UNIVERSITY OF TORONTO
FACULTY OF INFORMATION
The iSchool

INF 2178H S LEC0101 20231
Experimental Design for Data Science

Midterm Project

The Analysis of Arrests and Strip Searches (RBDC-ARR-TBL-001)

Group 39
Miaomiao Yang
Sheng Cheng

Feb 27th 2023

Table of Contents

List of Figures

List of Tables

Abstract

Analyzing arrest and strip search data in the Toronto area is particularly important due to the diversity of the population and the potential for disparities in policing practices. The research aims to investigate the potential biases and disparities in policing practices that may disproportionately affect certain communities and promote social harmony by raising and answering two research questions. The effect of occurrence category and perceived race on the frequency of strip searches and the influence of Sex and Age groups on the frequency of Strip Searches. The research methods involve exploratory data analysis, t-test, two-way ANOVA, and Tukey's HSD test. The output shows that the black strip search is highest while whites are lowest. The strip searches on women are more frequent than for men who are from 25 years old to 44 years old, no matter how old she is. The strip search frequency for men equal to or under 17 years old is higher than any other male age group. According to the results, some speculations were made. There might occur bias and stereotype during the law enforcement. These findings could have important implications for promoting social harmony and ensuring that policing practices are fair and equitable for all members of the community, regardless of race, sex, or age. Further research may be needed to better understand the underlying causes of these disparities and to develop strategies to address them.

Introduction

Analyzing arrest and strip search data in the Toronto area is particularly important due to the diversity of the population and the potential for disparities in policing practices. Toronto is a large and diverse city with a population of over 2.7 million people and over 250 different nationalities with 180 different dialects spoken (SAF, 2022). According to 2019 statistics, almost half of the population identifies as a visible minority (defined as either Asian, Black, Latin American, or Arab). In the City of Toronto, 51% of inhabitants were not born in Canada, and the Toronto Area as a whole has the highest percentage of non-Canadians of any metropolitan area in North America (Ryan, 2019). In recent years, the global pandemic and conflicts in certain nations have steadily altered the global pattern, influencing population migration. Due to the fact that the city of Toronto is home to a wide range of cultures and communities, it constantly attracts immigrants and refugees. Analyzing arrest and strip search data in the Toronto area is particularly important due to the diversity of the population and the potential for disparities in policing practices. And ensuring that policing practices are equitable and just for all residents is a crucial part of maintaining a safe and thriving community.

Analyzing arrests and strip searches data in Toronto can help to promote public safety by gaining insights into patterns of criminal activity, any law enforcement interventions that may be needed, and evaluating the effectiveness of current policing practices. For example, racial profiling and discrimination have been identified as issues in Toronto's policing practices, and analyzing data on arrests and strip searches can help identify the extent to which these practices are occurring and inform efforts to address them. Furthermore, analyzing Arrests and Strip Searches data in Toronto can help inform efforts to promote community safety and well-being. By identifying trends and patterns in criminal activity, law enforcement agencies can develop more effective strategies for preventing crime and responding to incidents when they occur.

The objective of the research is to investigate the factors to influence strip searches by identify potential biases and disparities in policing practices that may disproportionately affect certain communities. Two questions could be raised and analyzed to achieve the research objective by analyzing the Toronto Police Service's Arrests and Strip Searches data set (RBDC-ARR-TBL-001). The first research question is to examine the effect of occurrence category and perceived race on the frequency of strip searches. The second research question is to test the influence of Sex and Age groups on the frequency of Strip Searches.

Investigating these two research questions could provide insights into the potential biases or discriminatory practices within strip searches. By identifying patterns or disparities in the frequency of strip searches based on race, sex, or age, researchers and policymakers can work towards implementing measures to reduce such practices and ensure that strip searches are conducted fairly and impartially.

Literature Review

A strip search is an outcome that results from a series of decision points in an arrest process. According to the Video "*Race-Based Data Collection 2020 Findings – Strips Searches: Measurement & Outcomes"* (Strip Searches: Measurement & Outcomes, 2022) published by Toronto Police Service, we are able to understand the importance of applying data science to real-world case analysis and how findings to promote community safety and well-being. In 2020, there were 31,979 arrests made by the Toronto Police Service leading to 17,096 bookings of persons into custody (an individual may be arrested and booked more than once a year) that resulted in 7,114 strip searches. Intuitively, in 2020 from January to December, 53.5% of arrests resulted in a booking and 22% of arrests resulted in a strip search; 41.6% of bookings resulted in a strip search.

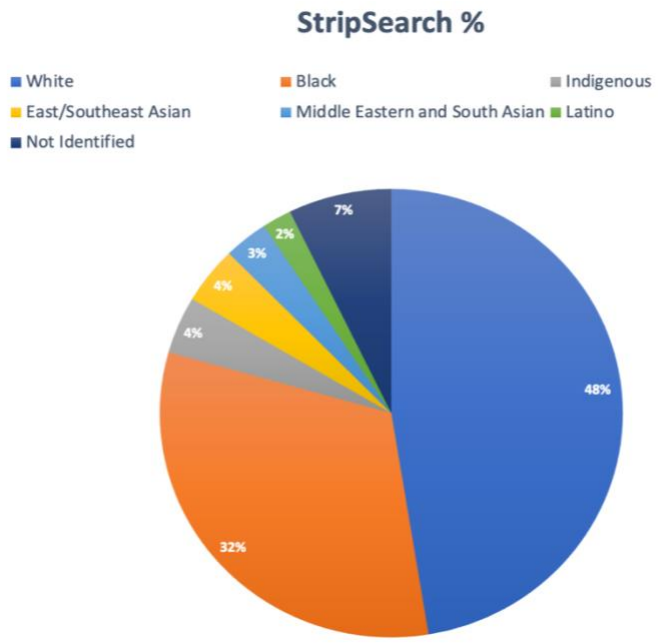

Figure 1.Strip Search Breakdown Graph by Perceived Race

Of the 7,114 strip search cases in 2020, a person may be searched more than once in the reporting period. The breakdown graph of strip searches by the perceived race of individuals is illustrated (*Figure 1*). The Toronto Police Service collaborated with the Community Advisory Panel and Service members to co-develop questions and hypotheses to explore arrests and strip search data in order to better comprehend strip search results and racial disparities.

Four discoveries resulted from the collaborative effort.

- The changes in policy and procedures led to a decline in the frequency of strip searches.
- The number of strip searches by race in 2020 decreased as a result of the new policy.
- There were differences by race after accounting for repeat arrests, including dug-related and weapons offenses.
- There were differences across arrest locations by division in strip searches conducted in 2020.

Among four findings, the first two of them will be listed for interpretation and learning. The graph below shows the total number of arrests and strip searches and illustrates the effect of the provincial pandemic response on all patterns in the months after the emergency declaration. Arrests dropped as people were required indoors and outside activities were prohibited. This also reflected in the decline in bookings and strip searches. But there was a slow rebound over the summer. The impact of changes in search policy initiatives took place

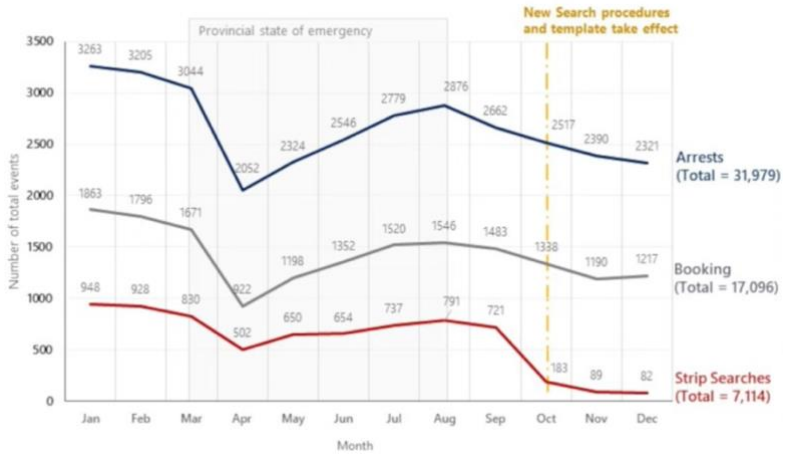

Figure 2. The Rate Change of Strip Search in 2020

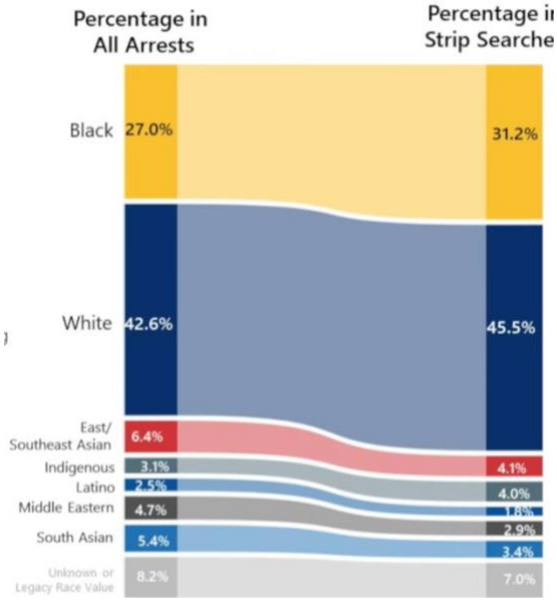

Figure 3. Differences in Arrests Rates VS. Strip Search Rates by Race in 2020.

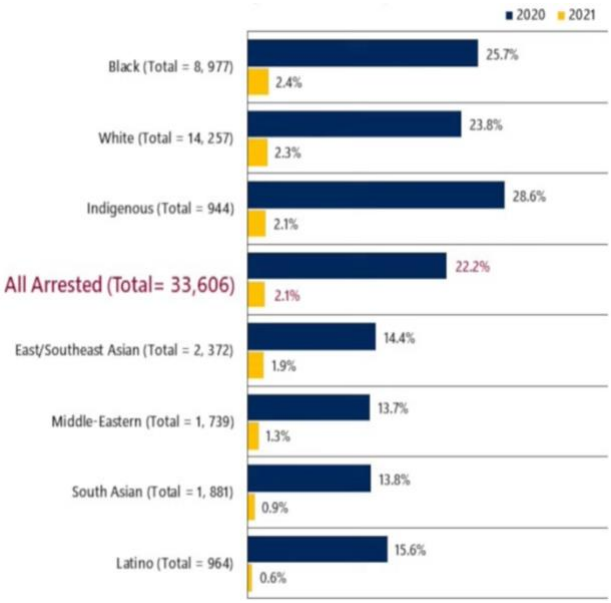

Figure 4. Comparation of Strip Search Rate in 2020 & 2021

in October; notably, there was a considerable drop in strip searches that started a month earlier as officers got educated.

In order to determine if a procedural adjustment lowered disproportionalities, comparing the strip searches by race group to that group's proportion in 2020 arrests and 2021 data. Break down the arrested population in 2020 by race can help people understand the broader process leading to a strip search. By comparing the proportion of each race group arrested with their proportion in strip searches, it clearly indicated Black, Indigenous, and White people were over-represented relative to their presence in arrests. The indigenous people were 1.3x over-represented in strip searches, while black and white people arrested were over-presented by 1.1x. It is essential to analyze 2020 data to provide a baseline for assessing the effects of changes and monitoring the progress. Comparing the findings to 2021 to determine the impact of policy changes on racial inequalities. As seen by the blue bars for 2020 and the yellow bars for 2021, the number and percentage of strip searches have decreased dramatically for all individuals (*Figure 4*). Yet, somewhat higher strip search rates were still detected among Black and White individuals arrested in 2021 compared to the national average.

Exploratory Data Analysis (EDA)

Data Cleaning

With specific research objective and questions, the dependent variable "`Strip Search`" and independent variables "`Perceived_race'`, "`Sex`", "`Age_group_at_arrest_`", and "`Occurrence_Category`" were identified. Through `value_counts ()` function to calculate the count of ubique values in each selected variable and thereby obtain the percentage of unique values. In the dataset, there are eight races, three genders, nine age groups, and over thirty occurrence categories. Carefully go through each printout, several possible repetitive terms were found.

- '`Aged 17 years and younger`' and '`Aged 17 years and under.`'
- '`Aged 65 years and older`' and '`Aged 65 and older.`'
- '`Break and Enter`' and '`Break & Enter.`'

In order to prevent an overestimation of the true significance of the relationship between variables and biased results in statistical models, it is essential for researchers to verify repetitive data. The '`EventID`' and '`ArrestID`' are used to check on the repeated observations through `duplicated ()` function.

After screening the dataset, we can conclude there are no repeated observations since 'NaN' means nothing. Updating the table by merging rows of each two similar terms and discarding columns that will not aim at research objective. For the purpose of further increasing the accuracy of results and performance of algorithms, examining 'NaN' values in 'StripSearch' column by the apply () function used to apply *lambda method* to each value. Next, creating a new 'stripbinary' column which contain binary values indicating whether a strip search was performed (1) or not (0). The cleaned dataset (*Table 1*) indicated the sum of strip searches were performed for each combination of the four columns.

Table 1. Dataset for Analysis

	Occurrence_Category	Perceived_Race	Age_group__at_arrest_	Sex	StripSearchFreq
0	Assault	Black	Aged 17 years and younger	F	0
1	Assault	Black	Aged 17 years and younger	M	0
2	Assault	Black	Aged 18 to 24 years	F	1
3	Assault	Black	Aged 18 to 24 years	M	3
4	Assault	Black	Aged 25 to 34 years	F	1

Descriptive Statistics

Descriptive statistics are important in analyzing and understanding a dataset, as they allow us to summarize and describe the characteristics and patterns of the data. It can be useful in two ways, one of which is to provide basic information about variables in a dataset and another of which is to highlight potential relationships between variables. The descriptive statistics of one dependent variable "Strip Search" and four independent variables "Perceived_race', "Sex", "Age _group_at_arrest_", and "Occurrence_Category" were organized in tables below. According to *Table 2*, the mean of 3.134646 indicates that the average frequency of strip searches is 3.13, while the standard deviation of 12.486678 indicates that the frequency of strip searches is quite variable, with values ranging from 0 to 177. The data provided in *Figure 5* shows percentage of strip searches for different perceived racial groups. Descriptive statistics for this data could include measures such as

Table 2. The Descriptive Statistics of Dependent Variable 'Strip Search Frequency'

	M	*SD*	*RANGE*
STRIP SEARCH FREQUENCY	3.134646	12.486678	0-177

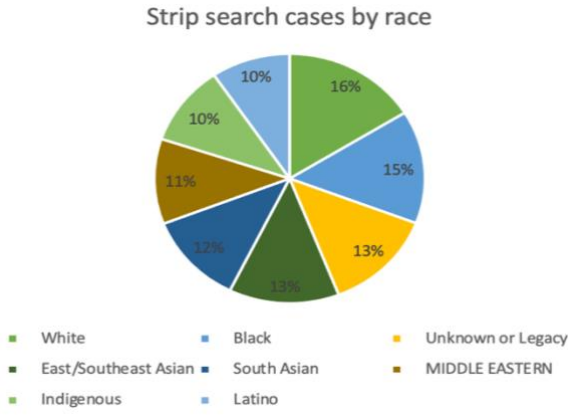

Figure 5. Strip Search Percentage of Race

the mean, standard deviation, and range, but these would not be applicable to this type of categorical data. For example, there were 402 strip searches for white people and 365 for black people. Whites represent 16.16% of all strip searches, while blacks represent 14.67%. there is no big difference among all race groups. Similar to the race, genders, age groups, and occurrence categories are categorical data. Male strip search frequency is higher than female. From *Figure 6*, the trend for strip search frequency increases with age increase and reaches the peak in the age group from 25 to 34 years old, then gradually decreases. *Figure 7 shows the strip searches case on all occurrence categories.* There were 110 strip searches for people arrested for "Robbery & Theft" which is the highest and represent 4.42% of all Strip searches. Full tables of variable are attached in **Appendix**.

Table 3. The Descriptive Statistics of Independent Variable 'Sex'

SEX	*FREQUENCY*	*%*
M	1450	58.279743
F	1030	41.398714
U	8	0.321543

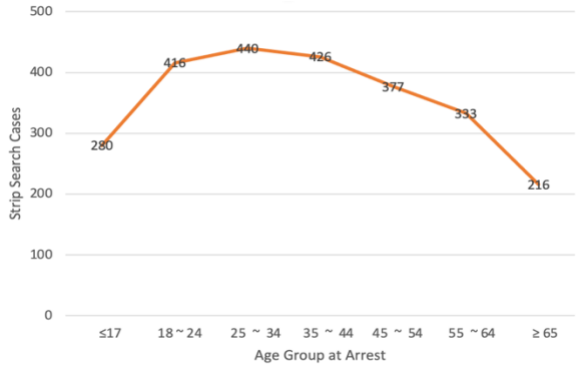

Figure 6. Strip Search Cases of different ' Age Group at Arrest'

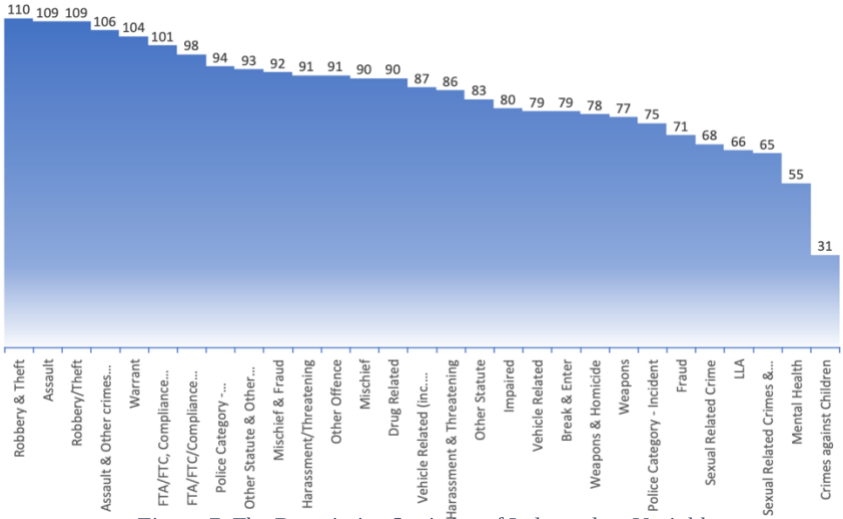

Figure 7. The Descriptive Statistics of Independent Variable ' Occurrence Category'

In this case, the stacked bar chart was used to shows the comparison of the composition of different occurrence category across race. Each bar is divided into multiple segments, with each segment representing a different occurrence category within the different race. The figure below (*Figure 8*) gives us a general idea of how much of each race is composed of each occurrence category. Obviously, blacks and whites have the highest cases of strip searches. There are ten occurrence categories in the black while fifteen occurrence categories in the white. Among all the categories, *'warrant', 'Robbery & Theft', 'FTA/FTC, Compliance Check & Parollee", Drug Related', and 'Assault & Other crimes against persons'* account for the majority of the proportion for blacks and whites which means those categories are more prevalent than other categories in blacks and whites.

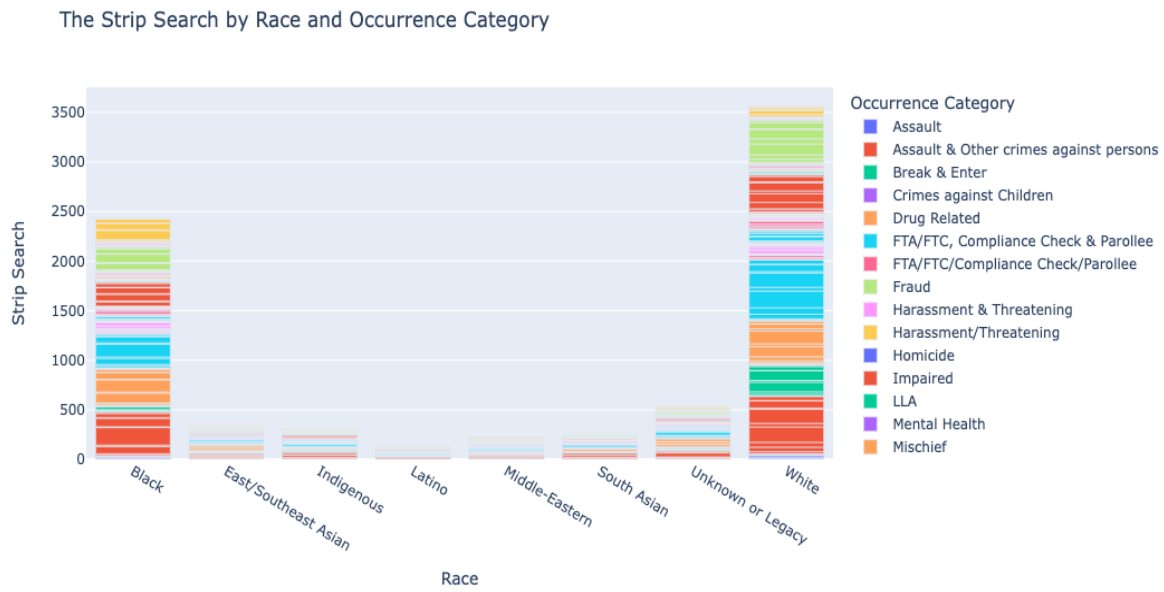

Figure 8. The Strip Search by Race and Occurrence Category (Stacked bar chart)

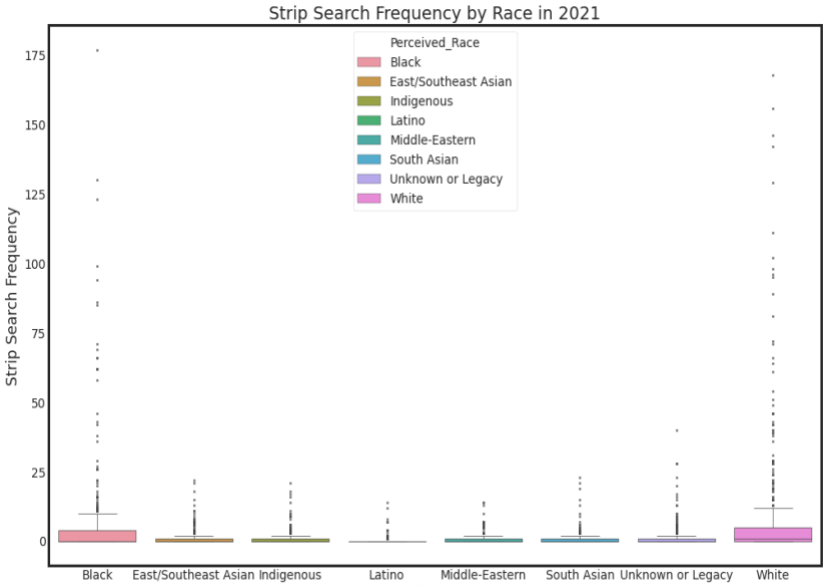

Figure 9. Strip Search Frequency by Race in 2021 (Boxplot)

From boxplot, except for White, all other racial medians are roughly 0. The distributions of all the races are right skewed, and there are some outliers in the distribution. The distributions of 'Unknown or Legacy,' 'East/Southeast Asian,' 'South Asian,' 'Middle Eastern,' and 'Indigenous' are roughly similar. The White median is greater than those of other races.

T – Test

Before performing a two-way ANOVA test, first, determine if there is a significant difference between the two groups. If there is no significant difference between the two groups, then the chosen variable may not be necessary to conduct further complex analysis. In this section, four two-sample t-tests will be conducted to examine whether there is a significant difference between the means of the two groups. In order to form two samples, two samples of each variable were extracted based on the characteristics of

the variable. Hypothesises were raised and listed. Importing *scipy.stats* module and use `ttest_ind()` function to perform the t-test successfully performing each t-test.

'Assault & Other crimes against persons' and 'FTA/FTC, Compliance Check/ Parollee'.

Random selecting 'Assault & Other crimes against persons' and 'FTA/FTC, Compliance Check/ Parollee' as two samples to examine the significance.

Ho: There is no mean difference in the strip search frequency between the people arrested for 'Assault & Other crimes against persons' and 'FTA/FTC, Compliance Check/ Parollee'.

Ha: There is mean difference in the strip search frequency between the people arrested by 'Assault & Other crimes against persons' and 'FTA/FTC, Compliance Check/ Parollee'.

Table 4. T-Test Results of 'Assault' and 'Robbery & Theft'

T-statistic	4.20360123043
p-value	3.94406434114e-05

A positive t-statistic value suggests that there is evidence to support the alternative hypothesis that the means of the two groups are not equal and that the mean of the 'Assault & Other crimes against persons' is greater than the mean of the 'FTA/FTC, Compliance Check/ Parollee". In this case, the significance level of 0.05 is used, meaning that the p-value of 3.94406434114e-05 is less than 0.05, thus rejecting the null hypothesis and concluding that there is a significant difference between the means of 'Assault & Other crimes against persons' and 'FTA/FTC, Compliance Check/ Parollee'.

Whites and Non-whites

Ho: There is no mean difference in the strip search frequency between the people who are White and those who are non-white.

Ha: There is mean difference in the strip search frequency between the people who are white and those who are non-White.

Table 5. T-Test Results of Whites and Non-whites

T-statistic	10.2677404968
p-value	2.98173369586e-24

In this case, the t-statistic of 10.267 indicates that there is a large difference between the means of the whites and non-whites. With a p-value of 2.98173369586e-24, the likelihood of detecting a difference as great or larger than the one seen by chance is exceedingly low if the null hypothesis is true. This provides strong evidence to reject the null hypothesis and conclude that the means of the two groups whites and non-whites are significantly different.

Age Groups under 35 and older 35 Years

H0: There is no mean difference of the strip search frequency between age groups under 35 and older 35 years.

H1: There is a mean difference of the strip search frequency between age groups under 35 and older 35 years.

Table 6. T-test Results of Age Group under 35 and older 35 Years.

T-statistic	2.83631882561
p-value	0.004609856353

In this case, the t-statistic of 2.836 indicates that there is a moderate difference between the means of the two groups. The p-value of 0.0046 is less than a chosen significance level, which provides evidence to reject the null hypothesis and conclude that the means of the two groups are significantly different.

Male and Female

H0: There is no mean difference of the strip search frequency between Male and Female.

H1: There is a mean difference of the strip search frequency between Male and Female.

Table 7. T-test Results of Male and Female

T-statistic	6.42558605834
p-value	1.5699842451e-10

In this case, the t-statistic of 6.425 indicates that there is a relatively large difference between the means of the Male and Female. The p-value of 1.5699842451e-10 is very small which is sufficient evidence to reject the null hypothesis and conclude that the means of the two gender groups male and female are significantly different.

Four two-sample t-tests were completed, all indicating a significant difference between the means of the two groups being compared. Further two-way ANOVA can be used to examine the effects of occurrence

category and perceived race on the frequency of strip searches and test the impact of the Sex and Age group on the frequency of strip searches.

Methodology

Dataset Description

The studied dataset Arrests and Strip Searches (RBDC-ARR-TBL-001) contains information on over 600,000 arrests on arrests and strip searches conducted by the Toronto Police Service between 2020 and 2021. The variables in the dataset include demographic information such as age, gender, and perceived race, as well as information on the location and reason for the arrest, actions taken by the police, and whether a strip search was conducted. Besides demographic information, it also includes variables such as the year of the arrest, event ID, arrest ID, and person ID, as well as variables related to actions taken during the arrest, such as whether the person was combative or resisted, and whether they were cooperative. The dataset also includes information on whether items were found during the search. By analyzing this dataset and answering the research questions, we can better understand potential disparities in policing practices and inform efforts to promote more equitable and just systems.

Statistical Methods

Before running two-way ANOVA for each group, confirm the assumptions of normality, homogeneity of variance, and independence of observations. If these assumptions are met, a two-way ANOVA can be used to test for the effects of the two categorical variables on the continuous outcome variable. The Tukey's HSD Tests followed after the two-way ANOVA test can be used to determine which groups are significantly different from each other. The test calculates the difference between the means of all possible pairs of groups and compares them to a critical value based on the sample size and number of groups.

Two-way ANOVA

The central limit theorem states that if the sample size is large enough (typically $n > 30$), the sampling distribution of the means will be approximately normal, regardless of the underlying distribution of the population (LaMorte, 2016). The sample size of the data used in this ANOVA test is 204 which is much more than 30. It is sufficient to assume that the data follows a normal distribution. Data from people who have been arrested by police, then it is likely that each individual in the sample is independent of the

other, assuming that each arrest is unrelated to the others. All the groups of the data are selected from the same population, so it is sufficient to conclude that the variance of each group is approximately the same. Thus, all the conditions are met, two-way ANOVA test could be processed.

Load in statistical package *statsmodels* library to set up and run the two-way ANOVA. The code in *Google Colab* fits a two-way ANOVA model with the frequency of strip searches as the dependent variable and perceived race, occurrence category, gender, and age group as the independent variables. The `anova_lm()` function can generate ANOVA tables, which show the main effects and interaction effects of perceived race and occurrence category and sex and age group on the frequency of strip searches.

Tukey's HSD Tests

As the results from the two-way ANOVA test indicates the significant effects of the independent variables on the dependent variable which would be detail presented in the section of **Results**, the following Tukey's HSD Tests can be applied to determine which groups are significantly different from each other.

To perform Tukey's HSD test, uses the *statsmodels* library same as the two-way ANOVA test. The `MultiComparison ()` function sets up multiple comparisons for the test. In the case of examining the effect of race and occurrence category on strip search frequency, the comparisons are based on the combination of the perceived race and occurrence category variables, which are concatenated into a single variable called '`Race_Occurrence.`' The `tukeyhsd ()` function is then applied to run the Tukey HSD test and calculate the confidence intervals and p-values for each pairwise comparison.

Results

Research Question 1: the effect of race and occurrence category on strip search frequency.

Table 8. Two-way ANOVA Test Result of Q1

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	12884.884134	7	4.996942	0.000033
C(Occurrence_Category)	8596.708586	1	23.337486	0.000003
C(Perceived_Race):C(Occurrence_Category)	9458.227929	7	3.668036	0.000975
Residual	69252.583933	188	NaN	NaN

The ANOVA table above (*Table 8*) shows that perceived race and occurrence category significantly affect the frequency of strip searches, with p-values of 0.000033 and 0.000003, respectively. The interaction between perceived race and occurrence category is also significant, with a p-value of 0.000975. A higher F-statistic indicates a stronger effect, while a lower p-value suggests a greater likelihood that the effect is not due to chance. The residual row in the ANOVA table shows the sum of squares and degrees of freedom for the error term, representing the amount of unexplained variation in the model. The F-statistic and p-value for the residual row are not applicable, as they are based on a comparison of the residual mean square to the error mean square, which is undefined.

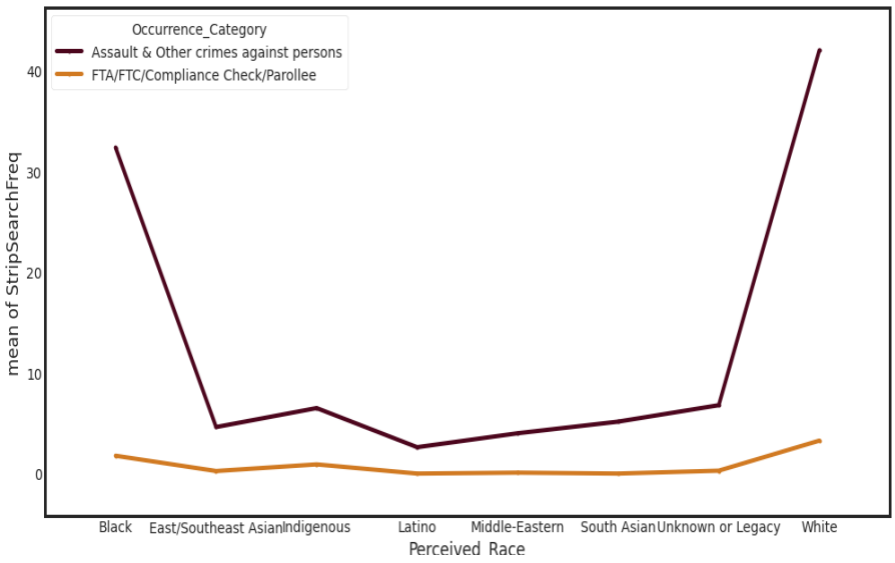

Figure 10. Interaction Plot of Race and Mean of Strip Search Frequency

Unparallel lines indicates that there is an interaction effect between perceived races and Occurrence Categories. In this case, the effect of perceived races on the strip search frequency depends on the level of Occurrence Categories. The line for 'Assault & Other crimes against persons' is consistently higher than the line for "FTA/FTC/Compliance Check/Parollee", it suggests that "Assault & Other crimes against persons" tend to get more strip search than "FTA/FTC/Compliance Check/Parollee", regardless of the age groups. The difference between the lines varies depending on the perceived races, it suggests that the effect of the races on the strip search frequency is different for "Assault & Other crimes against persons" and "FTA/FTC/Compliance Check/Parollee."

Table 9. Tukey's HSD Test 'TRUE' Result of Q1

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	lower	upper	p-adj	reject
Black_Assault & Other crimes against persons	Black_FTA/FTC/Compliance Check/Parollee	-30.5879	-56.264	-4.9118	0.0052	TURE
Black_Assault & Other crimes against persons	East/Southeast Asian_Assault & Other crimes against persons	-27.7418	-53.4178	-2.0657	0.0205	TURE
Black_Assault & Other crimes against persons	East/Southeast Asian_FTA/FTC/Compliance Check/Parollee	-32.1071	-58.3321	-5.8822	0.0034	TURE

The *Tukey HSD Test* exams multiple comparisons of means with a family-wise error rate (FWER) of 0.05. The first two columns of the table indicate the groups being compared. For each pair of groups, the table shows the difference in means between them (meandiff), the p-value (p-adj), the lower and upper bounds of a confidence interval for the difference in means, and whether or not the null hypothesis should be rejected at the FWER of 0.05. If the rejected column is "True," it means that the null hypothesis should be rejected; however, if it is "False," the null hypothesis cannot be rejected, meaning there is not enough evidence to conclude that the means are different.

Looking at specific row 1, it shows there is statistically significant evidence that the mean number of crimes committed for '`Assault & Other crimes against persons`' is lower than the mean number of crimes for '`FTA/FTC/Compliance Check/Parollee`' among Blacks (reject = TRUE, p-adj = 0.0052). The full Q1 results of Tukey's HSD Test can be checked on *Google Colab*. The fully organized Q1 TRUE result is attached in **Appendix**.

Make short of long, the strip search frequency of "`Assault & Other crimes against persons`" is higher than "`FTA/FTC/Compliance Check/Parollee`" for both blacks and whites. For both categories, the strip search frequency on blacks is higher than east/southeast Asian, south Asia, middle eastern, Latino, and Unknown races, while whites are lower than not only those mentioned races but also indigenous people.

Research Question 2: the effect of genders and age groups on strip search frequency.

The ANOVA table (*Table 10*) indicates age group at arrest has a significant main effect on the frequency of strip searches, with a p-value of 7.417785e-17. However, sex does not have a significant main effect, as indicated by the high p-value of 9.956259e-01. There is also a significant interaction effect between sex and age group at arrest, with a p-value of 9.578688e-03. A higher F-statistics and a lower p-value indicate a greater likelihood that the effect is not due to chance.

Table 10. Two-way ANOVA Test Result of Q2

	sum_sq	df	F	PR(>F)
C(Sex)	0.008894	2	0.000030	9.956259e-01
C(Age_group__at_arrest_)	18436.396157	6	20.770821	7.417785e-17
C(Sex):C(Age_group__at_arrest_)	4155.173265	12	2.340652	9.578688e-03
Residual	365399.611865	2470	NaN	NaN

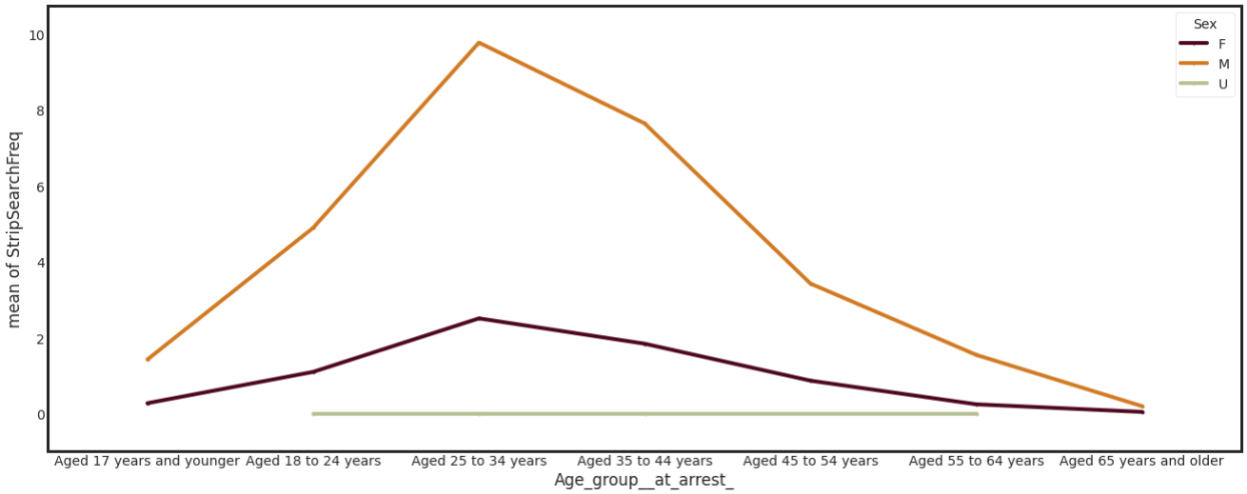

Figure 11. Interaction Plot of Age Group and Mean of Strip Search Frequency

The lines are not parallel here. It indicates that there is an interaction effect between age groups and Sex. In this case, the impact of age groups on the strip search frequency depends on the level of Sex. The line for females is consistently higher than for males, which suggests that females tend to get more strip searches than males, regardless of age group. The difference between the lines varies depending on the age groups; it suggests that the effect of the age groups on the strip search frequency differs for males and females.

The output from *Tukey's HSD Test* has several columns (*Table 11*). The first two columns indicate the groups being compared. The "meandiff" column shows the difference in means between the two groups being compared. The "p-adj" column shows the adjusted p-value for each comparison, which is used to determine whether the difference in means is statistically significant. The "lower" and "upper" columns show the lower and upper bounds of the confidence interval for the mean difference. Finally, the "reject" column indicates whether the null hypothesis of no difference in means is rejected at a significance level of 0.05 (i.e., whether the difference in means is statistically significant). If the "reject" column says "True," it means that the difference in means between the two groups is statistically significant. If it says "False," then the difference is not statistically significant.

Table 11. Tukey's HSD Test 'TRUE' Result of Q2

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
F_Aged 17 years and younger	F_Aged 18 to 24 years	0.8243	0.9	-4.2958	5.9443	False
F_Aged 17 years and younger	F_Aged 25 to 34 years	2.232	0.9	-2.8259	7.29	False
F_Aged 17 years and younger	F_Aged 17 years and younger	1.5651	0.9	-3.5892	6.7194	False

In the first row, the mean difference between the age groups "F_Aged 17 years and younger" and "F_Aged 18 to 24 years" is 0.8243, but the p-value is 0.9, which is greater than the significance level of 0.05. Therefore, there is no significant difference between the two age groups, and the null hypothesis that the means are equal cannot be rejected. The full Q2 results of Tukey's HSD Test can be checked on *Google Colab*. The full organized Q2 TRUE result was attached in **Appendix**.

Results of Research Question 2 could be concluded:

- For women, the strip searches are more frequent than for man, no matter how old she is.
- The strip search frequency for man equal or under 17 years old is higher than men from 25 years old to 44 years old.
- The strip search frequency for men from 25 years old to 44 years old and higher than men from 45 years old to equal or older than 65 years old.
- Men from 18 years old to 24 years old has higher strip search frequency than men from 25 years old to 34 years old and equal or older 65 years old.

Discussion

According to the results, there may be several possible reasons, like racial profiling, crime types, and bias for the observed differences in strip search frequency. Even if they have not committed more severe offenses, law enforcement personnel may be more inclined to strip-search some specific ethnic groups, such as blacks. The observed differences in strip search frequency between the two types of crimes may be since certain types, such as "Assault & Other crimes against persons", may be more likely to involve physical altercations and therefore require more extensive searches.

When we look at bias, we need to think about it in many ways, like historical data bias, reporting bias, and systemic bias. Historical prejudice in law enforcement might contribute to the over-policing and overcriminalization of specific ethnic groups. The observed differences in strip search frequency between different races and types of crimes may be due in part to differences in reporting practices. For instance, police may be more prone than others to record strip searches on members of specific ethnic groups or for certain sorts of offenses. Systemic biases in the criminal justice system may contribute to the observed differences in strip search frequency between different races and types of crimes. For example, certain

laws or policies may disproportionately affect certain racial groups, or there may be differences in the availability of legal resources or representation for different groups.

It was an unexpected result that females of all ages are searched more often than most males from 25 to 44 years old. This higher frequency of strip searches could be due to their generally lower representation in the criminal justice system, making them more noticeable to officers and more likely to be subject to searches. It could also be due to the stereotypes and biases of officers regarding women's criminal behavior, leading them to suspect women more often. This does not, however, rule out the possibility that certain police officers engage in deliberate sexual harassment throughout the enforcement process; this is the question that desires further investigation.

The higher strip search frequency for younger men (age 17 and under) are more likely to be subjected to strip searches, maybe because they are overrepresented in the criminal justice system unjustly targeted by law enforcement. Moreover, there may be a presumption that younger males are more likely to be engaged in criminal activities, resulting in more searches.

Conclusion

During the process of data cleaning, it was observed that all variables were in categorical form, posing a challenge in implementing certain statistical tests such as t-tests and ANOVA tests, which require continuous dependent variables. To address this issue, a continuous variable was created based on a binary variable by grouping it with several discrete variables. However, this method comes with limitations, including the introduction of bias in the analysis. Specifically, when the sum of the binary variable is zero, it is overlooked, irrespective of the number of zeros present in the sum. These limitations should be considered when interpreting the analysis results, and it is recommended to exercise caution in drawing conclusions based on the limitations of the approach used for creating the continuous variable. Further research is needed to explore alternative techniques for dealing with categorical variables in statistical analysis to overcome these limitations.

It is important to note that the above-mentioned **Discussion** is only speculation based on outputs. Possible biases in the data collection and analysis could include selective enforcement or profiling by law enforcement officers, inaccurate or incomplete reporting of search data, or systematic differences in how different groups are searched or treated by law enforcement. These biases could lead to overestimates

or underestimates of the true frequency of strip searches for different groups and age ranges. Further analysis and research would be needed to better understand the causes and potential biases in the observed patterns.

In general, the midterm project provides us with an excellent opportunity to explore the nature of the problem through data in a human-centered manner.

Reference

Greg Moreau, C. C. (2022, 8 2). *Police-reported crime statistics in Canada, 2021*. Retrieved from Statistics Canada: https://www150.statcan.gc.ca/n1/pub/85-002-x/2022001/article/00013-eng.htm

LaMorte, W. W. (2016, July 24). *Central Limit Theorem*. Retrieved from The Role of Probability: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html

Ryan, A. (2019, 3 22). *TORONTO GLOBAL*. Retrieved from HOW MULTICULTURAL IS TORONTO? LET US COUNT THE WAYS…: https://torontoglobal.ca/TG-Blog/March-2019/How-multicultural-is-Toronto-Let-us-count-the-way

SAF. (2022, 11 8). Retrieved from Why Toronto Is the Most Multicultural City in the World: https://www.studyabroadfoundation.org/blogs/why-toronto-most-multicultural-city-world#:~:text=Recognized%20by%20both%20the%20United,nationalities%20from%20around%20the%20world.

Strip Searches: Measurement & Outcomes. (2022, 6 14). *Race-Based Data Collection 2020 Findings*. City of Toronto, Ontario, Canada: Toronto Police Service.

Appendix

Table 12. Strip Search Frequency and Percentage of 'Occurrence Category', 'Perceived Race' and 'Age Group'

	Frequency	*%*
Occurrence_Category		
Robbery & Theft	110	4.421222
Assault	109	4.381029
Robbery/Theft	109	4.381029
Assault & Other crimes against persons	106	4.260450
Warrant	104	4.180064
FTA/FTC, Compliance Check & Parollee	101	4.059486
FTA/FTC/Compliance Check/Parollee	98	3.938907
Police Category - Administrative	94	3.778135
Other Statute & Other Incident Type	93	3.737942
Mischief & Fraud	92	3.697749
Harassment/Threatening	91	3.657556
Other Offence	91	3.657556
Mischief	90	3.617363
Drug Related	90	3.617363
Vehicle Related (inc. Impaired)	87	3.496785
Harassment & Threatening	86	3.456592
Other Statute	83	3.336013
Impaired	80	3.215434
Vehicle Related	79	3.175241
Break & Enter	79	3.175241
Weapons & Homicide	78	3.135048
Weapons	77	3.094855
Police Category - Incident	75	3.014469
Fraud	71	2.853698
Sexual Related Crime	68	2.733119
LLA	66	2.652733
Sexual Related Crimes & Crimes Against Children	65	2.612540
Mental Health	55	2.210611
Crimes against Children	31	1.245981
Homicide	30	1.205788
Perceived_Race		
White	402	16.157556
Black	365	14.670418
Unknown or Legacy	330	13.263666
East/Southeast Asian	328	13.183280
South Asian	294	11.816720
Middle-Eastern	275	11.053055
Indigenous	259	10.409968
Latino	235	9.445338
Age_group__at_arrest_		
Aged 25 to 34 years	440	17.684887
Aged 35 to 44 years	426	17.122186
Aged 18 to 24 years	416	16.720257
Aged 45 to 54 years	377	15.152733
Aged 55 to 64 years	333	13.384244
Aged 17 years and younger	280	11.254019
Aged 65 years and older	216	8.681672

Table 13. Tukey's HSD 'TURE' Result of Q1

Multiple Comparison of Means - Tukey HSD, FWER=0.05					
group1	group2	meandiff	lower	upper	p-adj
Black_Assault & Other crimes against persons	Black_FTA/FTC/Compliance Check/Parollee	-30.5879	-56.264	-4.9118	0.0052
Black_Assault & Other crimes against persons	East/Southeast Asian_Assault & Other crimes against persons	-27.7418	-53.4178	-2.0657	0.0205
Black_Assault & Other crimes against persons	East/Southeast Asian_FTA/FTC/Compliance Check/Parollee	-32.1071	-58.3321	-5.8822	0.0034
Black_Assault & Other crimes against persons	Indigenous_FTA/FTC/Compliance Check/Parollee	-31.4481	-58.3072	-4.5889	0.0068
Black_Assault & Other crimes against persons	Latino_Assault & Other crimes against persons	-29.7418	-55.4178	-4.0657	0.008
Black_Assault & Other crimes against persons	Latino_FTA/FTC/Compliance Check/Parollee	-32.3571	-59.2163	-5.498	0.0044
Black_Assault & Other crimes against persons	Middle-Eastern_Assault & Other crimes against persons	-28.3571	-54.0332	-2.6811	0.0155
Black_Assault & Other crimes against persons	Middle-Eastern_FTA/FTC/Compliance Check/Parollee	-32.2662	-59.1254	-5.4071	0.0046
Black_Assault & Other crimes against persons	South Asian_Assault & Other crimes against persons	-27.2033	-52.8794	-1.5272	0.0261
Black_Assault & Other crimes against persons	South Asian_FTA/FTC/Compliance Check/Parollee	-32.3571	-59.2163	-5.498	0.0044
Black_Assault & Other crimes against persons	Unknown or Legacy_Assault & Other crimes against persons	-25.5714	-50.7675	-0.3753	0.0428
Black_Assault & Other crimes against persons	Unknown or Legacy_FTA/FTC/Compliance Check/Parollee	-32.0714	-57.2675	-6.8753	0.0017
Black_Assault & Other crimes against persons	White_FTA/FTC/Compliance Check/Parollee	-29.0905	-53.8631	-4.3179	0.0065
Black_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	40.2308	14.5547	65.9068	0.001
East/Southeast Asian_Assault & Other crimes against persons	White_Assault & Other crimes against persons	37.3846	11.7085	63.0607	0.001
East/Southeast Asian_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	41.75	15.5251	67.9749	0.001
Indigenous_Assault & Other crimes against persons	White_Assault & Other crimes against persons	35.5	9.2751	61.7249	0.001
Indigenous_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	41.0909	14.2318	67.95	0.001
Latino_Assault & Other crimes against persons	White_Assault & Other crimes against persons	39.3846	13.7085	65.0607	0.001
Latino_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	42.0	15.1409	68.8591	0.001
Middle-Eastern_Assault & Other crimes against persons	White_Assault & Other crimes against persons	38.0	12.3239	63.6761	0.001
Middle-Eastern_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	41.9091	15.05	68.7682	0.001
South Asian_Assault & Other crimes against persons	White_Assault & Other crimes against persons	36.8462	11.1701	62.5222	0.001
South Asian_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	42.0	15.1409	68.8591	0.001
Unknown or Legacy_Assault & Other crimes against persons	White_Assault & Other crimes against persons	35.2143	10.0182	60.4104	0.001
Unknown or Legacy_FTA/FTC/Compliance Check/Parollee	White_Assault & Other crimes against persons	41.7143	16.5182	66.9104	0.001
White_Assault & Other crimes against persons	White_FTA/FTC/Compliance Check/Parollee	-38.7333	-63.5059	-13.9607	0.001

Table 14. Tukey's HSD 'TURE' Result of Q2

Multiple Comparison of Means - Tukey HSD, FWER=0.05					
group1	group2	meandiff	lower	upper	p-adj
F_Aged 17 years and younger	M_Aged 25 to 34 years	9.4885	4.5175	14.4596	0.001
F_Aged 17 years and younger	M_Aged 35 to 44 years	7.3609	2.3865	12.3354	0.001
F_Aged 18 to 24 years	M_Aged 25 to 34 years	8.6643	4.5175	12.8111	0.001
F_Aged 18 to 24 years	M_Aged 35 to 44 years	6.5367	2.3857	10.6876	0.001
F_Aged 25 to 34 years	M_Aged 25 to 34 years	7.2565	3.1867	11.3264	0.001
F_Aged 25 to 34 years	M_Aged 35 to 44 years	5.1289	1.0549	9.203	0.0016
F_Aged 35 to 44 years	M_Aged 25 to 34 years	7.9234	3.7344	12.1124	0.001
F_Aged 35 to 44 years	M_Aged 35 to 44 years	5.7958	1.6027	9.9889	0.001
F_Aged 45 to 54 years	M_Aged 25 to 34 years	8.8959	4.4642	13.3277	0.001
F_Aged 45 to 54 years	M_Aged 35 to 44 years	6.7683	2.3327	11.2039	0.001
F_Aged 55 to 64 years	M_Aged 25 to 34 years	9.5178	4.7209	14.3147	0.001
F_Aged 55 to 64 years	M_Aged 35 to 44 years	7.3902	2.5897	12.1906	0.001
F_Aged 65 years and older	M_Aged 25 to 34 years	9.7167	3.7494	15.684	0.001
F_Aged 65 years and older	M_Aged 35 to 44 years	7.5891	1.619	13.5593	0.0013
M_Aged 17 years and younger	M_Aged 25 to 34 years	8.339	4.0638	12.6141	0.001
M_Aged 17 years and younger	M_Aged 35 to 44 years	6.2113	1.9322	10.4905	0.001
M_Aged 18 to 24 years	M_Aged 25 to 34 years	4.8735	0.9165	8.8305	0.0024
M_Aged 18 to 24 years	M_Aged 65 years and older	-4.6928	-9.1384	-0.2473	0.0261
M_Aged 25 to 34 years	M_Aged 45 to 54 years	-6.3408	-10.3285	-2.353	0.001
M_Aged 25 to 34 years	M_Aged 55 to 64 years	-8.2169	-12.2614	-4.1724	0.001
M_Aged 25 to 34 years	M_Aged 65 years and older	-9.5663	-14.0157	-5.117	0.001
M_Aged 35 to 44 years	M_Aged 45 to 54 years	-4.2132	-8.2052	-0.2211	0.0261
M_Aged 35 to 44 years	M_Aged 55 to 64 years	-6.0893	-10.1381	-2.0406	0.001
M_Aged 35 to 44 years	M_Aged 65 years and older	-7.4387	-11.8919	-2.9855	0.001