

Uncovering High-Value Accounts in AWS SaaS Ecosystem

A Data-Driven Segmentation Strategy for Retention
and Growth



TABLE OF CONTENT

Company Background

Business Context & Stakeholder Involved

Data Understanding

Exploratory Data Analysis

Conclusion & Recommendation

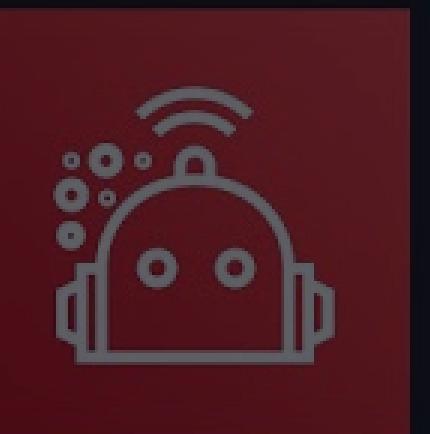
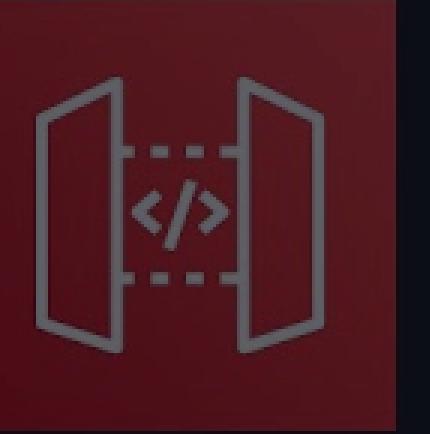
COMPANY BACKGROUND

COMPANY BACKGROUND

Amazon Web Services (AWS) is the cloud computing division of Amazon.com, Inc., providing Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) solutions for individuals, businesses, and institutions around the world.

Our mission:

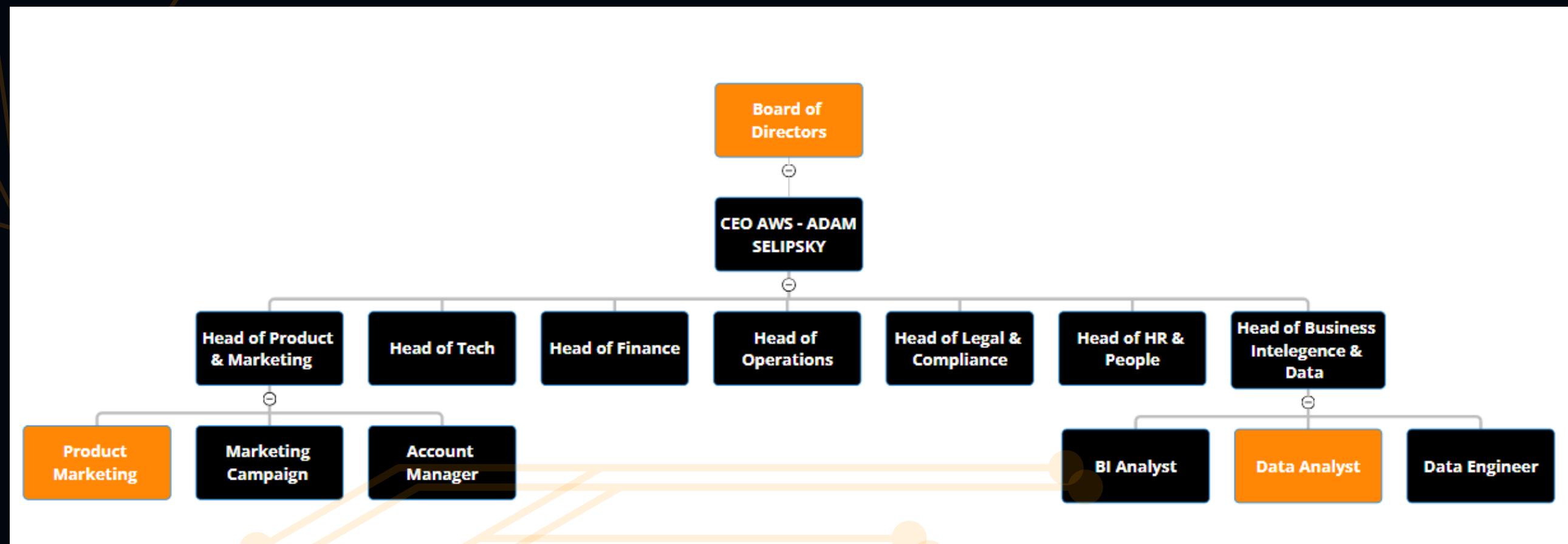
As a global SaaS leader, Amazon Web Services (AWS) drives digital transformation worldwide – empowering startups and enterprises alike to innovate and scale efficiently.

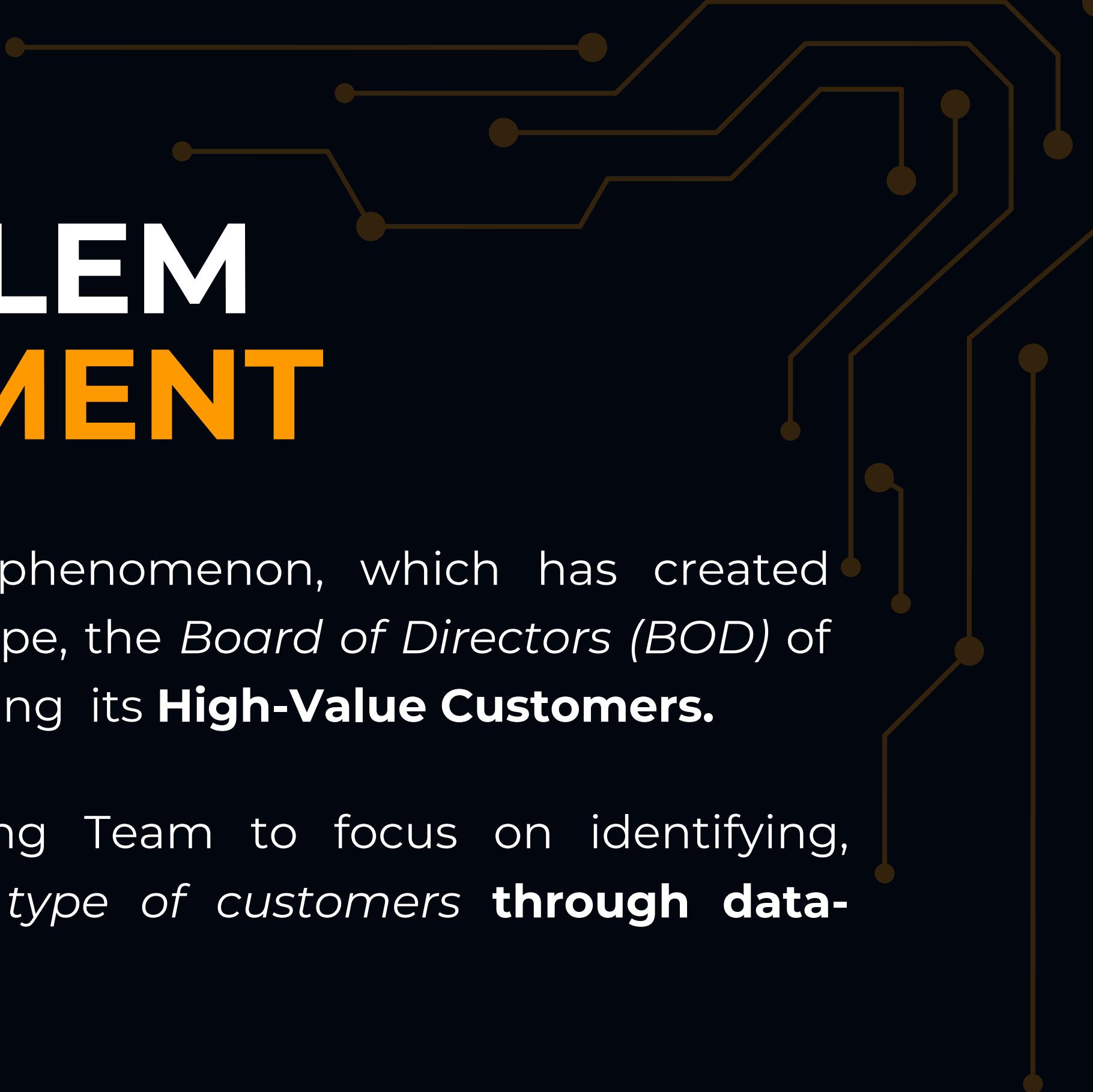


BUSINESS CONTEXT & **STAKEHOLDER INVOLVED**

STAKEHOLDER INVOLVED

For this project, the key stakeholders involved are the **Product Marketing Team** and the **Board of Directors**, who play essential roles in strategic decision-making and project direction.





PROBLEM STATEMENT

Following the recent **winter tech** phenomenon, which has created uncertainty in the global SaaS landscape, the *Board of Directors (BOD)* of AWS has raised concern over maintaining its **High-Value Customers**.

They directed the Product Marketing Team to focus on identifying, understanding, and retaining these *type of customers* **through data-driven insights**.

PROJECT AIM

1. **Identify customers** with the **highest value** (High-Value Customers) based on AWS historical transaction data.
2. Analyze the **characteristics** and **behavioral patterns** of potential High-Value Customers using relevant measurable factors.
3. Formulate the **most effective retention** and **promotional communication strategies** to **maintain** and **enhance the lifetime value (LTV)** of these High-Value Customers.

DATA UNDERSTANDING

DATA UNDERSTANDING

This dataset contains 9,994 SaaS sales transactions and 19 column. Each column represents **key attributes** such as *Order Date, Customer, Industry, Product, Sales, Discount, and Profit.*

No	Column Name	No	Column Name
1	Row ID	11	Customer ID
2	Order ID	12	Industry
3	Order Date	13	Segment
4	Date Key	14	Product
5	Contact Name	15	License
6	Country	16	Sales
7	City	17	Quantity
8	Region	18	Discount
9	Subregion	19	Profit

PRE- DESCRIPTIVE STATISTIC

To understand the overall structure and statistical profile of the dataset before applying the analysis of data.

- Dataset Size : 9,994 transaction records
- Top Country: United States (2,001 records)
- Top City: London (922 records)
- Top Region/Subregion: EMEA – NAMER
- Top Industry: Finance (2,127 records)
- Top Segment: SMB (5,191 records)
- Top Customer: Allianz (192 transactions)
- Most Common Product: ContactMatcher

PRE-DESCRIPTIVE STATISTIC

To understand the overall structure and statistical profile of the dataset before applying the analysis of data.

Sales

Ranges from 0.44 to 22,638.48, with a mean of 229.86 and a high standard deviation (623.25). **Outliers identified.**

Profit

Mean profit 28.65, with wide dispersion (SD 234.26, min -6,599.97 to max 8,399.98).

Strong variance in profitability across customers.

Discount

Mean discount 0.16 (16%), max 80%.

High discounts likely applied.

Quantity

Mean 3.79 units, range 1–14.

Shows mixed transaction sizes typical of SaaS product bundles.

DATA UNDERSTANDING

The following are the columns included in the dataset, categorized based on their context to provide a clearer understanding of each data dimension.

Transaction Details

- Row ID → Unique sequential number for each transaction record
- Order ID → Unique ID for each customer order
- Order Date → Date when the order was placed (date format)
- Date Key → Numeric representation of the date (YYYYMMDD)

Customer Information

- Contact Name → Name of the customer representative placing the order
- Customer → Company or organization name
- Customer ID → Unique customer identifier

Business Context

- Segment → Customer segment based on business scale (SMB, Enterprise, Strategic)
- Industry → Customer's industry or business sector (e.g., Energy, Tech, Healthcare)

Geographical Information

- Country → Country where the customer is located
- City → City where the customer operates
- Region → Major geographic region (e.g., AMER, EMEA, APJ)
- Subregion → More specific geographic division (e.g., NAMER, EU-WEST, JAPN)

Product & License Information

- Product → SaaS product sold (e.g., FinanceHub, ContactMatcher, Alchemy)
- License → Unique license code for the purchased product

Financial Performance

- Sales → Total sales value of the transaction
- Discount → Discount percentage applied (in decimal, e.g., 0.20 = 20%)
- Profit → Net profit after deducting costs and discounts
- Quantity → Number of product units or licenses purchased

DATA CLEANING & DATA PREPARATION



No Missing Value Detected

All columns show **False for missing data**, indicating that the dataset **is clean** and complete at the initial stage.



Order Date Conversion

The Order Date column was originally stored as an object (text) type. To enable time-based and trend analysis, it was **converted** into a proper **datetime format**.



Order Date Conversion

The Order Date column was originally stored as an object (text) type. To enable time-based and trend analysis, it was converted into a proper *datetime format*.



Column Ignored (Not Deleted)

Segment → **Retained** but **excluded** from **Exploratory Data Analysis (EDA)**. This column is used only for **outlier identification**.



Columns Removed

- Row ID → Dropped because it only serves as a sequential index and adds no analytical value.
- License → Removed since it does not contribute to the objectives of this analysis.



New Columns Added

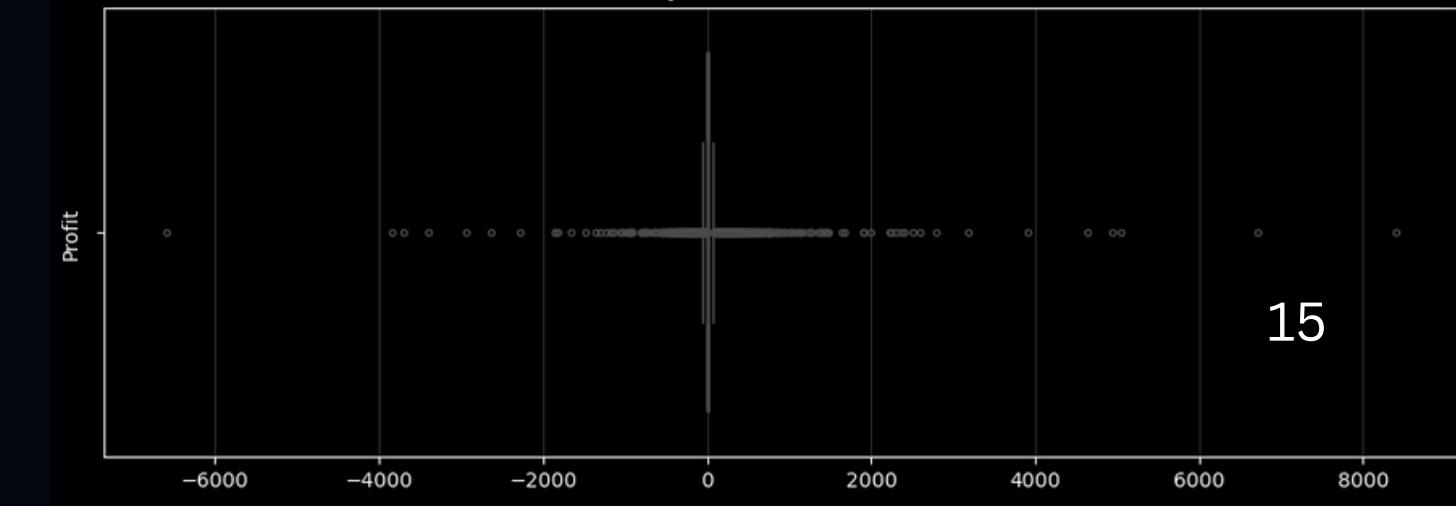
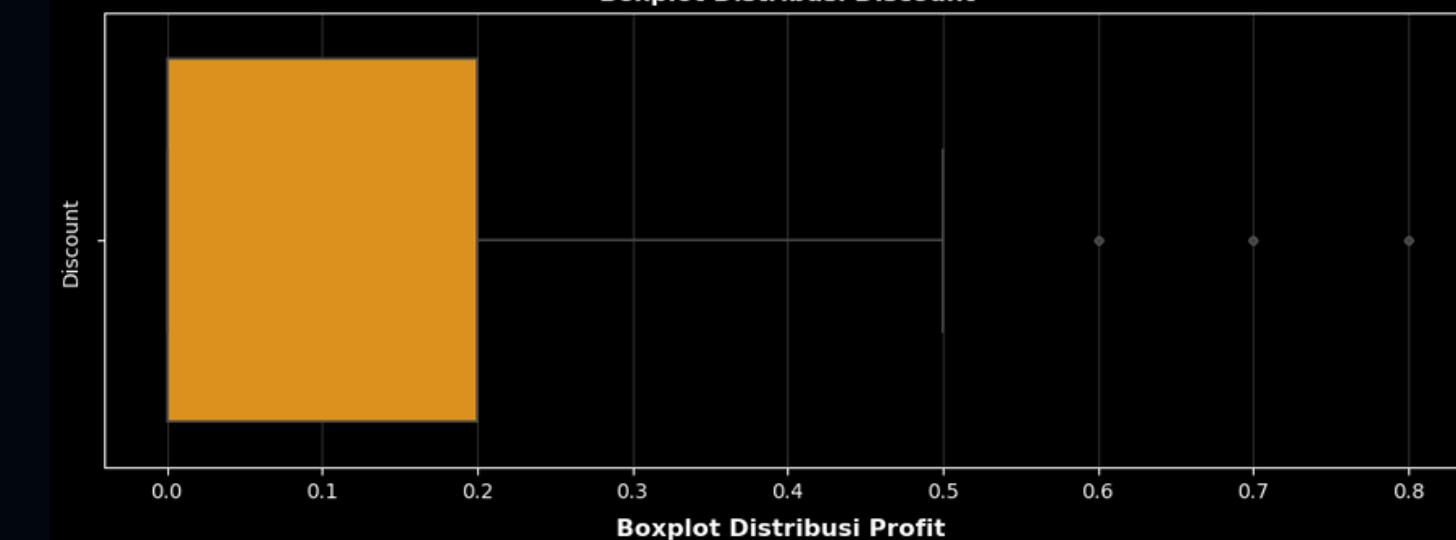
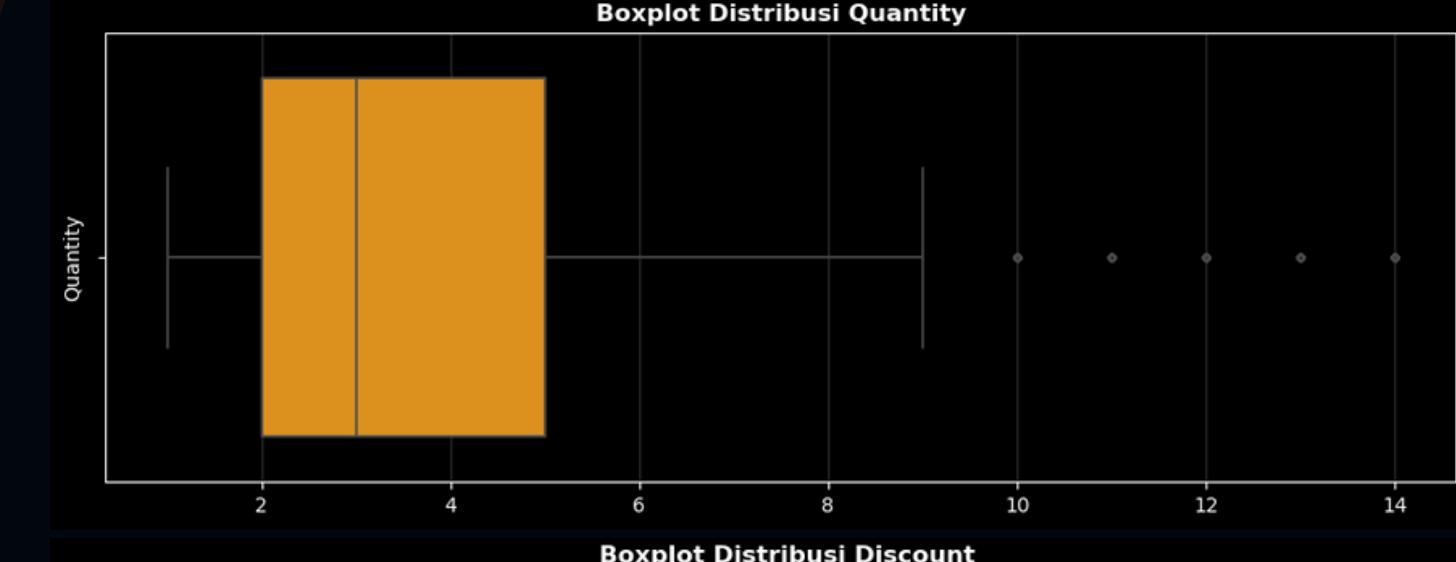
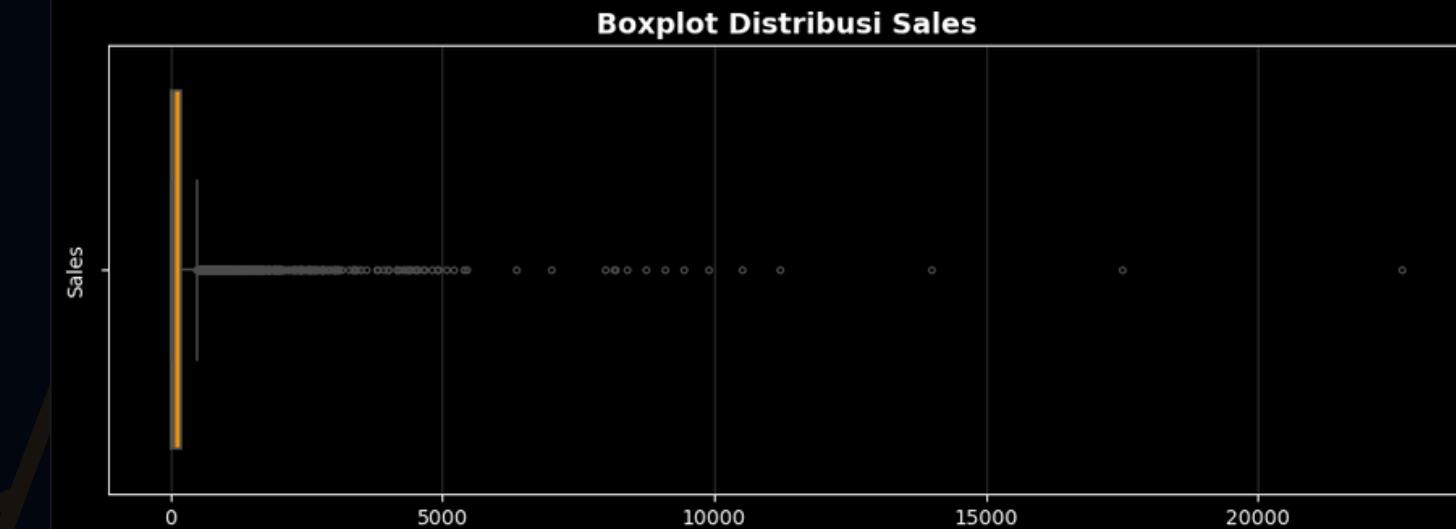
Year → Extracted from transaction dates to analyze yearly sales and profit trends. Another 10 more column also added for the LRFM model, which will be explained further in the EDA section.

OUTLIER IDENTIFICATION SUMMARY

Extreme values were detected in four transaction columns — **Sales**, **Quantity**, **Discount**, and **Profit**. These outliers provide valuable insights into **pricing strategies**, **customer scale differences**, and **profitability dynamics across segments**.

These outliers are not anomalies

The outliers observed here are not anomalies, but rather **a reflection of AWS SaaS market realities** and **its diverse profitability patterns**. Therefore, these values should not be removed, as they represent legitimate variations in normal business conditions.



FINAL DATASET

The final dataset consists of the original business and transactional variables, along with several newly engineered columns (highlighted in orange) created to support LRFM (Length, Recency, Frequency, Monetary) analysis.

Order ID	Industry	LRFM Score	Frequency
Order Date	Segment	L_Score	Monetary
Date Key	Product	R_Score	
Contact Name	Sales	M_Score	
Country	Quantity	R_Score	
City	Discount	F_Score	
Region	Profit	M_Score	
Subregion	Year	Length	
Customer ID	Customer Value	Recency	

EXPLORATORY DATA ANALYSIS

LIFETIME VALUE ANALYSIS USING THE LRFM METHOD

The LRFM model provides a comprehensive view of customer behavior:

- Length → How long the customer has been **active**.
- Recency → How **recently** they made their last transaction.
- Frequency → How **often** they make transactions.
- Monetary → **How much profit** they generate.

By analyzing these four dimensions, businesses can identify loyal, profitable customers and prioritize them in retention and growth strategies.

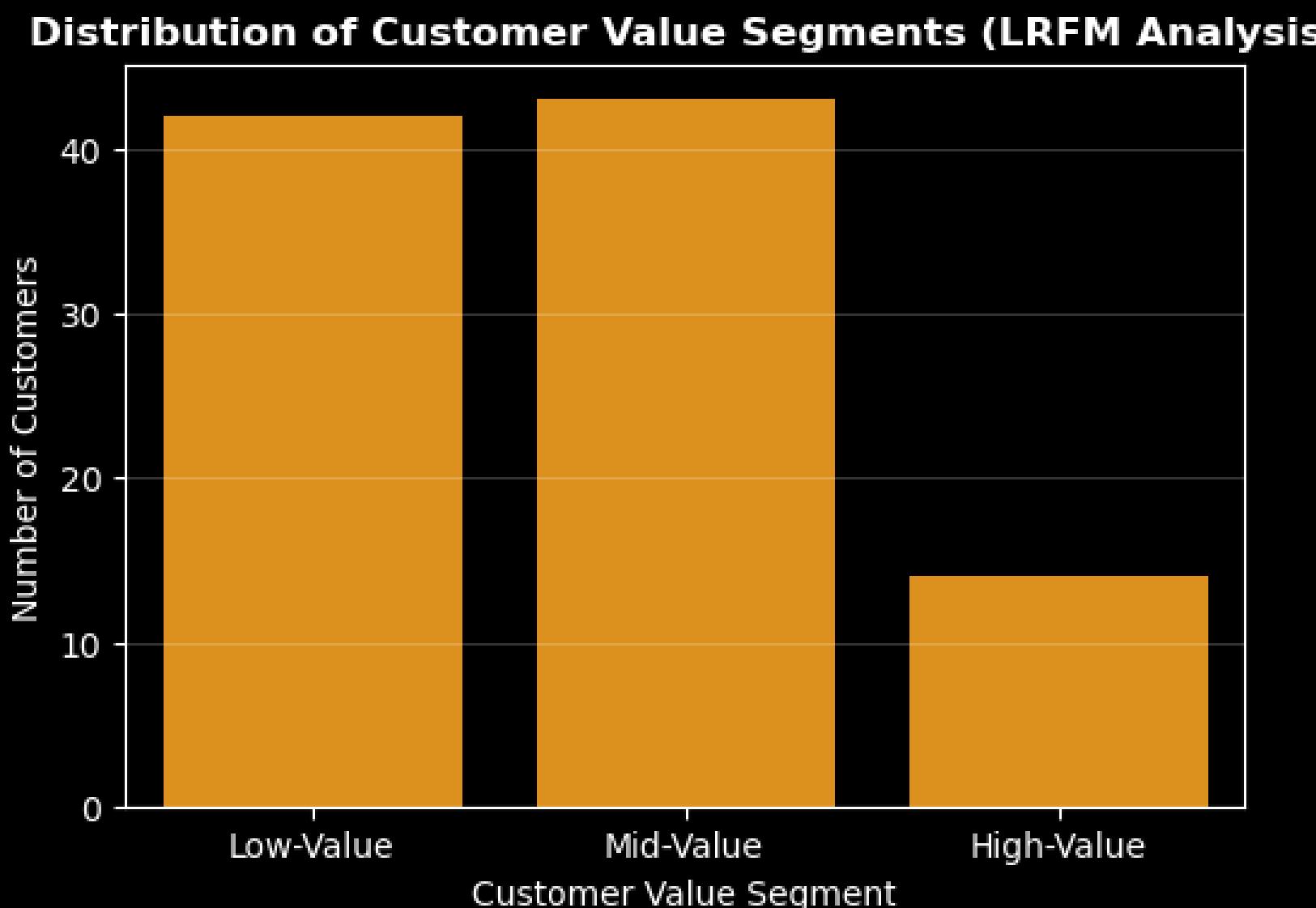


HIGH-VALUE CUSTOMER IDENTIFICATION

In the LRFM analysis , a customer is categorized as a **High-Value Customer** if they meet the following four characteristics:

1. **Recently active** in transactions (Low Recency)
2. Have maintained **a long-term relationship** with the company (High Length)
3. Purchase **frequently** (High Frequency)
4. And **generate substantial profit** (High Monetary).

DISTRIBUTION OF CUSTOMER VALUE SEGMENTS (LRFM ANALYSIS)



This chart illustrates the distribution of customers across LRFM-based value segments:

- Low-Value and Mid-Value segments dominate, representing **the largest proportion** of customers in the dataset.
- The High-Value segment is **significantly smaller (14)**, indicating that **only a limited group** of customers generate **the majority of profit and long-term value**.

DISTRIBUTION OF CUSTOMER VALUE SEGMENTS (LRFM ANALYSIS)

From a business perspective, this distribution highlights that:

- AWS should retain and nurture **High-Value Customers** due to their strong contribution to profitability
(Project Focus)
- The Mid-Value segment presents growth potential – these customers can be upgraded through targeted marketing and upselling.
- The Low-Value group can be addressed with automation or low-cost engagement strategies to maintain efficiency.

LIST OF HIGH-VALUE CUSTOMER

LRFM Descriptive Summary

Length: an average of **1,416 days (~3.8 years)**

Recency: An average of **4 days**

Frequency: An average of **68 transactions**,

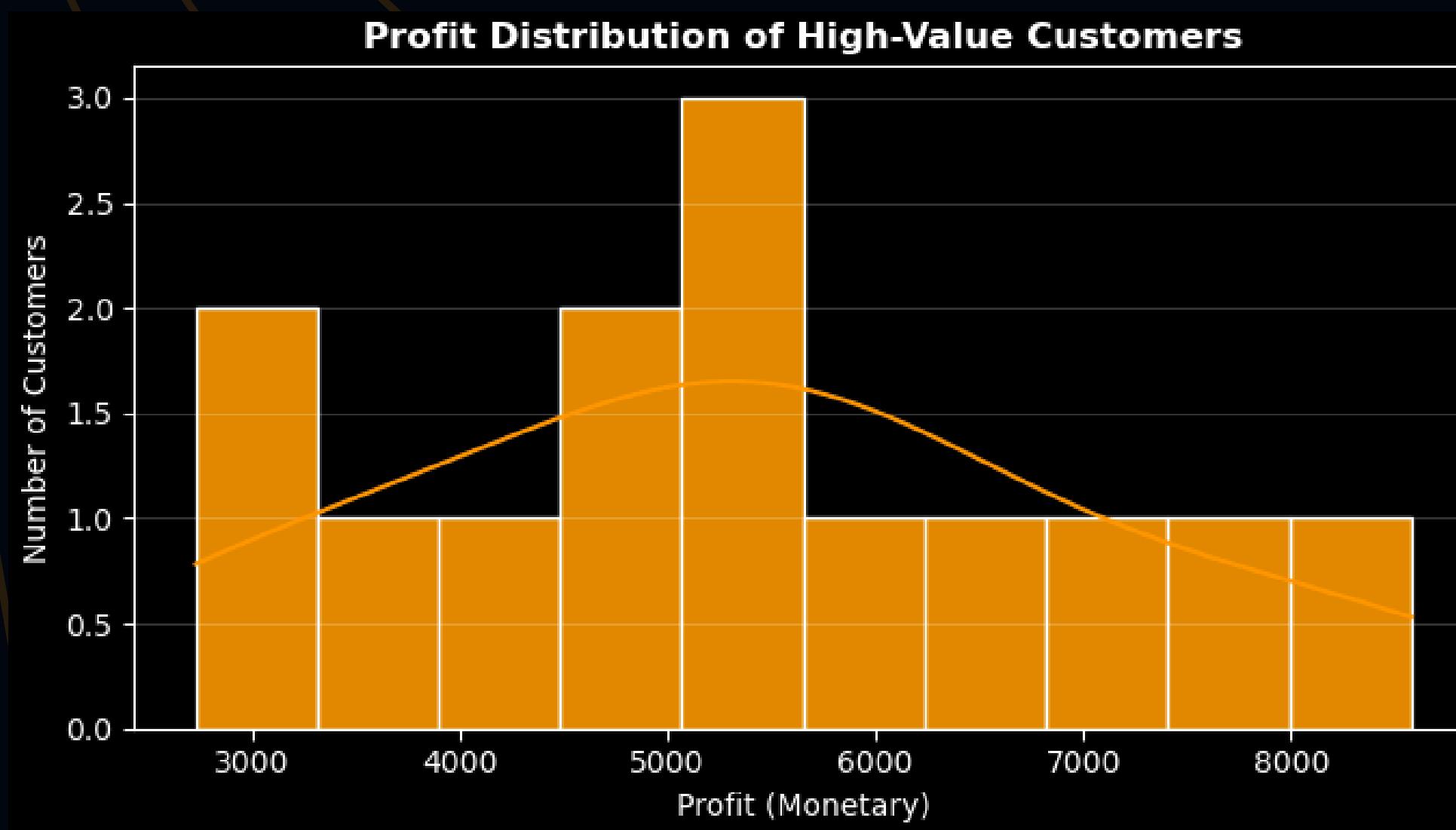
Monetary: An average profit of approximately **5,371**

*the LRFM scoring scale used in this analysis ranges from 1 to 4

Customer	L_Score	R_Score	F_Score	M_Score	LRFM_Score
Kroger	4	4	4	4	16
Johnson & Johnson	4	4	4	4	16
Mondelez International	4	4	4	4	16
Lukoil	4	4	3	4	15
Allianz	3	4	4	4	15
American Express	3	3	4	4	14
Chevron	3	4	4	3	14
BNP Paribas	2	4	4	4	14
Anthem	4	3	3	4	14
AmerisourceBergen	3	4	4	3	14
Lowes	4	3	3	4	14
Prudential Financial	4	3	4	3	14
Royal Dutch Shell	4	4	3	3	14
UnitedHealth Group	3	4	3	4	14

AWS High-Value Customers demonstrate long-term loyalty, recent and consistent engagement , and strong profitability, reflecting a stable 22 and highly valuable customer relationship.

DISTRIBUTION OF CUSTOMER VALUE SEGMENTS (LRFM ANALYSIS)



AWS **High-Value Customers** demonstrate **strong and stable profit distribution**, reflecting long-term relationships and high satisfaction levels.

They not only transact frequently but also contribute consistent profit values, indicating a healthy and well-managed premium customer base.

HIGH-VALUE CUSTOMER X INDUSTRY

Industry	Monetary	Frequency	Recency	Length
Consumer Products	8578	61	3	1423
Retail	7011	68	4.69	1432.45
Healthcare	4898.	62	4.29	1423.22
Finance	4840	83	4.49	1392.06
Energy	4454	64	3.81	1424.54

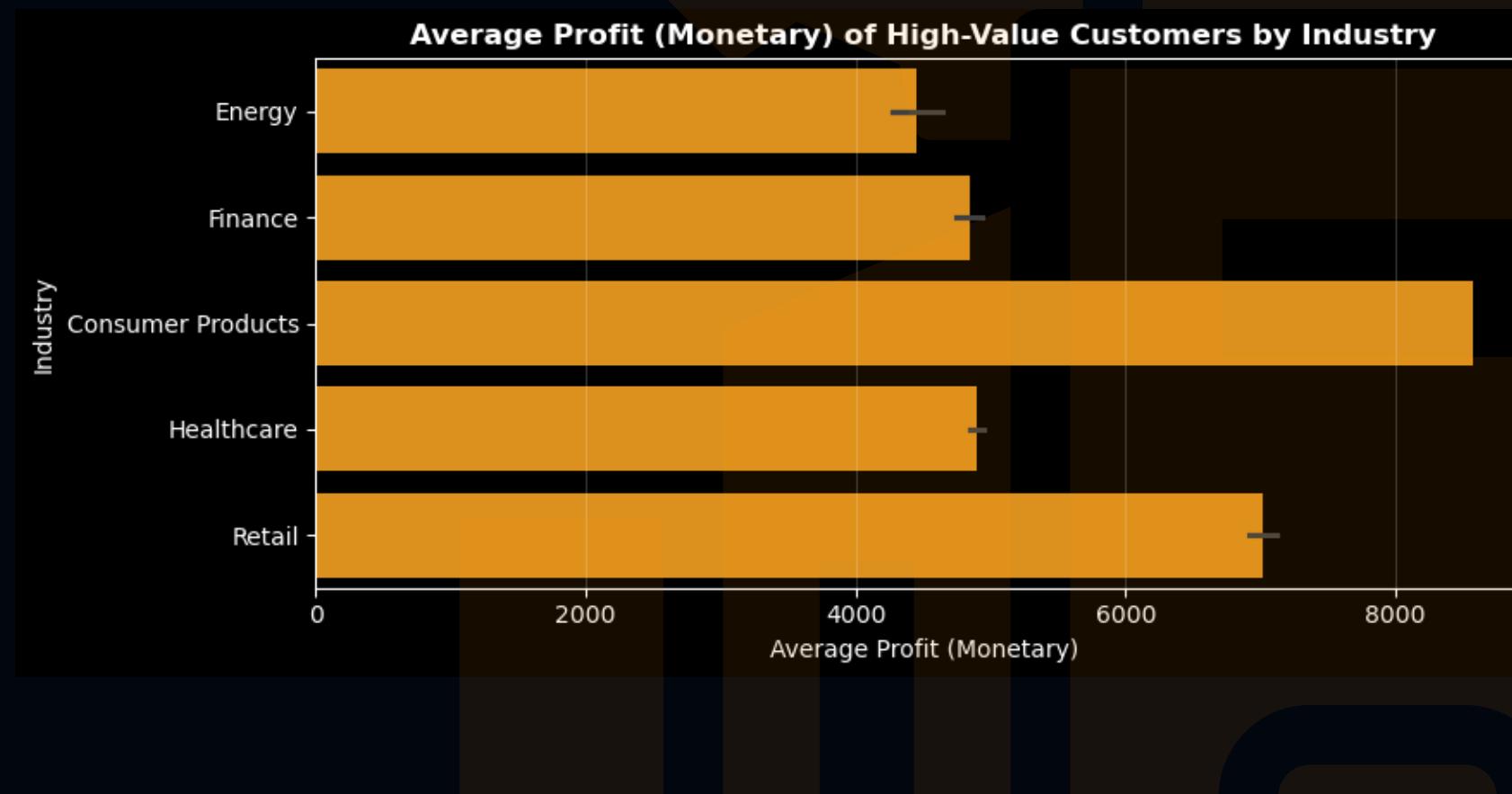
*agg. by mean

High-Value Customers are **found in only five industries** — **Energy, Finance, Consumer Products, Healthcare, and Retail.**

All industries show a **consistently high Length (~1,400 days)**, indicating that AWS High-Value Customers **maintain long-term relationships** of approximately four years with AWS services.

Meanwhile, profitability (Monetary) and transaction activity (Frequency) **vary across industries**, reflecting **diverse engagement levels** and **business dynamics**.

HIGH-VALUE CUSTOMER X INDUSTRY



High-Value Customers are concentrated in five industries, each showing distinct behavioral patterns:

1. Consumer Products : **Highest profits** (8,577) & **strong loyalty** (~4 years)
2. Finance : **Highest frequency** (83) but **lower profit margins**.
3. Healthcare : **Moderate Profit** (4898) and **steady engagement**
4. Energy : **Lower profit** (4,453) but **recent reactivation** (3.8) and **high length** (1424)
5. Retail : **High Profit** (7011) & the **most loyal** (Length 1432)

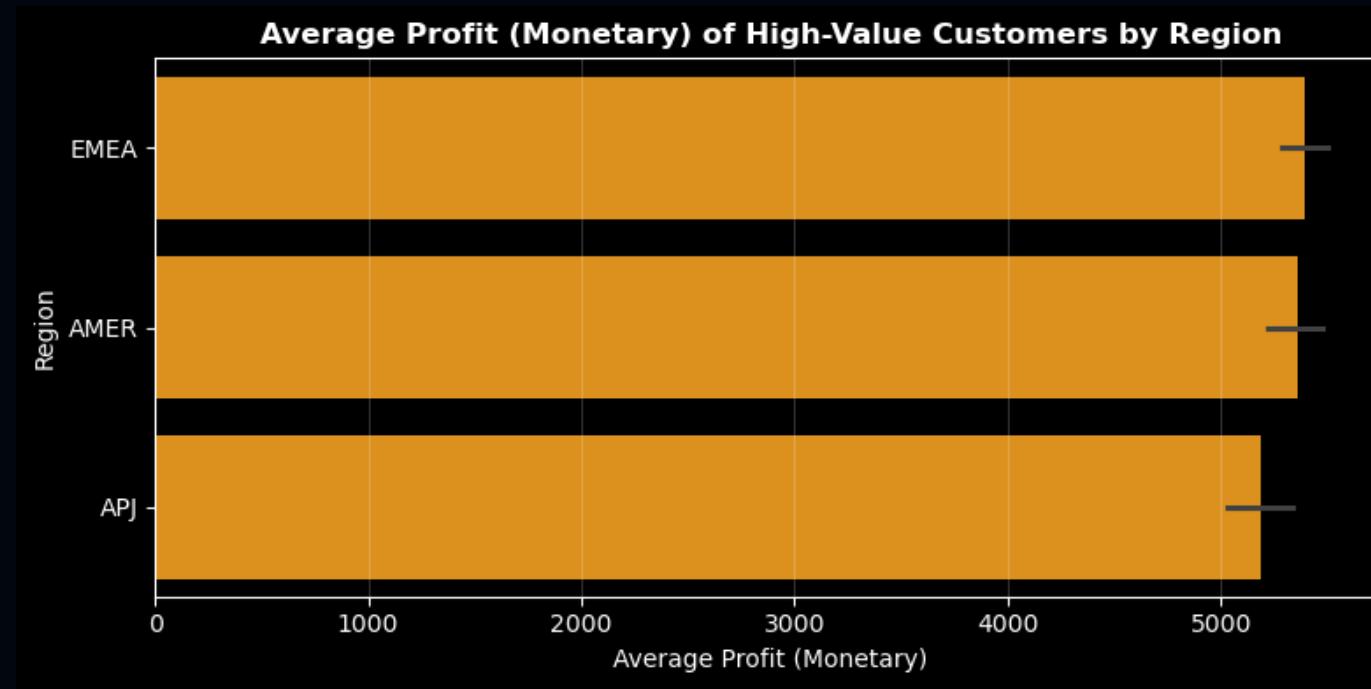
HIGH-VALUE CUSTOMER X INDUSTRY

Cross-Industry Behavioral Insights:

- Consumer Products & Retail → **Profit Leaders** : Large contracts, high profitability, and **strong loyalty**.
- Finance → **Activity Leader** : Frequent transactions but lower margins.
- Healthcare → **Steady Users** : Consistent engagement and stable profit.
- Energy → **Reactivated Clients** : Recently active again with growth potential.

AWS HIGH-VALUE CUSTOMERS ARE **DIVERSE IN BEHAVIOR** — WHILE ALL SHOW **STRONG LOYALTY** ($LENGTH \approx 1,400$ DAYS), THEIR PROFITABILITY, FREQUENCY, AND ENGAGEMENT PATTERNS **VARY** NOTABLY ACROSS INDUSTRIES

HIGH-VALUE CUSTOMER REGIONAL ANALYSIS



Region	Monetary	Frequency	Recency	Length
EMEA	53.915	698	43	14.143
AMER	53.566	702	42	14.155
APJ	51.860	714	42	14.125

*agg. by mean

1. EMEA (Europe, Middle East & Africa)

- Highest profit (5,391.49) — largest contribution to total profit.
- Stable activity (Frequency = 69.79) and recent engagement (Recency = 4.25).
- High Length (1,414 days) → strong long-term loyalty (~4 years).

2. AMER (North & South America)

- Second-highest profit (5,356.57)
- Highest Length (1,415.5 days) → longest engagement and strongest stability.
- High Frequency (70.15) and lowest Recency (4.19) → customers are active and recently transacted.

3. APJ (Asia-Pacific & Japan)

- Slightly lower profit (5,186) but highest Frequency (71.39) → frequent transactions, though smaller in value.
- Low Recency (4.21) → customers are recently active.
- Shorter Length (1,412) → newer customers but fast-growing adoption.

HIGH-VALUE CUSTOMER REGIONAL ANALYSIS

Key Takeaways:

- EMEA leads in **profitability**,
- AMER excels in **loyalty** and **stability**,
- APJ shows **high growth potential** with **strong engagement**.

Across all regions, AWS High-Value Customers **maintain long-term loyalty** (~1,400 days) and **active engagement**.

HIGH-VALUE CUSTOMER REGIONAL ANALYSIS (COUNTRY)

1) Monetary (Profit)

- Top 5 countries (highest profit):
 - a.Croatia ≈ 8,578
 - b.Israel ≈ 6,968
 - c.Turkey ≈ 6,917
 - d.Ukraine ≈ 6,428
 - e.UAE ≈ 6,092
- Eastern Europe & Middle East dominate the high-profit cluster.
- Mature SaaS markets (≈ 5,000–5,500): Canada, France, Germany, Japan, Singapore, South Korea, UK.
- Lowest: Qatar ≈ 2,828 → high-value customers with limited transaction volume.

2) Frequency (Activity)

- Top 5 countries (highest activity):
 - a.Netherlands (81)
 - b.Taiwan (76.2)
 - c.Poland (77.4)
 - d.Philippines (74.8)
 - e.India (74.2)
- These markets show frequent small-mid transactions.
- Lowest frequency: Qatar (57), Croatia (61), Turkey (61) → low frequency but high profit ⇒ large-value deals.

HIGH-VALUE CUSTOMER REGIONAL ANALYSIS (COUNTRY)

3) Recency

- Nearly all countries have Recency < 6, indicating recent activity.
- Most recent: Israel (2.56), Czech Republic (2.75), Taiwan (2.75) → high engagement.
- Longest since last transaction: China (5.67), Costa Rica (5.60), Greece (5.38) → re-engagement opportunity.

4) Length (Loyalty Duration)

- Consistent at ~1,400–1,440 days → average ~3.8 years with AWS.
- Longest: Austria (1,439), Ireland (1,431), Turkey (1,430) → long-standing, stable clients.
- Shortest: Netherlands (1,360), Indonesia (1,377) → newer but fast-growing bases.

HIGH-VALUE CUSTOMER REGIONAL ANALYSIS (COUNTRY)

Region	High-Value Characteristics
Western & Northern Europe	High Frequency & Length (Netherlands, Germany, UK, France, Finland)
Eastern Europe & Middle East	High Profit but Low Frequency (Croatia, Turkey, Israel, UAE)
Asia-Pacific	Very High Frequency (India, Taiwan, Indonesia, South Korea) but Moderate Profit
Americas	Balanced Profit & Frequency (USA, Canada, Brazil, Mexico)

HIGH-VALUE CUSTOMER REGIONAL ANALYSIS (COUNTRY)

High-Value customers are **active and long-tenured across countries**—but **profit leaders cluster in Eastern Europe/Middle East, activity leaders in APAC/EU**, and **a few markets signal re-engagement potential**.

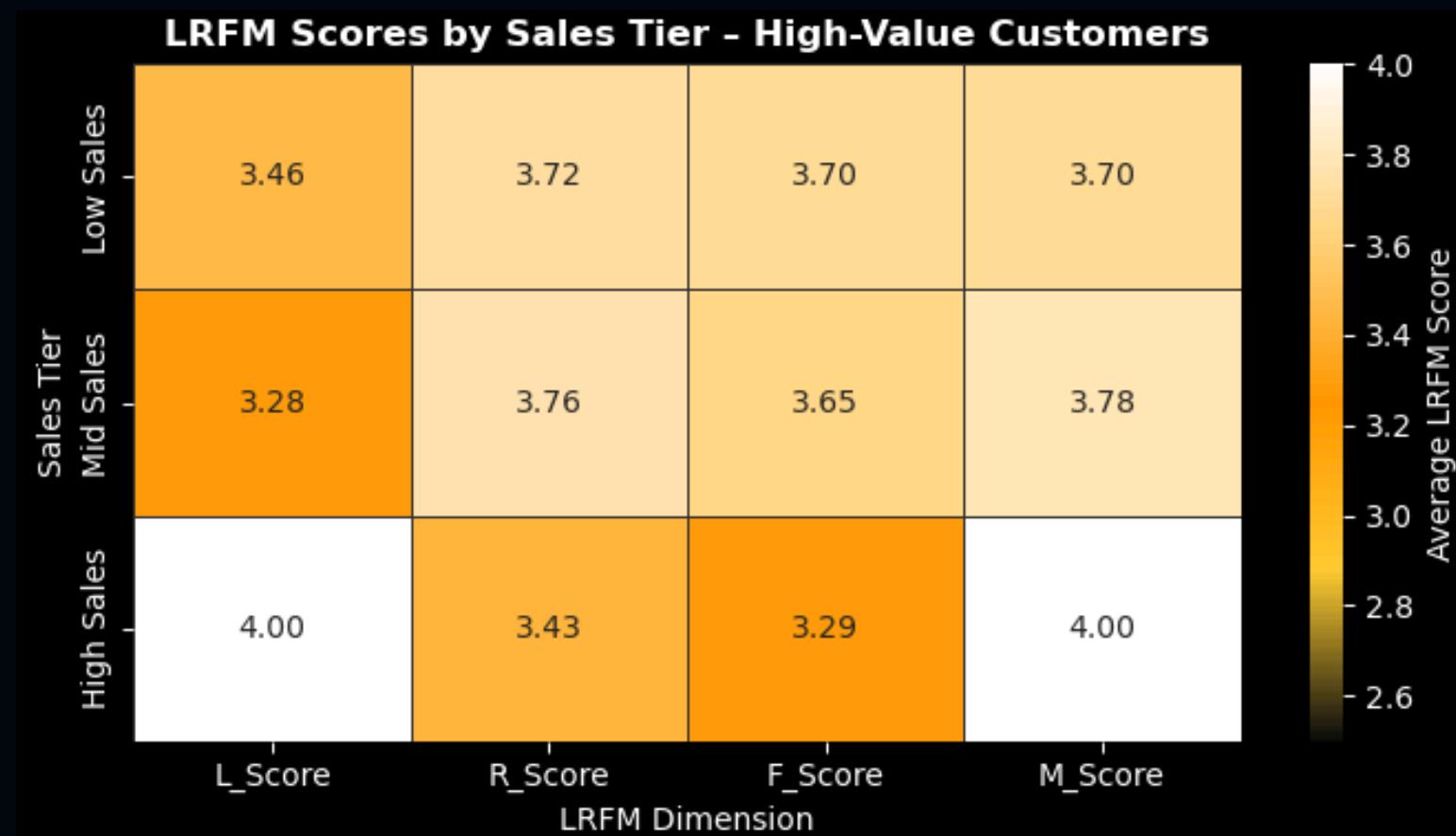
SALES VOLUME DISTRIBUTION AMONG HIGH-VALUE CUSTOMERS

Sales Tier	Number of Customers	Proportion	Key Insights
Low Sales (0-1000)	1,844	≈ 95.5%	Dominant segment — the majority of <i>High-Value</i> customers actually record low sales volume , reflecting small but loyal accounts.
Mid Sales (1000-5000)	82	≈ 4.2%	Minority segment with potential for growth — could upgrade to higher tiers with sustained engagement.
High Sales (>5000)	7	≈ 0.3%	Very few large customers — but they contribute a significant share of total profit .

The data shows an **extreme imbalance** across *sales tiers*, meaning that average-based or cross-tier comparisons (e.g., mean L_Score, R_Score, etc.) may be biased due to the small sample size, particularly within the High Sales segment.

LRFM SCORE BY SALES VOLUME AMONG HIGH-VALUE CUSTOMERS

Sales Tier	L_Score	R_Score	F_Score	M_Score
High Sales	4	3.43	3.29	4
Mid Sales	3.28	3.76	3.65	3.78
Low Sales	3.46	3.72	3.7	3.7



The LRFM analysis reveals distinct behavioral profiles among AWS High-Value Customers across sales tiers:

- High Sales customers : **longest relationships** and **highest profit**, **transact less frequently** and **less recently** — reflecting an **enterprise-contract model with large but infrequent renewals** or commitments.
- Mid Sales customers : **most active, high frequency** and **strong profit potential**, acting as **the core driver of sustained revenue**.
- Low Sales customers : **active** and **recent**, but **operate at smaller transaction scales** — representing a **promising growth segment** that can be nurtured through **upselling** or **usage-expansion programs**.

CONCLUSION & RECOMMENDATION

HIGH-VALUE CUSTOMER

High-Value Customers represent clients with **active engagement, strong loyalty, frequent transactions, and consistent profit contribution.**

They are the **top-priority segment** because they have a **direct impact** on the **company's revenue stability** and **business growth**, especially in SaaS enterprises like AWS.





HIGH-VALUE CUSTOMER

In the LRFM analysis (Length-Recency-Frequency-Monetary), a customer is categorized as a **High-Value Customer** if they meet the following four characteristics:

1. **Recently active** in transactions (Low Recency)
2. Have maintained **a long-term relationship** with the company (High Length)
3. Purchase **frequently** (High Frequency)
4. And **generate substantial profit** (High Monetary).

INDUSTRY BASED INSIGHT

Based on the *industry analysis*, High-Value Customers were found to contribute mainly within **five industry** categories — ***Consumer Products, Retail, Finance, Healthcare, and Energy.***

In the comparison across these five industries, it was revealed that AWS High-Value Customers **are not homogeneous**.

Each industry **exhibits distinct patterns** of **profitability, activity, and loyalty**.

INDUSTRY BASED INSIGHT

This table highlights that AWS **should apply a targeted and industry-specific strategy** to strengthen **retention** and **maximize the lifetime value (LTV)** of its High-Value Customers.

Industry	Key Characteristics	Recommended AWS Strategy
<i>Consumer Products</i>	High profit, low frequency	Segmented Retention – Premium retention programs & enterprise contracts
<i>Retail</i>	Strong loyalty, stable profit	Enterprise Loyalty Program – Loyalty initiatives & AI-based upselling
<i>Finance</i>	High frequency, low margin	Margin Growth Strategy – Margin optimization & compliance services
<i>Healthcare</i>	Consistent activity, loyal customers	Healthcare Specialization – Data governance & HIPAA compliance support
<i>Energy</i>	Low profit, newly active customers	Optimize Reactivation Program – Cross-sell IoT & predictive analytics solutions

REGIONAL INSIGHTS ON HIGH-VALUE CUSTOMERS

Based on the *statistical comparison*, there are no **extreme differences across regions**, but **AWS can optimize local potential by leveraging each region's unique strengths**:

- EMEA (Europe, Middle East & Africa) leads in **profitability**,
- AMER (North & South America) excels in **customer loyalty**, and
- APJ (Asia-Pacific & Japan) stands out in **transaction frequency** and **activity**.

REGIONAL INSIGHTS ON HIGH-VALUE CUSTOMERS (COUNTRY)

Globally, AWS High-Value Customers demonstrate strong engagement and loyalty across countries (Length ≈ 1,400 days; Recency < 5). However, regional differences emerge in profitability and transaction activity patterns:

Region	Customer Characteristics	Recommended Strategy
Eastern Europe / Middle East	High margins but low transaction frequency.	Develop Premium Enterprise Support with personalized account management and high SLA standards to retain large-value customers with infrequent transactions.
APAC (Asia-Pacific)	Relatively lower margins but high transaction frequency.	Drive Growth Expansion Programs , focusing on upselling and volume-based incentives to increase transaction value per customer.
Western Europe / Americas	Mature, stable markets with high customer retention.	Maintain loyal customer relationships through Modernization Credits , long-term contracts , and cloud infrastructure enhancement programs .

HIGH-VALUE CUSTOMER SALES INSIGHT BASED

Globally, AWS High-Value Customers demonstrate strong engagement and loyalty across countries (Length ≈ 1,400 days; Recency < 5). However, regional differences emerge in profitability and transaction activity patterns:

Sales Tier	Proportion	Behavioral Summary	Key Insight	Recommended Strategy
Low Sales	95.5%	Majority of High-Value Customers fall under this tier — small sales volume but high activity and loyalty.	Active, loyal, but not heavy spenders. Represent a large and engaged base.	Ideal for Growth / Enablement Programs to increase transaction size and cross-product adoption.
Mid Sales	4.2%	Minority segment with high retention and second-highest frequency (only 0.05 below Low Sales).	Active and loyal customers	Focus on Upselling and Enterprise Growth Expansion Programs to move them toward higher-value tiers.
High Sales	0.3%	Very small group (7 out of 1,933) but contributes the largest profit share (M_Score = 4) .	Few in number, but key revenue drivers with large, infrequent transactions.	Treat as VVIP accounts via Strategic Account Management and Personalized Relationship-Based Engagement .

BONUS INSIGHT

POTENTIAL MID-VALUE CUSTOMER (HIDDEN GEM)