**Project Documentation:**

**Customer Segmentation Using RFM Analysis**

*(An Interactive Data-Driven Dashboard for*

*Targeted Marketing Strategies)*

**Introduction**

Customer-centric strategies are essential in today’s competitive marketplace. One of the foundational pillars of successful marketing and customer relationship management is the ability to understand purchasing behaviors—such as how frequently customers buy, how much they spend, and how recently they interacted with the business.

**RFM Analysis**, which stands for **Recency**, **Frequency**, and **Monetary** analysis, is a proven statistical method to evaluate customer value. Recency measures how recently a customer has purchased, Frequency captures how often they purchase, and Monetary denotes the total value of their purchases.

This project presents a full-featured, interactive web application built using **Python** and **Streamlit** that allows businesses to:

* Upload their transactional datasets.
* Automatically compute RFM metrics.
* Segment customers based on behavioral patterns.
* Visualize insights through dynamic charts and maps.
* Generate tailored marketing actions.
* Export results for further use.

*A screenshot of a computer

AI-generated content may be incorrect.Homepage Section of the Dashboard*

**System Objectives**

The system is designed with the following goals:

* Automate the RFM scoring process for customer data.
* Group customers into insightful behavioral segments.
* Provide an intuitive dashboard with visualizations and summaries.
* Recommend marketing strategies for each customer segment.
* Allow users to download full results and summaries for campaign planning.

This tool can be applied across retail, e-commerce, subscription services, and any transaction-based domain.

**Dataset Format Requirements**

To ensure proper functionality, the system requires the input dataset (CSV format) to include the following columns:

* InvoiceNo – Unique ID for each transaction.
* InvoiceDate – Date and time of purchase.
* CustomerID – Unique identifier for each customer.
* Quantity – Number of items purchased in the invoice.
* UnitPrice – Price per unit of product.

An optional column:

* Country – Used to visualize geographical segment distribution.

The application automatically validates these fields upon upload. If the user does not provide a dataset, the system loads a default example file.

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

*File Upload Sidebar with Info on Required Fields*

**Data Preprocessing**

Upon uploading, the application performs several data cleaning steps to ensure consistency and accuracy:

* Removes records without CustomerID.
* Converts InvoiceDate to datetime format.
* Computes a new field TotalPrice as Quantity × UnitPrice.
* Ensures all customer IDs are integer type for proper aggregation.

These steps prepare the data for accurate grouping and scoring.

A screenshot of a computer

AI-generated content may be incorrect.

*Data Preprocessing Status on App UI*

**RFM Metric Computation**

The cleaned dataset is then grouped by CustomerID to compute:

* **Recency**: Days since last purchase (relative to the latest invoice date).
* **Frequency**: Number of distinct invoices.
* **Monetary**: Total amount spent across all transactions.

These metrics provide a behavioral fingerprint of each customer, which is crucial for segmentation.

**RFM Scoring and Customer Segmentation**

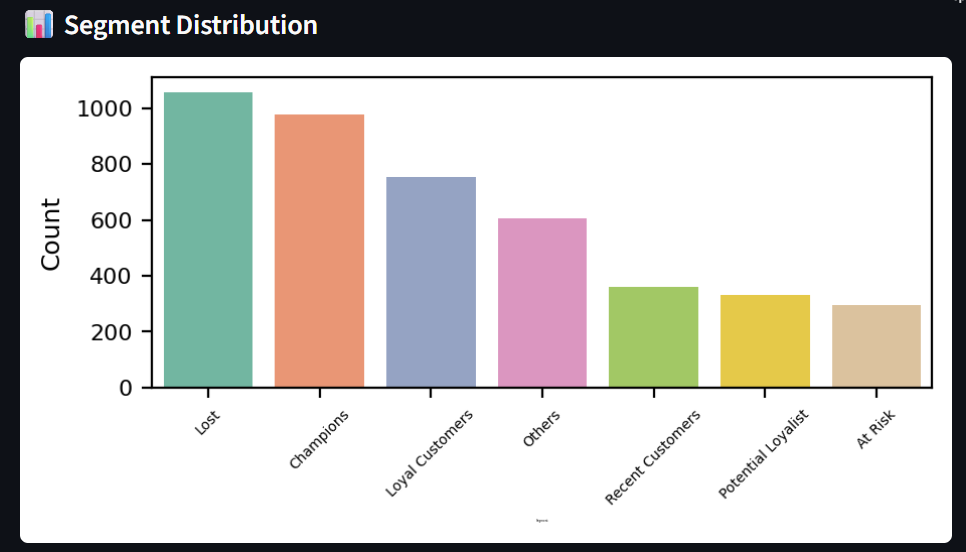
Each RFM metric is divided into **five quantile-based scores**:

* Higher R\_Score means more recent activity.
* Higher F\_Score indicates frequent purchases.
* Higher M\_Score signifies higher spending.

These scores are combined into a three-digit RFM\_Score (e.g., 543, 321), representing the customer's overall behavior.

Based on this score, customers are segmented into business-driven categories such as:

* **Champions** – Recently purchased, frequent, and high spenders.
* **Loyal Customers** – Frequent buyers with solid spending history.
* **Potential Loyalists** – Recent buyers with potential to be retained.
* **Recent Customers** – New customers with few transactions.
* **At Risk** – Former regulars who haven’t purchased recently.
* **Lost** – Inactive and low-value customers.
* **Others** – Customers who don’t fit other clear patterns.



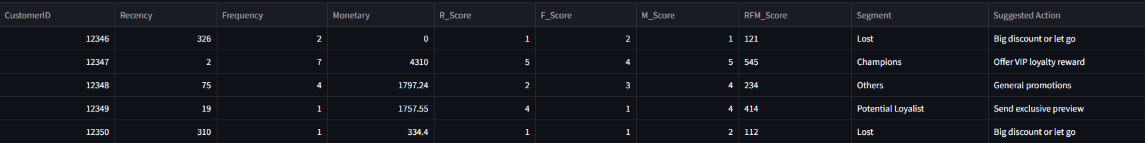
*Segment Distribution Bar Chart*

**Marketing Strategy Recommendations**

Each customer segment is mapped to a **suggested marketing action**:

* *Champions*: Reward with loyalty programs.
* *Loyal Customers*: Upsell premium products.
* *At Risk*: Target with win-back campaigns.
* *Lost*: Consider deep discounts or re-engagement offers.
* *Potential Loyalists*: Share exclusive previews or offers.

These recommendations are provided alongside each customer's segment label.

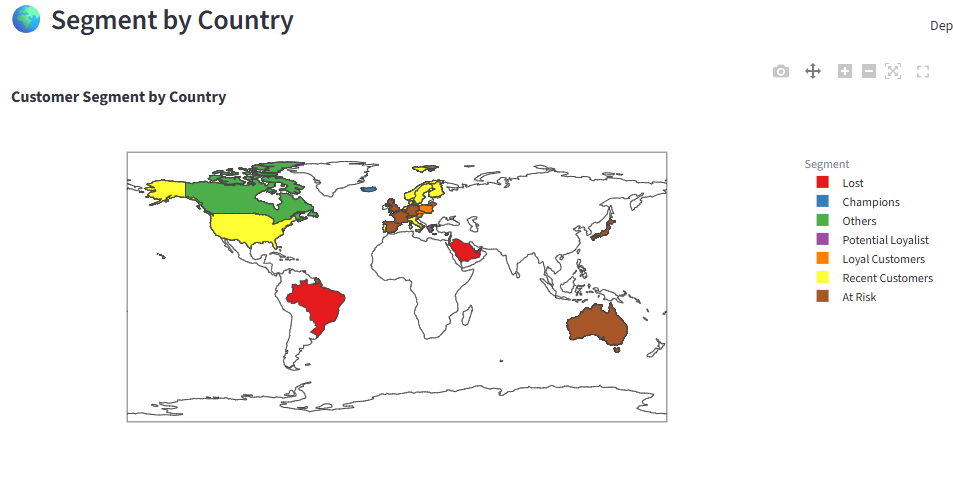
*RFM Table with Suggested Actions*

**Dashboard Visualizations and Analytics**

The dashboard includes a rich set of interactive charts that offer valuable insights into the RFM analysis:

* **Segment Distribution Chart**: Shows the number of customers in each segment.
* **Segment Summary Table**: Displays the average Recency, Frequency, and Monetary scores per segment.
* **Choropleth Map** (if Country is provided): Displays segment spread by region.
* **Scatter Plot (Monetary vs Frequency)**: Helps identify high-value repeat customers.
* **Box Plot (Recency by Segment)**: Highlights segment-wise recency trends.

These visualizations enable deeper exploration and understanding of customer behavior.



*Choropleth Map*

A screen shot of a graph

AI-generated content may be incorrect.

*A graph with different colored squares

AI-generated content may be incorrect.*  
*Scatter Plot and Box Plot Visuals*

**Export and Reporting Features**

The left sidebar of the dashboard includes convenient download options:

* **Full RFM Table** – Contains all customers, RFM scores, segments, and suggested actions.
* **Segment Summary Table** – Aggregated metrics by segment group.

These files can be shared with marketing or analytics teams for offline use.

A screenshot of a computer

AI-generated content may be incorrect.

*Sidebar Export Options*

**Error Handling and Flexibility**

The application is designed to be robust and user-friendly:

* Automatically displays an error if required columns are missing.
* Loads a fallback dataset if no CSV is provided.
* Skips optional plots if the Country column is absent.

These design choices make the dashboard usable in diverse situations with minimal setup.

**Conclusion**

This project demonstrates a complete end-to-end implementation of a customer segmentation strategy using RFM analysis. The system combines robust data processing, intuitive segmentation logic, visually compelling analytics, and actionable insights—all packaged in a user-friendly dashboard.

By enabling businesses to understand, analyze, and engage their customers effectively, this tool can enhance retention, maximize revenue opportunities, and support intelligent decision-making. Its adaptability makes it suitable for businesses of any scale that depend on customer transactions.