

```
In [ ]: # Tugas Besar Mata Kuliah Probabilitas dan Statistika

# Anggota
# - Brianaldo Phandiarta - 13520113
# - M Syahrul Surya Putra - 13520161
```

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import statistics
import scipy.stats as st
from statsmodels.stats.proportion import proportions_ztest
from IPython.display import display, Markdown

df = pd.read_csv("water_potability.csv",
                 names=["id",
                       "pH",
                       "Hardness",
                       "Solids",
                       "Chloramines",
                       "Sulfate",
                       "Conductivity",
                       "OrganicCarbon",
                       "Trihalomethanes",
                       "Turbidity",
                       "Potability"])
df.set_index("id")
```

## Soal 1

```
In [ ]: def dataDescription(data):
    dataFrame = df[data]

    display(Markdown("-- Deskripsi Data %s --" % data))
    display(Markdown("Mean: %s" % dataFrame.mean()))
    display(Markdown("Median: %s" % dataFrame.median()))
    display(Markdown("Modus: %s" % statistics.mode(dataFrame)))
    display(Markdown("Standar Deviasi: %s" % dataFrame.std()))
    display(Markdown("Variansi: %s" % dataFrame.var()))
    display(Markdown("Range: %s" % (dataFrame.max() - dataFrame.min())))
    display(Markdown("Nilai Minimum: %s" % dataFrame.min()))
    display(Markdown("Nilai Maksimum: %s" % dataFrame.max()))
    display(Markdown("Skewness: %s" % dataFrame.skew()))
    display(Markdown("Kurtosis: %s\n" % dataFrame.kurt()))
    display(Markdown("<br>-- Kuartil --"))
    display(Markdown("Kuartil 1: %s<br>Kuartil 2: %s<br>Kuartil 3: %s" % (dataFrame.qu
    display(Markdown("IQR: %s" % (dataFrame.quantile(0.75) - dataFrame.quantile(0.25))
```

```
In [ ]: dataDescription("pH")
```

-- Deskripsi Data pH --

Mean: 7.0871927687138285

Median: 7.029490455474185

Modus: 8.316765884214679

Standar Deviasi: 1.5728029470456655

Variansi: 2.4737091102355304

Range: 13.7725009497978

Nilai Minimum: 0.2274990502021987

Nilai Maksimum: 13.999999999999998

Skewness: 0.04853451405270669

Kurtosis: 0.6269041256617065

-- Kuartil --

Kuartil 1: 6.09078502142353

Kuartil 2: 7.029490455474185

Kuartil 3: 8.053006240791538

IQR: 1.9622212193680078

```
In [ ]: dataDescription("Hardness")
```

-- Deskripsi Data Hardness --

Mean: 195.96920903783524

Median: 197.20352491941043

Modus: 214.3733940856225

Standar Deviasi: 32.643165859429864

Variansi: 1065.5762773262472

Range: 243.84589036652147

Nilai Minimum: 73.4922336890611

Nilai Maksimum: 317.33812405558257

Skewness: -0.08532104172868622

Kurtosis: 0.5254804942991402

-- Kuartil --

Kuartil 1: 176.74065667669896

Kuartil 2: 197.20352491941043

Kuartil 3: 216.44758866727156

IQR: 39.7069319905726

```
In [ ]: dataDescription("Solids")
```

-- Deskripsi Data Solids --

Mean: 21904.673439053095

Median: 20926.88215534375  
Modus: 22018.41744077529  
Standar Deviasi: 8625.397911190576  
Variansi: 74397489.12637076  
Range: 56167.72980146483  
Nilai Minimum: 320.942611274359  
Nilai Maksimum: 56488.67241273919  
Skewness: 0.5910113724580447  
Kurtosis: 0.33732026745944976

-- Kuartil --

Kuartil 1: 15614.412961614333  
Kuartil 2: 20926.88215534375  
Kuartil 3: 27170.534648603603  
IQR: 11556.12168698927

```
In [ ]: dataDescription("Chloramines")
```

-- Deskripsi Data Chloramines --

Mean: 7.134322344600104  
Median: 7.1420143046226645  
Modus: 8.05933237743854  
Standar Deviasi: 1.5852140982642102  
Variansi: 2.512903737335613  
Range: 11.736129095114823  
Nilai Minimum: 1.3908709048851806  
Nilai Maksimum: 13.127000000000002  
Skewness: 0.013003497779569528  
Kurtosis: 0.5497821097667472

-- Kuartil --

Kuartil 1: 6.138326387572855  
Kuartil 2: 7.1420143046226645  
Kuartil 3: 8.109933216133502  
IQR: 1.9716068285606472

```
In [ ]: dataDescription("Sulfate")
```

-- Deskripsi Data Sulfate --

Mean: 333.211376415189

Median: 332.2141128069568

Modus: 356.88613564305666

Standar Deviasi: 41.21111102560979

Variansi: 1698.355671965137

Range: 352.03064230599716

Nilai Minimum: 129.00000000000003

Nilai Maksimum: 481.0306423059972

Skewness: -0.04572780443653543

Kurtosis: 0.7868544988131605

-- Kuartil --

Kuartil 1: 307.6269864860709

Kuartil 2: 332.2141128069568

Kuartil 3: 359.26814739141554

IQR: 51.641160905344634

In [ ]: `dataDescription("Conductivity")`

-- Deskripsi Data Conductivity --

Mean: 426.47670835257907

Median: 423.43837202443706

Modus: 363.2665161642437

Standar Deviasi: 80.70187180729437

Variansi: 6512.792113200974

Range: 551.7228828031471

Nilai Minimum: 201.6197367551575

Nilai Maksimum: 753.3426195583046

Skewness: 0.26801233302645316

Kurtosis: -0.23720600574806516

-- Kuartil --

Kuartil 1: 366.61921929632433

Kuartil 2: 423.43837202443706

Kuartil 3: 482.2097724598859

IQR: 115.5905531635616

```
In [ ]: dataDescription("OrganicCarbon")
```

-- Deskripsi Data OrganicCarbon --

Mean: 14.357939902048074

Median: 14.323285610653329

Modus: 18.436524495493305

Standar Deviasi: 3.3257700016987197

Variansi: 11.0607461041991

Range: 24.80670661116602

Nilai Minimum: 2.1999999999999886

Nilai Maksimum: 27.00670661116601

Skewness: -0.02021975629181238

Kurtosis: 0.031018388192253

-- Kuartil --

Kuartil 1: 12.122530374047727

Kuartil 2: 14.323285610653329

Kuartil 3: 16.683561746173808

IQR: 4.561031372126081

```
In [ ]: dataDescription("Trihalomethanes")
```

-- Deskripsi Data Trihalomethanes --

Mean: 66.40071666307466

Median: 66.48204080309809

Modus: 100.34167436508008

Standar Deviasi: 16.08110898232513

Variansi: 258.60206610141796

Range: 115.4229870670162

Nilai Minimum: 8.577012932983806

Nilai Maksimum: 124.0

Skewness: -0.05138268451619478

Kurtosis: 0.2230167810639787

-- Kuartil --

Kuartil 1: 55.94999302803186

Kuartil 2: 66.48204080309809

Kuartil 3: 77.2946128060674

IQR: 21.344619778035543

```
In [ ]: dataDescription("Turbidity")
```

-- Deskripsi Data Turbidity --

Mean: 3.9694969126303676

Median: 3.967373963531836

Modus: 4.628770536837084

Standar Deviasi: 0.7804710407083957

Variansi: 0.6091350453844462

Range: 5.044748555990993

Nilai Minimum: 1.45

Nilai Maksimum: 6.494748555990993

Skewness: -0.03226597968019271

Kurtosis: -0.049830796949249745

-- Kuartil --

Kuartil 1: 3.442881623557439

Kuartil 2: 3.967373963531836

Kuartil 3: 4.5146627202018825

IQR: 1.0717810966444437

## Soal 2

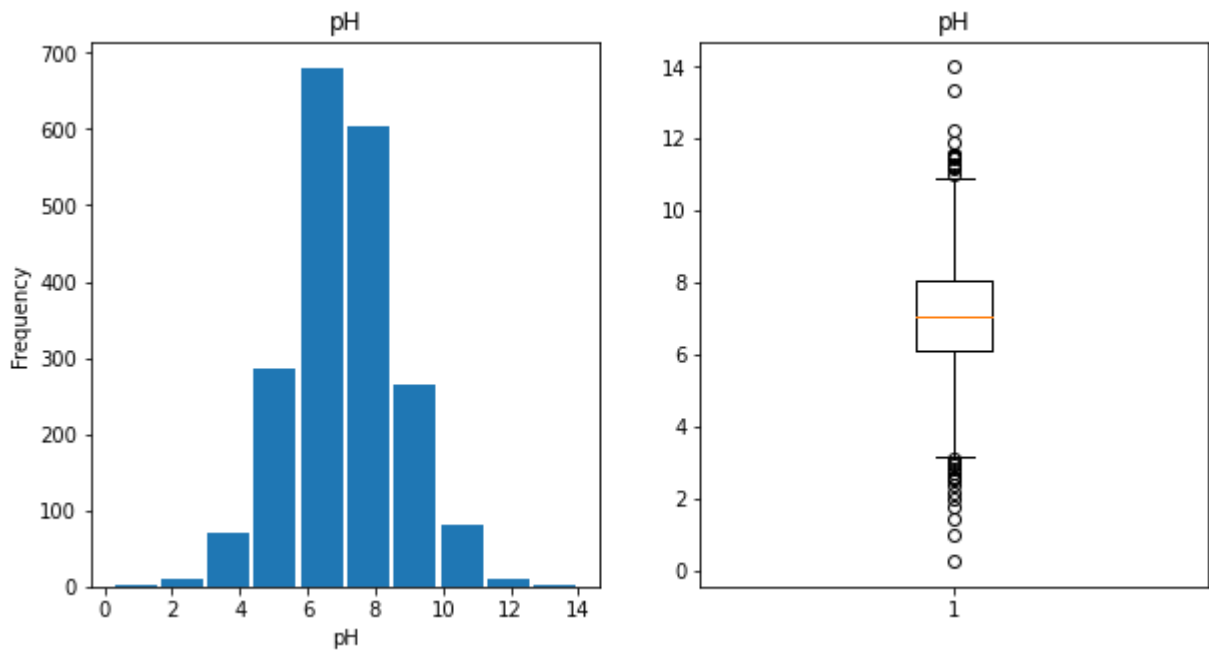
```
In [ ]: def plotHistogramAndBoxplot(name):
        fig = plt.figure(figsize=(10, 5))

        histogram = plt.subplot(1, 2, 1)
        plt.hist(df[name], rwidth=0.9)
        histogram.set_title(name)
        histogram.set_xlabel(name)
        histogram.set_ylabel("Frequency")

        boxplot = plt.subplot(1, 2, 2)
        plt.boxplot(df[name])
        boxplot.set_title(name)

        fig
```

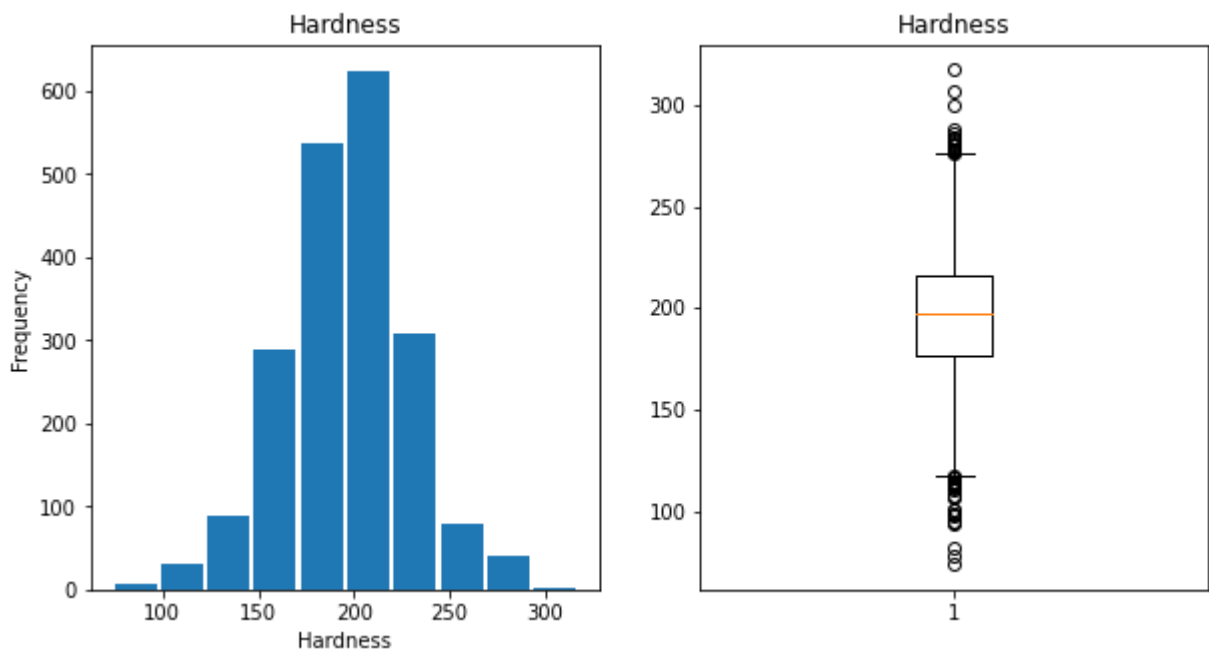
```
In [ ]: plotHistogramAndBoxplot("pH")
```



Berdasarkan plot histogram, histogram dari data pH berbentuk simetris.

Berdasarkan plot box plot, data dari median berada di tengah dengan whisker bagian atas dan bawah sama panjang. Terlihat juga bahwa terdapat pencilan yang banyak.

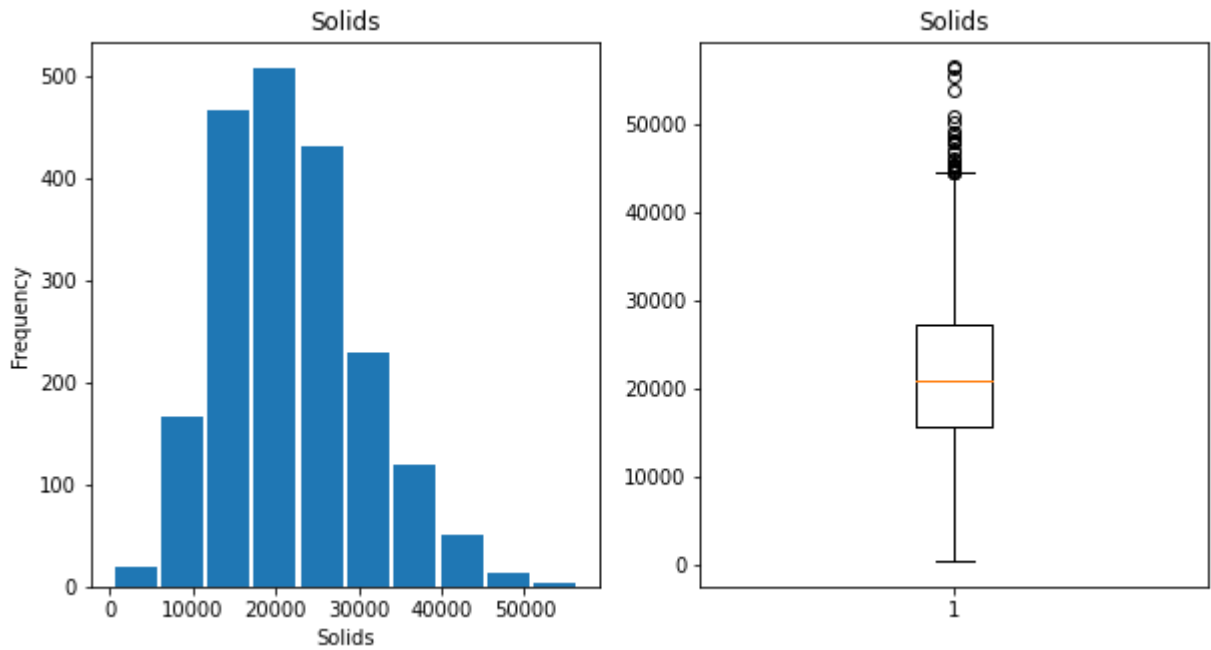
```
In [ ]: plotHistogramAndBoxplot("Hardness")
```



Berdasarkan plot histogram, histogram dari data Hardness berbentuk simetris, tetapi memiliki sedikit kecenderungan ke arah kiri (*left skewed*).

Berdasarkan plot box plot, data dari median berada di tengah dengan whisker bagian atas dan bawah sama panjang. Terlihat juga bahwa terdapat pencilan yang banyak.

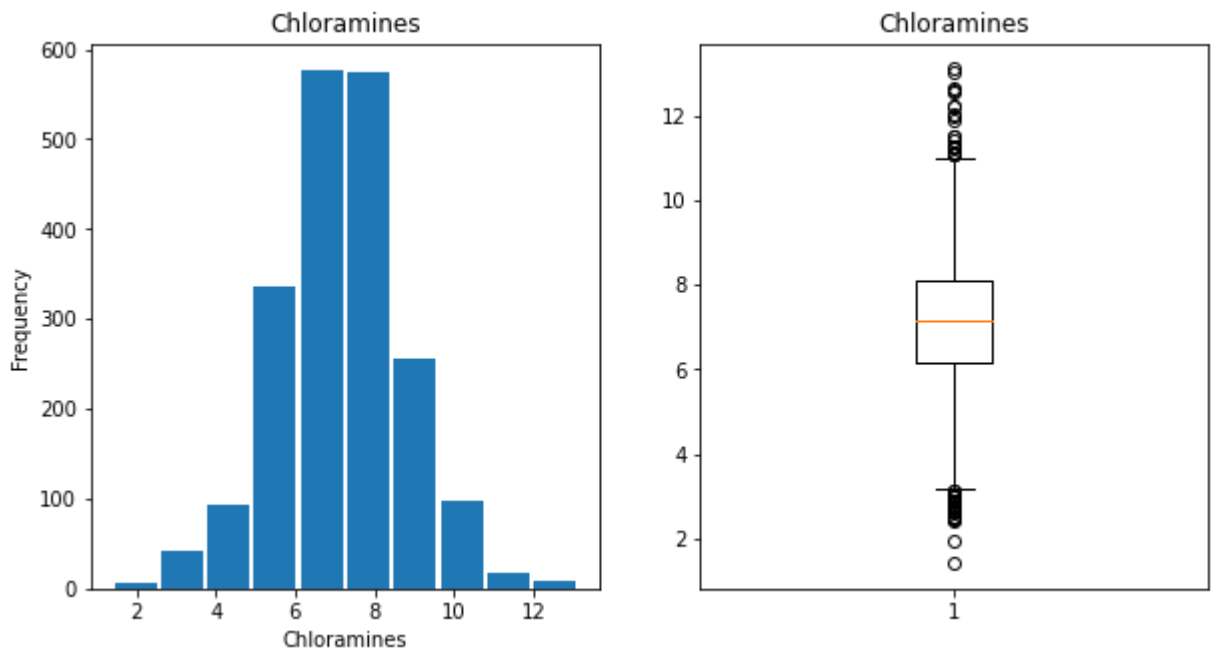
```
In [ ]: plotHistogramAndBoxplot("Solids")
```



Berdasarkan plot histogram, histogram dari data Solids memiliki kecenderungan ke arah kanan (*right skewed*).

Berdasarkan plot box plot, terlihat terdapat pencilan yang banyak pada batas atas yang berarti banyak data yang memiliki nilai lebih besar dari batas atas nilai maksimum.

```
In [ ]: plotHistogramAndBoxplot("Chloramines")
```

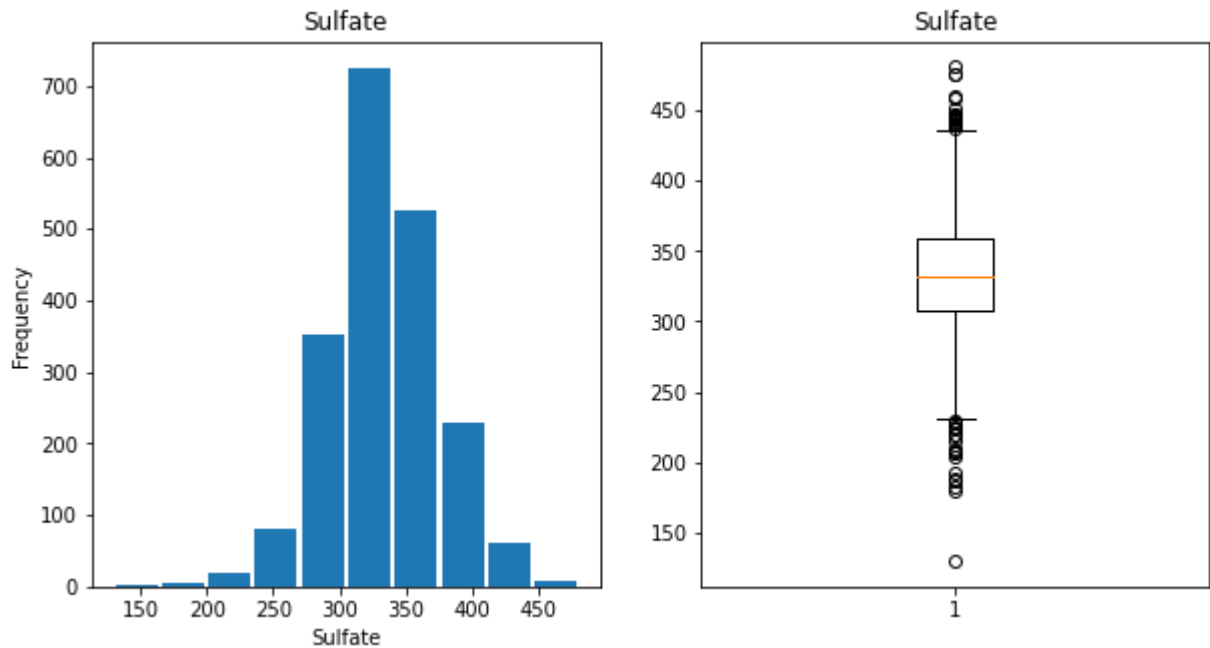


Berdasarkan plot histogram, histogram dari data Chloromines berbentuk simetris.

Berdasarkan plot box plot, data dari median berada di tengah dengan whisker bagian atas dan bawah sama panjang. Terlihat juga bahwa terdapat pencilan yang banyak.

```
In [ ]: plotHistogramAndBoxplot("Sulfate")
```

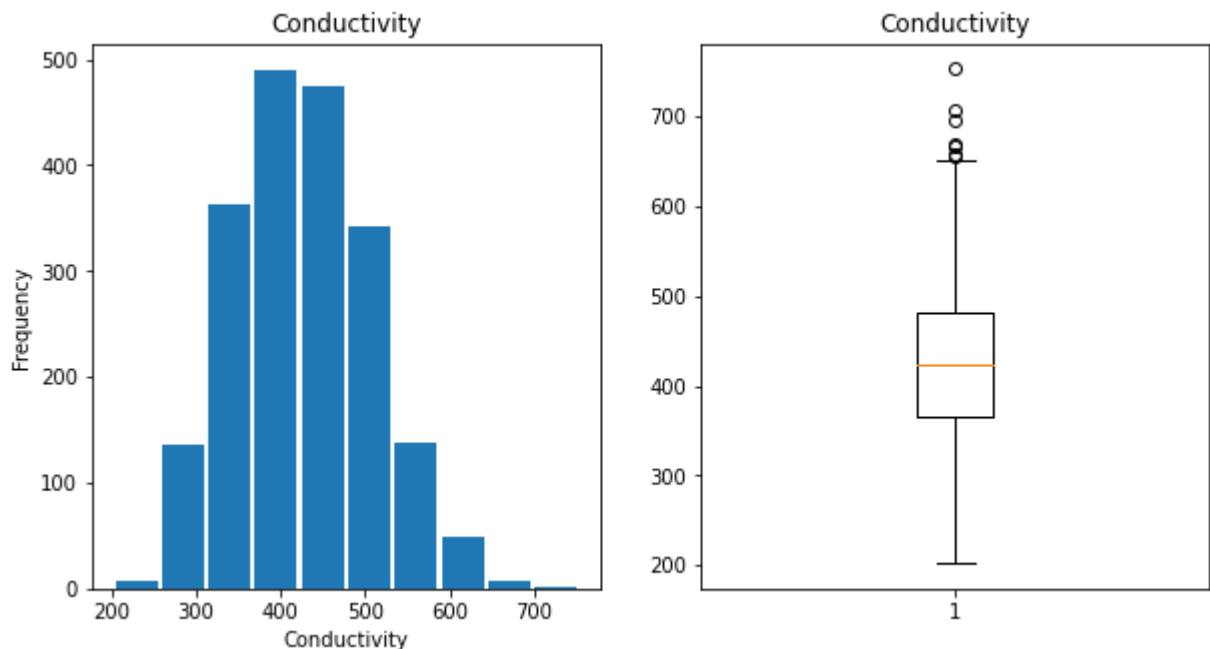




Berdasarkan plot histogram, histogram dari data Sulfate memiliki kecenderungan ke arah kiri (*right skewed*).

Berdasarkan plot box plot, terlihat terdapat pencilan yang banyak pada batas bawah yang berarti banyak data yang memiliki nilai lebih kecil dari batas bawah nilai minimum.

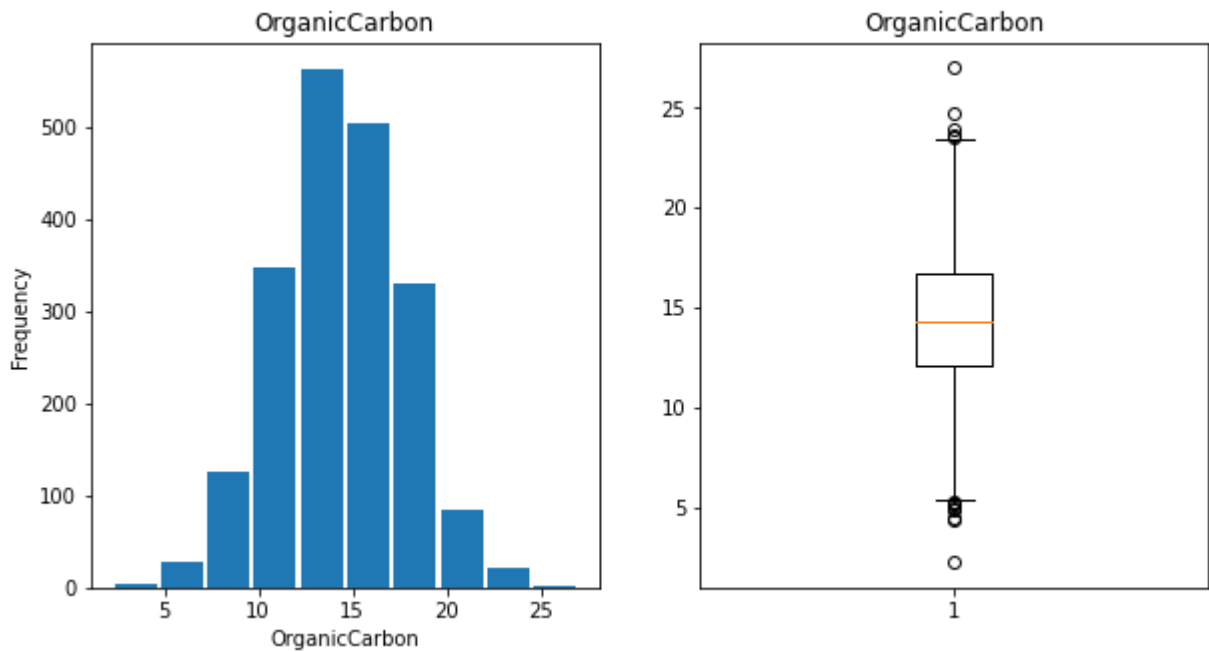
```
In [ ]: plotHistogramAndBoxplot("Conductivity")
```



Berdasarkan plot histogram, histogram dari data Conductivity memiliki kecenderungan ke arah kanan (*right skewed*).

Berdasarkan plot box plot, terlihat terdapat pencilan yang banyak pada batas atas yang berarti banyak data yang memiliki nilai lebih besar dari batas atas nilai maksimum.

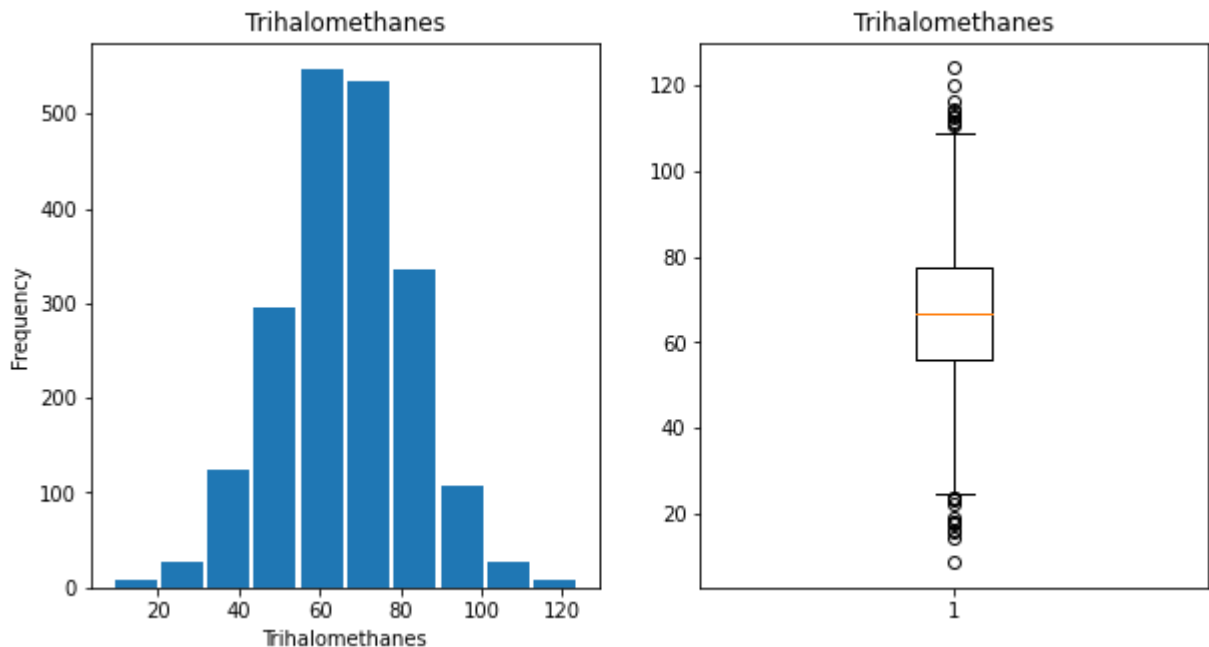
```
In [ ]: plotHistogramAndBoxplot("OrganicCarbon")
```



Berdasarkan plot histogram, histogram dari data OrganicCarbon berbentuk simetris.

Berdasarkan plot box plot, data dari median berada di tengah dengan whisker bagian atas dan bawah sama panjang. Terlihat juga bahwa terdapat pencilan yang banyak. Distribusi pencilan atas dan bawah juga terlihat simetris.

```
In [ ]: plotHistogramAndBoxplot("Trihalomethanes")
```

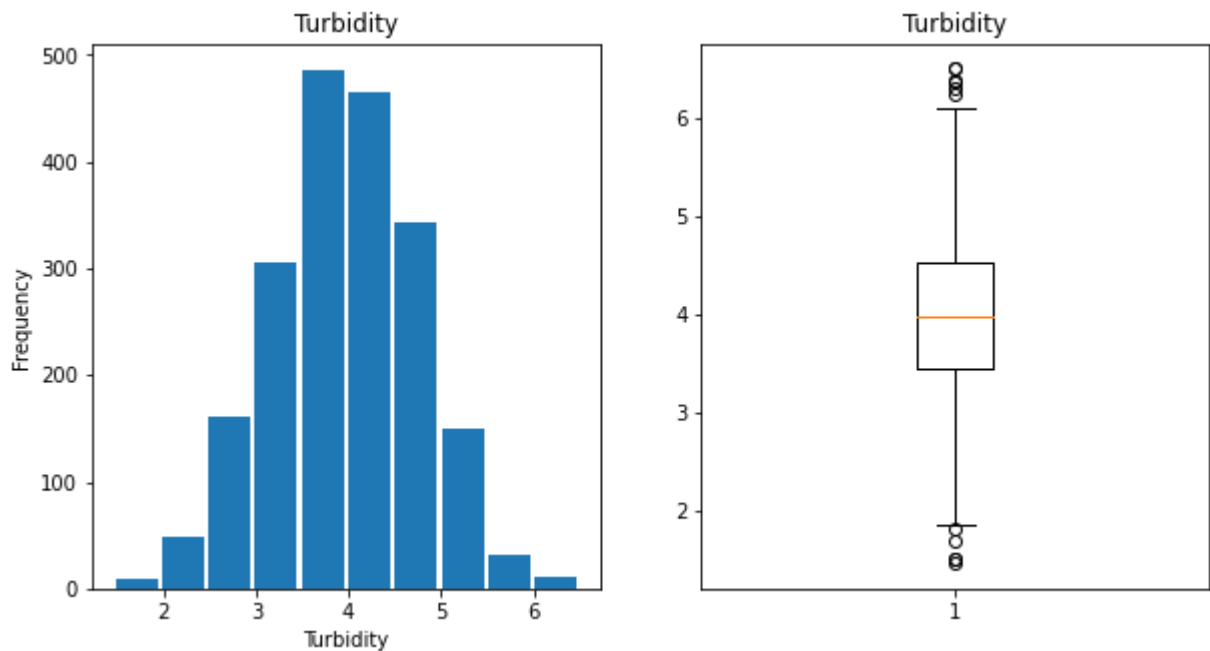


Berdasarkan plot histogram, histogram dari data Trihalomethanes berbentuk simetris.

Berdasarkan plot box plot, data dari median berada di tengah dengan whisker bagian atas dan bawah sama panjang. Terlihat juga bahwa terdapat pencilan yang banyak. Distribusi pencilan

atas dan bawah juga terlihat simetris.

```
In [ ]: plotHistogramAndBoxplot("Turbidity")
```



Berdasarkan plot histogram, histogram dari data Turbidity berbentuk simetris.

Berdasarkan plot box plot, data dari median berada di tengah dengan whisker bagian atas dan bawah sama panjang. Terlihat juga bahwa terdapat pencilan yang banyak. Distribusi pencilan atas dan bawah juga terlihat simetris.

### Soal 3

Data pH terlihat dari plot histogram yang berbentuk simetris dapat dikatakan memiliki distribusi normal. Akan tetapi, jika dilihat dari plot box plot, pencilan terlihat tidak simetris. Sehingga, data pH bukan termasuk distribusi normal.

Data Hardness terlihat dari plot histogram yang berbentuk simetris dapat dikatakan memiliki distribusi normal, walaupun memiliki sedikit kecenderungan ke arah kiri. Akan tetapi, jika dilihat dari plot box plot, pencilan terlihat tidak simetris. Sehingga, data Hardness bukan termasuk distribusi normal.

Data Solids terlihat dari plot histogram yang memiliki kecenderungan ke arah kanan. Sehingga, data Solids bukan termasuk distribusi normal.

Data Chloromines terlihat dari plot histogram yang berbentuk simetris dapat dikatakan memiliki distribusi normal. Akan tetapi, jika dilihat dari plot box plot, pencilan terlihat tidak simetris. Sehingga, data Chloromines bukan termasuk distribusi normal.

Data Sulfate terlihat dari plot histogram yang memiliki kecenderungan ke arah kiri. Sehingga, data Sulfate bukan termasuk distribusi normal.

Data Conductivity terlihat dari plot histogram yang memiliki kecenderungan ke arah kanan. Sehingga, data Conductivity bukan termasuk distribusi normal.

Data OrganicCarbon terlihat dari plot histogram yang berbentuk simetris dapat dikatakan memiliki distribusi normal. Jika dilihat dari plot box plot, pencilan terlihat terdistribusi secara simetris. Sehingga, data OrganicCarbon termasuk distribusi normal.

Data Trihalomethanes terlihat dari plot histogram yang berbentuk simetris dapat dikatakan memiliki distribusi normal. Jika dilihat dari plot box plot, pencilan terlihat terdistribusi secara simetris. Sehingga, data Trihalomethanes termasuk distribusi normal.

Data Turbidity terlihat dari plot histogram yang berbentuk simetris dapat dikatakan memiliki distribusi normal. Jika dilihat dari plot box plot, pencilan terlihat terdistribusi secara simetris. Sehingga, data Turbidity termasuk distribusi normal.

## Soal 4

**Melakukan test hipotesis 1 sampel, dengan menuliskan 6 langkah testing dan menampilkan juga boxplotnya untuk kolom/bagian yang bersesuaian.**

### a. Nilai Rata-rata pH di atas 7?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \mu = 7$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \mu > 7$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: display(Markdown(f'$t_{\alpha} > \{round(st.norm.ppf(1-0.05), 3)\} $'))
```

$$t_{\alpha} > 1.645$$

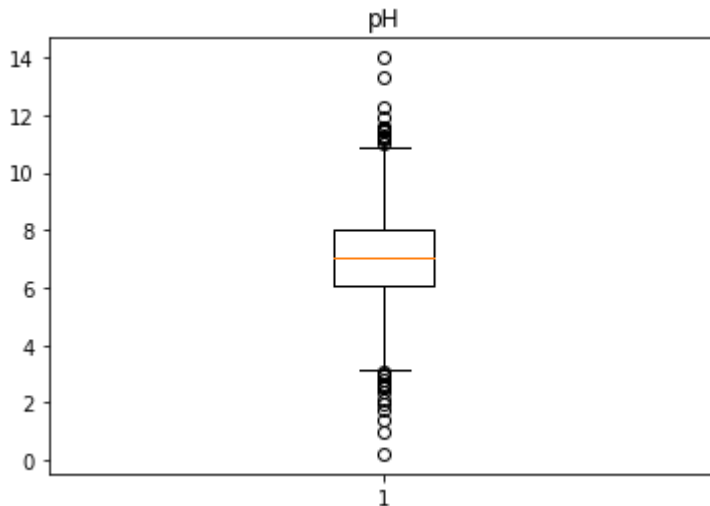
Tes statistik:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (1)$$

```
In [ ]: boxplot = plt.subplot(1,1,1)

plt.boxplot(df['pH'])
boxplot.set_title('pH')
```

```
Out[ ]: Text(0.5, 1.0, 'pH')
```



Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: t, p = st.ttest_1samp(df['pH'], popmean=7)
p = st.norm.sf(t)

display(Markdown(f'$t={round(t,3)}$'))
display(Markdown(f'$p={round(p,3)}$'))
```

$t = 2.485$

$p = 0.006$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if (p < 0.05):
    display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t={round(t,3)}>{round(st.t.ppf(1-p/2, df=df['pH'].count()-1)}$"))
    display(Markdown(f"**Tes signifikan:**\n\nTolak $H_0$ karena $p={round(p,3)}<\alpha=0.05$"))
    display(Markdown("**Jadi, pH rata-rata diatas 7**"))
else:
    display(Markdown(f"**Keputusan:**\n\nTidak menolak $H_0$ karena $t={round(t,3)}\leq{round(st.t.ppf(1-p/2, df=df['pH'].count()-1)}$"))
    display(Markdown(f"**Tes signifikan:**\n\nTidak menolak $H_0$ karena $p={round(p,3)}\geq\alpha=0.05$"))
    display(Markdown("**Jadi, pH rata-rata sama dengan 7**"))
```

**Keputusan:**

Tolak  $H_0$  karena  $t = 2.485 > 1.645$

**Tes signifikan:**

Tolak  $H_0$  karena  $p = 0.006 < \alpha = 0.05$

Jadi, pH rata-rata diatas 7

## b. Nilai Rata-rata Hardness tidak sama dengan 205?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \mu = 205$$

Pilih hipotesis alternatif  $H_1$  salah dari dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \mu \neq 205$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: crit = round(st.norm.ppf(1-0.05/2), 3)
display(Markdown("t<sub>\alpha/2</sub> < -%s dan t<sub>\alpha/2</sub> > %s " % (crit, crit)))
```

$$t_{\alpha/2} < -1.96 \text{ dan } t_{\alpha/2} > 1.96$$

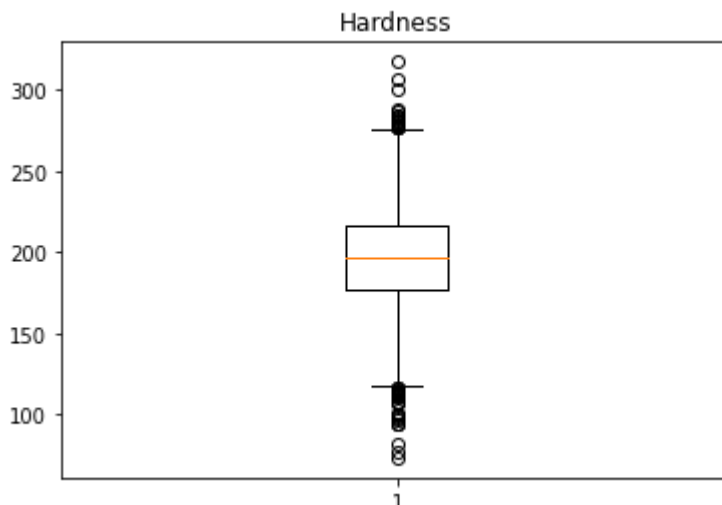
Tes statistik:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (2)$$

```
In [ ]: boxplot = plt.subplot(1,1,1)

plt.boxplot(df['Hardness'])
boxplot.set_title('Hardness')
```

```
Out[ ]: Text(0.5, 1.0, 'Hardness')
```



Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: t, p = st.ttest_1samp(df['Hardness'], popmean=205)

display(Markdown(f'$t={round(t,3)}$'))
display(Markdown(f'$p={round(p,3)}$'))
```

$$t = -12.403$$

$$p = 0.0$$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if (p < 0.05):
        if (t < -1 * st.norm.ppf(1-0.05/2)):
            display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t={round(t,3)} < -{round(1 * st.norm.ppf(1-0.05/2),3)}$"))
        else:
            display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t={round(t,3)} > {round(1 * st.norm.ppf(1-0.05/2),3)}$"))
            display(Markdown(f"**Tes signifikan:**\n\nTolak $H_0$ karena $p={round(p,3)} < \alpha=0.05$"))
            display(Markdown("**Jadi, Hardness rata-rata tidak sama dengan 205**"))
    else:
        display(Markdown(f"**Keputusan:**\n\nTidak menolak $H_0$ karena $t={round(t,3)}$ tidak berada di daerah kritis"))
        display(Markdown(f"**Tes signifikan:**\n\nTidak menolak $H_0$ karena $p={round(p,3)} \ge \alpha=0.05$"))
        display(Markdown("**Jadi, Hardness rata-rata sama dengan 205**"))
```

#### Keputusan:

Tolak  $H_0$  karena  $t = -12.403 < -1.96$

#### Tes signifikan:

Tolak  $H_0$  karena  $p = 0.0 < \alpha = 0.05$

**Jadi, Hardness rata-rata tidak sama dengan 205**

### c. Nilai Rata-rata 100 baris pertama kolom Solids bukan 21900?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \mu = 21900$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \mu \neq 21900$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: crit = round(st.norm.ppf(1-0.05/2), 3)
display(Markdown(f"$t_{\alpha/2} < -{crit}$ dan $t_{\alpha/2} > {crit}$"))
```

$$t_{\alpha/2} < -1.96 \text{ dan } t_{\alpha/2} > 1.96$$

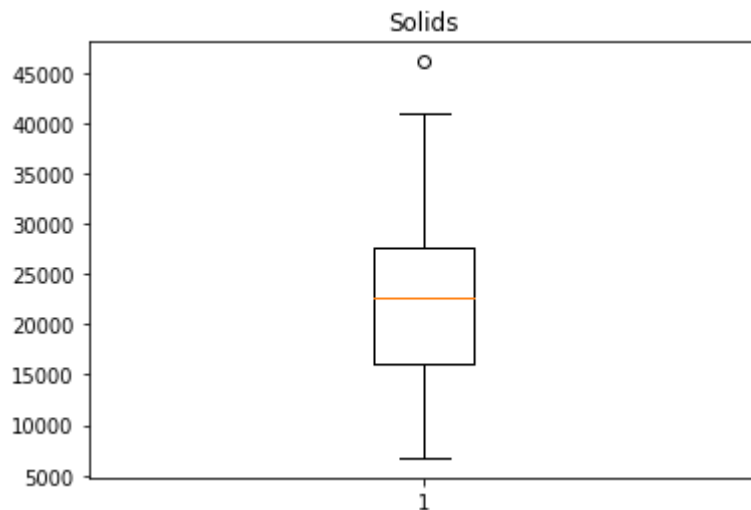
Tes statistik:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (3)$$

```
In [ ]: boxplot = plt.subplot(1,1,1)

plt.boxplot(df.iloc[:100]['Solids'])
boxplot.set_title('Solids')
```

```
Out[ ]: Text(0.5, 1.0, 'Solids')
```



Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: t, p = st.ttest_1samp(df['Solids'], popmean=21900)

display(Markdown(f'$t={round(t,3)}$'))
display(Markdown(f'$p={round(p,3)}$'))
```

$$t = 0.024$$

$$p = 0.981$$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if (p < 0.05):
    if (t < -1 * st.norm.ppf(1-0.05/2)):
        display(Markdown(f"***Keputusan:**\n\nTolak $H_0$ karena $t={round(t,3)}$ < -{round(1.96,3)}$"))
    else:
        display(Markdown(f"***Keputusan:**\n\nTolak $H_0$ karena $t={round(t,3)}$ > {round(1.96,3)}$"))
    display(Markdown(f"***Tes signifikan:**\n\nTolak $H_0$ karena $p={round(p,3)}$ < $\alpha=0.05$"))
    display(Markdown(f"***Jadi, 100 baris pertama Solids rata-rata tidak sama dengan 205**"))
else:
    display(Markdown(f"***Keputusan:**\n\nTidak menolak $H_0$ karena $t={round(t,3)}$ > {round(1.96,3)}$"))
    display(Markdown(f"***Tes signifikan:**\n\nTidak menolak $H_0$ karena $p={round(p,3)}$ > $\alpha=0.05$"))
    display(Markdown(f"***Jadi, 100 baris pertama Solids rata-rata sama dengan 205**"))
```



**Keputusan:**

Tidak menolak  $H_0$  karena  $t = 0.024$  berada pada  $-1.96 \leq t \leq 1.96$

**Tes signifikan:**

Tidak menolak  $H_0$  karena  $p = 0.981 \geq \alpha = 0.05$

**Jadi, 100 baris pertama Solids rata-rata sama dengan 205**

### d. Proporsi nilai Conductivity yang lebih dari 450, adalah tidak sama dengan 10%?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu, \sigma^2, p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : p = 0.1$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0, \theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : p \neq 0.1$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: crit = round(st.norm.ppf(1-0.05/2), 3)
display(Markdown("z<sub>\alpha/2</sub> < -%s dan z<sub>\alpha/2</sub> > %s" % (crit, crit)))
```

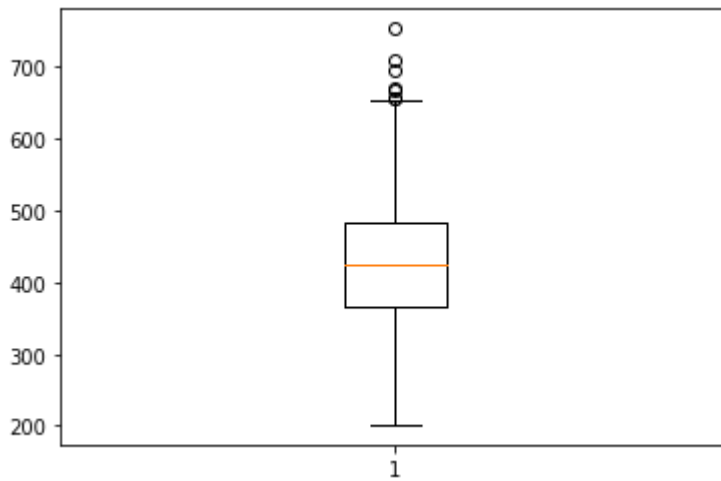
$$z_{\alpha/2} < -1.96 \text{ dan } z_{\alpha/2} > 1.96$$

Tes statistik:

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \quad (4)$$

```
In [ ]: plt.boxplot(df['Conductivity'])
boxplot.set_title('Conductivity')
```

```
Out[ ]: Text(0.5, 1.0, 'Conductivity')
```



Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: z, p = proportions_ztest(len(df[df["Conductivity"] > 450]), len(df), value=0.1, prop_v
display(Markdown(f'$z={round(z,3)}$'))
display(Markdown(f'$p={round(p,3)}$'))
```

$z = 40.446$

$p = 0.0$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if ((z < -1 * st.norm.ppf(1-0.05/2) or z > st.norm.ppf(1-0.05/2)) and p < 0.05):
        if (z < -1 * st.norm.ppf(1-0.05/2)):
            display(Markdown(f"***Keputusan:**\n\nTolak $H_0$ karena $z={round(z,3)}<{-round(
        else:
            display(Markdown(f"***Keputusan:**\n\nTolak $H_0$ karena $z={round(z,3)}>{round(
        display(Markdown(f"***Tes signifikan:**\n\nTolak $H_0$ karena $p={round(p,3)}<\alpha=0.05$
        display(Markdown("**Jadi, proporsi nilai Conductivity yang lebih dari 450 tidak sa
    else:
        display(Markdown(f"***Keputusan:**\n\nTidak menolak $H_0$ karena $z={round(z,3)}$ k
        display(Markdown(f"***Tes signifikan:**\n\nTidak menolak $H_0$ karena $p={round(p,3)}
        display(Markdown("**Jadi, proporsi nilai Conductivity yang lebih dari 450 sama der
```

**Keputusan:**

Tolak  $H_0$  karena  $z = 40.446 > 1.96$

**Tes signifikan:**

Tolak  $H_0$  karena  $p = 0.0 < \alpha = 0.05$

**Jadi, proporsi nilai Conductivity yang lebih dari 450 tidak sama dengan 10%**

**e. Proporsi nilai Trihalomethanes yang kurang dari 40, adalah kurang dari 5%?**

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : p = 0.05$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : p < 0.05$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: display(Markdown(f'$z_{\alpha} < -\{round(st.norm.ppf(1-0.05), 3)\}$'))
```

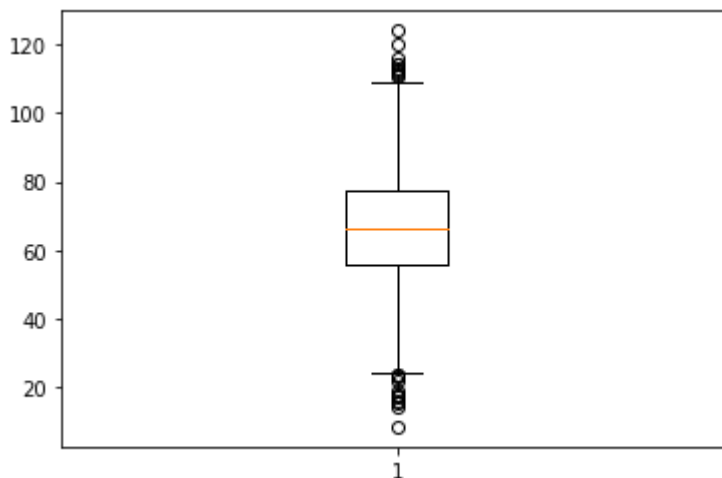
$$z_{\alpha} < -1.645$$

Tes statistik:

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \quad (5)$$

```
In [ ]: plt.boxplot(df['Trihalomethanes'])
        boxplot.set_title('Trihalomethanes')
```

```
Out[ ]: Text(0.5, 1.0, 'Trihalomethanes')
```



Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: z, p = proportions_ztest(len(df[df["Trihalomethanes"] < 40]), len(df), value=0.05, alt=
        display(Markdown(f'$z=\{round(z,3)\}$'))
        display(Markdown(f'$p=\{round(p,3)\}$'))
```

$$z = 0.549$$

$$p = 0.708$$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if ((z < -1 * st.norm.ppf(1-0.05/2)) or z > st.norm.ppf(1-0.05/2)) and p < 0.05:
        display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $z={round(z,3)} < -{round(st}
        display(Markdown(f"**Tes signifikan:**\n\nTolak $H_0$ karena $p={round(p,3)} < \alpha=0.05$
        display(Markdown("**Jadi, proporsi nilai Trihalomethanes yang kurang dari 40 kurang
    else:
        display(Markdown(f"**Keputusan:**\n\nTidak menolak $H_0$ karena $z={round(z,3)} \ge -{round(st}
        display(Markdown(f"**Tes signifikan:**\n\nTidak menolak $H_0$ karena $p={round(p,3)} \ge \alpha=0.05$
        display(Markdown("**Jadi, proporsi nilai Trihalomethanes yang kurang dari 40 adalah 5%"))
```

### Keputusan:

Tidak menolak  $H_0$  karena  $z = 0.549 \geq -1.645$

### Tes signifikan:

Tidak menolak  $H_0$  karena  $p = 0.708 \geq \alpha = 0.05$

**Jadi, proporsi nilai Trihalomethanes yang kurang dari 40 adalah 5%**

## Soal 5

**a. Data kolom Sulfate dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata kedua bagian tersebut sama?**

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \mu_1 - \mu_2 = 0$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: crit = st.norm.ppf(1-0.05/2)
        display(Markdown("t<sub>\alpha/2</sub> < -%s dan t<sub>\alpha/2</sub> > %s" % (crit, crit)))
```

$$t_{\alpha/2} < -1.959963984540054 \text{ dan } t_{\alpha/2} > 1.959963984540054$$

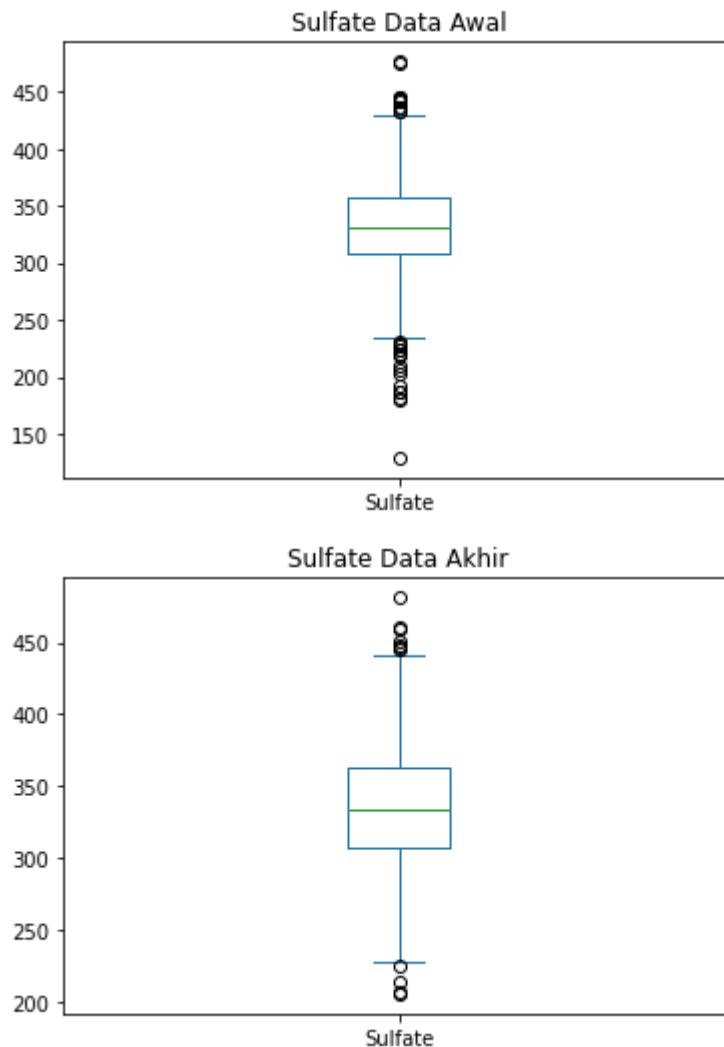
Tes Statistik

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p \sqrt{1/n_1 + 1/n_2}}$$

Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: df1 = df["Sulfate"][:1005]
df2 = df["Sulfate"][1005:]

df1.plot(kind="box").set_title("Sulfate Data Awal")
plt.show()
df2.plot(kind="box").set_title("Sulfate Data Akhir")
plt.show()
```



```
In [ ]: s1 = df1.var()
s2 = df2.var()
sp = ((1004 * s1 + 1004 * s2) / 2008)**0.5
x1 = df1.mean()
x2 = df2.mean()
t = (x1 - x2) / (sp*(1/1005 + 1/1005)**0.5)
p = 1 - st.t.cdf(t, 2008)

display(Markdown(f'$t={t}$'))
display(Markdown(f'$p={p}$'))
display(Markdown(f'Ambil keputusan dengan TOLAK $H_0$ jika nilai uji terletak di daerah
```

$$t = -2.0752690696871983$$

$$p = 0.9809556740463125$$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if (t < -1 * crit or t > crit and p < 0.05):
        if (t < -1 * crit):
            display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t = {t} < - {crit}$"))
        else:
            display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t = {t} > {crit}$"))
        display(Markdown(f"**Tes signifikan:**\n\nTolak $H_0$ karena $p = {p} < \alpha = 0.05$"))
        display(Markdown("**Jadi, rata-rata bagian awal dan akhir data Sulfate tidak sama"))
    else:
        display(Markdown(f"**Keputusan:**\n\nTidak menolak $H_0$ karena $t = {t}$ berada p"))
        display(Markdown(f"**Tes signifikan:**\n\nTidak menolak $H_0$ karena $p = {p} \ge \alpha$"))
        display(Markdown("**Jadi, rata-rata bagian awal dan akhir data Sulfate sama**"))
```

### Keputusan:

Tolak  $H_0$  karena  $t = -2.0752690696871983 < -1.959963984540054$

### Tes signifikan:

Tolak  $H_0$  karena  $p = 0.9809556740463125 < \alpha = 0.05$

**Jadi, rata-rata bagian awal dan akhir data Sulfate tidak sama**

## b. Data kolom OrganicCarbon dibagi 2 sama rata: bagian awal dan bagian akhir kolom. Benarkah rata-rata bagian awal lebih besar dari pada bagian akhir sebesar 0.15?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu, \sigma^2, p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \mu_1 - \mu_2 = 0.15$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0, \theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \mu_1 - \mu_2 \neq 0.15$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: crit = st.norm.ppf(1-0.05/2)
        display(Markdown("t<sub>\alpha/2</sub> < -%s dan t<sub>\alpha/2</sub> > %s" % (crit, crit)))
```

$$t_{\alpha/2} < -1.959963984540054 \text{ dan } t_{\alpha/2} > 1.959963984540054$$

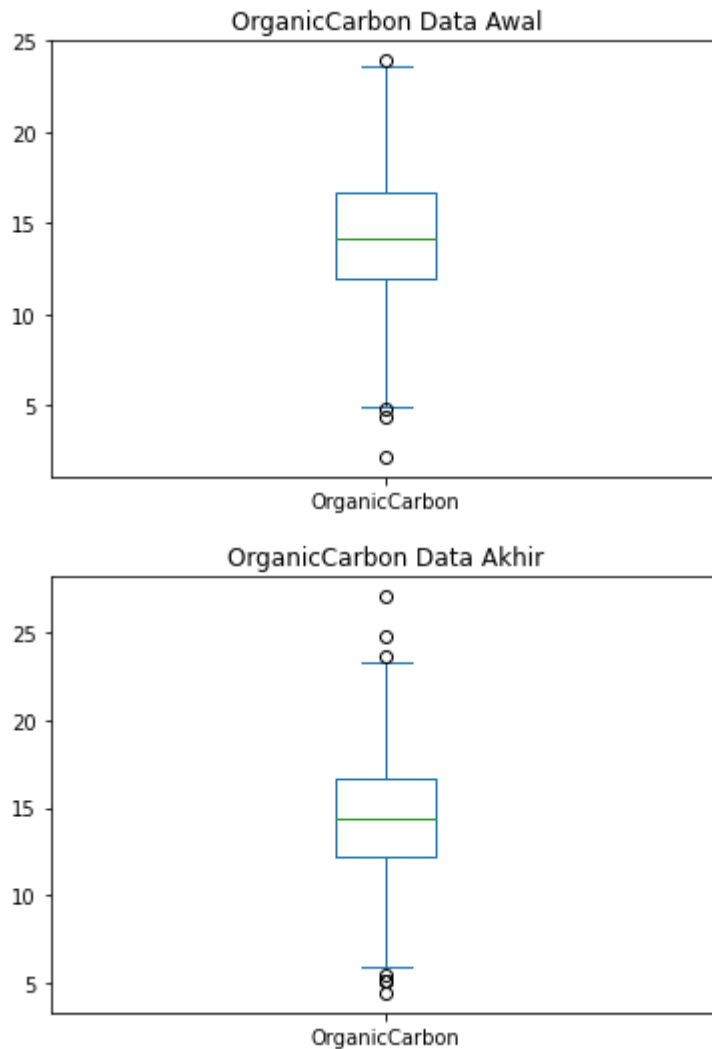
Tes Statistik

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p \sqrt{1/n_1 + 1/n_2}}$$

Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: df1 = df["OrganicCarbon"][:1005]
        df2 = df["OrganicCarbon"][1005:]
```

```
df1.plot(kind="box").set_title("OrganicCarbon Data Awal")
plt.show()
df2.plot(kind="box").set_title("OrganicCarbon Data Akhir")
plt.show()
```



```
In [ ]: s1 = df1.var()
s2 = df2.var()
sp = ((1004 * s1 + 1004 * s2) / 2008)**0.5
x1 = df1.mean()
x2 = df2.mean()
t = (x1 - x2 - 0.15) / (sp*(1/1005 + 1/1005)**0.5)
p = 1 - st.t.cdf(t, 2008)

display(Markdown(f'$t={t}$'))
display(Markdown(f'$p={p}$'))
display(Markdown(f'Ambil keputusan dengan TOLAK $H_0$ jika nilai uji terletak di daerah
```

$$t = -2.413145517798807$$

$$p = 0.9920477254406633$$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if (t < -1 * crit or t > crit and p < 0.05):
```

```

if (t < -1 * crit):
    display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t = {t} < - {crit}$"))
else:
    display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t = {t} > {crit}$"))
    display(Markdown(f"**Tes signifikan:**\n\nTolak $H_0$ karena $p = {p} < \alpha = 0.05$"))
    display(Markdown("**Jadi, rata-rata bagian awal data OrganicCarbon tidak lebih bes
else:
    display(Markdown(f"**Keputusan:**\n\nTidak menolak $H_0$ karena $t = {t}$ berada p
    display(Markdown(f"**Tes signifikan:**\n\nTidak menolak $H_0$ karena $p = {p} \ge \alpha$"))
    display(Markdown("**adi, rata-rata bagian awal data OrganicCarbon lebih besar sebe

```

### Keputusan:

Tolak  $H_0$  karena  $t = -2.413145517798807 < -1.959963984540054$

### Tes signifikan:

Tolak  $H_0$  karena  $p = 0.9920477254406633 < \alpha = 0.05$

**Jadi, rata-rata bagian awal data OrganicCarbon tidak lebih besar sebesar 0.15 dari bagian akhir datanya**

## c. Rata-rata 100 baris pertama kolom Chloramines sama dengan 100 baris terakhirnya?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \mu_1 - \mu_2 = 0$$

Pilih hipotesis alternatif  $H_1$  salah dari dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```

In [ ]: crit = st.norm.ppf(1-0.05/2)
display(Markdown("t<sub>\alpha/2</sub> < -%s dan t<sub>\alpha/2</sub> > %s" % (crit, crit)))

```

$$t_{\alpha/2} < -1.959963984540054 \text{ dan } t_{\alpha/2} > 1.959963984540054$$

Tes Statistik

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p \sqrt{1/n_1 + 1/n_2}}$$

Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

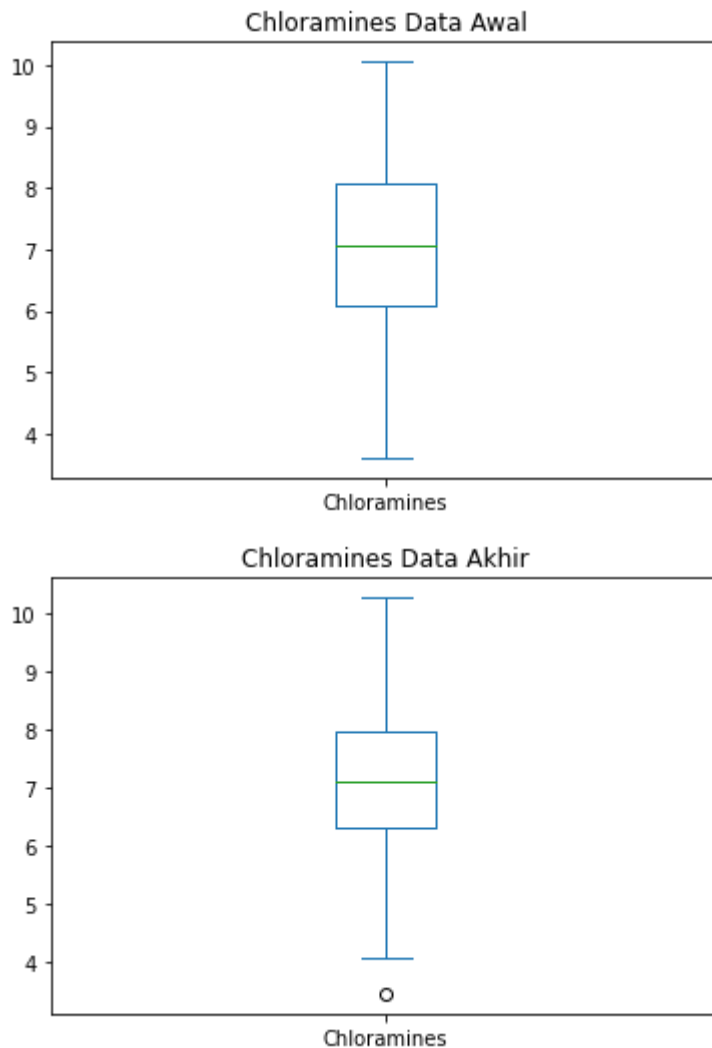
```

In [ ]: df1 = df["Chloramines"][:100]
df2 = df["Chloramines"][len(df)-100:]

```



```
df1.plot(kind="box").set_title("Chloramines Data Awal")
plt.show()
df2.plot(kind="box").set_title("Chloramines Data Akhir")
plt.show()
```



```
In [ ]: s1 = df1.var()
s2 = df2.var()
sp = ((99 * s1 + 99 * s2) / 198)**0.5
x1 = df1.mean()
x2 = df2.mean()
t = (x1 - x2) / (sp*(1/100 + 1/100)**0.5)
p = 1 - st.t.cdf(t, 98)

display(Markdown(f'$t={t}$'))
display(Markdown(f'$p={p}$'))
display(Markdown(f'Ambil keputusan dengan TOLAK $H_0$ jika nilai uji terletak di daerah
```

$$t = -0.7059424842236872$$

$$p = 0.7590506311645496$$

Ambil keputusan dengan TOLAK  $H_0$  jika nilai uji terletak di daerah kritis atau dengan tes signifikan, TOLAK  $H_0$  jika p-value lebih kecil dibanding tingkat signifikansi  $\alpha$  yang diinginkan.

```
In [ ]: if (t < -1 * crit or t > crit and p < 0.05):
        if (t < -1 * crit):
```

```

display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t = \{t\} < - \{crit\}$"))
else:
    display(Markdown(f"**Keputusan:**\n\nTolak $H_0$ karena $t = \{t\} > \{crit\}$"))
display(Markdown(f"**Tes signifikan:**\n\nTolak $H_0$ karena $p = \{p\} < \alpha = 0.05$"))
display(Markdown(f"**Jadi, rata-rata 100 data awal Chloramines tidak sama dengan rata-rata 100 data akhir Chloramines"))
else:
    display(Markdown(f"**Keputusan:**\n\nTidak menolak $H_0$ karena $t = \{t\}$ berada pada $-\{crit\} \le t \le \{crit\}$"))
display(Markdown(f"**Tes signifikan:**\n\nTidak menolak $H_0$ karena $p = \{p\} \ge \alpha = 0.05$"))
display(Markdown(f"**Jadi, rata-rata 100 data awal Chloramines sama dengan rata-rata 100 data akhir Chloramines"))

```

### Keputusan:

Tidak menolak  $H_0$  karena  $t = -0.7059424842236872$  berada pada  $-1.959963984540054 \leq t \leq 1.959963984540054$

### Tes signifikan:

Tidak menolak  $H_0$  karena  $p = 0.7590506311645496 \geq \alpha = 0.05$

**Jadi, rata-rata 100 data awal Chloramines sama dengan rata-rata 100 data akhir Chloramines**

## d. Proporsi nilai bagian awal Turbidity yang lebih dari 4, adalah lebih besar daripada, proporsi nilai yang sama di bagian akhir Turbidity ?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu, \sigma^2, p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : p_1 > p_2$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0, \theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : p_1 \leq p_2$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: display(Markdown(f'$z_\alpha > \{crit\}$'))
```

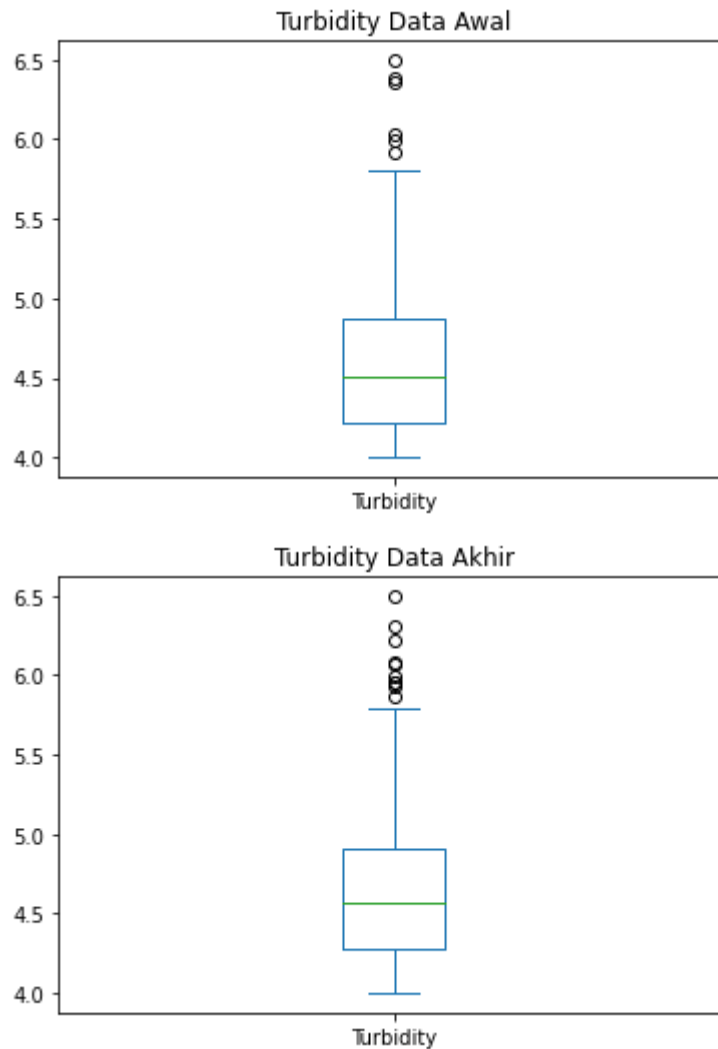
$$z_\alpha > 1.959963984540054$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}$$

Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: df1 = df["Turbidity"][:1005]
df1 = df1[df1 > 4]
df2 = df["Turbidity"][1005:]
df2 = df2[df2 > 4]
```

```
df1.plot(kind="box").set_title("Turbidity Data Awal")
plt.show()
df2.plot(kind="box").set_title("Turbidity Data Akhir")
plt.show()
```



```
In [ ]: crit = st.norm.ppf(1-0.05)

x1 = len(df1)
x2 = len(df2)
p1 = x1/1005
p2 = x2/1005

p = (x1 + x2)/(2010)
q = 1 - p

z = (p1 - p2)/((p*q*(2/1005))**0.5)

if z < 0:
    pValue = (st.norm.cdf(z))
else:
    pValue = (st.norm.sf(z))

display(Markdown(f'$z={z}$'))
display(Markdown(f'$p={pValue}$'))
```

$$z = -0.13388958661778735$$

$$p = 0.4467449424088169$$

```
In [ ]: if (z > crit and pValue < 0.05):
        display(Markdown(f"***Keputusan:**\n\nTolak $H_0$ karena $z = {z} > {crit}$"))
        display(Markdown(f"***Tes signifikan:**\n\nTolak $H_0$ karena $p={pValue} < \alpha = 0.05$"))
        display(Markdown("**Jadi, proporsi nilai bagian awal Turbidity yang lebih dari 4,
    else:
        display(Markdown(f"***Keputusan:**\n\nTidak menolak $H_0$ karena $z = {z}$ berada pada
        display(Markdown(f"***Tes signifikan:**\n\nTidak menolak $H_0$ karena $p={pValue} \ge \alpha = 0.05$"))
        display(Markdown("**Jadi, proporsi nilai bagian awal Turbidity yang lebih dari 4,

```

### Keputusan:

Tidak menolak  $H_0$  karena  $z = -0.13388958661778735$  berada pada  $z \leq 1.6448536269514722$

### Tes signifikan:

Tidak menolak  $H_0$  karena  $p = 0.4467449424088169 \geq \alpha = 0.05$

**Jadi, proporsi nilai bagian awal Turbidity yang lebih dari 4, lebih besar daripada, proporsi nilai yang sama di bagian akhir Turbidity**

## e. Bagian awal kolom Sulfate memiliki variansi yang sama dengan bagian akhirnya?

Tentukan Hipotesis nol ( $H_0 : \theta = \theta_0$ ), dimana  $\theta$  bisa berupa  $\mu$ ,  $\sigma^2$ ,  $p$ , atau data lain berdistribusi tertentu (normal, binomial, dsc.).

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Pilih hipotesis alternatif  $H_1$  salah dari  $\theta > \theta_0$ ,  $\theta < \theta_0$ , atau  $\theta \neq \theta_0$ .

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Tentukan tingkat signifikan  $\alpha$ .

$$\alpha = 0.05$$

Tentukan uji statistik yang sesuai dan tentukan daerah kritis.

Daerah kritis:

```
In [ ]: display(Markdown(f" < f<sub>(1-α/2)</sub> (v1,v2) atau f > f<sub>(α/2)</sub> (v1,v2)"))
```

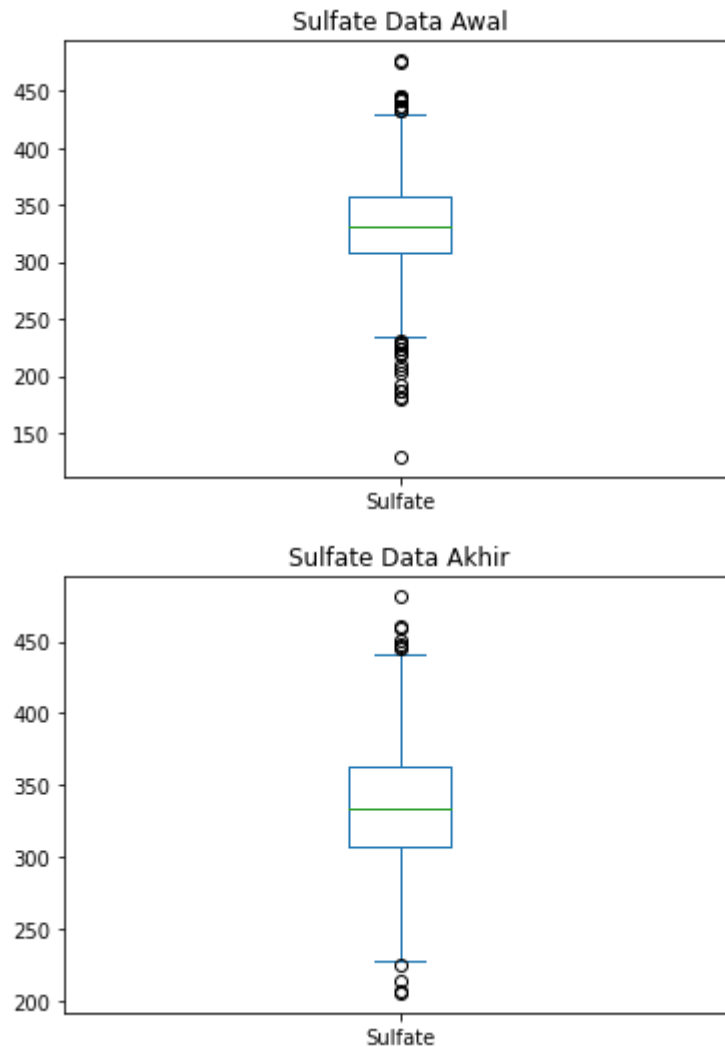
$$f < f_{(1-\alpha/2)}(v_1, v_2) \text{ atau } f > f_{(\alpha/2)}(v_1, v_2)$$

$$f = \frac{s_1^2}{s_2^2}$$

Hitung nilai uji statistik dari data sample. Hitung p-value sesuai dengan uji statistik yang digunakan.

```
In [ ]: df1 = df["Sulfate"][:1005]
        df2 = df["Sulfate"][1005:]
```

```
df1.plot(kind="box").set_title("Sulfate Data Awal")
plt.show()
df2.plot(kind="box").set_title("Sulfate Data Akhir")
plt.show()
```



```
In [ ]: s1 = df1.var()
s2 = df2.var()

f = s1**2/s2**2

critRight = st.f.ppf(1-0.05/2, 1004, 1004)
critLeft = st.f.ppf(0.05/2, 1004, 1004)

display(Markdown(f'$f={f}$'))
display(Markdown(f'$critLeft={critLeft}$'))
display(Markdown(f'$critRight={critRight}$'))
```

$f = 1.0307348049752798$

$critLeft = 0.883572344355818$

$critRight = 1.1317692392568777$

```
In [ ]: if (f < critLeft or f > critRight):
    if (f < critLeft):
```

```

display(Markdown(f"**Keputusan:**\n\nTolak  $H_0$  karena  $f = \{f\} < \{critLeft\}$ 
else:
    display(Markdown(f"**Keputusan:**\n\nTolak  $H_0$  karena  $f = \{f\} > \{critRight\}$ 
display(Markdown("**Jadi, bagian awal kolom Sulfate tidak memiliki variansi yang s
else:
    display(Markdown(f"**Keputusan:**\n\nTidak menolak  $H_0$  karena  $f = \{critLeft\} \leq$ 
display(Markdown("**Jadi, bagian awal kolom Sulfate memiliki variansi yang sama de

```

### Keputusan:

Tidak menolak  $H_0$  karena

$$f = 0.883572344355818 \leq 1.0307348049752798 \leq 1.1317692392568777$$

**Jadi, bagian awal kolom Sulfate memiliki variansi yang sama dengan bagian akhirnya**

## Soal 6

```

In [ ]: def korelasi(data):
        x = df[data]

        # Mendapatkan nilai korelasi
        kor = x.corr(other=df["Potability"])

        print("Nilai Koefisien Korelasi antara %s dan Potability: %s" % (data, kor))
        print("%s dan Potability " % data, end="")

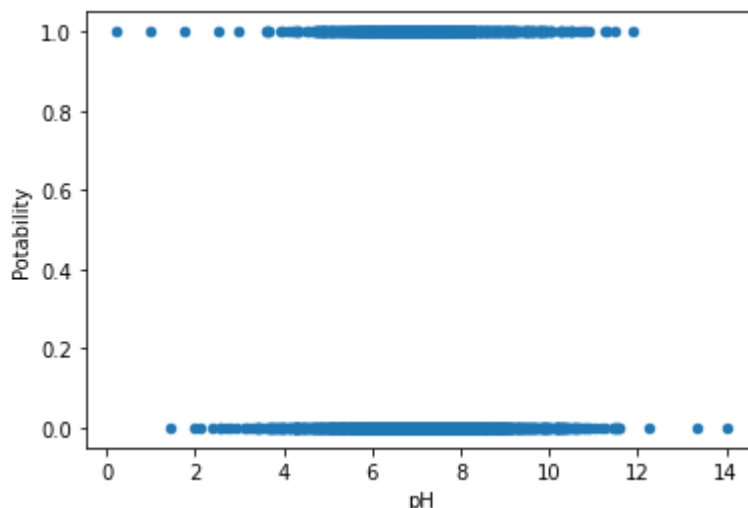
        if (kor > 0.25):
            print("memiliki korelasi positif")
        else:
            print("tidak memiliki korelasi")

        df.plot(kind="scatter", x=data, y="Potability").set_xlabel(data)

```

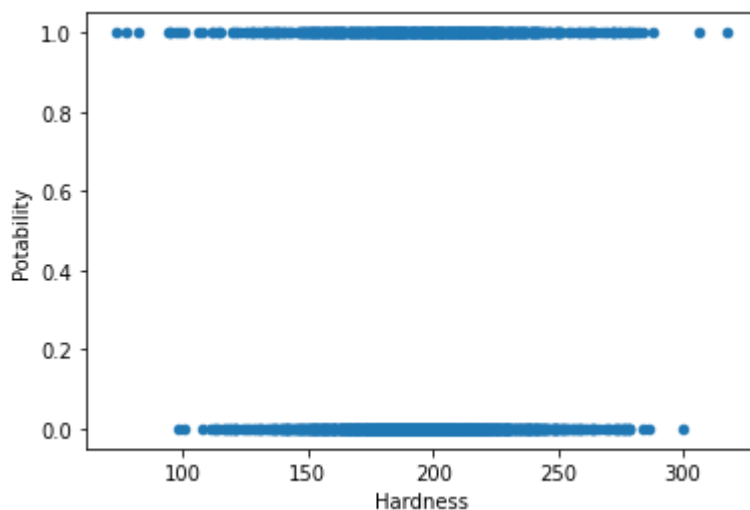
```
In [ ]: korelasi("pH")
```

Nilai Koefisien Korelasi antara pH dan Potability: 0.01547509440843348  
pH dan Potability tidak memiliki korelasi



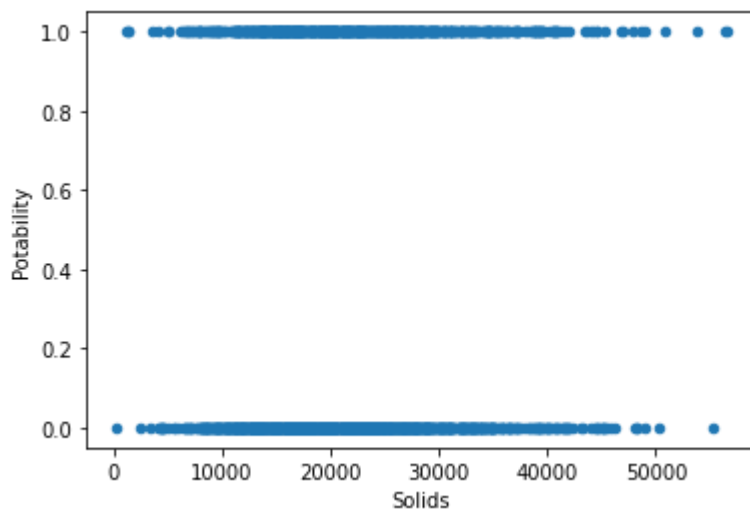
```
In [ ]: korelasi("Hardness")
```

Nilai Koefisien Korelasi antara Hardness dan Potability: -0.0014631528959479344  
Hardness dan Potability tidak memiliki korelasi



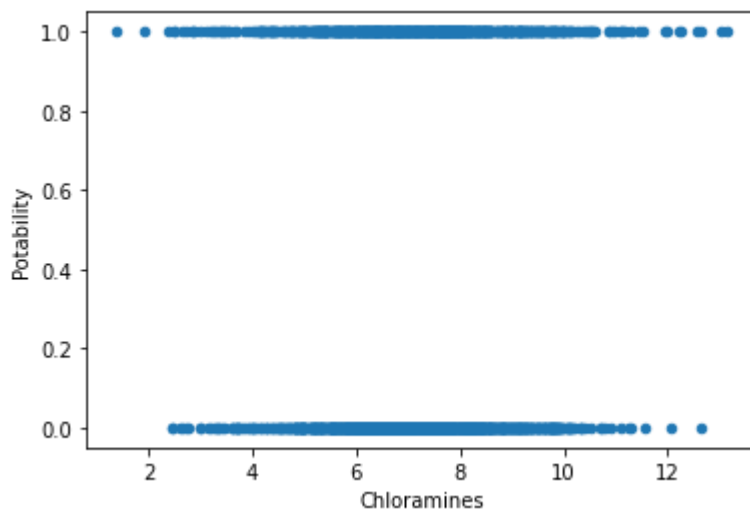
```
In [ ]: korelasi("Solids")
```

Nilai Koefisien Korelasi antara Solids dan Potability: 0.03897657818173466  
Solids dan Potability tidak memiliki korelasi



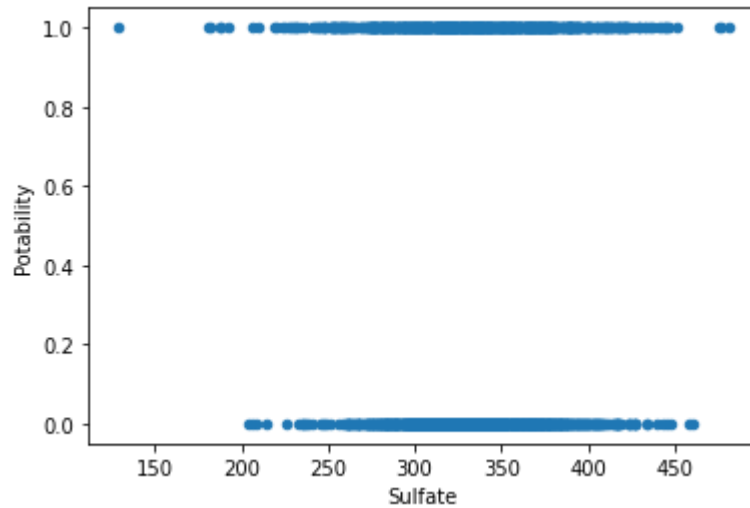
```
In [ ]: korelasi("Chloramines")
```

Nilai Koefisien Korelasi antara Chloramines dan Potability: 0.02077892184052409  
Chloramines dan Potability tidak memiliki korelasi



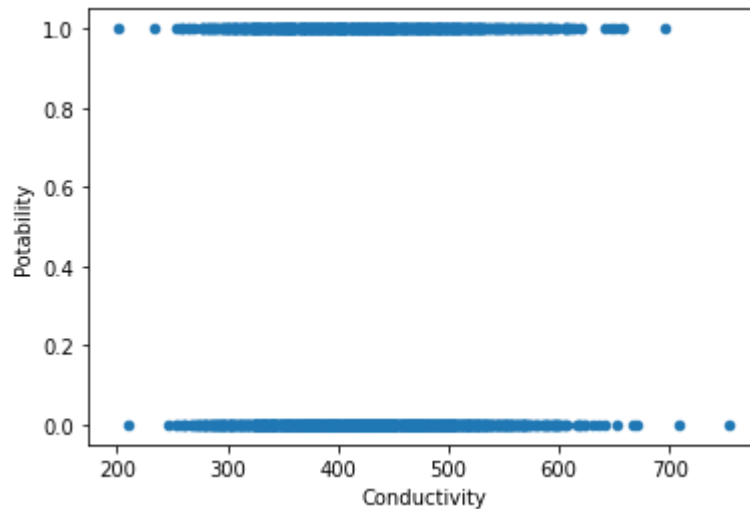
```
In [ ]: korelasi("Sulfate")
```

Nilai Koefisien Korelasi antara Sulfate dan Potability:  $-0.015703164419273778$   
Sulfate dan Potability tidak memiliki korelasi



```
In [ ]: korelasi("Conductivity")
```

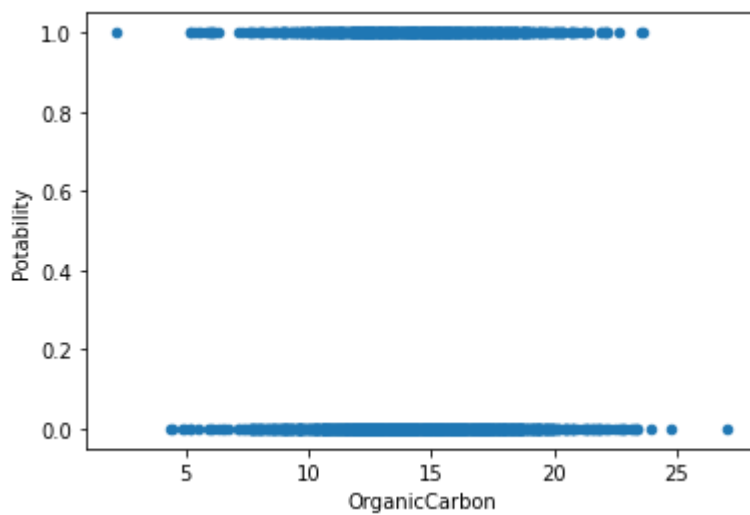
Nilai Koefisien Korelasi antara Conductivity dan Potability:  $-0.016257120111377067$   
Conductivity dan Potability tidak memiliki korelasi



```
In [ ]: korelasi("OrganicCarbon")
```

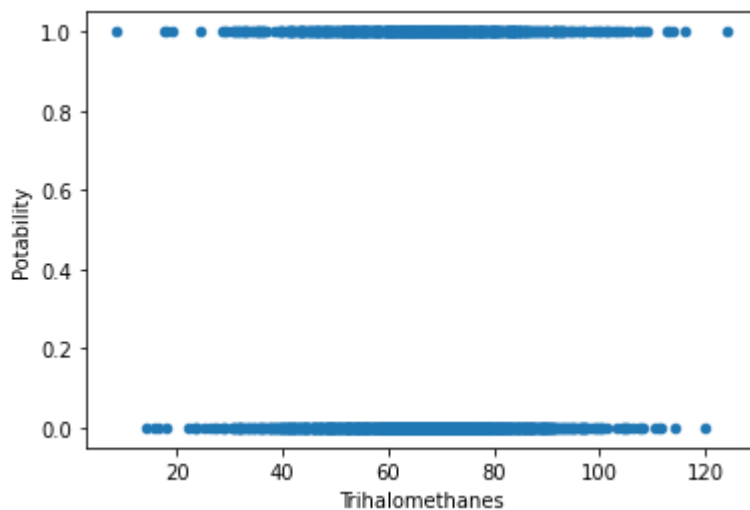
Nilai Koefisien Korelasi antara OrganicCarbon dan Potability:  $-0.015488461910747259$   
OrganicCarbon dan Potability tidak memiliki korelasi





```
In [ ]: korelasi("Trihalomethanes")
```

Nilai Koefisien Korelasi antara Trihalomethanes dan Potability: 0.009236711064712997  
Trihalomethanes dan Potability tidak memiliki korelasi



```
In [ ]: korelasi("Turbidity")
```

Nilai Koefisien Korelasi antara Turbidity dan Potability: 0.022331042640622665  
Turbidity dan Potability tidak memiliki korelasi

