Technical Report: Final Project DS5110 Introduction to Data Management and Processing

Team Members: Hemraj Mahadeshwar, Mika Nguyen, Jenny Taconet

December 8, 2024

Contents

1 Introduction

The development of a data visualization dashboard to monitor key metrics and performance indicators in a manufacturing plant is a highly relevant and valuable initiative in the domain of data management and processing. Manufacturing plants generate substantial volumes of operational data, such as machine performance metrics, production losses, and equipment downtime. A well-designed dashboard allows executives to quickly identify challenges, recognize opportunities, and take informed actions. By transforming raw data into actionable insights, this project aims to enhance decision-making processes through the application of data science principles.

Optimal decision making is critical in the manufacturing sector, which is likely to gain significantly from data-driven insights. Leveraging operational data can reduce waste, improve production quality, and minimize costs. A data visualization dashboard serves as a practical tool to bridge the gap between raw data and actionable solutions, enabling organizations to address real-world challenges effectively.

This project focuses on analyzing employee productivity in the garment industry, a labor-intensive sector where manual processes play an important role. Meeting the high global demand for garment products is highly dependent on the productivity and efficiency of employees. For decision makers, tracking, analyzing and predicting team productivity is essential to improve performance and achieve organizational goals. By leveraging data visualization techniques, including time-series charts and summary tables, this dashboard highlights key metrics and provides actionable insights into employee productivity, helping organizations drive better outcomes in a competitive market.

2 Literature Review: Monitoring Key Performance Indicators and Metrics in Manufacturing Plants

The increasing complexity of modern manufacturing processes, coupled with the transformative technologies of Industry 4.0, has made the monitoring of Key Performance Indicators (KPIs) a crucial tool for optimizing operations, enhancing sustainability, and improving real-time decision-making. Advanced tools such as the Internet of Things (IoT), edge computing, and sustainability frameworks have brought significant advancements in KPI management across various manufacturing domains.

2.1 Role of IoT and Edge Computing in KPI Measurement

The integration of IoT and edge cloud computing has revolutionized real-time data acquisition, processing, and analysis in manufacturing. IoT gathers detailed data from machines and equipment, creating comprehensive performance reports, while edge computing facilitates low-latency data processing, enabling real-time KPI monitoring. These technologies address longstanding issues of incompatible communication protocols and data silos by promoting open standards and seamless device integration. In a cement manufacturing plant, an IoT-enabled platform demonstrated its ability to monitor KPIs related to pressure, temperature, and machine performance. The platform's integration with predictive analytics powered by artificial intelligence (AI) and machine learning enhanced decision-making and reduced the need for manual interventions. Additionally, IoT systems improve inventory management, quality control, and worker safety through real-time alerts and automated responses.

2.2 Sustainability and KPI Optimization in Industry 4.0

Sustainability has emerged as a critical focus in manufacturing, aligning with global initiatives such as the United Nations' Sustainable Development Goals (SDGs). The integration of sustainability frameworks within manufacturing enables the optimization of resource use, reduction of environmental impact, and achievement of broader ecological and economic objectives.

2.3 Advances in Visualization and Decision Support

Effective visualization of KPIs is essential for enabling plant managers and stakeholders to make informed decisions. Modern platforms employ advanced human-machine interfaces (HMIs) and visualization tools like Tableau and Power BI to present KPIs on both static and mobile devices. These tools support hierarchical KPI visualization, tailored to various organizational levels, from machine operators to senior management.

Predictive analytics capabilities further enhance decision support by identifying potential breakdowns and optimization opportunities. For example, combining big data analytics with IoT allows manufacturing plants to predict energy demand, schedule maintenance, and optimize production processes.

3 Methodology

3.1 Data Collection

To achieve the objectives of this project, we selected a static dataset that provides detailed information about employee productivity in the garment industry. This dataset, hosted on OpenML, contains comprehensive attributes related to various productivity metrics of workers. Key attributes include data on production lines, efficiency rates, standard hours, target achievement percentages, and the impact of factors such as over-time, incentive systems, and idle times on productivity. The dataset also includes temporal data, such as working hours and dates, allowing for trend analysis over time.

The selection of this dataset aligns closely with the project's goals of understanding performance dynamics in the garment manufacturing sector. By leveraging this robust foundation, we aim to uncover actionable insights into employee performance and operational efficiency. These insights will be visualized and analyzed using advanced data analysis techniques to facilitate data-driven decision-making and optimize productivity within the sector. The dataset's structured nature also allows for the integration of predictive modeling to forecast productivity trends and assess the effectiveness of potential interventions.

3.2 Data Preprocessing

The data pre-processing for this dataset involved several key steps to ensure the data was clean, consistent, and ready for analysis. First, missing values (nulls) were identified and appropriately handled, ensuring that incomplete data did not interfere with the analysis. The key performance indicators (KPIs) were then derived by performing relevant calculations, such as overtime per worker and productivity ratios, to assess the performance of the teams and the overall department. Additionally, the data was aggregated to provide meaningful insights at the team and department levels, enabling a clearer understanding of the operational efficiency and productivity metrics. This comprehensive preprocessing approach ensured that the dataset was well-structured for further analysis and visualization. This preprocessing helped transform raw data into actionable insights, making it easier to assess performance across different production units and better inform business decisions.

3.3 Analysis Techniques

The given productivity attributes of the dataset were utilized to calculate additional productivity metrics, including the productivity gap, productivity ratio, and worker-related metrics such as overtime per worker, incentive per worker, idle time per worker, style changes per worker, and worker efficiency. These metrics were aggregated across different departments and teams to identify specific trends and patterns. After data cleaning and feature extraction, we visualized departmental and team-specific metrics in our monitoring dashboard using user-friendly bar and line charts, making it easy for users to view and understand the data.

4 Results

The following figures include the data visualization dashboard to monitor key metrics and performance indicators for worker productivity in the garment industry. A live demonstration version of the dashboard can be viewed and accessed through Grafana on our project repository.

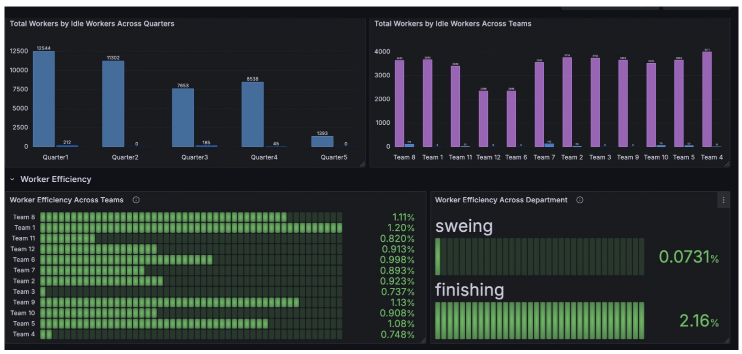

Figure 1: Worker Efficiency

Figure 2: Worker Metrics

Figure 3: Productivity

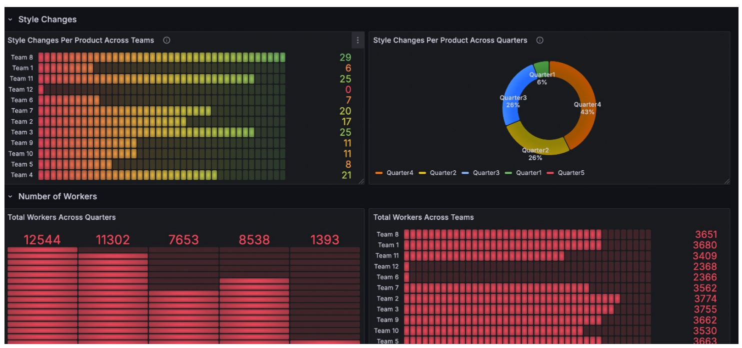

Figure 4: Style Changes and Other Worker Metrics

5 Discussion

The analysis of worker performance and team metrics across quarters reveals several key trends. The total number of workers fluctuates over time, beginning at 13,393 in Quarter 1 and gradually declining in subsequent quarters. This trend varies across teams, with some maintaining consistent workforce levels while others experience significant reductions. These fluctuations may reflect seasonal changes, project completions, or resource optimization efforts. Additionally, the number of idle workers peaks in certain quarters, such as Quarter 3 and Quarter 5, with significant variation across teams. This suggests inefficiencies in task allocation or workload distribution that need to be addressed to maximize productivity.

Worker efficiency across all teams is generally low, with metrics around 0.02 or less. Departments such as sewing and finishing exhibit notable variations in efficiency, but overall, there remains a significant gap between actual and targeted productivity. This discrepancy indicates a need for operational improvements, including enhanced training programs, workflow optimization, or performance-based incentives. Idle time per worker further highlights inefficiencies, ranging from 0 to 14.5 hours per worker in specific months. The spike in idle time during certain periods points to poor workload balancing or seasonal downtime, necessitating better resource planning.

Unfinished products present another area of concern. The number of incomplete items is highest in the early quarters, peaking at 517,189 in Quarter 1 before declining to 226,612 in Quarter 4. While the steady reduction indicates improvements in workflow management, the initial backlog suggests inefficiencies in earlier production stages. Overtime metrics reveal that hours worked beyond standard schedules decrease from 14.7 hours per worker in Quarter 1 to just 8.33 hours in Quarter 4. However, occasional spikes in overtime indicate short-term workload surges that may be due to inadequate planning or unforeseen demands.

While this analysis reveals idle time and productivity gaps, it does not leverage real-time data from IoT systems. The dataset's static nature limits the ability to simulate predictive interventions or real-time monitoring. The absence of IoT-generated data could explain why inefficiencies such as idle time and workload imbalances persist in these findings, suggesting that integrating IoT tools might mitigate these issues.

6 Conclusion

The analysis underscores several areas for improvement. Workforce allocation should be optimized by reallocating idle workers to high-demand projects, and productivity gaps can be addressed through skill development and workflow enhancements. Proactive task scheduling and predictive analytics can help minimize idle time and improve engagement. Efforts to streamline production workflows and manage unfinished products will further enhance efficiency. Finally, a review of overtime policies is recommended to identify and mitigate the causes of workload imbalances, ensuring better resource utilization and reduced dependency on overtime hours. These steps will contribute to sustained improvements in team performance and overall operational efficiency.

7 References

1. D. Georgakopoulos, P. P. Jayaraman, M. Fazia, M. Villari, and R. Ranjan, ”Internet of Things and Edge Cloud Computing Roadmap for Manufacturing,” IEEE Cloud Computing, vol. 3, no. 4, pp. 66-73, July-Aug. 2016, doi: 10.1109/MCC.2016.91.

2. J. Jose and V. Mathew, ”IoT Based Model for Data Analytics of KPI Platform in Continuous Process Industry,” Informatica, vol. 48, no. 1, 2024.

3. C. Favi, M. Marconi, M. Mandolini, and M. Germani, ”Sustainable life cycle and energy management of discrete manufacturing plants in the industry 4.0 framework,” Applied Energy, vol. 312, p. 118671, 2022.

Appendix A: Code

Data Preprocessing and Aggregation

```
df['productivity_gap'] = df['targeted_productivity'] - df['actual_productivity']
df['productivity_ratio'] = df['actual_productivity'] / df['targeted_productivity']

df['overtime_per_worker'] = df['over_time'] / df['num_workers']
df['incentive_per_worker'] = df['incentive'] / df['num_workers']
df['idle_time_per_worker'] = df['idle_time'] / df['num_workers']
df['style_changes_per_worker'] = df['style_changes'] / df['num_workers']
df['idle_time_ratio'] = df['idle_time'] / (df['over_time'] + df['smv'])
df['low_productivity_flag'] = df['actual_productivity'] < (0.7 * df['targeted_productivity'])
df['worker_efficiency'] = df['actual_productivity'] / (df['num_workers'] * df['smv'])
```

Figure 5: KPIs

```
team_metrics = df.groupby(['team', 'month']).agg({
    'actual_productivity': 'mean',
    'targeted_productivity': 'mean',
    'wip': 'sum',
    'over_time': 'mean',
    'worker_efficiency': 'mean',
    'idle_time': 'sum',
    'style_changes': 'sum',
    'num_workers': 'mean',
    'incentive': 'sum',
    'productivity_gap': 'mean',
    'productivity_ratio': 'mean',
    'overtime_per_worker': 'mean',
    'incentive_per_worker': 'mean',
    'idle_time_per_worker': 'mean',
    'style_changes_per_worker': 'mean',
    'idle_time_ratio': 'mean',
    'low_productivity_flag': 'sum'  # Assuming it’s a flag, summing could give the count of occurrences
}).reset_index()
```

Figure 6: Team Metrics

```
dept_metrics = df.groupby(['department', 'month']).agg({
    'actual_productivity': 'mean',
    'targeted_productivity': 'mean',
    'wip': 'sum',
    'over_time': 'sum',
    'worker_efficiency': 'mean',
    'idle_time': 'sum',
    'style_changes': 'sum',
    'num_workers': 'sum',
    'incentive': 'sum',
    'productivity_gap': 'mean',
    'productivity_ratio': 'mean',
    'overtime_per_worker': 'mean',
    'incentive_per_worker': 'mean',
    'idle_time_per_worker': 'mean',
    'style_changes_per_worker': 'mean',
    'idle_time_ratio': 'mean',
    'low_productivity_flag': 'sum'
}).reset_index()

dept_metrics
```

Figure 7: Department Metrics