**Afrimash Dataset Breakdown**

**RFM (Recency, Frequency and Monetary) Data**

| **Column Name** | **Description** |
| --- | --- |
| **Customer\_ID** | Unique identifier assigned to each customer across all transactions. It serves as the primary key linking the RFM dataset to the transaction dataset. |
| **Frequency** | Number of completed purchase transactions made by the customer during the observed period. Higher values indicate more active or loyal customers. |
| **Monetary** | Total revenue generated by the customer from all purchases (Net Sales aggregated across transactions). This measures the customer's overall value to the business. |
| **Avg\_Order\_Value** | The average amount spent per transaction by the customer. Calculated as Monetary ÷ Frequency. Indicates spending behaviour per purchase. |
| **Customer\_Lifetime\_Days** | Total number of days between the customer’s first and last purchase (or registration date and last purchase). Reflects how long the customer has been active. |
| **Purchase\_Rate** | Average purchase frequency per day, calculated as Frequency ÷ Customer\_Lifetime\_Days. Useful for predicting future purchase probability. |
| **Customer\_Type** | Classification of customer behaviour — typically **‘new’** (first-time buyers) or **‘returning’** (repeat customers). |
| **Attribution** | The traffic or marketing source that brought the customer (e.g., Direct, Organic: Google, Paid Ad, Referral). Useful for channel performance analysis. |
| **Total\_Items\_Sold** | The total number of items the customer purchased across all orders. Reflects purchase volume and complements the frequency metric. |

**Transaction Data**

| **Column Name** | **Description** |
| --- | --- |
| **Customer\_ID** | Unique identifier linking each transaction to the specific customer (matches the Customer\_ID in the RFM data). |
| **Order #** | Unique identifier for each completed order transaction. Can be used to count distinct purchases per customer. |
| **Date** | Timestamp showing when the order was placed (in YYYY-MM-DD HH:MM:SS format). Enables time-based analysis like recency or purchase intervals. |
| **Product(s)** | Description of the product(s) purchased in each transaction, often including quantity, brand, and variant details. Useful for product-level segmentation. |
| **Items sold** | Number of items included in the order. This, when summed by Customer\_ID, corresponds to Total\_Items\_Sold in the RFM data. |
| **Revenue** | Gross revenue from the transaction before deductions (e.g., taxes, discounts, or adjustments). Useful for sales trend analysis. |
| **Net\_Sales** | Actual revenue earned after cleaning adjustments like discounts, refunds, and taxes. This is the net figure used in the RFM’s **Monetary** value. |
| **Status** | Indicates the completion state of the transaction (e.g., *completed*, *cancelled*, *pending*). Only “completed” transactions contribute to the RFM calculations. |