Forecasting Report

AUTHOR
LUKE STUCKY

PUBLISHED
December 11, 2024

BUSINESS UNDERSTANDING

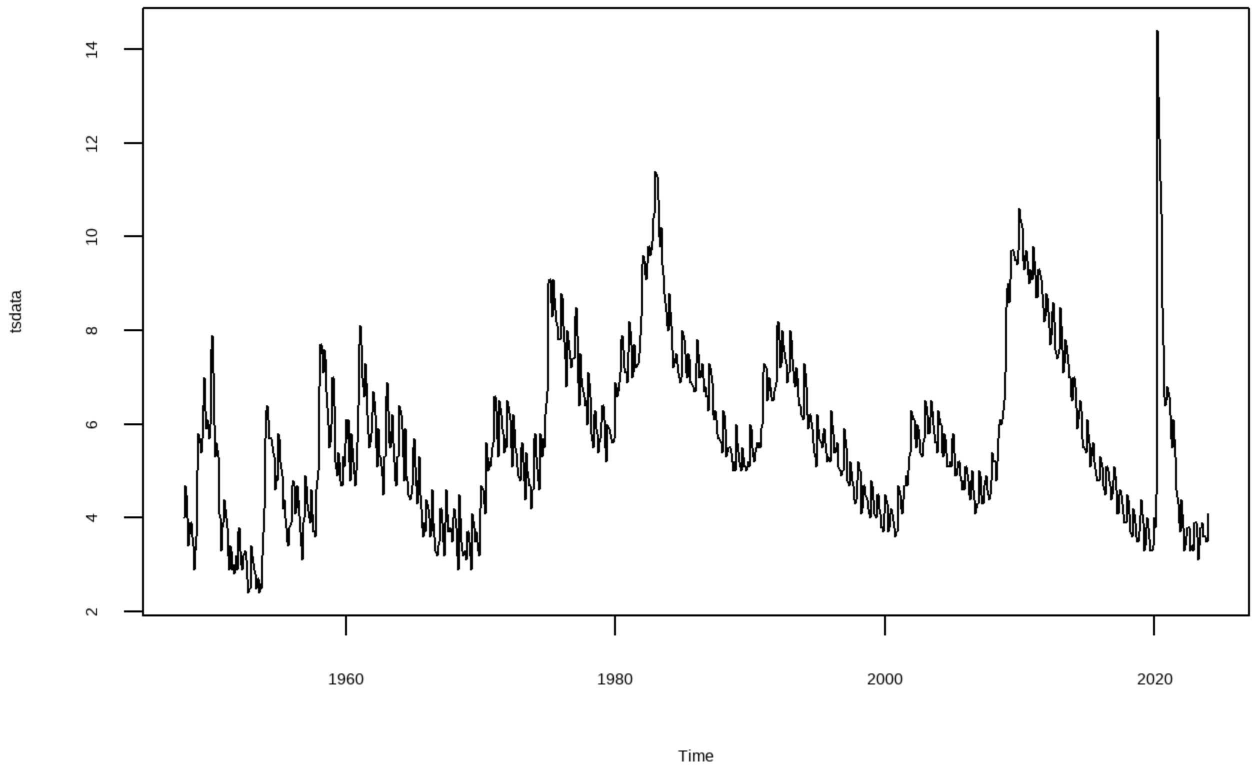
As the President's economic advisor, Luke Stucky's responsibility is to analyze historical unemployment rate data to provide actionable insights into labor market trends. These insights will enable the President to anticipate economic challenges and develop informed strategies for policy decisions.

To complete this project, Luke set out to answer the following two questions to aid in correctly informing the President:

What patterns and trends can be identified in historical unemployment rate data to understand labor market dynamics?
How might these historical insights guide proactive planning and decision-making during the President's term?

By focusing on historical data analysis, Luke aims to equip the President with a comprehensive understanding of past labor market conditions, enabling data-driven strategies to address current and future economic challenges while supporting workforce stability.

DATA UNDERSTANDING

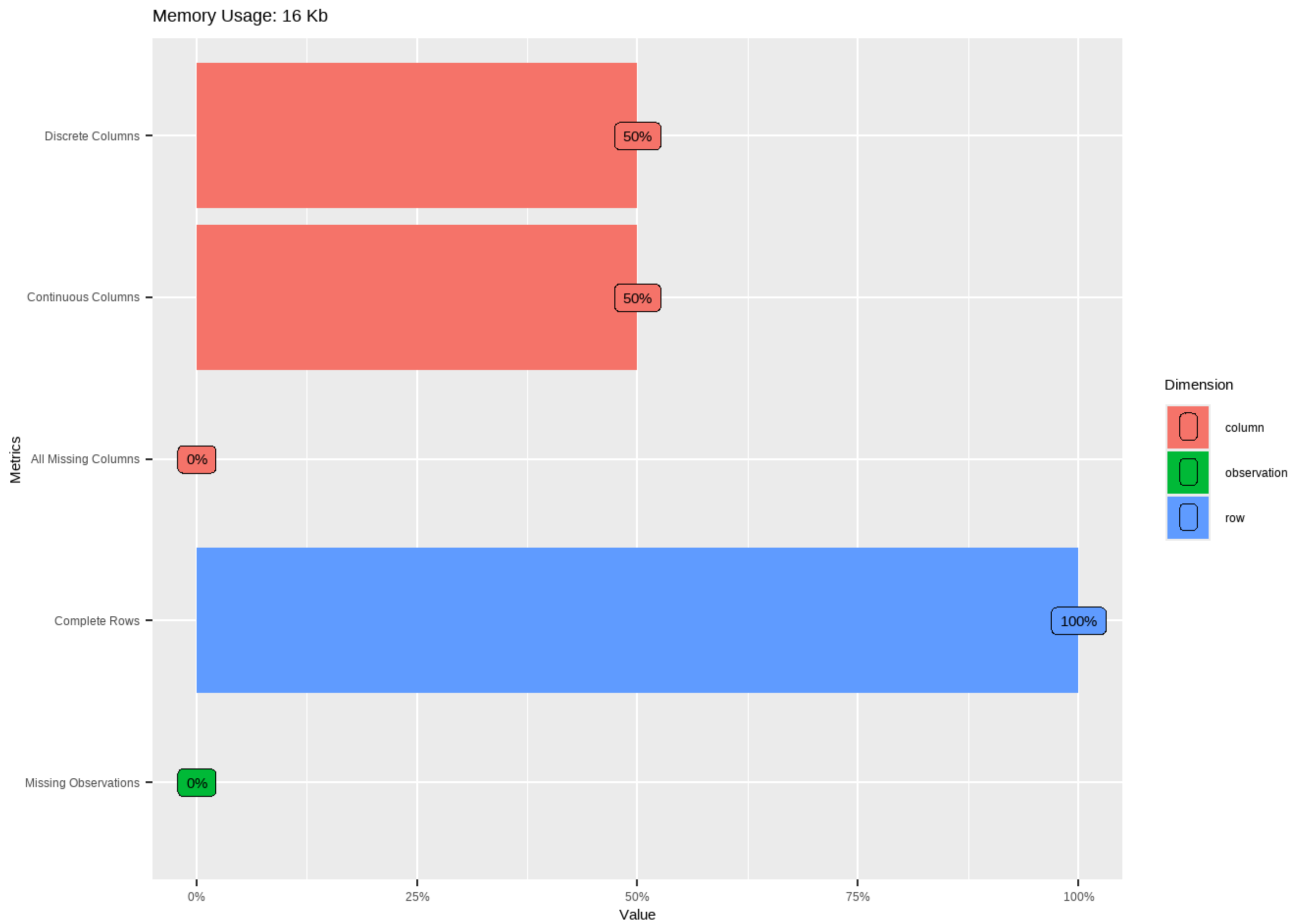


The time-series plot of unemployment rates shows a cyclical behavior. Theses cycles are linked to recessions and recoveries. When a recession hits, the unemployment rate is at its highest, but it begins to decline then until the next recession.

EDA

```
        date unratensa
1 1948-01-01       4.0
2 1948-02-01       4.7
3 1948-03-01       4.5
```

```
          date unratensa
921 2024-09-01       3.9
922 2024-10-01       3.9
923 2024-11-01       4.0
```



Outliers

```
  variables outliers_cnt outliers_ratio outliers_mean with_mean without_mean
1 unratensa           13       1.408451      11.14615  5.685157     5.607143
```

There are outliers in our data, but that is to be expected in unemployment rate data. We will leave the outliers in as we do not want to change anything about this data.

Data Preparation

Partitioning

Modeling

Regression with Trend and Seasonality

```
Call:
```

```
lm(formula = Unemployment ~ Trend + Seasonal, data = reg_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0443 -1.3170 -0.1895  1.0377  8.2970 

Coefficients:
              Estimate Std. Error t value             Pr(>|t|)    
(Intercept)  5.7390067  0.2167144  26.482 < 0.0000000000000002 ***
Trend        0.0011451  0.0002133   5.369          0.000000101 ***
Seasonal2    0.0078617  0.2746315   0.029             0.977169    
Seasonal3   -0.2564413  0.2746307  -0.934             0.350674    
Seasonal4   -0.6299549  0.2746302  -2.294             0.022030 *  
Seasonal5   -0.7784684  0.2746298  -2.835             0.004691 ** 
Seasonal6   -0.2190873  0.2746295  -0.798             0.425223    
Seasonal7   -0.3754956  0.2746294  -1.367             0.171879    
Seasonal8   -0.6884828  0.2746295  -2.507             0.012353 *  
Seasonal9   -0.8948911  0.2746298  -3.259             0.001162 ** 
Seasonal10  -1.0670889  0.2746302  -3.886             0.000110 ***
Seasonal11  -0.9261288  0.2746307  -3.372             0.000777 ***
Seasonal12  -0.8878002  0.2746315  -3.233             0.001271 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.698 on 900 degrees of freedom
Multiple R-squared:  0.07163,   Adjusted R-squared:  0.05925 
F-statistic: 5.787 on 12 and 900 DF,  p-value: 0.000000001038
```

Running a regression with seasonality and trend showed us that the trend and lots of seasons are statistically significant. However, the model is not able to explain very much variation.

Prediction

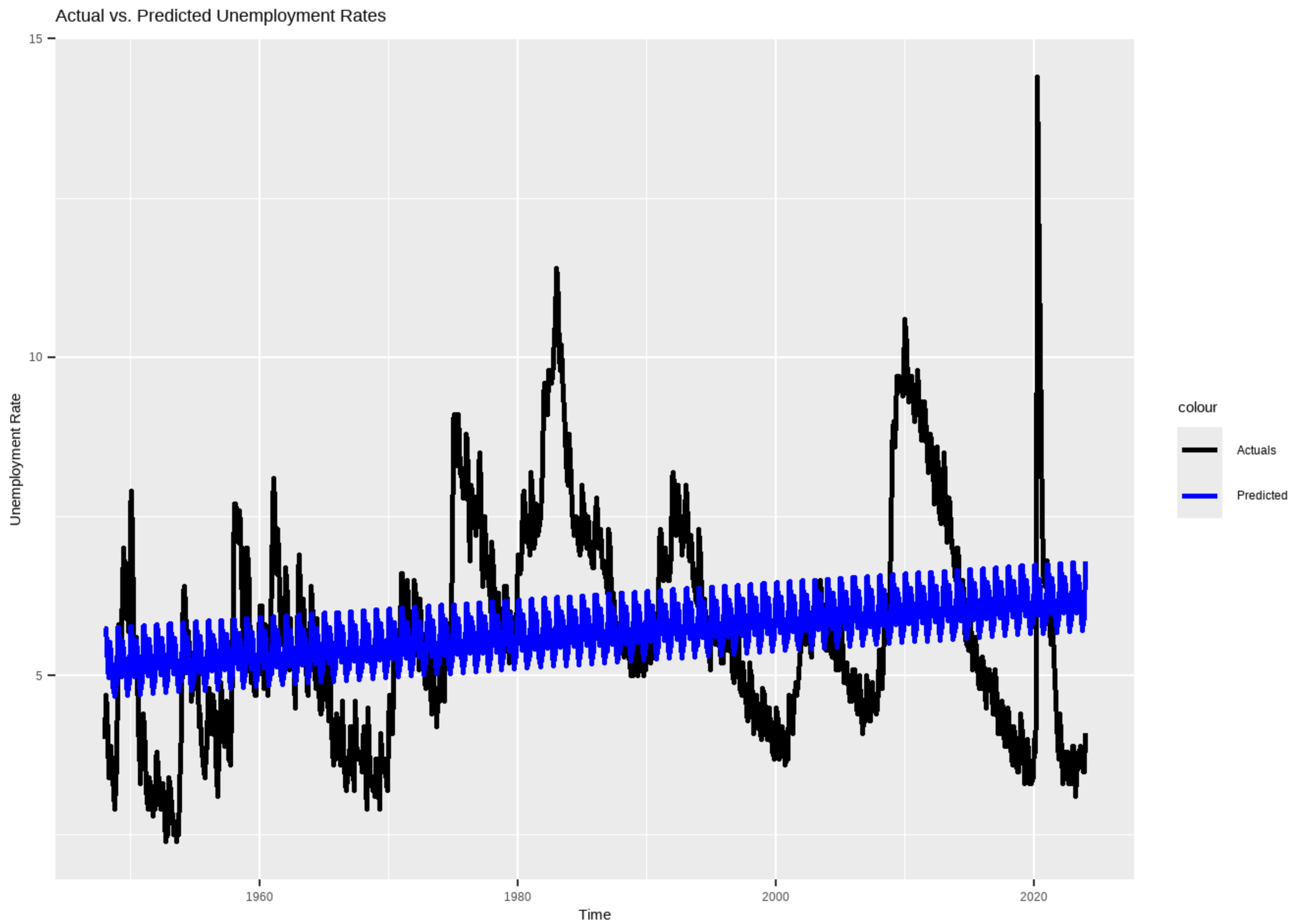
```
Training RMSE: 1.51

Training MAPE: 22.41 %

Validation RMSE: 2

Validation MAPE: 31.62 %
```

Displaying the MAPE and RMSE for both the training and the validation set shows that the validation set performs a lot worse than the training set. This is likely due to the two big recessions that hit in the validation set. Our MAPE for the training set is 22.41 percent but for the validation it is 31.62 percent. This is the same for the RMSE.



Plotting the predicted values shows that there is an upward trend in the unemployment rate. It is very slow but it is present.

Simple Moving Average

Calculate the RMSE and MAPE to assess fit

```
Training RMSE: 0.64
```

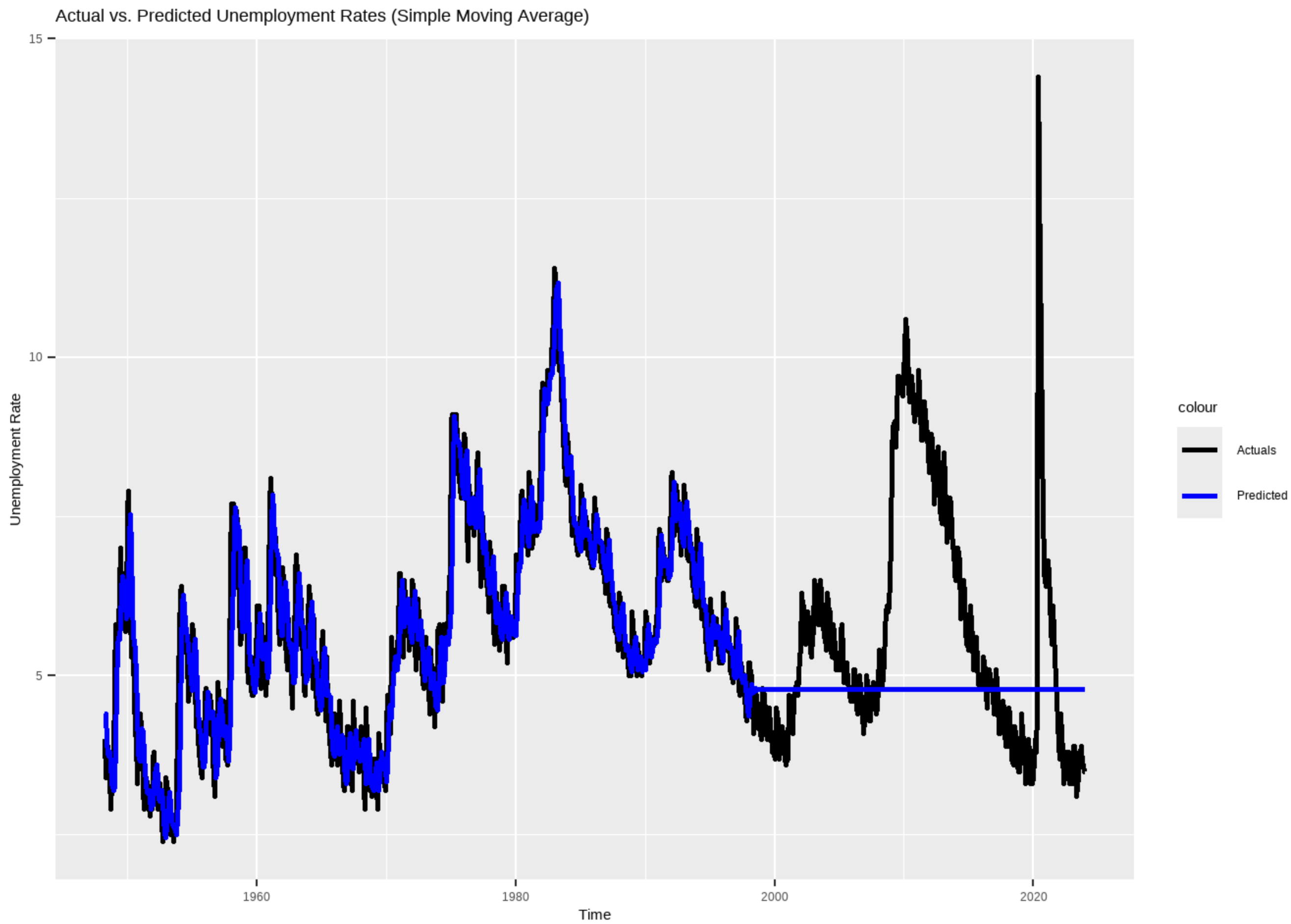
```
Training MAPE: 8.94 %
```

```
Validation RMSE: 2.13
```

```
Validation MAPE: 22.37 %
```

The validation RMSE performs worse in the simple moving average than in the regression, but the validation MAPE outperforms in this model by a landslide.

Actual vs Predicted



The plot of the predicted rate shows that it would remain constant a little under a 5% unemployment rate. Again, the 2008 recession and covid hit during the prediction, so that hurt the metrics significantly.

Holt-Winters Smoothing Model

Estimate the model using the computer

Validate

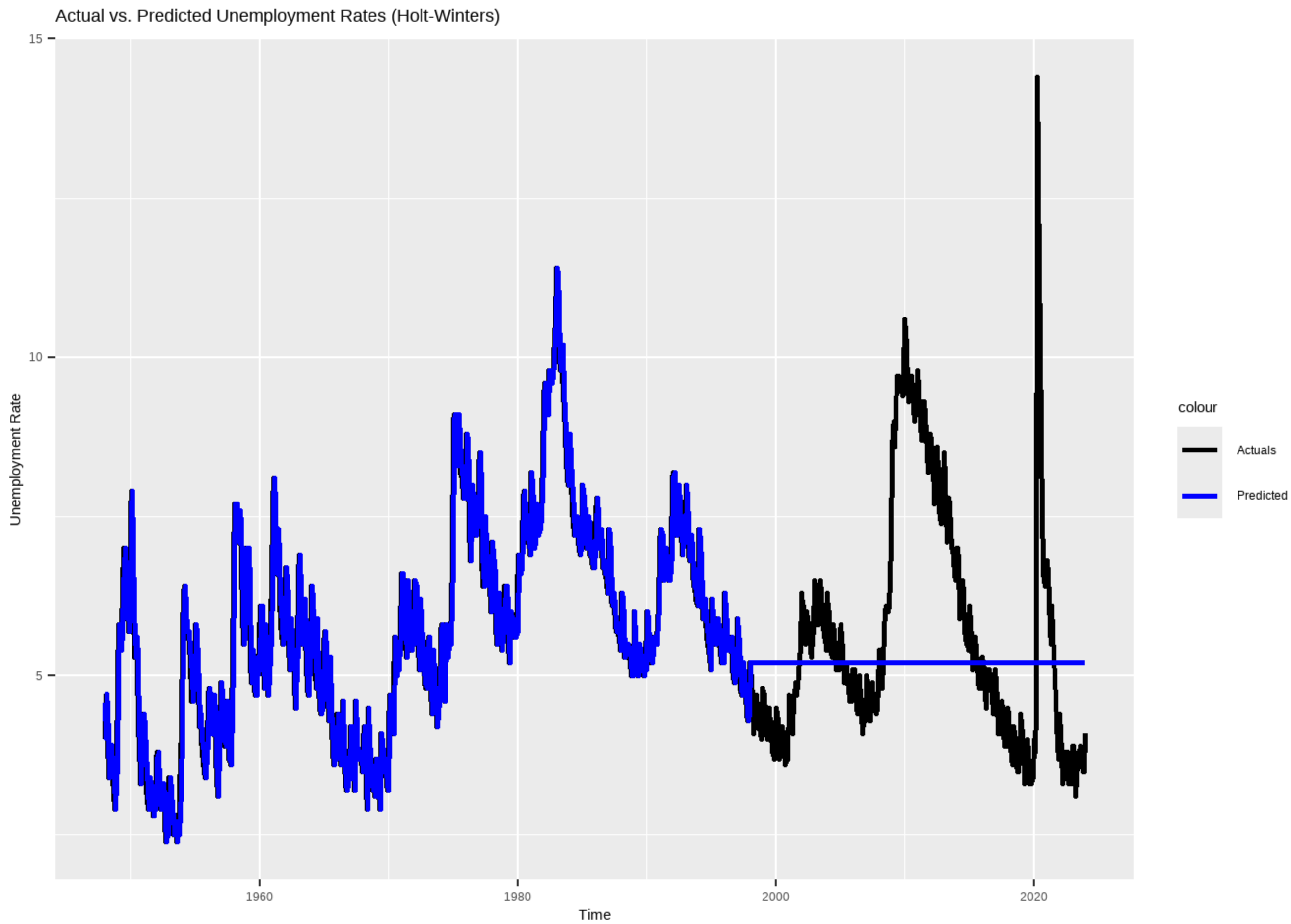
```
Training RMSE: 0.5

Training MAPE: 7.01 %

Validation RMSE: 2

Validation MAPE: 23.92 %
```

The Holt-Winters Smoothing method performs much better in the training set, but in the validation set, our MAPE is higher than in the simple moving average. Given this, it will be a close call for which model is used.



From the plot, it is clear that the training plot is much more accurate in this Holt-Winters model than in the simple moving average. In this, model, the predicted unemployment rate remains constant at a little over 5%.

Evaluation

Model	Set	RMSE	MAPE
Simple Moving Average	Training	0.6392698	8.942365
Simple Moving Average	Validation	2.1299945	22.373326
Holt-Winters	Training	0.5028948	7.008321
Holt-Winters	Validation	1.9961611	23.923588
Regression with Trend and Seasonality	Training	1.5148934	22.412612
Regression with Trend and Seasonality	Validation	1.9985052	31.619017

Comparison of Model Metrics for Training and Validation Sets

Considering the RMSE and the MAPE for both the training and validation data, the best model would be the Holt-Winters. This is due to it having the lowest RMSE in both and a competitively low MAPE.

Deployment

After evaluating each model, we determined that the Holt-Winters model provides the most accurate predictions, demonstrated by its lowest RMSE of 2 and competitive MAPE of 31.62 on the validation set of data.

Returning to the questions from the beginning, the Holt-Winters model can effectively address each question:

What patterns and trends can be identified in historical unemployment rate data to understand labor market dynamics? The Holt-Winters model shows that unemployment rates rise and fall in cycles, which match times of economic trouble and recovery. It catches seasonal changes and long-term trends well. During recessions, unemployment rates rise, and during recoveries, they fall.

How might these historical insights guide proactive planning and decision-making during the President's term? By identifying patterns in unemployment rates, the President can better anticipate economic instability and implement timely policies. While the Holt-Winters model can't predict exact timing of recessions and recoveries, it provides a good reference for expected trends and seasonal fluctuations. This helps the President prepare for potential unemployment spikes and supports workforce stability during recoveries. Despite its limitations, the model's insights guide proactive steps to manage the labor market effectively.

the Holt-Winters model, with its lowest RMSE of 2 and competitive MAPE of 31.62, offers the most accurate predictions among the evaluated models. Although it cannot predict the exact timing of recessions and recoveries, it serves as a reliable reference for expected trends and seasonal fluctuations, aiding in proactive planning and effective decision-making to manage labor market dynamics.

REFERENCES

R and Packages

```
R version 4.3.1 (2023-06-16 ucrt)


R Packages Used:

 [1] "dlookr"       "knitr"        "Metrics"      "forecast"     "smooth"
 [6] "greybox"      "janitor"      "rpart.plot"   "rpart"        "klaR"
[11] "MASS"         "pROC"         "gains"        "caret"        "lattice"
[16] "gridExtra"    "flextable"    "DataExplorer" "lubridate"    "forcats"
[21] "stringr"      "dplyr"        "purrr"        "readr"        "tidyr"
[26] "tibble"       "ggplot2"      "tidyverse"
```

Other References

Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2023). *Business analytics: Communicating with numbers* (2nd Ed.). McGraw-Hill.