**Analysis of Stop, Question, and Frisk Policies in New York City: A Comprehensive Study**

Santhi Kiran Chavali

College of Professional Studies, Northeastern University

Master of Professional Studies in Analytics, ALY6015

Professor Roy Wada

Due Date: 12/7/2022

**Abstract**

The stop-question-and-frisk program, sometimes known as stop-and-frisk, is a tactic used by the New York City Police Department to briefly hold, question, and occasionally search bystanders and suspects on the street for weapons and other contraband. In other regions of the country, this is referred to as the Terry stop. 12,404 people were stopped as a result of the stop-and-frisk program in 2016. The stop-and-frisk program used to be implemented on a considerably bigger scale. Over 100,000 persons were imprisoned annually between 2003 and 2013, reaching a peak of 685,724 in 2011. Over 99% of all non-lethal interpersonal violence in the city that was ever officially recorded was committed by the New York Police Department at its height.

Keywords: stop-and-frisk, Terry stop

**Introduction**

The New York Police Department (NYPD) donated this stop, question, and frisk dataset for this analysis. It comprises details regarding the NYPD's "stop, question, and frisk" program. A police officer may stop someone, question them, and conduct a weapon search if they are suspected of criminality. This controversial policy has received harsh criticism. The practice has been roundly denounced as racial profiling. The complete dataset from 2003 to 2019 contains information about this policy. It has a record of each stop, including details about the occasion, the area, the officer, the subject, and other elements.

**Analysis**

**Exploratory Data Analysis**

This data set has 12,012 rows and 31 variables. To start the analysis, we first check if the data set has any null values. On analyzing we found that there are no null values present in the data set. Now, we can proceed to the statistical analysis and visualization for exploratory data analysis. Before we start plotting tables and graphs to understand more about the data, we add 2 more variables to the graph which will further help in visualization and analysis. The 2 columns are “Decade” and “Area”. “Area” is a variable that groups the precincts into different areas of New York City.

**Summary Statistics of the population data set.**

Table no 1 shows the descriptive statistics of the entire population. Where we can see that the average number of people stopped by the police from the year 2003 to 2015 is almost 400, whereas the number of people frisked where only 209 out of which only 34 people were searched and only 24 people were arrested.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table no 1 Descriptive Stats Table for Population** | | | |
|  | mean | sd | median |
| stopped | 399.05 | 1255.47 | 41 |
| arrested | 24.38 | 65.52 | 3 |
| frisked | 209.17 | 720.52 | 13 |
| searched | 34.18 | 101.78 | 3 |

**Summary Statistics of the data set grouped by Gender**

Table no 2 shows the descriptive statistics of the data set grouped by gender male and female. The data set includes an equal number of males and females who were stopped by the NYPD. Here we can see that the average number of males arrested is 44 whereas only 4 females were arrested, bringing a male-to-female ratio of 11:1. Similar statistics are observed in frisked and searched as well where 402 males were frisked whereas only 15 females were frisked, and 64 males were searched whereas only 3 females were searched. This brings the male-to-female ratio of 134:5 frisked and 64:3 searched.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table no 2 Descriptive Statistics of the data set grouped by Gender** | | | | | | |
|  | Male (N=6006) | | | Female (N=6006) | | |
|  | mean | sd | median | mean | sd | median |
| stopped | 741.17 | 1702.81 | 157 | 56.91 | 137.71 | 12 |
| arrested | 44.11 | 87.75 | 10 | 4.64 | 10.35 | 1 |
| frisked | 402.97 | 980.74 | 72 | 15.37 | 37.19 | 3 |
| searched | 64.49 | 137.14 | 14 | 3.86 | 8.60 | 1 |

**Summary Statistics of the data set grouped by Race**

Table no 3 shows the descriptive statistics of the data set grouped by race. The different races given in this data set are "A": Asian and Pacific Islander, "B": Black, "I": American Indian/Alaskan Native, "P": White-Hispanic, "Q": Black-Hispanic, "W": White. Here we can see that the black population holds the highest average of arrests, searched, frisked, and stopped by the NYPD whereas the American Indian/Alaskan Native has the lowest average of arrests, searched, frisked, and stopped.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table no 3 Descriptive Statistics of the data set grouped by Race** | | | | | | |
| race | A | B | I | P | Q | W |
| stopped | 77.08 | 1297.81 | 9.78 | 152.96 | 608.42 | 248.22 |
| arrested | 4.82 | 76.96 | 0.47 | 10.37 | 37.89 | 15.77 |
| frisked | 32.62 | 704.39 | 4024 | 87.67 | 327.2 | 98.92 |
| searched | 6.28 | 108.07 | 0.72 | 14.09 | 55.29 | 20.62 |

**Percentage of people frisked, searched, and arrested**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table no 4 Percentage of people frisked, searched, and arrested** | | | | |
| **Race** | **Total\_stops** | **Frisked** | **Searched** | **Arrested** |
| Black | 2598217 | 54.27 | 8.33 | 5.93 |
| Black-Hispanic | 1218049 | 53.78 | 9.09 | 6.23 |
| White | 496939 | 39.85 | 8.31 | 6.35 |
| Asian | 154316 | 42.32 | 8.15 | 6.25 |
| American Indian/Alaskan Native | 19584 | 43.31 | 7.32 | 4.84 |
| White-Hispanic | 306232 | 57.32 | 9.21 | 6.78 |

**Total Arrests in the Areas within New York City**

**Chart, bar chart

Description automatically generated**

The Bronx has the highest number of arrests with total 54678 arrests followed by Manhattan North with 53027 arrests in total. The Bronx and Manhattan North can be termed as the most unsafe areas.

**Trend for arrests over Years**

**Chart

Description automatically generated**

Arrests increased from 2002 to 2010, then decreased significantly from 2010 to 2015, with fewer arrests than ever before. People of race Black(B) and Black-Hispanic(Q) have seen the highest number of arrests when compared to other races. Concerns about racial discrimination have recently increased, as evidenced by the graph, which shows that in 2015, the number of arrests among races differed only slightly. Furthermore, males were arrested far more frequently than females in the first decade than in the second.

**People Stopped Over Years**

**Chart, bar chart

Description automatically generated**

The year 2011 saw the most arrests, with 349k Black people and 174k Black-Hispanic people arrested. White people are arrested in significant numbers as well, with the highest number being 61.5k in 2011.

**People Frisked Over Years**

**Chart, bar chart

Description automatically generated**

In Brooklyn North, many people of the Black race are frisked. A significant number of Black-Hispanic as well as White people have been frisked in Queens North.

**Models**

**ANOVA testing**

Using a one-way ANOVA test, we are going to answer the following question. In one-way ANOVA testing, we primarily state the null and alternate hypotheses. In this situation, our confidence interval is 95 which states that our alpha value is 0.05

**Question: With a confidence interval of 95% (alpha =0.05) can we say that there is enough evidence to conclude that a difference in mean arrests exists among the different casts?**

**H0 = There is a difference in the mean arrests among the races.**

**Ha = There is no significant difference in the mean arrests among the races.**

**Alpha: 0.05**

Applying the ANOVA test:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table no. 4 One-way ANOVA test | | | | | |
|  | DF | Sum sq | Mean sq | f-stat | p-value |
| Type | **5** | 8351871 | 1670374 | 464.1 | 2e-16 \*\*\* |
| Residuals | **12006** | 43216021 | 3600 |  |  |

**Observations:** Since the p-value 2e-16 > the alpha value = 0.05, we reject the null hypothesis. At a 95% confidence interval, we can conclude that there is not enough evidence to conclude that there is a difference in the mean arrests among the races.

**Linear Regression Model**

Using linear regression, we are going to predict the number of arrests in the NYPD. First, we start by fitting all the variables in the regression model and selecting the variables that are correlated to the target variable “arrests”. Below is the table we obtained after feature selection.

|  |  |
| --- | --- |
| **Table no 5 Linear Regression Model** | |
|  | *Dependent variable:* |
|  |  |
|  | arrested |
|  | |
| searched | 0.121\*\*\* |
|  | (0.004) |
|  |  |
| contrabn | 1.264\*\*\* |
|  | (0.018) |
|  |  |
| weapnfnd | 0.147\*\*\* |
|  | (0.019) |
|  |  |
| pf\_weapn | -1.260\*\*\* |
|  | (0.040) |
|  |  |
| pf\_hcuff | 0.819\*\*\* |
|  | (0.015) |
|  |  |
| Constant | 1.308\*\*\* |
|  | (0.160) |
|  |  |

|  |  |
| --- | --- |
| **Table no 6 Linear Regression Model** | |
| Observations | 8,408 |
| R2 | 0.957 |
| Adjusted R2 | 0.957 |
| Residual Std. Error | 13.709 (df = 8402) |
| F Statistic | 37,523.720\*\*\* (df = 5; 8402) |
| P-value | 2.2e-16 |

**Interpretation:**

The above table is a linear regression model fitted to predict the arrests made by the NYPD. From this table, we can see that our degree of freedom is 8402, and the p-value is 2.2e-16 which is smaller than our confidence interval of 95% (alpha = 0.05). The obtained value of R-squared is 95.7%. and an f-statistic value of 37,523.720. The \*\*\* on the f-statistic value represents that the value is significant, and our fitted model is accurate. The predictor variables used for the model are **searched, contrabn, weapnfnd, pf\_weapn, pf\_hcuff.**

**Prediction**

Chart, scatter chart

Description automatically generatedAfter applying the model to the test data set we plotted the below graph, which represents the actual arrests vs predicted arrests. Here, we can see that the regression line passes from the middle of the values, and we can see that the line is almost straight. This concludes that our model can predict values accurately.

**Linear Regression Analysis Using subset**

Using linear regression, we are going to predict the number of arrests in the NYPD. First, we start by fitting all the variables in the regression model and selecting the variables that are correlated to the target variable “arrests”. Below is the table we obtained after feature selection.

|  |  |
| --- | --- |
| **Table no 7 Linear Regression Model** | |
|  | *Dependent variable:* |
|  |  |
|  | arrested |
|  | |
| searched | 0.132\*\*\* |
|  | (0.004) |
|  |  |
| contrabn | 1.337\*\*\* |
|  | (0.015) |
|  |  |
| pf\_weapn | -0.809\*\*\* |
|  | (0.033) |
| Pf\_hcuff | 0.821\*\*\* |
|  | (0.012) |
|  |  |
| cs\_vcrim | 0.039\*\*\* |
|  | (0.002) |
|  |  |
| rf\_attir | 0.073\*\*\* |
|  | (0.004) |
| rf\_vcact | -0.054 |
|  | (0.002) |
| rf\_rfcmp | -0.082 |
|  | (0.003) |
| Constant | 1.039\*\*\* |
|  | (0.130) |
|  |  |

|  |  |
| --- | --- |
| **Table no 8 Linear Regression Model** | |
| Observations | 12,012 |
| R2 | 0.959 |
| Adjusted R2 | 0.959 |
| Residual Std. Error | 13.275 (df = 12003) |
| F Statistic | 35,077.120\*\*\* (df = 8; 12003) |
| P-value | 2.2e-16 |

**Interpretation:**

The above table is a linear regression model fitted to predict the arrests made by the NYPD using subset analysis. From this table, we can see that our degree of freedom is 12012, and the p-value is 2.2e-16 which is smaller than our confidence interval of 95% (alpha = 0.05). The obtained value of R-squared is 95.9%. and an f-statistic value of 35,077.120. The \*\*\* on the f-statistic value represents that the value is significant, and our fitted model is accurate. Using subset analysis in linear regression we could improve our prediction accuracy by 0.2%. The predictor variables used for the model are **searched, contrabn, pf\_weapn, Pf\_hcuff, cs\_vcrim, rf\_attir , rf\_vcact, rf\_rfcmp.**

**Prediction**

Chart, scatter chart

Description automatically generatedAfter applying the model to the test data set, we plotted the below graph, which represents the actual arrests vs predicted arrests using subset analysis. Here, we can see that the regression line passes from the middle of the values, and we can see that the line is straighter than the regression line observed in the above linear regression chart. This concludes that our model can predict values more accurately than in linear regression.

**Logistic Regression Model**

Using logistic regression, we are going to predict whether the person involved in the NYPD incident is A MALE OR A FEMALE. First, we start by fitting all the variables in the regression model and selecting the variables that are correlated to the target variable “male”. Below is the table we obtained after feature selection. The below table shows the output we get once run the model.

Table 9 shows that all the 22 factors used to predict whether a man or female was involved in the incident with the NYPC have very low p-values, indicating that they are all significant in the model.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table no 9 Logistic Regression Model** | | | | | | | | | |
|  | *Dependent variable:* | | | | | | | |
|  | male | | | | | | | |
| Predictor |  | Predictor |  | | Predictor |  |  | |
| raceB | -4.895\*\*\* | summoned | | 0.072\*\*\* | cs\_vcrim | | | -0.156\*\*\* |
|  | (0.282) |  | | (0.011) |  | | | (0.012) |
| raceI | 0.677\*\*\* | contrabn | | 0.166\*\*\* | cs\_bulge | | | 0.161\*\*\* |
|  | (0.088) |  | | (0.036) |  | | | (0.018) |
| raceP | -1.106\*\*\* | weapnfnd | | 0.266\*\*\* | rf\_vcrim | | | 0.158\*\*\* |
|  | (0.118) |  | | (0.053) |  | | | (0.020) |
| raceQ | -2.378\*\*\* | pf\_hcuff | | -0.099\*\*\* | rf\_othsw | | | 0.028\*\*\* |
|  | (0.177) |  | | (0.024) |  | | | (0.006) |
| raceW | -1.124\*\*\* | cs\_objcs | | 0.461\*\*\* | rf\_rfcmp | | | 0.114\*\*\* |
|  | (0.125) |  | | (0.032) |  | | | (0.022) |
| stopped | -0.081\*\*\* | cs\_descr | | 0.108\*\*\* | rf\_verbl | | | -0.327\*\*\* |
|  | (0.004) |  | | (0.009) |  | | | (0.083) |
| searched | 0.092\*\*\* | cs\_casng | | 0.105\*\*\* | rf\_knowl | | | 0.421\*\*\* |
|  | (0.023) |  | | (0.007) |  | | | (0.066) |
| arrested | -0.174\*\*\* | cs\_lkout | | -0.076\*\*\* | rf\_furtv | | | -0.107\*\*\* |
|  | (0.020) |  | | (0.011) |  | | | (0.016) |
| frisked | 0.149\*\*\* | cs\_cloth | | 0.109\*\*\* | Constant | | | -0.890\*\*\* |
|  | (0.015) |  | | (0.020) |  | | | (0.068) |

**Confusion Matrix of Training Data Set.**

The confusion matrix of the logistic model fit on the training data set is shown in the table below.

|  |  |  |
| --- | --- | --- |
| **Table no.10: Confusion Matrix of the training data set** | | |
|  | Reference | |
| Prediction | **No** | **Yes** |
| **No** | 4048 | 838 |
| **Yes** | 156 | 3366 |

**Interpretation:** The training data set is where the confusion matrix mentioned above was found. False Positives are 156, False Positives are 838, False Negatives are 848, True Negatives are 4048, True Positives are 3366.

False Positives – If the predicted gender is Male and if it is false

False Negatives – If the predicted gender is not Male and if it is false.

**Accuracy, Precision, Recall, and Specificity of the Training Data set.**

**Metrics of the training model:**

|  |  |
| --- | --- |
| **Table no. 11 Metrics of the Model** | |
| **Metrics** | **Values** |
| Accuracy | 88.18% |
| Precision | 95.55% |
| Recall | 80.06% |
| Specificity | 96.29% |

**Interpretation:**

The model's accuracy is 88.18%, which is good, according to the measures mentioned above. The recall is taken from (TP/(TP+FN)), which is 80.06%, and specificity is derived from (TP/(TP+FP)), which is 95.55%. Taking into account the results, we can say that the model can reliably predict whether the person involved in the incident was a male or a female.

**Confusion Matrix of Test Data Set.**

The below table displays the confusion matrix of the logistic model run on the test data set.

|  |  |  |
| --- | --- | --- |
| **Table no.12: Confusion Matrix of the test data set** | | |
|  | Reference | |
| Prediction | **No** | **Yes** |
| **No** | 1722 | 355 |
| **Yes** | 80 | 1447 |

**Interpretation:**

The training data set is where the confusion matrix mentioned above was found. There are 1722 True Negatives, 1447 True Positives, 80 False Positives, and 355 False Negatives.

**Accuracy, Precision, Recall, and Specificity of the Training Data set.**

**Metrics of the training model:**

|  |  |
| --- | --- |
| **Table no. 13 Metrics of the Test Data** | |
| **Metrics** | **Values** |
| Accuracy | 87.93% |
| Precision | 94.76% |
| Recall | 80.30% |
| Specificity | 95.56% |

**Interpretation:**

The model's accuracy is 87.93% on the test data set, which is lower than it was on the training data set, as can be seen from the metrics mentioned above. Precision is obtained from (TP/(TP+FP)), recall is derived from (TP/(TP+FN)), and specificity is derived from (TP/(TP+FP)), which is 95.56%. We can infer from the metrics that the model is correctly forecasting.

**Plotting and interpreting ROC curve and calculating AOC.**

A receiver operating characteristic curve represents how well a classification model performs at each classification threshold.

**Interpretation:**

The ROC curve is depicted in the plot up top. Plotting Specificity (False Positive Rate) vs. Sensitivity yields the result (True Positive Rate). The ROC curve ought to be nearer 1. As a result, the peak in the above graph is closer to 1 and accurately represents the model, as can be seen from the graph.

**Calculating the Area Under the Curve:**

Chart, line chart

Description automatically generatedThe ROC curve's area under it is 94.29%. Since AOC is higher, we can conclude that the model correctly predicts that males as males and females as females

**Conclusion:**

Black and Latinos were more likely to be frisked than whites and, among those frisked, were less likely to be found with a weapon. Black and Latinos were disproportionately stopped regardless of the demographic makeup of the neighborhood.​ Between 2014 and 2017, the NYPD used force on over 21,000 black and Latino people and over 2,200 white people. Even among those stopped, black and Latino people were more likely to have force used against them than white people.​ Out of total stops among races, Black people were more frisked than other races, whereas whites were more likely to be found with weapons and get arrested.​ A person is likely to be arrested If he/she is suspected to be involved in violent crime carrying contraband with questionable attire and if he/she refuses to comply with officers.

**References**

1. New York City Police Department. Stop-Question and Frisk Data. <https://www.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page>
2. Powell.C(Jan 2020) *Oneway ANOVA Explanation and Example in R; Part 1* <https://datascienceplus.com/oneway-anova-explanation-and-example-in-r-part-1/>
3. STHDA(n.d.) *Chi-Square Test of Independence in R* <http://www.sthda.com/english/wiki/chi-square-test-of-independence-in-r>
4. Danielle Kutner(Mar 16, 2020). Stop and Frisk Data Visualization and Analysis. <https://medium.com/@daniellekutner/stop-and-frisk-data-visualization-and-analysis-504c9a41ab6c>
5. Hlavac, Marek (2022). stargazer: *Well-Formatted Regression and Summary Statistics Tables.*

*R package version 5.2.3.* <https://CRAN.R-project.org/package=stargazer>