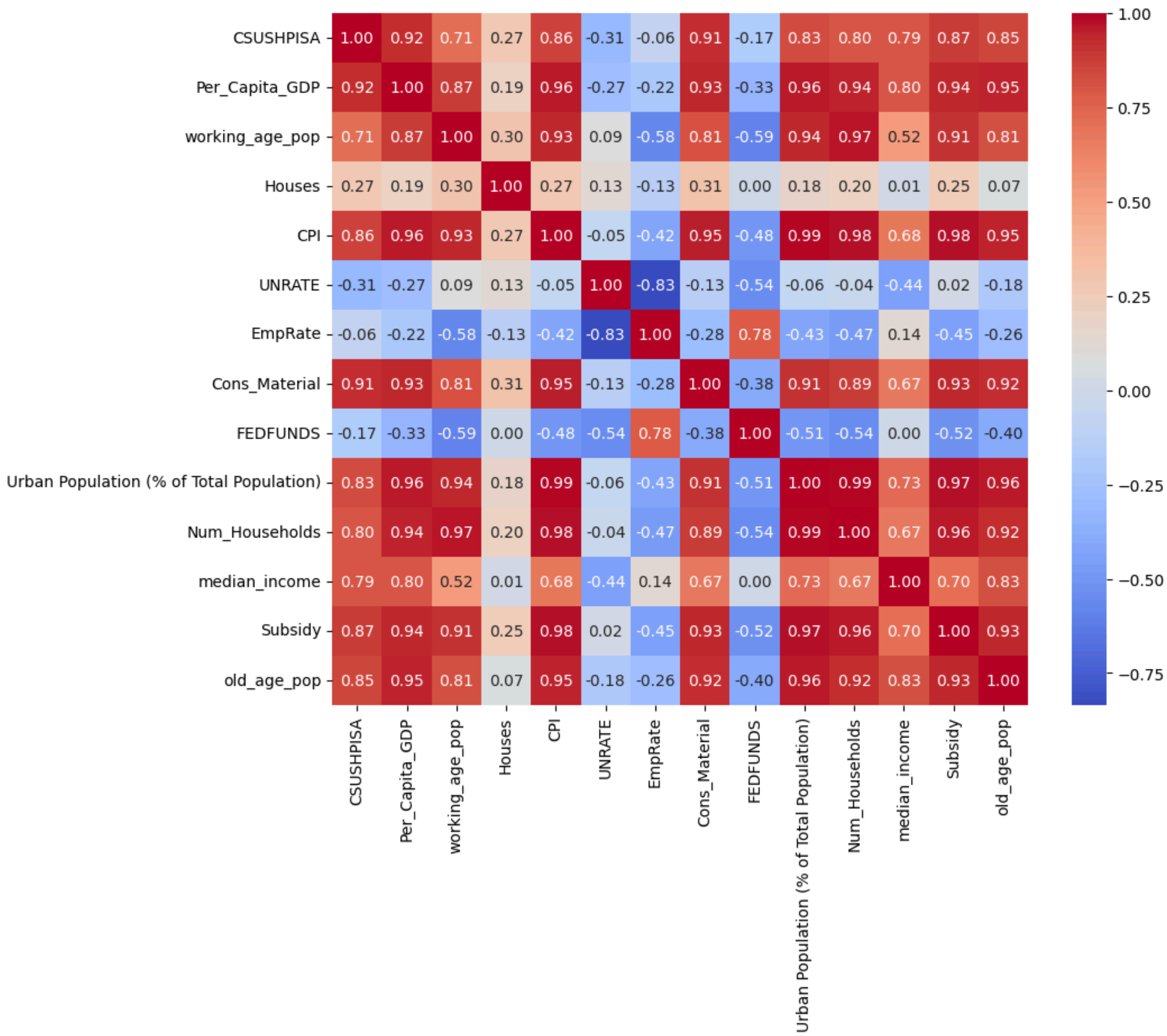
Detailed Insights from the Obtained Results

1. Correlation Heatmap



The heatmap shows the correlation coefficients between various variables.

i) Highly Correlated Variables:

- **CSUSHPISA and Per_Capita_GDP** (0.92): Suggests that higher GDP per capita is strongly associated with higher home prices.
- **CSUSHPISA and CPI** (0.86): Indicates that home prices are a major component of the CPI.
- **Per_Capita_GDP and CPI** (0.96): Reinforces the relationship seen in the pair plot, where higher economic output per person correlates with higher consumer prices.

ii) Interesting Observations:

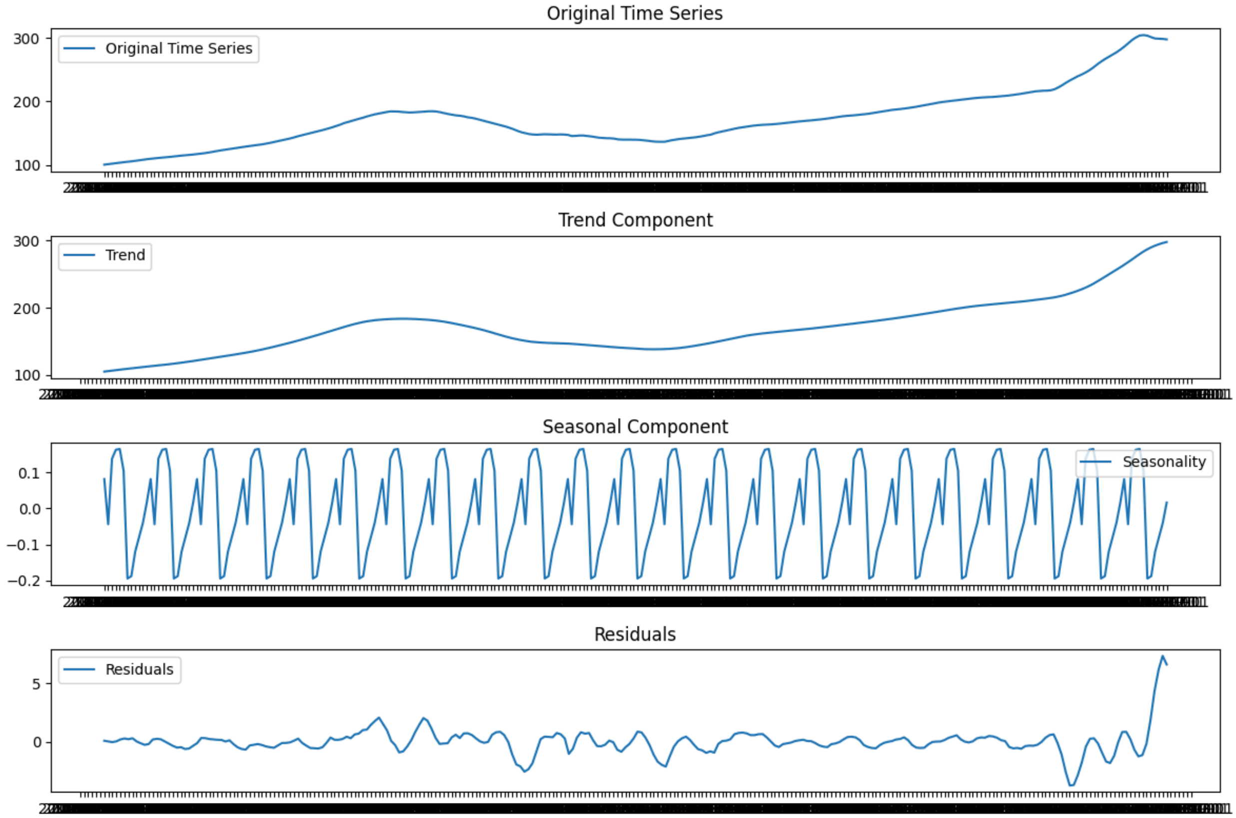
- **Urban Population**: Strongly correlated with several variables such as Per_Capita_GDP (0.96) and CPI (0.94), indicating that urbanization is linked with economic growth and price levels.
- **Subsidy**: Highly correlated with Per_Capita_GDP (0.94) and CPI (0.92), suggesting that subsidies play a significant role in economic and price dynamics.
- **Old Age Population**: Shows strong positive correlations with Per_Capita_GDP (0.87) and CPI (0.85), which could indicate the economic impact of an aging population.

iii) Negative Correlations:

- **UNRATE and EmpRate**: As expected, the unemployment rate is negatively correlated with the employment rate (-0.83), showing an inverse relationship.
- **FEDFUNDS and working_age_pop**: The Federal Funds Rate is negatively correlated with the working-age population (-0.59), potentially indicating that higher interest rates might affect employment demographics.

2. Time Series Decomposition

This decomposition splits the time series data into trend, seasonal, and residual components.



i) **Trend Component**:

- There is a clear upward trend over the period, indicating consistent growth in the variable under consideration (most likely CSUSHPISA).

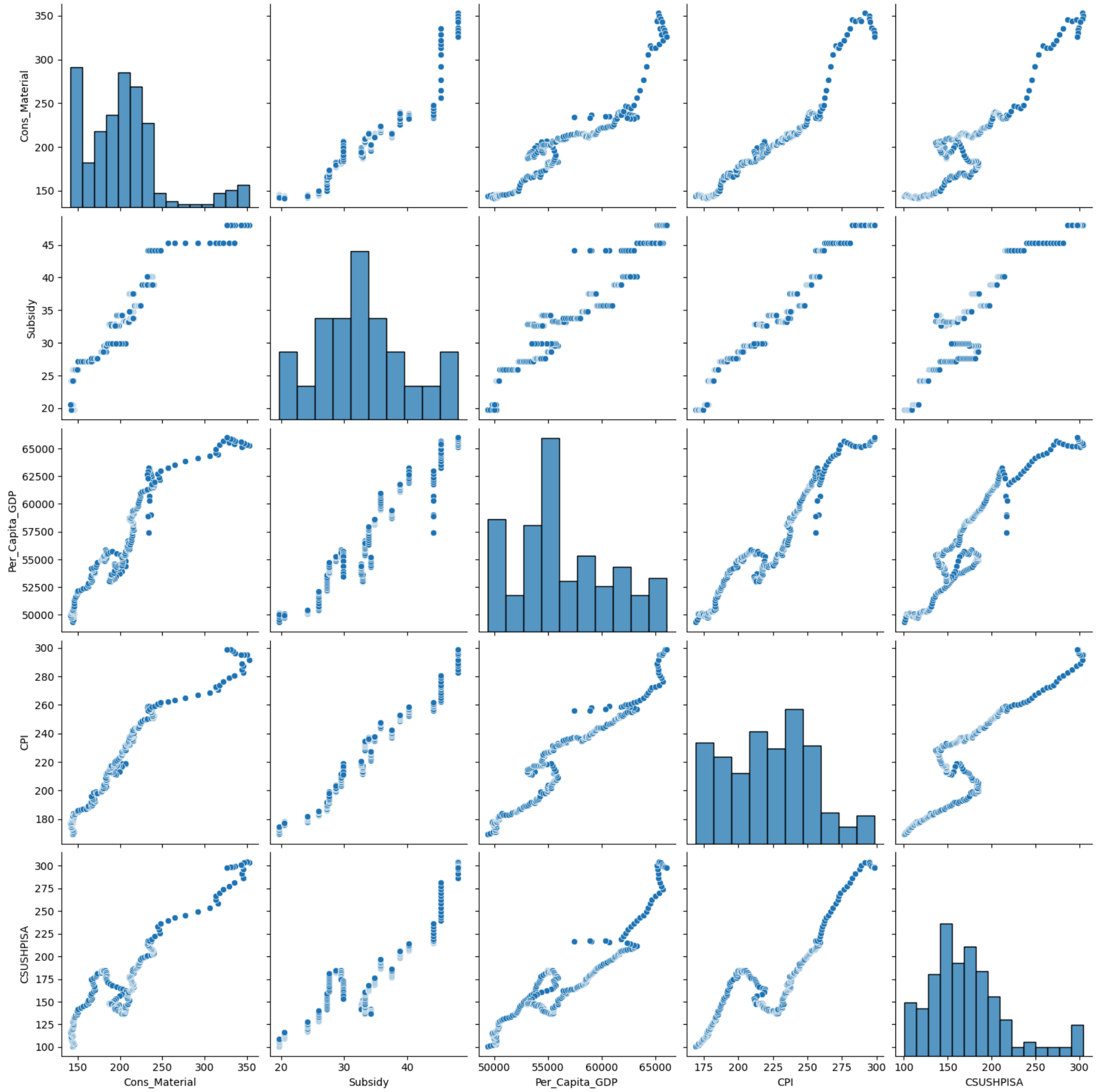
ii) **Seasonal Component**:

- The seasonality shows a regular pattern, suggesting predictable fluctuations over time. This might be due to annual cycles such as economic seasons or other regular events.

iii) **Residual Component**:

- The residuals appear to be relatively stable with some variability towards the end. The spike at the end indicates some unusual change not captured by the trend or seasonality, which could be due to an unexpected event or anomaly.

3. Pair Plot



The pair plot showcases the pairwise relationships between several variables. Here are the key insights:

i) Strong Positive Correlations:

- **Cons_Material and Per_Capita_GDP**: There is a strong positive relationship indicating that as the consumption of materials increases, the per capita GDP also increases.
- **Cons_Material and CPI**: As material consumption increases, the Consumer Price Index (CPI) also increases, showing a potential inflationary effect.
- **Cons_Material and CSUSHPISA**: There's a strong positive correlation with the S&P Case-Shiller U.S. National Home Price Index, suggesting that as material consumption increases, home prices also rise.

ii) Subsidy Relationships:

- **Subsidy and Per_Capita_GDP**: A positive correlation suggests that higher subsidies are associated with higher per capita GDP.
- **Subsidy and CPI**: Positive correlation indicating that subsidies might contribute to higher CPI levels.

iii) Other Observations:

- **Per_Capita_GDP and CPI**: Strong positive correlation, implying that higher GDP per capita is associated with higher consumer prices.
- **CSUSHPISA and CPI**: Home prices and CPI move together, showing that housing costs are a significant component of consumer price changes.

4. Histograms and Kernel Density Plots

These plots show the distribution of various features and the target variable (home prices).

Histograms and Kernel Density Plots

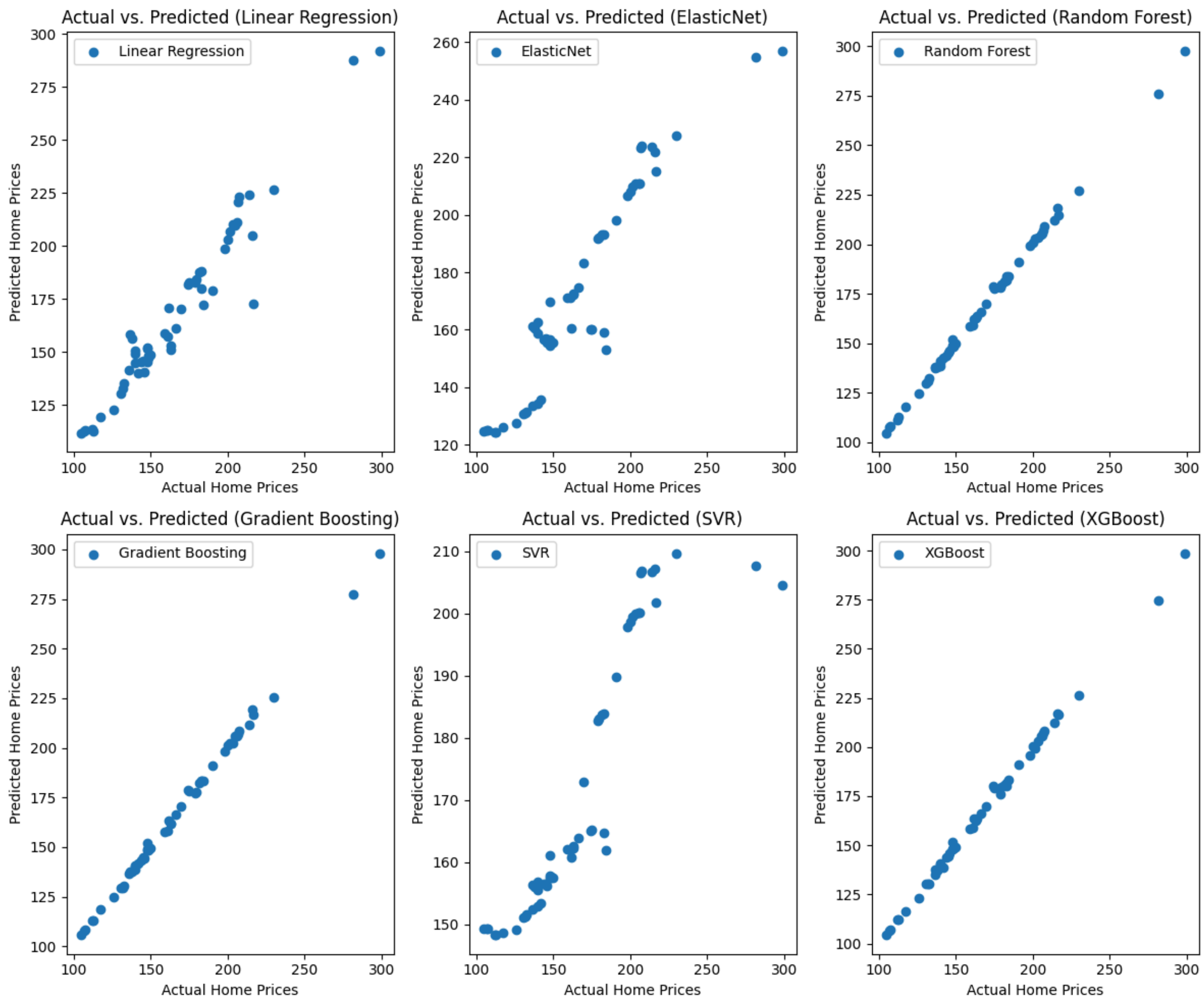
Feature Distributions:

- **CSUSHPISA (Home Prices)**:
    - The distribution shows multiple peaks and clusters, indicating variability in home prices.
    - **Insight**: Home prices have a varied distribution with several clusters, suggesting potential non-linear relationships with other features.
- **Per Capita GDP**:
    - Distribution shows a peak around 60,000 and a wider spread.
    - **Insight**: This suggests that GDP per capita has a significant range, which might influence home prices differently across its spectrum.
- **working_age_pop**:
    - Distribution shows a concentration at higher values.

    - **Insight**: The working-age population might have a substantial impact on home prices, reflecting the economic productivity of the population.
- **CPI, Cons_Material, median_income, Num_Households**:
    - These features show a varied spread and multiple peaks.
    - **Insight**: These economic indicators have complex distributions, suggesting they influence home prices in varied ways.
- **UNRATE, FEDFUNDS, EmpRate**:
    - Distributions show multiple peaks, indicating variability.
    - **Insight**: These features may have non-linear relationships with home prices, necessitating advanced modeling techniques to capture their effects.
- **Urban Population, old_age_pop, Subsidy**:
    - These features show more uniform distributions with less pronounced skewness.
    - **Insight**: These features might have a more straightforward relationship with home prices.

5. Actual vs. Predicted Home Prices

These scatter plots compare the actual home prices to the prices predicted by each model.



Model Comparisons:

- **Random Forest, Gradient Boosting, and XGBoost**:
  - The points are tightly clustered around the diagonal line (where predicted values equal actual values).
  - **Insight**: These models provide highly accurate predictions, with little deviation from actual values, demonstrating their strong predictive power.

Linear Regression:

- Points are generally close to the diagonal but with more dispersion compared to ensemble methods.

- **Insight**: Linear Regression provides fairly accurate predictions, but not as precise as the top-performing models.

ElasticNet:

- Shows significant dispersion, especially for higher home prices.
- **Insight**: ElasticNet has difficulty accurately predicting higher home prices, leading to larger errors.

SVR:

- Significant dispersion and deviation from the diagonal.
- **Insight**: SVR fails to provide accurate predictions, with large errors, especially for higher home prices.

6. Model Evaluation Metrics

This table summarizes the performance of different regression models based on Mean Squared Error (MSE) and R-squared (R^2) values.

Model Evaluation Metrics

Model	MSE	R-squared
Linear Regression	88.5	0.94
ElasticNet	200.79	0.87
Random Forest	2.17	0.998
Gradient Boosting	3.28	0.997
SVR	531.88	0.662
XGBoost	3.46	0.997

Model Performances:

- **Random Forest**:
    - **MSE: 2.17**: This indicates that the average squared difference between the predicted and actual home prices is very low, suggesting high accuracy.
    - **R-squared: 0.998**: This value is very close to 1, indicating that 99.8% of the variability in home prices is explained by the model.
    - **Insight**: Random Forest provides highly accurate predictions and explains almost all variability in the home prices, making it the best performing model.
- **Gradient Boosting**:
    - **MSE: 3.28**: Slightly higher than Random Forest but still very low.
    - **R-squared: 0.997**: Indicates that 99.7% of the variability is explained by the model.
    - **Insight**: Gradient Boosting is also highly effective, almost on par with Random Forest.
- **XGBoost**:
    - **MSE: 3.46**: Comparable to Gradient Boosting.
    - **R-squared: 0.997**: Similarly high R^2 value.
    - **Insight**: XGBoost performs similarly to Gradient Boosting, providing highly accurate predictions.
- **Linear Regression**:
    - **MSE: 88.5**: Higher than ensemble methods, indicating less accuracy.
    - **R-squared: 0.94**: Indicates that 94% of the variability is explained by the model.
    - **Insight**: While Linear Regression performs reasonably well, it is not as precise as ensemble methods.
- **ElasticNet**:
    - **MSE: 200.79**: Significantly higher, indicating poorer predictions.
    - **R-squared: 0.87**: Explains 87% of the variability.

    - **Insight**: ElasticNet struggles to capture the relationships in the data as effectively, resulting in lower accuracy.
- **Support Vector Regression (SVR)**:
    - **MSE: 531.88**: The highest MSE, indicating the poorest accuracy.
    - **R-squared: 0.662**: Explains only 66.2% of the variability.
    - **Insight**: SVR performs the worst among the evaluated models, indicating it is not suitable for this dataset.

Summary

- The pair plot and heatmap indicate strong positive relationships between material consumption, GDP per capita, CPI, and home prices.
- Subsidies appear to be positively correlated with economic growth and consumer prices.
- The time series decomposition highlights a clear upward trend and regular seasonal patterns, with some residual variability towards the end.
- Urbanization and aging populations have significant impacts on economic indicators like GDP and CPI.
- Negative correlations such as between the unemployment rate and employment rate, and between interest rates and working-age population, highlight important inverse relationships in the data.