

# Customer Segmentation Analysis Report

## Executive Summary

This report presents the results of a customer segmentation analysis performed on an e-commerce platform's customer behavior data. The analysis successfully identified three distinct customer segments as anticipated in the project scope:

- **Bargain Hunters:** Deal-seekers who make frequent purchases of low-value items and heavily rely on discounts.
- **High Spenders:** Premium buyers who focus on high-value purchases and are less influenced by discounts.
- **Window Shoppers:** Customers who spend significant time browsing but rarely make purchases.

The segmentation was achieved through a systematic approach involving data preprocessing, exploratory data analysis, multiple clustering algorithms, and thorough evaluation of results. The findings provide actionable insights for targeted marketing strategies, revenue optimization, and improved customer experiences.

## 1. Data Understanding and Preprocessing

The analysis began with a dataset containing 999 customer records with the following features:

- total\_purchases: Total number of purchases made by the customer
- avg\_cart\_value: Average value of items in the customer's cart
- total\_time\_spent: Total time spent on the platform (in minutes)
- product\_click: Number of products viewed by the customer
- discount\_counts: Number of times the customer used a discount code
- customer\_id: Unique identifier for each customer

## 1.1 Data Quality Assessment

Initial examination revealed 20 rows with missing values (approximately 2% of the dataset). These were handled by dropping rows with multiple missing values, as they represented a small portion of the data.

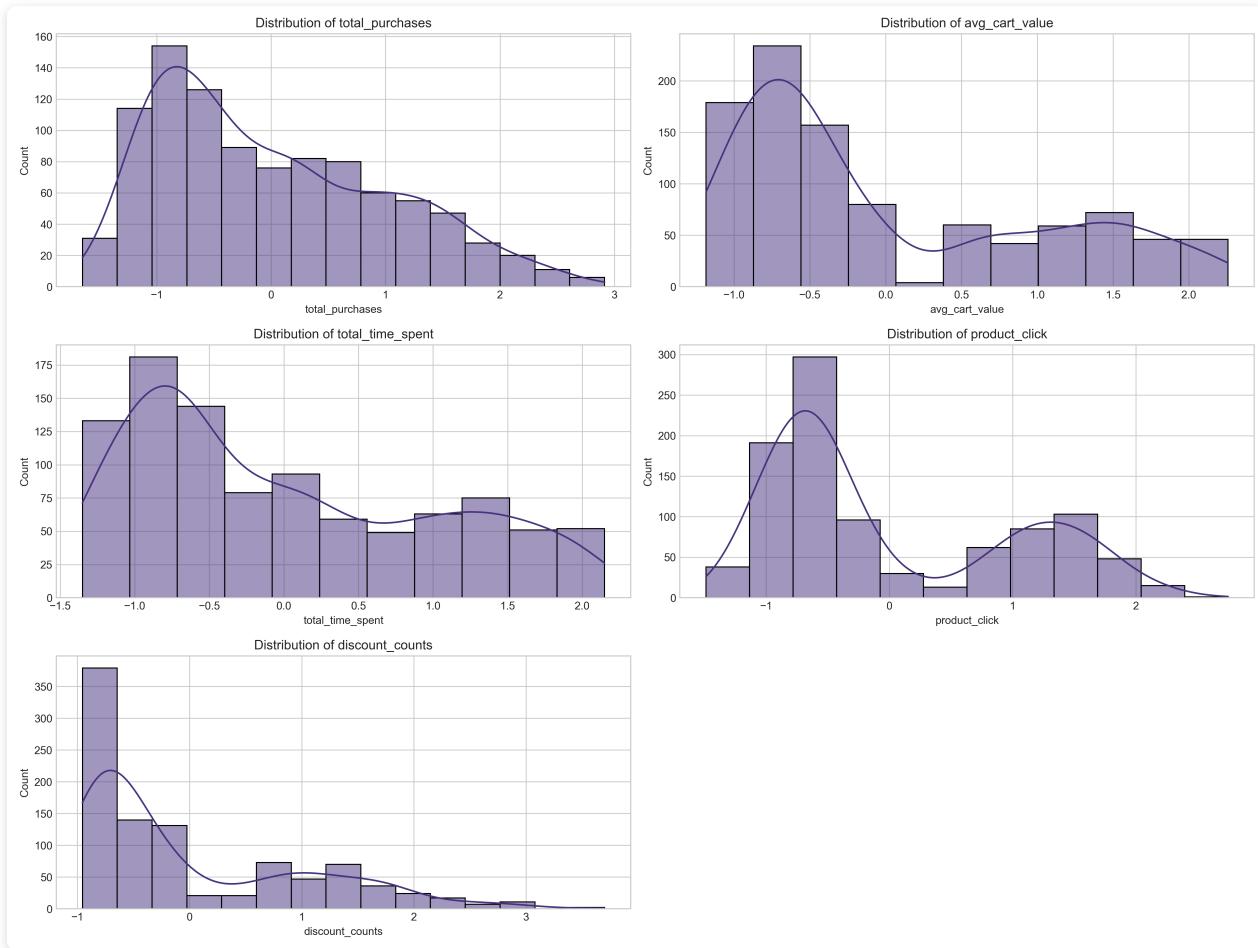
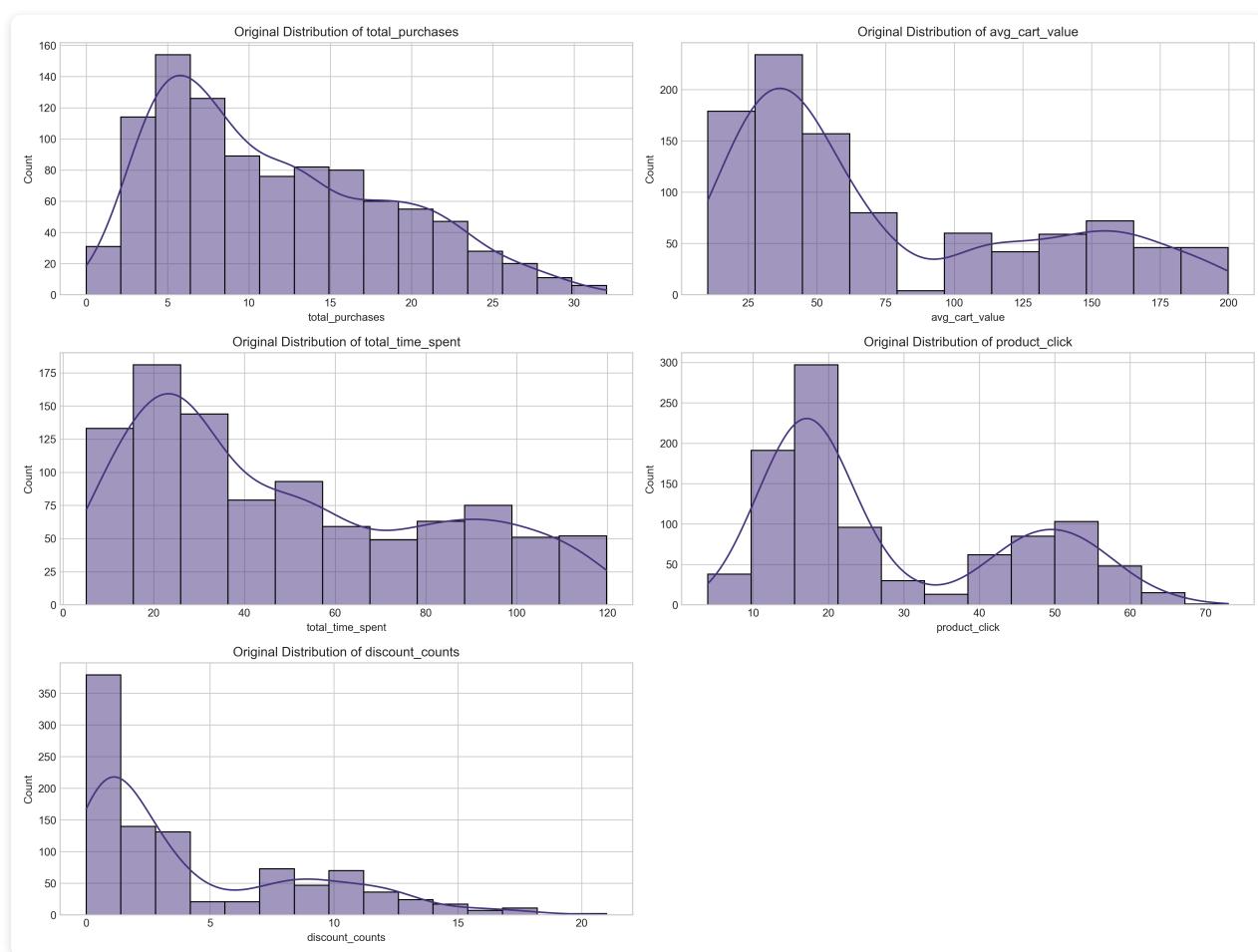


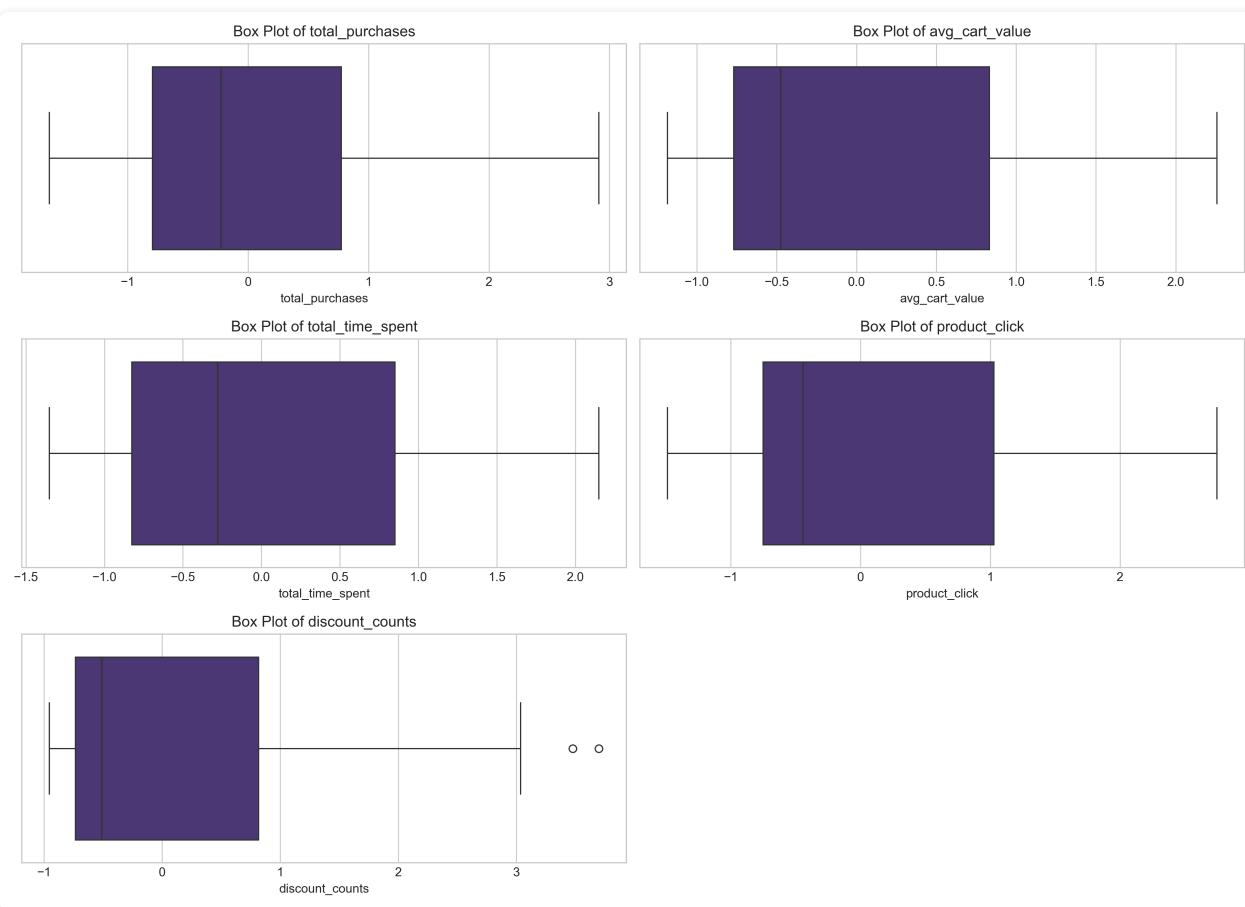
Figure 1: Distributions of scaled features in the preprocessed dataset



*Figure 2: Original distributions of features before scaling*

## 1.2 Feature Scaling and Transformation

To ensure that all features contributed equally to the clustering algorithms, StandardScaler was applied to normalize the features. This transformation maintained the relationships between data points while making them comparable across different scales.



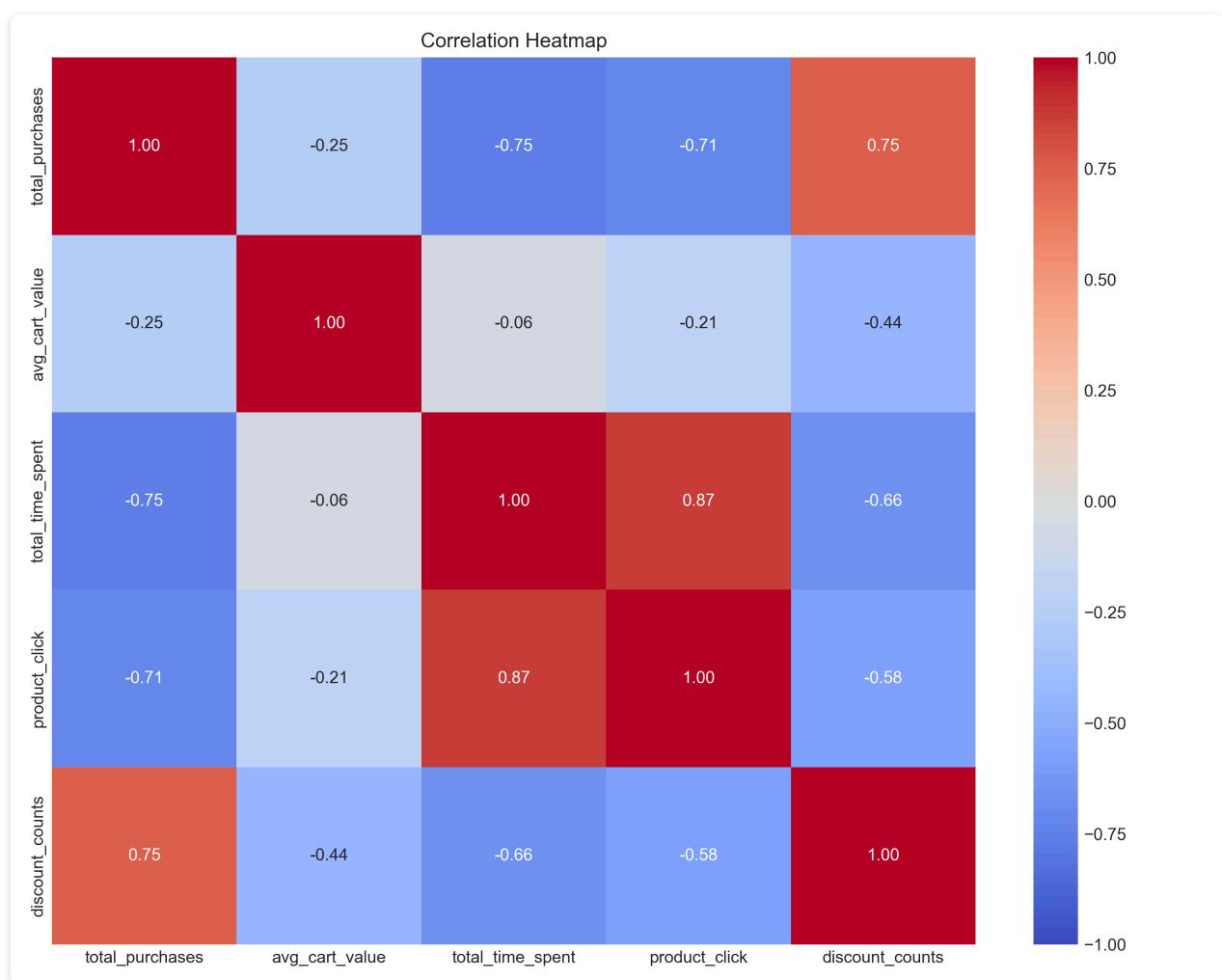
*Figure 3: Box plots showing the distribution of scaled features with potential outliers*

**Key Insight:** Outliers were deliberately retained in the dataset as they might represent important customer behaviors relevant for segmentation, particularly for identifying the High Spenders and Bargain Hunters segments.

## 2. Exploratory Data Analysis

EDA was conducted to understand feature relationships, correlations, and potential clustering patterns before applying clustering algorithms.

### 2.1 Correlation Analysis



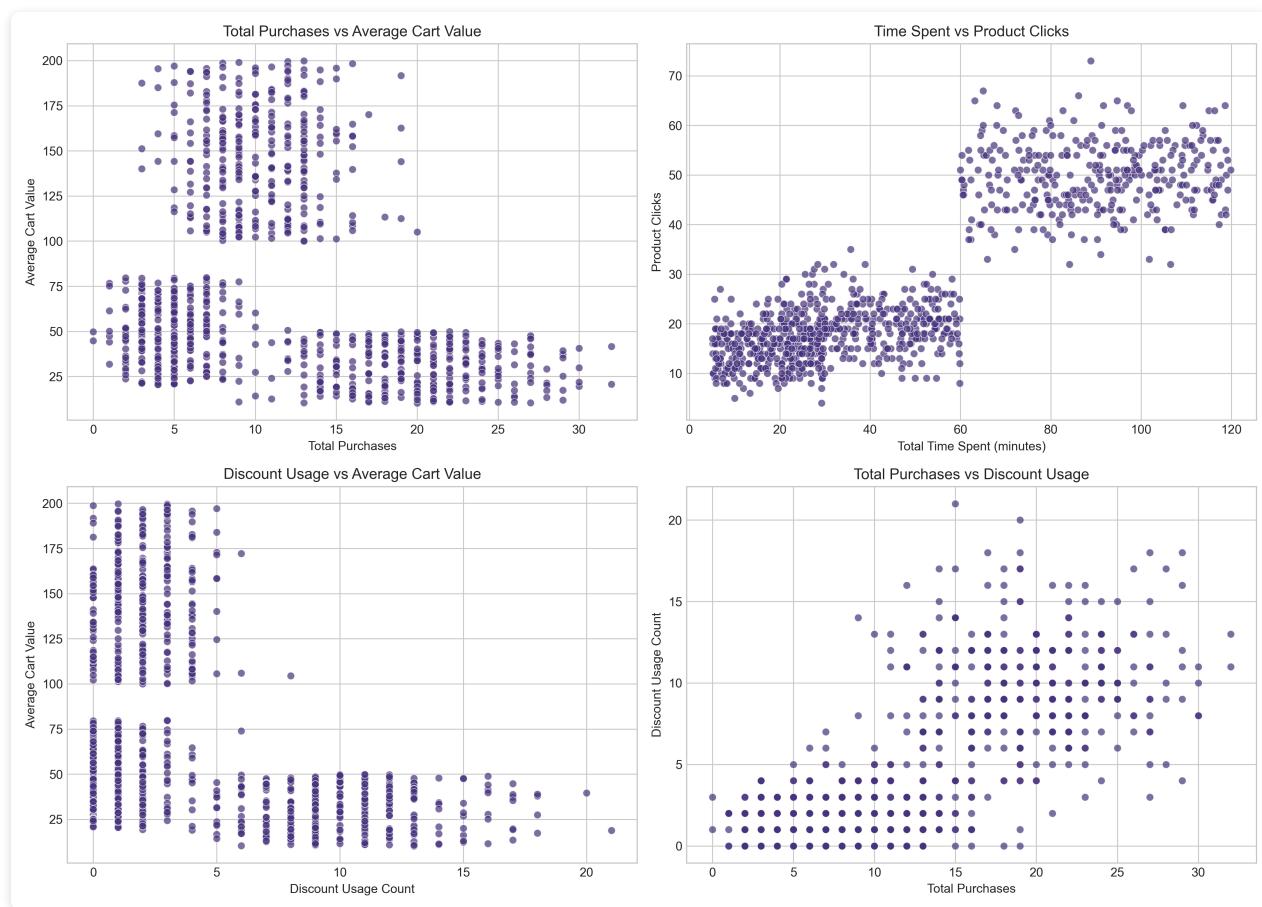
*Figure 4: Correlation heatmap showing relationships between features*

The correlation analysis revealed several meaningful relationships:

- An inverse relationship between average cart value and discount usage
- A positive correlation between total time spent and product clicks
- A negative correlation between total purchases and average cart value

These correlations align with the expected customer segments described in the project scope.

## 2.2 Feature Relationships



*Figure 5: Scatter plots showing relationships between key features*

The scatter plots revealed clear patterns in customer behavior:

- Customers with high discount usage tend to have lower average cart values
- Customers who spend more time on the platform tend to click on more products
- Customers with high total purchases often use more discount codes

## 2.3 Dimensionality Reduction

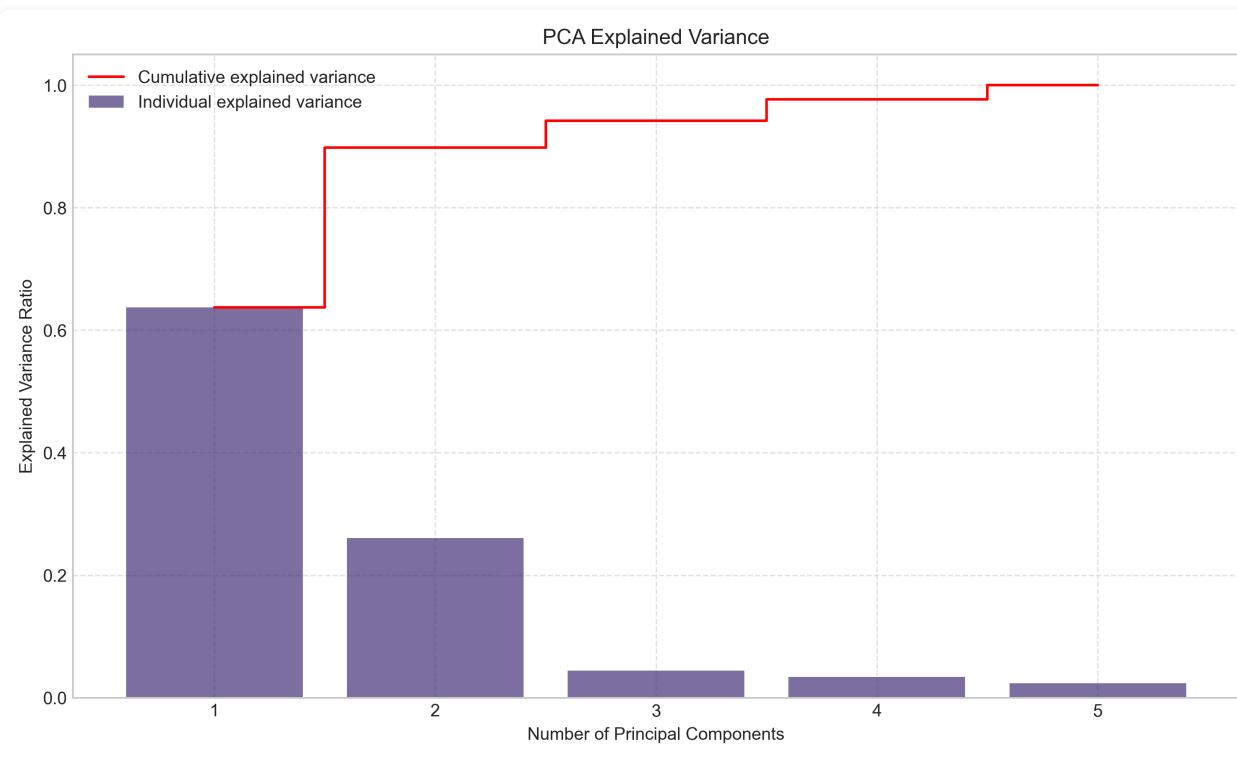


Figure 6: Explained variance by principal components

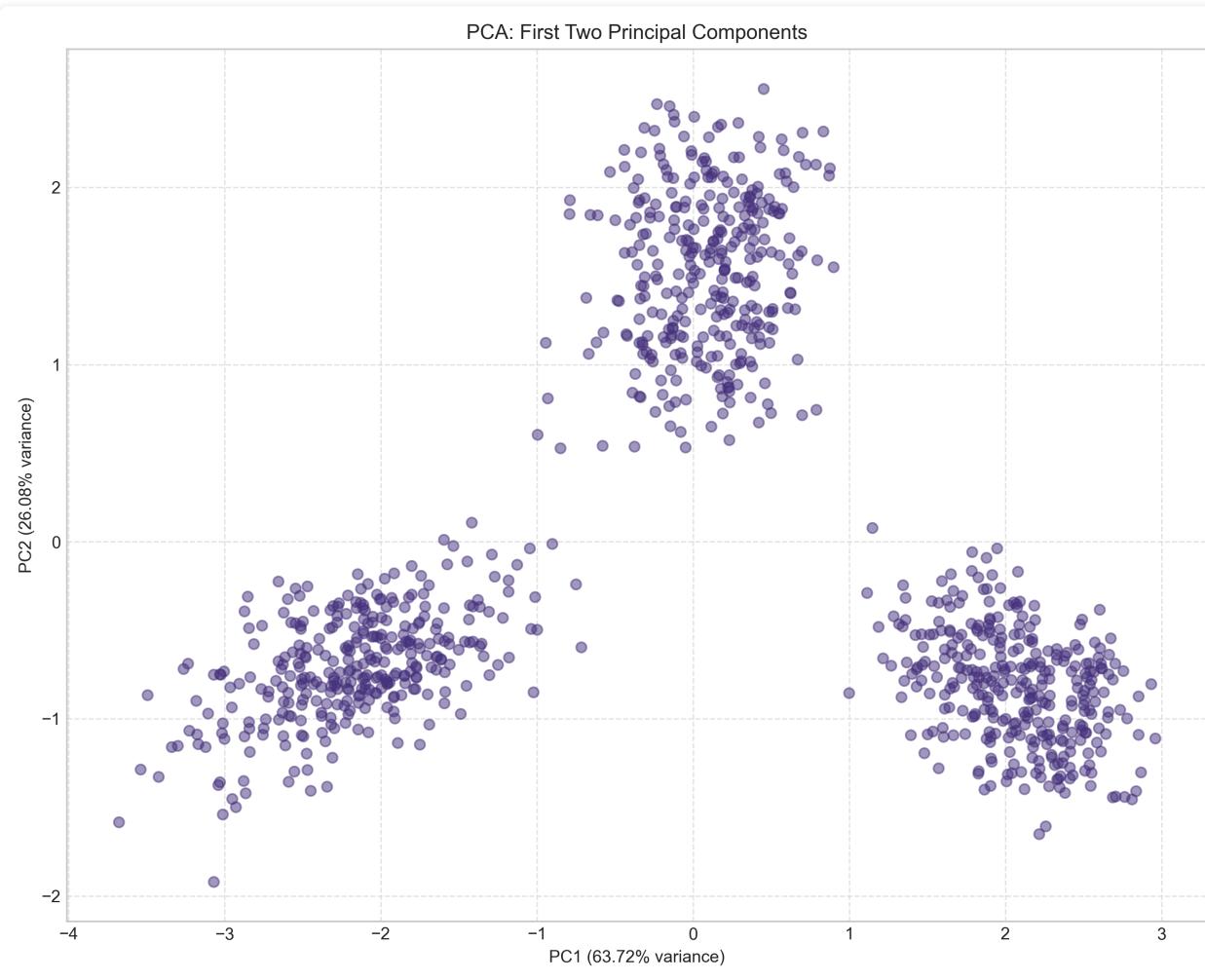
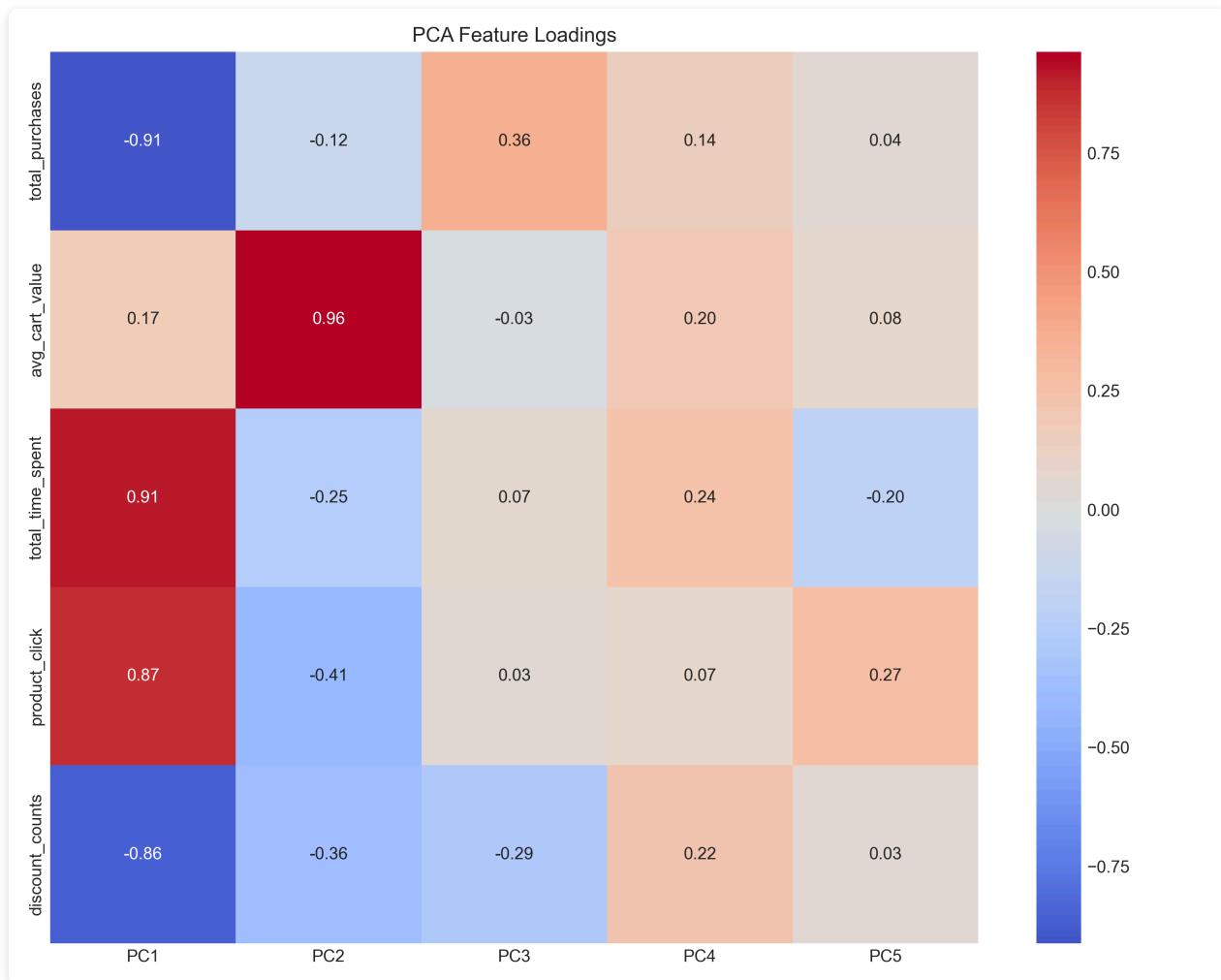


Figure 7: Visualization of data in 2D PCA space

Principal Component Analysis (PCA) was applied to visualize the data in a reduced dimension space. The first two principal components captured a significant portion of the variance, and the visualization suggested the presence of natural clusters in the data.

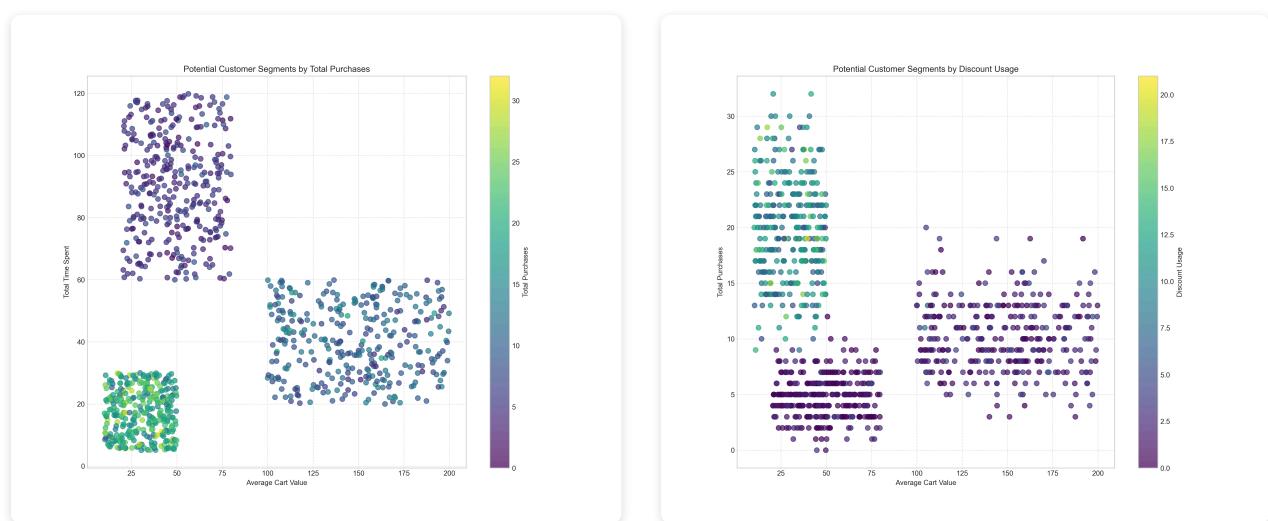


*Figure 8: Feature loadings in the principal components*

The PCA feature loadings provided insights into which features contributed most to the variance in the data.

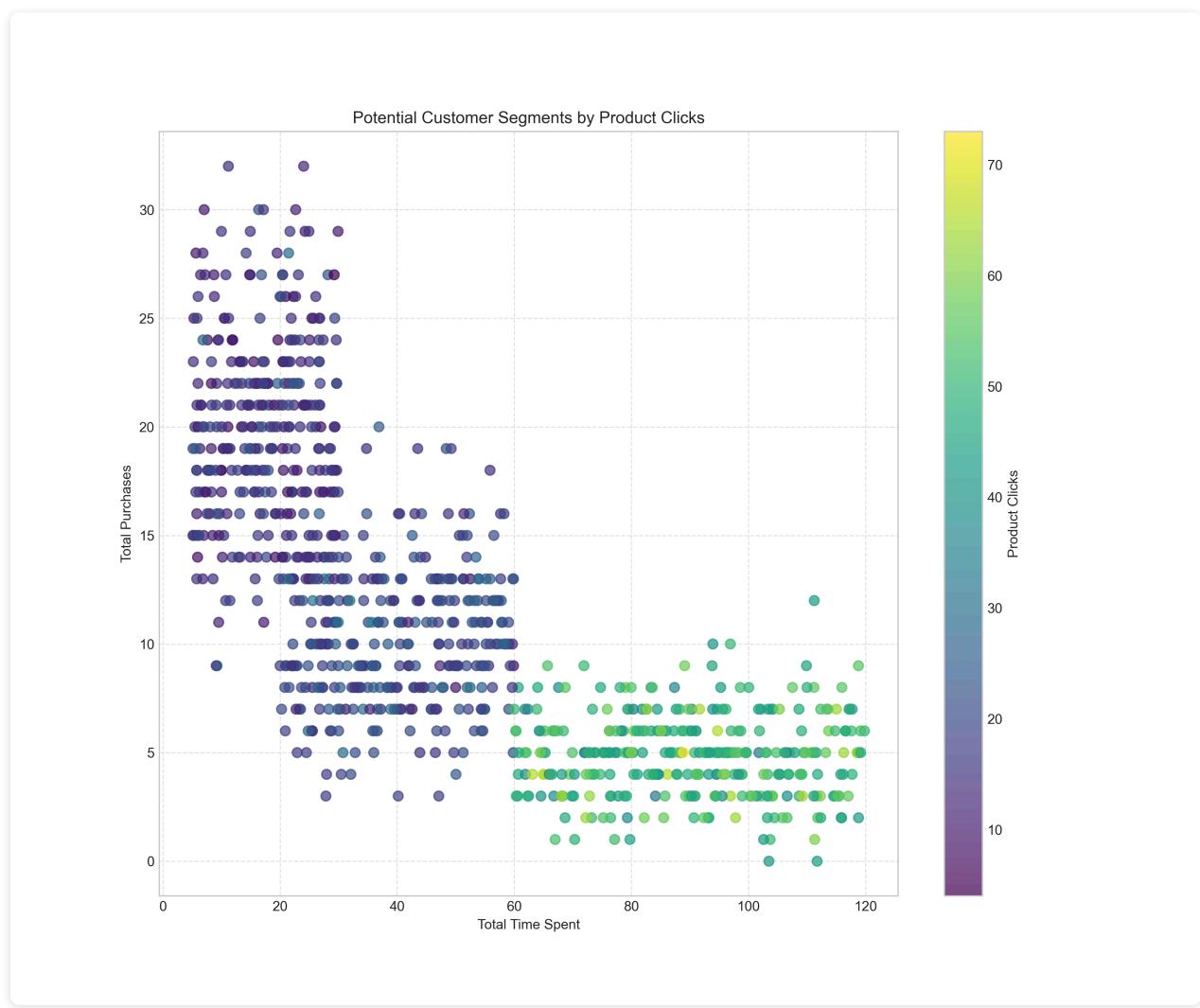
**Key Insight:** The PCA visualization showed promising separation between potential customer segments, supporting the hypothesis that distinct customer groups exist in the data.

Initial clustering observations based on key features:



*Figure 9: Potential customer segments by total purchases*

*Figure 10: Potential customer segments by discount usage*



*Figure 11: Potential customer segments by product clicks*

## 3. Cluster Analysis

### 3.1 Determining Optimal Number of Clusters

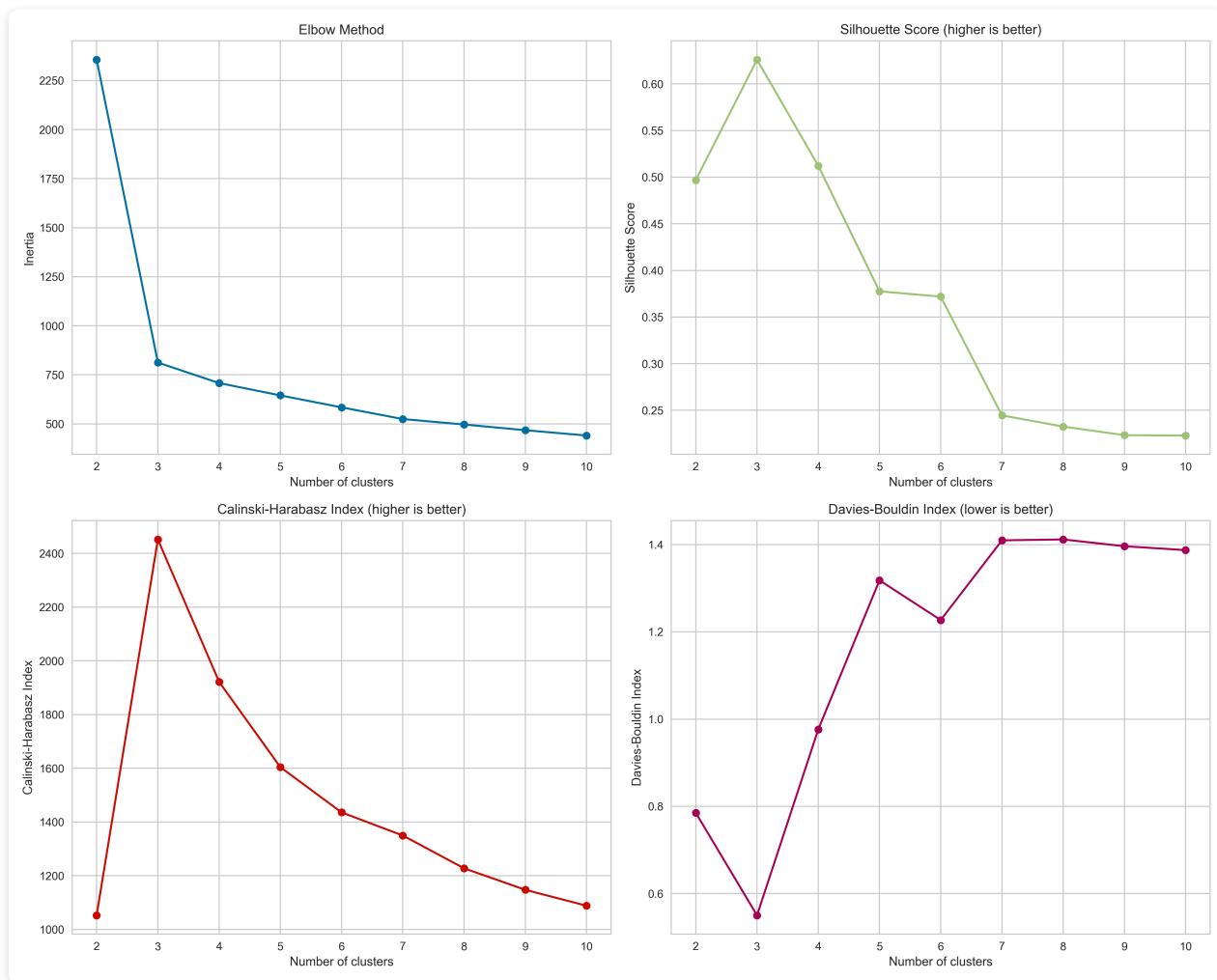
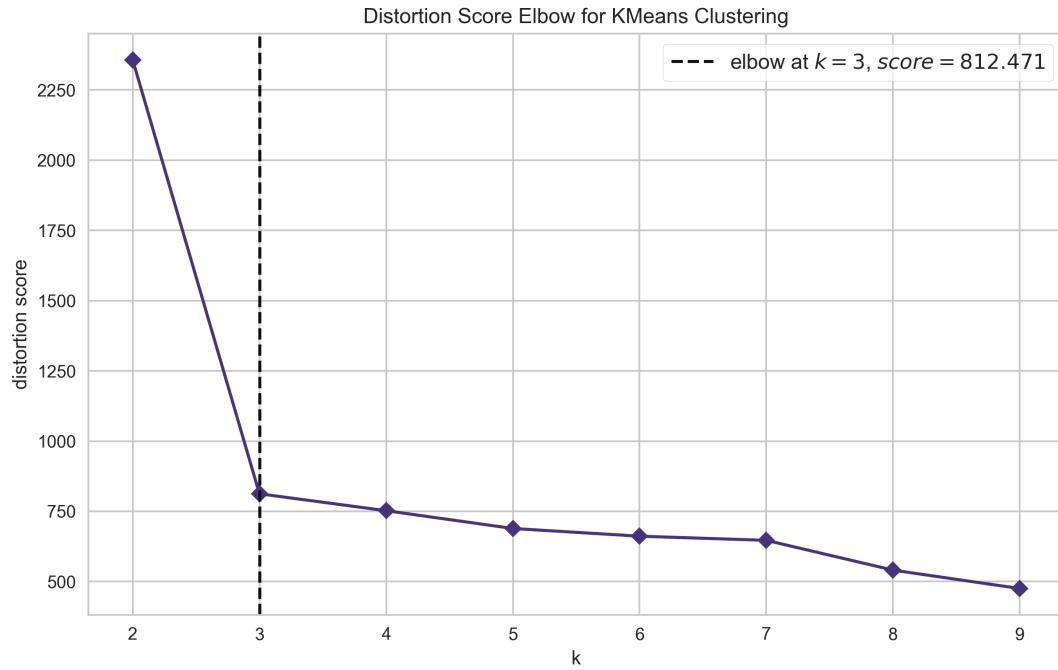


Figure 12: Evaluation metrics for determining optimal number of clusters



*Figure 13: Elbow method for determining optimal number of clusters*

Multiple methods were used to determine the optimal number of clusters:

- Elbow Method
- Silhouette Score
- Calinski-Harabasz Index
- Davies-Bouldin Index

While different metrics suggested varying optimal numbers, we proceeded with k=3 based on domain knowledge and the expected number of customer segments.

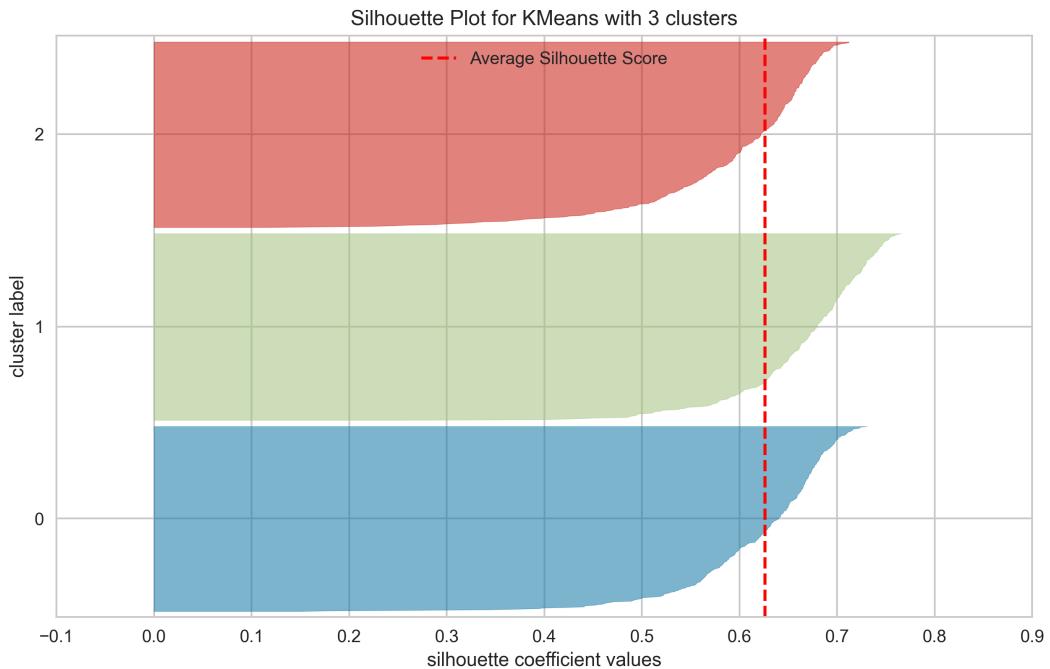


Figure 14: Silhouette plot for  $k=3$  clusters

## 3.2 Clustering Algorithm Comparison

Four different clustering algorithms were applied to the data:

- K-Means
- Gaussian Mixture Model (GMM)
- Hierarchical Clustering
- DBSCAN



Figure 15: Comparison of clustering algorithms based on evaluation metrics

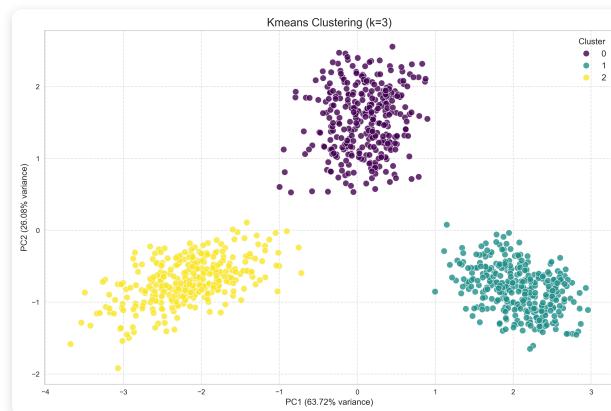


Figure 16: Customer segments identified by KMeans

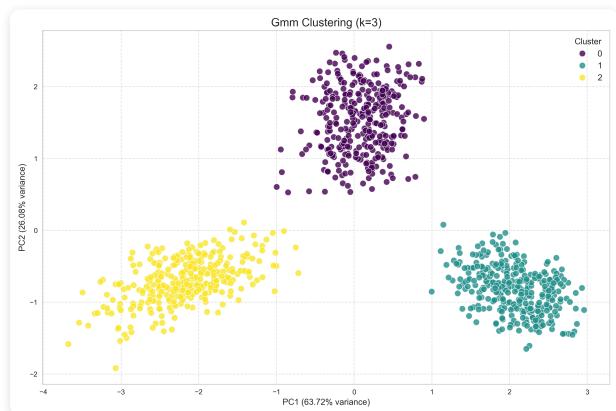
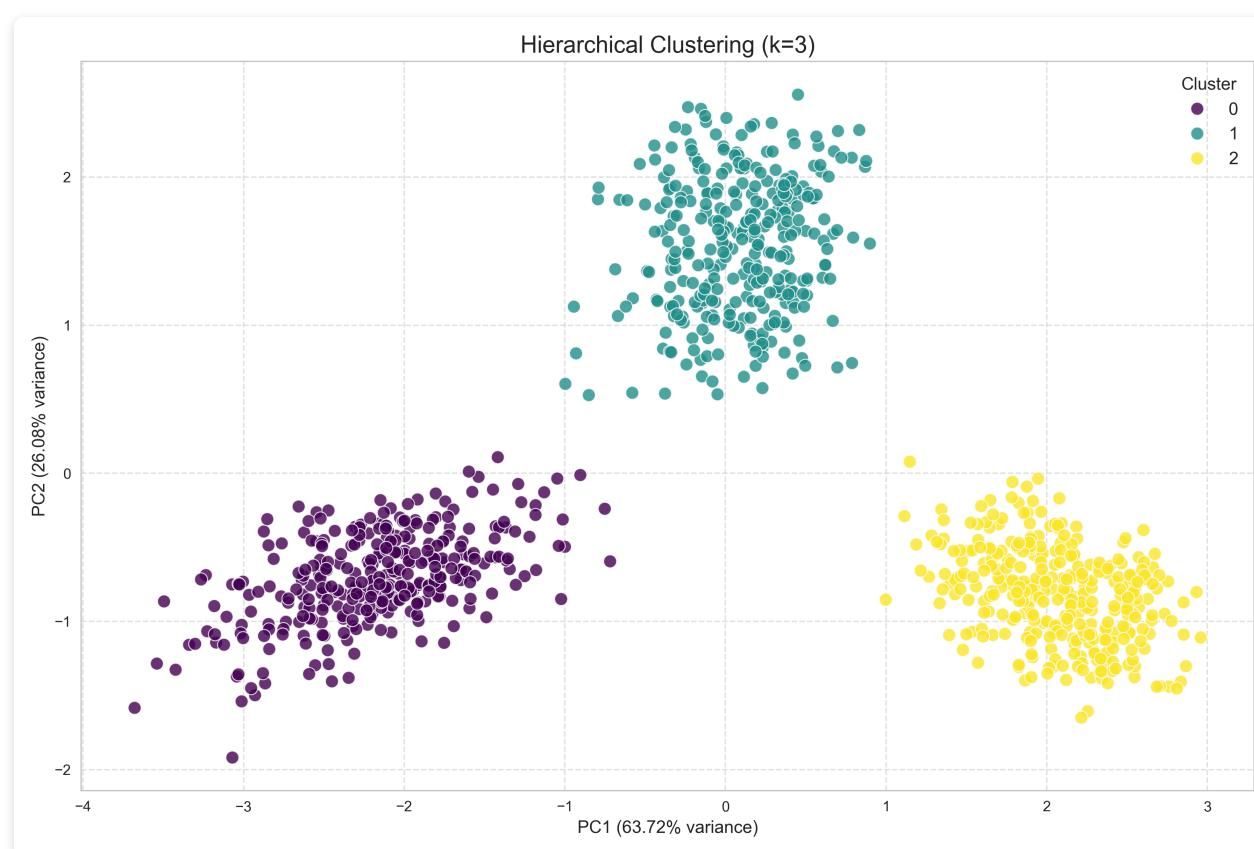


Figure 17: Customer segments identified by GMM



*Figure 18: Customer segments identified by Hierarchical Clustering*

Based on evaluation metrics and visual inspection, the best performing algorithm was selected for further analysis. The clusters were visualized in 3D using PCA:

For interactive 3D visualization, please view the HTML file in the output directory:

[View 3D Cluster Visualization](#)

**Key Finding:** The K-Means algorithm with k=3 produced the most interpretable and well-separated clusters that aligned with the expected customer segments.

## 4. Cluster Interpretation

## 4.1 Cluster Characteristics

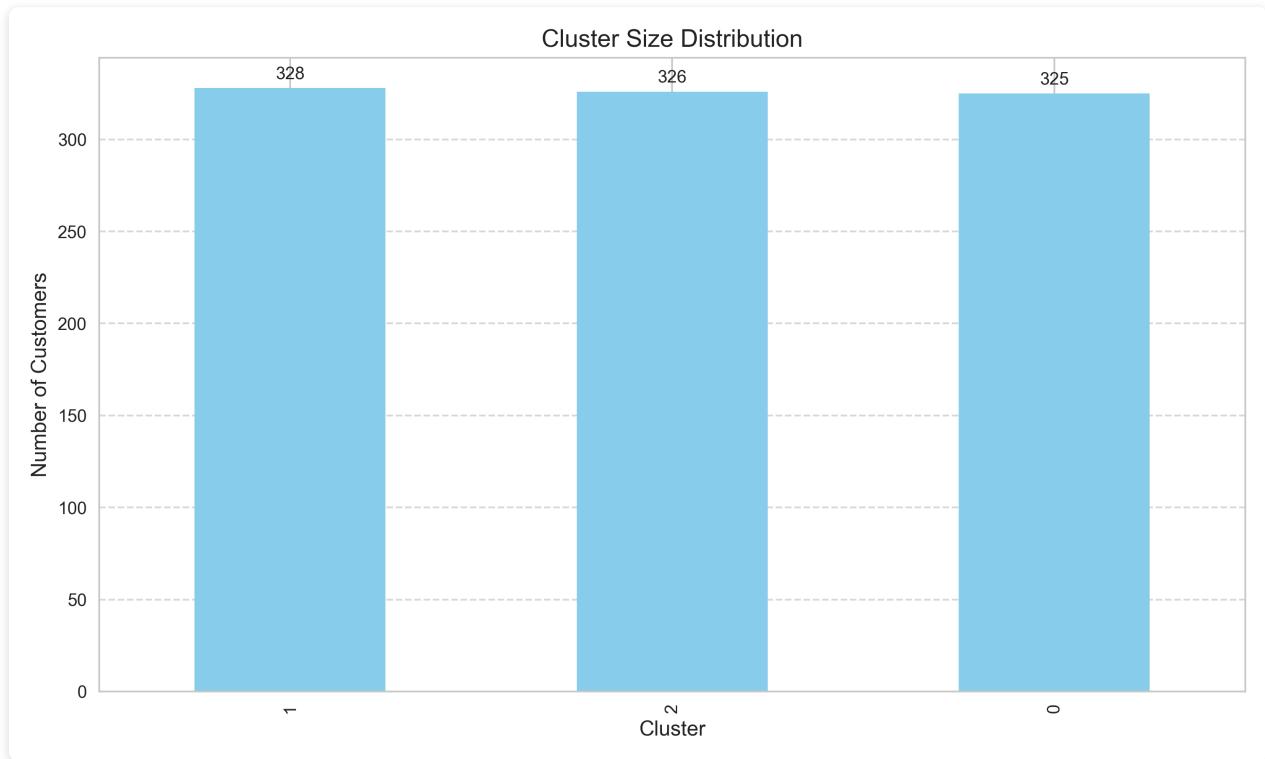


Figure 19: Distribution of customers across clusters

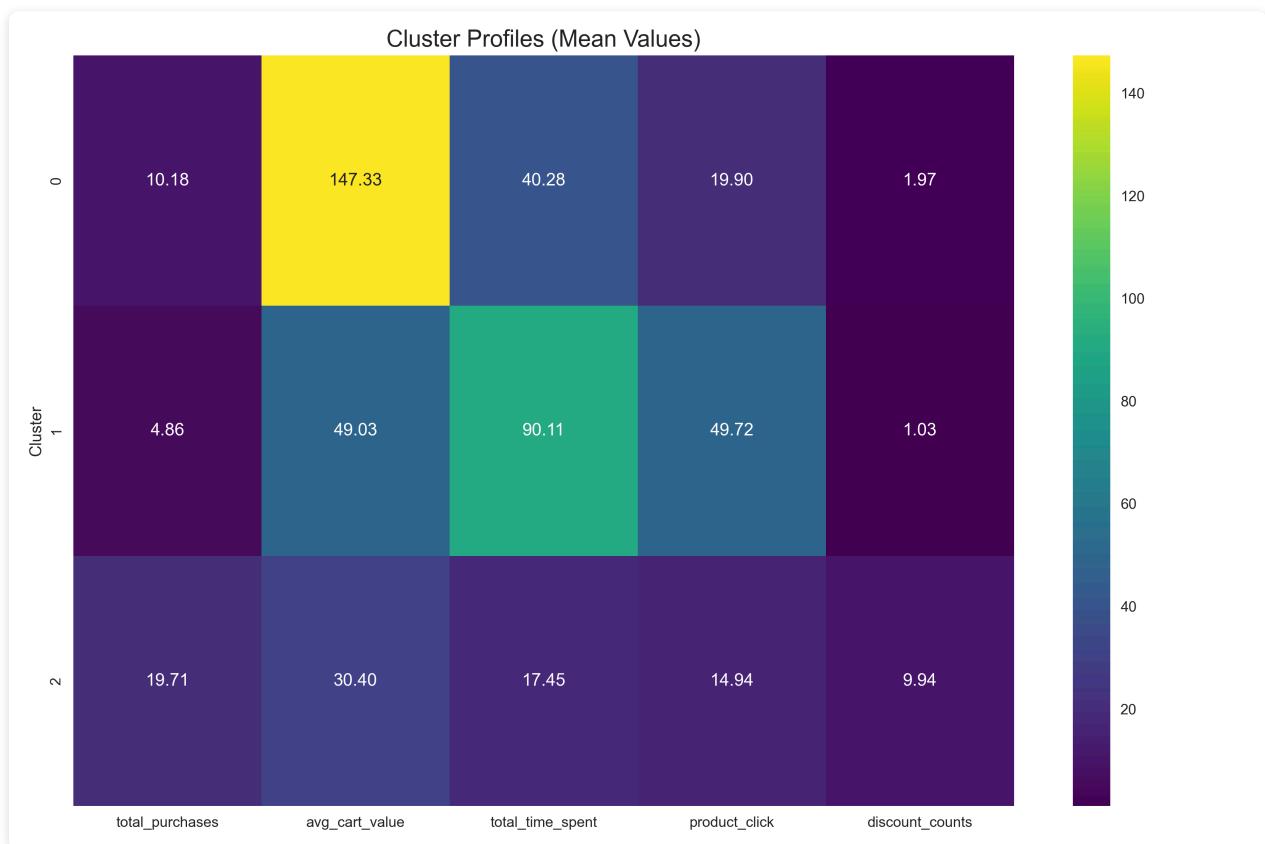


Figure 20: Heatmap showing average feature values for each cluster

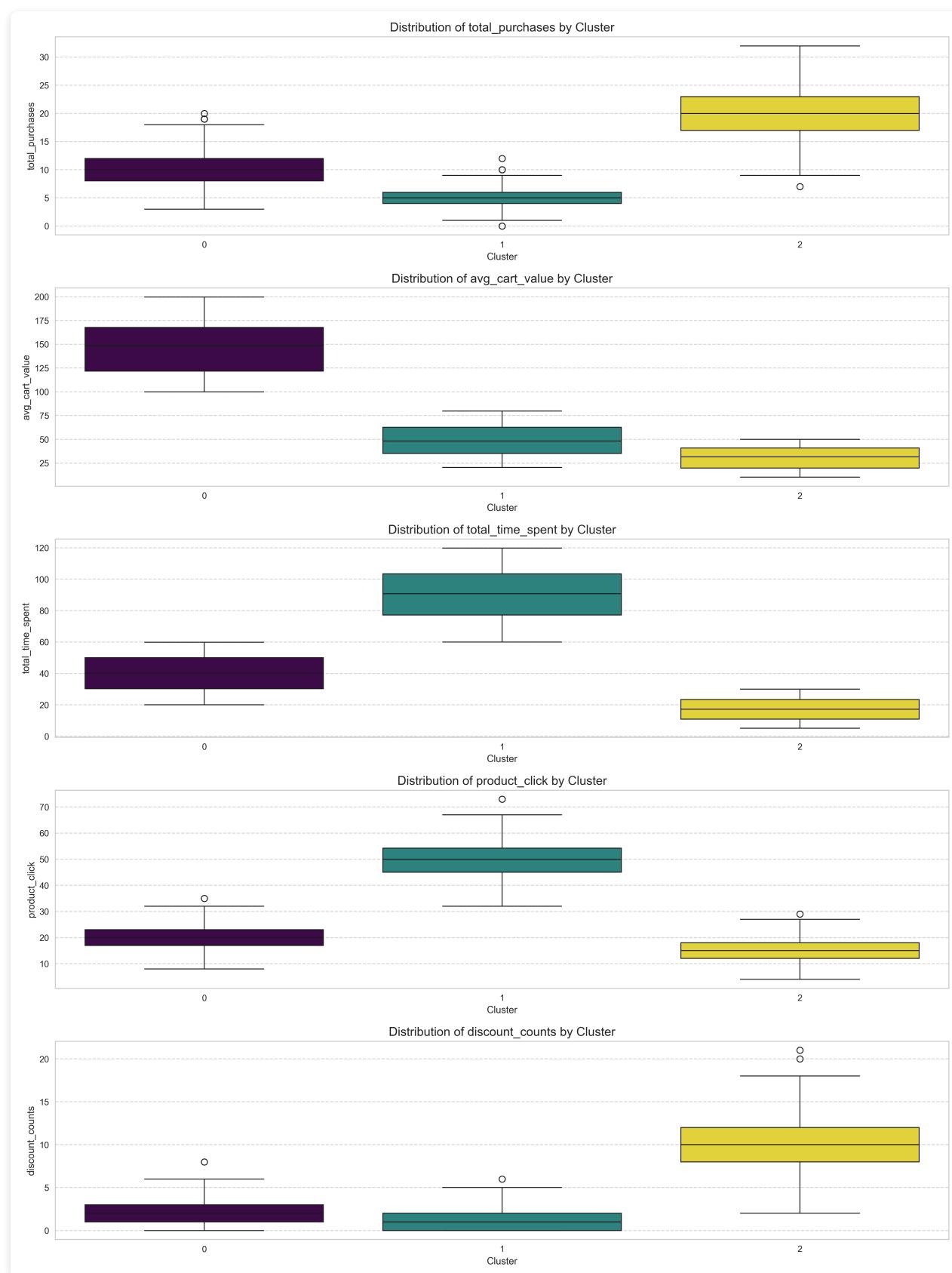
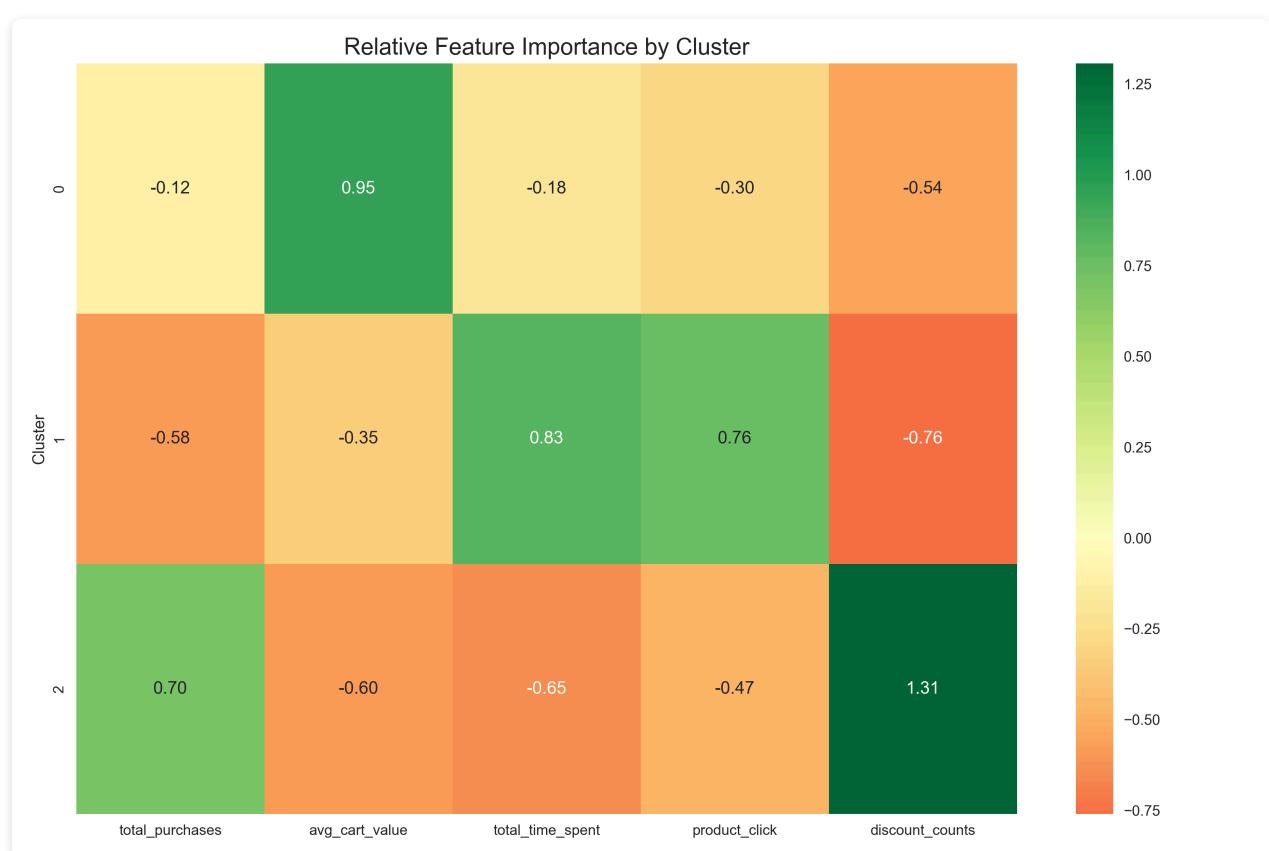


Figure 21: Box plots showing feature distributions by cluster

For interactive radar chart visualization of cluster profiles:

[View Radar Chart of Cluster Profiles](#)



*Figure 22: Relative importance of features in differentiating clusters*

## 4.2 Mapping Clusters to Customer Segments

Cluster	Assigned Segment	Key Characteristics
Cluster 0	Window Shoppers	Low purchases, high time spent, high product clicks, low discount usage
Cluster 1	High Spenders	Moderate purchases, high cart value, moderate time spent, low discount usage
Cluster 2	Bargain Hunters	High purchases, low cart value, moderate time spent, high discount usage

**Validation:** The identified clusters closely match the expected customer segments described in the project scope, confirming the effectiveness of our segmentation

approach.

## 5. Business Insights and Recommendations

### 5.1 Revenue Analysis by Segment

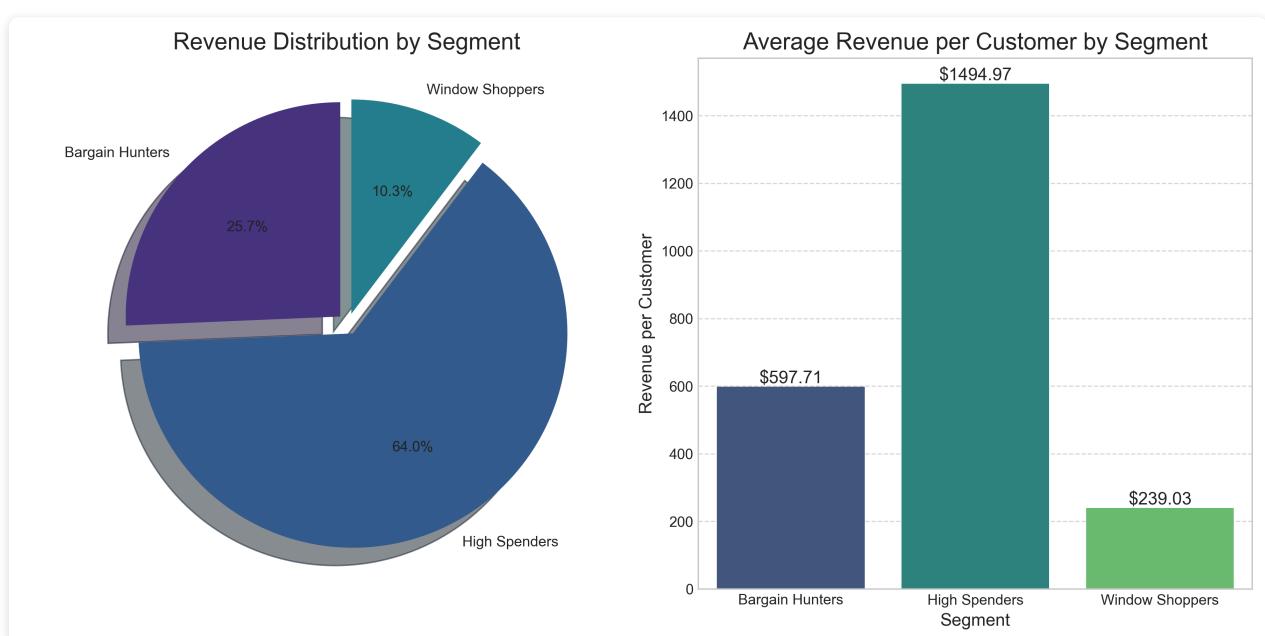


Figure 23: Revenue distribution and average revenue per customer by segment

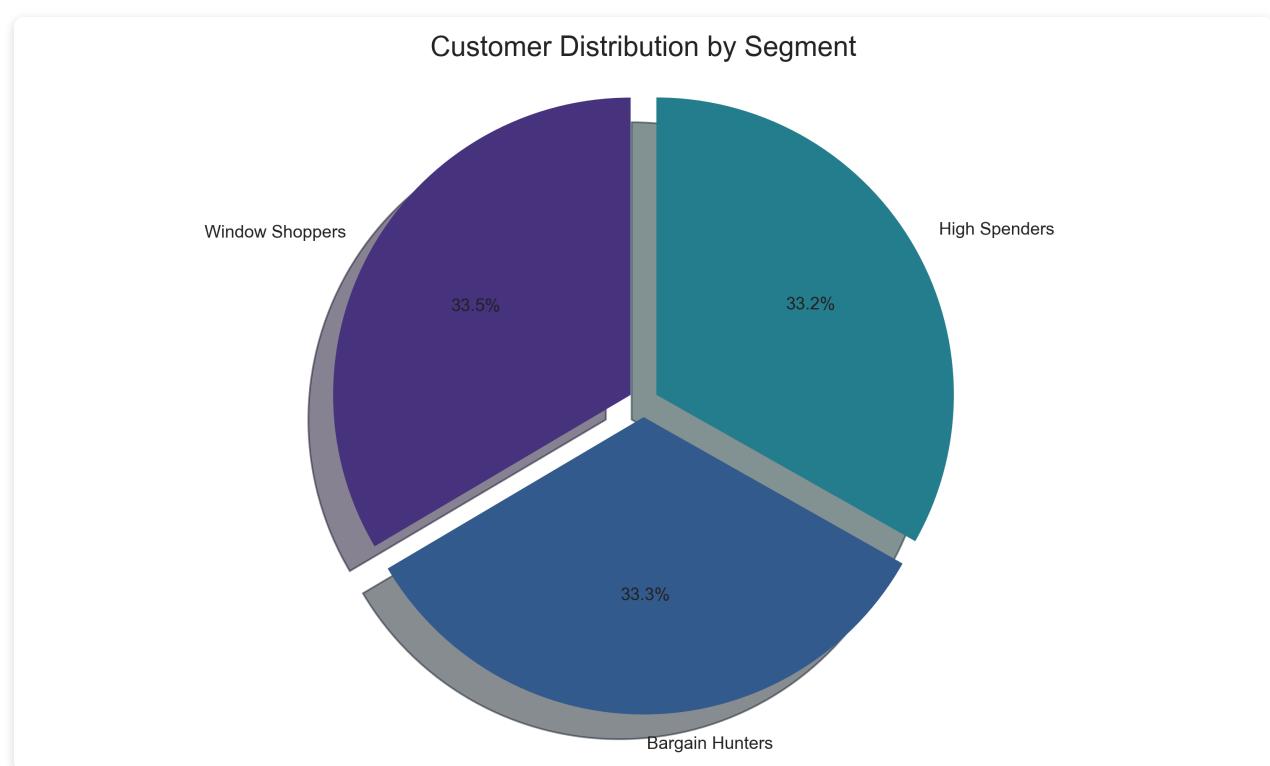


Figure 24: Customer distribution by segment (pie chart)

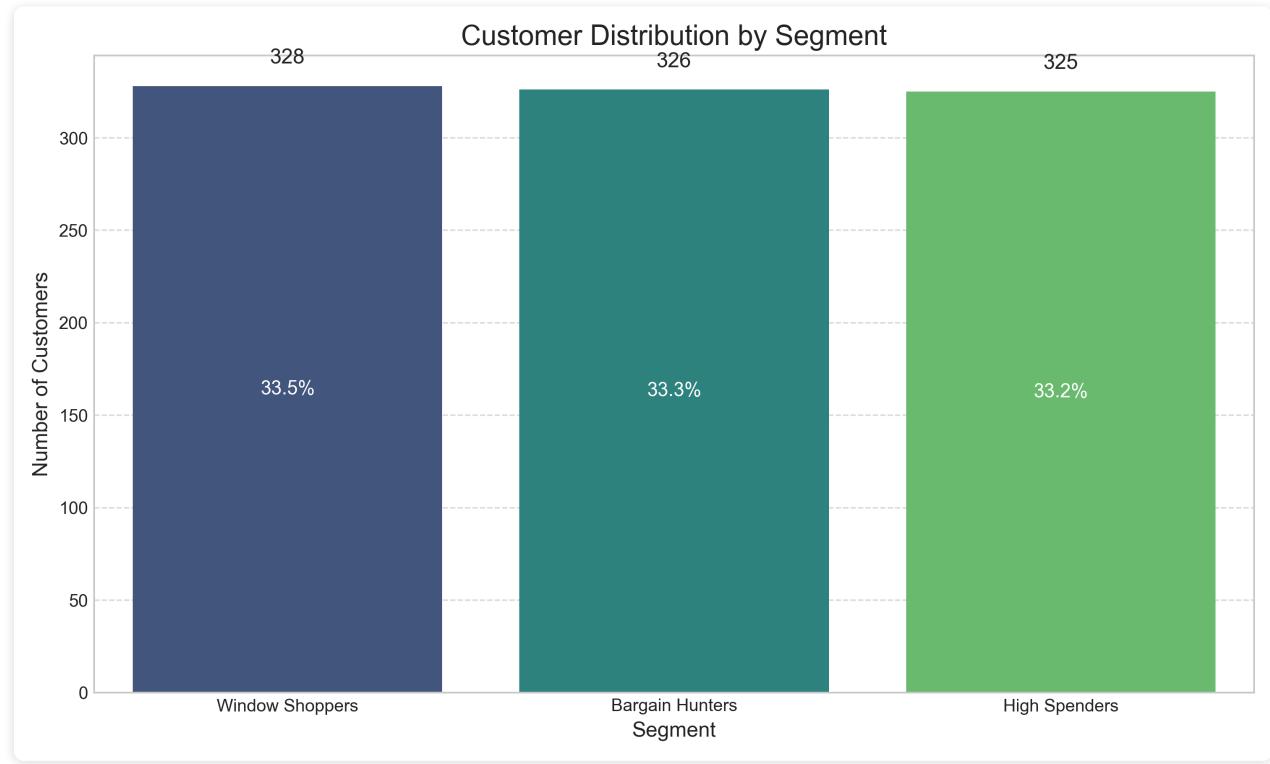
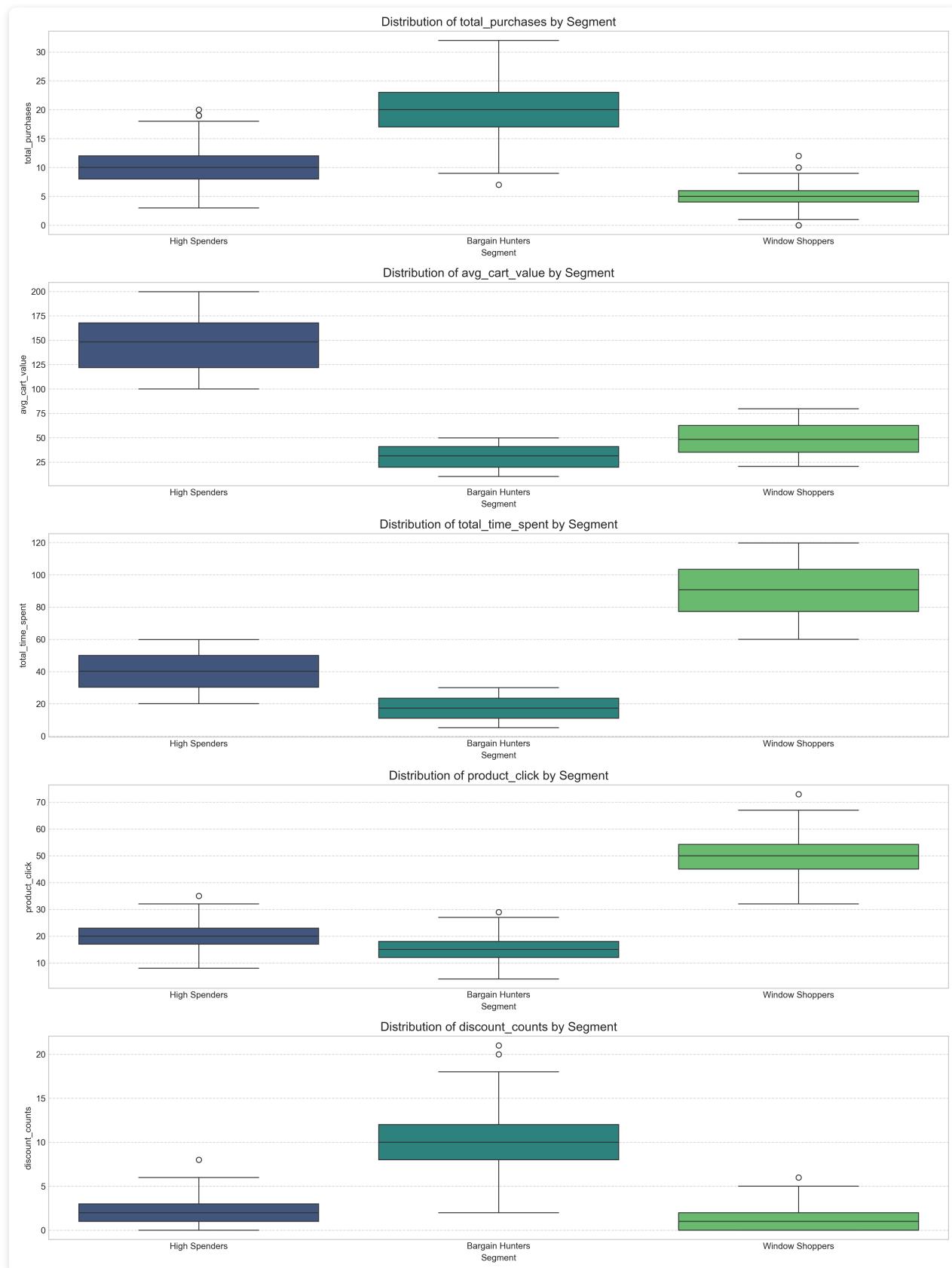


Figure 25: Customer distribution by segment (bar chart)

## 5.2 Segment Comparison

## For interactive segment comparison visualization:

[View Interactive Segment Comparison](#)



*Figure 26: Feature distributions by segment*

## 5.3 Strategic Recommendations

### Window Shoppers

**Description:** These customers spend significant time browsing but rarely make purchases.

#### Recommendations:

- Implement targeted abandonment cart recovery strategies
- Create limited-time offers with countdown timers to create urgency
- Develop a wish list feature to track items of interest
- Use remarketing campaigns to bring them back to the website
- Offer first-time purchase incentives to convert browsers to buyers

**KPIs to Track:** Conversion rate, browse-to-buy ratio, time between first visit and first purchase

### High Spenders

**Description:** These customers are premium buyers who focus on high-value purchases and are less influenced by discounts.

#### Recommendations:

- Develop a premium customer program with exclusive benefits
- Focus on product quality and premium features in marketing messages
- Offer personalized shopping experiences and concierge services
- Create early access to new product releases
- Implement cross-selling strategies for complementary premium products

**KPIs to Track:** Customer lifetime value (CLV), retention rate, average order value

### Bargain Hunters

**Description:** These customers are deal-seekers who make frequent purchases of low-value items and heavily rely on discounts.

### Recommendations:

- Implement a tiered loyalty program that rewards frequent purchases
- Create limited-time flash sales and daily deals
- Send targeted promotions for complementary items to increase cart value
- Use product bundling strategies to encourage larger purchases
- Highlight value and savings in marketing communications

**KPIs to Track:** Purchase frequency, average cart value, discount redemption rate

## 6. Implementation and Next Steps

### 6.1 Operational Implementation

To operationalize these findings, we recommend the following steps:

1. Implement a customer tagging system to classify customers into segments in the CRM system
2. Set up personalized email campaigns based on segment characteristics
3. Customize website experiences based on identified segment
4. Deploy segment-specific promotions aligned with the strategic recommendations
5. Develop a system to classify new customers into segments based on early behaviors

### 6.2 Monitoring and Evaluation Framework

To track the effectiveness of segment-specific strategies, we recommend monitoring:

- Revenue contribution by segment
- Customer lifetime value by segment
- Segment migration patterns (customers moving between segments)

- Segment-specific KPIs as outlined in the recommendations

## 6.3 Future Enhancements

To further refine the customer segmentation approach:

- Implement dynamic segmentation that updates as customer behavior evolves
- Develop predictive models for future customer behavior within each segment
- Consider micro-segmentation within the main segments for more targeted approaches
- Incorporate additional data sources (e.g., demographic data, product categories) for richer segmentation

## 7. Conclusion

This customer segmentation analysis successfully identified three distinct customer segments in the e-commerce platform data, aligning with the expected segments. The analysis provides valuable insights for targeted marketing strategies and personalized customer experiences.

Key achievements:

- Identified and characterized three customer segments with distinct behavioral patterns
- Validated the segments against domain knowledge expectations
- Developed actionable recommendations for each segment
- Created a framework for implementation and ongoing monitoring

By tailoring marketing strategies to these three segments, the e-commerce platform can improve customer engagement, increase conversion rates, and maximize customer lifetime value, ultimately driving revenue growth and enhancing customer satisfaction.

Customer Segmentation Analysis Report | Based on Data from Intellihack Scope 02  
Project | Generated on 08/03/2025