Digital Egypt Pioneers Initiative (DEPI) 

Power BI (Data analysis program)

Code: DEPI\_ALX2\_DAT2\_S3

Analytics Avengers

Task 2: Data Cleaning Report

Project Info :Optimizing supply chain process through Data analysis

Dataset used: Indian companies data that involving in the supply chain process

Date submitted 28/3/2025



Prepared By:

|  |  |
| --- | --- |
| Name | الإسم |
| Mohamed Ayman Mohamed Bayoumi (Team Leader) | محمد ايمن محمد بيومي |
| Ritage Hamada Farouk abdelghany | ريتاج حماده فاروق عبدالغني |
| Youssef Atef Fawzy Eshak | يوسف عاطف فوزي اسحاق |
| Mohammed Youssef Fakhry Ahmed | محمد يوسف فخري احمد |
| Nada El Sayed Ali Hassan | ندى السيد علي حسن |

Table of Contents

[• Checking the Data: 2](#_Toc193471117)

[• Steps We Did in Power Query: 2](#_Toc193471118)

[• Final Clean Dataset (Before & After) 4](#_Toc193471119)

[• Problems You Faced & How You Solved Them 5](#_Toc193471120)

# • Checking the Data:

1. There were no missing values in the data.
2. There were no duplicate rows in the data.
3. There were no strange numbers or outliers detected in the data.
4. All columns are in the correct format, with text, numbers, and dates properly aligned to their respective categories.

# • Steps We Did in Power Query:

**Step 1**: We noticed the table name was 'Suppy\_chain\_data.' **-How**:Renaming the Table: In Power Query, we changed the table name to 'Supply\_chain.'  
**-Why:** This was done to simplify the name and make it more concise while keeping it clear and relevant.**A screenshot of a computer

AI-generated content may be incorrect.**

**Step 2:** Promoted headers

**How:** In Power Query, used the Promote Headers feature to convert the first row of data into the column headers. **Why:** To ensure that the dataset has proper column headers, making it easier to analyze and reference.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 1 before Promoted headers

**A screenshot of a computer

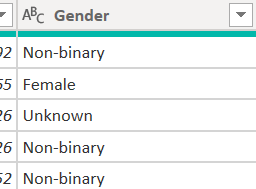
AI-generated content may be incorrect.**

Figure 2 after Promoted Headers

**Step 3:** The header name of the column 'Customer\_demographics' was changed to 'Gender'.

**-How**: This was done by selecting the column, right-clicking on the header, and renaming it to 'Gender'.

**-Why**: for better clarity and accuracy in representing the data.



**Step 3:** Identifying the Issue: In the "Gender" column, we found the word "Unknown."  
**-How:** Replacing Values: In Power Query, we replaced "Unknown" with "Prefer not to say."  
**-Why**: This was done to make the data more respectful and clear, providing a better option for users who may not want to disclose their gender.

A screenshot of a computer

AI-generated content may be incorrect.

**Step 4:** Identifying Duplicate Names: We noticed two columns with the same name, "Lead Time."  
**-How:** Renaming Columns: we renamed one column to "Supplier Lead Time" and the other to "Shipping Lead Time."  
 **-Why**: This was done to make the columns clear and avoid confusion in the analysis.

A screenshot of a computer

AI-generated content may be incorrect. **A screenshot of a computer

AI-generated content may be incorrect.**

**Step 5:** Identifying Large Decimal Numbers: We noticed that many columns contained very large decimal numbers.  
**-How:** Rounding Numbers: we applied rounding to these columns, limiting the decimal places to 3. **-Why**: This was done to make the data easier to read and work with, ensuring the values are more manageable and consistent.

A screenshot of a computer

AI-generated content may be incorrect.

# • Final Clean Dataset (Before & After)

Before: 101 row

After:100 row

# • Problems You Faced & How You Solved Them

**Problem:** Some Customer demographics were showing as "unkown"  
**Solution**: Replaced them with Prefer not say

**Problem:** The "Lead time" columns there was 2 coulmns with lead time name and the third with Manufacturing Lead time   
**Solution:** after Studying the data we found that data coulmn format is in grouped order as product, sales , supplier, shipping and, manufacturing so as manufacturing lead time came after manufacturing related data, we did for the rest and changed column names to “supplier lead time” ,“Shipping Lead time”

**Problem**: Many columns contained very large decimal numbers.  
**Solution**: We rounded the numbers to two decimal places to make the data easier to read and manage.

**Problem**: The table name was "Suppy\_chain\_data," which had a typo.  
**Solution**: We renamed the table to "Supply\_chain" to correct the typo.