EASTERN INTERNATIONAL UNIVERSITY

BECAMEX BUSINESS SCHOOL

SCLM 449 - Process Control and Improvement

CASE STUDY: HAPPY COW LTD

Milk Quality Management

Lecturer: Mr. Huynh Minh Binh

Prepared by: Tiramisu Group

Name	IRN
Vương Hoàng Cẩm Thư	1932300177
Nguyễn Thị Thu Phương	2032300167
Dinh Nho Thanh Binh	2132300011
Nguyen Do Xuan Ai	2332300353
Nguyen Ho Hoang Yen	1932300502
Nguyen Hoang Mai Thi	2132300453
Tran Kim Ngoc	2032300403

Quarter 3: 2024-2025

Table of content

I. Introduction

With Frodo at the helm, Happy Cow Ltd. has developed into a prosperous milk production business in the center of Middle Earth, tucked away between the serene Shire and the guardian Misty Mountains. After successfully catering to the local markets in Shire and Anor, the business is now looking to reach more discerning markets like Gondor, Rohan, and Mordor in order to grow its clientele. Creating a significant presence in the bright and rising kingdom of Binh Duong is the first stage in our ambitious diversification plan. But in order to compete in these new markets, Happy Cow Ltd. needs to make sure that its milk products live up to greater requirements. The challenge for the supply chain management team is to create a thorough plan to evaluate and improve milk quality using machine learning models based on seven important observable variables: pH, temperature, taste, odor, fat, turbidity, and color. This will allow for the precise classification of milk into low, medium, or high quality and safeguard the brand's reputation throughout the Middle East.

1. pH: This feature defines pH of the milk, which is in the range of 3 to 9.5.

2. Temperature: This feature defines the temperature of the milk, and its range is from 34'C to 90'C.

3. Taste: This feature defines the taste of the milk and takes the possible values: 1 (good) or 0 (bad).

4. Odor: This feature defines the odor of the milk and takes the possible values: 1 (good) or 0 (bad).

5. Fat: This feature defines fat of the milk and takes the possible values: 1 (good) or 0 (bad).

6. Turbidity: This feature defines the turbidity of the milk and takes the possible values: 1 (good) or 0 (bad).

7. Color: This feature defines the color of the milk, which is in the range of 240 to 255.

8. Grade: This is the target and takes the values: low_quality,medium_quality or high_quality.

II. Methodology

To evaluate and improve milk quality at Happy Cow Ltd., the combination of statistical process control techniques and machine learning, towards both real -time monitoring and predictive models was used. The techniques are selected based on their actual utility and their relevance to the nature of the data collected. First of all, control charts (charts $\bar{X}$ and S) are used to monitor the stability of the process over time for continuous variables such as pH, temperature and color. X-bar chart represents the average or arithmetic mean of a process's changes over time, calculated from a set of subgroup values, while S chart denotes the standard deviation of a process over time, derived from a set of values (Hessing, 2025). The control chart helps detect abnormal changes or trends that can indicate potential quality issues, allowing early intervention to the production process.

A box plot visually represents a dataset's distribution, highlighting key statistics like the median, quartiles, and possible outliers in a clear, concise way (GeeksforGeeks, 2024). It allows for summarizing data distribution, spotting outliers, and comparing multiple datasets compactly and visually (GeeksforGeeks, 2024). Description images (box chart) are built to visualize the distribution of the forecast variables on three levels of milk quality (low, medium, high). These charts are especially useful to detect foreign values and understand the central trend and the spread of variables such as taste, smell, fat, turbidity, pH, temperature and color.

Correlation analysis. A correlation matrix has been created to evaluate linear relationships between forecast variables. This helps identify potential multi -line issues and adds context for explaining regression models. For example, fat shows a moderate correlation with both smell and taste, indicating that sensory quality factors are connected to each other.
Based on seven criteria, milk quality is categorized as low, medium, or high using MLR polynomial logistics (MLR), a fundamental prediction technique. The aim variable is the categorization of more than two layers, which is why this model was selected. It enables the identification of the variable that is a statistical prediction variable. The percentage of milk samples that correspond to a particular kind is assessed using its characteristics, constructing a strong prediction model that has a high level of explanation power (Nagelkerke R2 = 0.837).

III. Control Charts for Process Monitoring

1. PH

Overview, The milk pH process is statistically stable and in control, which is a very positive finding. The process is predictable, and the average pH is centered at an ideal level for fresh milk.

However, while the process is stable, the amount of variation is quite high. This means that while the average quality is good, individual batches will predictably vary enough to potentially fall outside of optimal quality standards for freshness or safety. The primary goal should be to reduce this variation.

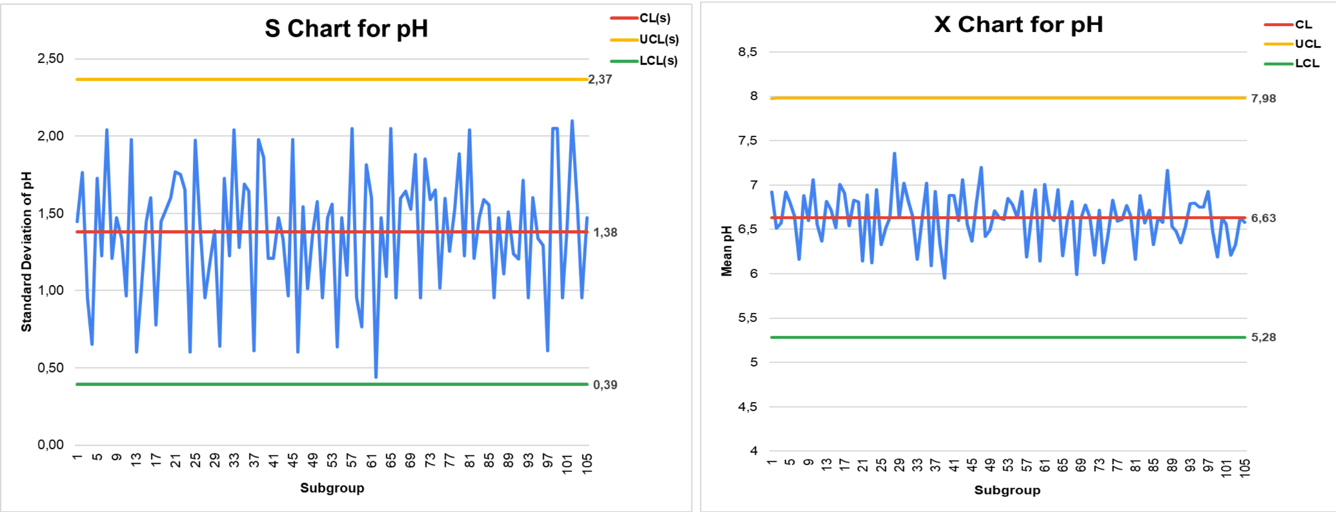

Figure 1: S chart & X chart for pH

The S-chart tracks the standard deviation (variation) within each subgroup. It must be analyzed first because if the variation is not stable, the control limits on the X-chart are meaningless.

Center Line (CLs): 1.38

Upper Control Limit (UCLs): 2.37

Lower Control Limit (LCLs): 0.39

All data points are within the upper and lower control limits. There are no obvious non-random patterns (e.g., long runs on one side of the average, trends, etc.). This indicates that the short-term variation of the process is stable and predictable. The way pH varies within a sample batch is consistent over time. This is excellent, as it suggests that factors like measurement consistency and short-term environmental conditions are stable.

The X-chart tracks the average pH of each subgroup. Since the S-chart is in control, we can confidently analyze the X-chart.

Center Line (CL): 6.63

Upper Control Limit (UCL): 7.98

Lower Control Limit (LCL): 5.28

All data points are within the control limits. The distribution of points around the center line appears random. This confirms that the process is stable.

There is an Excellent Centering which means the overall process average (CL) is 6.63. This is ideal, as the normal pH of fresh cow's milk is typically between 6.5 and 6.7 (Dung, 2021). Compared to Dutch Lady which is one of the leading milk brands in Vietnam, owned by FrieslandCampina Vietnam. The standard pH of Dutch Lady fresh milk is between 6.6 - 6.8, ensuring the quality and purity of the product (*Dutch Lady, n.d.*). Therefore, the Happy cow's process is centered perfectly in the optimal range.

Predictability:

Consistent Average Quality: Happy cows can reliably predict that the average pH of their milk will be around 6.63, which signifies high-quality, fresh milk.

No "Special Causes": The process is free from "special cause" variation. This means Happy cow don't have sudden, unpredictable events throwing your pH off (like a machine malfunction or a single contaminated batch). The variation that they see is inherent to their current system.

Even though the process is "in control," it may not be "capable" of meeting strict quality specifications all the time.

Wide Range of pH Values: The chart shows that subgroup averages fluctuate between approximately 5.9 and 7.4.

Low pH (approaching 6.0 and below): A lower pH indicates the milk is becoming more acidic. This is a primary indicator of bacterial growth and fermentation (souring), where bacteria convert lactose into lactic acid. Batches with a pH closer to 6.0 are at risk of having a shorter shelf life and developing an off-taste.

High pH (approaching 7.0 and above): A higher (more alkaline) pH is often an indicator of milk from cows with mastitis, an udder infection. This milk can have quality defects and may pose food safety concerns.

Control Limits vs. Specification Limits:

The control limits (5.28 to 7.98) show the expected variation of the process (the "voice of the process").

Quality standards for milk are the specification limits (the "voice of the customer/regulator"). For example, a customer might reject any milk with a pH below 6.4 or above 6.9.

Happy cow's process, while stable, is producing milk that regularly approaches and may sometimes exceed these critical quality specification limits.

Predictions and Recommendations

Based on these charts, it will predictably and consistently produce some batches of milk that are on the edge of being too acidic (soured) or too alkaline (potential mastitis). The overall quality will be good on average, but there will be a consistent level of "borderline" product. Based on analyzing above, we provide some recommendations:

First, Recommendation 1 (Investigate): they should not react to individual high or low points on the chart. Since the process is in control, these are not special events. Instead, focus on the overall system. Secondly, Recommendation 2 (Reduce Variation): The primary goal is to make the blue line on both charts "flatter" by tightening the variation. This requires fundamental process improvement, such as:

Herd Health: Improving screening and management of mastitis to reduce high-pH milk.

Sanitation & Cooling: Improving sanitation of milking equipment and ensuring rapid, consistent cooling to inhibit bacterial growth that causes low pH.

Feed Consistency: Ensuring consistent diet for the herd, as feed can influence milk composition.

By reducing the overall variation, Happy cows will tighten the control limits and ensure that virtually all their milk not only averages 6.63 but also stays consistently within a much tighter, optimal range (e.g., 6.5 to 6.8), dramatically improving overall quality and shelf-life predictability.

2. **Temperature**

Overview, The process for controlling temperature is out of statistical control. Both the process variation (S Chart) and the process average (X Chart) show clear signs of instability due to special cause variation. This means the process is unpredictable, and as a result, the quality of the milk will be inconsistent and unreliable.

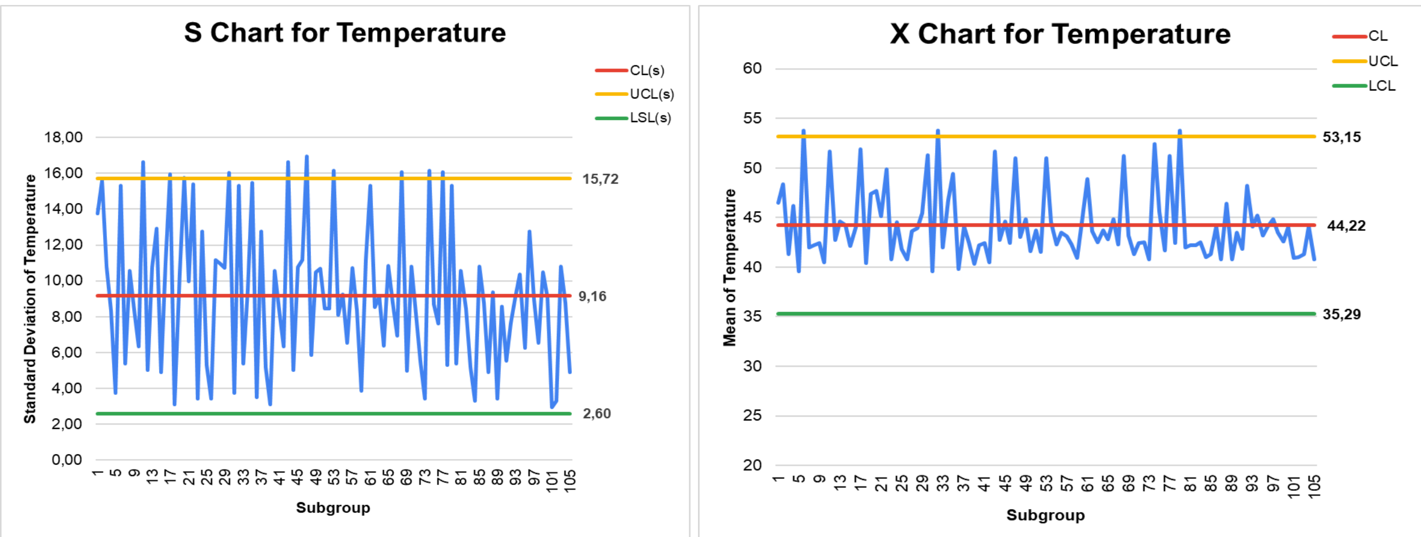

Figure 2: S chart & X chart for Temperature

As Happy Cow Ltd. grows into high-demand markets like Gondor and Mordor, keeping a constant temperature during milk processing is crucial for guaranteeing product safety, shelf life, and regulatory compliance. 105 subgroups' standard deviations are displayed in the S Chart for Temperature.

Center Line (CL): 9,16

Upper Control Limit (UCL): 15,72

Lower Control Limit (LCL): 2,60

There are multiple points that exceed the UCL of 15.72 (e.g., around subgroups 21, 37, 45, 65, 85, 93). This is the clearest signal that the process variation is out of control. The chart shows extreme and rapid fluctuations. The standard deviation does not hover around the center line; instead, it jumps wildly from low to high values. This indicates that the process consistency is unstable and unpredictable. An unstable S-chart means the temperature control system is erratic. Sometimes the temperature is very steady (low standard deviation), and other times it fluctuates wildly (high standard deviation).

High variation could mean parts of the milk are being overheated while other parts are underheated within the same batch. This can lead to, Underheated milk may not kill all harmful bacteria, leading to spoilage and safety risks. Overheated milk can have a "cooked" taste, denatured proteins, and reduced nutritional value.

Such variance could jeopardize the final product's sensory and microbiological integrity for a business like Happy Cow that strives to meet or beyond international quality standards (like Dutch Lady or Vinamilk).

The average temperature of milk samples is monitored using the X Chart for Temperature, which provides more evidence of this problem.

Center Line (CL): 44,22

Upper Control Limit (UCL): 53,15

Lower Control Limit (LCL): 35,29

While the most obvious issue is the severe oscillation pattern, a closer look reveals that at least one point also breaches the Upper Control Limit (UCL). This combination of signals indicates both systemic instability and specific, out-of-bounds events, making the process unpredictable and unreliable for producing consistent quality milk. There are some Points Outside the Control Limits (A Clear Signal of Instability). Upon close inspection, at least three data points, around subgroups 77, 33 and 9 , clearly exceed the Upper Control Limit (UCL) of 53.15. Other points come extremely close to or may just touch the UCL.

A single point outside the control limits is, by itself, sufficient evidence to declare a process out of control. It signals the presence of a "special cause" of variation—an event or factor that is not part of the normal process. This is a critical finding that demands immediate investigation. What happened at subgroup 77 to cause such a high average temperature?

Severe and Systematic Oscillation (Sawtooth Pattern). This is the most dominant visual pattern in the chart. The data does not vary randomly around the center line (CL). Instead, it swings in a rapid, predictable, up-and-down cycle from a high value to a low value.

This is a classic non-random pattern indicating a fundamental problem with the process control system. Random, or "common cause," variation should cluster around the average. This pattern suggests: Firstly, An operator or an automated controller is "chasing" the target. When the temperature is slightly high, they over-correct and make it too low, and vice versa.

Secondly, An automated PID (Proportional-Integral-Derivative) controller that is not properly tuned can easily induce this type of oscillation.Finally, There may be a significant delay between making an adjustment (e.g., turning on a heater) and seeing its effect, leading to overcompensation.

Based on this unstable and out-of-control process, the milk quality will be unpredictable and inconsistent. This leads to a high risk of defective product when the process is producing batches that are, on average, too hot and too cold. Batches corresponding to the high peaks, especially the one exceeding the UCL, were likely overheated. This can cause a cooked or burnt flavor, denature proteins (affecting texture and nutritional value), and reduce shelf life. Batches corresponding to the low points were likely under-heated. If this is a pasteurization step, this is a critical food safety failure. Harmful pathogens may survive, leading to rapid spoilage and posing a serious health risk to consumers.

We cannot look at this chart and predict that the next batch will be acceptable. The process's performance is erratic. While some batches may fall within an acceptable range by chance, many will not. Quality is being determined by luck rather than by a stable, controlled process.

Conclusion, The X Chart provides undeniable evidence that the temperature control process is broken. Both specific events (points outside the UCL) and systemic issues (oscillation) are present. To ensure milk quality and safety, the root causes of both these problems must be identified and eliminated.

Milk Brand Comparison

Dutch Lady and Vinamilk use PID systems with real-time monitoring and a tolerance of ±1°C to maintain strict temperature control. Regular calibrations are performed on these, and any differences are automatically corrected. By comparison, Happy Cow's process exhibits substantial instability and variations, with standard deviations rising to 15.72°C, which is significantly more than what is considered acceptable in the industry.

International food safety standards such as ISO 22000 and HACCP are met by both Dutch Lady and Vinamilk. They guarantee constant pasteurization; any variation results in prompt product rejection or corrective action. Operating under comparable norms would put Happy Cow's frequent swings below 60°C and above 70°C at danger for safety issues and regulatory violations.

Happy Cow has the danger of batches being overheated or under heated, whereas top brands constantly provide premium milk with a consistent flavor and a long shelf life. This may lead to concerns that rivals closely monitor by precise heat treatment, such as microbial contamination, bad flavor, or nutrient loss.

Another significant distinction is operational discipline. Dutch Lady and Vinamilk perform preventive equipment maintenance, regularly train employees, and adhere to stringent SOPs. The oscillation patterns and Happy Cow's lack of control imply that these measures are either nonexistent or very weak.

Happy Cow needs to enhance temperature management right now in order to be competitive. Upgrading automation, improving training, and imposing process discipline will help it meet high standards and thrive in competitive markets like Gondor and Mordor.

Predictions and Recommendations

Recommendation 1 (Stabilize Temperature variance): Although there isn't a specific cause for the variance, the observed range is sufficiently large to raise questions about the quality of the product. To lower variability, the business should improve cooling system maintenance, temperature control automation, and equipment calibration.

Recommendation 2 (Increase Process Capability): Happy Cow should determine if their temperature goals are in line with microbiological safety and regulatory requirements. For reliable pasteurization and spoiling avoidance, the average temperature should be raised closer to 63°C while maintaining strict control.

Recommendation 3 (SOPs and Training): Retraining employees on temperature management and standardizing operational procedures can help minimize human-induced variability.

By fixing these temperature control problems, Happy Cow Ltd. will improve the consistency and safety of its milk, bolstering its strategic growth into cutthroat markets while upholding its brand promise as the freshest and safest dairy in the Middle East.

3. Colour

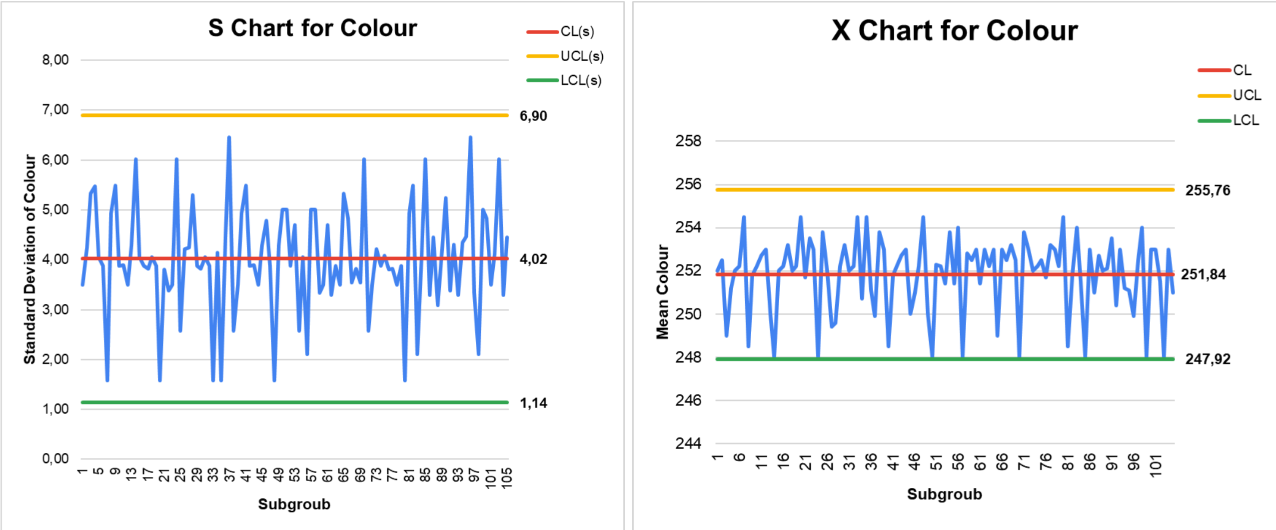

Figure 5: S chart & X chart for Color

The S-chart evaluates the short-term variation in the milk color process by monitoring the standard deviation of color values within each grouping. It is crucial to look at this chart first since any interpretation of the X-chart, which tracks averages, becomes untrustworthy if the variance is unstable.

Center Line (CLs): 4,02

Upper Control Limit (UCLs): 6,90

Lower Control Limit (LCLs): 1,14

Every data point is contained inside the control ranges, and no non-random patterns, including cycles, trends, or runs, are present. This implies that color variation within subgroups—also known as batch-to-batch variance—is steady and constant throughout time. Such stability suggests that controls are in place over external influencing elements, such as operator handling, sensor calibration, and illumination conditions. High short-term consistency suggests that consumers would perceive milk drawn from the same batch as having a uniform appearance, which is important because milk color has a significant impact on consumer perception. The color measuring method is dependable and free from unique sources of variation, such as sensor drift, abrupt changes in the environment, or sample contamination, as confirmed by the stability of variance. Insights from the accompanying X-chart, which examines mean values, can be trusted thanks to it.

The X-chart shows each subgroup's average color value. For identifying slow changes or patterns in the general look of milk over batches, this measure is crucial.

Center Line (CL): 251,84

Upper Control Limit (UCL): 255,76

Lower Control Limit (LCL): 247,92

The center line is randomly dispersed, and all average values fall inside the control boundaries. The process is in statistical control if there are no noticeable rising or negative trends, nor are there any extended periods of time above or below the average. The milk's color remains constant from batch to batch, indicating a dependable and well-run production system. It is confirmed that there are no special causes for the process by this stable mean and the previous result of stable variation. Significant interruptions like changes in the raw materials, flaws in the process, or inconsistent operator behavior are not affecting the milk's color. In the event that the present average (CL = 251.84) corresponds to industry or internal standards for fresh milk color, it falls comfortably within the anticipated range of standard visual quality.

Milk Brand Comparison

Milk color is a visual indicator of freshness, purity, and general quality that greatly influences consumer impression. According to Happy Cow's S-chart and X-chart evaluations, a statistically stable color control procedure has been devised. The process does not exhibit volatility due to specific causes, and both short-term and long-term averages stay within their respective control ranges. At 251.84, the average color value is stable and aesthetically pleasing. Nonetheless, Happy Cow's control range (247.92–255.76) is quite broad in contrast to the usual consumer expectations, which could vary from 249 to 254. Despite the process's dependability, this greater variation means that some batches could seem overly yellow or pale to pickier customers.

Dutch Lady, a well-known brand worldwide, on the other hand, has extremely sophisticated quality control procedures that give color consistency first priority. Dutch Lady, which is well-known for her high-quality milk products, uses automated color monitoring systems in real time under carefully regulated lighting. These devices guarantee that the color of milk stays within a very specific, customer-aligned range of roughly 250 to 253. The company's fresh milk consistently surpasses local quality criteria, including a total plate count that is much below statutory limits, and it complies with international standards like ISO certifications. Dutch Lady is able to maintain remarkable visual uniformity throughout all manufacturing batches thanks to this strict process control, which boosts customer loyalty and trust.

Vinamilk, a market leader in Southeast Asia, also shows a great dedication to visual uniformity by emphasizing process discipline and raw milk quality. Vinamilk strictly regulates upstream factors, such as cow health, feed quality, and centralized milk collection systems, to guarantee consistency rather than mainly depending on downstream corrections. Consequently, the milk hue stays in a more constrained range, usually between 248 and 252, which is in line with consumer expectations. Additionally, the business uses centralized lab testing and adheres to international standards such as ISO 17025 and BRC to guarantee that all of its products fulfill stringent quality requirements, including those related to appearance.

Happy Cow nevertheless trails Dutch Lady and Vinamilk by a small margin in terms of visual accuracy and customer alignment, despite having a reliable and statistically sound color control system. Vinamilk prioritizes consistency through raw milk quality and preventive measures, whereas Dutch Lady sets the standard with automated, real-time modifications and little variance. Happy Cow should think about enforcing more stringent environmental and input quality controls, investing in more sophisticated color monitoring equipment, and reducing its internal control limits in order to catch up to these leaders. With these enhancements, the brand would be able to more successfully compete in regions where visual product homogeneity is a crucial differentiation and better satisfy consumer expectations.

Predictions and Recommendations

With steady short-term fluctuation, Happy Cow can accurately forecast that the average milk color will stay around 251.84. This degree of consistency and predictability suggests that the procedure can be repeated and that customers will find the product's visual quality to be consistent. Although the "voice of the process" is represented by the control limits (247.92 to 255.76), the "voice of the customer"—also known as customer expectations—may be more stringent. Customers or authorities, for example, might mandate that milk have a color value that is strictly between 249 and 254. In that instance, some milk batches may fall outside of this allowed range even though they were made using a reliable procedure. Therefore, although the process is under control, it might not be entirely able to continuously satisfy all of the appearance-related expectations of the consumer. Milk that falls outside of this range could be viewed as being excessively yellow or pale, which could damage consumer confidence or result in product rejection.

Recommendation 1 (Pay Attention to Systemic Changes): The company should not get caught up in the details of either chart. The method is stable, therefore these variations are to be expected. Efforts should instead concentrate on improving the system generally in order to lower overall variation.

Recommendation 2 (Tighten Control Limits by Reducing Variability): It is crucial to concentrate on a few crucial operational enhancements in order to lessen variance and tighten control limits in milk color. First and foremost, sensor standardization is essential. To avoid inconsistent results from glare, shadows, or color shifts from ambient light, color measurement equipment should be calibrated on a regular basis and utilized under controlled lighting. Simultaneously, environmental controls should be implemented to prevent outside influences like sunshine or changing room temperatures, which might have an impact on the milk's quality and measurement accuracy. Strict cleaning procedures must be followed since even small residues or accumulation in processing or storage systems can subtly change the color of milk. Equipment sanitation is another important consideration. Furthermore, milk composition is directly impacted by feed consistency and cow health; changes in nutrition or health conditions, such as illness or stress, can alter the appearance of milk. Uniformity is supported by providing the herd with a consistent, high-quality food and preventative medical care. Lastly, cooling efficiency should not be disregarded because quick and steady cooling right after milking helps maintain the milk's natural color by avoiding discoloration from heat. When combined, these steps will drastically cut down on undesired color variability and guarantee that the process stays within more stringent, customer-aligned quality standards.

Recommendation 3 (Define and Comply with the Limits of the Specification): Establish precise specification limitations such as 249–254 in collaboration with the quality assurance team. Next, try to minimize process variance so that almost all output is inside this more constrained range. This method reduces the possibility of consumer discontent due to cosmetic variances and enhances product uniformity.

IV. Boxplot Analysis for Grade Prediction

Overview, The box: Represents the central 50% of the data (Interquartile Range or IQR).

The line inside the box: The median (the middle value).

The 'x' inside the box: The mean (the average value).

The "whiskers" (lines extending from the box): Show the range of the data, excluding outliers.

The dots/circles: Outliers, or data points that fall far outside the rest of the distribution.

1. Grade and pH

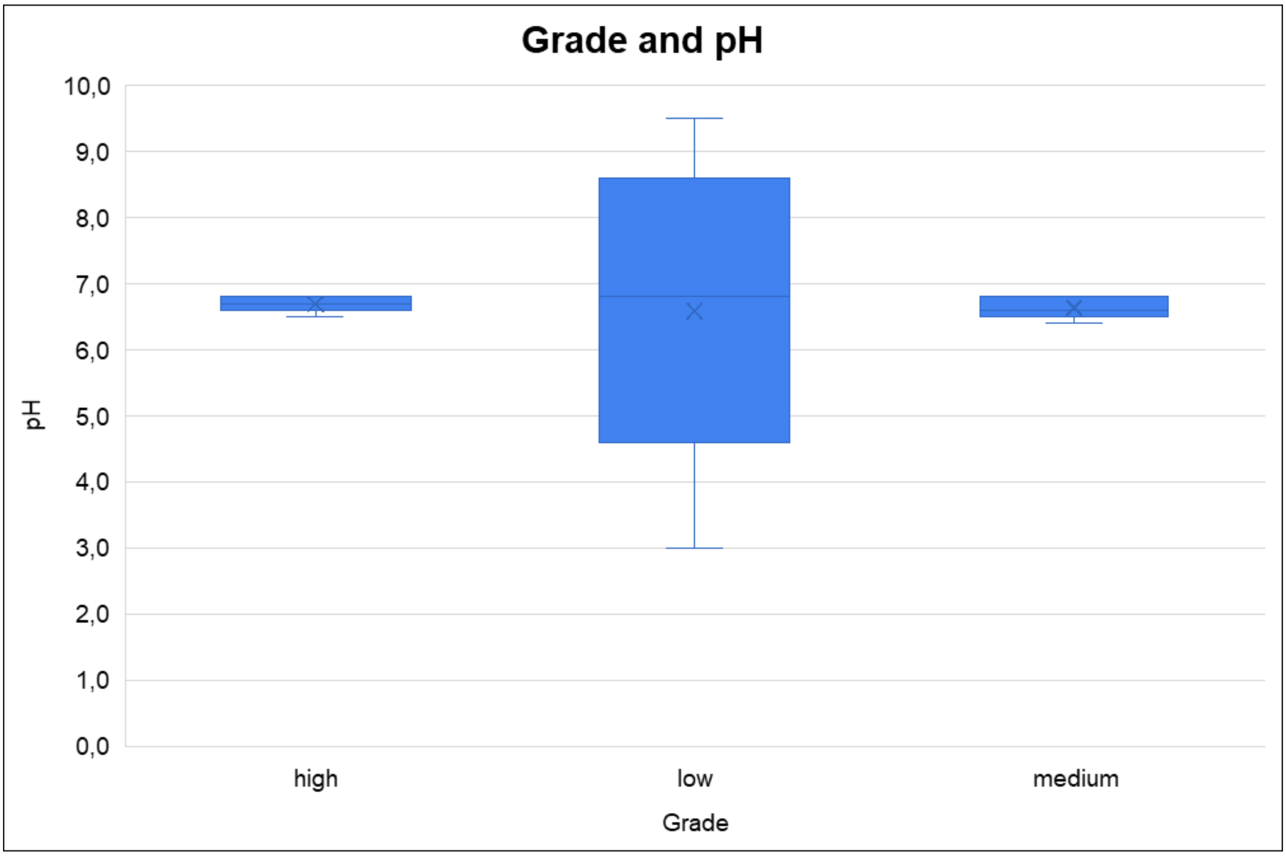

Figure 7: Grade and pH

The boxplot titled "Grade and pH" illustrates the distribution of milk pH levels across three quality categories: high, medium, and low. High-grade milk exhibits a stable and consistent pH, with a median around 6.7 and a very narrow range, reflecting strong process control. In contrast, low-grade milk displays extreme pH variability, ranging from approximately 4.5 to 8.5, with a median around 6.5, indicating a complete lack of control or spoilage, where the widespread is the defining characteristic. Medium-grade milk, similar to high-grade, shows a stable pH around 6.8 with a narrow range. The boxplot highlights that pH variability is the strongest predictor of low-grade milk, as highly erratic pH levels are almost certainly indicative of low quality, while high and medium grades are indistinguishable by pH alone due to their stable, near-neutral values.

2. Grade and temperature

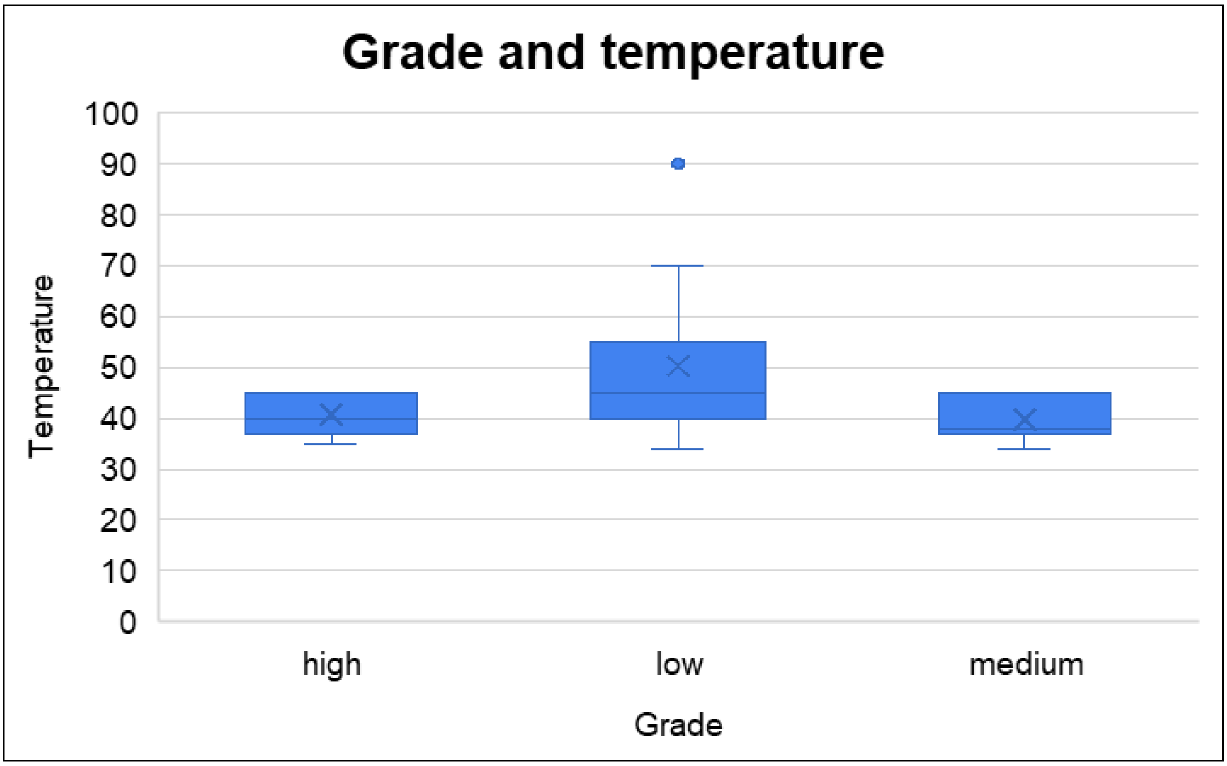

Figure 8: Grade and Temperature

The box plot illustrating the relationship between milk grade and temperature reveals distinct temperature distributions across different quality levels. High-grade milk is tightly controlled, with a median temperature around 45 degrees, indicating consistent management. Low-grade milk, however, has a higher median temperature of approximately 50 degrees and a wider range, including a high-temperature outlier, suggesting poor temperature control. Medium-grade milk exhibits the lowest median temperature, around 40 degrees, and is also tightly controlled. The plot indicates that higher temperatures are associated with lower quality, with low-grade milk being warmer and more variable, while medium-grade milk is kept the coolest. Temperature serves as a strong secondary predictor for distinguishing between milk grades.

3. Grade and taste

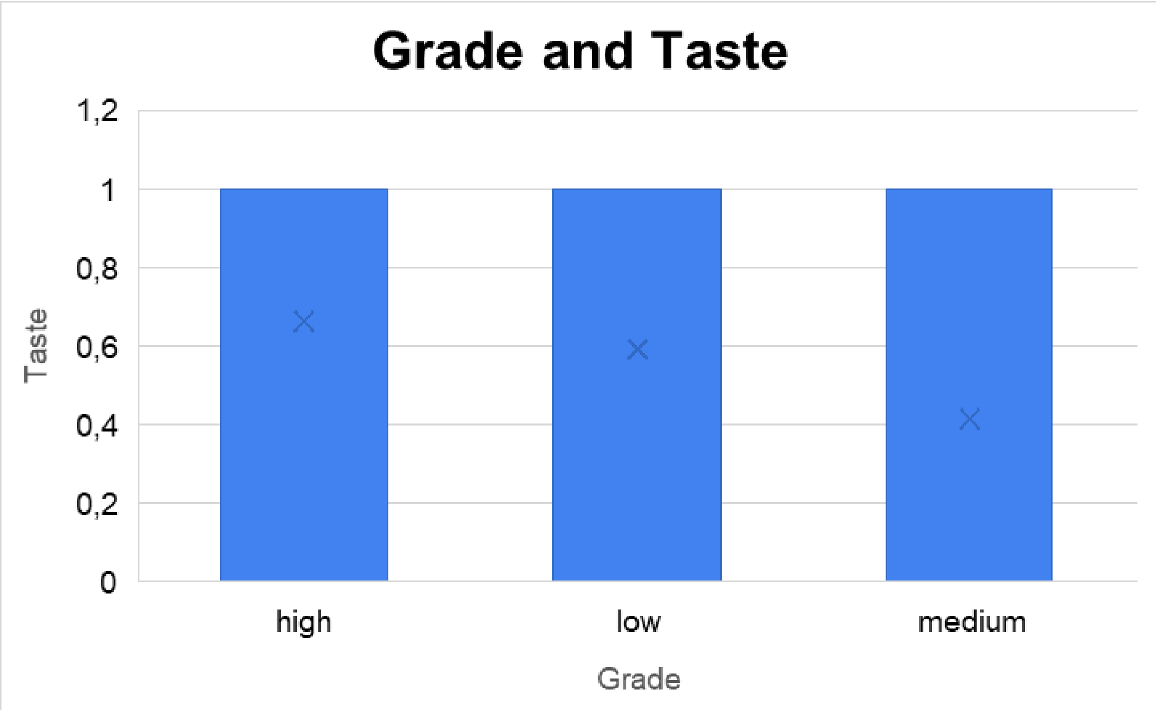

Figure 9: Grade and Taste

The box plot illustrating the relationship between milk grade and taste reveals that both high-grade and low-grade milk have a median taste value of 1, indicating that over 50% of samples in these categories were rated as having "good" taste. Medium-grade milk also has a median taste value of 1, but its mean ('x') is lower, suggesting a higher proportion of "bad" tasting samples compared to the high and low grades. The trend pattern shows that taste is a weak and counter-intuitive predictor of milk quality, as low-grade milk does not consistently exhibit bad taste according to this data, making it an ineffective metric for distinguishing between grades.

4. Grade and Odor

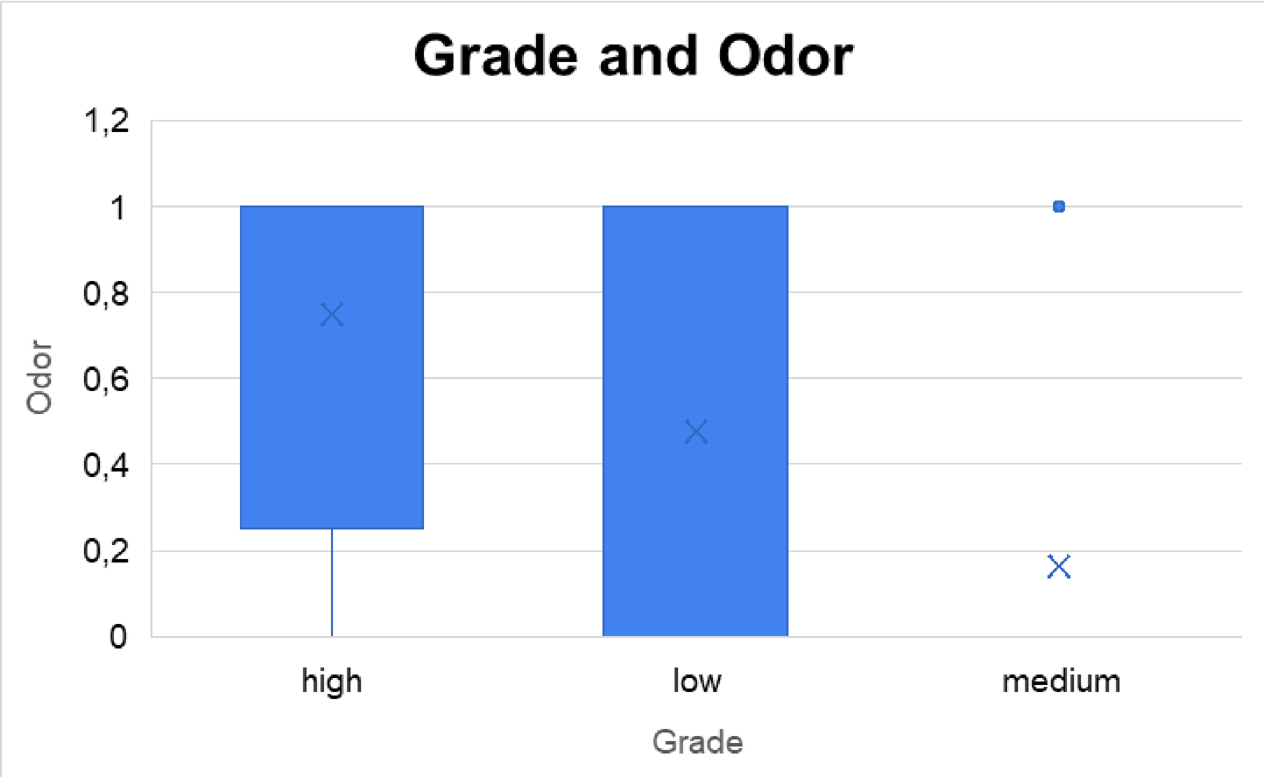

Figure 10: Grade and Odor

The box plot illustrating the relationship between milk grade and odor shows distinct patterns across quality levels. High-grade milk has a high median and mean odor value, indicating consistently good odor across samples. Low-grade milk has a lower mean compared to high-grade, suggesting more instances of bad odor. Medium-grade milk is particularly distinctive, with a median odor value of 0 and a very low mean, indicating that the vast majority of samples have a "bad" odor. The trend pattern highlights odor as an excellent predictor of milk quality, especially for identifying medium-grade milk, where a bad odor strongly indicates medium quality.

5. Grade and Fat

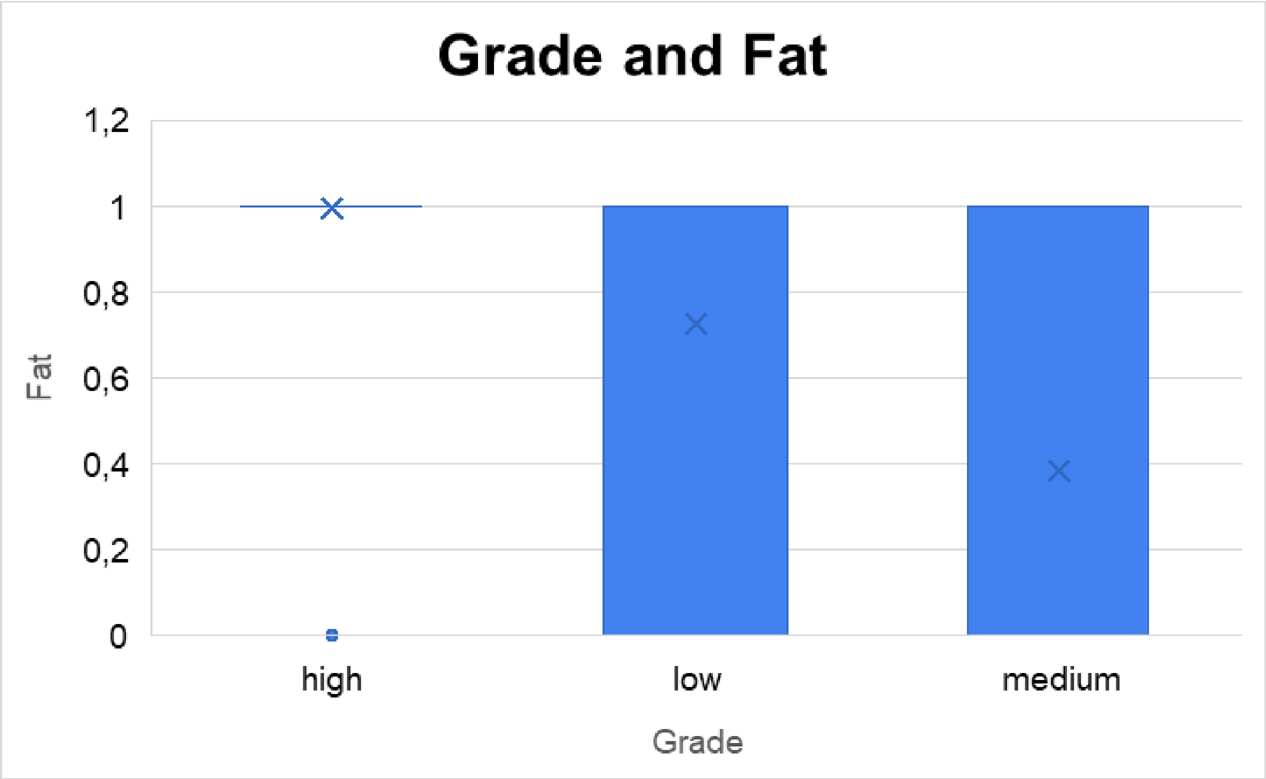

Figure 11: Grade and Fat

The box plot illustrating the relationship between milk grade and fat content reveals distinct patterns across quality levels. High-grade milk is highly consistent, with almost all samples having a fat value of 1, as indicated by the box plot collapsing into a single line at 1, where the 25th, 50th, and 75th percentiles align. Low-grade milk has a median fat value of 1 but shows more variability compared to high-grade milk. Medium-grade milk exhibits significant variability, with a nearly 50/50 split between fat values of 0 and 1. The trend pattern suggests that fat content is a good predictor of milk quality, with high-grade milk characterized by consistent fat content (value=1) and medium-grade milk distinguished by its inconsistent fat content.

6. Grade and Turbidity

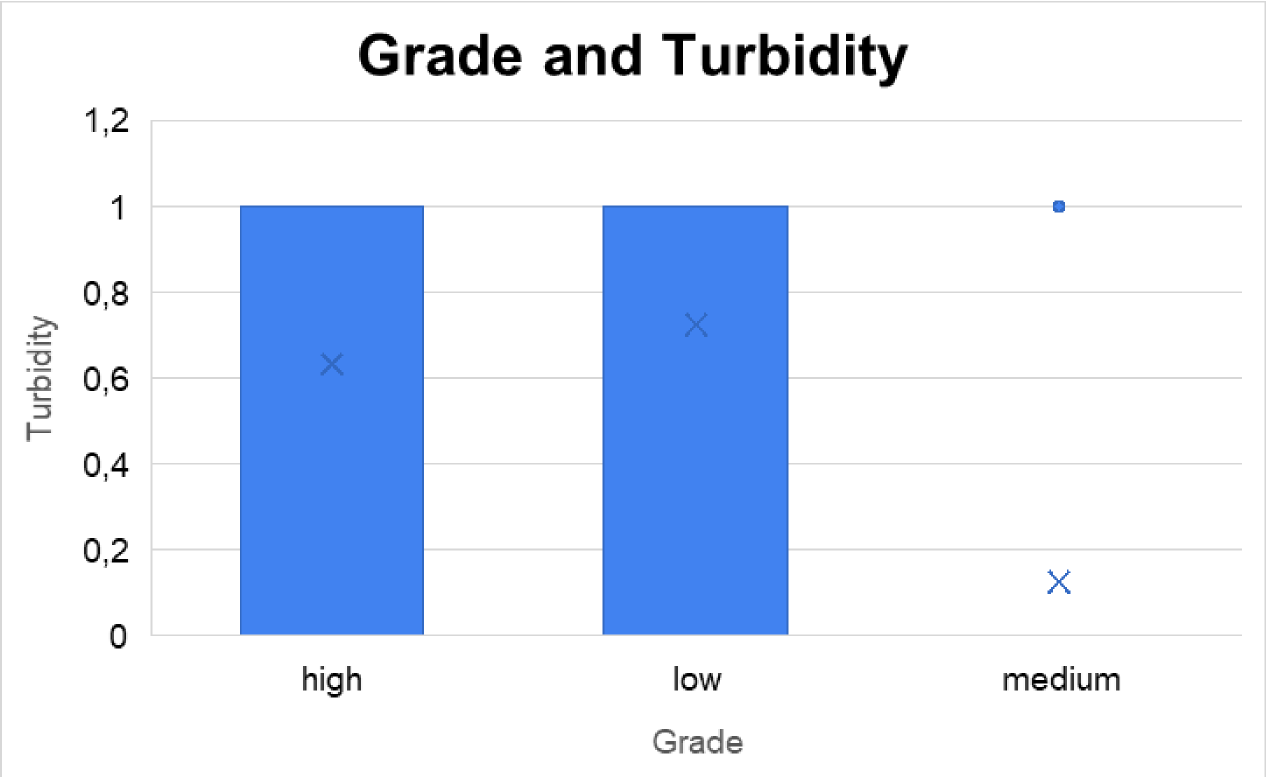

Figure 12: Grade and Turbidity

The chart illustrating the relationship between milk grade and turbidity shows a noticeable pattern. For high and low grades, the turbidity values cluster around 1, indicating good turbidity. In contrast, the medium grade displays a wider variation, with a significant portion of samples showing a turbidity value of 0 (bad). This suggests that turbidity may be a more distinguishing factor for identifying medium-grade milk, as it contains a mix of both good and bad turbidity, whereas high and low grades tend to have consistently good turbidity. Thus, turbidity could be a relevant feature when classifying or predicting the quality of milk.

7. Grade and Colour

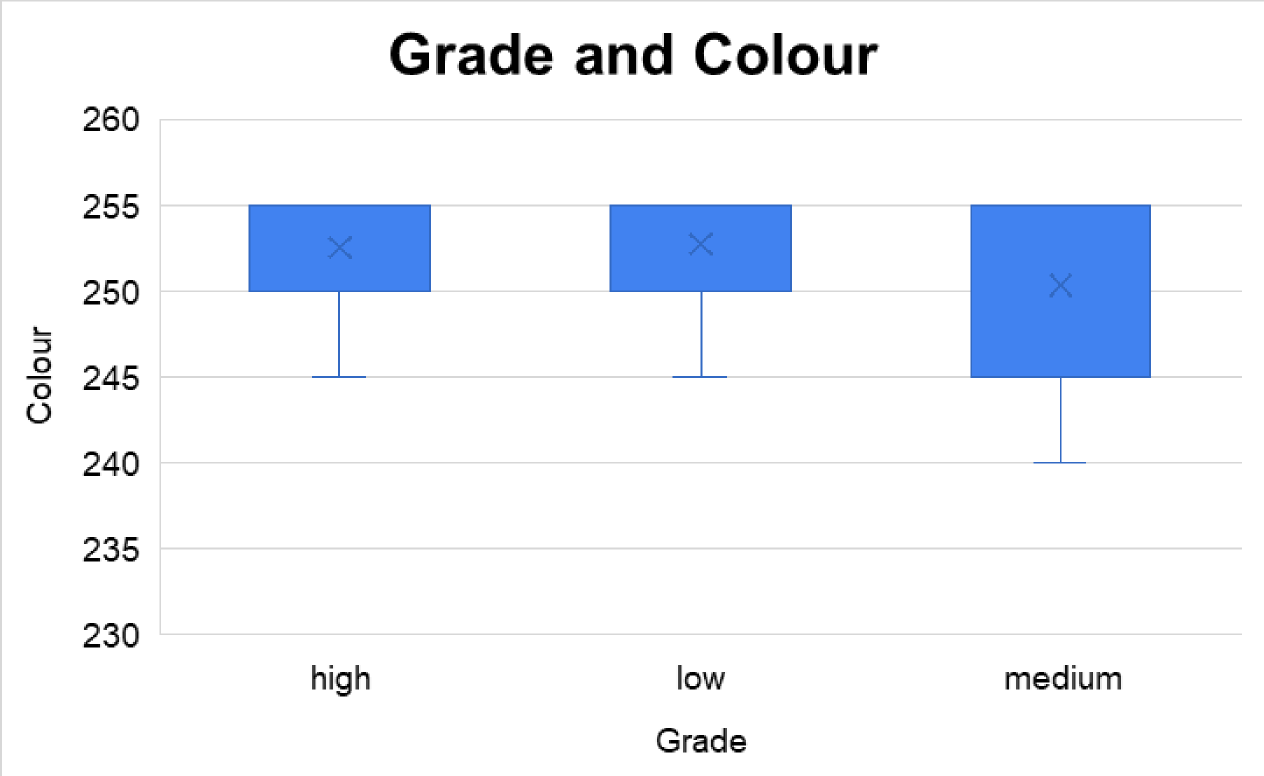

Figure 13: Grade and Color

The chart illustrates the relationship between milk grade and color. All three grades high, low, and medium exhibit color values within a narrow range (approximately 245 to 255). However, medium-grade milk shows greater variation in color, with a wider interquartile range and more spread in the data. In contrast, high and low grades have more consistent color values, tightly clustered around the upper end of the range. This suggests that color is not a strong distinguishing factor between high and low grades, but its variability might help identify medium-grade milk, which lacks consistency in color quality.

V. Distribution and Descriptive Visualizations

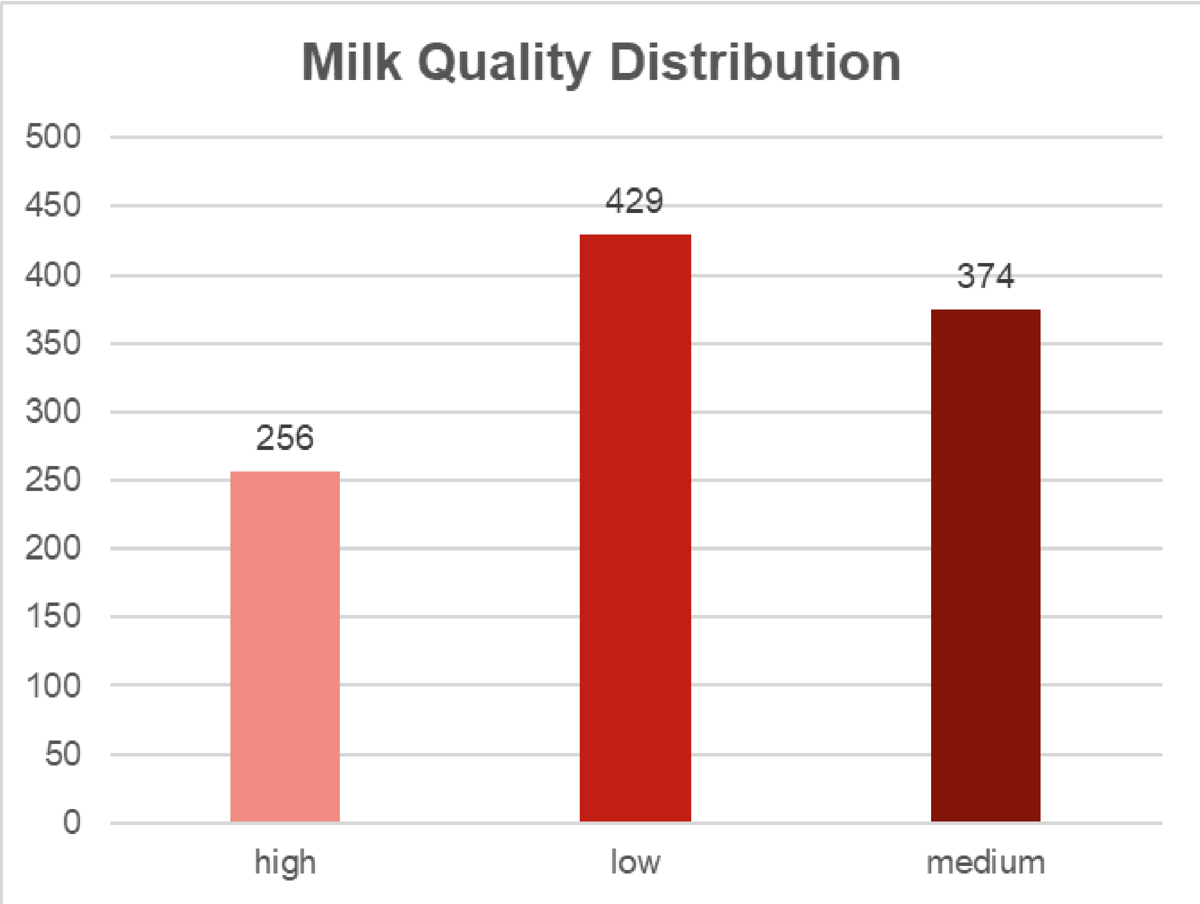

Figure 14: Milk Quality Distribution

The Milk Quality Distribution chart highlights Happy Cow Ltd's strategic emphasis on serving the extensive low and medium-quality market segments. The data shows a clear alignment between their current production and target audience, with 803 samples (75.8%) matching the quality levels acceptable to these customers. The key opportunity lies in achieving consistent and efficient production for low and medium grades while minimizing high-quality milk (256 samples) to control costs. By standardizing processes for these grades, Happy Cow can ensure reliability and cost-effectiveness. Understanding the factors that elevate a batch from low to medium quality allows for targeted, economical adjustments to increase medium-grade output, strengthening the value proposition and fostering brand loyalty in these market segments.

VI. Correlation Analysis

Correlation Matrix

	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
pH	1	0,244684059	-0,064052575	-0,081330527	-0,093428868	0,048384048	-0,164564923
Temprature	0,244684059	1	-0,109791641	-0,048869938	0,024073373	0,18510645	-0,00851136
Taste	-0,064052575	-0,109791641	1	0,017582248	0,324149101	0,055754708	-0,082653594
Odor	-0,081330527	-0,048869938	0,017582248	1	0,314504654	0,457935182	-0,039361312
Fat	-0,093428868	0,024073373	0,324149101	0,314504654	1	0,329263584	0,11415082
Turbidity	0,048384048	0,18510645	0,055754708	0,457935182	0,329263584	1	0,136436137
Colour	-0,164564923	-0,00851136	-0,082653594	-0,039361312	0,11415082	0,136436137	1

Figure 15: Correlation Matrix

pH & Temperature (r = 0.24): There is a slight positive association indicating that as temperature increases, pH levels tend to move away from neutral.

pH & Other Attributes: The correlations between pH and attributes such as Taste (-0.06), Odor (-0.08), Fat (-0.09), and Turbidity (+0.05) are negligible. This suggests pH is more of an indicator of spoilage rather than directly affecting sensory qualities. Significant pH variation in low-quality milk points to fundamental processing issues unrelated to other features.

Taste & Fat (r = +0.32): A moderate positive correlation shows that adequate fat content is crucial for good taste, enhancing the richness and flavor. However, in lower-quality milk, even correct fat levels don't guarantee a good taste due to adverse effects from spoilage factors like high temperature and abnormal pH.

Odor & Turbidity (r = +0.46): The strongest correlation demonstrates a direct link between clarity and smell. High turbidity, often due to bacterial spoilage or poor homogenization, correlates with poor odor, reinforcing that clear milk likely has a better odor, serving as an effective diagnostic indicator.

Fat, Odor, and Turbidity: Moderate positive links connect Fat to Odor (+0.31) and Turbidity (+0.33), forming a beneficial cluster. Maintaining optimal fat content positively impacts this group, enhancing overall milk quality.

Temperature & Taste (r = -0.110): A slight negative correlation suggests that higher temperatures slightly reduce the likelihood of favorable taste ratings, linking poor processing to an unsatisfactory customer experience.

Temperature & Turbidity (r = 0.185): Higher temperatures are slightly associated with increased turbidity, emphasizing the importance of maintaining cooler temperatures (around 40°C or lower) to ensure milk's physical and sensory quality remains stable.

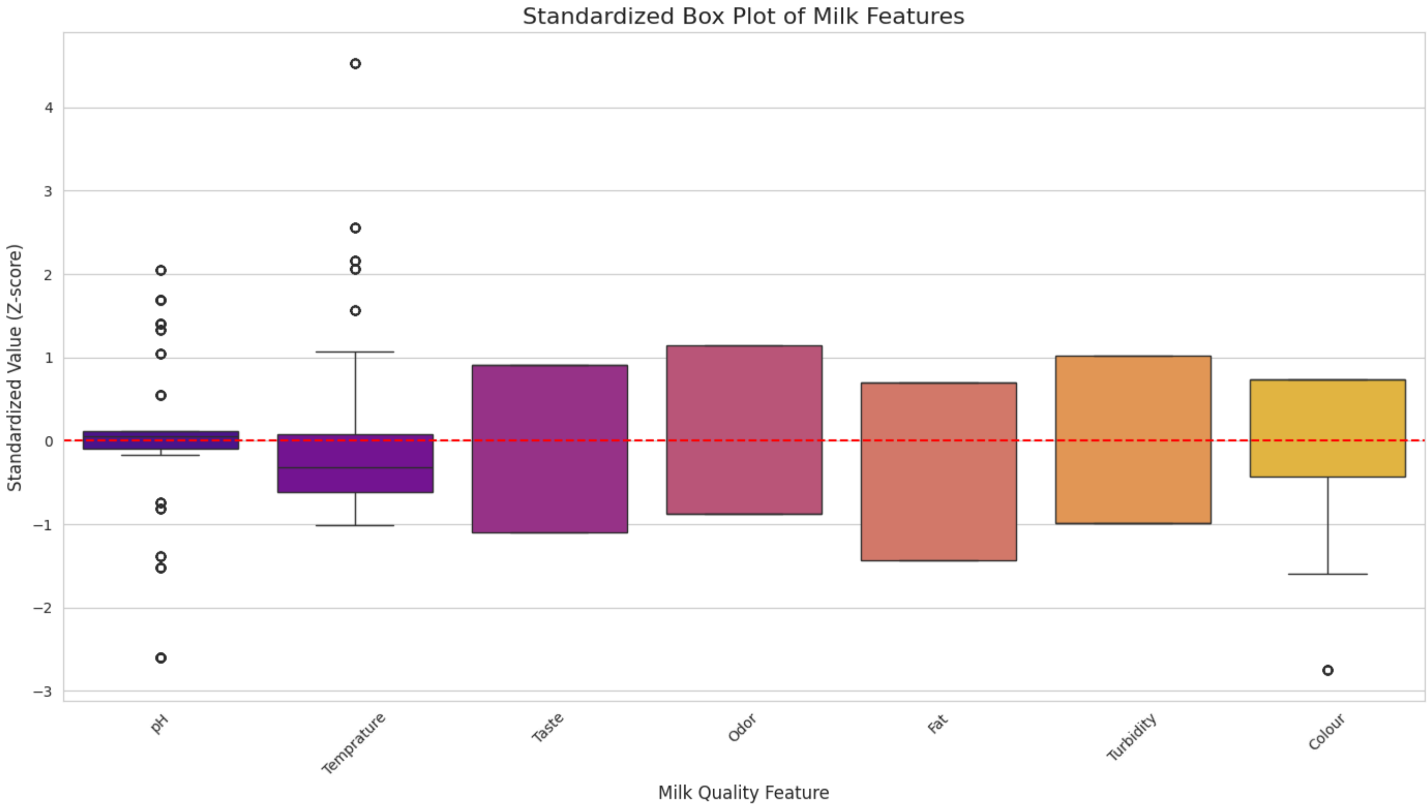

Figure 16: Standardized Box Plot of Milk Features

The Correlation Matrix indicates a slight but noteworthy positive correlation between Temperature and pH (0.24). While this figure alone isn't concerning, its true significance is illustrated by the Standardized Box Plot. The box plot reveals that these two features are the most unstable processes, both showing significant variability and numerous outliers. This instability in temperature regulation likely drives the instability in pH levels. Uncontrolled high temperatures can boost bacterial growth and chemical reactions, causing the pH to deviate from its optimal range. The Box Plot illustrates this disorder, and the Correlation Matrix confirms their connection. Thus, stabilizing Temperature becomes top priority, as it will directly stabilize pH levels, addressing the root of quality challenges.

The Correlation Matrix highlights the strongest relationship: a substantial positive correlation of 0.46 between Odor and Turbidity. This suggests that a batch of milk with a bad odor is very likely to be turbid (cloudy), and vice versa. The Standardized Box Plot for both features shows distributions including both "good" and "bad" outcomes, indicating areas needing improvement. This relationship offers a significant strategic benefit. While Odor is subjective, Turbidity is quantifiable and can be quickly assessed through visual inspection or basic equipment. Therefore, Turbidity can serve as an effective early indicator for Odor issues. If a process check reveals a turbid batch, it can be immediately flagged for further sensory evaluation or removal, preventing substandard products from reaching customers. This correlation serves as a cost-effective quality assurance shortcut.

The Correlation Matrix also shows that Fat is moderately correlated with Taste (0.32) and Odor (0.31). These connections are crucial since Fat is a compositional attribute that can be precisely controlled, whereas Taste and Odor are the final sensory experiences for the customer. The Standardized Box Plot indicates inconsistent outcomes for all three features. This provides a clear opportunity for improvement. By focusing on the specific, measurable action of standardizing Fat content, the data demonstrates that they can directly and positively influence both the taste and aroma of the end product. This approach transforms a subjective goal into an objective, data-driven process.

VII. Multinomial Logistic Regression (MLR)

1. Case Processing Summary

Case Processing Summary			
		N	Marginal Percentage
Grade	high	256	24.2%
	low	429	40.5%
	medium	374	35.3%
Taste	0	480	45.3%
	1	579	54.7%
Odor	0	601	56.8%
	1	458	43.2%
Fat	0	348	32.9%
	1	711	67.1%
Turbidity	0	539	50.9%
	1	520	49.1%
Valid		1059	100.0%
Missing		0	
Total		1059	
Subpopulation		83[a]	

a. The dependent variable has only one value observed in 83 (100.0%) subpopulations.

Table 1: Case Processing Summary

Table 1 shows the number of observations (N) and the marginal percentage for each category of the different variables. There are a total of 1059 valid cases (milk samples) in this dataset, meaning all these samples have complete information for the variables listed (Grade, Taste, Odor, Fat, Turbidity).

Grade (Target Variable - Milk Quality) is the variable to predict. It has three categories: high: 256 samples (24.2%), low: 429 samples (40.5%), medium: 374 samples (35.3%) The "low" quality milk is the most prevalent category, followed by "medium" and then "high." This indicates a class imbalance in the target variable. "Low" quality samples are almost twice as common as "high" quality samples. Taste is fairly balanced, with 480 samples (45.3%) rated as bad and 579 samples (54.7%) rated as good, showing a slight tilt toward good taste. Odor, however, leans toward negative ratings, with 601 samples (56.8%) classified as bad and 458 samples (43.2%) as good. Fat content shows a significant imbalance, with 348 samples (32.9%) rated as bad and a substantial 711 samples (67.1%) rated as good, indicating that over two-thirds of the samples have favorable fat content. Turbidity is nearly evenly split, with 539 samples (50.9%) rated as bad and 520 samples (49.1%) as good, reflecting a very balanced distribution.

Subpopulation Note: "The dependent variable has only one value observed in 83 (100.0%) subpopulations." This is a crucial warning. It suggests that for 83 specific combinations of predictor variable levels, all cases fell into a single grade category. This can sometimes lead to issues like perfect separation or overly optimistic model fit statistics for those specific combinations, potentially affecting the stability of some coefficients.

2. **Model Fitting Information**

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	2280,836			
Final	854,333	1426,503	14	<,001

Table 2: Model Fitting Information

-2 Log Likelihood: The final model (-2LL = 854.333) is a significant improvement over the intercept-only model (-2LL = 2280.836). Likelihood Ratio Test (Chi-Square): The Chi-Square value is 1426.503 with 14 degrees of freedom (df), and the significance (Sig.) is < .001. This highly significant result indicates that the model including the predictor variables (pH, Temperature, Colour, Taste, Odor, Fat, Turbidity) is significantly better at predicting milk grade than a model with no predictors (i.e., just predicting based on the overall proportions of grades). The predictors collectively contribute significantly to explaining the variation in milk grade.

3. **Pseudo R-Square**

Pseudo R-Square

Cox and Snell	,740
Nagelkerke	,837
McFadden	,625

Table 3: Pseudo R-Square

These values suggest how much of the "variance" in milk grade is explained by the model. Nagelkerke's R-square is often preferred as it can range from 0 to 1. A value of .837 is very high, indicating that the model has excellent explanatory power. The predictors account for a substantial portion of the variability in milk quality.

4. **Likelihood Ratio Tests for Individual Effect**

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	854,333[a]	,000	0	.
pH	956,964	102,631	2	<,001
Temperature	1456,182	601,849	2	<,001
Colour	1098,660	244,327	2	<,001
Taste	1058,761	204,428	2	<,001
Odor	1052,851	198,518	2	<,001
Fat	1182,045	327,712	2	<,001
Turbidity	1076,002	221,669	2	<,001

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Table 4: Likelihood Ratio Tests for Individual Effects

The table 4 shows the results of Likelihood Ratio Tests for a model predicting milk quality (Grade) based on the listed effects (pH, Temperature, Colour, Taste, Odor, Fat, Turbidity).

In simple terms, -2 Log Likelihood (-2LL) is a measure of how poorly a model predicts the outcome. Think of it as an "error" or "badness-of-fit" score. A lower -2LL value is better. It means the model fits the data more closely. A higher -2LL value is worse. It means the model is doing a poor job of explaining the data.

In this test, for each row, a "reduced model" is created. This is a temporary model where the variable listed in that row has been removed. For the pH row, the "Reduced Model" is a model with every variable except pH. For the Temperature row, the "Reduced Model" is a model with every variable except Temperature. The "-2 Log Likelihood of Reduced Model" Column shows the error score (-2LL) for each of those temporary "reduced models." Look at the Intercept row. Its -2 Log Likelihood is 854,333. This is the error score for the final, full model that includes all seven predictors. This is our best possible model and serves as our baseline for comparison. Now, let's look at the other rows. pH: When we remove pH from the model, the error score (-2LL) jumps from 854,333 to 956,964. This increase in error tells us that pH was doing a useful job. The model got worse without it. When we remove Temperature, the error score skyrockets to 1,456,182. This is a massive increase in error, indicating that Temperature is an extremely important predictor. The model is significantly crippled without it.

By comparing the values in this column, we can rank the importance of the variables. A higher value means removing that variable caused more "damage" (i.e., increased the error more), which means the variable is more important.

Ranking by "-2 Log Likelihood of Reduced Model" (from highest/most important to lowest/least important):

Temperature: 1,456,182 (Removing it hurts the model the most)

Fat: 1,182,045

Colour: 1,098,660

Turbidity: 1,076,002

Taste: 1,058,761

Odor: 1,052,851

pH: 956,964 (Removing it hurts the model the least, but still significantly)

It provides the same conclusion as the Chi-Square column—that Temperature and Fat are the most dominant predictors, but all seven factors are essential for an accurate model of milk quality. For Chi-Square Statistic, this value indicates the magnitude of the difference in the model's fit when an effect is removed. A higher Chi-Square value suggests that the variable has a stronger impact on the model.

Temperature (601,849) has the highest Chi-Square value, suggesting it is the most influential predictor of milk quality among the listed variables in this model. Fat (327,712) is the next most influential.

Colour (244,327) follows. Turbidity (221,669), Taste (204,428), Odor (198,518), pH (102,631) has the lowest Chi-Square value among the significant predictors, but it is still highly significant.

In conclusion, to assess or ensure good milk quality, it is crucial to monitor and control all these seven parameters, as each one significantly impacts the final grade of the milk according to this statistical analysis.

Significance (Sig. column): For every single independent variable (pH, Temperature, Colour, Taste, Odor, Fat, and Turbidity), the significance value (p-value) is "<.001". This is a very small p-value. This indicates that each of these variables makes a statistically significant contribution to the model predicting milk quality. In other words, changes in each of these factors are significantly associated with changes in the assessed milk quality. We can reject the null hypothesis that these variables have no effect on milk quality.

5. Parameter Estimates

Parameter Estimates

Grade[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
high	Intercept	-184,734	22,643	66,562	1	<,001			
	pH	,014	,163	,008	1	,930	1,014	,737	1,397
	Temperature	,185	,054	11,912	1	<,001	1,203	1,083	1,337
	Colour	,736	,086	73,195	1	<,001	2,087	1,763	2,470
	[Taste=0]	-3,626	,414	76,620	1	<,001	,027	,012	,060
	[Taste=1]	0[b]	.	.	0	.	.	.	.
	[Odor=0]	-5,879	,575	104,372	1	<,001	,003	,001	,009
	[Odor=1]	0[b]	.	.	0	.	.	.	.
	[Fat=0]	-10,811	1,207	80,275	1	<,001	2,018E-5	1,896E-6	,000
	[Fat=1]	0[b]	.	.	0	.	.	.	.
	[Turbidity=0]	-3,221	,531	36,843	1	<,001	,040	,014	,113
	[Turbidity=1]	0[b]	.	.	0	.	.	.	.
low	Intercept	-217,720	22,279	95,502	1	<,001			
	pH	-,728	,138	28,022	1	<,001	,483	,369	,632
	Temperature	,606	,052	134,016	1	<,001	1,833	1,654	2,031
	Colour	,816	,084	93,410	1	<,001	2,261	1,916	2,667
	[Taste=0]	-5,370	,475	127,822	1	<,001	,005	,002	,012
	[Taste=1]	0[b]	.	.	0	.	.	.	.
	[Odor=0]	-3,382	,544	38,586	1	<,001	,034	,012	,099
	[Odor=1]	0[b]	.	.	0	.	.	.	.
	[Fat=0]	-2,524	,411	37,763	1	<,001	,080	,036	,179
	[Fat=1]	0[b]	.	.	0	.	.	.	.
	[Turbidity=0]	-5,213	,489	113,818	1	<,001	,005	,002	,014
	[Turbidity=1]	0[b]	.	.	0	.	.	.	.

a. The reference category is: medium.

b. This parameter is set to zero because it is redundant.

Table 5: Parameter Estimates

The reference category for the dependent variable "Grade" is **medium**. So, the coefficients (B) and Odds Ratios (Exp(B)) are interpreted in relation to the "medium" grade.

For binary predictors (Taste, Odor, Fat, Turbidity), one category is shown (e.g., [Taste=0]), and the other (e.g., [Taste=1]) is the reference for that predictor. 0 means "bad" or "low" attribute and 1 means "good" or "high" attribute for Taste, Odor, Fat. For Turbidity, 0 likely means "low/clear" and 1 means "high/cloudy".

The key columns for thi analysis are:

Sig.: To determine if a factor's effect is statistically real (we'll use $p < 0.05$ as the threshold).

Exp(B): The odds ratio, showing how much a factor increases (>1) or decreases (<1) the odds of that grade.

95% Confidence Interval (CI): The range where we are 95% confident the true odds ratio lies. If this range does not include 1.0, the result is significant. A narrow range indicates a more precise estimate.

A. Predicting "High" Grade vs. "Medium" Grade

Intercept: -184.734 (baseline log-odds)

pH: B = 0.014, Sig. = .930. Not a significant predictor for distinguishing 'high' from 'medium' grade. (Exp(B) = 1.014, close to 1). The 95% CI [0.737, 1.397] contains 1.0, confirming that we cannot confidently say whether pH increases or decreases the odds.

Temperature: B = 0.185, Sig. < .001, Exp(B) = 1.203 with a 95% CI [1.763, 2.470].. For each unit increase in temperature, the odds of milk being 'high' grade versus 'medium' grade increase by a factor of 1.203 (or 20.3%). Higher temperatures are associated with 'high' grade relative to 'medium'.

Colour: B = 0.736, Sig. < .001, Exp(B) = 2.087 with a 95% CI [1.763, 2.470]. For each unit increase in colour score, the odds of milk being 'high' grade versus 'medium' grade increase by a factor of 2.087 by 108.7% (i.e., more than double). Higher colour scores strongly favor 'high' grade over 'medium'.

Taste [Taste=0] (Bad Taste): B = -3.626, Sig. < .001, Exp(B) = 0.027 with a 95% CI [0.012, 0.060]. Milk with Taste=0 (bad taste) has 0.027 times the odds of being 'high' grade versus 'medium' grade compared to milk with Taste=1 (good taste). In other words, "good taste" (Taste=1) dramatically increases the odds of being 'high' grade compared to 'medium'.

Odor [Odor=0] (Bad Odor): B = -5.879, Sig. < .001, Exp(B) = 0.003 with a 95% CI [0.001, 0.009]. Milk with Odor=0 (bad odor) has 0.003 times the odds of being 'high' grade versus 'medium' grade compared to milk with

Odor=1 (good odor). This is a 99.7% decrease in odds. Bad odor very strongly reduces the likelihood of high quality.

Fat [Fat=0] (Bad Fat Content): B = -10.811, Sig. < .001, Exp(B) = 2.018E-5 (effectively 0) with a 95% CI [1.896E-6, .000]. Milk with Fat=0 (low fat) has virtually zero odds of being 'high' grade versus 'medium' grade compared to milk with Fat=1 (high fat). "High fat" (Fat=1) is extremely strongly associated with 'high' grade.

Turbidity [Turbidity=0] (Bad Turbidity): B = -3.221, Sig. < .001, Exp(B) = 0.040 with a 95% CI [0.014, 0.113]. Milk with Turbidity=0 (low turbidity/clear) has 0.040 times the odds of being 'high' grade versus 'medium' grade compared to milk with Turbidity=1 (high turbidity/cloudy). This is a 96% decrease in odds. Bad turbidity strongly reduces the likelihood of high quality.

B. Predicting "Low" Quality Milk (vs. Medium Quality)

Intercept: -217.720 (baseline log-odds)

pH: B = -0.728, Sig. < .001, Exp(B) = 0.483 with a 95% CI [0.369, 0.632]. For each unit increase in pH, the odds of milk being 'low' grade versus 'medium' grade decrease by a factor of 0.483 (i.e., are 48.3% of what they were). Higher pH makes it less likely to be 'low' grade (favoring 'medium' or 'high'). Conversely, lower pH increases the odds of being 'low' grade. The confidence interval is entirely below 1.0, confirming this.

Temperature: B = 0.606, Sig. < .001, Exp(B) = 1.833 with a 95% CI [1.654, 2.031]. For each unit increase in temperature, the odds of milk being 'low' grade versus 'medium' grade increase by a factor of 1.833. Higher temperatures are strongly associated with 'low' grade relative to 'medium'.

Colour: B = 0.816, Sig. < .001, Exp(B) = 2.261 with a 95% CI [1.916, 2.667]. For each unit increase in colour score, the odds of milk being 'low' grade versus 'medium' grade increase by a factor of 2.261. Higher colour scores strongly favor 'low' grades over 'medium'. This is interesting: higher colour also favored 'high' over 'medium', suggesting 'medium' grade milk has the lowest colour scores.

Taste, Odor, Fat, and Turbidity:

All these "bad" characteristics (Taste=0, Odor=0, etc.) have Exp(B) values that are very small and CIs that are entirely below 1.0. For example, for Taste=0, Exp(B) = 0.005.

Taste [Taste=0] (Bad Taste): B = -5.370, Sig. < .001, Exp(B) = 0.005. Milk with Taste=0 ("bad taste") has 0.005 times the odds of being 'low' grade versus 'medium' grade compared to milk with Taste=1 ("good taste"). "Good taste" (Taste=1) strongly makes it less likely to be 'low' grade (favoring 'medium' or 'high').

Odor [Odor=0] (Bad Odor): B = -3.392, Sig. < .001, Exp(B) = 0.034. Milk with Odor=0 ("bad odor") has 0.034 times the odds of being 'low' grade versus 'medium' grade compared to milk with Odor=1 ("good odor"). "Good odor" (Odor=1) strongly makes it less likely to be 'low' grade.

Fat [Fat=0] (Bad Fat Content): B = -2.524, Sig. < .001, Exp(B) = 0.080. Milk with Fat=0 ("low fat") has 0.080 times the odds of being 'low' grade versus 'medium' grade compared to milk with Fat=1 ("high fat"). "High fat" (Fat=1) makes it less likely to be 'low' grade.

Turbidity [Turbidity=0] (Bad Turbidity): B = -5.213, Sig. < .001, Exp(B) = 0.005. Milk with Turbidity=0 ("low turbidity/clear") has 0.005 times the odds of being 'low' grade versus 'medium' grade compared to milk with Turbidity=1 ("high turbidity/cloudy"). "High turbidity" (Turbidity=1) strongly makes it less likely to be 'low' grade (favoring 'medium' or 'high').

Summary of Trends for Predicting Milk Quality:

To get "High" Grade (vs. Medium):

- **Strongly Favors High:** Good Taste (Taste=1), Good Odor (Odor=1), High Fat (Fat=1), High Turbidity (Turbidity=1), Higher Colour.
- **Moderately Favors High:** Higher Temperature.
- **No significant effect:** pH.

To get "Low" Grade (vs. Medium):

- **Strongly Favors Low:** Higher Temperature, Higher Colour, Lower pH.
- **Strongly Protects Against Low (i.e., favors Medium/High):** Good Taste (Taste=1), Good Odor (Odor=1), High Fat (Fat=1), High Turbidity (Turbidity=1).

General Patterns:

Good sensory attributes (Taste=1, Odor=1) and high fat (Fat=1) strongly push milk towards "high" grade and away from "low" grade.

High Turbidity (Turbidity=1) also pushes milk towards "high" grade and away from "low" grade. This might be correlated with fat content (e.g., high-fat milk is more turbid).

Temperature: Has a divergent effect. Higher temperatures increase the odds of both 'high' (vs medium) and 'low' (vs medium), but the effect is stronger for 'low' (Exp(B)=1.833 for low vs medium; Exp(B)=1.203 for high vs medium). This suggests an optimal temperature range might exist for 'medium' grade, or that temperature extremes push quality away from medium.

Colour: Similar to temperature, higher colour scores increase odds of both 'high' (vs medium) and 'low' (vs medium). This implies that 'medium' quality milk tends to have lower colour scores.

pH: Lower pH increases the odds of 'low' grade (vs. medium). Higher pH decreases the odds of being 'low' grade. pH doesn't significantly distinguish 'high' from 'medium'.

Recommendations:

The model demonstrates strong predictive power in identifying key drivers of milk quality, with binary predictors such as Taste, Odor, Fat, and Turbidity exerting significant influence. Specifically, the "good" or "high" states of these predictors strongly correlate with higher milk grades, effectively shifting milk away from lower quality classifications. Notably, the association of high turbidity (Turbidity=1) with better grades is intriguing, potentially indicating that high turbidity serves as a proxy for desirable traits, such as elevated fat content, which is also a robust predictor of high quality. However, caution is warranted for predictions involving rare combinations of these factors, as the model's reliability may decrease for such specific subpopulations.

This model provides a solid basis for understanding what factors drive milk quality classification and could be used to predict the grade of new milk samples based on these measured characteristics.

6. Integrating MLR and Correlation Insights for Predicting Milk Quality

The correlation matrix shows the pairwise linear relationships between the predictor variables themselves. This helps us understand:

Potential multicollinearity (if predictors are too highly correlated).

Interdependencies between predictors, which can add context to the MLR results.

- Understanding Turbidity's Positive Role

MLR: High Turbidity strongly increases the odds of "high" grade and decreases the odds of "low" grade (both vs. "medium").

Correlation: Turbidity is moderately positively correlated with Fat (r=0.329) and Odor (r=0.457).

The beneficial effect of high turbidity on milk quality likely stems, in part, from its association with higher fat content (a desirable trait) and better odor. However, since the MLR accounts for Fat and Odor as separate predictors, Turbidity still has a significant *unique* positive contribution to predicting higher quality, even beyond its association with fat and odor. Perhaps it reflects overall solid content or a specific visual appeal linked to richness.

- The Cluster of "Goodness" (Fat, Taste, Odor)

MLR: High Fat, good Taste, and good Odor are all very strong predictors of "high" quality milk and protective against "low" quality.

Correlation: Taste & Fat (r=0.324), Odor & Fat (r=0.314) show moderate positive correlations. Taste & Odor (r=0.017) are nearly uncorrelated.

While fat content is linked to better taste and odor, each of these sensory attributes (Taste, Odor) and the compositional attribute (Fat) brings unique, powerful predictive information to the model. The model benefits from including all three.

- Temperature's Complex Role

MLR: Higher temperature increases odds of *both* "high" (vs. medium) and "low" (vs. medium).

Correlation: Temperature has weak correlations with most other variables (e.g., with pH r=0.244, Turbidity r=0.185).

Temperature's divergent impact on pushing milk away from "medium" (towards either extreme) is not strongly explained by its correlation with other measured predictors. This suggests temperature has a distinct, non-linear influence that might relate to optimal processing/storage ranges or microbial activity affecting different quality grades differently.

- Colour's Complex Role

MLR: Higher colour scores increase odds of *both* "high" (vs. medium) and "low" (vs. medium).

Correlation: Colour has weak correlations with other predictors (e.g., pH r=-0.164, Fat r=0.114, Turbidity r=0.136).

Like temperature, the finding that "medium" quality milk has lower colour scores, while both "high" and "low" have higher scores, seems to be an independent effect of Colour not strongly driven by its relationship with other predictors. This warrants further investigation into what "colour" represents in this context for different grades.

- pH as a Differentiator for "Low" Quality

MLR: Lower pH significantly increases odds of "low" grade (vs. medium). Not significant for "high" vs. "medium."

Correlation: pH has weak negative correlations with Taste (-0.06), Odor (-0.08), Fat (-0.09), and Colour (-0.16).

While lower pH might be slightly associated with poorer sensory attributes or lower fat, its primary role in the model is to distinctly identify milk more likely to be "low" quality. This aligns with common knowledge where acidity (lower pH) is often an indicator of spoilage or poor quality.

Overall, to achieve "High" quality milk, prioritize ensuring good Taste, good Odor, high Fat content, and optimal Turbidity, as these factors strongly drive superior grades. Additionally, maintaining appropriate Colour and Temperature is important, though their ideal ranges appear more complex and require careful management. To avoid "Low" quality milk, it is critical to sustain good Taste, good Odor, high Fat, and optimal Turbidity, while also preventing pH from dropping too low and managing Temperature and Colour to avoid extreme values that could degrade milk quality.

7. Combining Boxplots and Regression Parameters

Key predictors of milk grade are Temperature, Colour, Taste, Odor, Fat, and Turbidity. pH was found to not be a significant predictor. By combining the statistical model with visual analysis from the boxplots, we can create clear profiles for each milk grade.

- High-Grade Milk: Characterized by low temperature, good taste, good odor, low fat, and low turbidity.
- Low-Grade Milk: Characterized by high temperature, high color value, bad taste, bad odor, high fat, and high turbidity.
- Medium-Grade Milk (the reference category): Generally falls between the high and low grades on most metrics.

Combine the visual evidence from the boxplots with the statistical results from the "Parameter Estimates" table to build a profile for each milk grade. The reference category for comparison is "medium" grade.

Profile of High-Grade Milk

(Comparing High Grade vs. Medium Grade)

Temperature: The boxplot clearly shows that high-grade milk has a lower temperature (median ~45°C) than medium-grade milk (median ~48°C). The MLR confirms temperature is a significant predictor.

Taste & Odor: The Exp(B) values for [Taste=0] (.027) and [Odor=0] (.003) are very small. This means having a bad taste or odor makes it extremely unlikely for milk to be high-grade compared to medium-grade.

Trend: High-grade milk must have good taste and good odor.

Fat & Turbidity: The boxplots for "Grade and Fat" and "Grade and Turbidity" show a stark difference. High-grade milk is almost exclusively in the "0" category for both Fat and Turbidity, while medium-grade milk is in the "1" category.

Trend: High-grade milk is characterized by low fat (Fat=0) and low turbidity (Turbidity=0).

pH & Colour: pH is not a significant predictor. While Colour is significant in the overall model, its effect in distinguishing high vs. medium grade is less pronounced visually than the other factors.

Profile of Low-Grade Milk

(Comparing Low Grade vs. Medium Grade)

Temperature: The boxplot shows low-grade milk has the highest temperature (median ~55°C). The MLR's Exp(B) of 1.833 confirms this: for every 1-degree increase in temperature, the odds of milk being "low" grade (vs. "medium") increase by 83.3%.

Trend: High temperature is a strong predictor of low-grade milk.

Colour: The boxplot shows low-grade milk has the highest colour values. The Exp(B) of 2.261 supports this powerfully: a higher colour value more than doubles the odds of the milk being "low" grade vs. "medium".

Trend: High colour value is a strong predictor of low-grade milk.

Taste & Odor: The Exp(B) values for [Taste=0] (.005) and [Odor=0] (.034) are very low. This indicates that having bad taste or bad odor is strongly associated with the "low" grade.

Trend: Low-grade milk is characterized by bad taste (Taste=0) and bad odor (Odor=0).

Fat & Turbidity: The boxplots show that low-grade milk, like medium-grade milk, is predominantly in the "1" category for both Fat and Turbidity. The MLR coefficients confirm this (the Exp(B) values for [Fat=0] and [Turbidity=0] are less than 1, meaning the "0" state makes the "low" grade less likely).

Trend: Low-grade milk is characterized by high fat (Fat=1) and high turbidity (Turbidity=1).

By synthesizing the statistical model and the visual data, we can establish clear, actionable rules to predict milk quality:

1. To identify HIGH-GRADE milk, look for:
    a. Temperature: Low (around 45°C)
    b. Taste & Odor: Good (Value = 1)
    c. Fat & Turbidity: Low (Value = 0)
2. To identify LOW-GRADE milk, look for:
    a. Temperature: High (above 50°C)
    b. Colour: High (darker)
    c. Taste & Odor: Bad (Value = 0)
    d. Fat & Turbidity: High (Value = 1)
3. pH is not a useful feature for distinguishing between milk grades in this dataset.

This combined analysis demonstrates a robust and reliable method for predicting milk quality, with clear patterns emerging from the data.

VIII. Strategic Recommendations

1. Apply automatic classification system according to quality accreditation threshold

Based on the factors that have a strong influence on quality such as FAT, Turbidity, Taste and Odor, the group proposed to develop a simple test table applied directly at the test line. Each lot of milk can be labeled based on the achievement or failure to meet the thresholds such as: turbidity = 1, taste = 1, odor = 1. The lots do not meet the full or low milk group instead of high -end. This helps reduce quality risks and increases resource allocation efficiency.

2. Strictly control the temperature factor in production and storage

Model results show that high temperatures have strong correlation with both cases: low milk (exp (b) = 1,833) and high type milk (exp (b) = 1.203), but negatively affects the classification of "Medium". This shows that the temperature oscillation pushes milk from average, so it is necessary to optimize the stable temperature area around 45 ° C to reduce standard deviation risks.

3. **Standardizing sensory evaluation through internal training**

Because the two elements of Taste and Odor have a strong and clear influence on the classification results (with EXP (B) very low when in a "bad" state), the company recommends that the company should organize regular sensory assessment training courses. This ensures consistency between assessments, especially in the environment expansion environment.

4. **Integrated sensor system in real time**

Factors like Turbidity, Color and Temperature can be automatically monitored by sensors. In particular, turbidity (exp (b) = 0.04 with high grade; = 0.005 with low grade) is an important indicator that is easy to measure, good representation for overall quality. Investing in continuous monitoring systems will help reduce the dependence on craft assessments and increase the reaction quickly.

5. **Prevention of damage through pH supervision**

Although the pH is not meant to distinguish between high and medium milk (Sig = 0.930), it plays a strong role in detecting low milk (exp (b) = 0.483). This shows that the low pH is an indicator of the degradation. Therefore, abnormal pH samples need to be warned early to remove before packaging.

6. **Consider high milk production to avoid wasting**

From the distribution chart, 40.5% are low milk and 35.3% are the average type - the two groups account for the majority of the target market of the business. The production of too much high milk not only increases costs but also does not match the current product positioning. Should optimize the process to increase Medium milk efficiency, reduce costs and increase profits.

IX. Conclusion

This project has proven the effectiveness of the application of data analysis and prediction models in milk quality management. By combining process control techniques (Control Charts) with a polynomial recovery model (MLR), the variables have a significant impact on product quality classification. The results show that factors such as Fat, Odor, Taste, Turbidity have a decisive role in determining high milk, while low pH, high temperatures, and dark colors are typical indicators of low milk. In particular, the pH does not help distinguish the high type but very sensitive to the type of low has provided a deep perspective on the characteristics of poor quality products. With a model with high explanation (Nagelkerke $R^2 = 0.837$), Happy Cow Ltd. Smart testing systems can be applied for quick classification, early detection of risks, thereby reducing losses and improving brand reliability. The implementation of strategic proposals such as temperature monitoring, standardization of sensory evaluation and pH factor control will contribute to building a modern quality control system, consistent with the goal of expanding the market in Binh Duong and new areas.

Reference

Bizhub. (2019, October 18). *Dutch Lady fresh milk quality levels, already very high, just got better.* Corporate News. https://bizhub.vn/dutch-lady-fresh-milk-quality-levels-already-very-high-just-got-better-post310127.html

Dung, M. (2021, April 16). *Độ pH của sữa tươi bao nhiêu là đảm bảo tiêu chuẩn an toàn*. Máy Đo Chuyên Dụng. https://maydochuyendung.com/tin-tuc/thiet-bi-kiem-tra-nuoc/do-ph-cua-sua-tuoi-la-bao-nhieu

Dutch Lady (n.d.). *Chỉ Số Ph - Nhân Tố Kiểm Định "Độ Tươi" Của Sữa Cô Gái Hà Lan*. https://www.dutchlady.com.vn/tin-tuc/52

GeeksforGeeks. (2024, March 12). *Box plot*. GeeksforGeeks. https://www.geeksforgeeks.org/box-plot/

Hessing, T. (2025, June 5). *X Bar S Control Chart*. Six Sigma Study Guide. https://sixsigmastudyguide.com/x-bar-s-chart/

Milk & Health (n.d.). *Questions About Pasteurization* . https://www.milkandhealth.com/en/questions-about-pasteurization/

New Product technologies - Vinamilk. (n.d.). https://www.vinamilk.com.vn/en/improvement-innovation/new-production-technologies

Periasamy, P. (n.d.). *Quality control*. Scribd. https://www.scribd.com/document/391988662/Quality-Control

Sustainable Development - Vinamilk. (n.d.). https://www.vinamilk.com.vn/en/sustainable-development

Vinamilk. (n.d.). *Product safety and quality*. https://www.vinamilk.com.vn/phat-trien-ben-vung/bao-cao/2022/en/product-safety-and-quality.html

Wordpress. (n.d.). *Food product analysis: Dutch Lady Full Cream Milk | UKEssays.com*. UKEssays.com. https://www.ukessays.com/essays/sciences/food-product-analysis-dutch-lady-cream-2575.php