Write-up

1.Introduction

With the economic downturn and inflation brought on by the pandemic, Toronto's streets are beginning to see large numbers of homeless people and the number of indiscriminate attacks is gradually increasing. In the first week of February alone, there were three cases of indiscriminate assault in the Toronto area. One of the victims died, while another was a student at the University of Toronto. The research team hopes to find out the impact of gender and age group on suspects (those searched), as well as the distribution area and gender of mentally unstable suspects (usually the perpetrators of indiscriminate attacks are mentally unstable people) through the strip search dataset released by the Toronto Police Service. The method we would use is exploratory data analysis, an approach to analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. We will also do power analysis in this part. Power analysis is essential before conducting experiments because it helps researchers determine the required sample size to detect an effect with a specified level of confidence. This ensures that the study is adequately powered and minimizes the chances of making Type II errors (failing to detect a true effect when it exists). The specific hypotheses related to the research questions will be tested by ANOVA, which is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing each group's means and includes spreading the variance into diverse sources. In addition, ANCOVA and logistic regression will be also used in this research

2. Literature Review

2.1 Strip Search

The ongoing issue of unlawful strip-searching of women, especially by male law enforcement officers, continues to plague Canada despite firm legal decisions denouncing such practices. In 1995, the strip-searching of female inmates at the Prison for Women by male officers garnered national attention. The Arbour Inquiry described these strip searches as "cruel, inhumane, and degrading" and determined they breached the women's Charter rights. The matter of strip-searching resurfaced as a significant legal and policy concern in 2001 when the Supreme Court of Canada, in R v Golden, declared that strip searches should not be conducted as a standard procedure. The court also established a legal threshold that must be satisfied before initiating a strip search and outlined 11 precautions to ensure that police carry out strip searches lawfully. However, in 2008, more than ten years after the Arbour Inquiry and seven years following Golden, another incident of a female detainee (SB) being strip-searched by male officers occurred. This event made national headlines in 2010 when a court-mandated the release of the strip-search video footage and granted a stay of proceedings due to the Charter violations. (Psutka & Sheehy, 2016)

Moreover, if women perceive strip-searching as a form of sexual assault, it is probable that most will not report the incident. For instance, in the case of Oklahoma City police officer Daniel Holtzclaw, who was prosecuted for raping at least 12 women in his custody, most victims only came forward after the police identified them as potential victims during the investigation. It is unrealistic to expect women who have undergone illegal strip-searching to report the misconduct to the very institution that enabled it. Women subjected to strip-searching by police may encounter similar obstacles to those strip-searched by prison guards when attempting to report the abuse.

Debra Parkes, a professor and former president of the Elizabeth Fry Society of Manitoba, identifies several factors that discourage women in prisons from reporting strip searches. These include a lack of knowledge about their rights, practical barriers like illiteracy, limited access to legal aid, the belief that filing a complaint would be fruitless, and fear of retaliation from correctional staff. The study referenced in this article took place in Manitoba, where over 73% of the female prison population is Aboriginal. The various forms of oppression, both individual and systemic, that Aboriginal women face may create additional barriers to reporting unauthorized strip-searching incidents to the authorities.

Kirkup (2014) explores the complex interaction between non-normative genders and sexualities within the intricate realm of criminal procedure. Based on an in-depth analysis of Forrester v. Peel (Regional Municipality) Police Services Board, a recent case where a transgender detainee alleged discrimination in services based on "sex," the article links strip searches to a broader system of bodily control. Transgender bodies are targeted not just because they are viewed as different, but also due to the meaning behind that difference: defiance of the systems that enforce a strict, essentialist gender binary. Consequently, transgender bodies become crucial sites for concurrent observation, normalization, and scrutiny by not only law enforcement but also the wider society.

2.2 Distribution area of mentally unstable suspects

The research (Wang, Lee, & Williams, 2019) investigates the spatial patterns of property and violent crime in neighborhoods within Canada's largest city, examining how specific offender characteristics and neighborhood socioeconomic conditions relate to crime rates. The spatial analyses confirm previous research findings that crime is concentrated in neighborhoods with particular characteristics rather than being randomly distributed. The spatial distribution of property and violent crimes follows a U-shaped pattern across the city, generally aligning with neighborhoods of lower socioeconomic status. The OLS and GWR models confirm the significant relationship between crime rates and various dimensions of the Ontario Marginalisation Index, supporting the social disorganization theory that links crime to neighborhood socioeconomic disadvantage. Property and violent crimes tend to cluster in similar neighborhoods over the three-year study period (2014-2016), with consistent crime concentrations in the city center and high concentrations in the northwest and northeast parts of the city.

Although the present study finds that distance to crime is not a significant predictor of neighborhood crime rates for each crime type, property crime offenders travel longer distances to crime locations compared to violent crime offenders. This finding is consistent with previous research and has implications for understanding the routine activity theory and rational choice framework. The study finds that violent crimes cluster in specific neighborhoods, which explains the shorter distance traveled for violent crimes compared to property crimes. The research is unique in considering both neighborhood socioeconomic characteristics and individual demographic characteristics of offenders charged with property and violent crimes at the neighborhood level from 2014 to 2016. The OLS regression finds that neighborhood socioeconomic conditions measured by the Ontario Marginalisation Index are significantly related to property and violent crime. The study also finds that individual characteristics of offenders are associated with property and violent crime in various ways. The relationships suggest the complexity of how offender characteristics are connected to property and violent crime occurrences and call for further exploration of other offender characteristics and contextual factors that might improve understanding of property and violent criminal behavior in Toronto.

Lam (2021) also proposes that 21.9% of Toronto's population lives below the poverty line. Neighborhoods like Bay Street Corridor, Moss Park, and Kensington-Chinatown have a higher percentage of residents living in poverty, while Waterfront Communities-The Island has a lower percentage compared to the city's average. Racial demographics also differ greatly across these neighborhoods. Bay Street Corridor and Kensington Chinatown have a significantly higher percentage of visible minorities compared to Toronto's average of 50.1%. Moreover, Bay Street Corridor has the highest percentage of non-permanent residents at 20.1%, as opposed to the city's average of 3.5%. In Moss Park, 2% of residents identify as Aboriginal, which is higher than the city's average of 0.9%. These figures highlight the economic disparities within the Greater Toronto Area neighborhoods. Historically, poverty has been a justification for increased police surveillance, often framed as a fight against drug abuse. Regrettably, these measures have disproportionately affected low-income communities, perpetuating the cycle of poverty. The lack of resources for struggling families, coupled with Toronto's unaffordable housing, exacerbates these negative circumstances. Additionally, the influence of racism in Canada can be linked to the overrepresentation of crime in the four neighborhoods discussed. The lasting consequences of racism, stemming from a long history of colonization and genocide, continue to impact

Indigenous communities. Indigenous citizens are more than twice as likely to have frequent, involuntary contact with police. Similarly, though with different historical backgrounds, black communities in Canada have faced comparable discriminatory practices, including police surveillance. Black Lives Matter Toronto recently protested the "Community Contracts Policy," which authorized the practice of requesting identification from individuals in public spaces. This policy disproportionately affected black communities, as they accounted for 27% of incidents while only representing 8.5% of Toronto's population.

3. Method

3.1 Dataset description

The primary source of the research is a dataset from Arrests and Strip Searches by Toronto Police Service; this dataset contains information on all arrests and strip searches. The dataset also includes information on whether a person was taken to a police station within 24 hours of their arrest. However, some records may indicate a strip search but not a booking due to issues with the booking template. In these cases, it should be assumed that a booking did take place. This database was initially collected and created by Toronto Police Service. The 'Arrests_and_Strip_Searches_(RBDC-ARR-TBL-001).csv' contains the records of 65,276 cases from 2020 to 2021. In this research, our main variables are:

strip search: a type of search conducted by a police officer, wherein they remove some or all of a person's clothing and visually inspect their body.

Age group: The age of the person who was arrested or strip-searched is their age at the time of the arrest, as reported to the arresting officer.

Year: 2020 and 2021

Sex: Male (M) and Female (F)

In order to make this dataset available, we compute the amount of each group (divided by year, sex, and age group) and create a new data frame to do the ANOVA test.

3.2 EDA Method

The method we would use is exploratory data analysis, an approach to analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. The specific hypotheses related to the research questions will be tested by ANOVA, ANCOVA, and logistic regression, which is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing each group's means and includes spreading the variance into diverse sources.

3.3 Power Analysis

Power analysis is essential before conducting experiments because it helps researchers determine the required sample size to detect an effect with a specified level of confidence. This ensures that the study is adequately powered and minimizes the chances of making Type II errors (failing to detect a true effect when it exists). Back to our research, we define the power analysis for the independent samples t-test using smp.TTestIndPower() firstly. After that, we solve for the required sample size using the solve_power() method with the desired effect size, power, and alpha (significance level) as input parameters. Now, we can plot the power curve for different effect sizes and sample sizes using the plot_power() method. This visualization helps to understand the relationship between sample size, effect size, and power. The power analysis performs for a specific example by defining the number of runs, effect size, and using the solve_power() method to find the required sample size. Then, we create a function get_t_result() that generates random samples and performs a two-sample t-test to test for a difference between groups. It runs a simulation with the determined sample size and store the p-values in a DataFrame. The power analysis also calculates the proportion of p-values below the alpha level (0.05), which represents the power of the test in this simulation. After that, we calculate the effect size (Cohen's d) using the pooled standard deviation and the means of the samples from a pilot study. The power analysis performs to find the required sample size for the given effect size, alpha, and power. Finally, we can view power curves for a range of effect sizes and sample sizes to visualize the relationship between these parameters.

4. Results

4.1 Exploratory Data Analysis (EDA)

4.1.1 Descriptive Analysis

A total of 65,276 cases were included in the dataset from 2020 to 2021, of which 2,179 were mentally unstable or suicidal suspects. The number of cases in which the suspect was a female mentally unstable person ranged from a low of 1 to a high of 105, while the number of male mentally unstable suspects ranged from a low of 6 to a high of 254. The average number of female cases was 13.9 and the average number of male cases was 46.4.

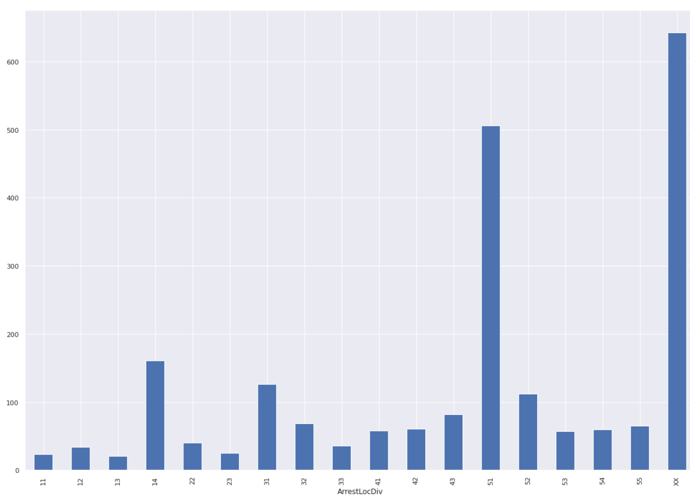

Figure 4-1

Block 51 had the highest number of mentally unstable persons of all cases recorded in the search, more than double that of Block 14, which ranked second. In addition, the bar chart also told us that the police have a large number of cases involving mentally unstable suspects where the location was not recorded or where it occurred is vague.

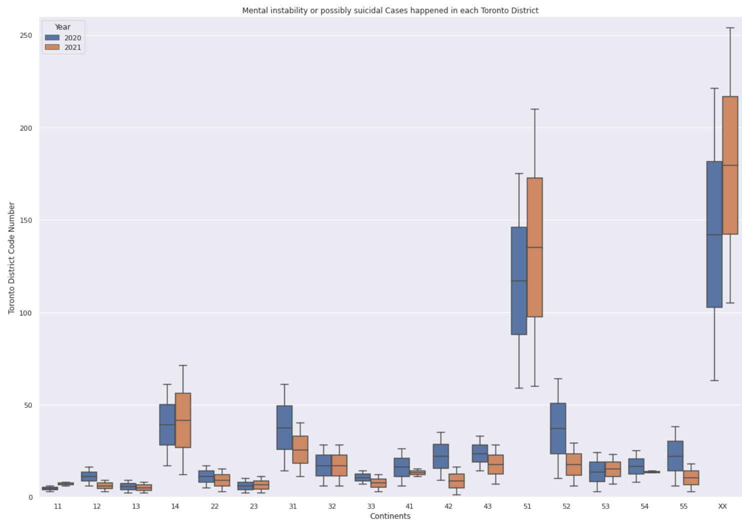

Figure 4-2

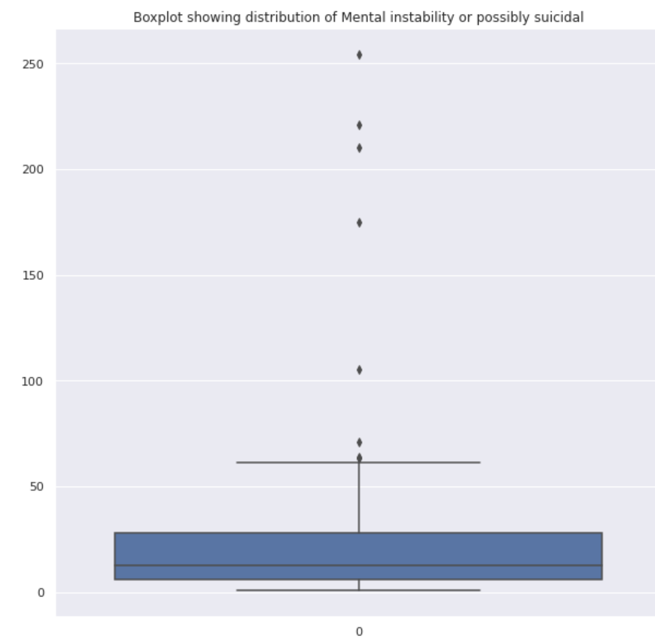

Figure 4-3

Box plots show the distribution of quantitative data to facilitate comparisons between variables or between levels of categorical variables. Boxes show the quartiles of the data set, while whisker extensions show the remainder of the distribution, except for points identified as "outliers" using the interquartile function method. The box plots show cases involving mentally unstable or suicidal people in 17 Toronto blocks (excluding unknown location XX) in 2020 versus 2021. The advantage of the box plot is that the observer can visualize the range between the median and quartiles of the corresponding data, as well as its outliers. The box plot shows that the median number of cases occurring in each neighborhood is relatively stable and stays below 50, except for Block 51. If we consider the cases involving mentally unstable people in 2020 and 2021 as a whole, there is a small difference between the upper and lower quartiles, and the larger outliers should be brought about by Block 51 and unknown locations.

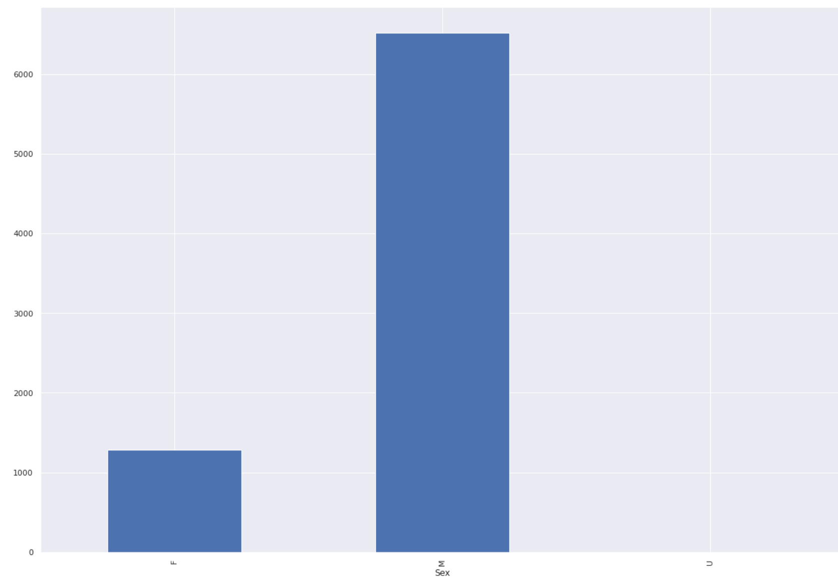

Figure 4-4 Strip Search on different sex

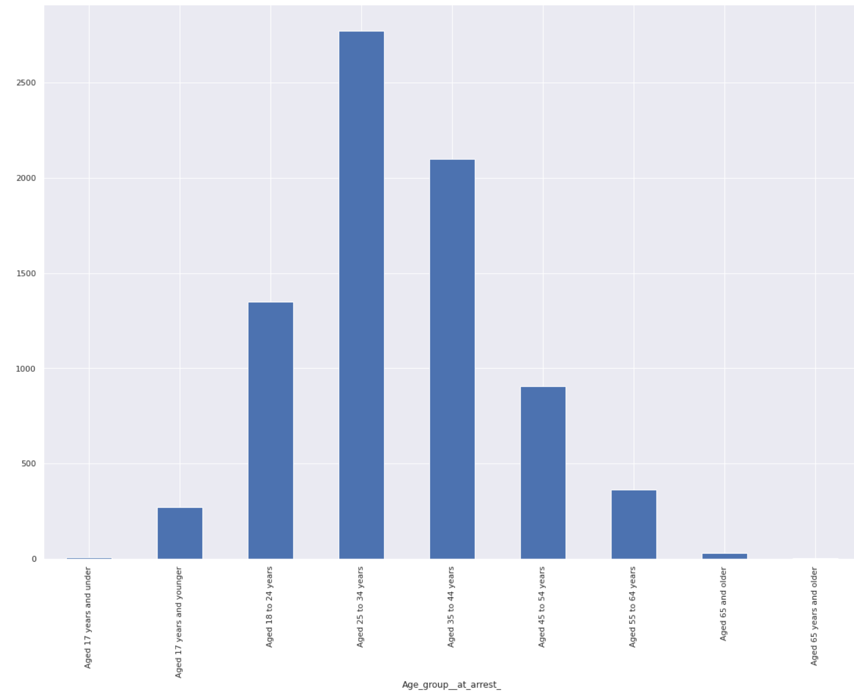

Figure 4-5 Strip Search on different age groups

Besides, in the 65276 cases, there were 7801 cases with a strip search, including 1283 for females and 6518 for males. For different age groups, 25-34 and 35-44 have the most cases of a strip search, while teenagers and elders have the least cases.

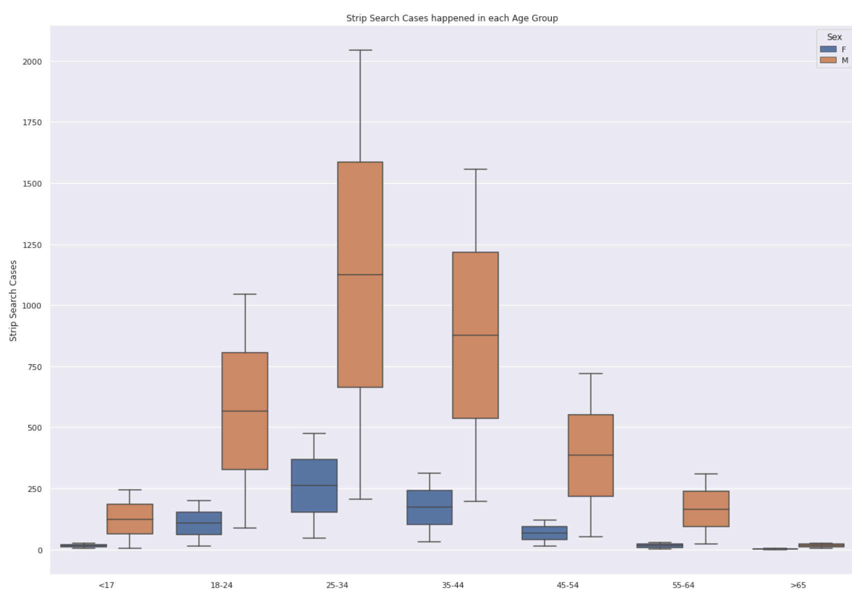

Figure 4-6

4.1.2 T-test

For the first T-test, our hypotheses are:

H0: There is no difference between men and women in all police-recorded cases involving mental instability and suicidal tendencies.

H1: There is a significant difference between men and women in all police-recorded cases involving mental instability and suicidal tendencies.

The results showed that there were more males (M=46.4, SD=63.4) than females (M=13.9, SD=22.1) in all police-recorded cases involving mental instability and suicidal ideation. With an alpha value determined to be 0.05, this is a statistically significant difference as the p-value (0.006) is less than 0.05, 95% CI [9.97, 55.08]. Therefore, we can reject the null hypothesis that there is no difference between men and women in all police-recorded cases involving mental instability and suicidal tendencies.

Then, we do unpaired two-sample T-tests to see whether Strip Search varies in different groups of a variable. Our hypotheses are:

H0: There is no significant difference in strip search cases between males and females.

H1: There is a significant difference in strip search cases between males and females.

As a result, we can see that the p-value is equal to 0.044, less than 0.05, so we can reject the null hypothesis, and conclude that there is a significant difference in strip search between males and females.

4.2 ANOVA test

Then, we presented the two-way ANOVA to test for the number of strip searches among different sex and age groups. In the first ANOVA test, our hypotheses are:

H0: There is no significant main effect of Sex or Age group at arrest on StripSearch.

H1: There is a significant main effect of Sex or Age group at arrest on StripSearch.

H2: There is a significant interaction effect between Sex and Age group at arrest on StripSearch.

	sum_sq	df	F	PR(>F)
C(Sex)	9.79E+05	1	3.893754	0.068542
C(Age_group__at_arrest_)	1.56E+06	6	1.033677	0.444279
C(Sex):C(Age_group__at_arrest_)	6.09E+05	6	0.403802	0.864393
Residual	3.52E+06	14	NaN	NaN

Table 4-1 Two-way ANOVA: Effect of Sex and Age group at arrest on StripSearch

From Table 4-1, all three p-values are higher than 0.05, so we cannot reject the null hypothesis. Therefore, there is no significant main effect of the Sex or Age group at arrest and no significant interaction effect between the two factors on StripSearch.

In the second ANOVA test, our hypotheses are:

H0: There is no significant main effect of Year or Age group at arrest on StripSearch.

H1: There is a significant main effect of Year or Age group at arrest on StripSearch.

H2: There is a significant interaction effect between Year and Age group at arrest on StripSearch.

	sum_sq	df	F	PR(>F)
C(Age_group__at_arrest_)	1.56E+06	6	1.383141	0.28775
C(Year)	1.48E+06	1	7.857849	0.014092
C(Age_group__at_arrest_):C(Year)	1.00E+06	6	0.887887	0.529137
Residual	2.63E+06	14	NaN	NaN

Table 4-2 Two-way ANOVA: effect of Year and Age group at arrest on StripSearch

From Table 4-2, two p-values are higher than 0.05, so we cannot reject the null hypothesis. Therefore, there is a significant main effect of Year on StripSearch, but no significant main effect of Age group at arrest on StripSearch.

In the third ANOVA test, our hypotheses are:

H0: There is no significant main effect of Sex or Year on StripSearch.

H1: There is a significant main effect of Sex or Year on StripSearch.

H2: There is a significant interaction effect between Sex and Year on StripSearch.

	sum_sq	df	F	PR(>F)
C(Sex)	9.79E+05	1	6.622096	0.016678
C(Year)	1.48E+06	1	9.98732	0.004228
C(Sex):C(Year)	6.64E+05	1	4.490743	0.044619
Residual	3.55E+06	24	NaN	NaN

Table 4-3 Two-way ANOVA: effect of Sex and Year on StripSearch

From Table 4-3, all three p-values are lower than 0.05, so we can reject the null hypothesis. Therefore, there is a significant main effect of Sex or Year on StripSearch, and also a significant interaction effect between Sex and Year on StripSearch.

Then, we presented the two-way ANOVA to test for the amount of all police-recorded cases involving mental instability and suicidal ideation among different sex and ArrestLocDiv. In the fourth ANOVA test, our hypotheses are:

H0: There is no significant main effect of Sex or ArrestLocDiv at arrest on mental instability cases.

H1: There is a significant main effect of Sex or ArrestLocDiv at arrest on mental instability cases.

H2: There is a significant interaction effect between Sex and ArrestLocDiv at arrest on mental instability cases.

	sum_sq	df	F	PR(>F)
C(ArrestLocDiv)	123491.9028	17	72.511372	1.02E-22
C(Sex)	19045.01389	1	190.10689	6.20E-16
C(ArrestLocDiv):C(Sex)	30601.23611	17	17.968284	1.03E-12
Residual	3606.5	36	NaN	NaN

Table 4-4 Two-way ANOVA: effect of Sex and ArrestLocDiv on mental instability cases

From Table 4-4, all three p-values are lower than 0.05, so we can reject the null hypothesis. Therefore, there is a significant main effect of Sex or ArrestLocDiv at arrest on mental instability cases, and also a significant interaction effect between Sex and ArrestLocDiv at arrest on mental instability cases.

4.3 Post-hoc tests

Then, we do Tukey's HSD as the post hoc tests for the strip search cases grouped by Sex and Year.

Group 1	Group 2	Meandiff	P-adj	Lower	Upper	Reject
female_2020	female_2021	-151.2857	0.8746	-718.202	415.6306	FALSE
female_2020	male_2020	681.8571	0.0143	114.9408	1248.7735	TRUE
female_2020	male_2021	-85.2857	0.9	-652.202	481.6306	FALSE
female_2021	male_2020	833.1429	0.0024	266.2265	1400.0592	TRUE
female_2021	male_2021	66	0.9	-500.9163	632.9163	FALSE
male_2020	male_2021	-767.1429	0.0053	-1334.0592	-200.2265	TRUE

Table 4-5 Tukey HSD: Means of StripSearch Grouped by Sex and Year

Thus, we would conclude that there is a statistically significant difference between the means of groups female_2020 and male_2020, groups female_2021 and male_2020, and groups male_2020 and male_2021, but not a statistically significant difference between the means of groups female_2020 and female_2021, groups female_2020 and male_2021, and groups female_2021 and male_2021.

4.4 Interaction Plot

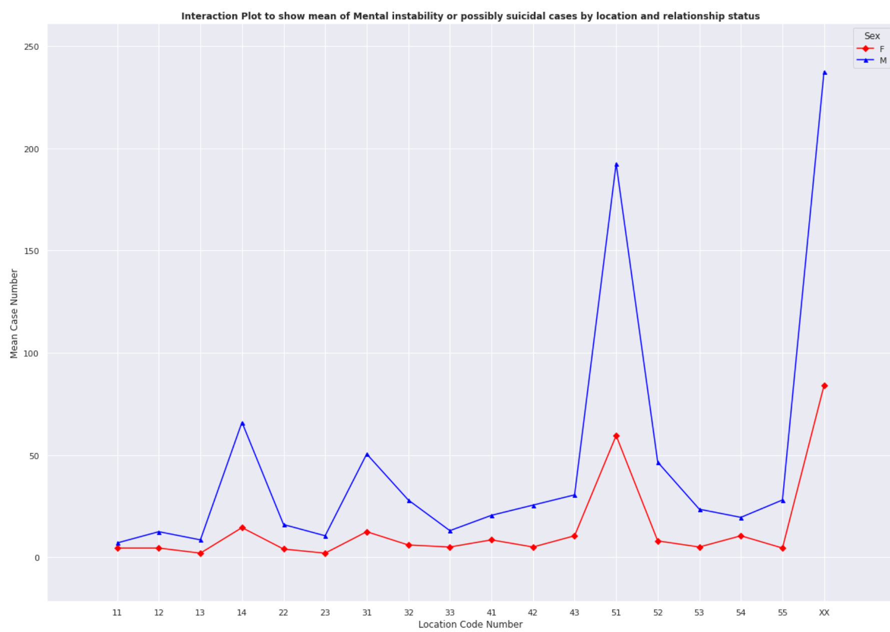

Figure 4-7

Although the interaction chart does not provide any information about statistically significant differences, the chart shows that. (a) males are more present in mentally unstable and suicidal cases compared to females; (b) the higher the number of cases in the neighborhood, the greater the difference in the percentage of males and females involved

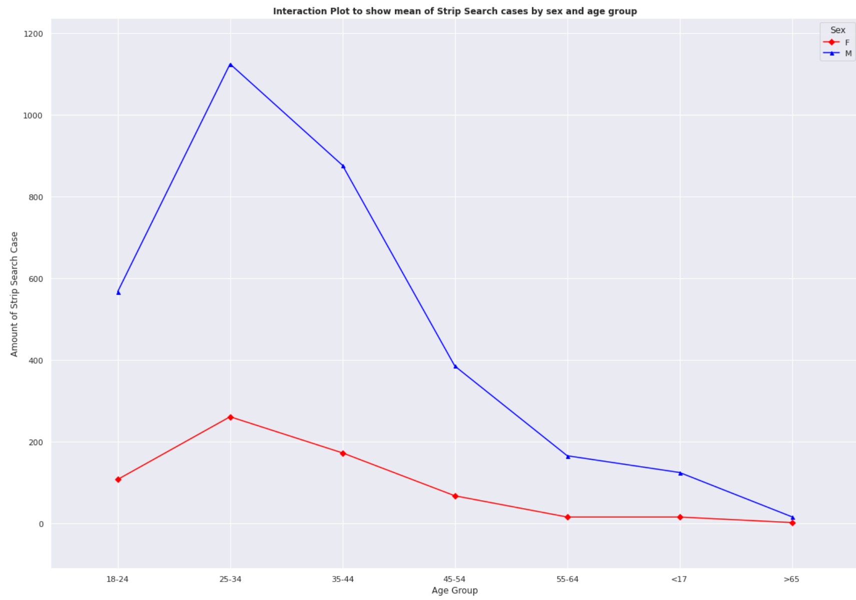

Figure 4-8

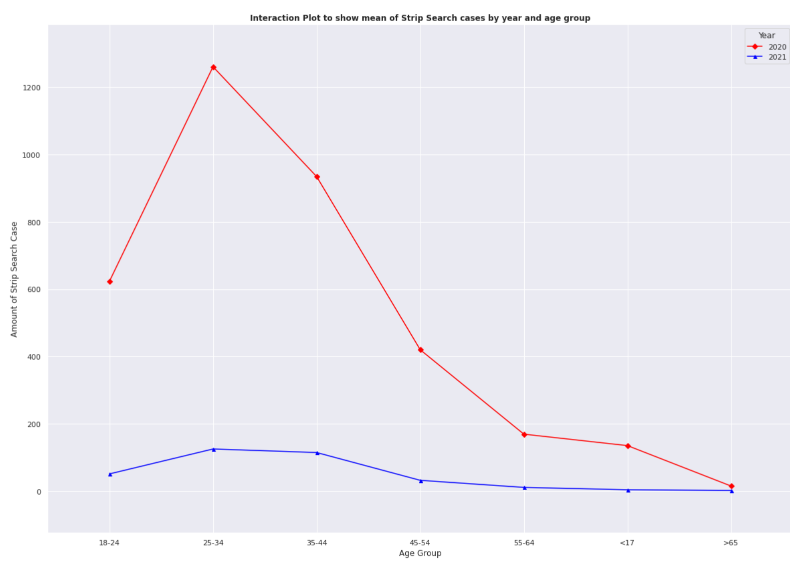

Figure 4-9

Based on the research, we found that although there is a significant difference in strip search between males and females, there is no significant main effect of the Sex or Age group at arrest and no significant interaction effect between the two factors on strip search. When we consider the Year and Age group as independent variables, only Year (instead of Age group and the interaction between Year and Age group) shows the main effect on StripSearch.

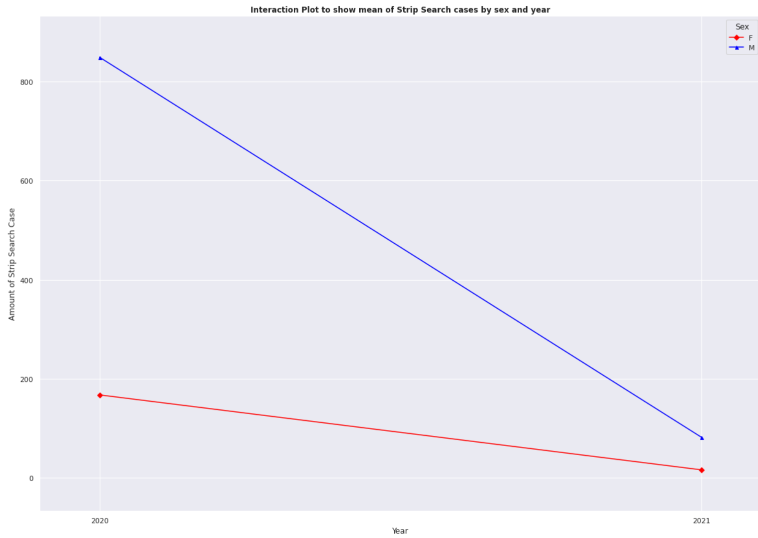

Figure 4-10

However, there is a significant main effect of Sex or Year on StripSearch, and also a significant interaction effect between Sex and Year on StripSearch, especially between the means of groups female_2020 and male_2020, groups female_2021 and male_2020, and groups male_2020 and male_2021.

4.5 ANCOVA

Firstly, we analyze the effect of Sex on the Mental_inst_count while controlling for the covariate Year.

OLS Regression Results						
Dep. Variable:	Mental_inst_count	R-squared:	0.108			
Model:	OLS	Adj. R-squared:	0.082			
Method:	Least Squares	F-statistic:	4.170			
Date:	Wed, 12 Apr 2023	Prob (F-statistic):	0.0195			
Time:	18:42:11	Log-Likelihood:	-379.06			
No. Observations:	72	AIC:	764.1			
Df Residuals:	69	BIC:	771.0			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	14.3750	9.758	1.473	0.145	-5.092	33.842
Sex[T.M]	32.5278	11.268	2.887	0.005	10.049	55.006
Year[T.2021]	-0.9167	11.268	-0.081	0.935	-23.395	21.562
Omnibus:	64.457	Durbin-Watson:	1.838			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	303.921			
Skew:	2.866	Prob(JB):	1.01e-66			
Kurtosis:	11.274	Cond. No.	3.19			

The analysis aimed to investigate the effect of sex (male or female) on the number of mental institution counts (Mental_inst_count) while controlling for the year (2020 or 2021). The results indicate that the overall model is statistically significant (F-statistic = 4.170, p = 0.0195) and can explain about 10.8% of the variance in mental institution counts (R-squared = 0.108). When comparing the two sexes, it appears that being male is associated with a significant increase in the number of mental institution counts (coef = 32.53, p = 0.005). Specifically, the number of mental institution counts for males is estimated to be 32.53 units higher than for females, on average, when controlling for the year. However, the effect of the year is not statistically significant (coef = -0.92, p = 0.935), which means that there is no substantial difference in mental institution counts between the years 2020 and 2021 when accounting for the effect of sex.

Secondly, we analyze the effect of Sex on the Mental_inst_count while controlling for the area.

OLS Regression Results 4						
Dep. Variable:	StripSearch	R-squared:	0.381			
Model:	OLS	Adj. R-squared:	0.164			
Method:	Least Squares	F-statistic:	1.756			
Date:	Wed, 12 Apr 2023	Prob (F-statistic):	0.153			
Time:	18:42:11	Log-Likelihood:	-206.35			
No. Observations:	28	AIC:	428.7			
Df Residuals:	20	BIC:	439.3			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-10.5139	13.051	-0.806	0.424	-36.69	15.663
Sex[T.M]	32.5278	5.988	5.432	0.000	20.517	44.538
ArrestLocDiv[T.12]	2.7500	17.964	0.153	0.879	-33.282	38.782
ArrestLocDiv[T.13]	-0.5000	17.964	-0.028	0.978	-36.532	35.532
ArrestLocDiv[T.14]	34.5000	17.964	1.920	0.060	-1.532	70.532
ArrestLocDiv[T.22]	4.2500	17.964	0.237	0.814	-31.782	40.282
ArrestLocDiv[T.23]	0.5000	17.964	0.028	0.978	-35.532	36.532
ArrestLocDiv[T.31]	25.7500	17.964	1.433	0.158	-10.282	61.782
ArrestLocDiv[T.32]	11.2500	17.964	0.626	0.534	-24.782	47.282
ArrestLocDiv[T.33]	3.2500	17.964	0.181	0.857	-32.782	39.282
ArrestLocDiv[T.41]	8.7500	17.964	0.487	0.628	-27.282	44.782
ArrestLocDiv[T.42]	9.5000	17.964	0.529	0.599	-26.532	45.532
ArrestLocDiv[T.43]	14.7500	17.964	0.821	0.415	-21.282	50.782
ArrestLocDiv[T.51]	120.2500	17.964	6.694	0.000	84.218	156.282
ArrestLocDiv[T.52]	21.5000	17.964	1.197	0.237	-14.532	57.532
ArrestLocDiv[T.53]	8.5000	17.964	0.473	0.638	-27.532	44.532
ArrestLocDiv[T.54]	9.2500	17.964	0.515	0.609	-26.782	45.282
ArrestLocDiv[T.55]	10.5000	17.964	0.584	0.561	-25.532	46.532
ArrestLocDiv[T.XX]	155.0000	17.964	8.628	0.000	118.968	191.032
Omnibus:	14.420	Durbin-Watson:	1.681			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	60.469			
Skew:	0.041	Prob(JB):	7.40E-14			
Kurtosis:	7.489	Cond. No.	21.5			

This analysis investigates the relationship between mental institution counts (Mental_inst_count) and several independent variables, including sex (male or female), arrest location division (ArrestLocDiv), and the interaction between them. The model explains a substantial proportion of the variance in mental institution counts, with an R-squared value of 0.806, indicating that 80.6% of the variability in the dependent variable can be accounted for by the predictors. The results show that sex has a significant effect on mental institution counts (coef = 32.53, $p < 0.001$). On average, the number of mental institution counts for males is estimated to be 32.53 units higher than for females. The arrest location division also has a significant impact on mental institution counts. Specifically, divisions with codes '14' (coef = 34.5, p = 0.060), '31' (coef = 25.75, p = 0.158), '51' (coef = 120.25, p < 0.001), '52' (coef = 21.5, p = 0.237), and 'XX' (coef = 155, p < 0.001) show noticeable differences in mental institution counts compared to the reference category (division '11'). The division with code 'XX' has the highest impact on mental institution counts, with an estimated 155 units higher than the reference category. It is essential to note that some coefficients are not statistically significant, as indicated by their p-values. These non-significant coefficients suggest that there may not be a substantial difference in mental institution counts across certain arrest location divisions when controlling for other factors. The model's assumptions, such as normality and homoscedasticity, are also evaluated. The Omnibus test (p = 0.001) and the Jarque-Bera test (p = 7.40e-14) both indicate that the residuals may not be normally distributed. Therefore, further investigation or model refinement might be necessary to ensure the validity of the analysis.

Thirdly, we analyze the effect of Sex on the StripSearch while controlling for the covariate Year.

OLS Regression Results 3

Dep. Variable:	StripSearch	R-squared:	0.368			
Model:	OLS	Adj. R-squared:	0.318			
Method:	Least Squares	F-statistic:	7.287			
Date:	Wed, 12 Apr 2023	Prob (F-statistic):	0.00321			
Time:	18:42:11	Log-Likelihood:	-206.62			
No. Observations:	28	AIC:	419.2			
Df Residuals:	25	BIC:	423.2			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	321.2500	134.339	2.391	0.025	44.573	597.927
Sex[T.M]	373.9286	155.122	2.411	0.024	54.449	693.408
Year[T.2021]	-459.2143	155.122	-2.96	0.007	-778.694	-139.735
Omnibus:	19.236	Durbin-Watson:	1.838			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26.703			
Skew:	1.579	Prob(JB):	1.59e-06			
Kurtosis:	6.593	Cond. No.	3.19			

This OLS regression analysis examines the relationship between the number of strip searches (StripSearch) and the independent variables sex (male or female) and year (2020 or 2021). The model explains 36.8% of the variance in the number of strip searches, with an R-squared value of 0.368. The results indicate that sex has a significant effect on the number of strip searches (coef = 373.93, p = 0.024). On average, the number of strip searches for males is estimated to be 373.93 units higher than for females. Year is also a significant factor in the number of strip searches (coef = -459.21, p = 0.007). The number of strip searches in 2021 is estimated to be 459.21 units lower than in 2020. The model's assumptions, such as normality and homoscedasticity, are also evaluated. The Omnibus test (p = 0.000) and the Jarque-Bera test (p = 1.59e-06) both suggest that the residuals may not be normally distributed. Therefore, further investigation or model refinement might be necessary to ensure the validity of the analysis.

Finally, we analyze the effect of Sex on the StripSearch while controlling for Age_group__at_arrest.

OLS Regression Results 4

Dep. Variable:	StripSearch	R-squared:	0.381			
Model:	OLS	Adj. R-squared:	0.164			
Method:	Least Squares	F-statistic:	1.756			
Date:	Wed, 12 Apr 2023	Prob (F-statistic):	0.153			
Time:	18:42:11	Log-Likelihood:	-206.35			
No. Observations:	28	AIC:	428.7			
Df Residuals:	20	BIC:	439.3			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	150.2857	242.844	0.619	0.543	-356.279	656.85
Sex[T.M]	373.9286	171.717	2.178	0.042	15.733	732.124
Age_group__at_arrest_[T.25-34]	355.5000	321.253	1.107	0.282	-314.622	1025.622
Age_group__at_arrest_[T.35-44]	187.2500	321.253	0.583	0.566	-482.872	857.372
Age_group__at_arrest_[T.45-54]	-111.0000	321.253	-0.346	0.733	-781.122	559.122
Age_group__at_arrest_[T.55-64]	-246.7500	321.253	-0.768	0.451	-916.872	423.372
Age_group__at_arrest_[T.<17]	-267.2500	321.253	-0.832	0.415	-937.372	402.872
Age_group__at_arrest_[T.>65]	-328.2500	321.253	-1.022	0.319	-998.372	341.872
Omnibus:	9.509	Durbin-Watson:	3.333			
Prob(Omnibus):	0.009	Jarque-Bera (JB):	8.034			
Skew:	1.024	Prob(JB):	0.0180			
Kurtosis:	4.640	Cond. No.	8.88			

This table presents the results of an OLS regression that investigates the relationship between the number of strip searches (StripSearch) and the independent variables sex, year, and age group at the time of arrest. The model explains 38.1% of the variance in the number of strip searches, with an R-squared value of 0.381. However, the adjusted R-squared, which accounts for the number of independent variables, is lower at 0.164. The results show that sex has a significant effect on the number of strip searches (coef = 373.93, p = 0.042). On average, the number of strip searches for males is estimated to be 373.93 units higher than for females. The other independent variables, age groups at the time of arrest, do not show statistically significant effects on the number of strip searches. Their p-values are all greater than 0.05, indicating that the observed relationships may be due to chance. The model's assumptions are also evaluated. The Omnibus test (p = 0.009) and the Jarque-Bera test (p = 0.0180) both suggest that the residuals may not be normally distributed. The Durbin-Watson statistic is 3.333, which could indicate the presence of negative autocorrelation among the residuals. Further investigation or model refinement might be necessary to ensure the validity of the analysis.

4.6 Logistic regression

In this study, we conducted a logistic regression analysis to investigate the relationship between the likelihood of a strip search (StripSearch) and various independent variables, including gender (Sex), age group at the time of arrest (Age_group__at_arrest_), and arrest year (Arrest_Year). Our dataset contained 52,201 observations.

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	5345.9212	94.627	56.495	0	5160.455	5531.387
Sex[T.M]	0.2364	0.038	6.224	0	0.162	0.311
Sex[T.U]	-2.1042	3.144	-0.669	0.503	-8.267	4.059
Age_group__at_arrest_[T.Aged 17 years and younger]	0.7002	0.39	1.797	0.072	-0.063	1.464
Age_group__at_arrest_[T.Aged 18 to 24 years]	1.2088	0.383	3.155	0.002	0.458	1.96
Age_group__at_arrest_[T.Aged 25 to 34 years]	1.2764	0.382	3.339	0.001	0.527	2.026
Age_group__at_arrest_[T.Aged 35 to 44 years]	1.318	0.382	3.446	0.001	0.568	2.067
Age_group__at_arrest_[T.Aged 45 to 54 years]	0.9386	0.384	2.446	0.014	0.187	1.691
Age_group__at_arrest_[T.Aged 55 to 64 years]	0.6617	0.387	1.711	0.087	-0.096	1.42
Age_group__at_arrest_[T.Aged 65 and older]	-0.5023	0.432	-1.163	0.245	-1.349	0.344
Age_group__at_arrest_[T.Aged 65 years and older]	0.0645	0.629	0.103	0.918	-1.168	1.297
Arrest_Year	-2.6478	0.047	-56.55	0	-2.74	-2.556

Table 4-6 Logistic Regression 1

The logistic regression model, with an accuracy of 0.88, revealed several significant associations between the independent variables and the likelihood of a strip search. Notably, males (Sex[T.M]) were found to have a significantly higher probability of being subjected to a strip search compared to the reference group (presumably females), with a coefficient of 0.2364 ($p < 0.001$). However, the results for the unknown gender category (Sex[T.U]) were not statistically significant ($p = 0.503$).

With respect to age groups at the time of arrest, all age groups except those aged 65 years and older (Age_group__at_arrest_[T.Aged 65 years and older]) and aged 65 and older (Age_group__at_arrest_[T.Aged 65 and older]) demonstrated a significantly higher likelihood of being subjected to a strip search compared to the reference group. The highest likelihood of a strip search was found among individuals aged 35 to 44 years (Age_group__at_arrest_[T.Aged 35 to 44 years]), with a coefficient of 1.3180 ($p = 0.001$).

Furthermore, our results indicated that the likelihood of a strip search has decreased over the years. The coefficient for the arrest year (Arrest_Year) was -2.6478, which was statistically significant ($p < 0.001$). In conclusion, our logistic regression analysis demonstrated that gender, age group at the time of arrest, and arrest year were significantly associated with the likelihood of a strip search. The findings suggest that males and younger age groups (below 65 years) are more likely to experience a strip search. Additionally, the likelihood of strip searches has decreased over time.

Then, we conduct a logistic regression analysis to investigate the factors influencing actions taken at arrest concerning individuals with mental health issues. The dependent variable was "Actions_at_arrest___Mental_inst," and 20 independent variables were considered in the analysis, with a total of 52,201 observations. The model's pseudo-R-squared value was 0.04299, suggesting that the model accounts for approximately 4.3% of the variance in the dependent variable.

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	109.4668	99.686	1.098	0.272	-85.915	304.848
Sex[T.M]	-0.2557	0.059	-4.35	0	-0.371	-0.14
Sex[T.U]	2.1265	0.896	2.374	0.018	0.371	3.882
ArrestLocDiv[T.12]	0.3708	0.298	1.246	0.213	-0.212	0.954
ArrestLocDiv[T.13]	0.4728	0.328	1.443	0.149	-0.169	1.115
ArrestLocDiv[T.14]	1.442	0.248	5.821	0	0.956	1.928
ArrestLocDiv[T.22]	0.5401	0.291	1.859	0.063	-0.029	1.11
ArrestLocDiv[T.23]	0.1448	0.319	0.454	0.65	-0.48	0.77
ArrestLocDiv[T.31]	1.3857	0.255	5.438	0	0.886	1.885
ArrestLocDiv[T.32]	0.8839	0.27	3.275	0.001	0.355	1.413
ArrestLocDiv[T.33]	1.0287	0.303	3.395	0.001	0.435	1.623
ArrestLocDiv[T.41]	0.5029	0.275	1.829	0.067	-0.036	1.042
ArrestLocDiv[T.42]	0.9098	0.274	3.315	0.001	0.372	1.448
ArrestLocDiv[T.43]	0.8163	0.265	3.08	0.002	0.297	1.336
ArrestLocDiv[T.51]	2.129	0.237	8.99	0	1.665	2.593
ArrestLocDiv[T.52]	1.2678	0.257	4.941	0	0.765	1.771
ArrestLocDiv[T.53]	1.2063	0.276	4.374	0	0.666	1.747
ArrestLocDiv[T.54]	1.1897	0.273	4.354	0	0.654	1.725
ArrestLocDiv[T.55]	1.2053	0.271	4.447	0	0.674	1.736
ArrestLocDiv[T.XX]	0.4917	0.235	2.09	0.037	0.031	0.953
Arrest_Year	-0.0562	0.049	-1.139	0.255	-0.153	0.04

Table 4-7 Logistic Regression 2

The model, with accuracy of 0.96, revealed several significant factors that influenced actions taken at arrest concerning individuals with mental health issues. Male individuals (Sex[T.M]) were less likely to be subjected to such actions compared to female individuals (coef = -0.2557, $p < 0.001$). Furthermore, individuals with an unknown sex (Sex[T.U]) were more likely to be subjected to these actions compared to female individuals (coef = 2.1265, $p = 0.018$).

Additionally, several divisions of arrest locations were identified as significant predictors. Individuals arrested in locations 14, 31, 32, 33, 41, 42, 43, 51, 52, 53, 54, and 55 were more likely to receive actions related to mental health issues compared to the reference location. Among these, location 51 showed the strongest association with the dependent variable (coef = 2.1290, $p < 0.001$).

It is important to note that the Arrest_Year variable was not significant in predicting the dependent variable (coef = -0.0562, $p = 0.255$), suggesting that there was no evident trend across years regarding actions taken at arrest for mental health issues.

In conclusion, this logistic regression analysis highlighted the influence of sex and arrest location on actions taken at arrest concerning individuals with mental health issues. However, the model's relatively low explanatory power implies that other factors not considered in this analysis may also contribute to these outcomes. Future research should aim to incorporate additional variables to better understand the complex interplay between factors influencing actions taken at arrest for individuals with mental health issues.

6. Discussion

The findings of this project have to be seen in the light of some limitations. Most importantly, the inequality in our group-level data. In this analysis, we chose to perform two-way ANOVA to test the differences between group means. However, the sample sizes of each group are not equal. Although it is not one of the assumptions made in an ANOVA, it may still lead to problems, such as it may reduce the statistical power or the robustness to unequal variance. Despite the data inequality, several improvements could be made for future studies. For example, a three-way ANOVA test could be used later to solve the problem of unbalanced data: regarding "Year", "Sex", and "Age Group" as the three categorical variables, which may show the difference among each group clearer and more detailed.

7. Conclusion

In conclusion, after the icon, t-test, and ANOVA tests, we were able to determine that: firstly, gender and year had a significant effect on the strip search data while different age groups did not. Second, among all strip searches, cases involving mentally unstable and suicidal individuals were significantly influenced by gender and district. Toronto's 14 and 51 Blocks were the most frequent locations for related cases. In addition, the number of strip searches decreases each year, but the number of mentally unstable suspects increases each year. This is the reason for the increase in indiscriminate attacks felt by Toronto citizens in the news. Based on the data set and preliminary analysis, the research team recommends that Toronto citizens take care to avoid Block 14 and Block 51 and pay more attention to men who are acting abnormally when walking outside.

This ANCOVA demonstrates that sex plays a significant role in mental institution counts and strip searches. Males tend to have higher mental institution counts and strip searches than females. While year did not affect mental institution counts, it did have a significant impact on the number of strip searches. The arrest location division was also found to be a significant factor in mental institution counts. Further investigation or model refinement might be necessary to address the potential violation of assumptions such as normality, homoscedasticity, and autocorrelation in the analyses. Future research may also explore other factors influencing mental institution counts and strip searches or investigate potential strategies to address disparities between different sexes and arrest location divisions.

The logistic regression analyses in this study revealed that gender, age group at the time of arrest, and arrest year were significantly associated with the likelihood of a strip search, while sex and arrest location were influential factors in actions taken at arrest for individuals with mental health issues. The findings suggest that males and younger individuals are more likely to experience a strip search, and the likelihood of strip searches has decreased over time. Meanwhile, males were found to be less likely to receive actions related to mental health issues at arrest compared to females, and arrest location played a role in determining these actions. Although these results offer valuable insights, the relatively low explanatory power of the second model implies that other factors not considered in this analysis may also contribute to the outcomes. Future research should aim to incorporate additional variables to better understand the complex interplay between factors influencing actions taken at arrest for individuals with mental health issues.

Reference:

Charron, M. (n.d.). Neighbourhood Characteristics and the Distribution of Police-reported Crime in the City of Toronto. In *peelregion.ca* (Issue 18). Retrieved April 10, 2023, from https://www.peelregion.ca/health/youth/pdfs/Neighbourhood-Characteristics.pdf

Kirkup, K. (2009). Indocile Bodies: Gender Identity and Strip Searches in Canadian Criminal Law*. *Canadian Journal of Law and Society / La Revue Canadienne Droit et Société*, *24*(1), 107–125. https://doi.org/10.1017/S0829320100009790

Lam, R. (2021). Ten Neighbourhoods in Toronto with the Highest Total Crime Experience Vastly Different Crime Rates. *Society-Mag.Com*. https://society-mag.com/goto/https://tellingstorieswithdata.com/inputs/pdfs/paper_one-2021-Rachael_Lam.pdf

Psutka, M., & Sheehy, E. (2016). STRIP-SEARCHING OF WOMEN IN CANADA: WRONGS AND RIGHTS. *The Canadian Bar Review*, *94*(2). https://cbr.cba.org/index.php/cbr/article/view/4376

Wang, L., Lee, G., & Williams, I. (2019). The Spatial and Social Patterning of Property and Violent Crime in Toronto Neighbourhoods: A Spatial-Quantitative Approach. *ISPRS International Journal of Geo-Information 2019, Vol. 8, Page 51*, *8*(1), 51. https://doi.org/10.3390/IJGI8010051