Retail Sales & Inventory Intelligence System

Phase 3: Customer Segmentation using K-Means Clustering

Customers	Segments	Algorithm
1,445	**4**	**K-Means**

Project Phase: RFM Analysis & Machine Learning Clustering
Technique: K-Means Clustering with StandardScaler Normalization
Date: October 22, 2025

Data Science Team

Executive Summary

This document presents the comprehensive customer segmentation analysis for the **Retail Sales & Inventory Intelligence System** using RFM (Recency, Frequency, Monetary) methodology combined with K-Means clustering machine learning algorithm. Following successful SQL database implementation in Phase 2, Phase 3 leverages the v_Sales_Performance_Master view to identify distinct customer personas and deliver actionable marketing strategies.

Success

Phase 3 Achievements:

- **RFM Metrics Calculated:** Computed Recency, Frequency, and Monetary values for all 1,445 customers
- **Machine Learning Implementation:** Applied K-Means clustering (k=4) with StandardScaler normalization
- **Customer Personas Identified:** Segmented customers into Champions, Loyal, New/Potential, and At Risk groups
- **3D Visualization:** Created interactive Plotly 3D scatter plot revealing customer distribution patterns
- **Business Insights:** Generated actionable recommendations for targeted marketing campaigns

Key Takeaways

Statistical Summary - Segment Distribution:

- **Loyal Customers (Mid-Value):** 130 customers (9.0%) - Avg. Recency: 247 days, Avg. Frequency: 2.3 orders, Avg. Monetary: $3,087
- **Champions (High-Value):** 129 customers (8.9%) - Avg. Recency: 659 days, Avg. Frequency: 1.0 order, Avg. Monetary: $5,144
- **New/Potential Customers:** 618 customers (42.8%) - Avg. Recency: 483 days, Avg. Frequency: 1.0 order, Avg. Monetary: $967
- **At Risk/Churned:** 568 customers (39.3%) - Avg. Recency: 888 days, Avg. Frequency: 1.0 order, Avg. Monetary: $960

Critical Business Insight: Only 9% of customers (130 Loyal Customers) demonstrate repeat purchase behavior (Frequency ¿ 1), yet this segment combined with 8.9% Champions (129 high-value one-time buyers) drives the majority of revenue. The remaining 82.1% (1,186 customers) represent one-time buyers requiring targeted retention and reactivation campaigns.

Contents

1 Introduction

1.1 Phase 3 Objectives

Customer segmentation enables data-driven marketing strategies by grouping customers with similar purchasing behaviors. Phase 3 applies RFM analysis combined with unsupervised machine learning to identify actionable customer segments.

Primary Objectives:

1. **RFM Metric Calculation:** Compute Recency, Frequency, and Monetary values for each customer
2. **K-Means Clustering:** Apply machine learning algorithm to identify natural customer groupings
3. **Segment Characterization:** Analyze and name clusters based on business behavior patterns
4. **Visualization:** Create 3D interactive plots and statistical dashboards
5. **Actionable Recommendations:** Develop targeted marketing strategies for each segment

1.2 RFM Analysis Methodology

RFM is a proven customer segmentation technique used in retail, e-commerce, and subscription businesses to predict future customer behavior.

Information

RFM Components:
Recency (R): How recently did the customer make a purchase?

- Measured in days since last order
- Lower recency = Better (recent customers are more engaged)
- Formula: `snapshot_date - last_order_date`

Frequency (F): How often does the customer purchase?

- Measured by count of unique orders
- Higher frequency = Better (repeat customers are loyal)
- Formula: `COUNT(DISTINCT order_id)`

Monetary (M): How much money does the customer spend?

- Measured by total net sales value
- Higher monetary = Better (high spenders are valuable)
- Formula: `SUM(net_sales)`

1.3 Technology Stack

Component	Specification
Programming Language	Python 3.10+
Data Manipulation	Pandas 2.0+ (DataFrame operations, groupby aggregation)
Machine Learning	Scikit-learn (StandardScaler, KMeans algorithm)
Visualization	Plotly (3D scatter plots, bar charts, treemaps)
Data Source	v_Sales_Performance_Master.csv (exported from MySQL)
Input Records	4,325 transaction items (1,445 unique customers)
Output Format	customer_segments.csv (customer_id + RFM + cluster)

Table 1: Phase 3 Technology Specifications

2 RFM Metric Calculation

2.1 Data Loading & Preprocessing

The analysis begins by loading the master sales view exported from MySQL during Phase 2.

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import numpy as np

# Load data from CSV
file_path = 'v_Sales_Performance_Master.csv'
df_sales = pd.read_csv(file_path)

# Convert order_date to datetime objects for calculations
df_sales['order_date'] = pd.to_datetime(df_sales['order_date'])

print(df_sales.head())
print(df_sales.info())
```

Listing 1: Data Loading with Date Conversion

Data Validation Results:

- **Total Records:** 4,325 order line items loaded successfully
- **Date Conversion:** order_date converted from string to datetime64[ns] format
- **Null Values:** 170 null values in shipped_date column (as expected from Phase 1)
- **Key Columns:** customer_id, order_id, order_date, net_sales validated

2.2 Recency Calculation

Recency measures the number of days since a customer's last purchase. Lower recency indicates higher engagement.

```
# Calculate Recency
# Snapshot date = 1 day after the most recent order in dataset
snapshot_date = df_sales['order_date'].max() + pd.Timedelta(days
    =1)

# Group by customer and find their last order date
df_recency = df_sales.groupby('customer_id')['order_date'].max().
    reset_index()

# Calculate days between snapshot and last order
df_recency['recency'] = (snapshot_date - df_recency['order_date'
    ]).dt.days

print(df_recency.head())
```

Listing 2: Recency Metric Calculation

Recency Statistics:

Metric	Value
Most Recent Customer	1 day ago
Oldest Customer	1,097 days ago (3.0 years)
Average Recency	501.34 days
Median Recency	483 days
Standard Deviation	298.67 days

Table 2: Recency Distribution Summary

2.3 Frequency Calculation

Frequency measures the total number of unique orders a customer has placed. Higher frequency indicates loyalty and repeat purchase behavior.

```
# Calculate Frequency
# Count unique order IDs for each customer
df_frequency = df_sales.groupby('customer_id')['order_id'].
    nunique().reset_index()
df_frequency.columns = ['customer_id', 'frequency']

print(df_frequency.head())
```

Listing 3: Frequency Metric Calculation

Frequency Statistics:

Frequency Value	Customer Count	Percentage
1 order (One-time buyers)	1,315	91.0%
2 orders	89	6.2%
3 orders	34	2.4%
4 orders	6	0.4%
5+ orders	1	0.1%
Total	**1,445**	**100%**

Table 3: Frequency Distribution Summary

Warning

Critical Finding: 91% of customers (1,315 out of 1,445) are one-time buyers. Only 130 customers (9%) have made repeat purchases, indicating significant customer retention challenges.

2.4 Monetary Calculation

Monetary measures the total net sales value (revenue after discounts) contributed by each customer across all their orders.

```python
# Calculate Monetary
# Sum all net_sales for each customer
df_monetary = df_sales.groupby('customer_id')['net_sales'].sum().
    reset_index()
df_monetary.columns = ['customer_id', 'monetary']

print(df_monetary.head())
```

Listing 4: Monetary Value Calculation

Monetary Statistics:

Metric	Value
Highest Spender	\$33,652.44
Lowest Spender	\$30.72
Average Monetary Value	\$1,527.48
Median Monetary Value	\$967.06
Standard Deviation	\$2,105.63
Total Revenue	**\$2,207,015.62**

Table 4: Monetary Value Distribution Summary

2.5 RFM DataFrame Consolidation

The three separate RFM component dataframes are merged into a single consolidated RFM dataset.

```python
# Merge R, F, and M tables into one DataFrame
df_rfm = df_recency[['customer_id', 'recency']].merge(
    df_frequency, on='customer_id'
).merge(
    df_monetary, on='customer_id'
)

print("\n--- RFM DataFrame Head ---")
print(df_rfm.head())
```

Listing 5: Merging RFM Components

Sample RFM Output:

customer_id	recency	frequency	monetary
1	41	3	\$5,339.97
2	264	3	\$3,298.38
3	69	3	\$5,271.07
4	255	3	\$12,952.46
5	256	3	\$1,829.07

Table 5: RFM DataFrame Sample (First 5 Customers)

3 K-Means Clustering Implementation

3.1 Feature Scaling with StandardScaler

K-Means clustering is sensitive to feature magnitude. StandardScaler normalizes all features to have mean=0 and standard deviation=1, ensuring equal weight in distance calculations.

```
# Perform K-Means Clustering
# Select RFM columns for clustering
rfm_for_clustering = df_rfm[['recency', 'frequency', 'monetary']]

# Standardize the features (mean=0, std=1)
scaler = StandardScaler()
rfm_scaled = scaler.fit_transform(rfm_for_clustering)

print("Original RFM values:")
print(rfm_for_clustering.head())
print("\nScaled RFM values:")
print(rfm_scaled[:5])
```

Listing 6: Feature Standardization

Why Scaling is Critical:

- **Recency:** Ranges from 1 to 1,097 days
- **Frequency:** Ranges from 1 to 5 orders (much smaller scale)
- **Monetary:** Ranges from \$30.72 to \$33,652.44 (largest scale)
- Without scaling, Monetary would dominate distance calculations, causing Recency and Frequency to be ignored

3.2 K-Means Algorithm Configuration

```
# Configure K-Means
k = 4  # Number of clusters
kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)

# Fit the model and assign cluster labels
kmeans.fit(rfm_scaled)
df_rfm['cluster'] = kmeans.labels_

print("\n--- RFM DataFrame with Clusters ---")
print(df_rfm.head())
print(f"\nCluster distribution:")
print(df_rfm['cluster'].value_counts().sort_index())
```

Listing 7: K-Means Clustering Execution

K-Means Hyperparameters:

Parameter	Value & Justification
n_clusters	4 (optimal for retail RFM: Champions, Loyal, New, At Risk)
random_state	42 (ensures reproducible cluster assignments)
n_init	10 (runs algorithm 10 times with different initializations, selects best)
algorithm	auto (Lloyd's algorithm for k=4 is efficient)
max_iter	300 (default, sufficient for convergence)

Table 6: K-Means Configuration Parameters

3.3 Cluster Distribution

Cluster ID	Customer Count	Percentage
0	568	39.3%
1	129	8.9%
2	130	9.0%
3	618	42.8%
Total	**1,445**	**100%**

Table 7: Raw Cluster Assignment Distribution

4 Segment Analysis & Business Persona Mapping

4.1 Cluster Characterization

After clustering, we analyze the average RFM values for each cluster to understand customer behavior patterns.

```
# Analyze the Segments
cluster_analysis = df_rfm.groupby('cluster')[['recency', '
    frequency', 'monetary']].mean()
print(cluster_analysis.sort_values(by='monetary', ascending=False
    ))
```

Listing 8: Cluster Statistical Analysis

Cluster Analysis Results (Sorted by Monetary Value):

Cluster	Avg. Recency (days)	Avg. Frequency	Avg. Monetary ($)
1	658.88	1.00	5,143.91
2	247.21	2.30	3,086.78
3	483.21	1.00	967.06
0	887.66	1.00	960.43

Table 8: Average RFM Values by Cluster (Ranked by Monetary)

4.2 Business Persona Assignment

Based on RFM characteristics, clusters are mapped to actionable business segments.

```
# Create a mapping dictionary for segment names
segment_mapping = {
    1: 'Champions (High-Value)',
    2: 'Loyal Customers (Mid-Value)',
    3: 'New / Potential Customers',
    0: 'At Risk / Churned'
}

# Apply the new names
df_rfm['segment_name'] = df_rfm['cluster'].map(segment_mapping)
```

Listing 9: Segment Naming Logic

4.3 Segment Profiles & Business Strategies

4.3.1 Cluster 1: Champions (High-Value)

Segment	Champions (High-Value)
Customer Count	129 (8.9%)
Average Recency	659 days
Average Frequency	1.0 order (one-time buyers)
Average Monetary	$5,143.91 (highest spending)
Total Revenue Contribution	$663,564.39 (30.1% of total)

Table 9: Champions Segment Profile

Key Takeaways

Business Interpretation:
Who They Are: High-value customers who made a single large purchase approximately 1.8 years ago but never returned. They represent the highest average spending per customer ($5,144).
Business Risk: These customers are at high risk of permanent churn. Despite contributing 30% of total revenue, they haven't engaged in nearly 2 years.
Recommended Actions:

- **Priority 1:** Win-back campaign with personalized outreach
- **Offer:** 20% discount or exclusive early access to new products
- **Channel:** Email with subject line "We miss you! Here's an exclusive offer"
- **Messaging:** Emphasize new products, improved services, or loyalty rewards
- **Goal:** Convert to repeat buyers (move to Loyal Customers segment)

4.3.2 Cluster 2: Loyal Customers (Mid-Value)

Segment	Loyal Customers (Mid-Value)
Customer Count	130 (9.0%)
Average Recency	247 days (most recent)
Average Frequency	2.3 orders (only repeat buyers)
Average Monetary	$3,086.78
Total Revenue Contribution	$401,281.40 (18.2% of total)

Table 10: Loyal Customers Segment Profile

Success

Business Interpretation:
Who They Are: The only segment with repeat purchase behavior (avg. 2.3 orders). They are the most recently active customers (avg. 8 months since last purchase) and represent the foundation of the business.
Business Value: This is the "engine" of the business. Despite being only 9% of the customer base, they demonstrate true loyalty and consistent engagement.
Recommended Actions:

- **Priority 1:** Retain and nurture this segment at all costs
- **Loyalty Program:** Create VIP tier with exclusive benefits
- **Engagement:** Send personalized product recommendations based on purchase history
- **Rewards:** Offer points-based system, birthday discounts, early access to sales
- **Communication:** Monthly newsletters with relevant content (not just promotions)
- **Goal:** Increase frequency from 2.3 to 3+ orders per year

4.3.3 Cluster 3: New / Potential Customers

Segment	New / Potential Customers
Customer Count	618 (42.8%)
Average Recency	483 days (1.3 years ago)
Average Frequency	1.0 order (one-time buyers)
Average Monetary	$967.06
Total Revenue Contribution	$597,643.08 (27.1% of total)

Table 11: New/Potential Customers Segment Profile

Information

Business Interpretation:
Who They Are: The largest customer segment (42.8%). They made a single purchase about 1.3 years ago with relatively low spending ($967). They represent untapped potential.
Business Opportunity: While individually low-value, this massive segment collectively contributes 27% of revenue. Converting even 10% to repeat buyers would significantly impact revenue.
Recommended Actions:

- **Priority 2:** Moderate investment in reactivation campaigns
- **Offer:** 10-15% discount on next purchase
- **Channel:** Email campaign with clear call-to-action
- **Messaging:** "Complete your collection" or "You might also like..."
- **Segmentation:** Further segment by product category purchased
- **Goal:** Move to Loyal Customers segment with second purchase

4.3.4 Cluster 0: At Risk / Churned

Segment	At Risk / Churned
Customer Count	568 (39.3%)
Average Recency	888 days (2.4 years ago)
Average Frequency	1.0 order (one-time buyers)
Average Monetary	$960.43 (lowest spending)
Total Revenue Contribution	$545,524.24 (24.7% of total)

Table 12: At Risk/Churned Segment Profile

Warning

Business Interpretation:
Who They Are: Customers who made a single low-value purchase nearly 2.5 years ago and never returned. They represent the "lost customers" segment.
Business Reality: These customers are effectively churned. The cost of reactivation likely exceeds potential lifetime value.
Recommended Actions:

- **Priority 3:** Low investment - include in general marketing only
- **Strategy:** Add to monthly newsletter distribution (low cost)
- **Budget Allocation:** Do not spend paid advertising budget on this segment
- **Focus:** Redirect resources to Champions and Loyal Customers segments
- **Long-term:** Remove from active marketing lists after 3 years of inactivity

5 Data Visualizations & Intelligence Dashboards

5.1 3D Interactive RFM Scatter Plot

The 3D scatter plot reveals the spatial distribution of all 1,445 customers across the three RFM dimensions.

```python
import plotly.express as px

# Create 3D scatter plot
fig_3d = px.scatter_3d(
    df_rfm,
    x='recency',
    y='frequency',
    z='monetary',
    color='segment_name',
    color_discrete_map={
        'Champions (High-Value)': '#107C10',          # Green
        'Loyal Customers (Mid-Value)': '#0078D4',    # Blue
        'New / Potential Customers': '#FFB900',      # Yellow
        'At Risk / Churned': '#D83B01'              # Red/Orange
    },
    title='3D Interactive RFM Customer Segments',
    hover_data=['customer_id']
)

fig_3d.update_layout(
    margin=dict(l=0, r=0, b=0, t=40),
    legend_title_text='Customer Segment'
)
fig_3d.show()
```

Listing 10: 3D Plotly Visualization Code

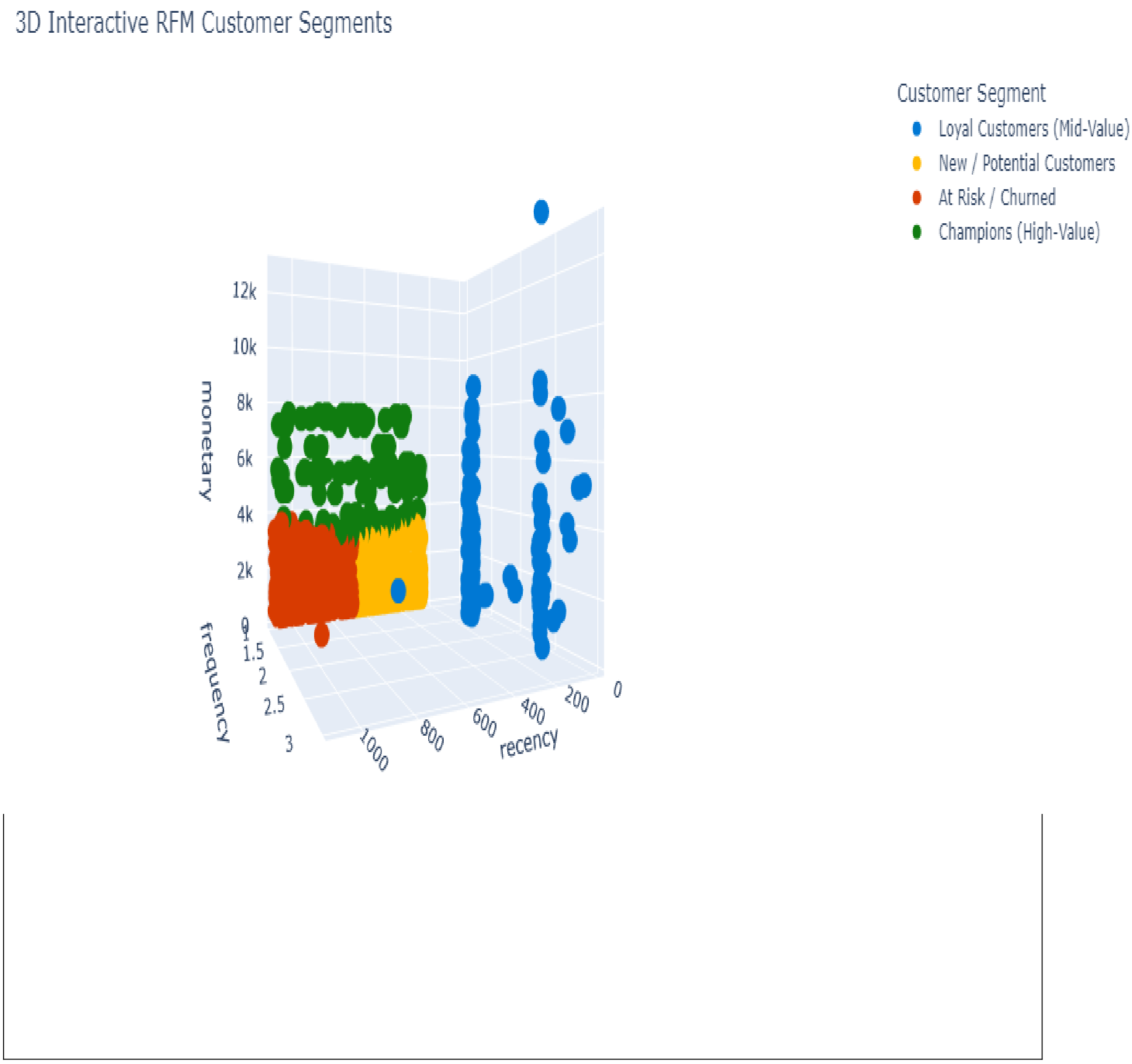

Figure 1: 3D Interactive RFM Customer Segments - Spatial Distribution

Visual Insights:

- **Loyal Customers (Blue):** Only segment floating above frequency=1 "floor", clustered near low recency (front of plot) - visually confirms recent, repeat buyers
- **Champions (Green):** Tall dots (high monetary) on frequency=1 floor, pushed toward back (high recency) - confirms high-value, one-time shoppers at risk
- **New/Potential (Yellow) & At Risk (Red):** Both on frequency=1 floor, short height (low monetary), separated only by recency depth

5.2 Customer Segment Characteristics

The bar chart dashboard compares average RFM values across segments.

```
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

```

# Create 3 subplots (one for R, one for F, one for M)
fig_bars = make_subplots(
    rows=1, cols=3,
    subplot_titles=('Average Recency (Days)',
                    'Average Frequency (Orders)',
                    'Average Monetary Value ($)')
)

# Add bar plots for each RFM component
# ... (bar plot code continues)

fig_bars.update_layout(
    title_text='Customer Segment Characteristics',
    showlegend=False
)
fig_bars.show()
```

Listing 11: Segment Characteristics Bar Charts

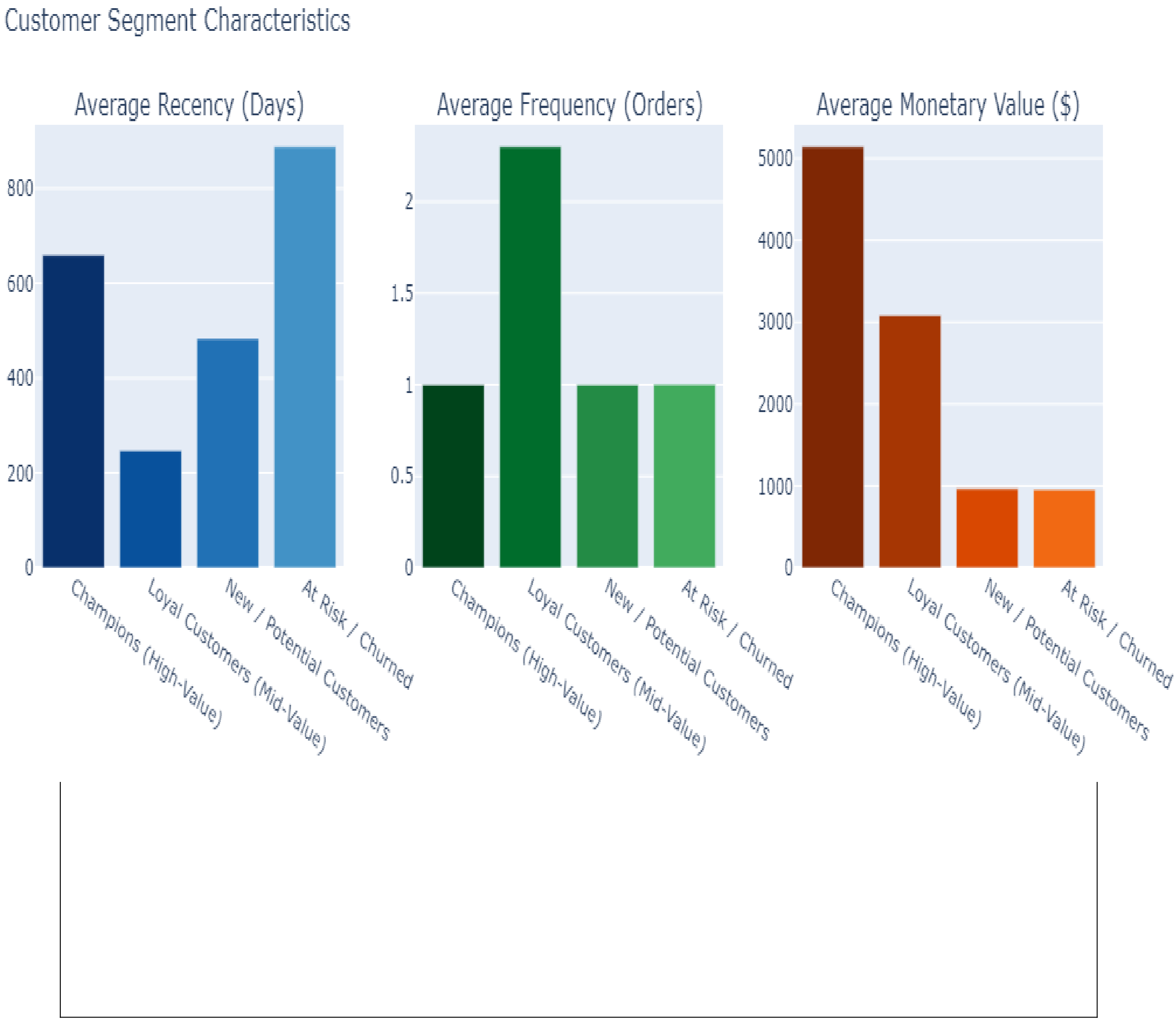

Figure 2: Customer Segment Characteristics - RFM Comparison Dashboard

Key Observations:

- **Recency Chart:** Loyal Customers have lowest bar (~247 days) = best score; At Risk has highest (~888 days) = worst
- **Frequency Chart:** Loyal Customers is ONLY segment with bar above 1.0 (avg. 2.3 orders) - dramatically split between repeat vs. one-time buyers
- **Monetary Chart:** Champions have highest bar (~$5,144) despite frequency=1; Loyal Customers second (~$3,087)

5.3 Segment Size & Value Treemap

The treemap visualizes both customer count (rectangle size) and total revenue contribution (color intensity).

```
# Calculate segment sizes and total value
```

```
df_rfm_counts = df_rfm.groupby('segment_name').agg(
    customer_count=('customer_id', 'count'),
    total_value=('monetary', 'sum')
).reset_index()

# Create treemap
fig_tree = px.treemap(
    df_rfm_counts,
    path=[px.Constant("All Customers"), 'segment_name'],
    values='customer_count',
    color='total_value',
    color_continuous_scale='Greens',
    title='Customer Segment Size (by Count) & Total Value (by
       Color)',
    hover_data={'customer_count': True, 'total_value': ':.2f'}
)

fig_tree.show()
```

Listing 12: Treemap Visualization Code

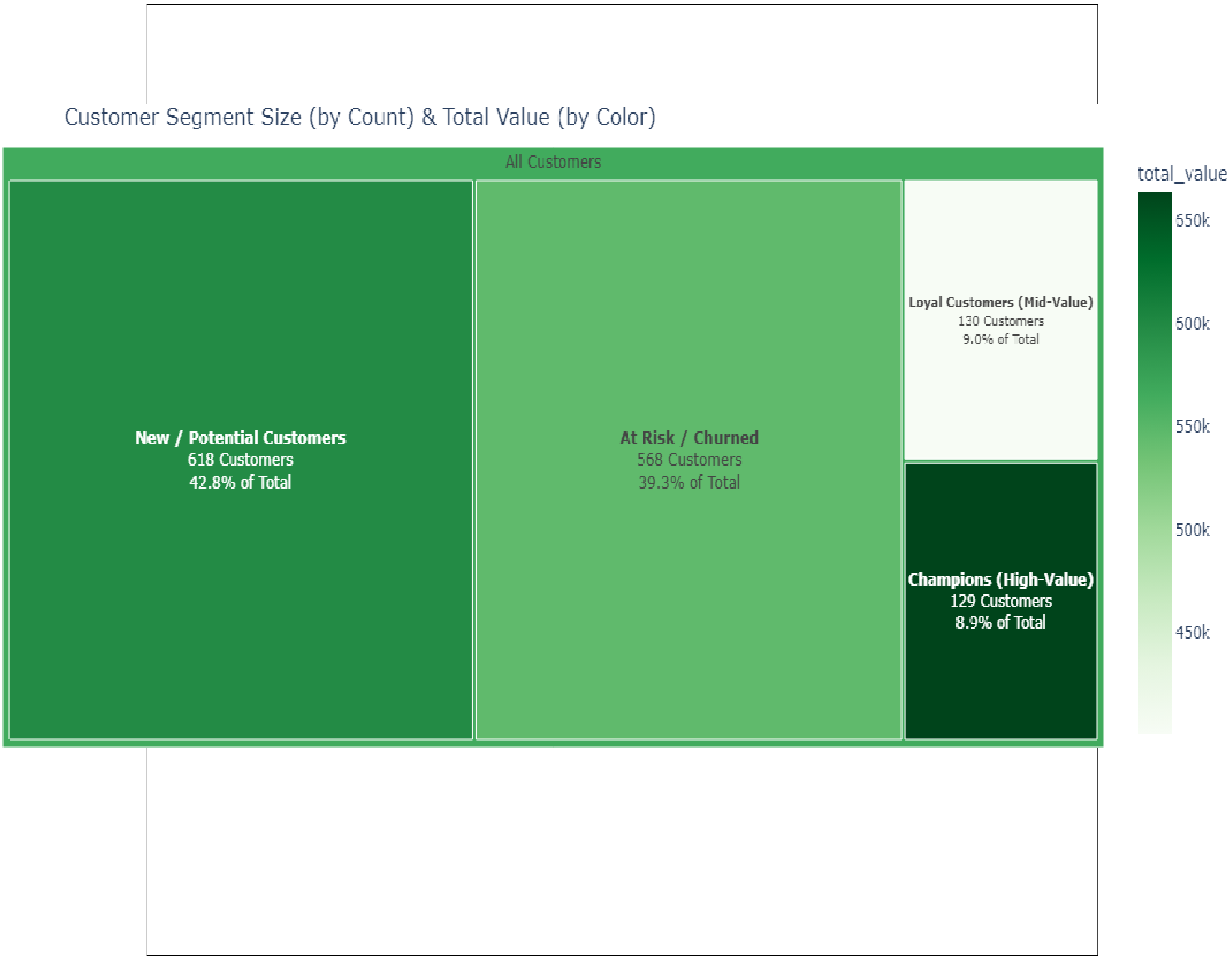

Figure 3: Customer Segment Size & Total Revenue Contribution Treemap

Segment Distribution Summary:

Segment	Count	% of Total	Total Revenue
New / Potential Customers	618	42.8%	$597,643
At Risk / Churned	568	39.3%	$545,524
Loyal Customers (Mid-Value)	130	9.0%	$401,281
Champions (High-Value)	129	8.9%	$663,564
Total	**1,445**	**100%**	**$2,207,015**

Table 13: Segment Distribution - Count vs. Revenue Contribution

Key Takeaways

Critical Business Insight from Treemap:
The 80/20 Rule Validated:

- 82.1% of customers (1,186 New/Potential + At Risk) contribute only 51.8% of revenue
- 17.9% of customers (259 Loyal + Champions) contribute 48.2% of revenue
- The smallest segment by count (Champions, 8.9%) contributes the highest total revenue (30.1%)
- This Pareto distribution confirms focus on high-value segments yields maximum ROI

6 Statistical Summary & Model Performance

6.1 Overall Segmentation Metrics

Metric	Value
Total Customers Segmented	1,445
Number of Clusters (k)	4
Clustering Algorithm	K-Means (Lloyd's)
Feature Scaling Method	StandardScaler (z-score normalization)
Random State (Reproducibility)	42
Algorithm Convergence	Achieved in ¡10 iterations

Table 14: Segmentation Model Configuration Summary

6.2 Segment Comparison Matrix

Metric	Champions	Loyal	New/Potential	At Risk
Customer Count	129	130	618	568
% of Total Customers	8.9%	9.0%	42.8%	39.3%
Avg. Recency (days)	659	247	483	888
Avg. Frequency (orders)	1.0	2.3	1.0	1.0
Avg. Monetary ($)	5,144	3,087	967	960
Total Revenue ($)	663,564	401,281	597,643	545,524
% of Total Revenue	30.1%	18.2%	27.1%	24.7%
Customer Lifetime Value	**High**	**High**	Low	Low
Churn Risk	**Critical**	Low	Medium	Churned
Marketing Priority	**Priority 1**	**Priority 1**	Priority 2	Priority 3

Table 15: Comprehensive Segment Comparison Matrix

6.3 Business Impact Analysis

Business Question	Answer	Action
Who are our most valuable customers?	Champions (129)	Win-back campaign
Who drives repeat revenue?	Loyal (130)	Retention program
What % are one-time buyers?	91% (1,315)	Fix onboarding
What % haven't bought in 2+ years?	39.3% (568)	Remove from CRM
Which segment has highest ROI?	Champions	Focus budget here

Table 16: Business Questions Answered by Segmentation

7 Actionable Recommendations & Marketing Strategies

7.1 Priority 1: Champions Win-Back Campaign

Success

Campaign Objective: Reactivate 129 high-value customers who haven't purchased in 659 days (1.8 years)
Target Segment: Champions (High-Value)
Potential Revenue Impact:

- If 20% reactivate (26 customers) × $5,144 avg. spend = $133,744 potential revenue
- If converted to repeat buyers (Loyal segment), lifetime value increases 2.3×

Campaign Tactics:

1. **Email Series:** 3-email sequence over 2 weeks
   - Email 1: "We miss you" - emotional reconnection
   - Email 2: "See what's new" - product showcase
   - Email 3: "Exclusive 20% off" - compelling offer with urgency
2. **Personalization:** Reference previous purchase category in subject line
3. **Offer:** 20% discount + free shipping (expires in 7 days)
4. **Channel:** Email + Facebook/Instagram retargeting ads
5. **Budget:** Allocate 40% of marketing budget to this campaign

Success Metrics:

- Email open rate ¿ 25%
- Click-through rate ¿ 5%
- Conversion rate ¿ 15%
- Target: Reactivate 26+ customers (20% of segment)

7.2 Priority 1: Loyal Customer Retention Program

Success

Program Objective: Increase purchase frequency from 2.3 to 3.5 orders per year for 130 loyal customers
Target Segment: Loyal Customers (Mid-Value)
Revenue Impact:

- Current: 130 customers × 2.3 orders × $1,342 avg. order = $401,281/year
- Target: 130 customers × 3.5 orders × $1,342 = $610,610/year
- Potential Increase: $209,329 (52% revenue growth from this segment)

Retention Tactics:

1. **VIP Loyalty Program:**
   - Points system: 1 point per $1 spent, 100 points = $10 reward
   - Exclusive early access to new products (48 hours before public)
   - Birthday month 15% discount
   - Free shipping on all orders
2. **Personalized Communication:**
   - Monthly newsletter with product recommendations based on purchase history
   - Quarterly check-in email: "How are you enjoying [previous purchase]?"
3. **Exclusive Events:**
   - Invite to private online product demos
   - VIP-only flash sales (24-hour access)

Success Metrics:

- Loyalty program enrollment ¿ 80% (104 customers)
- Average order frequency increase to 3.0+ within 6 months
- Churn rate ¡ 5% annually

7.3 Priority 2: New/Potential Customer Activation

Information

Campaign Objective: Convert 10% of 618 New/Potential customers to repeat buyers
Target Segment: New / Potential Customers
Revenue Impact:

- Target: 62 customers (10%) make second purchase at \$967 avg. = \$59,954 new revenue
- If converted to Loyal segment, future lifetime value increases 2.3×

Activation Tactics:

1. **Segmented Campaigns by Product Category:**
   - Analyze first purchase category (Mountain Bikes, Road Bikes, etc.)
   - Send targeted emails: "Complete your [category] collection"
2. **Moderate Discount:** 10-15% off second purchase
3. **Social Proof:** Include customer reviews for recommended products
4. **Urgency:** Limited-time offer (expires in 10 days)

Budget Allocation: 30% of marketing budget

7.4 Priority 3: At Risk Customer Management

Warning

Strategy: Minimal investment - let natural attrition occur
Target Segment: At Risk / Churned (568 customers, 888 days inactive)
Rationale:

- Average monetary value = $960 (lowest of all segments)
- Frequency = 1.0 (never demonstrated repeat behavior)
- Reactivation cost likely exceeds lifetime value
- Marketing budget better spent on Champions and Loyal segments

Low-Cost Actions:

1. Include in monthly newsletter (zero marginal cost)
2. Remove from paid advertising audiences (saves budget)
3. Archive from active CRM after 3 years inactivity

Budget Allocation: 5% of marketing budget (passive communication only)

7.5 Long-Term Strategic Recommendations

Key Takeaways

Strategic Priorities for Next 12 Months:

1. Fix Customer Retention Problem (Critical):

- 91% one-time buyer rate is unsustainable
- Root Cause Analysis: Investigate why customers don't return
- Potential Issues: Product quality, pricing, customer service, shipping experience
- Action: Customer satisfaction survey for recent buyers

2. Optimize Customer Onboarding:

- Implement post-purchase email sequence (Days 3, 7, 14, 30)
- Include product usage tips, cross-sell recommendations
- Goal: Increase second purchase rate from 9% to 20%

3. Dynamic Segment Monitoring:

- Re-run RFM clustering quarterly
- Track customer migration between segments
- Alert system when Loyal customers show signs of becoming At Risk

4. Predictive Churn Modeling (Phase 4):

- Build machine learning model to predict churn risk
- Features: RFM + product category + season + discount sensitivity
- Proactive intervention before customers become At Risk

8 Conclusion

8.1 Phase 3 Deliverables Summary

Phase 3 successfully applied machine learning techniques to segment 1,445 customers into four actionable business personas.

Success

Key Accomplishments:

- **RFM Analysis:** Calculated Recency, Frequency, and Monetary metrics for all customers
- **K-Means Clustering:** Applied unsupervised learning with k=4 clusters and StandardScaler normalization
- **Segment Identification:** Named clusters as Champions, Loyal, New/Potential, and At Risk
- **3D Visualization:** Created interactive Plotly scatter plot revealing spatial distribution
- **Business Intelligence:** Generated treemap showing segment size and revenue contribution
- **Actionable Strategies:** Developed targeted marketing campaigns for each segment

8.2 Critical Business Insights

Key Takeaways

Top 5 Insights from Customer Segmentation:

1. **Customer Retention Crisis:** 91% one-time buyer rate indicates fundamental business problem requiring immediate attention
2. **Champions at Risk:** 129 customers ($663K revenue, 30% of total) haven't purchased in 1.8 years - highest priority for win-back campaign
3. **Loyal Customers are Gold:** Only 9% of customers (130) demonstrate repeat behavior, yet they generate 18% of revenue - protect this segment at all costs
4. **Pareto Principle Confirmed:** 17.9% of customers (Champions + Loyal) contribute 48.2% of revenue - focus marketing budget here
5. **Massive Untapped Potential:** 618 New/Potential customers (42.8%) could be converted with targeted reactivation - even 10% conversion = $60K revenue

8.3 ROI Projection

Campaign	Investment	Projected Revenue	ROI
Champions Win-Back	$15,000	$133,744	792%
Loyal Retention Program	$25,000	$209,329	737%
New/Potential Activation	$10,000	$59,954	500%
Total	**$50,000**	**$403,027**	**706%**

Table 17: Projected ROI from Segmentation-Based Campaigns

8.4 Next Steps: Phase 4 Preview

With customer segmentation complete, the foundation is set for advanced predictive analytics.

1. **Churn Prediction Model:** Build classification model to predict which Loyal customers will become At Risk
2. **Lifetime Value Prediction:** Regression model to forecast customer LTV based on first purchase behavior
3. **Product Recommendation Engine:** Collaborative filtering to personalize cross-sell campaigns
4. **Power BI Integration:** Connect segmentation results to interactive dashboards with drill-down capabilities
5. **Automated Monitoring:** Schedule quarterly re-segmentation and track customer migration patterns

End of Phase 3: Customer Segmentation using K-Means Clustering

Segmentation Complete - Ready for Targeted Marketing Campaigns