As a good cluster solution is one where all the cases are near each other based on the "average dissimilarities" the following table output below shows cluster information for various proposed cluster solutions.

Table showing the cluster information for a 2-cluster solution:

CLUSTERS	SIZE	MAX_DISS	AVG_DISS	DIAMETER	SEPARATION
1	248	0.3116525	0.1249108	0.476838	0.01371724
2	252	0.3200633	0.1226568	0.504733	0.01371724

A 2-cluster solution shows the highest average dissimilarity in cluster 1 with 0.1249108. We then proceed to observe if the average dissimilarity in the clusters can be reduced by observing a 3-cluster solution.

Table showing the cluster information for a 3-cluster solution:

CLUSTERS	SIZE	MAX_DISS	AVG_DISS	DIAMETER	SEPARATION
1	166	0.2589343	0.09350649	0.4270656	0.1025366
2	169	0.2709354	0.09588464	0.4015534	0.1025366
3	165	0.2712634	0.09566797	0.4013531	0.1027971

Compared to a 2-cluster solution, the highest average dissimilarity observed in the 3-cluster solution is 0.09588464 which is smaller. Therefore, ideally, a 3-cluster solution would be preferred to the 2-cluster solution.

Table showing the cluster information for a 4-cluster solution:

CLUSTERS	SIZE	MAX_DISS	AVG_DISS	DIAMETER	SEPARATION
1	164	0.2589343	0.09274438	0.4270656	0.05872166
2	108	0.1609749	0.07890605	0.2711364	0.03137963
3	165	0.2712634	0.09566797	0.4013531	0.10279713
4	63	0.2613279	0.07365667	0.3534804	0.03137963

The highest average dissimilarity observed in the 4-cluster solution is 0.09566797 which isn't too far-fetched from the 3-cluster solution.

Table showing the cluster information for a 5-cluster solution:

CLUSTERS	SIZE	MAX_DISS	AVG_DISS	DIAMETER	SEPARATION
1	163	0.2589343	0.09252212	0.4270656	0.03832484
2	108	0.1609749	0.07890605	0.2711364	0.03137963
3	105	0.2403815	0.08192849	0.3406916	0.03467192
4	63	0.2613279	0.07365667	0.3534804	0.03137963
5	61	0.1988843	0.07859031	0.2958137	0.03467192

Table showing the cluster information for a 6-cluster solution:

CLUSTERS	SIZE	MAX_DISS	AVG_DISS	DIAMETER	SEPARATION
1	126	0.2652404	0.08110279	0.4270656	0.04842375
2	108	0.1609749	0.07890605	0.2711364	0.03137963
3	40	0.2369799	0.08191394	0.2850844	0.04842375
4	105	0.2403815	0.08192849	0.3406916	0.03467192
5	61	0.2613279	0.07123510	0.3308637	0.03137963
6	60	0.1988843	0.07802711	0.2776034	0.03467192

Table showing the cluster information for a 7-cluster solution:

CLUSTERS	SIZE	MAX_DISS	AVG_DISS	DIAMETER	SEPARATION
1	126	0.2652404	0.08110279	0.4270656	0.04842375
2	108	0.1609749	0.07890605	0.2711364	0.03137963
3	40	0.2369799	0.08191394	0.2850844	0.04842375
4	58	0.2145209	0.07129455	0.2958819	0.02471494
5	55	0.1905956	0.07482954	0.2764029	0.02471494
6	61	0.2613279	0.07123510	0.3308637	0.03137963
7	52	0.1988843	0.07687768	0.2555803	0.04008566

As observed from the 3-cluster solution downwards to a 7-cluster solution, the average dissimilarities only improved marginally and will continue to improve only marginally however, having many clusters solution isn't ideal as such a 3-cluster solution would be

preferred because the average dissimilarities in subsequent cluster solutions only decreased just marginally.

In conclusion, since a fewer number of clusters is to be preferred and the average dissimilarity in different cluster solutions are not significantly larger than each other, a 3-cluster solution would be ideal/preferred for the given dataset.

Output showing the characteristics of the clusters you have decided to extract, and, using this, report on how the clusters differ from one another

In the table output below shows the mean value of the continuous variables of each of the clusters extracted from the dataset

CLUSTERS	FIXED.ACIDITY	VOLATILE.ACIDITY	CITRIC.ACID	RESIDUAL.SUGAR	CHLORIDES
1	8.027711	0.5423795	0.2686145	2.432530	0.09656024
2	7.242604	0.5081065	0.2118935	2.106509	0.07207101
3	9.683030	0.5275758	0.3576364	3.069697	0.08886667

CLUSTERS	FREE.SULPHUR.DIOXIDE	TOTAL.SULPHUR.DIOXIDE	PH	SULPHATES	ALCOHOL
1	15.78916	51.91566	3.313916	0.6750602	10.18936
2	16.26923	42.56509	3.359586	0.6324852	11.20750
3	15.45455	49.64848	3.264182	0.6661818	10.05606

Additionally, the characteristics of the clusters extracted based on the categorical/ordinal variable “Density” in the dataset can be seen in the table below

	Density		
Cluster	High	Low	Medium
1	0	0	166
2	0	169	0
3	165	0	0

The table above shows the cluster solution classification by the “Density” variable which is an ordinal variable. The table shows the following:

- All 166 observations in Cluster 1 have a medium density
- All 169 observations in Cluster 2 have a low density

- All 165 observations in Cluster 3 have a high density

Furthermore, given that the dataset has ten (10) continuous variables, one assumption is that if the average value (mean) of a variable ordered by clusters differs significantly from one another, that variable is likely important in creating the clusters solution.

In that regard, the table below shows how the 3-cluster solution differs from one another in relation to all the variables in the dataset

N.B: the dataset is scaled which indicates that each continuous variable is averaged at a mean of 0. If the mean is above 0 that indicates the cluster is above average and vice versa

Variable	*Cluster Solution Insights/Interpretation*
Fixed.acidity	• Based on the mean score, each of the cluster (1 to 3) is significantly above average • The mean score of each cluster differs slightly from each other so we can conclude that this variable contributes to the cluster solution
Volatile.acidity	• Each cluster (1 to 3) is above average • The mean score of each cluster doesn't differ from each other therefore, the variable doesn't contribute to the cluster solution
Citric.acid	• Each cluster is slightly above average • Cluster 1 with a mean of approx. 0.27 and cluster 2 with a mean of 0.21 share some similarities, while cluster 3 with a mean of 0.36 differs from cluster 1 & 2 • We can conclude the variable slightly contributes to the cluster solution
Residual.sugar	• Each cluster is above average • Cluster 1 with a mean of approx. 2.43 and cluster 2 with a mean of 2.11 share some similarities, while cluster 3 with a mean of 3.07 differs from cluster 1 & 2
Chlorides	• Each cluster is marginally above average • Each cluster differs from each other with slightly different mean values • Hence, the variable contributes to the cluster solution
Free.sulphur.dioxide	• Each cluster is clearly above average • Cluster 1 & 3 with a mean of 15.79 and 15.45 respectively share some similarities, while cluster 2 with 16.27 differs
	• Each cluster is clearly above average

Total.sulphur.dioxide	• Based on the mean, each cluster differs significantly from the other • Hence, the variable highly contributes to the cluster solution
pH	• Each cluster is above average • Each cluster have slightly similar mean ranging between 3.26 to 3.36 • Hence, the variable doesn't contribute to the cluster solution
sulphates	• Each cluster is slightly above average • Each cluster have slightly similar mean ranging between 0.63 to 0.68 • Hence, the variable doesn't contribute to the cluster solution
Alcohol	• Each cluster is above average • Cluster 1 & 3 with a mean of 10.19 and 10.06 respectively share some similarities, while cluster 2 with a mean of 11.21 differs

In conclusion, while the 3 clusters in the cluster solution share some similarities in some of the continuous variable, they clearly differ in other variables such as the "fixed.acidity", "total.sulphur.dioxide" and "density".