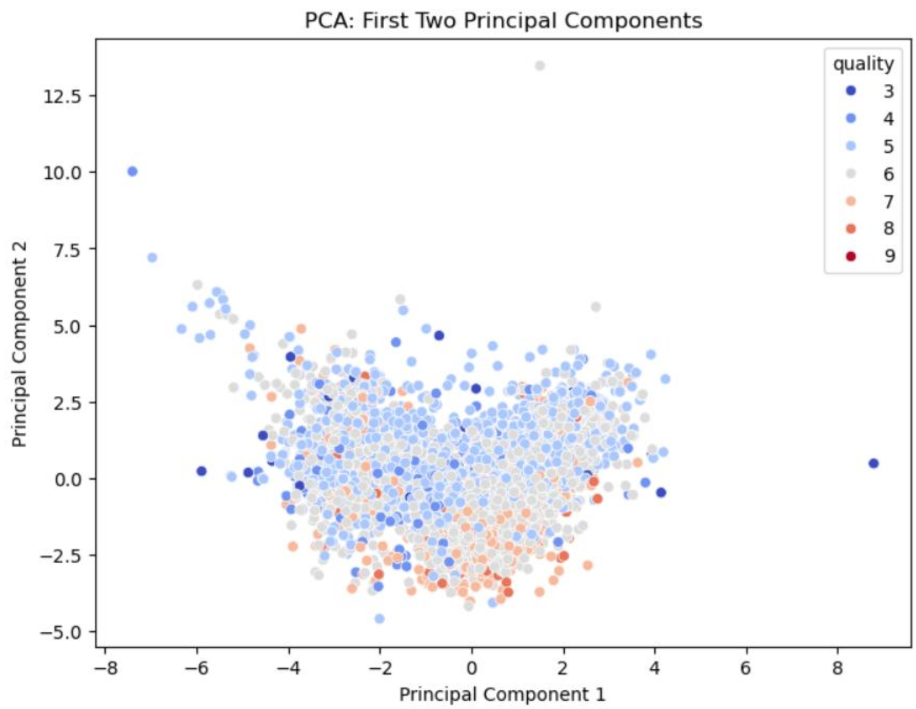
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

```
White Wine Dataset:
```

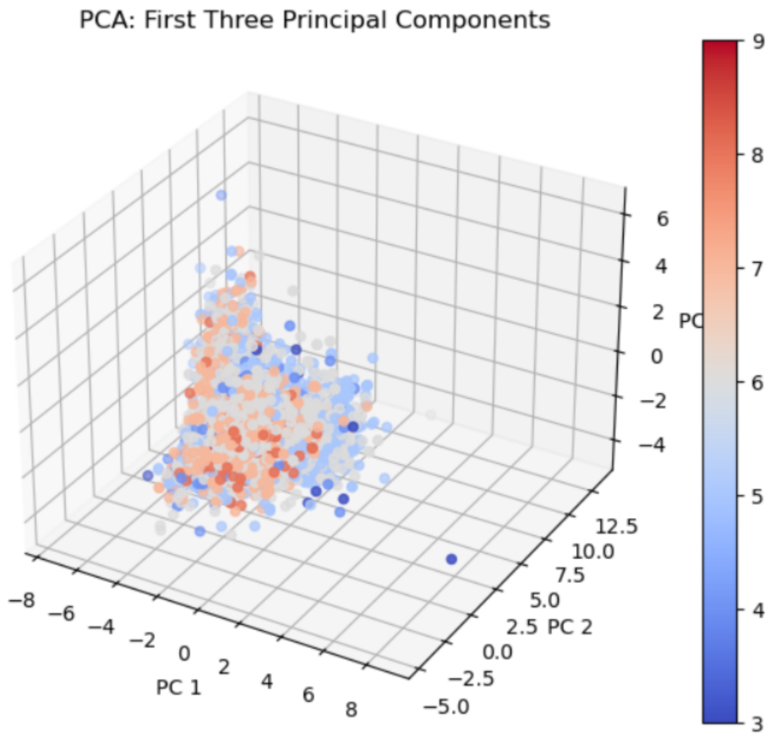
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
1	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
2	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
3	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
4	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6

```
Red Wine Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
```

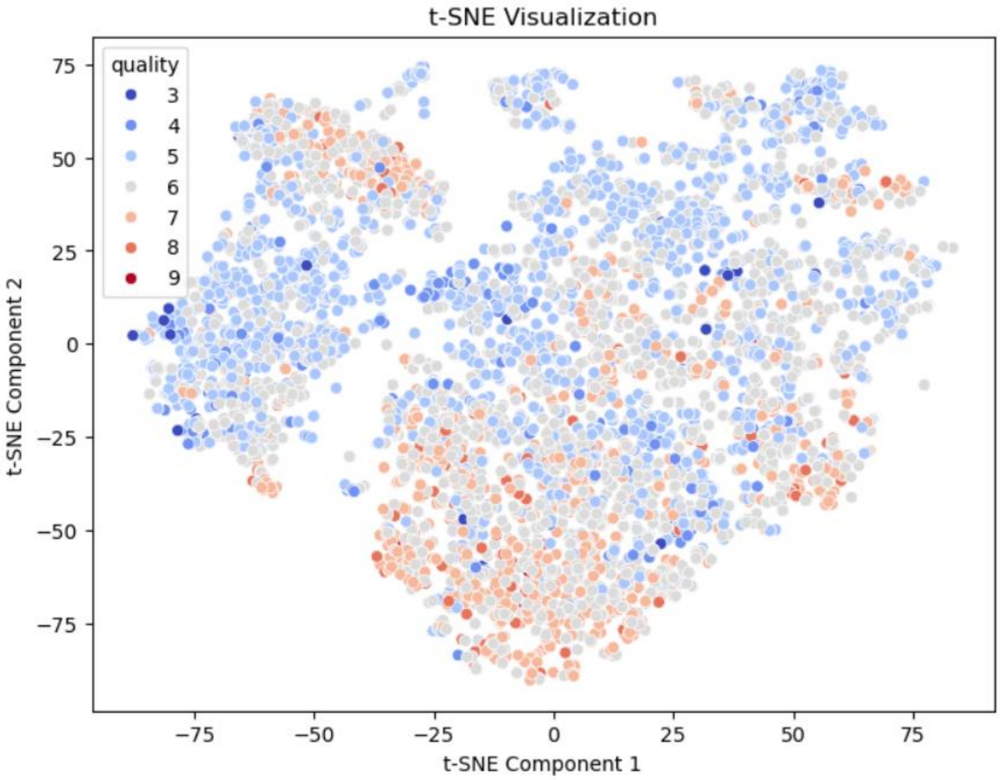
```
Missing values in dataset:
 fixed acidity           0
volatile acidity        0
citric acid             0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      0
sulphates               0
alcohol                 0
quality                 0
dtype: int64

Normalized Data (first 5 rows):
    fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  \
0        0.142473          2.188833    -2.192833       -0.744778   0.569958
1        0.451036          3.282235    -2.192833       -0.597640   1.197975
2        0.451036          2.553300    -1.917553       -0.660699   1.026697
3        3.073817         -0.362438     1.661085       -0.744778   0.541412
4        0.142473          2.188833    -2.192833       -0.744778   0.569958

   free sulfur dioxide  total sulfur dioxide   density        pH  sulphates  \
0            -1.100140             -1.446359  1.034993  1.813090   0.193097
1            -0.311320             -0.862469  0.701486 -0.115073   0.999579
2            -0.874763             -1.092486  0.768188  0.258120   0.797958
3            -0.762074             -0.986324  1.101694 -0.363868   0.327510
4            -1.100140             -1.446359  1.034993  1.813090   0.193097

    alcohol
```


PCA: First Two Principal Components
quality
3
4
5
6
7
8
9
Principal Component 2
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−8
−6
−4
−2
0
2
4
6
8
Principal Component 1

[ ]:



2.4 Discussion on PCA & Information Loss PCA reduces dimensions while preserving as much variance as possible. Explained v
principal component retains. Trade-off: Reducing dimensions means some information is lost, but visualization and computati
variance: If the first two PCs explain ~80% of the variance, PCA is effective. If they explain less, we might lose significant inform


t-SNE Visualization
quality
3
4
5
6
7
8
9
75
50
25
0
−25
−50
−75
t-SNE Component 2
−75
−50
−25
0
25
50
75
t-SNE Component 1

[ ]:

[ ]: