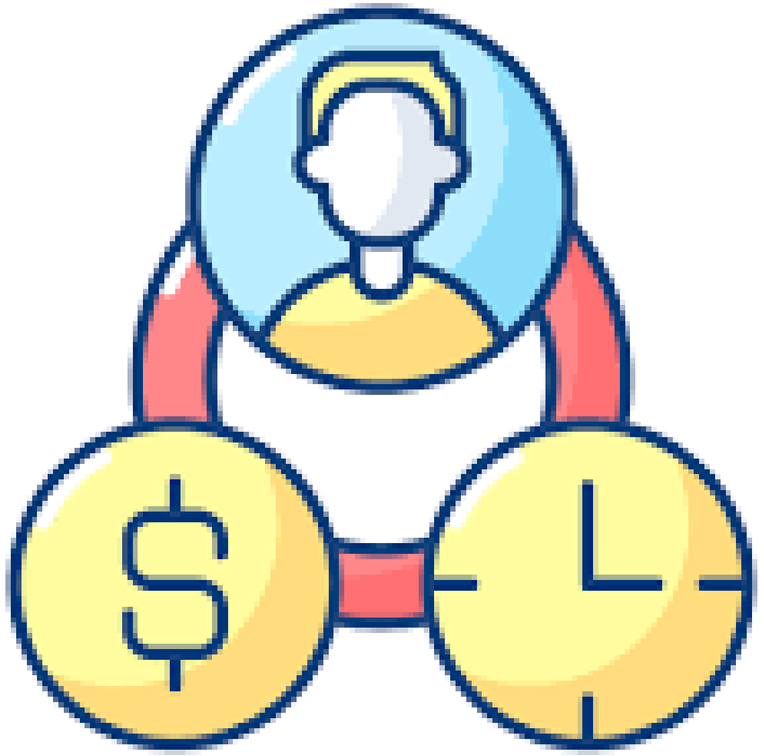


RFM SEGMENTATION ANALYSIS

A Comprehensive Analysis of Customer Value

ALISHBA RIZWAN

JUNE 2024

WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/

INTRODUCTION

Understanding customer behavior is essential for targeted marketing and customer relationship management.

The goal of this analysis is to segment customers based on their purchasing behavior using the RFM (Recency, Frequency, Monetary) model.

WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/

WHAT IS RFM SEGMENTATION?

01 Recency (R): How recently a customer made a purchase.

02 Frequency (F): How often a customer makes a purchase.

03 Monetary (M): How much money a customer spends on purchases.

RFM segmentation helps in identifying valuable customer segments for targeted marketing strategies.

DATASET OVERVIEW

01

The dataset used for this analysis includes customer transactions. Key Columns: Customer ID, Invoice Date, Quantity, Unit Price.

02

Preprocessing: Handling missing values, converting dates to the correct format, and calculating subtotals for each invoice.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerI	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/2010 8:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/2010 8:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	12/1/2010 8:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	12/1/2010 8:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/2010 8:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/2010 8:34	2.1	13047	United Kingdom

DATA PREPROCESSING

STEPS TAKEN

- Converted invoice dates to Date objects.
- Removed any duplicate or irrelevant entries.
- Ensured all numerical values are in a consistent format.

```r
# Load required libraries
library(tidyverse)
library(lubridate)
library(janitor)

# Load the dataset
df_salesData <- read.csv("E:/Data Analytics/df_sales_data.csv")

# Clean column names
df_salesData <- clean_names(df_salesData)

# Convert 'invoice_date' to a Date object
df_salesData$invoice_date <- as.Date(df_salesData$invoice_date, format = "%m/%d/%Y")

# Print the first few rows to check the data
print(head(df_salesData))

# Find the latest date in the dataset
latest_date <- max(df_salesData$invoice_date, na.rm = TRUE)
print(latest_date)
```

RFM CALCULATION

RECENCY: CALCULATED AS THE NUMBER OF DAYS SINCE THE LAST PURCHASE.

FREQUENCY: CALCULATED AS THE NUMBER OF UNIQUE INVOICES PER CUSTOMER.

MONETARY: CALCULATED AS THE TOTAL AMOUNT SPENT BY EACH CUSTOMER.

EXAMPLE CALCULATION: DETAILED EXAMPLE OF HOW EACH METRIC IS CALCULATED FOR A SAMPLE CUSTOMER.

```
Calculate recency, frequency, monetary value, and additional metrics
fm_data <- df_salesData %>%
group_by(customer_id) %>%
summarise(
  recency = as.numeric(difftime(latest_date, max(invoice_date), units = "days")),
  frequency = n_distinct(invoice_no),
  monetary = sum(quantity * unit_price),
  avg_order_value = mean(quantity * unit_price, na.rm = TRUE),
  total_items = sum(quantity, na.rm = TRUE),
  avg_days_between_purchases = ifelse(n_distinct(invoice_no) > 1,
   mean(difftime(sort(unique(invoice_date)), lag(sort(unique(invoice_date))), units =
  NA),
  clv = sum(quantity * unit_price, na.rm = TRUE) / n_distinct(invoice_no),
  first_purchase = min(invoice_date, na.rm = TRUE),
  customer_tenure = as.numeric(difftime(latest_date, min(invoice_date, na.rm = TRUE)
  avg_cart_size = mean(quantity, na.rm = TRUE),
  repeat_purchase_rate = ifelse(n_distinct(invoice_no) > 1, 1, 0)
) %>%
ungroup()
```

RFM CALCULATION

AVERAGE DAYS BETWEEN PURCHASES: MEASURES THE AVERAGE TIME BETWEEN REPEAT PURCHASES.
CUSTOMER LIFETIME VALUE (CLV): AVERAGE MONETARY VALUE OF PURCHASES DIVIDED BY THE NUMBER OF INVOICES.
DISCOUNT UTILIZATION RATE: PERCENTAGE OF TRANSACTIONS WHERE DISCOUNTS WERE USED.

```
  NA),
  clv = sum(quantity * unit_price, na.rm = TRUE) / n_distinct(invoice_no),
  first_purchase = min(invoice_date, na.rm = TRUE),
  customer_tenure = as.numeric(difftime(latest_date, min(invoice_date, na.rm = TRUE), units = "days")),
  avg_cart_size = mean(quantity, na.rm = TRUE),
  repeat_purchase_rate = ifelse(n_distinct(invoice_no) > 1, 1, 0)
) %>%
ungroup()
```

CREATING QUARTILES

Purpose: Quartiles help in ranking customers within each RFM metric.
Method: Each RFM metric is divided into four quartiles.
Quartile Ranges: Explanation of how quartile ranges are determined for each metric.

```
# Create quartiles for each RFM metric
rfm_data <- rfm_data %>%
  mutate(
    recency_quartile = ntile(recency, 4),
    frequency_quartile = ntile(frequency, 4),
    monetary_quartile = ntile(monetary, 4),
    avg_order_value_quartile = ntile(avg_order_value, 4),
    total_items_quartile = ntile(total_items, 4),
    clv_quartile = ntile(clv, 4),
    customer_tenure_quartile = ntile(customer_tenure, 4)
  )
```

RFM SEGMENTATION

COMBINING QUARTILES: RFM SCORES ARE CREATED BY COMBINING THE QUARTILES OF EACH METRIC.

Segments: Customers are segmented based on their RFM scores.

Example Segments: Champions, Loyal Customers, Potential Loyalists, etc.

```
rfm_data <- rfm_data %>%
  mutate(
    RFM_Score = paste(recency_quartile, frequency_quartile, monetary_quartile, sep = ""),
    RFM_Segment = case_when(
      RFM_Score %in% c("444", "443", "434", "433", "344", "343", "334") ~ "Champions",
      RFM_Score %in% c("442", "441", "432", "431", "424", "423", "422", "414", "413", "412", "411") ~ "Loyal Customers",
      RFM_Score %in% c("324", "323", "322", "314", "313", "312", "311") ~ "Potential Loyalists",
      RFM_Score %in% c("441", "431", "421", "411") ~ "New Customers",
      RFM_Score %in% c("344", "343", "334", "324", "323", "322") ~ "Promising",
      RFM_Score %in% c("242", "241", "232", "231", "224", "223", "222", "214", "213", "212", "211") ~ "Need Attention",
      RFM_Score %in% c("432", "431", "422", "421", "412", "411") ~ "About to Sleep",
      RFM_Score %in% c("144", "143", "142", "141", "134", "133", "132", "131", "124", "123", "122", "121") ~ "Churn Risk",
      RFM_Score %in% c("244", "243", "242", "241", "234", "233", "232", "231") ~ "High Spending New Customers",
      RFM_Score %in% c("111") ~ "Lost Low-Value Customers",
      RFM_Score %in% c("211", "212", "213") ~ "One-Time High Spenders",
      TRUE ~ "Other"
    )
```

WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/

DATA FRAME

RFM with addional metrices Alishba.R* | rfm_data | RFM Analysis Code.R | Untitled1*

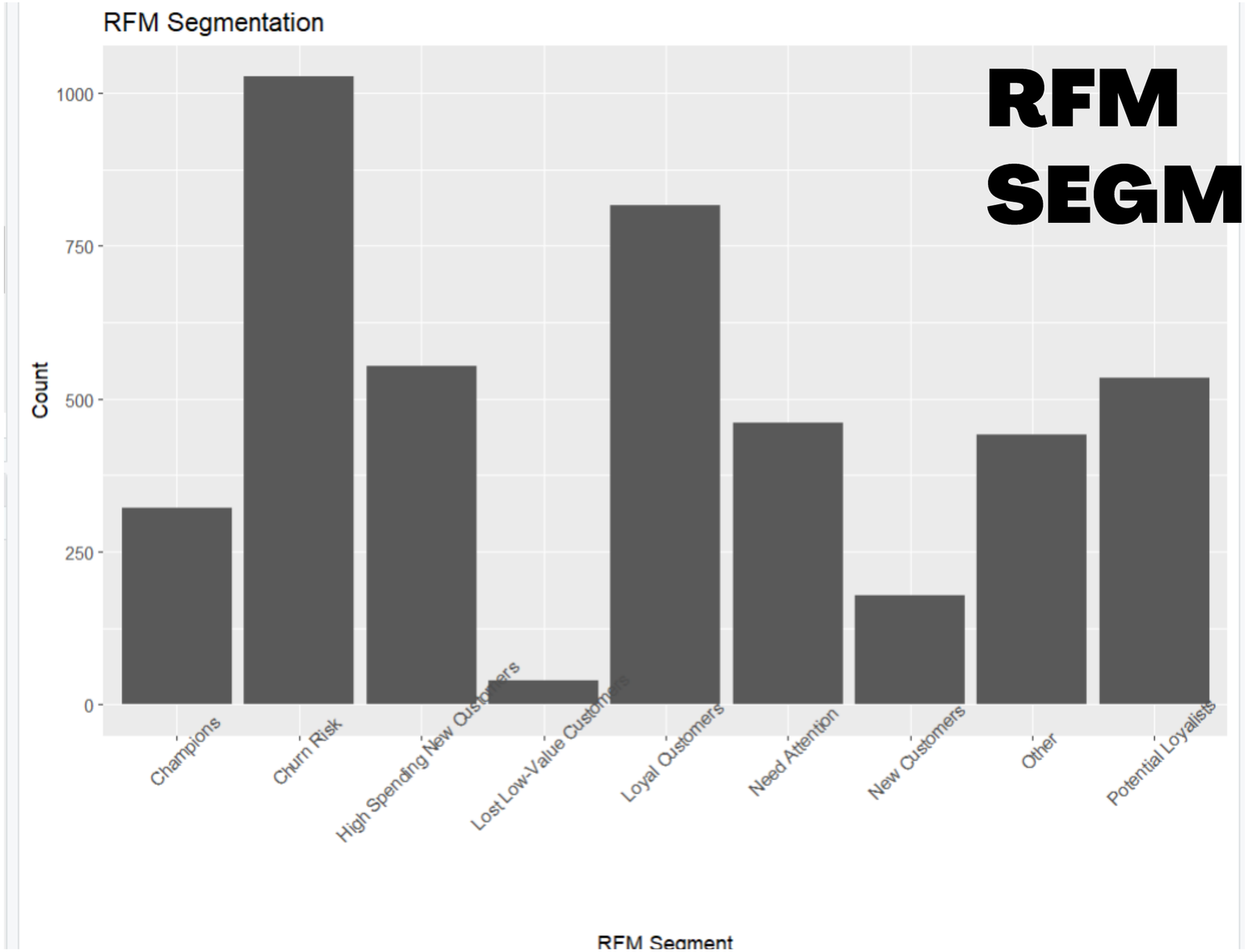
Filter

	customer_id	recency	frequency	monetary	avg_order_value	total_items	avg_days_between_purchases	clv	first_purchase	customer_tenure	avg_cart_size
1	12346	325	2	0.00	0.000000	0	NaN	0.00000	2011-01-18	325	0.000000
2	12347	2	7	4310.00	23.681319	2458	60.833333	615.71429	2010-12-07	367	13.505495
3	12348	75	4	1797.24	57.975484	2341	94.333333	449.31000	2010-12-16	358	75.516129
4	12349	18	1	1757.55	24.076027	631	NA	1757.55000	2011-11-21	18	8.643836
5	12350	310	1	334.40	19.670588	197	NA	334.40000	2011-02-02	310	11.588235
6	12352	36	11	1545.41	16.267474	470	43.333333	140.49182	2011-02-16	296	4.947368
7	12353	204	1	89.00	22.250000	20	NA	89.00000	2011-05-19	204	5.000000
8	12354	232	1	1079.40	18.610345	530	NA	1079.40000	2011-04-21	232	9.137931
9	12355	214	1	459.40	35.338462	240	NA	459.40000	2011-05-09	214	18.461538
10	12356	22	3	2811.43	47.651356	1591	151.500000	937.14333	2011-01-18	325	26.966102
11	12357	33	1	6207.67	47.386794	2708	NA	6207.67000	2011-11-06	33	20.671756
12	12358	1	2	1168.06	61.476842	248	149.000000	584.03000	2011-07-12	150	13.052632
13	12359	7	6	6245.53	24.588701	1612	64.800000	1040.92167	2011-01-12	331	6.346457
14	12360	52	3	2662.06	20.636124	1165	74.000000	887.35333	2011-05-23	200	9.031008
15	12361	287	1	189.90	18.990000	91	NA	189.90000	2011-02-25	287	9.100000
16	12362	3	13	5154.58	18.812336	2212	24.333333	396.50615	2011-02-17	295	8.072993
17	12363	109	2	552.00	24.000000	408	133.000000	276.00000	2011-04-11	242	17.739130
18	12364	7	4	1313.10	15.448235	1506	35.000000	328.27500	2011-08-19	112	17.717647
19	12365	291	3	320.69	13.943043	173	NaN	106.89667	2011-02-21	291	7.521739
20	12367	4	1	168.90	15.354545	173	NA	168.90000	2011-12-05	4	15.727273
21	12370	51	4	3545.69	21.231677	2353	103.000000	886.42250	2010-12-14	360	14.089820
22	12371	44	2	1887.96	29.967619	591	15.000000	943.98000	2011-10-11	59	9.380952

DATA FRAME

repeat_purchase_rate	recency_quartile	frequency_quartile	monetary_quartile	avg_order_value_quartile	total_items_quartile	clv_quartile	customer_tenure_quartile	RFM_Score	RFM_Segment
1	4	2	1	1	1	1	3	421	New Customers
1	1	4	4	4	4	4	4	144	Churn Risk
1	3	3	4	4	4	4	4	334	Champions
0	2	1	4	4	3	4	1	214	Need Attention
0	4	1	2	3	2	3	3	412	Loyal Customers
1	2	4	3	2	3	1	3	243	High Spending New Customers
0	4	1	1	3	1	1	2	411	Loyal Customers
0	4	1	3	3	3	4	2	413	Loyal Customers
0	4	1	2	4	2	4	2	412	Loyal Customers
1	2	2	4	4	4	4	3	224	Need Attention
0	2	1	4	4	4	4	1	214	Need Attention
1	1	2	3	4	2	4	2	123	Churn Risk
1	1	4	4	4	4	4	3	144	Churn Risk
1	3	2	4	3	4	4	2	324	Potential Loyalists
0	4	1	1	3	1	2	3	411	Loyal Customers
1	1	4	4	3	4	4	3	144	Churn Risk
1	3	2	2	4	3	3	2	322	Potential Loyalists
1	1	3	3	2	4	3	1	133	Churn Risk
1	4	2	2	2	2	1	3	422	Loyal Customers
0	1	1	1	2	2	2	1	111	Lost Low-Value Customers

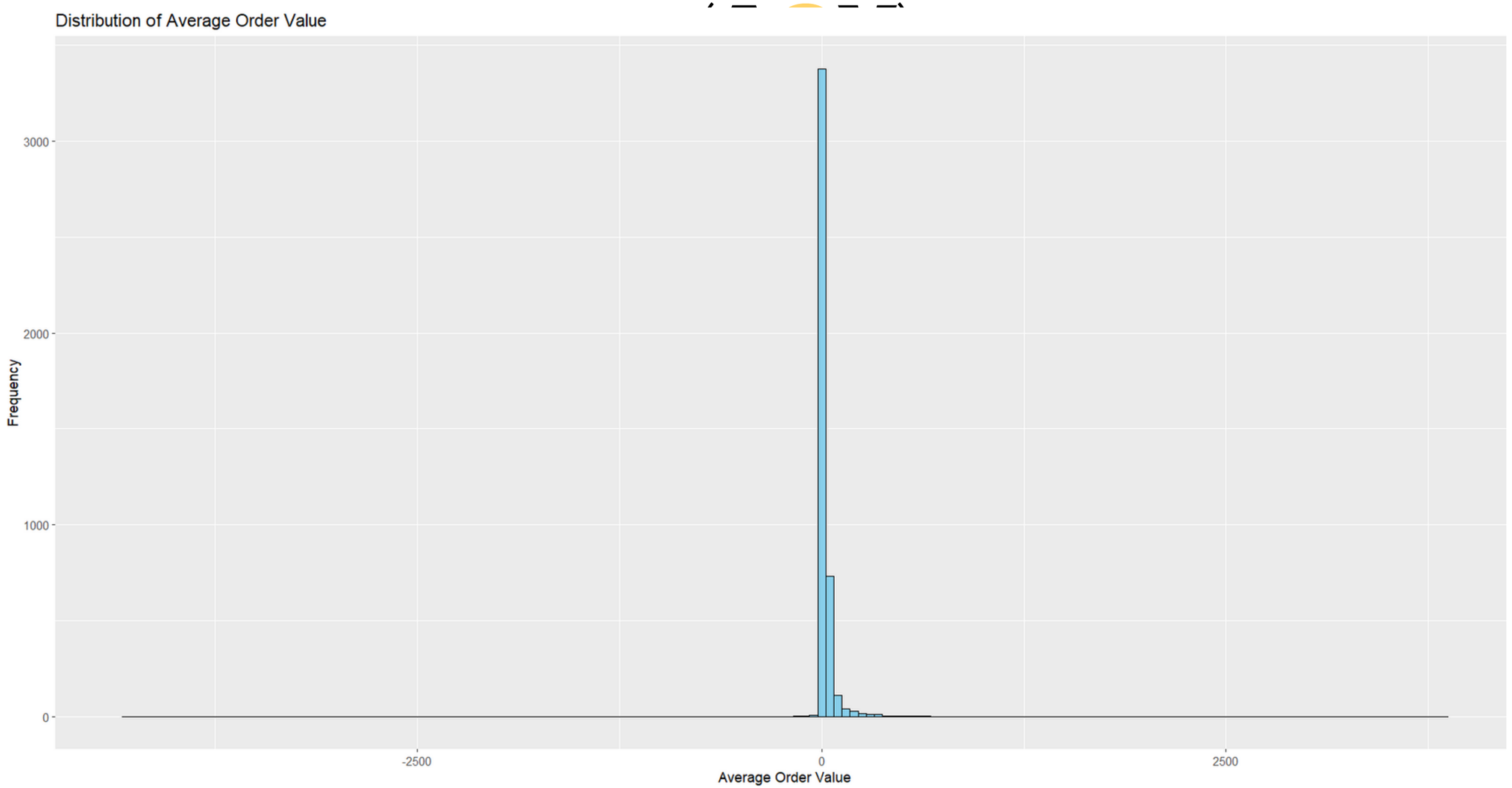
WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/



RFM SEGMENTATION

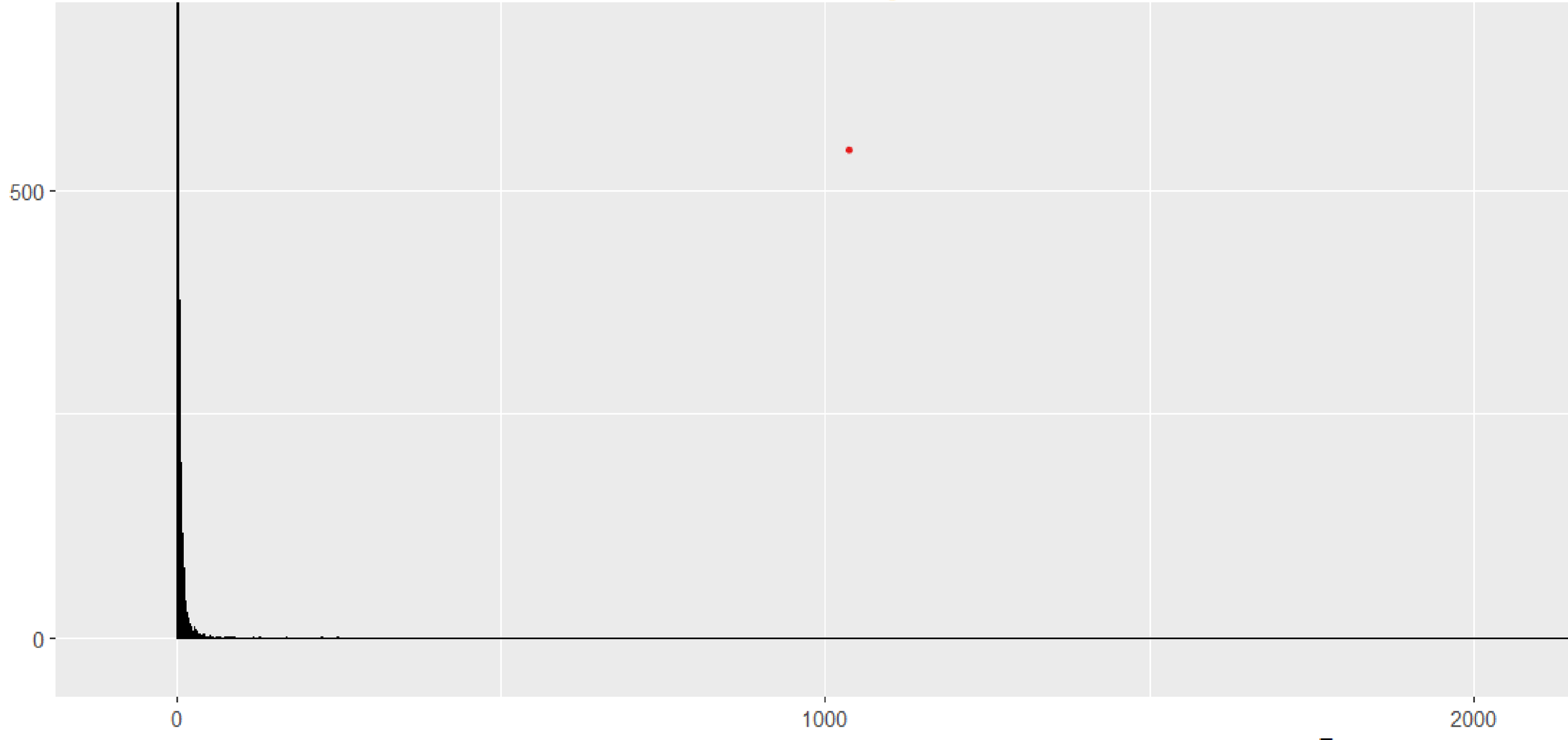
WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/

AVERAGE ORDER VALUE



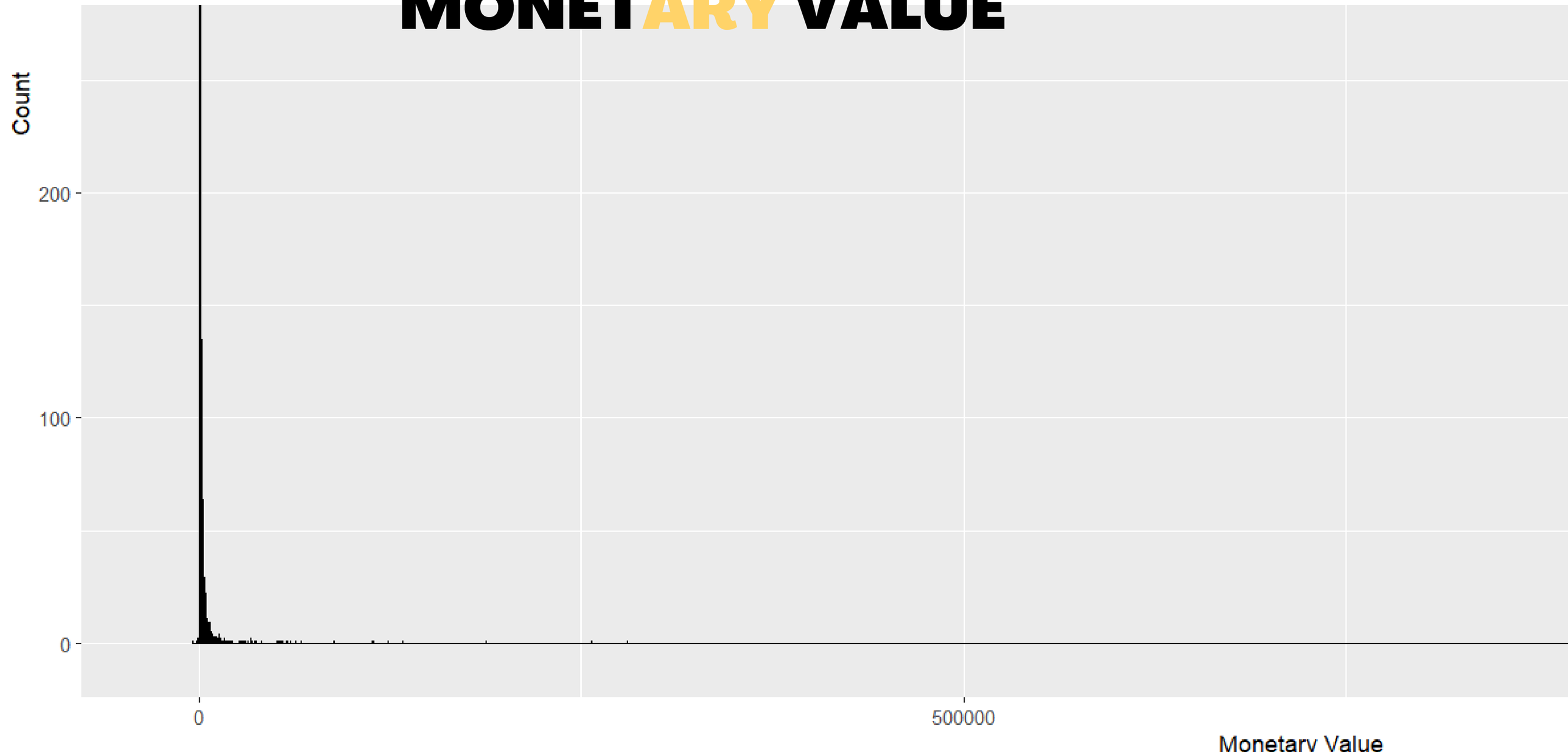
WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/

PURCHASE FREQUENCY



WWW.LINKEDIN.COM/IN/ALISHBA-RIZWAN--/

MONETARY VALUE



KEY INSIGHTS FROM RFM SEGMENTATION

- **IDENTIFICATION OF HIGH-VALUE CUSTOMERS:** RFM SEGMENTATION HIGHLIGHTS CUSTOMERS WHO FREQUENTLY PURCHASE, SPEND MORE, AND HAVE RECENTLY INTERACTED WITH THE BUSINESS.

- **TARGETED MARKETING STRATEGIES:** DIFFERENT CUSTOMER SEGMENTS REQUIRE DIFFERENT MARKETING APPROACHES.

- **EFFICIENT RESOURCE ALLOCATION:** BY FOCUSING MARKETING EFFORTS ON SPECIFIC CUSTOMER SEGMENTS, BUSINESSES CAN ALLOCATE RESOURCES MORE EFFICIENTLY.

RECOMMENDATIONS FOR MARKETING STRATEGIES:

1. **CHAMPIONS AND LOYAL CUSTOMERS:** IMPLEMENT LOYALTY PROGRAMS OR EXCLUSIVE OFFERS.

2. **POTENTIAL LOYALISTS**: OFFER INCENTIVES SUCH AS DISCOUNTS OR SPECIAL PROMOTIONS.

3. **NEW CUSTOMERS:** PROVIDE ONBOARDING EXPERIENCES.

4. **PROMISING CUSTOMERS**: NURTURE THESE CUSTOMERS WITH TARGETED PROMOTIONS.

5. **AT-RISK CUSTOMERS:** DEVELOP RE-ENGAGEMENT CAMPAIGNS.

RECOMMENDATIONS FOR MARKETING STRATEGIES (CONT.)

- **EXPAND ANALYSIS:** INCORPORATE ADDITIONAL METRICS SUCH AS AVERAGE ORDER VALUE, PURCHASE FREQUENCY, AND CUSTOMER LIFETIME VALUE TO GAIN DEEPER INSIGHTS INTO CUSTOMER BEHAVIOR.

- **REFINE SEGMENTATION:** CONTINUOUSLY MONITOR AND REFINE CUSTOMER SEGMENTS TO ADAPT TO CHANGING CUSTOMER BEHAVIORS AND MARKET CONDITIONS.

- **IMPLEMENT FEEDBACK LOOPS:** USE CUSTOMER FEEDBACK TO ENHANCE PRODUCT OFFERINGS AND CUSTOMER SERVICE, ENSURING CONTINUOUS IMPROVEMENT IN CUSTOMER SATISFACTION AND LOYALTY.

CONCLUSION

IN CONCLUSION, RFM SEGMENTATION PROVIDES A STRATEGIC FRAMEWORK FOR UNDERSTANDING AND OPTIMIZING CUSTOMER RELATIONSHIPS. BY LEVERAGING THE INSIGHTS GAINED FROM RFM ANALYSIS, BUSINESSES CAN IMPLEMENT TARGETED MARKETING STRATEGIES, IMPROVE CUSTOMER RETENTION, AND DRIVE REVENUE GROWTH.

THANK YOU!

FOR QUESTIONS OR COLLABORATION, PLEASE CONTACT ME AT:

Alishba Rizwan
www.linkedin.com/in/alishba-rizwan--/
alishbarizwanakbarmirza@gmail.com