There are seven clusters solution in the given dataset. The seven-cluster solution has the best average dissimilarity value with the highest value at 0.0797521 compared to the previous clusters in the given dataset.

However, the eight-cluster solution has the same highest value for the average dissimilarity while the six-cluster solution has its highest dissimilarity value at 0.08346951, confirming that the seven-cluster solution is the ideal solution for the given dataset.

The table below shows the R output of the 6-cluster solution.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster | size | max\_diss | av\_diss | diameter | separation |
| 1 | 116 | 0.13567353 | 0.06692318 | 0.22816617 | 0.030350687 |
| 2 | 40 | 0.1615634 | 0.07592112 | 0.21999561 | 0.037660043 |
| 3 | 127 | 0.17151395 | 0.06443843 | 0.29334507 | 0.037660043 |
| 4 | 109 | 0.30154089 | 0.08346951 | 0.37765927 | 0.024024973 |
| 5 | 57 | 0.21774149 | 0.06414895 | 0.27247656 | 0.030350687 |
| 6 | 51 | 0.17859673 | 0.07490735 | 0.25707695 | 0.024024973 |

The table below shows the R output of the 7-cluster solution.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Clusters | size | max\_diss | av\_diss | diameter | separation |
| 1 | 116 | 0.1356735 | 0.06692318 | 0.2281662 | 0.03035069 |
| 2 | 40 | 0.1615634 | 0.07592112 | 0.2199956 | 0.03766004 |
| 3 | 127 | 0.1715139 | 0.06443843 | 0.2933451 | 0.03766004 |
| 4 | 51 | 0.2916113 | 0.0797521 | 0.4354919 | 0.02867339 |
| 5 | 62 | 0.2020207 | 0.06950349 | 0.2829707 | 0.02947747 |
| 6 | 47 | 0.1245804 | 0.06384286 | 0.216808 | 0.2867339 |
| 7 | 57 | 0.2177415 | 0.06414895 | 0.2724766 | 0.03035069 |

The table below shows the R output for the 8-cluster solution.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster | size | max\_diss | av\_diss | diameter | separation |
| 1 | 116 | 0.13567353 | 0.06692318 | 0.22816617 | 0.030350687 |
| 2 | 38 | 0.1615634 | 0.07288083 | 0.21999561 | 0.033372372 |
| 3 | 69 | 0.11284556 | 0.05342591 | 0.1696449 | 0.01213971 |
| 4 | 51 | 0.29161126 | 0.0797521 | 0.43549191 | 0.028673389 |
| 5 | 62 | 0.20202069 | 0.06950349 | 0.28297069 | 0.029477469 |
| 6 | 47 | 0.12458035 | 0.06384286 | 0.21680804 | 0.028673389 |
| 7 | 57 | 0.21774149 | 0.06414895 | 0.27247656 | 0.030350687 |
| 8 | 60 | 0.14600806 | 0.06155769 | 0.20918386 | 0.01213971 |

The seven-cluster summary aggregate mean was estimated in the tables shown below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| cluster | fixed.acidity | volatile.acidity | citric.acid | residual.sugar | chlorides |
| 1 | 6.797413793 | 0.617068966 | 0.09482759 | 2.04612069 | 0.0768879 |
| 2 | 9.2475 | 0.377 | 0.4465 | 2.2975 | 0.1073 |
| 3 | 7.714173228 | 0.602952756 | 0.17393701 | 2.201968504 | 0.0855118 |
| 4 | 8.843137255 | 0.507941176 | 0.34215686 | 3.597058824 | 0.1106275 |
| 5 | 11.47903226 | 0.423709677 | 0.55177419 | 2.948387097 | 0.0965645 |
| 6 | 8.674468085 | 0.625851064 | 0.19914894 | 2.820212766 | 0.0935957 |
| 7 | 8.033333333 | 0.353333333 | 0.38631579 | 2.396491228 | 0.0723333 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| cluster | free.sulphur.dioxide | total.sulphur.dioxide | pH | sulphates | alcohol |
| 1 | 16.50862069 | 39 | 3.404052 | 0.58853448 | 10.79828 |
| 2 | 13.175 | 38.525 | 3.2105 | 0.78125 | 10.53833 |
| 3 | 14.48425197 | 53.42519685 | 3.324331 | 0.61047244 | 9.85643 |
| 4 | 29.09803922 | 86.70588235 | 3.278235 | 0.68117647 | 9.617647 |
| 5 | 9.983870968 | 29.58064516 | 3.164516 | 0.71209677 | 10.52581 |
| 6 | 11.36170213 | 32.74468085 | 3.326809 | 0.63425532 | 9.845745 |
| 7 | 16.5 | 40.33333333 | 3.292982 | 0.71964912 | 11.94561 |

To determine how the cluster characteristics that have been extracted scaled, we scale the mean output of the continuous variable such that the mean is zero across the whole dataset. The table below shows the scaled mean table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster | fixed.acidity | volatile.acidity | citric.acid | residual.sugar | chlorides |
| 1 | -1.277943235 | 0.987853097 | -1.3352362 | -1.050945592 | -1.0305924 |
| 2 | 0.381572916 | -1.05751082 | 0.81188666 | -0.586867572 | 1.0667888 |
| 3 | -0.656994844 | 0.867584383 | -0.8522365 | -0.763230804 | -0.4358428 |
| 4 | 0.10768602 | 0.058095093 | 0.17482353 | 1.812282481 | 1.29626786 |
| 5 | 1.893055987 | -0.659548974 | 1.45463432 | 0.614752381 | 0.32641162 |
| 6 | -0.006558618 | 1.062675709 | -0.698306 | 0.378126345 | 0.12166905 |
| 7 | -0.440818226 | -1.259148487 | 0.44443421 | -0.404117238 | -1.3447021 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster | free.sulphur.dioxide | total.sulphur.dioxide | pH | sulphates | alcohol |
| 1 | 0.10050201 | -0.345683791 | 1.49153949 | -1.2709008 | 0.44266154 |
| 2 | -0.4266553 | -0.369976404 | -0.952213393 | 1.55035189 | 0.11524691 |
| 3 | -0.2196186 | 0.392054559 | 0.484994794 | -0.9497408 | -0.7436544 |
| 4 | 2.09131156 | 2.094106757 | -0.096998483 | 0.08532862 | -1.0444177 |
| 5 | -0.9312799 | -0.827411657 | -1.532798338 | 0.53798542 | 0.09946848 |
| 6 | -0.7133985 | -0.665595462 | 0.516279122 | -0.6015721 | -0.7571139 |
| 7 | 0.09913879 | -0.277494002 | 0.089196808 | 0.64854779 | 1.8878091 |

For each variable, all the clusters that has a scaled mean below zero are below average, while those above zero are above average. For Instance, For the Fixed Acidity, clusters 1,3,6, and 7 are below average, while clusters 2,4, and 5 are above average.

For the Volatile Acidity, clusters 2,5, and 7 are below average, while clusters 1,3,4, and 6 are above average. Similarly for other clusters in each variable.

Furthermore, the ordinal variable in the dataset gives more insight about the clusters as shown below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Density | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 |
| High | 0 | 0 | 0 | 51 | 62 | 47 | 0 |
| Low | 116 | 0 | 0 | 0 | 0 | 0 | 57 |
| Medium | 0 | 40 | 127 | 0 | 0 | 0 | 0 |

The table above shows that clusters 4,5, and 6 have a high density with sample sizes of 51, 62, and 47 respectively, while clusters 1 and 7 have a Low density with sample sizes of 116 and 57 respectively. Similarly, clusters 2 and 3 have a medium density with sample sizes of 40 and 127, respectively.