

Clustering Report

K-Means Clustering:

- Customer Segmentation: K-Means groups customers based on purchase behavior (e.g., quantity, price), helping tailor marketing strategies for different segments.
- Efficiency and Simplicity: The algorithm is fast and easy to implement, suitable for large datasets.
- Interpretability: It provides clear, understandable clusters, making it easy to identify and target distinct customer types.
- Optimized k: The number of clusters ($k=6$) is chosen based on domain knowledge and evaluated using metrics like silhouette score and Davies-Bouldin index.

PCA (Principal Component Analysis):

- Visualization: PCA reduces high-dimensional data to 2D, allowing easy visualization of clusters in scatter plots.
- Simplification: It condenses the data into the most important components, improving the clarity of clustering results.
- Cluster Separation: PCA helps assess how well the clusters are separated, improving interpretation of the clustering model.

Number of Clusters Formed: 6

- While the Davies-Bouldin Index was lowest at $k=10$, we decided to go with $k=6$ because having 10 clusters for just 200 customers would create too many small, very specific groups. This could complicate our marketing efforts, as it would be difficult to create targeted campaigns for so many tiny segments. Additionally, $k=6$ doesn't make a significant difference in the Davies-Bouldin Index compared to $k=10$, meaning the clustering quality remains similar. We chose $k=6$ because it offers a practical balance: six clusters are distinct enough to represent different customer behaviors, yet manageable enough to allow for focused, effective marketing strategies. This way, we avoid unnecessary complexity and still gain valuable insights that can drive business decisions.

Silhouette Score for $k=6$: 0.3122

- This score reflects how well the clusters are grouped together and separated from each other. A score closer to 1 is ideal, meaning the clusters are distinct and easy to tell apart. With a score of **0.3122**, it suggests that the clusters are somewhat separated, but there's still room to improve the differentiation between customer segments. While the clustering is fairly good, some adjustments could be made to make the customer groups even clearer and more distinct.

Davies-Bouldin Index for k=6: 1.2105

- This index shows how similar each cluster is to its closest neighbor. A lower value is better, meaning the clusters are well-defined and distinct. With a value of **1.2105**, the clusters are quite clear and separate, with minimal overlap. This suggests the customer groups are well-defined and easy to differentiate from one another, making it a solid foundation for targeted marketing strategies.

Cluster Analysis Explanation

Cluster 1: Customers with slightly below-average spending and low purchase quantities. These customers might be occasional buyers or less engaged.

Cluster 2: Customers with higher-than-average spending and purchases, predominantly from South America. These could be moderately loyal customers worth nurturing.

Cluster 3: This group, also from South America, shows high engagement with above-average spending and purchases. These are likely your more profitable customers.

Cluster 4: Customers from South America with low engagement and low spending. They could be newer customers who haven't developed strong purchasing habits yet.

Cluster 5: Customers from Asia with below-average spending and fewer purchases. These might be less engaged and need more attention to increase loyalty.

Cluster 6: Customers from Asia showing relatively high spending but low purchases. They could be loyal customers who have been around for a while, showing potential for consistent engagement.

Actionable Insights

High-Value Customers (Cluster 3):

- Focus on retention with exclusive rewards and loyalty programs.

Moderately Engaged Customers (Cluster 2):

- Boost engagement with targeted promotions and premium offers.

Inactive Customers (Cluster 1 & 6):

- Run re-engagement campaigns with personalized offers and discounts.

New or Emerging Customers (Cluster 4 & 5):

- Use onboarding campaigns and incentives to drive initial engagement.

Loyal Customers (Cluster 5):

- Reward loyalty with VIP perks and early access to products.

