

Aim: Explore the descriptive statistics on the given dataset.

```
from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True)
```

Load the Dataset

	Gender	Height	Weight	bmi	Age
0	Male	174	80	26.4	25
1	Male	189	87	24.4	27
2	Female	185	80	23.4	30
3	Female	165	70	25.7	26
4	Male	149	61	27.5	28
5	Male	177	70	22.3	29
6	Female	147	65	30.1	31
7	Male	154	62	26.1	32
8	Male	174	90	29.7	27

	Height	Weight	bmi	Age
count	9.000000	9.000000	9.000000	9.000000
mean	168.222222	73.888889	26.177778	28.333333
std	15.368619	10.740629	2.639497	2.345208
min	147.000000	61.000000	22.300000	25.000000
25%	154.000000	65.000000	24.400000	27.000000
50%	174.000000	70.000000	26.100000	28.000000
75%	177.000000	80.000000	27.500000	30.000000
max	189.000000	90.000000	30.100000	32.000000

Measure of Central Tendency

Mean

```
df.mean()

<ipython-input-7-c61f0c8f89b5>:1: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a
df.mean()
Height      168.222222
Weight      73.888889
bmi         26.177778
Age         28.333333
dtype: float64

mean = df["Age"].mean()
mean

28.333333333333332
```

Median

```
df.median()

<ipython-input-8-6d467abf240d>:1: FutureWarning: The default value of numeric_only in DataFrame.median is deprecated. In
df.median()
Height      174.0
Weight      70.0
bmi         26.1
Age         28.0
dtype: float64

median = df["Age"].median()
median

28.0
```

Mode

```
df.mode()

Gender  Height  Weight  bmi  Age
0      Male   174.0    70.0  22.3  27.0
1      NaN    NaN     80.0  23.4  NaN
2      NaN    NaN     NaN   24.4  NaN
3      NaN    NaN     NaN   25.7  NaN
4      NaN    NaN     NaN   26.1  NaN
5      NaN    NaN     NaN   26.4  NaN
6      NaN    NaN     NaN   27.5  NaN
7      NaN    NaN     NaN   29.7  NaN
8      NaN    NaN     NaN   30.1  NaN

mode = df["Age"].mode()
mode

0      27
Name: Age, dtype: int64
```

Measure of spread

```
df.var()

<ipython-input-16-28ded241fd7c>:1: FutureWarning: The default value of numeric_only in DataFrame.var is deprecated. In a
df.var()
Height      236.194444
Weight      115.361111
bmi         6.966944
```

```
Age          5.500000  
dtype: float64
```

```
df.std()
```

```
<ipython-input-17-ce97bb7eaef8>:1: FutureWarning: The default value of numeric_only in DataFrame.std is deprecated. In a  
df.std()  
Height      15.368619  
Weight      10.740629  
bmi         2.639497  
Age         2.345208  
dtype: float64
```

```
var = df["Age"].var()  
var
```

```
5.5
```

```
sd = df["Age"].std()  
sd
```

```
2.345207879911715
```

```
M1 = df["Age"].max()  
M1
```

```
32
```

```
M2 = df["Age"].min()  
M2
```

```
25
```

```
range = M1 - M2  
range
```

```
7
```

```
Q1 = df["Age"].quantile(0.25)  
Q1
```

```
27.0
```

```
Q3 = df["Age"].quantile(0.75)  
Q3
```

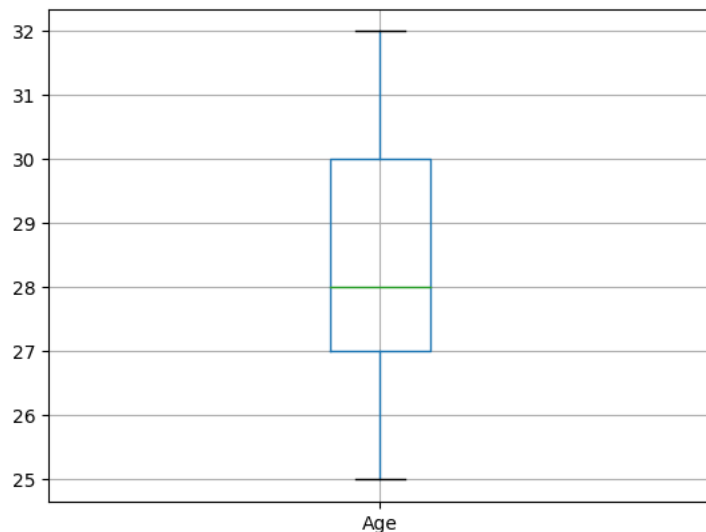
```
30.0
```

```
IQR = Q3 - Q1  
IQR
```

```
3.0
```

```
df.boxplot(column = "Age")
```

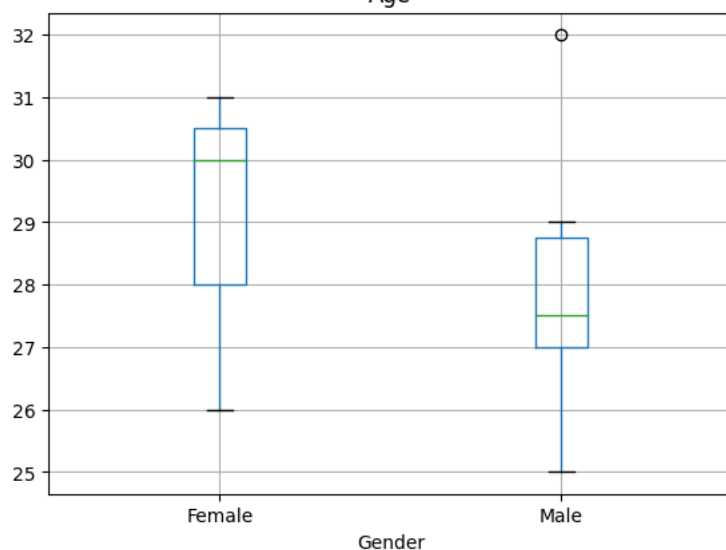
<Axes: >



```
df.boxplot(column = "Age", by = "Gender")
```

<Axes: title={'center': 'Age'}, xlabel='Gender'>

Boxplot grouped by Gender
Age



Measure of shape

```
Skew = df["Age"].skew()
Skew
```

0.232582599660668

```
Kurtosis = df["Age"].kurtosis()
Kurtosis
```

-1.041322314049585

Platykurtic (kurtois<3)

Leptokurtic (kurtois>3)

mesokurtic (kurtois=3)

```
Risk = sd/mean
Risk
```

0.08277204282041346

```
Harmonic_mean = stats.hmean(df["Age"])
Harmonic_mean
```

```
28.162486494356003
```

```
DQ = Harmonic_mean/mean
DQ
```

```
0.993970111565506
```

```
zscore = stats.zscore(df["Age"])
zscore
```

```
0    -1.507557
1    -0.603023
2     0.753778
3    -1.055290
4    -0.150756
5     0.301511
6     1.206045
7     1.658312
8    -0.603023
Name: Age, dtype: float64
```

Conclusion:

In this experiment we explore Descriptive Statistics for "BMI" Dataset. We perform all statistics operations like measures of central tendency, measures of spread and measures of shape on "Age" Column.

Measure of central Tendency: Measure of central tendency is used to describe the middle/centre value of the data. Mean, Median, Mode are measures of central tendency.

Measure of Spread: The measure of spread/dispersion is used to describe how data is spread. It also describes the variability of the dataset. Standard Deviation, Variance, Range, IQR, are used to describe the measure of spread.