DATA DRIVEN PRODUCTIVITY OPTIMIZATION IN THE APPAREL INDUSTRY

TABLE OF CONTENTS

INTRODUCTION

The ‘Apparel Industry’ is a labour-intensive and highly competitive sector that significantly contributes to the economies of many developing countries. In this environment, optimizing productivity is essential for maintaining profitability and sustaining growth. With increasing pressure from global brands to deliver high-quality products within tight deadlines and low margins, manufacturers must ensure maximum efficiency in resource utilization. Higher productivity enables firms to reduce operational costs, meet client demands faster, and improve overall output without compromising quality. Moreover, by minimizing delays and enhancing workflow efficiency, companies can gain a competitive edge in both domestic and international markets.

Beyond operational efficiency, productivity optimization also supports broader organizational goals, including employee well-being and environmental sustainability. A well-structured productivity system with clear targets and fair incentives can improve worker motivation, reduce idle time, and create a more stable work environment. Simultaneously, efficient production processes can reduce material waste and energy consumption, helping companies align with sustainability objectives. Thus, productivity optimization in the garment industry is not just about doing more—it's about working smarter, creating value for businesses, workers, and the environment alike.

Objectives of the Project:

1. *Exploratory Data Analysis*: Perform informative visualizations that effectively communicate the relationships between different variables and their impact on productivity.
2. *Key Factor Analysis*: Identify and analyse the primary variables influencing productivity. Determine which factors have the most significant impact on overall efficiency.
3. *Predictive Modelling*: Develop a robust predictive model to forecast actual productivity. This model should be adaptable to various scenarios and provide insights into future productivity trends. *(Note: The target variable for the predictive model is "actual_productivity".)*
4. *Business Recommendation*: Formulate data-driven strategies to improve overall productivity while maintaining a balance with worker satisfaction and well-being. Also, propose methods to increase productivity while considering environmental impact.

DATASET OVERVIEW

Dataset Link: Productivity Prediction of Garment Employees

The dataset used for this project is titled "Garment Workers Productivity" and contains real-world operational data collected from a garment manufacturing company. It comprises a total of 1197 records with 15 variables, capturing various aspects of production-related activities, workforce details, and performance metrics. The primary goal of the dataset is to understand and predict the actual productivity of garment workers across different teams and departments.

The dataset includes a mix of numerical and categorical variables. Numerical features provide quantitative insight into performance and operational metrics while Categorical variables describe contextual and organizational factors. The key variable of interest—actual_productivity—represents the measured output of a team relative to their target and serves as the target variable for predictive modelling. Overall, the dataset is rich in features that allow for comprehensive exploratory, diagnostic, and predictive analysis.

The dataset includes the following variables:

1. date: Production date (Format: MM-DD-YYYY)
2. day: Day of the week
3. quarter: Portion of the month (month divided into four quarters)
4. department: Specific department associated with the data point
5. team_no: Team number associated with the data point
6. no_of_workers: Number of workers in each team
7. no_of_style_change: Number of changes in the style of a particular product
8. targeted_productivity: Productivity goal set by management for each team daily (Range: 0-1)
9. smv: Standard Minute Value- allocated time for a task
10. wip: Work in progress- number of unfinished items
11. over_time: Amount of overtime worked by each team (in minutes)
12. incentive: Financial incentive provided (in BDT- Bangladeshi Taka)
13. idle_time: Time when production was interrupted (in minutes)
14. idle_men: Number of workers who were idle due to production interruption
15. Actual_productivity (target): The actual productivity delivered by the workers (Range: 0-1)

SOURCE CODE OF THE PROJECT: GOOGLE COLAB NOTEBOOK

DATA PROCESSING

The raw dataset was systematically processed to prepare it for analysis and modelling. The data processing involved the following key steps:

→ *Data Loading and Initial Exploration*

The dataset, '*garment_worker_productivity.csv*', was imported using Python's pandas library. An initial check revealed 1,197 rows and 15 columns, including features such as date, department, team, productivity metrics, and labor-related variables. The .head() and .info() functions were used to inspect data types, missing values, and overall structure.

→ *Data Cleaning and Validation*

To ensure consistency:

- Whitespace was stripped from categorical fields (e.g., 'department'), and null entries were dropped.
- The dataset was filtered to retain only those rows where targeted_productivity and actual_productivity were between 0 and 1.
- Missing values in the 'wip' column were imputed using the median, accounting for the likelihood of a skewed distribution.
- A copy of the cleaned dataset was created to preserve the original data and continue processing on the new dataframe df_new.

→ *Data Transformation*

Transformations were applied to prepare the data for analysis:

- The 'date' column was converted to datetime format.
- A new 'month' column was extracted from the date for time-series analysis.

→ *Encoding Categorical Variables*

To make categorical variables suitable for machine learning models, label encoding was applied to the following columns – 'department' (0 for sewing, 1 for finishing), 'quarter' (0 to 5), 'day' (0 for Monday to 5 for Saturday)

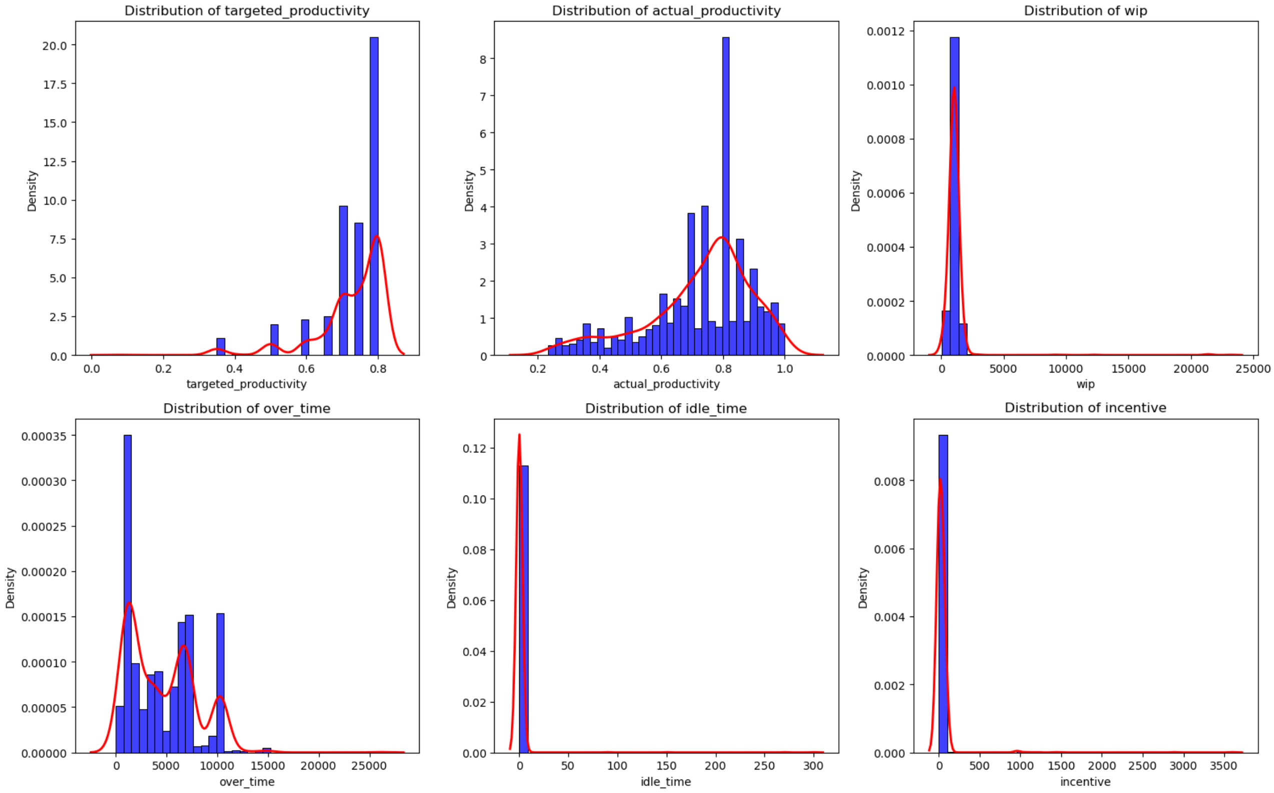
This conversion ensured that all features were in numerical form and ready for further analytical tasks. *(Below: Data info before and after processing)*

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1197 entries, 0 to 1196
Data columns (total 15 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   date                   1197 non-null   object
 1   quarter                1197 non-null   object
 2   department             1197 non-null   object
 3   day                    1197 non-null   object
 4   team                   1197 non-null   int64
 5   targeted_productivity  1197 non-null   float64
 6   smv                    1197 non-null   float64
 7   wip                    691 non-null    float64
 8   over_time              1197 non-null   int64
 9   incentive              1197 non-null   int64
 10  idle_time              1197 non-null   float64
 11  idle_men               1197 non-null   int64
 12  no_of_style_change     1197 non-null   int64
 13  no_of_workers          1197 non-null   float64
 14  actual_productivity    1197 non-null   float64
dtypes: float64(6), int64(5), object(4)
memory usage: 140.4+ KB
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1160 entries, 0 to 1196
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   date                   1160 non-null   datetime64[ns]
 1   quarter                1160 non-null   int64
 2   department             1160 non-null   int64
 3   day                    1160 non-null   int64
 4   team                   1160 non-null   int64
 5   targeted_productivity  1160 non-null   float64
 6   smv                    1160 non-null   float64
 7   wip                    1160 non-null   float64
 8   over_time              1160 non-null   int64
 9   incentive              1160 non-null   int64
 10  idle_time              1160 non-null   float64
 11  idle_men               1160 non-null   int64
 12  no_of_style_change     1160 non-null   int64
 13  no_of_workers          1160 non-null   float64
 14  actual_productivity    1160 non-null   float64
 15  month                  1160 non-null   int32
dtypes: datetime64[ns](1), float64(6), int32(1), int64(8)
memory usage: 149.5 KB
```

EXPLORATORY DATA ANALYSIS

1. Distribution of Key Variables



This visualization presents the distribution of six key variables using histograms overlaid with kernel density estimation (KDE) curves. Here's the interpretation of each plot:

→ *Targeted Productivity*

- Observation: The distribution is right-skewed with a concentration of values near 0.8.
- Interpretation: Most teams have high targeted productivity levels, indicating management expectations are consistently set high. There are few instances with low targets.

→ *Actual Productivity*

- Observation: The distribution appears approximately normal but slightly left-skewed. Most values cluster between 0.6 and 0.9.
- Interpretation: Actual productivity is generally high but often falls short of the targets. Some outliers or lower-performing teams are visible on the left tail.

→ *Work In Progress (WIP)*

- Observation: Highly right-skewed with most values near zero and a long tail stretching to 25,000.
- Interpretation: Most departments maintain low levels of WIP, but there are occasional large backlogs, possibly due to process bottlenecks or style changes.

→ *Over Time*

- Observation: Multimodal and right-skewed distribution, with peaks around 0, 5,000, and 10,000 minutes.
- Interpretation: Some teams don't work overtime at all, while others work in batches of significant extra hours, possibly reflecting deadline-based work or uneven workload distribution.

→ *Idle Time*

- Observation: Very right-skewed with a sharp peak near 0.
- Interpretation: Most teams have minimal idle time, suggesting efficient operations. However, a few instances of very high idle times indicate occasional severe disruptions.

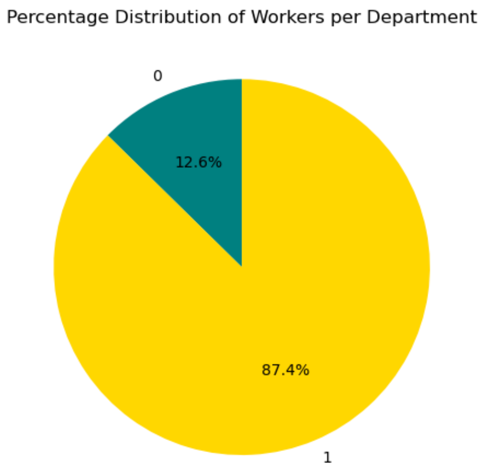
→ *Incentive*

- Observation: Right-skewed distribution with a dense concentration near 0 and a few instances of high incentive payouts.
- Interpretation: Most workers receive little to moderate incentive, with a few high outliers likely linked to exceptional performance or high over time.

Insights:

- The distributions suggest a production environment with high expectations but variable actual performance.
- Operational disruptions (idle time) and compensation (incentives) are generally minimal but show sporadic spikes.
- The dataset contains potential outliers and skewness, which should be addressed in modelling (e.g., normalization or transformation may be required).
- A gap between targeted and actual productivity is visually evident, hinting at potential process inefficiencies or over-ambitious targets.

2. Distribution of Workers per Department



This pie chart shows the percentage distribution of workers across two departments: Sewing (labelled as 0) and Finishing (labelled as 1).

Interpretation:

→ *Finishing Department (1)*:

- Accounts for 87.4% of the total workforce.
- This dominant share suggests that most of the production manpower is concentrated in the finishing operations, likely due to the labor-intensive nature of final assembly, quality checks, and packaging.

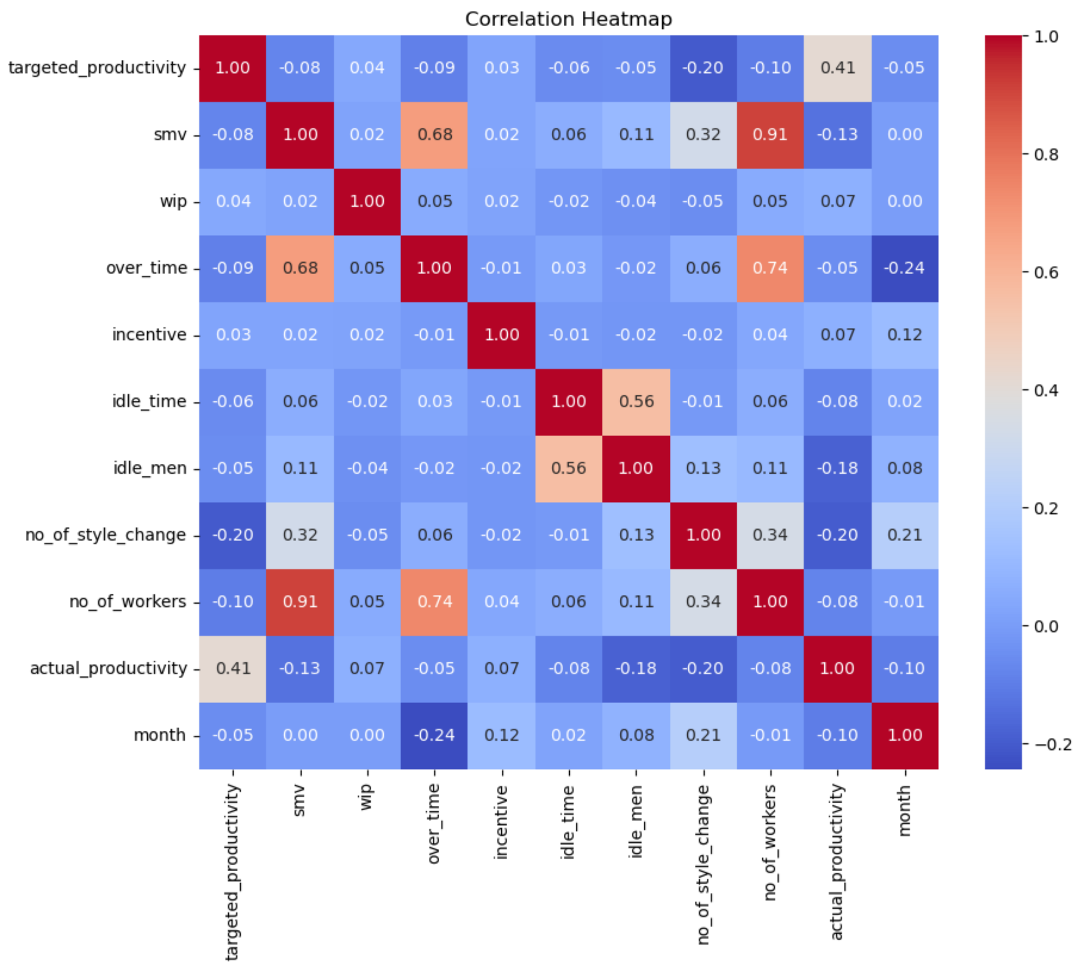
→ *Sewing Department (0)*:

- Comprises only 12.6% of the workforce.
- This relatively small proportion may indicate either high automation or a more specialized role requiring fewer workers in the sewing phase.

Insights:

- The imbalance implies a heavier operational focus or demand in the finishing stage.
- Resource allocation strategies should consider whether this distribution aligns with productivity outputs.
- If performance issues are observed, the finishing department may warrant deeper analysis due to its scale and impact.

3. Correlation between Variables (Heatmap)



Correlation Interpretation

→ *Strong Positive Correlations:*

- no_of_workers & smv (0.91): More workers are associated with higher SMV (Standard Minute Value), possibly due to more complex tasks requiring more manpower.
- no_of_workers & over_time (0.74): Larger teams tend to work more overtime, indicating demand pressure or inefficiencies with bigger teams.
- idle_time & idle_men (0.56): Naturally expected—when idle time increases, more workers remain idle.

→ *Moderate Positive Correlations*:

- actual_productivity & targeted_productivity (0.41): Shows that higher targets are somewhat aligned with higher actual productivity, though not very strongly.
- smv & over_time (0.68): Tasks with higher SMV often require more overtime—possibly due to underestimated time allocation.

→ *Moderate Negative Correlations:*

- no_of_style_change & targeted_productivity (-0.20): Frequent style changes disrupt productivity goals—indicating task switching or inefficiencies.
- no_of_style_change & actual_productivity (-0.20): Style changes are also associated with reduced actual productivity—confirming their negative operational impact.
- over_time & month (-0.24): Overtime reduces as months progress—possibly indicating better process optimization over time.

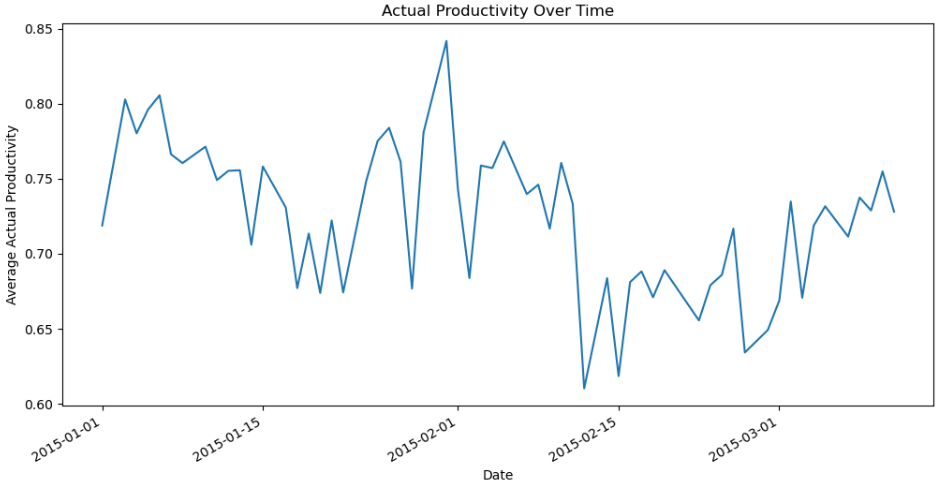
→ *Negligible or Weak Correlations:* Most other variables (e.g., wip, incentive, idle_time) show very low or near-zero correlations with actual_productivity, suggesting limited direct linear influence.

Insights:

- Productivity (actual and targeted) is moderately linked but negatively influenced by style changes.
- Workforce size and task complexity (SMV) are key drivers of overtime.
- Idle time is tightly linked with number of idle workers, as expected.
- There's room to explore non-linear relationships or latent variables, as many direct correlations are weak.

4. Actual Productivity Analysis

a. Time-Based Analysis

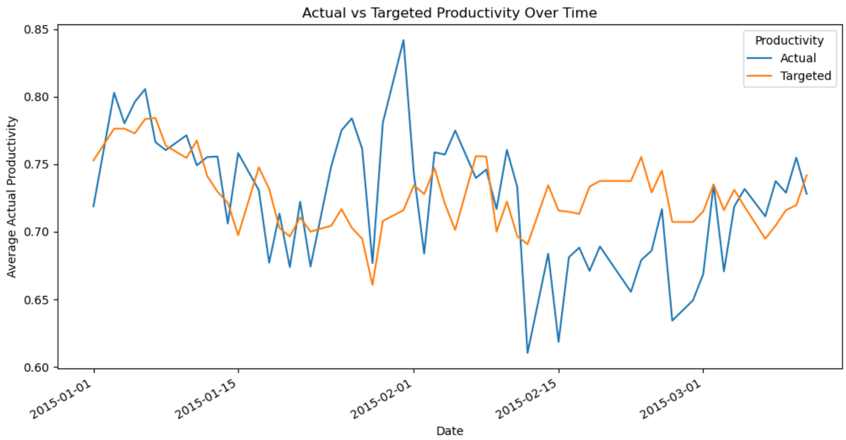


→ *Actual Productivity Over Time:*

- Fluctuating trend: Productivity varies frequently, without a consistent upward or downward trend.
- Highs: Peaks close to 0.84 in early February.
- Lows: Drops to around 0.61 in mid-February, indicating possible disruptions or inefficiencies during that period.
- Observation: The average actual productivity mostly ranges between 0.65 and 0.80, suggesting moderate productivity with potential for improvement.

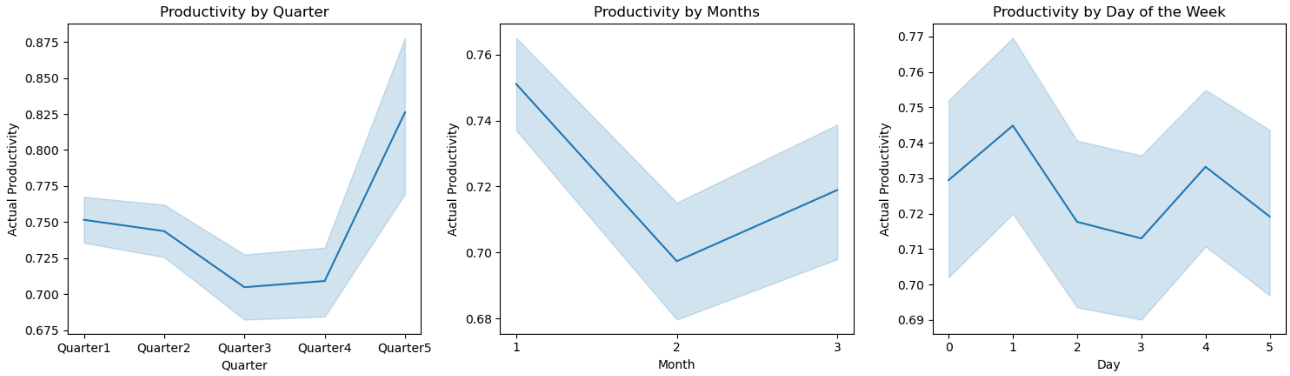
→ *Actual vs. Targeted Productivity Over Time*

- Actual productivity is often below targeted productivity, especially in the second half of February.
- Closer alignment is seen in early January and early February.
- Divergence increases later, indicating either:
  - Overestimation in targets
  - Underperformance due to operational or resource issues



Interpretation:
The gap between actual and targeted productivity suggests there are inefficiencies or constraints in achieving planned goals. Management may need to revisit how targets are set or address bottlenecks affecting actual performance.

b. Temporal Analysis



→ *Productivity by Quarter*

- Quarter 1 to Quarter 4 shows a gradual decline in actual productivity.
- Quarter 5 (likely a partial or extended period) exhibits a sharp increase in productivity, peaking around 0.82–0.87.
- Interpretation: This dip-and-rebound pattern could be due to operational challenges in the mid-period (e.g., style changes, worker fatigue, holidays) and recovery or optimization efforts in Quarter 5.

→ *Productivity by Month*

- Month 1 (January) has the highest average productivity (~0.75).
- Productivity drops in Month 2 (February) to around 0.70, with a wide variability range (suggesting inconsistent performance).
- Month 3 (March) shows a slight rebound in productivity.
- Interpretation: February likely had disruptions — possibly due to higher absenteeism, inefficiencies, or increased idle time. March shows recovery, but not to January levels.

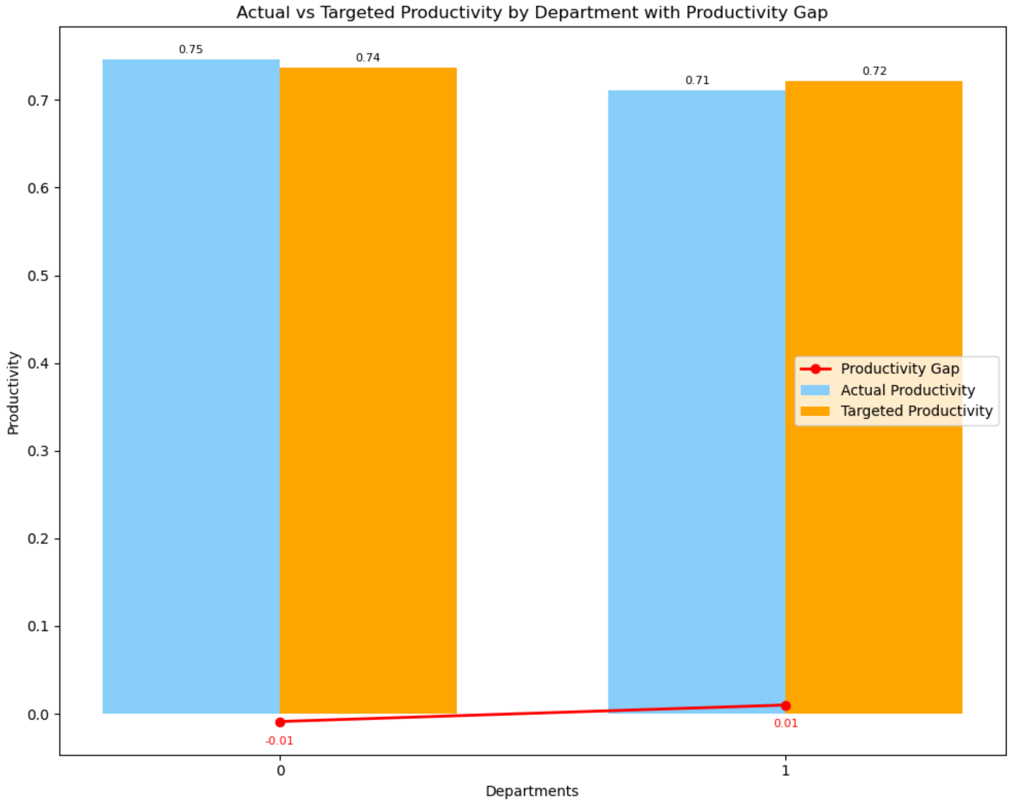
→ *Productivity by Day of the Week (0 = Monday, 5 = Saturday)*

- Highest productivity occurs on Tuesday (Day 1).
- Wednesdays and Thursdays (Days 2 & 3) have the lowest productivity.
- Slight recovery on Fridays and Saturdays (Days 4 & 5).
- Interpretation:
  - Mondays start strong, Tuesday peaks (possibly due to momentum).
  - Mid-week drop might suggest fatigue or reduced motivation.
  - Late-week partial recovery could result from task completion pressures or fewer interruptions.

Insights

- Productivity is time-sensitive: it dips mid-month, mid-week, and mid-quarter.
- Peaks occur at start of periods (Month 1, early week) and end periods (Quarter 5) — suggesting performance cycles.

c. ***Productivity Analysis by Department***



→ *Department: Sewing (Dept 0)*

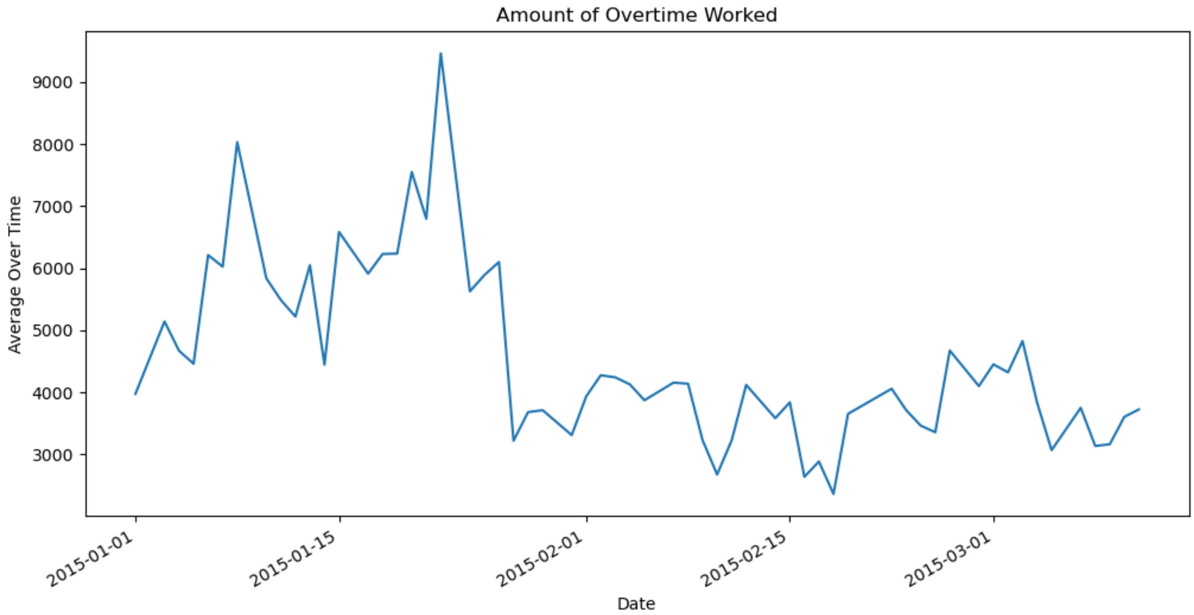
- Insight: The Sewing department slightly outperformed expectations.
- Possible Factors:
  - Efficient workflow or team coordination.
  - Experienced workers or fewer interruptions.
  - Effective utilization of available time.

→ *Department: Finishing (Dept 1)*

- Insight: The Finishing department slightly underperformed relative to the target.
- Possible Factors:
  - Bottlenecks at the end of the production line.
  - More idle time or delays (e.g., waiting for output from Sewing).
  - Quality checks or rework causing slowdowns.

5. Overtime Analysis

a. Time-Based Analysis



→ *Early January to Late January 2015:*

- Overtime levels show a consistent upward trend, peaking at over 9000 units near the end of January.
- This suggests high production demand or possibly staff shortages, requiring more effort to meet deadlines.

→ *Early February Onward:*

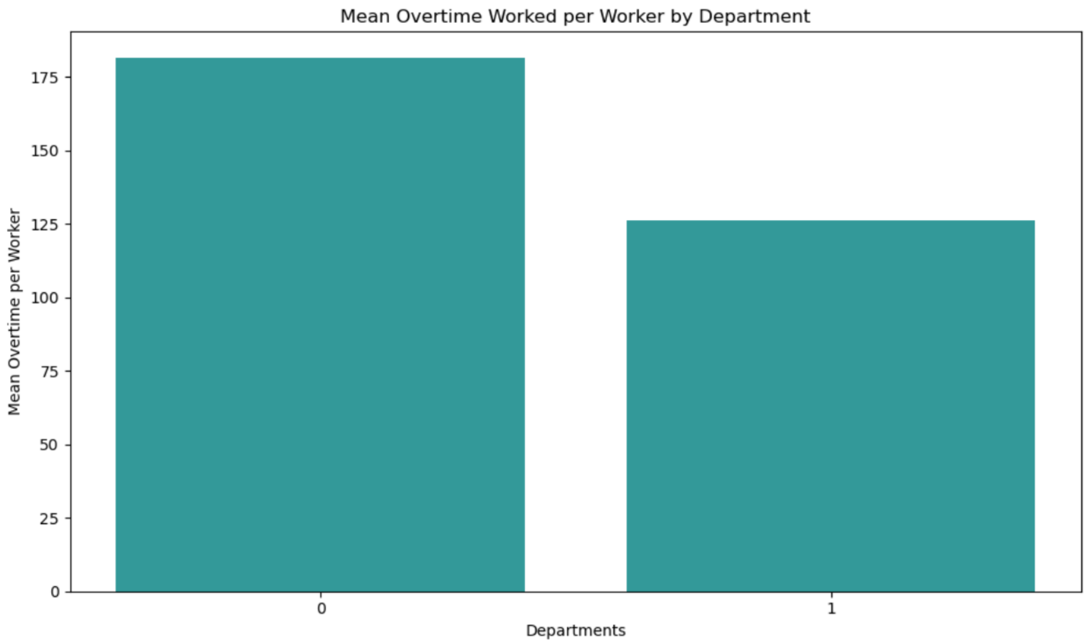
- A sharp drop is seen immediately after the peak — from ~9000 to around 3000–4000.
- From February through March, overtime levels remain consistently lower and stable, typically between 2500–4500.

Insights and Recommendations:

- The dramatic drop in overtime post-January may indicate:
  - Completion of a major order or seasonal rush.
  - Introduction of process efficiencies or workforce expansion.
  - Management intervention to control excessive overtime costs or burnout.
- The high overtime period in January could have contributed to the slightly better productivity observed in the Sewing department *(as discussed earlier).*
- Analyze output vs. overtime: Check if the high overtime correlated with a proportional rise in productivity/output.
- Evaluate impact on employees: High overtime can lead to fatigue, which might explain productivity dips in February, especially in Finishing.

- Sustainability check: Consider whether sustained high productivity can be achieved without relying heavily on overtime.

b. Department Wise Analysis



→ *Sewing Department (0):*

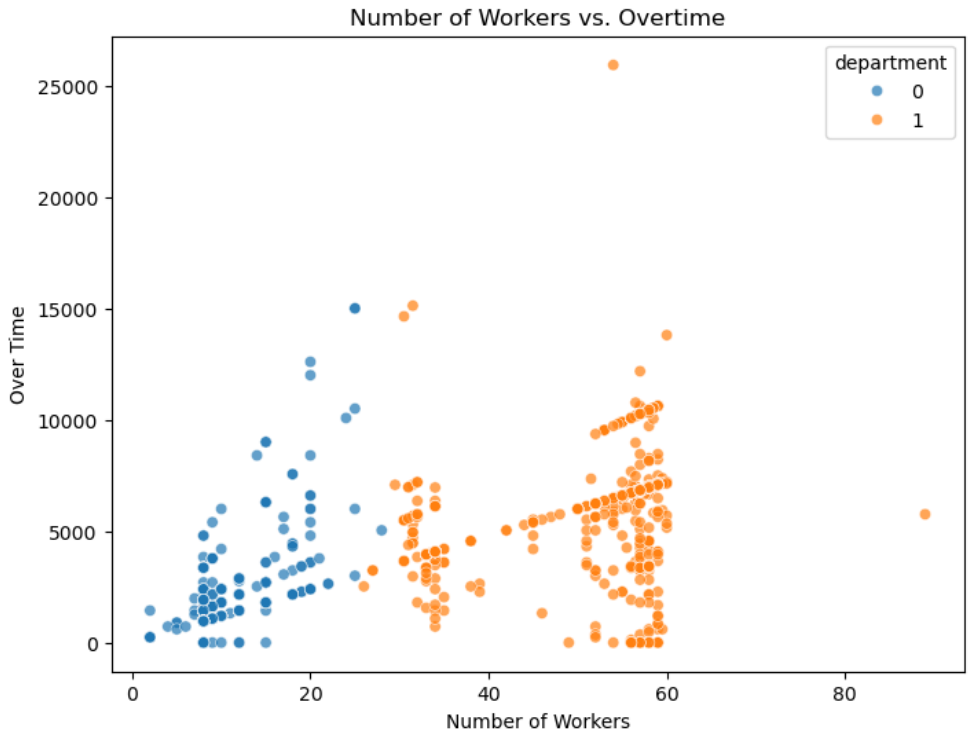
- Mean overtime per worker is ~180 hours.
- This is significantly higher than the Finishing department.

→ *Finishing Department (1):*

- Mean overtime per worker is ~125 hours.
- Workers here logged ~55 fewer hours of overtime on average compared to Sewing.

Insights and Recommendation:

- *Higher Overtime Load in Sewing:*
  - Suggests greater workload, bottlenecks, or labor-intensive operations in sewing.
  - May explain why Sewing slightly exceeded its productivity target earlier (0.75 actual vs. 0.74 target).
- *Lower Overtime in Finishing:*
  - May indicate more efficient workflows or less intense workload.
  - Aligns with the observed positive productivity gap (0.72 actual vs. 0.71 target), possibly achieved more efficiently.
- Evaluate sustainability: High overtime in Sewing might not be viable long-term; explore automation or workforce expansion.
- Benchmark productivity vs. overtime: Compare whether Sewing's higher overtime translates into proportionately higher output/productivity.



→ There is a <u>positive correlation</u> in both departments between the number of workers and overtime — meaning: As the number of workers increases, total overtime also tends to increase.

→ However, *<u>the strength of the correlation varies by department</u>*:

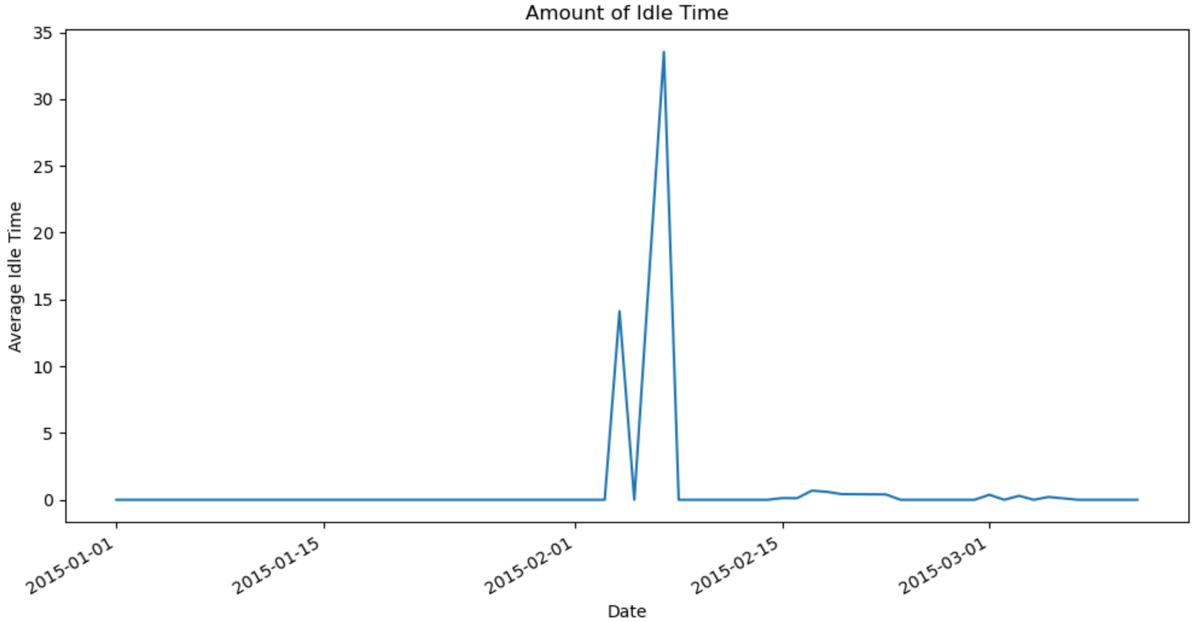
- <u>Sewing (Dept 0)</u>: The correlation is weaker and more scattered. Even small teams sometimes show high overtime, indicating inefficiency or uneven workload.
- <u>Finishing (Dept 1)</u>: Shows a stronger and more structured positive correlation. Larger teams tend to have more total overtime in a consistent way, implying a more systematic allocation of work.

Insights and Recommendations:

- *<u>Sewing Dept (0):</u>*
  - Fewer workers often work disproportionately higher overtime.
  - Suggests potential understaffing, workforce imbalance, or operational inefficiencies.
- *<u>Finishing Dept (1):</u>*
  - More stable scaling of overtime with worker count.
  - Indicates better workforce planning or load balancing.
- Investigate task allocation in Sewing — ensure workload is not concentrated among a few.
- Consider adding staff or optimizing schedules in Sewing to manage overtime spikes.

6. Idle Time Analysis

a. Time-Based Analysis



→ *Near-Zero Idle Time for Most Days:*

- For the majority of the time period (especially early January and post-February), average idle time is consistently zero or near zero.
- This suggests that workers were consistently engaged, with minimal downtime.

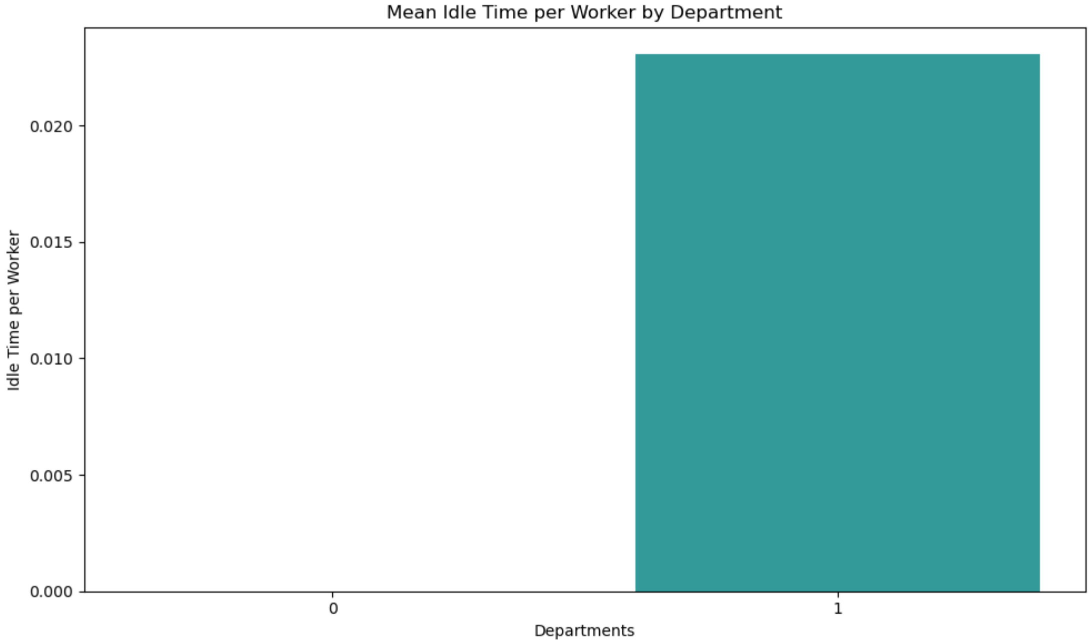
→ *Sharp Spikes in Early February:*

- There are two significant spikes in average idle time:
  - First Spike: Around early February (approx. Feb 5), reaching ~14 minutes.
  - Second Spike: Peaks at ~34 minutes, around Feb 8–9.
- These spikes are sudden and short-lived, with the idle time quickly returning to zero.

→ *Post-Spike Pattern:*

- After the second spike, there is a very slight increase in baseline idle time (still below ~2 minutes), with small fluctuations.
- Could be due to: Machine breakdowns or power outages, Material shortage or logistics delay, Operational bottlenecks (e.g., tasks queued due to upstream delays), Human resource factors like absenteeism or shift mismanagement

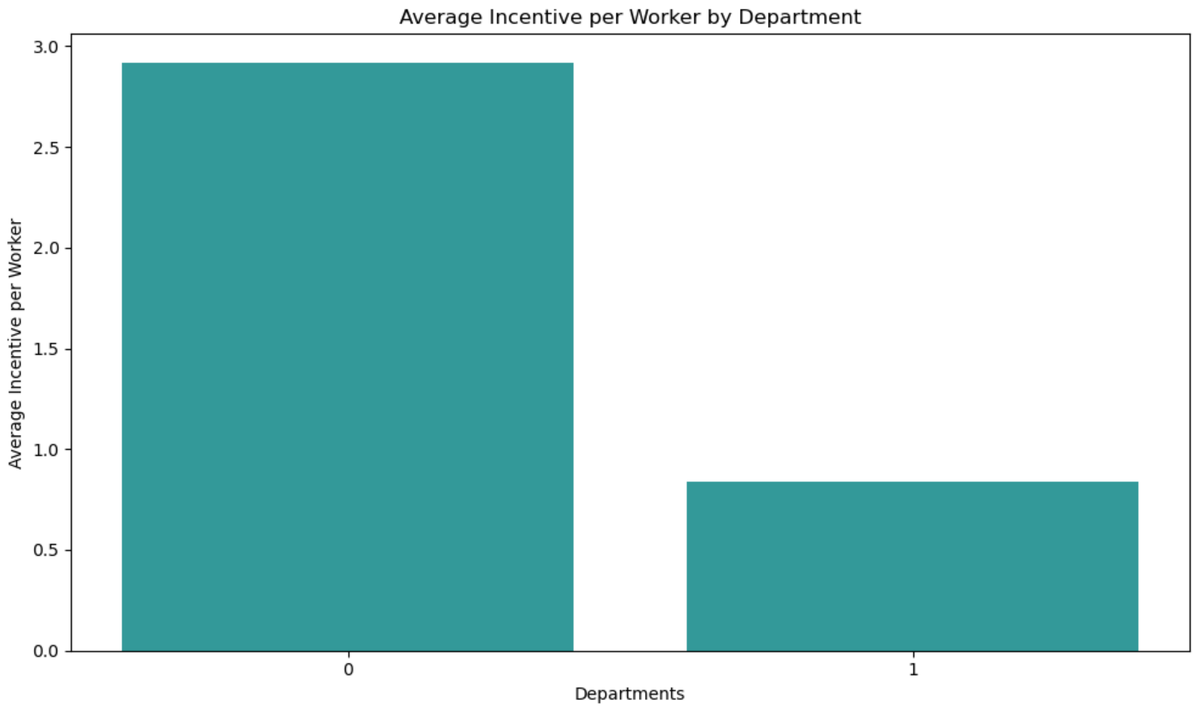
b. Department Wise Analysis



The graph concedes with the initial assessment of the Sewing Department working efficiently and even overtime at times, while the Finishing department might face some inefficiencies in terms of idle time, which is rather minor.

7. Incentive Analysis

a. Department-Wise Analysis



→ *Sewing Department:*

- Has a significantly higher average incentive per worker—close to ₹3 per unit (or arbitrary incentive unit).
- Suggests better performance recognition or more favorable incentive structures.

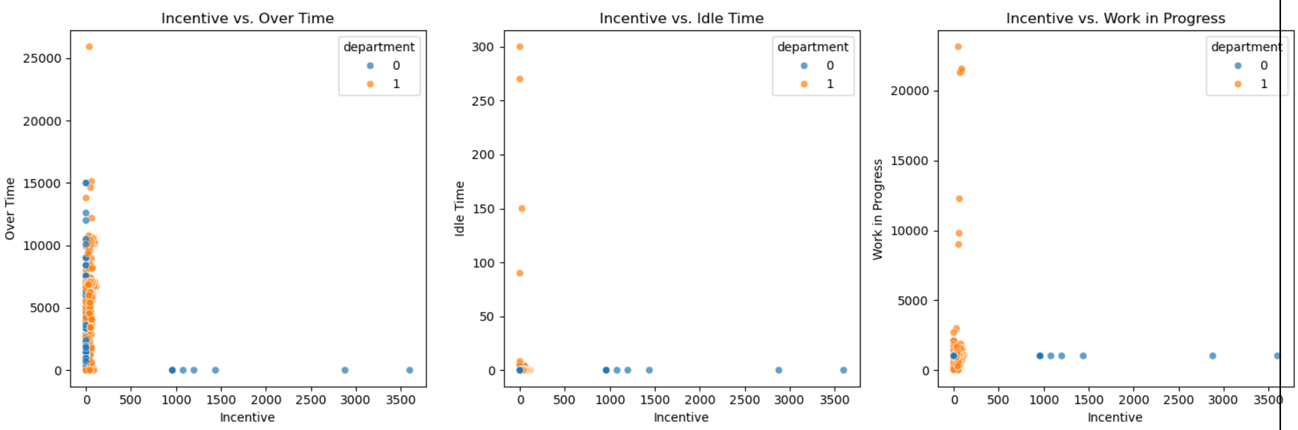
→ *Finishing Department:*

- Receives a much lower average incentive—less than ₹1 per unit.
- This large gap could point to fewer opportunities to earn incentives or less efficient output.

Insights and Recommendations:

- Department 0 (Sewing): Fewer workers, higher overtime per worker, and higher incentives → possibly a more efficient and motivated team, albeit with a heavier workload.
- Department 1 (Finishing): More workers, lower overtime per worker, and lower incentives → workload and rewards may be more evenly distributed, but potentially with lower individual motivation or recognition.
- Balance Incentives: Department 1 may benefit from revisiting its incentive structure to better align with individual contributions.

b. Assessing Correlations



→ *Incentive vs. Over Time*

- Department 1 (orange) has many data points clustered at low incentive levels, even when overtime hours are high (some >25,000).
- Department 0 (blue) shows a few outliers with very high incentives and relatively low overtime.
- There appears to be no strong linear correlation between incentives and overtime for either department.
- However, Department 0 seems to reward more per unit of overtime, implying that their incentive system is not purely based on hours worked.
- Department 1 may have a weaker link between overtime and incentives—indicating either a flat incentive system or one less responsive to overtime.

→ *Incentive vs. Idle Time*

- Most data points for both departments lie close to zero idle time, especially where incentives are higher.
- A few outliers (mostly Finishing Department) show high idle time but low incentives.
- There is a negative correlation (though weak): higher idle time is associated with lower incentives, particularly in Department 1.
- This supports the logic that workers with higher idle time contribute less to performance and thus receive lower incentives.

→ *Incentive vs. Work in Progress*

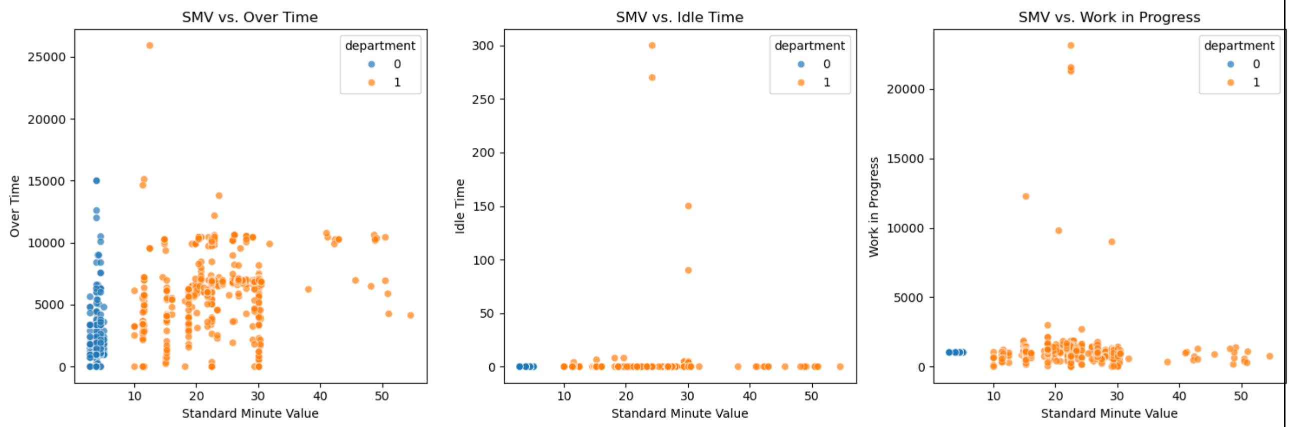
- Again, most incentives are clustered at the lower range, but Department 1 shows many instances of high WIP and low incentives.
- Department 0 again stands out with fewer data points but relatively higher incentive values even when WIP is moderate.
- No clear linear correlation.
- However, Department 0 seems to maintain a more direct incentive–WIP alignment: higher incentives correspond with meaningful WIP levels.
- In contrast, Department 1 has high WIP but still low incentives, suggesting a disconnect—possibly inefficiencies or poor performance measurement.

Insights:

- Department 0 appears to use incentives more strategically—possibly linked to output and efficiency (WIP, low idle time).
- Department 1 has weaker incentive alignment, where high overtime or WIP doesn't consistently lead to higher incentives.
- There is no strong statistical correlation in any of the scatter plots, but the qualitative patterns suggest:
    - Department 0 is more performance-driven.
    - Department 1 may have systemic inefficiencies or a misaligned incentive structure.

8. Analysing Standard Minute Value (SMV) or Task Complexity

a. Assessing Correlations



→ *SMV vs. Over Time*

- Department 0 (Sewing): Data points are clustered at lower SMV values (around 8-15), with overtime values generally below 15,000 minutes. Operates with more standardized task times (narrow SMV range), suggesting more uniform and predictable operations.
- Department 1 (Finishing): Shows a much wider distribution across SMV values (10-55), with overtime typically between 0-10,000 minutes, and occasional outliers reaching 25,000+ minutes. Works with varied task complexities (wide SMV range), indicating more diverse operations that require different time allocations.
- Distribution Pattern: No strong linear correlation is evident, but there's a noticeable clustering pattern by department.
- The lack of clear correlation suggests overtime is driven by factors beyond task complexity alone (possibly deadlines, capacity constraints, or worker availability).

→ *SMV vs. Idle Time*

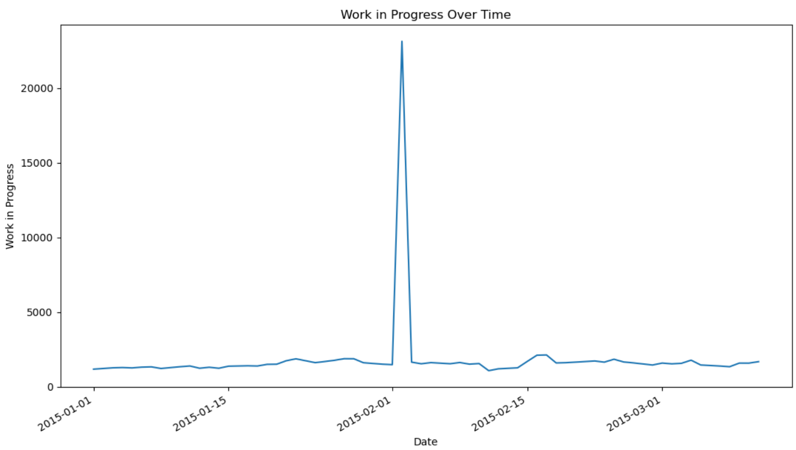
- Both Departments: The vast majority of data points show near-zero idle time across all SMV values.
- Department 1 (Finishing): Contains several notable outliers with high idle times (50-300 minutes), occurring across various SMV values. The occasional high idle time events in Department 1 likely represent specific operational disruptions (possibly material shortages, equipment failures, or workflow bottlenecks).
- Department 0 (Sewing): Shows consistently minimal idle time, with virtually no significant outliers. The consistent lack of idle time in Department 0 suggests better process control or fewer external dependencies.
- Low Overall Idle Time: Both departments maintain efficient operations with minimal downtime in most instances.

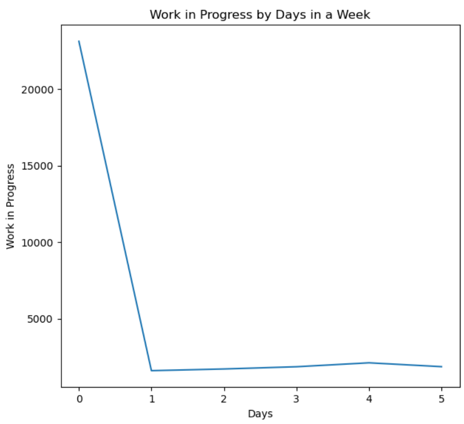
→ *SMV vs. Work in Progress*

- Department 0 (Sewing): Maintains consistently low WIP levels (below 1,000) across its narrow SMV range. The consistently low WIP suggests efficient throughput with minimal accumulation of unfinished items.
- Department 1 (Finishing): Shows varied WIP levels, with most below 2,000 but several significant outliers reaching 10,000-20,000 items. The occasional very high WIP values indicate periodic accumulation of unfinished items, suggesting potential capacity constraints or process bottlenecks.
- Distribution Pattern: No strong correlation between SMV and WIP is evident in either department.
- WIP Independence from Task Complexity: The lack of correlation between SMV and WIP suggests that WIP levels are more influenced by production scheduling, capacity, or external factors than by the complexity of the tasks.

9. Work in Progress (WIP) Analysis

a. *Time-Based Analysis*





→ *Work in Progress Over Time*

- Baseline WIP: For most of the time period (January through March), WIP levels remain consistently low and stable, typically below 2,000 items.
- Dramatic Spike: There is a single, extremely sharp spike occurring around February 1, 2015, where WIP suddenly jumps to approximately 22,000 items.
- Post-Spike Recovery: Immediately after the spike, WIP levels quickly return to the baseline, suggesting rapid resolution of the backlog.
- Stable Operations: The generally consistent low WIP levels indicate efficient production flow under normal circumstances, with minimal accumulation of unfinished items.
- Major Disruption Event: The isolated, extreme spike likely represents a significant operational disruption, such as:
    - A large batch of new orders arriving simultaneously

    - A temporary production stoppage (e.g., equipment failure, material shortage)
    - A transition between production seasons or major style changes
- Effective Resolution: The quick return to baseline suggests effective management intervention to clear the backlog, possibly through overtime, additional resources, or prioritization strategies.

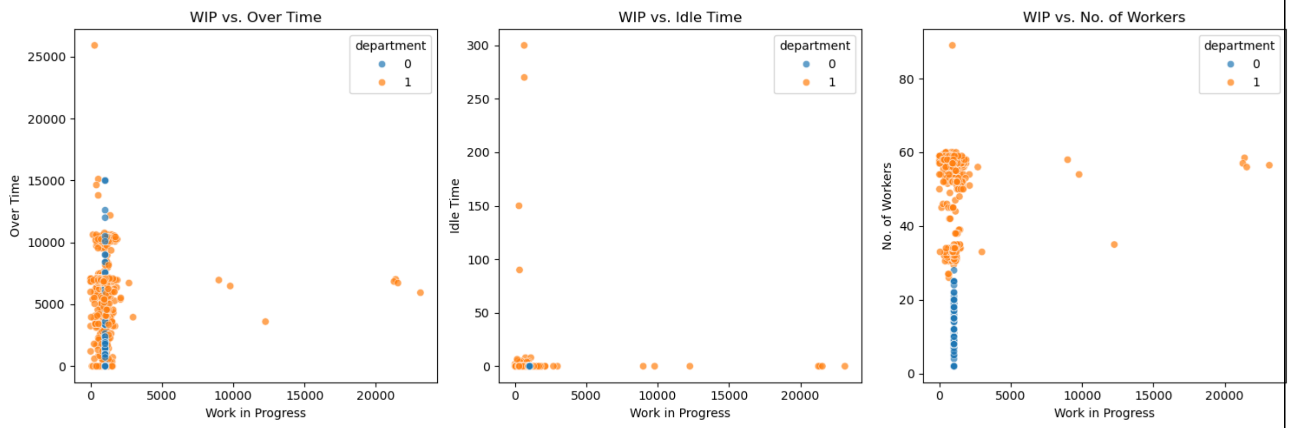
→ *Work in Progress by Days in a Week*

- Monday (Day 0): Shows extremely high WIP levels (approximately 22,000 items).
- Tuesday (Day 1): Dramatic decrease to around 1,500 items.
- Wednesday-Saturday (Days 2-5): Relatively stable and low WIP (approximately 1,500-2,000 items).
- Monday Accumulation: The significantly higher WIP on Mondays strongly suggests:
    - Weekend inventory buildup (work accumulating over non-production days)
    - Beginning-of-week order processing or material delivery
    - Monday being used as a planning or setup day, with actual production ramping up on Tuesday
    - Weekly Clearing Pattern: The sharp drop between Monday and Tuesday indicates an efficient system for processing the Monday backlog.
- Mid-to-Late Week Stability: The consistent low WIP from Tuesday through Saturday suggests steady production flow once the weekly cycle is established.

Insights and Recommendations:

- Process Vulnerability: Despite generally good WIP management, the system appears vulnerable to occasional extreme disruptions that can cause inventory to increase by a factor of 10+ times baseline.
- This WIP pattern may help explain earlier observations about productivity variations by day of week, where Tuesday showed peak productivity – likely reflecting the focused effort to process Monday's accumulated WIP.
- Consider strategies to smooth the Monday-Tuesday transition (e.g., staggered deliveries, weekend processing)
- Maintain the effective clearing strategies that allow quick recovery from WIP accumulation

b. *Assessing Correlations*



→ *WIP vs. Over Time*

- Concentration Pattern: Most data points from both departments are clustered at low WIP values (below 5,000), with overtime ranging from 0 to 15,000 minutes.
- Department 0 (Sewing): Consistently shows low WIP (all points below 2,500), with overtime values ranging from near zero to about 15,000 minutes.
- Department 1 (Finishing): Shows mostly low WIP values similar to Sewing, but with several outliers extending to much higher WIP (10,000-25,000).
- Inverse Relationship: For higher WIP values (above 5,000), overtime tends to be lower (below 7,000 minutes).
- Typical Operations: Under normal conditions, both departments maintain low WIP levels regardless of overtime worked.
- Strategic Response: The lower overtime during high WIP scenarios suggests that overtime is not the primary strategy for addressing WIP backlogs, or that these backlogs occur during periods when overtime cannot be effectively deployed.

→ *WIP vs. Idle Time*

- Zero-Idle Pattern: The vast majority of data points show near-zero idle time regardless of WIP level.
- Department 1 (Finishing): Contains all the significant idle time outliers, with values ranging from 50 to 300 minutes.
- Idle Time Distribution: High idle time instances occur primarily at very low WIP levels, with a few exceptions at higher WIP values
- Operational Efficiency: Generally, operations maintain minimal idle time across all WIP scenarios, indicating effective worker utilization.
- Finishing Department Challenges: The presence of high idle time events exclusively in Department 1 suggests:
    - Process bottlenecks or wait times unique to finishing operations
    - Possible material shortages or quality issues requiring production pauses
    - Less effective idle time management compared to the Sewing department

- Disruption Pattern: High idle time at low WIP suggests external disruptions (equipment failure, material shortages) rather than process-related bottlenecks that would lead to WIP accumulation.

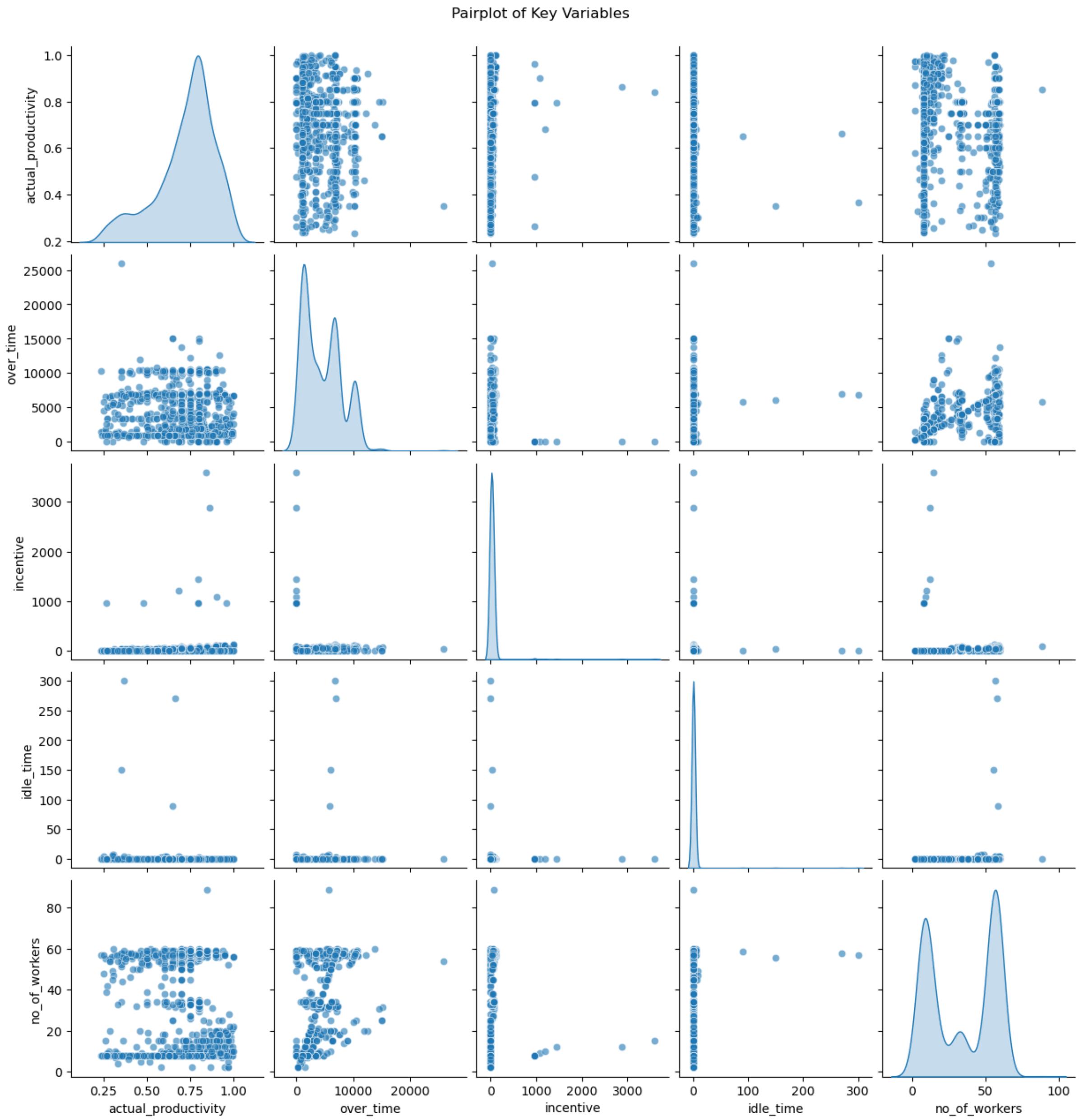
→ *WIP vs. No. of Workers*

- Department Staffing Differences: Clear separation between departments:
    - Department 0 (Sewing) consistently operates with smaller teams (2-25 workers)
    - Department 1 (Finishing) operates with larger teams (30-60 workers), with occasional teams of up to 85 workers
- Resource Allocation: The consistent pattern of smaller teams in Sewing and larger teams in Finishing reflects the structural differences in these production stages:
    - Sewing may be more machine-oriented with fewer workers per production line
    - Finishing likely involves more manual processes requiring larger teams

Insights and Recommendations:

- Overtime Strategy: The data suggests overtime is not consistently used to address WIP backlogs, as high-WIP scenarios don't show correspondingly high overtime.
- These patterns align with earlier observations showing that the Sewing department slightly outperformed its productivity targets while Finishing slightly underperformed.
- Analyze the specific process bottlenecks in Finishing that lead to occasional high WIP despite large team sizes
- Investigate the root causes of idle time in Finishing to improve worker utilization
- Consider workflow optimization rather than additional staffing to address WIP issues in Finishing
- Examine whether Sewing's process management strategies could be adapted for the Finishing department.

10. Pair Plot of Key Variables



→ *Departmental Differences*: The bimodal patterns in worker numbers clearly highlight the different operational structures of Sewing (smaller teams) and Finishing (larger teams).

→ *Productivity Drivers*: Actual productivity shows complex relationships with other variables, with no single factor showing strong linear correlation, suggesting multifaceted influences on performance.

→ *Incentive Misalignment*: The weak relationship between productivity and incentives suggests an opportunity for better alignment of reward structures with performance.

→ *Operational Efficiency:* The extremely low idle time across most observations confirms overall efficient operations, with disruptions being exceptional rather than routine.

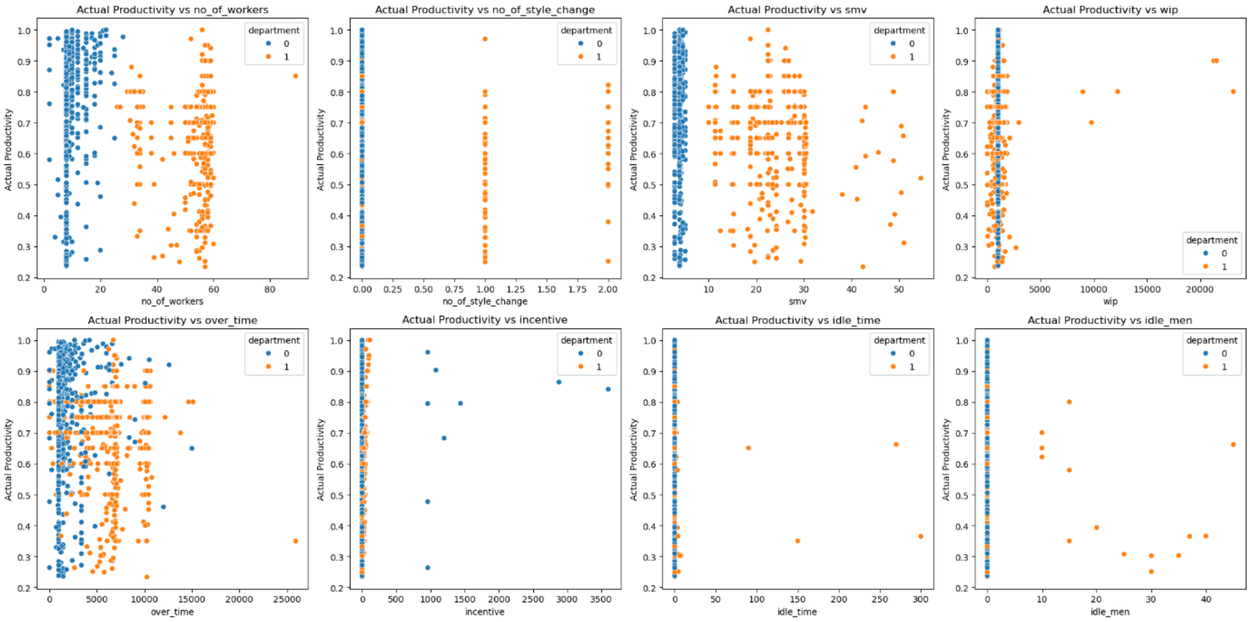
→ *Resource Allocation*: Larger teams tend to work more overtime, which aligns with earlier observations about the Finishing department having more workers but also more overtime.
→ *Productivity Variance*: Smaller teams (Sewing) show wider productivity variance, suggesting that small teams may be more affected by individual performance variations.

Insights:

- Performance Management: The weak productivity-incentive relationship suggests a need to reevaluate incentive structures to better reward high productivity.
- Resource Optimization: The strong worker-overtime relationship indicates potential for better workforce planning to reduce overtime costs.
- Process Improvement: The occasional high idle time events represent opportunities for targeted process improvement to eliminate these disruptions.
- Team Structure: The different productivity patterns between small and large teams suggest potential benefits in optimizing team sizes for maximum efficiency.

KEY FACTOR ANALYSIS

1. Finding Correlations of Variables with Productivity



→ *Actual Productivity vs No. of Workers*

- For both departments, the majority of observations are concentrated below 50 workers.
- Productivity seems widely spread for lower worker counts, but doesn't show a consistent increasing or decreasing trend.
- Department 1 has more data points at higher worker counts (40–80), but productivity remains mostly between 0.5 and 1.
- No strong correlation between number of workers and actual productivity.
- Productivity efficiency may depend more on task complexity or process flow rather than team size alone.

→ *Actual Productivity vs No. of Style Changes*

- The number of style changes is heavily skewed toward 0 and 1.
- A few observations exist with 2 style changes.
- No clear trend in productivity change with increasing style changes.
- No direct relationship detected between style changes and productivity.
- Style change could be a disrupting factor, but its impact appears inconsistent.

→ *Actual Productivity vs SMV (Standard Minute Value)*

- Department 0 (Sewing) shows more clustering at higher productivity levels across SMV values.
- Department 1 (Finishing) shows more spread in productivity values, especially with SMVs between 20–40.
- Potential weak inverse relationship between SMV and productivity for Department 1.

- Higher SMV may indicate complex tasks, slightly lowering productivity in some cases.
- Department 0 seems more efficient across SMV ranges.

→ *Actual Productivity vs WIP (Work in Progress)*

- Majority of WIP values are below 5000.
- No clear trend between WIP and productivity.
- No evident correlation between WIP and productivity.
- High WIP doesn't necessarily hinder or help productivity.

→ *Actual Productivity vs Over Time*

- Overtime ranges from 0 to over 25,000 units.
- Department 1 has significant observations with high overtime but generally lower productivity.
- Department 0 has lower overtime and higher productivity clusters.
- Inverse relationship: more overtime correlates with lower productivity, especially in Department 1.
- High overtime may be a sign of inefficiencies or production delays.
- Could imply burnout or poor planning in Department 1.

→ *Actual Productivity vs Incentive*

- Most incentive values are low or 0.
- Slight indication that higher incentives (above 1000) correspond to higher productivity, especially in Department 0.
- Positive correlation in Department 0 between incentive and productivity.
- Department 1 shows less clear linkage.
- Incentivization may be an effective motivator but inconsistently applied across departments.

→ *Actual Productivity vs Idle Time*

- Idle time ranges up to 300.
- High idle time is often associated with very low productivity.
- Strong inverse relationship between idle time and productivity.
- As idle time increases, productivity significantly drops.
- Reducing idle time could be a key driver of performance improvement.

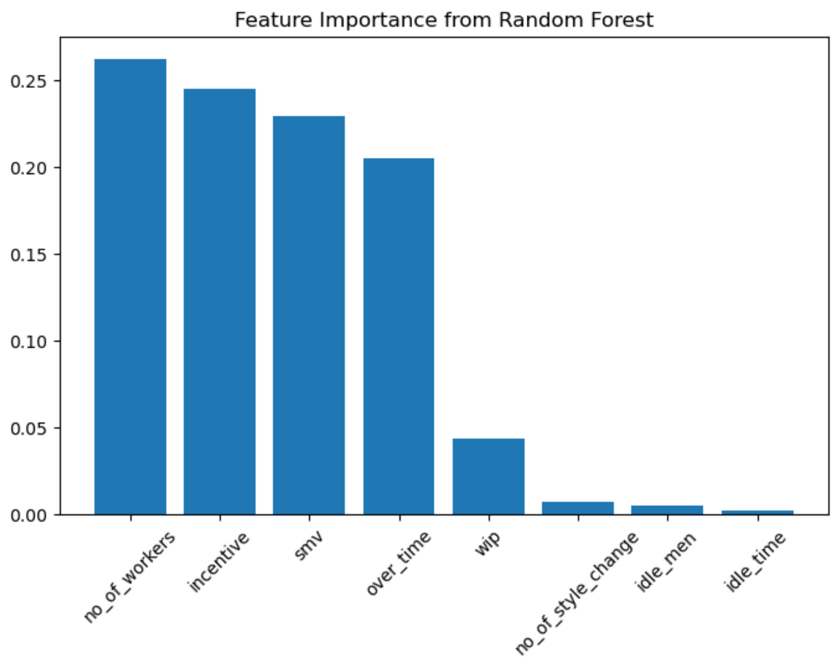
Insights and Recommendations

- No significant correlation between the number of workers and productivity indicates that simply increasing manpower does not guarantee better output. Focus on efficiency over size and capacity planning tools instead.
- Strong negative correlation between both idle_time and idle_men with productivity. Perform real-time tracking of idle resources using digital dashboards and implement task reallocation systems
- Higher over_time is associated with lower productivity, particularly in Department 1. Audit overtime policies—over-reliance may indicate process inefficiencies or poor

planning. Encourage a balanced workload distribution to reduce dependence on overtime.

- Leverage Incentive Schemes More Effectively: Tailor and promote fair, performance-linked and non-monetary incentives to boost productivity, especially in underperforming departments.
- Balance Task Complexity and Skill Readiness: Align complex tasks (high SMV) with experienced workers or training, and break them into manageable subtasks.
- Investigate Departmental Differences: Audit Department 1 for inefficiencies and apply Department 0's successful practices with department-specific KPIs.
- Reassess WIP Strategy: Optimize WIP levels using Kanban or pull systems, and review them with idle/overtime data to prevent hidden inefficiencies.

2. Feature Importance from Random Forest Regression



→ *Top 4 Most Important Features*

Rank	Feature	Importance	Interpretation
1	no_of_workers	~0.26	Most influential. Indicates how manpower scale affects productivity.
2	incentive	~0.24	A close second. Indicates reward systems strongly affect motivation and output.
3	smv (task complexity)	~0.23	Complex tasks significantly influence output—likely varies by worker skill.
4	over_time	~0.21	Indicates inefficiencies or stress in planning—affects productivity.

→ *Moderately Important*

Rank	Feature	Importance	Interpretation
5	wip	~0.045	Affects flow but less critical in prediction.

→ *Low Importance Features*

Rank	Feature	Importance	Interpretation
6	no_of_style_change	Very Low	Rare and minimal direct influence on productivity.
7	idle_men	Very Low	Despite visual correlation, possibly redundant with other features like no_of_workers.
8	idle_time	Lowest	Counterintuitive; may suggest redundancy with over_time or low variance in dataset.

Strategic Reconciliation & Final Insights

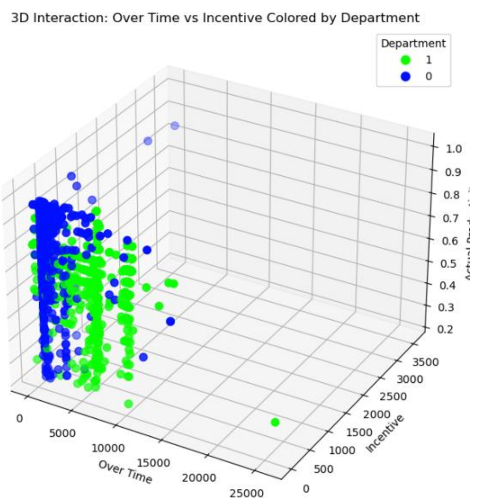
- Model confirms that workforce size, incentives, and task complexity are the primary levers to improve productivity.
- Visual trends on idle time/men don't translate into model importance—possibly due to overlap with other features (over_time or no_of_workers).
- Both visual and machine learning approaches agree that style changes and WIP levels play a minor role.

Strategic Priorities

- Optimize worker deployment based on skill and not just quantity.
- Revise incentive structures—especially in Department 1.
- Address overtime-related inefficiencies in scheduling and workload balancing.
- Upskill workers for high-SMV tasks or refine task design.
- Re-evaluate the utility of tracking idle metrics if they're redundant in predictive power.

3. Interactions of Variables and Productivity

a. *Overtime vs Incentive on Productivity*



→ *Low Overtime & High Incentive = Higher Productivity (optimal)*

- In the region where Over Time is low (below ~5000) and Incentive is moderately high, both departments exhibit relatively higher productivity (Z-axis).
- This reinforces the earlier insights that excessive overtime is counterproductive, and incentives play a motivating role.

→ *High Overtime = Low Productivity (Especially for Dept 1)*

- For Department 1 (green), high overtime (>10,000) correlates strongly with low productivity.
- Indicates diminishing returns or even negative impact of prolonged overtime for Dept 1, likely due to fatigue or poor planning.

→ *Low Incentive + High Overtime = Poor Performance Cluster*

- A dense cluster exists in the low incentive, high overtime quadrant for both departments, particularly visible for green points, all having very low actual productivity (~0.2–0.4).
- These are likely the least efficient combinations of resource use.

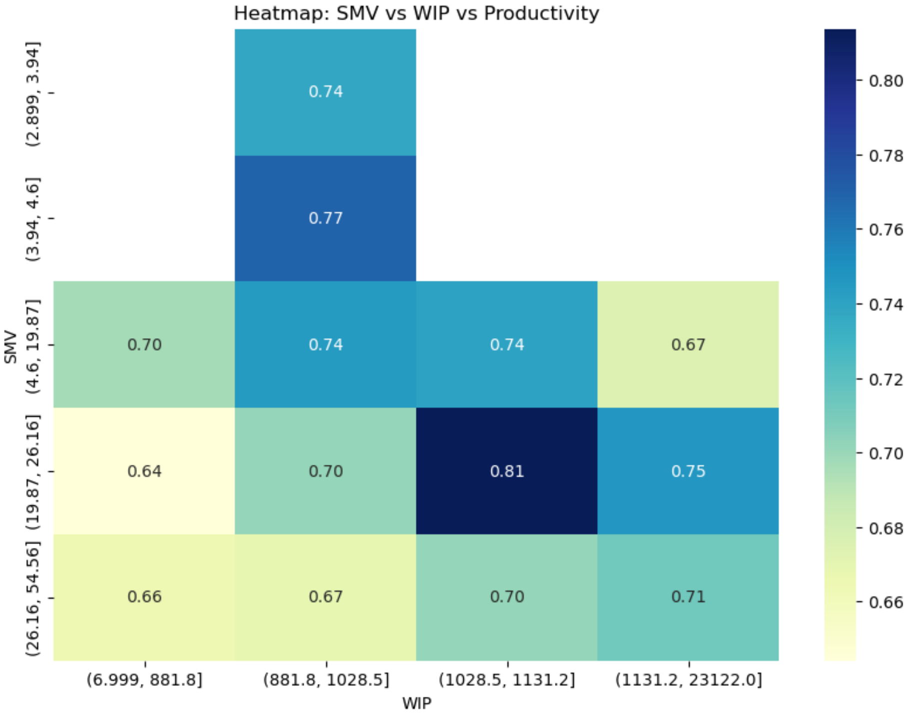
→ *Departmental Differences*

- Department 0 (blue) shows some higher productivity values even in moderate-to-high incentive settings, regardless of overtime levels.
- Suggests that Dept 0 responds better to incentive structures, or that they may have more efficient processes.

Insights & Recommendations

- Cap Overtime, Especially for Department 1. Consider setting a soft upper limit (~7000 units) on overtime for planning/scheduling.
- Introduce or Strengthen Incentive Programs. Incentives clearly correlate with better performance across departments. Review and potentially differentiate incentive schemes to fit department-wise behavioral patterns.
- Analyse Task Complexity vs Overtime: Since we know from earlier insights that smv (task complexity) is important, analyse if complex tasks are driving overtime. This can uncover process gaps or training needs.
- Department-Specific Interventions: Dept 1 needs stricter control on overtime and possibly workflow redesign. Dept 0 is comparatively stable but can still benefit from optimized incentive tuning.

b. Standard Minute Value (SMV) vs Work in Progress (WIP) on Productivity



→ *Productivity peaks at moderate SMV and moderate WIP*

- Highest productivity (0.81) occurs in the SMV range 19.87–26.16 and WIP range 1028.5–1131.2.
- There exists an optimal combination of task complexity (SMV) and workload (WIP) where productivity is maximized.
- These SMVs likely represent well-balanced task durations—not too simple (under-utilizing) nor too complex (overburdening).
- Mid-range WIP may indicate efficient workflow without bottlenecks or excess idle time.

→ *Very high or very low SMV correlates with lower productivity*

- Both extremes of SMV (2.9–3.94 and 26.16–54.56) show relatively lower productivity scores (between 0.66–0.74).
- Interpretation:
  - Low SMV = overly simple tasks may not leverage full workforce capacity.
  - High SMV = overly complex tasks may increase fatigue, errors, or inefficiency.

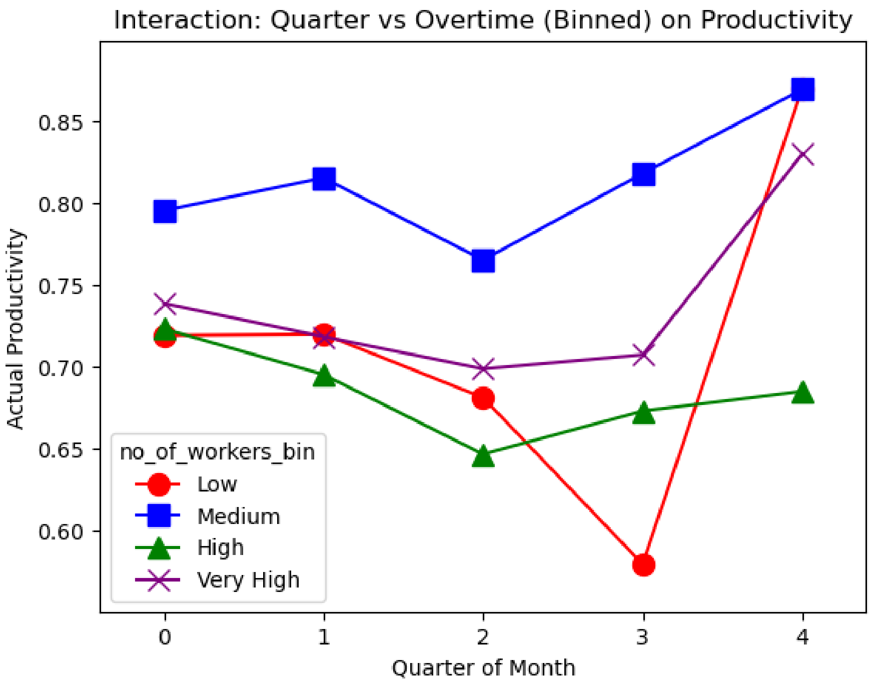
→ *Too little or too much WIP negatively affects productivity*

- Low WIP (<881.8) and very high WIP (>1131.2) generally lead to lower productivity.
- Interpretation: Low WIP: Idle resources, underutilization; High WIP: Congestion, multitasking inefficiencies, inventory buildup, possible workflow delays.

Insights & Recommendations

Observation	Strategic Insight	Recommendation
Highest productivity at mid SMV & WIP	A balance between task complexity and workload yields best results	Design tasks with moderate complexity (SMV ~20–26) and regulate WIP around 1030 units
Low SMV leads to underperformance	Simpler tasks may not challenge the workforce or utilize capacity fully	Avoid excessive micro-tasking; batch tasks into more substantial units
High SMV reduces productivity	Complex tasks slow down throughput	Break down high-SMV operations into smaller modular steps to boost flow
Extreme WIP disrupts output	Both idle and congested workflows lower productivity	Implement WIP caps and real-time monitoring systems to stabilize flow

c. ***Quarter vs Overtime on Productivity***



→ *Medium-sized teams consistently outperform others*

- The blue line (Medium workers) shows the highest and most stable productivity across all quarters, peaking at ~0.87 in Q4.
- Interpretation: Medium team sizes might strike a balance between coordination efficiency and manpower sufficiency.

→ *Low worker teams perform poorly—except in Q4*

- The red line (Low workers) has low and declining productivity till Q3, followed by a sharp rise in Q4 (from ~0.58 to 0.87).
- Interpretation:

        - Small teams may struggle with capacity or task overload early in the month.
        - Q4 spike could suggest end-of-month rush, deadline pressure, or performance incentives.

→ *High and Very High worker teams underperform consistently*

- Green (High) and Purple (Very High) lines remain flat or dip slightly, never crossing ~0.74 productivity.
- Interpretation:
    - Larger teams may suffer from coordination delays, communication gaps, or reduced accountability.
    - Crowding or "too many cooks" effect may reduce per-person output.

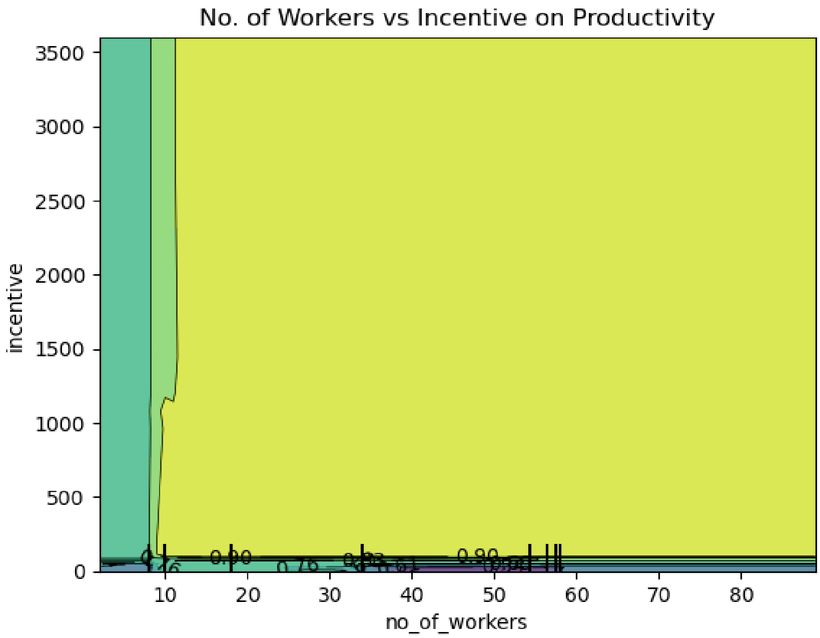
→ *Q4 tends to boost productivity overall*

- All lines (except High workers) rise noticeably in Q4.
- Interpretation: End-of-month urgency, quotas, or incentives could drive productivity up. Could also reflect improved rhythm or efficiency built over the month.

Insights & Recommendations

Observation	Insight	Recommendation
Medium teams perform best consistently	Team size optimization is crucial	Target medium team sizes (optimal range) for task assignments
Small teams spike late in month	Response to deadlines	Consider scheduling critical tasks for them in Q4 or rebalancing workloads earlier
High worker count does not equate to high productivity	Diminishing returns from scale	Avoid overstaffing, especially on tasks with limited parallelizability
Q4 boosts productivity	Time-driven motivation plays a role	Leverage this by aligning incentives or goals with month-end cycles

d. No. of Workers vs Incentive on Productivity



→ *Incentive is the dominant productivity driver*: A consistent high-productivity zone appears when incentives exceed ₹1000, even if the number of workers is moderate to high. This indicates a strong direct correlation between higher incentives and productivity gains.

→ *Worker count has diminishing standalone impact*: Merely increasing the number of workers without proportionately increasing incentives does not yield significant productivity improvements.

→ *Low incentive zones yield low productivity*: At lower incentive levels (below ₹1000), productivity remains suppressed, even as worker count increases.

Insights & Recommendations

→ *Prioritize Incentive-Based Productivity Models*

- Shift focus toward designing performance-linked incentive systems, especially for operational teams.
- Consider tiered incentive structures where productivity milestones unlock incremental rewards, thereby stimulating continuous performance.

→ *Optimize Manpower Allocation*

- Avoid simply scaling workforce numbers under the assumption of increased output. Without sufficient incentive motivation, adding workers has limited marginal benefit.
- Focus on right-sizing teams and directing resources toward financial motivation instead.

→ *Resource Efficiency Through Incentive Tuning*

- The company can likely achieve higher productivity with fewer workers by leveraging well-designed incentive schemes — a cost-effective approach.

- Run pilot programs that test different combinations of worker count and incentive levels to determine optimal workforce-incentive ratios.

→ *Incentive Strategy as a Retention & Morale Tool*

- Higher incentives not only enhance productivity but can also improve worker satisfaction and retention, particularly in labor-intensive environments.
- Consider integrating non-monetary incentives (recognition, skill-building opportunities) to further augment motivation.

→ *Data-Driven Incentive Benchmarking*

- Use this and similar analyses to continuously calibrate incentive policies, ensuring alignment with actual productivity outcomes.
- Create dashboards that track productivity elasticity to incentive changes for real-time feedback and planning.

ENVIRONMENTAL SUSTAINABILITY ANALYSIS

1. Derived Sustainability Metrics

1. **Productivity per Worker**

$$\text{Productivity per Worker} = \frac{\text{Actual Productivity}}{\text{Number of Workers}}$$

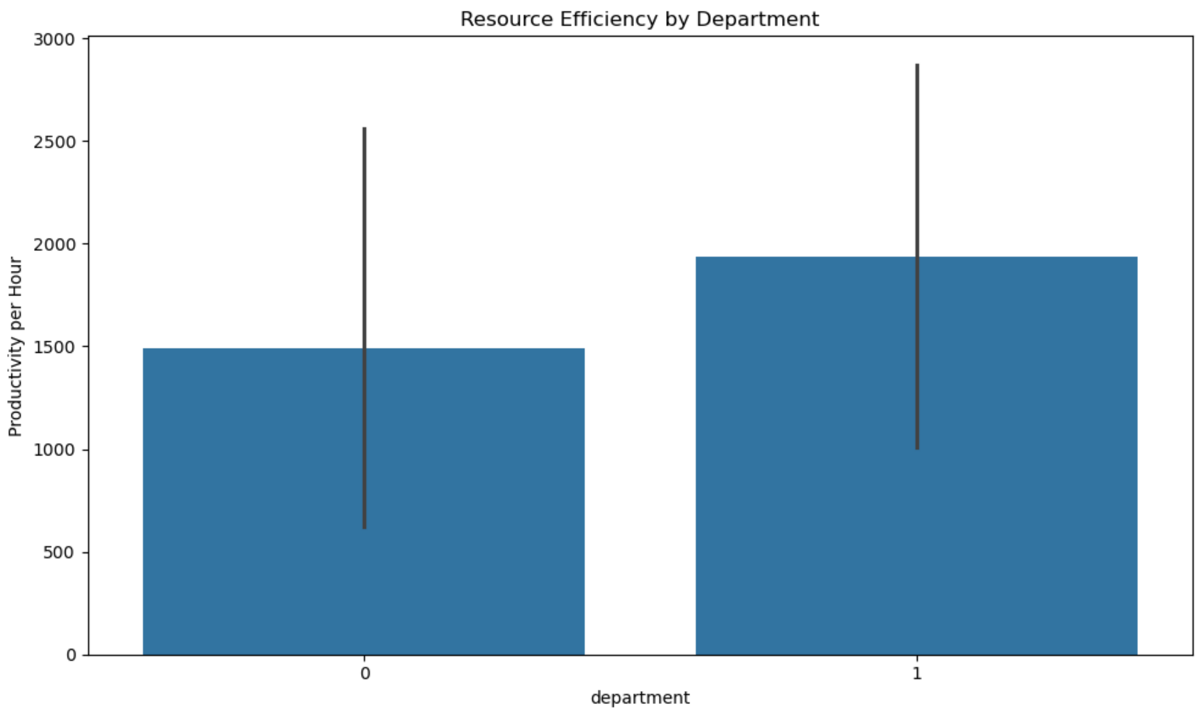
2. **Idle Time Ratio**

$$\text{Idle Time Ratio} = \frac{\text{Idle Time}}{\text{Overtime} + \varepsilon} \quad (\text{where } \varepsilon = 10^{-5} \text{ to avoid division by zero})$$

3. **Productivity per Hour**

$$\text{Productivity per Hour} = \frac{\text{Actual Productivity}}{\left(\frac{\text{Overtime}}{60}\right) + \varepsilon} \quad (\text{where } \varepsilon = 10^{-5} \text{ to avoid division by zero})$$

2. Resource Efficiency by Department



→ *This bar chart compares the average productivity per hour across two departments*.

- Department 1 outperforms Department 0 in productivity per hour, indicating higher time efficiency.
- The error bars are wide for both departments, but Department 1 has higher mean productivity and a slightly tighter lower bound, suggesting relatively better consistency in achieving efficient output.

→ *Sustainability Insights*:

- Department 1 demonstrates better resource utilization:
  - Higher productivity per hour implies that less time and potentially fewer energy resources are consumed to achieve the same or more output.
  - This aligns with sustainable operations principles: "doing more with less."

→ *Department 0 shows potential inefficiencies:*

- Lower average productivity per hour indicates longer time inputs for lower output, which can translate to higher energy consumption and operational emissions if machines or lighting are involved.
- Greater variation (seen in the error bars) may suggest inconsistent practices or variable idle times, pointing to process instability.

→ *Environmental cost per unit output is likely lower in Department 1:*

- From a sustainability standpoint, Department 1 is more eco-efficient — they are likely producing less waste, consuming less power per unit output, and reducing overall environmental load.

Recommendations

→ *Replicate Department 1's practices in Department 0:*

- Conduct process audits to understand the workflow, scheduling, and incentive mechanisms in Department 1.
- Identify best practices that can be transferred to Department 0 to standardize high-efficiency behaviours.

→ *Targeted interventions in Department 0:*

- Focus on training, workplace layout, or task distribution to reduce idle time and improve throughput.
- Investigate any equipment or procedural bottlenecks that may be dragging down efficiency.

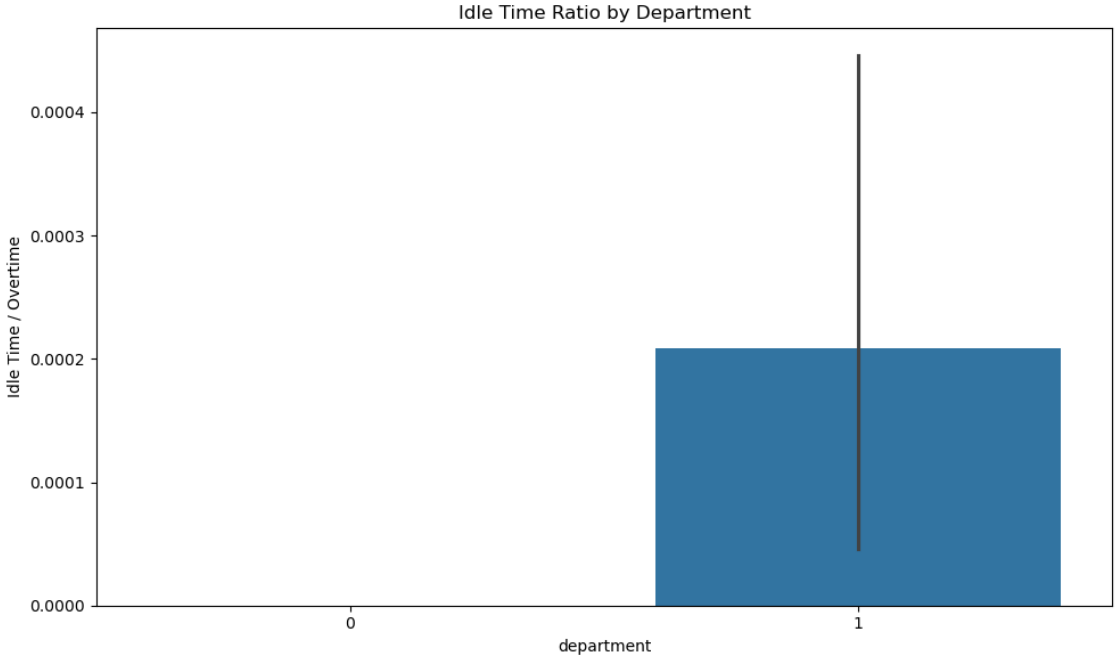
→ *Benchmark against sustainability KPIs:*

- Incorporate productivity-per-hour as an operational sustainability metric tied to energy or carbon intensity (e.g., kWh per unit output).
- Departments with lower productivity per hour could be flagged for energy audits.

→ *Promote uniform resource use practices:*

- High variance in both departments indicates inconsistent use of labor/time.
- Deploy standard operating procedures (SOPs) and monitor compliance using these sustainability-linked metrics.

3. ***Idle Time as Waste – Departmental View***



→ *This bar chart compares the Idle Time Ratio (defined as idle_time / over_time) across two departments:*

- Department 0 shows an idle time ratio essentially at zero, indicating negligible idle time relative to overtime.
- Department 1 shows a non-zero idle time ratio, though still quite low in absolute terms (~0.0002), with some variability (as reflected by the error bar).

→ *Operational Efficiency*

- Department 0 appears to maximize overtime utilization, minimizing downtime. This suggests tight scheduling, better workflow orchestration, or less bottlenecking during extended shifts.
- Department 1, despite higher overall productivity, exhibits slight inefficiencies in how overtime is managed—some portion of it is not translated into productive output, suggesting coordination or resource mismatch.

→ *Potential Sustainability Risk*

- Idle time during overtime equates to wasted energy, labor hours, and facility usage, all of which:
    - Increase energy footprint per unit of output
    - Can reduce employee morale or engagement (due to perception of inefficiency)
    - Lead to inefficient resource consumption
- From an environmental sustainability lens, minimizing idle time during overtime is crucial for lowering indirect emissions and operational waste.

Insights and Recommendations

Aspect	Recommendation
Resource Planning	Review task allocations during overtime in Department 1. Introduce dynamic shift planning or staggered scheduling.
Sustainability	Idle time in overtime contributes to scope 2 emissions (electricity, lighting, HVAC during idle hours) — minimizing this supports ESG goals.
Monitoring Systems	Install better real-time monitoring or workflow feedback loops to catch idle phases early.
Training	Train supervisors on adaptive work allocation during overtime shifts.

4. ***Sustainability Score (Efficiency to Waste)***

sustainability_score = productivity_per_hour × (1 − idle_time_ratio)

	department	sustainability_score
1	1	1617.14
0	0	1493.59

Insights

Department	Sustainability Score	Interpretation
Finishing (1)	1617.14	Highest performing department, indicating efficient use of labor and time with minimal idle time and strong output per hour.
Sewing (0)	1493.59	Slightly lower score due to lower productivity per hour, though its idle time is minimal, suggesting tight operational control.

Although Department 0 had almost zero idle time, it was outperformed because Department 1 had a significantly higher productivity per hour, which compensated for its slightly higher idle time ratio.

Strategic Sustainability Implications

Area	Recommendation
Finishing Department	Continue optimizing task throughput. Focus next on reducing idle time to enhance the already strong score.
Sewing Department	Explore ways to boost output per hour without sacrificing current efficiency—e.g., tool upgrades, lean methods, or automation.
Score Utility	Use this sustainability score as a benchmark KPI to monitor operational and environmental performance across departments.

MACHINE LEARNING APPLICATION

The following regressions were carried out:

Technique	R Squared Value	Mean Absolute Error (MAE)	Mean Squared Error (MRE)
Multiple Linear	0.077	0.119	0.024
Random Forest	0.442	0.079	0.014
Gradient Boosting (XGBoost)	0.495	0.079	0.013
Support Vector	0.350	0.098	0.016
Neural Networks	0.099	0.099	0.018
Decision Tree	0.460	0.081	0.014

According to the analysis, Gradient Boosting Regression (XGBoost) is the best fitting model because:

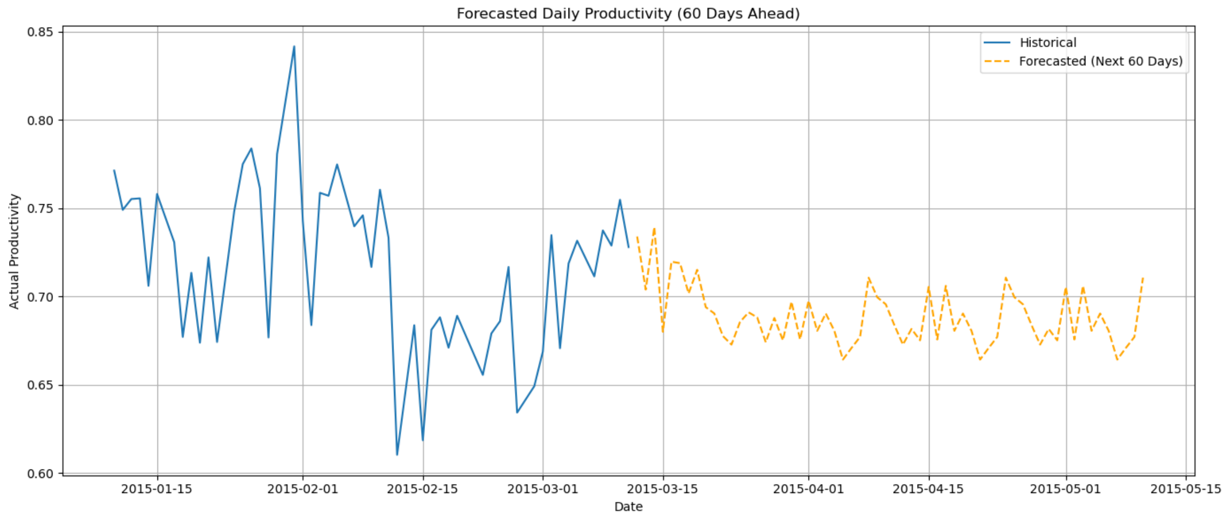
- Highest R^2 (0.495): This means XGBoost explains 49.5% of the variance in actual productivity — the best among all models.
- Lowest MSE (0.013): This indicates small average squared prediction errors, i.e., fewer extreme errors.
- Equal lowest MAE (0.079): XGBoost and Random Forest both perform best here in terms of average absolute error.

Thus, XGBoost Regression is used to carry out further simulation operations

Random Forest is a close second and a strong alternative, especially in this case as interpretability and feature importance *(as conducted earlier)* are crucial.

SIMULATION ANALYSIS

1. Forecasting (XGBoost Regression + lag)



Interpretations and Insights for Forecasted Value

- Predicted productivity values are less volatile.
- The range is narrower, roughly between 0.67 and 0.71.
- Indicates a flattening or stabilizing trend, likely due to smoothing from the forecasting model.
- The forecast suggests that future productivity is expected to stabilize at a moderately high level.
- No major increases or declines are predicted — suggesting no anticipated disruptions or improvements based on current trends.
- This could reflect a system in equilibrium or a model that's conservative in forecasting extremes.
- If higher productivity is a goal, intervention may be needed (e.g., incentive programs, reducing idle time).
- Stability can be good for predictability in resource planning, but might also hint at a ceiling in current operations.

2. Sensitivity (What-if Analysis) – Using XGBRegressor

Top 4 variables based on Feature Importance (Random Forest Regression) conducted earlier are considered here.

	feature	change_pct	new_prediction	change_in_productivity
0	no_of_workers	-50	0.8483	-0.0978
1	no_of_workers	-40	0.8635	-0.0826
2	no_of_workers	-30	0.9114	-0.0347
3	no_of_workers	-20	0.9114	-0.0347
4	no_of_workers	-10	0.9088	-0.0373
5	no_of_workers	10	0.9325	-0.0136
6	no_of_workers	20	0.9325	-0.0136
7	no_of_workers	30	0.9325	-0.0136
8	no_of_workers	40	0.9325	-0.0136
9	no_of_workers	50	0.9325	-0.0136
10	incentive	-50	0.7899	-0.1561
11	incentive	-40	0.8066	-0.1394
12	incentive	-30	0.8559	-0.0902
13	incentive	-20	0.8712	-0.0748
14	incentive	-10	0.9026	-0.0435
15	incentive	10	0.9461	0.0000
16	incentive	20	0.9878	0.0417
17	incentive	30	0.9878	0.0417
18	incentive	40	0.9878	0.0417
19	incentive	50	0.9878	0.0417
20	smv	-50	0.9262	-0.0199
21	smv	-40	0.9331	-0.0130
22	smv	-30	0.9354	-0.0107
23	smv	-20	0.9573	0.0112
24	smv	-10	0.8953	-0.0508
25	smv	10	0.9499	0.0038
26	smv	20	0.8967	-0.0494
27	smv	30	0.8422	-0.1039
28	smv	40	0.8422	-0.1039
29	smv	50	0.8422	-0.1039
30	over_time	-50	0.9440	-0.0020
31	over_time	-40	0.9422	-0.0039
32	over_time	-30	0.9419	-0.0042
33	over_time	-20	0.9445	-0.0016
34	over_time	-10	0.9461	-0.0000
35	over_time	10	0.9278	-0.0183
36	over_time	20	0.9256	-0.0205
37	over_time	30	0.9248	-0.0212
38	over_time	40	0.9248	-0.0212
39	over_time	50	0.9252	-0.0209

→ *no_of_workers*:

- Prediction increases up to a certain point but plateaus at ~0.9325.
- Change beyond ±10% leads to minimal change (≤ 1.3%).
- Insight: Number of workers shows diminishing returns — increasing beyond a point doesn't boost productivity further.

→ *incentive:*

- Positive correlation: as incentives increase, predicted productivity increases significantly.
  - e.g., +50% → productivity ↑ by ~4.2%
  - -50% → productivity ↓ by ~15.6%
- Insight: Strongest positive driver of productivity in this table. Incentive schemes are effective levers.

→ *smv (Standard Minute Value):*

- Impact is non-linear:
  - +10% shows a small increase (+0.0038), but +30% or more causes sharp drops (e.g., -10.4%).
- Insight: High SMV reduces productivity (possibly indicating inefficient tasks or overestimated work).

→ over_time:

- Increasing overtime shows a mild negative impact (max ~-2.2%).
- Insight: More overtime does not help productivity — may even hurt it, suggesting worker fatigue or diminishing efficiency.

Recommendations

Feature	Strategy	Reason
Incentive	Increase by 20–50%	Shows best improvement in productivity
No. of Workers	Keep stable or small tweak	Additional workers don't significantly help
SMV	Avoid high increases	Reduces productivity sharply after +20%
Overtime	Avoid increases	Leads to marginal or negative impact

3. Goal Seek Analysis

Where, target variable is 'actual_productivity' and the goal variable is 'targeted_productivity'

a. Using Linear Regression

```
Model R-squared: 0.04309702240677982
Model MSE: 0.027437160828433373
Optimal percentage change in no_of_workers to achieve target productivity: 5.07%
Optimal percentage change in over_time to achieve target productivity: 20.07%
Optimal percentage change in idle_time to achieve target productivity: -67.92%
Optimal percentage change in incentive to achieve target productivity: -71.58%
Optimal percentage change in smv to achieve target productivity: -10.32%
Optimal percentage change in wip to achieve target productivity: -36.89%
```

→ *Model Performance:*

- R^2 (0.043): Very low, indicating that the model explains only 4.3% of the variance in productivity.
- MSE (0.0274): This measures prediction error—fairly low, but the R^2 suggests poor model fit.

→ *Insights:*

- Large negative adjustments (e.g., -68% idle_time, -72% incentive) seem unrealistic or extreme, suggesting the model may not generalize well.
- The poor R^2 implies these changes are not reliable, and the model doesn't understand the complex relationships in the data well

b. ***Using Gradient Boosting Regression***

```
R-squared: 0.46293900718179615
MSE: 0.01373159843913397

Optimal percentage changes (to reach target productivity):
no_of_workers: 0.00%
over_time: 0.00%
idle_time: 0.00%
incentive: 0.00%
smv: 0.00%
wip: 0.00%
```

→ *Model Performance*

- $R^2 = 0.463$: The model explains about 46.3% of the variability in productivity. This is significantly better than the earlier linear (4%) and tree-based (33%) models.
- MSE = 0.0137: Indicates a low average error in predictions.
- This suggests the GBR is capturing non-linear relationships more effectively.

→ *Insights:*

- The model is already predicting productivity at or close to the target for the selected record. Hence, no changes are necessary.
- This contrasts with previous models where Changes were unrealistic (linear regression)
- A 0% change across the board means: "Productivity is already optimal for this case, or no marginal change in features will help further."

COMPREHENSIVE FINDINGS

→ *Productivity Trends:*
- Actual productivity is generally high but does not consistently meet targeted goals.
- Productivity fluctuates across time—highest at the beginning and end of quarters.
- Tuesday shows peak productivity while mid-week dips suggest possible fatigue.

→ *Influence of Key Variables:*
- No. of workers has a moderate impact but does not directly correlate with productivity improvements.
- Overtime negatively impacts productivity—suggesting fatigue or inefficiencies.
- Idle time and idle men show strong inverse relationships with productivity.
- Incentives significantly boost productivity, particularly in the Sewing department.
- Task complexity (SMV) impacts efficiency—higher complexity leads to lower productivity.
- Style changes disrupt performance, though their direct effect is relatively small.
- Work-in-progress (WIP) levels fluctuate, with large spikes hinting at bottlenecks.

→ *Departmental Differences:*
- Finishing department has larger teams but slightly underperforms compared to Sewing.
- Sewing department operates with smaller teams yet exhibits more efficient workflow.
- Finishing department incurs more idle time, despite having the majority of workers.

→ *Regression and Predictive Modeling Insights:*
- XGBoost performs best among tested models, explaining nearly 50% of productivity variance.
- Random Forest also ranks highly, offering strong feature importance insights.
- Linear models struggle due to multicollinearity and non-linear dependencies.
- Simulation suggests diminishing returns on increasing workers or overtime.

→ *Sustainability Insights:*
- Productivity per hour varies across departments—Finishing exhibits superior time efficiency.
- Higher overtime leads to wasted resources, increasing environmental footprint.
- Idle time during overtime reveals inefficiencies in resource utilization.

OVERALL BUSINESS RECOMMENDATIONS

→ *Address Productivity Gaps:*

- Revise productivity targets to ensure they are achievable without excessive overtime.
- Identify mid-week fatigue trends and optimize break schedules to maintain efficiency.
- Develop strategies for eliminating downtime spikes (e.g., improving shift planning).

→ *Workforce Strategy:*

- Rebalance manpower: Avoid excessive workforce scaling and focus on efficiency instead.
- Department-specific intervention: Implement workflow optimization for Finishing to reduce idle time.
- Minimize unnecessary overtime, as it negatively impacts worker efficiency.

→ *Incentive Structuring:*

- Incentives must be recalibrated to align with actual productivity drivers.
- Department 1 requires a revised reward system, as its incentive model is weaker.
- Experiment with tiered incentives, unlocking incremental rewards for performance.

→ *Reducing Idle Time:*

- Track idle metrics dynamically and develop real-time dashboards for supervisors.
- Implement task reallocation systems to assign work to idle employees.

→ *Work-in-Progress (WIP) Adjustments:*

- Introduce pull systems or Kanban frameworks to regulate WIP inventory.
- Investigate the February spike and introduce controls to prevent similar disruptions.

→ *Sustainability Integration:*

- Ensure time efficiency is optimized to reduce unnecessary energy consumption.
- Idle time minimization contributes to better ESG goals, cutting resource waste.
- Benchmark productivity per hour and adjust resource allocation accordingly.

CONCLUSION

This analysis highlights the apparel industry's productivity optimization landscape, identifying key factors that drive efficiency and areas requiring intervention. The data clearly shows that workforce size alone does not dictate output—instead, incentives, idle time, task complexity, and overtime efficiency are stronger determinants of productivity. While the Finishing department dominates the workforce size, inefficiencies such as idle time and weak incentive alignment hinder its ability to match Sewing's streamlined operations.

On the predictive modeling side, while XGBoost is the best fit among tested models, data preprocessing gaps remain—notably, noise removal and multicollinearity treatment are necessary before further refinements. A more robust strategy for reducing idle time, optimizing incentives, and controlling overtime dependency will be crucial for improving operational flow and long-term efficiency.

Ultimately, an integrated approach—balancing predictive analytics, incentive structuring, and process efficiency—will be key to transforming productivity in a sustainable and scalable manner.

HOW I LEVERAGED GENERATIVE AI FOR THIS PROJECT

I used a variety of Generative AI tools, which included – ChatGPT, Copilot, ClaudeAI, Gemini to aid and enhance this project. AI-assisted code generation helped speed up Python scripting for visualization and analysis, while also acting as a debugger by troubleshooting errors. It also helped in identifying key patterns and trend.

In the modeling phase, AI insights guided the selection and tuning of algorithms like Random Forest, XGBoost, and SVR, enhancing accuracy and uncovering deeper patterns. For reporting, AI tools contributed creative and clear visualizations, making insights easier to communicate.

Overall, generative AI improved efficiency, minimized common challenges, and added depth to the analysis.