

Test korelasi

Tentukan apakah setiap kolom non-target berkorelasi dengan kolom target, dengan menggambarkan juga scatter plot nya. Gunakan correlation test.

Daftar kolom non-target:

1. pH
2. Hardness
3. Solids
4. Chloramines
5. Sulfate
6. Conductivity
7. OrganicCarbon
8. Trihalomethanes
9. Turbidity

Daftar kolom target:

1. Potability

Metode test korelasi

1. Tes korelasi dengan metode pearson
 - $H_0 : \rho = 0$ (tidak ada korelasi antara kolom non-target dengan kolom target)
 - $H_1 : \rho \neq 0$ (ada korelasi antara kolom non-target dengan kolom target)
2. Beberapa jenis koefisien korelasi
 - $0 < \text{correlation coefficient} < 1$ (positive correlation)
 - $\text{correlation coefficient} = 1$ (strong positive correlation)
 - $-1 < \text{correlation coefficient} < 0$ (negative correlation)
 - $\text{correlation coefficient} = -1$ (strong negative correlation)
 - $\text{correlation coefficient} = 0$ (no correlation)
3. Scatter plot antara dua kolom

```
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats
from IPython.display import Markdown, display

df = pd.read_csv('water_potability.csv', index_col=0)

col_numeric = [
    'pH',
    'Hardness',
    'Solids',
    'Chloramines',
```

```

        'Sulfate',
        'Conductivity',
        'OrganicCarbon',
        'Trihalomethanes',
        'Turbidity'
    ]

df_potability = df['Potability']

for i in col_numeric:
    stat, p = stats.pearsonr(df[i], df_potability)
    display(Markdown("### Tes korelasi kolom {} dengan kolom  
potability : <br>".format(i)))
    isKorelasi = True
    if(p == 0):
        print("H0 diterima, berarti tidak ada korelasi antara kolom " +
i +
            " dengan kolom potability")
        isKorelasi = False
    else:
        print("H0 ditolak, berarti terdapat korelasi antara kolom " +
i +
            " dengan kolom potability")

    if(isKorelasi):
        coefcorr = df[i].corr(df_potability, method='pearson')
        print("Correlation Coeficient : " + str(coefcorr))
        if (coefcorr > 0 and coefcorr < 1):
            print("Correlation type : Positive Correlation.")
            print("Sehingga nilai " + i + " berbanding lurus dengan  
nilai potability")
        elif (coefcorr == 1):
            print("Correlation type : Strong Positive Correlation.")
            print("Sehingga nilai " + i + " berbanding lurus dengan  
nilai potability")
        elif (coefcorr > -1 and coefcorr < 0):
            print("Correlation type : Negative Correlation.")
            print("Sehingga nilai " + i + " berbanding terbalik dengan  
nilai potability")
        elif (coefcorr == -1):
            print("Correlation type : Strong Negative Correlation.")
            print("Sehingga nilai " + i + " berbanding terbalik dengan  
nilai potability")
        else:
            print("Correlation type : No Correlation")
    print()
    plt.scatter(df[i], df['Potability'])
    plt.title(i + " vs Potability")
    plt.xlabel(i)

```

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

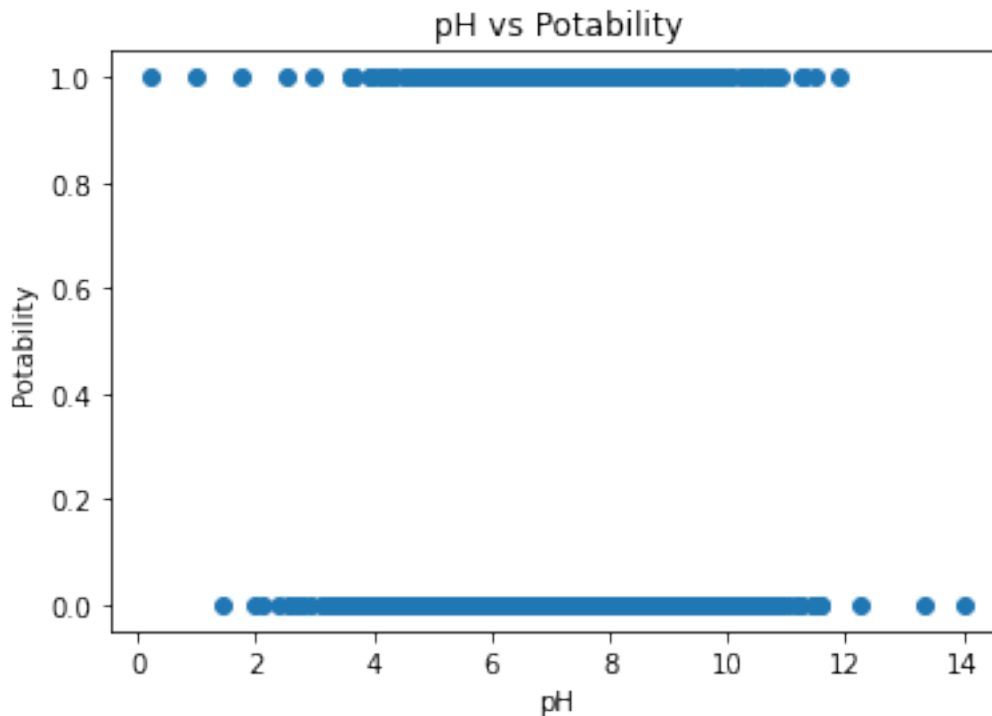
<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom pH dengan kolom potability

Correlation Coeficient : 0.01547509440843346

Correlation type : Positive Correlation.

Sehingga nilai pH berbanding lurus dengan nilai potability



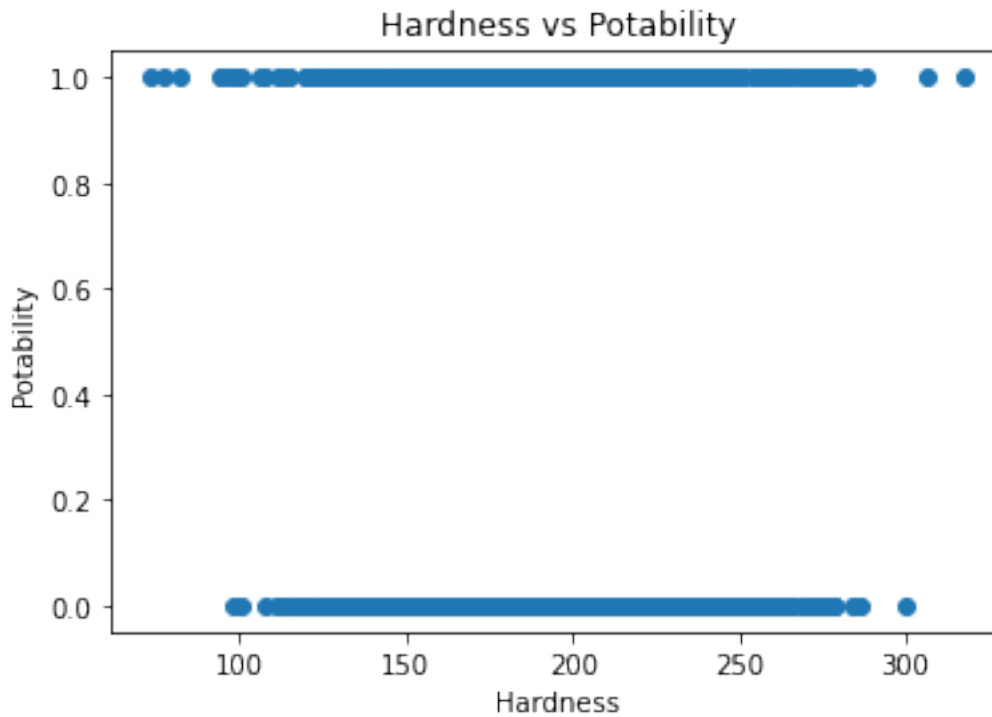
<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom Hardness dengan kolom potability

Correlation Coeficient : -0.0014631528959479338

Correlation type : Negative Correlation.

Sehingga nilai Hardness berbanding terbalik dengan nilai potability



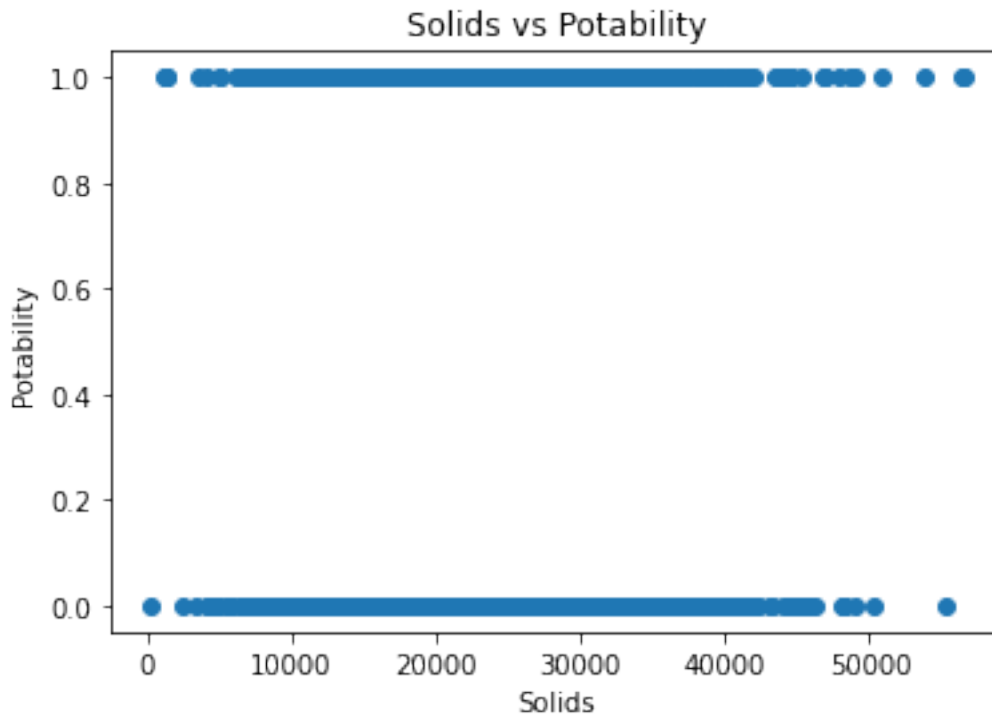
<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom Solids dengan kolom potability

Correlation Coefficient : 0.03897657818173466

Correlation type : Positive Correlation.

Sehingga nilai Solids berbanding lurus dengan nilai potability



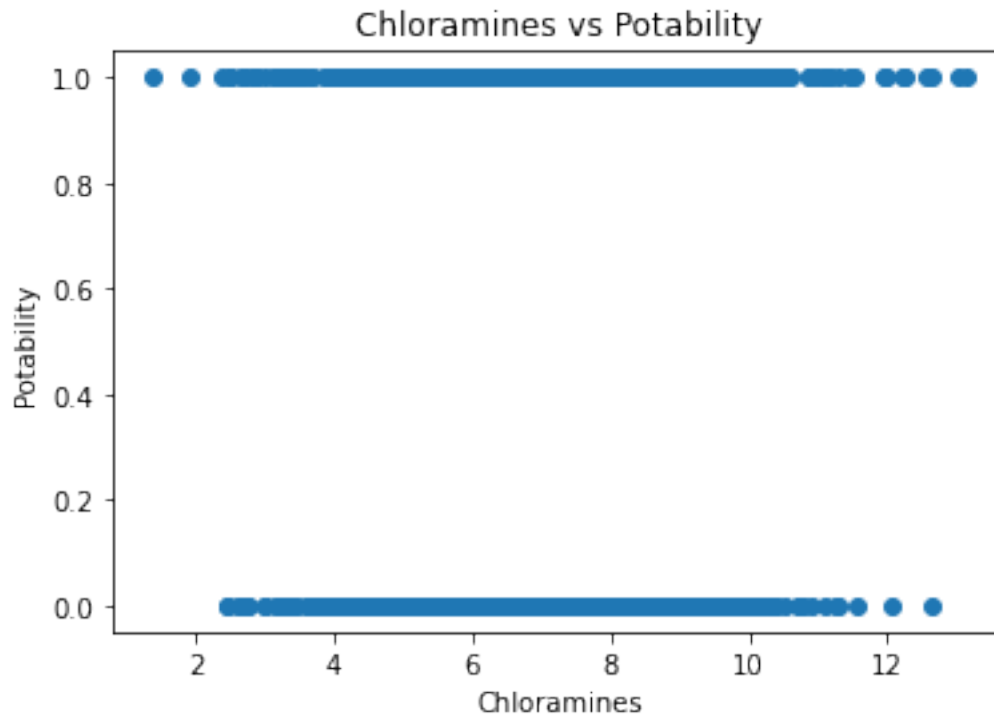
<IPython.core.display.Markdown object>

H_0 ditolak, berarti terdapat korelasi antara kolom Chloramines dengan kolom potability

Correlation Coefficient : 0.020778921840524076

Correlation type : Positive Correlation.

Sehingga nilai Chloramines berbanding lurus dengan nilai potability



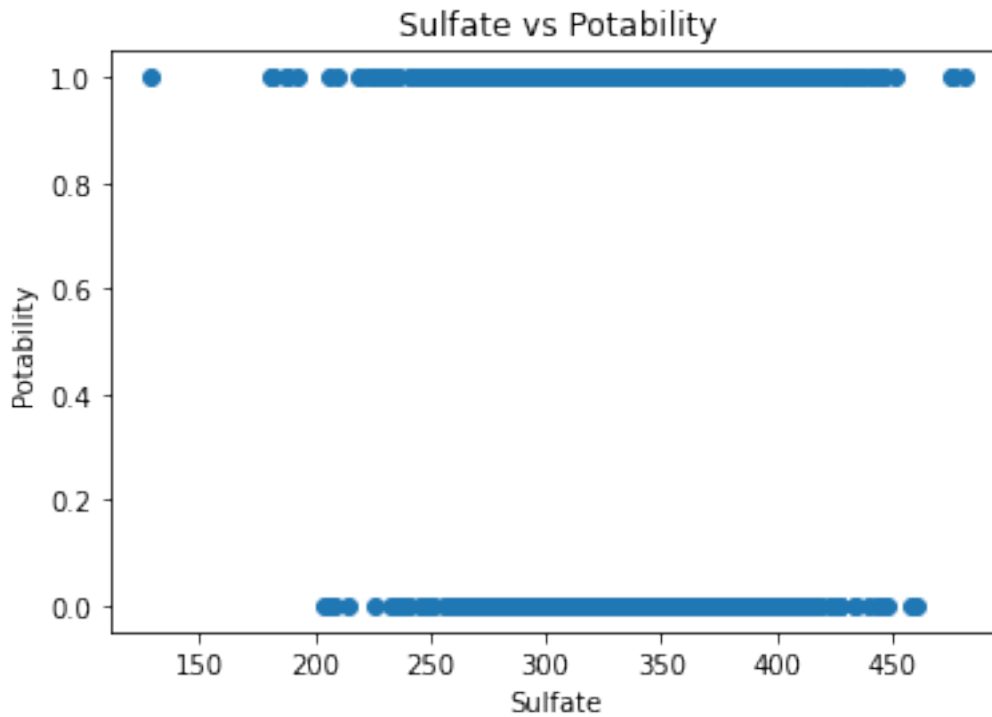
<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom Sulfate dengan kolom potability

Correlation Coefficient : -0.015703164419273802

Correlation type : Negative Correlation.

Sehingga nilai Sulfate berbanding terbalik dengan nilai potability



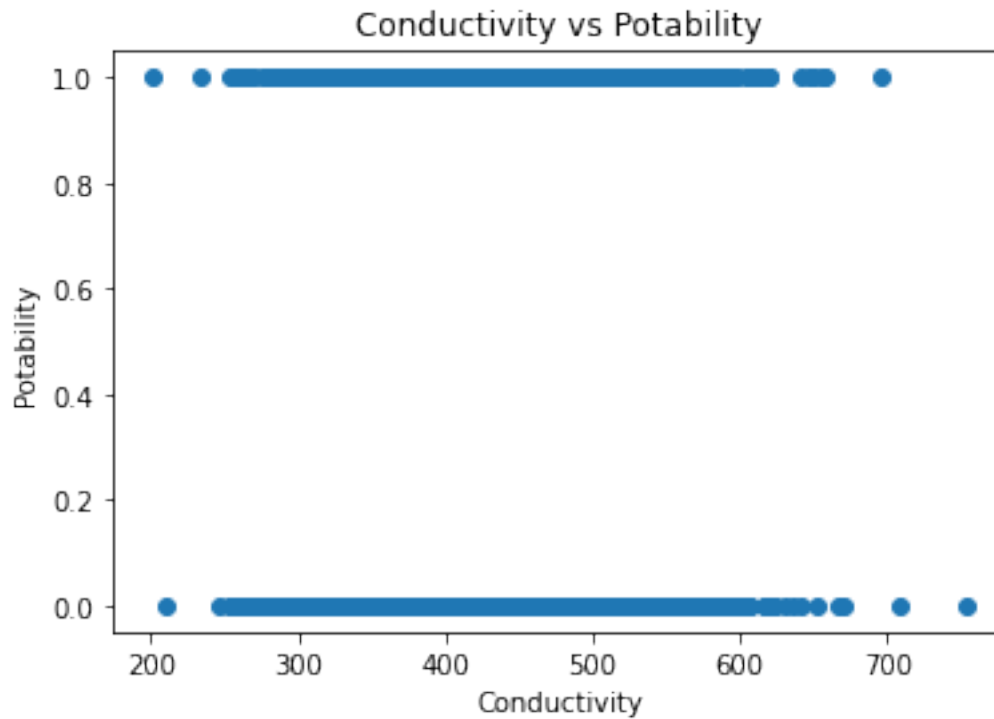
<IPython.core.display.Markdown object>

H_0 ditolak, berarti terdapat korelasi antara kolom Conductivity dengan kolom potability

Correlation Coefficient : -0.01625712011137705

Correlation type : Negative Correlation.

Sehingga nilai Conductivity berbanding terbalik dengan nilai potability



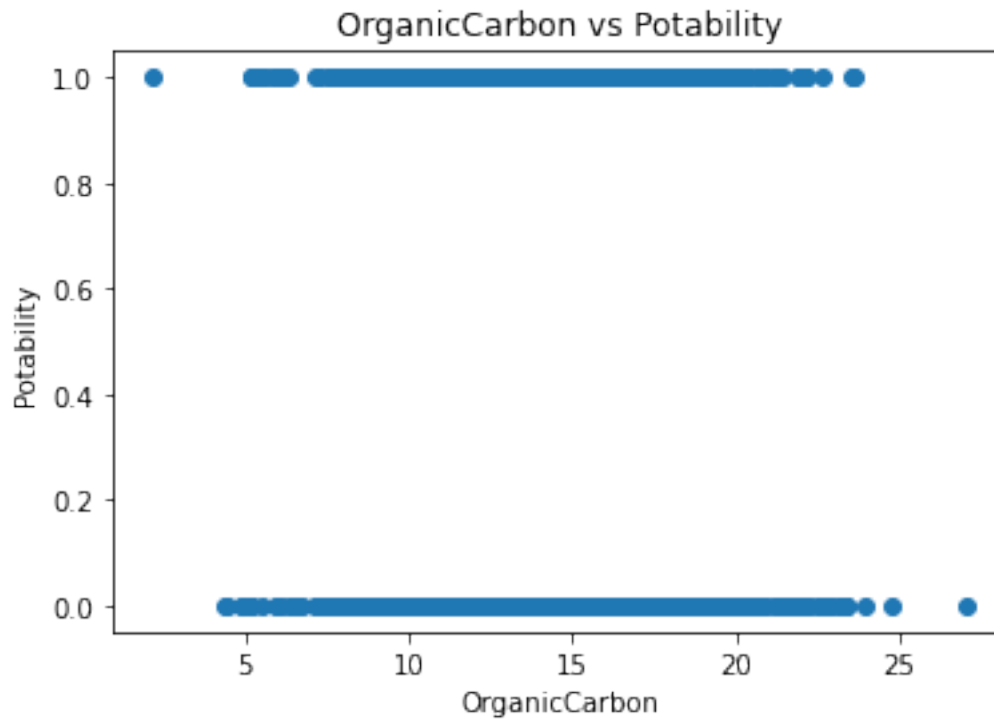
<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom OrganicCarbon dengan kolom potability

Correlation Coefficient : -0.015488461910747266

Correlation type : Negative Correlation.

Sehingga nilai OrganicCarbon berbanding terbalik dengan nilai potability



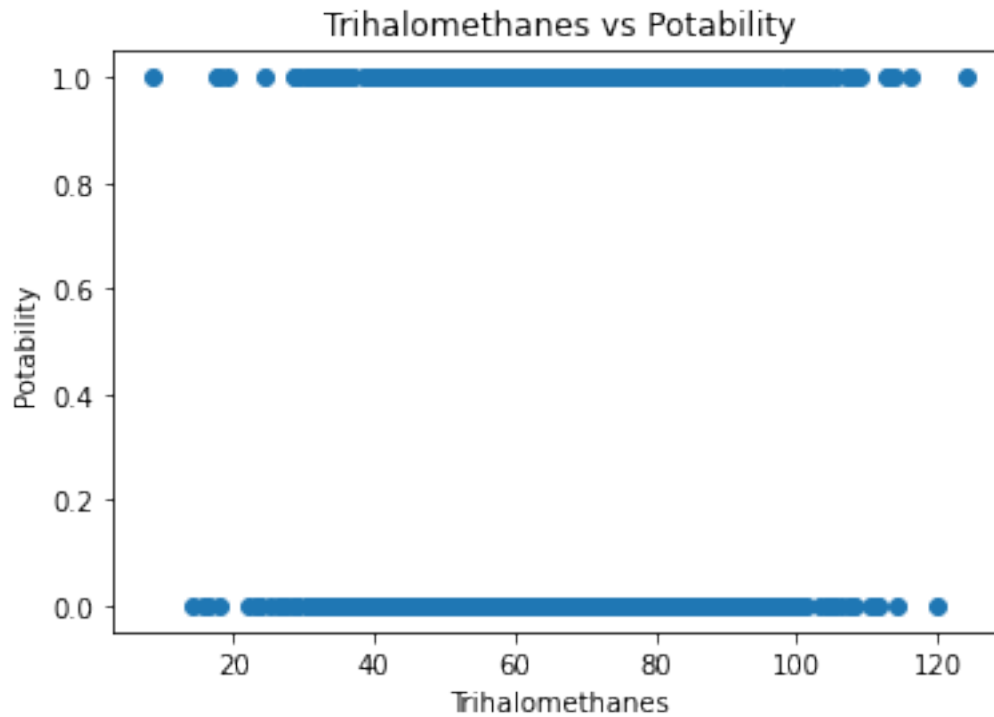
<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom Trihalomethanes dengan kolom potability

Correlation Coeficient : 0.009236711064713014

Correlation type : Positive Correlation.

Sehingga nilai Trihalomethanes berbanding lurus dengan nilai potability



<IPython.core.display.Markdown object>

H0 ditolak, berarti terdapat korelasi antara kolom Turbidity dengan kolom potability

Correlation Coefficient : 0.022331042640622658

Correlation type : Positive Correlation.

Sehingga nilai Turbidity berbanding lurus dengan nilai potability

