

PDCode\_PenalLaw Dataset:

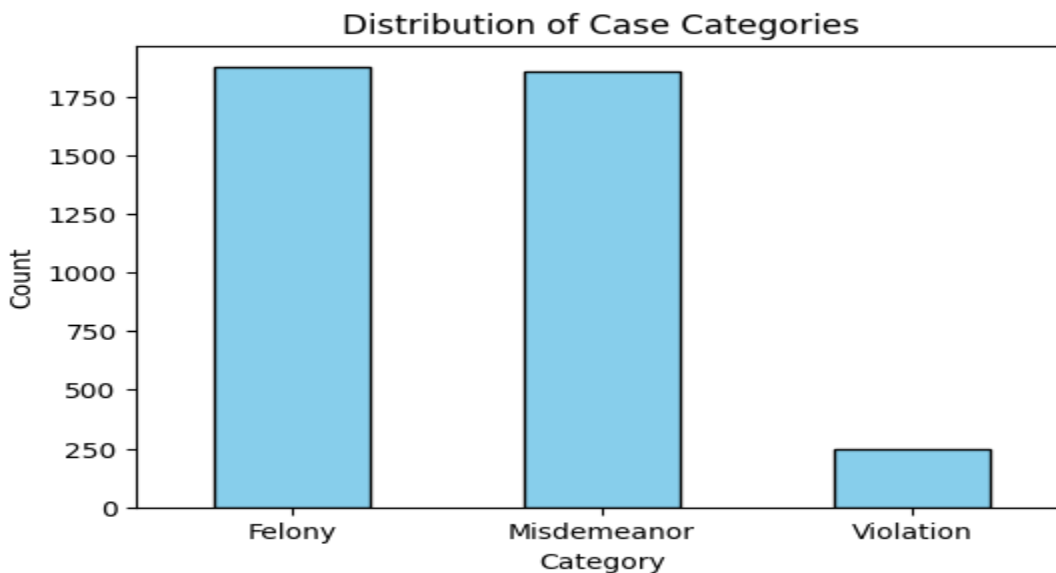
## Quick EDA Results

### 1. Dataset Size

- **Rows:** 4,671
- **Columns:** 5
- No missing values ✓

### 2. Categories (CATEGORY)

- **Felony (F):** 1,872 # Most serious crimes, punish more than 1 year in prison
- **Misdemeanor (M):** 1,856 # up to 1 year in jail, probation, fines (less serious than felonies but more serious than violations)
- **Violation (V):** 250 # least serious offenses, fines, tickets
- **I (Other / Special):** 682
- **Blank:** 11

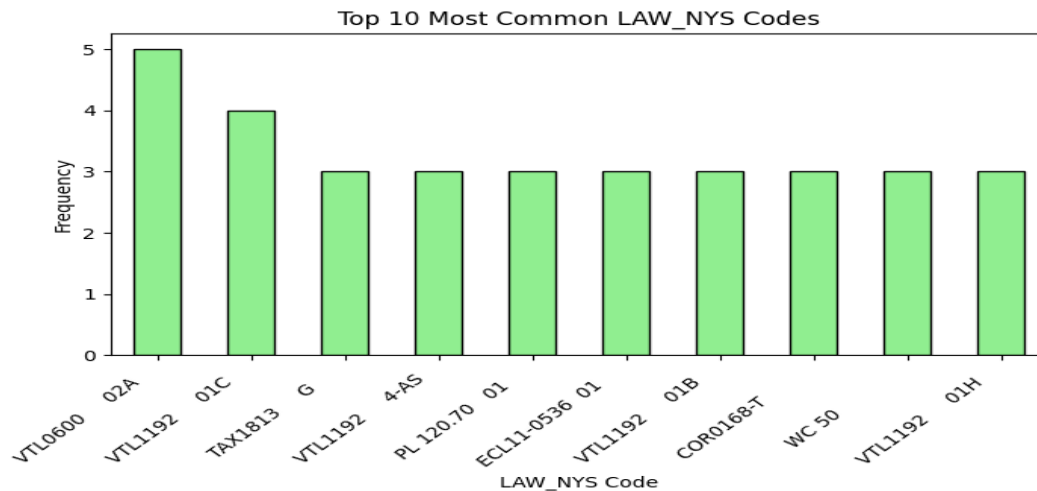


👉 Almost 50/50 split between **Felony and Misdemeanor**, with fewer Violations.

### 3. Legal Codes (LAW\_NYS) # law section reference

- 4,581 unique codes.

- Most frequent:
  - VTL0600 02A → 5 times
  - VTL1192 01C → 4 times
  - Several variations of **VTL1192** (alcohol/driving law) → 3 times each
  - TAX1813 G → 3 times



#### 4. Offense Descriptions

- **4,265 unique long descriptions** (LIT\_LONG).
- Common keywords: *VIOL, LAW, CORP, COURT, VEHICLE, TAX* → reflect law violation contexts.

This is a **reference/codebook dataset**, not actual crime/incident data.

That means it only explains **types of laws and their categories**, not **where or when they happened**.