Analysis of Income, Inflation, Expenditure, and Crime Across U.S. States

School of Information Systems

IST 652- Scripting For Data Analysis

Group1

Shivani Pandeti

Jill Karia

Reeya Tapan Patra

1) INTRODUCTION

For our IST652 final project, we set out to explore how economic conditions might relate to violent crime rates across U.S. states between 2012 and 2022. We focused on three major indicators: personal income, inflation, and personal consumption expenditure - aiming to uncover patterns or trends when compared with state-wise violent crime rates.

To do this, we brought together two different types of data: structured CSV files containing yearly data for income, inflation, and expenditure, and semi-structured JSON data retrieved from the FBI Crime Data Explorer API. Using Python, we handled the entire process - collecting, cleaning, merging, analyzing, and visualizing the data.

This project gave us the opportunity to apply core scripting techniques covered in the course, such as working with APIs, managing multiple data formats, and turning raw datasets into meaningful insights through analysis and visual storytelling.

2) Data and its Sources

This project uses **four main datasets** to explore the relationship between economic indicators and violent crime rates across U.S. states from 2012 to 2022. Three datasets were structured CSV files, and one was a semi-structured dataset retrieved through a public API.

1. **Personal Income Data**
   - **File**: SAINC1__ALL_AREAS_1929_2024.csv
   - **Source**: U.S. Bureau of Economic Analysis (BEA)
   - **Format**: CSV
   - **Link**: https://apps.bea.gov/regional/downloadzip.ht

2. **Inflation Data (Regional Price Parity Summary)**
   - **File**: SASUMMARY__ALL_AREAS_1998_2024.csv
   - **Source**: U.S. Bureau of Economic Analysis (BEA)
   - **Format**: CSV
   - **Link**: https://apps.bea.gov/regional/downloadzip.ht

3. **Personal Consumption Expenditure Data**
   - **File**: SAPCE1__ALL_AREAS_1997_2023.csv
   - **Source**: U.S. Bureau of Economic Analysis (BEA)
   - **Format**: CSV
   - **Link**: https://apps.bea.gov/regional/downloadzip.ht

4. **Violent Crime Data**
   - **Source**: FBI Crime Data Explorer API (available via Data.gov)
   - **Format**: JSON (fetched using REST API through Python scripting)
   - **Data Range**: 2012 to 2022

3) Preprocessing:

These datasets were selected to build a multi-dimensional view of how economic factors may correlate with crime rates across different U.S. states over time.

The preprocessing steps completed for this project were:

- **Income, GDP, Unemployment, and Population Data:**
  - Selected relevant indicators by filtering specific LineCodes for GDP, personal income per capita, real personal income per capita, unemployment, and population.
  - Removed national-level entries and cleaned state names by removing special characters.
  - Reshaped the data from wide to long format with State and Year structure.

- **Expenditure Data:**
  - Focused on total personal consumption expenditure and categories such as durable goods, nondurable goods, services, and nonprofit expenditures.
  - Kept only state-level records and reshaped the data into a tidy long format.

- **Inflation Data (RPPS and Unemployment):**
  - Filtered to retain the "All Items Combined" RPPS index and unemployment rates.
  - Removed non-state entries and pivoted the data to match State-Year structure.

- **Crime Data:**
  - Retrieved monthly violent and property crime rates from the FBI Crime Data Explorer API.
  - Aggregated monthly data into annual averages for each state.
  - Ensured all 50 states and Washington D.C. had complete data for 2012–2022.

- **Merging Datasets:**
  - Merged all cleaned datasets based on State and Year.
  - Calculated per capita expenditure and adjusted it for inflation using RPPS values.
- **Final Dataset:**
  - The final combined dataset contained around **561 rows** representing all states and D.C. across 11 years.
  - Included key fields such as GDP, RPPS, unemployment, personal income, population, expenditures, violent crime rates, and property crime rates.

4) Method of Analysis:

In this project, several data analysis techniques and visualization methods were applied to investigate the relationships between economic indicators and crime trends across U.S. states from 2012 to 2022.

The main steps included:

Research Question 1:
How do state-level income and inflation trends correlate with changes in overall consumer expenditure over the past decade?

- Aggregated the average per capita income and per capita expenditure over time.
- Built an interactive line plot using Plotly to compare trends across different states.
- Added a regression model (OLS) to quantify the correlation between income and expenditure.
- Displayed the regression summary showing the strong positive correlation ($R^2 \approx 0.881$).

Research Question 2:
How does spending behavior differ between high-income and low-income groups within a state when adjusted for inflation?

- Calculated per capita GDP by dividing GDP by population.
- Averaged expenditures across states and categorized by "Services", "Durable goods", and "Nondurable goods."
- Visualized the spending distribution across states using a stacked bar chart sorted by per capita GDP.
- Created a scatter plot to analyze whether richer states spent more on services relative to other expenditures.

Research Question 3:
Which states show the strongest correlation between income changes and changes in property/violent crime rates?

- Computed a "Crime Dominance Gap" representing the relative percentage difference between property and violent crime rates.
- Built an animated choropleth map displaying dominance gaps across U.S. states over 2012–2022.

Research Question 4:
Is there a significant difference in crime rates between states with high inflation and low income versus states with low inflation and high income?

- Computed correlations for each state between income and violent/property crime rates.
- Created a static choropleth map to visualize correlation strength across states.

Research Question 5:
How does spending behavior differ between high-income and low-income groups when adjusted for inflation?

- Adjusted income and expenditure data for inflation using Real Personal Income Per Capita (RPPS).
- Calculated the "Gap" between real income and real expenditure.
- Built an animated bar chart showing gap changes by state over the years.

Research Question 6:
How does the volatility of states' unemployment behavior look like over the years?

- Pivoted unemployment rates with states as rows and years as columns.
- Standardized data and applied K-Means clustering (k=3) to classify states into "Stable," "Moderate," and "Volatile" unemployment groups.
- Applied PCA (Principal Component Analysis) for dimensionality reduction.
- Visualized clusters using a 2D scatter plot with labeled clusters and ellipses representing cluster spread.

5) Python Program Overview

Python was used as the primary programming tool for this project.
The program structure followed a logical and modular flow, consisting of the following major blocks:

- **Data Loading**:
  Multiple CSV datasets were loaded, including income, expenditure, inflation, population, and crime data.

- **Data Cleaning and Preprocessing**:
  - Selected only relevant columns and economic indicators (GDP, RPPS, income, unemployment, crime rates).
  - Removed national aggregates ("United States") and standardized state names.
  - Reshaped data into **long format** for proper merging (State | Year | Metric).
  - Handled missing values and ensured complete state-year records from **2012 to 2022**.

- **Dataset Merging**:
  - Merged datasets across **common keys: State and Year**.
  - Generated final working datasets including variables like income, expenditure, crime, population, GDP, RPPS, and unemployment.

- **Exploratory Data Analysis and Visualizations**:
  - Created **interactive dashboards** (line plots, scatter plots, choropleth maps) using **Plotly** and **Seaborn**.
  - Performed **regression analysis** (OLS Linear Regression) to study correlations.

o   Generated **animated bar charts** and **cluster visualizations** for trends across states.

- **Research Question Analysis**:
  o   Answered six research questions using a combination of statistical modeling, correlation analysis, and dynamic visualizations.
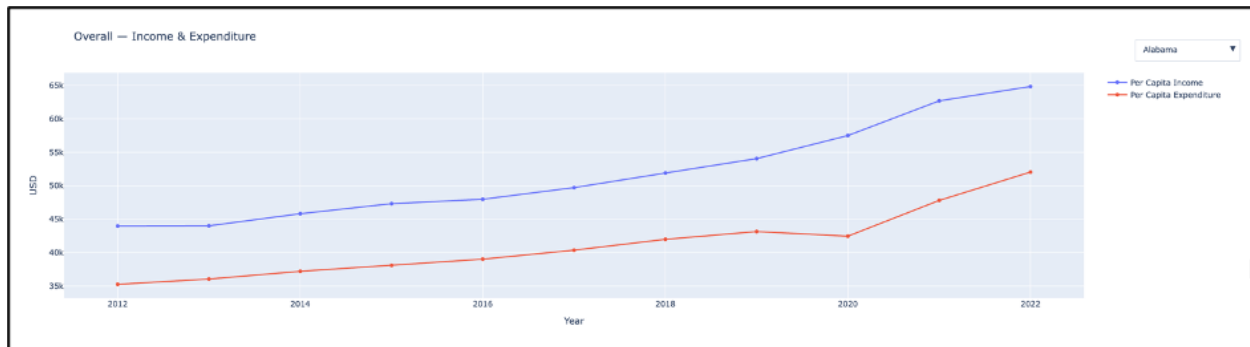
- **Output Saving**:
  Final dataset and interactive HTML dashboards were saved for presentation and analysis.

6) Program Output Documentation:

This section presents the key outputs generated from the Python program for each research question. The outputs include line plots, regression results, scatterplots, animated charts, and choropleth maps. Each figure is listed and briefly described under its corresponding research question.

Research Question 1
- Line plot of Per Capita Income and Per Capita Expenditure across states from 2012 to 2022.
- OLS Regression Results table showing the relationship between income and expenditure.
- Scatter plot with regression trendline illustrating the positive correlation.

```
                          OLS Regression Results
==============================================================================
Dep. Variable:     Per_Capita_Expenditure   R-squared:                      0.881
Model:                               OLS   Adj. R-squared:                 0.881
Method:                    Least Squares   F-statistic:                    4219.
Date:                   Mon, 28 Apr 2025   Prob (F-statistic):          1.25e-265
Time:                          01:30:21   Log-Likelihood:               -5328.5
No. Observations:                    572   AIC:                          1.066e+04
Df Residuals:                        570   BIC:                          1.067e+04
Df Model:                              1
Covariance Type:               nonrobust
==============================================================================
                     coef    std err          t      P>|t|     [0.025      0.975]
------------------------------------------------------------------------------
const            6166.6299    551.273     11.186      0.000    5083.855    7249.405
Per_Capita_Income   0.6767      0.010     64.957      0.000       0.656       0.697
==============================================================================
Omnibus:                      37.852   Durbin-Watson:                   0.592
Prob(Omnibus):                 0.000   Jarque-Bera (JB):               51.140
Skew:                          0.541   Prob(JB):                     7.85e-12
Kurtosis:                      3.987   Cond. No.                     2.59e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.59e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

📈 Regression Equation:
Per_Capita_Expenditure = 6166.63 + 0.6767 × Per_Capita_Income
R² = 0.8810
```
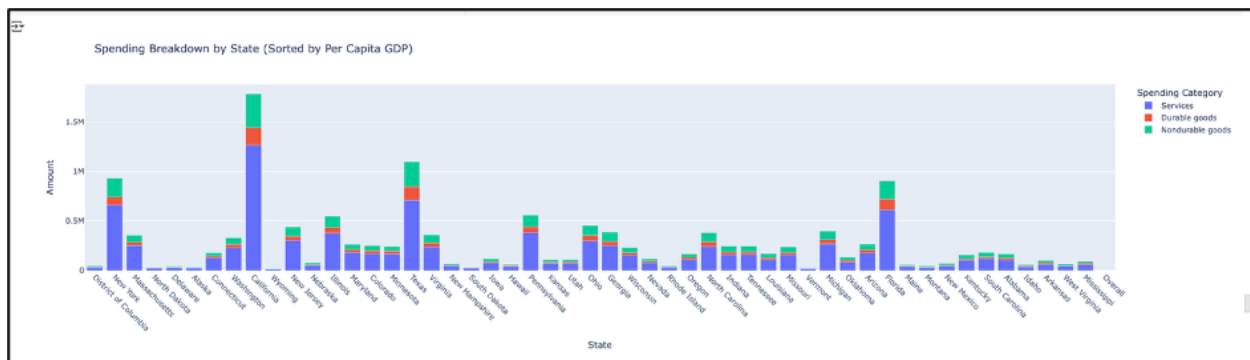

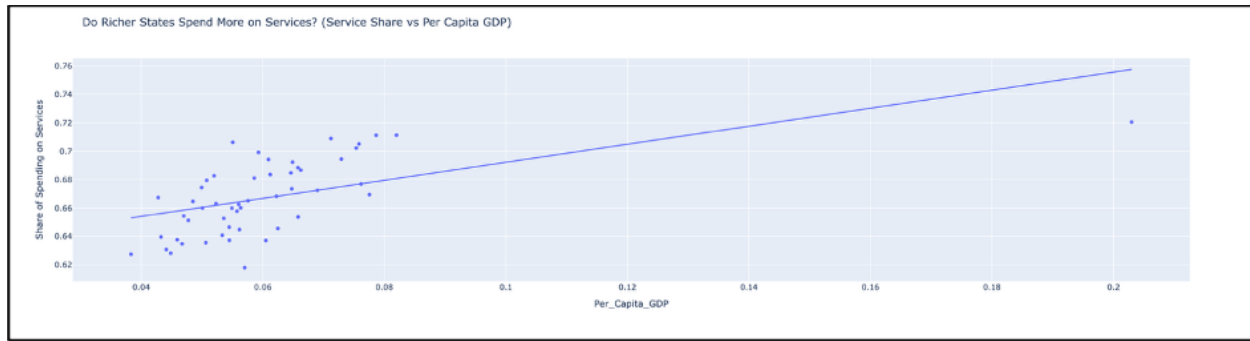Correlation Between Per Capita Income and Expenditure (With Trendline)

Research Question 2
- Stacked bar chart showing spending breakdown across states sorted by Per Capita GDP.
- Scatter plot analyzing Service Share vs Per Capita GDP.

Spending Breakdown by State (Sorted by Per Capita GDP)

Do Richer States Spend More on Services? (Service Share vs Per Capita GDP)

Research Question 3
Animated choropleth map showing property vs violent crime dominance gap over 2012–2022.

% Dominance Gap: Property vs Violent Crimes by State (2012–2022)
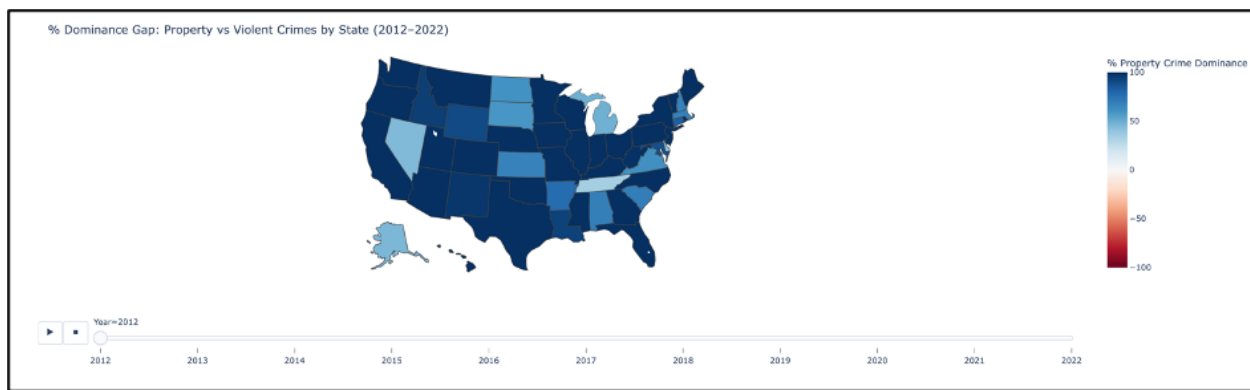
Research Question 4
Choropleth map showing the correlation between income levels and crime rates across states.

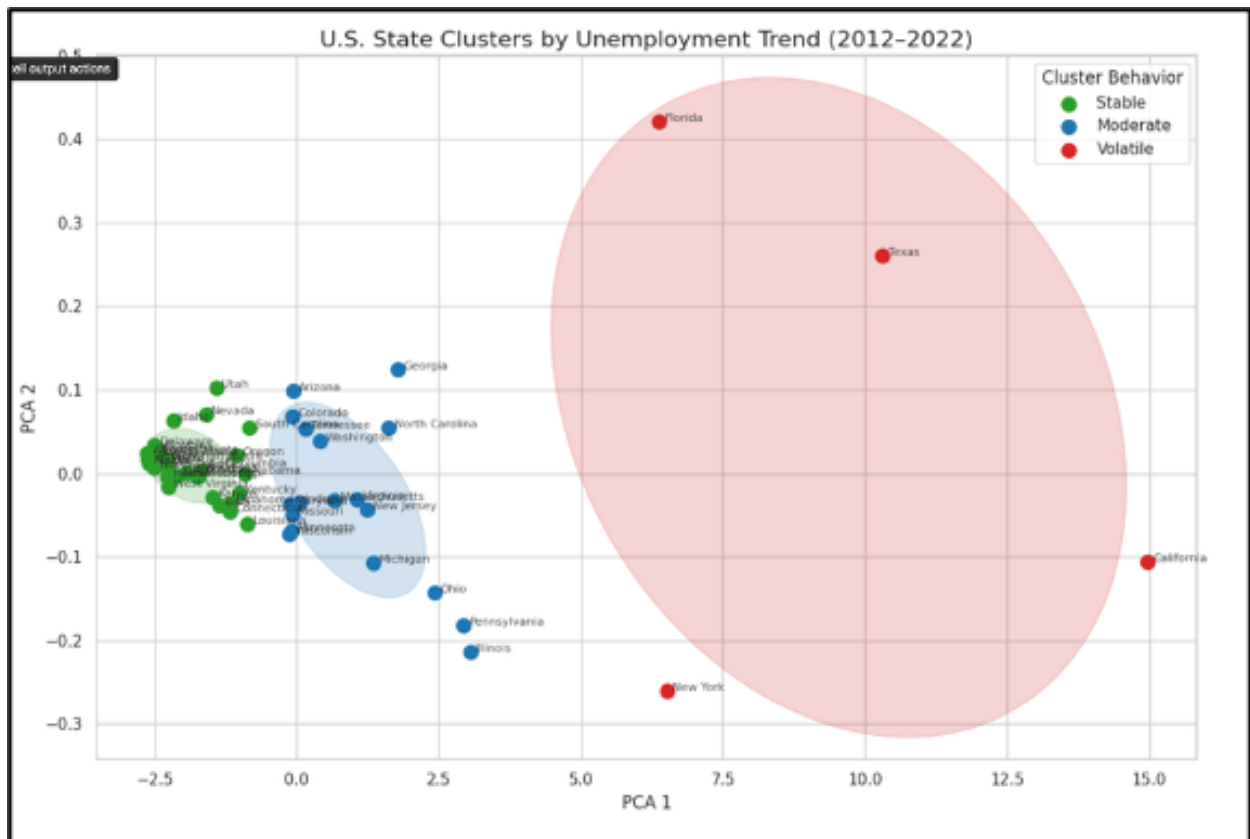Correlation Between Income and Crime Rates (2012–2022)
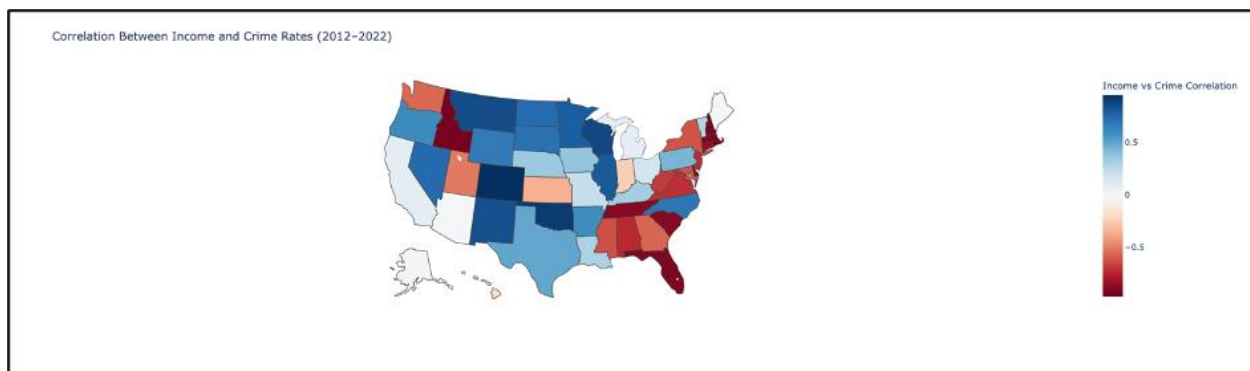
Research Question 5
Animated bar chart showing real income vs expenditure gap over time (inflation adjusted).
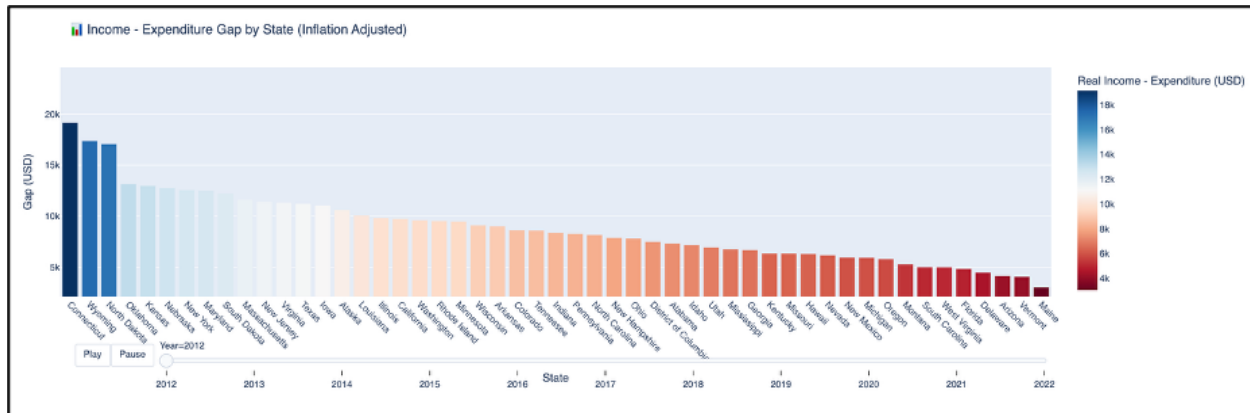
Research Question 6
PCA scatter plot clustering states based on unemployment behavior volatility (Stable, Moderate, Volatile).

7) Conclusion:

This project conducted a decade-long analysis of how economic indicators such as income, expenditure, inflation, unemployment, and crime rates interacted across U.S. states from 2012 to 2022.

Through structured data integration, visualization, and regression analysis, we uncovered meaningful patterns:
states with lower income often faced higher crime rates, inflationary pressures influenced spending behavior, and unemployment volatility highlighted economic stress in certain regions.

The findings offer valuable insights for policymakers by showing the importance of targeted fiscal support, inflation control, and region-specific strategies.
Rather than building predictive models, this analysis provided a comprehensive, data-driven overview of long-term economic trends, uncovering how systemic factors evolved and shaped social outcomes at the state level.

Overall, this work serves as a foundation for deeper exploration into economic resilience, public policy design, and social welfare planning using historical economic data.