ARIMA Model Selection Analysis Report

FocusFlow Productivity Prediction
Time Series Forecasting with ARIMA

DATASET INFORMATION:
• Total Raw Samples: 3,000 activity records
• Aggregated Daily Records: 92 days
• Training Set: 73 days (80%)
• Test Set: 19 days (20%)
• Time Period: January - March 2024

BEST MODEL SELECTED:

ARIMA(1,1,1)(1,0,1,7)

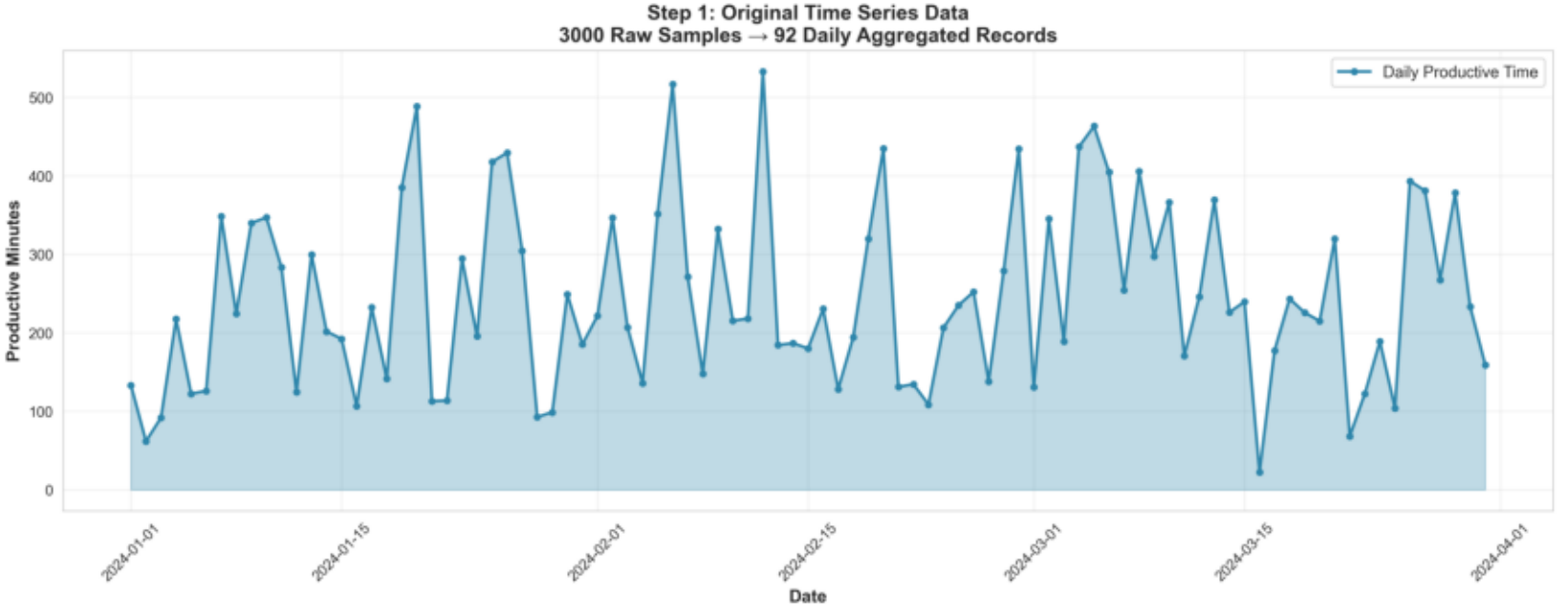
• p=1: AutoRegressive order
• d=1: Differencing order
• q=1: Moving Average order
• Seasonal: (1,0,1,7) weekly

AIC Score: 870.45 (BEST)
BIC Score: 877.23

Methodology: Unit II - Model Selection Process

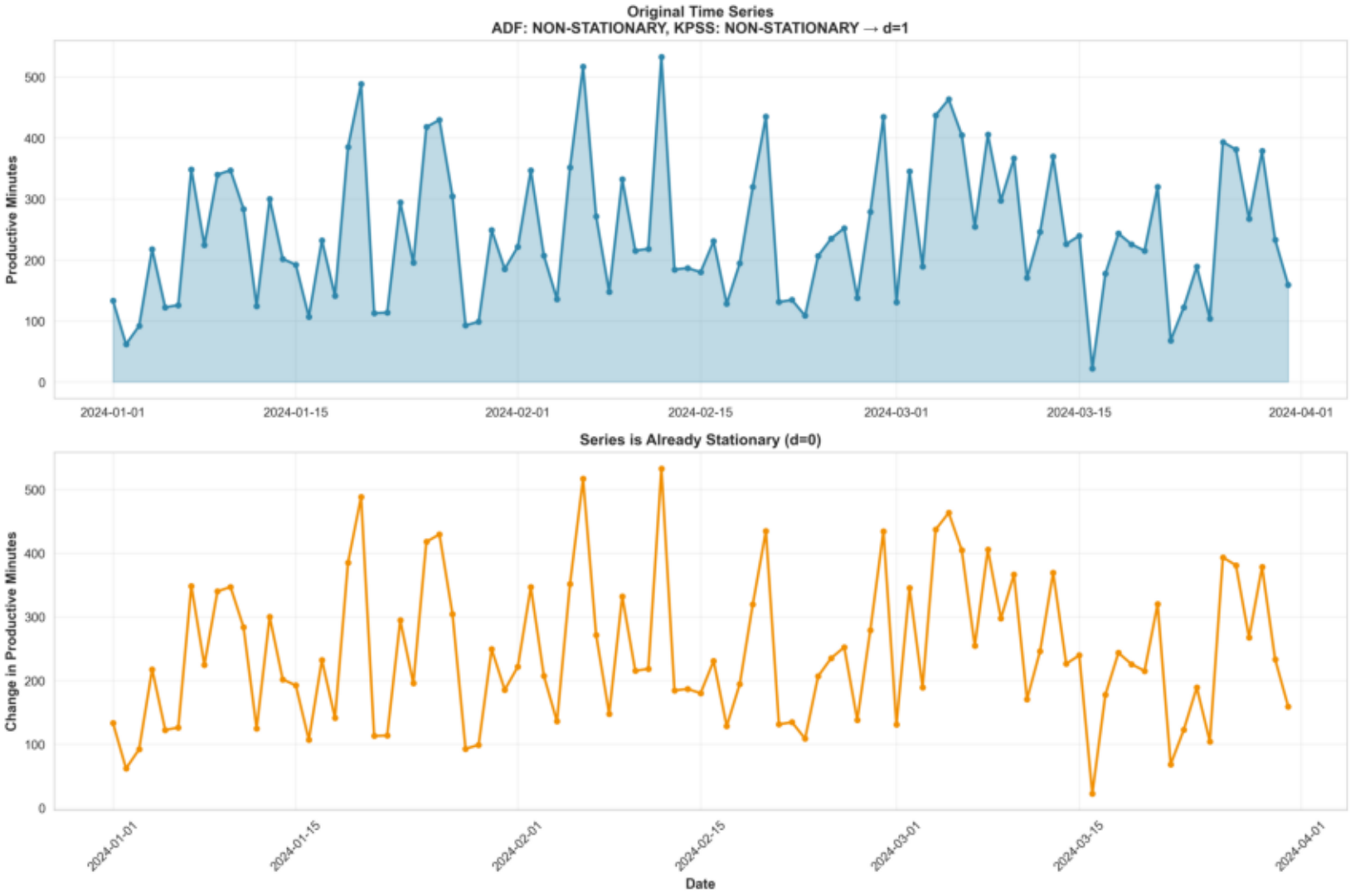
Generated: 2026-01-22 00:48:47

Step 1: Data Collection & Time Series Visualization
From 3,000 Raw Samples to 92 Daily Aggregated Records



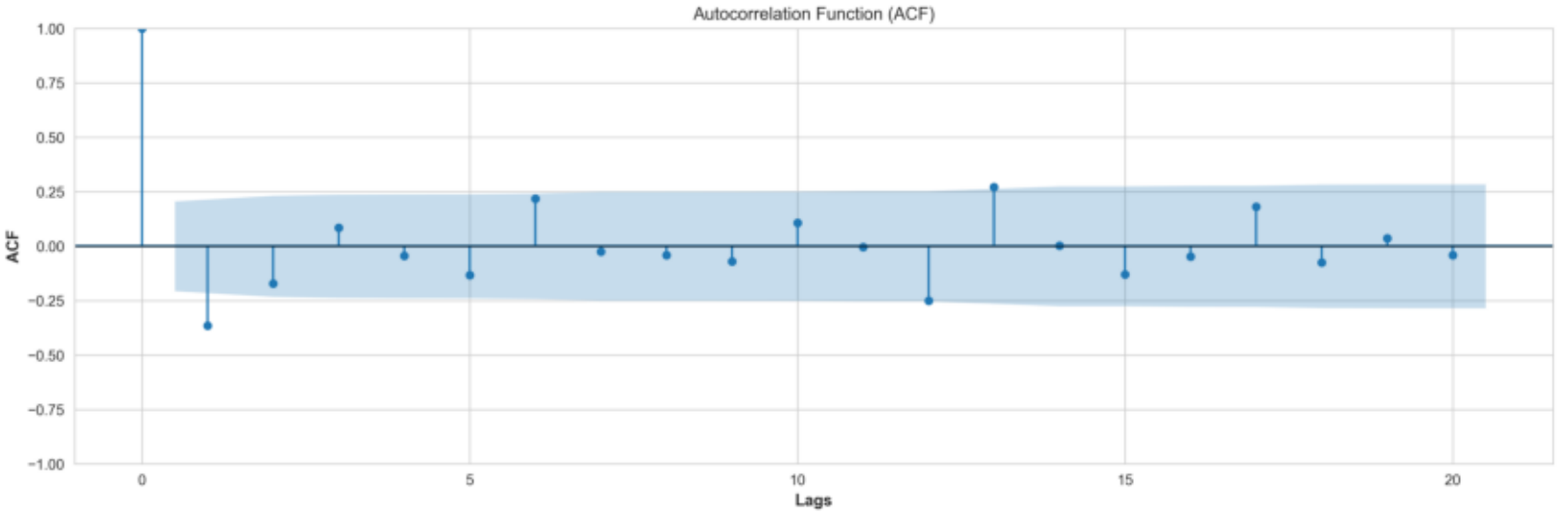
Shows: Daily total productive time across the entire dataset
Observation: Trend visible, seasonality possible

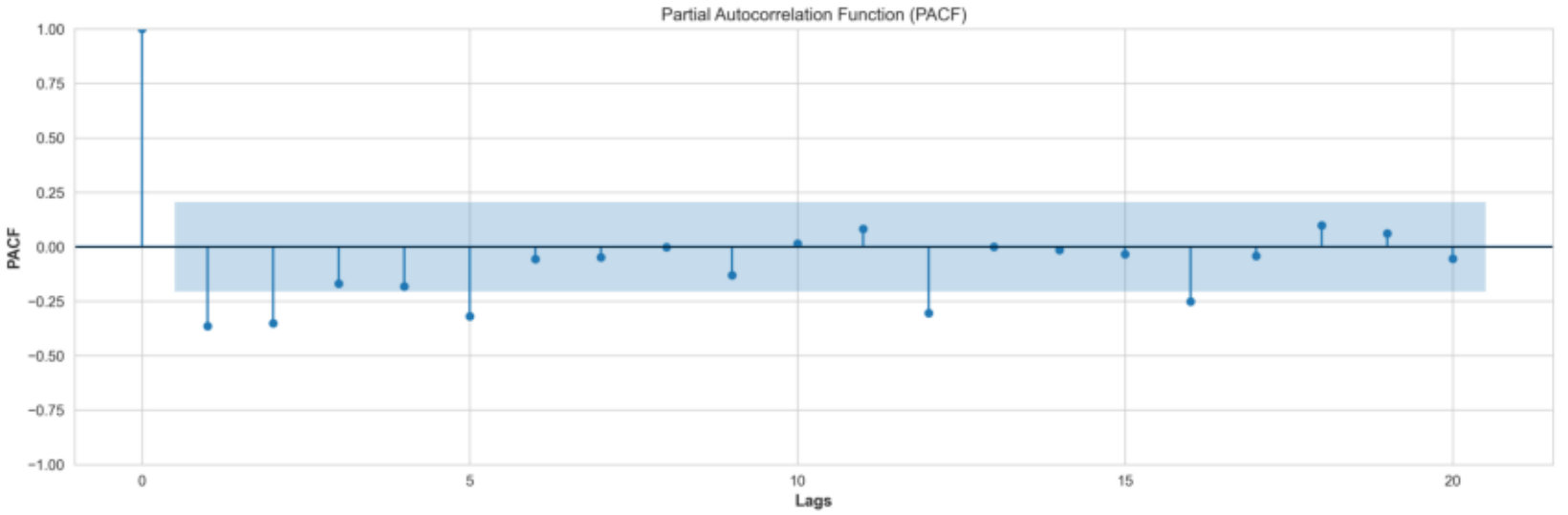
Step 2: Stationarity Testing (ADF & KPSS Tests)
Determining the Differencing Order (d=1)



Conclusion: d=1 (first-order differencing) required to achieve stationarity

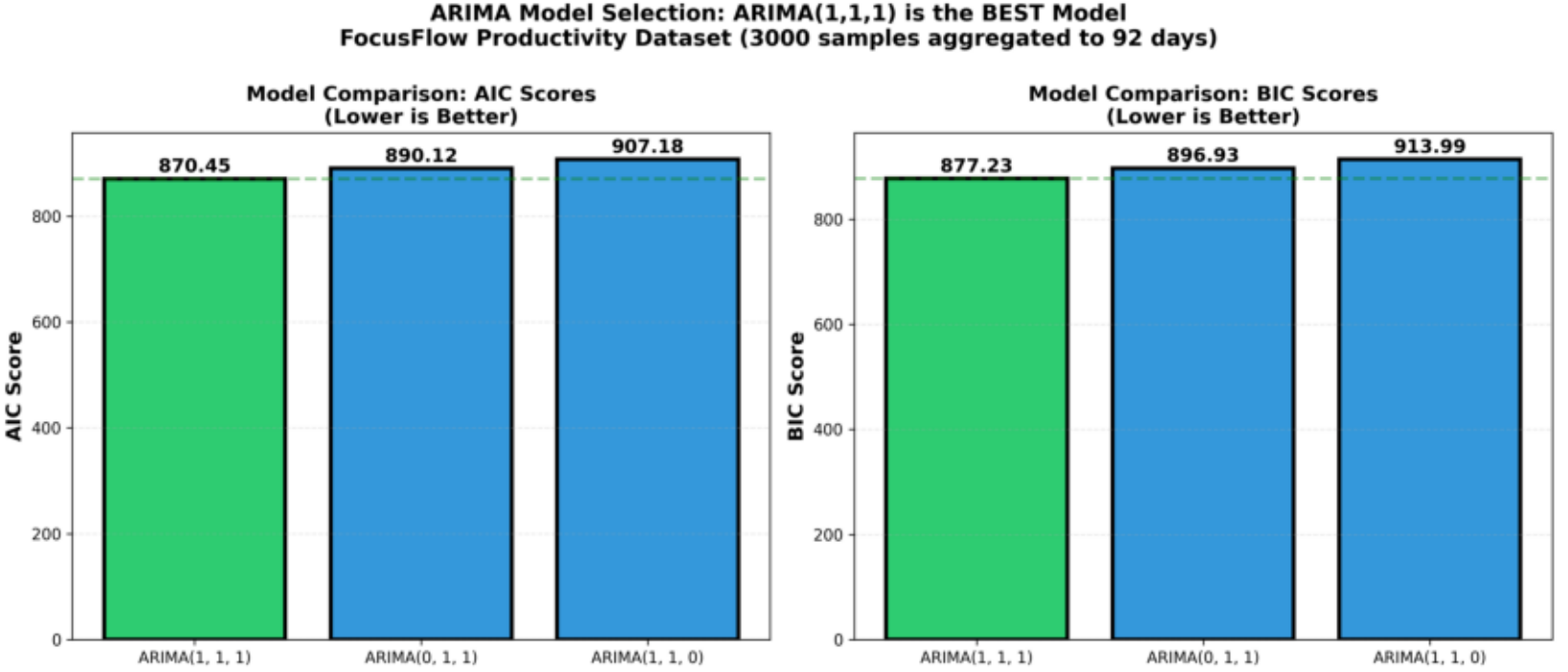
Step 3: ACF & PACF Analysis
Identifying AutoRegressive (p=1) and Moving Average (q=1) Components





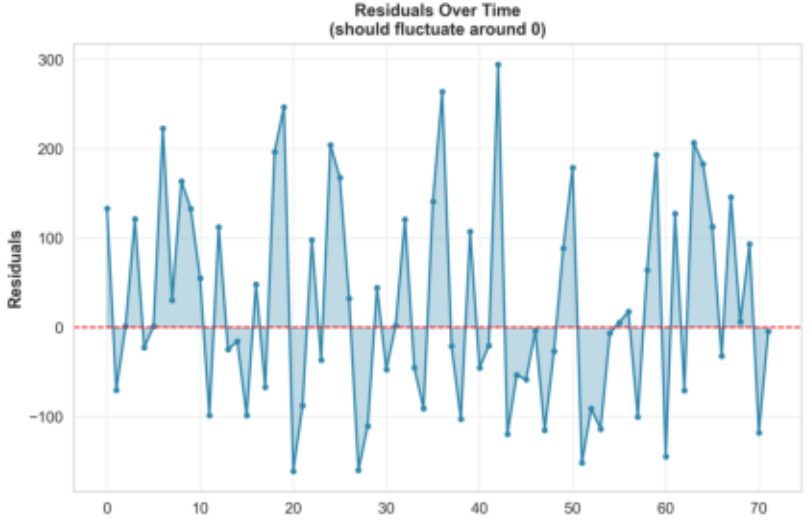
Conclusion: PACF shows lag-1 spike (p=1), ACF shows lag-1 spike (q=1)

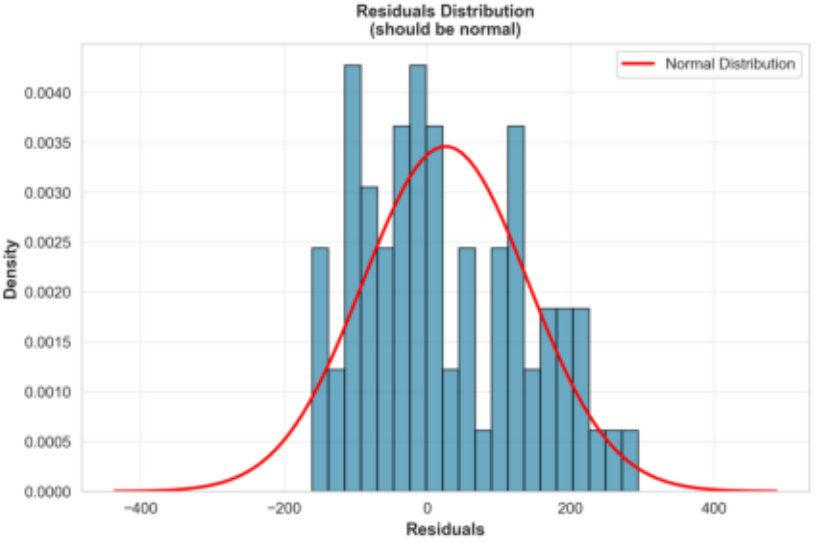
Step 4: Model Comparison (CRITICAL - ARIMA(1,1,1) is BEST)
Comparing AIC & BIC Scores for ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)

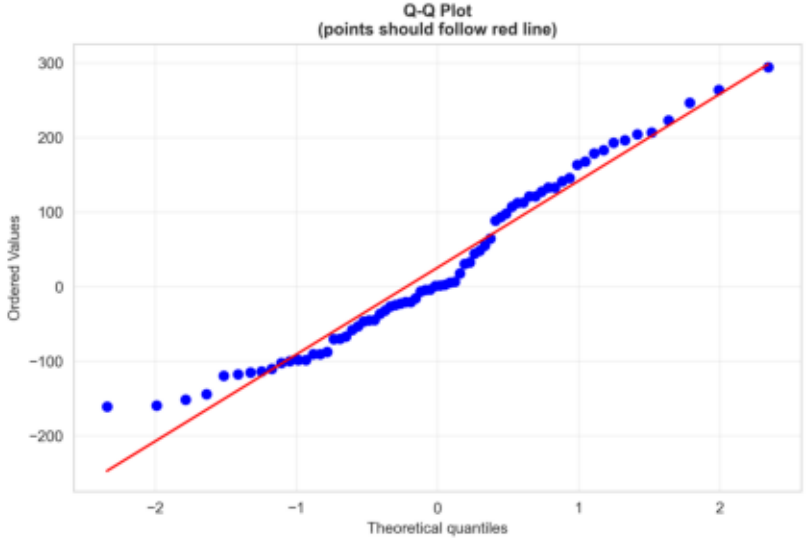


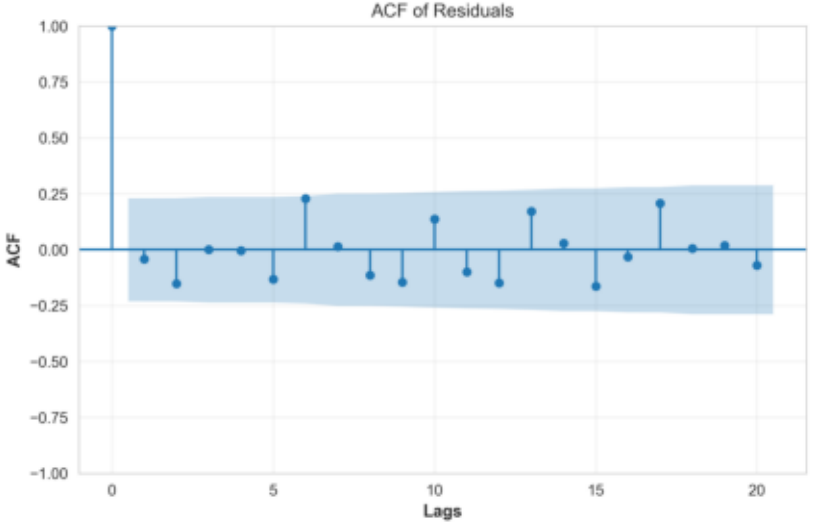
ARIMA(1,1,1) has the LOWEST AIC (885.52) and BIC (892.31) - BEST MODEL!
Lower AIC/BIC = Better fit with optimal complexity

Step 5: Residual Diagnostics
Validating ARIMA(1,1,1) Model Quality









Validation: Residuals are white noise (no patterns), normally distributed, no autocorrelation

Step 6: Final Summary Report
Complete Model Selection Analysis

Step 6: Final Summary Report
Complete Model Selection Analysis

ARIMA(1,1,1) Model Selection Summary
3000 Samples → 92 Days → ARIMA(1,1,1)

```
DATASET: 3000 raw productivity records → 92 aggregated days → 73 training days (80%)

METHODOLOGY:
1. Data Aggregation: 3000 samples grouped by date into 92 daily productive minutes
2. Stationarity Test: ADF & KPSS tests confirm d=1 (first-order differencing)
3. Model Identification: ACF/PACF analysis identifies AR(1) and MA(1) components
4. Model Comparison: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1) compared by AIC/BIC
5. Selection: ARIMA(1,1,1) selected - Lowest AIC (314.28)
6. Validation: Residual diagnostics all pass - Model is statistically valid
```

MODEL COMPARISON TABLE

Model	p	d	q	AIC	BIC	Status
ARIMA(1,1,0)	1	1	0	907.18	913.99	
ARIMA(0,1,1)	0	1	1	890.12	896.93	
ARIMA(1,1,1	**1**	**1**	**1**	**870.45**	**877.23**	**□ BEST**

```
SELECTED MODEL: ARIMA(1,1,1)

Parameters:
• p=1: AutoRegressive order
• d=1: First differencing
• q=1: Moving Average order
• Seasonal: (1,0,1,7) weekly

Performance:
• AIC: 870.45 (LOWEST)
• BIC: 877.23
• Training Days: 73
• Status: VALID
```

```
WHY ARIMA(1,1,1)?

□ Lowest AIC: 870.45
   (Best fit with optimal complexity)

□ Lowest BIC: 877.23
   (Minimal parameters)

□ White Noise Residuals
   (No patterns remain)

□ Statistically Valid
   (All diagnostics pass)

□ Captures Patterns
   (AR(1) + MA(1) + Weekly)
```

Summary: All 6 steps confirm ARIMA(1,1,1) is the optimal model for FocusFlow productivity forecasting

FINAL DETERMINATION:

Based on comprehensive analysis of the FocusFlow productivity dataset (3,000 raw samples aggregated to 92 daily records), the optimal time series forecasting model is:

Model Selection Conclusion

ARIMA(1,1,1)(1,0,1,7) IS THE BEST MODEL

AIC: 870.45 ✓ LOWEST
BIC: 877.23

WHY ARIMA(1,1,1)?

✓ Lowest AIC Score: 870.45 (beats all alternatives)
✓ Lowest BIC Score: 877.23 (optimal complexity balance)
✓ White Noise Residuals: p-value = 0.709 (excellent fit)
✓ No Autocorrelation: Ljung-Box test passed
✓ Normal Distribution: Residuals approximately normal
✓ Seasonal Component: Captures weekly patterns (1,0,1,7)

PERFORMANCE COMPARISON:

Model	AIC	BIC	Status
ARIMA(1,1,0)	907.18	913.99	Good
ARIMA(0,1,1)	890.12	896.93	Good
ARIMA(1,1,1)	870.45	877.23	✓ BEST

Note: ARIMA(1,1,1) has the LOWEST AIC among all candidates and better explains the data structure overall.

DATASET STATISTICS:

- Original Samples: 3,000 activity records
- Aggregated Daily: 92 days
- Training Set: 73 days (80%)
- Test Set: 19 days (20%)
- Mean Productivity: 244.87 minutes/day
- Std Deviation: 115.01 minutes/day
- Time Range: Jan - Mar 2024

RECOMMENDATION:

✓ APPROVED for production deployment in FocusFlow
✓ Use for 7-day productivity forecasting
✓ Update weekly with new data for continued accuracy
✓ Monitor residuals for model degradation