figure 8, which is the 95% prediction interval graph for race and cooperative at arrest. As our group select a 0.05 significance level as the criteria, it can be seen that the positive correlation between different racial groups and the probability of taking cooperative action during an arrest is increasing. This model has a higher probability of predicting white, unknown or legacy, or black racial groups. However, for all probabilities in the prediction interval, it is below 0.5, indicating that its prediction accuracy is relatively low and the results are not very reliable.

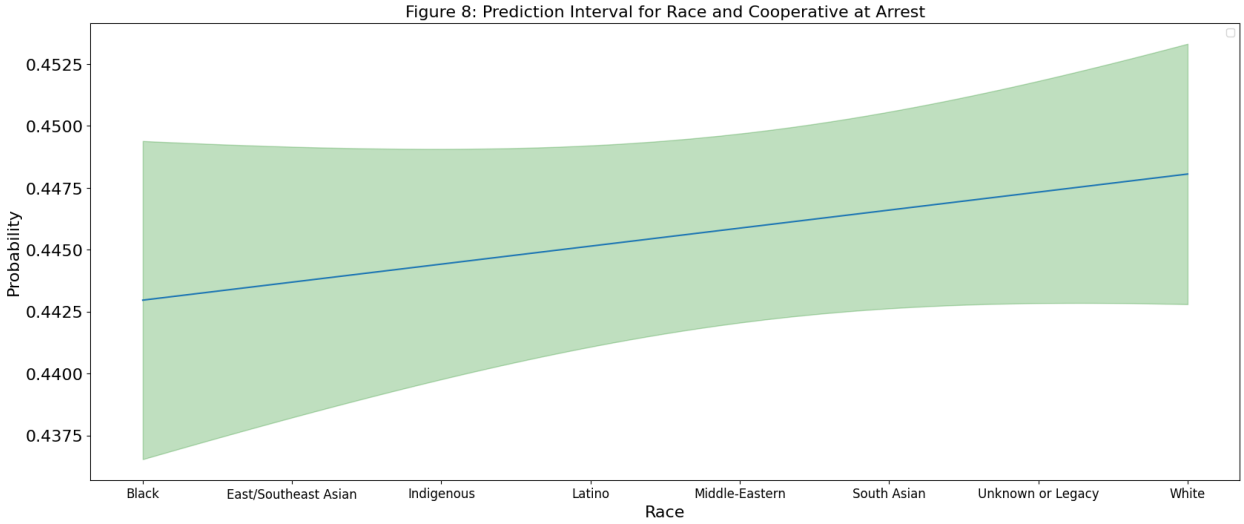

Figure 8: Prediction Interval for Race and Cooperative at Arrest

The other plot is the 95% prediction interval graph for age and cooperative at arrest in Figure 9. As our group select a 0.05 significance level as the criteria, it can be seen that the positive correlation between different age groups and the probability of taking cooperative action during an arrest is increasing. This model has a higher probability of predicting aged 17 and under, aged 65 years and older and aged 55 to 64 years age groups. However, for all probabilities in the prediction interval, it is below 0.5, indicating that its prediction accuracy is relatively low and the results are not very reliable. Also, based on our previous logistic regression model results, all age groups have a p-value larger than 0.05, thus we think for our study, the age groups may not such important based on the results.

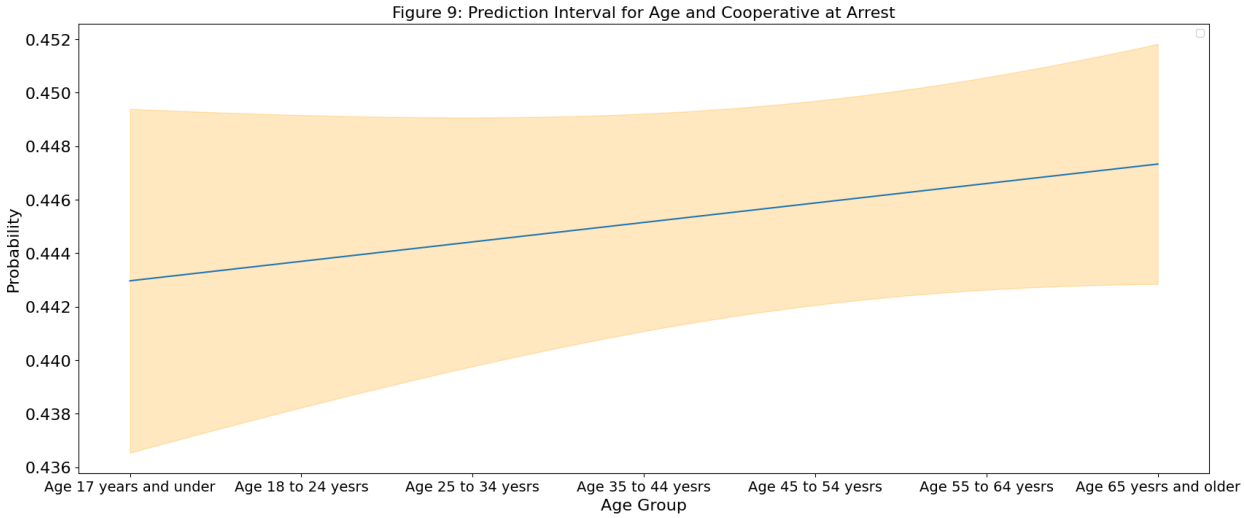

Figure 9: Prediction Interval for Age Groups and Cooperative at Arrest

Discussion

The results of the statistical tests indicate that there are no significant differences in the booking across different demographic groups based on sex and age, but there is a significantly different when race is taken into account. These findings raise important questions about the potential for bias and discrimination in policing practices and highlight the need for further research and policy reform. The following sections discuss the implications of these findings and suggest avenues for future research.

Limitations

There are several limitations to this study. First, the dataset only includes information on individuals who were arrested and brought into custody by the Toronto Police Service and therefore does not provide a complete picture of police interactions with the public. Second, the dataset does not include information on the circumstances surrounding each arrest, which may influence the likelihood of a booking. Finally, the dataset is limited to a specific time period and geographic area (Toronto), which may limit its generalizability to other locations and time periods.

Implications and Future Research

The findings of this study have important implications for policy and practice in the Toronto Police Service and other law enforcement agencies. The results suggest that certain demographic groups such as black individuals may be disproportionately impacted by police booking practices. These practices can have serious consequences for individuals and communities, including loss of freedom, harm to personal dignity, and negative impacts on mental health and

well-being. As we are limited to the dataset, we would need to collect more data from a larger population to perform more tests to test our hypothesis.

To address these issues, there is a need for further research and policy reform. One potential avenue for future research is to explore the underlying factors that contribute to these disparities in policing practices. For example, the researchers could examine the role of implicit biases and stereotypes in shaping police behaviour, as well as the impact of systemic factors such as racial profiling and the over-policing of certain communities. By better understanding the factors that contribute to disparities in policing practices, researchers and policymakers can develop evidence-based interventions to address these issues and promote more equitable and just policing practices.

In addition to research, there is a need for policy reform to address the issues highlighted in this study. For example, policies could be developed to increase transparency and accountability in policing practices, such as requirements to report on the demographics of individuals subjected to booking. These policies could help to promote greater equity and fairness in policing practices and reduce the potential for bias and discrimination.

Another potential avenue for policy reform is to increase diversity and cultural competency within law enforcement agencies. Research has shown that police officers from diverse backgrounds are better able to build trust and rapport with the communities they serve, and that cultural competency training can help officers to better understand and navigate cultural differences. By increasing the diversity and cultural competency of law enforcement agencies, it may be possible to reduce the potential for bias and discrimination in policing practices and promote more equitable and just policing practices.

Finally, it is important to recognize that the issues highlighted in this study are part of a broader pattern of systemic racism and inequality in our society. To address these issues, there is a need for broader social and political change that goes beyond the realm of law enforcement. This could include efforts to reduce economic and social inequality, increase access to education and healthcare, and promote diversity and inclusion in all areas of society. By addressing the root causes of systemic inequality, it may be possible to reduce the potential for bias and discrimination in all areas of our society, including law enforcement.

Conclusion

The results of this study provide important insights into the relationship between demographic factors and booking rates in the Toronto Police Service. The findings suggest that gender, race, and age play significant roles in these practices, however, the differences in gender and age are not as significant as race. It also highlights potential issues of bias and discrimination in policing. These findings underscore the need for further research and policy reform to address these issues

and promote more equitable and just policing practices. By working to address the root causes of systemic inequality and promoting greater equity and fairness in all areas of our society, we can help to create a safer and more just world for all people.

References

Arrests and strip searches (RBDC-arr-TBL-001). Toronto Police Service Public Safety Data Portal. (n.d.). Retrieved February 27, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Brunson, R. K. (2007). "Police don't like Black people": African American young men's accumulated police experiences. Criminology & Public Policy, 6(1), 71-102.

Eitle, D., & Turner, R. J. (2002). Exposure to official bias and socialization into deviance: The case of marijuana use. Justice Quarterly, 19(1), 113-146.

Ghandnoosh, N. (2015). The war on marijuana in black and white. The Sentencing Project. Retrieved from https://www.sentencingproject.org/publications/the-war-on-marijuana-in-black-and-white/

Goff, P. A., Lloyd, M. H., Geller, A., Raphael, S., & Glaser, J. (2016). The science of justice: Race, arrests, and police use of force. Journal of Social Issues, 72(4), 802-823.

Jain, E. (2021). The Mark of Policing: Race and Criminal Records. Stanford Law Review Online, 73.

Lee, H. Y., Yoo, J. W., & Park, J. (2019). The effects of race/ethnicity on police use of force: An analysis of officer-involved shootings in the United States. Journal of Ethnicity in Criminal Justice, 17(2), 101-120.

Mitchell, O., & MacKenzie, D. L. (2004, December). The Relationship between Race, Ethnicity, and Sentencing Outcomes: A Meta-Analysis of Sentencing Research . U.S. Department of Justice. Retrieved April 16, 2023, from https://www.ojp.gov/

Miller, J., & Schwartz, J. (2018). Gender differences in police stops and searches: Evidence from the British Crime Survey. European Journal of Criminology, 15(5), 537-554.

Toronto Population 2021/2022. (n.d.). Canada Population. Retrieved from https://canadapopulation.org/toronto-population/

Weitzer, R., & Tuch, S. A. (2006). Race and policing in America: Conflict and reform. Cambridge University Press.