New York Police Department Stop-and-Frisk Data Analysis

Northeastern University

College of Professional Studies

Final Project Proposal/Dataset Selection

Dataset: New York Police Department Stop-and-Frisk (Group 4)

Class: ALY6015 - 21454

Professor Wada Roy

Group Members:

Amryta Panda

Bennett Furman

Durga Bhanu Nayak

Context:

The New York Police Department's Stop-and-Frisk policy was a controversial strategy to stop "suspicious" individuals and check for weapons or contraband. In 1999, African American and Latino individuals were detained in 84% of the total stops, but they only made up 50% of New York City's population [1]. Weapons and contraband were found much more often on white individuals than black people. According to article one, weapons were half as likely to be found on black people. Contraband, such as illicit drugs or drug paraphernalia, was one third as likely to be found on black individuals than white people. A federal judge named Shira Scheindlin deemed the policy unconstitutional and claimed that it conflicted with the fourth amendment, which prevents excessive searches and seizures. According to article two, the policy operated between 1994 and 2013 under two mayors. Mayor Rudy Giuliana was in office from 1994 to 2001. Then, Mayor Michael Bloomberg was in office from 2002 to 2013. The end of the policy was close to ten years ago. This was not a policy from the 1950s or 1960s. It was modern systemic racism under the guise of a New York Police Department policy. Even though the policy was found to be unconstitutional, our group wants to analyse the data for ourselves and make our own conclusions. We look forward to analysing the data and determining whether certain races or genders were targeted more than others. Please find the specifics of our proposal below. Additionally, taking incidents into consideration the department was disregarded and mentioned distrustful. Media has played a main part in overshadowing the face of the NYPD and charging them of assault, guilty of attacking with physical force and of wrong suspect, sometimes, they are also portrayed as "murderers". In recent years, not only NYPD but police departments of different other states have been in the limelight and it's making a negative impact on citizens and the media is making that worse. Moreover, it's ironic to say that having available for the security of the entire state/ city/ town the negative portrayal of the department has been demoralising.

Questions from the dataset:

1. Do certain races get stopped, frisked, or arrested more?
2. Is stopping a person due to furtive movements related to arrest?
3. Is stopping a person due to clothes commonly worn during a crime related to arrest?
4. Are there more arrests according to precinct?
5. How many arrests resulted after stopping or frisking?
6. Did the stop and frisk policy reduce crimes in New York City during these years?
7. What is the conversion ratio of stop-arrest?
8. How many people were frisked for behaviour/ clothing and arrested?

Descriptive Statistics Tables:

(i) **For entire data set (n = 12012)**

Below is the descriptive statistics for some key numerical variables from entire sample

Variables	mean	sd	median	max	skew	kurtosis	se
stopped	399.05	1255.47	41	22972	7.44	78.3	11.46
arrested	24.38	65.52	3	926	5.72	43.8	0.6
frisked	209.17	720.52	13	13593	7.3	71.56	6.57
searched	34.18	101.78	3	1828	6.31	54.23	0.93
summoned	24.59	92.79	2	2071	9.81	133.01	0.85
contrabn	7.45	21.41	1	364	6.45	58.59	0.2
weapnfnd	4.88	14.99	0	264	6.79	65.11	0.14
cs_descr	70.03	187.52	8	2702	5.56	42.18	1.71
cs_cloth	16.93	72.1	1	1942	11.48	193.5	0.66
cs_furtv	175.83	649.95	12	14520	8.98	116.3	5.93
cs_bulge	35.15	168.57	1	4653	11.25	181.61	1.54
rf_othsw	29.02	116.31	2	4125	15.54	393.73	1.06
rf_furtv	135.16	500.63	7	10662	8.07	90.39	4.57
rf_bulge	33.29	152.61	1	3597	10.49	153.18	1.39

(ii) **Groupwise Descriptive statistics**

For males (n = 6006)

Below is the descriptive statistics of key numerical columns for male group

Variables	mean	sd	median	max	skew	kurtosis	se
stopped	741.17	1702.81	157	22972	5.36	40.49	21.97
arrested	44.12	87.76	10	926	4.09	22.1	1.13
frisked	402.97	980.74	72	13593	5.2	36.1	12.66
searched	64.49	137.15	14	1828	4.49	27.46	1.77
summoned	45.05	127.21	8	2071	7.12	69.23	1.64
contrabn	13.66	28.81	3	364	4.65	30.27	0.37
weapnfnd	9.37	20.19	2	264	4.89	33.73	0.26
cs_descr	130.33	250.38	32	2702	3.95	21.27	3.23

cs_cloth	32.42	99.45	5	1942	8.28	100.25	1.28
cs_furtv	331.51	890.64	54	14520	6.45	59.97	11.49
cs_bulge	68.3	233.61	8	4653	8.03	92.41	3.01
rf_othsw	53.12	159.73	10	4125	11.53	212.95	2.06
rf_furtv	261.38	684.74	40	10662	5.76	45.92	8.84
rf_bulge	64.82	211.07	9	3597	7.48	77.75	2.72

For females (n = 6006)

Below is the descriptive statistics of key numerical columns for female group

Variables	mean	sd	median	max	skew	kurtosis	se
stopped	56.92	137.72	12	2261	6.62	63.15	1.78
arrested	4.64	10.36	1	108	4.46	25.8	0.13
frisked	15.37	37.19	3	507	5.48	40.94	0.48
searched	3.86	8.61	1	103	4.5	26.84	0.11
summoned	4.14	14.14	0	324	10.19	155.17	0.18
contrabn	1.24	3.09	0	51	5.29	42.6	0.04
weapnfnd	0.39	1.19	0	19	5.66	47.13	0.02
cs_descr	9.72	19.09	2	210	4.07	22.44	0.25
cs_cloth	1.45	5.22	0	112	10.52	157.48	0.07
cs_furtv	20.15	56.71	3	936	7.5	78.95	0.73
cs_bulge	2	7.8	0	184	9.9	143.94	0.1
rf_othsw	4.93	19.65	1	608	15.74	370.25	0.25
rf_furtv	8.95	23.87	1	365	6.15	52.15	0.31
rf_bulge	1.75	6.51	0	123	8.77	105.86	0.08

Observations:

- In New York, average stops made by police for males is around 13 times higher than that of females.
- Average male arrests are also significantly higher than the females.
- Most of the suspicious activities have been carried out by males rather than females.

Outcome variables:

We plan to use whether someone was frisked, stopped, or arrested as potential outcome variables. We may just stick with one outcome variable. In this case, we will go with arrests. All of these outcome variables have binary outcomes. Either someone is arrested or not, stopped or not, and frisked or not. The data set only contains the number of arrests, so we will have to extrapolate information on whether someone was arrested or not. This will allow us to use a logistic regression, among other models, to predict the binary outcome of getting arrested or not.

Predictor variables:

Potential predictor variables include the year of the stop, the precinct in which the stop occurred, the sex of the person stopped, the perceived race of the person stopped, total number of summons issued, total number of contraband items found on the suspect, and total number of weapons found on the suspect. A host of more specific predictors are included in the data set, and are separated into activities involved in stops and frisks. We will have to decide which of these predictors we want to use in our models. Potential choices include stops due to fitting a relevant description, stops due to wearing clothes commonly used in a crime, stops due to exhibiting furtive movements, stops due to a suspicious bulge, frisks due to other suspicion of weapons, frisks due to furtive movements, and frisks due to a suspicious bulge.

Subgroups: Potential subgroups for subgroup analysis include gender, race, year, and precinct.

Analytical Methods (Models): From the dataset, we will be examining the dataset thoroughly and perform the required exploratory data analysis. Upon examining the required variables, we plan to use logistic regression to predict the outcome of the binary response variable. In our case, the binary response variable indicates whether someone was arrested or not. We will create dummy variables and split the data into testing and training sets. We also plan to use a multiple linear regression model. We will use the number of arrests as our dependent variable and choose from the predictor variables listed above in the predictor variables section. We will refine our model by adding or removing predictors to obtain the maximum r squared adjusted value. We will use other linear regression diagnostics, including the Residuals vs. Fitted, Normal Q-Q, Scale-Location, and Residuals vs. Leverage plots to evaluate the fit of our model. Other models we are considering using include Generalised Linear Models, Lasso Regression, and Ridge Regression.

Analytical Methods (Statistical Tests): We will be analysing if any significant difference in the number of arrests is present among multiple race groups using One-Way ANOVA. We will also use the Chi Square Goodness of Fit Test to determine the number of arrests per race is consistent with a hypothesised distribution, in which the number of arrests is equal for each race.

References:

1. NYPD’s Infamous Stop-and-Frisk Policy Found Unconstitutional. (2019, March 15). The Leadership Conference Education Fund. Retrieved March 2, 2022, from https://civilrights.org/edfund/resource/nypds-infamous-stop-and-frisk-policy-found-unconstitutional/
2. The Bridge Initiative. (2020, June 5). Factsheet: NYPD STOP AND FRISK POLICY. Georgetown University. Retrieved March 2, 2022, from https://bridge.georgetown.edu/research/factsheet-nypd-stop-and-frisk-policy/