**RFM Analysis**

RFM Analysis is used to understand and segment customers based on their buying behaviour.

RFM stands for recency, frequency, and monetary value, which are three key metrics that provide information about customer engagement, loyalty, and value to a business.

Using RFM Analysis, a business can assess customers:

* recency (the date they made their last purchase)
* frequency (how often they make purchases)
* and monetary value (the amount spent on purchases)

The given dataset is provided by an e-commerce platform containing customer transaction data including .

1. customer ID,
2. purchase date,
3. transaction amount,
4. product information,
5. Order ID,
6. location.

The platform aims to leverage RFM (recency, frequency, monetary value) analysis to segment customers and optimize customer engagement strategies

Calculate the Recency, Frequency, and Monetary values of the customers to move further:

* To calculate recency, we subtracted the purchase date from the current date and extracted the number of days using the datetime.now().date() function. It gives us the number of days since the **customer’s last purchase**, representing their **recency value.**
* We calculated the frequency for each customer. We grouped the data by ‘CustomerID’ and counted the number of unique ‘OrderID’ values to determine the number of purchases made by each customer. It gives us the **frequency value,** representing **the total number of purchases made by each customer.**
* We calculated the monetary value for each customer. We grouped the data by ‘CustomerID’ and summed the ‘TransactionAmount’ values to calculate the total amount spent by each customer. It gives us the **monetary value**, representing the **total monetary contribution of each customer.**

**Calculating RFM Scores:**

recency\_scores = [5, 4, 3, 2, 1] # Higher score for lower recency (more recent)

frequency scores = [1, 2, 3, 4, 5] # Higher score for higher frequency

monetary\_scores = [1, 2, 3, 4, 5] # Higher score for higher monetary value

* We assigned scores from 5 to 1 to calculate the recency score, where a higher score indicates a more recent purchase. It means that customers who have purchased more recently will receive higher recency scores.
* We assigned scores from 1 to 5 to calculate the frequency score, where a higher score indicates a higher purchase frequency. Customers who made more frequent purchases will receive higher frequency scores.
* To calculate the monetary score, we assigned scores from 1 to 5, where a higher score indicates a higher amount spent by the customer. bold text

**# Calculate RFM score by combining the individual scores #**

data['RFM\_Scores'] = data['RecencyScore'] + data['FrequencyScore'] + data['MonetaryScore']

**# Create RFM segments based on the RFM score #**

segment\_labels = ['Low-Value', 'Mid-Value', 'High-Value']

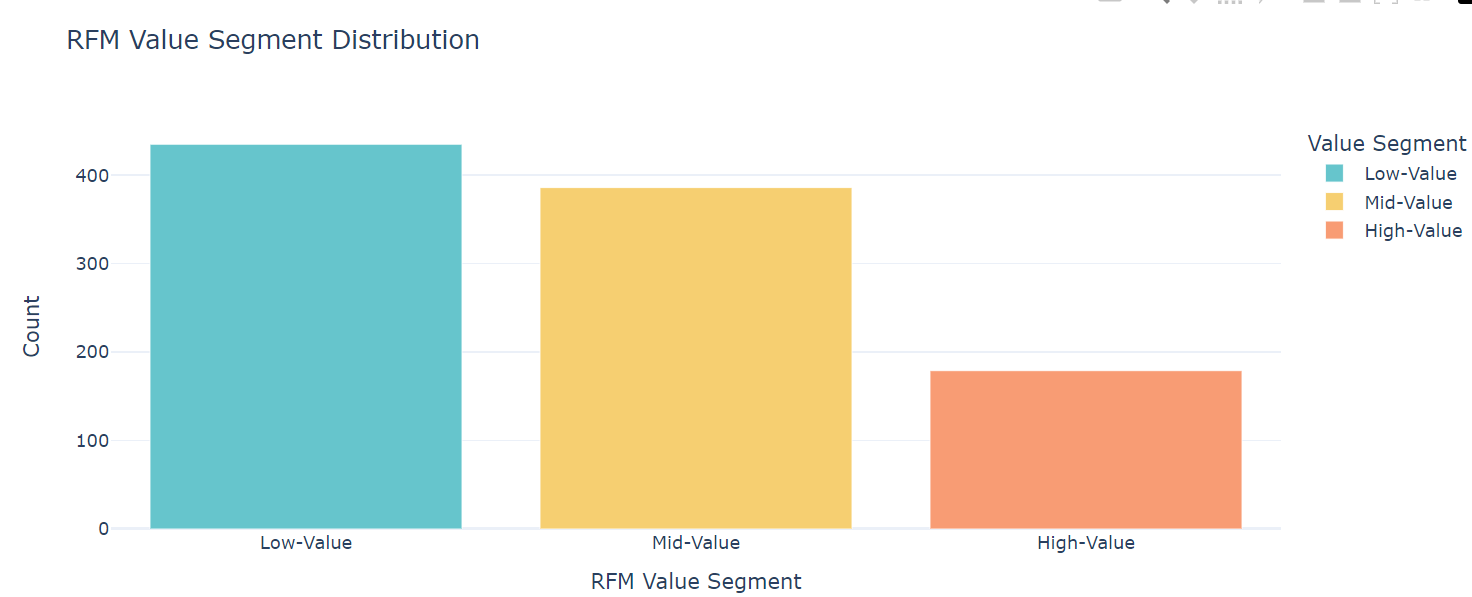
* We divided RFM scores into three segments, namely “Low-Value”, “Mid-Value”, and “High-Value”. Segmentation is done using the pd.qcut() function, which evenly distributes scores between segments.

**# RFM Segment Distribution #**

Low-Value 435

Mid-Value 386

High-Value 179



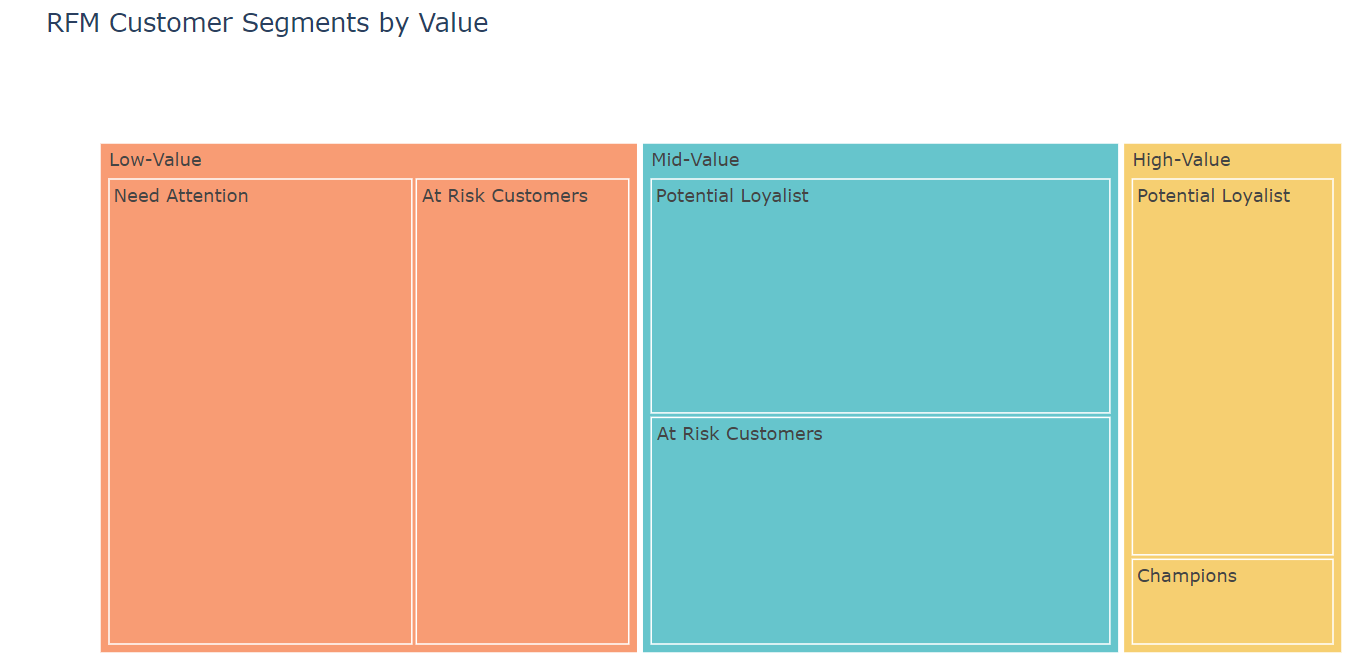
# Now let’s create and analyze {[ RFM Customer Segments }] that are broader classifications based on the RFM scores.

**RFM Customer Segments**

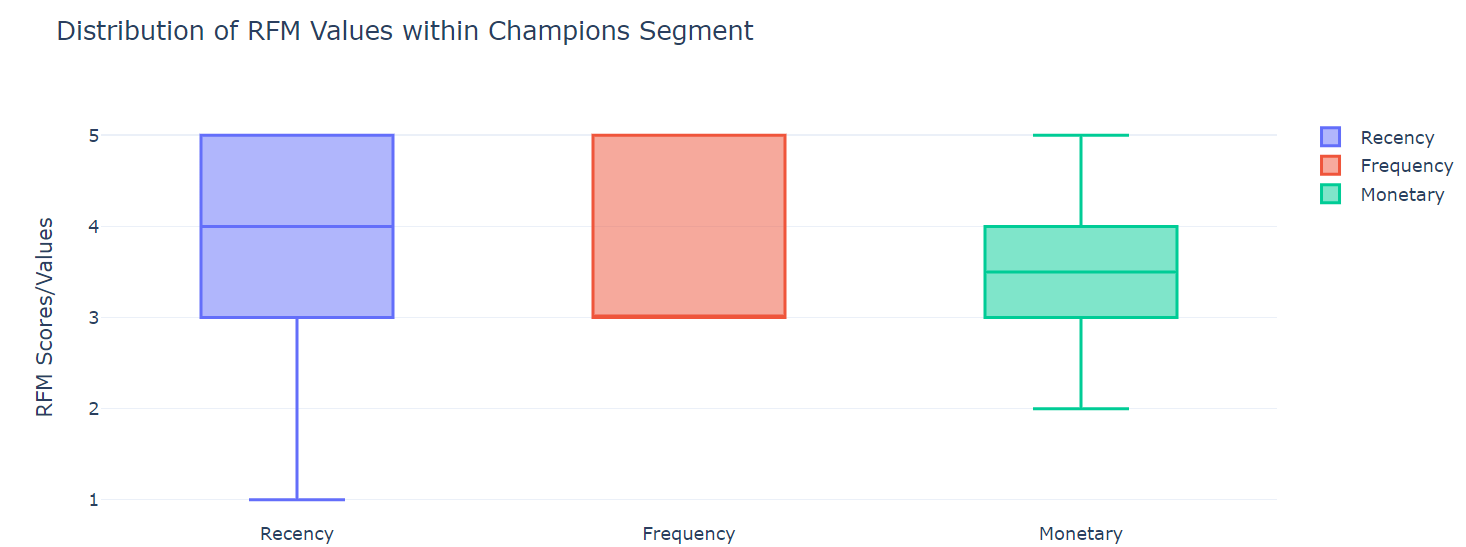
* These segments, such as “Champions”, “Potential Loyalists”, "At Risk Customers", "Need Attention" and “Lost” provide a more strategic perspective on customer behaviour and characteristics in terms of recency, frequency, and monetary aspects.

**Now let’s analyze the distribution of customers across different RFM customer segments within each value segment:**

|  | **Value Segment** | **RFM Customer Segments** | **Count** |
| --- | --- | --- | --- |
| **2** | Low-Value | Need Attention | 255 |
| **7** | Mid-Value | Potential Loyalist | 196 |
| **4** | Mid-Value | At Risk Customers | 190 |
| **0** | Low-Value | At Risk Customers | 180 |
| **11** | High-Value | Potential Loyalist | 145 |
| **9** | High-Value | Champions | 34 |
| **1** | Low-Value | Champions | 0 |



**Now let’s analyze the distribution of RFM values within the Champions segment:**

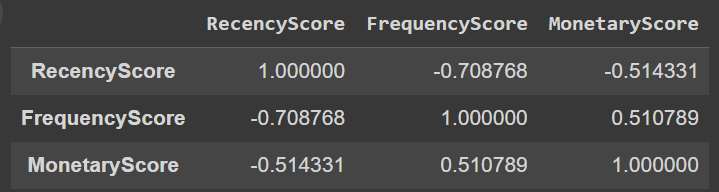
****

**Recency: Q3/Max/Upper Fence = 5, Median = 4, Q1 = 3, Lower fence/Min = 1**

**Frequency: Q3/Max/Upper Fence = 5, Median = 3, Lower fence/Min/Q1 = 3**

**Monetary: Max/Upper Fence = 5, Q3 = 4, Median = 3.5, Q1 = 3, Lower fence/Min = 2**

**# Now let’s analyze the correlation of the recency, frequency, and monetary scores within the champions segment:**

****

**Here we can see that Frequency and Monetary Scores are positively co-related and rest others are negatively co-related**

**Now let’s have a look at the number of customers in all the segments:**

RFM Customer Segments Count

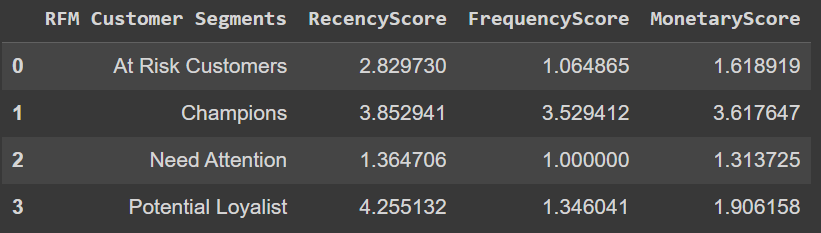
0 At Risk Customers 370

1 Potential Loyalist 341

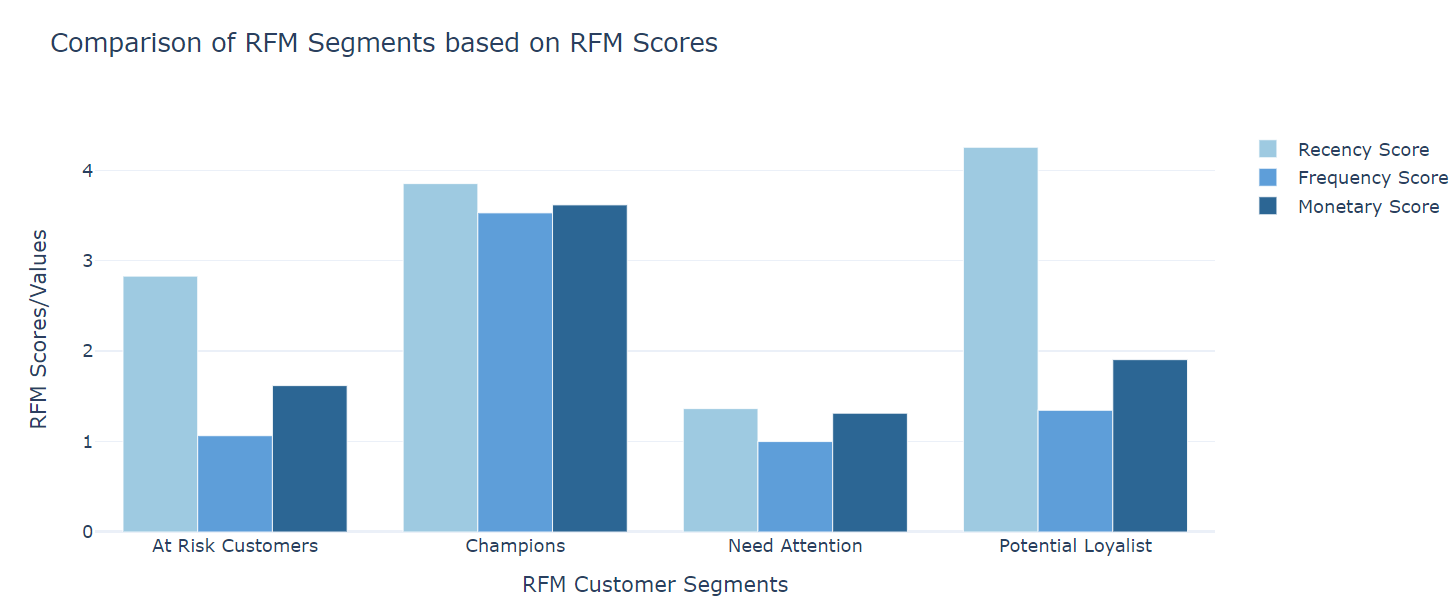
2 Need Attention 255

3 Champions 34

**# Now let’s have a look at the recency, frequency, and monetary scores of all the segments:**

****

**# Create a grouped bar chart to compare segment scores:**

****

* **Champions segments is performing good in three aspects of RFM.**
* **Potential loyalist & At Risk Customers have higher Recency score followed by Monetary and Frequency.**
* **Need Attention customers have lower Recency and Monetary Scores followed by Frequency as they are less frequent at the stores when compared to other segments of customers.**