DESKRIPSI STATISTIKA

Deskripsi: mean, median, modus, standar deviasi, variansi, range, nilai minimum, maksimum, kuartil, IQR, skewness dan kurtosis.

In [ ]:
```python
# Import Library Pandas
import pandas as pd

# Read csv file
df = pd.read_csv("../data/anggur.csv")

# Print descriptive statistics function
def print_descriptive_statistics(dataframe):
    # Mean
    print("Mean:", dataframe.mean())
    print("----------------------------------------")

    # Median
    print("Median:", dataframe.median())
    print("----------------------------------------")

    # Modus
    print("Modus:")
    all_modes = dataframe.mode().values.tolist()
    if (len(all_modes) == dataframe.count()):
        print("Ada", dataframe.count(), "modus pada kolom ini. Jumlah tersebut sama dengan jumlah nilai pada kolom ini.")
        print("Hal ini menandakan kolom ini memiliki nilai-nilai yang berbeda satu sama lain.")
    else:
        for mode in all_modes:
            print(mode)
    print("----------------------------------------")

    # Standar Deviasi
    print("Standar Deviasi:", dataframe.std())
    print("----------------------------------------")

    # Variansi
    print("Variansi:", dataframe.var())
    print("----------------------------------------")

    # Range
    print("Range:", dataframe.max() - dataframe.min())
    print("----------------------------------------")

    # Minimum
    print("Nilai Minimum:", dataframe.min())
    print("----------------------------------------")

    # Maximum
    print("Nilai Maksimum:", dataframe.max())
    print("----------------------------------------")

    # Kuartil
    print("Kuartil Bawah:", dataframe.quantile(0.25))
    print("Kuartil Tengah:", dataframe.quantile(0.50))
    print("Kuartil Atas:", dataframe.quantile(0.75))
    print("----------------------------------------")

    # IQR
    print("IQR:", dataframe.quantile(0.75) - dataframe.quantile(0.25))
    print("----------------------------------------")

    # Skewness
    print("Skewness:", dataframe.skew())
    print("----------------------------------------")

    # Kurtosis
    print("Kurtosis:", dataframe.kurtosis())
    print("========================================")
```

In [ ]:
```python
display(df)
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	5.90	0.4451	0.1813	2.049401	0.070574	16.593818	42.27	0.9982	3.27	0.71	8.64	7
1	8.40	0.5768	0.2099	3.109590	0.101681	22.555519	16.01	0.9960	3.35	0.57	10.03	8
2	7.54	0.5918	0.3248	3.673744	0.072416	9.316866	35.52	0.9990	3.31	0.64	9.23	8
3	5.39	0.4201	0.3131	3.371815	0.072755	18.212300	41.97	0.9945	3.34	0.55	14.07	9
4	6.51	0.5675	0.1940	4.404723	0.066379	9.360591	46.27	0.9925	3.27	0.45	11.49	8
...	...	...	...	...	...	...	...	...	...	...	...	...
995	7.96	0.6046	0.2662	1.592048	0.057555	14.892445	44.61	0.9975	3.35	0.54	10.41	8
996	8.48	0.4080	0.2227	0.681955	0.051627	23.548965	25.83	0.9972	3.41	0.46	9.91	8
997	6.11	0.4841	0.3720	2.377267	0.042806	21.624585	48.75	0.9928	3.23	0.55	9.94	7
998	7.76	0.3590	0.3208	4.294486	0.098276	12.746186	44.53	0.9952	3.30	0.66	9.76	8
999	5.87	0.5214	0.1883	2.179490	0.052923	16.203864	24.37	0.9983	3.29	0.70	10.17	7

1000 rows × 12 columns

Kolom Fixed Acidity

In [ ]:
```
# Fixed Acidity
print("====================o====================")
print("Deskripsi Statistika Kolom Fixed Acidity")
print("====================o====================")

df_fixed_acidity = df["fixed acidity"]

# Print Descriptive Statistics
print_descriptive_statistics(df_fixed_acidity)
```

```
====================o====================
Deskripsi Statistika Kolom Fixed Acidity
====================o====================
Mean: 7.152530000000006
----------------------------------------
Median: 7.15
----------------------------------------
Modus:
6.54
----------------------------------------
Standar Deviasi: 1.2015975764938276
----------------------------------------
Variansi: 1.4438367358358397
----------------------------------------
Range: 8.17
----------------------------------------
Nilai Minimum: 3.32
----------------------------------------
Nilai Maksimum: 11.49
----------------------------------------
Kuartil Bawah: 6.3774999999999995
Kuartil Tengah: 7.15
Kuartil Atas: 8.0
----------------------------------------
IQR: 1.6225000000000005
----------------------------------------
Skewness: -0.028878575532660055
----------------------------------------
Kurtosis: -0.019292120932933532
========================================
```

Kolom Volatile Acidity

In [ ]:
```
# Volatile Acidity
print("====================o====================")
print("Deskripsi Statistika Kolom Volatile Acidity")
print("====================o====================")

df_volatile_acidity = df["volatile acidity"]

# Print Descriptive Statistics
print_descriptive_statistics(df_volatile_acidity)
```

```
====================o====================
Deskripsi Statistika Kolom Volatile Acidity
====================o====================
Mean: 0.5208384999999999
----------------------------------------
Median: 0.52485
----------------------------------------
Modus:
0.5546
----------------------------------------
Standar Deviasi: 0.09584827405534954
----------------------------------------
Variansi: 0.009186891639389393
----------------------------------------
Range: 0.6652
----------------------------------------
Nilai Minimum: 0.1399
----------------------------------------
Nilai Maksimum: 0.8051
----------------------------------------
Kuartil Bawah: 0.4561
Kuartil Tengah: 0.52485
Kuartil Atas: 0.585375
----------------------------------------
IQR: 0.12927499999999997
----------------------------------------
Skewness: -0.1976986986092083
----------------------------------------
Kurtosis: 0.16185290336961788
========================================
```

Kolom Citric Acid

In [ ]:
```
# Citric Acid
print("====================o====================")
print("Deskripsi Statistika Kolom Citric Acid")
print("====================o====================")

df_citric_acid = df["citric acid"]
```

```
# Print Descriptive Statistics
print_descriptive_statistics(df_citric_acid)
```

```
====================o====================
Deskripsi Statistika Kolom Citric Acid
====================o====================
Mean: 0.27051699999999995
----------------------------------------
Median: 0.2722
----------------------------------------
Modus:
0.3019
----------------------------------------
Standar Deviasi: 0.04909837147076352
----------------------------------------
Variansi: 0.0024106500810810853
----------------------------------------
Range: 0.29290000000000005
----------------------------------------
Nilai Minimum: 0.1167
----------------------------------------
Nilai Maksimum: 0.4096
----------------------------------------
Kuartil Bawah: 0.2378
Kuartil Tengah: 0.2722
Kuartil Atas: 0.302325
----------------------------------------
IQR: 0.064525
----------------------------------------
Skewness: -0.045576058685017296
----------------------------------------
Kurtosis: -0.1046792495951605
========================================
```

Kolom Residual Sugar

In [ ]:
```
# Residual Sugar
print("====================o====================")
print("Deskripsi Statistika Kolom Residual Sugar")
print("====================o====================")

df_residual_sugar = df["residual sugar"]

# Print Descriptive Statistics
print_descriptive_statistics(df_residual_sugar)
```

```
====================o====================
Deskripsi Statistika Kolom Residual Sugar
====================o====================
Mean: 2.5671036825067572
----------------------------------------
Median: 2.519430272865794
----------------------------------------
Modus:
Ada 1000 modus pada kolom ini. Jumlah tersebut sama dengan jumlah nilai pada kolom ini.
Hal ini menandakan kolom ini memiliki nilai-nilai yang berbeda satu sama lain.
----------------------------------------
Standar Deviasi: 0.9879154365046932
----------------------------------------
Variansi: 0.9759769096842584
----------------------------------------
Range: 5.5182004097078625
----------------------------------------
Nilai Minimum: 0.032554525015195
----------------------------------------
Nilai Maksimum: 5.550754934723058
----------------------------------------
Kuartil Bawah: 1.896329943488683
Kuartil Tengah: 2.519430272865794
Kuartil Atas: 3.220873482829786
----------------------------------------
IQR: 1.3245435393411031
----------------------------------------
Skewness: 0.13263808618992312
----------------------------------------
Kurtosis: -0.04298003436476261
========================================
```

Kolom Chlorides

In [ ]:
```
# Chlorides
print("====================o====================")
print("Deskripsi Statistika Kolom Chlorides")
print("====================o====================")

df_chlorides = df["chlorides"]

# Print Descriptive Statistics
print_descriptive_statistics(df_chlorides)
```

```
====================o====================
Deskripsi Statistika Kolom Chlorides
====================o====================
Mean: 0.08119515250784973
----------------------------------------
Median: 0.0821669021645236
----------------------------------------
Modus:
Ada 1000 modus pada kolom ini. Jumlah tersebut sama dengan jumlah nilai pada kolom ini.
Hal ini menandakan kolom ini memiliki nilai-nilai yang berbeda satu sama lain.
----------------------------------------
Standar Deviasi: 0.020110647243996742
----------------------------------------
Variansi: 0.0004044381325724738
----------------------------------------
Range: 0.1256351302653488
----------------------------------------
Nilai Minimum: 0.0151224391657095
----------------------------------------
Nilai Maksimum: 0.1407575694310583
----------------------------------------
Kuartil Bawah: 0.06657363190977357
Kuartil Tengah: 0.0821669021645236
Kuartil Atas: 0.09531150148556258
----------------------------------------
IQR: 0.028737869575789013
----------------------------------------
Skewness: -0.05131929742072573
----------------------------------------
Kurtosis: -0.2465081359240382
========================================
```

Kolom Free Sulfur Dioxide

In [ ]:
```python
# Free Sulfur Dioxide
print("====================o====================")
print("Deskripsi Statistika Kolom Free Sulfur Dioxide")
print("====================o====================")

df_free_sulfur_dioxide = df["free sulfur dioxide"]

# Print Descriptive Statistics
print_descriptive_statistics(df_free_sulfur_dioxide)
```

```
====================o====================
Deskripsi Statistika Kolom Free Sulfur Dioxide
====================o====================
Mean: 14.907679251029792
----------------------------------------
Median: 14.860346236568924
----------------------------------------
Modus:
Ada 1000 modus pada kolom ini. Jumlah tersebut sama dengan jumlah nilai pada kolom ini.
Hal ini menandakan kolom ini memiliki nilai-nilai yang berbeda satu sama lain.
----------------------------------------
Standar Deviasi: 4.888099705756564
----------------------------------------
Variansi: 23.89351873341741
----------------------------------------
Range: 27.26784690109891
----------------------------------------
Nilai Minimum: 0.194678523326937
----------------------------------------
Nilai Maksimum: 27.462525424425845
----------------------------------------
Kuartil Bawah: 11.426716949457617
Kuartil Tengah: 14.860346236568924
Kuartil Atas: 18.313097915395005
----------------------------------------
IQR: 6.886380965937388
----------------------------------------
Skewness: 0.007130415991143398
----------------------------------------
Kurtosis: -0.36496364342685306
========================================
```

Kolom Total Sulfur Dioxide

In [ ]:
```python
# Total Sulfur Dioxide
print("====================o====================")
print("Deskripsi Statistika Kolom Total Sulfur Dioxide")
print("====================o====================")

df_total_sulfur_dioxide = df["total sulfur dioxide"]

# Print Descriptive Statistics
print_descriptive_statistics(df_total_sulfur_dioxide)
```

```
====================o====================
Deskripsi Statistika Kolom Total Sulfur Dioxide
====================o====================
Mean: 40.290150000000075
----------------------------------------
Median: 40.19
----------------------------------------
Modus:
35.2
37.25
39.64
40.61
41.05
41.59
44.51
----------------------------------------
Standar Deviasi: 9.965767376218295
----------------------------------------
Variansi: 99.3165193968969
----------------------------------------
Range: 66.80999999999999
----------------------------------------
Nilai Minimum: 3.15
----------------------------------------
Nilai Maksimum: 69.96
----------------------------------------
Kuartil Bawah: 33.785
Kuartil Tengah: 40.19
Kuartil Atas: 47.0225
----------------------------------------
IQR: 13.237500000000004
----------------------------------------
Skewness: -0.024060026812269975
----------------------------------------
Kurtosis: 0.06394978916172311
========================================
```

Kolom Density

In [ ]:
```python
# Density
print("====================o====================")
print("Deskripsi Statistika Kolom Density")
print("====================o====================")

df_density = df["density"]

# Print Descriptive Statistics
print_descriptive_statistics(df_density)
```

```
====================o====================
Deskripsi Statistika Kolom Density
====================o====================
Mean: 0.9959253000000002
----------------------------------------
Median: 0.996
----------------------------------------
Modus:
0.9959
0.9961
0.9965
0.997
----------------------------------------
Standar Deviasi: 0.0020201809426487133
----------------------------------------
Variansi: 4.081131041041044e-06
----------------------------------------
Range: 0.013799999999999923
----------------------------------------
Nilai Minimum: 0.9888
----------------------------------------
Nilai Maksimum: 1.0026
----------------------------------------
Kuartil Bawah: 0.9946
Kuartil Tengah: 0.996
Kuartil Atas: 0.9972
----------------------------------------
IQR: 0.0025999999999999357
----------------------------------------
Skewness: -0.07688278915513917
----------------------------------------
Kurtosis: 0.01636562128503849
========================================
```

Kolom pH

In [ ]:
```python
# pH
print("====================o====================")
print("Deskripsi Statistika Kolom pH")
print("====================o====================")

df_pH = df["pH"]

# Print Descriptive Statistics
print_descriptive_statistics(df_pH)
```

```
====================o====================
Deskripsi Statistika Kolom pH
====================o====================
Mean: 3.303610000000003
----------------------------------------
Median: 3.3
----------------------------------------
Modus:
3.34
----------------------------------------
Standar Deviasi: 0.10487548220040155
----------------------------------------
Variansi: 0.010998866766766742
----------------------------------------
Range: 0.7399999999999998
----------------------------------------
Nilai Minimum: 2.97
----------------------------------------
Nilai Maksimum: 3.71
----------------------------------------
Kuartil Bawah: 3.23
Kuartil Tengah: 3.3
Kuartil Atas: 3.37
----------------------------------------
IQR: 0.14000000000000012
----------------------------------------
Skewness: 0.14767259510827038
----------------------------------------
Kurtosis: 0.0809095518741838
========================================
```

Kolom Sulphates

In [ ]:
```python
# Sulphates
print("====================o====================")
print("Deskripsi Statistika Kolom Sulphates")
print("====================o====================")

df_sulphates = df["sulphates"]

# Print Descriptive Statistics
print_descriptive_statistics(df_sulphates)
```

```
====================o====================
Deskripsi Statistika Kolom Sulphates
====================o====================
Mean: 0.5983899999999999
----------------------------------------
Median: 0.595
----------------------------------------
Modus:
0.59
----------------------------------------
Standar Deviasi: 0.10081900799141184
----------------------------------------
Variansi: 0.010164472372372365
----------------------------------------
Range: 0.6699999999999999
----------------------------------------
Nilai Minimum: 0.29
----------------------------------------
Nilai Maksimum: 0.96
----------------------------------------
Kuartil Bawah: 0.53
Kuartil Tengah: 0.595
Kuartil Atas: 0.67
----------------------------------------
IQR: 0.14
----------------------------------------
Skewness: 0.1491989008699043
----------------------------------------
Kurtosis: 0.06481928180859686
========================================
```

Kolom Alcohol

In [ ]:
```python
# Alcohol
print("====================o====================")
print("Deskripsi Statistika Kolom Alcohol")
print("====================o====================")

df_alcohol = df["alcohol"]

# Print Descriptive Statistics
print_descriptive_statistics(df_alcohol)
```

```
====================o====================
Deskripsi Statistika Kolom Alcohol
====================o====================
Mean: 10.592279999999985
----------------------------------------
Median: 10.61
----------------------------------------
Modus:
9.86
10.31
----------------------------------------
Standar Deviasi: 1.5107060052287598
----------------------------------------
Variansi: 2.282232634234237
----------------------------------------
Range: 8.989999999999998
----------------------------------------
Nilai Minimum: 6.03
----------------------------------------
Nilai Maksimum: 15.02
----------------------------------------
Kuartil Bawah: 9.56
Kuartil Tengah: 10.61
Kuartil Atas: 11.622499999999999
----------------------------------------
IQR: 2.0624999999999982
----------------------------------------
Skewness: -0.01899140432111647
----------------------------------------
Kurtosis: -0.13173155932281988
========================================
```

Kolom Quality

In [ ]:
```python
# Quality
print("====================o====================")
print("Deskripsi Statistika Kolom Quality")
print("====================o====================")

df_quality = df["quality"]

# Print Descriptive Statistics
print_descriptive_statistics(df_quality)
```

```
====================o====================
Deskripsi Statistika Kolom Quality
====================o====================
Mean: 7.958
----------------------------------------
Median: 8.0
----------------------------------------
Modus:
8
----------------------------------------
Standar Deviasi: 0.9028017783827452
----------------------------------------
Variansi: 0.8150510510510475
----------------------------------------
Range: 5
----------------------------------------
Nilai Minimum: 5
----------------------------------------
Nilai Maksimum: 10
----------------------------------------
Kuartil Bawah: 7.0
Kuartil Tengah: 8.0
Kuartil Atas: 9.0
----------------------------------------
IQR: 2.0
----------------------------------------
Skewness: -0.08905409122491781
----------------------------------------
Kurtosis: 0.10829100232871003
========================================
```