Final Project Write Up

Zaid Abdulaziz, Bryan Zhu

Group 17

INF2178 Experimental Design for Data Science

Faculty of Information, University of Toronto

Shion Guha

Introduction

The Toronto Police Service (TPS) is the primary police force responsible for ensuring law enforcement in the city of Toronto. The TPS is the largest police service on the municipal level in Canada, and the third largest in the country overall, trailing only the Royal Canadian Mounted Police and the Ontario Provincial Police. Their self-stated mission is: "*We are dedicated to delivering police services in partnership with our communities to keep Toronto the best and safest place to be.*" Alongside providing policing services, the TPS has a responsibility in maintaining professionalism and accountability among its task force, in order to ensure that they treat all individuals with equal dignity and respect. As a service that is publicly funded through tax dollars from those they are meant to protect, the TPS has an obligation to fulfill the needs of the community it serves and to aid in simply making their municipality a better place for people to live.

In recent years, particularly following the death of George Floyd in Minneapolis, Minnesota, USA in 2020, there has been rising concern with excessive force exhibited by police officers, discrimination towards members of certain minority groups, and lack of accountability within police departments. While problems with discrimination and profiling by police forces had always been of concern, this event dramatically pushed forth this issue into the limelight. Protests were held all over the globe in a widespread movement to demand changes to police departments in cities around the world.

The growing concern for discrimination and misuse of police power has been of interest to researchers as well. In 2021, the U.S. Department of Justice published research pertaining to data from 2018 which concluded that black and Hispanic individuals were overrepresented in crime arrests relative to their proportion of population, while white people were underrepresented in crime arrests relative to their proportion of the population. Canada is anecdotally considered to be a diverse and accepting nation, but nevertheless a report by the Ontario Human Rights Commission found that black individuals in Toronto were also overrepresented as police targets. The TPS, along with most police departments around the world, typically denies allegations of internal biases and discrimination, and thus many public opinions of the TPS have declined negatively over the past few years.

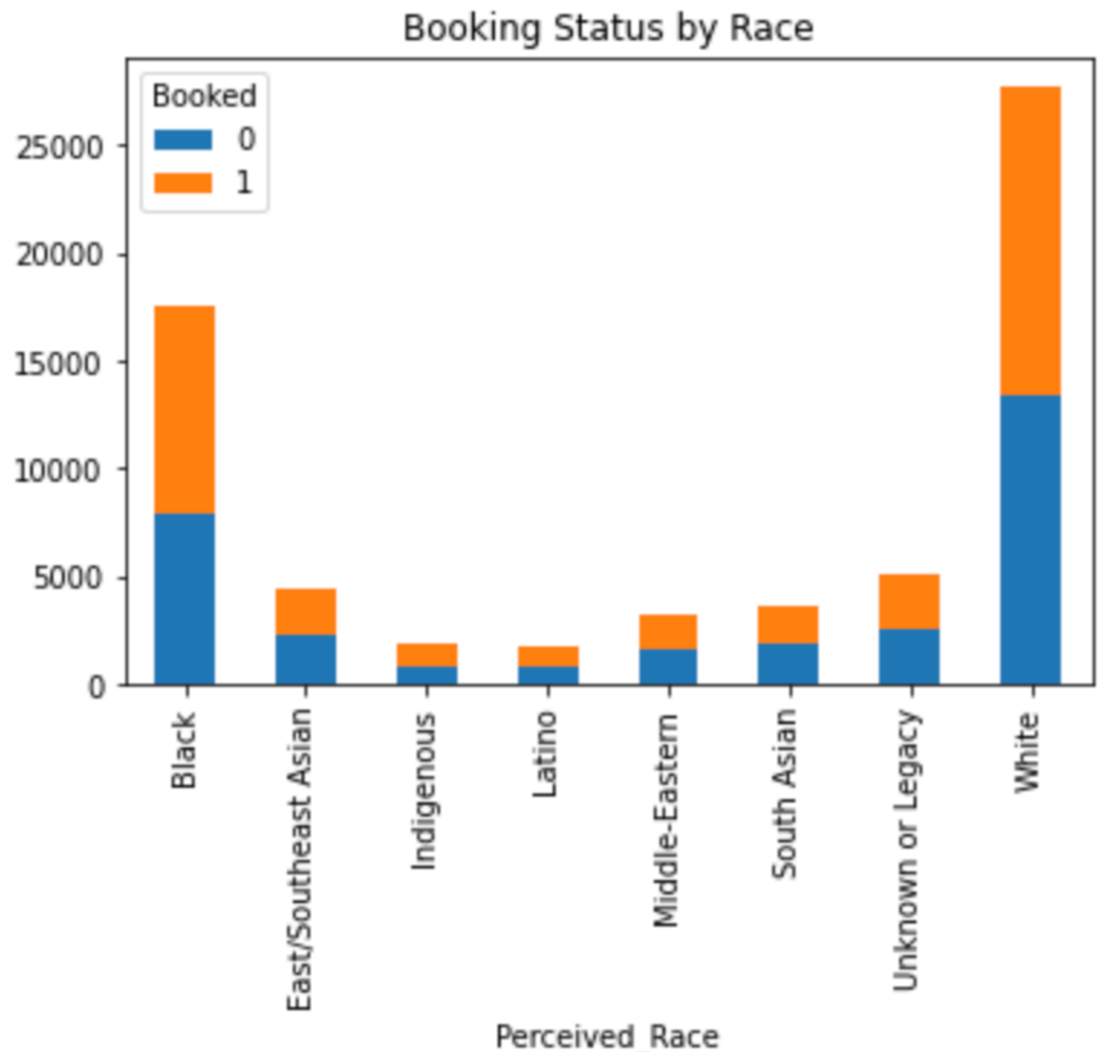
In this report, we aim to explore the dataset "Arrests and Strip Searches (RBDC-ARR-TBL-001)" to investigate whether there are disproportionate investigations and actions towards certain groups of people. We will aim to answer two research questions in specific. The first is what are differences in rate of police actions towards people of differing demographics. The second is whether there's a statistically significant difference in police actions based on those individual demographics. With these research questions, we will try to gain a better understanding of the Toronto Police Service in seeing if there is underlying discrimination within this police department.

Exploratory Data Analysis

The dataset consisted of 65,276 individual records of Toronto police arrests of various individuals. Demographic information, such as perceived race and sex, as well as police actions taken, such as bookings and strip searches, are noted in the columns. More information regarding the actual dataset will be provided in an upcoming section in this report.

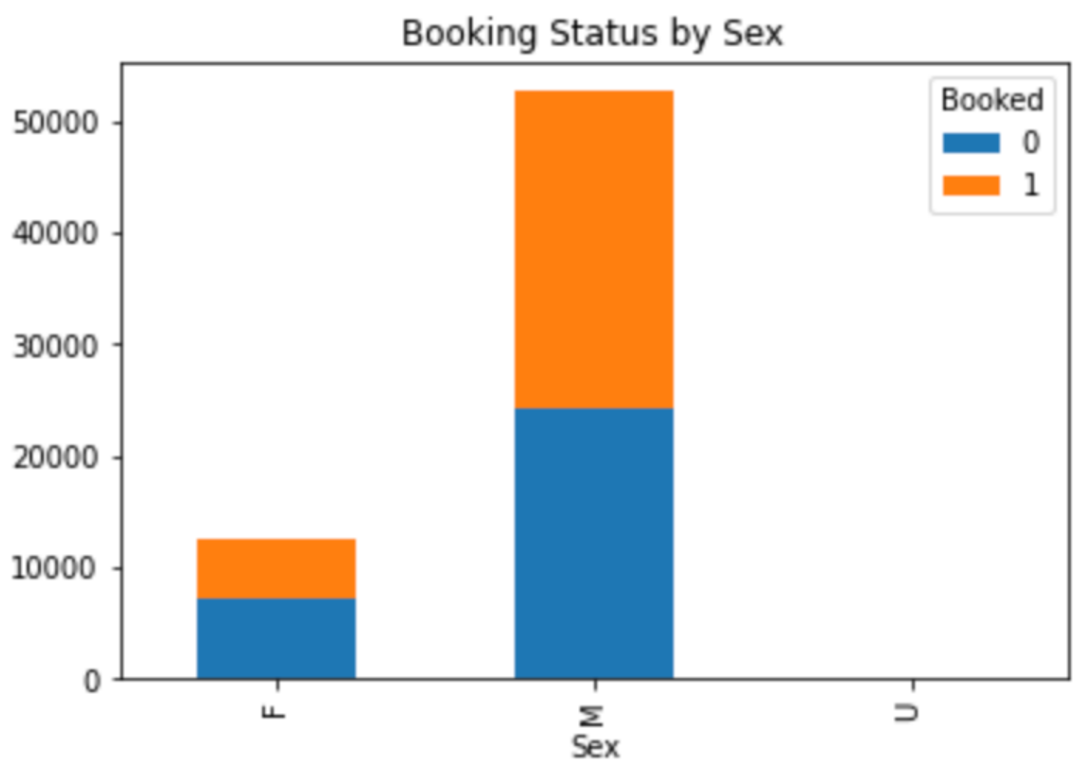
The first thing we wanted to examine was whether an individual was booked or not. The action of booking means entering a suspect's information, such as physical measurements, fingerprints, and mugshot into a police system following an arrest. Although related, this process is distinct from the action of getting arrested.

Figure 1



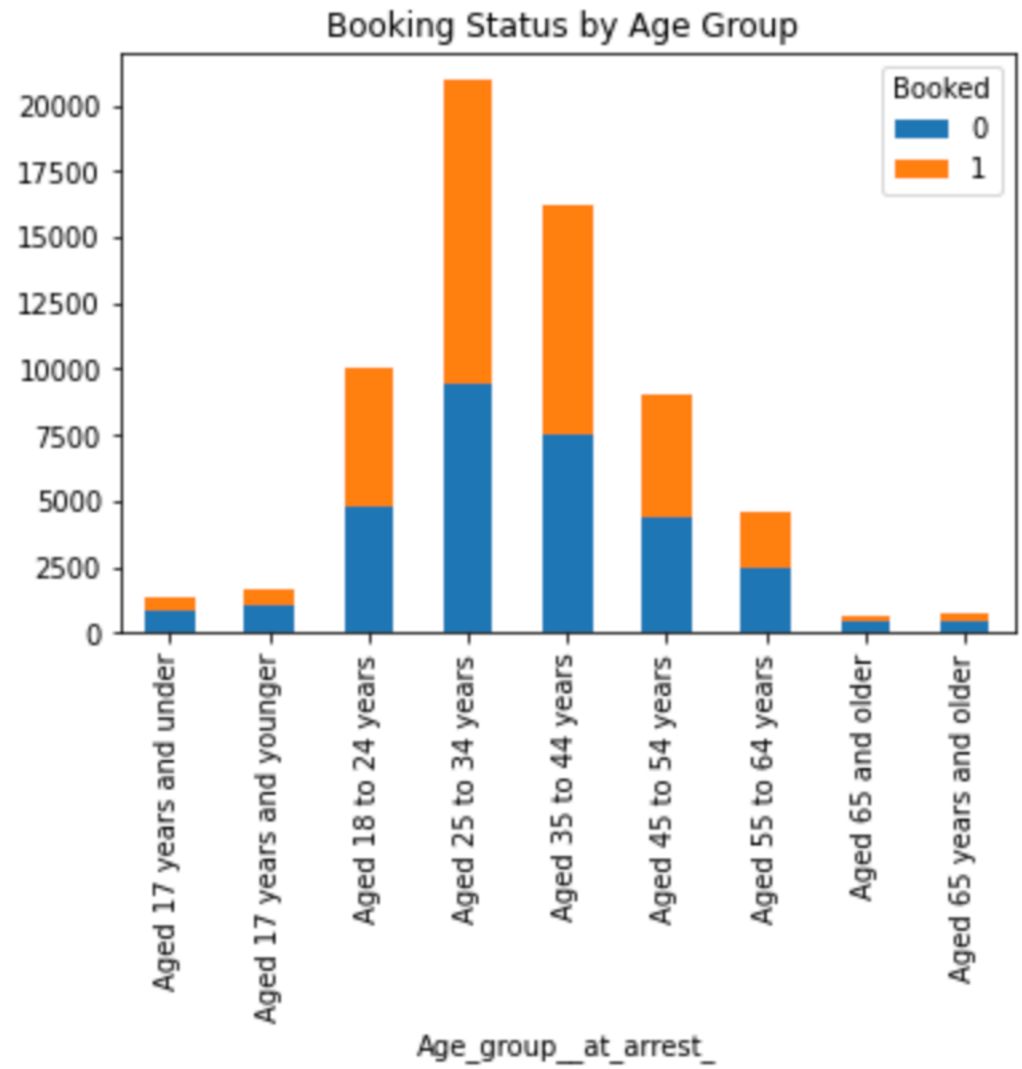
In Figure 1 above, we examine the different statuses of bookings based on perceived race. The state of whether the individual was booked or not following their arrest is indicated by the colour of the bars in the histogram; orange indicates a booking and blue indicates a non-booked individual. From this graph, there seems to be similar booking rates among the different rates. The booking rate is roughly one in every two arrests for each race, indicated by the fact that the proportion of coloured bars are roughly equal in each column. One will clearly notice however, that the bars for people of black perceived race and white perceived race are substantially higher than the bars for the other perceived races. This indicates that there are more total arrests of people belonging to those perceived races. Although the total number of arrests and bookings of black individuals is lower than that of white individuals, one should take into consideration the fact that the number of white individuals living in Toronto far outnumber the number of black individuals living in Toronto.

Figure 2



In Figure 2, the rates of booking status in relation to sex are examined. Unlike the previous Figure 1, in this figure it shows that males have a greater chance of being booked than females do. The proportion of males being booked following arrests is greater than 50%, while the proportion of females being booked is less than 50%. It is also noteworthy that there is a clear disparity in rates of arrests by sex; males are far more likely to be arrested than females are, as indicated by the extreme difference in bar heights in Figure 2.

Figure 3



In Figure 3, we observe the booking status in relation to age group. We note that in addition to individuals aged 17 and below and 65 and above having lower arrest rates in general, they are also less likely to be booked than people belonging to other age groups. Individuals aged 25 to 34 years both have the greatest arrest numbers and are most likely to get booked following the arrest.

Figure 4

In Figure 4, we wanted to look at booking rates compared to actions concealed. If an individual is in the actions concealed column, it means that at the arrest they had concealed items. Individuals having no concealed items seemed to have no differences in booking rates. The number of people who were found to have concealed items at arrest was too small to be able to observe any difference in booking rate in this figure.

Figure 5

In Figure 5, we look at booking status in relation to combative actions. What this means is individuals in the right column were noted to have combative actions during their arrest, such as biting or spitting. Individuals without combative actions did not have any noteworthy differences in booking rates. Individuals who had combative actions however were more likely to be booked following their arrest, as indicated by the greater proportion of orange in the right-side column.

Following this, we conducted a series of t-tests to determine whether there were statistically significant differences between some of these areas of interest.

Figure 6

```
[ ] # t test for booked and sex
    tStat, pValue = stats.ttest_ind(table1['Sex'][table1['Booked'] == 1],
                                    table1['Sex'][table1['Booked'] == 0], equal_var = False)
    print("P-Value:{0} T-Statistic:{1}".format(pValue,tStat))

    P-Value:0.5822700263828706 T-Statistic:0.5517787523981469
```

Figure 6 shows the t-test performed for booking status in relation to sex. The null hypothesis for the t-test is that there is no significant difference between booking rate between males and females. Since the p-value is 0.58, far above the 0.05 threshold for significance, we fail to reject the null hypothesis.

Figure 7

```
[ ] # t test for booked and actions concealed
    tStat, pValue = stats.ttest_ind(table1['Actions_at_arrest___Concealed_i'][table1['Booked'] == 1],
                                    table1['Actions_at_arrest___Concealed_i'][table1['Booked'] == 0], equal_var = False)
    print("P-Value:{0} T-Statistic:{1}".format(pValue,tStat))

    P-Value:0.5474659919656182 T-Statistic:0.6034870468926516
```

Figure 7 shows the t-test for booking status in relation to whether the suspect had concealed items. The null hypothesis for this t-test is that there is no significant difference in booking rate between people who had concealed items and people who did not have concealed items. The p-value for the test is 0.55, so we fail to reject the null hypothesis.

Figure 8

```
[ ] # t test for booked and actions combative
    tStat, pValue = stats.ttest_ind(table1['Actions_at_arrest___Combative__'][table1['Booked'] == 1],
                                    table1['Actions_at_arrest___Combative__'][table1['Booked'] == 0], equal_var = False)
    print("P-Value:{0} T-Statistic:{1}".format(pValue,tStat))

    P-Value:0.15116604656296828 T-Statistic:1.445819223466016
```

Figure 8 shows the t-test for booking status in relation to whether the suspect performed combative actions. The null hypothesis for this test is that there is no significant difference in bookings between those who had and those who did not have combative actions. The p-value for this test is 0.15, so although it's close, again we fail to reject the null hypothesis.

Methods

The dataset in this project was provided by the Toronto Police Service Public Safety Data Protocol. It is self-described as having information relating to all arrests and strip searches. There are a total of 65,276 listings, and 24 recorded attributes (variables). The variables include a wide variety of demographic data (perceived race, age group, sex), arrest actions (strip search, bookings, arrest categories, actions at arrest), and dates of arrests. Most of the data within this dataset is categorical. The dataset consists of arrests within the city of Toronto, as well as arrests outside of the city limits that were conducted by the Toronto Police Service. As of the time of writing of this report, the information and data within the TPS data table was last updated on November 10, 2022.

As mentioned, the report first began with conducting exploratory data analysis in order to gain a brief overview of some of the general trends within the table. Primarily, we were interested in booking status following arrests, and observed some of the relations booking status had with the other attributes of the data set. After this, some t-tests were performed to observe whether there were any significant results.

Next, we constructed a logistical regression model to observe our predictive analytics. We were interested in the variable of perceived race and sex, and how they related to the booking rate after arrests. The logistic regression model created is shown in Figure 9.

Figure 9

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

# logistic regression
X = table1[['Perceived_Race', 'Sex']]
y = table1['Booked']

# Slit data into training and testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Define the logistic regression model and fit it to the training data
logreg = LogisticRegression()
logreg.fit(X_train, y_train)

# get class labels
y_pred = logreg.predict(X_test)

# get accuracy of model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

After these, we conducted two ANCOVA tests to similarly observe the effect of perceived race and sex on the booking rates following arrest. The benefit of ANCOVAs is that they allow for the controlling of a covariate factor, something that normally wouldn't be possible when making field observations. Typically, the controlling of external factors needs to be done through randomised control experiments, which this dataset is not representative of. Hence, ANCOVA can provide great utility since it allows for us to examine the effects of our desired variables (perceived race and sex) on the variable in question (booking rate). The codes for the ANCOVA tests are shown in Figures 10 and 11.

Figure 10

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols



# Perform ANCOVA with 'Race' and 'Sex' as independent variables and 'booked' as the dependent variable
model = ols('Booked ~ Perceived_Race + Sex', table1).fit()

# Print  table
print(sm.stats.anova_lm(model, typ=2))
```

Figure 11

```
# Perform ANCOVA with  'StripSearch' as the continuous covariate
model = ols('Booked ~ Perceived_Race + Sex + StripSearch ', table1).fit()

# Print  table
print(sm.stats.anova_lm(model, typ=2))
```

Lastly, a power analysis is conducted to estimate the smallest sample size that would be required for proper data analysis of a dataset such as this. From this power analysis, we can deduce whether our previous statistical tests have strong basis given the sample size of the dataset compared to the sample size given from the power analysis. Figure 12 shows the code used for the power analysis.

Figure 12

```
import statsmodels.stats.power as smp

alpha = 0.05
power = 0.8
effect_size = 0.5

# power analysis
power_analysis = smp.TTestIndPower()
sample_size = power_analysis.solve_power(effect_size=effect_size, alpha=alpha, power=power)

print(f"Sample size: {sample_size}")
```

Results

From the logistic regression model, the accuracy of the model was calculated. The obtained accuracy value for this model was 0.55. This tells us that the model might not be a good predictor of future values since the accuracy is rather low.

The ANCOVA test results, in which both tests booking status is the dependent variable, are shown in the following tables below:

ANCOVA Table 1 (from Figure 10)

	Sum of Squares	df	F	PR(>F)
Perceived_race	21.742418	1.0	87.815860	7.404166e-21
Sex	108.070720	1.0	436.488856	1.304860e-96
Residual	16161.008483	65273.0	NaN	NaN

ANCOVA Table 2 (from Figure 11)

	Sum of Squares	df	F	PR(>F)
Perceived_race	11.028680	1.0	48.891973	2.730430e-12
Sex	88.592377	1.0	392.744743	3.777482e-87
Stripsearch	1437.446372	1.0	6372.438877	0.000000e+00
Residual	14723.562111	65272.0	NaN	NaN

In Table 1, the p-values for both sex and perceived race, shown in the farthest righthand column, are both extremely small decimal values well below 0.05. This means that the results are significantly significant and we can reject the null hypothesis that perceived race or sex do not impact booking rates after arrest.

Similarly in Table 2, which adds in strip searches as a continuous covariate, the p-valaues for perceived race and sex remain below the 0.05 threshold. These similar results from both ANCOVA tests gives us greater confidence in suggesting that perceived race and sex do result in an impact in booking rates following arrests by Toronto Police Service members.

The value for sample size from the power analysis was 63.77. This means that the minimum sample size for this "experiment" should have been 64. Given that this dataset contains over 65,000 listings, we can operate under the assumption that our sample size from the dataset is more than sufficient, giving us greater confidence in our results.

Discussion

Overall, the findings suggest that there is a statistically significant difference in booking rates conducted by Toronto Police Service members. Factors such as race and sex may affect these booking rates and appear to be disproportionate, and upon conducting the appropriate statistical tests a statistically significant correlation was found.

These findings are not surprising, given that there is a wide array of previous research in the literature that suggests similar behaviours in the field. There have been many studies conducted, including the ones mentioned earlier in the introduction, that display differing police treatment to individuals based on several sociocultural factors such as race, sex, and income. There is lots of data, including the data within this own dataset, that suggest that black individuals are overrepresented when it comes to arrest numbers relative to their proportion in the population of interest. It is fairly common knowledge that police tend to prioritize efforts in lower-income regions that have historically higher crime rates. Within these lower-income regions however, there tend to be a greater proportion of minority groups who live there, meaning that police will be spending more resources and effort patrolling regions where minorities make up a large part of the habitants. If a police presence is more common, it is likely that it in itself results in more arrests and subsequently more bookings simply due to the fact that there is a greater police presence in the area.

Additionally, regardless of race, individuals living in lower-income areas are more likely to either commit, experience, or be arrested than individuals living in higher-income areas. This inflates the number of arrests from low-income regions, which again, are typically populated by minority groups. This also then in turn slows down socioeconomic progression of lower-income regions, and a vicious cycle of people unable to escape poverty is perpetuated. Social, cultural, and historical aspects of a community are not factored into consideration when performing and conducting data analysis, which is why municipalities should always take caution if relying heavily on solely statistics to make administrative decisions, as they may not always tell the whole story.

Our statistical tests show a bias in booking rates, but do not show a way that this bias may be removed. This is a multi-faceted issue in which many different socioeconomic and cultural elements must be taken into consideration to truly fix the problem; our project simply aims to show that the bias exists within the Toronto Police Service, and that in particular members of the black community are disproportionately more prone to bookings.

Conclusion

To summarize, when conducting exploratory data analysis we found patterns of high arrest frequency of black individuals, and high booking rates for males and for individuals aged 25 to 34 years old. When performing further statistical tests, such as ANCOVA, these findings were shown to be statistically significant, suggesting that the answer to our research question is that police actions towards people of differing demographics are on average not equal. Specifically, black individuals and males are more likely to be booked following their arrest.

One important limitation of this project is that the dataset only contained arrest information for a two-year period, which was specifically during the Covid-19 pandemic, which was known to have disproportionally displaced certain groups of individuals more than others. It would be interesting to see if there's a larger dataset that included information from a longer time range, particularly data from before the year 2020. Perhaps this would change the results of some of the statistical tests conducted and provide opportunities for other aspects of data analysis as well.

The Toronto Police Service is not perfect but remains an integral part of municipal governance and law enforcement. People, to no fault of their own, all have their own inherent biases that arise for various reasons such as environment and social upbringing. While this is unavoidable, there are steps individuals may take to try to recognize their own biases and try to take them into consideration. We ultimately wish that the Toronto Police Service take care and stay true to their mission of keeping Toronto the best place to live.

References

Beck, Allen J. (2021). Race and Ethnicity of Violent Crime Offenders and Arrestees, 2018. Bureau of Justice Statistics.

Owusu-Bempah, A. (2014). Black Males' Perceptions of and Experiences with the Police in Toronto.

Toronto Police Service. (n.d.). Mission, Vision & Values. Toronto Police Service, from https://www.tps.ca/mission-vision-values/

Toronto Police Service Public Data Portal. (2022). Arrests and Strip Searches (RBDC-ARR-TBL-001), from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about