# Linear Regression with Backward Selection to predict U.S. Unemployment Rate in December 2024

**Presented by Shapley Powerpuff**

Aigul, Feby, Kaylie, Olivia, Swetha

**December 11, 2024**

# Research Outline

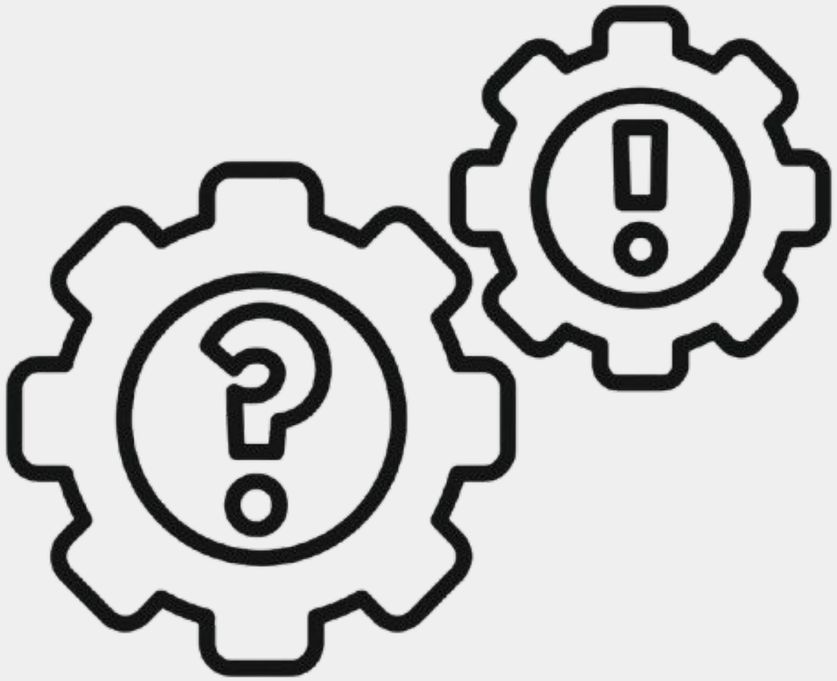
01 Research Problem and Analysis Scope

02 Descriptive Summary

03 Methodology (Model Overview and Data Pre-processing)

04 Interpretation of Model Result and Goodness-of-Fit

05 Conclusion and Prediction



# 01 Research Problem and Analysis Scope

# Research Problem and Analysis Scope

Unpredictable fluctuations in U.S. employment rate (U-3) create challenges for policymakers to design labor market policies, allocate resources and implement timely interventions. This unpredictable changes has an impact on national economic stability and public confidence.

By using **linear regression with backward selection** on lagged macroeconomic variables, we will predict U-3 Unemployment Rate for December 2024 to help policymakers make prediction and identify factors that affect the unemployment rate.

# Literature Review

## Key Findings

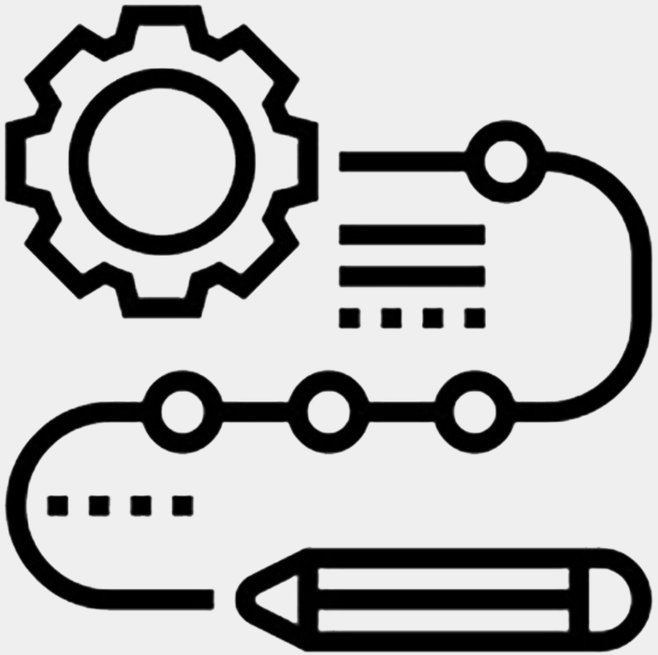
- **Collins (2009),** conducted a statistical analysis to identify key factors influencing unemployment rates, highlighting the importance of demographic and economic variables. Explored various regression techniques for predicting unemployment rates, including forward selection, backward elimination, and stepwise selection. Emphasized the use of R-squared values and residual plots to assess model fit and diagnose potential issues.
- **Wolf-Powers (2013),** found unemployment strongly linked to public capital investment showing limited effects.
- **Blanchflower, Bryson (2022),** highlighted lagged consumer expectations as strong predictors of unemployment trends 6–18 months ahead, outperforming traditional forecasting methods.
- **Capistrano (2023)**, used dummy variables for years 2014-2024 in a regression model to account for time-specific effects on unemployment rates.

## Our Approach

Linear Regression with backward selection and lagged variables on macroeconomic indicators.

**Gaps Addressed by Our Model:**

- Employs **backward elimination** to refine predictors and improve regression model accuracy.
- Incorporates **lagged variables** for key predictors to handle autocorrelation.
- Create **dummy variables** for years to capture yearly effects.
- Trained on more **recent data.**



# 02 Descriptive Summary

# Data Overview

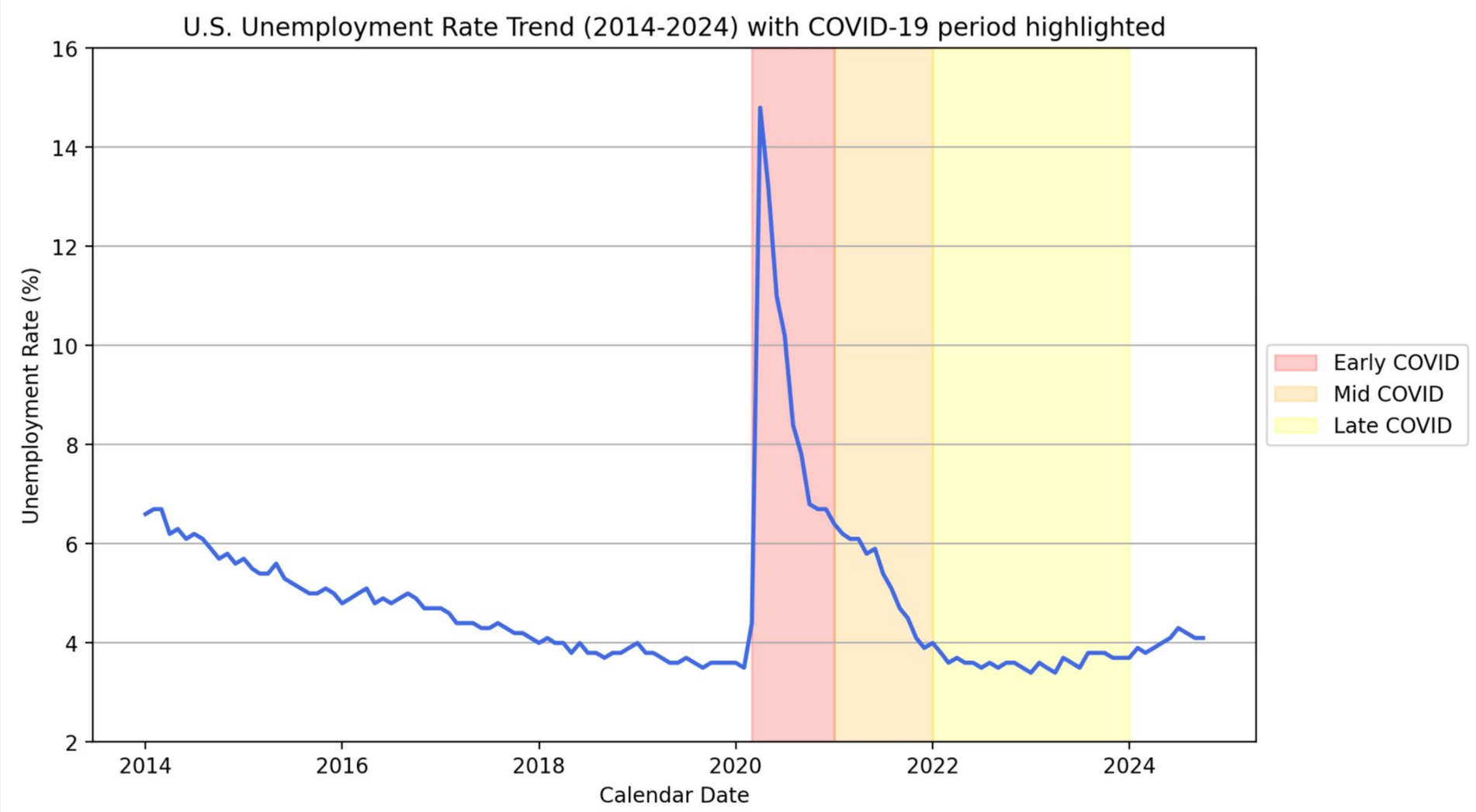
| *Key Variables* | Description | Unit | Variable Significance |
|---|---|---|---|
| UNRATE | Unemployment rate | % | Primary indicator of labor market health |
| GDPC1 | Real GDP | USD | Shows total economic output with inflation adjustment |
| BBKMGDP | Brave-Butters-Kelley Real Gross Domestic Product | % | Economic measure that track monthly growth of the U.S economy |
| ICT_INVESTMENT | Industrial and Commercial Property Investment | USD | Shows business sector growth |
| IP_INVESTMENT | Intellectual Property Investments | USD | Indicates knowledge economy development |
| INFLATION_ADJ | Adjusted Inflation | % | Used for trend analysis |
| INFLATION_NOT_ADJ | Raw inflation | % | Used for trend analysis |
| CPIAUCSL | Consumer Price Index | Index | Measures consumer price changes |
| FEDFUNDS | Federal Funds Rate | % | Central bank policy tool |
| LFPR | Labor Force Participation Rate | % | Shows overall workforce participation |

The data retrieved from different data sources and covered monthly records between January 2014 and October 2024 with 10 variables spanning across 11 years.

**Data source:**

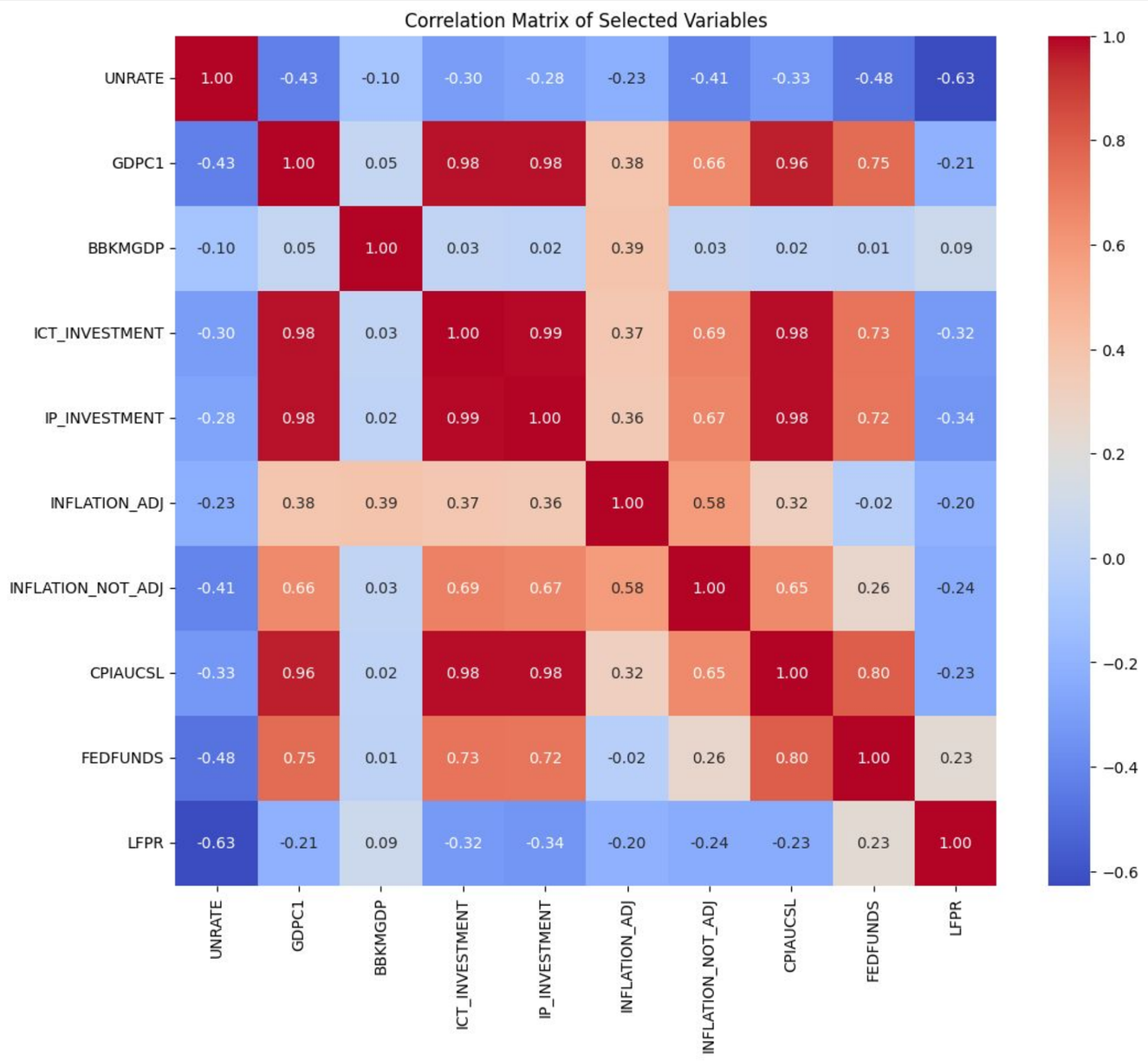
- Unemployment rate, U.S. Bureau of Labor Statistics
- Labor Force Participation rate, U.S. Bureau of Labor Statistics
- Consumer Price Index, U.S. Bureau of Labor Statistics
- Federal Reserve Interest Rate, FRED Economic Data
- Gross Domestic Product, FRED Economic Data
- Private Fixed Investment, FRED Economic Data

**A steady decline in unemployment rate before 2020, a sharp spike during the COVID-19 pandemic, and a rapid recovery followed by stabilization, reflecting the labor market's resilience and recovery post-crisis.**




*World Health Organization CDC Covid Timeline reference: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline*

# Correlation within variables




### Economic Growth vs. Unemployment

Higher GDP and investments (ICT, IP) are strongly tied to lower unemployment rates.

### Inflation vs. Monetary Policy

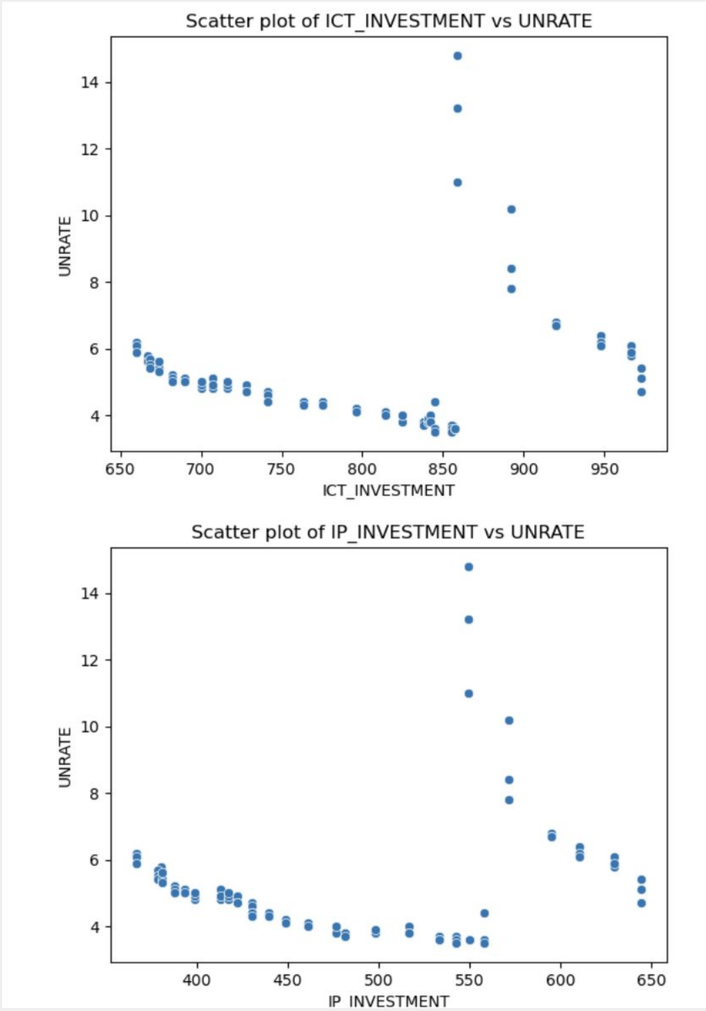
Inflation (adjusted and non adjusted) measures are negatively correlated with interest rates, reflecting monetary policy's role.

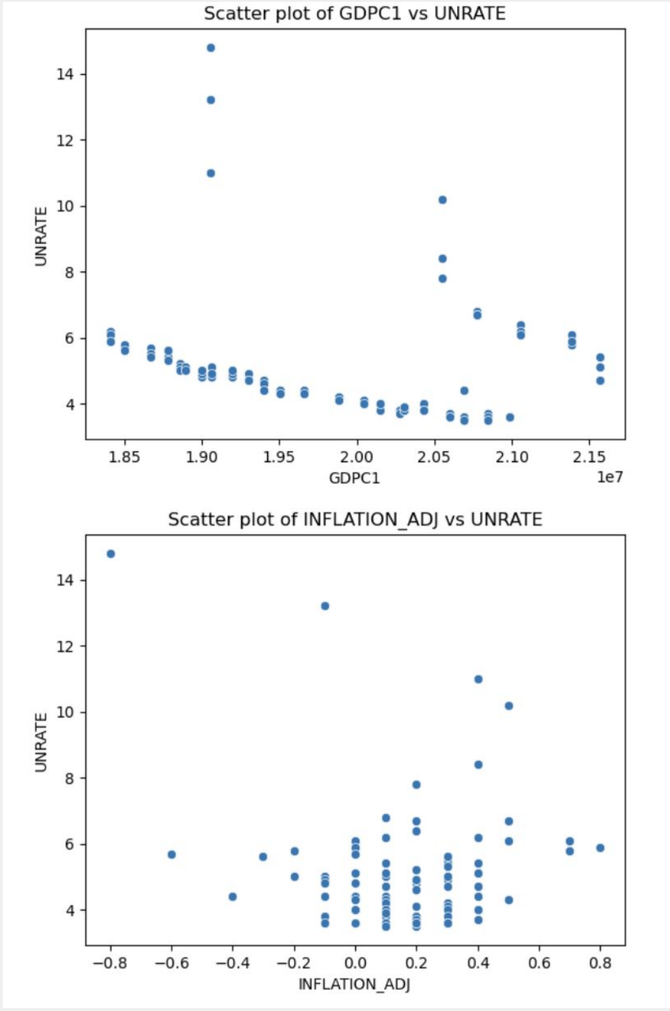
### Labor Force Participation

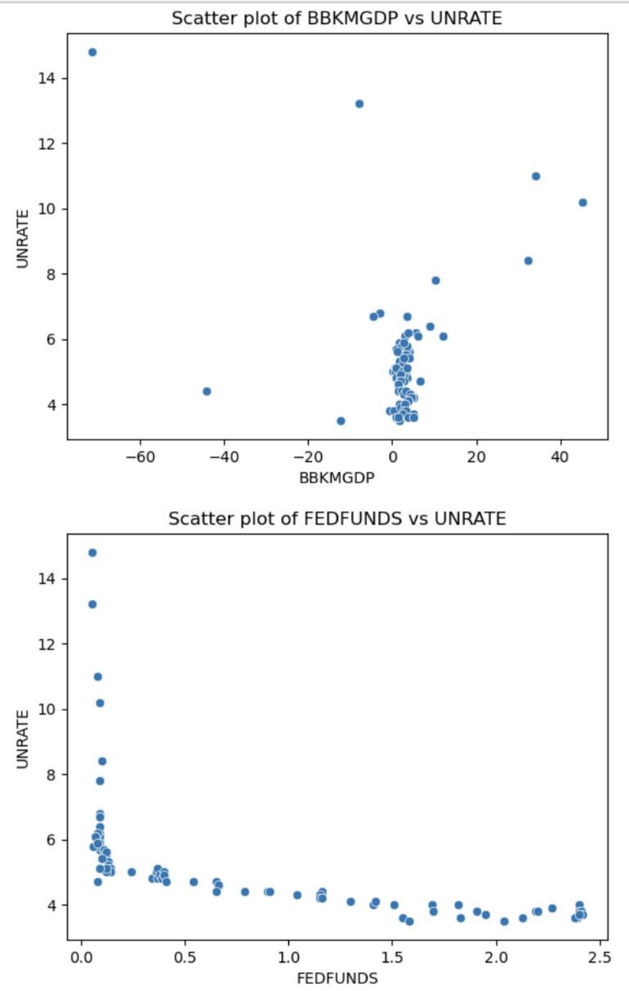
Participation rates (overall and by demographic) consistently show strong negative correlations with unemployment, underlining their importance in economic recovery.

# Linearity Check

## Generally, we observe linearity between dependent and independent variables

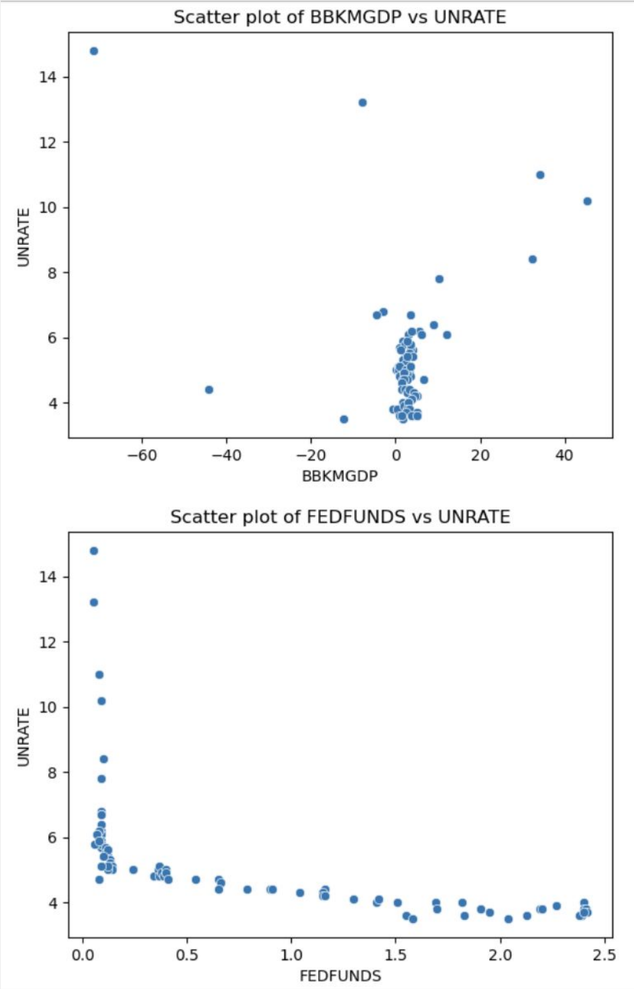


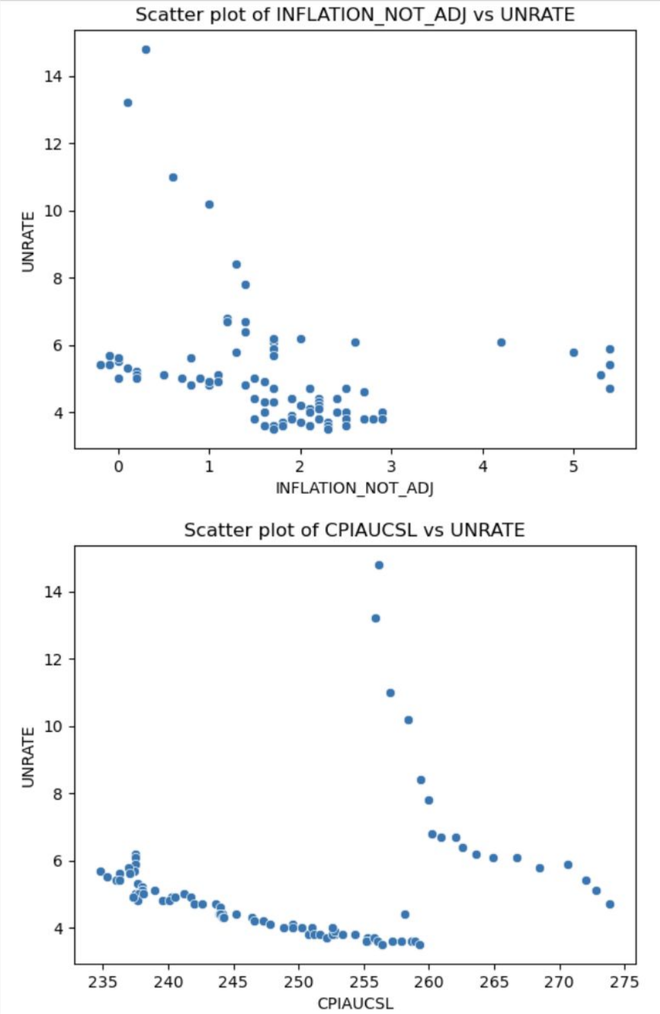



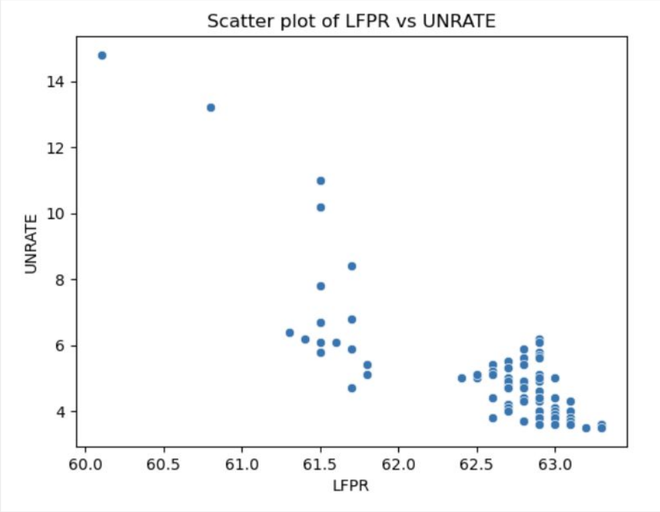



# Linearity Check

## Generally, we observe linearity between dependent and independent variables

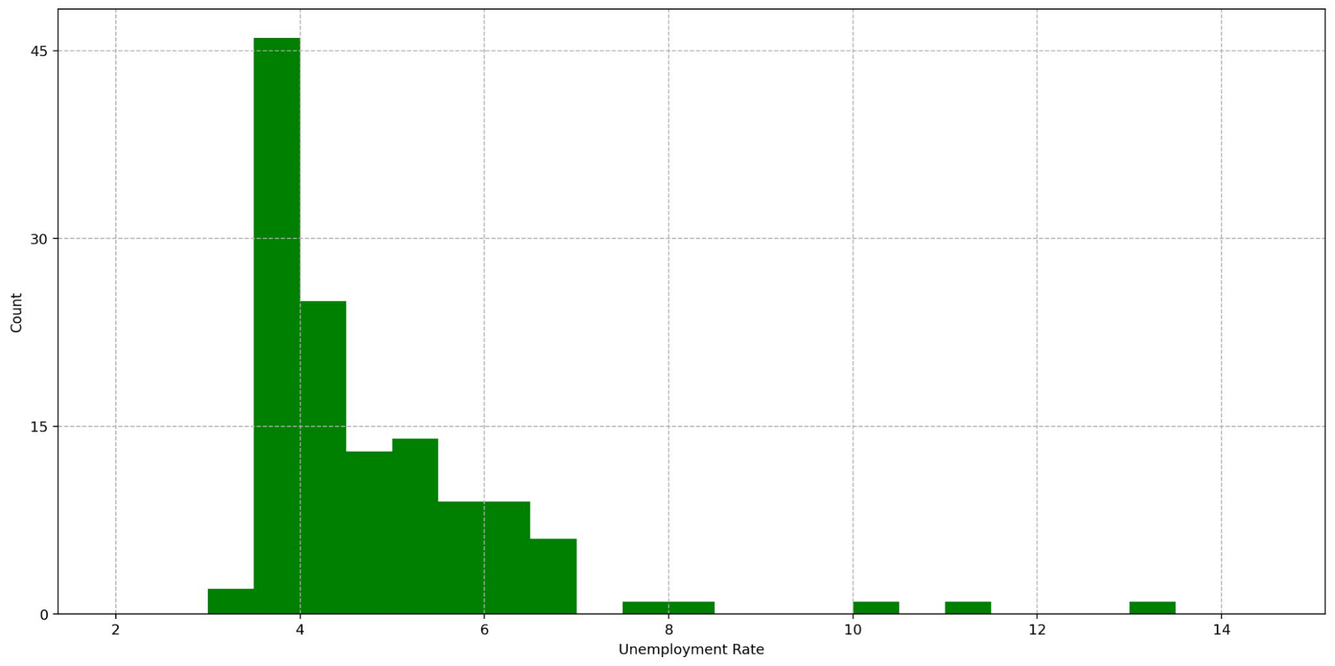


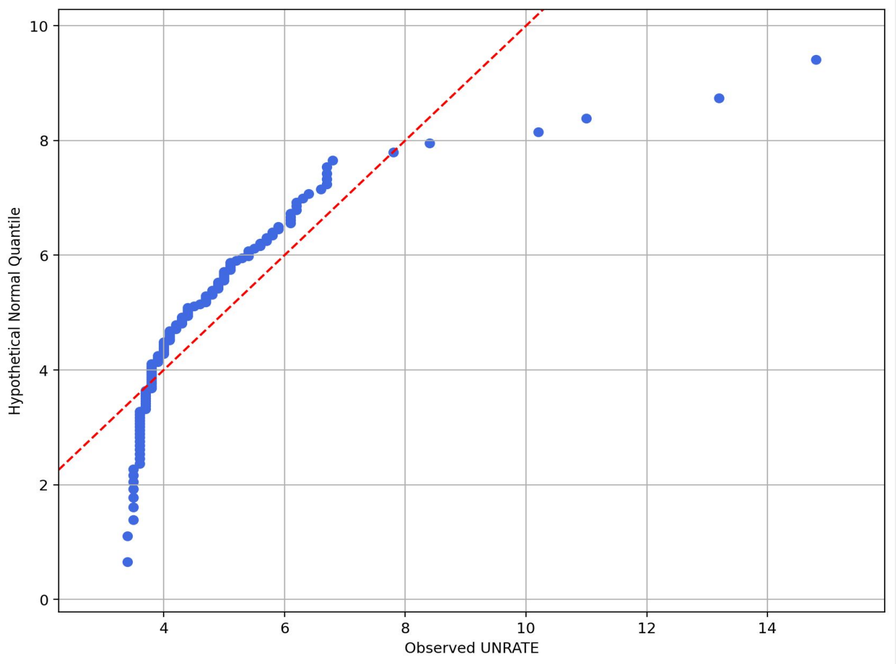



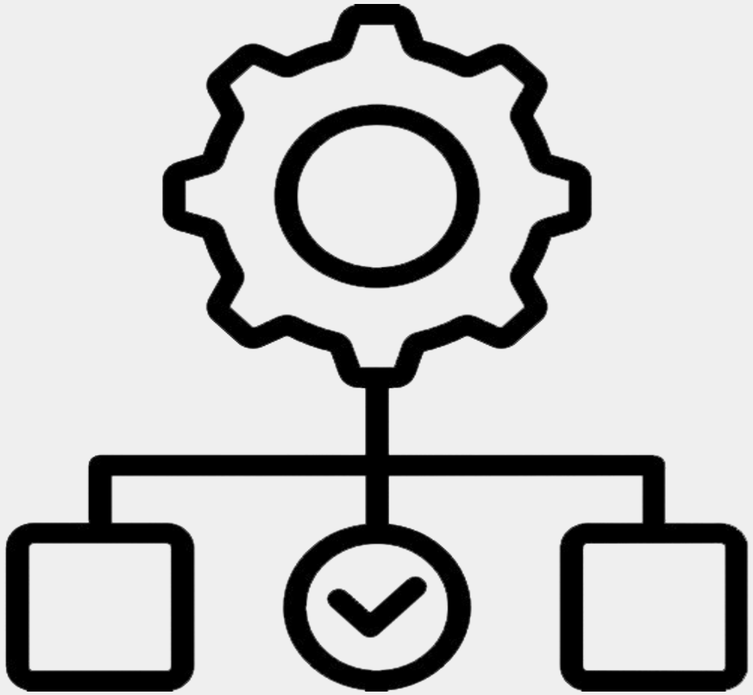
# Normality

Unemployment rate distribution is **skewed to the right**, based on the QQ plot, the observed number deviate the hypothetical quantile, especially in the lower quantile. Shapiro & Anderson tests also support our argument



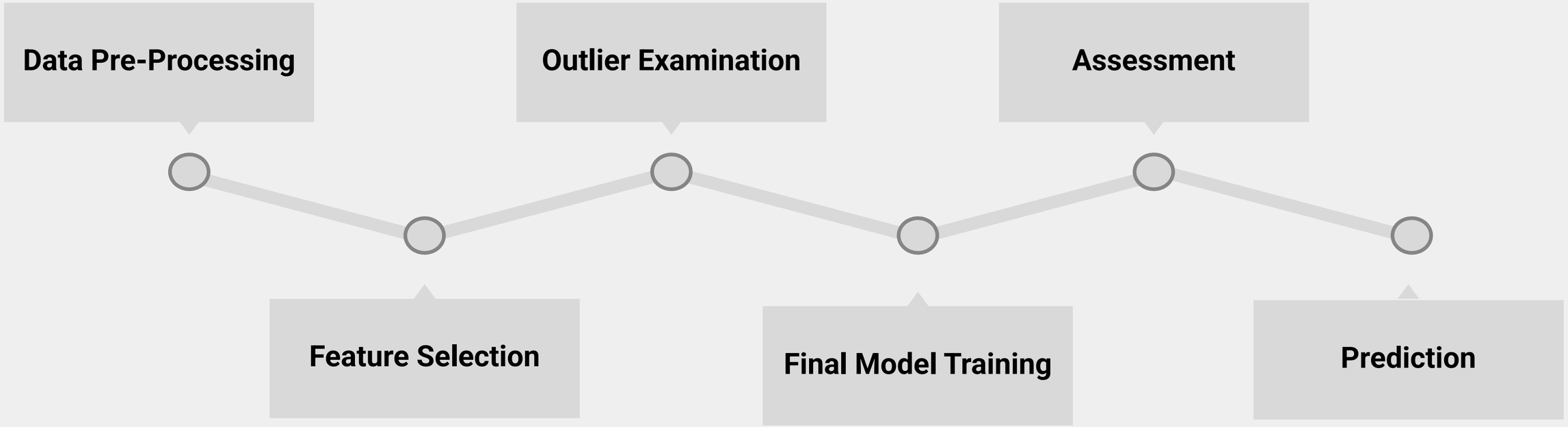




```
Shapiro Test Before Transformation =  0.6834465691739086
p-value =  2.2589890527869732e-15
Anderson Test Before Transformation =  9.449848683776622
Critical Values =  [0.56  0.637 0.765 0.892 1.061], p-values =  [0.15  0.1   0.05  0.025 0.01 ]
```



# 03 Methodology

# Model Overview



# Data Pre-Processing and Feature Selection

## Data Pre-Processing

- Data with quarterly frequency is assume to be same for each month within its quarter to **ensure consistency in frequencies** with the monthly data.
- Creating **dummy variables for the year** with 2024 as reference year, to help capture unique effects of each year on the dependent variable.
- **Box-Cox Transformation** to adjust normality of unemployment rate.

## Feature Selection

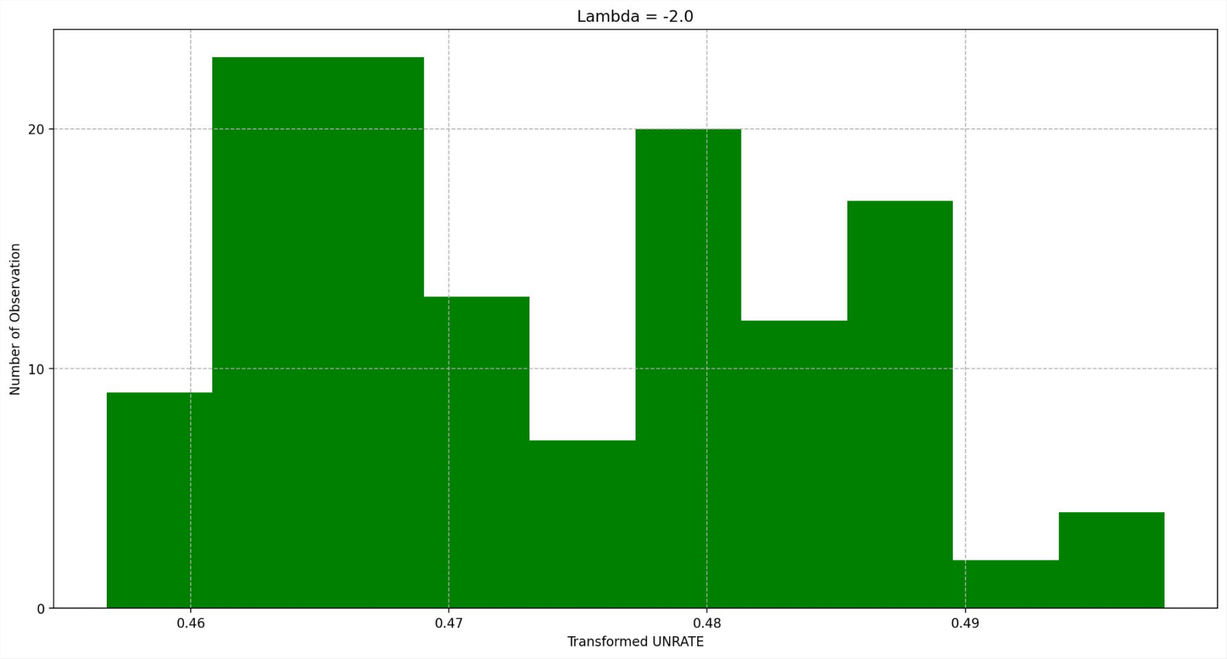
- Based on the cross correlation plots, unemployment rate is highly correlated with other variables at lag 0, 1, and 2, suggesting lagged effects of the potential predictors on unemployment. (*)
  - ➔ Taking the **lagged values** of the variables (Lag = **2 months**) as our data is only available up to October 2024- two months before the desired prediction, without compromising on the model's quality.

- **Backward selection** to select the predictors that best optimize model's performance.

**Detailed plots can be found in the Appendix*

# Box-Cox Transformation

We applied the Box-Cox transformation to obtain the transformed UNRATE. Based on Shapiro-Wilk and Anderson test, although the Unemployment rate still does not follow a normal distribution, **the transformation made our data closer to a normal distribution.**




```
Shapiro Test After Transformation =
0.9485281965740976
p-value =  8.750301391717689e-05

Anderson Test After Transformation=
2.201062406563267

Critical Values =
[0.56  0.637 0.765 0.892 1.061]
p-values =
[0.15  0.1   0.05  0.025 0.01 ]
```

# Backward Selection Summary

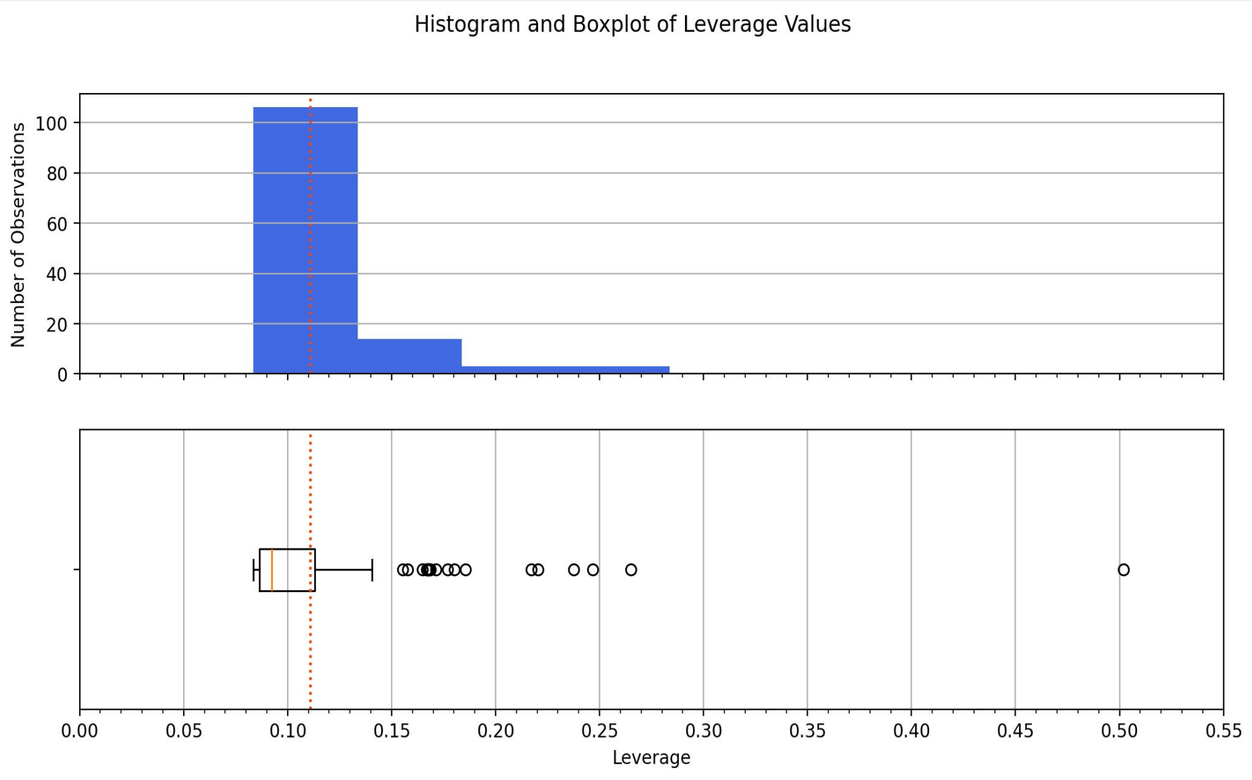
| | Step | Variable Removed | Residual Sum of Squares | N Non-Aliased Parameters | F Stat | F DF1 | F DF2 | F Sig |
|---|---|---|---|---|---|---|---|---|
| **0** | 0 | None | 0.000705 | 18 | NaN | NaN | NaN | NaN |
| **1** | 1 | LFPR | 0.000705 | 17 | 0.006904 | 1.0 | 108.0 | 0.933936 |
| **2** | 2 | GDPC1 | 0.000705 | 16 | 0.013699 | 1.0 | 109.0 | 0.907041 |
| **3** | 3 | INFLATION_ADJ | 0.000706 | 15 | 0.013295 | 1.0 | 110.0 | 0.908414 |
| **4** | 4 | FEDFUNDS | 0.000708 | 14 | 0.457556 | 1.0 | 111.0 | 0.500175 |

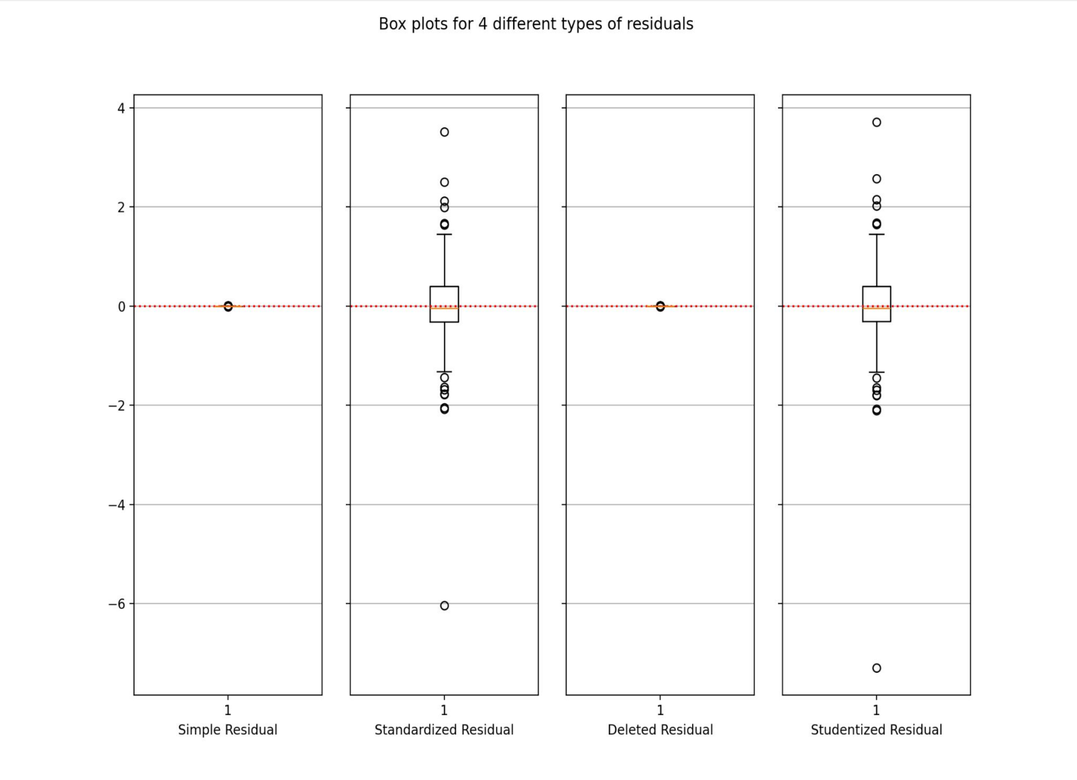
The final predictors in the model are:

- IP_INVESTMENT
- INFLATION_NOT_ADJ
- BBKMGDP
- Dummies for the year

# Outlier & High Influential Observations

Based on leverage & residual, we dropped **four (4) data points** that are considered as outliers and high leverages (mainly data from 2019)



# 04 Model Result and Goodness-of-Fit

# Model Specification

The linear regression models denoted by

$$UNRATE_t = X_{t-2}\beta + \varepsilon$$

Where

- y is 127x1 observation (dependent variable) matrix
- X is 127x14 regressor variables matrix, including variables IP_INVESTMENT, INFLATION_NOT_ADJ, BBKMGDP, and the dummies for the year.
- β is 14x1 parameter regression vector
- ε is 127x1 random error vector
- and $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 \mathbf{I}_n)$ is i.i.d

## Assumptions

- Normality
- Linearity
- Homoscedasticity
- No Multicollinearity
- No outlier

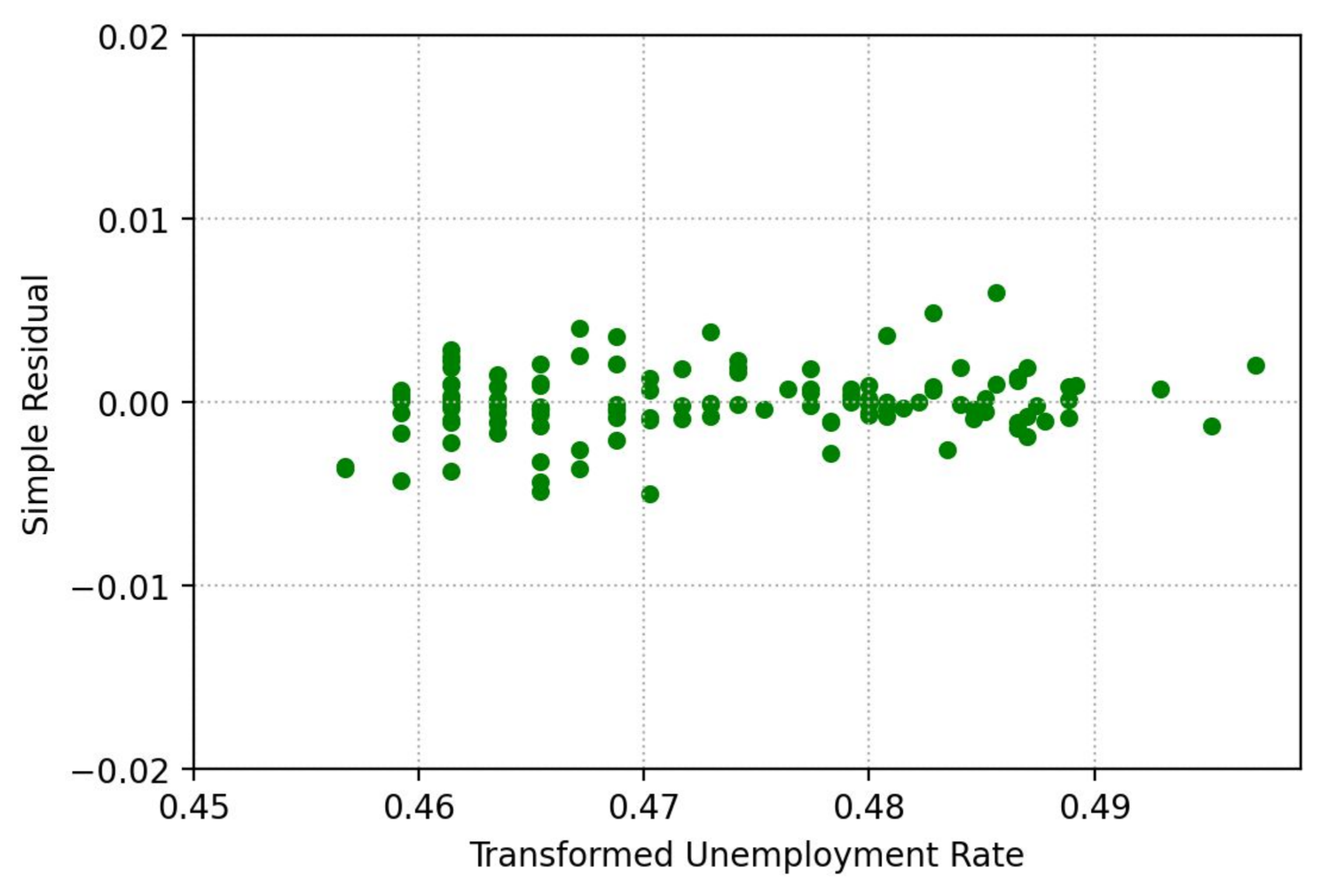
# Model Result and Interpretation

| | Estimate | Standard Error | t | Significance | Lower 95 CI | Upper 95 CI |
|---|---|---|---|---|---|---|
| **Intercept** | 0.563187 | 0.011145 | 50.530891 | 1.793693e-77 | 0.541097 | 0.585277 |
| **IP_INVESTMENT** | -0.000108 | 0.000014 | -7.625813 | 9.648672e-12 | -0.000136 | -0.000080 |
| **INFLATION_NOT_ADJ** | -0.002295 | 0.000245 | -9.354998 | 1.266775e-15 | -0.002781 | -0.001809 |
| **BBKMGDP** | -0.000076 | 0.000028 | -2.663272 | 8.911396e-03 | -0.000132 | -0.000019 |
| **2014** | -0.033936 | 0.006121 | -5.544324 | 2.079506e-07 | -0.046067 | -0.021804 |
| **2015** | -0.040083 | 0.005768 | -6.949381 | 2.812925e-10 | -0.051514 | -0.028651 |
| **2016** | -0.036866 | 0.005424 | -6.797219 | 5.920613e-10 | -0.047616 | -0.026116 |
| **2017** | -0.038490 | 0.005126 | -7.509548 | 1.734315e-11 | -0.048649 | -0.028332 |
| **2018** | -0.039130 | 0.004557 | -8.586037 | 7.016519e-14 | -0.048162 | -0.030097 |
| **2019** | -0.039168 | 0.003757 | -10.425981 | 4.524246e-18 | -0.046614 | -0.031722 |
| **2020** | -0.007714 | 0.003249 | -2.374265 | 1.933284e-02 | -0.014154 | -0.001275 |
| **2021** | -0.005431 | 0.002566 | -2.116385 | 3.658815e-02 | -0.010517 | -0.000345 |
| **2022** | -0.006880 | 0.002088 | -3.294379 | 1.330994e-03 | -0.011019 | -0.002741 |
| **2023** | -0.007963 | 0.001086 | -7.335356 | 4.154744e-11 | -0.010115 | -0.005812 |

Variables like IP_INVESTMENT, INFLATION_NOT_ADJ, and BBKMGDP have negative coefficients, suggesting an inverse relationship with unemployment rate.
For example, as INFLATION_NOT_ADJ increases, unemployment is expected to decrease, holding other variables constant. (Consistent with economic theory like Philip Curve)

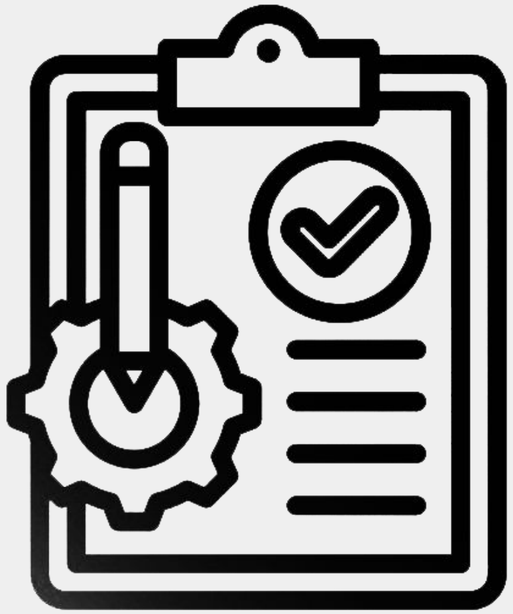
The coefficients of the yearly dummies confirm observed patterns regarding Unemployment rate before and after COVID-19

# Simple Residual



# Goodness-of-Fit

| Metric | Value | Interpretation |
|---|---|---|
| **R-squared** | 0.9663 | Explains 96.6% of the variability in unemployment rate predictions. |
| **MSE** | 3.378 | Shows the average squared difference between predicted and observed values. |
| **RMSE** | 0.0024 | Indicates the standard deviation of prediction errors. |
| **MAE** | 0.0013 | Represents the average absolute difference between predicted and actual values. |
| **Pearson Correlation** (between observed and predicted values) | 0.983 | Strong alignment between observed and predicted unemployment rates. |



# 05 Conclusion and Prediction

# Conclusion

We have applied the concepts and model learnt in this course to **create a linear regression model for prediction with high accuracy**. Below is the summary of the steps:

- Conducted exploratory data analysis to understand relationships between variables and identify potential transformations.
- Applied a Box-Cox transformation to normalize the dependent variable and improve model assumptions.
- Used backward selection to identify the most significant predictors and simplify the model.
- Built a linear regression model to predict the US unemployment rate.
- Verified model assumptions (e.g., linearity, homoscedasticity, and normality of residuals), performed model diagnostics, and assessed goodness-of-fit to ensure robustness and reliability.

**Limitation and Future Improvement of the model**

We focused on the accuracy of prediction and we believe there is a room for improvement on model interpretability, consider treating the multicollinearity using any other method like Ridge Regression, PCA in the future.

# The Unemployment rate of December 2024

Our Linear Regression with Backward Selection Model predicted the U.S. Unemployment Rate for December 2024 is **4.14%**

The 95% Confidence Interval is [4.0227%, 4.2696%]

# References


1. C. Lee, (2021) CIN (Computers, Informatics, Nursing) journal, published by Lippincott Williams & Wilkins, https://cdn-links.lww.com/permalink/cin/a/cin_2021_09_20_lee_cin-d-21-00253_sdc3.pdf
2. R. Collins (2009), "Factors related to the unemployment rate: A statistical analysis", published by University of Northern Iowa
3. L. Wolf-Powers (2013), "Predictors of Employment Growth and Unemployment in U.S. Central Cities, 1990-2010", published by University of Pennsylvania
4. D. Blanchflower, A. Bryson (2022), "The Economics of Walking About and Predicting Unemployment in the USA", published by Cambridge University Press
5. I. Kalish, R.Gibbard (2024), "United States Economic Forecast", published by Deloitte
6. T. Burton, G. Ehrlich, K.Henson (20240, "The U.S. Economic Outlook for 2024–2026", published by University of Michigan
7. M. Capistrano, (2023), "Exploring the Drivers of Unemployment and Forecasting the Unemployment Rate: A Time Series and Regression Analysis", published by International Journal of Research Publication and Reviews

06

# Response of Proffesor Feedback

# Is Our Model Useless?

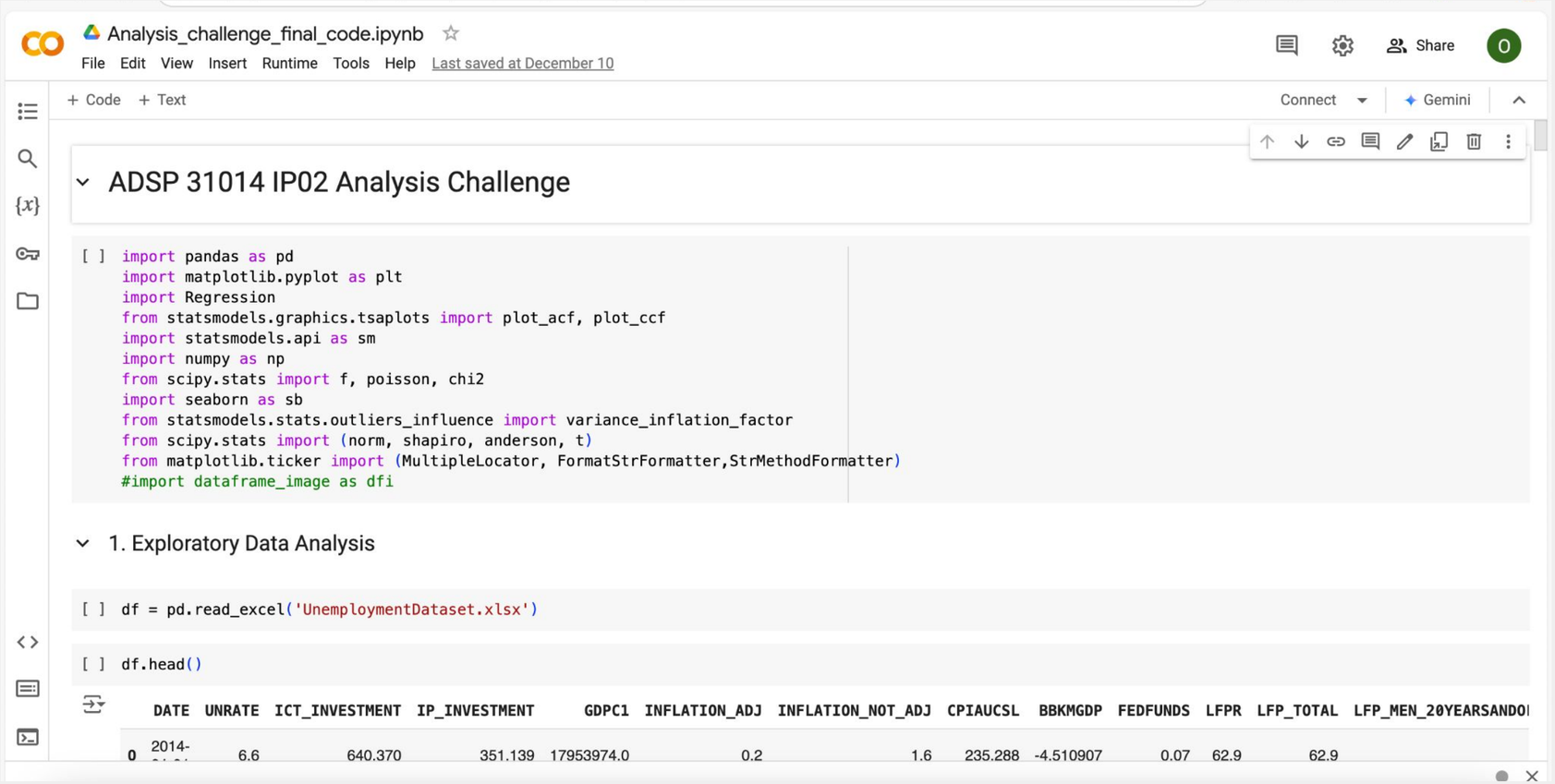
| No | Removal Threshold | Reference Year | Issues | Final Model | R-Squared | MSE | RMSE | MAE | UNRATE Prediction |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.1 | 2024 | Multico in inflation, IP investment, year 2022 | IP_Investment, BBKMGDP, Inflation_Not_adj, dummy year | 0.9673 | 3.45 x 10^-6 | 0.0018 | 0.0013 | 4.13% |
| 2 | 0.05 | 2024 | Multico in inflation, IP investment, year 2022 | IP_Investment, Inflation_Not_adj, dummy year | 0.9664 | 3.54 x 10^-6 | 0.0024 | 0.0013 | 4.13% |
| 3 | 0.1 | 2023 | Multico in inflation, IP investment | P_Investment, BBKMGDP, Inflation_Not_adj, dummy year | 0.9673 | 3.45 x 10^-6 | 0.0018 | 0.0013 | 4.13% |
| 4 | 0.05 | 2023 | Multico in inflation, IP investment | IP_Investment, Inflation_Not_adj, dummy year | 0.9664 | 3.54 x 10^-6 | 0.0024 | 0.0013 | 4.13% |
| 5 | 0.1 | 2023 | Drop IP_Investment, VIF inflation_NOT_Adj=5 | Inflation_not_adj, BBGKMGDP | 0.9636 | 3.84 x 10^-6 | 0.0024 | 0.0014 | 4.17% |
| 6 | 0.1 | 2023 | DROP INFLATION, multico in all variables | FEDFUNDS, LFPR, IP_Investment | 0.9664 | 3.54 x 10^-6 | 0.0024 | 0.0013 | 4.13% |

With the same approach, we tried to check our model using different year references and it shown that either using reference yeat t (2024) or t-1 (2023), the model can help predict the unemployment rate. And using t-1 (2023) as reference year provide features that posses less multicolinearity compare to t (2024).

With similar approach, we can use the model to predict unemployment rate in 2025 using 2024 as reference year.

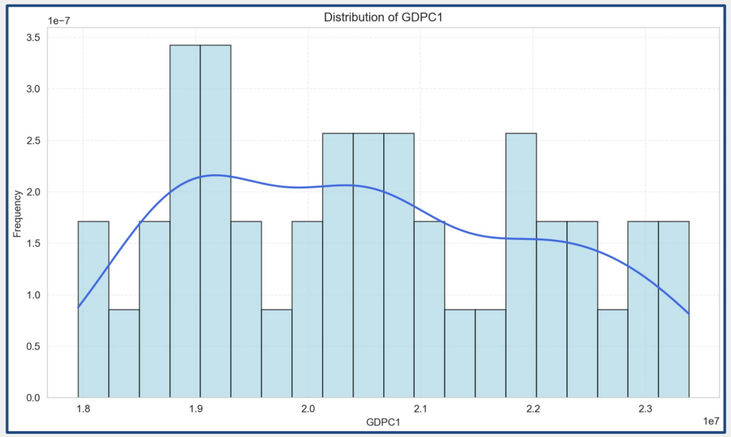
# 06 Appendix

# Code

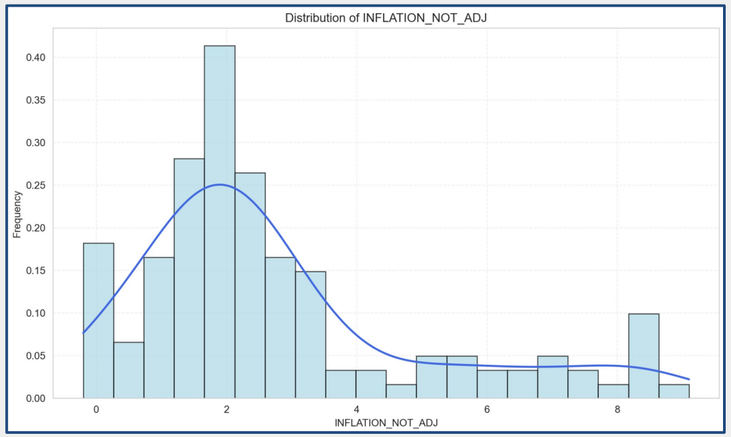



[Analysis Challenge Final Code (Shapley PowerPuff)]

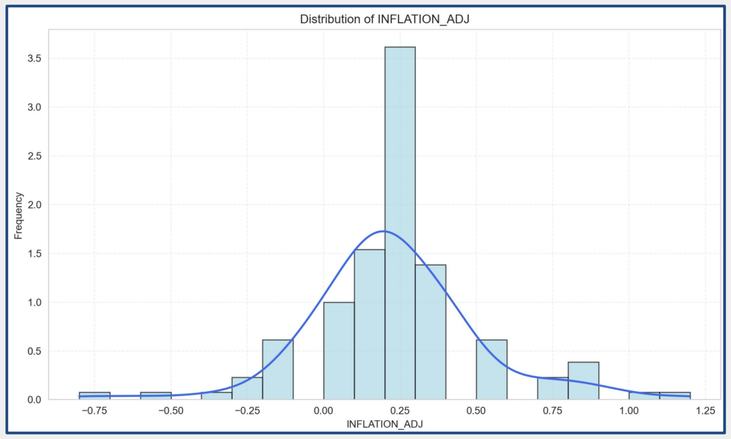
# Distribution of Variables



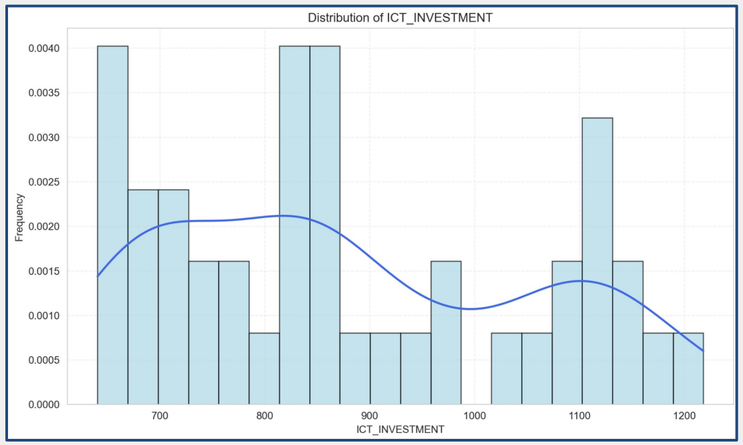

**GDPC1:** Approximately symmetric distribution, no significant skewness.



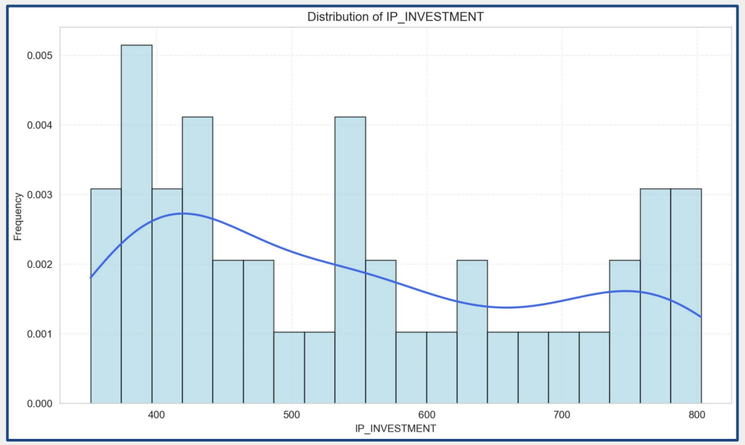

**INFLATION_NOT_ADJ:** Distribution is right-skewed, most values concentrated around 2 and a few higher outliers extending beyond 8.



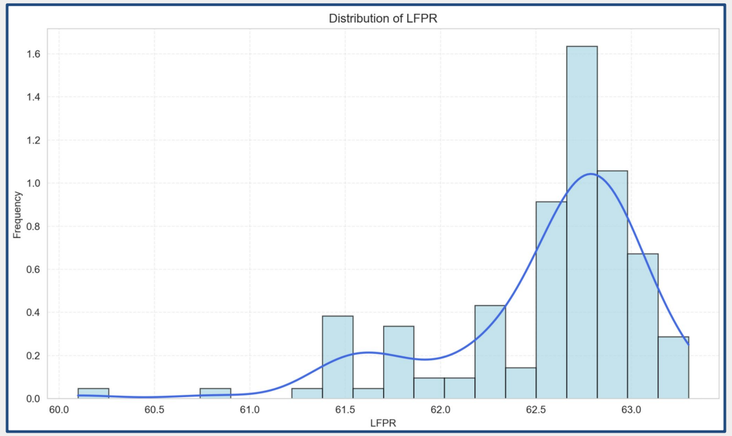

**INFLATION_ADJ:**Normal distribution centered around 0.25 with a sligh right skew. A few values extending toward 1.0




**ICT_INVESTMENT:** Distribution is multimodal with distinct peaks around 700, 900, and 1100.

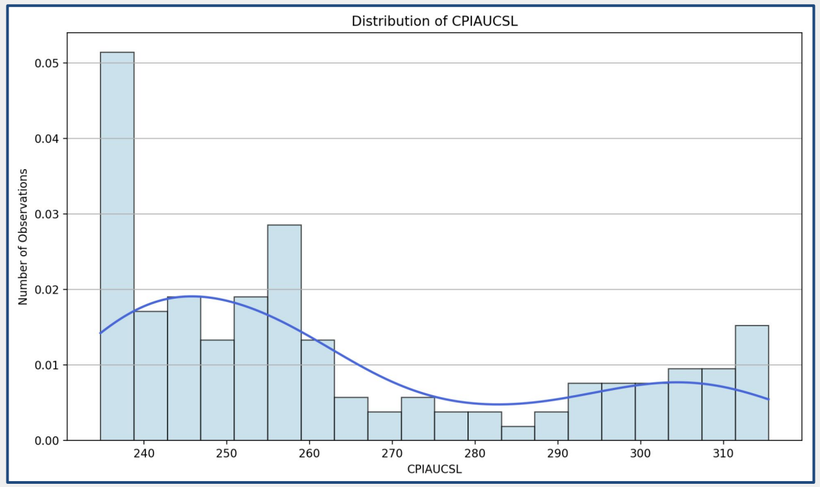



**IP_INVESTMENT:** Distribution is with prominent peaks around 400, 600, and 800.

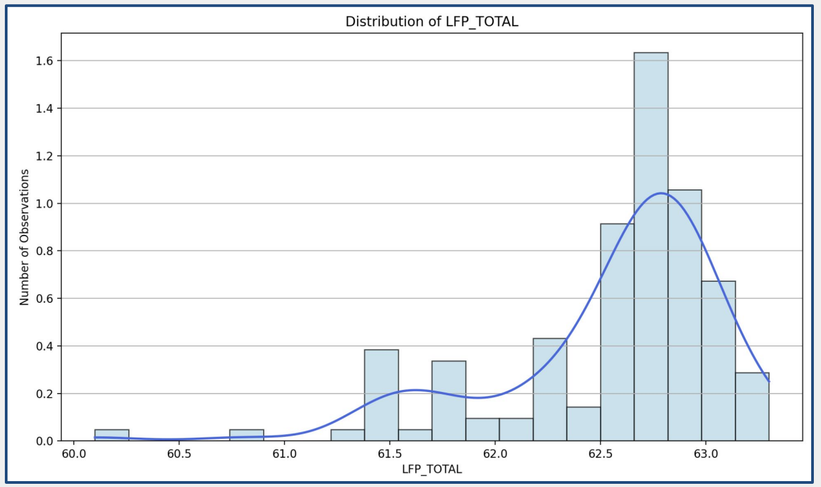



**LFPR:** Distribution is is right-skewed, with the majority of values concentrated between 62.0 and 63.0, and a small tail extending below 61.0

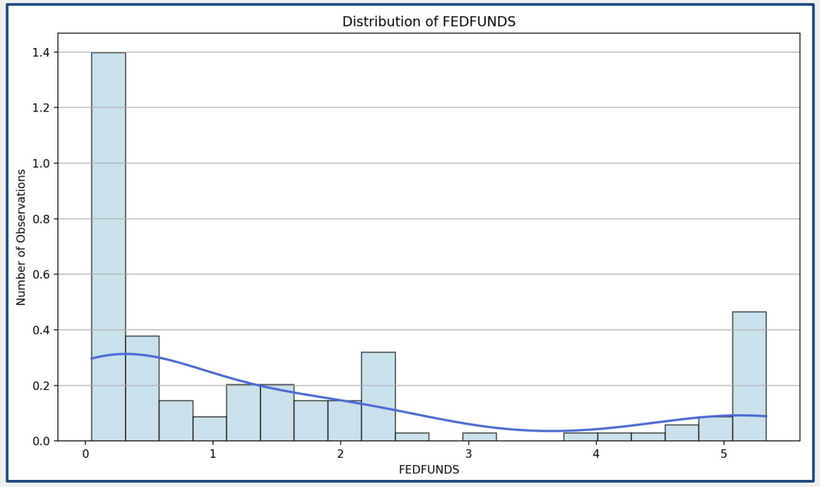
# Distribution of Variables



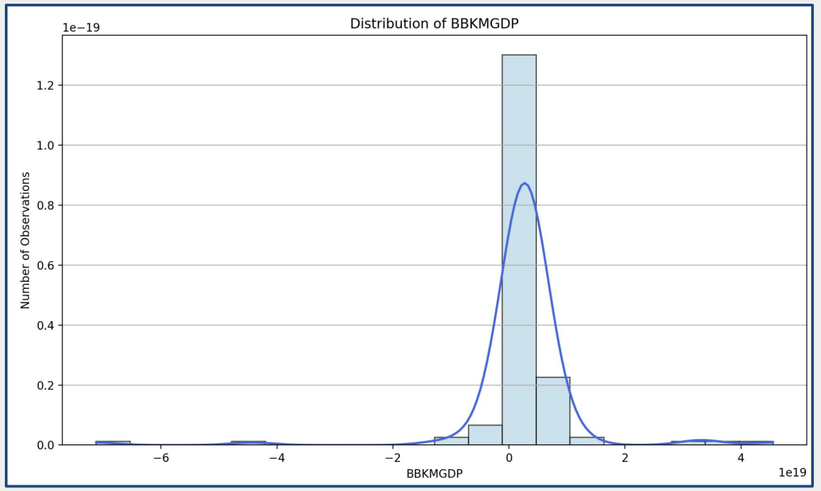

**CPIAUCSL:** Asymmetric distribution with skewness and peaks at lower and higher values.




**LFP_TOTAL:** Right-skewed distribution with a peak around 62.5, indicating most observations are clustered near the higher range.

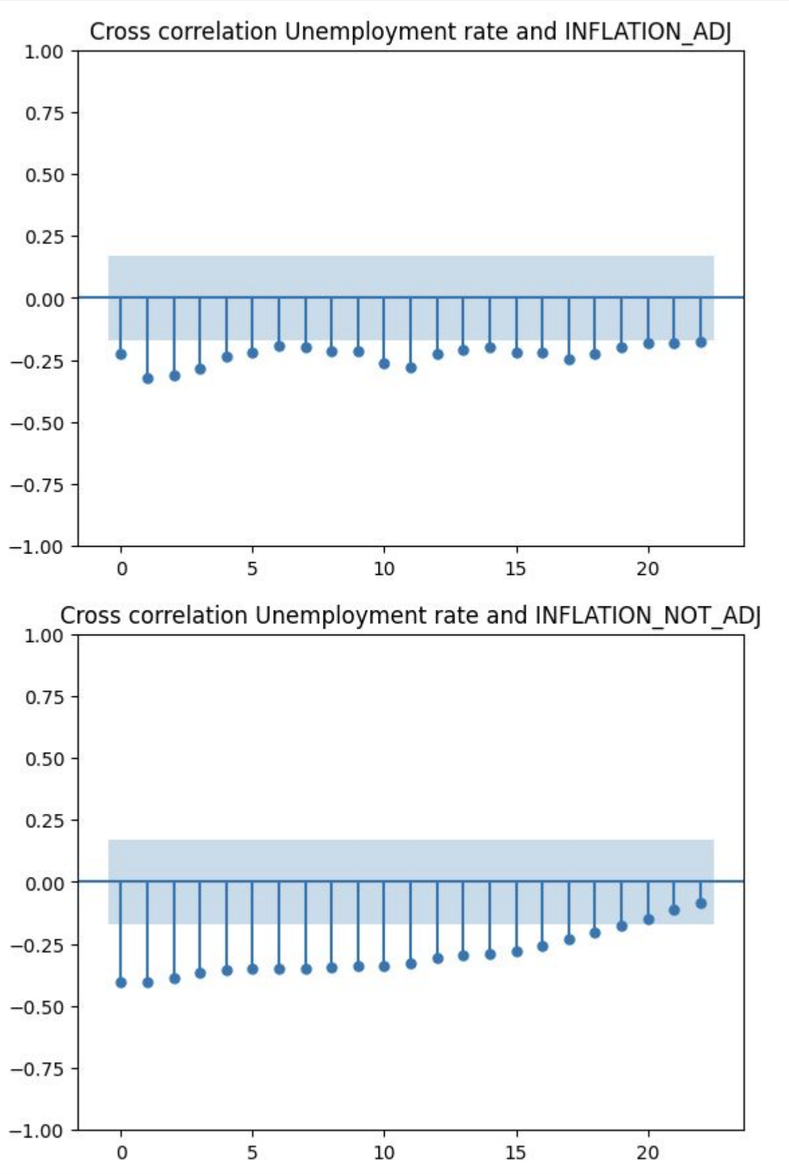


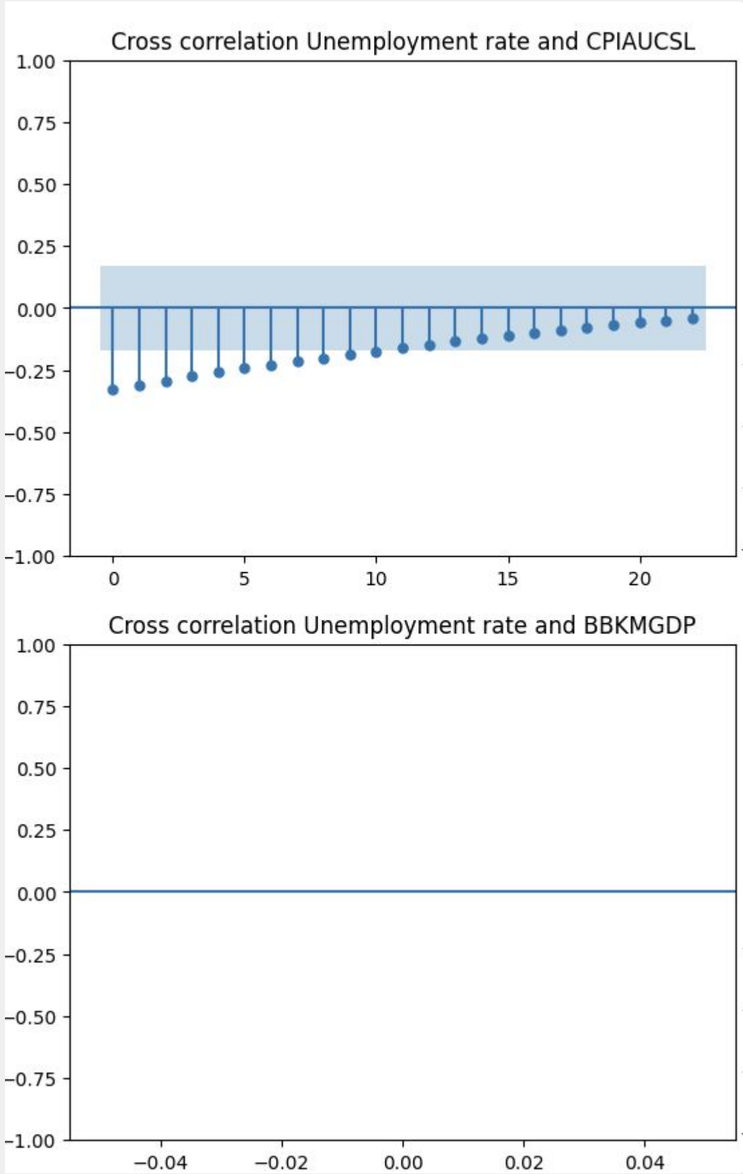

**FEDFUNDS:** Distribution is highly right-skewed with the majority of observations concentrated near lower values around 0 and a long tail extending to higher values.

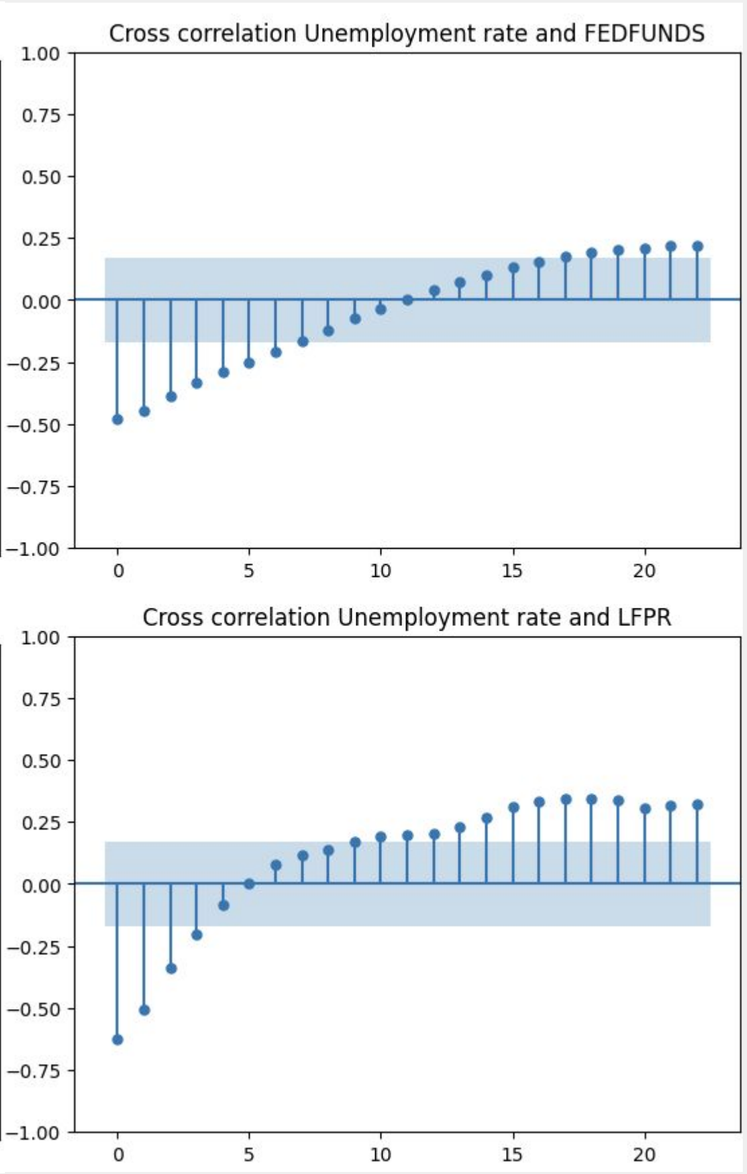



**BBKMGDP:** Distribution shows a highly skewed pattern with the majority of values concentrated near zero, including a small of extreme negative and positive outliers.

# Cross Correlation Between Lag Variables



# Model Research (Regression)

| Independent Variables | Outlier | Covid Data | Multicollinearity (if any) | Method | Model performance (RMSE, MSE) | Prediction of Unemployment Rate |
|---|---|---|---|---|---|---|
| ['ICT_INVESTMENT', 'IP_INVESTMENT','GDPC1','INFLATION_ADJ', 'CPIAUCSL','FEDFUNDS','LFPR'] | | | - | Linear Regression | MSE: 0.44<br>MAE: 0.49<br>RMSE: 0.67 | 4.83 |
| ['ICT_INVESTMENT', 'IP_INVESTMENT','GDPC1','INFLATION_ADJ_LOG', 'CPIAUCSL','FEDFUNDS','LFPR_LOG'] | Log Transformation on UNRATE, LFPR, INFLATION | | | Linear Regression | MSE: 0.40<br>MAE: 0.46<br>RMSE: 0.63 | 4.83 |
| Intercept, GDPC1, FEDFUNDS, LFPR, UNRATE_lag1, UNRATE_lag3, UNRATE_lag6, BBKMGDP | | N/A | N/A | Forward selection | Rsquare: 93.01%<br>RMSE: 0.45 | 3.8% |
| 'IP_INVESTMENT','INFLATION_NOT_ADJ','BBKMGDP','LFPR','LFP_TOTAL','LFP_HISPANICORLATINO' + year dummies | | Dummy variables for year | N/A | Backward Selection | 0.7386 | 4.283 |
| GDPC1, ICT_INVESTMENT, FEDFUNDS_LAG1, FEDFUNDS, LFPR, CPIAUCSL | | | GDPC1 and GDPC1_LAG1, FEDFUNDS and FEDFUNDS_LAG1 | Forward Selection used | RMSE = 1.5459 | 3.48% |
| ['ICT_INVESTMENT', 'IP_INVESTMENT', 'GDPC1', 'INFLATION_ADJ', 'INFLATION_NOT_ADJ', 'CPIAUCSL', 'FEDFUNDS', 'LFPR', 'LFP_TOTAL', 'LFP_MEN_20YEARSANDOLDER', 'LFP_WOMEN_20YEARSANDOLDER', 'LFP_16TO19YEARSOLD', 'LFP_WHITE', 'LFP_BLACKORAFRICANAMERICAN', 'LFP_ASIAN', 'LFP_HISPANICORLATINO', 'ICT_INVESTMENT_lagged1', 'ICT_INVESTMENT_lagged3', 'IP_INVESTMENT_lagged1', 'IP_INVESTMENT_lagged3', 'GDPC1_lagged1', 'GDPC1_lagged3', 'INFLATION_ADJ_lagged1', 'INFLATION_ADJ_lagged3', 'INFLATION_NOT_ADJ_lagged1', 'INFLATION_NOT_ADJ_lagged3', 'CPIAUCSL_lagged1', 'CPIAUCSL_lagged3', 'FEDFUNDS_lagged1', 'FEDFUNDS_lagged3', 'LFPR_lagged1', 'LFPR_lagged3', 'LFP_TOTAL_lagged1', 'LFP_TOTAL_lagged3', 'LFP_MEN_20YEARSANDOLDER_lagged1', 'LFP_MEN_20YEARSANDOLDER_lagged3', 'LFP_WOMEN_20YEARSANDOLDER_lagged1', 'LFP_WOMEN_20YEARSANDOLDER_lagged3', 'LFP_16TO19YEARSOLD_lagged1', 'LFP_16TO19YEARSOLD_lagged3', 'LFP_WHITE_lagged1', 'LFP_WHITE_lagged3', 'LFP_BLACKORAFRICANAMERICAN_lagged1', 'LFP_BLACKORAFRICANAMERICAN_lagged3', 'LFP_ASIAN_lagged1', 'LFP_ASIAN_lagged3', 'LFP_HISPANICORLATINO_lagged1', 'LFP_HISPANICORLATINO_lagged3'] | | | | Forward selection and PCA | 0.87 | 4.2 |

# Dummy Year as Intercept

| Metric | Value |
|---|---|
| **MSE** | 3.3797 |
| **RMSE** | 0.0024 |
| **MAE** | 0.0013 |
| **Intercept** | 0.5569 |
| **Year dummy variables** | for 2014: -0.030607<br>for 2023: -0.037243 |

## Summary of the results

- The inclusion of **dummy year variables** enhances the model's ability to account for time-specific effects.
- **Year dummy variables indicate significant variations** in unemployment rates across years, reflecting year-to-year changes in economic conditions.

# Model Overview

## Box Cox Transformation

Many statistical techniques, such as hypothesis tests and confidence intervals, rely on the assumption that the response variable follows a normal distribution. When unemployment values deviate significantly from normality, the validity of these inferences can be compromised. Applying a Box-Cox transformation **adjusts the data to better approximate normality**, stabilizing variance and meeting the assumptions of linear regression.

## Backward Selection

Backward selection was used to select the predictors of our model. It began with a model that included all predictors and iteratively removed the least significant ones. This process continued until all remaining **predictors were statistically significant or the model reached optimal performance**. This approach ensures the final model is parsimonious, avoids overfitting, and focuses only on the most impactful predictors.

## Linear Regression

Linear regression was used to model the relationship between the target variable and the selected predictors. This technique assumes a linear relationship and **estimates the coefficients that best fit the data** by minimizing the sum of squared errors. This approach is simple, interpretable, and efficient, making it suitable for understanding the impact of individual predictors on the target variable.

# Multicollinearity (VIF)

There is a potential multicollinearity from independent variables, especially IP_INVESTMENT & INFLATION_NOT_ADJ, as they have high VIF

| | feature | VIF |
|---|---|---|
| 0 | IP_INVESTMENT | 18.534585 |
| 1 | BBKMGDP | 1.093604 |
| 2 | INFLATION_NOT_ADJ | 23.139918 |
| 3 | 2014 | 1.363116 |
| 4 | 2015 | 1.753180 |
| 5 | 2016 | 1.487056 |
| 6 | 2017 | 1.535546 |
| 7 | 2018 | 1.660292 |
| 8 | 2019 | 1.793722 |
| 9 | 2020 | 2.075821 |
| 10 | 2021 | 2.867328 |
| 11 | 2022 | 6.859435 |
| 12 | 2023 | 2.752344 |

## Unemployment Trends Across Economic Periods

| Economic Period | Average Unemployment | Model captures |
|---|---|---|
| Pre-Pandemic (2014-2019) | 4-6% | Normal economic cycles |
| Pandemic (2020-2021) | above 14% | Extreme deviation |
| After Pandemic (2022-2024) | approx. 4% | Normalization trend |

# Implication on Policy

1. Lagged variables (e.g., GDP, CPI) suggesting a predictive power for unemployment trends..

   **Policy Recommendation:** we can use these indicators as early alert to implement timely interventions in the labor market.

2. The inverse relationship between inflation and unemployment

   **Policy Recommendation:** Central banks can consider moderate inflation support job creation.

3. Negative coefficients for variables such as IP Investment suggest that increased investments can drive down unemployment rates.

   **Policy Recommendation:** Introduce incentives, such as tax credits or grants, to encourage private sector investment in industries like technology and manufacturing.

4. Yearly dummy variables highlight significant effect on major economic event (COVID-19)

   **Policy Recommendation:** create targeted programs for affected sectors in yearly basis