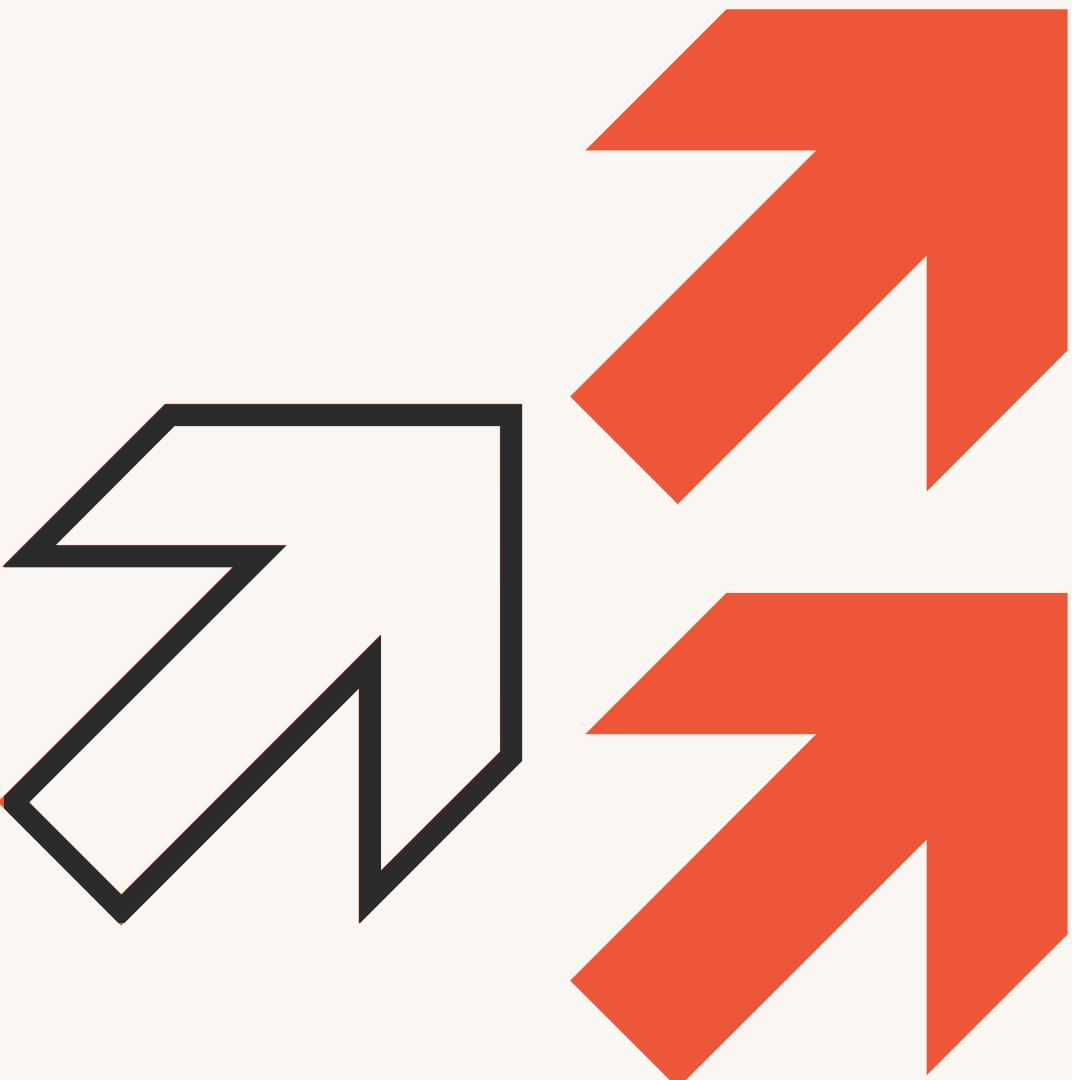


INSY 662 - FINAL REPORT

POLITICAL CRIME PATTERNS IN MEXICO



TABLE OF CONTENT



Executive Summary

01

Introduction

- Background
- Project Overview
- Potential Stakeholders

02

Dataset Exploration

- Dataset Description
- Key Variables
- Pre-processing steps
- Exploratory Analysis

03

Model Building

- Modelling Steps
- Scaling
- Feature Selection
- Tuning

04

Results & Conclusions

- Results Iteration
- Sample Prediction
- Final Takeaways
- Societal Impact
- Future Directions

BACKGROUND

Between 2018 and October 1, 2024, **2,133 incidents of threats, murders, armed attacks, disappearances, and kidnappings against political figures or facilities** have been recorded during election periods in Mexico.



Organizations such as Data Cívica, México Evalúa, and Animal Político aim to gain a deeper understanding of **political-criminal violence**.



“crime one directed against a particular government or political system”

SECTION 01

PROJECT OBJECTIVE

INSY 662 - FINAL PROJECT REPORT



PROJECT OVERVIEW

THE MAIN QUESTION



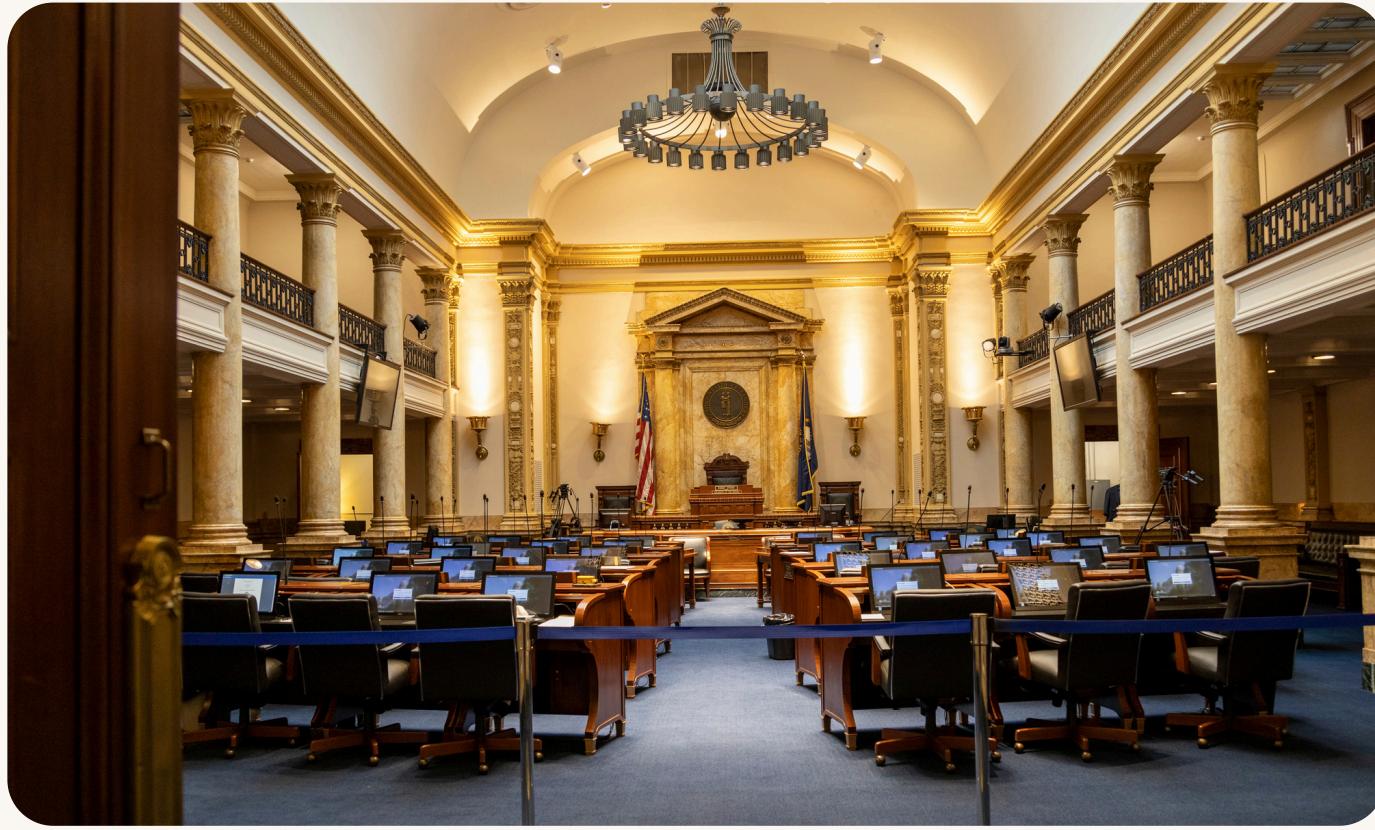
How can past incidents of political violence and macroeconomic factors help predict future political crimes in Mexico?

THE OBJECTIVE

Predict the number and type of political crimes in Mexico by analyzing past violence against politicians and key macroeconomic factors.

THE IMPACT

- Identify patterns to forecast political crime trends.
- Develop a predictive model to reduce risks for political figures.
- Provide insights into socioeconomic factors contributing to political violence.



POTENTIAL STAKEHOLDERS

- Government and Public Officials
 - Law Enforcement Agencies
 - Policy Makers
- Political parties (Leaders, candidates, & campaign managers)
- Judicial Bodies (Prosecutors)
- Non-Governmental Organizations
- Journalists and Media Outlets
- General Public (citizens and community groups)

EXECUTIVE SUMMARY



What indicators are most effective in predicting political crime in Mexico? Our data was sourced from an organization called “Voting between Bullets” which aims to make voting fair without the influence of violence. We found it difficult to predict political violence due to the variability and the continuously evolving political landscape, however, it is core to democracy and the proper execution of the vote. With our research, we can best protect political figures and assist the proper, just, and peaceful undertaking of elections. The key insights we identified are as follows:

#1

MODEL

- Merge of Crime and economic data.
- KNN classification

#2

MAJOR POLITICAL PARTIES
AT RISK

- National Action Party (PAN)
- Institutional Revolutionary Party (PRI)

#3

GENDER

- **Both** as threatened

SECTION 02

DATA EXPLORATION

INSY 662 - FINAL PROJECT REPORT



DATA DESCRIPTION

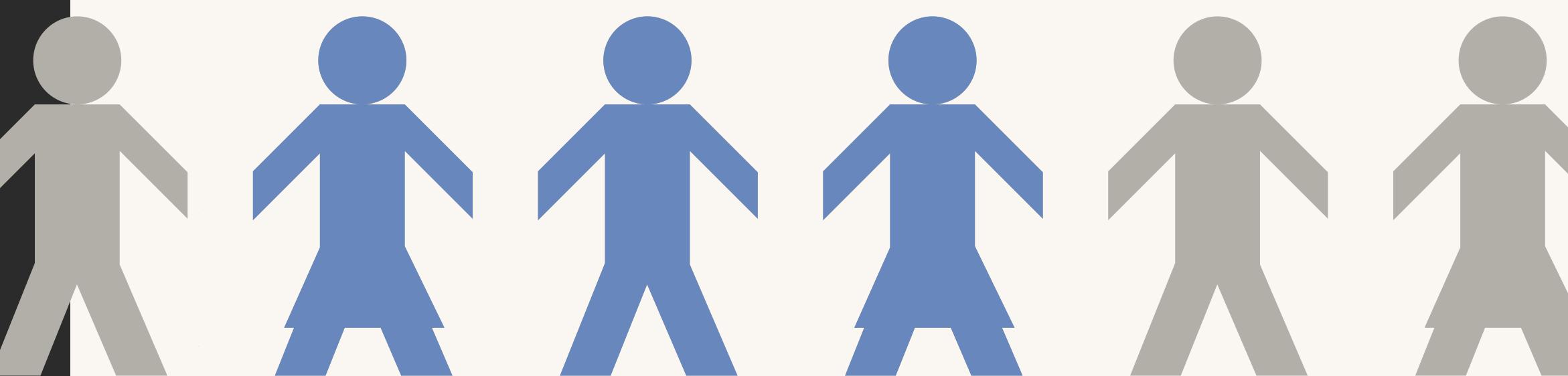


We utilized three distinct datasets to explore the relationship between political violence, demographic and economic factors:

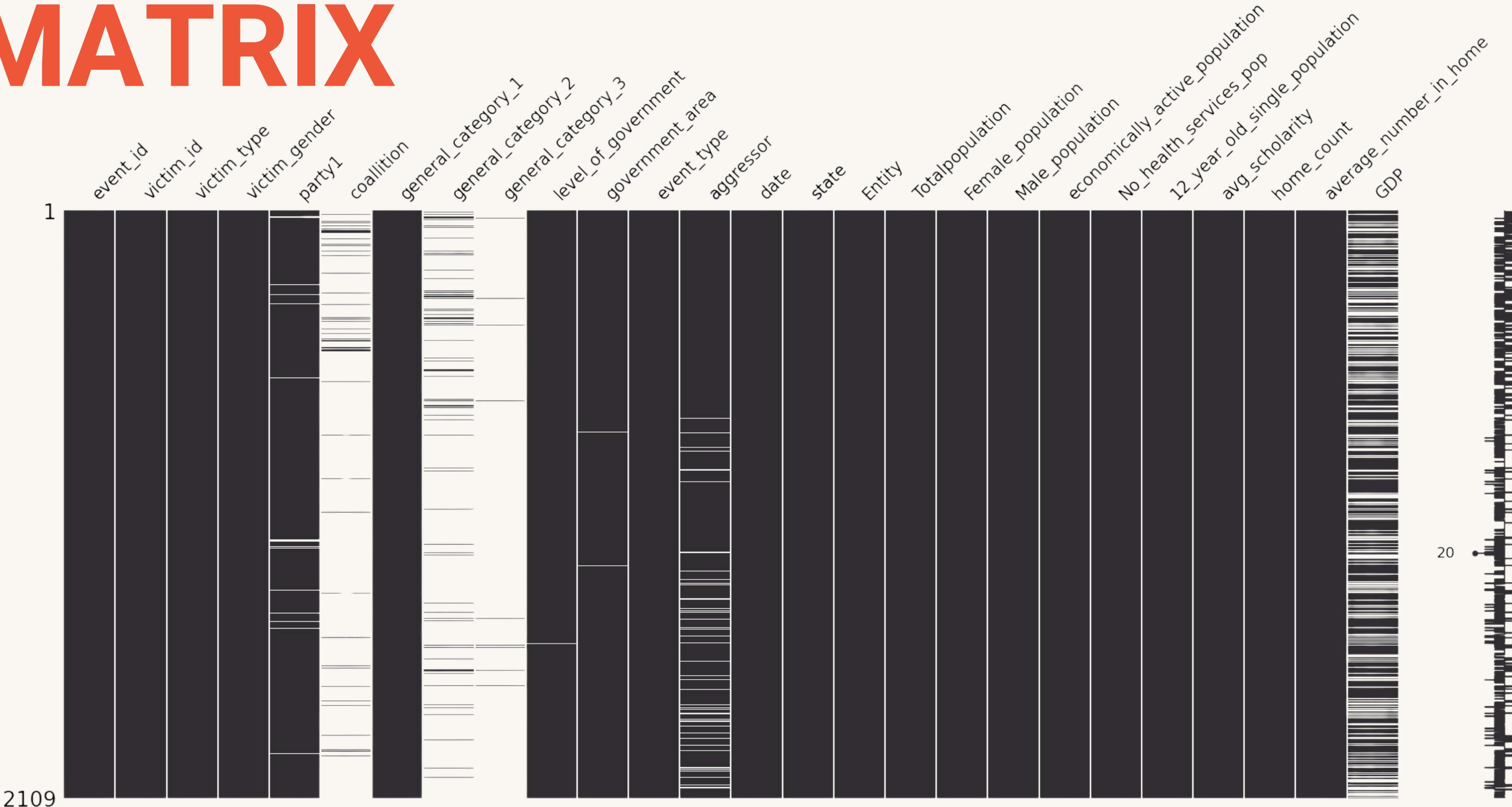
- **Political Violence:** Sourced from Datacivica, a non-profit data organization, this dataset includes information on various political crimes in Mexico between 2017 and 2024.
- **Economic Data:** Gross Domestic Product (GDP) information collected by Mexico's Central Bank (Banxico).
- **Demographic Data:** Provided by Mexico's National Statistical Office, this dataset offers granular demographic details across different country regions like scholarly, gender, age, etc.

KEY VARIABLES

General Crime Data	Political Party Data	Demographic Data
Victim Name	Political Party	GDP
Victim Gender	General Category	Female Population
Event Type	Government Area	Health Service Availability
Victim Type	State	12 Year Old Proportion
		Average Scholarly
		Home Count
		Average Number In Home



PRE-PROCESSING MATRIX



PREPROCESSING STEPS

- 1 **STANDARDIZE FORMATTING**
- 2 **DATA CLEANSING**
- 3 **HANDLE MISSING DATA**
- 4 **DETECT OUTLIERS & NORMALIZE DATA**
- 5 **GAIN PRELIMINARY INSIGHTS**

Rename columns and convert variables to consistent formats (e.g., English labels, numeric scales).

Remove duplicates and irrelevant identifiers that don't contribute to analysis.

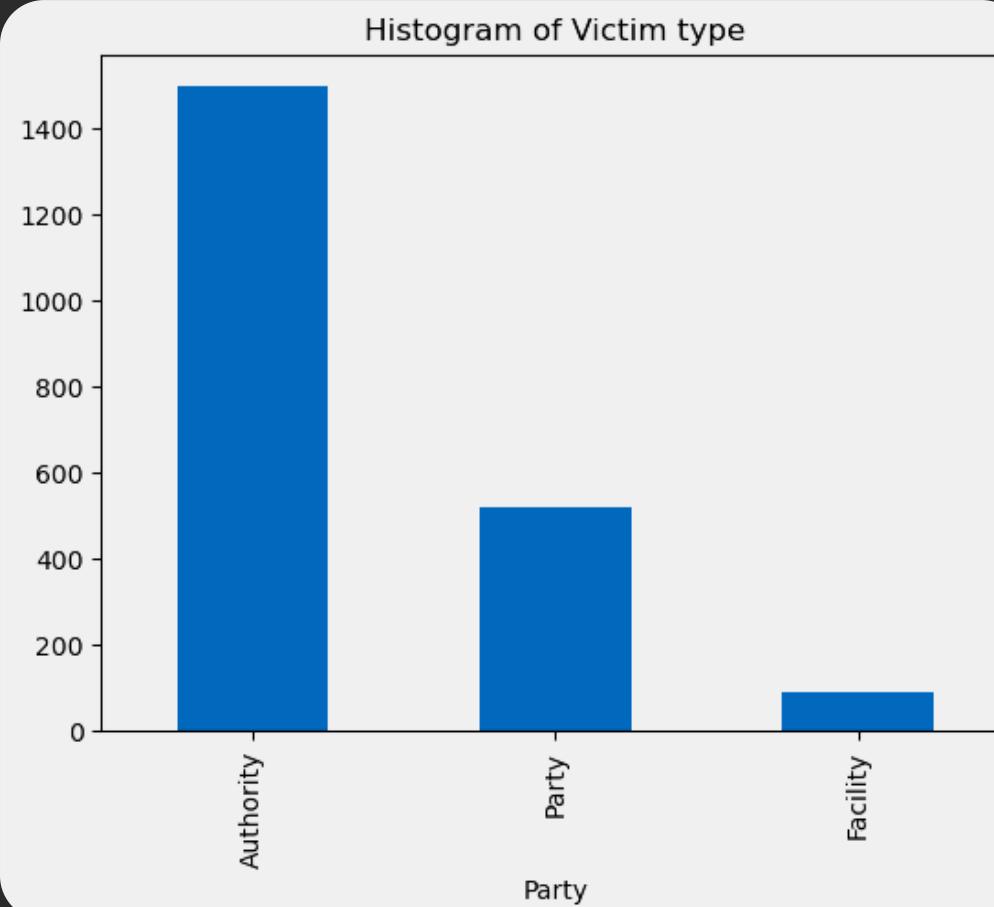
Address missing values by removing null rows or imputing values using statistical techniques.

Identify outliers using statistical methods and normalize data for consistent scaling.

Create frequency tables, histograms, and heat-maps to visualize data.

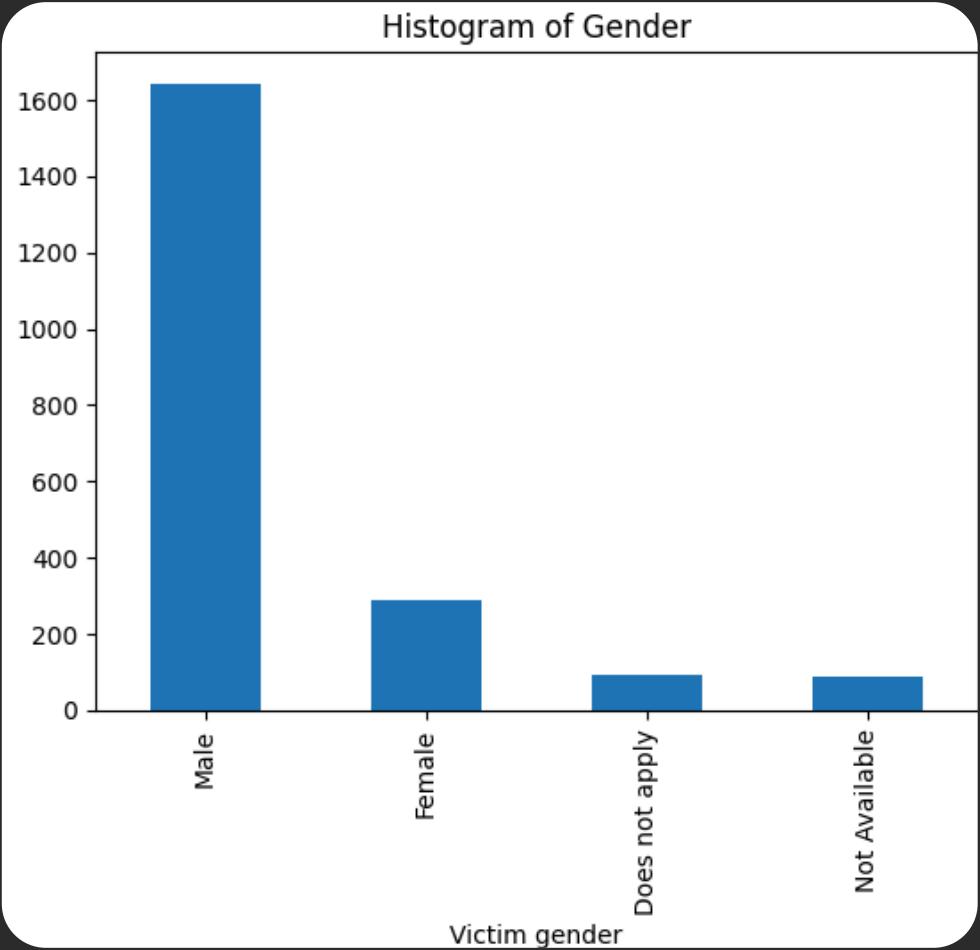
EXPLORATORY ANALYSIS I

VICTIM TYPE HISTOGRAM



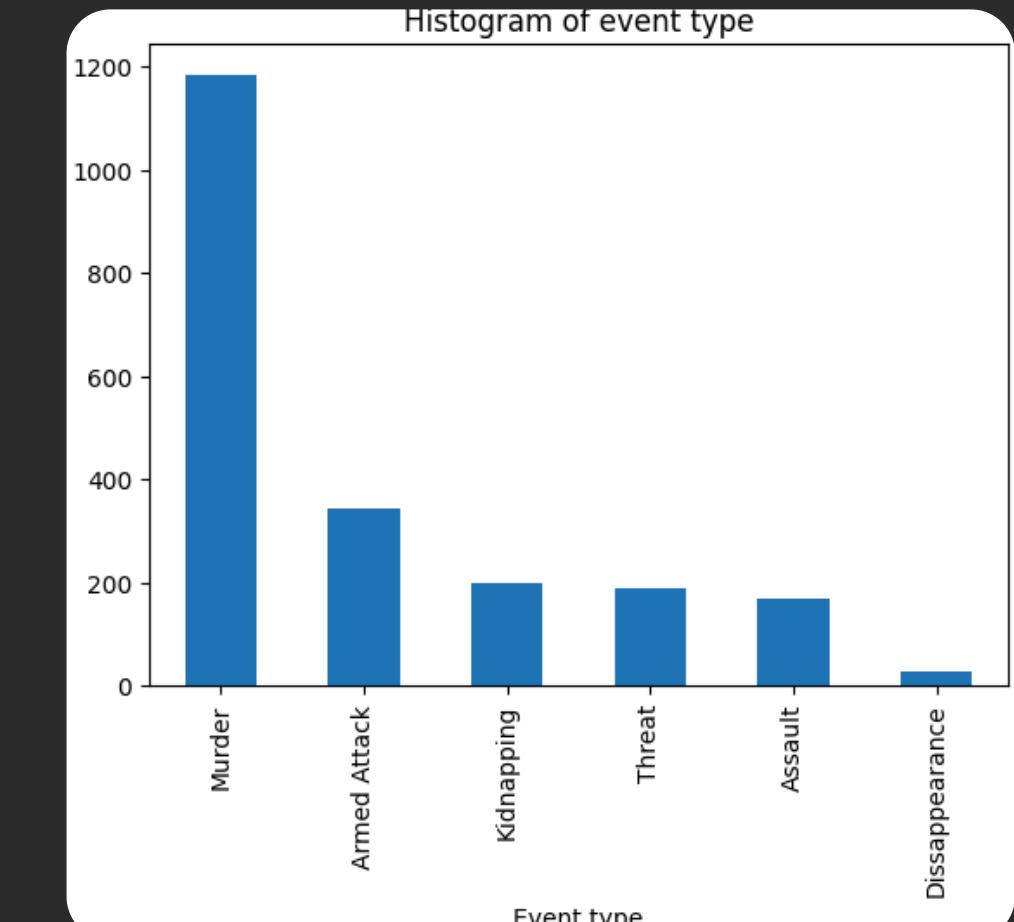
We can identify three main categories no further processing was considered.

GENDER HISTOGRAM



We stayed with three values; “not available” changed by the mode, “does not apply” represents candidates from the LBTQ++ community

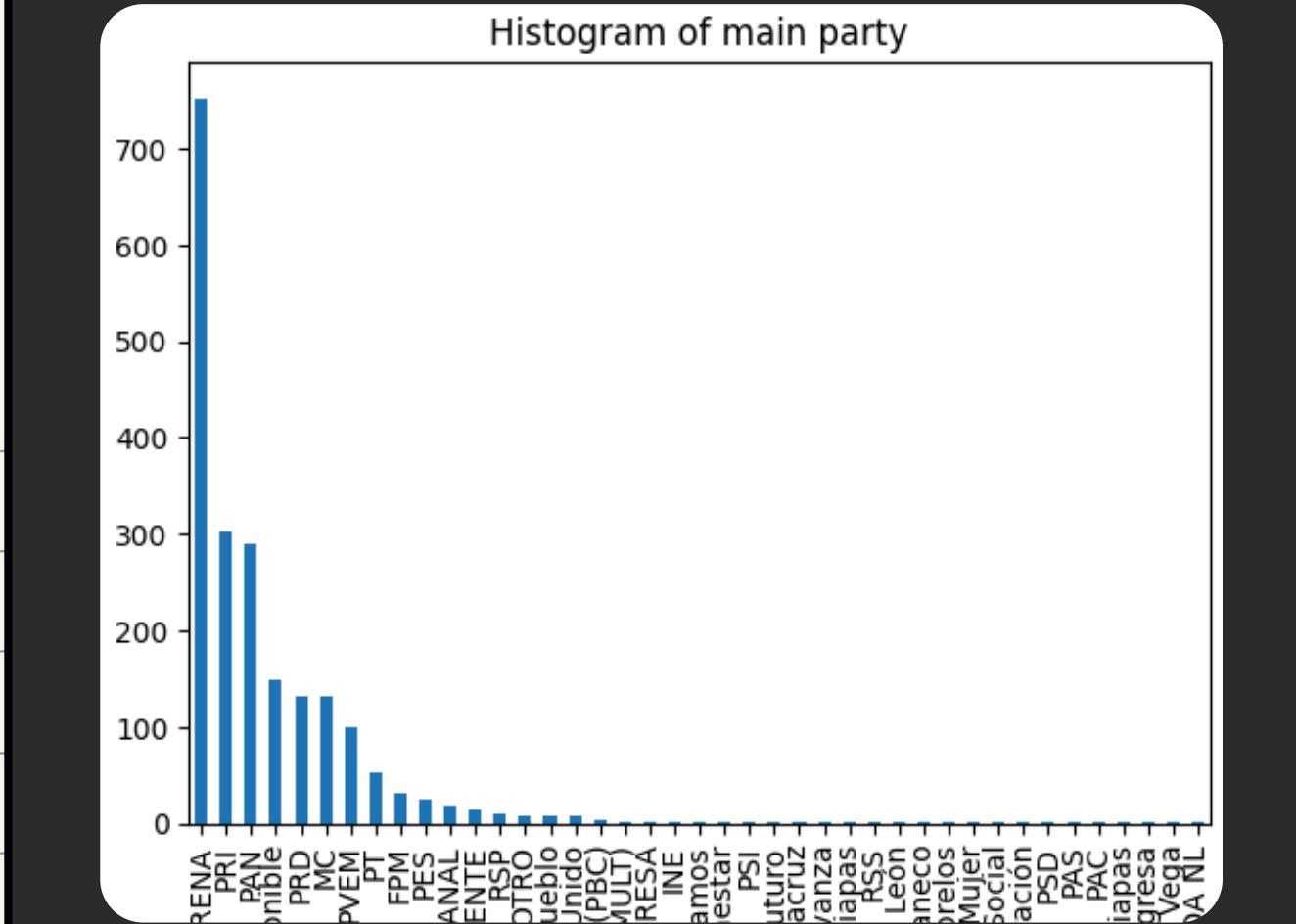
EVENT TYPE HISTOGRAM



We can identify six different categories and no further processing was considered.

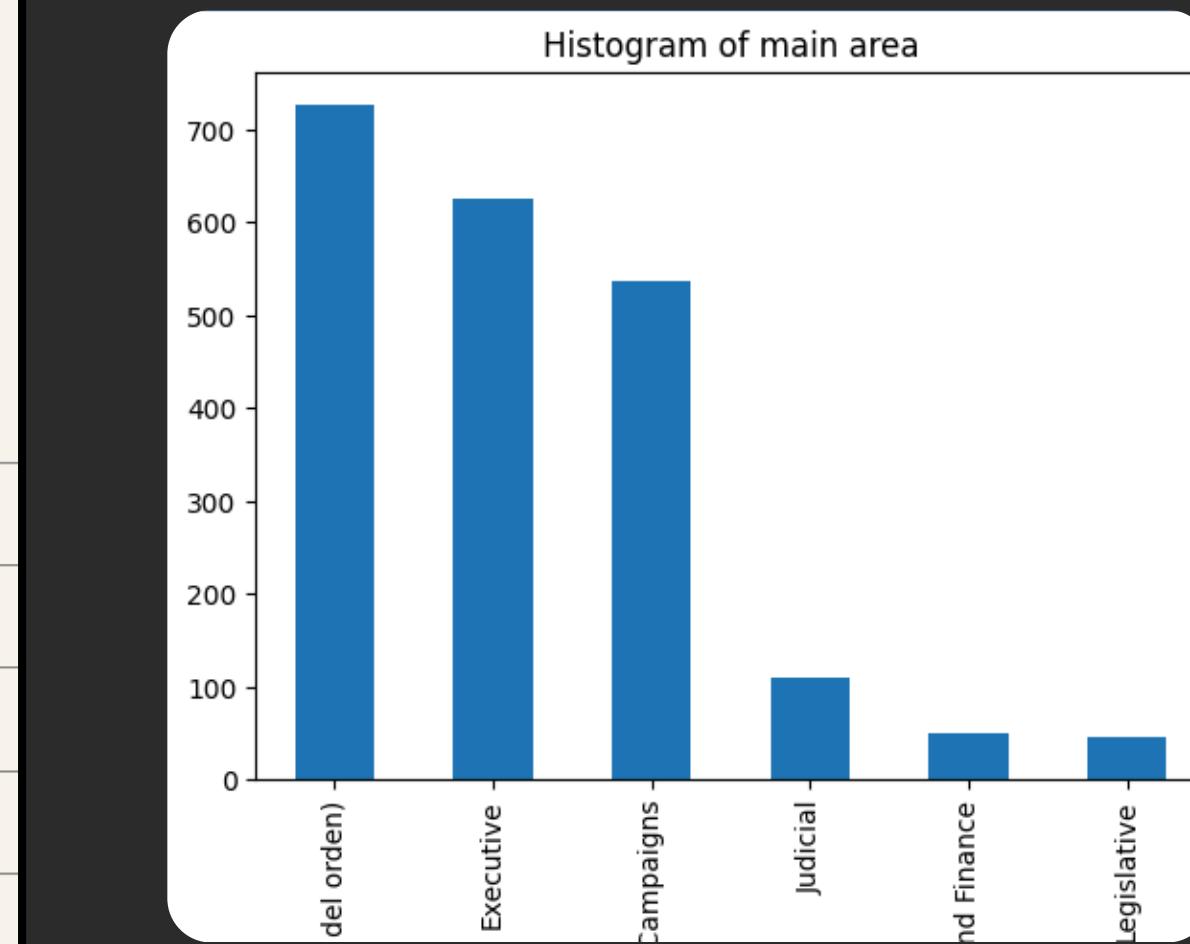
EXPLORATORY ANALYSIS II

MAIN PARTY HISTOGRAM



The data for the main party had a high level of granularity. We chose to group the top 10 parties while consolidating the remaining parties into an "Others" category.

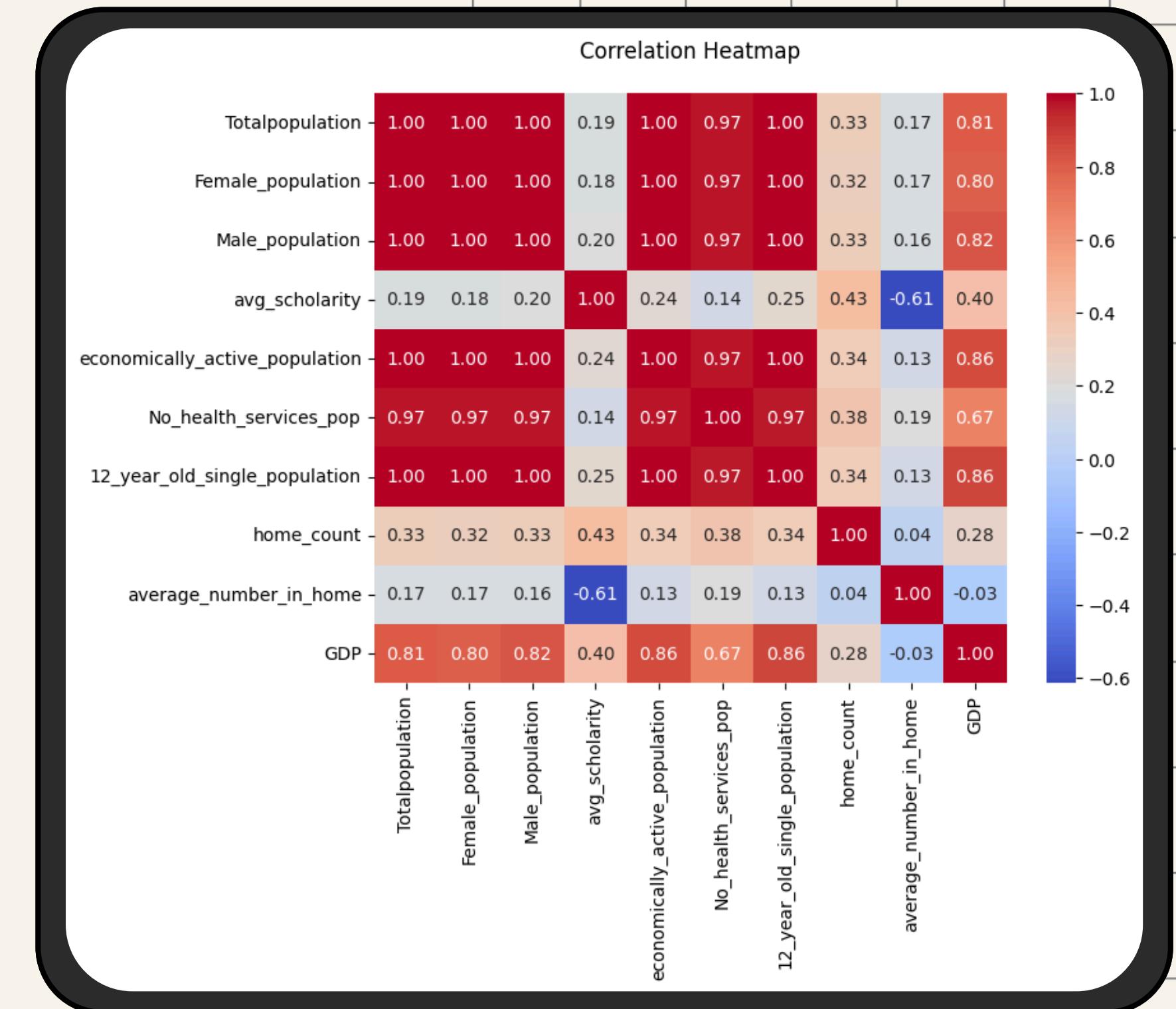
AREA HISTOGRAM



For the six categories in area, no further process was done.

CORRELATION MATRIX

- Some variables show a correlation of 1 due to overlapping population aggregations.
- Retained key variables: gender distribution, lack of access to health services, and total population.
- Omitted economically active population as it overlaps with GDP data.



POST-PROCESSING MATRIX



SECTION 03

MODEL BUILDING

INSY 662 - FINAL PROJECT REPORT



MODELING STEPS

We explored multiple models with various feature selection techniques. KNN without feature selection delivered comparable performance to other methods while being computationally efficient, making it the preferred choice for stakeholders.

#1

SCALING

Applied normal scaling to standardize the dataset for better model performance.

#2

FEATURE SELECTION

Utilized Decision Trees and Neural Networks to evaluate key features for analysis.

#3

MODEL TUNING

Focused on optimizing KNN clustering parameters to achieve the best results.

INITIAL MODELING

DATA SPLITTING

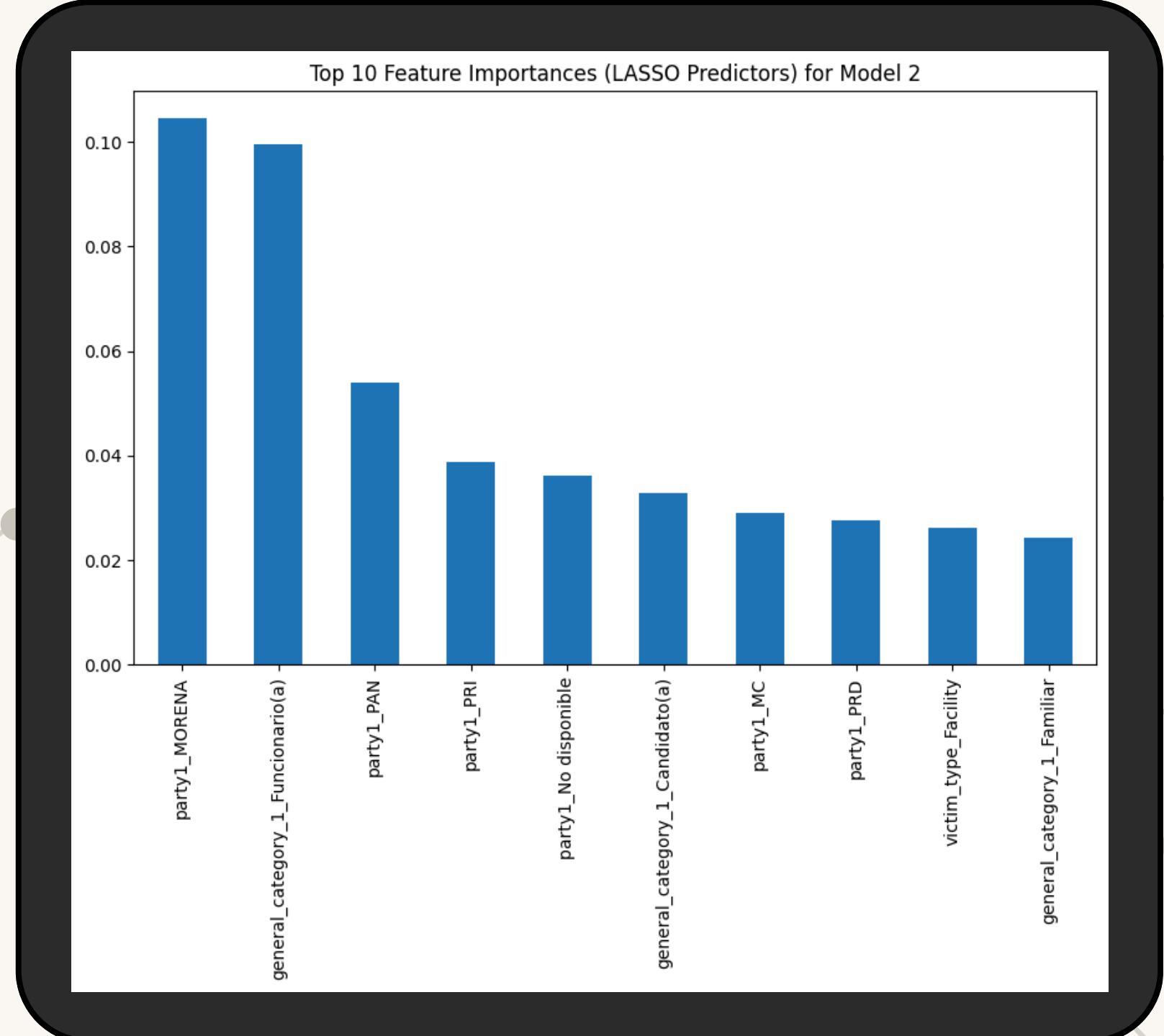
- Split the dataset into training (70%) and testing (30%) sets without scaling.

MODEL TESTING

- Trained three models on the raw data:
 - Random Forest Classifier
 - LASSO Regression
 - K-Nearest Neighbors (KNN)
 - Grid Search

MODEL EVALUATION

- KNN achieved the highest initial accuracy of 57%.



MODEL-TUNING

DATA SPLITTING

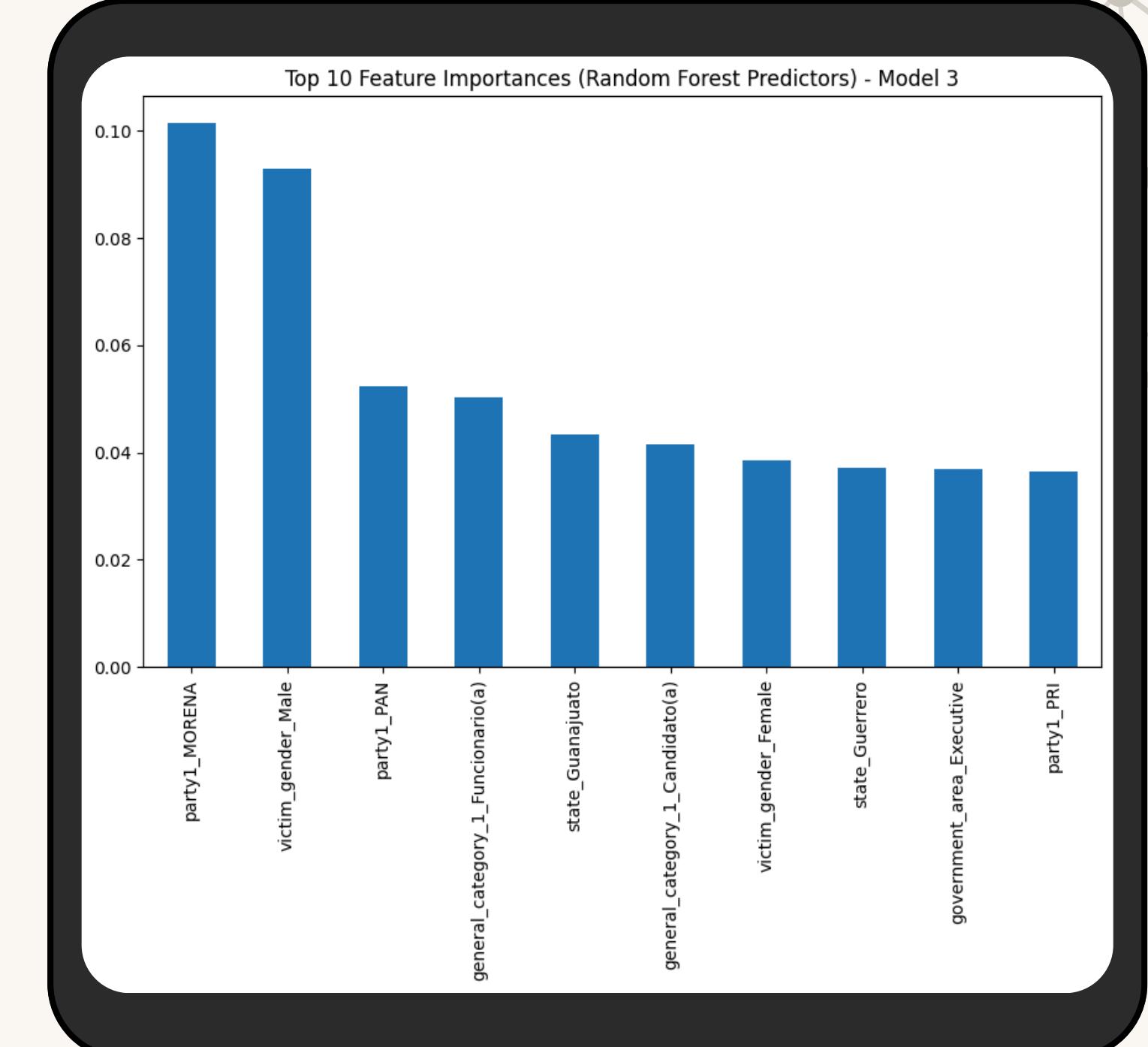
- Split the dataset into training (70%) and testing (30%) sets without scaling.

KNN TESTING WITH FEATURE SELECTION

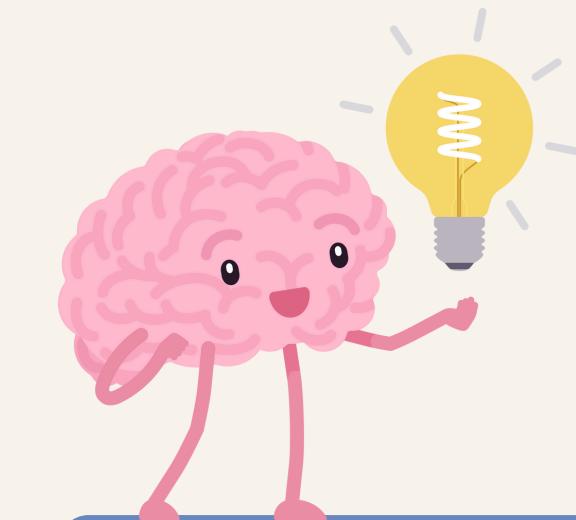
- Trained 3 KNN models with features from:
 - Random Forest Classifier
 - LASSO Regression

MODEL EVALUATION

- **KNN no feature** selection and with random forest achieved the same accuracy of 65%.



FEATURE SELECTION TO BOOST PERFORMANCE



FEATURE SELECTION METHODS

- **LASSO Regression:** Eliminated less important features by shrinking coefficients to zero.
- **Random Forest Feature Importances:** Ranked features based on their contribution to the target variable.

RETRAINING MODEL

- Retrained the models using selected features from LASSO and Random Forest methods to improve efficiency and accuracy.



KNN WITHOUT FEATURE SELECTION

- Trained KNN on the original dataset without feature selection.

Key Observation:

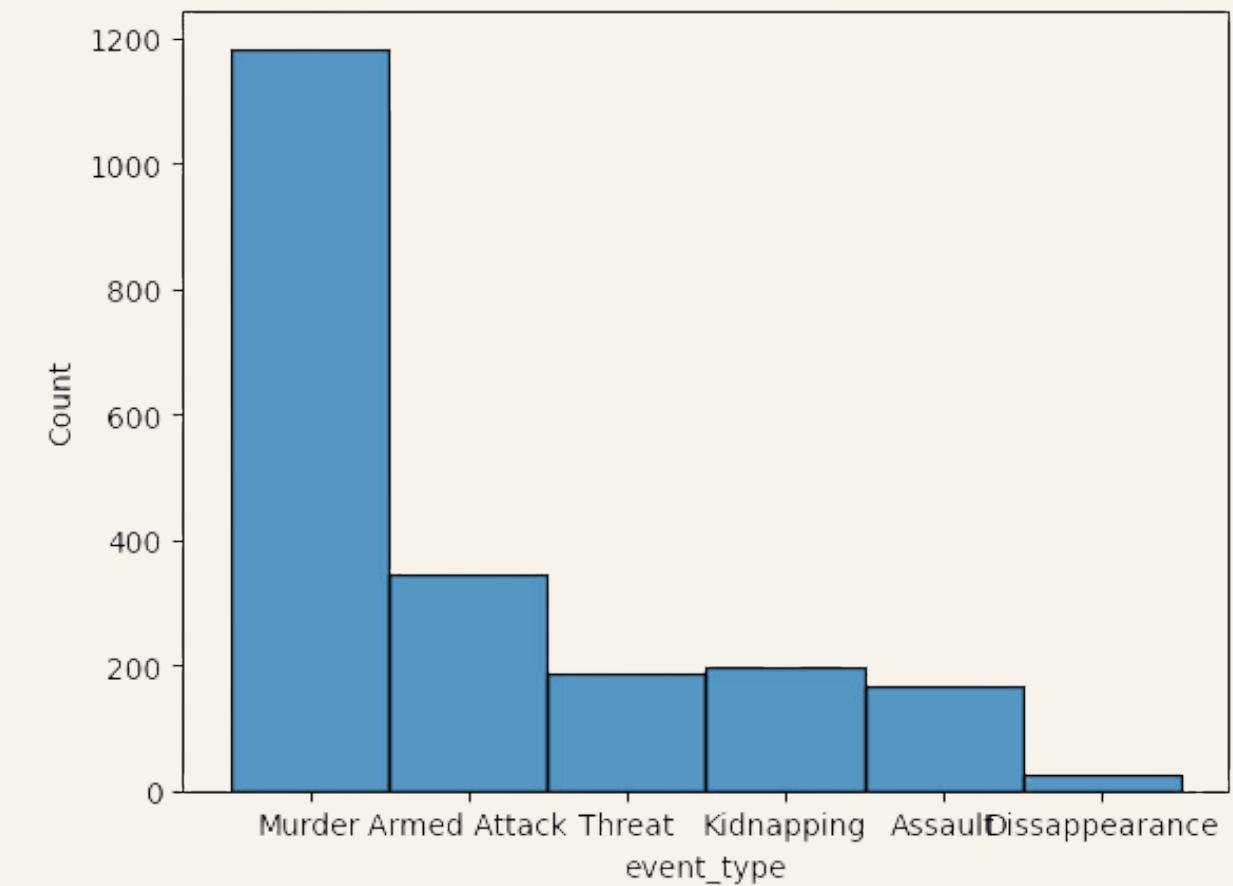
- KNN performed comparably to models using feature selection, making it a computationally efficient choice.

KNN MODEL

WITHOUT FEATURE SELECTION

- Significant misclassification persisted due to an imbalance in the target variable.
- Applied **SMOTE** to increase all non-murder categories to 350 data points each.

BEFORE



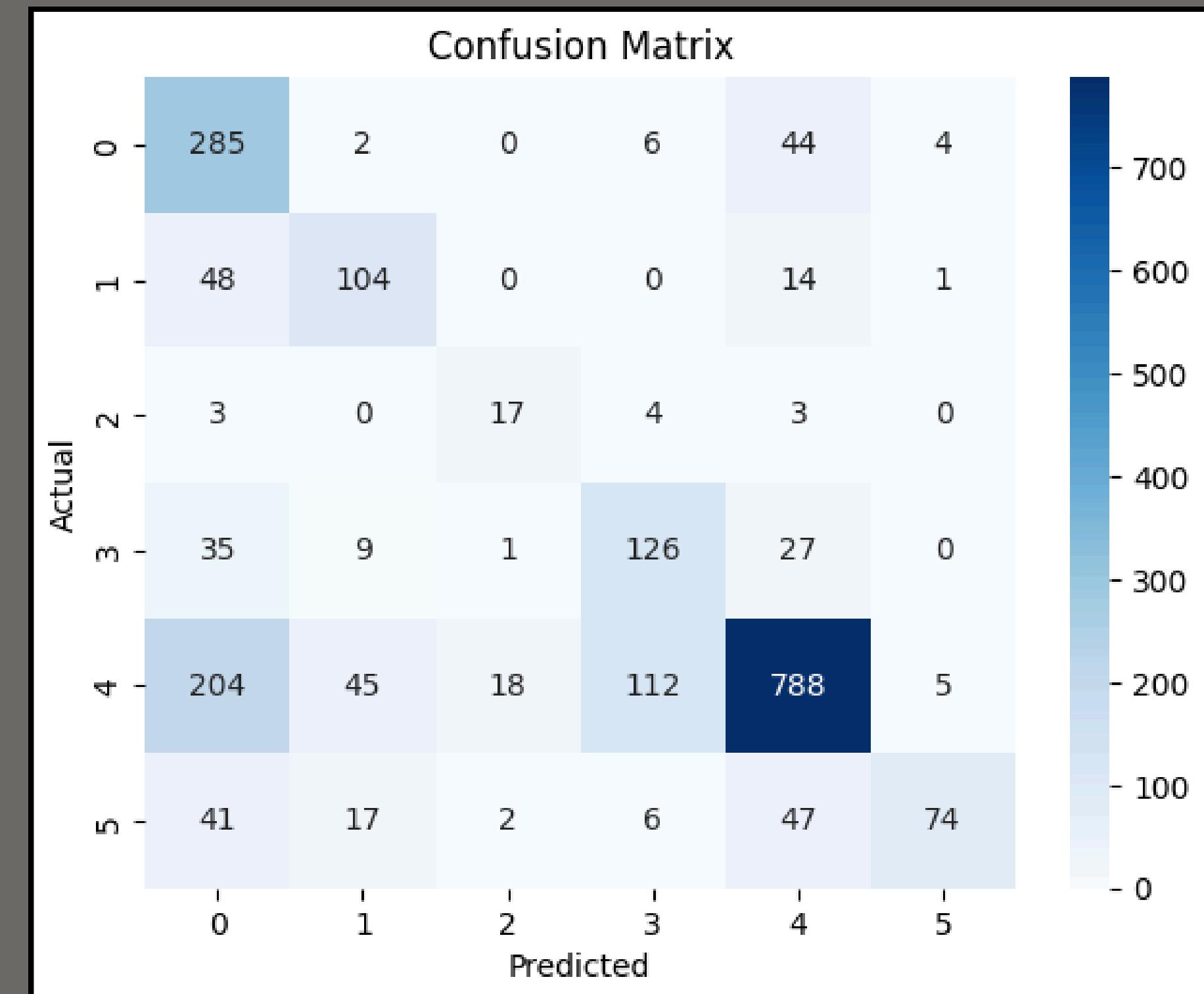
Test set accuracy: 0.6663479923518164				
Classification Report:				
	precision	recall	f1-score	support
0	0.46	0.84	0.60	341
1	0.59	0.62	0.60	167
2	0.45	0.63	0.52	27
3	0.50	0.64	0.56	198
4	0.85	0.67	0.75	1172
5	0.88	0.40	0.55	187
accuracy			0.67	2092
macro avg	0.62	0.63	0.60	2092
weighted avg	0.73	0.67	0.68	2092

KNN MODEL

WITHOUT FEATURE SELECTION

- Significant misclassification persisted due to an imbalance in the target variable.
- Applied **SMOTE** to increase all non-murder categories to 350 data points each.

BEFORE



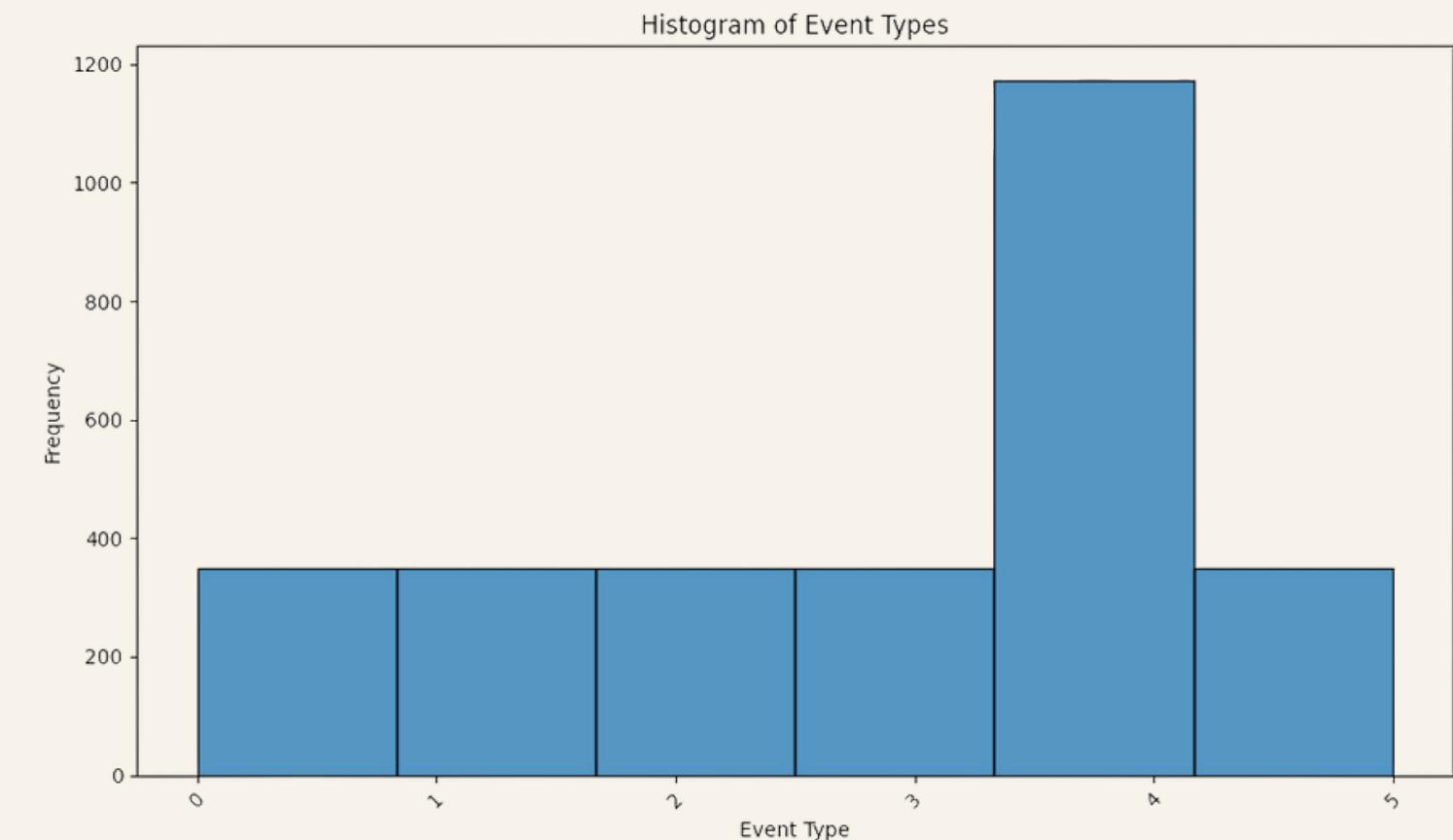
accuracy	0.67	2092
macro avg	0.62	0.63
weighted avg	0.73	0.67

KNN MODEL

WITHOUT FEATURE SELECTION

- Significant misclassification persisted due to an imbalance in the target variable.
- Applied **SMOTE** to increase all non-murder categories to 350 data points each.
- **Increased accuracy by 6%** and decreased misclassification.

AFTER



Test set accuracy: 0.7231348391512663

Classification Report:

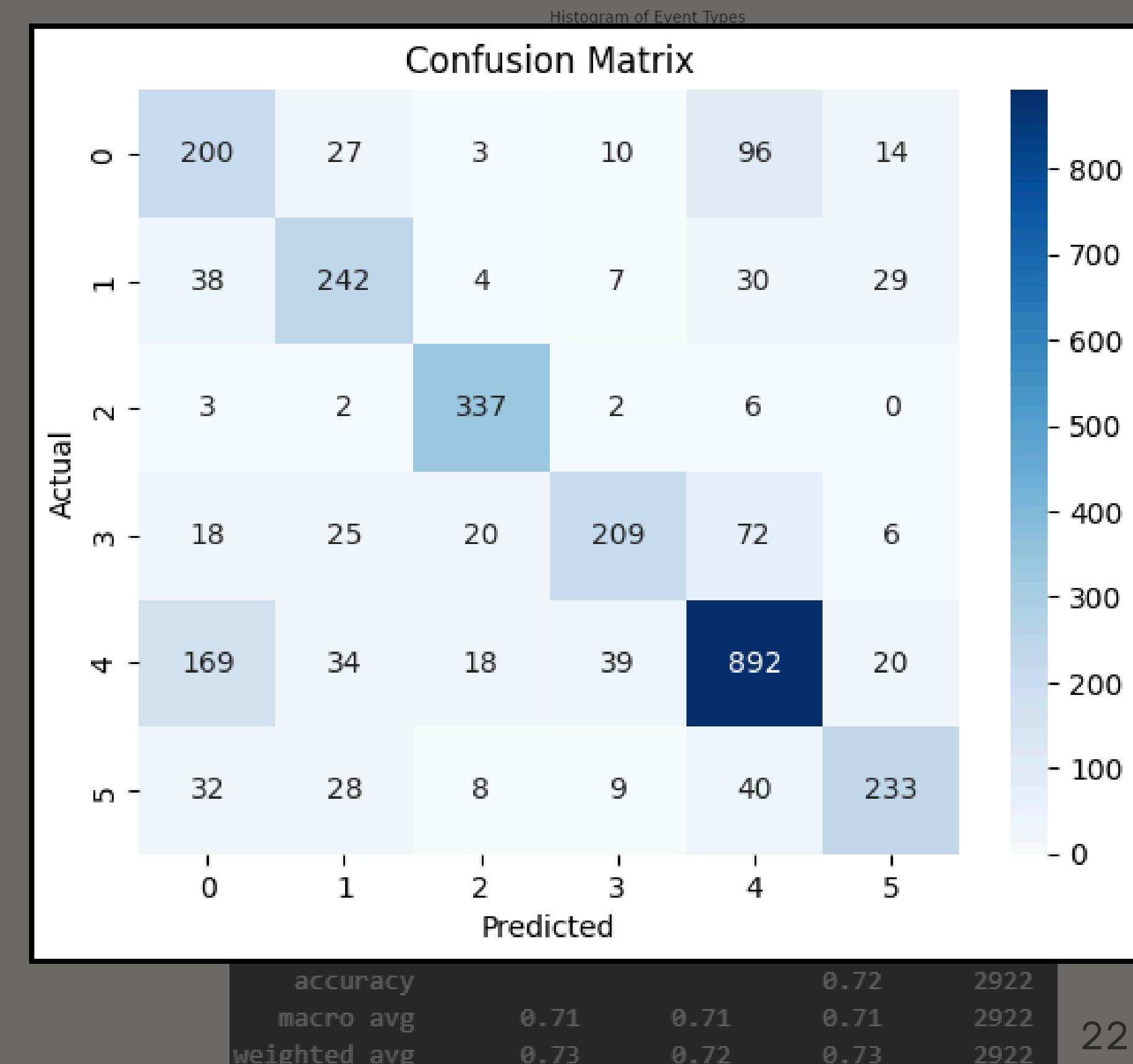
	precision	recall	f1-score	support
0	0.43	0.57	0.49	350
1	0.68	0.69	0.68	350
2	0.86	0.96	0.91	350
3	0.76	0.60	0.67	350
4	0.79	0.76	0.77	1172
5	0.77	0.67	0.71	350
accuracy			0.72	2922
macro avg	0.71	0.71	0.71	2922
weighted avg	0.73	0.72	0.73	2922

KNN MODEL

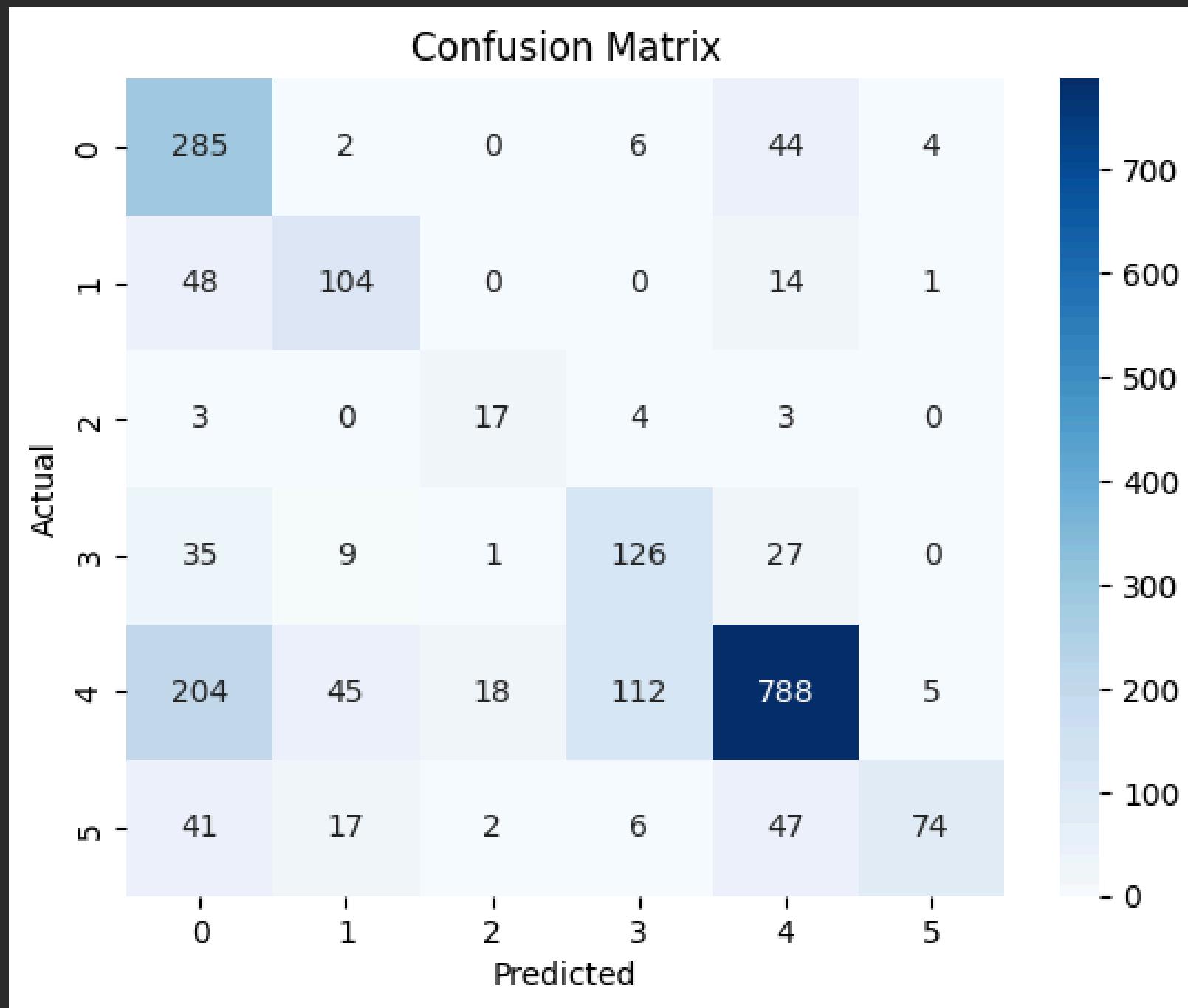
WITHOUT FEATURE SELECTION

- Significant misclassification persisted due to an imbalance in the target variable.
- Applied **SMOTE** to increase all non-murder categories to 350 data points each.
- Increased accuracy by 6% and decreased misclassification.

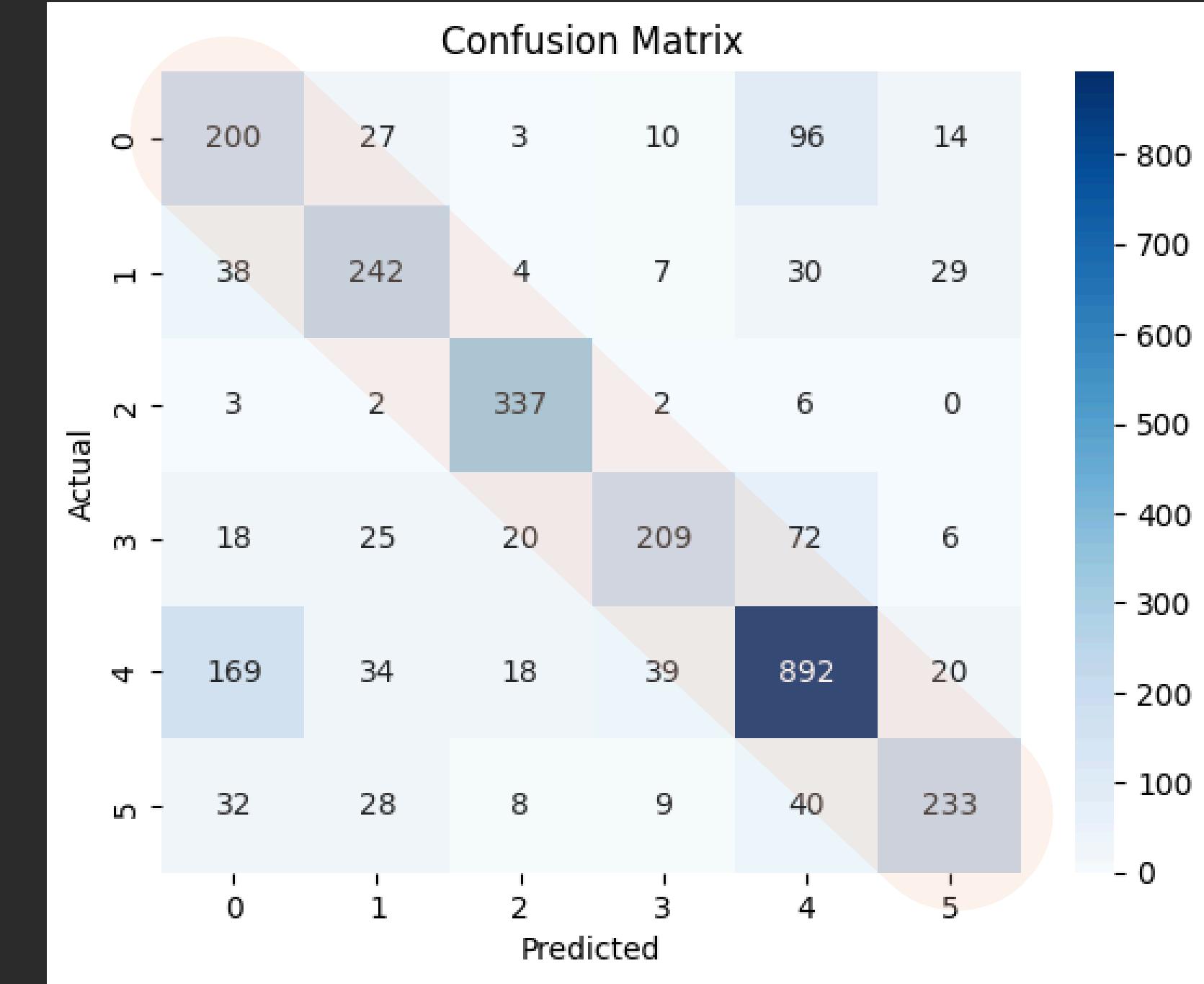
AFTER



BEFORE



AFTER



SECTION 04

RESULTS & CONCLUSION

INSY 662 - FINAL PROJECT REPORT

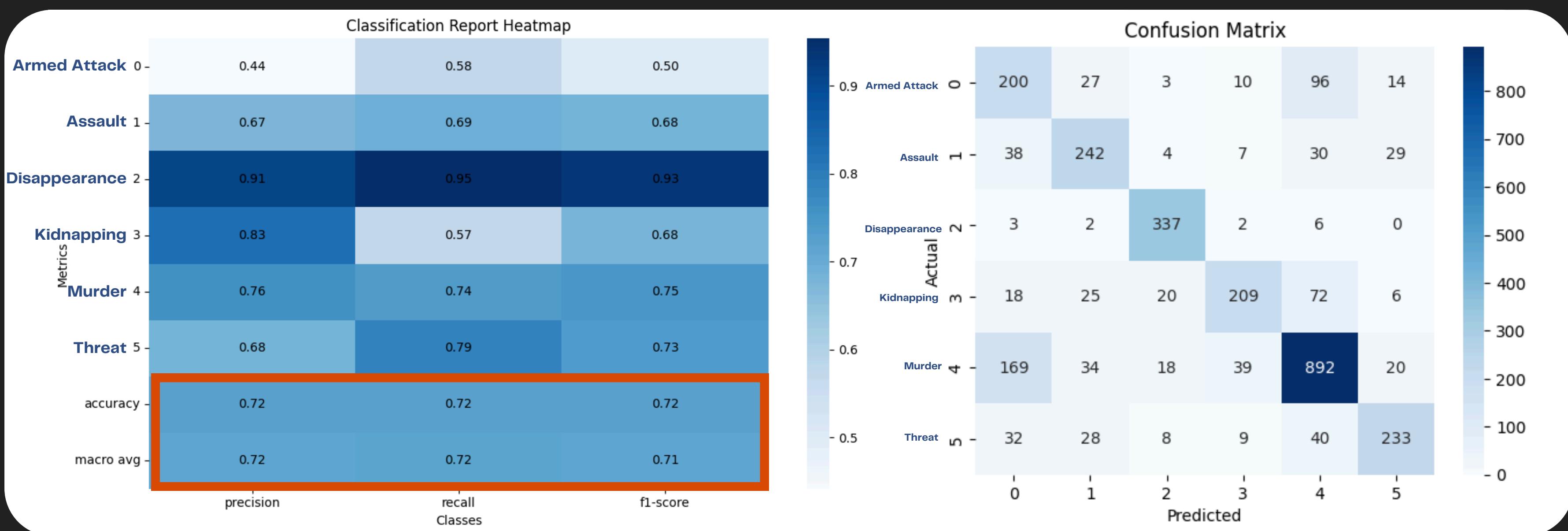


FINAL MODEL

KNN Model

Accuracy: 72%

• KNeighborsClassifier
KNeighborsClassifier(n_neighbors=3)

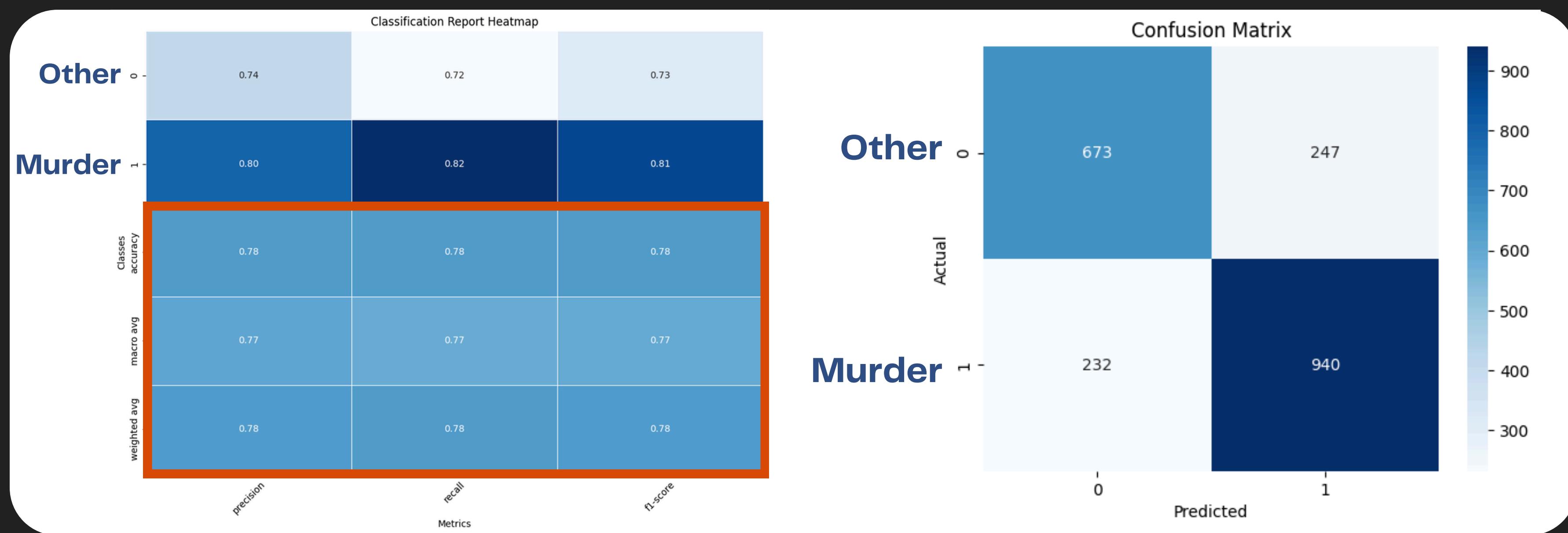


FINAL MODEL

KNN Model(BINARY)

Accuracy: 77%

• KNeighborsClassifier
KNeighborsClassifier(n_neighbors=3)



FEATURE IMPORTANCE

STATE

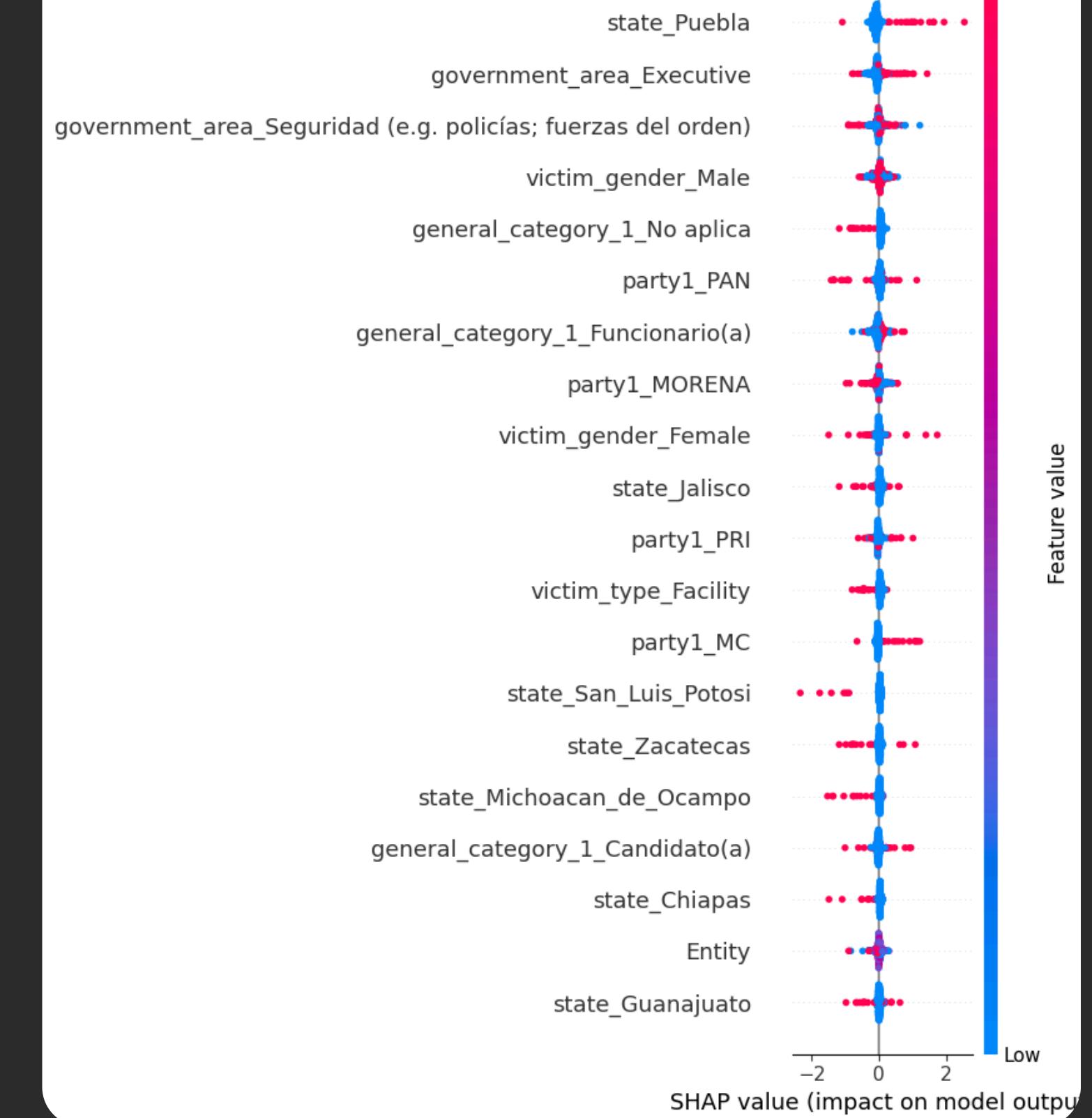
- Puebla
- Jalisco
- Zacatecas

DEMOGRAPHIC AT RISK

- Both Gender
- Executive
- Government official (Fonctionario)

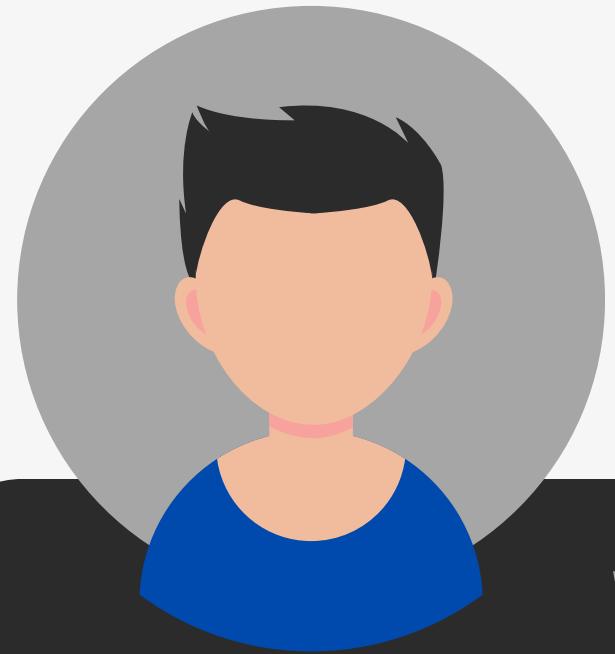
PARTYS AT RISK

- National Action Party (PAN)
- Institutional Revolutionary Party (PRI)



SAMPLE PREDICTION





Victim Name: John Doe

Event Date: January 1, 2024

EVENT TYPE:
PENDING PREDICTION

Victim Type: Authority

Victim Gender: Male

Political Context:

- Party Affiliation: PRD (no coalition)
- First Political Office: Pre-candidate for PRI mayor of Atoyac
- Government Level: Municipal
- Government Area: Executive

Event Location:

- State: Guerrero
- Municipal: Atoyac de Álvarez
- INEGI Codes: State - 99999, Municipal - 99999

Demographics:

- Total Population: 1,000,000
- Male/Female Population: 500,000 each
- Economically Active Population: 600,000
- Population W/out Health Services: 200,000
- 12-Year-Old Single Population: 150,000

Socioeconomic Indicators:

- Average Scholarly: 8.5 years
- Homes Count: 250,000
- Average Household Size: 4.0
- GDP: \$500,000



Victim Name: John Doe

Event Date: January 1, 2024

EVENT TYPE:
PREDICTION COMPLETE

Victim Type: Authority

Victim Gender: Male

Political Context:

- Party Affiliation: PRI
- First Political Position: Mayor
- for PRI mayoral candidate

DISAPPEARANCE

- Government Level: Municipal
- Government Area: Executive

- Population W/out Health Services: 200,000
- 12-Year-Old Single Population: 150,000

Event Location:

- State: Guerrero
- Municipal: Atoyac de Álvarez
- INEGI Codes: State - 99999,
Municipal - 99999

Socioeconomic Indicators:

- Average Scholarly: 8.5 years
- Homes Count: 250,000
- Average Household Size: 4.0
- GDP: \$500,000

FINAL TAKEAWAYS

MODELLING INSIGHTS

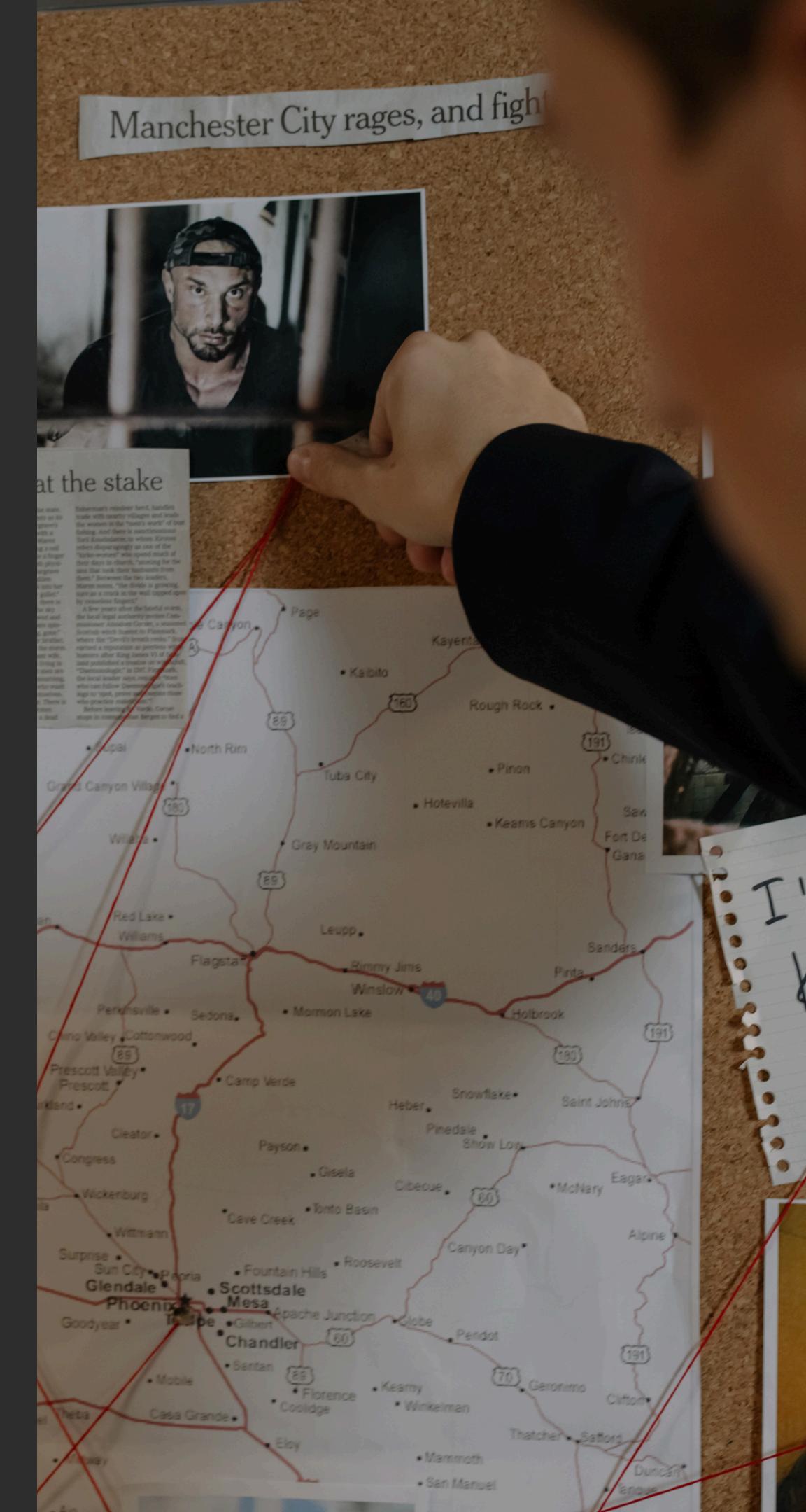
- Effective feature selection and optimization were crucial.
- Preprocessing and model tuning significantly improved performance.
- SMOTE addressed class imbalance effectively.

LIMITATIONS

- Assumes attacks; additional data could enhance model accuracy.

KEY VARIABLES

- Political party, state location, and individual roles were the most critical factors for predicting political violence.



SOCIETAL IMPACT



GOVERNMENT AGENCIES

Use model predictions to develop region-specific policies that counter political violence and address vulnerabilities exploited by organized crime.



LAW ENFORCEMENT

Focus resources on high-risk zones and threats while safeguarding political figures with personalized risk profiles.



POLITICAL PARTIES

Improve campaign security by identifying high-risk areas and providing tailored risk assessments.



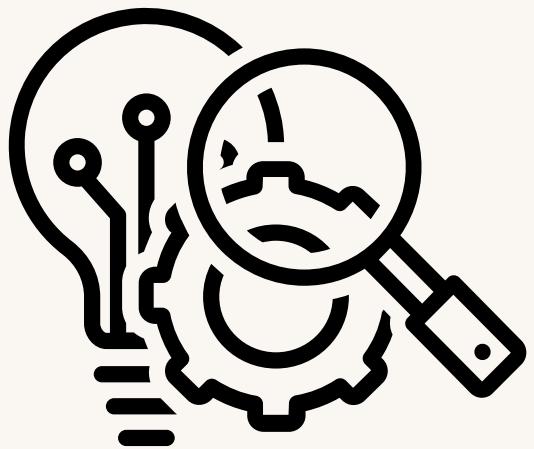
POLICY MAKERS

Tackle root causes like unemployment and instability through targeted policies to reduce crime and foster long-term stability.

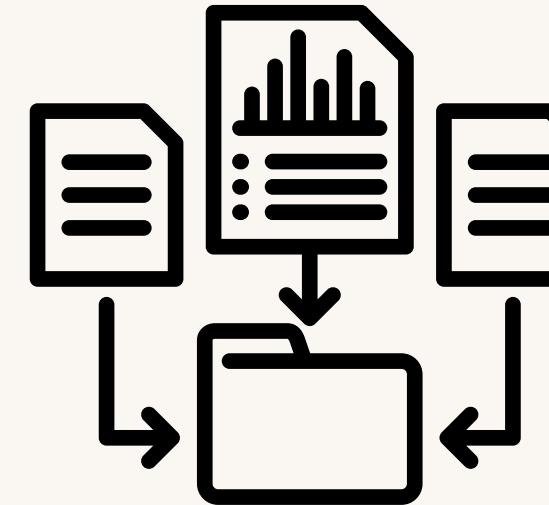
FUTURE DIRECTIONS



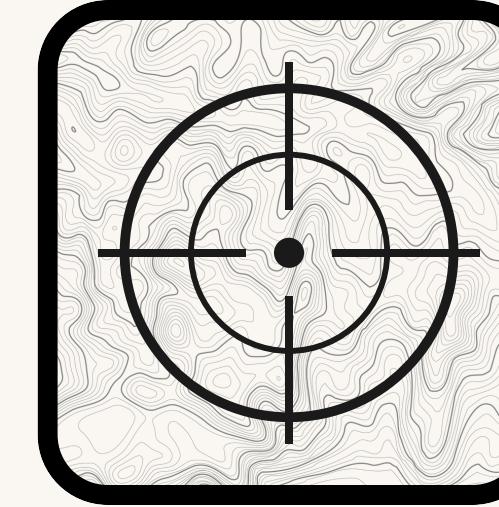
Additional enhancements that can make our model more robust, actionable, and widely applicable.



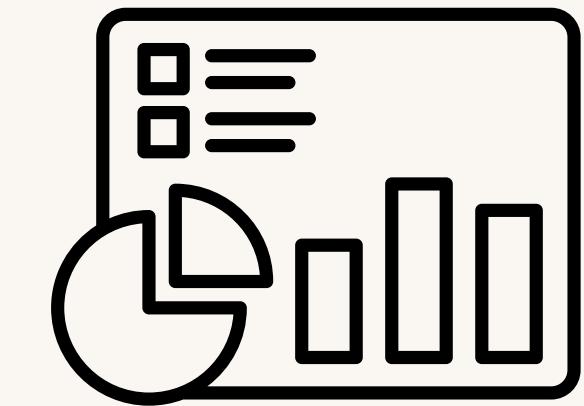
Improve Model Accuracy.



Enhance Law Enforcement Collaboration.



Broaden The Project Scope.
(unharmed politician)



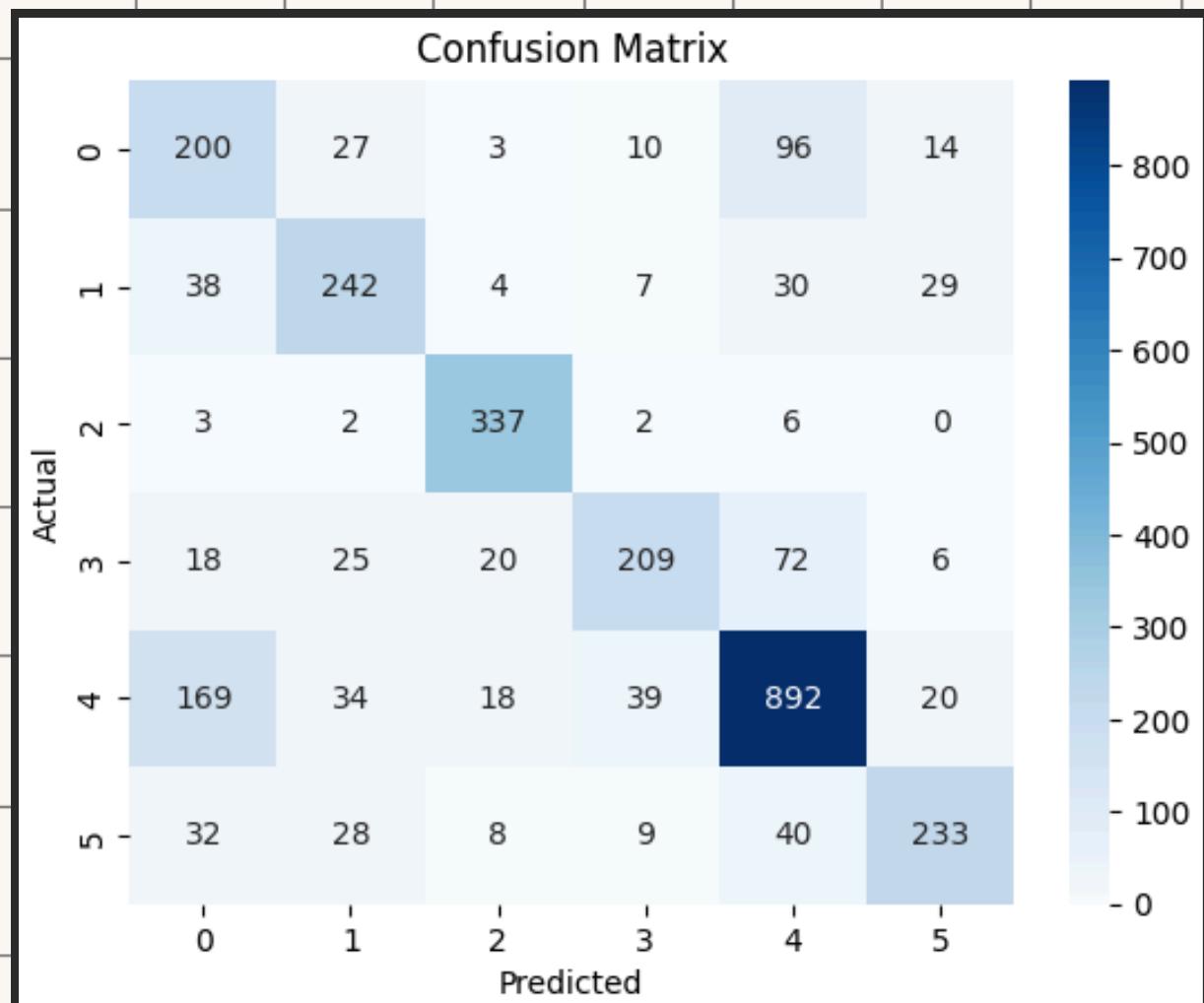
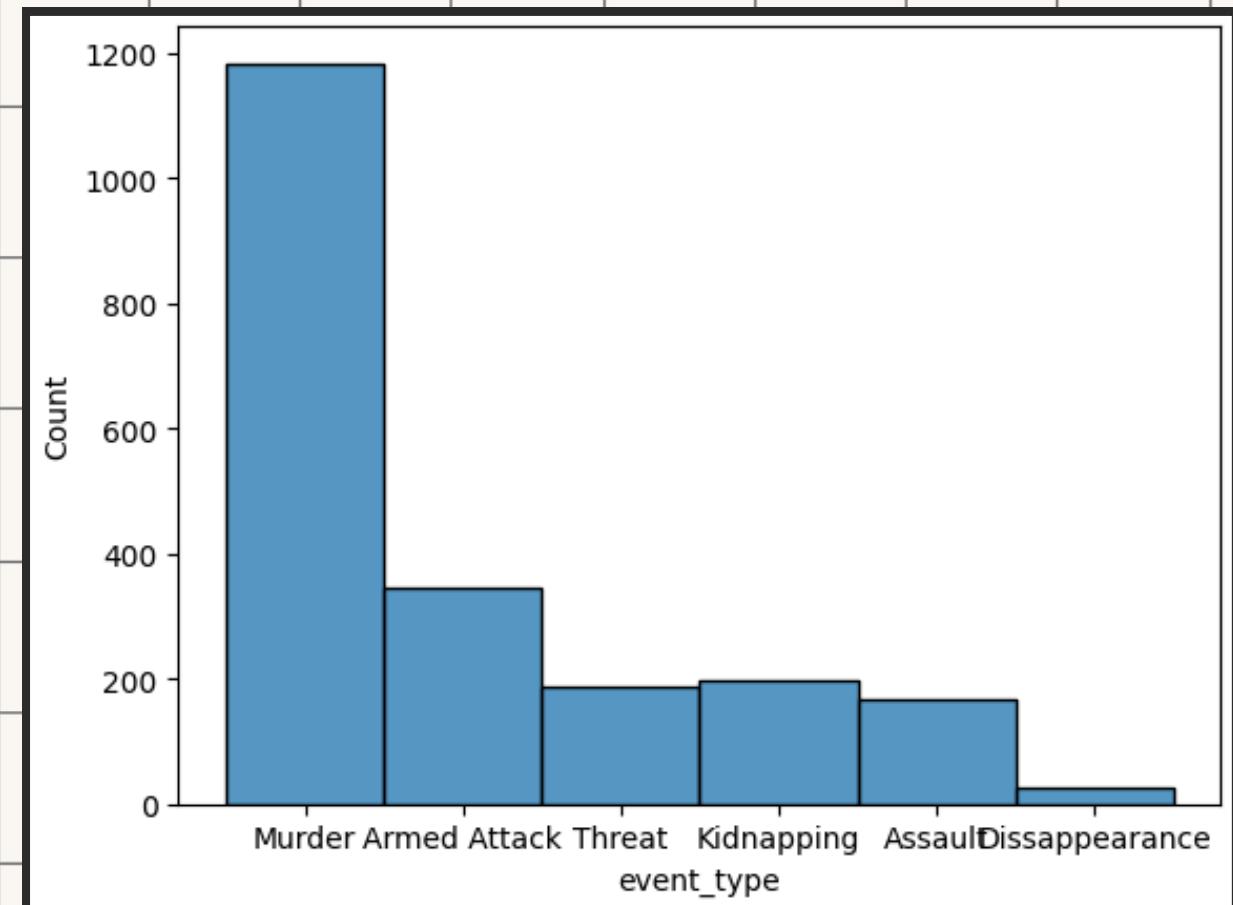
Integrate Explainability Tools.



**THANK YOU.
QUESTIONS & ANSWERS**

APPENDIX

- **Selected KNN without feature selection.**
 - However there was still a lot of misclassification.
 - Over-predicting 4 (Murder).
 - Caused by unbalanced amount of target variable.
 - Performed SMOTE to boost all non-murder to 350 data points each.
 - increased accuracy by 4% and decreased misclassification.
- Applying Permutation Importance.
 - Identified significant features affecting KNN.
 - Tested thresholds (0, 0.1, 0.001, 0.005, 0.0001).
 - Selected threshold of 0.001 for optimal results.
- **Final Model Evaluation**
 - **Retrained KNN with initially selected features.**
 - **Achieved accuracy of 72%, indicating significant improvement.**



REFERENCES

99

- Data Cívica. (n.d.). Votar entre balas. Data Cívica. Retrieved November 23, 2024, from <https://votar-entre-balas.datacivica.org/>
- Merriam-Webster. (n.d.). Political crime. In Merriam-Webster.com dictionary. Retrieved November 23, 2024, from <https://www.merriam-webster.com/dictionary/political%20crime>
- Santiago, R. (2017). ¿Cuáles son las causas de la violencia en México? Forbes México. Retrieved from <https://www.forbes.com.mx/cuales-son-las-causas-de-la-violencia-en-mexico/>
- Valencia, S. (2018). Gore capitalism. Semiotext(e).

REFERENCES

66²²

- Becerra Romero, A. T. (2018). Investigación documental sobre la narcocultura como objeto de estudio en México. *Culturales*, 6, e349. <https://doi.org/10.22234/recu.20180601.e349>
- Donatello, B. (2024, March 21). ¡Despoblación por causa del narco! *El Economista*. <https://www.eleconomista.com.mx/opinion/Despoblacion-por-causa-del-narco-20240321-0022.html>
- Zepeda G (2024). “Fracaso de México en lograr empleo digno aumenta los reclutamientos del narcotráfico”. *El Financiero*. Retrieved from: <https://www.elfinanciero.com.mx/nacional/2024/05/14/fracaso-de-mexico-en-lograr-empleo-digno-aumenta-los-reclutamientos-del-narcotrafico/>
- Gonzales, M (2022). “Los niños del narco en México | "Finalmente encontré algo que me hacía sentir mejor que la droga: matar". BBC NEWS.
- Sampieri, H (2006). “Metodología de la investigación”. McGrawHill. Fourth Edition.