Cleaning Columns

Originally there were 224 columns in the mass shooter table, using the logic below I reduced them to 75. I didn't include my logic for every column, below are just the highlights.

1. Personally Identifiable & Redundant Information

- **Shooter's Name** (Shooter_Last_Name, Shooter_First_Name) – Not relevant for analysis, anonymization is preferred.

- **Street-Level Data** (Street_Number, Street_Name, Zip_Code) – **City and state are enough** for geographic analysis; these may introduce privacy concerns.

- **Full_Date, Day, Day_of_Week, Month, Year** – Keep Full_Date and extract time-based features programmatically.

---

2. Excessive Detail on Personal Background

- Height, Weight – No established link to mass violence.

- Birth_Order_Code, Birth_Order – The impact of birth order on mass shootings is weakly studied.

- Number_of_Siblings, Older_Siblings, Younger_Siblings – Unlikely to provide useful insights.

- Sexual_Orientation_Code, Sexual_Orientation – Not analyzing hate crimes, so not directly related.

- Education_Code, School_Performance_Code, School_Performance_Specified – Not analyzing academic struggles as a factor

- Mother_Violent_Treatment – Specific and potentially redundant with Parental_Substance_Abuse, Parent_Criminal_Record, Family_Member_Incarcerated.

---

3. Redundant or Overlapping Criminal & Law Enforcement History

- Part_I_Crimes_Code, Part_II_Crimes_Code – Not analyzing specific crime types

- Domestic_Abuse_Specified_Code, Domestic_Abuse_Specified – Redundant due to History_of_Domestic_Abuse column.

---

4. Excessively Detailed Mental Health History

May not add significant predictive value beyond a **general mental health or crisis indicator**:

- FASD__Fetal_Alcohol_Spectrum_Disorder_ – Extremely specific condition with limited data availability.

- Health_Issues, Health_Issues___Specify – Not testing a hypothesis linking physical health to violence.

- Head_Injury___Possible_TBI – May be relevant but should be assessed based on available cases.

- Voluntary_or_Involuntary_Hospitalization_Code, Voluntary_or_Involuntary_Hospitalization – **Collapsed into a general prior hospitalization column.**

- Voluntary_or_Mandatory_Counseling_Code, Voluntary_or_Mandatory_Counseling – **Same thing.**

- Psychiatric_Medication_Specified, Medication_Category – **Collapsed into binary psychiatric medication column**.

---

6. Excessive Detail on Firearm Use

- Firearm_Proficiency_Code, Firearm_Proficiency –

- Specify_Other_Weapons_or_Gear – Unless analyzing **weapon type impact**, it may not add significant value.

Creating North Star metrics

How the Overall Composite Risk Score was defined and calculated

1. Family Background

These columns provide insights into the shooter's upbringing and family environment:

- **Parental_Divorce_Separation:** Whether the shooter's parents were divorced or separated.

- **Parental_Death_in_Childhood:** Whether the shooter experienced the death of a parent during childhood.

- **Parental_Substance_Abuse:** Whether the shooter's parents had a history of substance abuse.

- **Parent_Criminal_Record:** Whether the shooter's parents had a criminal record.

- **Family_Member_Incarcerated:** Whether a family member was incarcerated.

---

2. Childhood Trauma and Abuse

These columns provide insights into the shooter's childhood experiences:

- **Childhood_Trauma:** Whether the shooter experienced trauma during childhood.

- **Physically_Abused:** Whether the shooter was physically abused.

- **Sexually_Abused:** Whether the shooter was sexually abused.

- **Emotionally_Abused:** Whether the shooter was emotionally abused.

- **Neglected:** Whether the shooter was neglected during childhood.

- **Bullied:** Where the shooter was bullied.

---

3. Psychological and Behavioral History

These columns provide insights into the shooter's mental health and behavioral history:

- **Mental_Illness:** Whether the shooter had a diagnosed mental illness.

- **Substance_Use:** Whether the shooter had a history of substance use.

- **History_of_Physical_Altercations:** Whether the shooter had a history of physical altercations.

- **History_of_Animal_Abuse:** Whether the shooter had a history of animal abuse.

- **History_of_Domestic_Abuse:** Whether the shooter had a history of domestic abuse.

- **History_of_Sexual_Offenses:** Whether the shooter had a history of sexual offenses.

- **Gang_Affiliation:** Whether the shooter was affiliated with a gang.

- **Terror_Group_Affiliation:** Whether the shooter was affiliated with a terrorist group.

- **Known_Hate_Group_or_Chat_Room_Affiliation_Code:** Whether the shooter was affiliated with a hate group or extremist chat room.

---

4. Recent Stressors and Crisis Indicators

These columns provide insights into the shooter's recent life circumstances and mental state:

- **Recent_or_Ongoing_Stressor_Code, Recent_or_Ongoing_Stressor:** Whether the shooter experienced recent or ongoing stressors.

- **Signs_of_Being_in_Crisis:** Whether the shooter exhibited signs of being in crisis.

- **Crisis_Six_Months_or_Less:** Whether the shooter exhibited signs of crisis within the last six months.

- **Paranoia:** Whether the shooter exhibited paranoia.

- **Rapid_Mood_Swings:** Whether the shooter exhibited rapid mood swings.

- **Isolation:** Whether the shooter was socially isolated.

- **Abusive_Behavior:** Whether the shooter exhibited abusive behavior.

- **Suicidality_Code, Suicidality:** Whether the shooter exhibited suicidal tendencies.

Determining the weights for Composite risk Score

- **Psychological and Behavioral History** (e.g., mental illness, history of violence, substance abuse) is often a stronger predictor.

  40 points

- **Recent Stressors** (e.g., crisis indicators, paranoia, suicidality) may also be highly relevant.

  30 points

- **Childhood Trauma and Abuse** can have long-term effects but may be less directly predictive than recent behaviors.

  20 points

- **Family Background** (e.g., parental divorce, criminal record) may provide context but is often less directly predictive.

   10 points

Creating the Fiream Acquistion Risk Score

We can break this down into two components:

1. **Acquisition Risk** (higher risk for illegal, stolen, or modified firearms)
2. **Firearm Risk** (higher risk for weapons with modifications or high-capacity magazines)

1. Acquisition Risk (0-3 points)

- **3 Points** → **Illegal_Purchase = 1** OR **Theft = 1** (High Risk: Black market, stolen)
- **2 Points** → **Assembled_with_Legal_Parts = 1** OR **Modified = 1** (Moderate Risk: Ghost guns, modifications)
- **1 Point** → **Gifted = 1** OR **Legal_Purchase = 1** (Lower Risk, but loopholes exist)
- **0 Points** → **Unknown = 1** OR no other factors met

2. Firearm Risk (0-2 points)

- **2 Points** → **Large_Capacity_Magazine = 1** OR **Extended_Magazine = 1** (Higher lethality)
- **1 Point** → **Modified = 1** (Gun was altered for increased firepower)
- **0 Points** → No modifications

Interpretation

- **Score 5** → Highest risk (illegally obtained, stolen, or assembled + high-capacity magazine)
- **Score 3-4** → Moderate risk (legal but modified, or illegally purchased but unmodified)
- **Score 1-2** → Lower risk (legal, no modifications)
- **Score 0** → Unknown or no risk factors

Creating Motive Risk Score

Broken down into **three weighted categories**, each capturing distinct motive types:

1. Hate-Based Motives (Highest Risk: 8 points each)

- **Racism/Xenophobia** (Motive__Racism_Xenophobia_Code = 1)

- **Religious Hate** (Motive__Religious_Hate_Code = 1)

- **Misogyny** (Motive__Misogyny = 1) → **6 points**

- **Homophobia** (Motive__Homophobia = 1) → **6 points**

- **Documented Prejudices** (Known_Prejudices not null or "No evidence") → **4 points**

2. Personal Grievances (Moderate Risk: 4–5 points each)

- **Employment Issues** (Motive__Employment_Issue = 1) → **5 points**

- **Relationship Issues** (Motive__Relationship_Issue = 1) → **5 points**

- **Interpersonal Conflict** (Motive__Interpersonal_Conflict = 1) → **4 points**

3. Other Motives (Variable Risk)

- **Fame-Seeking** (Motive__Fame_Seeking = 1) → **7 points** (high due to copycat risk)

- **Economic/Legal Issues** (Motive__Economic_Issue or Legal_Issue = 1) → **3 points each**

- **Other Documented Motives** (Motive__Other_Code = 1) → **2 points**

Risk Tiers:

Score Range	Risk Level	Intervention Focus
≥15	High Risk (Hate/Extremism)	Counter-terrorism, hate crime prevention
10–14	Moderate Risk (Grievance)	Workplace/domestic violence programs
<10	Situational/Opportunistic	General mental health screening

<mark>Creating the intervention opportunity rate</mark>

The **IOR** score is based on three main factors:

1. **Prior System Contact (0-5 points)**

   - **Law Enforcement Contact:** Known_to_Police_or_FBI, Criminal_Record

   - **Mental Health System Contact:** Prior_Hospitalization, Suicidality, Prior_Counseling, Psychiatric_Medication

   - **School or Community Interventions:** History_of_Physical_Altercations, History_of_Domestic_Abuse, Gang_Affiliation

2. **Signs of Crisis Before the Incident (0-5 points)**

   - **Behavioral Red Flags:** Signs_of_Being_in_Crisis, Crisis_Six_Months_or_Less, Paranoia, Rapid_Mood_Swings, Isolation

   - **Recent Stressors:** Recent_or_Ongoing_Stressor, Abusive_Behavior

3. **Pre-Incident Leakage (0-5 points)**

   - **Warning Signs Given:** Leakage (if leakage occurred, +3 points)

   - **Who Was Warned:** Leakage_Who (e.g., authorities vs. friends, higher points if authorities were warned)

   - **How the Leakage Occurred:** Leakage_How (direct threats or manifestos get higher weight)

Interpretation

- **Score 12-15 → Severe missed intervention** (multiple system contacts, crisis indicators, and direct leakage to authorities)

- **Score 8-11 → Moderate missed intervention** (some system contact, crisis indicators, and general leakage)

- **Score 4-7 → Limited intervention opportunities** (some warning signs but no major system contact)

- **Score 0-3 → No clear intervention opportunity**

Creating victim impact severity score:

1. Fatalities Impact (Weighted Heavily)

- Each **death** contributes **5 points**

2. Injuries Impact (Moderate Weight)

- Each **injury** contributes **2 points**

3. Years of Life Lost Impact (Weighted by Age)

- **Total Years Lost ÷ 10** (to scale appropriately)

Interpretation

- **Score 50+ → Mass casualty event** (high fatalities, young victims)
- **Score 30-49 → Severe impact** (moderate deaths + injuries + younger victims)
- **Score 15-29 → Significant event** (higher injuries than fatalities)
- **Score < 15 → Lower impact event**

This metric quantifies the **human cost per incident**, which can inform policymaking, intervention strategies, and victim support initiatives.