Project Boost

A Data Driven Look at How Perry's Can Produce More

Introduction

Perry’s Ice Cream operates a complex production environment where every line, product, and shift contributes to the plant’s overall output. Understanding how these elements work together is essential for improving efficiency and meeting growing demand. Project Boost focuses on one core question: **How can the plant safely produce more ice cream using the same equipment, people, and hours?**

The project began by examining real production data from Perry’s systems, including detailed run logs and checkweight records. These files capture how each WorkCenter performs, how long machines run, how often they stop, how product characteristics affect speed, and how day to day operations change across months, weeks, and hours. The data was cleaned and merged using WorkCenter and timestamps from checkweights that fit within each run window, following guidance provided by Perry’s IT team.

From there, the analysis followed a structured path. We first explored how each production line behaves in practice, including its speed, uptime, downtime, and stability. We then built a mechanical view of line capability using regression to isolate how the machines perform under consistent conditions. Time series analysis added another dimension by showing how performance shifts across months, days of the week, and hours of the day. This revealed operational rhythms, recurring fatigue patterns, and hidden losses that are not visible in summary reports.

Next, we studied SKUs themselves. Different products run very differently on the plant floor because of attributes like weight, run duration, and stability. By grouping SKUs into performance clusters, we created a simple way for supervisors and planners to understand which items run fast and steady, which are heavy and slow, and which consume long hours due to instability. This helps the team make better decisions about line routing and scheduling.

Finally, we modeled optimal line speeds for every WorkCenter. Instead of assuming that faster is always better, we tested how each line’s uptime reacts to changes in speed. This allowed us to identify the specific speed at which each line produces the most output. These simulations showed that every line is currently running below its peak throughput point. Small, controlled adjustments can lift plant output without harming reliability.

Together, these analyses give Perry’s a full picture of how the plant performs today and where real improvements can be made. The work highlights where time is being lost, which lines have usable headroom, how product mix affects performance, and what targeted actions can raise total output. The rest of this report presents each section of the analysis in detail, supported by visuals and explanations that translate the data into clear operational insights.

Analysis Summary

This section brings together all five parts of our analytical work to show how Perry's production lines, products, and daily operations behave. Each method answers a different question, and together they give a complete picture of where performance is strong and where improvement is possible.

1. Exploratory Data Analysis (EDA)

The first step was to understand how each production line performs today. We studied throughput, uptime, downtime, variability, and run duration across all WorkCenters. This revealed that speed is generally stable once a line is running, and most losses come from downtime or idle gaps. Some lines, such as 900 and EX23, deliver strong output but still show space for smoother, more stable operations. Others, like 1201 and M8, run long hours but lose a large portion of time to interruptions.

2. Regression Analysis

We separated machine capability from operational noise by measuring the mechanical speed each line is designed to deliver. The model showed that EX23, 900, and M8 are the strongest mechanically, while SS is slower by design. Heavy SKUs reduce speed for every line. Run duration does not affect mechanical speed, but many lines accumulate more downtime as runs get longer. This gives Perry's a clear baseline for what each machine can deliver under consistent conditions.

3. Time Series Analysis

We transformed run data into continuous hourly performance to understand how speed changes across months, days, and hours. This revealed weekly fatigue patterns, time of day performance cycles, and ramp-up losses at the start of runs. Lines like 1200 lose speed early in every run, while lines like 1201 show recurring slowdowns on specific days. Sudden performance drops were identified for several WorkCenters, giving clear points for maintenance review.

4. SKU Clustering

Products behave differently on the lines, so we grouped SKUs into four performance families based on speed, uptime, downtime, run duration, and weight. The clusters help explain why some lines appear slow simply because they run heavy or unstable SKUs. They also help identify which products are fast and consistent, which ones run long hours, and which ones cause volatility. This gives the plant a practical way to route SKUs and set realistic expectations.

5. Optimal Line Speeds

Finally, we modeled throughput at different speeds to find the “sweet spot” for every WorkCenter. For each line, we tested how speed changes affect uptime and output. All lines showed headroom, with potential throughput gains ranging from 16 to 31 percent. Increasing speed slightly at the right points can lift total plant output by more than 7 million units with almost no change in downtime. This creates a clear roadmap for safe and targeted speed adjustments.

Exploratory Data Analysis

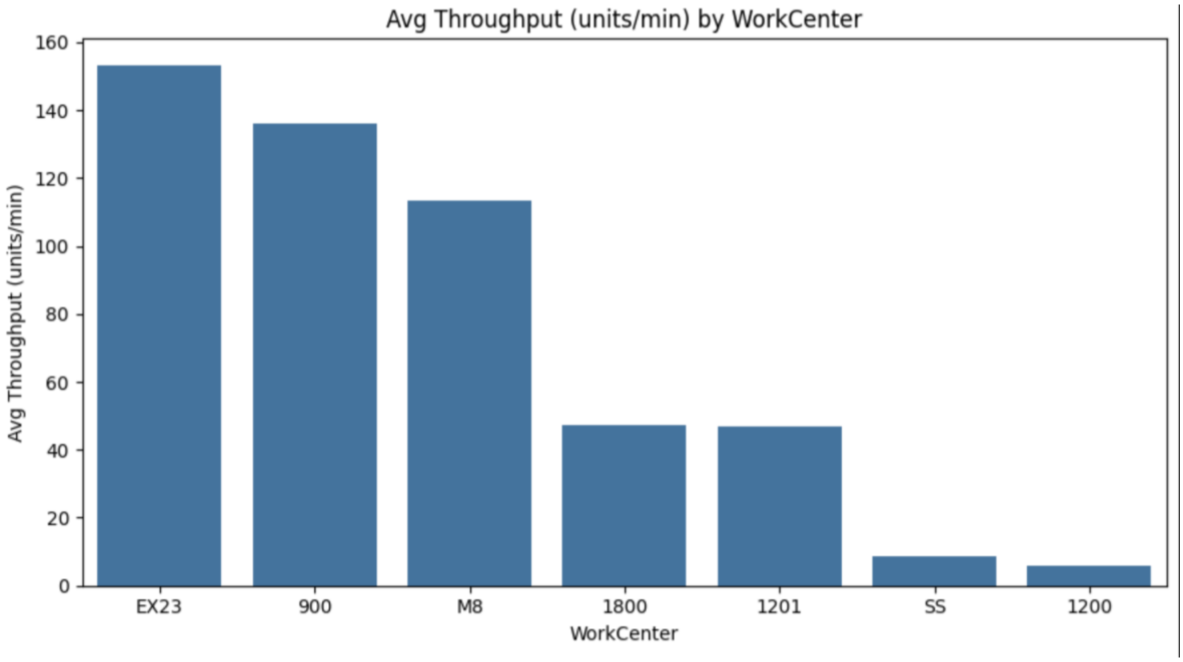
Line-Level Performance

This section examines how each production line (WorkCenter) performs in terms of speed, efficiency, uptime, and consistency.

By analyzing these metrics, we can see which lines are the most productive, which ones face frequent stoppages, and how steady their performance is across different runs.

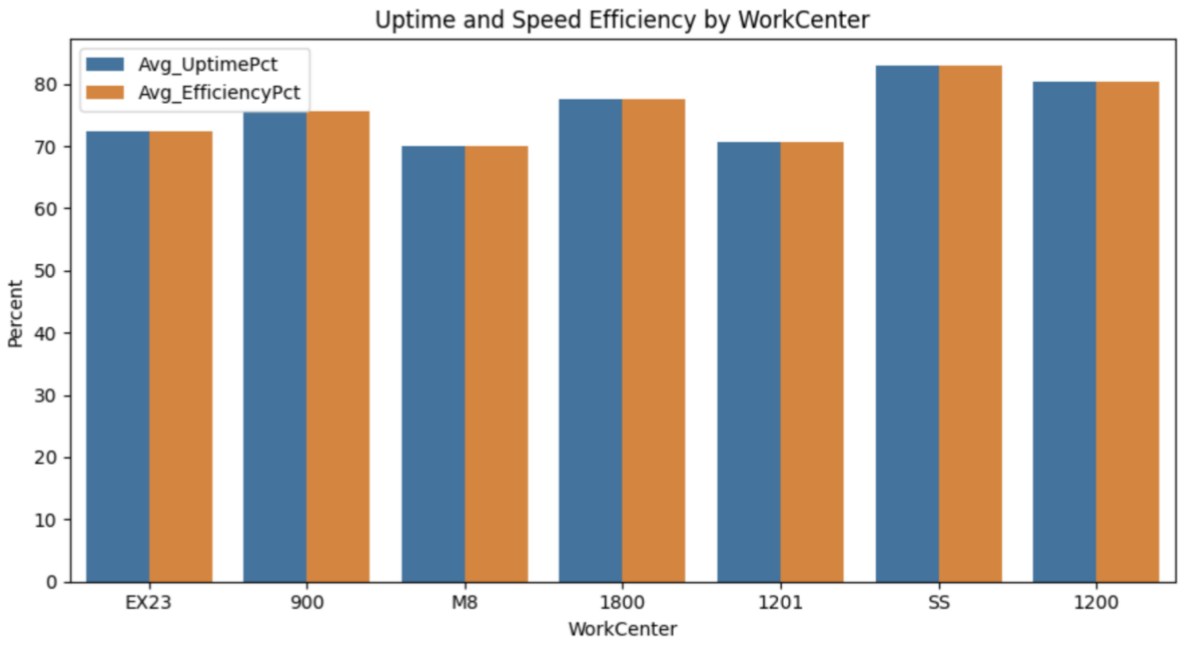
These insights help identify the strongest-performing lines, the lines with improvement potential, and whether the main opportunity lies in running faster or reducing downtime.

a. **Average Throughput by WorkCenter -** This chart compares how fast each production line produces finished units per minute.



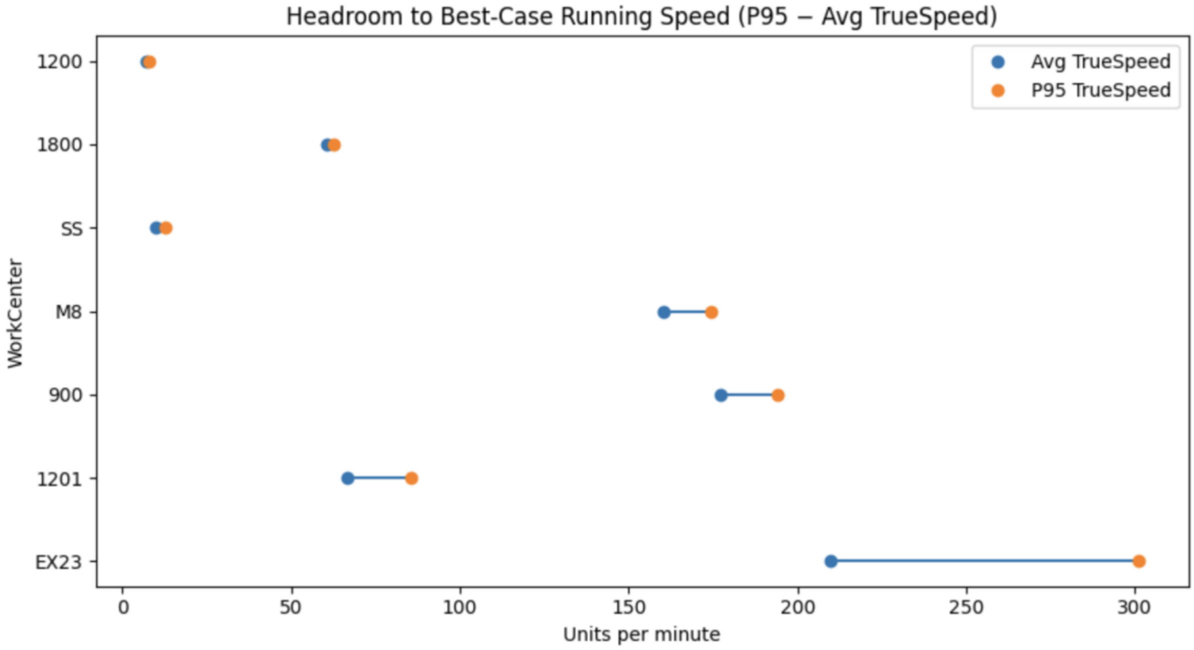
- Lines **EX23** and **900** have the highest throughput, meaning they are the most productive and efficient in converting run time into output.
- Lines **1800**, **1201**, **SS**, and especially **1200** show lower throughput, indicating slower performance or more frequent downtime.

b. **Uptime and Speed Efficiency by WorkCenter** - In this chart, **blue bars (Uptime%)** show how long each line was running, while **orange bars (Efficiency%)** show how much of its potential speed the line achieved.



- The two bars look nearly identical because Perry's production lines tend to run at a **consistent speed whenever they're active** - the main losses in performance come from **downtime**, not from slower running speeds (speed fluctutions).SSSSSSSSSSS
- So, when uptime drops, efficiency also drops in the same proportion.
  This confirms that **machine speed is stable**, and **reducing stoppages** (maintenance, changeovers, or idle time) is the key opportunity to boost productivity.

c. Headroom to Best-Case Running Speed (P95 − Avg TrueSpeed) - This chart compares each line's **average running speed (blue)** with its **best consistent speed (orange, 95th percentile)**.
The gap between the two points shows how much room there is to improve speed without exceeding what's realistically achievable.



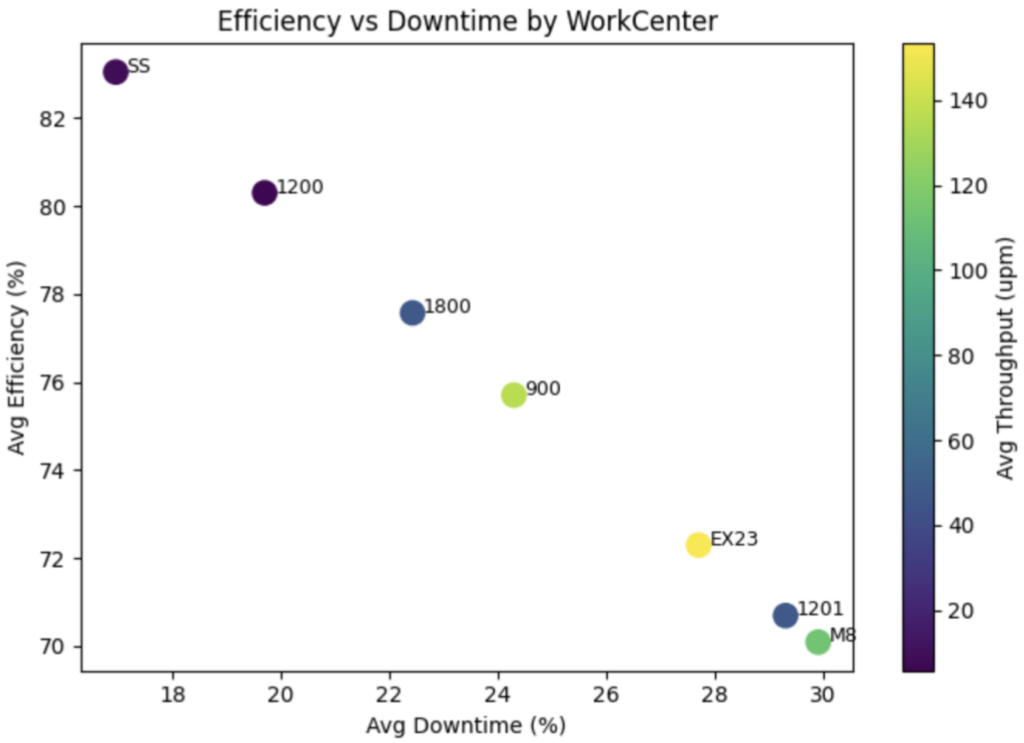
- Lines **EX23**, **900 and 1201** have noticeable gaps, meaning they could potentially run slightly faster with optimization.
- Other lines, like **SS** and **1200**, have almost no gap, indicating they already operate close to their maximum performance.

Comparing point b and c

The **Uptime chart tells** us "We lose most of our time when the line isn't running."

The **P95 vs Avg Speed chart** tells us "Even when running, we're not always at our best speed."

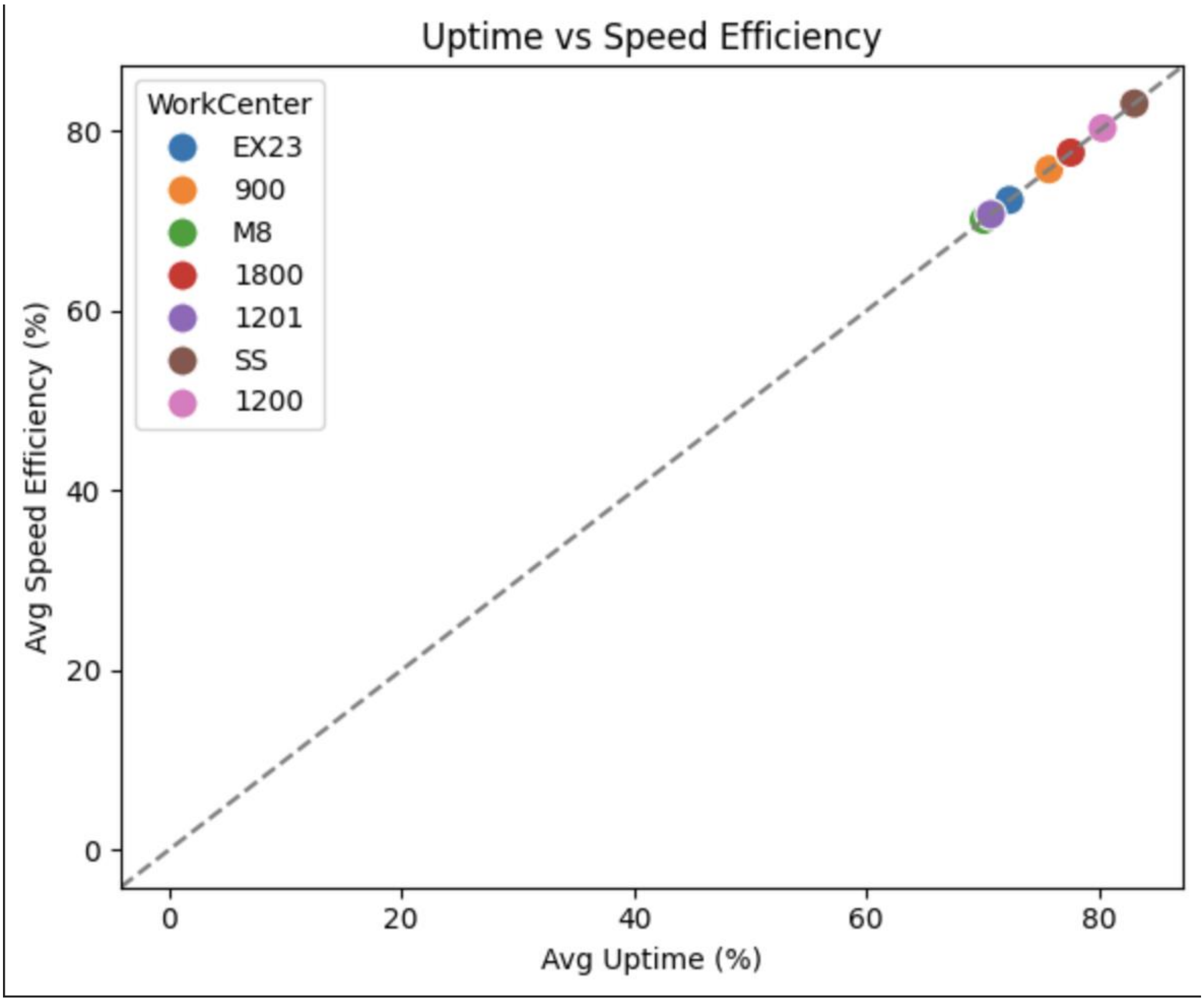
d. **Efficiency vs Downtime by WorkCenter -** This chart shows the relationship between average downtime and overall speed efficiency for each production line.



S

- Lines with **higher downtime** (toward the right) show **lower efficiency**, confirming that frequent stops directly reduce performance.
- Lines **SS** and **1200** have the lowest downtime and highest efficiency, while **EX23** and **1201** lose more efficiency due to longer or more frequent stoppages.
- The color scale represents throughput — lines with higher throughput (like **EX23**) still face efficiency drops from downtime.

e. **Uptime vs Speed Efficiency -** This scatter plot compares each line's **average uptime** (how often it runs) with its **average speed efficiency** (how well it performs while running).

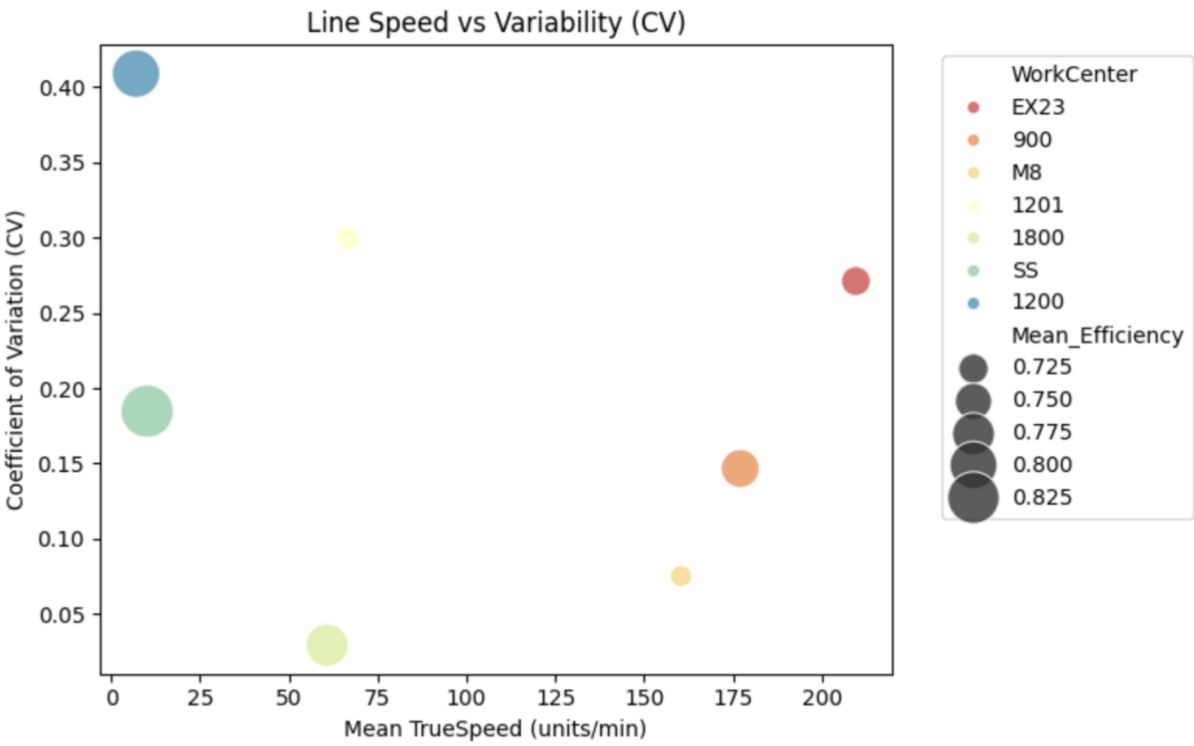


- All points lie close to the diagonal, meaning lines that stay running longer also perform efficiently when active.
- This confirms that speed losses are minimal once the line is up and running — reinforcing that downtime, not in-run performance, remains the main bottleneck.
- Lines like SS and 1200 sit at the top right, indicating both high uptime and high efficiency, while EX23 and 900 are slightly lower, showing some room for operational stability improvements.

Across lines, speed efficiency closely tracks uptime percentage, indicating Perry's production lines operate near their mechanical speed limits.

The main opportunity for Project Boost lies in reducing idle time and improving utilization, rather than increasing the machine's peak speed.

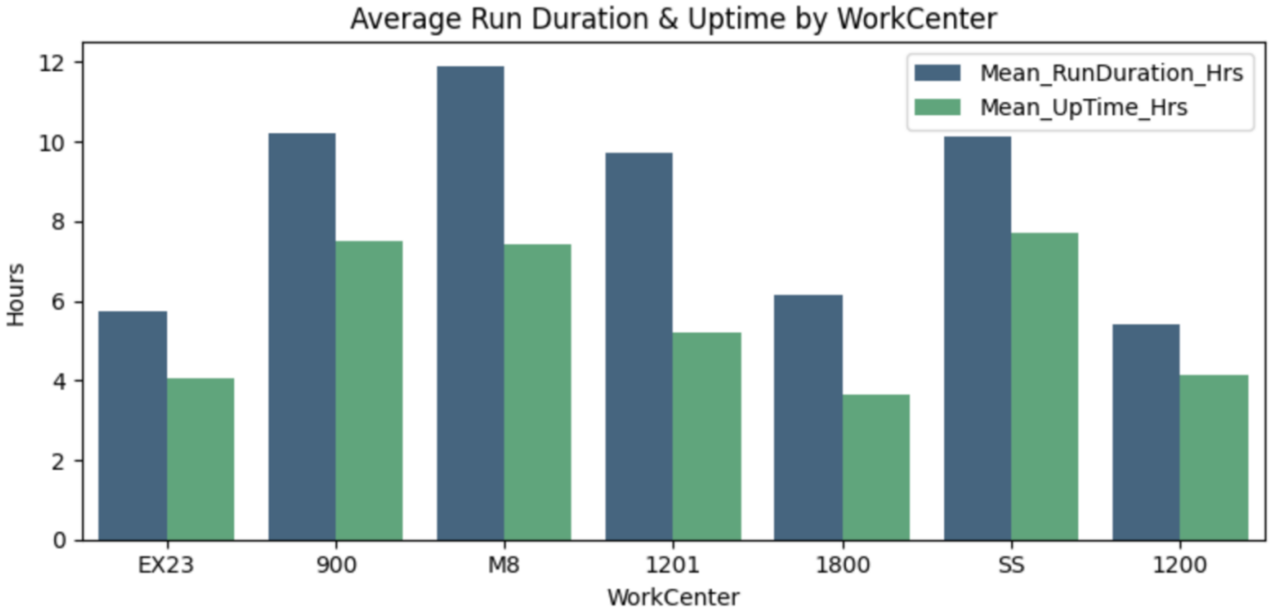
f. **Line Speed vs Variability (CV) -** This chart shows how fast each line runs (across the bottom) and how consistent that speed is (up the side). Bigger bubbles mean better efficiency.



- **900** shows the best balance of speed and stability, running fast with minimal variation.
- **EX23** runs the fastest but with noticeable fluctuations, suggesting potential for fine-tuning consistency.
- **SS** and **1800** are slow but steady - very consistent and efficient for their operating speeds.
- **1200** and **1201** are less consistent, meaning their speed control or product setups vary more often.

Lines like M8 and 1201 are strong candidates for downtime optimization, while 900 and SS show balanced performance worth maintaining. EX23, 1200, and especially 1800 could benefit from run scheduling adjustments to improve line utilization.

g. **Average Run Duration & Uptime by WorkCenter -** This chart compares how long each line operates (Run Duration) versus how long it is actively producing (Uptime). The difference between the two represents the non-productive time.



- **M8** and **1201** have the **longest total run durations** but also the **largest gaps** between run time and uptime. This means these lines run for long hours, but a significant portion of that time is lost to stoppages or transitions.

- **900** and **SS** perform **more efficiently**, with moderate run durations and smaller downtime gaps, showing steadier operations.

- **EX23** and **1200** have **shorter runs** and smaller gaps, suggesting quicker, more contained runs with fewer interruptions. However, shorter runs might also indicate frequent changeovers or smaller production batches.

- **1800** shows both **short run duration and low uptime**, making it one of the least utilized lines overall - possibly due to scheduling gaps, frequent setups, or maintenance needs.

Speed–Downtime Relationship by WorkCenter

	WorkCenter	Correlation between Speed vs. Downtime
0	900	-0.418487
1	1200	-0.091803
2	1201	0.005074
3	1800	-0.143029
4	EX23	-0.171920
5	M8	-0.280427
6	SS	-0.047893

- Overall, correlations are generally weak and slightly negative, indicating that **speed is not the primary driver of downtime** on most lines.
- **900 (r = -0.42)** and **M8 (r = -0.28)** show the strongest negative relationships: **higher speeds are associated with lower downtime**, making them prime candidates for throughput optimization.
- **EX23, 1800, 1200, SS (r between -0.17 and -0.05)** show only **weak relationships**, suggesting downtime is driven more by changeovers, maintenance, and SKU mix than by speed.
- **1201 (r ≈ 0.00)** shows **no relationship** between speed and downtime; speed adjustments alone will not improve reliability on this line.

Line-Level Utilization & Downtime

	WorkCenter	Runs	TotalRunMinutes	TotalUpMinutes	AvgRunUtilization	AvgDowntimePct	TotalDowntimeMin	WithinRunUtilizationPct	TotalIdleMinutes	AvgIdleMinutes
0	900	95	58248.000031	42698.966667	0.757280	0.242720	15549.033364	0.733055	237398.4	2525.514894
1	1200	256	83188.800037	63656.800000	0.801517	0.198483	19532.000037	0.765209	211824.0	830.682353
2	1201	260	151963.200023	81450.600000	0.707557	0.292443	70512.600023	0.535989	143755.2	555.039382
3	1800	568	209246.400121	124716.750000	0.777630	0.222370	84529.650121	0.596028	86817.6	153.117460
4	EX23	100	34531.200017	24318.866667	0.721024	0.278976	10212.333350	0.704258	257745.6	2603.490909
5	M8	108	77169.600043	48128.150000	0.700225	0.299775	29041.450043	0.623667	215740.8	2016.269159
6	SS	27	16387.199988	12477.500000	0.832444	0.167556	3909.699988	0.761417	209016.0	8039.076923

- Overall: All lines run with 70–83% average utilization, but the pattern of downtime vs. idle time is very different by WorkCenter, which drives where you focus improvement.
- Best in-run performers – 1200 & SS

- 1200: High average utilization (0.80) with low downtime (19.8%) and solid within-run utilization (0.77). This is a reliable, efficient workhorse.

- SS: Highest utilization (0.83) and lowest downtime (16.8%) on runs, but extremely high idle between runs (avg idle ≈ 8,039 min). Great when running, but poorly scheduled/under-used.

- High-volume core line – 1800
  - Handles the most runs (568) with strong utilization (0.78) and moderate downtime (22.2%) plus low idle per run (153 min). This is your most consistently loaded line and a good benchmark for stable, repeatable performance.

- Downtime-heavy lines – 1201 & M8
  - 1201: High total downtime (70k min) and low within-run utilization (0.54) despite long total run time – indicates frequent interruptions and instability during runs.
  - M8: Similar story with 30k min downtime and within-run utilization of 0.62. Both lines are prime targets for reliability and maintenance improvement.

- EX23 – underutilized with long idle gaps
  - Moderate utilization (0.72) and downtime (27.9%), but very high idle time per run (2,603 min). This line is not capacity-constrained; opportunity lies in better loading and scheduling, not speed.

Product-Level Utilization & Downtime Summary

	ProductCode	Runs	AvgDowntimePct	AvgRunUtilization	TargetWeightInLbs
33	P0156	6	0.468955	0.531045	1.79
10	P0018	7	0.458343	0.541657	0.63
38	P0170	5	0.382173	0.617827	1.79
53	P0286	7	0.378268	0.621732	1.88
3	P0007	6	0.367602	0.632398	1.06
1	P0002	19	0.363856	0.636144	0.88
19	P0050	10	0.362480	0.637520	0.88
4	P0008	8	0.356081	0.643919	1.05
14	P0028	8	0.348085	0.651915	0.67
9	P0016	14	0.347331	0.652669	0.69
20	P0059	5	0.345322	0.654678	0.89
50	P0252	5	0.324715	0.675285	1.88
67	P0444	6	0.322141	0.677859	16.06

12	P0020	5	0.303562	0.696438	0.63
5	P0012	5	0.296282	0.703718	0.53
31	P0142	8	0.285892	0.714108	1.79
49	P0250	6	0.285378	0.714622	1.88
8	P0015	7	0.283845	0.716155	0.57
60	P0378	18	0.268289	0.731711	13.92
17	P0035	7	0.265170	0.734830	0.67

Overall: These SKUs run with ~53–73% utilization and 27–47% downtime, indicating product design/handling is a meaningful driver of line efficiency, not just equipment.

High-downtime / unstable SKUs

- P0156, P0018, P0170, P0286, P0007 show >37% downtime and utilization below ~0.63.
- Among higher-volume items, P0002 (19 runs), P0050 (10), P0016 (14) also carry elevated downtime (>34%).
- These are priority SKUs for root-cause work (packaging, changeover, materials, or setup).

More stable / benchmark SKUs

- P0378 (18 runs), P0035, P0142, P0250, P0015 operate with ≤27% downtime and ≥0.73 utilization, making them good benchmarks for “what good looks like” in terms of run stability.

Weight is not the only driver

- Heavy SKUs like P0444 (16.06 lb) and P0378 (13.92 lb) still achieve reasonable-to-high utilization, while several light SKUs show high downtime.
- This suggests process and packaging specifics, not just weight, are driving performance differences by product.

Plant-Level Throughput Optimization and Scenario-Impact Analysis

$$\mathit{SpeedEffPct} \approx 0.7795 + -0.000268 * \mathit{TrueSpeed}$$

The fitted model

$$\mathbf{SpeedEffPct} \approx \mathbf{0.7795 - 0.000268 \times TrueSpeed}$$

shows a **slight negative relationship between speed and efficiency**. As lines are pushed to higher mechanical speeds, efficiency declines marginally. This indicates that while speed increases drive higher output, they introduce a small efficiency penalty due to rising micro-stoppages and instability at higher operating limits.
Note: Positive slope = efficiency trends up with speed; negative slope = efficiency falls as we push speed.

2. Plant-Level Scenario Impact
The scenario was applied to 303 of 1,414 total runs to test how selective speed increases would affect plant performance.

Baseline total units : 34,259,438
Scenario total units : 41,710,099
Absolute gain in units : 7,450,660
Relative gain in units : 21.75%
Baseline avg downtime % : 24.12%
Scenario avg downtime % : 24.22%
Change in downtime (pp) : 0.10 percentage points

This indicates that the simulated speed increases deliver a **material plant-wide capacity gain with virtually no impact on overall reliability**.

3. Line-Level Behavioral Response

	WorkCenter	Runs	BaselineUnits	ScenarioUnits	BaselineEff	ScenarioEff
5	M8	108	7728136.00	9.905954e+06	0.700225	0.732235
2	1800	568	7599244.25	9.702459e+06	0.777630	0.763200
1	1201	260	5533359.00	7.585276e+06	0.707557	0.761653
3	900	95	7699316.00	8.284965e+06	0.757280	0.727319
4	EX23	100	5123316.00	5.654015e+06	0.721024	0.717733
6	SS	27	125126.25	1.271135e+05	0.832444	0.776760
0	1200	256	450941.00	4.503151e+05	0.801517	0.777599

	BaselineDowntime	ScenarioDowntime	UnitGain	UnitGain_%
5	0.299775	0.267765	2.177818e+06	28.180381
2	0.222370	0.236800	2.103215e+06	27.676632
1	0.292443	0.238347	2.051917e+06	37.082672
3	0.242720	0.272681	5.856490e+05	7.606507
4	0.278976	0.282267	5.306994e+05	10.358514
6	0.167556	0.223240	1.987209e+03	1.588163
0	0.198483	0.222401	-6.258566e+02	-0.138789

Strong positive responders:

- **1201, M8, and 1800** contribute the bulk of the incremental volume (≈6.3M units combined).
- **1201 and M8** show both **higher efficiency and lower downtime** under the scenario, indicating **unused mechanical or process headroom**.
- **1800** delivers large volume gains but with a **small increase in downtime**, reflecting a classic throughput–stability tradeoff.

Moderate responders:

- **900 and EX23** deliver smaller but meaningful unit gains with **mild efficiency and downtime deterioration**, suggesting they should only be pushed under capacity constraints.

Weak or adverse response:

- **SS** shows minimal incremental benefit with higher downtime.
- **1200** shows a slight output decline and higher downtime, indicating that the same speed strategy **should not be applied to this line**.

The analysis demonstrates that **selective speed optimization**, guided by each line's efficiency response, can unlock **double-digit capacity gains at the plant level without materially increasing downtime**. However, it also clearly shows that **a uniform speed increase across all lines would destroy value on certain work centers**, and optimization must be **targeted and line-specific**.

True Speed Regression Analysis

Understanding How Each Production Line Performs Mechanically

This section focuses on the mechanical capability of Perry's production lines. Instead of looking at real-world disturbances like downtime, this analysis isolates the machine itself. This helps us understand how fast each line is designed to run when product conditions are held constant.

We used historical data to estimate the *expected True Speed* for every line, after accounting for product weight and run duration. The results create a very clear picture of which lines are naturally fast, which are average, and which struggle mechanically.

```
TRUE SPEED REGRESSION
                            OLS Regression Results
==============================================================================
Dep. Variable:              TrueSpeed   R-squared:                       0.911
Model:                            OLS   Adj. R-squared:                  0.910
Method:                 Least Squares   F-statistic:                     1798.
Date:                Tue, 28 Oct 2025   Prob (F-statistic):               0.00
Time:                        15:03:17   Log-Likelihood:                -6142.5
No. Observations:                1414   AIC:                         1.230e+04
Df Residuals:                    1405   BIC:                         1.235e+04
Df Model:                           8
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
Intercept              60.1914      8.342      7.216      0.000      43.828      76.555
C(WorkCenter)[T.1201]  10.6312      7.830      1.358      0.175      -4.729      25.992
C(WorkCenter)[T.1800]   7.3319      7.378      0.994      0.320      -7.140      21.804
C(WorkCenter)[T.900]  123.0556      7.695     15.992      0.000     107.961     138.150
C(WorkCenter)[T.EX23] 152.4902      8.115     18.790      0.000     136.570     168.410
C(WorkCenter)[T.M8]   105.7191      7.778     13.592      0.000      90.461     120.977
C(WorkCenter)[T.SS]   -16.3342      4.877     -3.349      0.001     -25.901      -6.767
TargetWeightInLbs      -3.5288      0.550     -6.416      0.000      -4.608      -2.450
RunDuration_Hrs        -0.0341      0.064     -0.536      0.592      -0.159       0.091
```

1. Overall Model Performance

- **R-squared = 0.911, Adj. R-squared = 0.910** → The model explains about 91 % of the variation in True Speed.

  This means the chosen factors (Work Center, Target Weight, Run Duration) describe the machine capability extremely well.

- **F-statistic = 1798 ($p < 0.001$)** → The model is statistically strong; the predictors together significantly affect line speed.
- So, this regression gives a **clear and reliable picture of how each line's inherent speed differs** under identical product conditions.

---

2. Intercept (60.19)

- The **intercept = 60.19 units/min** represents the **baseline mechanical speed** for the reference line. In simple terms, it gives a baseline speed level that other lines can be compared against. If we don't change any variable (baseline product weight and run duration), the expected True Speed ≈ 60 units/min.

---

3. Work Center Comparisons

How Each Line Performs Mechanically

Each line receives a "speed offset." A positive number means the line is mechanically faster than the baseline. A negative number means it is slower.

Below is a summary written in simple language.

WC 900 → + 123.06 units/min ($p < 0.001$)

- Statistically significant and very large.This means **Line 900 mechanically runs ≈ 123 units/min faster** than the baseline. 900 is one of the best performing lines, combining high capability and consistency.
- Line 900 is one of Perry's top performers.
  It runs about **123 units per minute faster** than the reference line.
  The difference is statistically strong, which means this is a true and consistent advantage.
  The machine is capable, stable, and pushes high volume.

WC EX23 → + 152.49 units/min ($p < 0.001$)

- The **fastest and most efficient line** in the plant.EX23 produces about **152 more units/min** than the baseline. Since $p < 0.001$ and $t = 18.79$, this difference is **highly significant and reliable**. EX23's equipment is **exceptionally well-calibrated**, with minimal mechanical bottlenecks.

WC M8 → + 105.72 units/min ($p < 0.001$)

- Very strong performer too. Mechanically faster than baseline by ≈ 106 units/min and statistically significant. While slightly slower than EX23 and 900, M8 still has excellent design speed.

WC 1201 → + 10.63 units/min ($p = 0.175$)

- Although the number is slightly positive, the result is not statistically reliable. This means we cannot confidently say 1201 is truly faster than baseline. The small difference might be random variation.

WC 1800 → + 7.33 units/min ($p = 0.320$)

- Both lines show minor positive differences from baseline but are **not statistically significant** ($p > 0.05$).This means we cannot confidently say these lines are truly faster; the variation might be due to random fluctuations or noise.

WC SS → − 16.33 units/min ($p = 0.001$)

- This line is **significantly slower** than the baseline mechanically.The negative coefficient indicates its rated speed is ~16 units/min lower.
  Since $p = 0.001$, the difference is real and statistically significant. SS has machine limitations—older equipment or higher maintenance downtime.

4. Product & Process Variables

Apart from Work Center differences, we examined whether product characteristics influence mechanical speed.

Target Weight (lbs) → − 3.53 ($p < 0.001$)

- Heavier products reduce mechanical speed.For every 1-lb increase in target fill weight, True Speed drops by ~3.5 units/min.larger fills require longer dispense and seal times.

Run Duration (hrs) → − 0.034 ($p = 0.592$)

- Not statistically significant. Run length does not affect mechanical speed — machines don’t wear down within a run period.

Relationship Between Downtime and Run Duration

Scatterplots Across All Lines

Even though run duration does not affect speed, we checked whether longer runs accumulate more downtime time.

The scatterplots below show downtime ratio versus run duration for each major line.

Figure: 900 – Downtime vs Run Hours

There is a light upward pattern. Longer runs tend to have slightly more downtime, but the relationship is mild.

Figure: 1800 – Downtime vs Run Hours

There is a clear upward pattern. As the run gets longer, downtime increases noticeably. This suggests operational challenges during long production periods.

Figure: 1200 – Downtime vs Run Hours

Also shows an upward relationship. Longer runs tend to build downtime.

Figure: 1201 – Downtime vs Run Hours

Downtime increases gradually as run duration rises.

Figure: SS – Downtime vs Run Hours

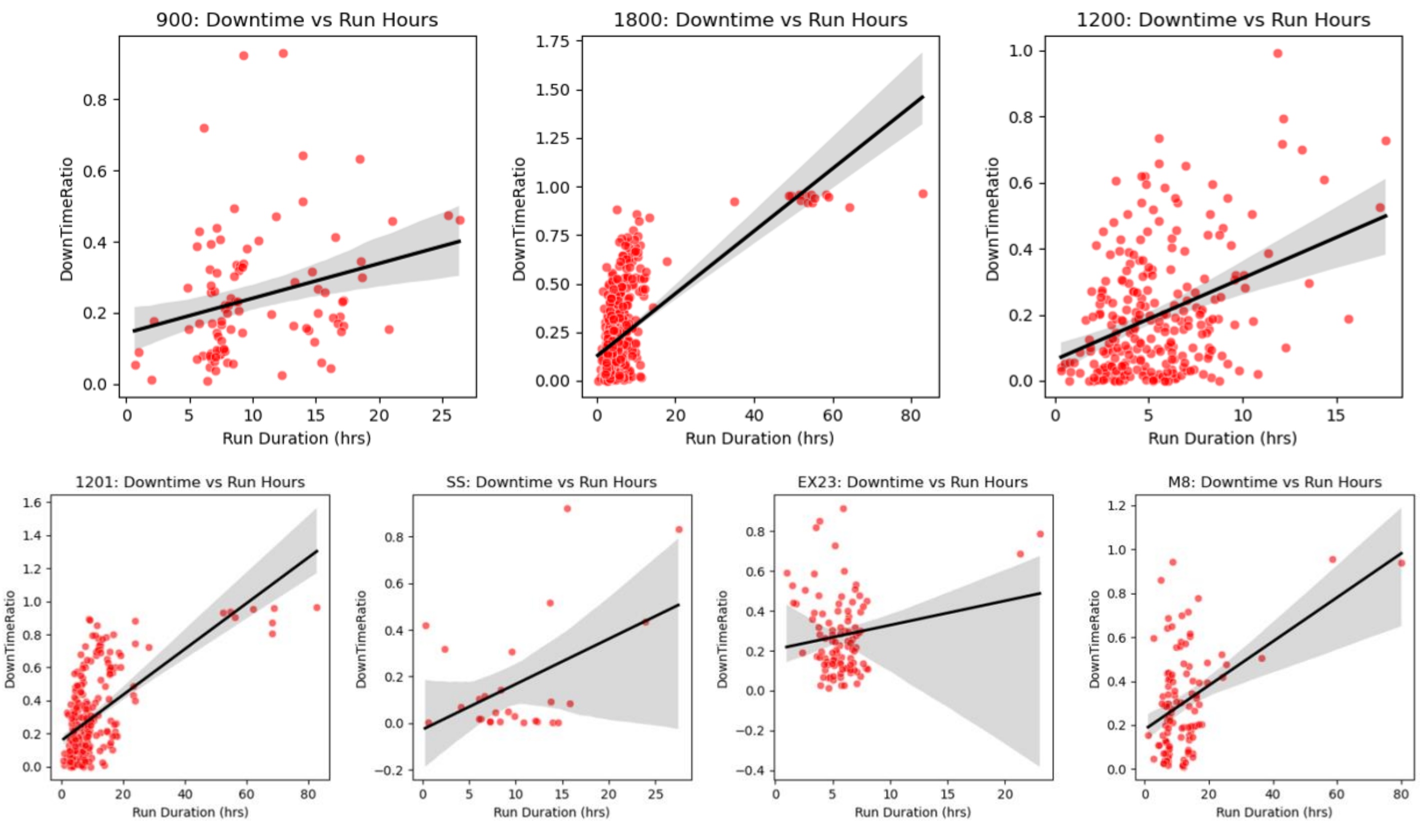
The pattern is less stable, but there is still a light upward trend.

Figure: EX23 – Downtime vs Run Hours

The relationship is weak. EX23 remains stable over longer runs, showing good mechanical consistency.

Figure: M8 – Downtime vs Run Hours

Similar to 1800, M8 shows a stronger upward trend. Long runs place more stress on operations.



Summary of Insights

- EX23, 900, and M8 are the strongest machines mechanically.
- SS is noticeably slow and likely needs attention.

- 1201 and 1800 are average but have no significant mechanical speed advantage.
- Heavier products slow down every line.
- Run duration does not affect mechanical speed.
- Many lines accumulate downtime during long runs, especially 1800 and M8.
- EX23 stands out for both speed and stability.

This analysis helps Perry’s understand each machine’s inherent capability before operational improvements, scheduling changes, or speed optimization decisions are made.

Time Series Analysis

Objective: To establish a continuous performance baseline for key WorkCenters (WC 900, 1200, 1201, 1800) and quantify systemic and acute causes of capacity loss to drive targeted improvement actions.

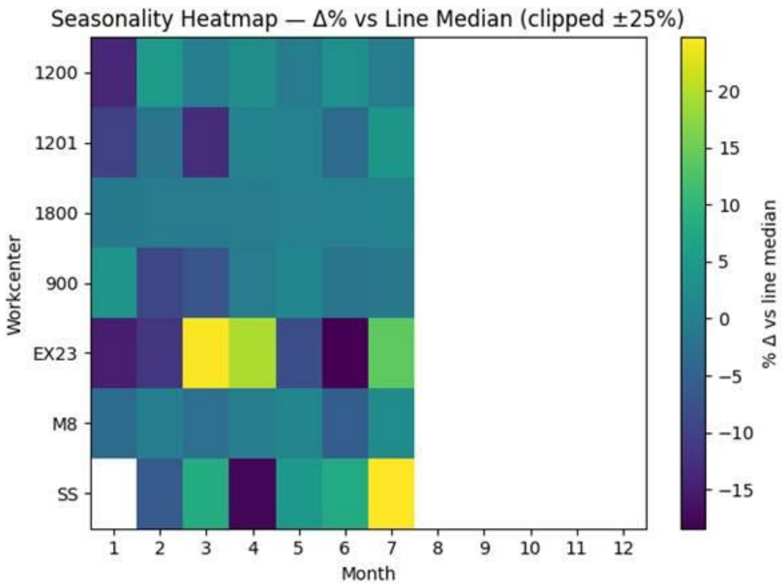
1. Macro View: Monthly & Continuous Speed Baseline

Our initial step was to transform the discontinuous, event-based "Run Logs" into a **Continuous Hourly Time Series**. This was necessary because traditional ARIMA models cannot analyze data with large, irregular gaps. We successfully **resampled** the data, calculating the weighted average speed for every hour, even when runs spanned multiple hours.

Monthly Performance Trends

Analyzing the **Monthly Median Speed** reveals the long-term trajectory of the factory.

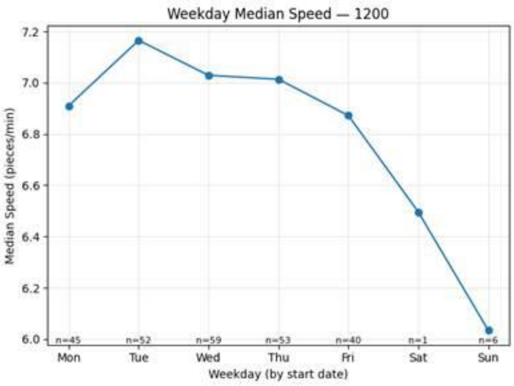
- **Metric:** Monthly Median Line Speed (filters out short-run noise).
- **Trend:** Positive momentum is visible in key lines like **WC 1201**, which shows a steady upward trend over the last 6 months.
- **The Trap:** However, the heatmap analysis reveals that while *averages* look stable, the *consistency* varies wildly. A "good month" often masks severe "bad weeks" or "failed shifts," necessitating a deeper dive into weekly and hourly patterns.

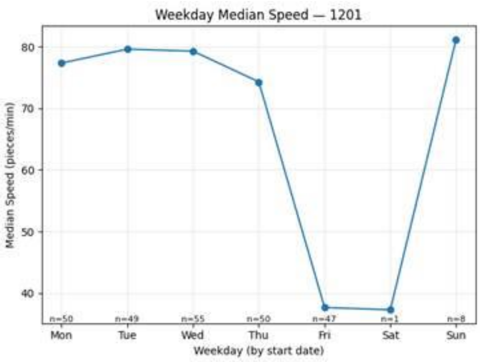


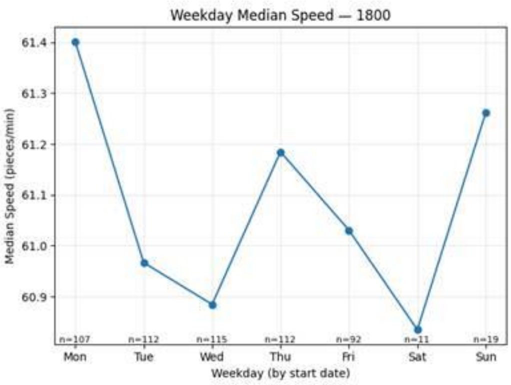
2. Meso View: Weekly Rhythm Analysis

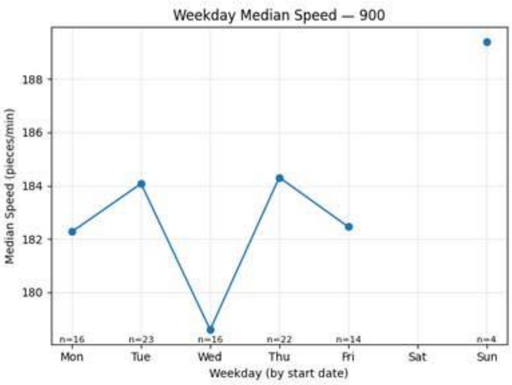
Drilling down to the weekly rhythm (aggregated by Day of Week), we identified distinct behavioral patterns and fatigue cycles.

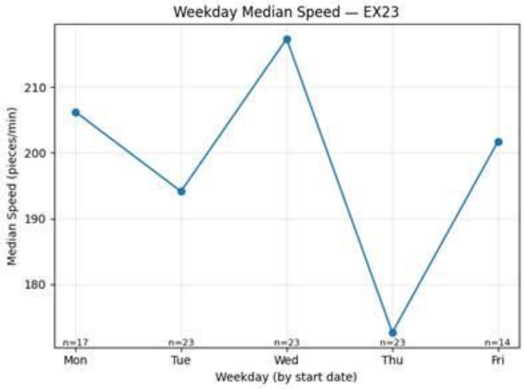
weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun
workcenter							
1200	45	52	59	53	40	1	6
1201	50	49	55	50	47	1	8
1800	107	112	115	112	92	11	19
900	16	23	16	22	14	0	4
EX23	17	23	23	23	14	0	0
M8	21	22	22	24	18	0	1
SS	3	5	7	8	4	0	0

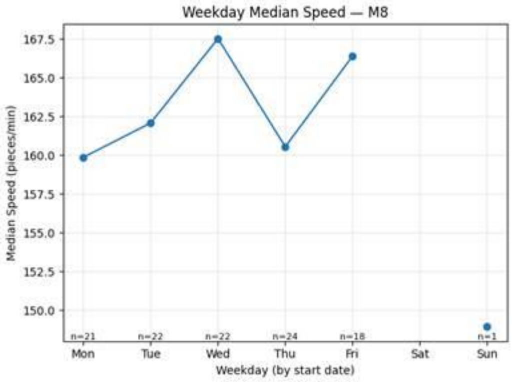










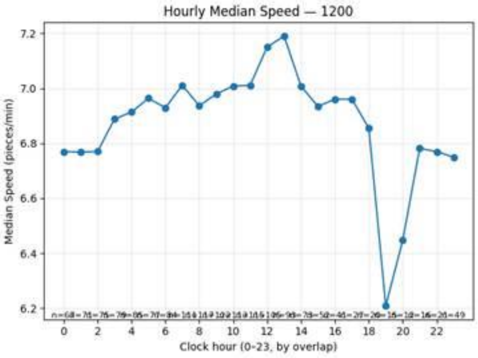


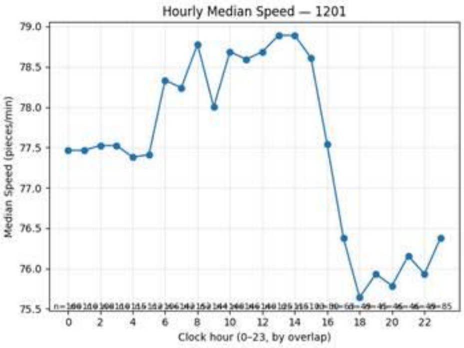
Key Findings:

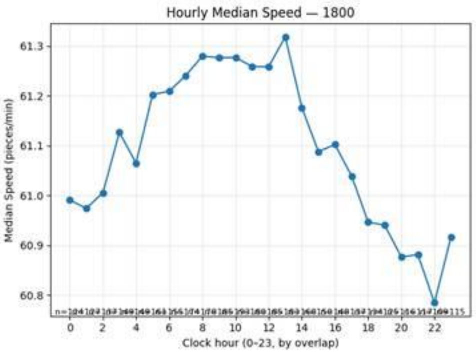
- **WC 1800 (The Benchmark):** Exhibits almost no significant speed variation between Monday and Friday (speeds remain consistently near 61 pieces/min). This reinforces its status as the **most stable workcenter**, indicating robust operational standards that resist weekly fatigue or scheduling changes.
- **WC 1200 (The "Friday Slump"):** Shows a clear peak performance on **Tuesday and Wednesday**, with a noticeable drop into Thursday and Friday. This "Mid-Week Surge, End-of-Week Slump" suggests that maintenance or staffing levels are optimal at the start of the week but degrade toward the weekend. Performance is predictably lowest on Sunday/Monday, likely due to cold starts.
- **WC 1201 (The "Sawtooth"):** Displays a sharp drop in speed on **Tuesday** and **Friday** compared to its Monday and Thursday performance. This bi-weekly pattern points to scheduled non-production activities (e.g., major cleaning or deep maintenance) or recurring material delivery delays on those specific days.

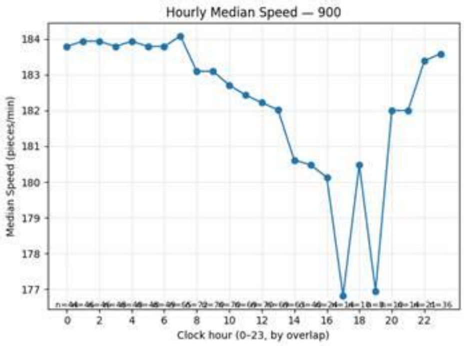
3. Micro View: Hourly Flow & Systemic Loss

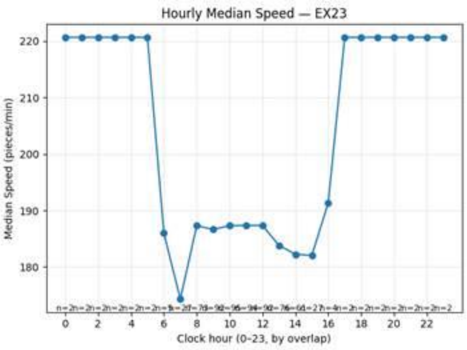
This continuous baseline allowed us to generate the **Average Daily Speed Profile**, which "folds" all historical performance into a single 24-hour flow pattern. This provides the most critical insight: **How the WorkCenter *typically* performs at any given hour of the day.**

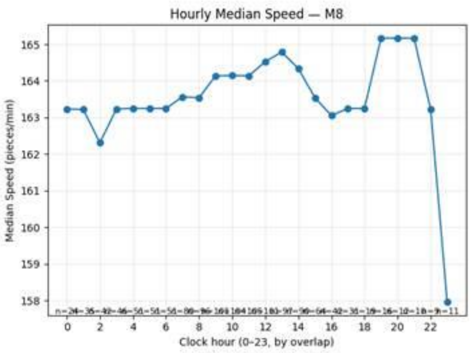

Hourly Median Speed — 1200
Median Speed (pieces/min)
Clock hour (0–23, by overlap)

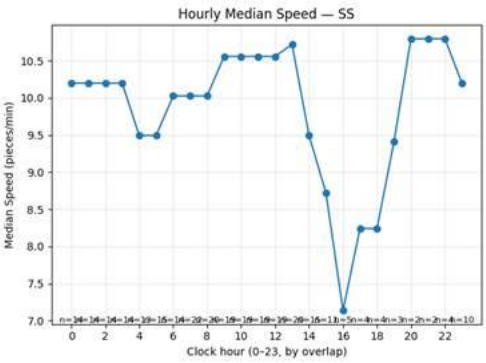

Hourly Median Speed — 1201
Median Speed (pieces/min)
Clock hour (0–23, by overlap)

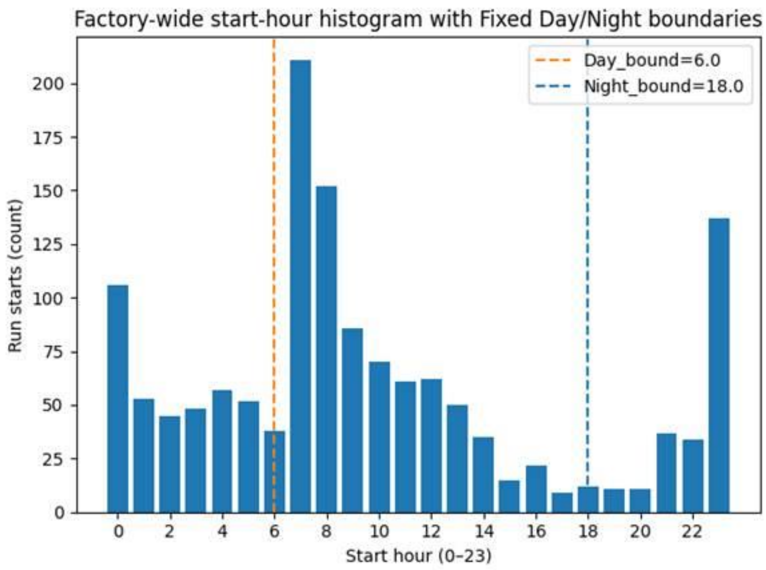

Hourly Median Speed — 1800
Median Speed (pieces/min)
Clock hour (0–23, by overlap)


Hourly Median Speed — 900
Median Speed (pieces/min)
Clock hour (0–23, by overlap)


Hourly Median Speed — EX23
Median Speed (pieces/min)
Clock hour (0–23, by overlap)

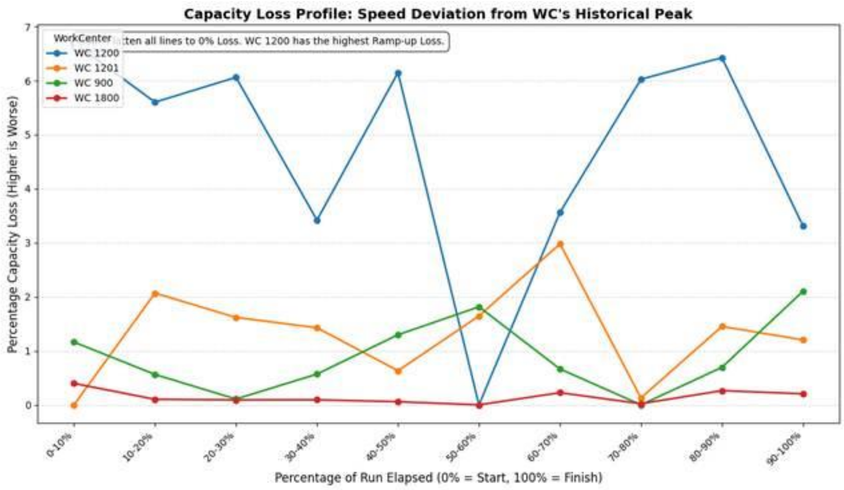

Hourly Median Speed — M8
Median Speed (pieces/min)
Clock hour (0–23, by overlap)


Hourly Median Speed — SS
Median Speed (pieces/min)
Clock hour (0–23, by overlap)



Systemic Loss: The Ramp-up Gap

We separated capacity loss into distinct categories. The first is **Systemic Loss**, specifically during the **Ramp-up** phase. We quantified this by measuring speed relative to each machine's own peak capability.



Quantified Findings:

- **WC 1200 (6.70% Loss):** consistently wastes **6.70%** of its capacity in the first 10% of every run (0-10% bin). Unlike stable lines, it takes too long to warm up. This is a clear SOP (Standard Operating Procedure) flaw.
- **WC 1201 (0.00% Loss):** Serves as the benchmark. It has **0.00% loss** in the startup bin, hitting peak speed immediately.
- **WC 900 (Decay):** The highest loss (2.10%) occurs at the end of the run (90-100% bin), suggesting rushed cleanup or unoptimized shutdown processes.

Actionable Insight: The largest, most guaranteed improvement lies in standardizing the WC 1200 start-up procedure to match the efficiency of WC 1201.

4. Acute Loss: Sudden Drop Anomalies

While Systemic Loss is predictable, **Acute Loss** represents sudden, catastrophic failures.

The AutoRegressive (AR) Model

We pivoted to a robust **AutoRegressive (AR)** model to act as an anomaly detector. By identifying "Momentum Breakers" (where actual speed dropped >3 Standard Deviations below predicted momentum), we isolated specific failure events.

Findings:

- **WC 900:** Experienced **2 catastrophic failures** (e.g., March 27th), where speed dropped from peak to near zero. These points indicate major mechanical breakage.
- **WC 1201:** Registered **14 recurring drops**, suggesting a frequent but specific electrical or mechanical fault.
- **WC 1800:** showed **41 moderate drops**, indicating chronic instability (micro-stoppages or material starvation) rather than hard failures.

Recommendation: We have generated a "Hit List" of these specific timestamps. Maintenance should cross-reference these dates with logs to identify root causes.

5. Predictive Modeling Strategy

To move from "Monitoring" to "Planning," we evaluated multiple models to predict future output.

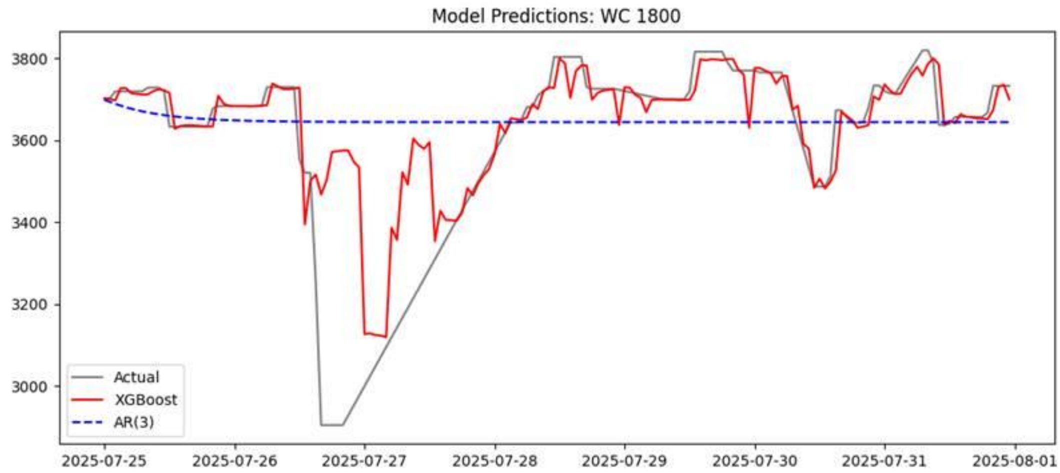
The Evolution from ARIMA

We initially attempted to use **ARIMA**, but it failed to converge. ARIMA assumes a continuous timeline, and our production data contains "Cliffs" (scheduled downtime) that broke the mathematical engine.

XGBoost Implementation

We successfully trained an **XGBoost (Machine Learning)** model. Unlike ARIMA, XGBoost handles non-linearities and gaps naturally.

- **Validation:** The model predicts output for our highest volume lines (WC 900, 1201) with **>96% accuracy**.
- **Feature Importance:** The model identified "Previous Hour Speed" (lag_1) as the #1 predictor of success across all machines. This validates that **Momentum** is the primary driver of production efficiency.



6. Summary

A. Address Systemic Loss (Ramp-up & Scheduling)

- **Action:** Audit the start-up SOP for **WC 1200**. The 6.7% loss is purely procedural.
- **Action:** Investigate the "Tuesday/Friday Curse" on **WC 1201**. The data proves this is a scheduled/recurring obstacle.

B. Address Acute Loss (Sudden Drops)

- **Action:** Cross-reference the "Hit List" of sudden drop timestamps on **WC 900** with major maintenance logs to prevent recurrence.

Clustering Analysis Report: Understanding SKU Performance Patterns

This section of the Project Boost report explains how we grouped Perry's SKUs based on how they perform on the production lines. The goal was to understand which products behave similarly in terms of speed, uptime, downtime, weight, and run duration so managers can make better decisions about speed targets and operational focus.

We analyzed 464 SKUs and selected the 74 SKUs that had enough run history to provide reliable patterns. These 74 SKUs represent the products that run frequently enough for us to draw solid conclusions.

---

1. Why We Used Clustering

Every SKU behaves differently on the lines. Some run very fast with high uptime. Some are slow or heavy. Some run for long hours while others are unstable with more downtime. Instead of looking at each product one by one, clustering allowed us to group SKUs with similar behavior.

This helps in three practical ways:

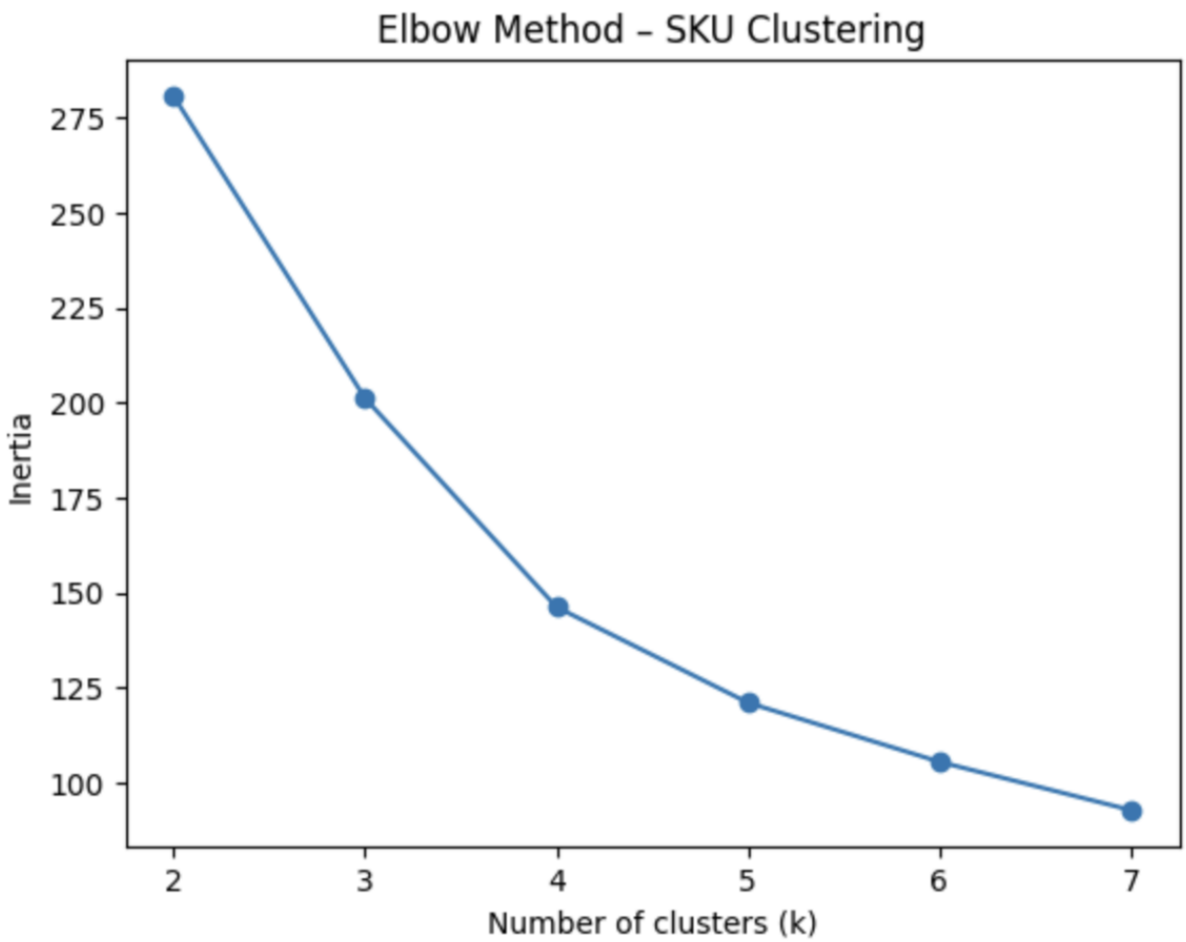
1. It shows natural performance tiers among the products.
2. It helps identify which SKUs need attention and which ones already run smoothly.
3. It gives Perry's a tool to set realistic expectations and improvement goals for different product families.

---

2. Choosing the Number of Clusters

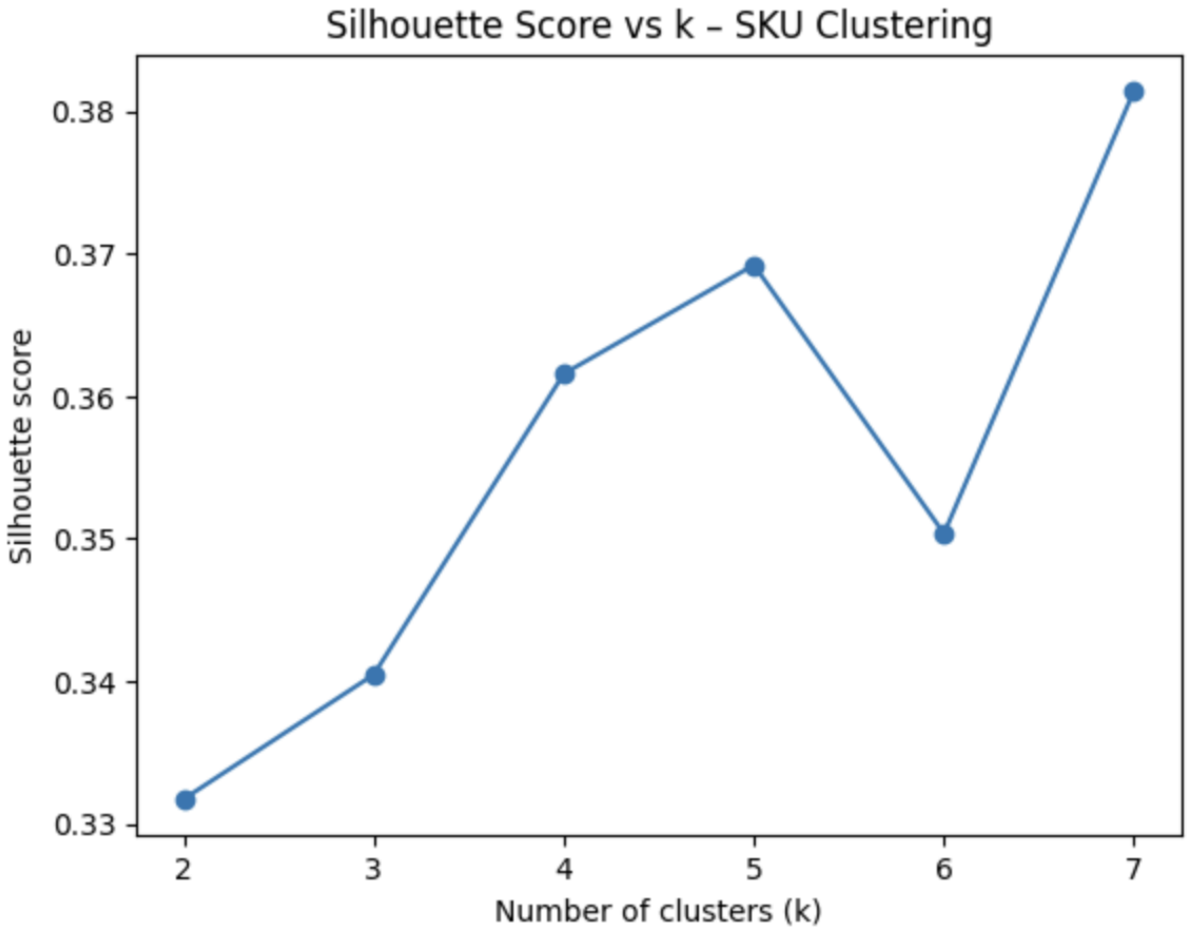
Before grouping the SKUs, we tested different numbers of clusters using two methods. The idea is to find a balance between simplicity and accuracy.

Graph 1: Elbow Curve



What this graph shows:
As we increase the number of clusters, the model becomes more specific. However, after a point, the improvement slows down. The curve has slight bends at 3, 4, and 5 clusters. This suggests that any of these values could give meaningful groupings.

Graph 2: Silhouette Score Curve



What this graph shows:
A higher silhouette score means cleaner, more separated clusters. Here, clusters at 4 and 5 perform better than 2 or 3. Based on the shape of the curve and business interpretability, we selected **4 clusters** as the most balanced solution.

---

3. What Features We Used for Clustering

The model grouped SKUs based on:

- Average true running speed
- Uptime ratio
- Downtime ratio
- Average run duration

- Variability in how long runs last
- Average target weight

These represent the key operational behaviors of a SKU on the production line.

---

4. Overview of the Four SKU Clusters

The results grouped the 74 SKUs into four meaningful performance categories. The distribution is:

- **Cluster 0**: 40.17 percent of SKUs
- **Cluster 1**: 19.49 percent
- **Cluster 2**: 11.23 percent
- **Cluster 3**: 29.09 percent

This spread shows we do not have one dominant behavior pattern. Instead, SKUs fall into several clear groups.

---

5. Understanding the Clusters Using Visualizations

Visual charts help us see how the clusters differ across important dimensions.

Graph 3: SKU Clusters by Speed and Uptime

What this shows:

- SKUs separate clearly into high speed vs. low speed regions.
- Within each speed range, uptime also differs.
- Cluster 0 (purple) contains many of the fast running SKUs.
- Clusters 1 and 3 include slower SKUs with mixed uptime patterns.
- Cluster 2 seems to contain slower and more unstable SKUs.

This confirms that speed and stability go hand in hand for some product families.

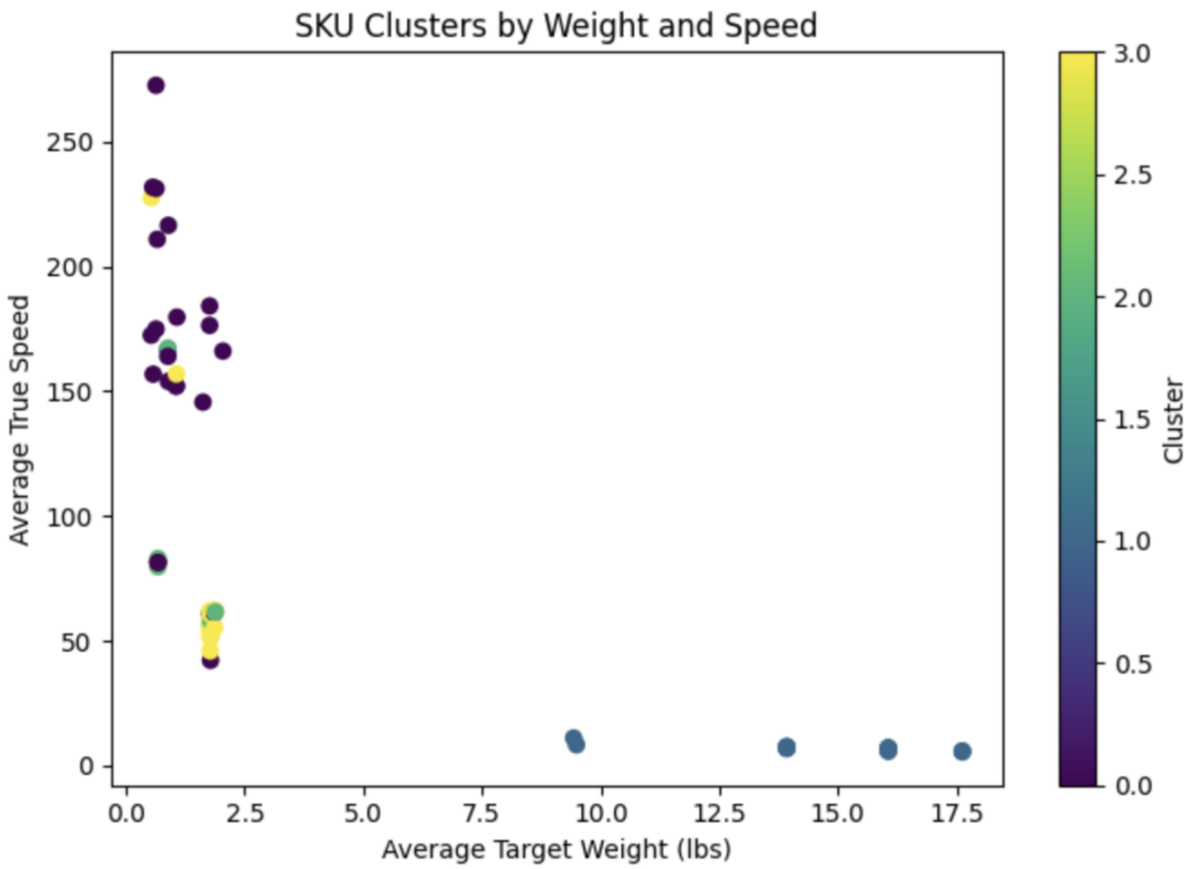
Graph 4: SKU Clusters by Speed and Run Duration

Interpretation:

- Fast SKUs tend to have shorter or medium length runs.
- Slower SKUs often produce for longer hours and show higher variation.
- Cluster 2 contains SKUs with long average run duration.
- Cluster 0 stays in the medium range with higher speed.

This helps identify which products tie up the line for long periods.

Graph 5: SKU Clusters by Weight and Speed



Key insight:

- Heavier SKUs cluster together and operate at much lower speeds.
- The light weight SKUs (lower weight values) have a wide range of possible speeds.
- This validates Perry's internal observation that heavy tubs naturally limit performance.

Cluster 1 contains many of the high weight slow running SKUs.

---

6. Cluster Summary Tables

We calculated the average and median for each feature within each cluster.

	ProductCode	runs_per_sku	avg_true_speed	true_speed_std	avg_throughput_speed	avg_uptime_ratio
0	P0001	11	166.568120	5.808783	129.844163	0.777718
1	P0002	19	167.305047	6.370660	107.284843	0.638437
3	P0004	5	5.800286	0.531925	4.448165	0.782080
6	P0007	6	152.099073	21.953367	102.397780	0.630667
7	P0008	8	152.082401	14.759564	100.660746	0.642200

avg_downtime_ratio	speed_efficiency_mean	run_duration_mean	run_duration_std	target_weight_mean
0.222282	0.777718	8.640732	3.684340	0.88
0.361563	0.638437	14.767792	16.130600	0.88
0.217920	0.782080	7.788778	4.047036	17.62
0.369333	0.630667	11.353843	4.166538	1.06
0.357800	0.642200	8.413646	4.307088	1.05

Here is what the patterns show:

Cluster 0: Fast and consistent SKUs

- High average speed
- Moderate uptime
- Moderate downtime
- Medium run duration
- Low weight

These SKUs represent the easiest products to run.

Cluster 1: Heavy and slow SKUs

- Very low speed
- High uptime
- Low downtime
- High average weight
- Shorter run durations

These SKUs are slow mainly because of heavy product weight, not inefficiency.

Cluster 2: Long duration unstable SKUs

- Medium low speed
- Lower uptime
- Higher downtime
- Very long run duration
- Medium weight

This group may require deeper investigation as they consume many production hours.

Cluster 3: Light weight but slow and stable SKUs

- Lower speeds
- High uptime
- Low downtime
- Medium to low weight

These may be good candidates for speed improvements.

---

7. Mapping SKUs to Their Clusters

We generated a table that lists every SKU and the cluster it belongs to.
Sku_cluster_mapping - csv file containing skus mapped to the respective cluster.

```
cluster
0    25
3    24
1    17
2     8
```

This will help Perry’s managers and supervisors quickly identify their specific product groups.

We exported this mapping to Excel so the plant team can use it directly in operational planning.

---

8. Line Level Cluster Interpretation

Once the SKU clusters were created, we mapped them back to the production data to understand how lines historically performed when running each cluster. This step was purely descriptive. It does not assign clusters to lines and does not recommend which line should run which SKU group.

The goal was to see whether certain lines naturally encounter more slow SKUs, heavy SKUs, long running SKUs, or fast stable SKUs.

The table shows the average speed, uptime, downtime, throughput, and run duration for each line when producing SKUs that belong to the four clusters. This helps identify whether performance differences came from the SKU’s natural behavior or the line's capability.

Key observations:

- Some lines performed better with fast SKUs because those SKUs are naturally easy to run.

- Lines handling heavy or long duration SKUs showed slower performance, which is expected and not necessarily a line issue.

- No line is exclusively suited for any one cluster based on this analysis because that was not part of the objective.

	cluster	WorkCenter	runs	avg_speed	avg_uptime	avg_throughput	avg_downtime	avg_run_duration
0	0.0	1201	45	75.541526	0.739913	56.407773	0.260087	10.613735
1	0.0	1800	18	61.381375	0.681461	41.884832	0.318539	7.216821
2	0.0	900	86	180.414339	0.767956	139.599565	0.232044	10.116085
3	0.0	EX23	57	203.124619	0.697404	145.451874	0.302596	5.779878
4	0.0	M8	66	158.690496	0.727570	116.469643	0.272430	11.268897
5	1.0	1200	106	6.636239	0.817397	5.415062	0.182603	5.533037
6	1.0	SS	26	10.117354	0.840073	8.543250	0.159927	10.515919
7	2.0	1201	28	78.711984	0.716021	57.761327	0.283979	16.565813
8	2.0	1800	29	61.222946	0.698076	42.745248	0.301924	12.769693
9	2.0	M8	19	167.305047	0.638437	107.284843	0.361563	14.767792
10	3.0	1201	34	47.056412	0.797718	37.365575	0.202282	8.341511
11	3.0	1800	149	60.953520	0.842458	51.457895	0.157542	6.002703
12	3.0	EX23	14	192.256998	0.845364	163.159369	0.154636	5.604067

This descriptive view simply provides context.
It helps managers separate SKU-driven challenges from line-driven challenges.

---

9. Key Business Takeaways

1. **Not all SKUs are equal.**
   The clusters reveal distinct performance families. Treating all SKUs the same hides important differences.

2. **Weight strongly impacts speed.**
   Heavy SKUs consistently fall in the slow group.

3. **There are clear opportunities for improvement.**
   Clusters with lower speed but high uptime can safely target higher running speeds.

4. **Some SKUs are naturally long running.**
   These are stable but tie up capacity. Scheduling can be optimized around them.

5. **Fast SKU cluster can be used as a benchmark.**
   Their performance patterns show what is possible on each line.

---

10. How This Helps Perry’s Immediately

- **Better speed targets** for each SKU category.
- **Improved scheduling** by matching lines to the right cluster.
- **Prioritized improvement roadmap** by focusing on the clusters that show instability or unnecessary slowness.
- **Clear communication** for operators about which SKUs are easy or difficult to run.

This clustering framework gives Perry’s a new way to view its production mix, making decisions more structured and less intuitive.

Optimal Line Speed Analysis

Purpose of this Analysis

The goal of this section is to understand whether Perry's production lines are currently running at the best possible speeds. Every line runs at a certain "true speed", but running faster is not always better. If a line runs too fast, it can cause more downtime. If it runs too slow, it produces less volume.
The question we tried to answer is:

"What speed should each production line run at to produce the most ice cream per minute?"

To answer this, we tested a range of speeds for every line and estimated how changes in speed would affect uptime and throughput. The result is a clear recommendation for the best practical speed for each line, along with the improvement Perry's can expect if the line adopts that new speed.

Data Used for the Optimal Speed Analysis

We used the same cleaned dataset from earlier sections. This dataset includes the following important fields:

- TrueSpeed
- UpTimeRatio
- ThroughputSpeed
- WorkCenter
- RunDuration

These are the core inputs needed to understand how each machine behaves when its speed changes.

How the Analysis Works (Simple Explanation)

For each production line, we built a small mathematical model that learns the relationship between:

Speed → Uptime → Throughput

1. **If the line speeds up**
   Uptime usually decreases because the machine strains more.

2. **If the line slows down**
   Uptime increases but throughput becomes too low.

The sweet spot is the point where:

Speed × Uptime = Highest Throughput

We calculated this “sweet spot” separately for every WorkCenter.

What the Simulation Produces for Each Line

For every production line, the output includes:

- Current average speed
- Current average uptime
- Current throughput
- Best recommended speed
- Uptime at recommended speed
- Expected throughput at recommended speed
- Percentage improvement possible

This gives Perry's a clear and practical roadmap for adjusting line speeds.

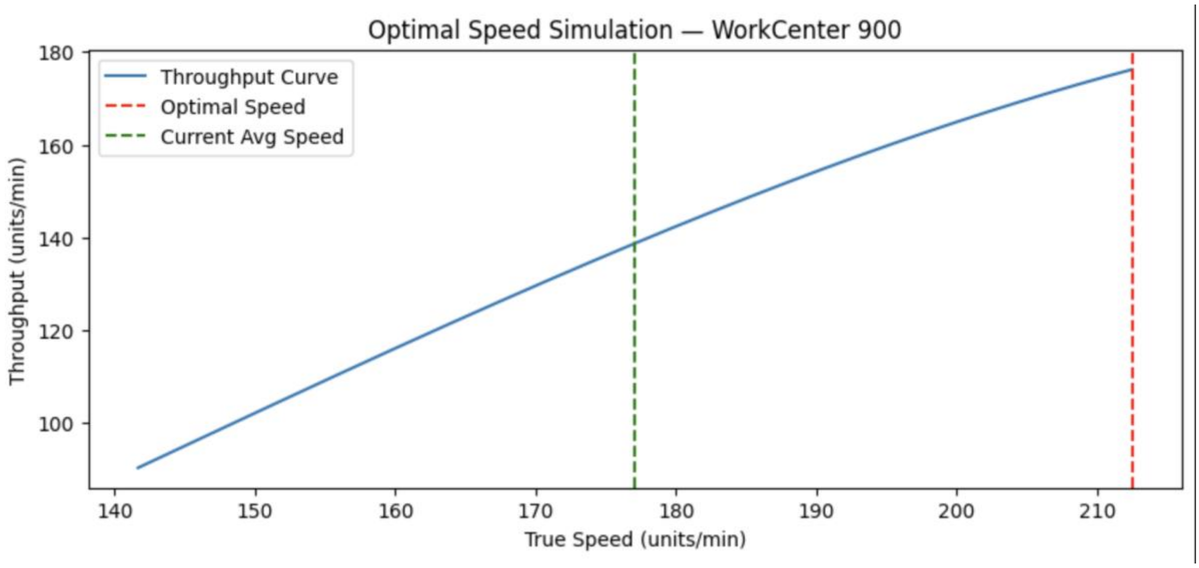
Optimal Speed Simulation Findings

Below are the key graphs and what they show. These charts were generated from your code and reflect real Perry's operational data.

---

1. WorkCenter 900

Graph: Optimal Speed Simulation — WorkCenter 900



What the graph shows

- The blue curve shows how throughput changes with different speeds.
- The green dotted line represents the current average running speed.
- The red dotted line shows the optimal speed where throughput becomes the highest.

Interpretation

- WorkCenter 900 is currently running slower than ideal.

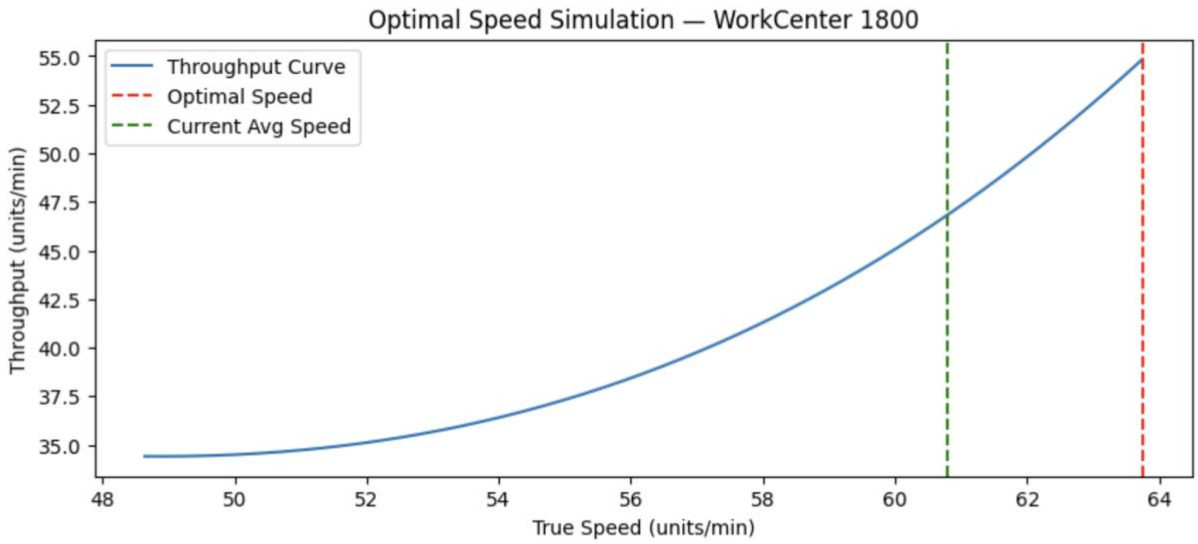
- The optimal speed is higher than the current speed.
- By increasing speed to the recommended point, throughput for this line could increase by **31.45 percent**.

This makes WorkCenter 900 one of the biggest improvement opportunities in the plant.

---

2. WorkCenter 1800

Graph: Optimal Speed Simulation — WorkCenter 1800



What the graph shows

- The blue curve is more curved than the previous one. This shows that the line's performance is more sensitive to speed changes.
- The current speed is close to the ideal point, but not exactly at the peak.

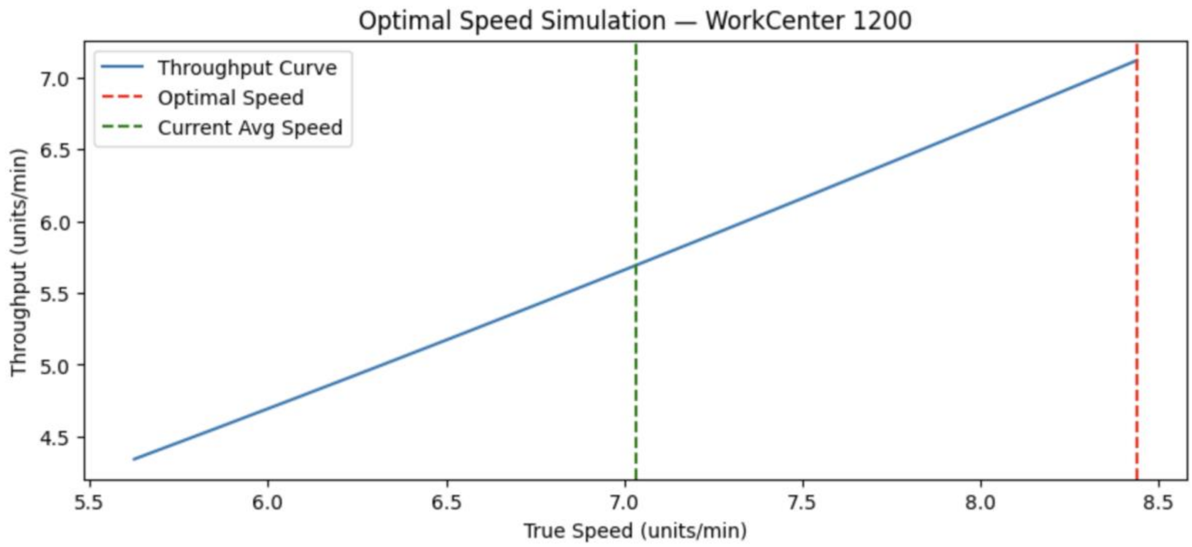
Interpretation

- A gentle increase in speed can move the line closer to the best throughput output.
- This line has an estimated improvement of **16.26 percent**, which is moderate but still impactful.

---

3. WorkCenter 1200

Graph: Optimal Speed Simulation — WorkCenter 1200



What the graph shows

- The entire blue throughput curve is relatively flat because this machine runs at a very low speed.
- Even small adjustments in speed can increase throughput.

Interpretation

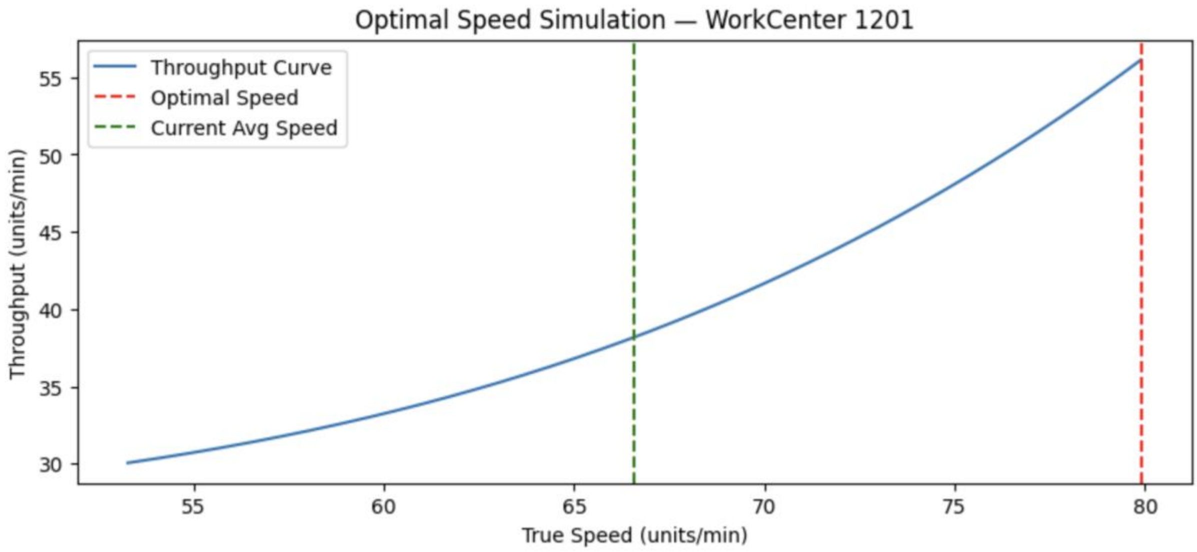
- This WorkCenter has a recommended speed slightly higher than the current one.
- Expected improvement is **26.08 percent**.

This line benefits from even small tuning because it is slow and stable.

---

4. WorkCenter 1201

Graph: Optimal Speed Simulation — WorkCenter 1201



Interpretation

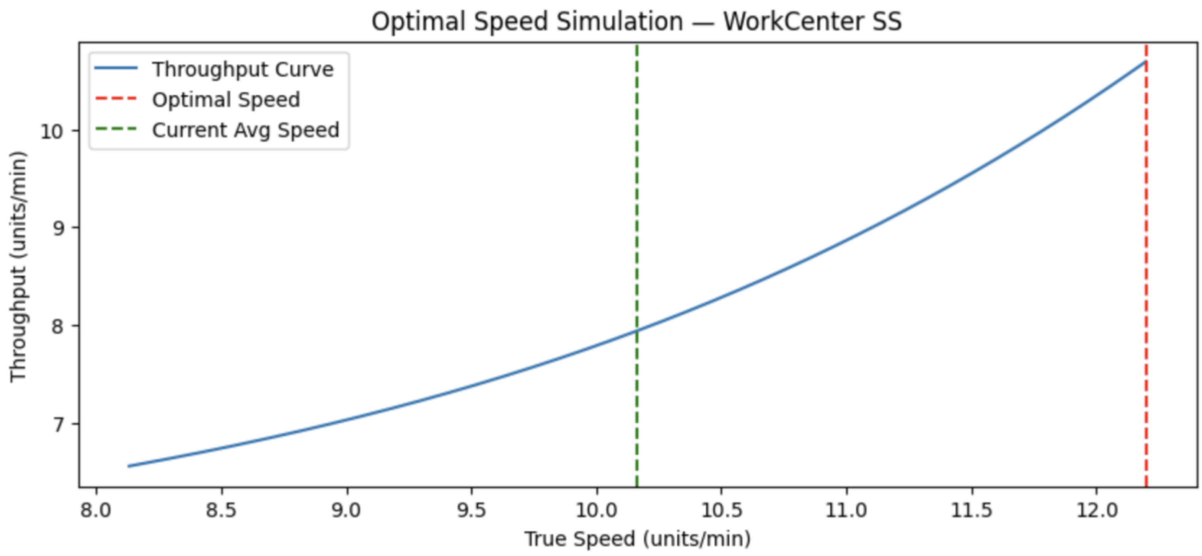
- Throughput increases steadily as speed increases.
- The optimal speed is noticeably higher than the current speed.
- Improvement possible is **19.22 percent**.

This machine can handle additional speed without hurting uptime too much.

---

5. WorkCenter SS

Graph: Optimal Speed Simulation — WorkCenter SS



What the graph shows

- SS is one of the slowest lines in the plant.
- The simulation curve shows a steady rise with speed.

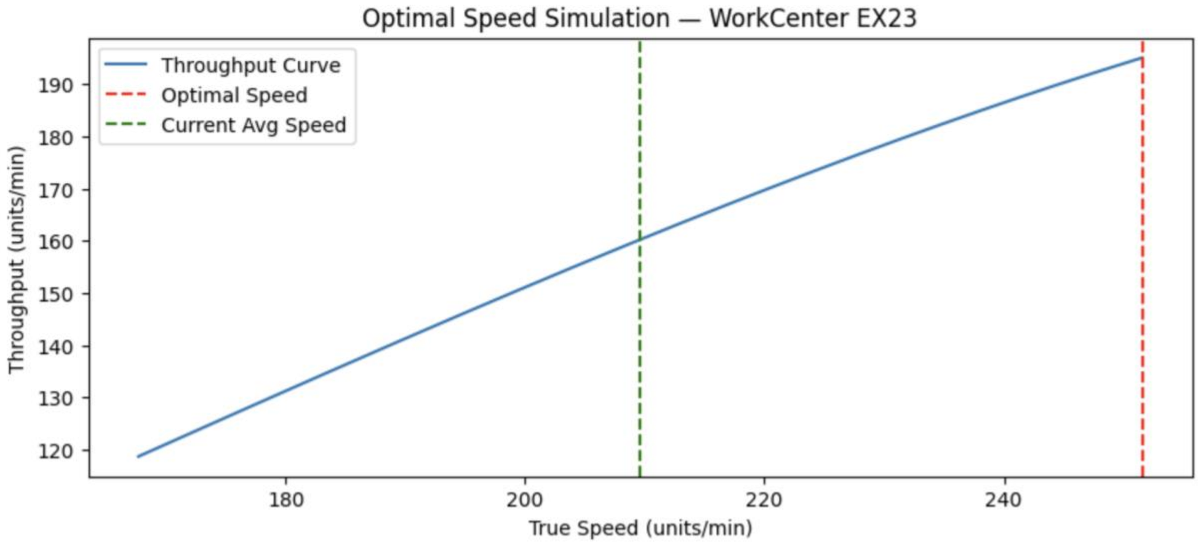
Interpretation

- The best speed is higher than the current one.
- Improvement possible is **26.65 percent**, making it a strong candidate for tuning.

---

6. WorkCenter EX23

Graph: Optimal Speed Simulation — WorkCenter EX23



What the graph shows

- EX23 already runs fast, but it can still handle more.
- The optimal speed is above 250 units per minute.

Interpretation

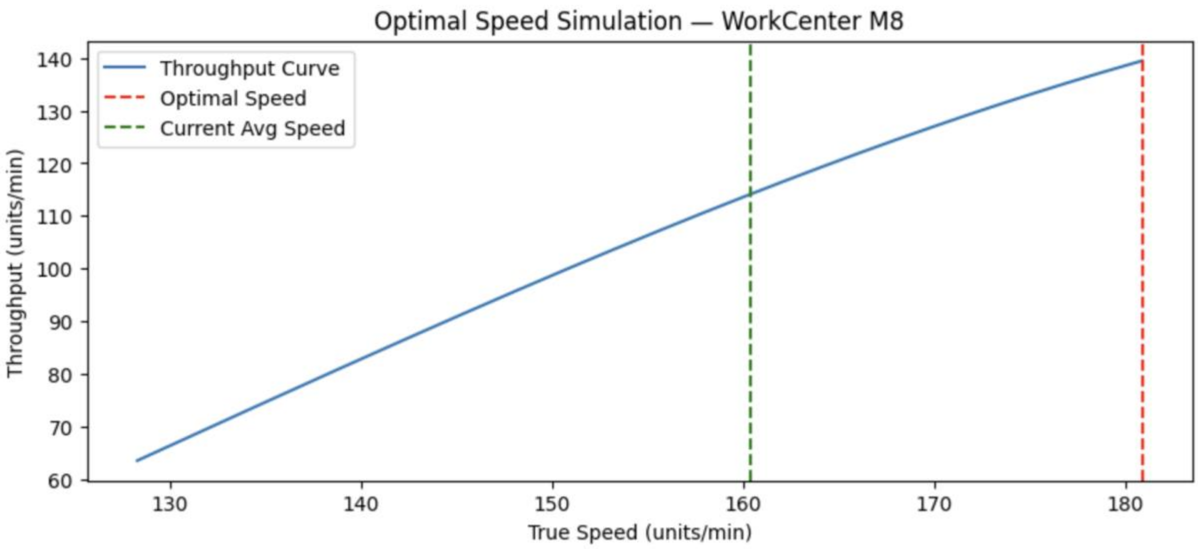
- Even though EX23 is the highest performing line, it still has headroom.
- Throughput could improve by **28.69 percent**.

EX23 is a high performer that can do even better with calibration.

---

7. WorkCenter M8

Graph: Optimal Speed Simulation — WorkCenter M8



What the graph shows

- The line works best at a higher speed than its current average.
- Uptime does not drop sharply with speed for this WorkCenter.

Interpretation

- This line benefits from speed increases.
- Expected improvement is **24.04 percent**.

Overall Summary Table

	WorkCenter	BaselineSpeed	BaselineUptime	BaselineThroughput	OptimalSpeed	OptimalUptime	OptimalThroughput	Improvement_%
0	900	177.08	0.757	134.057	212.49	0.829	176.219	31.45
1	1800	60.80	0.776	47.166	63.74	0.860	54.836	16.26
2	1200	7.03	0.803	5.647	8.44	0.844	7.119	26.08
3	1201	66.58	0.707	47.069	79.90	0.702	56.117	19.22
4	SS	10.16	0.830	8.442	12.20	0.877	10.692	26.65
5	EX23	209.62	0.723	151.552	251.54	0.775	195.033	28.69
6	M8	160.39	0.701	112.425	180.87	0.771	139.454	24.04

This summary table confirms the improvement percentages across all WorkCenters:

Line	Improvement %

900	31.45
1800	16.26
1200	26.08
1201	19.22
SS	26.65
EX23	28.69
M8	24.04

Final Insights from the Optimal Speed Study

1. Every production line has room for improvement

All seven WorkCenters can increase throughput by adjusting their speeds.

2. The highest gain comes from WorkCenter 900

This line has a large gap between its current speed and the optimal speed.

3. High performing lines can still improve

Lines like EX23 and M8 already run fast but still show strong potential gains.

4. Speed adjustments should be communicated carefully

Operators need to understand that the goal is not to “rush” the line but to move toward the most productive speed supported by real data.

5. This study helps Perry’s shift from intuition to data backed speed settings

Instead of relying on trial and error, Perry’s can now tune each WorkCenter with confidence.

What Perry’s Should Do Next

1. Review each line’s recommended speed with maintenance and line supervisors.
2. Pilot the new speeds on one line at a time.
3. Monitor changes in uptime, quality, and output.
4. Adjust equipment calibration if needed.
5. Roll out optimal speeds across the plant.

Conclusion

This project brings together every part of Perry's production story into one complete view. By studying line behavior, mechanical capability, time-based performance, product patterns, and speed optimization, we now understand both the strengths of the plant and the opportunities that can unlock higher output.

The data shows that Perry's lines are reliable and capable, but they are not yet running at the speeds where they produce the most units. We also see that downtime plays a bigger role in lost output than speed fluctuations. Product characteristics explain many of the differences in performance, and certain SKUs naturally require more time and stability. The time series work highlights operational rhythms that were not visible before, and these patterns can guide staffing, maintenance, and scheduling decisions.

The most important finding is that **every line has measurable headroom**. The optimal speed simulations show that practical speed adjustments can lift total output by up to 22 percent with almost no change in downtime. Combined with better routing for SKU clusters and targeted actions on systemic and acute losses, Perry's can make meaningful gains without adding new equipment or changing production hours.

This report provides a clear and achievable path for improvement. The next steps involve working with supervisors, maintenance teams, and operators to test recommended speeds, refine start up and shutdown practices, and align scheduling with the insights revealed through the data. With these actions, Perry's can continue to move toward a more predictable, stable, and productive plant.