Arrests and Strip Searches in Toronto:

A Case Study on Age and Race

Midterm Report

Jean Paul Ngezigihe

Jian Zhang

INF2178H S LEC0101: Experimental Design

Professor Shion Guha

February 28th, 2023

Introduction

"Stop and frisk" is a police practice of temporarily detaining, questioning, and at times searching civilians and suspects on the street for weapons. It became a subject of racial profiling criticism as ninety percent of those stopped in 2017 were African-American or Latino, mostly aged 14–24 (New York Civil Liberties Union, 2019). Seventy percent of those stopped were later found to be innocent (New York Civil Liberties Union, 2019). A similar method also appeared in Canada, specifically in Ontario, as "Carding". Local police enforcement and city officials have largely used carding as a way to "protect the streets" and provide a false sense of security to win votes while negating its psychological impacts on victims (Johnson & al., 2021). It was officially discontinued in the summer of 2014. However, some argue the effect of this colonial method is still pervasive today in Toronto Police Service. We therefore ask whether it is the case. In this report, we will investigate the effect of demographic attributes on the Toronto Police arrests and strip searches, using the publicly available dataset "Arrests and Strip Searches".

Our first research question examines: "*Is there a significant difference in strip search based on perceived race*?"
The second research question examines: "*Is there a significant difference in strip search based on perceived race and age?"*

Literature review

According to the Toronto Police Service, a strip search refers to a search conducted by a police officer on a person, which includes the removal of some or all clothing and a visual inspection of the body. Scholars from the International Journal for Crime, Justice and Social Democracy describe strip searches as a police misuse of power with potential harm on individuals and the society. While scholars agree on strip searches being abusive, others have questioned why governments are funding police departments more than ever (Spurrier, M. (2023). Police departments are now increasingly turning to algorithms to respond to such criticisms (Joh, E. E. (2017). By using artificial intelligence to decide which neighborhoods to increase efforts.
, they argue their practices to be fair and non-discriminatory. However, such practices still find themselves rooted in colonial practices (Barabas & al., 2020). Users of algorithmic tools for

decision making need to develop more robust frameworks for understanding their work and inquire their own perspective and position of power (Barabas & al., 2020). This report aims to use a studying up lenses while investigating our research questions,

Exploratory Data Analysis

Descriptive statistics

Using df.head(), we see the dataset consists of 25 columns with information on individual characteristics such as sex, age group as well as information about the arrest; action at arrest, whether items were found, etc. More information on the dataset in the following section. Furthermore, we identified there were 65276 entries in this dataset and obtained information about the column data types and counts. See Table 1 below for first 10 variables:

Table 1: Data type and count for first 10 variables

#	Column	Non-Null	Count	Datatype
0	Arrest_Year	65276	Non-Null	int64
1	Arrest_Month	65276	Non-Null	object
2	EventID	65276	Non-Null	int64
3	ArrestID	64807	Non-Null	object
4	PersonID	65276	Non-Null	int64
5	Perceived_Race	65276	Non-Null	object
6	Sex	65276	Non-Null	object
7	Age_group_at_arrest	65276	Non-Null	object
8	Youth_at_arrest	65276	Non-Null	object
9	ArrestLocDiv	65276	Non-Null	int64
10	StripSearch	65276	Non-Null	int64

This table informs our method choice as we know the datatypes of our variables we choose to investigate.

Next, we performed a descriptive summary of our numeric variables

Table 2: Descriptive statistics of some of the numeric variables in the dataset

index	StripSearch	Booked	Actions_at_arrest___ Concealed_i	Actions_at_arrest__ _Combative__	ItemsFound
count	65276.0	65276.0	65276.0	65276.0	7801.0
mean	0.11950793553	0.519501807	0.004075004595869845	0.044135670077823395	0.37315728752
std	0.3243877323	0.49962336	0.0637060523598573	0.2053980501659194	0.4836743893
min	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0
50%	0.0	1.0	0.0	0.0	0.0
75%	0.0	1.0	0.0	0.0	1.0
max	1.0	1.0	1.0	1.0	1.0

From this table, we see many of our numeric variables range from 0 to 1. Therefore, descriptive statistics such as minimum or maximum do not provide much information.

Preprocessing

From this initial exploratory analysis, we realized there were many variables we did not need. Therefore, we preprocessed the data to remove unnecessary variables. Additionally, our dependent variable of interest is strip search. However, it is non-continuous. Therefore, we performed some preprocessing to make it a continuous variable and perform our General Linear Model (GLM). We then create different tables for analysis. See below examples for strip search and booking.

Table 3: Strip search by race and group

```
print(strip)
```

```
     Arrest_Year Perceived_Race      Age_group__at_arrest_  StripSearch
0           2020          Black  Aged 17 years and younger          139
1           2020          Black        Aged 18 to 24 years          552
2           2020          Black        Aged 25 to 34 years          774
3           2020          Black        Aged 35 to 44 years          405
4           2020          Black        Aged 45 to 54 years          168
..           ...            ...                        ...          ...
107         2021          White        Aged 25 to 34 years          101
108         2021          White        Aged 35 to 44 years          126
109         2021          White        Aged 45 to 54 years           31
110         2021          White        Aged 55 to 64 years           12
111         2021          White    Aged 65 years and older            3

[112 rows x 4 columns]
```

Table 4: Booked by race and age group

```
     Arrest_Year Perceived_Race      Age_group__at_arrest_  Booked
0           2020          Black  Aged 17 years and younger     307
1           2020          Black        Aged 18 to 24 years    1060
2           2020          Black        Aged 25 to 34 years    1670
3           2020          Black        Aged 35 to 44 years     896
4           2020          Black        Aged 45 to 54 years     403
..           ...            ...                        ...     ...
107         2021          White        Aged 25 to 34 years    1780
108         2021          White        Aged 35 to 44 years    1892
109         2021          White        Aged 45 to 54 years    1109
110         2021          White        Aged 55 to 64 years     514
111         2021          White    Aged 65 years and older     111

[112 rows x 4 columns]
```

Then, we combined both tables to form our data frame “Strip search and Booked by race and age group” which we used for more exploratory data analysis.

Table 5: Dataframe combining Strip search and Booked by race and age group

```
    Arrest_Year Perceived_Race      Age_group__at_arrest_  StripSearch  \
0          2020          Black  Aged 17 years and younger          139
1          2020          Black        Aged 18 to 24 years          552
2          2020          Black        Aged 25 to 34 years          774
3          2020          Black        Aged 35 to 44 years          405
4          2020          Black        Aged 45 to 54 years          168
..          ...            ...                        ...          ...
107        2021          White        Aged 25 to 34 years          101
108        2021          White        Aged 35 to 44 years          126
109        2021          White        Aged 45 to 54 years           31
110        2021          White        Aged 55 to 64 years           12
111        2021          White    Aged 65 years and older            3

     Booked
0       307
1      1060
2      1670
3       896
4       403
..      ...
107    1780
108    1892
109    1109
110     514
111     111

[112 rows x 5 columns]
```

Figure 1: Missing or duplicated values

Figur

```
6] # Determine if there are any missing datapoints or duplicate rows in the dataset
   print("\nAre there any missing datapoints in the dataset?", df.isnull().values.any())
   print("Number of duplicated rows:", df.duplicated().sum())
```

```
Are there any missing datapoints in the dataset? False
Number of duplicated rows: 0
```

```
7] #check missing values
   df.isna().sum()
```

```
Arrest_Year             0
Perceived_Race          0
Age_group__at_arrest_   0
StripSearch             0
Booked                  0
dtype: int64
```

We determine if there are any missing data points or duplicate rows in our dataset. There were none. Then, verified for missing values. There were none as well.

Data Visualization

We identified which variables had a higher correlation with each other.

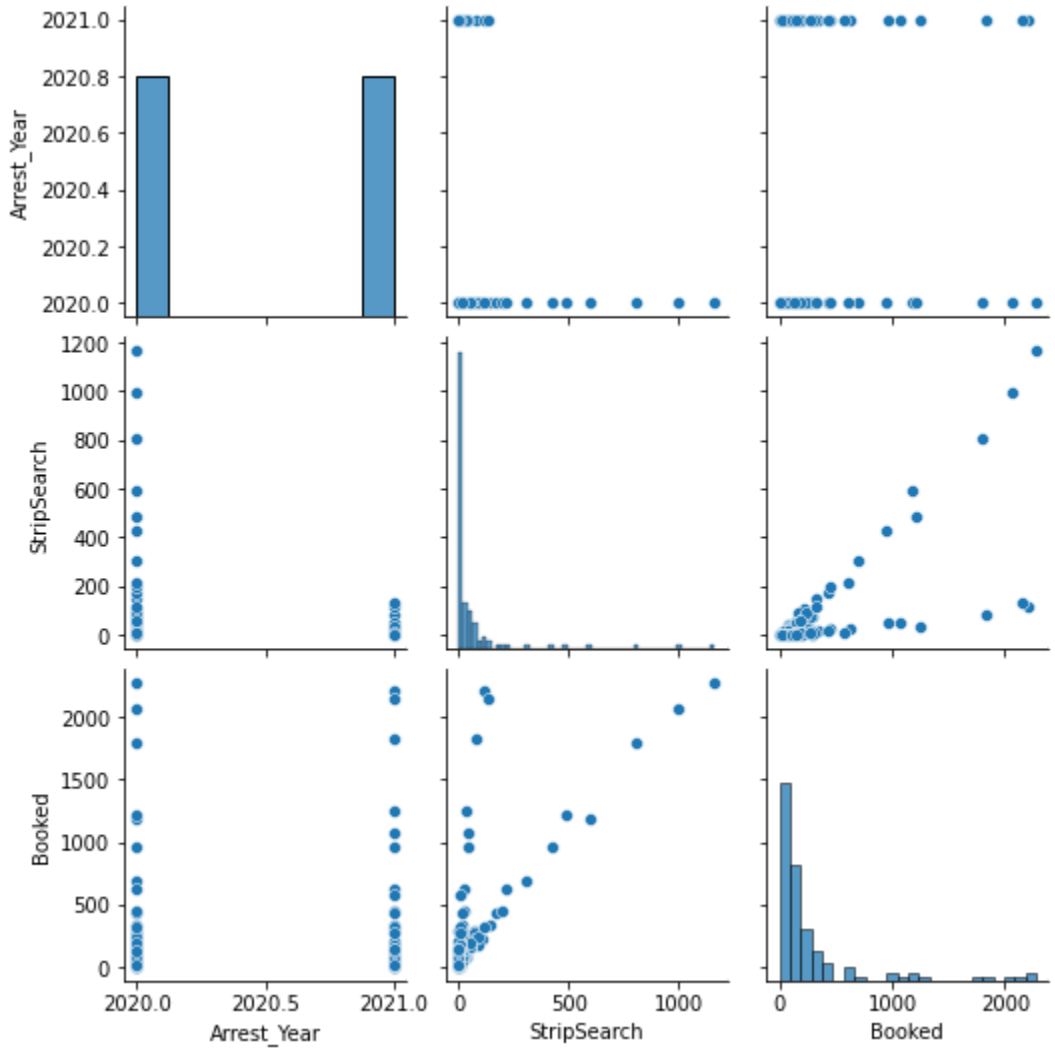
Figure 2: HeatMap

Starting with the first variable from top left we see Arrest year has a correlation value of -0.32 and -0.0051 with StripSearch or Booked. This indicates no correlation with these variables. This informs us that police officers' arrests and bookings are not impacted by the year.

StripSearch has a correlation value of -0.32 and 0.73 with Arrest_year and Booked. This indicates no correlation with Arrest_Year and a positive correlation with Booked. This informs us that police strip searches are correlated to being subsequently booked. However, we may not make inference here since this is not causation and more analysis is required.

Booked has a correlation value of -0.0051 and 0.73 with Arrest_year and StripSearch. This indicates no correlation with Arrest_Year and a positive correlation with StripSearch. This informs us that police bookings are correlated to having been striped searched before. Again, this information cannot tell us whether strip search causes bookings, but informs it could be a potential analysis step.

Figure 3: Pairplots between numeric variables



The pair plots allow us to graphically see the relationship between our numeric variables. There is no trend in the plots associated with Arrest_Year. In the Strip Search and Booked plots, we see the data points moving upward to the right. This displays there is a positive relationship between StripSearch and Booked. This confirms our previous analysis from the heat map and informs us we might need to investigate them further.

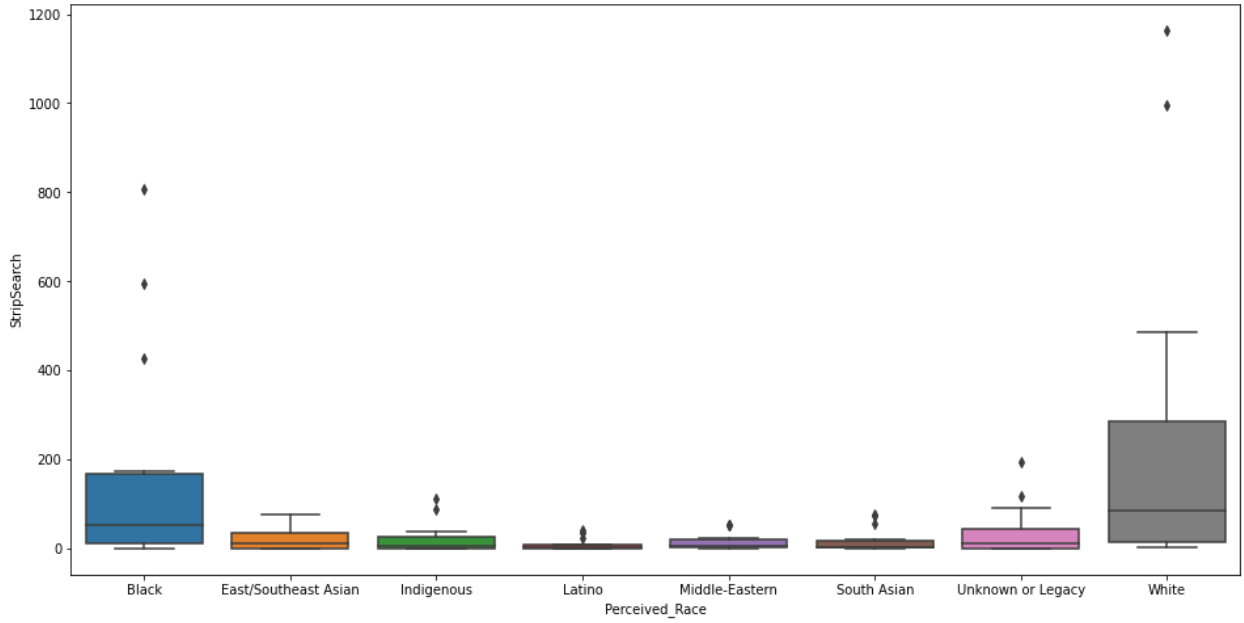
Box Plots

According to the Arrest and Strip Search (RBDC-ARR-TBL-001) data set, there may be some difference in the likelihood of being booked for ethnic groups that are not white compared to

white race. That is, racial groups that are non-white may have a greater likelihood of being booked compared to White racial groups.

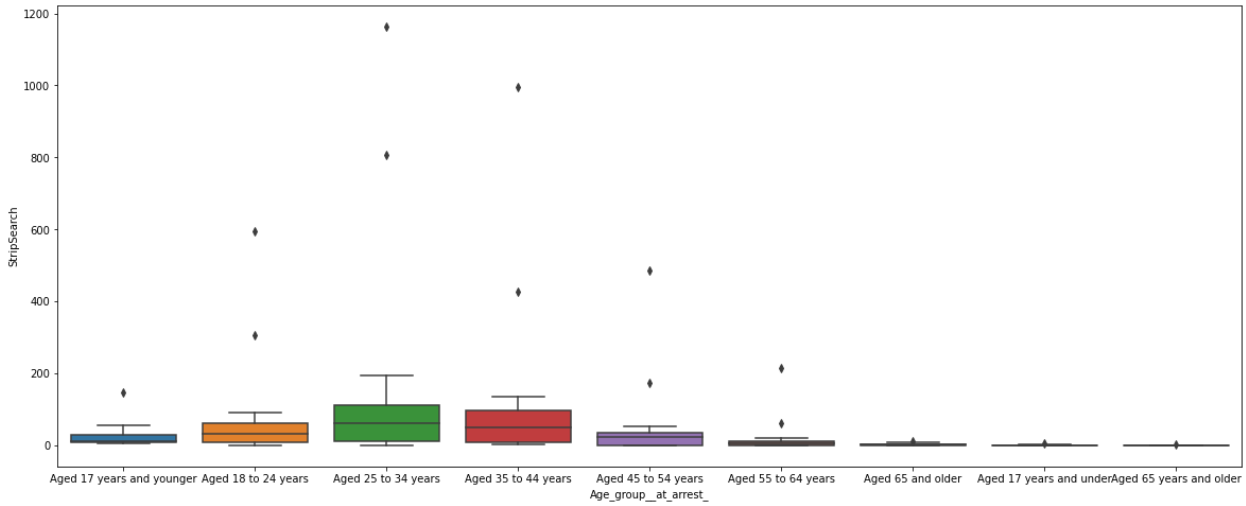
Based on the Arrest and Strip Search (RBDC-ARR-TBL-001) dataset, we found some variation in the likelihood of strip searches based on race and age. In detail, ethnic groups that are not white may have a higher probability of being strip searched compared to white race. Also, younger groups are more likely to be strip searched than older groups.

Figure 3: Distribution of Race in Strip Search Boxplot



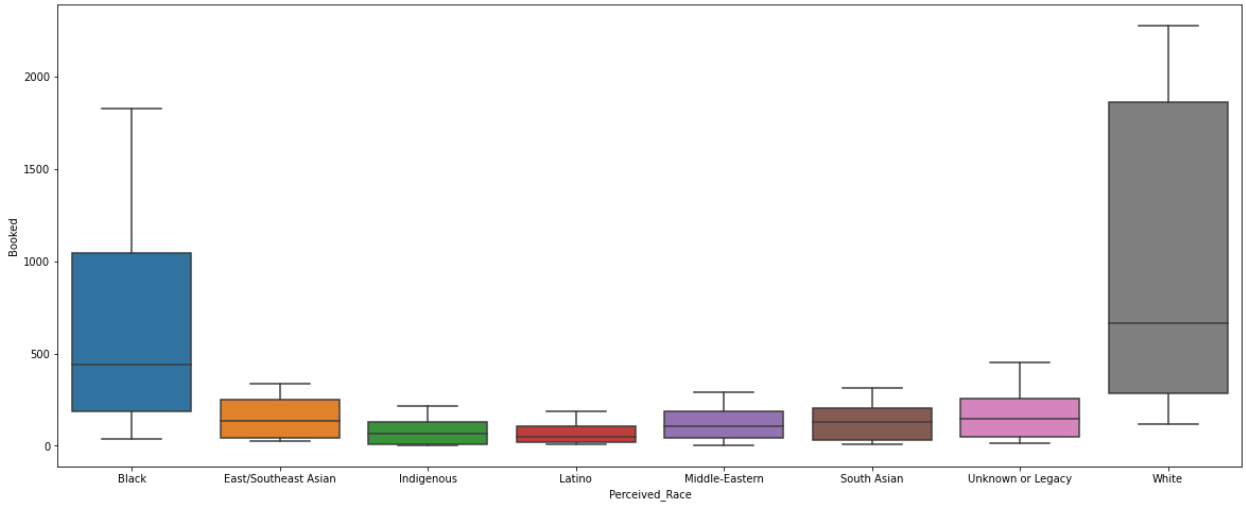
From the boxplot, we can observe the distribution of strip search results by perceived race, and we find that there are differences in the probability of being strip-searched by different perceived races. For example, the boxes for non-whites are higher and may have some outliers than those for whites, indicating that non-whites are more likely and more often strip-searched at the time of arrest than whites. This finding also implies that there may be some racial bias in the law enforcement practices of the officers.

Figure 4: Boxplot of Strip Searches by Age Group at Arrest



The boxplot expresses the distribution of strip searches of individuals by age group at the time of the arrest. We can see that the probability distribution is wider for the younger group compared to the older group, suggesting that the younger group has a relatively greater variation in the probability of being strip-searched, while the older group is likely to receive a relatively consistent search procedure. In addition, we can see boxes for the 18-24 and 25-34 age groups (with the 26-35 age group having the highest IQR) compared to the other age groups), which also suggests that these two age groups are more likely to be strip searched compared to the other age groups. Overall, there was some variability across age groups regarding the probability of being strip-searched, suggesting that age could be a factor in the decision to strip-search.

Figure 5: Boxplot of Booked by Perceived Race at Arrest



This Boxplot shows the variable "Booked" distribution across the different categories of "Perceived_Races". We can discover from the graph that the Black group has the highest median value in the arrested, so perceived race and their IQR are also larger, indicating a more significant variation in Book for this group. This boxplot also suggests that there may be some racial differences in the likelihood of being registered by the officers.

T-tests

From the EDA, we see a high proportion of strip search on perceived race as Black. We wonder if there is a significant difference in the means compared to the overall groups. Additionally, we see a high count of strip searches for young adults. We are curious to see if there is significant difference in the means compared to the overall groups. In this section, we will perform T-Tests to test these hypotheses.

Hypothesis for t-tests

Is there a significant difference in the means of Black compared to the overall population?

H0: There is no significant difference in the means of Black compared to the overall population
H1: There is a significant difference in the means of Black compared to the overall population

Is there a significant difference in the means of Indigenous compared to the overall population?

H0: There is no significant difference in the means of Indigenous compared to the overall population
H1: There is a significant difference in the means of Indigenous compared to the overall population

* The hypothesis for the other t-test would follow the same convention *

Assumptions check
As we will be using t-test, we need to check if it meets its assumptions.

1. **The assumption for a t-test follows a continuous or ordinal scale**. Yes, our data measurement strip search is continuous.

2. **The second assumption made is that of a simple random sample.** Yes, our sample collected a representative, randomly selected portion of the total population.
3. **The data is a normal distribution, bell-shaped distribution curve.**

Figure 6: Shapiro test result

```
#Normally distributed
from scipy.stats import shapiro
import scipy.stats as stats

normal = strip1['StripSearch']
shapiro(normal) #Results
```

```
ShapiroResult(statistic=0.29892998933792114, pvalue=0.0)
```

The p-value is 0.0 which is less than the alpha(0.05) then we reject the null hypothesis. We have sufficient evidence to say that the sample does not come from a normal distribution. Although this assumption is not met, we will continue our test and note it as a limitation.

4. **The data is a reasonably large sample size**. Yes, our data is reasonably large with above 500 data points.
5. **Data sample is independent.** Individuals arrested are independent of each other. Yes, our data is independent.

The assumptions are violated. However, we will still run the statistical test and note it as a limitation.

Table 6: One sample one-tail 'greater' T-Test results

Group	P Value	Significant/Not Significant
Black	0.003	Significant
Latino	1.0	Not Significant
Indigenous	1.0	Not Significant
Middle-Eastern	1.0	Not Significant

Figure 7: Confidence interval results

```
[12] # Confidence level
     alpha = 0.05  # 95% confidence
     # Degrees of freedom
     dof = sample_size - 1
     # Percent-point function (aka quantile function) of the t-distribution
     t = stats.t.ppf(1 - (alpha / 2), dof)
     # Margin of error
     d = t * sem
     # Intervals
     upper_ci = x_bar + d
     lower_ci = x_bar - d

     print(f'Sample mean = {x_bar:.2f}, 95% CI [{lower_ci:.2f}, {upper_ci:.2f}]')

     Sample mean = 22.13, 95% CI [12.93, 31.33]
```

Using a two-tailed t test, we see most groups are significantly different. We use a one-tail t-test, alternative 'greater' to determine if the mean of the group is greater than the population mean.

The results indicate that the mean strip search for the Black group is higher than the mean of the population (M=9.04). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.003) is less than 0.05, 95% CI [12.93, 31.33]. Therefore, we can reject the null hypothesis that there is no difference in strip search for the Black group and the population. However, it is important to note we do have a limitation on this test since it fails the normally distributed assumption

Method

Dataset description

The Arrest and Strip Search dataset (RBDC-ARR-TBL-001) collects demographic information on individuals arrested and strip-searched by relevant law enforcement agencies. Each row represents an arrest with demographic information and information about whether they were strip searched or booked at a police station within 24 hours. The dataset contains a mix of integer and string. The datatypes are mostly categorical such as age and arrest month with a few continuous columns. The dataset can be found here.

The goal of our method analysis is to demonstrate statistically whether arrestees who resulted in strip searches were influenced by their age. For example, are younger individuals perceived as more likely to engage in criminal activity, increasing the likelihood of being strip searched.
Another research question examines whether people of different racial backgrounds are more likely to be arrested and booked by the Toronto Police Service. Our methods provide us with a quantitative argument to answer our research question and potentially make other correlation claims on potential bias and discrimination in the Toronto Police department.

It is worth mentioning that the "perceived race" values in the arrest and strip search dataset (RBDC-ARR-TBL-001) are essentially based on the personal opinions of the officers involved in the arrest and strip search at the time. For example, the accuracy of perceived race is largely limited by the officers' biases and assumptions about race at the time and/or time of day. In other words, because of some bias of the officers at the time, it is likely that the actual race of the arrestee i not be inaccurate to the race registered may include for error

ANOVA Tests

One Way

So far, the EDA demonstrated some perceived groups such as Black are searched more often than other groups (Figure 3). Additionally, younger age groups seemed to be striped search more often (figure 4). T-Test is a good start to find correlation between groups. However, we need a more robust model to formulate conclusions. With ANOVA (Analysis of Variance), we can determine whether the variation between the means of the groups is larger than the variation within the groups, and whether this difference is statistically significant. We will perform a one-way ANOVA to answer our first research question. The independent variable will be perceived race. The dependent variable will be strip search.

H0: $\mu 1 = \mu 2 = \mu 3 = \mu..$ (where μ = mean) There is no significant difference in strip search for perceived race
H1: There is a significant difference in strip search for perceived race

Two Way

We are also interested in the combined effect on age groups. We will run a two-way ANOVA to answer our second research question. It will inform if there is a significant effect of each factor on the dependent variable, as well as whether there is an interaction effect between the two factors. The two factors will perceive race and age groups at arrests as independent variables. Strip search will be our dependent variable.

H0: There is no significant difference in strip search by perceived race.
H1: There is no significant difference in strip search by age.
H2: There is no significant interaction between perceived race and age, or their main effects, on strip search.

Assumptions checks

1. **Independence**: The observations within each group of our dataset are independent of each other. This assumption is passed.
2. **Normality**: The data within each group is not normally distributed. See Figure 6. This assumption fails.
3. **Homogeneity of Variance**: The variance of the data within each group is equal.

Figure 8: Leve's test for Homogeneity results

```
#Homogeneity
from scipy.stats import levene


# define alpha
alpha =0.05

# now we pass the groups and center value from the following
# ('trimmed mean', 'mean', 'median')
w_stats, p_value =levene(Black, White, Indigenous, Latino, South_east,
                         Middle_Eastern, South_asian, center ='mean')
print(p_value)

if p_value > alpha :
  print("We do not reject the null hypothesis")
else:
  print("Reject the Null Hypothesis")
```

```
6.225354902749196e-43
Reject the Null Hypothesis
```

The p-value is 6.225354902749196e-43. This is below the alpha of 0.05. Therefore we reject the null hypothesis that the variance among groups is equal. This assumption fails.

4. **Random Sampling**: The groups are formed by random sampling from the population. Indeed, the arrests came from arrests that were not controlled. This assumption is passed.

The assumptions are violated. However, we will still run the statistical test and note it as a limitation.

Post-hoc tests e.g, Tukey's HSD

We will run a post-hoc Tukey HSD (Honestly Significant Difference) test for each ANOVA. It will allow us to identify which specific groups in our dataset are significantly different from each other after conducting our ANOVA test that showed a significant difference between the means of some or all of the groups.

Results/ Findings (includes interaction plots)

Table 7: One-Way ANOVA results

```
from scipy.stats import f_oneway

# One-way ANOVA
statistic, pvalue = f_oneway(Black, White, Indigenous, Latino, South_east,
                             Middle_Eastern, South_asian)

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

One-way ANOVA: s = 15.703367169787976, p = 5.389566540799918e-17
```

The results show an F-statistic value of 15.703367169787976, which demonstrates a statistically significant difference in the mean number of strip searches. Furthermore, the P-value of 5.389566540799918ee-17 is less than 0.05, indicating strong evidence for the rejection of the null hypothesis. Overall, the results based on the one-way ANOVA showed significant differences in the number of strip searches based on perceived race.

Post Hoc Test - Tukey HSD

Figure 9: One way ANOVA Post Hoc Test - Tukey HSD results

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-18.2202	0.001	-29.909	-6.5313	True
Black	Indigenous	-17.9708	0.001	-30.0827	-5.859	True
Black	Latino	-19.7919	0.001	-31.9037	-7.68	True
Black	Middle-Eastern	-18.9766	0.001	-30.8643	-7.0889	True
Black	South Asian	-18.7929	0.001	-30.6508	-6.9351	True
Black	Unknown or Legacy	-16.6745	0.001	-28.2595	-5.0896	True
Black	White	8.3623	0.3517	-3.173	19.8977	False
East/Southeast Asian	Indigenous	0.2493	0.9	-11.8625	12.3612	False
East/Southeast Asian	Latino	-1.5717	0.9	-13.6835	10.5401	False
East/Southeast Asian	Middle-Eastern	-0.7564	0.9	-12.6441	11.1313	False
East/Southeast Asian	South Asian	-0.5727	0.9	-12.4306	11.2851	False
East/Southeast Asian	Unknown or Legacy	1.5457	0.9	-10.0393	13.1306	False
East/Southeast Asian	White	26.5825	0.001	15.0471	38.1179	True
Indigenous	Latino	-1.8211	0.9	-14.3416	10.6995	False
Indigenous	Middle-Eastern	-1.0058	0.9	-13.3096	11.2981	False
Indigenous	South Asian	-0.8221	0.9	-13.0971	11.4529	False
Indigenous	Unknown or Legacy	1.2963	0.9	-10.7153	13.3079	False
Indigenous	White	26.3332	0.001	14.3694	38.297	True
Latino	Middle-Eastern	0.8153	0.9	-11.4886	13.1191	False
Latino	South Asian	0.999	0.9	-11.276	13.274	False
Latino	Unknown or Legacy	3.1174	0.9	-8.8942	15.129	False
Latino	White	28.1542	0.001	16.1904	40.118	True
Middle-Eastern	South Asian	0.1837	0.9	-11.8702	12.2376	False
Middle-Eastern	Unknown or Legacy	2.3021	0.9	-9.4835	14.0877	False
Middle-Eastern	White	27.339	0.001	15.6021	39.0758	True
South Asian	Unknown or Legacy	2.1184	0.9	-9.6371	13.8738	False
South Asian	White	27.1553	0.001	15.4487	38.8619	True
Unknown or Legacy	White	25.0369	0.001	13.6068	36.4669	True

To know which perceived races differed significantly in the mean number of strip searches, we tested using Tukey's HSD test. The Black group has a p-value of 0.001 against all the groups except White. This is lower than the significant threshold of 0.05. Therefore, the Tukey HSD test confirms that the Black group were significantly more likely to be strip searched on average than all other minority groups. Furthermore, we can find other groups that have a p-value of lower than 0.05 such as Indigenous compared to White.

Table 8: Two-Way ANOVA results

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	94496.413806	7.0	18.862960	2.608622e-23
C(Age_group__at_arrest_)	43707.975219	8.0	7.634195	7.329143e-10
C(Perceived_Race):C(Age_group__at_arrest_)	77742.947466	56.0	1.939838	7.798278e-05
Residual	550343.419196	769.0	NaN	NaN

We conducted a two-way ANOVA to explore the effect of perceived race and age group on the likelihood of being strip-searched. The two-way ANOVA included two independent variables, "Perceived_Race" and "age group at the time of arrest," with the dependent variable being "strip search".

The p-values for Perceived Race and Age group at arrest is 2.608622e-23 and 7.329143e-10 respectively. This is less than 0.05 which implies that the means of both the factors possess a statistically significant effect on Strip searches. The p-value for the interaction effect is 7.798278e-05. This is lower than 0.05 which depicts that there is a significant interaction effect between Perceived Race and Age group. We reject all three null hypotheses.

Two-Way ANOVA Post Hoc Test - Tukey HSD

Figure 10: Two-way ANOVA Post Hoc Test - Tukey HSD results

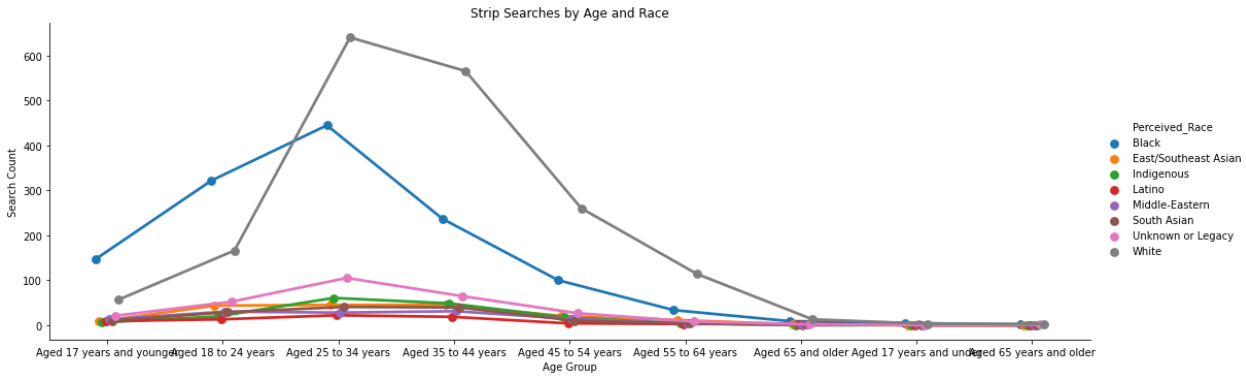
```
Black / Aged 17 years and under                South Asian / Aged 35 to 44 years    -11.0    0.9 -155.0553 133.0553  False
Black / Aged 17 years and under                South Asian / Aged 45 to 54 years     -6.0    0.9 -150.0553 138.0553  False
Black / Aged 17 years and under                South Asian / Aged 55 to 64 years     19.5    0.9 -124.5553 163.5553  False
Black / Aged 17 years and under                  South Asian / Aged 65 and older      3.0    0.9 -163.3407 169.3407  False
Black / Aged 17 years and under           South Asian / Aged 65 years and older    194.0 0.0028   27.6593 360.3407   True
Black / Aged 17 years and under      Unknown or Legacy / Aged 17 years and under    213.0  0.001   46.6593 379.3407   True
Black / Aged 17 years and under    Unknown or Legacy / Aged 17 years and younger     45.0    0.9 -121.3407 211.3407  False
Black / Aged 17 years and under          Unknown or Legacy / Aged 18 to 24 years     68.0    0.9  -76.0553 212.0553  False

  Black / Aged 17 years and under      Unknown or Legacy / Aged 65 years and older     21.0    0.9 -145.3407 187.3407  False
  Black / Aged 17 years and under                  White / Aged 17 years and under     27.0    0.9 -139.3407 193.3407  False
  Black / Aged 17 years and under                White / Aged 17 years and younger    168.0 0.0431    1.6593 334.3407   True
  Black / Aged 17 years and under                      White / Aged 18 to 24 years     41.0    0.9 -103.0553 185.0553  False
```

Table 9: Two-Way ANOVA results filtered for TRUE

group1							group2							meandiff	p-adj	lower	upper	reject
Black	/	Aged	17	years	and	under	White	/	Aged	17	years	and	younger	168	0.0431	1.6593	334.3407	TRUE
Black	/	Aged	17	years	and	younger	White	/	Aged	17	years	and	younger	177	0.018	10.6593	343.3407	TRUE
Black	/	Aged	18	to	24	years	White	/	Aged	17	years	and	younger	162	0.0062	17.9447	306.0553	TRUE
Black	/	Aged	25	to	34	years	White	/	Aged	17	years	and	younger	168	0.0028	23.9447	312.0553	TRUE
Black	/	Aged	35	to	44	years	White	/	Aged	17	years	and	younger	172.5	0.0015	28.4447	316.5553	TRUE
Black	/	Aged	45	to	54	years	White	/	Aged	17	years	and	younger	176	0.001	31.9447	320.0553	TRUE
Black	/	Aged	55	to	64	years	White	/	Aged	17	years	and	younger	174.5	0.0011	30.4447	318.5553	TRUE
Black	/	Aged	65	and	older	South	Asian	/	Aged	65	years	and	older	205	0.001	38.6593	371.3407	TRUE
Asian	/	Aged	17	years	and	younger	White	/	Aged	17	years	and	younger	180	0.0132	13.6593	346.3407	TRUE

The Black - Aged 17 under, 18-24, 25-34, up to 65 years old, has a p-value of lower than 0.005 against the White group 17 . This is lower than the significant threshold of 0.05. Therefore, we reject the null hypothesis. We find similar results for the following minority groups; Indigenous, Latino and Middle-Eastern.The Tukey HSD test confirms that the race and age group were significantly more likely to be strip searched on average.

Figure 11 : Interaction plot of the relationship between the number of strip searches conducted by age and race



The interaction plot above illustrates the relationship between the number of strip searches conducted by age and race. This interaction plot includes the number of strip searches plotted on the y-axis and the age group plotted on the x-axis and uses different colors to indicate the different perceived races of the searched individuals.

The interaction plot shows that for most age groups, the number of strip searches increases with age for arrestees younger than 35 years old, but for arrestees older than 35 years old, it just starts to decrease slowly with age.

In addition, the interaction plot shows that the perceived Black race is relatively more likely to be strip searched than other perceived races. Overall, we can use the information expressed in the plot to understand that perceived race and age group have an interactive effect on the rate of strip searches. This is because the probability of being strip searched varies for different perceived racial groups depending on their age.

In general, the p-values for the three effects of C(Perceived_Race), C(Age group at arrest), and C(Perceived_Race):C(Age_group__at_arrest_) were all less than the significance level (0.05), indicating that we can reject the null hypothesis. Therefore, we can conclude that there is a significant difference in the mean value of "StripSearch" in different levels of each factor and their combination.

Discussion

Overall, the results of our statistical analysis reveals there are significant differences in strip searches based on demographic attributes of Race. The two way ANOVA also shows there is a significant difference in the means of strip search based on Race and Age as well as a combined effect on strip search. From the Tukey HSD test on our Two-Way ANOVA, we see the groups with significant differences in means are minority groups. Indeed, the results demonstrate that while “Carding” practices are discontinued, the Toronto Police still dispositionally target vulnerable communities. Further investigation should be done to identify the reasons behind this discriminatory approach. It is known that police tend to increase more efforts in certain racialized communities, which therefore increase the number of arrests (Giwa, S, 2018). We recommend the Toronto Police to reassess their arrests and strip searches practices based on the evidence brought by this paper that there exists some relationships between strip search and protected demographic attributes. Debiasing techniques such as intra-processing methods could be helpful in debiasing (Savani & al., 2020). The “Artificial Intelligence Fairness 360” toolkit could be implemented as well as it provide a good starting bias framework for organizations (Bellamy & al., 2019)

Limitations

While we have concluded a correlation between our dependent variable strip search with independent variables race and age, we do not have evidence for causation.

Next, we should perform an Analysis of Covariance to establish the relationship between the movement of our variables while controlling for the effects of the covariates . This would reduce the variability in our outcome variable, which can improve the accuracy of our statistical test and increase the power of our analysis.

Conclusion

In conclusion, we showed that the perceived race and age group at the time of arrest have a significant effect on the average number of strip searches. For minority groups, the probability of being strip-searched significantly differed from that of the dominant white group. We conducted a One-Way Anova and Two-way Anova tests to see the correlation between perceived race, age group, and strip search. As detailed in our report, we reject the null hypothesis H0 at the 95% significance level, thus indicating that both age group and perceived race are related to the likelihood of being strip-searched and that age group and perceived race have some influence on each other. Subsequently, we obtained further validation on which group differed through Tukey's HSD test

Further analysis on the biases using an Analysis of Covariance could be conducted further to establish causal claims for the outcome variable strip-search. Additionally, as seen in the EDA, there is a correlation between strip search and "Booked that could be explored. The purpose of doing so is to gain a more comprehensive understanding of the factors that influence the rate of strip searches to provide as much fairness as possible in making arrests.

In summary, there are significant differences in strip search rates based on perceptions of race and age group. This demonstrates potential biases in the Toronto Police. We urge the Toronto Police to investigate further on their practices using our recommendation. Ultimealty, we hope it could lead the way for more organizations to adopt more fair bias free practices across Canada.

References

Barabas, C., Doyle, C., Rubinovitz, J. B., & Dinakar, K. (2020, January). Studying up: reorienting the study of algorithmic fairness around issues of power. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 167-176).

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM Journal of Research and Development, 63(4/5), 4-1.

Giwa, S. (2018). Community policing in racialized communities: A potential role for police social work. Journal of Human Behavior in the Social Environment, 28(6), 710-730.

Grewcock, M., & Sentas, V. (2021). Strip searches, police power and the infliction of harm: An analysis of the New South Wales strip search regime. International journal for crime, justice and social democracy, 10(3), 191-206.

Joh, E. E. (2017). Feeding the machine: Policing, crime data, & algorithms. Wm. & Mary Bill Rts. J., 26, 287.

Johnson, V. E., Nadal, K. L., Sissoko, D. G., & King, R. (2021). “It’s not in your head”: Gaslighting,‘Splaining, victim blaming, and other harmful reactions to microaggressions. Perspectives on psychological science, 16(5), 1024-1036.

Spurrier, M. (2023). The police are perpetrating harm, so why is the government giving them more power?. bmj, 380.

Stop-and-Frisk Data". New York Civil Liberties Union. January 2, 2012. Retrieved November 30, 2019

Savani, Y., White, C., & Govindarajulu, N. S. (2020). Intra-processing methods for debiasing neural networks. Advances in Neural Information Processing Systems, 33, 2798-2810.