

Class Activity - Simple Data Mining

The following is a sample dataset with a hidden pattern designed for class activity. The dataset contains sales transactions with product categories, dates, and revenue. A specific **time-based pattern** and **product-revenue relationship** are embedded.

Dataset: Sales Data (Hidden Pattern)

Transaction_ID	Date	Product_Category	Region	Revenue (\$)
1	2023-01-01	Electronics	North	1200
2	2023-01-02	Clothing	South	800
3	2023-01-03	Electronics	East	1100
4	2023-02-01	Electronics	North	1400
5	2023-02-02	Clothing	West	750
6	2023-03-01	Electronics	North	1300
7	2023-03-02	Clothing	South	780
8	2023-03-15	Electronics	North	1200
9	2023-03-20	Electronics	East	1350
10	2023-04-01	Clothing	South	810
11	2023-04-02	Electronics	North	1200
12	2023-04-10	Clothing	West	790
13	2023-05-01	Electronics	East	1450
14	2023-05-02	Clothing	South	820
15	2023-05-20	Electronics	North	1500

Class Activity

Your task is to:

- 1. Identify the Pattern:**
 - Analyze the dataset using techniques like visualization, clustering, or association rule mining to discover the patterns.
- 2. Explain the Findings:**
 - Write a summary of the identified patterns, trends, and insights.
- 3. Methodology:**
 - Use any data mining technique you prefer (e.g., manual sorting, calculating averages, clustering by region/product, or visualizing trends).

Answer: Hidden Patterns

1. Time-Based Pattern:

- Electronics revenue spikes in **March and May**, particularly in the **North region**.
- Clothing sales have consistent but slightly increasing revenue over time.

2. Region-Product Relationship:

- Electronics dominate the **North and East** regions in revenue.
- Clothing has steady sales in the **South**.

3. Revenue Threshold:

- Transactions with revenues over **\$1200** are mostly associated with **Electronics in the North or East regions**.