Process Improvement Tracker

Detailed Project Report

Author: Devin Richmond
Date: June 2025

Table of Contents

1. Executive Summary

In a simulated semiconductor fab, we measured the impact of doubling tool capacity on wafer-lot wait times ("queue time") and daily throughput. By comparing a **Before** phase (one etch tool) with an **After** phase (two identical etch tools), we found:

- **Average queue time** decreased by **50 %** (2.00 h → 1.00 h)
- **Median queue time** dropped by **1.00 h** (2.00 h → 1.00 h)
- **On-time lots** (≤ 2 h wait) rose from **50 %** to **98 %**
- **Daily lots processed** increased by ≈ **7 %** (~650 → ~700)

This project showcases end-to-end data skills: simulating "dirty" logs, cleaning and wrangling, simple statistical analysis, clear visualizations, and plain-English storytelling.

---

2. Introduction

Semiconductor fabs consist of many sequential tool steps; any delay at one step cascades through the line, increasing work-in-progress (WIP) and cost. Queue time—the waiting period before a lot enters a tool—is a critical metric. Reducing queue times improves on-time delivery and maximizes expensive tool utilization. This report demonstrates how adding a second etch tool halves typical wait times and modestly boosts throughput.

---

3. Data Description

- **Phases**:
  - **Before** (2025-04-01 to 2025-05-01): one etch tool, mean simulated queue ≈ 2 h
  - **After** (2025-05-02 to 2025-06-01): two etch tools, mean simulated queue ≈ 1 h
- **Rows**: ~39 200 lots after initial cleaning
- **Columns**:
  - `lot_id` (string): Unique identifier
  - `phase` (Before/After)
  - `arrival_time`, `start_time`, `end_time` (timestamps)
  - `tool_id` (Etch-1 or Etch-2)
  - `operator` (OP-A, OP-B, OP-C)
  - `queue_time_hr`, `cycle_time_hr` (computed durations)

Dirty elements introduced for realism: missing/malformed timestamps, swapped start/end times, duplicates, and extreme outliers.

4. Methodology

4.1 Data Simulation

- Generated arrival times uniformly over each phase's month.
- Simulated queue and service (cycle) times using normal distributions (mean & SD set per phase).
- Introduced "dirt" by randomly:
    - Removing timestamps (1 %)
    - Formatting some fields as strings (2 %)
    - Swapping some start/end pairs (0.5 %)
    - Injecting outliers (1 %)
    - Duplicating rows (0.5 %)

4.2 Data Cleaning

1. **Parse Dates**: Used `pd.to_datetime(..., errors='coerce')` to convert all timestamp columns, dropping rows with any `NaT` (~2.5 %).
2. **Compute Durations**:
    - `queue_time_hr` = (start_time − arrival_time) in hours
    - `cycle_time_hr` = (end_time − start_time) in hours

4.3 Outlier Detection

- Applied the **1.5 × IQR rule** on both `queue_time_hr` and `cycle_time_hr`.
- Flagged and dropped ~1 % of rows where either metric lay beyond [Q1 − 1.5 IQR, Q3 + 1.5 IQR].

4.4 Exploratory Data Analysis

- **Summary Statistics** by phase: count, mean, median, standard deviation, and key percentiles.
- **Visualizations**:
    - Overlaid histograms of queue times (cleaned vs. raw; then Before vs. After).
    - Boxplots comparing phases.
    - Line chart of daily lots processed in each phase.
- **Key Metrics**: computed percent change, on-time lot percentages (≤ 2 h).

5. Results

5.1 Queue-Time Distribution

Insight: After removing outliers, waits cluster around 1 h instead of 2 h; extreme waits (> 3.5 h) are eliminated.

5.2 Queue-Time Comparison by Phase

Insight: Median queue time halved (2 h → 1 h); the middle 50 % of waits tightened from [1.5–2.5 h] to [0.7–1.3 h].

5.3 Throughput Analysis

Insight: Daily output increased from ~650 to ~700 lots, and daily counts became less volatile.

5.4 Key Performance Indicators

KPI	Before	After	Change
Average queue time (hr)	2.00	1.00	–50 %
Median queue time (hr)	2.00	0.99	–1.01 hr
On-time lots (≤ 2 h wait)	50 %	98 %	+48 pp
Average daily lots processed	650 approx.	700 approx.	+7 %

---

6. Discussion

- **Business Impact:** Halving queue times slashes WIP holding costs and speeds order fulfillment. A 7 % rise in throughput translates to dozens more lots per day without capital-heavy tool purchases beyond the second tool.
- **Data Skills Demonstrated:** End-to-end handling of messy logs, robust outlier treatment, clear visual storytelling in plain language.
- **Limitations:** Simulation may not capture real-world scheduling nuances or maintenance events. Real fab data may exhibit different distributions or time dependencies.

---

7. Conclusions & Recommendations

This analysis confirms that increasing tool capacity can dramatically reduce wait times and modestly boost throughput. For a real fab, I would recommend:

1. **Pilot Implementation:** Add a secondary tool in a single production line to validate these gains.
2. **Monitor Real Metrics:** Instrument live queue measurements and compare with model predictions.
3. **Further Optimizations:** Explore operator scheduling and preventive maintenance timing to smooth queues.

---

8. Next Steps

- **Cost–Benefit Analysis:** Calculate ROI for capital investment.
- **Shift-Level Analysis:** Model impact of operator staffing patterns.
- **Quality Integration:** Incorporate defect and yield data to balance speed with quality.
