

**PROJECT REPORT**

Online Retail-Customer Segmentation Analysis

**SUBMITTED BY:**

ANKHEE BISWAS

**Problem Statement**

An online retail store is trying to understand the various customer purchase patterns for their firm, you are required to give enough evidence-based insights to provide the same.

**Project Objective**

The objective of this project is to:

1. Using the above data, find useful insights about the customer purchasing history that can be an added advantage for the online retailer.
2. Segment the customers based on their purchasing behaviour

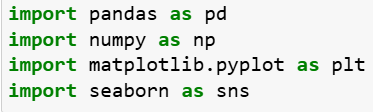
**Data Description**

The dataset available is: online\_retail.csv contains 387961 rows and 8 columns.

|  |  |
| --- | --- |
| **Feature Name** | **Description** |
| Invoice | Invoice Number |
| Stock Code | Productid |
| Description | Product Description |
| Quantity | Quantity of the product |
| Invoice Date | Date of the invoice |
| Price | Price of the Product per unit |
| CustomerID | Customerid |
| Country | Region of Purchase |

**Data Pre-processing Steps and Inspiration**The preprocessing of the data included the following steps:

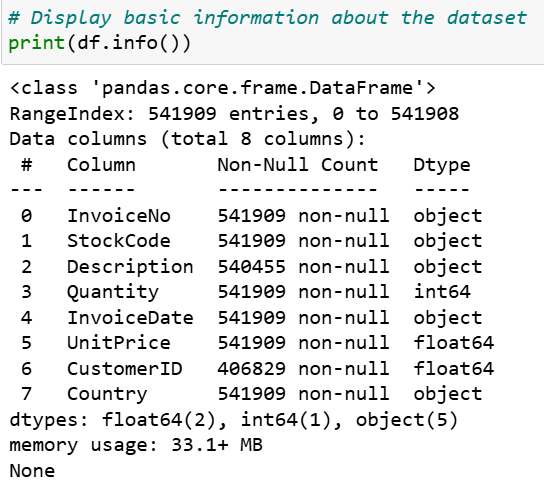
**Step 1:** Importing the necessary libraries.

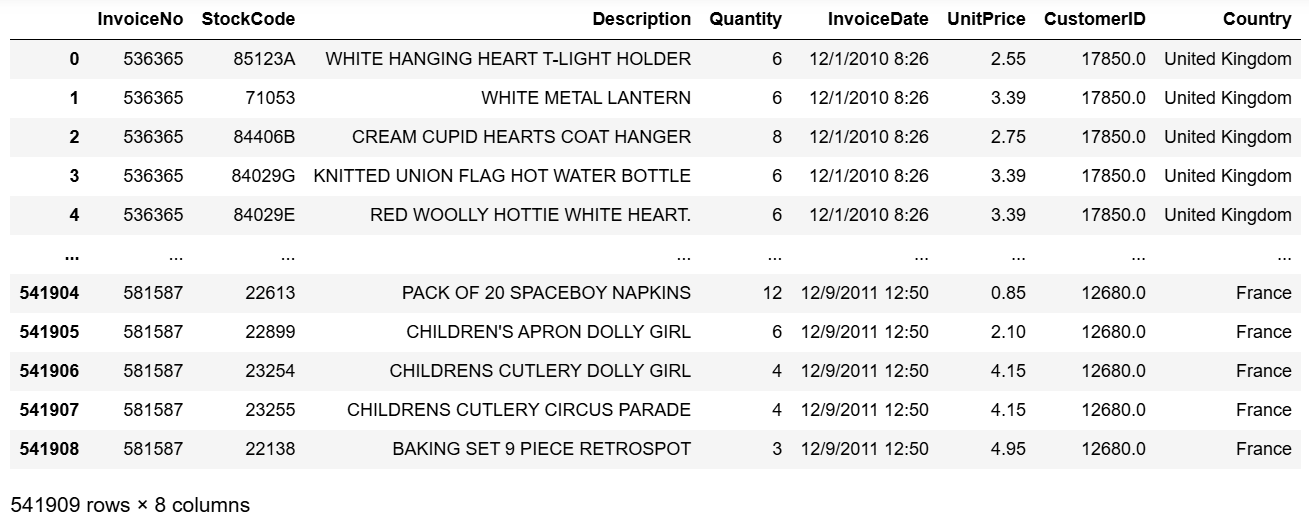


**Step 2:** Loading the dataset with the correct encoding.

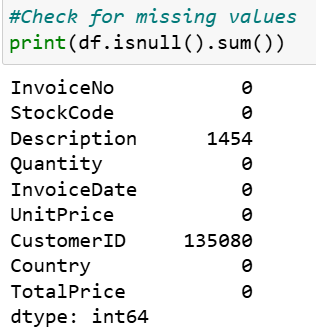


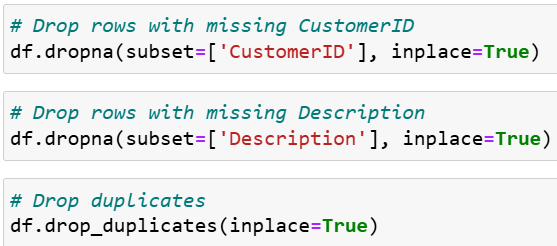
**Step 3:** Displaying the dataset with basic information & data type.





**Step 4:** Checking for missing values. Dropping the rows with the missing values at the same time working with the duplicate values.





**Choosing the Algorithm for the Project**

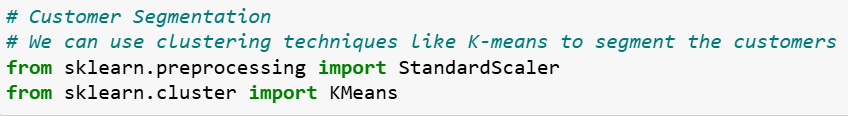
I have chosen K-Means clustering algorithm for analysing online retail customer segmentation in this project due to several advantages it offers.

1. Simplicity and Efficiency
2. Interpretability
3. Scalability
4. Versatility

K-Means clustering is often the go-to algorithm for customer segmentation in online retail due to its simplicity, efficiency, and interpretability, the choice of algorithm should be guided by the specific dataset characteristics and business requirements.

**Motivation and Reasons For Choosing the Algorithm**

K-Means is straightforward to implement and understand, making it a good starting point for customer segmentation. It is computationally efficient, especially with large datasets, due to its linear time complexity. It works well with large datasets, such as those typical in online retail scenarios, where customer behaviour data can be extensive. The results of K-Means clustering are easy to interpret. Each customer is assigned to a cluster, and these clusters can be analysed to understand distinct customer segments. It can handle various types of data attributes commonly found in retail datasets, such as purchase frequency, average spending, and recency of purchases.



**Assumptions**

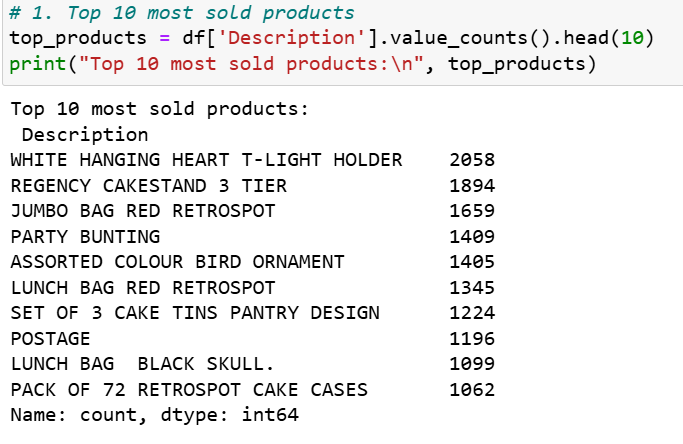
The following assumptions were made in order to create the model for the project includes an extensive exploratory data analysis with the RFM Analysis and segmenting customers based on K-Means cluster techniques on the provided Online Retail dataset to understand the following:

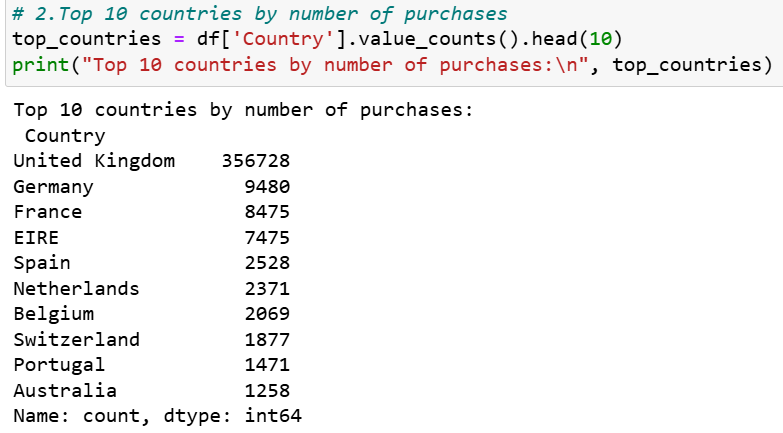
* Insights about customer purchasing history
* Customer Acquisition Over Time
* Customer Retention Rate Over Time
* Average Order Value (AOV) Trend Over Time
* Customer Lifetime Value (CLV) Segmentation.
* RFM Analysis for customer segmentation

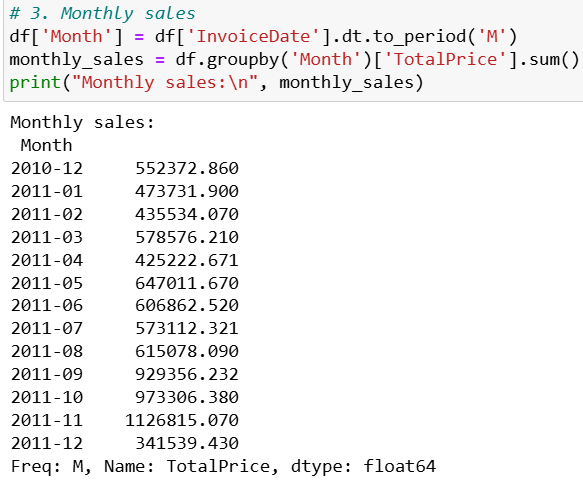
**Model Evaluation and Technique**

The following techniques and steps were involved in the evaluation of the model:

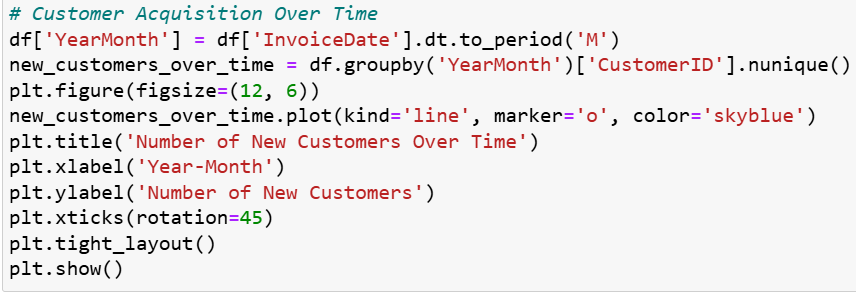
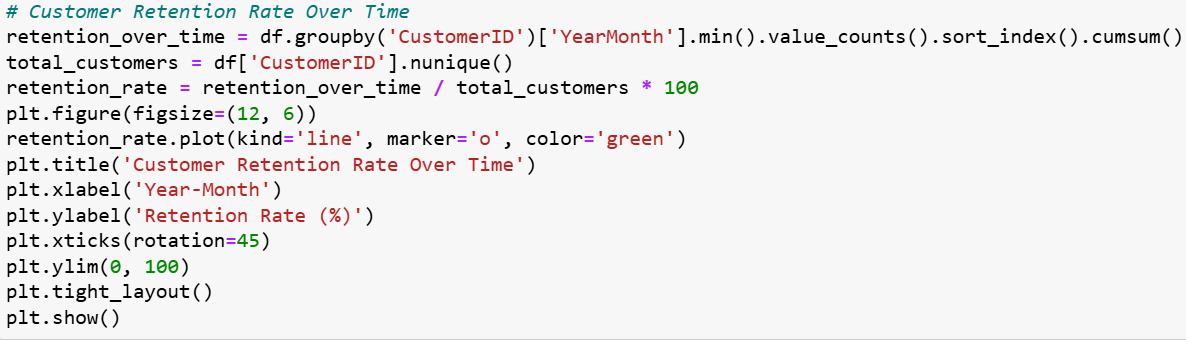
**Step 1:** Analysing the customer purchase behaviour and digging up the insights by EDA Analysis



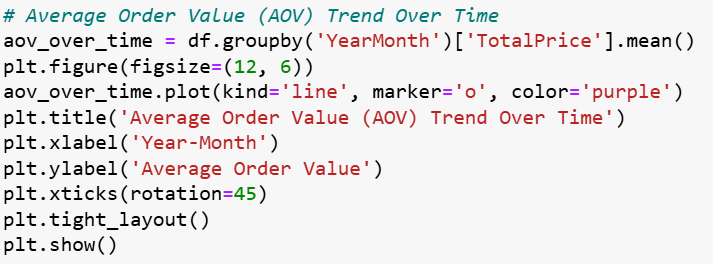


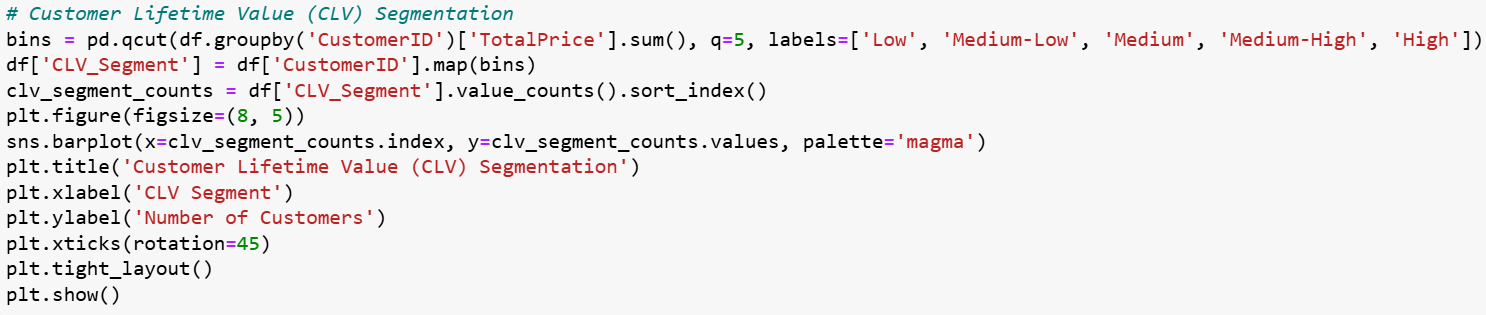


**Step 2:** Visualizing the Customer Acquisition Over Time & Retention Rate Over Time



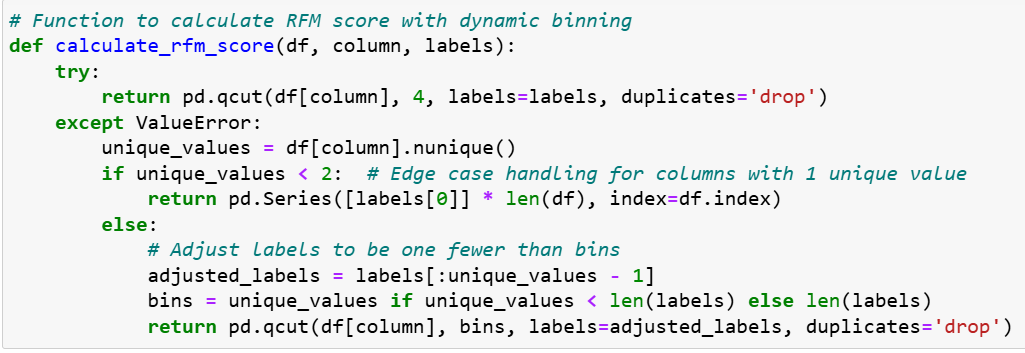
**Step 3:** Visualizing the Average Order Value (AOV) Trend Over Time & Customer Churn Analysis.

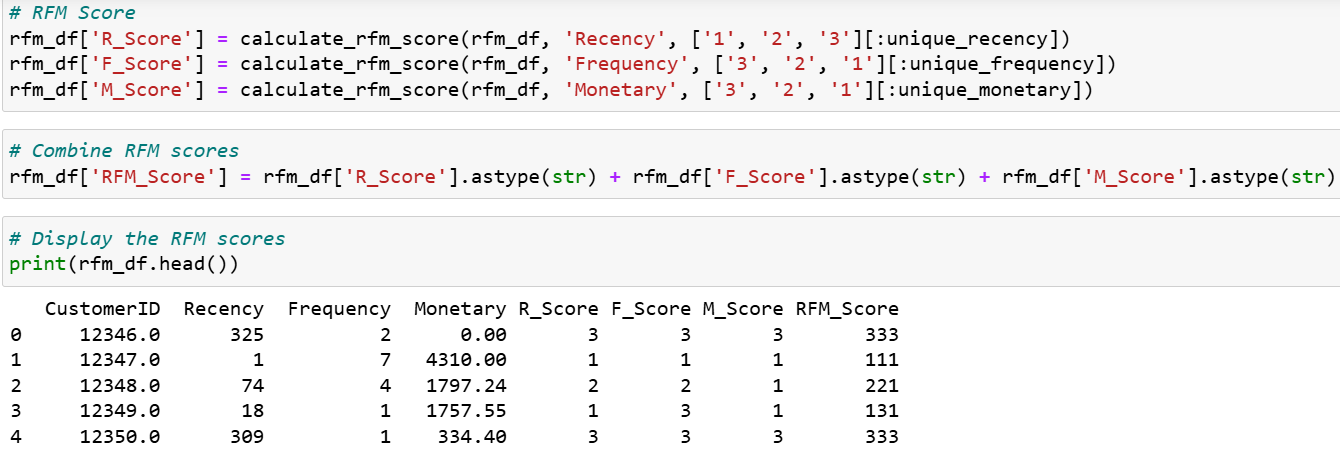


**Step 4:** Visualizing the Customer Lifetime Value (CLV) Segmentation.

**Step 5:** RFM Analysis for customer segmentation. In which we need to calculate the Recency, Frequency, Monetary. We take the calculate\_rfm\_score function to calculate the RFM Score in evaluate customer behaviour and segment customers based on their purchasing patterns.

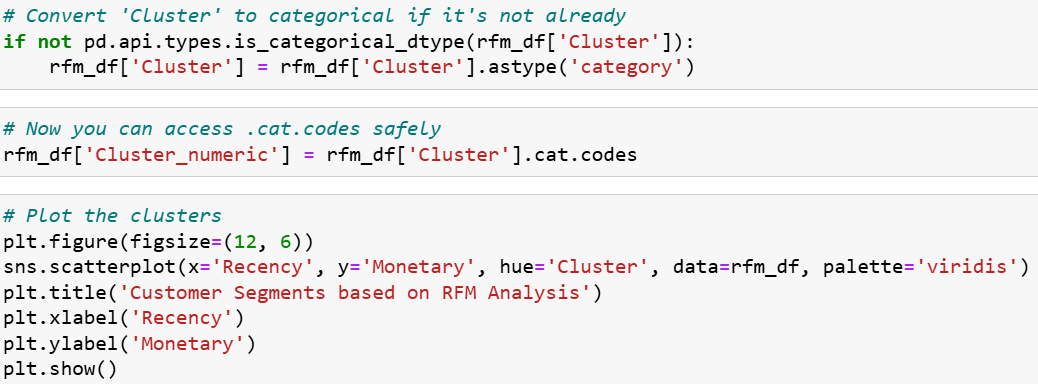


****

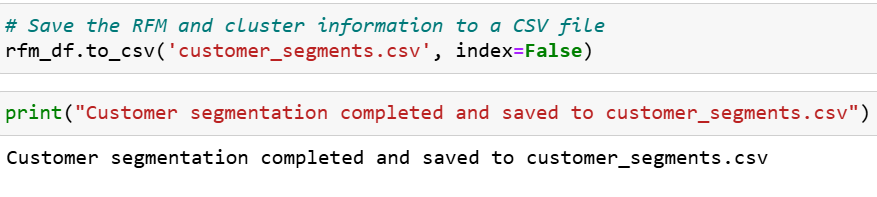
****

**Step 6:** Next, we use the clustering techniques like K-means to segment the customers.

**Step 7:** Lastly, we plot the clusters based on the RFM Analysis.

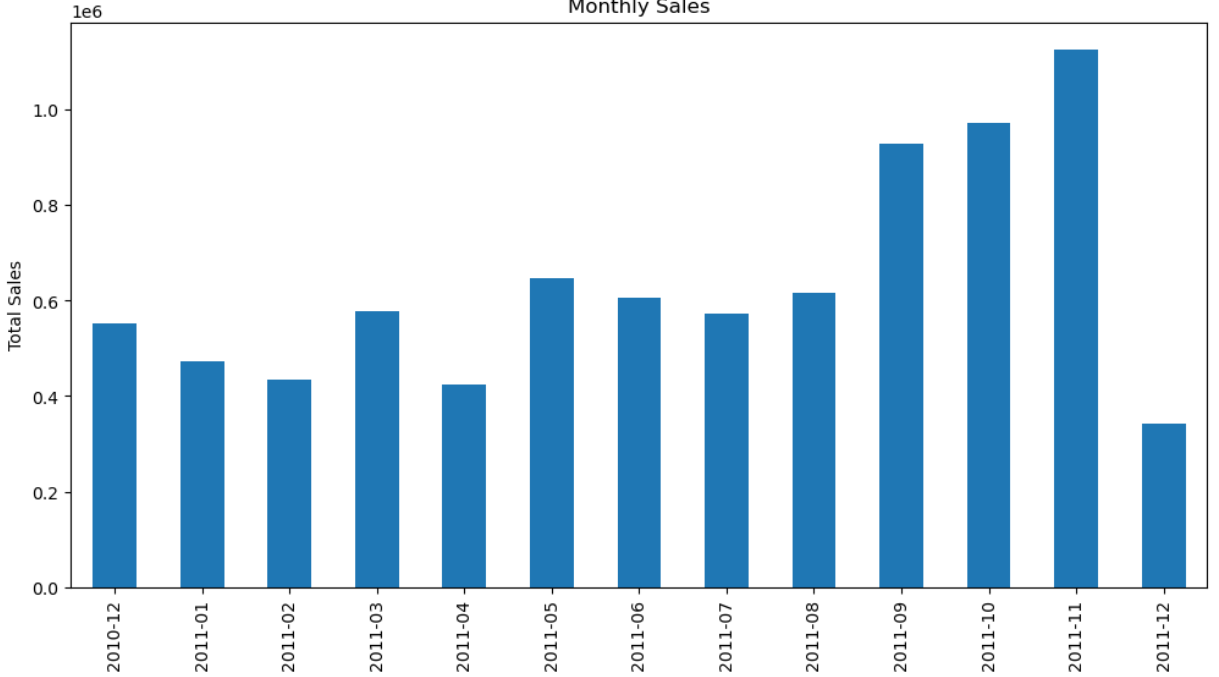


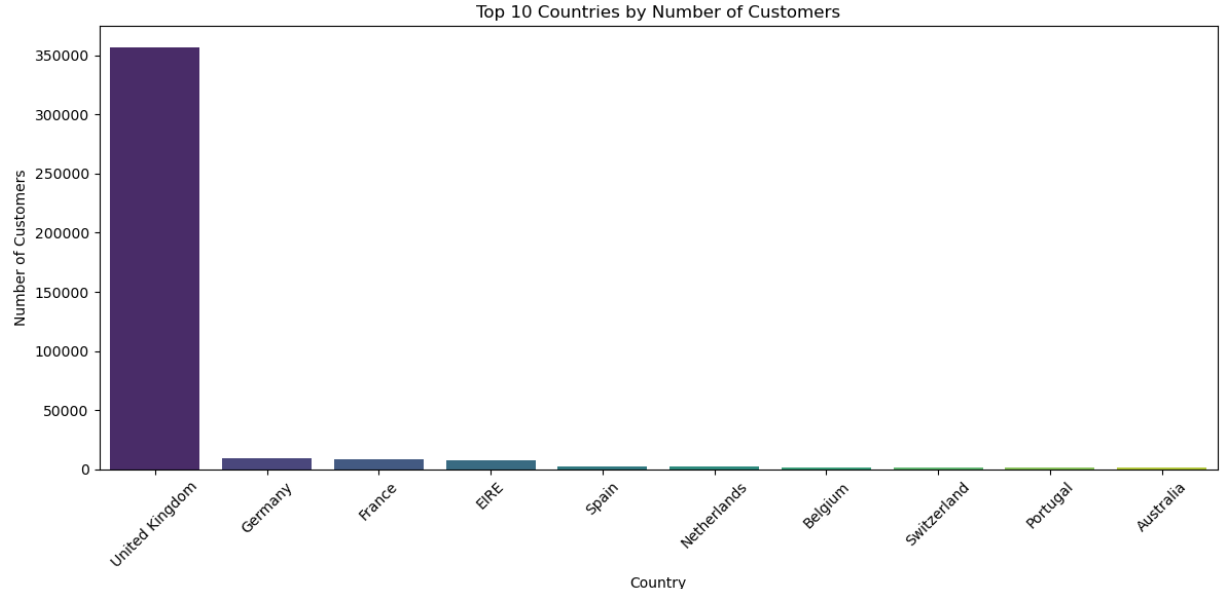
**Step 8:** We save the RFM and cluster information to a CSV file(customer\_segments.csv).

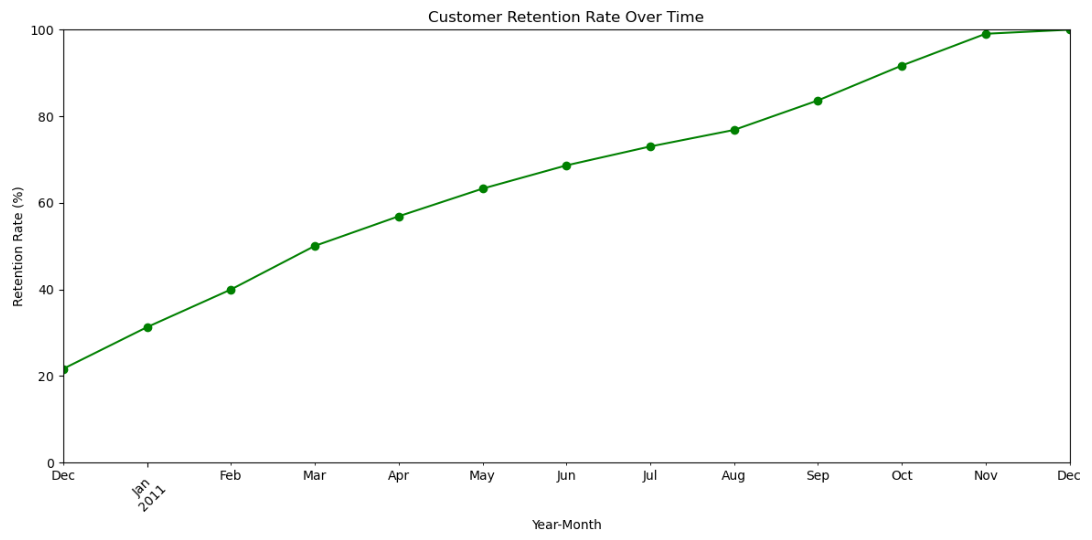
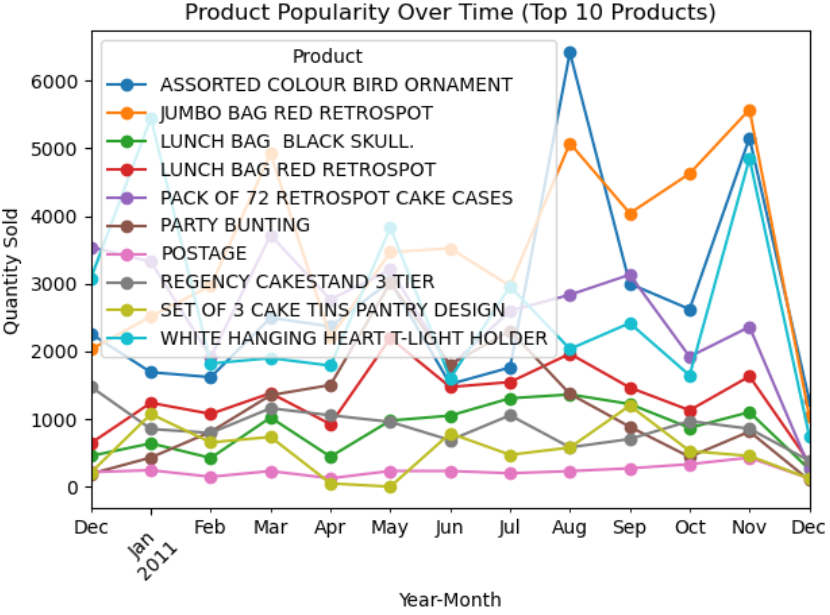
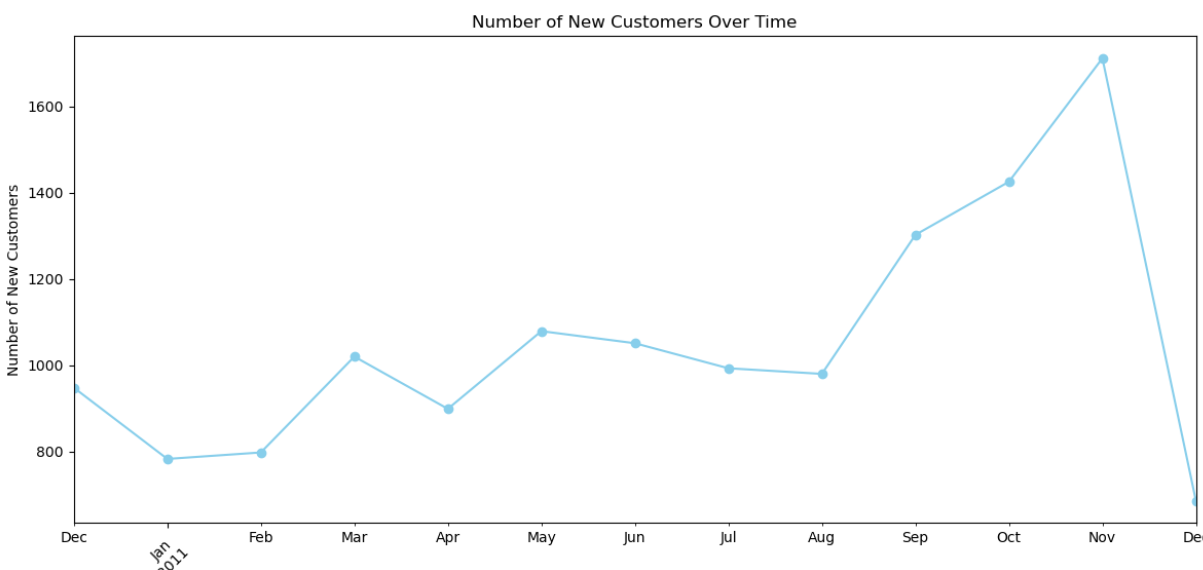


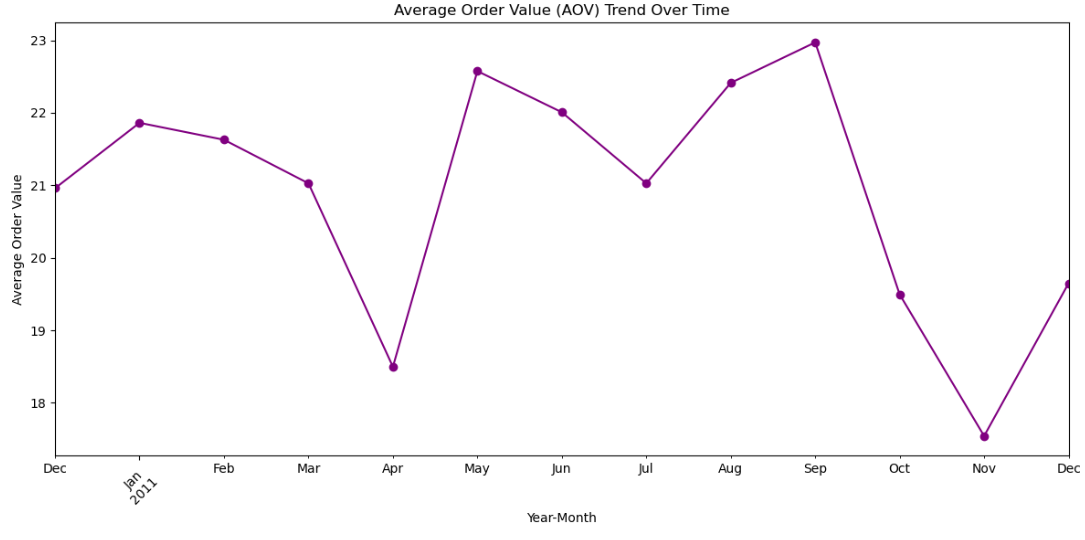
**Inferences from the Project**

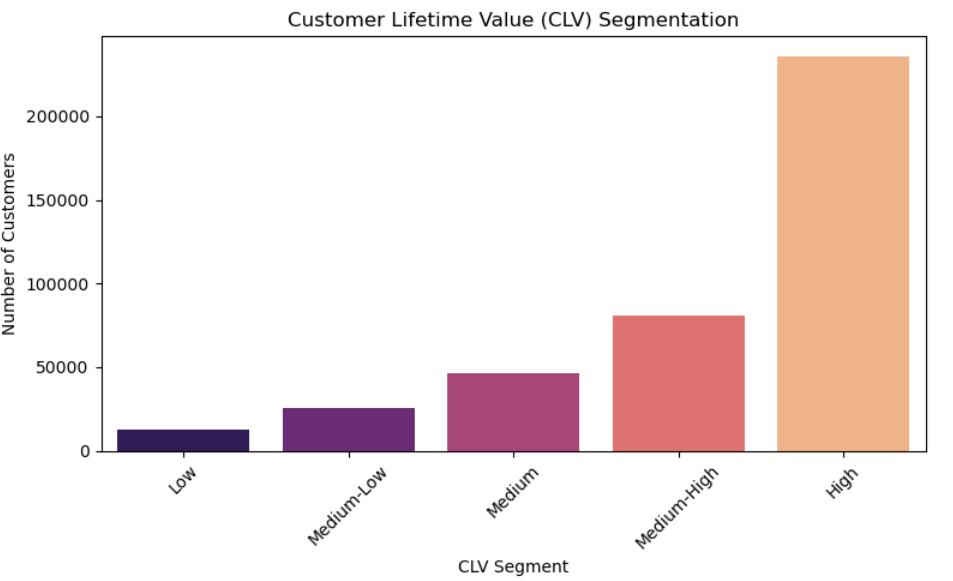
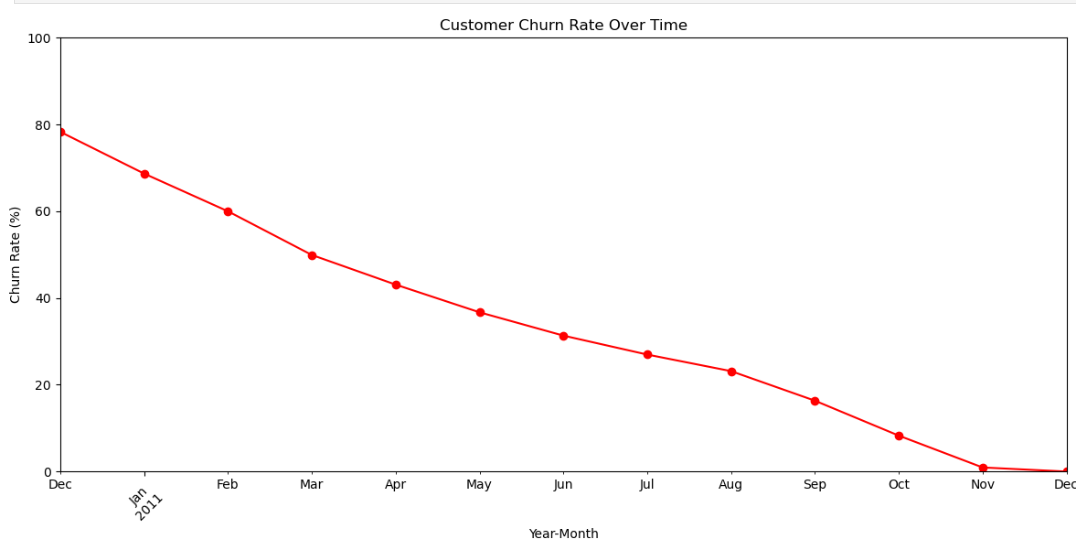
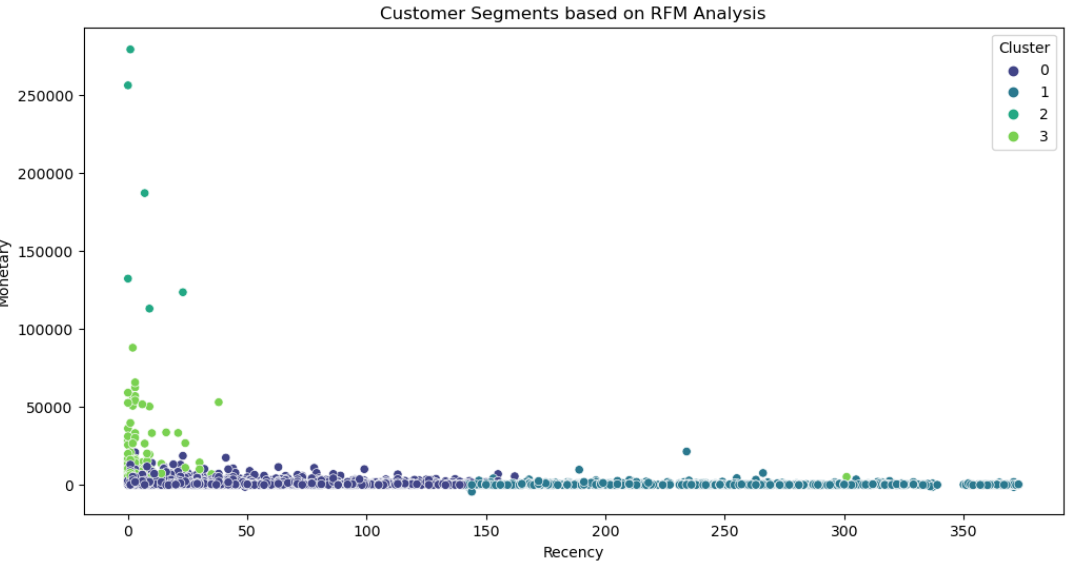
The model performance, inferences, based on the visualizations created on the online retail analysis.





****

****

****

**Future Possibilities**

The future possibilities, limitations are as follows:

* **Targeted Marketing Campaigns:** To Develop tailored marketing campaigns for each customer segment based on their RFM characteristics. For example, high-value customers (high Monetary) can be targeted with exclusive offers, while recent but infrequent buyers (low Recency, low Frequency) can be encouraged to make repeat purchases.
* **Customer Retention Programs:** Identify customers with low recency and frequency who were once high spenders (high Monetary) and create retention strategies to win them back.
* **Product Recommendations:** Use purchasing behaviour data to recommend products that similar customer segments have bought, increasing cross-sell and up-sell opportunities.
* **Personalized Communication:** Customizing communication strategies to match the preferences and behaviours of different customer segments, improving customer satisfaction and loyalty.
* **Inventory Management:** Analysing purchasing patterns to optimize inventory levels, ensuring popular products are always in stock and reducing overstock of less popular items.

**Conclusion**

"In conclusion, the analysis of the online retail dataset provides significant insights into customer purchasing behaviours, which can be leveraged to enhance business strategies and operations. By segmenting customers based on Recency, Frequency, and Monetary (RFM) metrics, we gain a nuanced understanding of different customer groups.

The RFM segmentation not only highlights current purchasing patterns but also informs future strategic initiatives. For instance, targeted marketing campaigns can be developed based on each segment's characteristics, personalized communication can enhance customer relationships, and inventory management can be optimized to ensure popular products are always available. By leveraging these insights, the online retail store can improve customer satisfaction, increase revenue, and build a loyal customer base, ultimately leading to sustained business growth.