**NAME:- SAHIL KHAN**

**COURSE:- BIG DATA ANALYTICS**

**REGISTRATION NO:- 2023PGDM1215**

**Mini Project Summary: MapReduce Simulation for Customer Purchase Pattern Analysis**

**Objective**

The objective of this project is to simulate the MapReduce framework using a retail transaction dataset in Excel and visualize insights using Power BI. From a business perspective, the goal is to extract actionable insights such as:

* The top-selling products
* The highest revenue-generating categories
* The most frequent or loyal customers

This simulation helps understand how Map (key-value generation) and Reduce (aggregation) principles can be used for large-scale data analysis, even on a smaller scale with Excel and BI tools.

**Dataset Overview**

A dataset of 50 customer transactions was created, consisting of the following fields:

* Transaction ID
* Customer ID
* Product
* Category
* Quantity
* Price
* Date
* Revenue (derived field: Quantity × Price)

Products belonged to categories such as *Stationery, Education, Electronics, Office Supplies,* and *Lifestyle*.

Method Used: Map & Reduce Simulation

* **Map Phase (Key-Value Pairs):**

We simulated the Map phase by generating the following key-value pairs:

* Product → Quantity
* Category → Revenue
* Customer ID → Count of Purchases
* **Reduce Phase (Aggregation):**

Using Pivot Tables in Excel and Visuals in Power BI, we aggregated the above key-value pairs to:

* Sum total quantity for each product
* Sum total revenue for each category
* Count the number of purchases per customer

**Key Insights**

1. Which product sold the most?

Based on the bar chart, the product "Ruler" had the highest quantity sold, followed closely by "Keyboard" and "Workbook".

2. Which category earned the most revenue?

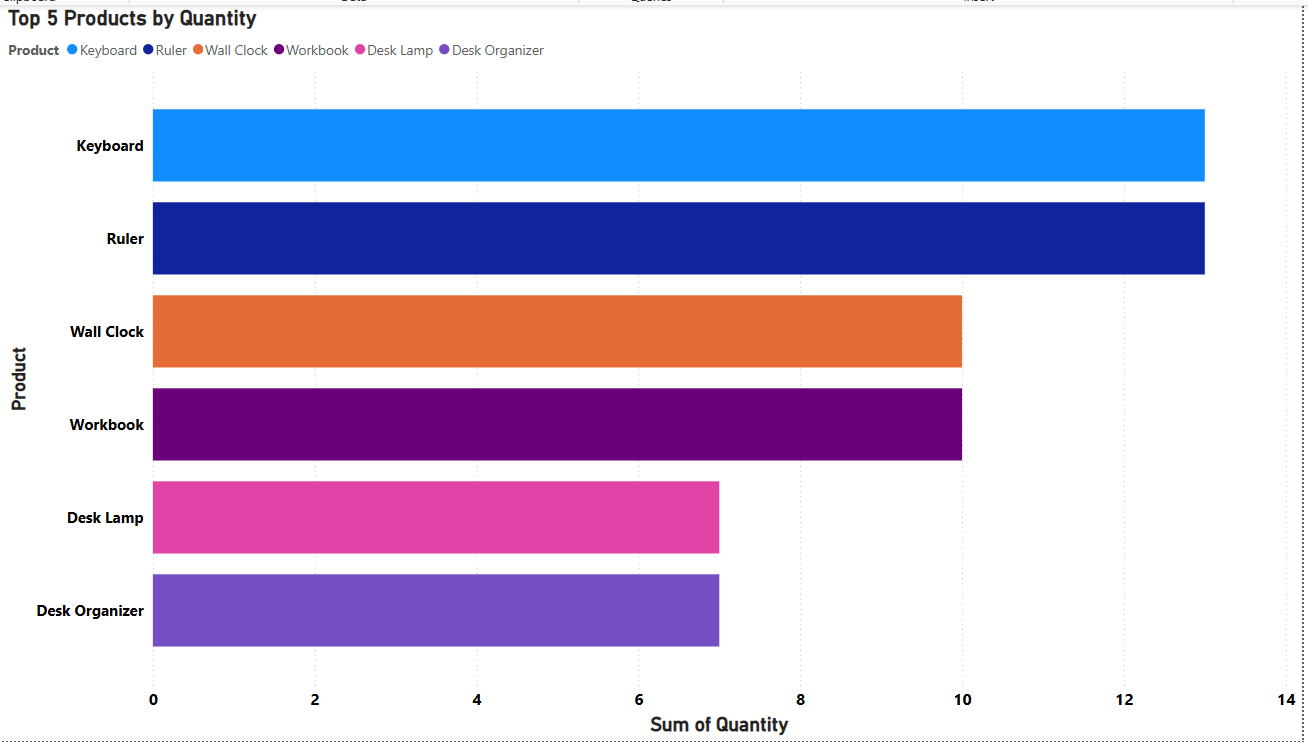
From the pie chart, the "Electronics" category accounted for the highest percentage of total revenue, making it the most profitable category.

3. Who are your most loyal customers?

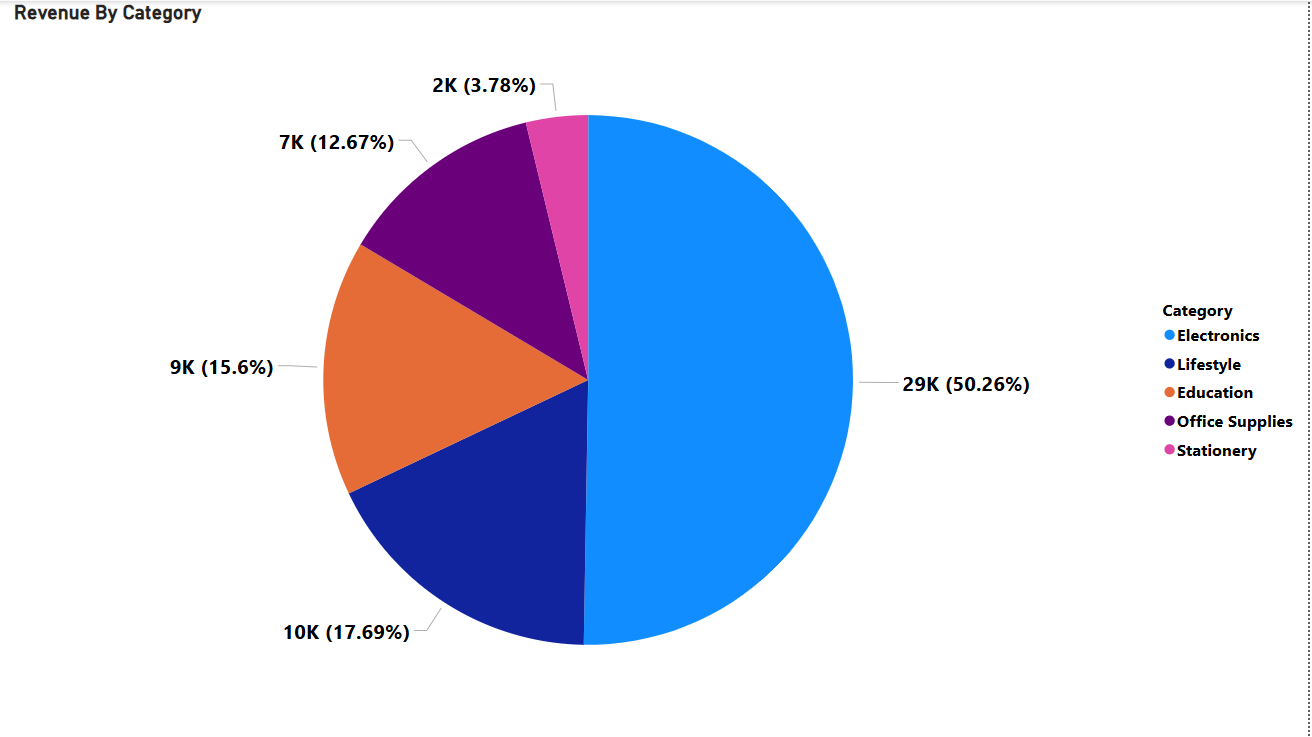
The bar chart for customer frequency revealed that customers like C105, C110, and C102 made the most transactions, identifying them as the most loyal or active buyers.

Sample Visuals

Below are examples of visuals created in Power BI:

**Bar Chart – Top 5 Products by Quantity Sold**

**Pie Chart of Revenue By Category**

****

**Conclusion**

This mini project demonstrates how Big Data principles like MapReduce can be applied practically using simple tools such as Excel and Power BI. It highlights how businesses can extract key insights from structured transaction data, improving decisions related to product strategy, customer loyalty, and category performance.