Ministry of Communications and Information Technology

Digital Egypt Pioneers

Digital Egypt Pioneers

FINAL PROJECT

Data Analyst Specialist Track

I Learn - CAI1_DAT1_G12d

Amr Desouky

Youssef Yasser

Ahmed Rabie

Alaa Marouf

Digital Egypt Pioneers

FINAL PROJECT

Supply Chain

Under Supervision

Eng. Ahmed Samir

01

Week 1

Build Data Model, Data Cleaning and Preprocessing

- **Deliverables:**
  - Cleaned dataset ready for analysis.
  - Data preprocessing notebook.

02

Week 2

Analysis Questions Phase

- **Deliverables:**
  - Set of analysis questions that can be answered via the dataset

03

Week 3

Forecasting Questions Phase

- **Deliverables:**
  - Visualization plots answering forecasting questions

04

Week 4

Visualization Dashboard and Recommendation

- **Deliverables:**
  - Visualization dashboard.
  - Recommendation for the business

Digital Egypt Pioneers

Supply Chain Project Over View

Supply Chain Project

Content

1 **Build Data Model, Data Cleaning and Preprocessing**

2 Analysis Questions Phase

3 Forecasting Questions Phase

4 Visualization Dashboard and Recommendation

Digital Egypt Pioneers

1. Build Data Model, Data Cleaning and Preprocessing

1 . 1 Data Exploration

Data Overview

The provided dataset, likely named "supply_data," is a CSV file containing 100 rows and 24 columns. This indicates a relatively small but potentially informative dataset for analyzing supply chain-related information.

Data Types

The data within the columns is represented using a mix of data types : **Numeric**: used for quantities, costs, times, and rates, which is appropriate for numerical calculations .**Categorical**: used for categorical variables such as product types, shipping carriers, supplier names, inspection results, transportation modes, and routes, suggesting that these columns contain distinct categories or textual descriptions.

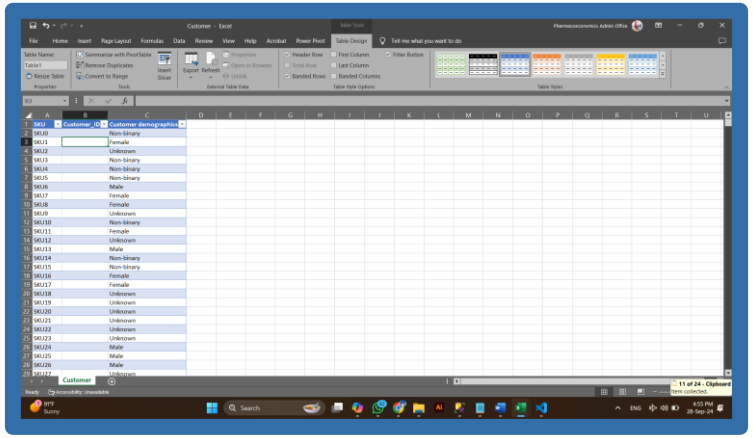
1. Build Data Model, Data Cleaning and Preprocessing

Digital Egypt Pioneers

1 . 2 Data Preparation

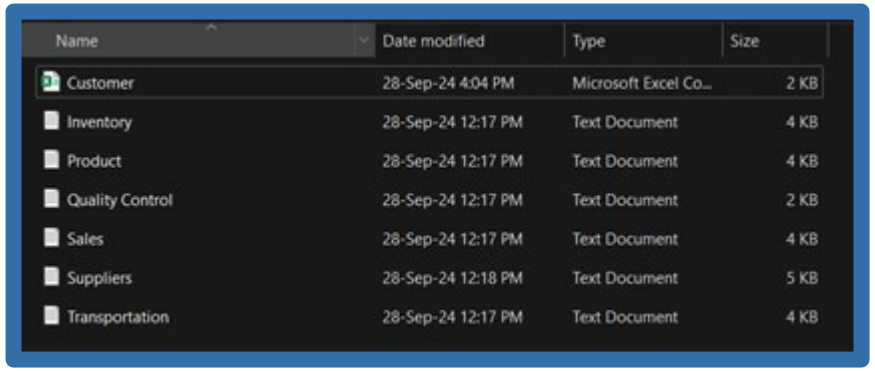
The initial data was extracted from an Excel file with 24 columns. This data was separated into 7 main CSV files, each representing a specific aspect of the supply chain

- Customers
- Inventory
- Products
- Quality Control
- Sales
- Suppliers
- Transportation



1 . 3 CSV File Creation

The separated data was saved as tab-delimited CSV files. Below is an example of how to create a Transportation table in SQL Server:

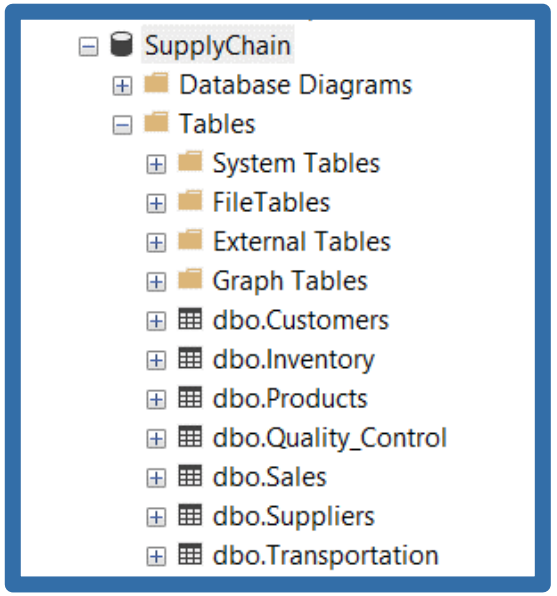


1. Build Data Model, Data Cleaning and Preprocessing

Digital Egypt Pioneers

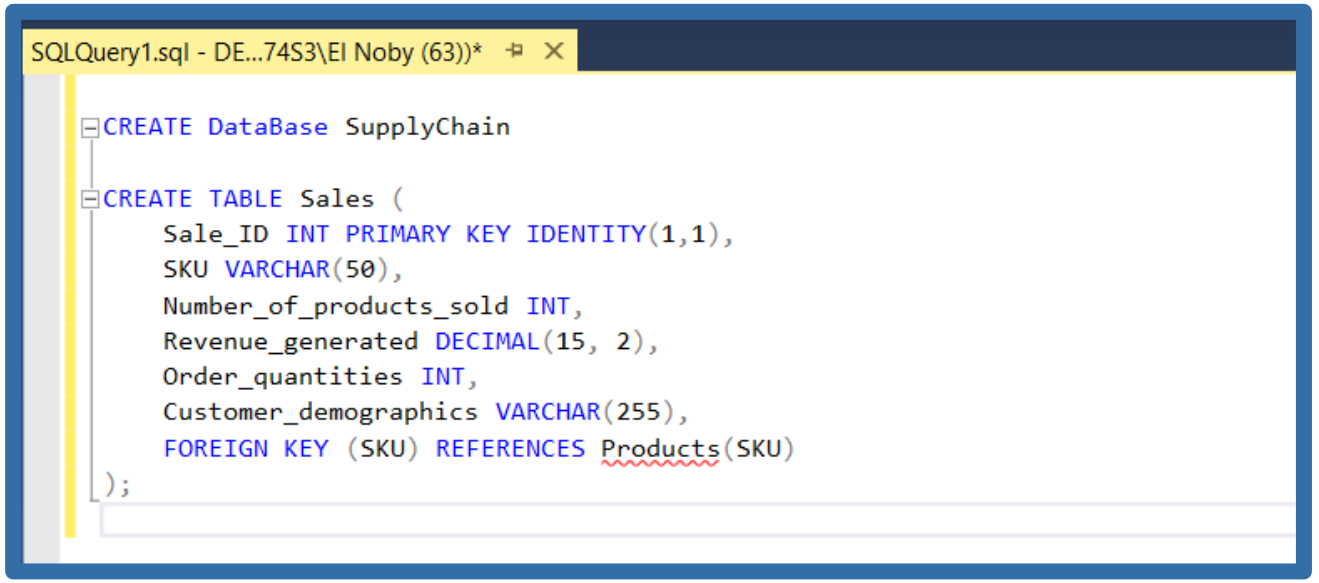
1 . 4 Creating the Database

A new database called Supply Chain was created in SQL Server.



1 . 5 Creating Tables

For each of the CSV files, corresponding tables were created in the SQL Server database. For example, the Products table was created as follows



1. Build Data Model, Data Cleaning and Preprocessing

Digital Egypt Pioneers

1 .6 Data Ingestion

Data was imported into each table using the BULK INSERT command. Below is an example for the Transportation table

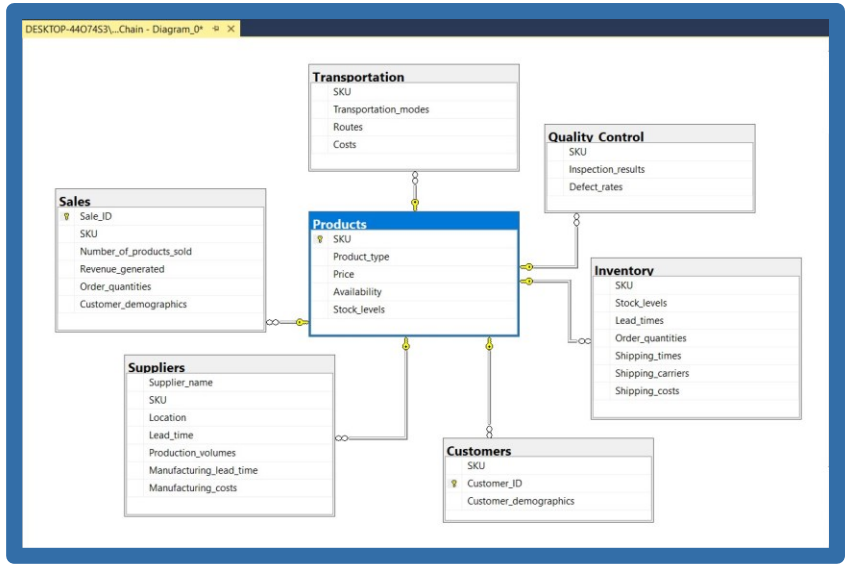
SQLQuery1.sql - DE...74S3\El Noby (63))*

```
BULK INSERT Transportation
FROM 'D:\Personal\learn\data analysis\DEPI\Final Project\SQL\Transportation.txt'
WITH (
    FIELDTERMINATOR = ' ',
    ROWTERMINATOR = '\n',
    FIRSTROW = 2
);

select * from Transportation
```

1 .7 Data Relationships

Once the tables were populated, relationships were established among them, leveraging foreign keys to ensure referential integrity

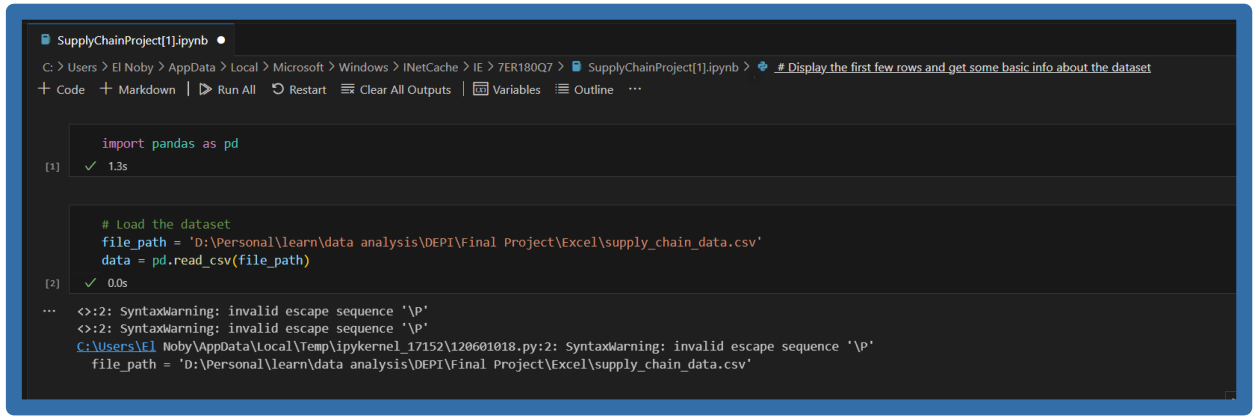


1. Build Data Model, Data Cleaning and Preprocessing

Digital Egypt Pioneers

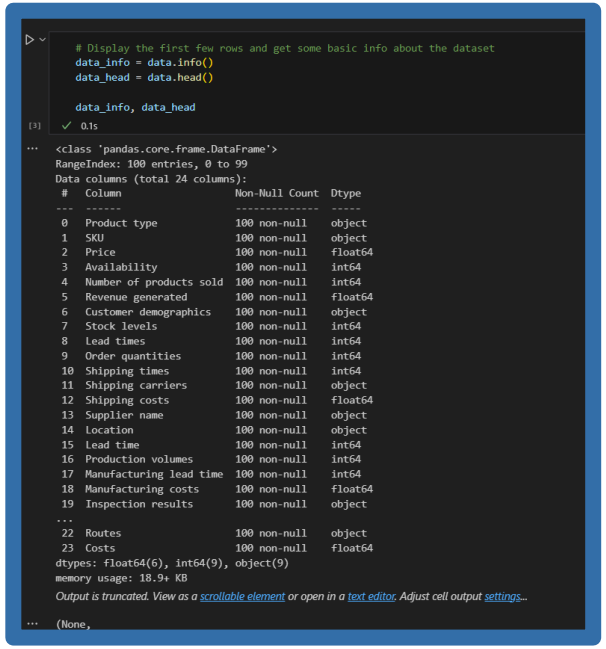
1 . 8 Data Cleaning and Preprocessing in Python

Loading Data into Python ,The CSV files were loaded into Python for data cleaning and preprocessing using the Pandas package:



1 . 9 . 1 Data Discovering

Data Discovering find the data types and columns heads:



1. Build Data Model, Data Cleaning and Preprocessing

1 . 9 . 2 Data Discovering

Data Discovering find Nulls and Duplicated Values:

```
# Step 6: Check for missing values and duplicates
missing_values = data.isnull().sum()
duplicates = data.duplicated().sum()
```

[4] ✓ 0.0s

```
missing_values
```

[6] ✓ 0.0s

```
Product type                0
SKU                         0
Price                       0
Availability                0
Number of products sold     0
Revenue generated           0
Customer demographics       0
Stock levels                0
Lead times                  0
Order quantities            0
Shipping times              0
Shipping carriers           0
Shipping costs              0
Supplier name               0
Location                    0
Lead time                   0
Production volumes          0
Manufacturing lead time     0
Manufacturing costs         0
Inspection results          0
Defect rates                0
Transportation modes        0
Routes                      0
Costs                       0
dtype: int64
```

1. Build Data Model, Data Cleaning and Preprocessing

1 . 10 Data Cleaning

Data cleaning involved stripping whitespace from string fields:

```
data.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
```

[8]

Python

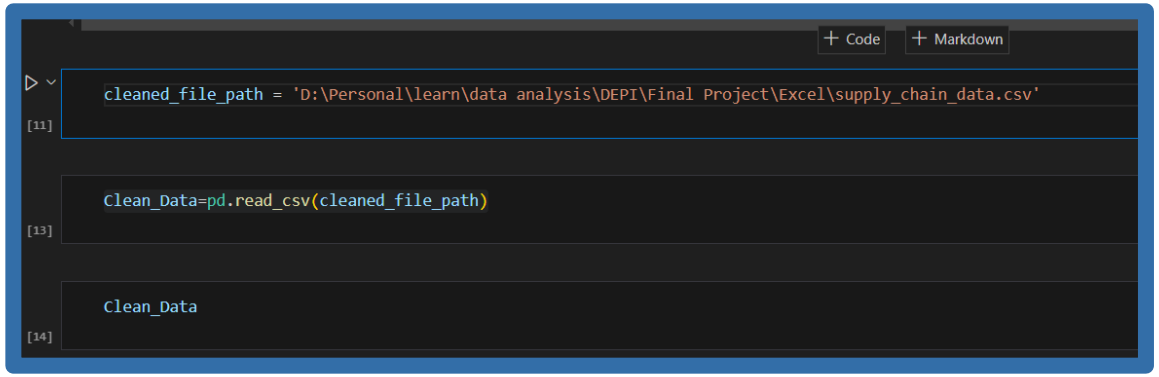
	Product type	SKU	Price	Availability	Number of products sold	Revenue generated	Customer demographics	Stock levels	Lead times	Order quantities	...	Location	Lead time	Production volumes	Manufacturing lead time	Manufacturing costs	Inspection results	Defect rates	Transportation modes	Routes	Costs
0	haircare	SKU0	69.808006	55	802	8661.996792	Non-binary	58	7	96	...	Mumbai	29	215	29	46.279879	Pending	0.226410	Road	Route B	187.752075
1	skincare	SKU1	14.843523	95	736	7460.900065	Female	53	30	37	...	Mumbai	23	517	30	33.616769	Pending	4.854068	Road	Route B	503.065579
2	haircare	SKU2	11.319683	34	8	9577.749626	Unknown	1	10	88	...	Mumbai	12	971	27	30.688019	Pending	4.580593	Air	Route C	141.920282
3	skincare	SKU3	61.163343	68	83	7766.836426	Non-binary	23	13	59	...	Kolkata	24	937	18	35.624741	Fail	4.746649	Rail	Route A	254.776159
4	skincare	SKU4	4.805496	26	871	2686.505152	Non-binary	5	3	56	...	Delhi	5	414	3	92.065161	Fail	3.145580	Air	Route A	923.440632
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	haircare	SKU95	77.903927	65	672	7386.363944	Unknown	15	14	26	...	Mumbai	18	450	26	58.890686	Pending	1.210882	Air	Route A	778.864241

1. Build Data Model, Data Cleaning and Preprocessing

1 . 11 Saving the Cleaned Data

Finally, the cleaned data was saved to a new

CSV file:



[* For Accessing the Python File]

1 . 12 Conclusion

the complete process from data extraction to cleaning and preprocessing. The structured approach ensures that data integrity is maintained, and the cleaned data is ready for analysis

Supply Chain Project

Content

Digital Egypt Pioneers

2. Analysis Questions Phase

2 . 1 Introduction

This section outlines the process of conducting a comprehensive data analysis to uncover valuable insights from the provided dataset. The analysis will focus on formulating and answering key questions that are relevant to the organization's decision-making process.

By leveraging SQL, we will explore the dataset to identify patterns, trends, and correlations that can inform strategic initiatives and enhance overall business performance.

We have sperate the Qeustion Regarding Each part:

2 . 2 Sales and Revenue 2 . 3 Customer Segmentation 2 . 4 Transportation 2 . 5 Supplier 2 . 6 Quality Control

2. Analysis Questions Phase

Digital Egypt Pioneers

2 . 2 Sales and Revenue

2 . 2 . 1 What is the Impact of Product Category on Sales Performance?

SQLQuery1.sql - DE...74S3\El Noby (62))*

```
SELECT Product_Type,
SUM(S.Revenue_generated) AS Total_Revenue ,
SUM(S.Number_of_products_sold) AS Total_Orders
FROM Sales S
INNER JOIN Products ON S.SKU = Products.SKU
GROUP BY Product_Type
Order By Total_Revenue DESC;
```

100 %

Results | Messages

	Product_Type	Total_Revenue	Total_Orders
1	skincare	241628.17	20731
2	haircare	174455.42	13611
3	cosmetics	161521.27	11757

2 . 2 . 2 How do lead times and shipping times affect sales performance?

SQLQuery1.sql - DE...74S3\El Noby (62))*

```
SELECT TOP 10 i.Lead_times, i.Shipping_times,
SUM(s.Revenue_generated) AS total_revenue
FROM Inventory i
JOIN Sales s ON i.sku = s.sku
GROUP BY i.Lead_times, i.Shipping_times
Order By total_revenue DESC;
```

100 %

Results | Messages

	Lead_times	Shipping_times	total_revenue
1	5	7	19198.42
2	18	7	19037.37
3	27	8	15405.41
4	17	1	13027.50
5	26	7	11828.05
6	24	8	10766.33
7	6	5	10094.36
8	4	1	9866.47
9	1	5	9692.32
10	17	4	9655.14

2. Analysis Questions Phase

Digital Egypt Pioneers

2 . 3 Customer Segmentation

2 . 3 . 1 What is Sales Trends Based on Customer Demographics?

SQLQuery1.sql - DE...74S3\El Noby (62))*

```
SELECT P.Product_type ,C.Customer_demographics,
SUM(S.Number_of_products_sold) AS Number_Of_Products
FROM Sales S
JOIN Products P ON S.SKU = P.SKU
JOIN Customers C ON C.SKU = P.SKU
GROUP BY P.Product_type, C.Customer_demographics
ORDER BY P.Product_type, Number_Of_Products;
```

100 %

Results | Messages

	Product_type	Customer_demographics	Number_Of_Products
1	cosmetics	Male	2304
2	cosmetics	Non-binary	2607
3	cosmetics	Unknown	2834
4	cosmetics	Female	4012
5	haircare	Female	936
6	haircare	Male	2292
7	haircare	Non-binary	2820
8	haircare	Unknown	7563
9	skincare	Male	2911
10	skincare	Unknown	4814
11	skincare	Non-binary	5153
12	skincare	Female	7853

2 . 3 . 2 How do customer demographics influence purchasing behavior?

SQLQuery1.sql - DE...74S3\El Noby (62))*

```
SELECT C.Customer_demographics,
SUM(S.Revenue_generated) AS Total_Revenue ,
SUM(S.Number_of_products_sold) AS Total_Order
FROM Sales S
JOIN Customers C ON C.SKU = S.SKU
GROUP BY C.Customer_demographics
ORDER BY Total_Revenue;
```

100 %

Results | Messages

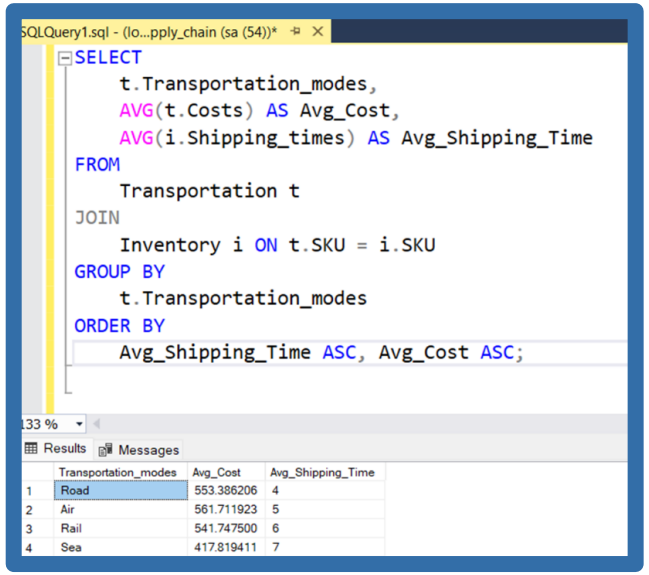
	Customer_demographics	Total_Revenue	Total_Order
1	Non-binary	116365.81	10580
2	Male	126634.42	7507
3	Female	161514.49	12801
4	Unknown	173090.14	15211

2. Analysis Questions Phase

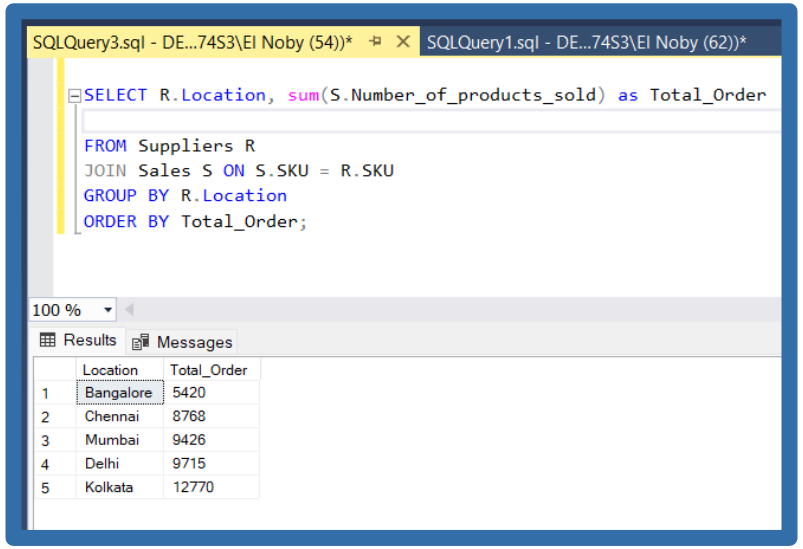
Digital Egypt Pioneers

2 . 4 Transportation

2 . 4 . 1 Which transportation modes are associated with the lowest costs and fastest delivery times?



2 . 4 . 2 Which Location have the largest amount of order shipping from?



2. Analysis Questions Phase

Digital Egypt Pioneers

2 . 5 Supplier

2 . 5 . 1 Average each Supplier Performance For lead time?

SQLQuery3.sql - DE...74S3\El Noby (54))* | SQLQuery1.sql - DE...74S3\El Noby (62))*

```
SELECT S.Supplier_name, AVG(S.Lead_time) AS Average_Lead_Time
FROM Suppliers S
GROUP BY S.Supplier_name;
```

100 %

Results | Messages

	Supplier_name	Average_Lead_Time
1	Supplier 1	14
2	Supplier 2	18
3	Supplier 3	20
4	Supplier 4	15
5	Supplier 5	18

2 . 5 . 2 What are the Location trends in Supplier orders number?

SQLQuery4.sql - DE...74S3\El Noby (65))* | SQLQuery

```
SELECT S.Supplier_name, S.Location,
Count(S.SKU) as total_order
FROM Suppliers S
GROUP BY S.Location ,S.Supplier_name
order by S.Supplier_name ,total_order ;
```

100 %

Results | Messages

	Supplier_name	Location	total_order
1	Supplier 1	Chennai	4
2	Supplier 1	Delhi	4
3	Supplier 1	Bangalore	5
4	Supplier 1	Mumbai	6
5	Supplier 1	Kolkata	8
6	Supplier 2	Kolkata	3
7	Supplier 2	Chennai	3
8	Supplier 2	Mumbai	5
9	Supplier 2	Bangalore	5
10	Supplier 2	Delhi	6

2. Analysis Questions Phase

Digital Egypt Pioneers

2 . 6 Quality Control

2 . 6 . 1 the overall Inspection results regarding each category

SQLQuery5.sql - DE...74S3\El Noby (63))* | SQLQuery4.sql - DE...74S3\El Noby (6

```
select P.Product_type , Q.Inspection_results ,
count( Q.Inspection_results) AS Total_Inspection_results
FROM Products P
join Quality_Control Q ON Q.SKU = P.SKU
Group by P.Product_type , Q.Inspection_results
order by P.Product_type , Q.Inspection_results
```

100 %

Results | Messages

	Product_type	Inspection_results	Total_Inspection_results
1	cosmetics	Fail	10
2	cosmetics	Pass	6
3	cosmetics	Pending	10
4	haircare	Fail	13
5	haircare	Pass	6
6	haircare	Pending	15
7	skincare	Fail	13
8	skincare	Pass	11
9	skincare	Pending	16

2 . 6 . 2 the Supplier with the highest average Defect rates

SQLQuery5.sql - DE...74S3\El Noby (63))* | SQLQuery4.sql -

```
select TOP 1 S.Supplier_name  ,
avg ( Q.Defect_rates) AS AVG_Defect_rates
FROM Suppliers S
join Quality_Control Q ON Q.SKU = S.SKU
Group by S.Supplier_name
order by AVG_Defect_rates DESC
```

100 %

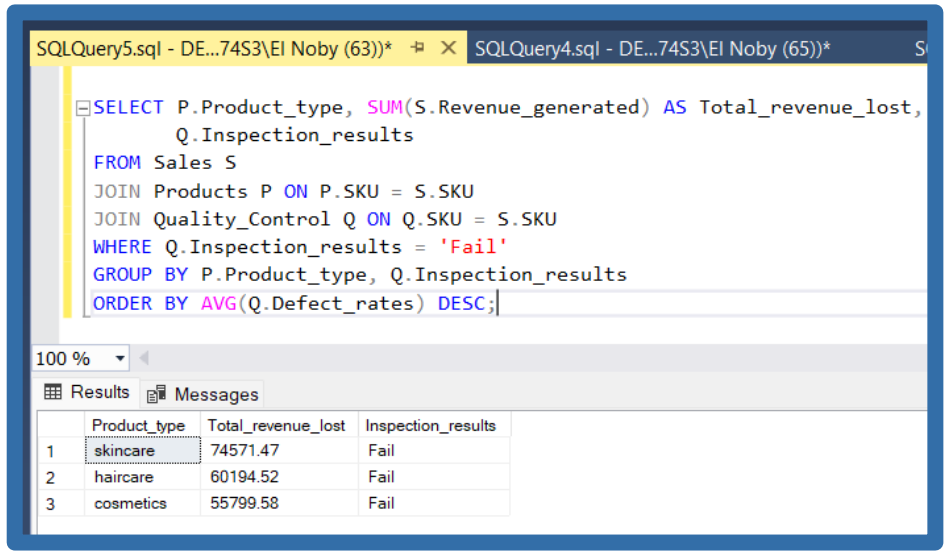
Results | Messages

	Supplier_name	AVG_Defect_rates
1	Supplier 5	2.665000

2. Analysis Questions Phase

2 . 6 Quality Control

2 . 6 . 3 total revenue lost regarding failed Inspection results



2 . 7 Conclusion

This analysis provides a comprehensive overview of the dataset, leveraging SQL to extract, clean, and analyze the data. The findings reveal valuable insights that can inform strategic decision-making.

* For Accessing the Python File

Supply Chain Project

Content

Digital Egypt Pioneers

3. Forecasting Questions Phase

Digital Egypt Pioneers

Using Linear Regression Packages in Python for Forecasting

- The goal is to use linear regression to analyze trends in a given dataset and answer forecasting questions. This involves:
- Identifying key trends: Determining the patterns or relationships within the data.
- Formulating forecasting questions: Creating specific questions based on the identified trends
- Applying linear regression: Using a Python library to model the relationship between variables and predict future values.
- Visualizing results: Creating plots to illustrate the trends and forecasts

* For Accessing the Python File

Supply Chain Project

Content

Digital Egypt Pioneers

4. Visualization Dashboard and Recommendation

4 . 1 Introduction

In this final phase of our project, we will transition from data analysis to data visualization.

By creating a comprehensive **Tableau dashboard**, we aim to transform the insights gained from our SQL queries into a visually compelling and interactive narrative. This dashboard will serve as a **powerful tool for stakeholders** to explore and understand the key findings of our analysis

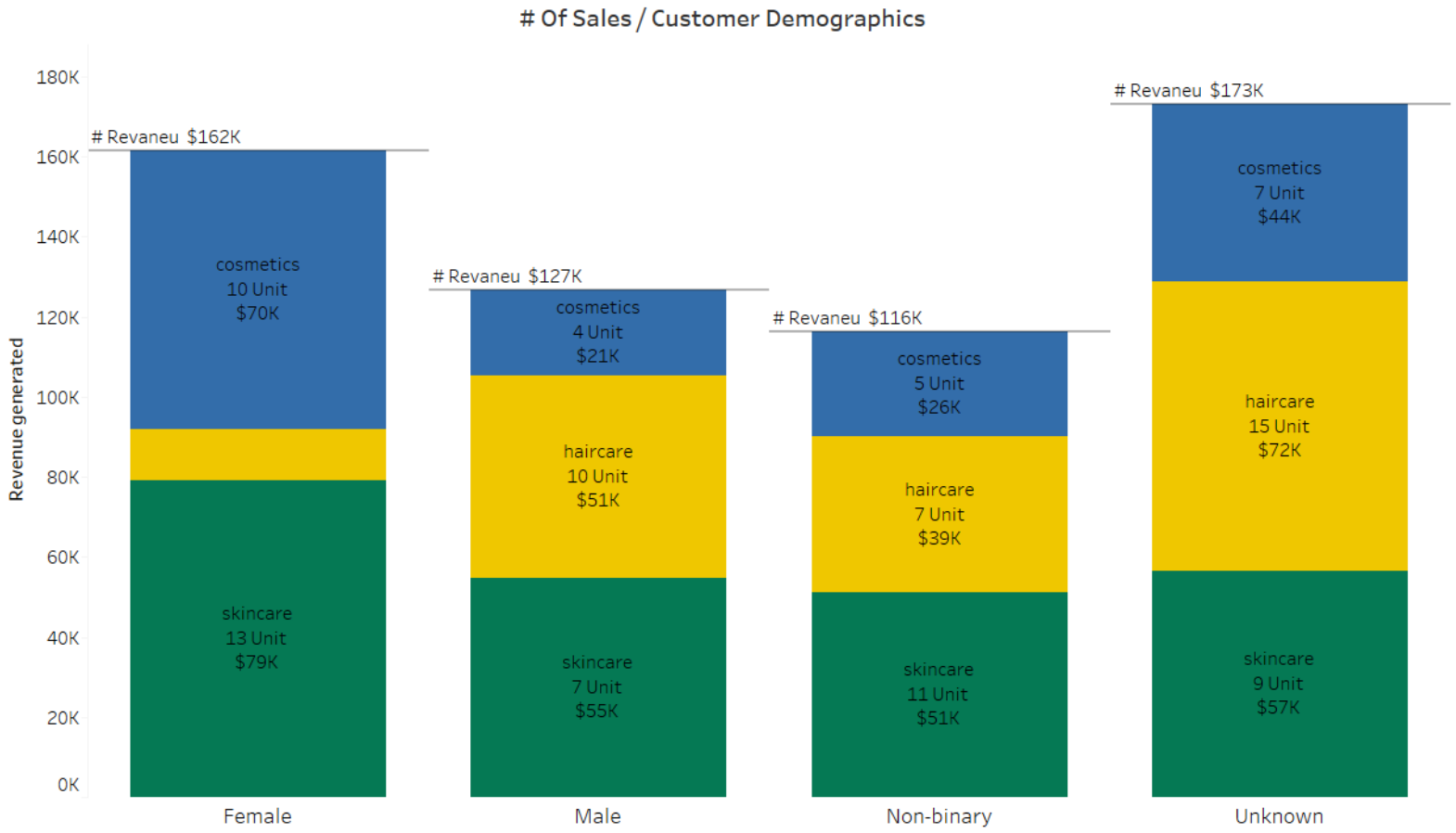
We have Organize Information: Grouped related visualizations into logical sections

We have visualize each visualization with two pieces of information at least To achieve maximum benefit

Store	Shipping	Manufacturing
Product type	Product type	Product type
SKU	SKU	SKU
Price	Order quantities	Production volumes
Availability	Shipping times	Manufacturing lead time
Number of products sold	Shipping carriers	Manufacturing costs
Revenue generated	Shipping costs	Inspection results
Customer demographics	Supplier name	Defect rates
Stock levels	Location	costs
Lead times	Lead time	
	Transportation modes	
	Routes	

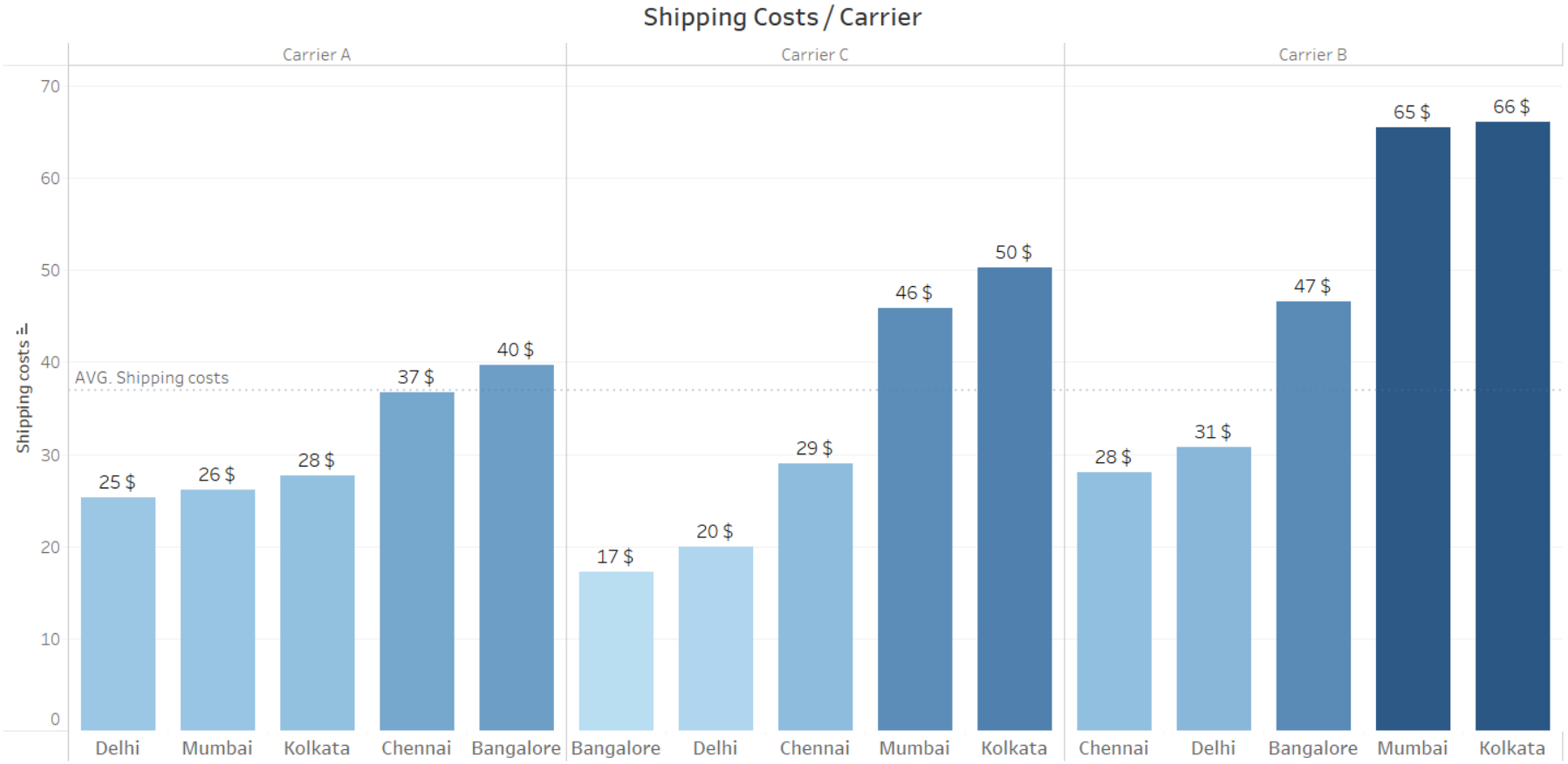
* For Tableau full Dashboard click here

4. Visualization Dashboard and Recommendation



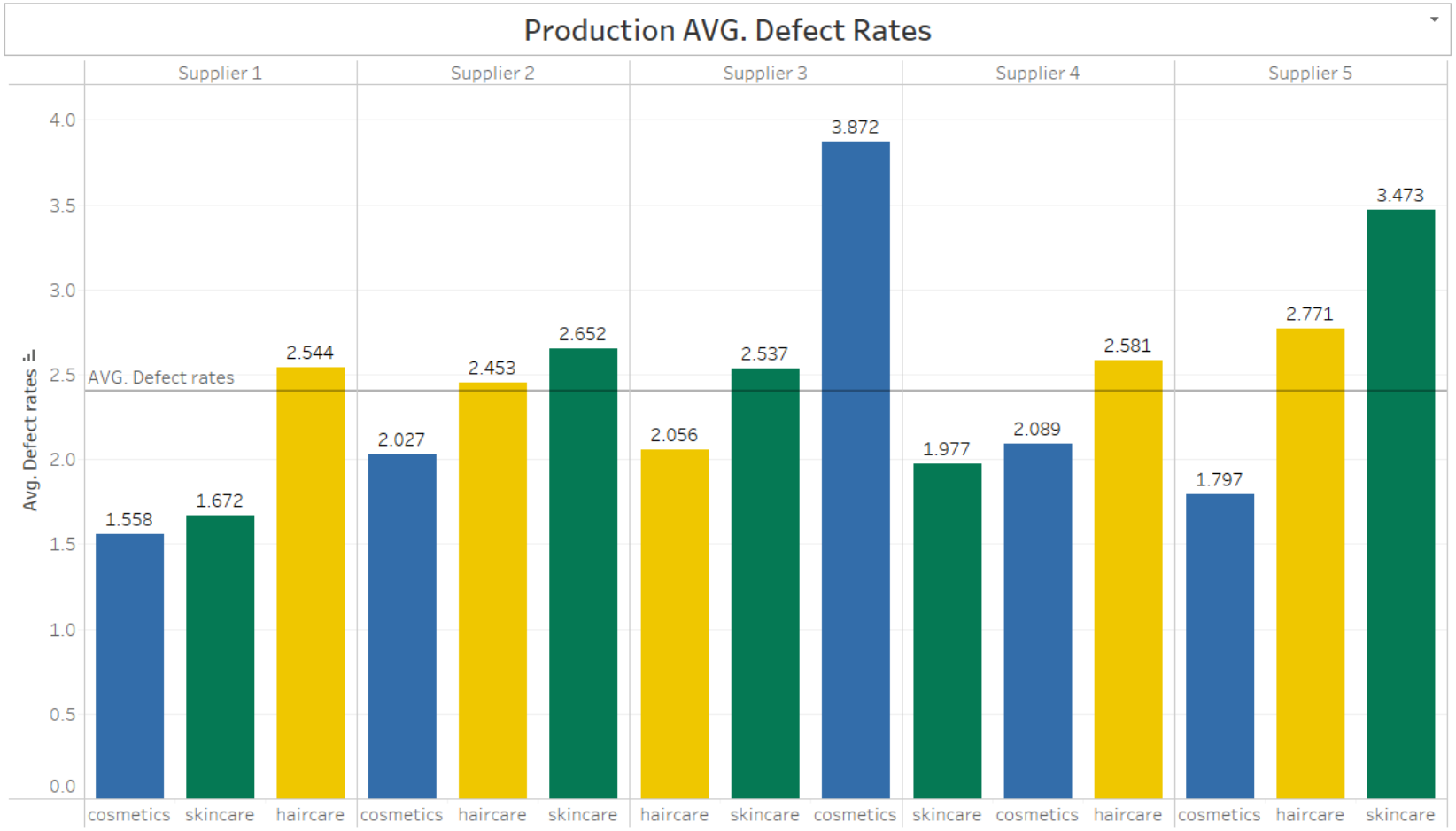
Store

4. Visualization Dashboard and Recommendation



Shipping

4. Visualization Dashboard and Recommendation



Manufacturing

4. Visualization Dashboard and Recommendation

Digital Egypt Pioneers

1. Target Skincare Products to Female:

Our analysis suggests that **skincare products are a popular category among female customers**. Consider focusing marketing efforts and product development on this segment to maximize sales and customer satisfaction.

2. Increase Skincare Product Stock:

Given the current stock levels falling below availability thresholds, it is **recommended to increase the inventory of skincare products**. This will help ensure product availability, prevent stockouts, and meet customer demand effectively.

Store Recommendation

4. Visualization Dashboard and Recommendation

1. Prioritize Road Transportation for Cost and Speed:

Our analysis indicates **that Road transportation is the most cost-effective and efficient mode for shipping**. Consider utilizing road transportation whenever possible to optimize shipping costs and delivery times.

2. Partner with Carrier A for Optimal Shipping Costs:

Based on our findings, **Carrier A consistently offers shipping costs that are below or within the overall average**.
Partnering with Carrier A can help reduce shipping expenses and improve overall supply chain efficiency.

4. Visualization Dashboard and Recommendation

Digital Egypt Pioneers

1. Address Quality Issues with Supplier 5 :

Supplier 5 has been identified as having **the highest average defect rates.** It is recommended to conduct a thorough investigation into the root causes of these defects and implement corrective actions to improve product quality.

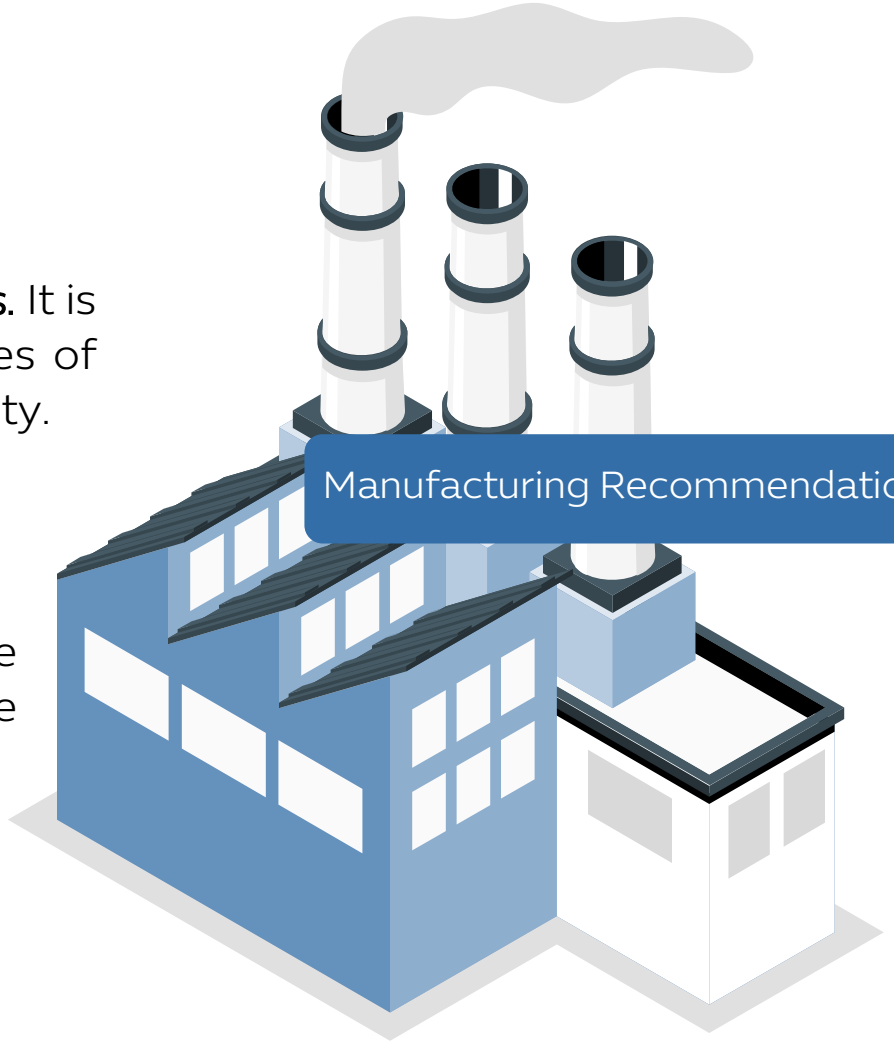
2. Optimize Lead Times with Supplier 3 :

Supplier 3 has been identified as having the longest average lead times.

With the lowest order Quantity, explore ways to expedite the procurement process and delivery timelines from Supplier 3 to improve supply chain efficiency and reduce costs.

3. Mitigate Revenue Loss from Failed Inspections :

The analysis reveals **that skincare products experience the highest revenue loss due to failed inspections, followed by hair care and cosmetics**. To minimize financial impact, focus on implementing stricter quality control measures for these product categories



FINAL PROJECT

Data Analyst Specialist Track

THANKS.

* For Full Project Attachment