Data Science Documentation

# Datasets:

* Training
* Validation

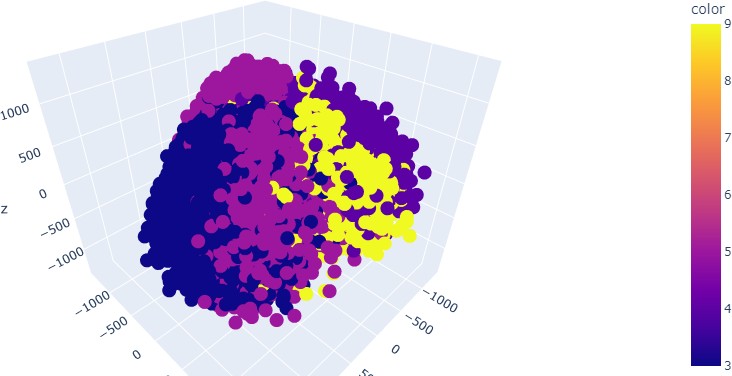
# Reduction by using:

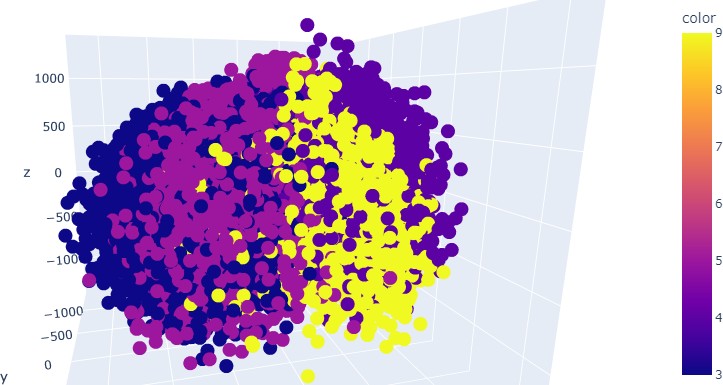
* PCA
* Kernel PCA
* LLE
* TSNE

**Variance preserved in PCA transformations:** [0.12592273 0.08201049 0.05513535]

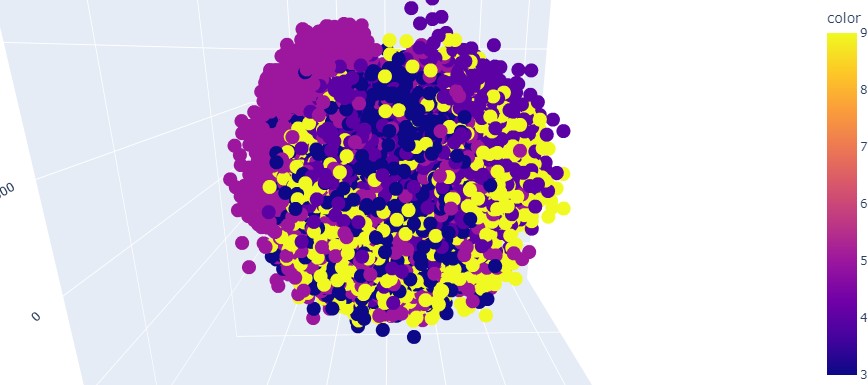
# 3D Scatter Plots

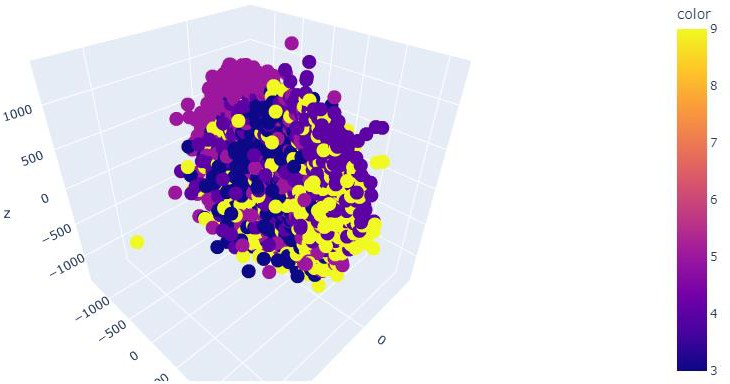
## PCA:



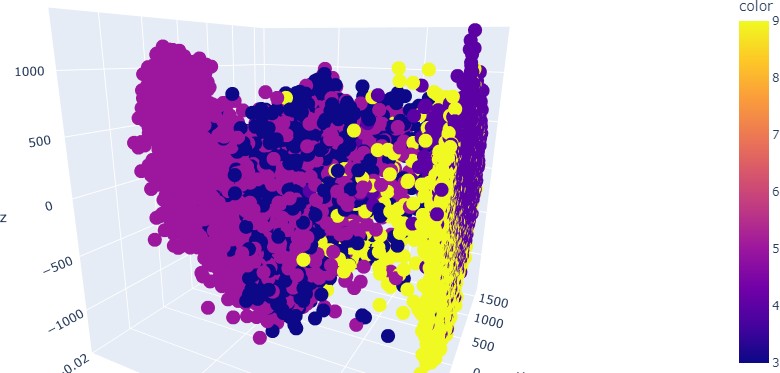


Kernel PCA:



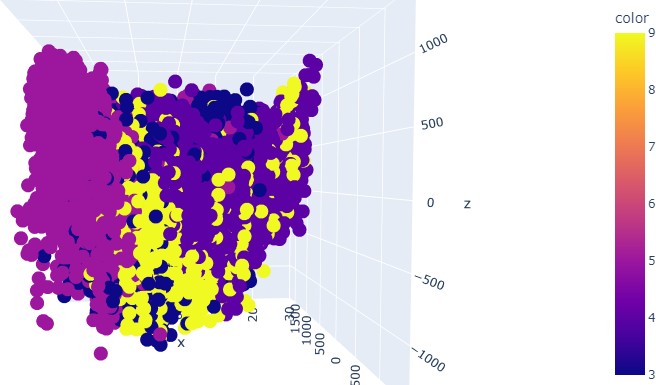


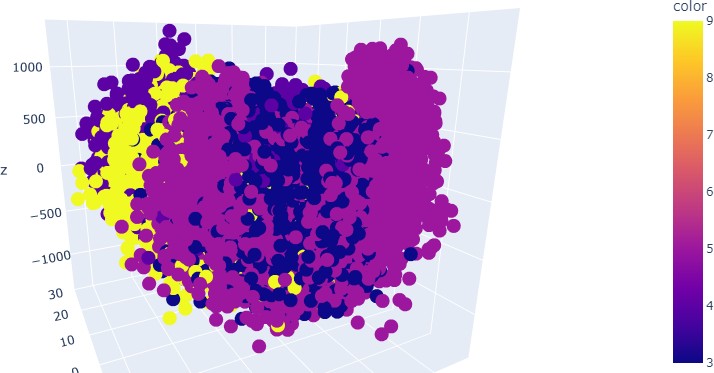
LLE:

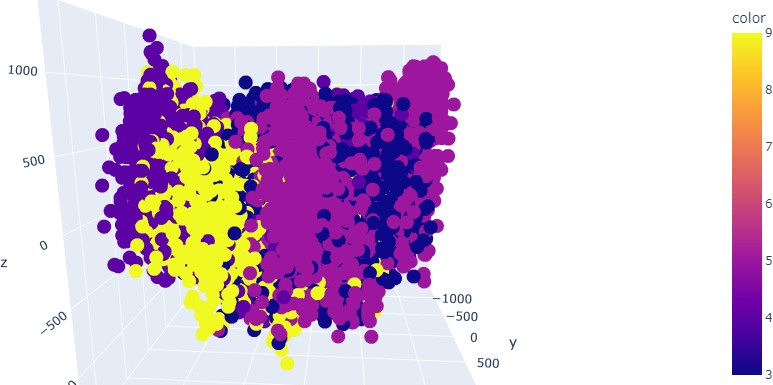




TSNE:







Verdict:

Upon the observation of the visualization of different reduction techniques, we can infer that the

**TSNE** and **PCA** separated the digits better than the rest.

# Logistic Regression:

## PCA:

**Training Accuracy:** 61.835475578406175

**Validation Accuracy:** 61.64911379398921

## Kernel PCA:

**Training Accuracy:** 26.020565552699225

**Validation Accuracy:** 27.819162599537634

## LLE:

**Training Accuracy:** 64.1388174807198

**Validation Accuracy:** 65.81042897508348

## TSNE:

**Training Accuracy:** 95.12082262210797

**Validation Accuracy:** 24.017467248908297

# SVM:

## PCA:

**Training Accuracy:** 50.75064267352185

**Validation Accuracy:** 50.52658618032366

## Kernel PCA:

**Training Accuracy:** 25.922879177377894

**Validation Accuracy:** 25.276136655535574

## LLE:

**Training Accuracy:** 44.44215938303342

**Validation Accuracy:** 43.77087079373234

## TSNE:

**Training Accuracy:** 96.09254498714654

**Validation Accuracy:** 24.96789108656563

## Verdict:

Upon observation of the result of the accuracies, we can note that on the training dataset, **TSNE**

performed well and on the validation, **LLE** performed better.