

Banking Client Segmentation & Persona Analysis Using K-Medoids Clustering

OBJECTIVE: TO SEGMENT BANK CLIENTS USING CLUSTERING TECHNIQUES AND DEVELOP CLIENT PERSONAS TO ENABLE INFORMED AND EFFECTIVE TARGETING STRATEGIES.

1. DATA

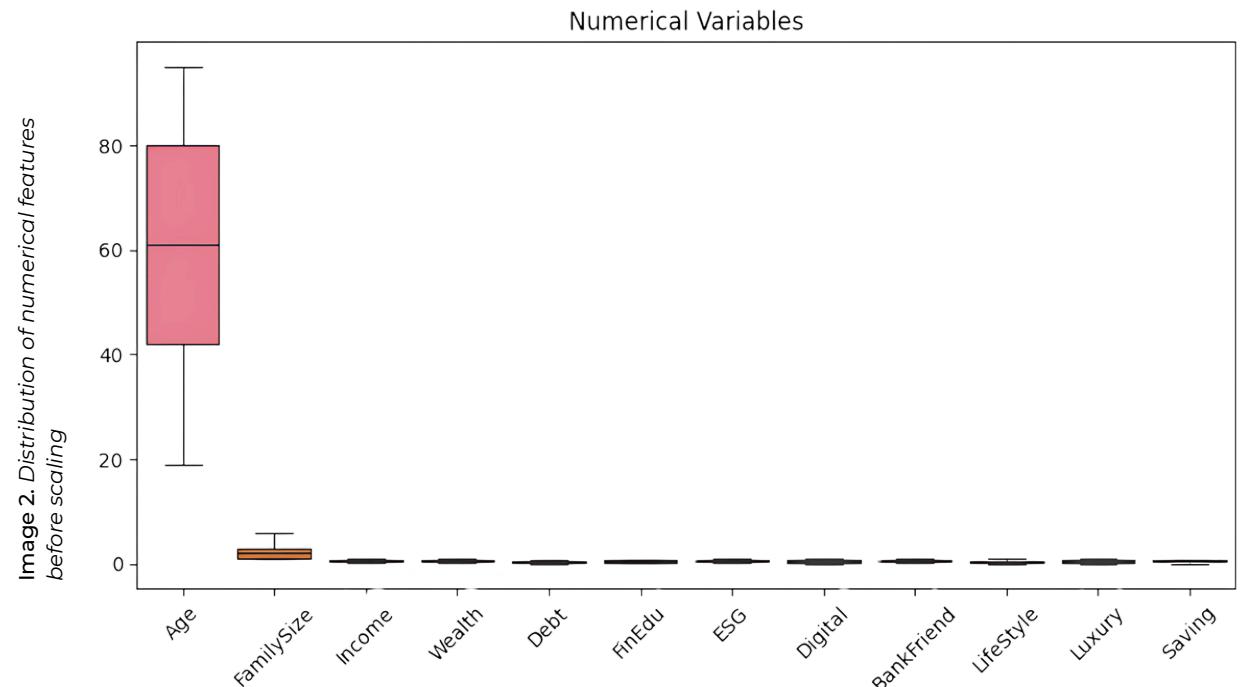
SOURCE: DATASET OF BANK CLIENTS WITH FEATURES INCLUDING BOTH NUMERICAL AND CATEGORICAL VARIABLES.

ID	Age	Gender	Job	Area	CitySize
DROPPED	Integer	Binary/Integer	Categorical/Integer	Categorical/Integer	Categorical/Integer
FamilySize	Income	Wealth	Debt	FinEdu	ESG
Integer	Continuous/Float	Continuous/Float	Continuous/Float	Continuous/Float	Continuous/Float
Digital	BankFriend	LifeStyle	Luxury	Saving	Investments
Continuous/Float	Continuous/Float	Continuous/Float	Continuous/Float	Continuous/Float	Categorical/Integer

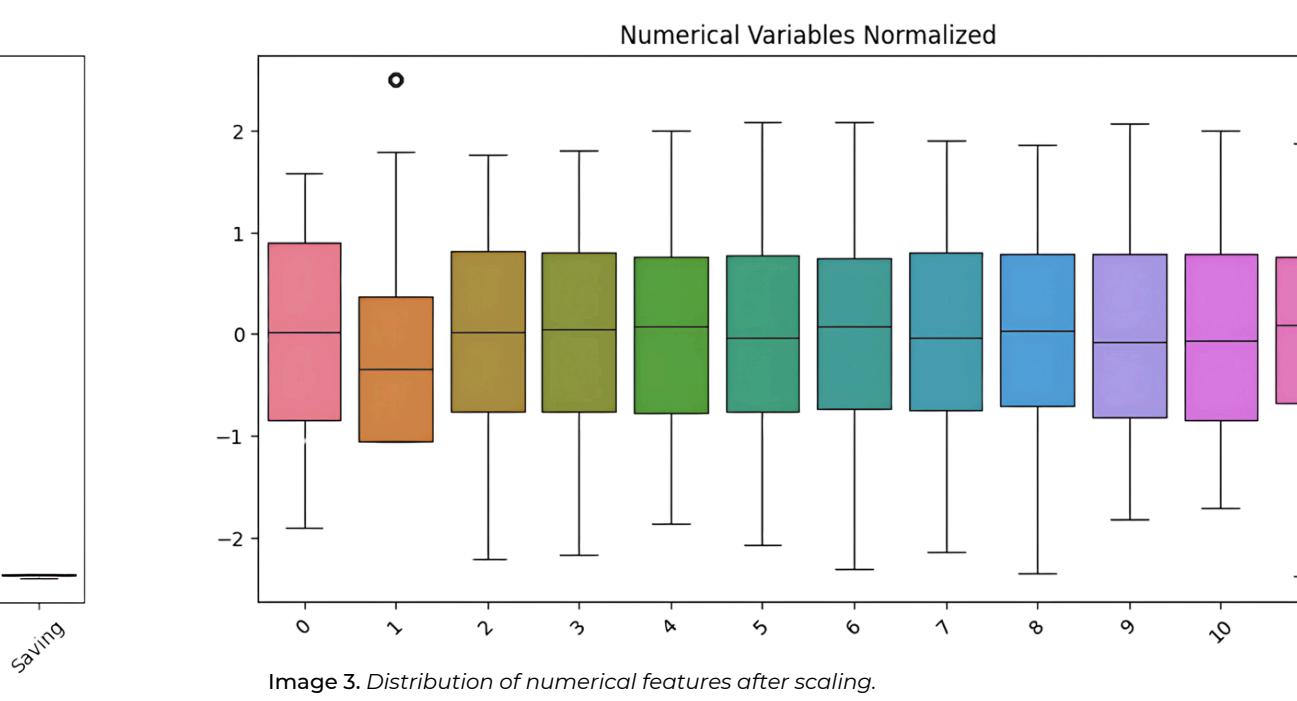
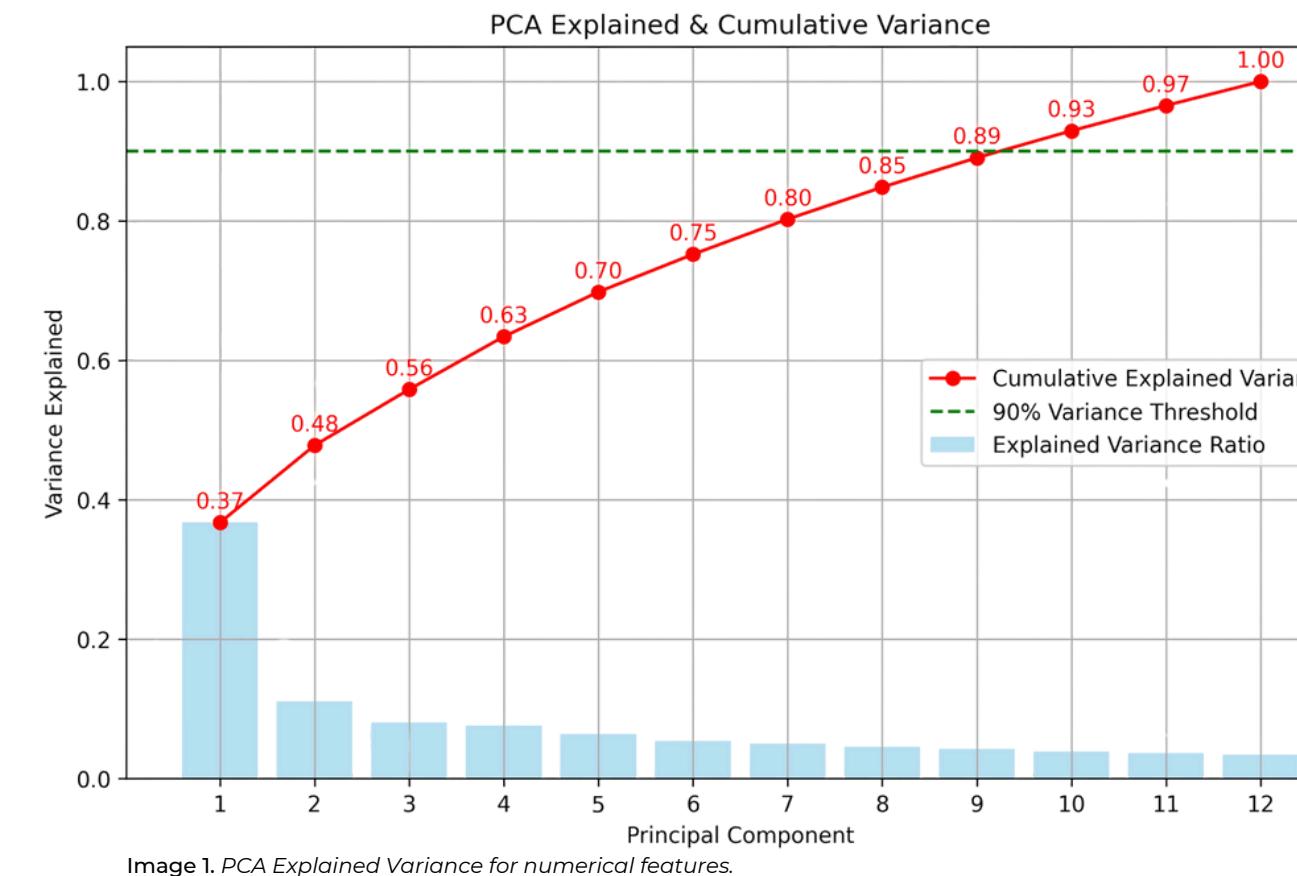
Image 1. Source Data Schema. For more details about the data, check the data file "Dataset1_BankClients.xlsx" and the metadata file "BankClients_MetaData.xlsx" in the source code repo.

2. PREPROCESSING

- FEATURE SELECTION:** SELECTING MEANINGFUL COLUMNS AS FEATURES
- OUTLIER TREATMENT:** WINSORIZATION (CLIPPING EXTREME 1% VALUES).
- FEATURE REDUCTION:** CORRELATION-BASED REDUCTION.
- SCALING & DIMENSIONALITY REDUCTION:** STANDARD SCALER FOLLOWED BY PCA (RETAINING $\geq 90\%$ EXPLAINED VARIANCE).
- CATEGORICAL ENCODING:** ONE-HOT ENCODING FOR CATEGORICAL FEATURES.
- ADDITIONAL FEATURE GROUPS:** PROFESSIONAL PROFILING AND FINANCIAL EDUCATION FEATURES TO BE PROCESSED SEPARATELY WITH DIFFERENT WEIGHTS.



PCA HAS BEEN APPLIED TO THE NUMERICAL FEATURES OUT OF WHICH IT RETAINED A VARIANCE OF **0.929...**



3. CLUSTERING

WHY K-MEDOIDS?

IT IS A ROBUST CLUSTERING METHOD THAT MINIMISES THE SUM OF DISSIMILARITIES BETWEEN POINTS AND THEIR MEDOID. UNLIKE K-MEANS, IT WORKS WELL WITH ARBITRARY DISTANCE METRICS AND IS LESS SENSITIVE TO OUTLIERS.

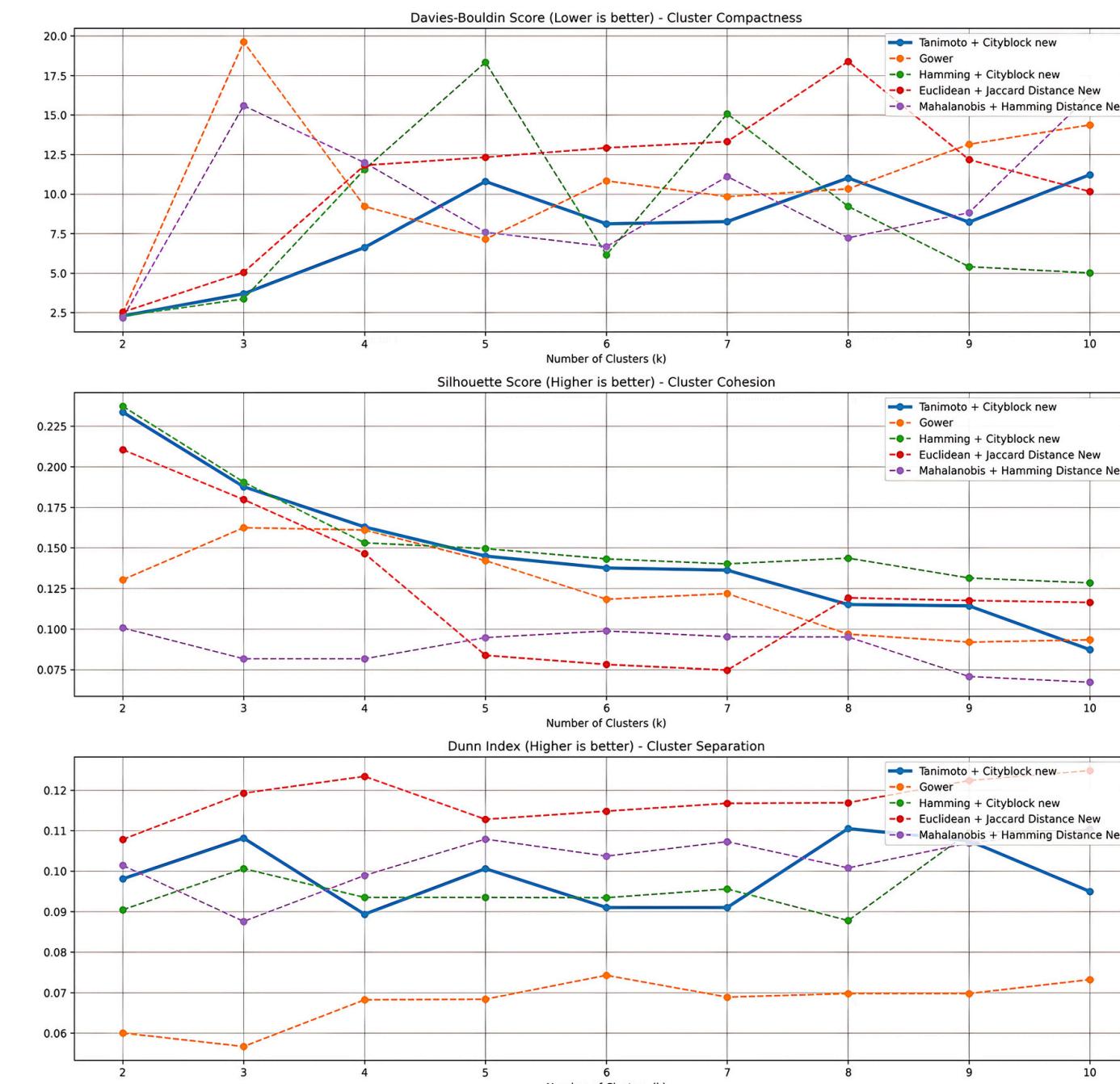


Image 4. PCA Explained Variance for numerical features.

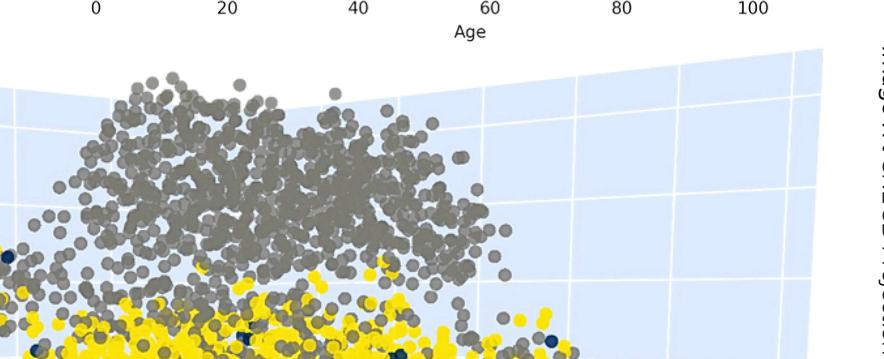
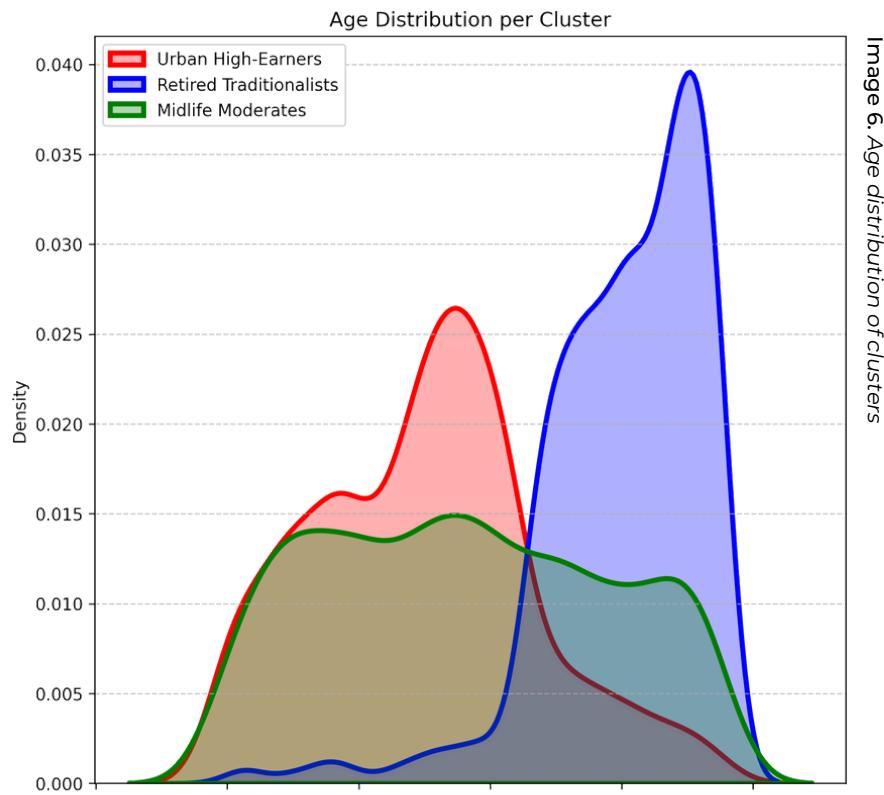
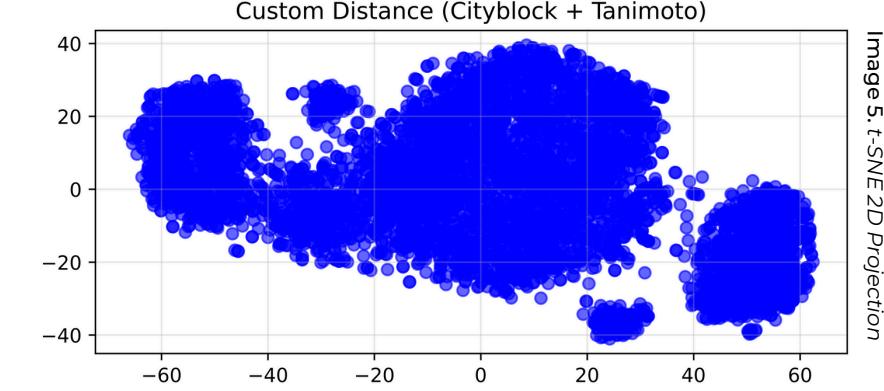
DISTANCE METRIC: TANIMOTO + CITYBLOCK NEW
SEVERAL DISTANCE METRICS, BOTH STANDARD AND CUSTOM, WERE EXPLORED.

WE SETTLED FOR THE CUSTOM "**TANIMOTO + CITYBLOCK NEW**" METRIC, WHICH:

- MEASURES NUMERICAL DIFFERENCES USING THE **CITYBLOCK** (MANHATTAN) DISTANCE.
- HANDLES CATEGORICAL FEATURES VIA THE **TANIMOTO** (JACCARD) DISTANCE.
- LEVERAGES TWO SEPARATE GROUPS OF NUMERICAL FEATURES WITH DIFFERENT WEIGHTS YIELDING WELL-SEPARATED AND COMPACT CLUSTERS.

OPTIMAL K

THE PERFORMANCE METRICS INDICATE THAT A **THREE-CLUSTER SOLUTION** IS OPTIMAL, AS IT MAXIMISES THE SEPARATION BETWEEN CLUSTERS WHILE MINIMISING THE VARIANCE WITHIN EACH CLUSTER. THIS OUTCOME SUGGESTS THAT OUR CLIENT DATA NATURALLY GROUPS INTO THREE DISTINCT SEGMENTS



4. RESULTS



URBAN HIGH-EARNER

PROFESSIONALS AROUND 49 YEARS OLD LIVING IN LARGE CITIES IN THE NORD REGION, TYPICALLY IN SMALL HOUSEHOLDS. THEY ENJOY HIGH INCOME AND WEALTH, STRONG FINANCIAL LITERACY, AND HIGH DIGITAL ADOPTION, WITH ABOVE-AVERAGE SPENDING ON LIFESTYLE AND LUXURY. THEIR FINANCIAL NEEDS ARE BEST MET BY DIGITAL OR HYBRID BANKING SOLUTIONS OFFERING ADVANCED WEALTH MANAGEMENT AND PREMIUM SERVICES.

MIDLIFE MODERATE



PROFESSIONALS AROUND 56 YEARS OLD LIVING IN MEDIUM-SIZED CITIES IN THE NORD REGION, TYPICALLY IN SLIGHTLY LARGER HOUSEHOLDS. THEY EXHIBIT MODERATE INCOME AND WEALTH, BALANCED FINANCIAL LITERACY, AND AVERAGE DIGITAL ADOPTION, SUGGESTING A NEED FOR A HYBRID SERVICE MODEL THAT COMBINES DIGITAL CONVENIENCE WITH PERSONALISED IN-PERSON ADVICE.

RETIRED TRADITIONALIST

