Customer Segmentation Clustering Analysis

Introduction

After spending considerable time analyzing the eCommerce dataset and testing different clustering approaches, I've found some really interesting patterns in our customer base. I focused on finding natural groupings in our customer data using various clustering techniques, and I wanted to share my findings in this report.

What I Found

Looking at the numbers first, I ended up with 2 distinct customer groups, which might seem simple but actually makes a lot of sense when you look at the data. Here's why I chose 2 clusters:

The Numbers Behind My Decision

- Davies-Bouldin Index came in at 0.0113 (this is really good - the closer to 0, the better)
- Got a Silhouette Score of 0.9928 (nearly perfect - it goes up to 1.0)
- Calinski-Harabasz Score hit 323,155.91 (way higher than other cluster numbers I tried)

I actually tested anywhere from 2 to 10 clusters, but the 2-cluster solution kept coming up as the winner. When I tried more clusters, the quality metrics got worse - for example, with 10 clusters, the Silhouette Score dropped to about 0.20, which isn't great.

Breaking Down the Clusters

What's really interesting is how the customers split up:

- Cluster 0: 188 customers (our main group)
- Cluster 1: 12 customers (a small but distinct group)

What Makes These Groups Different

Looking at the feature importance chart, I noticed three main things that really separate these groups:

1. How often they buy (transaction count/recency scored 771.78 - huge!)
2. Purchase patterns (frequency score of 66.77)
3. How much they typically buy at once (avg basket size score of 4.95)

Interestingly, where customers are located (region) didn't matter much in separating these groups.

Visual Analysis

The PCA visualization really helped me understand what's going on. You can clearly see the two groups separated on the plot, with no messy overlap. The smaller group (those 12 customers) stands out pretty distinctly from our main customer base.

What This Means for Business

From working with this data, I think we've found something really valuable here. Those 12 customers in Cluster 1 are doing something different from everyone else - they might be our power users or high-value customers. I'd recommend:

1. Taking a closer look at those 12 customers to understand what makes them special
2. Maybe creating different marketing approaches for each group
3. Keeping an eye on how these groups change over time
4. Diving deeper into how each group typically shops with us

Final Thoughts

While I initially thought we might need more customer segments, the data really pushed me toward this two-group solution. It's clean, it's backed by strong statistical measures, and most importantly, it makes practical sense for our business.

The next step would probably be to look more closely at the shopping patterns within each group - I think there's still more we could learn there.