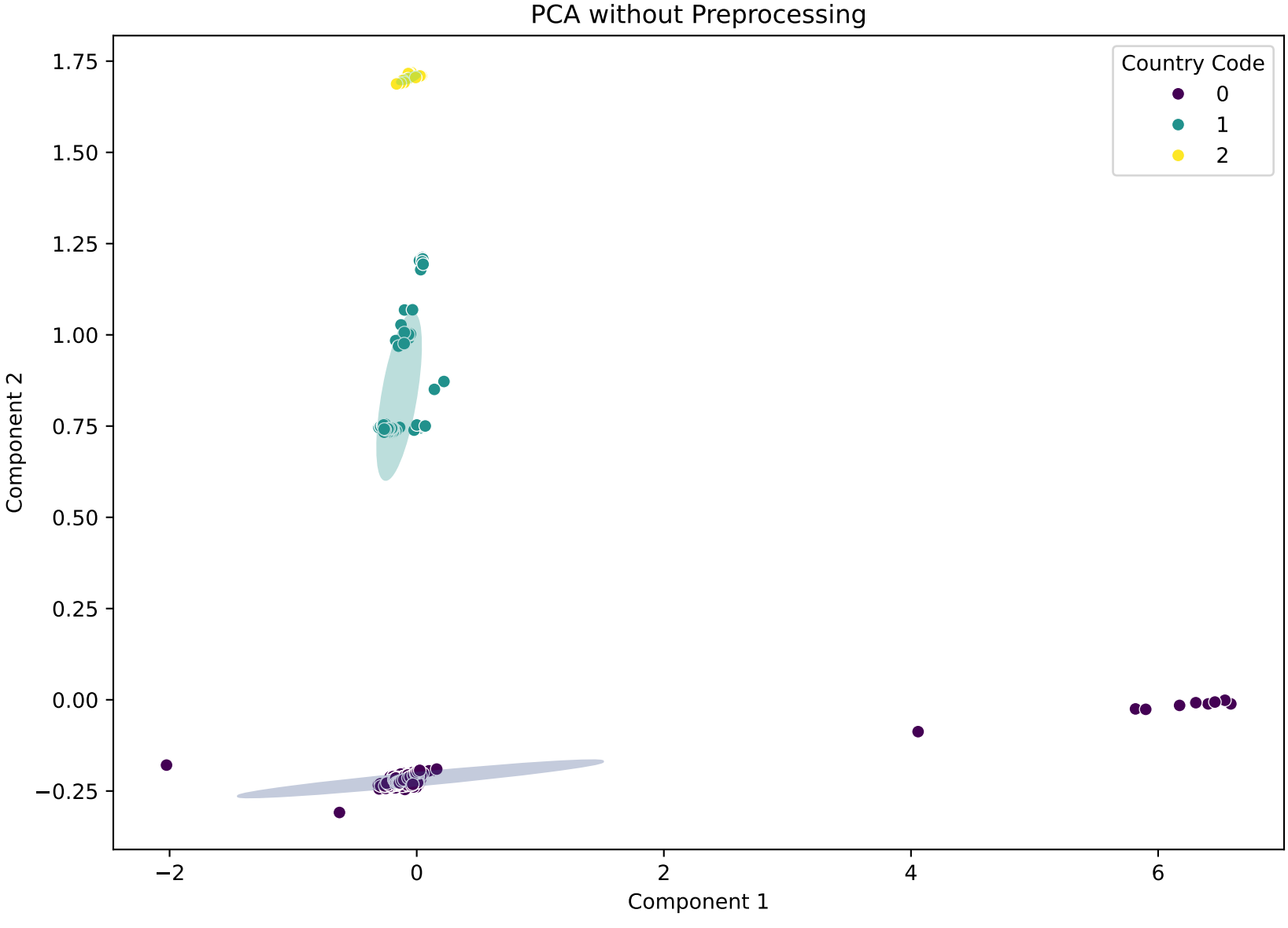
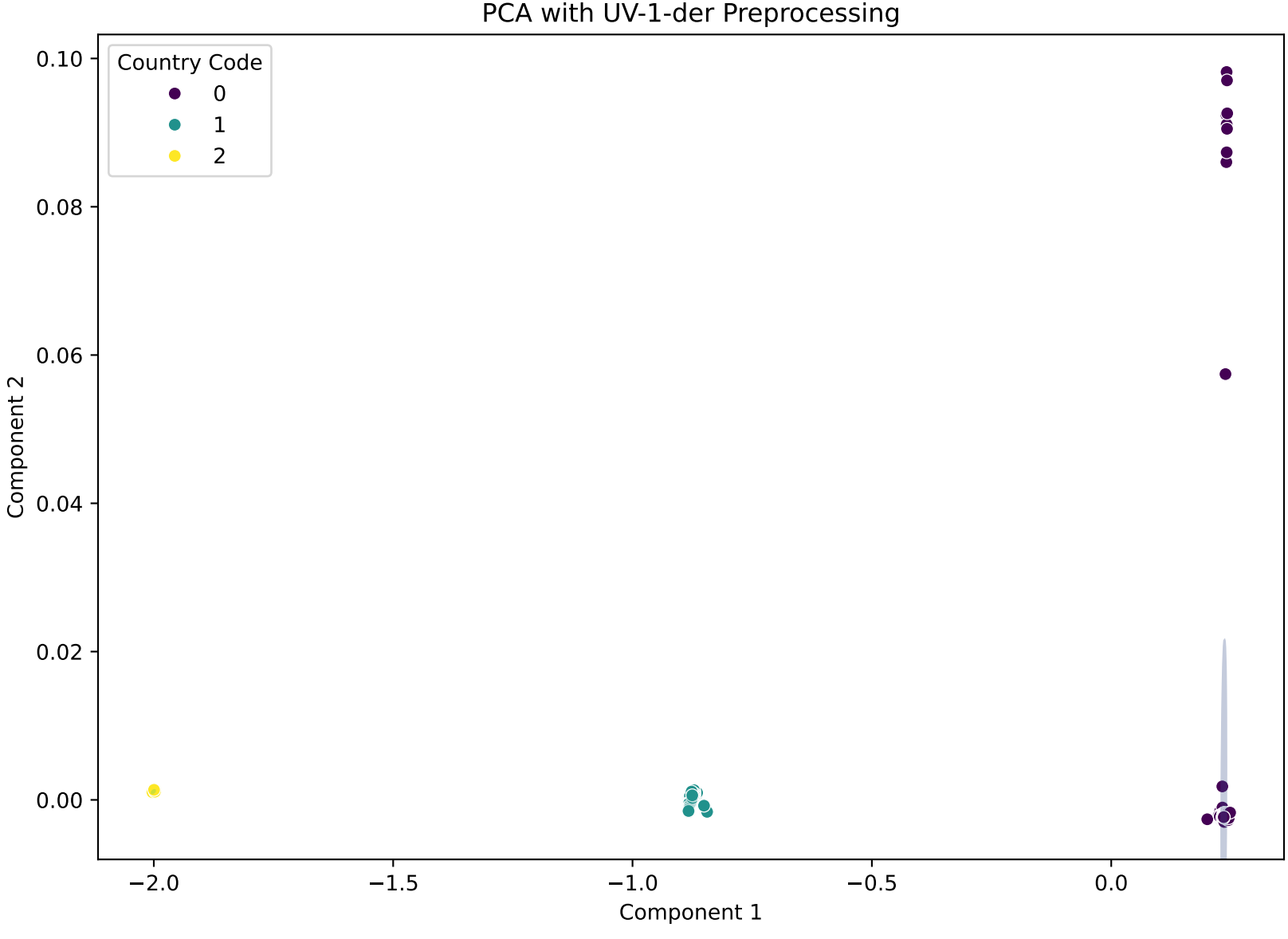
Clustering Analysis Report

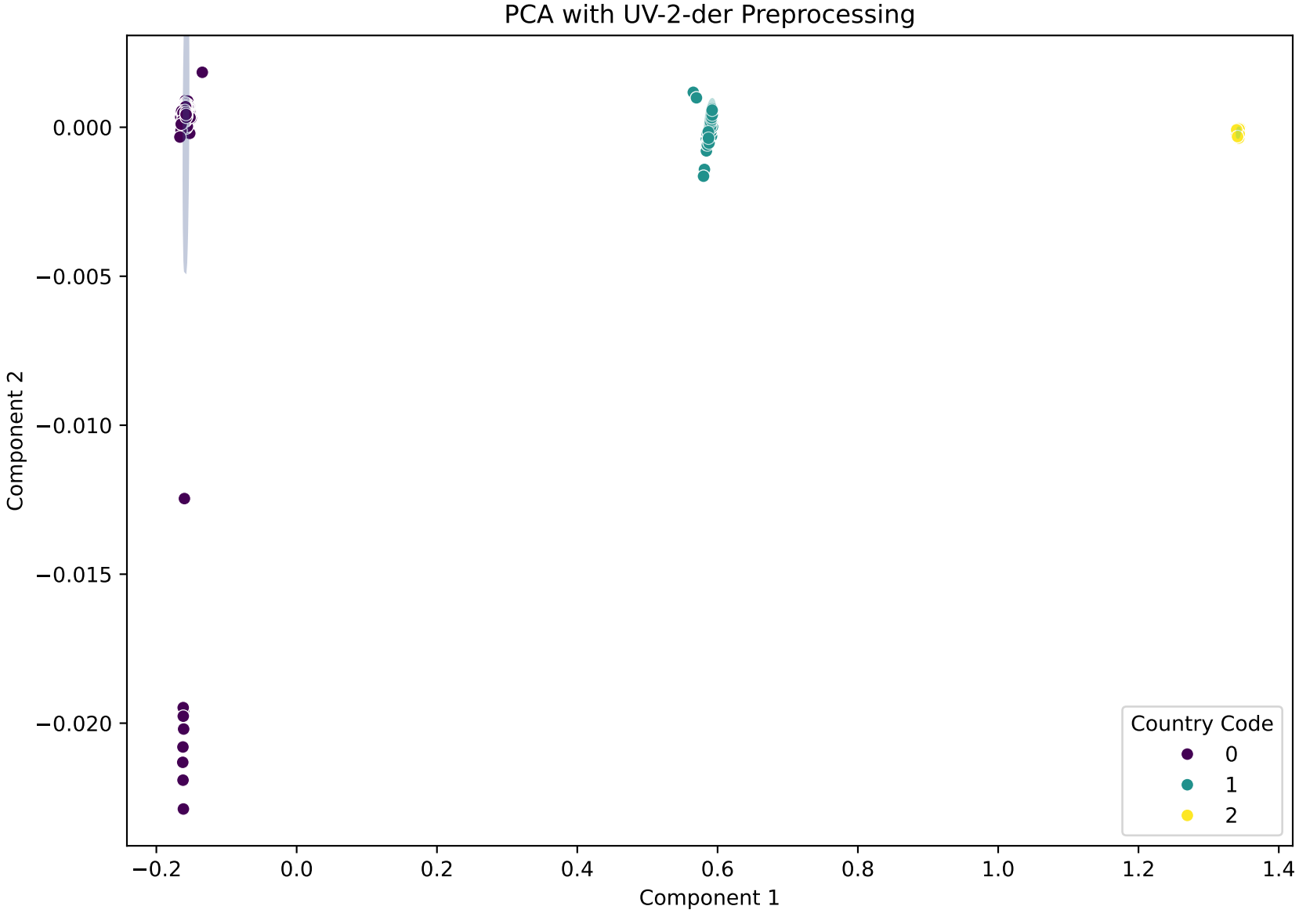
This report evaluates different preprocessing methods and their effectiveness in clustering FTIR spectral data.

Evaluation Metrics

Preprocessing Method	Explained Variance Ratio	Silhouette Score
UV-SNV	[0.7442217 0.2557783]	-0.09591964002830583
UV-MSC	[0.7442217 0.2557783]	-0.09591964002830554
UV-1-der	[9.99381129e-01 6.18871257e-04]	0.9935890147687199
UV-2-der	[9.99929423e-01 7.05768906e-05]	0.9957407692057924
Par-MSC	[0.7442217 0.2557783]	-0.09591964002830557


PCA without Preprocessing
Country Code
0
1
2
Component 2
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
Component 1
−2
0
2
4
6


PCA with UV-1-der Preprocessing
Country Code
0
1
2
0.10
0.08
0.06
0.04
0.02
0.00
Component 2
−2.0
−1.5
−1.0
−0.5
0.0
Component 1


PCA with UV-2-der Preprocessing
Component 2
0.000
−0.005
−0.010
−0.015
−0.020
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Component 1
Country Code
0
1
2

Conclusions

Based on the evaluation metrics, UV-1-der and UV-2-der preprocessing methods provide the best clustering results.
PCA without preprocessing also performs well.