

# **Experimental Data Analysis of Toronto Police Service Strip Search Rates**

Ragave Vicknarajah (1003959610)

Kailai Yang (1005536939)

INF2178: Experimental Design for Data Science

Dr. Shion Guha

February 28, 2023

## **Introduction**

Police have the power to conduct strip searches to ensure safety and security when they suspect a person may be in possession of a weapon or other contraband. They can also be used to keep illegal items like narcotics or weapons out of correctional facilities. (Babbar, 2021) However, concerns have been raised about the use of strip searches and the potential for abuse or infringement of civil liberties. They are a highly invasive procedure that can be traumatic and degrading to the person being searched. As a result, there is a lot of debate regarding whether or not to conduct strip searches, with concerns like privacy, human rights, and abuse potential being raised. Strip searches can be a serious violation of privacy, especially for vulnerable populations such as minors. To investigate the impartiality of Toronto police officers' requests for strip searches during stops of various minority groups, we will examine the demographics of those asked to strip search from arrest. We hypothesize that the Toronto police maintain consistent fairness in requesting strip searches for various minority suspects during the 2020-2021. We will use the dataset containing information related to all arrests and strip searches data publicly available on the Toronto Police Service Public Safety Data Portal during the 2020-2021 years for this project.

## **Literature Review**

A strip search is a sort of search in which a person's clothing is partially or completely removed in order to look for contraband or other objects. The use of strip searches is debatable since they may be viewed as an invasion of privacy and dignity, especially if they are performed in a way that is humiliating or public. (Babbar, 2021) This can be especially traumatic for vulnerable populations such as minors, the elderly, and survivors of sexual violence. (Psutka & Sheehy, 2016).

In addition, there is a growing awareness of systemic bias and discrimination in law enforcement and the criminal justice system. Racial and ethnic discrepancies can be found at every level of the criminal justice process, from arrest to sentence, according to a growing corpus of studies. For instance, research has shown that compared to white defendants, black and Hispanic defendants are more likely to face more serious charges and to be sentenced to lengthier terms of imprisonment. (Smith et al., 2020) Research shows that people of color are disproportionately subjected to strip searches, even when they are not suspected of carrying contraband or weapons. These searches can have a traumatic impact on individuals who have already experienced systemic discrimination and racial profiling. (Alexander, 2012)

Strip searches are often a topic of concern when it comes to gender issues because of the specific issues that arise when strip searches are conducted on people of different genders. For example, women may be more likely to experience trauma and emotional distress as a result of strip searches due to prior experiences of sexual violence and objectification. (Lemke, 2022) Because strip searches can be a traumatic and invasive experience, especially for minors who

may find greater physical and psychological effects. According to the report by the National Prison Rape Elimination Commission (NPREC) found that over half of the juvenile facilities surveyed reported conducting strip searches on all youth entering the facility, regardless of individualized suspicion. (2009) There is a need to strengthen protections for vulnerable populations, including clear guidance for minors on when and how to conduct strip searches. Therefore, police officers should be more cautious when making strip searches on suspects, especially when dealing with some vulnerable groups.

## **Dataset Description**

The dataset we used in this project shows information related to arrests and strip searches conducted by police officers, as well as indicators of whether a person was booked at a police station within 24 hours following the arrest event. It could be used to study patterns of arrest and strip search behavior among police officers, as well as the relationship between strip searches and booking rates. This dataset has 65,276 Records in which the basic information such as perceived race, gender, and age group of each captured person is recorded during the 2020-2021 years. The dataset - Arrests and Strip Searches (RBDC-ARR-TBL-001) - is available on Toronto Police Service Public Safety Data Portal and can be found through the following link: <https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about>. The dataset includes the following information: Arrest Year, Arrest Month, EventID, ArrestID, PersonID, Perceived Race, Sex, Age group (at arrest), Youth at arrest (under 18 years), ArrestLocDiv, StripSearch, Booked, Occurrence Category, Actions at Arrest - Concealed items, Actions at Arrest - Combative, violent or spitter/biter, actions at arrest - resisted, defensive or escape risk, Actions at arrest - Mental instability or possibly suicidal, Actions at arrest - Assaulted officer, Actions at arrest - Cooperative, SearchReason-CauseInjury, SearchReason-AssistEscape, SearchReason-PossessWeapons, SearchReason-PossessEvidence, ItemsFound. The strip search indicator is a binary variable indicating whether a person was subject to a search that involved the removal of some or all clothing and a visual inspection of the body. The booking status is a binary variable indicating whether a person was booked within 24 hours following the arrest event. Some records may indicate that a person was strip searched, but the data does not indicate a booking; in those cases, it is assumed that a booking took place. The age of the person arrested and/or strip searched is the age they provided to the arresting officer at the time of arrest.

## **Research Objective and Questions**

The objective of our project is to analyze the arrest and strip search dataset provided by the Toronto Police. Our focus is to examine whether individuals are treated equally when it comes to being asked to undergo a strip search, with a specific focus on variables such as race, age, sex, and youth status. Through our analysis, we aim to identify any disparities or biases that may exist in the strip search practices of the Toronto Police, and to provide recommendations for any necessary changes that could help to ensure that all individuals are treated fairly and with

respect during interactions with law enforcement. By shedding light on potential issues in strip search practices, our research aims to contribute to broader efforts towards promoting equity and justice within the criminal justice system.

We propose some research questions that can help guide our analysis of the Toronto Police arrest and strip search dataset to determine whether certain demographic groups are being treated differently when it comes to strip searches. Based on our literature review and preliminary analysis of the data set, we have identified the possibility of bias and differential treatment of minorities in the current justice system. In order to further explore this issue, we have formulated the following research questions.

- RQ1: Are individuals of certain races more likely to be subjected to strip searches by the Toronto Police based on the arrest year?
- RQ2: Is there a significant difference in the frequency of strip searches for male, females, and unknown in Toronto Police arrest cases?

We believe an in-depth study of this dataset will help us explore these research questions and help us understand the reasonableness of the Toronto police in asking suspects to strip search.

## **Descriptive Statistics**

### *Data cleaning*

The original dataset provided on Toronto Police Service Public Safety Data Portal was available in a CSV file which contains a total 65, 276 Records. The entire dataset includes only categorical variables and binary variables, and does not provide any numerical, which does not allow us to use the variables in the dataset directly as dependent variables for analysis using anova test. In order to run the ANOVA and t-test on the dataset it needs to be cleaned since there are columns with N/A values and no continuous variables. To begin we loaded the dataset into Google Collab and right away checked for any columns that contained N/A values. The columns that included N/A values were the following: ArrestID, Age\_group\_at arrest\_, Perceived\_Race, and Occurance\_Category. The rows that contained N/A values were dropped from the dataset. After cleaning the dataset we were left with 64,615 records and 10 columns of data.

In order to include a continuous variable we decided to transfer the categorical variables to numerical variables. Arrest year includes two levels, 2020 and 2021, which we transformed into 0 to represent 2020 and 1 to represent 2021. Perceived Race includes a total eight categories, which we transformed into 0 to represent Black, 1 to represent East/Southeast Asian, 2 to represent Indigenous, 3 to represent Latino, 4 to represent Middle-Eastern, 5 to represent South Asian, 6 to represent Unknown or Legacy and 7 to represent White. Due to the relatively small sample of non-white groups, we merged others other than white in what is collectively referred

to as Non-White. In the Non-White column 0 means the person is white and 1 represents non-white people. Sex includes 3 levels, Male, Female and Unknown, which we encoded Female to 0, Male to 1 and Unknown to 2. The original datasets contain 9 age groups, which they divided 17 years and younger and 65 years and older twice. We therefore combined all those under 17 years of age, and the same for those over 65 years of age. The new merged age group contains seven groups, which we transformed into 0 to represent Aged 17 and under, 1 to represent Aged 18 to 24 years, 2 to represent Aged 25 to 34 years, 3 to represent Aged 35 to 44 years, 4 to represent Aged 45 to 54 years, 5 to represent Aged 55 to 64 years, 6 to represent Aged 65 and older.

The final dataset after cleaning up has 10 columns of data: Arrest Year, Arrest Month, Perceived Race, Sex, Age group at arrest, youth at arrest under 18 years, Non white, StripSearch, Booked, Occurrence Category and Items Found. Since this dataset uses only binary variables and our research question focuses more on strip search, we counted the number of times strip search occurred on each group and named it Search count. Because the number of groups in the sample varies, the mere fact that search count is not sufficient to account for the fairness of police in asking suspects to strip search. We counted the proportion of strip searches for further analysis, and this data was named Search rate. Search rate is the proportion that the arrest people are asked to do the strip search in the total group. Search rate equals to  $\text{stripe\_saerch\_count} / (\text{stripe\_saerch\_count} + \text{not\_Strip\_search\_count})$ . The new merged dataframe after grouped includes the following columns: Arrest Year, Perceived Race, Sex, Age group at arrest, youth at arrest under 18 years, Non white, Search count and Search rate.

### *Data Summary*

The original dataset is grouped by the Arrest Year, Perceived Race, Sex, Age group at arrest, youth at arrest under 18 years and Non white columns. After grouped the dataset we were left with 230 records and 8 columns of data. From Figure 1, we can find the mean of these 230 search count records is 31.88 and mean of search rate is 7.28%. The standard deviation is 100.318 in search count and 9.36% in search rate. The max group has the 804 search count and 55.56% search rate.

	<b>Search_count</b>	<b>search_rate</b>
<b>count</b>	230.000000	230.000000
<b>mean</b>	31.873913	7.280802
<b>std</b>	100.318126	9.365414
<b>min</b>	0.000000	0.000000
<b>25%</b>	0.000000	0.000000
<b>50%</b>	3.000000	2.679115
<b>75%</b>	17.750000	13.596931
<b>max</b>	804.000000	55.555556

Fig.1. Dataset Summary

### *Race*

The perceived races included in the dataset are: White, Black, Unknown or Legacy, East/Southeast Asian, South Asian, Middle-Eastern, Indigenous, and Latino. According to the graph and the frequency counts (Figure 2) majority of the people included in the dataset were perceived as White compared to the other races included in the dataset. Figure 3 includes a count bar graph of the Perceived\_Race category from the dataset to show the counts of whether the person got strip search.

White	27407
Black	17352
Unknown or Legacy	5002
East/Southeast Asian	4388
South Asian	3594
Middle-Eastern	3213
Indigenous	1907
Latino	1752
Name: Perceived_Race, dtype: int64	

Fig.2. Strip Search Counts on Each Races

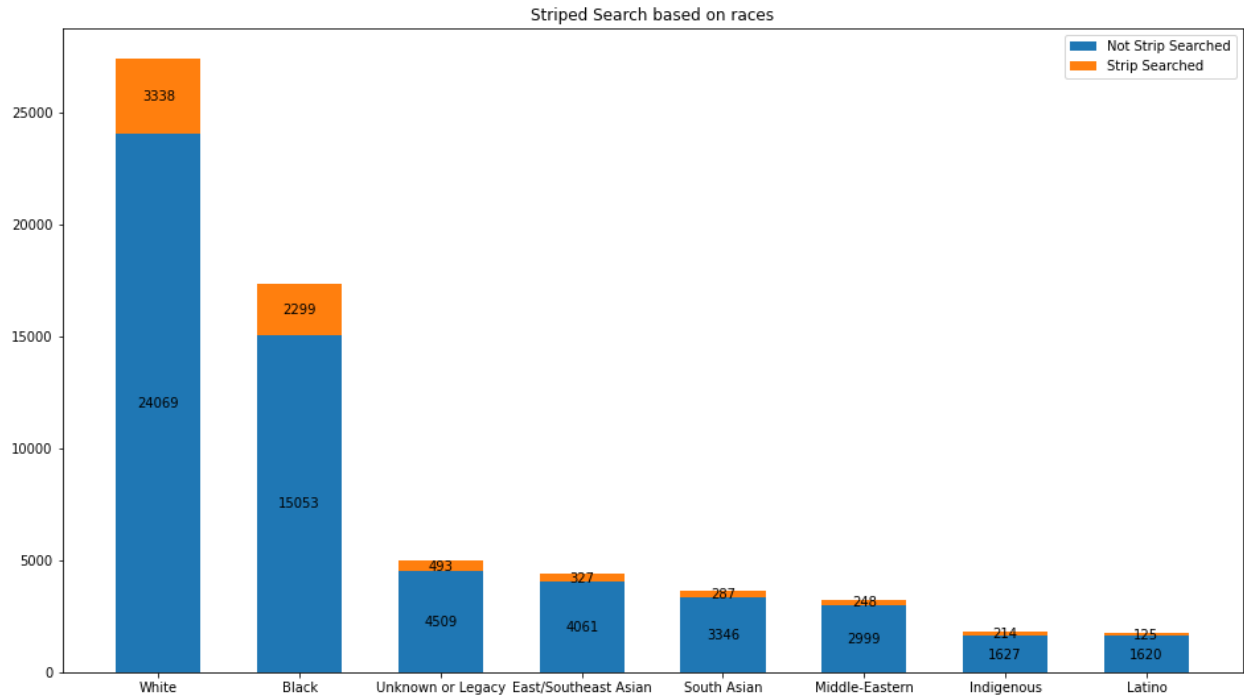


Fig.3. Strip Search Proportion on Each Races

### Gender

The gender included in the dataset are: Female, Male and Unknown. Figure 4 includes the total count number in each gender group. According to the counts and bar chart (Figure 5), we can find the proportion of not strip searched and strip searched in each gender group. 6,123 Males and 1,208 females are requested to strip search. In this cleaned dataset, it only contains 9 records about unknown gender and all of them were not searched.

```
M      52106
F      12500
U           9
Name: Sex, dtype: int64
```

Fig.4. Strip Search Counts on Each Gender Group

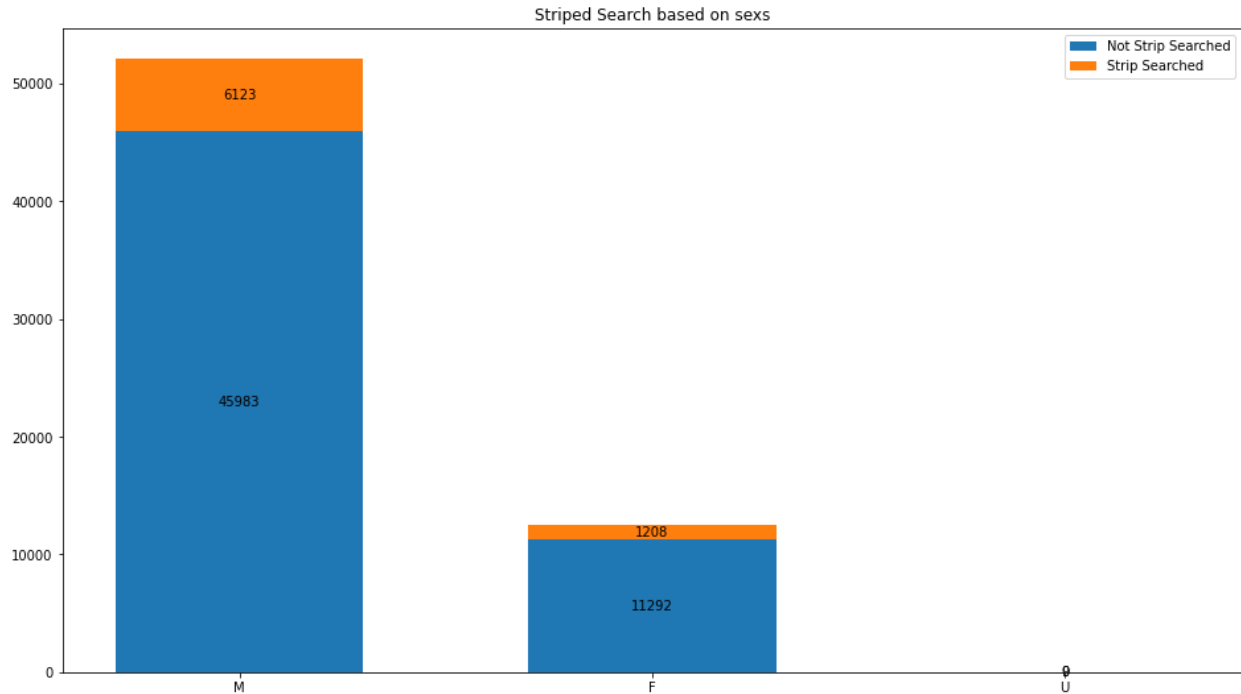


Fig.5. Strip Search Proportion on Each Gender Group

### *Correlation Plot*

The correlation heatmap (Figure 6) provides a visualization of the strength of relationships between the numerical variables (Kumar, 2022). Using the correlation heatmap we are able to have an idea of the strength of the relationship of the variables (Kumar, 2022). A value closer to one is considered a positive correlation between the two variables, a value closer to negative one is considered a negative correlation, and if the value is zero there is no correlation between the variables (Kumar, 2022).

According to Figure 6 there are no variables that have a strong positive correlation, meaning that there are no variable relationships close to one. The most notable positive relationships are between the perceived race of the individual and age group at arrest, 0.18. However, this relationship is closer to zero than one suggesting that there is no real correlation between the variables. There is a negative correlation between age group at arrest and youth at arrest under 18 years. However, the correlation does not have much significance since the variables only have a relationship because the data collected for those variables are relatively the same.



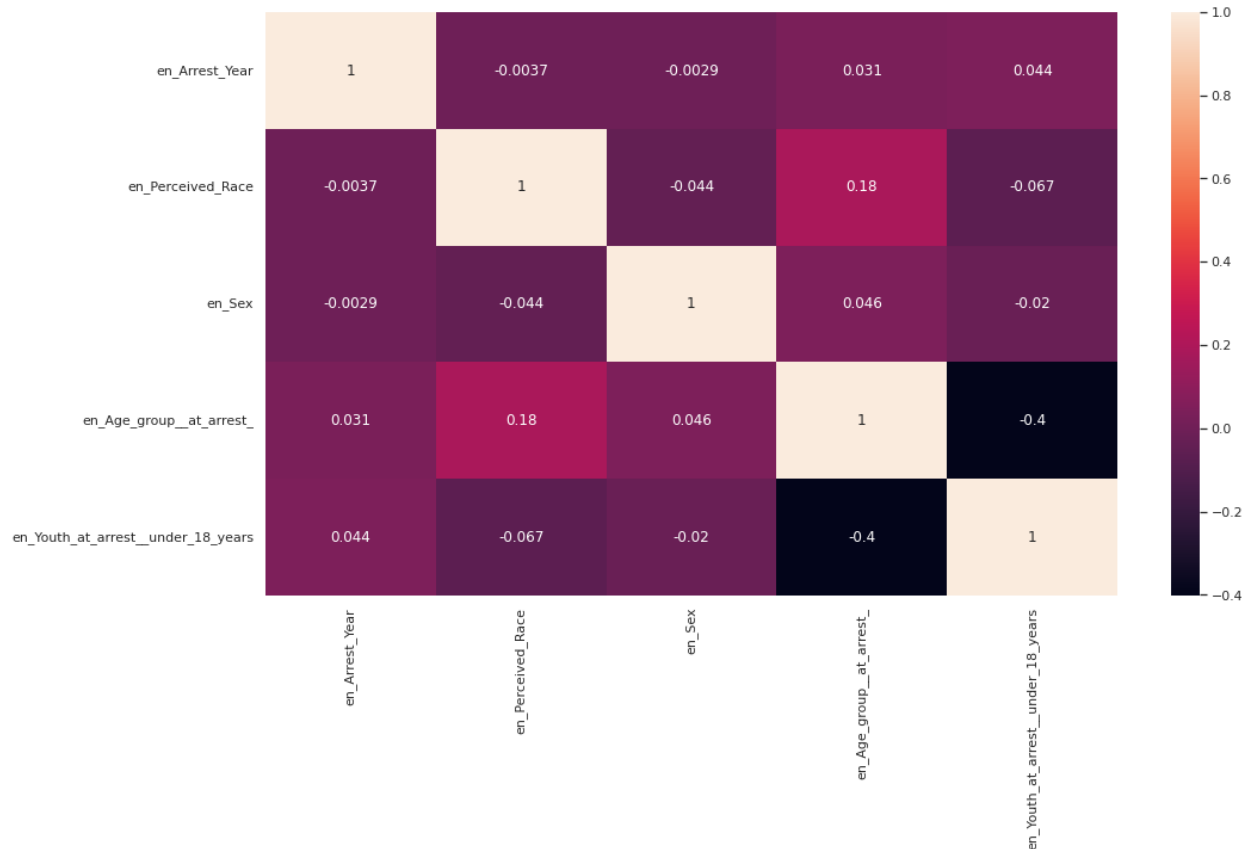


Fig.6. Correlation Plot

## T-Tests

A t-test is a statistical test that compares the means of two samples (Bevans, 2022). It is used to determine if there is a significant difference between the means of two groups and how they are related (Bevans, 2022). The test uses a null hypothesis that the difference in the group means is zero and an alternative hypothesis that the difference in group means is different from zero. (Bevans, 2022) For this paper six t-tests were carried out to examine the relationship between two selected variables of our choosing from the dataset (can be found in the results section).

- T-Test 1(p-value:  $0.469 > 0.05$ ):  
Hypothesis: The arrest white people and non white people has the same search rate  
Null Hypothesis ( $H_0$ ): Mean search rate white = Mean search rate non white  
Alternative Hypothesis ( $H_a$ ): Mean search rate white  $\neq$  Mean search rate non white
- T-Test 2 (p-value: F to M - 0.011; F to U - 0.055; M to U - 0.023):  
Hypothesis: Female and Male has the same search rate  
 $H_0$ : Mean search rate female = Mean search rate male = mean search rate unknown  
 $H_a$ : Mean search rate female  $\neq$  Mean search rate male  $\neq$  mean search rate unknown

- T-Test 3 (p-value:  $0.0847 > 0.05$ ):  
Hypothesis: white female and white male has the same search rate  
H0: Mean search rate white female = Mean search rate white male  
Ha: Mean search rate white female  $\neq$  Mean search rate white male
- T-Test 4 (p-value:  $0.761 > 0.05$ ):  
Hypothesis: adult and youth has the same search rate  
H0: Mean search rate adult = Mean search rate youth  
Ha: Mean search rate adult  $\neq$  Mean search rate youth
- T-Test 5 (p-value:  $0.653 > 0.05$ ):  
Hypothesis: white and black has the same search rate  
H0: Mean search rate white = Mean search rate black  
Ha: Mean search rate white  $\neq$  Mean search rate black
- T-Test 6 (p-value:  $5.289e-31 > 0.05$ ):  
Hypothesis: In 2020 and 2021 has the same search rate  
H0: Mean search rate white = Mean search rate black  
Ha: Mean search rate white  $\neq$  Mean search rate black

## ANOVA Tests

To answer the research questions for this paper both a one-way and two-way analysis of variance (ANOVA) test will be conducted. A one-way ANOVA is a statistical test used to determine whether there are any statistically significant differences between the means of three or more independent groups (Laerd Statistics, 2022). It is typically used when you have a single independent variable, or factor, and the goal is to investigate if variations, or different levels, of that factor have an effect on the dependent variable (Laerd Statistics, 2022). A two-way ANOVA is a statistical test used to estimate how the mean of a quantitative variable changes according to the levels of two categorical variables (Scribbr, 2022). It is an extension of the one-way ANOVA that examines the influence of two different categorical independent variables on a dependent variable (Scribbr, 2022).

Prior to conducting an ANOVA, there are a series of assumptions that must be met. The assumptions for a one-way ANOVA are: 1) the dependent variable should be measured at the continuous level (i.e., they are interval or ratio variables (Mackenzie, 2018). 2) The samples should be independent (Mackenzie, 2018). 3) The populations from which the samples are obtained must be normally distributed (Mackenzie, 2018). 4) Homogeneity of variance, meaning that the variance of each group is equal. For a two-way ANOVA, the assumptions are similar to those of a one-way ANOVA with an additional assumption that there is no interaction between the independent variables (Technology Networks 2018).

## Results

### *Research Question One*

Research question one is focused on whether individuals of certain races are more likely to be subjected to strip searches by the Toronto Police based on the arrest year. There are three null hypotheses as follows, 1) there is no difference in the means of perceived race on search rate, 2) there is no difference in means of year on search rate, and 3) there is no interaction between the effect of perceived race and year on search rate. The alternative hypothesis is for one and two the means are equal, and there is an interaction between the effect of perceived race and year on search rate.

Prior to running the two-way ANOVA tests were run in order to determine if the assumptions were fulfilled. In order to check the assumption of homogeneity of variance the Levene test was carried out. The p-value (0.00) was less than 0.05, therefore it is significant and we can conclude that the factors do not have equal variances. In order to verify the normality assumption a boxplot was created (Figure 7). According to the boxplot the distance of both whiskers on each of the boxes are not equal and there are many outliers, therefore we can conclude that the data is not normally distributed. As for independent observations, based on how the data was collected by the TPS we assume that there is no relationship between the observations within or between the groups. Although all assumptions were not fulfilled we will still carry out the two-way ANOVA.

Two t-tests were run prior to the two-way ANOVA. The t-test for white people against black had a result of a p-value greater than 0.05 (0.653), therefore we can conclude that the mean search rate is the same in these two groups. The t-test for the year 2020 and year 2021 had a very small p-value ( $5.28e-31$ ) less than 0.05, therefore we can conclude that the mean search rate differs between the two groups.

In order to answer research question two we decided to conduct a two-way ANOVA using the Perceived Race, Arrest Year, and Search Rate variables. There are two factors that will be evaluated: Perceived Race which has 8 levels and Year which has 2 levels. Using the two-way ANOVA we are able to evaluate how perceived race and year affects the search rate. Search rate is the proportion that the arrest people are asked to do the strip search in the total group. We count the total number that people are asked to do the strip search (strip\_search\_count) and total number that people are not searched for (not\_strip\_search\_count). Search rate equals to  $\text{strip\_search\_count} / (\text{strip\_search\_count} + \text{not\_Strip\_search\_count})$

From the two-way ANOVA three hypotheses will be tested: 1) effect of perceived race on search rate, 2) effect of year on search rate, and 3) effect of perceived race and year on search rate. As there are 8 and 2 levels for perceived race and year, this is an 8 x 2 factorial design yielding 16 unique combinations for measurement of the response variable. The boxplot in Figure 7 provides a visualization of how the data is distributed by perceived race and year. Based on the distribution of the data there looks to be potential differences between the groups.

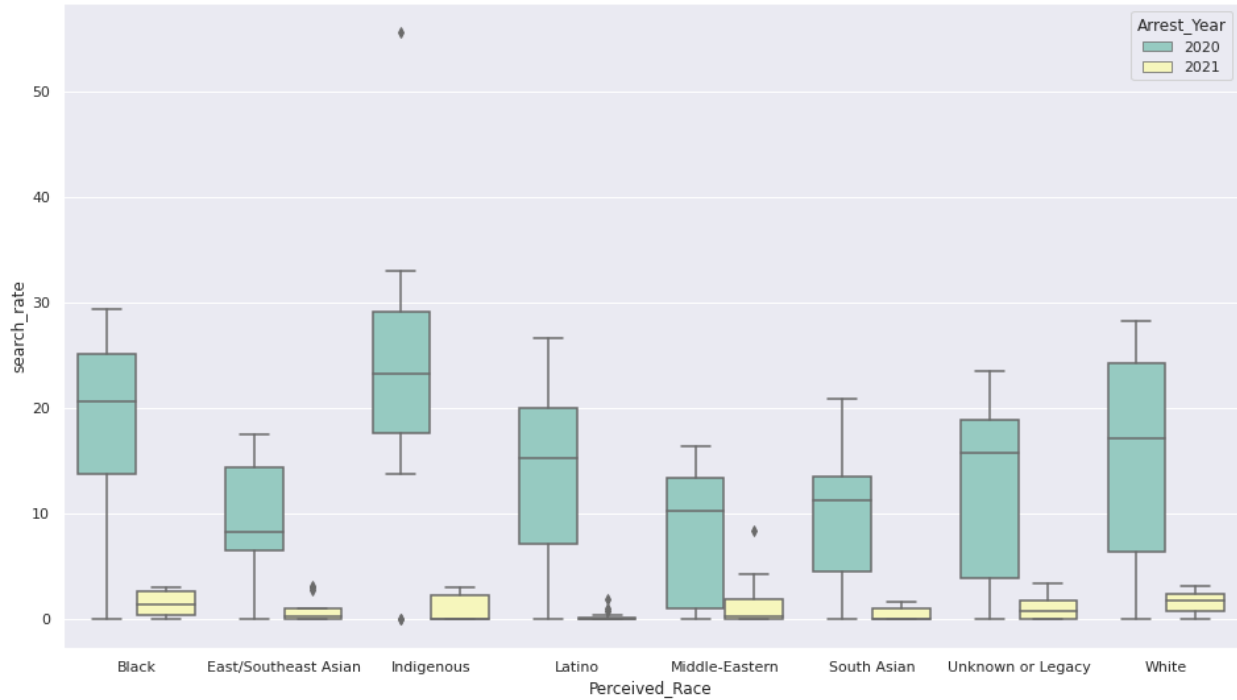


Fig.7. Boxplot of Perceived Races and Year

The results from the two-way ANOVA can be found in Figure 8. The p-value obtained from the ANOVA analysis for perceived race, arrest year, and interaction are all statistically significant with a p-value of less than 0.05. Therefore, the perceived race significantly affects the search count, arrest year significantly affects the search count, and interaction of both perceived race and arrest year significantly affects the search count.

	sum_sq	df	F	PR(>F)
<b>C(Perceived_Race)</b>	1297.575355	7.0	4.586232	8.520835e-05
<b>C(Arrest_Year)</b>	8951.400772	1.0	221.468752	7.362095e-35
<b>C(Perceived_Race):C(Arrest_Year)</b>	1195.590896	7.0	4.225772	2.175314e-04
<b>Residual</b>	8649.526164	214.0	NaN	NaN

Fig.8. Two-way ANOVA Results

Because the interaction is significant an interaction plot was created (refer to Figure 9). The interaction plot helps to visualize the means of the response of the two factors (perceived race and arrest year) on a graph. The x-axis demonstrates perceived race since it is the factor with the most levels. Based on Figure 9 the interaction effect is significant between perceived

race and arrest year because the two lines are not parallel to each other. The interaction is ordinal because the lines do not cross each other.

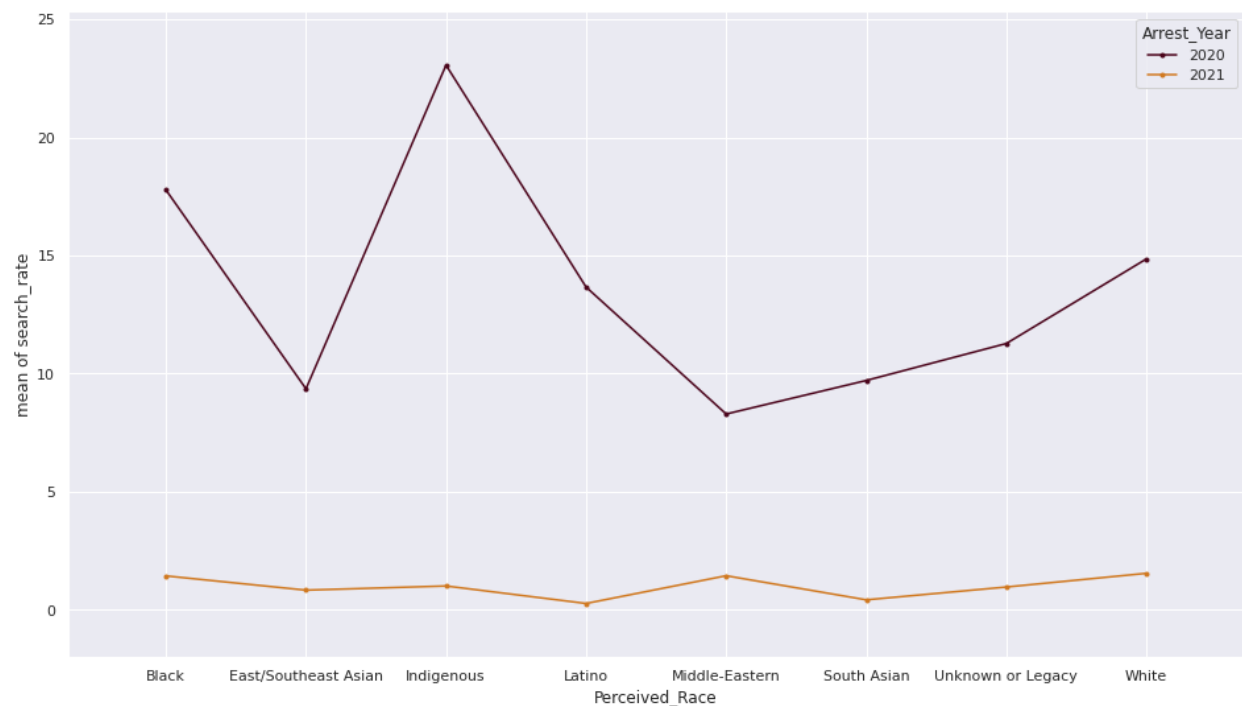


Fig.9. Interaction Plot

Due to the findings of the two-way ANOVA being significant a post-hoc test, Tukey's HSD, was carried out. The post-hoc test will allow us to see which perceived race and arrest year are significantly different from each other. The level of significance to be used is 0.001. Figure 10 provides comparisons between perceived races. The only group with a significant p-value is Indigenous and Middle Eastern.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Black	East/Southeast Asian	4.511982	-0.689241	9.713205	3.755408	0.142166
1	Black	Indigenous	2.426062	-2.775161	7.627285	2.019257	0.824285
2	Black	Latino	3.119559	-2.081665	8.320782	2.596467	0.582866
3	Black	Middle-Eastern	4.740543	-0.460680	9.941766	3.945644	0.102984
4	Black	South Asian	4.540926	-0.660297	9.742149	3.779499	0.136510
5	Black	Unknown or Legacy	2.988863	-2.084964	8.062690	2.550149	0.602237
6	Black	White	1.192165	-3.881662	6.265992	1.017176	0.900000
7	East/Southeast Asian	Indigenous	6.938044	1.736821	12.139267	5.774665	0.001595
8	East/Southeast Asian	Latino	1.392424	-3.808799	6.593647	1.158941	0.900000
9	East/Southeast Asian	Middle-Eastern	0.228561	-4.972662	5.429784	0.190236	0.900000
10	East/Southeast Asian	South Asian	0.028944	-5.172279	5.230167	0.024091	0.900000
11	East/Southeast Asian	Unknown or Legacy	1.523119	-3.550708	6.596946	1.299551	0.900000
12	East/Southeast Asian	White	3.319817	-1.754010	8.393644	2.832525	0.483332
13	Indigenous	Latino	5.545621	0.344397	10.746844	4.615725	0.027538
14	Indigenous	Middle-Eastern	7.166605	1.965382	12.367828	5.964901	0.001000
15	Indigenous	South Asian	6.966988	1.765765	12.168211	5.798756	0.001492
16	Indigenous	Unknown or Legacy	5.414925	0.341098	10.488752	4.620107	0.027274
17	Indigenous	White	3.618227	-1.455599	8.692054	3.087134	0.367445
18	Latino	Middle-Eastern	1.620985	-3.580239	6.822208	1.349176	0.900000
19	Latino	South Asian	1.421368	-3.779855	6.622591	1.183031	0.900000
20	Latino	Unknown or Legacy	0.130695	-4.943132	5.204522	0.111511	0.900000
21	Latino	White	1.927393	-3.146434	7.001220	1.644485	0.900000
22	Middle-Eastern	South Asian	0.199617	-5.001606	5.400840	0.166145	0.900000
23	Middle-Eastern	Unknown or Legacy	1.751680	-3.322147	6.825507	1.494563	0.900000
24	Middle-Eastern	White	3.548378	-1.525449	8.622205	3.027537	0.394383
25	South Asian	Unknown or Legacy	1.552063	-3.521764	6.625890	1.324247	0.900000
26	South Asian	White	3.348761	-1.725066	8.422588	2.857220	0.472299
27	Unknown or Legacy	White	1.796698	-3.146451	6.739846	1.573500	0.900000

Fig.10. Tukey's HSD Results on Perceived\_Race and Arrest\_Year

Figure 11 provides the main effect for the arrest years, which has a p-value of 0.001 therefore it is significant. This means that the alternative hypothesis is true and year has an effect on search rate.

	<b>group1</b>	<b>group2</b>	<b>Diff</b>	<b>Lower</b>	<b>Upper</b>	<b>q-value</b>	<b>p-value</b>
<b>0</b>	2020	2021	12.471741	10.819083	14.124399	21.036346	0.001

Fig.11. Tukey's HSD Results on Arrest\_Year

Figure 12 provides the interaction effect between perceived race and arrest year, here every possible condition is compared. There is a significant interaction for two of the groups in each factor level. The interaction between Black, 2020 and Black, 2021 both have a significant p-value and the size of difference suggests that the interaction effect is influenced by the factor. Additionally the interactions between Black, 2020 and East/South East Asian, 2021, and Black, 2020 and Indigenous, 2021 are significant.

	<b>group1</b>	<b>group2</b>	<b>Diff</b>	<b>Lower</b>	<b>Upper</b>	<b>q-value</b>	<b>p-value</b>
<b>0</b>	(Black, 2020)	(Black, 2021)	16.340336	8.008030	24.672643	9.616914	0.001000
<b>1</b>	(Black, 2020)	(East/Southeast Asian, 2020)	8.420506	0.088200	16.752812	4.955790	0.044764
<b>2</b>	(Black, 2020)	(East/Southeast Asian, 2021)	16.943795	8.611488	25.276101	9.972072	0.001000
<b>3</b>	(Black, 2020)	(Indigenous, 2020)	5.279076	-3.053231	13.611382	3.106938	0.671650
<b>4</b>	(Black, 2020)	(Indigenous, 2021)	16.767288	8.434981	25.099594	9.868191	0.001000

Fig.12. Interaction Effect Between Perceived\_Race and Arrest\_Year

### *Research Question Two*

Research question two is focused on whether there is a significant difference in the frequency of strip searches for males, females, and unknown in the dataset. The null hypothesis is as follows, the mean search rate for females, males, and unknown is all the same. The alternative hypothesis is as follows, the mean search rate for females is not all the same.

Prior to running the ANOVA two tests were run in order to check that they fit the assumptions. The first test that was run was the Levene test to determine whether the three groups have equal variances. The result of the Levene test was a p-value that is greater than 0.05, the p-value was 4.32. In order to check the normality assumption, the data was plotted on a normal probability plot (Q-Q plots) for each grouped variable (refer to Figures 13, 14, 15). The Q-Q plot for females and males does not show a straight-line pattern therefore we assume that

the data is normally distributed. However, the data in the Q-Q plot for unknown falls on the straight line. As for independent observations, based on how the data was collected by the TPS we assume that there is no relationship between the observations within or between the groups. Although not all assumptions were fulfilled we will still carry out the two-way ANOVA.

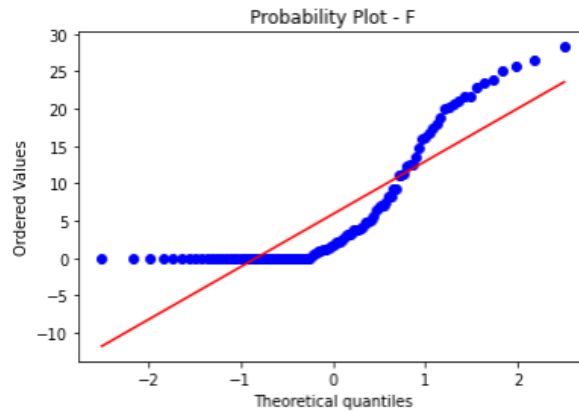


Fig.13. Q-Q plots on Female

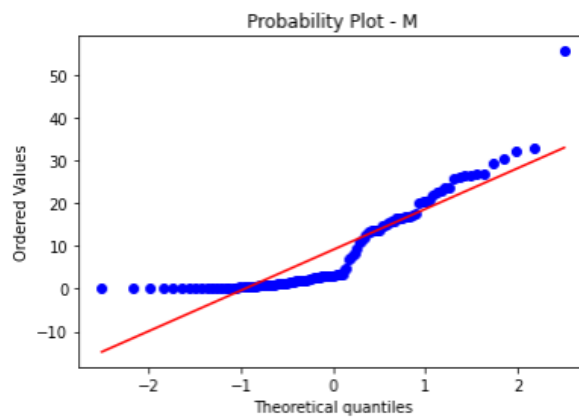


Fig.14. Q-Q plots on Male



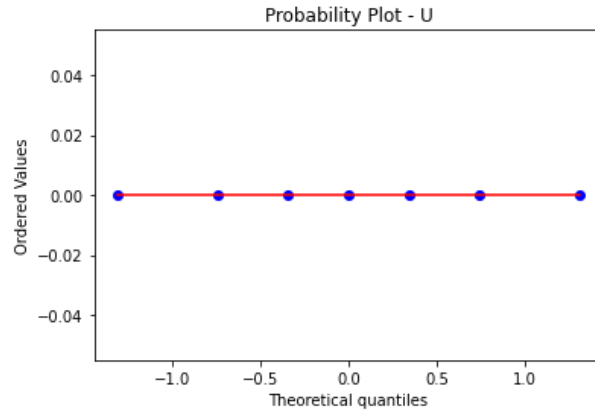


Fig.15. Q-Q plots on Unknown Gender

The figure 16 displays the distribution of search rates for each gender group: female, male and unknown. Due to the data of the unknown gender group are all not searched, so there are no any detailed elements that can be found in the unknown group. According to Figure 16, we can find the medium line for both two female and male groups are almost the same and male group has more samples.

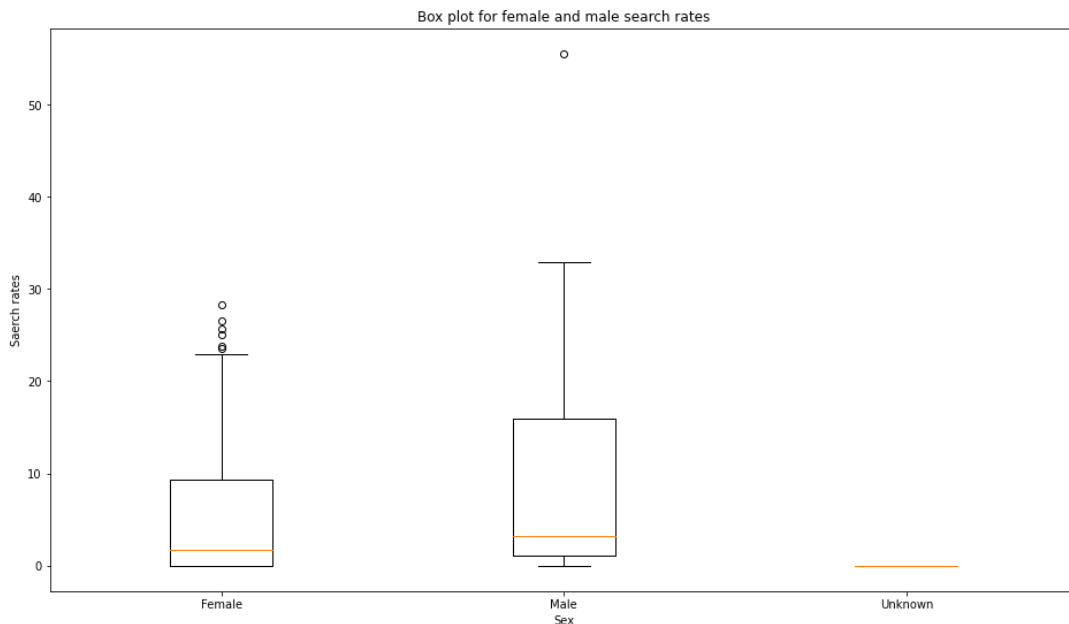


Fig.16. Boxplots of Each Gender Group

The result of the one-way ANOVA was an F-statistic of 5.677 and a p-value of 0.004. Due to the p-value being smaller than 0.05 we can reject the null hypothesis, therefore there is a difference in the mean search rate for male, females, and unknown individuals. We will carry out further a Tukey HSD post-hoc test to validate the results of ANOVA and t-test were significant.

Of the three t-tests that were conducted for research question two, only two returned with a p-value of less than 0.05. The t-test demonstrated for males against females had a result of a p-value smaller than 0.05 (0.011), therefore we can conclude that the mean search rate differs between the two groups. The t-test for males against unknown had a result of a p-value smaller than 0.05 (0.023), therefore we can conclude that the mean search rate differs between the two groups. The t-test for females against unknown had a p-value greater than 0.05 (0.055), therefore we can conclude that the mean search rate does not differ between the two groups.

A post-hoc test, Tukey's HSD was conducted because research question two resulted in a significant result, a p-value less than 0.05 (refer to Figure 17). According to the result of Tukey's HSD the p-value for the difference in means between female and male is 0.026, meaning it is significant. The p-value for the difference in means between female and unknown is 0.228, meaning it is non-significant. The P-value for the difference between male and unknown is 0.031, meaning it is significant. Thus, we conclude that there is a statistically significant difference between the means of female and male, and, male and unknown.

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
F	M	3.2094	0.026	0.3087	6.1101	True
F	U	-5.8975	0.2278	-14.3377	2.5427	False
M	U	-9.1068	0.0309	-17.5448	-0.6689	True

Fig.17. Tukey's HSD Results on Each Gender Group

For the purposes of our research questions we will be conducting both one-way and two-way ANOVA. A one-way ANOVA is a statistical test that compares the variance in the group means within a sample whilst considering only one independent variable or factor. There are two possible hypotheses for a one-way ANOVA.

## Discussion

### *Research Question One*

The results of the post-hoc Tukey's test were surprising, specifically the results demonstrated in Figures 10 and 11. Tukey's post-hoc test (Figure 10) revealed a significant pairwise difference between the perceived races Indigenous and Middle Eastern suggesting that these groups experience different search rates compared to the other perceived races in the dataset. In a search of the literature there are no articles focusing on the strip search rates for

Indigenous and Middle Eastern individuals. It is surprising that Tukey's HSD did not report that there are different search rates between the minority races (ex. Black and East/Southeast Asian) and White. There is an enormous amount of articles available that focus on the disparity in strip search rates based on race. Research has even demonstrated that people of color, such as Black and Middle Eastern individuals, are more likely to be stopped and searched by police than White individuals due to the use of racial profiling. The disparity in strip search rates between minority races and Whites can be based on social and historical context in which policing takes place. For example, there may be disparities in policing resources and practices in different neighborhoods which can lead to a disproportionate impact on minority communities.

Tukey's post-hoc test (Figure 11) demonstrates that there is a significant pairwise difference between the effect of arrest year on search rate. Based on the data the search rate is higher in 2020 than 2021. The result of the Tukey's test for the interaction effect between arrest year and perceived race on search rate showed that there was a significant difference between Black, 2020 and Black, 2021. There is no literature to support why there may have been more strip searches in 2020 than in 2021 based on perceived race. However, factors such as types or frequency of crimes committed, changes in the Toronto Police Service policies, or changes to the public perceptions of policing practices should be considered.

### *Research Question Two*

It is not surprising that the result of the one-way ANOVA for research question 2 was significant. Strip searches are typically conducted based on an individual's perceived risk factor. Men are often considered a higher risk due to a higher rate of arrests, convictions, and repeat offenses. It is also thought that men are more likely to conceal weapons and contraband due to their greater physical strength and a higher probability of escape. However, this practice can be seen as discriminatory. It should be noted that there are a larger number of males included in the dataset in comparison to females and unknown, therefore this could have had an effect on the results of the one-way ANOVA.

The Toronto Police Service (TPS) states that all searches, including strip searches, must be conducted in accordance with the law, with respect for the individual's privacy and dignity, and in accordance with the guidelines set out in the Ontario Police Services Act (2019). The TPS acknowledges that males may be subject to more extensive searches due to the higher rate of crime involving males (2019). However, the TPS has taken steps to ensure that the searches are conducted in a fair and even-handed manner, regardless of gender (2019). The TPS has implemented a policy that requires all searches to be authorized by a supervisor before being performed, and that all searches must be conducted in accordance with the law (2019).

## **Conclusion**

In conclusion, the findings of this data analysis report suggest that there are significant differences in strip search rates based on perceived race and arrest year. The results of Tukey's

post-hoc test were surprising, as there were no significant differences found between strip search rates for minority races and Whites, despite the large amount of literature on racial disparities in policing. However, the significant pairwise difference between Indigenous and Middle Eastern individuals suggests that more research needs to be conducted in this area. Additionally, the one-way ANOVA revealed that men are subject to more extensive strip searches than females and unknown genders.

Future research should further look into and understand the underlying factors that may contribute to differences in strip search rates based on race, and to develop evidence based-policies and procedures to ensure that strip searches are conducted fairly and without bias. This may include providing training and education to police officers, engaging with community members to build trust, and implementing reforms to reduce the use of racial profiling and other discriminatory practices.

### *Limitations*

There were two limitations to this study, the first one is that the dataset that was used for this paper did not include any continuous variables. The dataset only contains categorical variables and binary variables, which limit the type of analyses that can be performed. It could be more difficult to identify their relationships between variables that exist along a continuum, such as the age group in this dataset. It makes it harder to detect subtle trends or to make accurate classifications. Without the continuous variables, it is more difficult to account for individual differences in the data, as discrete categories may not capture the full range variation that exists within a population. This limits the generalizability of the results and makes it more difficult to draw conclusions about the underlying phenomena being studied. In this case, we can only group some variables to generate a new continuous variable to conduct analysis.

The second limitation is that the dataset did not fulfill the assumptions to run a one-way and two-way ANOVA. Assumptions for both the one-way and two-way ANOVA need to be met in order for the results to be valid. When the assumptions are not met the results may be biased or inaccurate. The only assumption that was fulfilled for both research question one and two was that the samples should be independent. Therefore, we are unable to guarantee the reliability of the results in this paper.

## References

- Babbar, M. (2021). *Scholarly Commons: Northwestern pritzker school of law*. Site. Retrieved February 25, 2023, from <https://scholarlycommons.law.northwestern.edu/>
- Psutka, M., & Sheehy, E. (2016, December 1). *Strip-searching of women in Canada: Wrongs and rights*. The Canadian Bar Review. Retrieved February 25, 2023, from <https://cbr.cba.org/index.php/cbr/article/view/4376>
- Smith, B. W., Stinson, P. M., & Paternoster, R. (2020). *Racial and ethnic disparities in the American criminal justice system: A systematic review and meta-analysis*. *Journal of Criminal Justice*, 68, 101657. Retrieved February 25, 2023, from <https://doi.org/10.1016/j.jcrimjus.2020.101657>
- Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press. Retrieved February 25, 2023, from <https://www.thenewpress.com/books/new-jim-crow>
- Lemke, M. (2022, July 18). *Policing toronto: Strip searching in a divided city - the bullet*. Socialist Project. Retrieved February 25, 2023, from <https://socialistproject.ca/2022/07/policing-toronto-strip-searching-in-a-divided-city/>
- Preventing and Addressing Sexual Abuse in Juvenile Facilities: A Call to Action*. Report of the National Prison Rape Elimination Commission. (2009, June). Retrieved February 25, 2023, from <https://www.ojp.gov/pdffiles1/226680.pdf>
- Bevans, R. (2022, December 19). *An introduction to T tests: Definitions, formula and examples*. Scribbr. Retrieved February 28, 2023, from <https://www.scribbr.com/statistics/t-test/>
- Bevans, R. (2022, November 17). *Two-way ANOVA: Examples & when to use it*. Scribbr. Retrieved February 28, 2023, from <https://www.scribbr.com/statistics/two-way-anova/>
- Kumar, A. (2022, April 16). *Correlation Concepts, Matrix & Heatmap using Seaborn*. Data Analytics. Retrieved February 28, 2023, from <https://vitalflux.com/correlation-heatmap-with-seaborn-pandas/#:~:text=A%20correlation%20heatmap%20is%20a,necessarily%20imply%20a%20causal%20relationship.>
- Law document english view*. Ontario.ca. (2018, November 19). Retrieved February 28, 2023, from <https://www.ontario.ca/laws/statute/s19001>
- One-way ANOVA*. One-way ANOVA - An introduction to when you should run this test and the test hypothesis | Laerd Statistics. (n.d.). Retrieved February 28, 2023, from <https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php>

*One-way vs two-way ANOVA: Differences, assumptions and hypotheses.* Informatics from Technology Networks. (n.d.). Retrieved February 28, 2023, from <https://www.technologynetworks.com/informatics/articles/one-way-vs-two-way-anova-definition-differences-assumptions-and-hypotheses-306553>

*Two-way ANOVA in SPSS statistics.* Two-way ANOVA in SPSS Statistics - Step-by-step procedure including testing of assumptions | Laerd Statistics. (n.d.). Retrieved February 28, 2023, from <https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php>