Arrests and Strip Searches
Data Analysis Report

INF 2178 Midterm Project

Group 46

Zidong Zhou (1008517758)

Gary Bhanot (999669327)

Table of Contents

1. Literature Review

Arrests and Strip Searches (RBDC-ARR-TBL-001) dataset[1] provided by Toronto Police Service's Public Safety Data Portal contains all related information to all arrests and strip searches performed by the Toronto Police. A strip search is a kind of search performed by an officer which involves the removal of some or all pieces of clothing for a police officer to inspect buttocks, genitals, breasts or undergarments visually.[2] It is often done before booking a person with the pretext of looking for concealed contraband and weapons.

As per the Star article title, 'Clearly, we were doing it wrong' [3]. Toronto Police's Interim Chief admitted that Toronto Police were doing about 22,000 strip searches year after year, and they were doing it wrong. This article came out after OIPRD published Summary of Ontario Jurisprudence Involving Strip Searches Post R. v. Golden, a landmark 2001 Supreme Court of Canada decision on the legality of Strip Searches. OIPRD's study focussed on 89 cases between 2002 and December 31, 2018, where the strip searches were considered a violation of an individual's charter rights.

2. Introduction

Strip searches have been dubbed "Inherently degrading" by Canada's top court [2]. This report aims to examine and identify any patterns in strip searches, arrests, and any associated

aggression by an individual. Since Toronto is a multicultural city, this report will analyze trends between different races and gender along with the arrests and strip search data.

3. Exploratory Data Analysis

Since the data is timestamped, we could explore, visualize data and see trends over the years. Still, for this study, we are more interested in finding insights if a person's perceived race or gender plays any role in them being strip searched and booked as well as seeing if there is a pattern between aggression levels across different age groups and gender while being strip searched and booked.

3.1 Aggression levels across different age groups and gender

If you look at Figure 3.1.1, we can clearly see that females tend to be more aggressive with an exception of women aged between 25-34 years. Elderly and younger people tend to be less aggressive than the other age groups as you can see that the aggression level for age groups 65 and older and 17 and younger age group is less than 0.35. Women aged 18-24 tend to be the most aggressive, and overall, people aged 25-44 are relatively more aggressive than others.

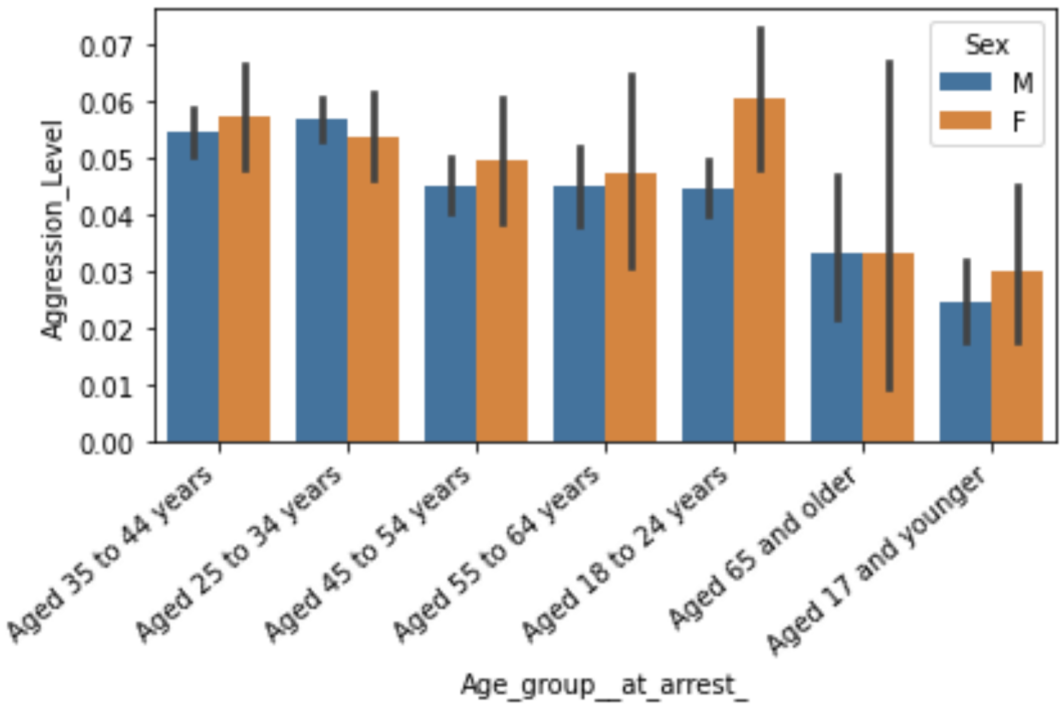

Figure 3.1.1 Barplot showing aggression levels across different age groups and gender.

3.2 Strip Search levels across different age groups and gender

Figure 3.2.1 depicts the total number of strip searches performed for men and women and unknown gender across different age groups. Just looking at this barplot, It seems that men are more likely to be strip searched when being booked than women. Elderly are barely strip searched when booked and people aged between 25-34 are more likely to be strip searched.

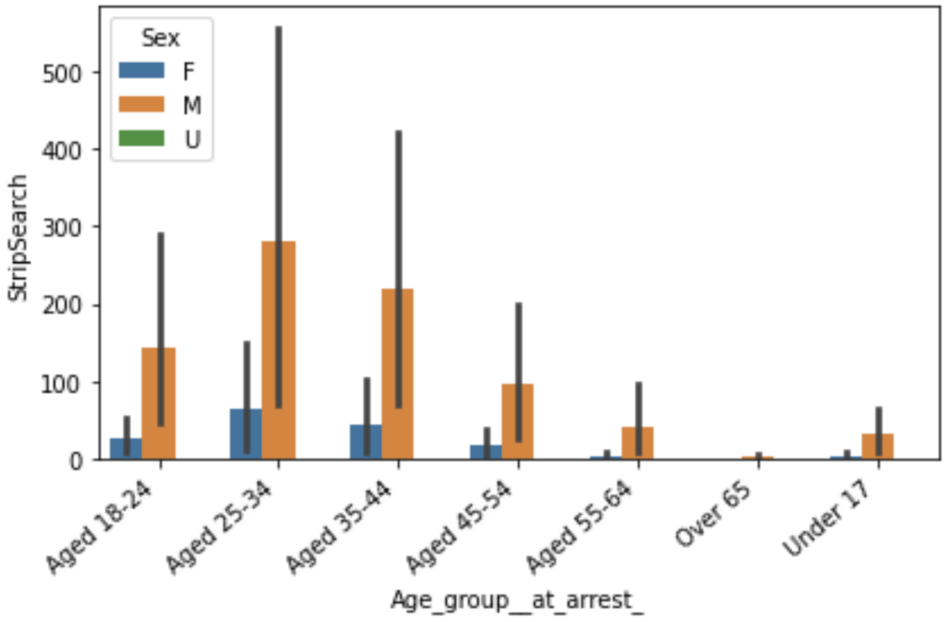

Figure 3.2.1 Barplot showing Strip Search levels across different age groups and gender.

3.3 Strip Search and Booked levels across different Races and gender

Figure 3.3.1 and Figure 3.3.2 are very identical in the sense that most of the people being strip searched are already on their way to get booked but you can see that the Black and White people are more likely to be strip searched and booked and all other races fall within the first quartile, so there is a big difference between atleast 2 groups versus other races but we will see how statistically significant this difference is when we design our research and formalize our research questions.

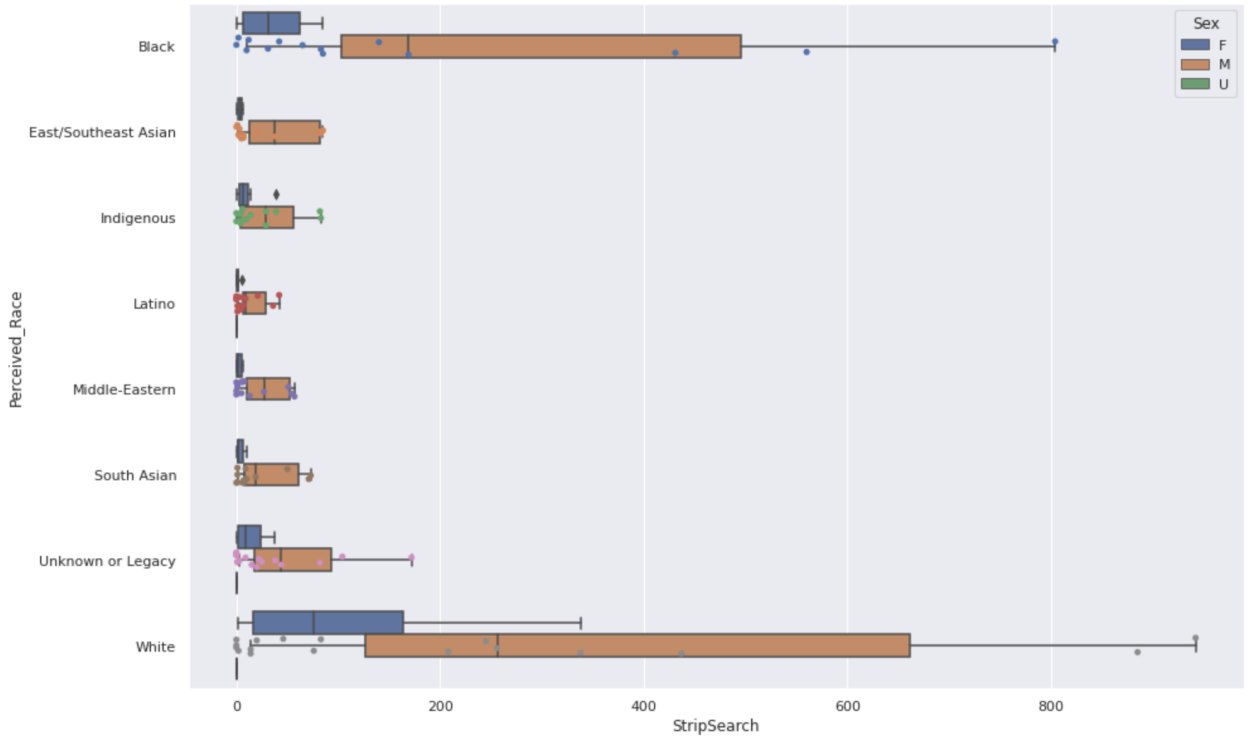

Figure 3.3.1 Boxplot showing Strip Search levels across different races and gender.

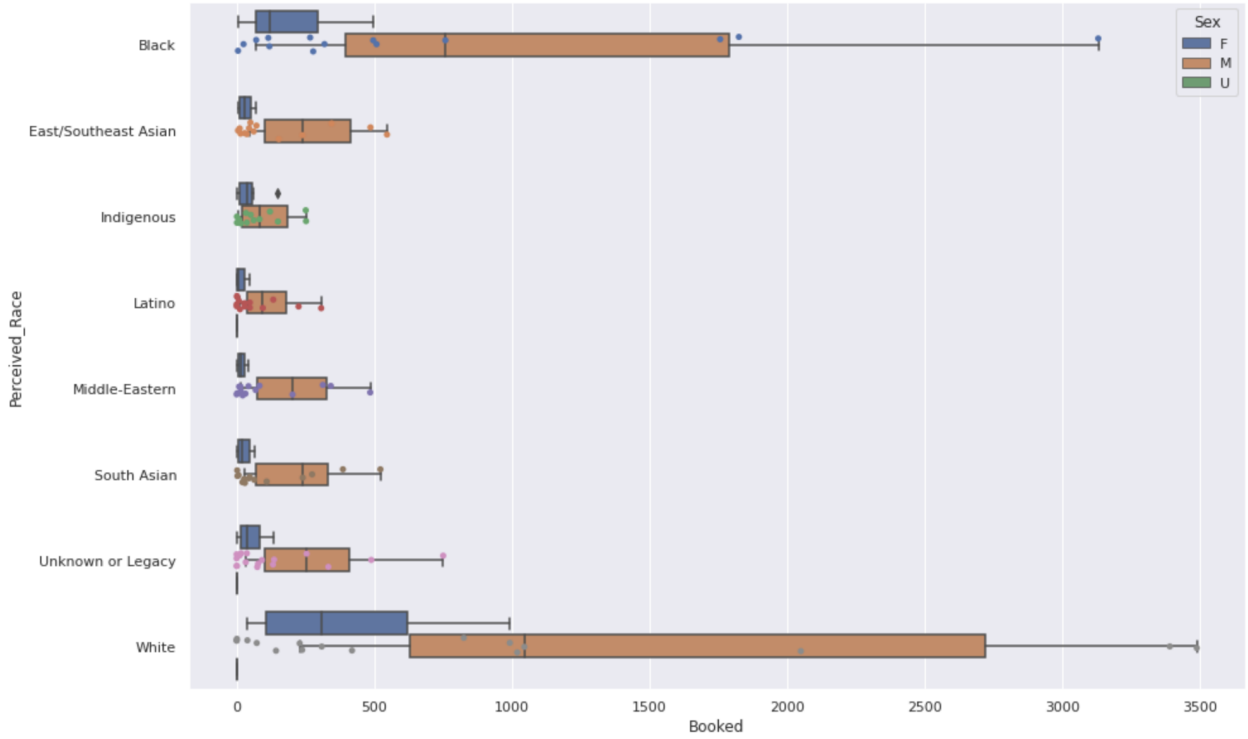

Figure 3.3.2 Boxplot showing Booked levels across different races and gender.

Figure 3.3.3 shows another angle of how much difference there is between Black, White and other races and we can see that the Female tend to be less likely to be strip searched compared to males.

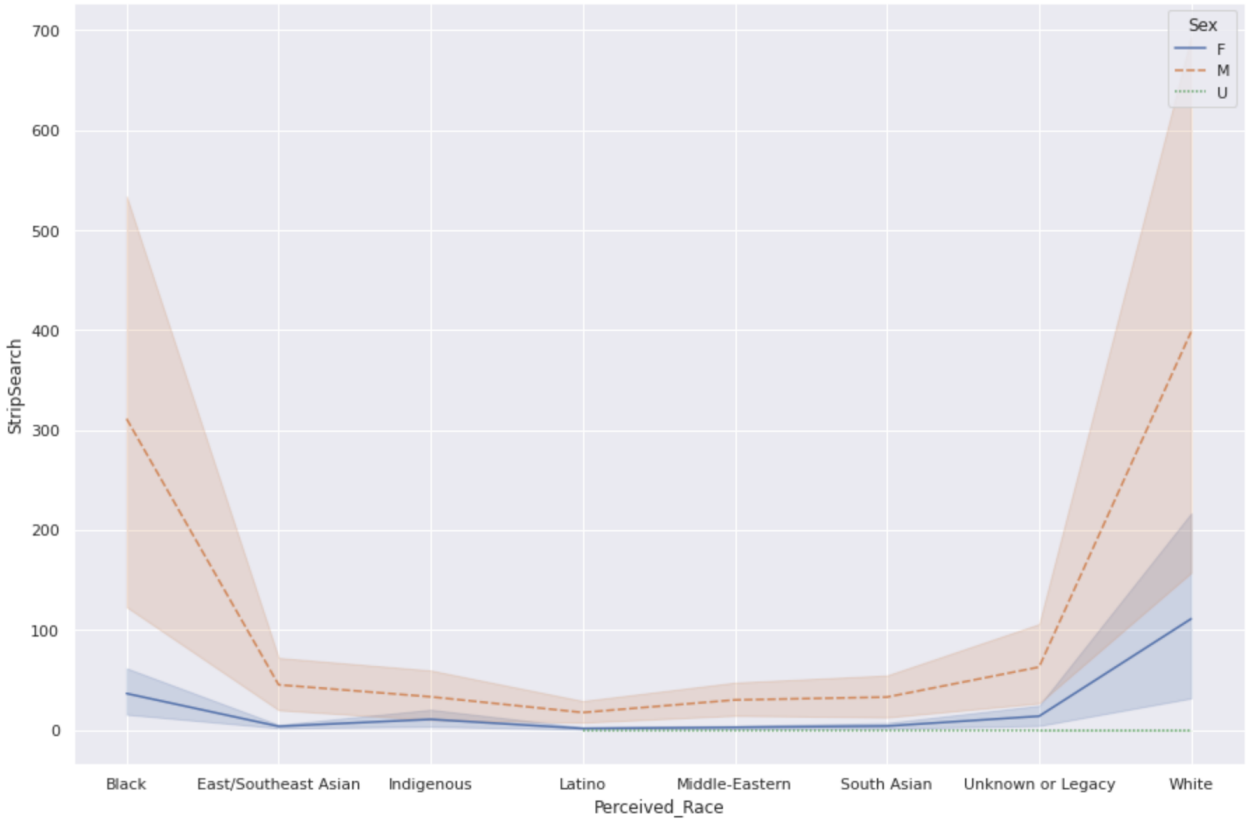

Figure 3.3.3 Lineplot showing Strip Search levels across different races and gender.

3.4 Relationship between Strip searches and getting arrested

As we know that an individual is strip searched in the pretext of getting booked to see if they are hiding any contraband or weapons, there seems to be a linear relation (figure 3.4.1) between Strip searches and arrests as expected.

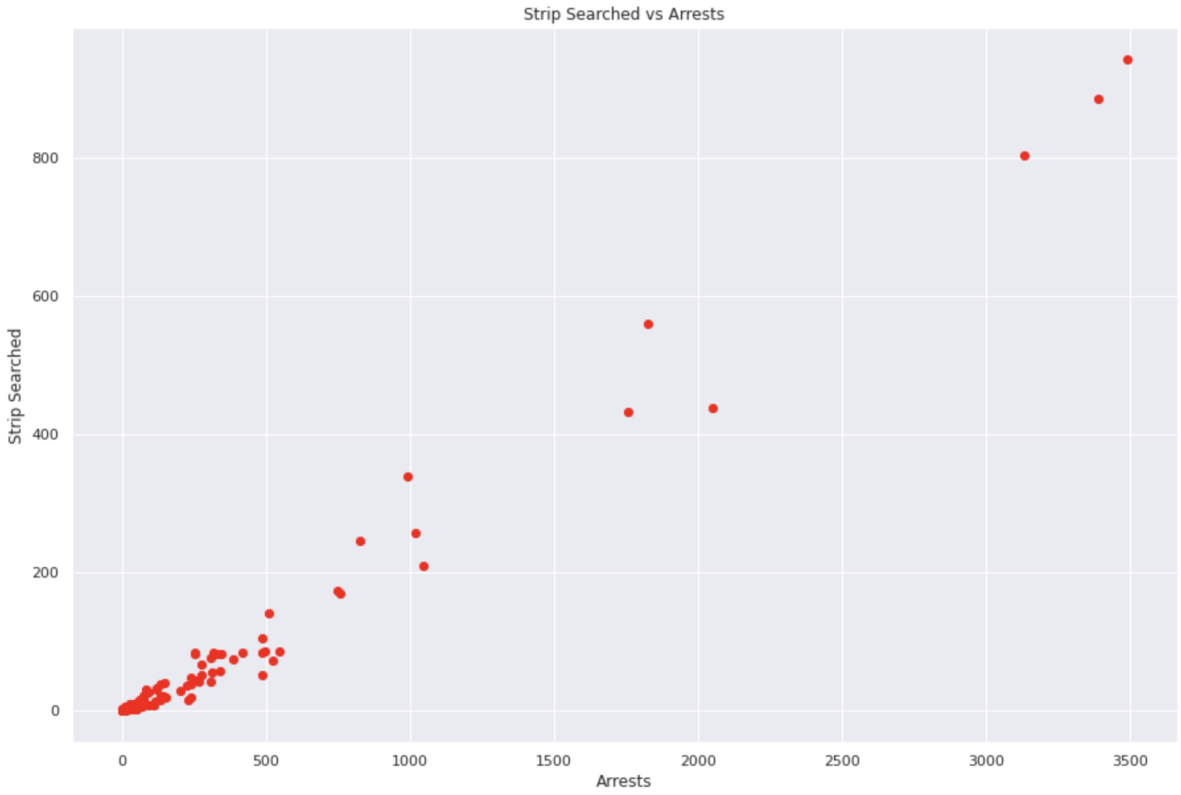

Figure 3.4.1 Scatterplot showing Strip Search and arrests

4. Research Questions/Research Design

This report aims to find if there is any association between the perceived race and gender and the strip searches, along with analyzing the aggression levels of individuals during these encounters with police across different age groups.

To put it formally, and in the statistical framework, below are the hypotheses we will be testing in this report.

RQ1: Do men and women across all age groups have same level of displayed aggression when being strip searched and booked?

H_0:All the means of the aggression levels for each combination of gender and age group are the same.

H_1:At least one aggression level of a combination of gender and age group differs from the others.

RQ2: Do people of all races and gender are equally likely to be strip searched and booked?

H_0:All the means of the strip searches and arrests for each combination of gender and perceived race are the same.

H_1:At least one strip search/arrest level of a combination of gender and percieved race differs from the others.

5. Methods

5.1 Dataset Description

This dataset contains data about arrests (arrest ID, event ID, time, location, category, action at arrest), strip searches (strip search indicator, reasons, items found), and demographic information of arrestees (person ID, perceived race, gender, age group) in Toronto. Whether an arrestee was booked within one day after an arrest event is also indicated in this dataset. All the variables in this dataset are nominal (IDs, race, sex, arrest location, occurrence category), ordinal (arrest month, arrest year, age group), and binary (youth at arrest, various actions at arrest, strip search reasons, and other indicators).

To investigate the underlying relations between the gender as well as the age of an arrestee and how aggressive he/she is when arrested, only the columns “Sex”, “Age_group__at_arrest_”, “Actions_at_arrest___Combative__”, and “Actions_at_arrest___Assaulted_o” are of interest. However, there are some duplicate and incomplete data in these 4 columns. For instance, there are 9 different values in the age group (at arrest) column: aged 17 years and younger, aged 17 years and under, aged 18 to 24 years, aged 25 to 34 years, aged 35 to 44 years, aged 45 to 54 years, aged 55 to 64 years, aged 65 years and older, aged 65 and older. To implement the experiment more conveniently and accurately, it is necessary to merge “aged 17 years and younger” and “aged 17 years and under” as well as “aged 65 years and older” and “aged 65 and older”, respectively. Additionally, some rows with unclear sex represented by “U” cannot be analyzed in this research and thus will be deleted.

5.2 ANOVA

For RQ1, to determine the gender of sex and age of a person have a significant effect on how aggressive he/she is at arrest. To measure the aggressiveness level, we calculate the sums of the values in the “Actions_at_arrest___Combative__” and “Actions_at_arrest___Assaulted_o” columns for each row as the new column “Aggression_Level” where the values vary from 0 to 2. The higher the aggression level is, the more aggressive the arrestee is. As mentioned in section 5.1, we merged some of the duplicate groups like “aged 65 and older” and aged 65 years and older” in the “Age_group__at_arrest_” column. As a result, there are 7 age groups in this column. We also

deleted rows with missing age groups or genders that are not “M” or “F” and columns that are not interested in this research. As a result, we have 65243 rows and 3 columns.

In the first place, a 2 (sex) by 7 (age group) two-way ANOVA test with interaction and another two-way ANOVA without interaction were conducted. In these two tests, the two categorical explanatory variables are sex and age group, while the aggression level serves as the quantititave outcome. Some functions from the *statsmodels*, *pandas*, and *matplotlib* modules are used to carry out ANOVA tests and make an interaction plot. We also assume that the aggression levels for each combination of sex and age group are normally distributed and have equal variances. Obviously, the gender and the age of a person are independent of each other. The hypothses are listed as follows:

H_0*:All the means of the aggression levels for each combination of sex and age group are the same.*

H_1*:At least one aggression level of a combination of sex and age group differs from the others.*

After that, using functions from the *scipy* model, a one-way ANOVA where the sex and aggression levels are used as the categorical variable and the quantitative outcome, respectively, was implemented to study whether there is statisitcal evidence that gender has a measurable effect on the aggressiveness level. Similarly, it is assumed that the aggression levels for both genders are normally distributed and have equal variances. Again, the gender and the age of a person are independent of each other. The hypothses of this test are:

H_0: $The\ means\ of\ the\ aggression\ levels\ for\ both\ genders\ are\ the\ same.$

H_1: $The\ means\ of\ the\ aggression\ levels\ for\ both\ genders\ are\ not\ the\ same.$

For RQ2, to see if people of all races and gender are equally likely to be strip searched and booked, we created a subset of the data called stripSearchs which contains all the relevant columns to our research namely, Perceived_Race, Sex, Age_group__at_arrest_. We than ran another ANOVA with and without interaction where Sex and Perceived race are the two categorical variables and the strip search is the quantitative variable.

After running the 2 Way Anova, we factorized the Sex column and ran a 1 Way anova between sex and the StripSearch variable to study whether there is statistical evidence that gender has a measureable effect on being strip searched. Similarly to RQ1, it is assumed that the strip search levels for both genders are normally distributed and have equal variances. Again, the gender and the perceived race of a person are independent of each other.

5.3 Post-Hoc Test

From ANOVA tests we will know if we have any statistically significant differences between the means of our categorical variables but to know which pairs have statistically significant difference we will be running Tukey's HSD test [4], it is to show the difference between the specific groups.

6. Results and Findings

6.1 RQ1

ANOVA

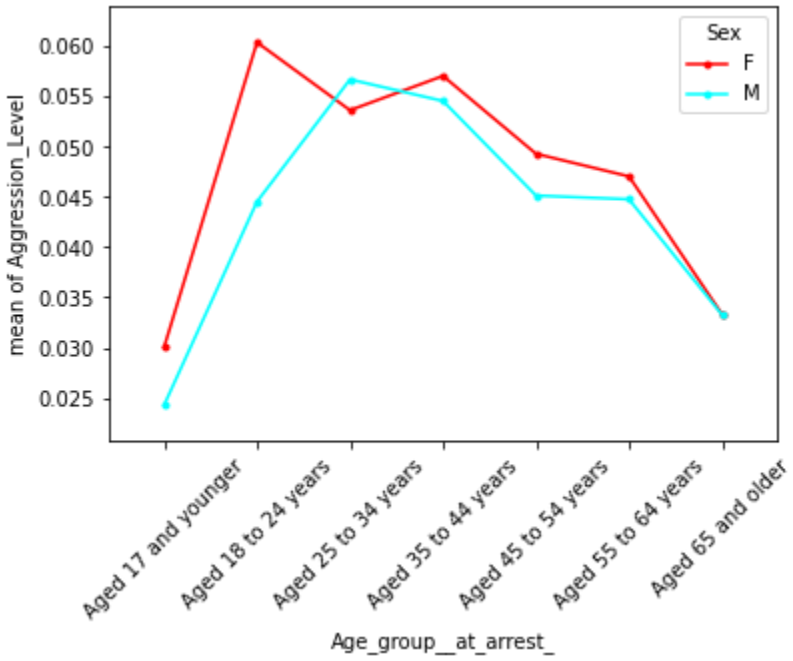

Figure 6.1.1 Interaction Plot

The diagram above (Figure 6.1.1) is the interaction plot of aggression levels by age group, while the lines of different colors are used for sex of gender. Each point is a sample mean of the outcome for a combination of one age group and gender. As we can see from this diagram, the lines for both genders are the lowest for arrestees aged 17 and younger, increase with the age and reach their peaks within the range from 18 to 34 before start to

decreasing. Overall, these two graphs have generally similar trends. However, they differ from each other in some aspects. For example, the female line peaks at the age group of 18 to 24 and then fluctuates a little bit until the age of 45 to 54 where the decline initiates, while the male group reaches its highest point at the age of 25 to 34 and starts dropping right after that. Moreover, the aggression levels of females are higher than those of males at most points, especially for the age from 18 to 24, but for the age group of 25 to 34, male arrestees are relatively more aggressive.

The following table (6.1.2) summarizes the two-way ANOVA with interaction. There are 2 and 7 categories of sex and age group, respectively, so the degree of freedom (df) for them are 1 (2 minus 1) and 6 (7 minus 1), while the df of the interaction is also 6 (7 minus 1 multiplied by 2 minus 1). The df of the residual is determined by the sample size minus the product of 2 and 7 which is 65229. As shown by this table, the F-statistic of the age group is much larger than those of the sex and interaction, and its p-value is also significantly smaller than those of the latter. As the interaction p-value (0.3056) is quite large using a significance level of $\alpha = 0.05$, we cannot reject the additive model, and that's why we rerun the ANOVA without interaction.

	sum_sq	df	F	PR(>F)
C(Sex)	0.100795	1.0	1.784902	1.815532e-01
C(Age_group__at_arrest_)	3.666315	6.0	10.820709	4.534492e-12
C(Sex):C(Age_group__at_arrest_)	0.404768	6.0	1.194627	3.056266e-01
Residual	3683.523975	65229.0	NaN	NaN

Table 6.1.2 Showing ANOVA Results

The table (6.1.3) below shows the results of the two-way ANOVA without interaction. In this case, the df of residual equals the sample size minus the df of two variables and 1 (65243 – 1 – 6 – 1). Again, the p-value of the age group is much smaller than the p-value of the sex, and its F-statistic is greatly larger than the latter.

	sum_sq	df	F	PR(>F)
C(Sex)	0.100795	1.0	1.784870	1.815571e-01
C(Age_group__at_arrest_)	3.666315	6.0	10.820515	4.536962e-12
Residual	3683.928743	65235.0	NaN	NaN

Table 6.1.3 Showing ANOVA Results

To further study if the gender has a statistically significant effect on the aggression level at arrest, we implement a one-way ANOVA on sex. According to the calculation results, the F-statistic is about 1.714, and the p-value is around 0.19.

Tukey's Test

Tukey's results show that there is a significant difference between Aged 17 and Younger and Aged 25-34, followed by Aged 17 and younger and Aged 35-44 but you can see that the difference even though statistically significant is not a lot i.e. an individual aged 25-34 is 0.035 times more likely to get more aggressive that someone who is 17 years and younger.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Aged 17 and younger	Aged 18 to 24 years	0.0222	0.001	0.0077	0.0367	True
Aged 17 and younger	Aged 25 to 34 years	0.0305	0.001	0.0169	0.0441	True
Aged 17 and younger	Aged 35 to 44 years	0.0294	0.001	0.0156	0.0432	True
Aged 17 and younger	Aged 45 to 54 years	0.0202	0.001	0.0056	0.0349	True
Aged 17 and younger	Aged 55 to 64 years	0.0195	0.0083	0.0031	0.0359	True
Aged 17 and younger	Aged 65 and older	0.0076	0.9	-0.0154	0.0307	False
Aged 18 to 24 years	Aged 25 to 34 years	0.0083	0.0622	-0.0002	0.0168	False
Aged 18 to 24 years	Aged 35 to 44 years	0.0072	0.1995	-0.0017	0.0161	False
Aged 18 to 24 years	Aged 45 to 54 years	-0.0019	0.9	-0.0121	0.0082	False
Aged 18 to 24 years	Aged 55 to 64 years	-0.0027	0.9	-0.0152	0.0098	False
Aged 18 to 24 years	Aged 65 and older	-0.0145	0.3594	-0.035	0.006	False
Aged 25 to 34 years	Aged 35 to 44 years	-0.0011	0.9	-0.0084	0.0063	False
Aged 25 to 34 years	Aged 45 to 54 years	-0.0102	0.0112	-0.019	-0.0014	True
Aged 25 to 34 years	Aged 55 to 64 years	-0.011	0.069	-0.0224	0.0004	False
Aged 25 to 34 years	Aged 65 and older	-0.0228	0.0126	-0.0427	-0.0029	True
Aged 35 to 44 years	Aged 45 to 54 years	-0.0092	0.0512	-0.0183	0.0	False
Aged 35 to 44 years	Aged 55 to 64 years	-0.0099	0.1596	-0.0216	0.0018	False
Aged 35 to 44 years	Aged 65 and older	-0.0218	0.0231	-0.0418	-0.0017	True
Aged 45 to 54 years	Aged 55 to 64 years	-0.0008	0.9	-0.0135	0.0119	False
Aged 45 to 54 years	Aged 65 and older	-0.0126	0.5403	-0.0332	0.008	False
Aged 55 to 64 years	Aged 65 and older	-0.0118	0.6614	-0.0337	0.01	False

Table 6.1.4 Tukey's Results

6.1 RQ2

ANOVA

For RQ2, When we run a 2-Way ANOVA with interaction, from the following table we can see that the P-Value is statistically significant for all three, Sex, Perceived Race and the interaction of Sex and Perceived Race which means that all three have measurable effect on being strip searched.

	sum_sq	df	F	PR(>F)
C(Sex)	5.740725e+06	2.0	13.810559	0.000005
C(Perceived_Race)	1.521349e+07	7.0	10.456953	0.000005
C(Sex):C(Perceived_Race)	9.086819e+06	14.0	3.122902	0.001659
Residual	2.078383e+07	100.0	NaN	NaN

Table 6.1.5 ANOVA Results

A One-Way ANOVA between Sex and Strip Search also returns a P-value of less than 0.05 showing sex has a statistically significant and measurable difference when it comes to being strip searched.

Tukey's Post HOC Test

If you look at Table 6.1.6, we can see that the biggest difference of being strip searched is between White and Latino where a white person is 200 times more likely to be strip searched, followed by white vs Middle-Eastern and White vs South Asians.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-149.5	0.1168	-316.9357	17.9357	False
Black	Indigenous	-152.0	0.1044	-319.4357	15.4357	False
Black	Latino	-165.0571	0.0489	-329.6786	-0.4357	True
Black	Middle-Eastern	-157.5714	0.0811	-325.0071	9.8643	False
Black	South Asian	-155.5	0.0888	-322.9357	11.9357	False
Black	Unknown or Legacy	-142.3277	0.1191	-302.206	17.5505	False
Black	White	35.9076	0.9	-123.9707	195.7858	False
East/Southeast Asian	Indigenous	-2.5	0.9	-169.9357	164.9357	False
East/Southeast Asian	Latino	-15.5571	0.9	-180.1786	149.0643	False
East/Southeast Asian	Middle-Eastern	-8.0714	0.9	-175.5071	159.3643	False
East/Southeast Asian	South Asian	-6.0	0.9	-173.4357	161.4357	False
East/Southeast Asian	Unknown or Legacy	7.1723	0.9	-152.706	167.0505	False
East/Southeast Asian	White	185.4076	0.0115	25.5293	345.2858	True
Indigenous	Latino	-13.0571	0.9	-177.6786	151.5643	False
Indigenous	Middle-Eastern	-5.5714	0.9	-173.0071	161.8643	False
Indigenous	South Asian	-3.5	0.9	-170.9357	163.9357	False
Indigenous	Unknown or Legacy	9.6723	0.9	-150.206	169.5505	False
Indigenous	White	187.9076	0.0098	28.0293	347.7858	True
Latino	Middle-Eastern	7.4857	0.9	-157.1357	172.1072	False
Latino	South Asian	9.5571	0.9	-155.0643	174.1786	False
Latino	Unknown or Legacy	22.7294	0.9	-134.1991	179.658	False
Latino	White	200.9647	0.0033	44.0362	357.8933	True
Middle-Eastern	South Asian	2.0714	0.9	-165.3643	169.5071	False
Middle-Eastern	Unknown or Legacy	15.2437	0.9	-144.6346	175.122	False
Middle-Eastern	White	193.479	0.0069	33.6007	353.3573	True
South Asian	Unknown or Legacy	13.1723	0.9	-146.706	173.0505	False
South Asian	White	191.4076	0.0079	31.5293	351.2858	True
Unknown or Legacy	White	178.2353	0.0101	26.2899	330.1807	True

Table 6.1.6 Tukey's Results

If you look at Table 6.1.7, we can see that males are 93 times more likely to be strip searched than female.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	93.4643	0.0045	24.9176	162.011	True
0	2	-22.9107	0.9	-168.3202	122.4987	False
1	2	-116.375	0.1432	-261.7845	29.0345	False

Table 6.1.7 Tukey's Results

7. Discussion and Conclusion

According to the interaction plot in the last section, the two graphs overall have similar trends and patterns, which is consistent with the calculation results of the two-way ANOVA with interaction that the interaction p-value (0.3056) is not significant using $\alpha = 0.05$, meaning that we have no evidence to reject the additive model and conclude that the effects of the age on the aggression level are the same for both genders, and vice versa. This is the reason why we carry out one more two-way ANOVA without interaction which gives similar results. The p-value of the age group (4.53e-12) is significant, which indicates that the age group does affect how aggressive an arrestee is when arrested. Because the age group has 7 levels (6 df), we can directly check the estimated means table below to see which age group tends to be more or less aggressive. The results show that arrestees who are 25 to 44 years old are more aggressive than the others, while those who are 17 and younger or 65 and older are the least aggressive at attest. This is not a contradiction with the observation that some youth are more violent at arrest as it is only a statistical calculation and conclusion. Nonetheless, the p-value of the sex (0.182) is not significant, so there is no clear evidence that gender plays a role in the aggressiveness of a person at arrest.

Age_group__at_arrest_	Aggression_Level
Aged 17 and younger	0.025641
Aged 18 to 24 years	0.047814
Aged 25 to 34 years	0.056099
Aged 35 to 44 years	0.055046
Aged 45 to 54 years	0.045886
Aged 55 to 64 years	0.045118
Aged 65 and older	0.033283

As mentioned before, a one-way ANOVA was carried out to confirm that gender does not significantly affect the aggression level at arrest. Using a significance level $\alpha = 0.05$, the p-value (0.19) is not statistically significant. Hence, we retain the null hypothesis that the means of aggressiveness level for both genders can be considered the same, consistent with the results of the two-way ANOVA tests.

Although the statistical calculations tell us there is no interaction between the sex and the age group, the interaction plot shows that the first half of the graphs are not parallel at all as they have different peaks, slopes, and trends, which is different than elder age groups. It is possible that there might be an interaction between the two variables in terms of younger age groups, and the ANOVA results could be different if we do the test on only younger arrestees. Similarly, we can see from the interaction plot that females are generally more aggressive than males, especially for those aged from 18 to 24, in contrast to the ANOVA result that gender does not statistically affect aggression level. Again, the results might be different if we study the younger and the elder groups separately.

For RQ2, we can reject the null hypothesis i.e. All the means of the strip searches and arrests for each combination of gender and perceived race are the same as the p-value for both Sex

and Perceived Race along with the interaction of both is less than 0.05. The tukeys test also confirms the findings and demonstrates the groups where the largest difference lies.

When it comes to the limitations of this study, the accuracy of these conclusions might be affected if the assumptions we made are not true. To do these ANOVA tests, it is assumed that the aggression levels, Strip Search levels for each gender and each combination of gender and age group as well as Perceived Race are normally distributed and have the same variances, which is not necessarily correct and thus lowers the accuracy of the results. Calculating the sum of indicators of whether combative, violent, or spitter/biter and assaulted officers to measure how aggressive an arrestee is might be too rough and subjective for this model, as there may be other aspects to measure the aggressiveness. This experiment can be improved by looking for a more comprehensive, accurate, and normally-distributed measure to represent the degree of aggression.

8. References

1. Toronto Police Service. (2022, November 10). Arrests and Strip Searches (RBDC-ARR-TBL-001). Retrieved from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about.

2. Wendy Gillis. (2020, November 24). Retrieved from https://www.thestar.com/news/gta/2020/11/23/clearly-we-were-doing-it-wrong-toronto-police-doing-far-fewer-strip-searches-under-new-rules-interim-chief-says.html

3. Gerry McNeilly. (2019 March). Summary of Ontario Jurisprudence Involving Strip Searches Post R. v. Golden. Retrieved from https://www.oiprd.on.ca/wp-content/uploads/OIPRD-Summary-of-Ontario-Jurisprudence-Involving-Strip-Searches-Accessible.pdf

4. Renesh Bedre. (2022, March, 6). Tukey's HSD. Retrieved from https://www.reneshbedre.com/blog/anova.html