Monitoring analyst test

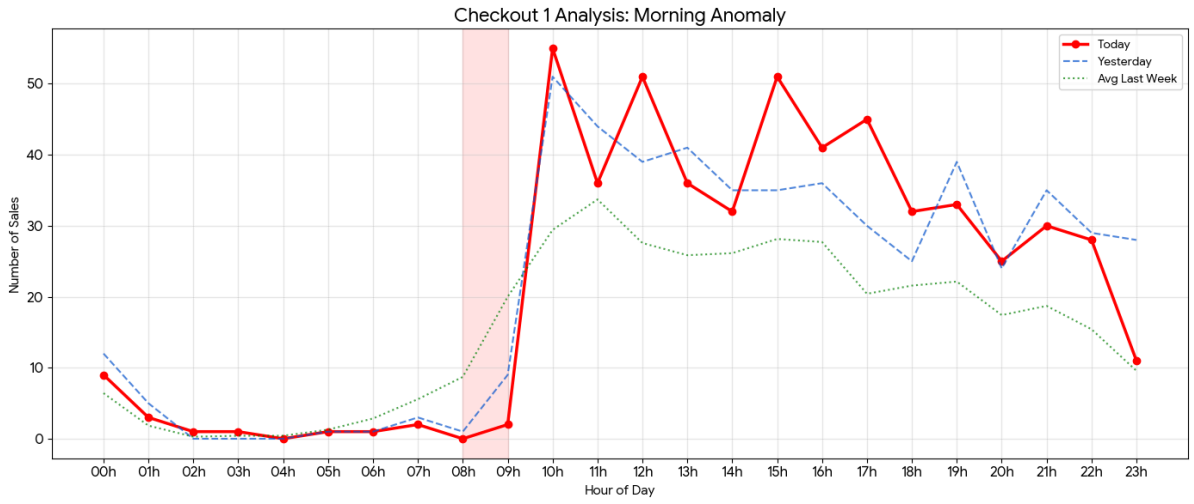
Analysis of hypothetical data.

The following analysis details the identification of two critical availability incidents based on the sharp deviation between the actual transactional volume and the historical baseline (`avg_last_week`).

1. Data Analysis & Conclusions

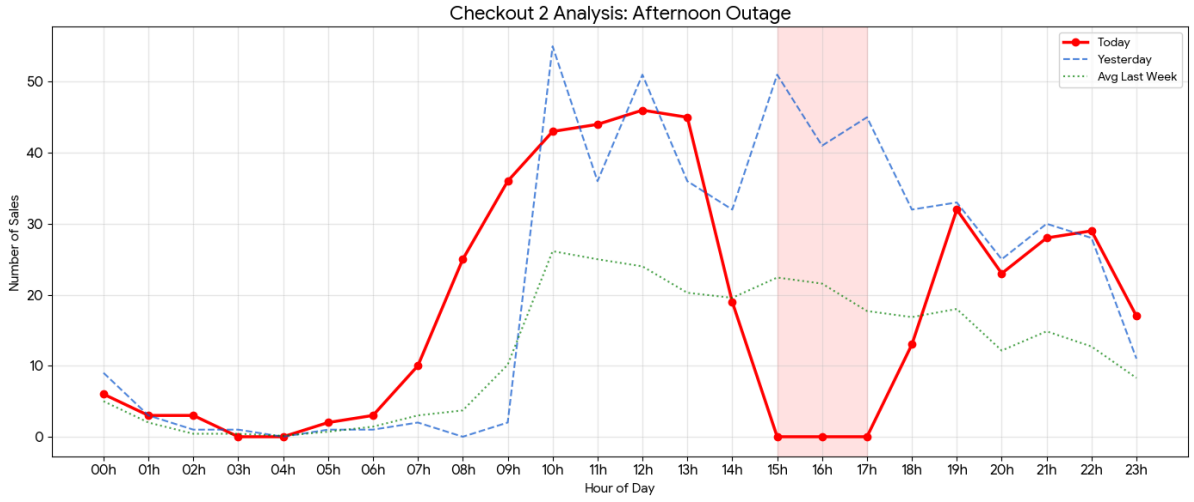
By cross-referencing the `today` column (current real-time data) against the `avg_last_week` column (expected baseline behavior), I identified two distinct and critical incidents.

Scenario A: Checkout 1 — Severe Service Degradation



- **The Anomaly:** Between 08:00 and 09:00, there was a drastic drop in sales volume. The red-shaded area (08h-09h) highlights the trough where performance dropped nearly 100% relative to expectations.
    - At 08:00, sales hit **0** (expected average was ~8.7).
    - At 09:00, sales were **2** (expected average was ~20).
- **The Rebound Effect (Catch-up):** At 10:00, sales surged to 55, significantly surpassing the historical average of 29. Here, the red peak overtakes the green baseline, confirming the system recovery theory.
- **Conclusion:** This indicates a system failure or severe latency during the morning ramp-up period. The spike at 10:00 suggests that the system came back online and processed a backlog of queued transactions, or that users retried their failed attempts immediately after service stabilization.

Scenario B: Checkout 2 — Total Unavailability (Hard Down)



- **The Anomaly:** Between 15:00 and 17:00, sales volume flatlined to exactly zero.
- **Context:** This represents a peak business window. The historical baseline (`avg_last_week`) indicates an expected volume of 22 to 28 sales per hour during this period.
- **Visual Evidence:** Between 15:00 and 17:00, while the green (`Avg`) and blue (`Yesterday`) lines signal high activity, the red line drops to the x-axis. The gap between the red and green lines visually represents the direct financial loss incurred during the outage.
- **Conclusion:** Unlike Checkout 1, this scenario depicts a **Total Outage**. No processing occurred for three consecutive hours, resulting in irreversible revenue loss with no immediate recovery (volume at 18:00 returned weak and below average).

2. SQL Query & Anomaly Explanation

To detect such incidents in a production database and generate automated alerts, relying solely on `sales = 0` is insufficient (as zero sales at 04:00 AM is normal behavior). We must compare **Actual vs. Expected** performance.

Below is a SQL query designed to identify these anomalies by calculating the percentage deviation.

```
/* ANOMALY MONITORING QUERY
Objective: Alert if current volume is < 50% of the historical average
during business hours (08h-22h).
*/
```

```sql
SELECT
    time AS hour_of_day,
    today AS current_sales,
    avg_last_week AS expected_baseline,
    yesterday AS previous_day_sales,
    -- Drop Calculation (Drop %)
    ROUND(((avg_last_week - today) / avg_last_week) * 100, 1) AS
drop_percentage,
    CASE
        -- Critical Alert: Drop > 80% or Zero Sales during peak hours
        WHEN today = 0 AND avg_last_week > 5 THEN 'CRITICAL: OUTAGE'
        -- Warning Alert: Drop > 50%
        WHEN today < (avg_last_week * 0.5) THEN 'WARNING: DEGRADATION'
        ELSE 'NORMAL'
    END AS system_status
FROM
    checkout_sales_monitor
WHERE
    -- Focus on business hours to avoid false positives during early
morning
    time BETWEEN '08h' AND '22h'
    -- Filter only where anomalies exist for the report
    AND today < (avg_last_week * 0.5)
ORDER BY
    time;
```

3. The Importance of Data Triangulation (Today, Yesterday, Avg)

As an analyst, understanding the specific utility of each comparison column is crucial for accurate diagnosis:

- **Seasonality (Same Day Last Week):** Consumer behavior varies by day of the week (e.g., a Monday behaves differently than a Saturday). Comparing against the "Same Day Last Week" helps confirm if the current pattern is consistent with weekly cycles.
- **Recent Trend (Yesterday):** Useful for spotting immediate shifts, but risky as a standalone metric.
    - *Example:* If "Yesterday" was a holiday, volume would be naturally low. Using it as a baseline would make "Today's" normal volume appear as a false positive anomaly.
- **The Golden Baseline (Avg Last Week/Month):** This is the most reliable metric for alerting. Averages smooth out outlier spikes and troughs.

- *Context:* In Checkout 1, knowing that the 08:00 average is **8.71** is what allows us to confirm that **0** represents a failure, not just "low traffic."

Executive Summary for Leadership:

"We detected severe instability in **Checkout 1** during the morning (partial outage) and a total interruption in **Checkout 2** during the afternoon (critical outage). The monitoring system, utilizing historical averages as a baseline, confirmed that the deviation was **technical** rather than organic (low demand).

Recommended Action: Immediate investigation of system logs at **08:00** and **15:00** to identify the root cause."