Effect of Multiple Factors on the Level of Cooperation at the Time of Arrest

Laien Zhou
Yuelin Liu

A midterm paper submitted in conformity with the requirements
for the course of INF2178

Faculty of Information

University of Toronto

Table of Contents

1 Introduction

1.1 Background

Arrest and strip search procedures are important components of the nowadays criminal justice system, designed to maintain public safety and uphold the law. However, studies have shown that these procedures can disproportionately affect certain groups, particularly those who belong to racial minorities (Engel 2003). In Canada, racial profiling has been identified as a pervasive issue in policing, with racialized individuals more likely to be arrested by police officers than non-racialized individuals (O'Neil 2020).

The Toronto Police Service has recognised the need to address systematic prejudice within its ranks and has instituted a number of programmes targeted at decreasing racial disparities in enforcement (O'Neil 2020). However, we are not sure how effective the efforts have been in reducing the impact of race on the arrest and strip search process. In addition to race, there are other variables such as age, gender, and location(in Toronto or not) at the time of arrest that may also affect the level of cooperation of individuals.

To address these issues, this study will focus on examining the relationship between various factors and the level of cooperation at the time of the arrest. By analyzing data from Toronto Police Service, we aim to provide new insights into the factors that influence the arrest and strip search process, and to help develop better strategies to reduce incidents of force during arrests, which would ultimately lead to a safer and more effective community policing approach.

1.2 Literature Review

Kaminski and Sorensen (1995) discovered that police officers who encounter non-White suspects are more likely to be attacked. Engel (2003) also found that non-White suspects were more likely to resist arrest when confronted by White officers. Walker (1999) explained this phenomenon by stating, “To the extent that officers stereotype young African-American males as potential suspects, they may provoke higher rates of antagonistic behavior that, in turn, results in higher rates of arrest” (pp. 226-227).

In addition to race, gender, and age were also examined as possible factors of the level of cooperation during an arrest. According to the study, men have a higher rate of criminal behavior compared to women and that offending decreases with age (U.S. Department of Justice 2009). Also, Bierie (2015) discovered that female offenders were less likely to assault police personnel than male offenders.

Besides, two studies have discovered evidence to support the expected age group impacts. According to McCluskey et al. (1999), suspects who looked to be under the age of 21 were more likely to fight capture. Bierie (2015) also discovered that confrontations between police and criminals involving at least one juvenile were more likely to result in an officer being assaulted.

Furthermore, Perreault (2019) reported that police services in rural areas served 16% of the population in Canadian provinces in 2017 but reported a higher percentage of violent crimes, property crimes, Criminal Code traffic offenses, and other Criminal Code violations compared to urban areas.

1.3 Research Questions/Objective

The main objective of this study is to investigate the factors that influence the level of cooperation between individuals and police officers at the time of the arrest. Specifically, we aim to answer the following 2 research questions:

1. What is the effect of ***perceived race*** and ***sex*** on the level of cooperation at the time of arrest?
2. How does ***location in Toronto or not*** at arrest and ***age group*** affect the level of cooperation at the time of the arrest?

2 Exploratory data analysis

2.1 Overview

The dataset used in this study is the "Arrests and Strip Searches" dataset (RBDC-ARR-TBL-001) from the Toronto Police Service Public Safety Data Portal. It provides relevant information about arrests and strip searches and is conducted by police officers within Toronto. A strip search is defined as a search that involves the removal of clothing and a visual examination of the body. The dataset includes 65,276 records, with 24 attributes including the year and month of arrest, the ID of arrest, the race, gender, age group of the suspects, the arrest location, and the actions taken during an arrest, etc. The location of the arrest is recorded at the Division level and refers to the area within the Division boundaries where the arrest took place. The age is the individual’s age at the time of the arrest, as reported by the arresting officer.

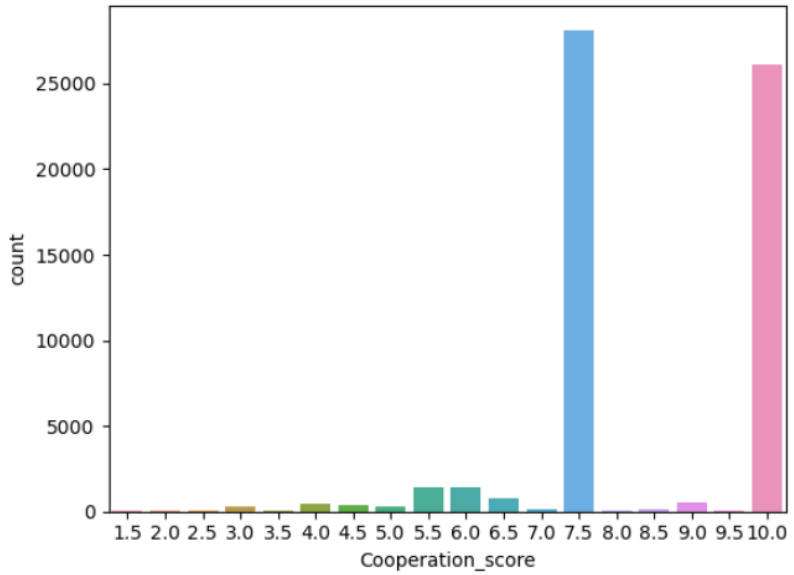
2.2 Descriptive statistics

For the problem to be studied in this paper, the dataset underwent the following data cleaning process. To begin with, the samples with a "U"(Uncertain) attribute under the *"Sex"* category were removed as

their count was relatively small in comparison to other categories and would not be statistically significant. Additionally, the samples with an "Unknown or Legacy" attribute under the *"Perceived_Race"* category were removed. The category *'ArrestLocDiv'* has been renamed to *'intoronto'*, while we set its value to "1" for rows where it is not equal to "XX" and to "0" for rows where it is equal to "XX". This is done to represent whether an arrest location is in Toronto or not. For the *'Age_group__at_arrest_'* category, we reassigned the original age group values to a smaller set of standardized categories to make data analysis and interpretation easier. After that, we created a new continuous variable *"Cooperation_score"* (defined in the *Method* section). Lastly, all the samples containing "N/A" values were purged from the dataset to ensure accurate statistical calculations.

The figure below shows the distribution of Cooperation Scores (defined in the *Method* section) in the sample. Approximately half of the ratings are concentrated at a score of 7.5, with close to half but fewer concentrated at a score of 10. The remaining ratings have a similar normal distribution structure with centers around 5.5-6 and 9 for scores below 7.5 and between 7.5-10, respectively. These findings indicate that the majority of individuals in the dataset were willing to cooperate, as a score of 7.5 implies that the individuals did not show explicit cooperative behavior but also no obvious resistance at the time of arrest, while a score of 10 implies that the individuals cooperate fully with the police at the time of the arrest.

Figure 1: Distribution of the Cooperation Scores



The four box plots below show the distribution of the dependent variable (*Cooperation_Scores*) when grouped by 4 different independent variables, respectively:

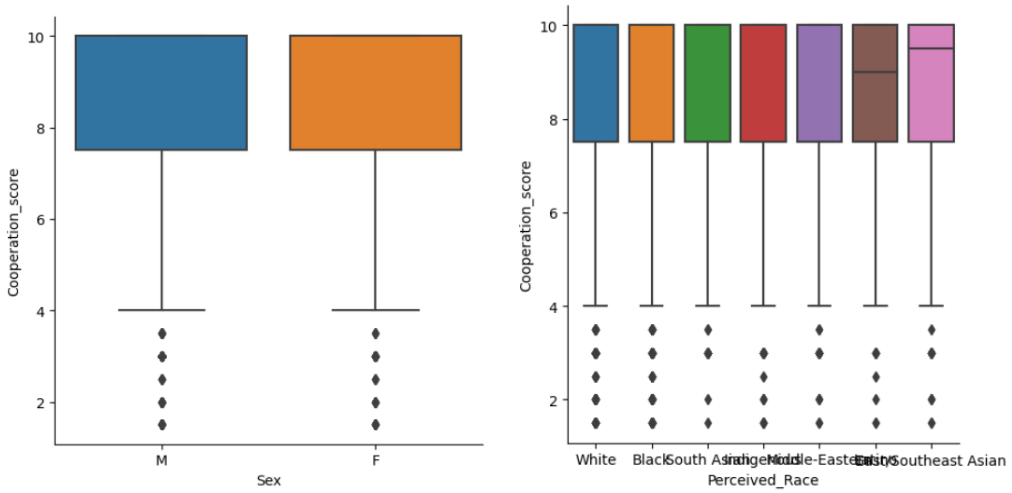
- In the top left plot, the distribution of cooperation scores for individuals of different genders is presented. It can be observed that the boxes for both genders are concentrated in the interval of 7.5 to 10 points, which corresponds to the main concentration of Cooperation scores in Figure 1. The cases with scores below 4 points for both genders are classified as outliers.

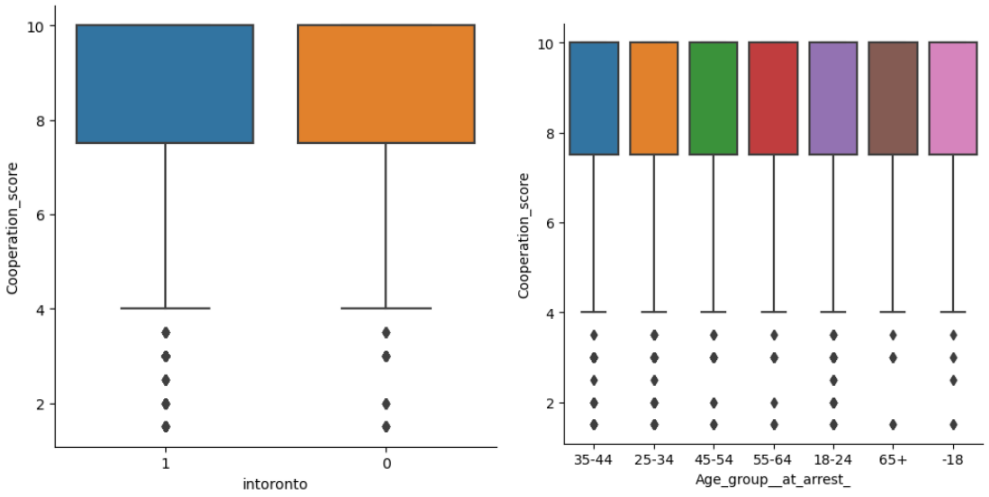
- The box plot in the upper right corner displays the distribution of cooperation scores across various races. The boxes for all races are concentrated in the range of 7.5 to 10 points, with cases below 4 points classified as outliers. However, there are noticeable differences in the mean cooperation scores among different race groups. Latino individuals have a lower mean cooperation score compared to other races, while individuals from East/Southeast Asian backgrounds have a higher mean cooperation score than other races. Moreover, the distribution of outliers for cooperation scores varies among different race groups.

- The box plot in the lower left corner displays the distribution of cooperation scores for individuals who were in the Toronto area at the time of arrest and those who were not. Both boxes are concentrated within the 7.5-10 point range, but the distribution of outliers (under a score of 4) for both groups is different. Outliers for individuals not in the Toronto area are relatively more widely spread than those for individuals in the Toronto area.

- The bottom right box plot displays the distribution of cooperation scores for individuals in different age groups. The boxes for all age groups are mainly distributed in the range of 7.5-10 points, with scores below 4 points being categorized as outliers. However, the distribution of outliers varies among the age groups, with the greatest distance between outliers observed in individuals over 65 years old, and a relatively distant distribution of outliers in those aged 45-64 years old.

Figure 2: Boxplots of Sex, Perceived_Race, Intoronto (location), and Age_group_at_arrest_ with Cooperation_Score





2.3 T-tests

T-test is used to compare the means of two groups for each factor, including the Sex, Perceived_Race, Intoronto (location), and Age_group_at_arrest_:

1. **Sex vs Cooperation_Score**
    - H0 (null hypothesis): There is no difference in cooperation score for male and female.
    - Ha (alternative hypothesis): There is a significant difference in cooperation scores for male and female.

Appendix: Table 1 - Sex vs Cooperation score (code)

Cooperation score			
	Mean	SD	
male	8.4239	1.5548	
female	8.3748	1.5643	
satatistic	3.0514	p-value	0.0023
CI	(0.0176, 0.0807)	DoF	17659.60

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

- Interpretation: The results indicate that the mean cooperation score for male (M=8.4239, SD=1.5548) is higher than the mean cooperation score for female (M=8.3748, SD=1.5643). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0023) is less than 0.05, 95% CI [0.0176, 0.0807].

- Therefore, we can REJECT the null hypothesis that there is no difference in cooperation scores for male and female.

2. In_toronto vs Cooperation_score

- H0 (null hypothesis): There is no difference in cooperation score for individuals arrested inside Toronto and outside Toronto.
- Ha (alternative hypothesis): There is a significant difference in cooperation scores for individuals arrested inside Toronto and outside Toronto.

Appendix: Table 2 - In_toronto vs Cooperation score

Cooperation score			
	Mean	SD	
inside toronto	8.3928	1.6443	
outside toronto	8.4401	1.4453	
satatistic	-3.7616	p-value	0.0002
CI	(-0.0721, -0.0227)	DoF	60067.41

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

- Interpretation: The results indicate that the mean cooperation score for individuals arrested inside Toronto (M=8.3928, SD=1.6443) is lower than the mean cooperation score for individuals arrested inside Toronto (M=8.4401, SD=1.4453). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0002) is less than 0.05, 95% CI [-0.0721, -0.0227].
- Therefore, we can REJECT the null hypothesis that there is no difference in cooperation score for individuals arrested inside Toronto and outside Toronto.

3. Perceived_Race vs Cooperation_score

- H0: There is no difference in cooperation score for white and non-white individuals.
- Ha: There is a significant difference in cooperation scores for white and non-white individuals.

Appendix: Table 3 - Perceived race vs Cooperation score

Cooperation score		
	Mean	SD
white	8.4067	1.5566

not white	8.4210	1.5569	
satatistic	-1.1238	p-value	0.2611
CI	(-0.0393, 0.0106)	DoF	58703.24

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

- Interpretation: The results indicate that the mean cooperation score for white individuals (M=8.4067, SD=1.5566) is lower than the mean cooperation score for non-white individuals (M=8.4210, SD=1.5569). With alpha established at 0.05, this is NOT a statistically significant difference as the p-value (0.2611) is greater than 0.05, 95% CI [-0.0393, 0.0106].
- Therefore, we CANNOT reject the null hypothesis that there is no difference in cooperation score for white and non-white individuals.

4. Age group vs Cooperation_score

- H0 (null hypothesis): There is no difference in cooperation score for individuals younger than 35 and older or equal to 35.
- Ha (alternative hypothesis): There is a significant difference in cooperation score for individuals younger than 35 and older or equal to 35.

Appendix: Table 4 - Age group vs Cooperation score

	Cooperation score		
	Mean	SD	
< 35	8.3852	1.5650	
≥ 35	8.4314	1.5517	
satatistic	-3.4998	p-value	0.0005
CI	(-0.0720, -0.0203)	DoF	45914.73

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

- Interpretation: The results indicate that the mean cooperation score for individuals younger than 35 (M=8.3852, SD=1.5650) is lower than the mean cooperation score for individuals older or equal to 35 (M=8.4314, SD=1.5517). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.0005) is less than 0.05, 95% CI [-0.0720, -0.0203].
- Therefore, we can REJECT the null hypothesis that there is no difference in cooperation score for individuals younger than 35 and older or equal to 35.

3 Method

3.1 Dataset

Table 5. Variables and Value Categories	
Variable	Categories
Independent Variables	
1. Sex	Male 48505 Female 11682
2. Age_group__at_arrest_	-18 2825 18-24 9183 25-34 19299 35-44 15048 45-54 8395 55-64 4225 65+ 1212
3. in Toronto or not (intoronto)	Yes 32694 No 27493
4. Perceived_Race	White 27708 Black 17518 East/Southeast Asian 4412 South Asian 3613 Middle-Eastern 3237 Indigenous 1932 Latino 1767
Dependent Variable	
Cooperation_score	Domain: [0,10] Total count: 60187

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

3.2 Measurement

3.2.1 Dependent Variable

The ***Cooperation_score*** (cooperation score at arrest) is a composite index that is intended to reflect the level of cooperation that a suspect displayed during the process of being arrested. The index is calculated based on the values assigned to each of the six behaviors at arrest.

Cooperation Score = 7.5 + 2.5 * df["Actions_at_arrest___Cooperative"] - 0.5 * df["Actions_at_arrest___Concealed_i"] - 2 * df["Actions_at_arrest___Combative__"] - 1.5 *

df["Actions_at_arrest___Resisted__d"] - 1 * df["Actions_at_arrest___Mental_inst"] - 2.5 * df["Actions_at_arrest___Assaulted_o"]

There are six specific types of actions at arrest:

1. Actions at arrest - Concealed items
2. Actions at arrest - Combative, violent or spitter/biter
3. Actions at arrest - Resisted, defensive or escape risk.
4. Actions at arrest - Mental instability or possibly suicidal
5. Actions at arrest - Assaulted officer
6. Actions at arrest – Cooperative

The cooperation coefficient has a linear positive relationship with the level of cooperation, meaning that the higher the index, the more cooperative the suspect was with the police. Conversely, lower indexes indicate less cooperation.

Each action at arrest is given a value from [-2.5, 2.5], depending on their level of cooperation (+) and level of threat (-). The cooperation score is set to be 0 to 10, with a starting value of 7.5. It is then adjusted based on the values assigned to each of the six behaviors at arrest, with each behavior being paired with a different coefficient based on the classification of the egregiousness and impact of the behavior. For example, if a suspect concealed items, this would result in a decrease in the cooperation score of 0.5. If the inmate was combative, violent or a spitter/biter, this would result in a decrease of 2.0 in the cooperation score. The final cooperation score is the sum of the starting value of 7.5 and the adjustments made based on each of the six behaviors at arrest.

Limitation: It's important to note that this scoring system is subjective and based on the judgment of the person assigning the values to each behavior. The values assigned to each behavior and the coefficients used to adjust the cooperation score may be adjusted based on the specific needs of the organization using the scoring system.

3.2.2 Independent Variables

The independent variables in this study include "Sex", "Age_group_at_arrest_", "in Toronto or not (intoronto)" and "Perceived_Race":

1) ***Sex*** refers to the gender of the individuals in the study, with two categories: Male and Female. The number of males in the study is 48505, while the number of females is 11682.
2) ***Age_group_at_arrest_*** refers to the age range of individuals at the time of their arrest. The age ranges included in the study are -18, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. The number of individuals in each age range are 2825, 9183, 19299, 15048, 8395, 4225, and 1212, respectively.

3) ***intoronto*** refers to whether the individuals were arrested in Toronto or not, with two categories: "Yes" and "No." The number of individuals arrested in Toronto is 32694, while the number of individuals not arrested in Toronto is 27493.
4) ***Perceived_Race*** refers to the race of the individuals in the study. The races included are White, Black, East/Southeast Asian, South Asian, Middle-Eastern, Indigenous, and Latino. The number of individuals in each race are 27708, 17518, 4412, 3613, 3237, 1932, and 1767, respectively.

These independent variables will be analyzed to determine their impact on the dependent variable.

Note: We found that in the t-test, *Perceived_Race* has a p-value greater than 0.05, indicating that we fail to reject the null hypothesis that there is no difference in cooperation score for white and non-white individuals. However, we still decided to include it in the subsequent two-way ANOVA analysis. Our reason is that even if the t-test does not detect a significant difference between the two groups, it may still have an effect when combined with another variable in a two-way ANOVA test. The variability in the data may also be explained by the interaction of the two variables, rather than the individual variables themselves. We will discuss more in the *Discussion* Part.

3.3 Statistical Tests

3.3.1 Two-way ANOVA tests

Two-way ANOVA Tests are used to compare the means of two or more groups across two categorical variables. The test allows us to examine the effect of each variable on the cooperation score while taking into account the influence of other variables. Besides, it may reveal significant effects that were not detected by a t-test.

Hypothesis 1: Perceived_Race and Sex vs Cooperation score

Q: What is the effect of ***perceived race*** and ***sex*** on the level of cooperation at the time of arrest?

H0:

1) There is no difference in the mean cooperation score for any perceived race type.
2) There is no difference in the mean cooperation score for either sex type.
3) There is no interaction effect between perceived race and sex.

Ha:

1) There is a difference in the mean cooperation score by perceived race type.
2) There is a difference in the mean cooperation score by sex type.
3) There is an interaction effect between perceived race and sex.

Table 6 : Result of the Two-way ANOVA Tests of Hypothesis 1

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	595.531029	6.0	41.12739	2.57E-50
C(Sex)	12.18898	1.0	5.050628	2.46E-02
C(Perceived_Race):C(Sex)	24.486327	6.0	1.691026	1.19E-01
Residual	145219.0752	60173.0	NaN	NaN

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

The results indicate the following:

- Since the p-value (2.57e-50) is less than 0.05, we REJECT the null hypothesis that there is no difference in the mean cooperation score for any perceived race type.
- Since the p-value (2.46e-02) is less than 0.05, we REJECT the null hypothesis that there is no difference in the mean cooperation score for either sex type.
- Since the p-value (0.119) is greater than 0.05, we CANNOT reject the null hypothesis that there is no interaction effect between perceived race and sex.

Hypothesis 2: In Toronto or not and Age group at arrest vs Cooperation score

Q: How does ***location in Toronto or not*** at arrest and ***age group*** affect the level of cooperation at the time of arrest?

H0:

1) There is no difference in the mean cooperation score for either arrested in Toronto or not.
2) There is no difference in the mean cooperation score for any age group at arrest type.
3) There is no interaction effect between arrested in Toronto or not and the age group at arrest.

Ha:

1) There is a difference in the mean cooperation score by arrested in Toronto or not.
2) There is a difference in mean cooperation score by age group at arrest type.
3) There is an interaction effect between arrested in Toronto or not and age group at arrest type.

Table 7: Result of the Two-way ANOVA Tests of Hypothesis 2

	sum_sq	df	F	PR(>F)
C(intoronto)	28.067518	1.0	11.59764	6.61E-04

C(Age_group__at_arrest_)	181.964605	6.0	12.53145	3.58E-14
C(intoronto):C(Age_group__at_arrest_)	21.299441	6.0	1.466839	1.85E-01
Residual	145625.028	3.0	NaN	NaN

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

The results indicate the following:

- Since the p-value (2.57e-50) is less than 0.05, we REJECT the null hypothesis that there is no difference in the mean cooperation score for either arrested in Toronto or not.
- Since the p-value (2.46e-02) is less than 0.05, we REJECT the null hypothesis that there is no difference in the mean cooperation score for any age group at arrest type.
- Since the p-value (0.1865) is greater than 0.05, we CANNOT reject the null hypothesis that there is no interaction effect between arrested in Toronto or not and the age group at arrest.

3.3.2 Post-hoc Tests

Based on the above Two-way ANOVA test with interaction, it is found that the mean of each group is not equal for each independent variable. In order to find out which groups differ from each other, the post-hoc test (Tukey's HSD) is performed as follows:

1) Sex vs cooperation score

Table 8 : Result of the Tukey's HSD Test for Sex vs Cooperation Score

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
F	M	0.0491	0.0022	0.0177	0.0806	True

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Interpretation:

- The adjusted p-value for the mean difference between female and male is 0.0022.
- There is a significant difference in mean cooperation score between groups in sex type.

2) Perceived Race vs cooperation score

Table 9 : Result of the Tukey's HSD Test for Perceived Race vs Cooperation Score

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	0.3181	0.001	0.241	0.3953	True
Black	Indigenous	-0.1877	0.001	-0.2975	-0.0779	True
Black	Latino	0.2521	0.001	0.1378	0.3664	True
Black	Middle-Eastern	0.1574	0.001	0.0698	0.2451	True
Black	South Asian	0.1719	0.001	0.0882	0.2556	True
Black	White	0.0663	0.001	0.0221	0.1105	True
East/Southeast Asian	Indigenous	-0.5059	0.001	-0.6308	-0.3809	True
East/Southeast Asian	Latino	-0.066	0.712	-0.195	0.0629	False
East/Southeast Asian	Middle-Eastern	-0.1607	0.001	-0.2667	-0.0547	True
East/Southeast Asian	South Asian	-0.1462	0.001	-0.249	-0.0435	True
East/Southeast Asian	White	-0.2519	0.001	-0.3261	-0.1776	True
Indigenous	Latino	0.4398	0.001	0.2891	0.5906	True
Indigenous	Middle-Eastern	0.3451	0.001	0.2135	0.4768	True
Indigenous	South Asian	0.3596	0.001	0.2305	0.4887	True
Indigenous	White	0.254	0.001	0.1462	0.3618	True
Latino	Middle-Eastern	-0.0947	0.3764	-0.2302	0.0408	False
Latino	South Asian	-0.0802	0.5627	-0.2132	0.0528	False
Latino	White	-0.1858	0.001	-0.2982	-0.0734	True
Middle-Eastern	South Asian	0.0145	0.9997	-0.0964	0.1253	False
Middle-Eastern	White	-0.0911	0.0265	-0.1762	-0.0061	True
South Asian	White	-0.1056	0.0023	-0.1867	-0.0246	True

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Interpretation (some):

- The adjusted p-value for the mean difference between individuals with perceived race as **Black** and those with perceived race as **White** is 0.001, we can conclude that there is a significant difference in mean cooperation score between individuals with perceived race as black and those with perceived race as White.
- The adjusted p-value for the mean difference between individuals with perceived race as **East/Southeast Asian** and those with perceived race as **Latino** is 0.712, we can conclude that there is no significant difference in mean cooperation score between individuals with perceived race as East/Southeast Asian and those with perceived race as Latino.
- The adjusted p-value for the mean difference between individuals with perceived race as **Indigenous** and those with perceived race as **Middle-Eastern** is 0.001, we can conclude that there is a significant difference in mean cooperation score between Indigenous individuals and those with perceived race as Middle-Eastern.
- The adjusted p-value for the mean difference between individuals with perceived race as **Latino** and those with perceived race as **South Asian** is 0.5537, we can conclude that there is no significant difference in mean cooperation score between individuals with perceived race as Latino and those with perceived race as South Asian.
- The adjusted p-value for the mean difference between individuals with perceived race as **South Asian** and those with perceived race as **White** is 0.0023, we can conclude that there is a significant difference in mean cooperation score between South Asian individuals and those with perceived race as White.

Conclusion:

- There is a significant difference in mean cooperation score between at least two groups in perceived race type.

3) Arrested in Toronto or not vs cooperation score

Table 10 : Result of the Tukey's HSD Test for Intoronto vs Cooperation Score

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-0.0474	0.001	-0.0724	-0.0224	True

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Interpretation:

- The adjusted p-value for the mean difference between individuals arrested in Toronto and those arrested not in Toronto is 0.001.

- There is a significant difference in mean cooperation score between groups arrested in Toronto or not.

4) Age group at arrest vs cooperation score

Table 11 : Result of the Tukey's HSD Test for Age Group vs Cooperation Score

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
-18	18-24	-0.1427	0.0004	-0.2414	-0.044	True
-18	25-34	-0.1888	0.001	-0.2812	-0.0964	True
-18	35-44	-0.1513	0.001	-0.2454	-0.0573	True
-18	45-54	-0.0749	0.2877	-0.1747	0.0249	False
-18	55-64	-0.0771	0.3894	-0.1886	0.0344	False
-18	65+	0.0252	0.9992	-0.1323	0.1827	False
18-24	25-34	-0.0461	0.2259	-0.1043	0.012	False
18-24	35-44	-0.0086	0.9996	-0.0694	0.0521	False
18-24	45-54	0.0678	0.06	-0.0015	0.137	False
18-24	55-64	0.0656	0.2603	-0.0197	0.1508	False
18-24	65+	0.1679	0.0076	0.0277	0.3081	True
25-34	35-44	0.0375	0.287	-0.0124	0.0874	False
25-34	45-54	0.1139	0.001	0.0539	0.1739	True
25-34	55-64	0.1117	0.0005	0.0338	0.1896	True
25-34	65+	0.214	0.0001	0.0781	0.3498	True
35-44	45-54	0.0764	0.0058	0.0139	0.1389	True
35-44	55-64	0.0742	0.0888	-0.0057	0.1541	False
35-44	65+	0.1765	0.0028	0.0395	0.3135	True
45-54	55-64	-0.0022	0.9	-0.0887	0.0843	False
45-54	65+	0.1001	0.3563	-0.0409	0.2411	False
55-64	65+	0.1023	0.4031	-0.0472	0.2518	False

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

Interpretation (some):

- The adjusted p-value for the mean difference between individuals **less than 18** and those **in 18-24** is 0.001, we can conclude that there is a significant difference in mean cooperation score between individuals less than 18 and those greater than 18 and lower than 24.
- The adjusted p-value for the mean difference between individuals in **18-24** and those **greater than 65** is 0.0076, we can conclude that there is a significant difference in mean cooperation score between individuals in 18-24 and those greater than 65.
- The adjusted p-value for the mean difference between individuals in **25-34** and those in **35-44** is 0.2871, we can conclude that there is no significant difference in mean cooperation score between individuals in 25-34 and those in 35-44.
- The adjusted p-value for the mean difference between individuals in **35-44** and those in **65+** is 0.0028, we can conclude that there is a significant difference in the mean cooperation score between individuals in 35-44 and those greater than 65.
- The adjusted p-value for the mean difference between individuals in **55-64** and those **less than 18** is 0.3911, we can conclude that there is no significant difference in mean cooperation score between individuals in 55-64 and those less than 18.

Conclusion:

- There is a significant difference in mean cooperation score between at least two groups in age group at arrest type.

4 Results/ Findings

4.1 Interaction plots

After performing the ANOVA tests to assess whether there are significant differences in mean cooperation scores across different groups, interaction plots are used to help further interpret the results by visually displaying the relationship between variables. Specifically, the interaction plots display the levels of an independent variable on the X-axis and have a separate line for the means of each level of the other variable. The Y-axis is the dependent variable cooperation score. The interaction plots indicate how the effect of one variable differs for groups in the other variable.

1) **Interaction plots (Perceived Race / Sex vs Cooperation score)**

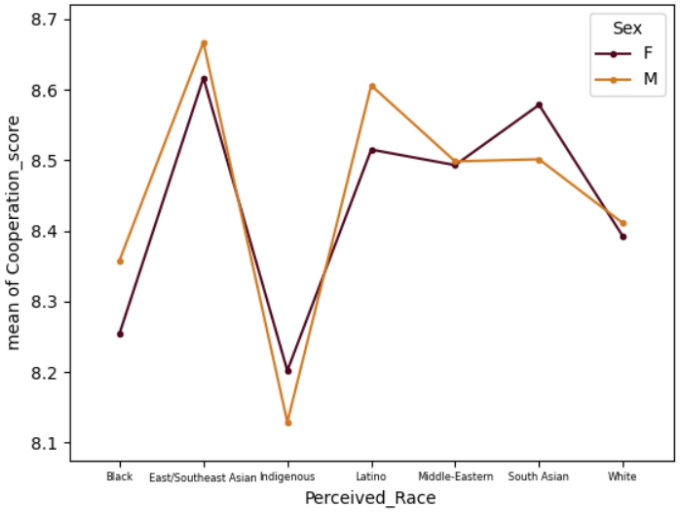


Figure 3: Interaction plot for Perceived Race and Sex with the Cooperation Score

While the interaction plot does not provide any information on statistically significant difference, the plot showed:

- For individuals in the age group with perceived race as **Black**, **East/Southeast Asian**, **Latino**, and **White**, 'Male' individuals tend to have a higher cooperation score compared to those in the sex type as 'Female'; for individuals with race as **Indigenous** and **South Asian**, individuals in the sex type as 'Male' tend to have a lower cooperation score compared to those in the sex type as 'Female'; for individuals with perceived race as **Middle-Eastern**, the mean cooperation score between 'Male' and 'Female' is very close to each other;

There are noticeable differences in mean cooperation scores between individuals with perceived race as **Latino, Indigenous** and **South Asian** of different sex types.

Additionally, we found that **Black** individuals and **Indigenous** individuals(the lowest score) have lower mean cooperation scores compared to other race groups, regardless of gender. In contrast, East/Southeast Asians had the highest mean cooperation scores.

2) Interaction plots (Age group at arrest / Arrested in Toronto or not vs Cooperation score)

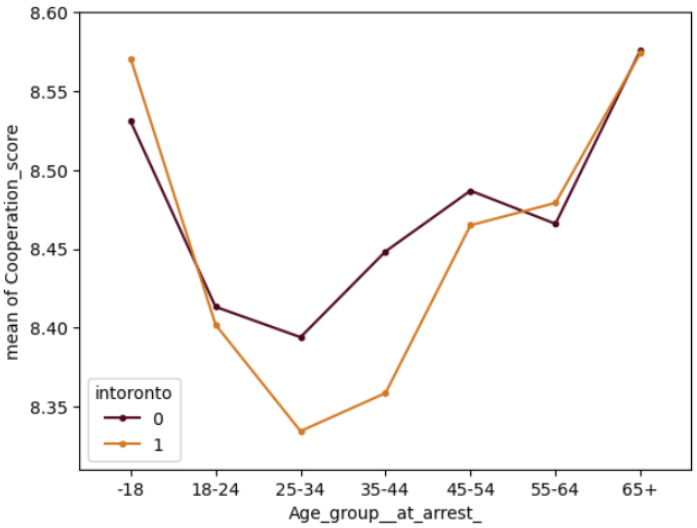


Figure 4: Interaction plot for Age group and in Toronto or not with the Cooperation Score

While the interaction plot does not provide any information on statistically significant difference, the plot showed:

- For individuals in the age group **lower than 18** and **between 55 to 64**, individuals arrested inside Toronto tend to have a **higher** cooperation score compared to those arrested outside Toronto; for individuals in the age group **greater than 18** and **lower than 55**, individuals arrested inside Toronto tends to have a **lower** cooperation score compared to those arrested outside Toronto; for individuals in the age group that are **greater than 65**, individuals from both arrested in Toronto group and not arrested in Toronto group tend to have approximately the same cooperation score.

There are noticeable differences in mean cooperation scores based on whether the individuals were arrested in Toronto or not, within the age groups of **25-34**, **35-44**, and **45-54**. Additionally, we found that individuals in the **25-34** and **35-44** age groups had lower mean cooperation scores compared to other age groups, especially when they were arrested in Toronto. In contrast, individuals that are **younger than 18** and **older than 65** (highest) had higher mean cooperation scores, regardless of the location of the arrest.

4.2 Interpretation of the model

4.2.1 Model of Hypothesis 1

Based on the Two-way ANOVA test in the previous session, the final model formula of hypothesis 1 is fixed as follows:

Cooperation score ~ Perceived Race + Sex

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.3105	0.018	467.822	0.000	8.276	8.345
C(Perceived_Race)[T.East/Southeast Asian]	0.3180	0.026	12.153	0.000	0.267	0.369
C(Perceived_Race)[T.Indigenous]	-0.1822	0.037	-4.881	0.000	-0.255	-0.109
C(Perceived_Race)[T.Latino]	0.2513	0.039	6.480	0.000	0.175	0.327
C(Perceived_Race)[T.Middle-Eastern]	0.1553	0.030	5.221	0.000	0.097	0.214
C(Perceived_Race)[T.South Asian]	0.1708	0.028	6.015	0.000	0.115	0.226
C(Perceived_Race)[T.White]	0.0681	0.015	4.534	0.000	0.039	0.098
C(Sex)[T.M]	0.0362	0.016	2.247	0.025	0.005	0.068

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

We can interpret the parameters as follows (partially):

- **Intercept**: The cooperation score is 8.31 for the individuals with perceived race as Black and in the sex type as female.
- The coefficient for whether Pervceived_Race.**East/Southeast_Asian**: Compared with individuals with perceived race as Black, the cooperation score among those participants with perceived race as East/Southeast Asian is 0.32 higher, holding other variables constant. Since the p-value is less than 0.05, the result is statistically significant.
- The coefficient for whether Pervceived_Race.**Indigenous**: Compared with individuals with perceived race as Black, the cooperation score among those participants with perceived race as Indigenous is 0.18 lower, holding other variables constant. Since the p-value is less than 0.05, the result is statistically significant.
- The coefficient for whether Pervceived_Race.**White**: Compared with individuals with perceived race as Black, the cooperation score among those participants with perceived race as White is 0.07 higher, holding other variables constant. Since the p-value is less than 0.05, the result is statistically significant.
- The coefficient for Sex.M: Compared with individuals who are **female**, the cooperation score among those participants who are **male** is 0.04 higher, holding other variables constant. Since the p-value is greater than 0.05, the result is NOT statistically significant.

4.2.1 Model of hypothesis 2

Based on the Two-way ANOVA test in the previous session, the final model formula of hypothesis 2 is fixed as follows:

Cooperation score ~ Intoronto + Age_group_at_arrest

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.5706	0.030	286.679	0.000	8.512	8.629
C(intoronto)[T.1]	-0.0434	0.013	-3.405	0.001	-0.068	-0.018
C(Age_group__at_arrest_)[T.18-24]	-0.1400	0.033	-4.181	0.000	-0.206	-0.074
C(Age_group__at_arrest_)[T.25-34]	-0.1855	0.031	-5.916	0.000	-0.247	-0.124
C(Age_group__at_arrest_)[T.35-44]	-0.1479	0.032	-4.636	0.000	-0.211	-0.085
C(Age_group__at_arrest_)[T.45-54]	-0.0720	0.034	-2.128	0.033	-0.138	-0.006
C(Age_group__at_arrest_)[T.55-64]	-0.0754	0.038	-1.995	0.046	-0.150	-0.001
C(Age_group__at_arrest_)[T.65+]	0.0245	0.053	0.458	0.647	-0.080	0.129

Source: Arrests and Strip Searches (RBDC-ARR-TBL-001)

We can interpret the parameters as follows (partially):

- **Intercept**: The cooperation score is 8.57 for the individuals who are arrested outside Toronto and in the age group of less than 18 at arrest.
- The coefficient for whether being arrested inside Toronto or not **(Intoronto.1**): Compared with individuals who are arrested outside Toronto, the cooperation score among those participants who are arrested inside Toronto is 0.04 lower, holding other variables constant. Since the p-value is less than 0.05, the result is statistically significant.
- The coefficient for Age_group_at_arrest.**18-24**: Compared with individuals who are lower than 18 at arrest, the cooperation score among those participants who are greater than 18 and lower than 24 at arrest is 0.14 lower, holding other variables constant. Since the p-value is less than 0.05, the result is statistically significant.
- The coefficient for Age_group_at_arrest.**35-44**: Compared with individuals who are lower than 18 at arrest, the cooperation score among those participants who are greater than 35 and lower than 44 at arrest is 0.15 lower, holding other variables constant. Since the p-value is less than 0.05, the result is statistically significant.
- The coefficient for Age_group_at_arrest.**65+**: Compared with individuals who are lower than 18 at arrest, the cooperation score among those participants who are greater than 65 at arrest is 0.02 higher, holding other variables constant. Since the p-value is greater than 0.05, the result is NOT statistically significant.

4.3 Results

In the exploratory data analysis part, with the help of descriptive statistics and box plots, we found that the majority of individuals in the dataset were willing to cooperate with the Toronto police, with a concentration of cooperation ratings at 7.5 and 10. The following four T-tests showed that there is a significant difference in cooperation scores for different sexes, age groups and locations (with p-value of 0.0023, 0.0005, and 0.0002, respectively).

For perceived race, we found there is no difference in the score for white and non-white individuals (p-value = 0.2611). However, we still decided to include this variable in the two-way ANOVA analysis. We considered that even if the t-test does not detect a significant difference between the two groups, it may still have an effect when combined with another variable in a two-way ANOVA test. Also, when interpreting the findings of a statistical test, the p-value should not be the only element to consider: test power and group number can also be essential factors. Therefore, it may still be worthwhile to include *Perceived_Race* in a two-way ANOVA analysis, especially when our sample size is big enough.

For the Two-way ANOVA tests, we tested 2 hypotheses: what is the effect of perceived race and sex on the level of cooperation, and how does the location in Toronto or not and age group affect the level of cooperation at the time of the arrest. Interaction plots are also used here to further interpret the results by visually displaying the relationship between variables, and help gain a better understanding of how the cooperation score is affected by the different levels of the 4 categorical variables. Based on the Two-way ANOVA test with interaction, it is found that the mean of each group is not equal for each independent variable.

Hypothesis 1: Perceived_Race and Sex vs Cooperation score

There is a significant difference in the mean cooperation score for perceived race and sex. However, there is no interaction effect between them (p-value = 0.119). The interaction plot showed that for individuals in the age group with perceived race as Black, East/Southeast Asian, Latino, and White, 'Male' individuals tend to have a higher cooperation score compared to those in the sex type as 'Female'. For individuals as Indigenous and South Asian, 'Male' individuals tend to have a lower cooperation score compared to 'Female'. For Middle-Eastern individuals, the mean cooperation score for both groups is very close to each other. Black and Indigenous individuals have lower mean cooperation scores compared to other race groups, regardless of gender, while East/Southeast Asians had the highest mean cooperation scores. Additionally, significant differences in mean cooperation scores were observed between individuals with perceived race as Latino, Indigenous, and South Asian of different sex types.

Based on the Two-way ANOVA test, the final model formula of *Hypothesis 1* is: **Cooperation score ~ Perceived Race + Sex.** The intercept shows that the cooperation score is 8.31 for individuals with perceived race as Black and in the sex type as female. There are statistically significant differences in mean cooperation score between individuals with perceived race as East/Southeast Asian, Indigenous, Latino, Middle-Eastern, South Asian, and White compared to those with perceived race as Black, while holding other variables constant. The coefficient for whether an individual is male compared to female is not statistically significant.

Hypothesis 2: In Toronto or not and Age group at arrest vs Cooperation score

There is a significant difference in the mean cooperation score for those arrested in Toronto and those who are not, as well as for different age groups at arrest. However, there is no interaction effect between being arrested in Toronto and the age group at arrest (p-value = 0.1865). The interaction plot showed individuals in the 25-34 and 35-44 age groups had lower mean cooperation scores compared to other age groups, particularly when they were arrested in Toronto. Younger individuals (less than 18 years old) and those over 65 had the highest mean cooperation scores regardless of the location of the arrest. For individuals that are greater than 18 and lower than 55,

individuals arrested inside Toronto tended to have a lower cooperation score compared to those arrested outside Toronto.

Based on the Two-way ANOVA test, the final model formula of *Hypothesis 2* is: **Cooperation score ~ Intoronto + Age_group_at_arrest.** The intercept shows that individuals who are arrested outside Toronto and in the age group of less than 18 have a cooperation score of 8.57. Also, being arrested inside Toronto leads to a statistically significant decrease in cooperation score by 0.04. Individuals in age groups between 18 and 64 have statistically significant lower cooperation scores compared to those who are less than 18 years old. However, the coefficient for individuals who are 65 years old or older is not statistically significant, indicating that there is no significant difference in cooperation scores compared to those who are less than 18 years old.

For further study, in order to find out which groups differ from each other, post-hoc tests (Tukey's HSD) are performed: for sex versus cooperation score, there is a significant difference in mean cooperation score between female and male groups(the adjusted p-value is 0.0022).

When analyzing perceived race, we found significant differences in mean cooperation score for each race group. Specifically, for Black individuals, East/Southeast Asian individuals, Indigenous individuals, and white individuals, the adjusted p-value for the mean difference between them and all other race groups is 0.001, indicating a significant difference (except for the pair of East/Southeast and Latino). For Latino individuals, there is a significant difference in mean cooperation score compared to those with perceived race as black, Indigenous, and white. Middle-Eastern individuals do not show a significant difference in mean cooperation score compared to those with perceived race as South Asian and Latino.

For the location of arrest, the adjusted p-value of 0.001 shows that there is a significant difference in mean cooperation score between individuals arrested in Toronto and those not.

Significant differences in mean cooperation scores for each age group have also been found. Individuals less than 18 years old showed a significant difference in mean cooperation score compared to those in 18-24, 25-34, and 35-44 age groups; individuals in the 18-24 age group showed a significant difference in mean cooperation score compared to those less than 18 and greater than 65 years old; individuals in the 55-64 age group showed a significant difference compared to those in the 25-34 age group; individuals 65+ showed a significant difference compared to those in the 18-24, 25-34, and 35-44 age groups, etc.

5 Discussion

5.1 Limitations

The findings of this project have to be seen in light of some limitations.

Firstly, the dataset used in this study is limited to a single city, Toronto, which means that the results may not be generalizable to other locations. And it only covers a relatively short period of 2 years (*'Arrest_Year'*), which may not capture long-term trends or changes over time.

Besides, there are some issues with the accuracy of the data used in the study. The dataset only relies on self-reported data from officers, which may be biased or incomplete. For example, it includes indicators of whether a person was booked at a police station within 24 hours following an arrest, but due to issues with the booking template, there are records where a person was strip searched but the data does not indicate a booking. This could impact the validity of the study's findings. Additionally, some of the data in the *'ArrestLocDiv'* column is marked as 'XX'. It is unclear whether this means that the data is an accurate representation that it is outside of the City of Toronto or it just cannot be geo-coded. The age of the person arrested is given as a range, not the exact age, and the arrest month *'Arrest_Month'* is also given as a period of months. These could impact the accuracy of the study's findings and limit the ability to draw strong conclusions from the data.

Another limitation of the study is related to the calculation of the cooperation score. The scoring system used in the study is subjective and is fully based on the judgment of the person assigning the values to each behavior at arrest. The coefficients used to adjust the cooperation score may be adjusted based on the specific needs of the law enforcement agency using the system.

Finally, there are some missing values for the *Search Reason* and *'ItemsFound'* columns of the data, which could lead to incomplete or biased results. Overall, these limitations should be taken into consideration when interpreting the findings of the study and applying them to make policy decisions.

5.2 Further Improvement

To address the limitations, several improvements could be made for our future studies.

Firstly, future research could include data from multiple cities to see if the results are consistent across different locations. We can also use data from a longer period to capture long-term trends and changes in policing patterns, and use larger sample sizes to increase the statistical power of the analysis. To improve the accuracy of the data, we could consider using additional data sources to supplement the officer's self-reports. Efforts should also be made to minimize missing data in future studies.

Secondly, compared to the use of continuous cooperation scores as the dependent variable in this paper, choosing the probability of cooperation occurrence in logarithmic form as the dependent variable may be a more relevant and accurate approach. The dependent variable can be set to range from 0 to 1 (0 for not occurring, 1 for occurring), and the number in the interval represents the expected probability of cooperation for the actor. This way the probabilistic form of representation allows for an intuitive observation of the results of the statistical analysis compared to an abstract score that still requires a secondary analysis. Again, the chi-square test in this case would be a better

test than the t-test because it directly analyzes the proportions of the various cases of the dependent variable among the subgroups.

For future studies, we could also investigate additional variables that may impact the cooperation scores. By doing so, it could provide more comprehensive and accurate insights into the dynamics of arrests and strip searches in the criminal justice system.

6 Conclusion

This study aims to study the relationship between the level of cooperation during an arrest and several indicators, including the race, gender, age group of the individual, and whether the location in Toronto or not at the time of the arrest, based on the 'Arrests and Strip Searches' dataset from the Toronto Police Service.

The two research questions we had were: 1. What is the effect of ***perceived race*** and ***sex*** on the level of cooperation at the time of arrest? and 2. How does ***location in Toronto or not*** at arrest and ***age group*** affect the level of cooperation at the time of arrest? Through exploratory data analysis, t-tests, two-way ANOVA tests, and post-hoc tests, we found that perceived race, sex, age group, and the location at arrest are all significant factors that impact cooperation scores.

Specifically, the results show that the majority of individuals were willing to cooperate with the Toronto police. Female individuals tend to have a lower cooperation score compared to male individuals, and individuals arrested in Toronto tend to have a lower cooperation score compared to those arrested outside Toronto. There are significant differences in mean cooperation scores for each race group, with Black and Indigenous individuals having lower mean cooperation scores compared to other race groups, and East/Southeast Asian individuals having the highest mean cooperation scores.

While the findings of this study provide valuable insights, there are limitations that need to be addressed in future research. The dataset used in this study relies on self-reported data from officers, which may be biased or incomplete, and the calculation of the cooperation score is subjective. To address these limitations, future research could expand the scope of the dataset to include multiple cities, use larger sample sizes, and consider alternative data collection methods. Efforts should also be made to refine the scoring system used to calculate the cooperation score.

Overall, this study highlights the importance of understanding the factors that influence the level of cooperation at arrest. It could help Toronto police and other law enforcement agencies to identify strategies for reducing conflicts during the arrest process. Also, by developing a more efficient allocation of police resources, law enforcement can better focus on areas with a higher potential for resistance, and develop better strategies to improve cooperation during arrests, which would ultimately lead to a safer and more effective community policing approach.

References

Bierie, David M. 2015. ‘‘Assault of Police.’’ *Crime & Delinquency*. doi: 10.1177/0011128715574977.

Engel, R. S. (2003). *Explaining suspects' resistance and disrespect toward police.* Journal of Criminal Justice, 31, 475-492.

McCluskey, J. D., S. D. Mastrofski, and R. B. Parks. 1999. ‘‘To Acquiesce or Rebel: Predicting Citizen Compliance with Police Requests.’’ *Police Quarterly* 2:389-416.

Kaminski, R., & Sorensen, D. (1995). *A multivariate analysis of individual, situational and environmental factors associated with police assault injuries.* American Journal of Police, 14, 3-48.

O'Neil, J., & Hanson, A. (2020). Toronto police launch review of use of force, racial profiling in wake of protests. *CBC News*. https://www.cbc.ca/news/canada/toronto/toronto-police-review-use-of-force-racial-profiling-1.5623215

Perreault, Samuel. (2019). *Police-reported crime in rural and urban areas in the Canadian provinces, 2017.* Statistics Canada. https://www150.statcan.gc.ca/n1/pub/85-002-x/2019001/article/00009-eng.htm

U.S. Department of Justice, Federal Bureau of Investigation. (2009). *Uniform Crime Reporting Program Data: Arrests by Age, Sex, and Race, 2009*. ICPSR30761-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-30. doi: 10.3886/ICPSR30761.

Walker, S. (1999). *The police in America (3rd ed.).* Boston: McGraw-Hill.

Appendix

```
def welch_dof(x, y):
    '''
    Takes in final grades for binary features
    Returns degrees of freedom
    '''
    dof = (x.var()/x.size + y.var()/y.size)**2 / (
        (x.var()/x.size)**2 / (x.size-1) + (y.var()/y.size)**2 / (y.size-1))
    return dof
```

1. Conduct T-test on 'Sex' category (M is male, F is female):

```
M_score = df[df["Sex"] == "M"]['Cooperation_score']
F_score = df[df["Sex"] == "F"]['Cooperation_score']
print('Ttest results for Cooperation score:',stats.ttest_ind
      (M_score, F_score, equal_var=False)) #p-value is significant

# calculate mean and standard deviation of groups
print('Mean score for M_score, F_score:',M_score.mean(), F_score.mean())
print('Standard deviation for M_score, F_score:',M_score.std(), F_score.std())

# compute confidence interval
cm = sms.CompareMeans(sms.DescrStatsW(M_score), sms.DescrStatsW(F_score))
print('Confidence interval for Cooperation_score:',cm.tconfint_diff(usevar='unequal'))

# welch dof
print('DOF for Cooperation_score:',welch_dof(M_score, F_score))
```

Figure 3. Code Implementation of t-test on sex vs cooperation score

2. Conduct T-test on 'intoronto' category (1 is in toronto, 0 is not in toronto):

```
int_score = df[df["intoronto"] == "1"]['Cooperation_score']
not_score = df[df["intoronto"] == "0"]['Cooperation_score']
print('Ttest results for Cooperation score:',stats.ttest_ind
      (int_score, not_score, equal_var=False)) #p-value is significant

# calculate mean and standard deviation of groups
print('Mean score for int_score, not_score:',int_score.mean(), not_score.mean())
print('Standard deviation for int_score, not_score:',int_score.std(), not_score.std())

# compute confidence interval
cm = sms.CompareMeans(sms.DescrStatsW(int_score), sms.DescrStatsW(not_score))
print('Confidence interval for Cooperation_score:',cm.tconfint_diff(usevar='unequal'))

# welch dof
print('DOF for Cooperation_score:',welch_dof(int_score, not_score))
```

Figure 4. Code Implementation of t-test on in_toronto vs cooperation score

3. Conduct T-test on 'Perceived_Race' category:

```
White_score = df[df["Perceived_Race"] == "White"]['Cooperation_score']
notW_score = df[df["Perceived_Race"] != "White"]['Cooperation_score']
print('Ttest results for Cooperation score:',stats.ttest_ind
      (White_score, notW_score, equal_var=False)) #p-value is NOT significant

# calculate mean and standard deviation of groups
print('Mean score for White_score, notW_score:',White_score.mean(), notW_score.mean())
print('Standard deviation for White_score, notW_score:',
      White_score.std(), notW_score.std())

# compute confidence interval
cm = sms.CompareMeans(sms.DescrStatsW(White_score), sms.DescrStatsW(notW_score))
print('Confidence interval for Cooperation_score:',cm.tconfint_diff(usevar='unequal'))

# welch dof
print('DOF for Cooperation_score:',welch_dof(White_score, notW_score))
```

Figure 5. Code Implementation of t-test on perceived race vs cooperation score

4. Conduct T-test on 'Age_group__at_arrest_' category:

```
younger_score = df[(df["Age_group__at_arrest_"] == "25-34") |
                   (df["Age_group__at_arrest_"] == "-18")]['Cooperation_score']
older_score = df[(df["Age_group__at_arrest_"] != "25-34") &
                 (df["Age_group__at_arrest_"] != "-18")]['Cooperation_score']
print('Ttest results for Cooperation score:',stats.ttest_ind
      (younger_score, older_score, equal_var=False)) #p-value is significant

# calculate mean and standard deviation of groups
print('Mean score for younger_score, older_score:',younger_score.mean(),
      older_score.mean())
print('Standard deviation for younger_score, older_score:',younger_score.std(),
      older_score.std())

# compute confidence interval
cm = sms.CompareMeans(sms.DescrStatsW(younger_score), sms.DescrStatsW(older_score))
print('Confidence interval for Cooperation_score:',cm.tconfint_diff(usevar='unequal'))

# welch dof
print('DOF for Cooperation_score:',welch_dof(younger_score, older_score))
```

Figure 6. Code Implementation of t-test on age group at arrest vs cooperation score