



# Multivariate Analysis of Global Development Indicators

# Research Objectives & Questions



## Reduce Dimensionality

Use PCA to reduce dimensions while retaining variance.



## Validate Results

Validate PCA findings using Factor Analysis.



## Cluster Countries

Group countries based on latent economic structures.



## Predict Income Groups

Predict income levels using derived latent variables (SVM).

# Data & Preprocessing

Analyzing 10 global development indicators across 265 countries from 2000–2022.

Data from World Bank – World Development Indicators (WDI), resulting in 6095 country-year observations.

## Indicators

GDP, population, GDP per capita, inflation, unemployment, trade, government spending, current account, and FDI.

## Preprocessing

- Median imputation for missing values.
- Outliers retained as economically meaningful.

## Standardization

All variables were standardized before analysis for a consistent scale.

Data reshaped to country-year (2000-2022). Sample-to-feature ratio: 609.5:1 for robust PCA.

# Why Principal Component Analysis (PCA)?

## → **Interrelated Indicators**

Global development indicators are strongly interrelated (e.g., GDP, population, trade, income).

## → **Avoids Multicollinearity**

Using raw indicators would cause multicollinearity and unstable analysis.

## → **Extracts Latent Dimensions**

PCA extracts independent latent economic dimensions from correlated indicators.

## → **Reflects Real Structure**

These latent dimensions reflect real economic structure, not individual metrics.

PCA was essential to uncover the true economic dimensions underlying global development and to support reliable downstream analysis.

## **PCA also enables the following:**



Country clustering



Income group classification



Enables economic interpretation of results, not just prediction

# PCA Results & Interpretation



## PC1: Global Integration

Trade openness and interconnectedness.



## PC2: Economic Scale

Size and overall development level.



## PC3: Macroeconomic Stress

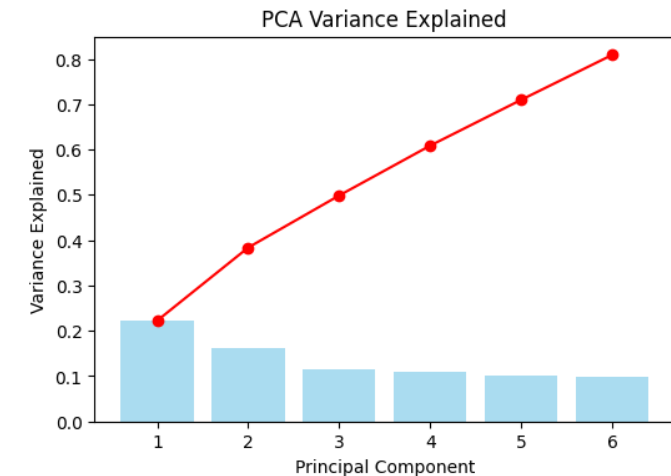
Indicators of instability and volatility.



## PC4: Government Role

Extent of government economic involvement.

Indicator	PC1	PC2	PC3	PC4
GDP	-0.04	0.45	-0.09	-0.09
Population	0.09	0.43	0.06	-0.04
GDP_per_capita	-0.13	0.24	-0.35	-0.14
Inflation	-0.05	-0.08	0.08	-0.03
Unemployment	0.18	-0.18	0.42	0.42
Exports_pctGDP	0.48	-0.06	0.16	-0.09
Imports_pctGDP	0.47	0.01	0.19	-0.05
Gov_Expense_pctGDP	0.36	-0.01	-0.06	0.67
Current_Account_pctGDP	-0.26	-0.07	-0.47	-0.14
FDI_pctGDP	0.32	0.14	0.17	-0.23



Six principal components explain 80 % of the total variance. The first four components explain 61% and are used for classification.

**Key Understanding: Economic development is fundamentally multidimensional.**

# Country Clustering Analysis



## Trade-Oriented Middle-Income Economies

Middle-income economies with high trade activity on a global scale.



## Small High-Income Economies

Small populations, wealthy economies, and a focus on openness and globalization.



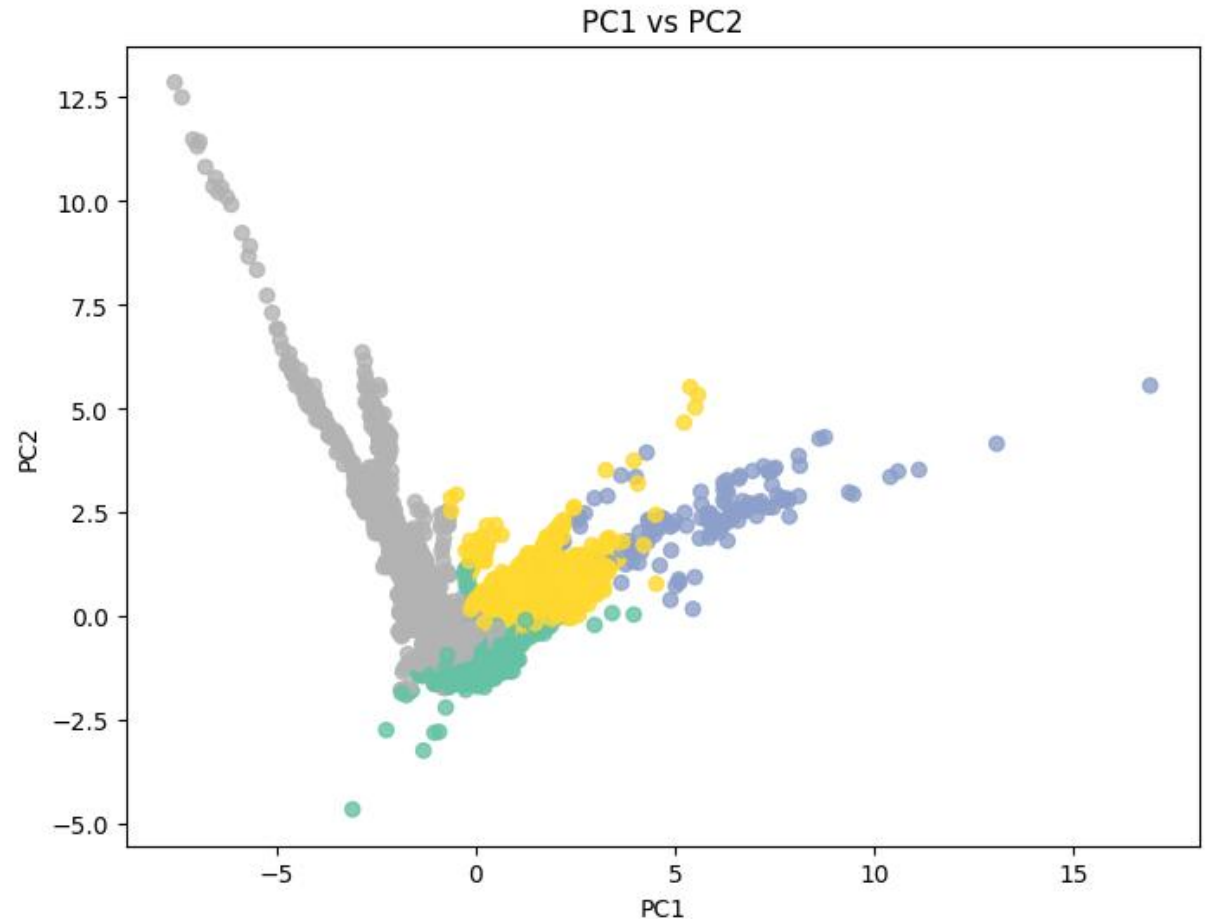
## Large Developing Economies

Highly populated countries with lower income levels and moderate engagement in international trade



## Very Large Middle-Income Economies

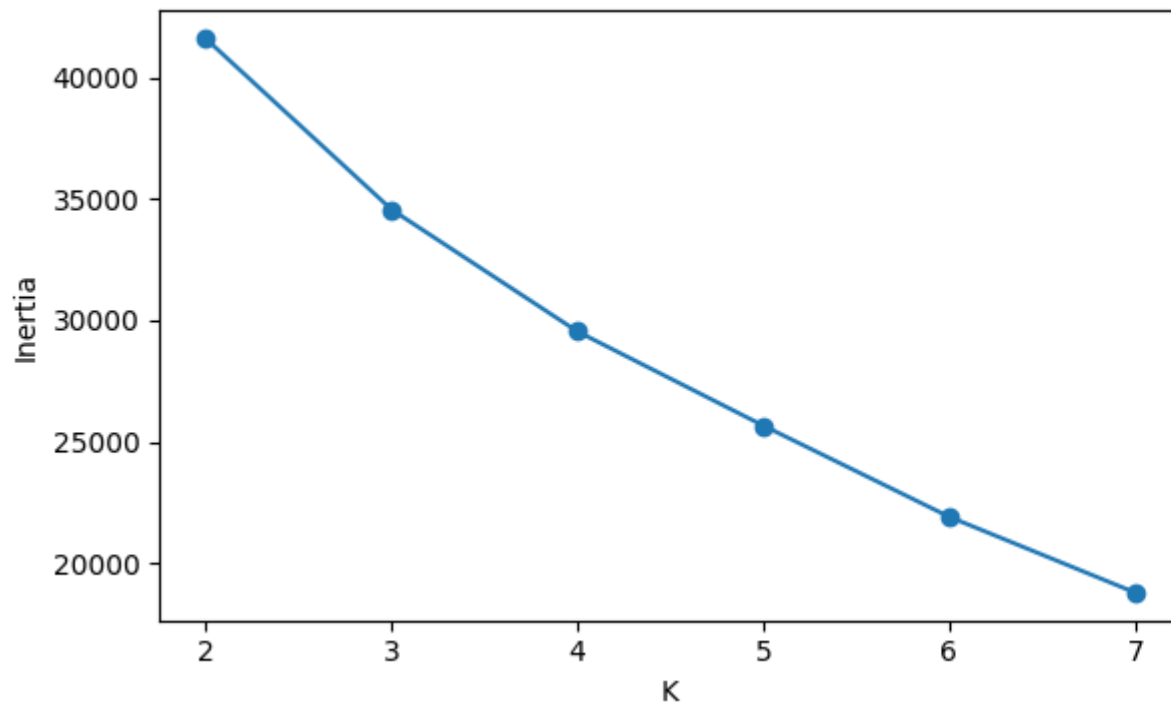
Large economies with mid-range income levels, balanced trade, and relatively higher inflation.



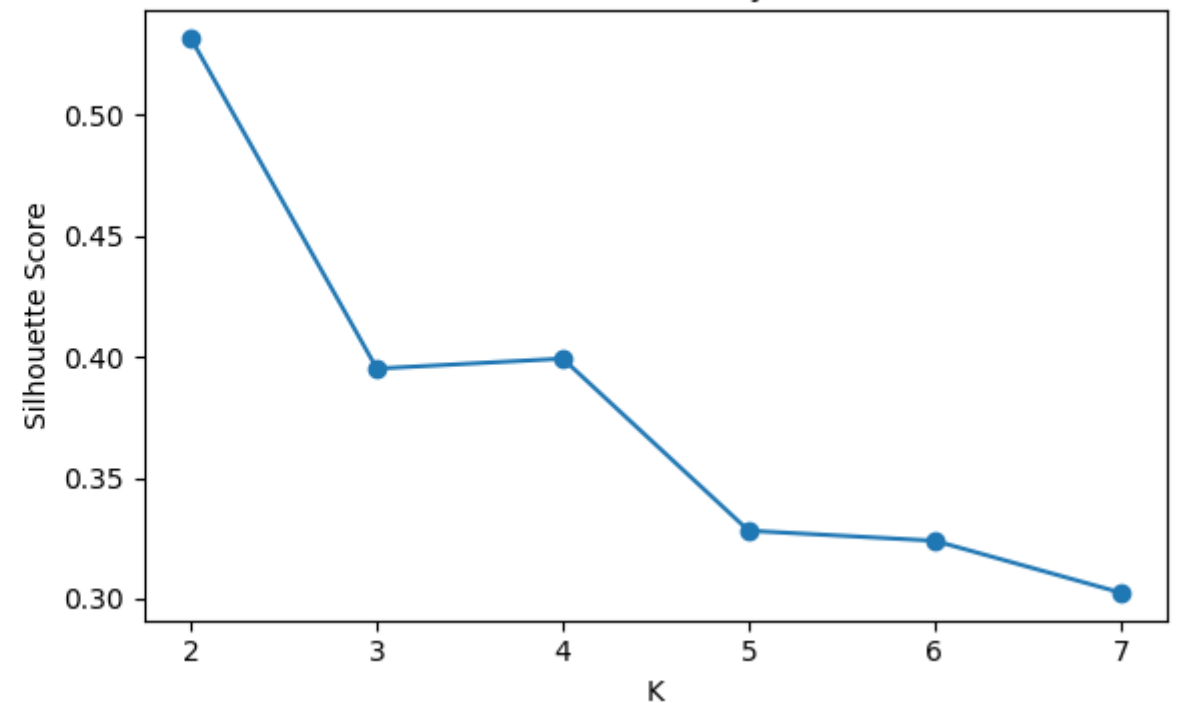
# Choice of Number of Clusters

K-Means clustering applied in PCA space. **Optimal clusters: 4**, stable and economically interpretable.

Elbow Method

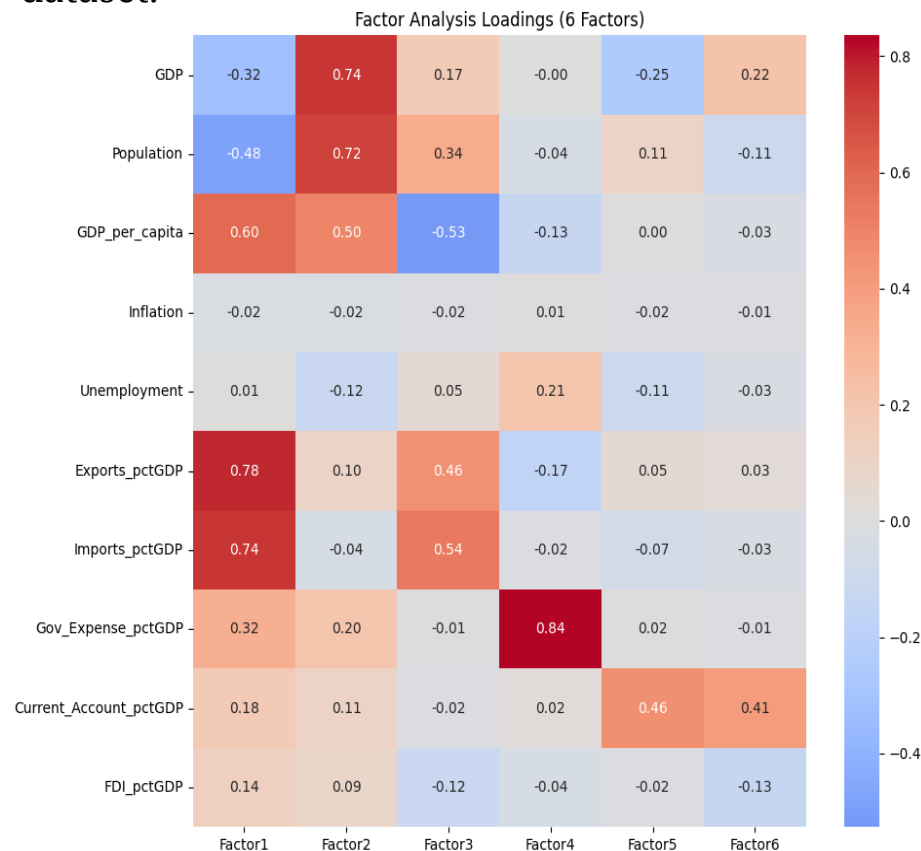


Silhouette Analysis



# Validation with Factor Analysis

- **Purpose:** Both PCA and Factor Analysis (FA) were applied to the same set of 10 development indicators to uncover the underlying structure in the data.
- **Method:** Despite using different mathematical approaches, both methods revealed three major latent patterns in the dataset.



## Overall Conclusion

PCA and Factor Analysis both reveal the same core structure in the data:

1. **Trade intensity**
2. **Economic size**
3. **Income and macroeconomic stability**

Some amount of agreement between the two methods increases confidence that these are real, meaningful latent dimensions in global development patterns.

Factor Analysis Loadings (first few rows)

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
GDP	-0.319691	0.740521	0.167613	-0.000306	-0.246316	0.217561
Population	-0.479693	0.716145	0.339041	-0.037839	0.109876	-0.109879
GDP_per_capita	0.601270	0.499711	-0.526784	-0.131476	0.001158	-0.030397
Inflation	-0.024493	-0.024991	-0.015491	0.011367	-0.019729	-0.012706
Unemployment	0.013122	-0.122327	0.049552	0.206012	-0.106777	-0.029944

PCA Loadings (first few rows)

	PC1	PC2	PC3	PC4	PC5	PC6
GDP	-0.302049	0.615489	0.161652	0.001603	-0.100130	0.000806
Population	-0.360474	0.549869	0.167235	-0.120539	-0.134184	-0.001558
GDP_per_capita	0.278227	0.325664	-0.341965	0.368883	0.072905	0.037452
Inflation	-0.026583	-0.069063	0.114660	0.160472	-0.186893	0.959296
Unemployment	0.046433	-0.186103	0.658182	0.359412	0.011726	-0.144647



# Income Group Classification (SVM)



## Method

Support Vector Machine (SVM) with RBF kernel.



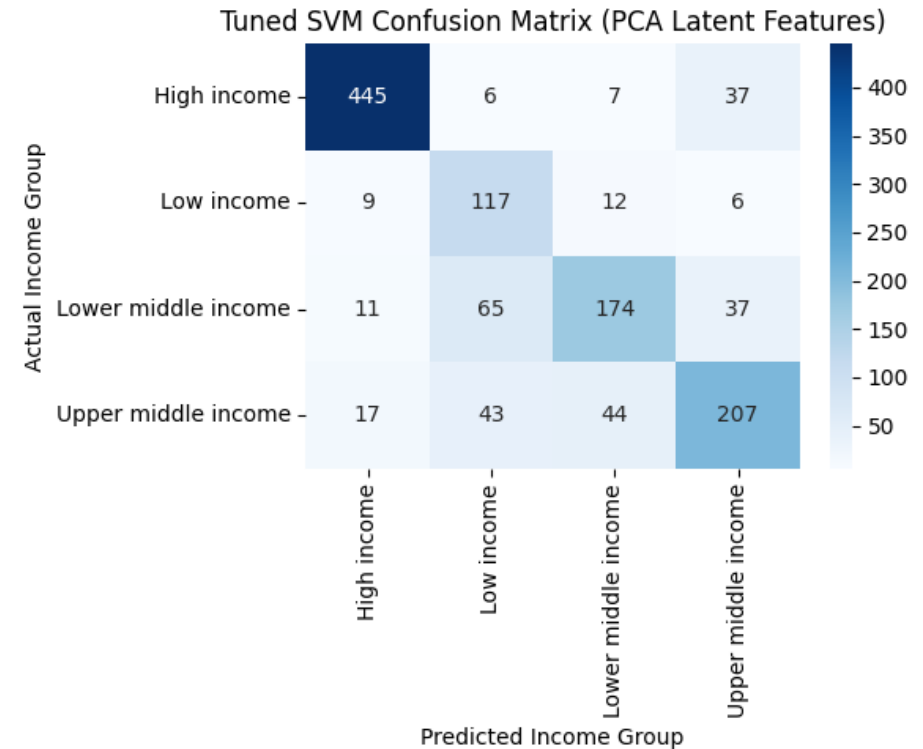
## Pipeline

StandardScaler used for SVM. Class imbalance handled with weighting.



## Tuning

Hyperparameters tuned via GridSearchCV with 5-fold cross-validation.



Base Accuracy 63%

Tuned Accuracy 76%

# Conclusions & Takeaways

## Key Dimensions

Reduced 10 indicators to 6 latent dimensions, explaining 80% variance.

## Validated Findings

PCA results confirmed by Factor Analysis and regression.

## Meaningful Clusters

Identified four distinct country clusters.

## Strong Prediction

Achieved robust income classification performance.

Development patterns are multidimensional and best captured by latent variables

PCA extracts these dimensions, enabling more meaningful country classification than any single indicator.