Manufacturing Analytics Technical Challenge

Analysis of Manufacturing Production Data: Insights for Operations Improvement

Mat Yohannes

November 13, 2024

1 Introduction

This paper analyzes production data for a label printing facility that produces beer bottle labels. Using various visualizations constructed in Preset.io, the goal is to identify key metrics and offer actionable insights that an Operations Supervisor can use to make better decisions about the facility's performance. By evaluating metrics like production output, reject rates, cycle times, and downtime, this report helps identify opportunities for process improvement and waste reduction.

2 Visualization Selection & Rationale

2.1 Overall Equipment Effectiveness (OEE) by Device and Shift

Why: OEE is a useful metric in manufacturing which provides insights into equipment productivity.
Use: Supervisors can identify underperforming shifts or devices, highlighting areas that need maintenance or operation corrections.
Formula:

$$\frac{\frac{\text{duration}}{\sum(\text{CycleTime} \times \text{equipment\_cycles})} \times \frac{\text{good\_count}}{\text{target\_count}} \times \text{good\_count}}{\text{good\_count} + \text{reject\_count}} \tag{1}$$

2.2 Cycle Time Efficiency per Operation

Why: This chart shows how efficiently operations are performed relative to cycle time expectations.
Use: Supervisors can pinpoint operations with high delays.
Formula:

$$\frac{\text{CycleTime} \times \text{good\_count}}{\text{CycleTime} \times (\text{good\_count} + \text{reject\_count})} \tag{2}$$

2.3 Yield Rate by Shift and Device

Why: Chart helps visualize quality control by displaying the ratio of successful outputs per shift and device.
Use: Decisions on shift-specific training or device adjustments can be made to reduce defect rates.
Formula:

$$\frac{\text{good\_count}}{\text{good\_count} + \text{reject\_count}} \tag{3}$$

2.4 Production Phases Time Breakdown by Device and Shift

Why: Breaks down time spent in different production phases.
Use: Identifies shifts or devices that are experiencing long downtimes which may require scheduling or equipment changes.

2.5 Labor Utilization by Shift and Device

Why: Measures labor usage to understand productivity against time invested in production.
Use: Supervisors can allocate labor more effectively.
Formula:

$$\frac{\sum(\text{labor})}{\sum(\text{duration})} \tag{4}$$

2.6 Target Count Achievement by Shift and Device

Why: Tracks target goals to assess shift or device productivity.
Use: Supervisors can identify where target production falls short and allow changes to meet goals.
Formula:

$$\frac{\text{good\_count}}{\text{target\_count}} \tag{5}$$

2.7 Partial Cycle Lost Time by Device and Operation Type

Why: Displays wasted time in partial cycles which show inefficiencies in production.
Use: Specific devices or operations with recurring cycle losses can be addressed.

2.8 Performance Impact by Process State and Device

Why: Analyze different process states impact on device performance.
Use: Supervisors can reduce downtime by addressing states frequently reducing production.

2.9 Production Rate by Area and Operation

Why: Shows how different areas and operations contribute to production speed.
Use: Supervisors can optimize high-output areas and adjust low-output processes to improve rates.
Formula:

$$\frac{\sum(\text{duration})}{\sum(\text{equipment\_cycles})} \tag{6}$$

2.10 Downtime Distribution by Team and Shift

Why: Chart shows how downtime is distributed across teams and shifts.
Use: Supervisors can identify patterns in team-related downtime, allowing them to focus teaming or reallocation teams.

3 Challenges

Uploading the CSV files to Preset.io was challenging due to row limitations, which reduced over 100,000 rows to fewer than 6,000, affecting data completeness. Attempts to use PostgreSQL and MySQL databases were hindered by port connection issues (port forwarding was not allowed by my internet service provider). To address this, I uploaded the files to AWS S3, created tables in AWS Athena, and connected Athena to Preset.io, enabling access to the full dataset. However, the costs started exceeding the free-tier limit.

After reconsidering, I opted to work with a sample size instead of the entire dataset. For a population of 100,000 rows, a sample size of at least 384 was statistically sufficient, so I decided on a limit of 500. This decision allowed me to continue with meaningful analysis while controlling costs. I then focused on refining key metrics to ensure the visualizations provided valuable, actionable insights for operational improvements.

Certain graphs, such as Downtime Distribution by Team and Shift, Partial Cycle Lost Time by Device and Operation Type, Performance Impact by Process State and Device, and Production Rate by Area and Operation, do not display proper visualizations due to the limit of the sample size. The reduced dataset impacted the completeness of these visualizations, limiting their ability to provide detailed insights.

A strong understanding of the label printing domain—including production processes, team roles, and typical performance benchmarks—would have helped make the analysis more precise.

4 SQL Query

```
SELECT p.deviceKey, duration, equipment_cycles,
good_count, reject_count, target_count,
labor, part_display_name,
performance_impact_display_name,
process_state, process_state_reason_display_name,
production_phase_display_name,
shift_display_name, shift_hour_display_name,
team_display_name,
CycleTime, partial_cycle_lost_time, area,
Location, Operation, Type
FROM ProductionMetric AS p
LEFT JOIN DeviceProperty as d
ON p.deviceKey = d.deviceKey
WHERE duration >= 0
AND good_count >= 0
AND labor >= 0
AND partial_cycle_lost_time >= 0
AND part_display_name LIKE 'Part%'
AND shift_display_name IN ('First Shift',
'Second Shift', 'Third Shift')
AND shift_hour_display_name NOT LIKE 'No Hour'
AND team_display_name IN ('Team 1',
'Team 2', 'Team 3')
AND target_count > 0
AND (CycleTime * equipment_cycles) > 0
AND (good_count + reject_count) > 0
AND process_state LIKE 'down'
ORDER BY RANDOM()
LIMIT 500;
```