Customer Segmentation for Enhanced Retention and Personalized Marketing

Analisis Data Nasabah PT Bank Mandiri

Background

- Context: Competition in the banking industry is intensifying with the rise of digital banks. PT Bank Mandiri must retain its loyal customer base while improving the efficiency of its marketing spend. Traditional one-size-fits-all campaigns are no longer effective, as conversion rates have declined by 10% compared to the previous year.
- Project Objective: To identify distinct customer segments based on behavioral patterns, enabling the Marketing and Product teams to design personalized campaigns, relevant loyalty programs, and proactive measures to prevent customer churn.

Selanjutnya >

Current Focus

- Current Situation: Acquisition costs are rising while customer engagement remains
- stagnant.Objective: To segment customers for more personalized and targeted
- campaigns.Methodology: Analyze transaction data, CRM interactions, and digital behavior patterns.

Selanjutnya ›

Current Focus

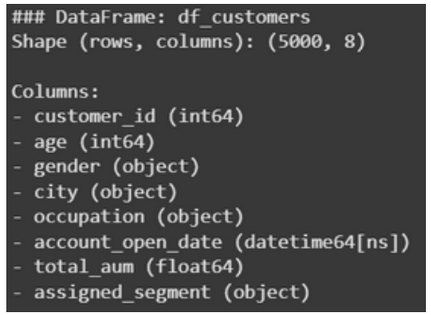
01 Data Sources (Customers, Transactions, CRM, Digital Logs)

02 Preprocessing & Feature Engineering

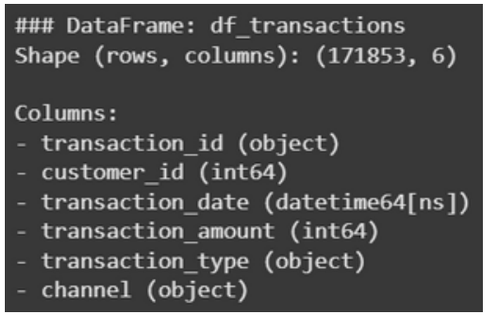
03 Clustering / Segmentasi

04 Analisis & Rekomendasi

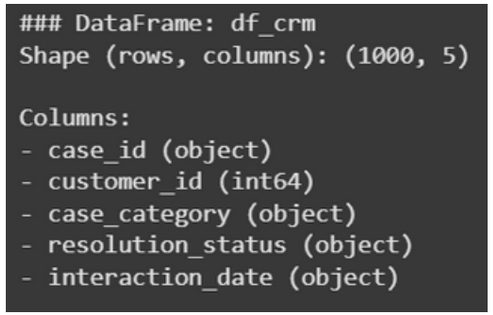
Dataset



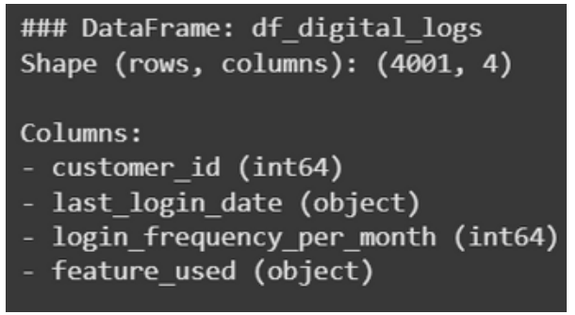
customers.csv



transactions.csv



crm_cases.csv



digital_logs.csv

The dataset used in this analysis is not sourced from PT Bank Mandiri's internal data, but is a synthetic dataset created specifically for this project. This approach was chosen to simulate real-world data conditions without compromising customer privacy or confidentiality. Each persona is designed with distinct behavioral and demographic characteristics, such as age range, transaction volume, occupation type, and predefined AUM (Assets Under Management).

The dataset was programmatically generated using Python, leveraging libraries such as Pandas and Faker. We defined five distinct customer personas, each with unique behavioral profiles—ranging from high-value loyal customers to those at risk of churn.

Data Processing & Feature

df_customers (Data Demografi Nasabah)

5,000

5 segmen: Potential Loyalists, At-Risk Customers, Dormant, The Champions, Newcomers

99 cities in indonesia

18–65 years

df_transactions (Data Transaksi)

There are 171,853 transactions recorded from 4,923 active customers who are regularly engaging in financial activities.

The transaction period spans from 2021 to 2025, capturing customer activity and behavioral trends over a five-year timeframe.

Three primary types of transactions: Transfer, Payment, and Cash Withdrawal.

df_digital_logs (Data Aktivitas Digital)

4,001 customers are actively engaged in digital banking activities.

login frequency varies significantly, ranging from 1 to 50 times per month.

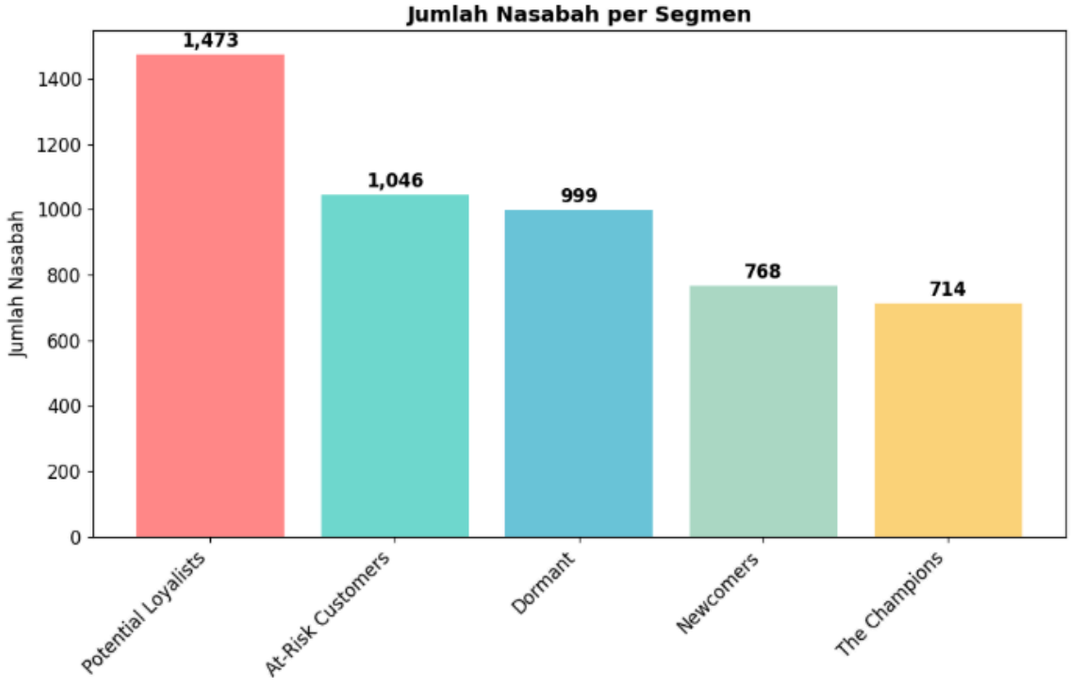
df_crm (Data Interaksi Layanan)

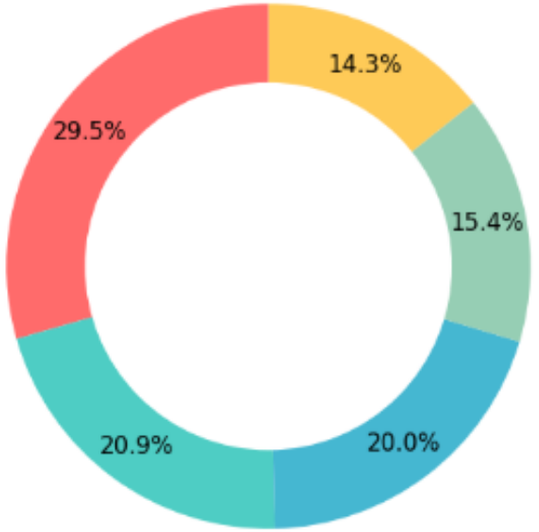
Only 20% of customers engage with the CRM platform, indicating limited direct interaction.

Customer cases fall into two main categories: Complaints and Product Inquiries.

Data Processing & Feature Engineering

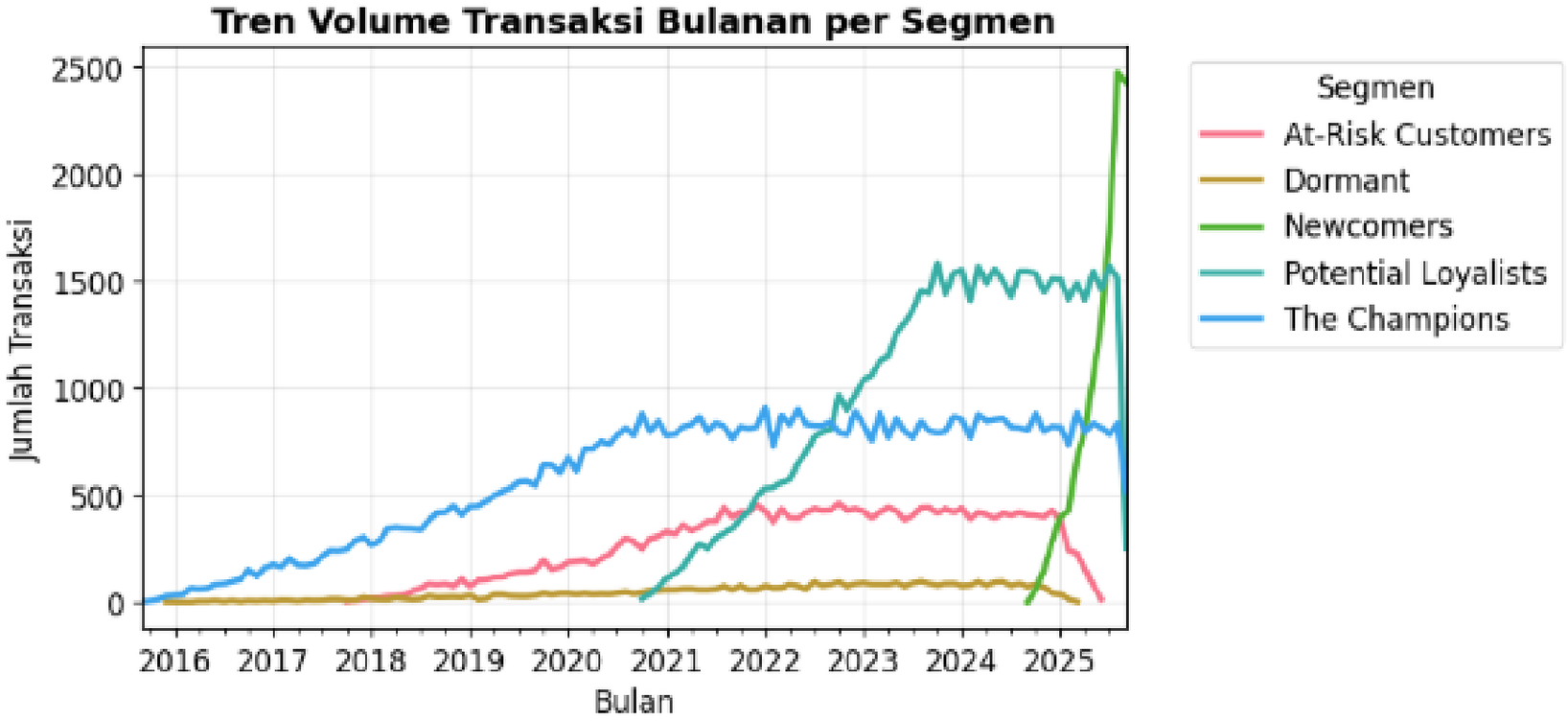
Jumlah Nasabah per Segmen





Potential Loyalists : 1473 (29.5%)
At-Risk Customers : 1046 (20.9%)
Dormant : 999 (20%)
Newcomers : 768 (15.4%)
The Champions : 714 (14.28%)

Data Processing & Feature Engineering



This chart illustrates the monthly transaction frequency trends across different customer segments. It serves as a valuable tool for monitoring and comparing spending behavior among various customer groups over time.

Data Processing & Feature Engineering

At-Risk Customers

 Age: 42.4
 Total AUM: 59.28M
 Tx/Month: 0.36
 Avg Tx Value: 503.8K
 CS Interactions: 0.21
⚠ Complaints: 0.08
 Pending Cases: 0.05

Gender: Pria
City: Binjai
Occupation: Wiraswasta
Channel: ATM

Potential Loyalists

 Age: 36.2
 Total AUM: 124.53M
 Tx/Month: 1.00
 Avg Tx Value: 578.5K
 CS Interactions: 0.20
⚠ Complaints: 0.07
 Pending Cases: 0.08

Gender: Pria
City: Cimahi
Occupation: Karyawan Swasta
Channel: ATM

Dormant

 Age: 44.8
 Total AUM: 2.54M
 Tx/Month: 0.07
 Avg Tx Value: 101.3K
 CS Interactions: 0.19
⚠ Complaints: 0.07
 Pending Cases: 0.05

Gender: Pria
City: Parepare
Occupation: PNS
Channel: ATM

Newcomers

 Age: 24.1
 Total AUM: 12.98M
 Tx/Month: 7.84
 Avg Tx Value: 267.2K
 CS Interactions: 0.21
⚠ Complaints: 0.08
 Pending Cases: 0.06

Gender: Wanita
City: Sungai Penuh
Occupation: Karyawan Swasta
Channel: ATM

The Champions

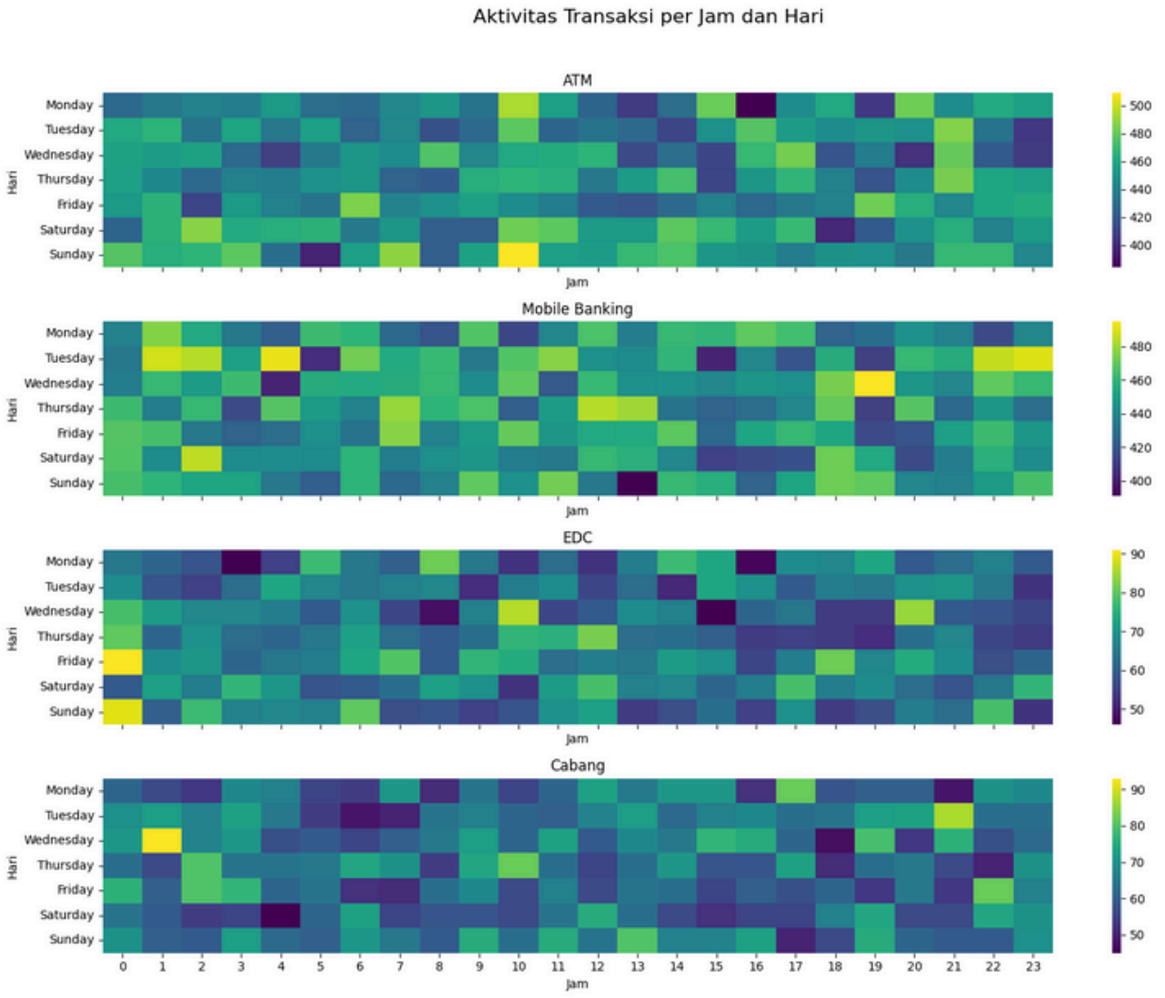
 Age: 47.7
 Total AUM: 629.91M
 Tx/Month: 1.15
 Avg Tx Value: 2809.7K
 CS Interactions: 0.20
⚠ Complaints: 0.07
 Pending Cases: 0.04

Gender: Wanita
City: Bau-Bau
Occupation: Wiraswasta
Channel: ATM

Data Processing & Feature Engineering

- The Champions are the oldest customer segment, with an average age of around 48 years. They hold the highest average AUM (~IDR 630 million) and demonstrate both high transaction frequency and value, making them the most valuable segment.
- Potential Loyalists are younger, averaging 36 years old, with a strong AUM (~IDR 125 million) and fairly high transaction activity they show promising potential to become Champions
- At-Risk Customers, averaging over 40 years old, maintain a decent AUM (~IDR 59 million) but exhibit lower transaction frequency and value compared to Champions and Loyalists, indicating a decline in engagement.
- Newcomers are the youngest group, averaging 24 years old, with a low AUM (~IDR 13 million) but the highest monthly transaction frequency, suggesting they are just beginning their banking journey.
- Dormant customers, with an average age of 45 years, have the lowest AUM (~IDR 2.5 million) and the least transaction activity, making them the most inactive segment.

Segment Characteristics

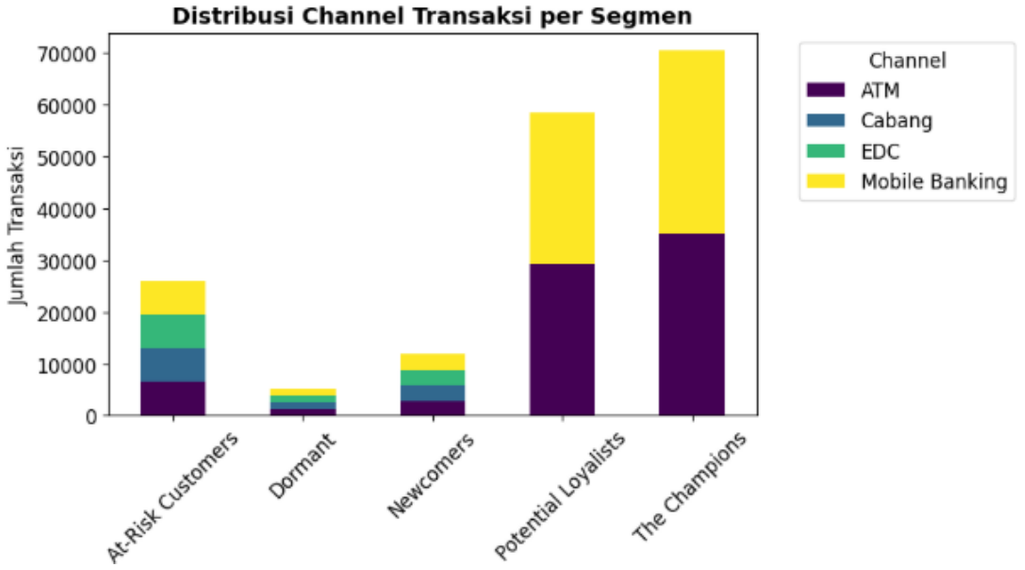


ATM

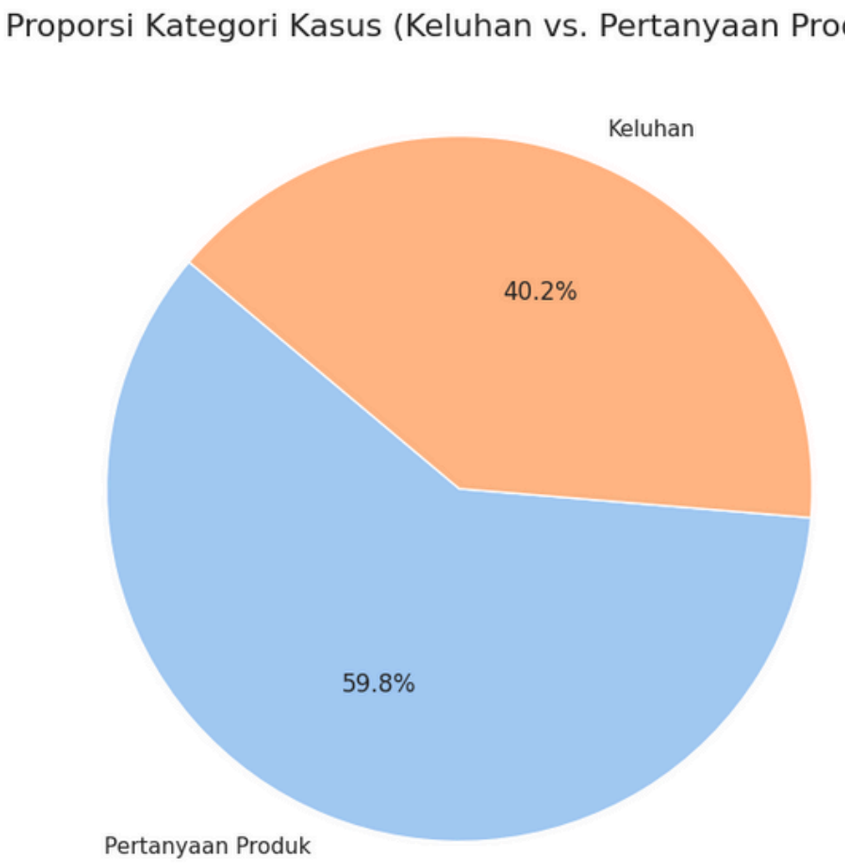
ATMs are busiest during weekday commute hours—both in the morning and evening. Usage drops significantly over the weekend.

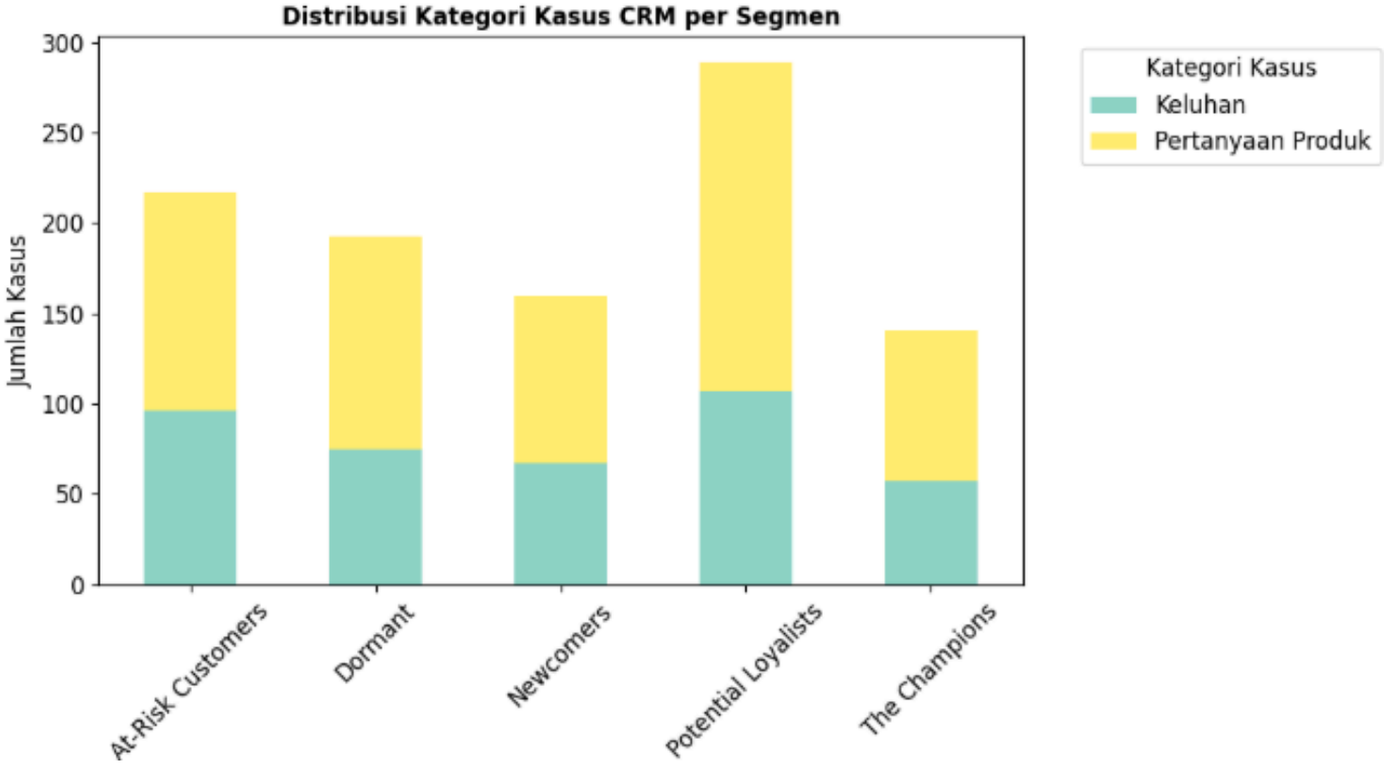
Mobile Banking

Mobile banking activity remains high throughout the week, peaking during evening hours. It is the preferred channel for transactions outside of working hours and on weekends.



Segment Characteristics

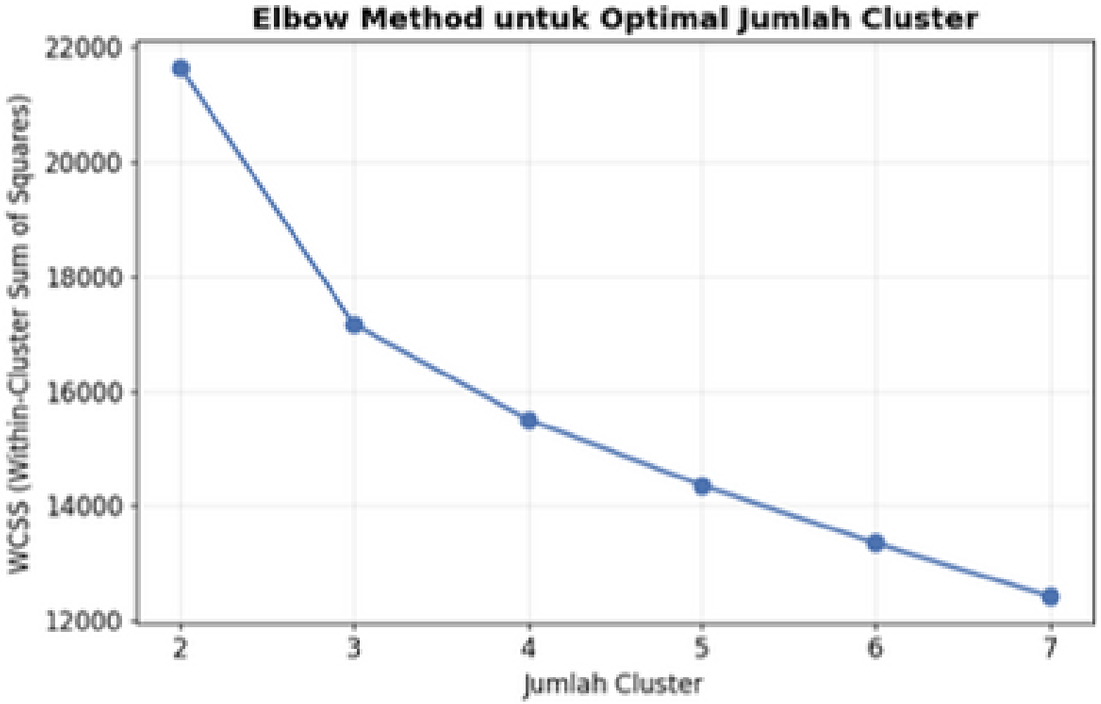




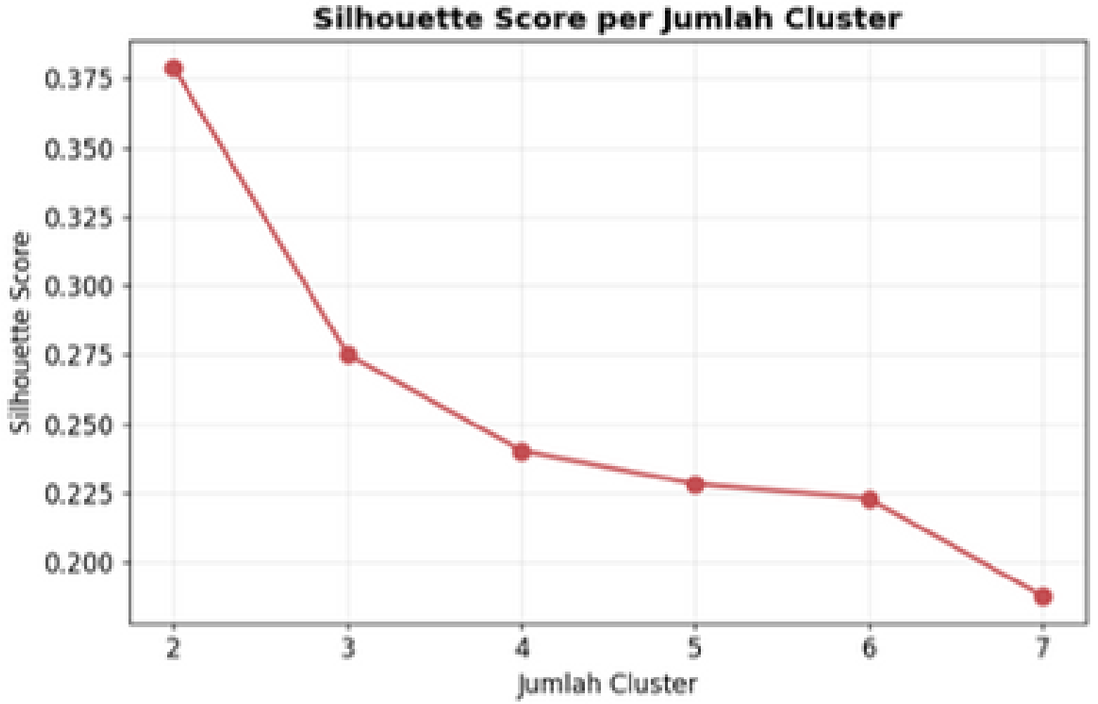
59.8% of cases involve product inquiries, while 40.2% are complaints. This indicates that the majority of customer interactions are not merely problem-driven, but stem from curiosity or a need for additional information. This presents a significant opportunity for education and upselling. Product-related questions can be strategically redirected into sales and engagement opportunities, rather than being viewed solely as a service burden.

This chart highlights the behavioral differences across clusters based on RFM. Cluster 3 demonstrates the strongest profile with frequent and high-value transactions, making it the top priority for retention and loyalty programs. In contrast, Cluster 1 shows high churn risk, while Clusters 0 and 2 still hold potential for cross-sell and upsell opportunities.

CRM Interaction Analysis

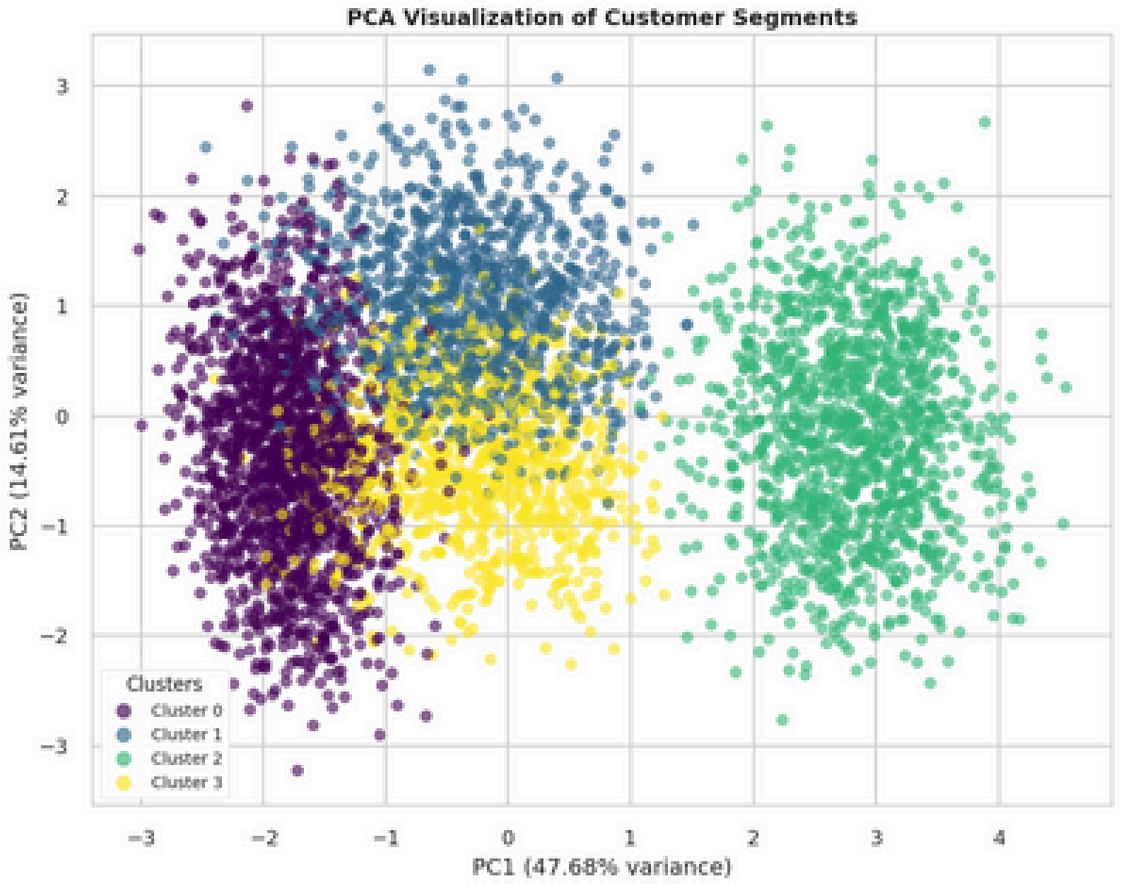


The left chart (Elbow Method) illustrates the relationship between the number of clusters and the WCSS (Within-Cluster Sum of Squares). As the number of clusters increases, the WCSS decreases due to more evenly distributed data. The "elbow point" on the graph indicates the optimal number of clusters, which in this case appears to be around 3 to 4.

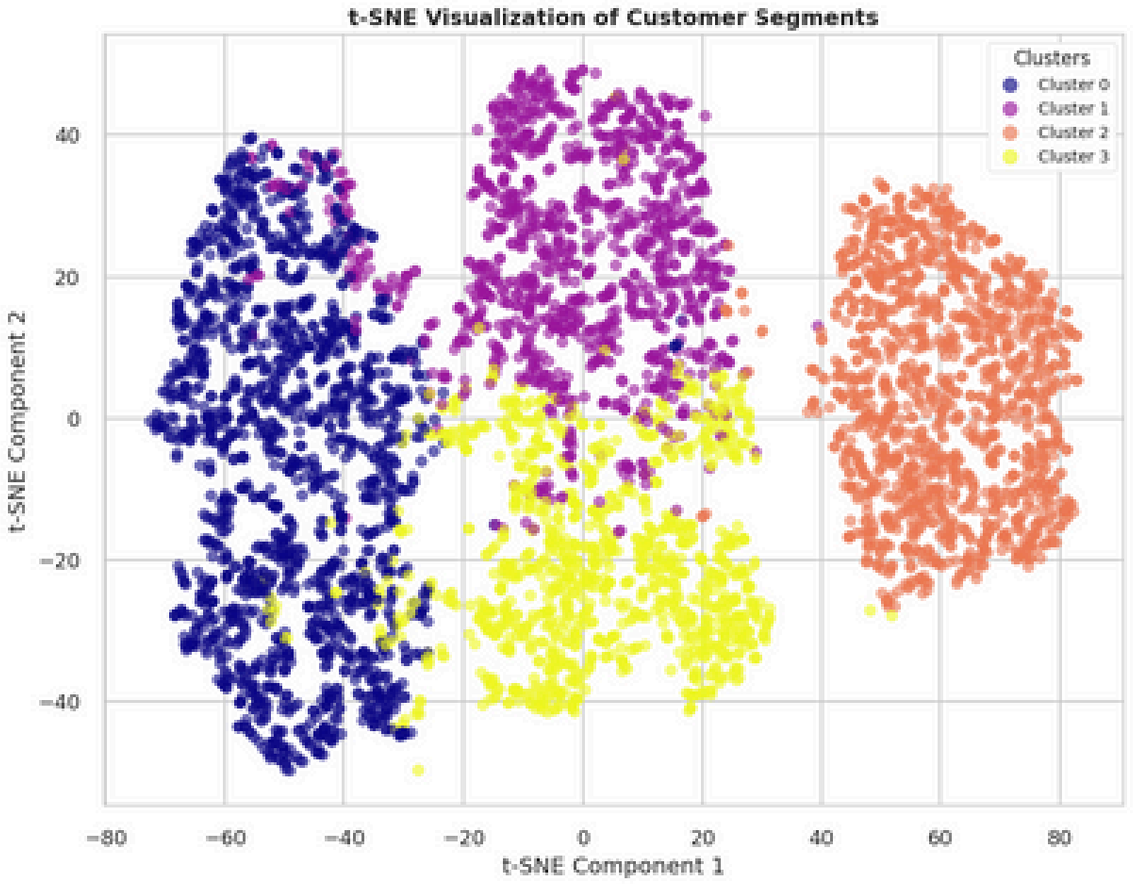


The right chart (Silhouette Score) measures the quality of separation between clusters, with values ranging from -1 to 1. The higher the score, the better the cluster separation. In this chart, the highest score is observed with 2 clusters, although 3 clusters also show reasonably good separation.

CRM Interaction Analysis

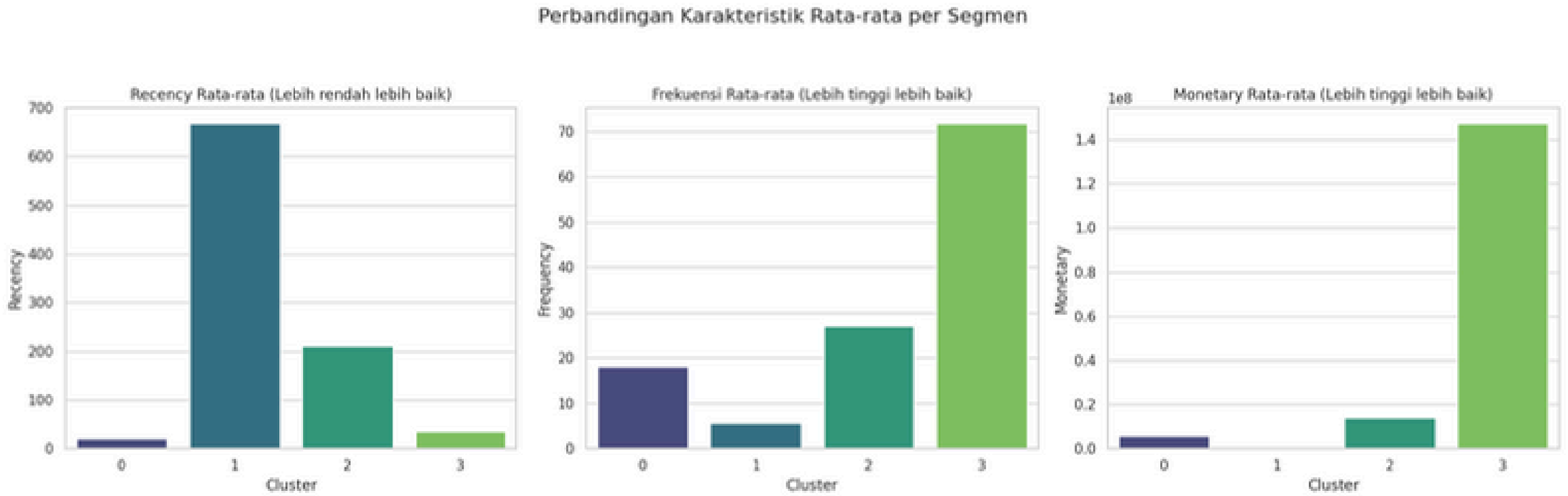


The PCA chart reduces the data into two principal components (PC1 and PC2). Each point represents an individual customer, with colors indicating their respective clusters. The distinct color groupings form separate patterns, suggesting that customers within each cluster share relatively homogeneous characteristics.



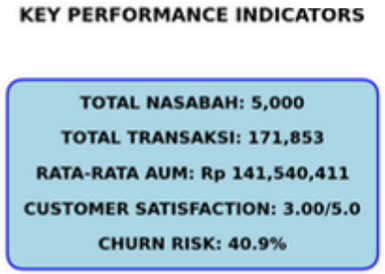
The t-SNE chart projects the data into a 2D space with a focus on local proximity. Different colors represent the formed clusters, with similar data points tending to group together. This visualization provides a clearer and more intuitive separation between clusters compared to PCA.

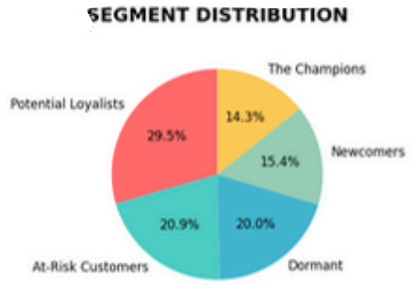
CRM Interaction Analysis

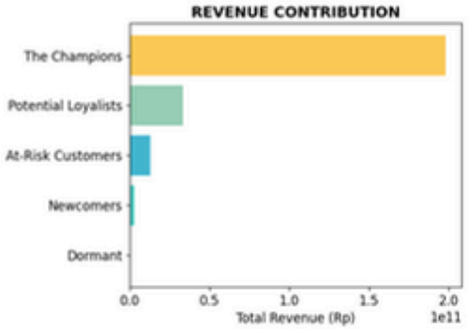


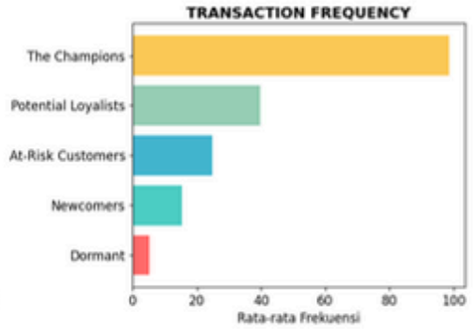
This chart highlights the behavioral differences across clusters based on RFM analysis. Cluster 3 demonstrates the healthiest profile, with frequent, high-value, and active transactions—making it a top priority for retention efforts. In contrast, Cluster 1 shows a high risk of churn, while Clusters 0 and 2 still present opportunities for optimization through cross-selling and upselling strategies.

CRM Interaction Analysis









- The Champions – High Value, High Priority This is the top-performing segment with the highest contribution to the business. The focus is on retention and value maximization through premium services, dedicated relationship managers, and exclusive events. Goal: Maintain satisfaction and loyalty among elite customers.

- Potential Loyalists – Growing Stars Customers with strong potential to graduate into the Champion segment. Key strategies include upselling, financial education, and intensive engagement to boost transaction value and loyalty. Goal: Transform them into high-value customers.

- At-Risk Customers – Critical Segment Customers showing signs of churn and requiring special attention. Implement proactive retention campaigns, win-back incentives, and enhanced personal communication. Goal: Minimize the loss of valuable customers.

- Newcomers – Fresh Entrants New customers currently in the onboarding phase. They need structured onboarding, welcome offers, and product education to feel comfortable quickly. Goal: Build early engagement for long-term loyalty.

- Dormant – Sleeping Customers Passive or inactive customers. Focus on low-cost reactivation campaigns and assess viability before investing further. Goal: Revive potential customers or understand the reasons behind churn.
-

Strategic Recommendations

🎯 Strategic Recommendations

- Champions → Maximize retention and offer premium services
- Potential Loyalists → Development programs and upselling initiatives
- At-Risk Customers → Proactive win-back campaigns
- Newcomers → Structured onboarding and early engagement
- Dormant → Cost-efficient reactivation and churn evaluation

📈 Expected Business Impact

- Marketing effectiveness increases up to 5x through personalization
- Churn rate reduced by 15-25% via targeted retention strategies
- Customer Lifetime Value (CLV) grows by 20-30%
- Marketing budget allocation becomes more efficient, aligned with segment priorities

🔮 Next Steps

- Launch segmented campaigns based on strategic recommendations
- Monitor segment performance regularly
- Re-cluster customers every 6 months to reflect updated data
- Integrate with marketing automation systems for real-time execution