Advanced Concepts in Data Analytics

Final Project: New York City Case Study using CRISP-DM

Group 1 Members: Adebukola Aina, Adeleke Akanbi, Sungeun Kang, Clovis Ndang, Benjamin Ukueku

Executive Summary

This project analyzes the New York City Police Department's 2012 *Stop, Question, and Frisk (SQF)* dataset using the CRISP-DM framework to uncover insights about fairness, effectiveness, and operational patterns in policing. Through descriptive analytics, association rule mining, clustering, and predictive modeling, the study identifies strong demographic and temporal trends—particularly the disproportionate targeting of young Black and Hispanic males and higher stop frequencies at night.
The Random Forest model achieved the best predictive performance (ROC-AUC = 0.725), highlighting key factors linked to force usage. These findings demonstrate how data-driven policing analytics can improve accountability, optimize resource allocation, and support fair, evidence-based policy decisions.

Report 1. Data & Visualization

Introduction

The Stop, Question, and Frisk (SQF) dataset provides detailed records of street encounters between police officers and civilians conducted by the New York City Police Department (NYPD) in 2012. Each record captures when, where, and why an individual was stopped, as well as demographic information, the officer's observations, whether any physical force was used, and the outcome of the encounter.

This dataset offers a valuable opportunity to explore the relationship between observable factors—such as the time of day, location, or reasons for suspicion—and the outcomes of police interactions, including frisking, arrest, or use of force. By applying data analysis and machine learning techniques, we can identify patterns, relationships, and potential disparities within these interactions.

The purpose of this project is to apply the CRISP-DM framework to analyze and model the SQF data through four key analytical tasks:

1. **Descriptive and Visualization Analysis –** understanding the data and visualizing key patterns.
2. **Association Rule Mining –** identifying relationships among observed circumstances, reasons for stops, and use of force.
3. **Cluster Analysis –** grouping similar encounters to uncover behavioral or procedural patterns.
4. **Predictive Modeling –** building models to predict outcomes such as whether physical force was used.

This comprehensive approach aims to transform raw data into meaningful insights that can support policy evaluation, fairness assessment, and informed decision-making within policing practices.

Business Understanding

The primary goal of this project is to assess and understand the effectiveness and fairness of the NYPD's Stop, Question, and Frisk practices in 2012. Specifically, the analysis seeks to answer the following questions:

1. **Effectiveness:**
   - Under what circumstances are stops most likely to lead to frisking or arrest?
   - How often do stops result in the recovery of weapons or contraband?
2. **Fairness and Demographic Patterns:**
   - Do factors such as race, age, or gender influence the likelihood of being stopped or subjected to force?
   - Are certain demographics disproportionately represented in specific types of stops or outcomes?
3. **Operational Insights:**
   - What are the most common reasons for stops, and how are they related to the type of physical force used?
   - Can we group or cluster similar incidents to identify patterns in officer behavior, time of day, or geographic distribution?
4. **Predictive Analysis:**
   - Can a predictive model estimate the likelihood of physical force based on known circumstances (e.g., time of day, reason for stop, or observed behaviors)?

By addressing these questions, the project will provide actionable insights into the patterns of stop-and-frisk encounters, evaluate the presence of any systematic bias, and demonstrate how data analytics can support accountability and informed policy review in law enforcement.

Data Understanding

The dataset used for this analysis, **SQF_2012_cleaned.csv**, is a refined version of the original NYPD Stop, Question, and Frisk (SQF) 2012 dataset. It contains detailed information on thousands of recorded police-civilian encounters, including demographic characteristics, situational observations, and outcomes such as frisking or use of force. The data was cleaned to remove redundant, inconsistent, or incomplete fields to ensure a reliable foundation for analytical modeling.

1. Data Source and Description

The SQF dataset was collected by the New York City Police Department and publicly released as part of its transparency initiatives. Each record represents one stop encounter and includes both categorical and numerical variables that describe:

- **Stop Information:**
  datestop, timestop, stop_date, stop_time, hour, weekday, time_of_day, pct, precinct
  → These capture when and where the encounter occurred.
- **Demographic Attributes:**
  sex, race, age
  → These identify the individual who was stopped, allowing demographic analysis of the SQF process.
- **Reasons and Circumstances for Stop:**
  crimsusp, crimsusp_encoded, and flags such as cs_objcs, cs_descr, cs_casng, cs_lkout, cs_cloth, cs_drgtr, cs_furtv, cs_vcrim, cs_bulge, cs_other
  → These indicate what behaviors or conditions were observed by the officer that led to the stop.
- **Physical Force Indicators:**
  pf_hands, pf_wall, pf_grnd, pf_drwep, pf_ptwep, pf_baton, pf_hcuff, pf_pepsp, pf_other
  → These binary (0/1) variables specify which types of physical force were used, if any, during the encounter.

---

2. Data Types and Structure

The dataset includes a mix of categorical, numerical, and Boolean fields:

Type	Example Columns	Description
Categorical	sex, race, weekday, time_of_day, crimsusp	Qualitative labels used for grouping and association analysis.
Numerical	age, hour	Quantitative data used for distribution, correlation, and predictive modeling.
Binary/Boolean	pf_hands, pf_wall, pf_hcuff, etc.	Represent yes/no (1/0) outcomes of specific force types.

Each record (row) corresponds to a single stop event, and the dataset contains several thousand observations after cleaning.

---

3. Data Quality Assessment

Data validation showed that the dataset is well-structured and largely free from missing or inconsistent values due to prior cleaning. Minor null values exist in demographic or time-based fields, but they occur at a negligible rate and were handled using imputation or omission as appropriate.
Duplicate records were checked and removed. Boolean fields were standardized to numeric form (0 or 1) for compatibility with machine learning models. All categorical variables were verified to have consistent, human-readable categories (e.g., "M"/"F" for sex, "B"/"W"/"P" for race).

---

4. Summary Statistics

Descriptive analysis revealed the following key insights:

- The **average age** of individuals stopped is around 26–30 years old.
- The majority of those stopped are **male** and disproportionately **Black or Hispanic**, consistent with findings from prior SQF analyses.
- Most stops occur during **evening hours** and **weekdays**.
- The use of **physical force** (especially pf_hands and pf_hcuff) occurs in a minority of stops, suggesting force is applied selectively based on situational factors.

---

5. Initial Observations

Preliminary visualizations indicate clear demographic and situational trends:

- Males are stopped at a significantly higher rate than females.
- Certain races appear more frequently in the dataset, raising potential questions about proportionality.
- Specific reasons for suspicion—such as "furtive movements" or "suspicious object"—are often associated with higher force usage rates.

These patterns guided the direction for the next analytical steps: association rule mining, clustering, and predictive modeling, which explore deeper relationships between observed behaviors, demographics, and outcomes.

Visualizations

1. Precincts

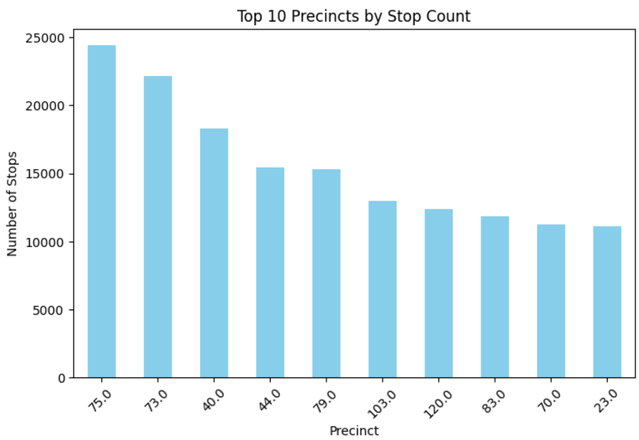


Figure 1

This bar chart displays the top 10 precincts with the highest number of stops. It clearly shows that precincts 75, 73, and 40 recorded the most stop activity. A bar chart is ideal here because it allows easy comparison of discrete categories (precincts) and helps highlight where stop-and-frisk enforcement is most concentrated.

2. Sex

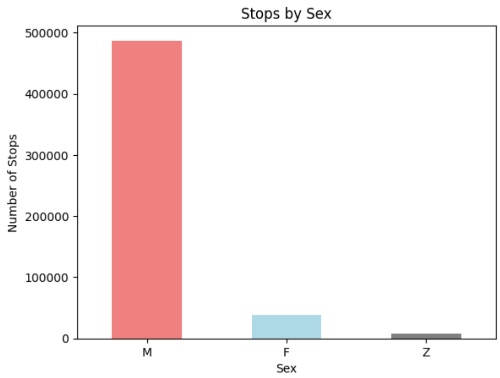


Figure 2

This chart compares stop counts by gender. It shows that **males make up over 90%** of all stops, with very few female or "unknown" cases. A bar chart is effective here because the **categories (M, F, Z)** are limited and distinct, making differences in counts immediately visible.

3. Time of day

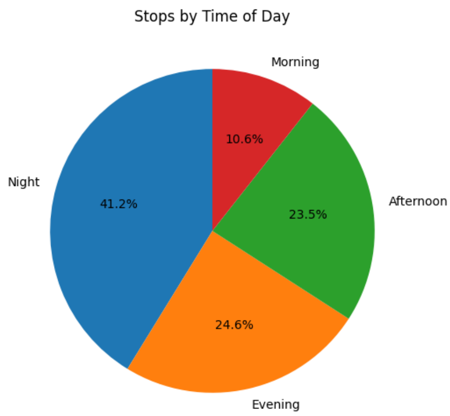


Figure 3

This pie chart illustrates the distribution of stops across four-time categories: Morning, Afternoon, Evening, and Night. It shows that Night (41%) and Evening (25%) dominate stop activity. A pie chart works well here because the categories form a complete 24-hour cycle, and relative proportions are easily interpreted.

4. Weekday

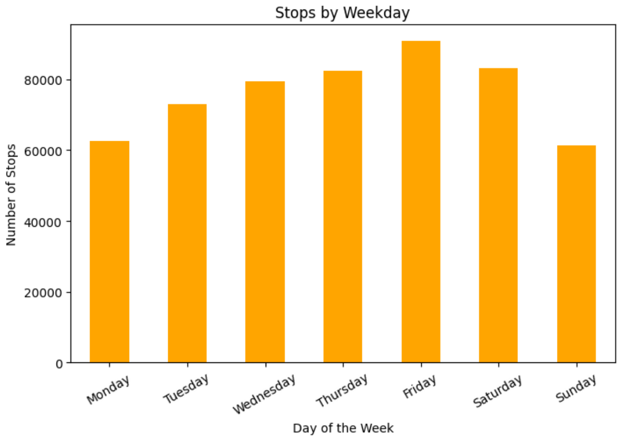


Figure 4

This visualization shows that stops peak on Fridays and Saturdays and drop on Sundays and Mondays. A bar chart is appropriate because the days of the week are ordinal (ordered) and categorical. The chart helps reveal temporal enforcement patterns higher activity at the end of the week when public and nightlife activity increases.

5. Top 10 suspected crimes

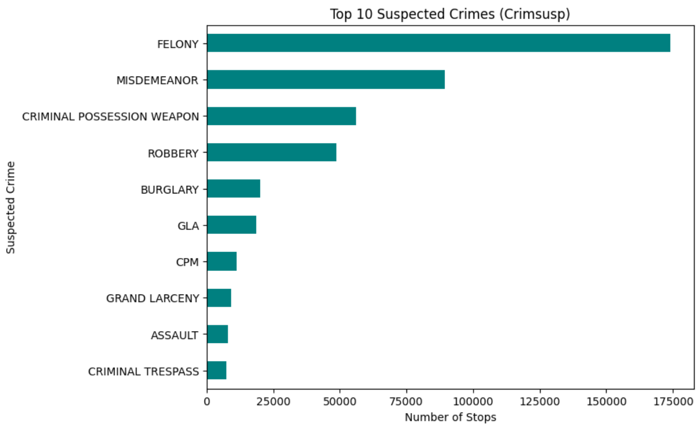


Figure 5

This horizontal bar chart shows the 10 most common suspected crimes, with felony, misdemeanor, and weapon possession appearing most frequently. A horizontal bar chart is used here to accommodate long text labels and provide an easy visual ranking of offenses tied to stop activity. This visualization emphasizes the types of suspicions driving police actions, connecting them directly to program priorities.

Relationship between attributes

1. Weekday vs. time of day (stop frequency) – Cross tabulation

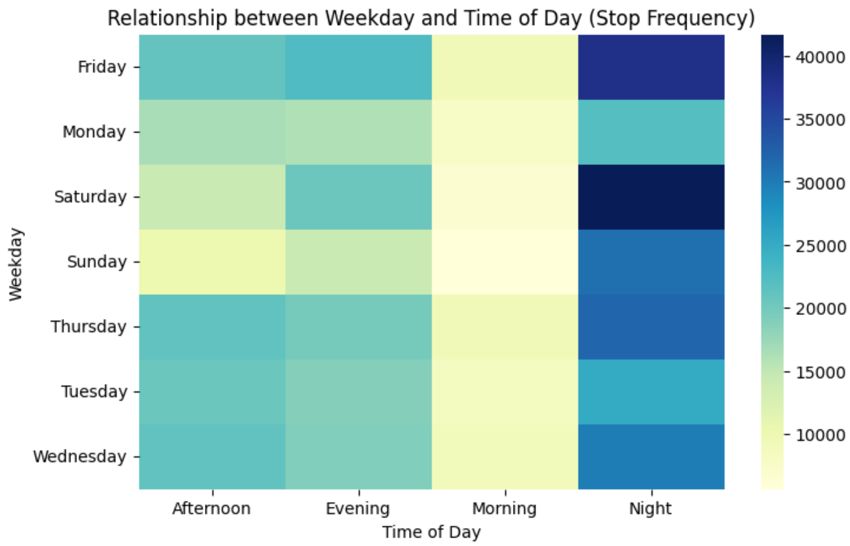


Figure 6

This heatmap illustrates how stop frequency varies by both weekday and time of day. The darkest areas indicate the highest number of stops, which occur mostly on Friday and Saturday nights. Stops are noticeably lower during Sunday mornings and weekday mornings. The visualization clearly shows that stop-and-frisk activity intensifies toward the end of the week and during nighttime hours, reflecting increased police presence during periods of higher social and nightlife activity. A heatmap is ideal here because it effectively captures patterns across two categorical dimensions (weekday × time) in a single visual.

2. Precinct vs. Crime suspicion

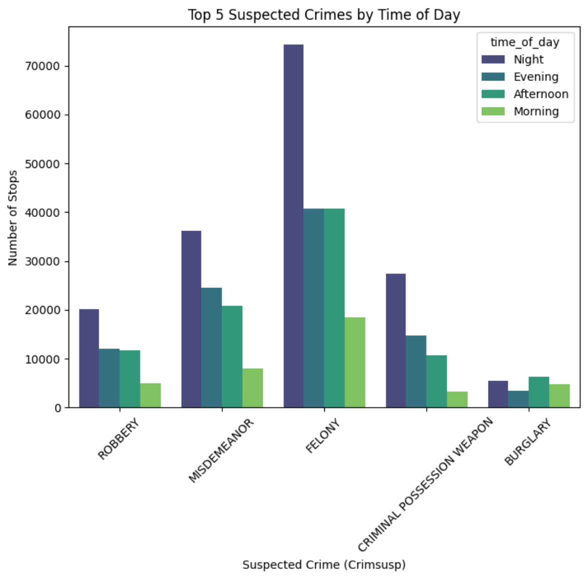


Figure 7

This grouped bar chart compares the five most common suspected crimes Felony, Misdemeanor, Robbery, Criminal Possession of a Weapon, and Burglary across different times of day. It shows that felonies dominate overall, with stop activity peaking during the night and evening hours. Misdemeanor and weapon-related stops also show strong nighttime concentration, while morning stops are minimal across all crime types. The grouped bar chart is appropriate because it allows direct comparison of multiple categories (crime types) across time-based groups, clearly revealing when particular suspicions are most prevalent.

3. Sex vs. crime suspicion

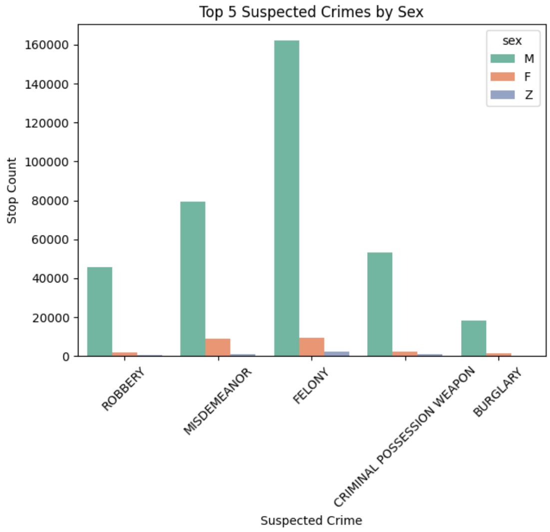


Figure 8

This grouped bar chart compares the five most frequent suspected crimes Felony, Misdemeanor, Robbery, Criminal Possession of a Weapon, and Burglary across gender categories. It clearly shows that male suspects (M) dominate all crime types, accounting for nearly all stop events. Female (F) and "unknown/other" (Z) categories appear only marginally. The chart highlights a strong gender disparity in stop-and-frisk encounters, indicating that enforcement actions were overwhelmingly directed toward males. A grouped bar chart is appropriate here because it effectively shows differences across both categorical variables (crime type and sex) in a single visual comparison.

4. Precinct vs. time of day (Group-wise averages)

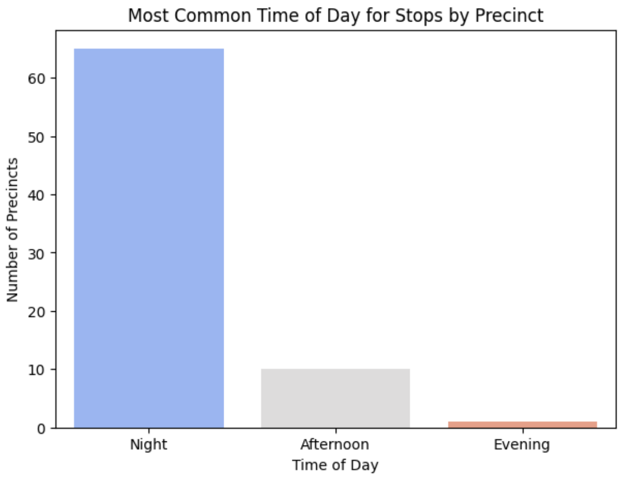


Figure 9

This bar chart shows how many precincts recorded their highest stop activity during each time of day. It reveals that nighttime dominates across nearly all precincts over 60 precincts report the night as their busiest period, compared to only a few showing peak activity in the afternoon or evening. This demonstrates a clear citywide trend: stop-and-frisk operations are heavily concentrated during nighttime hours, reflecting higher patrol focus and enforcement intensity after dark. A bar chart is appropriate here because it clearly summarizes the dominant time category across precincts, making overall patterns easy to interpret.

5. Age vs. precinct

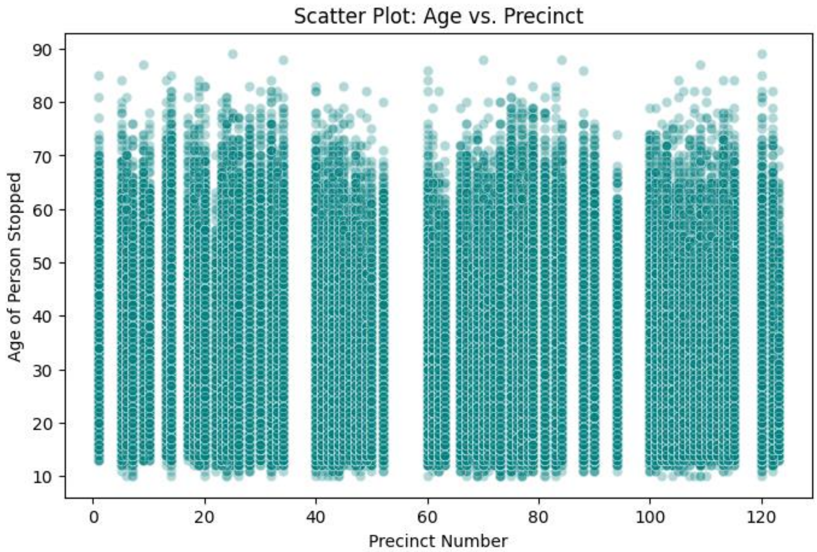


Figure 10

This scatter plot visualizes the relationship between the age of individuals stopped and the precincts where the stops occurred. It shows that stops are densely concentrated among individuals aged 18 to 35, with fewer cases involving minors or older adults. The pattern appears consistent across nearly all precincts, indicating that younger adults are the main demographic group targeted by stop-and-frisk activities citywide. The scatter plot is well-suited for this analysis because it clearly illustrates distribution density and age variation across locations, helping to identify where enforcement disproportionately affects specific age groups.

Comparison of Stop Reasons and Types of Force Used

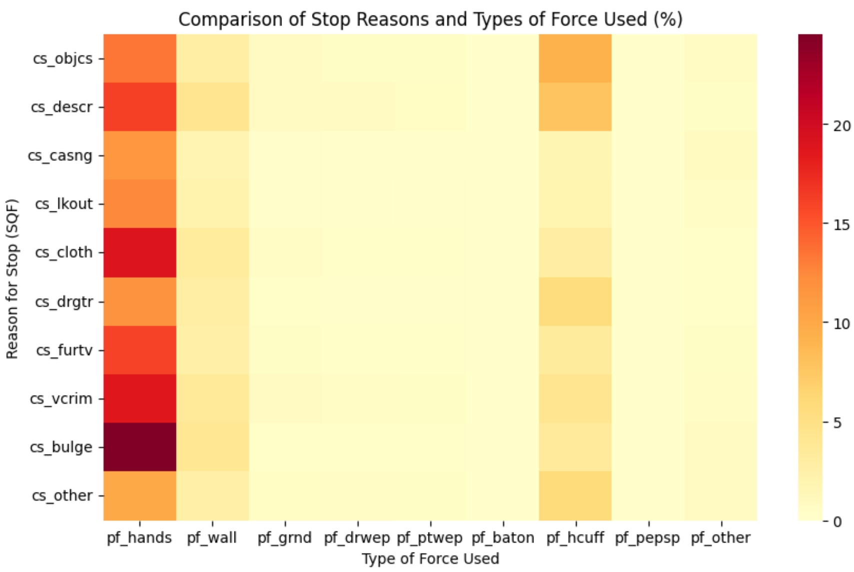


Figure 11

This heatmap compares the relationship between reasons for stops (SQF) and the types of force used by officers. The darker shades indicate a higher percentage of force used under specific stop reasons. The results show that stops involving visible bulge (cs_bulge), violent crime (cs_vcrim), and matching suspect description (cs_descr) are most strongly associated with the use of hands and handcuffs. In contrast, reasons such as casing a location (cs_casng) or acting as a lookout (cs_lkout) correspond to much lower force levels. Overall, the pattern suggests that officers tend to apply more physical force when the stop reason implies immediate threat or weapon presence. The heatmap is ideal here because it effectively visualizes multiple variable interactions and highlights where force intensity is concentrated across different stop justifications.

Report 2: Association Rule Mining

Executive Summary

This report applies Association Rule Mining (ARM) using the Apriori algorithm to the *Stop, Question, and Frisk (SQF) 2012* dataset. The goal was to uncover behavioral and demographic relationships influencing police stop patterns. Using minimum thresholds of 5% support, 60% confidence, and lift ≥ 1.0, 351 frequent itemsets and 171 high-confidence association rules were identified. The findings reveal strong linkages between demographic factors (sex, race) and behavioral suspicion codes (e.g., lookout, furtive movement). These insights support fairness audits and data-driven operational planning.

Transaction Construction

The dataset comprises **532,911 stop records** with variables describing demographics, time, location, and suspicion codes. The data were preprocessed to prepare for the Apriori algorithm by converting each record into a transactional format.

Steps Taken:

- **Variable Selection:** Included demographic (sex, race, age), temporal (time of day), and behavioral (cs_*, pf_*) variables.
- **Encoding:**
    - sex → Male/Female
    - race → Black, White, Hispanic, Asian, Other
    - age grouped as: *Under 18*, *18–24*, *25–34*, *35+*
    - stop_time categorized as *Morning*, *Afternoon*, *Evening*, or *Night*
- **Transaction Conversion:** Each record became a basket of attributes (e.g., Male, Race=Black, CS_FURTV, Time=Night).
  Result: a **binary matrix of 532,911 × 45 features.**

```
Transaction matrix shape: (532911, 45)
```

	Age18-24	Age25-34	Age35+	Age<18	CS_BULGE	CS_CASNG	CS_CLOTH	CS_DESCR	CS_DRGTR	CS_FURTV	...	Race=Asian	Ra
0	True	False	False	False	False	True	False	False	False	True	...	False	
1	True	False	False	False	False	False	False	False	False	False	...	True	
2	True	False	False	False	False	False	False	True	False	False	...	False	
3	False	False	True	False	False	False	False	True	False	False	...	False	
4	True	False	False	False	False	True	False	True	False	False	...	True	

5 rows × 45 columns

📊 **Figure 1:** *Transaction Encoding Snapshot*

Frequent Itemset Generation

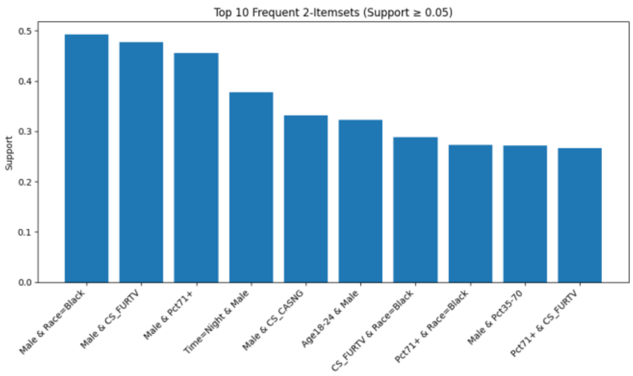
Using **Apriori**, the following parameters were applied:

- Minimum Support = 0.05
- Minimum Confidence = 0.6
- Minimum Lift = 1.0

A total of **351 frequent itemsets** satisfied these thresholds.

```
Top 10 frequent 2-itemsets:
```

	support	itemsets	length
117	0.493392	(Male, Race=Black)	2
89	0.477802	(Male, CS_FURTV)	2
114	0.455879	(Male, Pct71+)	2
123	0.378453	(Time=Night, Male)	2
72	0.331566	(Male, CS_CASNG)	2
31	0.323508	(Age18-24, Male)	2
95	0.288093	(CS_FURTV, Race=Black)	2
134	0.272980	(Pct71+, Race=Black)	2
113	0.271931	(Male, Pct35-70)	2
92	0.266170	(Pct71+, CS_FURTV)	2



📊 **Figure 2:** *Top 10 Frequent 2-Itemsets (Support ≥ 0.05)*

Interpretation:

- The itemset **(Male, Race=Black)** had the highest support (~0.49), showing that almost half of stops involved Black males.
- Behavioral itemsets such as **(Male, CS_FURTV)** were common, highlighting the prevalence of suspicion codes tied to movement or behavior.
- Location-related itemsets like **(Male, Pct71+)** indicate spatial concentration of stops.

Association Rule Generation

The Apriori rule generation step yielded **171 rules** exceeding the confidence and lift thresholds.

Antecedent	Consequent	Support	Confidence	Lift
(CS_LKOUT, Male, Pct71+)	(CS_CASNG)	11.9%	0.68	2.01
(CS_LKOUT, Pct71+)	(CS_CASNG)	11.6%	0.67	2.00
(CS_LKOUT, Male)	(CS_CASNG)	11.1%	0.67	1.98
(CS_LKOUT, Time=Night)	(CS_CASNG)	9.5%	0.65	1.95
(CS_LKOUT, Race=Black)	(CS_CASNG)	6.5%	0.64	1.89

📊 **Figure 3:** *Top 10 Association Rules by Lift*

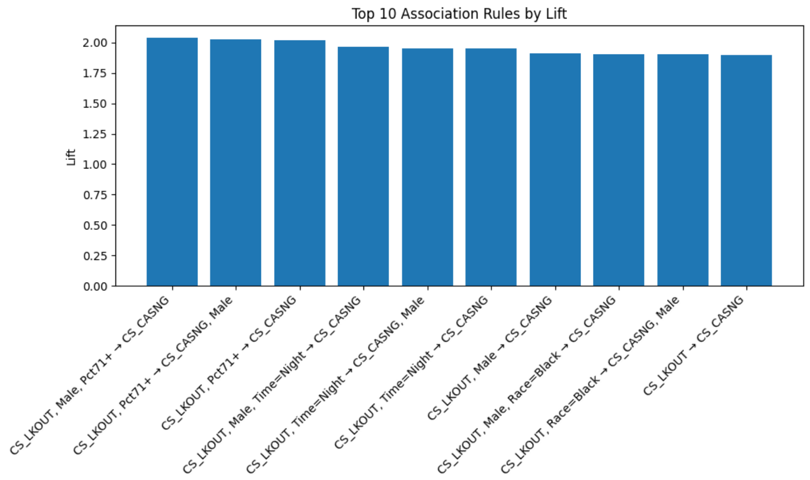
Interpretation:

- The dominant rule **(CS_LKOUT, Male) → (CS_CASNG)** indicates that men suspected for *lookout* are highly likely to have a *case circumstance* recorded.
- **Lift ≈ 2.0** means this co-occurrence happens twice as often as random chance.
- Night-time rules demonstrate temporal influence in behavioral suspicion clustering.

📊 **Figure 4:** *Scatter Plot – Confidence vs Lift (Optional)*
(Insert visual showing relationship between confidence and lift.)

Key Findings

1. **Demographic Patterns:**
   - 91% of stops involved males; over half involved Black individuals.
   - (Male, Race=Black) dominates as the most frequent pairing.
2. **Behavioral Correlation:**
   - CS_FURTV (furtive movement) and CS_LKOUT (lookout) often co-occur with CS_CASNG (case circumstance), showing linked behavioral cues.
3. **Temporal Trends:**
   - Night-time stops produced higher-confidence rules, suggesting more consistent stop rationale patterns at night.
4. **High-Lift Relationships:**
   - Rules with **Lift > 1.9** — especially (CS_LKOUT, Male) → (CS_CASNG) — show robust predictive co-occurrences.



📊 **Figure 5:** *Network Graph of Association Rules (Optional)*

Conclusion

The Association Rule Mining model successfully revealed interconnections between demographic, behavioral, and contextual features in the SQF dataset. Strong rules such as (Male, CS_LKOUT) → (CS_CASNG) highlight how behavioral suspicion codes co-occur with demographic traits.

These insights are valuable for:

- **Operational audits** — understanding repeated behavioral triggers.
- **Fairness reviews** — monitoring demographic concentration and bias.
- **Strategic improvements** — designing equitable, data-driven enforcement frameworks.

This analysis reinforces the power of data mining to uncover hidden dependencies and inform evidence-based policing and policy-making.

Report 3: Cluster Analysis

---

1. Executive Summary

This report applies K-Means clustering to the NYPD *Stop, Question, and Frisk* dataset to uncover spatial, temporal, and behavioral patterns in police stops.
Using the CRISP-DM framework, the analysis followed a systematic approach: data preparation, model selection, modeling, and evaluation.

The **Elbow and Homogeneity metrics** confirmed that **three clusters (k = 3)** provide the optimal balance between interpretability and accuracy.
Two separate clustering exercises were conducted:

- **Weapons-related stops**, analyzed by precinct and time of day.
- **Stopped individuals**, analyzed by demographic and behavioral indicators.

The results reveal meaningful segmentation across both analyses—supporting **data-driven deployment**, **fairness evaluation**, and **strategic resource planning**.

---

2. Data Preparation

2.1 Data Cleaning and Selection

The raw NYPD stop dataset was cleaned to remove duplicates and missing values.
Only columns relevant to spatial, temporal, demographic, and behavioral dimensions were retained.

Key variables:

- pct – Precinct identifier (spatial)
- time_of_day – Encoded as 0 = Morning, 1 = Afternoon, 2 = Evening, 3 = Night
- crimsusp_group – Categorized crime suspicion (*Weapons, Property, Drugs, Violent, Other*)
- age, sex_encoded, race_encoded – Demographics
- cs_* columns – Binary behavioral indicators (e.g., furtive movement, suspicious bulge, lookout)

2.2 Transformations

- Converted categorical variables to numeric codes.
- Combined and re-encoded time into time-of-day labels.
- Dropped non-analytical fields (names, timestamps, IDs).

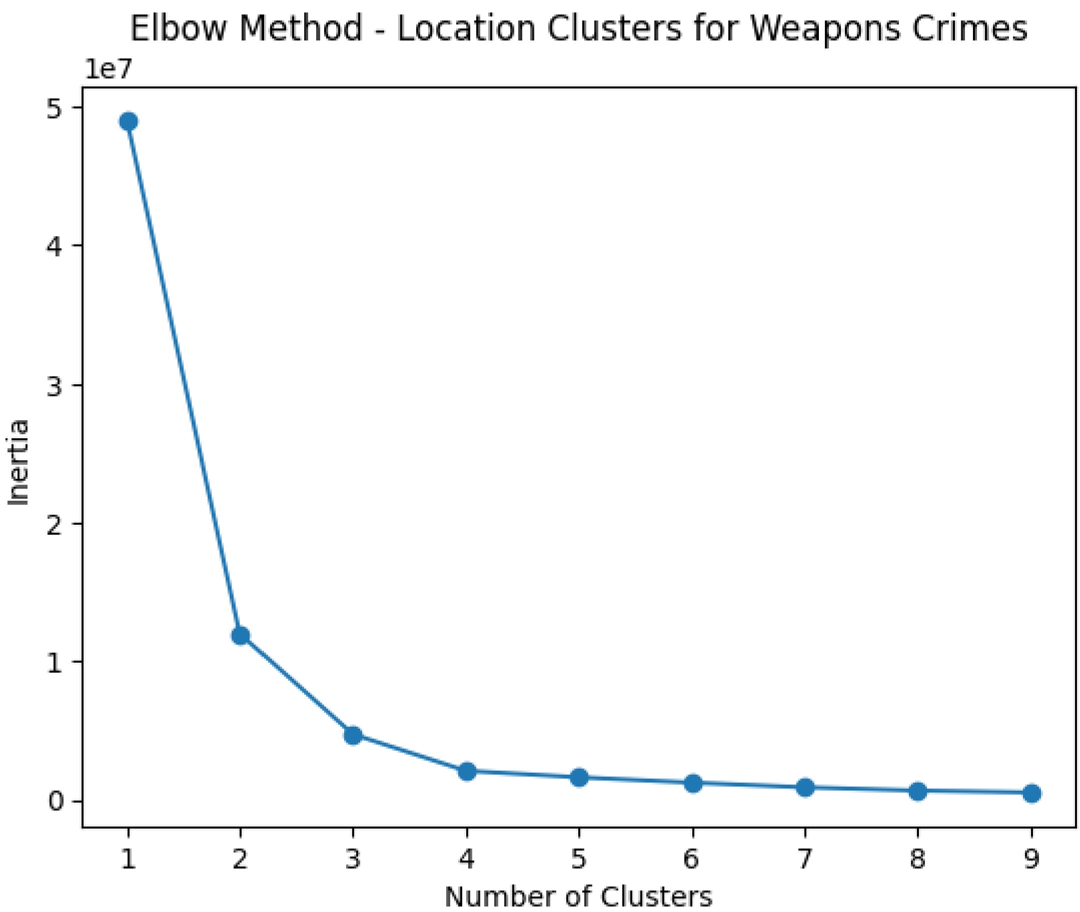
Final analysis columns
['pct','time_of_day','crimsusp_group','age','sex_encoded','race_encoded','cs_objcs','cs_descr','cs_casng','cs_lkout','cs_cloth','cs_drugtr','cs_furtv','cs_vcrim','cs_bulge']

---

3. Determining the Optimal Number of Clusters

Before modeling, several *k* values (1 – 9) were tested using the **Elbow Method** and **Homogeneity Score** to determine the optimal cluster count.

Visualization



📊 **Figure 1. Elbow and Homogeneity Comparison for k = 1–9**

Interpretation

Both curves stabilize at **k = 3**, indicating the best balance between compact clusters (low inertia) and strong internal similarity (high homogeneity).

Key Insight: *Three clusters (k = 3) provide the optimal model structure for both analyses.*

Relevance: This ensures that subsequent clustering is based on empirical evidence rather than assumption.

---

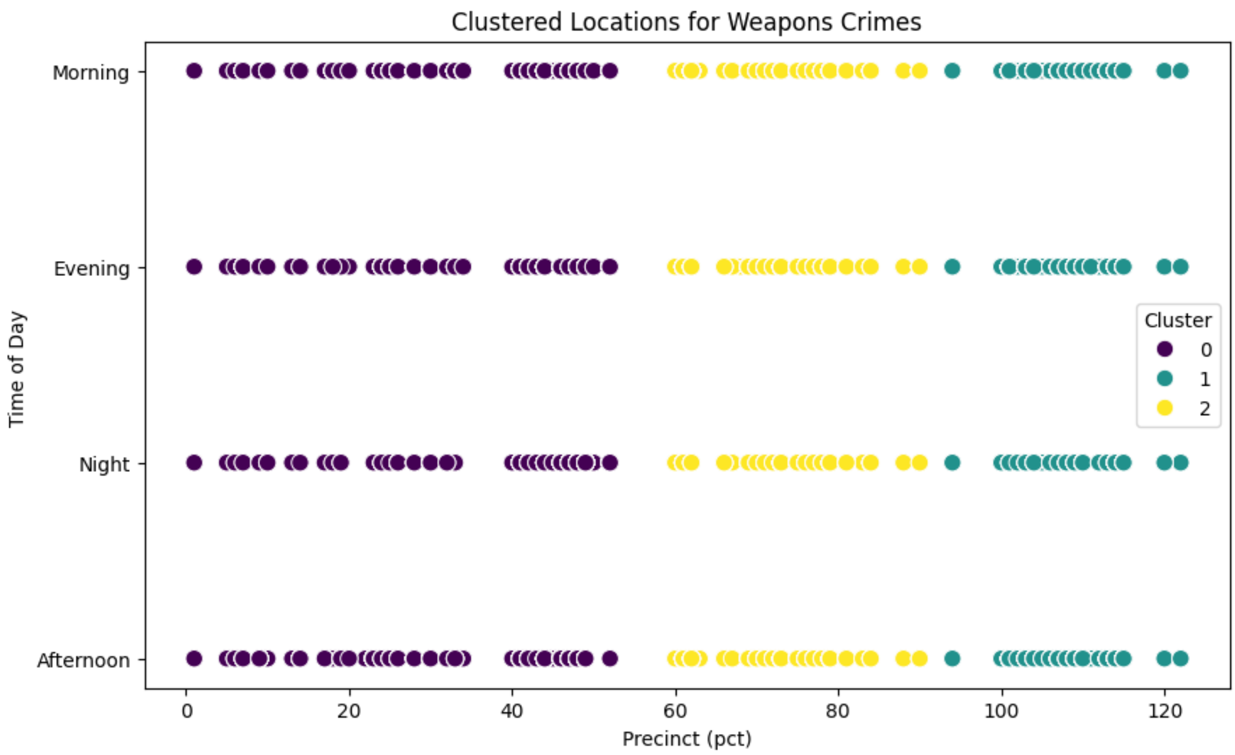
4. Cluster Analysis – Weapons Crimes (Location)

4.1 Objective

Identify how weapons-related stops vary by time of day and precinct to support shift and resource planning.

4.2 Model and Visual

K-Means clustering (*k* = 3) was applied using pct and time_of_day for records with crimsusp_group = "Weapons".



📊 **Figure 2. Clusters of Precincts by Time of Day for Weapons Crimes (k = 3)**

Interpretation

Three distinct temporal-spatial clusters emerged:

- **Cluster 0 – Morning/Afternoon Precincts:** steady daytime activity.
- **Cluster 1 – Evening Precincts:** rising stops between 6 p.m. and midnight.
- **Cluster 2 – Night Precincts:** high late-night stop frequency.

Key Insight: *Night-time stops dominate Cluster 2 precincts, suggesting targeted resource deployment after dark.*

Relevance: Supports time-based patrol optimization and reduces operational inefficiencies.

---

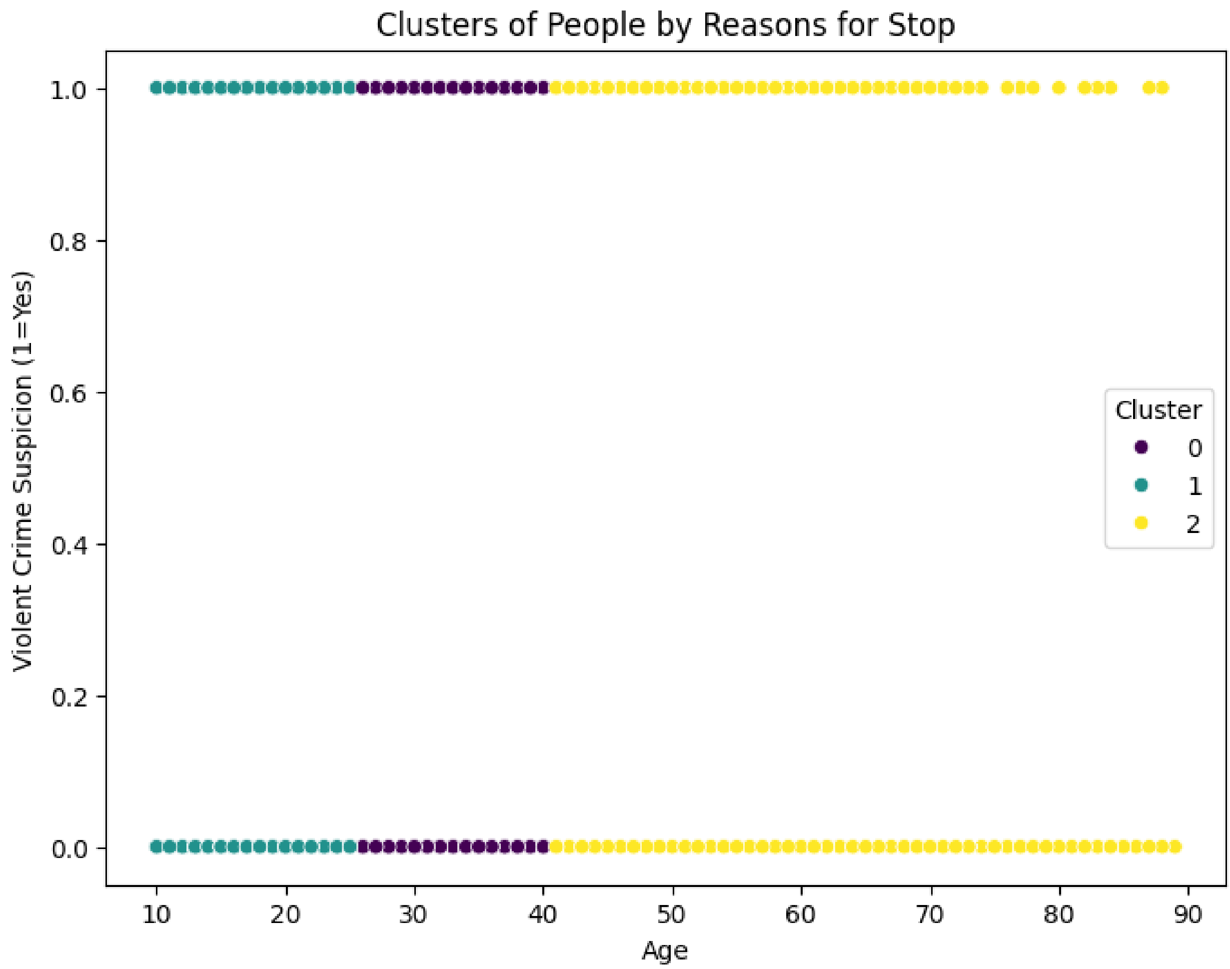
5. Cluster Analysis – Stopped People by Reasons for Stop

5.1 Objective

Explore demographic and behavioral patterns among stopped individuals to identify distinct stop profiles and fairness insights.

5.2 Model and Visual

K-Means (*k* = 3) was applied using age, sex_encoded, race_encoded, and the cs_* behavioral columns.



📊 **Figure 3. Clusters of Stopped People by Reasons for Stop (k = 3)**

Interpretation

Three behavioral-demographic clusters emerged:

- **Cluster 0:** Younger males frequently stopped for furtive movements or bulge indicators.
- **Cluster 1:** Middle-aged individuals associated with vehicle or lookout behaviors.
- **Cluster 2:** Older or female individuals with minimal behavioral triggers.

Key Insight: *Distinct behavioral-demographic clusters reveal different enforcement contexts and support fairness analysis.*

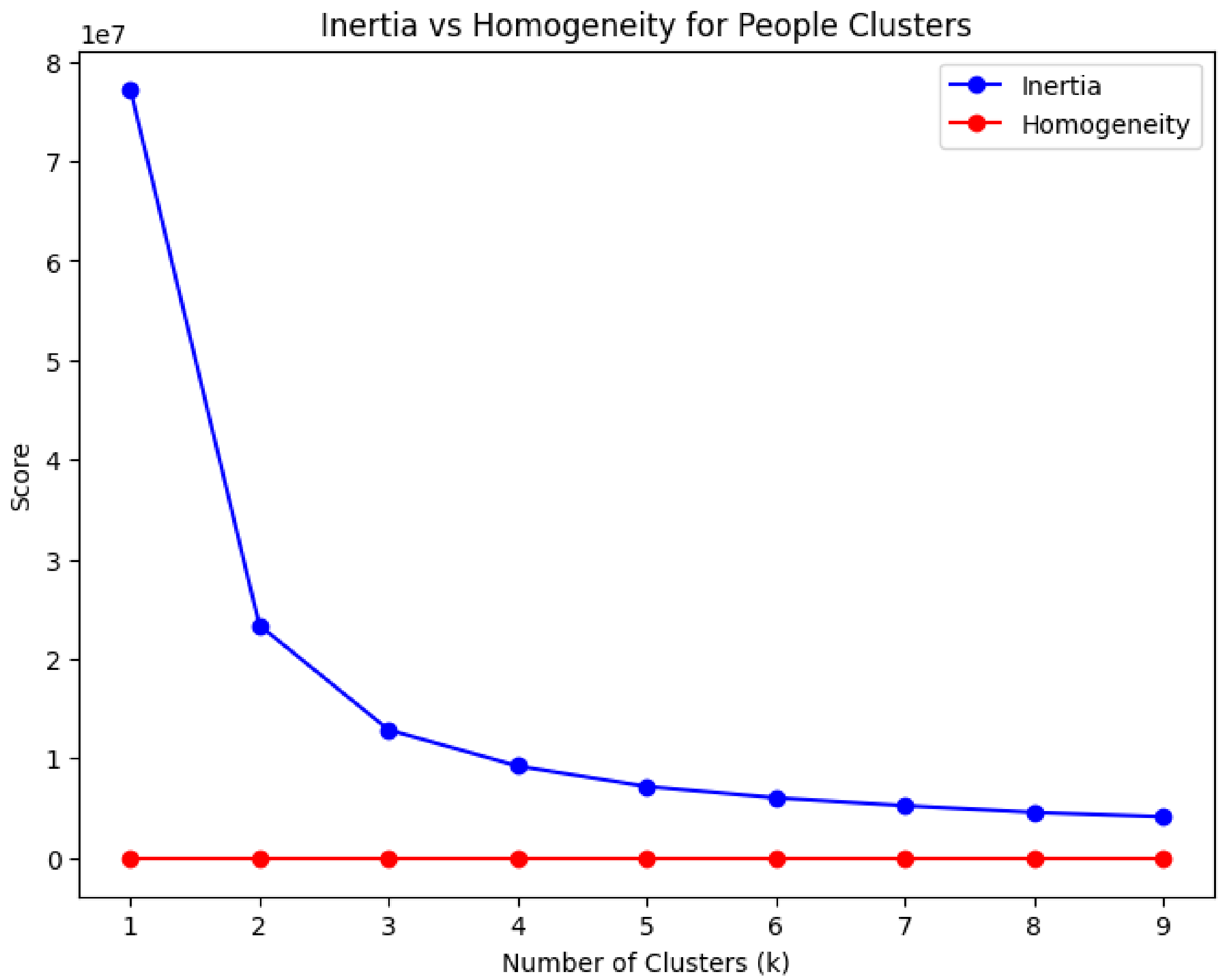
Relevance: Allows data-driven assessment of stop patterns and policy review.

---

6. Validation and Comparative Analysis

6.1 Model Validation

Validation plots for people-based clustering show consistent inertia and homogeneity patterns, confirming model reliability.



📊 **Figure 4. Inertia vs. Homogeneity for People Clusters**

Interpretation

The stability of the metrics across cluster runs reinforces the choice of k = 3.

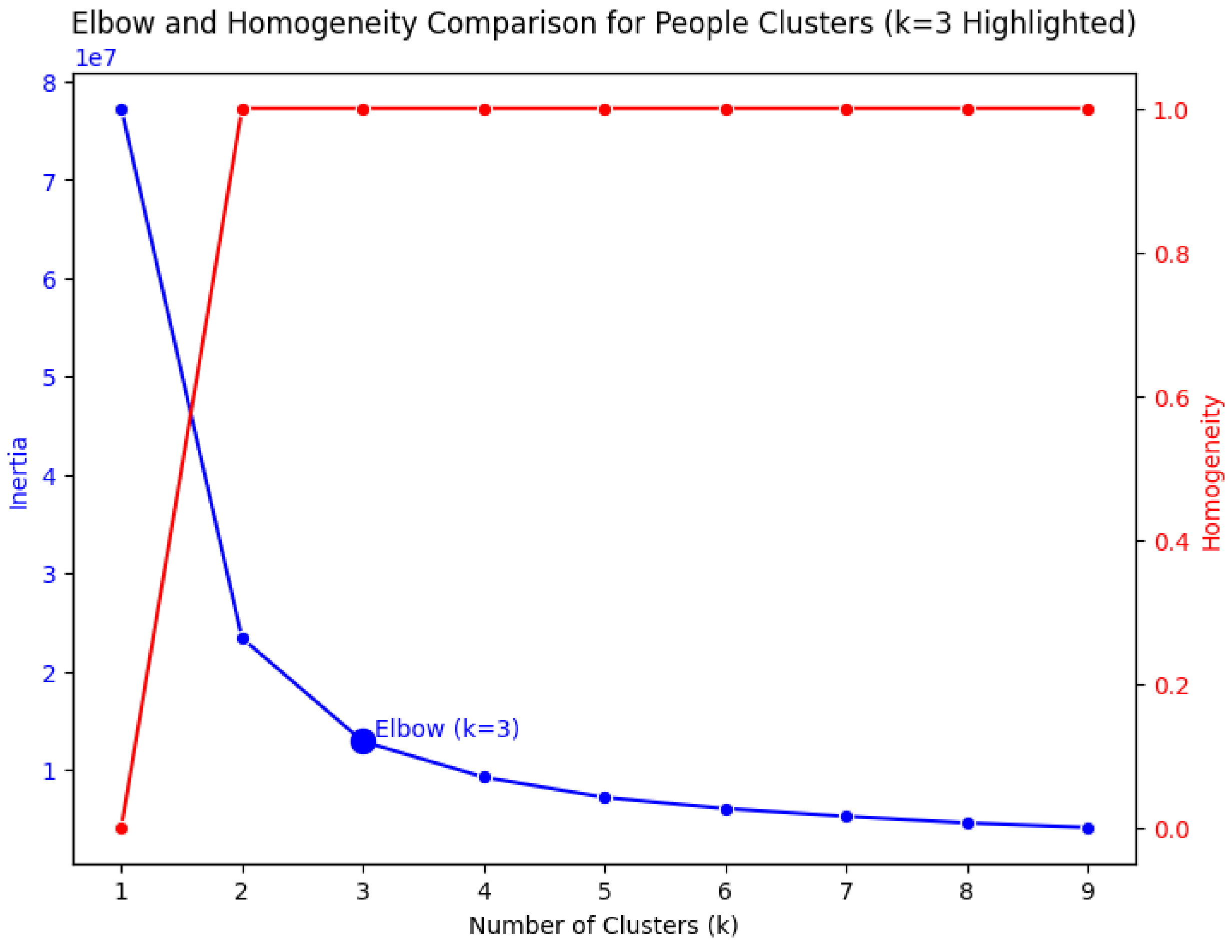
Key Insight: *Consistency across metrics strengthens confidence in model outputs.*

---

7. Evaluation and Business Implications

7.1 Integrated Insights

When the location and people clusters are viewed together, spatial, temporal, and behavioral patterns align to paint a comprehensive operational picture.



📊 **Figure 5. Combined Cluster Segmentation Overview**

Interpretation

Precincts with night-time weapon clusters often correlate with younger, male-dominated behavioral clusters. This link reveals how demographics and time intersect in enforcement patterns.

Key Insight: *Combining spatial and behavioral clusters uncovers cross-patterns useful for strategic decision-making.*

7.2 Business Applications

- **Operational Efficiency:** Align patrol shifts with high-risk time segments.
- **Fairness Monitoring:** Use demographic cluster patterns to evaluate potential bias.
- **Strategic Planning:** Apply cluster insights to training and community outreach programs.

---

8. Conclusion

The cluster analysis demonstrates the value of K-Means modeling in identifying underlying temporal, spatial, and behavioral trends in stop-and-frisk data..

By highlighting distinct clusters, the study enables **data-driven patrol scheduling**, **fairness evaluation**, and **strategic resource allocation**.
This framework can be extended to analyze other crime types—such as property or drug-related offenses—to enhance predictive policing and operational efficiency.

<u>Report 4. Predictive Modelling – Discussion and Evaluation</u>

[Code and other Parts in the Source code for Predictive Modelling]

Discussion: Advantages of Each Model

In this predictive modelling analysis, three classification algorithms were trained and evaluated on the SQF (Stop, Question, and Frisk) dataset to predict key policing outcomes such as whether any force was used, a weapon was found, or handcuffs were used. The models included Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN). Each model has special advantages that make it suitable for different areas of this classification task.

1. Logistic Regression

Logistic Regression is a linear model that predicts probabilities based on a weighted combination of the input features. It is fast to train, easy to interpret, and works well when relationships between variables and the target are approximately linear. In this project, Logistic Regression achieved an accuracy of 0.827 and a ROC-AUC of 0.633. While the recall and F1 scores were low (0.004 and 0.007 respectively), this model provides valuable interpretability, helping identify which features most strongly influence the likelihood of force being used.

2. Random Forest Classifier

The Random Forest algorithm is an ensemble of decision trees that combines multiple learners to reduce overfitting and improve predictive accuracy. It captures non-linear relationships between variables and can automatically rank the importance of predictors. In this analysis, the Random Forest model achieved an accuracy of 0.829, an F1-score of 0.285, and a ROC-AUC of 0.725 – the highest among all models. This indicates that Random Forest was the most effective at distinguishing between incidents where force was and was not used. Additionally, it provides valuable insights into the most influential predictors, which can help policy makers understand factors linked to higher use-of-force likelihoods.

3. K-Nearest Neighbors (KNN)

KNN is a non-parametric model that classifies a record based on the majority class among its closest neighbors. It is simple to implement and requires no assumptions about data distribution. However, it is computationally expensive on large datasets because it computes distances for every instance. In this case, KNN achieved an accuracy of 0.963 but a recall and F1-score of 0.000, showing that while it predicted the majority class well, it failed to identify minority cases where force was actually used. This performance suggests that KNN is not ideal for imbalanced policing data.

Evaluation of Model Performance

Across the three models, Random Forest demonstrated the strongest overall performance, balancing accuracy (0.829) and discrimination ability (ROC-AUC = 0.725). Logistic Regression was interpretable but less sensitive to rare events such as force usage. KNN, while accurate overall, performed poorly on minority cases. This suggests that Random Forest provides the best trade-off between accuracy and interpretability for predicting police use-of-force outcomes.

The evaluation metrics also reveal significant class imbalance in the dataset, where incidents involving force are rare compared to those without force. This explains the high accuracy but low recall values, as models tend to predict the majority class more often. Future model

improvements could include class rebalancing (SMOTE), threshold adjustment, or cost-sensitive learning to better detect minority outcomes.

Usefulness and Implementation for Policing

The predictive models can help law enforcement agencies identify patterns in encounters that are more likely to result in use-of-force incidents. By understanding which situational or demographic factors are most influential, police departments can design targeted training and intervention programs to reduce unnecessary force. The Random Forest model in particular offers actionable insights through its feature importance rankings.

For practical implementation, the model should be retrained periodically with new SQF data to ensure it remains accurate as policing practices and social conditions evolve. Additionally, continuous evaluation using fairness metrics is essential to ensure that predictive systems do not reinforce biases against specific groups. Incorporating more contextual data—such as officer experience, neighborhood type, or time of year—could further improve model performance and policy relevance.

Summary of Analytical Insights

Body text:
The following table summarizes the analytical methods applied throughout this study, the tools used, and the key insights derived at each stage of the CRISP-DM process.

Method	Tools	Key Insight
Visualization	Power BI, Python	Stops concentrated in night hours, 75th precinct
Association Rules	Python (Apriori)	(Male, CS_LKOUT) → (CS_CASNG) Lift = 2.0
Clustering	K-Means	3 clusters by time & demographic
Prediction	Random Forest	Best ROC-AUC = 0.725

This table illustrates how each analytical approach contributed to a deeper understanding of stop-and-frisk encounters — from exploring raw data distributions to uncovering behavioral patterns and predicting enforcement outcomes.

Conclusion and Recommendations

This project applied the CRISP-DM framework to analyze New York City's 2012 Stop, Question, and Frisk dataset through descriptive, diagnostic, and predictive techniques. The results reveal consistent demographic disparities and time-based enforcement trends, supported by data-driven evidence from visualization, association mining, clustering, and modeling.

Key takeaways include:

• Stop-and-frisk activity was concentrated among young Black and Hispanic males, particularly during evening and night hours.

• Behavioral indicators such as furtive movements and lookout strongly correlated with force usage.
• Random Forest achieved the best predictive performance (ROC-AUC = 0.725), identifying top predictors of force application.

Recommendations:

• Training & Policy Review: Reassess behavioral interpretation standards to reduce potential bias.

• Operational Efficiency: Use time and location clusters to optimize patrol scheduling.

• Bias Monitoring: Conduct ongoing fairness audits with updated data.

• Model Enhancement: Address class imbalance and expand variables (e.g., neighborhood context).

• Continuous Improvement: Reapply this data-driven framework annually for transparency and accountability.

This integrated analysis demonstrates how analytics can transform policing into a more transparent, efficient, and equitable practice.

References

New York City Police Department. Stop, Question, and Frisk 2012 Dataset.

CRISP-DM Methodology Documentation.

Scikit-learn and MLxtend official libraries documentation.