Data-Driven Customer Insights: A

Comprehensive Segmentation Analysis

IE6400 – Foundations Data Analytics Project Report 2

**GROUP NUMBER 10**

Gagan Yadav (+1) 617 858 4659

Dennis Jose (+1) 862 440 7362

Stephy Romichan (+1) 575 842 9980

**ABSTRACT:**

This project delves into e-commerce analytics using the "Online Retail" dataset from the UCI Machine Learning Repository, emphasizing customer segmentation through RFM (Recency, Frequency, Monetary) analysis. By analyzing transaction data spanning December 2010 to 2011, customers are categorized based on their purchasing recency, frequency, and spending patterns. The goal is to generate actionable insights to guide targeted marketing and customer engagement strategies. Additionally, exploratory data analysis is employed to identify deeper trends within customer segments, aiding effective decision-making and business growth. The findings result in distinct customer clusters with specific behaviors, facilitating the creation of tailored marketing strategies. Ultimately, the insights aim to improve customer retention, streamline marketing efforts, and foster business success.

**INTRODUCTION:**

This project explores e-commerce analytics using the "Online Retail" dataset from the UCI Machine Learning Repository. The dataset contains transaction records from December 2010 to 2011 for a UK-based online retail business specializing in unique, all-occasion gifts, catering primarily to wholesalers. It provides a comprehensive overview of customer interactions during this period.

The focus is on leveraging RFM (Recency, Frequency, Monetary) analysis, a popular framework for evaluating customer purchasing behavior. By analyzing when customers last made purchases, how frequently they buy, and their spending levels, we aim to create actionable customer segments. These segments will serve as the foundation for targeted marketing and personalized engagement strategies, with the ultimate goal of enhancing customer retention and driving better business outcomes.

Through an in-depth RFM analysis, the project seeks to uncover patterns in customer transactions, enabling data-driven decision-making and the optimization of marketing efforts. By dissecting the dataset, we aim to gain nuanced insights into customer behavior that will support effective business strategies for improving engagement and fostering growth.

In summary, this study combines rigorous RFM analysis with exploratory data analysis to generate valuable insights. These insights will inform data-driven strategies to enhance customer retention, refine product offerings, and drive business growth through more effective and tailored marketing efforts.

**OBJECTIVE:**  
The primary aim of this project is to conduct a detailed RFM analysis to uncover patterns in customer transactions. RFM segmentation, a widely used approach, categorizes customers based on their purchase recency, frequency, and monetary value. By assigning unique RFM scores to each customer, the analysis seeks to deliver actionable insights for enhancing targeted marketing efforts and strengthening customer retention strategies.

**DATASET:**  
The dataset provides an in-depth view of customer interactions, encompassing extensive transactional data. It serves as a vital resource for analyzing customer behavior, preferences, and engagement with the online retail platform, enabling a comprehensive understanding of user patterns.

**PROJECT SCOPE:**  
Beyond RFM analysis, the project includes exploratory data analysis (EDA) to identify patterns and trends within customer segments. These insights aim to guide marketing strategies, enhance customer experiences, and promote sustainable business growth.

**METHODOLOGY:**  
The project begins with data cleansing to ensure the dataset's accuracy and reliability. An RFM analysis follows, categorizing customers by recency, frequency, and monetary value metrics. This is complemented by EDA to reveal deeper patterns and trends, providing a richer perspective on customer behavior.

**ANTICIPATED RESULTS:**  
The project aims to identify distinct customer segments, each with unique behaviors and preferences. These insights will enable businesses to design tailored marketing strategies, refine product offerings, and improve customer engagement. Ultimately, the findings will foster stronger customer relationships, drive business growth, and enhance the customer experience in the competitive e-commerce sector.

**DATA PROCESSING:**  
The initial phase involves refining the dataset to ensure high-quality analysis. Covering transactions from December 1, 2010, to December 9, 2011, the dataset originates from a UK-based non-store online retailer. A thorough review of this data ensures a robust foundation for analysis and actionable insights.

**DATASET OVERVIEW AND INITIAL OBSERVATIONS:**

The dataset offers a detailed snapshot of 541,909 transactions, captured across eight key columns: *InvoiceNo*, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *UnitPrice*, *CustomerID*, and *Country*. Each column provides critical insights into customer transactions:

* **InvoiceNo:** Unique identifiers for each transaction.
* **StockCode:** Unique identifiers for individual products.
* **Description:** Textual details of the products.
* **Quantity:** The number of units purchased per transaction.
* **InvoiceDate:** The date and time of each transaction.
* **UnitPrice:** The monetary value per unit of a product.
* **CustomerID:** Unique identifiers for customers.
* **Country:** Indicates the geographic location of customers.

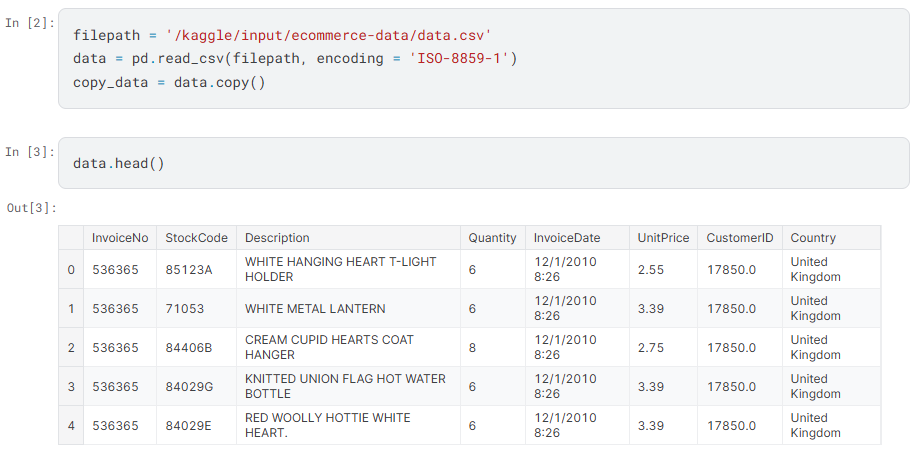
Upon reviewing data types:

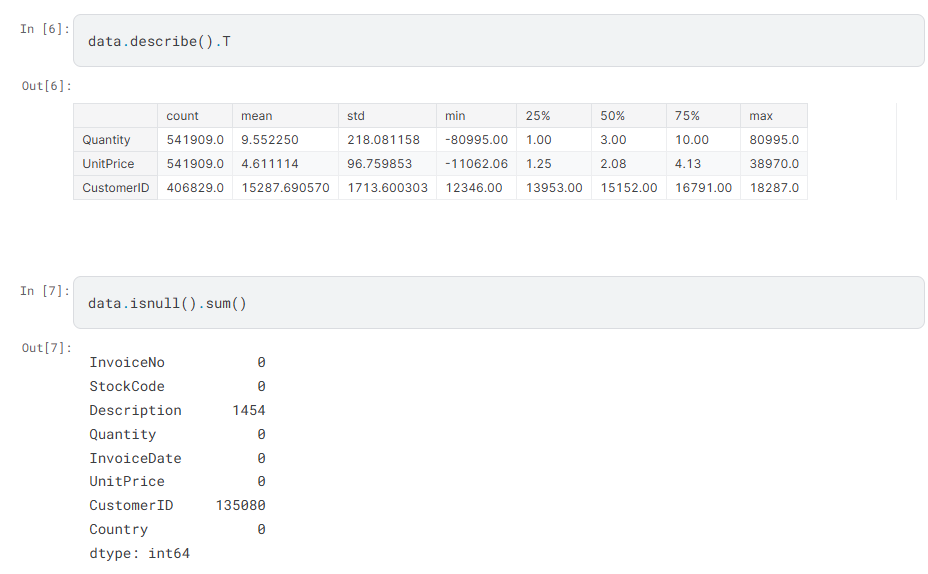
* Two columns, likely *UnitPrice* and a monetary metric, are of type *float64*.
* One column, *Quantity*, is an *int64* type, representing whole numbers.
* Five columns, including text and categorical data such as product descriptions and customer identifiers, are of type *object*.

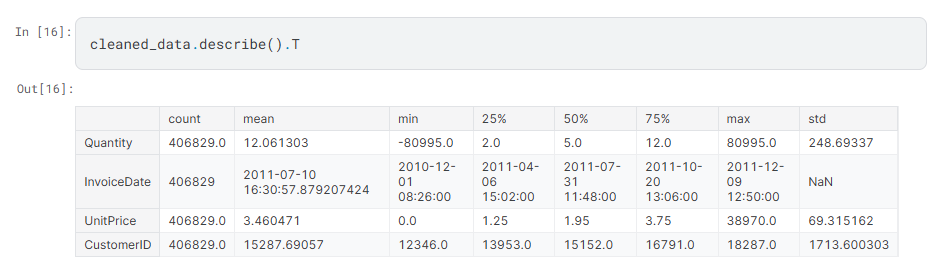
This dataset structure provides a solid foundation for analyzing customer behavior, transactions, and geographic trends.

**DATA PREPROCESSING:**

The CustomerID and InvoiceDate columns are utilized to calculate recency. The most recent transaction date in the dataset is determined, and the time difference between this date and each customer's transaction date is computed. This difference is then recorded in a new column named Diff.

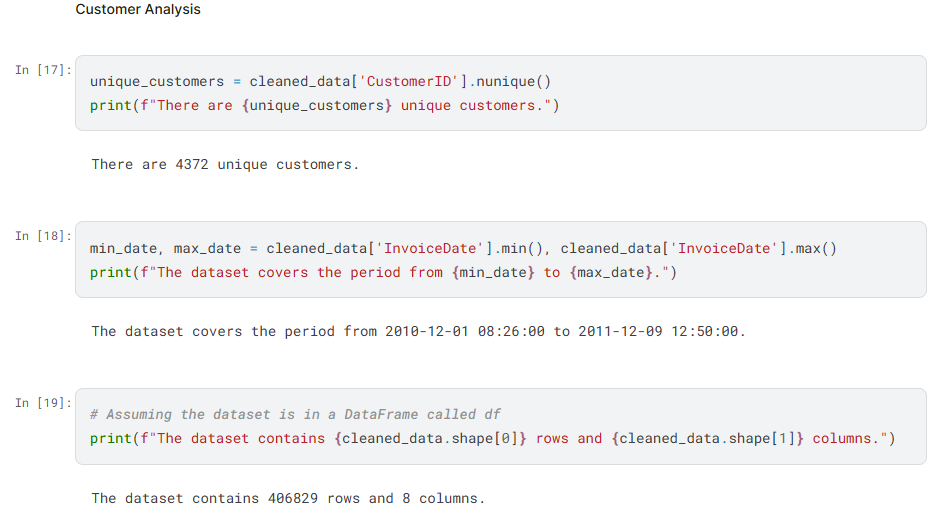


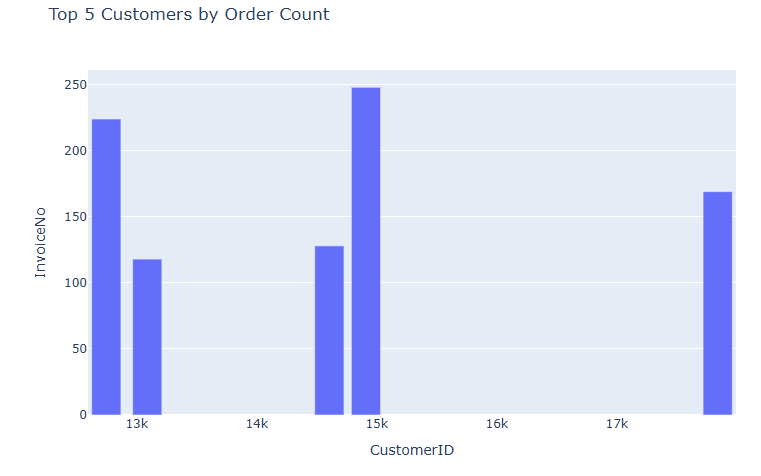
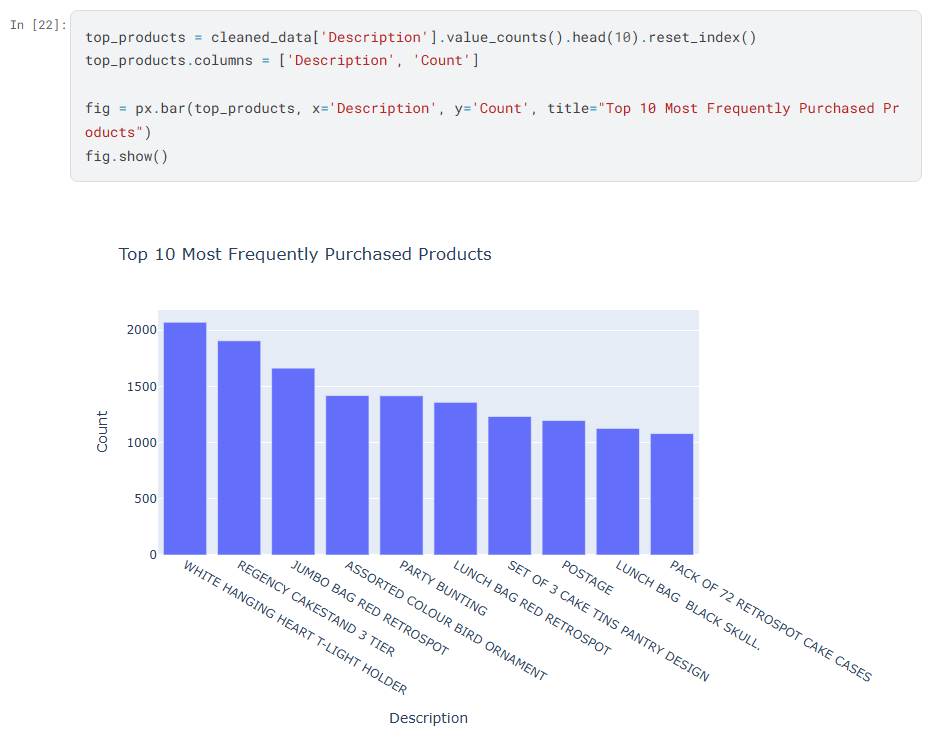
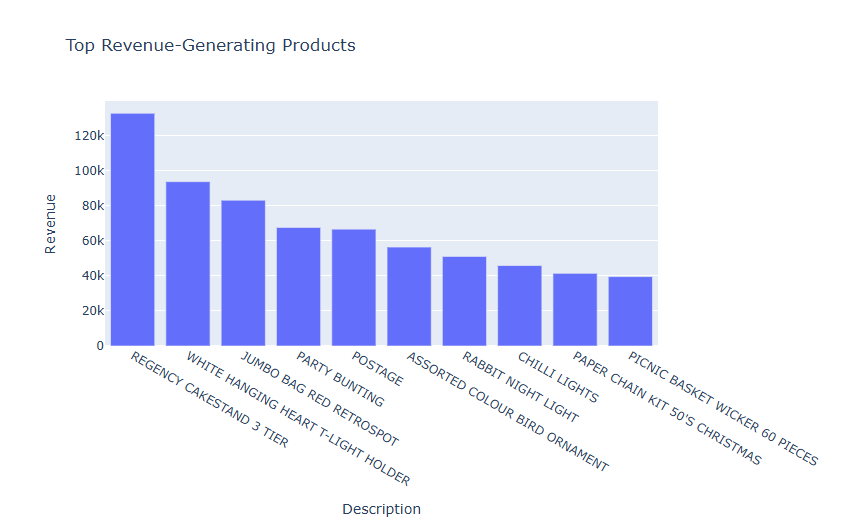
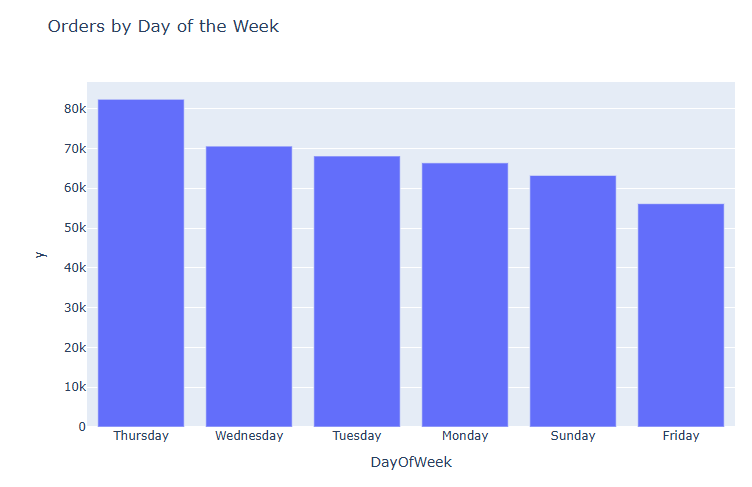
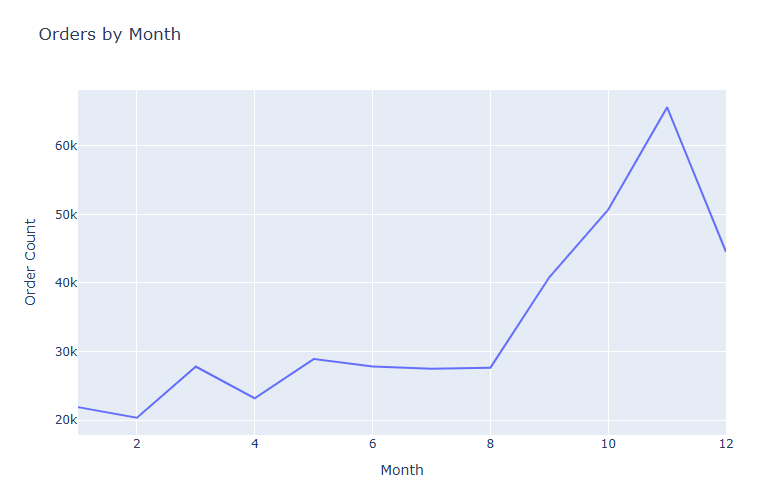
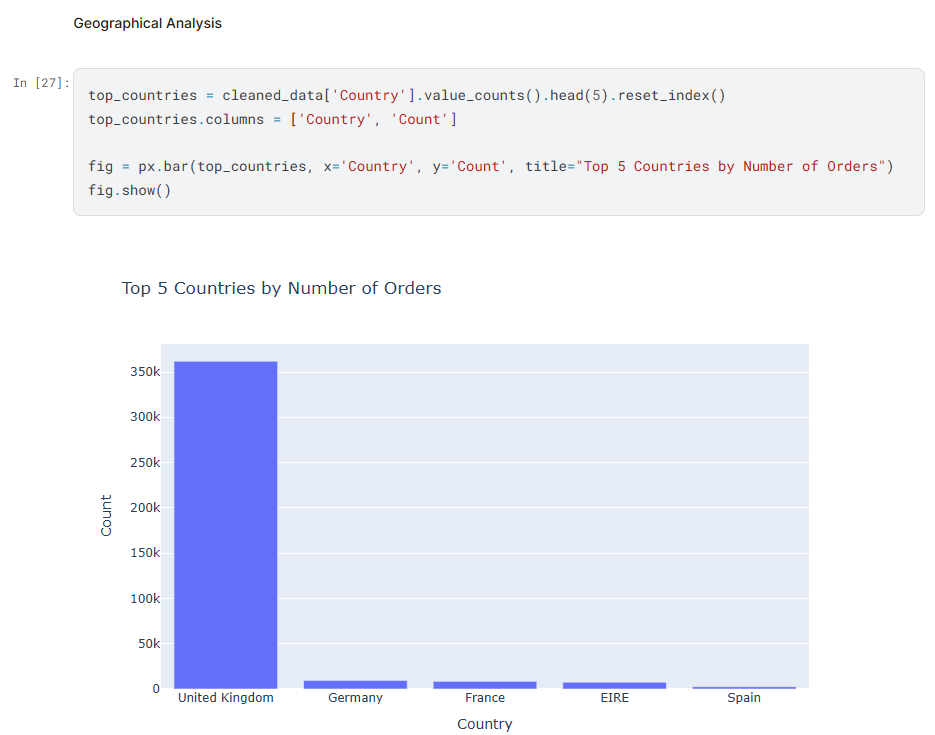
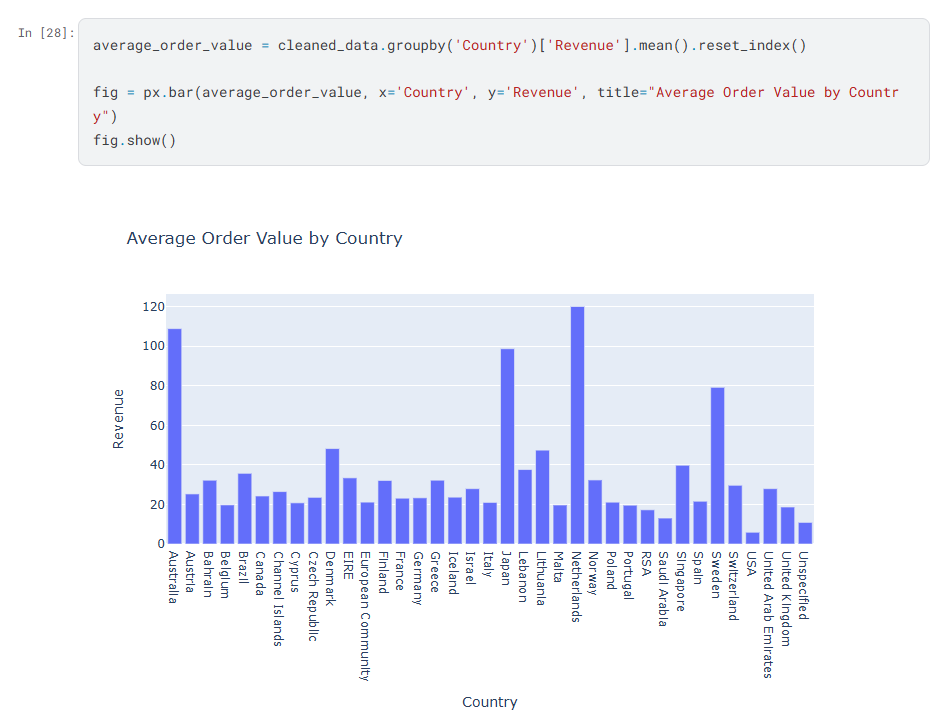
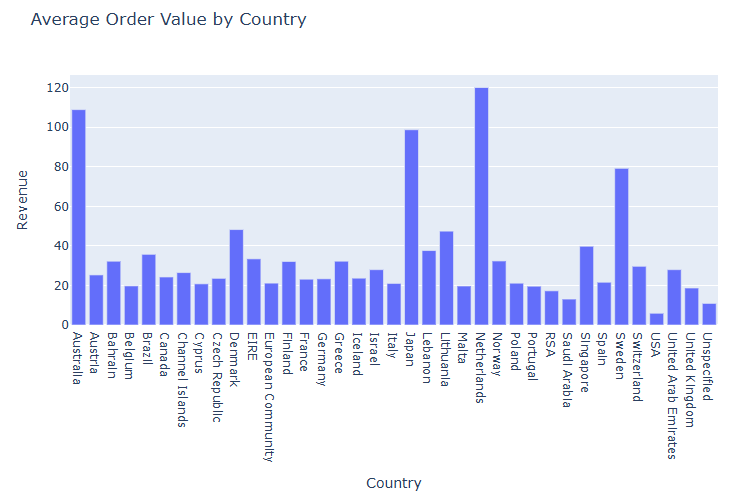


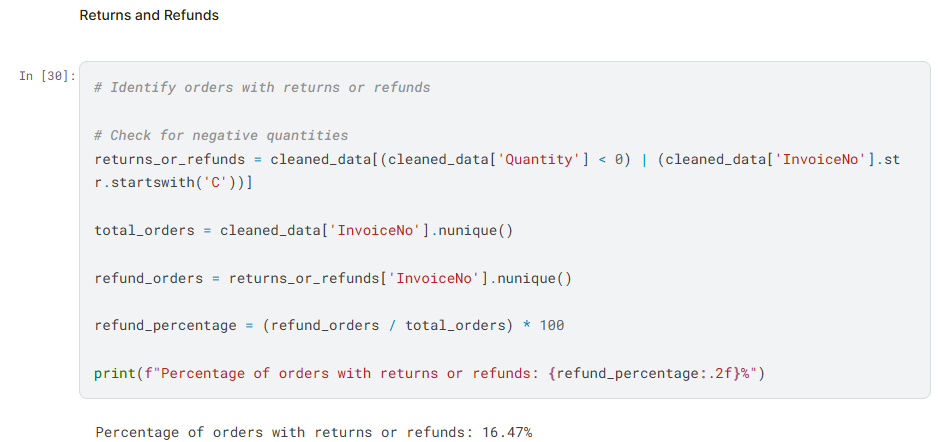
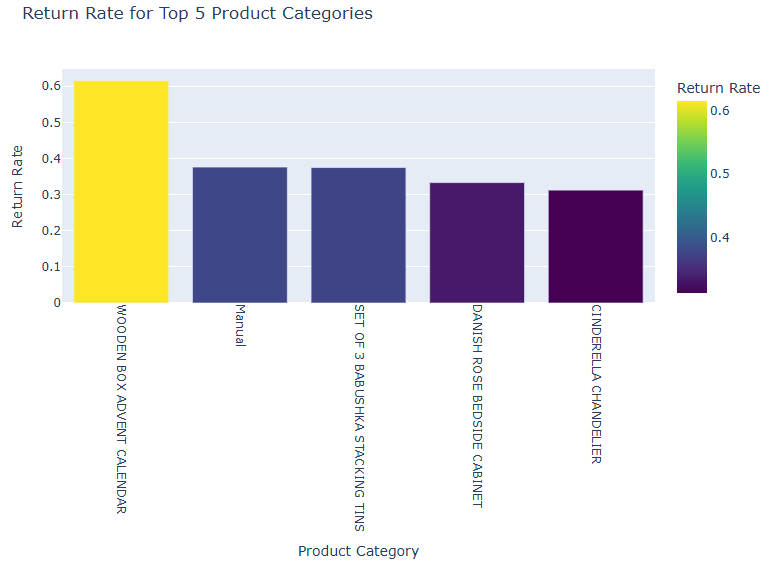
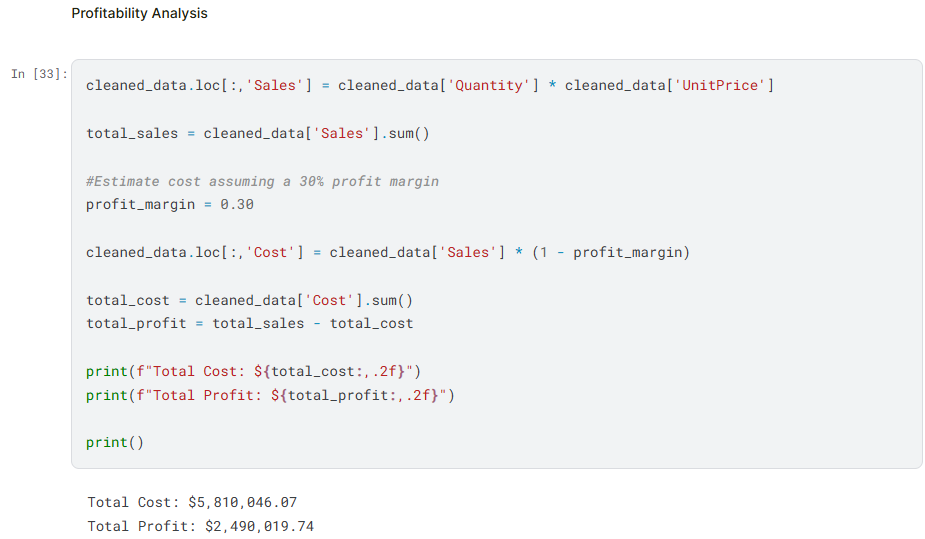


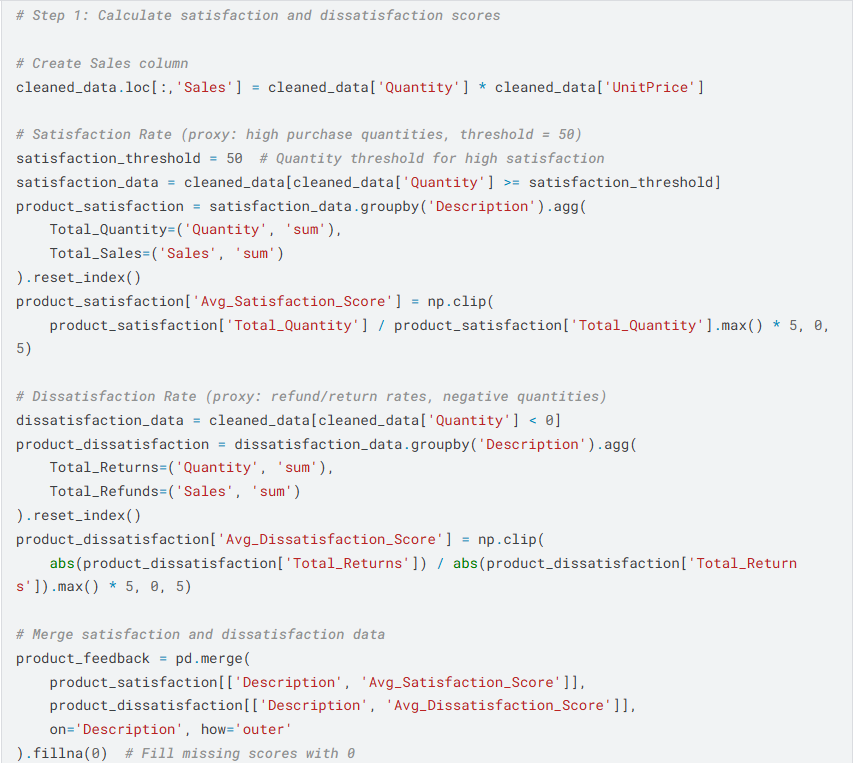
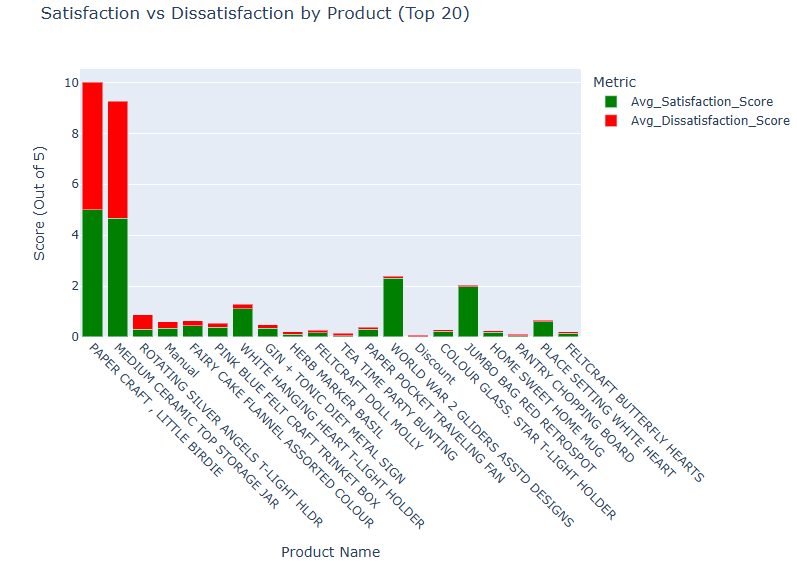
**EXPLORATORY DATA ANALYSIS:**

Exploratory data visualization involves creating charts and graphs to uncover patterns, trends, and insights within the dataset. It helps identify outliers, correlations, and distribution patterns across key variables, providing a clearer understanding of customer behavior and transaction dynamics.



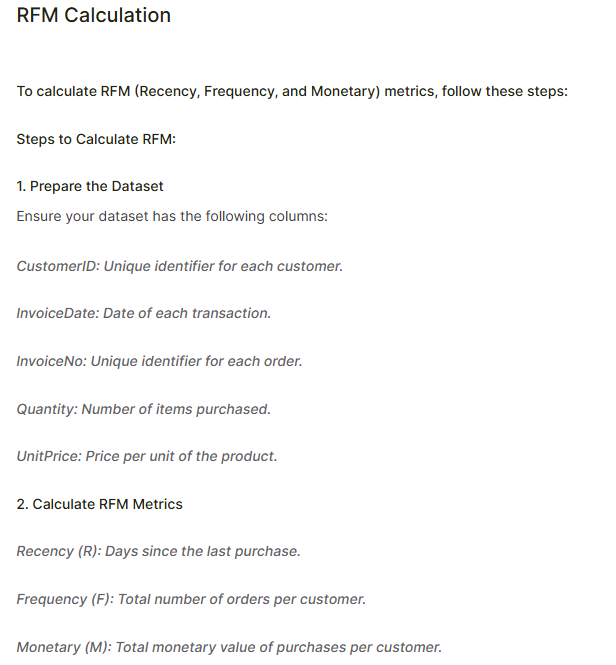
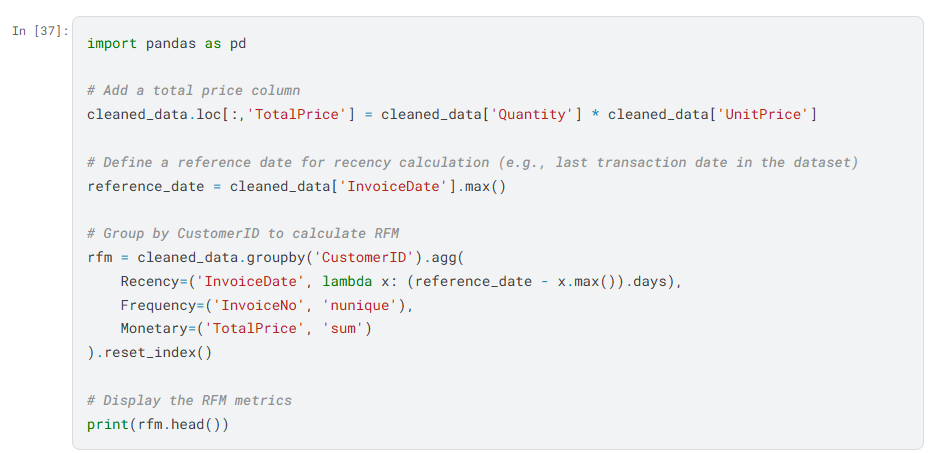
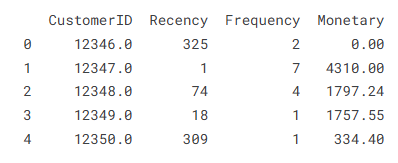
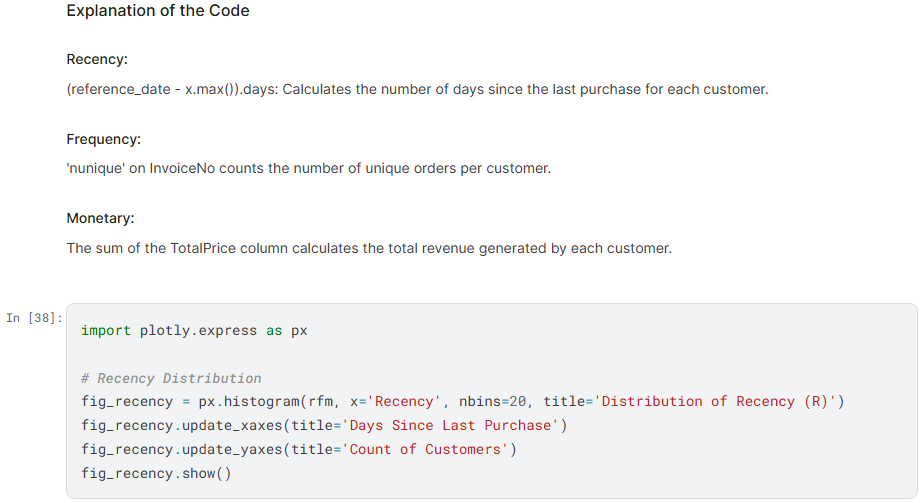
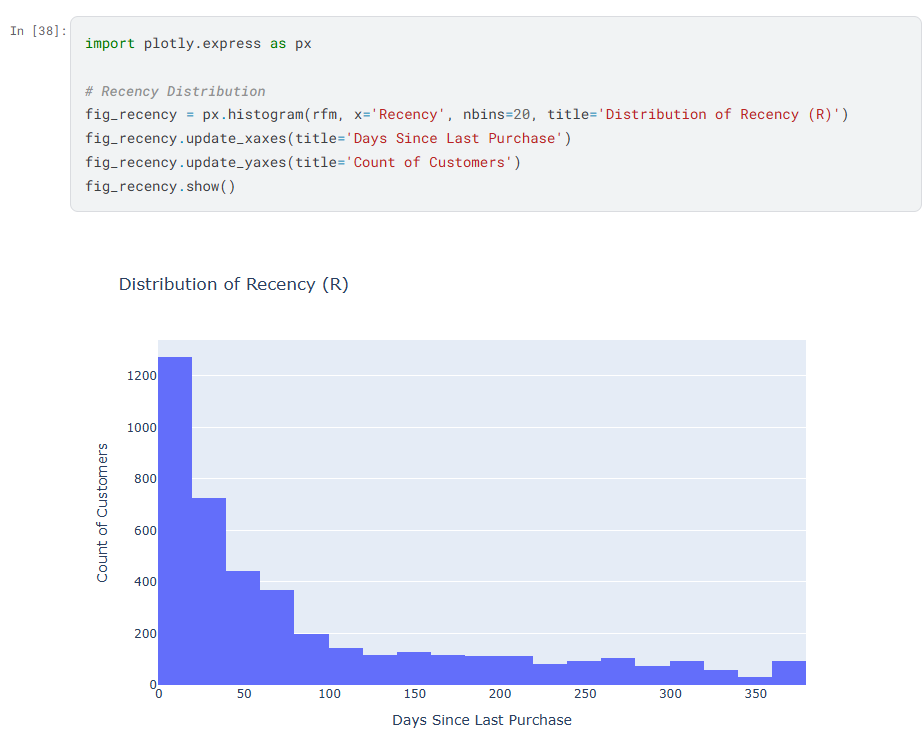
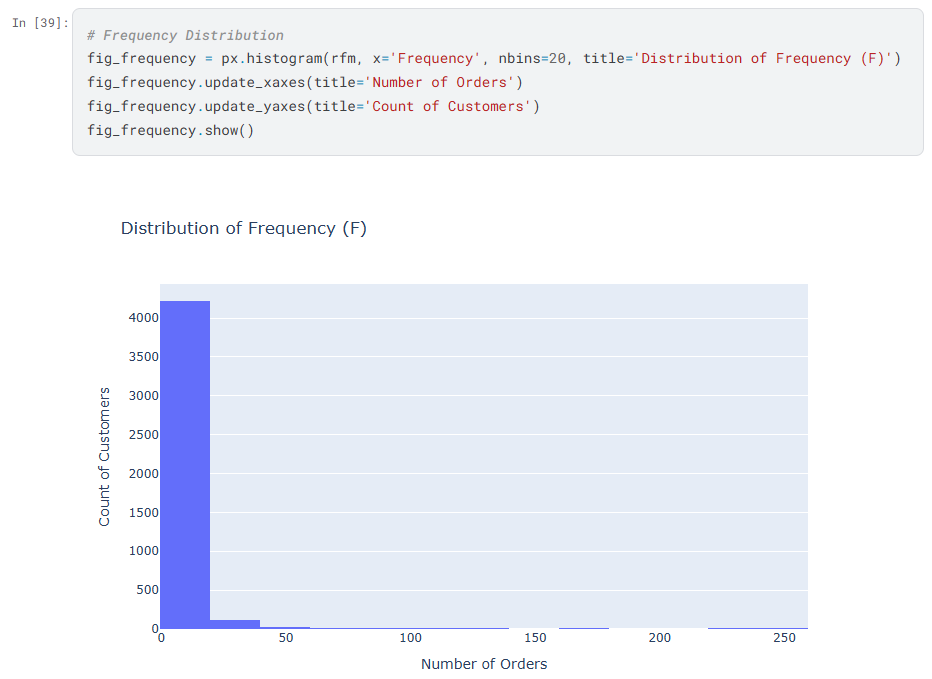
    

**RFM CALCULATION:**

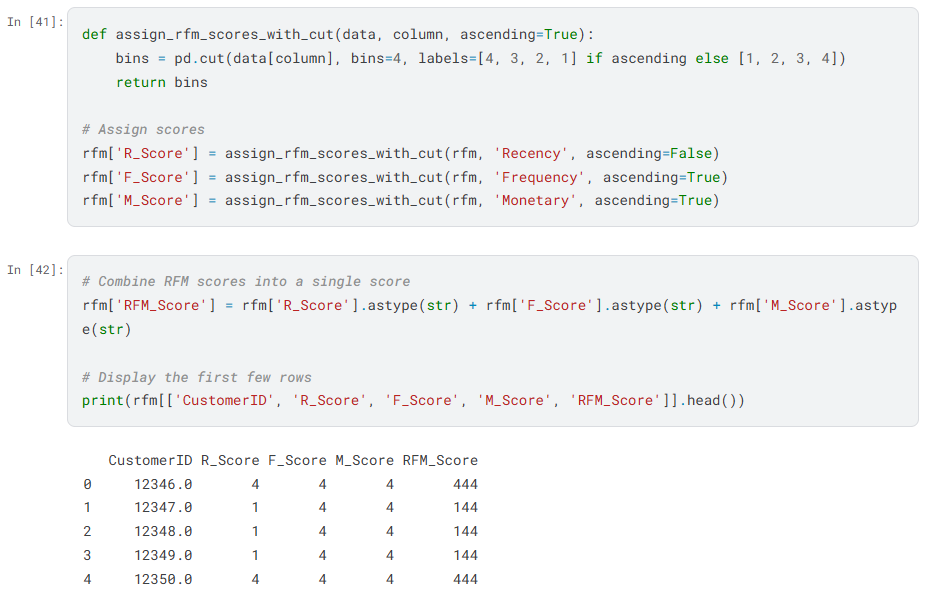
RFM (Recency, Frequency, Monetary) calculation involves assigning scores to customers based on their purchasing behavior:

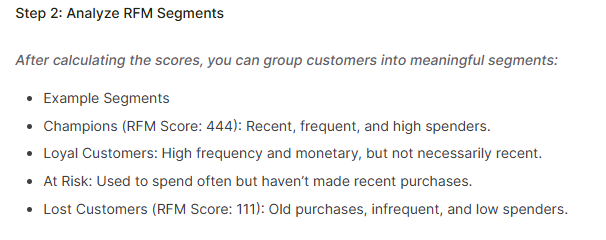
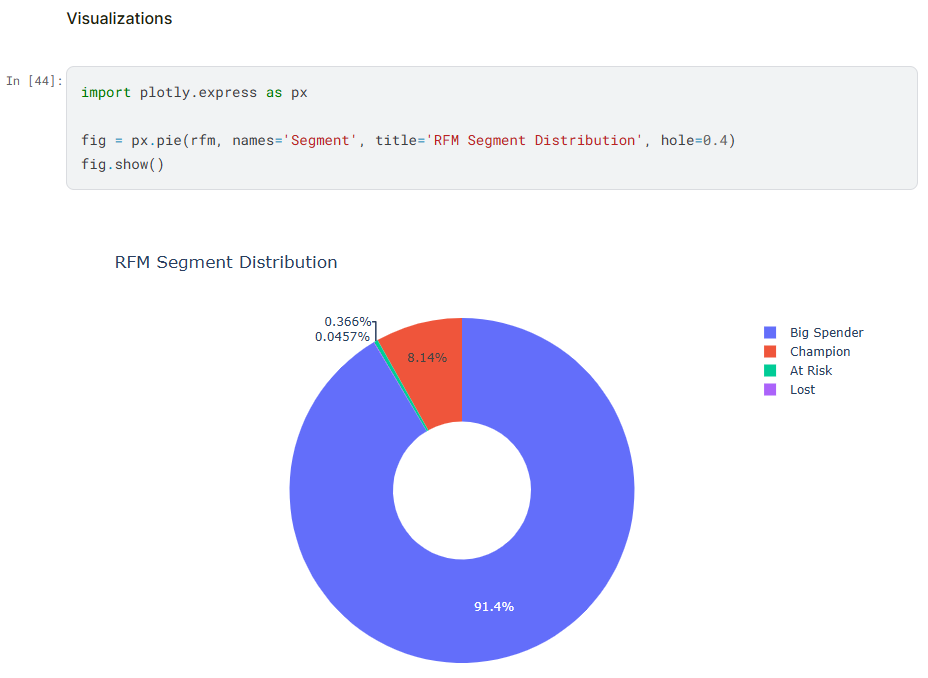
* **Recency:** Calculated as the time elapsed since a customer's last purchase.
* **Frequency:** Represents the total number of purchases made by each customer.
* **Monetary:** Measures the total spending of each customer.

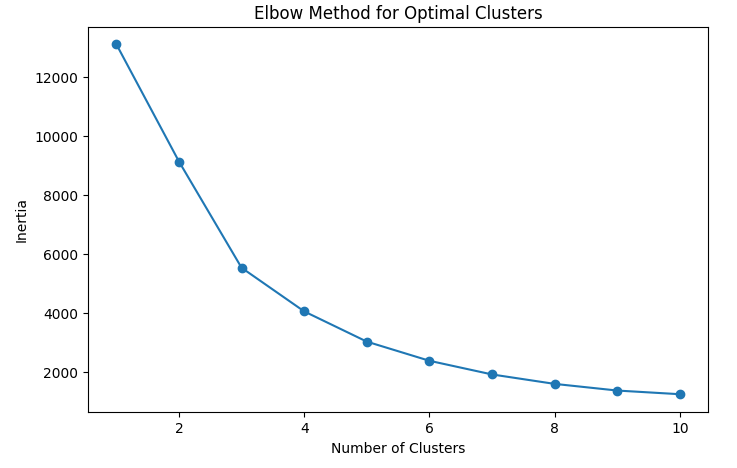
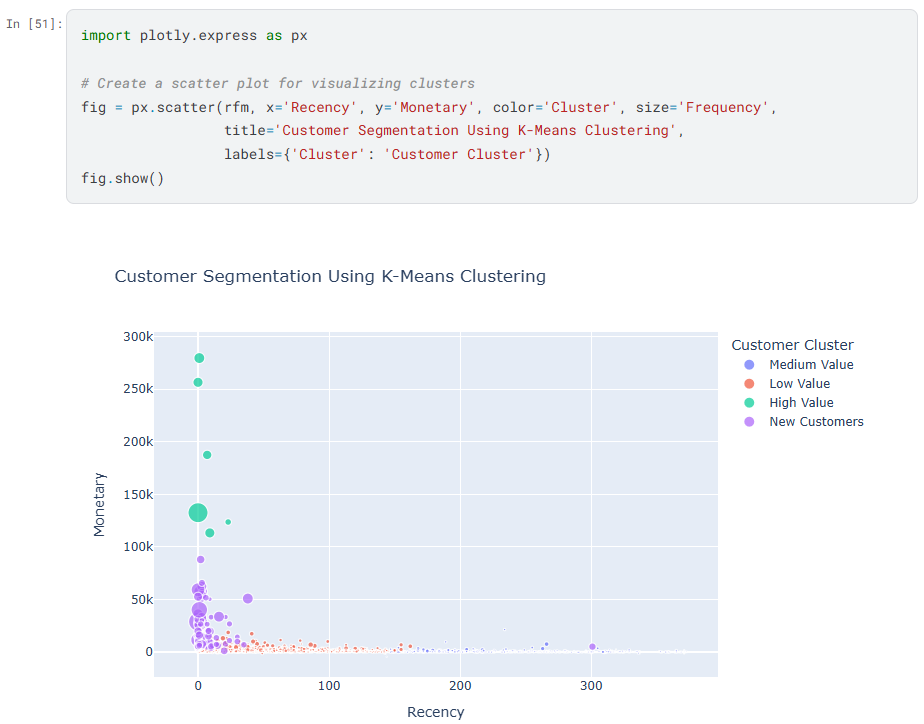
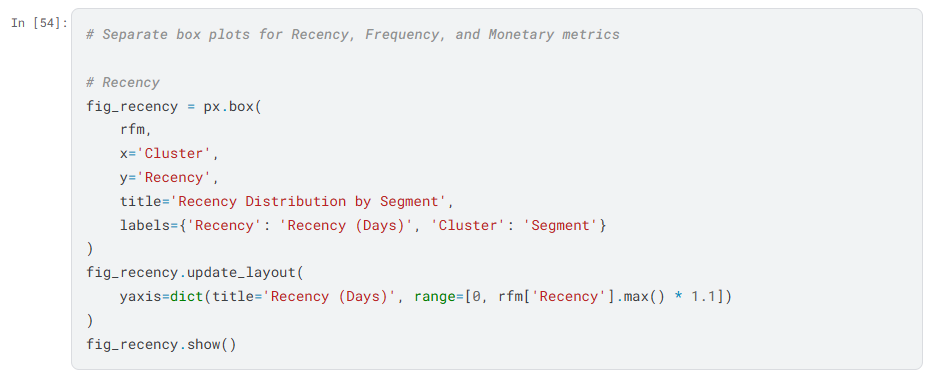
These metrics are computed from the dataset and used to segment customers, enabling targeted marketing strategies and improved customer retention.

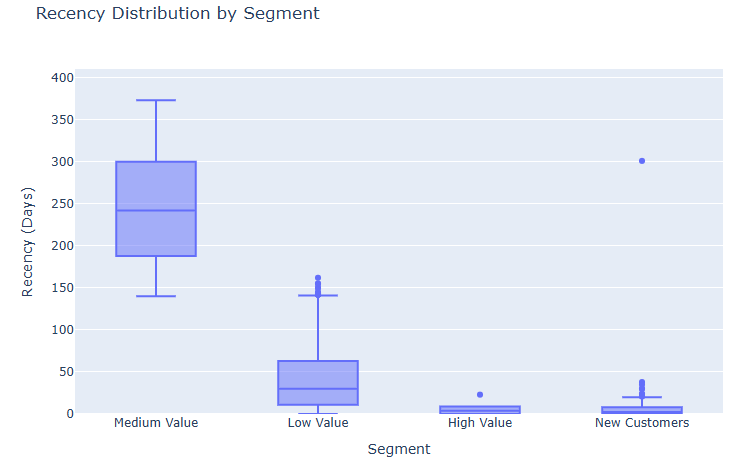
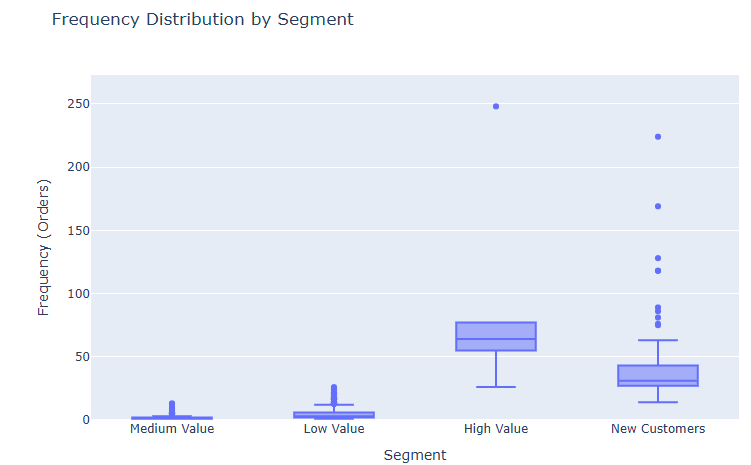
     

**RFM SEGMENTATION:**

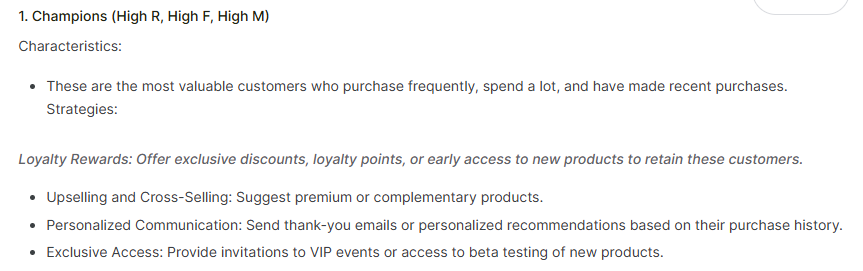
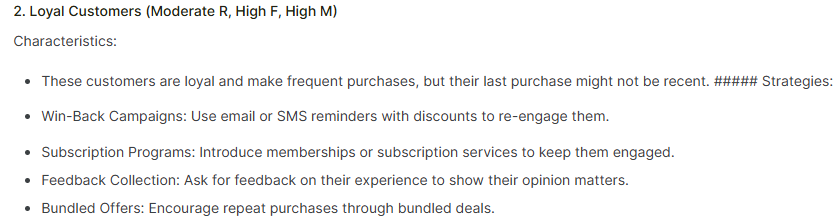
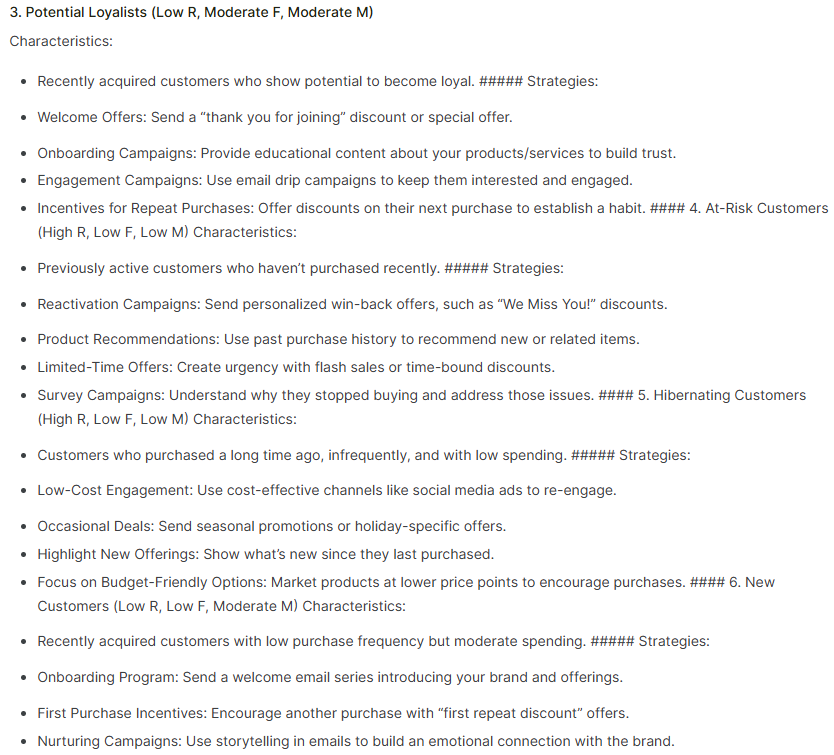
RFM segmentation groups customers based on their purchasing behavior, categorized by recency, frequency, and monetary value. These segments help identify patterns such as high-value customers, frequent buyers, or those at risk of inactivity. Businesses can use these insights to develop targeted marketing strategies and enhance customer engagement.

**CUSTOMER SEGMENTATION:**      

**SEGMENT PROFILING:**  

**MARKETING RECOMMENDATION:**

**CONCLUSION:**

RFM (Recency, Frequency, Monetary) analysis is a practical and insightful way to understand customer behavior. It helps businesses figure out who their most valuable customers are, who buys frequently, and who might be losing interest. By looking at when someone last shopped, how often they make purchases, and how much they spend, companies can create personalized marketing strategies that truly connect with their audience. This approach not only strengthens customer relationships but also boosts loyalty and helps businesses grow by focusing on what their customers need most.