**Project Objective**

**Predict** whether an **arrest was made** after a Terry Stop using features like weapon presence, time of day, officer/subject demographics, etc.

We’ll also explore **bias analysis** regarding race and gender as an optional ethical component.

**Plan**

1. **Data Preparation**
2. **Exploratory Data Analysis (EDA)**
3. **Feature Engineering**
4. **Model Training (e.g., Logistic Regression, Random Forest, XGBoost)**
5. **Model Evaluation**
6. **Bias & Fairness Analysis (Optional)**
7. **Non-Technical PowerPoint Slides**

I’ll begin by preprocessing the dataset:

* Convert time and date
* Clean target and categorical values
* Encode categories
* Handle missing values

Let's proceed with the preprocessing and class balancing.

✅ Preprocessing complete: we now have **63,708 rows** and **229 features** ready for modeling after encoding and cleaning.

**Next Step: Model Training & Evaluation**

We’ll:

* Split the data
* Train models (Logistic Regression, Random Forest, XGBoost)
* Evaluate with metrics like accuracy, precision, recall, F1-score, and ROC-AUC

| **Column Name** | **Encoding Recommendation** |
| --- | --- |
| 'Subject Age Group' | Label or One-Hot |
| 'Stop Resolution' | Label or One-Hot |
| 'Weapon Type' | Label or One-Hot |
| 'Officer Gender' | One-Hot |
| 'Officer Race' | One-Hot |
| 'Subject Perceived Race' | One-Hot |
| 'Subject Perceived Gender' | One-Hot |
| 'Initial Call Type' | One-Hot |
| 'Final Call Type' | One-Hot |
| 'Call Type' | One-Hot |
| 'Officer Squad' | One-Hot |
| 'Precinct', 'Sector', 'Beat' | One-Hot |
| 'Reported\_month\_year' | One-Hot or convert to Period (for time series) |