L.Louw

Data Stimulation

Sales Data Simulation for Harmony Electronics – Assessment Specifications

L.Louw

Task: Data Analytics / Data Simulation and Reporting

Tools Required: Anaconda, Jupyter Notebook, Python libraries (Pandas, NumPy, Matplotlib, Seaborn)

Assessment Type: Practical / Hands-on Assignment

Submission Format: Jupyter Notebook (.ipynb) and .csv dataset

Data Creation Task

You are tasked with simulating sales data for **Harmony Electronics**, a retail store specializing in electronics and home appliances. Generate a .csv dataset with **at least 250 sales transactions**. Each row should represent a single transaction. Your dataset must include the following columns:

1.   **Transaction ID** – Unique identifier for each sale.

2.   **Customer ID** – Unique identifier for each customer.

3.   **Customer Age Group** – Categorized as 18–25, 26–35, 36–50, 51+

4.   **Product Category** – Electronics, Home Appliances, Mobile Devices, Gaming, Audio, Accessories

5.   **Product Name** – Specific item purchased (e.g., "Samsung Galaxy S23", "Sony WH-1000XM5")

6.   **Unit Price (R)** – The price per unit of the product

7.   **Quantity Purchased** – Number of units bought per transaction

8.   **Total Sale Amount (R)** – Calculated as Unit Price × Quantity Purchased

9.   **Date of Purchase** – Randomly distributed dates over the last 12 months; consider seasonal trends (e.g., more electronics sold in December due to holiday shopping)

10.   **Payment Method** – Cash, Credit Card, Debit Card, EFT

Additional Guidelines:

•   Some customers may have **multiple transactions**; others may appear only once.

•   Prices should reflect realistic values for each category.