📊 Supply Chain Analytics Project – Unilever (BizLearner)

By: Md Shahar Shitol
Role Targeting: Customer Service & Channel Excellence (CSCE)

🎯 Objective

Use data analytics to identify supply chain inefficiencies, cost drivers, and supplier performance bottlenecks for Unilever Bangladesh.

📁 Dataset Overview

- 100 SKUs across haircare, skincare, cosmetics
- 23 columns including price, revenue, defects, shipping, suppliers
- No missing values, clean data

Below is a sample snapshot:

🧹 Step 1: Data Cleaning

- Standardized column names
- Converted categorical columns for efficiency
- Exported cleaned CSV for later use

```
import pandas as pd
from google.colab import files

uploaded = files.upload()

# Load data
df = pd.read_excel("Supply Chain Analytics Uniliver.xlsx")

# Basic overview
print(df.info())
print(df.describe())
print(df.head())

# Rename columns for consistency
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")

# Check for missing/null values
missing = df.isnull().sum()
print("Missing values:\n", missing)

# Convert categorical fields
df['product_type'] = df['product_type'].astype('category')
```

```
df['customer_demographics'] = df['customer_demographics'].astype('category')
df['inspection_results'] = df['inspection_results'].astype('category')

# Save cleaned data
df.to_csv("cleaned_supply_chain_data.csv", index=False)
```

Choose Files Supply Cha…Uniliver.xlsx

- **Supply Chain Analytics Uniliver.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 27237 bytes, last modified: 6/23/2025 - 100% done

```
Saving Supply Chain Analytics Uniliver.xlsx to Supply Chain Analytics Uniliver (2).xlsx
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 23 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  Lead time                100 non-null    int64  
 15  Production volumes       100 non-null    int64  
 16  Manufacturing lead time  100 non-null    int64  
 17  Manufacturing costs      100 non-null    float64
 18  Inspection results       100 non-null    object 
 19  Defect rates             100 non-null    float64
 20  Transportation modes     100 non-null    object 
 21  Routes                   100 non-null    object 
 22  Costs                    100 non-null    float64
dtypes: float64(6), int64(9), object(8)
memory usage: 18.1+ KB
None
            Price  Availability  Number of products sold  Revenue generated  \
count  100.000000    100.000000               100.000000         100.000000   
mean    49.462461     48.400000               460.990000        5776.048187   
std     31.168193     30.743317               303.780074        2732.841744   
min      1.699976      1.000000                 8.000000        1061.618523   
25%     19.597823     22.750000               184.250000        2812.847151   
50%     51.239831     43.500000               392.500000        6006.352023   
75%     77.198228     75.000000               704.250000        8253.976921   
max     99.171329    100.000000               996.000000        9866.465458   

       Stock levels  Lead times  Order quantities  Shipping times  \
count    100.000000  100.000000        100.000000      100.000000   
mean      47.770000   15.960000         49.220000        5.750000   
std       31.369372    8.785801         26.784429        2.724283   
min        0.000000    1.000000          1.000000        1.000000   
25%       16.750000    8.000000         26.000000        3.750000   
50%       47.500000   17.000000         52.000000        6.000000   
75%       73.000000   24.000000         71.250000        8.000000   
max      100.000000   30.000000         96.000000       10.000000   

       Shipping costs   Lead time  Production volumes  \
count      100.000000  100.000000          100.000000   
mean         5.548149   17.080000          567.840000   
std          2.651376    8.846251          263.046861   
min          1.013487    1.000000          104.000000   
25%          3.540248   10.000000          352.000000   
50%          5.320534   18.000000          568.500000   
75%          7.601695   25.000000          797.000000   
```

```
max          9.929816   30.000000          985.000000

       Manufacturing lead time  Manufacturing costs  Defect rates       Costs
count                100.00000           100.000000    100.000000  100.000000
mean                  14.77000            47.266693      2.277158  529.245782
std                    8.91243            28.982841      1.461366  258.301696
min                    1.00000             1.085069      0.018608  103.916248
25%                    7.00000            22.983299      1.009650  318.778455
50%                   14.00000            45.905622      2.141863  520.430444
75%                   23.00000            68.621026      3.563995  763.078231
max                   30.00000            99.466109      4.939255  997.413450
  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  ...  Supplier name Lead time  Production volumes  \
0                96  ...     Supplier 3        29                 215
1                37  ...     Supplier 3        23                 517
2                88  ...     Supplier 1        12                 971
3                59  ...     Supplier 5        24                 937
4                56  ...     Supplier 1         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
0      0.226410                  Road  Route B  187.752075
1      4.854068                  Road  Route B  503.065579
2      4.580593                   Air  Route C  141.920282
3      4.746649                  Rail  Route A  254.776159
4      3.145580                   Air  Route A  923.440632

[5 rows x 23 columns]
Missing values:
 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
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
```

```
# Clean column names
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")

# Convert relevant columns to category type
cat_cols = ['product_type', 'customer_demographics', 'inspection_results',
            'shipping_carriers', 'supplier_name', 'transportation_modes', 'routes']
for col in cat_cols:
    df[col] = df[col].astype('category')

# Preview categorical distributions
for col in cat_cols:
    print(f"\nUnique values in '{col}':")
    print(df[col].value_counts())

# Export cleaned CSV
df.to_csv("cleaned_supply_chain_data.csv", index=False)
print("✅ Cleaned CSV exported successfully.")
```

```
Unique values in 'product_type':
product_type
skincare     40
haircare     34
cosmetics    26
Name: count, dtype: int64

Unique values in 'customer_demographics':
customer_demographics
Unknown       31
Female        25
Non-binary    23
Male          21
Name: count, dtype: int64

Unique values in 'inspection_results':
inspection_results
Pending    41
Fail       36
Pass       23
Name: count, dtype: int64

Unique values in 'shipping_carriers':
shipping_carriers
Carrier B    43
Carrier C    29
Carrier A    28
Name: count, dtype: int64

Unique values in 'supplier_name':
supplier_name
Supplier 1    27
Supplier 2    22
Supplier 4    18
Supplier 5    18
Supplier 3    15
Name: count, dtype: int64

Unique values in 'transportation_modes':
transportation_modes
Road    29
```

```
Rail    28
Air     26
Sea     17
Name: count, dtype: int64

Unique values in 'routes':
routes
Route A    43
Route B    37
Route C    20
Name: count, dtype: int64
✅ Cleaned CSV exported successfully.
```

🧠 Step 2: Exploratory Insights Using SQL

We explore key business questions such as:

- Top SKUs by revenue
- Supplier defect rates
- Shipping cost efficiency by carrier

Double-click (or enter) to edit

```
import sqlite3

# Create SQLite DB in memory
conn = sqlite3.connect(':memory:')
cursor = conn.cursor()

# Load CSV into a new SQL table
df.to_sql("supply_chain_data", conn, index=False, if_exists='replace')

# Preview the first 5 rows using SQL
query = "SELECT * FROM supply_chain_data LIMIT 5;"
pd.read_sql(query, conn)
```

	product_type	sku	price	availability	number_of_products_sold	revenue_generated	customer
0	haircare	SKU0	69.808006	55	802	8661.996792	
1	skincare	SKU1	14.843523	95	736	7460.900065	
2	haircare	SKU2	11.319683	34	8	9577.749626	
3	skincare	SKU3	61.163343	68	83	7766.836426	
4	skincare	SKU4	4.805496	26	871	2686.505152	

5 rows × 23 columns

```
query = """
SELECT sku, product_type, SUM(revenue_generated) AS total_revenue
FROM supply_chain_data
GROUP BY sku, product_type
ORDER BY total_revenue DESC
LIMIT 5;
"""
pd.read_sql(query, conn)
```

	sku	product_type	total_revenue
0	SKU51	haircare	9866.465458
1	SKU38	cosmetics	9692.318040
2	SKU31	skincare	9655.135103
3	SKU90	skincare	9592.633570
4	SKU2	haircare	9577.749626

```
query = """
SELECT supplier_name, ROUND(AVG(defect_rates), 2) AS avg_defect_rate
FROM supply_chain_data
GROUP BY supplier_name
ORDER BY avg_defect_rate ASC;
"""
pd.read_sql(query, conn)
```

	supplier_name	avg_defect_rate
0	Supplier 1	1.80
1	Supplier 4	2.34
2	Supplier 2	2.36
3	Supplier 3	2.47
4	Supplier 5	2.67

```
query = """
SELECT shipping_carriers, ROUND(AVG(shipping_costs), 2) AS avg_shipping_cost
FROM supply_chain_data
GROUP BY shipping_carriers
ORDER BY avg_shipping_cost DESC;
"""
pd.read_sql(query, conn)
```

	shipping_carriers	avg_shipping_cost
0	Carrier C	5.60
1	Carrier A	5.55
2	Carrier B	5.51

```
query = """
SELECT sku, product_type, availability, revenue_generated
FROM supply_chain_data
WHERE availability < 30
ORDER BY revenue_generated DESC
LIMIT 10;
"""
pd.read_sql(query, conn)
```

	sku	product_type	availability	revenue_generated
0	SKU31	skincare	28	9655.135103
1	SKU67	skincare	16	9473.798033
2	SKU52	skincare	1	9435.762609
3	SKU18	haircare	23	9364.673505
4	SKU99	haircare	17	9185.185829
5	SKU60	skincare	16	8864.084350
6	SKU14	skincare	26	8653.570926
7	SKU64	skincare	11	8458.730878
8	SKU29	cosmetics	3	8318.903195
9	SKU71	cosmetics	14	8180.337085

```
query = """
SELECT sku, supplier_name, manufacturing_costs, defect_rates
FROM supply_chain_data
ORDER BY manufacturing_costs DESC
LIMIT 10;
"""
pd.read_sql(query, conn)
```