## About Dataset

## Introduction

This dataset provides insights into factors influencing defect rates in a manufacturing environment. Each record represents various metrics crucial for predicting high or low defect occurrences in production processes.

## Variables Description

### Production Metrics

**ProductionVolume**: Number of units produced per day.

* **Data Type**: Integer.
* **Range**: 100 to 1000 units/day.  
  **ProductionCost**: Cost incurred for production per day.
* **Data Type**: Float.
* **Range**: $5000 to $20000.

### Supply Chain and Logistics

**SupplierQuality**: Quality ratings of suppliers.

* **Data Type**: Float (%).
* **Range**: 80% to 100%.  
  **DeliveryDelay**: Average delay in delivery.
* **Data Type**: Integer (days).
* **Range**: 0 to 5 days.

### Quality Control and Defect Rates

**DefectRate**: Defects per thousand units produced.

* **Data Type**: Float.
* **Range**: 0.5 to 5.0 defects.  
  **QualityScore**: Overall quality assessment.
* **Data Type**: Float (%).
* **Range**: 60% to 100%.

### Maintenance and Downtime

**MaintenanceHours**: Hours spent on maintenance per week.

* **Data Type**: Integer.
* **Range**: 0 to 24 hours.  
  **DowntimePercentage**: Percentage of production downtime.
* **Data Type**: Float (%).
* **Range**: 0% to 5%.

### Inventory Management

**InventoryTurnover**: Ratio of inventory turnover.

* **Data Type**: Float.
* **Range**: 2 to 10.  
  **StockoutRate**: Rate of inventory stockouts.
* **Data Type**: Float (%).
* **Range**: 0% to 10%.

### Workforce Productivity and Safety

**WorkerProductivity**: Productivity level of the workforce.

* **Data Type**: Float (%).
* **Range**: 80% to 100%.  
  **SafetyIncidents**: Number of safety incidents per month.
* **Data Type**: Integer.
* **Range**: 0 to 10 incidents.

### Energy Consumption and Efficiency

**EnergyConsumption**: Energy consumed in kWh.

* **Data Type**: Float.
* **Range**: 1000 to 5000 kWh.  
  **EnergyEfficiency**: Efficiency factor of energy usage.
* **Data Type**: Float.
* **Range**: 0.1 to 0.5.

### Additive Manufacturing

**AdditiveProcessTime**: Time taken for additive manufacturing.

* **Data Type**: Float (hours).
* **Range**: 1 to 10 hours.  
  **AdditiveMaterialCost**: Cost of additive materials per unit.
* **Data Type**: Float ($).
* **Range**: $100 to $500.

### Target Variable

**DefectStatus**: Predicted defect status.

* **Data Type**: Binary (0 for Low Defects, 1 for High Defects).

### Defect Instances

The dataset focuses on defect instances more because they do not occur often. However, non-defect instances were added too for this reason the dataset is imbalanced, consider balancing it before proceeding with machine learning techniques.

## Data Conclusion

This dataset encompasses a comprehensive collection of metrics vital for predicting defect rates in manufacturing operations. It includes production volumes, supply chain quality, quality control assessments, maintenance schedules, inventory management details, workforce productivity metrics, energy consumption patterns, additive manufacturing specifics, and more.

## Dataset Usage and Attribution Notice

This dataset, shared by Rabie El Kharoua, is original and has never been shared before. It is made available under the CC BY 4.0 license, allowing anyone to use the dataset in any form as long as proper citation is given to the author. A DOI is provided for proper referencing. Please note that duplication of this work within Kaggle is not permitted.

## Exclusive Synthetic Dataset

This dataset is synthetic and was generated for educational purposes, making it ideal for data science and machine learning projects. It is an original dataset, owned by Mr. Rabie El Kharoua, and has not been previously shared. You are free to use it under the license outlined on the data card. The dataset is offered without any guarantees. Details about the data provider will be shared soon.