Darina Diaz

Professor Rebin Muhammad

Data 110

October 29, 2024

### **Project 1- Murders**

#### **1. Introduction**

* **Dataset Overview**: This dataset comprises data on murders across U.S. states, including each state's population, region, and total number of murders. The dataset allows us to explore the distribution of population and murder totals and identify regional patterns or correlations.
* **Purpose**: Understanding murder totals in relation to population and regional distribution can provide insights into social factors that may influence crime and inform public policy or further research.

#### **2. Exploratory Data Analysis (EDA)**

* **Summary Statistics**: We began with summary statistics to understand the range, mean, and spread of population and murder totals across states. This analysis revealed significant variation in both population sizes and murder totals, indicating a right-skewed distribution.
* **Visualizations and Key Observations**:
  + **Population and Murder Totals Distribution**:
    - **Visualizations**: Histograms with kernel density estimates (KDE) for population and total murders.
    - **Rationale**: These visualizations reveal the distribution shape, showing both population and murders are right-skewed, with most states clustered in lower ranges but a few significant outliers (highly populated states) influencing the overall average.
    - **Insights**: The right-skew indicates a concentration of lower-population states with relatively low murder counts, with outliers in highly populated areas.
  + **Regional Analysis**:
    - **Visualizations**: Boxplots for population and total murders segmented by region.
    - **Rationale**: Boxplots provide a clear view of variability and outliers within each region, helping to pinpoint which regions might have higher populations or more concentrated murder totals.
    - **Insights**: The South exhibits the highest variation in both population and murder totals, suggesting some states in this region experience more incidents, potentially due to larger populations in certain states.
  + **Relationship Between Population and Murder Totals**:
    - **Visualization**: Scatterplot with region-based color coding.
    - **Rationale**: This scatterplot helps illustrate the positive correlation between population and murder totals while allowing us to differentiate trends by region.
    - **Insights**: A strong positive relationship exists between population and murder totals across regions, confirming that more populous states tend to report higher murder counts. The South and West show a few outliers, reinforcing the need for focused policy or further analysis in these areas.

#### **3. Conclusion**

* **Summary of Findings**:
  + Population size strongly correlates with total murders, highlighting a potential scaling factor where larger populations have higher murder counts.
  + Regional differences suggest that the South has greater variability in both population and murders, indicating that certain states may need focused policy attention to manage crime rates effectively.
* **Recommendations for Further Research**:
  + A deeper analysis could examine specific factors (e.g., economic, social, or demographic data) within high-population states or regions like the South to understand influences on murder totals.
  + Further exploration of other crime metrics, if available, would offer a more comprehensive view of crime patterns across different regions and state populations.