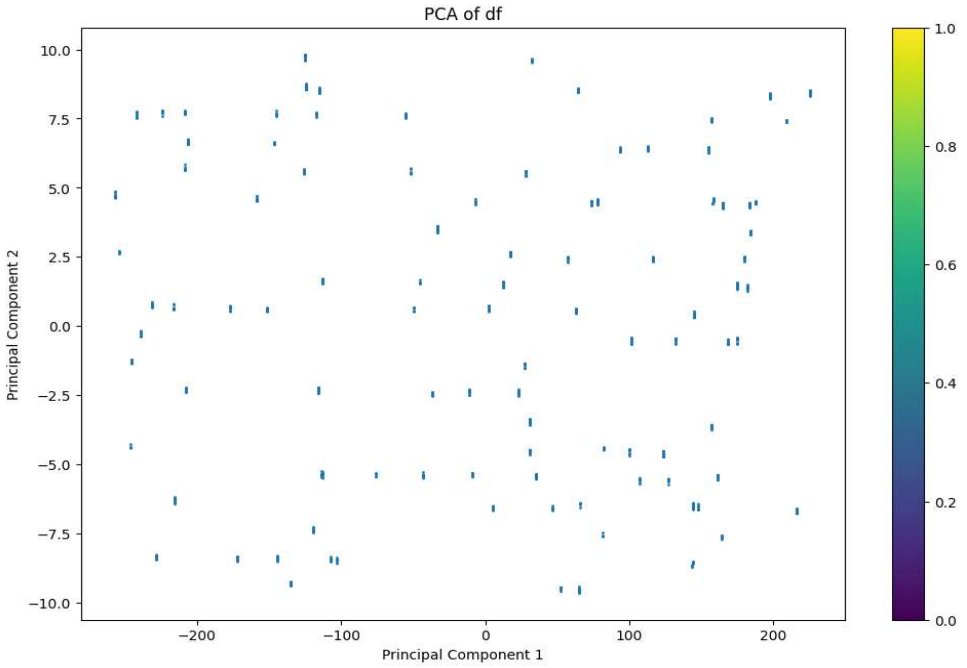
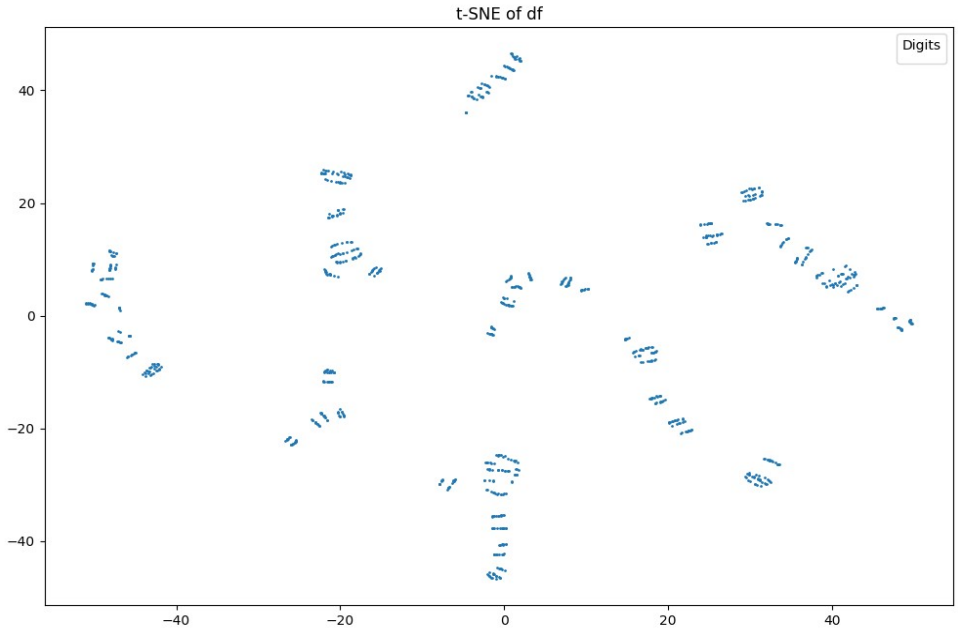
REPORT



After applying PCA for dimensionality reduction to two dimensions, the results did not provide clear inferences or visible patterns, making it difficult to assess whether clusters could even be formed.

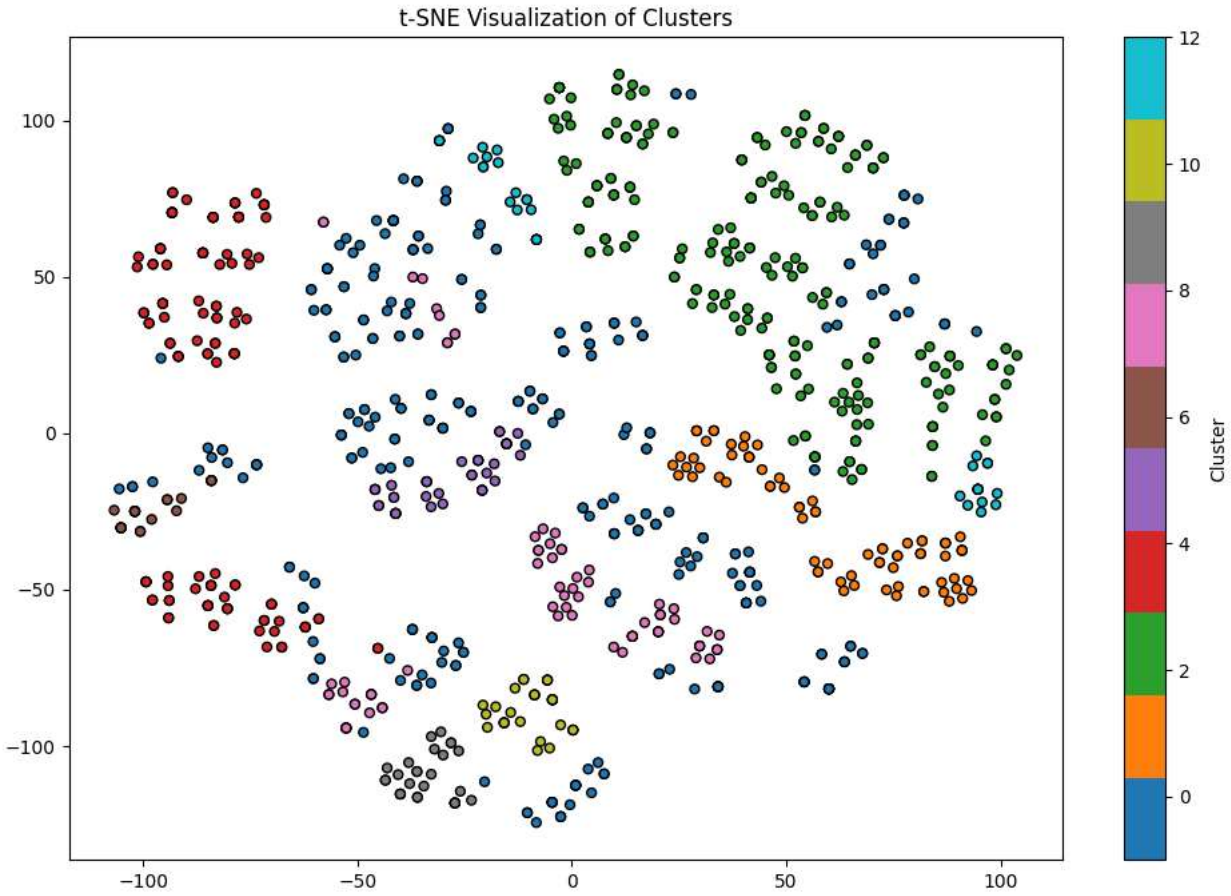


However, when using t-SNE, well-defined clusters were visible, indicating a potential structure in the data. Based on these observations, decided to proceed with DBSCAN

(Density-Based Spatial Clustering of Applications with Noise) whicch is better suited for identifying clusters of varying shapes and densities, especially in high-dims data.

EXPERIMENT 1

After encoding the dataset and performing scaling on all columns, DBSCAN was applied with the params eps=0.34 and min_samples=15. However, the resulting clusters were not feasible, as evidenced by the t-SNE visualization, where a color parameter was used to differentiate the clusters. The output did not reveal clear or meaningful groupings, indicating that the maybe chosen parameter values for DBSCAN may not have been optimal for the dataset. Number of noise pts suggested clearly Not Good.



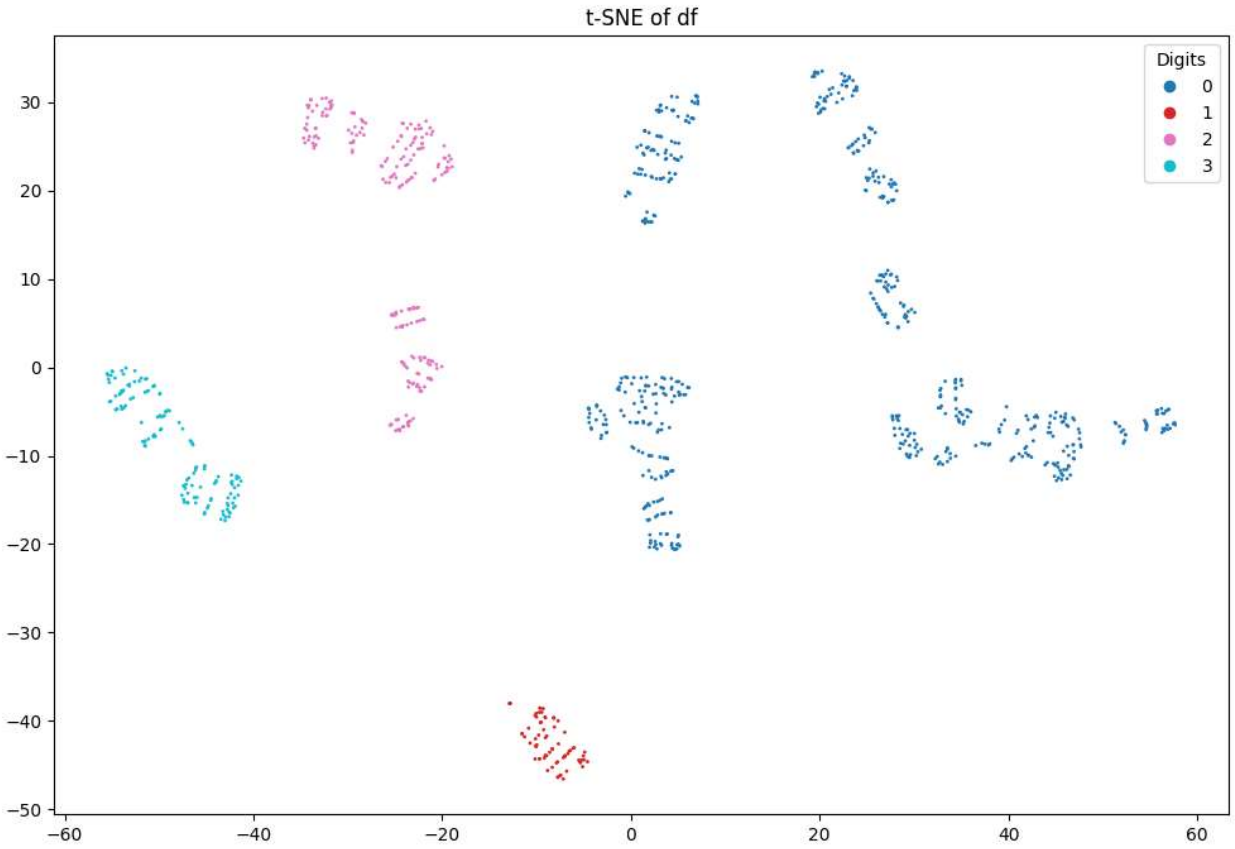
No of Clusters : 13

Davies-Bouldin Index: 2.0612663475112556

EXPERIMENT 2

Upon inspecting the t-SNE visualization, noticed that the points did not exhibit well-defined clusters, even without clustering ie they weren't separated well enough before even. As a

result, I decided to use the previous t-SNE output and apply DBSCAN on that. To make the data as previous scaled only the Price column, as it was the one that was out of range compared to the other features. After applying the scaling and running DBSCAN again with eps=0.15 and min_samples=5, the DB Index was 0.4471 with no data labelled as Noise!



No. of Clusters: 4

Davies-Bouldin Index: 0.4470408820021904

Experiments and Evaluation

Experiments	*Clusters*	*DB - Index*	*Noise*
EXP1: All features scaled	13	2.061	285
EXP2: Only Price col Scaled	4	0.4470	0