👥 ***Customer Segmentation using RFM Analysis (Python)***

📌 Project Overview

This project performs customer segmentation using **RFM (Recency, Frequency, Monetary) analysis** on a realistic e-commerce transactional dataset.
The objective is to identify high-value customers, loyal customers, and at-risk customers to support data-driven marketing and retention strategies.

The project follows a complete **end-to-end analytics workflow**, starting from raw transactional data to actionable business insights and exportable customer segments.

🎯 Objectives

- Segment customers based on purchasing behavior
- Identify high-value and loyal customers
- Detect at-risk and inactive customers
- Translate analytical insights into actionable marketing strategies
- Provide marketing-ready customer segment exports

🛠️ Tools & Technologies

- **Language:** Python
- **Environment:** Google Colab (Cloud-based Jupyter Notebook)
- **Libraries:**
    - pandas
    - numpy
    - matplotlib
    - seaborn

📊 Dataset

- **Type:** Synthetic but realistic e-commerce transaction data
- **Records:** ~8,000 transaction-level rows
- **Key Characteristics:**
  - Multiple transactions per customer
  - Missing customer IDs (guest checkouts)
  - Negative quantities (returns)
  - Noisy pricing and transactional logs

Key Fields

- InvoiceNo
- InvoiceDate
- CustomerID
- Country
- Quantity
- UnitPrice
- TotalPrice

The dataset was intentionally designed to be **messy and realistic**, requiring data cleaning before analysis, similar to real-world business datasets.

🔍 Data Cleaning & Preprocessing

- Removed transactions with missing `CustomerID`
- Filtered out returns and invalid transactions (negative quantity or price)
- Removed duplicate records
- Recomputed total transaction value after cleaning
- Defined a snapshot date for accurate recency calculation

📐 RFM Feature Engineering

For each customer:

- **Recency:** Days since last purchase
- **Frequency:** Number of unique purchases
- **Monetary:** Total spending

Customers were scored using **quantile-based RFM scoring (1–5)** and combined into a composite RFM score.

👥 Customer Segmentation

Customers were grouped into meaningful business segments:

- **Champions**
- **Loyal Customers**
- **Potential Loyalists**
- **At Risk**
- **Lost Customers**

Segmentation was validated using:

- ➔ RFM score distributions
- ➔ Heatmaps
- ➔ 3D RFM visualizations

📊 Visualizations

- ❖ RFM Heatmap (Recency vs Frequency vs Monetary)
- ❖ 3D Scatter Plot of customer clusters
- ❖ Segment-wise customer distribution
- ❖ Monetary value comparison across segments

🎯 Marketing Strategy Recommendations

Each customer segment was mapped to targeted marketing actions:

- ❖ Loyalty and premium offers for Champions
- ❖ Upsell and engagement campaigns for Loyal Customers
- ❖ Win-back campaigns for At Risk customers
- ❖ Selective re-engagement or suppression for Lost Customers

---

📤 Deliverables

❖ Segmented customer dataset exported as CSV
❖ Optional segment-wise customer lists for marketing teams

📌 Conclusion

This project demonstrates how RFM analysis can convert raw transactional data into actionable customer insights.
The results highlight the importance of customer retention, personalized marketing, and targeted engagement strategies to maximize customer lifetime value and optimize marketing ROI.

🚀 Future Enhancements

★ Time-based RFM (monthly or quarterly snapshots)
★ Integration with clustering models (K-Means)
★ Campaign impact and churn prediction analysis
★ Dashboard development (Power BI / Tableau)

☁️ Execution Environment

The entire project was developed and executed using **Google Colab**, ensuring reproducibility and cloud-based accessibility.
