'*winequality-red.csv'* and '*winequality-white.csv'* datasets include a number of measurable chemical features of any wine, such as fixed acidity, volatile acidity, citric acid, and residual sugar:

```
      fixed acidity  volatile acidity  ...  alcohol  quality
0               7.4             0.700  ...      9.4        5
1               7.8             0.880  ...      9.8        5
2               7.8             0.760  ...      9.8        5
3              11.2             0.280  ...      9.8        6
4               7.4             0.700  ...      9.4        5
...             ...               ...  ...      ...      ...
1594            6.2             0.600  ...     10.5        5
1595            5.9             0.550  ...     11.2        6
1596            6.3             0.510  ...     11.0        6
1597            5.9             0.645  ...     10.2        5
1598            6.0             0.310  ...     11.0        6

[1599 rows x 12 columns]
      fixed acidity  volatile acidity  ...  alcohol  quality
0               7.0              0.27  ...      8.8        6
1               6.3              0.30  ...      9.5        6
2               8.1              0.28  ...     10.1        6
3               7.2              0.23  ...      9.9        6
4               7.2              0.23  ...      9.9        6
...             ...               ...  ...      ...      ...
4893            6.2              0.21  ...     11.2        6
4894            6.6              0.32  ...      9.6        5
4895            6.5              0.24  ...      9.4        6
4896            5.5              0.29  ...     12.8        7
4897            6.0              0.21  ...     11.8        6

[4898 rows x 12 columns]
```

The datasets were loaded and merged into one pandas dataframe, log-transformed, and scaled:

```
18  def get_transformed_dataframe():
19      df1 = pd.read_csv('winequality-red.csv', delimiter=';')
20      df2 = pd.read_csv('winequality-white.csv', delimiter=';')
21      df3 = pd.concat([df1, df2])
22      df3 = log_transform(df3)
23      df3 = df3.apply(scale, axis=0)
24      return df3, len(df1)
25
26  def log_transform(df):
27      df[['citric acid']] += 1
28      df = df.transform(lambda x: np.log(x))
29      return df
30
31  def scale(df):
32      return (df - df.mean()) / df.std()
```

1st and 2nd Principal Components were built from the scaled attributes, using a fully connected feedforward Neural Network autoencoder with two units in the hidden layer and a linear activation functions:

```
34  def build_model(x):
35      encoder = models.Sequential([layers.Dense(2, input_shape=[12])])
36      decoder = models.Sequential([layers.Dense(12, input_shape=[2])])
37      autoencoder = models.Sequential([encoder, decoder])
38      autoencoder.compile(loss='mse', optimizer=optimizers.SGD(lr=0.1))
39      autoencoder.fit(x, x, epochs=20)
40      return encoder.predict(x)
```

Observations of the two Principal Components are plotted in the plane. Red dots represent red wines, yellow dots represent white wine in the two distinct clusters shown in the PCA 2D projection below: