Unsupervised learning final Project

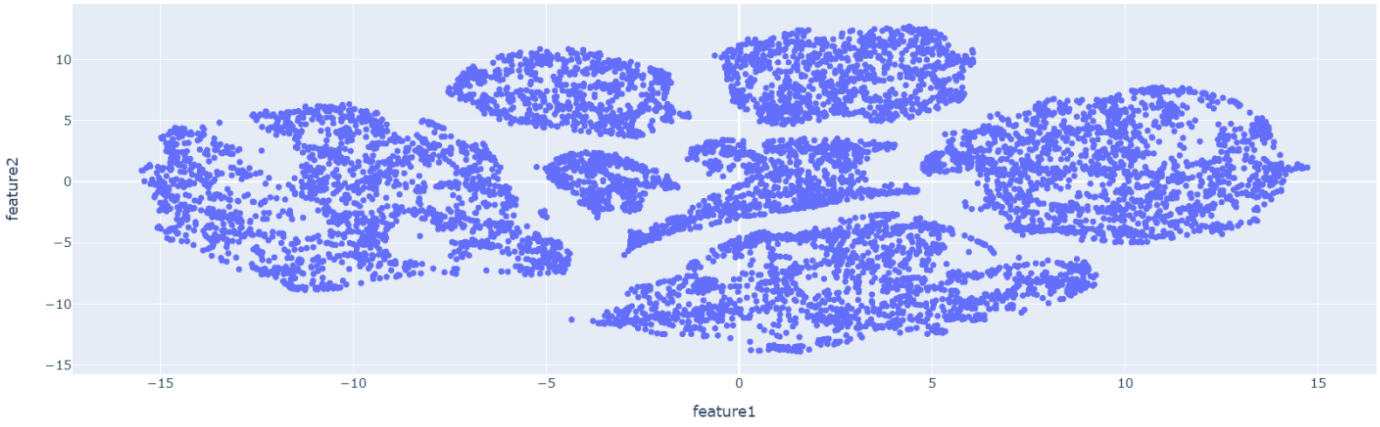
By: *Fatma Alzahraa Marzouk Gaber*

- Evaluate the difference between data transformation techniques(notebook):

  At this part using different Data Transformation techniques

  - Feature Scaling:
    1. MinMax Scaler.
    2. Standard Scaler.
    3. MaxAbsScaler.
    4. Robust Scaler.
    5. Quantile Transformer Scaler.
  - Feature Transformation:
    1. Log Transform.
    2. Power Transformer Scaler.

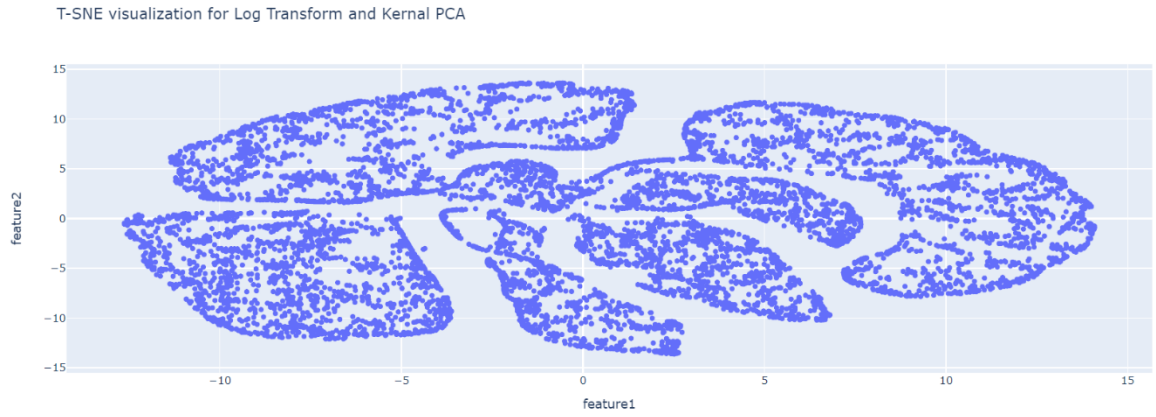
The Best Case for our problem is Log Transform. It separates Data as seen below:

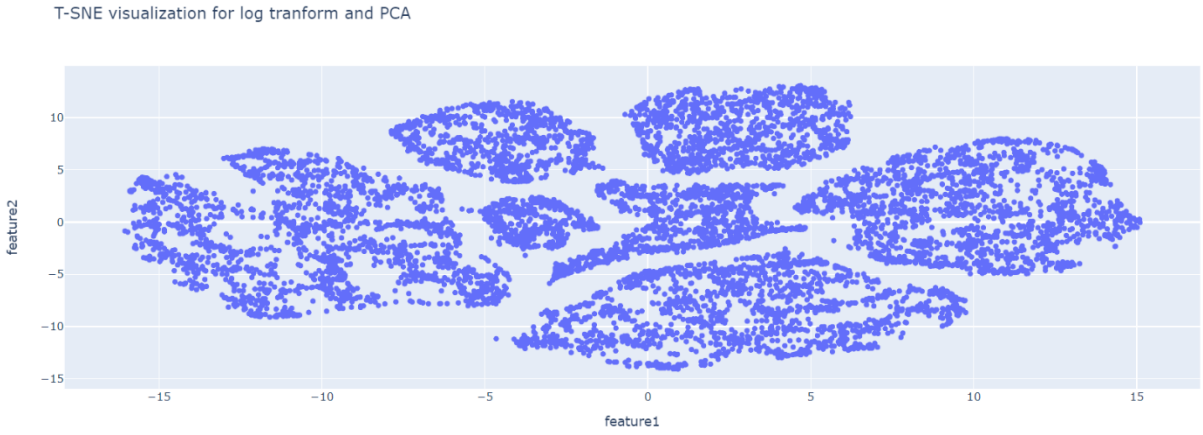


- Is PCA better than Kernel PCA? (notebook)

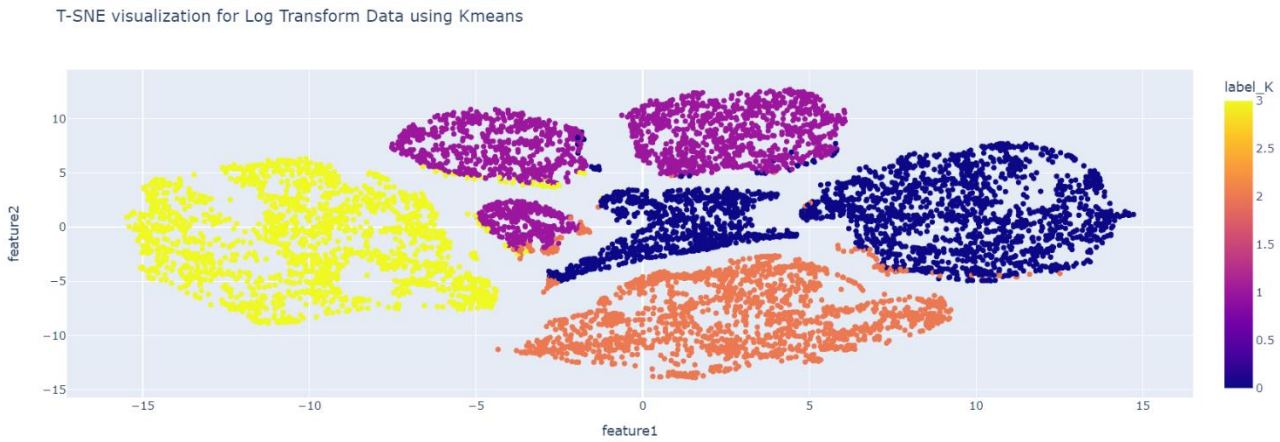
  At This part working on Stander scaler data, Log Transformed data using n_components = 11 for stander scaler and 7 for Log Transformed data.
  for kernel PCA using Radial Basis Function Kernel.
  The results for stander scaler data almost same.

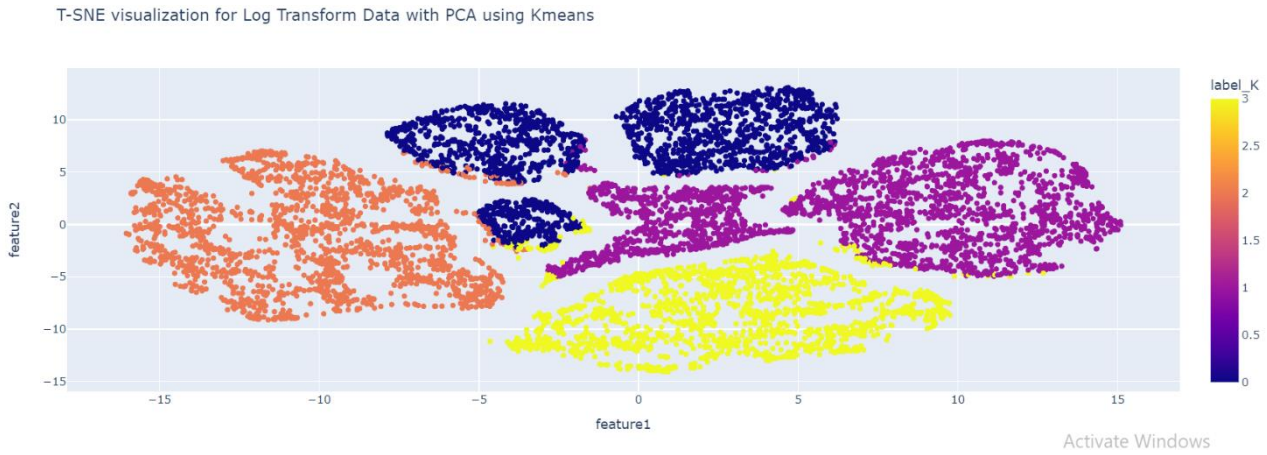
For log Data Kernel PCA adds boundaries for data compared to PCA.

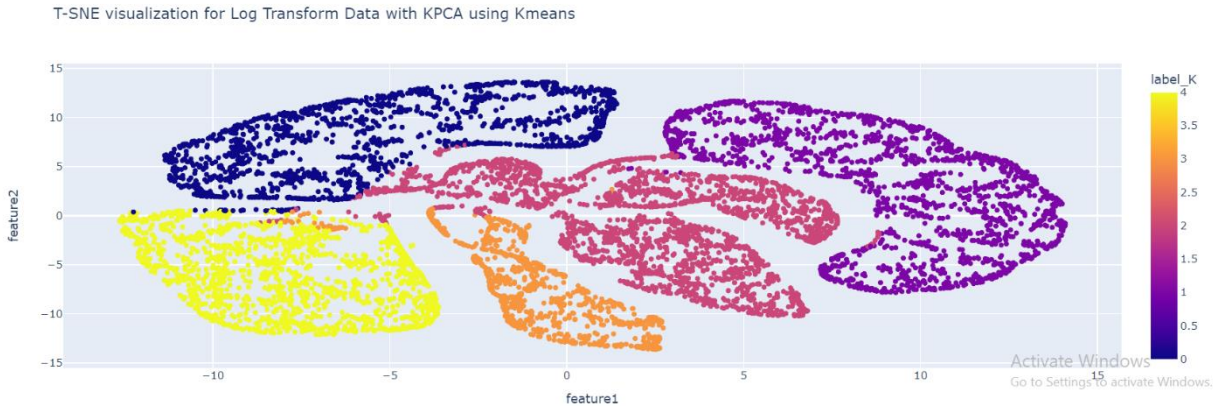




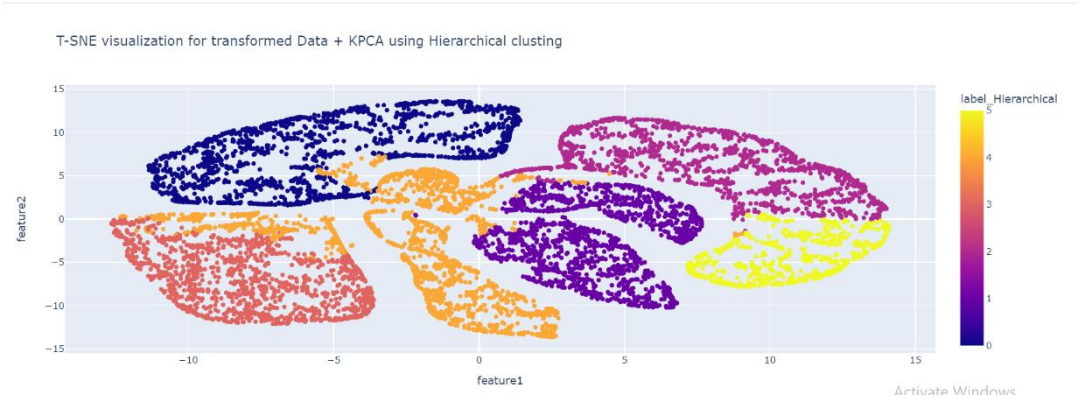
- Try all unsupervised algorithms that you studied:
    1. Kmeans and Hierarchical Clustering (notebook).
    At this part using Stander scaler data and Log transformed Data.
    Results Kmeans on Log Data:



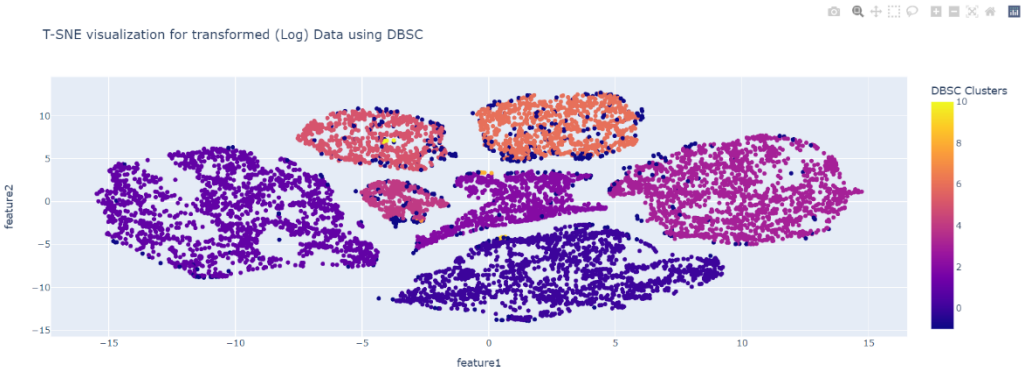


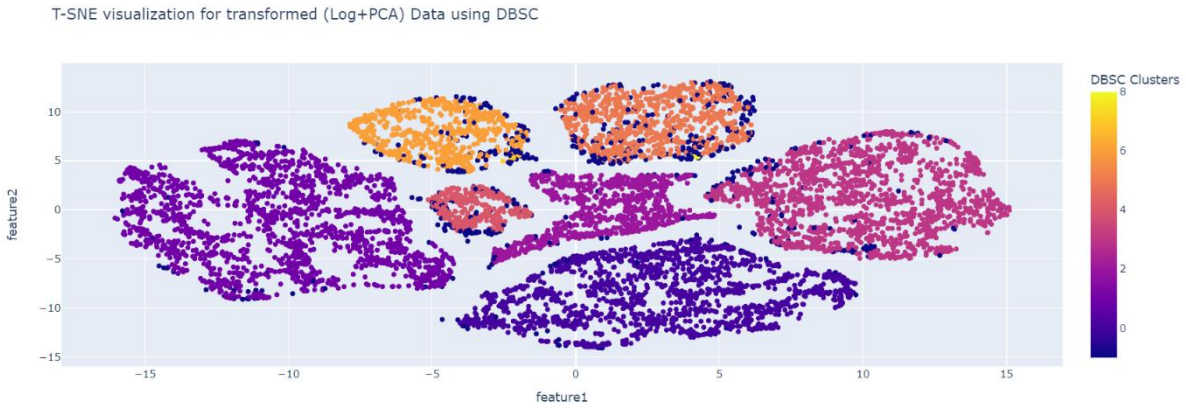


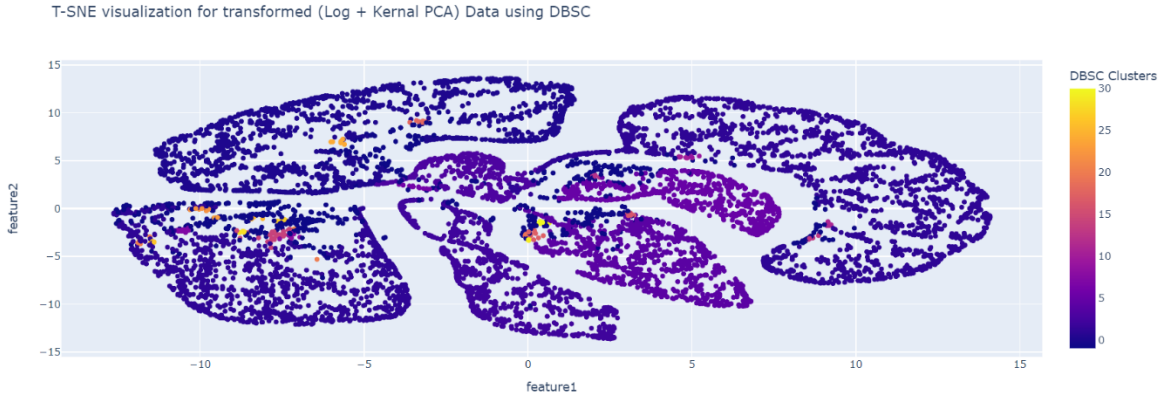
Results Hierarchical on Log Data:
Kernel PCA is best visualization in Hierarchical Clustring.



2. DBSCA (notebook)
   Results:

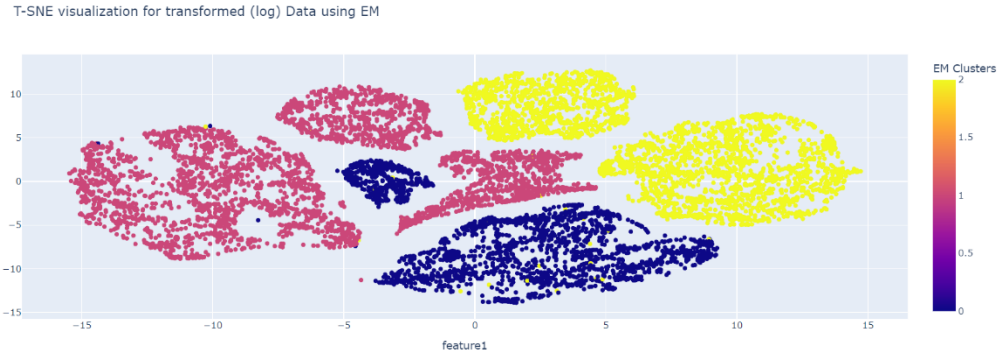


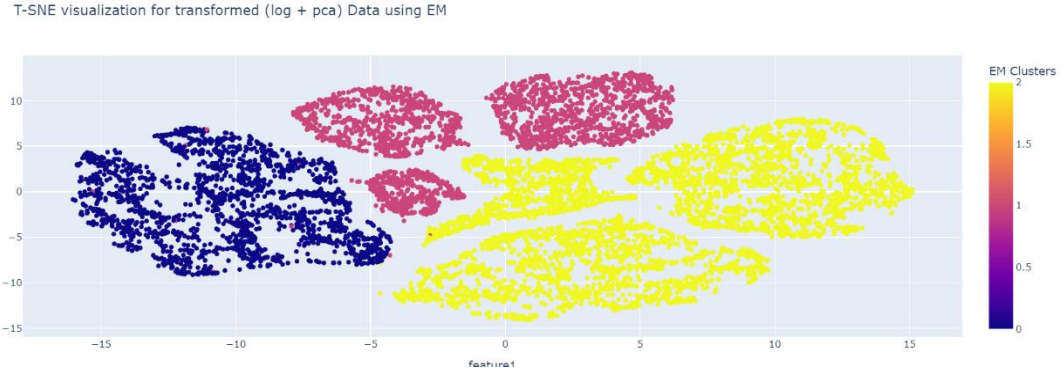


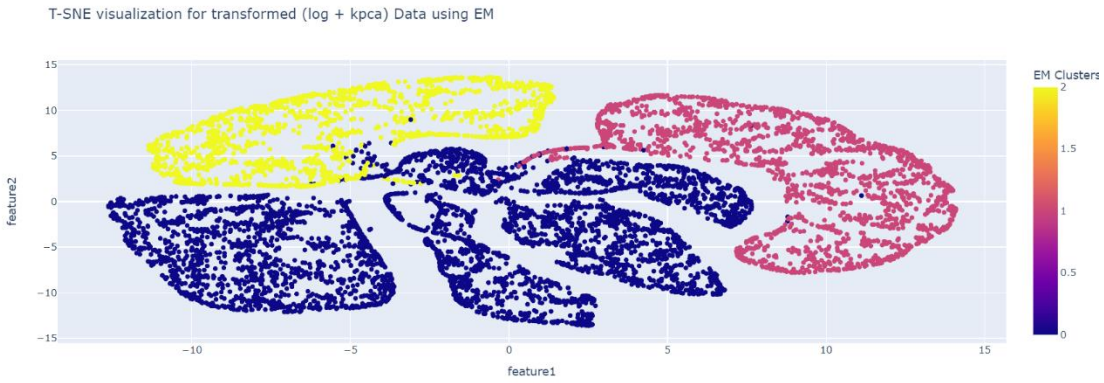


---

3. EM (notebook):
   Results:







anomaly detection algorithm (notebook).

Results: