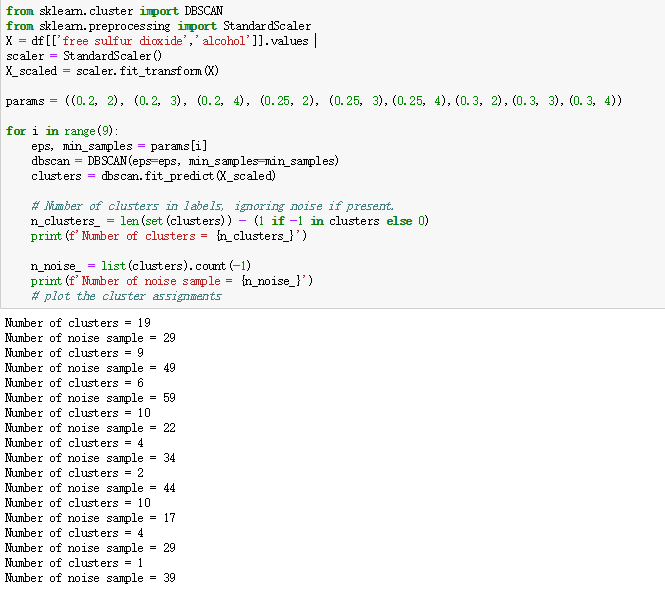
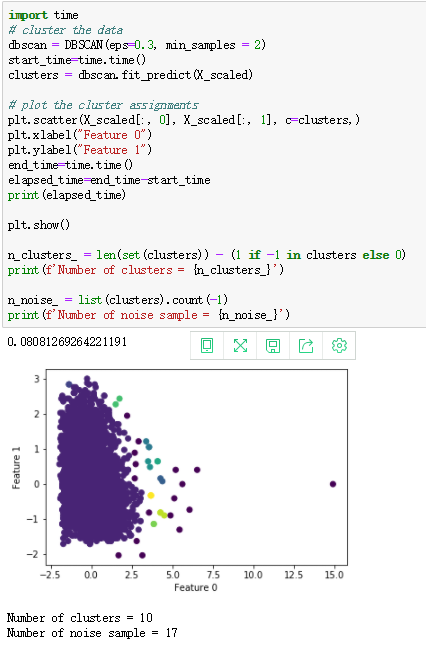
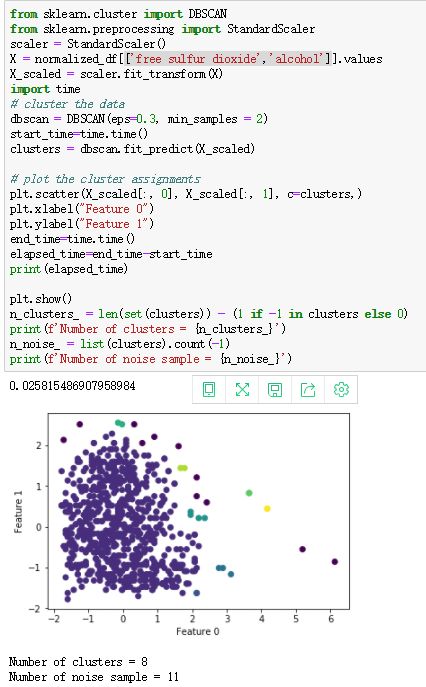
DBSCAN

First we choose ‘free sulfur dioxide' and 'alcohol’ two features.To get the appropriate eps and min\_samples we write a for loop.



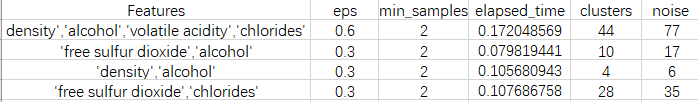
As we can see eps=0.3 and min\_samples=2 has more clusters(10) and less noise(17)



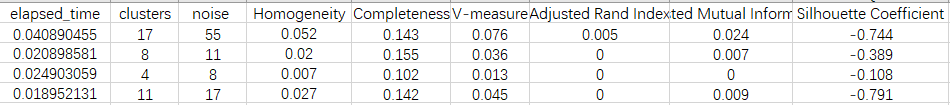


After we get normalized data and run DBSCAN again (with same eps and min\_samples)We can find that clusters and noise reduced and there is a large decentralized cluster. Clusters are easier to identify than before data balancing, indicating that data balancing has had some effect.

We make a table of different features (before databalancing) :



After databalancing:



Conclusion：From this table, we can see that, if there are more features, the appropriate EPS will be larger, the computing time will be increased, clusters and noises will be larger, but appropriate min\_samples are generally the same. Data Balancing can reduce noise most of the time, but sometimes noise will increase a little.According to the data of EVALUATION METRICS, we can see that It is considered that DBSCAN is not suitable for this dataset.