



**Inspectorate
of Policing**

**Service d'inspection
des services policiers**

**Office of the Inspector General
of Policing**

777 Bay Street

7th Floor

Toronto ON M5G 2C8

**Bureau de l'inspecteur général des services
policiers**

777, rue Bay

7e étage

Toronto ON M5G 2C8

**Comprehensive Data Analysis and Machine Learning Modelling
to Identify Patterns And Influential Factors of Incidents In
Ontario 2023: A Dual-dataset Approach**

Presentation by: Sanjir Inam Salsabil

Friday - 23rd of August, 2024

Flow of the Presentation

- Objective and Data Overview
- Data Exploration, Visualizations and Integration
- Data cleaning, Processing and Standardization
- Exploratory Data Analysis (EDA)
- Potential Case Studies

- Time Series Analysis of Incident Types

- Impact of Individual Attributes on “Use of Force”

- Identification of trends, Significant Correlation Analysis

- Statistical Analysis
- Application of Machine Learning Modeling on cleaned Data

- Models Selection

- Finding Top most Important Features in Predicting Incident types

- Model Evaluation Based on Performance metrics, classification reports and comparing with other models

- Fine-tuning model hyperparameters for better performance and find the best model

- Conclusions and Discussion on Future works

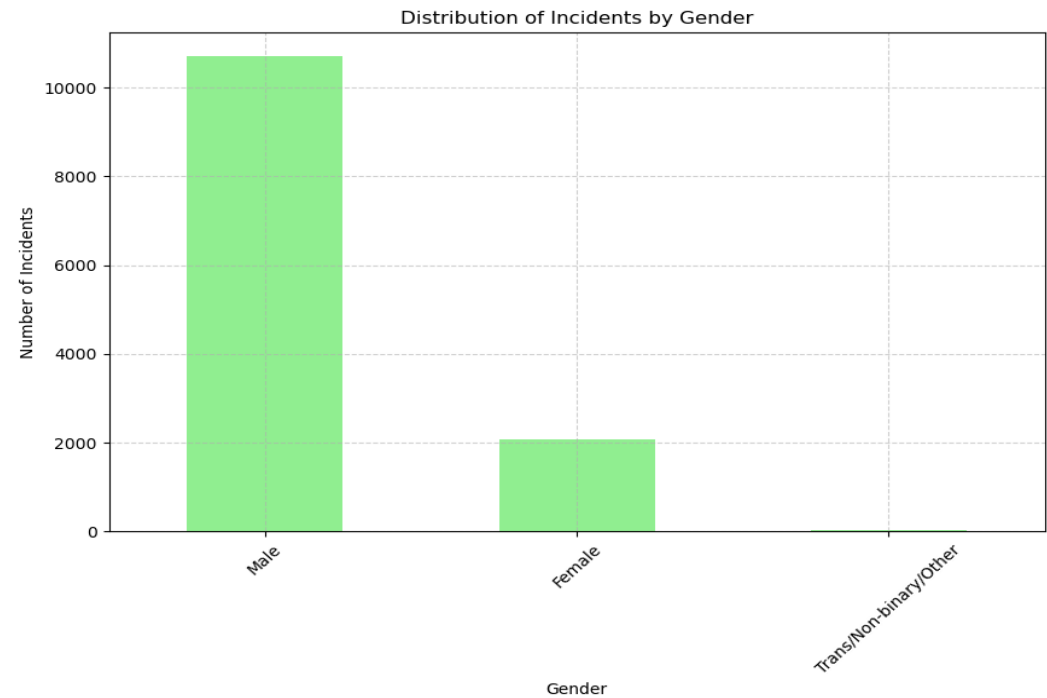
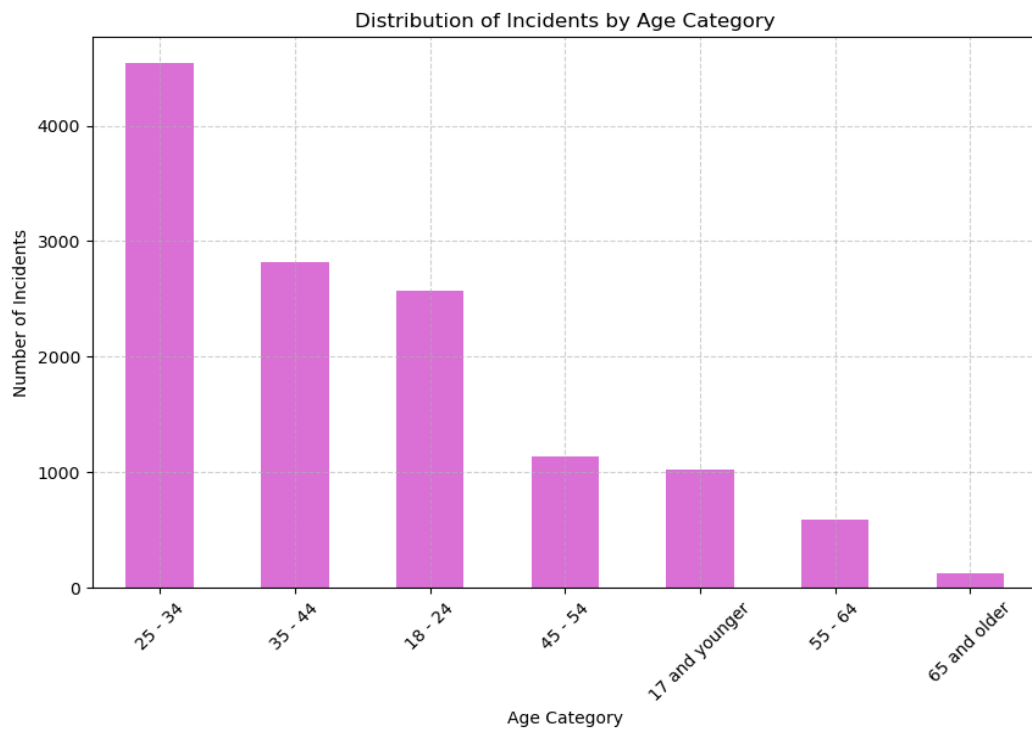
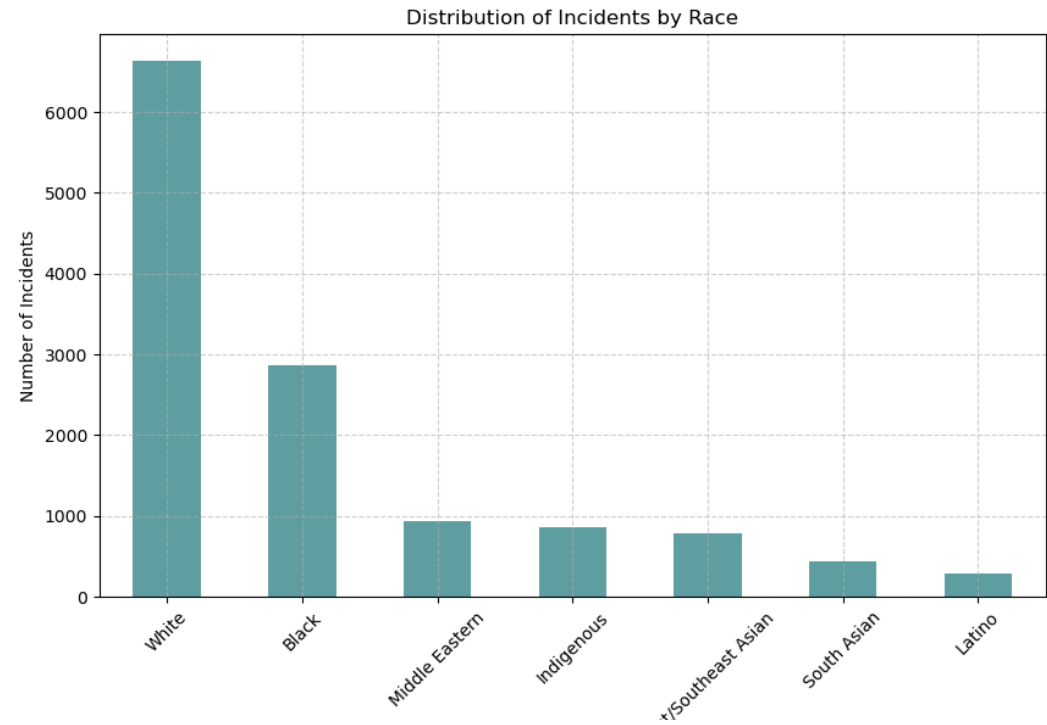
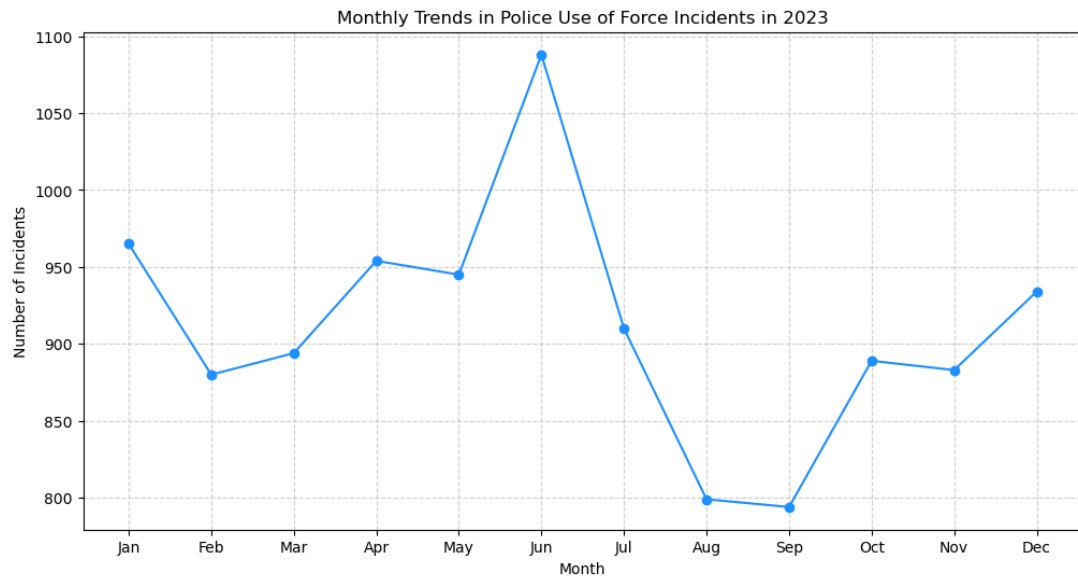
The Primary Idea

Data-Driven Insights for Strategic Decision-Making

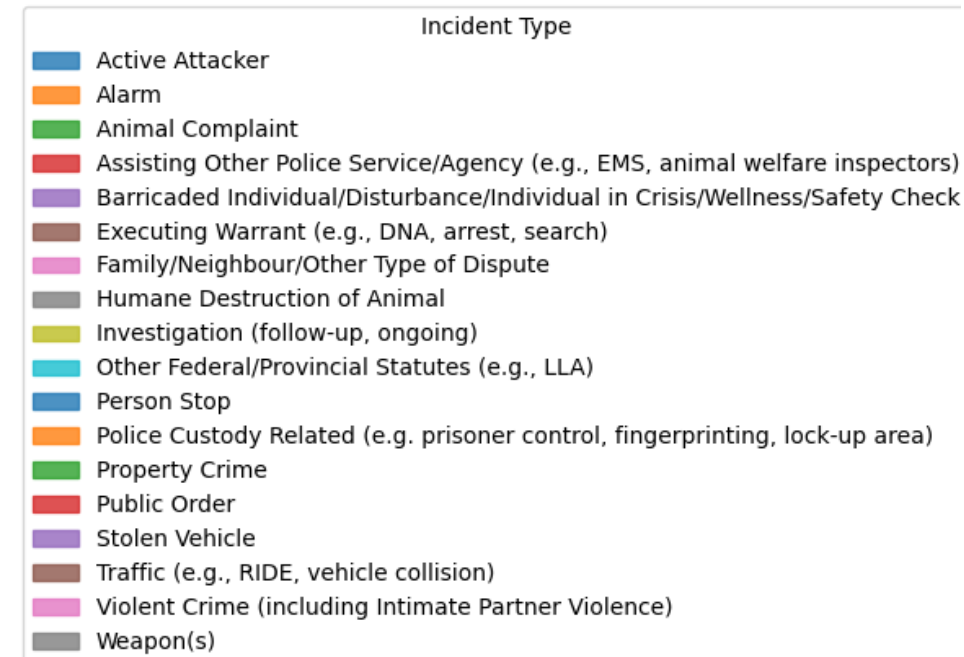
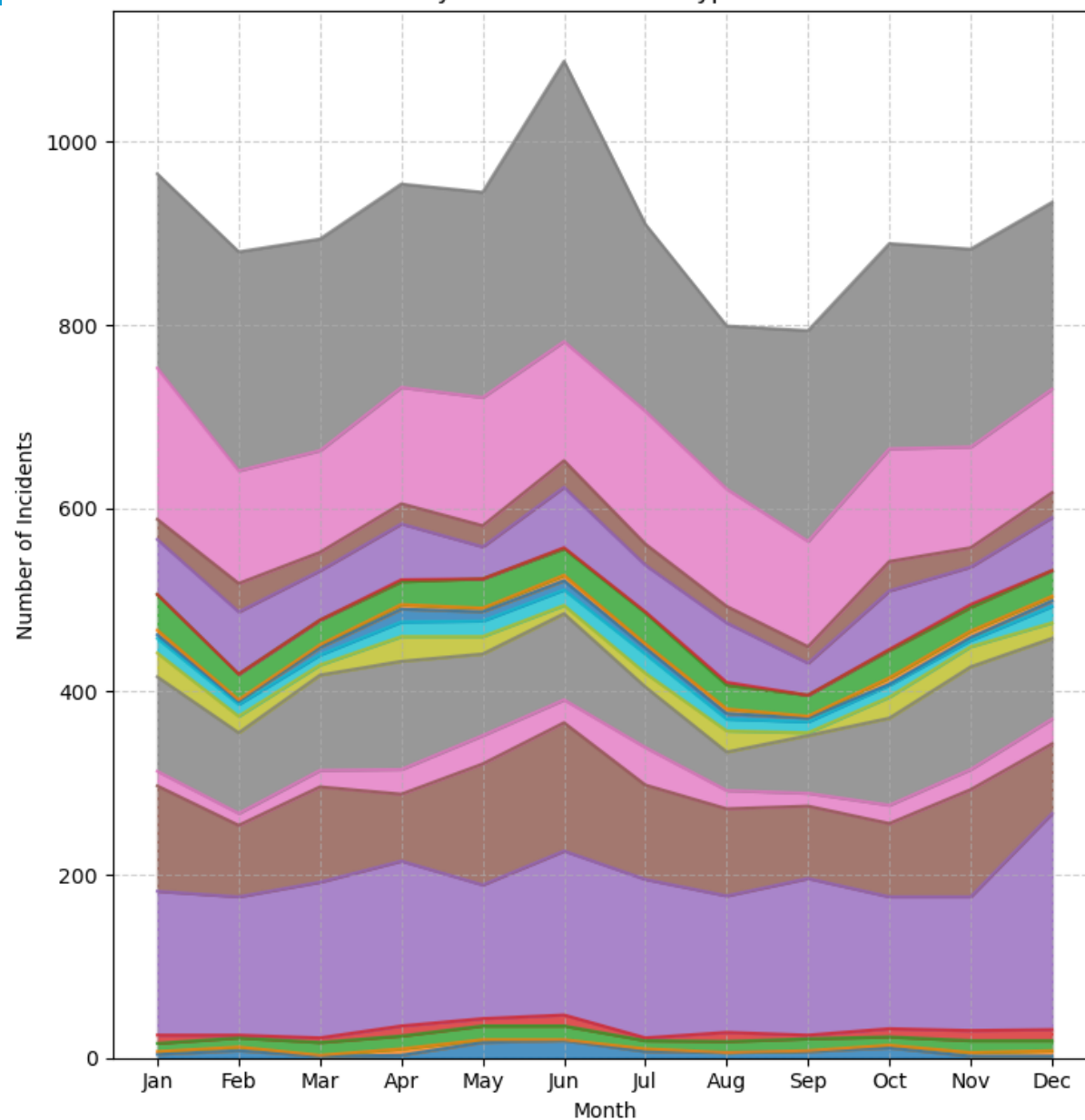
Data Overview

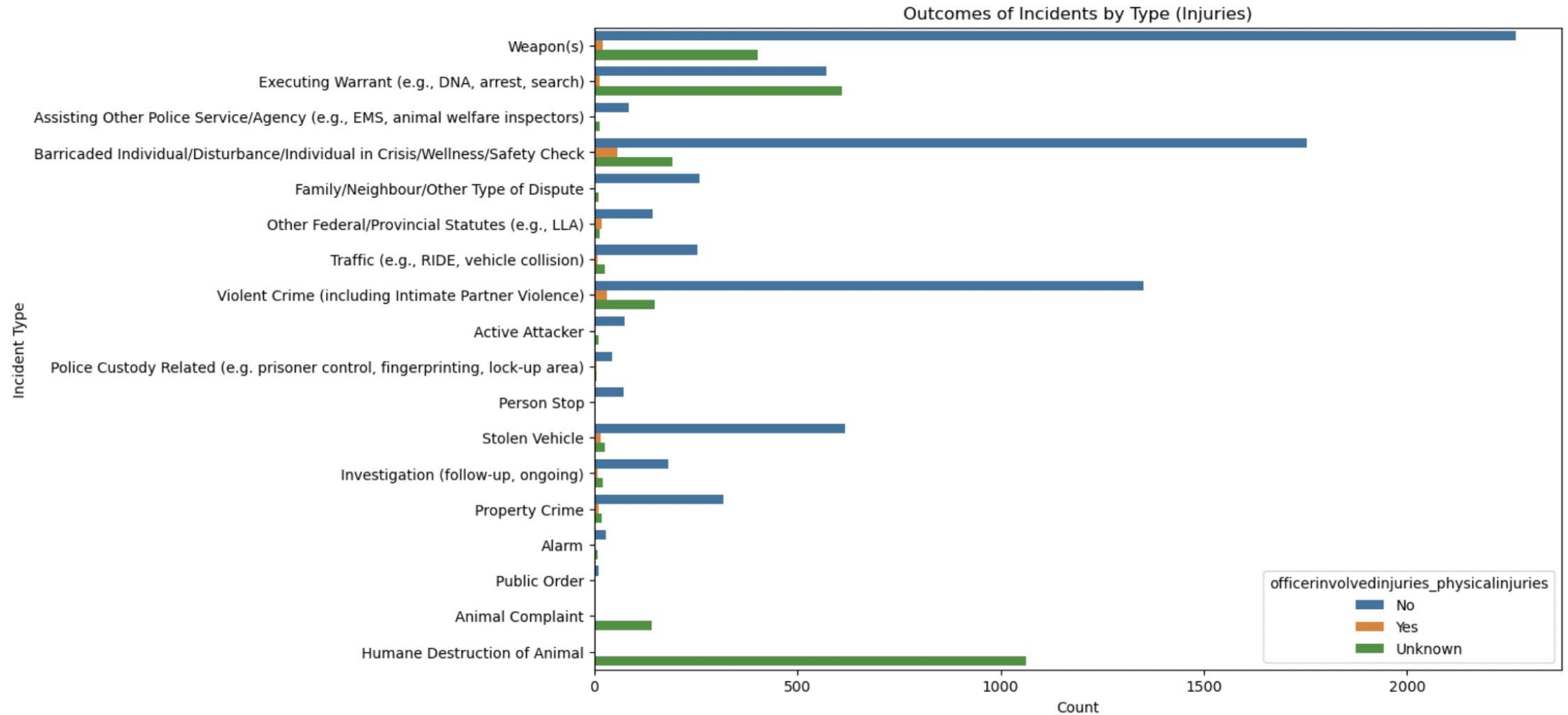
```
#Exploring the datasets  
main_record_df_shape = main_record_df.shape  
individual_record_df_shape = individual_record_df.shape  
  
main_record_df_shape, individual_record_df_shape  
  
((10935, 65), (12805, 112))
```



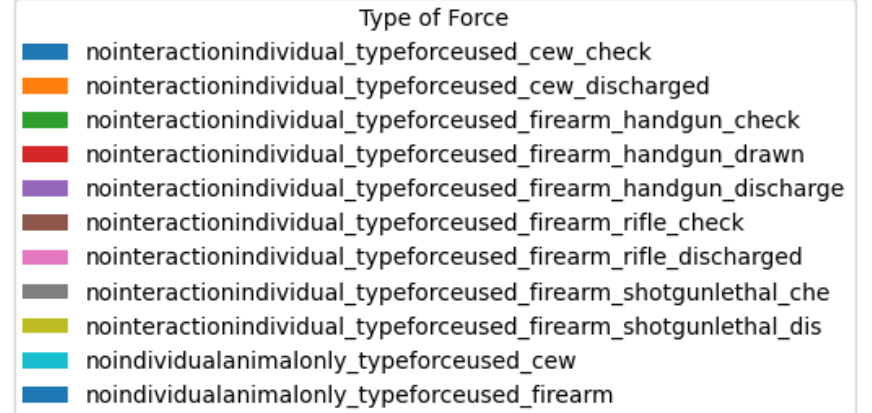
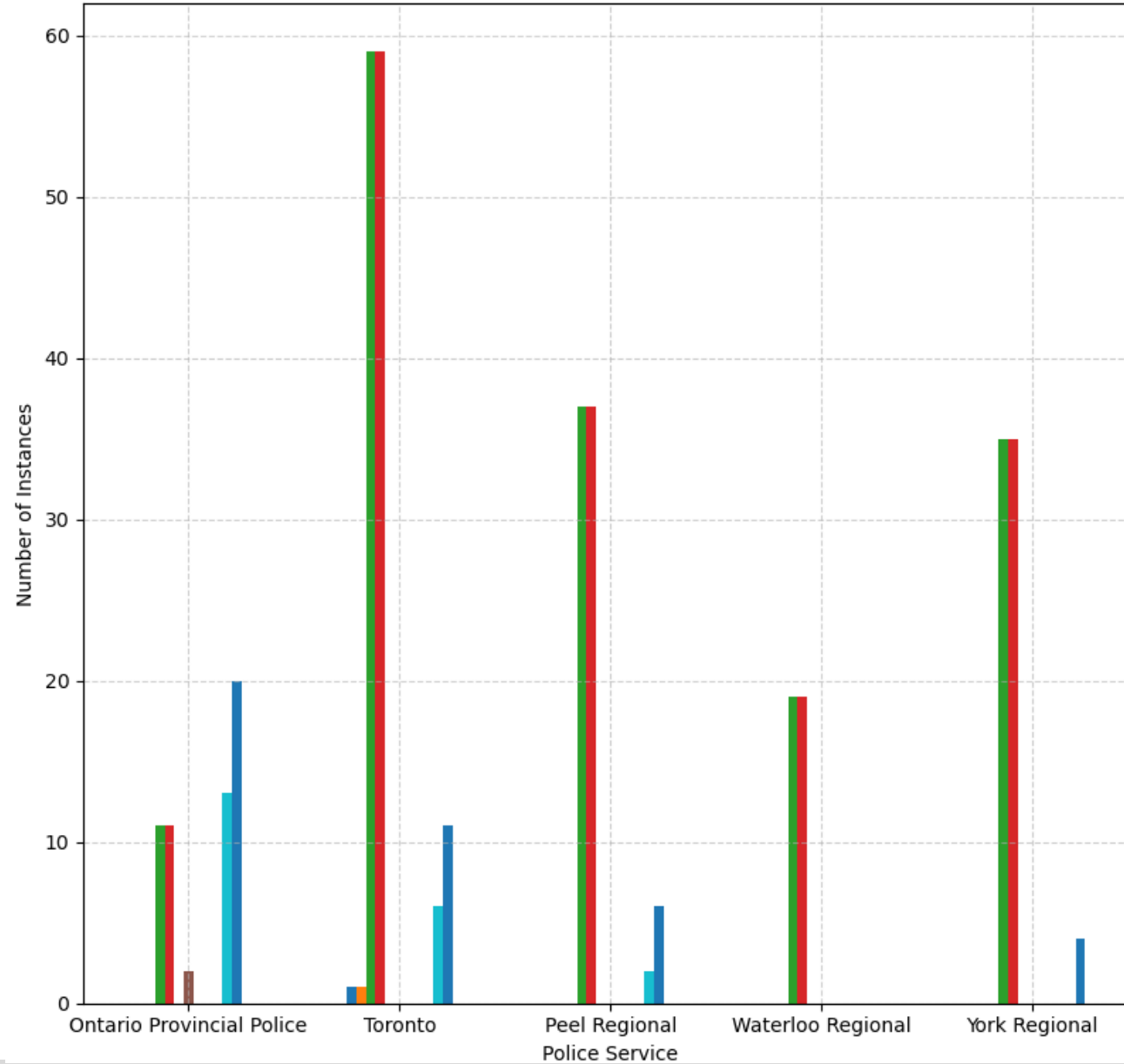


Monthly Trends in Incident Types in 2023





Types of Force Used by Top 5 Police Services in 2023



Data Integration - Cleaning - Processing

Compared to the initial combined dataset, which included 176 columns with mixed data types, the Cleaned data reduce to 42 columns

```
# Check the resulting shape of the cleaned data  
cleaned_data_shape = combined_data_cleaned.shape  
cleaned_data_shape
```

```
(12805, 42)
```

```
combined_data.shape
```

```
(12805, 176)
```

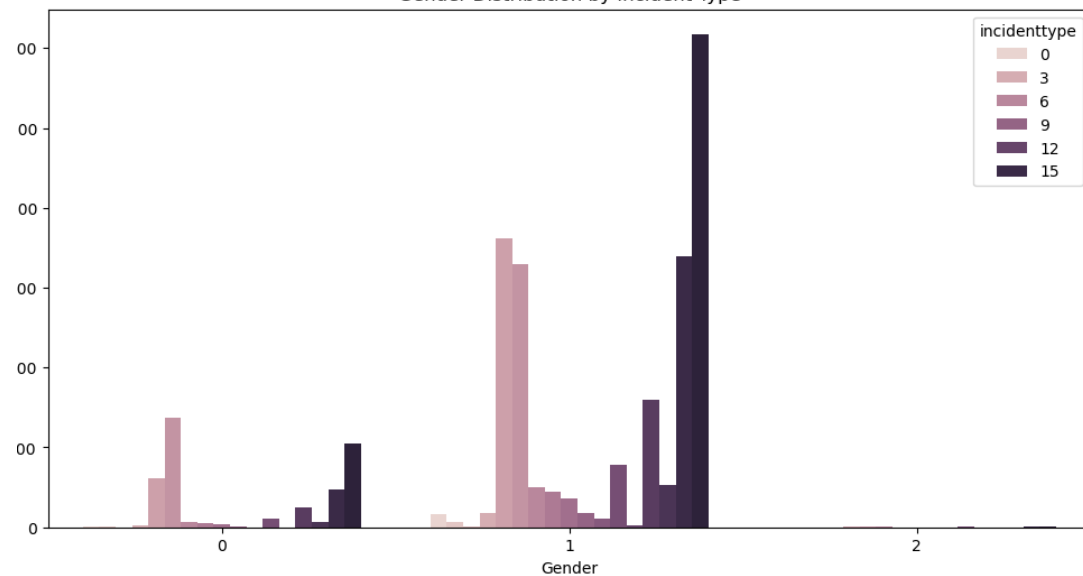
2 datasets use a common identifier, 'BatchFileName,' by integrating incident-level and individual-level information



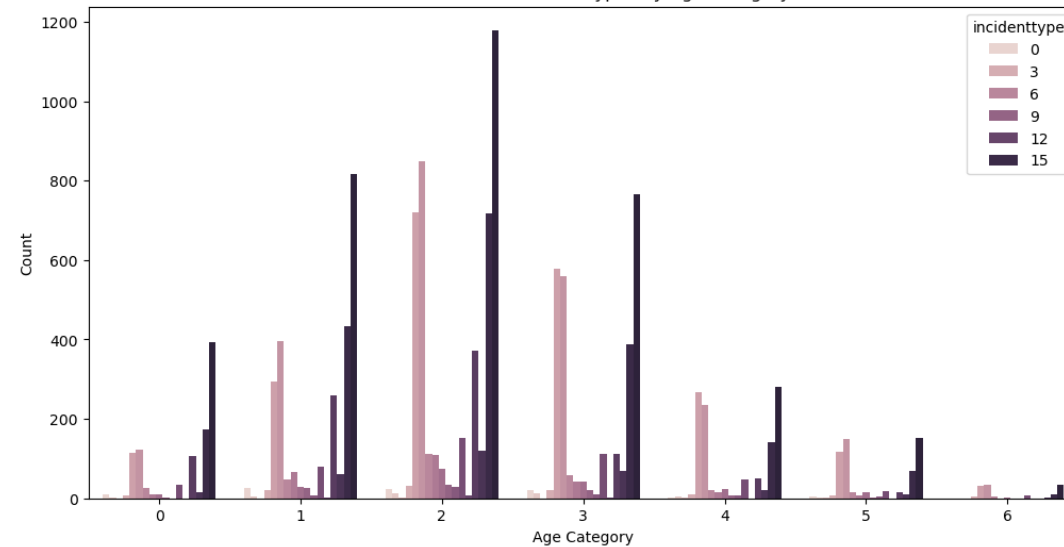
EDA

In depth Explanatory Data
Analysis

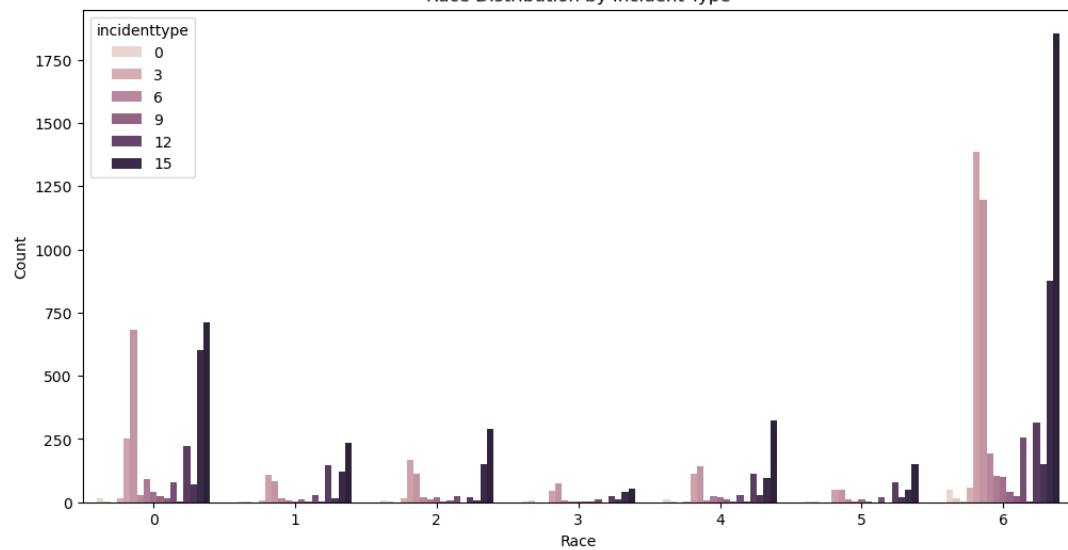
Gender Distribution by Incident Type



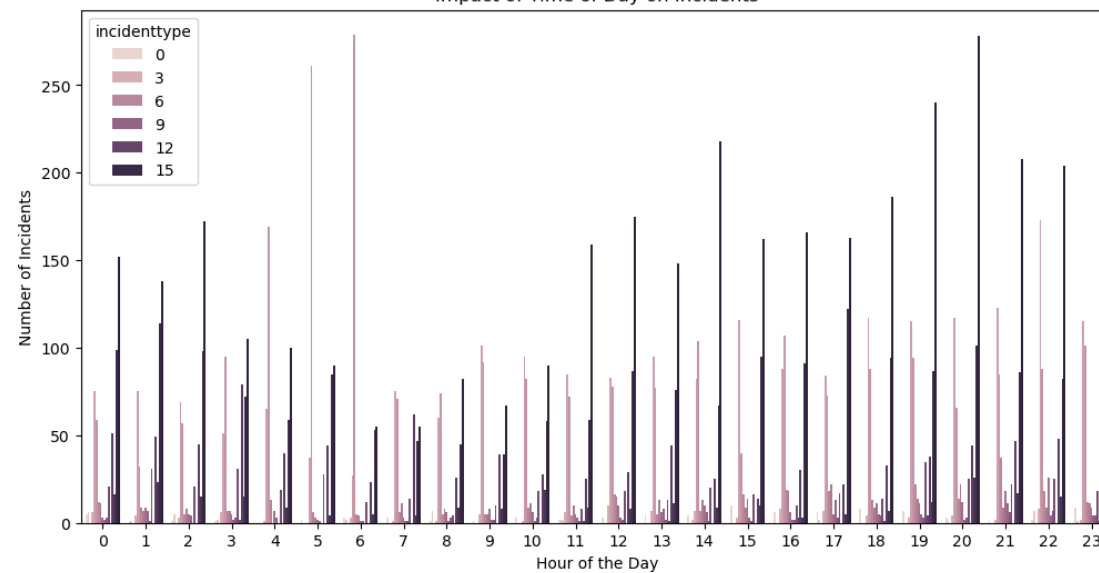
Distribution of Incident Types by Age Category



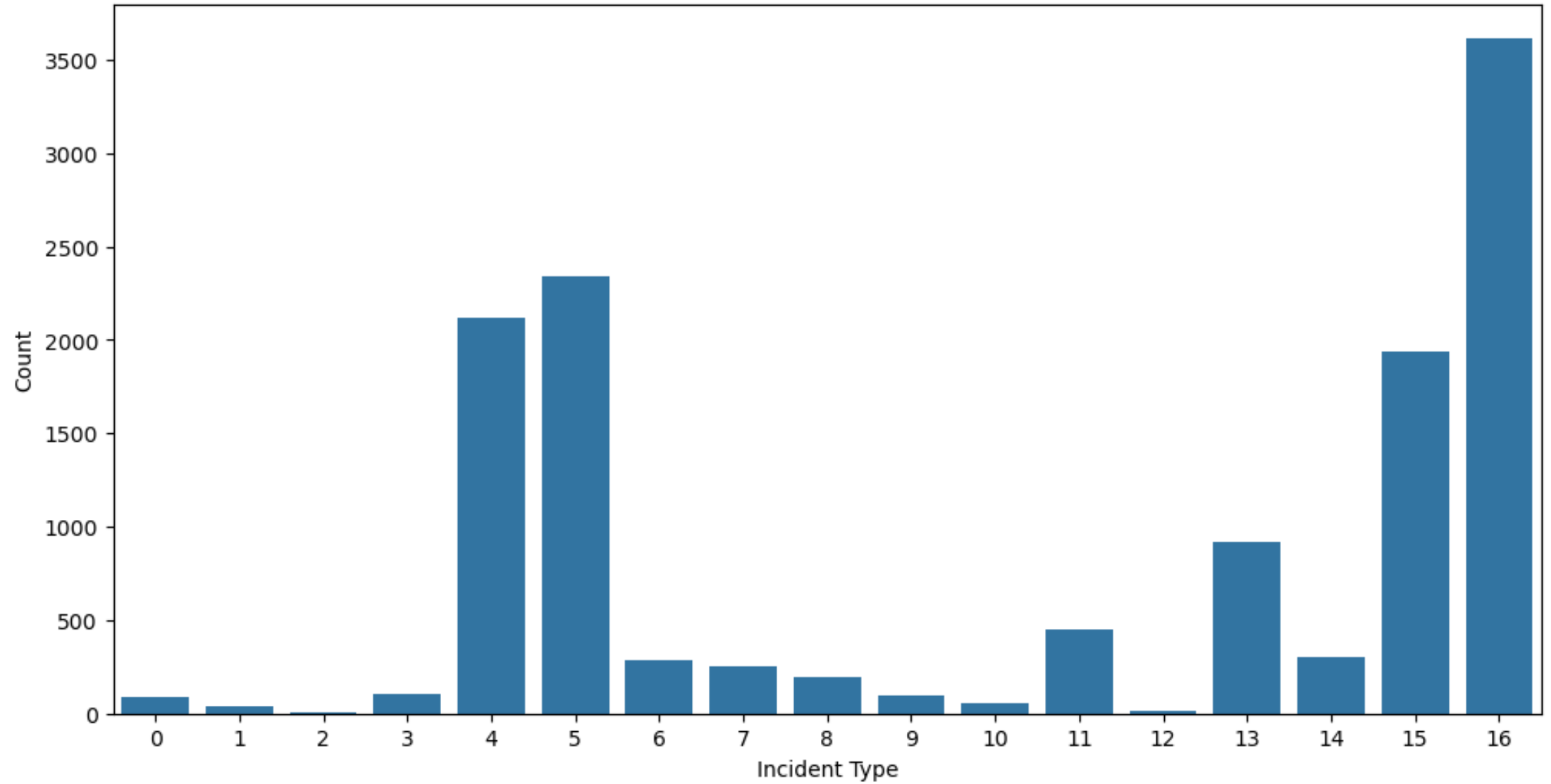
Race Distribution by Incident Type

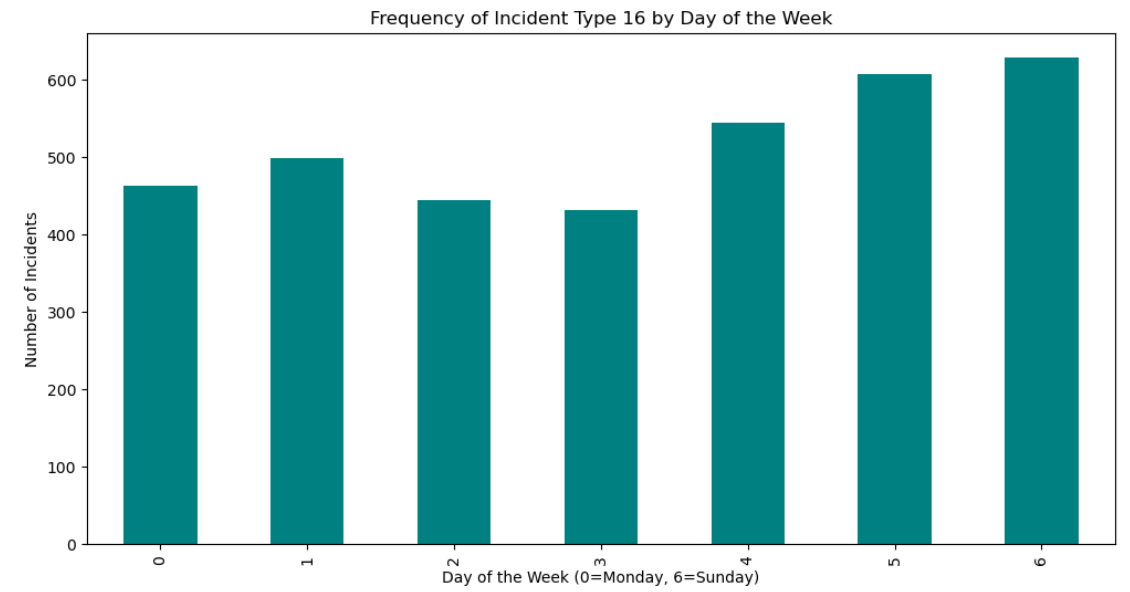
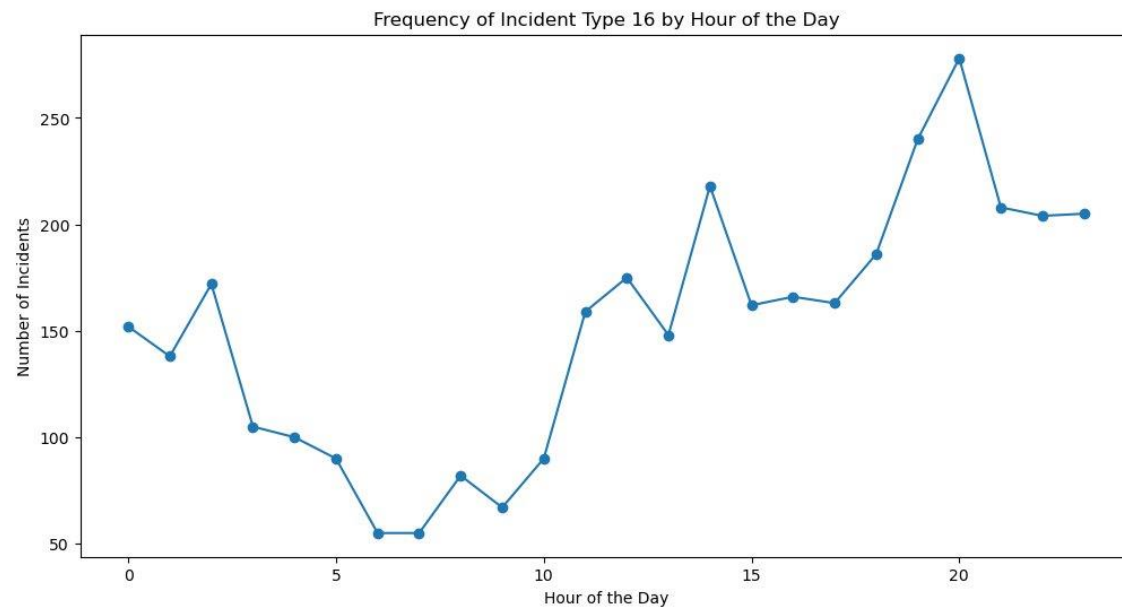


Impact of Time of Day on Incidents

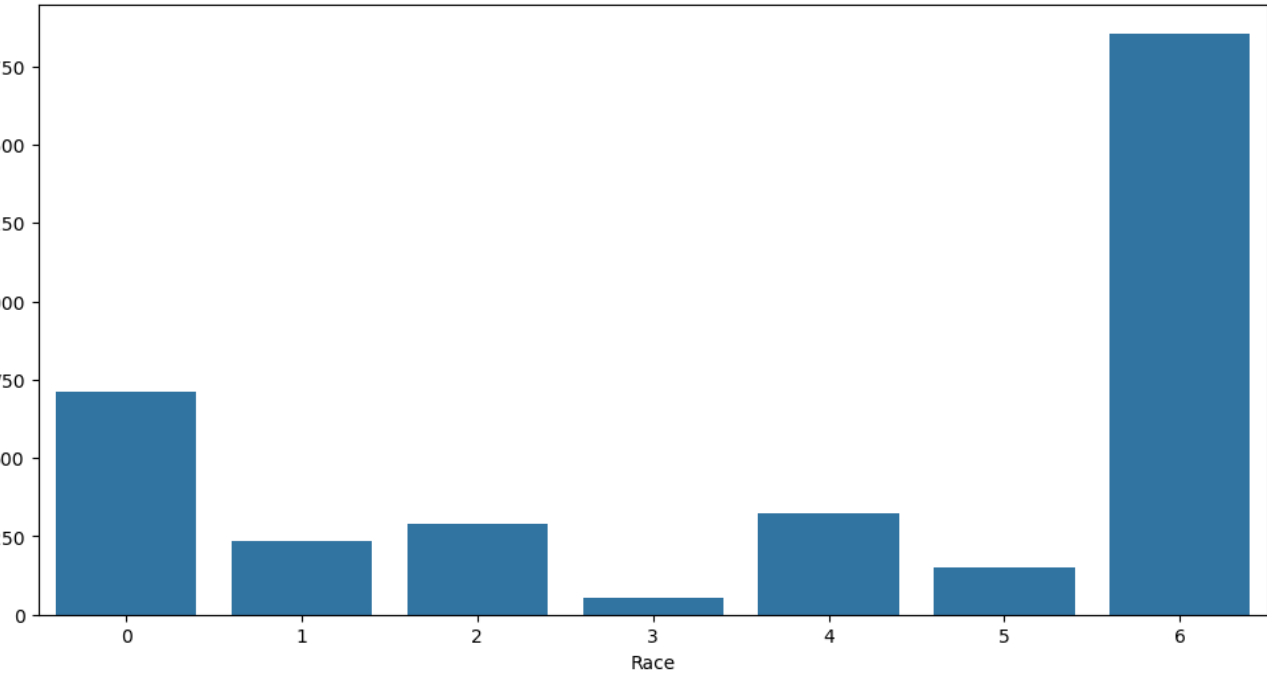


Distribution of Incident Types

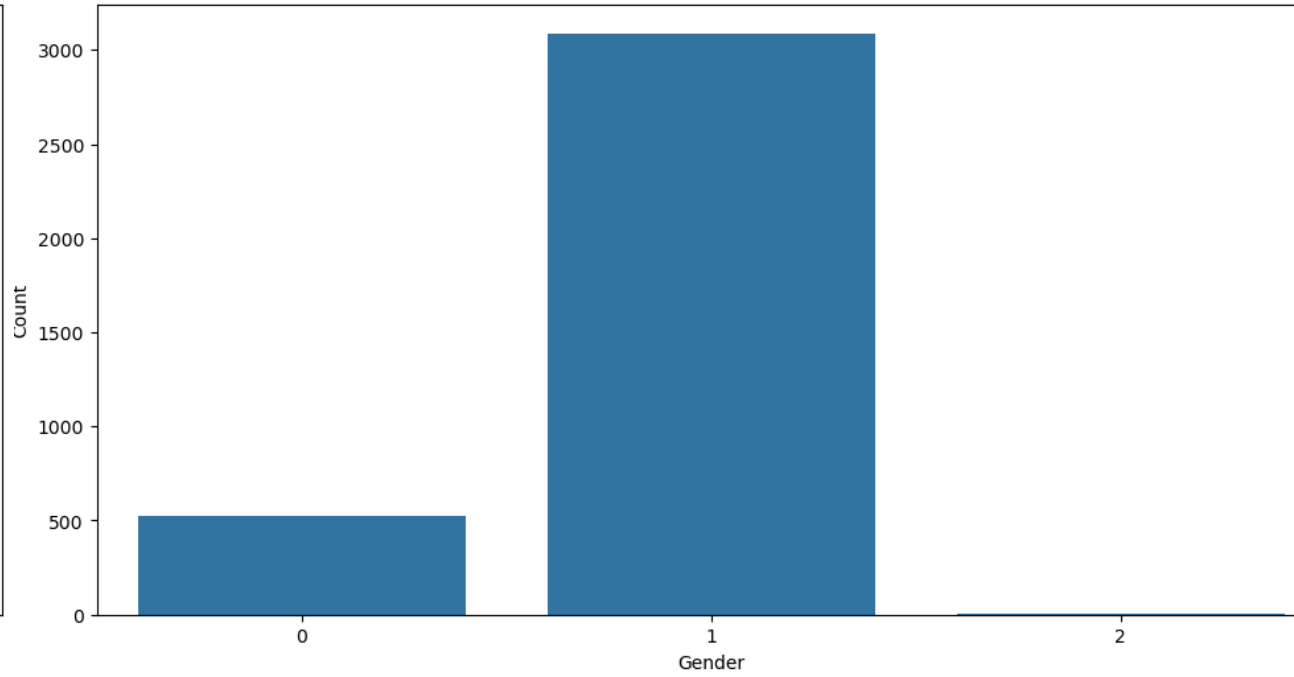




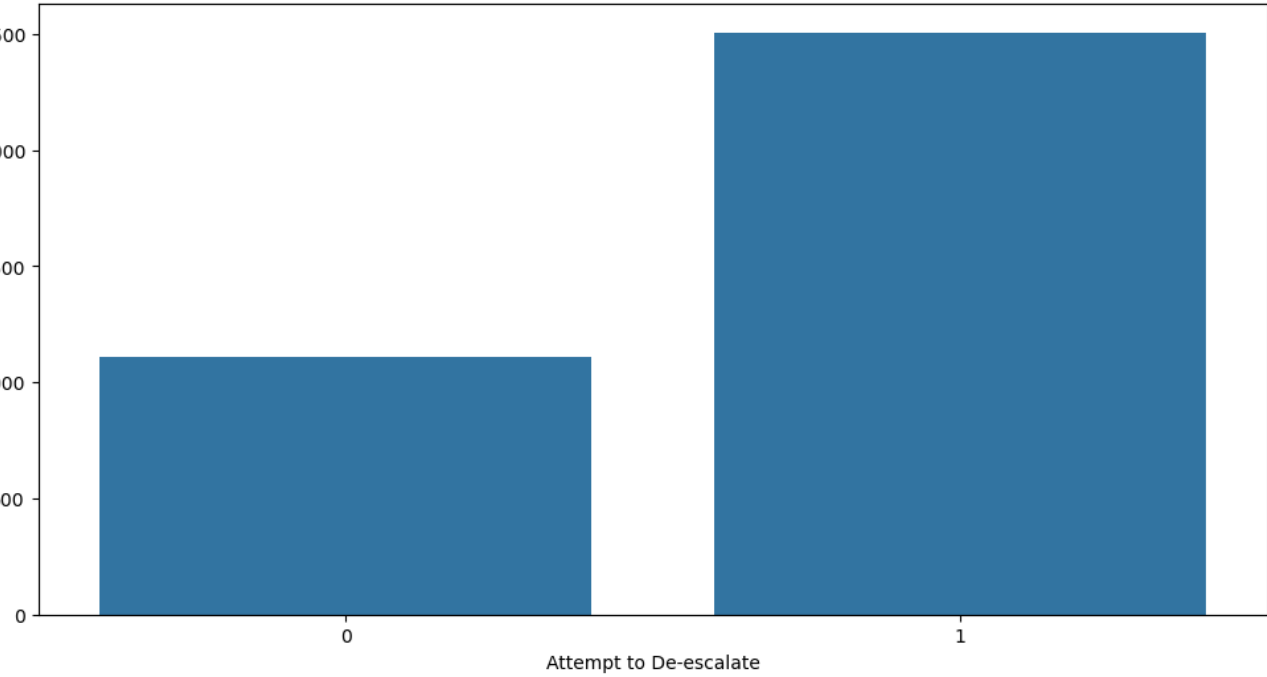
Demographic Breakdown of Incident Type 16 by Race



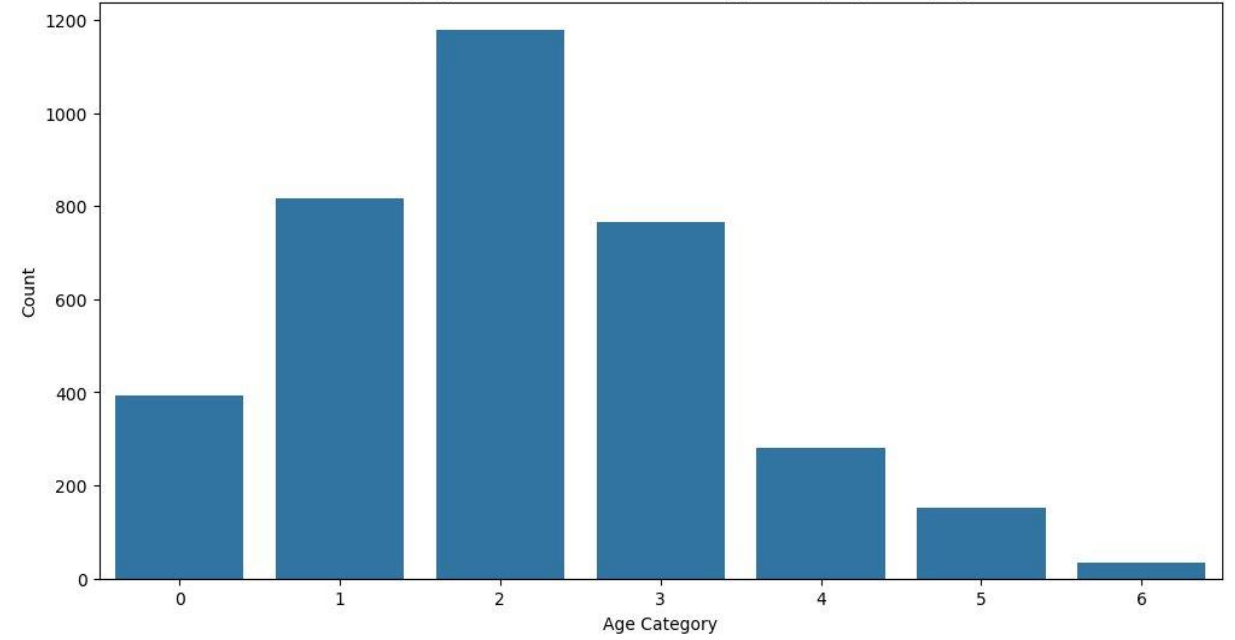
Demographic Breakdown of Incident Type 16 by Gender



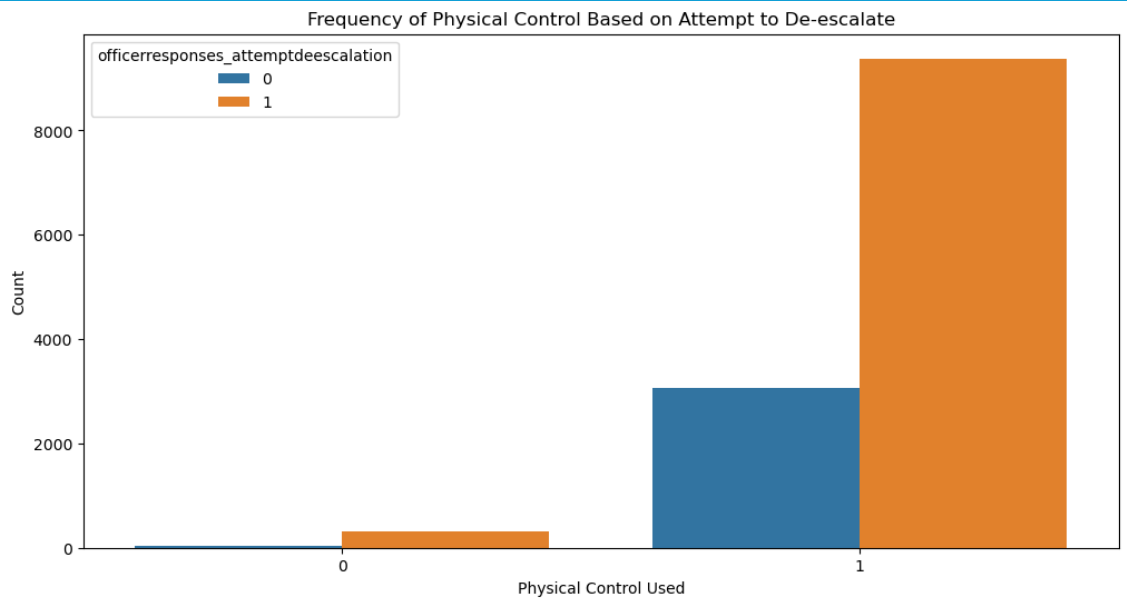
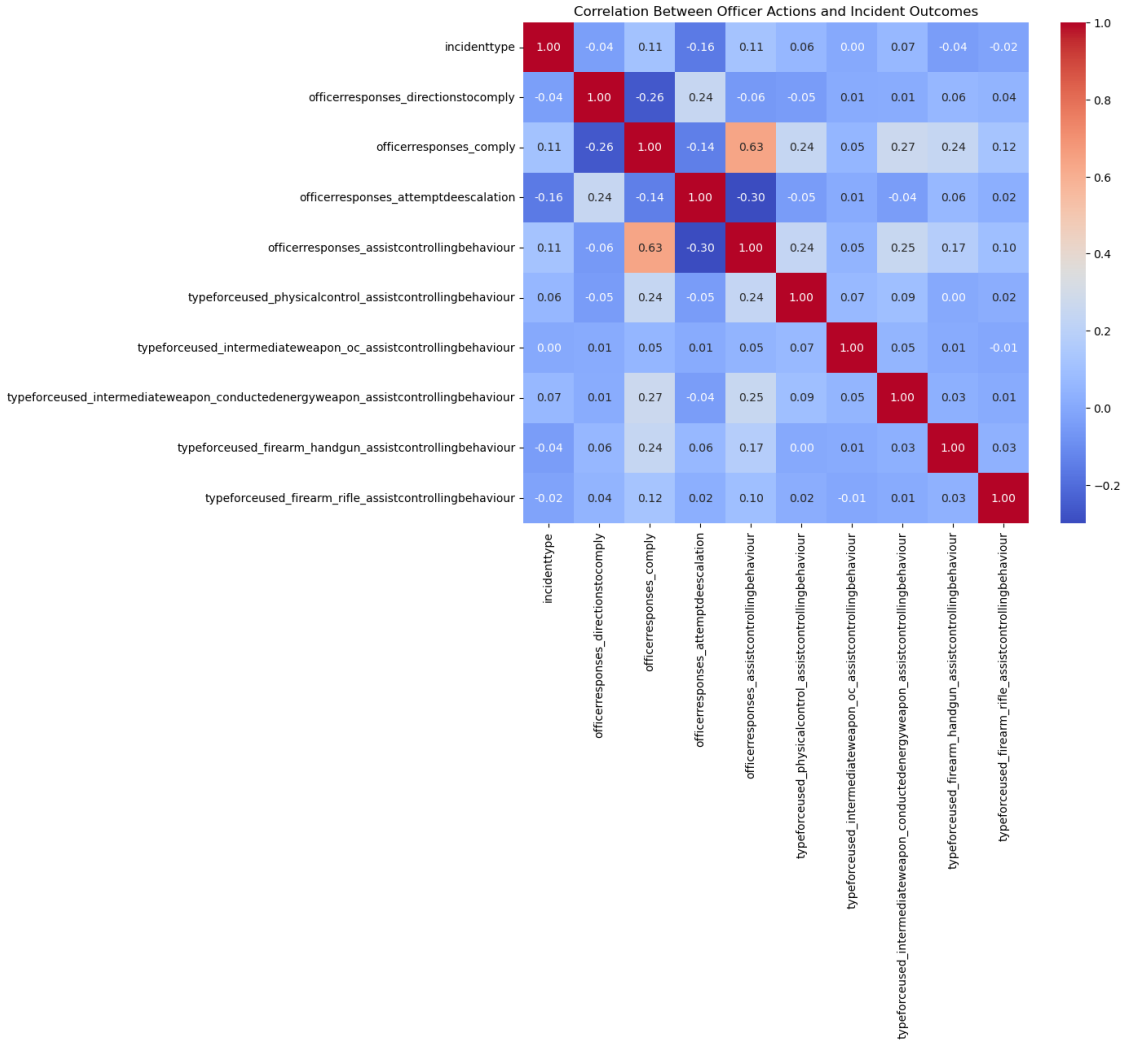
Officer Response: Attempt to De-escalate in Incident Type 16



Demographic Breakdown of Incident Type 16 by Age Category



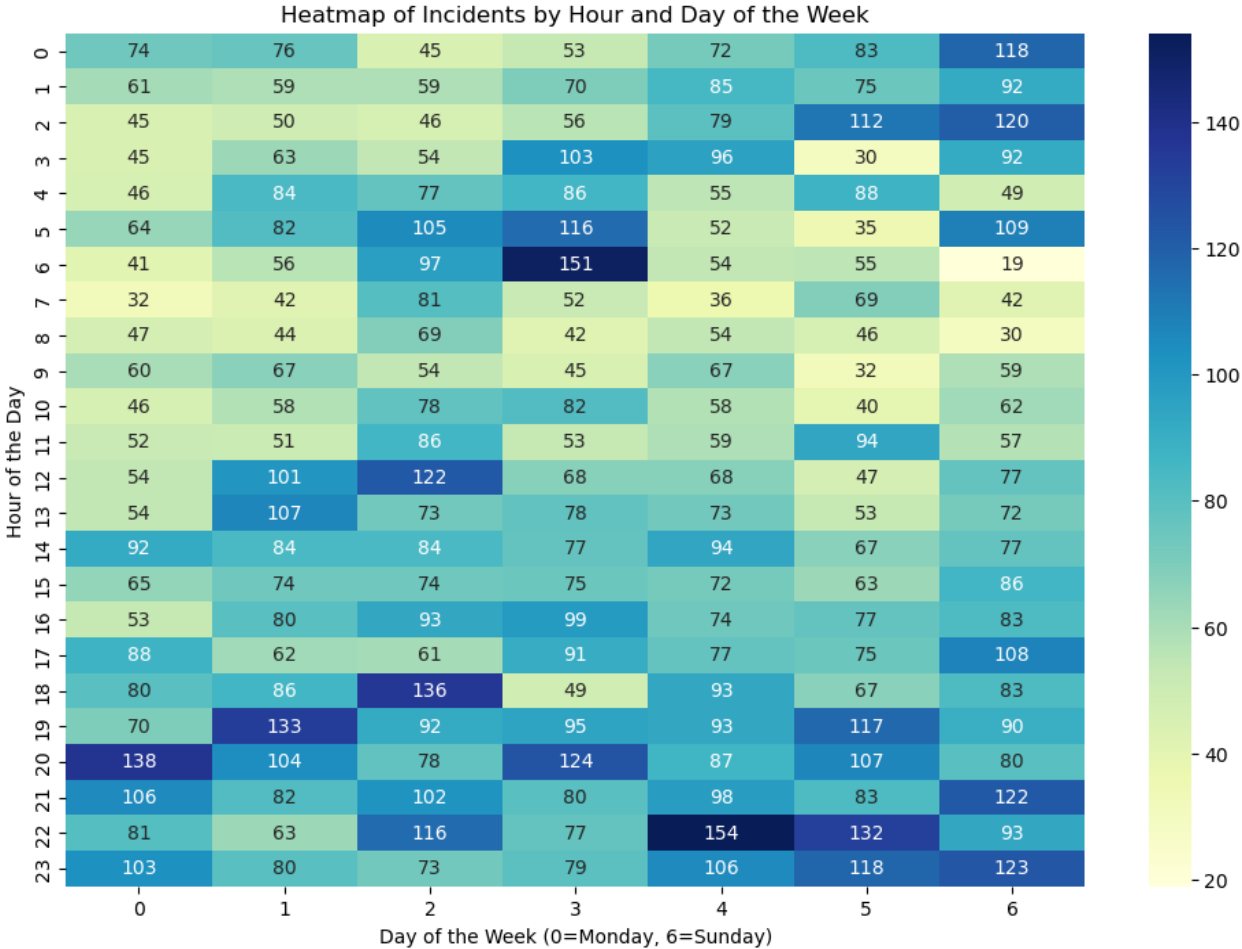
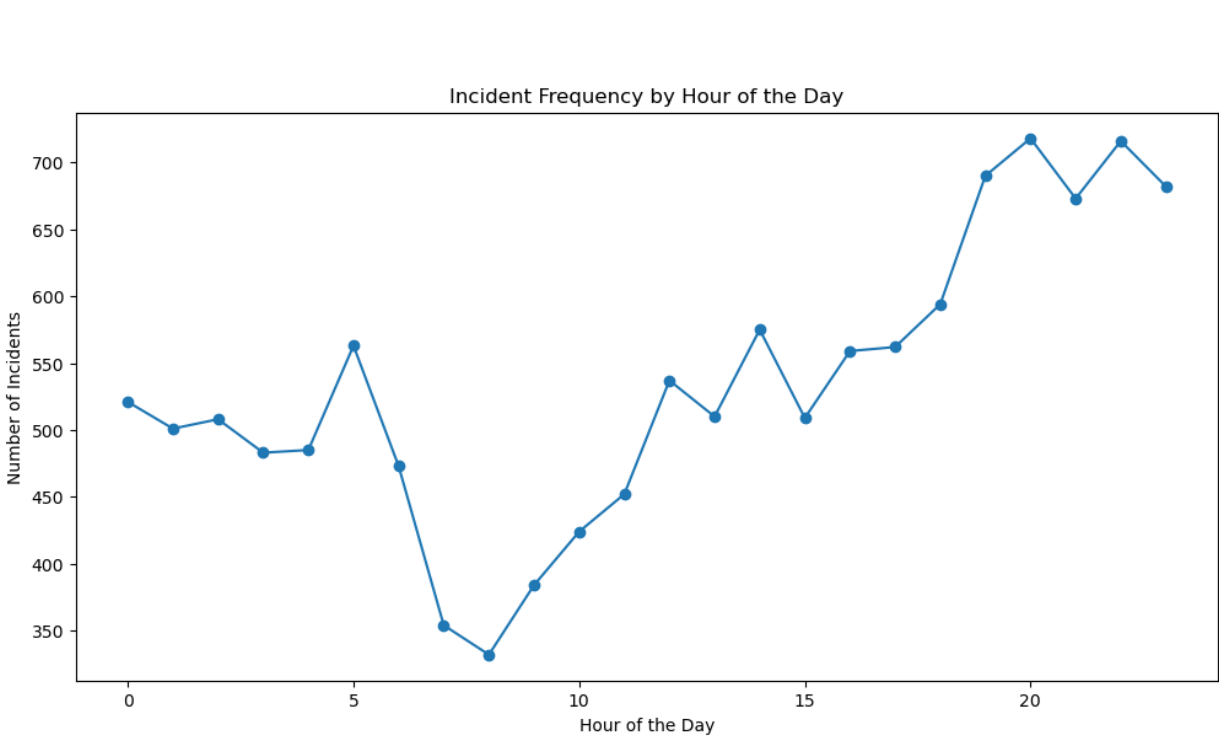
Case Study of Correlation Analysis between Officer Actions and Incident Outcomes



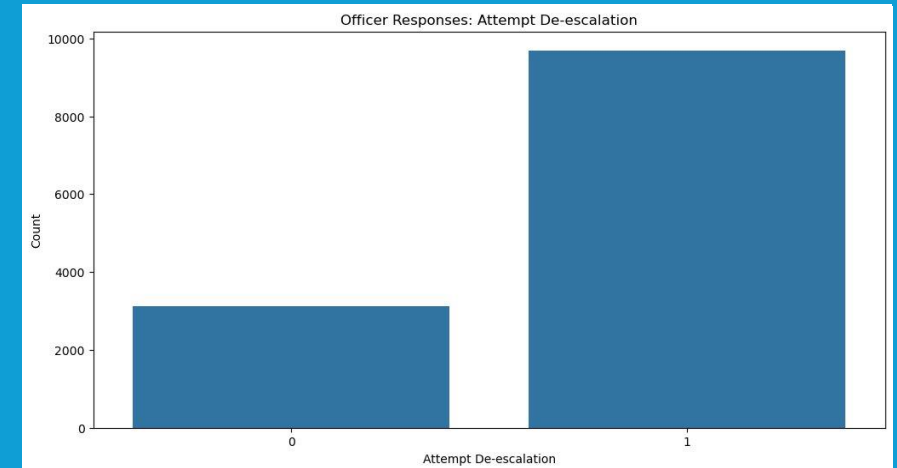
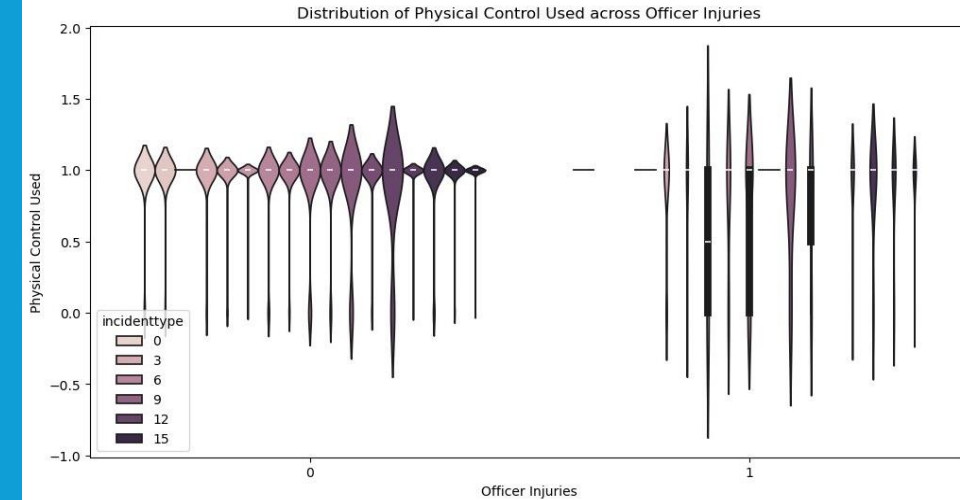
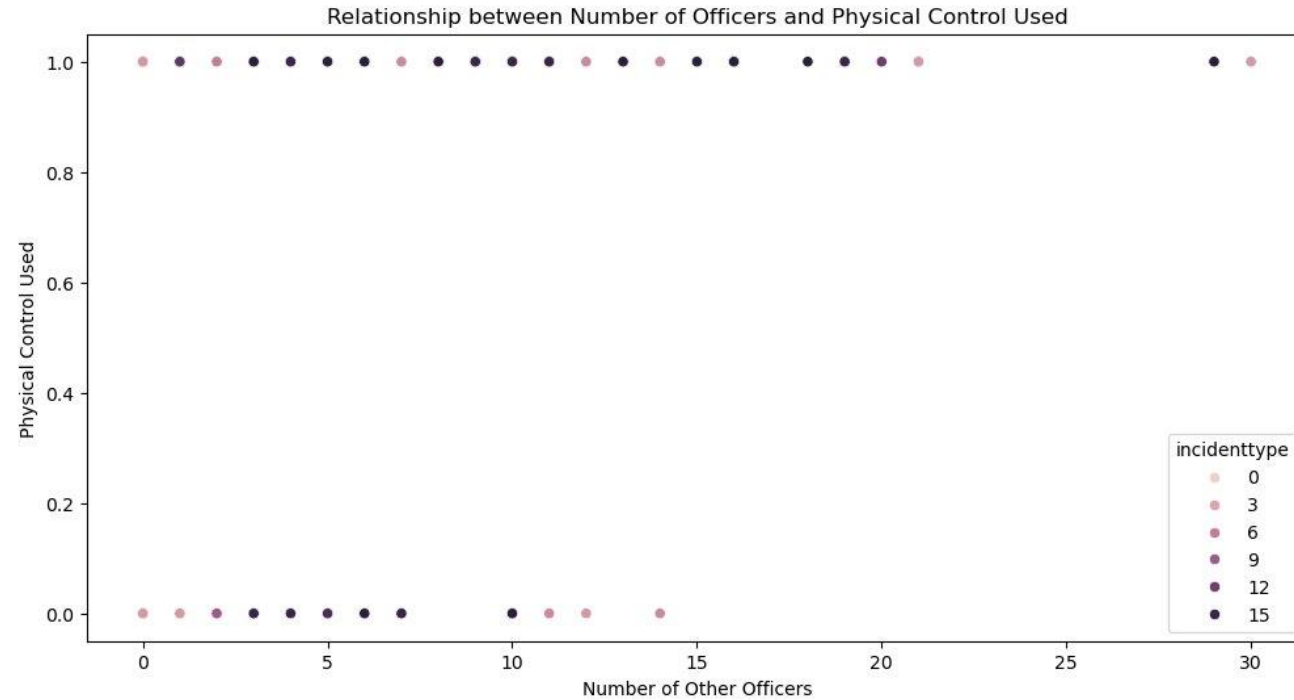
Investigate how different officer responses correlate with the outcome of Incidents

Case study of Time Series Analysis of Incident Types

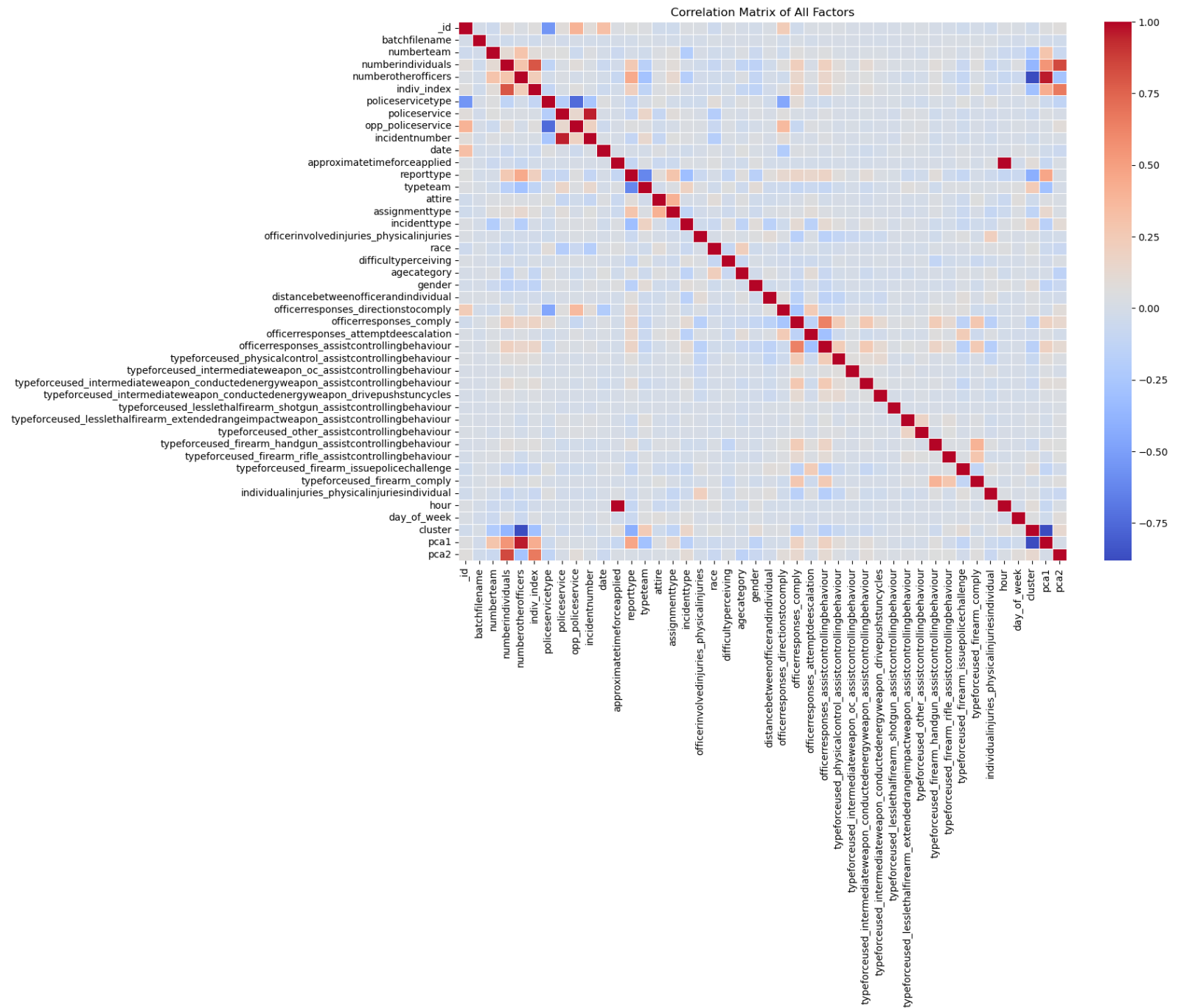
Investigate how the frequency, type of incidents vary over the time (Time of day, weeks and Month)



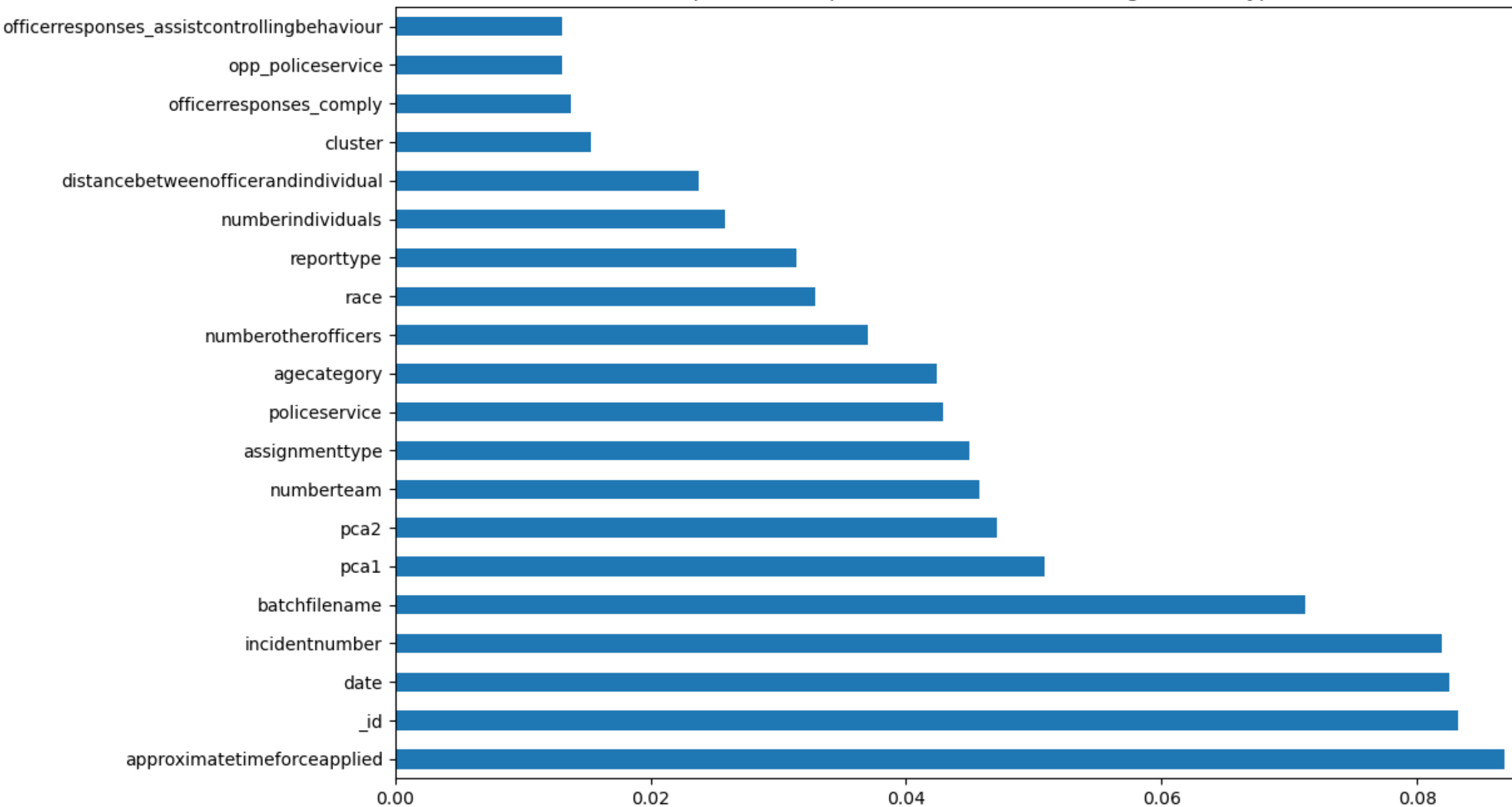
Investigate Severity, Relationship, Impact Likelihood and the Distribution of different levels of Officer Injuries



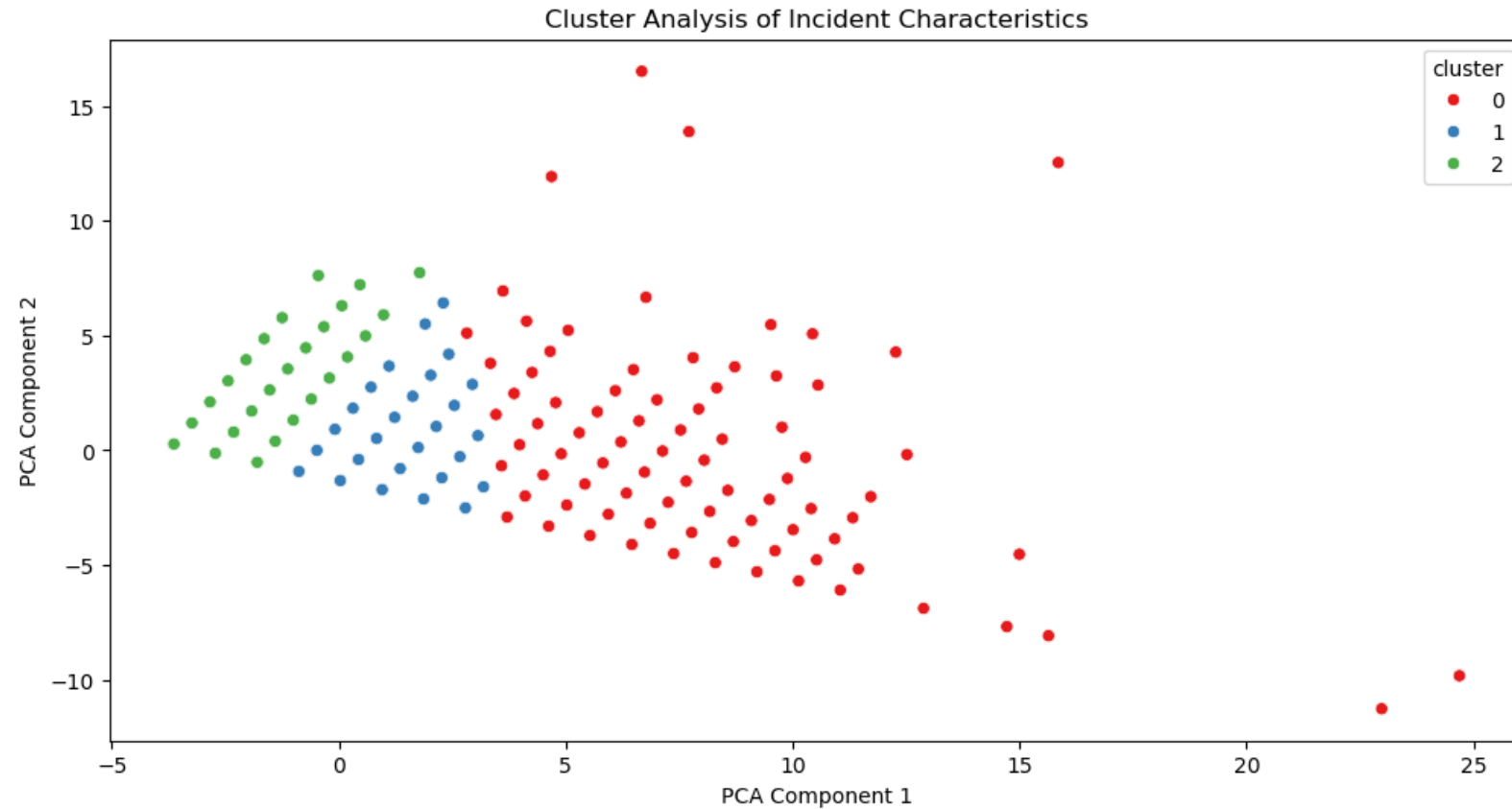
Heatmap Analysis | Feature Engineering with Visualization to Demonstrate Meaningful and Influential Relationship between Variables

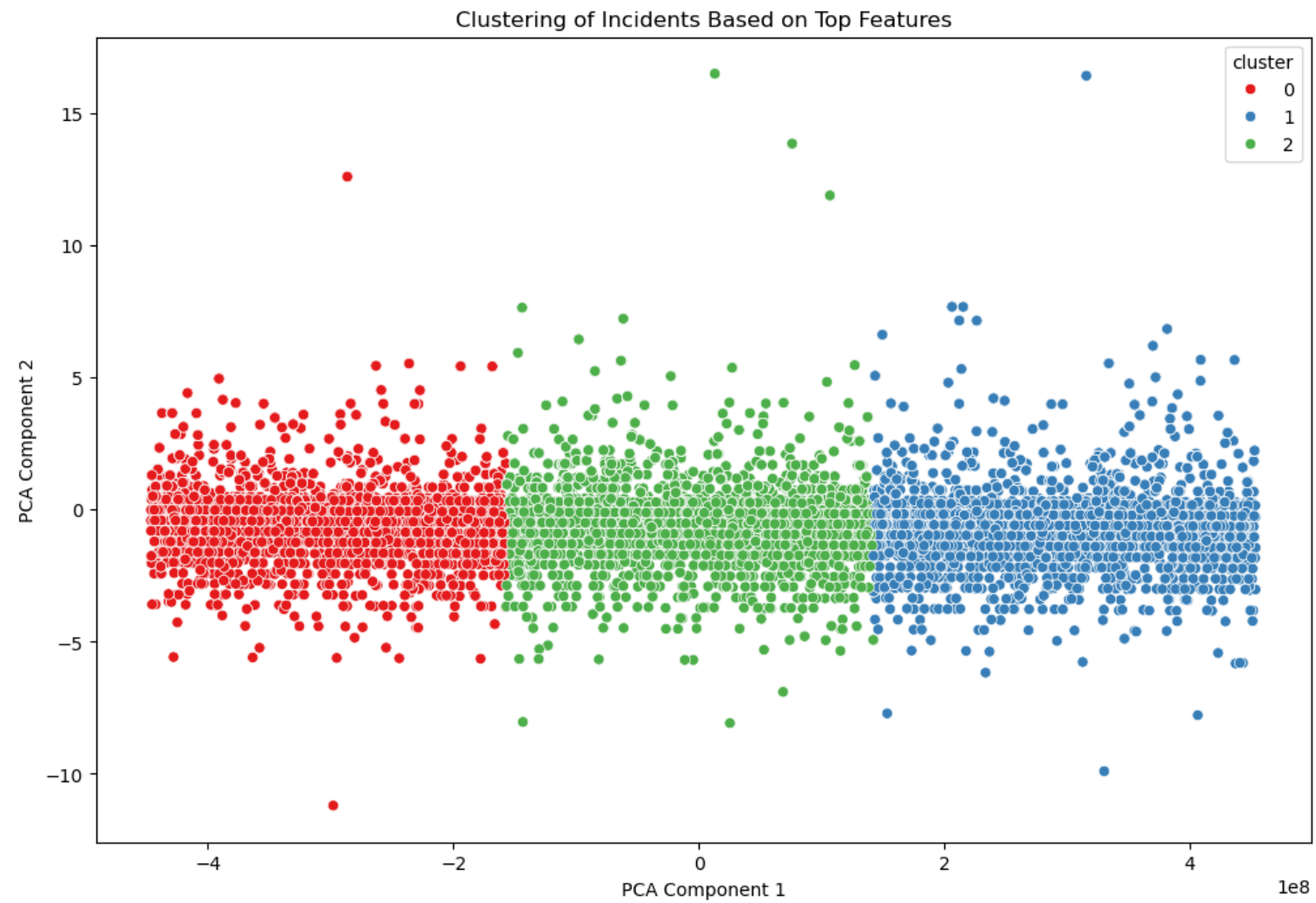


Top 20 Most Important Features in Predicting Incident Type

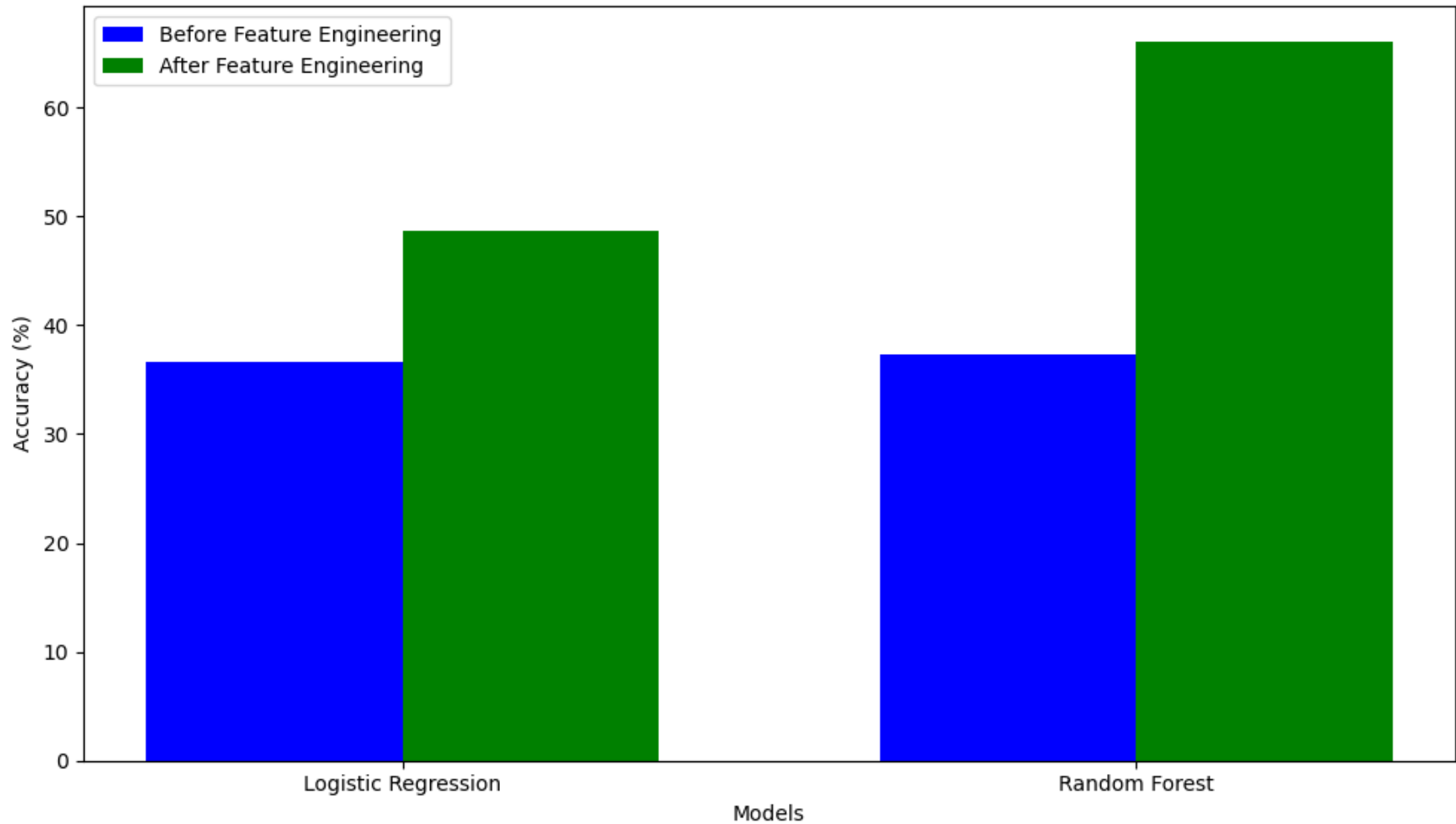


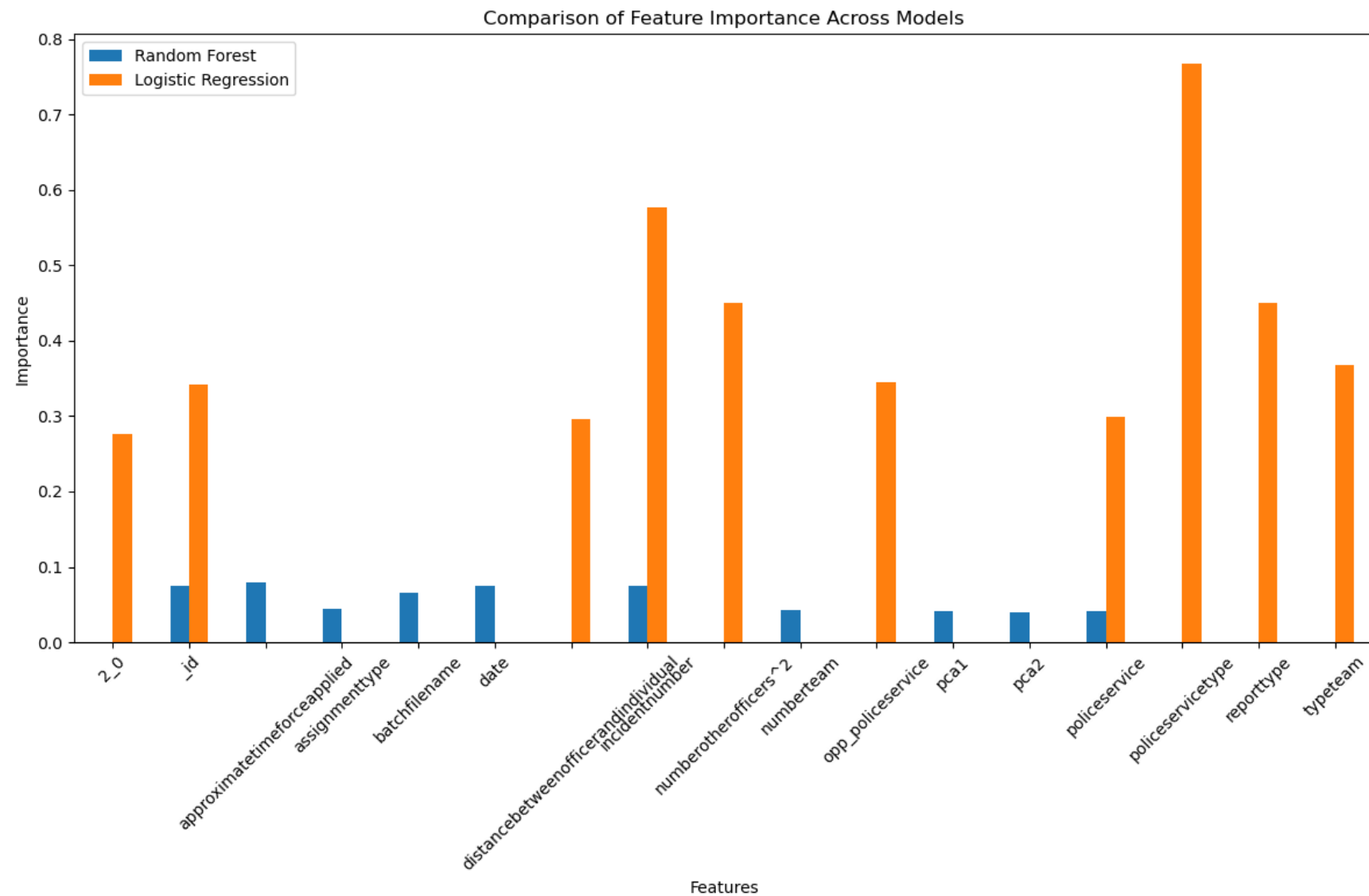
Applying K-means Cluster Analysis of Incident Characteristics and Using Principal Component Analysis (PCA)

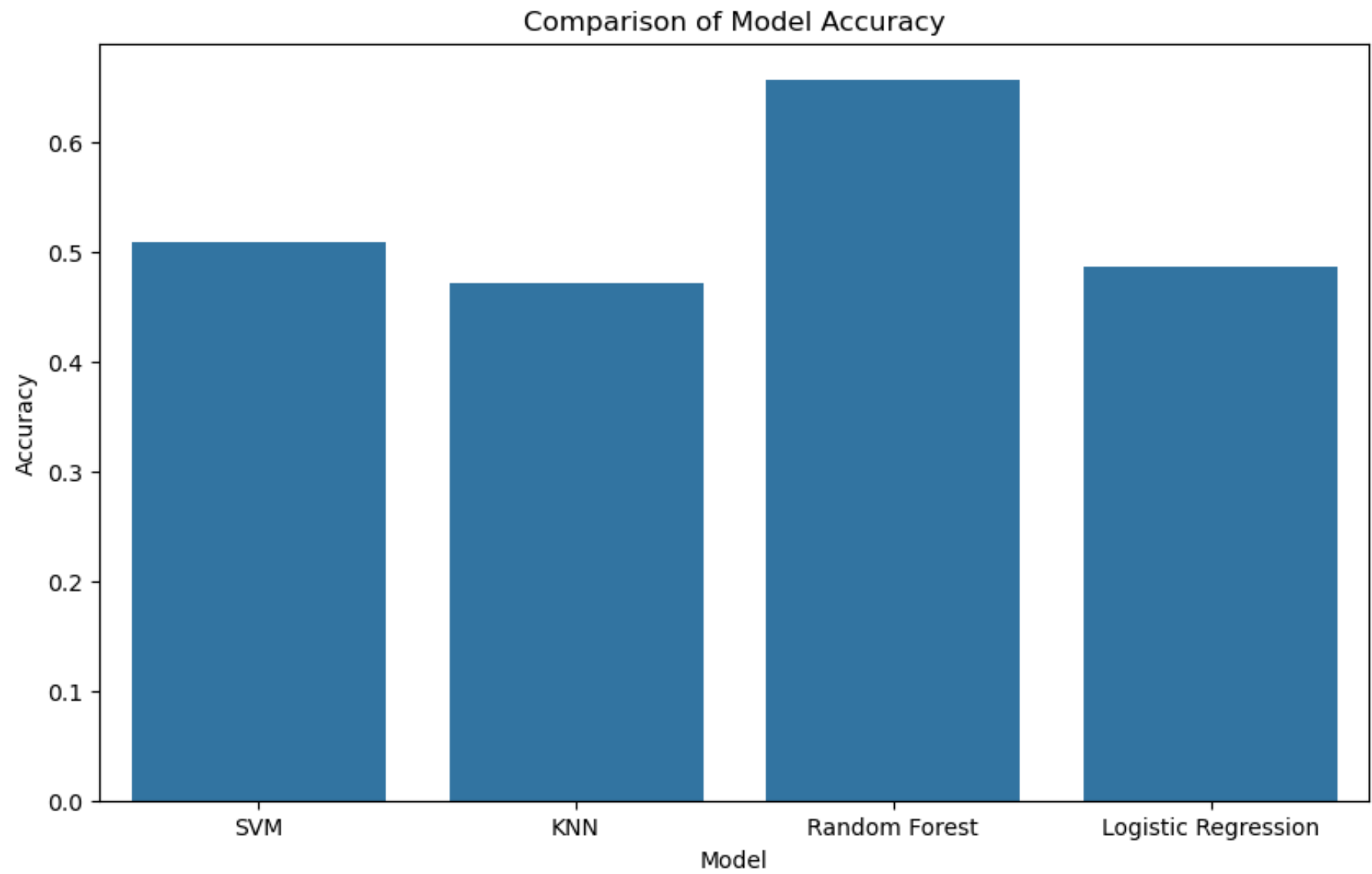




Model Accuracy Before and After Feature Engineering







Strategic Recommendation

- Improving the quality and completeness of data collection processes will further enhance the reliability of future analysis
- Monitor model performance, Validate Model (Check with VIF) for potential multi-collinearity and robust to overfitting while in Production
- Regular reporting to senior management will ensure that the insights from this analysis are consistently applied to decision-making processes
- Implementing a real-time dashboard continuously monitors key metrics related to police use of force will allow for timely interventions and Adjustments to policies

