

Clustering Analysis of Vietnamese Banks (Q3 - 2023)

Course: Machine Learning for Business

Lecturer: Mr. Doan

Quarter 3, 2024-2025



TEAM MEMBERS

Trần Phương Thảo
2132300447

Phạm Tuấn Kiệt
2032300089

Thái Gia Huy
2132300332

TABLE OF CONTENT

1. Introduction
2. Data overview
3. Clustering method
4. K-Means – How many bank groups?
5. The bank segments (applied K-Means & PCA)
6. Cluster profiles and interpretation
7. Strategic Recommendations

1. INTRODUCTION

INTRODUCTION

1. Business problem

Vietnamese banks vary significantly in scale, profitability, and operational focus. However, there is currently no standardized segmentation that helps regulators, investors, or the banks themselves understand peer groups effectively.

2. Proposed solution

To address this gap, we propose a clustering-based approach that groups Vietnamese banks based on key financial indicators.

3. Objective

To use clustering techniques on Q3 2023 Income Statement (KQKD) data to group similar banks, understand their unique profiles, and suggest targeted strategic directions.

2. DATA OVERVIEW

DATA OVERVIEW

Data Focus: Analyzed Q3 2023 Income Statement (KQKD) metrics for 27 Vietnamese banks. These 21 key indicators reflect core profitability and operational performance.

Key Preparations:

- Shaping the data (Pivoting): Organized data so each bank was a row, and financial metrics were columns – ready for analysis.
- Ensuring fair comparison (Standardization): Applied StandardScaler so all metrics contributed equally to the grouping process.

3. CLUSTERING METHOD

CLUSTERING METHOD

Main Method:

K-Means Clustering: Primarily used K-Means, an algorithm that sorts banks into a pre-set number of groups by finding natural similarities.

4. K-MEANS

HOW MANY BANK GROUPS?

HOW MANY BANK GROUPS?

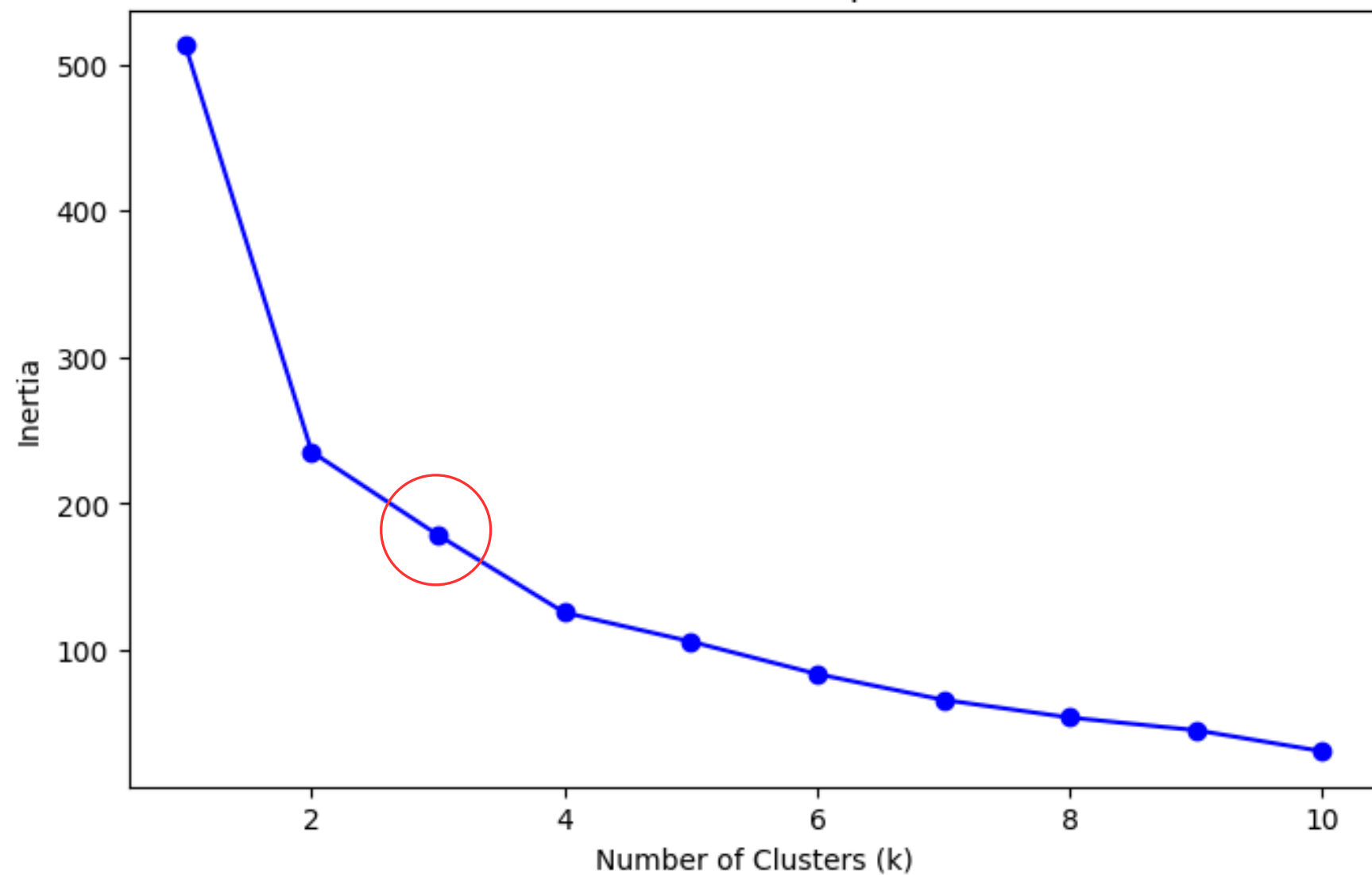
K-Means: Finding the Right Number of Segments ('k')

To determine the best number of groups, we used two standard methods:

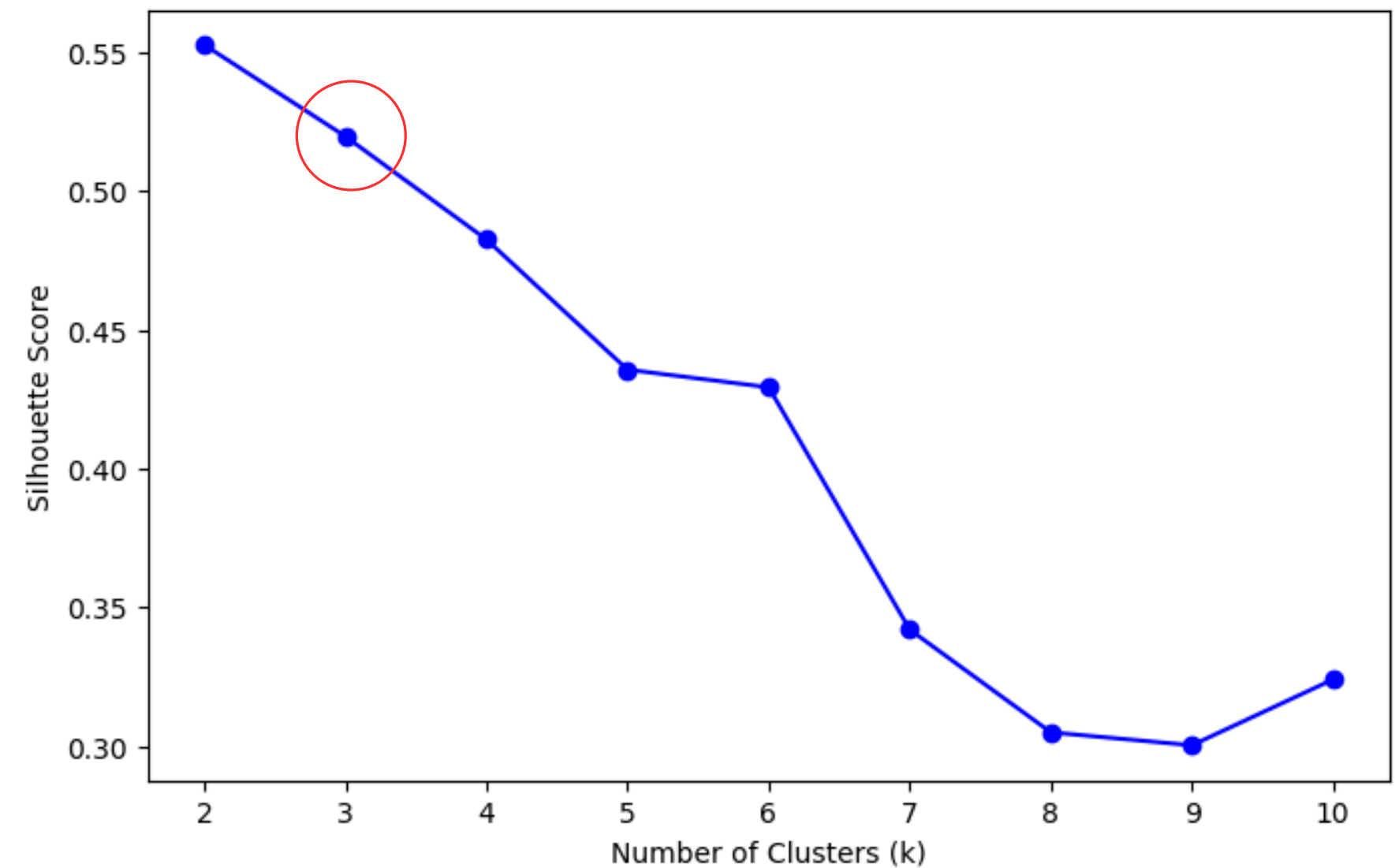
- **Elbow Method:** Looks for the point where adding more groups gives diminishing returns.
- **Silhouette Scores:** Measures how distinct and well-separated the groups are.

HOW MANY BANK GROUPS?

Elbow Method for Optimal k



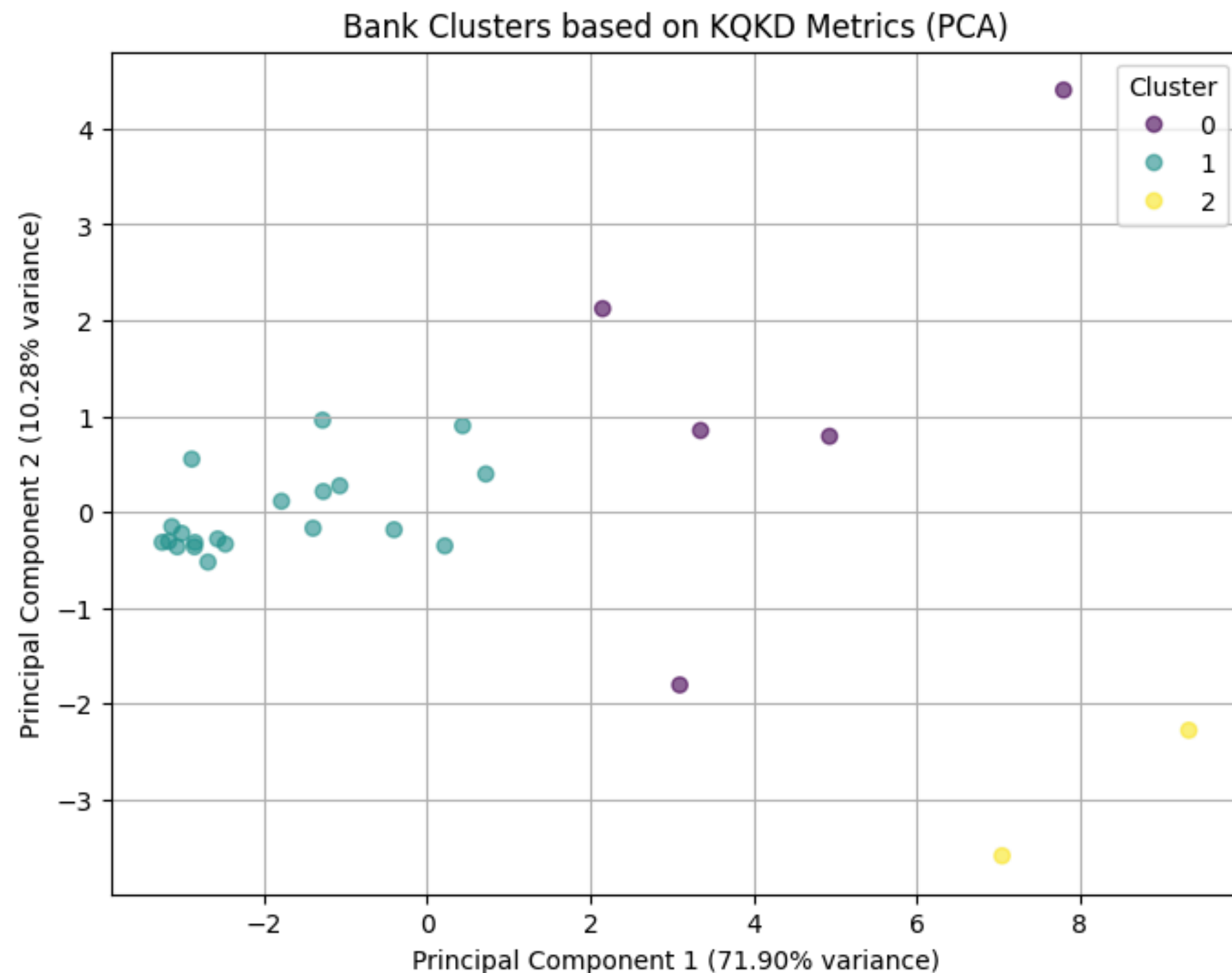
Silhouette Score vs. k



=> k=3 was chosen for the final K-Means model.

5. THE BANK SEGMENTS (APPLIED K-MEANS & PCA)

THE BANK SEGMENTS



27 banks were segmented into 3 distinct clusters.

- **Cluster 0:** 5 banks
- **Cluster 1:** 20 banks
- **Cluster 2:** 2 banks

THE BANK SEGMENTS

Cluster 0: Top-tier



Cluster 1: Moderate performers



Cluster 2: Income generators



6. CLUSTER PROFILES AND INTERPRETATION

CLUSTER PROFILES

Cluster 0: Top-tier



High profitability & income leaders

Number of banks: 5

Key financial traits:

- Highest average Profit After Tax.
- Very strong Total Operating Income & Net Interest Income.
- Moderate Operating Costs & Credit Loss Provisions.

=> Top-tier financial performance

CLUSTER PROFILES

Cluster 1: Moderate performers



Smaller operations

Number of banks: 20 (largest group)

Key financial traits:

- Lowest average Profit After Tax & overall Income.
- Lowest magnitude of Operating Costs & Credit Provisions (consistent with smaller scale or different focus).

=> Represents banks with smaller operational scales

CLUSTER PROFILES

Cluster 2: Income generators



High-risk players

High income, high cost & Provision leaders

Number of banks: 2

Key financial traits:

- Highest Total Operating Income & Net Interest Income (indicating very large operations).
- BUT, also the highest Operating Costs and significantly the highest provisions for Credit Losses.

7. STRATEGIC RECOMMENDATIONS

RECOMMENDATIONS

Cluster 0: Top-tier



Lower Current Risk. Strong financials provide a buffer.

Recommendation:

- Leverage brand & market position.
- Invest in FinTech & digital customer experience.
- Explore prudent diversification (e.g., wealth management).

RECOMMENDATIONS

Cluster 1: Moderate performers



Lower profitability can mean more vulnerability.

Recommendation:

- Identify and dominate specialized market segments.
- Invest in cost-effective technology to improve cost-to-income ratio.
- Develop targeted customer acquisition & retention strategies.

RECOMMENDATIONS

Cluster 2: Income generators



Higher Current Risk. Significant credit risk is evident from high provisions, and cost pressures are notable despite high income.

Recommendation:

- Intensively manage non-performing loans (NPLs) & strengthen credit processes.
- Conduct the review of operating costs for significant savings.



THANK YOU

