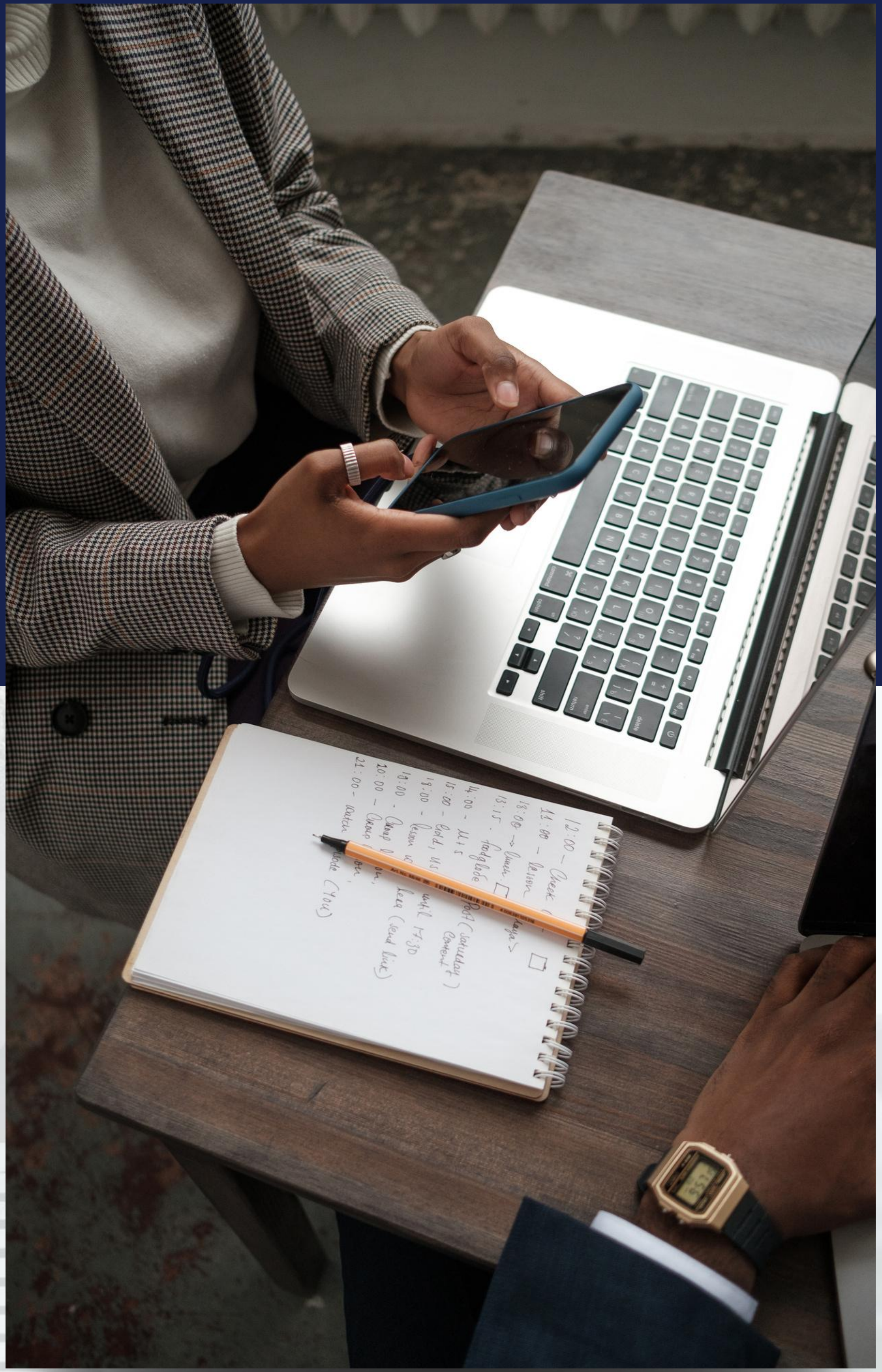
OCT, 10 2025

Continuous **Improvement**

DEFECT REDUCTION IN MANUFACTURING PROCESS

BY: TRIKA YULIANA

Intro-DUCTION



Project Description

This project analyzes factors incluencing defect rates in a manufacturing environment during Q2 2024, based on 3.240 production instances. The target variable is DefectStatus, where 0 for Low Defects and 1 for High Defects.

Features in the Dataset

Production Volume, Production Cost, Supplier Quality, Delivery Delay, Defect Rate, Quality Score, Maintenance Hours, Downtime Percentage, Inventory Turnover, Stockout Rate, Worker Productivity, Safety Incidents, Energy Consumption, Energy Efficiency, Additive Process Time, Addtivite Material Cost, Defect Status.

Source: Kaggle by Rabie Al Kharoua, licensed under Attribution 4.0 International (CC BY 4.0)

Tools & Concept:

1. Tools : Python (pandas, numpy, scipy, statsmodels, matplotlib)
2. Concept : Six Sigma (DMAIC framework), hypothesis testing, logistic refression, and probability-based improvement simulation

OUTLINE

1. DEFINE
   A. CREATING PROBLEM STATEMENT
   B. OBJECTIVE STATEMENTS
   C. SCOPE VS SCOPE CREEP
   D. CREATING TO CRITICAL TO QUALITY (CTQ) METRICS

2. MEASURE
   CALCULATING BASELINE METRICS FOR KEY FACTORS:
   A. CURRENT SIGMA LEVEL
   B. CURRENT KEY FACTORS OVERVIEW

3. ANALYZE
   A. PARETO CHART
   B. HYPOTHESIS TEST
   C. LOGISTIC REGRESSION

4. IMPROVE
   A. MEASURES AND IMPLICATIONS
   B. IMPLEMENTATION PLANNING

5. CONTROL

1. DEFINE

A. Problem Statement

In the Q2 2024, Electronic manufacturing process recorded a total of 3,240 production batches. Among them, 2,723 (84.0%) were classified as high defect occurrence, with quality score average < 80.0%, significantly exceeding the acceptable target of ≤10% and quality score ≥85.0%. Preliminary analysis indicates that factors such as supplier quality, delivery delays, maintenance hours, and worker productivity contributed to the issue, resulting in an estimated average production cost of $12,500 per batch.

B. Objective Statements

The goal of this project is to identify and mitigate the main root causes of high defect occurences in the manufacturing process during Q2 2024, aiming to gradually reduce the defect rate from 84.0% toward the acceptable target of ≤10%.

Measurable Target:

a. Identify key factors of defects occurences through data-driven analysis
b. Achieve an initial 8-10% reduction in defect rate and and 5-10% increase in avg quality score within the first quarter of improvement initiatives
c. Reduce average production cost per defective batch by USD 1,500 – 2,000 within the same period

C. Scope vs Scope Creep

SCOPE	SCOPE CREEP
This project focuses on factors directly influencing defect rates based on preliminary assumptions, including: - Supplier quality - Delivery delays - Maintenance hrs - Worker productivity	The following areas are excluded from this improvement phase: - Energy consumption - Production costs - External supplier logistics - Product-specific issues

D. Critical to Quality Metrics (CTQ)

DRIVER	CTQ METRICS
- Supplier Quality	- %, target high
- Delivery Delays	- Days, target low
- Machine Maintenance	- Hours, target optimal
- Worker Productivity	- %, target high

2. MEASURE

A. Current Sigma Level

Based on the initial analysis of 3,240 recorded data points, 2,723 were identified as high defects (Defect Status = 1), while the remaining units were classified as good. Using this data, the process yield is calculated as:

$$\left(\frac{\text{\# of opportunities - \# of defects)}}{\text{\# of opportunities}} \right) \text{X } 100 = \text{Yield}$$

((3,240 – 2,733)/3240) *100 = a yield of 15.96
This corresponds to a **DPMO of 840,432** and a **Sigma Level of 0.5**.

2. MEASURE

B. Current Key Factors Overview

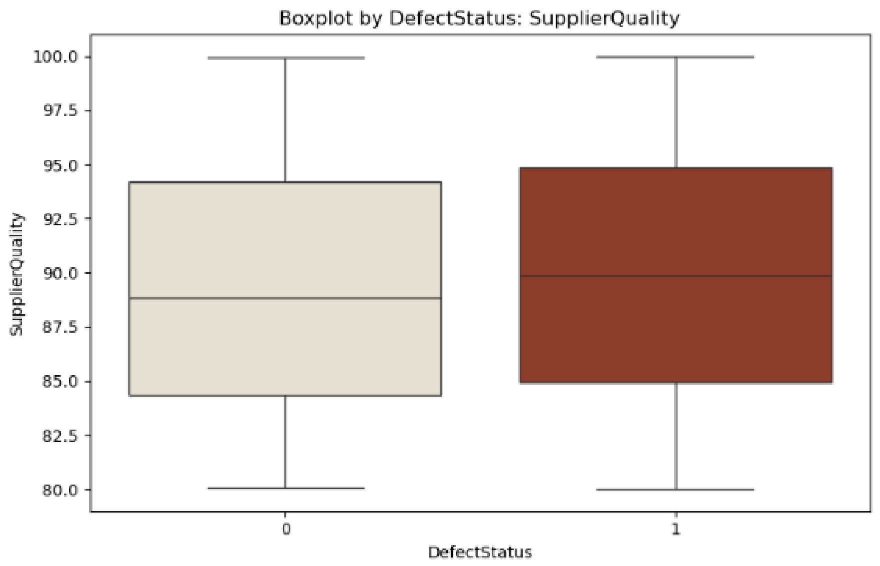
AVERAGE AND VARIATION

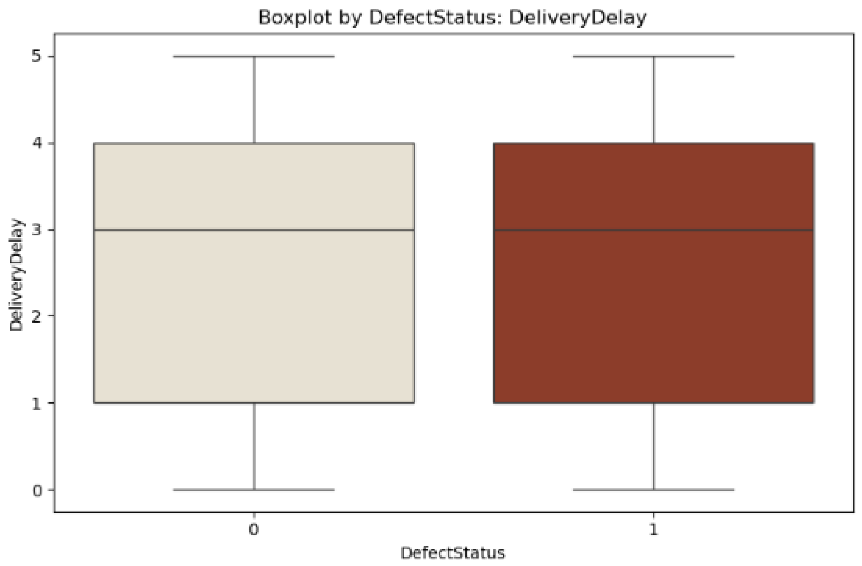
Defect Status	Supplier Quality		Delivery Delay		Maintenance Hours		Worker Productivity	
	Mean	Std	Mean	Std	**Mean**	**Std**	Mean	Std
0	89.33	5.79	2.54	1.69	**6.79**	**5.10**	90.11	5.72
1	89.93	5.75	2.56	1.71	**12.37**	**6.81**	90.03	5.73

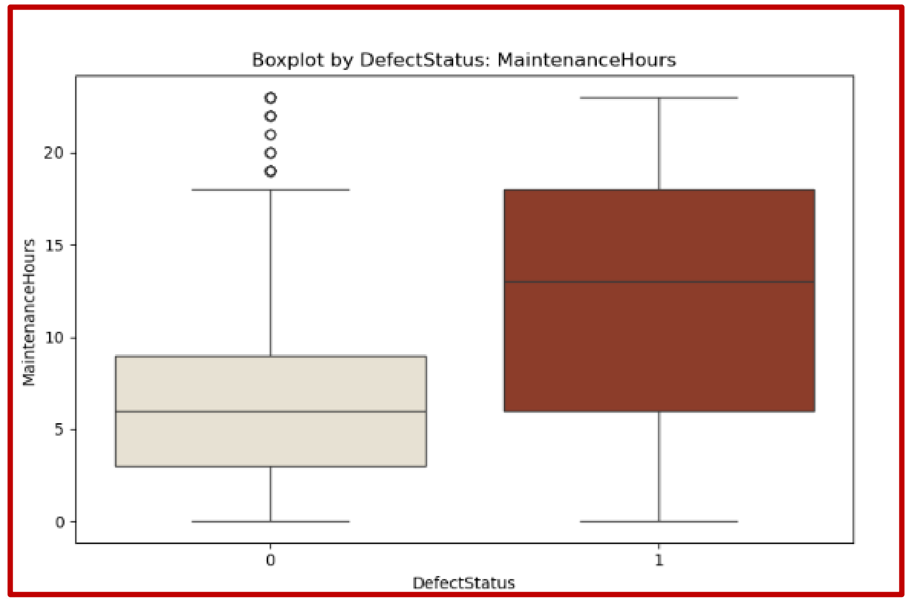
2. MEASURE

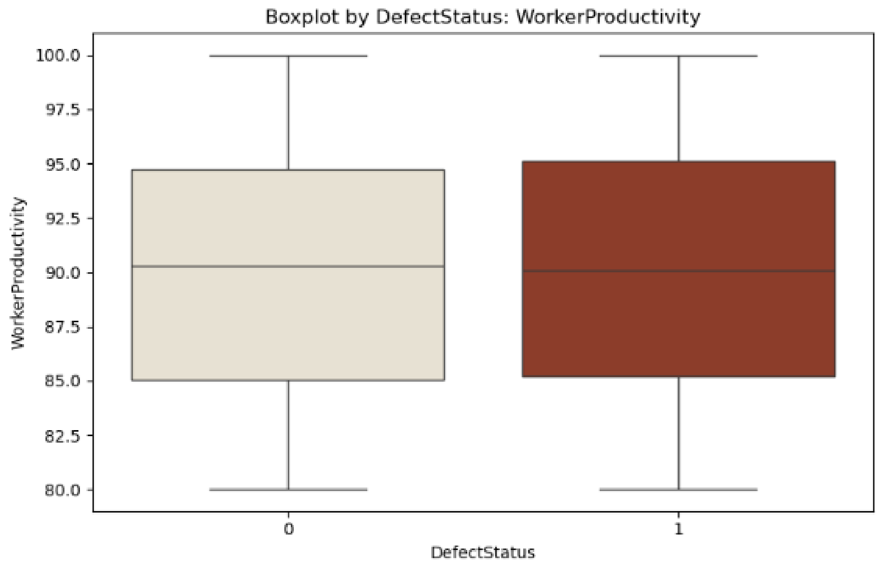
B. Current Key Factors Overview

BOX PLOT BY KEY FACTORS





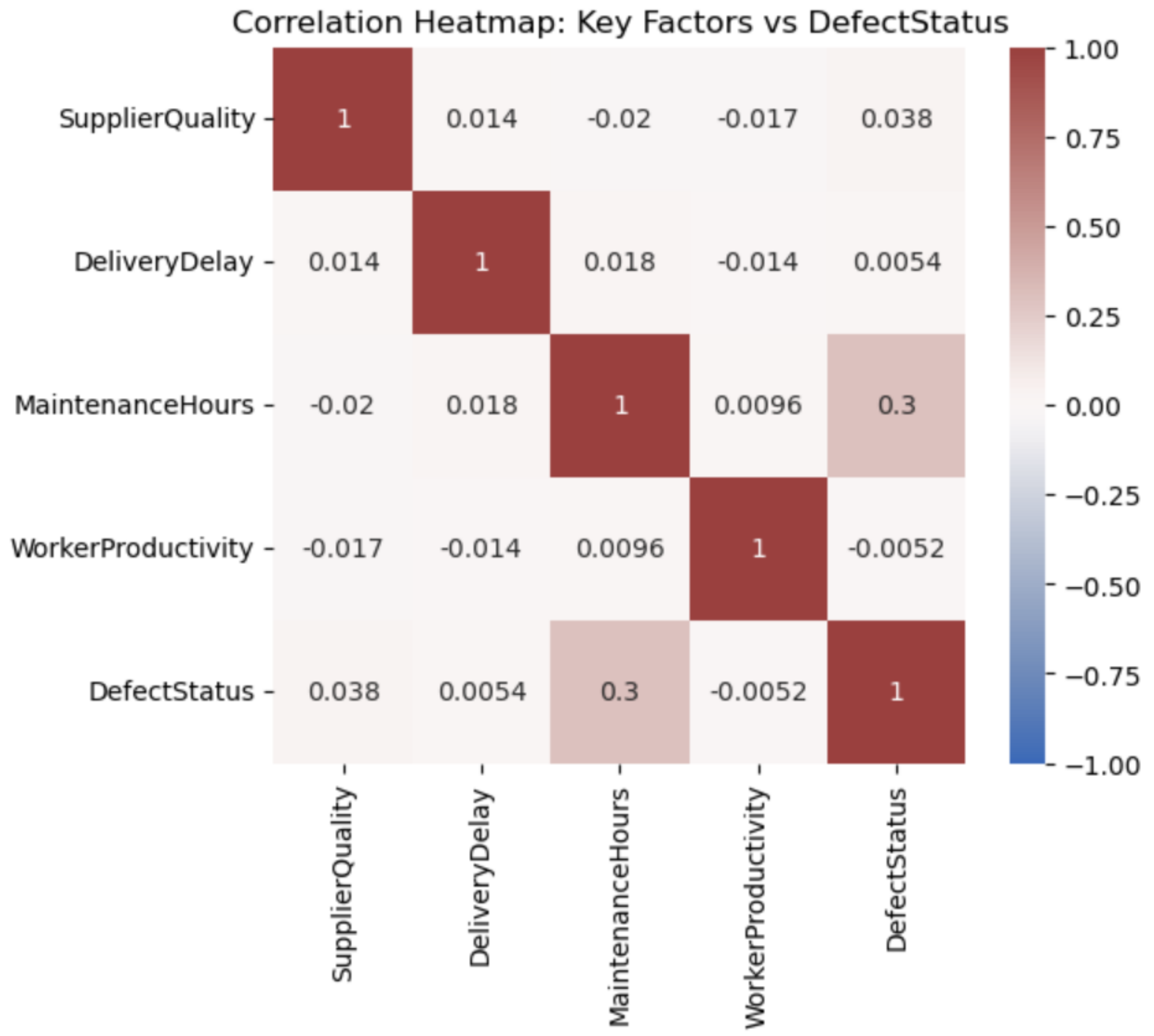




2. MEASURE

B. Current Key Factors Overview

KEY FACTORS VS TARGET



2. MEASURE

C. Initial Finding from Key Factors Overview

1. From the sigma level indicates that the process performance is **significantly below the desired Six Sigma target**, suggesting a high defect rate and substantial room for improvement.

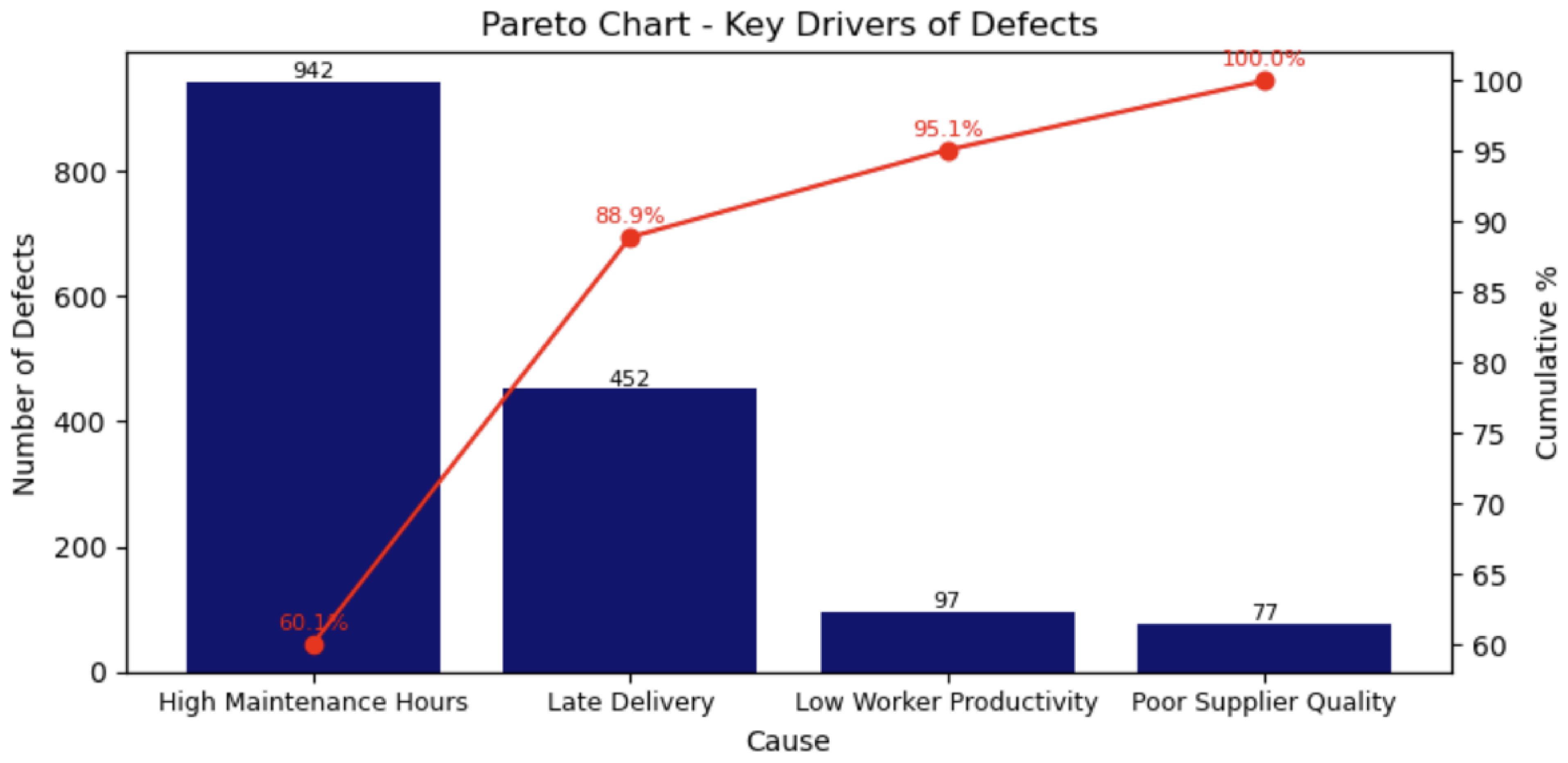
2. Among the 4 key factors, maintenance hours shows the most significant difference between low defect and high defect groups.
   Low defects, mean = 6.79 hrs, std = 5.10
   High defects, mean = 12.37 hrs, std = 6.81
   Other factors show mnimal variation between both groups.

3. Boxplot Distribution
   The median and IQR for most factors are nearly identical between defect groups, except for maintenance hours, where the high defect group clearly shifts upward.

4. Correlation Key Factors vs Target
   Supplier Quality → **0.04** (weak positive)
   Delivery Delay → **0.01** (very weak)
   Maintenance Hours → **0.30** (**moderate positive**)
   Worker Productivity → **–0.01** (negligible)
   Maintenance Hours has the strongest relationship with defect occurrence, suggesting that higher maintenance time tends to associate with higher defect rates.

3. ANALYZE

A. Pareto Chart



1. Based on Pareto Principle (80/20 rule), high maintenance and late delivery contribute nearly 90% of total defects.
2. High maintenance hours shows both a significant mean difference and moderate correlation (r = 0.3), confirming it as the primary driver of defects.
3. Late delivery, while having a low correlation, appears frequently in defect cases, possibly indicating indirect or compounding effects, such as production delays or rushed processes increasing defect likelihood.
4. Focus improvement: maintenance process optimization and review delivery scheduling alignment to minimize secondary impact on defect occurrence.

3. ANALYZE

From the previous findings, we identified Maintenance Hours as the main key driver affecting defect rate.
So, in this phase, we'll focus our improvement efforts there, while the other factors will be reviewed later since their correlation to defects isn't significant.

B. Hypothesis Testing

1. Normality Test using Shapiro-Wilk
   Low Defects normality p-value: 0.0000
   High Defects normality p-value: 0.0000
   Data is not distributed normally

2. Mann-Whitney U test (Non-parametric t-test)
   The Mann-Whitney U Test was conducted to compare the median Maintenance Hours between high defect and low defect groups.

   Result:
   Mann-Whitney U Statistic: 1033362.5, p-value: 3.71e-64

   This indicating a **statistically significant difference** between the two groups ($p < 0.05$), confirms that Maintenance Hours are significantly higher in batches with high defects.

C. Logistic Regression

```
Optimization terminated successfully.
         Current function value: 0.392174
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:           DefectStatus   No. Observations:                 3240
Model:                          Logit   Df Residuals:                     3238
Method:                           MLE   Df Model:                            1
Date:                Fri, 10 Oct 2025   Pseudo R-squ.:                  0.1066
Time:                        09:42:35   Log-Likelihood:                -1270.6
converged:                       True   LL-Null:                       -1422.2
Covariance Type:            nonrobust   LLR p-value:                 6.874e-68
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.4008      0.082      4.909      0.000       0.241       0.561
MaintenanceHours     0.1335      0.008     15.841      0.000       0.117       0.150
====================================================================================

Odds Ratio:
 const               1.493005
MaintenanceHours    1.142788
dtype: float64
```

$$\mathrm{logit}(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1 X$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

1. Analysis found that maintenance duration has a strong link with defect rates (p value < 0.001, Pseudo R^2 = 0.107).
2. The positive coefficient for Maintenance Hours **indicates that each additional hour increases the likelihood of defects.**
3. Based on model predictions, the probability of defect occurring is:
   - **Low Defect (0) = 77%**
   - **High Defect (1) = 85%**

4. IMPROVE

In this improvement we set as follows:
1. High defect (1), target maintenance process = 8 hrs (it shows the realistic maintenance window for a single shift)
2. Low defect (0), target maintenance process = 4 hrs (represents a targeted efficiency improvement without increasing defect risk based on historical data)

Finding

```
Optimization terminated successfully.
         Current function value: 0.330778
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:           DefectStatus   No. Observations:                 3240
Model:                          Logit   Df Residuals:                     3238
Method:                           MLE   Df Model:                            1
Date:                Fri, 10 Oct 2025   Pseudo R-squ.:                  0.2464
Time:                        10:27:41   Log-Likelihood:                -1071.7
converged:                       True   LL-Null:                       -1422.2
Covariance Type:            nonrobust   LLR p-value:                1.836e-154
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
const               -0.8223      0.103     -7.946      0.000      -1.025      -0.619
MaintenanceHours     0.4851      0.021     23.503      0.000       0.445       0.526
====================================================================================

Odds Ratio:
 const               0.439423
MaintenanceHours    1.624281
dtype: float64
```

Implications

Defect Status	Prob Before	Prob After	Reduction (%)
0	0.77	0.70	9.15
1	0.85	0.78	8.62
Overall	0.84	0.77	8.7

Quality Score	Before	After	Δ Improvement
Overall	80.13	80.60	+0.46

Production Cost	Before	After	Δ Saving
Overall	12,423.02	12,400.05	$ −22.97/batch or $74,426.10 saving for total production

1. Maintenance optimization has a stronger impact on quality (Pseudo R^2 increased from 0.107 → 0.246).
2. Probability of defects decreased in low defects (9.15%) and high defects (8.62%) and overall 8.70%
3. The continuous improvement also successfully improved the quality up to 0.46 point resulted in an estimated cost saving of bout $22.97 per batch or $74k across total production.
4. This demonstrates that reducing excessive maintenance hours can meaningfully reduce defect rates and production cost while improving quality score, without disrupting operational feasibility.

4. IMPROVE

Implementation Planning

Problem Statement	Validated Root Cause	Potential Solutions	Practical Method	Effectiveness	Feasibility	Cost Benefit	Overall	Take Action?
High defect rates	Excessive maintenance hours leading to increased downtime and inefficiency	Optimize maintenance scheduling to reduce unnecessary downtime and set to maintenance hour threshold (e.g., 4–8 hours per cycle) based on performance data	Adjust maintenance SOP to schedule preventive maintenance only within defined threshold; monitor via maintenance logs and QC dashboard	8	8	9	576	✅ Yes
		Implement predictive maintenance system using equipment runtime data	Integrate maintenance data with IoT sensors to trigger alerts before equipment failure	9	5	6	270	⚠️ later
		Conduct operator training	Train maintenance team to identify non-critical vs critical maintenance tasks	7	9	8	504	✅ Yes

Notes: This implementation plan is based on simulation or analytical assumption from the dataset.
In real-world practice, the improvement planning should be designed with multiple stakeholders including process owners, black belts, SME, etc to ensure operational alignment.

5. CONTROL

Objectives: Ensure that the improvement implemented during improve phase are sustained and the process remains stable. KPI can be used such as defect rate or quality score using Statistical Process Control (SPC).

In our simulation, we use defect rate to measure on SPC as follows:

Defect Status	Mean	Std	Sigma Level	Status Label
0	2.01	1.00	-0.01	Low Defects
1	2.89	1.31	-0.68	High Defects

DefectStatus	Mean	Std	SigmaLevel	Status_Label
0	0.74	0.44	2.87	Low Defects
1	0.77	0.42	2.93	High Defects

This shows that, the sigma levels have increased from near to 0 to 2.87-2.93, indicating that the process is now well below USL.

Note: in real operations, control chart would be used to continuously monitor defects or quality scores per batch or time period. In this simulation, the expected values from logistic regression model demonstrate the potential improvement but do not capture real time process variation.

OVERALL FINDING & CONCLUTION

1. From 3.240 production cycle, 2,723 (84.0%) were classified as high defect occurrence. Key factors analyzed included supplier quality, delivery delay, maintenance hours, and worker productivity.
2. Among these factors, maintenance hours played were the most significant contributor high defects with the probability for low defects = 77% and high defects = 85%.
3. Improvement measures involved optimizing maintenance hours to the target levels → low defects = 4 hrs and high defects = 8 hrs.
4. This improvement results:
   - Stronger impact on quality (Pseudo R^2 increased from 0.107 → 0.246).
   - Probability of low defects decreased by 9.15%, high defects 8.62%, and overall reduction of 8.70%
   - improved the quality score up to 0.46 point resulted in an estimated cost saving of bout $22.97 per batch or $74k across total production.
5. Recommended implementation planning for the initial improvement includes: 1). optimize maintenance scheduling to reduce unnecessary downtime, and 2). conduct operator training for the worker.
6. Control and monitoring: result should be continuously monitored per batch per production period to see track the variance ensure the effectiveness of the implemented measures.

THANK YOU