RFM Superstore Customer Segmentation

Presented By: Patrick Nguyen

Course: Analytical Tool & Applications

Date: November 13th2025

How to better retain your customer by identify the most matter customer group

AGENDA

Unique and professional presentation design

Why customer retention matters

- **Affordability**

  It's 6 to 7 times more expensive to acquire a new customer than it is to retain an existing customer.

- **ROI**

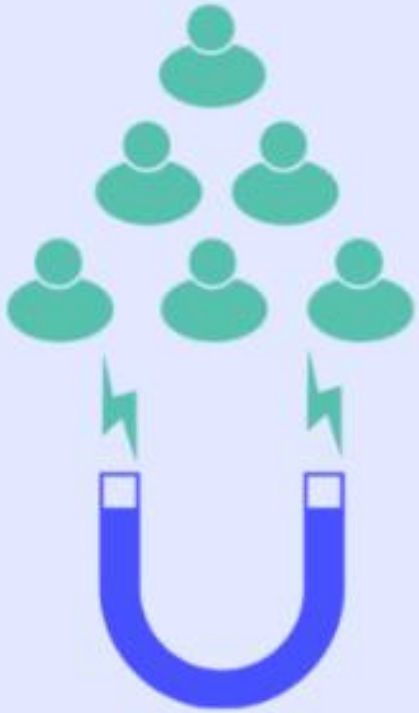
  A 5% increase in customer retention can increase company revenue by 25-95%.

- **Loyalty**

  Retained customers buy more often and spend more than newer customers.

- **Referrals**

  Satisfied, loyal customers are more likely to sing a company's praises and refer their friends and family



Information Source: HubSpot Blog

MKT FUNCTION ARISES

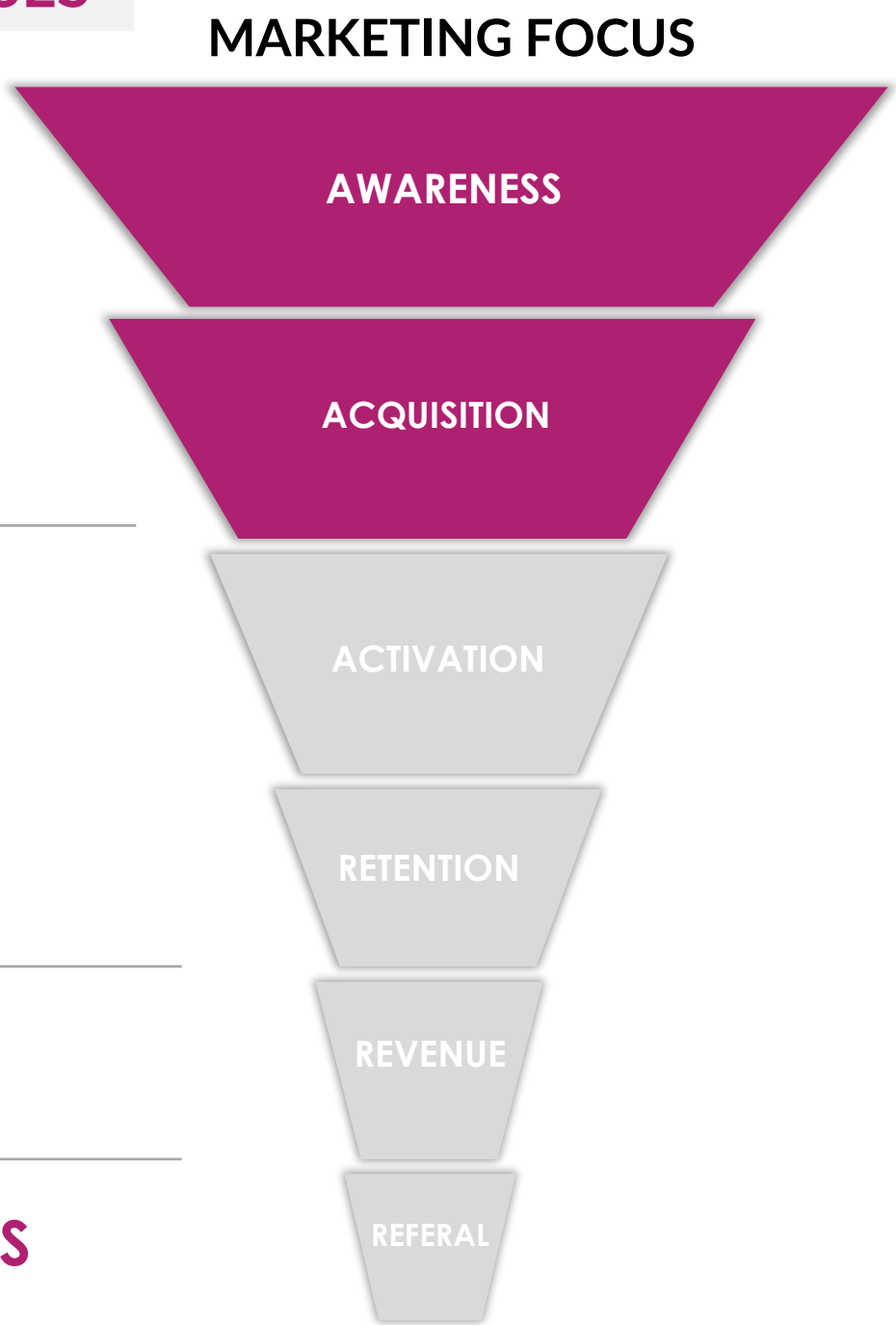
DATA

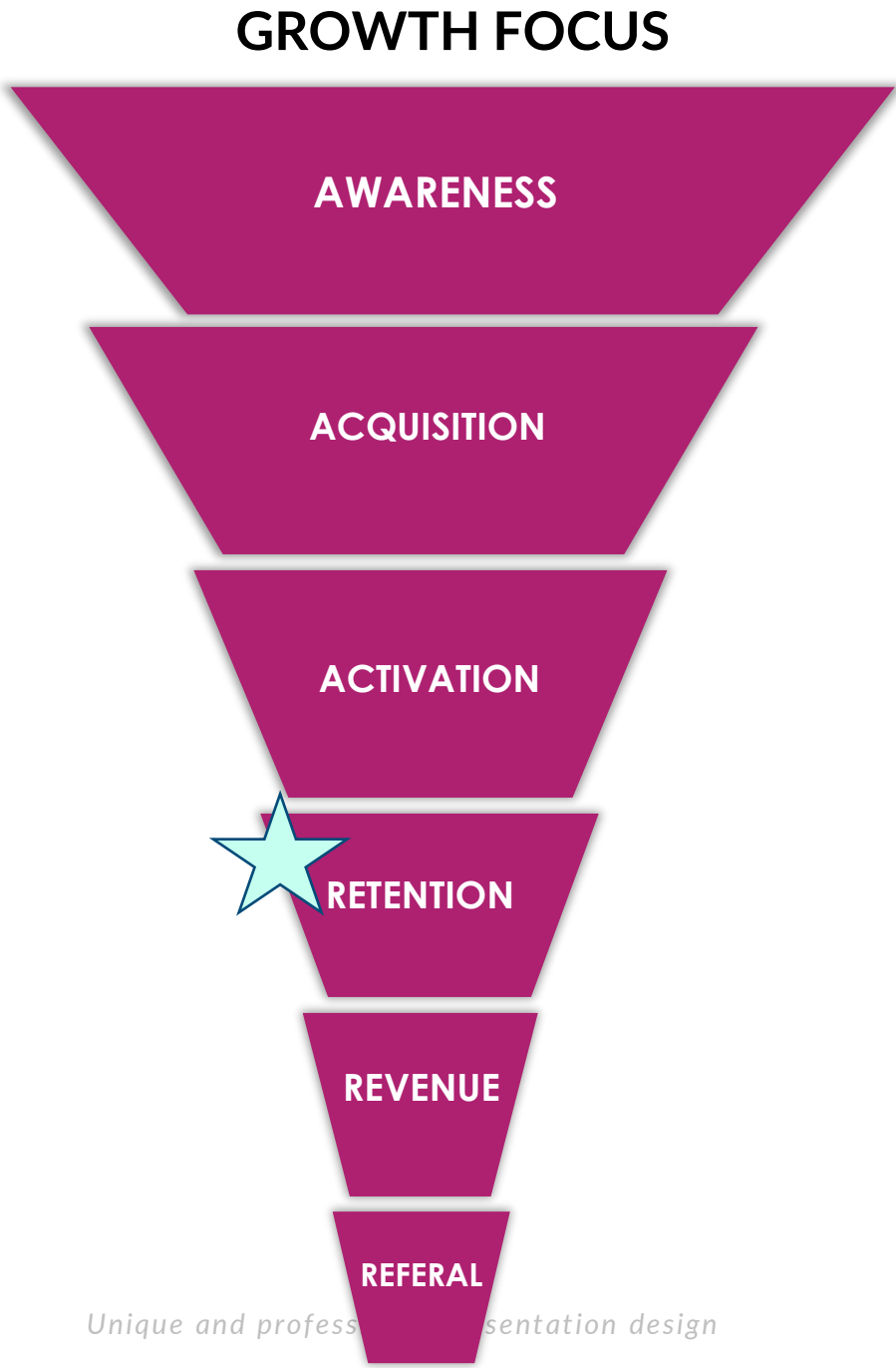
MARKETING

PRODUCT

SALES EFFORT

CLIENT SUCCESS





Unique and profess sentation design

01

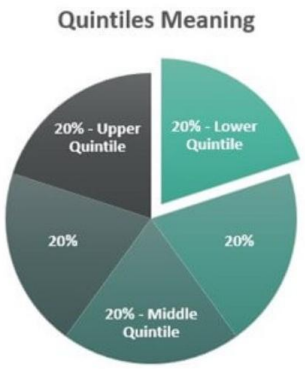
Business Understanding

- SuperStore is a global retail company. The Marketing Department wants to run marketing campaigns during the **Christmas and New Year holidays** to thank customers for their past support of the company. In addition, potential customers can be upgraded to become loyal customers.

- The Marketing Director also proposed a plan to use the RFM model **to segment customers and then launch appropriate marketing campaigns.** Analyze the current situation of the company and give suggestions to the Marketing team

02. Data Understanding

Segment	RFM Score
Champions	555, 554, 544, 545, 454, 455, 445
Loyal	543, 444, 435, 355, 354, 345, 344, 335
Potential Loyalist	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
New Customers	512, 511, 422, 421, 412, 411, 311
Promising	525, 524, 523, 522, 521, 515, 514, 513, 425,424, 413,414,415, 315, 314, 313
Need Attention	535, 534, 443, 434, 343, 334, 325, 324
About To Sleep	331, 321, 312, 221, 213, 231, 241, 251
At Risk	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
Cannot Lose Them	155, 154, 144, 214,215,115, 114, 113
Hibernating customers	332, 322, 233, 232, 223, 222, 132, 123, 122, 212, 211
Lost customers	111, 112, 121, 131,141,151



r ID	Customer Na	Segment	Country	City	State	Postal Code
0	Claire Gute	Consumer	United State:	Henderson	Kentucky	42420
0	Claire Gute	Consumer	United State:	Henderson	Kentucky	42420
5	Darrin Van H	Corporate	United State:	Los Angeles	California	90036
5	Sean O'Donn	Consumer	United State:	Fort Lauderd	Florida	33311
5	Sean O'Donn	Consumer	United State:	Fort Lauderd	Florida	33311
0	Brosina Hoffr	Consumer	United State:	Los Angeles	California	90032
0	Brosina Hoffr	Consumer	United State:	Los Angeles	California	90032
0	Brosina Hoffr	Consumer	United State:	Los Angeles	California	90032
0	Brosina Hoffr	Consumer	United State:	Los Angeles	California	90032
0	Brosina Hoffr	Consumer	United State:	Los Angeles	California	90032
0	Brosina Hoffr	Consumer	United State:	Los Angeles	California	90032
0	Andrew Aller	Consumer	United State:	Concord	North Carolir	28027
0	Irene Maddo	Consumer	United State:	Seattle	Washington	98103
5	Harold Pawla	Home Office	United State:	Fort Worth	Texas	76106
5	Harold Pawla	Home Office	United State:	Fort Worth	Texas	76106
5	Pete Kriz	Consumer	United State:	Madison	Wisconsin	53711
0	Alejandro Gr	Consumer	United State:	West Jordan	Utah	84084
5	Zuschuss Dor	Consumer	United State:	San Francisc	California	94109
5	Zuschuss Dor	Consumer	United State:	San Francisc	California	94109
5	Zuschuss Dor	Consumer	United State:	San Francisc	California	94109

- Dataset includes 4 different related tables including: **transaction information, products information, returned orders of customers purchasing products & segmentation table** from 2014 to 2017 and RFM classification

02

Data Understanding

Comment

- There are no duplicate records in orders_pre.
- The Order ID' column is not unique => We have to count distinct 'Order ID' instead of count
- We have to filter orders that were not returned.

Transaction information dataframe

```
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Row ID       9994 non-null   int64
 1   Order ID     9994 non-null   object
 2   Order Date   9994 non-null   datetime64[ns]
 3   Ship Date    9994 non-null   datetime64[ns]
 4   Ship Mode    9994 non-null   object
 5   Customer ID  9994 non-null   object
 6   Channel      9994 non-null   object
 7   Postal Code  9994 non-null   int64
 8   Product ID   9994 non-null   object
 9   Sales        9994 non-null   float64
 10  Quantity     9994 non-null   int64
 11  Unit Cost    9994 non-null   float64
dtypes: datetime64[ns](2), float64(2), int64(3), object(5)
memory usage: 937.1+ KB
```

]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Channel	Postal Code	Product ID	Sales	Quantity	Unit Cost
0	646	CA-2017-126221	2017-12-30	2018-01-05	Standard Class	CC-12430	Home Office	47201	OFF-AP-10002457	209.300	2	76.3945
1	907	CA-2017-143259	2017-12-30	2018-01-03	Standard Class	PO-18865	Consumer	10009	FUR-BO-10003441	323.136	4	77.7546
2	908	CA-2017-143259	2017-12-30	2018-01-03	Standard Class	PO-18865	Consumer	10009	TEC-PH-10004774	90.930	7	12.6003

Returned orders dataframe: We have to filter orders that were not returned before RFM anaalysis

:
```
returned.info()
returned.head(5)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 296 entries, 0 to 295
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Returned  296 non-null    object
 1   Order ID  296 non-null    object
dtypes: object(2)
memory usage: 4.8+ KB
```

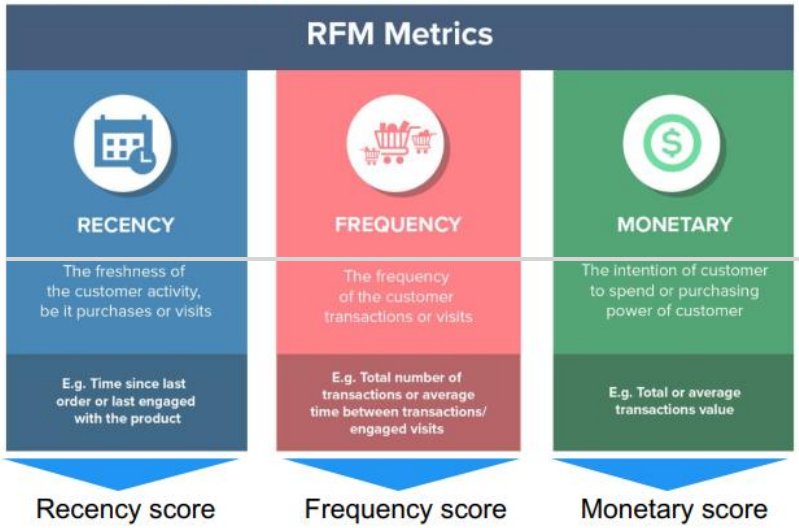
:

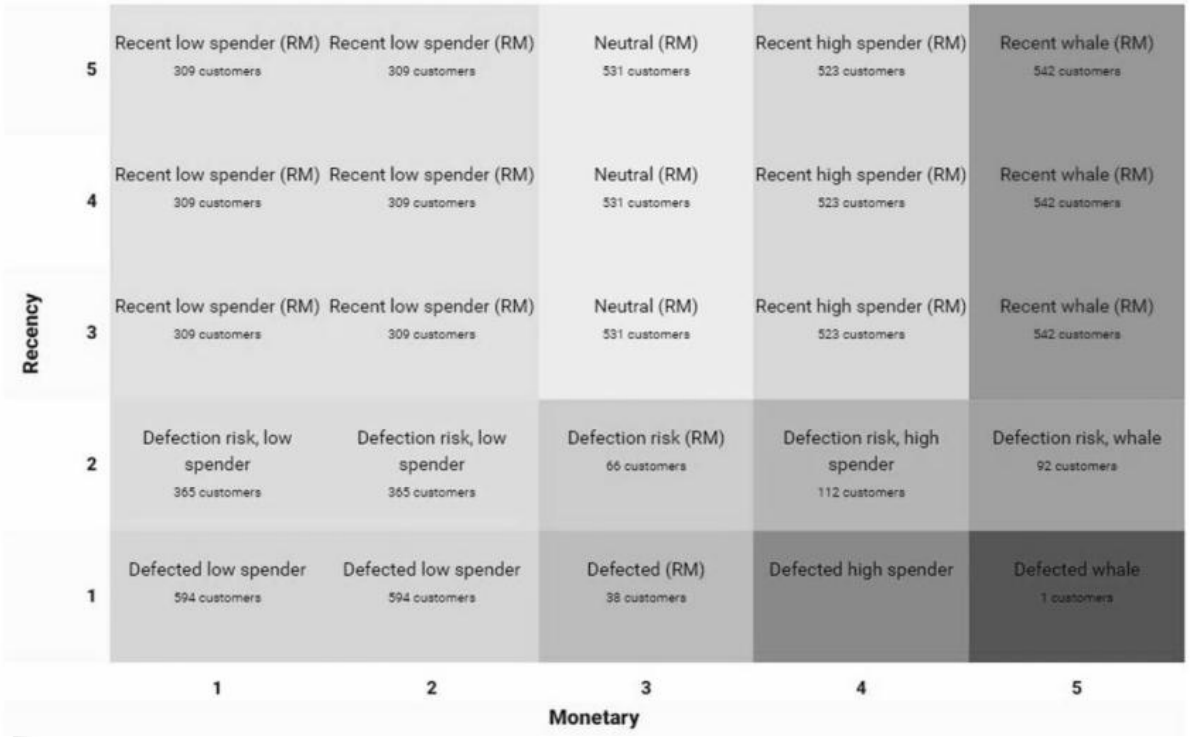
	Returned	Order ID
0	Yes	CA-2017-153822
1	Yes	CA-2017-129707
2	Yes	CA-2014-152345
3	Yes	CA-2015-156440
4	Yes	US-2017-155999

03

RFM MODEL

Use RFM model to identify potential users for sales incremental





- RFM is a method used for analyzing customer value. It is commonly used in database marketing and direct marketing and has received particular attention in retail and professional services industries
- It ranks a customer in each of these three categories, generally on a scale of 1 to 5 (the higher the number, the better the result). The “best” customer would receive a top score in every category

03

RFM SEGMENTATION

```
# Caculate Recency, Frequency, Monetory
rfm = orders.groupby('Customer ID').agg({'Order Date':'max',
                                         'Order ID':'nunique',
                                         'Sales':'sum'}).reset_index()

rfm['Order Date'] = (pd.to_datetime('2017-12-31') - rfm['Order Date']).dt.days
rfm.columns = ['customer_id','recency','frequency','monetory']
rfm
```

	customer_id	recency	frequency	monetory
0	AA-10315	185	5	5563.560
1	AA-10375	20	9	1056.390
2	AA-10480	260	4	1790.512
3	AA-10645	483	5	5073.975

```
# Get R, F, M score using qcut

lab_des=[5,4,3,2,1]
lab_asc=[1,2,3,4,5]

rfm['r'] = pd.qcut(rfm['recency'], q=5, labels=lab_des)
rfm['f'] = pd.qcut(rfm['frequency'], q=5, labels=lab_asc)
rfm['m'] = pd.qcut(rfm['monetory'], q=5, labels=lab_asc)

# concatenate R, F, M score into RFM score
rfm['rfm_score'] = rfm['r'].astype(str) + rfm['f'].astype(str) + rfm['m'].astype(str)

rfm.head(5)
```

	customer_id	recency	frequency	monetory	r	f	m	rfm_score
0	AA-10315	185	5	5563.560	2	2	5	225
1	AA-10375	20	9	1056.390	5	5	2	552
2	AA-10480	260	4	1790.512	2	1	3	213
3	AA-10645	483	5	5073.975	1	2	5	125

```
# Assign segmentation for each customer_id

cus_segmentation = rfm.merge(segmentation, on ='rfm_score', how = 'left')
cus_segmentation.head(5)
```

	customer_id	recency	frequency	monetory	r	f	m	rfm_score	segment
0	AA-10315	185	5	5563.560	2	2	5	225	At Risk
1	AA-10375	20	9	1056.390	5	5	2	552	Potential Loyalist
2	AA-10480	260	4	1790.512	2	1	3	213	About To Sleep
3	AA-10645	483	5	5073.975	1	2	5	125	At Risk
4	AB-10015	416	3	886.156	1	1	2	112	Lost customers

Define & Calculate metrics for : R (recency), F (frequency), M (Monetary)

Getting RFM score using pd.qcut() >> divided into 5 quantiles (quintiles). Each contain ~ 20% of the customer

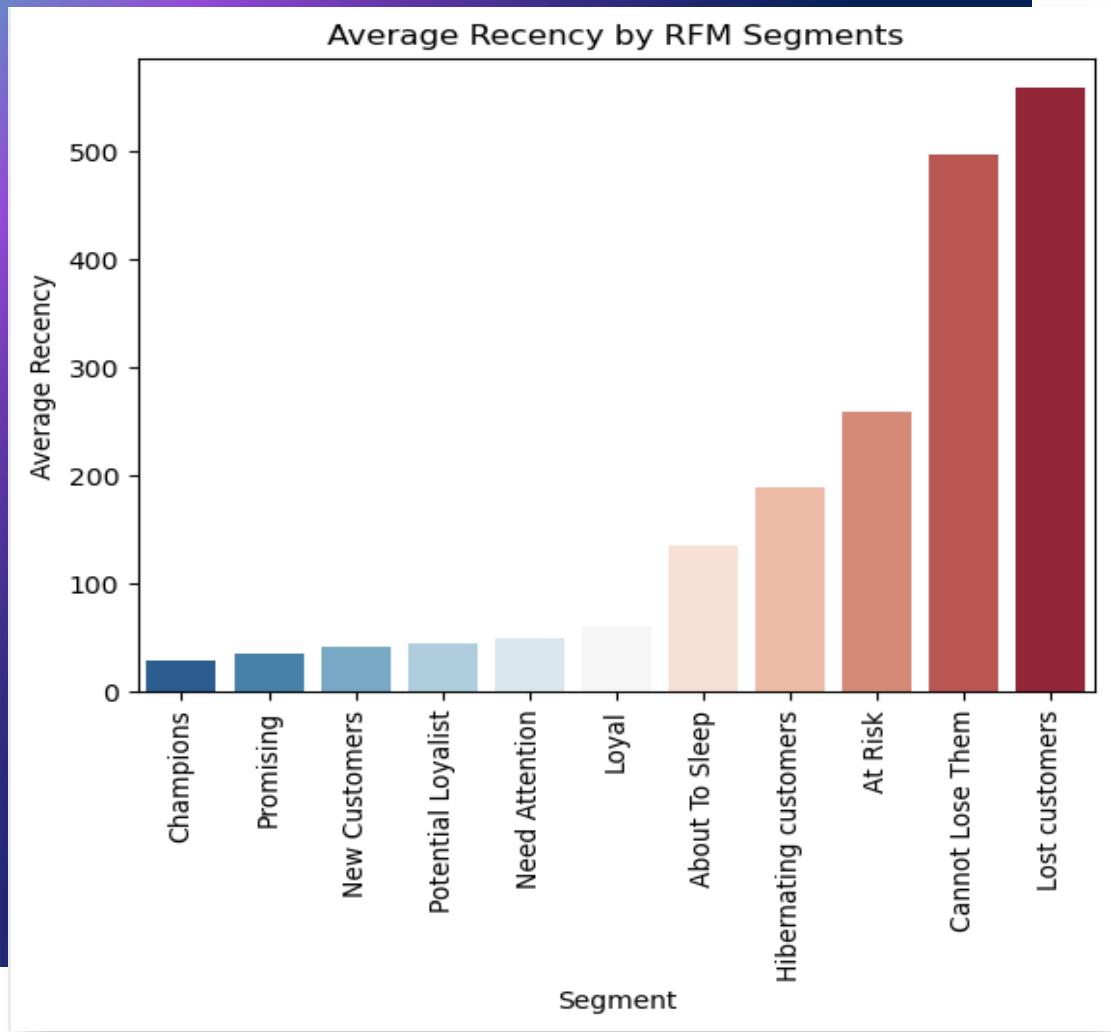
Mapping RFM score with predefine segment to complete our RFM model

03

Key Visualizations

Comment:

- Segments with the most recent purchase date:Champions (28.5 days), Promises (35.3 days) and New Customers (41.5 days)
- Segments that have not returned to buy for a long time: Lost customers (558 days) and Cannot Lose Them (496.5 days)
- >> Mean of Recency ~ 166 days, while the median is 83 days, this represents a lot of high Recency values. The larger this index, the higher the customer's tendency to leave

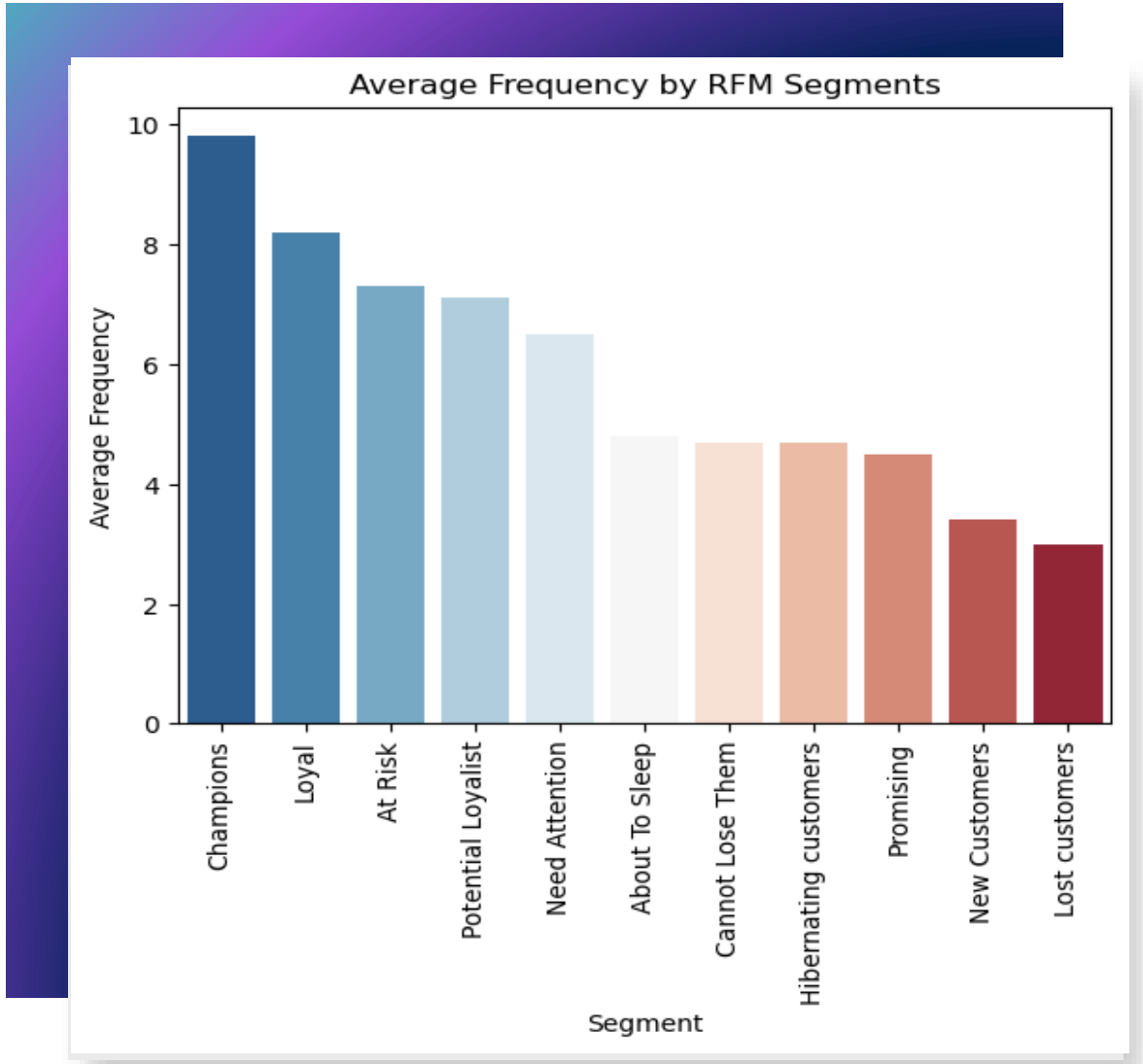


03

Key Visualizations

Comment:

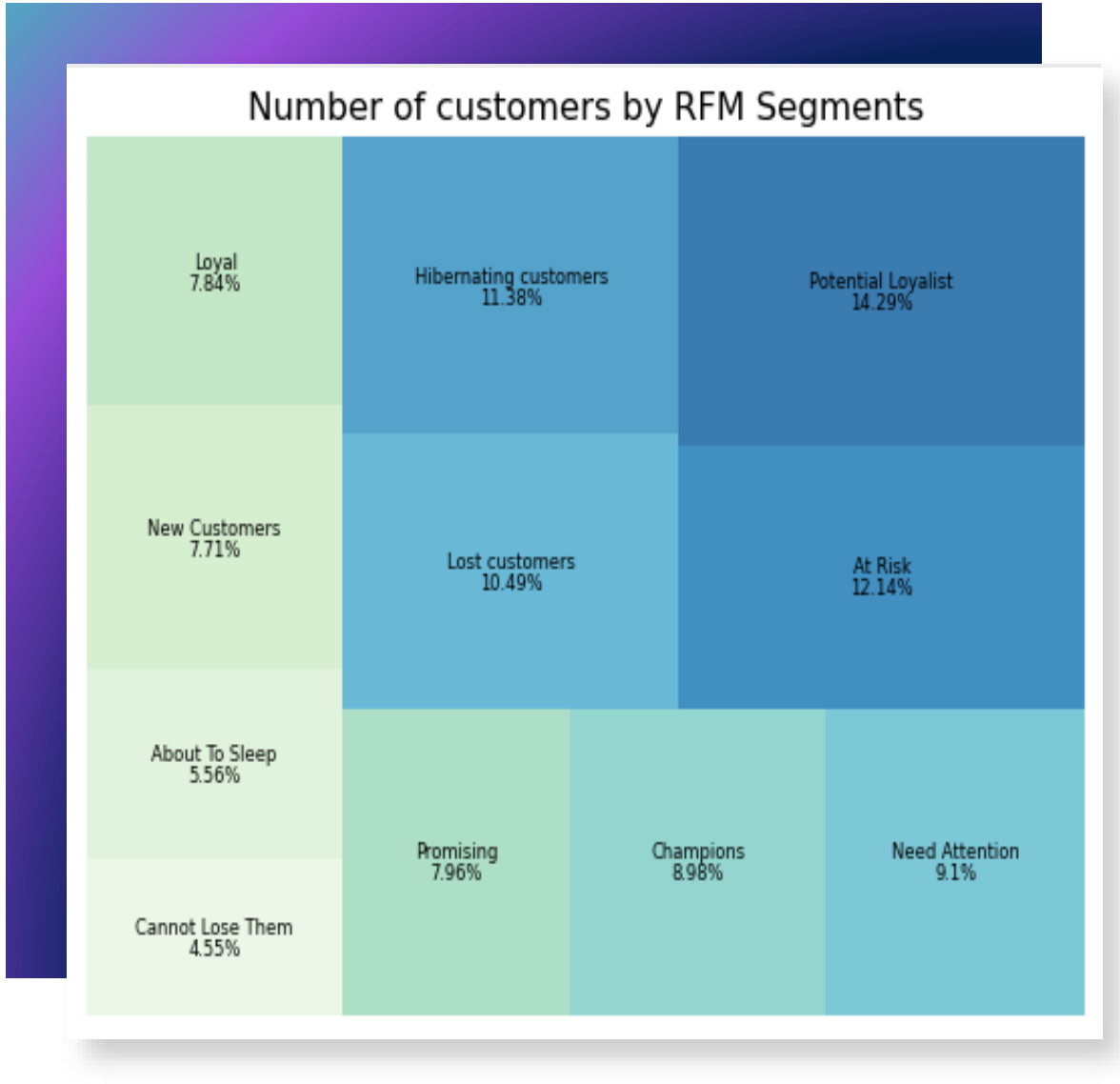
- Most frequent purchase segments: Champions (9.8x), Loyal (8.2x), At Risk (7.3x)
- Segments with the lowest number of orders: Lost customers (3.0 times), New Customers (3.4 times)
- The number of orders of the Champions segment is 3.3 times higher than that of the Lost customers segment
- >> Mean of Frequency is 6 times, which is low for a retail company. This means customer loyalty is not high



03

Key Visualizations

- The Treemap showing 11 segments:
- 2 segments with the highest proportion of customers are Potential Loyalist (14.29%) and At Risk (12.14%)
- The two most positive segments account for less than 17% of the proportion of customers (Champions, 8.98%, and Loyal, 7.84%)



03

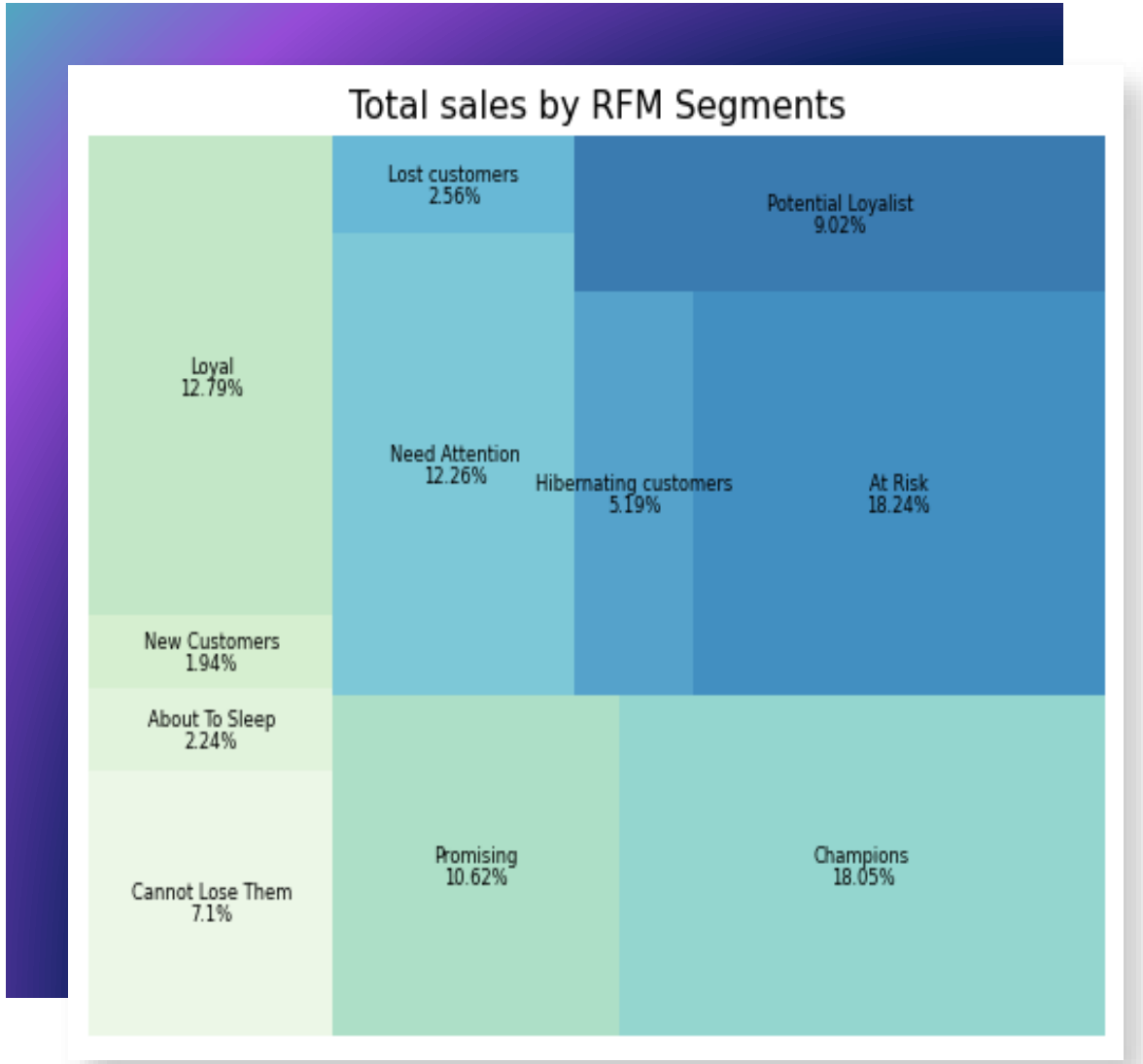
Key Visualizations

Champions + Loyal ≈ 40% of revenue, despite small population share

Negative segments such as Hibernating and Lost accounted for a high proportion of customers, 11.38% and 10.49%, respectively. However, these two groups account for less than 8% of revenue

>>Potential Loyalist (the ideal segment) has the highest proportion of customers (14.29%), but its revenue proportion is only 9.02%. Meanwhile, the negative segment, At Risk, accounts for 1[illegible]24% of revenue

=> In this Christmas - New Year marketing campaign, SuperStore needs to prioritize their efforts to promote the Potential Loyalist group to become Loyal and Champions, and find ways to reconnect with customers in the At Risk group



04

Insights & Recommendations

01

RFM Health vs Retail Benchmarks

Frequency = 6 orders (retail average: 10–14) → below industry

Recency = 166 days (healthy brands: <60–90 days) → high churn risk

Monetary stable, but engagement signals declining

Strategic Focus:

- Lift Recency & Frequency to align with industry norms
- Prioritize segments with highest potential: Potential Loyalists & At Risk

02

For Potential Loyalists:

Cart-value booster offers → target > $238.68 avg

Introduce +10–20% value bundles (benchmark: bundle increases AOV by 12–18%)

Strengthen loyalty program (top retailers see +25% retention lift)

03

For At Risk:

Investigate high-return categories → fix root cause (industry standard return rate: 8–10%, likely higher here)

Survey + reward (points/voucher) → expected win-back rate: 12–18%

Free shipping upgrade (industry: reduces churn by 10–15%)

Personalized recommendation emails → focus on Binders, Paper, Phones, Storage

Unique and professional presentation design

THANK YOU

References

- Hughes, A. M. (2020). *Strategic Database Marketing*. McGraw-Hill.
- IBM SPSS Modeler (2023). *RFM Analysis and Customer Segmentation Techniques*.
- Kaggle (2024). *SuperStore Dataset – E-commerce Retail Transactions*.
- McKinsey & Company (2022). *The Value of Personalization in Retail* (conceptual support for retention/upsell impact).
- Python Library Docs: *pandas*, *NumPy*, *Matplotlib*, *Seaborn*.
- Mercy University (2025). *ANLC 763 Course Materials*.