TRƯỜNG ĐẠI HỌC QUỐC TẾ MIỀN ĐÔNG
EASTERN INTERNATIONAL UNIVERSITY

MILK QUALITY MANAGEMENT

Course: SCLM 449 - Process Control and Improvement
Lecturer: Mr. Huynh Minh Binh
Group: Dream
Quarter: 1, 2025 - 2026

TEAM MEMBERS

NAME	IRN
1. Phan Đoàn Kháng Uyên	2032300232
2. Nguyễn Trần Linh Vy	2032300242
3. Nguyễn Thị Phương Anh	1832309008
4. Phạm Thị Ánh Tuyết	2032300527
5. Nguyễn Nhật Mai Yến	2132000189
6. Nguyễn Tuấn Vũ	2132300235
7. Nguyễn Thị Diễm Quỳnh	2132300258
8. Vũ Thị Mỹ Duyên	2132300019
9. Huỳnh Ngọc Anh Thư	2132300465
10. Huỳnh Thuỵ Bảo Trâm	2132300228

TABLE OF CONTENT

1. INTRODUCTION
2. ANALYSIS OF CURRENT QUALITY SITUATION
3. IDENTIFY THE CURRENT SITUATION AND SPECIFICATION
4. BUILDING THE MILK PREDICTIVE MODEL
5. MANAGEMENT AND QUALITY IMPROVEMENT PROPOSALS
6. CONCLUSION & RECOMMENDATION

1. INTRODUCTION THE PROJECT UNDERSTANDING

1. High-Risk Market Expansion

- HAPPY COW Ltd. is pursuing an aggressive expansion into Bình Dương market standards , which is identified as a demanding high-end market.
- Consumers here expect premium quality and high reliability.

2. Current Quality Problem

- Production data reveals **a polarization effect**: simultaneous output of both very high and very low quality products.
- This indicates serious inconsistency and lack of **process control.**

3. Potential Market Impact

- Risk of brand damage and consumer rejection.
- Loss of trust and credibility, especially in sensitive, high-value markets.

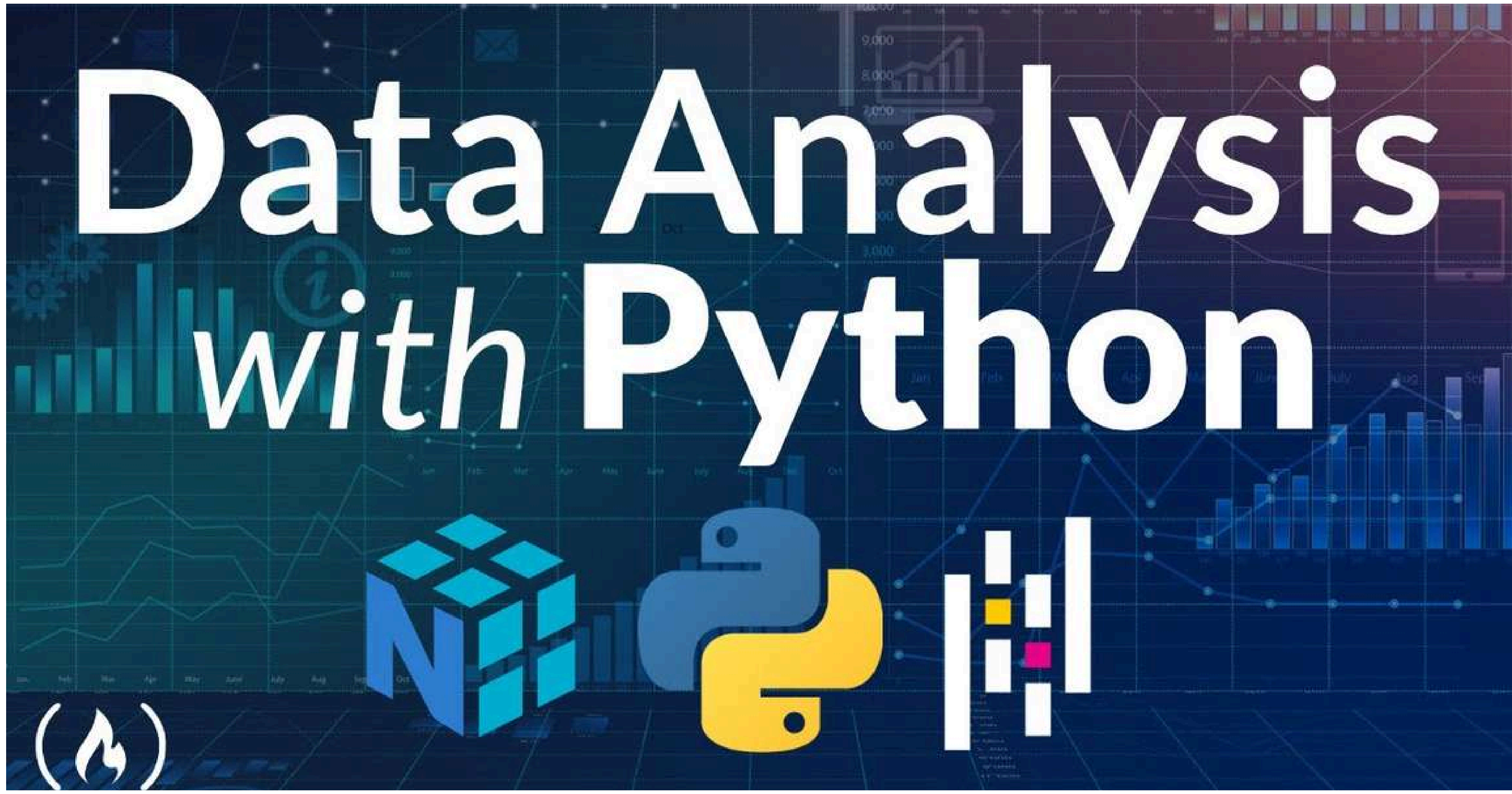
4. Urgent Need for Quality System Upgrade

- The Quality Management Team must shift from **reactive testing to proactive process control.**
- Stabilizing the process will **reduce uncontrolled variation**, ensure **consistent quality**, and meet market expectations.

2. ANALYSIS OF CURRENT QUALITY SITUATION

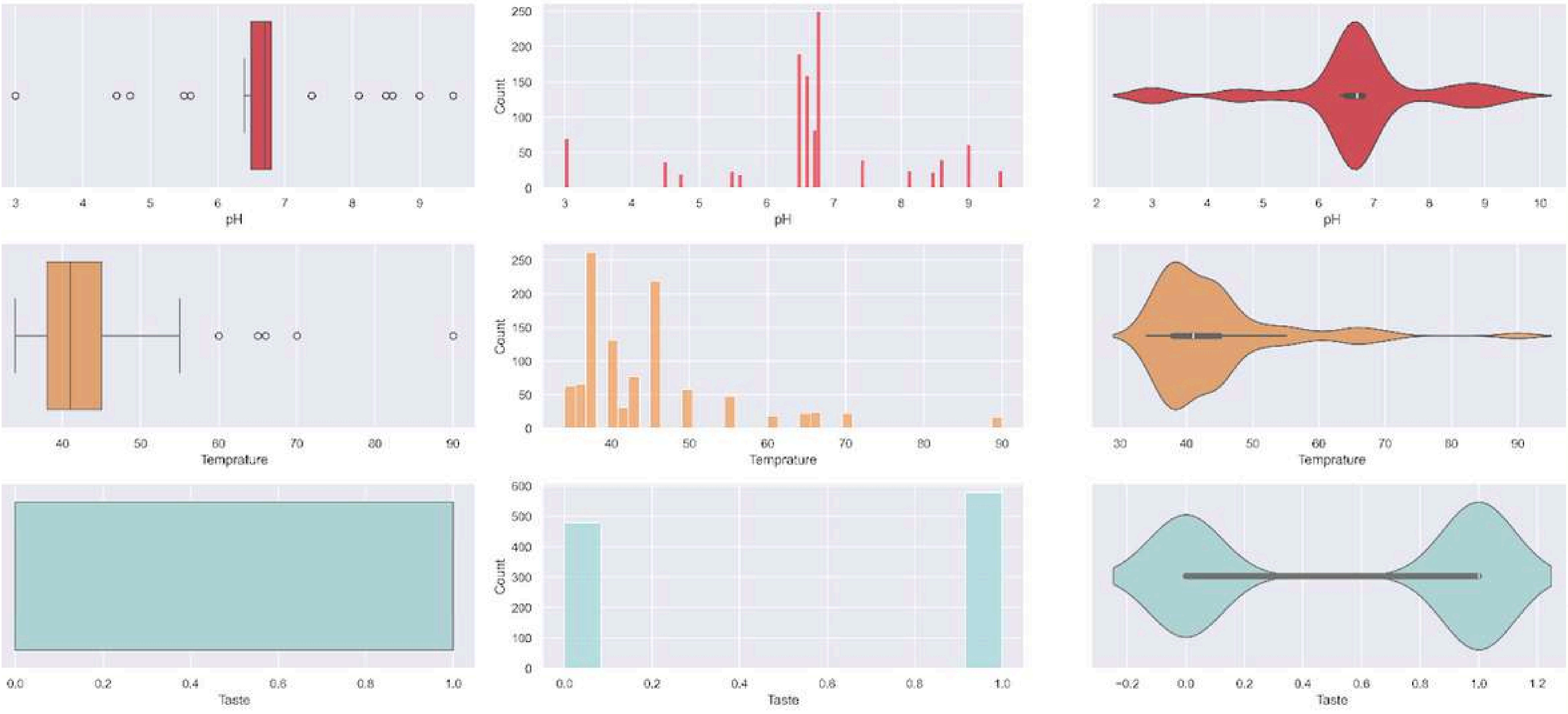
Dataset Overview

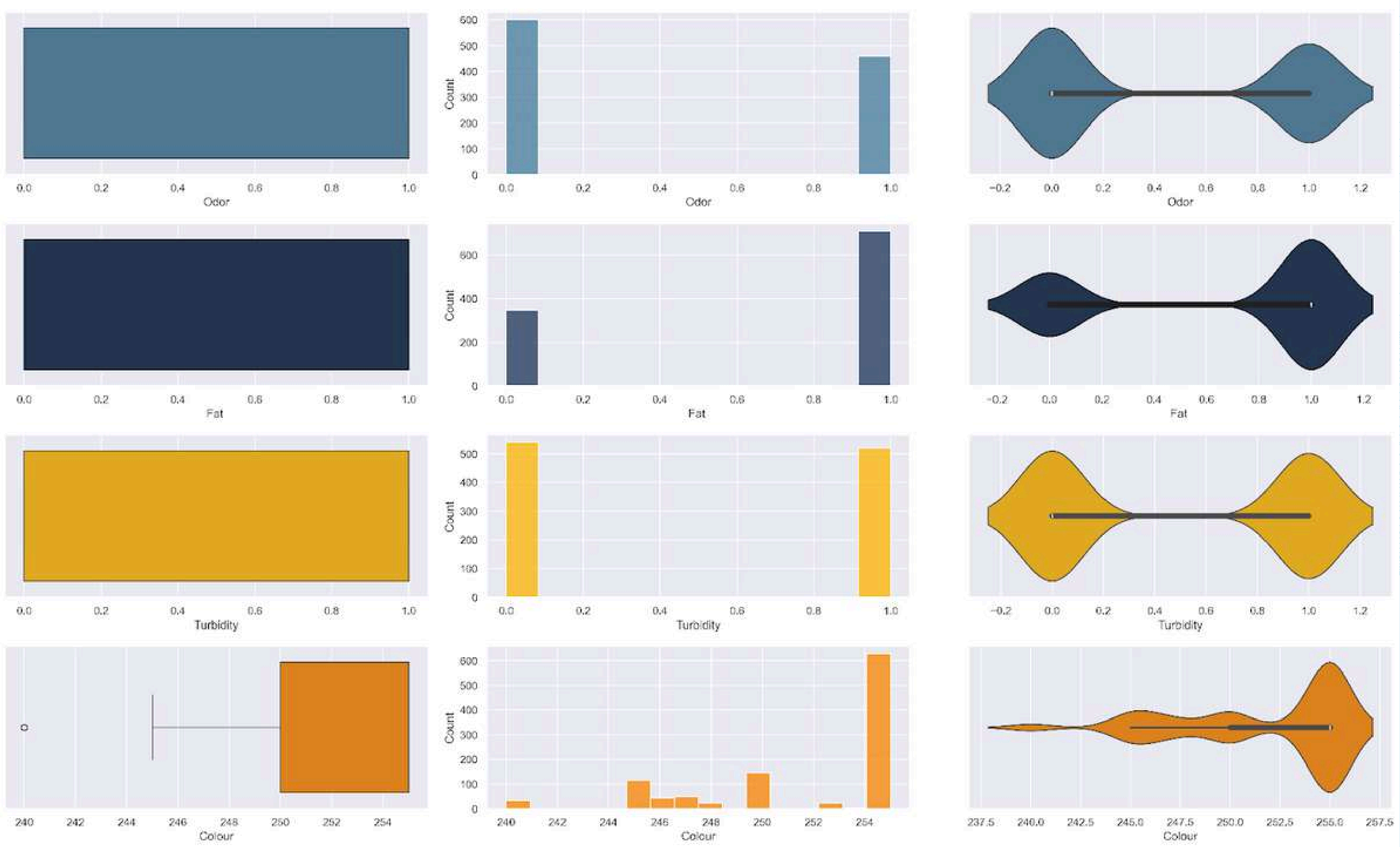
The dataset contains **1,059** records, and all observations are complete with no missing values across all **8 variables (pH, Temperature, Taste, Odor, Fat, Turbidity, Colour, and Grade).** This indicates that the dataset provided by the instructor is fully **non-null** and ready for analysis ***without requiring* any *imputation or data-cleaning for missing data.***



1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1.1. Input variables overview





- **pH and Temperature** have a wide range of values and many outliers. This means that these two factors change a lot during production, and the process control is not very stable.

- The sensory variables, such as **Taste, Smell, Fat Content, and Turbidity,** have skewed distributions. This shows that most values are concentrated in one area instead of being spread evenly.

- **Both pH and Temperature** show two peaks in their distribution. This suggests that there are differences between production batches, and the process is not consistent.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1. 2 Statistical Analysis (Descriptive Statistic)

	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
count	1059.00	1059.000	1059.000	1059.000	1059.000	1059.000	1059.000
mean	6.63	44.227	0.547	0.432	0.671	0.491	251.840
std	1.40	10.098	0.498	0.496	0.470	0.500	4.307
min	3.00	34.000	0.000	0.000	0.000	0.000	240.000
25%	6.50	38.000	0.000	0.000	0.000	0.000	250.000
50%	6.70	41.000	1.000	0.000	1.000	0.000	255.000
75%	6.80	45.000	1.000	1.000	1.000	1.000	255.000
max	9.50	90.000	1.000	1.000	1.000	1.000	255.000

The descriptive statistics show many important features about the stability of the production process and the sensory quality of milk.

The median, range, and IQR values help to see how much each variable changes, which supports the evaluation of quality control.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1. 2 Statistical Analysis (Descriptive Statistic)

	Temprature	Turbidity
count	1059.000	1059.000
mean	44.227	0.491
std	10.098	0.500
min	34.000	0.000
25%	38.000	0.000
50%	41.000	0.000
75%	45.000	1.000
max	90.000	1.000

- The average **temperature** is about 44.23°C, and it ranges from 34°C to 90°C, showing that the heating or cooling process is not stable.
- **Turbidity** also shows an imbalance.
  - About 49.1% of samples are turbid, and 50.9% are clear. This means the two groups are almost equal, so the distribution is nearly split in half.
  - This may indicate problems with the filtration system or possible microbiological contamination in some batches.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1. 2 Statistical Analysis (Descriptive Statistic)

	Taste	Odor	Fat
count	1059.000	1059.000	1059.000
mean	0.547	0.432	0.671
std	0.498	0.496	0.470
min	0.000	0.000	0.000
25%	0.000	0.000	0.000
50%	1.000	0.000	1.000
75%	1.000	1.000	1.000
max	1.000	1.000	1.000

- Most samples have a good **fat content**, with an average value of 0.671.
- **Taste** shows moderate stability, with about 54.7% of samples having a good taste.
    - May come from changes in fat quality or microbial activity during storage.
- **Odor** is not as good. The average is 0.432, which means 43.2% good odor and 56.8% bad odor.
    - Caused by poor handling, contamination, or bad storage of raw materials.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1. 2 Statistical Analysis (Descriptive Statistic)

	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
count	1059.00	1059.000	1059.000	1059.000	1059.000	1059.000	1059.000
mean	6.63	44.227	0.547	0.432	0.671	0.491	251.840
std	1.40	10.098	0.498	0.496	0.470	0.500	4.307
min	3.00	34.000	0.000	0.000	0.000	0.000	240.000
25%	6.50	38.000	0.000	0.000	0.000	0.000	250.000
50%	6.70	41.000	1.000	0.000	1.000	0.000	255.000
75%	6.80	45.000	1.000	1.000	1.000	1.000	255.000
max	9.50	90.000	1.000	1.000	1.000	1.000	255.000

Overall, four variables are the most important for milk quality:

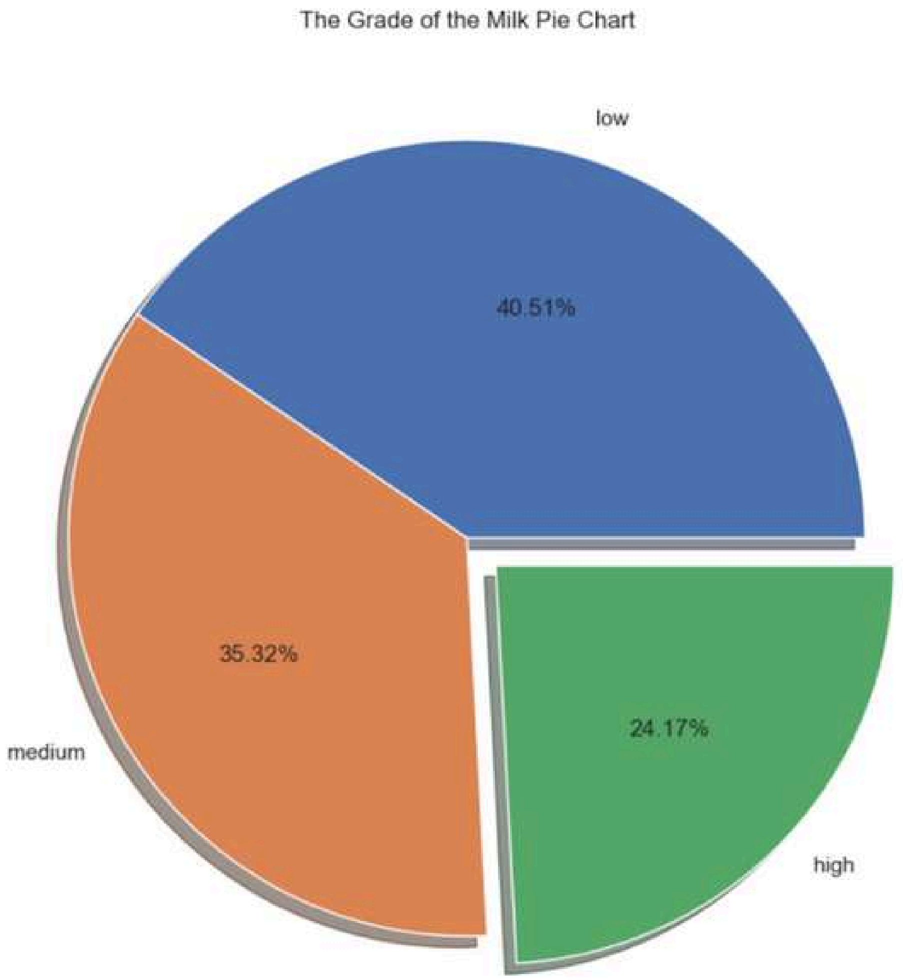
- Temperature.
- Fat Content.
- Odor.
- Turbidity.

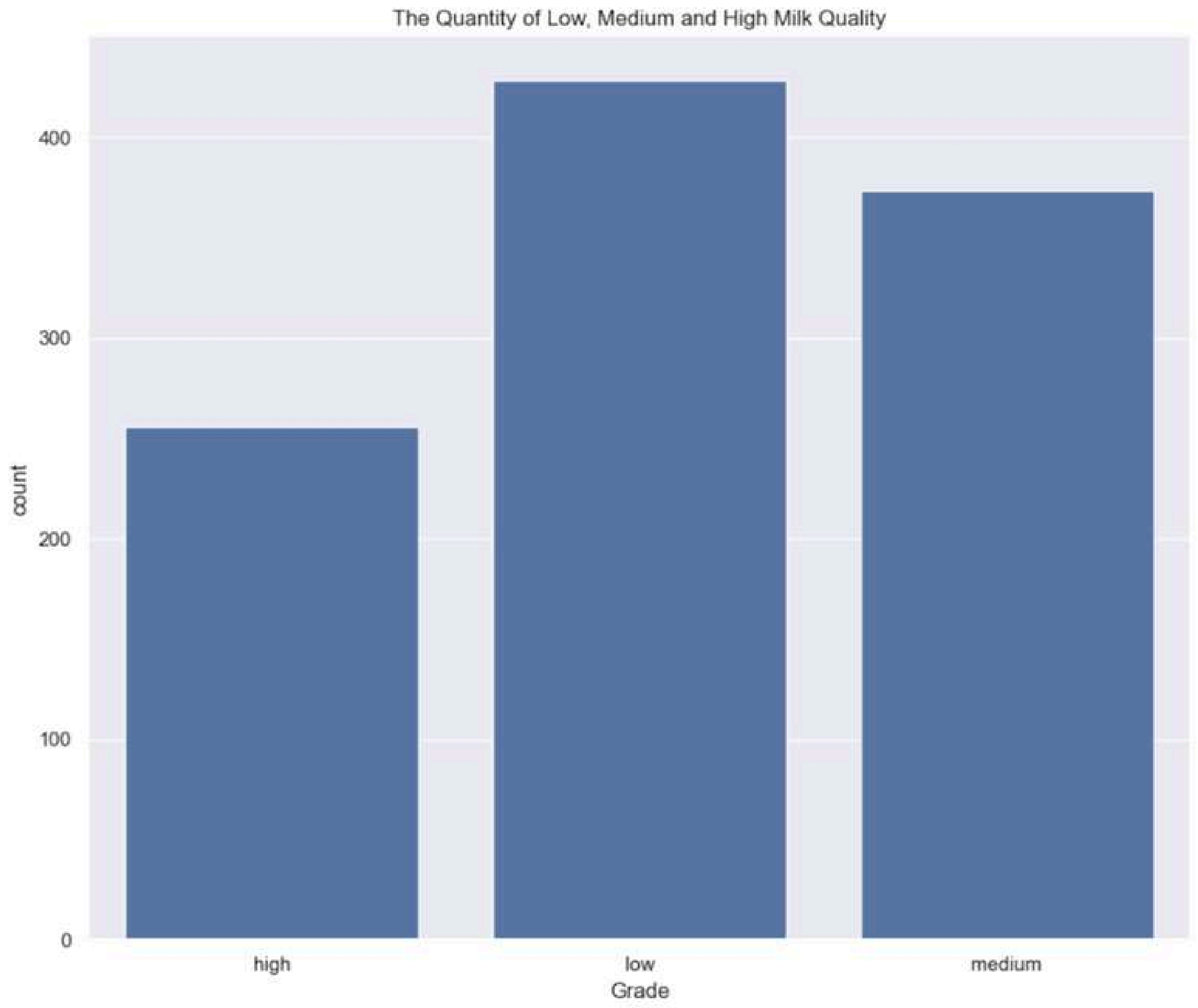
The other three variables have weaker effects.

- **pH,** the range is wide (from 3.0 to 9.5), but most samples are between 6.5 and 6.8.
- **Color** is very stable, with values between 240 and 255, showing almost no variation. So color is not a concern for quality control.
- Although **taste** is important for consumers, it is often affected by other technical factors like fat content, odor, or turbidity.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1.3. Distribution of the target variable (Grade)





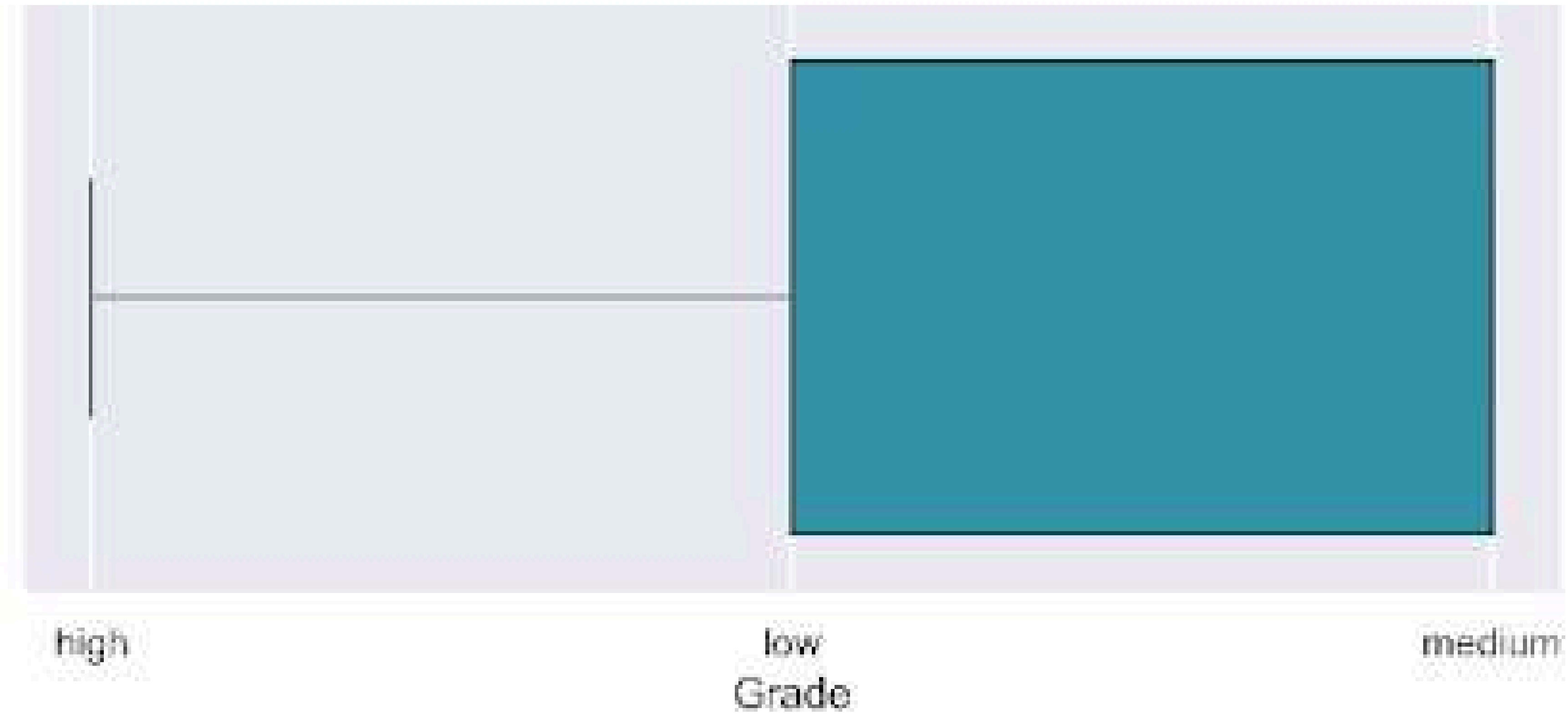
- **Low-quality milk** has the highest number of samples, with 439 samples, which is about 40.5% of the total.
- **The average-quality group** is also large, with 360 samples.
- **The high-quality group** has only 227 samples, which is 24.2% of the total.

The product mainly produces this low-quality group of products instead of high-quality products.

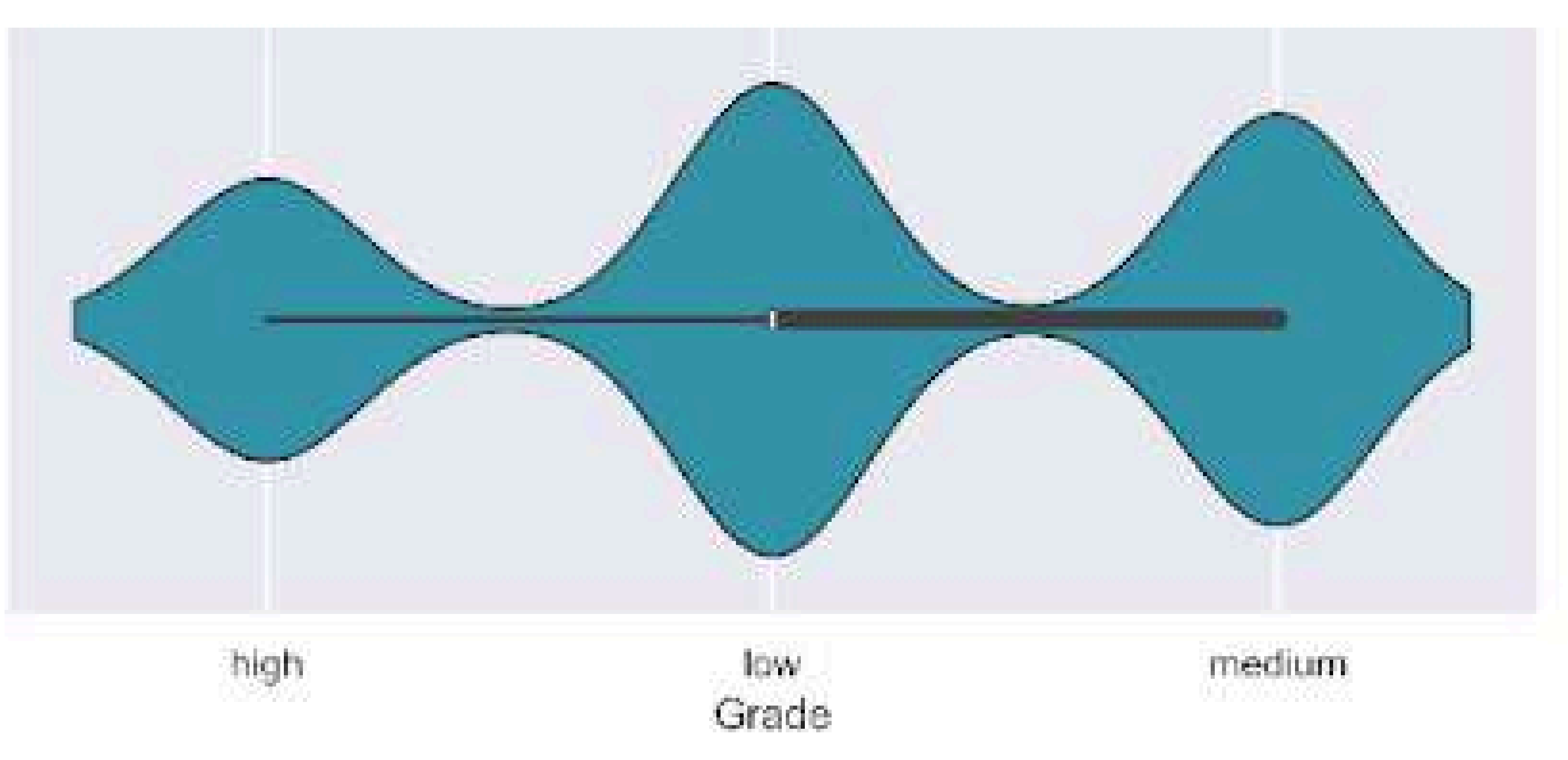
1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1.3. Distribution of the target variable (Grade)

The box plots



The violin plots



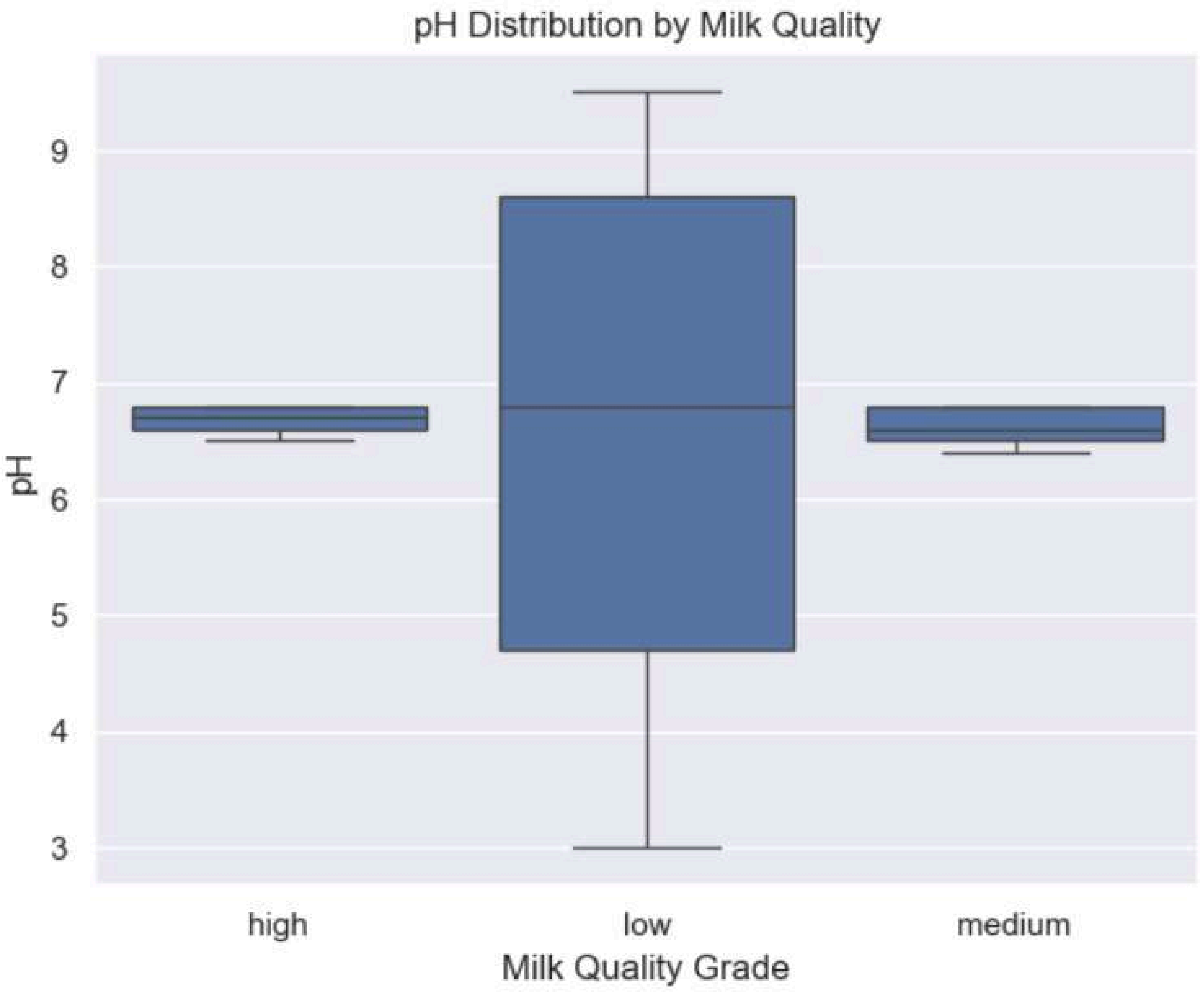
For **low and average-quality milk**, the variability is **much higher**. This suggests that the production process has some inconsistencies or biases, making it unstable.

In contrast, **the high-quality milk group** shows less variation, indicating the process is **more stable and consistent when it works well**.

EIU
TRƯỜNG ĐẠI HỌC QUỐC TẾ MIỀN ĐÔNG
EASTERN INTERNATIONAL UNIVERSITY

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1.4.Characteristics of the Three Milk Quality Groups.



High-quality group, the pH is very stable, staying between 6.65 and 6.75, with almost no unusual values.
=> The milk has a stable chemical environment, not much affected by microorganisms or chemical reactions.

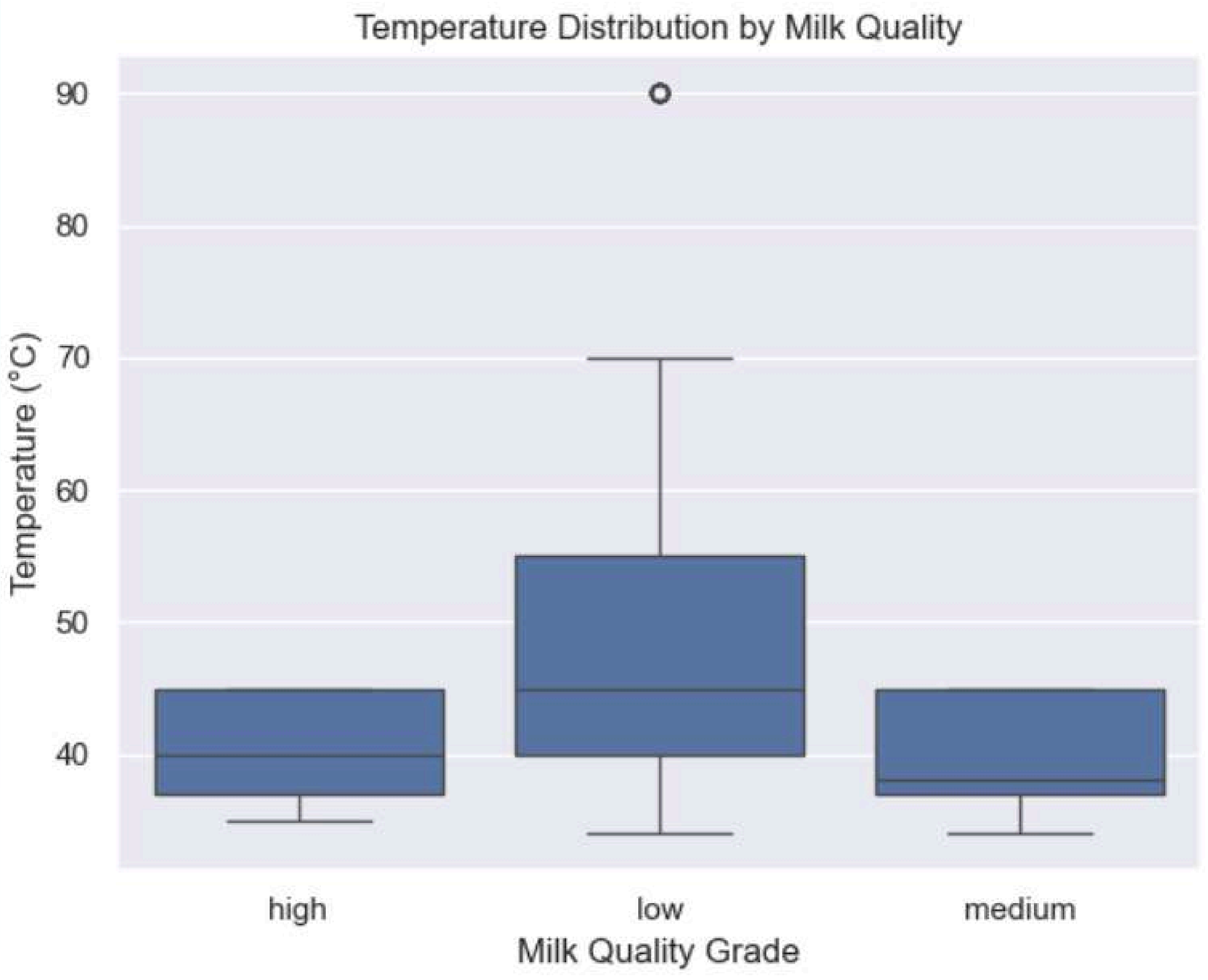
Medium-quality group, the pH is also similar, between 6.5 and 6.8, but the values are more spread out.
=> The process is not completely stable yet.

Low-quality group, the pH changes a lot from 3.0 to 9.5, which is far outside the normal range for milk.
=> Poor control during production, maybe due to bad storage, contamination, or chemical mistakes. Such large changes make the milk quality worse and show that the materials or equipment may not be stable.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

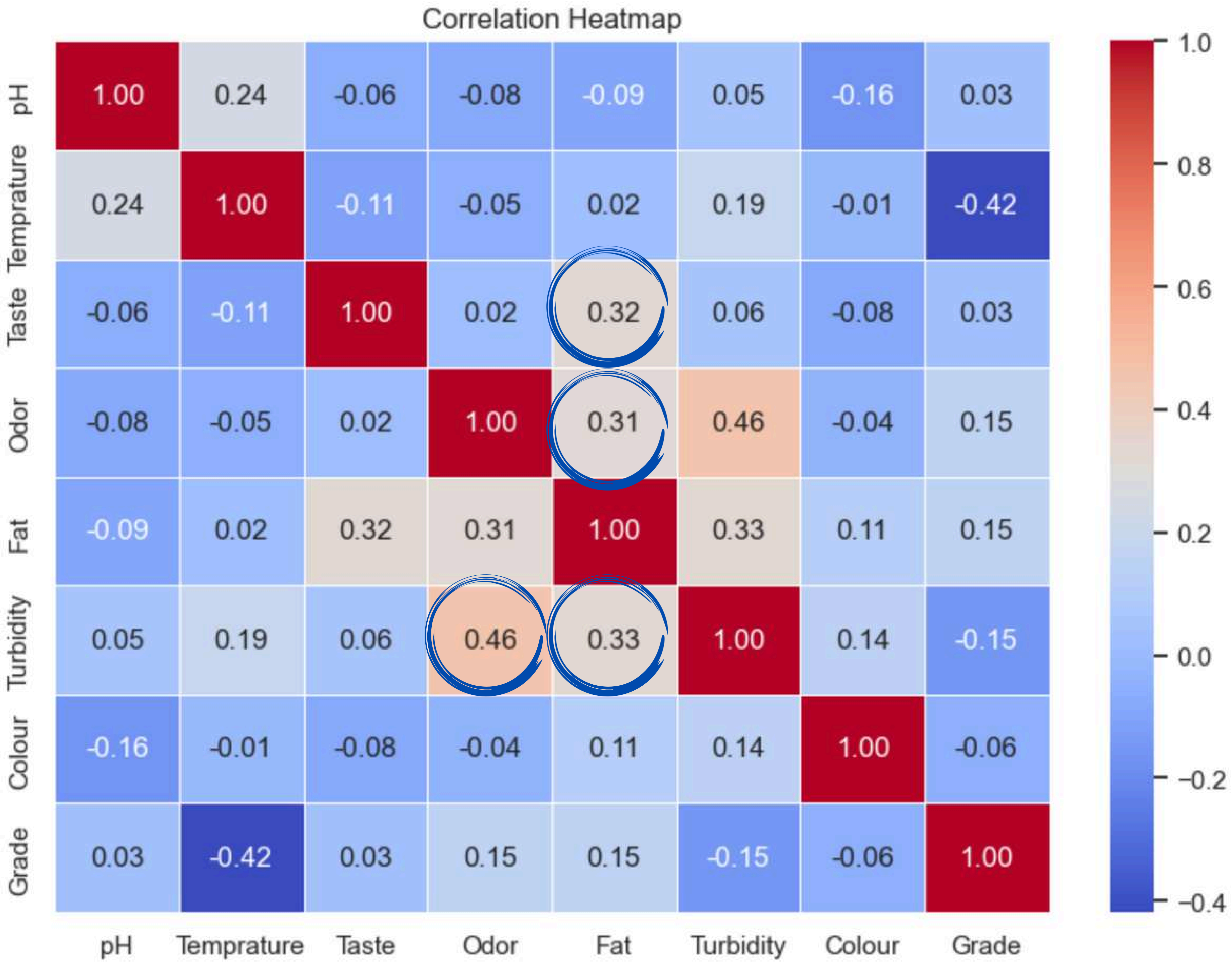
1.4.Characteristics of the Three Milk Quality Groups.

- **High-quality group** has very stable temperatures, which are in the right range for heating and sterilization.
- **Medium-quality group** is also similar, but it shows a bit more variation. This means the process is less uniform between batches.
- **Low-quality group** shows the biggest temperature variation and even some outliers.
  - These very high temperatures can cause protein damage, mineral imbalance, and loss of milk structure.
  - There may be equipment problems, or workers may not follow the correct technical process.



1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1.5. Correlation Analysis

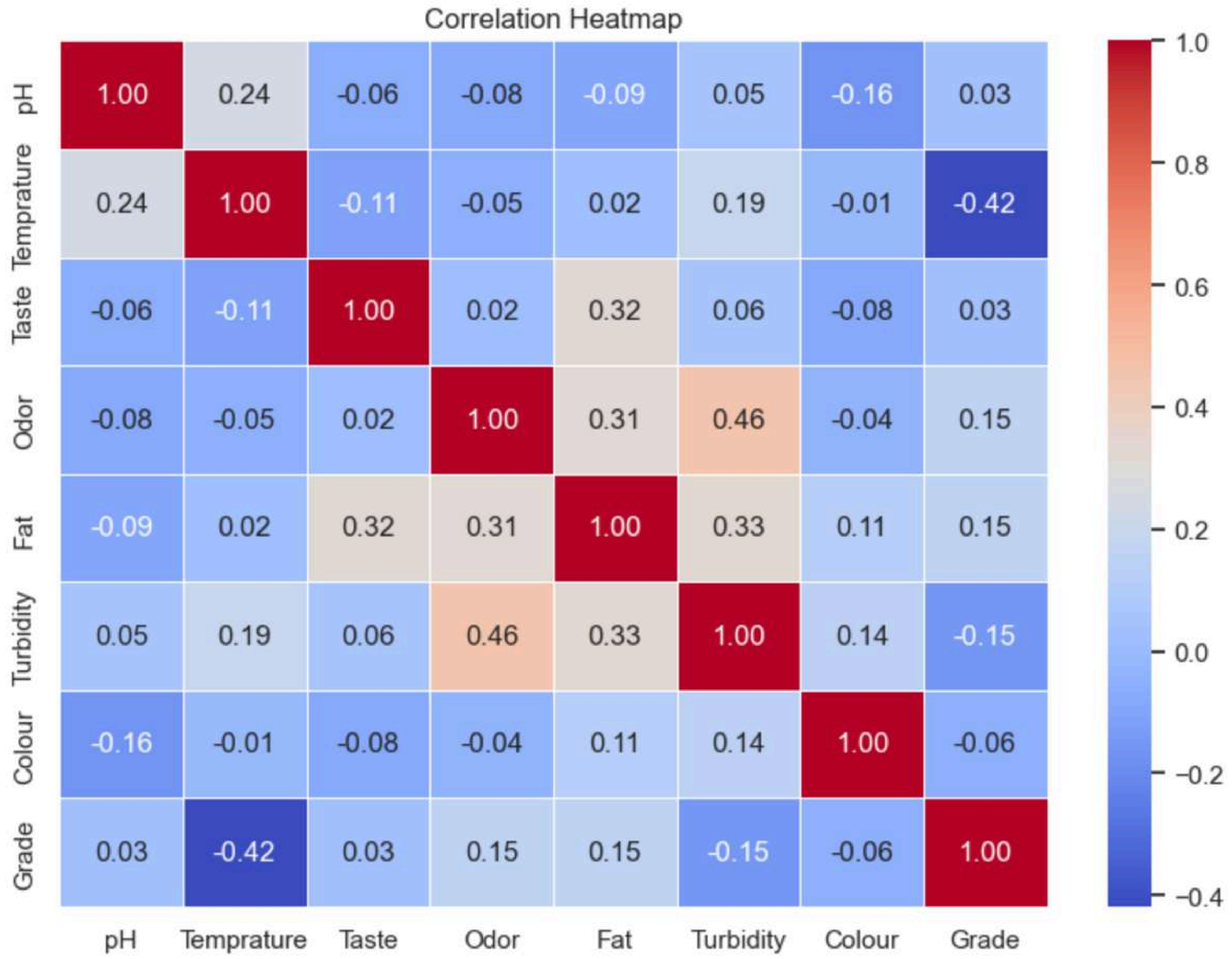


- The **strongest correlation** is between **odor and turbidity,** with a correlation value of r = 0.46.
  - Milk samples with higher turbidity are more likely to have a bad odor.
  - This can happen because of problems in filtration, contamination, or microbiological issues during production.

- **Fat content** also has a **moderate correlation** with **taste** (r = 0.32). It is also related to **odor** (r = 0.31) and **turbidity** (r = 0.33).
  - Even though correlation does not mean causation, it still shows that some variables change together.
  - For example, fat content, odor, and turbidity move together, so they may be affected by similar production factors.

1. OVERVIEW AND DESCRIPTIVE DATA ANALYSIS OF MILK QUALITY DATA

1.5. Correlation Analysis

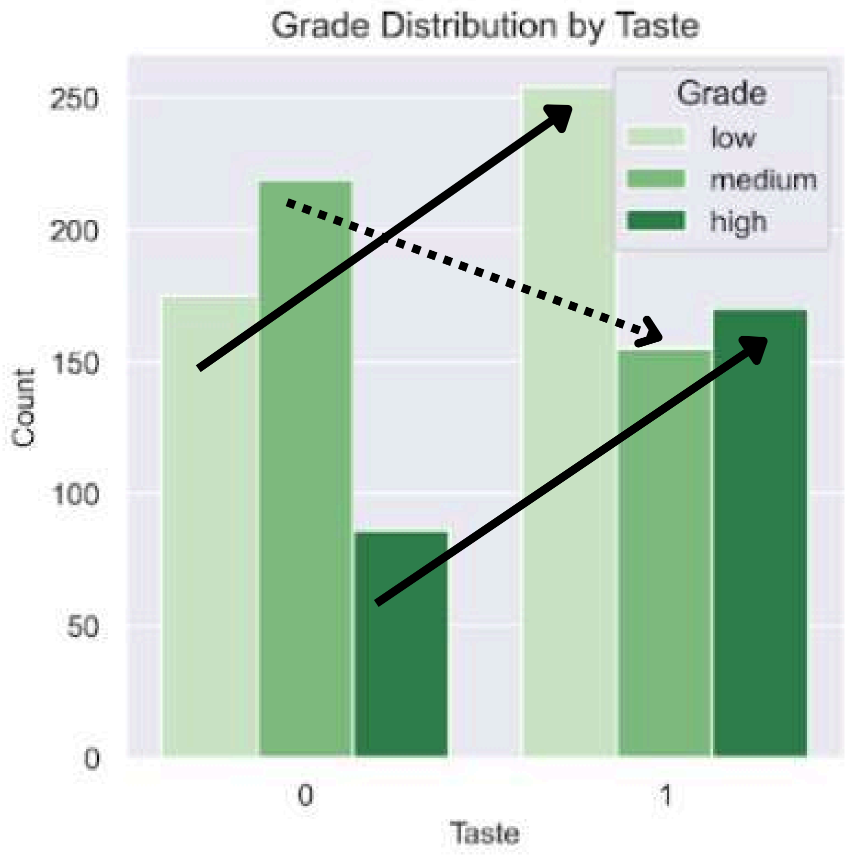


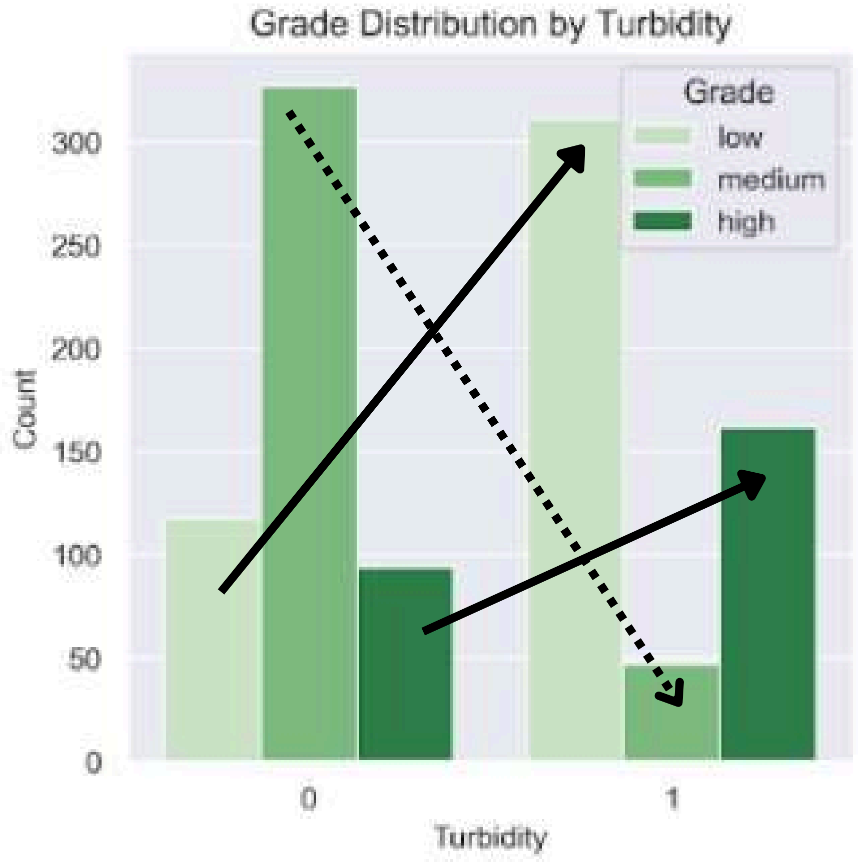
- pH, color, and taste have very weak correlations with the overall quality (less than |r| = 0.1).

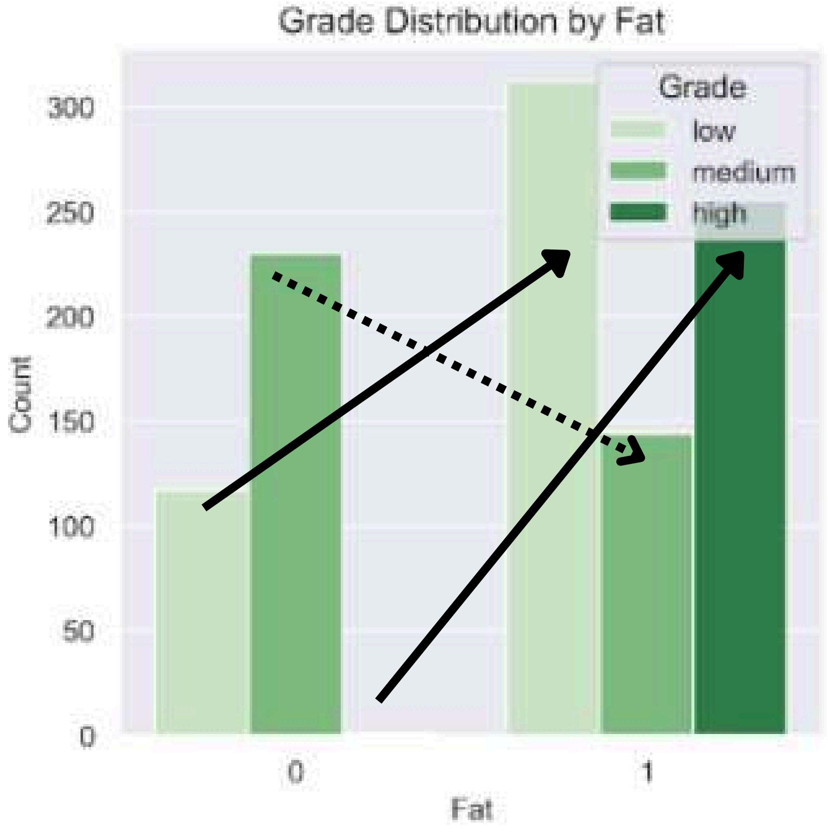
These variables are either well-controlled during production, or they are not very important in deciding the final quality grade.

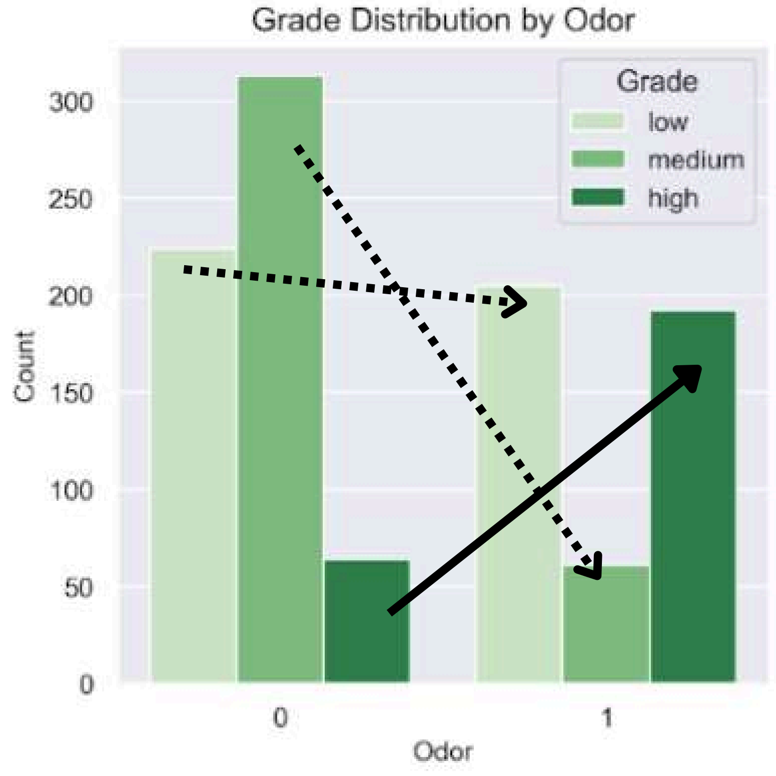
2. PROCESS PERFORMANCE EVALUATION

2.1.Binary variables





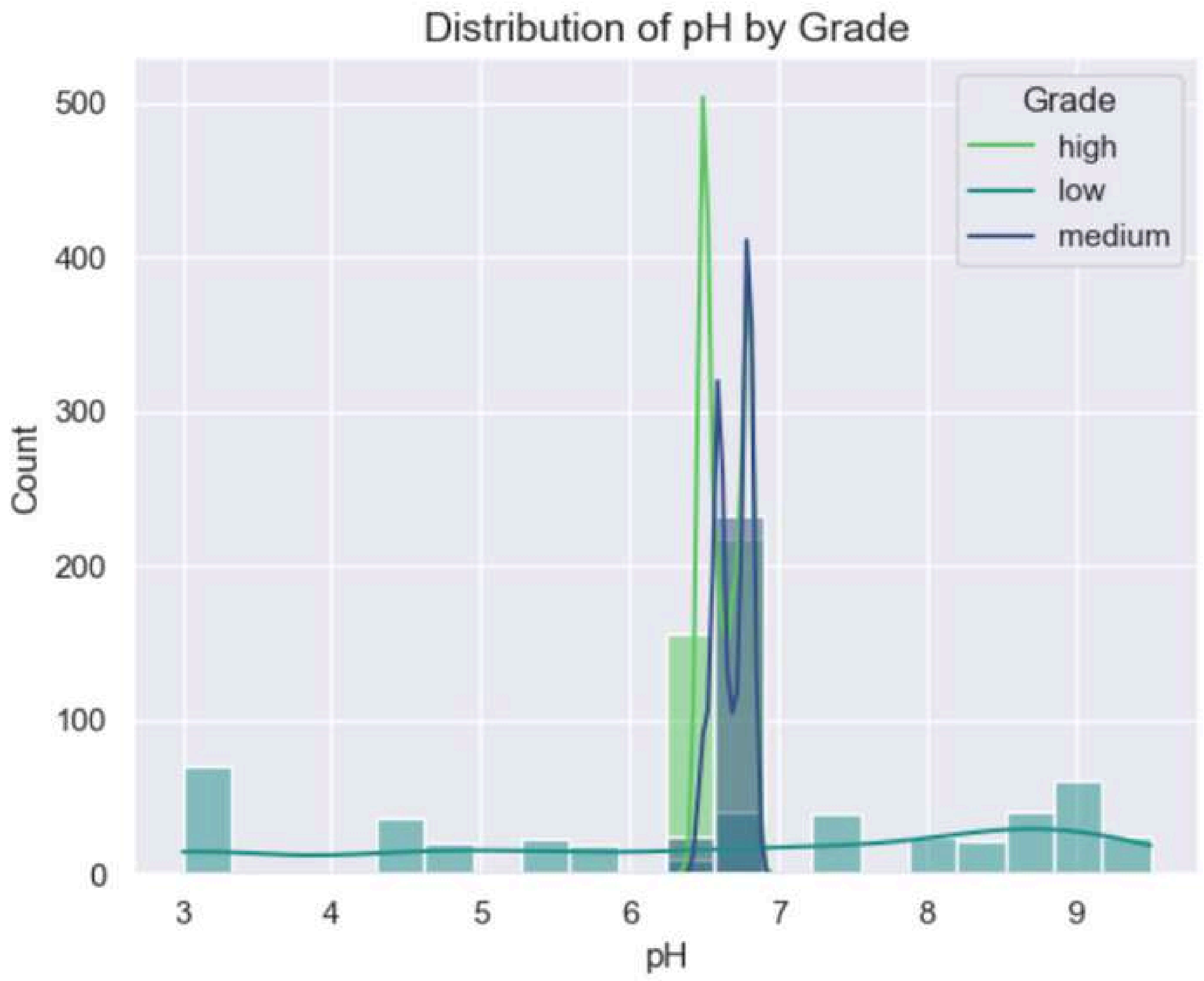


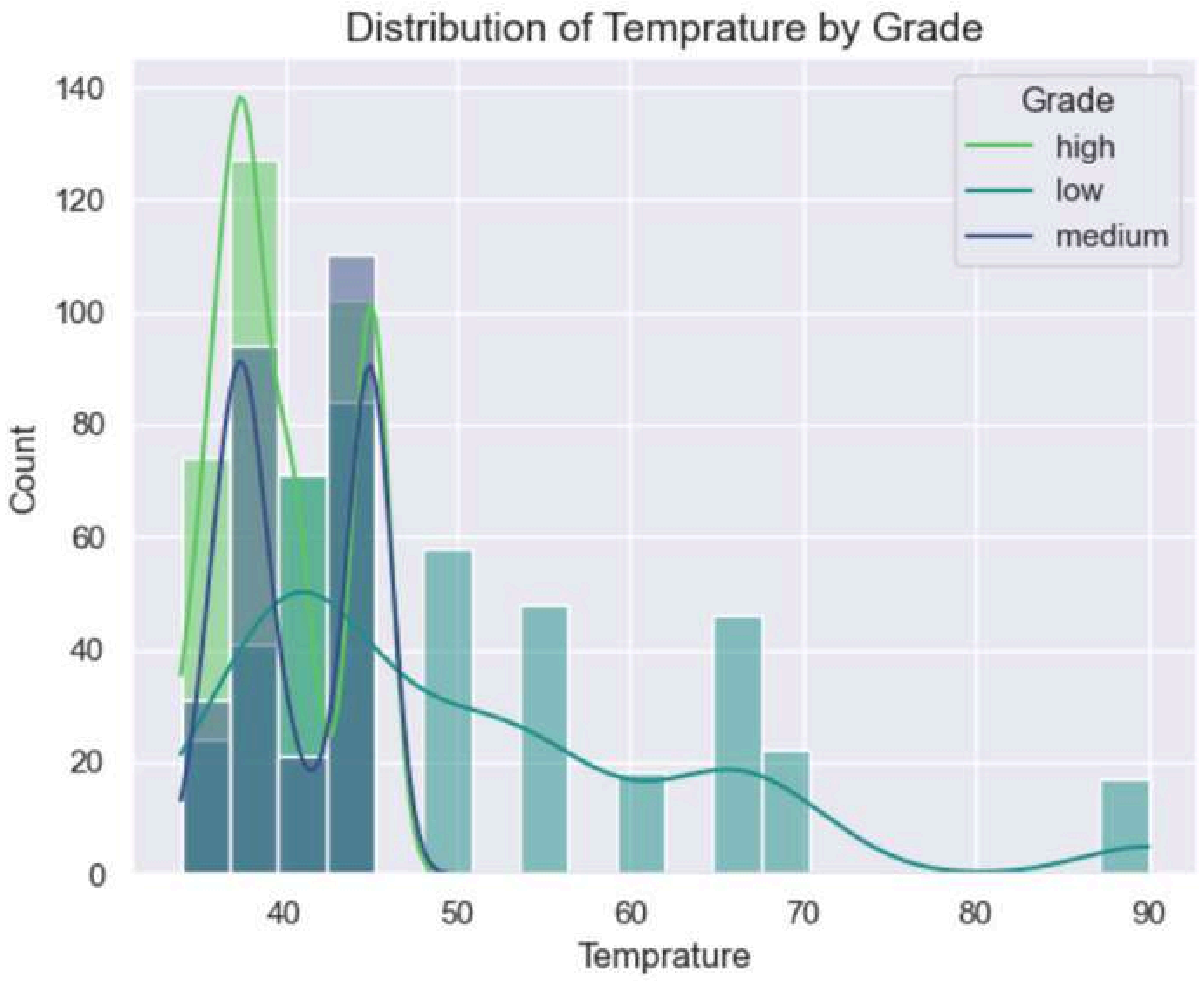


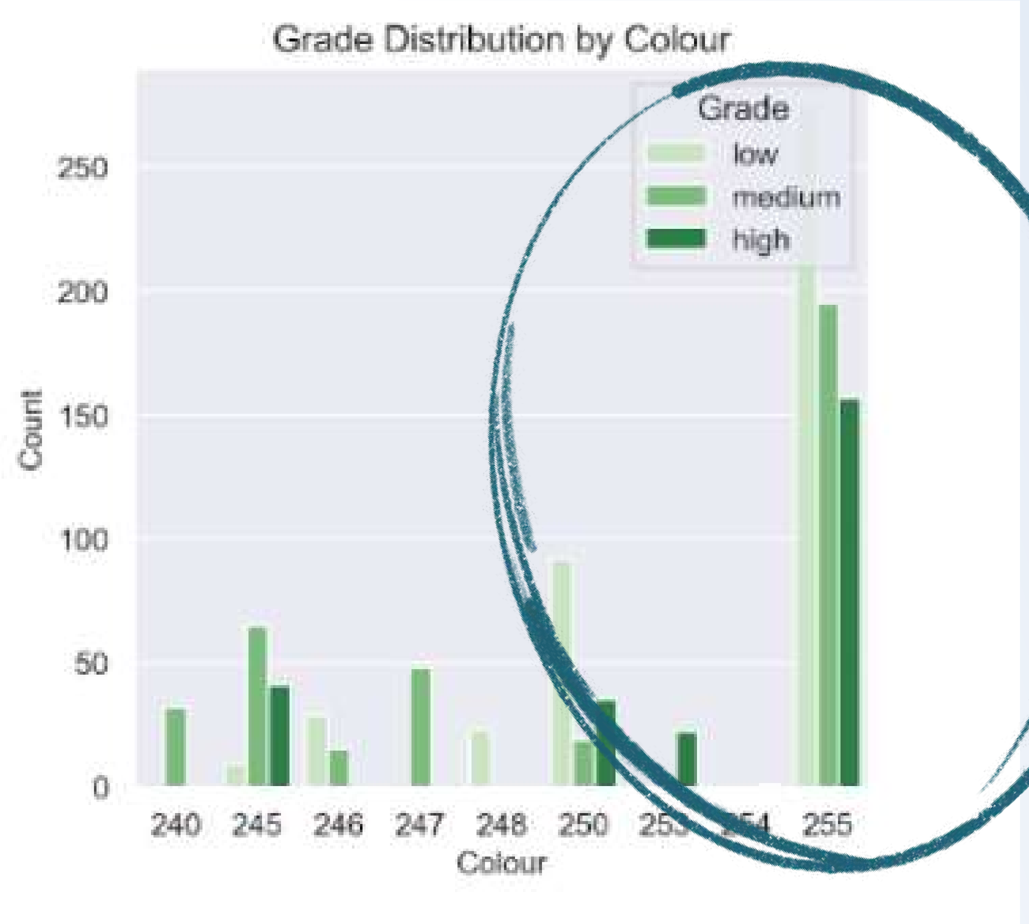
- Grade distributions show clear polarization, with many products rated High and Low, but very few rated Medium.
- This pattern reflects an unstable production process and significant variation between batches.
- Likely caused by uneven processing or storage conditions, leading some batches to perform extremely well while others fall sharply in quality.
- That is the <u>noticeable polarization effect.</u>
- The results highlight the need to standardize and stabilize the manufacturing process to ensure more consistent, predictable product quality.

PROCESS VARIATION ON TEMPERATURE AND PH

2.2 Continuous Variables Analysis (Temperature, pH, Colour)







- pH: Has minimal impact; all grades cluster tightly (6.5–6.8). It is not a major distinguishing factor due to regulation.
- Temperature: The most important quality indicator; shows clear differentiation.
- High Quality: Produced at lower temperatures (38°C - 42°C).
- Low Quality: Spread across a vast range (up to 80°C -90°C), confirming poor control.
- Colour: Indicates a process split. High quality is in the upper range (254–255); low quality is in the lower range (240–248), suggesting defects (overheating/oxidation).

Assessment: Does the current quality meet the higher standardization requirements for market expansion?

- Conclusion: NO. The manufacturing system has Systemic Instability and Inadequate Control.
- Critical Imbalance: Output is dominated by approximately 40% low-quality milk versus only approximately 24% high-quality milk, failing to meet premium market standards.
- Polarization Effect: High variability in critical factors (Temperature, Odor, Turbidity, Fat) creates polarization: batches are either unmistakably high or clearly low quality, preventing reliable medium-grade production.

→ Action Needed: Substantial improvements in process control, monitoring, and standardization are essential before pursuing demanding market expansion.

3. IDENTIFY THE CURRENT SITUATION AND SPECIFICATION

A. THE HYPOTHESIS TESTING ON VARIABLE INDEPENDENCE

The hypothesis testing phase aims to explain the statistical relationship between the variables related to milk quality: sensory factors, physical characteristics, and process aspects.
Using a dataset with both categorical and numeric variables, statistical tests including Chi-Square, ANOVA, and Pearson were conducted

Rationale for Using Multiple Tests:

- **Chi-Square tests** assess dependency among **categorical** sensory and physical variables.
- **ANOVA** evaluates whether **continuous variables** differ significantly across Grade levels.
- **Pearson correlation** identifies linear associations among **continuous variables.**

	Variable 1	Variable 2	p-value	Conclusion	Test
0	Taste	Turbidity	7.966269e-02	Independent	Chi-Square
1	Odor	Taste	6.101446e-01	Independent	Chi-Square
2	Odor	Turbidity	8.049347e-50	NOT independent	Chi-Square
3	Fat	Taste	1.035843e-25	NOT independent	Chi-Square
4	Fat	Odor	2.736139e-24	NOT independent	Chi-Square
5	Fat	Turbidity	1.752000e-26	NOT independent	Chi-Square
6	Fat	Grade	8.519530e-59	NOT independent	Chi-Square
7	Grade	Taste	2.532332e-10	NOT independent	Chi-Square
8	Grade	Odor	2.363640e-48	NOT independent	Chi-Square
9	Grade	Turbidity	5.699600e-69	NOT independent	Chi-Square
10	NaN	NaN	6.405242e-01	Independent	ANOVA
11	NaN	NaN	9.787957e-66	NOT independent	ANOVA
12	NaN	NaN	1.327858e-16	NOT independent	ANOVA
13	pH	Temprature	6.679096e-16	NOT independent	Pearson
14	pH	Colour	7.214234e-08	NOT independent	Pearson
15	Temprature	Colour	7.820435e-01	Independent	Pearson

A. THE HYPOTHESIS TESTING ON VARIABLE INDEPENDENCE

Results of Hypothesis Testing

1. Chi-Square Tests for Categorical Variables

- Strong statistical dependence among sensory & physical variables.
- Significant links:

Odor ↔ Turbidity ($p \approx 8.01\times10^{-5}$)

Fat ↔ Taste ($p \approx 1.03\times10^{-25}$)

Fat ↔ Odor ($p \approx 2.73\times10^{-24}$)

Fat ↔ Turbidity ($p \approx 1.75\times10^{-26}$)

- Grade depends on Taste, Odor, Fat, Turbidity.
- Defects tend to occur together: when one characteristic deteriorates, others often deteriorate as well.
- Low Fat often accompanies defective Odor and Taste.
- Taste ↔ Turbidity: statistically independent **($p \approx 0.079$).**

→ Indicates no direct link between poor taste and cloudiness.

- Grade’s dependence on multiple variables reinforces their importance in defining overall product quality.

A. THE HYPOTHESIS TESTING ON VARIABLE INDEPENDENCE

Results of Hypothesis Testing

2. ANOVA for Continuous Variables Across Grade Levels

Significant Differences Across Grades

- Temperature varies significantly across Grade levels.
- Colour also shows significant variation.

→ Both are important process-related contributors to milk quality.

Non-Significant Variable

- pH does NOT differ across grades (**p ≈ 0.641**).

→ pH is not a direct driver of Grade classification in this dataset.

Results of Hypothesis Testing

3. Pearson Correlation Analysis

Strong Linear Relationships

- pH ↔ Temperature ($p \approx 6.68 \times 10^{-16}$)
- pH ↔ Colour ($p \approx 7.21 \times 10^{-8}$)

→ Indicates strong associations between acidity and these process variables.

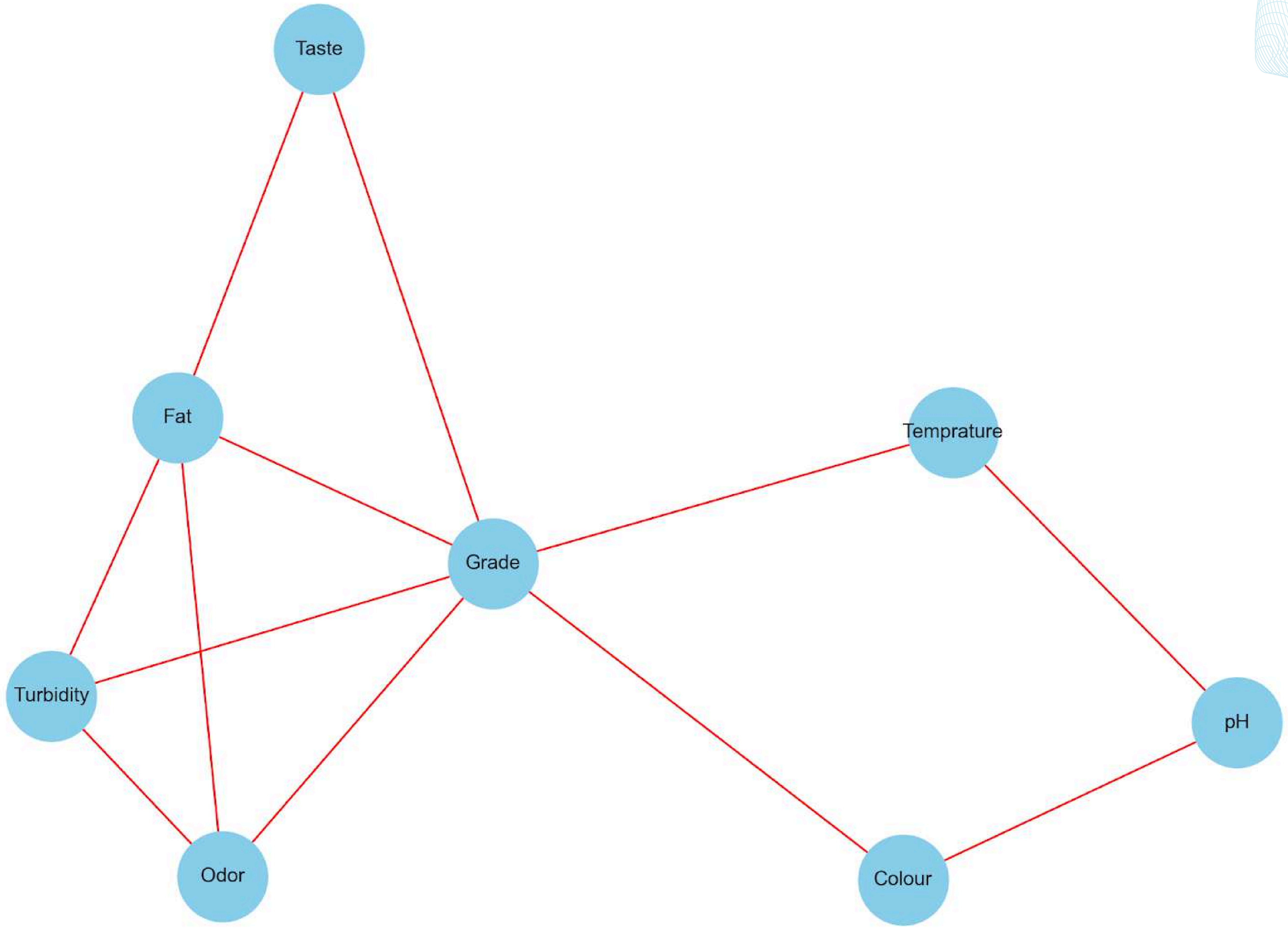
Independent Variables

- Temperature ↔ Colour are statistically independent

($p \approx 0.782$).

→ They influence quality through separate mechanisms, not a shared source.

Variable Dependence Network Graph



A. THE HYPOTHESIS TESTING ON VARIABLE INDEPENDENCE

Critical-to-Quality (CTQ) Variables

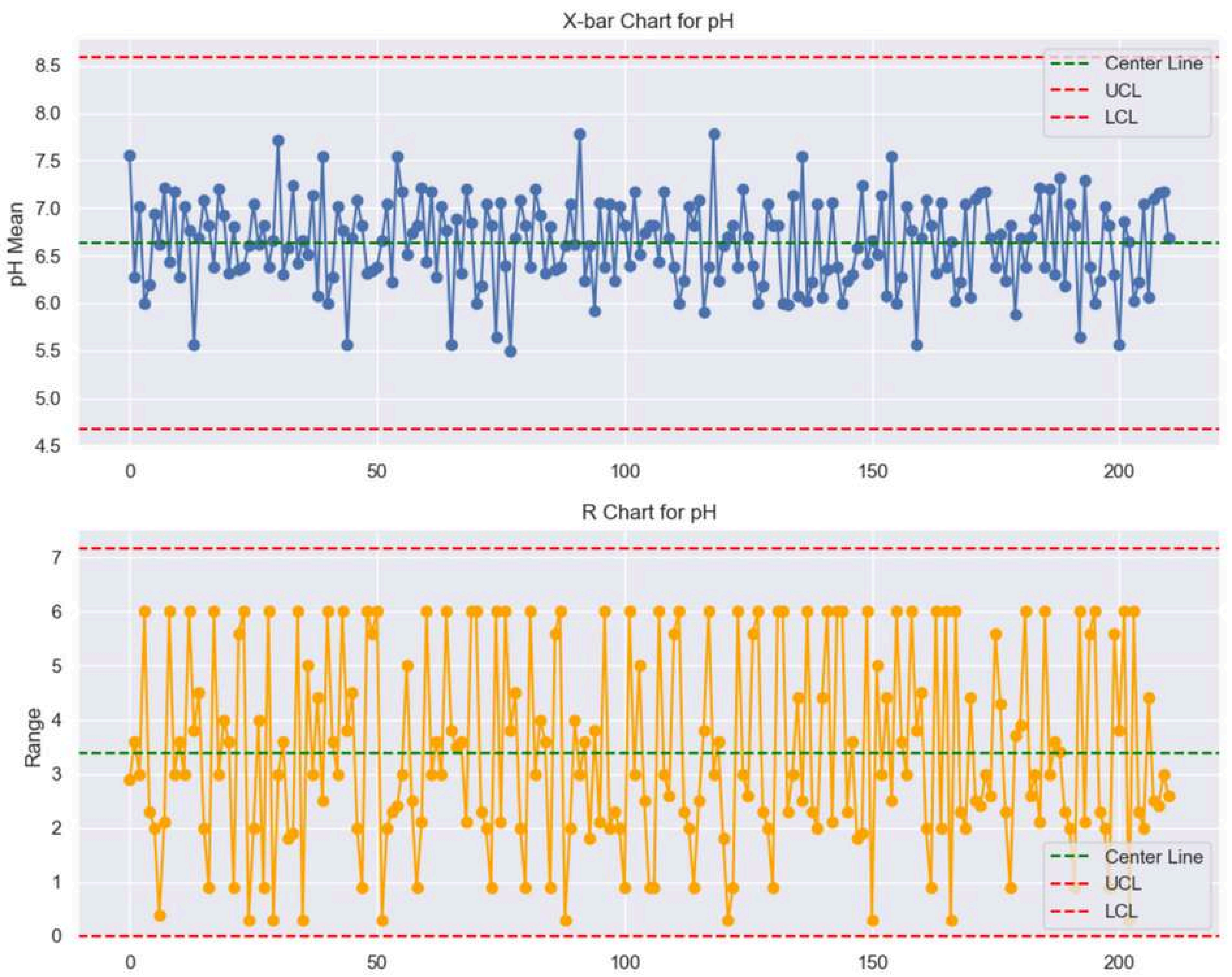
- Based on Chi-square, ANOVA, and Pearson analyses, several variables emerged as CTQs contributing the strongest to the quality classification process.
- Basic CTQs: Odor, Fat, Turbidity, Temperature. All show remarkably high statistical correlations with Grade and require high quality management.
- Secondary CTQs: Taste, Colour moderately associated with Grade and still important in defining product quality levels.
- Non-CTQ: pH - despite correlations with Temperature and Colour, pH does not define distinct Grades and its effect arises indirectly.

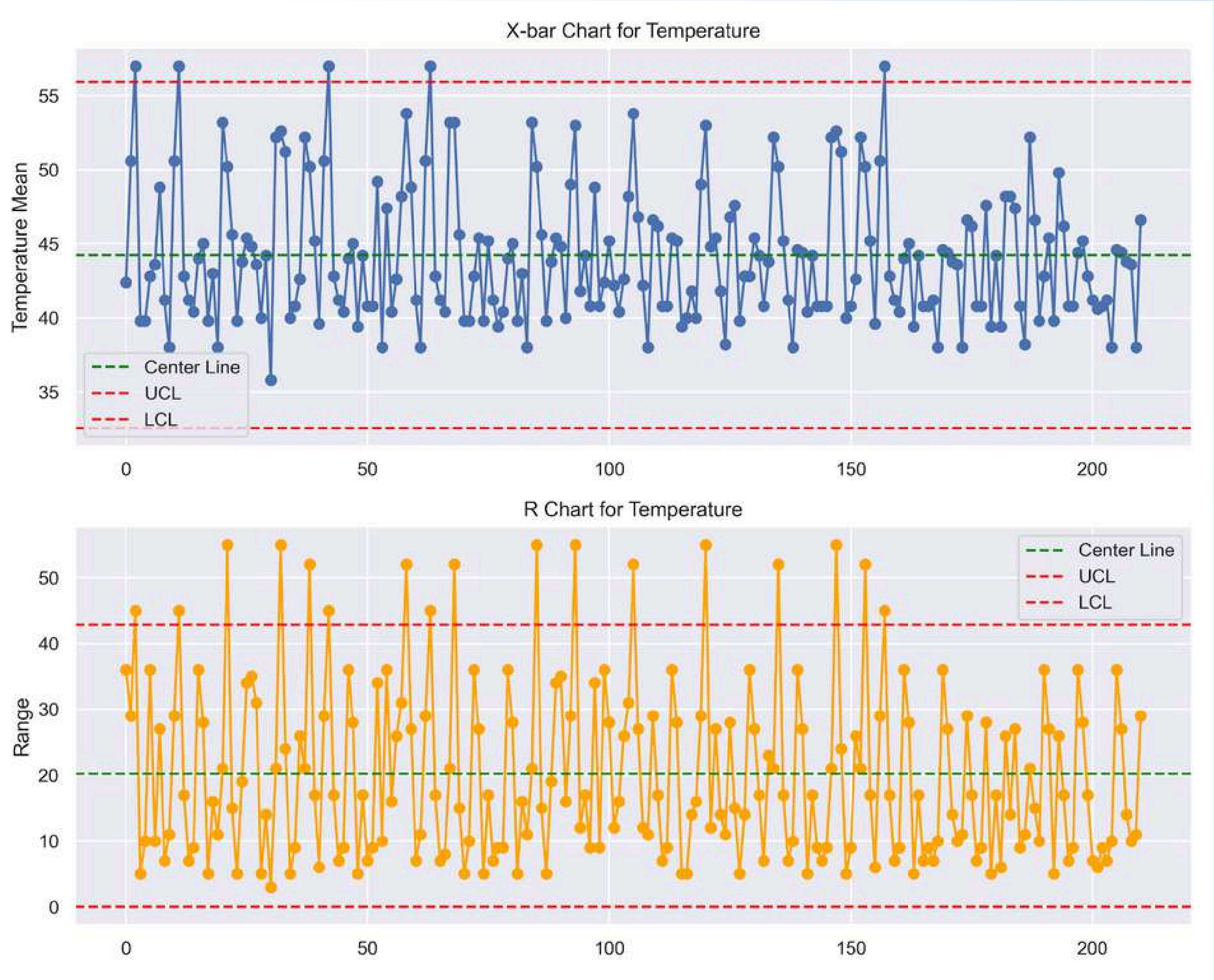
Failure Pathways

- Process Variation Path: **pH → Temperature → Colour → Grade** . Temperature changes trigger changes in pH and Colour, leading to Grade degradation; indicates malfunctioning equipment, uncontrolled temperatures, and unmanaged operations.
- Sensory Defect Path: **Odor → Turbidity → Fat → Taste → Grade** . Poor Odor leads to increased Turbidity, unstable Fat, and impaired Taste, resulting in lower Grades; reflects issues with filtration and raw material quality.
- Together, these pathways reveal the complex aspects of quality failure and the need for process stabilization and better management of quality characteristics.

B. PROCESS VARIATION ON TEMPERATURE AND PH

Evaluation of Current Process Stability

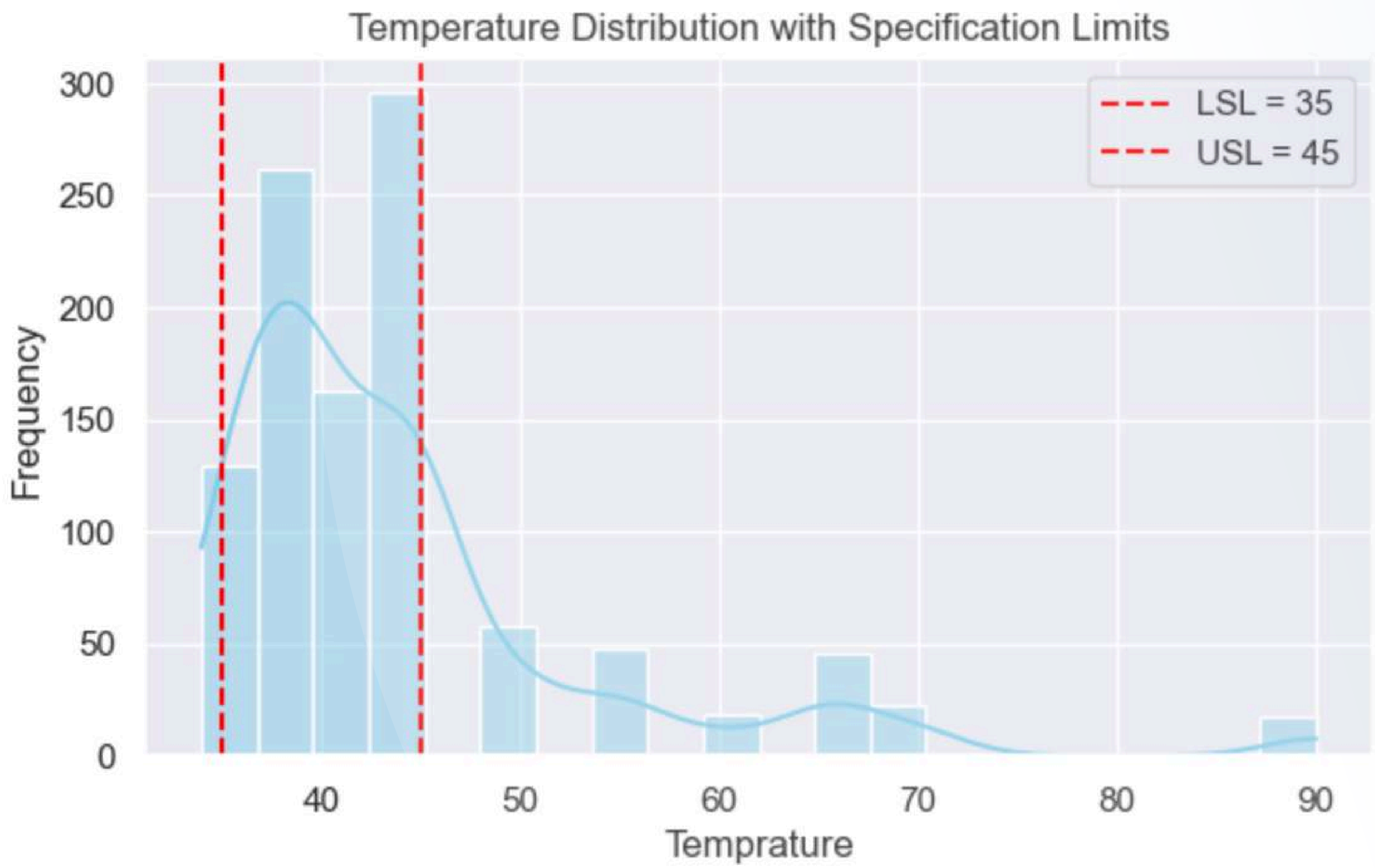




- The X-bar chart shows subgroup means fluctuating widely from 36°C to 57°C, frequently touching or exceeding the UCL.
- This indicates large instability and absence of control in the pasteurization temperature.
- In a stable process, points should remain close to the center line and within limits—this is not observed.
- The R chart also displays excessive variation, with ranges repeatedly approaching the upper control limit.
- Both charts confirm the presence of special-cause variation, likely due to inconsistent heating, operator error, or equipment malfunction.
- As a result, the process cannot consistently maintain the required 38–42°C high-quality range.

B. PROCESS VARIATION ON TEMPERATURE AND PH

Specification Setting for High-Grade Milk



Justification for Spec Limits (38–42°C):

- High-grade milk ranges **35°C–45°C**, but distribution is not uniform.
- Highest-density cluster is **38°C–42°C**
- values at **35–37°C** and **43–45°C** are low-density tails caused by process noise or measurement variation.

→ Therefore, specification limits (LSL–USL) are set at **38°C–42°C**, reflecting the stable natural cluster.

B. PROCESS VARIATION ON TEMPERATURE AND PH

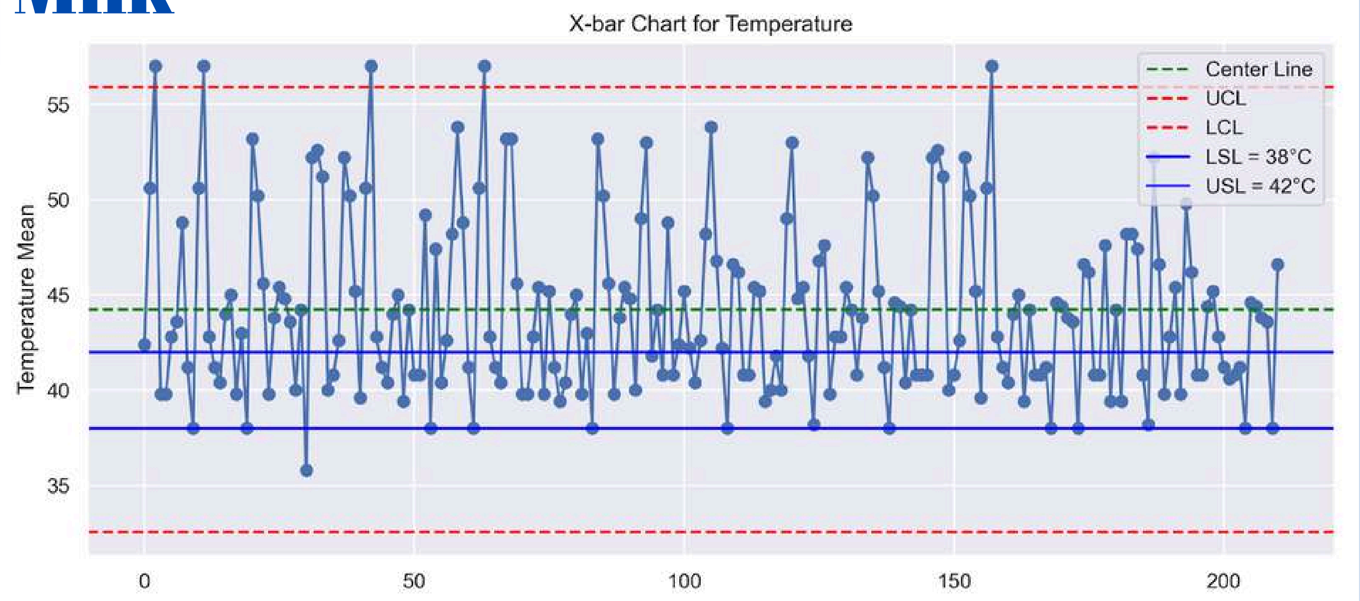
Specification Setting for High-Grade Milk

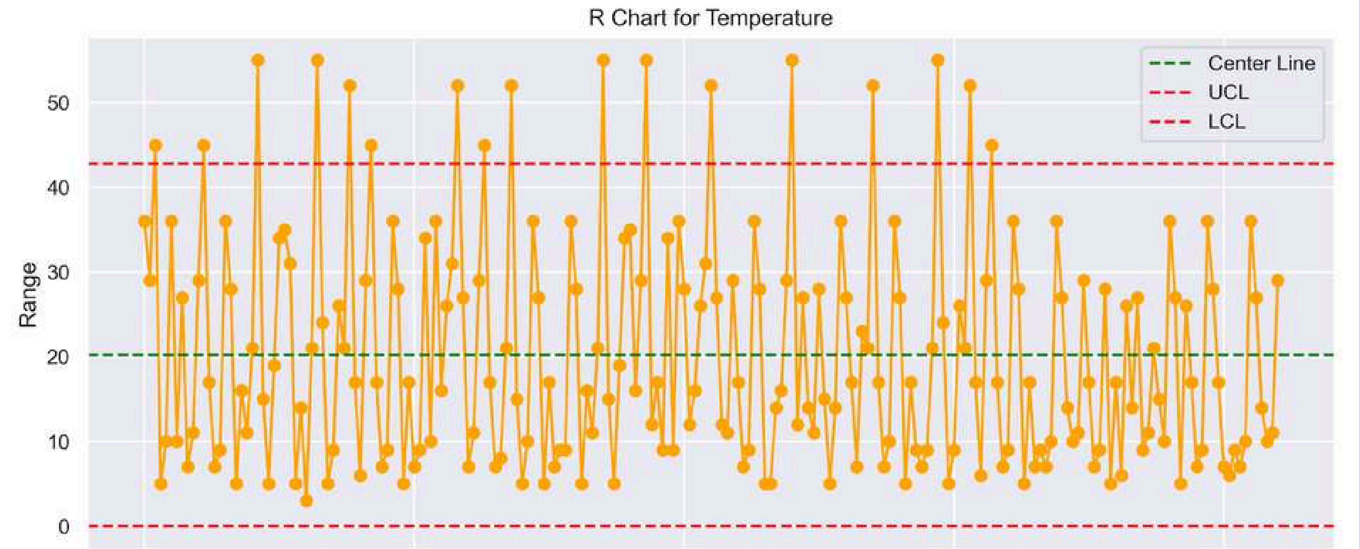
COMBINED X-BAR & R CHART EVALUATION

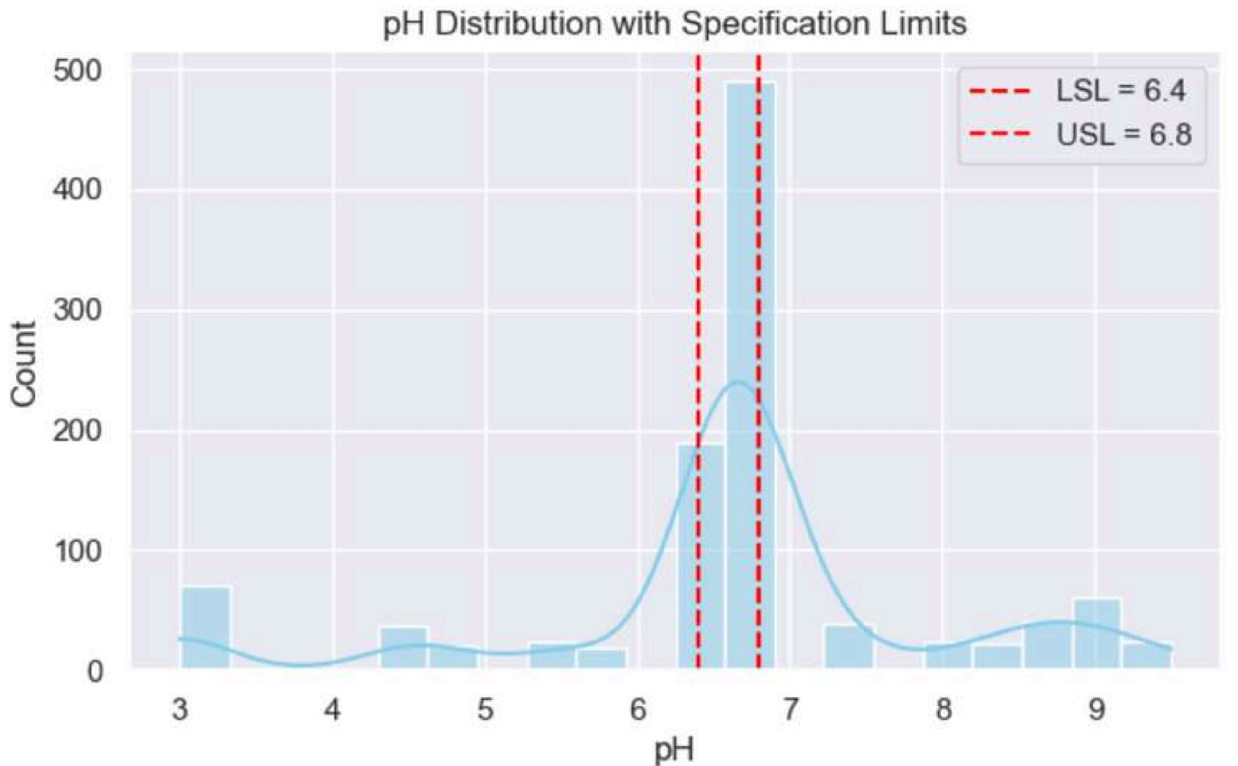
- Subgroup averages range **38°C–56°C**, with most points **outside 38–42°C**, showing **high variability** and **constant failure** to meet high-grade requirements.
- Frequent temperatures above the upper specification limit cause **quality deterioration**, reduced nutritional value, and other defects.

PROCESS INSTABILITY

- Subgroup means drift in and out of control limits → indicates **lack of control** and **special-cause variation** (heating system inconsistency, operator error, equipment malfunction).
- R chart also shows large and irregular variation, confirming uncontrolled short- and long-term variability

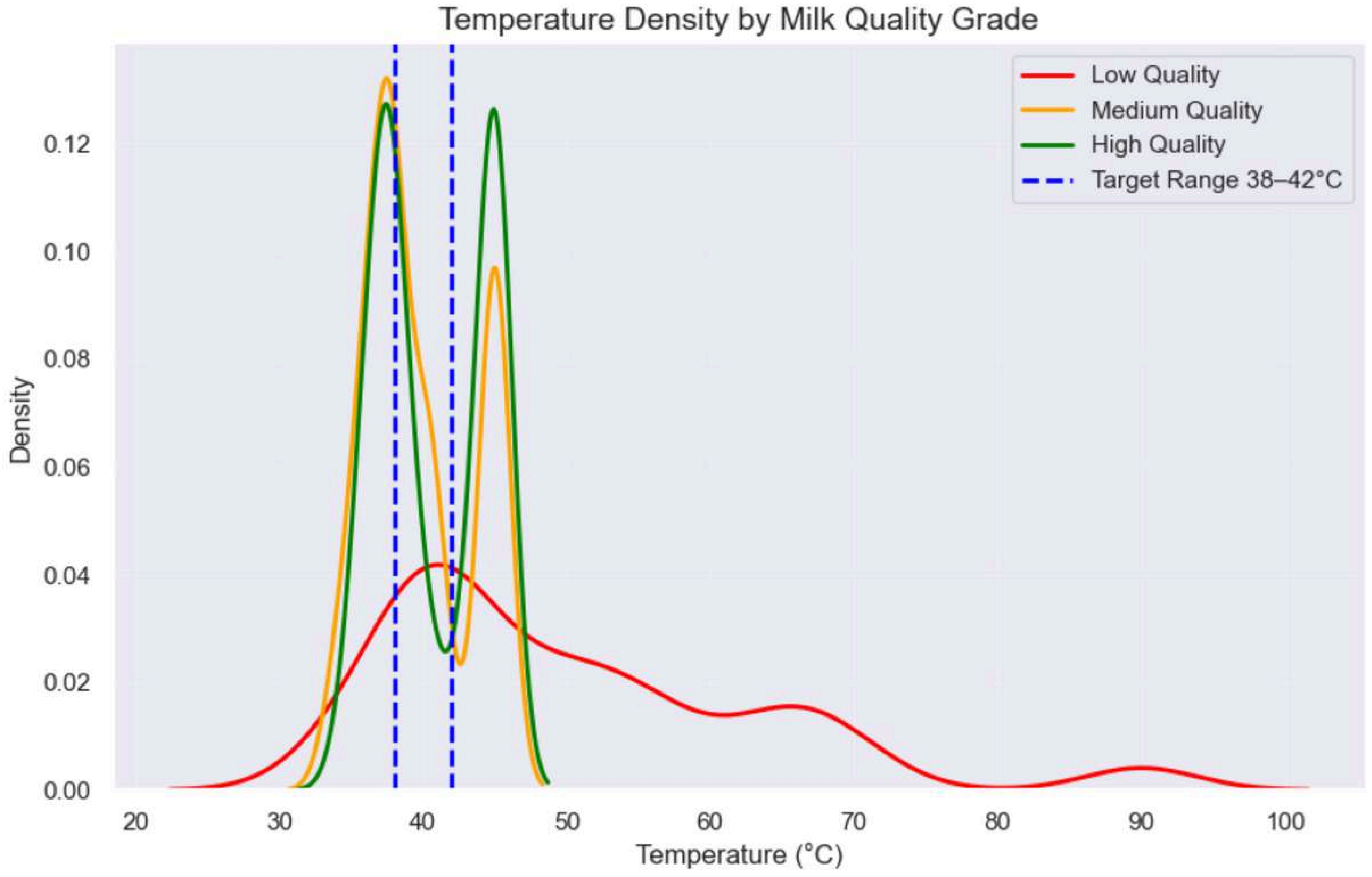






THE REASON FOR SETTING HIGH GRADE

- **The medium-quality** range spans from 34-45°C **does not form a consistent temperature cluster** and **appears only when the process fluctuates between high and low states**.
- The target temperature of 38–42°C is selected because it **represents the only stable and statistically valid operating zone** identified in the high-quality group.
- Setting the target at the **high-quality range aligns** with both the **voice of the customer and the natural behavior of the process**, while ensuring meaningful improvements in stability, capability, and product quality.
- Targeting medium quality would not enhance process control and would fail to meet market and quality requirements.



Medium is a "by-product", not a steady state of the process. It does not represent a realistic process target for the plant to aim for.

C. PROCESS CAPABILITY ANALYSIS (CP AND CPK)

Formulas

Cp = (USL – LSL) / (6σ)

Cpk = Z(min) / 3, where

Z(USL) = (USL – X̄) / σ

Z(LSL) = (X̄ – LSL) / σ

Given Values

- Mean Temperature = 44.23°C
- Standard Deviation = 10.10°C
- Cp = 0.066
- Cpk = –0.073

Interpretation

- **Cp = 0.066 < 1** → process variation is far wider than the allowable tolerance; even if centered, temperature would not fit within 38–42°C.
- **Cpk = –0.073** → process mean lies **outside** specification limits; biased above USL = 42°C → **systematic deviation**, not random noise.

Overall

- Temperature control process is **not capable** (Cp < 1).
- Process is **not centered** and consistently **violates specification limits** (Cpk < 0).
- Excessive variation caused by **uncontrolled process shifts, equipment instability, and inconsistent operations.**
- Current process **cannot reliably produce high-grade milk,** even before considering other CTQ attributes.

Conclusion

- Results confirm findings from X-bar and R charts: the temperature system is **highly unstable** and requires **urgent corrective actions** (equipment calibration, heating control, operator discipline).
- Without improvement, the process cannot meet **high-grade or export quality** standards.

EIU
TRƯỜNG ĐẠI HỌC QUỐC TẾ MIỀN ĐÔNG
EASTERN INTERNATIONAL UNIVERSITY
15 Year Anniversary 2010 - 2025

EVALUATION

Overall Findings

• Analytical results (distribution, hypothesis testing, CTQs, SPC) show the process is neither stable nor capable of producing high-quality milk consistently.

• Only 24% of output is high grade → process cannot meet premium market requirements.

Key Issues

• Large variation—especially in Temperature—causing frequent shifts outside acceptable conditions.

• Temperature control often violates limits and specs.

• Fluctuations indicate equipment/operator inconsistency.

Interdependencies

• Hypothesis testing confirms that quality variables are highly correlated—a deviation in one (e.g., Temperature) often triggers deterioration in others.

• Current system operates reactively with high variation and low capability.

• To target high-grade markets, the company must stabilize CTQs—particularly Temperature—standardize processes, calibrate equipment, and enhance real-time monitoring.

D. IMPROVEMENT AREAS.

Root Cause Identification

Six major causes were selected based on:

• Findings from SPC, hypothesis testing, and model analysis

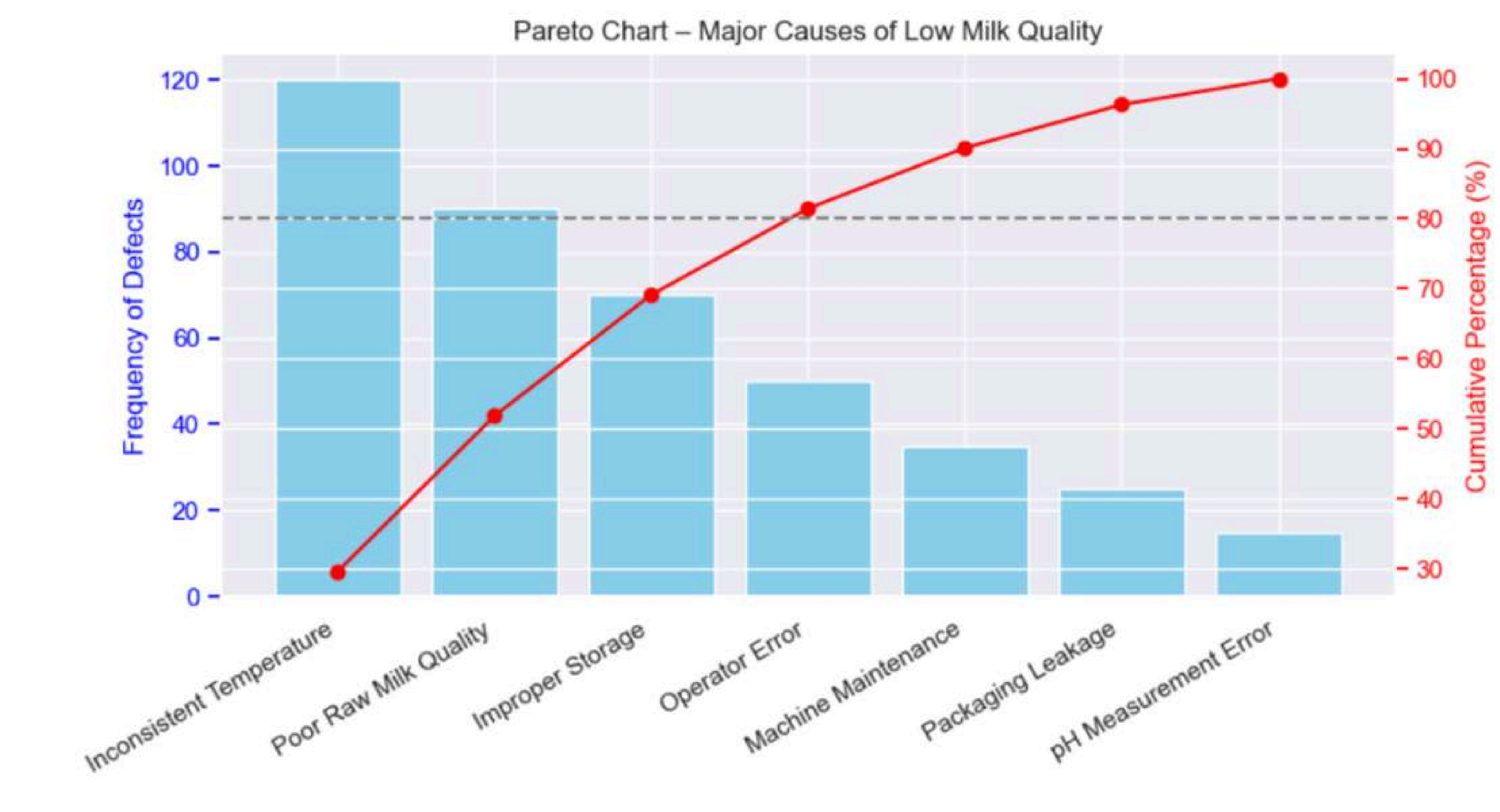
• 6M Framework (Man, Machine, Method, Material, Measurement, Environment)

Key Insight – 80/20 Principle

Three causes contribute to **>80% of total defects**:

1. Inconsistent Temperature Control
2. Poor Raw Milk Quality
3. Improper Storage Practices

These indicate that quality issues mainly originate from process instability and raw material variation.



Two Failure Pathways

• Process Variation Path: Temperature → pH → Colour → Grade

• Sensory Defect Path: Odor – Turbidity – Fat – Taste → Grade

Improvement Focus

• Strengthen temperature monitoring

• Tighten raw milk sourcing & supplier control

• Standardize storage procedures

V. BUILDING THE MILK QUALITY PREDICTIVE MODEL

Purpose

- Automatically classify milk into Low – Medium – High quality
- Reduce human subjectivity and errors
- Support expansion into high-demand markets

Why is it needed?

- Manual grading is unstable, biased, and labor-dependent
- Only 24.17% of samples meet High Quality, below market requirements

Benefits of the Model

- Standardizes quality assessment
- Detects and separates defective batches before packaging
- Improves speed, accuracy, and operational efficiency

Role in Quality Management

- Part of the Output Control stage
- Complements process improvements such as temperature and sensory control

V. BUILDING THE MILK QUALITY PREDICTIVE MODEL

B. METHODOLOGY

1. Data Preparation

Dataset split into:

- X (features): pH, temperature, flavor, odor, fat, turbidity, color
- y (target): Quality grade (Low–Medium–High)

X variables align with CTQs identified from hypothesis testing & SPC charts → ensures meaningful learning.

2. Standardization

- Although Decision Trees don’t require scaling, normalization was applied for pipeline consistency and future model testing.
- Reduces imbalance between continuous and binary variables → prevents feature dominance.
- Scaler fitted on training set only to avoid data leakage and maintain valid model evaluation.

V. BUILDING THE MILK QUALITY PREDICTIVE MODEL

B. METHODOLOGY

3. Train–Test Split

- Dataset split 80% training / 20% testing using a fixed random seed (42) for reproducibility.
- Final partition:
  - ✓ X_train = 847 ✓ X_test = 212
  - ✓ y_train = 847 ✓ y_test = 212
- Ensures evaluation reflects true generalization performance on unseen data.

C. MODEL RESULTS AND EVALUATION

1. pH = Primary Splitter (Root Node)
- pH gives the strongest first separation.
- pH > 7.1 → stable Medium branch.
- pH ≤ 7.1 → needs more CTQ splits.

2. Temperature = Main Quality Driver
- Second most important variable.
- High-quality appears only in tight temperature ranges.
- Temperature variation → shifts samples to Medium/Low.

3. Odor, Fat, Turbidity = Key Sensory Refiners
- Odor: off-odor → Low; clean → Medium/High.
- Fat: low fat → Medium/Low; higher fat → High.
- Turbidity: clear → High; cloudy → Low.

4. Colour = Secondary High vs Medium Splitter
- Helps refine High-quality classification in deeper nodes.

5. Quality Polarisation
- Some branches are very pure (stable High quality).
- Mixed nodes show unstable conditions → need process improvement.

CONFUSION MATRIX – MODEL PERFORMANCE

- The model correctly classified 210/212 test samples.
- Only 2 misclassifications, showing extremely strong predictive ability.
- Low-quality: 77/78 correct (1 misclassified as High).
- Medium-quality: 86/86 correct (100% accuracy).
- High-quality: 47/48 correct (1 misclassified as Medium).

ACCURACY & F1-SCORE OVERVIEW

- Accuracy: 99.06%
- F1-Score (Low): 0.9935
- F1-Score (Medium): 1.0000
- F1-Score (High): 0.9792
- Macro F1: 0.9890
- Weighted F1: 0.9906

Accuracy: 0.9906

Classification Report:

	precision	recall	f1-score	support
0	1.0000	0.9872	0.9935	78
1	0.9885	1.0000	0.9942	86
2	0.9792	0.9792	0.9792	48
accuracy			0.9906	212
macro avg	0.9892	0.9888	0.9890	212
weighted avg	0.9906	0.9906	0.9906	212

RAW MILK — DECISION TREE — GOOD OR BAD

MODEL EVALUATION & GENERALIZATION ABILITY

·The model shows very high accuracy, but Decision Trees may overfit.
·Dataset is highly separable (polarization effect) → classes are very clear.
·To ensure real-world reliability:

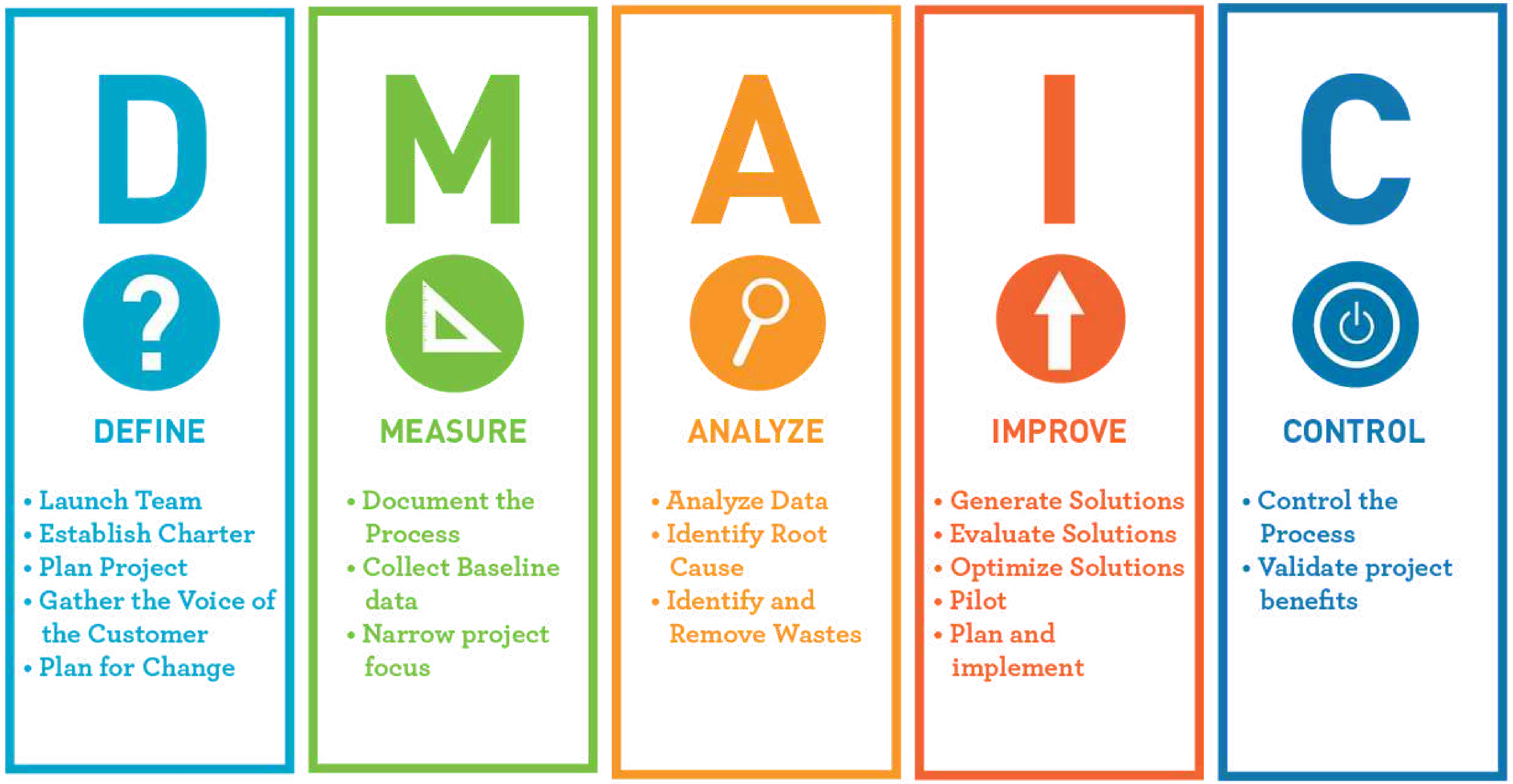
- Use k-fold cross-validation
- Apply pruning
- Test model on new production batches
- Monitor accuracy over time

EIU TRƯỜNG ĐẠI HỌC QUỐC TẾ MIỀN ĐÔNG EASTERN INTERNATIONAL UNIVERSITY
15 Year Anniversary 2010 - 2025

VI. MANAGEMENT AND QUALITY IMPROVEMENT PROPOSALS

Rationale for Using the DMAIC Framework

- The production problems are complex and multi-factor
- A data-driven approach is required
- The company must not only improve but also sustain improvements



How the DMAIC Framework was applied

DEFINE:

- **Problems:** a high percentage of Low-grade milk (around 40.5%) and unstable production performance.
- The business need is to produce consistent high-grade milk to meet the requirements of premium markets.
- Mentioned in sections I

MEASURE:

- Section II measured the current process performance by analyzing the distribution and variation of Temperature, Odor, Turbidity, Fat, Colour, and pH.
- Grade distribution results also provided a baseline for improvement.

EIU
TRƯỜNG ĐẠI HỌC QUỐC TẾ MIỀN ĐÔNG
EASTERN INTERNATIONAL UNIVERSITY
15 Year Anniversary 2010 - 2025

How the DMAIC Framework was applied

ANALYZE:

- Root causes were explored using ANOVA, Chi-Square tests, correlation analysis, SPC (X-bar and R charts), process capability indices, and the Decision Tree model.
- **Temperature** was identified as the **strongest CTQ**, followed by Odor, Turbidity, Fat, and Taste.
- The Decision Tree achieved **99.06% accuracy**, confirming that these CTQs reliably determine product quality.

IMPROVE:

- The Improve phase is organized into three main areas: **Input Control, Process Control, and Output Control** because the Analyze phase showed that these areas contain the main sources of variation.

Table 1: Feasibility of Input Improve ment Actions

Improvement Action	Cost Level	Implementation Difficulty	Notes
Raw Milk Acceptance Standards	Low	Easy	Quick tests and digital logs are inexpensive.
Supplier Scorecard	Low	Easy–Moderate	Mainly administrative work; no new equipment needed.
Storage & Handling (FIFO, cleaning logs, chilling)	Low–Moderate	Easy	Uses existing equipment; changes are mostly procedural.

How the DMAIC Framework was applied

Table 2: Feasibility of Process Improvement Actions

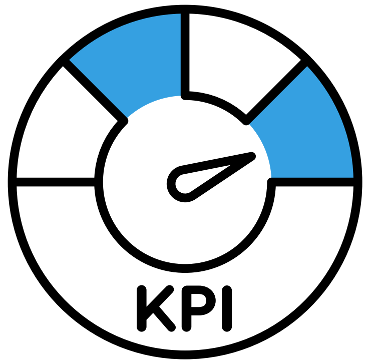
Improvement Action	Cost Level	Difficulty	Notes
PID Temperature Controllers	Moderate	Moderate	Most expensive action but crucial for quality.
SOP Updates & Training	Low	Easy	Requires time, not money.
Real-Time Monitoring Dashboard	Low–Moderate	Moderate	Can be built using simple tools; sensors may add minor cost.

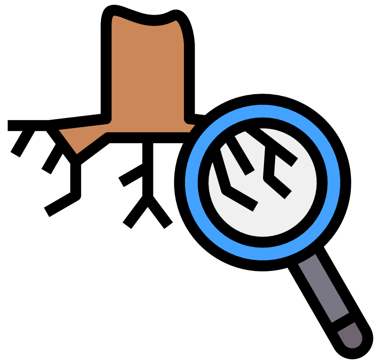
How the DMAIC Framework was applied?

Table 3: Feasibility of Output Improvement Actions

Improvement Action	Cost Level	Difficulty	Notes
Decision Tree Model Integration	Very Low	Easy	Requires basic software setup and staff training; improves consistency in output decisions.
Standardized Final Inspection	Low	Easy	Only requires procedure writing and training.

How the DMAIC Framework was applied ?





CONTROL:

- **Key Performance Indicators (KPIs):** %Temperature readings within 38–42°C, % High-grade batches, Supplier compliance rate, Reduction in process variation, Frequency of deviations and reprocessing events
- **Routine Monitoring and Audit Activities:** SPC chart review for Temperature and other CTQs, Supplier performance review and quality meetings, SOP updates and sanitation audits, Equipment calibration and process capability review (Cp, Cpk)
- **Continuous Learning and Prevention of Recurrence:** apply root-cause analysis tools such as the 5 Whys or the Fishbone diagram. Encouraging small improvement ideas (Kaizen activities) supports a culture of continuous learning.

VII. CONCLUSION AND RECOMMENDATIONS

CONCLUSION

Current process is unstable: 40.51% Low quality, only 24.17% High quality

Main causes:

- **Uncontrolled temperature** (avg. 44.23°C, extremes up to 90°C)
- **Inconsistent raw milk** (odor, turbidity, low fat)
- **Inadequate cleaning**

Situation: “Sometimes excellent, often bad”

Predictive model proves excellent milk is achievable

With better control + tools, consistent high-quality milk is possible daily

EIU
TRƯỜNG ĐẠI HỌC QUỐC TẾ MIỀN ĐÔNG
EASTERN INTERNATIONAL UNIVERSITY

15 Year Anniversary 2010 - 2025

RECOMMDENDATIONS

1

Strict Control of Critical Factors

2

Automated Real-Time Monitoring

3

Deploy DMAIC Improvement Cycle

4

Leverage Predictive Model

THANK YOU !