INF2178 Experimental Design for Data Science

Steve Li

Bhanvi Gupta

February 28th, 2023

Insights from Arrests Made by the Toronto Police Service, 2020-2021

Midterm paper submitted in conformity with the requirements for the course of INF2178H (Winter 2023)

Link to Google Colab notebook for the code:
https://colab.research.google.com/drive/1b86Cg1xwLRYdOMyMYngP1ozO3MhncevY?usp=sharing

Faculty of Information

University of Toronto

Table of Contents

1. Introduction

Justice in policing has been a key social and ethical issue that continues to garner people's attention and spark debates over the years. As citizens with conscience and a moral compass, it is natural for us to turn our attention to injustices in policing, such as police brutality and the disproportional use of policing on certain racial or ethnic groups. While many sensational and outrageous events, such as the shooting of Michael Brown in Ferguson, Missouri in 2014 and the killing of George Floyd in Minneapolis, Minnesota in 2020, take place every few years and cause public attention and outrage, many instances of police injustice get considerably less attention and thus get overlooked. To continue the scrutiny on police forces and their use of violence and resources, we can use statistical analysis as a tool to examine for any injustices. For this project, a dataset on arrests and strip searches released by the Toronto Police Service (TPS) is used. This project aims to explore the relationship between the arrests and a person's demographics such as race, gender, and age group.

1.1 Literature Review

Arrests in Toronto have been the focus of a number of studies in recent years. In this literature review, I will summarize the main findings from some of the key studies in this area. One area of research has been focused on the relationship between race and arrests in Toronto. A study by Owusu-Bempah and Wortley (2011) found that black and brown people were overrepresented in Toronto's jail population, and that this was due to systemic factors such as discrimination, poverty, and biased policing. Similarly, a study by Tanovich (2016) found that racial profiling was a significant problem in Toronto, with police more likely to stop and search individuals from racialized communities.

Another area of research has explored the impact of arrests on individuals and communities in Toronto. A study by Roach (2016) found that even a brief period of detention can have a significant impact on an individual's life, including their employment and housing opportunities. Similarly, a study by Clarke and Wortley (2019) found that the overrepresentation of black people in Toronto's criminal justice system had a negative impact on their mental health and well-being.

A number of studies have also examined the use of police discretion in arrests in Toronto. A study by Brian C. Renauer, Emma Covelli (2011) found that police officers often used their discretion when deciding whether or not to make an arrest, and that this discretion could be influenced by factors such as the offender's demeanor, the seriousness of the offense, and the officer's personal biases.

In general, younger age groups tend to have higher rates of arrests for crimes such as vandalism, drug offenses, and property crimes. However, older age groups may be more likely to be arrested for white-collar crimes, such as fraud or embezzlement.

According to data from the Toronto Police Service, in 2020, the highest number of arrests were made for individuals between the ages of 18-24, accounting for 34.7% of all arrests. The next highest age group was individuals between the ages of 25-34, who accounted for 27.5% of all arrests. This suggests that younger age groups are more likely to be involved in criminal activity than older age groups. It was also found that males were more likely to be arrested for crimes than females. Males accounted for 74.1% of all arrests, while females accounted for 25.7%. This trend is consistent with national and international patterns in which males are overrepresented in the criminal justice system.

There are also gender-specific crimes that occur in Toronto, such as sexual assault and domestic violence. In general, women are more likely to be victims of these crimes, while men are more likely to be perpetrators. The Toronto Police Service reports that in 2020, 85% of all reported sexual assaults and 80% of all reported domestic violence incidents involved female victims.

It's important to note that the overrepresentation of males in the criminal justice system and the higher rates of victimization among women are complex issues that are influenced by a variety of social, cultural, and economic factors. It's crucial to address the root causes of these issues in order to create safer communities for all individuals.

In conclusion, arrests in Toronto have been studied from various perspectives, including the relationship between race and arrests, the impact of arrests on individuals and communities, the use of police discretion in arrests, and the use of alternatives to arrests. The findings from these studies highlight the need for ongoing research and policy interventions to address issues related to arrests in Toronto, particularly those related to racial discrimination and bias in policing.

1.2 Research Questions and objectives

The objective of our study is to find out the relationship between demographic factors such as gender, youth status, and race and the number of arrests. We will be running multiple exploratory data analysis as well as descriptive data analysis tests on the dataset to draw meaningful conclusions. The research questions are listed as follows:

- We wish to find out if there is a difference in the number of arrests between genders (RQ1).
- We wish to find out if there is a difference in the number of arrests between youths and non-youths (RQ2).
- We wish to find out if there is a difference in the number of arrests based on the person’s perceived race (White, Black or Latino) (RQ3).
- We wish to find out if there is a difference in the number of arrests based on the person’s age group (18 to 24 years, 25 to 34 years, or 35 to 44 years) (RQ4).
- We wish to find out if there is a difference in the number of arrests based on the person’s perceived race and age group, and if an interaction effect exists between perceived race and group (RQ5).

2. Exploratory Data Analysis (EDA)

2.1 Descriptive statistics

```
[ ] #View the descriptive statistics for the number of arrests
    NumofArrests.describe()
```

```
count    37329.000000
mean         1.736103
std          2.012614
min          1.000000
25%          1.000000
50%          1.000000
75%          2.000000
max         54.000000
Name: PersonID, dtype: float64
```

Figure 2.1.1 Number of Arrests

```
[ ] #create a boxplot for the number of arrests
    plt.figure(figsize= (20,10))
    plt.boxplot(NumofArrests, vert=False)

    # show plot
    plt.show()
```

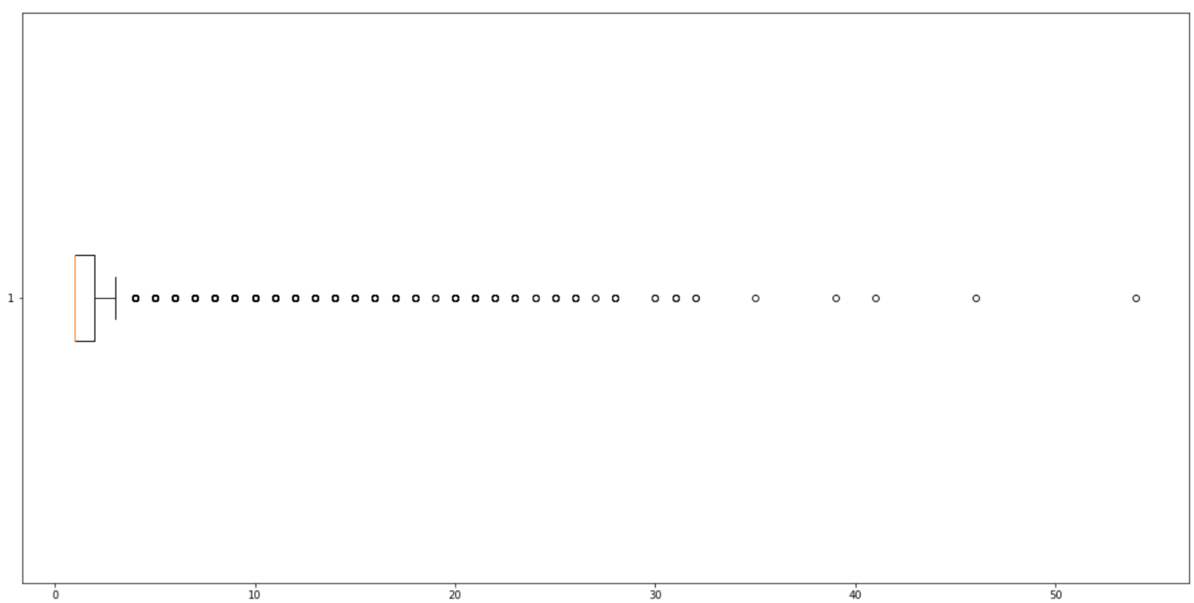

Figure 2.1.2 Horizontal Boxplot for Number of Arrests

Figure 2.1.1 shows that, based on the number of unique PersonIDs in the NumofArrests pandas series, a total number of 37329 people have been arrested. The highest number of times a person has been arrested is 54, and the lowest number of times a person has been arrested is 1. The average number of times a person is arrested is 1.73.

From the boxplot in Figure 2.1.2, we can see that there are a lot of outliers for the number of arrests, including some extreme outliers. The presence of extreme outliers drives up the mean and may affect the effectiveness of the t-tests. To deal with the extreme outliers, we decided to exclude the numbers of arrests that are greater than or equal to 30.

```
[15] #remove the outliers from NumofArrests
     validNumofArrests = NumofArrests[NumofArrests < 30]
     print(validNumofArrests)

     335617    28
     318274    28
     329126    28
     325445    28
     321623    27
               ..
     309949     1
     325498     1
     318570     1
     324886     1
     310583     1
     Name: PersonID, Length: 37314, dtype: int64

[16] validNumofArrests.describe()

     count    37314.000000
     mean         1.724983
     std          1.912217
     min          1.000000
     25%          1.000000
     50%          1.000000
     75%          2.000000
     max         28.000000
     Name: PersonID, dtype: float64
```

Figure 2.1.3 Valid Number of Arrests

Figure 2.1.3 shows the descriptive statistics of the number of arrests after the extreme outliers greater than or equal to 30 have been removed. There are a total of 37314 valid arrests. The mean number of arrests is 1.72. The standard deviation of the number of arrests is 1.91. The 25th percentile is 1. The median is 1, and the 75th percentile is 2.

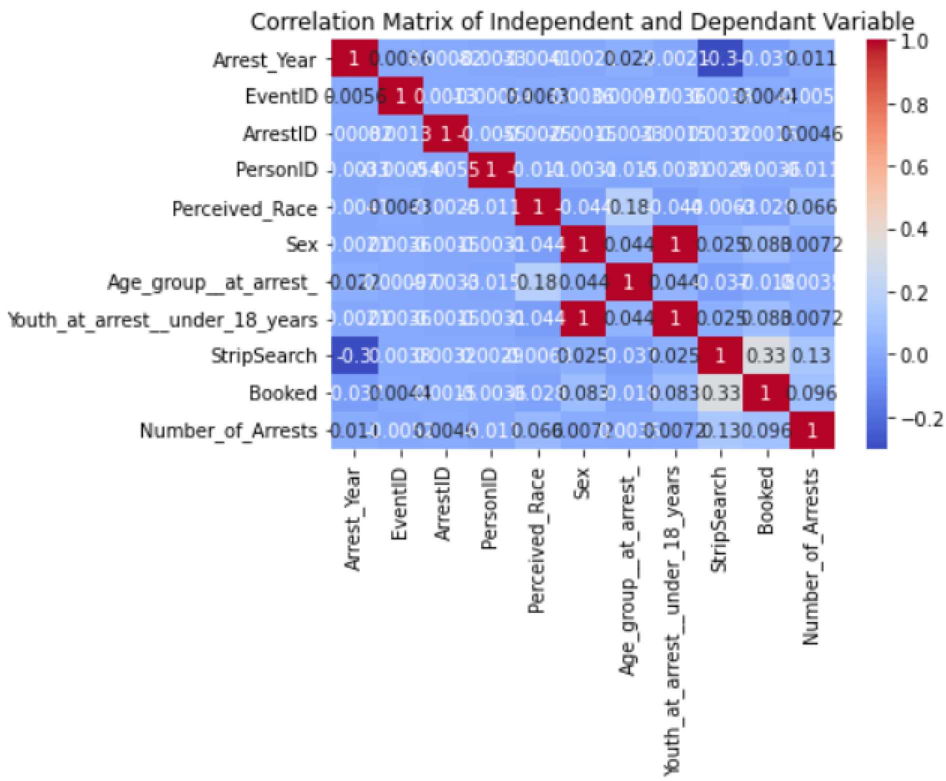

Figure 2.1.4 Correlation Matrix

From the correlation matrix, we can see that youth_at _arrest_under_18_years and sex are highly correlated. While this is not going to change our estimates, it can still lead to multicollinearity. Since we want to study the relationship between age group, sex and race in further sections, we will not be dropping these variables for goodness of fit.

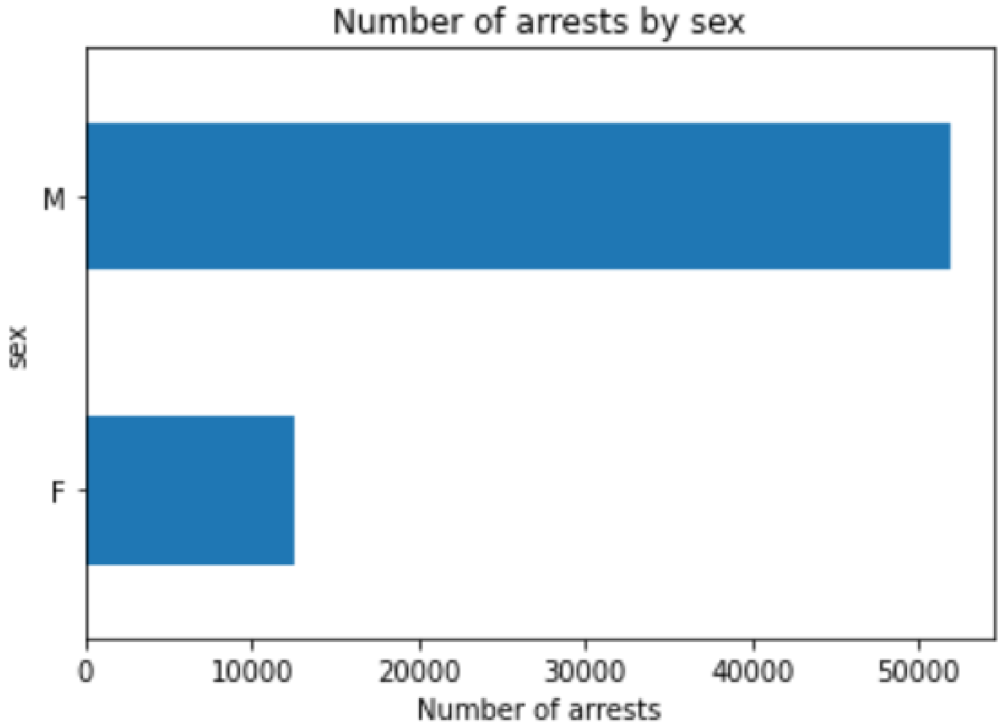

Fig 2.1.5 Number of Arrests by sex

From this chart (1- male, 0- female), we can see that the number of arrests for males is much higher than that of females. We also grouped data on the basis of the three parameters we are studying i.e race, sex and age groups to see the frequency of each of those arrests being reported. We found out that the number of arrests committed are more by young people than older people. So, according to the exploratory analysis, the probability of arresting a black young male is the highest.

2.2 T-tests

2.2.1 T-test #1

The purpose of the first t-test is to investigate the following research question (RQ1): Is there a difference in the number of arrests between males and females? The person's sex (male or female) is the independent variable measured with two categories, and the number of arrests is the dependent, continuous variable. Because

Figure 2.1.5 shows that, in the dataset, the number of arrests for males is much higher than that of females. To ensure the accuracy of the t-test, we wanted to make sure the sample size of each group is the same. Thus, we took random samples of 10000 from each sex group for the t-test.

For the first t-test, the null and alternative hypotheses are as follows:

- Null hypothesis: On average, males and females DO NOT differ in the number of arrests. (H_0: μ1 = μ2)
- Alternative hypothesis: On average, males and females DO differ in the number of arrests. (H_a: μ1 ≠ μ2)

2.2.2 T-test #2

The purpose of the second t-test is to investigate the following research question (RQ2): Is there a difference in the number of arrests between youths and non-youths? The person's youth status (youth or not a youth) is the independent variable with two categories, and the number of arrests is the dependent, continuous variable. In the dataset, the number of arrests for non-youths is much higher than that of youths. To ensure the accuracy of the t-test, we wanted to make sure the sample size of each group is the same. Thus, we took random samples of 10000 from each group for the t-test.

For the 2nd t-test, the null and alternative hypotheses are as follows:

- Null hypothesis: On average, non-youths and youths DO NOT differ in the number of arrests. (H_0: μ1 = μ2)
- Alternative hypothesis: On average, non-youths and youths DO differ in the number of arrests. (H_a: μ1 ≠ μ2)

3. Method

3.1 Dataset Description

The dataset used for this study was released by the Toronto Police Service in November 2022. It contains information on arrests and strip searches in the years 2020 and 2021. Although not explicitly stated as so, it is inferred that the dataset includes all arrests and strip searches that took place in these two years. No exact date and time is included in the dataset – only the year (2020 and 2021) and month range (January to March, April to June, July to September, or October to December) are given. The dataset also includes the event ID and arrest ID of each arrest or strip search event. Regarding personal information, for each person involved in the arrest or strip search event, the person ID, perceived race, sex, age group at arrest, and their youth status (whether they are 17 years or younger) are recorded. Regarding the location of the arrest, a two-digit location code is included for the applicable entries.
The dataset also includes information on whether the person has undergone a strip search and whether they were booked at a police station within 24 hours of the arrest. Regarding the arrest event itself, the dataset includes the category of the arrest. The person's actions during the arrest, for instance, if they were cooperative, or if they assaulted the officer, is included. For strip searches, the reason that prompted the search was recorded.

As a part of our data preprocessing, we removed duplicate entries and created a continuous variable called the 'Number_of_Arrests' which calculates the number of arrests per person using PersonID. We also eradicated all duplicate entries and null entries present in the dataset. Since we have a high number of entries in the dataset, we did not go for data substitution.

3.2 ANOVA

To investigate the remaining three research questions, RQ3, RQ4 and RQ5, we conducted three ANOVAs, including two one-way ANOVAs and one two-way ANOVA.

3.2.1 One-way ANOVA #1

The research question investigated in the first one-way ANOVA is whether there is a difference in the number of arrests based on a person's perceived race (White, Black or Latino) (RQ3). The person's perceived race is the independent variable with three categorical and independent groups (White, Black or Latino), and the number of arrests is the dependent, continuous variable. We noticed that the numbers of persons in each racial group are not equal. To ensure a balanced design, we need to make sure the sample sizes for each group are equal. Thus, we selected 1000 random samples from each group.

The null and alternative hypotheses for the first one-way ANOVA are as follows:

- Null hypothesis H_0: $\mu1 = \mu2 = \mu3$. The means of the number of arrests for the three perceived races are equal.
- Alternative hypothesis H_a: at least one perceived race's mean number of arrests is different from those of the other racial groups.

3.2.2 One-way ANOVA #2

The research question investigated in the second one-way ANOVA is whether there is a difference in the number of arrests based on a person's age group (RQ4). The person's age group is the independent variable with three categorical and independent groups (18 to 24 years, 25 to 34 years, or 35 to 44 years), and the number of arrests is the dependent, continuous variable. We noticed that the numbers of persons in each racial group are not equal. To ensure a balanced design, we need to make sure the sample sizes for each group are equal. Thus, we selected 5000 random samples from each age group.

The null and alternative hypotheses for the second one-way ANOVA are as follows:

- Null hypothesis H_0: μ1 = μ2 = μ3. The means of the number of arrests for the three age groups are equal.
- Alternative hypothesis H_a: at least one age group's mean number of arrests is different from those of the other age groups.

3.2.3 Two-way ANOVA

The research question investigated in the two-way ANOVA is whether there is a difference in the number of arrests based on the person's perceived race and age group, and whether an interaction effect exists between perceived race and age group (RQ5). The first independent variable is the person's perceived race with three categorical and independent groups (White, Black or Latino), and the second independent variable is the person's age group with three categorical and independent groups (18 to 24 years, 25 to 34 years, or 35 to 44 years).

The three sets of hypotheses for the two-way ANOVA are stated as follows:

- Set 1
  - Null hypothesis H_0: $\mu_a 1 = \mu_a 2 = \mu_a 3$. The means of the number of arrests for the three perceived races are equal.
  - Alternative hypothesis H_a: at least one perceived race's mean number of arrests is different from those of the other racial groups.
- Set 2
  - Null hypothesis H_0: $\mu_\beta 1 = \mu_\beta 2 = \mu_\beta 3$. The means of the number of arrests for the three age groups are equal.
  - Alternative hypothesis H_a: at least one age group's mean number of arrests is different from those of the other age groups.
- Set 3

- H_0: The effect of one independent variable does not depend on the effect of the other independent variable, i.e. there is no interaction between factor A (perceived race) and factor B (age group).
- H_a: There is an interaction between factor A (perceived race) and factor B (age group).

3.3 Post hoc tests

Because ANOVAs are only able to show whether one of the group's means differ from those of other groups, Post hoc tests, specifically, Tukey's HSD were conducted on all three ANOVAs to show which groups' means differed from those of other groups. The results of the Tukey's tests are reported in the next section.

4. Results

4.1 T-tests

```
[22] #conduct the 1st t-test
     rp.ttest(group1= df_ttest1['Number_of_Arrests'][df_ttest1['Sex'] == 'M'], group1_name= "Male",
              group2= df_ttest1['Number_of_Arrests'][df_ttest1['Sex'] == 'F'], group2_name= "Female")
```

```
(   Variable        N      Mean        SD        SE  95% Conf.  Interval
 0      Male  51857.0  3.860867  4.502916  0.019774   3.822111  3.899624
 1    Female  12509.0  3.777680  4.710832  0.042120   3.695119  3.860241
 2  combined  64366.0  3.844701  4.544149  0.017911   3.809595  3.879807,
                Independent t-test     results
 0  Difference (Male - Female) =        0.0832
 1          Degrees of freedom =    64364.0000
 2                           t =        1.8378
 3       Two side test p value =        0.0661
 4      Difference < 0 p value =        0.9670
 5      Difference > 0 p value =        0.0330
 6                   Cohen's d =        0.0183
 7                   Hedge's g =        0.0183
 8               Glass's delta1 =        0.0185
 9          Point-Biserial r =        0.0072)
```

Figure 4.1.1

For t-test #1, an independent samples t-test was run to determine if there were differences in the number of arrests between males and females. The mean number of arrests of males (M = 3.83, SD = 4.50) was not significantly different from that of females (M = 3.81, SD = 4.73), t(19998) = 0.28, p = 0.78 > 0.05.

Thus, we fail to reject the null hypothesis and conclude that the number of arrests does not differ significantly by sex.

```
[26] #conduct the 2nd t-test
     rp.ttest(group1= df_ttest2['Number_of_Arrests'][df['Youth_at_arrest__under_18_years'] == 'Not a youth'], group1_name= "Not a Youth",
              group2= df_ttest2['Number_of_Arrests'][df['Youth_at_arrest__under_18_years'] != 'Not a youth'], group2_name= "Youth")

(      Variable        N      Mean        SD        SE  95% Conf.  Interval
 0  Not a Youth  61343.0  3.906575  4.598286  0.018566   3.870186  3.942964
 1        Youth   3023.0  2.589150  2.992475  0.054427   2.482433  2.695867
 2     combined  64366.0  3.844701  4.544149  0.017911   3.809595  3.879807,
                         Independent t-test     results
 0  Difference (Not a Youth - Youth) =       1.3174
 1                Degrees of freedom =   64364.0000
 2                                 t =      15.5906
 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 =       0.2905
 7                         Hedge's g =       0.2905
 8                    Glass's delta1 =       0.2865
 9                  Point-Biserial r =       0.0613)
```

Figure 4.1.2

For t-test #2, an independent samples t-test was run to determine if there were differences in the number of arrests between non-youths and youths. The number of arrests of non-youths (M = 3.91, SD = 4.66) was higher than that of youths (M = 2.60, SD = 3.01), a statistically significant difference, t(19998) = 23.68, p = 0.00 < 0.05.

Thus, we can reject the null hypothesis and conclude that the number of arrests differ significantly by youth status.

4.2 One-way ANOVA

4.2.1 One-way ANOVA #1

```
[73] from scipy.stats import f_oneway

     # One-way ANOVA for arrests grouped by race
     statistic, pvalue = f_oneway(arrests_w_anova, arrests_b_anova, arrests_l_anova)

     print(f'One-way ANOVA: s = {statistic}, p = {pvalue}')

     One-way ANOVA: s = 33.98759890983767, p = 2.536751930066254e-15
```

Figure 4.2.1

From the results of the first one-way ANOVA, we can see that the p value is much smaller than 0.05. Therefore, we can reject the null hypothesis. We can conclude that there are significant differences in the means of the number of arrests of people from different perceived races.

4.2.2 One-way ANOVA #2

```
[93] # One-way ANOVA for arrests grouped by age group

     statistic, pvalue = f_oneway(age_35_44_anova, age_25_34_anova, age_18_24_anova)

     print(f'One-way ANOVA: s = {statistic}, p = {pvalue}')

     One-way ANOVA: s = 149.72580797615132, p = 4.126534831566302e-65
```

Figure 4.2.2

From the results of the second one-way ANOVA, we can see that the p value is much smaller than 0.05. Therefore, we can reject the null hypothesis. We can conclude that there are significant differences in the means of the number of arrests of people from different age groups.

4.3 Two-way ANOVA

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	9103.933969	2.0	197.753402	4.189182e-86
C(Age_group__at_arrest_)	7690.078205	2.0	167.041977	6.540313e-73
C(Perceived_Race):C(Age_group__at_arrest_)	1349.898219	4.0	14.661078	5.724755e-12
Residual	767548.559662	33345.0	NaN	NaN

Figure 4.3.1

From the results of the two-way ANOVA, we can see that the p-value for both perceived race and age group at arrest is much smaller than 0.05. Thus, we can reject the null hypotheses in both set 1 and set 2 and conclude that there are significant differences in the means of the number of arrests of people from different age groups or different perceived races. The p-value for the interaction effect is also smaller than 0.05. Thus, we reject the null hypothesis in set 3 and conclude that there is significant interaction effect between perceived race and age group.

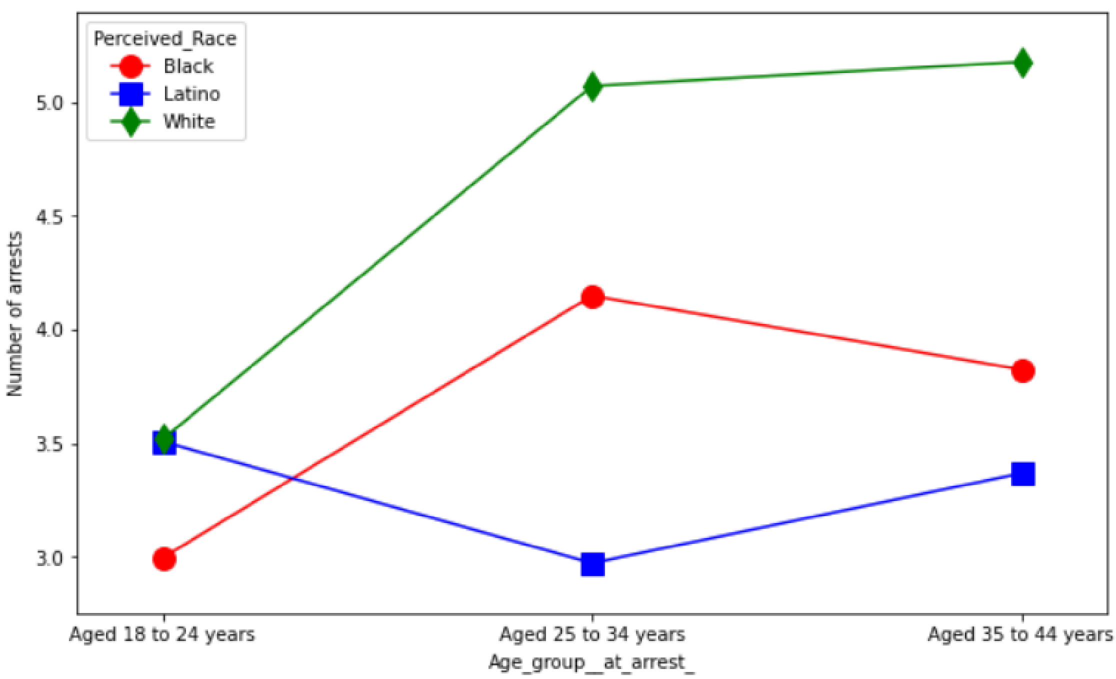

Figure 4.3.2 Interaction Plot

From the interaction plot, we can see the interactive effect of age group and race on a sample population. Since the lines are not parallel, we can say that there is a cumulative effect of both age and race on the number of arrests.

The mean number of arrests for age group {18,24} is approximately 3250.

The mean number of arrests for age group {25,34} is approximately 3900.

The mean number of arrests for age group {35,44} is approximately 4100.

The green trend line depicts the number of arrests made on White individuals. We see that the number of arrests sharply increases as the age group changes from {18,24} to {25, 34}. The increase in the number of crimes is not that sharp for the next age group.

The red trend line depicts the number of arrests made on Black individuals. We see that the number of arrests sharply increases as the age group changes from {18,24} to {25, 34}. The number of arrests declines steadily for the next age group.

The blue trend line depicts the number of arrests made on Latino individuals. We see that the number of arrests decreases as the age group changes from {18,24} to {25, 34}. The number of arrests increases steadily for the next age group.

4.4 Tukey's HSD Tests

4.4.1 Tukey's HSD Test for One-way ANOVA #1

```
[ ] #Tukey test for one-way anova #1
    tukey1 = pairwise_tukeyhsd( df_race_age['Number_of_Arrests'],df_race_age['Perceived_Race'],  alpha=0.05)
    print(tukey1.summary())
```

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj  lower   upper  reject
---------------------------------------------------
 Black Latino  -0.5231 0.001 -0.8432 -0.2029   True
 Black  White   1.1527 0.001  1.0247  1.2808   True
Latino  White   1.6758 0.001  1.3596  1.9919   True
---------------------------------------------------
```

Figure 4.4.1

From the results of the Tukey's test, we observe that

- P-value for the difference in means between Black and Latino: 0.001
- P-value for the difference in means between Black and White: 0.001
- P-value for the difference in means between Latino and White: 0.001

Thus, we would conclude that there is a statistically significant difference between the means of arrest numbers of Black persons and Latino persons, between the means of arrest numbers of Black persons and white persons, and between the means of arrest numbers of Latino persons and White persons.

4.4.2 Tukey's HSD Test for One-way ANOVA #2

```
[ ] #Tukey test for one-way anova #2
    tukey2 = pairwise_tukeyhsd( df_race_age['Number_of_Arrests'],df_race_age['Age_group__at_arrest_'],  alpha=0.05)
    print(tukey2.summary())
```

```
             Multiple Comparison of Means - Tukey HSD, FWER=0.05
=============================================================================
      group1              group2       meandiff p-adj   lower  upper  reject
-----------------------------------------------------------------------------
Aged 18 to 24 years Aged 25 to 34 years   1.3799  0.001  1.2133 1.5466   True
Aged 18 to 24 years Aged 35 to 44 years   1.4718  0.001  1.2981 1.6455   True
Aged 25 to 34 years Aged 35 to 44 years   0.0919 0.2716 -0.0478 0.2316  False
-----------------------------------------------------------------------------
```

Figure 4.4.2

From the results of the Tukey's test, we observe that

- P-value for the difference in means between age group 18 to 24 years and age group 25 to 34 years: 0.001
- P-value for the difference in means between age group 18 to 24 years and age group 35 to 44 years: 0.001

- P-value for the difference in means between age group 25 to 34 years and age group 35 to 44 years: 0.272

Thus, we would conclude that there is a statistically significant difference between the means of arrest numbers of persons aged 18 to 24 and persons aged 25 to 34, between the means of arrest numbers of persons aged 18 to 24 and persons aged 35 to 44, but not a statistically significant difference between the means of arrest numbers of persons aged 25 to 34 and persons aged 35 to 44.

4.4.3 Tukey's HSD Test for Two-way ANOVA

```
[ ] #tukey test for 2 way anova
    mc = MultiComparison(df_race_age['Number_of_Arrests'], df_race_age['Age_group__at_arrest_'].astype(str) + ',' +  df_race_age['Perceived_Race'].astype(str))
    tukey_anova2way = mc.tukeyhsd()
    print(tukey_anova2way)
```

```
                   Multiple Comparison of Means - Tukey HSD, FWER=0.05
============================================================================================
          group1                     group2           meandiff p-adj   lower   upper  reject
--------------------------------------------------------------------------------------------
 Aged 18 to 24 years,Black Aged 18 to 24 years,Latino   0.5119 0.6618  -0.373  1.3967  False
 Aged 18 to 24 years,Black  Aged 18 to 24 years,White   0.5218  0.001  0.1426   0.901   True
 Aged 18 to 24 years,Black  Aged 25 to 34 years,Black   1.1533  0.001  0.8452  1.4615   True
 Aged 18 to 24 years,Black Aged 25 to 34 years,Latino  -0.0236    0.9  -0.668  0.6208  False
 Aged 18 to 24 years,Black  Aged 25 to 34 years,White    2.076  0.001  1.7824  2.3696   True
 Aged 18 to 24 years,Black  Aged 35 to 44 years,Black   0.8281  0.001  0.4799  1.1763   True
 Aged 18 to 24 years,Black Aged 35 to 44 years,Latino   0.3722 0.8099 -0.3725   1.117  False
 Aged 18 to 24 years,Black  Aged 35 to 44 years,White   2.1817  0.001  1.8852  2.4783   True
Aged 18 to 24 years,Latino  Aged 18 to 24 years,White   0.0099    0.9 -0.8892  0.9091  False
Aged 18 to 24 years,Latino  Aged 25 to 34 years,Black   0.6415 0.3533 -0.2301   1.513  False
Aged 18 to 24 years,Latino Aged 25 to 34 years,Latino  -0.5355 0.7807 -1.5747  0.5037  False
Aged 18 to 24 years,Latino  Aged 25 to 34 years,White   1.5641  0.001  0.6976  2.4306   True
Aged 18 to 24 years,Latino  Aged 35 to 44 years,Black   0.3162    0.9 -0.5703  1.2028  False
Aged 18 to 24 years,Latino Aged 35 to 44 years,Latino  -0.1396    0.9 -1.2439  0.9646  False
Aged 18 to 24 years,Latino  Aged 35 to 44 years,White   1.6698  0.001  0.8023  2.5374   True
 Aged 18 to 24 years,White  Aged 25 to 34 years,Black   0.6315  0.001  0.2844  0.9787   True
 Aged 18 to 24 years,White Aged 25 to 34 years,Latino  -0.5454 0.2091 -1.2093  0.1185  False
 Aged 18 to 24 years,White  Aged 25 to 34 years,White   1.5542  0.001  1.2199  1.8885   True
 Aged 18 to 24 years,White  Aged 35 to 44 years,Black   0.3063 0.2413 -0.0768  0.6894  False
 Aged 18 to 24 years,White Aged 35 to 44 years,Latino  -0.1495    0.9 -0.9112  0.6121  False
 Aged 18 to 24 years,White  Aged 35 to 44 years,White   1.6599  0.001  1.3231  1.9968   True
 Aged 25 to 34 years,Black Aged 25 to 34 years,Latino  -1.1769  0.001 -1.8029 -0.5509   True
 Aged 25 to 34 years,Black  Aged 25 to 34 years,White   0.9227  0.001  0.6719  1.1735   True
 Aged 25 to 34 years,Black  Aged 35 to 44 years,Black  -0.3252 0.0345 -0.6382 -0.0123   True
 Aged 25 to 34 years,Black Aged 35 to 44 years,Latino  -0.7811  0.025   -1.51 -0.0522   True
 Aged 25 to 34 years,Black  Aged 35 to 44 years,White   1.0284  0.001  0.7742  1.2826   True
Aged 25 to 34 years,Latino  Aged 25 to 34 years,White   2.0996  0.001  1.4806  2.7186   True
Aged 25 to 34 years,Latino  Aged 35 to 44 years,Black   0.8517 0.0015   0.205  1.4984   True
Aged 25 to 34 years,Latino Aged 35 to 44 years,Latino   0.3958    0.9  -0.527  1.3186  False
Aged 25 to 34 years,Latino  Aged 35 to 44 years,White   2.2053  0.001  1.5849  2.8257   True
 Aged 25 to 34 years,White  Aged 35 to 44 years,Black  -1.2479  0.001 -1.5465 -0.9492   True
 Aged 25 to 34 years,White Aged 35 to 44 years,Latino  -1.7038  0.001 -2.4266 -0.9809   True
 Aged 25 to 34 years,White  Aged 35 to 44 years,White   0.1057    0.9 -0.1307  0.3421  False
 Aged 35 to 44 years,Black Aged 35 to 44 years,Latino  -0.4559 0.6015 -1.2026  0.2908  False
 Aged 35 to 44 years,Black  Aged 35 to 44 years,White   1.3536  0.001  1.0521  1.6551   True
Aged 35 to 44 years,Latino  Aged 35 to 44 years,White   1.8095  0.001  1.0854  2.5336   True
--------------------------------------------------------------------------------------------
```

Figure 4.4.3

From the results of the Tukey's test, we can see that there are significant differences ($p < 0.05$) between:

- Aged 18 to 24 years, Black and Aged 18 to 24 years, White,
- Aged 18 to 24 years, Black and Aged 25 to 34 years, Black,
- Aged 18 to 24 years, Black and Aged 25 to 34 years, White,
- Aged 18 to 24 years, Black and Aged 35 to 44 years, Black,
- Aged 18 to 24 years, Black and Aged 35 to 44 years, White,
- Aged 18 to 24 years, Latino and Aged 25 to 34 years, White,
- Aged 18 to 24 years, Latino and Aged 35 to 44 years, White,
- Aged 18 to 24 years, White and Aged 25 to 34 years, Black,
- Aged 18 to 24 years, White and Aged 25 to 34 years, White,
- Aged 25 to 34 years, Black and Aged 25 to 34 years, Latino,
- Aged 25 to 34 years, Black and Aged 25 to 34 years, White,
- Aged 25 to 34 years, Black and Aged 35 to 44 years, Black,
- Aged 25 to 34 years, Black and Aged 35 to 44 years, Latino,
- Aged 25 to 34 years, Black and Aged 35 to 44 years, White,
- Aged 25 to 34 years, Latino and Aged 25 to 34 years, White,
- Aged 25 to 34 years, Latino and Aged 35 to 44 years, Black,
- Aged 25 to 34 years, Latino and Aged 35 to 44 years, White,
- Aged 25 to 34 years, White and Aged 35 to 44 years, Black,
- Aged 25 to 34 years, White and Aged 35 to 44 years, Latino,
- Aged 35 to 44 years,Black and Aged 35 to 44 years,White,
- Aged 35 to 44 years,Latino and Aged 35 to 44 years,White

There are no statistically significant differences between the rest of the age and race groups.

5. Discussion

From the results of the exploratory data analysis, we were able to find out the susceptibly most impactful subgroups from the parameters we wanted to study and shortlisted them as a part of our tests. The chosen independent variables for this study were race, age and sex which were proven to be statistically significant using t tests. From the visualizations, we could say that black, young men were more likely to be arrested for a crime. To study these findings using statistical methods, we found out that there are significant differences in mean number of arrests between levels of the various independent variables. From there we went on to study two way anovas to compare the means of more than 2 levels in the independent variables, namely . To further find out the larger value in between means, we did a tukey test on both one way and two anovas. We found out that among the 3 age group levels selected, one was not significant (35 to 44). From the tukey test for race, we can say that all levels (i.e latino, white and black) have statistcial significance. Later on when we analyse the interaction plot of the two-way anova, we can see that the number of crimes committed are the highest by white people over the age of 35.

To fulfill the research objective of finding out the relationship between the number of arrests and demographic factors such as gender, youth status, race, and age, we first started with the t-tests, which examine gender and youth status. From the results of the first t-test, we were able to answer RQ1 by stating that there is no significant difference in the number of arrests between males and females; males and females are just as likely to be arrested. From the second t-test, we were able to answer RQ2 that non-youths have a significantly higher mean number of arrests than youths, that is, those who are over 17 years of age are more likely to be arrested than those under 17. We then moved on to examine the effects of race and age group on the number of arrests through two one-way ANOVAs. The results from the one-way ANOVAs provide answers for RQ3 and RQ4 by showing that the mean number of arrests differs significantly between the race and age groups that we selected. The results from our two-way ANOVA answers RQ5 and reinforces our earlier findings that the number of arrests differ significantly by race and age group.

To obtain the finer results that show which race and age groups experience more arrests, we resort to Tukey's tests. Tukey's test for the first ANOVA revealed that black and Latino people are more likely to be targets of arrests than white people. Meanwhile, black people have a slightly less likelihood to be arrested when compared to Latino people. This shows that out of the three racial groups that we chose, Latinos have the highest likelihood to be arrested. Tukey's test for the second ANOVA revealed that those aged between 18 and 24 years are more likely to be arrested than the other two age groups. For the remaining two age groups, those aged 25 to 34 years are more likely to be arrested than those aged 35 to 44 years. Thus, we can infer that within an age range of 18 to 44 years, the younger one is, the more likely they are subjected to an arrest. Lastly, Tukey's test for the two-way ANOVA indicated that the most notable differences are between Latino people aged 25 to 34 and white people aged 35 to 44, between black people aged 18 to 24 and white people aged 35 to 44, and between Latino people aged 25 to 34 and white people aged 25 to 34. These results correspond to the trend we observed from the first two Tukey's tests that being younger and being black or Latino makes one a more likely target of arrests than being older and being white. If we combine these findings with the earlier finding from the second t-test, which showed non-youths have a higher mean number of arrests than youths, we can infer that the age group that is most likely to be arrested is between 18 and 24 years.

This study is limited in a few ways. Firstly, we were only able to base our study on the dataset itself. We were not able to ascertain if the dataset includes all arrests made by the Toronto Police Service in the years 2020 and 2021, thus, our findings are limited to what the dataset includes and cannot be generalized for the City of Toronto for these two years. Secondly, for all three ANOVA tests, we only included three racial groups, white, black, and Latino, and three age groups, 18 to 24, 25 to 34, and 35 to 44. A more comprehensive analysis of the dataset should include more age groups. Thirdly, we could improve on how we report on the results for the Tukey's test for our two-way ANOVA. Our discussion of the results were hampered by the fact that there are too many combinations of race and age groups. A different way of reporting and summarizing the results might better help us to track the most noteworthy trends.

6. Conclusion

While data-driven analysis pedagogy can help us predict the likelihood of arrest of a person,it can also create biases in people's minds and lead to unwarranted arrests by the police. But they also help us realise that only a subsegment of numbers being referred to without doing an analysis of other valid factors can lead to the wrong answers. This could be easily seen from our study. In the beginning of the study, from the data obtained from the number of arrests grouped by demographic factors, we could see that the number of arrests were higher for males compared to females, blacks compared to other races and young people compared to other age groups. Later, on doing a 2- way anova, we could see a strong interaction between the age and race parameters which further led to completely different findings. We found that older white people had a higher number of arrests as compared to the rest of the categories. Therefore, it is important to use the right tools and methods to draw conclusions and provide the right predictions.

This study aimed to find out the effects of several demographic factors, such as gender, youth status, race, and age, on a person's number of arrests. Our five research questions and their corresponding hypotheses were tested by t-tests, one-way and two-way ANOVAs, and Tukey's Tests. Among the demographic factors that we chose, we revealed that youth status, race, and age had significant effects on the number of arrests, while sex did not. Through a finer look enabled by Tukey's tests, we were able to observe a general trend that being in the youngest age group of 18 to 24 years and being Latino or Black makes one most likely to be the target of an arrest. However, being younger does not make one more likely to get arrested across all age groups. A t-test revealed that youths younger than 17 years are less likely to be arrested than those over 17 years. Combining the observation from Tukey's tests with our earlier findings, we were able to infer that being of Latino heritage and being between 18 and 24 years of age makes one the most likely target of being arrested by the Toronto Police Service (TPS). Further research can be conducted on datasets on arrests made by the

TPS from other years or made by police forces in other Canadian cities to examine whether similar trends exist.

References

- *The usual suspects: police stop and search practices in Canada* by Scot Wortley and Akwasi Owusu-Bempah (https://www.researchgate.net/publication/238046161_The_Usual_Suspects_Police_Stop_and_Search_Practices_in_Canada)
- *Environmental Criminology and Crime Analysis: Situating the Theory, Analytic Approach and Application* by Richard Wortley and Lorraine Mazerolle (https://www.researchgate.net/publication/29468939_Environmental_Criminology_and_Crime_Analysis_Situating_the_Theory_Analytic_Approach_and_Application)
- *Annual Statistical Report* by the Toronto Police 2020 https://data.torontopolice.on.ca/pages/asr-archives
- *Examining the relationship between police experiences and perceptions of police bias by* Brian C. Renauer, Emma Covelli (2011) .
- *Domestic Violence Prevention and Reduction in British Columbia (2000-2010)* by Katherine R. Rossiter (https://www.jibc.ca/sites/default/files/research/pdf/Domestic-violence-prevention-reduction_REPORT.pdf)