INF 2178 - Midterm Assignment

By: Group 42

Sophia Qiu, Kaicheng Huang

Introduction

Police arrest decisions have a profound impact on the individual being arrested. Arrests could result in negative consequences including but not limited to potential harm to reputation, loss of employment, and ultimately deprivation of liberty. It's imperative to ensure the fairness of arrest decisions to uphold the principles of justice and maintain public trust in law enforcement and the criminal justice system. Many different influential factors can affect a police officer's decision to make an arrest and whether the arrested individual will be booked. Building on an existing body of research, the current study examines the effect of gender and age in shaping the individual's interaction with law enforcement.

Previous research has identified a link between individual demographics, socioeconomic status, and arrest decisions. Visher (1983) found that white female offenders of higher social status were offered more leniency on both the decision to arrest, and in the case of arrest. Visher describes such leniency as stemming from the police's perception of women as less dangerous and needing protection. Other research studies also identified similar biases in arrest decisions citing demographic and social factors as important influencers. Police officers may hold preconceived knowledge about the behaviour and characteristics of different offenders based on their group membership. Individuals exhibiting behaviours inconsistent with the existing schema are more likely to be arrested than those who confirmed the expectations. This bias is particularly prevalent among juvenile offenders (Sealock & Simpson, 1998). Additionally, the public perception of biases and unfairness of police procedures are found to negatively impact an individual's attitude toward the criminal justice system and law enforcement (Pollock & Menard, 2015). These findings highlight the importance of understanding the effect demographic and social factors have on law enforcement practices. Using an arrest dataset published by Toronto Police Service, the current study will explore the effect of time of arrest on the number of arrests, and whether age and gender play a role in mean arrest events.

The research questions for the present study are the following:

RQ1: Is there a significant difference in the number of arrests between different quarters?

Where the null hypothesis for this research question will be:

Null hypothesis (H_0): There is no significant difference in the number of arrests between different quarters.

Alternative hypothesis (H_1): There is a significant difference in the number of arrests between different quarters.

RQ2: Is there a significant difference in the number of arrests events between age groups, while accounting for the effects of gender ?

Since this research question has two independent variables, the hypotheses can be initiated as the following:

Null hypothesis (H_0)	Alternative hypothesis (H1)
There is no significant difference in the number of arrests between different age groups.	There is a significant difference in the number of arrests between different age groups.
There is no significant difference in the number of arrests between different genders.	There is a significant difference in the number of arrests between different genders.
There is no significant difference in the number of arrests between different age groups, while accounting for the effects of gender.	There is a significant difference in the number of arrests between different age groups, while accounting for the effects of gender.

Dataset

Data Overview

The present study is based on the Arrests and Strip Searches dataset obtained from Toronto Police Service Public Safety Data Portal (Toronto Police Service, 2022). The data contains comprehensive information on all arrests and strip searches conducted by the Toronto Police Service from 2020 to 2021. The dataset allows us to gain insight into patterns and trends surrounding arrests and strip search events, including the demographics of those who are arrested and searched. The demographic information contains the individual's personal ID number,

perceived race, sex, age group, and the indicator of whether they are a minority. Indicators regarding the individuals were also recorded in the dataset, such as whether the police conducted a strip search on the individual at arrest, and whether the person was booked at a police station within 24 hours of the arrest event. Other information related to the arrest event was also indicated in the dataset, including the location of Division boundaries where the arrest took place, the category of occurrence leading to the arrest, the individual's action at arrest, the research reason if the individual has been strip-searched, as well as the number of items found through the strip search. This dataset analyzes the frequency and distribution of arrests and strip searches in Toronto, while allowing for the examination of potential factors that may influence police decision-making in these instances.

Data Cleaning

We conducted a comprehensive data cleaning process to ensure the accuracy, completeness, and consistency of the existing dataset. We commenced by grouping the data and creating new columns to calculate the number of arrests for each individual. Specifically, we grouped the data by the PersonID column and utilized the count method to determine the frequency of each PersonID value in the dataset. We then added a new column, "num_arrests," and assigned the corresponding count values to each row. Additionally, we created a new column, "Age_group," to classify individuals as either youth or adults, based on the "Youth_at_arrest__under_18_years" column in the original column. Rows with a value of "Not a youth" in the original dataset were assigned to "Adult" in the new column, and rows with values "Youth (aged 17 and younger)" or "Youth (aged 17 years and under)" are combined and assigned to "Youth". Subsequently, we used the group_by method to group the data by age group. We counted the number of arrests in each group, providing a comprehensive view of the number of arrests associated with different age groups.

Furthermore, we created a new column to indicate the gender of each individual and grouped the data by gender, enabling us to count the number of arrests in each group. We also created a new column to record the quarter in which the individual was arrested and grouped the data by the quarter of each arrest event, providing a quarterly view of the number of arrests. We then created a new data frame for further analysis and removed all null values from the dataset to

ensure the validity of the data. The new DataFrame contains ten columns and 65102 rows with no null values. Overall, these cleaning processes enhanced the quality of the dataset and facilitated further analyses.

Exploratory Data Analysis

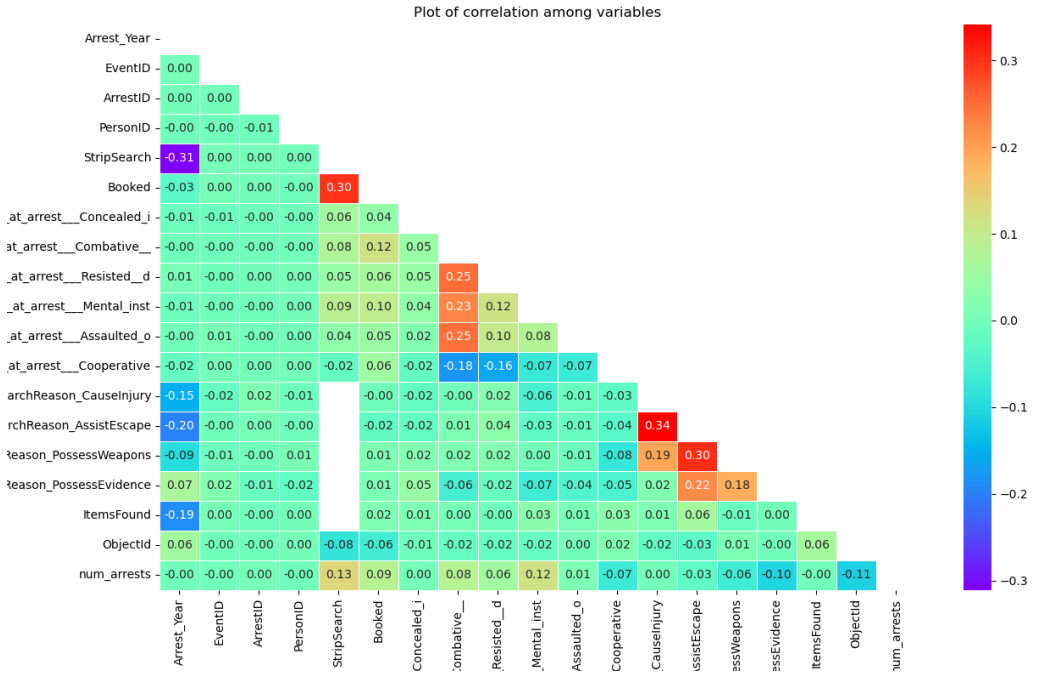
Correlation

A correlation heatmap was plotted to discover the correlation among the variables in the dataset (see Figure 1). However, due to the lack of continuous variables in the data frame, we faced challenges finding the significant association between different variables.

One thing worth noting on the heatmap is that "Booked" is positively associated with "StripSearch" with a correlation of 0.3. Yet, the dataset description stated that all individuals who have been strip-searched are booked within 24 hours. Thus, this inconsistency in the dataset reduces the validity of the correlation between "Booked" and "StripSearch."

Figure 1

Correlation Plot Among Each Variable

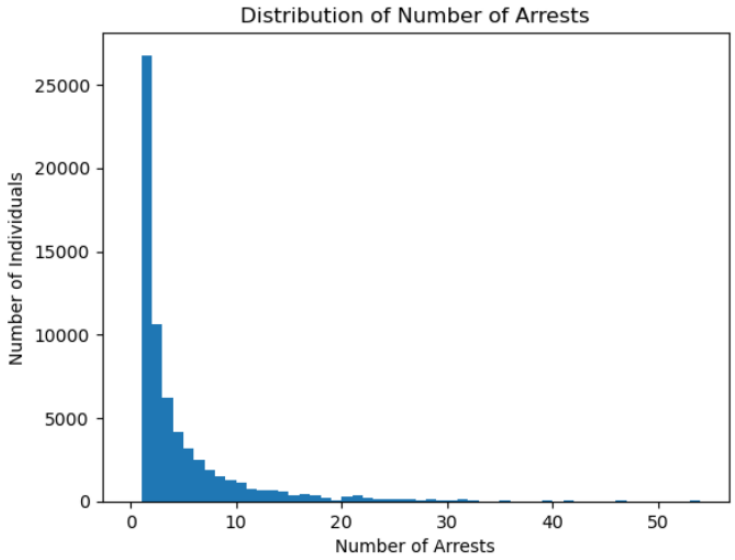


T-Tests and Distributions

Moreover, a histogram was mapped to visualize the distribution of the number of arrests with the number of individuals (see Figure 2). The maximum number of arrests for individuals was 54 times (M = 4.13, SD = 5.4). The histogram also showed the skewness of the data.

Figure 2

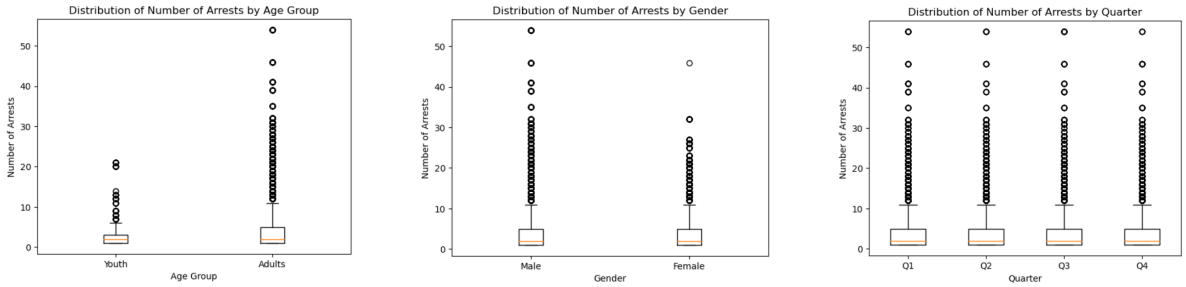
Histogram of Number of Arrests



Subsequently, we conducted t-tests to reveal the significant difference between different age groups in the number of arrests, t(DF) = -15.93, p < .001. The result suggested that adults (M = 4.20) have a significantly higher number of arrests compared to youth (M = 2.61). A box plot was plotted to visualize the distribution of the number of arrests by age group. Another t-test was conducted to compare the average number of arrests between males and females. The results indicated a significant difference between the two genders, t(DF) = 5.18, p < .001, where males (M = 4.19) have a significantly higher number of arrests than females (M = 3.91). A box plot was plotted to visualize the distribution of the number of arrests by gender. Lastly, we plotted a boxplot to visualize the distribution of the number of arrests by the quarter of the year (see Figure 3).

Figure 3

Boxplot of the Number of Arrests by Different Age Groups, Genders, and Quarters.



Shapiro-Wilk Test

A Shapiro-Wilk test was conducted to assess the normality of the distribution of the dataset. The test statistic is 0.616 and the p-value is less than 0.05, which suggests the null hypothesis should be rejected, and the dataset is not normally distributed. However, the accuracy of the Shapiro-Wilk test may be affected when the sample size is larger than 5000, the p-value may inflate and be inaccurate with the current dataset (n = 65277).

Methods

The results obtained from the descriptive statistics further guided our direction to investigate our research questions.

RQ1 - Is there a significant difference in the number of arrests between different quarters?

From the boxplot above, we could not conclude whether there is any statistical significance of the mean number of arrests in different quarters of the year. Therefore, we planned to investigate RQ1 using one-way ANOVA to test if there is any significant difference between the four quarters. The one-way ANOVA helps to compare the means of three or more groups, which allows us to compare the number of arrests in four different quarters. If the one-way ANOVA test yields a statistically significant result, we will proceed to conduct a Tukey's HSD for the post-hoc test to identify which specific quarters are significantly different from each other.

RQ2 - Is there a significant difference in the number of arrests events between age groups, while accounting for the effects of gender ?

Based on the t-tests conducted, we found statistically significant differences in the number of arrests between age groups (youth and adults) and genders (male and female). Therefore, we planned to conduct a two-way ANOVA to examine whether there is a main effect of age groups (youth and adult), genders (male and female), and their interaction effect on the number of arrests. The two-way ANOVA not only analyzes the effect of age group and gender on the number of arrests independently, but also investigates their combination. If the result of the two-way ANOVA is significant, we will proceed to a post-hoc test, like Tukey's HSD, to identify which specific groups exhibit a significant difference from each other.

Results and Findings

RQ1 - Is there a significant difference in the number of arrests between different quarters?

One-Way ANOVA

We implemented a one-way ANOVA to answer our first research question, which aimed to reveal the effect of quarters of the year (Q1, Q2, Q3, and Q4) on the number of arrests. However, the results showed no significant difference between the means of the quarters ($F(3, 65098) = 1.03$, $p > .05$). The mean number of arrests for each quarter was as follows: Q1 = 4.15 (SD = 5.43), Q2 = 4.23 (SD = 5.38), Q3 = 4.15 (SD = 5.43), and Q4 = 4.15 (SD = 5.43).

The results obtained from the one-way ANOVA test showed no significance. This indicated that there is not enough evidence to reject the null hypothesis, meaning that there is no significant difference in means between quarters. Without a significant result from the one-way ANOVA, there is no basis for conducting a post-hoc test, as it is designed to determine which specific groups are significantly different from each other.

RQ2 - Is there a significant difference in the number of arrests events between age groups, while accounting for the effects of gender ?

Two-Way ANOVA

We conducted a two-way ANOVA to investigate how the changes in age group and gender will affect the mean of the number of arrests (see Table 1). The results indicated that age group ($F(1, 65098) = 248.27$, $p < .001$) has a significant main effect on the number of arrests, meaning that there is a significant difference in the number of arrests between the two age groups (youth and adult) compared. The findings also suggested that gender ($F(1, 65098) = 22.40$, $p < .001$) has a significant main effect on the number of arrests, indicating that there is a significant difference in the number of arrests between the two genders (male and female) compared. Additionally, the results showed that there was a significant interaction effect between age group and gender, $F(1, 65098) = 5.77$, $p = .016$. This indicates that the relationship between age group and the number of arrests depends on gender, meaning that the effect of age group on the number of arrests may differ for males and females. The interaction effect cannot be interpreted in isolation from the main effects of age group and gender, but it suggests that the relationship between age group and the number of arrests is not the same for males and females. It is essential to conduct follow-up analyses to explore the interaction effect's nature further.

Table 1

Two-way ANOVA results for RQ2

Predictor	Sum of Squares	*df*	F	*p*
(Intercept)	167.82	1.00	5.77	.016
Age group	7221.36	1.00	248.27	<.001
Gender	651.47	1.00	22.40	<.001
Residual	1893519	65098.00		

Tukey's HSD

To further identify which groups are significantly different in comparison, we conducted a Tukey's HSD post-hoc test to examine differences between multiple means with a family-wise error rate of 0.05 (see Table 2). The results showed significant differences in the number of arrests between all comparison groups, except between YouthFemale and YouthMale (mean difference = -0.2811, p = 0.6089, 95% CI [-0.8691, 0.3069]). Expressly, the results indicated that the number of arrests for AdultMale is significantly higher compared to Adultfemale (mean difference = 0.2844, p < .001, 95% CI [0.1429, 0.4259]), YouthFemale (mean difference = -1.4352, p < .001, 95% CI [-1.9511, -0.9194]), and YouthMale (mean difference = -1.7163, p < .001, 95% CI [-2.0119, -1.4208]). Furthermore, AdultFemale showed a significantly higher mean in the number of arrests compared to YouthFemale (mean difference = -1.1509, p < .001, 95% CI [-1.6786, -0.6232]), and YouthMale (mean difference = -1.432, p < .001, 95% CI [-1.7477, -1.1162]).

Table 2

Multiple Comparison of Means - Tukey's HSD post-hoc test results for RQ2

				95% Confidence Interval	
Group 1	Group 2	Mean Difference	*p*	Lower	Upper
AdultFemale	AdultMale	0.2844	>.001	0.1429	0.4259
	YouthFeamle	-1.1509	>.001	-1.6786	-0.6232
	YouthMale	-1.432	>.001	-1.7477	-1.1162
AdultMale	YouthFeamle	-1.4352	>.001	-1.9511	-0.9194
	YouthMale	-1.7163	>.001	-2.0119	-1.4208
Youth Female	YouthMale	-0.2811	0.6089	-0.8691	0.3069

Note. The mean difference is significant at the .05 level.

Normality Test

A quantile-quantile plot (Q-Q plot) was generated to assess the normality of the data, where the points are located more closely to the diagonal line indicating the data is more normally distributed. The Q-Q plot showed non-linear points, representing that the data deviates from the normal distribution and is skewed to the left (see Figure 4).

Figure 4

Normal QQ Plot for Two-Way Anova Test

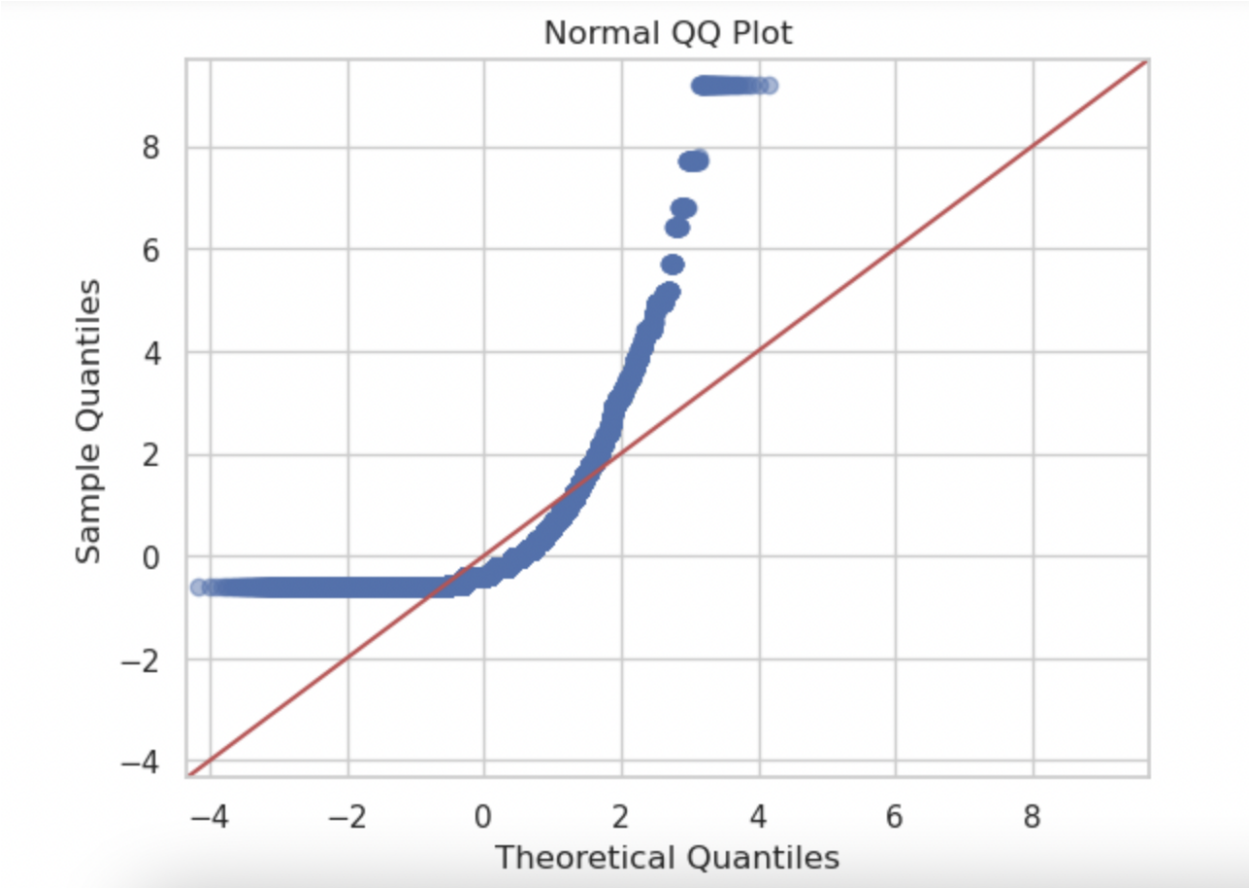
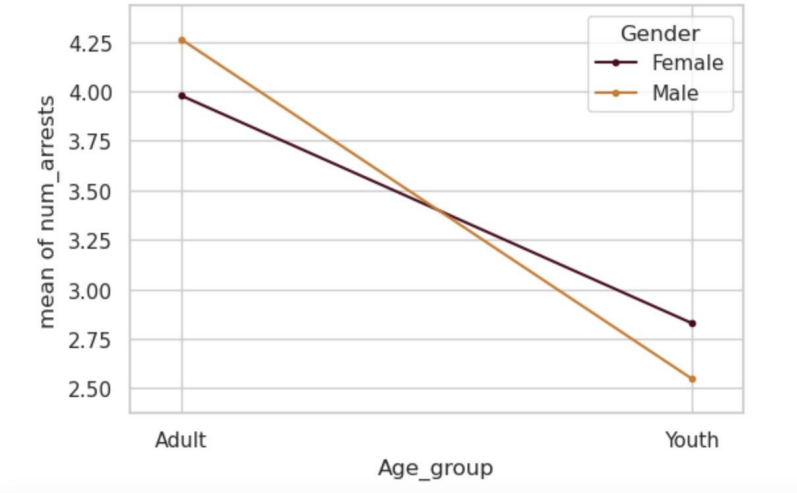


Figure 5

Interaction plot



Levene's Test

A Levene test was conducted to assess the homogeneity of variances between age groups and genders. The result indicated a statistical significance with a test statistic of 86.40 ($p < .001$). We rejected the null hypothesis since the p-value was less than 0.05. Therefore, the assumption of homogeneity of variances was violated, meaning that the variances are not equal across age groups and genders. An alternative test that does not assume homogeneity of variances is suggested in the follow-up research.

Discussion

The one-way ANOVA result indicated that the number of arrests did not significantly differ across the four quarters. One possible explanation for this finding is that it may be difficult to draw inferences from quarterly data due to weather fluctuations. According to the meteorological definition, spring covers the period from March to May, summer from June to August, fall from September to November, and winter from December to February. Based on the long-term temperature trend in the Greater Toronto Area, summer is the hottest season, while

winter is the coldest (Mohsin & Gough, 2010). We posited a hypothesis that the number of arrests in Toronto is associated with the temperature, with a higher number of arrests during the warmer seasons. However, the present study may not fully address this hypothesis since the four quarters do not align with the meteorological seasons. In addition, due to the dataset's limitation, which recorded the Arrest_Month in three-month intervals (e.g., January-March), we were unable to split the data into meteorological seasons. As a result, the analysis may be less precise in identifying the impact of temperature on the number of arrests. Further research with more comprehensive data could provide more robust results.

The two-way ANOVA results indicated that both age group and gender have a significant main effect on the number of arrests in Toronto, implying that they are significant predictors for differences in the number of arrests. The interaction effect between age group and gender was also found to be significant, suggesting that the relationship between age group and the number of arrests may differ between males and females, and the relationship between gender and the number of arrests may differ between youth and adults. Further post-hoc tests were conducted to explore this interaction effect. Notably, the results indicated that gender is not a significant factor in predicting the number of arrests for youth, as no difference was found between youth females and youth males. This suggests that any observed differences between males and females in youth may be attributed to chance or other factors not accounted for in the study. Moreover, adult males exhibited a significantly higher number of arrests compared to the other three groups (adult females, youth males, and youth females). Additionally, adult females had a significantly higher number of arrests compared to youth males and females. These findings suggest that gender and age group are important factors to consider when examining differences in the number of arrests in Toronto.

The present study has certain limitations that should be taken into account when interpreting the results. One potential limitation is that the age range is broad for both youth and adults. The youth group ranges from 0 to 17 years old, which means that there is no differentiation between pre-adolescents and adolescents, as the latter group typically ranges from 12-17 years old. Furthermore, the adult group is accounting for anyone who is above 17 years old. The lack of information on the specific ages of the adult group may limit the ability to draw

more precise conclusions about the age-related differences in the number of arrests. Additionally, the number of youth in the dataset is relatively small (n = 3043) compared to the number of adults (n = 62235), which may limit the generalizability of the findings to the larger population. Therefore, a follow-up study that further differentiate the age group by years is suggested to draw a more robust conclusion for the effect of age on the number of arrests.

Another limitation is the inconsistencies identified in the dataset. Specifically, some rows showed the same personID, suggesting that the same individual was involved in multiple arrest events. However, those rows contain different perceived races with the same personID, indicating a lack of consistency across the dataset. Moreover, the dataset contains a miscode in the "Booked" column. Based on the description, all individuals who have been strip-searched are booked within 24 hours. However, the data contains rows that showed positive in strip search and negative in booked. These discrepancies in the data raise concerns about the reliability and validity of the dataset and may limit the extent to which accurate conclusions can be drawn from the data.

Another important limitation of the present study is that the two-way ANOVA result does not meet the assumption of normal distribution and homogeneity of variances. These violations of assumptions can potentially lead to inaccurate conclusions or misleading results, raising questions about the reliability and validity of the findings. In particular, violations of the normality assumption may affect the accuracy of the p-values, which may result in either type I or type II errors. Violations of the homogeneity of variance assumption may lead to unequal error variances, which can affect the F-test's significance and the test's power. Future research with a dataset that addresses these issues is suggested to improve the accuracy and generalizability of the findings.

Conclusion

In conclusion, this study investigated the differences in the number of arrests between different quarters, as well as between different age groups (youth and adults) while accounting for the effects of gender (males and females). The results showed that there was no significant difference in the number of arrests between different quarters, indicating that the number of

arrests was evenly distributed throughout the year. Moreover, a main effect was found between the mean number of arrests between different age groups, with adults having a higher number of arrests compared to the youth. Additionally, the results revealed a significant main effect of gender, with males having a higher number of arrests than females. Furthermore, the interaction effect between age group and gender was also found to be statistically significant, indicating that the effect of age group on the number of arrests may also depend on gender. These findings have important implications for law enforcement agencies and policymakers, as they suggest that efforts should be made to address the factors that contribute to higher rates of arrests among adults and males, and to explore the unique factors that contribute to the relationship between age groups and the number of arrests among males and females. Future research is suggested to conduct follow-up analyses to further investigate the interaction effect and explore other potential factors that may contribute to differences in the number of arrests.

References

Mohsin, T., & Gough, W. A. (2010). Trend analysis of long-term temperature time series in the Greater Toronto Area (GTA). Theoretical and applied climatology, 101, 311-327.

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.

Sealock, M. D., & Simpson, S. S. (1998). Unraveling bias in arrest decisions: The role of juvenile offender type-scripts. Justice quarterly, 15(3), 427-457.

Toronto Police Service. (2022). Arrests and Strip Searches (RBDC-ARR-TBL-001). PUBLIC SAFETY DATA PORTAL.

Visher, C. A. (1983). Gender, police arrest decisions, and notions of chivalry. Criminology, 21(1), 5-28.