Customer Segmentation for Targeted Marketing Campaigns

---

1. Introduction

This project aims to segment customers based on their purchasing behavior using **RFM analysis** and **machine learning techniques**. The insights derived from this analysis will help the business develop targeted marketing strategies to improve customer engagement, retention, and revenue.

---

2. Data Exploration and Cleaning

Key Steps

- **Handled Missing Values:** Removed rows with missing `CustomerID` to ensure accurate customer-level analysis.
- **Removed Canceled Orders:** Filtered out transactions with negative `Quantity` or `UnitPrice` (indicating canceled orders or returns).
- **Created New Features:** Calculated `TotalSpend` as `Quantity * UnitPrice` for each transaction.

Insights

- The cleaned dataset contains **397,884 transactions** after removing invalid entries.
- The majority of transactions are from the **United Kingdom**, indicating it is the primary market.

---

3. Exploratory Data Analysis (EDA)

Key Visualizations

1. **Distribution of Customers by Country:**
    - The **United Kingdom** dominates, with over 350,000 transactions.
    - Other notable markets include **Germany**, **France**, and **Ireland**.
2. **Top 10 Most Popular Products:**
    - The most popular products are **WHITE HANGING HEART T-LIGHT HOLDER**, **REGENCY CAKESTAND 3 TIER**, and **JUMBO BAG RED RETROSPOT**.
    - These products are primarily related to **home decor** and **baking**, indicating strong customer interest in these categories.
3. **Distribution of Total Spend:**

- Most transactions fall in the **0–0–200** range, with a peak around **0–0–50**.
- High-value transactions (e.g., >$1,000) are rare but contribute significantly to revenue.

Insights

- The UK is the primary market, and marketing efforts should focus on retaining and engaging UK customers.
- Popular products like **WHITE HANGING HEART T-LIGHT HOLDER** can be highlighted in marketing campaigns to drive sales.

---

4. Customer Segmentation

RFM Analysis

Customers were segmented based on:

- **Recency (R):** How recently a customer made a purchase.
- **Frequency (F):** How often they purchase.
- **Monetary (M):** How much they spend.

Machine Learning for Segmentation

- Applied **K-Means clustering** to group customers into segments.
- Identified key clusters with the following characteristics:

Cluster	Recency (Days)	Frequency	Monetary ($)
0	43.81	3.69	1,358.77
1	248.47	1.55	480.42
2	1.50	135.83	58,381.12
3	15.72	22.30	13,533.99
4	7.67	42.83	190,863.46

Insights

- **Cluster 2 and 4:** These clusters represent **high-value customers** with very high monetary values and frequent purchases. They are likely **bulk buyers** or **wholesale customers**.
- **Cluster 0:** Represents **moderately active customers** with moderate recency, frequency, and monetary values. These customers have the potential to become loyal with targeted engagement.
- **Cluster 1:** Represents **inactive customers** with low frequency and monetary values. These customers are at risk of churn and require re-engagement strategies.

- **Cluster 3:** Represents **frequent buyers** with high monetary values. These customers are highly engaged and should be retained with loyalty programs.

---

5. Predictive Modeling

Churn Prediction

- Built a **Random Forest model** to predict customer churn based on RFM metrics.
- Achieved the following performance:

Metric	Precision	Recall	F1-Score	Support
Class 0	1.00	1.00	1.00	865
Class 1	1.00	1.00	1.00	437
Accuracy	1.00	-	-	1,302

Insights

- The model achieved **perfect accuracy** (1.00) on the test set, indicating excellent performance.
- **Recency** and **Frequency** are the most important factors in predicting churn.
- The model can be used to **identify at-risk customers** with high confidence and target them with retention campaigns.

---

6. Customer Lifetime Value (CLV)

Calculation

- Calculated CLV for each customer using the formula:
  CLV = (Average Purchase Value × Purchase Frequency × Customer Lifespan)

Segment-Level CLV

Segment	CLV ($)
151	15,957.67
152	2,075.93
153	996.87
251	13,228.04
351	34,670.01
451	95,147.06

Segment	CLV ($)
551	1,765,910.00
At Risk	333.79
Champions	10,870.32
Loyal Customers	13,520.72
Potential Loyalists	31,147.37

Insights

- The **551 Segment** has an exceptionally high CLV (**$1,765,910**), making it the most valuable customer group. These customers are likely **wholesale buyers** or **bulk purchasers**.
- The **451 Segment** also has a high CLV (**$95,147.06**), indicating significant revenue potential.
- The **At Risk Segment** has the lowest CLV (**$333.79**), highlighting the need for re-engagement strategies.
- **Potential Loyalists** and **Loyal Customers** have moderate CLV values, suggesting opportunities to increase their lifetime value through targeted campaigns.

---

7. Actionable Recommendations

Marketing Strategies by Segment

1. **High-Value Clusters (Cluster 2 and 4):**
    - **Strategy:** Focus on retaining these customers with exclusive perks and personalized service.
    - **Actions:**
        - Offer bulk purchase discounts or loyalty rewards.
        - Assign a dedicated account manager for personalized support.
2. **At-Risk Cluster (Cluster 1):**
    - **Strategy:** Re-engage these customers with win-back campaigns.
    - **Actions:**
        - Send personalized emails with special discounts or free shipping offers.
        - Highlight new products or features that might interest them.
3. **High-CLV Segments (551 and 451):**
    - **Strategy:** Maximize revenue from these segments by encouraging repeat purchases.
    - **Actions:**
        - Offer loyalty rewards or VIP perks.
        - Upsell or cross-sell complementary products.
4. **Potential Loyalists and Loyal Customers:**
    - **Strategy:** Build loyalty and increase CLV through targeted campaigns.
    - **Actions:**

    - Offer personalized discounts or rewards.
    - Encourage them to join a loyalty program.

Business Impact

- By targeting high-value segments and re-engaging at-risk customers, the company can increase revenue and reduce churn.

---

8. Conclusion

Summary

This project successfully segmented customers using **RFM analysis** and **machine learning techniques**. The insights were used to develop targeted marketing strategies that can improve customer engagement, retention, and revenue.

Next Steps

1. Implement the proposed marketing campaigns.
2. Monitor performance and refine strategies based on results.
3. Expand the analysis to include additional data (e.g., demographics, website behavior).

---

9. Appendix

Tools and Technologies

- **Programming Languages:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly

Dataset Source

- Online Retail Dataset

Code Repository

- GitHub Link