Customer Segmentation Analysis

Executive Summary

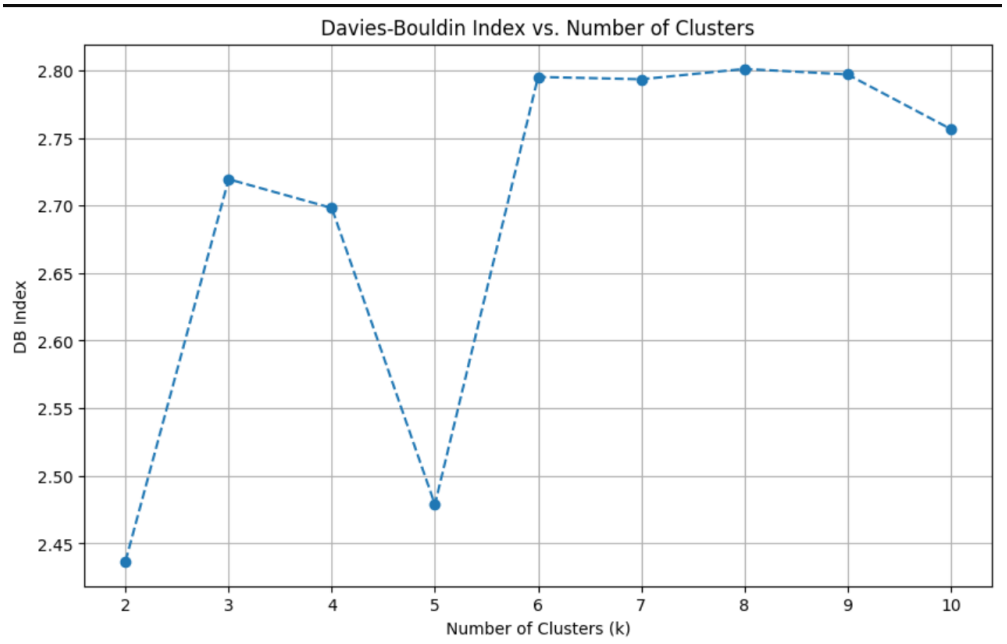
This analysis explores customer segmentation patterns using two distinct clustering approaches: KMeans and KMedoids. We analyzed a complex dataset comprising 111 features that capture customer behavior across transactions, geographic distributions, and product preferences. These features include both quantitative measurements (transaction quantities and prices) and categorical variables (regional information and product preferences) transformed through one-hot encoding.

Methodology and Results

KMeans Clustering Analysis

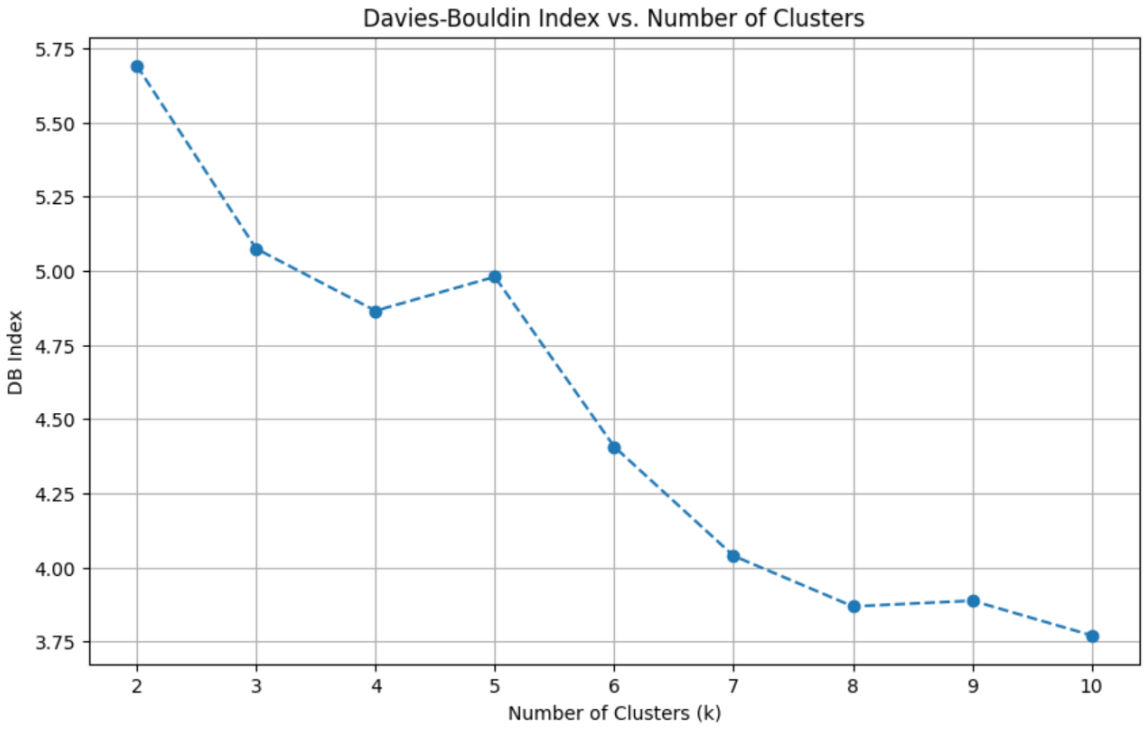
Our KMeans analysis revealed a remarkably clear binary segmentation of customers. The Davies-Bouldin Index reached its optimal value of 2.437 at k=2, suggesting that customers naturally fall into two distinct behavioral groups. As we tested additional clusters, the DB Index showed an interesting pattern: it peaked at k=5 (2.795) and maintained relatively stable values afterward, indicating that additional clusters beyond this point provided diminishing returns.

The progression of DB Index scores across different k values tells an interesting story: k=2: 2.437 | k=3: 2.719 | k=4: 2.698 | k=5: 2.795 | k=6: 2.793 | k=7: 2.801 | k=8: 2.797 | k=9: 2.757 | k=10: 2.801



KMedoids Clustering Analysis

KMedoids painting presented a more nuanced picture of customer segmentation. The algorithm identified ten distinct customer segments as optimal, achieving a DB Index of 3.770. What's particularly noteworthy is the consistent improvement in clustering quality as k increased, demonstrated by the steadily decreasing DB Index scores: k=2: 5.691 | k=3: 5.074 | k=4: 4.865 | k=5: 4.979 | k=6: 4.407 | k=7: 4.038 | k=8: 3.868 | k=9: 3.887 | k=10: 3.770



Comparative Analysis

The stark contrast between KMeans (k=2) and KMedoids (k=10) optimal clusters reveals fundamental differences in how these algorithms interpret our customer base. Consider the following aspects:

Algorithm Performance Characteristics

KMeans offers a straightforward, binary classification that might align well with high-level business strategies (such as premium versus value-seeking customers). Its Davies-Bouldin Index of 2.437 suggests relatively well-defined cluster boundaries, though its sensitivity to outliers warrants careful consideration.

KMedoids, with its ten-cluster solution, provides a more granular view of customer segments. While its higher Davies-Bouldin Index (3.770) might initially seem less favorable, this complexity could capture subtle customer behavior patterns that a simpler model might miss.

Dimensionality Considerations

Working with 111 features presents significant challenges for both algorithms. Our visualization through PCA, while necessary for human interpretation, inevitably sacrifices some information. The differences in cluster solutions between KMeans and KMedoids might partially reflect how each algorithm handles this high-dimensional space.