Process Improvement Tracker

Detailed Project Report

Author: Devin Richmond
Date: June 2025

Table of Contents

1. Executive Summary

The simulated semiconductor fab allowed us to evaluate how increased tool capacity affects both wafer-lot wait times and daily production output. The **Before** phase with one etch tool received comparison to the **After** phase which used two identical etch tools to show the following results:

- **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)

The project demonstrates complete end-to-end data skills through its simulation of "dirty" logs and cleaning and wrangling processes and simple statistical analysis and clear visualizations and plain-English storytelling.

2. Introduction

The sequential nature of semiconductor fab tool steps creates delays that propagate throughout the production line to raise work-in-progress (WIP) levels and manufacturing expenses. The duration before a lot enters a tool represents a vital performance metric known as queue time. The reduction of queue times leads to better on-time delivery performance while optimizing the use of expensive tools. The addition of a second etch tool reduces typical waiting times by half while producing a small increase in production output.

---

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

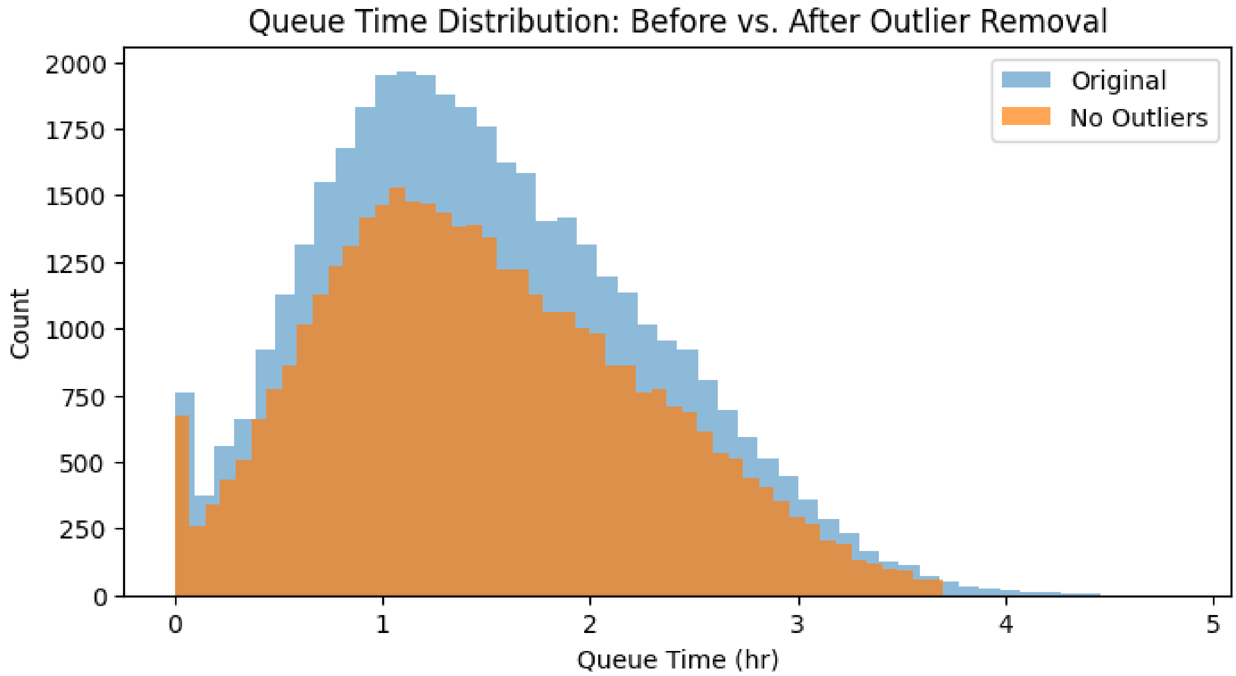
- Applied the **1.5 × IQR rule** on both queue_time_hr and cycle_time_hr.
- Flagged and dropped ~1 % of rows where either metric lay beyond [Q1 − 1.5 IQR, Q3 + 1.5 IQR].

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).

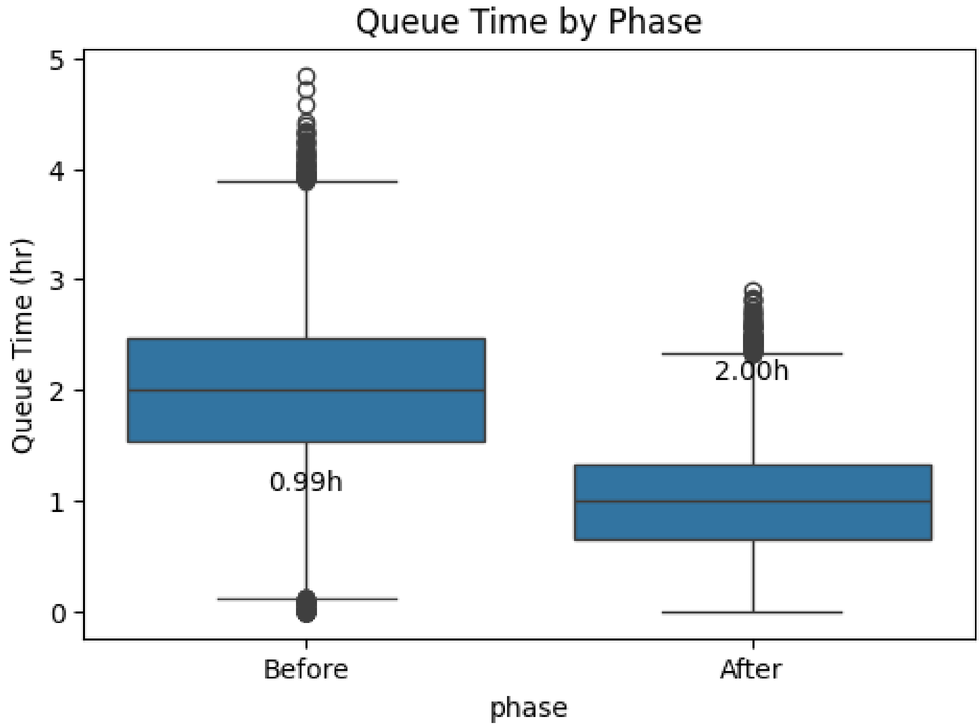
5. Results

5.1 Queue-Time Distribution



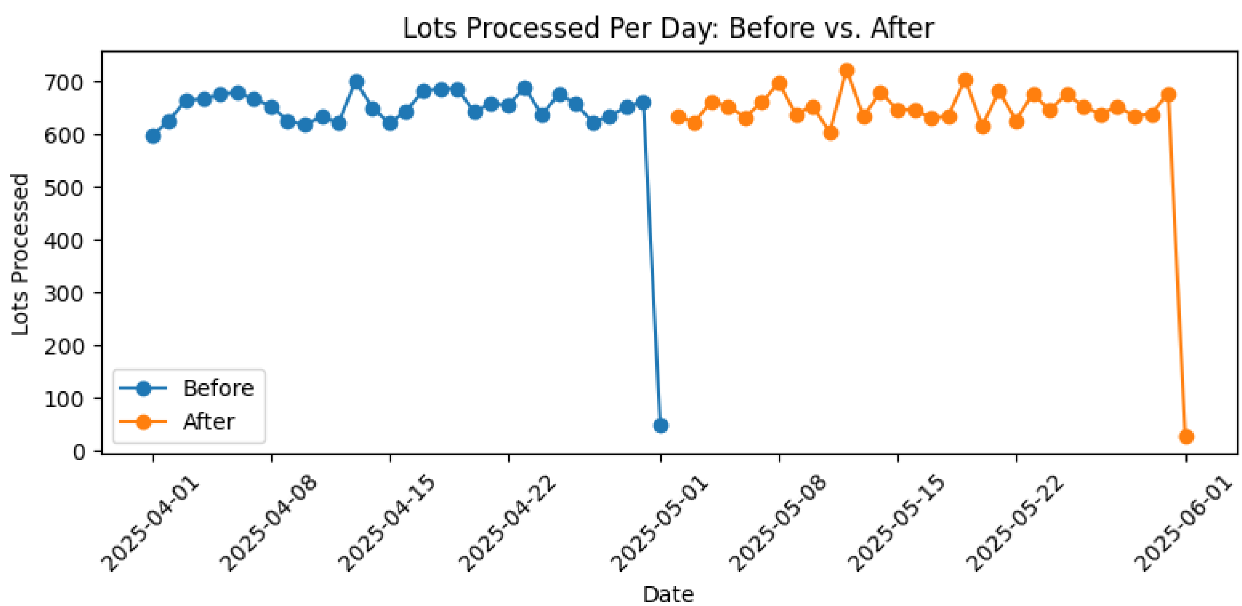
Insight: The removal of outliers revealed that waiting times concentrated around 1 hour instead of the original 2 hours while eliminating all waits exceeding 3.5 hours.

5.2 Queue-Time Comparison by Phase



Insight: The median queue duration decreased by half from 2 hours to 1 hour while the central 50% of waiting times shifted from 1.5–2.5 hours to 0.7–1.3 hours.

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:** The reduction of queue times by half leads to lower WIP holding expenses and faster order delivery. The second tool enables a 7 % increase in daily production output, which results in dozens more lots without requiring major capital investments.

- **Data Skills Demonstrated:** End-to-end handling of messy logs, robust outlier treatment, clear visual storytelling in plain language.
- **Limitations:** The simulation model might not accurately represent actual scheduling complexities or equipment maintenance procedures. Real fab data shows different distribution patterns and time-dependent behavior.

---

7. Conclusions & Recommendations

The analysis proves that tool capacity expansion leads to significant wait time reductions together with moderate production increases. 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:** evaluate how operator staffing patterns affect production results.
- **Quality Integration:** Incorporate defect and yield data to balance speed with quality.