## **Important Note on Data Accuracy and Sourcing**

All analyses and findings in this report are based strictly on publicly available, reputable datasets and official statistics. Every claim or conclusion is supported by cited sources and data-driven reasoning.

This report uses a cleaned and revised version of the original dataset.  
 As part of the data cleaning process, some rows were removed due to missing state identifiers, and certain columns had missing values filled or revised.  
 **As a result, some states may be underrepresented or missing entirely, which could slightly skew comparisons and calculations. All metrics and summaries are calculated based on the available (cleaned) data.**

The responsibility of the analyst includes:

* Verifying all data sources are current and reliable.
* Clearly referencing each dataset or source.
* Avoiding unsupported or speculative claims.
* Correcting errors or updating findings as new/verified data becomes available.

**Disclaimer:** These findings are for educational or analytical purposes only and are not intended as policy, legal, or financial advice. Interpretations should be made in light of the noted data limitations.

## **Summary of Data Cleaning Actions**

### **1. Deleted Rows with Missing State Abbreviation or State Name**

* Rows 2 to 39 were deleted from the dataset.  
  + Reason: These rows were missing values in either the state\_abbr column, the state\_name column, or both.
  + Without a state abbreviation or name, these records could not be included in state-level analysis, as their geographic context was unknown.

### **2. rape\_legacy Column Cleaning**

* All missing values in the rape\_legacy column were filled with 0.  
  + This allows for accurate calculation and aggregation, ensuring no blanks disrupt the analysis.

### **3. rape\_revised Column Cleaning**

* All missing values in the rape\_revised column were filled with 0.  
  + This ensures every row in the dataset contains a valid numeric value for this field.

### **4. State Name Completion**

* Where state\_name was missing but state\_abbr was present, the full state name was filled in based on the abbreviation.  
  + This did not alter the dataset’s integrity, but ensured all records have a valid and readable state name.

## **How This Affects the Analysis**

* The present analysis is conducted only on the cleaned dataset described above.
* Any statistics, findings, or conclusions are based on this revised, fully-audited dataset.
* All cleaning actions are documented here for full transparency and reproducibility.

Great! Here’s a **clean citation you can paste into your analysis or README** to document the exact dataset and source used, based on your screenshot and the official description:

## **Dataset Used**

**Title:** Summary Reporting System (SRS) – Estimated Crimes  
 **Source:** FBI Crime Data Explorer  
 **URL:**[CDE](https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads) https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads **File Name:** estimated\_crimes\_1979\_2023.csv  
 **Data Type:** Estimated crimes (state and national level)  
 **Years Covered:** 1979–2023  
 **Last Modified:** September 23, 2024  
 **File Type:** CSV  
 **File Size:** 200.2 KB  
 **Description:** This dataset contains FBI-estimated crime totals for the nation and all 50 states, 1979 to the current available year. The data reflects what the FBI has traditionally included in its annual publications.

**You can also include your screenshot as an appendix or visual proof of provenance in your case study or documentation.**

Let me know if you need a full sample “Data Sources” section or anything else formalized!

