

TIME SERIES FORECASTING OF MAY-JUNE SALES FOR CH-CLEAN WITH R

Welcome to this end-to-end analytical workflow, spanning from data extraction to the forecasting of May-June sales for Clean City.



Presented by Oumayma Chakroun
MBM_BA_M1



OUTLINE



Company
introduction



Data collection &
exploration



Preprocessing &
Stationarity Check



Model Implementation
SARIMA, ETS models , Prophet,
XGboost



Model Evaluation &
Comparison



Conclusion



1.COMPANY INTRODUCTION

Company overview:

Founded in December, 24, 2021

Page 04

01

Boumhel el bassatine, in front kiosque shell

Business model: B2C

Small size comapny: (8 workers)

02

Product Portfolio:

Detergents, Cosmetics, Decoration, Party supplies.

Revenue concentration: Detergents and Cosmetics,

03

Market positionning:

Target customers (households, middle-income)

Competitive positioning: price-oriented



04

Sales Structure

Data frequency: daily

Sales channels: in-store only

Structural constraints (stockouts, supply issues)

05

Seasonality & Event Sensitivity

Key demand drivers:

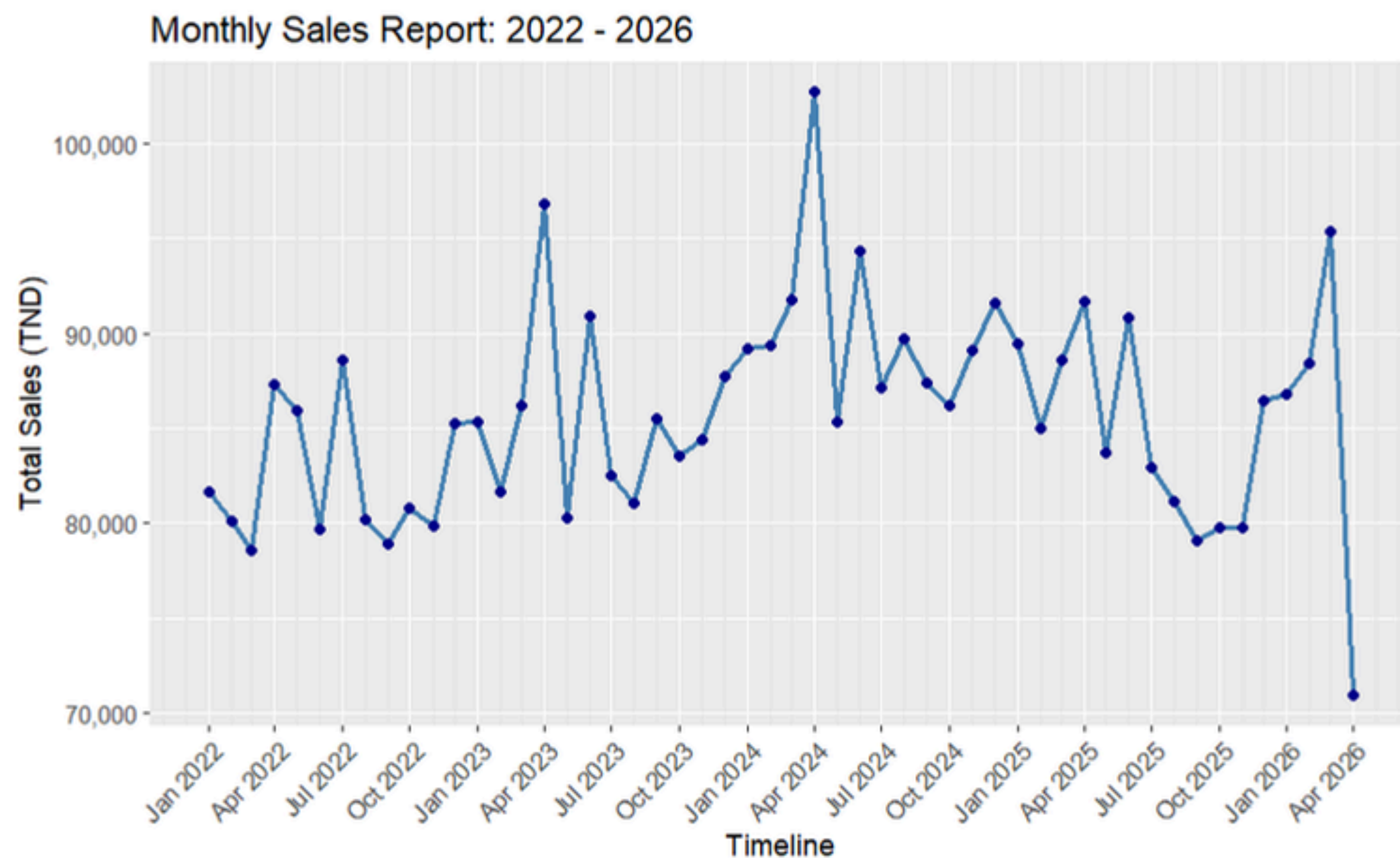
- Religious events (e.g., Eid)
- Social/educational events (e.g., Bac results)
- Promotions, discounts, and pay cycles



2.DATA COLLECTION & EXPLORATION

A.DATA PLOTTING

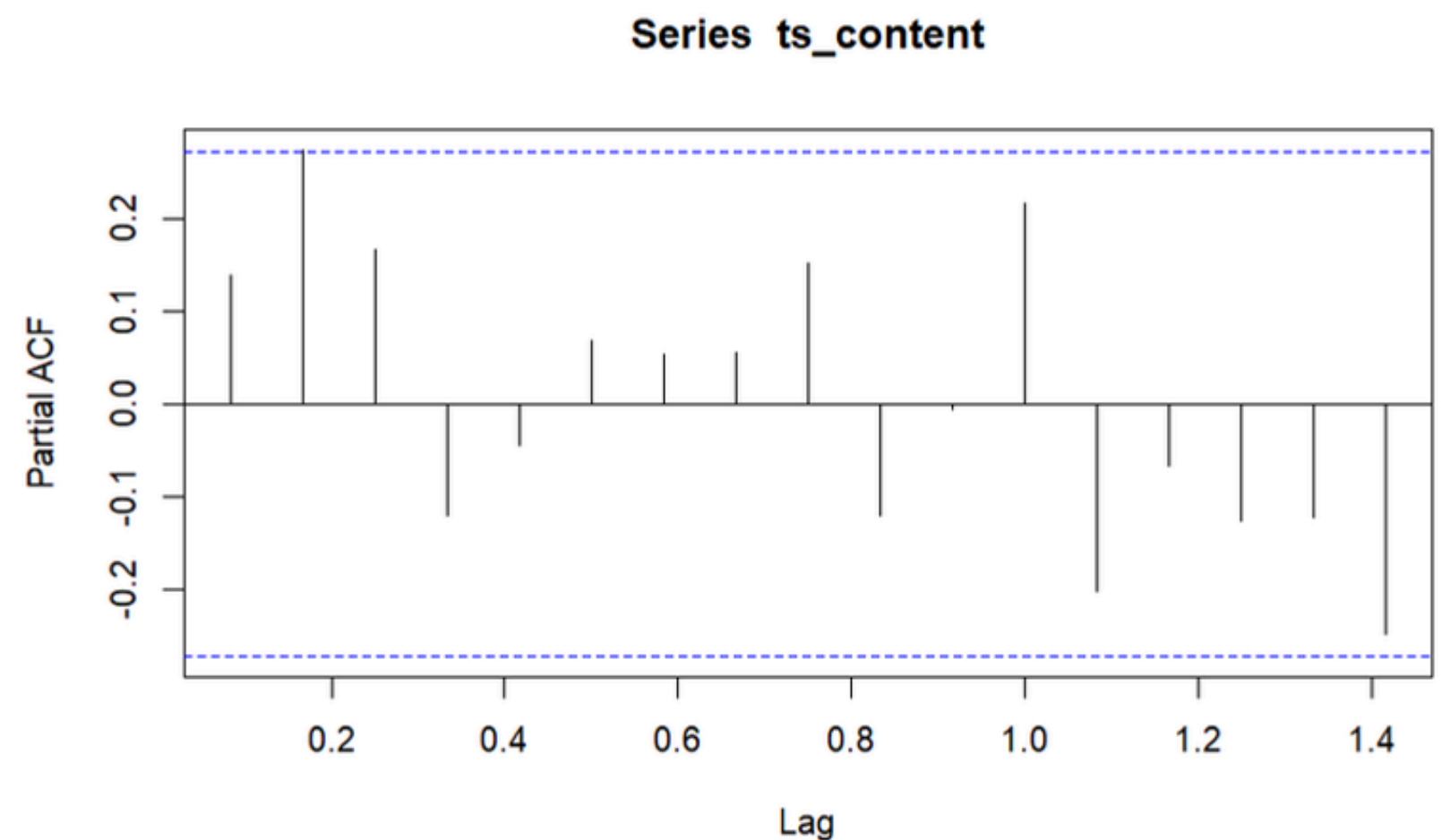
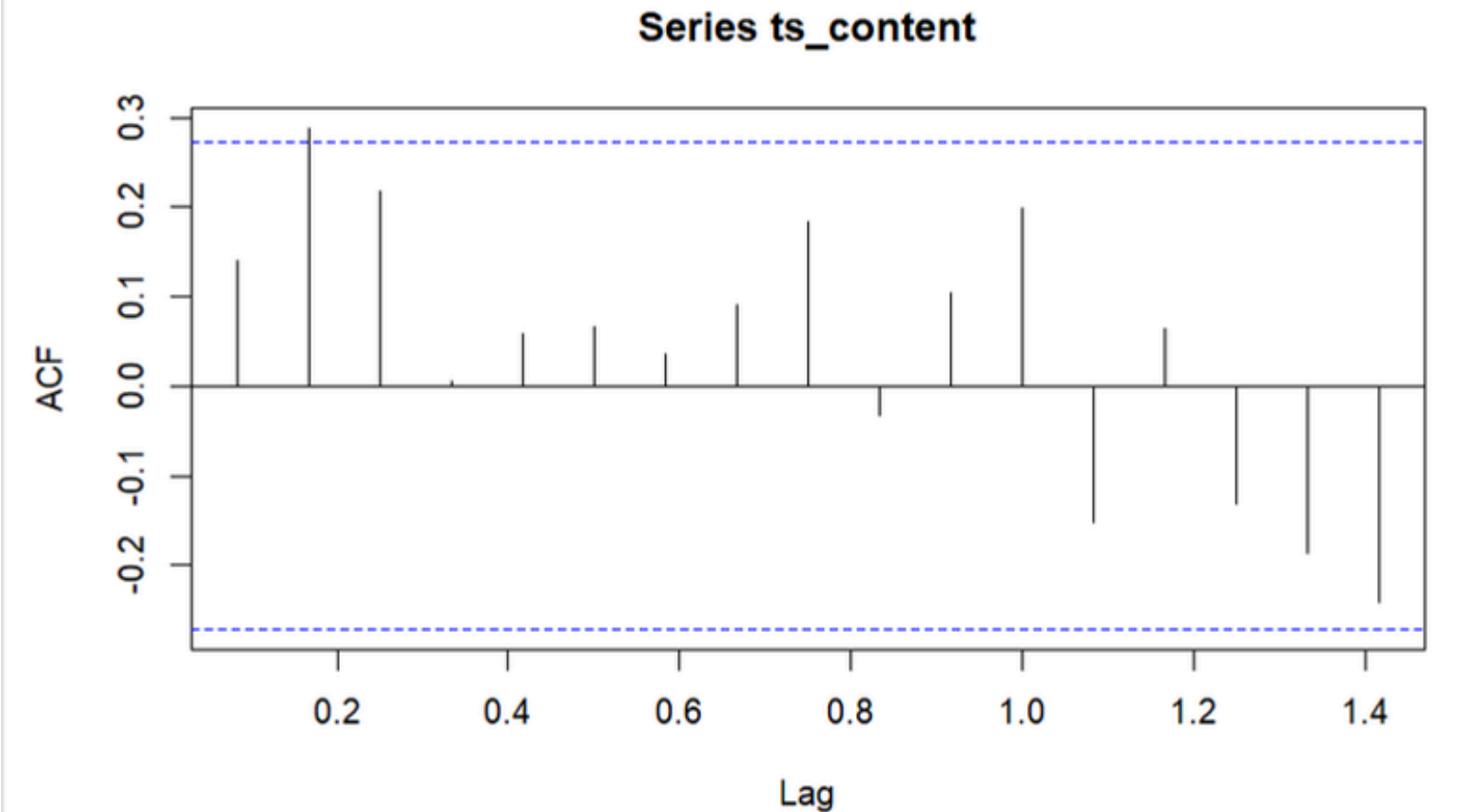
The raw data, originally recorded at a daily frequency, is aggregated to a **monthly** level for forecasting purposes.
The dataset was extracted from the **store's ERP system** and spans the period from **January 2022 to April 2026**.



- High variance with strong short-term fluctuations
- Presence of multiple local peaks and abrupt corrections
- Non-linear and unstable trend, no seasonality captured.
- Phases: growth → sharp decline
- The recurring pattern in April (Two successive EID-El-Fitr)
- Highest pick explained by launching a promotion campaign on mosaïque fm
- Short-term spikes driven by exogenous demand shocks
- Significant drop in last two periods (inflation, collecting money for Eid-EL-Adhha)

B.CORRELATION CHECK

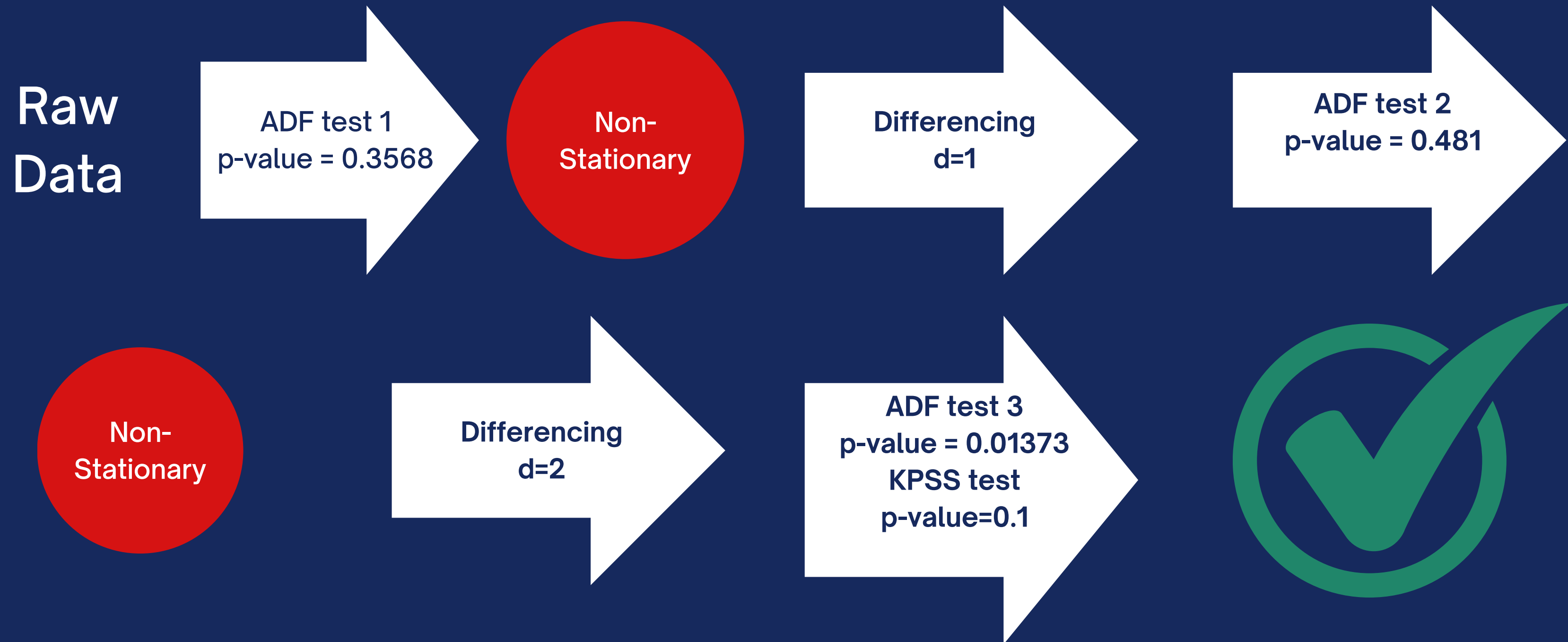
- Most autocorrelation coefficients lie within 95% confidence bounds
- No persistent structure across lags
- Minor exceedances around lags 0.2, not systematic
- No identifiable AR(p) or MA(q) pattern
- Residual behavior is close to white noise
- Further diagnostics required before confirming independence





3.PREPROCESSING & STATIONARITY CHECK

A. STATIONARITY



Model	AIC	BIC
ARIMA(0,0,3)(0,0,1)[12] with non-zero mean	1041.16	1052.87
ARIMA(3,2,0)(1,0,0)[12]	1024.76	1034.32
ARIMA(0,1,1)(0,0,1)[12]	1021.6	1027.3

B.ARIMA MODEL SELECTION

C. MODEL VALIDATION

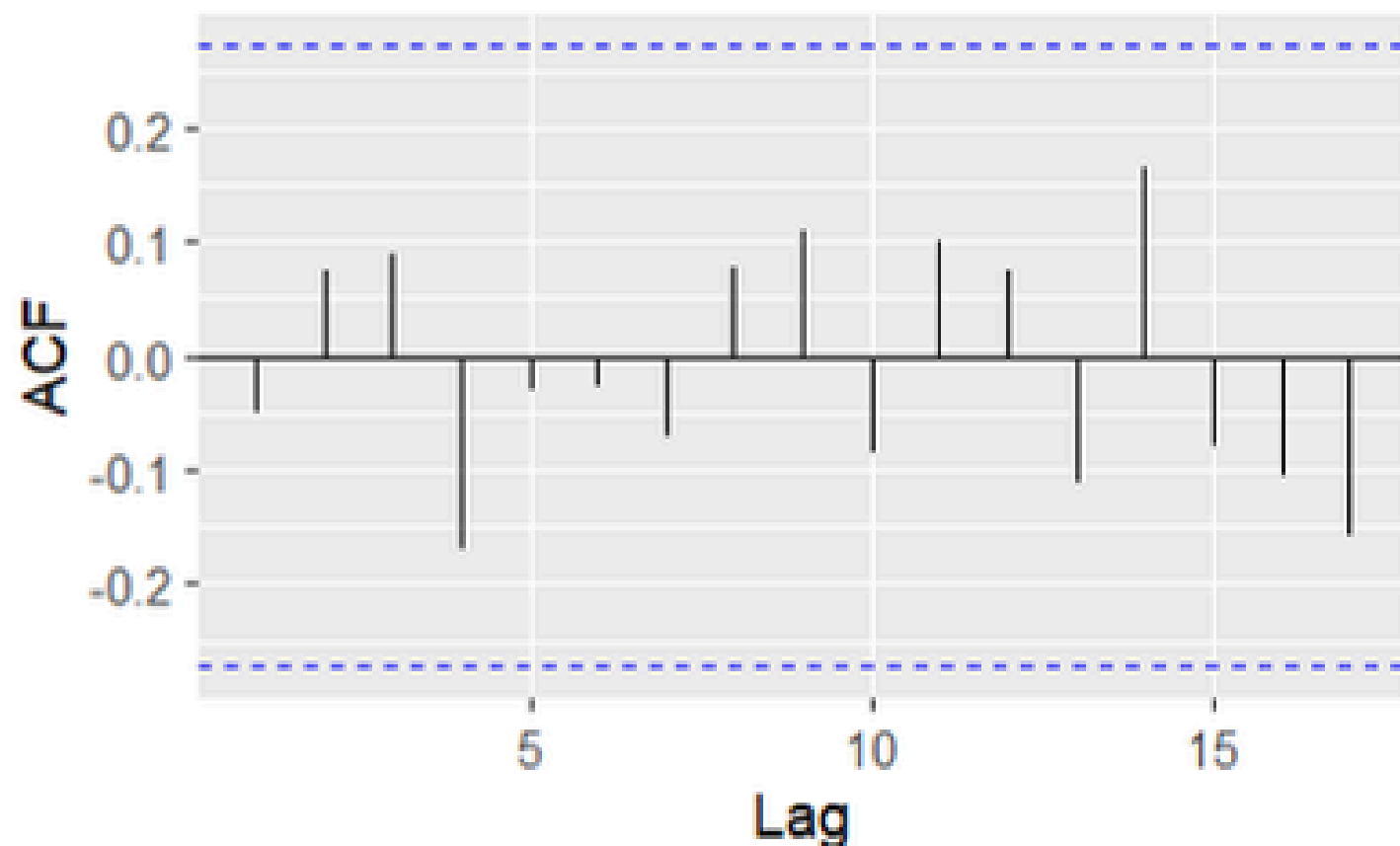
Page 11

Shapiro-Wilk normality test

```
data: residuals(m3)  
W = 0.9729, p-value = 0.2793
```

Ljung-Box test

```
data: Residuals from ARIMA(0,1,1)(0,0,1)[12]  
Q* = 4.6491, df = 8, p-value = 0.7943
```



- no autocorrelation
- residuals are normally distributed
- Residuals ACF reflects white noise

ARIMA(0,1,1)(0,0,1)[12]

READY FOR FORECASTING!



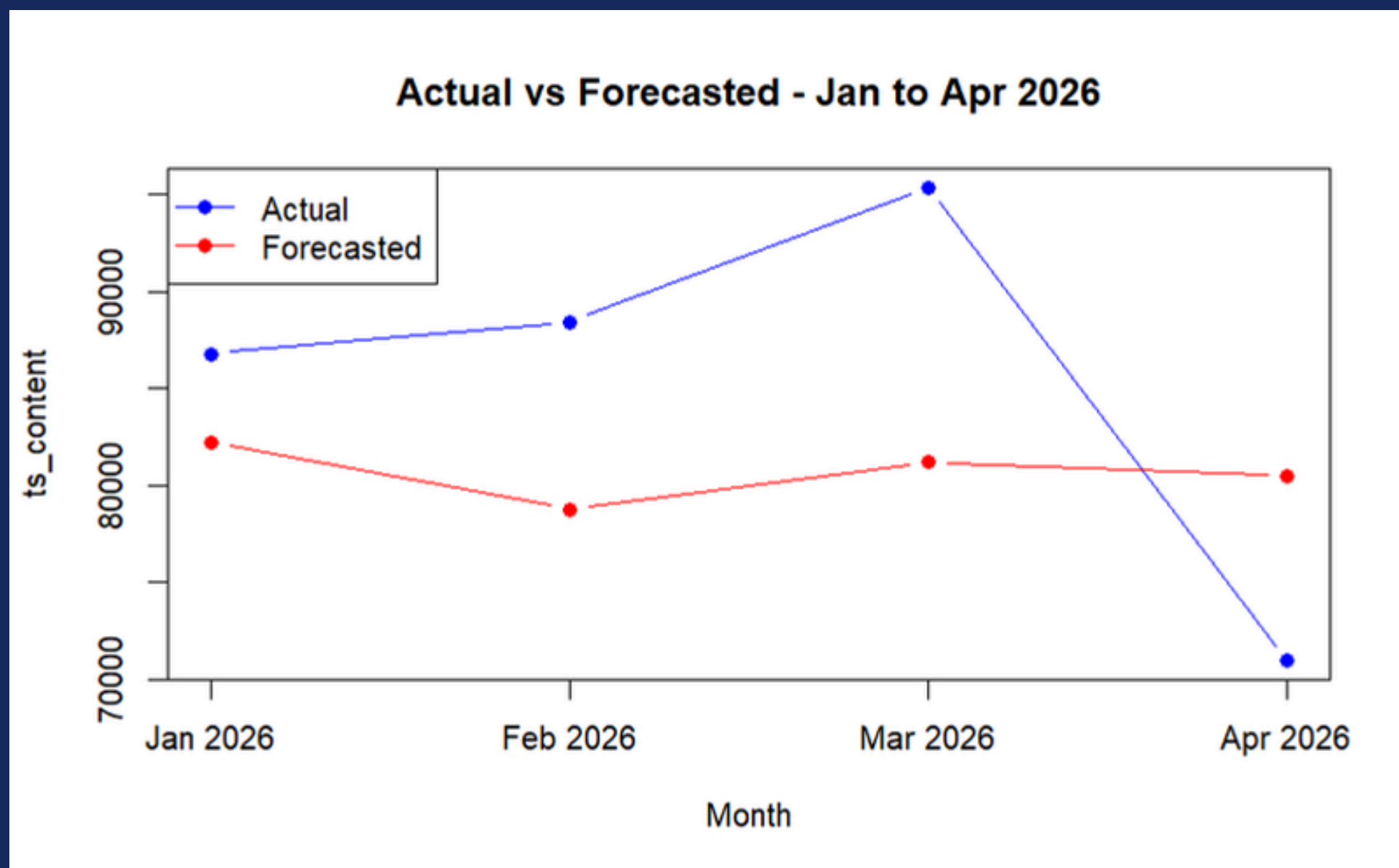
4.MODEL IMPLEMENTATION



Train set: from January 2022 to December 2025

Test set: January to April 2026

Page 14



Errors

MAE: 9481.14

RMSE: 10070.77

MAPE: 11.12 %

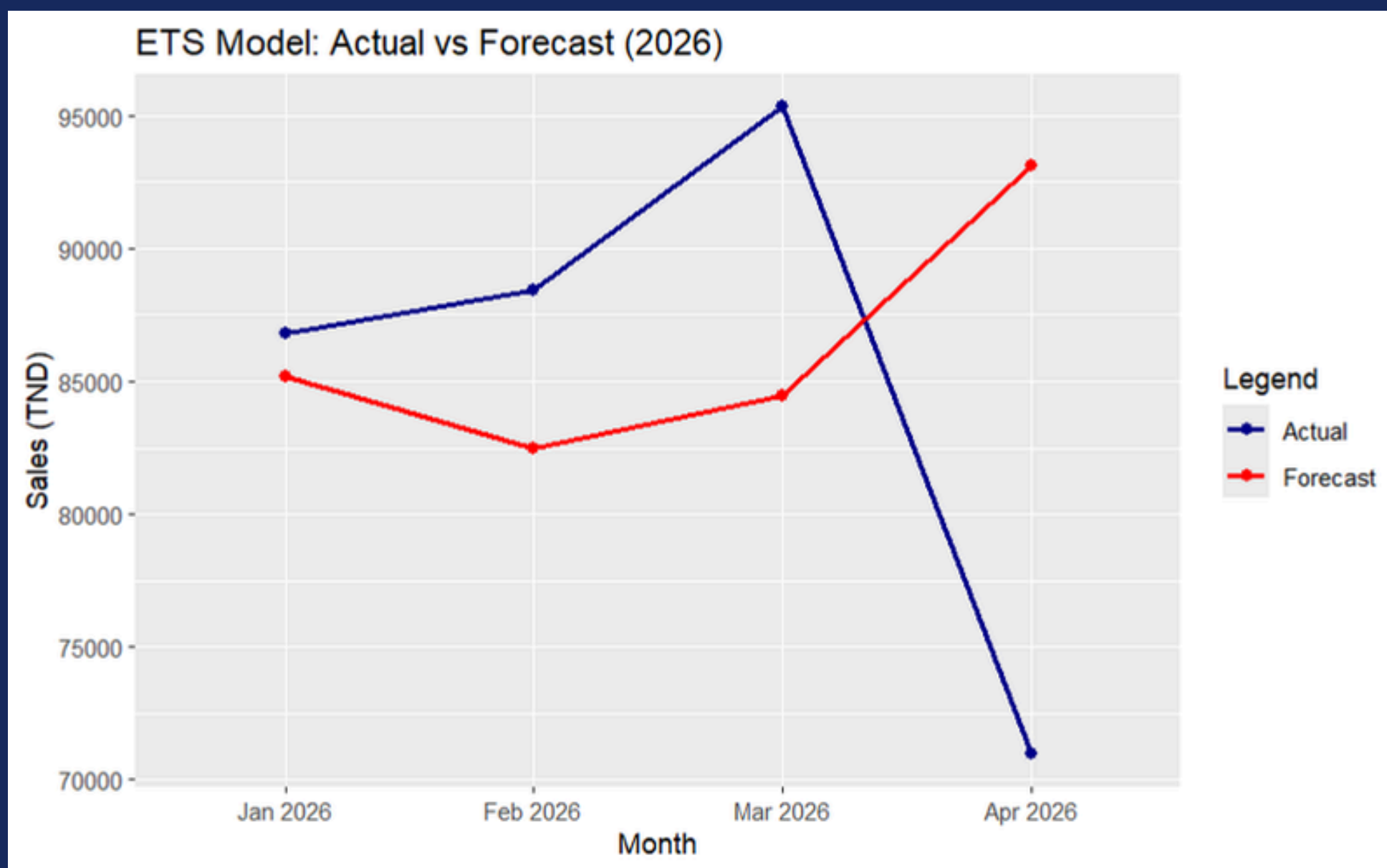
A.SARIMA MODEL



Train set: from January 2022 to December 2025

Test set: January to April 2026

Page 15



Errors

MAE: 10159.32

RMSE: 12726.1

MAPE: 12.82 %

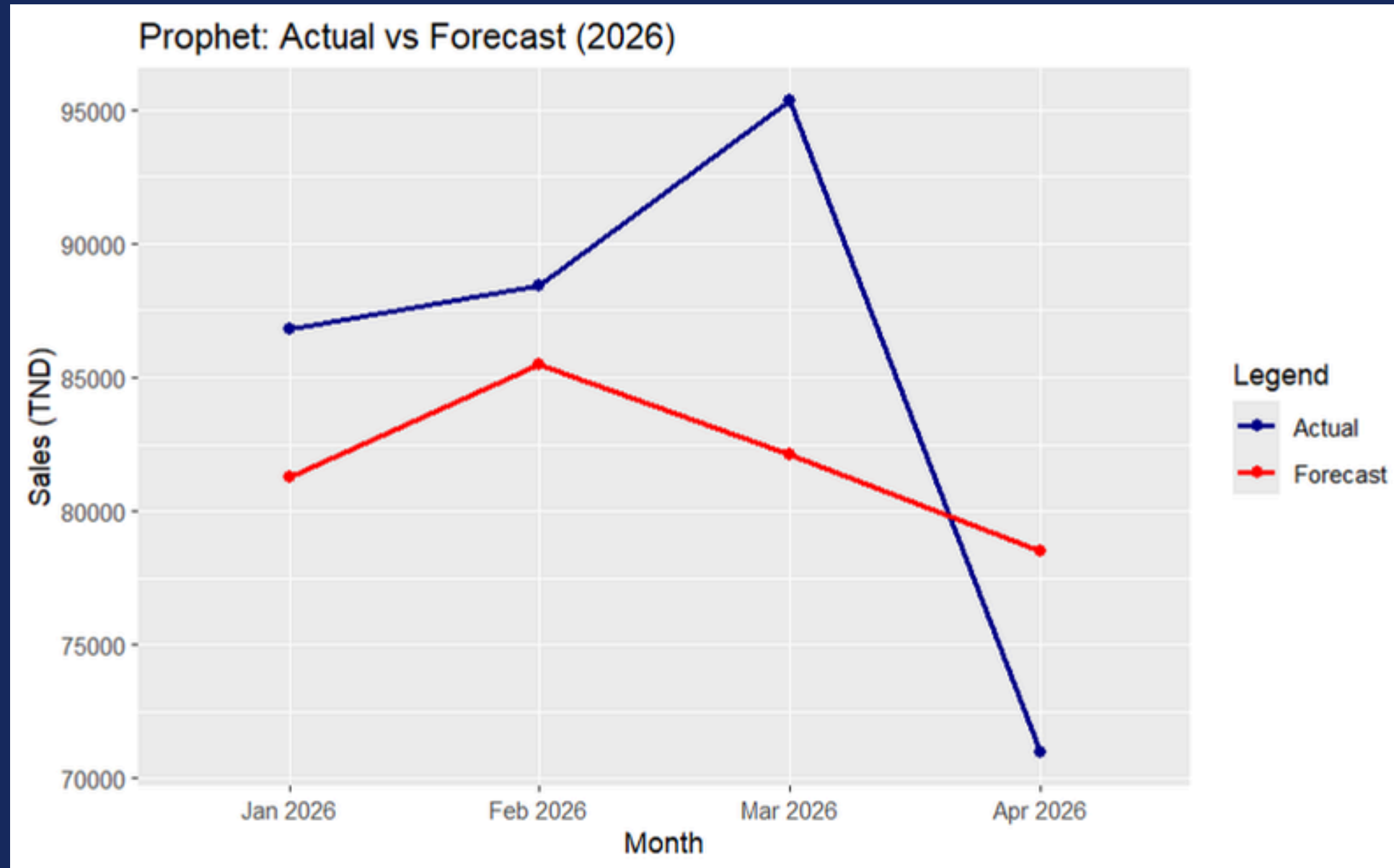
B.ETS MODEL



Train set: from January 2022 to December 2025

Test set: January to April 2026

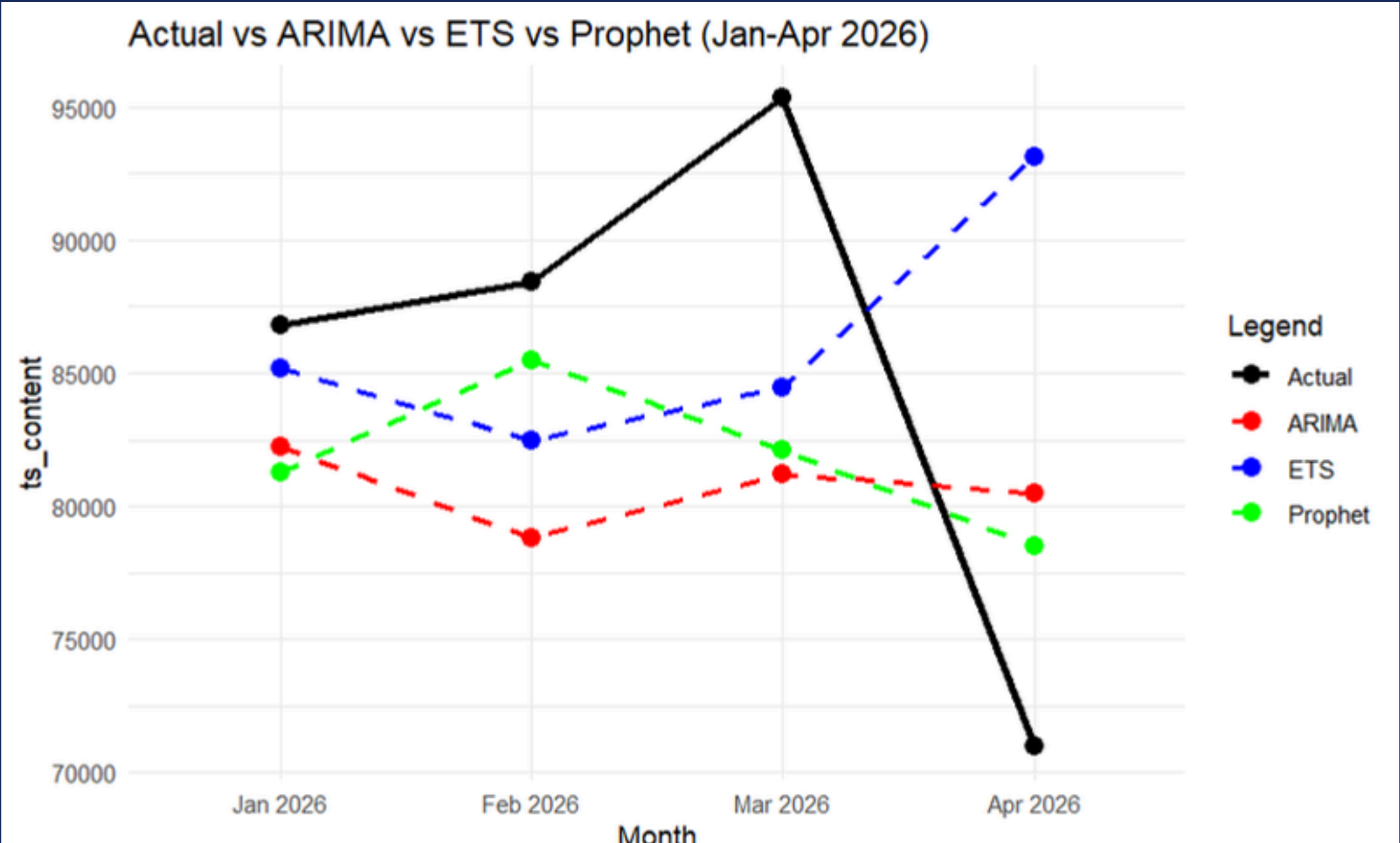
Page 16



Errors

MAE: 7295.46
RMSE: 8220.33
MAPE: 8.53%

C.PROPHET MODEL



Comparison of different models errors

Model	MAE	MAPE	RMSE
ARIMA	9481.14	10070.77	11.12 %
ETS	10159.32	12726.1	12.82 %
Prophet	7295.46	8220.33	8.53%

BEST MODEL

D.COMPARING MODELS

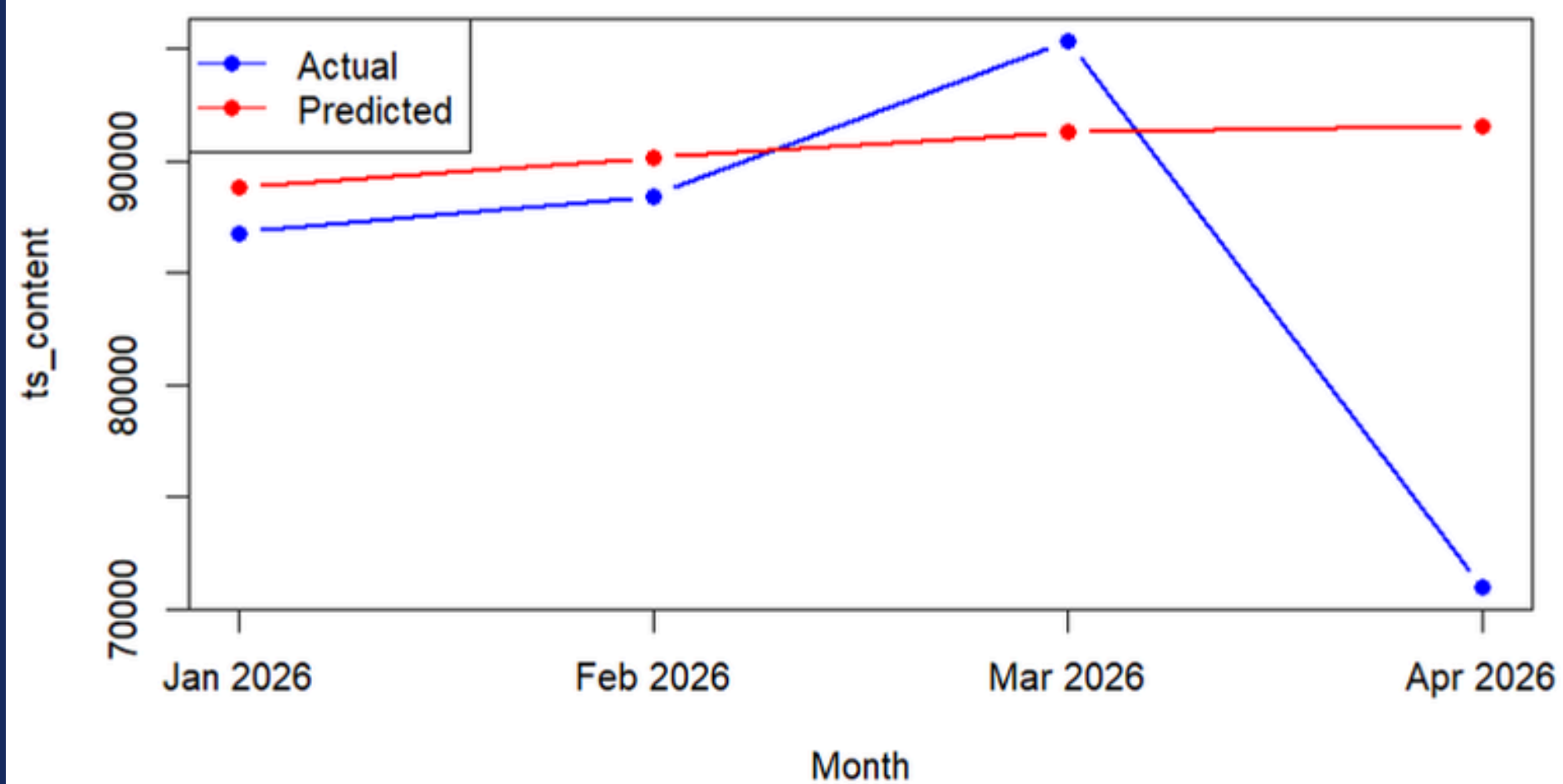


Train set: from January 2022 to December 2025

Test set: January to April 2026

Page 18

XGBoost Forecast vs Actual



Errors

XGBoost MAE: 7116.93
XGBoost RMSE: 10587.33
XGBoost MAPE: 9.41 %

E.XGBOOST MODEL



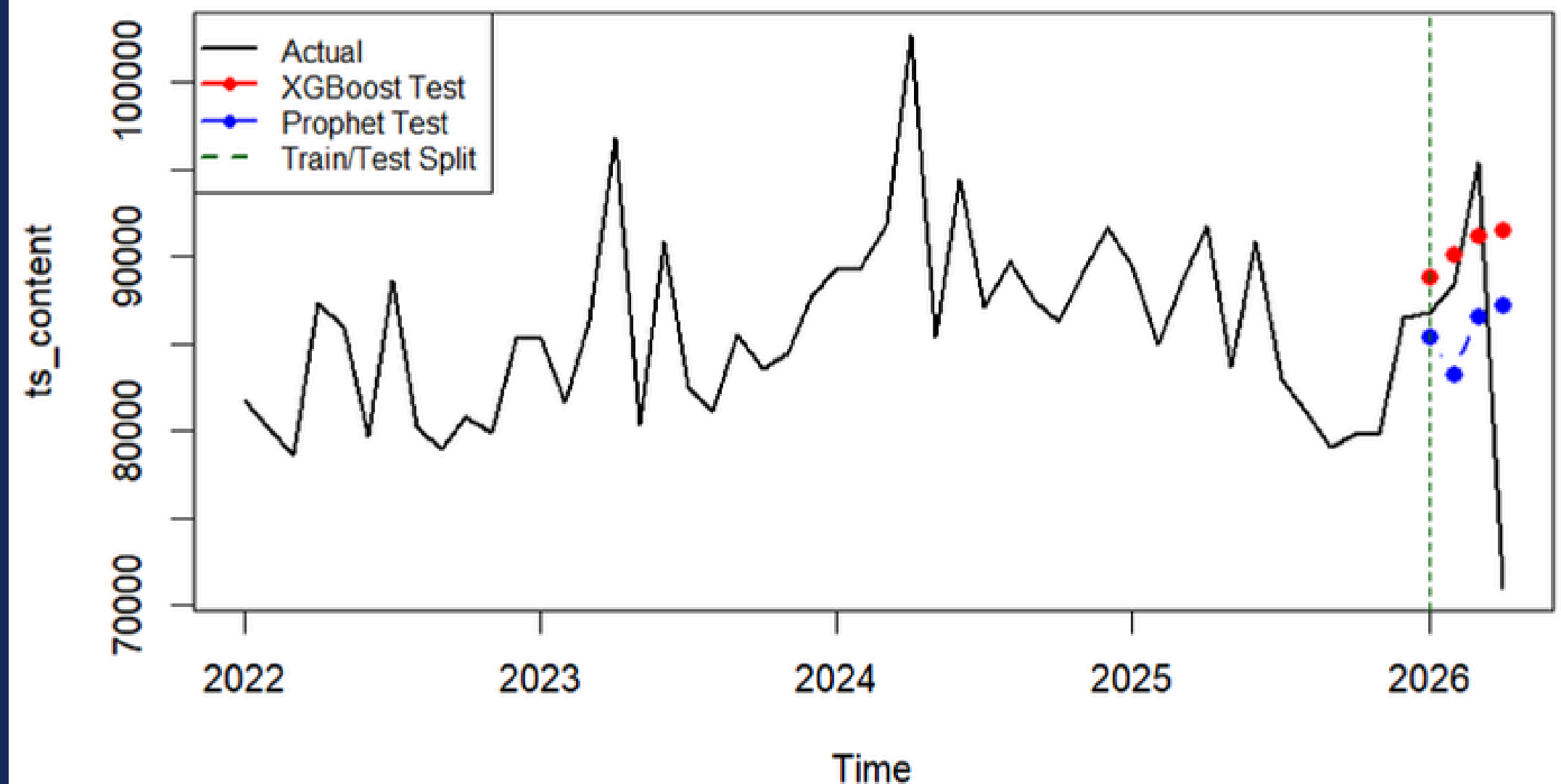
5.MODEL EVALUATION & COMPARISON

XGBoost

- Predictions follow the actual trend more closely
- Better at capturing short-term patterns using lag features
- Red points sit closer to the actual line

Prophet

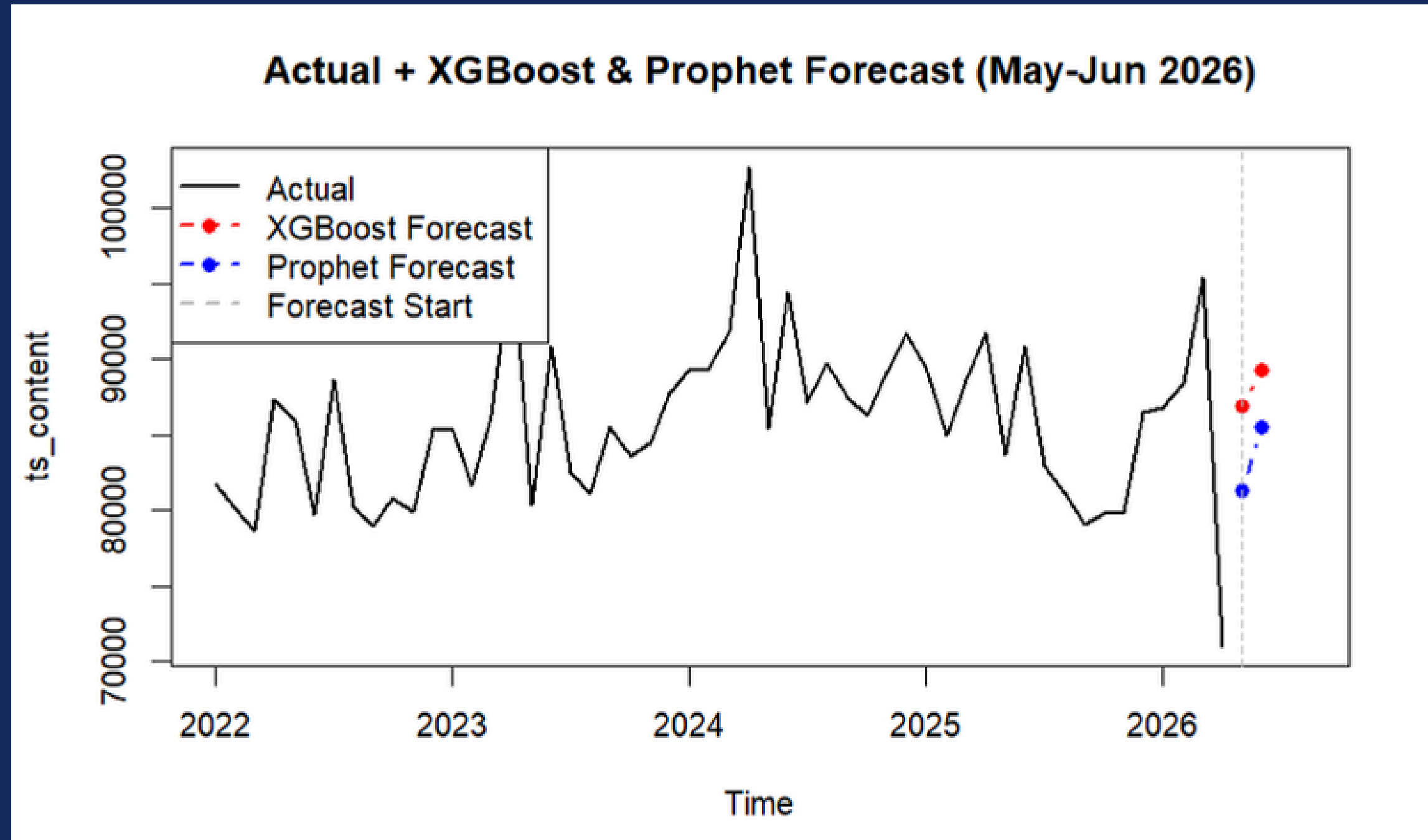
- Tends to smooth out fluctuations
- Blue points show more conservative estimates
- Better at capturing seasonality and holidays automatically

Actual vs XGBoost & Prophet (Test Period)

MODEL EVALUATION

- Both models forecast sales recovery in May–Jun 2026
- XGBoost predicts higher sales driven by recent lag momentum
- Prophet predicts conservative growth based on seasonal trends
- Disagreement between models signals forecast uncertainty

→ **Recommend monitoring
May actuals before trusting
June forecast**



MODEL COMPARISON



6.CONCLUSION

Model	Strengths	Limitation
ARIMA	Statistically validated, white noise residuals	Linear only
ETS	Handles seasonality automatically	Sensitive to outliers
Prophet	Trend + seasonality decomposition	Oversmooths short-term
XGBoost	Best test period fit, captures nonlinearity	Needs feature engineering

MODEL PERFORMANCE SUMMARY

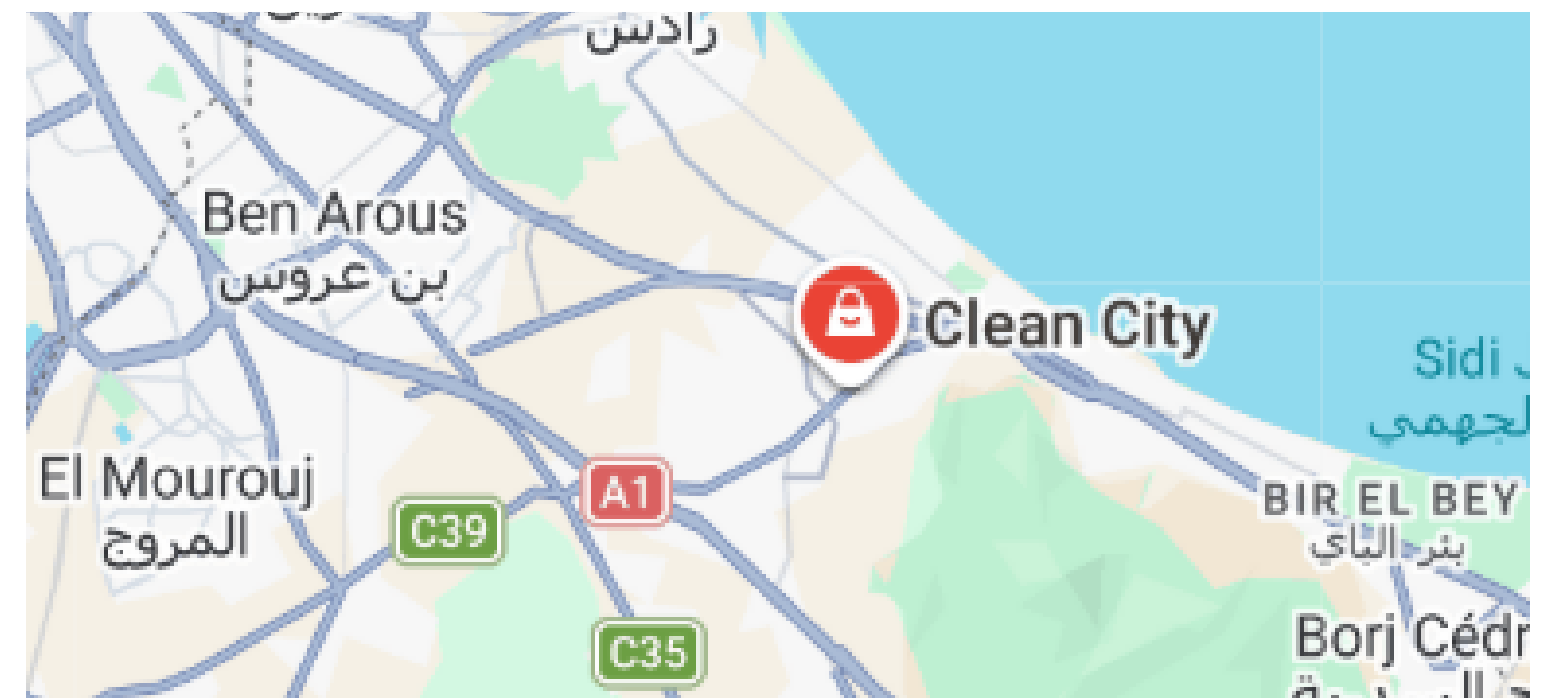
LEARNING OUTCOMES

- Real data is never clean — handling missing values, outliers, and irregular patterns.
- Classical models (ARIMA/ETS) and ML models (XGBoost) require completely different data formats.
- Models can disagree on forecasts — uncertainty must be communicated, not hidden.
- Short datasets (48 observations) severely limit model complexity and reliability.



THANK YOU FOR YOUR ATTENTION

You are welcome to visit us at Clean City!!!



CleanCityBoumhal



+216 92 004 917



cleancity