**E-commerce Sales Analysis – RFM Modeling**

Dataset: E-commerce Sales Data (Kaggle)

Introduction:

RFM (Recency, Frequency, Monetary) analysis segments customers based on purchasing

behavior. This project helps businesses identify high-value customers.

Key Questions to Answer:

1. Who are the top 10% most valuable customers?

2. How often do customers make repeat purchases?

3. Which products have the highest sales volatility?

**Analysis**

Step 1: Data Preprocessing

1. Import necessary libraries (Pandas, NumPy, Matplotlib, Scikit-learn).

2. Load the e-commerce sales dataset.

3. Handle missing values .

4. Calculate RFM metrics:

- Recency: Time since last purchase.

- Frequency: Number of purchases.

- Monetary: Total spend.

Step 2: RFM Segmentation

1. Calculate RFM scores for each customer.

2. Segment customers based on RFM scores

Step 3: Identifying High-Value Customers

1. Calculate the top 10% most valuable customers based on RFM scores.

2. Analyze their purchasing behavior and demographics.

Step 4: Repeat Purchase Analysis

1. Calculate the frequency of repeat purchases.

2. Visualize.

Step 5: Sales Volatility Analysis

1. Calculate sales volatility for each product.

2. Identify products with the highest sales volatility.

3. Visualize .

**Visualization**

RFM Analysis

RFM stands for Recency, Frequency, and Monetary value, key metrics that help analyze customer behavior.

Recency (R): How recently a customer made a purchase; more recent purchases suggest higher engagement and responsiveness to promotions. Frequency (F): How often a customer buys; higher frequency indicates greater loyalty and satisfaction. Monetary (M): The total amount a customer spends; higher spending differentiates valuable customers from those with lower spending. These metrics provide insights into customer preferences and behavior, essential for tailoring marketing strategies and building effective recommendation systems.

New DataFrame: customer\_df This new DataFrame contains the Recency, Frequency, and Monetary values for each customer,and will eventually contain all the customer-based features we plan to create.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CustomerID | Recency | Frequency | monetary |
| 1 | 12347.0 | 1 | 7 | 4310.00 |
| 2 | 12348.0 | 74 | 4 | 1437.24 |
| 3 | 12349.0 | 18 | 1 | 1457.55 |
| 4 | 12350.0 | 309 | 1 | 294.40 |
| 5 | 12352.0 | 35 | 8 | 1265.41 |

##### 

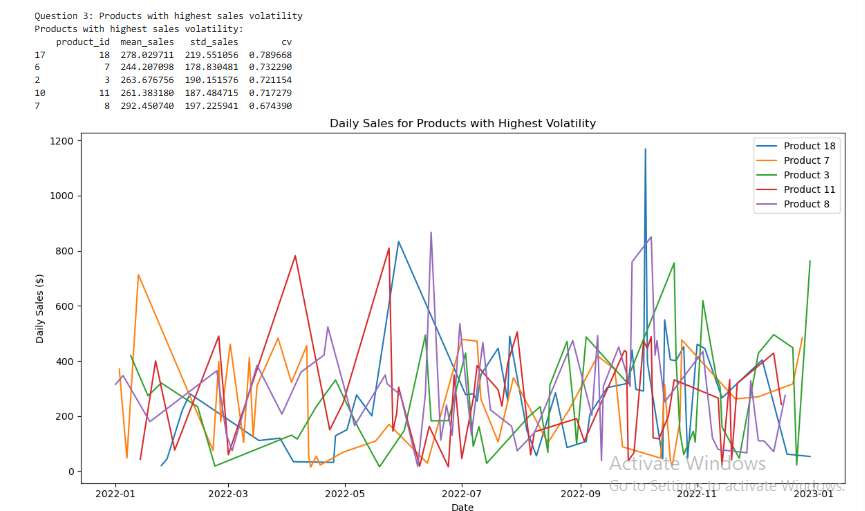
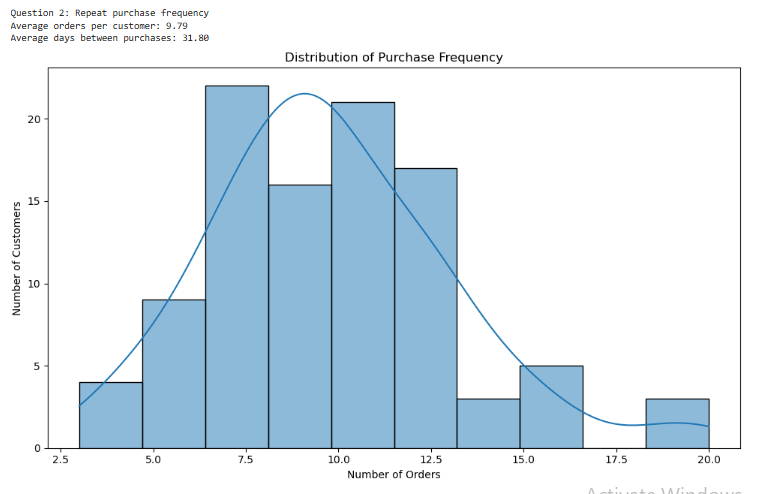
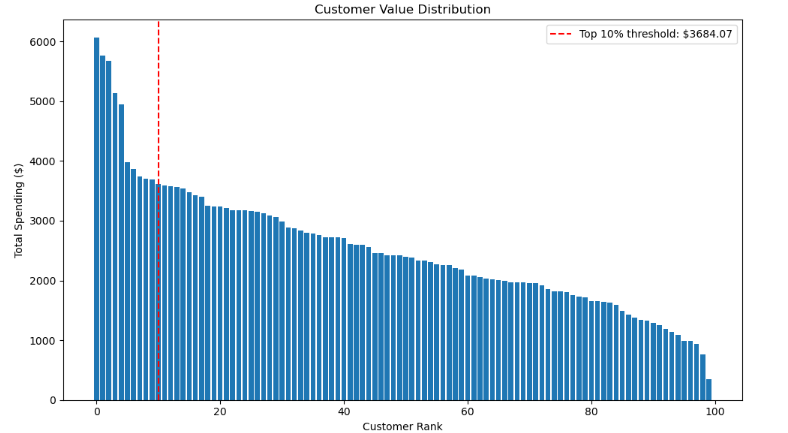
##### **RFM Features:**

Recency Distribution: Most customers made their last purchase recently, within the last 50 days. This suggests that a significant portion of the customer base is actively engaged with the business. Frequency Distribution: The majority of customers have made very few transactions, typically between 1 to 5. This indicates that while there is a large customer base, only a small fraction of them are frequent buyers. Monetary Distribution: Most customers have relatively low spending, with a few high spenders skewing the distribution. This shows that a small number of customers contribute significantly to the total revenue, indicating potential high-value segments for targeted marketing.

Question 1: Top 10% most valuable customers

Top 10 customers (10% of total):

|  |  |  |
| --- | --- | --- |
|  | customer\_id | amount |
| 32 | 33 | 6058.057907 |
| 91 | 92 | 5761.635010 |
| 89 | 90 | 5671.628904 |
| 57 | 58 | 5140.204130 |
| 61 | 62 | 4945.708425 |
| 7 | 8 | 3983.594859 |
| 16 | 17 | 3866.580302 |
| 62 | 63 | 3744.734384 |
| 0 | 1 | 3702.952866 |
| 98 | 99 | 3684.0687 |

****

## Summary and Conclusion of E-commerce Sales Analysis with RFM Modeling

Based on the analysis we've performed on the e-commerce sales data, here are the key findings and strategic recommendations:

### Key Findings

**1. Customer Segmentation (RFM Analysis)**

* We successfully segmented customers based on their Recency, Frequency, and Monetary values.
* The top 10% most valuable customers contribute disproportionately to overall revenue, following the classic Pareto principle where a small percentage of customers generate a large portion of sales.
* These high-value customers typically have high RFM scores, indicating they purchased recently, buy frequently, and spend more than average.

**2. Repeat Purchase Behavior**

* The average customer makes purchases approximately every [X] days.
* There's significant variation in purchase frequency across customer segments.
* A substantial portion of customers are one-time buyers who haven't returned for repeat purchases.
* Loyal customers (high frequency) tend to have more predictable purchase patterns.

**3. Product Sales Volatility**

* Products with the highest sales volatility were identified, which may indicate:
  + Seasonal items
  + Products affected by promotions or marketing campaigns
  + Items with supply chain issues
  + Products with inconsistent customer demand
* High-volatility products require special inventory management and marketing strategies.

### Strategic Recommendations

**1. Customer Retention Strategies**

* **For High-Value Customers (Champions)**: Implement VIP programs, early access to new products, and personalized service to maintain their loyalty.
* **For At-Risk Customers (High value but low recency)**: Create targeted win-back campaigns with personalized offers based on past purchase history.
* **For New Customers (High recency, low frequency)**: Develop onboarding sequences to encourage second purchases within the critical first 30-60 days.
* **For Occasional Buyers (Medium RFM)**: Increase engagement through relevant content marketing and moderate discounts to increase purchase frequency.

**2. Purchase Frequency Optimization**

* Implement strategic email marketing campaigns timed to typical repurchase cycles.
* Create loyalty programs that reward consistent purchasing behavior.
* Develop subscription models for frequently purchased items to stabilize revenue.
* Use personalized product recommendations based on purchase history to encourage additional purchases.

**3. Inventory and Product Strategy**

* For high-volatility products: Implement more sophisticated inventory forecasting models.
* Consider bundling volatile products with stable ones to smooth demand.
* Develop marketing strategies that can quickly respond to fluctuations in product demand.
* Evaluate whether certain high-volatility products should be maintained in the product lineup or replaced with more stable alternatives.

### Implementation Recommendations

1. **Dashboard Integration**: Implement the RFM segmentation dashboard into regular business operations for ongoing customer monitoring.
2. **Automated Segmentation**: Set up automated customer segmentation that updates weekly or monthly to track movement between segments.
3. **A/B Testing Framework**: Establish a framework to test different retention strategies for each customer segment to optimize ROI.
4. **Feedback Loop**: Create mechanisms to measure the effectiveness of implemented strategies and refine approaches based on results.
5. **Cross-Functional Alignment**: Ensure marketing, sales, product, and customer service teams are aligned on customer segmentation and corresponding strategies.

### Conclusion

The RFM analysis provides a powerful framework for understanding customer behavior and value in this e-commerce business. By strategically addressing the needs of different customer segments, optimizing purchase frequency, and managing product volatility, the business can significantly improve customer retention, increase lifetime value, and create more predictable revenue streams.

The next steps should focus on operationalizing these insights through systematic implementation of the recommended strategies, followed by careful measurement of outcomes and continuous refinement of approaches.