This project explores dimensionality reduction techniques—PCA, t-SNE, and UMAP—applied to the MNIST handwritten digits dataset to visualize and analyze clusters in a 2D space. By comparing each technique based on execution time and cluster separation quality (measured through Silhouette Score), the project provides insights into the effectiveness of these methods for high-dimensional data visualization. PCA serves as a baseline linear method, while t-SNE and UMAP offer non-linear approaches that excel at uncovering complex patterns in data. The final output includes visualizations and metrics that highlight the trade-offs in efficiency and clustering quality among the three techniques, making it a valuable study in dimensionality reduction for machine learning and data analysis.

Techniques Used

1. **Principal Component Analysis (PCA)**: A linear technique that projects data into two dimensions, preserving as much variance as possible.
2. **t-Distributed Stochastic Neighbor Embedding (t-SNE)**: A non-linear technique that excels at creating clusters in high-dimensional data.
3. **Uniform Manifold Approximation and Projection (UMAP)**: A non-linear technique that preserves local and global data structures, often faster than t-SNE.

Comparison Aspects

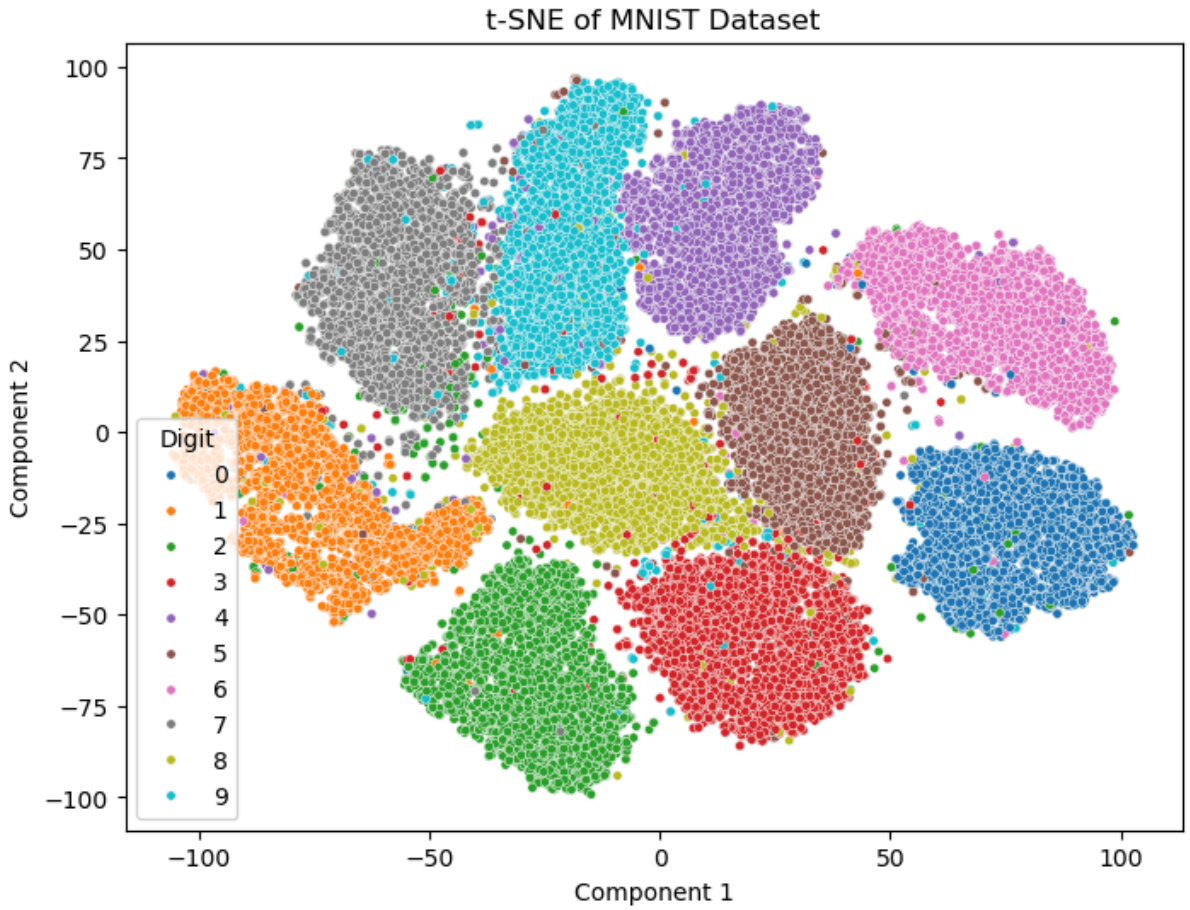
The notebook compares PCA, t-SNE, and UMAP based on:

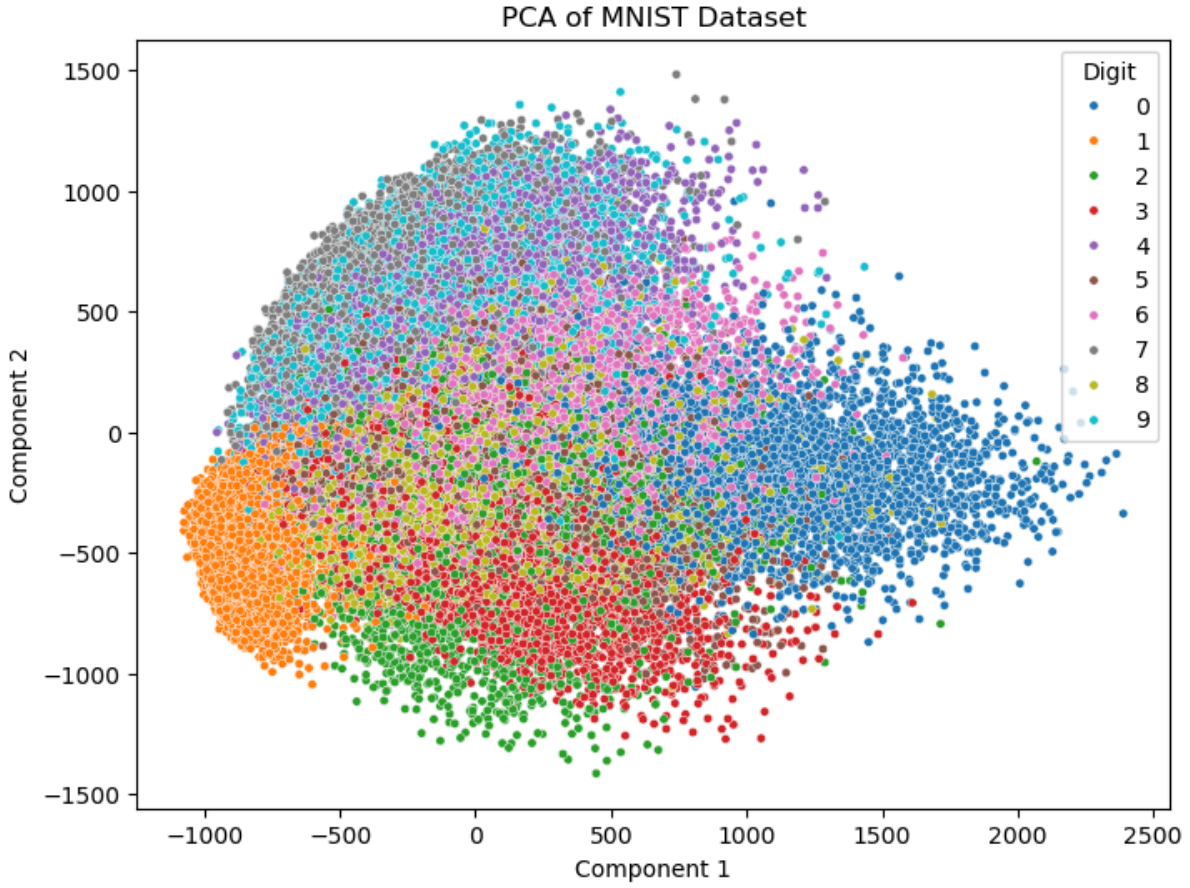
1. **Execution Time**: The time taken by each technique to reduce the dimensionality, helping assess computational efficiency.
2. **Silhouette Score**: Quantifies cluster separation, with higher scores indicating better-defined clusters.

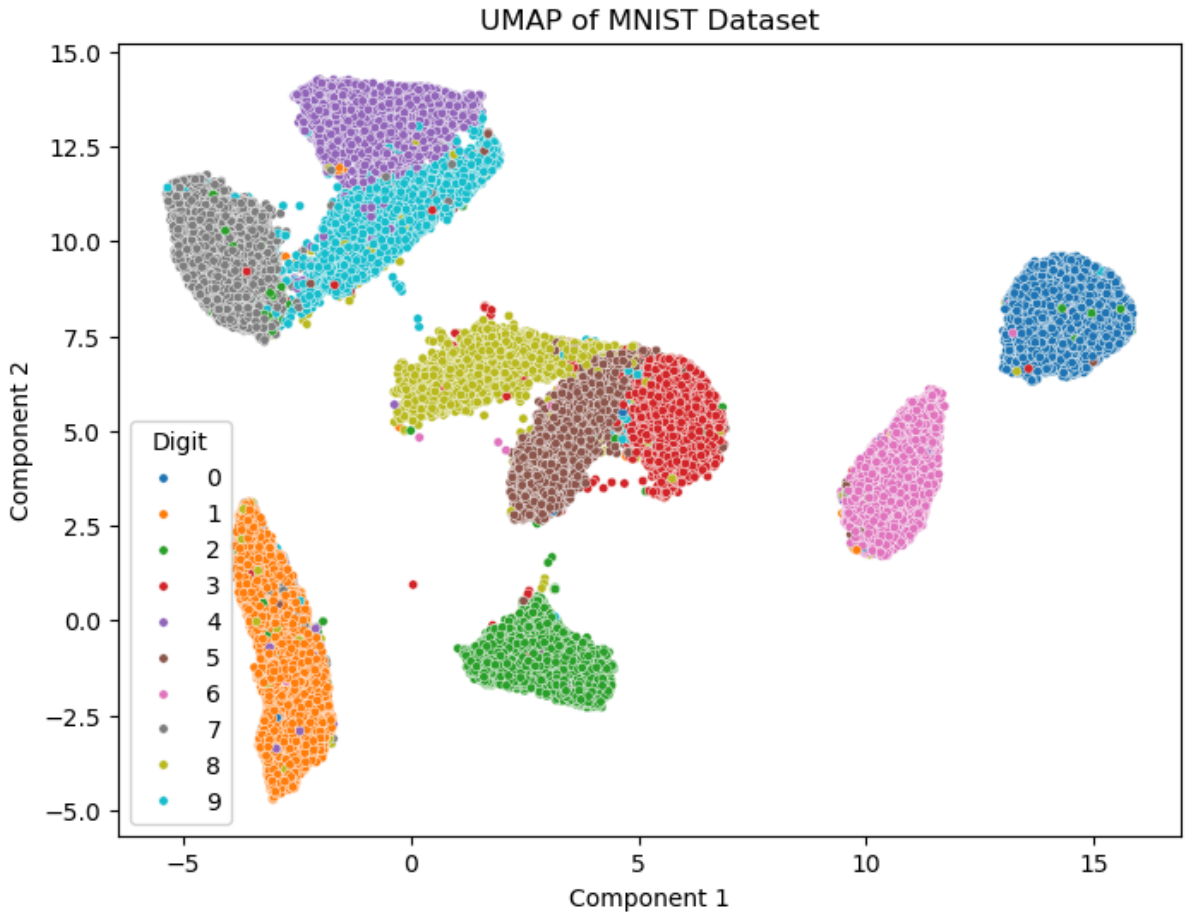
Results

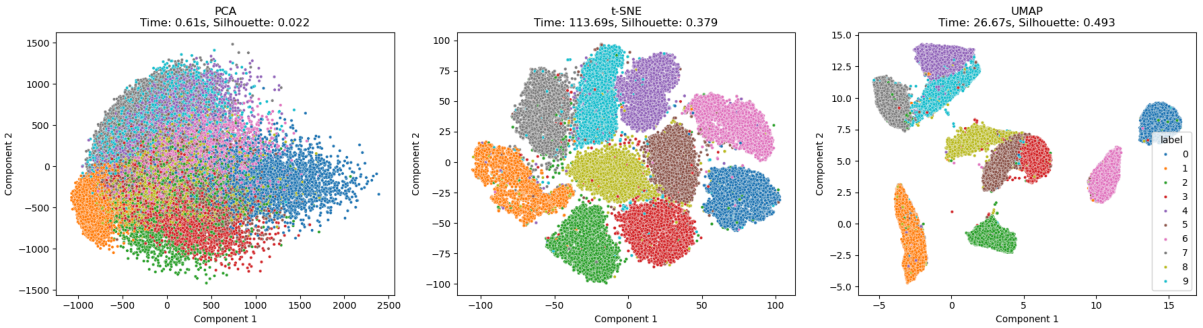
Each technique is visualized in a combined plot, showing clusters of digits in 2D space. The results include:

- **Execution Time**: Provides insight into computational efficiency.
- **Silhouette Score**: Indicates how well-separated and distinct the clusters are for each technique.


t-SNE of MNIST Dataset
Component 2
Component 1
Digit
0
1
2
3
4
5
6
7
8
9
100
75
50
25
0
−25
−50
−75
−100
−100
−50
0
50
100


PCA of MNIST Dataset
Component 2
Component 1
Digit
0
1
2
3
4
5
6
7
8
9
1500
1000
500
0
−500
−1000
−1500
−1000
−500
0
500
1000
1500
2000
2500






```
PCA Time: 0.61 seconds, Silhouette Score: 0.022
t-SNE Time: 113.69 seconds, Silhouette Score: 0.379
UMAP Time: 26.67 seconds, Silhouette Score: 0.493
```

PCA t-SI