CSE-472
Assignment-4

ID: 1905061
Name: Nazmus Sakib

Submission Folder Tree

```
.
├── 1905061_report.pdf
├── em
│   ├── em_data.txt
│   └── em.py
└── pca
    ├── pca_data.txt
    └── pca.py
```

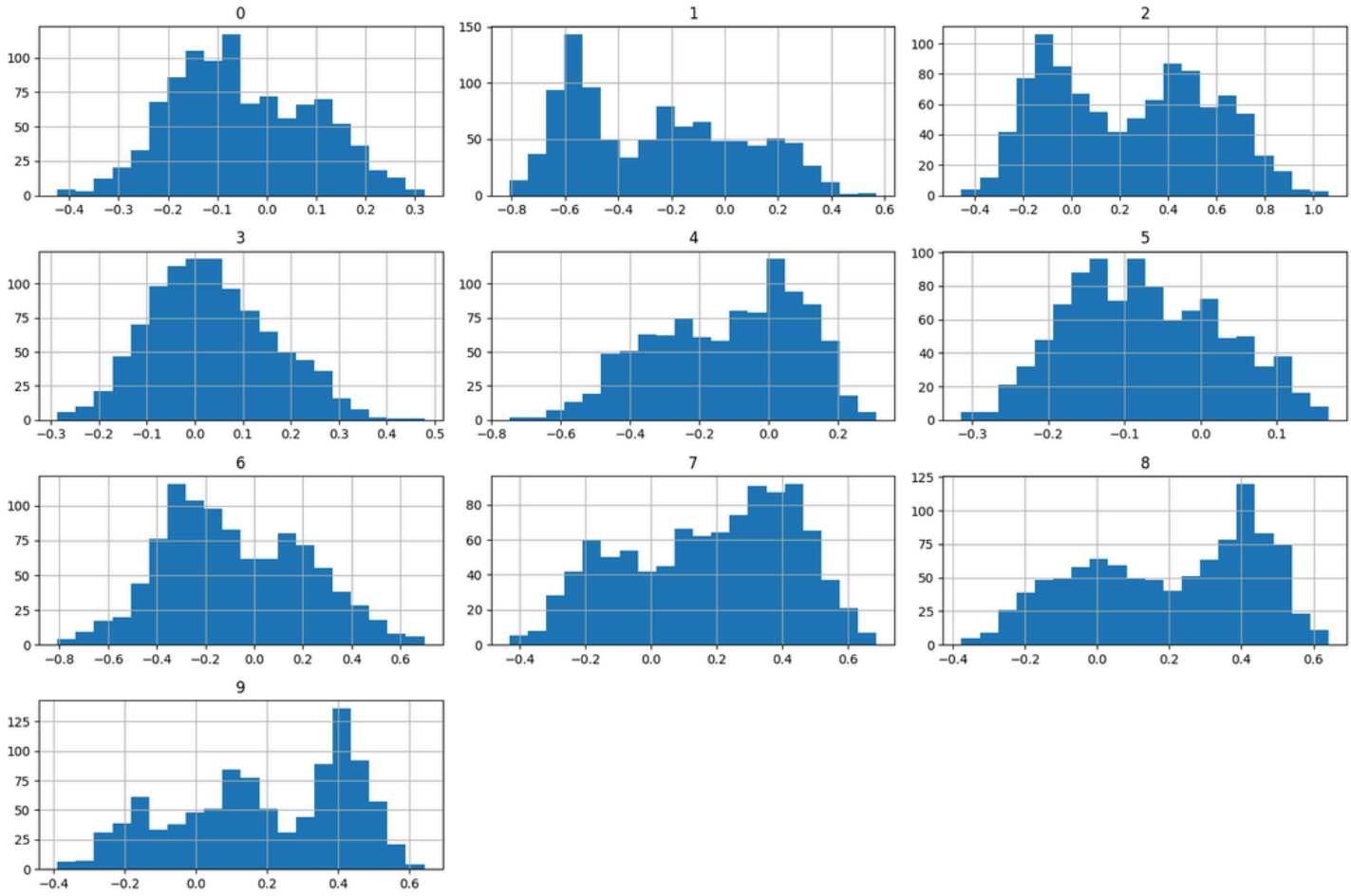
Run Commands:

(required packages: umap-learn pandas matplotlib seaborn scikit-learn numpy)

- python pca.py
- python em.py

Plots for PCA (Principal Component Analysis)

1) **Feature distribution plot**

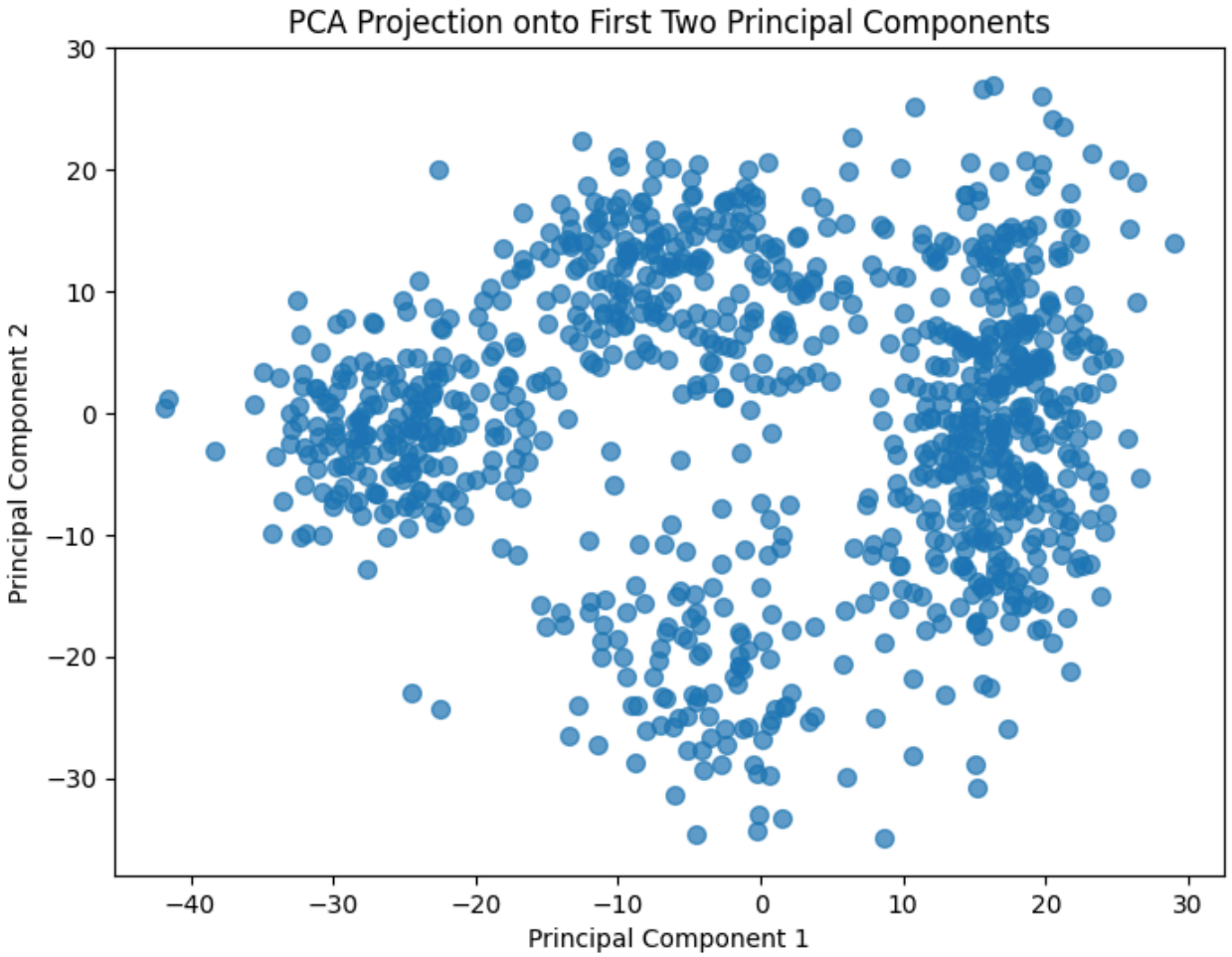


→ The distributions vary across features, with some showing symmetry (approximately normal distribution) and others skewness.

→ Feature 3 appears nearly normally distributed, while features like 9 exhibit bimodal behavior, potentially indicating clustering or categorical splits in the dataset.

→ The variance across features is noticeable, suggesting that standardization is needed

2) **PCA 2D Scatter Plot**

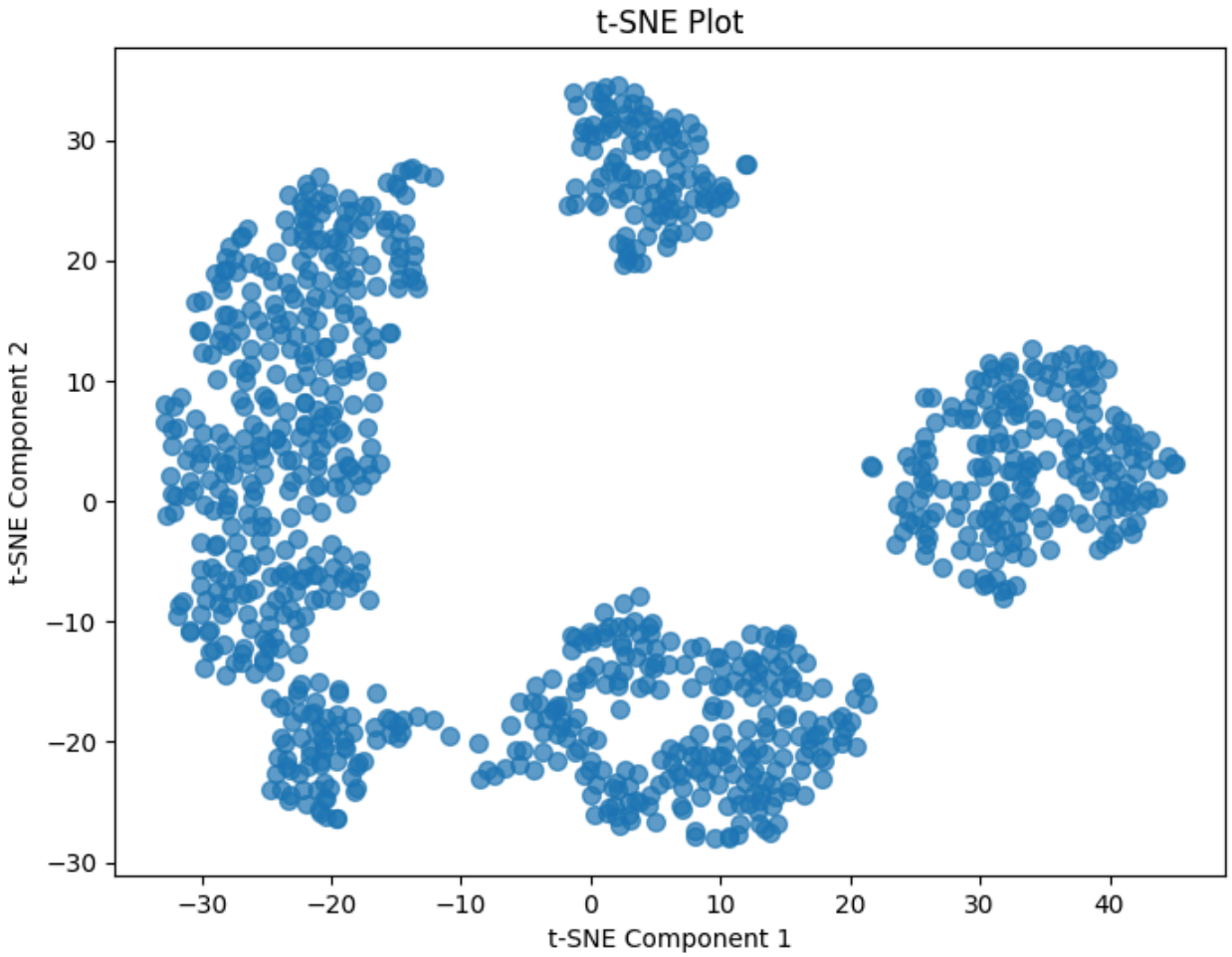


→ It shows the projection of the data onto the first two principal components.

→ The plot suggests that the first two components capture significant variance, as distinct clusters or groupings are visible.

→ Overlapping clusters imply shared characteristics among some data points.

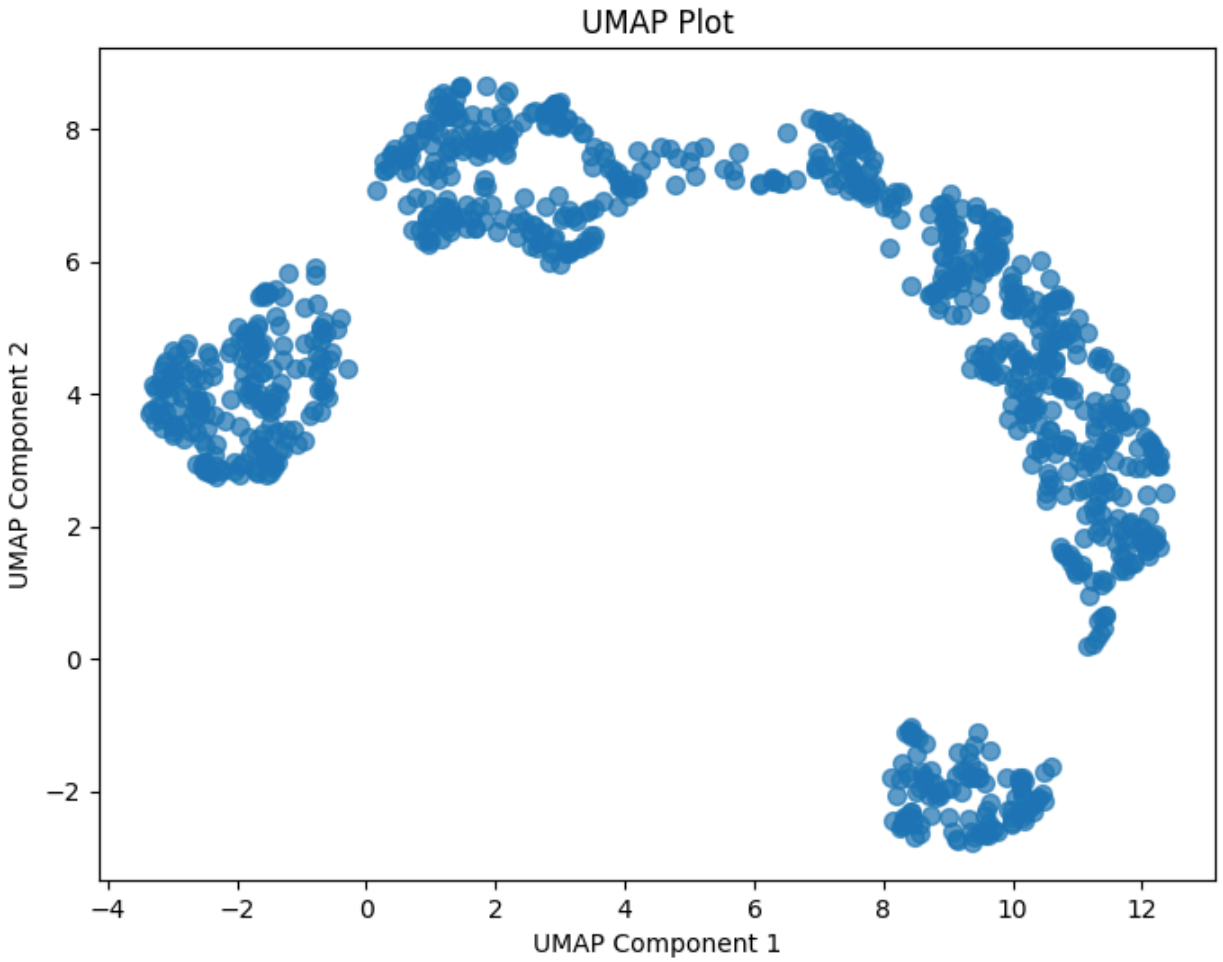
3) **t-SNE Plot**



→ This indicates that t-SNE captured local structures (i.e., relationships between nearby points) more effectively than PCA, which focuses on global variance.

→ Four distinct clusters are visible, suggesting meaningful groupings within the dataset.

4) **UMAP Plot**



→ UMAP preserves both local and global structures effectively, as shown by the more continuous arrangement of points while still forming clusters.

Expectation-maximization (EM) algorithm

Parameter	Value
Mean number of children in families with family planning advice (lambda1)	1.783
Mean number of children in families without family planning advice (lambda2)	4.911
Proportion of families with family planning advice (pi)	0.356
Proportion of families without family planning advice	0.644