

Task 2 & 3 - Summary Report

Task 2: Data Preparation and Integration

This task focused on cleaning, formatting, and joining two datasets for further analysis.

Steps Performed:

- Loading Data: Two Excel sheets were loaded into df1 and df2.
- Join Strategy: Both datasets contained a common column "Primary Key", which was used to perform a left join, keeping all records from df1.
- Data Cleaning:
 - Checked for and noted any null values.
 - Dropped duplicate records if present.
 - Stripped whitespace from column names to standardize naming.
 - Converted columns like Cost and Segment Total to float data types.
 - Filled missing values:
 - Coverage → replaced with "Unknown".
 - Cause, Correction in df1 → filled with "Not Mentioned".

Task 3.1 – Trend Analysis

This analysis involved identifying temporal patterns and data relationships through visual exploration.

Steps Performed:

- Date Handling: Ensured Order Date was converted to proper datetime format.
- Monthly Trends: Grouped data by month and counted records to assess trends.

Visualizations:

- Line plot to show monthly record count.
- Heatmap showing correlation between numeric variables, with interpretation of:
 - Pearson Correlation Coefficient scale (from -1 to 1).
 - Use of Coolwarm color map to distinguish strength and direction of relationships.

Task 3.2 – Root Cause Identification

This sub-task explored how component failures contribute to revenue and resource consumption.

Steps Performed:

- Average Revenue by Failure Component:
 - Bar chart showing the top 8 failure components contributing the most revenue loss.
- Actual Hours by Failure Component:
 - Boxplot showing variation in actual hours spent across different failure components.

Summary of Key Actions

- Standardized and merged two datasets via primary key.
- Cleaned and preprocessed text, numeric, and date fields.
- Used visual tools (line plot, heatmap, bar chart, boxplot) to explore:
 - Monthly trends
 - Variable relationships
 - Revenue and time distribution by failure component

Task 2 & 3 - Conclusion

The work in Tasks 2 and 3 successfully transformed raw data into meaningful insights. In Task 2, the data was cleaned, merged using a common key, and prepared for analysis by handling missing values and standardizing formats.

In Task 3, trend analysis uncovered monthly patterns and variable relationships using visual tools like line plots and heatmaps. The root cause analysis identified key failure components responsible for high revenue loss and time usage, using bar charts and boxplots.

Together, these steps provide a clear view of problem areas, helping teams focus on improving efficiency and reducing costs.

References and links (Deliverables)

1. Problem Statement:

| | |
|-------------------|----------------------|
| PO Assignment.pdf | LINK |
|-------------------|----------------------|

2. Datasets:

| | |
|--------------------------|----------------------|
| Data For Task 2 | LINK |
| merged_task2_output.xlsx | LINK |

3. Analysis:

| | |
|--|----------------------|
| task 2 & 3_analysis.ipynb (Jupyter Notebook) | LINK |
| task 2 & 3_analysis.pdf (PDF) | LINK |
| task 2 & 3_analysis.py (Python Script) | LINK |

To explore more of my work, visit my GitHub profile: <https://github.com/priyanshubiswas-tech>