**CUSTOMER SEGMENTATION IN E-COMMERCE BUSINESSES**

**Objective**

To identify and categorize customers based on their purchasing behavior using the RFM model — Recency, Frequency, and Monetary value — to enable personalized marketing strategies and improve customer retention.

**Dataset Overview**

* **Source**: E-commerce transactions dataset from various sources
* **Fields Used**:
  + CustomerID: Unique customer identifier
  + InvoiceNo: Invoice number per transaction
  + InvoiceDate: Date of purchase
  + Quantity: Number of items bought
  + UnitPrice: Price per item
* **Derived Field**:
  + TotalPrice: Quantity × UnitPrice (used to calculate Monetary value)

**Project Workflow**

* **Data Cleaning**
* Removed records with missing customer IDs and zero or negative quantities.
* Added a new column for total purchase value (Quantity × UnitPrice).
* **RFM Metric Calculation (SQL)**
* **Recency**: Number of days since the last purchase (relative to analysis date).
* **Frequency**: Total number of distinct purchase transactions per customer.
* **Monetary**: Total money spent by each customer across all purchases.
* Grouped data by customer to calculate these metrics.
* **RFM Scoring**
* Divided each metric into 5 segments (quantiles or quintiles).
* Assigned scores from 1 to 5 for:
  + **Recency**: Lower days = higher score (more recent = better).
  + **Frequency** & **Monetary**: Higher value = higher score.
* Combined R, F, and M scores into a 3-digit RFM Score (e.g., 555, 231).
* **Customer Segmentation**

Created customer segments based on RFM Score combinations:

| **Segment** | **RFM Score Patterns** | **Description** |
| --- | --- | --- |
| Champions | 555, 554, 545 | Recent, frequent, and high spenders |
| Loyal Customers | 444–555 range | Regular and high-value customers |
| New Customers | R=5, F ≤ 2 | New but recent customers |
| At Risk | R=2 or lower | Once active, now inactive |
| Lost | 111 | Long inactive, low-frequency and spend |
| Potential Loyalists | F ≥ 3, M ≥ 3 | Medium engagement with strong potential |

**Key Insights**

* **Champions (31.3%)** generated nearly **40% of total revenue**.
* **Loyal Customers (24%)** showed steady purchasing behavior — ideal for loyalty programs.
* **At Risk Customers (22.6%)** can be re-engaged via special offers.
* **Lost Customers (22.1%)** may require win-back campaigns or exit surveys.