DeFi Portfolio Clustering & Analysis Assignment

MSc Banking & Finance

Due: Before Week 8 or Week 9 Session

Assignment Overview

Apply dimensionality reduction and clustering techniques to real financial data to construct and analyze cluster-based investment portfolios. **This is an individual assignemnt!!**

Total Points: 100 (+ 30 bonus points)
Format: Jupyter notebook with code, analysis, and visualizations
Due Date: Before Week 8 Session

1 Part 1: Real Data Application (50 points)

1.1 Data Collection (15 points)

- Download historical prices for minimum 50 assets/tokens
- Time period: minimum 24 months of daily data
- Required features: prices
- Optional features : volume, market cap, volatility
- Document data sources and preprocessing steps

1.2 Principal Component Analysis (15 points)

- Apply PCA to your feature matrix
- Determine optimal number of components (explain variance threshold)
- Visualize explained variance ratio
- Interpret the principal components

1.3 K-Means Clustering (20 points)

- Implement K-Means on PCA-transformed data
- Use elbow method and silhouette score to determine optimal k
- Assign assets/tokens to clusters
- Visualize clusters in 2D/3D space
- Document cluster characteristics and asset/token compositions

2 Part 2: Portfolio Construction (30 points)

2.1 Portfolio Design (10 points)

- Construct equal-weighted or optimized portfolio from each cluster
- Document portfolio composition and weights
- Explain rationale for weight allocation

2.2 Performance Metrics (15 points)

Calculate and report:

- Cumulative returns
- Sharpe ratio
- Maximum drawdown
- Volatility (annualized)
- Alpha and Beta (if applicable)

2.3 Visualizations (5 points)

- Portfolio equity curves
- Returns distribution
- Risk-return scatter plot
- Cluster composition breakdown

3 Part 3: Written Analysis (20 points)

3.1 Results Interpretation (8 points)

- Explain what each cluster represents economically
- Discuss which clusters performed best/worst and why
- Identify patterns in asset/token characteristics within clusters

3.2 Investment Recommendations (7 points)

- Provide actionable investment insights based on your analysis
- Discuss risk-adjusted performance across clusters
- Suggest portfolio allocation strategies

3.3 Limitations & Caveats (5 points)

- Discuss data quality issues
- Address model assumptions and violations
- Mention market conditions and survivorship bias
- Suggest improvements for future analysis

4 Submission Requirements

4.1 Format

Jupyter notebook (.ipynb) with:

- Clear markdown sections following assignment structure
- Well-commented Python code
- Inline visualizations
- Professional presentation quality

4.2 Code Requirements

- Reproducible (set random seeds)
- Modular functions for key operations
- Error handling where appropriate
- Libraries: pandas, numpy, scikit-learn, matplotlib/seaborn, yfinance/ccxt

4.3 Documentation

- README section with data sources and setup instructions
- Comments explaining complex operations
- References for methods and formulas used

Component	Points	Criteria
Data Quality & Preprocessing	15	Completeness, handling missing data, feature engineering
PCA Implementation	15	Correct application, interpretation, visualization
Clustering Analysis	20	Method selection, parameter tuning, cluster validation
Portfolio Construction	10	Sound methodology, clear documentation
Performance Metrics	15	Accurate calculations, comprehensive metrics
Visualizations	5	Professional, informative, properly labeled
Written Analysis	20	Depth of insight, clarity, actionable recommendations

5 Evaluation Rubric

6 Academic Integrity

- All code must be your own or properly attributed
- Collaboration is permitted but submissions must be individual
- Cite all data sources, libraries, and external references

Submission Deadline: Before Week 8 Session