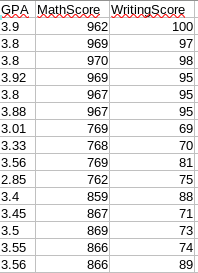
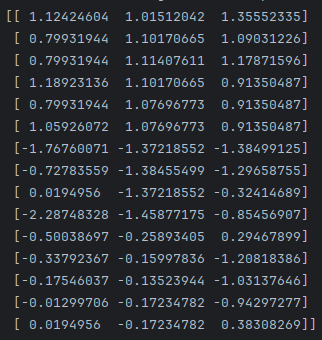
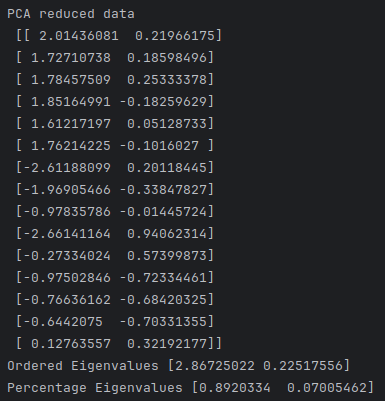
1. Clean and prepared data

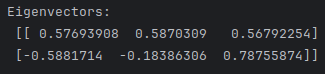


2. Normalized data:

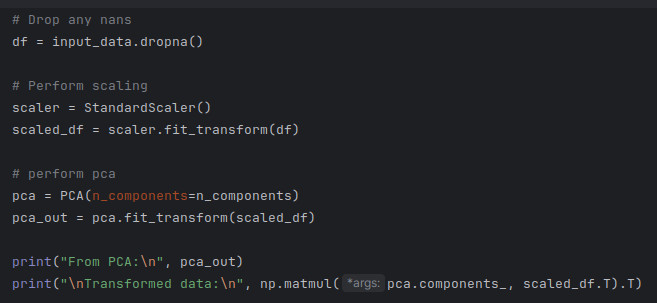
3.

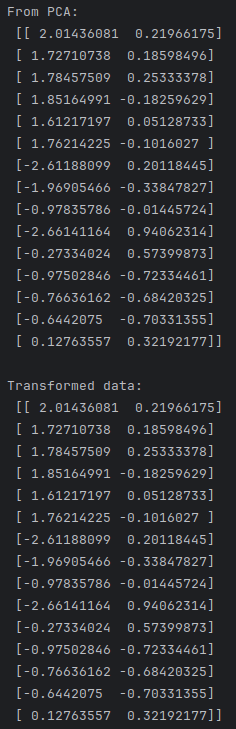


f) The percentage retained is 0.89 + 0.07 = 0.96 = 96%

g)

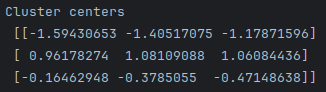
h) We can perform a matrix multiplication using numpy to show that multiplying the eigenvectors by the data produces the same result. The following page contains a code snippet. The page after that shows the output from a print statement. The data is loaded into the df variable, and we perform PCA. n\_components is 2. We then print out the result from pca and the matrix multiplication using np.matmul() of the eigenvectors by the original scaled data.





4)

a)



The centroids tell us the average value of all the data points of a cluster for each dimension.

b)

