Report 1

DATA UNDERSTANDING AND VISUALIZATION OF NEW YORK CITY'S STOP-QUESTION-FRISK (SQF) DATA

By

PRAJWAL NAGARAJ

ABSTRACT

The Data Understanding report provides a thorough analysis of the Stop, Question, and Frisk (SQF) program data for the year 2012. The goal is to uncover insights about the program's effectiveness and identify any prevalent trends or patterns in the dataset.

The key findings of the analysis include observations on data quality, such as missing values and outliers, which may affect the reliability and validity of the dataset. Additionally, the report describes important attributes within the dataset, shedding light on factors like demographic information, stop locations, and outcomes of encounters.

The analysis also presents initial insights into attribute relationships, offering preliminary understanding of how various factors may influence the likelihood of stops or the outcomes of these encounters.

Based on the findings, recommendations are provided to guide future research and policy decisions related to the SQF program and its impact on policing and community relations in New York City. These recommendations aim to inform further analysis and potential actions going forward.

- **Business Understanding:**

1. **What is the purpose of the SQF program?**

The Stop, Question, and Frisk (SQF) program is a strategy employed by law enforcement agencies, notably the New York City Police Department (NYPD), to enhance public safety and prevent crime. This program involves officers stopping, questioning, and sometimes frisking individuals suspected of engaging in criminal activity or behaving suspiciously. The overarching goals of the SQF program include preventing and reducing crime, gathering intelligence on criminal activities, and promoting community safety.

2. **How would you define and measure the effectiveness of such a program?**

To measure the effectiveness of the SQF program, we need to consider various factors:

- **Crime Rates:** We'll analyze changes in crime rates, particularly in areas where SQF activities are concentrated, to assess the program's impact on crime prevention and reduction.
- **Arrest and Conviction Rates:** Tracking the number of arrests and subsequent convictions resulting from SQF encounters will help us understand the program's effectiveness in apprehending criminals.
- **Community Feedback:** Soliciting feedback from community members, including residents and advocacy groups, will provide valuable insights into the perceived effectiveness of the program and its impact on community relations.
- **Legal Compliance:** We'll evaluate the program's adherence to legal standards, such as constitutional rights and anti-discrimination laws, to ensure accountability and transparency.

3. **What data would you need to be able to judge its effectiveness?**

To judge the effectiveness of the SQF program, we require the following data:

- **Stop, Question, and Frisk Data:** Detailed records of stops, questions, and frisks conducted by law enforcement officers, including information on the individuals stopped, reasons for the stops, outcomes, and officer identifiers.
- **Crime Data:** Comprehensive statistics on reported crimes, arrests, and convictions, broken down by location and time, to identify trends and correlations with SQF activities.

- **Community Surveys:** Surveys or feedback mechanisms to collect community perceptions and experiences regarding the SQF program, including trust in law enforcement, feelings of safety, and satisfaction with police practices.
- **Legal Compliance Records:** Documentation of legal challenges, complaints, lawsuits, and settlements related to the SQF program, along with any policy changes or court rulings affecting its implementation.

➢ **Data Understanding:**

1. **Description of Data:**

The dataset contains various attributes including demographic information (e.g., age, race, gender), details of police encounters (e.g., reasons for stop, force used), and outcomes of stops (e.g., arrests, frisks).

- Below is the table with meaning and type of data (e.g., scale, values) for each attribute in the data file.

Column	Data Type	Meaning
year	int64	YEAR OF STOP (CCYY)
pct	int64	PRECINCT OF STOP (FROM 1 TO 123)
ser_num	int64	UF250 SERIAL NUMBER
datestop	int64	DATE OF STOP (MM-DD-YYYY)
timestop	int64	TIME OF STOP (HH:MM)
recstat	object	RECORD STATUS
inout	object	WAS STOP INSIDE OR OUTSIDE ?
trhsloc	object	WAS LOCATION HOUSING OR TRANSIT AUTHORITY ?
perobs	float64	PERIOD OF OBSERVATION (MMM)
crimsusp	object	CRIME SUSPECTED
perstop	object	PERIOD OF STOP (MMM)
typeofid	object	STOPPED PERSON'S IDENTIFICATION TYPE
explnstp	object	DID OFFICER EXPLAIN REASON FOR STOP ?
othpers	object	WERE OTHER PERSONS STOPPED, QUESTIONED OR FRISKED ?
arstmade	object	WAS AN ARREST MADE ?
arstoffn	object	OFFENSE SUSPECT ARRESTED FOR
sumissue	object	WAS A SUMMONS ISSUED ?
sumoffen	object	OFFENSE SUSPECT WAS SUMMONSED FOR
compyear	int64	COMPLAINT YEAR (IF COMPLAINT REPORT PREPARED)
comppct	int64	COMPLAINT PRECINCT (IF COMPLAINT REPORT PREPARED)
offunif	object	WAS OFFICER IN UNIFORM ?
officrid	object	ID CARD PROVIDED BY OFFICER (IF NOT IN UNIFORM)
frisked	object	WAS SUSPECT FRISKED ?
searched	object	WAS SUSPECT SEARCHED ?
contrabn	object	WAS CONTRABAND FOUND ON SUSPECT ?
adtlrept	object	WERE ADDITIONAL REPORTS PREPARED ?
pistol	object	WAS A PISTOL FOUND ON SUSPECT ?
riflshot	object	WAS A RIFLE FOUND ON SUSPECT ?
asltweap	object	WAS AN ASSAULT WEAPON FOUND ON SUSPECT ?
knifcuti	object	WAS A KNIFE OR CUTTING INSTRUMENT FOUND ON SUSPECT ?

machgun	object	WAS A MACHINE GUN FOUND ON SUSPECT ?
othrweap	object	WAS ANOTHER TYPE OF WEAPON FOUND ON SUSPECT
pf_hands	object	PHYSICAL FORCE USED BY OFFICER - HANDS
pf_wall	object	PHYSICAL FORCE USED BY OFFICER - SUSPECT AGAINST WALL
pf_grnd	object	PHYSICAL FORCE USED BY OFFICER - SUSPECT ON GROUND
pf_drwep	object	PHYSICAL FORCE USED BY OFFICER - WEAPON DRAWN
pf_ptwep	object	PHYSICAL FORCE USED BY OFFICER - WEAPON POINTED
pf_baton	object	PHYSICAL FORCE USED BY OFFICER - BATON
pf_hcuff	object	PHYSICAL FORCE USED BY OFFICER - HANDCUFFS
pf_pepsp	object	PHYSICAL FORCE USED BY OFFICER - PEPPER SPRAY
pf_other	object	PHYSICAL FORCE USED BY OFFICER - OTHER
radio	object	RADIO RUN
ac_rept	object	ADDITIONAL CIRCUMSTANCES - REPORT BY VICTIM/WITNESS/OFFICER
ac_inves	object	ADDITIONAL CIRCUMSTANCES - ONGOING INVESTIGATION
rf_vcrim	object	REASON FOR FRISK - VIOLENT CRIME SUSPECTED
rf_othsw	object	REASON FOR FRISK - OTHER SUSPICION OF WEAPONS
ac_proxm	object	ADDITIONAL CIRCUMSTANCES - PROXIMITY TO SCENE OF OFFENSE
rf_attir	object	REASON FOR FRISK - INAPPROPRIATE ATTIRE FOR SEASON
cs_objcs	object	REASON FOR STOP - CARRYING SUSPICIOUS OBJECT
cs_descr	object	REASON FOR STOP - FITS A RELEVANT DESCRIPTION
cs_casng	object	REASON FOR STOP - CASING A VICTIM OR LOCATION
cs_lkout	object	REASON FOR STOP - SUSPECT ACTING AS A LOOKOUT
rf_vcact	object	REASON FOR FRISK- ACTIONS OF ENGAGING IN A VIOLENT CRIME
cs_cloth	object	REASON FOR STOP - WEARING CLOTHES COMMONLY USED IN A CRIME
cs_drgtr	object	REASON FOR STOP - ACTIONS INDICATIVE OF A DRUG TRANSACTION
ac_evasv	object	ADDITIONAL CIRCUMSTANCES - EVASIVE RESPONSE TO QUESTIONING
ac_assoc	object	ADDITIONAL CIRCUMSTANCES - ASSOCIATING WITH KNOWN CRIMINALS
cs_furtv	object	REASON FOR STOP - FURTIVE MOVEMENTS
rf_rfcmp	object	REASON FOR FRISK - REFUSE TO COMPLY W OFFICER'S DIRECTIONS
ac_cgdir	object	ADDITIONAL CIRCUMSTANCES - CHANGE DIRECTION AT SIGHT OF OFFICER
rf_verbl	object	REASON FOR FRISK - VERBAL THREATS BY SUSPECT
cs_vcrim	object	REASON FOR STOP - ACTIONS OF ENGAGING IN A VIOLENT CRIME
cs_bulge	object	REASON FOR STOP - SUSPICIOUS BULGE
cs_other	object	REASON FOR STOP - OTHER
ac_incid	object	ADDITIONAL CIRCUMSTANCES - AREA HAS HIGH CRIME INCIDENCE
ac_time	object	ADDITIONAL CIRCUMSTANCES - TIME OF DAY FITS CRIME INCIDENCE
rf_knowl	object	REASON FOR FRISK - KNOWLEDGE OF SUSPECT'S PRIOR CRIM BEHAV
ac_stsnd	object	ADDITIONAL CIRCUMSTANCES - SIGHTS OR SOUNDS OF CRIMINAL ACTIVITY

ac_other	object	ADDITIONAL CIRCUMSTANCES - OTHER
sb_hdobj	object	BASIS OF SEARCH - HARD OBJECT
sb_outln	object	BASIS OF SEARCH - OUTLINE OF WEAPON
sb_admis	object	BASIS OF SEARCH - ADMISSION BY SUSPECT
sb_other	object	BASIS OF SEARCH - OTHER
repcmd	object	REPORTING OFFICER'S COMMAND (1 TO 999)
revcmd	int64	REVIEWING OFFICER'S COMMAND (1 TO 999)
rf_furt	object	REASON FOR FRISK - FURTIVE MOVEMENTS
rf_bulg	object	REASON FOR FRISK - SUSPICIOUS BULGE
offverb	object	VERBAL STATEMENT PROVIDED BY OFFICER (IF NOT IN UNIFORM)
offshld	object	SHIELD PROVIDED BY OFFICER (IF NOT IN UNIFORM)
forceuse	object	REASON FORCE USED
sex	object	SUSPECT'S SEX
race	object	SUSPECT'S RACE
dob	int64	SUSPECT'S DATE OF BIRTH (CCYY-MM-DD)
age	int64	SUSPECT'S AGE
ht_feet	int64	SUSPECT'S HEIGHT (FEET)
ht_inch	int64	SUSPECT'S HEIGHT (INCHES)
weight	int64	SUSPECT'S WEIGHT
haircolr	object	SUSPECT'S HAIRCOLOR
eyecolor	object	SUSPECT'S EYE COLOR
build	object	SUSPECT'S BUILD
othfeatr	object	SUSPECT'S OTHER FEATURES (SCARS, TATOOS ETC.)
addrtyp	object	LOCATION OF STOP ADDRESS TYPE
rescode	object	LOCATION OF STOP RESIDENT CODE
premtype	object	LOCATION OF STOP PREMISE TYPE
premname	object	LOCATION OF STOP PREMISE NAME
addrnum	object	LOCATION OF STOP ADDRESS NUMBER
stname	object	LOCATION OF STOP STREET NAME
stinter	object	LOCATION OF STOP INTERSECTION
crossst	object	LOCATION OF STOP CROSS STREET
aptnum	object	LOCATION OF STOP APT NUMBER
city	object	LOCATION OF STOP CITY
state	object	LOCATION OF STOP STATE
zip	object	LOCATION OF STOP ZIP CODE
addrpct	object	LOCATION OF STOP ADDRESS PRECINCT
sector	object	LOCATION OF STOP SECTOR
beat	object	LOCATION OF STOP BEAT
post	object	LOCATION OF STOP POST
xcoord	object	LOCATION OF STOP X COORD
ycoord	object	LOCATION OF STOP Y COORD
dettypcm	object	DETAILS TYPES CODE
linecm	object	COUNT >1 ADDITIONAL DETAILS
detailcm	object	CRIME CODE DESCRIPTION

2. **Data Quality Verification:**

- We utilized the 'duplicated' function to verify the presence of duplicate entries, finding none.
- For missing values, we employed the 'isnull().sum()' method, revealing a few instances. These were subsequently eliminated using 'dropna()'.
- Outliers within the dataset were detected via z-scores, pinpointing anomalies in 'perobs', 'age', 'weight', ser_num and 'height', which were then rectified.
- Extreme values in 'age' were replaced with the mode, with special consideration for ages below 5 and above 100, amounting to 1777 data points.
- In the case of 'height', we unified measurements into inches, combining feet and inches into a single column.
- 'Ser_num', serving as a serial number, was disregarded due to its negligible significance.
- 'Perobs', originally in float format representing the period of observation in minutes, was converted to integers, and values exceeding 60 minutes were replaced with the mode.
- Similarly, weight values below 20 and above 500 were addressed using the mode, encompassing 1015 data points.
- We even did the box plots for outliers to analyze them.

3. **Key Statistics:**

We provided simple statistics such as count, unique, top and frequency for categorical columns, and for numerical columns we did simple statistics like count, mean, median, and range for important attributes.

Categorical Columns:

1. **recstat, inout, trhsloc, crimsusp, perstop, typeofid, explnstp, othpers, arstmade, arstoffn, sumissue, sumoffen, offunif, officrid, frisked, searched, contrabn, adtlrept, pistol, riflshot, asltweap, knifcuti, machgun, othrweap, pf_hands, pf_wall, pf_grnd, pf_drwep, pf_ptwep, pf_baton, pf_hcuff, pf_pepsp**: These columns represent various aspects of police stops and encounters, such as stop outcome, location, reasons for stop, actions taken by officers, and presence of weapons or contraband.

2. **Unique Values and Frequencies**: The 'unique' count signifies the number of distinct categories within each column, while 'top' indicates the most frequent category, and 'freq' denotes its frequency.

Numerical Columns:

1. **year, pct, ser_num, datestop, timestop, perobs, compyear, comppct, revcmd, dob, age, ht_feet, ht_inch, weight**: These columns encompass numerical data related to the year of the incident, precinct, serial number, date and time of the stop, duration of observation, and various physical attributes of the individuals stopped, such as age, height, and weight.

2. **Mean, Standard Deviation, Min, Max, Quartiles**: These statistics provide insights into the distribution and range of numerical values within each column.

Observations:

Missing Values: There don't seem to be any counts provided for missing values within the descriptions. It would be essential to check for missing data, especially in categorical columns, as they might indicate incomplete records or data entry issues.

Outliers: While the descriptions provide statistical summaries for numerical columns, they don't explicitly mention outliers. Investigating outliers, particularly in age, height, and weight, could be valuable for data quality assessment and anomaly detection.

Data Integrity: Some columns, such as 'arstoffn' and 'sumissue', appear to have significant missing data, as indicated by the large counts of missing values. Understanding the reasons behind these missing values and their potential impact on analyses is crucial.

Age and Physical Attributes: The mean age of individuals stopped is approximately 28 years, with a wide standard deviation of 24 years, indicating variability in age across the dataset. The average height is around 5 feet 2 inches, with an average weight of approximately 169 pounds.

4. Visualizations:

Most important attributes like age, height, weight, and length of observation before engaging. Visualizations including histograms, box plots, and pie charts were utilized to explore the distributions and relationships between attributes.

Visualization for Distribution of Age:

Chart Attributes:

1. **X-axis (Age)**: Represents different age groups, segmented into bins (e.g., 0-10, 10-20 years, etc.). This axis helps in identifying the age distribution across the population.
2. **Y-axis (Count)**: Indicates the number of individuals within each age group. This helps in quantifying the population within each age segment.
3. **Bars**: Each bar represents an age group, where the height of the bar corresponds to the count of individuals in that age group. Blue is used for the bars, which provides a clear visual distinction for each segment.

Interpretation:

- **Clarity in Distribution**: Histograms are ideal for showing the distribution of data over a continuous interval or certain time. In this case, it helps in understanding how the population is spread across different ages.
- **Ease of Analysis**: This visualization makes it easy to identify which age groups are the most or least populous. For instance, the peak around the age of 20 suggests a high concentration of young adults.

- **Comparative Analysis**: The histogram allows for quick comparison between different age groups, making it evident how the population decreases with increasing age.

Visualization for Distribution of Weight:

Chart Attributes:

1. **X-axis (Weight)**: The x-axis represents weight, divided into bins of 100 units each, ranging from 0 to 500. This binning helps in grouping the weight data into manageable intervals, making it easier to observe the distribution pattern.
2. **Y-axis (Count)**: The y-axis quantifies the number of occurrences (frequency) of each weight bin. The values range from 0 to 200,000, indicating the total count of data points falling within each weight range.
3. **Bars**: Each bar in the histogram represents a weight interval. The height of the bar corresponds to the count of data points within that specific weight range. For example, the tallest bar is between 100 and 200, suggesting that most data points fall within this weight range.

Interpretation:

- **Simplicity and Clarity**: Histograms are straightforward and provide a clear visual summary of data distribution. They allow viewers to easily identify the most common ranges of data (mode) and observe the shape of the distribution (e.g., normal, skewed).
- **Effective for Large Data Sets**: When dealing with large datasets, histograms help in summarizing the data into a form that is easy to interpret, avoiding information overload that might come from listing all data points.
- **Identifying Outliers and Gaps**: This visualization helps in spotting any outliers or unusual gaps in the data. For instance, very high or very low bars at the extremes of the histogram might indicate outliers.
- **Comparative Analysis**: If comparing weight distributions across different groups or conditions, histograms can be aligned side-by-side to provide a visual comparison.

Visualize Length of observation before Engaging:

Chart Attributes:

- **X-axis (Perobs): This axis represents the number of observations before engaging, ranging from 0 to over 60. The values are spaced at intervals (0, 10, 20, 30, 40, 50, 60), suggesting that the data is grouped into these bins.**
- **Y-axis (Count): This axis quantifies the number of occurrences for each bin of Perobs. The scale goes up to 400,000, indicating a high volume of data points collected.**

- **Bars: Each bar represents the count of observations that fall into each Perob bin. The height of the bar corresponds to the count on the y-axis.**

Interpretation:

- **The tallest bar is at 0 Perobs, with a count close to 380,000, indicating that a vast majority of observations involve immediate engagement without any prior observation.**
- **The count drastically decreases as the number of Perobs increases, with the second bar (around 10 Perobs) significantly lower, and further reductions as the number of Perobs continues to increase.**

Distribution of Stopped Individuals by Race:

Chart Type: Pie Chart

Attributes:

- **Black: This category occupies the largest portion of the chart, indicating that Black individuals constitute the majority of the stops.**
- **White-Hispanic: The second largest segment, showing a significant number of stops among White-Hispanic individuals.**
- **White: Represents a smaller proportion compared to Black and White-Hispanic.**
- **Black-Hispanic: This segment is slightly smaller than White but larger than the remaining categories.**
- **Other: The smallest segment, indicating fewer stops among individuals of other races.**
- **Asian/Pacific Islander and American Indian/Alaskan: These categories are very small, suggesting minimal stops among these races.**
- **Reason for Visualization Choice: A pie chart is effective here as it provides a clear visual representation of the proportionate distribution of stopped individuals by race, allowing for immediate visual comparison of the categories.**

Distribution of Stops by City:

Chart Type: Pie Chart

Attributes:

- **Bronx: The largest segment, indicating the highest number of stops occur in this area.**
- **Brooklyn: The second largest, showing a substantial number of stops but less than the Bronx.**
- **Manhattan: Represents a significant portion but is smaller than Bronx and Brooklyn.**
- **Queens: Smaller than Manhattan, indicating fewer stops.**

Staten Island: The smallest segment, suggesting the fewest stops among the listed cities.

Reason for Visualization Choice: Similar to the first chart, the pie chart here effectively illustrates the distribution of stops across different cities, highlighting areas with higher and lower frequencies of stops.

Frisk by Race, Sex, City:

1. Frisks by Race

- **Chart Type: Stacked bar chart**
- **Attributes:**
    - **X-axis: Represents different racial categories (Black, White-Hispanic, White, Black-Hispanic, Asian/Pacific Islander, Other, American Indian/Alaskan Native).**
    - **Y-axis: Represents the count of individuals.**
    - **Colors: Blue bars represent individuals who were frisked (Y), and orange bars represent individuals who were not frisked (N).**
- **Interpretation: This chart is useful for visualizing the distribution of frisks among different racial groups and comparing the proportion of frisked vs. not frisked within each group. The choice of a stacked bar chart allows for an easy comparison of the two categories (frisked and not frisked) across different races.**

2. Frisks by City

- **Chart Type: Stacked bar chart**
- **Attributes:**
    - **X-axis: Represents different cities (Bronx, Manhattan, Brooklyn, Queens, Staten Island).**
    - **Y-axis: Represents the count of individuals.**
    - **Colors: Blue bars represent individuals who were frisked (Y), and orange bars represent individuals who were not frisked (N).**
- **Interpretation: This chart provides a geographical breakdown of frisks, showing which cities have higher occurrences. The use of a stacked bar chart here helps in understanding the proportion of frisked vs. not frisked individuals in each city, highlighting potential areas with higher police activity or different policing policies.**

3. Frisks by Sex

- **Chart Type: Stacked bar chart**
- **Attributes:**

- **X-axis: Represents different sex categories (Male, Female, Other).**
- **Y-axis: Represents the count of individuals.**
- **Colors: Blue bars represent individuals who were frisked (Y), and orange bars represent individuals who were not frisked (N).**

- **Interpretation:** This chart focuses on the distribution of frisks based on sex, providing insights into how different genders are affected by frisking practices. The stacked bar chart format is particularly effective for showing the disparity in frisking practices among different sex categories.

Why These Visualizations Were Chosen:

The use of stacked bar charts in all three visualizations provides a clear method to compare two categories (frisked vs. not frisked) across different groups (race, city, sex). This format is straightforward and allows for immediate visual comparisons, making it easier to spot trends and disparities in the data. Each chart addresses a specific demographic factor, offering a comprehensive view of the data from multiple angles, which is crucial for understanding the broader context of frisking practices.

Further Visualization like Distribution of trshloc data that is inside or outside or neighthe, also arrested summons issued or nan plot was made.

5. **Exploring Relationships:**

Scatter plots, correlation matrices, and cross-tabulations were used to investigate relationships between different attributes, such as the correlation between age, height, and weight, and the distribution of frisks by race, sex, and location.

Scatter Plots

1. **Age vs. Height:**
    - The scatter plot shows a wide distribution of heights across different ages, with no clear trend indicating that height does not vary significantly with age once adulthood is reached.
    - The histogram on the y-axis shows the distribution of heights, which is relatively uniform across the range, with a slight peak around 70 inches.
2. **Age vs. Weight:**
    - This plot shows a more dispersed distribution, suggesting that weight varies more significantly with age compared to height.
    - The histogram on the y-axis for weight shows a distribution that is slightly right-skewed, indicating that higher weights are less frequent but present.
3. **Height vs. Weight:**

- There is a visible trend where weight increases with height, which is expected as taller individuals generally weigh more.
- The plot shows a broad range of weights for similar heights, indicating variability in body composition among individuals of the same height.

Correlation Matrix

- The matrix quantifies the relationships observed in the scatter plots:
  - **Age and Height:** Correlation coefficient of 0.07, indicating a very weak positive relationship.
  - **Age and Weight:** Correlation coefficient of 0.24, suggesting a weak positive correlation where weight slightly increases with age.
  - **Height and Weight:** Correlation coefficient of 0.45, showing a moderate positive correlation, which is the strongest relationship observed among the three.

Analysis

- The correlation matrix and scatter plots together provide a clear visual and statistical representation of how these attributes interrelate.
- The strongest correlation is between height and weight, which is intuitive as taller individuals typically have more body mass.
- The weak correlation between age and weight might be influenced by factors such as changes in metabolism, lifestyle, and health conditions over a person's lifespan.
- The very weak correlation between age and height after reaching adulthood is expected as height does not change significantly in adult years.

This analysis helps in understanding how these physical attributes correlate and vary across a population, useful in fields such as health sciences, sports science, and demography.

Group-wise averages of numerical attributes by crimsusp, forceuse, and type of force used was made as below.

Criminal Suspicion (crimsusp): This column indicates different types of criminal suspicions, such as CPW (Criminal Possession of a Weapon), FEL (Felony), Assault, etc.

- **Force Use (forceuse)**: Describes whether force was used or not, along with the reason for its use.
- **Physical Force Used (pf_hands, pf_wall, pf_grnd, pf_drwep, pf_ptwep, pf_baton, pf_hcuff, pf_pepsp, pf_other)**: These columns indicate whether specific types of physical force were used during the encounter. For instance, 'pf_hands' could represent whether physical force with hands was used or not.

- **Height (height)**: Represents the height of individuals associated with each criminal suspicion.

Observations:

- **Force Use Patterns**: It seems that force was used in various situations, such as self-defense or suspected flight. This data could provide insights into the circumstances under which law enforcement officers employ force during interactions.
- **Height Variability**: The heights associated with different criminal suspicions vary. This could be due to factors such as demographics or the nature of the offenses.
- **Force Tactics**: Different types of physical force were used in some cases, like the use of batons or handcuffs. Understanding the distribution of these tactics across different criminal suspicions could shed light on law enforcement practices.

5. **Comparison of Reasons for SQF and Type of Force Used:**

A cross-tabulation was performed to compare the reasons for SQF with the type of force used by officers. The group-wise averages of numerical attributes by reasons for SQF and type of force used were also calculated to identify patterns.

Attributes:

1. **Reason for SQF (Y-axis)**: This attribute categorizes the reasons for initiating an SQF. It is divided into two categories:
    - **N (No)**: Indicates that the specific reason was not a factor in the SQF.
    - **Y (Yes)**: Indicates that the specific reason was a factor in the SQF.
2. **Type of Force Used (X-axis)**: This attribute categorizes the type of force used during the SQF. The types include:
    - **Defense of Other**
    - **Defense of Self**
    - **Other**
    - **Overcome Resistance**
    - **Suspected Flight**
    - **Suspected Weapon**
3. **Count (Color Intensity)**: The intensity of the color in each cell represents the count of incidents. Darker shades indicate a higher number of incidents, as shown in the color bar on the right side of the heatmap.

Analysis:

- **High Incidence Areas**: The darkest cells (indicating the highest counts) are observed in the "N" row for "Defense of Self" and "Overcome Resistance". This suggests that these types of force are more commonly used regardless of the specific reason for the SQF being affirmed (Y) or not (N).
- **Low Incidence Areas**: The lighter shades in the "Y" row across most types of force suggest that when a specific reason is affirmed as a factor in the SQF, the frequency of using force is generally lower. This could indicate a more cautious approach when a specific reason is identified.
- **Specific Observations**:
    - **Defense of Self**: This type of force is predominantly used when no specific reason for the SQF is affirmed, suggesting a possible default defensive action by officers.
    - **Suspected Weapon**: When a suspected weapon is the reason for the SQF (Y), there is a noticeable increase in the use of force, which is intuitive given the potential threat.

Why This Visualization:

A heatmap is chosen for its ability to visually represent the density of data through color saturation, making it easier to spot trends, correlations, and outliers across two dimensions. This format is particularly useful for quickly understanding the relationship between the reason for an SQF and the type of force used, facilitating immediate visual insights into areas that may require further investigation or policy evaluation.

Conclusion:

The heatmap effectively highlights the relationship between the reasons for SQF and the corresponding police response in terms of force used. It reveals patterns that suggest certain types of force are more commonly used irrespective of the affirmed reasons for the SQF, while others are more contingent on specific conditions being met. This analysis can be crucial for law enforcement training and policymaking, aiming to optimize the appropriateness of force used during SQF encounters.