WORKFLOW/ARCHITECTURE DIAGRAM

Raw Dataset (categorical + numeric)

Data Cleaning
- Handle nulls
- Remove duplicates
- Normalize categorical formatting

Feature Engineering / Encoding
- OneHotEncoder → nominal categories
- OrdinalEncoder → ordered categories
- LabelEncoder → label-like single features

Numeric Scaling (StandardScaler)

Feature Combination
Pandas concat / NumPy hstack

PCA Dimensionality Reduction (1D)
- Reduce all features → 1 latent axis
- Remove noise & multi-collinearity

KMeans Clustering
- Train on PCA(1D) feature
- Generate cluster labels

Save Artifacts with Pickle
- PCA (fitted)
- KMeans model

Model Testing
- Load pickled objects
- Transform test dataset (no refitting)
- PCA → KMeans → cluster assignments

Visualisation
- PCA 1D scatter/stripplot
- Cluster distribution