

PROJECT REPORT

BEHAVIOURAL SEGMENTATION AND PURCHASE ANALYSIS IN VARIOUS RETAIL STORES

By: YUGESH MADALA

Regd.mailid:yugeshyogi972@gmail.com

CONTENTS

1.Problem Statement

2.Project Objective

3.Data Description

4.Data Preprocessing Steps

5.Data Visualization Inferences

6.RFM analysis

7.Assumptions

8.Behavioural Segmentation of Customers and Interpretations

9.Business Insights

10.Future possibilities of Project

11.References

1.PROBLEM STATEMENT:

- The online retail sector has experienced exponential growth, leading to vast amounts of transactional data.
- Analysing this data is crucial for understanding customer behaviour, improving sales strategies, and optimizing marketing efforts.
- The problem is to identify key customer segments, understand their behaviour, and derive actionable insights to enhance business decision-making.

2.PROJECT OBJECTIVE:

The primary objective is to analyse customer purchasing patterns from an online retail dataset and segment customers based on their behaviour. The Primary goals include:

1.Customer Segmentation:

- Utilize RFM (Recency, Frequency, Monetary) analysis to segment customers based on their purchasing behavior.
- Identify different customer groups such as loyal customers, high spenders, and potential churners.

2.Understanding Customer Behaviour:

- Gain insights into customer purchasing patterns and trends.
- Determine the recency of customer purchases, the frequency of their transactions, and their overall monetary contribution.

3.Targeted Marketing Strategies:

- Develop tailored marketing strategies for different customer segments.
- Enhance customer engagement and retention by delivering personalized offers and promotions.

4.Revenue Maximization:

- Identify high-value customers who contribute significantly to revenue.
- Develop strategies to nurture these customers and increase their lifetime value.

5.Customer Retention:

- Detect potential churners and create retention strategies to prevent customer attrition.
- Understand the factors leading to customer churn and address them proactively.

6.Business Decision Making:

- Provide actionable insights to the business for making informed decisions.
- Use RFM analysis to guide product recommendations, inventory management, and promotional activities.

By achieving these goals, the project aims to **enhance the overall customer experience, improve customer loyalty, and drive business growth through data-driven decision making.**

3.DATA DESCRIPTION:

The dataset used for this analysis consisting of historical transactions from an online retail store. The key features of the dataset include:

- **InvoiceNo:** A unique identifier assigned to each transaction. Each row in the dataset corresponds to a single transaction.
- **StockCode:** A unique identifier assigned to each distinct product.
- **Description:** A textual description of the product.
- **Quantity:** The quantity of each product per transaction. This represents the number of units sold per transaction.
- **InvoiceDate:** The date and time when a transaction was generated.
- **UnitPrice:** The unit price of the product. This represents the price per unit in GBP (British pounds).
- **CustomerID:** A unique identifier assigned to each customer. Each customer can have multiple transactions associated with their ID
- **Country:** The country where the customer resides.
- **Revenue:** The Revenue of each transaction. This represents the Quantity x Unit Price in GBP (British pounds).

ADDITIONAL CALCULATED VARIABLES REQUIRED FOR RFM ANALYSIS:

- **Recency:** The number of days since the customer's last purchase.
- **Frequency:** The total number of transactions made by the customer.
- **Monetary:** The total monetary value of the customer's purchases.

4.DATA PREPROCESSING STEPS:

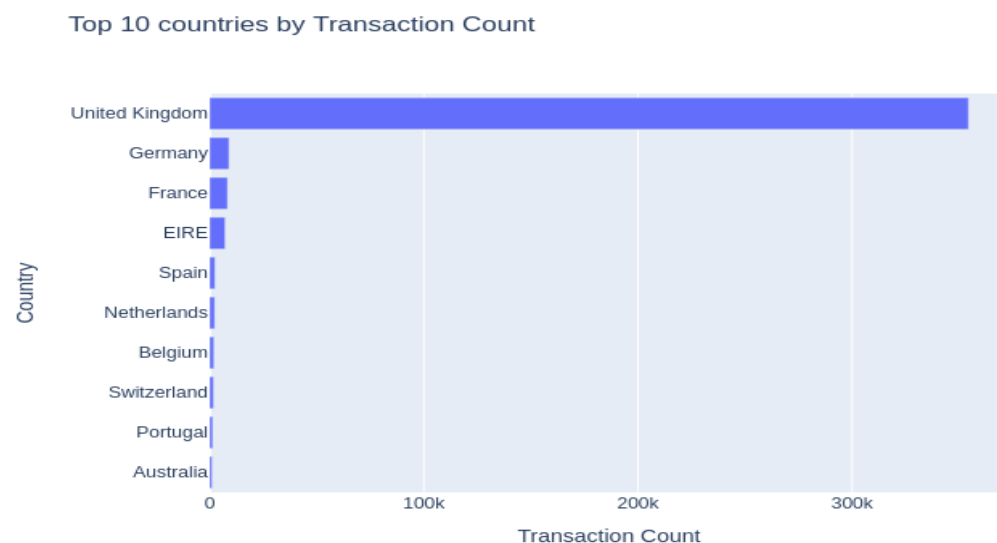
To ensure data quality and reliability, the following pre-processing steps were undertaken:

- **Checking Null Values:** Identified missing values, especially in the CustomerID column, which are critical for segmentation. Rows with missing CustomerID were dropped as they do not contribute to meaningful customer insights.
- **Handling Negative Values:** Transactions with negative quantities, typically representing returns or errors, were removed to focus on genuine purchase data.
- **Feature Engineering:** Created a Revenue feature by multiplying Quantity and UnitPrice to represent the monetary value of each transaction.
- **Date and Time Transformation:** Extracted features such as month, day, hour, and weekend vs. weekday to understand temporal purchasing patterns.

5.DATA VISUALIZATION INFERENCES:

1.Top 10 Countries by Transactions:

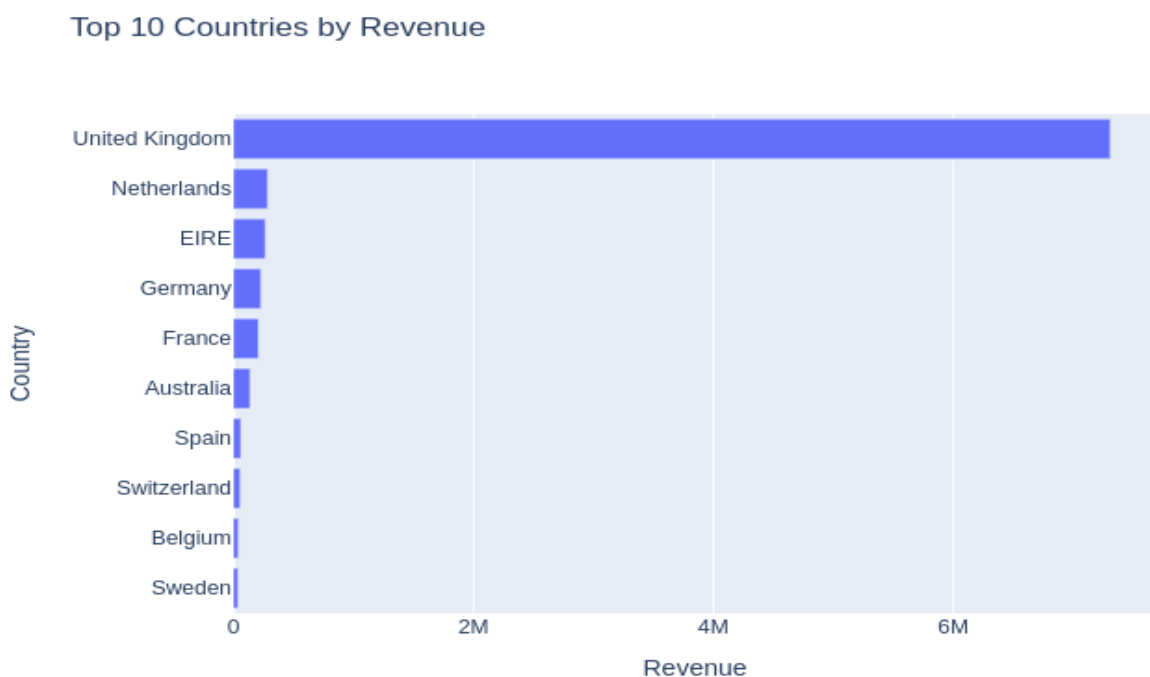
A bar plot displaying the countries with the highest number of transactions, highlighting key markets.



Inference: From the above bar plot, we can infer that that the highest transactions have been done by United Kingdom, The home town of the Company.

2. Top 10 Countries by Revenue:

A visualization to showcase the countries contributing the most to the total revenue.



Inferences:

1. United Kingdom has the highest revenue compared to other countries owing to highest transaction count.
2. Surprisingly, Netherlands which has less transaction count than Germany, France and EIRE have the highest revenue compared to these three countries.
3. Subsequently, Australia which stands in the bottom of top 10 countries by transactions has revenue better than Spain, Belgium and Switzerland.

3. Month-to-Month Transactions:

A Visualization showing transaction trends over the months, identifying peaks and troughs.



Inferences: The above bar graph infers that there has been a general increase in the number of transactions from the month of August and a peak increase in the month of November which shows there might be a seasonality factor too. There has been a significant decrease in the number of transactions in the month of December.

4. Month-to-Month Revenue:

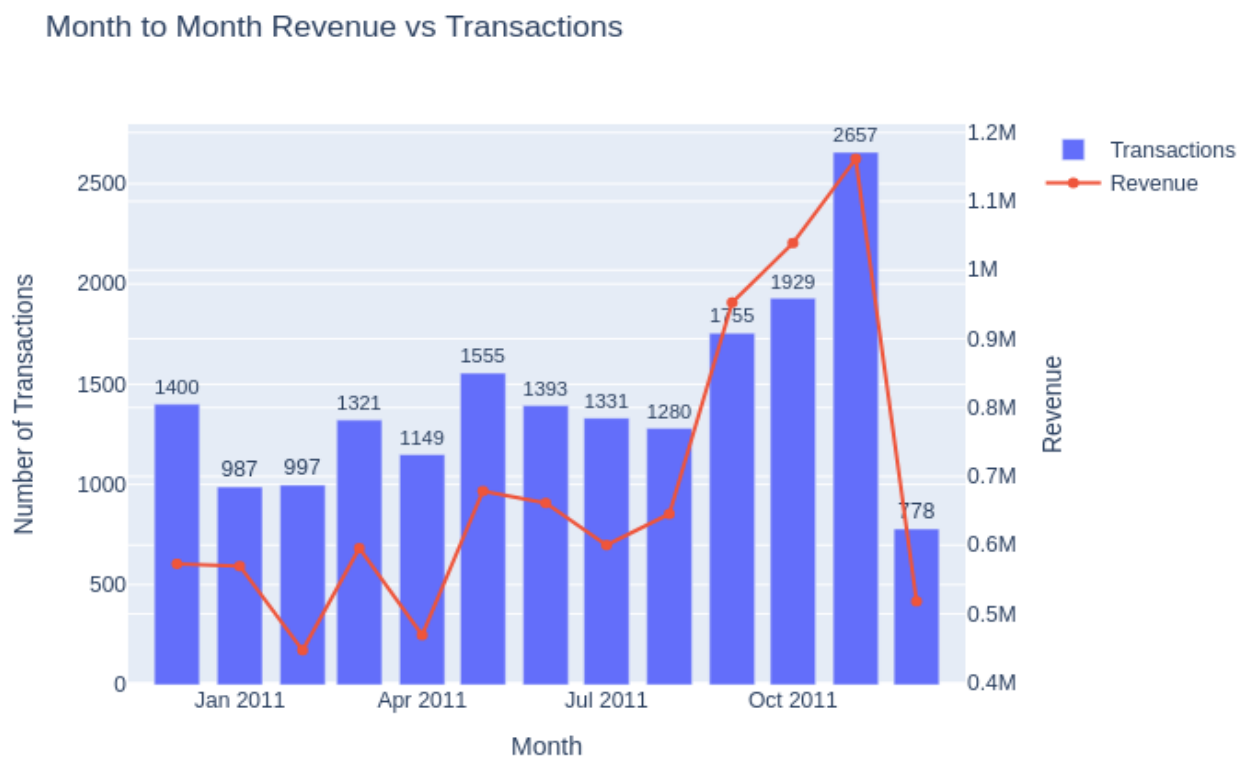
A Visualization showing revenue trends over the months, identifying peaks and troughs



Inference: The above visualization shows that the Revenue has been increased in the month of November similar to number of transactions and has been decreased sharply in the month of December.

5. Month-to-Month Revenue vs. Transactions:

A combined plot to analyse the relationship between the number of transactions and the revenue generated each month.

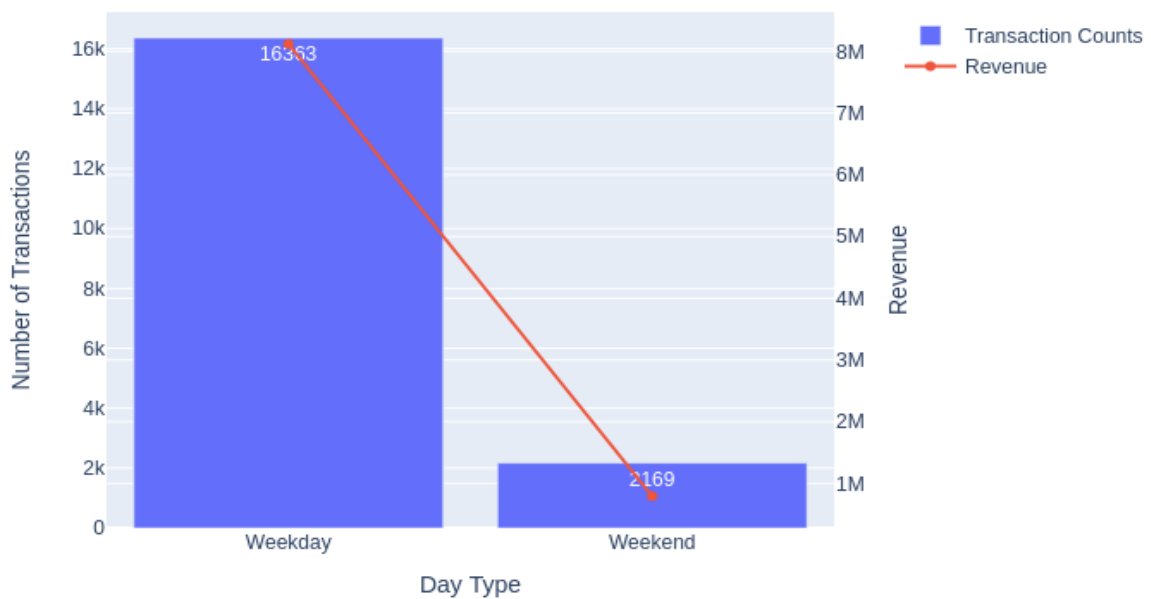


Inferences: From the above plot we can infer that both Transactions and Revenue went along during month-to-month observation.

6. Weekends vs. Weekdays Transactions and Revenue:

Insights into how purchasing behaviour changes on weekends compared to weekdays.

Weekends vs Weekdays comparison by Transactions and Revenues

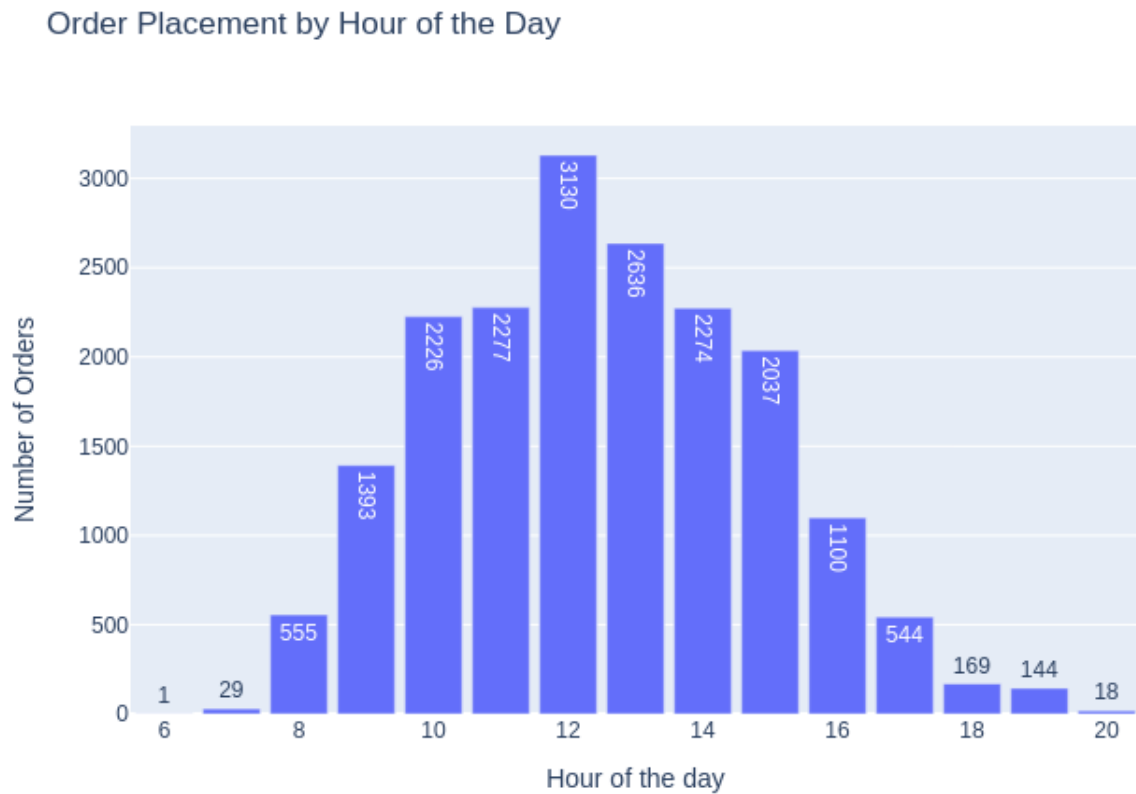


Inferences:

From the above visualization we can infer that There has been a higher activity on Weekdays. This suggests that customers are highly active in making Purchases during the workweek. The lower activity on the weekends could indicate an opportunity to increase sales by targeting marketing campaigns.

7.Order Placement by Hour:

Analysis of the time of day with peak order placements, aiding in targeted marketing strategies.



Inference:

From the above plot, we can say that the number of orders has been following Normal distribution with a greater number of orders at the 12th hour of the day.

6.RFM ANALYSIS:

STEPS PERFORMED:

1.Calculation of RFM Metrics:

- The RFM analysis starts with calculating three key metrics for each customer: Recency, Frequency, and Monetary value.
- Recency measures the number of days since a customer's last purchase, indicating how recently they engaged with the store.
- Frequency counts the total number of transactions made by a customer, reflecting their purchasing regularity.
- Monetary represents the total spending amount by the customer, showing their overall value to the business.
- These metrics provide a comprehensive view of customer behaviour by combining recent engagement, loyalty, and contribution to revenue.

2. RFM Scoring:

- After calculating the RFM metrics, each customer is scored on a scale of 1 to 5 for Recency, Frequency, and Monetary value.
- The scoring is typically done by dividing the metric values into quintiles or percentiles, where a higher score signifies better performance (e.g., a recency score of 5 means the customer has recently made a purchase, while a frequency score of 5 indicates they are frequent buyers).
- This scoring approach allows businesses to rank customers effectively, creating a detailed profile of their buying behaviour.
- The combined RFM score (e.g., RFM = 555) helps identify top-tier customers, such as frequent buyers who spend significantly and have made recent purchases.

3.Customer Segmentation:

- Based on the RFM scores, customers were segmented using specific rules implemented in the code. The segmentation logic classified customers into various groups based on their RFM scores.
- 'Champions' are customers who scored the highest in all three metrics (RFM score = 555), representing the most valuable and engaged customers. 'Loyal Customers' are those with high scores in one or more metrics, indicating consistent purchasing behaviour.
- 'Potential Loyalists' have slightly lower scores, showing promise of becoming loyal with targeted engagement.

- 'New Customers' scored well initially but have room for further engagement, while 'Need Attention' customers require proactive measures to encourage more frequent purchases.
- Lastly, 'At Risk' customers scored the lowest, indicating declining engagement and requiring re-engagement strategies. This segmentation helps in tailoring marketing strategies and optimizing customer relationship management.

4. Visualization of Segmented Customers:

- To gain insights into the customer segments, a bar graph was used to display the number of customers in each group.
- The visualization clearly shows the distribution of customers across segments such as 'Champions', 'Loyal Customers', 'Potential Loyalists', 'New Customers', 'Need Attention', and 'At Risk'.
- By examining the bar graph, it is easy to identify which segments have the highest and lowest customer counts.
- For instance, the 'Champions' and 'Loyal Customers' segments highlight a strong base of engaged and high-value customers, while the 'At Risk' and 'Need Attention' segments reveal opportunities for re-engagement and targeted marketing strategies.

This simple yet powerful visualization allows for a quick assessment of the customer landscape, helping businesses prioritize actions to nurture valuable customers and regain those at risk of churning.

7. ASSUMPTIONS:

1)The data is representative of typical purchasing behaviour, not influenced heavily by one-time events or promotions:

- The dataset is assumed to capture regular purchasing behaviour without significant distortion from special events such as flash sales, holiday promotions, or one-time discounts.
- This assumption is crucial as such events can lead to temporary spikes in sales or unusual buying patterns that do not reflect the usual behaviour of customers.
- The focus is on identifying consistent trends rather than short-term fluctuations that could mislead the segmentation results.
- It ensures that the insights derived from the data are applicable to everyday business decisions rather than being specific to special circumstances.
- By excluding the influence of these irregular events, the analysis remains relevant and actionable for routine customer behaviour management and strategy formulation.

2)Customers with missing IDs are considered unimportant for behavioural segmentation:

- Missing CustomerID values indicate transactions that cannot be attributed to specific individuals, making them unsuitable for behavioural analysis.
- These transactions are assumed to be minor in impact and are removed from the dataset to maintain the quality and accuracy of the segmentation process.
- Including unidentified transactions could dilute the effectiveness of RFM scoring, leading to less precise customer insights.
- This approach prioritizes the integrity of customer-related data, ensuring that the analysis focuses only on identifiable and trackable purchasing behaviour.
- The decision to exclude these entries helps streamline the segmentation, making it more accurate and meaningful by concentrating on complete and reliable data.

3)Negative quantities are exclusively return transactions and are not included in the final analysis:

- Negative quantities typically represent returns, cancellations, or corrections, which do not contribute positively to the understanding of purchasing behavior.
- These transactions are assumed to reflect non-purchasing activities that could skew the Monetary component of RFM analysis if included.
- Excluding negative quantities helps to ensure that the RFM metrics focus solely on actual purchases, enhancing the validity of customer segmentation.
- This assumption allows for a clear separation of buying actions from non-revenue-generating activities, leading to more accurate insights into customer value.

- By removing these anomalies, the analysis remains focused on positive engagements, thereby supporting better decision-making for marketing and customer retention strategies.

8. BEHAVIOURAL SEGMENTATION OF CUSTOMERS AND INTERPRETATIONS:

1) The below bar graph illustrates the segmentation of customers based on their RFM scores, effectively categorizing them into distinct groups such as 'Champions', 'Loyal Customers', 'Potential Loyalists', 'New Customers', 'Need Attention', and 'At Risk'.

2) This visualization provides a clear and concise overview of the distribution of customers within each segment, highlighting the relative size of each group.



1. Potential Loyalist (1147 customers):

- **Characteristics:** These customers are on the verge of becoming loyal customers. They have made recent purchases and have a moderate to high frequency of transactions.
- **Strategy:** Engage these customers with personalized marketing campaigns, loyalty programs, and incentives to convert them into loyal customers.

2. Loyal Customers (1130 customers):

- **Characteristics:** These customers have high frequency and high monetary value. They consistently make purchases and contribute significantly to the revenue.
- **Strategy:** Maintain and strengthen relationships with these customers by providing exclusive offers, early access to new products, and superior customer service to ensure their continued loyalty.

3. New Customers (889 customers):

- **Characteristics:** These are recent customers with a low to moderate number of purchases. They have recently engaged with the company and are in the early stages of the customer lifecycle.
- **Strategy:** Nurture these customers with welcome emails, onboarding programs, and targeted promotions to encourage repeat purchases and move them towards becoming loyal customers.

4. Need Attention (641 customers):

- **Characteristics:** These customers have a moderate frequency and monetary value but haven't made recent purchases. They are at risk of churning if not re-engaged.
- **Strategy:** Re-engage these customers with win-back campaigns, personalized offers, and reminders about the benefits of your products or services to rekindle their interest.

5. Champions (348 customers):

- **Characteristics:** These customers are the best and most valuable. They have high recency, frequency, and monetary values. They are the top spenders and frequent buyers.
- **Strategy:** Reward these customers with VIP programs, special recognition, and exclusive benefits to maintain their high level of satisfaction and encourage them to become brand advocates.

6. At Risk (183 customers):

- **Characteristics:** These customers have a high monetary value but have not made recent purchases. They are at high risk of churning and need immediate attention.
- **Strategy:** Implement targeted retention strategies such as personalized re-engagement campaigns, special discounts, and feedback requests to understand and address their reasons for disengagement.

9.BUSINESS INSIGHTS:

1.Enhancing Loyalty Among Potential Loyalists and Loyal Customers:

Since 'Potential Loyalists' and 'Loyal Customers' are the largest segments, prioritizing strategies to convert potential loyalists into fully loyal customers and maintaining the loyalty of current loyal customers can significantly boost revenue.

2. Implementing Re-engagement Strategies:

Segments like 'Need Attention' and 'At Risk' require dedicated re-engagement efforts to reduce the likelihood of churn and recover potential lost revenue.

3. Onboarding and Nurturing New Customers:

Onboarding and nurturing new customers is crucial to enhance their experience, build loyalty, and move them into more valuable segments over time.

4. Rewarding and Retaining Champion Customers:

Keeping champion customers happy through exclusive rewards and recognition is essential to encourage repeat purchases and foster positive word-of-mouth.

10.FUTURE POSSIBILITIES OF PROJECT:

1.Predictive Modelling: Enhancing Customer Retention and Purchase Likelihood

- Future work can involve developing advanced predictive models to forecast customer behaviour, such as churn and future purchase likelihood.
- By utilizing machine learning algorithms, businesses can identify early signs of disengagement and proactively address them through personalized retention strategies.
- Predictive models can also help in forecasting customer lifetime value, enabling the business to allocate resources more effectively towards high-value customers.

2. Integration with Marketing Systems: Automating Targeted Campaigns:

- Integrating customer segmentation data with marketing systems opens up the potential for automated and highly targeted marketing campaigns.
- By syncing segmentation insights with tools like CRM (Customer Relationship Management) and marketing automation platforms, businesses can deliver personalized content, offers, and recommendations to each customer segment.
- This seamless integration will help in optimizing marketing efforts, reducing manual intervention, and ensuring that the right message reaches the right customer at the right time, thereby enhancing the overall effectiveness of marketing strategies.

3. Real-Time Analytics: Adapting to Dynamic Customer Behaviour:

- Incorporating real-time data analytics into the project can significantly enhance its responsiveness to changing customer behaviour.
- By leveraging real-time data feeds, businesses can monitor and react to customer activities as they happen, such as browsing patterns, purchase behaviour, or even abandonment signals.
- Real-time analytics allows for immediate adjustments to marketing strategies, dynamic pricing, or instant personalized recommendations, creating a more agile and customer-centric approach.

11.REFERENCES:

Online Tools and Resources:

- **OpenAI. (2024). ChatGPT - Assisting in Data Analysis and Segmentation Techniques. Retrieved from <https://openai.com/chatgpt>.**
- **Google. (n.d.). Various Articles and Resources on Customer Segmentation and RFM Analysis. Retrieved from <https://www.google.com>.**
- **Kaggle. (n.d.). RFM Segmentation Analysis [Notebook].**