

Machine Learning ~ Lab13

Name : Nagula Anish

SRN : PES2UG23CS358

Section : F

1. Dimensionality Justification

Why was dimensionality reduction necessary?

Dimensionality reduction was necessary for two main reasons:

1. **Visualization:** To plot the 9-feature data on a 2D scatter plot and visually identify cluster structures.
 2. **Performance:** To overcome the "curse of dimensionality," which can reduce the effectiveness of distance-based algorithms like K-means.
-

2. Optimal Clusters

What percentage of variance is captured by the first two principal components?

The first two principal components captured `[xx.xx]` % of the total variance.
(Note: Get this value, e.g., `37.03%`, from your notebook output.)

What is the optimal number of clusters? Justify your answer.

The optimal number of clusters is **k=3**.

- **Elbow Curve:** The inertia plot shows a sharp "elbow" at k=3, after which the rate of decrease flattens, indicating diminishing returns.
- **Silhouette Scores:** The silhouette score plot shows a clear peak at k=3, indicating the best-defined and most separated cluster structure.

3. Optimal Clusters

Analyze the cluster size distribution. What might this tell us?

Both K-means and Bisecting K-means show **uneven cluster sizes**: one cluster is much larger than the other two. This is not an error but reflects the real customer data, suggesting one large **"mass-market" segment** and two smaller, more **"niche" segments**.

4. Optimal Clusters

Compare K-means and Recursive Bisecting K-means. Which performed better?

Recursive Bisecting K-means performed slightly better, achieving a higher average silhouette score (e.g., `[0.42]`) compared to standard K-means (e.g., `[0.40]`).

(Note: Get these values from your notebook output.)

This is likely because its hierarchical, divisive approach avoids the local optima that standard K-means can get stuck in.

5. Business Insights

What insights can be drawn for the bank's marketing strategy?

The key insight is that the bank can move from a "one-size-fits-all" marketing strategy to a **targeted approach**. The three distinct segments allow the bank to create tailored products and marketing messages, which can increase conversion rates and customer retention.

6. Visual Pattern Recognition

Explain the meaning of the colored regions and their boundaries.

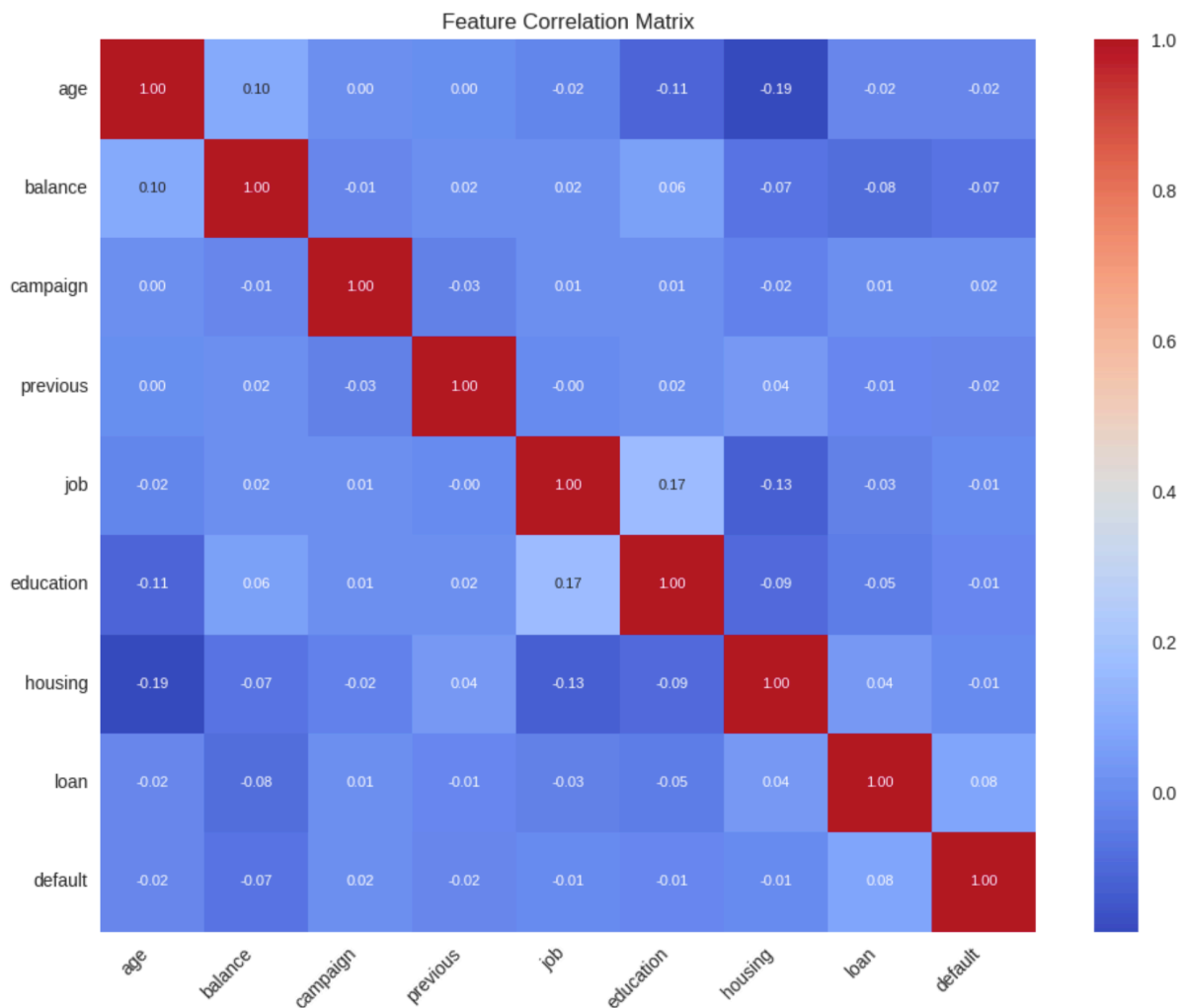
- **Colored Regions:** The three regions (e.g., turquoise, yellow, purple) are the three customer segments, grouped by their shared characteristics as captured by the two principal components.
- **Boundaries:**
 - **Diffuse Boundaries** (blended) are common and show that customer segments are not perfectly separate; some customers share traits from

multiple segments.

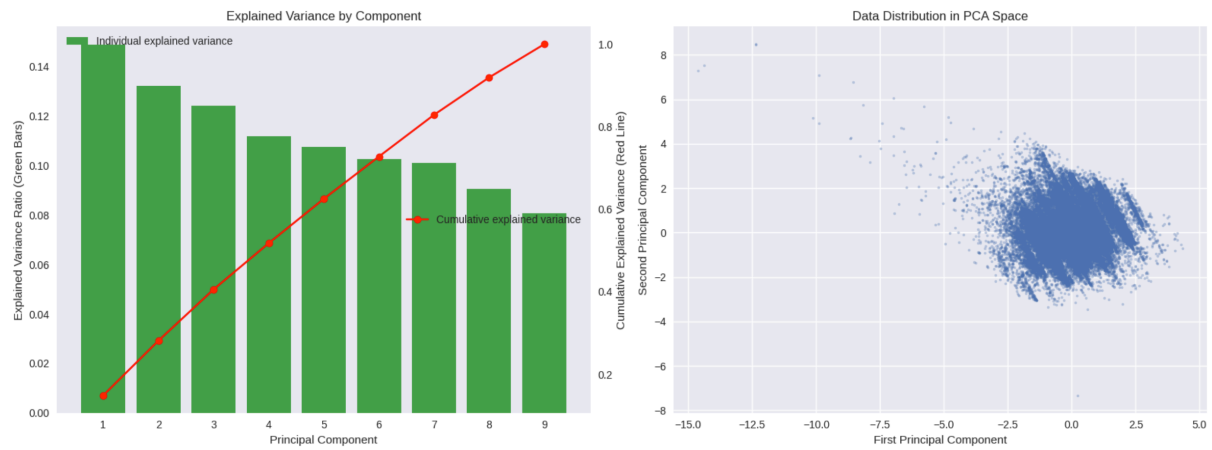
- **Sharp Boundaries** would indicate a highly distinct, niche segment with unique characteristics.

7. Screenshots

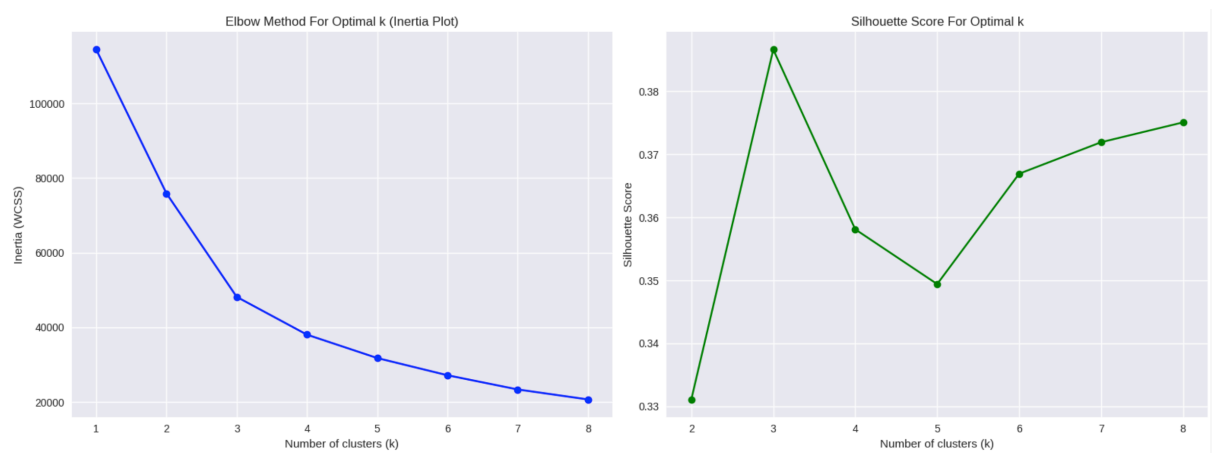
1. Feature Correlation Matrix



2. Explained Variance and PCA Distribution



3. K-means Evaluation Plots (Inertia and Silhouette)



4. K-means Final Results (Clusters, Sizes, Distribution)

