```
'''
Data analysis on wine quality using Python numpy package
'''

import csv
import numpy as np
import statistics as st

# on Windows do, with open('C:\downloads\winequality-red.csv', 'r') as f:
with open('/users/staff/downloads/winequality-red.csv', 'r') as f:
    wines = list(csv.reader(f, delimiter=';'))

    # convert wines data into numpy array
    wines = np.array(wines[1:], dtype=np.float) # skips the header wines[1:]


print("\nDisplay data")
print(wines)

print("\nThere are", wines.shape[0], "rows of wine data")
print("\nDisplay numbers of rows and columns")
print(wines.shape)

print("\nDisplay data type")
print(wines.dtype)

print("\nDisplay row 2 col 3")
print(wines[2,3])

print("\nDisplay first three items from fourth column")
print(wines[0:3,3])

print("\nDisplay entire row 3 - third wine data")
third_wine = wines[3,:]
print(third_wine)

print("\nDisplay row 3 col 1")
print(third_wine[1])

wines_quality = wines[:,11] # select col 12 only
print("\nDisplay col 12")
print(wines_quality)

print("\nMin wine quality score:", wines[:,11].min())
print("Max score:", wines[:,11].max())
print("Average score:", wines[:,11].mean())
```

```
print("Median score:", st.median(wines[:,11]))
print("Std. Dev.:", wines[:,11].std())

# Determine wines where the quality is > 7
high_quality = wines[:,11] > 7
print("\nHigh quality wines, score > 7")
print("There are", wines[high_quality,:].shape[0]) # print how many of them
print(wines[high_quality,:][:2,:]) # print first two lines only

# Determine wines where the quality is > average
average = wines[:,11].mean()
avg_quality = wines[:,11] > average
print("\nNumber of wines quality greater than average...")
print("There are", wines[avg_quality,:].shape[0]) # print how many of them
```

Test-Run

```
Display data
[[ 7.4   0.7   0.   ... 0.56  9.4   5.  ]
 [ 7.8   0.88  0.   ... 0.68  9.8   5.  ]
 [ 7.8   0.76  0.04 ... 0.65  9.8   5.  ]
 ...
 [ 6.3   0.51  0.13 ... 0.75  11.    6.  ]
 [ 5.9   0.645 0.12 ... 0.71  10.2   5.  ]
 [ 6.    0.31  0.47 ... 0.66  11.    6.  ]]

There are 1599 rows of wine data

Display numbers of rows and columns
(1599, 12)

Display data type
float64

Display row 2 col 3
2.3

Display first three items from fourth column
[1.9 2.6 2.3]

Display entire row 3 - third wine data
[11.2   0.28  0.56  1.9   0.075 17.   60.    0.998 3.16  0.58
 9.8   6.  ]

Display row 3 col 1
```

0.28

Display col 12
[5. 5. 5. ... 6. 5. 6.]

Min wine quality score: 3.0
Max score: 8.0
Average score: 5.6360225140712945
Median score: 6.0
Std. Dev.: 0.8073168769639513

High quality wines, score > 7
There are 18
[[7.9 0.35 0.46 3.6 0.078 15. 37. 0.9973 3.35
  0.86 12.8 8.]
 [10.3 0.32 0.45 6.4 0.073 5. 13. 0.9976 3.23
  0.82 12.6 8.]]

Number of wines quality greater than average...
There are 855