# BeerBo Data Analyst Challenge – Analysis Summary

This report summarizes a complete statistical analysis conducted on manufacturing datasets provided by BeerBo Printing.   
The analysis integrates three datasets: ProductionMetric, Quality, and DeviceProperty to extract actionable insights about production efficiency, quality control, and downtime patterns.

## Key Findings

• Line 4 had the highest reject rate (2.54%) but also the highest efficiency (good units/hour).  
• Unplanned downtime accounted for nearly 76% of all recorded downtime.  
• The most frequent cause of unplanned stops was 'Security Alarm'.  
• 'Detected by Max WIP' and 'Cosmetic Defect' were the most common reject reasons.  
• Smart-Logistics device type had the highest reject rate among all device types.

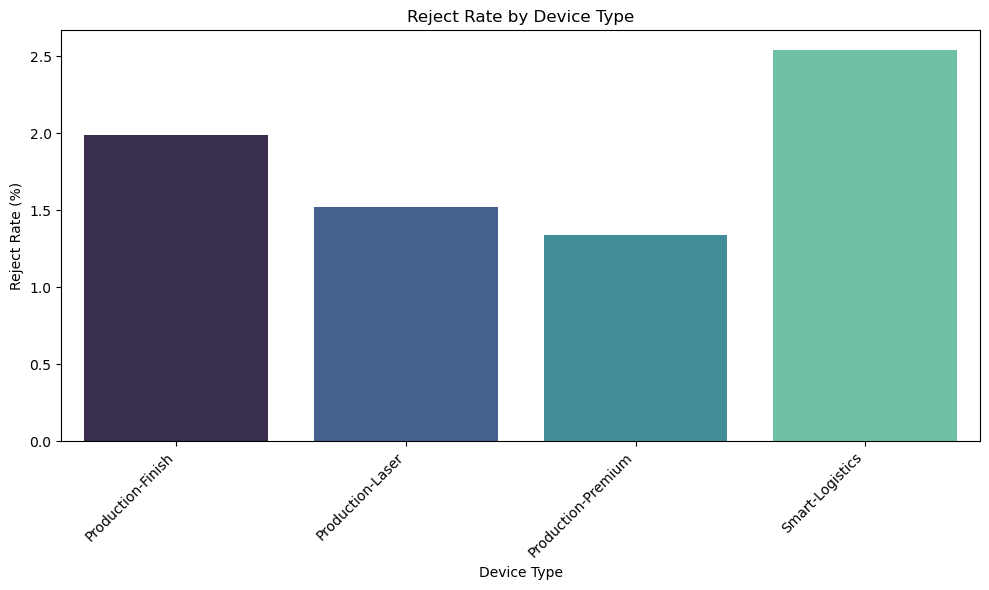
## Methodology

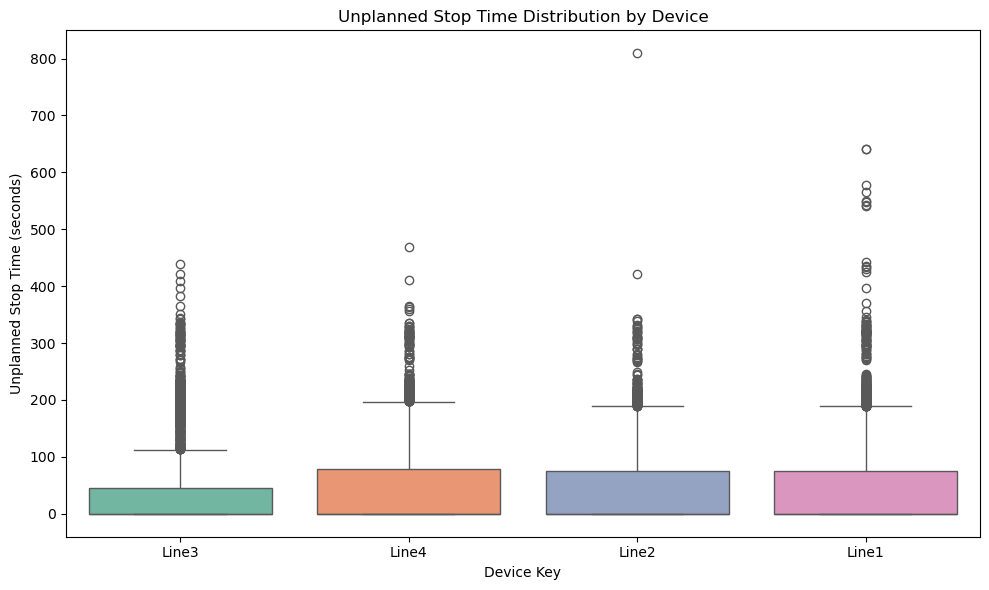
The analysis was performed using a combination of SQL Server for structured queries and Python for data cleaning, merging, and visualization. The datasets were joined using common keys such as deviceKey and prodmetric\_stream\_key.

## Visualizations

The following charts were generated as part of the statistical analysis:

### Bar Chart – Reject Rate by Device



Box Plot – Unplanned Downtime by LineScatter Plot – Unplanned Stop Time vs Reject Count

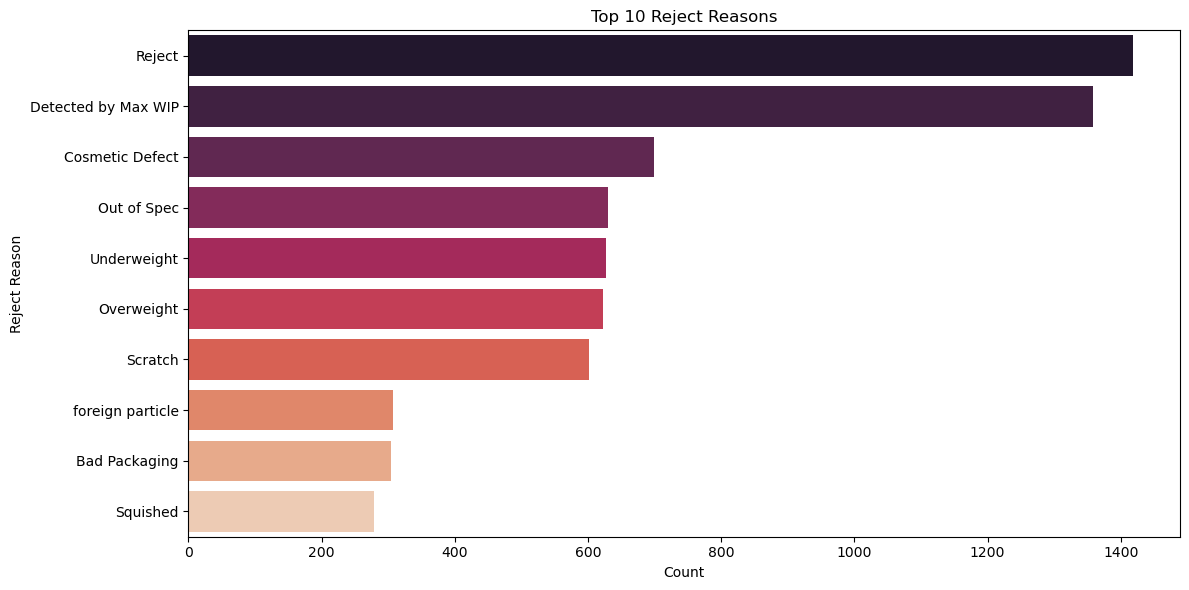
### 

### Histogram – Distribution of Good Count

### Bar Chart – Good Count per Hour by Device

### 

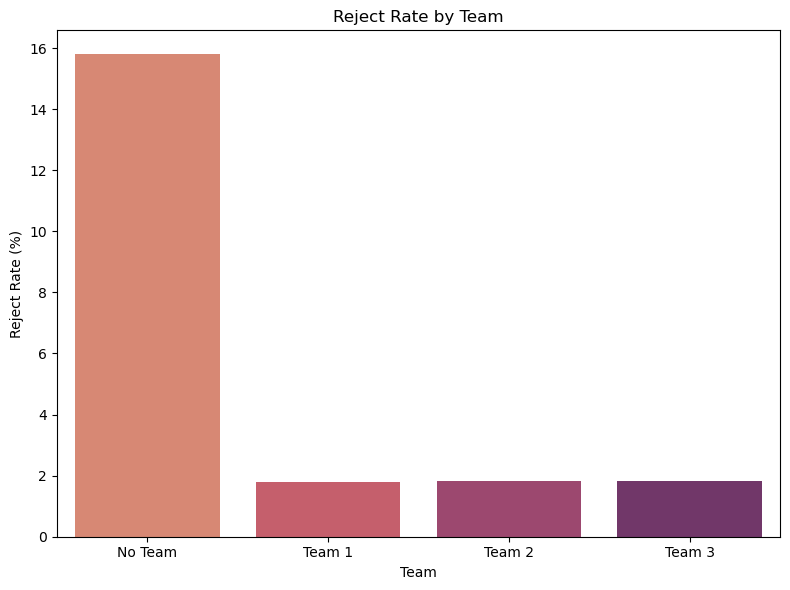
### Bar Chart – Top 10 Reject Reasons



### Bar Chart – Reject Rate by Shift

### 

### Bar Chart – Reject Rate by Team



## SQL Query Summary

All core SQL queries used to analyze the data have been included in the separate file 'beerbo\_sql\_queries.sql'.   
These queries cover basic exploration, downtime analysis, reject rate breakdowns, and joins across the three datasets.