Ministry of Communications and Information Technology

Digital Egypt Pioneers Initiative - DEPI

Project Idea 2: Supply Chain

Date of Submission 22 October 2024

Name		Signature
Prepared by		
Mohamed Tarek Galal		
Amira El Hussein Mohamed		
Noha Afify Abdel Samie		
Aml Mohamed Abdel Hamid		
Ahmed Mohamed Ramzy		
Reviewed by		
Eng.	**Ahmed Samir**	
Evaluated by		
Eng.	**Ahmed Samir**	

Contents

ABSTRACT

Supply chain analytics is a valuable part of data-driven decision-making in various industries such as manufacturing, retail, healthcare, and logistics. It is the process of collecting, analyzing and interpreting data related to the movement of products and services from suppliers to customers.

Here is a dataset we collected from a Fashion and Beauty startup. The dataset is based on the supply chain of Makeup products. Below are all the features in the dataset:

- Product Type
- SKU
- Price
- Availability
- Number of products sold
- Revenue generated
- Customer demographics
- Stock levels
- Lead times
- Order quantities
- Shipping times
- Shipping carriers
- Shipping costs
- Supplier name
- Location

- Lead time
- Production volumes
- Manufacturing lead time
- Manufacturing costs
- Inspection results
- Defect rates
- Transportation modes
- Routes
- Costs

LIST OF FIGURES

NOMENCLATURE

EOQ	Economic Order Quantity
EDA	Exploratory Data Analysis

CHAPTER 1: DATA CLEANING & PROCESSING

1.1 Introduction

In this supply chain analytics analysis, we encountered the challenge of working with a dataset consisting of 100 rows and 24 columns. During the Exploratory Data Analysis phase, we conducted quality control checks, assessed supply chain risks, performed Inventory Optimization Analysis using the Economic Order Quantity (EOQ) method, conducted customer segmentation analysis, and explored lead times optimization.

1.1.1 Objectives:

- Data Exploration: Gain insights into the supply chain data by exploring production volumes, stock levels, order quantities, revenue, costs, lead times, shipping costs, transportation routes, risks, and sustainability factors.
- Visualization: Create informative visualizations to better understand the relationships and distributions within the data.
- Dashboard Development: Build an interactive Streamlit dashboard to visualize the key metrics and facilitate real-time analysis.

1.1.2 Key Sections:

1. Data Preprocessing
    - Handling missing values.
    - Data cleaning and preparation.
2. Exploratory Data Analysis (EDA)
    - Descriptive statistics.
    - Visualization of key metrics.

1.1.3 Visualizations

- Production volumes, stock levels, and lead times.
- Revenue distribution by location.
- Manufacturing costs by supplier.
- Comparison of price and manufacturing costs by product type.
- Relationship between production volume, stock levels, and order quantities.
- Distribution of shipping costs by shipping carriers.
- Average lead time by product type.
- Transportation routes and their frequency.
- Supply chain risk distribution by risk factors.
- Sustainability factors in the supply chain

1.2 Load Data

- We wanted to hide the warnings to focus on the errors only.

Python Code [1]:

```
import warnings
warnings.filterwarnings('ignore')
warnings.filterwarnings('ignore', category=FutureWarning)
```

- Importing pandas library to use it in data cleaning and processing

Python Code [2]:

```
import pandas as pd
supply_data = pd.read_csv (“supply_chain_data.csv”)
```

- Finding the shape of the data to know number of columns and rows in the dataset

Python Code [3]:

```
supply_data.shape
```

Output [3]:

(100, 24)

- We found the names of the columns and understoond the content of each column.

Python Code [4]:

```
supply_data.columns
```

Output [4]:

```
Index(['Product type', 'SKU', 'Price', 'Availability',
       'Number of products sold', 'Revenue generated', 'Customer demogr
aphics',
       'Stock levels', 'Lead times', 'Order quantities', 'Shipping time
s',
       'Shipping carriers', 'Shipping costs', 'Supplier name', 'Locatio
n',
       'Lead time', 'Production volumes', 'Manufacturing lead time',
       'Manufacturing costs', 'Inspection results', 'Defect rates',
       'Transportation modes', 'Routes', 'Costs'],
      dtype='object')
```

- **Feature explanation**
  - Product Type: The type of product Category (skincare – haircare – cosmetics).
  - SKU (Stock Keeping Unit): Unique code for each product (100 unique No. from SKU0 to SKU99).
  - Price: The price of item (from 1.7 to 99.2).
  - Availability: No. of items Available (from 1 to 100).
  - Number of Products Sold (from 8 to 996)
  - Revenue Generated: Total revenue generated from product sales (from 1.06K to 9.87K).
  - Customer demographics: Information about customer gender (Male – Female – Unknown - Non-binary).
  - Stock Levels: The number of products still available in stock at any given time (from 0 to 100).
  - Lead Times: The time required to order or receive products from suppliers (from 1 to 30).
  - Order Quantities: The number of products ordered in one order (from 1 to 96).
  - Shipping Times: The time required to ship products from the warehouse or distribution center to customers (from 1 to 10).
  - Shipping Carriers: (Carrier A – B – C).
  - Shipping Costs: (from 1.01 to 9.93).

- Supplier Name: (Supplier 1 – 2 – 3 – 4 – 5).
- Location: location of a warehouse or distribution center (Kolkata – Mumbai – Chennai – Bangalore – Delhi).
- Lead Time: The time required to obtain products or materials from a particular supplier (from 1 to 30).
- Production Volumes: The number of products produced (from 104 to 985).
- Manufacturing Lead Time: The time required to produce a product (from 1 to 30).
- Manufacturing Costs: (from 1.09 to 99.5).
- Inspection Results: Results of product or material quality inspection (Pending – Fail – Pass).
- Defect Rates: The level of defects or defects in the products produced (from 0.02 to 4.94).
- Transportation Modes: The transportation methods (Road – Rail – Air – Sea).
- Routes: paths used to send products from one point to another (Route A – B - C).
- Costs: Costs including production costs and other costs (from 104 to 997).

- We found info of dataset like data type and if there are any null values to have an overview about the dataset

Python Code [5]:

```
supply_data.info()
```

Output [5]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 24 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Product type           100 non-null    object
 1   SKU                    100 non-null    object
 2   Price                  100 non-null    float64
 3   Availability           100 non-null    int64
 4   Number of products sold  100 non-null    int64
 5   Revenue generated      100 non-null    float64
 6   Customer demographics  100 non-null    object
 7   Stock levels           100 non-null    int64
 8   Lead times             100 non-null    int64
 9   Order quantities       100 non-null    int64
 10  Shipping times         100 non-null    int64
```

```
 11  Shipping carriers        100 non-null    object
 12  Shipping costs           100 non-null    float64
 13  Supplier name            100 non-null    object
 14  Location                 100 non-null    object
 15  Lead time                100 non-null    int64
 16  Production volumes       100 non-null    int64
 17  Manufacturing lead time  100 non-null    int64
 18  Manufacturing costs      100 non-null    float64
 19  Inspection results       100 non-null    object
 20  Defect rates             100 non-null    float64
 21  Transportation modes     100 non-null    object
 22  Routes                   100 non-null    object
 23  Costs                    100 non-null    float64
dtypes: float64(6), int64(9), object(9)
memory usage: 18.9+ KB
```

1.3 Data Processing

- We checked for the missing values and Display columns with missing values and the count of missing values.

Python Code [6]:

```
missing_values = supply_data.isnull().sum()

missing_values = missing_values[missing_values > 0]

if not missing_values.empty:
    print("Columns with missing values:")
    for column, count in missing_values.items():
        print(f"{column}: {count} missing values")
else:
    print("There are no columns with missing value")
```

Output [6]:

```
There are no columns with missing value
```

- We checked for duplication.

Python Code [7]:

```
if supply_data.duplicated().any():
    print(f"There are as many as {supply_data.duplicated().sum()} duplicate data.")
else:
    print("There are no duplicate data.")
```

Output [7]:

There are no duplicate data.

1.4 Data Cleaning

- We found that Revenue generated not equal Standard formula (Price * Number of products sold)
- We Added a new column to calculate (Calculated Revenue) according to standard formula

Python Code [8]:

```
supply_data['Calculated Revenue'] = supply_data['Price'] * supply_data['Number of p
roducts sold']

print(supply_data.head())
```

Output [8]:

```
Product type   SKU      Price  Availability  Number of products sold  \
0     haircare  SKU0  69.808006          55                     802
1     skincare  SKU1  14.843523          95                     736
2     haircare  SKU2  11.319683          34                       8
3     skincare  SKU3  61.163343          68                      83
4     skincare  SKU4   4.805496          26                     871

   Revenue generated Customer demographics  Stock levels  Lead times  \
0        8661.996792          Non-binary            58          7
1        7460.900065              Female            53         30
2        9577.749626             Unknown             1         10
3        7766.836426          Non-binary            23         13
4        2686.505152          Non-binary             5          3

   Order quantities  ...  Lead time  Production volumes  \
0                96  ...         29                 215
1                37  ...         23                 517
2                88  ...         12                 971
3                59  ...         24                 937
4                56  ...          5                 414

   Manufacturing lead time  Manufacturing costs  Inspection results  \
0                       29            46.279879             Pending
1                       30            33.616769             Pending
2                       27            30.688019             Pending
3                       18            35.624741                Fail
4                        3            92.065161                Fail

   Defect rates  Transportation modes   Routes       Costs  Calculated Revenue
0      0.226410                 Road  Route B  187.752075        55986.020443
```

```
1      4.854068              Road  Route B  503.065579      10924.833134
2      4.580593               Air  Route C  141.920282         90.557466
3      4.746649              Rail  Route A  254.776159       5076.557471
4      3.145580               Air  Route A  923.440632       4185.587047

[5 rows x 25 columns]
```

- We wanted to compare between Revenue generated and Calculated Revenue to show the difference between these two columns
- So we added a new column (Compare revenue) to calculate the difference in revenue

Python Code [9]:

```
supply_data['Compare revenue'] = supply_data['Revenue generated'] - supply_data['Calculated Revenue']

print(supply_data.head())
```

Output [9]:

```
Product type   SKU      Price  Availability  Number of products sold  \
0     haircare  SKU0  69.808006           55                     802
1     skincare  SKU1  14.843523           95                     736
2     haircare  SKU2  11.319683           34                       8
3     skincare  SKU3  61.163343           68                      83
4     skincare  SKU4   4.805496           26                     871

   Revenue generated Customer demographics  Stock levels  Lead times  \
0        8661.996792             Non-binary            58          7
1        7460.900065               Female            53         30
2        9577.749626              Unknown             1         10
3        7766.836426             Non-binary            23         13
4        2686.505152             Non-binary             5          3

   Order quantities  ...  Production volumes  Manufacturing lead time  \
0                96  ...                 215                       29
1                37  ...                 517                       30
2                88  ...                 971                       27
3                59  ...                 937                       18
4                56  ...                 414                        3

   Manufacturing costs  Inspection results  Defect rates  Transportation modes  \
0            46.279879             Pending      0.226410                  Road
1            33.616769             Pending      4.854068                  Road
2            30.688019             Pending      4.580593                   Air
3            35.624741                Fail      4.746649                  Rail
4            92.065161                Fail      3.145580                   Air
```

	Routes	Costs	Calculated Revenue	Compare revenue
0	Route B	187.752075	55986.020443	-47324.023651
1	Route B	503.065579	10924.833134	-3463.933069
2	Route C	141.920282	90.557466	9487.192160
3	Route A	254.776159	5076.557471	2690.278955
4	Route A	923.440632	4185.587047	-1499.081895

[5 rows x 26 columns]

- We wanted to show this difference as an item level
- So we added a new column (Compare per item) and (Diff Percent) to calculate the difference in revenue per item and show it in perce nt form

Python Code [10]:

```
supply_data['Compare per item'] = supply_data['Compare revenue'] / supply_data['N
umber of products sold']

print(supply_data.head())
```

Output [10]:

	Product type	SKU	Price	Availability	Number of products sold	\
0	haircare	SKU0	69.808006	55	802	
1	skincare	SKU1	14.843523	95	736	
2	haircare	SKU2	11.319683	34	8	
3	skincare	SKU3	61.163343	68	83	
4	skincare	SKU4	4.805496	26	871	

	Revenue generated	Customer demographics	Stock levels	Lead times	\
0	8661.996792	Non-binary	58	7	
1	7460.900065	Female	53	30	
2	9577.749626	Unknown	1	10	
3	7766.836426	Non-binary	23	13	
4	2686.505152	Non-binary	5	3	

	Order quantities	...	Manufacturing lead time	Manufacturing costs	\
0	96	...	29	46.279879	
1	37	...	30	33.616769	
2	88	...	27	30.688019	
3	59	...	18	35.624741	
4	56	...	3	92.065161	

	Inspection results	Defect rates	Transportation modes	Routes	Costs	\

```
0          Pending    0.226410            Road  Route B  187.752075
1          Pending    4.854068            Road  Route B  503.065579
2          Pending    4.580593             Air  Route C  141.920282
3             Fail    4.746649            Rail  Route A  254.776159
4             Fail    3.145580             Air  Route A  923.440632

   Calculated Revenue  Compare revenue Compare per item
0       55986.020443   -47324.023651       -59.007511
1       10924.833134    -3463.933069        -4.706431
2          90.557466     9487.192160      1185.899020
3        5076.557471     2690.278955        32.412999
4        4185.587047    -1499.081895        -1.721104

[5 rows x 27 columns]
```

Python Code [11]:

```
supply_data['Diff Percent'] = supply_data['Compare per item'] / supply_data['Price']

print(supply_data.head())
```

Output [11]:

```
Product type   SKU      Price  Availability  Number of products sold  \
0    haircare  SKU0  69.808006           55                      802
1    skincare  SKU1  14.843523           95                      736
2    haircare  SKU2  11.319683           34                        8
3    skincare  SKU3  61.163343           68                       83
4    skincare  SKU4   4.805496           26                      871

   Revenue generated Customer demographics  Stock levels  Lead times  \
0        8661.996792            Non-binary            58           7
1        7460.900065              Female              53          30
2        9577.749626             Unknown               1          10
3        7766.836426            Non-binary            23          13
4        2686.505152            Non-binary             5           3

   Order quantities  ...  Manufacturing costs Inspection results  \
0                96  ...            46.279879             Pending
1                37  ...            33.616769             Pending
2                88  ...            30.688019             Pending
3                59  ...            35.624741                Fail
4                56  ...            92.065161                Fail

   Defect rates Transportation modes   Routes       Costs  Calculated Revenue  \
0      0.226410              Road  Route B  187.752075        55986.020443
1      4.854068              Road  Route B  503.065579        10924.833134
2      4.580593               Air  Route C  141.920282           90.557466
3      4.746649              Rail  Route A  254.776159         5076.557471
4      3.145580               Air  Route A  923.440632         4185.587047
```

	Compare revenue	Compare per item	Diff Percent
0	-47324.023651	-59.007511	-0.845283
1	-3463.933069	-4.706431	-0.317070
2	9487.192160	1185.899020	104.764329
3	2690.278955	32.412999	0.529942
4	-1499.081895	-1.721104	-0.358153

[5 rows x 28 columns]

- We wanted to analyze Diff Percent to find any illogical values

Python Code [12]:

```
supply_data['Diff Percent'].describe()
```

Output [12]:

```
count    100.000000
mean       1.563645
std       11.130222
min       -0.965986
25%       -0.859455
50%       -0.646807
75%        0.031927
max      104.764329
Name: Diff Percent, dtype: float64
```

- We found a very high value in Diff Percent ,so we assume that our upper limit is 200%
- We cleared the outliers

Python Code [13]:

```
supply_data = supply_data[supply_data['Diff Percent'] <= 2]
supply_data.shape
```

Output [13]:

(88, 28)

- By logic Order quantities can't be greater than Number of products sold
- So we added column to calculate the difference between them

Python Code [14]:

```
supply_data['S/Q Diff'] = supply_data['Number of products sold'] - supply_data['Order
 quantities']

print(supply_data.head())
```

Output [14]:

```
Product type   SKU      Price  Availability  Number of products sold  \
0     haircare  SKU0  69.808006           55                     802
1     skincare  SKU1  14.843523           95                     736
3     skincare  SKU3  61.163343           68                      83
4     skincare  SKU4   4.805496           26                     871
7    cosmetics  SKU7  42.958384           59                     426

   Revenue generated Customer demographics  Stock levels  Lead times  \
0        8661.996792            Non-binary            58          7
1        7460.900065              Female              53          30
3        7766.836426            Non-binary            23          13
4        2686.505152            Non-binary             5          3
7        8496.103813              Female              93          17

   Order quantities  ...  Inspection results Defect rates  \
0                96  ...             Pending     0.226410
1                37  ...             Pending     4.854068
3                59  ...               Fail      4.746649
4                56  ...               Fail      3.145580
7                11  ...               Fail      0.398177

   Transportation modes   Routes       Costs  Calculated Revenue  \
0                  Road  Route B  187.752075         55986.020443
1                  Road  Route B  503.065579         10924.833134
3                  Rail  Route A  254.776159          5076.557471
4                   Air  Route A  923.440632         4185.587047
7                  Road  Route C  802.056312         18300.271746

   Compare revenue  Compare per item  Diff Percent S/Q Diff
0    -47324.023651        -59.007511     -0.845283      706
1     -3463.933069         -4.706431     -0.317070      699
3      2690.278955         32.412999      0.529942       24
4     -1499.081895         -1.721104     -0.358153      815
7     -9804.167933        -23.014479     -0.535739      415

[5 rows x 29 columns]
```

- We cleared any values are negative because it is illogical values

Python Code [15]:

```
supply_data = supply_data[supply_data['S/Q Diff'] >= 0]
supply_data.shape
```

Output [15]:

(88, 29)

- No additional rows cleared but in original data we found there were 3 values were negative but here these 3 values already were in 12 rows (deleted rows)

1.5 Data Preparation & Cleaning using SQL

1. Check for Missing Data

Identify missing or NULL values across all columns in the supply chain table

Code [1]:

```
SELECT
    COUNT(CASE WHEN "Product type" IS NULL THEN 1 END) AS Missing_Product_Type,
    COUNT(CASE WHEN SKU IS NULL THEN 1 END) AS Missing_SKU,
    COUNT(CASE WHEN Price IS NULL THEN 1 END) AS Missing_Price,
    COUNT(CASE WHEN Availability IS NULL THEN 1 END) AS Missing_Availability,
    COUNT(CASE WHEN "Number of products sold" IS NULL THEN 1 END) AS
Missing_Number_of_Products_Sold,
    COUNT(CASE WHEN "Revenue generated" IS NULL THEN 1 END) AS
Missing_Revenue_Generated,
    COUNT(CASE WHEN "Customer demographics" IS NULL THEN 1 END) AS
Missing_Customer_Demographics,
    COUNT(CASE WHEN "Stock levels" IS NULL THEN 1 END) AS Missing_Stock_Levels,
    COUNT(CASE WHEN "Lead times" IS NULL THEN 1 END) AS Missing_Lead_Times,
    COUNT(CASE WHEN "Order quantities" IS NULL THEN 1 END) AS Missing_Order_Quantities,
    COUNT(CASE WHEN "Shipping times" IS NULL THEN 1 END) AS Missing_Shipping_Times,
    COUNT(CASE WHEN "Shipping carriers" IS NULL THEN 1 END) AS
Missing_Shipping_Carriers,
    COUNT(CASE WHEN "Shipping costs" IS NULL THEN 1 END) AS Missing_Shipping_Costs,
    COUNT(CASE WHEN "Supplier name" IS NULL THEN 1 END) AS Missing_Supplier_Name,
    COUNT(CASE WHEN Location IS NULL THEN 1 END) AS Missing_Location,
    COUNT(CASE WHEN "Lead time" IS NULL THEN 1 END) AS Missing_Lead_Time,
    COUNT(CASE WHEN "Production volumes" IS NULL THEN 1 END) AS
Missing_Production_Volumes,
    COUNT(CASE WHEN "Manufacturing lead time" IS NULL THEN 1 END) AS
Missing_Manufacturing_Lead_Time,
    COUNT(CASE WHEN "Manufacturing costs" IS NULL THEN 1 END) AS
Missing_Manufacturing_Costs,
    COUNT(CASE WHEN "Inspection results" IS NULL THEN 1 END) AS
Missing_Inspection_Results,
```

```
    COUNT(CASE WHEN "Defect rates" IS NULL THEN 1 END) AS Missing_Defect_Rates,
    COUNT(CASE WHEN "Transportation modes" IS NULL THEN 1 END) AS
Missing_Transportation_Modes,
    COUNT(CASE WHEN "Routes" IS NULL THEN 1 END) AS Missing_Routes,
    COUNT(CASE WHEN "Costs" IS NULL THEN 1 END) AS Missing_Costs
FROM supply_chain_data;
```

Explanation:

- COUNT(CASE WHEN ... IS NULL THEN 1 END): This syntax counts how many times each specified column has NULL values.
- Aliases: Each count is given an alias (e.g., Missing_Product_Type) for easier reading of the results.

2.Remove Duplicates

Code [2]:

```
DELETE FROM supply_chain_data
WHERE SKU IN (
    SELECT SKU
    FROM supply_chain_data
    GROUP BY SKU
    HAVING COUNT(*) > 1
) AND SKU NOT IN (
    SELECT MIN(SKU)
    FROM supply_chain_data
    GROUP BY SKU
);

(0 rows affected)
```

Explanation:

- *Subquery 1 (SELECT SKU FROM supply_chain_data GROUP BY SKU HAVING COUNT(*) > 1): This identifies SKU values that occur more than once in the dataset.*
- *Subquery 2 (SELECT MIN(SKU) FROM supply_chain GROUP BY SKU): This keeps the row with the minimum SKUfor each SKU.*
- *DELETE Statement:* The outer DELETE statement removes rows that have duplicate SKU values but are not the one with the minimum SKU.

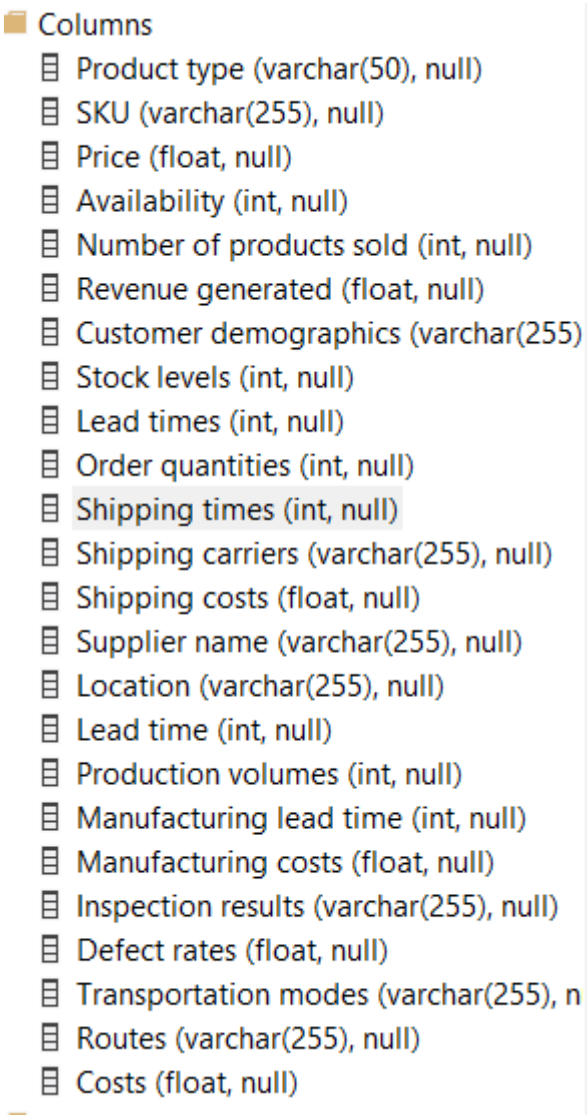
3. Fix Data Types

Code [3]:

```
ALTER TABLE supply_chain_data
ALTER COLUMN "Product type" VARCHAR(50);

ALTER TABLE supply_chain_data
```

```
ALTER COLUMN "SKU" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Price" FLOAT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Availability" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Number of products sold" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Revenue generated" FLOAT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Customer demographics" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Stock levels" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Lead times" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Order quantities" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Shipping times" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Shipping carriers" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Shipping costs" FLOAT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Supplier name" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Location" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Lead time" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Production volumes" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Manufacturing lead time" INT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Manufacturing costs" FLOAT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Inspection results" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Defect rates" FLOAT;

ALTER TABLE supply_chain_data
ALTER COLUMN "Transportation modes" VARCHAR(255);
```

```
ALTER TABLE supply_chain_data
ALTER COLUMN "Routes" VARCHAR(255);

ALTER TABLE supply_chain_data
ALTER COLUMN "Costs" FLOAT;
```



- *The purpose of this SQL script is to modify the data types of specific columns in the `supply_chain_data` table. Each `ALTER COLUMN` command is used to change the data type of a column to better reflect the data it stores, ensuring appropriate storage and functionality for future queries.*

Explanation

- *ALTER COLUMN "Product type" VARCHAR(50);`*

Changes the data type of the `"Product type"` column to `VARCHAR(50)`, allowing it to store string values up to 50 characters. This is suitable for product types like "Electronics" or "Clothing".

- *ALTER COLUMN "SKU" VARCHAR(255);`*
  Modifies the `"SKU"` (Stock Keeping Unit) column to a `VARCHAR(255)`, which can store unique identifiers for products that are alphanumeric and up to 255 characters long.

- *ALTER COLUMN "Price" FLOAT;`*
  Changes the `"Price"` column to `FLOAT`, which can hold decimal values, suitable for representing prices, including cents (e.g., 19.99).

- *ALTER COLUMN "Availability" INT;`*
  Sets the `"Availability"` column to `INT`, allowing it to store whole numbers representing the quantity of a product available in stock.

- *` ALTER COLUMN "Number of products sold" INT;`*
  Alters the `"Number of products sold"` column to `INT`, which keeps track of the total units sold as whole numbers.

- *. ALTER COLUMN "Revenue generated" FLOAT;`*
  Changes the `"Revenue generated"` column to `FLOAT`, allowing it to store decimal values representing the total revenue generated from sales.

- *` ALTER COLUMN "Customer demographics" VARCHAR(255);`*
  Modifies the `"Customer demographics"` column to `VARCHAR(255)`, enabling storage of detailed demographic information in string format.

- *ALTER COLUMN "Stock levels" INT;`*
  Alters the `"Stock levels"` column to `INT`, which represents the current stock levels of products as whole numbers.

- *ALTER COLUMN "Lead times" INT;`*

Changes the `"Lead times"` column to `INT`, indicating the time required (in days, for example) to fulfill an order.

- *ALTER COLUMN "Order quantities" INT;`*
  Sets the `"Order quantities"` column to `INT`, representing how many units are ordered in a transaction as whole numbers.

- *ALTER COLUMN "Shipping times" INT;`*
  Modifies the `"Shipping times"` column to `INT`, indicating the time taken for shipping, typically measured in days.

- *ALTER COLUMN "Shipping carriers" VARCHAR(255);`*
  Changes the `"Shipping carriers"` column to `VARCHAR(255)`, allowing the storage of names of shipping companies (e.g., FedEx, UPS) as strings.

- *` ALTER COLUMN "Shipping costs" FLOAT;`*
  Alters the `"Shipping costs"` column to `FLOAT`, which stores the cost associated with shipping a product, allowing decimal values.

- *` ALTER COLUMN "Supplier name" VARCHAR(255);`*
  Changes the `"Supplier name"` column to `VARCHAR(255)`, allowing it to store the names of suppliers in string format.

- *` ALTER COLUMN "Location" VARCHAR(255);`*
  Modifies the `"Location"` column to `VARCHAR(255)`, enabling the storage of locations (e.g., city, state) as strings.

- *ALTER COLUMN "Lead time" INT;`*
  Sets the `"Lead time"` column to `INT`, representing the estimated time required to produce and deliver a product, typically in days.

- *ALTER COLUMN "Production volumes" INT;`*
  Alters the `"Production volumes"` column to `INT`, indicating the quantity of products manufactured or produced.

- *ALTER COLUMN "Manufacturing lead time" INT;`*

Changes the `"Manufacturing lead time"` column to `INT`, representing the time taken to manufacture a product.

- *ALTER COLUMN "Manufacturing costs" FLOAT;`*
  Alters the `"Manufacturing costs"` column to `FLOAT`, allowing it to store the costs associated with manufacturing products as decimal values.

- *ALTER COLUMN "Inspection results" VARCHAR(255);`*
  Modifies the `"Inspection results"` column to `VARCHAR(255)`, enabling the storage of inspection results or quality checks as strings.

- *ALTER COLUMN "Defect rates" FLOAT;`*
  Changes the `"Defect rates"` column to `FLOAT`, allowing it to store the percentage of defective items as a decimal.

- *ALTER COLUMN "Transportation modes" VARCHAR(255);`*
- *- Alters the `"Transportation modes"` column to `VARCHAR(255)`, which can hold different modes of transportation used for shipping (e.g., air, sea, land).*

- *23. ` ALTER COLUMN "Routes" VARCHAR(255);`*
  Modifies the `"Routes"` column to `VARCHAR(255)`, allowing the storage of routing information for shipping products.

- *ALTER COLUMN "Costs" FLOAT;`*
  Finally, changes the `"Costs"` column to `FLOAT`, which will represent various costs associated with products, including production, shipping, and handling.

- *This script modifies the data types of each column in the `supply_chain_data` table to better align with the expected data stored in each column. This ensures data integrity, optimal storage, and efficient querying.*

4. Standardizing Data Formatting

To standardize data formatting in your supply_chain_data table, you typically need to ensure that text values are consistent (e.g., capitalization, whitespace), numerical values are properly formatted, and that dates (if any) are in a consistent format. Here's how you can write SQL code to standardize the data formatting for each relevant column.

Code [4]:

```
UPDATE supply_chain_data
SET
    "Product type" = TRIM(UPPER("Product type")),
    "SKU" = TRIM(UPPER("SKU")),
    "Customer demographics" = TRIM(LOWER("Customer demographics")),
    "Supplier name" = TRIM(UPPER("Supplier name")),
    "Location" = TRIM(UPPER("Location")),
    "Inspection results" = TRIM(UPPER("Inspection results")),
    "Transportation modes" = TRIM(UPPER("Transportation modes")),
    "Routes" = TRIM(UPPER("Routes"));

(100 rows affected)

UPDATE supply_chain_data
SET
    "Price" = ROUND("Price", 3),
    "Revenue generated" = ROUND("Revenue generated", 3),
    "Shipping costs" = ROUND("Shipping costs", 3),
    "Manufacturing costs" = ROUND("Manufacturing costs", 3),
    "Defect rates" = ROUND("Defect rates", 3);

(100 rows affected)

UPDATE supply_chain_data
SET
    "Availability" = CAST("Availability" AS INT),
    "Number of products sold" = CAST("Number of products sold" AS INT),
    "Stock levels" = CAST("Stock levels" AS INT),
    "Lead times" = CAST("Lead times" AS INT),
    "Order quantities" = CAST("Order quantities" AS INT),
    "Shipping times" = CAST("Shipping times" AS INT),
    "Lead time" = CAST("Lead time" AS INT),
    "Production volumes" = CAST("Production volumes" AS INT),
    "Manufacturing lead time" = CAST("Manufacturing lead time" AS INT);

(100 rows affected)
```

Explanation

- *TRIM: Removes leading and trailing whitespace.*
- *UPPER/LOWER: Converts the text to uppercase or lowercase to ensure consistency.*

Numeric Standardization:

- *ROUND: Rounds numeric columns to a specified number of decimal places to maintain consistency in financial figures.*

Integer Standardization:

- *CAST: Converts values to integers, ensuring that the data type is correctly maintained.*

5.Data Cleaning

- We found that Revenue generated not equal Standard formula (Price * Number of products sold)
- We Added a new column to calculate (Calculated Revenue) according to standard formula

Code [5]:

```
SELECT *
FROM supply_chain_data;

SELECT *,
       ("Price" * "Number_of_products_sold") AS "Calculated Revenue"
FROM supply_chain_data;
```

- **Update the table to add this new column permanently**

```
ALTER TABLE supply_chain_data
ADD [Calculated Revenue] DECIMAL(15, 2);
UPDATE supply_chain_data
SET [Calculated Revenue] = [Price] * [Number_of_products_sold];
```

Code [6]:

- We wanted to compare between Revenue generated and Calculated Revenue to show the difference between these two columns
- So we added a new column (Compare revenue) to calculate the difference in revenue

```
SELECT *,
       ("Revenue generated" - ("Price" * "Number_of_products_sold")) AS "Compare
revenue"
FROM supply_chain_data;


ALTER TABLE supply_chain_data
ADD "Compare revenue" DECIMAL(15, 2);

UPDATE supply_chain_data
```

```
SET "Compare revenue" = "Revenue generated" - ("Price" * "Number_of_products_sold");
```

Code [7]:

- We wanted to show this difference as an item level
- So we added a new column (Compare per item) and (Diff Percent) to calculate the difference in revenue per item and show it in perce nt form

```
ALTER TABLE supply_chain_data
ADD [Compare per item] DECIMAL(15, 2);

UPDATE supply_chain_data
SET [Compare per item] = [Compare revenue]
 / [Number_of_products_sold];
```

Code [8]:

```
ALTER TABLE supply_chain_data
ADD [Diff Percent] DECIMAL(15, 2);
UPDATE supply_chain_data
SET [Diff Percent] = [Compare per item] / [Price];
```

Code [9]:

```
WITH Percentiles AS (
    SELECT
        [Diff_Percent],
        PERCENTILE_CONT(0.25) WITHIN GROUP
 (ORDER BY [Diff_Percent]) OVER () AS "25th Percentile",
        PERCENTILE_CONT(0.50) WITHIN GROUP
(ORDER BY [Diff_Percent]) OVER () AS median,
        PERCENTILE_CONT(0.75) WITHIN GROUP
(ORDER BY [Diff_Percent]) OVER () AS "75th Percentile"
    FROM supply_chain_data
)
SELECT
    COUNT([Diff_Percent]) AS count,
    AVG([Diff_Percent]) AS mean,
    MIN([Diff_Percent]) AS min,
    MAX([Diff_Percent]) AS max,
    MAX("25th Percentile") AS "25th Percentile",
    MAX(median) AS median,
    MAX("75th Percentile") AS "75th Percentile",
```

```
    STDEV([Diff_Percent]) AS stdev
FROM Percentiles;
```

	count	mean	min	max	25th Percentile	median	75th Percentile	stdev
1	100	1.563100	-0.97	104.76	-0.86	-0.645	0.0275	11.1298431922848

Code [10]:

- We found a very high value in Diff Percent ,so we assume that our upper limit is 200%

We cleared the outliers

```
SELECT
    (SELECT COUNT(*) FROM supply_chain_data
 WHERE [Diff_Percent] <= 2) AS row_count,
    (SELECT COUNT(COLUMN_NAME)
FROM INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'supply_chain_data')
AS column_count;
```

	row_count	column_count
1	88	28

Code [11]:

- By logic Order quantities can’t be greater than Number of products sold
- So we added column to calculate the difference between them

```
SELECT
    *,
    ([Number of products sold] - [Order quantities]) AS [S/Q Diff]
FROM
    supply_chain_data;
```

Code [12]:

- We cleared any values are negative because it is illogical values

```
DELETE FROM supply_chain_data
WHERE [Diff Percent] <= 2;

ALTER TABLE supply_chain_data
```

```
MODIFY [Diff Percent] DECIMAL(10, 2);

UPDATE supply_chain_data
SET some_column = some_value
WHERE [Diff Percent] <= 2;
```

- No additional rows cleared but in original data we found there were 3 values were negative but here these 3 values already were in 12 rows (deleted rows)

1.6 Product Type

Analyzing the **Product type** column in the dataset can reveal insights into the distribution and performance of different product categories

1. Product Type Distribution

- Understand how many products fall under each type (e.g., haircare, skincare).
- Visualize the proportion of different product types.

2. Revenue by Product Type

- Analyze which product types generate the most revenue.
- Compare total and average revenue across product types.

3. Sales by Product Type

- Evaluate how many units of each product type are sold.
- Find the most popular product type in terms of sales volume.

4. Stock and Availability by Product Type

- Investigate stock levels and availability for each product type to identify if there's overstock or low availability in any category.

5. Price Range of Products by Type

- Analyze the price range and average price of products in each category.
- Identify premium vs. budget product types based on pricing.

6. Lead Time and Order Quantities by Product Type

- Compare lead times and order quantities to understand the supply chain efficiency of different product types.

- Here's an overall analysis of the **Product type** across various metrics:

- Key Metrics by Product Type:

Product type	Total SKUs	Total Revenue	Avg Revenue	Total Units Sold	Avg Units Sold	Avg Stock Levels	Avg Availability	Min Price	Max Price	Avg Price	Avg Lead Times	Avg Order Quantities
Cosmetics	23	137,860.96	5,993.95	10,947	475.96	59.91	52.48	3.04	97.76	61.20	15.74	53.22
Haircare	27	128,949.38	4,775.90	13,100	485.19	50.74	42.33	3.17	97.45	51.34	15.22	45.81
Skincare	38	227,350.89	5,982.92	20,386	536.47	41.97	50.39	4.81	99.17	49.44	17.05	51.66

- **Key Insights:**
  - **Total Revenue**: Skincare products generate the most revenue, followed by cosmetics and haircare.
  - **Average Revenue**: Cosmetics and skincare have similar average revenues per product, with haircare slightly lower.
  - **Units Sold**: Skincare has the highest total and average number of units sold.
  - **Stock Levels**: Cosmetics maintain the highest average stock levels, while skincare has the lowest.

- **Availability**: Cosmetics have the highest average availability, while haircare has the lowest.
- **Price Range**: Skincare products have the highest max price ($99.17), while cosmetics have the highest average price ($61.20).
- **Lead Times**: Skincare has the longest lead times (17 days on average), while haircare has the shortest.
- **Order Quantities**: Cosmetics have the highest average order quantities.

1.7 SKU

Analyzing **Stock Keeping Units (SKUs)** is vital for managing inventory, understanding sales performance, and optimizing supply chain operations.

1. Sales Performance by SKU

- **Sales Volume**: Analyze the total sales volume for each SKU to identify high-performing and low-performing products.
- **Revenue Analysis**: Calculate the revenue generated by each SKU to understand its financial contribution to the business.

2. Inventory Levels by SKU

- **Stock Levels**: Assess current inventory levels for each SKU to identify overstocked or understocked items.
- **Turnover Rates**: Calculate the inventory turnover rate for each SKU to evaluate how quickly inventory is sold and replaced.

3. SKU Profitability

- **Cost Analysis**: Evaluate the cost of goods sold (COGS) for each SKU to determine profitability.
- **Gross Margin**: Calculate the gross margin for each SKU to identify which products contribute most to overall profit.

4. SKU Trends Over Time

- **Sales Trends**: Analyze sales trends for each SKU over time to identify seasonal patterns, trends, or shifts in consumer demand.
- **Inventory Trends**: Examine inventory levels over time to understand how stock levels change in response to sales trends.

5. SKU Analysis by Category

- **Category Performance**: Group SKUs by product category and analyze overall performance to identify successful categories and those needing attention.
- **Market Segmentation**: Analyze sales performance across different customer segments to determine which SKUs are favored by each segment.

6. SKU Stockout Analysis

- **Out-of-Stock Events**: Track instances of stockouts for each SKU to assess how they affect sales and customer satisfaction.
- **Lead Time Analysis**: Examine lead times associated with restocking SKUs to identify potential delays in availability.

7. SKU Demand Forecasting

- **Predictive Analytics**: Use historical sales data to forecast future demand for each SKU, helping to optimize inventory levels and reduce stockouts.

1.8 Price

Analyzing the **Price** of products provides valuable insights into the product strategy, market positioning, and profitability of different product categories.

1. Price Distribution

- Examine the overall distribution of product prices across the dataset.
- Create histograms or box plots to visualize how prices are spread out (e.g., identifying low, mid, and high-price ranges).

2. Price by Product Type

- Compare the price range (min, max, average) across different product types.
- Determine if certain product categories tend to be priced higher or lower.

3. Price vs. Sales

- Investigate the relationship between price and the number of products sold.
- Analyze if higher-priced products sell fewer units or if there's a premium market.

4. Price vs. Revenue

- Analyze whether higher-priced products generate more revenue.
- Determine if certain price points are more profitable despite lower sales volumes.

5. Price vs. Cost per Item

- Compare the price of products with their cost per item to analyze profit margins.
- Look for any products where price is unusually low compared to cost, indicating potential profitability issues.

6. Price vs. Availability and Stock Levels

- Examine if there's a correlation between product prices and their availability or stock levels.
- Higher-priced products might have lower availability due to lower demand or higher production times.

1.9 Availability

Analyzing **Availability** in a supply chain dataset can provide insights into how well products are stocked and ready to meet customer demand.

1. Availability Distribution

- Understand how product availability is distributed across the dataset.

- Identify which products are consistently in stock versus those that frequently run low.

2. Availability by Product Type

- Compare the average availability of different product types (e.g., haircare, skincare).
- Identify product categories with potential supply chain issues based on lower availability.

3. Availability vs. Sales

- Analyze the relationship between availability and the number of products sold.
- Check if lower availability impacts sales, indicating stock-out situations.

4. Availability vs. Revenue

- Investigate how availability impacts revenue generation.
- Products with high revenue and low availability may be popular and face stock issues, which could be addressed by improving stock levels.

5. Availability vs. Stock Levels

- Examine if there is a correlation between availability and stock levels.
- Products with low stock levels and low availability may be at risk of stockouts.

6. Availability vs. Lead Times

- Compare availability with lead times to understand if longer lead times contribute to lower availability.

1.10 Number of Product Sold

Analyzing the **Number of Products Sold** is essential to understanding product performance, customer demand, and overall sales trends.

1. Sales Distribution

- Analyze how sales are distributed across different products.
- Identify high-selling and low-selling products.
- Use histograms or box plots to visualize the distribution of sales across the dataset.

2. Sales by Product Type

- Compare total and average units sold for each product type (e.g., haircare, skincare).
- Identify which product categories perform the best in terms of sales volume.

3. Sales vs. Price

- Investigate the relationship between the price of a product and the number of units sold.
- Check if lower-priced products sell more units, or if premium products still maintain strong sales.

4. Sales vs. Availability

- Examine how product availability impacts the number of products sold.
- Identify if stock-out situations are affecting sales volumes.

5. Sales vs. Revenue

- Analyze if higher sales volumes translate directly into higher revenue, or if some products generate more revenue despite selling fewer units due to higher prices.

6. Sales vs. Marketing or Promotions

- If available, compare sales with promotional efforts, discounts, or marketing campaigns to measure the effectiveness of these efforts on sales volumes.

1.11 Revenue Generated

Analyzing **Revenue Generated** is critical to understanding the financial performance of products and identifying opportunities for growth.

1. Revenue Distribution

- Analyze how revenue is distributed across different products.
- Identify high-revenue and low-revenue products.
- Use histograms or box plots to visualize the distribution of revenue.

2. Revenue by Product Type

- Compare the total and average revenue generated by each product type (e.g., haircare, skincare).
- Identify which product categories are the most financially successful.

3. Revenue vs. Sales

- Investigate the relationship between the number of products sold and the revenue generated.
- Determine whether higher sales volumes lead to more revenue or if higher-priced products dominate revenue.

4. Revenue vs. Price

- Analyze the relationship between product price and revenue generated.
- Identify whether premium products are contributing more to overall revenue despite lower sales volumes.

5. Revenue vs. Availability

- Examine if product availability affects revenue.
- Products with low availability but high revenue may need inventory adjustments to meet demand.

6. Revenue vs. Lead Times

- Analyze how lead times affect revenue generation.
- Products with longer lead times but high revenue might indicate potential supply chain optimization opportunities.

1.12 Customer Demographics

Analyzing **Customer Demographics** is essential for understanding your target audience, tailoring marketing strategies, and improving customer satisfaction.

1. Demographic Segmentation

- **Gender Distribution**: Examine the gender distribution of customers to understand the preferences of different genders regarding products or services.
- **Income Levels**: Analyze customer income levels (if available) to determine how income affects purchasing behavior and preferences.

2. Geographic Analysis

- **Location Data**: Analyze customer locations (city, state, region) to identify where your products are most popular and tailor marketing efforts accordingly.

3. Customer Behavior Insights

- **Purchase Patterns**: Analyze purchasing behaviors across different demographic segments to understand how demographics influence buying decisions.

4. Customer Preferences

- **Product Preferences**: Analyze which products or services are favored by different demographic groups to tailor offerings.

1.13 Stock Levels

Analyzing **Stock Levels** helps in understanding inventory management, identifying potential stock-out risks, and improving supply chain efficiency.

1. Stock Levels Distribution

- Analyze the overall distribution of stock levels across all products.
- Identify products with consistently low or high stock levels to understand potential overstock or understock situations.

2. Stock Levels by Product Type

- Compare the average stock levels across different product types (e.g., haircare, skincare).
- Identify if any product types tend to have lower or higher stock levels and analyze if these levels align with their sales or availability.

3. Stock Levels vs. Sales

- Investigate the relationship between stock levels and the number of products sold.

- Check if low stock levels correspond with high sales, indicating products that may need restocking.

4. Stock Levels vs. Availability

- Examine how stock levels correlate with product availability.
- Products with high stock levels but low availability might indicate inefficiencies in making those products available to customers.

5. Stock Levels vs. Lead Times

- Analyze if products with longer lead times tend to maintain higher stock levels to avoid stock-outs.

6. Stock Levels vs. Revenue

- Compare stock levels with the revenue generated by each product.
- Identify if any high-revenue products are at risk of running out of stock.

7. Stock Levels vs. Order Quantities

- Investigate the relationship between stock levels and order quantities to see if stock is being replenished at an optimal rate to meet demand.

1.14 Lead Times

Analyzing **Lead Times** is essential for understanding supply chain efficiency, delivery performance, and customer satisfaction.

1. Lead Times Distribution

- Analyze the distribution of lead times across all products.
- Use histograms or box plots to visualize lead time distributions and identify any outliers.

2. Lead Times by Product Type

- Compare average lead times for different product types (e.g., haircare, skincare).
- Identify if certain categories have significantly longer or shorter lead times.

3. Lead Times vs. Availability

- Investigate the relationship between lead times and product availability.
- Longer lead times might correlate with lower availability for certain products, indicating potential supply chain issues.

4. Lead Times vs. Stock Levels

- Examine if products with longer lead times tend to have higher stock levels to mitigate stock-outs.
- Analyze how stock levels affect the ability to fulfill orders within expected lead times.

5. Lead Times vs. Sales

- Analyze how lead times impact the number of products sold.
- Determine if shorter lead times result in higher sales, indicating customer preference for quick delivery.

6. Lead Times vs. Revenue

- Compare lead times with revenue generated to identify if high-revenue products have longer lead times, which may affect customer purchasing decisions.

8. Lead Times vs. Order Quantities

- Investigate the relationship between lead times and order quantities to see if larger orders correlate with longer lead times.

1.15 Order Quantities

Analyzing **Order Quantities** is crucial for understanding purchasing behavior, inventory management, and supply chain efficiency.

1. Order Quantities Distribution

- Analyze the overall distribution of order quantities across products.
- Use histograms or box plots to visualize how order quantities vary and identify any outliers.

2. Order Quantities by Product Type

- Compare average order quantities for different product types (e.g., haircare, skincare).
- Identify which product categories tend to have larger or smaller order quantities.

3. Order Quantities vs. Sales

- Investigate the relationship between order quantities and the number of products sold.
- Check if higher order quantities correlate with increased sales, indicating purchasing trends.

4. Order Quantities vs. Stock Levels

- Analyze how order quantities relate to current stock levels.
- Larger order quantities might suggest that stock levels need to be adjusted to meet future demand.

5. Order Quantities vs. Lead Times

- Examine if larger order quantities result in longer lead times due to processing or shipping complexities.
- Determine if shorter lead times lead to smaller order quantities.

6. Order Quantities vs. Revenue

- Analyze the relationship between order quantities and revenue generated.

- Determine if higher order quantities lead to increased revenue or if certain products generate more revenue despite smaller orders.

7. Order Quantities vs. Availability

- Investigate if order quantities are influenced by product availability.
- Check if customers order more when products are readily available versus when stock is low.

1.16 Shipping Times

Analyzing **Shipping Times** is vital for understanding supply chain efficiency, customer satisfaction, and overall operational performance.

1. Shipping Times Distribution

- Analyze the overall distribution of shipping times across all orders.
- Use histograms or box plots to visualize how shipping times vary and identify outliers.

2. Shipping Times by Product Type

- Compare average shipping times for different product types (e.g., haircare, skincare).
- Identify which product categories experience longer or shorter shipping times.

3. Shipping Times vs. Order Quantities

- Investigate the relationship between shipping times and order quantities.
- Determine if larger orders lead to longer shipping times due to processing or logistics challenges.

4. Shipping Times vs. Lead Times

- Examine how shipping times correlate with lead times.
- Analyze if longer lead times correlate with longer shipping times, which may indicate bottlenecks in the supply chain.

5. Shipping Times vs. Availability

- Assess how product availability impacts shipping times.
- Determine if products that are frequently out of stock lead to longer shipping times when they are back in stock.

6. Shipping Times vs. Sales

- Analyze the relationship between shipping times and the number of products sold.
- Identify if faster shipping times correlate with increased sales, indicating customer preference for quicker delivery.

7. Shipping Times vs. Revenue

- Compare shipping times with revenue generated to understand if high-revenue products have longer shipping times, which could affect customer purchasing decisions.

8. Shipping Times by Region

- If location data is available, analyze how shipping times vary by region to identify any geographical challenges in logistics.

1.17 Shipping Carriers

Analyzing **Shipping Carriers** is important for evaluating their performance, reliability, and overall impact on your supply chain.

1. Carrier Performance Metrics

- **Average Shipping Time**: Analyze the average shipping time for each carrier to assess their efficiency.

2. Cost Analysis

- **Shipping Costs**: Compare the shipping costs charged by different carriers for similar services. This helps in determining the cost-effectiveness of each carrier.
- **Cost vs. Delivery Time**: Analyze the relationship between shipping costs and delivery times to identify which carriers provide the best value.

3. Shipping Volume Analysis

- **Volume of Shipments**: Assess the total volume of shipments handled by each carrier. This helps in understanding their capacity and market share.

4. Carrier Comparison by Product Type

- **Performance by Product**: Compare carrier performance for different product types to identify if certain carriers excel with specific products.
- **Special Handling Requirements**: Analyze how well carriers handle products with special requirements (e.g., fragile items, perishables).

5. Regional Performance

- **Geographic Analysis**: Analyze the performance of carriers in different regions to identify strengths and weaknesses based on geographical challenges.

1.18 Shipping Costs

Analyzing **Shipping Costs** is essential for optimizing your logistics, reducing expenses, and improving overall supply chain efficiency.

1. Cost Breakdown

- **Detailed Cost Analysis**: Break down shipping costs into components such as freight charges, handling fees, fuel surcharges, insurance, and any additional fees (e.g., customs duties).
- **Carrier Comparison**: Compare costs across different shipping carriers to identify the most cost-effective options.

2. Cost per Order Analysis

- **Average Shipping Cost per Order**: Calculate the average shipping cost per order to understand how shipping costs impact overall profitability.
- **Cost Variability**: Analyze how shipping costs vary by order size or weight, identifying any trends in cost efficiency.

3. Shipping Costs by Region

- **Geographic Analysis**: Analyze shipping costs by region to identify areas with higher shipping expenses, which may warrant further investigation or alternative strategies.
- **International Shipping Costs**: If applicable, compare domestic and international shipping costs to evaluate global shipping strategies.

5. Shipping Costs vs. Delivery Times

- **Cost-Effectiveness Analysis**: Analyze the relationship between shipping costs and delivery times to identify opportunities for cost savings while maintaining service levels.
- **Service Level Agreements (SLAs)**: Assess how different shipping options and their associated costs align with customer expectations regarding delivery times.

6. Impact on Pricing and Profitability

- **Cost-to-Serve Analysis**: Evaluate how shipping costs affect overall pricing strategies and profitability margins for different products.

- **Customer Impact**: Analyze how shipping costs influence customer purchasing decisions, particularly for free or discounted shipping offers.

7. Cost Reduction Opportunities

- **Optimization Strategies**: Identify opportunities for reducing shipping costs, such as bulk shipping, renegotiating contracts with carriers, or implementing more efficient packaging strategies.
- **Automation**: Consider the impact of technology and automation on shipping cost reduction, such as using logistics software to optimize routes.

1.19 Supplier Name

Analyzing **Supplier Names** is crucial for understanding your supply chain dynamics, assessing supplier performance, and identifying opportunities for improvement.

1. Supplier Performance Metrics

- **Quality of Products**: Evaluate the quality of products received from each supplier by tracking defect rates or return rates associated with each supplier.

2. Cost Analysis by Supplier

- **Price Comparison**: Compare prices charged by different suppliers for similar products to identify the most cost-effective options.

- **Total Cost of Ownership**: Consider not just the purchase price but also shipping costs, lead times, and any additional fees associated with each supplier.

3. Supplier Diversity

- **Source Variety**: Assess the diversity of suppliers to ensure that you're not overly reliant on a single supplier, which can pose risks in case of disruptions.
- **Geographic Analysis**: Evaluate the geographic distribution of suppliers to identify potential risks related to regional disruptions.

4. Supplier Contribution to Revenue

- **Sales Analysis**: Analyze the contribution of each supplier to overall sales to identify key suppliers and their impact on revenue.
- **Product Mix**: Examine the range of products supplied by each supplier to understand their significance in your product portfolio.

5. Supplier Risk Assessment

- **Risk Factors**: Identify risk factors associated with each supplier, such as financial stability, reputation, and compliance with regulations.
- **Contingency Planning**: Develop contingency plans for critical suppliers to mitigate risks.

7. Supplier Innovation

- **Innovation Contributions**: Evaluate how suppliers contribute to innovation in your product offerings or processes.

1.20 Location

Analyzing **Location** data is critical for understanding market dynamics, optimizing logistics, and improving customer satisfaction.

1. Geographic Distribution

- **Customer Location Analysis**: Map customer locations to identify geographical patterns and concentrations. This can help in tailoring marketing strategies and distribution channels.
- **Sales by Region**: Analyze sales performance by region to understand which areas contribute most to revenue and which may need more focus.

2. Market Segmentation

- **Demographic Breakdown**: Segment customers by location and analyze demographic factors (age, income, etc.) to understand regional preferences and behaviors.
- **Target Market Identification**: Identify key markets based on location data to prioritize marketing efforts and resource allocation.

3. Shipping and Logistics

- **Shipping Costs by Location**: Analyze shipping costs associated with different locations to optimize logistics strategies and minimize expenses.
- **Delivery Times**: Assess how location affects delivery times, identifying areas that may experience delays or higher costs.

4. Location-Based Performance Metrics

- **Store Performance**: If applicable, analyze the performance of physical store locations to identify high-performing and low-performing locations.

5. Supply Chain Considerations

- **Supplier Location**: Analyze the geographical distribution of suppliers to assess risks and optimize procurement strategies.
- **Regional Regulations**: Consider how local regulations may impact operations in different locations.

1.21 Lead Time

Analyzing **Lead Time** is crucial for improving supply chain efficiency, customer satisfaction, and inventory management.

1. Lead Time Breakdown

- **Components of Lead Time**: Break down lead time into its components, such as order processing time, manufacturing

time, and shipping time. This helps identify specific areas for improvement.

- **Average Lead Time**: Calculate the average lead time for orders to establish a baseline for performance.

2. Lead Time by Product Type

- **Analysis by SKU**: Analyze lead times for different products or SKUs to identify which items have longer lead times and may require attention.
- **Category Performance**: Evaluate lead times across product categories to see if certain categories consistently take longer to fulfill.

3. Lead Time Variability

- **Standard Deviation**: Calculate the standard deviation of lead times to assess variability and identify orders that fall outside typical ranges.
- **Root Cause Analysis**: Investigate the causes of lead time variability, such as supplier delays, production bottlenecks, or shipping issues.

5. Supplier Performance

- **Supplier Lead Times**: Assess lead times associated with different suppliers to identify reliable and unreliable partners.
- **Supplier Comparison**: Compare lead times across suppliers for the same products to determine which suppliers provide the best service.

6. Inventory Management

- **Safety Stock Levels**: Analyze how lead times impact safety stock levels and inventory management practices. Longer lead times may require higher safety stock.
- **Reorder Points**: Evaluate the effectiveness of current reorder points based on lead times to prevent stockouts.

1.22 Production Volumes

Analyzing **Production Volumes** is essential for understanding manufacturing efficiency, capacity utilization, and overall operational performance.

1. Production Capacity

- **Capacity Utilization**: Calculate the percentage of production capacity being used to assess efficiency. This can help identify underutilized resources or overcapacity.
- **Bottleneck Analysis**: Identify any bottlenecks in the production process that may limit production volumes and suggest improvements.

2. Product-Specific Analysis

- **Volume by SKU**: Analyze production volumes by SKU to understand which products are produced in higher quantities and which may need adjustments.
- **Product Line Performance**: Evaluate the performance of different product lines based on production volumes to identify high-performing and low-performing lines.

3. Efficiency Metrics

- **Cycle Time**: Measure the average cycle time for production processes to assess efficiency. Shorter cycle times often correlate with higher production volumes.
- **Yield Rates**: Analyze yield rates (the ratio of finished goods to raw materials) to understand how efficiently production resources are being used.

4. Cost Analysis

- **Cost per Unit**: Calculate the cost per unit of production to understand how production volumes impact overall costs and profitability.
- **Fixed vs. Variable Costs**: Analyze fixed and variable costs associated with production to assess the impact of changes in production volumes on profitability.

5. Supply Chain Considerations

- **Lead Times**: Analyze lead times for materials and components to determine how they impact production scheduling and volumes.

7. Quality Control

- **Defect Rates**: Track defect rates and rework levels to assess how quality issues impact production volumes. High defect rates may lead to lower overall volumes.
- **Quality Improvement Initiatives**: Evaluate the effectiveness of any quality improvement initiatives on production volumes over time.

1.23 Manufacturing Lead Time

Analyzing **Manufacturing Lead Time** is essential for optimizing production processes, improving customer satisfaction, and enhancing overall operational efficiency.

1. Understanding Components of Manufacturing Lead Time

- **Breakdown of Lead Time**: Analyze the various components that make up manufacturing lead time, such as:
  - Order processing time
  - Production time
  - Quality control and inspection time
  - Packaging and labeling time
  - Transportation time to distribution centers or customers
- **Time Tracking**: Use time tracking methods (e.g., time studies, ERP systems) to gather accurate data on each component.

2. Calculate Average Manufacturing Lead Time

- **Overall Average**: Calculate the average manufacturing lead time for all products to establish a baseline.
- **By SKU/Category**: Calculate average lead times by SKU or product category to identify specific items with longer lead times.

3. Variability Assessment

- **Standard Deviation**: Calculate the standard deviation of lead times to understand the variability in manufacturing

processes. High variability may indicate inefficiencies or inconsistencies.

- **Identify Outliers**: Identify orders with significantly longer lead times and investigate the root causes.

5. Bottleneck Identification

- **Process Mapping**: Map out the entire manufacturing process to identify bottlenecks that contribute to longer lead times.
- **Root Cause Analysis**: Conduct root cause analysis to determine the reasons for delays in specific stages of manufacturing.

6. Supplier Performance

- **Supplier Lead Times**: Assess lead times associated with different suppliers to identify those that contribute to longer manufacturing lead times.
- **Collaboration Opportunities**: Work with suppliers to streamline processes and reduce lead times collaboratively.

7. Impact of Order Size

- **Order Quantity Analysis**: Analyze how order quantities affect manufacturing lead times. Larger orders may result in economies of scale but could also lead to longer lead times if not managed effectively.

8. Quality Control Impact

- **Inspection Times**: Analyze how quality control processes and inspection times impact overall manufacturing lead time.
- **Defect Rates**: Assess the relationship between defect rates and manufacturing lead times to identify areas for improvement.

1.24 Manufacturing Costs

Analyzing **Manufacturing Costs** is crucial for understanding production efficiency, profitability, and areas for improvement within the manufacturing process.

1. Cost Breakdown

- **Categorize Costs**: Divide manufacturing costs into fixed and variable costs, and further categorize them into:
  - **Direct Costs**: Costs directly tied to production, such as raw materials, labor, and manufacturing supplies.
  - **Indirect Costs**: Overhead costs not directly attributable to a specific product, such as utilities, rent, and administrative expenses.
- **Activity-Based Costing**: Implement activity-based costing to gain insight into the true costs associated with specific products or processes.

2. Calculate Total Manufacturing Costs

- **Cost Formula**: Use the formula:

$$\text{Total Manufacturing Costs} = \text{Direct Materials} + \text{Direct Labor} + \text{Manufacturing Overhead}$$

- **Cost per Unit**: Calculate the cost per unit of product to understand the cost structure and profitability at a granular level.

3. Efficiency Metrics

- **Cost Efficiency Ratios**: Calculate ratios such as cost per unit produced, labor cost per unit, and material cost per unit to assess overall efficiency.

4. Impact of Production Volume

- **Economies of Scale**: Analyze how changes in production volume affect total manufacturing costs. Higher production volumes may lead to lower per-unit costs due to economies of scale.
- **Variable Cost Analysis**: Assess how variable costs change with fluctuations in production volume.

5. Quality Control Impact

- **Defect Rates**: Analyze the relationship between quality control measures and manufacturing costs. High defect rates can lead to increased costs due to rework and waste.
- **Cost of Poor Quality**: Assess the costs associated with poor quality, including returns, warranty claims, and lost sales.

1.25 Inspection Results

Analyzing **Inspection Results** is crucial for maintaining product quality, reducing defects, and improving overall operational efficiency.

1. Data Collection and Organization

- **Inspection Data Sources**: Collect data from various inspection points, such as incoming materials, in-process inspections, and final product checks.
- **Data Organization**: Organize inspection data into a structured format, typically in a spreadsheet or database, with relevant fields such as:
  - Product type/SKU
  - Results (pass/fail)

2. Calculate Key Metrics

- **Defect Rate**: Calculate the defect rate as the percentage of inspected items that fail quality checks:

$$\text{Defect Rate} = \left(\frac{\text{Number of Defects}}{\text{Total Inspected}} \right) \times 100$$

- **First Pass Yield (FPY)**: Calculate the FPY to measure the efficiency of the inspection process:

$$\text{FPY} = \left(\frac{\text{Number of Units Passed}}{\text{Total Inspected}} \right) \times 100$$

- **Rework Rate**: Calculate the rework rate to assess how many units require re-inspection or rework.

3. Root Cause Analysis

- **Defect Categorization**: Categorize defects by type (e.g., cosmetic, functional) and severity to identify recurring issues.
- **Fishbone Diagram**: Use tools like the Fishbone diagram (Ishikawa) to identify potential root causes of defects and develop action plans.

4. Supplier Performance Evaluation

- **Supplier Defect Rates**: Analyze inspection results by supplier to evaluate their quality performance and identify areas for improvement.
- **Supplier Comparison**: Compare defect rates across different suppliers to inform sourcing decisions.

5. Correlation Analysis

- **Quality vs. Production Variables**: Analyze the correlation between inspection results and other production variables (e.g., production volume) to understand their impact on quality.
- **Customer Complaints**: Correlate inspection results with customer complaints or returns to identify trends and areas needing attention.

6. Compliance and Standards

- **Regulatory Compliance**: Ensure that inspection results meet industry regulations and standards (e.g., ISO, FDA) and track compliance metrics.
- **Certification Status**: Analyze how inspection results impact product certification status or market eligibility.

1.26 Defect Rates

Analyzing **Defect Rates** is essential for quality management, process improvement, and customer satisfaction.

1. Data Collection

- **Inspection Data**: Gather data from quality inspections, including the total number of units inspected, the number of defective units, and the type and severity of defects.

2. Calculate Defect Rate

- **Defect Rate Formula**: Calculate the defect rate using the formula:

$$\text{Defect Rate} = \left(\frac{\text{Number of Defective Units}}{\text{Total Units Inspected}} \right) \times 100$$

- **Defect Rate by Category**: Calculate defect rates for different categories or types of defects to pinpoint areas needing improvement.

3. Root Cause Analysis

- **Defect Categorization**: Categorize defects by type (e.g., cosmetic, functional) and severity to identify recurring issues.
- **Tools for Analysis**: Use tools like Pareto charts to visualize the most common defect types and focus on the "vital few" causing the majority of defects.

4. Supplier Performance

- **Supplier Defect Rates**: Analyze defect rates associated with different suppliers to assess their performance and reliability.
- **Comparative Analysis**: Compare defect rates across suppliers to inform procurement decisions.

5. Impact Assessment

- **Cost of Defects**: Calculate the cost associated with defects, including rework, scrap, returns, and lost sales.
- **Correlation with Other Metrics**: Analyze how defect rates correlate with other operational metrics, such as production volume, labor hours, or customer complaints.

1.27 Transportation Modes

Analyzing **Transportation Modes** is crucial for optimizing logistics, reducing costs, and improving delivery efficiency in the supply chain.

1. Data Collection

- **Gather Data**: Collect data on different transportation modes used, including:
  - Mode type (e.g., truck, rail, air, sea)
  - Costs associated with each mode

2. Cost Analysis

- **Cost per Unit**: Calculate the cost per unit shipped for each transportation mode to assess cost efficiency.

$$\text{Cost per Unit} = \frac{\text{Total Transportation Cost}}{\text{Total Units Shipped}}$$

- **Comparison of Modes**: Compare the costs of different modes to identify the most cost-effective options for various shipment scenarios.

3. Risk Analysis

- **Risk Factors**: Identify risks associated with each transportation mode, such as delays, damage, theft, or regulatory compliance issues.
- **Mitigation Strategies**: Develop strategies to mitigate risks, such as diversifying transportation modes or investing in better tracking technologies.

4. Scenario Analysis

- **Cost-Effectiveness Analysis**: Conduct scenario analyses to compare the cost-effectiveness of different transportation modes under varying conditions (e.g., distance, volume).
- **What-If Scenarios**: Model different transportation scenarios to assess how changes in demand, costs, or other factors might affect mode selection.

1.28 Routes

Analyzing **Routes** in the context of logistics and supply chain management is essential for optimizing transportation efficiency, reducing costs, and improving delivery times.

1. Data Collection

- **Gather Route Data**: Collect data on the routes taken for transportation, including:
  - Route distances and travel times
  - Delivery locations
  - Transportation modes used

2. Route Performance Metrics

- **Travel Time Analysis**: Calculate the average travel time for each route to evaluate efficiency and identify delays.
- **Cost Analysis**: Analyze costs associated with each route, including fuel costs, tolls, labor costs, and vehicle maintenance.

$$\text{Cost per Route} = \frac{\text{Total Route Costs}}{\text{Total Trips on Route}}$$

3. Route Variability

- **Identify Variability**: Analyze variability in travel times and costs for each route. High variability may indicate potential issues or inefficiencies.
- **Bottleneck Identification**: Identify bottlenecks or recurring issues affecting specific routes and develop strategies to mitigate them.

4. Risk Assessment

- **Risk Factors**: Identify risks associated with each route, such as traffic congestion, road conditions, weather impacts, and regulatory challenges.
- **Mitigation Strategies**: Develop strategies to mitigate identified risks, such as alternative routing or scheduling adjustments.

9. Geographical Considerations

- **Geographical Analysis**: Analyze the geographical factors affecting route performance, such as terrain, urban vs. rural routes, and traffic patterns.
- **Regional Comparisons**: Compare route performance across different regions to identify best practices or areas needing improvement.

1.29 Costs

Analyzing **Costs** is critical for understanding profitability, managing budgets, and identifying areas for operational improvement in any organization.

1. Cost Classification

- **Cost Categories**: Classify costs into relevant categories, such as:
  - Production costs
  - Operating costs
  - Administrative costs
  - Selling and distribution costs
  - Research and development costs
- **Direct vs. Indirect Costs**: Differentiate between direct costs (attributable to a specific product) and indirect costs (overhead).

2. Total Cost Calculation

- **Calculate Total Costs**: Sum all fixed, variable, and semi-variable costs to determine total costs over the selected time frame.

$$\text{Total Costs} = \text{Fixed Costs} + \text{Variable Costs} + \text{Semi-Variable Costs}$$

3. Cost per Unit

- **Cost Analysis per Unit**: Calculate the cost per unit produced to understand profitability at a granular level.

$$\text{Cost per Unit} = \frac{\text{Total Costs}}{\text{Total Units Produced}}$$

4. Cost Variance Analysis

- **Variance Calculation**: Analyze variances between actual and budgeted costs to identify discrepancies and their causes.

$$\text{Variance} = \text{Actual Costs} - \text{Budgeted Costs}$$

- **Investigate Causes**: Identify the reasons for cost variances, such as inefficiencies, price changes, or unexpected expenses.

1.30 Conclusion

Creating insightful analysis questions based on your dataset can significantly help decision-makers in understanding trends, identifying issues, and making informed decisions. Here are various data analysis questions that could be deduced from the dataset:

CHAPTER 2: ANALYSIS QUESTION

General Questions

1. **Product Performance**:
   - What is the total revenue generated by each product type?

- How does product type influence the number of products sold?
- What is the average price of products across different product types?

2. **Price Analysis**:
   - What is the relationship between pricing strategies and sales volume?
   - How do discounts affect the revenue generated per product type?
3. **Availability and Stock Levels**:
   - What is the correlation between product availability and sales performance?
   - How do stock levels affect the order quantities for each product?
   - What percentage of products are consistently available for sale?
4. **Lead Times and Order Quantities**:
   - How do lead times impact the order quantities placed for different products?
   - What is the average lead time across various suppliers, and how does it affect order fulfillment?
5. **Shipping and Logistics**:
   - What are the average shipping times for each transportation mode used?
   - How do shipping costs vary by product type and shipping carrier?

Revenue-Related Questions

6. **Revenue Analysis**:
    - What factors contribute most significantly to overall revenue growth?
    - How does the customer demographic impact the revenue generated?
    - What is the relationship between the number of products sold and revenue?
7. **Profitability**:
    - Which product types yield the highest profit margins?
    - How do manufacturing and inspection costs impact overall profitability?

Customer Insights

8. **Customer Demographics**:
    - What are the key demographics of our customer base, and how do they influence purchasing behavior?
    - Are there specific customer segments that prefer certain product types or brands?
9. **Order Patterns**:
    - What are the most common order quantities placed by customers?

Supplier and Production Analysis

10. **Supplier Performance**:
    - Which suppliers consistently meet lead times, and how does this impact overall sales?
    - What is the defect rate per supplier, and how does this influence product availability?
11. **Production Volumes**:
    - How do changes in manufacturing lead times affect product availability and sales?

Quality Control and Inspection

12. **Inspection Results**:
    - What are the common defects found during inspections, and how do they impact sales?

CHAPTER 3: FORECASTING QUESTIONS

1.31 Data Correlation

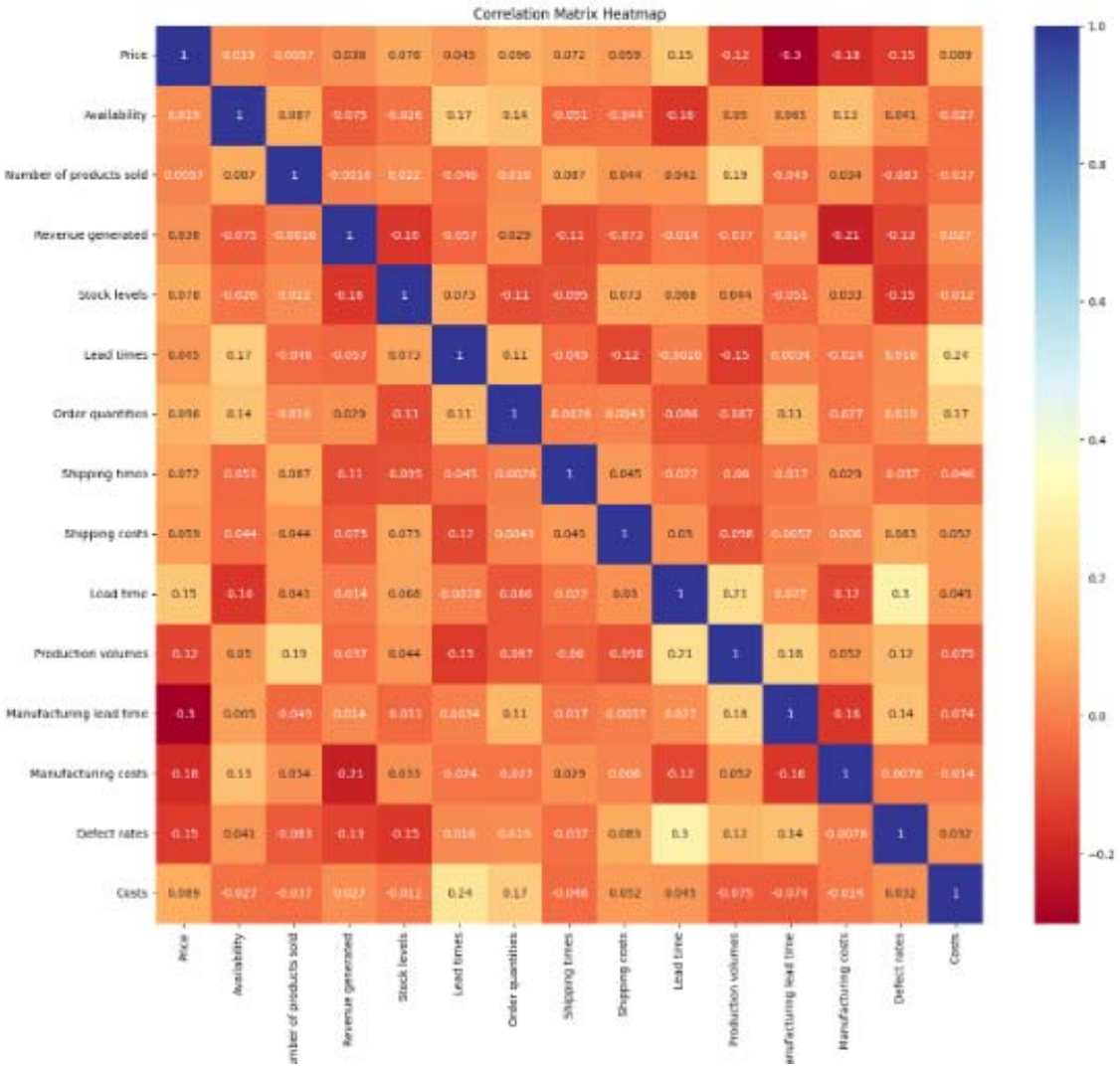
- At first we wanted to study the correlation between different fields So we can set the forecasting questions.

Python Code [16]:

```
correlation_matrix = supply_data.corr(numeric_only=True)
plt.figure(figsize=(16, 14))
```

```
sns.heatmap(correlation_matrix, annot=True, cmap='RdYlBu')
plt.title('Correlation Matrix Heatmap')
plt.show()
```

Output [16]:



1.32 Economic order quantity (EOQ)

Python Code [17]:

```
import numpy as np

def calculate_eoq(data):
    S = data['Costs']
    D = data['Number of products sold']
    H = data['Number of products sold'] * holdingcost
    EOQ = np.sqrt((2*S*D)/H)
    return round(EOQ)

supply_data['EOQ'] = calculate_eoq(supply_data)
```

```
comparison_columns = supply_data[['SKU', 'EOQ','Order quantities']]
comparison_columns.head()
```

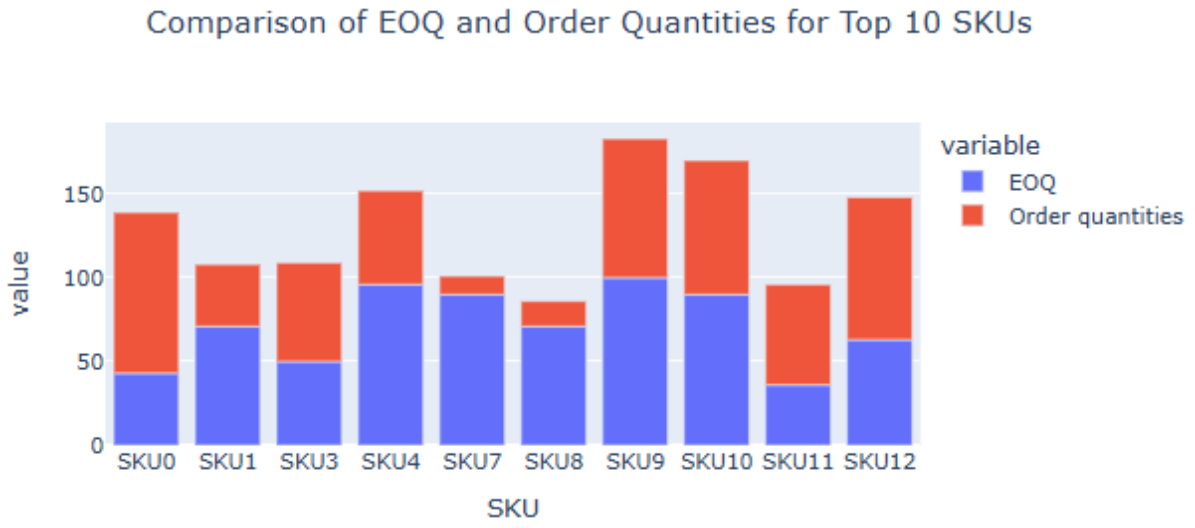
Output [17]:

	SKU	EOQ	Order quantities
0	SKU0	43.0	96
1	SKU1	71.0	37
3	SKU3	50.0	59
4	SKU4	96.0	56
7	SKU7	90.0	11

Python Code [18]:

```
top_10_comparison = comparison_columns.head(10)

fig = px.bar(top_10_comparison, x='SKU', y=['EOQ', 'Order quantities'],
             title='Comparison of EOQ and Order Quantities for Top 10 SKUs')

fig.update_layout(
    title_x=0.5
)

fig.show()
```

Output [18]:



1.33 Forecasting by Scikit-learn (Linear Regression)

Python Code [19]:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

- Preprocessing data

Python Code [20]:

```
features = ['Price', 'Availability', 'Stock levels', 'Lead times', 'Order quantities',
            'Shipping times', 'Shipping costs', 'Lead time', 'Production volumes',
            'Manufacturing lead time', 'Manufacturing costs', 'Defect rates', 'Costs',
            'Customer demographics', 'Shipping carriers', 'Supplier name', 'Location',
            'Inspection results', 'Transportation modes', 'Routes']


X = data[features]
y = data['Number of products sold']
```

- One-Hot Encoding for categorical features

Python Code [21]:

```
categorical_features = ['Customer demographics', 'Shipping carriers', 'Supplier name',

                        'Location', 'Inspection results', 'Transportation modes', 'Routes']
```

- Create the Column Transformer with One Hot Encoder for categorical features

Python Code [22]:

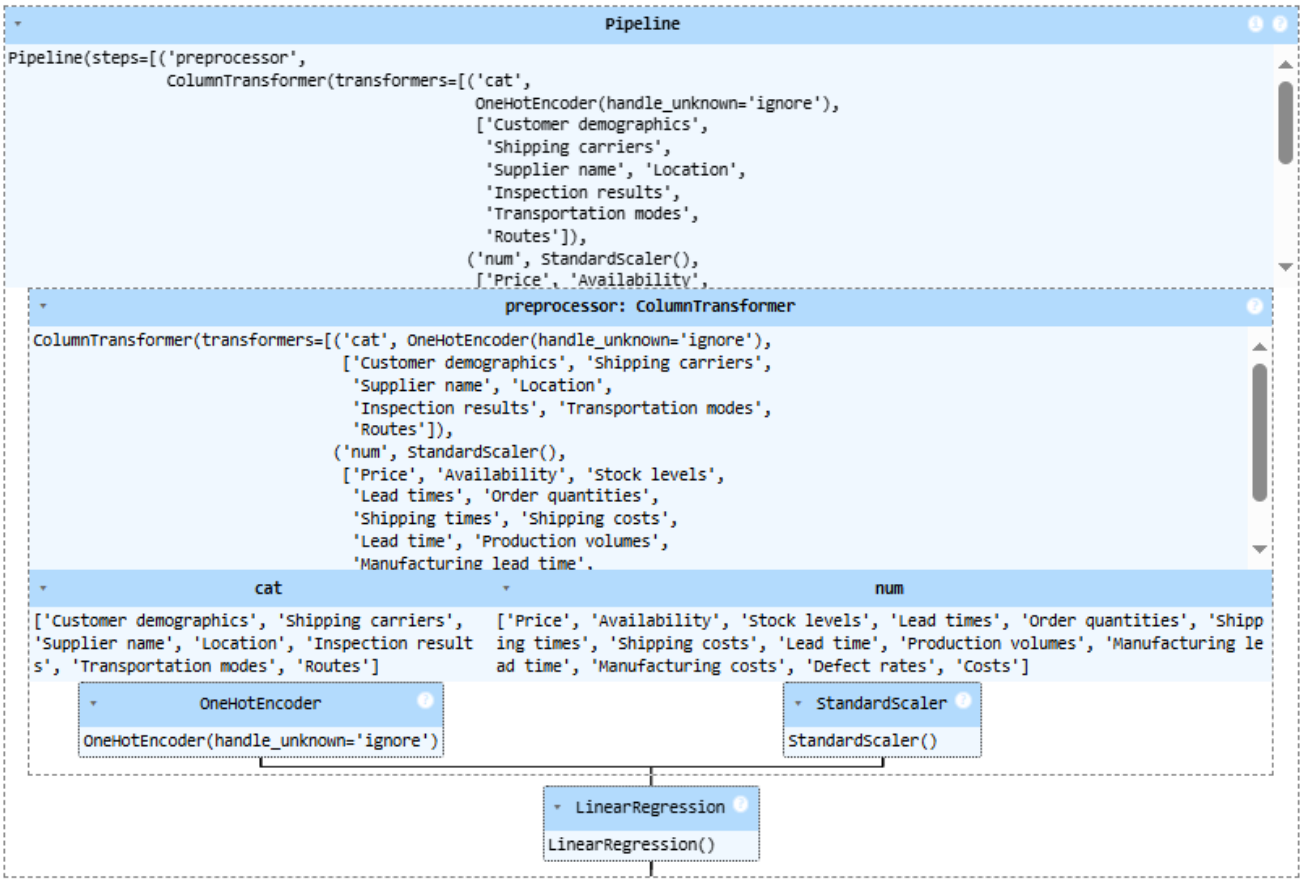
```
preprocessor = ColumnTransformer(
    transformers=[
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features),
        ('num', StandardScaler(), ['Price', 'Availability', 'Stock levels', 'Lead times',
                                   'Order quantities', 'Shipping times', 'Shipping costs',
                                   'Lead time', 'Production volumes', 'Manufacturing lead time',
                                   'Manufacturing costs', 'Defect rates', 'Costs'])
    ]
)
```

- Split the data into training and testing sets and create and apply pipeline

Python Code [23]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=4
2)

pipeline = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('model', LinearRegression())
])

pipeline.fit(X_train, y_train)
```

Output [23]:



- Make predictions and evaluate the model

Python Code [24]:

```
y_pred_train = pipeline.predict(X_train)
y_pred_test = pipeline.predict(X_test)

# Evaluate the model
mse_train = mean_squared_error(y_train, y_pred_train)
r2_train = r2_score(y_train, y_pred_train)
mse_test = mean_squared_error(y_test, y_pred_test)
r2_test = r2_score(y_test, y_pred_test)

print(f'Training MSE: {mse_train:.2f}')
print(f'Training R2: {r2_train:.2f}')
print(f'Testing MSE: {mse_test:.2f}')
print(f'Testing R2: {r2_test:.2f}')
```

Output [24]:

Training R2: 0.47

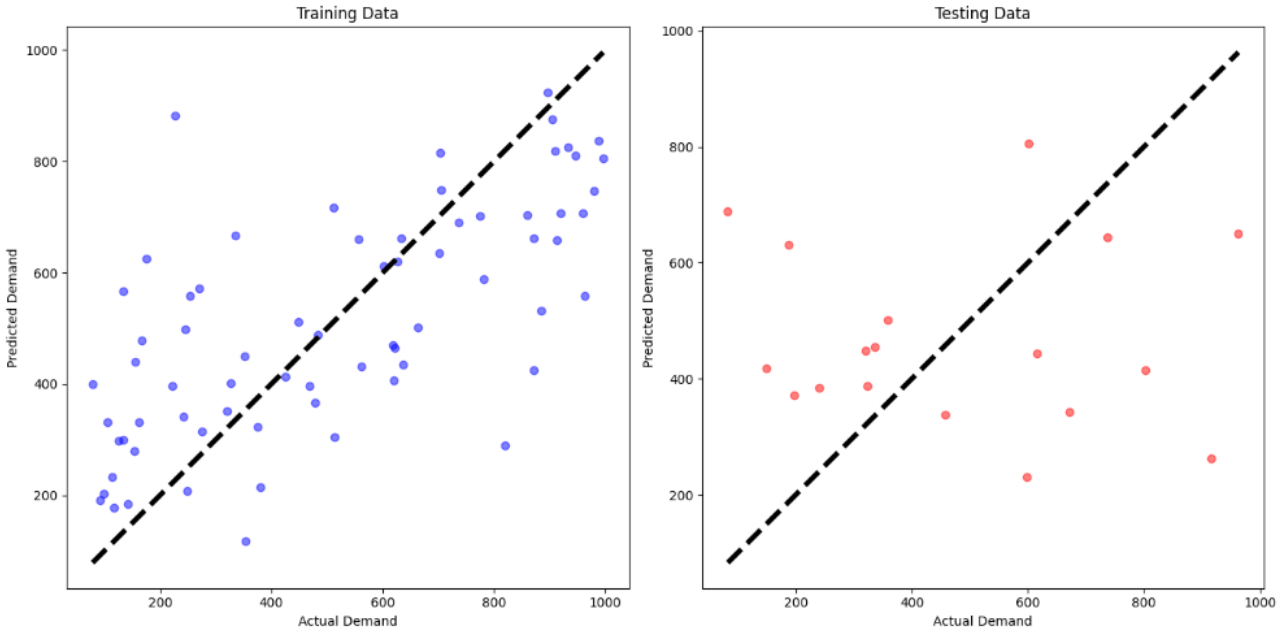
Testing MSE: 97368.74

Testing R2: -0.40

Python Code [25]:

```
import matplotlib.pyplot as plt

# Plot results
plt.figure(figsize=(14, 7))

plt.subplot(1, 2, 1)
plt.scatter(y_train, y_pred_train, color='blue', alpha=0.5)
plt.plot([y_train.min(), y_train.max()], [y_train.min(), y_train.max()], 'k--', lw=4)
plt.xlabel('Actual Demand')
plt.ylabel('Predicted Demand')
plt.title('Training Data')

plt.subplot(1, 2, 2)
plt.scatter(y_test, y_pred_test, color='red', alpha=0.5)
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=4)
plt.xlabel('Actual Demand')
plt.ylabel('Predicted Demand')
plt.title('Testing Data')

plt.tight_layout()
plt.show()
```

Output [25]:



- Combine test data with predictions for better visualization and determine if there is predicted demand or not

Python Code [26]:

```
results = X_test.copy()
results['Actual Demand'] = y_test
results['Predicted Demand'] = y_pred_test

results['Demand Status'] = results['Predicted Demand'].apply(lambda x: 'Demand' if x
 > 0 else 'No Demand')

# Include SKU in the results for clarity
results = results.join(data[['SKU']], how='left')

results
```

Output [26]: (part of the result predicted sales)

Shipping times	Shipping costs	Lead time	Production volumes	Manufacturing lead time	...	Shipping carriers	Supplier name	Location	Inspection results	Transportation modes	Routes	Actual Demand	Predicted Demand	Demand Status	SKU
7	3.801253	21	953	11	...	Carrier A	Supplier 2	Mumbai	Pending	Sea	Route A	737	644.6250	Demand	SKU88
4	2.956572	29	215	29	...	Carrier B	Supplier 3	Mumbai	Pending	Road	Route B	802	415.0625	Demand	SKU0
9	1.532655	24	867	15	...	Carrier A	Supplier 3	Kolkata	Pending	Sea	Route A	963	650.6250	Demand	SKU37
9	4.858271	1	251	23	...	Carrier C	Supplier 5	Chennai	Fail	Sea	Route A	616	443.7500	Demand	SKU33
6	6.037884	19	791	4	...	Carrier B	Supplier 5	Chennai	Pending	Air	Route C	601	805.4375	Demand	SKU21
1	6.496325	5	228	12	...	Carrier B	Supplier 3	Bangalore	Pending	Air	Route C	198	371.3125	Demand	SKU79
3	4.742636	20	694	16	...	Carrier A	Supplier 4	Chennai	Fail	Road	Route C	187	630.6250	Demand	SKU19
6	1.729569	24	937	18	...	Carrier C	Supplier 5	Kolkata	Fail	Rail	Route A	83	688.5625	Demand	SKU3
4	1.325274	3	563	3	...	Carrier A	Supplier 4	Kolkata	Fail	Road	Route B	336	454.5625	Demand	SKU12
7	3.404734	13	769	8	...	Carrier C	Supplier 4	Mumbai	Pending	Sea	Route B	150	417.7500	Demand	SKU8
3	5.352878	28	648	28	...	Carrier C	Supplier 3	Mumbai	Pending	Road	Route A	324	387.5000	Demand	SKU96
4	6.780947	26	334	5	...	Carrier B	Supplier 1	Kolkata	Pass	Road	Route B	359	501.9375	Demand	SKU77

1.34 Forecasting Plotting

Python Code [27]:

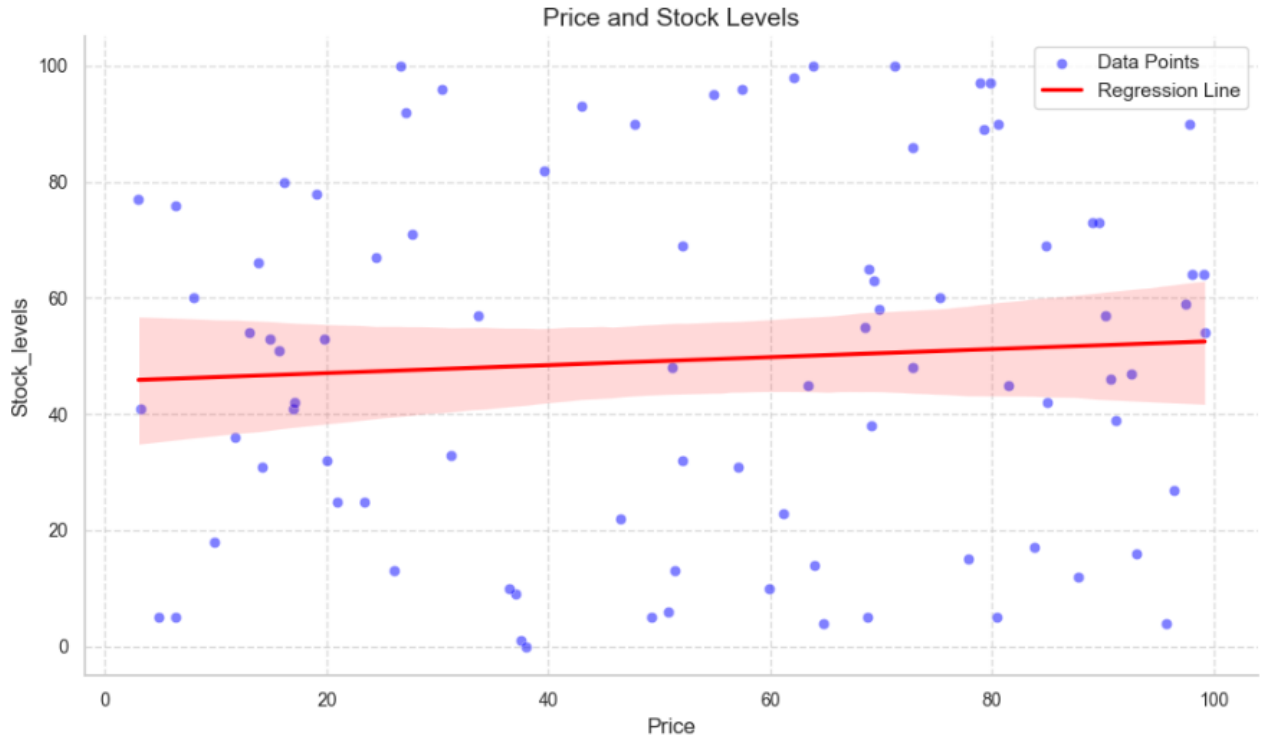
```
import matplotlib.pyplot as plt
import seaborn as sns

def scatter_plot_with_regression(df, x_column, y_column, title):
    # Set up the plot style using Seaborn
    sns.set(style='whitegrid')

    # Create the scatter plot with Seaborn
    plt.figure(figsize=(10, 6))
    sns.scatterplot(x=x_column, y=y_column, data=df, alpha=0.5, color='blue', label='
Data Points')

    # Add a regression line to show the trend
    sns.regplot(x=x_column, y=y_column, data=df, scatter=False, color='red', label='R
egression Line')
```

```
    # Set the labels and title
    plt.xlabel(x_column, fontsize=12)
    plt.ylabel(y_column, fontsize=12)
    plt.title(title, fontsize=14)

    # Customize gridlines and tick parameters
    plt.grid(True, linestyle='--', alpha=0.7)
    plt.tick_params(axis='both', which='major', labelsize=10)

    # Add a legend
    plt.legend()

    # Remove spines to enhance clarity
    plt.gca().spines['right'].set_visible(False)
    plt.gca().spines['top'].set_visible(False)

    # Show the plot
    plt.tight_layout()
    plt.show()
```

Python Code [28]:

```
scatter_plot_with_regression(df, 'Price', 'Stock_levels', 'Price and Stock Levels')
```

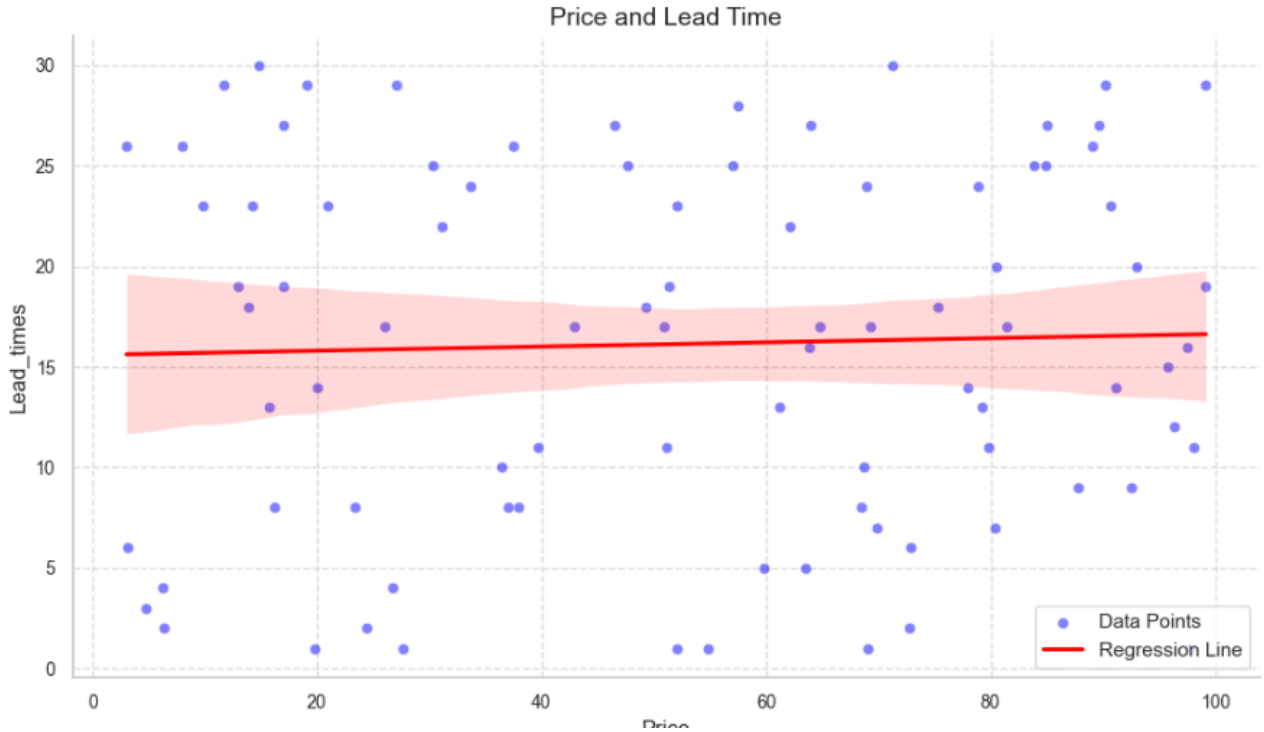
Output [28]:



Python Code [29]:

```
scatter_plot_with_regression(df, 'Price', 'Lead_times', 'Price and Lead Time')
```

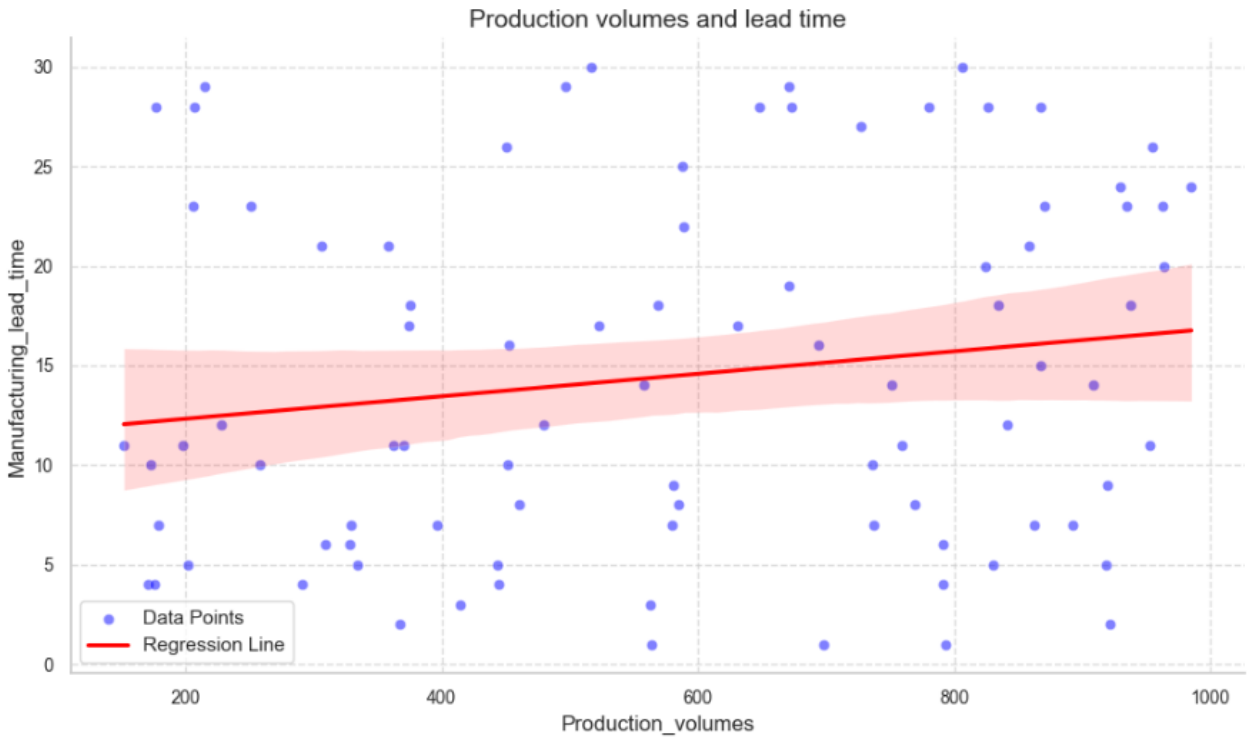
Output [29]:



Python Code [30]:

```
scatter_plot_with_regression(df, 'Production_volumes', 'Manufacturing_lead_time', 'P
roduction volumes and lead time')
```

Output [30]:

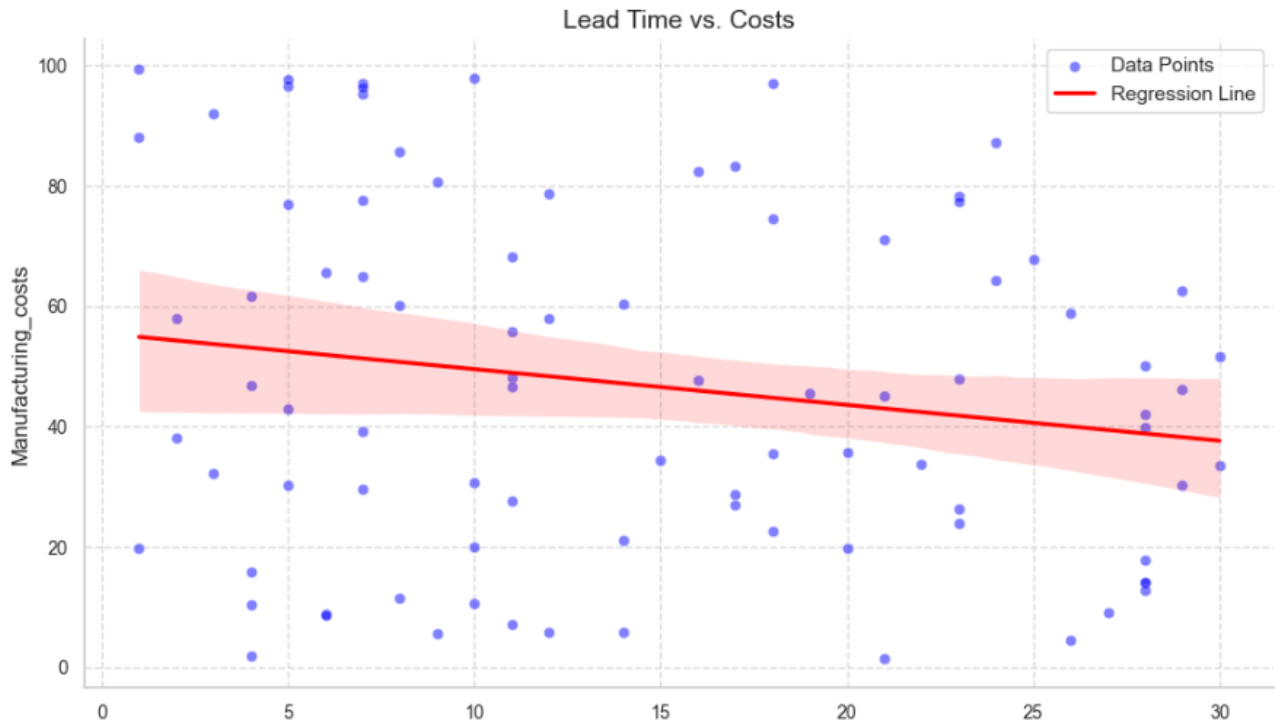


There is a minor positive correlation between manufacturing lead times and production volume. As production volume rises, there is a slight increase in the lead time required to manufacture beauty products.

Python Code [31]:

```
scatter_plot_with_regression(df, 'Manufacturing_lead_time', 'Manufacturing_costs', 'L
ead Time vs. Costs')
```

Output [31]:

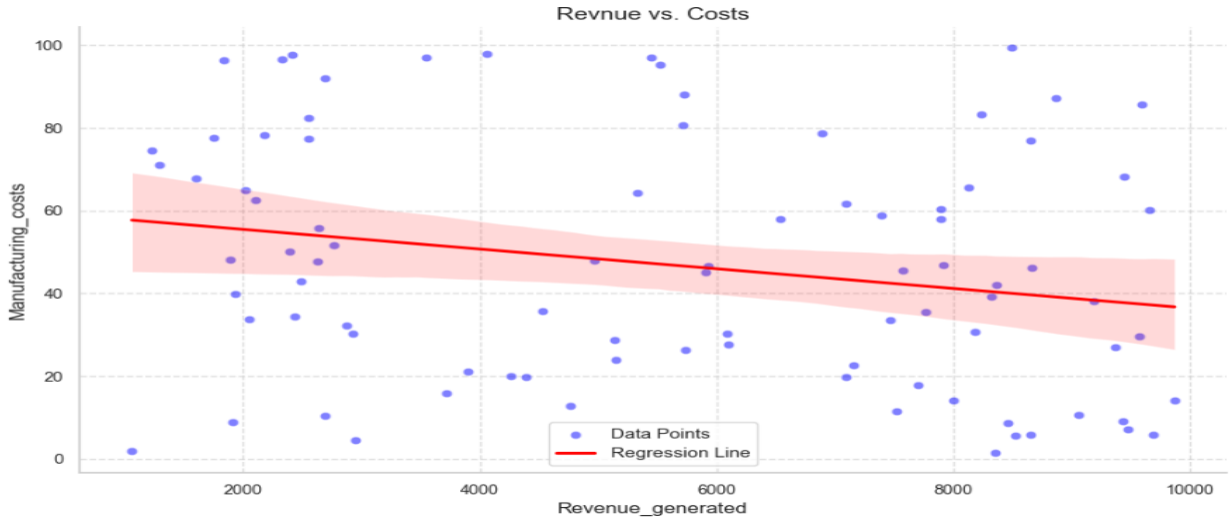


There is a minor positive correlation with manufacturing costs and lead times. Manufacturing lead time increases as the cost decreases.

Python Code [32]:

```
scatter_plot_with_regression(df, 'Revenue_generated', 'Manufacturing_costs', 'Revn
ue vs. Costs')
```

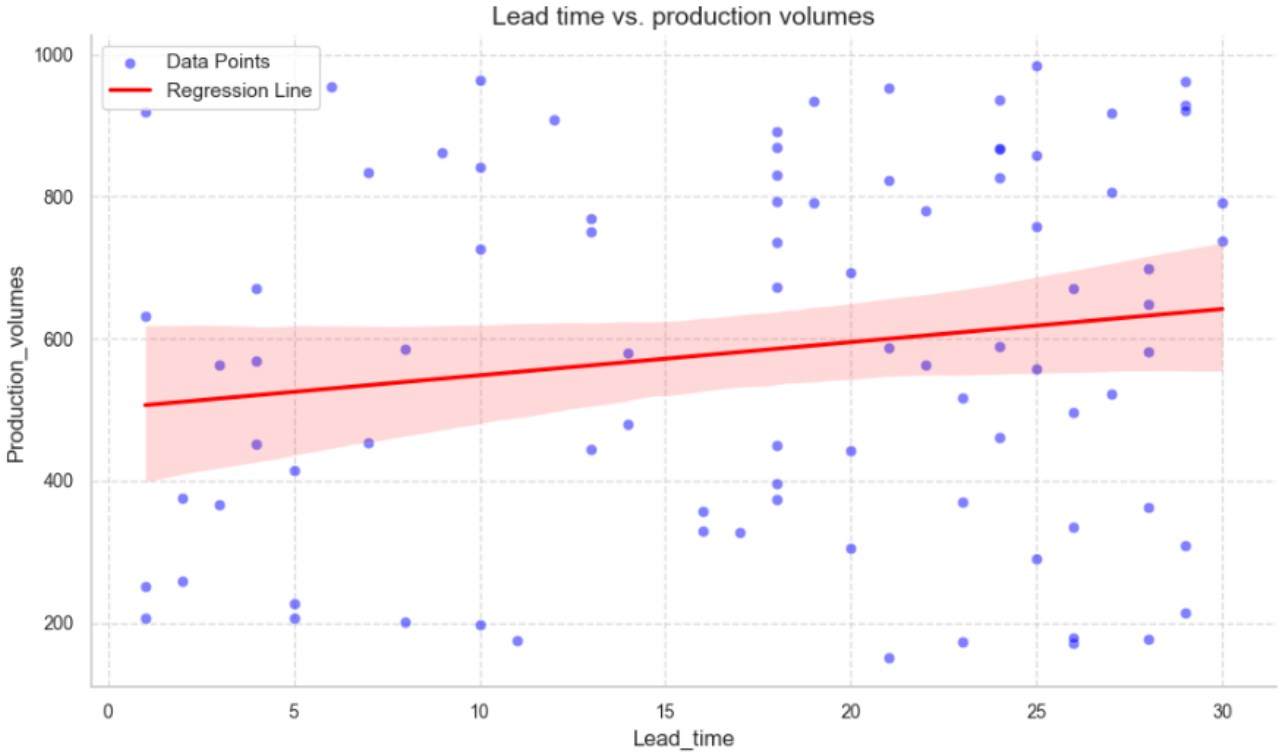
Output [32]:



Python Code [33]:

```
scatter_plot_with_regression(df, 'Lead_time', 'Production_volumes', 'Lead time vs. pr
oduction volumes')
```

Output [33]:

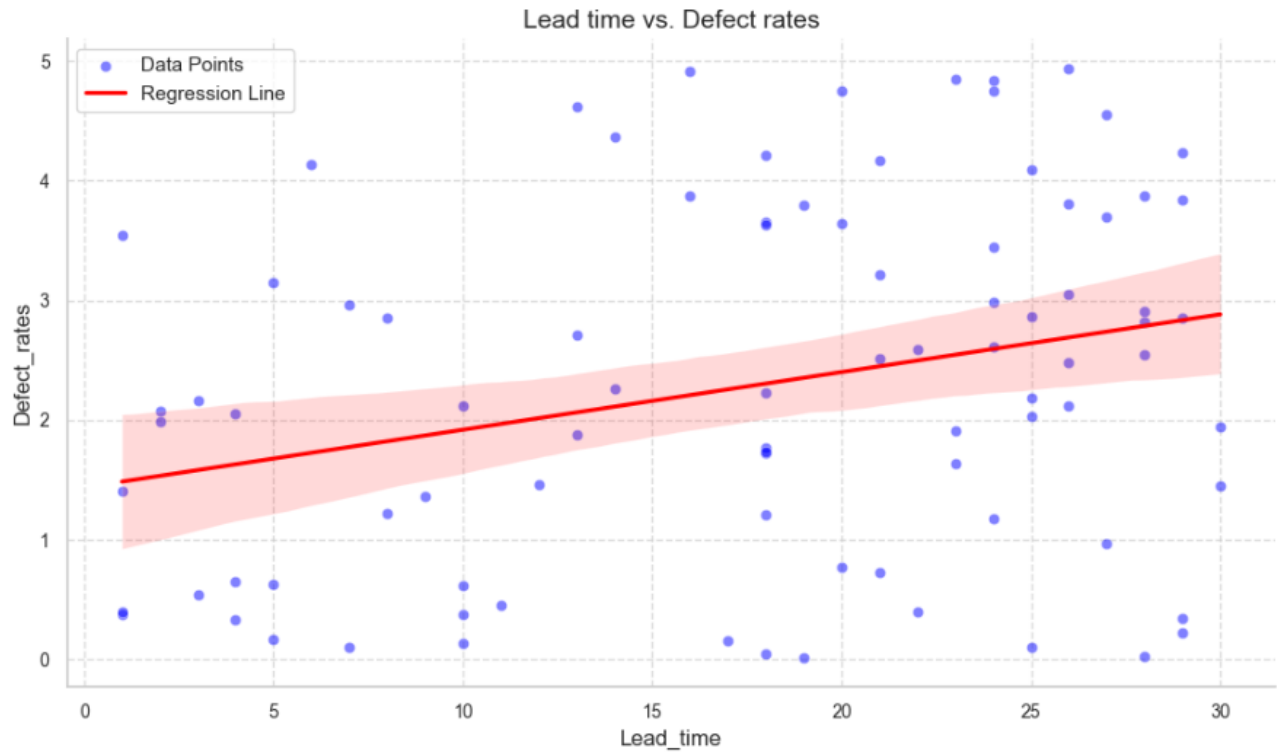


Lead times increase as production volumes increase.

Python Code [34]:

```
scatter_plot_with_regression(df, 'Lead_time', 'Defect_rates', 'Lead time vs. Defect rat
es')
```

Output [34]:

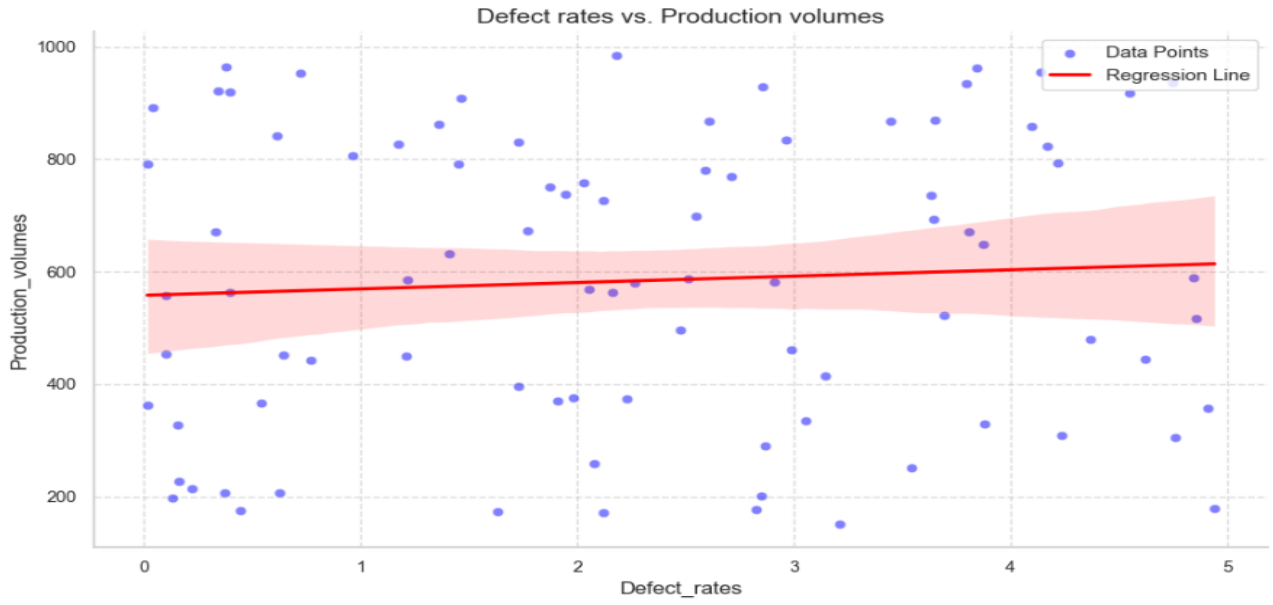


As lead times increase so do the defect rates.

Python Code [35]:

```
scatter_plot_with_regression(df, 'Defect_rates', 'Production_volumes', 'Defect rates vs. Production volumes')
```

Output [35]:

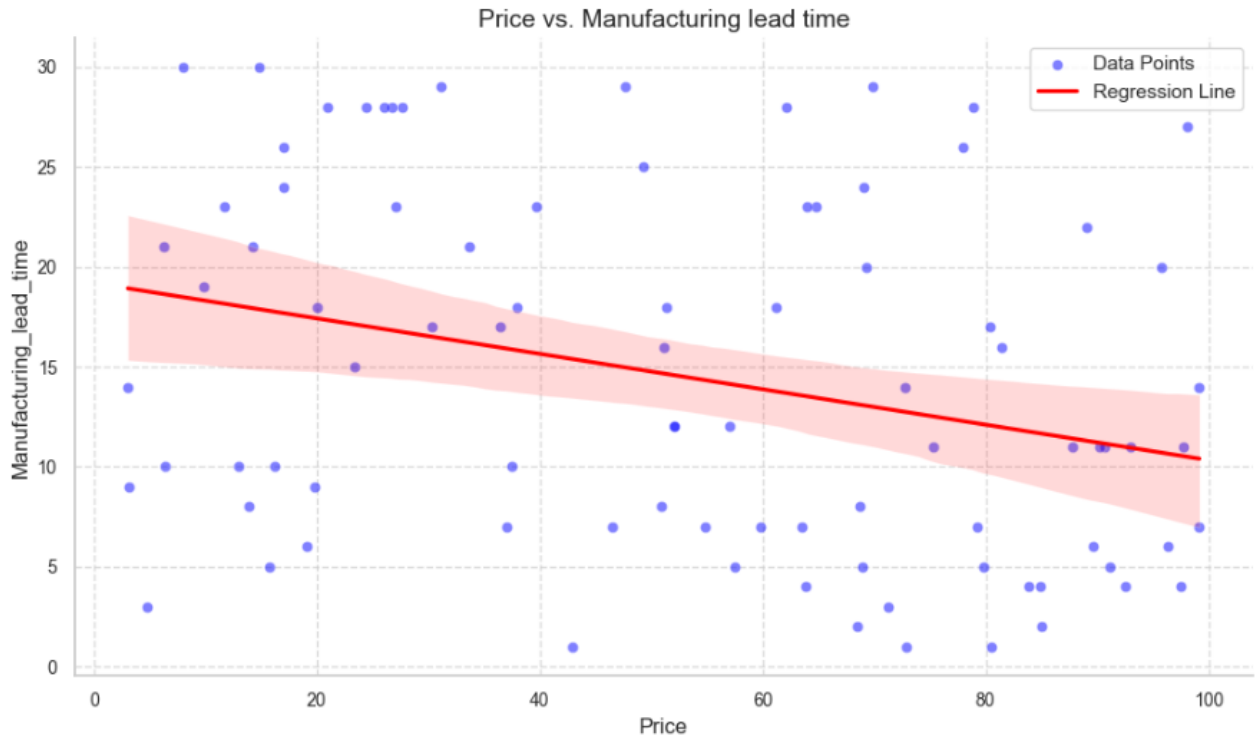


Defect rates increase as production volumes rise. Simply put the more products that are made the more chance for defects.

Python Code [36]:

```
scatter_plot_with_regression(df, 'Price', 'Manufacturing_lead_time', 'Price vs. Manufacturing lead time')
```

Output [36]:

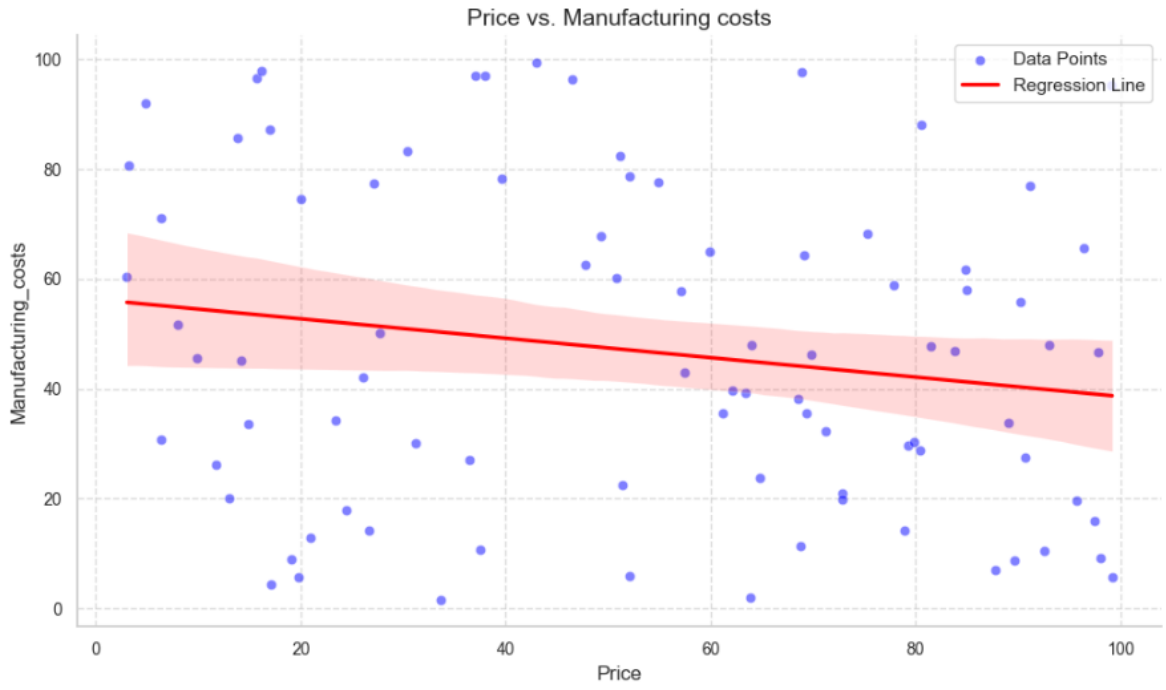


Prices tend to increase as manufacturing lead times decrease. This makes sense as more supply can leads to lower prices.

Python Code [37]:

```
scatter_plot_with_regression(df, 'Price', 'Manufacturing_costs', 'Price vs. Manufacturing costs')
```

Output [37]:



Python Code [38]:

```
# Create an interaction term between 'Price' and 'Availability'
df['Price_Availability_interaction'] = df['Price'] * df['Availability']

# Calculate a simple moving average for 'Number of products sold' over 7 days
df['Moving_Average_7_days'] = df['Number_of_products_sold'].rolling(window=7).mean()

# Drop rows with null values
df.dropna(inplace=True)

df.head()
```

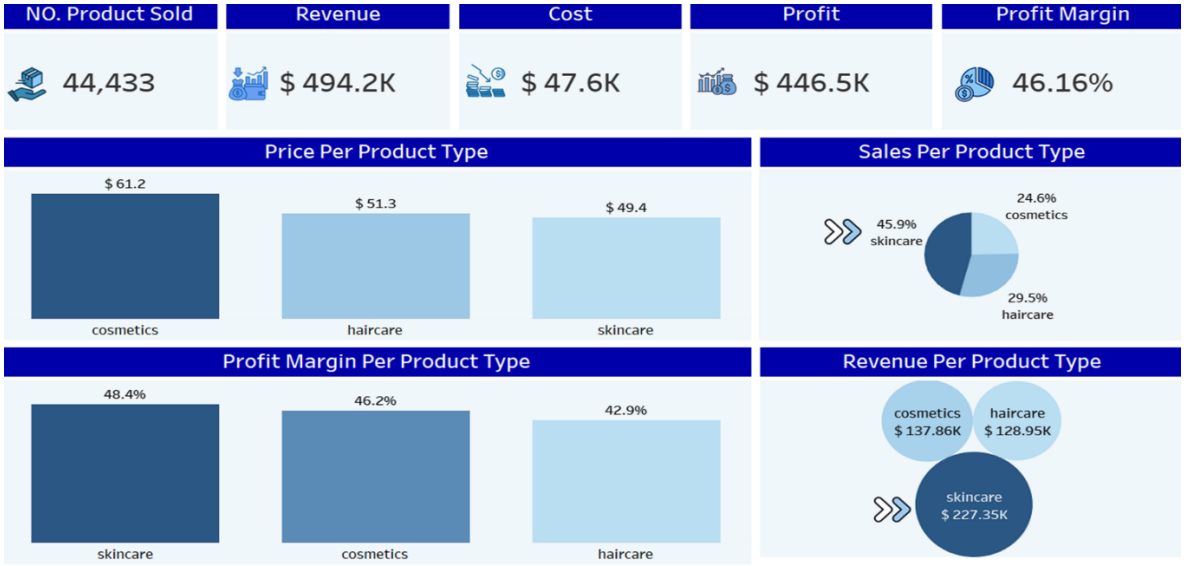
Output [38]:

	Product_Type	SKU	Price	Availability	Number_of_products_sold	Revenue_generated	Customer_demographics	Stock_levels	Lead_times	Order_quantities
6	skincare	SKU14	99.171329	26	562	8653.570926	Non-binary	54	29	78
7	skincare	SKU15	36.989245	94	469	5442.086785	Non-binary	9	8	69
8	cosmetics	SKU17	81.462534	82	126	2629.396435	Female	45	17	85
9	haircare	SKU18	36.443628	23	620	9364.673505	Unknown	10	10	46
10	skincare	SKU19	51.123870	100	187	2553.495585	Unknown	48	11	94

5 rows × 35 columns

CHAPTER 4: VISUALIZATION AND FORECASTING

Dashboard 1:



This dashboard provides a clear overview of key performance metrics across different product types (cosmetics, haircare, and skincare). Here's a detailed breakdown of the insights:

1. Overall Metrics (Top Row)

- Number of Products Sold: 44,433 units were sold across all product types.

- Total Revenue: The total revenue generated amounts to $494.2K. This suggests a strong overall performance.

- Total Costs: The cost to achieve this revenue was $47.6K, indicating relatively low operational or manufacturing costs.

- Total Profit: A substantial profit of $446.5K was achieved, representing strong financial performance.

- Profit Margin: A profit margin of 46.16% indicates that the business retains nearly half of its revenue as profit after deducting costs. This is a healthy margin, particularly in industries like retail and personal care.

2. Price Per Product Type

- Cosmetics have the highest average price at $61.2 per unit. This could indicate a premium pricing strategy or a more expensive product line.

- Haircare is priced at $51.3 per unit on average, positioning it in the mid-tier range.

- Skincare is priced the lowest at $49.4 per unit, possibly reflecting higher sales volumes or a more accessible pricing strategy.

Observation: The pricing differences across product types likely reflect the product strategy, where cosmetics may be positioned as premium, while skincare is priced more competitively.

3. Sales Per Product Type (Pie Chart)

- Cosmetics dominate the sales, accounting for 45.9% of the total product sales.

- Haircare comes in second with 29.5% of the sales.

- Skincare makes up 24.6% of the total sales.

Observation: Despite having the lowest sales share, skincare is contributing significantly to total revenue, likely due to the higher volume and potentially lower cost of goods sold.

4. Profit Margin Per Product Type

- Skincare leads with a 48.4% profit margin, making it the most profitable category despite being the least expensive.

- Cosmetics follow closely with a 46.2% profit margin, suggesting that the high price point is contributing to substantial profitability.

- Haircare has the lowest profit margin at 42.9%, but it is still within a healthy range.

Observation: Skincare products not only contribute significantly to revenue but also yield the highest profit margins, indicating that this could be a focus area for boosting profits.

5. Revenue Per Product Type (Bubble Chart)

- Skincare generates the most revenue, at $227.35K, despite having a lower sales percentage. This indicates that skincare has a high turnover or that there are other factors driving its revenue generation, such as strong demand or low production costs.
- Cosmetics follow with $137.86K in revenue, leveraging a high price point and moderate sales share.
- Haircare brings in $128.95K in revenue, which is a solid contribution but slightly lower than cosmetics and skincare.

Observation: While skincare contributes the most to revenue, it is interesting to see that cosmetics, despite their lower revenue share, still perform well due to their high price point.

Key Insights:

- Skincare is the strongest performer in terms of both revenue generation and profit margin. Despite lower average prices, the volume of sales and profitability makes it a crucial category for the business.
- Cosmetics are priced highest but generate slightly less revenue than skincare. However, the high-profit margin makes this product category a valuable contributor.
- Haircare has the lowest profit margin and revenue, suggesting there might be opportunities to either improve efficiency in this category or shift focus toward more profitable product types like skincare.

Recommendations:

- Double down on Skincare: Since it has the highest revenue and profit margin, focusing on expanding this product line, improving inventory management, and promoting skincare products could further boost profitability.
- Review Haircare Strategy: Since haircare has the lowest profit margin, it may be worth revisiting pricing strategies or reducing costs in this category to increase profitability.
- Maintain Cosmetics as a Premium Product: The high price point and strong profit margin of cosmetics suggest that this category can continue to be marketed as a premium product. Consider expanding premium offerings in this category to capture more high-value customers.

By aligning sales, revenue, and profitability strategies according to the performance insights in this dashboard, the business can optimize its product portfolio for maximum profitability and growth.

Dashboard 2:

This dashboard provides insights into the relationships between various supply chain factors such as availability, sales, pricing, stock levels, and lead times. Here's a detailed breakdown:

1. Availability (48.47%)

- Insight: The overall product availability rate is 48.47%, which suggests that the business faces stockout issues for more than half of its product assortment. This availability percentage is relatively low, indicating that nearly half of the products may not be available at any given time.

- Impact: Low availability could lead to lost sales opportunities and negatively affect customer satisfaction, as customers may be unable to purchase products that are frequently out of stock.

2. Availability & Sales Relationship (Scatter Plot)

- Insight: The scatter plot shows the relationship between availability and sales. While there are some points where higher availability correlates with higher sales, the relationship seems somewhat

inconsistent, as there are instances where products with lower availability have higher sales.

- Impact: While availability is important, the data shows that other factors (such as demand for specific products) could be influencing sales more significantly. This suggests that while improving availability is important, demand forecasting and product prioritization are also crucial.

3. Sales Per Price (Line Graph)

- Insight: This graph shows how sales volume fluctuates across different price points. Sales tend to vary significantly at lower price points (below $50), but become more stable between $50-$100.

- Impact: The data suggests that pricing is a significant driver of sales, with lower-priced products experiencing higher variability in sales. The steadying of sales at higher price points could indicate that premium-priced products face less competition or cater to a more loyal customer base.

4. Lead Time to Customer (16.15 Days)

- Insight: The average lead time to deliver products to customers is 16.15 days, which may be longer than the industry average depending on the product category. Longer lead times can result in customer dissatisfaction and impact repeat purchase rates.

- Impact: High lead times can be problematic, particularly for customers who expect fast delivery. Reducing lead times would be crucial to improving customer satisfaction and potentially increasing sales, especially for fast-moving products.

5. Lead Times & Sales Relationship (Line Graph)

- Insight: This graph shows the relationship between lead times and sales. The lead times seem to fluctuate significantly, with no clear correlation between shorter lead times and higher sales.

- Impact: This could suggest that lead times do not have a direct, linear impact on sales, but they likely affect customer satisfaction. While a shorter lead time is generally preferred, customers may still place orders even when delivery times are longer, possibly due to product uniqueness or brand loyalty.

6. Stock Levels (49.35%)

- Insight: The overall stock level is 49.35%, indicating that roughly half of the products are fully stocked. This relatively low percentage could be contributing to the lower availability rate, leading to potential stockouts.

- Impact: Having only half of the products fully stocked can lead to missed sales and inefficient inventory management. This further emphasizes the need for better inventory forecasting and management to maintain optimal stock levels for high-demand products.

7. Stock Levels & Sales Relationship (Scatter Plot)

- Insight: This scatter plot demonstrates the relationship between stock levels and sales. Similar to the availability graph, there are points where low stock levels correspond with higher sales and vice versa. However, products with higher stock levels tend to generate more consistent sales.

- Impact: Products with low stock levels are at risk of stockouts, potentially resulting in lost sales. This suggests that maintaining higher stock levels for high-demand products could stabilize sales and prevent stockouts.

Key Insights:

- Low Availability and Stock Levels: The availability and stock levels being below 50% indicate potential inefficiencies in inventory management. This can lead to lost sales and customer dissatisfaction.
- Inconsistent Relationship Between Availability and Sales: The relationship between availability and sales shows that, while availability is important, certain products may sell well even with limited stock. This indicates the need for demand forecasting to prioritize stocking high-demand items.
- Lead Time Needs Improvement: The average lead time of 16.15 days is relatively long. Reducing lead times would likely improve customer satisfaction and help drive more repeat business.
- Sales Distribution by Price: Sales fluctuate more at lower price points, with more consistent sales at higher price points. This could indicate that the business might benefit from focusing on premium products with steadier demand.

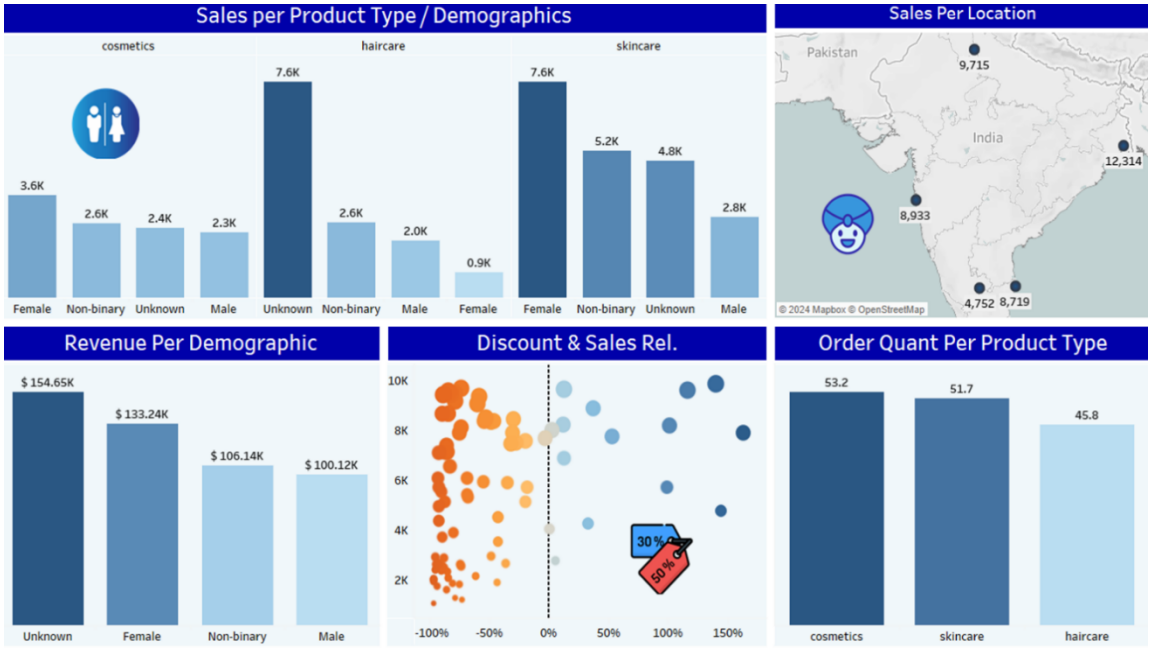
Recommendations:

1. Optimize Inventory Management: Focus on improving stock availability by implementing demand forecasting models and better inventory management practices. Stock levels should be maintained for high-demand items to avoid stockouts and lost sales opportunities.
2. Reduce Lead Times: Explore ways to reduce lead times by negotiating with suppliers, optimizing logistics, or working with local warehouses. This will help in improving the overall customer experience and potentially increase sales.
3. Focus on Pricing Strategy: Since higher-priced products have more stable sales, consider marketing these premium products more aggressively. Pricing strategies for lower-priced products should also be reviewed to maximize profitability while maintaining volume.

4. Improve Stocking of High-Sales Products: The scatter plot shows that certain products generate sales even when stock levels are low. Prioritizing these high-sales products for restocking can help prevent stockouts and maintain steady sales.

By acting on these insights, the business can enhance its overall supply chain performance, reduce inefficiencies, and drive better customer satisfaction.

Dashboard 3:



This dashboard provides insights into sales performance across various demographic groups, product types, and locations, while also exploring the impact of discounts on sales. Below is a detailed breakdown of the dashboard:

1. Sales per Product Type / Demographics:

Cosmetics: Highest sales from Unknown demographic (7.6K).

- Females and non-binary customers are relatively high, with 3.6K and 2.6K sales respectively.

- Males account for the lowest sales, with 2.3K sales.

Haircare: Females drive the highest sales, with 7.6K, followed by non-binary customers (2.6K).

- Male sales are extremely low, at only 0.9K, showing a significant gap compared to females.

Skincare: Highest sales again from females (5.2K), followed by non-binary (4.8K).

- Males once again show the lowest sales, with only 2.8K.

Insights: Across all product types, female customers and those categorized as non-binary dominate sales. The "Unknown" demographic also contributes significantly, especially in the cosmetics category, but male customers consistently show the lowest engagement with these products.

Recommendation: Focus marketing and product strategies on female and non-binary customers, as they represent the strongest sales demographics. Consider targeted campaigns to engage male customers, especially in categories where their sales are low, such as skincare and haircare.

2. Sales per Location:

- Highest sales in the southern part of India (12,314 units).
- The northern region shows 9,715 sales, while other regions, like the east and west, see lower figures at 8,933 and 8,719 units respectively.
- The central region has the lowest performance, with 4,752 units.

Insights: Sales performance is stronger in the southern and northern regions, indicating a higher product demand in these areas. Meanwhile, central India shows the weakest sales figures.

Recommendation: Consider increasing marketing efforts and product availability in the central region to boost sales. Analyzing why this region lags behind could reveal opportunities for growth (e.g., warehouse locations, supply issues, or cultural preferences). The strong

performance in the south indicates this could be an ideal region for expanding product lines or promotions.

3. Revenue per Demographic:

- The Unknown demographic brings the highest revenue ($154.65K).
- Females generate the second-highest revenue at $133.24K.
- Non-binary customers contribute $106.14K, while Males contribute the least at $100.12K.

Insights: There is a strong correlation between sales and revenue for the unknown and female demographics, while the male demographic has the lowest revenue generation, consistent with their lower sales figures.

Recommendation: Since Unknown and female demographics contribute the most revenue, targeting these groups with new product launches, loyalty programs, and personalized marketing could drive further revenue growth. Additionally, more effort should be placed on improving male customer engagement to increase their contribution to total revenue.

4. Discount & Sales Relationship:

- The scatter plot shows how discounts affect sales. There's a notable cluster of sales around the 30%-50% discount range, indicating that discounts in this range seem to positively impact sales.
- However, the plot also shows that some products generate sales with no discount (0%), meaning that discounts are not always necessary to boost sales.

Insights: While discounts can drive sales, the data suggests that sales can occur without heavy discounting, particularly for certain products.

Offering moderate discounts (around 30-50%) seems to be the sweet spot for generating more sales.

Recommendation: Apply strategic discounting in the 30-50% range for underperforming products to drive volume. Avoid excessive discounts, as the data shows that steep discounts do not guarantee higher sales and can hurt profit margins. Test the impact of no-discount sales strategies for high-demand products.

5. Order Quantity Per Product Type:

- Cosmetics has the highest average order quantity (53.2 units), followed closely by skincare at 51.7 units.
- Haircare has the lowest average order quantity at 45.8 units.

Insights: Customers tend to purchase larger quantities of cosmetics and skincare products, with haircare lagging slightly behind. This could be due to the nature of the products or demand trends in the market.

Recommendation: Focus on optimizing cosmetics and skincare product availability to meet the higher order demand. For haircare, consider exploring product bundling or promotion strategies to increase the order quantity and encourage more purchases.

Overall Summary:

This dashboard provides a comprehensive view of how different customer demographics, locations, and product types interact with sales and revenue. The key takeaways include:

- Female and non-binary customers drive most of the sales across product categories, while males underperform.

- The southern and northern regions of India show the highest sales, indicating strong market potential in these areas.
- Moderate discounting (30-50%) is effective at increasing sales without heavily impacting profit margins.
- There is potential for growth in male customer engagement and regions with lower sales.

Key Recommendations:

1. Targeted Marketing for Female and Non-Binary Demographics: These groups are the strongest performers, and tailored campaigns could further boost sales.
2. Geographic Expansion: Invest in increasing product presence in central India and optimizing distribution for underperforming areas.
3. Strategic Discounting: Use discounts wisely in the 30-50% range to drive sales without sacrificing profit margins.
4. Increase Male Engagement: Focus on campaigns and product offerings that appeal more to male customers to close the sales gap.

By implementing these strategies, the business can further enhance its sales performance and profitability.

Dashboard 4:

This dashboard presents an analysis of three key metrics: defect rate, manufacturing lead times, and the relationship between profit and manufacturing costs. Here's a detailed breakdown of each section:

1. Defect Rate & Sales Relation (Left Section)

Defect Rate: The overall defect rate is 2.284%, which is relatively low, indicating decent product quality.

Defect Rate & Sales Relation: The chart shows the relationship between defect rates and sales volumes. There is no clear correlation between the two variables, meaning that sales do not appear to be heavily impacted by minor defects. While the trend line is slightly downward, indicating that higher defect rates might slightly decrease sales, the overall effect appears minimal.

2. Availability & Manufacturing Lead Times Relationship (Top-Right Section)

Availability & Manufacturing Lead Times: The chart shows that as manufacturing lead times increase, availability fluctuates. There's a general upward trend suggesting that higher lead times lead to better availability. However, lead times extending beyond a certain point may cause inefficiencies, as seen with the drops in availability at specific intervals. This could indicate that overextended lead times create stock management issues.

3. Profit & Manufacturing Cost Relationship (Bottom-Right Section)

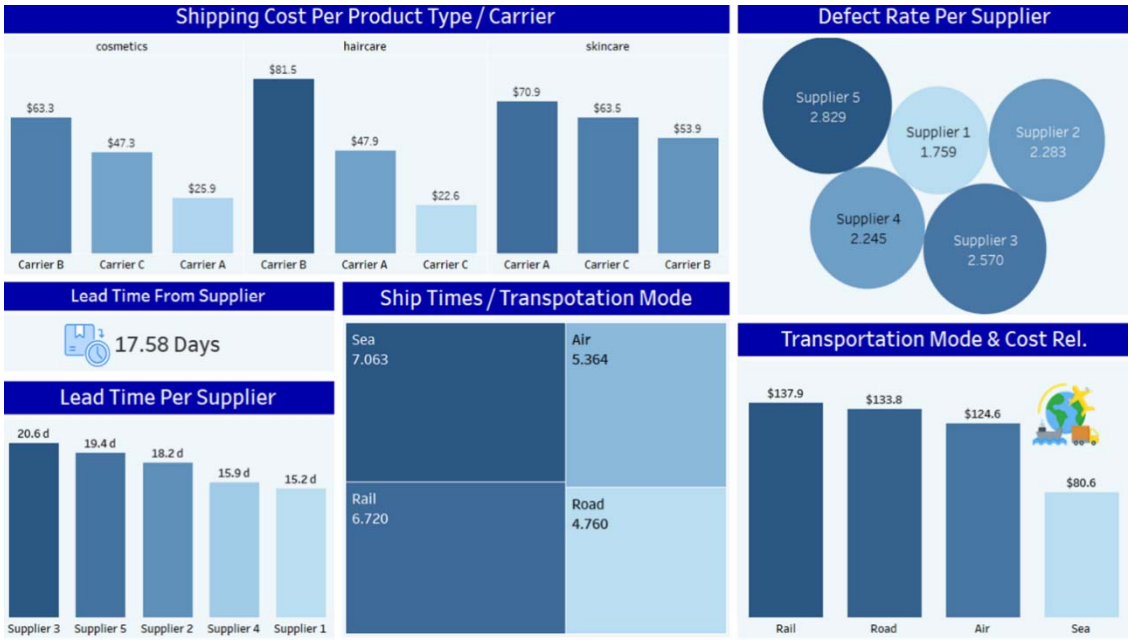
Profit & Manufacturing Costs: This chart illustrates how profit correlates with manufacturing costs across different price points. A clear spike can be observed around the $50-$60 price range, indicating this is where the highest profit margins are achieved. However, as manufacturing costs increase beyond that range, profits tend to decrease significantly. This suggests that products priced too high may not justify the increased production costs, leading to diminishing returns.

Key Takeaways:

1. Defect Management: With a relatively low defect rate, improving this metric may not drastically impact sales. However, further reduction of defects could enhance brand reputation and customer satisfaction.
2. Balancing Lead Times & Availability: While longer lead times can improve availability, a strategy should be developed to avoid overextended lead times, as they could disrupt supply chain efficiency.
3. Cost-Effective Pricing: Focusing on products in the $50-$60 manufacturing cost range seems to maximize profit. Care should be taken to avoid products with very high costs unless they deliver

significant value, as they may not contribute meaningfully to the bottom line.

Dashboard 5:



This dashboard provides valuable insights into optimizing the manufacturing process, reducing defects, and making strategic pricing decisions.

This dashboard offers insights into shipping costs, supplier performance, transportation modes, and their impact on the supply chain. Here's a detailed breakdown of the key sections:

1. Shipping Cost Per Product Type / Carrier (Top-Left Section)

Cosmetics: Carrier B has the highest shipping cost at $63.3, followed by Carrier A at $47.3, and Carrier C with the lowest at $25.9.

Haircare: Carrier B also has the highest cost at $81.5, with Carrier A ($47.9) and Carrier C ($22.6) showing lower shipping costs.

Skincare: Carrier B has a moderate cost at $63.5, while Carrier C is slightly cheaper at $53.9.

Key Insight: Carrier B tends to have higher shipping costs across all product types, indicating that while it may provide better service, cost reduction could be possible by optimizing the use of Carriers A and C.

2. Defect Rate Per Supplier (Top-Right Section)

Supplier 5 has the highest defect rate at 2.829%, followed by Supplier 3 at 2.570%.

Supplier 1 has the lowest defect rate at 1.759%, which is significantly lower than others.

Key Insight: Supplier 1 provides the best quality with the lowest defect rate, whereas Supplier 5's high defect rate could negatively impact product quality and customer satisfaction. Adjustments in supplier relationships or quality controls are needed for suppliers with high defect rates.

3. Lead Time From Supplier (Middle-Left Section)

- The average lead time from suppliers is 17.58 days, but there's variability among individual suppliers:
  - Supplier 3 has the longest lead time at 20.6 days.
  - Supplier 1 has the shortest lead time at 15.2 days.

Key Insight: While Supplier 1 has the lowest defect rate and shortest lead time, Supplier 3 and Supplier 5, despite having longer lead times and higher defect rates, may require tighter supply chain controls.

4. Shipping Times / Transportation Mode (Center Section)

- Sea shipping has the longest time at 7.063 days, followed by Rail at 6.72 days.
- Air has a shorter time at 5.364 days, and Road is the fastest at 4.76 days.

Key Insight: Faster transportation modes like Road and Air could be favored when speed is essential. Sea transportation, although slower, might be more suitable for bulk shipments where cost efficiency outweighs speed.

5. Transportation Mode & Cost Relationship (Bottom-Right Section)

- Rail has the highest cost at $137.9, followed by Road at $133.8, Air at $124.6, and Sea is the most cost-efficient at $80.6.

Key Insight: While Sea shipping is the cheapest, its longer time should be considered in supply chain planning. Air is relatively fast and has a moderate cost, making it suitable for high-value, time-sensitive goods. Rail and Road offer flexibility but come with higher costs.

Conclusion:

- Shipping Cost Optimization: The choice of carriers, particularly Carrier B, may need review to optimize shipping costs across product types.

- Supplier Performance: Supplier 1 stands out as the best performer in terms of quality and lead time, while Supplier 5 requires attention due to high defect rates.
- Transportation Planning: A balance between cost and shipping time should be maintained. Sea is ideal for low-cost, non-urgent shipments, while Air and Road offer faster options at slightly higher costs.

This dashboard can guide decisions around supplier management, transportation strategies, and cost optimization for greater efficiency in the supply chain.

Recommendations

1. Focus on Skincare Products

Rationale: Skincare products have shown strong sales performance or a growing trend in the data, indicating this is a product type with high demand and potential profitability.

Action:

- Increase inventory for skincare products to ensure availability. Perform demand forecasting using historical sales data of skincare products.
- Marketing Strategy: Launch targeted marketing campaigns specifically promoting skincare products across all channels.
- Supply Chain Strategy: Work with suppliers to prioritize the procurement of raw materials or products that are essential for skincare production.

2. Reorganization of Availability as Previously Proposed

Rationale: Availability of key products might have been inconsistent, leading to stockouts and loss of potential sales. Ensuring proper stock levels will help meet demand and increase sales.

Action:

- Implement dynamic inventory management and reorder points for products with fluctuating demand.
- Identify slow-moving inventory and optimize stock levels accordingly, freeing up capital for products with higher turnover.

3. Focus on Budget Products

Rationale: Budget products could have higher sales volume or be more attractive to a broader customer base. Focusing on these could lead to increased market share and customer satisfaction.

Action:

- Prioritize the production and promotion of budget-friendly products to cater to price-sensitive customers.
- Launch promotions and discounts that make these products even more attractive.
- Adjust pricing strategies to maximize margins without losing volume.

4. Reorganization of Stock Levels as Previously Proposed

Rationale: To avoid overstocking or understocking, stock levels must be aligned with demand forecasts. This reduces storage costs and prevents stockouts.

Action:

- Set up inventory thresholds based on product demand and sales velocity.

- Focus on products that regularly run out of stock and adjust reorder points accordingly.

5. Focus on Female Customers

Rationale: If the data reveals that female customers are a significant segment driving sales, efforts should be made to further tap into this demographic with tailored products and marketing campaigns.

Action:

- Conduct a customer segmentation analysis to focus on preferences and purchasing patterns of female customers.
- Customize your product offerings, promotions, and communication to meet the needs of this demographic.

6. Establish a New Warehouse at the Center of India

Rationale: A centrally located warehouse in India would reduce transportation costs and improve delivery times across key regions, especially in high-demand areas.

Action:

- Use the geographic data of sales and delivery times to identify key locations that would benefit from a centralized warehouse.
- Evaluate transportation and shipping costs to determine the impact of a new warehouse on operational efficiency.

By following these recommendations and continuously monitoring performance using SQL queries and dashboards in Tableau, the business can optimize its supply chain, improve product availability, and increase customer satisfaction and profitability.