INF2178 Experimental Design

Midterm Project

Group 66
Ke Wu, Ruiying Xu

Table of Contents

Part 1: Introduction

In 2020, the murder of George Floyd by a Minneapolis police officer who knelt on his neck for over nine minutes triggered global protests and brought the issues of police brutality and systemic racism to the forefront of public consciousness (Murder of George Floyd, 2023). These concerns are not new, as racial discrimination has long been a source of contention in police systems worldwide, and marginalized groups have called for fairness in the arrest process. Studies suggest that factors such as gender, age, and other demographic information may also impact fairness in police encounters. Our dataset includes information on the police's perception of suspects' actions and judgments on the necessity of strip-search during arrest and the race, age, and gender of these suspects. We aim to investigate whether our dataset's findings provide additional evidence to support the notion that fairness in police encounters is closely linked to suspects' demographic information, particularly race.

1.1 Literature Reviews

In the study of all areas of police-public contact, race has been identified as one of the most commonly studied predictors of dissatisfaction with the police. Pollock and Menard (2015). In another study conducted by Gabbidon and Higgins in 2009, it was observed that the perception of fairness in police contact situations is still closely tied to race, particularly for African Americans. The study revealed that African Americans believe that Black individuals are treated less fairly by police compared to White individuals. White individuals surveyed in the study did not perceive any difference in fairness based on race. This contrast in perceptions based on race raises some essential questions about the factors contributing to the perception of fairness in police encounters. Additionally, it is important to recognize that other variables may also play a role in shaping an individual's perception of fairness in police encounters. Factors such as age, gender, and other demographic variables may also influence how individuals perceive and interact with the police (Bradford et al., 2009).

1.2 Research Objectives and Questions

The objectives of our research are to examine how demographic factors such as gender, age, and race affect suspects' performances when getting arrested and police judgments on the necessity of strip-search in Toronto to contribute to the existing literature and provide a further understanding of the relationship between demographic factors and police encounters. Specifically, we wish to examine the effects of age and race on suspects' actions at arrest as well as the effects of gender and race on the police's reasons for strip-searching suspects. We conducted exploratory data analysis including descriptive statistics and t-tests to get an overview of our dataset, and formed 2 research questions based on the results of our EDA and the literature reviews as follows:

- Research Question 1: How do age and race affect suspects' actions at arrest in Toronto separately and interactively?
- Research Question 2: How do gender and race affect the police's judgments on the necessity of strip-search of the suspects? To be more specific, how do gender and race affect police's reasons for strip-searching the suspects?

By answering our research questions, we aim to enhance the understanding of the complex relationship between demographic factors and police encounters, boost the Toronto police's

efficiency by better predicting suspects' reactions and the need for strip-searching, inform training and education for police officers, and improve their ability to respond to different racial suspects in different ages, ultimately reducing the likelihood of accidental injuries to marginalized groups.

Part 2: EDA

2.1 Brief Dataset Description

The data set was provided by Toronto Police Service and was available in a CSV file on Public Safety Data Portal at the link in the appendix. This data set includes all information about arrests and strip searches in Toronto in 2020 and 2021. It has 65276 records of arrests and 24 attributes, which include Arrest Year, Arrest Month, EventID, ArrestID, PersonID, Perceived Race, Sex, Age Group (at arrest), Youth at Arrest (under 18 years old), ArrestLocDiv, StripSearch, Booked, Occurrence Category, Actions at Arrest - Concealed Items, Actions at Arrest - Combative, violent or spitter/biter, Actions at Arrest - Resisted, defensive or escape risk, Actions at Arrest - Mental instability or possibly suicidal, Actions at Arrest - Assaulted officer, Actions at Arrest - Cooperative (6 possible actions at arrests in total), SearchReason - CauseInjury, SearchReason - AssistEscape, SearchReason - PossessWeapons, SearchReason - PossessEvidence (4 possible search reasons in total), and ItemsFound. Data is provided in binary forms 0,1 (eg. StripSearch, Booked, Actions at Arrest, SearchReasons, and ItemsFound), numeric integer forms (eg. IDs and ArrestLocDiv), or text forms (indicating different categories, eg. Age Group, Sex, and Perceived Race).

2.2 Data Preprocessing

Among the 65276 records, we found records with duplicate values in PersonID and null values. For duplicate values, an example would be that the suspect with PersonID 310375 appeared 10 times in the data set in different cases. We considered the duplicated values to be errors in data collection, since among the 10 records of PersonID 310375, three different racial groups were stated, including indigenous, unknown/legacy, and white, and we perceived that one person cannot belong to multiple racial groups simultaneously. Thus, we deleted all the records with duplicated PersonID. For null values, we removed all the records with null values in racial or age or ArrestID or occurrence category attributes and converted all null values in actions at arrest, search reasons, and items found to 0.

We then modified some variables, deleting the unknown or legacy" group in the Perceived Race attribute, deleting sex type "U" (since only 3 records had U as the sex type, which is extremely unbiased), replacing "aged 17 years and under" with "aged 17 and younger" and "aged 65 and older" with "aged 65 years and older". We also exchanged the values of 0 and 1 in the cooperative actions to make sure that the value 1 indicates uncooperativeness since our study focused more on negative suspects' reactions when getting arrested. To ensure consistency, we renamed the attribute to "Actions at Arrest - Uncooperative".

Most importantly, we created two new continuous variables. By aggregating the number of 1's in attributes "Actions at Arrest - Concealed Items", "Actions at Arrest - Combative, violent or spitter/biter", "Actions at Arrest - Resisted, defensive or escape risk", "Actions at Arrest - Mental instability or possibly suicidal", "Actions at Arrest - Assaulted officer", and "Actions at Arrest - Uncooperative", we created the number of (negative) actions at arrest. By aggregating the number of 1's in attributes "SearchReason - CauseInjury", "SearchReason - AssistEscape", "SearchReason - PossessWeapons", and "SearchReason - PossessEvidence", we created the number of search reasons. The number of (negative) actions at arrest acted as the measurement for the outcome in our first research question, suspects actions at arrest; while the number of search reasons acted as the measurement for the outcome in our second research question, police's judgments on the necessity of strip search the suspects/police's reasons for strip searching the suspects.

Moreover, we changed the data types into categorical data and quantitative data (integers) correspondingly. Finally, we formed a new and clean dataset by deleting some unwanted columns and keeping the ones closely related to our research questions, which include PersonID, Perceived Race, Sex, Age group at arrest, StripSearch, Number of actions at arrest, and the number of search reasons. Below is part of the new dataset we created for this research. (Table 1. A subset of our final dataset)

Table 1. A subset of our final dataset

	PersonID	Perceived_Race	Sex	Age_group__at_arrest_	StripSearch	Number_actions_at_arrest	Number_search_reason
13	328630	Indigenous	M	Aged 45 to 54 years	0	1	0
22	324224	Black	M	Aged 45 to 54 years	0	0	0
23	302537	White	M	Aged 25 to 34 years	0	1	0
24	305813	Black	M	Aged 25 to 34 years	0	1	0
25	323720	White	M	Aged 35 to 44 years	0	1	0

2.3 Descriptive Statistics

To briefly understand the data set, we first produced two count charts showing the number of suspects in different racial groups and age groups. From Figure 2.1 we can tell that there are 7 different racial groups and the suspects are not evenly distributed into each group. We see the number of white suspects ranked first (9944), followed by black suspects (6800), while the number of indigenous suspects ranked last (343). This is reasonable given the uneven population distribution in Toronto, nevertheless not a good phenomenon for our further analysis. Figure 2.2 shows that there are 7 levels of age groups and the age group from 25 to 34 tends to have the most suspects (7075). There is a non-negligible difference in the number of suspects in different age groups and we see a right-skewed distribution from the chart. We also created a frequency table for sex groups (Table 2.1), from which we find that there are 18733 male suspects and 5289 female suspects, indicating large gender disparity in the dataset.

Figure 2.1. The number of suspects in each racial group in this dataset

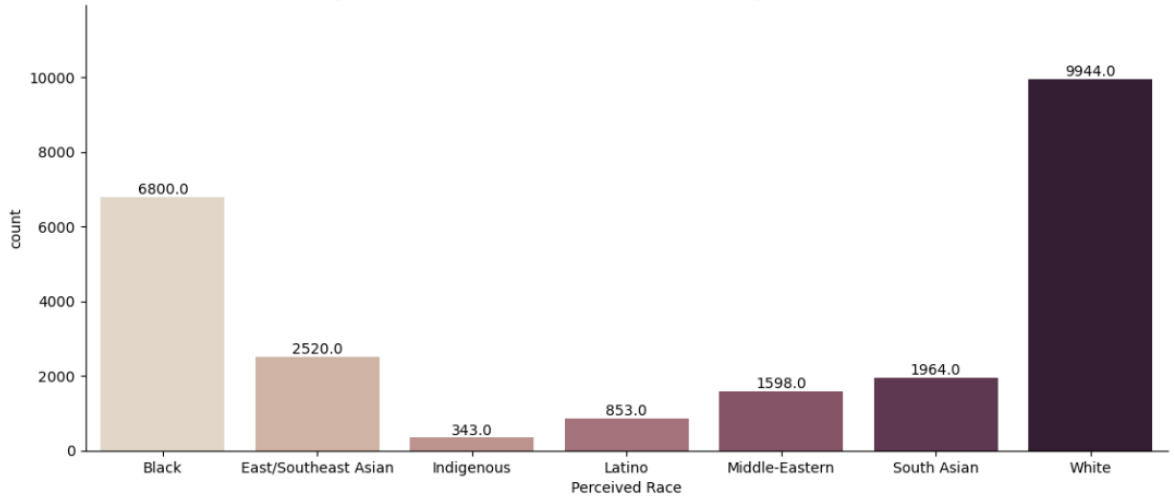


Figure 2.2.The number of suspects in each age group at arrested in this dataset

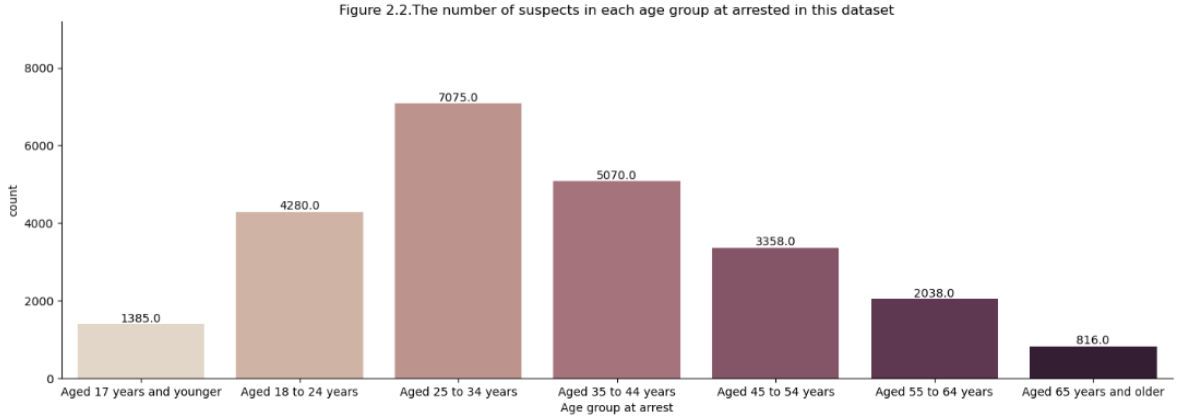


Table 2.1. The number of female and male suspects in this dataset

	count	pct
M	18733	0.779827
F	5289	0.220173

Figure 2.3. The total number of suspects in each number of actions(0,1,2,3,4) in this dataset

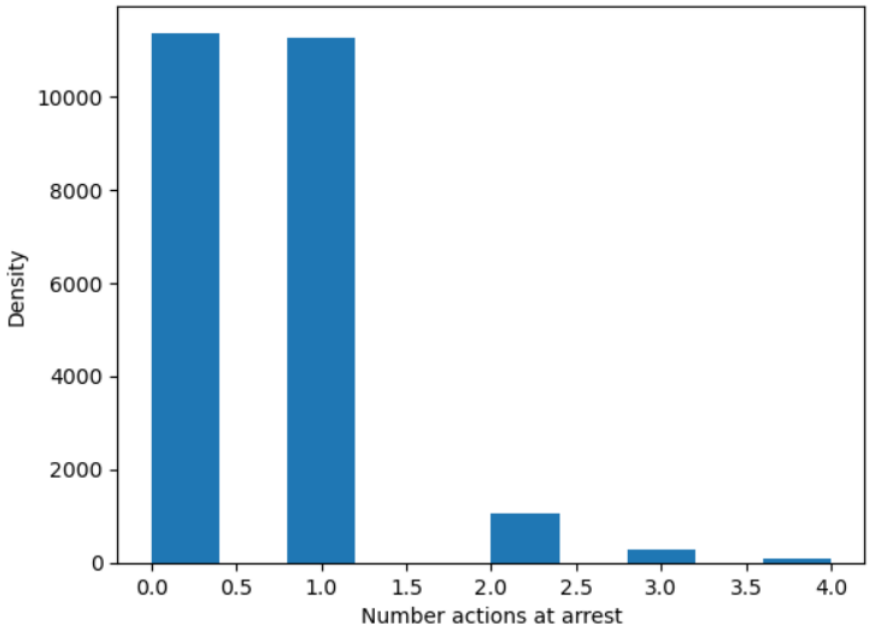
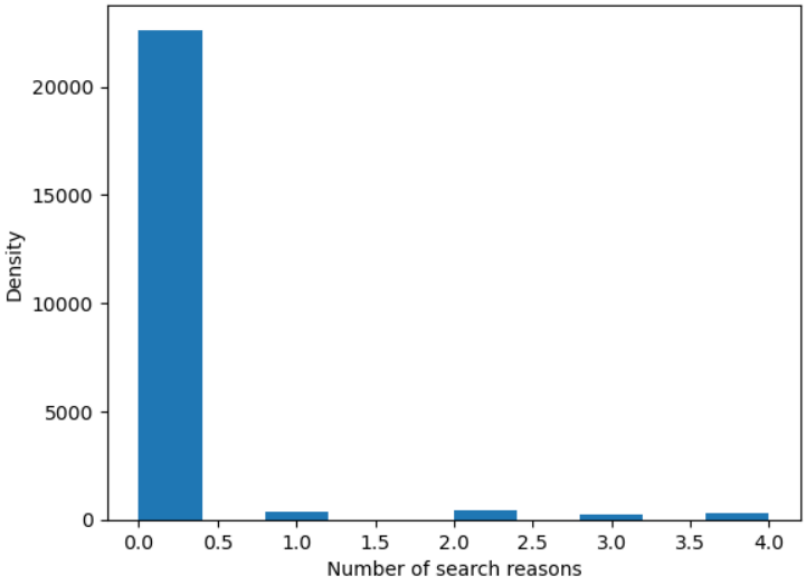


Figure 2.4. The total number of suspects in each number of search reasons (0,1,2,3,4) in this dataset



The density plot for the number of actions at arrest (Figure 2.3) indicates that the most suspects have either zero or one action at arrest. There is a vast decline in the number of suspects with the increase in the number of actions, showing that the number of actions is not normally distributed. The density plot for the number of search reasons (Figure 2.4) illustrates that for most suspects, the police have no search reason based on their judgment (since only 6% people were strip-searched after arrested), and there is a significant drop in the number of suspects who get strip-searched for at least one reason. This suggests that the number of search reasons does not follow a normal distribution.

Due to the limited range of values for both the numbers of actions at arrest and the numbers of search reasons, ranging from 0 to 4 (with no instances of 5 or 6 in the number of actions at arrest), the resulting boxplots exhibit a skewed distribution and are subject to bias (Figure 2.5 & 2.6). Therefore, we decided to include a summary of statistics (Table 2.2 & 2.3) to visualize the relationships between continuous and categorical variables, in addition to the boxplots. The summary of statistics (Table 2.2&2.3) informs that the mean numbers of actions at arrest are different among different racial groups and age groups and the mean numbers of search reasons are different among different racial groups and gender groups, which stimulateds our interest in further investigation.

Figure 2.5. The total number of suspects in each number of actions at arrest(0,1,2,3,4) for each racial group in this dataset

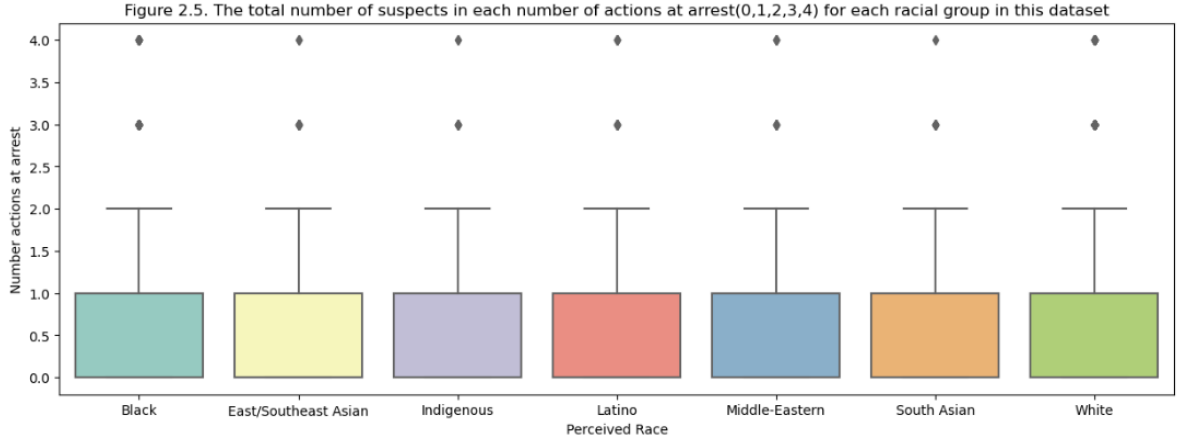


Table 2.2. The summary statistics of number of actions at arrest(0,1,2,3,4) for each racial group in the dataset

Perceived_Race	count	mean	std	min	25%	50%	75%	max
Black	6800.0	0.630000	0.678183	0.0	0.0	1.0	1.0	4.0
East/Southeast Asian	2520.0	0.525000	0.586478	0.0	0.0	0.0	1.0	4.0
Indigenous	343.0	0.784257	0.737538	0.0	0.0	1.0	1.0	4.0
Latino	853.0	0.512309	0.672324	0.0	0.0	0.0	1.0	4.0
Middle-Eastern	1598.0	0.577597	0.618985	0.0	0.0	1.0	1.0	4.0
South Asian	1964.0	0.587576	0.597131	0.0	0.0	1.0	1.0	4.0
White	9944.0	0.616452	0.680703	0.0	0.0	1.0	1.0	4.0

Figure 2.6. The total number of suspects in each number of actions at arrest(0,1,2,3,4) for each age group in this dataset

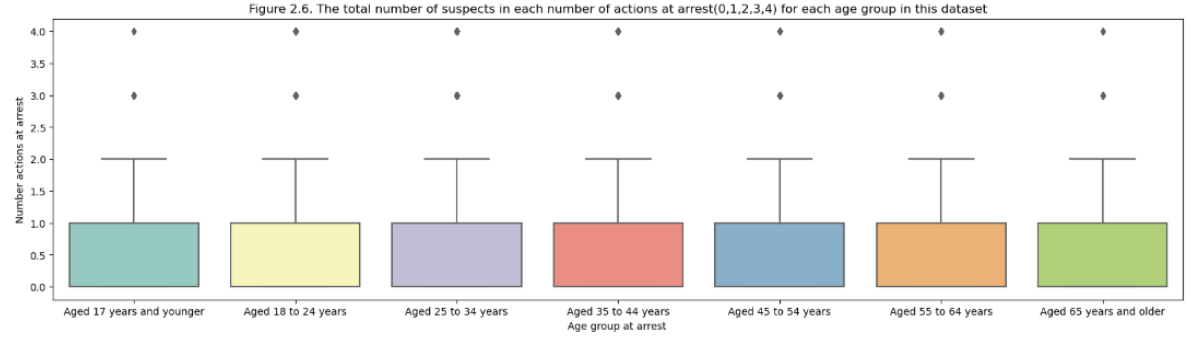


Table 2.3. The summary statistics of number of actions at arrest(0,1,2,3,4) for each age group in the dataset

Age_group__at_arrest_	count	mean	std	min	25%	50%	75%	max
Aged 17 years and younger	1385.0	0.561011	0.592032	0.0	0.0	1.0	1.0	4.0
Aged 18 to 24 years	4280.0	0.641822	0.674012	0.0	0.0	1.0	1.0	4.0
Aged 25 to 34 years	7075.0	0.621060	0.680212	0.0	0.0	1.0	1.0	4.0
Aged 35 to 44 years	5070.0	0.583432	0.657627	0.0	0.0	1.0	1.0	4.0
Aged 45 to 54 years	3358.0	0.582787	0.641733	0.0	0.0	1.0	1.0	4.0
Aged 55 to 64 years	2038.0	0.610402	0.670092	0.0	0.0	1.0	1.0	4.0
Aged 65 years and older	816.0	0.542892	0.623171	0.0	0.0	0.0	1.0	4.0

In line with our prior observations, we have also conducted the box plots (Figure 2.7 & 2.8) and summary statistics (Table 2.4&2.5) to examine the mean differences of the number of search reasons among different gender and racial groups. As a result, we discovered that there is a mean differences in number of search reasons among gender and racial groups which calls for further explorations.

Figure 2.7. The total number of suspects in each number of search reasons(0,1,2,3,4) for each racial group in this dataset

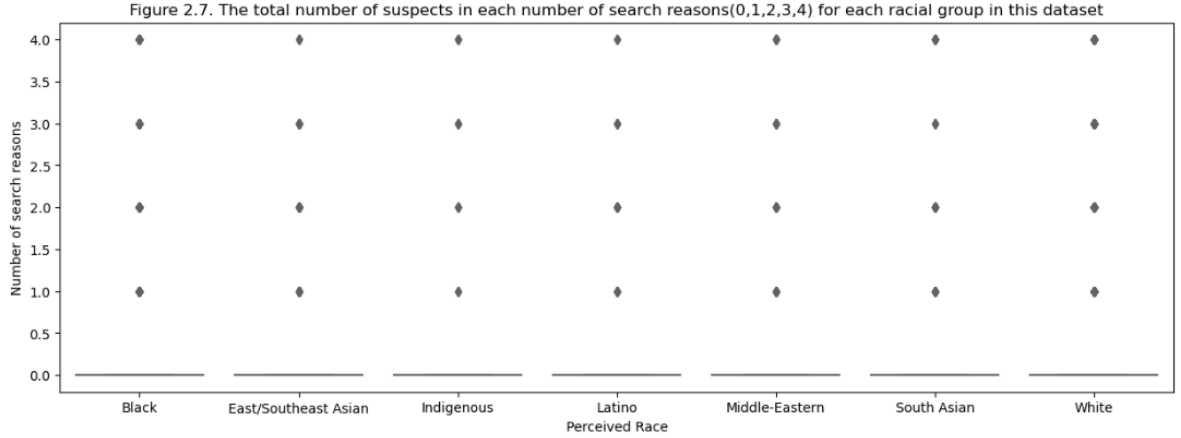


Table 2.4 The summary statistics of number of search reasons(0,1,2,3,4) for each racial group

Perceived_Race	count	mean	std	min	25%	50%	75%	max
Black	6800.0	0.192206	0.718895	0.0	0.0	0.0	0.0	4.0
East/Southeast Asian	2520.0	0.061508	0.371400	0.0	0.0	0.0	0.0	4.0
Indigenous	343.0	0.177843	0.717679	0.0	0.0	0.0	0.0	4.0
Latino	853.0	0.079719	0.442171	0.0	0.0	0.0	0.0	4.0
Middle-Eastern	1598.0	0.074468	0.443515	0.0	0.0	0.0	0.0	4.0
South Asian	1964.0	0.081976	0.474410	0.0	0.0	0.0	0.0	4.0
White	9944.0	0.134352	0.603595	0.0	0.0	0.0	0.0	4.0

Figure 2.8. The total number of suspects in each number of search reasons(0,1,2,3,4) for Female and male in this dataset

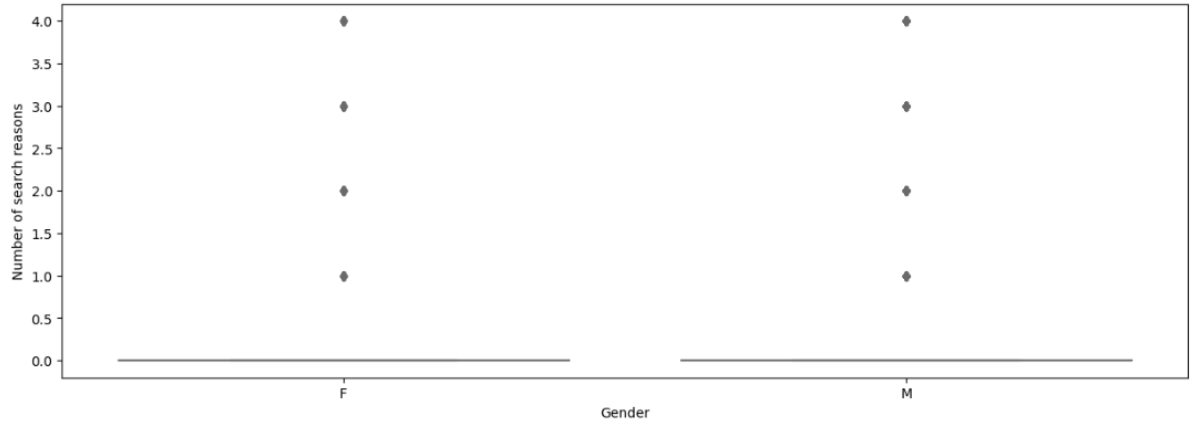


Table 2.5 The summary statistics of number of search reasons(0,1,2,3,4) for female and male

Sex	count	mean	std	min	25%	50%	75%	max
F	5289.0	0.104746	0.524724	0.0	0.0	0.0	0.0	4.0
M	18733.0	0.141622	0.618737	0.0	0.0	0.0	0.0	4.0

2.4 T-tests

Welch's t-test 1: Testing the differences in the mean number of actions at arrest for Indigenous and Latino racial groups

From Table 2.2 we found the mean number of actions at arrest for Indigenous suspects is around 0.784, which is higher than the mean number of actions at arrest for Latino suspects (around 0.512). Therefore, we decided to further explore whether there exist differences in the mean numbers of actions at arrest (outcome variable) between Indigenous suspects and Latino suspects (two-level explanatory variables). We tested the following hypothesis:

H0 (null hypothesis): the population means of the numbers of actions at arrest between the two independent groups of suspects, the Indigenous suspects and Latino suspects, are equal.

Ha (alternative hypothesis): the population means of the numbers of actions at arrest between the two independent groups of suspects, the Indigenous suspects and Latino suspects, are different.

The result of the Welch's t-test 1 is stated as follows. (Table 2.6.) We found that the population mean numbers of actions at arrest of the Latino suspects (Mean= 0.512, SD=0.672, 95% CI= [0.467, 0.557]) is lower than that of the Indigenous suspects (Mean=0.784, SD=0.738, 95% CI= [0.706, 0.863]). We set our significance level to 0.05. The p-value we received (around 0.0000) is smaller than the significance level, indicating a significant difference between the means. Therefore, we rejected H0.

Table 2.6. The result of t-test for the mean of number of action at arrest between indigenous and latino racial groups

	Satterthwaite t-test	results
0	Difference (Latino - Indigenous) =	-0.2719
1	Degrees of freedom =	582.6274
2	t =	-5.9122
3	Two side test p value =	0.0000
4	Difference < 0 p value =	0.0000
5	Difference > 0 p value =	1.0000
6	Cohen's d =	-0.3932
7	Hedge's g =	-0.3929
8	Glass's delta1 =	-0.4045
9	Point-Biserial r =	-0.2379

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	Latino	853.0	0.512309	0.672324	0.023020	0.467127	0.557492
1	Indigenous	343.0	0.784257	0.737538	0.039823	0.705927	0.862586
2	combined	1196.0	0.590301	0.702207	0.020305	0.550464	0.630138

Welch's t-test 2: Testing the mean differences in the number of actions at arrest for age 18-24 and age 17 and younger

From Table 2.3, we found the mean numbers of actions at arrest for suspects aged 17 and younger is around 0.561, which is lower than the mean numbers of actions at arrest for suspects aged 18-24 (around 0.642). Therefore, we decided to further explore whether there exist differences in the mean numbers of actions at arrest (outcome variable) between suspects aged 17 and younger and aged 18-24 (two-level explanatory variables). We tested the following hypothesis:

H0: the population means of the numbers of actions at arrest between the two independent groups of suspects, the suspects aged 17 and younger and the suspects aged 18-24, are equal.

Ha: the population means of the numbers of actions at arrest between the two independent groups of suspects, the suspects aged 17 and younger and the suspects aged 18-24, are different.

The result of the Welch's t-test 2 is stated as follows (Table 2.7). We found that the population mean number of actions at arrest of the suspects aged 17 and younger (Mean=0.561, SD=0.592, CI=[0.529, 0.592]) is lower than that of the suspects aged 18-24 (Mean=0.642, SD=0.674, CI= [0.622, 0.662]). The p-value we received (around 0.0000) is smaller than the significance level, indicating a significant difference between the means. Therefore, we rejected H0.

Table 2.7. The result of t-test for the mean of number of action at arrest between Aged 18 to 24 years and Aged 17 years and younger groups

	Satterthwaite t-test	results
0	Difference (Aged 18 to 24 years - Aged 17 year...	0.0808
1	Degrees of freedom =	2638.3094
2	t =	4.2638
3	Two side test p value =	0.0000
4	Difference < 0 p value =	1.0000
5	Difference > 0 p value =	0.0000
6	Cohen's d =	0.1234
7	Hedge's g =	0.1234
8	Glass's delta1 =	0.1199
9	Point-Biserial r =	0.0827

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	Aged 18 to 24 years	4280.0	0.641822	0.674012	0.010303	0.621624	0.662021
1	Aged 17 years and younger	1385.0	0.561011	0.592032	0.015908	0.529804	0.592218
2	combined	5665.0	0.622065	0.655787	0.008713	0.604985	0.639146

Welch's t-test 3: Testing the mean differences in the number of search reasons for black and East/ South Asian suspects

From Table 2.8, we found the mean numbers of search reasons for Black suspects is around 0.192, which is higher than the mean numbers of search reasons for East/South East Asian suspects (around 0.062). Therefore, we decided to further explore whether there exist differences in the mean number of search reasons (outcome variable) between East/South East Asian suspects and Black suspects (two-level explanatory variables). We tested the following hypothesis:

H0: the population means of the number of search reasons between the two independent groups of suspects, the East/South East Asian suspects and the Black suspects, are equal.

Ha: the population means of the number of search reasons between the two independent groups of suspects, the East/South East Asian suspects and the Black suspects, are different.

The result of the Welch's t-test 3 is stated as follows. (Table 2.8) We found that the population mean number of search reasons of the Black suspects (Mean=0.192, SD=0.719, CI=[0.175, 0.209]) is higher than that of the East/South East Asian suspects (Mean=0.062, SD=0.371, CI=[0.047, 0.076]). The p-value we received (around 0.000) is smaller than the significance level, indicating a significant difference between the means. Therefore, we rejected H0.

Table 2.8. The result of t-test for the mean of number of search reason between East/Southeast Asian and Black groups

	Satterthwaite t-test	results
0	Difference (East/Southeast Asian - Black) =	-0.1307
1	Degrees of freedom =	8382.8458
2	t =	-11.4305
3	Two side test p value =	0.0000
4	Difference < 0 p value =	0.0000
5	Difference > 0 p value =	1.0000
6	Cohen's d =	-0.2030
7	Hedge's g =	-0.2030
8	Glass's delta1 =	-0.3519
9	Point-Biserial r =	-0.1239

:

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	East/Southeast Asian	2520.0	0.061508	0.371400	0.007398	0.047000	0.076016
1	Black	6800.0	0.192206	0.718895	0.008718	0.175116	0.209296
2	combined	9320.0	0.156867	0.646307	0.006695	0.143744	0.169990

Welch's t-test 4: Testing the mean differences in the number of search reasons for male and female suspects

From Table2.5, we found the mean numbers of search reasons for male suspects is around 0.142, which is higher than the mean numbers of search reasons for female suspects (around 0.105). Therefore, we decided to further explore whether there exist differences in the mean numbers of search reasons (outcome variable) between male and female suspects (two-level explanatory variables). We tested the following hypothesis:

H0: the population means of the numbers of search reasons between the two independent groups of suspects, the male suspects and the female suspects, are equal.

Ha: the population means of the numbers of search reasons between the two independent groups of suspects, the male suspects and the female suspects, are different.

The result of the Welch's t-test 4 is stated as follows. (Table 2.9) We found that the population mean number of search reasons of the male suspects (Mean=0.142, SD=0.619, CI=[0.133, 0.150]) is higher than that of the female suspects (Mean=0.105, SD=0.525, CI=[0.091, 0.119]). The p-value we received (around 0.0000) is smaller than the significance level, indicating a significant difference between the means. Therefore, we rejected H0.

Table 2.9. The result of t-test for the mean of number search reason between female and male groups

	Satterthwaite t-test	results
0	Difference (F - M) =	-0.0369
1	Degrees of freedom =	9827.2275
2	t =	-4.3310
3	Two side test p value =	0.0000
4	Difference < 0 p value =	0.0000
5	Difference > 0 p value =	1.0000
6	Cohen's d =	-0.0615
7	Hedge's g =	-0.0615
8	Glass's delta1 =	-0.0703
9	Point-Biserial r =	-0.0436

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	F	5289.0	0.104746	0.524724	0.007215	0.090601	0.118890
1	M	18733.0	0.141622	0.618737	0.004521	0.132761	0.150483
2	combined	24022.0	0.133503	0.599490	0.003868	0.125921	0.141084

Assumption Checks for T-tests

Since all of our compared groups do not have equal variance with each other, we choose to use Welch's t-test. The assumptions of Welch's t-test are 1) at least 2 levels of categorical independent variables; 2) a quantitative dependent variable; 3) normality in the distribution of the groups; 4) independence of error. Through creating QQ plots, we found that all assumptions except for assumption 3 (normality) are met.

T-tests Concluding Remarks

The above t-tests show that the mean numbers of actions at arrest is significantly different depending on suspects' race and age, while the mean numbers of search reasons is significantly different depending on suspects' race and gender. These are important results that indicate the necessity of further research.

Part 3: Research Design and Methods

Based on the results of EDA, especially results from the t-tests, we decided to use inferential statistical testing to further study our research questions.

Research Question 1

From the first and second t-tests stated in **Section 2.4**, we see significant differences in the mean numbers of actions at arrest between Indigenous suspects and Latino suspects and between suspects aged 17 and younger and aged 18-24, indicating that different racial groups and age groups may have an impact on the number of actions at arrest. To further explore the question, we conducted two One-Way ANOVA (number of actions at arrest ~ racial group and number of actions at arrest ~ age group) and a Two-way ANOVA (an interaction model), hoping to explore the effect of each categorical independent variable (race and age) on the quantitative dependent variable (number of actions at arrest) and how the two independent variables interact with each other in affecting the outcome. Based on the results of ANOVA, we also conducted Tukey's HSD, performing multiple pairwise comparisons (post-hoc comparisons) to find out the pairs of significant treatments.

Research Question 2

From the third and fourth t-tests stated in **Section 2.4**, we see significant differences in the mean numbers of search reasons between Black suspects and East/South East Asian suspects and between male suspects and female suspects, indicating that different racial groups and gender groups may have an impact on the number of search reasons. To further explore the question, we conducted a Two-way ANOVA, hoping to explore the effect of each categorical independent variable (race and gender) on the quantitative dependent variable (number of search reasons) and how the two independent variables interact with each other in affecting the outcome. Based on the results of ANOVA, we also conducted Tukey's HSD, performing multiple pairwise comparisons (post-hoc comparisons) to find out the pairs of significant treatments.

Part 4: Results/ Findings

Research Question 1

One-Way ANOVA 1: Number of Actions at Arrest ~ Racial Group

We conducted a One-Way ANOVA to see how different levels of racial groups affect a suspect's number of actions at arrest. We made the null hypothesis and the alternative hypothesis as the follows.

H0: All of the population means of the number of actions at arrest are equal among different racial groups.

Ha: There exists at least one pair of racial groups, whose population means of the number of actions at arrest is not equal.

The results are shown as follows in Table 4.1. The F-statistics for the perceived race is 15.967, with a p-value of 2.055941e-18, which is significantly smaller than our pre-set significance level (0.05). Therefore, we reject the null hypothesis and reach the conclusion

that there exists at least one pair of racial groups, whose population means of the number of actions at arrest is not equal.

Table 4.1. The result for one-way anova for the mean of number of actions at arrest among different racial groups

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	41.821172	6.0	15.966735	2.055941e-18
Residual	10483.624003	24015.0	NaN	NaN

To further identify the pairs of significantly different treatments, we performed multiple pairwise comparison (post hoc comparison) analysis using Tukey's honestly significant difference (HSD) test.

We found 11 pairs of treatments with p-values smaller than 0.05, indicating that these 11 pairs of racial groups are significantly different in the population means of the numbers of actions at arrest, most of which involves indigenous suspects. The 11 pairs of racial groups are shown below. (Table 4.2)

Table 4.2. The statistically significant tukey tests results for the mean of number of actions at arrest among all different racial groups paisrs

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Indigenous	Black	0.154257	0.046446	0.262067	5.966304	0.001000
1	Indigenous	White	0.167804	0.060816	0.274793	6.540143	0.001000
2	Indigenous	South Asian	0.196680	0.082674	0.310686	7.193747	0.001000
3	Indigenous	Middle-Eastern	0.206660	0.090729	0.322590	7.433250	0.001000
4	Indigenous	East/Southeast Asian	0.259257	0.147136	0.371377	9.641997	0.001000
5	Indigenous	Latino	0.271947	0.147391	0.396503	9.104182	0.001000
9	Black	East/Southeast Asian	0.105000	0.059567	0.150433	9.636884	0.001000
10	Black	Latino	0.117691	0.046927	0.188454	6.935138	0.001000
13	White	East/Southeast Asian	0.091452	0.048004	0.134900	8.777008	0.001000
14	White	Latino	0.104143	0.034637	0.173648	6.247880	0.001000
16	South Asian	East/Southeast Asian	0.062576	0.003938	0.121215	4.449889	0.027587

One-Way ANOVA 2: Number of Actions at Arrest ~ Age Group

We conducted a One-Way ANOVA to see how different levels of age groups affect a suspect's number of actions at arrest. We made the null hypothesis and the alternative hypothesis as follows.

H0: All of the population means of the number of actions at arrest are equal among different age groups.

Ha: There exists at least one pair of age groups, whose population means of the number of actions at arrest is not equal.

The results are shown as follows (Table 4.3). The F-statistics for the age is 6.674, with a p-value of 4.518947e-07 (smaller than 0.05). Therefore, we reject the null hypothesis and reached the conclusion that there exists at least one pair of age groups, whose population means of the number of actions at arrest is not equal.

Table 4.3. The result for one-way anova for the mean of number of actions at arrest among different age groups

	sum_sq	df	F	PR(>F)
C(Age_group__at_arrest_)	17.523117	6.0	6.674609	4.518947e-07
Residual	10507.922058	24015.0	NaN	NaN

To further identify the pairs of significantly different treatments, we performed multiple pairwise comparison (post hoc comparison) analysis using Tukey's honestly significant difference (HSD) test.

We found 7 pairs of treatments with p-values smaller than 0.05, indicating that these 7 pairs of racial groups are significantly different in the population means of the numbers of actions at arrest, among which suspects aged 35-44 and aged 18-24 tend to have the most significant difference. This is reasonable since 18-24 is when people first become adults. The 7 pairs of racial groups are shown below. (Table 4.4)

Table 4.4. The statistically significant tukey tests results for the mean of number of actions at arrest among all different age groups paisrs

	group1	group2	Diff	Lower	Upper	q-value	p-value
3	Aged 45 to 54 years	Aged 18 to 24 years	0.059035	0.014072	0.103998	5.474951	0.002091
6	Aged 25 to 34 years	Aged 35 to 44 years	0.037628	0.001740	0.073517	4.371986	0.032733
9	Aged 25 to 34 years	Aged 65 years and older	0.078168	0.006060	0.150276	4.520315	0.023566
10	Aged 25 to 34 years	Aged 17 years and younger	0.060049	0.002741	0.117358	4.369264	0.032927
12	Aged 35 to 44 years	Aged 18 to 24 years	0.058390	0.017905	0.098876	6.013952	0.001000
18	Aged 18 to 24 years	Aged 65 years and older	0.098930	0.024428	0.173433	5.537061	0.001753
19	Aged 18 to 24 years	Aged 17 years and younger	0.080812	0.020517	0.141106	5.588798	0.001512

Two-Way ANOVA: Number of Actions at Arrest ~ Racial Group + Age Group

We conducted a Two-Way ANOVA (an interaction model) to explore how the two independent variables (race and age) interact with each other in affecting the outcome (number of actions at arrest). Since we have already conducted two One-Way ANOVAs, this Two-Way ANOVA focuses more on the interaction of the effects of the independent variables. We hereby made the null hypothesis and the alternative hypothesis as follows.

H0: The effect of race on population means of the number of actions at arrest does not depend on the effect of age (in other words, there is no interaction effect).
Ha: There is an interaction effect between race and age on the number of actions at arrest.

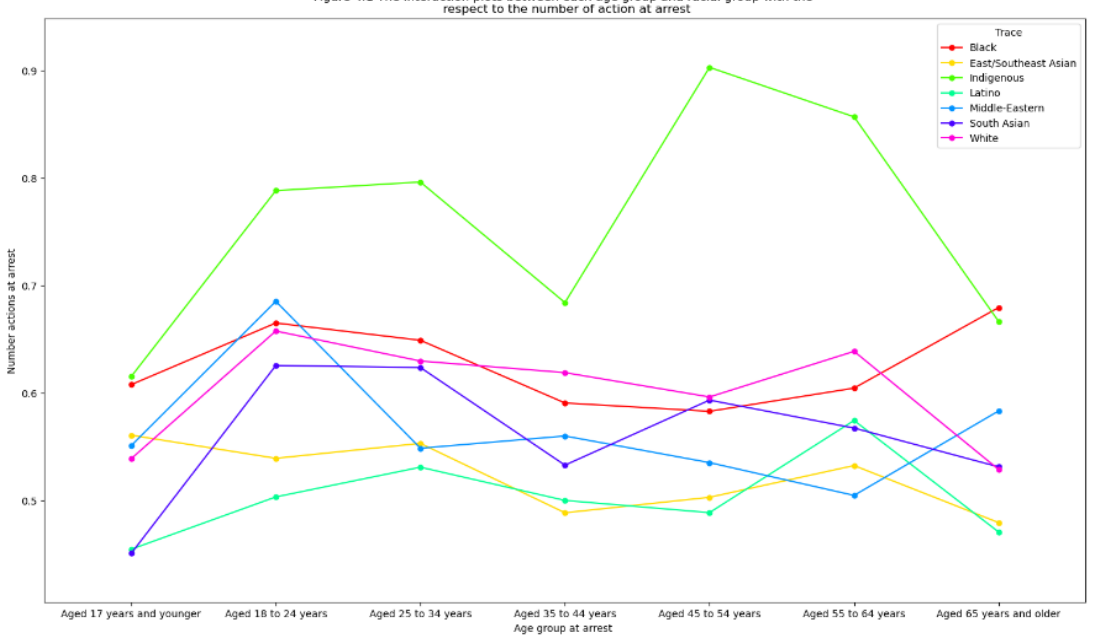
The results are shown as follows (Table 4.5). The F-statistics is 0.892, with a p-value of 6.539827e-01 (greater than 0.05), indicating that the interaction between the two explanatory variables (race and age) is not significant. Therefore, we failed to reject the null hypothesis and reached the conclusion that overall, there is no interaction effect between race and age on the number of actions at arrest.

Table 4.5. The result for two-way anova for the mean of number of actions at arrest among different age groups and different racial groups

	df	sum_sq	mean_sq	F	PR(>F)
C(Age_group__at_arrest_)	6.0	18.381704	3.063617	7.027308	1.732971e-07
C(Perceived_Race)	6.0	42.679759	7.113293	16.316432	7.544843e-19
C(Perceived_Race):C(Age_group__at_arrest_)	36.0	14.000102	0.388892	0.892038	6.539827e-01
Residual	23973.0	10451.242197	0.435959	NaN	NaN

We also created Figure 4.1 to provide a more holistic view of interaction effects. However, we found that some pairs of (race, age) variables have interactions with each other's effect on the number of actions at arrest. For example, white suspects aged 65 and older experience a sharp decrease in the number of actions taken during arrest compared to those aged 55 to 64 years old. In contrast, Black suspects in the same age range see a dramatic increase in actions taken (from the age group of 55-64 years to the age group of 65 years and older). In addition, black suspects who are aged 65 years and older have the highest mean of the number of actions at arrest.

Figure 4.1 The interaction plots between each age group and racial group with the respect to the number of action at arrest



To further identify the pairs of significantly different treatments, we performed multiple pairwise comparison (post hoc comparison) analysis using Tukey's honestly significant difference (HSD) test. We found 21 pairs of treatments with p-values smaller than 0.05, indicating that these 21 pairs of (racial, age) groups are significantly different in the population means of the numbers of actions at arrest, which includes Black suspects aged 25-34 and East/Southeast Asian suspects aged 35-44 years. This may be caused by the significant difference between the racial groups. The 21 pairs of (racial, age) groups are shown below (Table 4.6).

Table 4.6. The statistically significant tukey tests results for the mean of number of actions at arrest among all different age groups and racial groups paisrs

	group1	group2	Diff	Lower	Upper	q-value	p-value
18	(Indigenous, Aged 45 to 54 years)	(White, Aged 65 years and older)	0.374441	0.019037	0.729845	5.934865	0.022554
19	(Indigenous, Aged 45 to 54 years)	(White, Aged 17 years and younger)	0.364347	0.012109	0.716585	5.826775	0.030310
22	(Indigenous, Aged 45 to 54 years)	(South Asian, Aged 35 to 44 years)	0.370379	0.012059	0.728699	5.822706	0.030645
34	(Indigenous, Aged 45 to 54 years)	(East/Southeast Asian, Aged 45 to 54 years)	0.400302	0.037275	0.763329	6.211522	0.010172
35	(Indigenous, Aged 45 to 54 years)	(East/Southeast Asian, Aged 25 to 34 years)	0.350096	0.001339	0.698854	5.654745	0.047437
36	(Indigenous, Aged 45 to 54 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.414720	0.061429	0.768011	6.612595	0.002908
38	(Indigenous, Aged 45 to 54 years)	(East/Southeast Asian, Aged 18 to 24 years)	0.364010	0.010278	0.717742	5.796804	0.032867
39	(Indigenous, Aged 45 to 54 years)	(East/Southeast Asian, Aged 65 years and older)	0.423887	0.013121	0.834652	5.813066	0.031456
42	(Indigenous, Aged 45 to 54 years)	(Latino, Aged 25 to 34 years)	0.372281	0.006092	0.738470	5.726842	0.039536
43	(Indigenous, Aged 45 to 54 years)	(Latino, Aged 35 to 44 years)	0.403226	0.022676	0.783775	5.968788	0.020540
45	(Indigenous, Aged 45 to 54 years)	(Latino, Aged 18 to 24 years)	0.399958	0.004013	0.795903	5.690216	0.043423
81	(Indigenous, Aged 25 to 34 years)	(East/Southeast Asian, Aged 45 to 54 years)	0.293536	0.008165	0.578908	5.794293	0.033088
83	(Indigenous, Aged 25 to 34 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.307954	0.035076	0.580833	6.357206	0.006547
384	(Black, Aged 25 to 34 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.160609	0.032893	0.288326	7.083923	0.001000
496	(Black, Aged 18 to 24 years)	(East/Southeast Asian, Aged 45 to 54 years)	0.162274	0.005501	0.319048	5.830768	0.029980
498	(Black, Aged 18 to 24 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.176692	0.044013	0.309372	7.501764	0.001000
636	(White, Aged 25 to 34 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.141293	0.015278	0.267308	6.316056	0.007430
668	(White, Aged 35 to 44 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.130498	0.002394	0.258603	5.738398	0.038362
699	(White, Aged 55 to 64 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.150313	0.012154	0.288472	6.128665	0.013002
729	(White, Aged 18 to 24 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.169218	0.031834	0.306602	6.938396	0.001000
1044	(Middle-Eastern, Aged 18 to 24 years)	(East/Southeast Asian, Aged 35 to 44 years)	0.196888	0.016110	0.377665	6.135125	0.012759

Research Question 2

Two-Way ANOVA: Number of Search Reasons ~ Racial Group + Gender Group

To explore the effect of categorical independent variables (race and gender) on the quantitative dependent variable (number of search reasons) and how the two independent variables interact with each other in affecting the outcome. We conducted a Two-Way ANOVA (an interaction model). We made three sets of null hypothesis and alternative hypotheses as follows.

H0_1: All of the population means of the numbers of search reasons are equal among different racial groups.

Ha_1: There exists at least one pair of racial groups, whose population means of the numbers of search reasons is not equal.

H0_2: All of the population means of the numbers of search reasons are equal among different gender groups.
Ha_2: There exists at least one pair of gender groups, whose population means of the numbers of search reasons is not equal.

H0_3: The effect of race on population means of the numbers of search reasons does not depend on the effect of gender (In other words, there is no interaction effect).
Ha_3: There is an interaction effect between race and gender on the number of search reasons.

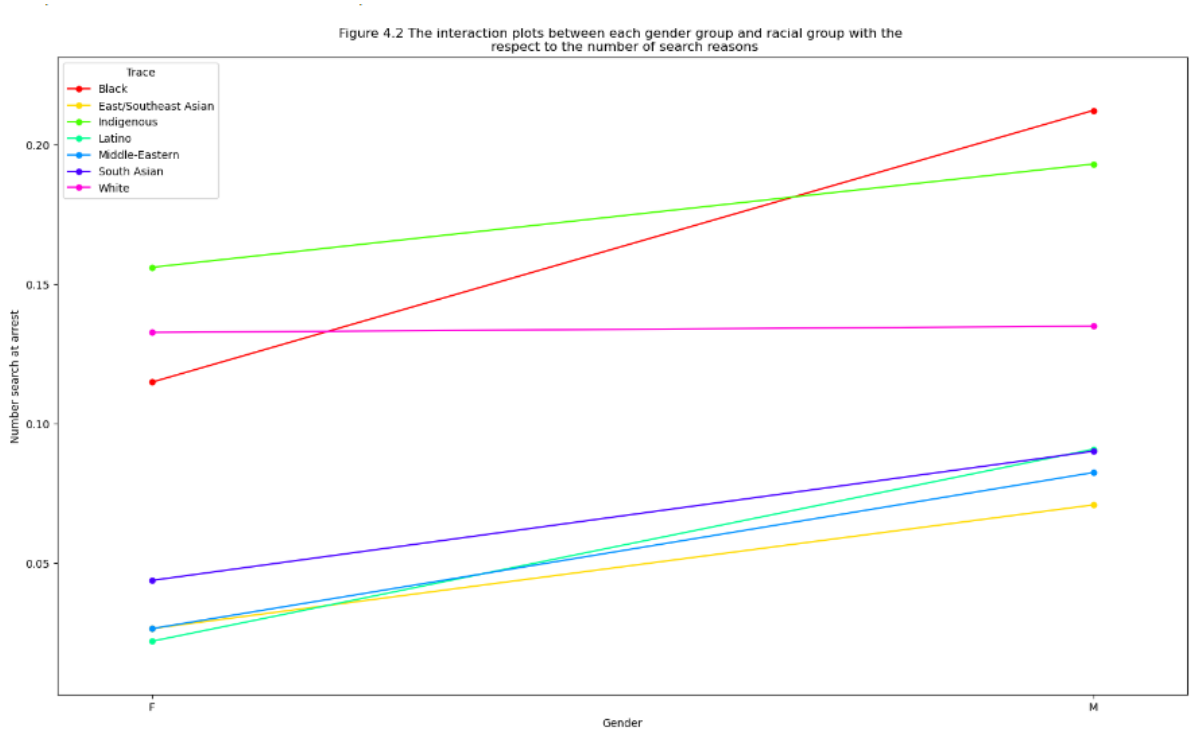
Table 4.7. The result for two-way anova for the mean of number of search reasons among different gender groups and different racial groups

	df	sum_sq	mean_sq	F	PR(>F)
C(Sex)	1.0	6.769257	6.769257	18.965286	1.336695e-05
C(Perceived_Race)	6.0	51.588292	8.598049	24.088972	1.357479e-28
C(Perceived_Race):C(Sex)	6.0	6.512685	1.085448	3.041076	5.657223e-03
Residual	24008.0	8569.147437	0.356929	NaN	NaN

The results are shown in Table 4.7. The F-statistics for race is 24.089, with a p-value of 1.357479e-28 (smaller than the significant level of 0.05). The F-statistics for gender is 18.965, with a p-value of 1.336695e-05 (smaller than the significant level of 0.05). The F-statistics for interaction is 3.041, with a p-value of 5.657223e-03 (smaller than 0.05). Therefore, we reject all three null hypotheses and reach the conclusions that

1. There exists at least one pair of racial groups, whose population means of the numbers of search reasons is not equal.
2. There exists at least one pair of gender groups, whose population means of the numbers of search reasons is not equal.
3. There is an interaction effect between race and gender on the number of search reasons.

To visualize the interaction effect, we created an interaction plot as follows (Figure 4.2). From the interaction plot, we noted that for the white suspects, the numbers of search reasons are almost the same between males and females. However, regarding the black suspects, the number of search reasons for male suspects is far higher than that of female suspects.



To further identify the pairs of significantly different treatments, we performed multiple pairwise comparison (post hoc comparison) analysis using Tukey's honestly significant difference (HSD) test. For the main effect of race on the mean numbers of search reasons, we found 9 pairs of treatments with p-values smaller than 0.05, indicating that these 9 pairs of racial groups are significantly different in the population means of the numbers of reasons. The 9 pairs of racial groups are shown below. (Table 4.8)

Table 4.8. The statistically significant tukey tests results for the mean of number of search reasons among all different racial groups paisrs

	group1	group2	Diff	Lower	Upper	q-value	p-value
4	Indigenous	East/Southeast Asian	0.116335	0.014953	0.217717	4.784867	0.012704
6	Black	White	0.057854	0.030134	0.085573	8.702806	0.001000
7	Black	South Asian	0.110230	0.065105	0.155356	10.185897	0.001000
8	Black	Middle-Eastern	0.117738	0.068766	0.166709	10.025245	0.001000
9	Black	East/Southeast Asian	0.130698	0.089616	0.171780	13.265991	0.001000
10	Black	Latino	0.112487	0.048501	0.176473	7.330618	0.001000
11	White	South Asian	0.052377	0.008879	0.095874	5.021057	0.007061
12	White	Middle-Eastern	0.059884	0.012409	0.107360	5.259755	0.003775
13	White	East/Southeast Asian	0.072844	0.033558	0.112131	7.731659	0.001000

For the main effect of gender on the mean numbers of search reasons, we already conducted the t-test, showing that the difference in the population means between male and female suspects is significant.

For the interaction effects, we found 16 pairs of (gender, race) groups are significantly different. The 13 pairs of racial groups are shown below. (Table 4.9)

Table 4.9. The statistically significant tukey tests results for the mean of number of search reasons among all different racial groups and different gender paisrs

	group1	group2	Diff	Lower	Upper	q-value	p-value
10	(Indigenous, M)	(East/Southeast Asian, F)	0.166604	0.000866	0.332342	4.768208	0.047360
25	(Black, M)	(Black, F)	0.097465	0.037399	0.157531	7.696769	0.001000
26	(Black, M)	(White, M)	0.077380	0.041547	0.113213	10.243177	0.001000
27	(Black, M)	(White, F)	0.079632	0.031220	0.128043	7.802418	0.001000
28	(Black, M)	(South Asian, M)	0.122233	0.065479	0.178987	10.216049	0.001000
29	(Black, M)	(South Asian, F)	0.168569	0.056986	0.280152	7.165918	0.001000
30	(Black, M)	(Middle-Eastern, M)	0.129879	0.069276	0.190482	10.165691	0.001000
31	(Black, M)	(Middle-Eastern, F)	0.185869	0.050101	0.321638	6.493810	0.001000
32	(Black, M)	(East/Southeast Asian, M)	0.141482	0.088940	0.194024	12.772771	0.001000
33	(Black, M)	(East/Southeast Asian, F)	0.185836	0.094542	0.277129	9.655626	0.001000
34	(Black, M)	(Latino, M)	0.121519	0.041819	0.201218	7.232320	0.001000
35	(Black, M)	(Latino, F)	0.190403	0.017043	0.363763	5.209739	0.016475
51	(White, M)	(East/Southeast Asian, M)	0.064102	0.013536	0.114668	6.013145	0.001738
52	(White, M)	(East/Southeast Asian, F)	0.108456	0.018285	0.198626	5.705280	0.004313
59	(White, F)	(East/Southeast Asian, M)	0.061851	0.001713	0.121989	4.878501	0.036894
60	(White, F)	(East/Southeast Asian, F)	0.106204	0.010338	0.202071	5.254914	0.014676

Assumption Checks for ANOVAs

The assumptions for ANOVAs are 1. Errors are approximately normally distributed; 2. Errors should be independent; 3. Variances are equal between treatment groups; 4. Observations are sampled independently from each other; 5. The dependent variable should be continuous. We have already checked the assumption of the normality of outcomes in the EDA, and since we have deleted all the duplicated suspects, the independence of errors is met. Moreover, we have tested the assumption of normality of errors using QQ-Plots and the assumption of equal variance using Levene's tests. Since we followed a similar procedure in checking the assumptions, we only include how we check the assumptions for Two-Way ANOVA: Number of Actions at Arrest ~ Racial Group + Age Group as an example. The QQ-Plot (Figure 4.3) shows that the errors are not normally distributed. In Levene's test, we made a null hypothesis that samples from populations have equal variances. The table below (Table 4.10) shows that the p-value is 0.0483, which is smaller than the significant level. Therefore, we reject the null hypothesis and conclude that the assumption of equal variance is not met.

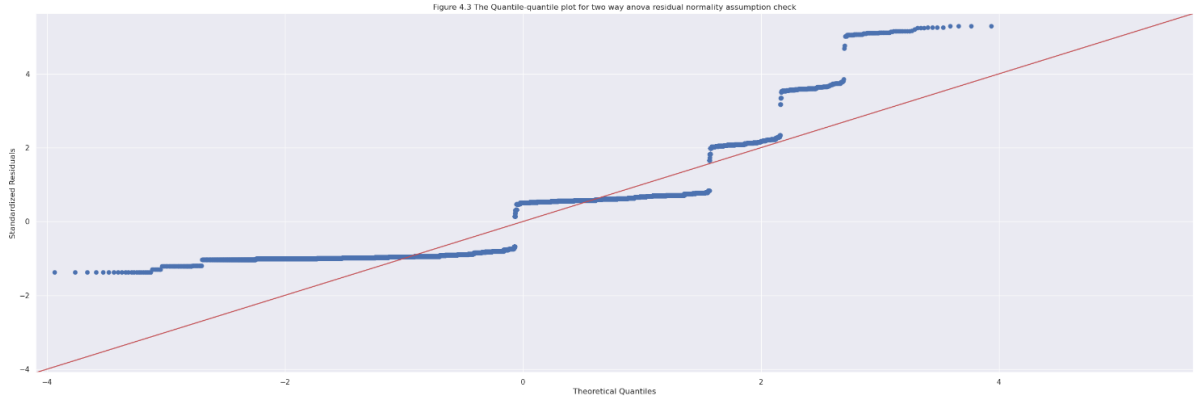


Table 4.10. the test result for Levene's test to check the assumption of homogeneity of variances

	Parameter	Value
0	Test statistics (W)	1.3623
1	Degrees of freedom (Df)	48.0000
2	p value	0.0483

To conclude, the results of the assumption checks for ANOVAs are as follows:

1. All four ANOVAs do not meet the assumption of the normality of errors;
2. All four ANOVAs meet the assumption of the independence of errors
3. All four ANOVAs do not meet the assumption of equal variance;
4. All four ANOVAs do not meet the assumption of the normality of outcomes;
5. All four ANOVAs meet the assumption of the continuity of dependent variables.

Part 5: Discussion and Conclusion

Conclusion

After conducting our research, we have arrived at the conclusion that demographic factors such as gender, age, and race significantly impact suspects' behavior during arrests and police judgments regarding the necessity of strip-search. Our results reveal that gender and race have a noticeable effect on the number of negative actions taken during arrests, although we did find a few exceptions, we did not observe an overall interactional effect between these factors. Furthermore, our investigation highlights that gender and race play a role in the number of search reasons given by police, with Black suspects having a higher population mean number of search reasons compared to White suspects. This discovery indicates that potential bias and unfairness towards different racial groups may exist during police encounters, consistent with previous literature. Additionally, we noted a strong interaction between gender and race in their influence on the mean number of search reasons. In light of these findings, our research emphasizes the importance of recognizing the influence of demographic factors in police encounters and highlights the need for future exploration into possible biases and discrimination.

Limitation

Our research exhibits several limitations that can be attributed to its variable design and the dataset used. Firstly, the unbalanced nature of the dataset is evident, with a disproportionate representation of the white racial group. The sample distribution among different racial groups is uneven, with a significantly lower number of minority subjects compared to white suspects. Although this may be an unavoidable issue given the current population structure in Toronto, it raises concerns about the generalizability of the findings. Additionally, the two continuous outcome variables created are constrained with values ranging from 0 to 4, which restricts the range of possible outcomes and does not follow a normal distribution. Secondly,

some of the assumptions underlying the t-tests and ANOVAs utilized in the analysis are not met, including the assumptions of normality and equal variance of quantitative outcomes. Such violations can limit the robustness of our models and negatively impact our results. To mitigate this issue and enhance the quality of our models, further studies may consider re-creating the outcome variable using more advanced statistical techniques. Finally, the outcome variables have limitations in explaining suspects' performance during arrest and police judgment regarding the necessity of strip-search. For example, the quantification of negative actions during an arrest may fail to capture the differences in severity among different actions. Consequently, further studies may adopt logistic regression techniques to explore the effect of race, gender, and age on specific actions during an arrest, which can provide a better understanding of how suspects behave based on their demographic information during arrests. Despite these limitations, this study contributes to highlighting the importance of demographic factors in police encounters and suggests the need for future studies to investigate potential biases and discrimination.

Appendix

The data set was provided by Toronto Police Service and was available in a CSV file on Public Safety Data Portal, and it can be retrieved through this link below: https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about.

Reference

Bradford, B., Jackson, J., & Stanko, E. (2009). Contact and confidence: Revisiting the impact of public encounters with the police. Policing & Society, 19(1), 20–46.

Gabbidon, S., & Higgins, G. (2009). The role of race/ethnicity and race relations on public opinion related to the treatment of Blacks by the police. Police Quarterly, 12(1), 102–115

Pollock, W., & Menard, S. (2015). Perceptions of Unfairness in Police Questioning and Arrest Incidents: Race, Gender, Age, and Actual Guilt. *Journal of Ethnicity in Criminal Justice*, *13*(3), 237–253. https://doi.org/10.1080/15377938.2015.1015197

Wikimedia Foundation. (2023, February 16). *Murder of George Floyd*. Wikipedia. Retrieved February 27, 2023, from https://en.wikipedia.org/wiki/Murder_of_George_Floyd