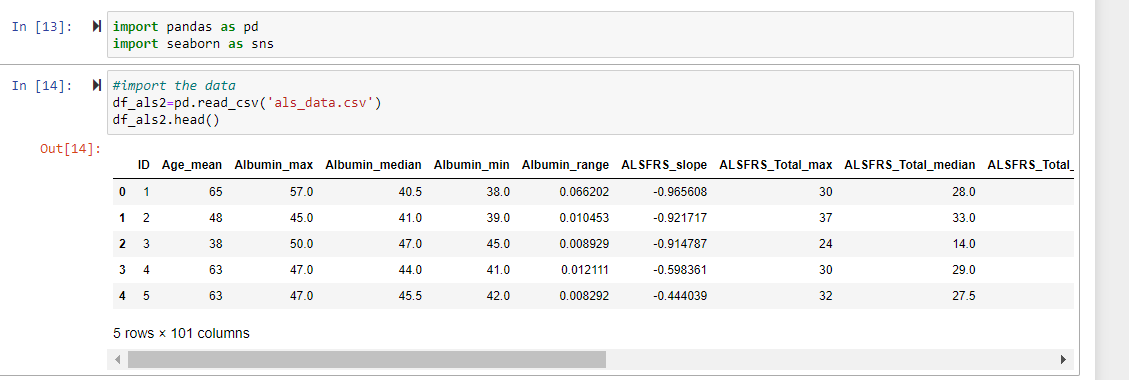
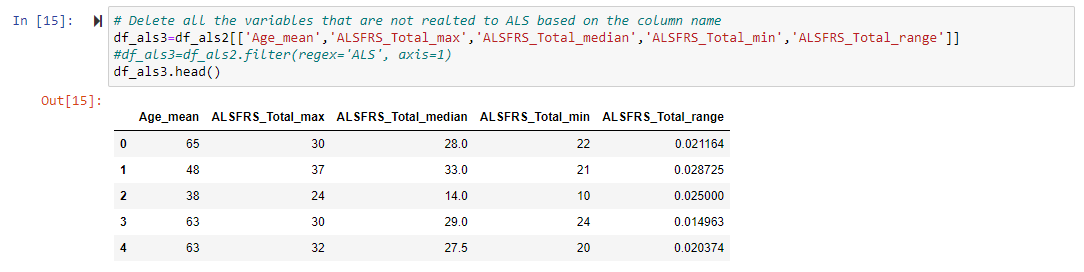
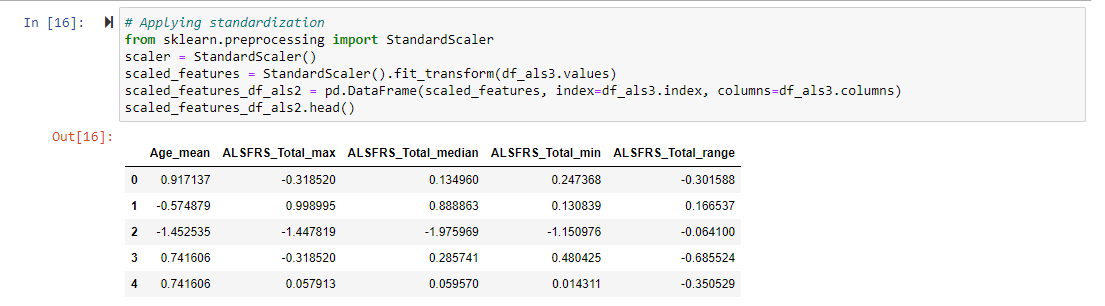
Load the data into the data frame.



Delete all the variables that are not related to ALS based on the column name

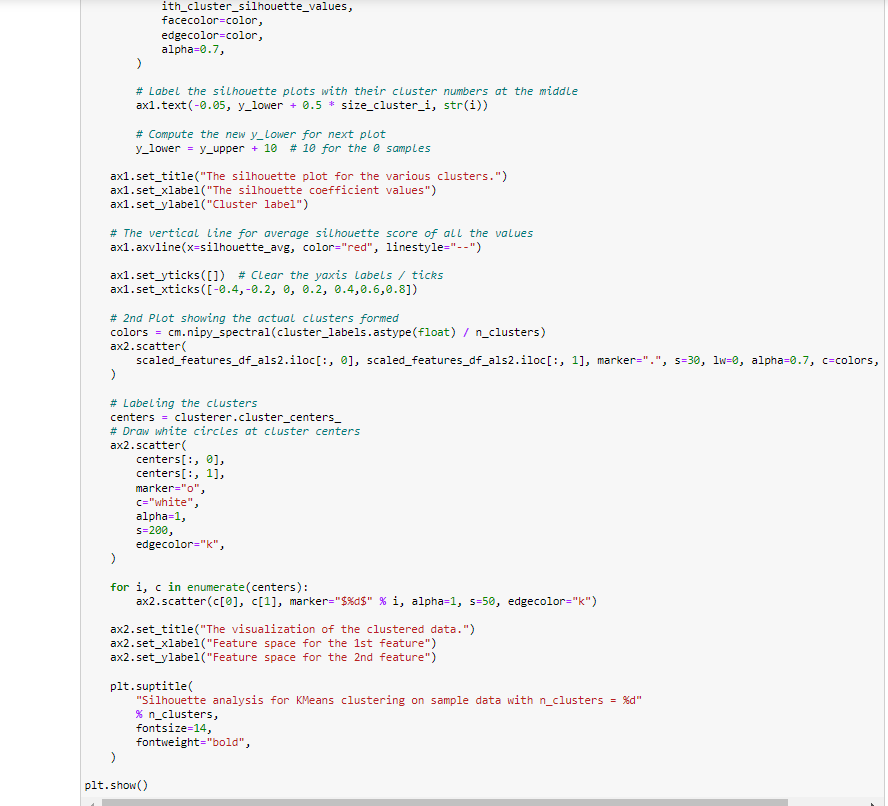


Apply standardization to data



Create a plot of the cluster silhouette score versus the number of clusters in a K-means cluster.





For n\_clusters = 2 The average silhouette\_score is : 0.3252456195301293

For n\_clusters = 3 The average silhouette\_score is : 0.29516078297360104

For n\_clusters = 4 The average silhouette\_score is : 0.24883051672684994

For n\_clusters = 5 The average silhouette\_score is : 0.26255252710997407

For n\_clusters = 6 The average silhouette\_score is : 0.2456132376507167

A picture containing logo

Description automatically generated

Logo

Description automatically generated with low confidence

A picture containing text

Description automatically generated

A picture containing text

Description automatically generated

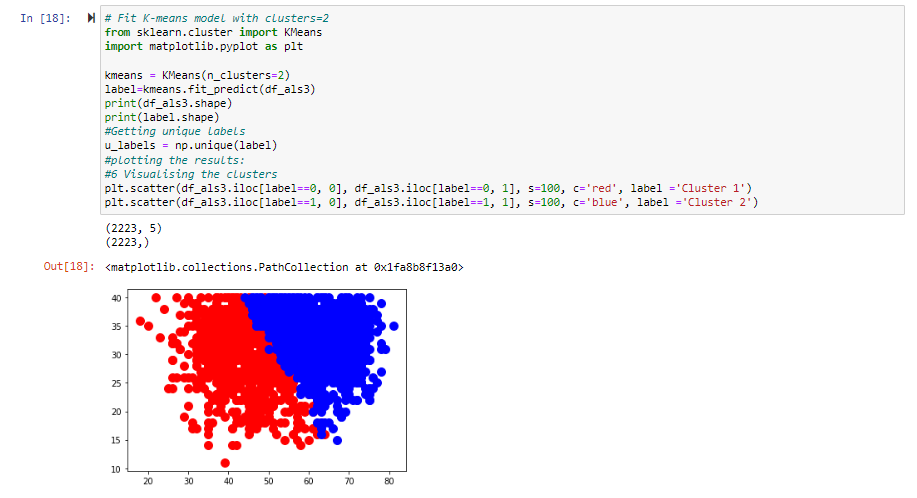
A picture containing chart

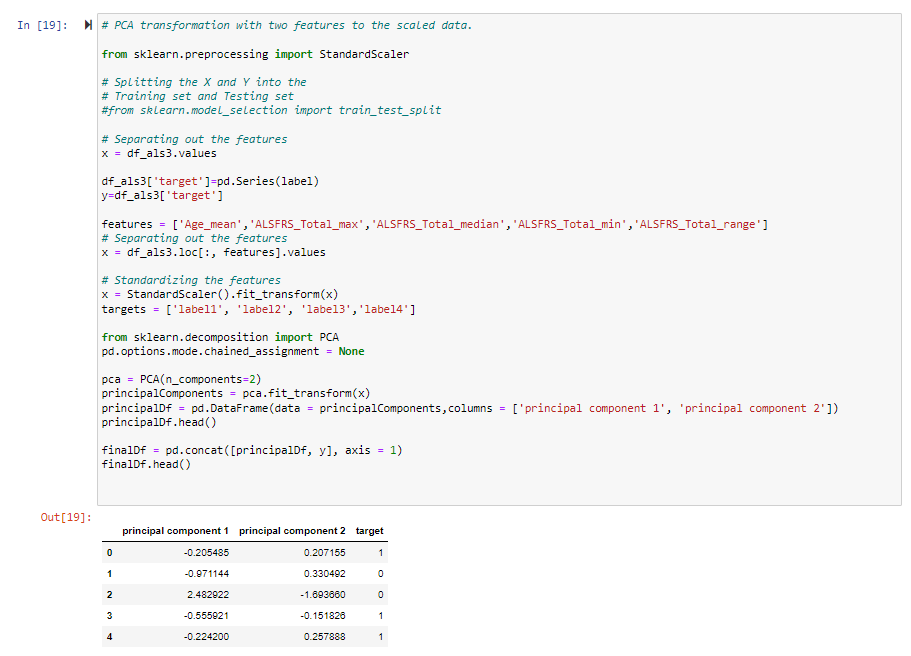
Description automatically generated

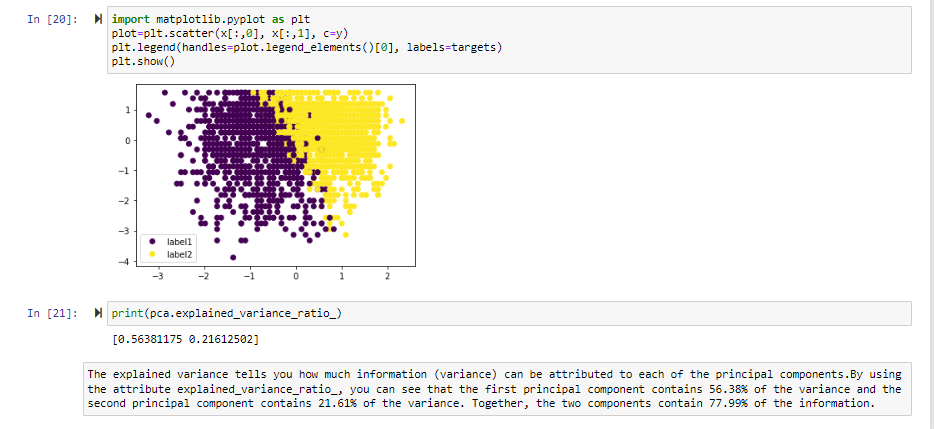
A picture containing chart

Description automatically generated

Based on the above plots. Cluster size=2 seems to be most appropriate than others as it stands well against all the three measuring criteria (scores below average Silhouette score, Wide fluctuations in the size of the plot, and non-uniform thickness).







The explained variance tells you how much information (variance) can be attributed to each of the principal components. By using the attribute explained\_variance\_ratio\_, you can see that the first principal component contains 56.38% of the variance and the second principal component contains 21.61% of the variance. Together, the two components contain 77.99% of the information.