Exploratory data analysis to assess the impact of COVID-19 on the Toronto crime statistics

Submitted by:
Dailiang Pan (1003559053)
Giriraj Heda (1007992415)

Master of Information
Department of Information, University of Toronto
INF2178H: Experimental Design for Data Science
Dr. Shion Guha
Feb 28, 2022

Introduction

According to the report, it reveals that the unemployment rate has rebounded and reached a record low of 4.9% in mid-2022 (Macklem, 2022), which indicates that the Canadian economy has finally recovered from the COVID-19 pandemic. The pandemic has affected many aspects of human society around the world, which means it results in a long-term societal impact and it has reshaped the structure of human society. Alexandre White, a sociologist, and historian from Johns Hopkins believed that the COVID-19 pandemic has not only affected global health systems but also the global economy and society. In his discussion, he mentioned "*…have reverberated long after the disease stops spreading*" (White, 2020). Hence, even though most countries have already lifted their restrictions in relation to the pandemic, we still need to examine the ongoing and subsequent impacts that result from the COVID-19 pandemic and discover how it shapes our economy and society. In this project, we took a look at the dataset about crime in Toronto, the data set is built by the Toronto Police service including data related to arrest and strip searches in 2020 and 2021. Based on the dataset, we were able to analyze what was the trend in police-reported crime in Toronto from 2020 to 2021 under the effects of the COVID-19 pandemic and its subsequent year. We tried to discover the potential relationship between variables in terms of arrest and strip search happening in Toronto, and to examine how they were interconnected and articulated during the pandemic. The reason why we focused on crime is that we believed the statistics of crime that can be used to measure social stability. On the other hand, according to the data, the US also faced challenges in terms of social security such as the high homicide rate during the pandemic (Rosenfeld et al., 2020). Hence, we wanted to see if the Canadian crime statistics had a similar trend as the US. In conclusion, we wanted to discover how crime statistics changed during the pandemic for two consecutive years from 2020 to 2021, and we eventually generated two research questions:

Research Questions

Research Question 1: Did the year of covid-19 outbreak have an effect on the number of people booked for any offense and did the perceived race have any effect on it?

Research Question 2: Did the year of covid-19 outbreak have an effect on an individual's cooperation level while being arrested and did the perceived race have any effect on it?

Literature Review

To date, the effect of the COVID-19 pandemic is no longer considered only a public health crisis; it also has social impacts that cause social changes on a global scale, especially in the global economy (Pak et al., 2020). It has increased the risk of global economic depression due to a disruption in global trade and social activities (Yu et al., 2021). Moreover, another research reveals that it negatively influences society by exacerbating social inequality, violence, and disorder (Basu, 2021). This is because governments are failed to control and manage countries during the pandemic with outdated regulations, laws, or policies. As a result, many researchers aim to discover how the pandemic affects society by analyzing social phenomena. According to the article, the authors focus on weekly crime rate change among different cities in the US and find that homicide and auto-theft increased during the pandemic in 2020 (Meyer et al., 2022). While COVID-19 disrupted daily life, did it play a role in the crime statistics in Toronto between 2020 and 2021 and if race had any effect pertaining to these statistics is our focus.

Exploratory Data Analysis:

Dataset Description

The dataset we focused on is constructed by the Toronto Police service, and it contains information in terms of arrests and strip searches that took place in the City of Toronto, and the data collection period is from 2020 to 2021. A strip search is when a police officer requires a person to remove some or all clothing to check if someone is carrying concealed or illegal items, usually drugs or weapons. The data also contains related information on booking which refers to the process of a criminal suspect being brought to jail after that suspect's arrest. The location of arrest is formed by different Divisions in Toronto which indicates the division where each arrest took place. Further, the dataset has multiple data elements, which include arrest year and month, perceived race, sex, age group, actions at arrest, arrest location, occurrence category, search reason, items found, and book and strip search counts.

The dataset contains over 65,000 records but also has the problem of missing values, hence, the first step in our data processing and cleaning is going to remove all null data. The problem of missing values is caused by different reasons, and that commonly exists in real-life datasets. For instance, according to the data dictionary, values in '**ArrestLocDiv**' are indicated by 'XX' in case the suspect is arrested outside of the Toronto area. As an initial step, we sum up all null or XX values, then we use '996' to replace all null and inconsistent values for the dataset to make our dataset more meaningful. Next, we analyzed the dataset to gather detailed information, in which we utilize the cat-plot to create multiple categorical variable plots to visualize our data. What's more, we observed that some variables like 'EventID', 'ArrestID', and 'PersonID' are meaningless; we would not use them in analysis.

Data Cleaning and Visualization

The first figure we have is a bar chart showing suspects count by perceived race and sex and that provides a basic insight into our dataset. In terms of perceived race, the variable has 9 different types of races including White, Unknown or Legacy, Black, South Asian, Indigenous, Middle Eastern, Latino, East/Southeast Asian, and any invalided values were marked as 996. Based on that, the number of arrests obviously varied among perceived races, we found that male is more likely to be arrested compared to female; meanwhile, white people are more likely to be arrested compared to other races.

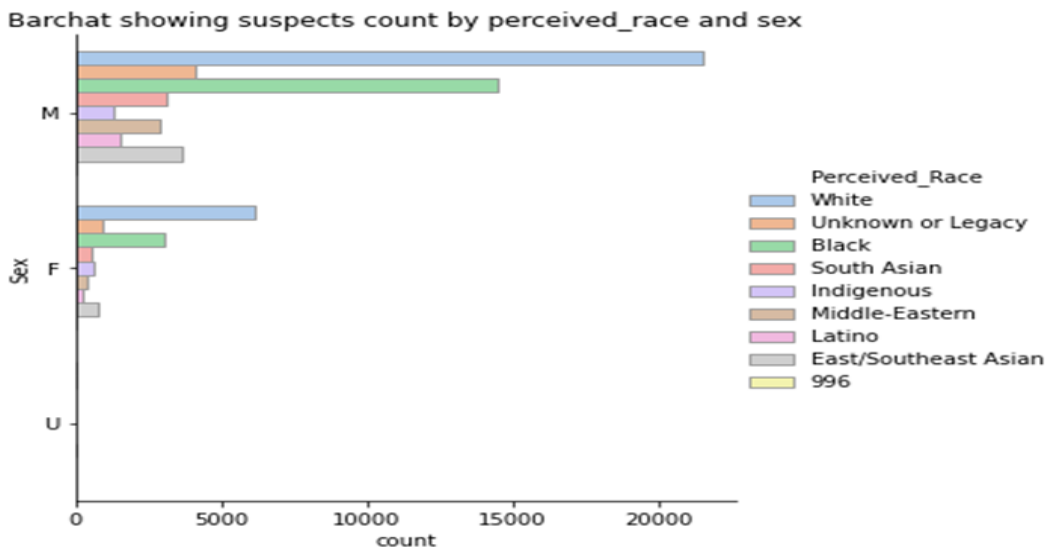

Fig: Bar chart showing suspects count by perceived_race and sex

We also looked at how arrest counts differed for each perceived race and the contribution of occurrence category for each perceived race. From figure below, we observed that suspects with a perceived race of white accounted for most of the total arrest counts, meanwhile, suspects with a perceived race of black accounted for the second-most of the total arrest counts. Following the observations, we decided to further examine white suspects and black suspects in terms of occurrence category and sex next.

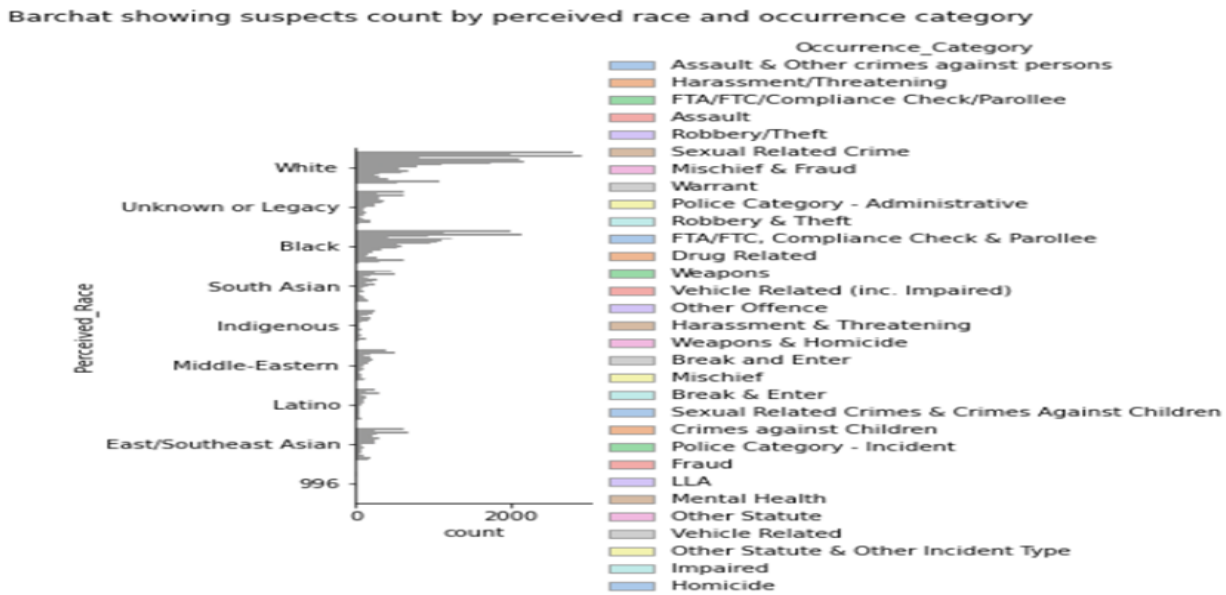

Fig: Bar chart showing suspects count by perceived_race and occurrence_category

Figure below provides a graphical depiction of the contributions from each occurrence category to total arrest counts in relation to white suspects in different sex groups. We compared different occurrence category contributions in terms of suspects with different sex groups, and we found that they are most likely to be arrested when they committed an assault, in terms of male suspects. Similarly, in terms of female suspects, they are most likely to be arrested when they committed an assault too. Therefore, we observed that the occurrence category of assault accounted for most of the total arrest counts regarding white suspects.

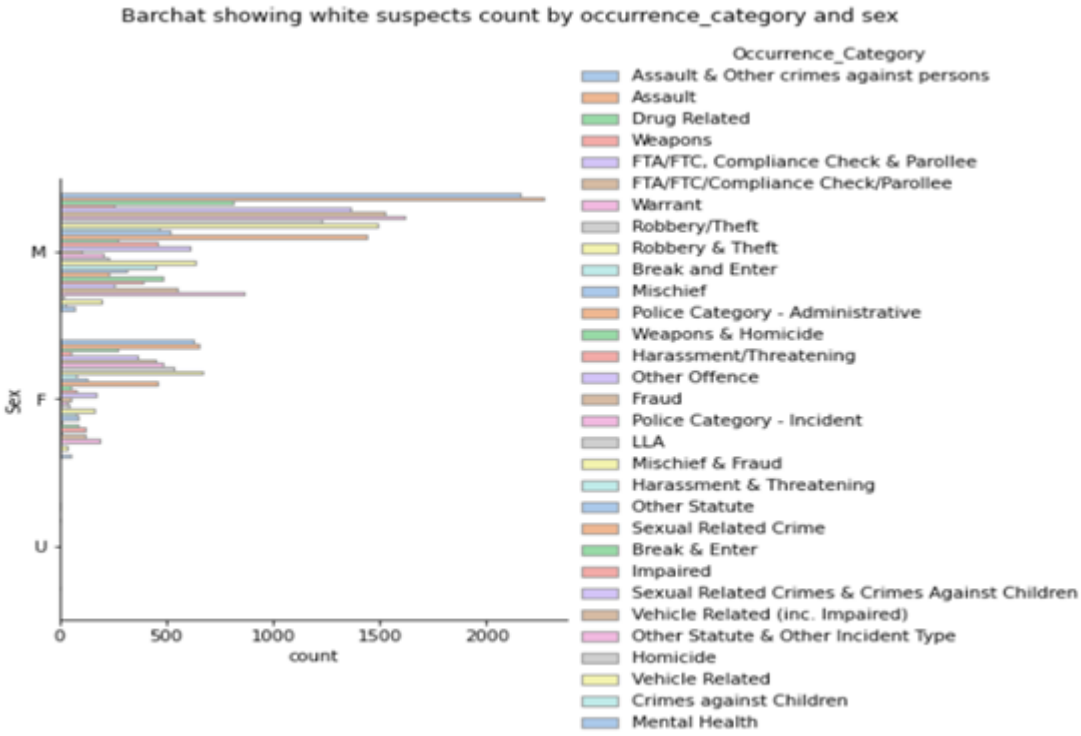

Fig: Bar chart showing suspects count by occurrence_category and sex for white suspects

Moreover, we also examined the contribution of the occurrence category within black suspects. Based on figure below, we obtain the observation of black suspects, which is like what we had found from white suspects, that they are most likely to be arrested when they committed an assault.

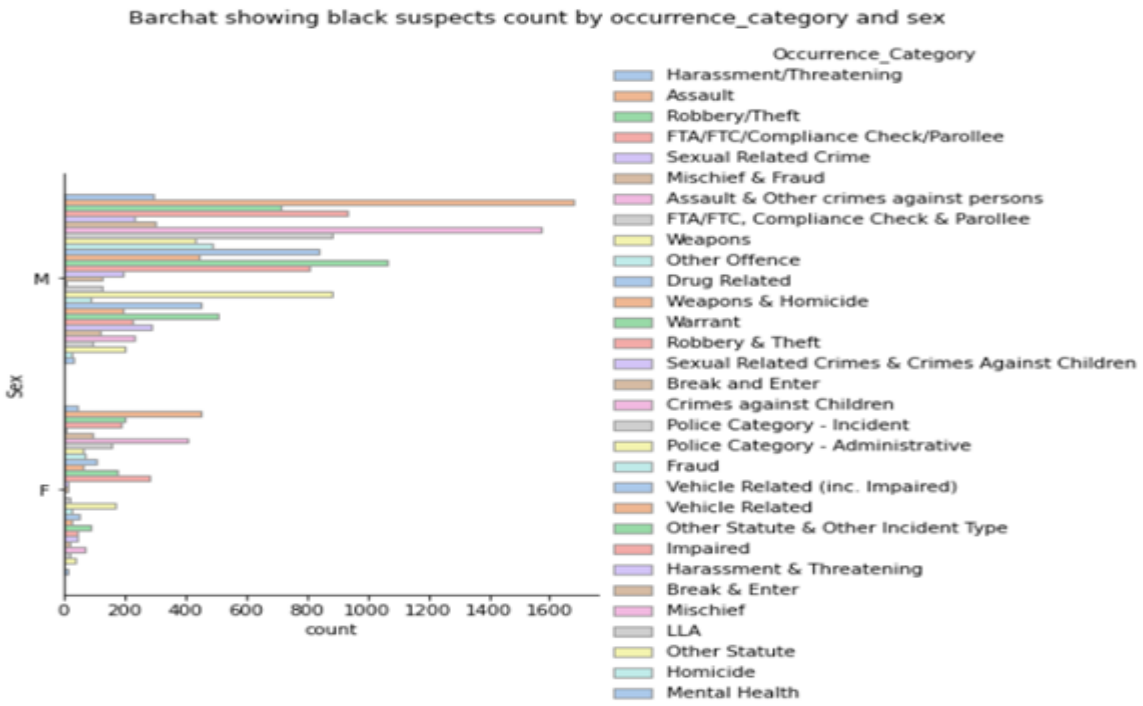

Fig: Bar chart showing suspects count by occurrence_category and sex for black suspects

Additionally, the distribution of crime type as assault for each perceived race and both sexes are shown in figure below. For each perceived race, we concluded that male suspects are more likely to be arrested due to an assault committed, compared to female suspects.

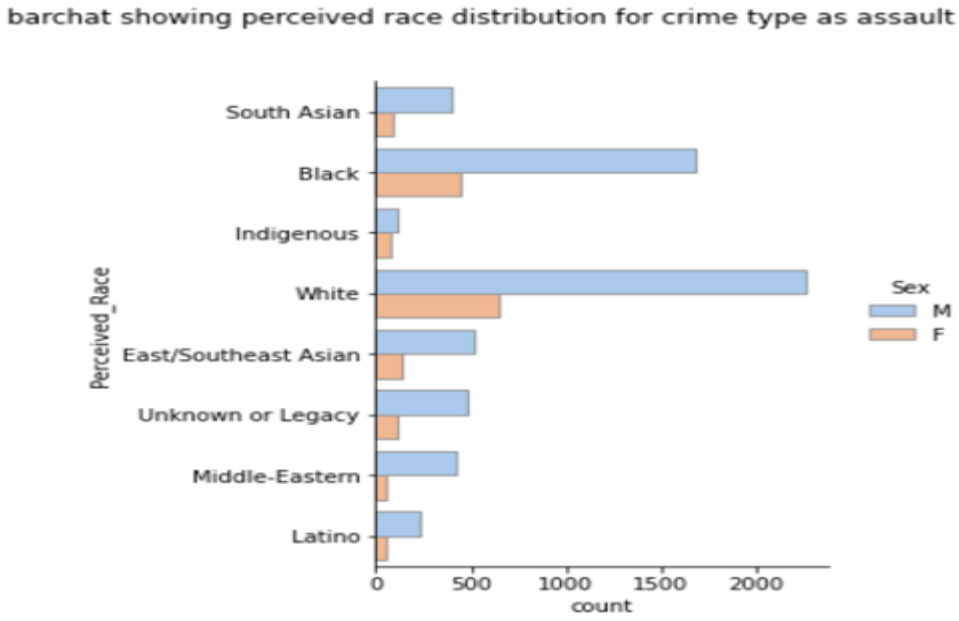

Fig: Bar chart showing perceived_race distribution for crime type as assault

Since the arrest took place at different police divisions, we wanted to check if there was any location-based correlation with race signifying some sort of race-based bias with the location and the race of the person being arrested. Also, we wanted to check if there was a certain age group that was arrested based on race.

Based on the computations, we get r=0.0035 indicating the relationship between perceived race and location might exist but is very weak. Similarly, we get r = -0.0015 from the correlation computation for age and location which indicates that the relationship between age and location might exist in an inverse way but still is regarded as very weak.

In conclusion, we realized that we still need to seek other potential relationships for other variables. Hence, we expanded the scope of our analysis and tried to explore our dataset referring to different years. By doing so, we produced two subsets after data filtration, they included the data in terms of book count collected in 2020 and 2021, respectively.

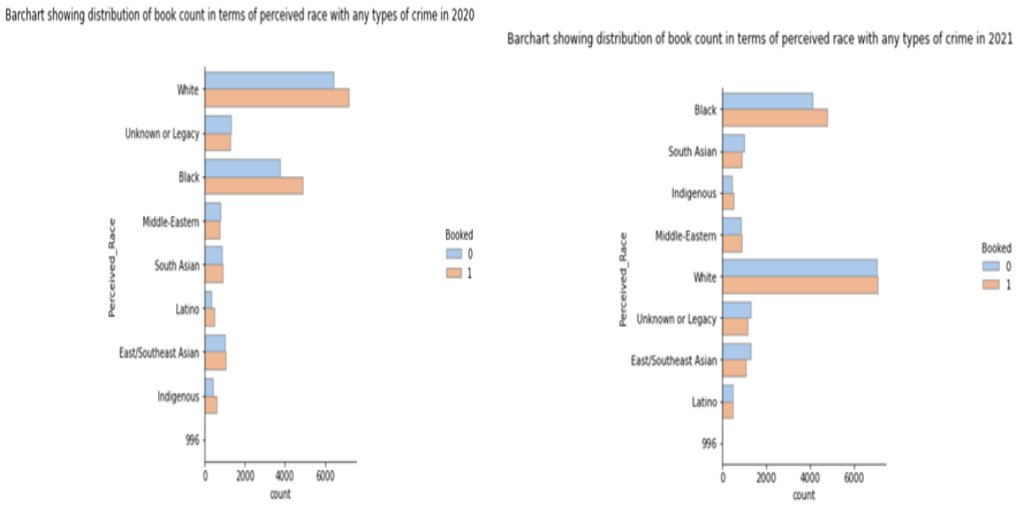

Fig: Bar chart showing booked count for perceived_race with types of crime in 2020 & 2021

Later, we produced two bar charts for showing the distribution of book count in terms of perceived race for both sexes with any types of crime in 2020 and 2021. As shown in figures above, we marked suspects who were booked as 1 and were not booked as 0. In the case of white and black suspects, as in 2020, the proportion of suspects who were arrested and booked that were higher than those who were arrested but not booked. However, in 2021, a trend that the proportion of white suspects who were arrested and booked and who were arrested but not booked was nearly the same. Further, we took a deeper dive into the subsets we created for each year, and we created frequency tables for each year and both sexes to see the exact number of booked counts in terms of perceived race.

Descriptive Statistics

From all tables, we had the exact number of cases for each race, in which white suspects were the maximum who were booked among perceived races in 2020 and 2021 for both sexes.

Perceived Race	Booked Count
Black	4206
East/Southeast Asian	917
Indigenous	419
Latino	421
Middle-Eastern	682
South Asian	803
Unknown or Legacy	1080
White	5772

Table: Number of males booked based on race for the year 2020

Perceived Race	Booked Count
Black	679
East/Southeast Asian	120
Indigenous	164
Latino	55
Middle-Eastern	59
South Asian	98
Unknown or Legacy	207
White	1413

Table: Number of females booked based on race for the year 2020

From the tables above we can see, the number of arrests made between males and females is significantly different for the year 2020. Generating statistics for the year 2021.

Perceived Race	Booked Count
Black	4120
East/Southeast Asian	947
Indigenous	337
Latino	429
Middle-Eastern	827
South Asian	785
Unknown or Legacy	986
White	5686

Table: Number of males booked based on race for the year 2021

Perceived Race	Booked Count
Black	667
East/Southeast Asian	110
Indigenous	154
Latino	60
Middle-Eastern	64
South Asian	91
Unknown or Legacy	163
White	1388

Table: Number of females booked based on race for the year 2021

The statistics obtained for both years show that the number of people being booked for different offenses has not really changed and are very close to each other. Generating further statistics to solidify the theory.

Next, we constructed four summary tables for displaying mean, std, and IQR values for both sexes in terms of booked count in 2020 and 2021. As shown in the following tables, the mean booked count of the male suspect from 2020 to 2021 was slightly decreased; meanwhile, the mean booked count of the female suspect from 2020 to 2021 was slightly decreased too.

	Male 2020	**Female 2020**	**Male 2021**	**Female 2021**
count	8	8	8	8
mean	1787.5	349.375	1764.625	337.125
std	2032.422692	475.298684	1995.013565	469.1282
min	419	55	337	60
25%	616.75	88.25	696	84.25
50%	860	142	887	132
75%	1861.5	325	1769.5	289
max	5772	1413	5686	1388

Table: representing statistics for males, females in 2020 and males, females in 2021

For this reason, we wanted to further examine the relationship between booked number and year, we created a boxplot showing distribution for the number of suspects booked categorized by perceived race and year. Besides, we also created a scatter plot to see if we could have the same relationships among those variables. In conclusion, based on previous data examinations, it was reasonable to believe that a strong relationship might exist between perceived race and booked counts, so we would further identify it based on T-test.

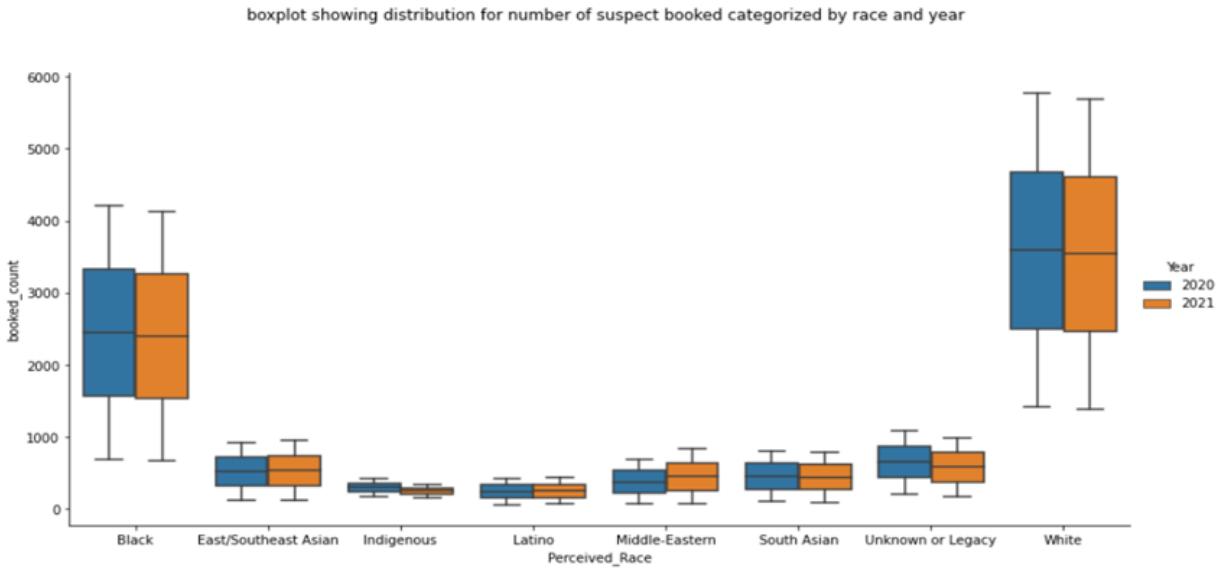

Fig: Boxplot showing distribution for the number of suspects booked, categorized by race & year

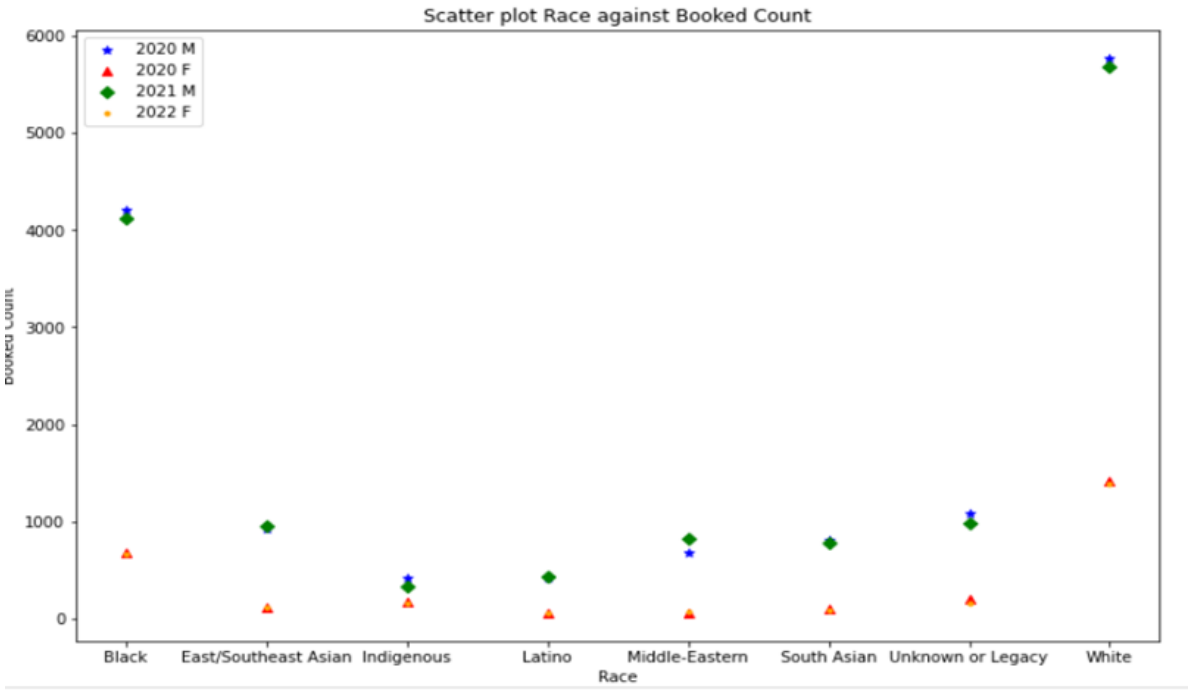

Fig: Scatterplot showing distribution for the no. of suspects booked, categorized by race & year

As we are also interested in the relationship between perceived race and cooperation level at arrest, in this case, we believed that only the suspects who were recorded as 'Actions_at_arrest___Cooperative' would be considered cooperative. Subsequently, we created a box plot and a scatterplot to gather the distribution for the mean of suspects being cooperative during arrest based on year. As shown in figure below, we observed that East/Southeast Asian suspects were more likely to be cooperative at arrest, compared to other perceived races in both 2020 and 2021. In terms of the comparison between years, we found that suspects were more likely to be cooperative at arrest in 2020, compared to 2021.

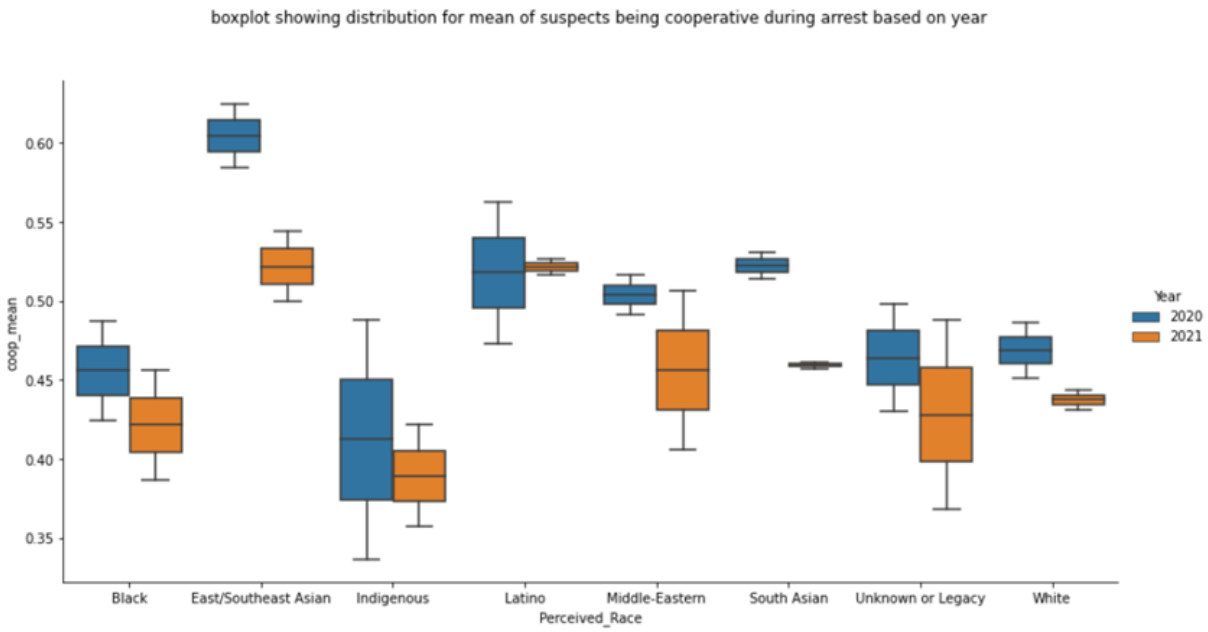

Fig: Boxplot showing distribution for mean of suspects being cooperative during arrest

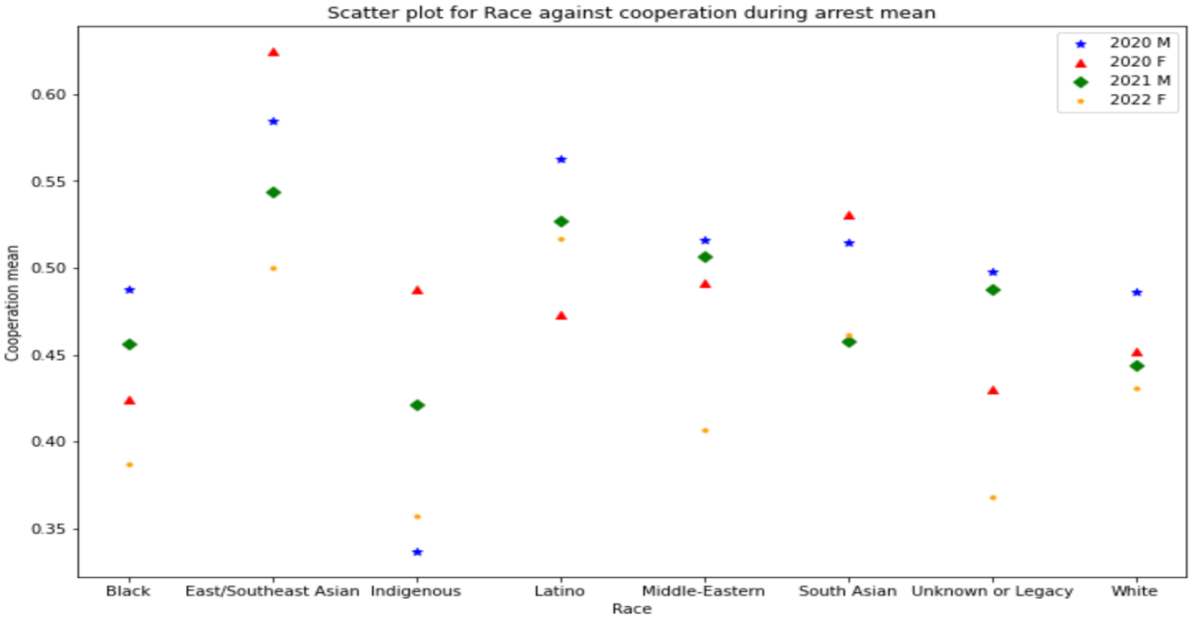

Fig: Scatterplot showing distribution for mean for suspects being cooperative during arrest

Shapiro-Wilk test

Firstly, we perform a Shapiro-Wilk test to check if our data in terms of our two research questions are normally distributed, we proposed the following hypothesis for both research questions:

H0 (Null Hypothesis): The sample is normally distributed

H1 (Alternative Hypothesis): The sample is not normally distributed

Setting an alpha level of 0.05, the results suggest that the count of suspects booked is not normally distributed with (p=8.13e-08), we reject the null hypothesis as the p-value is less than 0.05. However, for our second research question, the results suggest that the mean cooperation level of people while being arrested is normally distributed with (p=0.97), we fail to reject the null hypothesis as the p-value is greater than 0.05.

T-Test

As the above plots show data distribution very similar to each other based on the year, we ran a Welch's T-Test with the continuous variable in the data subsets created for the research questions. Before running the T-Tests, we checked the following assumptions were fulfilled: (1) a nominal two-level explanatory variable; (2) a quantitative outcome variable; (3) normality assumption; (4) independence of errors.

Research Question 1:

H0 (Null Hypothesis): The population mean of the two independent groups, the number of people booked in the year 2020 and the number of people booked in the year 2021, are equal.

H1 (Alternative Hypothesis): The population mean of the two independent groups, the number of people booked in the year 2020 and the number of people booked in the year 2021, are different.

In this trial, setting an alpha level of 0.05, since we fail to reject the null hypothesis as the p-value is greater than 0.05 (p=0.98), the results indicate that the number of people booked in 2020 was similar to the number of people booked in 2021; the mean of booked count in 2020 was 1068.44 that was higher than 1050.88 in 2021; the standard deviation for 2020 was 1607.68 that was higher than 1582.24 for 2021; the confidence interval for 2020 and 2021 was (-1134.13, 1169.25); we also performed a Welch DOF test, for which the results were 29.99.

T-Test concluding remark

As the p-value is greater than the set alpha level of 0.05, we fail to reject the null hypothesis which suggests that the number of people who were booked during the year of the COVID-19 outbreak was the same as the year 2021. This test was conducted with the ***limitation*** that the data was not normally distributed. However, as the sample size is greater than 50, the data distribution would have not a significant effect on the outcome of the test.

Research Question 2:

H0 (Null Hypothesis): The population mean of the two independent groups, mean co-operation level while being arrested in the year 2020 and mean co-operation level while being arrested in the year 2021, are the same.

H1 (Alternative Hypothesis): The population mean of the two independent groups, mean co-operation level while being arrested in the year 2020 and mean co-operation level while being arrested in the year 2021, are different.

In this trial, setting an alpha level of 0.05, since we fail to reject the null hypothesis as the obtained p-value is greater than 0.05 (p=0.09), the results indicated that the cooperation level in 2020 was like the cooperation level in 2021; the confidence interval for 2020 and 2021 was (-0.01, 0.08); we also created a Welch DOF test, the results was 29.16.

T-Test concluding remark

As the p-value is greater than the set alpha level of 0.05, we fail to reject the null hypothesis which suggests that the mean cooperation level for people being arrested during the year of the COVID-19 outbreak was the same as in the year 2021.

Research Design and Methods

We are using inferential statistics tests to explore further our research questions based on the findings from the descriptive statistics and the t-tests for both research questions.

Research Question 1:

Based on the boxplot and the scatterplot, we could observe that the number of people booked in the years 2020 and 2021 is very close to each other. The t-test also confirms these results as the obtained p-value is greater than 0.05 and therefore we cannot reject the null hypothesis. Expanding the scope of analysis to check if perceived race had any effect on the number of people being booked for any offense using one-way ANOVA. We will look if the combination of races has any effect on the mean difference in the number of people booked.

H0 (Null Hypothesis): the mean number of people being booked for any offense for all the eight race groups is equal.

H1 (Alternative Hypothesis): at least one mean group is different from the others for any of the race combinations for the number of people booked for any offense

The results of one-way ANOVA indicate that there exists a difference in at least one mean group of races on the number of people being booked. With the alpha level set at 0.05, this is a statistically significant difference as the obtained p-value (0.00293) is less than 0.05. Therefore, we reject the null hypothesis. Now we perform the post hoc test (Tukey's HSD) to identify which race group has the mean difference in the number of people being booked.

Results of Tukey's HSD-

Group 1	Group 2	Mean Difference	P-value	Reject Null Hypothesis?
Black	East/Southeast Asian	-1894.5	0.3503	FALSE
Black	Indigenous	-2149.5	0.2138	FALSE
Black	Latino	-2176.75	0.2019	FALSE
Black	Middle-Eastern	-2010	0.2827	FALSE
Black	South Asian	-1973.75	0.3028	FALSE
Black	Unknown or Legacy	-1809	0.407	FALSE
Black	White	1146.75	0.8446	FALSE
East/Southeast Asian	Indigenous	-255	0.9	FALSE
East/Southeast Asian	Latino	-282.25	0.9	FALSE
East/Southeast Asian	Middle-Eastern	-115.5	0.9	FALSE
East/Southeast Asian	South Asian	-79.25	0.9	FALSE
East/Southeast Asian	Unknown or Legacy	85.5	0.9	FALSE
East/Southeast Asian	White	3041.25	0.0241	TRUE
Indigenous	Latino	-27.25	0.9	FALSE
Indigenous	Middle-Eastern	139.5	0.9	FALSE
Indigenous	South Asian	175.75	0.9	FALSE
Indigenous	Unknown or Legacy	340.5	0.9	FALSE
Indigenous	White	3296.25	0.0119	TRUE
Latino	Middle-Eastern	166.75	0.9	FALSE
Latino	South Asian	203	0.9	FALSE
Latino	Unknown or Legacy	367.75	0.9	FALSE
Latino	White	3323.5	0.011	TRUE
Middle-Eastern	South Asian	36.25	0.9	FALSE
Middle-Eastern	Unknown or Legacy	201	0.9	FALSE

Middle-Eastern	White	3156.75	0.0175	TRUE
South Asian	Unknown or Legacy	164.75	0.9	FALSE
South Asian	White	3120.5	0.0194	TRUE
Unknown or Legacy	White	2955.75	0.0303	TRUE

The highlighted records in the table above represent the race groups where there exists a statistically significant difference in the number of people being booked. As we can see, the race groups, East/Southeast Asian & White, Indigenous & White, Latino & White, Middle-Eastern & White, South Asian & White and Unknown or Legacy & white have huge differences in their mean value for the number of people being booked.

One-Way ANOVA concluding remark

The above one-way ANOVA results presented some interesting facts. We found that there are some combinations of race groups that have statistically significant differences in the mean number of people being booked. This was further explored using Tukey's HSD test to determine which groups had the difference. The race groups East/Southeast Asian, Indigenous, Latino, Middle-Eastern, South Asian & Unknown or Legacy versus White had a mean difference of approximately 3000 people being booked.

Based on the test results from the t-test and one-way ANOVA for the data subset capturing the number of people booked for any offense, we know that there is no significant difference between the number of people booked in the year 2020 and 2021. However, there are some combinations of race groups where there is a statistically significant difference in the number of people booked. We now want to check if there is any interaction effect between the year and perceived race on the booked count. Performing two-way ANOVA to check if there exists any such effect using variables, (1) booked count (dependent variable), (2) perceived race (independent variable with 8 levels) and (3) year (independent variable with 2 levels).

Hypothesis 1-

H0 (Null Hypothesis): There is no difference in group means of people booked for any offense based on race for both years

H1 (Alternative Hypothesis): There exists a difference in group means of people booked for any offense based on race for both years

Hypothesis 2-

H0 (Null Hypothesis): There is no difference in group means of people booked for any offense for both years based on race

H1 (Alternative Hypothesis): There exists a difference in group means of people booked for any offense for both years based on race

Hypothesis 3-

H0 (Null Hypothesis): The effect of race does not depend on the effect of year on the number of people being booked, i.e., there is no interaction effect between race and year on the number of people being booked for any offense

H1 (Alternative Hypothesis): The effect of race does depend on the effect of year on the number of people being booked, i.e., there is an interaction effect between race and year on the number of people being booked for any offense

Results of two-way ANOVA-

	Sum Squared	Degree of Freedom	F-value	PR(>F)
Perceived Race	4.284098e+07	7	2.926133	0.035686
Year	2.467531e+03	1	0.001180	0.973025
Perceived Race : Year	1.581422e+04	7	0.001080	1.000000
Residual	3.346472e+07	16	NaN	NaN

The results of two-way ANOVA indicate that there exists a difference in the number of people booked for offense based on perceived race for both years. We can reject the null hypothesis and with the alpha level set at 0.05, this is a statistically significant difference as the obtained p-value (0.035686) is less than 0.05. However, for the second hypothesis, as the p-value (0.973025) is greater than 0.05, we cannot reject the null hypothesis. There is no difference in group means of people booked for any offense for both years based on race. Lastly, the result of the interaction effect shows that the effect of race does not depend on the effect of the year on the number of people booked for any offense. Therefore, there is no interaction effect between the perceived race and year on the number of booked counts.

Research Question 2:

We can see from the boxplot and scatterplot that there are some differences in the mean levels of cooperation during arrests made in 2020 and 2021. Yet, the t-test reveals that the mean cooperation level when being arrested in the years 2020 and 2021 are similar as the obtained p-value (0.0857) is more than 0.05, therefore we are unable to reject the null hypothesis. Using one-way ANOVA, we further examined whether perceived race had any impact on the mean amount of cooperation while being detained for any offense. We'll check to see if the combination of races affects the mean level of cooperation while being arrested.

H0 (Null Hypothesis): the mean level of cooperation for people being arrested for any offense for all 8 race groups is equal.

H1 (Alternative Hypothesis): at least one mean group is different from the others for any of the race combinations for the mean cooperation level while being arrested for any offense.

The results of one-way ANOVA show a difference in the mean level of cooperation when being arrested for at least one mean group of races. This difference is statistically significant as the alpha level is set at 0.05 because the resulting p-value (0.00279) is less than 0.05. Thus, we reject the null hypothesis. We now do the posthoc test to determine whether the racial group has the mean difference in the mean amount of cooperation while being arrested (Tukey's HSD).

Results of Tukey's HSD-

Group 1	Group 2	Mean Difference	P-value	Reject Null Hypothesis?
Black	East/Southeast Asian	0.1247	0.0245	TRUE
Black	Indigenous	-0.0379	0.9	FALSE
Black	Latino	0.0812	0.302	FALSE
Black	Middle-Eastern	0.0415	0.9	FALSE
Black	South Asian	0.0523	0.7616	FALSE
Black	Unknown or Legacy	0.0074	0.9	FALSE
Black	White	0.0145	0.9	FALSE
East/Southeast Asian	Indigenous	-0.1626	0.0018	TRUE
East/Southeast Asian	Latino	-0.0435	0.9	FALSE
East/Southeast Asian	Middle-Eastern	-0.0832	0.2751	FALSE
East/Southeast Asian	South Asian	-0.0724	0.4398	FALSE
East/Southeast Asian	Unknown or Legacy	-0.1173	0.0397	TRUE
East/Southeast Asian	White	-0.1102	0.0621	FALSE
Indigenous	Latino	0.1191	0.0355	TRUE
Indigenous	Middle-Eastern	0.0794	0.3265	FALSE
Indigenous	South Asian	0.0902	0.1939	FALSE
Indigenous	Unknown or Legacy	0.0453	0.8738	FALSE
Indigenous	White	0.0524	0.7608	FALSE
Latino	Middle-Eastern	-0.0396	0.9	FALSE
Latino	South Asian	-0.0288	0.9	FALSE
Latino	Unknown or Legacy	-0.0738	0.4165	FALSE

Latino	White	-0.0667	0.5329	FALSE
Middle-Eastern	South Asian	0.0108	0.9	FALSE
Middle-Eastern	Unknown or Legacy	-0.0341	0.9	FALSE
Middle-Eastern	White	-0.027	0.9	FALSE
South Asian	Unknown or Legacy	-0.0449	0.8795	FALSE
South Asian	White	-0.0379	0.9	FALSE
Unknown or Legacy	White	0.0071	0.9	FALSE

The racial groupings in which there is a statistically significant difference in the mean level of cooperation while being arrested are represented by the records in the table above that have been highlighted. As we can see, there is a mean difference in the levels of cooperation when being arrested between the Black and East/Southeast Asian, East/Southeast Asian and Indigenous, East/Southeast Asian and South Asian, and Indigenous and Latino race groups.

One-Way ANOVA concluding remark

The one-way ANOVA results above reveal some intriguing information. We discovered that several race group combinations exhibit statistical differences in the mean level of cooperation while being arrested. Tukey's HSD test was used to identify which groups had the difference to further investigate this. There was a statistically significant mean difference between the racial categories Black and East/Southeast Asian, East/Southeast Asian, and Indigenous, East/Southeast Asian, and South Asian, and Indigenous and Latino.

We now know that there is no statistically significant difference between the mean level for the years 2020 and 2021 based on the test results from the t-test. Using one-way ANOVA, we found that there are race group combinations where the mean level of cooperation when being arrested does differ statistically significantly across 4 racial groups. We now want to determine whether the perceived race and the year have any interaction effects on the level of cooperation. Performing two-way ANOVA to check if there exists any such effect using variables, (1) cooperation means (dependent variable), (2) perceived race (independent variable with 8 levels) and (3) year (independent variable with 2 levels).

Hypothesis 1-

H0 (Null Hypothesis): There is no difference in group means for cooperation level of people being arrested for any offense based on race for both years

H1 (Alternative Hypothesis): There exists a difference in group means for the cooperation level of people being arrested for any offense based on race for both years

Hypothesis 2-

H0 (Null Hypothesis): There is no difference in group means for the cooperation level of people being arrested for both years based on race

H1 (Alternative Hypothesis): There exists a difference in group means for the cooperation level of people being arrested for both years based on race

Hypothesis 3-

H0 (Null Hypothesis): The effect of race does not depend on the effect of year on the co-operation of people being arrested, i.e., there is no interaction effect between race and year on the co-operation level of people being arrested for any offense

H1 (Alternative Hypothesis): The effect of race does depend on the effect of year on the co-operation of people being arrested, i.e., there is an interaction effect between race and year on the co-operation level of people being arrested for any offense

Results of two-way ANOVA-

	Sum Squared	**Degree of Freedom**	**F-value**	**PR(>F)**
Perceived Race	0.072988	7	4.226957	0.008066
Year	0.012337	1	5.001134	0.039925
Perceived Race : Year	0.004756	7	0.275447	0.954652
Residual	0.039468	16	NaN	NaN

The results of the two-way ANOVA show that for both years, there is a difference in the mean level of cooperation while being arrested for any offense based on perceived race. With the alpha level set at 0.05, we can rule out the null hypothesis, and the resultant p-value (0.035686) is less than 0.05, indicating that the difference is statistically significant. As the p-value (0.039925) is smaller than our alpha level of 0.05 and the null hypothesis is rejected, we can see that there is a difference in group means for the number of people being arrested for both years based on race. Finally, the interaction effect results demonstrate that race's influence on the mean level of cooperation during an arrest is independent of the influence of year. Thus, the perceived race and year have no interaction effect on the mean cooperation level when arrested.

Results and Findings for Research Questions

Research Question 1

We now use an interaction plot to represent the above findings-

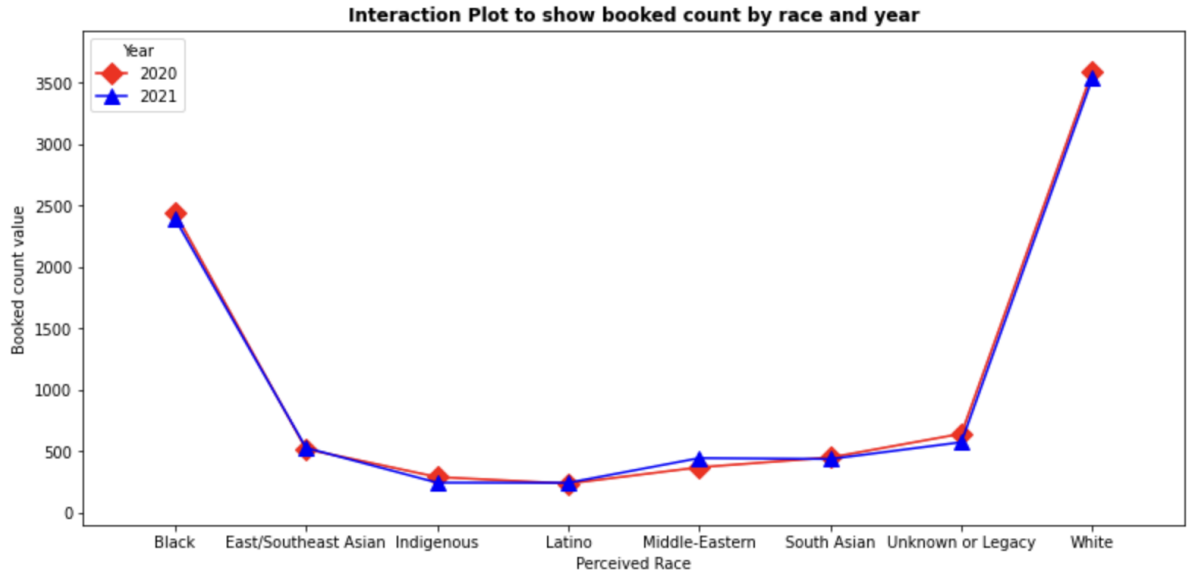

Fig: Interaction plot representing booked count based on race and year

The interaction plot above represents: (1) people with perceived race as White and Black constitute a very high proportion of individuals booked for an offense; (2) there is no difference in the number of individuals booked for both the years 2020 and 2021; (3) individuals with perceived race as White has the maximum count for people booked for both the years.

Two-way ANOVA concluding remark-

The two-way ANOVA results presented some interesting facts as we rejected the null hypothesis for hypothesis 1 where we obtained a p-value of 0.035686 which is lower than our threshold of 0.05. However, we cannot reject the null hypothesis for hypothesis 2 and 3 as their obtained p-values are 0.973025 and 1.000000 respectively which is above the set threshold limit. We found that tough there exist a difference in group means of people booked for any offense based on race for both years, perceived race had no effect on the year i.e., there was no interaction effect between them. This was further confirmed by the interaction plot which shows both lines mapped in a similar way. Also, from the results for two-way ANOVA, we can see that the p-value is less than the alpha level for categorical variable 'race'. Since, Tukey's HSD test is already performed for the variable, it was not repeated after two-way ANOVA. Based on the Tukey's HSD test to identify mean differences among perceived race groups, there were 6 out of 28 comparisons that showed statistically significant differences involving the white group. In other words, there were significant differences between the mean scores of white participants and participants of other perceived races in those 6 comparisons.

Research Question 2

We now use an interaction plot to represent the above findings-

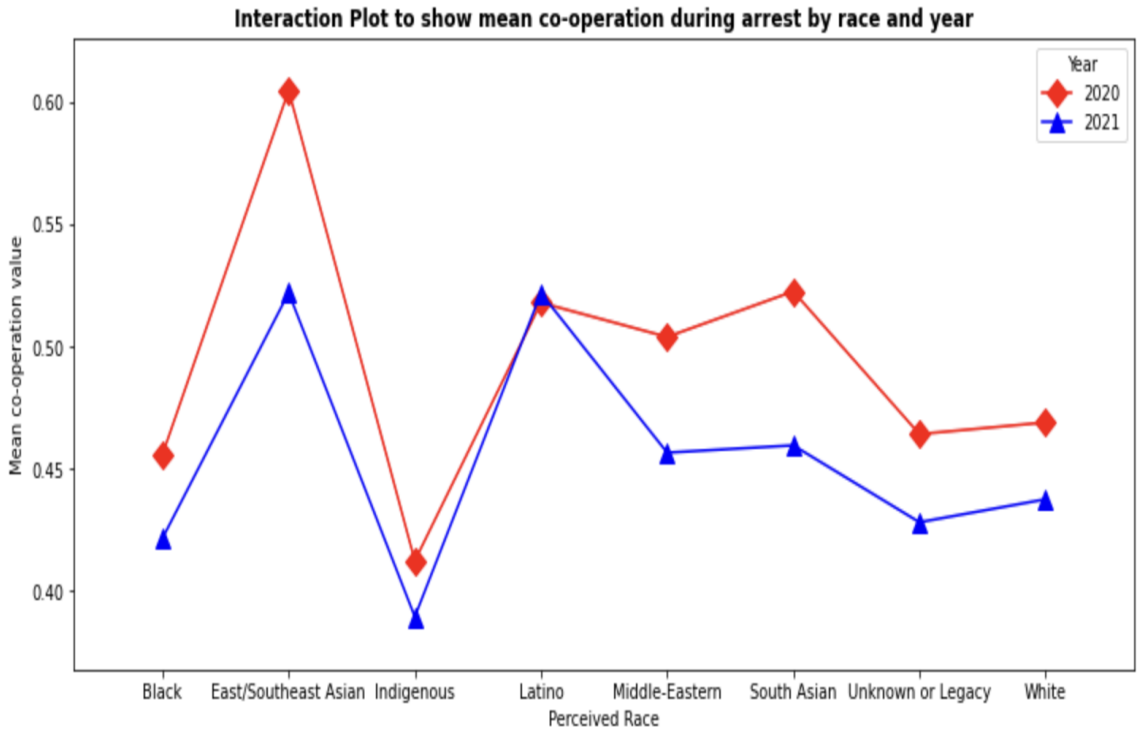

Fig: Interaction plot representing mean co-operation level based on race and year

The interaction plot above represents: (1) there is a huge drop in the co-operation level for perceived race, East/Southeast Asian from 2020 to 2021; (2) there is no change in the mean co-operation levels while being arrested for Latinos for both the years; (3) there is a drop in mean co-operation level for all the races from 2020 and 2021.

Two-way ANOVA concluding remark-

The two-way ANOVA results presented some interesting facts as we rejected the null hypothesis for hypothesis 1 and 2 as we obtained a p-value of 0.008066 and 0.039925 respectively which is lower than our set threshold level of 0.05. Although, we could not reject the null hypothesis for hypothesis 3 as we got the p-value of 0.954652 which is greater than 0.05. We discovered that though there is a difference between races and years in the mean level of cooperation when being arrested, perceived race has no influence on the year i.e., there is no interaction effect between them. Except for Latinos, all racial groups experience a decline in cooperation between 2020 and 2021, according to the interaction plot. Also, from the results for two-way ANOVA, we can see that the p-value is less than the alpha level for categorical variable 'race' and 'year'. As the variable year has only two levels, no post-hoc test was performed. Tukey's HSD test is already performed for the variable race therefore, it was not repeated after two-way ANOVA. Based on the Tukey's HSD test conducted above to find which combination of race groups had a statistically significant difference in their mean cooperation level, we found that there are 4 such groups out of 28.

Discussion

Many negative effects of the COVID-19 pandemic have been observed in various facets of life, including crime rates. To effectively design policies to lessen the effects of the epidemic, politicians and law enforcement officials must first understand how the pandemic has influenced crime trends. As a result, in our study, we examined how COVID-19 affected Toronto's crime statistics, paying particular attention to the factors that affected how many people were arrested and how cooperative they were.

During the pandemic, we wanted to see how race and the year affected crime trends. To evaluate the data and pinpoint the crucial factors that affect crime patterns, we applied statistical models. The findings of our study will give a thorough analysis and insightful information about how COVID-19 has affected Toronto's crime patterns. This information can be used to create plans that will lower crime rates and increase public safety. The creation of policies and programmes to aid the most vulnerable groups during the pandemic can also benefit from this information.

Overall, our study provides important insights into the complex relationship between COVID-19 and crime patterns in Toronto. It highlights the importance of considering demographic characteristics and other relevant factors when analyzing crime rates. We hope to contribute to the development of evidence-based strategies to mitigate the negative impacts of the pandemic on public safety.

Conclusion

Every element of life has been significantly impacted by the COVID-19 pandemic, including law enforcement procedures. Understanding how the pandemic has changed police procedures and how race may have an impact on those procedures is crucial. In our investigation, we explored the significance of race in these procedures as well as the effect of COVID-19 on arrest and strip search statistics in Toronto. We analysed the data using a variety of quantitative techniques to answer our research questions. For the years 2020 and 2021, we used statistics on people booked for an offense in Toronto, together with information on the race of those involved. The data on the average level of cooperation of the people during arrests was also calculated.

According to our analysis, COVID-19 had no significant impact on the amount of people who were arrested for any offence. Between the two years and the race, the count stayed largely the same. But we did see that the distribution of people by race was disproportionately high for White people, followed by Black people. This suggests that there might be racial discrepancies in Toronto's application of law enforcement practises.

Except for Latinos, who were detained in 2021, we saw a decline in the mean level of cooperation across all racial groups in response to our second research question. This shows that the connection between law enforcement personnel and those they apprehend may have been negatively impacted by the way activities were carried out during the pandemic.

Overall, our research offers crucial insights into how COVID-19 has affected Toronto's law enforcement practises and how race has shaped those practises. Our findings can assist in the development of evidence-based strategies to enhance public safety and address racial disparities in law enforcement practises by drawing attention to potential racial disparities and the detrimental effects of the pandemic on the relationship between law enforcement officials and those they arrest.

Reference

Meyer, M., Hassafy, A., Lewis, G., Shrestha, P., Haviland, A. M., & Nagin, D. S. (2022). Changes in crime rates during the COVID-19 pandemic. *Statistics and Public Policy*, *9*(1), 97–109. https://doi.org/10.1080/2330443x.2022.2071369

Pak, A., Adegboye, O. A., Adekunle, A. I., Rahman, K. M., McBryde, E. S., & Eisen, D. P. (2020). Economic consequences of the COVID-19 outbreak: The need for epidemic preparedness. *Frontiers in Public Health*, *8*. https://doi.org/10.3389/fpubh.2020.00241

Yu, Z., Razzaq, A., Rehman, A., Shah, A., Jameel, K., & Mor, R. S. (2021). Disruption in global supply chain and socio-economic shocks: A Lesson from covid-19 for sustainable production and consumption. *Operations Management Research*, *15*(1-2), 233–248. https://doi.org/10.1007/s12063-021-00179-y

Basu, R. (2021). *Democracy and public policy in the post-covid-19 world choices and outcomes*. Routledge.

Macklem, T. (2022, November 10). Restoring labour market balance and price stability. Bank of Canada. Retrieved February 26, 2023, from https://www.bankofcanada.ca/2022/11/restoring-labour-market-balance-and-price-stability/

White, A. (2020, April 9). How pandemics shape society. The Hub. Retrieved February 26, 2023, from https://hub.jhu.edu/2020/04/09/alexandre-white-how-pandemics-shape-society/

Rosenfeld, R., Abt, T., & Lopez, E. (2020). Online.ucpress.edu. Retrieved February 26, 2023, from https://online.ucpress.edu/fsr/article/33/1-2/72/115493/Pandemic-Social-Unrest-and-Crime-in-U-S-Cities