project ex 1

Idan pogrebinsky 325069565 Ilai avni 212778229

```
import pandas as pd
import matplotlib.pyplot as plt
```

Part 1

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. They contain information which was measured using physicochemical tests. The classes are ordered and not balanced (e.g. there are munch more normal wines than excellent or poor ones). Number of Instances: red wine - 1599; white wine - 4898.

- fixed acidity: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)
- volatile acidity: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste
- citric acid: found in small quantities, citric acid can add 'freshness' and flavor to wines
- density: the density of water is close to that of water depending on the percent alcohol and sugar content
- chlorides: the amount of salt in the wine
- pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale
- total sulfur dioxide: amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2 concentrations over 50 ppm, SO2 becomes evident in the nose and taste of wine
- free sulfur dioxide: the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine

- residual sugar: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet
- sulphates: the amount of sulphates in the wine, protects against oxidation and could affect the color and taste of wine.
- alcohol: The ethyl alcohol percentage in the wine
- quality: The quality of the wine

```
#load the datsets
red_wine_df = pd.read_csv('datasets/winequality-red.csv', sep=";")
white_wine_df = pd.read_csv('datasets/winequality-white.csv',sep=';' )
whole_wine_df = pd.concat([red_wine_df, white_wine_df], ignore_index=True)
#save them as proper CSVs
# red_wine_df.to_csv('datasets/edited/winequality-red.csv', sep=",")
# white_wine_df.to_csv('datasets/edited/winequality-white.csv', sep=",")
```

```
red_wine_df.columns
```

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol', 'quality'],
      dtype='object')
```

Question 1

Is there a connection between residual sugars and alcohol in red wine?

Hypothesis: the lower the residual sugar, the higher the alcohol content.

Question 2

Do red wines have higher citric acid values than white wines?

Hypothesis: red wine has higher values

Question 3

Are pH values of white wine different from pH values of red wine?

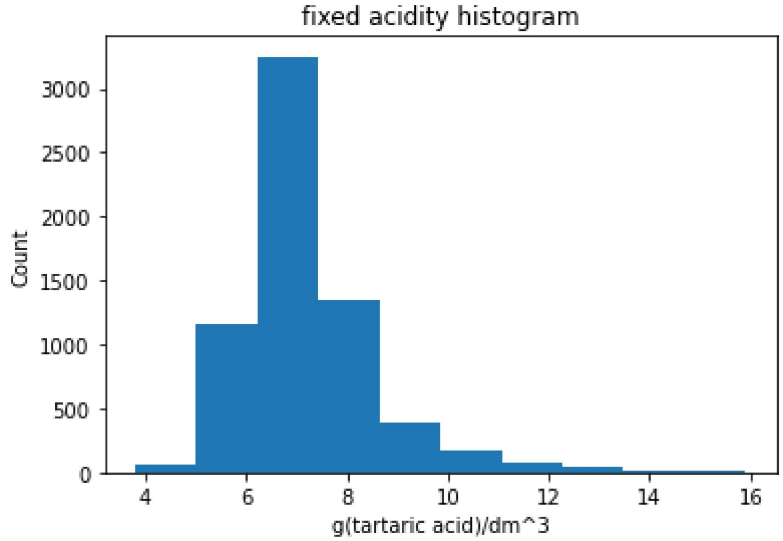
Part 2

```
label_dict = {"fixed acidity": "g(tartaric acid)/dm^3",
              "volatile acidity": "g(acetic acid)/dm^3",
              "citric acid": "g/dm^3",
              "residual sugar": "g/dm^3",
              "chlorides": "g(sodium chloride)/dm^3",
              "free sulfur dioxide": "mg/dm^3",
              "total sulfur dioxide": "mg/dm^3",
              "density": "g/cm^3",
              "pH": "pH",
              "sulphates": "g(potassium sulphate)/dm^3",
              "alcohol": "% vol.",
              "quality": "rating"}


for column in whole_wine_df.columns:
    y = whole_wine_df[column]
    plt.hist(y)
    plt.title(f'{column} histogram')
    plt.xlabel(label_dict[column])
```

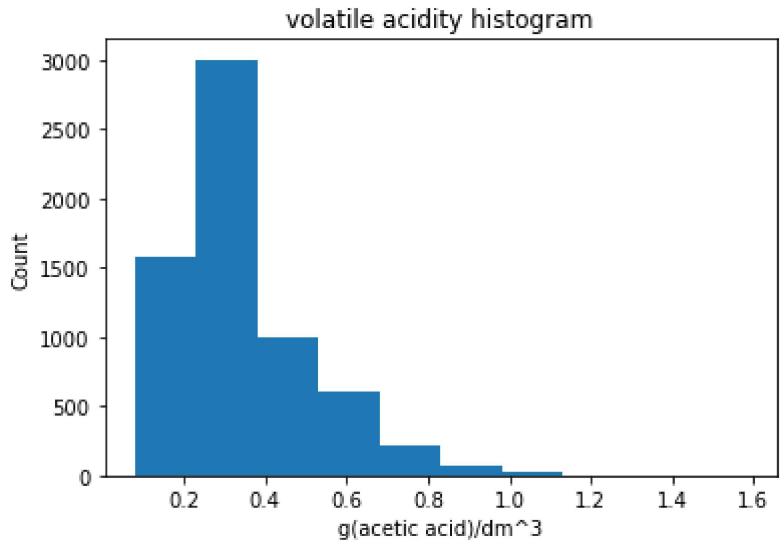
```
    plt.ylabel("Count")
    plt.show()
```

Download



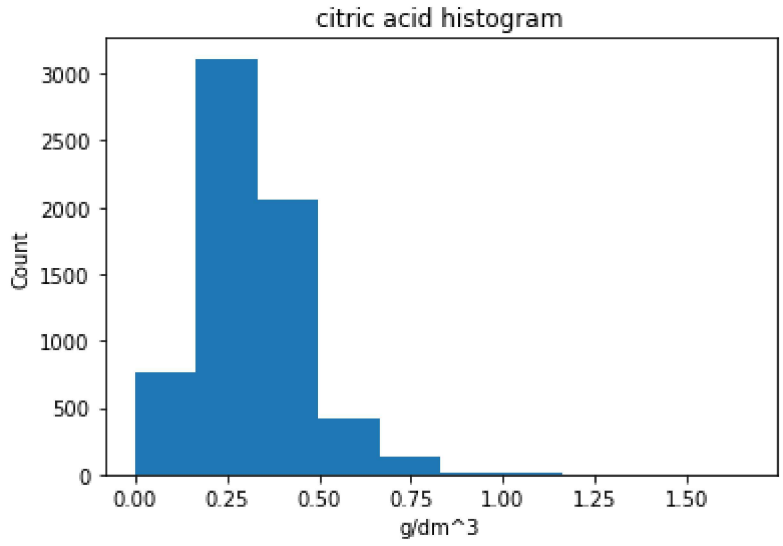

```
<Figure size 432x288 with 1 Axes>
```

Download



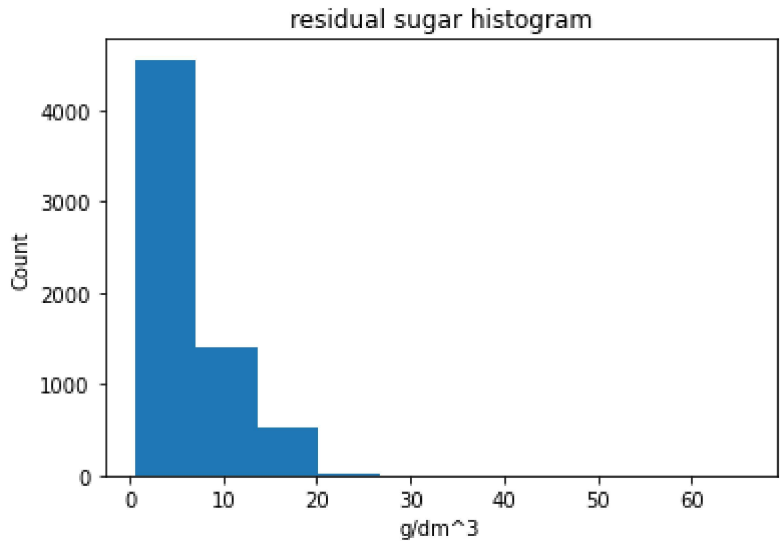

```
<Figure size 432x288 with 1 Axes>
```

Download



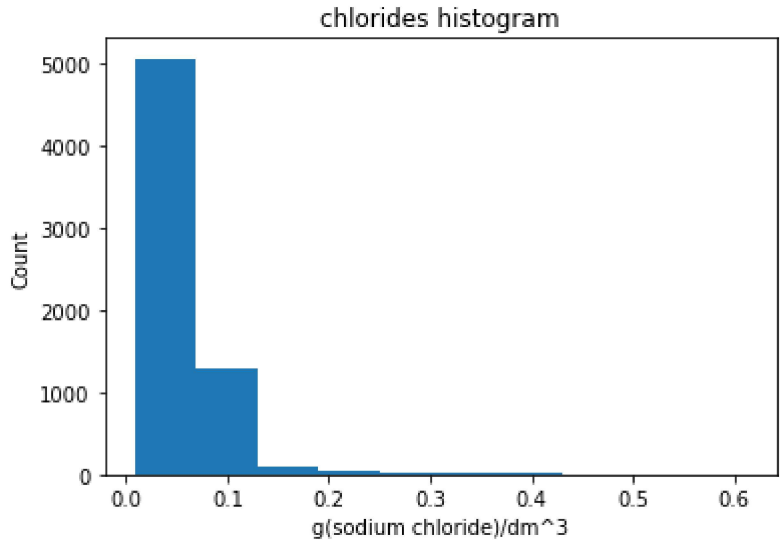

```
<Figure size 432x288 with 1 Axes>
```

Download



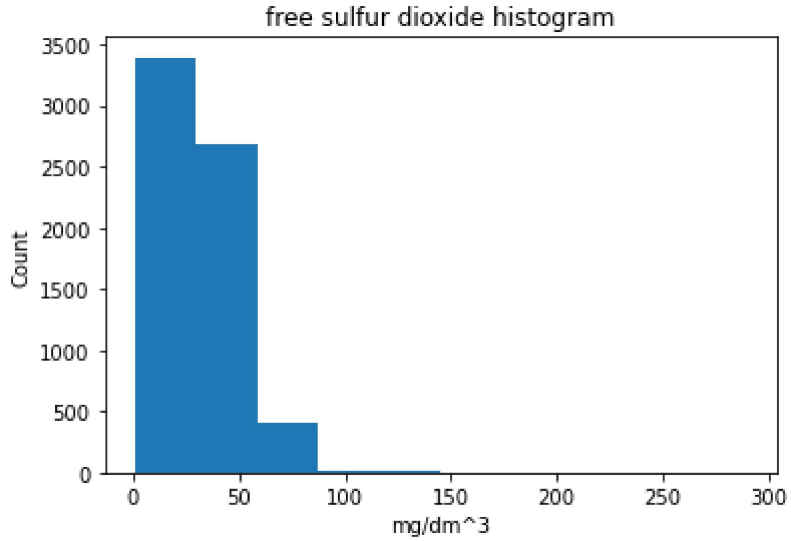

```
<Figure size 432x288 with 1 Axes>
```

Download



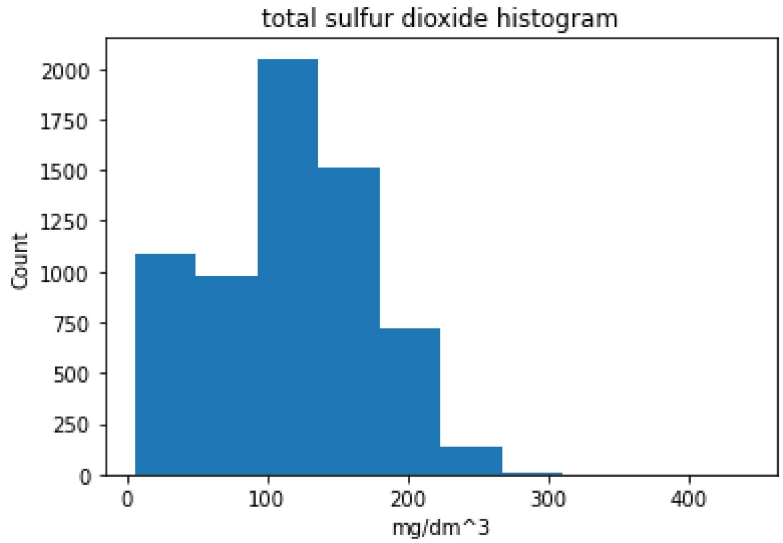

```
<Figure size 432x288 with 1 Axes>
```

Download



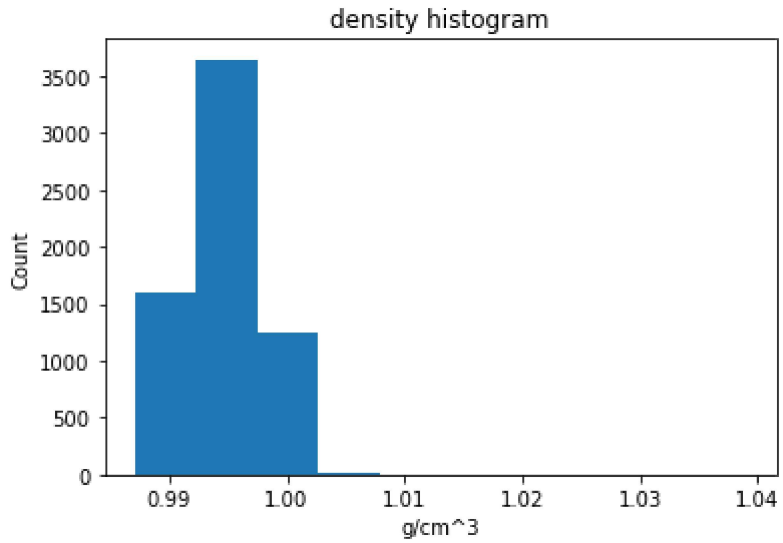

```
<Figure size 432x288 with 1 Axes>
```

Download



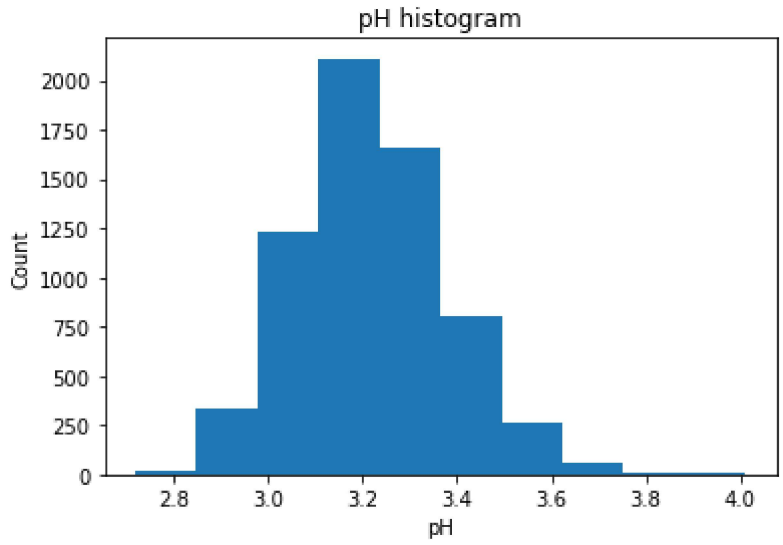

```
<Figure size 432x288 with 1 Axes>
```

Download



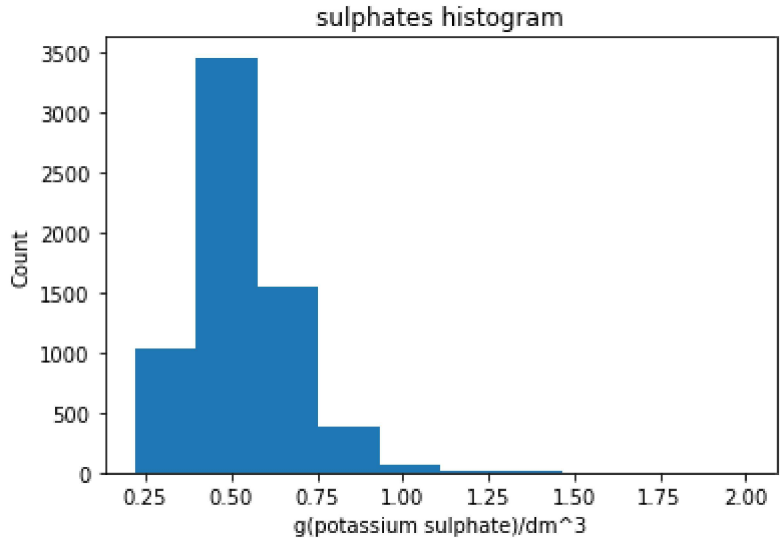

```
<Figure size 432x288 with 1 Axes>
```

Download



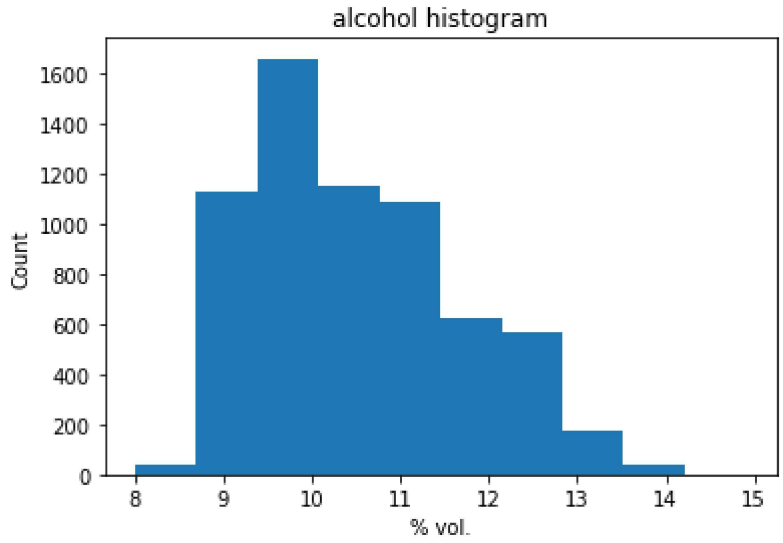

```
<Figure size 432x288 with 1 Axes>
```

Download



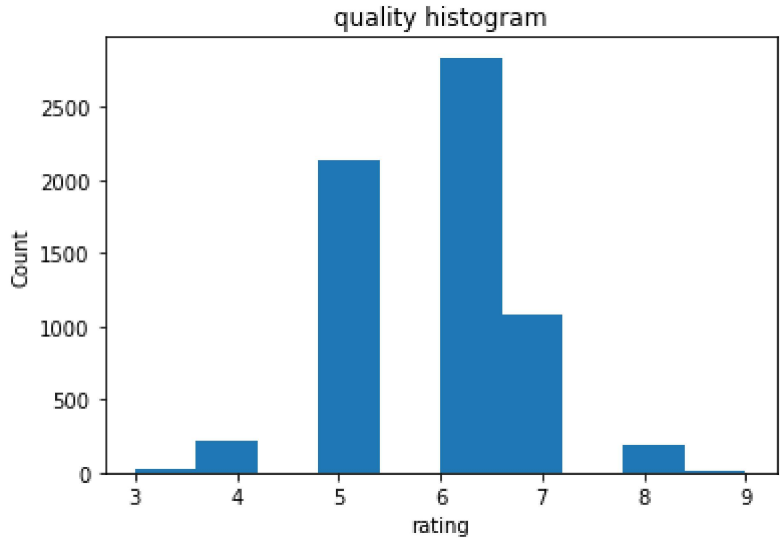

```
<Figure size 432x288 with 1 Axes>
```

Download




```
<Figure size 432x288 with 1 Axes>
```

Download




```
<Figure size 432x288 with 1 Axes>
```