Superstore RFM Analysis

Important Links

Dataset	Superstore
Google Collab	Superstore - RFM Analysis.ipynb
Tableau	Superstore RFM Analysis: Insights into Customer Behavior
Github	https://github.com/hijirdella/Superstore-RFM-Analysis
LinkedIn	https://www.linkedin.com/in/hijirdella/
Email	hijirdw@gmail.com

Hijir Della Wirasti

Business Intelligence

Table of contents

Objective

The primary objectives of this analysis are:

1. Perform **Exploratory Data Analysis (EDA)** to identify key patterns and visualize customer behavior.
2. Apply **RFM Analysis** (Recency, Frequency, and Monetary) to segment customers and identify valuable customer groups.
3. Derive actionable insights and recommendations to solve the main business problem.
4. Develop an optional interactive dashboard using **Tableau Public** to showcase data insights effectively.
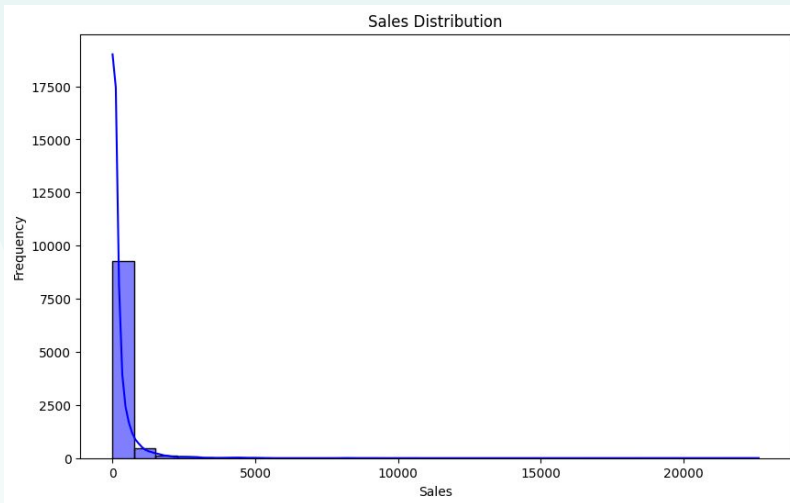
01

EDA
(Exploratory Data Analysis)

python™

Sales Distribution

Profit Distribution

The **Profit Distribution** graph indicates that the majority of profits are clustered around **zero**, with very few outliers on both the **negative (losses)** and **positive (profits)** ends.
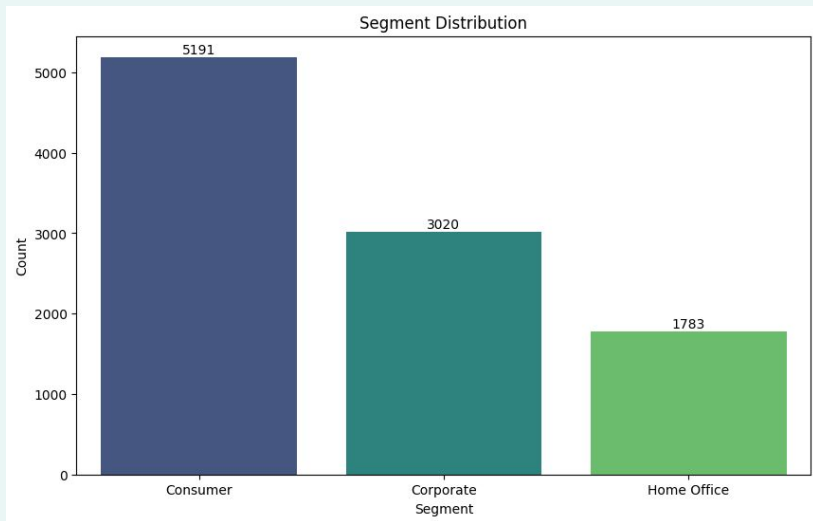**Key Highlights**:
- Most transactions generate **low profits** or **break even**.
- A small number of transactions lead to **significant profits** or **losses**.
- Further analysis is required to identify factors contributing to these outliers, such as **discounts**, **product categories**, or **regions**.
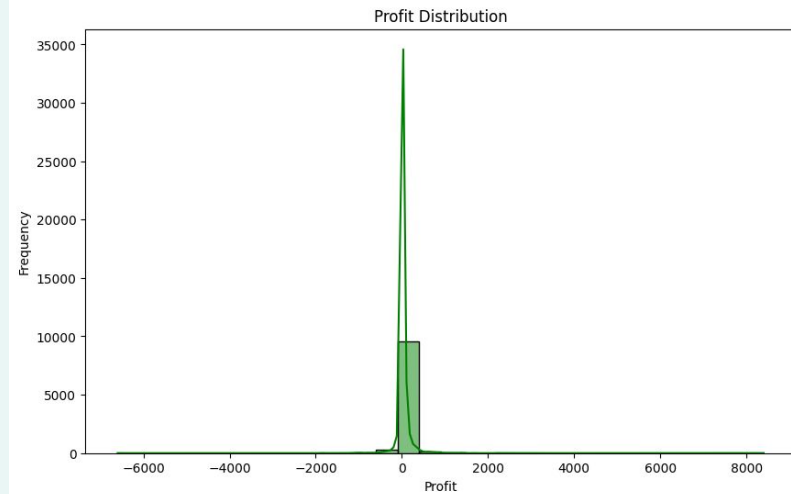
The **Sales Distribution** graph shows a **highly skewed distribution**, where the majority of sales are concentrated at **lower values**.
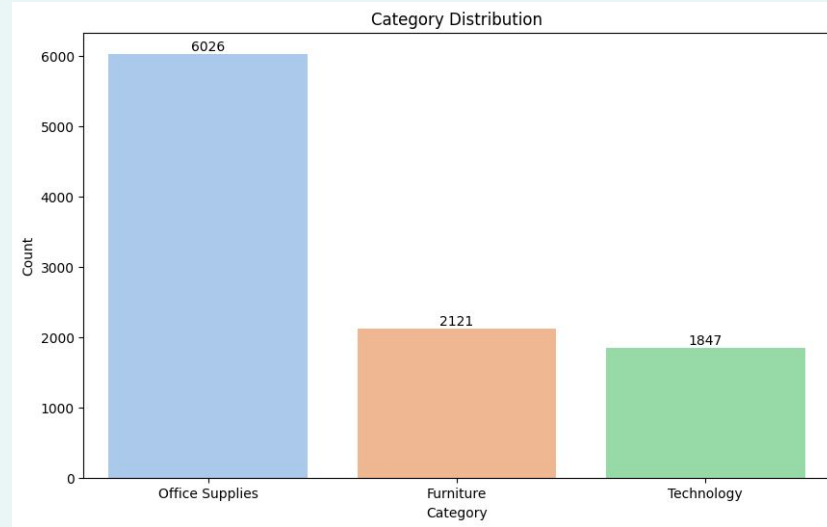**Key Highlights**:
- Most transactions involve **low sales amounts**.
- A **long tail** exists with very few **high-value transactions**.
- These **high-value sales** can be analyzed to identify drivers such as **specific products**, **customer segments**, or **geographical regions** for optimization.

Segment Distribution

Category Distribution

The **Segment Distribution** graph reveals that the majority of customers fall into the **Consumer** segment, followed by **Corporate** and then **Home Office**.

- **Key Highlights**:
  - The **Consumer segment** dominates with **5,191 customers**, accounting for the largest portion of the customer base.
  - The **Corporate segment** includes **3,020 customers**, significantly fewer than the Consumer segment.
  - The **Home Office segment** has the smallest customer base with **1,783 customers**.

The **Category Distribution** graph highlights the distribution of transactions across different product categories.

- **Key Highlights**:
  - **Office Supplies** dominates the transactions with **6,026 entries**, making it the most frequently purchased category.
  - **Furniture** follows with **2,121 entries**, indicating moderate interest in this category.
  - **Technology** has the fewest transactions at **1,847 entries**, suggesting a potential opportunity to boost sales in this category.

Sales vs Profit by Segment
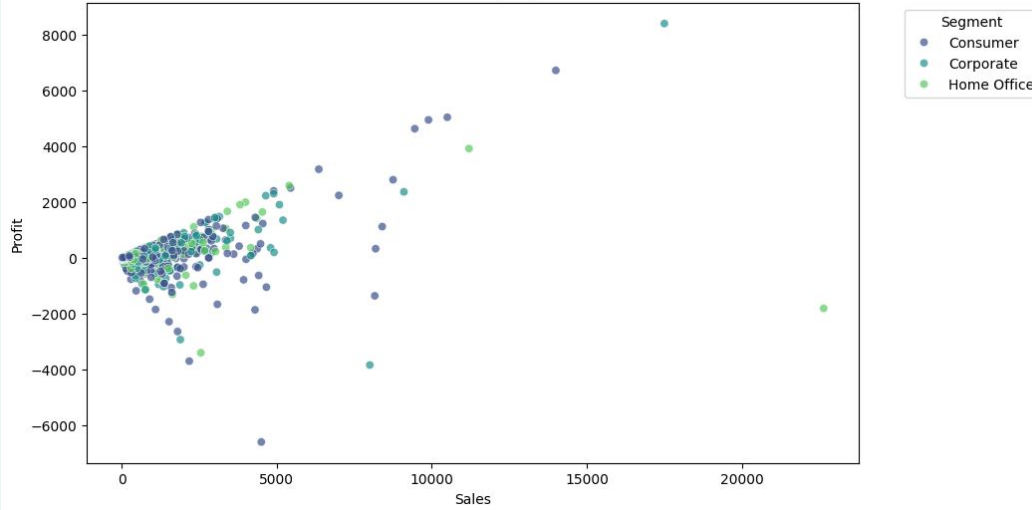
Key Highlights:

- There is a **positive correlation** between sales and profit, but it varies significantly by segment.
- The **Consumer segment** dominates the cluster with moderate sales and positive profit, indicating customers with **balanced Frequency and Monetary scores**.
- The **Corporate segment** exhibits some high-sales, high-profit outliers, aligning with **Big Spenders** and **Potential Customers**.
- The **Home Office segment** has several low-profit and negative-profit transactions, suggesting customers who might belong to the **Lost Cheap** or **Lost Customers** segments.

Actionable Recommendations:

- Focus on the **Corporate segment's outliers** for upselling and premium product promotions.
- Address the **negative profit transactions in Home Office** with better pricing strategies or improved operational efficiencies.
- For the **Consumer segment**, leverage loyalty programs to maximize repeat purchases from high Frequency score customers.

Sales by Category and Sub-Category
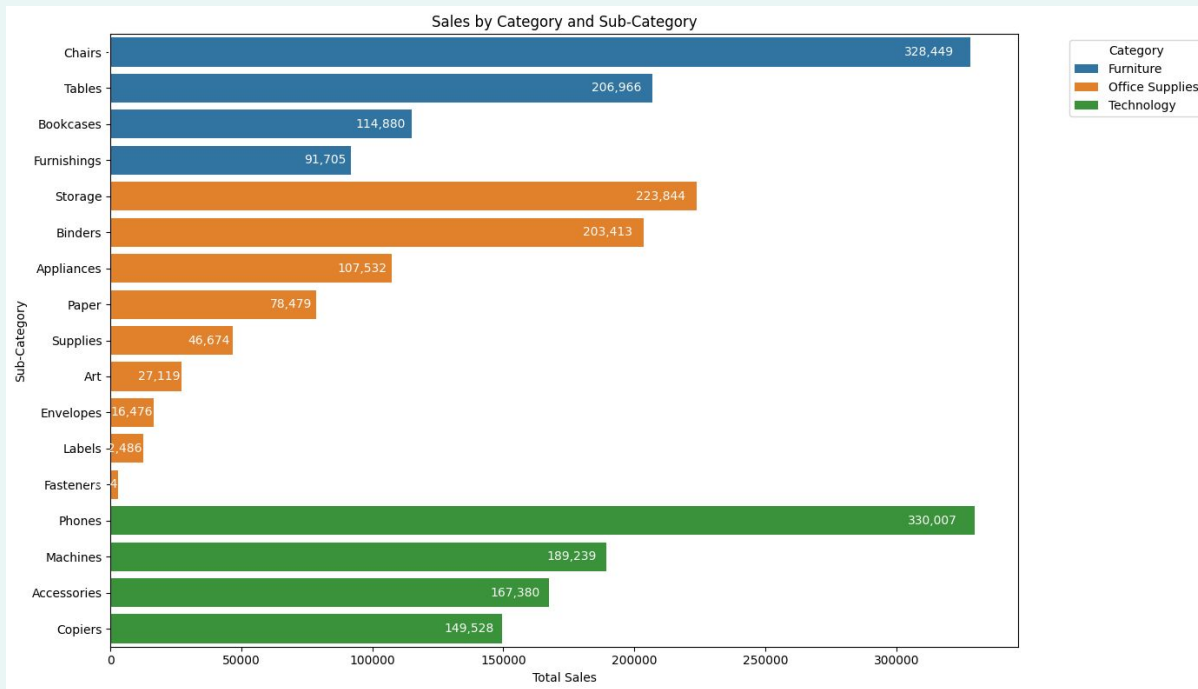
Key Observations:
- **Phones (Technology)** and **Chairs (Furniture)** are the top sub-categories by sales, with Phones generating the highest revenue at **$330,007**. These sub-categories likely attract **Big Spenders** and **Loyal Customers**.
- **Storage (Office Supplies)** also shows significant sales, suggesting it appeals to high **Frequency** customers.
- Low-performing sub-categories like **Fasteners**, **Labels**, and **Envelopes** indicate segments with lower **Monetary** value or low purchase engagement.

RFM Connection:
- High-sales sub-categories align with customers scoring **high in Monetary (AOV Quartile = 1)**.
- Sub-categories like **Art** and **Supplies** may cater to **Lost Cheap** or **Almost Lost** customers, requiring targeted promotions or incentives to boost engagement.

Recommendations:
- Focus promotions on high-value sub-categories like Phones, Chairs, and Storage to retain **Big Spenders** and maximize profitability.
- Experiment with discounts or bundles in low-performing sub-categories to revive engagement among **Lost Customers** or **Potential Customers**.

Sales by Region

Sales by Region

Region	Total Sales
West	725,458
East	678,781
Central	501,240
South	391,722

Key Observations:

- **West Region** leads in sales with **$725,458**, followed by **East ($678,781)** and **Central ($501,240)**. The **South Region** has the lowest sales at **$391,722**.
- Higher sales in the **West and East** may correlate with regions hosting **Loyal Customers** or **Big Spenders**.
- The **South Region** may have a higher concentration of **Lost Customers** or **Almost Lost** segments.

RFM Connection:

- Regions with higher sales likely host customers with **high Monetary and Frequency scores**.
- The **South Region's** lower sales suggest more customers with **low Frequency and Monetary scores**, needing tailored re-engagement strategies.

Recommendations:

- Invest in retention strategies for the **West and East Regions** to sustain **Loyal Customers**.
- Develop promotional campaigns in the **South Region** to revive interest among **Lost Customers** or attract **Potential Customers**.
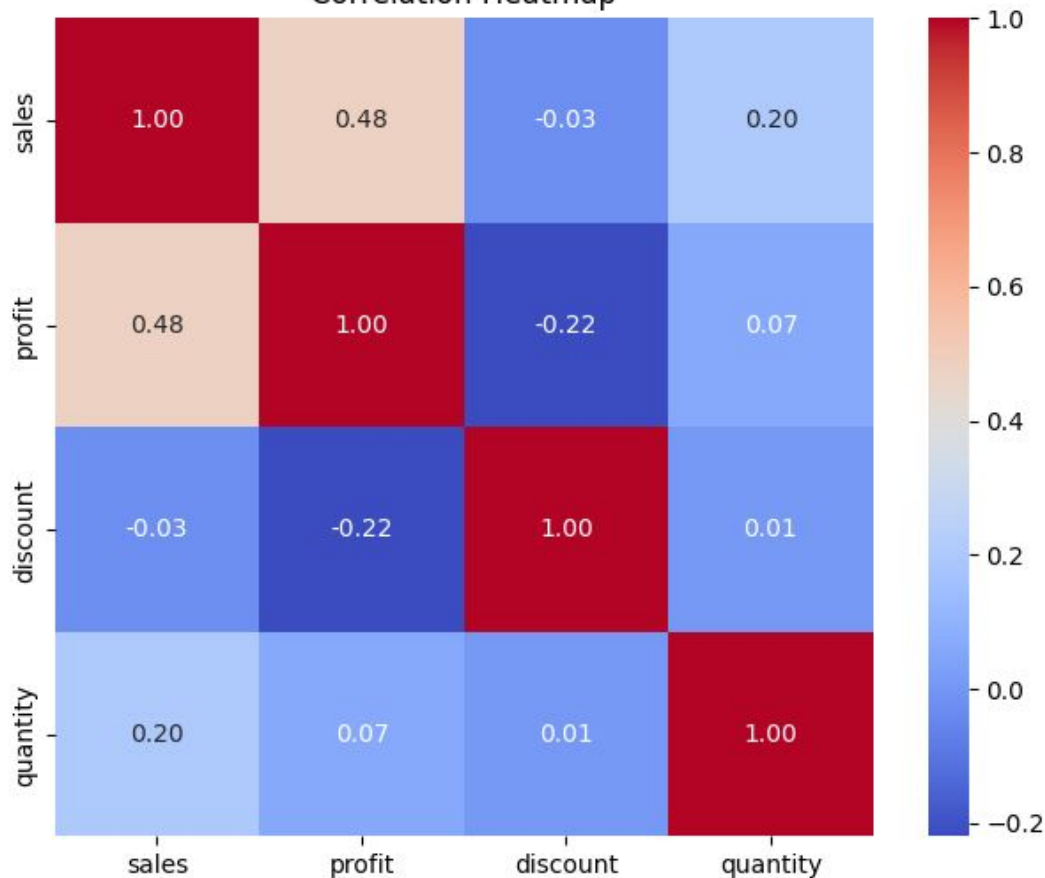
Correlation Heatmap

Key Observations:

- **Sales and Profit** have a moderate positive correlation (**0.48**), indicating higher sales tend to result in higher profits.
- **Discount and Profit** show a **negative correlation (-0.22)**, suggesting that heavy discounting might reduce profitability.
- **Quantity has a weak correlation with Sales (0.20)** but almost no impact on Profit, indicating that selling larger quantities doesn't necessarily translate to higher profits.

RFM Connection:

- **High Sales and Profit correlation** suggests **Big Spenders and Loyal Customers** likely contribute significantly to profitability.
- **Negative Discount-Profit relationship** indicates segments with higher discounts, such as **Lost Cheap or Almost Lost**, may not be contributing positively to overall profit.
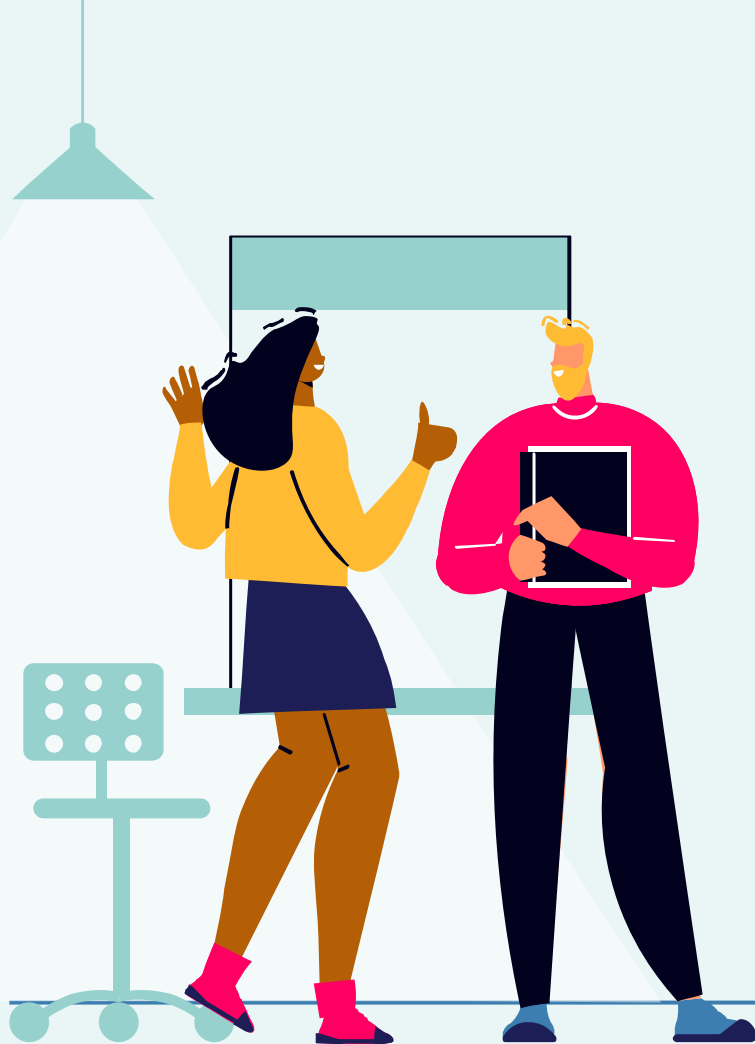
Recommendations:

- Focus on **upselling or cross-selling** to **Big Spenders** with minimal discounts to maintain profitability.
- Develop tailored offers for **Lost Customers** that balance **engagement and profitability**, avoiding excessive discounts.

02

RFM Analysis

Insights:

1. **Skewed Distribution**:
   - A higher concentration of customers is found in scores like **444** (highest recency, lowest frequency and monetary) and **442**, suggesting many customers fall into the **Lost Customers** or **Lost Cheap** categories.
   - The **low-frequency and low-monetary combinations dominate**, indicating a significant portion of the customer base requires re-engagement or minimal return on investment.

2. **Potential Focus**:
   - Scores like **111** or **121** are underrepresented, which correspond to **Best Customers** or high-value customer segments. This emphasizes the need to expand and retain such valuable customers.

3. **Opportunity Segments**:
   - Mid-range scores (e.g., **221**, **231**) indicate **Potential Customers** who can be nurtured to move up to **Loyal Customers** or **Big Spenders**.

RFM Score Distribution
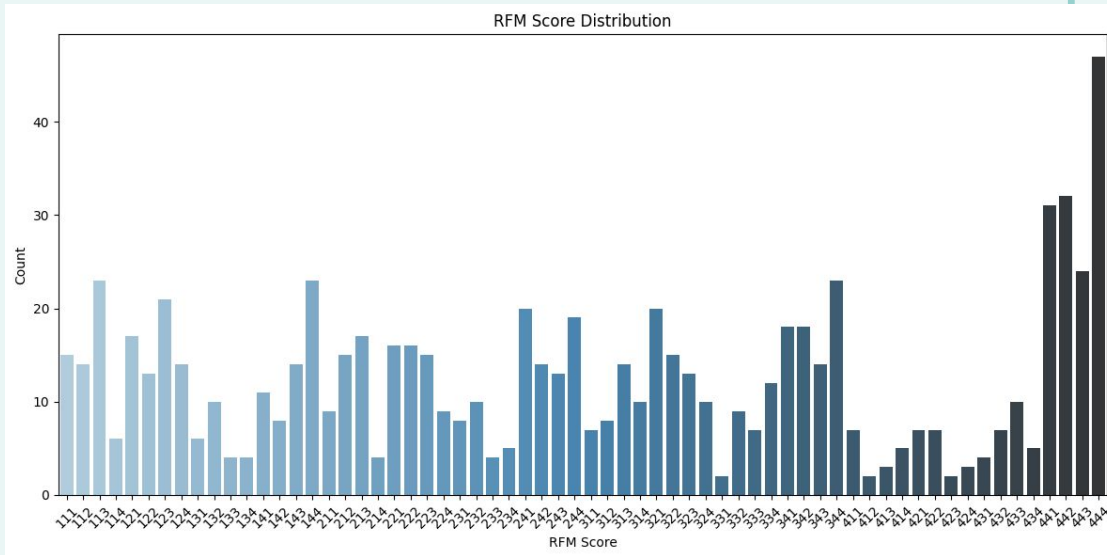
Recommendations:

- **Re-engage Lost Customers** with targeted campaigns focusing on promotions or loyalty offers.
- **Expand Best Customers** by identifying common characteristics and acquiring new customers with similar profiles.
- **Convert Potential Customers** through personalized upselling or cross-selling strategies.

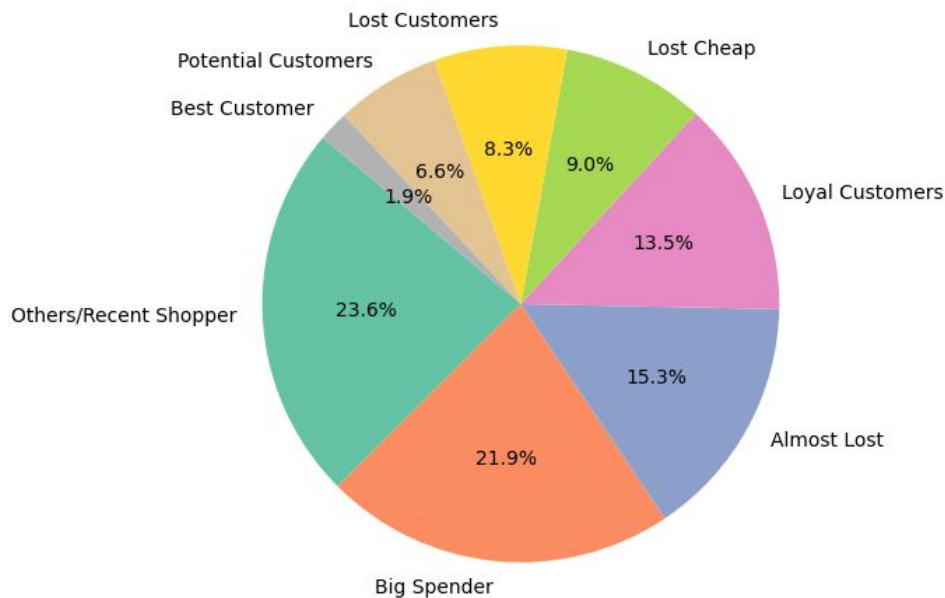Customer Segmentation Distribution

Segment Descriptions and Strategic Insights

1. Others/Recent Shopper (23.6%)

- **Description**: These customers have made recent purchases but show moderate frequency and spending levels.
- **Strategy**: Focus on engagement campaigns like newsletters, personalized emails, or limited-time offers to encourage repeat purchases and build loyalty.

2. Big Spender (21.9%)

- **Description**: Customers with high average order value (AOV), indicating significant spending but not necessarily frequent purchases.
- **Strategy**: Provide exclusive perks, early access to sales, and loyalty programs to maintain their high spending and deepen the relationship.

3. Almost Lost (15.3%)

- **Description**: Customers with declining recency and engagement but moderate frequency or spending in the past.
- **Strategy**: Launch win-back campaigns with strong incentives such as discounts or free shipping to revive their engagement.

4. Loyal Customers (13.5%)

- **Description**: These customers have the highest frequency of purchases, demonstrating consistent loyalty.
- **Strategy**: Reward their loyalty with special discounts, priority support, or exclusive memberships to ensure retention and word-of-mouth advocacy.

Customer Segmentation Distribution
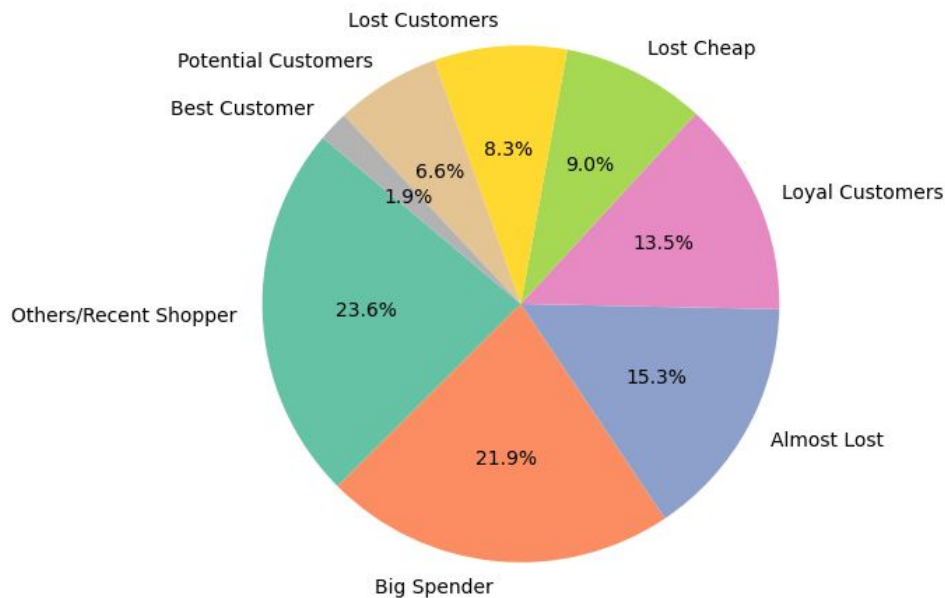
Segment Descriptions and Strategic Insights

5. Lost Cheap (9.0%)

- **Description**: Customers who were infrequent buyers with low AOV and have not returned in a long time.
- **Strategy**: Minimize effort on this segment or re-engage selectively with low-cost, high-impact campaigns.

6. Lost Customers (8.3%)

- **Description**: Customers who used to purchase more frequently but have not engaged in a significant period.
- **Strategy**: Use reactivation strategies like personalized emails or targeted ads with deep discounts to rekindle their interest.

7. Potential Customers (6.6%)

- **Description**: Customers with moderate to high recency, frequency, and AOV, but not yet classified as loyal or high spenders.
- **Strategy**: Nurture this group through personalized communications and cross-selling to move them to the **Loyal Customer** or **Big Spender** categories.

8. Best Customers (1.9%)

- **Description**: The most valuable segment with high recency, frequency, and AOV, indicating strong loyalty and high spending.
- **Strategy**: Retain this segment by offering them VIP treatment, early access to new products, and exclusive offers to ensure long-term engagement.

Number of Customers per Segment
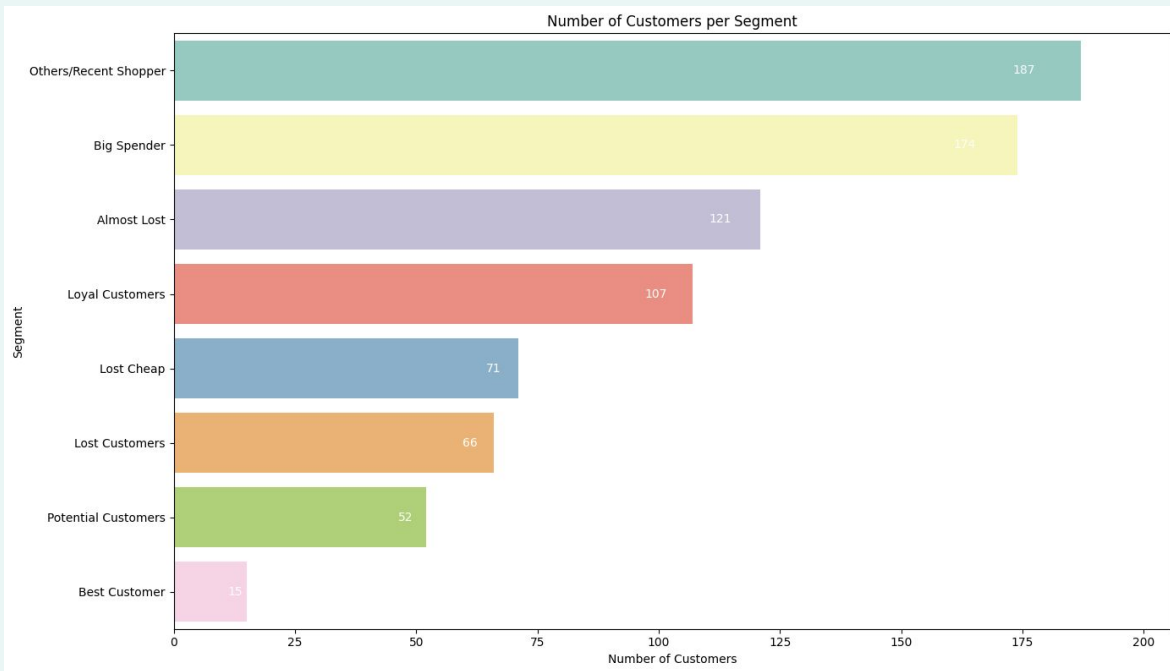
Recommendations:

- **Focus efforts on "Big Spender" and "Loyal Customers"** for revenue maximization.
- Use **win-back campaigns** for "Almost Lost."
- Gradually convert "Others/Recent Shoppers" into **Loyal Customers.**

Recency vs Frequency by Segment

Insights:

1. **Recency Distribution**:
   - Customers in **Lost Customers** and **Lost Cheap** segments generally have higher recency values, indicating they haven't interacted recently with the business.
   - **Big Spenders** and **Loyal Customers** cluster towards lower recency values, showing consistent and recent engagement.
2. **Frequency Patterns**:
   - **Loyal Customers** demonstrate the highest frequency, emphasizing their repeated purchases.
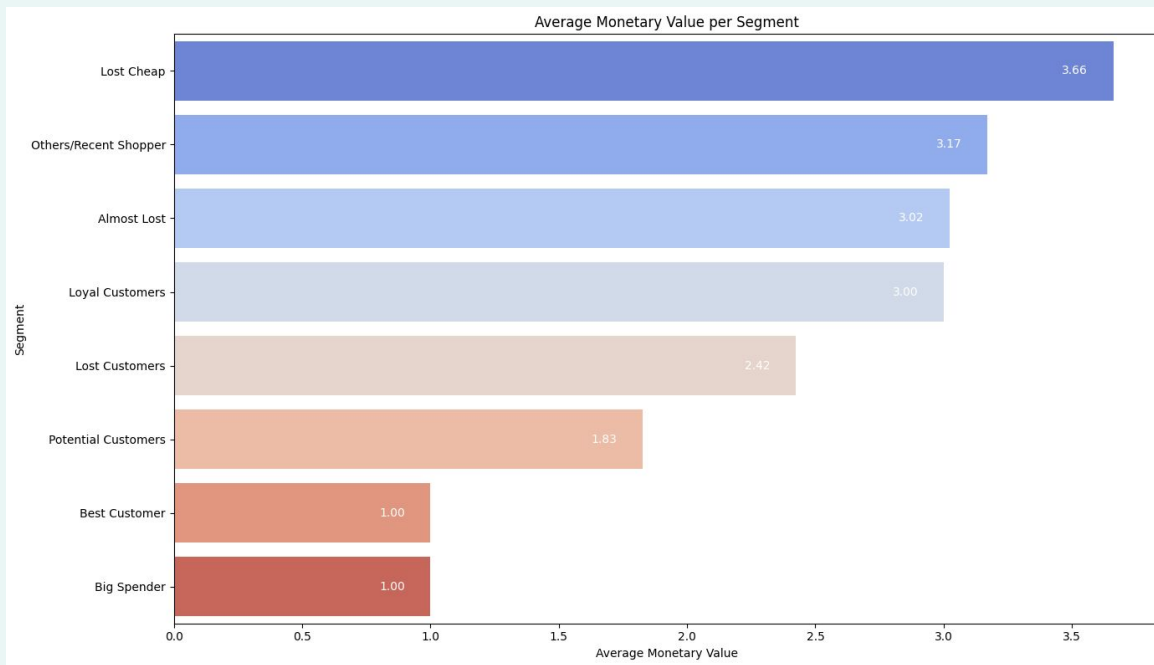   - **Almost Lost** and **Potential Customers** have mid-range frequency, suggesting an opportunity to retain and encourage them to buy more frequently.
3. **Segment Overlap**:
   - **Others/Recent Shoppers** are spread across various recency values but exhibit lower frequency, highlighting their sporadic engagement.

Recommendations:

- **Reactivation Campaigns**: Focus on **Lost Customers** with targeted offers to reduce recency and bring them back into active engagement.
- **Loyalty Rewards**: Strengthen relationships with **Loyal Customers** by introducing exclusive rewards, further increasing their frequency.
- **Upsell Opportunities**: Target **Potential Customers** with personalized offers to elevate their engagement level toward **Loyal Customers** or **Big Spenders**.

Average Monetary Value per Segment

Insights:

1. **Lost Cheap** has the **highest average monetary value (3.66)** despite being in the least desirable segment, likely due to higher single-order values before they disengaged.
2. **Others/Recent Shoppers (3.17)** and **Almost Lost (3.02)** maintain mid-to-high spending averages, indicating potential for re-engagement.
3. **Loyal Customers (3.00)** balance high frequency and steady spending, confirming their value.
4. **Best Customers (1.00)** and **Big Spenders (1.00)** have a lower monetary average due to RFM quartile normalization.

Recommendations:

- **Lost Cheap**: Explore reactivation campaigns, focusing on smaller, frequent offers to reignite interest.
- **Others/Recent Shoppers**: Provide exclusive deals to encourage continued engagement.
- **Loyal Customers**: Offer loyalty rewards to sustain their high frequency and monetary contribution.

Ternary Chart of RFM Segmentation
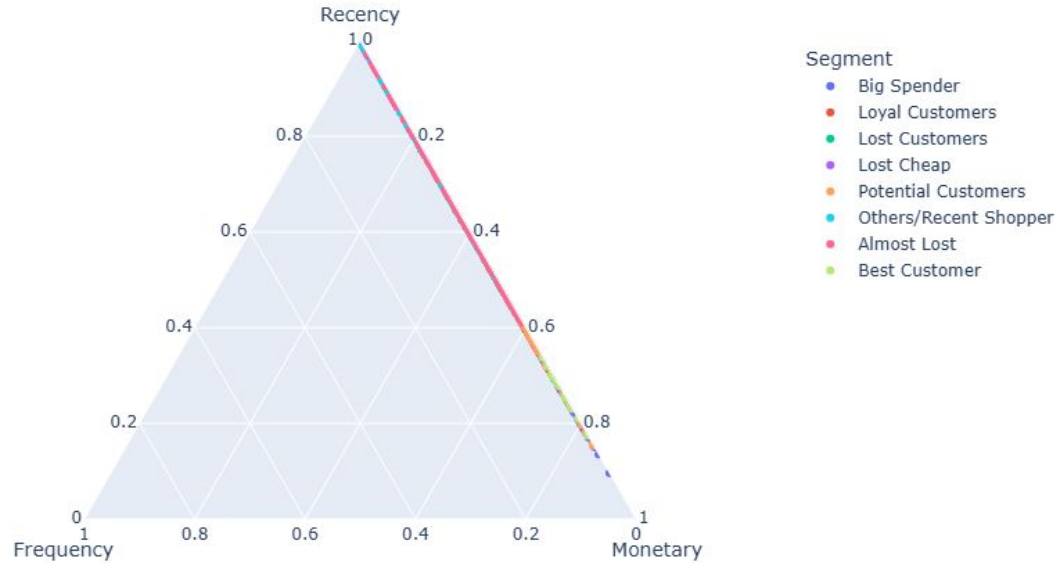
Insights from Ternary Chart:
1. **Big Spenders** and **Best Customers** focus on high **Frequency** and **Monetary** values but lower recency.
2. **Loyal Customers** show balanced engagement across all dimensions.
3. **Almost Lost** and **Lost Customers** need re-engagement due to low frequency and spending.

Actions:
- **Re-engage** lost segments with offers.
- **Reward** Loyal Customers to sustain engagement.
- **Upsell** Big Spenders for more frequent purchases.

"Your most unhappy customers are your greatest source of learning."

—Bill Gates

03

Dashboard

RFM Dashboard

Welcome **Hijir Della Wirasti!**

Year of Order Date	Month	Segment	State	City	RFM Score
2014 ——○—— 2017	(All)	(All)	(All)	(All)	(All)

SALES	PROFIT	#ORDERS	#CUSTOMERS
2,297,201	286,397	37,873	793

EXPORT

PDF

Top RFM Segment

Segment
- Others/Recent Shopper
- Big Spender
- Almost Lost
- Loyal Customers
- Lost Cheap
- Lost Customers
- Potential Customers
- Best Customer

Distinct count of C...
(0, 50, 100, 150)

RFM Segmentation

- Others/Recent Shopper 23.58%
- Almost Lost 15.26%
- Loyal Customers 13.49%
- Big Spender 21.94%
- Lost Cheap 8.95%
- Potential
- Lost Customers 8.32%

RFM Score

Frequency (0K, 2K, 4K)
Recency (0K, 500K, 1000K)

Sales by State

Washington 138,641
Montana 5,589
North Dakota 920
Minnesota 29,863
Oregon 17,431
Idaho 4,382
South Dakota 1,316
Wyoming 1,603
Nebraska 7,465
Iowa 4,580
Michigan 76,270
Nevada 16,729
Utah 11,220
Colorado 32,108
Kansas 2,914
Missouri 22,205
Indiana 53,555
Pennsylvania 116,512
District of Columbia 2,865
California 457,688
Arizona 35,283
New Mexico 4,784
Oklahoma 19,683
Kentucky 36,592
South Carolina 8,482
Texas 170,188
Mississippi 10,771

© 2025 Mapbox © OpenStreetMap

Sales by Top 10 Cities

City
- New York City
- Los Angeles
- Seattle
- San Francisco
- Philadelphia
- Houston
- Chicago
- San Diego

Sales (0K, 50K, 100K, 150K, 200K, 250K)

RFM Dashboard

Year of Order Date	Month	RFM Score	Year	RFM Score	Segment
2014 ——— 2017	(All)	(All)	2022	(All)	(All)

SALES	PROFIT	#ORDERS	#CUSTOMERS
2,297,201	286,397	37,873	793

Trend

90K
80K 733,215
70K
60K 609,206
50K
40K 470,533
30K
20K 484,247
10K

Metric Type

2014 2015 2016 2017

Year of Order Date

Top RFM Segment

Segment

- Others/Recent Shopper
- Big Spender
- Almost Lost
- Loyal Customers
- Lost Cheap
- Lost Customers

0   50   100   150

Distinct count of Customer Id

EXPORT

Sales by Customer Type

- Consumer
- Corporate
- Home Office

Sales by Category

- Furniture
- Office Supplies
- Technology

Sales by Sub Category

- Phones
- Chairs
- Storage
- Tables
- Binders
- Machines
- Accessories
- Copiers
- Bookcases
- Appliances

0K   200K

Sales

Sales by Top 10 Products

- Canon image...
- Fellowes PB5...
- Cisco TelePre...
- HON 5400 Se...
- GBC DocuBin...
- GBC Ibimaste...
- Hewlett Packa...
- HP Designjet ...
- GBC DocuBin...
- High Speed A...

0K 20K   4

Sal
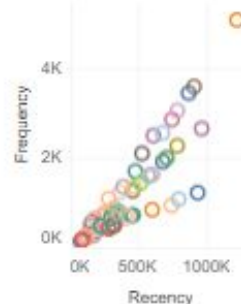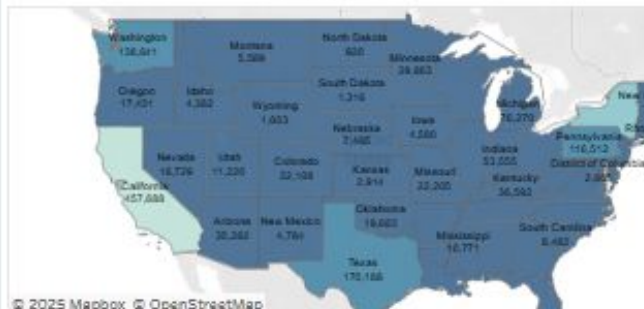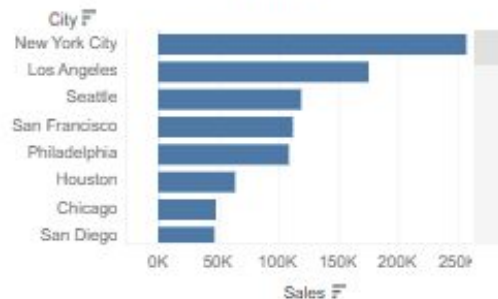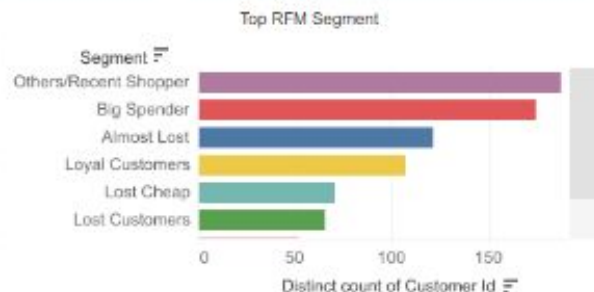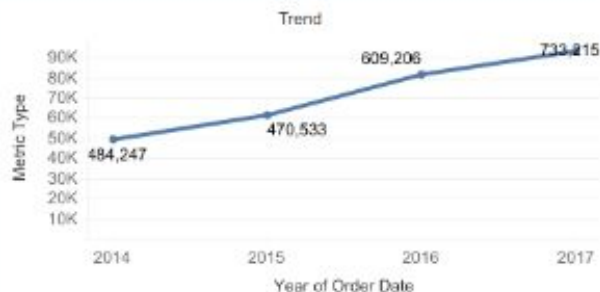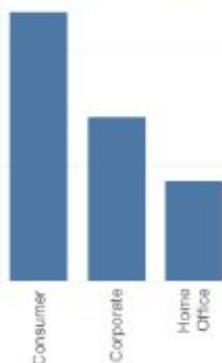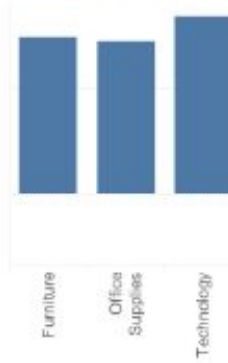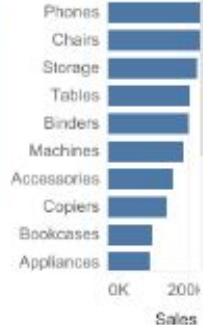
Sales by Top 10 Customers

- Sean Miller
- Tamara Chand
- Raymond Buch
- Tom Ashbrook
- Adrian Barton
- Ken Lonsdale
- Sanjit Chand
- Hunter Lopez
- Sanjit Engle
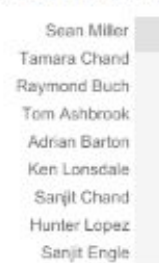
04

Insights and Recommendations

Insights and Recommendations

Segment	Insights
Best Customers	Top-performing customers with the highest recency, frequency, and monetary value. They are the backbone of revenue.
Big Spenders	Customers contributing high revenue but with varying purchase frequency. A crucial segment for maintaining profitability.
Loyal Customers	Customers with consistent purchases, indicating strong brand loyalty and frequent engagement.
Potential Customers	Customers showing potential for higher engagement with moderate recency and frequency.
Almost Lost	Customers showing declining engagement; they are at risk of churn.
Lost Customers	Customers who have churned, requiring significant efforts to re-engage.
Others/Recent Shoppers	Largest segment, contributing moderate revenue but low engagement frequency. Strong potential for growth.
Lost Cheap	Lowest monetary value and high churn risk. Requires cost-efficient reactivation strategies.
General Strategies	Strategies applicable to all segments to optimize customer value and profitability.
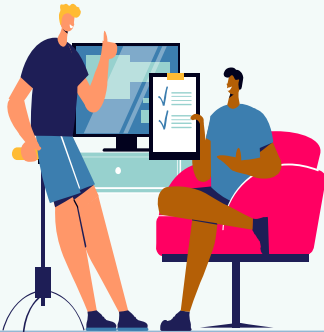
Insights and Recommendations

Segment	Recommendations
Best Customers	Implement VIP programs, exclusive offers, and early access to new products to reward loyalty.
Big Spenders	Focus on upselling and cross-selling through personalized product recommendations.
Loyal Customers	Maintain frequent engagement through targeted campaigns and loyalty rewards.
Potential Customers	Encourage purchases with tailored offers and nurture them towards higher engagement.
Almost Lost	Design reactivation campaigns with time-limited offers to prevent churn.
Lost Customers	Collect feedback to address issues and design tailored reactivation strategies.
Others/Recent Shoppers	Encourage repeat purchases through educational content and first-time buyer discounts.
Lost Cheap	Use low-cost automated campaigns to re-engage or allocate resources to higher-value segments.
General Strategies	Leverage CRM systems for dynamic segmentation, optimize regional strategies, and monitor discount impact.

Insights and Recommendations

- **Overall Insights**

  The analysis reveals **distinct customer behaviors** based on RFM segmentation. Segments such as **"Big Spenders"** and **"Best Customers"** exhibit consistent purchase patterns, higher monetary value, and frequent engagement, making them the **most valuable segments**. In contrast, segments like **"Lost Customers"** and **"Lost Cheap"** contribute minimally, indicating **potential churn or minimal interest**. The **"Others/Recent Shopper"** segment, while sizable, suggests **untapped opportunities**, as they show moderate recency and frequency but lower monetary value. These insights emphasize the importance of **targeted strategies** to maximize value across varying customer groups.

- **Recommendations**

  Focus **retention and loyalty programs** on **"Best Customers"** and **"Big Spenders"** to sustain and grow their contributions. For **"Others/Recent Shopper,"** implement **personalized offers or upselling strategies** to boost their average transaction size. Re-engage **"Almost Lost"** customers through **timely incentives or reminders** to revive their engagement. For **"Lost Customers"** and **"Lost Cheap,"** consider cost-effective strategies such as **email campaigns or discounts** to re-establish interest. Tailored communication and marketing efforts based on these insights will drive **profitability** and **enhance customer satisfaction** across all segments.

Customer retention is the cornerstone of profitability

Who I Am

Hijir Della Wirasti

Leveraging business intelligence and data analytics to deliver actionable insights and optimize strategic decision-making.

Thanks!

Do you have any questions?
hijirdw@gmai.com

https://www.linkedin.com/in/hijirdella/