CUSTOMER SEGMENTATION ANALYSIS REPORT

Oasis Infobyte Task 1 Project 2

March 20th, 2025.

Bola Abimbola

Outline

1. Introduction

Customer segmentation is a key analytical approach used to identify distinct groups within a customer base. For this project, we performed an exploratory data analysis (EDA) and applied segmentation techniques to classify customers into different categories based on their purchasing behavior. The objective is to provide insights that can drive strategic marketing and customer retention effort

2. Data Overview

The dataset used for this analysis contains the following columns:
- **Customer ID** – Unique identifier for each customer
- **Date** – Transaction date
- **Product Category** – Category of the purchased item
- **Quantity** – Number of units purchased
- **Unit Price** – Price per unit of the product
- **Amount Spent** – Total amount spent per transaction
- **Gender** – Gender of the customer
- **Age** – Age of the customer
- **Location** – Geographical location of the customer

Customer ID	Gender	Age	Location	Purchase Frequency	Total Amount	F_Score	M_Score	Last Purchase Date	Recency	R_Score	RFM_Score
CUST029	Female	64	Miami, FL	34	29871	2	2	22/01/2025	70	1	
CUST015	Female	54	Los Angeles, CA	25	20580	1	2	10/02/2025	51	2	
CUST013	Male	61	Los Angeles, CA	33	27927	2	2	21/01/2025	71	1	
CUST001	Male	38	Houston, TX	49	35453	3	3	11/02/2025	50	2	
CUST025	Female	43	New York, NY	33	30453	2	3	18/02/2025	43	2	
CUST007	Male	44	Houston, TX	31	24162	2	2	03/01/2025	89	1	
CUST009	Male	45	Los Angeles, CA	28	19972	1	1	13/03/2025	20	3	
CUST024	Female	26	Los Angeles, CA	30	21038	2	2	26/02/2025	35	2	
CUST012	Male	64	Los Angeles, CA	33	20031	2	2	25/02/2025	36	2	
CUST008	Male	59	Los Angeles, CA	29	22733	1	2	06/03/2025	27	3	
CUST011	Male	32	Miami, FL	22	16437	1	1	19/02/2025	42	2	
CUST019	Female	56	New York, NY	36	31934	2	3	02/02/2025	59	2	
CUST017	Female	38	Houston, TX	41	32974	3	3	08/02/2025	53	2	
CUST003	Male	29	Chicago, IL	43	30219	3	3	15/02/2025	46	2	
CUST027	Female	37	Houston, TX	29	22904	1	2	15/03/2025	18	3	
CUST005	Male	61	Houston, TX	43	36412	3	3	12/03/2025	21	3	
CUST010	Female	33	Houston, TX	27	17551	1	1	21/02/2025	40	2	
CUST026	Female	19	Houston, TX	45	33685	3	3	07/03/2025	26	3	
CUST028	Female	45	Houston, TX	45	34617	3	3	11/03/2025	22	3	
CUST002	Female	50	Miami, FL	30	22112	2	2	21/11/2024	132	1	
CUST016	Male	24	Los Angeles, CA	38	30001	2	3	22/02/2025	39	2	
CUST023	Female	31	Miami, FL	36	34474	2	3	17/02/2025	44	2	

Overview of Customer Table

Order ID	Customer ID	Date	Product Category	Quantity	Unit Price	Amount Spent	Gender	Age	Location	DateOnly	Mon
ORD0004	CUST001	Wednesday, 4 January 2023	Clothing	5	100	500	Male	38	Houston, TX	4/1/2023	Ja
ORD0009	CUST001	Monday, 9 January 2023	Beauty	1	151	151	Male	38	Houston, TX	9/1/2023	Ja
ORD0015	CUST017	Sunday, 15 January 2023	Sports	4	39	156	Female	38	Houston, TX	15/1/2023	Ja
ORD0019	CUST001	Thursday, 19 January 2023	Electronics	5	290	1450	Male	38	Houston, TX	19/1/2023	Ja
ORD0032	CUST001	Wednesday, 1 February 2023	Beauty	4	435	1740	Male	38	Houston, TX	1/2/2023	Fe
ORD0045	CUST001	Tuesday, 14 February 2023	Clothing	2	130	260	Male	38	Houston, TX	14/2/2023	Fe
ORD0051	CUST001	Monday, 20 February 2023	Electronics	2	446	892	Male	38	Houston, TX	20/2/2023	Fe
ORD0067	CUST017	Wednesday, 8 March 2023	Clothing	1	498	498	Female	38	Houston, TX	8/3/2023	M
ORD0107	CUST001	Monday, 17 April 2023	Electronics	1	210	210	Male	38	Houston, TX	17/4/2023	Ap
ORD0108	CUST001	Tuesday, 18 April 2023	Clothing	4	454	1816	Male	38	Houston, TX	18/4/2023	Ap
ORD0124	CUST001	Thursday, 4 May 2023	Home & Kitchen	2	85	170	Male	38	Houston, TX	4/5/2023	M
ORD0135	CUST017	Monday, 15 May 2023	Home & Kitchen	4	30	120	Female	38	Houston, TX	15/5/2023	M
ORD0137	CUST017	Wednesday, 17 May 2023	Sports	2	47	94	Female	38	Houston, TX	17/5/2023	M
ORD0145	CUST017	Thursday, 25 May 2023	Home & Kitchen	3	37	111	Female	38	Houston, TX	25/5/2023	M
ORD0147	CUST017	Saturday, 27 May 2023	Beauty	5	415	2075	Female	38	Houston, TX	27/5/2023	M
ORD0149	CUST017	Monday, 29 May 2023	Sports	5	438	2190	Female	38	Houston, TX	29/5/2023	M
ORD0151	CUST017	Wednesday, 31 May 2023	Beauty	1	351	351	Female	38	Houston, TX	31/5/2023	M
ORD0158	CUST001	Wednesday, 7 June 2023	Home & Kitchen	3	266	798	Male	38	Houston, TX	7/6/2023	Ju
ORD0159	CUST001	Thursday, 8 June 2023	Electronics	2	81	162	Male	38	Houston, TX	8/6/2023	Ju
ORD0173	CUST017	Thursday, 22 June 2023	Beauty	5	476	2380	Female	38	Houston, TX	22/6/2023	Ju
ORD0184	CUST001	Monday, 3 July 2023	Sports	2	300	600	Male	38	Houston, TX	3/7/2023	Ju
ORD0204	CUST001	Sunday, 23 July 2023	Sports	1	350	350	Male	38	Houston, TX	23/7/2023	Ju

Overview of Order Table

3. Data Cleaning and Preprocessing

To ensure data quality and reliability, the following preprocessing steps were carried out:
1. Handling Missing Values: Checked for missing or null values in key columns (e.g., Amount Spend, Customer ID, Date)
2. Date Formatting: Extracted Month from the Date column to enable time-based analysis using the formula:

```
1  Month = FORMAT([Date], "mmm")
```

3. Duplicate Removal: Removed duplicate entries to avoid skewed analysis

4. Customer Segmentation Approach

We applied the **RFM (Recency, Frequency, Monetary) Analysis**, a popular technique for segmenting customers based on their transaction history:
• **Recency (R)** – How recently a customer made a purchase.
• **Frequency (F)** – How often a customer makes a purchase.
• **Monetary (M)** – Total amount spent by the customer.

4.1 Calculating RFM Scores

1.Recency Score:

1. Calculated the number of days since the last purchase.
2. Customers with the most recent transactions received higher scores.

```
R_Score =
IF([Recency] <= 30, 3,
    IF([Recency] <= 60, 2, 1))
```

2.Frequency Score:

1. Counted the total number of transactions per customer.
2. Higher transaction counts received higher scores.

```
F_Score =
IF(CustomerData[Purchase Frequency] >= 40, 3,
    IF(CustomerData[Purchase Frequency] >= 30, 2, 1))
```

3.Monetary Score:

1. Summed up the total spending per customer.
2. Customers who spent more received higher scores.

```
R_Score =
IF([Recency] <= 30, 3,
    IF([Recency] <= 60, 2, 1))
```

4. RFM Score Calculation:

The final **RFM Score** was computed by summing **R, F** and **M Score**
Customers were classified into three segments based on their scores:

```
RFM_Score = [R_Score] + [F_Score] + [M_Score]
```

High-Value Customers (Highest RFM Scores)
Loyal Customers (Moderate RFM Scores)
At-Risk Customers (Lowest RFM Scores)

```
Customer_Segment =
SWITCH(TRUE(),
        [RFM_Score] >= 8, "High-Value Customer",
        [RFM_Score] >= 6, "Loyal Customer",
        [RFM_Score] >= 4, "At-Risk Customer",
        "Lost Customer")

```

5. Data Visualization & Dashboard Creation

The dashboard was created in **Power BI** to visualize customer segmentation insights. The key components include:

5.1 Filters and KPIs

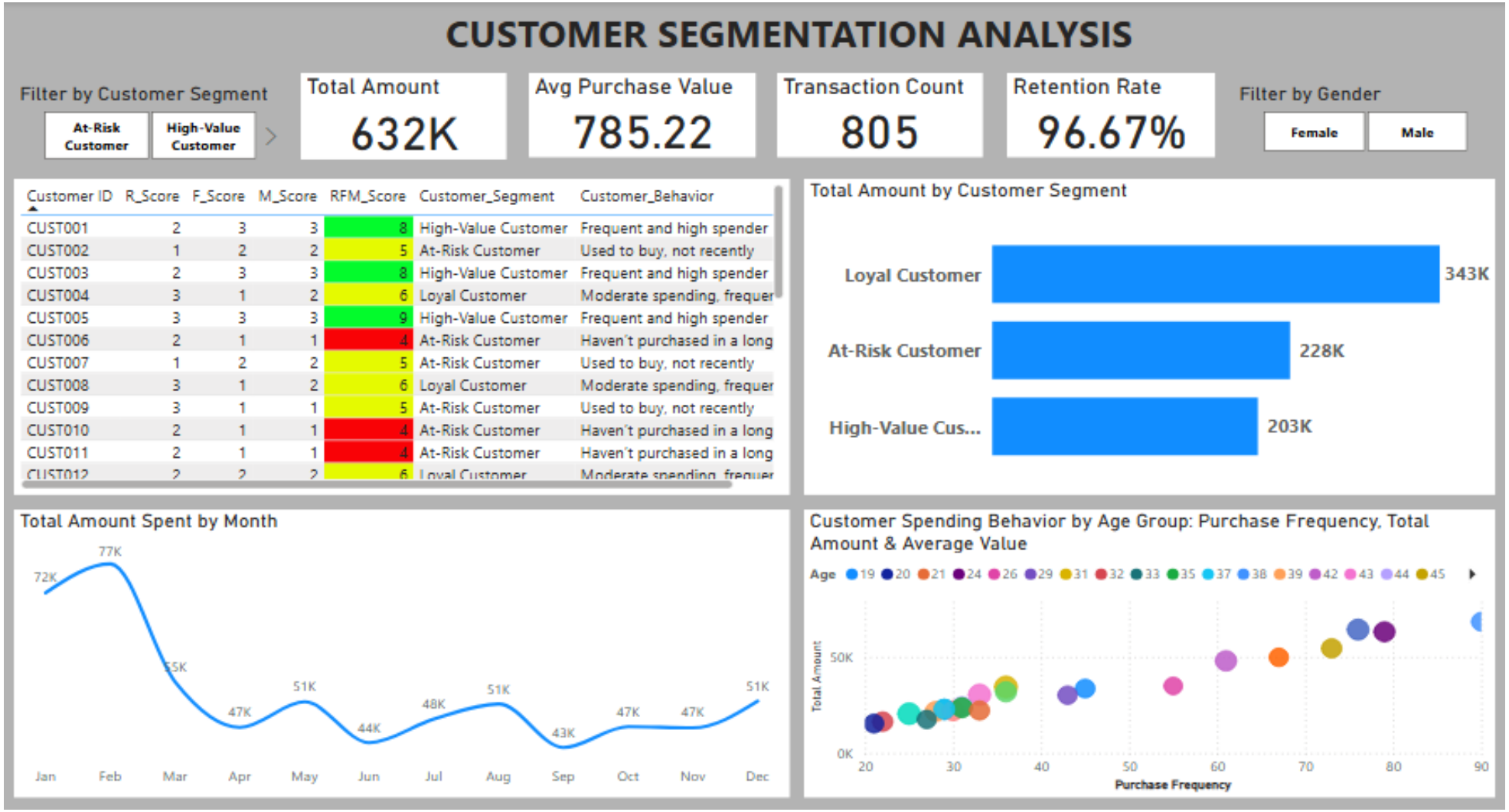
- **Filters:**
  - Customer Segment (High-Value, Loyal, At-Risk)
  - Gender Filter (Male, Female)

- **KPI's:**
  - Total Amount Spent
  - Average Purchase Value
  - Transaction Count
  - Retention Rate

5.2 Key Visualizations

- **Customer Segmentation Table**: Displays Customer ID, RFM scores, and assigned segment.
- **Total Amount by Customer Segment**: A bar chart comparing spending across customer groups.
- **Trend Analysis**: A line chart showing the **Sum of Amount Spent by Month**.
- **Purchase Frequency vs. Total Spending**: A scatter plot analyzing the correlation between frequency and spending habits.

5.3 Dashboard Overview



Dashboard Overview

Link to Dashboard: https://app.powerbi.com/view?r=eyJrIjoiYzNiZDBmMTgtNDA2Ny00NWMyLTkwYjQtOWEzZTRjMWY4NzgwIiwidCI6ImYyZGM3MTdkLTBlZmEtNGIzYi04N2VmLWU3MTBhMTMwOTVlNSJ9

6. Findings & Insights

6.1 Segment Performance

- **Loyal Customers contribute the most to revenue**, followed by At-Risk and High-Value Customers.
- **At-Risk Customers require retention efforts** to prevent churn.

6 .2 Spending Behavior

- Spending decreases over time, with a noticeable drop in later months.
- Customers who purchase frequently tend to have higher total spending.

7. Recommendations

Based on the analysis, we propose the following actions:

- **For High-Value Customers:**
    - Offer exclusive deals and loyalty rewards.
    - Provide early access to new products.
- **For Loyal Customers:**
    - Engage them with targeted marketing campaigns to maintain their status.
    - Offer personalized recommendations.
- **For At-Risk Customers:**
    - Implement re-engagement strategies like personalized discounts.
    - Send reminder emails and follow-ups to encourage repeat purchases.

8. Conclusion

This analysis provides a structured approach to customer segmentation using RFM analysis. The Power BI dashboard enables data-driven decision-making, helping businesses optimize marketing strategies, improve customer retention, and maximize revenue growth.

Prepared by: Bola