



**Supclair**  
Clarity in Every Chain



# *Capstone Project*

# *Technical Report*

**Date of Submission:** 29/3/2025

**Submitted by:** Supclair

# **Contents**

Introduction .....	3
Project's Objective.....	3
Plan and Approach .....	4
Business Questions .....	5
Data Cleaning.....	5
Objectives of Data Cleaning .....	5
Data Sources .....	7
Cleaning steps .....	7
Product Table.....	7
Transportation Table.....	8
Customer Table .....	8
Supplier Table .....	9
Report Table .....	9
Data Model.....	10
DAX Measures .....	11
Data Visualization & Insights.....	12
Sales Overview Page .....	12
Inventory & Costs Page .....	21
Supplier & Manufacturing I Page .....	21
Supplier & Manufacturing II Page .....	22
Insights Page .....	22

# ***Introduction***

In today's competitive business environment, efficient inventory management, supplier evaluation, and logistics optimization are critical for maintaining a seamless supply chain. This project aims to leverage data analytics and visualization tools to gain actionable insights into stock availability, sales performance, supplier efficiency, and transportation logistics. By integrating advanced analytical techniques, the project seeks to enhance decision-making and operational efficiency within the supply chain framework.

## ***Project's Objective***

The primary objective of this analysis is to develop a data-driven approach for evaluating key supply chain components, including inventory management, sales, supplier performance, and logistics efficiency. The study focuses on:

- Identifying patterns in stock turnover and availability.
- Assessing supplier reliability through defect rates and lead times.
- Analyzing the impact of pricing on sales performance.
- Evaluating transportation and shipping efficiency.
- Developing predictive insights for demand forecasting and supplier segmentation.

## **Plan and Approach**

The project follows a structured workflow consisting of five key steps:

### **1. Data Cleaning & Preprocessing**

- Handle missing values in Availability, Stock Level, and Lead Time.
- Standardize column formats for consistency.
- Remove duplicate entries.

### **2. Exploratory Data Analysis (EDA)**

- Compute summary statistics (mean, median, distribution of key variables).
- Perform correlation analysis (e.g., Lead Time vs. Defect Rate).
- Detect anomalies or trends in stock availability.

### **3. KPI Calculation & Metrics**

- Stock Turnover Rate = Total Products Sold / Average Stock Level
- Stockout Rate = (Orders with Zero Availability) / Total Orders
- Revenue per Product = Total Revenue / Number of Products Sold
- Average Shipping Time = Total Shipping Time / Total Orders
- Defect Rate per Supplier = Total Defective Products / Total Products Received

### **4. Data Visualization**

- Develop a Stock Availability Dashboard.
- Track Revenue & Sales Trends over time.
- Compare Supplier Performance based on Lead Time and Defect Rates.

# ***Business Questions***

The analysis is structured to address critical business questions from an inventory, sales, and logistics perspective:

## **Sales & Revenue Analysis**

- What are the best-selling products?
- How do price changes impact sales volume?
- What is the revenue generated per product?

## **Supplier & Logistics Performance**

- Which suppliers have the shortest lead times?
- Is there a correlation between lead time and defect rates?
- Are certain suppliers responsible for more defective products?
- How efficient are the shipping processes?

# ***Data Cleaning***

## ***Objectives of Data Cleaning***

The primary objective of this data cleaning process is to ensure the accuracy, consistency, and reliability of the **Supplier Quality Dataset**. By addressing common data quality issues, this process aims to enhance the validity of subsequent analysis and reporting. The specific goals of data cleaning include:

- **Handling Missing Data:** Identifying and addressing any missing values in key columns to prevent gaps in analysis.

- **Correcting Inconsistencies:** Standardizing naming conventions, data formats, and measurement units to ensure uniformity across records.
- **Removing Duplicates:** Eliminating redundant records to avoid double counting and biased results.
- **Resolving Errors:** Identifying and correcting incorrect values, such as negative defect counts or unrealistic downtime durations.
- **Ensuring Data Integrity:** Verifying relationships between different columns, such as supplier IDs matching the correct supplier names.

By conducting a thorough data cleaning process, the dataset will be optimized for accurate supplier quality analysis, ensuring meaningful insights for decision-making.

## ***Data Sources***

The dataset used in this analysis originates from supply chain management systems and consists of records related to lead times, stock levels, shipping details, and supplier performance.

The dataset is provided in CSV format (Supply chain.csv) for seamless import into data analysis tools, structured into multiple fields capturing essential supply chain metrics needed for evaluation and optimization.

## ***Cleaning steps***

The steps will be in this template for each table:

**Table name:**

**Notes:**

**Changes made:**

**Data type changes:**

---

The flat file “supply\_chain\_dataset” was transformed to the Power Query.

Our approach was to normalize the given dataset for better overall performance and data integrity. This was done by dividing the flat files into 1-fact table and 4-dimension tables. The fact table is the **Report Table**, and the dimension tables are **Products Table**, **Supplier Table**, **Customer Table**, and **Transportation table**.

To normalize the dataset, we followed these steps:

1. Make a copy of the **Report Table** as **Duplicate**.
2. Choose the **columns needed** for the table and **remove the others**.
3. Remove the **duplicates** as needed to make a **unique dimension table**.
4. Add an **indexed column** if needed.
5. Change the **Data types** if needed.
6. Go back to the **Report Table** and merge the new table with the **intended column** if needed. Choose the **ID column**.
7. Remove the **normalized column** from the **Report Table** leaving just the **ID column**

## Product Table

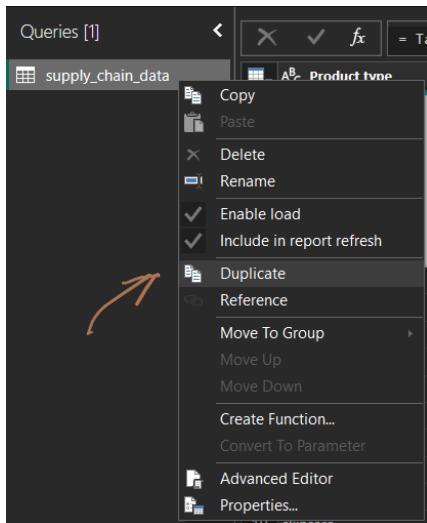


Figure 1: Duplicate Report Table

AB <sub>C</sub> Product type	AB <sub>C</sub> SKU	\$ Price
Valid 100%	SKU0	
Error 0%	SKU1	
Empty 0%	SKU2	
3 distinct, 0 unique	SKU3	
1 haircare	SKU4	
2 skincare	SKU5	
3 haircare	SKU6	
4 skincare	SKU7	
5 skincare	SKU8	
6 haircare	SKU9	
7 skincare	SKU10	
8 cosmetics	SKU11	
9 cosmetics		
10 skincare		
11 skincare		
12 skincare		

Figure 2: Removing other columns

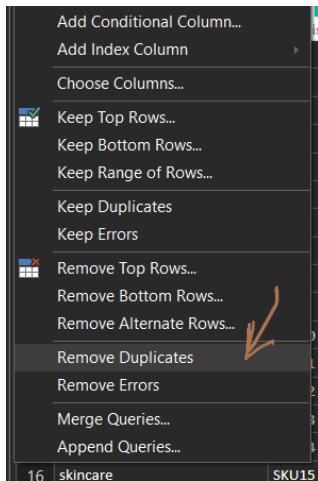


Figure 3: Removing Duplicates

AB <sub>C</sub> Product type	AB <sub>C</sub> SKU	\$ Price
Valid		
Error		
Empty		
3 distinct, 0 unique		
1 haircare		
2 skincare		
3 haircare		

Figure 4: Remove Product Type from Report Table

	A <sup>B</sup> <sub>C</sub> SKU	A <sup>B</sup> <sub>C</sub> Product type
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	100 distinct, 100 unique	3 distinct, 0 unique
1	SKU0	haircare
2	SKU1	skincare

Figure 5: Final Product Table

Instead of creating an indexed column, SKU is used.

There is not any missing data or inconsistency

## Transportation Table

	A <sup>B</sup> <sub>C</sub> TransportationID	A <sup>B</sup> <sub>C</sub> Shipping carriers	A <sup>B</sup> <sub>C</sub> Routes	A <sup>B</sup> <sub>C</sub> Transportation modes
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	33 distinct, 33 unique	3 distinct, 0 unique	3 distinct, 0 unique	4 distinct, 0 unique
1	1	Carrier A	Route B	Air
2	2	Carrier B	Route C	Air
3	3	Carrier C	Route B	Air

There is no missing data or inconsistency.

Data Type Changes:

Transportation ID column is changed from Whole number to Text.

## Customer Table

	A <sup>B</sup> <sub>C</sub> GenderID	A <sup>B</sup> <sub>C</sub> Customer demographics
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	4 distinct, 4 unique	4 distinct, 4 unique
1	1	Non-binary
2	2	Female
3	3	Unknown
4	4	Male

There is no missing data or inconsistency.

Data Type Changes:

Gender ID column is changed from Whole number to Text.

## Supplier Table

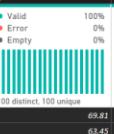
	A <sub>1</sub> <sup>B</sup> <sub>C</sub> Supplier ID	A <sub>1</sub> <sup>B</sup> <sub>C</sub> Supplier name	A <sub>1</sub> <sup>B</sup> <sub>C</sub> Location
	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  25 distinct, 25 unique	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  5 distinct, 0 unique	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  5 distinct, 0 unique
1	1	Supplier 1	Bangalore
2	2	Supplier 1	Mumbai

There is no missing data or inconsistency.

Data Type Changes:

Supplier ID column is changed from Whole number to Text.

## Report Table

	A <sub>1</sub> <sup>B</sup> <sub>C</sub> GenderID	A <sub>1</sub> <sup>B</sup> <sub>C</sub> SKU	A <sub>1</sub> <sup>B</sup> <sub>C</sub> SupplierID	A <sub>1</sub> <sup>B</sup> <sub>C</sub> TransportationID	\$ Price	A <sub>1</sub> <sup>B</sup> <sub>C</sub> Availability	A <sub>1</sub> <sup>B</sup> <sub>C</sub> NumberOfProductsSold	\$ RevenueGenerated	A <sub>1</sub> <sup>B</sup> <sub>C</sub> Stocklevels						
	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  4 distinct, 0 unique	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  100 distinct, 100 unique	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  25 distinct, 1 unique	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  33 distinct, 8 unique	100\$	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  100 distinct, 100 unique	68.81	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  43 distinct, 37 unique	55	802	8,662.00	<ul style="list-style-type: none"><li>Valid 100%</li><li>Error 0%</li><li>Empty 0%</li></ul>  96 distinct, 92 unique	253	8,218.90	8,128.03
1	1	SKU10	13	23	68.81										
2	2	SKU29	5	29	63.45										
3	3	SKU20	4	1	96.34										

The ID columns were added to the report and the normalized columns were removed.

There is no missing data, duplicates or inconsistency.

Data type Changes:

Price column is changed from Decimal number to fixed decimal number (currency).

RevenueGenerated column is changed from Decimal number to fixed decimal number (currency).

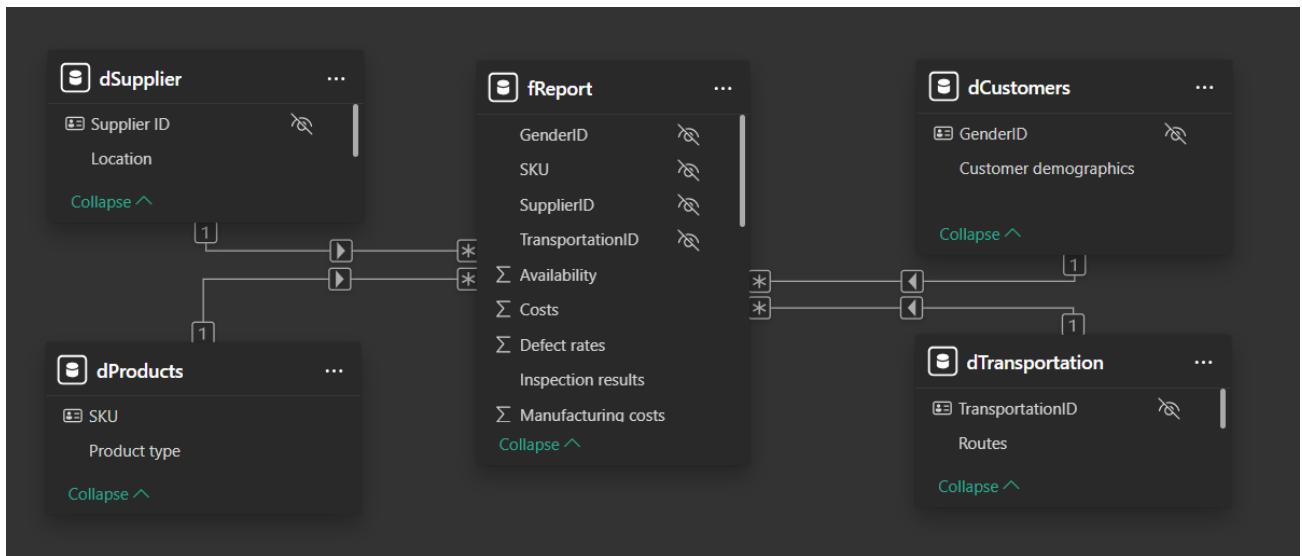
**ShippingCost column** is changed from **Decimal number** to **fixed decimal number (currency)**.

**ManufacturingCost column** is changed from **Decimal number** to **fixed decimal number (currency)**.

**Costs column** is changed from **Decimal number** to **fixed decimal number (currency)**.

## Data Model

**Star schema** for better performance with fact table in the middle (Report Table) and the dimensions tables connected to it (Supplier Table, Product Table, Customer Table, Transportation Table)



# **DAX Measures**

## Average Lead Time Variance

```
Average Lead Time Variance =  
VAR AverageLeadTime =  
    AVERAGE (  
        fReport[Supplier Lead time]  
    )  
VAR LeadTimeStdDev =  
    STDEV.P (  
        fReport[Supplier Lead time]  
    )  
RETURN  
    LeadTimeStdDev / AverageLeadTime
```

## Average Manufacturing Lead Time

```
Average Manufacturing Lead Time =  
    AVERAGE(  
        fReport[Manufacturing lead time]  
    )
```

## Average Order Lead Time

```
Average Order Lead time =  
    AVERAGE(  
        fReport[Order Lead time]  
    )
```

## Average Supplier Lead Time

```
Average Supplier Lead Time =  
    AVERAGE(  
        fReport[Supplier Lead time]  
    )
```

## Average Shipping Cost

```
Average shipping cost =  
    AVERAGE(  
        fReport[Shipping costs]  
    )
```

## Average Selling Price

```
Average Selling Price =  
    DIVIDE (  
        [Total Revenue],  
        [Total Sold Quantity]  
    )
```

## Cost per Defective Product

```
CostperDefectiveProduct =  
    DIVIDE(  
        [Total Costs],  
        [HighDefectCount],  
        BLANK()  
    )
```

## General Costs

```
General Costs =  
    SUM(  
        fReport[Costs]  
    )
```

## Manufacturing Cost

```
Manufacturing Cost =  
    SUM(  
        fReport[Manufacturing Costs]  
    )
```

## Revenue per Shipment

```
RevenueperShipment =  
    DIVIDE(  
        [Total Revenue],  
        [Total Shipments],  
        BLANK()  
    )
```

## Total Costs

```
Total Costs =  
    SUM ( fReport[Costs] ) + SUM ( fReport[Manufacturing costs] )  
    + SUM ( fReport[Shipping costs] )
```

## Total Profit

```
Total Profit =  
[Total Revenue] - [Total Costs]
```

## Total Revenue

```
Total Revenue =  
    SUM(  
        fReport[RevenueGenerated]  
    )
```

## Defect Rate (%)

```
Defect Rate (%) =  
(AVERAGE(fReport[Defect rates])/100)
```

## High Defect Count

```
HighDefectCount =  
    CALCULATE(  
        COUNT(  
            fReport[Inspection results]  
,  
            fReport[Inspection results] = "Fail"  
)
```

## Inspection Pass Rate (%)

```
Inspection Pass Rate % =  
VAR PassedInspections =  
    CALCULATE (  
        COUNT ( fReport[Inspection results] ),  
        fReport[Inspection results] = "Pass"  
)  
VAR TotalCompletedInspections =  
    CALCULATE (  
        COUNT ( fReport[Inspection results] ),  
        fReport[Inspection results] IN { "Pass", "Fail" }  
)  
RETURN  
    DIVIDE ( PassedInspections, TotalCompletedInspections )
```

## Inventory Turnover

```
InventoryTurnover =  
    VAR COGS = [Total Costs]  
    VAR AvgInventory = AVERAGE(fReport[StockLevels])  
    RETURN  
        DIVIDE(  
            COGS,  
            AvgInventory,  
            0  
)
```

## Manufacturing Efficiency

```
ManufacturingEfficiency = DIVIDE(
    SUM(fReport[Production volumes]),
    SUM(fReport[Manufacturing lead time])
) / 100
```

## Manufacturing Waste Ratio (%)

```
Manufacturing Waste Ratio % =
DIVIDE([Cost per Defective Product] * [High Defect Count], [Manufacturing Cost], BLANK()) *
100
```

## Production Volume

```
Manufacturing Waste Ratio % =
DIVIDE (
    [CostperDefectiveProduct] * [HighDefectCount],
    [Manufacturing Cost],
    BLANK ()
) * 100
```

## Stock Availability (%)

```
Stock Availability % =
AVERAGE (
    fReport[Availability]
) / 100
```

## Supplier Performance

```
Supplier Performance =
VAR AvgDefectRate =
    AVERAGEX ( VALUES ( dSupplier[Supplier ID] ), [Defect Rate (%)] )
VAR AvgLeadTime =
    AVERAGEX (
        VALUES ( dSupplier[Supplier ID] ),
```

```

[Average Manufacturing Lead Time]
)
VAR AvgStockAvailability =
    AVERAGEX ( VALUES ( dSupplier[Supplier ID] ), [Stock Availability %] )
VAR AvgManufacturingEfficiency =
    AVERAGEX ( VALUES ( dSupplier[Supplier ID] ), [ManufacturingEfficiency] )
VAR SupplierDefectRate = [Defect Rate (%)]
VAR SupplierLeadTime = [Average Manufacturing Lead Time]
VAR SupplierStockAvailability = [Stock Availability %]
VAR SupplierEfficiency = [ManufacturingEfficiency]
RETURN
SWITCH (
    TRUE (),
    -- Intermediate Supplier
    SupplierDefectRate <= AvgDefectRate * 1.1
        && SupplierLeadTime <= AvgLeadTime * 1.1
        && SupplierStockAvailability >= AvgStockAvailability * 0.95
        && SupplierEfficiency >= AvgManufacturingEfficiency * 0.95, "✓ Good",
    -- Acceptable Supplier
    SupplierDefectRate <= AvgDefectRate * 1.3
        && SupplierLeadTime <= AvgLeadTime * 1.3
        && SupplierStockAvailability >= AvgStockAvailability * 0.85
        && SupplierEfficiency >= AvgManufacturingEfficiency * 0.85, "● Acceptable",
    -- Bad Supplier (Needs improvement)
    SupplierDefectRate > AvgDefectRate * 1.3
        || SupplierLeadTime > AvgLeadTime * 1.3
        || SupplierStockAvailability < AvgStockAvailability * 0.75
        || SupplierEfficiency < AvgManufacturingEfficiency * 0.75, "✗ Bad",
    -- Extremely Bad Supplier (Very poor performance)
    "⚠️ Extremely Bad"
)

```

## Highest Supplier

```

HIghestSupplier =
    VAR SupplierSummary = (
        ADDCOLUMNS(
            SUMMARIZE(
                dSupplier,
                dSupplier[Supplier name]
            ),
            "Total Revenue", [Total Revenue]
        ))
    VAR MaxSupplier = TOPN(

```

```

    1,
    SupplierSummary,
    [Total Revenue],
    DESC
)
VAR MaxSupplierName = SELECTCOLUMNS(
    MaxSupplier,
    "Supplier", dSupplier[Supplier name]
)
RETURN
    MaxSupplierName & " " & FORMAT(
        SELECTCOLUMNS(
            MaxSupplier,
            "Revenue", [Total Revenue]
        ),
        "$#,##K"
)

```

## Highest Gender

```

HighestGender =
VAR GenderSummary = (
    ADDCOLUMNS(
        SUMMARIZE(
            dCustomers,
            dCustomers[Customer demographics]
        ),
        "Total Revenue", [Total Revenue]
    ))
VAR MaxGender = TOPN(
    1,
    GenderSummary,
    [Total Revenue],
    DESC
)
VAR MaxGenderName = SELECTCOLUMNS(
    MaxGender,
    "Gender", dCustomers[Customer demographics]
)
RETURN
    MaxGenderName & " " & FORMAT(
        SELECTCOLUMNS(
            MaxGender,
            "Revenue", [Total Revenue]
        ),
        "$#,##K"
)

```

```
)
```

## Highest Product Type Name

```
HighestProductName =  
    VAR ProductSummary = (  
        ADDCOLUMNS(  
            SUMMARIZE(  
                dProducts,  
                dProducts[Product type]  
>),  
            "Total Revenue", [Total Revenue]  
>))  
    VAR MaxProduct = TOPN(  
        1,  
        ProductSummary,  
        [Total Revenue],  
        DESC  
>)  
    VAR MaxProductName = SELECTCOLUMNS(  
        MaxProduct,  
        "ProductType", dProducts[Product Type]  
>)  
    RETURN  
        MaxProductName & " " & FORMAT(  
            SELECTCOLUMNS(  
                MaxProduct,  
                "Revenue", [Total Revenue]  
>),  
            "$#,##K"  
>)
```

## Total Order Quantity

```
Total Order Quantity =  
    SUM(  
        fReport[Order quantities]  
>)
```

## Total Shipments

```
Total Shipments =  
    SUM (  
        fReport[Shipping times]  
    )
```

## Total Shipping Costs

```
Total Shipping Costs =  
    SUM (  
        fReport[Shipping costs]  
    )
```

## Total Sold Quantity

```
Total Sold Quantity =  
    SUM (  
        fReport[NumberOfProductsSold]  
    )
```

# **Data Visualization & Insights**

This section presents the key visuals in the dashboard and the insights they provide. The dashboard consists of four main pages, each designed to address specific business questions and provide actionable insights.

## **Sales Overview Page**

### **Key Insights from Visuals:**

- **KPIs** (Total Revenue, Total Costs, Total Profit, Total Sold Quantity, Total Order Quantity, Total Shipments, Average Selling Price)
  - Provides a high-level summary of overall sales performance, profitability, and operational efficiency.
- **Donut Chart: Total Revenue by Product Type**
  - Highlights which product categories generate the most revenue, helping prioritize stock management and promotions.
- **Donut Chart: Total Revenue by Customer Gender**
  - Identifies revenue distribution across different customer segments, guiding targeted marketing strategies.
- **Donut Chart: Total Revenue by Supplier**
  - Reveals supplier contribution to revenue, helping evaluate supplier relationships.
- **Scatter Plot: Price vs. Number of Products Sold**
  - Displays pricing trends and demand elasticity, helping set optimal pricing strategies.
- **Bar Chart: Top 20 Profitable Products**
  - Identifies the most profitable SKUs, guiding inventory prioritization and promotions.
- **Filters: Supplier Name, Gender, Location, Product Type**
  - Enables dynamic analysis by refining insights based on key attributes.

## **Inventory & Costs Page**

### **Key Insights from Visuals:**

- **KPIs** (Total Revenue, Total Costs, General Costs, Total Shipping Costs, Manufacturing Cost, Total Shipments, Stock Availability %)
  - Gives a financial and operational overview of inventory costs and stock availability.
- **Table:** Products with Stock Availability < 20%
  - Identifies products at risk of stockouts, ensuring proactive inventory management.
- **Decomposition Tree:** Total Costs Breakdown (by Product Type, Manufacturing Cost, Shipping Costs, General Costs)
  - Analyzes cost distribution to pinpoint areas for cost optimization.
- **Column Chart:** General Costs by Product Type
  - Highlights cost-intensive product categories for better financial planning.
- **Column Chart:** Total Shipping Costs by Shipping Carrier & Transportation Mode
  - Evaluates logistics efficiency and cost differences between transportation methods.
- **Column Chart:** Manufacturing Cost by Supplier & Product Type
  - Identifies high-cost suppliers and product types requiring cost control.

## **Supplier & Manufacturing I Page**

### **Key Insights from Visuals:**

- **KPIs** (Average Order Lead Time, Average Supplier Lead Time, Average Manufacturing Lead Time, High Defect Count, Inspection Pass Rate, Manufacturing Efficiency, Manufacturing Waste Ratio)
  - Provides a supplier and production performance overview.
- **Column Chart:** Products Count by Total Lead Time (Bins)
  - Identifies bottlenecks in supply chain efficiency by analyzing lead time distribution.
- **Column Chart:** Average Supplier Lead Time by Supplier (Grouped by Location)
  - Compares supplier performance based on geographical distribution.
- **Table:** Supplier Performance (Good, Bad, Acceptable)
  - Classifies suppliers based on multiple KPIs, supporting supplier evaluation.

- **Tree Map: Inspection Results by Products Count**
  - Analyzes inspection outcomes to identify quality control issues.

## ***Supplier & Manufacturing II Page***

### ***Key Insights from Visuals:***

- **KPIs (Same as Supplier & Manufacturing I Page)**
  - Extends supplier and manufacturing performance analysis.
- **Line Chart: Average Manufacturing Lead Time vs. Production Volume**
  - Shows how production volume affects manufacturing efficiency and delays.
- **Line Chart: Average Supplier Lead Time vs. Production Volume**
  - Identifies trends in supplier lead times with production scaling.
- **Line Chart: Order Quantities vs. Order Quantities (10 Bins)**
  - Visualizes demand patterns and variations.
- **Column Chart: Total Shipments by Shipping Carrier**
  - Evaluates shipping carrier performance in fulfilling orders.
- **Column Chart: Total Order Quantity by Routes**
  - Highlights high-traffic shipping routes for logistics optimization.
- **Tree Map: Total Shipping Costs by Shipping Carrier**
  - Helps compare and optimize transportation expenses.

## ***Insights Page***

The page is for all the analysis and insights, to be easily accessible by dashboard users.

These insights enable strategic decision-making by providing a clear understanding of sales performance, cost distribution, supplier efficiency, and logistics operations.