Report on Customer Segmentation

Prepared by
Parth Darshan
IIT Jodhpur
CSE, 2026

Overview

This report presents the findings and methodologies used in performing cluster analysis on the provided dataset. The analysis involved customer segmentation using K-Means clustering, supported by Principal Component Analysis (PCA) for visualization and quality evaluation metrics for validation.

Key Metrics

- **Number of Clusters:** 9
- **Davies-Bouldin Index:** 0.4498 (indicating excellent cluster separation)
- **Silhouette Score:** 0.5499 (indicating moderately well-defined clusters)

Methodology

1. Preprocessing and EDA

- **Objective:** To clean and explore the data for patterns, trends, and inconsistencies.
- **Steps Taken:**
  - Identified and handled missing values.
  - Standardized features to ensure uniform scaling.
  - Analyzed patterns and distributions for actionable insights.

2. Optimal Clusters Identification

- Applied the **K-Means clustering algorithm** to the dataset.
- Utilized the **Elbow Method** to determine the optimal number of clusters. Nine clusters were identified as the most suitable choice for segmentation.

3. Dimensionality Reduction with PCA

- **Objective:** To reduce dimensionality and improve interpretability.
- **Steps Taken:**
  - Performed **Principal Component Analysis (PCA)** to project high-dimensional data into 2D space.
  - Retained maximum variance to highlight cluster separations effectively.

4. Cluster Evaluation

- Evaluated the quality and separation of clusters using the following metrics:
  - **Davies-Bouldin Index:** 0.4498 — indicating excellent separation between clusters.
  - **Silhouette Score:** 0.5499 — indicating moderately strong cluster definitions.

5. Cluster Visualization

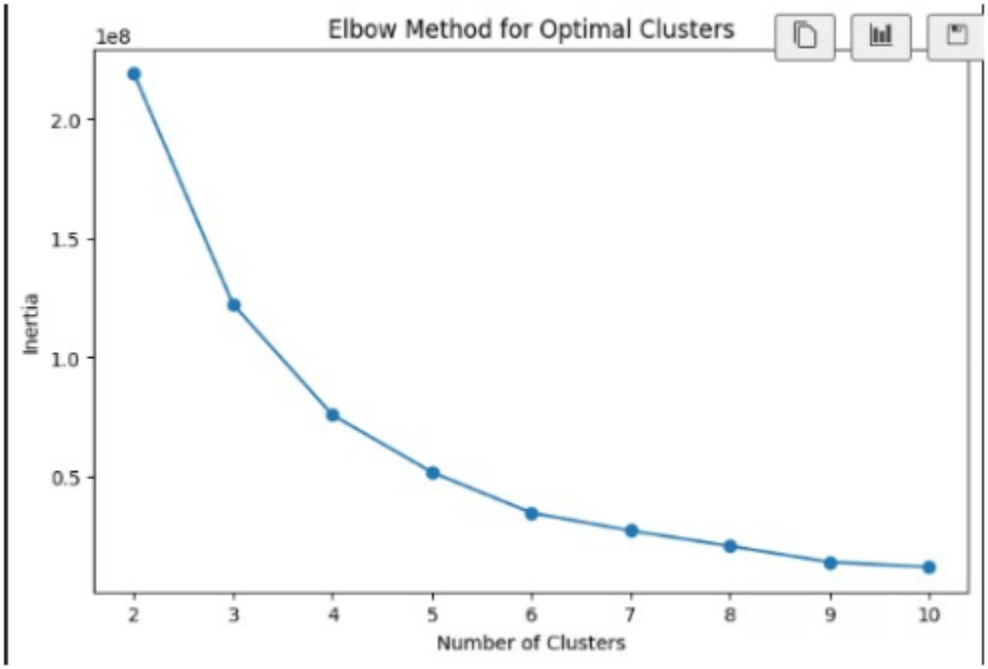
- Visualized clusters in 2D space using PCA projection.

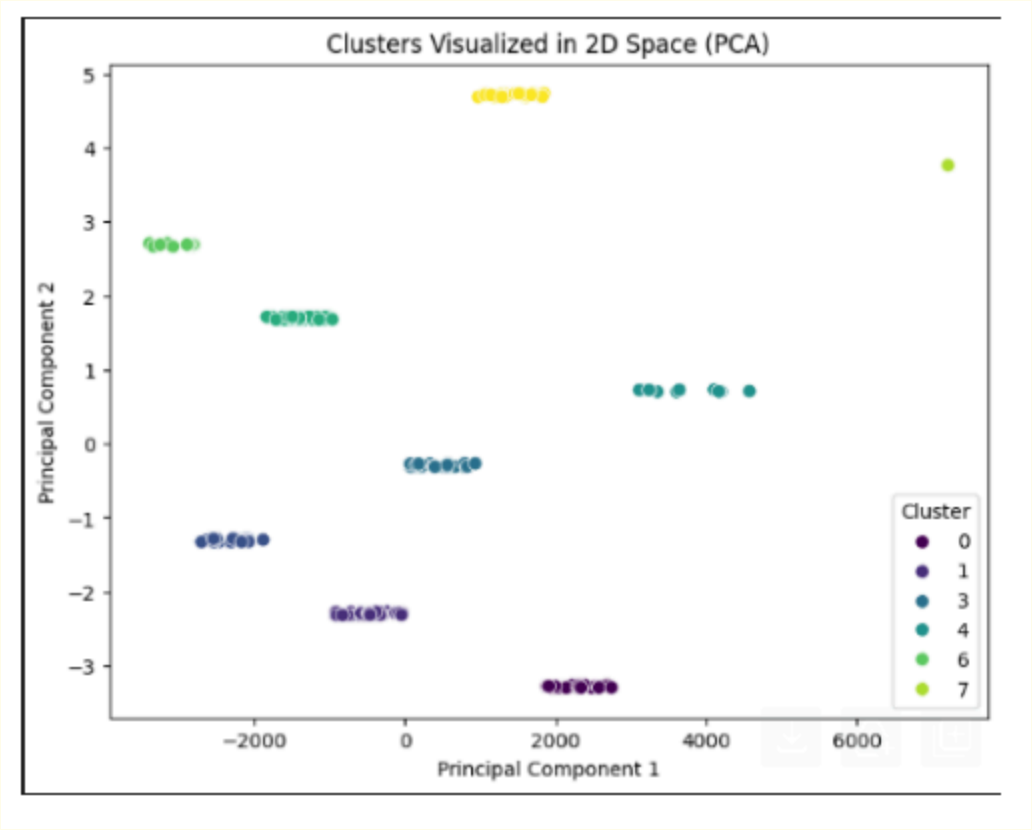
- Highlighted the spatial distribution and separations between clusters, providing a clear understanding of customer segmentation.

Findings and Insights

- Customers were segmented into **9 meaningful groups** based on transaction values and regional patterns.
- These clusters provide actionable insights to design targeted business strategies:
    - **High-value segments** can be approached with premium offerings.
    - **Regional patterns** can inform localized marketing campaigns.

Images:





Conclusion

The clustering analysis successfully segmented customers into well-defined groups. The combination of K-Means clustering and PCA visualization ensured robust insights into customer behavior, enabling data-driven strategies. The metrics confirm the quality of clustering, with clear separations and moderate cluster definitions.

For further details or refinements, please feel free to reach out.