

# **An Exploratory Analysis of the Relationship between Race, Sex and Strip Searches using the RBDC Dataset**

By  
Lisa Tang  
Shamina Vastani

A final paper submitted in conformity with the  
requirements for the course of INF2178H

Faculty of Information  
University of Toronto

## **Abstract**

In 2020, the Toronto Police Services conducted 7,114 strip searches (Forester & Jacobs, 2022). While being in violation of the Human Rights Charter, unless judiciously conducted, the preliminary literature review indicates that strip searches have been regularly performed as an integral part of the Toronto Police Services policing practices. A preliminary analysis of both scholarly and literature in the public domain, indicates that both the frequency and the manner in which these searches were conducted and the associated data have played a significant role in systemically “stigmatizing communities”. The Race Based Data Collection Strategy was undertaken in response to the public outcry against systemic racism and the Service’s role in perpetuating negative and stigmatized stereotypes of racialized communities. That the TPS has conducted more than “hundred’s of strip searches” which were not in accordance with procedure is indicative of a “decoupling” of practice from policy (Barabas et al., 2020) (Aragon et al., 2022), and is worrisome. This line of thinking has motivated this study to investigate the relationship between the perceptions of race and sex and their interaction in policing decisions which result in strip searches.

## Introduction

*“In 2001, in a landmark decision on the constitutionality of strip search practices (R. v. Golden), the Supreme Court of Canada defined what amounts to a strip search, and when and how it can lawfully be done. Despite this decision on the legality of strip searches, courts in Ontario continue to regularly find that police officers unlawfully or unreasonably conduct strip searches, sometimes resulting in the exclusion of evidence or the stay of charge.”*

*“... Strip searches are “inherently humiliating and degrading...” (R. v. Golden, [2001] 3 SCR 679, para 90. (CanLII) <http://canlii.ca/t/51xm>). Individuals who have experienced strip searches have described them as “demeaning,” “upsetting” and “devastating.” Some suffer varying degrees of psychological harm as a result of being strip searched. This is especially so for individuals who have been traumatized in the past or are otherwise vulnerable...*

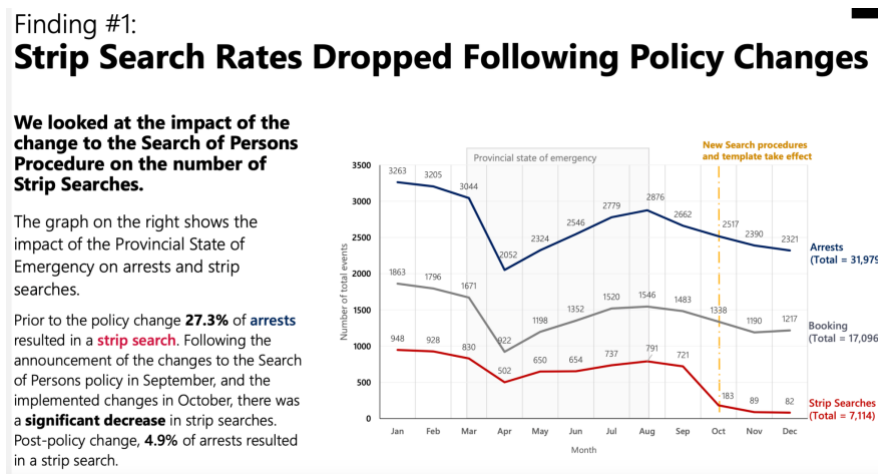
*...Often, those involved in the justice system do not fully understand the impact on those who are affected. Police officers must only conduct or authorize the use of strip searches in compliance with the law, including section 8 of the Canadian Charter of Rights and Freedoms, which states, “everyone has the right to be secure against unreasonable search and seizure.” (Breaking The Golden Rule, A Review of Police Strip Searches in Ontario, p. 5.6)*

In 2020, the Toronto Police Services conducted 7,114 strip searches (Forester & Jacobs, 2022). These are the findings stated in the detailed report on the *Race & Identity Based Data-Collection Strategy* - a living document put together by Professors Dr. Forester & Dr. Jacobs, the TPS and relevant community agencies including those that represent Black, Indigenous, minority groups and marginalized communities (June 2022). According to this report, while the arrested population (total number of arrests) that year was 31, 979, a staggering nearly 42% of the 17, 096 offenders who were taken into custody (54% of the total arrested population) were strip searched.

While being in violation of the Human Rights Charter, unless judiciously conducted, the preliminary literature review indicates that strip searches have been regularly performed as an integral part of the Toronto Police Services policing practices ([Toronto police officer strips naked “hundreds” of people | The Star](#)). Historically, reasons for being subjected to one could include offences such as impaired driving to suspicion of drug trafficking, to court restrictions violations among others ([Breaking the Golden Rule](#), p. 3, 4). Reasons that necessitate the administration of one are stated by the Service as being a proactive measure employed to ensure officer safety as well as of the inmates, upon the offender joining others in the cells. Essentially, the stripsearch is deployed as a pre-emptive measure to forestall any illegal activity behind bars. These could include carrying contraband or weaponry in one’s privates or underwear for the purpose of escaping, or

doing intentional harm to an inmate or officer on duty. However, the Supreme Court of Canada has ruled that proper procedure must be followed should it be deemed necessary for a strip search to be conducted. The ruling clearly states that the Service must establish “probable cause” which include reasons (possession) as justification, circumstances (location) regarding where one should be conducted and the manner (one piece of clothing and body area at a time) ([R v Golden - Wikipedia](#)). In other words, enforcing strip searches as part of “routine procedure” and without just cause is unconstitutional in Canada.

More recently, following the implemented changes to the *Search of Persons* policy in October 2020, a significant decrease in the number of stripsearches was noted. As per the RBDC Strategy report, the number of individuals subjected to stripsearches fell from 27.3% to 4.9%. See fig. 1 below for reference (Forester & Jacobs, 2022, p.75).



**Figure 1 Source.** Forester & Jacobs, 2022, p.75

## **Literature Review / Background**

Historically, mandatory strip searches have been enforced as routine procedure by the Toronto Police Services. A preliminary analysis of both scholarly and literature in the public domain, indicates that both the frequency and the manner in which these searches were conducted and the associated data have played a significant role in systemically “stigmatizing communities”. As can be summarized from the *OIPRD’s Breaking the Golden Rule* report, the fact that the records of such searches were not consistently and correctly maintained further exacerbated the situation. This is because even though there was public outcry and sufficient motivation to prompt investigations into complaints, the lack of insufficient data, especially race-based data impeded the cause (*OIPRD\_Breaking-the-Golden-Rule\_Report\_Accessible.pdf*, p. 9,10). The implementation of the changes to the *Search of Persons policy* and the RBDC strategy is corrective measure that is designed as a systematic data-based approach to address systemic racism in the Services (Forester & Jacobs, 2022). This Race Based Data Collection Strategy was undertaken in response to the public outcry against systemic racism and the Service’s role in perpetuating negative and stigmatized stereotypes of racialized communities. The Strategy can be described as a methodology designed to identify and address inequality, and racist practices in the use and abuse of racialized data of citizens to target and perpetuate harm.

The Human Rights Charter is central to and guides policy especially as it relates to ensuring equitable treatment of all citizens. That the TPS has conducted more than “hundred’s of strip searches” which were not in accordance with procedure is indicative of a “decoupling” of practice from policy (Barabas et al., 2020) (Aragon et al., 2022), and is worrisome. The accounts of Forrester (Kirkup, 2009) and Golden (2001) and the infamous and heart breaking George Floyd incident only make vivid the violence that racist attitudes and biases can result in. This line of thinking motivated this study to investigate the relationship between the perceptions of race and sex and their interaction in policing decisions which result in strip searches.

### **Research Question**

This research is interested in statistically examining the following three research questions:

- 1) What kind of relationship/ influence if any does the variable Sex of the offender have on the stripsearch outcome?
- 2) What kind of relationship/ influence if any does the variable Perceived\_Race of the offender have on the stripsearch outcome?
- 3) What kind of a relationship/ influence do variables Perceived\_Race and Sex, when taken together have on the stripsearch outcome?

## Method

The study applies a quantitative analysis approach to investigating whether an offender will be stripsearched. The data analysis was conducted in Python using the Pandas and Numpy libraries for processing the data whereas seaborn and matplotlib packages were used for the graphical presentation of the data (data visualization).

### Data set

This research uses the "Arrests and Strip Searches (RBDC\_ARR\_TBL\_001)" dataset, maintained by the Toronto Police Service, contains information on individuals who were arrested and strip searched in Toronto between January 1, 2010, and December 31, 2019 and can be accessed through [Toronto Police Service Public Safety Data Portal](#)

The purpose of this dataset is to provide transparency to the public regarding the activities of the Toronto Police Services in conducting arrests and strip searches and to identify potential issues or areas for improvement in police practices. The dataset contains 65,276 entries that provide information about different characteristics of each arrest and strip search, such as the arrestee's age, sex, and perceived race, as well as the location and cause for the arrest. There are no null or duplicate values in the dataset.

	Arrest_Year	Arrest_Month	EventID	ArrestID	PersonID	Perceived_Race	Sex	Age_group_at_arrest	Youth_at_arrest_under_18_years	ArrestLocDiv	
	0	2020	July-Sept	1005907	6017884.0	326622	White	M	Aged 35 to 44 years	Not a youth	54
	1	2020	July-Sept	1014562	6056669.0	326622	White	M	Aged 35 to 44 years	Not a youth	54
	2	2020	Oct-Dec	1029922	6057065.0	326622	Unknown or Legacy	M	Aged 35 to 44 years	Not a youth	54
	3	2021	Jan-Mar	1052190	6029059.0	327535	Black	M	Aged 25 to 34 years	Not a youth	XX
	4	2021	Jan-Mar	1015512	6040372.0	327535	South Asian	M	Aged 25 to 34 years	Not a youth	XX
	...	...	...	...	...	...	...	...	...	...	...
	65271	2021	Oct-Dec	1055609	6044336.0	316123	Indigenous	F	Aged 25 to 34 years	Not a youth	XX
	65272	2021	Oct-Dec	1032758	6031692.0	307736	South Asian	M	Aged 35 to 44 years	Not a youth	54
	65273	2021	Oct-Dec	1021067	6064396.0	324057	White	F	Aged 45 to 54 years	Not a youth	XX
	65274	2021	Oct-Dec	1008998	6008662.0	331870	Unknown or Legacy	M	Aged 17 years and under	Youth (aged 17 years and under)	XX
	65275	2021	Oct-Dec	1033395	6032145.0	310583	Latino	M	Aged 18 to 24 years	Not a youth	XX

**Figure 2.** "Arrests and Strip Searches (RBDC\_ARR\_TBL\_001)" dataset

## Measurement

The measurement variables selected in this study were used to determine the causality between race of the individual, their sex and the chance of being stripsearched. The assumption underpinning the analysis is that stripsearches are not part of “routine procedure”, but are deemed necessary as per the situation or criminal activity being investigated. Data Types that were non-numeric were factorized for convenience of calculation and data presentation. The explanation and the meaning of each variable is derived from the accompanying metadata literature pertaining to the dataset (*rbdc-resources*).

### Independent Variables

The independent variables for this study are *Sex* and *Perceived\_Race*. Both variables are of categorical datatype in the original dataset. While the datatype for *Perceived\_Race* is retained, that of *Sex* has been factored. Each has multiple levels. The levels of *Perceived\_Race* are self-explanatory and listed as: *White, Black, Unknown or Legacy, East/Southeast Asian, South Asian, Middle-Eastern, Indigenous, Latino*. The variable *Sex* has three levels: *Male, Female, Unknown* where each value corresponds to the appropriate letter *M, F, U*. After factorizing, the string values are presented as 0 for Male or M, 1 for Female or F, and 2 for Unknown or U. *ArrestLocDiv* is the area where the arrest was made and under which divisional unit (*rbdc-resources*). While this variable is not actively engaged as a predictor, it serves to inform as a proxy to the *Perceived\_Race* variable. In that, it provides information of the area in which the arrests were made thus also pointing to the demographic details of the locality ( *Race & Identity Based Data Collection Strategy Understanding Use of Force & Strip Searches in 2020 Detailed Report Toronto Police Service*. 2022, p. 62) *ArrestLocDiv* is of datatype *int* with 17 unique values: each stands for a divisional unit of the Service. This variable was factorized to ensure privacy as well as to facilitate calculation. (see Appendix 2C for details)

### Dependent Variable:

The response variable *StripSearch* is a categorical variable that classifies 0 = False and 1 = True to indicate whether a strip search was performed. There is only one dependent variable for this study.



## Data Analysis

The Exploratory Data Analysis was performed on a subset of the original dataset. The new dataset *tbd* comprises 65276 rows and 5 columns, namely, *Perceived\_Race*, *Sex*, *ArrestLocDiv*, *Occurrence\_Category* and *StripSearch*. *Occurrence\_Category* was dropped as it is not relevant to the study at this time. As discussed above, variables *Perceived\_Race* and *Sex* are the key Predictor Variables and *StripSearch* is the Response Variable. *ArrestLocDiv* is useful as a proxy variable. There are no null or duplicate values.

	Perceived_Race	Sex	ArrestLocDiv	Occurrence_Category	StripSearch
0	White	M	54	Assault & Other crimes against persons	0
1	White	M	54	Assault & Other crimes against persons	0
2	Unknown or Legacy	M	54	Assault & Other crimes against persons	0
3	Black	M	XX	Harassment/Threatening	0
4	South Asian	M	XX	FTA/FTC/Compliance Check/Parolee	0
...	...	...	...	...	...
65271	Indigenous	F	XX	Vehicle Related	0
65272	South Asian	M	54	Assault	0
65273	White	F	XX	Assault	0
65274	Unknown or Legacy	M	XX	Robbery/Theft	0
65275	Latino	M	XX	Mischief	0

65276 rows x 5 columns

**Figure 3.** Dataset *tpd*: subset of "Arrests and Strip Searches (RBDC\_ARR\_TBL\_001)" dataset

The following assumptions informed the study:

- The RBDC Strategy is a methodology employed to unpack biases against racial and gender stereotyping that inform policing decisions to conduct strip searches. Its purpose is to facilitate the monitoring, identifying and addressing the practices that perpetuate systemic racism that continue to harm and stigmatize racialized communities; (*Race & Identity Based Data Collection Strategy Understanding Use of Force & Strip Searches in 2020 Detailed Report Toronto Police Service*. 2022)
- The dataset serves as a tool to identify racist practices perpetuated as a result of prevailing biases in the TPS. It is not intended for the purpose of identification and violation of privacy of individuals in the Service nor is the data collected intended to be used for punishment; (*Race & Identity Based Data Collection Strategy Understanding Use of Force & Strip Searches in 2020 Detailed Report Toronto Police Service*. 2022)
- Racial bias is operationalized by the variable *Perceived\_Race* which makes it a key predictor in the analyses;

- Strip searches are not part of “routine procedure”;
- The data input for *Perceived\_Race* is not acquired through self-identification. According to the accompanying metadata literature and the section on Strip Searches, *Race & Identity Based Data Collection Strategy Understanding Use of Force & Strip Searches in 2020 Detailed Report Toronto Police Service* ( 2022), this datapoint is determined and entered by the Officer In Charge;
- It is assumed that *Sex* is confirmed with the offender and can thus be assumed as self-identified. However, according to the literature review, in cases where confirmation is not determined to be reliable, the OIC has the last word on the value selection of the *Sex* of the individual (*Race & Identity Based Data Collection Strategy Understanding Use of Force & Strip Searches in 2020 Detailed Report Toronto Police Service, 2022*)

Note: there is no explanation in the description of this variable or practice in the accompanying metadata literature.

The following insights were gleaned:

- The mean value of the population *StripSearch* is 0.119787, while its median and IQR values are 0.00. This indicates that there are higher numbers of False values in the dataset than True values. It can be inferred that fewer strip searches were recorded as well as the chances of getting stripsearched were less than probable at that time. This could be attributed to the changes made to the *Search of Persons* policy. (See fig 1 for more information.

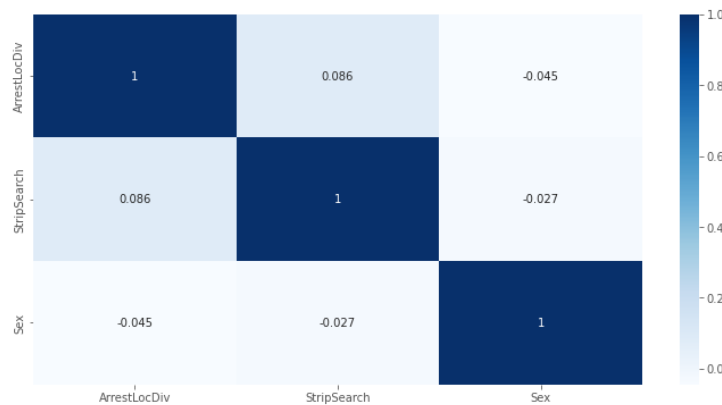
	Sex	ArrestLocDiv	StripSearch
count	65107.000000	65107.000000	65107.000000
mean	0.193543	5.051700	0.119787
std	0.395428	5.119414	0.324715
min	0.000000	0.000000	0.000000
25%	0.000000	1.000000	0.000000
50%	0.000000	2.000000	0.000000
75%	0.000000	9.000000	0.000000
max	2.000000	17.000000	1.000000

**Figure 4.** *Descriptive Statistics of dataset tpd*

- The mean value of the population *Sex* is 0.1935, with more males than females who were subjected to strip searches. Note, that the data accounts for all individuals stripsearched using binary classification as explained in the metadata literature. To state emphatically,

gender is not captured in the dataset only the sex of the individual is. This can prove to be particularly problematic in cases where an individual may be in transition or undergoing gender reassignment as discussed by Kirkup (2009).

- The correlation matrix in figure 4 indicates that *Sex* of an individual is negatively correlated with the chances of someone being stripsearched. However, the location of where an arrest is made (*ArrestLocDiv*), i.e the locality where a community or racialized population resides, has a positive correlation with *StripSearch*. This finding can be used to attest that the higher number of arrests made in a location lead to the subsequent stripsearches of those arrested resulting in certain groups being over-represented in the data. This information can be used to test whether racial biases towards certain groups in the population result in members of those communities being stripsearched and over-represented as being subjected to strip searches .

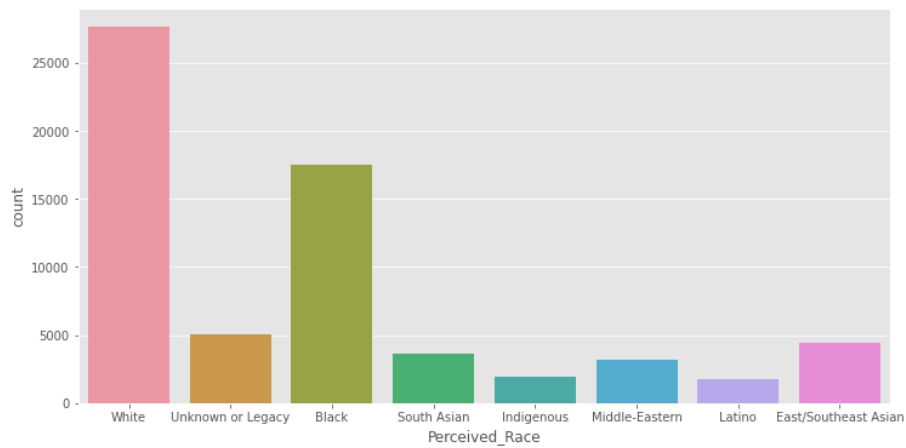


**Figure 5.** *Correlations Matrix of Numeric Variables*

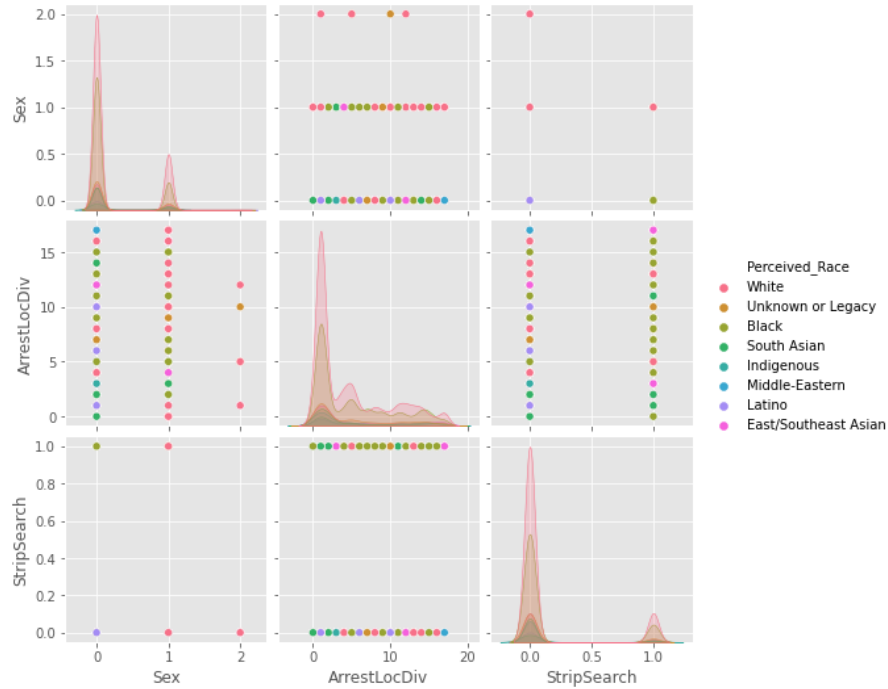
In conclusion, it can be inferred that while fewer strip searches were recorded, the chances of getting stripsearched were less than probable based on *Sex* of the individual. This could be attributed to the changes made to the *Search of Persons* policy. However, there is the likelihood that the ones that were recorded were influenced by racial bias as is indicated by the correlation between *ArrestLocDiv* and *StripSearch*.

#### Analysis of Predictor Variable *Perceived\_Race*

- The bar graph below presents the demographic distribution of racial groups presented in the data. The pairs plot shows the relationships between both categorical and continuous variables.
- Each group serves as a level of the categorical variable *Perceived\_Race*. White and Black groups have a significantly larger presence, with those identified as White are present nearly twice as much as those identified as Black. However as the pairs plot in fig.6 indicates the Black community is over-represented in arrests and strip searches thereby suggesting an over-representation and a strong case for racial bias. This is because given that the information about an individual's race is used as a key identifier, and is dependent on the perception of the officer, it can be attributed to having significant influence in determining a positive outcome for StripSearch.



**Figure 6.** Bar Graph showing the different levels of *Perceived\_Race*

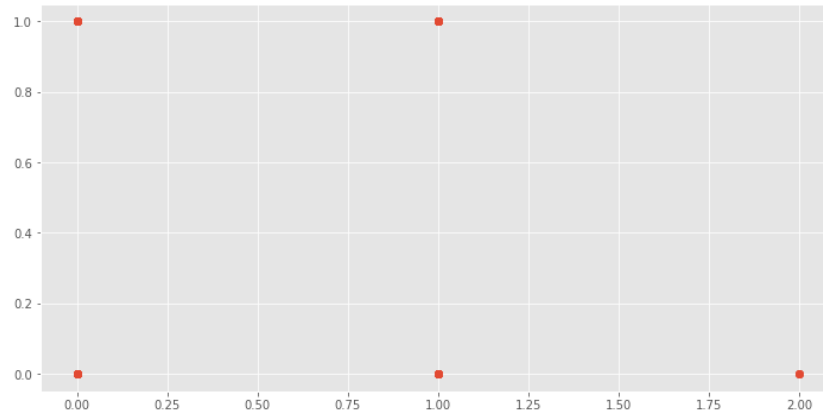


**Figure 7.** Pairsplot showing relationships between both *Perceived\_Race* and *Sex*, *ArrestLocDiv* and *StripSearch* variables.

- Indigenous population is under-represented in the data. In that their population count is significantly lower than the Black and White populations yet comparatively over-represented in the StripSearch count. Such a skewed representation is demonstrative of systemic bias resulting in continued stigmatizing and negative stereotyping of communities.

#### An Analysis of Predictor Variable *Sex*

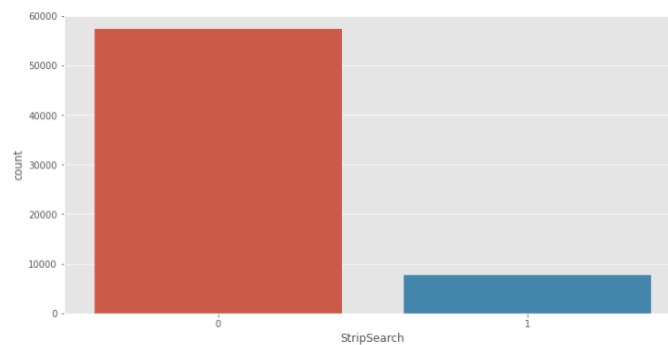
- It can be inferred from the pairs plot in figure 6 nearly twice as many males were arrested as females. It must be pointed out that the differentiation is on the basis of *Sex* and not on the gender of the person. Those identified as Unknown were significantly few. As discussed earlier, *Sex* does not appear to have a significant impact on whether a strip search will be performed as it is negatively correlated. That is, an individual who identifies as ‘Male’ or ‘Female’ has an equal chance of getting stripsearched. The scatterplot below confirms this finding.



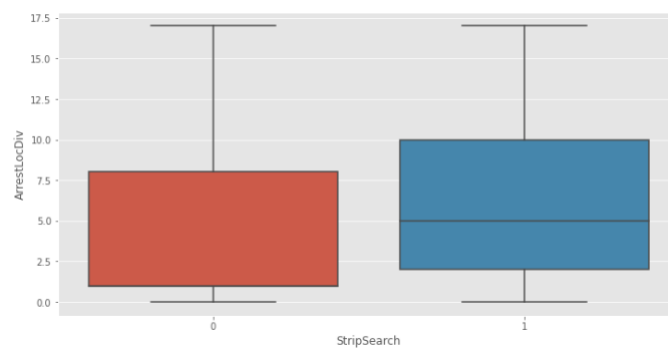
**Figure 8.** Scatterplot showing the relationship between Sex and StripSearch

Analysis for dependent variable *StripSearch*

- The bar plot and box plots below present the number of strip searches that were performed. While those that were performed were fewer, the times where an arrest led to a strip search is more than 5 times. Note that these are measured against the proxy variable for *Perceived\_Race*. Referring back to the pairs plot, a more meaningful analysis indicates that even though the StripSearch counts were fewer, the majority of those stripsearched were identified as Black.



**Figure 9.** Bar Graph showing True and False Counts of StripSearch



**Figure 10.** *Boxplot showing values of StripSearch v/s ArrestLocDiv*

From the discussion above it can be concluded that *Perceived\_Race* and *Sex* have a significant influence in who is subjected to a stripsearch.

## **T-test**

The T-test is a statistical tool utilized to compare the means of two distinct groups. By employing the t test, we can find out whether the two groups exhibit any significant differences. To examine the variables we want to explore, we will conduct a total of five t tests.

### **Sex(Male, Female) and getting strip search**

Our initial inquiry focused on whether there was a discrepancy in the likelihood of undergoing a strip search between female and male. To test our hypothesis, we formulated the following statements:

*Null hypothesis (H0): on average, there is no discernible difference in the probability of a strip search occurring between males and females.*

*Alternative hypothesis (H1): on average, there is a difference between the sexes in the chance of undergoing a strip search.*

Upon analyzing the t test results, we obtained a p-value of 4.3e-13. This value is much smaller than the significance level of 0.05, indicating that we can reject the null hypothesis with a high degree of certainty. Therefore, we can confidently state that there is a true difference in means between the sexes regarding the likelihood of a strip search occurring. Note this finding is different from that presented by the scatterplot in fig. 8.

### **Perceived race(White, Black) and strip search**

Our second investigation examines whether there is a variation in the likelihood of a strip search between white individuals and black individuals. To test our hypothesis, we formulated the following statements:

*Null hypothesis (H0): On average, white individuals and black individuals DO NOT differ in the chance of getting strip search.*

*Alternative hypothesis(H1): On average, white individuals and black individuals DO differ in the chance of getting strip search.*

After analyzing the t test output, we obtained a p-value of 0.0019. This value is lower than the significance level of 0.05, indicating that we can reject the null hypothesis with a high degree of confidence. Therefore, we can confidently state that there is a genuine difference in means between white individuals and black individuals regarding the likelihood of undergoing a strip search.

### **Perceived race(Indigenous, Black) and strip search**

Our third investigation aims to determine if there is a difference in the probability of a strip search occurring between indigenous individuals and black individuals. To test our hypothesis, we created the following statements:

*Null hypothesis (H0): On average, indigenous individuals and black individuals DO NOT differ in the chance of getting strip search.*

*Alternative hypothesis(H1): On average, indigenous individuals and black individuals DO differ in the chance of getting strip search.*

Upon examining the t test results, we obtained a p-value of 0.026. This value is smaller than the significance level of 0.05, indicating that we can reject the null hypothesis and state that there is a difference in means between indigenous individuals and black individuals regarding the likelihood of a strip search occurring.

### **Black male and Black Female and getting strip search**

Our final investigation is focused on determining if there is a difference in the likelihood of a strip search occurring between black female individuals and black male individuals. To test our hypothesis, we formulated the following statements:

*Null hypothesis (H0): On average, black female individuals and black male individuals DO NOT differ in the chance of getting strip search.*

*Alternative hypothesis(H1): On average, black female individuals and black male individuals DO differ in the chance of getting strip search.*



Upon analyzing the t test output, we obtained a p-value of  $1.62e-28$ . This value is way smaller than the significance level of 0.05, indicating that we can reject the null hypothesis with a high level of confidence. Therefore, we can confidently state that there is a difference in means between black female individuals and black male individuals regarding the likelihood of undergoing a strip search.

In conclusion, the results of our t tests demonstrate that there is a statistically significant difference in the likelihood of a strip search occurring depending on the sex and race of the individuals involved. These findings suggest that there may be bias and discrimination present in the strip search practices of the population under study. Further investigation and intervention may be necessary to address and rectify these issues.

## Results

We conducted two one-way ANOVA tests for the first two research questions and one two-way ANOVA test for the last research question.

### One-way ANOVA Test

Using one-way ANOVA tests to see if there is a significant difference between sex and getting strip search and perceived race and strip search, we were able to address answers for the first research question. The null hypothesis in this context posits that the average of getting strip searched is the same across all categories of independent variables, namely sex and perceived race. On the other hand, the alternative hypothesis proposes that there exists a difference in at least one mean from the others.

Upon conducting our study, we found a statistically significant difference in the mean of receiving a strip search based on sex ( $F(2) = 24.3$ ,  $p < 0.001$ ). Further examination via a Tukey post-hoc test exposed significant pairwise difference between female and male, with an average difference of 0.001 bushels/acre ( $p < 0.01$ ). However, there was no statistical difference in the mean of getting stripsearched between female and unknown and male and unknown.

The results of our statistical analysis show that there is a statistically significant difference in the mean rates of individuals receiving strip searches based on their perceived race. Specifically, our analysis revealed that the mean rates of strip searches differed significantly across racial groups ( $F(7) = 55.2$ ,  $p < 0.001$ ). To further explore these differences, we conducted a Tukey post-hoc test to compare the pairwise mean differences between the racial groups. Our analysis revealed that there were statistically significant differences in the mean rates of receiving a strip search between Black and Asian individuals ( $p = 0.001$ ), Black and White individuals ( $p < 0.05$ ), Indigenous and White individuals ( $p < 0.01$ ), Indigenous and Asian individuals ( $p = 0.001$ ), Latino and White individuals ( $p = 0.001$ ), and Asian and White individuals ( $p = 0.001$ ). These findings suggest that these groups experience different rates of strip searches based on their perceived race. However, our analysis also revealed that there was no statistically significant difference in the mean rates of receiving a strip search between Asian and Eastern individuals and between Black and Indigenous individuals, as the p-values were 0.9, which are much larger than the significance level of 0.05.

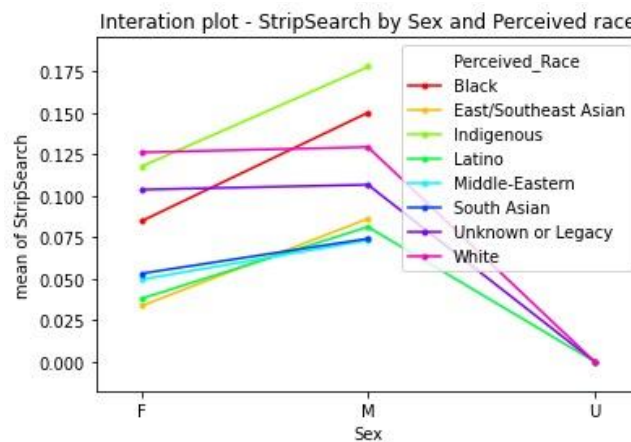
Therefore, we cannot conclude that there is a significant difference in the mean rates of strip searches between these two groups.

## Two-way ANOVA Test

In our study, we aimed to investigate the influence of both gender and perceived race on the likelihood of being subjected to a strip search by police officers. To analyze the data, we conducted a two-way ANOVA test, which allowed us to assess the impact of both variables on the outcome of interest. Our findings indicate that there was a significant interaction between *Perceived\_Race* and *Sex*, as evidenced by the p-value being much lower than the level of significance set at 0.01.

To further explore the relationship between the variables, we carried out Tukey post-hoc tests, which were similar to the ones we performed in the one-way ANOVA analysis. Our results confirmed that there is a statistically significant difference between *Perceived\_Race* and *StripSearch*, as well as between *Sex* and *StripSearch*.

While the two-way ANOVA did not provide any new substantial information, we constructed an interaction plot below to visually depict the relationship between *Perceived\_Race*, *Sex*, and the likelihood of being strip-searched. Our analysis of the plot showed that the Indigenous and Black group had the highest mean rate of being strip-searched by police officers, followed by Black males.



**Figure 11.** Interaction plot

## Discussion

To address our first research question, we conducted a statistical analysis using a one-way ANOVA test to examine the association between gender and the likelihood of undergoing a strip search. The results of the analysis showed a significant difference in the mean values of strip searches received based on sex, with males having a greater chance of being subjected to strip searches than females. This suggests that male individuals were disproportionately impacted by police officers in terms of strip search practices.

Moving on to our second research question regarding the relationship between *Perceived\_Race* and *StripSearch*, we conducted another one-way ANOVA test. The results of this analysis showed that the mean rates of *StripSearch* differed significantly across racial groups. The Tukey hsd results revealed statistically significant differences between various racial groups, with Black and Asian individuals, Black and White individuals, Asian and White individuals, and Indigenous and other race groups showing significant differences in mean rates. These findings suggest that strip searches are racially biased, with some racial groups, such as Black individuals, being disproportionately impacted by police officers.

Finally, a two-way ANOVA was performed to answer our third research question about the impact of the interaction of *Sex* and *Perceived\_Race* on the likelihood of undergoing a strip search. The analysis revealed a significant interaction between *Perceived\_Race* and *Sex*, with Indigenous males and Black males having the high mean rates of being strip-searched by police officers. These findings provide valuable insights into the possible disproportionate impact of perceived race and sex on police officers' decisions to conduct strip searches.

In summary, our study highlights the need to address potential racial and sex-related biases in strip search practices by police officers. It is important to acknowledge and rectify any discriminatory practices in the criminal justice system to ensure fair and just treatment for all individuals.

## Limitations

Although the dataset from Toronto Police Service regarding arrests and strip searches provides valuable information for research and analysis, there are several limitations to consider. First, the dataset contains information only about individuals arrested and strip-searched by Toronto Police Service officers, which may not be representative of the population as a whole (ie. Services across Canada). The data set cannot be used to draw conclusions about the prevalence of strip-searches or arrest rates in the wider population. Also, the data set covers only a limited time range from 2020 to 2021, and only includes data from the Toronto Police Service. As a result, this data set may not be generalizable to other police forces or more recent time periods. Second, the majority of the variables in this dataset are categorical, which may present certain limitations in terms of identifying meaningful relationships and patterns in the data. Numerical variables can provide a more nuanced and detailed understanding of the data. But in this case, most of the variables in this dataset are categorical, the options for conducting exploratory data analysis (EDA) are somewhat limited. For example, certain types of visualizations such as box plots and line graphs require descriptive statistics such as mean and standard deviation, which may not be available for this dataset especially for key predictor variables such as *Perceived\_Race*. Furthermore, while the study investigated several variables and generated three research questions, it is not comprehensive enough to account for all potentially influencing factors. Given the complexity of the issue of strip searches by police officers, there may be other factors beyond the ones investigated in this study that also play a role. Lastly, the t-test result for Black and Indigenous individuals suggests that there is a significant difference between the mean rates of strip searches received by these two groups. However, the Tukey HSD test, which compares mean rates across multiple groups, did not find a significant difference between Black and Indigenous individuals. While the p-values for both tests are close to 0.05, it is important to conduct further studies to obtain a more accurate conclusion. Therefore, the results of this study should be interpreted with caution and further research may be needed to fully understand the relationships between gender, race, and strip searches.

## Future Work

As discussed earlier, the limited number of variables restricts the data analyses to a particular view. To get a more holistic picture to measure the efficacy of the RDBC strategy to systematically identify racist practices, feature engineering should be considered in future work.

## Conclusion

The study's objective is to examine the connections between the likelihood of being subjected to a strip search and the independent variables of *Perceived\_Race* and *Sex*. The dataset used is the Arrests and Strip Searches dataset from the Toronto Police Service Public Safety Data Portal. The EDA provides a quick overview of the dataset, with analyses conducted on each variable of interest and visual aids such as pairs plots and box plots. The t-tests demonstrate significant gender and racial differences in the likelihood of being strip-searched, which suggest potential discrimination and bias in the strip search practices under study. The statistical analysis results indicate that males are more likely to undergo strip searches than females, and certain racial groups, such as Black individuals, are more likely to be subjected to strip searches than White and Asian individuals. Additionally, the study reveals a significant interaction between perceived race and gender concerning the likelihood of strip searches. Although it would be inaccurate and inappropriate to attribute the chance of being strip-searched solely to race and gender, since many other factors may impact an officer's decision, including the location, time of day, and potential weapons, for instance. All relevant variables must be considered to arrive at a sound conclusion. Nonetheless, the research suggests potential racial bias, highlighting the importance of conducting strip searches in a just and impartial manner, free from discrimination or bias.

# Appendix 1

## One - way ANOVA

Sex

Source of Variance	Sum of Squares	df	F	Sig.
Between Groups	5.1	2	24.3	2.9e-11 ***
Within Groups	6863.6	65273		
Total	6868.7	65275		

## Tukey multiple comparisons of means

95% family-wise confidence level

group1	group2	meandiff	p-adj	lower	upper	reject
F	M	0.0221	0.001	0.0146	0.0296	TRUE
F	U	-0.1017	0.6055	-0.3551	0.1517	FALSE
M	U	0.1238	0.4879	-0.3771	0.1296	FALSE

## One - way ANOVA

Perceived\_Race

Source of Variance	Sum of Squares	df	F	Sig.
Between Groups	40.4	7	55.2	3.8e-79 ***
Within Groups	6827.5	65264		
Total	6867.9	65271		

## Tukey multiple comparisons of means

95% family-wise confidence level

group1	group2	meandiff	p-adj	lower	upper	reject
Black	Asian	-0.06	0.001	-0.08	-0.04	TRUE
Black	White	-0.01	0.0229	-0.02	-0.0008	TRUE
Indigenous	White	-0.03	0.0025	-0.05	-0.0065	TRUE
Indigenous	Asian	-0.09	0.001	-0.11	-0.06	TRUE
Latino	Asian	-0.003	0.9	-0.032	0.0249	FALSE
Latino	White	0.054	0.001	0.03	0.08	TRUE
Asian	White	0.06	0.001	0.04	0.07	TRUE

## Two - way ANOVA

Dependent Variable: Strip Search

Source of Variance	Sum of Squares	df	F	Sig.
Sex	6.3	2	30.2	7.7e-14 ***
Perceived Race	15.4	7	21.1	2.0e-17 ***
Interaction	11	14	7.5	4.1e-12 ***
Error	6812.6	65253		
Total	6845.3	65276		

## Appendix 2

### A: Distribution by Perceived\_Race

```
1 # check distribution of values
2 tpd.Perceived_Race.value_counts()
```

```
White          27723
Black          17526
Unknown or Legacy 5056
East/Southeast Asian 4415
South Asian    3613
Middle-Eastern 3237
Indigenous     1934
Latino         1768
Name: Perceived_Race, dtype: int64
```

```
1 #tpd['Perceived_Race'] = pd.factorize(tpd['Perceived_Race'])[0]
```

```
1 # calculate percentage of Perceived_Race
2 tpd.Perceived_Race.value_counts(normalize=True)
```

```
White          0.424730
Black          0.268507
Unknown or Legacy 0.077460
East/Southeast Asian 0.067460
South Asian    0.055353
Middle-Eastern 0.049592
Indigenous     0.029630
Latino         0.027087
Name: Perceived_Race, dtype: float64
```

### B: Datatypes of each variable in subset tpd

```
Perceived_Race      object
Sex                 int64
ArrestLocDiv        int64
Occurrence_Category object
StripSearch         int64
dtype: object
```

### C: ArrestLocDiv factorization

```
t[19]: XX      29643
      51      5025
      14      3117
      41      2640
      43      2607
      52      2340
      31      2266
      32      2088
      12      1902
      22      1849
      42      1846
      11      1789
      23      1738
      55      1552
      54      1437
      53      1376
      13      1102
      33       955
      Name: ArrestLocDiv, dtype: int64

[20]: 1 tpd.ArrestLocDiv.isnull().sum()
t[20]: 0

[21]: 1 # step 3: factorize ArrestLocDiv to change from object to int (XX to a numeric value
      2 tpd['ArrestLocDiv'] = pd.factorize(tpd['ArrestLocDiv'])[0]

[22]: 1 tpd.ArrestLocDiv.unique()
t[22]: array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
              17])

[23]: 1 # step 4: check if XX is changed to numeric value
      2 tpd.ArrestLocDiv.value_counts()
t[23]: 1      29643
      5      5025
      4      3117
      11      2640
      14      2607
      3      2340
      7      2266
      17      2088
      9      1902
      12      1849
      2      1846
      8      1789
      15      1738
      13      1552
      0      1437
      6      1376
      10      1102
      16       955
      Name: ArrestLocDiv, dtype: int64
```



## References

- Aragon, C., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-Centered Data Science*. MIT Press.
- Barabas, C., Doyle, C., Rubinovitz, J., & Dinakar, K. (2020). Studying up. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.  
<https://doi.org/10.1145/3351095.3372859>
- Data Analysis Framework for Equity*. (n.d.). www.tps.ca. Retrieved February 28, 2023, from [https://www.tps.ca/media/filer\\_public/bd/31/bd318aa9-a7fb-43f8-9453-dfd8a8c7873f/eba9dc03-6917-45ae-8e72-4c2883ffd52b.pdf](https://www.tps.ca/media/filer_public/bd/31/bd318aa9-a7fb-43f8-9453-dfd8a8c7873f/eba9dc03-6917-45ae-8e72-4c2883ffd52b.pdf)
- Foster, Lorne, and Les Jacobs. n.d. "Race & Identity Based Data Collection Strategy." Toronto Police Service. Accessed February 25, 2023.  
[https://www.tps.ca/media/filer\\_public/93/04/93040d36-3c23-494c-b88b-d60e3655e88b/98ccfdad-fe36-4ea5-a54c-d610a1c5a5a1.pdf](https://www.tps.ca/media/filer_public/93/04/93040d36-3c23-494c-b88b-d60e3655e88b/98ccfdad-fe36-4ea5-a54c-d610a1c5a5a1.pdf).
- Government of Canada, D. of J. (2022, December 15). *Overrepresentation of black people in the Canadian Criminal Justice System*. Government of Canada, Department of Justice, Electronic Communications. Retrieved February 28, 2023, from <https://www.justice.gc.ca/eng/rp-pr/jr/obpccjs-spnsjpc/index.html>
- Kirkup, K. (2009). Indocile Bodies: Gender Identity and Strip Searches in Canadian Criminal Law. *Canadian Journal of Law and Society*, 24(1), 107–125.  
<https://doi.org/10.1017/s0829320100009790>
- Koehrsen, W. (2018, July 6). *Visualizing Data with Pairs Plots in Python - Towards Data Science*. Medium. <https://towardsdatascience.com/visualizing-data-with-pair-plots-in-python-f228cf529166>
- OIPRD\_Breaking-the-Golden-Rule\_Report\_Accessible.pdf*. (n.d.). www.oiprd.on.ca. Retrieved February 28, 2023, from [https://www.oiprd.on.ca/wp-content/uploads/OIPRD\\_Breaking-the-Golden-Rule\\_Report\\_Accessible.pdf](https://www.oiprd.on.ca/wp-content/uploads/OIPRD_Breaking-the-Golden-Rule_Report_Accessible.pdf)
- Race & Identity Based Data Collection Strategy Understanding Use of Force & Strip Searches in 2020 Detailed Report Toronto Police Service*. (2022).  
[https://www.tps.ca/media/filer\\_public/93/04/93040d36-3c23-494c-b88b-d60e3655e88b/98ccfdad-fe36-4ea5-a54c-d610a1c5a5a1.pdf](https://www.tps.ca/media/filer_public/93/04/93040d36-3c23-494c-b88b-d60e3655e88b/98ccfdad-fe36-4ea5-a54c-d610a1c5a5a1.pdf)
- rbdc-resources*. (n.d.). Data.torontopolice.on.ca. Retrieved February 28, 2023, from <https://data.torontopolice.on.ca/pages/rbdc-resources>
- Service, T. P. (2022, June 15). *Race-Based Data Collection*. ArcGIS StoryMaps.  
<https://storymaps.arcgis.com/stories/d648e2827cb74ea6bda9c96bd22b5c08>
- Toronto Police Service, P. M. (2022, November 10). *Arrests and strip searches (RBDC-arr-TBL-001)*. Toronto Police Service Public Safety Data Portal. Retrieved February 28, 2023, from <https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about>
- Z. (2022, July 16). *Pandas: How to Convert object to int*. Statology.  
<https://www.statology.org/pandas-convert-object-to-int/>