**Name- Sanchari Maji**

**Reg No- 2023PGDM1225**

**Mini Project Report**

**Simulating MapReduce for Customer Purchase Pattern Analysis**

**1. Objective**

The objective of this project is to simulate the MapReduce framework using a sample retail transaction dataset. It analyses customer purchase patterns to identify top-selling products, revenue contributions by product category, and the most frequent buyers.

**2. Business Scenario**

Assuming the role of a data analyst in a retail company, the task is to analyse transactional data using MapReduce principles. This simulation helps derive valuable insights into sales trends and customer behaviour, aiding in managerial decision-making.

**3. Dataset Description**

A dataset of 50 manually simulated retail transactions was used. Each record includes:

* Transaction ID
* Customer ID
* Product
* Category
* Quantity
* Price
* Date

**4. Tools Used**

* Microsoft Excel
* Tableau (for graphs and chart)
* Word (for documentation)

**5. Methodology**

The analysis is divided into two main phases:

**🧩 Map Phase:**

Key-value pairs were extracted as follows:

* Product → Quantity
* Category → Revenue (Quantity × Price)
* CustomerID → Frequency of transactions

**🧮 Reduce Phase:**

Values were aggregated to generate insights:

* Sum of quantity sold for each product
* Total revenue for each category
* Number of purchases per customer

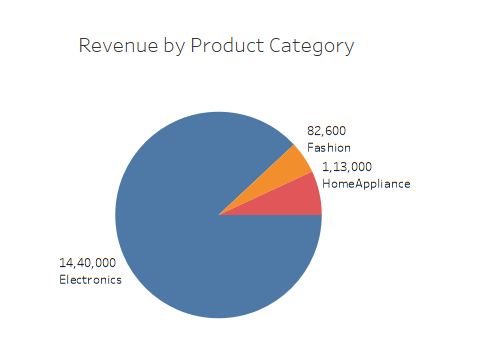
**6. Results**

**📊 Top 5 Selling Products by Quantity**

****

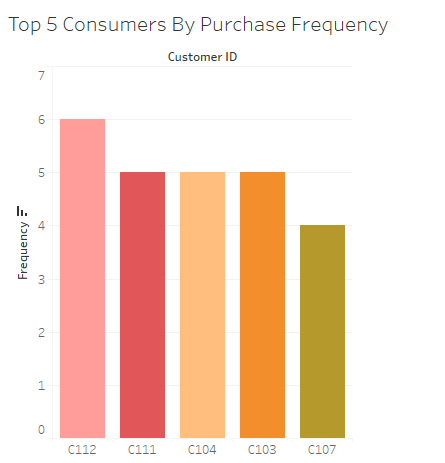
* Tablet – 17 units
* Shirt – 14 units
* Coffee Maker – 11 units
* Jacket – 8 units
* Jeans – 8 units

**🥧 Revenue Distribution by Category**

****

* Electronics – ₹14,40,000
* Home Appliance – ₹1,13,000
* Fashion – ₹82,600

**📊 Top 5 Customers by Purchase Frequency**

****

* C112 – 6 purchases
* C111 – 5 purchases
* C104 – 5 purchases
* C103 – 5 purchases
* C107 – 4 purchases

**7. Conclusion**

This project successfully demonstrates how MapReduce can be applied to a practical business problem using simple tools like Excel and Python. The insights gained from this simulation can help management in:

* Optimizing inventory
* Personalizing marketing
* Enhancing customer relationships

It highlights how big data techniques like MapReduce can be simplified for real-world business decisions.