## Input Data

About this file

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:

| **No.** | **Column Name** | **Column importance** | **Statistical description** |
| --- | --- | --- | --- |
| **1** | **Product type** | This column indicates the type of product (e.g., haircare, skincare, cosmetics). It’s useful for segmenting and analyzing performance by product category. | Product type:  Product type  skincare 40  haircare 34  cosmetics 26  Name: count, dtype: int64 |
| **2** | **SKU** | SKU (Stock Keeping Unit) is a unique identifier for each product. It’s useful for tracking individual products but may not be needed for aggregate analysis. You can decide whether to keep it based on your analysis goals. |  |
| **3** | **Price** | The price of each product is important for revenue analysis and pricing strategies. | * + Mean: ~49.46   + Range: 1.70 (min) to 99.17 (max)   + The prices vary significantly across products. |
| **4** | **Availability** | This column shows the availability of products (e.g., stock levels). It’s useful for inventory management analysis. | * + Mean: ~48.40   + Range: 1 (min) to 100 (max)   + Availability is evenly distributed, with some products having very low or very high availability. |
| **5** | **Number of products sold** | This column is crucial for sales analysis and understanding product performance. | * + Mean: ~460.99   + Range: 8 (min) to 996 (max)   + Some products sell significantly more than others. |
| **6** | **Revenue generated** | Revenue is a key metric for analyzing the financial performance of products. | * + Mean: ~5776.05   + Range: 1061.62 (min) to 9866.47 (max)   + Revenue varies widely, indicating some products are more profitable. |
| **7** | **Customer demographics** | This column provides information about the customer base (e.g., gender, age group). It’s useful for customer segmentation and targeted marketing. | Customer demographics:  Customer demographics  Unknown 31  Female 25  Non-binary 23  Male 21  Name: count, dtype: int64 |
| **8** | **Stock levels** | This column indicates the current stock levels, which is important for inventory management and demand forecasting. | * + Mean: ~47.77   + Range: 0 (min) to 100 (max)   + Some products are out of stock, while others are fully stocked. |
| **9** | **Lead times** | Time from customer order received to order delivery in days, It’s important for analyzing supply chain efficiency. | * + Mean: ~15.96   + Range: 1 (min) to 30 (max)   + Lead times vary, which could impact supply chain efficiency. |
| **10** | **Order quantities** | This column shows the quantity of products ordered. It’s useful for demand analysis and inventory planning. | * + Mean: ~49.22   + Range: 1 (min) to 96 (max)   + Order quantities vary significantly. |
| **11** | **Shipping times** | Shipping time is important for analyzing delivery performance and customer satisfaction, Time taken to ship products to customers in days | * + Mean: ~5.75   + Range: 1 (min) to 10 (max)   + Shipping times are relatively consistent. |
| **12** | **Shipping carriers** | This column indicates the shipping carrier used. It’s useful for analyzing carrier performance and costs. | Shipping carriers:  Shipping carriers  Carrier B 43  Carrier C 29  Carrier A 28  Name: count, dtype: int64 |
| **13** | **Shipping costs** | Shipping costs are important for analyzing logistics expenses and profitability. | * + Mean: ~5.55   + Range: 1.01 (min) to 9.93 (max)   + Shipping costs vary, which could impact profitability. |
| **14** | **Supplier name** | This column identifies the supplier. It’s useful for analyzing supplier performance and reliability. | Supplier name:  Supplier name  Supplier 1 27  Supplier 2 22  Supplier 5 18  Supplier 4 18  Supplier 3 15  Name: count, dtype: int64 |
| **15** | **Location** | This column indicates the location of suppliers or customers. It’s useful for geographic analysis and logistics planning. | Location:  Location  Kolkata 25  Mumbai 22  Chennai 20  Bangalore 18  Delhi 15  Name: count, dtype: int64 |
| **16** | **Lead time** | Time taken by suppliers to deliver raw materials in days | * Mean 17.1 * Quantiles 1 Min   10 25%  18 50%  25 75%  30 Max   * Lead times vary, which could impact supply chain efficiency. |
| **17** | **Production volumes** | This column shows the volume of products produced. It’s useful for production planning and capacity analysis. | * + Mean: ~567.84   + Range: 104 (min) to 985 (max)   + Production volumes vary significantly. |
| **18** | **Manufacturing lead time** | This column indicates the time taken to manufacture products. It’s useful for production efficiency analysis.(Time from production start to completion in days) | * Mean 14.8 * Quantiles 1 Min   7 25%  14 50%  23 75%  30 Max |
| **19** | **Manufacturing costs** | This column shows the cost of manufacturing. It’s important for cost analysis and profitability | * + Mean: ~47.27   + Range: 1.09 (min) to 99.47 (max)   + Manufacturing costs vary widely. |
| **20** | **Inspection results** | This column indicates the results of product inspections (e.g., Pass, Fail). It’s useful for quality control analysis. | Inspection results:  Inspection results  Pending 41  Fail 36  Pass 23  Name: count, dtype: int64 |
| **21** | **Defect rates** | This column shows the defect rate of products. It’s important for quality control and process improvement. | * + Mean: ~2.28   + Range: 0.02 (min) to 4.94 (max)   + Defect rates are relatively low but vary across products. |
| **22** | **Transportation modes** | This column indicates the mode of transportation (e.g., Road, Air, Sea). It’s useful for logistics analysis. | Transportation modes:  Transportation modes  Road 29  Rail 28  Air 26  Sea 17  Name: count, dtype: int64 |
| **23** | **Routes** | This column indicates the transportation routes. It’s useful for optimizing logistics and reducing costs. | Routes:  Routes  Route A 43  Route B 37  Route C 20  Name: count, dtype: int64 |
| **24** | **Costs** | This column is ambiguous. It’s unclear whether it refers to transportation costs, manufacturing costs, or something else. | * + Mean: ~529.25   + Range: 103.92 (min) to 997.41 (max)   + Costs vary significantly, which could impact profitability. |