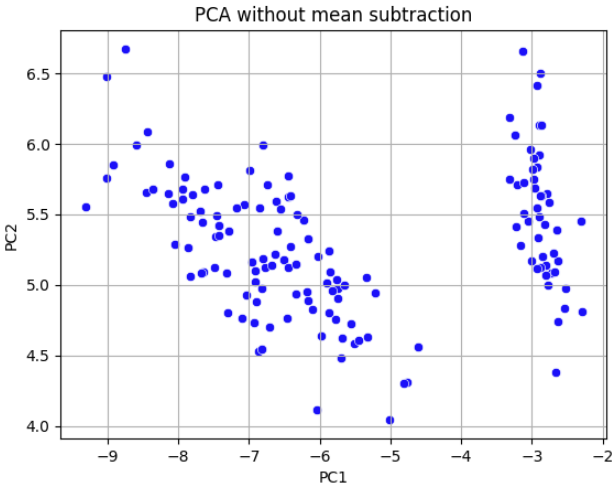
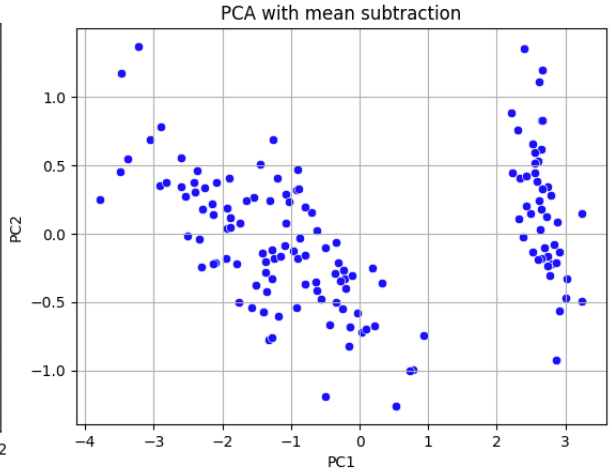
Unsupervised Feature Extraction Report

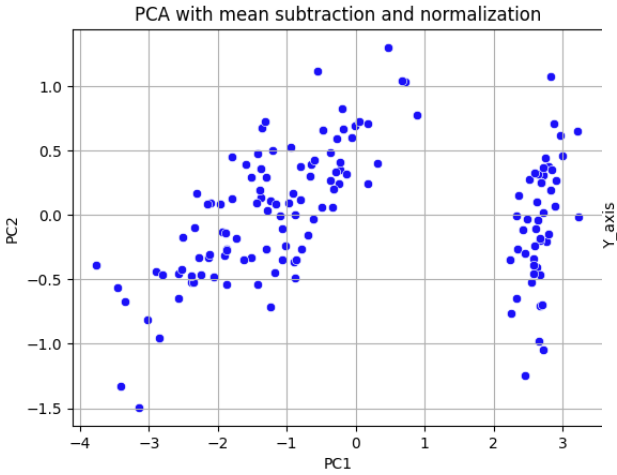
Results –

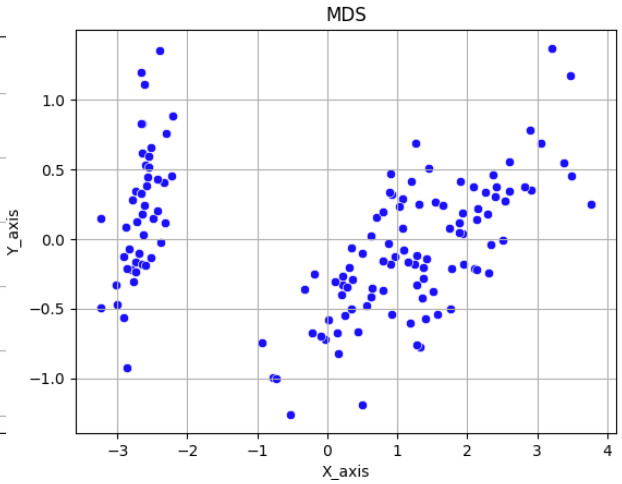
Running PCA1, PCA2, PCA3 and MDS on Iris and wine data sets results in the following –

Iris dataset –

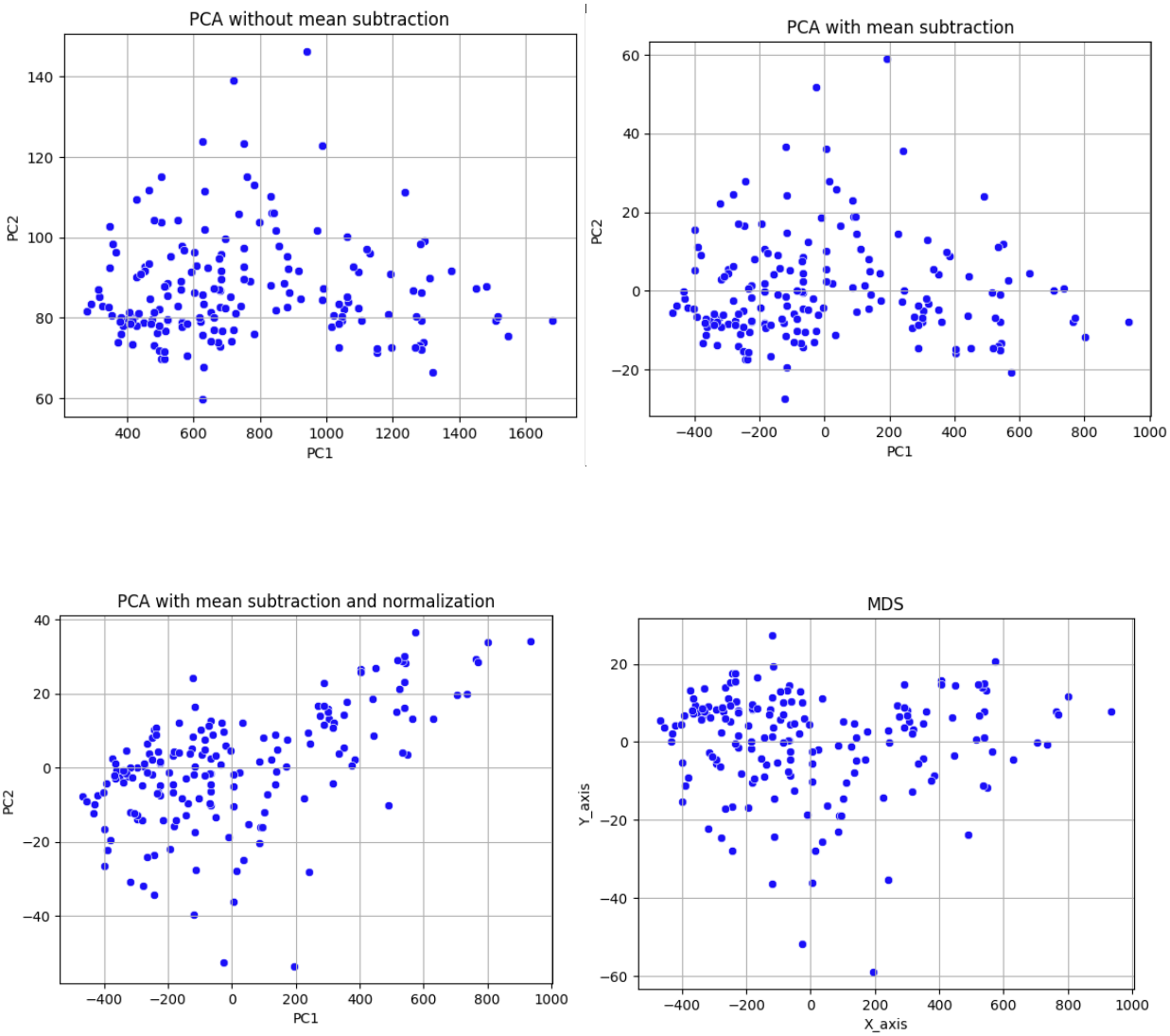




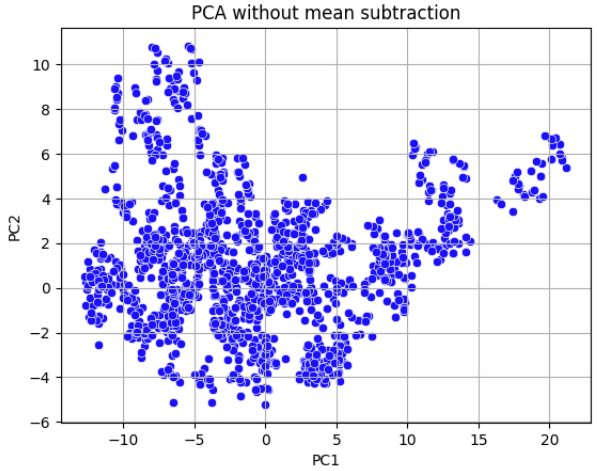


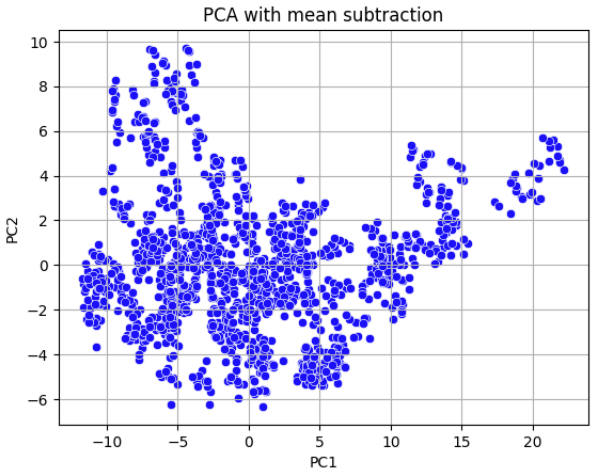


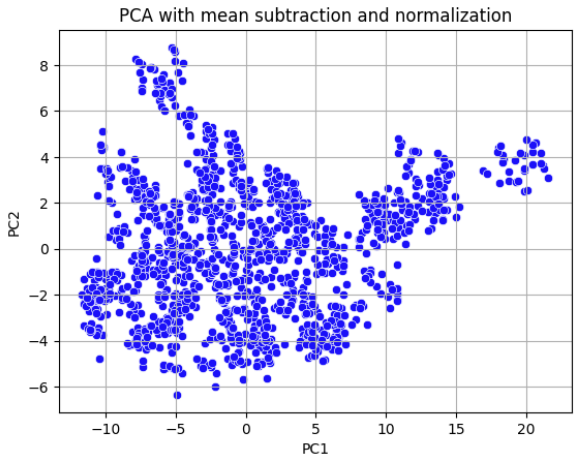
Wine dataset –

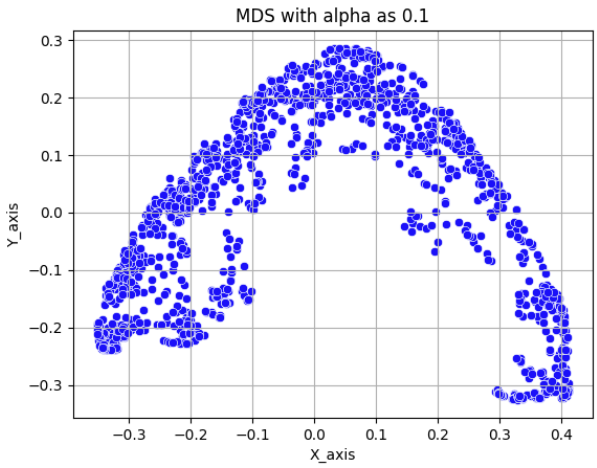


Bank Note Authorization (taken from https://www.kaggle.com/datasets/ritesaluja/bank-note-authentication-uci-data)-

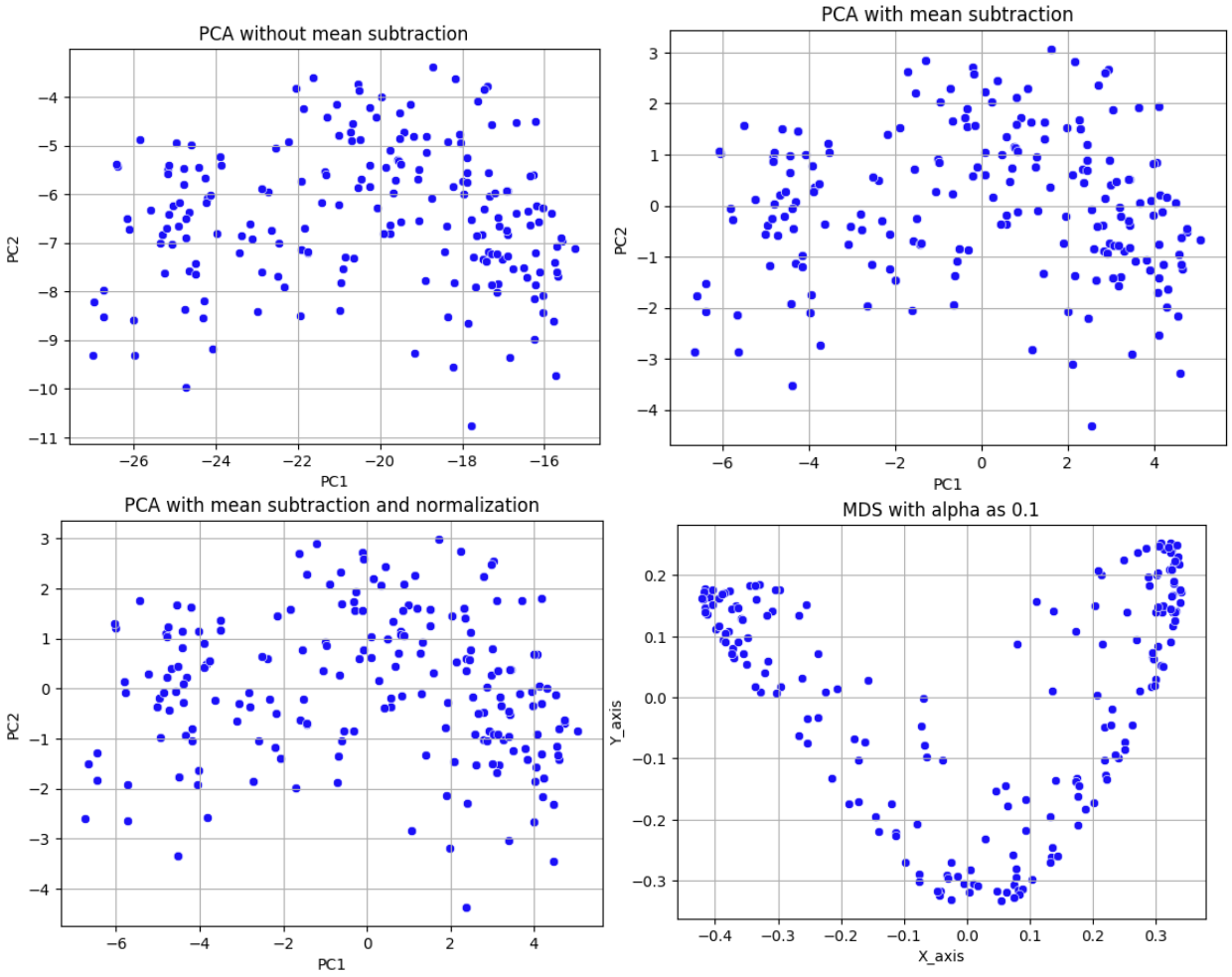








Seeds Dataset (https://www.kaggle.com/datasets/jmcaro/wheat-seedsuci)-



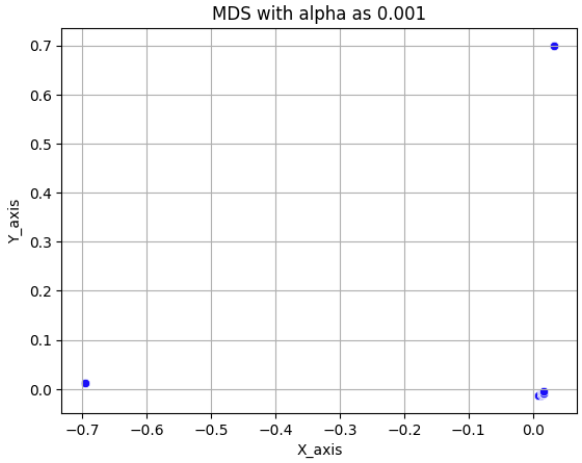
Standard Centered PCA vs Normalized Centered PCA –

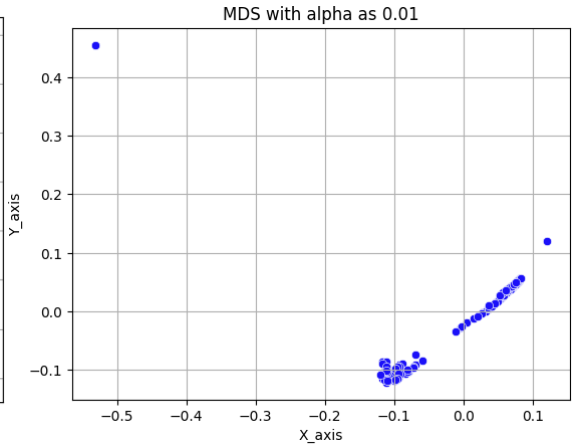
In normalized PCA, each term in the covariance matrix is scaled by the inverse of the squared norm of the centered data point, which can help mitigate the effect of outliers. Normalized PCA also has less sensitivity to scale letting us use this features with completely different dimensions.

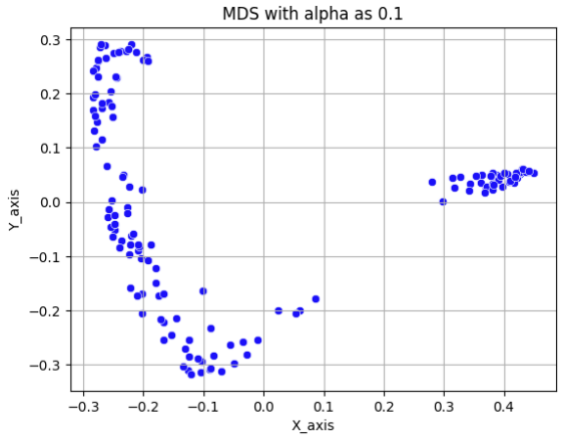
Normalized PCA is more useful in scenarios where we have a lot of outliers that which we want to ignore. Normalized PCA is also preferred when the scale of the features is contrasting like for example count, and weight in Kgs, Height in cm.

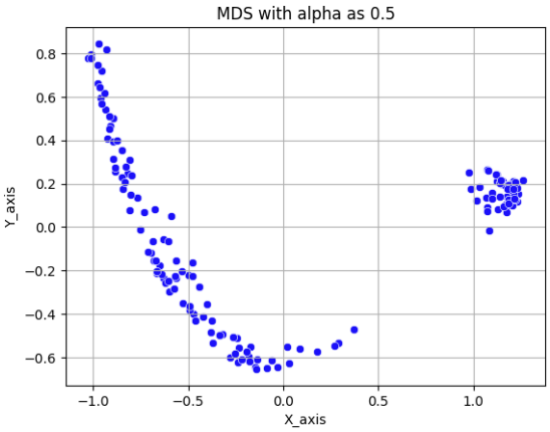
The Effect of Alpha on MDS –

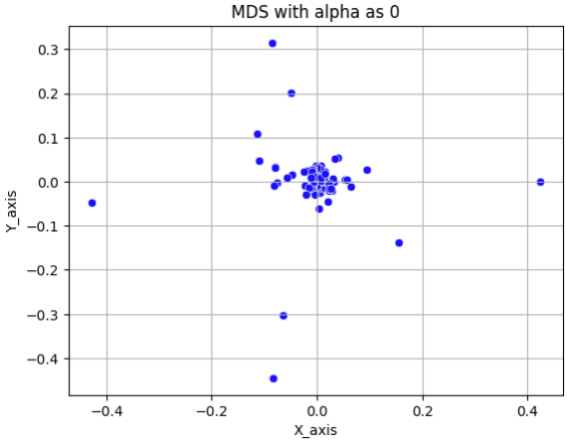
Below is the results of MDS with varying with alpha -

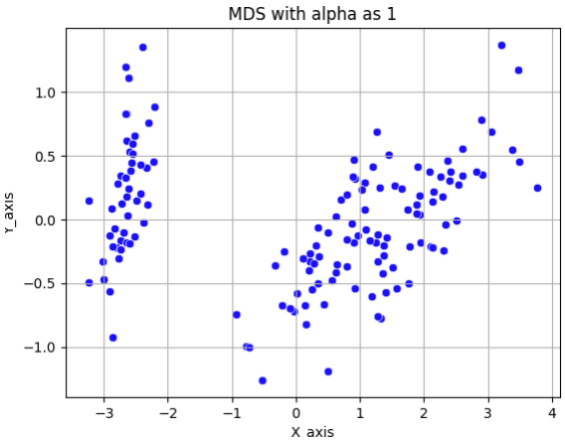


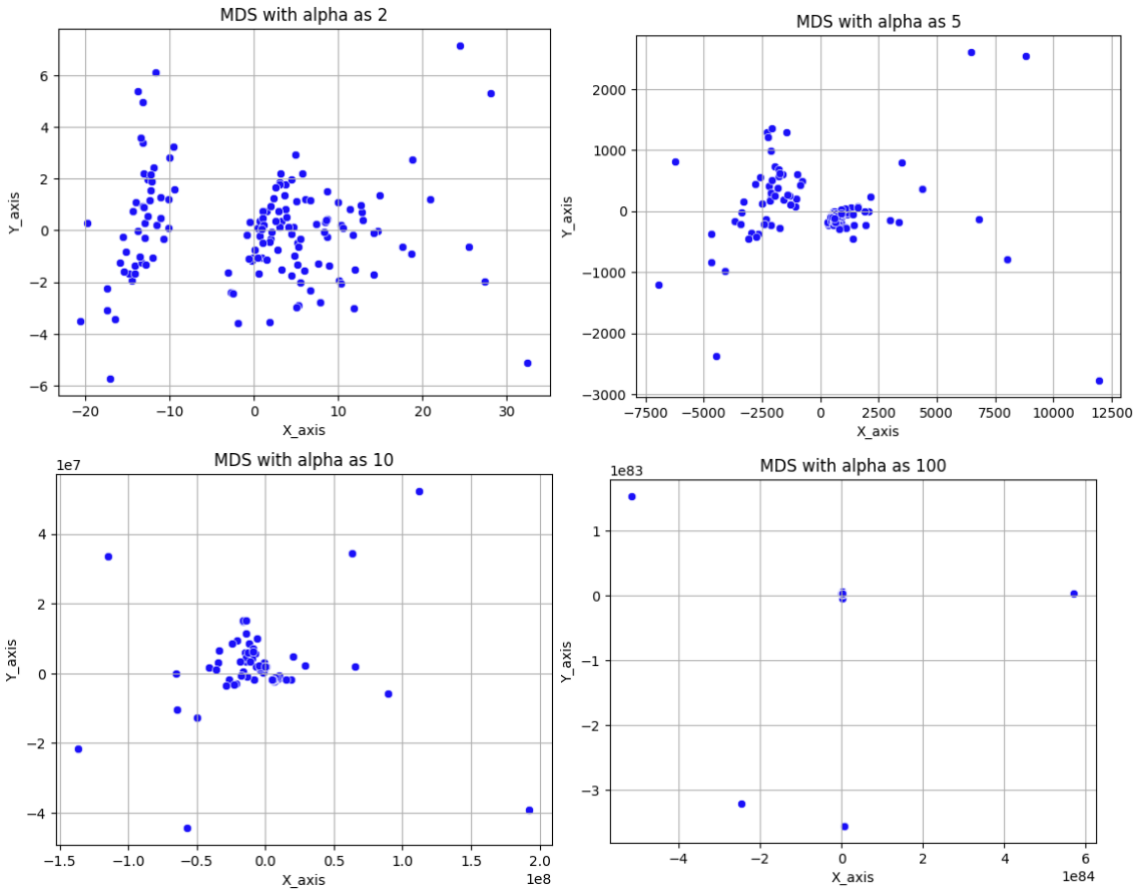












When α is small, the effect of distances becomes less pronounced and larger distances are more compressed than smaller ones. This can lead to a representation that emphasizes local structures or small-scale variations in the data. On the other hand, larger α values amplify the differences between distances. This can exaggerate the separation between points that are already far apart.

We can use a smaller alpha can help reduce the effect of noise on the data, and we can use it to preserve the relationship between points. Larger Alpha can help in the separation between distinct groups. But as seen too large or two small of an alpha basically renders MDS useless.