The Statistical Tests of Arrests and Strip Searches Dataset for Factors Impact Arrest Cooperativeness Portion and Strip Search Proportion

Group 19:

- Teng-Chieh Chuang (Student# 1004744752)
- Danni Yang (Student# 1009535558)

Introduction

This report demonstrates the study of arrests and strip search dataset, under the rule of the 8-step data analysis cycle. Firstly, the group decided to do the research from two different perspectives. Then the group members divided the work and did the data cleaning as per the requirements of each hypothesis. Since the target feature "proportion" was not provided from the original dataset, we calculated the proportion results and set it as a new column. Since we are required to apply T-test, ANOVA test, and post-hoc test to explore the data, there is not much we could do to select models and do the evaluation. The data visualization section transfers the mathematical result to a piece of more direct and understandable information. We did the assumption check to see whether the research question was answered or not. Finally, adjustments were applied a thousand times to get the accepted results.

Research Questions

- RQ1: Whether the proportion of strip search differs from race and sex groups?
- RQ2: Is there a statistically significant difference in cooperativeness between different race groups? Which race group exhibits the highest level of cooperativeness?

Given the data on how criminals of different backgrounds have behaved and their actions at arrest, we want to investigate any differences in cooperativeness between these criminals. Given our preliminary analysis, we want to focus on how cooperative scores may differ between criminals of different race groups. Besides, since the dataset provides detailed personal information about the crimes while the strip search does not happen every time, we would like to

find out whether the race and sex information would impact the police decision-making for performing strip search or not.

Literature Review

Helmer (2001) stated that strip searches without reasonable suspicion are unconstitutional and should be prohibited since the strip search was commonly applied in cases in many American cities where the police have the permit to do so for anyone arrested for any crime. He suggested that strip searches of felony detainees are permissible only if there is reasonable suspicion that the detainee is carrying contraband or weapons and should be taken a balancing test before searching. On the other hand, Newburn, Shiner, and Hayman (2004) explored the potential racial disparities in the use of strip searches for suspects in custody in the UK and found that the strip searches are used more frequently and invasively on Black and Asian detainees compared to White detainees. They suggested that this is due to unconscious biases and stereotypes held by police officers, and the lack of clear guidelines and training on when and how to conduct strip searches. There is a need for better oversight and regulation of strip searches, as well as greater awareness and training on issues of race and unconscious bias within the police force. Simons (1979) specifically focused on the female groups that argued that strip searches are used as a method to abuse women in jail for interrogative purposes.

However, with the decades of evolution of laws and policies and the popularization of human privacy rights, strip searches have been used more prudently in recent years. We are expecting to see any positive changes based on the analysis of the arrest and strip search dataset from the Toronto Police.

Exploratory Data Analysis (EDA)

Data Cleaning for RQ1:

We checked the null values and did the value count for each race group and gender group before we started the data cleaning. We only keep the columns which are related to the research question and remove the null values in them. There are 5056 records for unknown or legacy race groups and nine records for gender group “U”, which refers to unknown, so we removed the rows of them and thus could be more focused on the explainable data. We merged the East/Southeast Asian group and the South Asian group as a new Asian group since there are no obvious differences in macroscopic features and racial information between people from these two areas.

Then we filter the dataset by screening the strip search results and grouping the records with the same arrest year, month, race and sex. Table 1 shows the head of the dataset with a strip search, and Table 2 shows arrested cases without a strip search.

Table 1: Arrest records with a strip search.

	Arrest_Year	Arrest_Month	Perceived_Race	Sex	StripSearch
0	2020	Apr-June	Asian	F	12
1	2020	Apr-June	Asian	M	132
2	2020	Apr-June	Black	F	46
3	2020	Apr-June	Black	M	480
4	2020	Apr-June	Indigenous	F	13

Table 2: Arrest records without strip search.

	Arrest_Year	Arrest_Month	Perceived_Race	Sex	StripSearch
0	2020	Apr-June	Asian	F	92
1	2020	Apr-June	Asian	M	570
2	2020	Apr-June	Black	F	261
3	2020	Apr-June	Black	M	1054
4	2020	Apr-June	Indigenous	F	44

Since the target feature “proportion” is not provided in the original dataset, we calculated the ratio of the two record numbers and set the result as a new column.

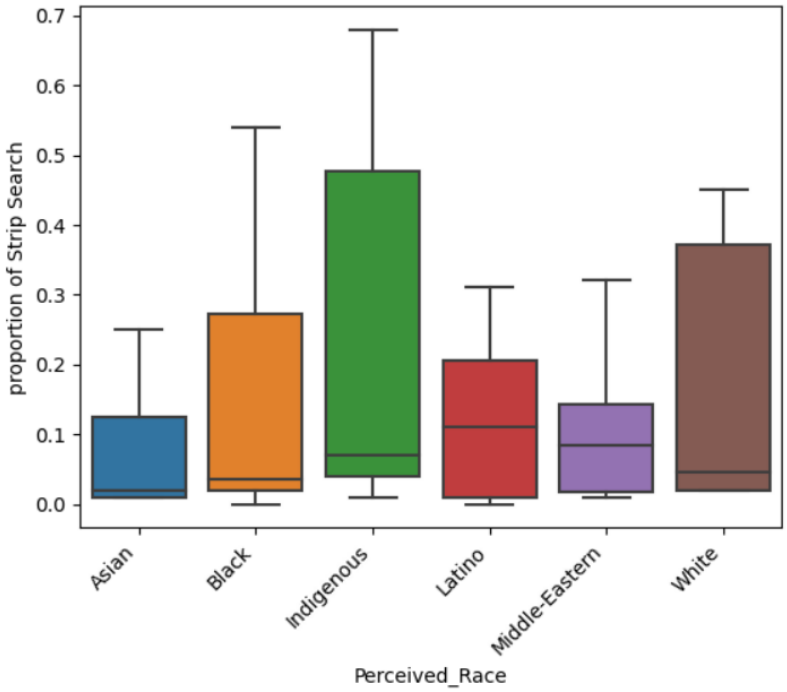
Table 3: Arrest records with strip search proportion.

	Arrest_Year	Arrest_Month	Perceived_Race	Sex	StripSearch_x	StripSearch_y	portion
0	2020	Apr-June	Asian	F	12	92	0.13
1	2020	Apr-June	Asian	M	132	570	0.23
2	2020	Apr-June	Black	F	46	261	0.18
3	2020	Apr-June	Black	M	480	1054	0.46
4	2020	Apr-June	Indigenous	F	13	44	0.30
5	2020	Apr-June	Indigenous	M	59	87	0.68

Data Visualization for RQ1

We created a box plot for both gender groups and perceived race groups to see the big picture of the data distribution. For the perceived race groups, the median values of each group proportion are all around 0.1, which does not differ a lot. The indigenous have the largest interquartile range (IQR) from 0.02 to 0.48, which is almost three times of the IQR of Asian and Middle East groups. There are no outliers shown in the box plot, which refers to balanced and centralized data distribution.

Figure 1: Box plot for perceived race group and strip search proportion.



On the other hand, the box plot of gender groups contains a few outliers which means the far distance of strip search proportions for the gender groups. The median value for both males and females are around 0.1, which means most of the arrest records are concentrated in the zone (0, 0.1). The IQR of male groups is a bit larger than the female groups that the male spread of the middle half of the data distribution is more incompact.

Figure 2: Box plot for perceived gender group and strip search proportion.

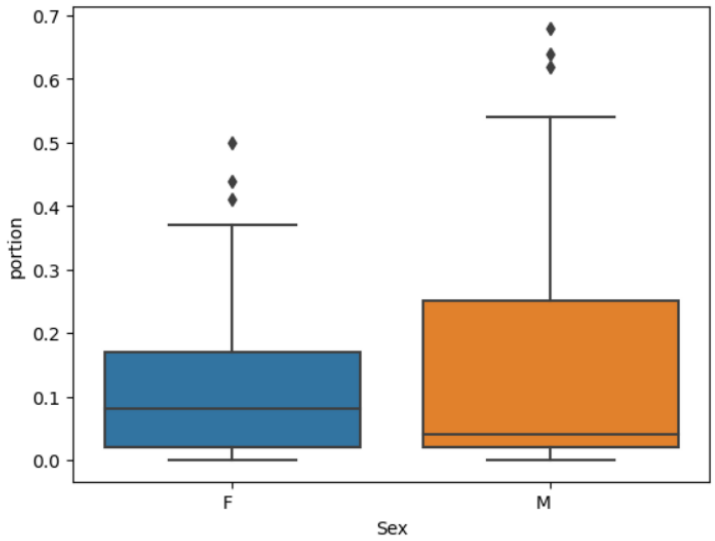


Figure 3 shows the distribution of strip search proportion. It is obvious to see that it's a right-skewed distribution which indicates the mode of the proportion is less than the mean value. Almost half of the data are located in the zone (0, 0.1). To conclude, the statistical summary of perceived race groups and gender groups is shown in Table 4 and Table 5.

Figure 3: Distribution of strip search proportion.

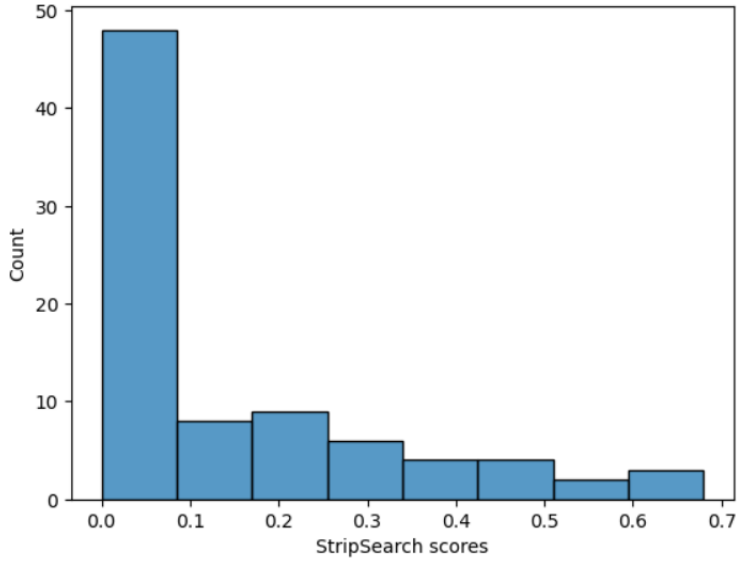


Table 4: Statistic summary of perceived race groups and strip search proportion.

Perceived_Race	count	mean	std	min	25%	50%	75%	max
Asian	15.0	0.080000	0.093427	0.01	0.0100	0.020	0.1250	0.25
Black	16.0	0.156250	0.194487	0.00	0.0200	0.035	0.2725	0.54
Indigenous	14.0	0.247857	0.266175	0.01	0.0400	0.070	0.4775	0.68
Latino	11.0	0.116364	0.112718	0.00	0.0100	0.110	0.2050	0.31
Middle-Eastern	12.0	0.100000	0.097887	0.01	0.0175	0.085	0.1425	0.32
White	16.0	0.170000	0.185257	0.02	0.0200	0.045	0.3725	0.45

Table 5: Statistic summary of perceived gender groups and strip search proportion.

Sex	count	mean	std	min	25%	50%	75%	max
F	37.0	0.127568	0.139272	0.0	0.02	0.08	0.17	0.50
M	47.0	0.162766	0.202717	0.0	0.02	0.04	0.25	0.68

T-test for RQ1:

Based on the statistical results in the previous steps, we found that the Indigenous group stands out in the box plot. Since the size of each group differs a lot, we choose the Welch t-test to check whether the means of the two populations are equal. Thus, we choose indigenous as the primary group and perform the five Welch t-tests for indigenous groups and each other group, as well as a Welch t-test for gender groups.

Indigenous and Latino

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Latino, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Latino, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.25, SD = 0.27) are higher than Latino (m = 0.12, SD = 0.11). The p-value of these two groups is 0.11, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that the strip search proportion does not differ from Indigenous and Latino groups. The t-statistic value of this pair group is 1.7, which means there is not much evidence to support that there is a significant difference between these two groups.

Indigenous and Black

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Black, are equal.

- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Black, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.25, SD = 0.27) are higher than Black (m = 0.16, SD = 0.19). The p-value of these two groups is 0.30, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that strip search proportion does not differ from Indigenous and Black groups. The t-statistic value of this pair group is 1.06, which means there is not much evidence to support there is a significant difference between these two groups.

Indigenous and White

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and White, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and White, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.25, SD = 0.27) are higher than White (m = 0.17, SD = 0.19). The p-value of these two groups is 0.37, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that the strip search proportion does not differ from Indigenous and White groups. The t-statistic value of this pair group is 0.92, which means there is not much evidence to support there is a significant difference between these two groups.

Indigenous and Asian

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Asian, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Asian, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.25, SD = 0.27) are higher than Asians (m = 0.08, SD = 0.09). The p-value of these two groups is 0.04, which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the proportion means of strip search proportion of the Indigenous group and Asian group are not equal. The t-statistic value of this pair group is 2.24, which means there is great evidence to support there are significant differences between these two groups.

Indigenous and Middle East

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Middle East, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Middle East, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.25, SD = 0.27) are higher than the Middle East (m = 0.10, SD = 0.10). The p-value of these two groups is 0.70, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that the strip search proportion does not differ from Indigenous and Middle East groups. The t-statistic value of this pair group is 1.93, which means there is not much evidence to support there is a significant difference between these two groups.

Female and Male

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Female and Male, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Female and Male, are not equal.

The mean value and the standard deviation of the Female group (m = 0.13, SD = 0.14) are lower than those the Male (m = 0.16, SD = 0.20). The p-value of these two groups is 0.35, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that strip search proportion does not differ from gender groups. The t-statistic value of this pair group is 0.94, which means there is not much evidence to support there is a significant difference between these two groups.

To conclude, we only find that the strip search proportion in the Indigenous group has a significant difference from the Asian group. Since the result did not fit the expectation, we did more Welch t-tests for other race groups. However, the Asian and Indigenous group is the only pair we found that has a significant difference in the strip search portion. We would then proceed with a Two-Way ANOVA test for further investigation about the mean of strip search proportion among both gender and perceived race groups.

Data Cleaning for RQ2:

First, given the dataset source explanation, we change the value of booked from 0 to 1 for all rows where the value for strip search is 1, meaning that a strip search took place. Next, we recognized for the Age column, there are two columns representing people under age 17, and also there are two columns representing people aged 65 and over. We removed the redundant column by combining the two columns that represent the same group.

Table 6

	Arrest_Year	Arrest_Month	Perceived_Race	Sex	Age_group__at_arrest_	Actions_at_arrest___Cooperative	Arrests	portion
0	2020	Apr-June	Black	F	Aged 17 years and younger	16	36	0.44
1	2020	Apr-June	Black	F	Aged 18 to 24 years	30	87	0.34
2	2020	Apr-June	Black	F	Aged 25 to 34 years	23	99	0.23
3	2020	Apr-June	Black	F	Aged 35 to 44 years	22	52	0.42
4	2020	Apr-June	Black	F	Aged 45 to 54 years	6	26	0.23

To account for the unequal representation of each race in the dataset, here we first combined the data rows by five columns, “Arrest_Year”, “Arrest_Month”, “Perceived_Race”, “Sex”, and “Age_group_at_arrest_”. We summed the values (1 and 0) in the Action at arrest cooperative column as the aggregated number of cooperative instances. Next, we calculate the cooperative scores for each race group by dividing the number of cooperative instances by the total number of arrests for each group and name the new column 'portion'.

Data Visualization for RQ2

Figure 4: Visualise the overall pattern between Race groups and cooperativeness.

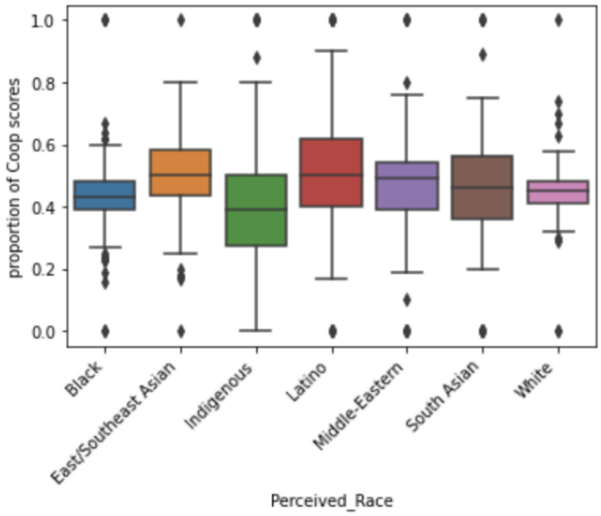
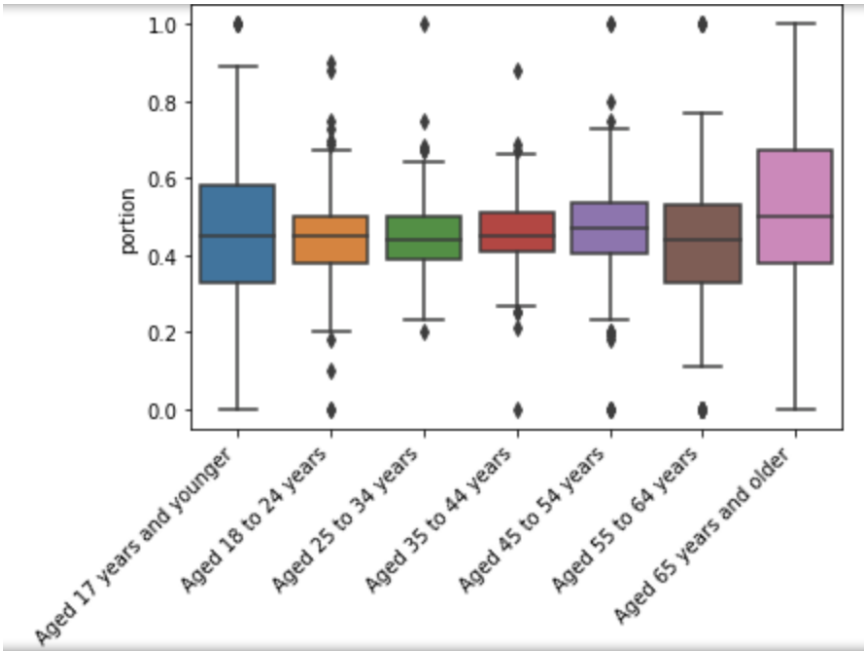


Figure 5: Visualise the overall pattern between age group and cooperativeness.



We created a box plot to visualize the overall pattern of the cooperative score between different race groups and age groups. First, we can see relatively consistent and uniform behaviours in the black and white groups shown by a comparatively short plot. The median mostly falls between 0.4 and 0.6, but the distribution varies. We can visually spot a difference between groups, especially between the Indigenous and Latino groups, where the upper quartile of the Latino group reaches over 0.6, and the lower quartile of the indigenous group goes as low as 0.3. From the plot, it seemed possible that there is a difference in cooperativeness between race groups. We also created a box plot of cooperative scores between different age groups, and the plot did not reveal any obvious trends or differences between groups.

Table 7: Summary statistics of cooperative scores of different race groups

Perceived_Race	count	mean	std	min	25%	50%	75%
Black	110	0.433818	0.141376	0.0	0.39	0.43	0.48
East/Southeast Asian	112	0.504196	0.142662	0.0	0.4375	0.5	0.5825
Indigenous	98	0.411939	0.248163	0.0	0.275	0.39	0.5
Latino	101	0.505545	0.242052	0.0	0.4	0.5	0.62
Middle-Eastern	107	0.464766	0.195855	0.0	0.39	0.49	0.54
South Asian	106	0.457453	0.218122	0.0	0.36	0.46	0.56
White	116	0.448621	0.105393	0.0	0.41	0.45	0.48

Again, from this table, we can see the mean cooperativeness score varies across different perceived race groups. The highest mean score is for Latinos and the lowest for Indigenous. There

is also a wide variation in the standard deviation of cooperativeness scores within each perceived race group.

T-test for RQ2:

Driven by the summary statistics of cooperative scores of different race groups, we selected the Indigenous Latino as the primary target for testing since they stood out in the boxplot, we also incorporated Black and White groups into the tests for comparisons. More specifically, we performed Welch's t-test on the race group pairs of Indigenous versus Latino, Black and white, and Latino versus Black and white, five tests in total.

Latino and Indigenous

We computed the mean cooperative scores for the Latino group and the Indigenous group and found that the score for Latinos is much higher than for Indigenous (0.51 and 0.41, respectively). We conducted a Welch's T-test to analyze whether the cooperative scores (outcome variable) differ between these two race groups. The hypothesis being tested are the following:

- H0(Null Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Indigenous, are equal.
- HA (Alternative Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Indigenous, are not equal.

The results indicate that the mean cooperative score of the Latino group (M=0.51, SD=0.24) is higher than the mean cooperative score of the Indigenous group(M=0,41, SD-0,25). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.008) is less than 0.05. Therefore, we reject the null hypothesis and conclude that the population means of cooperative scores of the two independent groups, Latino and Indigenous, are not equal.

Latino and White

We computed the mean cooperative scores for the Latino group and White group and found that the score for Latinos is also higher than for White (0.51 and 0.45, respectively). We conducted a Welch's T-test to analyze whether the cooperative scores (outcome variable) differ between these two race groups. The hypothesis being tested are the following:

- H0(Null Hypothesis): The population means of cooperative scores of the two independent groups, Latino and White, are equal.

- HA(Alternative Hypothesis): The population means of cooperative scores of the two independent groups, Latino and White, are not equal.

The results indicate that the mean cooperative score of the Latino group(M=0.51, SD=0.24) is higher than the mean cooperative score of the white group(M=0.45, SD=0.11). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.03) is less than 0.05. Therefore, we reject the null hypothesis and conclude that the population means of cooperative scores of the two independent groups, Latino and White, are not equal.

Latino and Black

We computed the mean cooperative scores for the Latino group and the Black group and found that the score for Latinos is also higher than White (0.51 and 0.43, respectively). We conducted a Welch's T-test to analyze whether the cooperative scores (outcome variable) differ between these two race groups. The hypothesis being tested are the following:

- H0(Null Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Black, are equal.
- HA (Alternative Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Black, are not equal.

The results indicate that the mean cooperative score of the Latino group (M=0.51, SD=0.14) is higher than the mean cooperative score of the White group (M=0.45, SD-0.25). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.01) is less than 0.05. Therefore, we reject the null hypothesis and conclude that the population means of cooperative scores of the two independent groups, Latino and Black, are not equal.

The same Welch's T-test was performed for White vs. Indigenous and Black vs. Indigenous. However, we did not find a significant difference in cooperative scores between the indigenous group and either the Black or White groups, as indicated by the p-values of 0,443 and 0.175.

The t-tests we performed overall suggest that the proportion of cooperative individuals in the Latino group is significantly different from that in the Black, White and Indigenous group. From here, to further investigate the difference in mean cooperative scores amongst different race groups, we would proceed with a one-way ANOVA test.

Methods

Dataset Description

For our project, we used the Police service dataset that's made public and available on the Public Safety Data Portal of the Toronto Police service. This dataset shows Policing information related to arrests and strip searches. The dataset also includes criminals' information such as demographic information, the occurrence of crime, and their actions at arrest. The data has 65276 records of arrests within the timeframe from 2020 to 2022. Some features of interest include gender, race groups, strip search conducted, and their actions at arrest. Note that the dataset does not contain any continuous data but only binary data (0 and 1). For example, for the strip search column, the value of 1 indicates a strip search was conducted, and if a strip search did not take place, the value equals 0. Same for columns such as "Action_at_Arrest_Cooperative", a value of 1 indicates the person was cooperative, and 0 if the person was not.

For the first research question, we focused on the strip search column and found that the strip search does not frequently happen that, only around 13.6% of the arrested cases applied strip searches, which is much lower than that decades ago. As for the second research question, we discovered that there are more records of people Aged between 18 to 44, with 25 to 34 being the largest age group. In addition, more than half of the total records are of the White and Black race group, with 27723 being White. We also noted that there are substantially more records of the arrest of males than females at approximately 5 to 1 ratio.

Given that the dataset lacks continuous data, we considered performing some data wrangling, including combining columns and aggregating the number of the potential explanatory variables, such as strip search.

Two-Way ANOVA for RQ1

As per the result of the Welch t-test, we selected the Asian group and the Indigenous group as one impact factor and gender groups as the other one to proceed with a Two-way ANOVA test to see whether there are significant differences among the groups.

Two-way ANOVA as the extension of One-way ANOVA, would examine the influence of two different categorical independent variables on one continuous dependent variable. However, since the dataset has no continuous data but only has binary data instead, we would hold reservations about the test results.

- H0(Null hypothesis): $\mu1 = \mu2 = \mu3 = \mu4 = \mu5 = \mu6$. The strip search proportion means that all the six perceived race groups are equal.
- H1(Alternative hypothesis): At least one mean is different from the others.

Post-hoc test (Tukey's HSD) for RQ1

The Welch t-test shows the significant difference between strip search proportion and perceived race group in a particular pair. We would perform the Tukey test as the support explanation of the Two-Way ANOVA test.

One-Way ANOVA for RQ2

After performing Welch's T-test, and we discovered a significant difference in mean cooperative scores between selected pairs, we proceeded with a one-way ANOVA to test for significant differences among multiple race groups.

The one-way ANOVA test compares the variance between groups to the variance within groups. In this case, the null hypothesis is that there is no significant difference in cooperativeness scores between the different race groups, and the alternative hypothesis is that there is at least one race group with a different cooperativeness score than the others.

Post-hoc test (Tukey's HSD) for RQ2

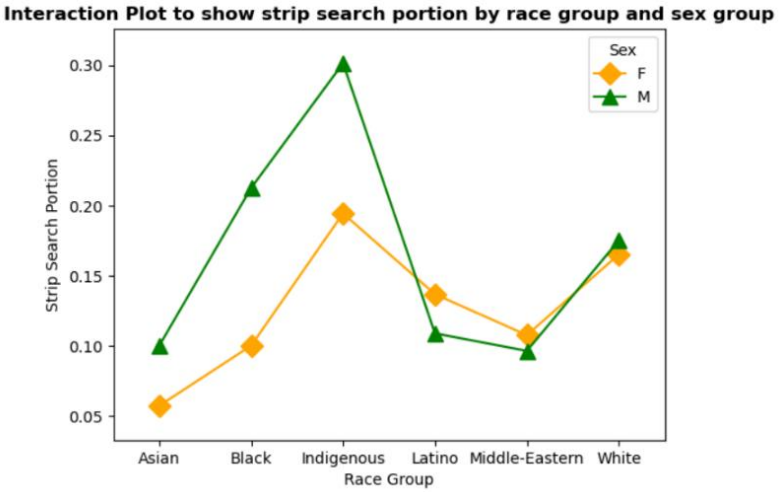
The one-way ANOVA test has informed us there is a significant difference in cooperativeness scores between the different race groups, and therefore race is a significant factor in determining cooperativeness scores. Therefore, we proceeded to follow up with a Tukey's HSD post-hoc test to test which specific groups are significantly different from one another.

Findings

Interaction Plot

For research question 1, as shown in Figure 6, the lines intersect significantly which refers to there exists an interaction effect between perceived race, gender, and strip search proportion. Since the two lines have similar trends with many overlaps, the Middle East group and the White group are virtually identical. On the flip side, the strip search proportion of male suspects in Indigenous groups is much higher than that of the female suspects.

Figure 6: Interaction Plot to show strip search portion by race group and sex group.



Two-Way ANOVA for RQ1

Table 8: Two-way ANOVA

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	0.274205	5.0	1.748874	0.134475
C(Sex)	0.043449	1.0	1.385571	0.243031
C(Perceived_Race):C(Sex)	0.056650	5.0	0.361313	0.873307
Residual	2.257767	72.0	NaN	NaN

We set the alpha level to 0.05. The ANOVA table shows that the sum of squares for the perceived race variable is 0.27. The F-statistic is 1.74, which refers to the ratio of the mean sum of squares to the mean square error. There are 5 degrees-of-freedom values in the race group. The p-value of the race group is 0.13, which is larger than 0.05. Therefore, we can NOT reject the null hypothesis. That is, we conclude there is no significant difference in the strip search proportion and perceived race group.

The sum of squares for the sex variable is 0.04. The F-statistic is 1.39 which refers to the ratio of the mean sum of squares to the mean square error. There is only one degree-of-freedom

value in the gender group. The p-value of race group is 0.24, which is larger than 0.05. Therefore, we can NOT reject the null hypothesis. That is, we conclude there is no significant difference in the strip search proportion and gender group.

In a word, the Two-way ANOVA test shows there is no significant difference in the strip search proportion and neither perceived race group nor gender group. Since we still need to do the Post-hoc test as per the assignment requirements, we would choose the perceived race group since the difference is a bit bigger.

Post-hoc test (Tukey's HSD) for RQ1

It is obvious to see all the values in the rejected column are false. The Asian and Indigenous group is the closest pair but still can not tell a significant difference. We would conclude that the race and gender group would NOT impact the strip search proportion in arrested cases.

Figure 6: Tukey's HSD for RQ1

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Asian	Black	0.0762	0.8257	-0.1063	0.2588	False
Asian	Indigenous	0.1679	0.11	-0.0209	0.3566	False
Asian	Latino	0.0364	0.9949	-0.1653	0.238	False
Asian	Middle-Eastern	0.02	0.9997	-0.1767	0.2167	False
Asian	White	0.09	0.7025	-0.0926	0.2726	False
Black	Indigenous	0.0916	0.7028	-0.0943	0.2775	False
Black	Latino	-0.0399	0.9917	-0.2389	0.1591	False
Black	Middle-Eastern	-0.0562	0.9575	-0.2502	0.1377	False
Black	White	0.0137	0.9999	-0.1659	0.1934	False
Indigenous	Latino	-0.1315	0.4238	-0.3362	0.0732	False
Indigenous	Middle-Eastern	-0.1479	0.2676	-0.3477	0.052	False
Indigenous	White	-0.0779	0.824	-0.2638	0.108	False
Latino	Middle-Eastern	-0.0164	0.9999	-0.2284	0.1957	False
Latino	White	0.0536	0.9689	-0.1453	0.2526	False
Middle-Eastern	White	0.07	0.8977	-0.124	0.264	False

One-Way ANOVA for RQ2

Table 9: One-way ANOVA Table.

Source	Sum of Squares	Degrees of Freedom	F-statistic	p-value
Perceived_Race	0.747156	6	3.477071	0.002137
Residual	26.6094	743		

The ANOVA table shows that the sum of squares (SS) for the perceived race variable is 0.747 with 6 degrees of freedom (df), which results in an F-statistic of 3.477 and a p-value of 0.002137. This means that there is a significant difference in cooperativeness scores between the different race groups. The residual SS is 26.609 with 743 df, which means that most of the variation in cooperativeness scores is due to differences within the groups rather than between them. Overall, the result suggests that race is a significant factor in determining cooperativeness scores, but there may be other factors that are also important.

Post-hoc test (Tukey's HSD) for RQ2

The one-way ANOVA test has informed us there is a significant difference in cooperativeness scores between the different race groups, and therefore race is a significant factor in determining cooperativeness scores. Therefore, we proceeded to follow up with a Tukey's HSD post-hoc test to test which specific groups are significantly different from one another.

Table 10: Tukey's HSD for RQ2

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	0.0704	0.0834	-0.0047	0.1455	False
Black	Indigenous	-0.0219	0.9	-0.0996	0.0558	False
Black	Latino	0.0717	0.0875	-0.0054	0.1488	False
Black	Middle-Eastern	0.0309	0.8908	-0.045	0.1069	False
Black	South Asian	0.0236	0.9	-0.0525	0.0998	False
Black	White	0.0148	0.9	-0.0597	0.0893	False
East/Southeast Asian	Indigenous	-0.0923	0.0082	-0.1696	-0.0149	True
East/Southeast Asian	Latino	0.0013	0.9	-0.0754	0.0781	False
East/Southeast Asian	Middle-Eastern	-0.0394	0.6932	-0.1151	0.0362	False
East/Southeast Asian	South Asian	-0.0467	0.5281	-0.1226	0.0291	False
East/Southeast Asian	White	-0.0556	0.2877	-0.1297	0.0185	False
Indigenous	Latino	0.0936	0.0093	0.0143	0.1729	True
Indigenous	Middle-Eastern	0.0528	0.42	-0.0254	0.1311	False
Indigenous	South Asian	0.0455	0.5906	-0.0329	0.1239	False
Indigenous	White	0.0367	0.7686	-0.0401	0.1134	False
Latino	Middle-Eastern	-0.0408	0.6862	-0.1184	0.0368	False
Latino	South Asian	-0.0481	0.5253	-0.1259	0.0297	False
Latino	White	-0.0569	0.2912	-0.1331	0.0192	False
Middle-Eastern	South Asian	-0.0073	0.9	-0.084	0.0694	False
Middle-Eastern	White	-0.0161	0.9	-0.0911	0.0588	False
South Asian	White	-0.0088	0.9	-0.084	0.0663	False

The table shows the mean differences (meandiff), the p-values adjusted for multiple comparisons (p-adj), the lower and upper bounds of the confidence intervals (lower and upper), and whether or not to reject the null hypothesis of no difference between the means (reject).

For each pair of groups, if the p-adj value is less than the significance level (e.g. 0.05), then we reject the null hypothesis that the means of the two groups are equal. If reject is True, then we conclude that there is a significant difference between the means of the two groups.

From this particular result, we can see that the only significant difference in means is between East/Southeast Asian and Indigenous, and between Indigenous and Latino, as their p-adj values are less than 0.05. All other comparisons do not have significant differences between their means.

Discussion

For research question 1, we find that the strip search proportion does not tell a significant difference between either perceived race group or gender group. Although in the Welch test, we found a particular pair of Asian and Indigenous groups significantly differs from the strip search

proportion. However, both the Two-way ANOVA test and the Tukey test reject the null hypothesis. We repeated thousands of times to see whether we could do the data cleaning step to improve the result, nevertheless, after all the attempts we would say that the strip search proportion would not be impacted by the perceived race and sex.

The Two-way ANOVA test shows a comparison result about the variance both between and within groups, but the final findings show that the impact of the perceived race group may be stronger than the gender group. The Tukey test compared all pairs of race groups and adjusted the significant level. But all the values in the rejected column of the Tukey test are false, which refers to the result after the reduction of type error in the Welch t-test. The p-value is larger than 0.05 for all the pairs. Although the Asian and Indigenous group is the closest pair, it still can not tell a significant difference.

In our study of research question 2, our research question was, “Is there a statistically significant difference in cooperativeness between different race groups? And which ace group exhibits the highest level of cooperativeness?”. Before we performed our tests, we visualized the overall pattern between Race groups and their cooperatives by creating a boxplot. The boxplot for the Indigenous group and Latino group showed a big difference and fueled our further investigation. We selected the indigenous and Latino group to perform Welch’s t-test, as well as the black and white groups, which both had a narrow boxplot. We aimed to compare the means between Indigenous and the other three selected groups, and Latinos with the other groups. From our t-tests, the interesting pattern that we discovered is that the proportion of cooperative individuals in the Latino group is significantly different from all the groups that it was compared with. This further informed our research question and allowed us to move on to the next step which is one-way ANOVA.

With one-way ANOVA we were able to compare the variance between groups to the variance within groups. The ANOVA table that we generated showed a significant difference in cooperativeness scores between the different race groups. Overall, the result suggests that race is a significant factor in determining cooperativeness scores. Note that from the table, we see a residual SS of 26.6 with 743 df suggesting that most of the variation in cooperativeness scores is due to differences within the groups rather than between them. This could indicate other factors such as individual differences within each race group.

Precedingly, we conducted a post-hoc test, Tukey's HSD, to test all possible pairwise comparisons simultaneously. The result here, however, showed that significant differences in means are only shown between East/Southeast Asian and Indigenous and between Indigenous and Latino, while all other comparisons do not have significant differences between their means. The difference between Indigenous and Latino aligns with our findings earlier in our t-test, but the difference shown between Latino with black and white was not shown here in Tukey's test. Tukey HSD test compares all possible pairs of group means and adjusts the significance level for multiple comparisons, it reduces Type I error and may be more conservative in detecting the significant differences. This might be the cause of detecting very few statistical differences in the mean between most of the pairs.

Limitations

For our study, we performed ANOVA tests and post-hoc tests to test the difference in means between different groups. However, our original dataset consists of only 1's and 0's which is not the most appropriate case for adopting ANOVA method. Also, we investigated the instance of strip search. There were 7801 instances of a strip search being conducted out of a total of 65276 instances in the dataset, suggesting that strip searches are relatively infrequent, and this low frequency of strip search may make it more difficult to detect statistically significant differences in the number of strip searches between groups, especially if the differences are relatively small.
We also recognize the potential for sampling bias, meaning our sample may not be representative of the population and thus making the result not generalizable. Our dataset did not provide location data which would be crucial and valuable data for our analysis. The missing location data prevent us from investigating the geographic area where the sample is drawn from.

Potential confounding variables would also be of limitations. As indicated from our ANOVA result, although there is a statistically significant difference in mean between groups, the large RSS suggests that the variation comes more from within the group. There may be other variables that could impact the cooperative scores that we did not measure or control for in our analysis, such as socioeconomic status or educational level. Lastly is causality, it is important to remember that our study may show statistically significant differences between groups but did not necessarily indicate any causal relationship.

Conclusion

The answer to whether the proportion of strip searches differs from race and sex groups is "NO" on the base of our data analysis of the dataset "Arrests_and_Strip_Searches" from Toronto police. We believe it could be regarded as a good change on comparison of the previous study results since the appeals about better oversight and regulation of strip searches, greater awareness and training on issues of race and unconscious bias within the police force has been accepted and impacted the strip search to become a fair method in resolve with reasonable suspicions. Or we may say that the strip search is used as a more fair and objective method in modern interrogations in Toronto.

As for the second research, we found that there is a statistically significant difference in cooperativeness between different race groups. The East/Southeast Asian and Indigenous, and the Indigenous and Latino, shows significant differences to arrest cooperativeness. That may indicate suspects from those race groups may be more cooperative when they were arrested, which would also be considered a reason for the reduction of strip search usage.

Reference

Helmer, G. M. (2001). Strip Search and the Felony Detainee: A Case for Reasonable Suspicion. BUL Rev., 81, 239.

Newburn, T., Shiner, M., & Hayman, S. (2004). Race, Crime and Injustice? Strip Search and the Treatment of Suspects in Custody 1. British Journal of Criminology, 44(5), 677-694.

Simons, P. E. (1979). Strip-Search: The Abuse of Women in Police Stations. Barrister, 6, 8.