

葡萄酒品質分析



第二組

B11108004顏詩妮 B11108032陳又濤
B11108049詹芸婷 B11108060王子晏



目錄

Contents

- 1 界定問題 決定目標
- 2 探索性資料分析
- 3 分析及建立模型
- 4 結果呈現
- 5 管理意涵
- 6 評估成效

產業介紹：葡萄酒



- 葡萄酒至少已具有**8000年的歷史**
- 生產地有**新世界**、**舊世界**之分
- 全球在2019年
創造了3,700億美元的收入

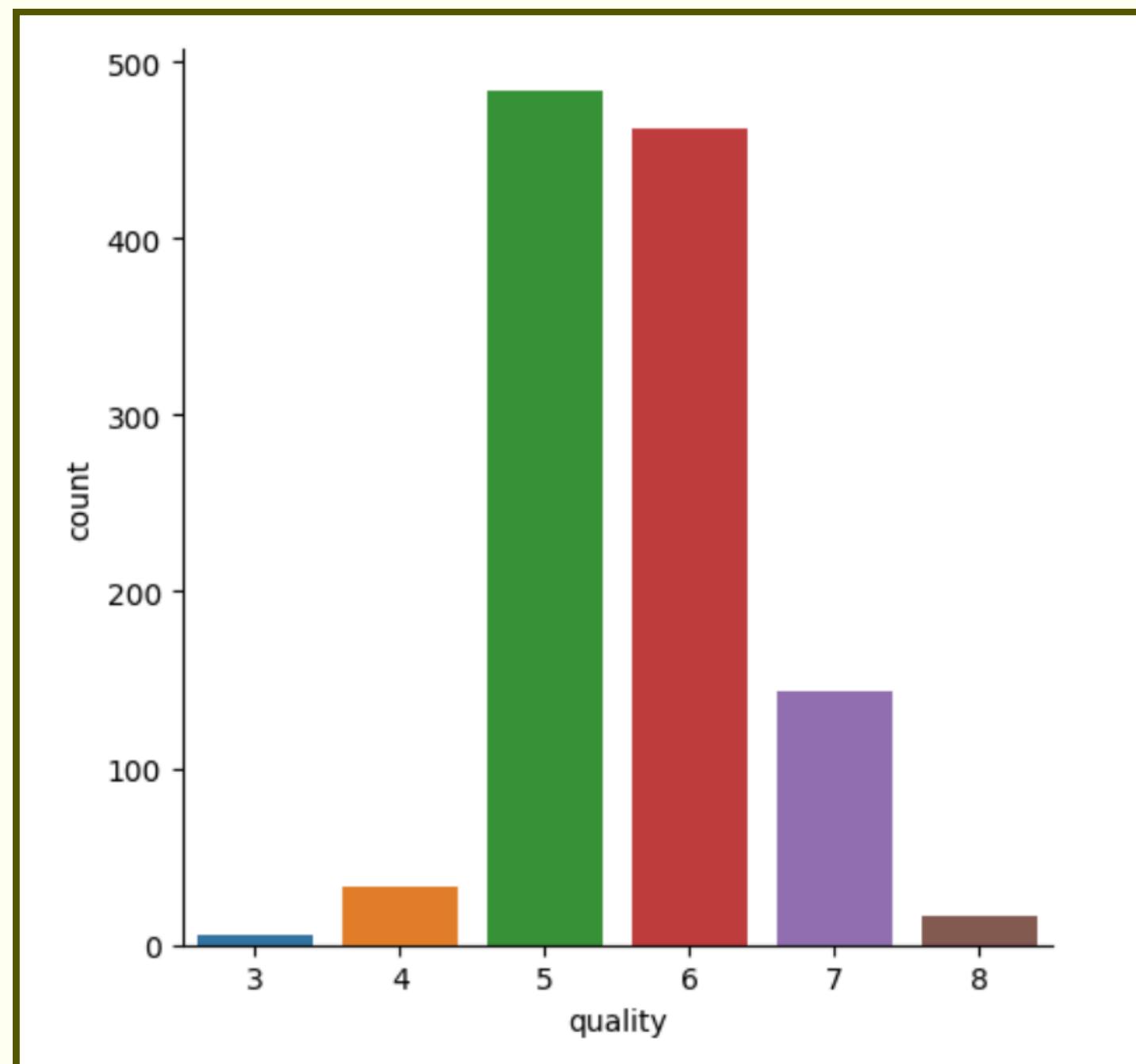
OI

界定問題 決定目標

角色定位：葡萄酒產品管理顧問

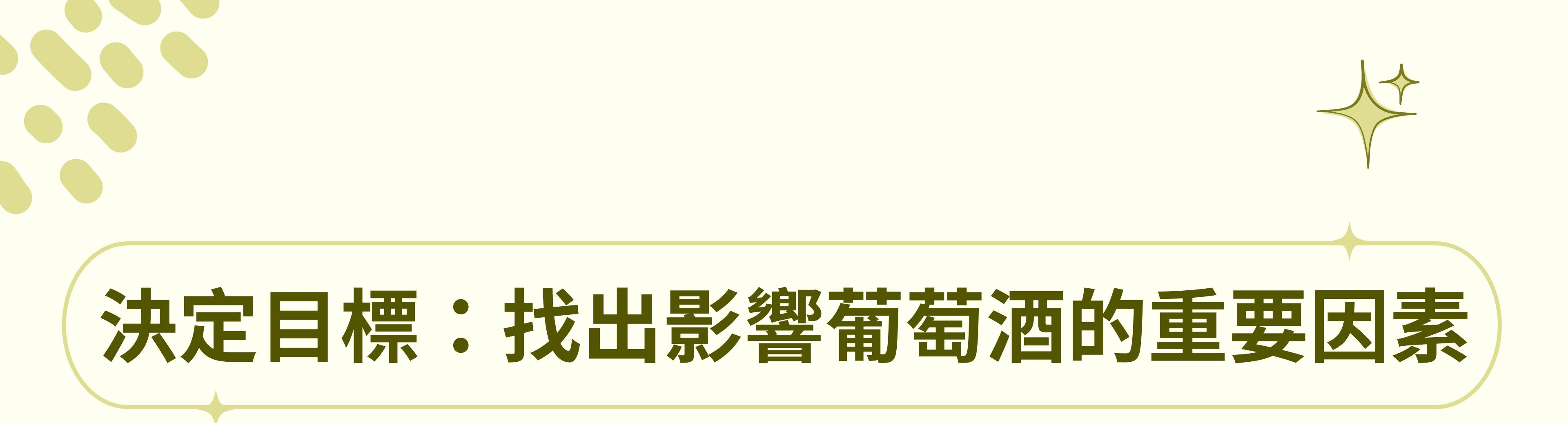
主要以分析葡萄酒的品質來進行產品管理

界定問題：葡萄酒的品質偏低



左圖為葡萄酒的品質分布
可以看出在滿分10分下
葡萄酒大概分布在5分-6分

可分析出品質偏低！



決定目標：找出影響葡萄酒的重要因素

根據葡萄酒的各項成分數據，分析出影響葡萄酒的重要因素，讓我們藉由主要控管重要因素來確保葡萄酒的品質。



02

探索性資料分析

資料介紹



共有1143列，12欄



11項因子影響著葡萄酒品質



無任何缺值

資料介紹

FIXED ACIDITY 非揮發性酸

葡萄酒中的非揮發性一般為有機酸，這些物質無色無味，像是酒石酸、蘋果酸、檸檬酸、琥珀酸

VOLATILE ACIDITY 揮發性酸

當葡萄酒暴露於空氣中，會氧化產生醋酸，因為是氣態，含量較多的時候我們可以明顯聞到一股刺鼻的酸味



CITRIC ACID 檸檬酸

一般檸檬酸含量只有酒石酸含量的1/20，檸檬酸能去除葡萄酒中多余的鐵和銅

RESIDUAL SUGAR 剩餘糖分

成酒中沒被酵母轉化成酒精，殘留下來的糖分就是剩餘糖份

資料介紹

CHLORIDES

氯化物

葡萄酒中鹽的含量

FREE SULFUR DIOXIDE

游離二氧化硫

它可以防止微生物的生長和葡萄酒的氧化，其中未解離的和分子態的二氧化硫具有最强的殺真菌能力



DENSITY

密度

酒精的密度接近水的密度，密度取決於酒精的百分比和糖含量

TOTAL SULFUR DIOXIDE

總二氧化硫

在葡萄酒中，二氧化硫 (SO_2) 以游離形式和結合形式存在，過量的游離 SO_2 可能會對感官產生影響

資料介紹

PH 酸鹼值

大部分葡萄酒的PH值在2.9至3.9之間

ALCOHOL 酒精

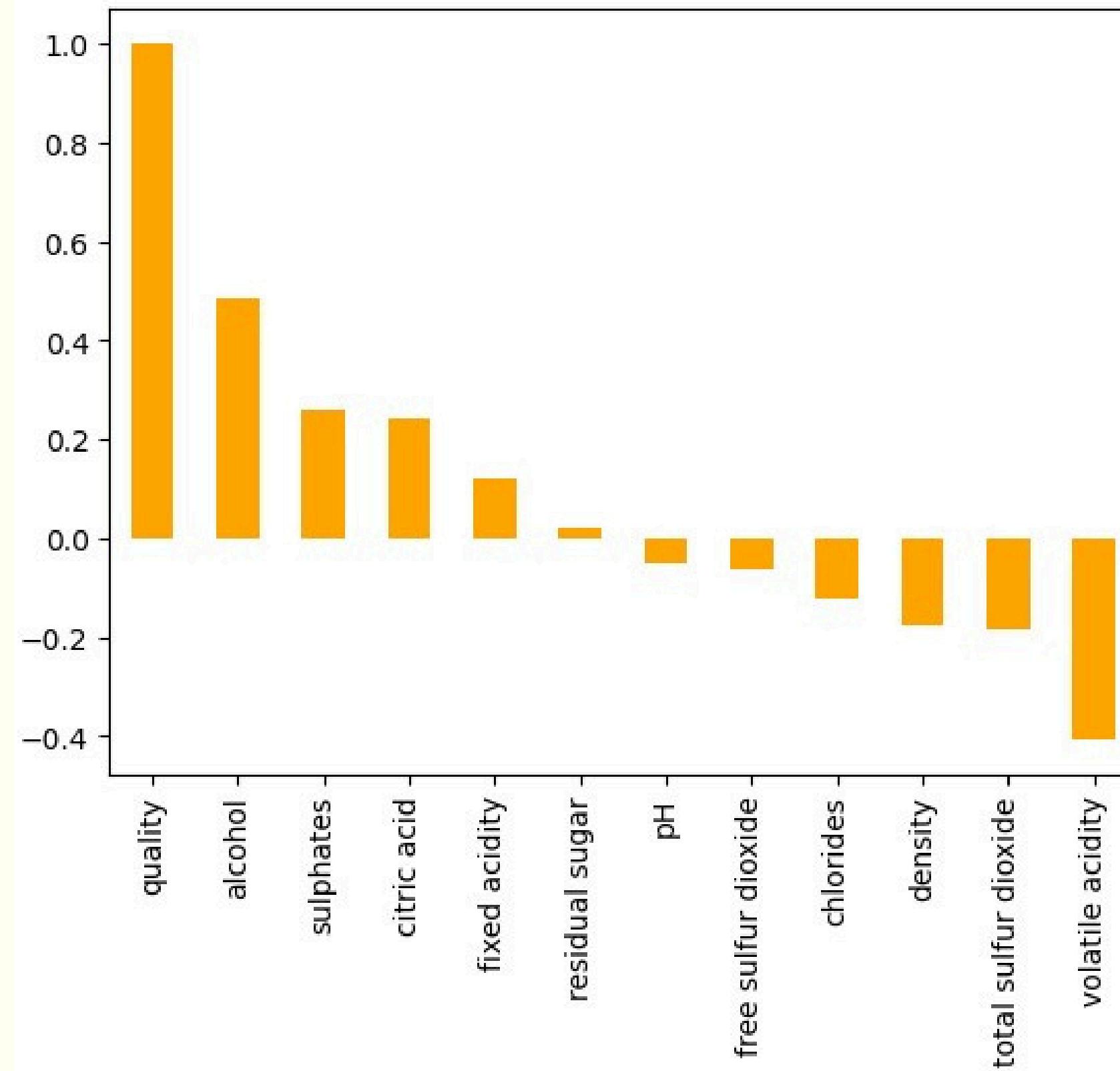
葡萄酒中的酸性元素



SULPHATES 亞硫酸鹽

發酵過程中會出現的天然副產品，也是一種葡萄酒添加劑，可以促進二氧化硫氣體 (SO_2) 水平，作為抗菌劑和抗氧化劑。

因素與品質之相關變數

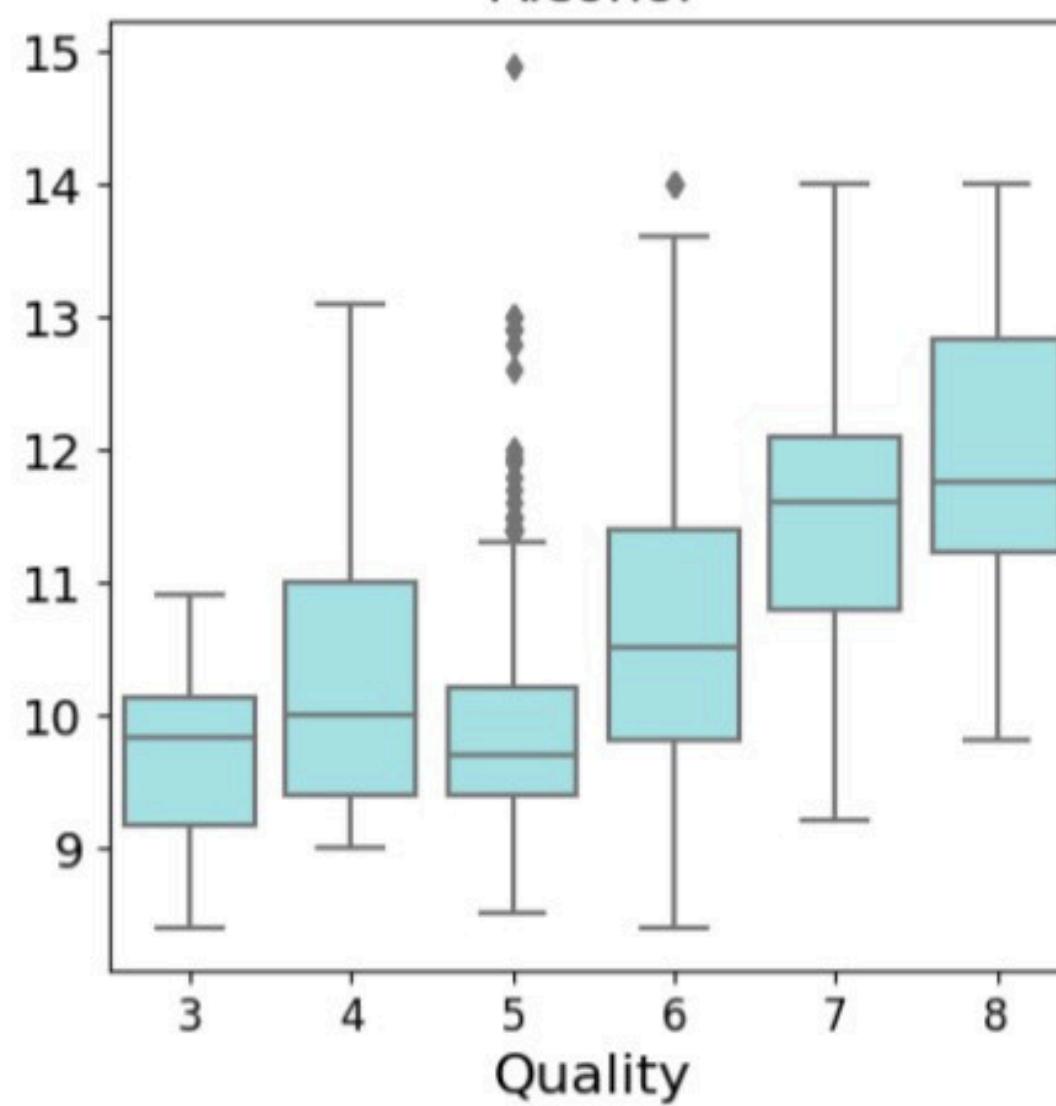


分析葡萄酒品質與其他成分之間的相關性，了解哪些成分可能影響葡萄酒的品質

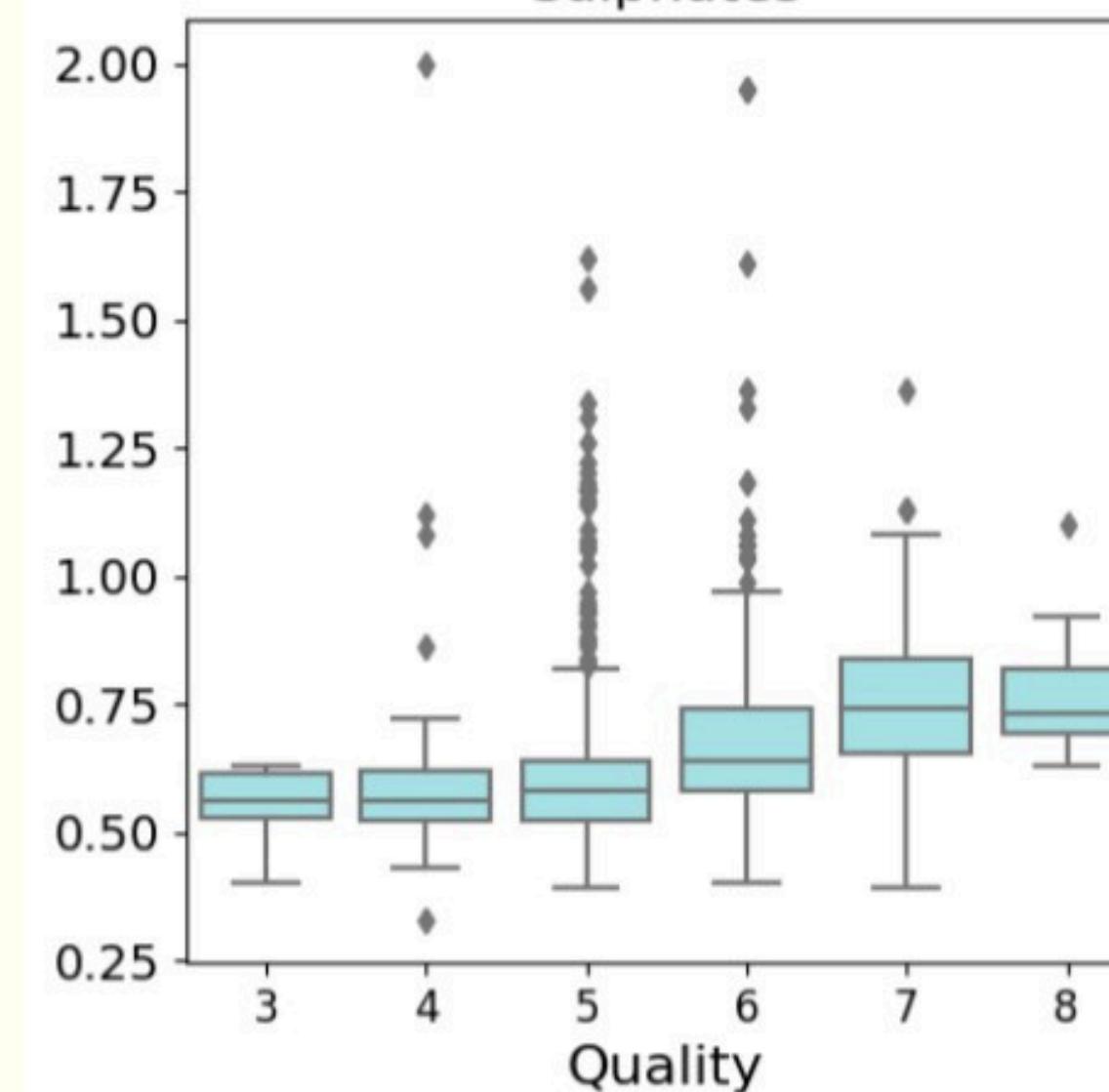
不同理化性質在品質中的分布情形

正相關

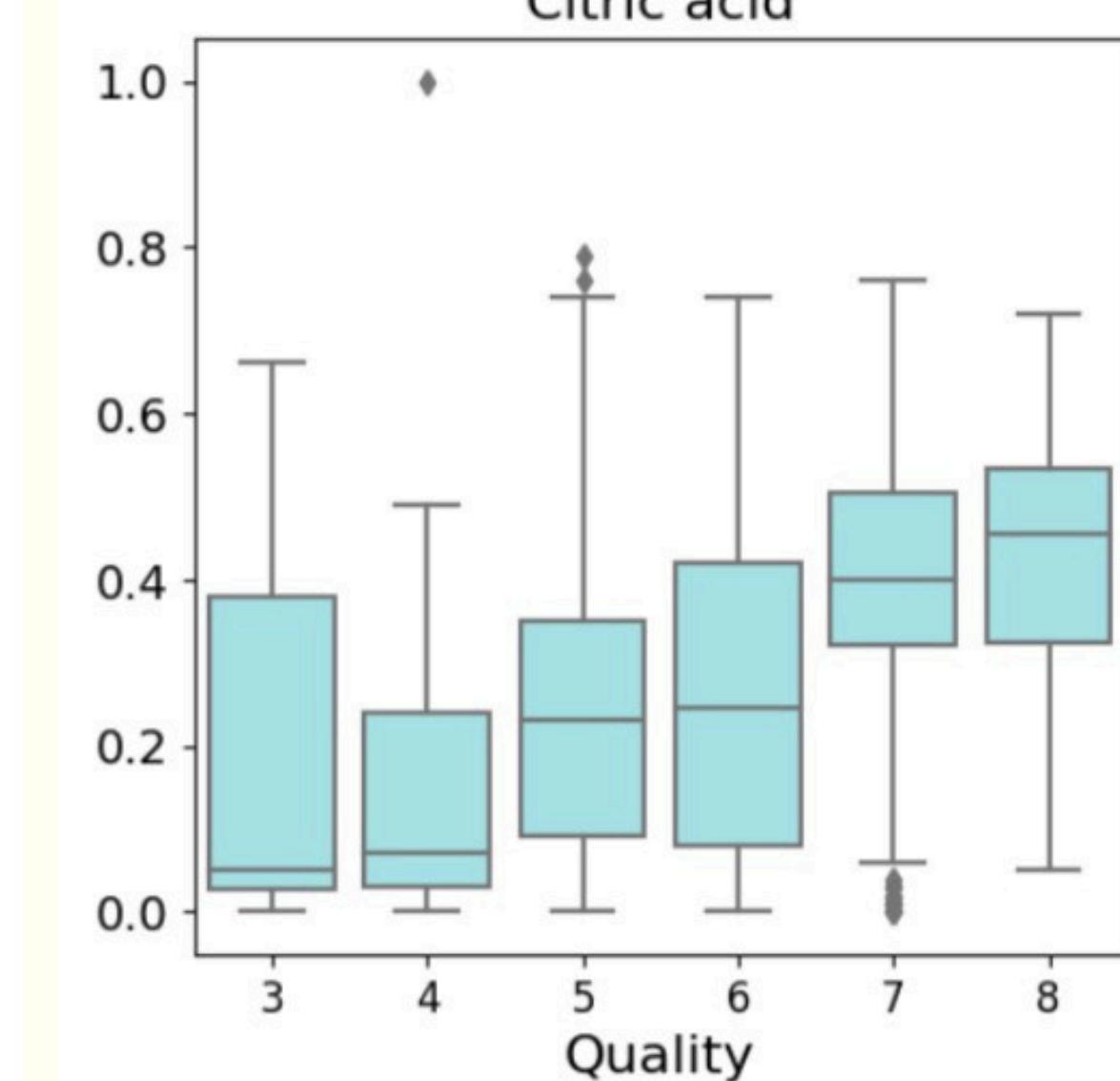
Alcohol



Sulphates



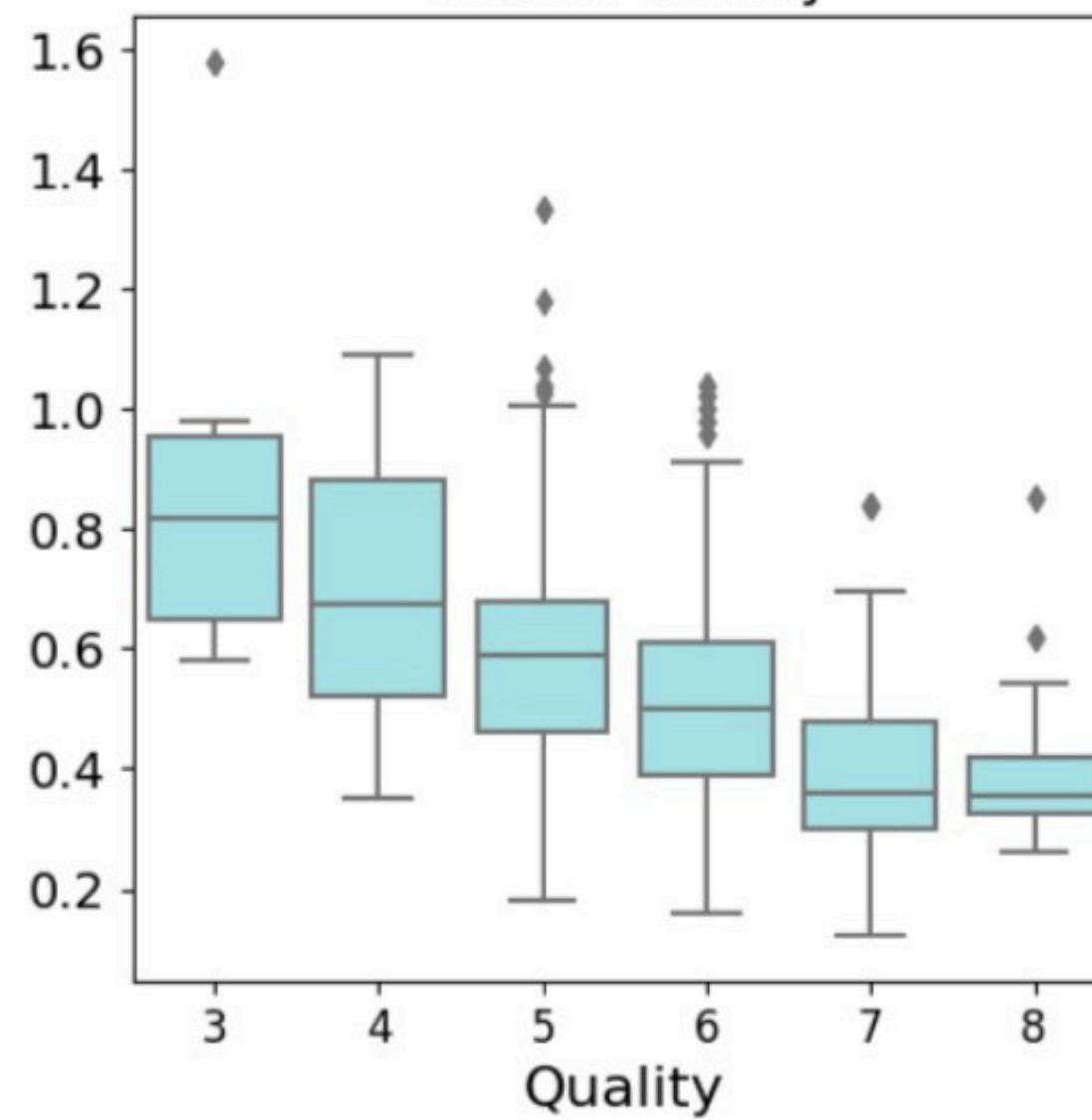
Citric acid



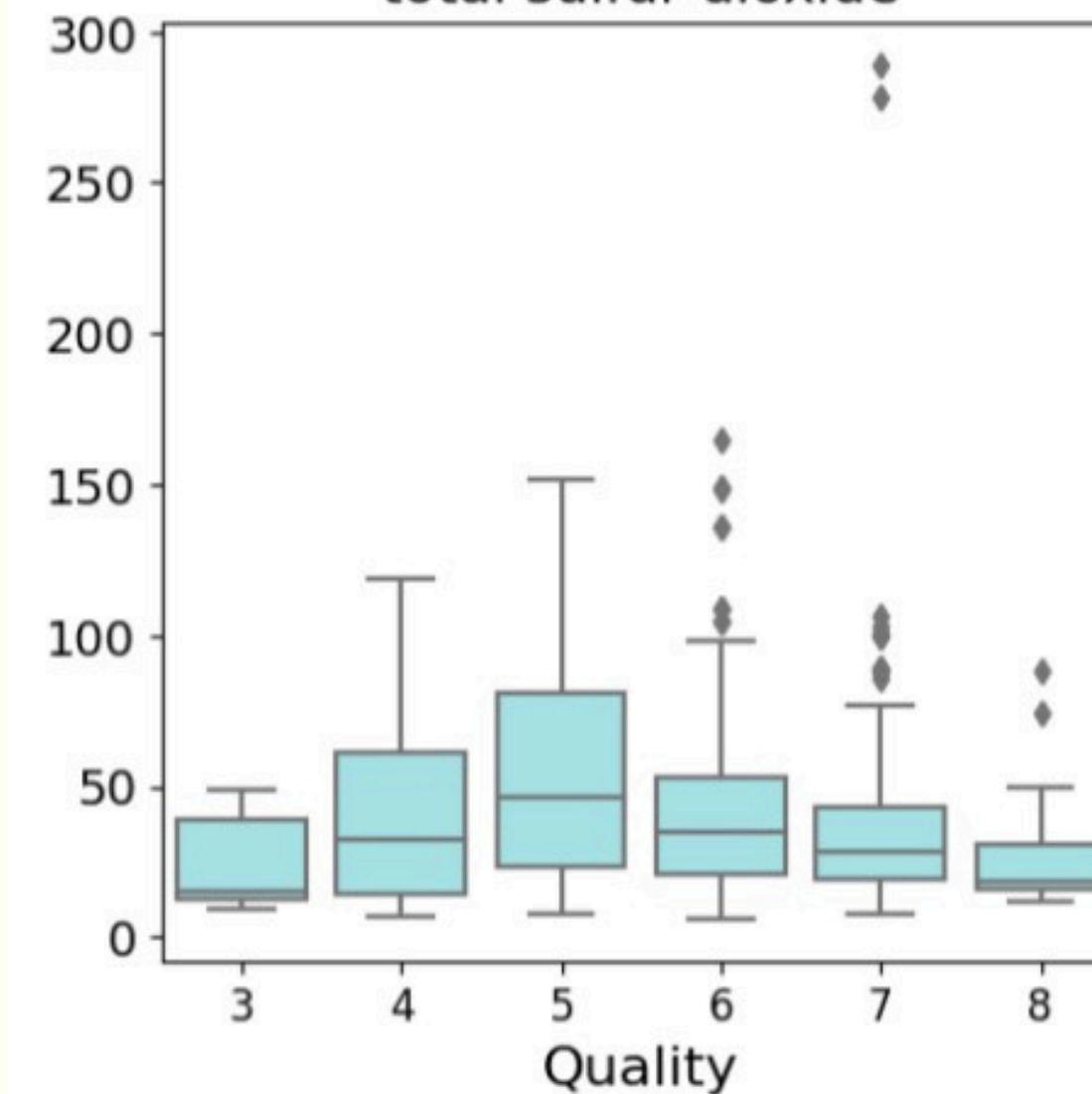
不同理化性質在品質中的分布情形

負相關

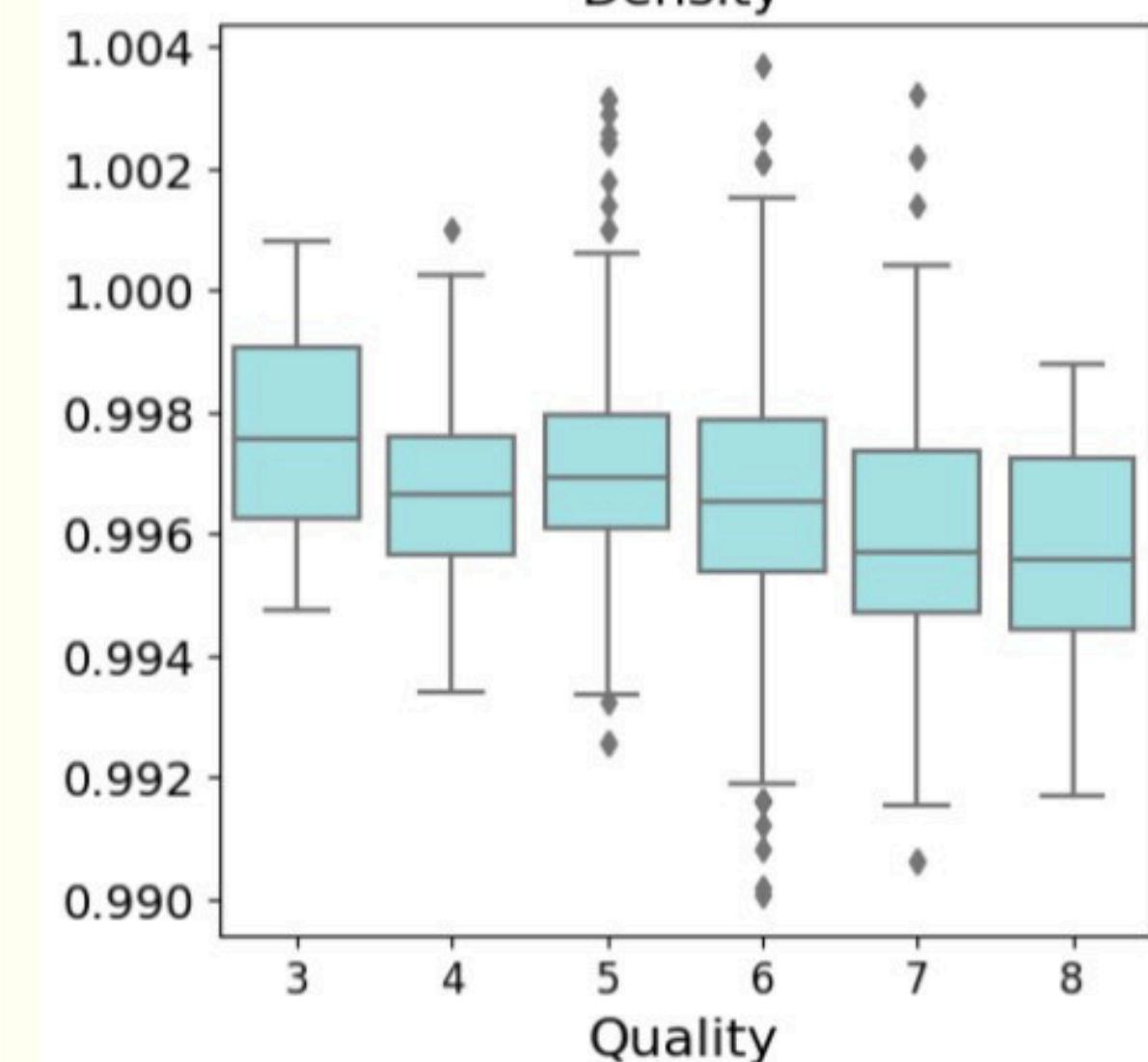
Volatile acidity



total sulfur dioxide



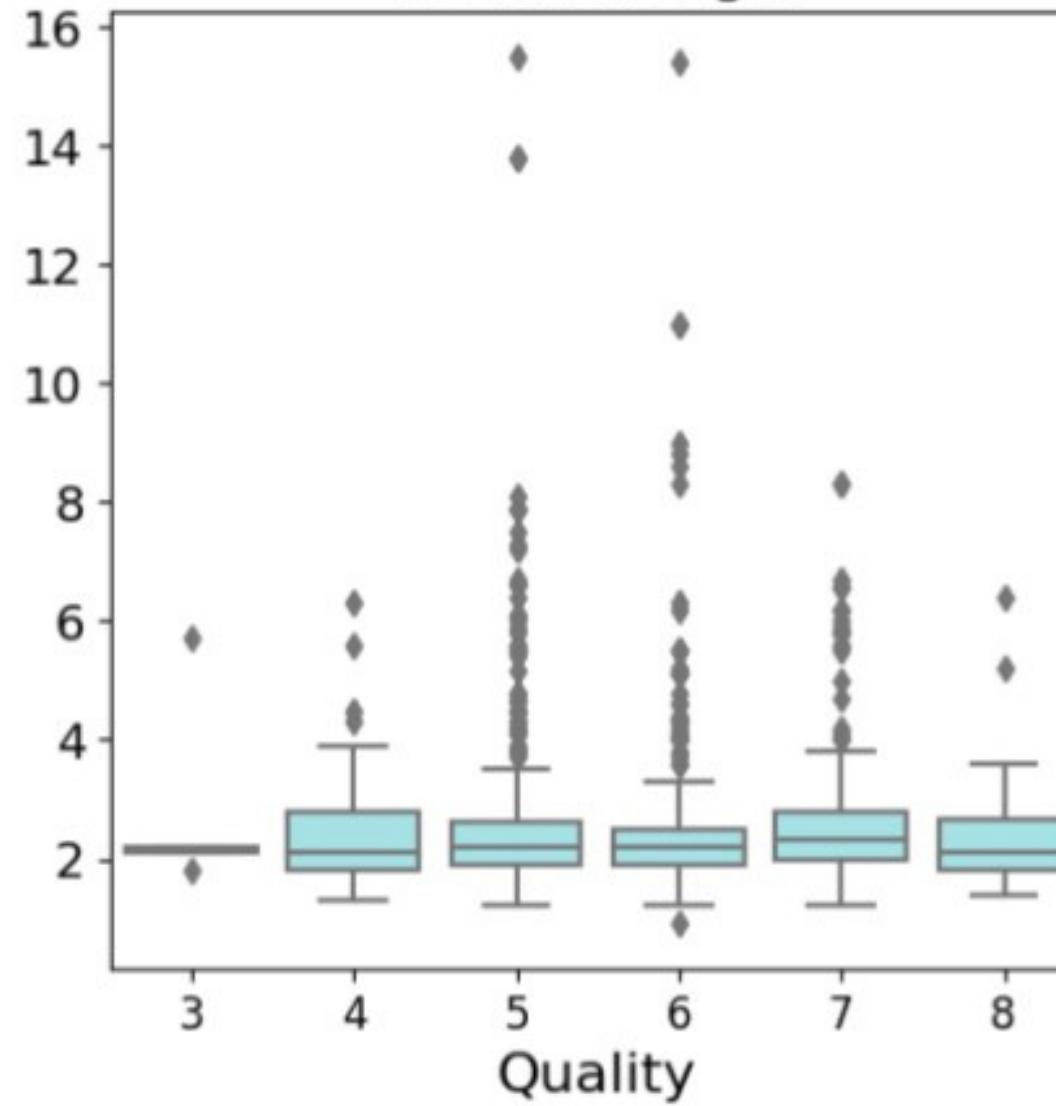
Density



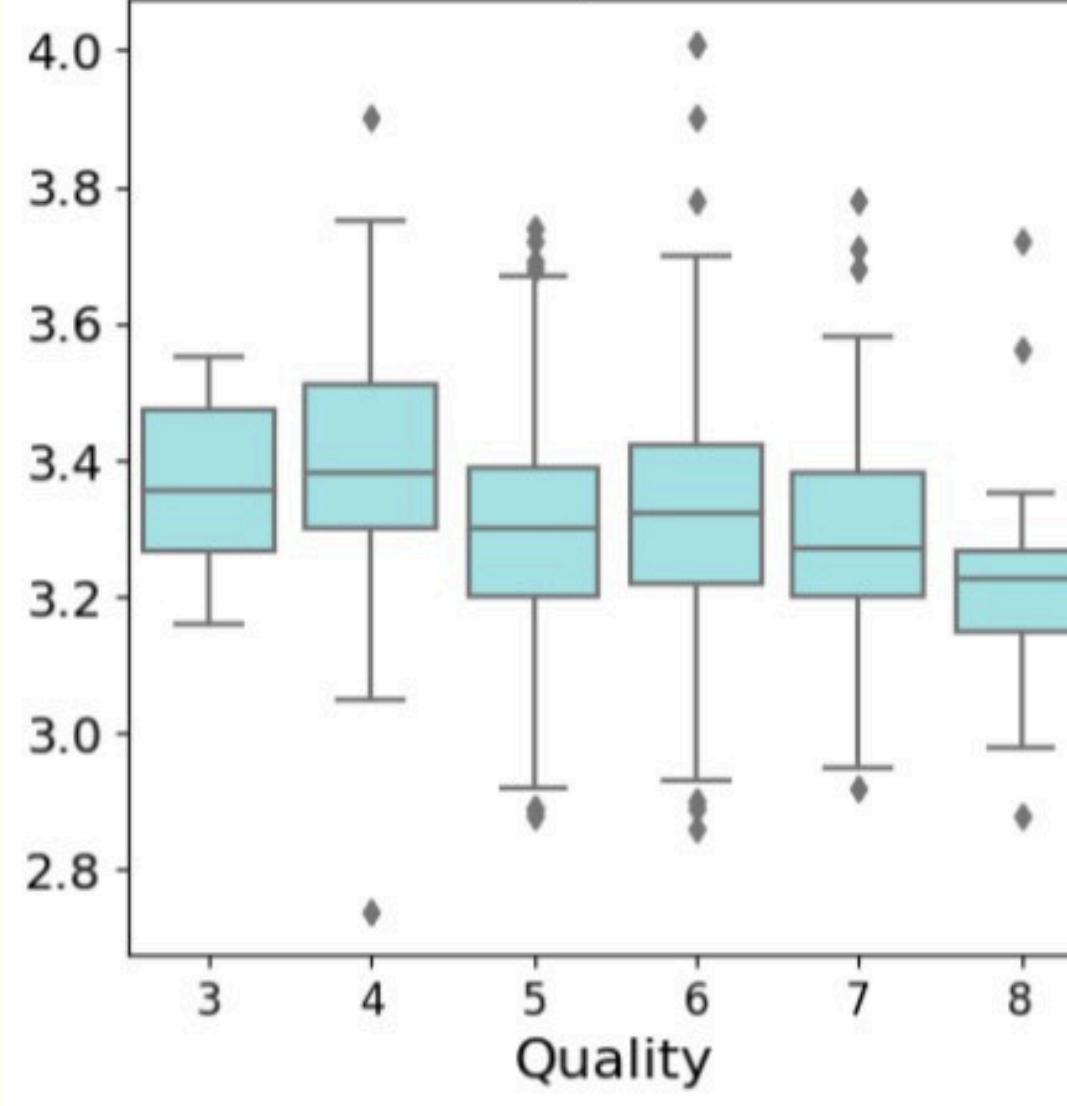
不同理化性質在品質中的分布情形

較無太大關係

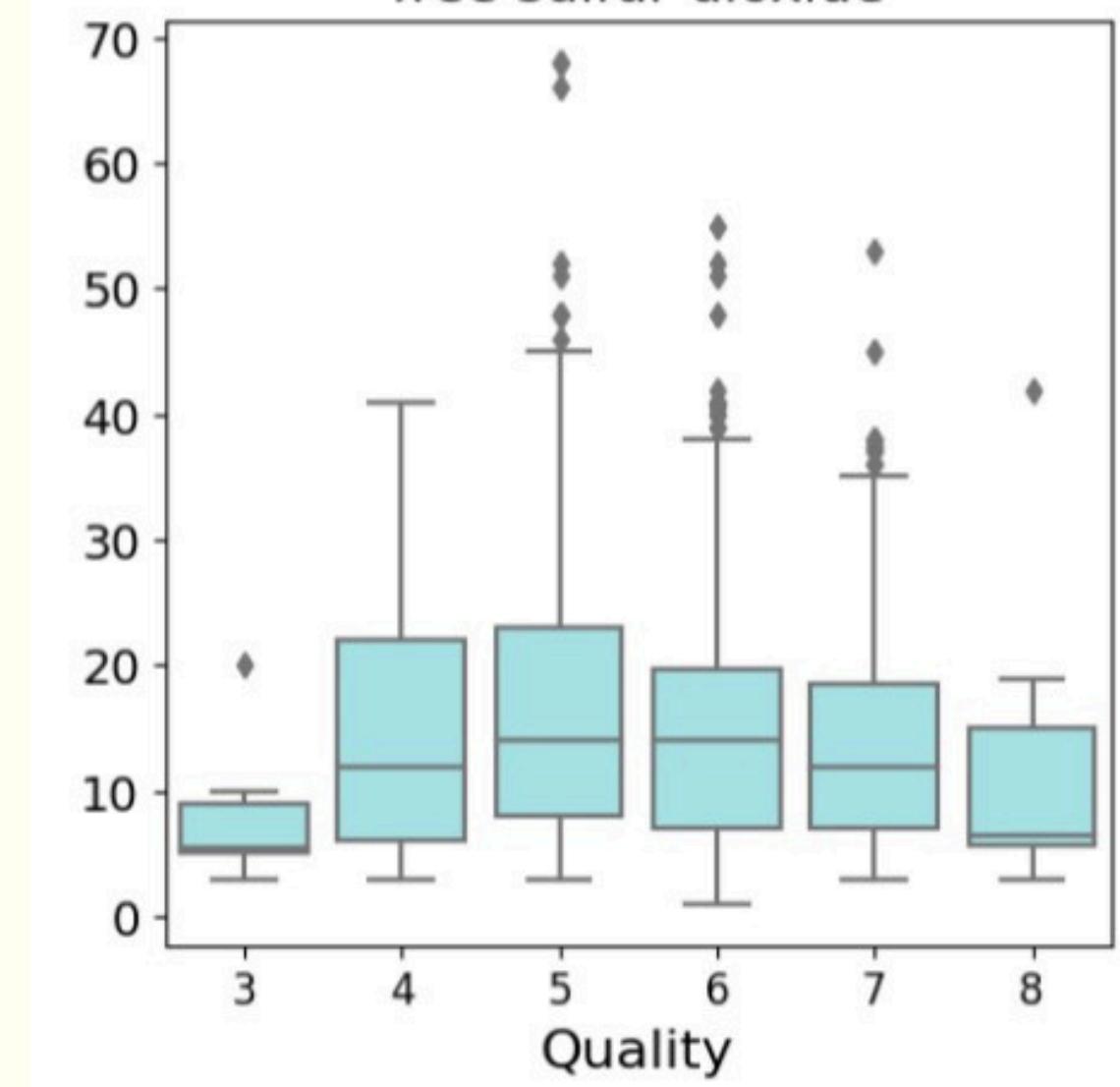
Residual sugar



pH



free sulfur dioxide





03

分析及建立模型

RESULTS

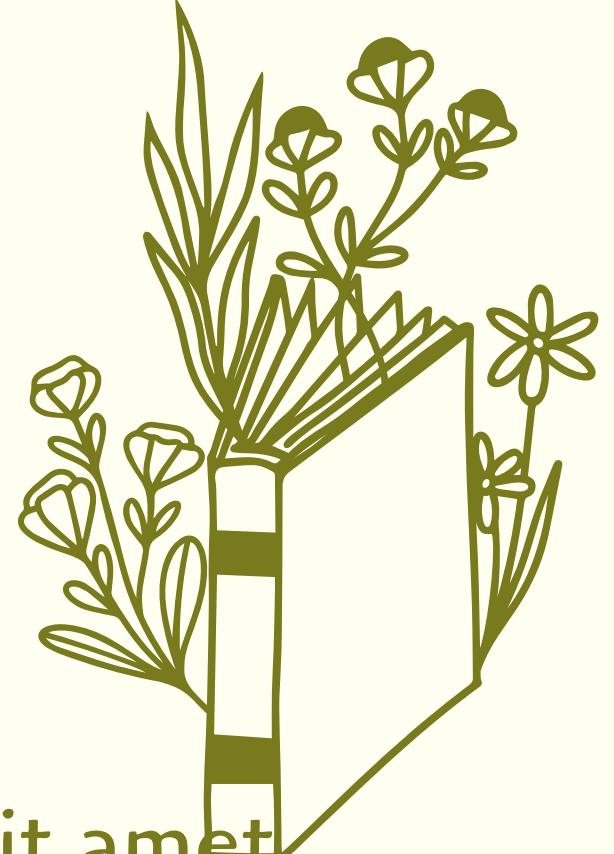
```
df[\"quality\"] = np.where(df[\"quality\"] >= 7, 1, 0)  
df['quality'].value_counts()
```

```
0    984  
1    159
```

Name: quality, dtype: int64

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit.**

- Elaborate on what you want to discuss.



04

結果呈現

數據探索和摘要統計

執行儲存格 | 執行更高版本 | 偵錯儲存格 | 前往 [5]

```
✓ #%%  
# 查看數據的前幾行  
print(df.head())
```

執行儲存格 | 執行更高版本 | 偵錯儲存格 | 前往 [6]

```
✓ #%%  
# 查看數據的基本統計信息  
print(df.describe())
```

數據探索和摘要統計

```
✓ # 查看數據的前幾行 ...
fixed acidity  volatile acidity  citric acid  residual sugar  chlorides \
0           7.4            0.70        0.00          1.9      0.076
1           7.8            0.88        0.00          2.6      0.098
2           7.8            0.76        0.04          2.3      0.092
3          11.2            0.28        0.56          1.9      0.075
4           7.4            0.70        0.00          1.9      0.076

free sulfur dioxide  total sulfur dioxide  density  pH  sulphates \
0             11.0              34.0  0.9978  3.51      0.56
1            25.0              67.0  0.9968  3.20      0.68
2            15.0              54.0  0.9970  3.26      0.65
3            17.0              60.0  0.9980  3.16      0.58
4             11.0              34.0  0.9978  3.51      0.56

alcohol  quality
0       9.4      5
1       9.8      5
2       9.8      5
3       9.8      6
4       9.4      5
```

```
✓ # 查看數據的基本統計信息 ...
...   fixed acidity  volatile acidity  citric acid  residual sugar \
count  1143.000000    1143.000000  1143.000000  1143.000000
mean   8.311111     0.531339    0.268364    2.532152
std    1.747595     0.179633    0.196686    1.355917
min    4.600000     0.120000    0.000000    0.900000
25%   7.100000     0.392500    0.090000    1.900000
50%   7.900000     0.520000    0.250000    2.200000
75%   9.100000     0.640000    0.420000    2.600000
max   15.900000    1.580000    1.000000    15.500000

chlorides  free sulfur dioxide  total sulfur dioxide  density \
count  1143.000000    1143.000000  1143.000000  1143.000000
mean   0.086933     15.615486    45.914698    0.996730
std    0.047267     10.250486    32.782130    0.001925
min    0.012000     1.000000    6.000000    0.990070
25%   0.070000     7.000000    21.000000    0.995570
50%   0.079000     13.000000    37.000000    0.996680
75%   0.090000     21.000000    61.000000    0.997845
max   0.611000     68.000000    289.000000   1.003690

pH  sulphates  alcohol  quality
count  1143.000000  1143.000000  1143.000000  1143.000000
mean   3.311015    0.657708    10.442111   5.657043
std    0.156664    0.170399    1.082196   0.805824
min    2.740000    0.330000    8.400000   3.000000
25%   3.205000    0.550000    9.500000   5.000000
50%   3.310000    0.620000    10.200000   6.000000
75%   3.400000    0.730000    11.100000   6.000000
max   4.010000    2.000000    14.900000   8.000000
```

單變量分析

```
執行儲存格 | 執行更高版本 | 偵錯儲存格 | 前往 [11]  
# %  
# 遍歷每個特徵，印出單變量統計信息  
for column in df.columns:  
    if column != "Id":  
        print_uvariate_statistics(df, column)  
        print("*" * 100)
```

單變量分析

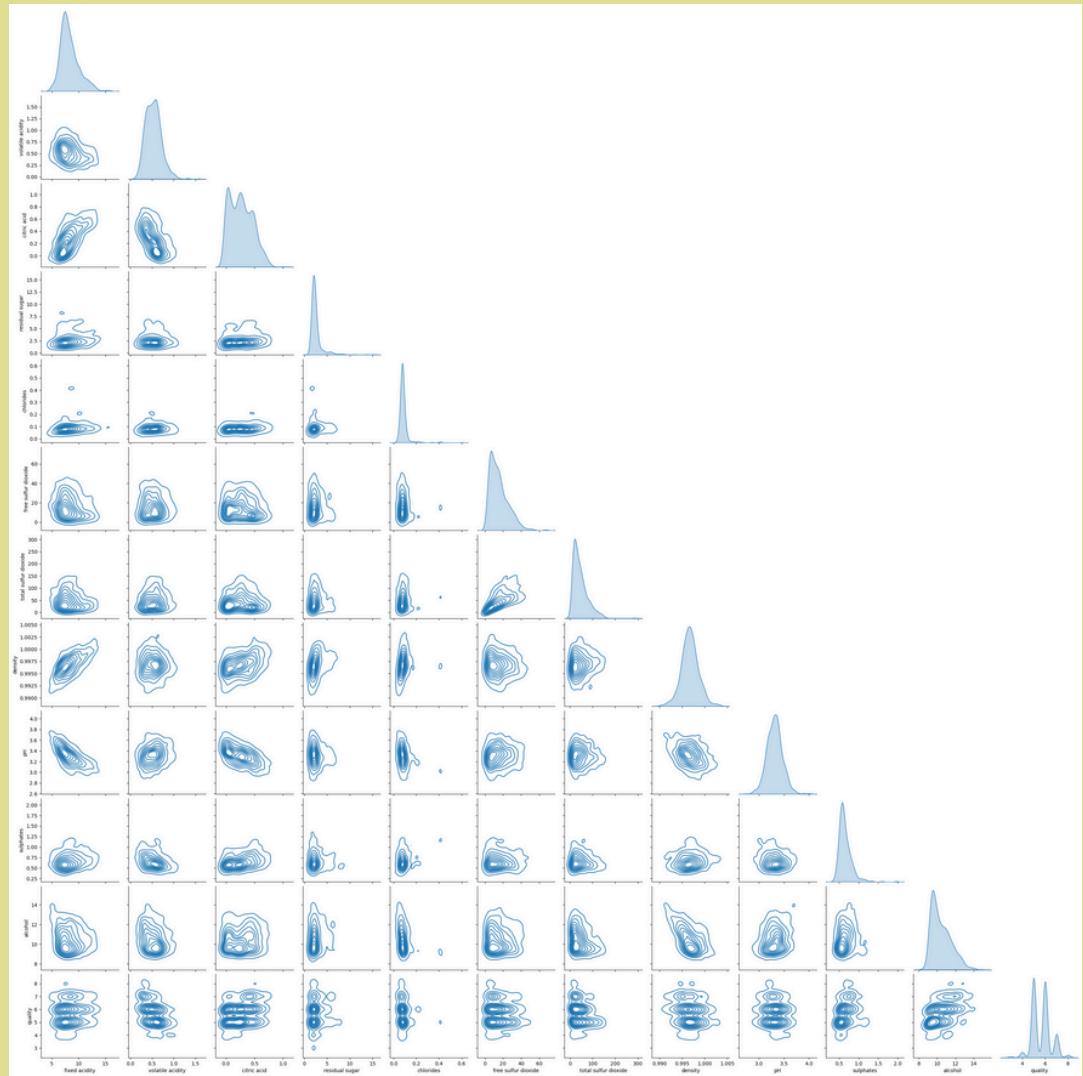
```
✓ # 遍歷每個特徵，印出單變量統計信息 ...
...
Mean of fixed acidity is 8.311111111111112
Median of fixed acidity is 7.9
Skewness of fixed acidity is 1.0449299563724053
Std of fixed acidity is 1.747595017169539
IQR of fixed acidity is 2.0
*****
Mean of volatile acidity is 0.5313385826771653
Median of volatile acidity is 0.52
Skewness of volatile acidity is 0.6815474143781255
Std of volatile acidity is 0.17963319302252445
IQR of volatile acidity is 0.2475
*****
Mean of citric acid is 0.2683639545056868
Median of citric acid is 0.25
Skewness of citric acid is 0.3715607833706083
Std of citric acid is 0.1966858523482191
IQR of citric acid is 0.32999999999999996
*****
Mean of residual sugar is 2.5321522309711284
Median of residual sugar is 2.2
Skewness of residual sugar is 4.361096404323471
Std of residual sugar is 1.355917466682676
IQR of residual sugar is 0.7000000000000002
*****
Mean of chlorides is 0.0869326334208224
...
Skewness of quality is 0.2867917004538591
Std of quality is 0.8058242481000936
IQR of quality is 1.0
*****
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

成對圖分析

```
執行儲存格 | 執行更高版本 | 偵錯儲存格 | 前往 [10]
▽ # %%
▽ import seaborn as sns
  import matplotlib.pyplot as plt

# 成對圖 with KDE
sns.pairplot(df, kind="kde", corner=True)
plt.show()
```

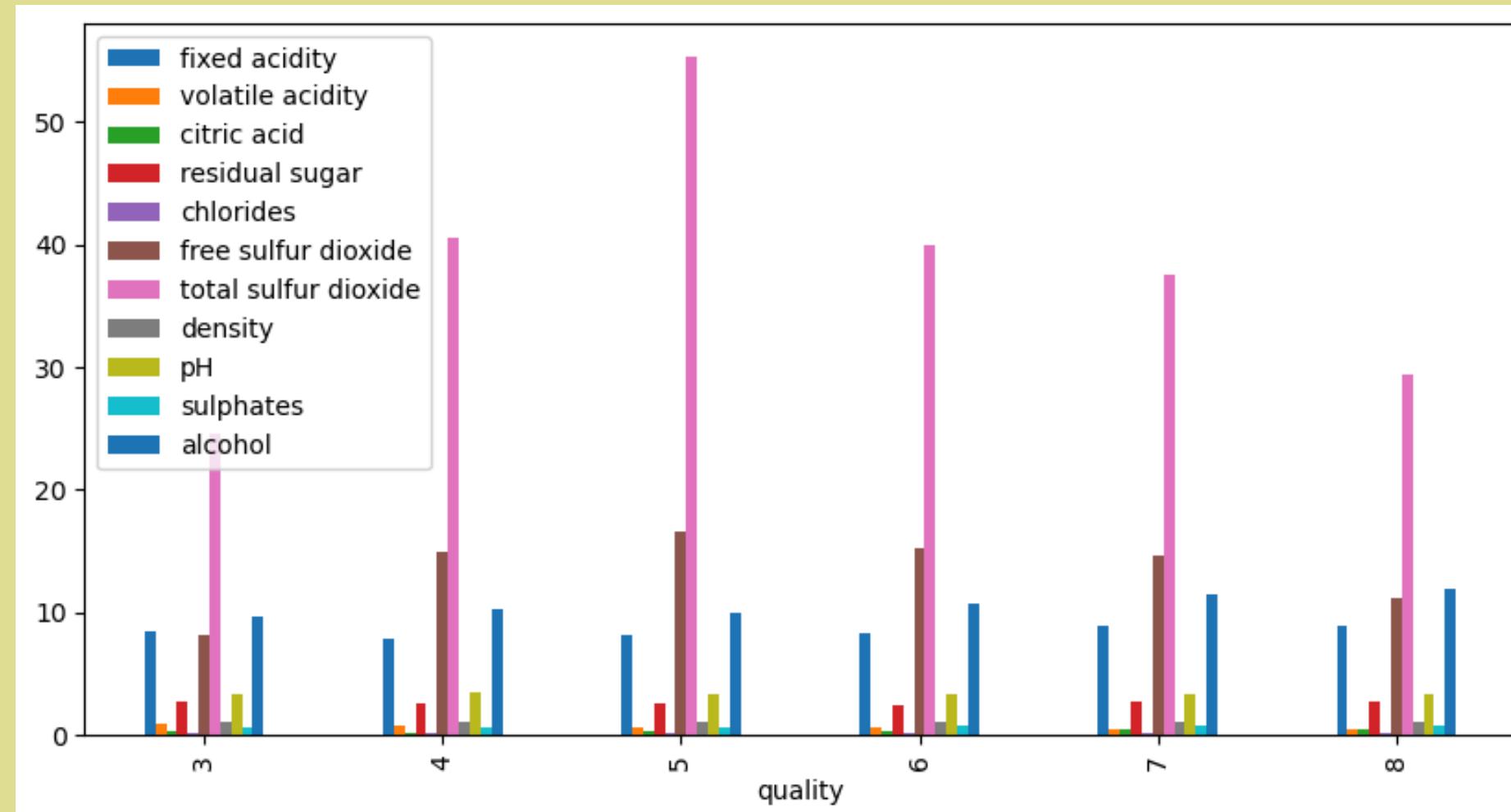
成對圖分析



品質等級平均值條形圖

```
執行儲存格 | 執行更高版本 | 偵錯儲存格  
# %%  
# 根據品質等級計算平均值  
average_quality = df.groupby("quality").mean()  
  
# 繪製條形圖  
average_quality.plot(kind="bar", figsize=(10, 5))  
plt.show()
```

品質等級平均值條形圖



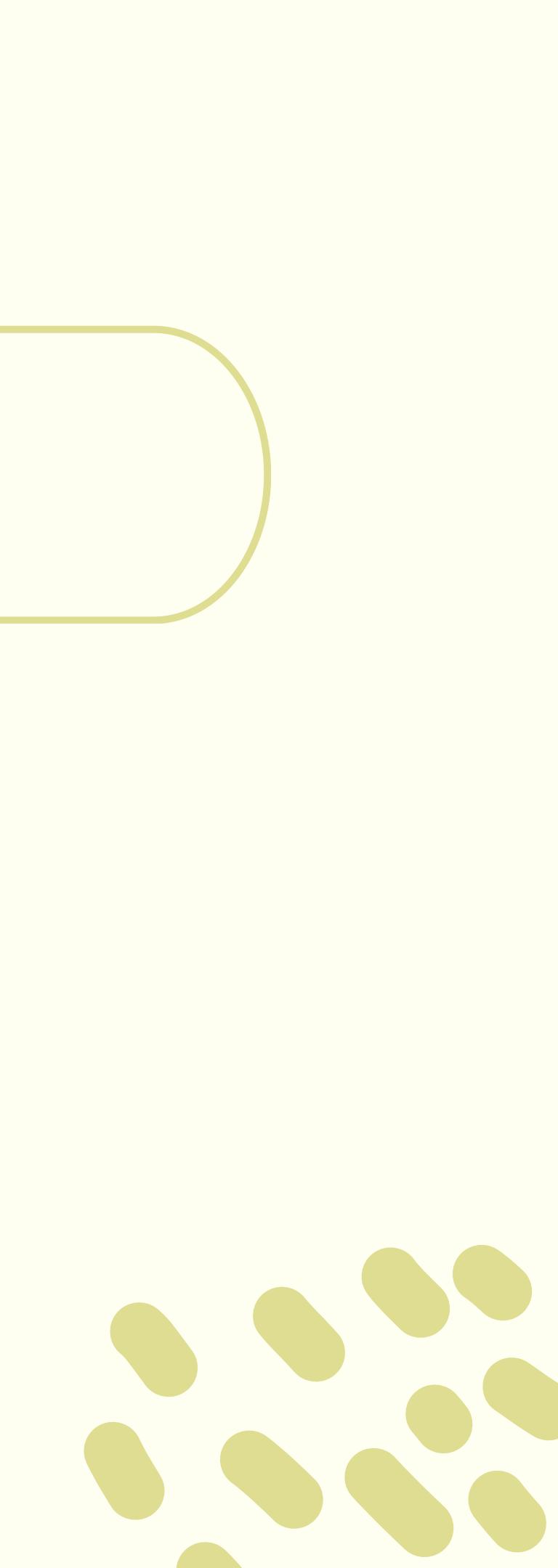


05

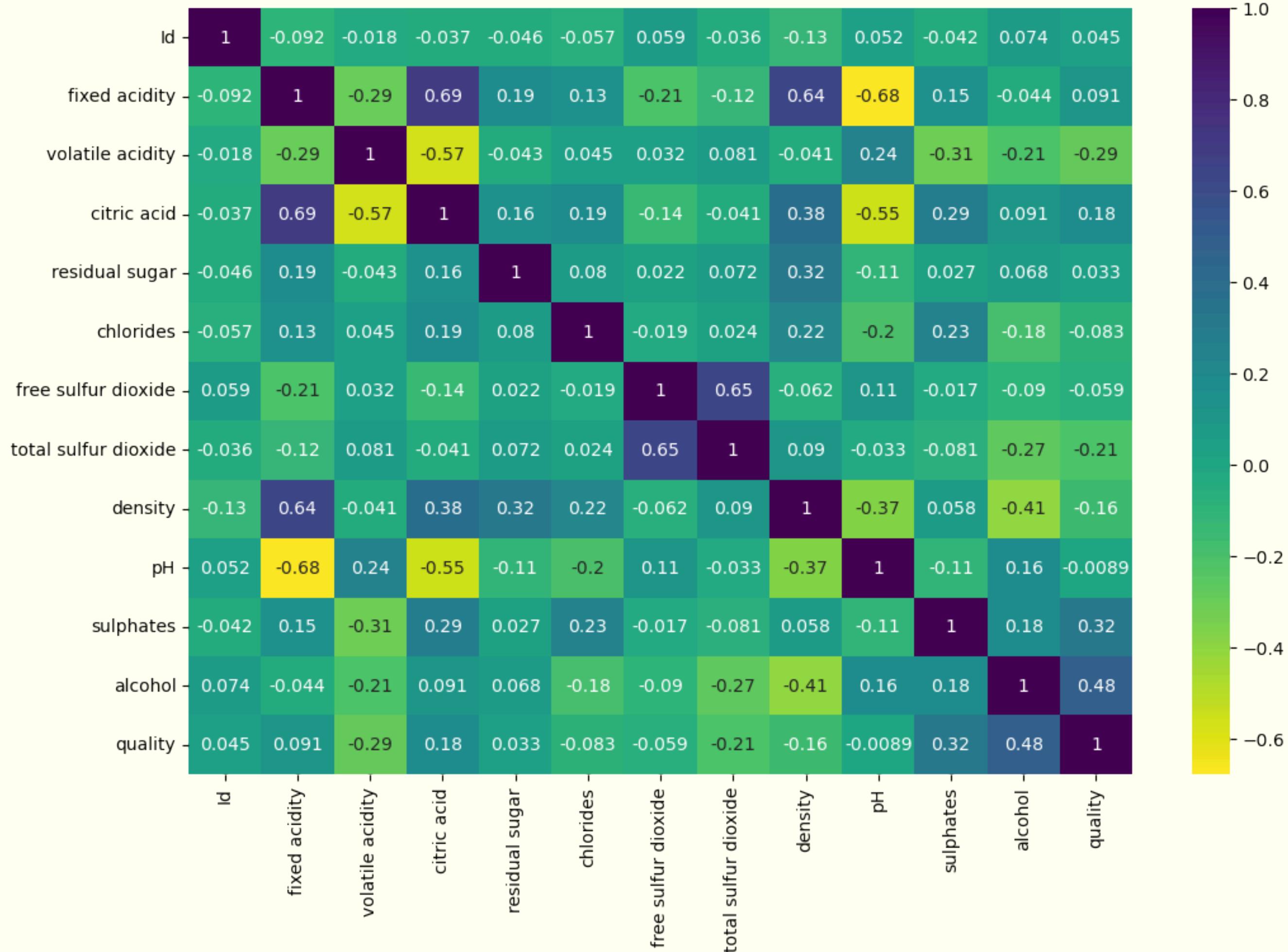
管理意涵



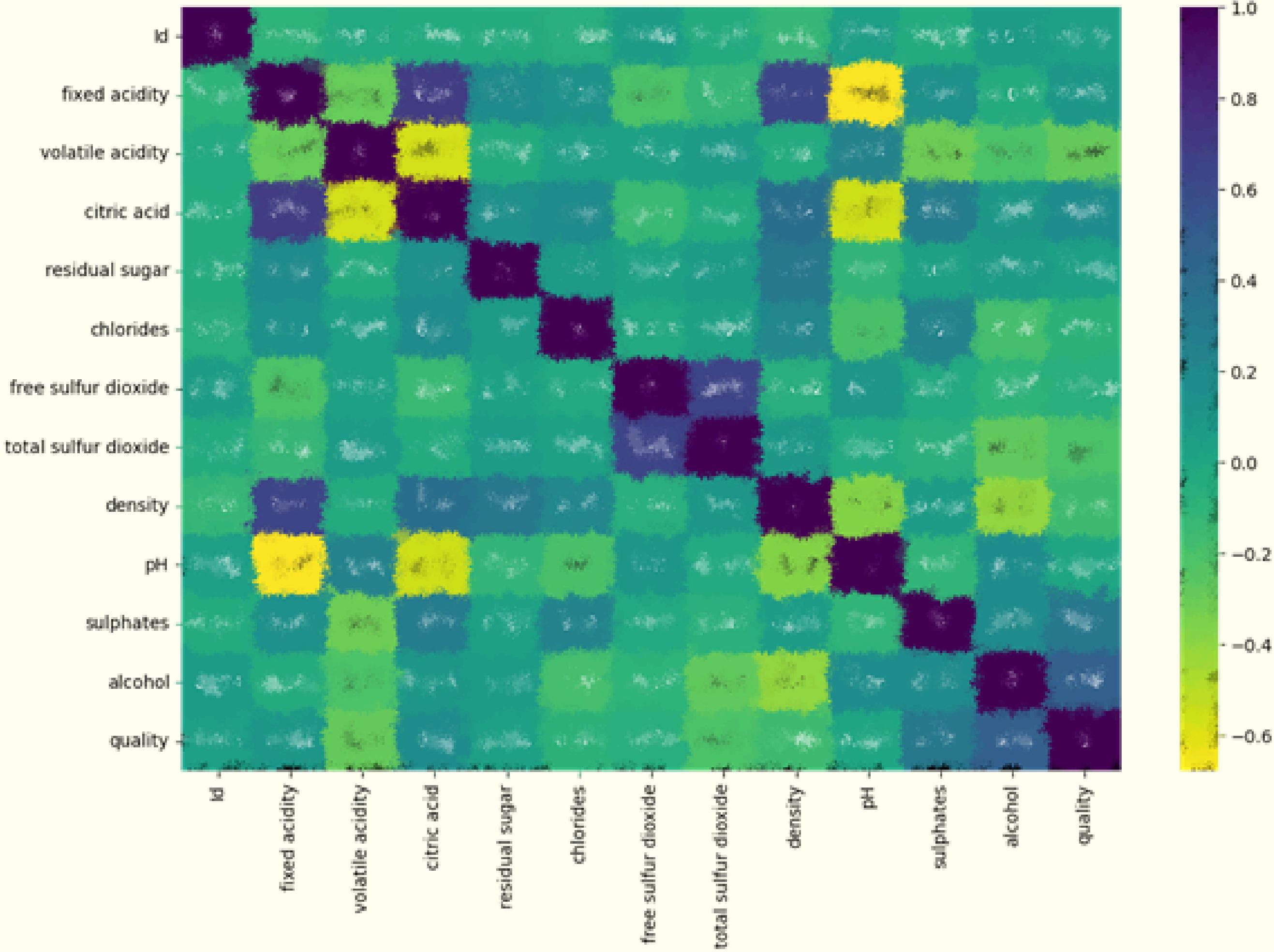
影響品質的三大重要因素

- Alcohol：酒精越多，品質越高
 - Sulphates：亞硫酸鹽越多，品質越高
 - Citric acid：檸檬酸越多，品質越高
- 

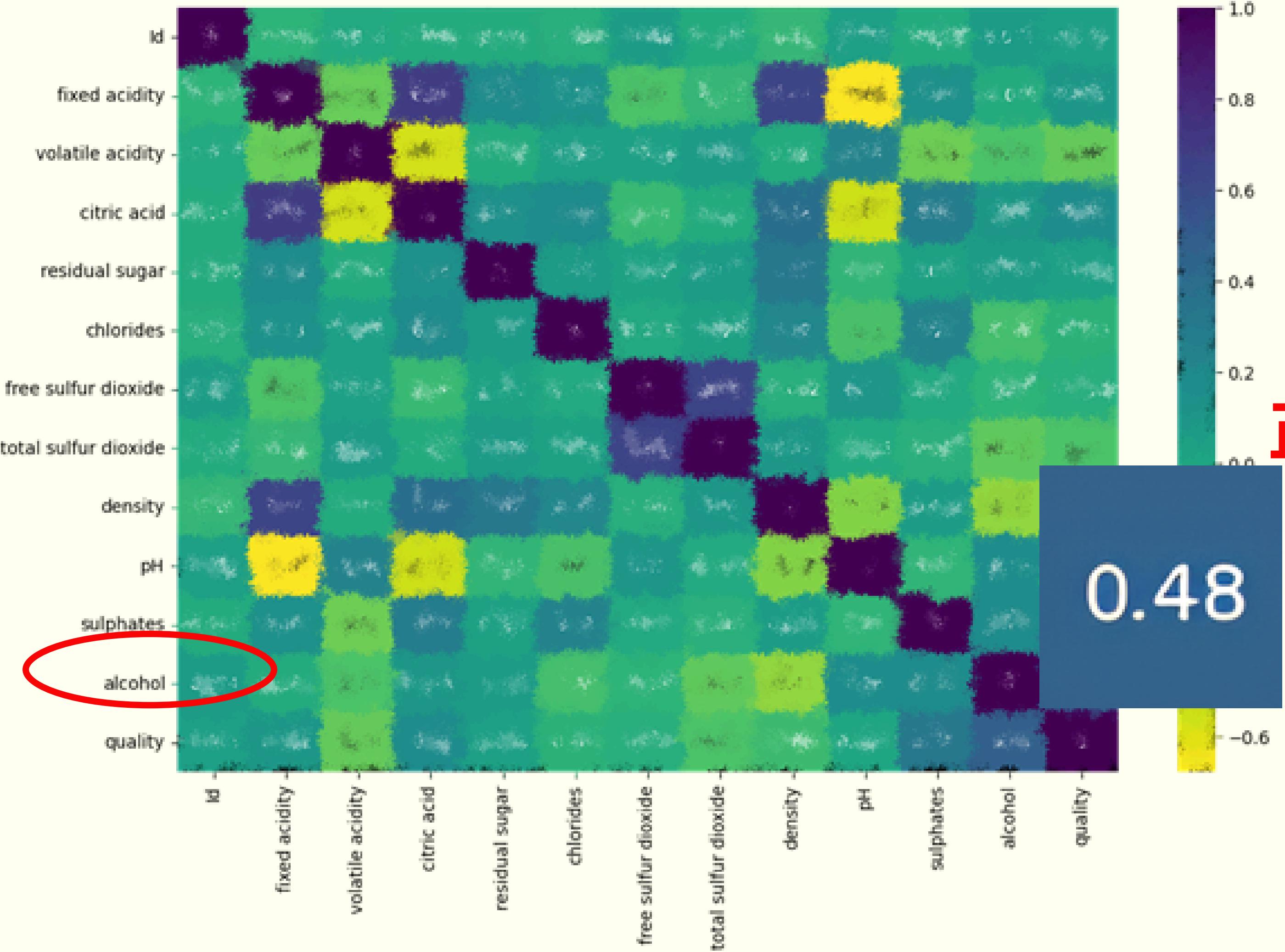
Correlation between Variables



Correlation between Variables

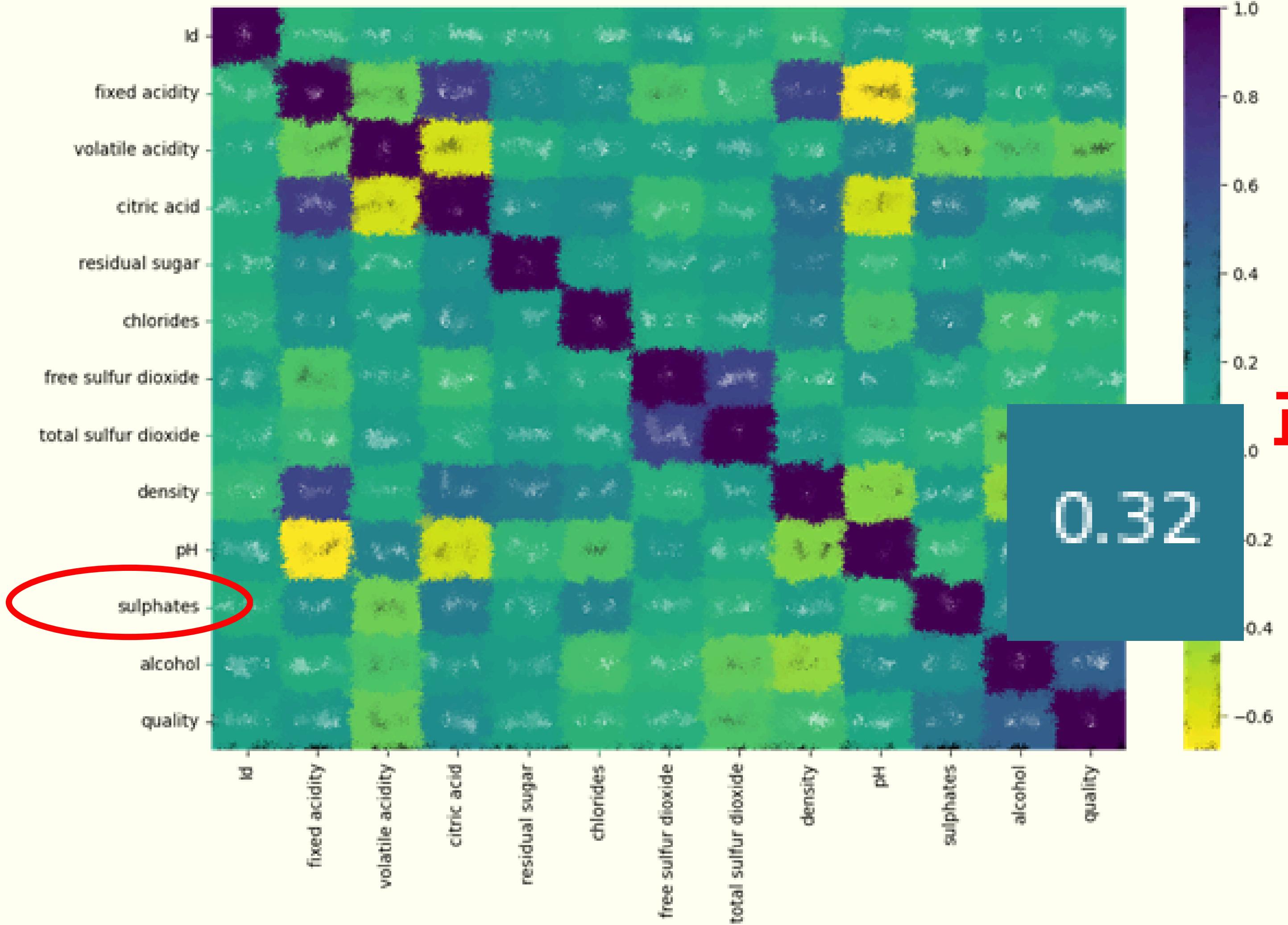


Correlation between Variables



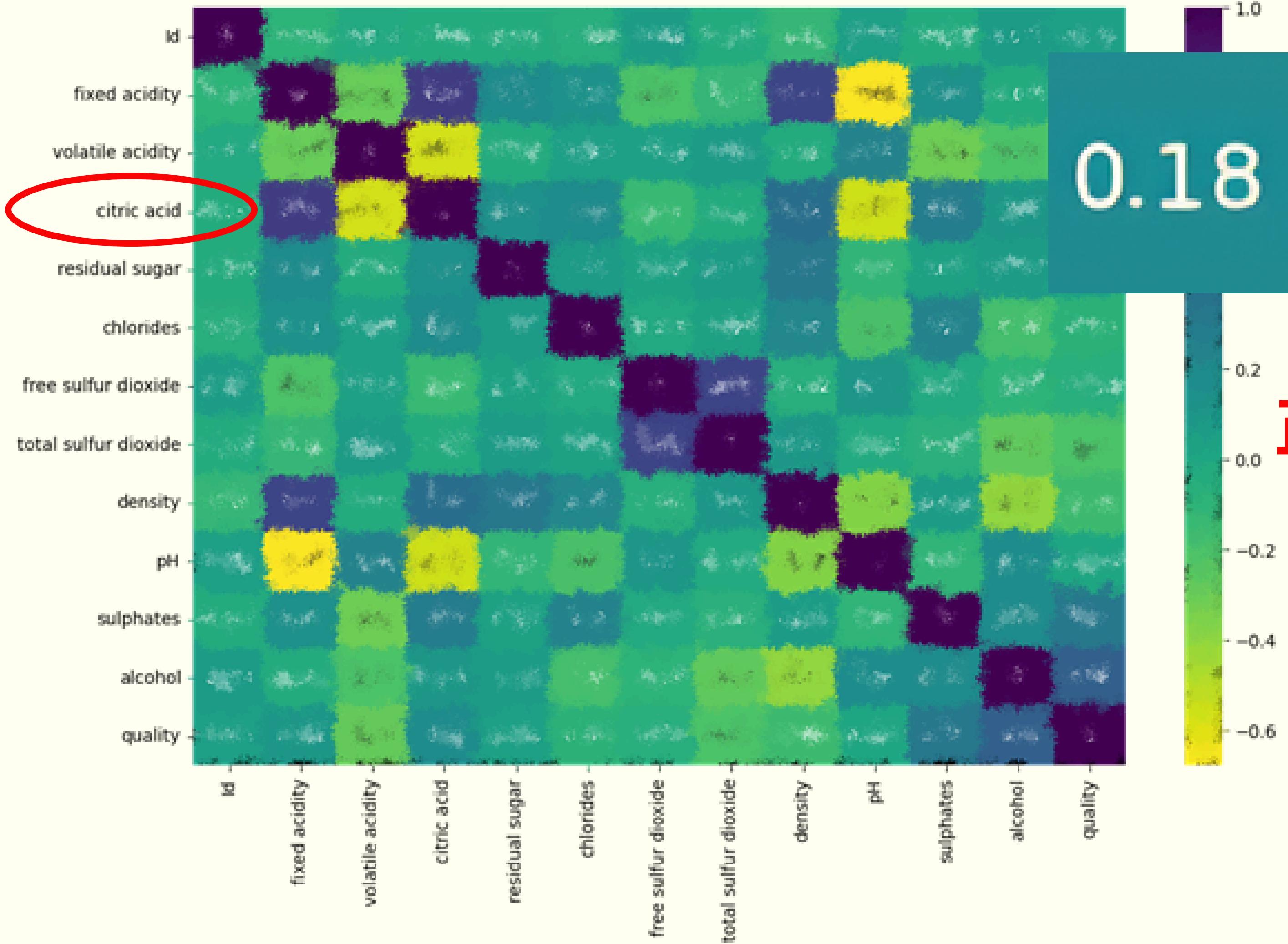
呈現
正相關

Correlation between Variables



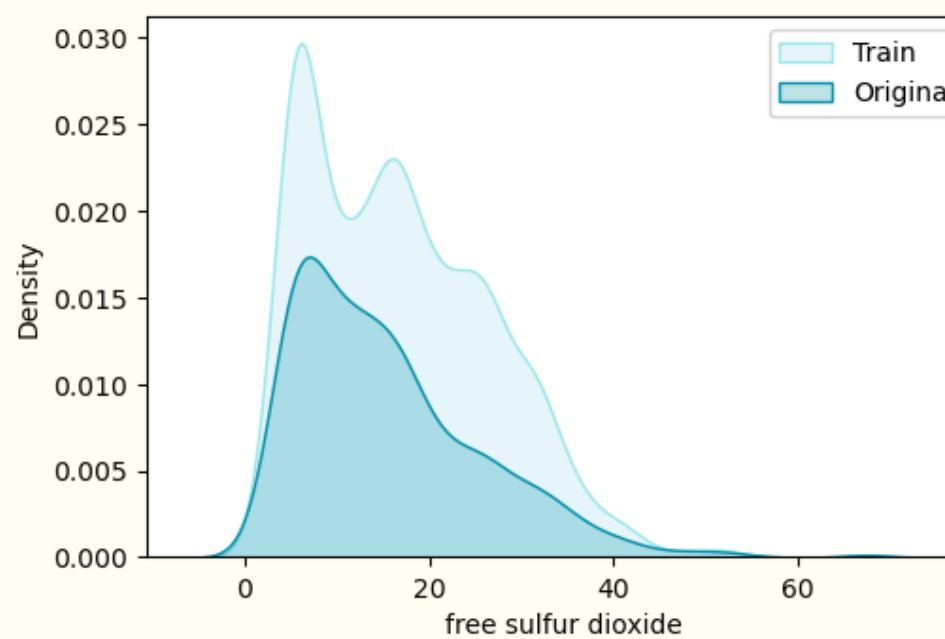
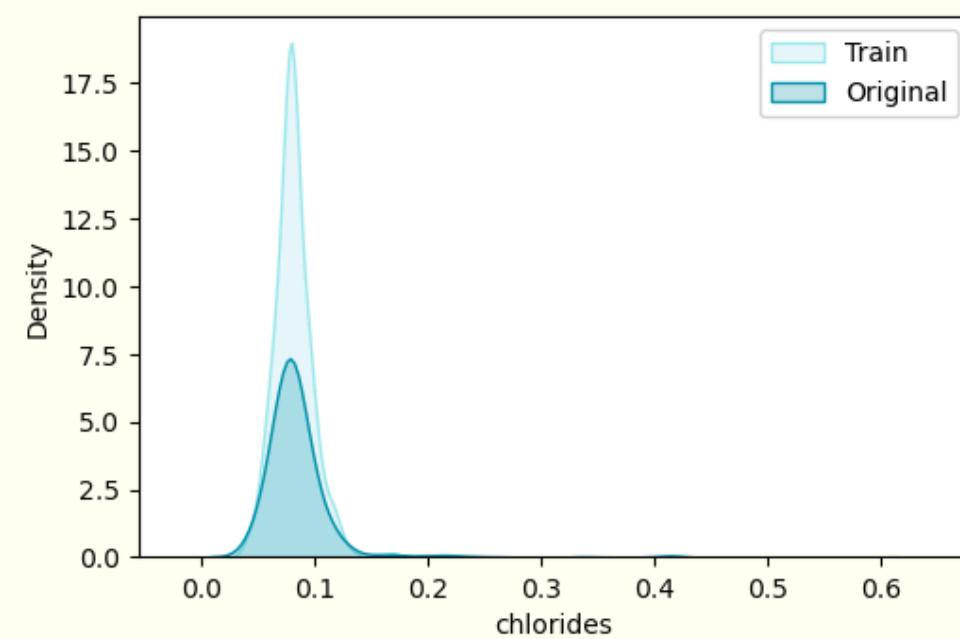
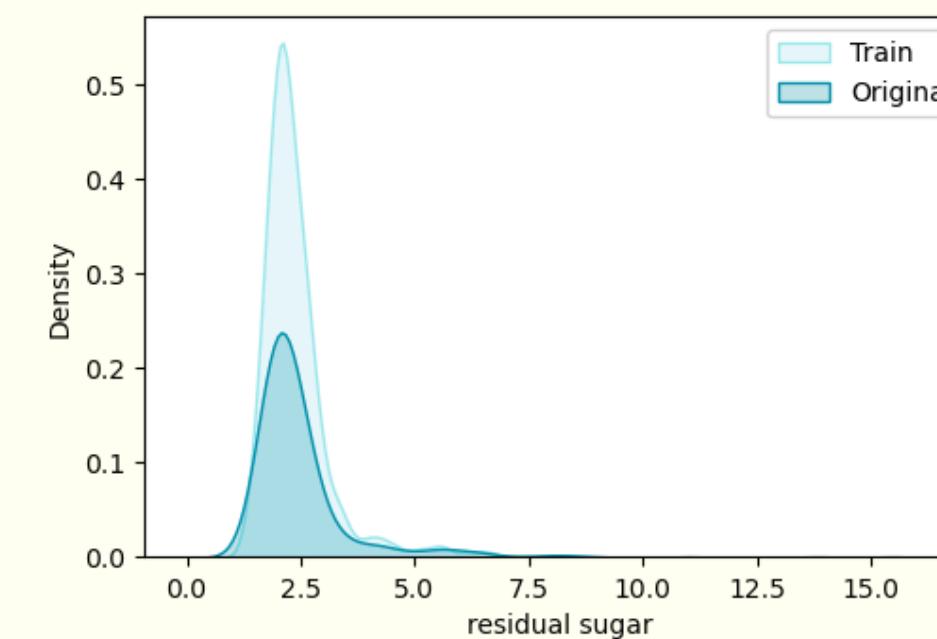
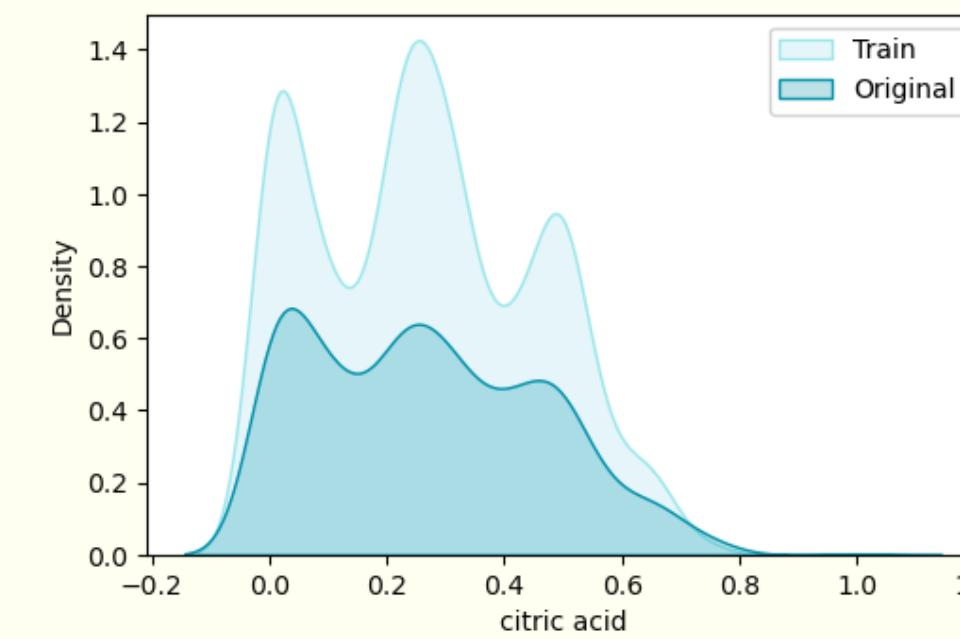
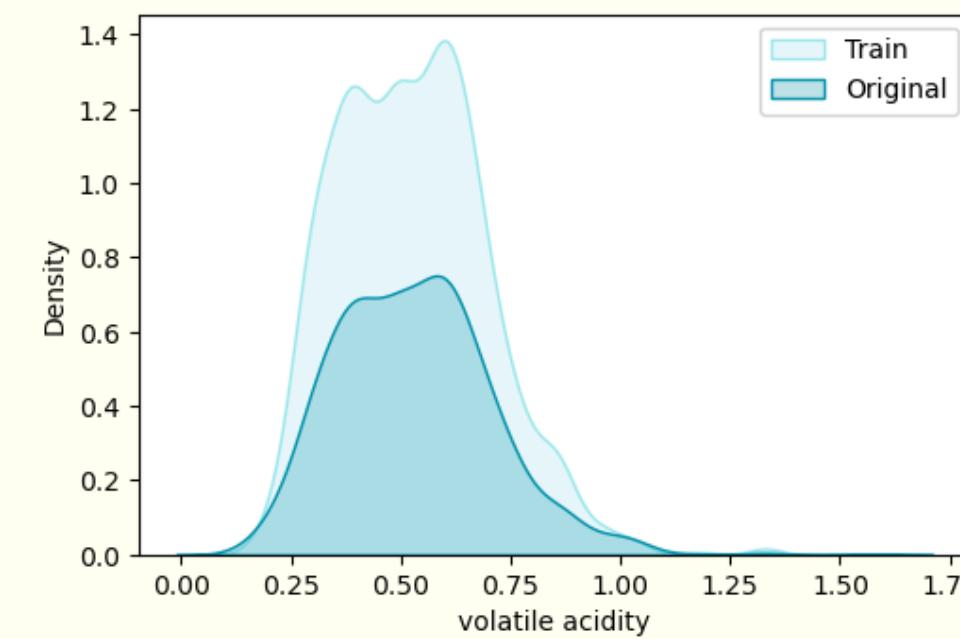
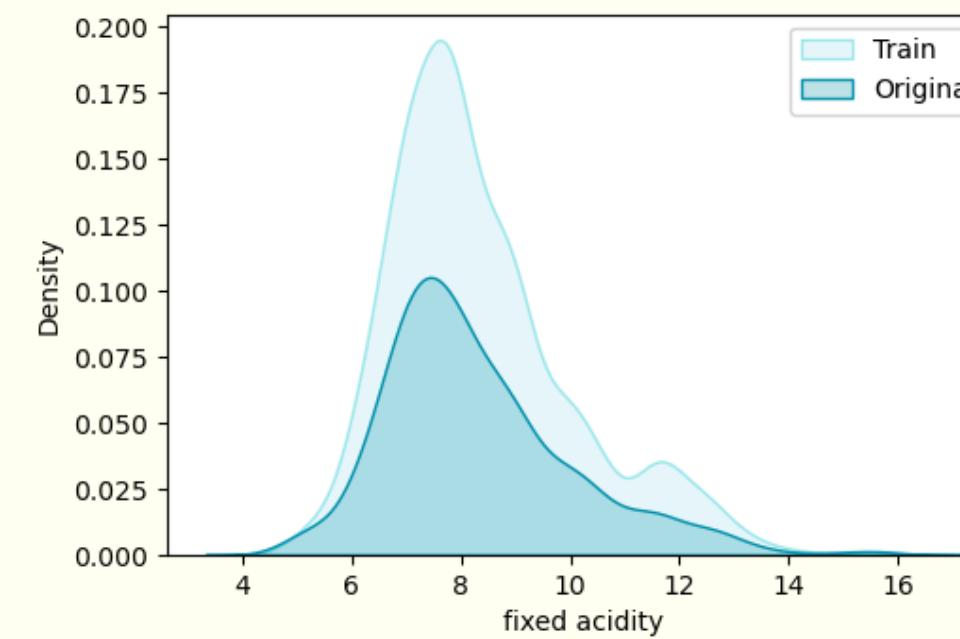
呈現
正相關

Correlation between Variables

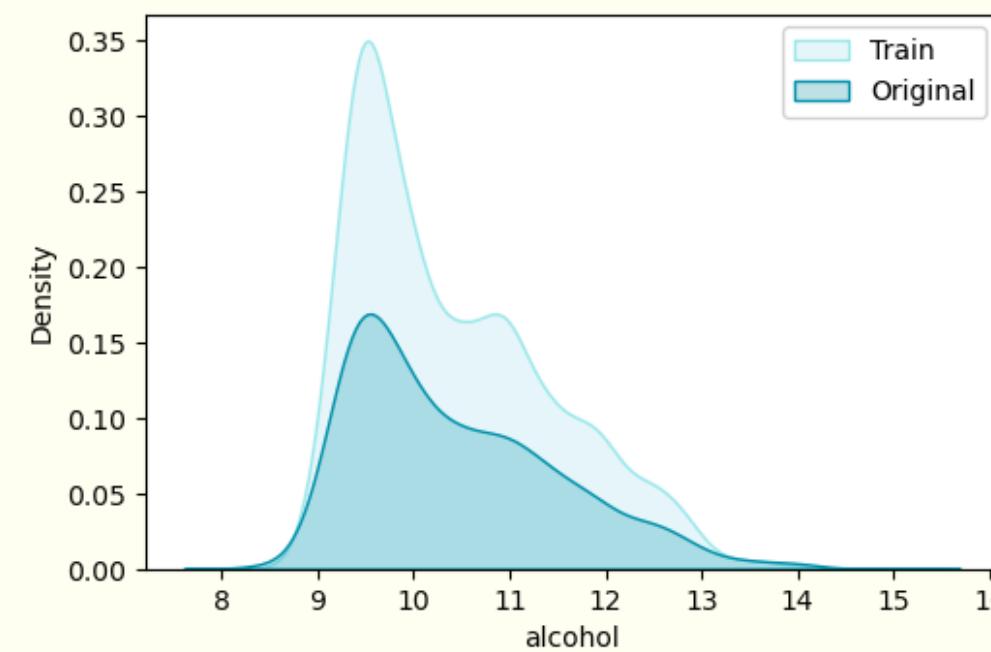
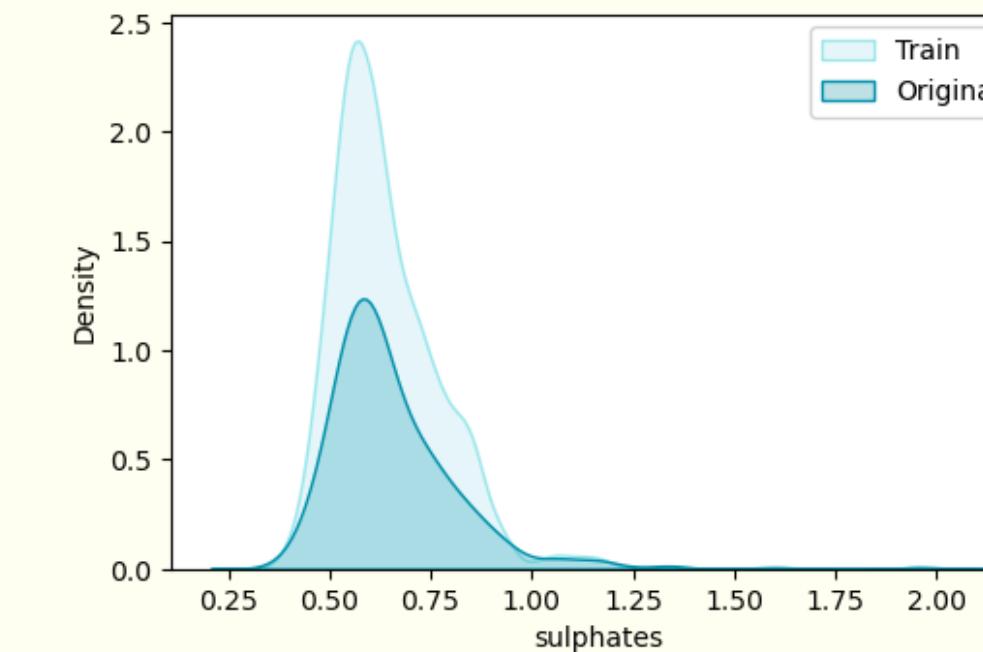
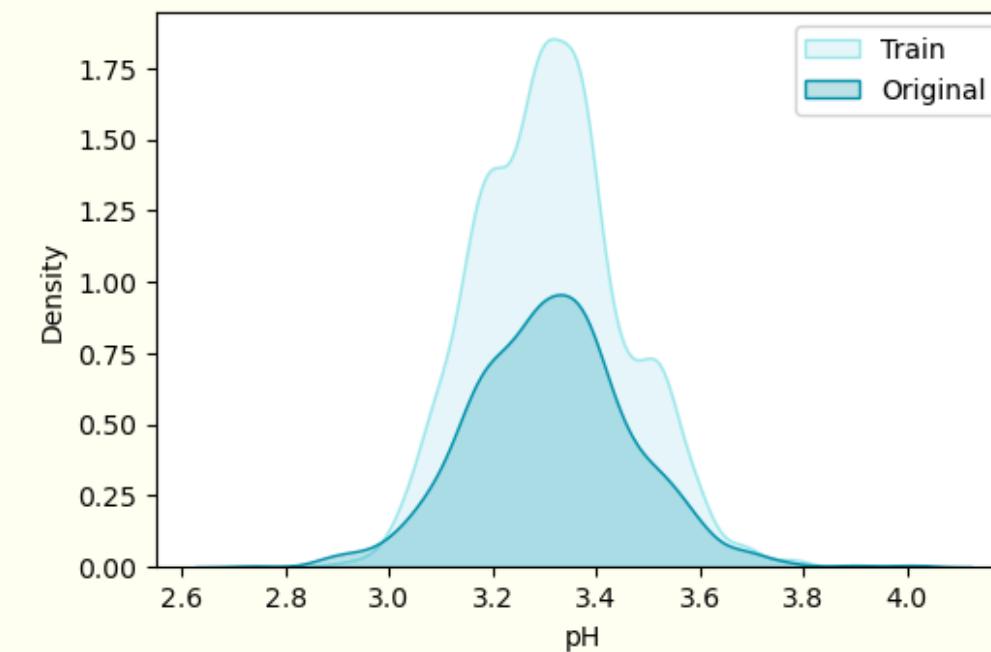
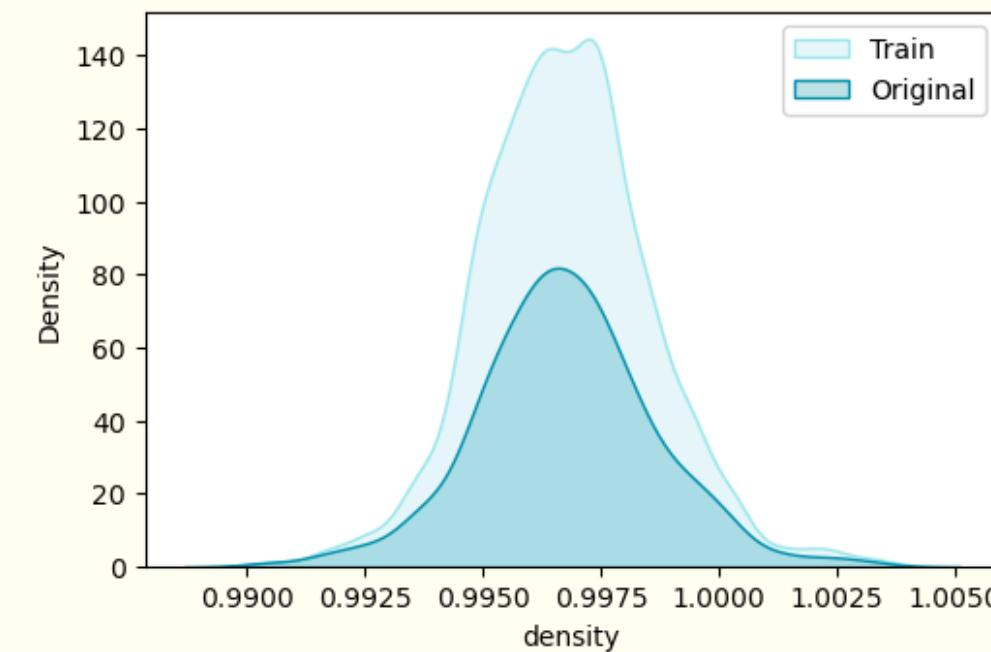
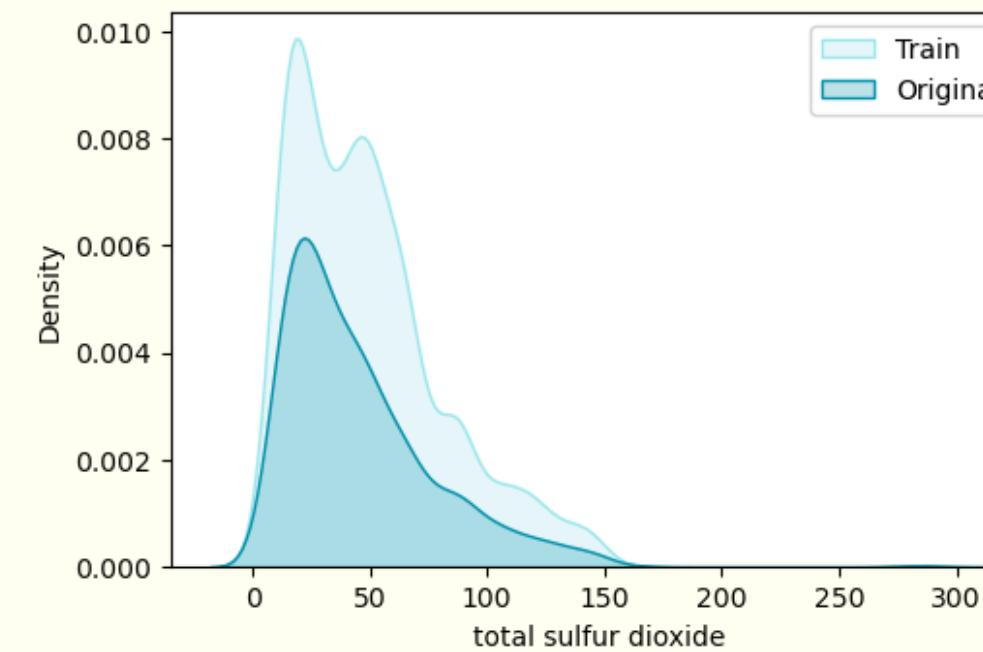


呈現
正相關

各因素分布狀況

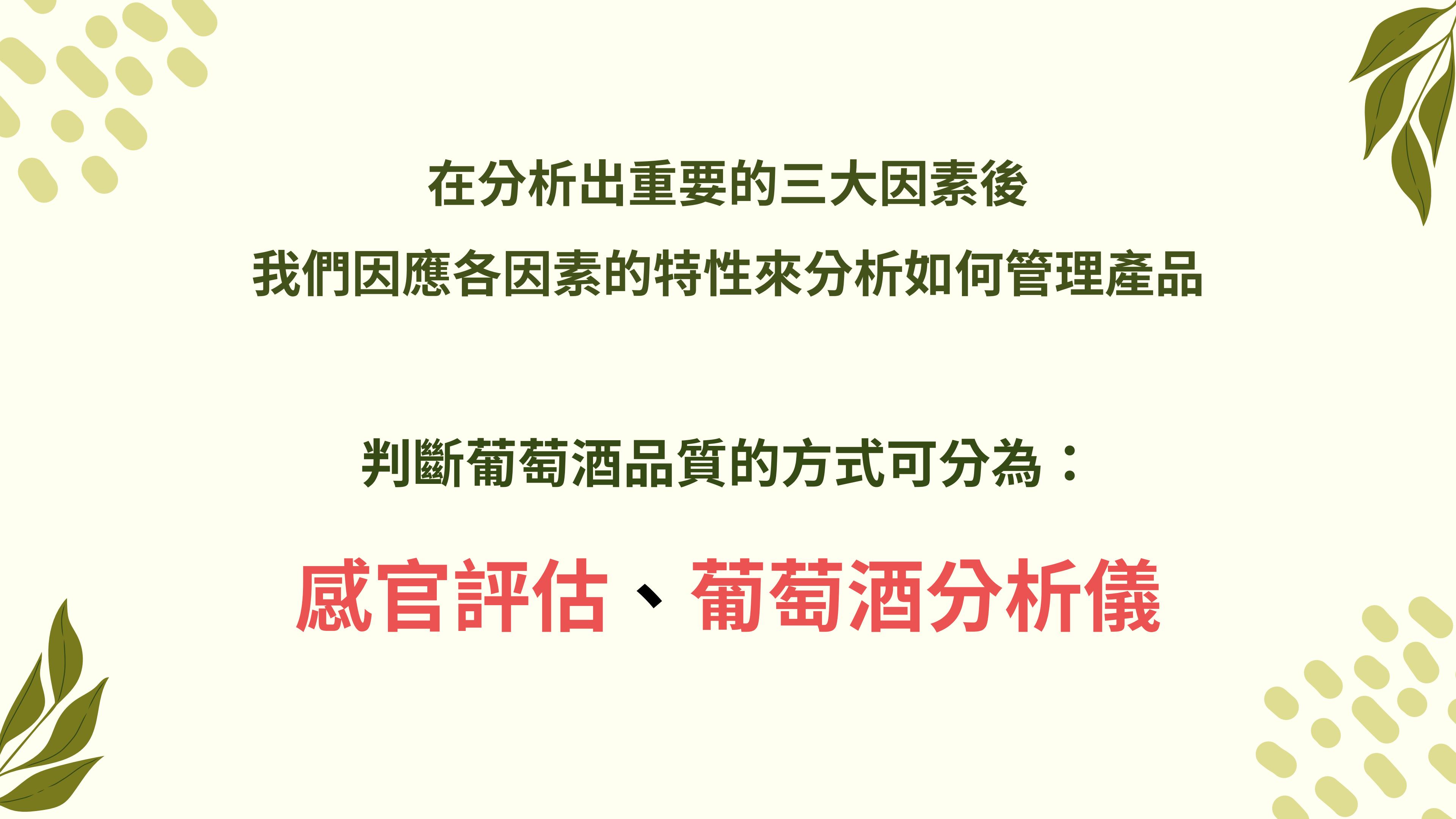


各因素分布狀況



06

評估成效



在分析出重要的三大因素後
我們因應各因素的特性來分析如何管理產品

判斷葡萄酒品質的方式可分為：

感官評估、葡萄酒分析儀

感官評估

檸檬酸

酸度、口感、果味突顯

亞硫酸鹽

香氣、風味、色澤

酒精

熱感、口感、和其他成分的平衡



葡萄酒分析儀

重複性為同一實驗室，分析人員用相同的分析法在短時間內對同一樣品重複測定結果之間的相對標準偏差。

要素	範圍	重複性
檸檬酸	pH3-4	0.01
酒精	濃度6-20	0.02



THANK
YOU