

# **Possible Racism Issue in Rate of Being Strip Searched**

Group 12

Anning Wang, Jiani Gu

Master of Information

INF2178 Experimental Design for Data Science

Prof. Shion Guha

Apr. 15, 2023

Link of Database: <https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about>

## **Table of Contents**

1. Introduction	3
1.1 Objective	3
1.2 Research Questions	4
2. Literature Review	4
3. Exploratory Data Analysis (EDA)	5
3.1 Descriptive statistic	5
3.2 Power Analysis	7
3.3 T-tests	9
4. Method	13
4.1 Dataset description	13
4.2 Methodology	13
5. ANOVA tests	14
5.1 One-way ANOVA	14
5.2 Two-way ANOVA	15
5.3 Post-hoc tests – Tukey’s HSD	16
6. ANCOVA Test	16
7. Logistic Regression	17
8. Results and Findings	18
9. Discussion	19
10. Conclusion	20
Reference & Appendix	21

## **1. Introduction**

### **1.1 Objective**

The definition of strip search is “the process of removing the clothes of a prisoner, or someone thought to have committed a crime, by a police officer or government official in order to find any illegal things, such as drugs” (2022). Strip search is not a necessary part of the police arrest or interrogation process, and because of the specific nature of this search, it is always more controversial in terms of public opinion. In 2020, a 15-year-old black schoolgirl was strip-searched by Metropolitan Police officers at the girl’s secondary school in Hackney, because a teacher mistakenly thought she was carrying marijuana. No other adults were present during the entire strip search, and the girl's parents were not informed. According to the search review, her intimate body parts were exposed, and she was made to take off her sanitary towel. Finally, no drugs were found (BBC, 2022). The incident was extremely traumatic for the girl's mental health. It was perceived by the public as racist and child abuse. The data states that there were 25 strip-searches of under-18s in the same borough in 2020-2021, and only two under-18s searches were white (BBC, 2022). Therefore, the various discussions about race, underage and strip searches that have arisen from this incident have aroused our interest and concern about the investigation of whether there is racism in strip searches and the association with arrest locations in Toronto.

## 1.2 Research Questions

Our study will seek to examine how the race of suspects and the arrest location affect strip searches during the arrest. We propose to investigate the below research questions for our project.

RQ 1: What is the rate of strip searches on different races of suspects in arrests in Toronto?

RQ 2: How does arrest location affect the likelihood of being strip searched for different races of suspects?

We formed these research questions based on information gained from our literature review and exploratory data analysis (please see the below section).

## 2. Literature Review

In 2022, the Toronto Police Service released a report on arrest records and strip search data, and this was a race- and identity-based strip search. The data shows that in 2020, one in three people strip searched will be black, even though the black population makes up only 10% of the city's population (CBC News, 2022). One third of all Native Americans arrested have been strip searched (CBC News, 2022). This report shows the potential for racism in Toronto police enforcement, and the public is unhappy with the results shown in this report. After Golden case (2001) which is a landmark case on strip searches, the Supreme Court of Canada acknowledged the basic intrusiveness of strip searches, and also acknowledged that strip searches “‘represent a significant invasion of privacy’ and are often a ‘humiliating, degrading

and traumatic experience.” (Lemke, 2022). “The Supreme Court also recognized that Black and Indigenous people suffer disproportionate harm due to the racial trauma associated with being strip searched” (Lemke, 2022). All of this evidence demonstrates that strip searches should be used judiciously. The current reports and data show that the police are highly racist in their selection of subjects for strip searches. Therefore, this is a topic that deserves attention and investigation.

### **3. Exploratory Data Analysis (EDA)**

#### **3.1 Descriptive statistic**

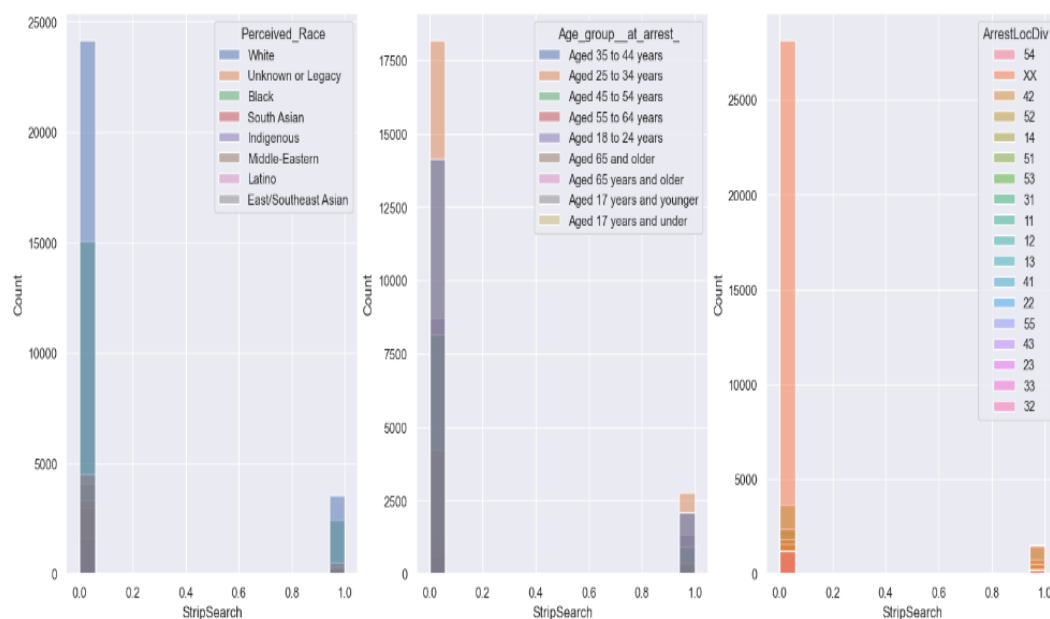
The dataset *Arrests and Strip Searches* contains 65,276 records (rows) and 25 columns. The entire dataset spans from January 2020 to December 2021. Since most columns in the entire dataset are categorical data with no numerical meaning, the dataset cannot directly summarize the valid mean, max, min, standard deviation, and other basic information.

Among all the arrested records, the Age groups are divided into 9 groups (actually 7 groups, because there are two groups with different titles but can be merged with the other two groups), among which the group of aged 25 to 34 years is the most frequently recorded with 20949 records, in addition, there are 3042 records are youth which is 17 years and under. In terms of gender, the records are only divided into three groups: male, female and unknown, with more records for males, 52,650 records in total and accounting for about 81%. In terms of perceived race, there are 8 categories of different races which include one category for unknown or legacy. This

is the largest proportion of whites, with a total of 27,723 records, accounting for about 42%, while blacks ranked second, accounting for about 27 percent.

For a more comprehensive understanding of the dataset, we created barplots of the strip search situation based on three groupings of perceived races, age groups, and arrest locations (please check Figure 1 below).

Figure 1. Stacked column chart of distribution of strip-search based on perceived races, age groups and arrest locations

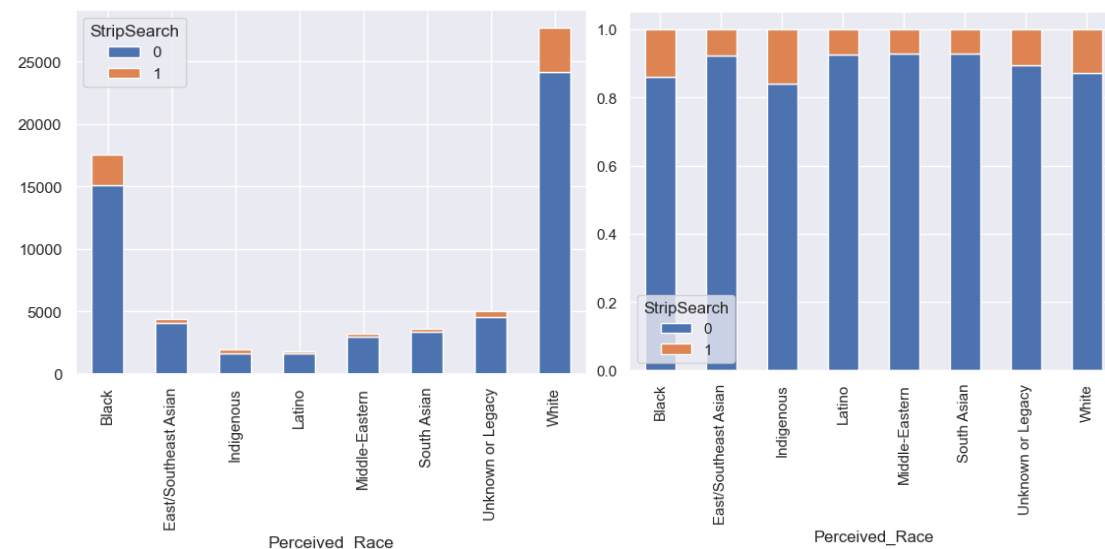


We also used a two-way frequency table to show the frequency of strip searches of arrestees with different races (see Figure.2) and a bar chart to show the statistics of strip search more visually. As Figure 3 shown below, the plot on the left shows the total statistics, while the plot on the right shows the rate of strip searches based on different races. These table and plots can refer to the RQ1 to represent the rate of strip search on different races of suspects in arrest.

Figure 2: Two-way frequency table of strip searches situation

Race	No strip search	Strip search
Black	15084	2434
East/Southeast Asian	4071	341
Indigenous	1626	306
Latino	1636	132
Middle eastern	3009	228
South Asian	3356	257
Unknown/Legacy	4519	536
White	24147	3566

Figure 3: Stacked column chart of frequency and rate of strip search situation

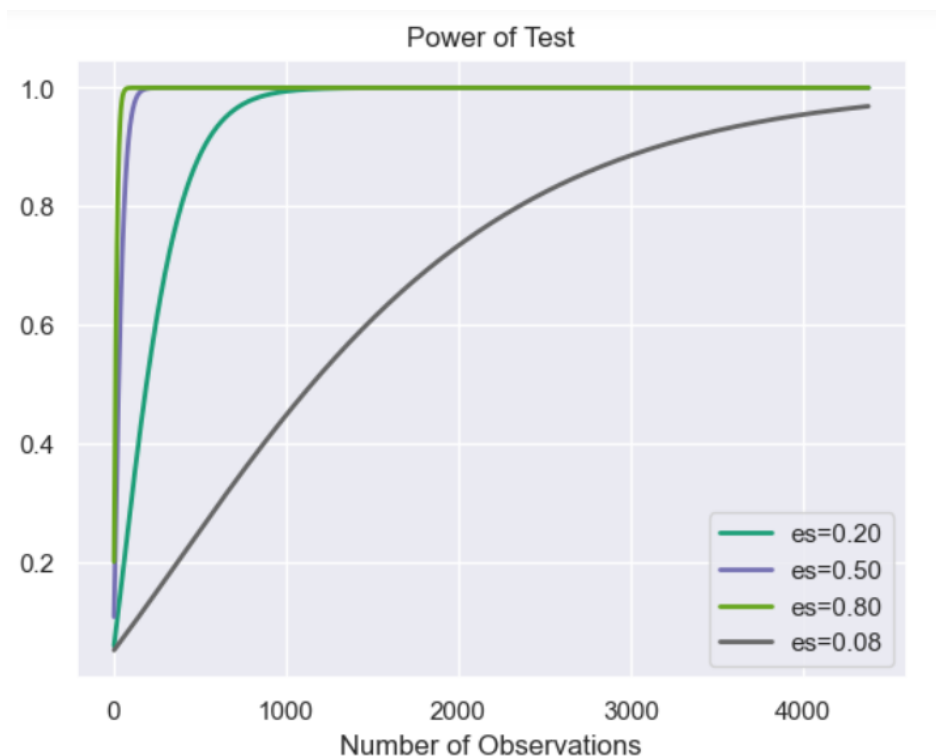


### 3.2 Power Analysis

The purpose of power analysis is to calculate the estimation smallest sample size needed for an experiment, given the required significance level, statistical power and effect size (TIBCO, 2023). Since a statistical power of 80% (0.8) is the minimum threshold for the test result to be accepted in many fields of research, we choose to calculate the smallest sample size that we needed based on the requirement of achieve a statistical power of 0.8. This power analysis is to estimate the sample size before the t-test to analyze whether the arrest location effect on the likelihood of being strip searched differed between black group and non-black group.

First, we used Cohen's metric to calculate the effect size of the explanatory variable, the result is 0.08. Then we use the effect size 0.08 and the statistical power of 0.8 to estimate the smallest sample size we needed. The result indicated that the black group sample size required is 4378, and the non-black group sample size required is 1607. The actual sample size for black group is 17526 and for non-black group is 47746. Our sample size is significantly larger than the minimum sample size, which means our t-test result is reliable.

Figure 4: Power of Test for stipe search vs black and non-black groups



This graph is the comparison about our effect size and the three levels of effect size. Power level of 0.1 to 0.3 is considered a small effect size, 0.3 to 0.5 is the moderate effect size and larger than 0.5 is considered the large difference effect. Our effect size is 0.08 which is smaller than 0.1, which conclude as the trivial effect size.



### 3.3 T-tests

To further explore the relationship between the categories in the dataset, several t-tests were conducted on the categorical attributes in the dataset. The following content will summarize and analyze each of these t-tests.

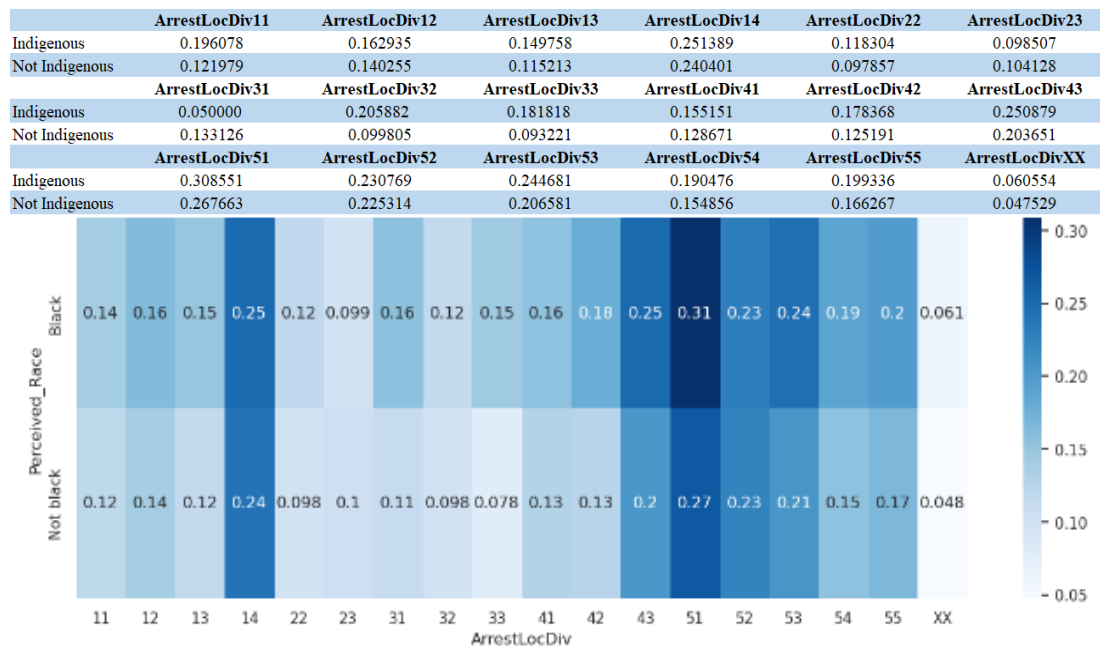
#### **Arrest location and Black group**

H0 (null hypothesis): Arrest location have no effect on the likelihood of being strip searched for both black and non-black group.

H1: Arrest location have effects on the likelihood of being strip searched for both black and non-black group.

For this t-test, we categorized multiple races into black group and non-black group to test the existence of black-specific effects. Then we created the cross-tabulation table and plot to check the distribution situation (see Figure 4 below). The p-value of this t-test is 0.895 which is much higher than the significance level of 0.05. Therefore, it is not statistically significant, and it should not reject the null hypothesis. It means we do not have enough evidence to conclude that the arrest location and black/non-black group are associated.

Figure 5: Cross-tabulation table and plot of arrest location and black group



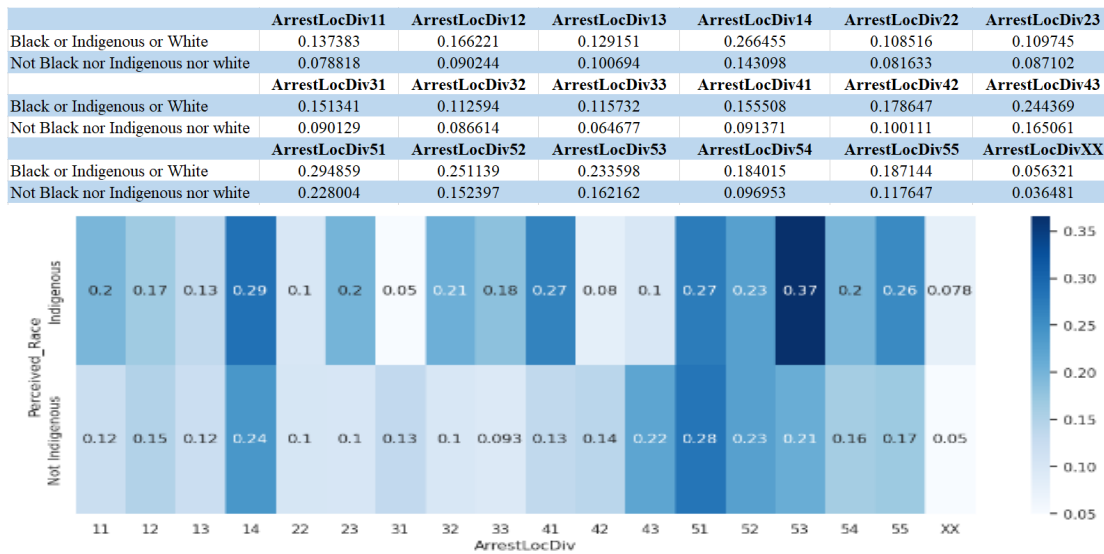
### Arrest location and indigenous group

H0 (null hypothesis): Arrest location have no effect on the likelihood of being strip searched for both indigenous and non-indigenous group.

H1: Arrest location have effects on the likelihood of being strip searched for both indigenous and non-indigenous group.

For arrest location and indigenous group, we categorized multiple races into indigenous group and non-indigenous group to test the existence of indigenous-specific effects. We also created the cross-tabulation table and plot to check the distribution situation (see Figure 5 below). The p-value of this t-test is 0.18 which is higher than the significance level of 0.05. Therefore, it is not statistically significant, and it should not reject the null hypothesis. It means we do not have enough evidence to conclude that the arrest location and indigenous/non-indigenous group are associated.

Figure 6: Cross-tabulation table and plot of arrest location and indigenous group



### Arrest location and white group

H0 (null hypothesis): Arrest location have no effect on the likelihood of being strip searched for both white and non-white group.

H1: Arrest location have effects on the likelihood of being strip searched for both white and non-white group.

For arrest location and white group, we repeated the steps as before. The p-value of this t-test is 0.25 which is higher than the significance level of 0.05. Therefore, it is not statistically significant, and it should not reject the null hypothesis. It means we do not have enough evidence to conclude that the arrest location and white/non-white group are associated.

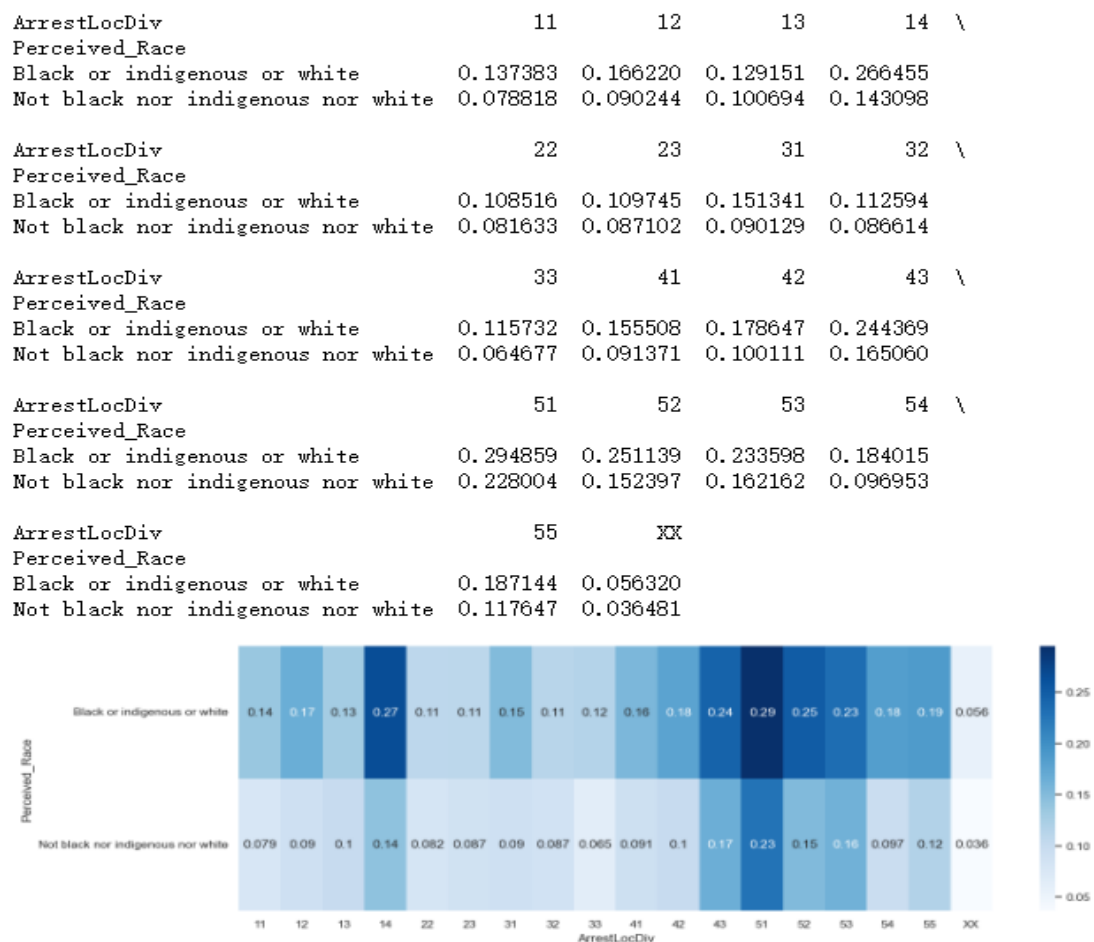
### Arrest location and both black, indigenous & white group

H0 (null hypothesis): Arrest location have no effect on the likelihood of being strip searched for both black or indigenous or white group and not black nor indigenous nor white group.

H1: Arrest location have effects on the likelihood of being strip searched for both black or indigenous or white group and not black nor indigenous nor white group.

For arrest location and black or indigenous or white group, we repeated the steps as before, the corresponding cross-tabulation table and plot shown below (Figure 6). The p-value of this t-test is 0.002 which is lower than the significance level of 0.05. Therefore, it is statistically significant, and the null hypothesis should be rejected. It means we have enough evidence to conclude that the arrest location and black or indigenous or white group are associated.

Figure 7: Cross-tabulation table and plot of arrest location and black or indigenous or white group



## **Gender and indigenous group**

H0 (null hypothesis): Gender has no effect on the likelihood of being strip searched for both indigenous and non-indigenous group.

H1: Gender have effects on the likelihood of being strip searched for both indigenous and non-indigenous group.

For gender and indigenous group, we repeated the steps as before. The p-value of this t-test is 38 which is much higher than the significance level of 0.05. Therefore, it is not statistically significant, and it should not reject the null hypothesis. It means we do not have enough evidence to conclude that the gender and indigenous/non-indigenous group are associated.

## **4. Method**

### **4.1 Dataset description**

The dataset of *Arrest and Strip Searches* was sourced from Public Safety Data Portal on Toronto Police Service. This dataset includes information related to all arrests and strip searches from January 2020 to December 2021, which contained 65276 records in total. This dataset is used within the police department to identify police officer behavior norms and the presence of human rights issues such as racism and is also available to the general public for monitoring and alerting purposes.

### **4.2 Methodology**

From the literature review and t-tests results above, a significant association between black, indigenous & white group, other races group and arrest location is

revealed. For further investigation, the one-way ANOVA, two-way ANOVA and post-hoc test are conducted to determine the independence and differences between means selected dependent and independent variables. For the deeper investigation, we also conducted the ANCOVA Test and logistic regression for our analysis.

## **5. ANOVA Tests**

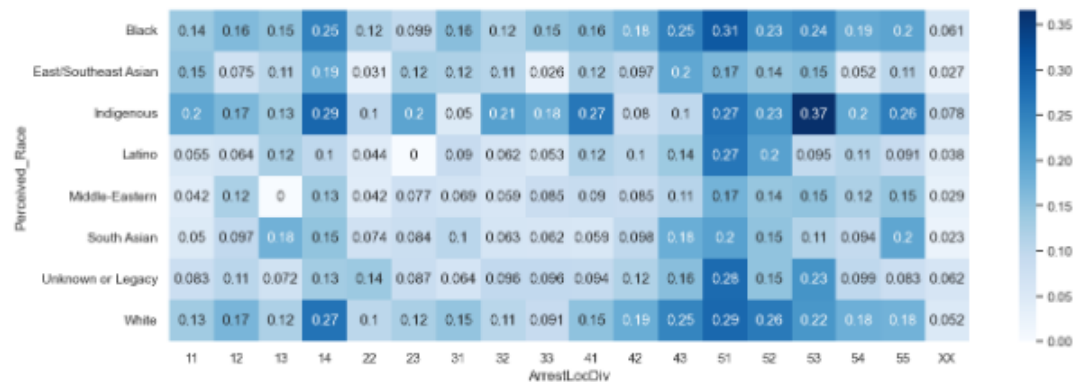
### **5.1 One-way ANOVA**

We use the statistical test Analysis of Variance to compare multiple group's means. We are trying to find if there is a significant difference in 8 different races group (which contain one unknown group) and arrest location. For the null hypothesis (H0), the mean strip search rate for all categories of independent variables (different races) is equal. For the alternative hypothesis (H1), the mean strip search rate for at least one category of independent variables is significantly different from the mean values for the other categories of independent variables.

The p-value of one-way ANOVA test is 0.895 which is much higher than the significance level of 0.05. It means the null hypothesis should not be rejected. We conclude that the mean strip search rate for all categories of independent variables has no significantly different.

The figure 7 shown below is the cross-tabulation plot of the strip search rate for all different races group and arrest location.

Figure 7: Cross-tabulation plot of all different races group and arrest location



## 5.2 Two-way ANOVA

We conduct two-way ANOVA to check whether there is a statistically significant difference between the mean value of many groups (At least 3 groups) that has been divided into factor "Race" and factor "Arrest Location". The primary aim of conducting a two-way ANOVA is to determine the impact of "Race" and "Arrest Location" on a response variable and to ascertain whether a relationship exists between these factors and the response variable. To better run two-way ANOVA, we created a "pandas" DataFrame of three variables: "Race", "Location" and "strip rate". Then we ran the Statsmodels library provides us with the `anova_lm()` function to perform the two-way ANOVA. The result of the operation is as follows:

	df	sum_sq	mean_sq	F	PR(>F)
C(Race)	1.0	0.001687	0.001687	0.763161	0.397079
C(Location)	1.0	0.000196	0.000196	0.088884	0.769980
C(Race):C(Location)	1.0	0.000013	0.000013	0.006000	0.939355
Residual	14.0	0.030941	0.002211	NaN	NaN

Following are the p-values for each of the factors in the output:

- The race p-value is equal to 0.397079
- The location p-value is equal to 0.769980
- The race \* location: p-value is equal to 0.939355

The p-values for race and arrest location are more significant than 0.05, implying that the means of both factors don't have a statistically significant effect on the strip search rate. The p-value for the interaction effect (0.939355) is more significant than 0.05, which shows no significant interaction effect between race and arrest location.

### 5.3 Post-hoc test – Tukey’s HSD

After ANOVA test, Tukey’s HSD test (Honestly Significant Difference) was conducted. The purpose of Tukey’s HSD test is to focus on the largest value of the difference between two group means (Zaiontz, 2023). The result of this test is as follows:

Multiple Comparison of Means \_ Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
a	b	-0.0541	0.3844	-0.1591	0.051	FALSE
a	c	0.0126	0.9455	-0.0925	0.01174	FALSE
b	c	0.0667	0.2475	-0.0384	0,1717	FALSE

The group a represent Races in the dataset, group b represents Arrest Location and group c represents strip search rate. According to the above table, there is no significant difference between the means of these three groups, which the null hypothesis should not be reject. We do not have enough evidence to conclude that these groups are associated.

## 6. ANCOVA Test

We conducted Analysis of Covariance (ANCOVA) test for analyzing the differences between 3 or more group means while controlling for the effects of at least one continuous covariate (Frost, 2023). The dependent variable is the rate of strip



search, the independent variable is the perceived races and the covariate is the arrest location in the dataset.

To better run ANCOVA test, we created a "pandas" DataFrame of three variables: "Race", "Location" and "strip rate". Then we ran the Statsmodels library provides us with the ANCOVA function to perform the one-way ANCOVA. The result of the operation is shown in the appendix.

According to the results table, the Prob (F-statistic) is 0.185 which is greater than 0.05. There is no evidence of relationship between any perceived races with the rate of strip search. The p-value of the covariate "Location" is 0.108 which is also greater than 0.05. There do not have the significant relationship between arrest location and the strip search rate.

## **7. Logistic Regression**

Logistic regression estimates the probability of an event occurring based on the independent variables (IBM, 2023). We conducted the odds ratio logistic regression on our research. We performed two sets of odds ratio logistic regressions for our study, and the two respective independent variables were perceived races and arrest location.

Prior to computing the logistic regression, we created a "pandas" DataFrame of three variables: "Race", "Location" and "strip rate". Then we ran the Statsmodels library provides us with the "smf.logit" function to perform the logistic regression.

The results of the operation is shown in the Appendix.

As the result table shown, the coefficient value of factor “race” is 0.1719, which means the not indigenous group increase the rate of strip search probability to about 54% of the case where the perceived race is the indigenous group. But the p-value of logistic regression between perceived races and the rate of strip search is 0.906 which is much higher than 0.05. There has no evidence that this set of regression is valid.

For the set of arrest location regression. The result table shown that the p-value of this set test is 0.827, which is much higher than 0.05. There do not have enough evidence that arrest location has effect on the rate of strip search.

## **8. Result and Findings**

### **Research Question 1:**

For the rate of strip searches in arrests, the total rate is about 11.9% without considering any factors. Based on races, the rate in white group is about 12.8%, and the white group who had strip-searched in total arrest records is about 5.5%. The strip search rate in black group is about 13.9%, and the black group who had strip-searched in total records is about 3.7%. For the indigenous group, the strip search rate is about 18.8% which is the highest value in different races. The strip search rates for the remaining races were more evenly split, with East/Southeast Asian 7.7%; Latino 7.5%; Middle-Eastern 7%; South Asian 7.1%.

From the above probability data, we can see that the three races with a higher rate of strip searches are indigenous, black and white, and among them, the public may

pay more attention to appealing to the rights of black people, while ignoring the fact that the indigenous group may also be targeted.

#### Research Question 2:

According to the result in t-tests (hypothesis 4), we conclude that the arrest location and black or indigenous or white group are associated. Especially in arrest location 14 and 52, the cross-tabulation plot (Figure 6) shows that the rate differences are large in these two locations. However, according to the result of ANOVA tests and Tukey's HSD test, we do not have enough evidence to proof that arrest locations and races are associated in strip search rate. It means for the period 2020-2021, there is no more obvious data to prove that the Toronto Police Service is targeting a specific race in terms of strip searches. However, there is a higher rate of searches for Blacks and Indigenous, who are a smaller population base in the city of Toronto.

For the addition test we conducted, ANCOVA test and Logistic Regression, all the tests' result reflected that we do not have enough evidence to proof that arrest locations and races are associated in strip search rate. This can conclude that the previous findings are still effective.

## **9. Discussion**

### Limitation & Future works

To prevent discriminatory design, a good data analysis requires the use of fair and scientifically representative datasets. As Ruha Benjamin (2019) noted, discriminatory designs reinforce social inequalities by explicitly amplifying racial hierarchies or by

ignoring social divisions. In our selected dataset, there may be a data imbalance problem. We noticed that the number of statistics for each race differed, leading to potentially biased results. For instance, the experiment may become too focused on black and white populations, ignoring other racial groups that require analysis. Therefore, in future studies, we should aim to select datasets that are similar or identical and have a uniform representation of all groups to ensure the accuracy of our analysis.

## **10. Conclusion**

In this report, we are primarily interested in investigating whether the Toronto police were seriously racist in a strip search, a search that needs to be used with caution. The findings deny this doubt. This is good news. But there is no denying that Blacks and Indigenous, the smaller races in Toronto's urban population, still have high numbers in the rate of strip searches. Toronto police should keep a record of this dataset to confirm to the public that officials are really paying attention to this issue and to reassure the general public.

## Reference

- BBC News (2022). *Campaigners say police strip-search of black schoolgirl was 'sexual assault'*. BBC News. <https://www.bbc.com/news/uk-england-london-60769462>
- BBC News (2022). *Met Police apologises for strip-search of Hackney schoolgirl*. BBC News. <https://www.bbc.com/news/uk-england-london-60757031>
- Benjamin, R. (2019). *Introduction: The New Jim Code, In Race After Technology*. Cambridge, Medford MA: Polity Press, 1-48.
- CBC News (2022). *'We do not accept your apology,' activist tells Toronto's police chief after race-based data released*. CBC. <https://www.cbc.ca/news/canada/toronto/toronto-police-race-based-data-use-force-strip-searches-1.6489151>
- Frost, J. (2023). *ANCOVA: Uses, Assumptions & Example*. Statistics By Jim. <https://statisticsbyjim.com/anova/ancova/>
- Lemke, M. (2022). *Strip searches are ineffective, unnecessary and target racialized Canadians*. The Conversation. <https://theconversation.com/strip-searches-are-ineffective-unnecessary-and-target-racialized-canadians-185187>
- Meera (2023). *Power Analysis, Statistical Significance & Effect Size*. Meera. <https://meera.seas.umich.edu/power-analysis-statistical-significance-effect-size.html>
- R. v. Golden (2001). Supreme Court Judgments. <https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/1924/index.do>

*Strip Search* (2023). Cambridge Dictionary. <https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/1924/index.do>

TIBCO (2023). *What is Power Analysis?* TIBCO. <https://www.tibco.com/reference-center/what-is-power-analysis>

Zaiontz, C. (2023). *Tukey HSD (Honestly Significant Difference)*. Real Statistics. <https://real-statistics.com/one-way-analysis-of-variance-anova/unplanned-comparisons/tukey-hsd/>

## Appendix

### 1. ANCOVA result table

OLS Regression Results						
=====						
Dep. Variable:	rate	R-squared:	0.229			
Model:	OLS	Adj. R-squared:	0.110			
Method:	Least Squares	F-statistic:	1.925			
Date:	Sun, 16 Apr 2023	Prob (F-statistic):	0.185			
Time:	02:46:10	Log-Likelihood:	28.934			
No. Observations:	16	AIC:	-51.87			
Df Residuals:	13	BIC:	-49.55			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
Intercept	0.1963	0.031	6.250	0.000	0.128	0.264
Race[T.Not black]	-0.0205	0.022	-0.933	0.368	-0.068	0.027
Location	-0.0024	0.001	-1.726	0.108	-0.005	0.001
=====						
Omnibus:	9.885	Durbin-Watson:	2.643			
Prob(Omnibus):	0.007	Jarque-Bera (JB):	6.596			
Skew:	1.476	Prob(JB):	0.0370			
Kurtosis:	4.086	Cond. No.	63.8			
=====						

### 2. Logistic Regression result table

Optimization terminated successfully.

Current function value: 0.239615

Iterations 6

Logit Regression Results						
Dep. Variable:	rate	No. Observations:	16			
Model:	Logit	Df Residuals:	14			
Method:	MLE	Df Model:	1			
Date:	Tue, 11 Apr 2023	Pseudo R-squ.:	inf			
Time:	17:07:34	Log-Likelihood:	-3.8338			
converged:	True	LL-Null:	0.0000			
Covariance Type:	nonrobust	LLR p-value:	1.000			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.7409	0.992	-1.754	0.079	-3.686	0.204
Race[T.weekly]	-0.1719	1.449	-0.119	0.906	-3.012	2.668

Optimization terminated successfully.

Current function value: 0.239349

Iterations 6

Logit Regression Results						
Dep. Variable:	rate	No. Observations:	16			
Model:	Logit	Df Residuals:	14			
Method:	MLE	Df Model:	1			
Date:	Sun, 16 Apr 2023	Pseudo R-squ.:	inf			
Time:	03:00:57	Log-Likelihood:	-3.8296			
converged:	True	LL-Null:	0.0000			
Covariance Type:	nonrobust	LLR p-value:	1.000			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.4283	1.907	-0.749	0.454	-5.167	2.310
Location	-0.0205	0.094	-0.219	0.827	-0.204	0.163