

Data Collection and Preprocessing Phase

Date	18 March 2025
Team ID	LTVIP2025TMID26704
Project Title	Toy Craft Tales: Tableau's Vision into Toy Manufacturer Data
Maximum Marks	3 Marks

Data Quality Report

The dataset is structurally sound with no missing or duplicate records. However, we identified moderate-severity issues in production quantities, sales revenue, and defect reporting. Implementing the proposed resolution plans will ensure comprehensive accuracy and data reliability for further analysis.

Data Quality Overview

- **Total Records:** 500
- **Columns Analyzed:** 17
- **Missing Values:** None
- **Duplicate Records:** None

Key Data Quality Issues and Resolutions

<u>Data Quality Issue</u>	<u>Severity</u>	<u>Resolution Plan</u>
No missing values detected	Low	No action required
No duplicate records	Low	No action required
Outliers in Units_Produced	Moderate	Validate with manufacturing logs; review extreme values.
Outliers in Net_Sales	Moderate	Cross-check with revenue records; verify large variances.
Defective_Units range (50-200)	Moderate	Confirm with quality control team for reporting accuracy.
Discount_Offered (0-20%)	Moderate	Ensure discount accuracy by validating with sales logs.

Technical Insights and Checks

1. Outlier Detection:

- Used descriptive statistics (mean, min, max, and standard deviation) to identify data points exceeding $\pm 3\sigma$ range.

2. Data Consistency:

- Cross-referenced 'Units_Sold' with 'Net_Sales' for logical consistency.

3. Range Validation:

- Verified 'Discount_Offered' values align with policy limits (0-20%).

4. Data Integrity:

- Ensured unique 'Toy_ID' for each record to maintain dataset integrity

Data Set Link: [toy manufacturing sales data.xlsx](#)

By addressing the identified data quality issues through the outlined resolution plans, we will enhance the accuracy and integrity of the toy manufacturing sales dataset. This will provide a solid foundation for reliable analysis, better decision-making, and improved business insights.