Time Series Analysis Report

Date: August 5, 2025
Prepared by: Shadab Arif Ansari
Objective: To identify Anomalies and building forecasting model

---

1. Missing Month Detection

We identified **missing monthly data** based on volume trends and system logs. These months had abnormally low or zero activity, indicating either data ingestion issues or actual operational gaps:

Month	Missing Entries
2025-03	44
2025-04	91
2025-05	128
2025-06	71

Insight:
The number of missing entries peaked in **May 2025**, indicating a major data disruption or system anomaly. Recommend cross-verifying with upstream logs or ETL pipelines.

---

2. Spike/Anomaly Detection (Z-Score Method)

Using **Z-score–based statistical anomaly detection** on the response volume, we found significant spikes where the response average deviated from the norm.

Detected Spikes (Sample):

Timestamp	Volume	Resp_Avg	Z-Score	Anomaly
2025-03-24 03:00:00	886	13.43	2.57	True
2025-04-08 02:00:00	2927	14.53	3.07	True
2025-04-28 08:00:00	9250	14.03	2.84	True
2025-05-14 16:00:00	2262	15.81	3.66	True
2025-05-19 16:00:00	4890	21.23	6.16	True
2025-05-26 03:00:00	3609	12.88	2.31	True

Insight:

- The **highest spike** occurred on **2025-03-26** , with a Z-score of **26.01** — a strong anomaly.
- April and May show frequent spikes, possibly correlating with missing entries or campaign surges.

Recommendations

- **Data Gaps:**
  - Investigate and backfill missing data in **March–June 2025** from raw sources or backups.

    - Automate anomaly logs when missing records cross a threshold (e.g., 20 per month).

- **Anomalies:**

    - Validate anomalies with external events (e.g., promotions, outages, product launches).

    - Consider modeling using robust methods like **STL decomposition**, **Isolation Forest**, or **AutoEncoder-based anomaly detection** for better precision.

Time Series Model Comparison for Forecasting Hourly Response Times

Models Chosen

We selected two models for time series forecasting:

1. **Simple Moving Average (SMA):**

    - A basic, intuitive model that forecasts by averaging previous values.

    - Best for short-term trends without complex seasonality or trend shifts.

2. **Holt-Winters Exponential Smoothing (Additive):**

    - Captures **trend** and **seasonality** in time series data.

    - Suitable for time series with **hourly seasonality** (e.g., 24-hour cycles).

📊 Evaluation Metrics Used

We evaluated both models using three standard error metrics:

Metric	Description	Why it matters
MAE (Mean Absolute Error)	Average of absolute errors	Interpretable in original units; robust to outliers
RMSE (Root Mean Squared Error)	Square root of mean squared error	Penalizes larger errors more, sensitive to variance
MAPE (Mean Absolute Percentage Error)	Mean of absolute percentage errors	Expresses error as a % of actuals; useful for interpretability

---

Results

Model	MAE	RMSE	MAPE
Simple Moving Average	0.9940	1.4246	14.67%
Holt-Winters (Additive)	0.98	1.4200	13.14%

Interpretation

- **MAPE Comparison**: Holt-Winters has **lower MAPE (13.14%)** than SMA (14.67%) → better relative forecasting accuracy.
- **RMSE** is **slightly better** (lower) for Holt-Winters, indicating better error handling.
- **SMA** is simpler but doesn't account for **seasonality**, which is present in the data (based on ACF and hourly periodicity).
- Holt-Winters captures **daily cycles (24-hour seasonality)**, giving it a clear edge.

Conclusion

Best Model: Holt-Winters Exponential Smoothing

- It outperforms Simple Moving Average in MAPE and RMSE.
- Captures both **trend and seasonality**, crucial for hourly data.
- More robust for time series with predictable patterns.