**Predictive Analysis of Production Quality Scores**

**Problem Statement:** The goal of the project was to analyze a manufacturing dataset to understand which factors affect the quality score of produced items. The dataset contained machine performance metrics, environmental conditions, and operational parameters. The challenge was to clean and preprocess the data, explore patterns visually, detect and handle outliers, and finally use regression analysis to identify significant predictors of product quality.

**Data Understanding & Preprocessing:** The dataset contained 1000 observations with features such as Machine ID, Shift, Temperature, Humidity, Speed, Production Time, and Pressure. First, the data was inspected for missing values and anomalies. Then, extreme values (outliers) in continuous variables like Temperature, Humidity, Speed, and Production Time were identified using the Interquartile Range (IQR) method. Outliers were clipped to a maximum or minimum threshold instead of being removed, to preserve data size while reducing the impact of extreme, unrealistic readings (e.g., a sudden spike in temperature from a sensor glitch). Categorical variables such as MachineID and Shift were converted into dummy variables for regression analysis.

**Exploratory Data Analysis (EDA):** Several plots were generated:

1. **Histograms** showed most continuous variables had approximately normal or slightly skewed distributions.
2. **Boxplots** highlighted outliers, particularly in Temperature and Speed, justifying the clipping step.
3. **Scatterplots** revealed negative trends between Temperature/Humidity and QualityScore, suggesting that higher temperatures or humidity levels tend to lower quality.
4. **Bar charts** for MachineID and Shift indicated that certain machines (e.g., M2 and M4) produced consistently lower quality scores compared to others.

**Statistical Modeling (OLS Regression):** An Ordinary Least Squares (OLS) regression model was fitted with QualityScore as the dependent variable. Key results:

* **R-squared = 0.277**, meaning ~27.7% of quality score variation is explained by the included features.
* **MachineID impact:** M2 (-3.90) and M4 (-4.44) significantly reduced quality compared to the baseline machine, while M3 and M5 showed no significant difference.
* **Environmental impact:** Temperature (-0.69) and Humidity (-0.40) had strong, statistically significant negative effects.
* **Operational parameters:** Higher Speed (-0.048) and longer Production Time (-0.147) were associated with lower quality.
* **Shift impact:** No shift showed a statistically significant difference, meaning time of day did not strongly affect quality.
* **Pressure:** Showed a slight positive coefficient (0.41) but was not statistically significant.

**Interpretation of Results:** The findings suggest that machine differences, environmental conditions, and operational speeds all play a role in quality outcomes. Machines M2 and M4 might need maintenance or recalibration. Higher temperatures and humidity levels could require environmental controls. Reducing production speed slightly could improve quality. Although shifts were not significant, consistency in processes across times of day is recommended.

**Conclusion & Recommendations:** The analysis highlights the importance of monitoring machine-specific performance and environmental control in manufacturing quality. To improve product quality:

1. Inspect and maintain underperforming machines.
2. Install better temperature and humidity control systems.
3. Optimize operational speed for quality rather than maximum throughput.
4. Continue monitoring quality trends over time for early detection of performance drops.