# Project Proposal: Customer Segmentation Using RFM Analysis

## 1. Project Title

Customer Segmentation and Marketing Insights with RFM Analysis on Online Retail Data

## 2. Introduction

Retail companies generate large volumes of sales data daily. However, raw transaction data is not directly useful unless transformed into actionable insights. Customer segmentation helps businesses understand their customers’ purchasing patterns and tailor marketing campaigns accordingly.  
  
One of the most effective segmentation techniques is RFM Analysis, which evaluates customers on three dimensions:  
- Recency (R): How recently a customer made a purchase.  
- Frequency (F): How often they purchased.  
- Monetary (M): How much they spent.  
  
This project will apply RFM analysis on the UCI Online Retail dataset, which contains transactions from a UK-based online store between 2010 and 2011. The project aims to identify customer segments such as Champions, Loyal Customers, At Risk, and Lost Customers, and provide marketing recommendations for each group.

## 3. Objectives

- To clean and preprocess the Online Retail dataset.  
- To calculate Recency, Frequency, and Monetary (RFM) values for each customer.  
- To assign RFM scores and segment customers into meaningful groups.  
- To provide marketing recommendations for each segment.  
- To visualize customer distribution using charts and heatmaps.

## 4. Research Questions

Customer Behavior:  
1. Who are the most valuable (high-spending and recent) customers in the dataset?  
2. How frequently do customers return for repeat purchases?  
3. What proportion of customers are inactive or lost?  
  
Segmentation Insights:  
4. What distinct customer groups can be identified using RFM analysis?  
5. How do Recency, Frequency, and Monetary values vary across different customer segments?  
6. How does rule-based segmentation compare with clustering-based segmentation (e.g., K-Means)?  
  
Marketing Strategy:  
7. What marketing strategies should be applied to Champions, Loyal Customers, At Risk, and Lost Customers?  
8. How can personalized campaigns increase customer retention and spending?  
9. What is the potential impact of re-engaging inactive customers on overall revenue?  
  
Business Value:  
10. How can RFM analysis improve decision-making for customer relationship management (CRM)?  
11. Can RFM metrics be used to predict Customer Lifetime Value (CLV)?  
12. How does customer segmentation contribute to better resource allocation in marketing campaigns?

## 5. Methodology

a. Data Preparation:  
- Dataset: Online Retail Dataset (UCI Machine Learning Repository)  
- Clean missing values, duplicates, and invalid transactions.  
- Create Total Price = Quantity × Unit Price.  
  
b. Feature Engineering:  
- Recency: Days since the customer’s last purchase.  
- Frequency: Number of transactions (Invoice count).  
- Monetary: Total spending of each customer.  
  
c. RFM Scoring:  
- Rank customers into 1–5 scores for each metric using quantile ranking.  
- Combine into a single RFM Score.  
  
d. Customer Segmentation:  
- Rule-based segmentation (Champions, Loyal, Potential Loyalist, At Risk, Lost, etc.).  
- (Optional Advanced) Apply Clustering (K-Means) for data-driven segmentation.  
  
e. Visualization:  
- Bar charts for customer count per segment.  
- Heatmaps for Recency vs Frequency vs Monetary values.  
- Boxplots for spending patterns across groups.  
  
f. Insights & Marketing Strategy:  
- Recommend personalized marketing strategies for each customer segment.

## 6. Tools and Technologies

- Python: Pandas, NumPy (data cleaning & analysis)  
- Seaborn, Matplotlib: Visualization  
- Scikit-learn: Optional clustering (K-Means)  
- Excel / Power BI / Tableau: Optional dashboards

## 7. Expected Outcomes

- Cleaned dataset with RFM metrics per customer.  
- Segmented customer base with clear categories.  
- Visualizations that highlight customer behavior.  
- Marketing recommendations to improve retention and revenue.  
- A complete project report with findings.

## 8. Deliverables

- Python Jupyter Notebook with analysis.  
- Visualization outputs (charts, heatmaps).  
- Final Project Report (PDF/Word).  
- (Optional) GitHub repository with README and code.

## 9. Timeline (Tentative)

- Week 1: Data collection & cleaning.  
- Week 2: Feature engineering (RFM metrics).  
- Week 3: Scoring & segmentation.  
- Week 4: Visualization, insights & report writing.

## 10. Conclusion

This project will showcase how RFM analysis transforms raw sales data into actionable business insights. By identifying high-value, loyal, and at-risk customers, businesses can design targeted marketing strategies to boost sales, retention, and overall customer satisfaction.