Unemployment Rate Forecasting: Colombia vs. the U.S.

Time Series Analysis for Energy and Environment Applications

ENERGY 797.01.Sp25

Apr 8, 2025

Team Members:

Mazhar Bhuyan

Jisup Kwak

Aye Nyein Thu

Yuqi Yang

Contents

Introduction and Data Description

Why Does Unemployment Matter?

- A key economic and social indicator.
- Affects both individual well-being and long-term economic stability.
- Critical for policy design, economic planning, and social support systems.

The COVID-19 Shock

- In 2020, over 255 million full-time jobs were lost globally (ILO, 2021).
- Global labor markets are still recovering from the shock, with different recovery paths.

Motivation

- Understand how unemployment behaves in different economies (developed vs. developing).
- Explore how labor market structures shape recovery after COVID-19.
- Forecast unemployment trends over the next 12 months.
- Provide insights to support better policy-making.

Data Description

U.S.

- ❑ Formal, structured labor market.
- ❑ Stronger unemployment insurance.
- ❑ Stable institutions and comprehensive safety nets.

Colombia

- ❑ Less regulated and more unstable labor market.
- ❑ Large informal sector.
- ❑ Weaker safety nets and social protection.

Data Source

- ❑ Monthly data from International Labor Organization (ILOSTAT).
- ❑ U.S.: 2001–2024; Colombia: 2010–2024.

Key Variables

- ❑ **Unemployment Rate**
  = (Unemployed / Labor Force) × 100
- ❑ Unemployed: Not working but actively seeking work in the past 4 weeks.
- ❑ The rate adjusts for changes in the size of the labor force, making it easier to compare across countries and over time.

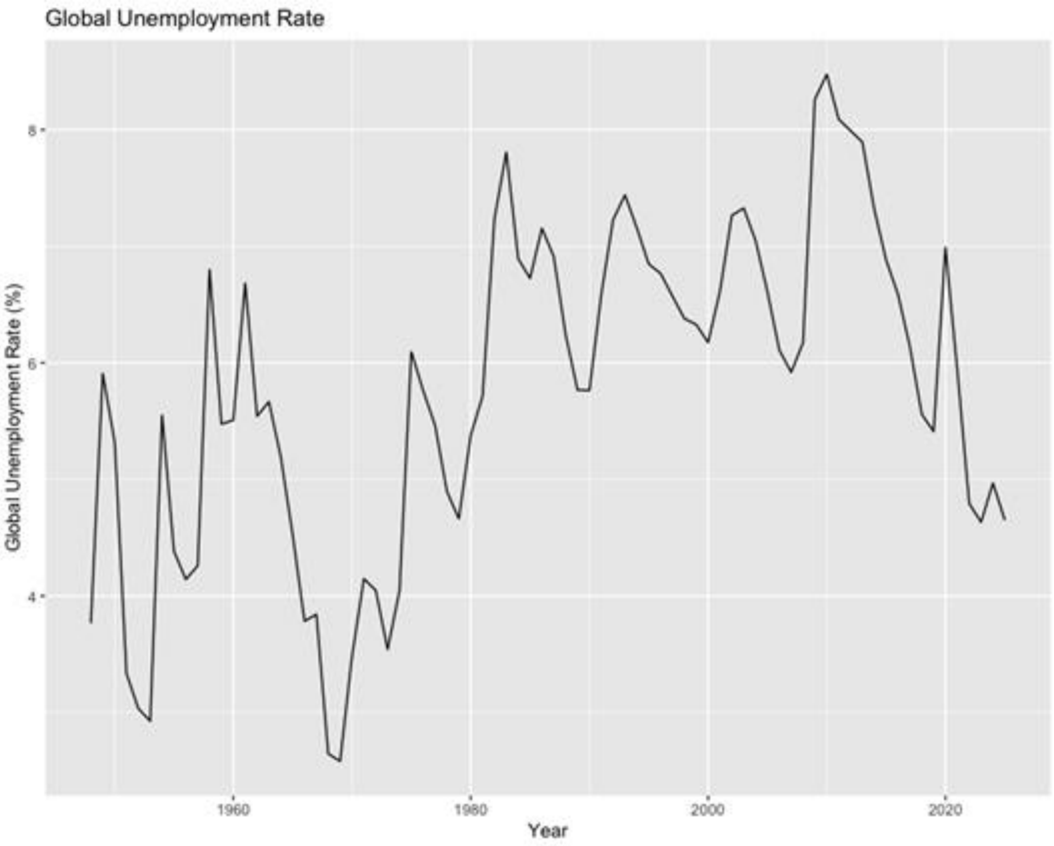
Forecasting Processes

1. Data Sources Identification
2. Data Wrangling
   a. Combining different variables
   b. Filtering specific years for forecasting
   c. Handling missing values and outliers
3. Descriptive Statistics
   a. Global Trends (No. of countries: 91 using weighted average)
   b. Incorporation of employment and population data
4. Time Series Analysis (No. of Countries: Colombia and US)
   a. Transform into time series
   b. Decomposition
   c. Testing Models and Residuals Analysis
   d. Performance Checks
   e. Forecasting 2025 with the average of best three models

No.	Time Series Models	Rationale
1	Seasonal Naive (SNAIVE)	Simple to use, Easy to understand
2	Simple Moving Average (SMA)	Smoothing out variations in the series using equal weights
3	Simple Exponential Smoothing Model (SMA)	Smoothing out variations in the series using weighted average and percentage of forecast error
4	SARIMA	Accounting for both seasonal and non-seasonal data
5	ARIMA	Accounting for the trends and autoregressive parts
6	STL + ETS Model	Handling complex/ irregular trend and seasonality
7	ARIMA + Fourier Model	Handling complex/ irregular seasonality while accounting for autoregressive components
8	TBATS	Handling complex/ irregular seasonality
9	Neural Network	Handling complex/ irregular trend and seasonality with non-linear relationship
10	State Space Exponential Smoothing	Handling both trend and seasonality and capturing changes in patterns over time
11	State Space with BSM Model	Capturing uncertainty or noise in the data, especially for non-linear patterns

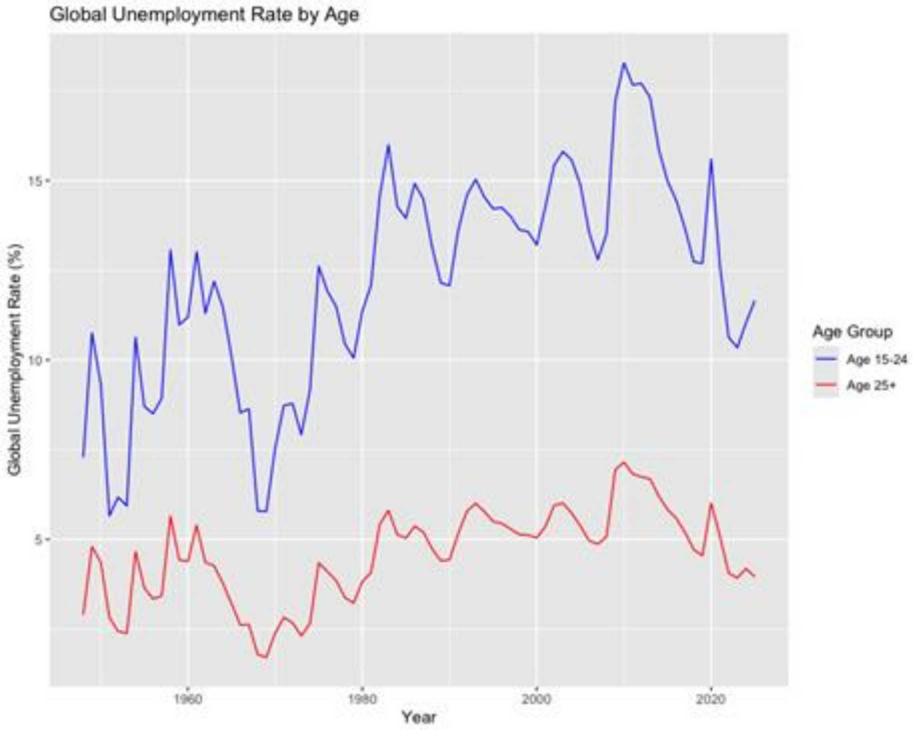
Global Unemployment Overview

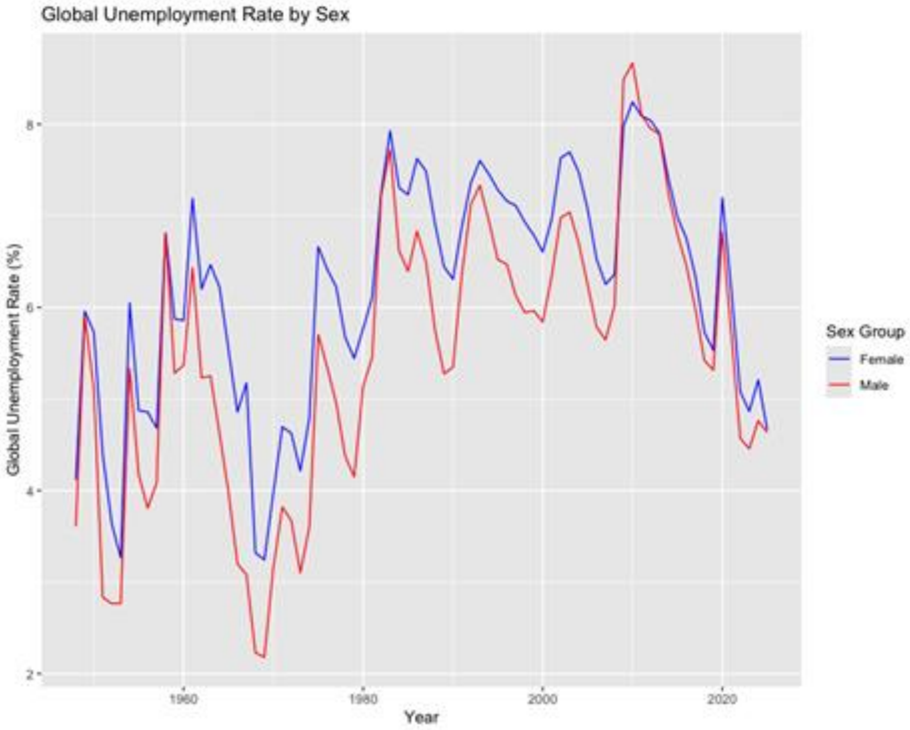
Global Unemployment Rate Using Weighted Average (January 1960 to February 2025)



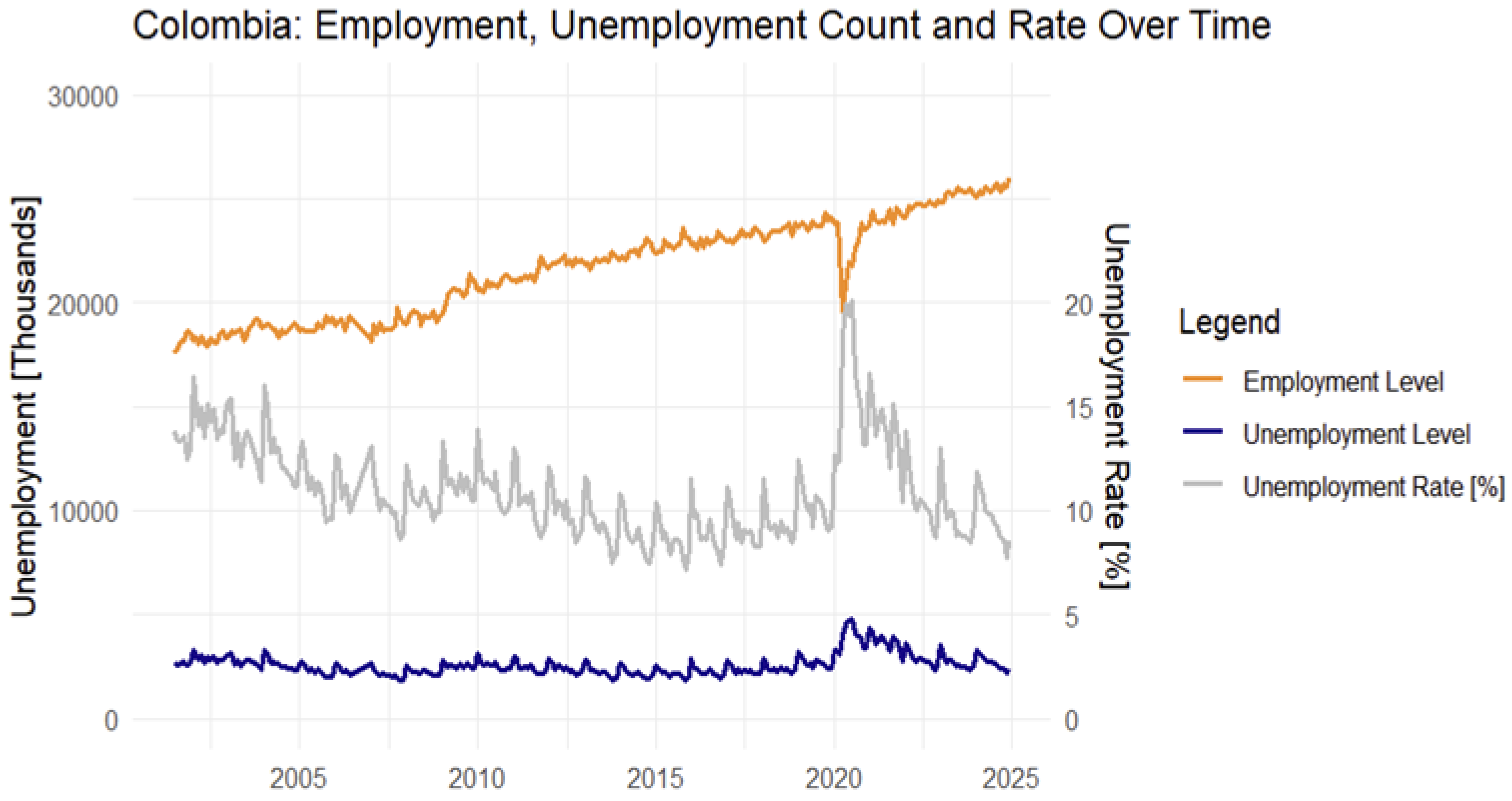
- The global unemployment rate is around 6%.
- Unemployment remains steady.
- This is close to historical lows, suggesting a resilient global labor market after COVID-19.
- Most economies have recovered job losses from the pandemic.

Global Unemployment Rate Using Weighted Average By Age Group and Sex Group



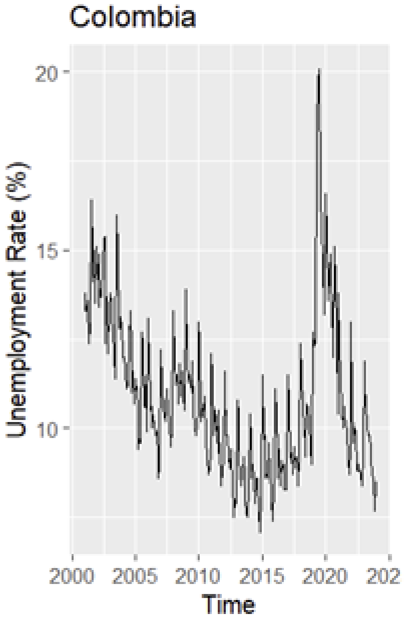


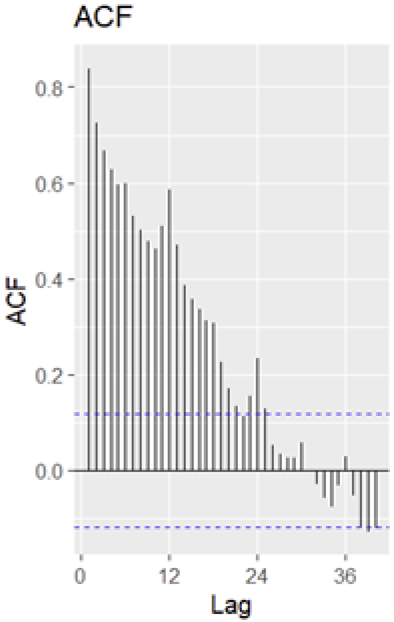
Forecasting Unemployment Rate in Colombia

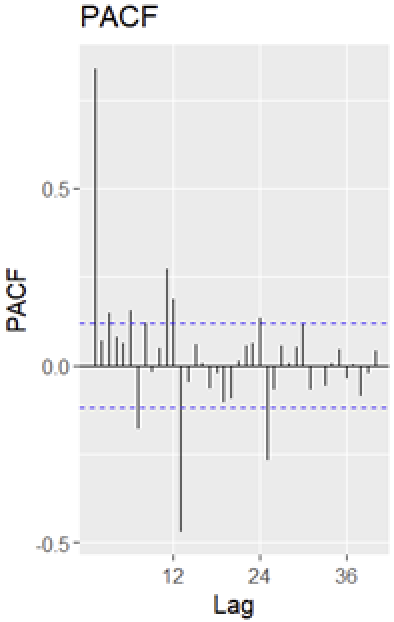


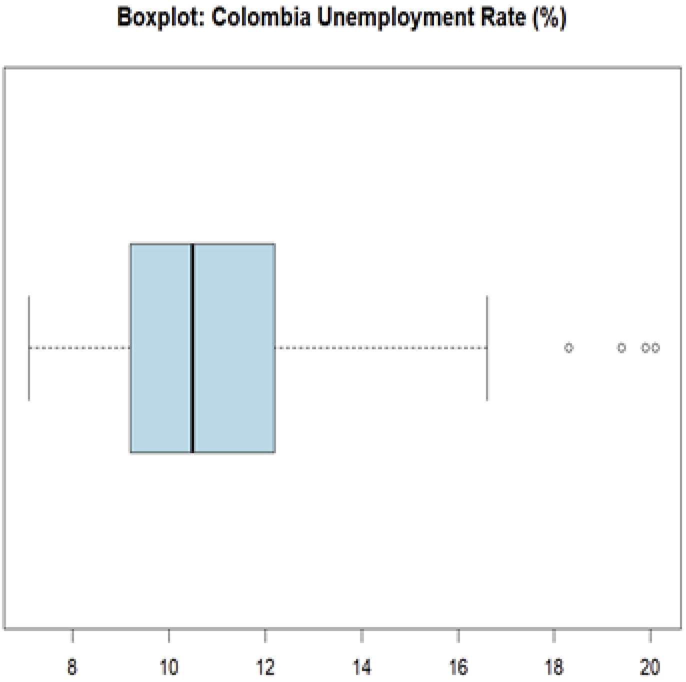
Unemployment rate shows a gradual decline with seasonal variation. Employment steadily increased, except for a sharp drop in 2020 (COVID-19). Clear labor market recovery observed after the pandemic.

Unemployment Rate in the Colombia - Initial Data Exploration



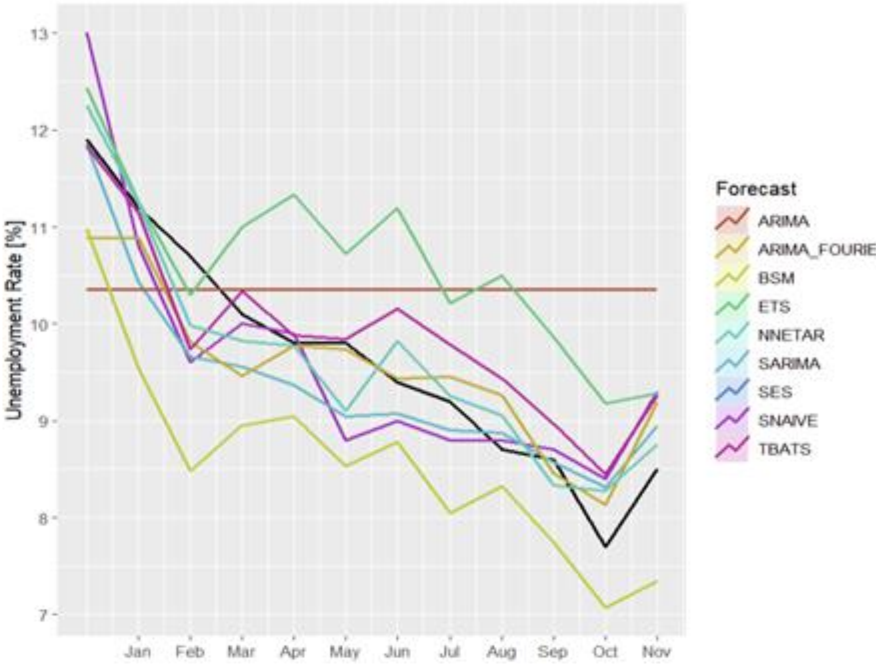






ACF and PACF indicate strong autocorrelation and seasonality.
Boxplot suggests a few outliers above 18%, especially around and 2020.
ADF Test (p = 0.098): Fails to reject the null → Evidence of unit root (non-stationary).

Test Result of Unemployment Rate in the Colombia



Forecast Accuracy for Unemployment Rate (%) Data (Average Based on RMSE and MAPE Only)

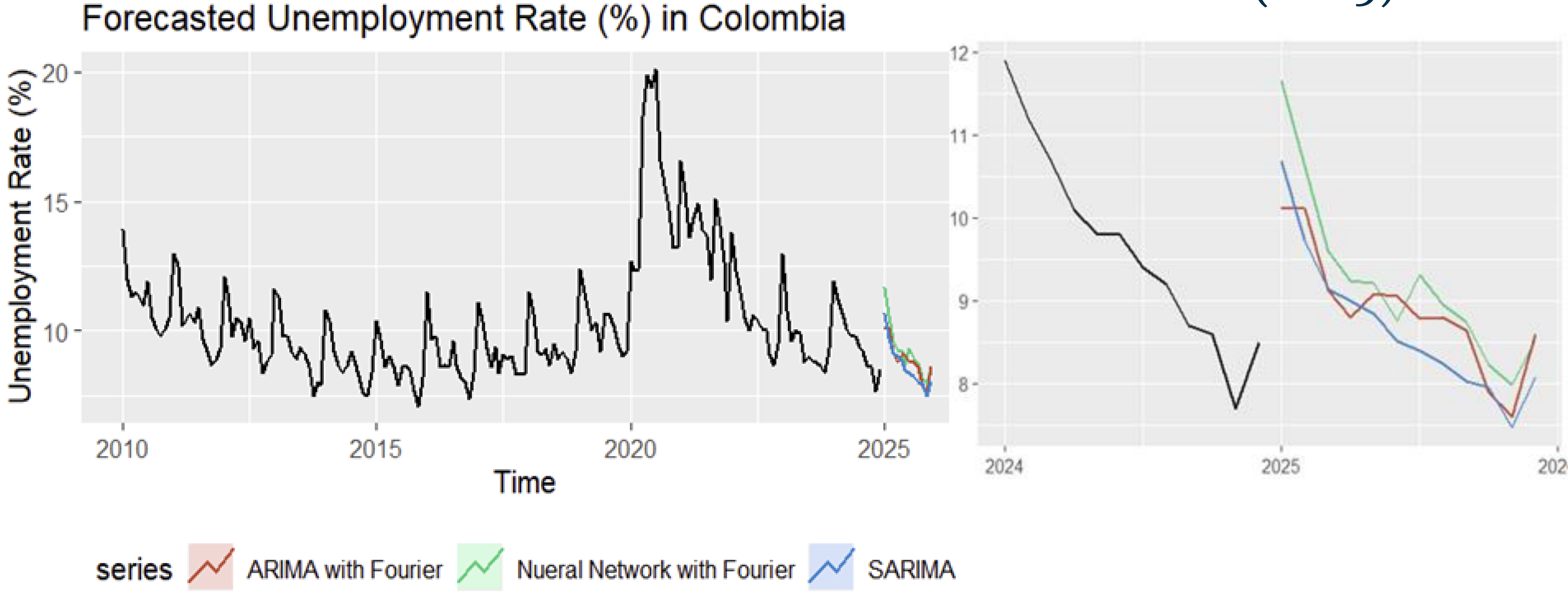
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.04167	0.65000	0.52500	0.21418	5.39401	0.20959	1.08035	3.02201
SMA	-0.71747	1.36577	1.17540	-9.00137	12.97873	0.70623	2.68494	7.17225
SES	-0.71741	1.36573	1.17537	-9.00072	12.97836	0.70623	2.68486	7.17204
SARIMA	0.25130	0.54309	0.45737	2.20043	4.74253	0.57114	0.96922	2.64281
ARIMA	-0.71747	1.36577	1.17540	-9.00137	12.97873	0.70623	2.68494	7.17225
ETS	-0.97220	1.16604	1.03968	-10.74747	11.37813	0.40279	2.24920	6.27209
ARIMA_FOURIER	0.09311	0.53181	0.42178	0.43652	4.32257	0.42562	0.84523	2.42719
TBATS	-0.26941	0.55797	0.44981	-3.28226	4.95462	0.38850	1.09180	2.75629
NNETAR	0.05220	0.42944	0.35110	0.33193	3.70552	0.02635	0.77632	2.06748
SSES	-0.34560	0.56963	0.48211	-3.98432	5.25354	0.15949	1.08906	2.91159
BSM	1.06391	1.16751	1.06391	10.87847	10.87847	0.22363	2.03562	6.02299
Average of 3	0.13220	0.43645	0.35979	0.98963	3.75271	0.46146	0.78750	2.09458

The dataset spans July 2001 to December 2024 from ILOSTAT
Training: Jan 2010 – Dec 2023 | Validation: Jan 2024 – Dec 2024
Due to two breaks in series(Jan 2007, Jan 2010) with missing data (Jul–Dec 2006)

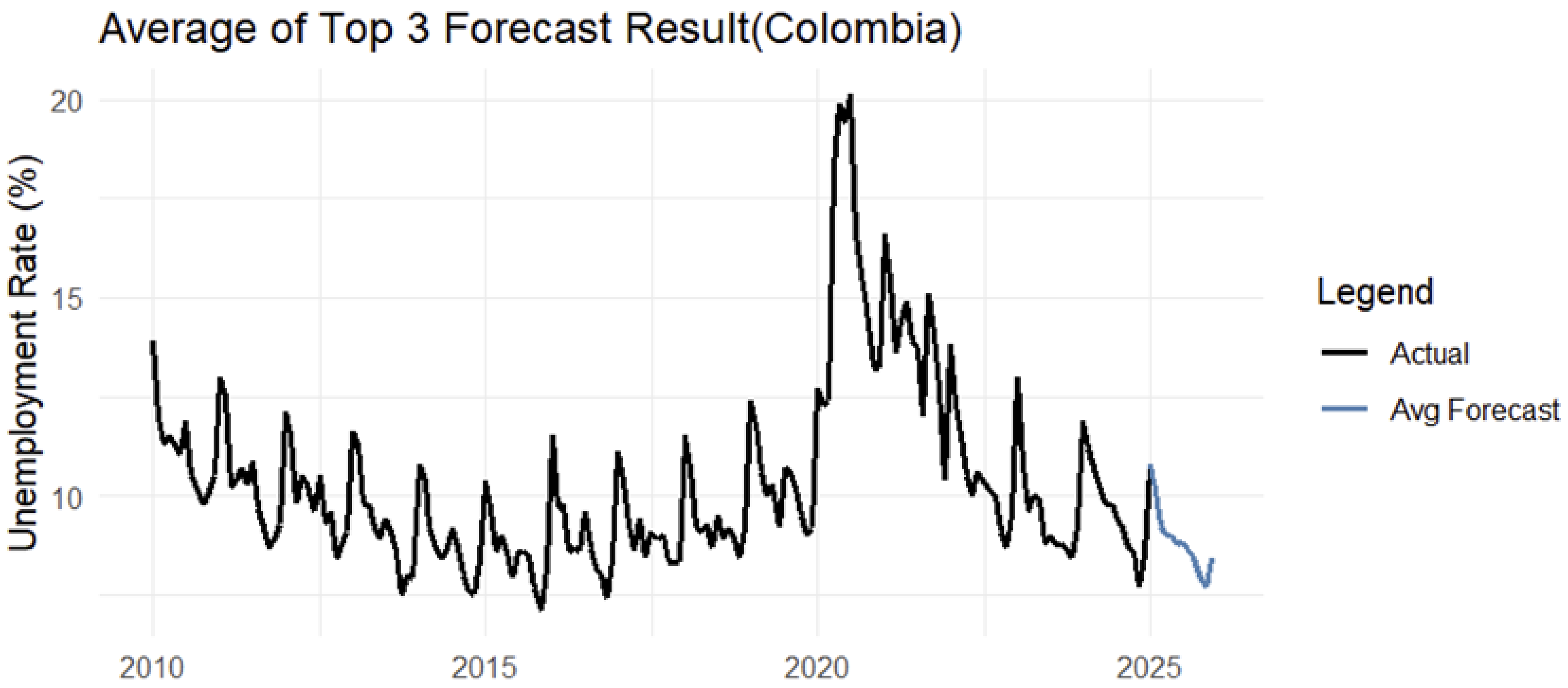
Top 3 Models (Average of RMSE and MAPE)

1. **Neural Network With Fourier**
2. **ARIMA with Fourier**
3. **SARIMA**

Forecast Result of Unemployment Rate in the Colombia(2025)

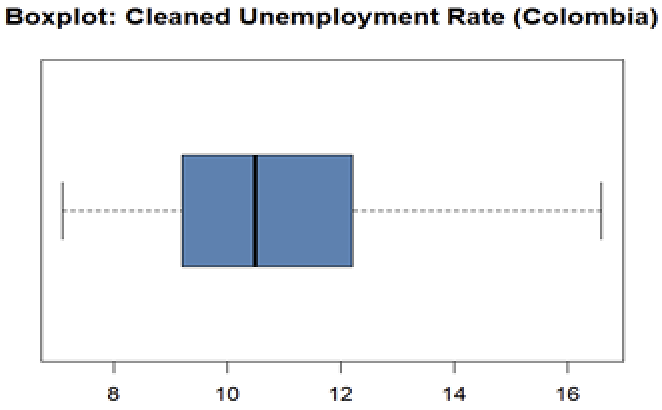


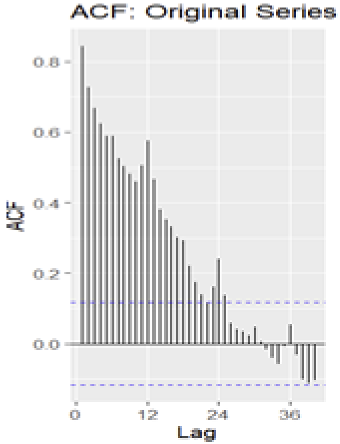
All models show a similar seasonal pattern with downward trend in 2025
NN with Fourier shows the highest forecasts with clear seasonality.
ARIMA with Fourier gives slightly lower values, capturing seasonality well.
SARIMA gives smoother forecasts, lower in mid-year months.

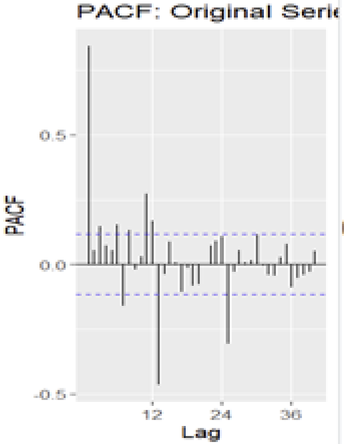


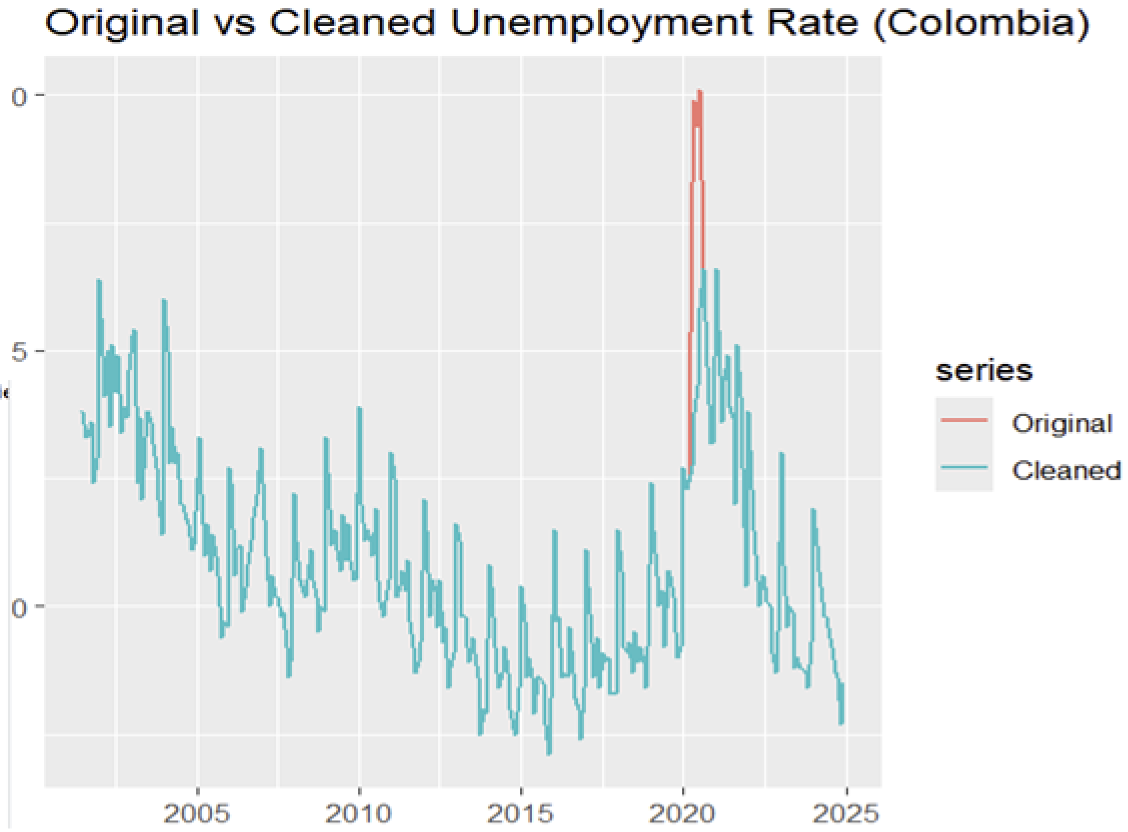
Monthly estimates for 2025 range from 7.5% to 11.6%.
Annual average unemployment rate for 2025 is projected at 7.7%.
Forecast continues the post-COVID decline in unemployment.
Seasonal fluctuations remain visible, with expected end-year increases.

Unemployment Rate in the Colombia – Outliers removed Series

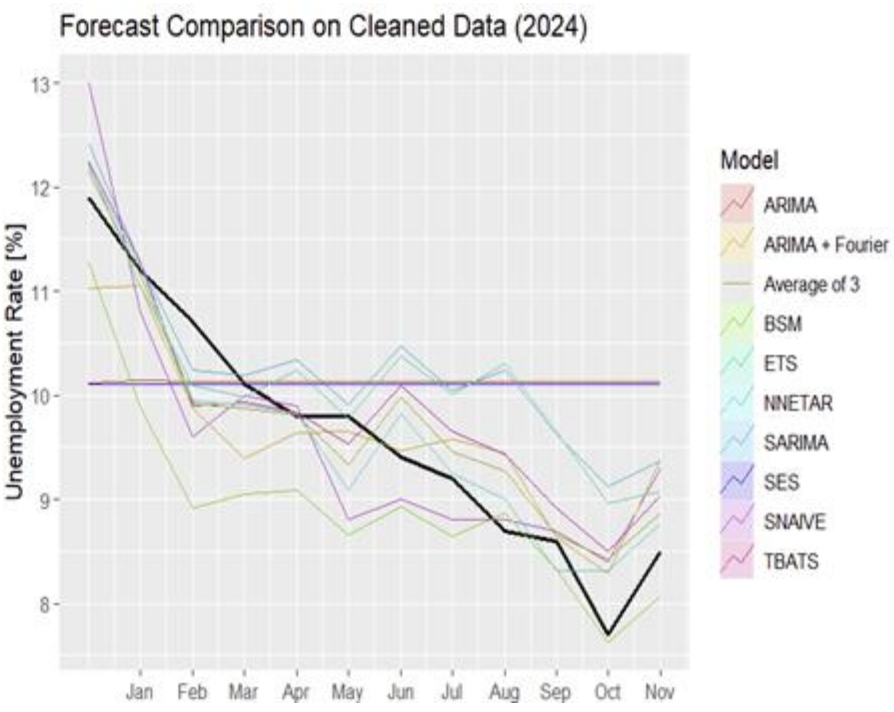








Test Result in the Colombia(2024, Outlier removed Series)



Forecast Accuracy on Cleaned Data (Average Based on RMSE and MAPE Only)

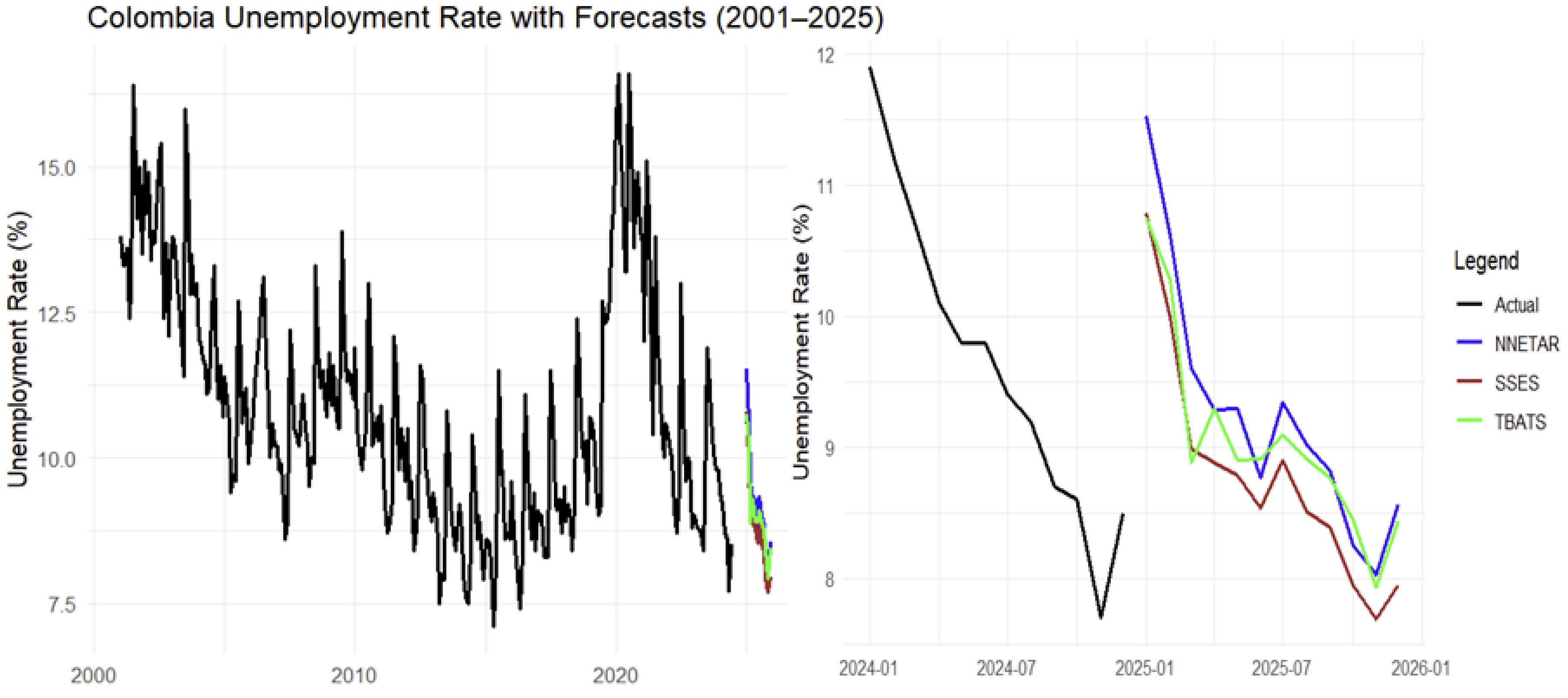
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.042	0.650	0.525	0.214	5.394	0.210	1.080	3.022
SMA	-0.583	1.300	1.108	-7.582	12.159	0.706	2.501	6.729
SES	-0.467	1.252	1.050	-6.362	11.454	0.706	2.352	6.353
SARIMA	-0.635	0.857	0.711	-7.239	7.954	0.574	1.708	4.405
ARIMA	-0.500	1.267	1.068	-6.712	11.668	0.705	2.394	6.467
ETS	-0.512	0.811	0.661	-5.978	7.378	0.613	1.609	4.095
ARIMA + Fourier	0.014	0.566	0.464	-0.438	4.870	0.479	1.002	2.718
TBATS	-0.229	0.510	0.435	-2.761	4.742	0.392	0.977	2.626
NNETAR	0.069	0.410	0.323	0.551	3.391	-0.037	0.734	1.901
SSES	-0.086	0.462	0.387	-1.288	4.175	0.349	0.872	2.319
BSM	0.689	0.868	0.718	6.783	7.108	0.562	1.446	3.988
Average of 3	-0.082	0.433	0.355	-1.166	3.814	0.234	0.807	2.123

The dataset spans July 2001 to December 2024 from ILOSTAT
Training: Jan 2010 – Dec 2023 | Validation: Jan 2024 – Dec 2024
Due to two breaks in series(Jan 2007, Jan 2010) with missing data (Jul–Dec 2006)

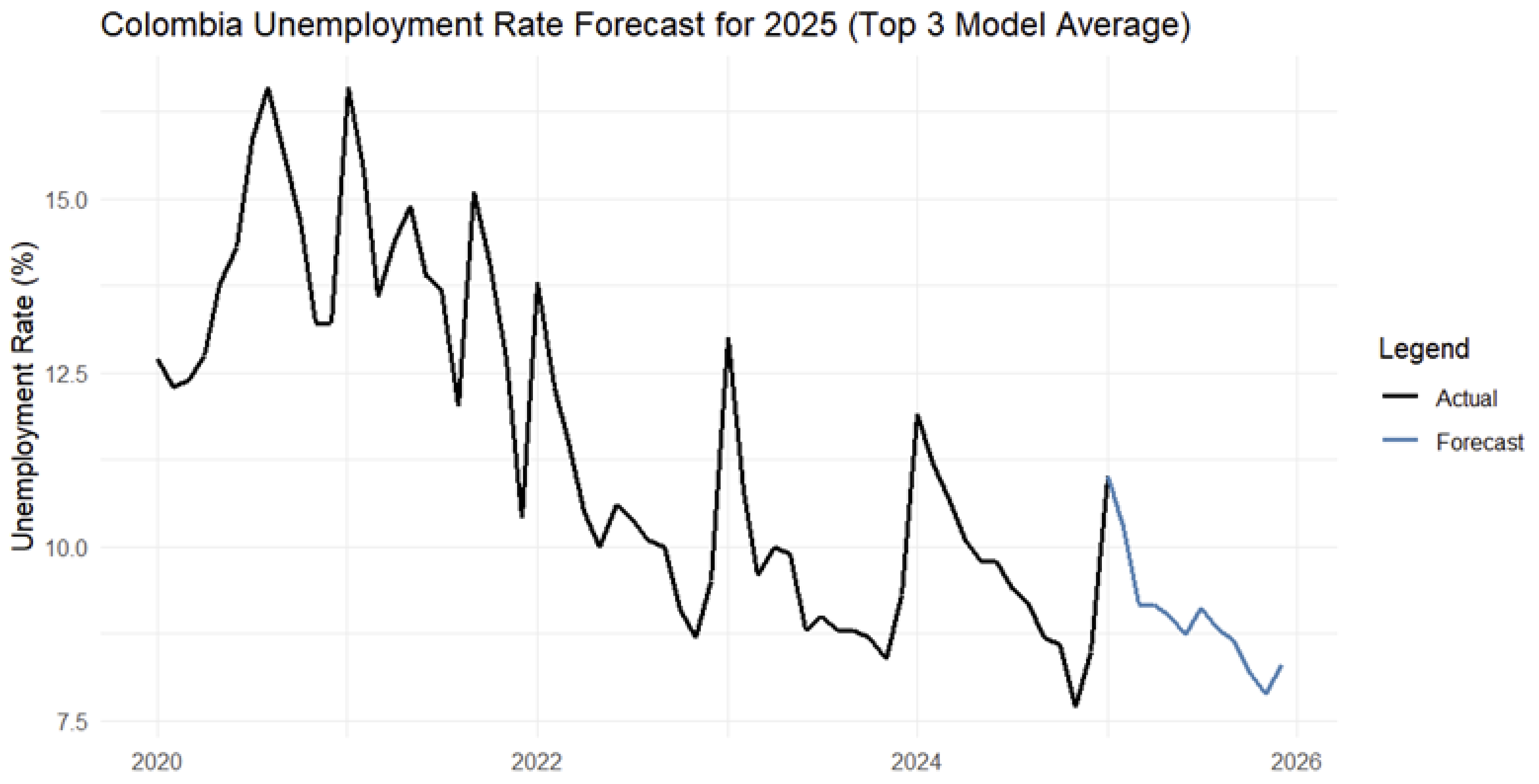
Top 3 Models (Average of RMSE and MAPE)

1. **Neural Network With Fourier**
2. **SSES**
3. **TBATS**

Forecast of Unemployment Rate (2025, Outlier removed))

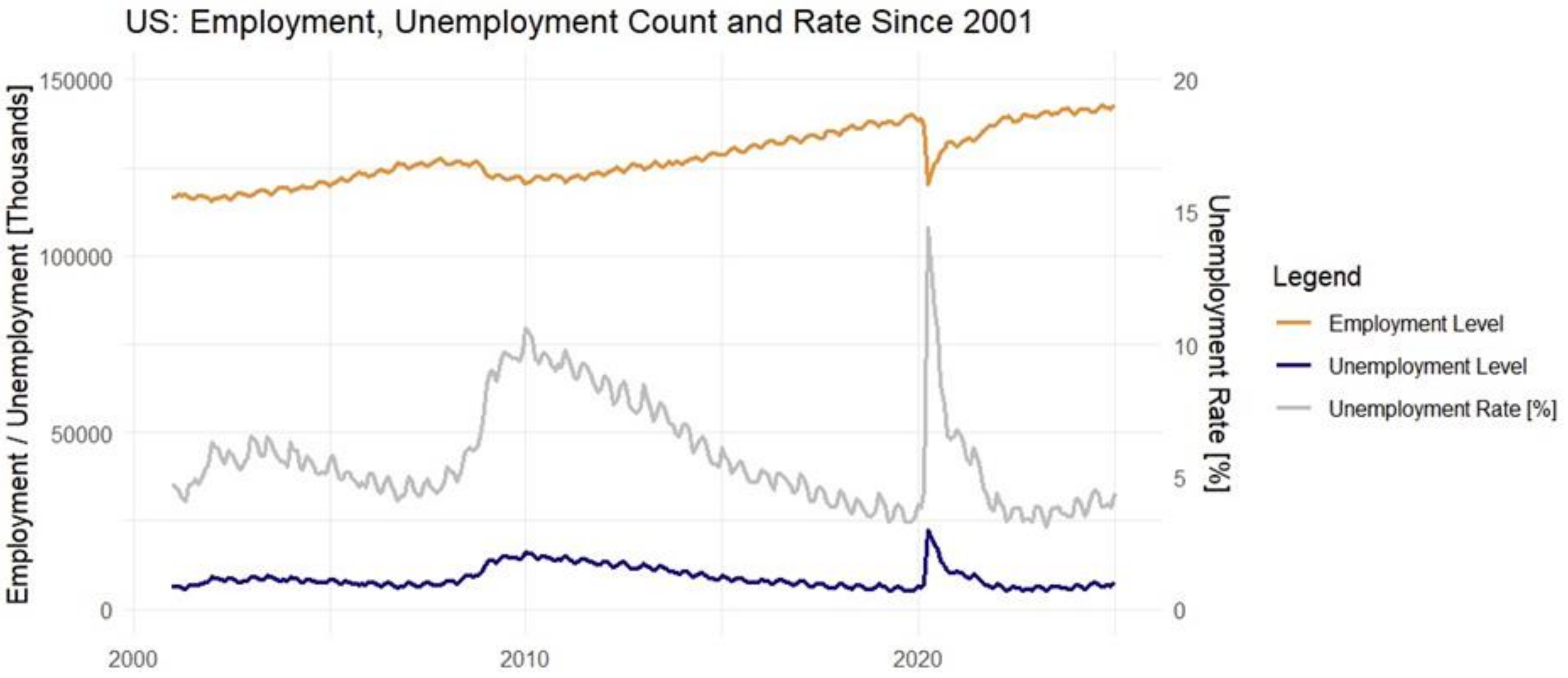


All models show a similar seasonal pattern with downward trend in 2025



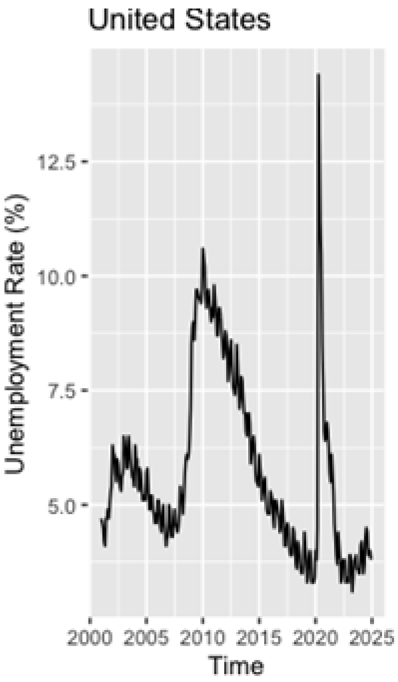
Monthly estimates for 2025 range from 7.8% to 11.0%.
Annual average unemployment rate for 2025 is projected at 9.0%.
Forecast continues the post-COVID decline in unemployment.
Seasonal fluctuations remain visible, with expected end-year increases.

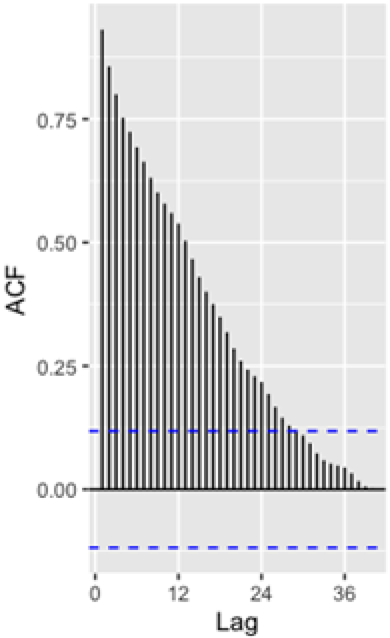
Forecasting Unemployment Rate in the United States

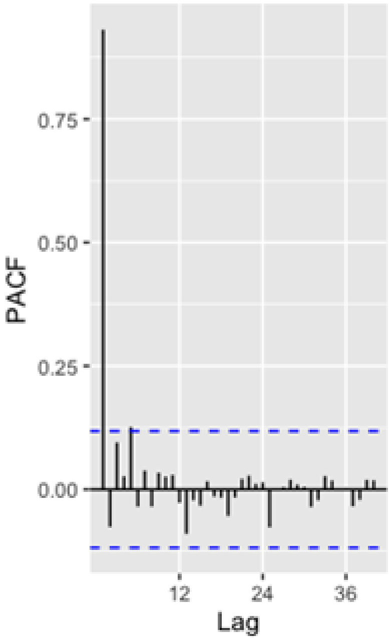


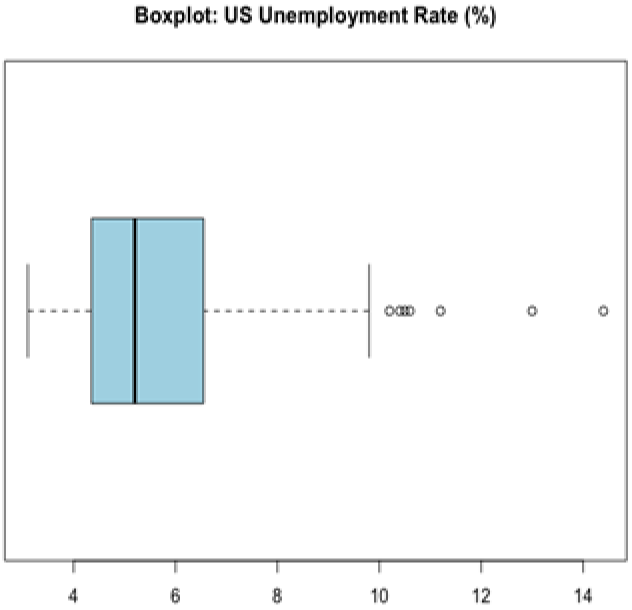
- Duration: January 2001 to December 2024
- Raw Data: International Labour Organization (ILO)
- Global Financial Crisis in 2008 and a slow recovery identified onwards
- COVID-19 induced spike in 2020 and a sharp recovery onwards

Forecasting Unemployment Rate in the US - Initial Plots



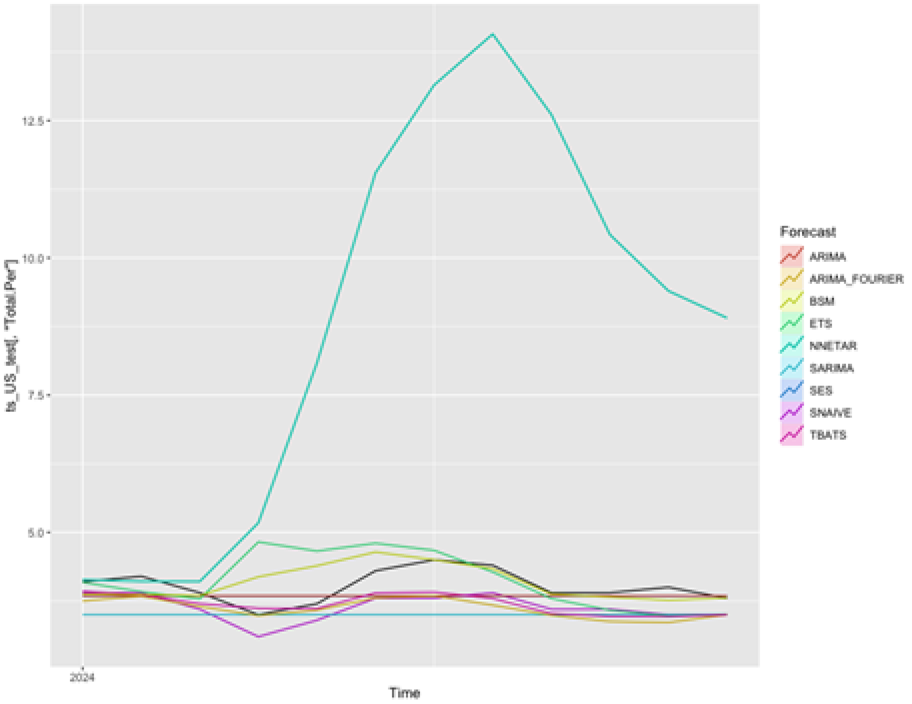






- Strong autocorrelation is identified the ACF and PACF plots.
- Evidence of Unit Root/ Stochastic (ADF Test: p-value at 0.43)
- Evidence of deterministic trend (Seasonal MK: tau at -0.29 and p-value < 0.0001)
- Evidence of Outliers (Box Plot and also Grubbs Test)
    - Both Original Series and Outlier-removed Series will be tested.

Forecasting Unemployment Rate in the US – Original Series



Forecast Accuracy for Unemployment Rate (%) (Average based on RMSE and MAPE)

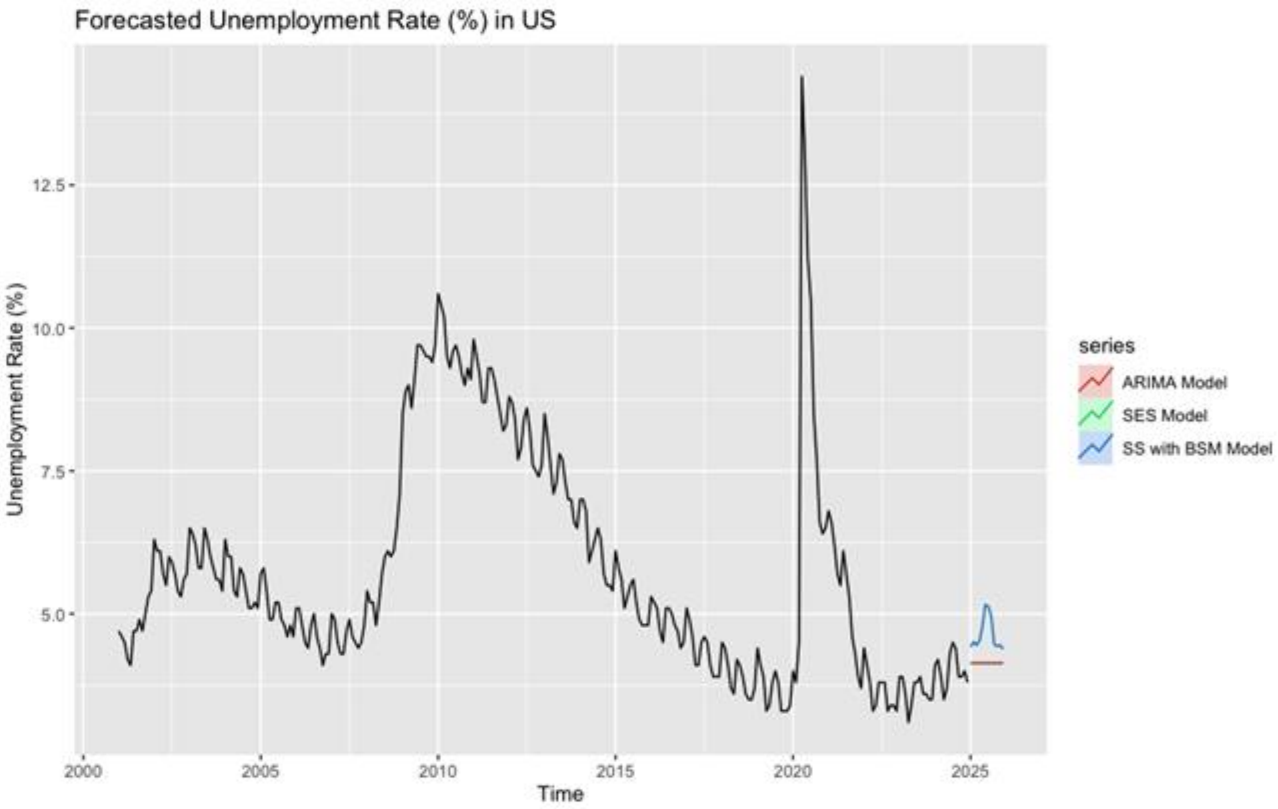
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.38333	0.40620	0.38333	9.46470	9.46470	0.28333	1.35284	4.93545
SMA	0.17216	0.33070	0.26108	3.80965	6.29631	0.44396	1.06312	3.31350
SES	0.17215	0.33070	0.26108	3.80953	6.29625	0.44396	1.06311	3.31347
SARIMA	0.51667	0.58878	0.51667	12.42930	12.42930	0.44396	1.85999	6.50904
ARIMA	0.17216	0.33070	0.26108	3.80965	6.29631	0.44396	1.06312	3.31350
ETS	-0.09918	0.54307	0.39400	-2.91283	10.30378	0.55949	1.95874	5.42342
ARIMA_FOURIER	0.40391	0.45420	0.40391	9.80344	9.80344	0.50899	1.46803	5.12882
TBATS	0.32666	0.38544	0.34616	7.88226	8.43937	0.53705	1.26470	4.41240
NNETAR	-5.43114	6.79032	5.44900	-132.66718	133.09243	0.81002	22.36934	69.94137
SSES	1.52521	1.71249	1.52948	38.05395	38.15815	0.73123	5.77401	19.93532
BSM	-0.06016	0.32935	0.22841	-1.82484	5.93544	0.57974	1.18115	3.13239

- Training Window: Jan-2001 to Dec-2023
- Testing Window: Jan-2024 to Dec-2024
- No missing values in the data set
- NN failing to capture a very high spike in the data set

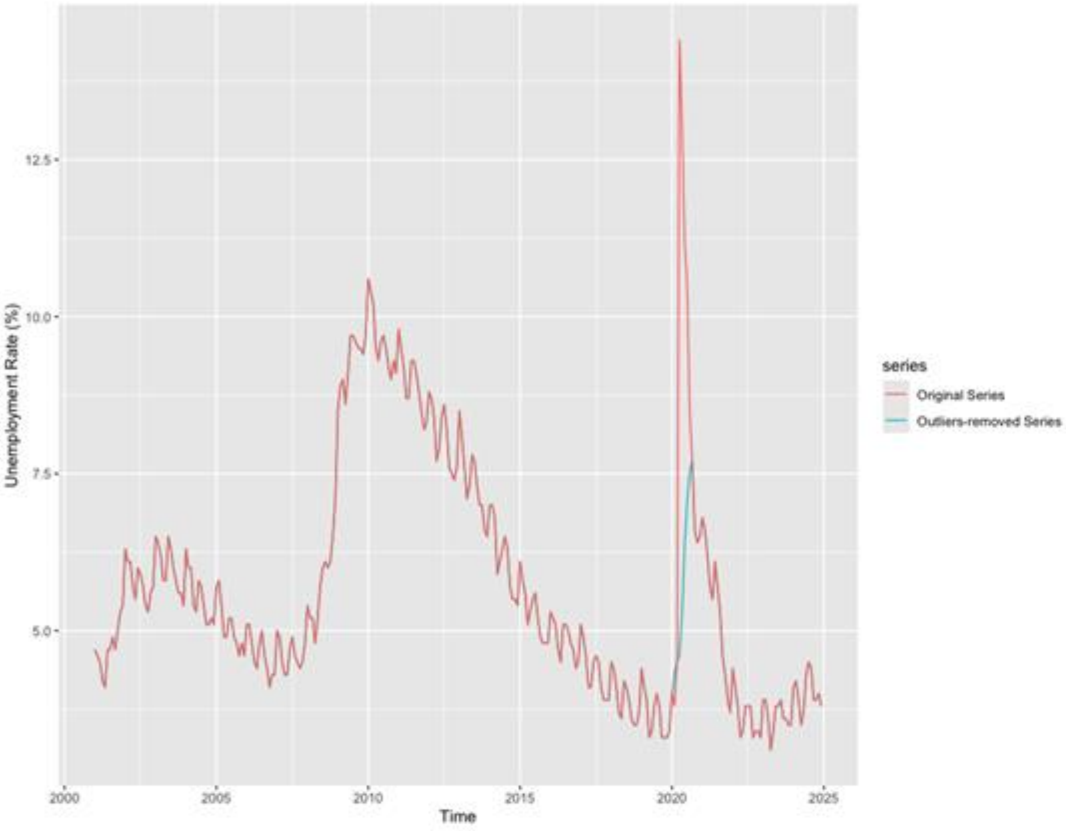
Top 3 Models

1. State Space with BSM
2. SES
3. ARIMA

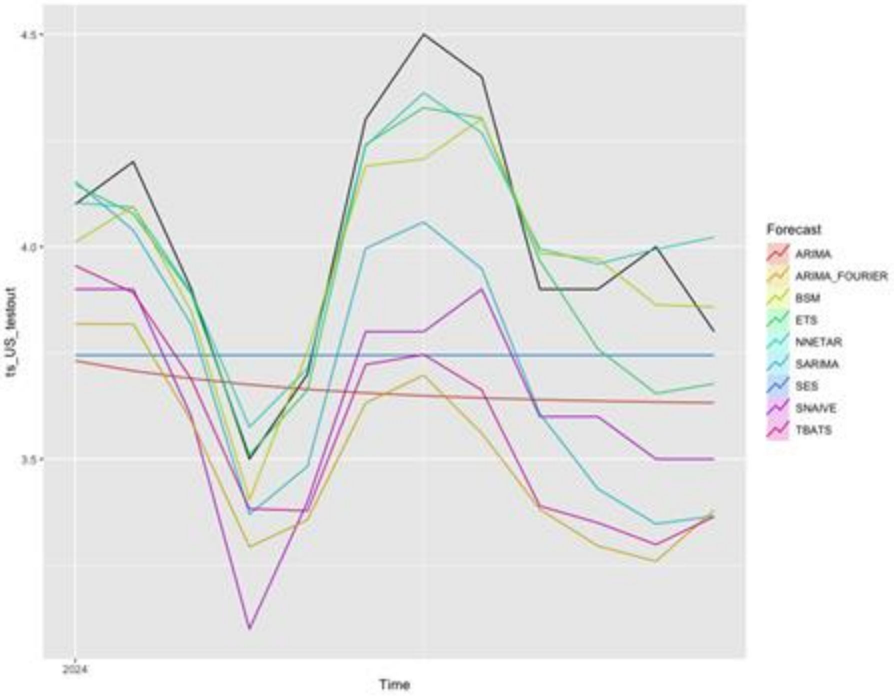
Forecasting Unemployment Rate in the US – Original Series



Forecasting Unemployment Rate in the US - Outliers-removed Series



Forecasting Unemployment Rate in the US - Outliers-removed Series



Forecast Accuracy for Unemployment Rate (%) (Average based on RMSE and MAPE)

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.38333	0.40620	0.38333	9.46470	9.46470	0.28333	1.35284	4.93545
SMA	0.27186	0.39196	0.32013	6.30418	7.67176	0.44396	1.24293	4.03186
SES	0.27185	0.39195	0.32013	6.30406	7.67168	0.44396	1.24292	4.03182
SARIMA	0.29883	0.35509	0.30781	7.39350	7.61242	0.58243	1.20095	3.98375
ARIMA	0.35322	0.45308	0.38245	8.34064	9.17576	0.46153	1.44837	4.81442
ETS	0.08226	0.13550	0.10280	2.00098	2.52297	0.12960	0.46055	1.32923
ARIMA_FOURIER	0.50967	0.55032	0.50967	12.48990	12.48990	0.55403	1.82003	6.52011
TBATS	0.44741	0.49729	0.44741	10.95395	10.95395	0.62666	1.65645	5.72562
NNETAR	0.09010	0.13690	0.10890	2.11306	2.60782	0.17670	0.44949	1.37236
SSES	0.41563	0.47631	0.42271	10.27258	10.44526	0.65134	1.60031	5.46079
BSM	0.05951	0.12113	0.10407	1.37766	2.53923	0.04463	0.39014	1.33018

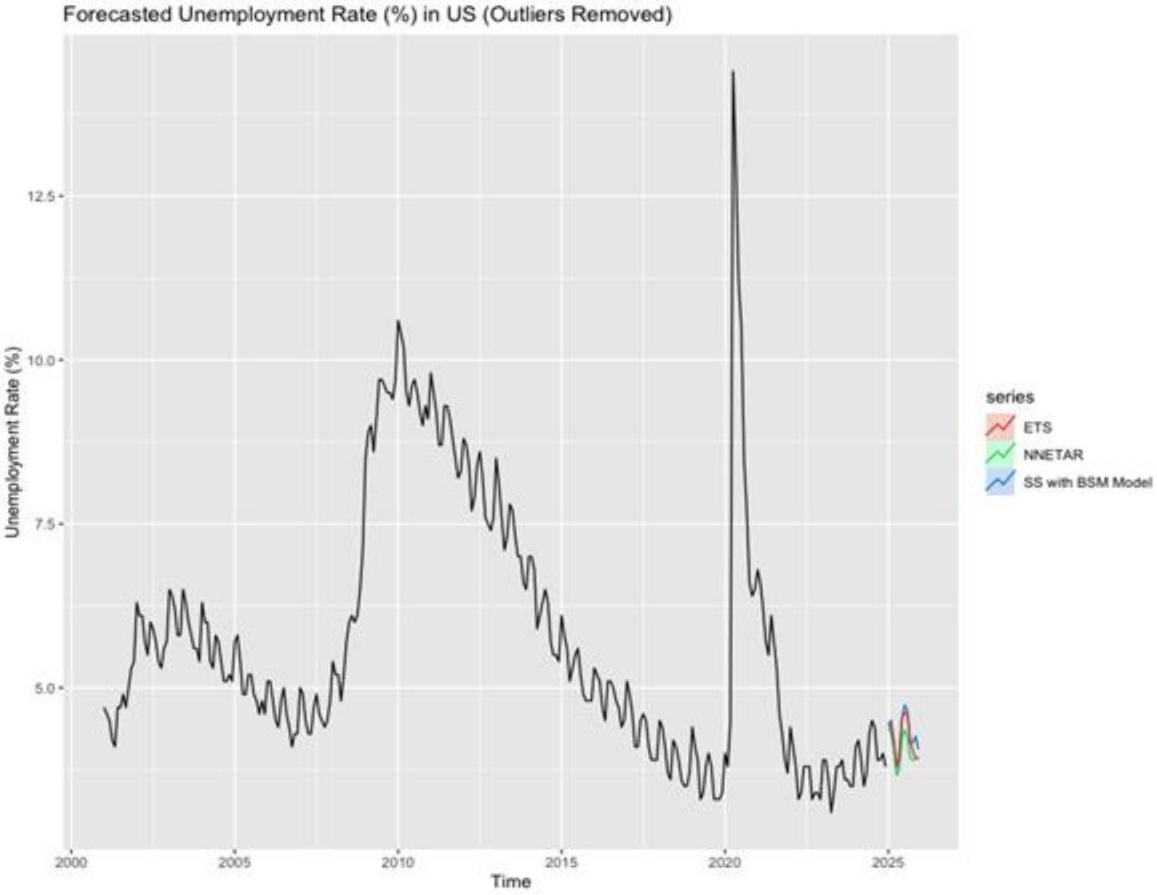
Top 3 Models with Original Series

1. State Space with BSM
2. SES
3. ARIMA

Top 3 Models with Outliers-removed Series

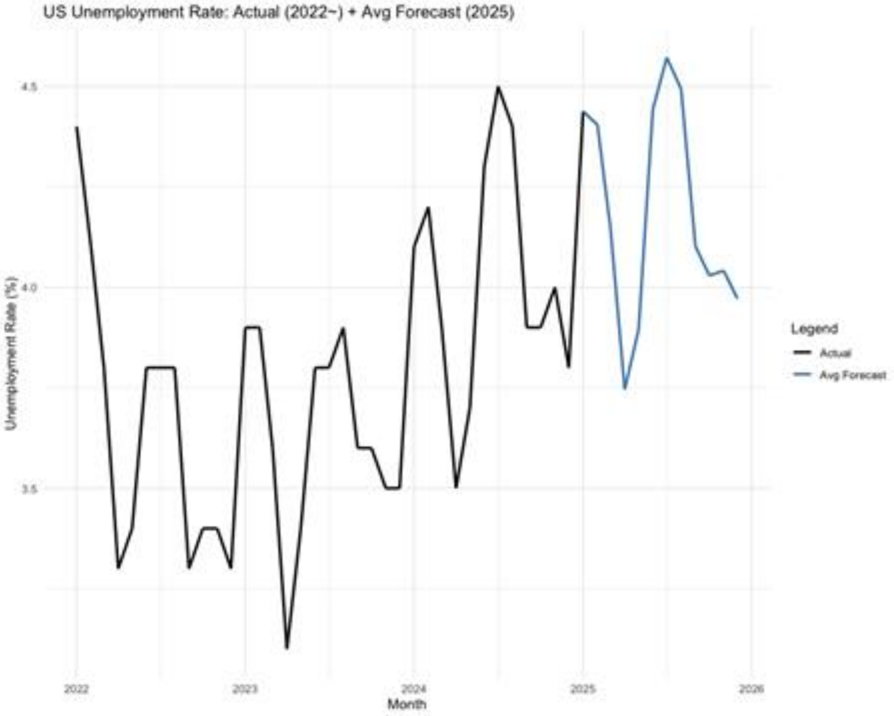
1. State Space with BSM
2. Neural Network
3. ETS

Forecasting Unemployment Rate in the US - Outliers-removed Series



- All models could capture both seasonality and trends.
- NN slightly underestimates.

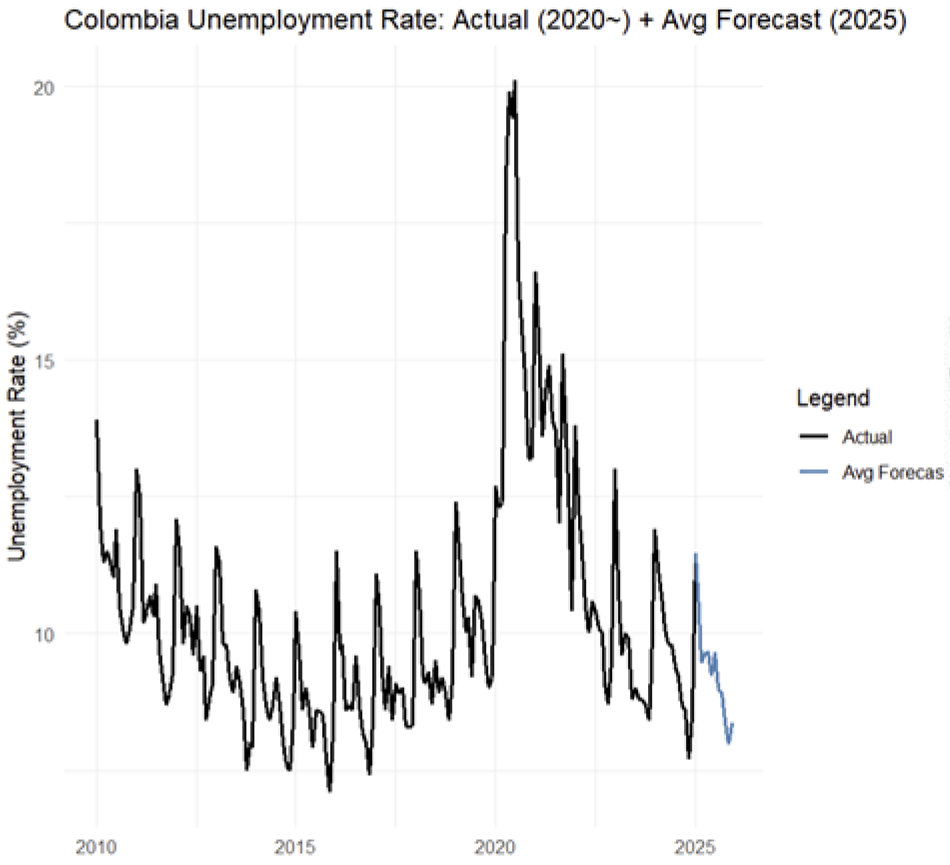
Forecasting Unemployment Rate in the US - Outliers-removed Series

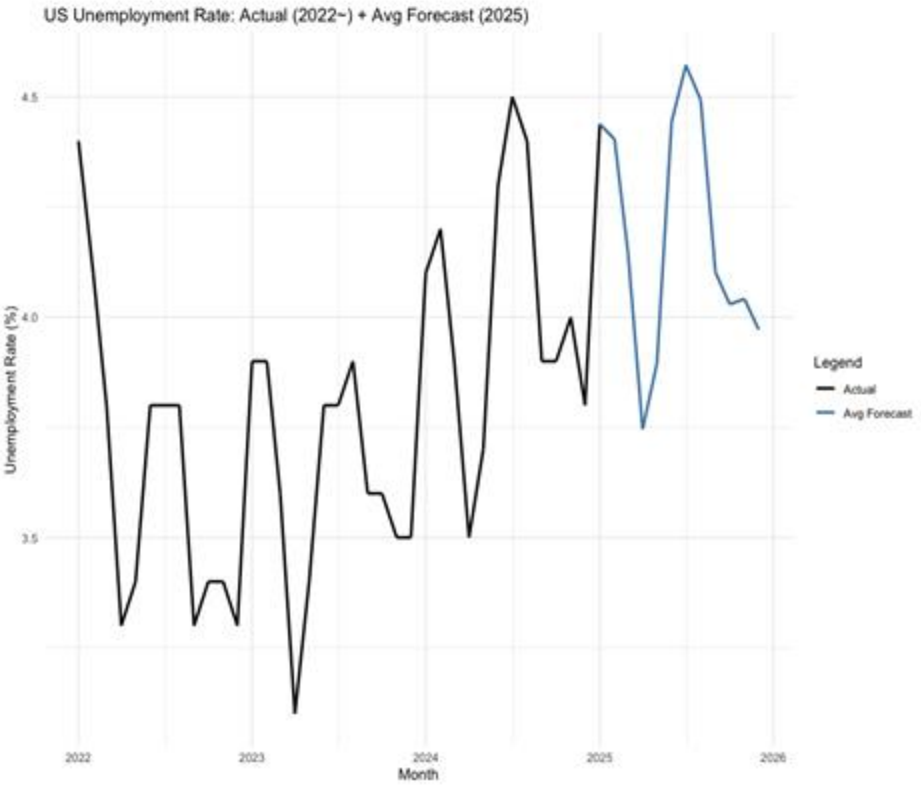


ILO Yearly Estimate for 2025: 4.3%

- Monthly Estimate for 2025: 3.75% to 4.54%
- Yearly Average Estimate for 2025: 4.17%
- Forecast able to capture seasonal fluctuations and increasing trend

Unemployment Rate in Colombia Vs. United States





Conclusion, Recommendation and Limitations

Conclusion: Colombia

- Unemployment is higher than the U.S. but shows a downward trend
- GDP growth (~0.6% in 2023) limits labor absorption despite declining unemployment.
- Structural breaks in 2007–2010 (survey change) and 2020 (COVID).
- Labor market is shaped by seasonality, informality, and data disruptions.
- Government efforts to formalize economy: tax reform, 'Colombia Goes Formal'.
- Complex dynamic models [NNAR + Fourier, ARIMA + Fourier, SARIMA] proved to be efficient due to complex non linear seasonality

Key Takeaway: Colombia's unemployment trends require careful modeling due to breaks and informality.

Conclusion: United States

- Unemployment is low and stable, with a clear recovery path post-COVID.
- GDP growth averaged around 2.4% (2023–2024), supporting labor market resilience.
- No major structural break beyond COVID; data is continuous and consistent.
- The U.S. is on the right track, showing a typical unemployment cycle with healthy labor market churn.
- High job openings and quit rates reflect strong worker mobility and confidence.
- The structured labor market and institutional support drive predictability.
- Simpler models (ETS, BSM) were sufficient due to minimal seasonality and stable trends.

Key Takeaway: The U.S. labor market supported effective forecasting using consistent, low-complexity models.

Recommendation : Colombia

- Continuation of the policies taken
- Reduce non-wage labor costs.
- Expand and enforce formalization programs.
- Target youth and women in employment policies.
- Improve data continuity; adjust for breaks.

Key Takeaway: Reducing unemployment greatly depends on reducing informality

Recommendation: USA

Preserve conditions that enable labor mobility and job matching efficiency.

Limitations

- Modeled total unemployment only — skipped disaggregation by age/gender for simplicity
- Informality not directly captured, especially relevant for Colombia.
- Colombia's structural breaks (2007–2010, 2018, COVID-19) disrupted trend continuity.
- Nonlinear seasonality in Colombia required adaptive models, limiting comparability with the U.S.
- Exogenous variables (GDP, inflation) were not included:
  - Unemployment is often leading, while these are lagging.
  - Including them would require forecasting additional series, adding complexity.
- COVID shock treated as an outlier rather than a dummy — better for modeling underlying trend behavior

Thank You!

Questions and Answers