**Task 2 - Data Analysis and Insights Generation using Python**

**Assignment By –** **Bhushan Dhawas**

**Google Collab link – (**[**Link**](https://colab.research.google.com/drive/14XeQuZgFb4jWEjJlKC9vbk9WfAUUcBKp?usp=sharing)**)**

**Assignment Files –** ([link](https://drive.google.com/drive/folders/1tiJ0XCZorPyZqyQzWGeD8SUvCFQER3dv?usp=sharing))

**My Portfolio Website –** ([link](https://bhushan0016.github.io/My-Portfolio-website/))

**Data Analysis Report:**

**1. Column-Wise Analysis:**

**Analysis -**

The dataset was examined through a **detailed column-wise analysis** to understand its structure and significance. The key columns in the dataset included:

* **Free-text fields:** "Complaint," "Cause," and "Correction," which contained descriptions of failure conditions, affected components, and corrective actions.
* **Predefined categories:** "Root Cause" and "Symptom Condition," which helped standardize the classification of issues and resolutions.

Each column was evaluated for **data type, unique values, and significance.** During the review, inconsistencies in categorical values, missing data, and formatting issues were identified, necessitating a **data cleaning process** before further analysis.

**2. Data Cleaning Summary:**

**Analysis -**

To ensure accuracy and consistency in the dataset, the following **data cleaning steps** were implemented:

* **Handling missing values:** Missing entries were identified and either replaced or marked for further review.
* **Standardizing column names:** Inconsistencies in column names were removed to ensure uniformity.
* **Correcting categorical discrepancies:** Values within categorical columns were cleaned to match predefined taxonomy.
* **Text preprocessing:** Since the dataset primarily consisted of text-based information, **text normalization** techniques such as **lowercasing, whitespace removal, and keyword extraction** were applied.

These steps ensured that the dataset was properly formatted and structured for further analysis.

**3. Visualizations:**

**Analysis –**

**Top 5 Critical Columns & Their Importance**

* + **Date/Time →** Tracks trends, seasonality, and anomalies over time.
  + **Category/Type →** Helps segment data and identify key patterns.
  + **Performance Metric (e.g., Sales, Revenue, Defect Rate) →** Measures success and tracks KPIs.
  + **Customer Feedback/Comments →** Provides qualitative insights via sentiment analysis.
  + **Geographic Location/Region →** Identifies regional trends for better decision-making.

**To identify key trends, visualizations were created for:**

* + **Common failure symptoms and root causes:**
* Recurring **Root Causes** such as "Not Installed," "Not Tightened," and "Loose" highlighted potential weaknesses in assembly and quality control.
* Frequent Symptom Conditions such as "Loose," "Missing," "Crushed," and "Oil Running" pointed to potential maintenance concerns.
  + **Fix Condition trends:**
* "Retightened" and "Installed" were the most common corrective actions, suggesting that many issues could have been prevented through better quality checks and installation procedures.

Visual representations helped to highlight these patterns and provide actionable insights.

**4. Generated Tags & Key Takeaways:**

**Analysis –**

Key tags were extracted from free-text fields to summarize failure patterns and corrective measures. Commonly identified tags included:

1. **Failure-related keywords:** "leaks," "cracks," "damages"
2. **Affected components:** "sensors," "valves," "pressure lines"
3. **Corrective actions:** "replace," "repair," "adjust"

**Key Takeaways:**

* Persistent failures in specific components suggest the need for **improved quality control measures.**
* Contamination issues in sensors, pressure line damage, and improper lubricant usage indicate areas requiring **process improvements.**
* Leakage-related complaints emphasize the importance of **better sealing techniques.**
* Implementing **predictive maintenance strategies** could reduce recurring failures and improve operational efficiency.

By structuring the dataset and analyzing failure trends, the insights generated can aid in **enhancing product reliability, optimizing maintenance strategies, and improving quality control processes.** This analysis has strengthened my understanding of **data processing, text analysis, and structured problem-solving**, reinforcing the value of data-driven decision-making in industrial maintenance and repair operations.

**5.Summary and Insights:**

**Analysis –**

**Tags Generated & Insights:**

* Extracted key **themes** from the free-text column (e.g., failure conditions, customer concerns).
* **Common words** indicate frequently occurring issues or key focus areas.
* Sentiment analysis suggests trends in **positive vs. negative feedback**.

**Actionable Recommendations:**

* **Improve Data Consistency** – Standardize categorical values and remove duplicates.
* **Address Key Trends** – Focus on categories with **high failure rates or customer complaints**.
* **Optimize Performance Metrics** – Identify underperforming regions/products and take corrective actions.
* **Enhance Customer Experience** – Leverage **feedback analysis** to improve service/product quality.

**Data Discrepancies & Approach:**

* **Missing Values** – Handled using **mean/mode imputation** for numerical/categorical data.
* **Inconsistent Formatting** – Fixed by **lowercasing & removing extra spaces**.
* **Outliers** – Removed using **IQR method** for numerical stability.