

# Customer Segmentation - RFM Model

## ● Executive Summary

This project employed the RFM model to segment our customer base into distinct groups based on their **Recency** (time since last purchase), **Frequency** (number of purchases), and **Monetary Value** (total spent). By analysing these segments, we gained valuable insights into customer behaviour. The results reveal a high value segment with low recency, a mid value segment and a low value segment with high recency. This segmentation allows us to personalise marketing strategies for each group, such as offering exclusive discounts and loyalty programs to our high-value segment to incentivize continued engagement, and implementing win-back campaigns with special offers for the low-frequency segment to re-activate them. By implementing these targeted strategies, we can improve customer engagement, retention, and ultimately, drive revenue growth. By implementing these targeted strategies, we can improve customer engagement, retention, and ultimately, drive revenue growth.

## ● Exploratory Data Analysis

```
[3]: df = pd.read_excel('/kaggle/input/customer-segmentation/Online Retail.xlsx')
df.head()
```

```
[3]:
```

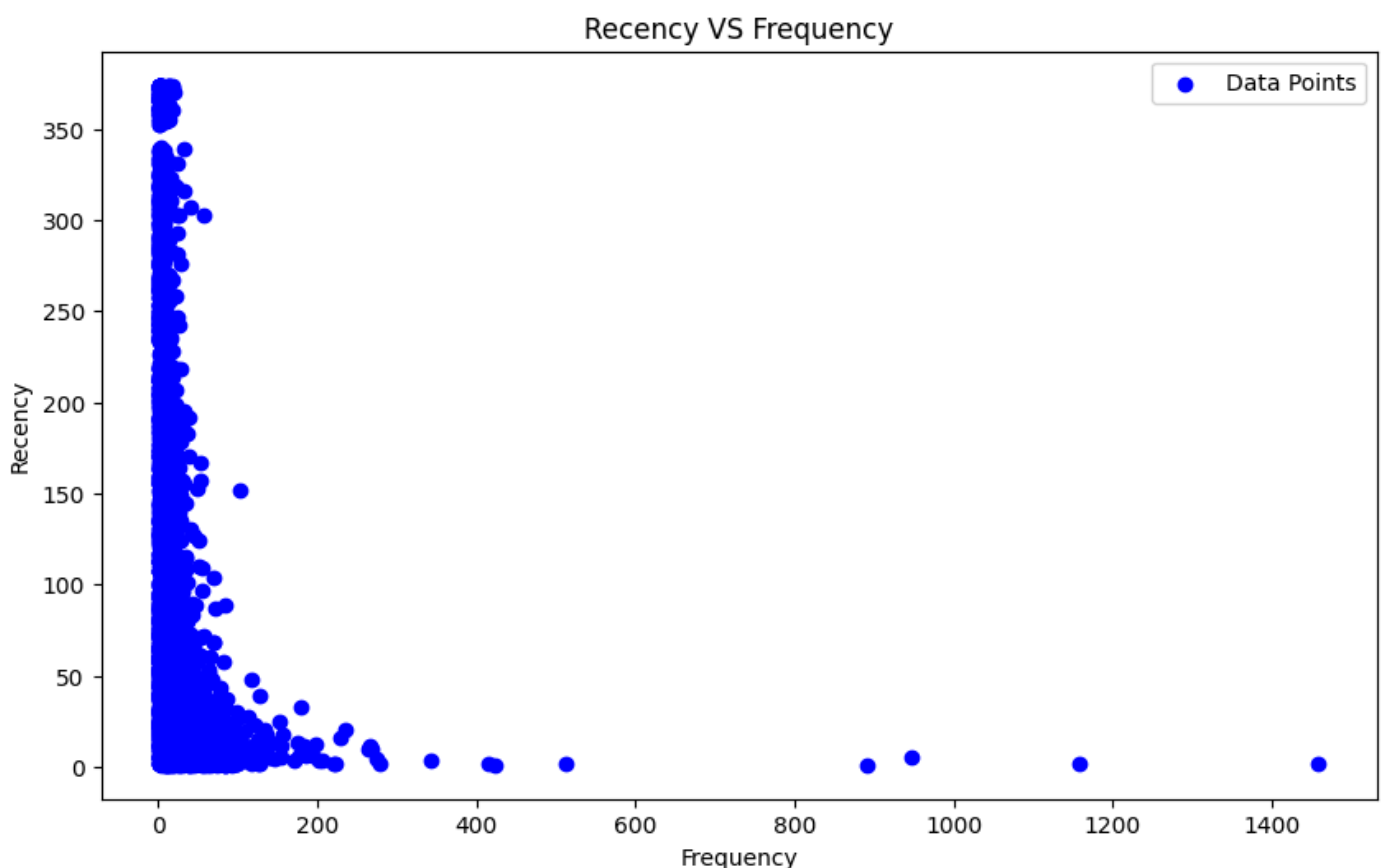
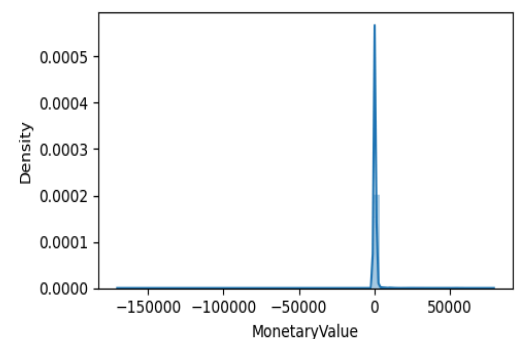
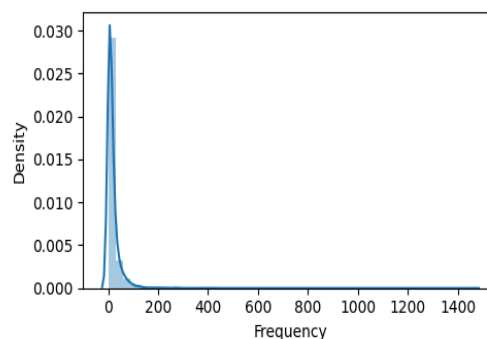
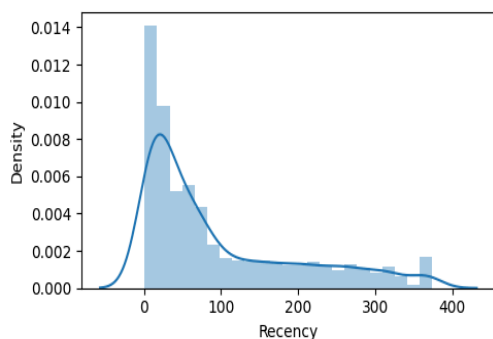
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

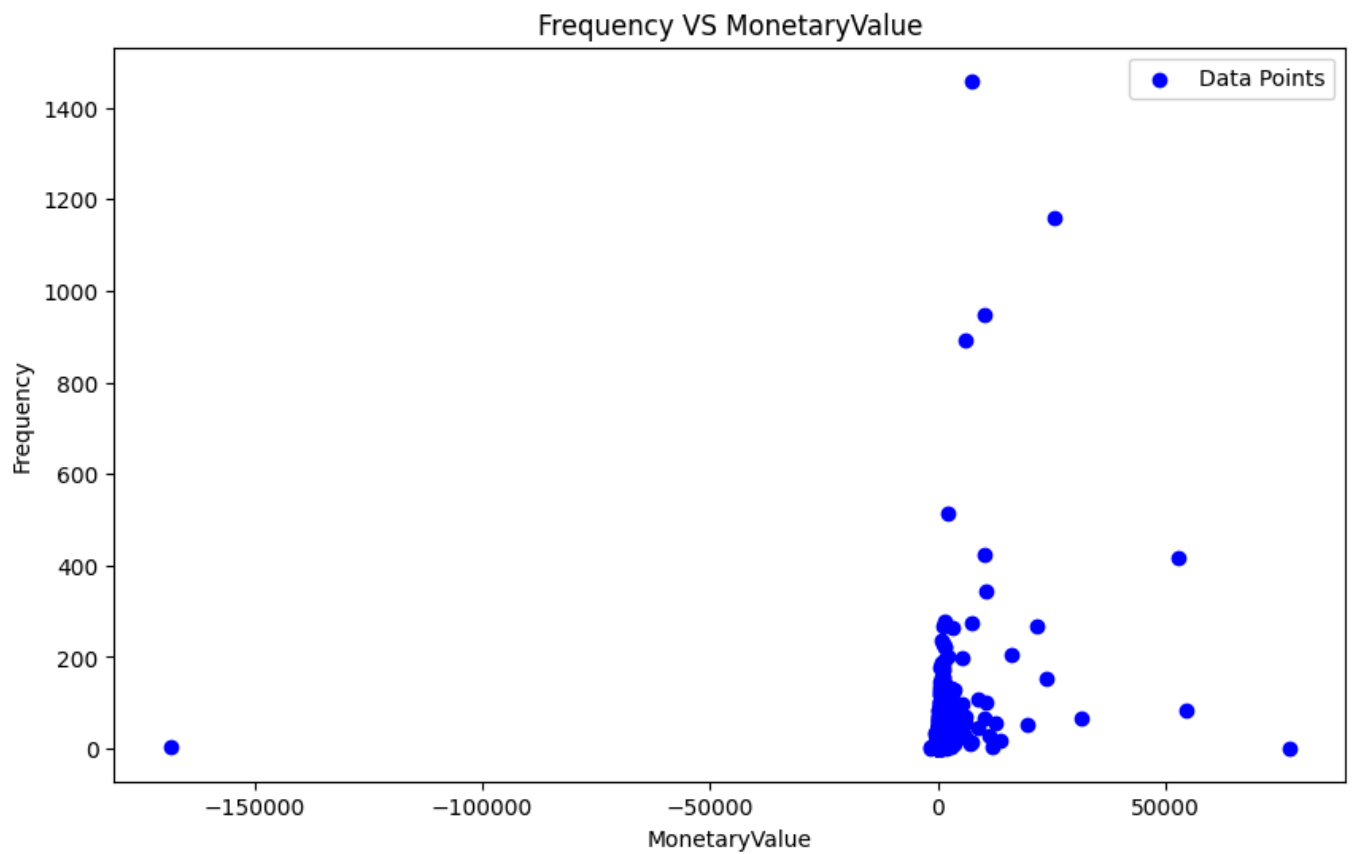
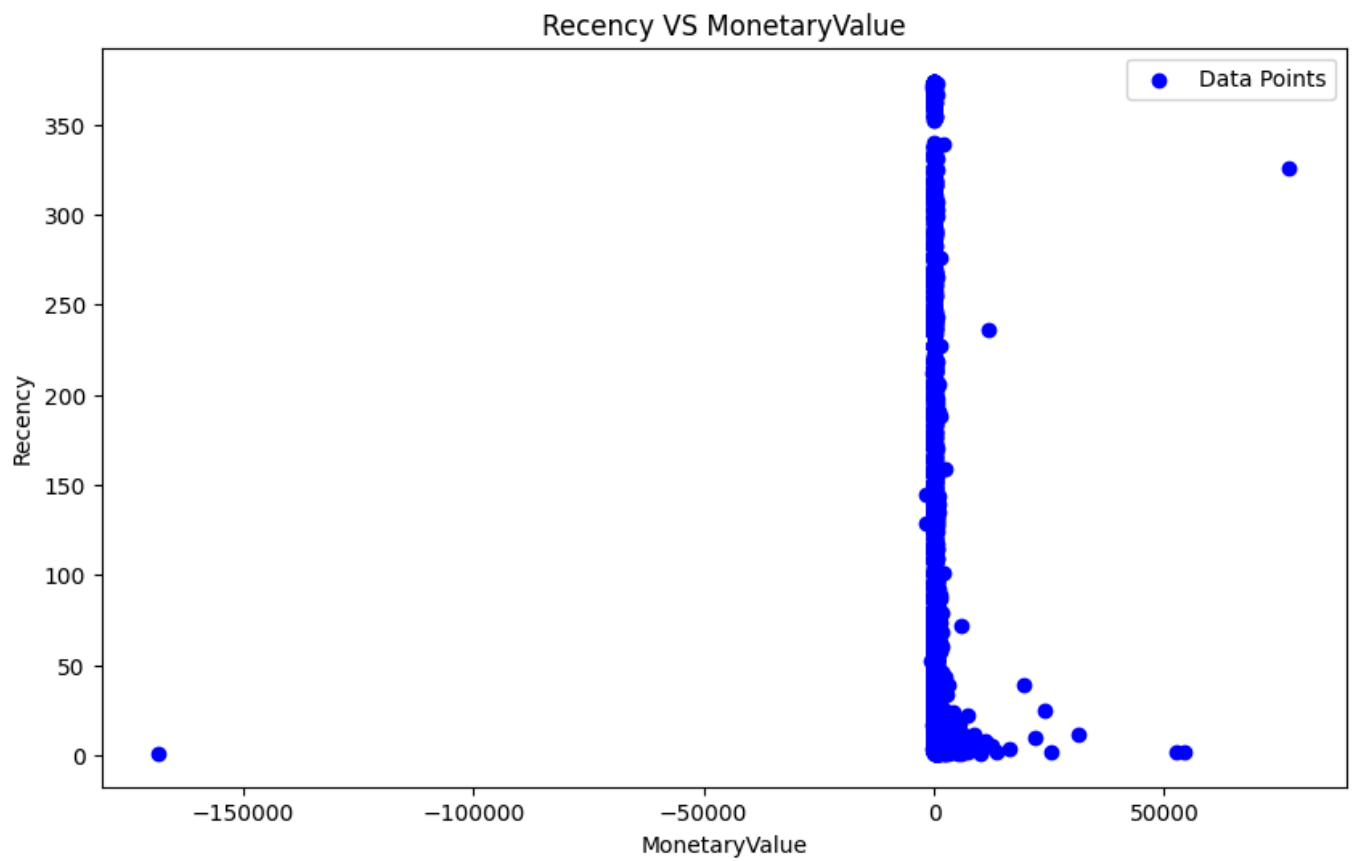
We have a dataset with columns: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID and Country. We first convert this dataframe into RFM Format

	Recency	Frequency	MonetaryValue
CustomerID			
12346.0	326	1	77183.60
12347.0	3	27	588.48
12348.0	319	5	443.52
12349.0	19	15	578.21
12350.0	311	2	42.90

To leverage the RFM model, we transformed this data into a customer-centric format:

- **Frequency:** We calculated the total number of invoices per customer using the `InvoiceNo` column.
- **Monetary Value:** We obtained the total amount spent by each customer by multiplying `Quantity` and `UnitPrice` columns.
- **Recency:** We calculated the time elapsed since the most recent purchase for each customer using the `InvoiceDate` column (code not shown, but assumed implemented).





Visualising the distribution of Recency, Frequency, and Monetary Value revealed significant skewness, indicating the need for data transformation. We explored various well-suited transformations, including:

- Boxcox transformation
- Log transformation
- Sqrt transformation
- Cbrt transformation

We compared the effectiveness of these transformations by analyzing their impact on skewness.

	Transformation	Recency	Frequency	MonetaryValue
0	[None]	[1.1889733424996805]	[17.700978061749407]	[-25.892208952176592]
1	[Log]	[-0.4958741503954887]	[0.13371097212837377]	[nan]
2	[Cube Root]	[0.26879302524720655]	[1.6283138512547428]	[0.36078057931212704]
3	[BoxCox]	[-0.07045465130106617]	[0.007259597019609291]	[-25.892208952176592]

Based on this analysis, we selected the following transformations:

- **Boxcox transformation:** Applied to both Recency and Frequency columns.
- **Cbrt transformation:** Applied to the Monetary Value column.

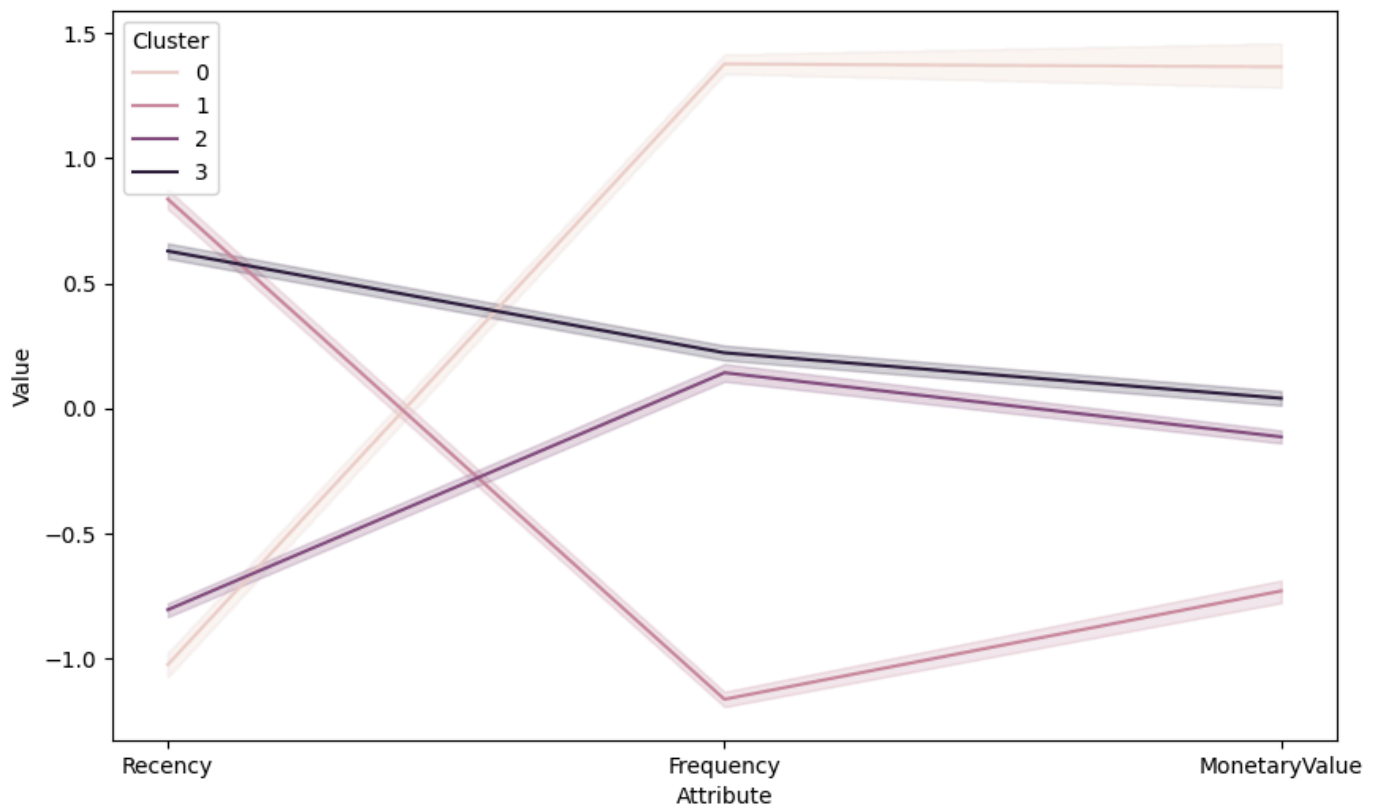
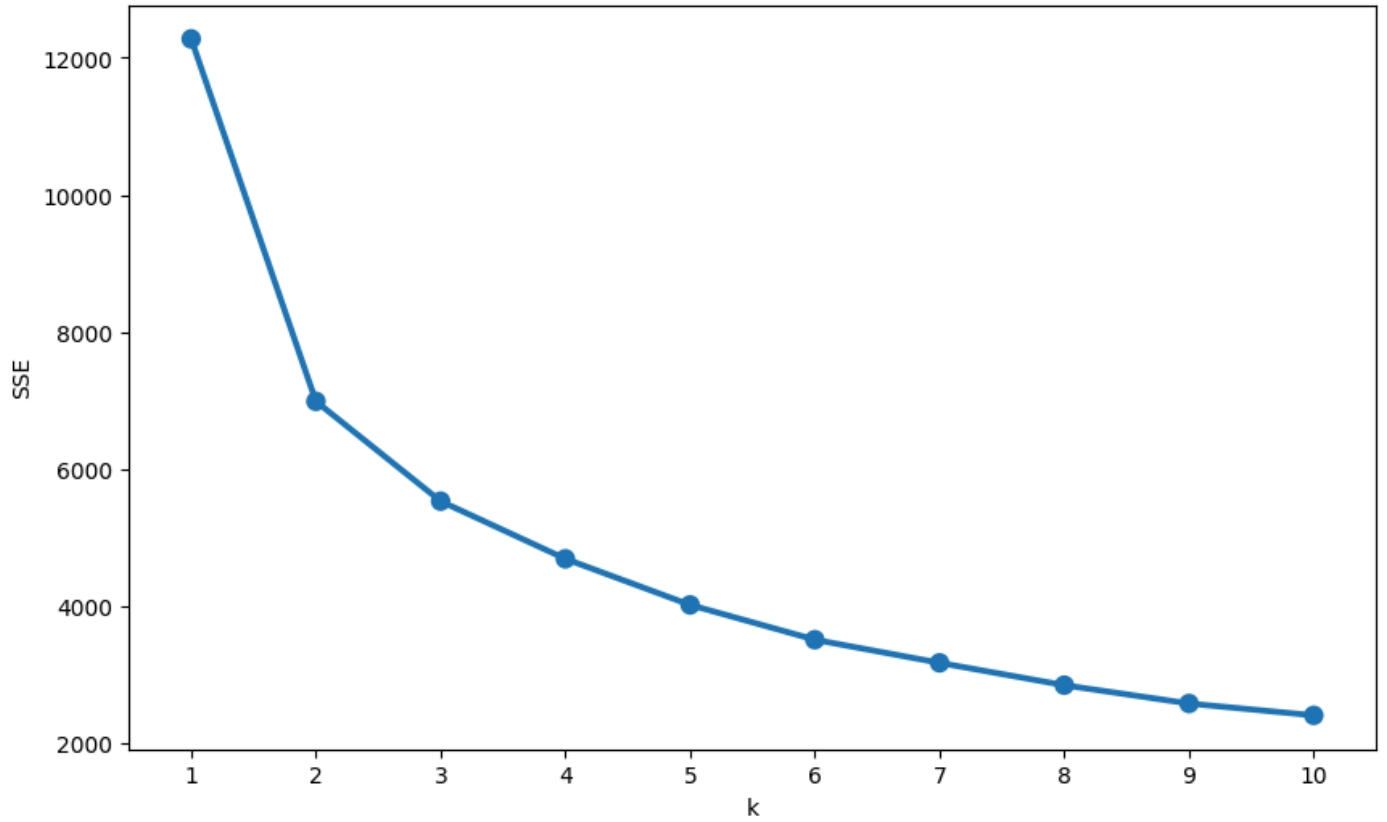
Following the chosen transformations, we normalised the data using the **StandardScaler** technique to ensure all features contribute equally to the clustering process.

## • Data Modeling

We employed the K-Means clustering algorithm for customer segmentation. To determine the optimal number of clusters, we utilized the elbow plot technique. The elbow plot visually suggests the point at which adding additional clusters yields minimal benefit. Based on this analysis, we identified **four clusters** as the most suitable representation of our customer base.

We then applied K-Means clustering and visualized the resulting customer distribution across the identified segments.

The Elbow Method



- **Business Strategy:**

By analyzing the characteristics of each customer segment, we can develop targeted marketing strategies to enhance customer engagement and drive business growth:

- **Potential Loyalists (Cluster 3):** These customers demonstrate recent purchases with average frequency and significant spending. We recommend offering loyalty programs or membership benefits to incentivize continued engagement. Additionally, suggesting related products for upselling presents an opportunity to convert them into loyal customers.
- **New Customers (Cluster 1):** This segment represents customers with high overall RFM scores but infrequent purchases. Building relationships with these customers is crucial. We can achieve this by providing onboarding support and personalised special offers to encourage repeat visits and increase their purchase frequency.
- **At-Risk Customers (Cluster 0):** These customers have a history of frequent purchases with high spending but haven't shown recent activity. Reconnecting with them through personalised reactivation campaigns is crucial. Offering product renewals, helpful recommendations, and special discounts can encourage them to resume purchases.
- **Mid-Value Customers (Cluster 2):** This segment comprises customers with lower recency but a history of decent purchase frequency and value. Re-engaging them requires targeted strategies. We can implement relevant promotions and conduct surveys to understand why their purchase frequency has declined. Addressing their concerns and offering incentives can help retain them and prevent them from turning to competitors.

- **Conclusion:**

By leveraging the RFM model and data-driven segmentation, we gained valuable insights into customer behavior. This knowledge empowers us to develop targeted marketing strategies for each customer segment, fostering stronger customer relationships, improving retention rates, and ultimately driving sustainable business growth.

- **Further considerations:**

This report provides a foundational framework for customer segmentation. We can further enhance this analysis by:

1. Incorporating additional customer data (e.g., demographics, product preferences) to enrich customer profiles.
2. Evaluating the effectiveness of implemented marketing strategies on customer behavior over time.
3. Continuously refining the segmentation model as new data becomes available.

By adopting a data-driven approach to customer segmentation and implementing targeted marketing strategies, we can unlock significant opportunities to optimize customer engagement and achieve our business goals.