2178 Final Project

Group 14

Youwen Xu (1006675493)
Jiayi Li (1005191374)

Faculty of Information
University of Toronto
INF2178: Experimental Design for Data Science
Prof.Shion Guha

April 16th, 2023

Introduction

1.1 Background and Literature Review

Strip searches and arrests are considered part of the police investigation and enforcement process. Also, this is required to be conducted only by specialized law enforcement agencies and officers. A strip search is defined as a visual inspection of an individual's entire body by police officers, and the investigation includes the removal of clothing (Toronto Police Service, 2022). The reasons for conducting strip searches and making arrests can vary depending on the specific circumstances of the area. One investigation (Derek Willis, 2020) indicated that approximately 4,000 officers in the New York City Police Department had misconduct charges. There is also a report (Barrison Law, 2021) that discussed illegal strip searches by police officers in Ontario, which would involve an invasion of privacy and have a negative impact on society.

Furthermore, a report (McNeilly, 2019) on police strip searches in Ontario mentions that there are approximately 22,000 strip searches per year, with the Toronto Police Service accounting for the majority. This report indicates that over 40% of arrests in Toronto have involved a strip search. In comparison to the other regions of Ontario, Toronto's strip search rate was also 40 times higher during 2014-2015. In this report, the importance of conducting strip searches legally is explored, along with recommendations on how to enhance the search process. Considering the case-specific and complex nature of strip searches and arrests, we will explore how influences arising from different states in police investigations predict the occurrence of strip searches and bookings. Finally, we hypothesize that analyzing the state of the person being arrested at the moment can help government departments improve the search process and reduce the potential negative impact on society. For this project, we will use a dataset from the Toronto Police Service about strip searches and arrests.

1.2 Research Ojective and Questions

Based on the above literature, our study will investigate how sex, race, arrested year, arrested status, and booked are interrelated with strip searches. In the EDA (Exploratory data analysis) section, the descriptive statistics and t-tests will explain in detail. Through the background research and the information provided by the dataset, we have formed two research questions.

1. How are the demographic attributes of arrestees (such as race, sex) and arrested year distributed? Do these attributes influence the strip searches of arrestees?

2. How does action at arrest affect whether the arrestee will be booked? (This includes situations where the person may not be strip searched).

Since these different factors may influence whether a person is strip-searched or only booked at the time of the arrest, we believe these questions allow for a more in-depth study of the dataset and analysis of valuable insights.

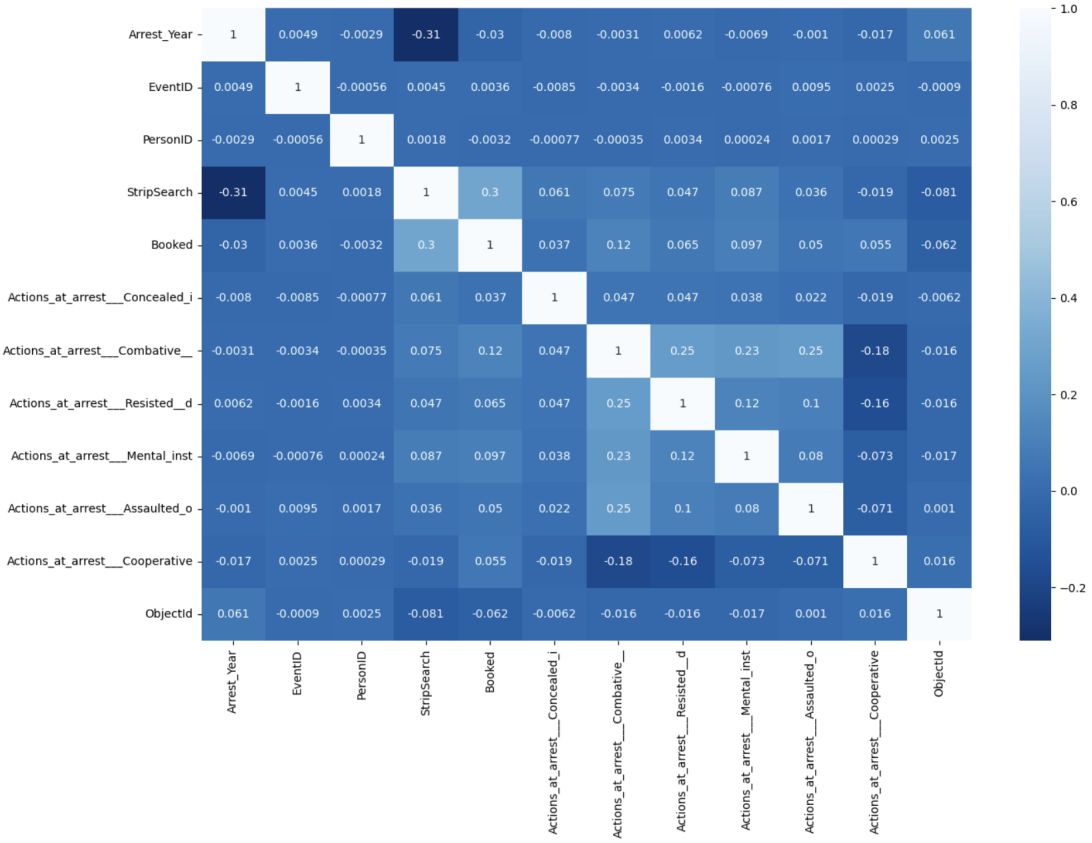
EDA (Exploratory data analysis)

2.1 Descriptive statistics

First of all, it is necessary to have a basic understanding of the dataset provided by the Toronto Police Service. We conducted a correlation plot to illustrate the correlation between every two variables.

The value of the correlation coefficient is represented by the shade of the color. A darker color indicates a stronger correlation, while a lighter color indicates a weaker correlation. Therefore, as shown in Figure 1, there is a relatively high correlation with the variables that are compared to strip search and booked.

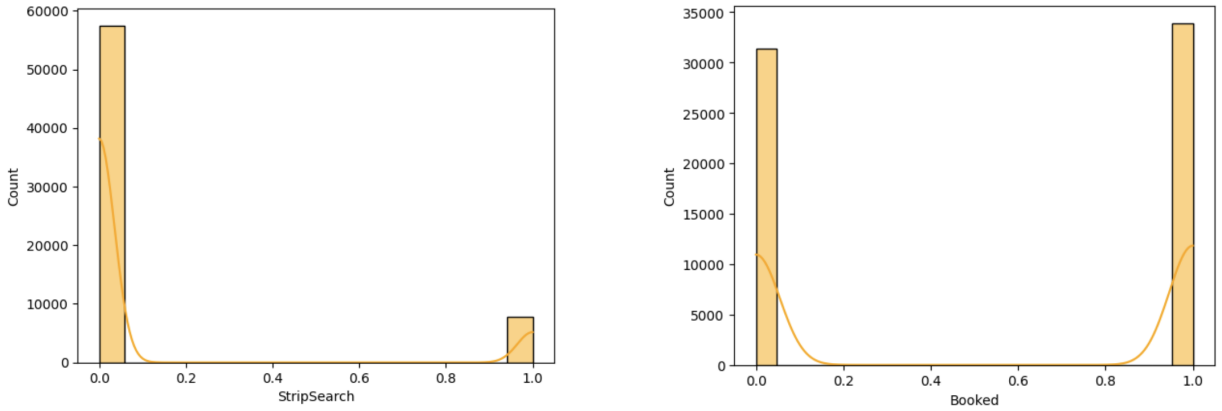
Figure 1. Correlation of Each Variable



Then, in order to explore variables with strong correlations, we plotted the histogram of strip-search and booked respectively to see their distribution (see Figure 2).

For the strip search, we can see that the tail of this distribution is longer on the right, and the left side of the distribution contains most of the data. Thus, the distribution is skewed to the right (positively skewed). This means that, there were relatively more data that were not strip searched (0 = 'No'). For the Booked, the data show that the number of booked (1 = 'Yes') is slightly higher than the number of unbooked (0 = 'No').

Figure 2: The Histogram of Strip-Search and The Histogram of Booked



Furthermore, we also created a graph on the race categories to get an insight into the background composition of the arrestees and the proportion of distribution of different races.

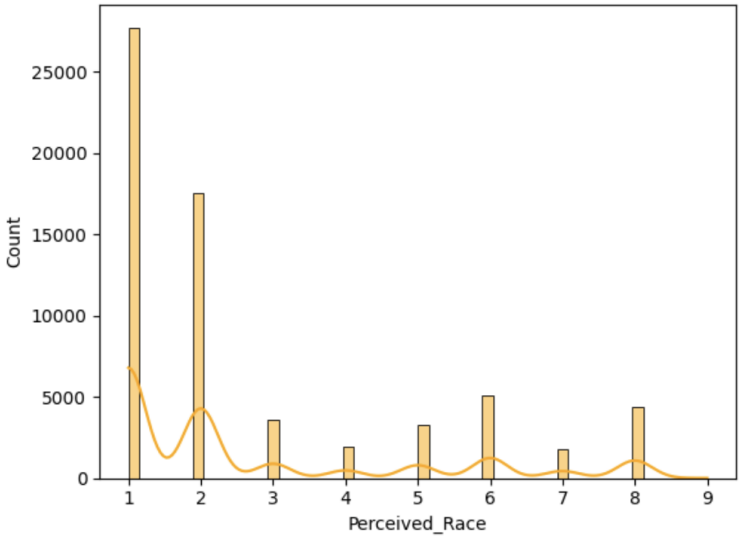
In Table 1, we can see the categories corresponding to all races. And Figure 3 shows that the highest number of races is white, with approximately 28,000, and the second highest is black, with an estimated 17,500. The relatively low number of races is Indigenous and Latino.

Table 1. Race Categories

Race	Categories
White	1
Black	2
South Asian	3
Indigenous	4
Middle-Eastern	5
Unknown or Legacy	6
Latino	7
East/Southeast Asian	8
Unknown	9

Source: Arrests and Strip Searches Dataset From Toronto Police Service.

Figure 3: The Histogram of Perceived Race



In the end, we also conducted a descriptive analysis of the sample size, mean, and standard deviation of some attributes in the dataset.

As shown in Table 2, the sample size of the data is 6,5276, except for race which is 65272. Here it may be due to the unknown status of some races. In addition, the mean and standard

deviation are not standardized. Therefore, it indicates that this dataset can provide statistically significant information.

Table 2. Descriptive Statistics

Variables	N	Mean	SD
Sex	6,5276	1.1936	0.3954
Race	6,5272	2.6902	2.2188
Arrest Year	6,5276	2020.5	0.4999
Actions at arrest (Concealed)	6,5276	0.0041	0.0637
Actions at arrest (Combative)	6,5276	0.0441	0.2054
Actions at arrest (Resisted)	6,5276	0.0383	0.1920
Actions at arrest (Assaulted)	6,5276	0.0064	0.0795
Booked	6,5276	0.5195	0.4996
Strip Search	6,5276	0.1195	0.3244

Source: Arrests and Strip Searches Dataset From Toronto Police Service.

2.2 T-Test

The t-test is useful to confirm whether there has a significant difference between the two groups and to analyze the relationship between the data as well as to reach a conclusion. If the p-value is less than 0.05, which indicates that there has a statistically significant difference between the two groups.

Sex and Strip Search

The data set showed that the number of males was higher than females in the strip search. We conducted a t-test to analyze whether the strip searches differed by sex. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, by sex, are equal).

H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups, by sex, are not equal)

The results we obtained show that the p-value is around 6.104e-12. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Race and Strip Search

The data set showed that the number of white was higher than black in the strip search. We conducted a t-test to analyze whether the strip searches differed by race. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, by race, are equal).

H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups, by race, are not equal)

The results we obtained show that the p-value is about 0.00174. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Arrest Year and Strip Search

We conducted a t-test to analyze whether the strip searches differed by arrested year. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, by arrested year, are equal).

H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups, by arrested year, are not equal)

The results we obtained show that the p-value is about 0. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Action at arrest (Concealed) and Booked

Like the above test, we conducted a t-test to analyze whether the circumstances of being booked differed based on actions that were concealed at the time of the arrest. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, concealed or not at the time of arrest, are equal).
H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups, concealed or not at the time of arrest, are not equal)

The results we obtained show that the p-value is about 3.426e-21. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Actions at arrest (Combative) and Booked

Like the above test, we conducted a t-test to analyze whether the circumstances of being booked differed based on actions that were combative at the time of the arrest. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, combative or not at the time of arrest, are equal).
H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups, combative or not at the time of arrest, are not equal)

The results we obtained show that the p-value is about 1.850e-193. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Actions at arrest (Resisted) and Booked

Like the above test, we conducted a t-test to analyze whether the circumstances of being booked differed based on actions that were resisted at the time of the arrest. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, resisted or not at the time of arrest, are equal).

H_A(Alternative Hypothesis) $\neq \mu_0$ (The means of the two groups, resisted or not at the time of arrest, are not equal)

The results we obtained show that the p-value is about 3.099e-61. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Actions at arrest (Assaulted) and Booked

Like the above test, we conducted a t-test to analyze whether the circumstances of being booked differed based on actions that were assaulted at the time of the arrest. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups, assaulted or not at the time of arrest, are equal).
H_A(Alternative Hypothesis) $\neq \mu_0$ (The means of the two groups, assaulted or not at the time of arrest, are not equal)

The results we obtained show that the p-value is about 9.372e-37. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can reject.

Method

3.1 Dataset description

The dataset is borrowed from the Toronto Police Service. This dataset presents all relevant information about arrests and strip searches, such as the demographics of the person arrested, occurrence category, actions at arrest, and the reason for the search.

This dataset was released on November 10, 2022, and contains a total of 65,276 records. The dataset can be found at the following link: https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about. (Toronto Police Service, 2022). It is important to note that in this dataset, booked is defined as whether or not the person was booked at the police station within 24 hours of the arrest. And if there is a record of a strip search, it will also be a registered event by default.

This dataset allows for an analysis of which specific situations are more likely to result in a strip search or arrest by exploring different types of arrests.

There are many attributes included in the dataset, but we primarily want to examine the relationship between variables of sex, race, arrest year, actions at arrest, strip search, and booked.

3.2 ANOVA Test

To determine whether the difference between the means of the strip search and the demographics of the arrestees (sex and race) is considered statistically significant, we will use the Anova Test.

As seen in Table 3, the p-value equals 4.258e-12, and the F-statistic equals 7.529. Since p is less than $\alpha = 0.05$, it can be shown that there is enough evidence to reject the null hypothesis of Anova. It also follows that there is a statistically significant difference between the demographics of the arrestees and the means of the two groups of strip searches. On the other hand, since the F-statistic is a large number, which demonstrates that the differences between the means of the groups are equally statistically significant.

Table 3. The ANOVA Table (Strip search as outcome variable)

	Df	Sum sq	F value	Pr(>F)
Sex: Race	16	12.577	7.529	4.258e-12***
Residuals	65255	6813.260	-	-

Source: Arrests and Strip Searches Dataset From Toronto Police Service.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

As seen in Table 4, for actions at arrest are concealed and combative, the p-value equals 6.547e-03, and the F-statistic equals 7.394. Since p is less than $\alpha = 0.05$, it can be shown that there is enough evidence to reject the null hypothesis of Anova. It also follows that there is a statistically significant difference between the actions at arrest and the means of the two groups of booked. On the other hand, since the F-statistic is a large number, which demonstrates that the differences between the means of the groups are equally statistically significant.

In addition, for actions at arrest are Resisted and Assaulted, the p-value equals 1.304e-02, and the F-statistic equals 6.164. Since p is less than $\alpha = 0.05$, it can be shown that there is enough evidence to reject the null hypothesis of Anova. It also follows that there is a statistically significant difference between the actions at arrest and the means of the two groups of booked. On the other hand, since the F-statistic is a large number, which demonstrates that the differences between the means of the groups are equally statistically significant.

Table 4. The ANOVA Table (Booked as outcome variable)

	Df	Sum sq	F value	Pr(>F)
Concealed : Combative	1	1.819	7.394	6.547e-03***
Residuals	65272	16057.85	-	-
Resisted : Assaulted	1	1.529	6.164	1.304e-02***
Residuals	65272	16194.32	-	-

Source: Arrests and Strip Searches Dataset From Toronto Police Service.
*** p < 0.001; ** p < 0.01; * p < 0.05.

3.3 Post-hoctests (Tukey's HSD)

Tukey's HSD is useful for comparing means between several groups and can be used to determine whether the differences between each group are statistically significant.

As shown in Table 5, while booked as the dependent variable, all other independent variables are indicated the null hypothesis are being rejected. (The differences between the different groups were statistically significant)

Table 5. The Tukey HSD Table (Booked as dependent variable)

	Concealed	Combative	Resisted	Assaulted
Reject	True	True	True	True

Source: Arrests and Strip Searches Dataset From Toronto Police Service.

As shown in Table 6, while strip search as the dependent variable, sex and some race (independent variables) groups are indicated the null hypothesis are rejected. (The differences between the different groups were statistically significant)

Note that some groups of race also indicated that the null hypothesis are not being rejected. (The differences between the different groups were not statistically significant)

Table 6. The Tukey HSD Table (Strip search as dependent variable)

	Sex(1:2)	Race(1:9)	Race(2:4)	Race(2:9)	Race(3:5)	Race(3:7)	Race(3:8)	Race(3:9)
Reject	True	False	False	False	False	False	False	False

	Race(4:9)	Race(5:7)	Race(5:8)	Race(6:9)	Race(7:8)	Race(7:9)	Race(8:9)
Reject	False	False	False	False	False	False	False

Note: Other group of Race are all True.
Source: Arrests and Strip Searches Dataset From Toronto Police Service.

3.4 ANCOVA Test

We will use the ANCOVA test to determine whether the difference between the two or more groups while controlling for one continuous covariate (arrest year) is considered statistically significant.

As seen in Table 7, for the representative sex and race, the p-value are all less than $\alpha = 0.05$. Also, since 0 is not lie on the confidence interval, it shows that there is enough evidence to reject the null hypothesis of Ancova. It also follows that there is a statistically significant difference between the demographics of the arrestees and strip searches while controlling for one continuous covariate (arrest year)

Table 7. The ANCOVA Table (Strip search as outcome variable)

	Std err	T	P>|t|	[0.025 0.975]
Sex	0.003	-6.960	0.000	-0.027 -0.015
Race	0.005	-10.253	0.000	–0.066 -0.045

Source: Arrests and Strip Searches Dataset From Toronto Police Service.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Moreover, as seen in Table 8, for actions at arrest are concealed, combative, resisted and assaulted, the p-values are all less than $\alpha = 0.05$. Then, we can see that 0 is not lie on any confidence intervals. This is shown that there is enough evidence to reject the null hypothesis of Ancova. It also follows that there is a statistically significant difference between the actions at arrest and booked while controlling for one continuous covariate (arrest year)

Table 8. The ANCOVA Table (Booked as outcome variable)

	Std err	T	P>|t|	[0.025 0.975]
Concealed	0.031	9.396	0.000	0.228 0.348
Combative	0.009	29.760	0.000	0.263 0.300
Resisted	0.010	16.582	0.000	0.149 0.188
Assaulted	0.025	12.669	0.000	0.263 0.359

Source: Arrests and Strip Searches Dataset From Toronto Police Service.
*** p < 0.001; ** p < 0.01; * p < 0.05.

3.5 Power Analysis

Power analysis is a way to determine the minimum sample size required to detect a significant effect with a desired level of statistical power. It is also used to determine whether differences between groups or conditions are statistically significant. The graph of power will be shown in the result section and indicate the relationship between the sample size, effect size, and power.

As shown in Table 9, where alpha was set at 0.05 and power was set at 0.80. The dependent variable in this analysis is Strip search, and the results show that sex has a significant effect (effect size = 3.311). However, the sample size is small (sample size =3), which can be limit the generalizability of the findings. Despite this limitation, the high power value of 0.99 indicates that the probability of correctly detecting the significant effect is very high.

For race, the results show there is a signification effect (effect size = 7.925). However, the sample size is small (sample size =10), which can be limit the generalizability of the findings. Despite this limitation, the high power value of 1 indicates that the probability of correctly detecting the significant effect is very high.

Table 9. The Power Analysis Table (Strip search as outcome variable)

	Effect Size	Sample Size	Power
Sex	3.311	3	0.99
Race	7.925	10	1

As shown in Table 10, where alpha was set at 0.05 and power was set at 0.80. The dependent variable in this analysis is Booked, and the results show that concealed has a relatively moderate effect (effect size = -1.032). However, the sample size is not large enough (sample size =16), which can be limit the generalizability of the findings. Despite this limitation, the high power value of 0.83 indicates that the probability of correctly detecting the significant effect is very high.

For combative, the results show there is a small effect (effect size = -0.951). However, the sample size is not large enough (sample size =18), which can be limit the generalizability of the findings. Despite this limitation, the high power value of 0.77 indicates that the probability of correctly detecting the significant effect is high.

For rersisted, the results show there is a small effect (effect size = -0.963). However, the sample size is not large enough (sample size =18), which can be limit the generalizability of the findings. Despite this limitation, the high power value of 0.77 indicates that the probability of correctly detecting the significant effect is high.

For assaulted, the results show there is a relatively moderate effect (effect size = -1.027). However, the sample size is not large enough (sample size =16), which can be limit the generalizability of the findings. Despite this limitation, the high power value of 0.82 indicates that the probability of correctly detecting the significant effect is very high.

Table 10. The Power Analysis Table (Booked as outcome variable)

	Effect Size	Sample Size	Power
Concealed	-1.032	16	0.83
Combative	-0.951	18	0.77
Resisted	-0.963	18	0.77
Assaulted	-1.027	16	0.82

3.6 Logistics regression model

"Logistics regression model is a statistical model that in its basic form uses a logistic function to model a binary response variable" (Caetano, 2020). We focus only on the action at the time of arrest, while using Booked as a dependent variable.

The logistics regression model we are using is :

- $log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{Concealed} + \beta_2 x_{Combative} + \epsilon$
- $log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{Resisted} + \beta_2 x_{Assaulted} + \epsilon$

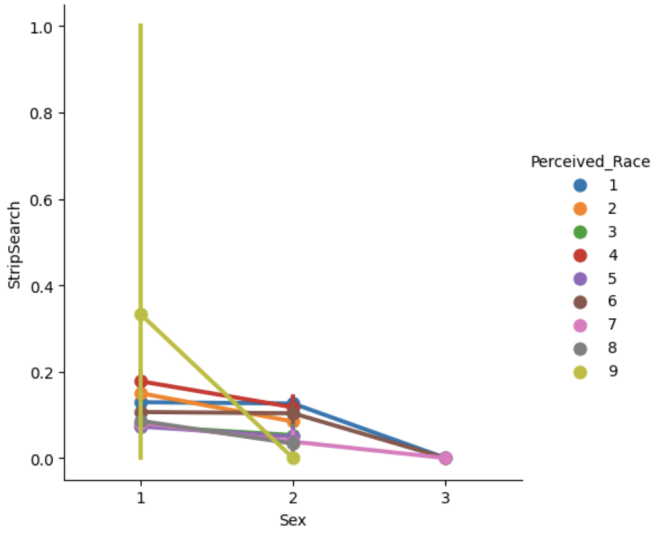
The interpretation and example of odds ratio will be present in the result section.

Result

4.1 Interaction plots

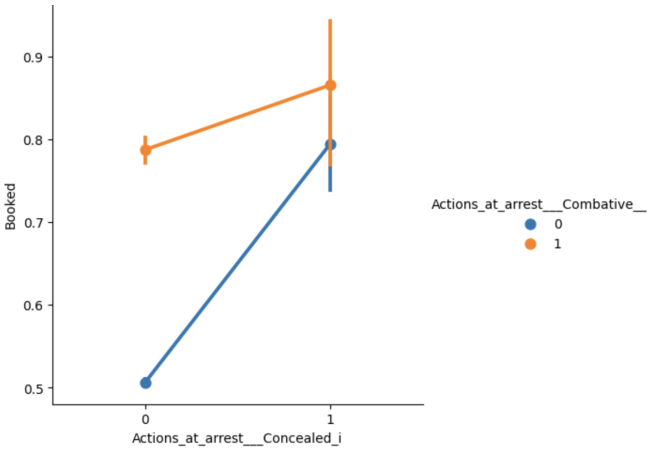
As shown in Figure 4, we can see how the relationship between the different variables changes. It is worth noting that the trend for all race variables is decreasing, which means that regardless of race, there are more males than females in terms of sex. This also includes a smaller fraction of unknown sexes. (1 = male, 2 = female, 3 = unknown on x-axis.) On the other hand, we found that except for the unknown race (9), the indigenous (4) and Black races had higher records of strip searches.

Figure 4: The Interaction Plot of Strip Search and Sex with Perceived Race.



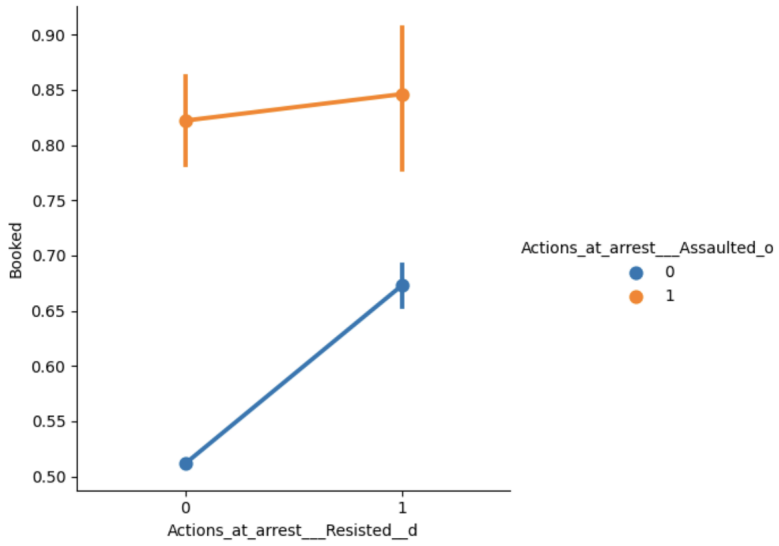
As shown in Figure 5, We found that when actions at the time of arrest were concealed, they usually also carried stronger combative actions. This indicates that the chance of being booked increases in this arrested condition. (0 = No, 1 = Yes)

Figure 5: The Interaction Plot of Booked and Actions at Arrest (Concealed) with Action at Arrest (Combative)



As shown in Figure 6, we found that arrests with assaulted issues had a high record of being booked. We also observed that when the action at the time of arrest was resisted, it also usually carried a high level of assaulted issue. This shows that in this case, the arrest is usually booked. (0 = No, 1 = Yes)

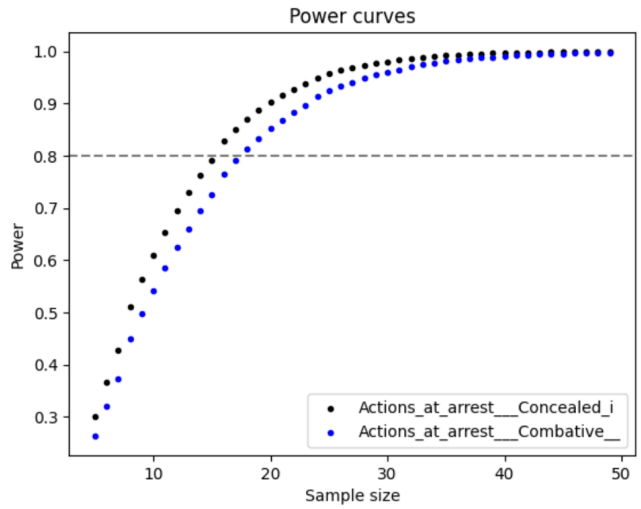
Figure 6: The Interaction Plot of Booked and Actions at Arrest (Resisted) with Action at Arrest (Assaulted)



4.2 Power graph

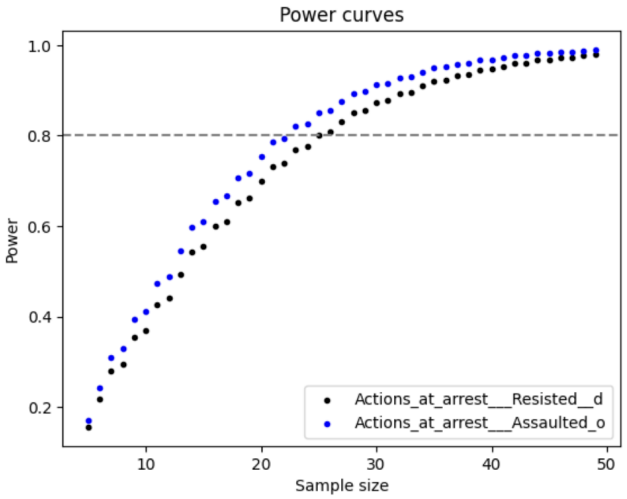
As shown in Figure 7, when the sample size of concealed and combative increase, the power of these two independent variables are also increases. This means that there is a higher probability of correctly detecting the significant effect.

Figure 7: The Power Curves of Concealed and Combative



As shown in Figure 8, when the sample size of resisted and assaulted increase, the power of these two independent variables are also increases. This means that there is a higher probability of correctly detecting the significant effect.

Figure 8: The Power Curves of Resisted and Assaulted



4.3 Logistic Regression Model

- $log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{Concealed} + \beta_2 x_{Combative} + \epsilon$

Table 11. Summary Table of the logistics regression model1 (Booked as outcome variable)

	Coef	Std.error	Z-value	Pr(>)|z|	[0.025 0.975]	Odds ratio
Concealed	1.279	0.173	7.381	0.000	0.940 1.619	3.595
Combative	1.271	0.051	25.020	0.000	1.171 1.371	3.564

- $log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_{Resisted} + \beta_2 x_{Assaulted} + \epsilon$

Table 12. Summary Table of the logistics regression model2 (Booked as outcome variable)

	Coef	Std.error	Z-value	Pr(>)|z|	[0.025 0.975]	Odds ratio
Resisted	0.690	0.048	14.262	0.000	0.596 0.786	1.995
Assaulted	1.435	0.148	9.720	0.000	1.145 1.724	4.198

4.3.1 Interpretation

As shown in Table 11:

The β_0 is the intercept of logistic regression model. The $\beta_1$is the coefficient of concealed of the logistic regression model, which is 1.279. The standard error of β_1 is 0.173. The null hypothesis, H_0 is $\beta_1 = 0$, while the alternative hypothesis, H_a is $\beta_1 \neq 0$. Since the p-value is 0, which is smaller than 0.05 and 0 is not lie on the confident interval, there is a very strong evidence against the null hypothesis. The odds ratio is about 3.595, meaning a positive association exists. This implies that as the concealed variable increases, the odds of the outcome also increase.

The $\beta_2$is the coefficient of concealed of the logistic regression model, which is 1.271. The standard error of β_2 is 0.051. The null hypothesis, H_0 is $\beta_2 = 0$, while the alternative hypothesis, H_a is $\beta_2 \neq 0$. Since the p-value is 0, which is smaller than 0.05 and 0 is not lie on the confident interval, there is a very strong evidence against the null hypothesis. The odds ratio is about 3.564, meaning a positive association exists. This implies that as the combative variable increases, the odds of the outcome also increase.

As shown in Table 12:

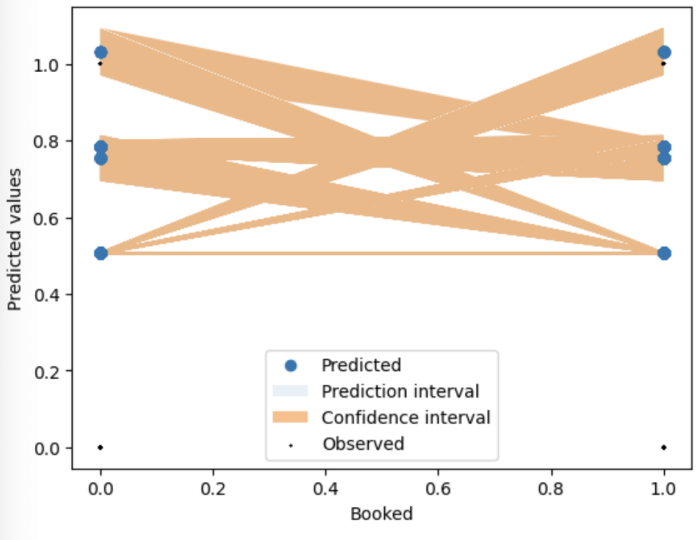
The β_0 is the intercept of logistic regression model. The $\beta_1$is the coefficient of concealed of the logistic regression model, which is 0.690. The standard error of β_1 is 0.048. The null hypothesis, H_0 is $\beta_1 = 0$, while the alternative hypothesis, H_a is $\beta_1 \neq 0$. Since the p-value is 0, which is smaller than 0.05 and 0 is not lie on the confident interval, there is a very strong evidence against the null hypothesis. The odds ratio is about 1.995, meaning a positive association exists. This implies that as the resisted variable increases, the odds of the outcome also increase.

The $\beta_2$is the coefficient of concealed of the logistic regression model, which is 4.198. The standard error of β_2 is 0.148. The null hypothesis, H_0 is $\beta_2 = 0$, while the alternative hypothesis, H_a is $\beta_2 \neq 0$. Since the p-value is 0, which is smaller than 0.05 and 0 is not lie on the confident interval, there is a very strong evidence against the null hypothesis. The odds ratio is about 3.564, meaning a positive association exists. This implies that as the assaulted variable increases, the odds of the outcome also increase.

4.4 Predicted Values and Confidence Interval Plot

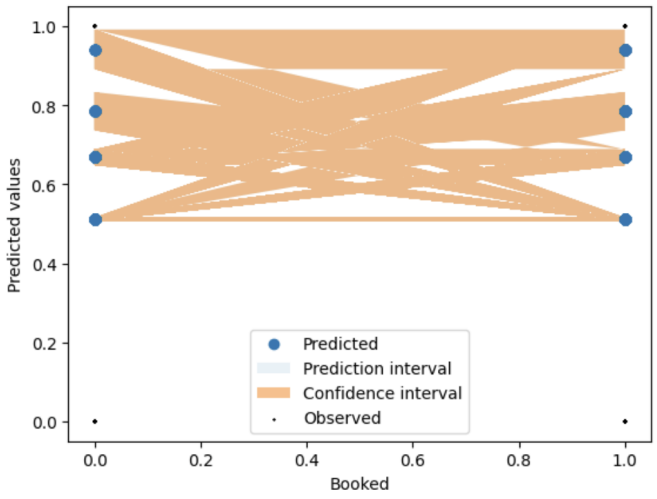
As shown in Figure 9, we can see the accuracy score is around 0.504 and 0 is not lie on any confidence intervals.

The Figure 9: The Plot of Booked and Predicted Values (Concealed and Combative)



As shown in Figure 10, we can see the accuracy score is around 0.498 and 0 is not lie on any confidence intervals.

The Figure 10: The Plot of Booked and Predicted Values (Resisted and Assaulted)



Discussion and Conclusion

Based on a dataset of strip searches and arrests provided by the Toronto Police Service (Toronto Police Service, 2022), we want to analyze the conditions under which strip searches and booked are more likely to occur. This will allow us to reach more accurate conclusions to help the community or government improve their policies and reduce the negative social impact of unnecessary searches that may occur. In our study, we explore the demographics (sex and race) of arrestees to determine whether they are likely to be strip searched. As well as by studying the actions taken at the time of the arrest to determine if they would be booked. However, some limitations may be present in the data set that may make our results slightly biased.

With the Anova Test and Tukey's HSD, we confirmed the selected variables were statistically significant among each other on the condition that they were analyzed to improve the accuracy of the results. The obtained analysis also demonstrated the existence of meaningful differences and relationships between the variables.

With the Ancova test, power analysis and logistic regression model, we also confirmed that there is a statistically significant relationship between the selected independent variables and the predictor and indicated that the probability of correctly detecting a significant effect is very high.

Based on these results, we expect to obtain the features and reasons for the population being strip searched and booked. Then make suggestions for improving the process. For instance, by analyzing the results, we suggest establishing specific search procedures for the racial groups that occur with high frequency, reducing the unfair treatment that may arise while also improving the accuracy of the searches. This is because responsible and ethical search procedures are the only way to minimize bias and unintended consequences.

5.1 Limitation

Fristly, there are some information about race and sex that shown as unknown, which may result in building models that are not good enough for their degree.

Secondly, the records collected may lead to slightly biased results due to certain issues with the booked templates in the dataset.

Lastly, the sample size of the variables may not be large enough to affect the relationship between the sample size and the power to detect the effect size.

5.2 Conclusion

Strip searches and arrests have always happened every day in our society. The reasons for these arrests and the demographic characteristics of the people involved have also been the focus of governmental and social attention. In this study, we applied a dataset provided by the Toronto Police Service, which included 6,5276 records, to analyze the influence of the demographics of arrestees on strip searches and actions on being booked at the time of arrest, respectively. Firstly, we performed an EDA on the data to obtain the distribution and correlation between the different variables in the dataset. After that, we performed t-tests, ANOVA tests, and Tukey's HSD. In this process, we show that there was a difference between each group's means. And confirm that they were statistically significant with each other. Then, the ANCOVA test and power analysis demonstrated the relationship between the variables and the probability of correctly detecting a significant effect. In other words, these

tests were effectively used to assess the probability of detecting a significant effect in relation to the variables under consideration. Lastly, the most intuitive relationships between the data were shown through interaction plots and the relationship between our chosen independent variable (actions at arrest) and the predictor (Booked) was also confirmed to be statistically significant by the power curve and logistic regression model. we can conclude that there is a relationship between demographic attributes (such as race and sex) and the likelihood of being strip searched during an arrest. Specifically, more males than females are recorded as being strip searched, and White, Black, and Indigenous individuals have the highest proportion of being strip searched. Additionally, the research suggests that the likelihood of being booked is influenced by the action taken during arrest, including situations where the person may not be strip searched. Overall, the research provides evidence that demographic attributes and arrestee behavior can impact the likelihood of being strip searched or booked during an arrest.

Rerefence

Arrests and strip searches (RBDC-arr-TBL-001). (2022, November 10). Retrieved February 25, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Barrison Law. (2021, April 06). Strip searches in Ontario are occurring too often. Retrieved February 25, 2023, from https://criminallawoshawa.com/strip-searches-in-ontario-are-occurring-too-often/

Caetano, S. (2020). Introduction to Logistic Regression. Lecture.

Derek Willis, E. (2020, July 26). Mark Dibenedetto - the NYPD files. Retrieved February 25, 2023, from https://projects.propublica.org/nypd-ccrb/officer/19606-mark-dibenedetto

McNeilly, G. (2019, March). Breaking the golden rule a review of police - office of the independent ... Retrieved February 25, 2023, from http://www.oiprd.on.ca/wp-content/uploads/OIPRD_Breaking-the-Golden-Rule_Report.pdf