Day Objectives

Identifying and Eliminating Outliers

- **Outliers are observations that are significantly different from other data points**
- **Outliers can adversely affect the training process of a machine learning algorithm, resulting in a loss of accuracy.**
- **Need to use the mathematical formula and retrieve the outlier data.**

  interquartile range(IQR) = Q3(quantile(0.75)) – Q1(quantile(0.25))

In [4]:

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

In [2]:

```
adv = pd.read_csv("https://raw.githubusercontent.com/APSSDC-Data-Analysis/Data-Analysis-batch8/main/Datasets/Advertising.csv")
adv.head()
```

Out[2]:

	TV	radio	newspaper	sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

In [3]:

```
adv.shape
```

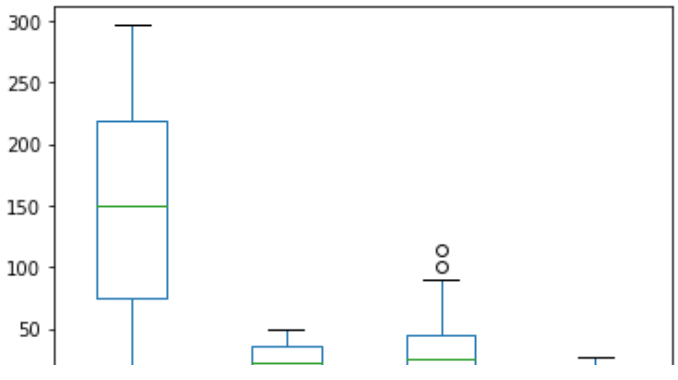
Out[3]:

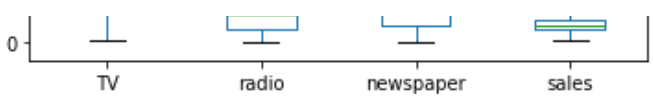
```
(200, 4)
```

Identifing Outliers

In [6]:

```
adv.plot( kind = "box")
plt.show()
```





In [7]:

```
adv.describe()
```

Out[7]:

	TV	radio	newspaper	sales
count	200.000000	200.000000	200.000000	200.000000
mean	147.042500	23.264000	30.554000	14.022500
std	85.854236	14.846809	21.778621	5.217457
min	0.700000	0.000000	0.300000	1.600000
25%	74.375000	9.975000	12.750000	10.375000
50%	149.750000	22.900000	25.750000	12.900000
75%	218.825000	36.525000	45.100000	17.400000
max	296.400000	49.600000	114.000000	27.000000

In [8]:

```
# interquartile range(IQR) = Q3(quantile(0.75)) - Q1(quantile(0.25))
```

In [9]:

```
Q1 = adv.quantile(0.25)
Q1
```

Out[9]:

```
TV           74.375
radio         9.975
newspaper    12.750
sales        10.375
Name: 0.25, dtype: float64
```

In [10]:

```
Q3  = adv.quantile(0.75)
Q3
```

Out[10]:

```
TV           218.825
radio         36.525
newspaper     45.100
sales         17.400
Name: 0.75, dtype: float64
```

In [11]:

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

Out[11]:

```
TV           144.450
radio         26.550
newspaper     32.350
sales          7.025
dtype: float64
```

In [12]:

```
min_value = (Q1 - 1.5*IQR)
max_value = (Q3 + 1.5*IQR)
```

In [15]:

```
outliers_df = adv[((adv < min_value )  | (adv > max_value)).any(axis = 1)]
outliers_df
```

Out[15]:

	TV	radio	newspaper	sales
16	67.8	36.6	114.0	12.5
101	296.4	36.3	100.9	23.8

In [20]:

```
filter_df = adv[~((adv < min_value )  | (adv > max_value)).any(axis = 1)]
filter_df
```

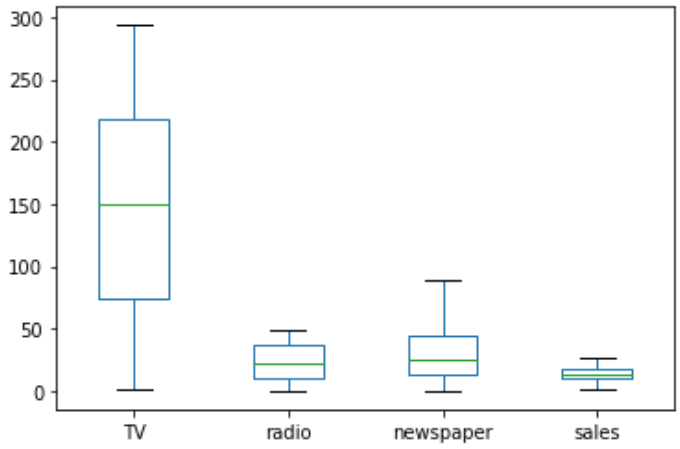
Out[20]:

	TV	radio	newspaper	sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9
...	...	...	...	...
195	38.2	3.7	13.8	7.6
196	94.2	4.9	8.1	9.7
197	177.0	9.3	6.4	12.8
198	283.6	42.0	66.2	25.5
199	232.1	8.6	8.7	13.4

198 rows × 4 columns

In [18]:

```
filter_df.plot(kind = "box")
plt.show()
```



In [ ]:

```

```

In [ ]:

```

```

Preprocessing Techniques

- Data Preprocessing is a technique that is used to convert the raw data into a clean data set

Data preprocessing steps

- Loading data
- Exploring data
- Cleaning data
- Transforming data
  - will learn data preprocessing techniques with scikit-learn, one of the most popular frameworks used for industry data science
  - The scikit-learn library includes tools for data preprocessing and data mining. It is imported in Python via the statement import sklearn.

Data Imputation

- if the dataset is missing too many values, we just don't use it
- if only a few of the values are missing, we can perform data imputation to substitute the missing data with some other value(s).
- There are many different methods for data imputation
  - Using the mean value
  - Using the median value
  - Using the most frequent value
  - Filling in missing values with a constant

Feature Scaling

1. Standardizing Data

- Data scientists will convert the data into a standard format to make it easier to understand.
- The standard format refers to data that has 0 mean and unit variance (i.e. standard deviation = 1), and the process of converting data into this format is called data standardization.
- improve the performance of models
- it rescales the data to have mean = 0 and varience(statistical measure that provides indicator of data's dispresion) = 1
- Standardization rescales data so that it has a mean of 0 and a standard deviation of 1.
- The formula for this is: (–)/
  - We subtract the mean () from each value (x) and then divide by the standard deviation ()

2. Data Range

- Scale data by compressing it into a fixed range
- One of the biggest use cases for this is compressing data into the range [0, 1]
- Classifier is MinMaxScaler

3. Normalizing Data

- Want to scale the individual data observations (i.e. rows)
- Used in classification Problems and data mining
- **ex : columns : salary,ex_yr,position_levels**
- when clustering data we need to apply L2 normalization to each row

- L2 normalization applied to a particular row of a data array
- L2 norm of a row is just the square root of the sum of squared values for the row

4. Robust Scaling

- Deal with is outliers (data point that is significantly further away from the other data points)
- Robustly scale the data, i.e. avoid being affected by outliers
- Scaling by using data's median and Interquartile Range (IQR)
- Here mean affected but median remains same
- Subtract the median from each data value then scale to the IQR

Data Imputation

- Handling Missing Data
  - There are many different methods for data imputation
    - Using the mean value
    - Using the median value
    - Using the most frequent value
    - Filling in missing values with a constant

In [27]:

```
titanic = pd.read_csv("https://raw.githubusercontent.com/APSSDC-Data-Analysis/Data-Analysis-batch8/main/Datasets/titanic.csv")
titanic.head()
```

Out[27]:

	survived	pclass	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked
0	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [28]:

```
titanic.shape  # 1000 emp details
```

Out[28]:

```
(891, 11)
```

In [29]:

```
titanic.isnull().sum()
```

Out[29]:

```
survived      0
pclass        0
name          0
sex           0
age         177
sibsp         0
parch         0
ticket        0
fare          0
cabin       687
```

```
embarked      2
dtype: int64
```

In [30]:

```
titanic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   survived  891 non-null    int64  
 1   pclass    891 non-null    int64  
 2   name      891 non-null    object 
 3   sex       891 non-null    object 
 4   age       714 non-null    float64
 5   sibsp     891 non-null    int64  
 6   parch     891 non-null    int64  
 7   ticket    891 non-null    object 
 8   fare      891 non-null    float64
 9   cabin     204 non-null    object 
 10  embarked  889 non-null    object 
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

In [24]:

```
from sklearn.impute import SimpleImputer
```

In [25]:

```
si = SimpleImputer(strategy= "mean")
si
```

Out[25]:

```
SimpleImputer()
```

In [32]:

```
si.fit_transform(titanic[["age"]])
```

Out[32]:

```
array([[22.        ],
       [38.        ],
       [26.        ],
       [35.        ],
       [35.        ],
       [29.69911765],
       [54.        ],
       [ 2.        ],
       [27.        ],
       [14.        ],
       [ 4.        ],
       [58.        ],
       [20.        ],
       [39.        ],
       [14.        ],
       [55.        ],
       [ 2.        ],
       [29.69911765],
       [31.        ],
       [29.69911765],
       [35.        ],
       [34.        ],
       [15.        ],
       [28.        ],
       [ 8.        ],
       [38.        ],
       [29.69911765],
```

```
       [19.        ],
       [29.69911765],
       [29.69911765],
       [40.        ],
       [29.69911765],
       [29.69911765],
       [66.        ],
       [28.        ],
       [42.        ],
       [29.69911765],
       [21.        ],
       [18.        ],
       [14.        ],
       [40.        ],
       [27.        ],
       [29.69911765],
       [ 3.        ],
       [19.        ],
       [29.69911765],
       [29.69911765],
       [29.69911765],
       [29.69911765],
       [18.        ],
       [ 7.        ],
       [21.        ],
       [49.        ],
       [29.        ],
       [65.        ],
       [29.69911765],
       [21.        ],
       [28.5       ],
       [ 5.        ],
       [11.        ],
       [22.        ],
       [38.        ],
       [45.        ],
       [ 4.        ],
       [29.69911765],
       [29.69911765],
       [29.        ],
       [19.        ],
       [17.        ],
       [26.        ],
       [32.        ],
       [16.        ],
       [21.        ],
       [26.        ],
       [32.        ],
       [25.        ],
       [29.69911765],
       [29.69911765],
       [ 0.83      ],
       [30.        ],
       [22.        ],
       [29.        ],
       [29.69911765],
       [28.        ],
       [17.        ],
       [33.        ],
       [16.        ],
       [29.69911765],
       [23.        ],
       [24.        ],
       [29.        ],
       [20.        ],
       [46.        ],
       [26.        ],
       [59.        ],
       [29.69911765],
       [71.        ],
       [23.        ],
       [34.        ],
```

```
       [34.        ],
       [28.        ],
       [29.69911765],
       [21.        ],
       [33.        ],
       [37.        ],
       [28.        ],
       [21.        ],
       [29.69911765],
       [38.        ],
       [29.69911765],
       [47.        ],
       [14.5       ],
       [22.        ],
       [20.        ],
       [17.        ],
       [21.        ],
       [70.5       ],
       [29.        ],
       [24.        ],
       [ 2.        ],
       [21.        ],
       [29.69911765],
       [32.5       ],
       [32.5       ],
       [54.        ],
       [12.        ],
       [29.69911765],
       [24.        ],
       [29.69911765],
       [45.        ],
       [33.        ],
       [20.        ],
       [47.        ],
       [29.        ],
       [25.        ],
       [23.        ],
       [19.        ],
       [37.        ],
       [16.        ],
       [24.        ],
       [29.69911765],
       [22.        ],
       [24.        ],
       [19.        ],
       [18.        ],
       [19.        ],
       [27.        ],
       [ 9.        ],
       [36.5       ],
       [42.        ],
       [51.        ],
       [22.        ],
       [55.5       ],
       [40.5       ],
       [29.69911765],
       [51.        ],
       [16.        ],
       [30.        ],
       [29.69911765],
       [29.69911765],
       [44.        ],
       [40.        ],
       [26.        ],
       [17.        ],
       [ 1.        ],
       [ 9.        ],
       [29.69911765],
       [45.        ],
       [29.69911765],
       [28.        ],
       [61.        ],
```

```
       [ 4.        ],
       [ 1.        ],
       [21.        ],
       [56.        ],
       [18.        ],
       [29.69911765],
       [50.        ],
       [30.        ],
       [36.        ],
       [29.69911765],
       [29.69911765],
       [ 9.        ],
       [ 1.        ],
       [ 4.        ],
       [29.69911765],
       [29.69911765],
       [45.        ],
       [40.        ],
       [36.        ],
       [32.        ],
       [19.        ],
       [19.        ],
       [ 3.        ],
       [44.        ],
       [58.        ],
       [29.69911765],
       [42.        ],
       [29.69911765],
       [24.        ],
       [28.        ],
       [29.69911765],
       [34.        ],
       [45.5       ],
       [18.        ],
       [ 2.        ],
       [32.        ],
       [26.        ],
       [16.        ],
       [40.        ],
       [24.        ],
       [35.        ],
       [22.        ],
       [30.        ],
       [29.69911765],
       [31.        ],
       [27.        ],
       [42.        ],
       [32.        ],
       [30.        ],
       [16.        ],
       [27.        ],
       [51.        ],
       [29.69911765],
       [38.        ],
       [22.        ],
       [19.        ],
       [20.5       ],
       [18.        ],
       [29.69911765],
       [35.        ],
       [29.        ],
       [59.        ],
       [ 5.        ],
       [24.        ],
       [29.69911765],
       [44.        ],
       [ 8.        ],
       [19.        ],
       [33.        ],
       [29.69911765],
       [29.69911765],
       [29.        ],
```

```
       [22.        ],
       [30.        ],
       [44.        ],
       [25.        ],
       [24.        ],
       [37.        ],
       [54.        ],
       [29.69911765],
       [29.        ],
       [62.        ],
       [30.        ],
       [41.        ],
       [29.        ],
       [29.69911765],
       [30.        ],
       [35.        ],
       [50.        ],
       [29.69911765],
       [ 3.        ],
       [52.        ],
       [40.        ],
       [29.69911765],
       [36.        ],
       [16.        ],
       [25.        ],
       [58.        ],
       [35.        ],
       [29.69911765],
       [25.        ],
       [41.        ],
       [37.        ],
       [29.69911765],
       [63.        ],
       [45.        ],
       [29.69911765],
       [ 7.        ],
       [35.        ],
       [65.        ],
       [28.        ],
       [16.        ],
       [19.        ],
       [29.69911765],
       [33.        ],
       [30.        ],
       [22.        ],
       [42.        ],
       [22.        ],
       [26.        ],
       [19.        ],
       [36.        ],
       [24.        ],
       [24.        ],
       [29.69911765],
       [23.5       ],
       [ 2.        ],
       [29.69911765],
       [50.        ],
       [29.69911765],
       [29.69911765],
       [19.        ],
       [29.69911765],
       [29.69911765],
       [ 0.92      ],
       [29.69911765],
       [17.        ],
       [30.        ],
       [30.        ],
       [24.        ],
       [18.        ],
       [26.        ],
       [28.        ],
       [43.        ],
```

```
[26.        ],
[24.        ],
[54.        ],
[31.        ],
[40.        ],
[22.        ],
[27.        ],
[30.        ],
[22.        ],
[29.69911765],
[36.        ],
[61.        ],
[36.        ],
[31.        ],
[16.        ],
[29.69911765],
[45.5       ],
[38.        ],
[16.        ],
[29.69911765],
[29.69911765],
[29.        ],
[41.        ],
[45.        ],
[45.        ],
[ 2.        ],
[24.        ],
[28.        ],
[25.        ],
[36.        ],
[24.        ],
[40.        ],
[29.69911765],
[ 3.        ],
[42.        ],
[23.        ],
[29.69911765],
[15.        ],
[25.        ],
[29.69911765],
[28.        ],
[22.        ],
[38.        ],
[29.69911765],
[29.69911765],
[40.        ],
[29.        ],
[45.        ],
[35.        ],
[29.69911765],
[30.        ],
[60.        ],
[29.69911765],
[29.69911765],
[24.        ],
[25.        ],
[18.        ],
[19.        ],
[22.        ],
[ 3.        ],
[29.69911765],
[22.        ],
[27.        ],
[20.        ],
[19.        ],
[42.        ],
[ 1.        ],
[32.        ],
[35.        ],
[29.69911765],
[18.        ],
[ 1.        ],
```

```
       [36.        ],
       [29.69911765],
       [17.        ],
       [36.        ],
       [21.        ],
       [28.        ],
       [23.        ],
       [24.        ],
       [22.        ],
       [31.        ],
       [46.        ],
       [23.        ],
       [28.        ],
       [39.        ],
       [26.        ],
       [21.        ],
       [28.        ],
       [20.        ],
       [34.        ],
       [51.        ],
       [ 3.        ],
       [21.        ],
       [29.69911765],
       [29.69911765],
       [29.69911765],
       [33.        ],
       [29.69911765],
       [44.        ],
       [29.69911765],
       [34.        ],
       [18.        ],
       [30.        ],
       [10.        ],
       [29.69911765],
       [21.        ],
       [29.        ],
       [28.        ],
       [18.        ],
       [29.69911765],
       [28.        ],
       [19.        ],
       [29.69911765],
       [32.        ],
       [28.        ],
       [29.69911765],
       [42.        ],
       [17.        ],
       [50.        ],
       [14.        ],
       [21.        ],
       [24.        ],
       [64.        ],
       [31.        ],
       [45.        ],
       [20.        ],
       [25.        ],
       [28.        ],
       [29.69911765],
       [ 4.        ],
       [13.        ],
       [34.        ],
       [ 5.        ],
       [52.        ],
       [36.        ],
       [29.69911765],
       [30.        ],
       [49.        ],
       [29.69911765],
       [29.        ],
       [65.        ],
       [29.69911765],
       [50.        ],
```

```
[29.69911765],
[48.        ],
[34.        ],
[47.        ],
[48.        ],
[29.69911765],
[38.        ],
[29.69911765],
[56.        ],
[29.69911765],
[ 0.75      ],
[29.69911765],
[38.        ],
[33.        ],
[23.        ],
[22.        ],
[29.69911765],
[34.        ],
[29.        ],
[22.        ],
[ 2.        ],
[ 9.        ],
[29.69911765],
[50.        ],
[63.        ],
[25.        ],
[29.69911765],
[35.        ],
[58.        ],
[30.        ],
[ 9.        ],
[29.69911765],
[21.        ],
[55.        ],
[71.        ],
[21.        ],
[29.69911765],
[54.        ],
[29.69911765],
[25.        ],
[24.        ],
[17.        ],
[21.        ],
[29.69911765],
[37.        ],
[16.        ],
[18.        ],
[33.        ],
[29.69911765],
[28.        ],
[26.        ],
[29.        ],
[29.69911765],
[36.        ],
[54.        ],
[24.        ],
[47.        ],
[34.        ],
[29.69911765],
[36.        ],
[32.        ],
[30.        ],
[22.        ],
[29.69911765],
[44.        ],
[29.69911765],
[40.5       ],
[50.        ],
[29.69911765],
[39.        ],
[23.        ],
[ 2.        ],
```

```
        [29.69911765],
        [17.        ],
        [29.69911765],
        [30.        ],
        [ 7.        ],
        [45.        ],
        [30.        ],
        [29.69911765],
        [22.        ],
        [36.        ],
        [ 9.        ],
        [11.        ],
        [32.        ],
        [50.        ],
        [64.        ],
        [19.        ],
        [29.69911765],
        [33.        ],
        [ 8.        ],
        [17.        ],
        [27.        ],
        [29.69911765],
        [22.        ],
        [22.        ],
        [62.        ],
        [48.        ],
        [29.69911765],
        [39.        ],
        [36.        ],
        [29.69911765],
        [40.        ],
        [28.        ],
        [29.69911765],
        [29.69911765],
        [24.        ],
        [19.        ],
        [29.        ],
        [29.69911765],
        [32.        ],
        [62.        ],
        [53.        ],
        [36.        ],
        [29.69911765],
        [16.        ],
        [19.        ],
        [34.        ],
        [39.        ],
        [29.69911765],
        [32.        ],
        [25.        ],
        [39.        ],
        [54.        ],
        [36.        ],
        [29.69911765],
        [18.        ],
        [47.        ],
        [60.        ],
        [22.        ],
        [29.69911765],
        [35.        ],
        [52.        ],
        [47.        ],
        [29.69911765],
        [37.        ],
        [36.        ],
        [29.69911765],
        [49.        ],
        [29.69911765],
        [49.        ],
        [24.        ],
        [29.69911765],
        [29.69911765],
```

```
        [44.        ],
        [35.        ],
        [36.        ],
        [30.        ],
        [27.        ],
        [22.        ],
        [40.        ],
        [39.        ],
        [29.69911765],
        [29.69911765],
        [29.69911765],
        [35.        ],
        [24.        ],
        [34.        ],
        [26.        ],
        [ 4.        ],
        [26.        ],
        [27.        ],
        [42.        ],
        [20.        ],
        [21.        ],
        [21.        ],
        [61.        ],
        [57.        ],
        [21.        ],
        [26.        ],
        [29.69911765],
        [80.        ],
        [51.        ],
        [32.        ],
        [29.69911765],
        [ 9.        ],
        [28.        ],
        [32.        ],
        [31.        ],
        [41.        ],
        [29.69911765],
        [20.        ],
        [24.        ],
        [ 2.        ],
        [29.69911765],
        [ 0.75      ],
        [48.        ],
        [19.        ],
        [56.        ],
        [29.69911765],
        [23.        ],
        [29.69911765],
        [18.        ],
        [21.        ],
        [29.69911765],
        [18.        ],
        [24.        ],
        [29.69911765],
        [32.        ],
        [23.        ],
        [58.        ],
        [50.        ],
        [40.        ],
        [47.        ],
        [36.        ],
        [20.        ],
        [32.        ],
        [25.        ],
        [29.69911765],
        [43.        ],
        [29.69911765],
        [40.        ],
        [31.        ],
        [70.        ],
        [31.        ],
        [29.69911765],
```

```
       [18.        ],
       [24.5       ],
       [18.        ],
       [43.        ],
       [36.        ],
       [29.69911765],
       [27.        ],
       [20.        ],
       [14.        ],
       [60.        ],
       [25.        ],
       [14.        ],
       [19.        ],
       [18.        ],
       [15.        ],
       [31.        ],
       [ 4.        ],
       [29.69911765],
       [25.        ],
       [60.        ],
       [52.        ],
       [44.        ],
       [29.69911765],
       [49.        ],
       [42.        ],
       [18.        ],
       [35.        ],
       [18.        ],
       [25.        ],
       [26.        ],
       [39.        ],
       [45.        ],
       [42.        ],
       [22.        ],
       [29.69911765],
       [24.        ],
       [29.69911765],
       [48.        ],
       [29.        ],
       [52.        ],
       [19.        ],
       [38.        ],
       [27.        ],
       [29.69911765],
       [33.        ],
       [ 6.        ],
       [17.        ],
       [34.        ],
       [50.        ],
       [27.        ],
       [20.        ],
       [30.        ],
       [29.69911765],
       [25.        ],
       [25.        ],
       [29.        ],
       [11.        ],
       [29.69911765],
       [23.        ],
       [23.        ],
       [28.5       ],
       [48.        ],
       [35.        ],
       [29.69911765],
       [29.69911765],
       [29.69911765],
       [36.        ],
       [21.        ],
       [24.        ],
       [31.        ],
       [70.        ],
       [16.        ],
```

```
[30.        ],
[19.        ],
[31.        ],
[ 4.        ],
[ 6.        ],
[33.        ],
[23.        ],
[48.        ],
[ 0.67      ],
[28.        ],
[18.        ],
[34.        ],
[33.        ],
[29.69911765],
[41.        ],
[20.        ],
[36.        ],
[16.        ],
[51.        ],
[29.69911765],
[30.5       ],
[29.69911765],
[32.        ],
[24.        ],
[48.        ],
[57.        ],
[29.69911765],
[54.        ],
[18.        ],
[29.69911765],
[ 5.        ],
[29.69911765],
[43.        ],
[13.        ],
[17.        ],
[29.        ],
[29.69911765],
[25.        ],
[25.        ],
[18.        ],
[ 8.        ],
[ 1.        ],
[46.        ],
[29.69911765],
[16.        ],
[29.69911765],
[29.69911765],
[25.        ],
[39.        ],
[49.        ],
[31.        ],
[30.        ],
[30.        ],
[34.        ],
[31.        ],
[11.        ],
[ 0.42      ],
[27.        ],
[31.        ],
[39.        ],
[18.        ],
[39.        ],
[33.        ],
[26.        ],
[39.        ],
[35.        ],
[ 6.        ],
[30.5       ],
[29.69911765],
[23.        ],
[31.        ],
[43.        ],
```

```
       [10.        ],
       [52.        ],
       [27.        ],
       [38.        ],
       [27.        ],
       [ 2.        ],
       [29.69911765],
       [29.69911765],
       [ 1.        ],
       [29.69911765],
       [62.        ],
       [15.        ],
       [ 0.83      ],
       [29.69911765],
       [23.        ],
       [18.        ],
       [39.        ],
       [21.        ],
       [29.69911765],
       [32.        ],
       [29.69911765],
       [20.        ],
       [16.        ],
       [30.        ],
       [34.5       ],
       [17.        ],
       [42.        ],
       [29.69911765],
       [35.        ],
       [28.        ],
       [29.69911765],
       [ 4.        ],
       [74.        ],
       [ 9.        ],
       [16.        ],
       [44.        ],
       [18.        ],
       [45.        ],
       [51.        ],
       [24.        ],
       [29.69911765],
       [41.        ],
       [21.        ],
       [48.        ],
       [29.69911765],
       [24.        ],
       [42.        ],
       [27.        ],
       [31.        ],
       [29.69911765],
       [ 4.        ],
       [26.        ],
       [47.        ],
       [33.        ],
       [47.        ],
       [28.        ],
       [15.        ],
       [20.        ],
       [19.        ],
       [29.69911765],
       [56.        ],
       [25.        ],
       [33.        ],
       [22.        ],
       [28.        ],
       [25.        ],
       [39.        ],
       [27.        ],
       [19.        ],
       [29.69911765],
       [26.        ],
       [32.        ]])
```

In [33]:

```
titanic[["age"]].mean()
```

Out[33]:

```
age    29.699118
dtype: float64
```

In [34]:

```
si = SimpleImputer(strategy= "median")
si.fit_transform(titanic[["age"]])
```

Out[34]:

```
array([[22.  ],
       [38.  ],
       [26.  ],
       [35.  ],
       [35.  ],
       [28.  ],
       [54.  ],
       [ 2.  ],
       [27.  ],
       [14.  ],
       [ 4.  ],
       [58.  ],
       [20.  ],
       [39.  ],
       [14.  ],
       [55.  ],
       [ 2.  ],
       [28.  ],
       [31.  ],
       [28.  ],
       [35.  ],
       [34.  ],
       [15.  ],
       [28.  ],
       [ 8.  ],
       [38.  ],
       [28.  ],
       [19.  ],
       [28.  ],
       [28.  ],
       [40.  ],
       [28.  ],
       [28.  ],
       [66.  ],
       [28.  ],
       [42.  ],
       [28.  ],
       [21.  ],
       [18.  ],
       [14.  ],
       [40.  ],
       [27.  ],
       [28.  ],
       [ 3.  ],
       [19.  ],
       [28.  ],
       [28.  ],
       [28.  ],
       [28.  ],
       [18.  ],
       [ 7.  ],
       [21.  ],
       [49.  ],
       [29.  ],
       [65.  ],
       [28.  ],
```

[21.],
[28.5],
[5.],
[11.],
[22.],
[38.],
[45.],
[4.],
[28.],
[28.],
[29.],
[19.],
[17.],
[26.],
[32.],
[16.],
[21.],
[26.],
[32.],
[25.],
[28.],
[28.],
[0.83],
[30.],
[22.],
[29.],
[28.],
[28.],
[17.],
[33.],
[16.],
[28.],
[23.],
[24.],
[29.],
[20.],
[46.],
[26.],
[59.],
[28.],
[71.],
[23.],
[34.],
[34.],
[28.],
[28.],
[21.],
[33.],
[37.],
[28.],
[21.],
[28.],
[38.],
[28.],
[47.],
[14.5],
[22.],
[20.],
[17.],
[21.],
[70.5],
[29.],
[24.],
[2.],
[21.],
[28.],
[32.5],
[32.5],
[54.],
[12.],
[28.],
[24.],

[28.],
[45.],
[33.],
[20.],
[47.],
[29.],
[25.],
[23.],
[19.],
[37.],
[16.],
[24.],
[28.],
[22.],
[24.],
[19.],
[18.],
[19.],
[27.],
[9.],
[36.5],
[42.],
[51.],
[22.],
[55.5],
[40.5],
[28.],
[51.],
[16.],
[30.],
[28.],
[28.],
[44.],
[40.],
[26.],
[17.],
[1.],
[9.],
[28.],
[45.],
[28.],
[28.],
[61.],
[4.],
[1.],
[21.],
[56.],
[18.],
[28.],
[50.],
[30.],
[36.],
[28.],
[28.],
[9.],
[1.],
[4.],
[28.],
[28.],
[45.],
[40.],
[36.],
[32.],
[19.],
[19.],
[3.],
[44.],
[58.],
[28.],
[42.],
[28.],
[24.],

[28.],
[28.],
[34.],
[45.5],
[18.],
[2.],
[32.],
[26.],
[16.],
[40.],
[24.],
[35.],
[22.],
[30.],
[28.],
[31.],
[27.],
[42.],
[32.],
[30.],
[16.],
[27.],
[51.],
[28.],
[38.],
[22.],
[19.],
[20.5],
[18.],
[28.],
[35.],
[29.],
[59.],
[5.],
[24.],
[28.],
[44.],
[8.],
[19.],
[33.],
[28.],
[28.],
[29.],
[22.],
[30.],
[44.],
[25.],
[24.],
[37.],
[54.],
[28.],
[29.],
[62.],
[30.],
[41.],
[29.],
[28.],
[30.],
[35.],
[50.],
[28.],
[3.],
[52.],
[40.],
[28.],
[36.],
[16.],
[25.],
[58.],
[35.],
[28.],
[25.],

[41.],
[37.],
[28.],
[63.],
[45.],
[28.],
[7.],
[35.],
[65.],
[28.],
[16.],
[19.],
[28.],
[33.],
[30.],
[22.],
[42.],
[22.],
[26.],
[19.],
[36.],
[24.],
[24.],
[28.],
[23.5],
[2.],
[28.],
[50.],
[28.],
[28.],
[19.],
[28.],
[28.],
[0.92],
[28.],
[17.],
[30.],
[30.],
[24.],
[18.],
[26.],
[28.],
[43.],
[26.],
[24.],
[54.],
[31.],
[40.],
[22.],
[27.],
[30.],
[22.],
[28.],
[36.],
[61.],
[36.],
[31.],
[16.],
[28.],
[45.5],
[38.],
[16.],
[28.],
[28.],
[29.],
[41.],
[45.],
[45.],
[2.],
[24.],
[28.],
[25.],

[36.],
[24.],
[40.],
[28.],
[3.],
[42.],
[23.],
[28.],
[15.],
[25.],
[28.],
[28.],
[22.],
[38.],
[28.],
[28.],
[40.],
[29.],
[45.],
[35.],
[28.],
[30.],
[60.],
[28.],
[28.],
[24.],
[25.],
[18.],
[19.],
[22.],
[3.],
[28.],
[22.],
[27.],
[20.],
[19.],
[42.],
[1.],
[32.],
[35.],
[28.],
[18.],
[1.],
[36.],
[28.],
[17.],
[36.],
[21.],
[28.],
[23.],
[24.],
[22.],
[31.],
[46.],
[23.],
[28.],
[39.],
[26.],
[21.],
[28.],
[20.],
[34.],
[51.],
[3.],
[21.],
[28.],
[28.],
[28.],
[33.],
[28.],
[44.],
[28.],

[34.],
[18.],
[30.],
[10.],
[28.],
[21.],
[29.],
[28.],
[18.],
[28.],
[28.],
[19.],
[28.],
[32.],
[28.],
[28.],
[42.],
[17.],
[50.],
[14.],
[21.],
[24.],
[64.],
[31.],
[45.],
[20.],
[25.],
[28.],
[28.],
[4.],
[13.],
[34.],
[5.],
[52.],
[36.],
[28.],
[30.],
[49.],
[28.],
[29.],
[65.],
[28.],
[50.],
[28.],
[48.],
[34.],
[47.],
[48.],
[28.],
[38.],
[28.],
[56.],
[28.],
[0.75],
[28.],
[38.],
[33.],
[23.],
[22.],
[28.],
[34.],
[29.],
[22.],
[2.],
[9.],
[28.],
[50.],
[63.],
[25.],
[28.],
[35.],
[58.],

[30.],
[9.],
[28.],
[21.],
[55.],
[71.],
[21.],
[28.],
[54.],
[28.],
[25.],
[24.],
[17.],
[21.],
[28.],
[37.],
[16.],
[18.],
[33.],
[28.],
[28.],
[26.],
[29.],
[28.],
[36.],
[54.],
[24.],
[47.],
[34.],
[28.],
[36.],
[32.],
[30.],
[22.],
[28.],
[44.],
[28.],
[40.5],
[50.],
[28.],
[39.],
[23.],
[2.],
[28.],
[17.],
[28.],
[30.],
[7.],
[45.],
[30.],
[28.],
[22.],
[36.],
[9.],
[11.],
[32.],
[50.],
[64.],
[19.],
[28.],
[33.],
[8.],
[17.],
[27.],
[28.],
[22.],
[22.],
[62.],
[48.],
[28.],
[39.],
[36.],

[28.],
[40.],
[28.],
[28.],
[28.],
[24.],
[19.],
[29.],
[28.],
[32.],
[62.],
[53.],
[36.],
[28.],
[16.],
[19.],
[34.],
[39.],
[28.],
[32.],
[25.],
[39.],
[54.],
[36.],
[28.],
[18.],
[47.],
[60.],
[22.],
[28.],
[35.],
[52.],
[47.],
[28.],
[37.],
[36.],
[28.],
[49.],
[28.],
[49.],
[24.],
[28.],
[28.],
[44.],
[35.],
[36.],
[30.],
[27.],
[22.],
[40.],
[39.],
[28.],
[28.],
[28.],
[35.],
[24.],
[34.],
[26.],
[4.],
[26.],
[27.],
[42.],
[20.],
[21.],
[21.],
[61.],
[57.],
[21.],
[26.],
[28.],
[80.],
[51.],

[32.],
[28.],
[9.],
[28.],
[32.],
[31.],
[41.],
[28.],
[20.],
[24.],
[2.],
[28.],
[0.75],
[48.],
[19.],
[56.],
[28.],
[23.],
[28.],
[18.],
[21.],
[28.],
[18.],
[24.],
[28.],
[32.],
[23.],
[58.],
[50.],
[40.],
[47.],
[36.],
[20.],
[32.],
[25.],
[28.],
[43.],
[28.],
[40.],
[31.],
[70.],
[31.],
[28.],
[18.],
[24.5],
[18.],
[43.],
[36.],
[28.],
[27.],
[20.],
[14.],
[60.],
[25.],
[14.],
[19.],
[18.],
[15.],
[31.],
[4.],
[28.],
[25.],
[60.],
[52.],
[44.],
[28.],
[49.],
[42.],
[18.],
[35.],
[18.],
[25.],

[26.],
[39.],
[45.],
[42.],
[22.],
[28.],
[24.],
[28.],
[48.],
[29.],
[52.],
[19.],
[38.],
[27.],
[28.],
[33.],
[6.],
[17.],
[34.],
[50.],
[27.],
[20.],
[30.],
[28.],
[25.],
[25.],
[29.],
[11.],
[28.],
[23.],
[23.],
[28.5],
[48.],
[35.],
[28.],
[28.],
[28.],
[36.],
[21.],
[24.],
[31.],
[70.],
[16.],
[30.],
[19.],
[31.],
[4.],
[6.],
[33.],
[23.],
[48.],
[0.67],
[28.],
[18.],
[34.],
[33.],
[28.],
[41.],
[20.],
[36.],
[16.],
[51.],
[28.],
[30.5],
[28.],
[32.],
[24.],
[48.],
[57.],
[28.],
[54.],
[18.],

[28.],
[5.],
[28.],
[43.],
[13.],
[17.],
[29.],
[28.],
[25.],
[25.],
[18.],
[8.],
[1.],
[46.],
[28.],
[16.],
[28.],
[28.],
[25.],
[39.],
[49.],
[31.],
[30.],
[30.],
[34.],
[31.],
[11.],
[0.42],
[27.],
[31.],
[39.],
[18.],
[39.],
[33.],
[26.],
[39.],
[35.],
[6.],
[30.5],
[28.],
[23.],
[31.],
[43.],
[10.],
[52.],
[27.],
[38.],
[27.],
[2.],
[28.],
[28.],
[1.],
[28.],
[62.],
[15.],
[0.83],
[28.],
[23.],
[18.],
[39.],
[21.],
[28.],
[32.],
[28.],
[20.],
[16.],
[30.],
[34.5],
[17.],
[42.],
[28.],
[35.],

```
       [28.  ],
       [28.  ],
       [ 4.  ],
       [74.  ],
       [ 9.  ],
       [16.  ],
       [44.  ],
       [18.  ],
       [45.  ],
       [51.  ],
       [24.  ],
       [28.  ],
       [41.  ],
       [21.  ],
       [48.  ],
       [28.  ],
       [24.  ],
       [42.  ],
       [27.  ],
       [31.  ],
       [28.  ],
       [ 4.  ],
       [26.  ],
       [47.  ],
       [33.  ],
       [47.  ],
       [28.  ],
       [15.  ],
       [20.  ],
       [19.  ],
       [28.  ],
       [56.  ],
       [25.  ],
       [33.  ],
       [22.  ],
       [28.  ],
       [25.  ],
       [39.  ],
       [27.  ],
       [19.  ],
       [28.  ],
       [26.  ],
       [32.  ]])
```

In [35]:

```
titanic[["age"]].median()
```

Out[35]:

```
age    28.0
dtype: float64
```

In [38]:

```
si = SimpleImputer(strategy= "most_frequent")
si.fit_transform(titanic[["age"]])
```

Out[38]:

```
array([[22.  ],
       [38.  ],
       [26.  ],
       [35.  ],
       [35.  ],
       [24.  ],
       [54.  ],
       [ 2.  ],
       [27.  ],
       [14.  ],
       [ 4.  ],
       [58.  ],
       [20.  ],
```

[20.],
[39.],
[14.],
[55.],
[2.],
[24.],
[31.],
[24.],
[35.],
[34.],
[15.],
[28.],
[8.],
[38.],
[24.],
[19.],
[24.],
[24.],
[40.],
[24.],
[24.],
[66.],
[28.],
[42.],
[24.],
[21.],
[18.],
[14.],
[40.],
[27.],
[24.],
[3.],
[19.],
[24.],
[24.],
[24.],
[24.],
[18.],
[7.],
[21.],
[49.],
[29.],
[65.],
[24.],
[21.],
[28.5],
[5.],
[11.],
[22.],
[38.],
[45.],
[4.],
[24.],
[24.],
[29.],
[19.],
[17.],
[26.],
[32.],
[16.],
[21.],
[26.],
[32.],
[25.],
[24.],
[24.],
[0.83],
[30.],
[22.],
[29.],
[24.],
[28.],
[17.],

[17.],
[33.],
[16.],
[24.],
[23.],
[24.],
[29.],
[20.],
[46.],
[26.],
[59.],
[24.],
[71.],
[23.],
[34.],
[34.],
[28.],
[24.],
[21.],
[33.],
[37.],
[28.],
[21.],
[24.],
[38.],
[24.],
[47.],
[14.5],
[22.],
[20.],
[17.],
[21.],
[70.5],
[29.],
[24.],
[2.],
[21.],
[24.],
[32.5],
[32.5],
[54.],
[12.],
[24.],
[24.],
[24.],
[45.],
[33.],
[20.],
[47.],
[29.],
[25.],
[23.],
[19.],
[37.],
[16.],
[24.],
[24.],
[22.],
[24.],
[19.],
[18.],
[19.],
[27.],
[9.],
[36.5],
[42.],
[51.],
[22.],
[55.5],
[40.5],
[24.],
[51.],
[16.],

[16.],
[30.],
[24.],
[24.],
[44.],
[40.],
[26.],
[17.],
[1.],
[9.],
[24.],
[45.],
[24.],
[28.],
[61.],
[4.],
[1.],
[21.],
[56.],
[18.],
[24.],
[50.],
[30.],
[36.],
[24.],
[24.],
[9.],
[1.],
[4.],
[24.],
[24.],
[45.],
[40.],
[36.],
[32.],
[19.],
[19.],
[3.],
[44.],
[58.],
[24.],
[42.],
[24.],
[24.],
[28.],
[24.],
[34.],
[45.5],
[18.],
[2.],
[32.],
[26.],
[16.],
[40.],
[24.],
[35.],
[22.],
[30.],
[24.],
[31.],
[27.],
[42.],
[32.],
[30.],
[16.],
[27.],
[51.],
[24.],
[38.],
[22.],
[19.],
[20.5],
[18.]

[18.],
[24.],
[35.],
[29.],
[59.],
[5.],
[24.],
[24.],
[44.],
[8.],
[19.],
[33.],
[24.],
[24.],
[29.],
[22.],
[30.],
[44.],
[25.],
[24.],
[37.],
[54.],
[24.],
[29.],
[62.],
[30.],
[41.],
[29.],
[24.],
[30.],
[35.],
[50.],
[24.],
[3.],
[52.],
[40.],
[24.],
[36.],
[16.],
[25.],
[58.],
[35.],
[24.],
[25.],
[41.],
[37.],
[24.],
[63.],
[45.],
[24.],
[7.],
[35.],
[65.],
[28.],
[16.],
[19.],
[24.],
[33.],
[30.],
[22.],
[42.],
[22.],
[26.],
[19.],
[36.],
[24.],
[24.],
[24.],
[23.5],
[2.],
[24.],
[50.],
[24.],

[24.],
[24.],
[19.],
[24.],
[24.],
[0.92],
[24.],
[17.],
[30.],
[30.],
[24.],
[18.],
[26.],
[28.],
[43.],
[26.],
[24.],
[54.],
[31.],
[40.],
[22.],
[27.],
[30.],
[22.],
[24.],
[36.],
[61.],
[36.],
[31.],
[16.],
[24.],
[45.5],
[38.],
[16.],
[24.],
[24.],
[29.],
[41.],
[45.],
[45.],
[2.],
[24.],
[28.],
[25.],
[36.],
[24.],
[40.],
[24.],
[3.],
[42.],
[23.],
[24.],
[15.],
[25.],
[24.],
[28.],
[22.],
[38.],
[24.],
[24.],
[40.],
[29.],
[45.],
[35.],
[24.],
[30.],
[60.],
[24.],
[24.],
[24.],
[25.],
[18.],
[19.],

[19.],
[22.],
[3.],
[24.],
[22.],
[27.],
[20.],
[19.],
[42.],
[1.],
[32.],
[35.],
[24.],
[18.],
[1.],
[36.],
[24.],
[17.],
[36.],
[21.],
[28.],
[23.],
[24.],
[22.],
[31.],
[46.],
[23.],
[28.],
[39.],
[26.],
[21.],
[28.],
[20.],
[34.],
[51.],
[3.],
[21.],
[24.],
[24.],
[24.],
[33.],
[24.],
[44.],
[24.],
[34.],
[18.],
[30.],
[10.],
[24.],
[21.],
[29.],
[28.],
[18.],
[24.],
[28.],
[19.],
[24.],
[32.],
[28.],
[24.],
[42.],
[17.],
[50.],
[14.],
[21.],
[24.],
[64.],
[31.],
[45.],
[20.],
[25.],
[28.],
[24.],

[24.],
[4.],
[13.],
[34.],
[5.],
[52.],
[36.],
[24.],
[30.],
[49.],
[24.],
[29.],
[65.],
[24.],
[50.],
[24.],
[48.],
[34.],
[47.],
[48.],
[24.],
[38.],
[24.],
[56.],
[24.],
[0.75],
[24.],
[38.],
[33.],
[23.],
[22.],
[24.],
[34.],
[29.],
[22.],
[2.],
[9.],
[24.],
[50.],
[63.],
[25.],
[24.],
[35.],
[58.],
[30.],
[9.],
[24.],
[21.],
[55.],
[71.],
[21.],
[24.],
[54.],
[24.],
[25.],
[24.],
[17.],
[21.],
[24.],
[37.],
[16.],
[18.],
[33.],
[24.],
[28.],
[26.],
[29.],
[24.],
[36.],
[54.],
[24.],
[47.],
[34.],

[34.],
[24.],
[36.],
[32.],
[30.],
[22.],
[24.],
[44.],
[24.],
[40.5],
[50.],
[24.],
[39.],
[23.],
[2.],
[24.],
[17.],
[24.],
[30.],
[7.],
[45.],
[30.],
[24.],
[22.],
[36.],
[9.],
[11.],
[32.],
[50.],
[64.],
[19.],
[24.],
[33.],
[8.],
[17.],
[27.],
[24.],
[22.],
[22.],
[62.],
[48.],
[24.],
[39.],
[36.],
[24.],
[40.],
[28.],
[24.],
[24.],
[24.],
[19.],
[29.],
[24.],
[32.],
[62.],
[53.],
[36.],
[24.],
[16.],
[19.],
[34.],
[39.],
[24.],
[32.],
[25.],
[39.],
[54.],
[36.],
[24.],
[18.],
[47.],
[60.],
[22.]

[22.],
[24.],
[35.],
[52.],
[47.],
[24.],
[37.],
[36.],
[24.],
[49.],
[24.],
[49.],
[24.],
[24.],
[24.],
[44.],
[35.],
[36.],
[30.],
[27.],
[22.],
[40.],
[39.],
[24.],
[24.],
[24.],
[35.],
[24.],
[34.],
[26.],
[4.],
[26.],
[27.],
[42.],
[20.],
[21.],
[21.],
[61.],
[57.],
[21.],
[26.],
[24.],
[80.],
[51.],
[32.],
[24.],
[9.],
[28.],
[32.],
[31.],
[41.],
[24.],
[20.],
[24.],
[2.],
[24.],
[0.75],
[48.],
[19.],
[56.],
[24.],
[23.],
[24.],
[18.],
[21.],
[24.],
[18.],
[24.],
[24.],
[32.],
[23.],
[58.],
[50.],

[50.],
[40.],
[47.],
[36.],
[20.],
[32.],
[25.],
[24.],
[43.],
[24.],
[40.],
[31.],
[70.],
[31.],
[24.],
[18.],
[24.5],
[18.],
[43.],
[36.],
[24.],
[27.],
[20.],
[14.],
[60.],
[25.],
[14.],
[19.],
[18.],
[15.],
[31.],
[4.],
[24.],
[25.],
[60.],
[52.],
[44.],
[24.],
[49.],
[42.],
[18.],
[35.],
[18.],
[25.],
[26.],
[39.],
[45.],
[42.],
[22.],
[24.],
[24.],
[24.],
[48.],
[29.],
[52.],
[19.],
[38.],
[27.],
[24.],
[33.],
[6.],
[17.],
[34.],
[50.],
[27.],
[20.],
[30.],
[24.],
[25.],
[25.],
[29.],
[11.],
[24.]

[24.],
[23.],
[23.],
[28.5],
[48.],
[35.],
[24.],
[24.],
[24.],
[36.],
[21.],
[24.],
[31.],
[70.],
[16.],
[30.],
[19.],
[31.],
[4.],
[6.],
[33.],
[23.],
[48.],
[0.67],
[28.],
[18.],
[34.],
[33.],
[24.],
[41.],
[20.],
[36.],
[16.],
[51.],
[24.],
[30.5],
[24.],
[32.],
[24.],
[48.],
[57.],
[24.],
[54.],
[18.],
[24.],
[5.],
[24.],
[43.],
[13.],
[17.],
[29.],
[24.],
[25.],
[25.],
[18.],
[8.],
[1.],
[46.],
[24.],
[16.],
[24.],
[24.],
[25.],
[39.],
[49.],
[31.],
[30.],
[30.],
[34.],
[31.],
[11.],
[0.42],
[27.],

[27.],
[31.],
[39.],
[18.],
[39.],
[33.],
[26.],
[39.],
[35.],
[6.],
[30.5],
[24.],
[23.],
[31.],
[43.],
[10.],
[52.],
[27.],
[38.],
[27.],
[2.],
[24.],
[24.],
[1.],
[24.],
[62.],
[15.],
[0.83],
[24.],
[23.],
[18.],
[39.],
[21.],
[24.],
[32.],
[24.],
[20.],
[16.],
[30.],
[34.5],
[17.],
[42.],
[24.],
[35.],
[28.],
[24.],
[4.],
[74.],
[9.],
[16.],
[44.],
[18.],
[45.],
[51.],
[24.],
[24.],
[41.],
[21.],
[48.],
[24.],
[24.],
[42.],
[27.],
[31.],
[24.],
[4.],
[26.],
[47.],
[33.],
[47.],
[28.],
[15.],
[20.],

```
       [20.  ],
       [19.  ],
       [24.  ],
       [56.  ],
       [25.  ],
       [33.  ],
       [22.  ],
       [28.  ],
       [25.  ],
       [39.  ],
       [27.  ],
       [19.  ],
       [24.  ],
       [26.  ],
       [32.  ]])
```

In [41]:

```
titanic[["age"]].mode()
```

Out[41]:

	age
0	24.0

In [42]:

```
si = SimpleImputer(strategy= "constant", fill_value = -1)
si.fit_transform(titanic[["age"]])
```

Out[42]:

```
array([[22.  ],
       [38.  ],
       [26.  ],
       [35.  ],
       [35.  ],
       [-1.  ],
       [54.  ],
       [ 2.  ],
       [27.  ],
       [14.  ],
       [ 4.  ],
       [58.  ],
       [20.  ],
       [39.  ],
       [14.  ],
       [55.  ],
       [ 2.  ],
       [-1.  ],
       [31.  ],
       [-1.  ],
       [35.  ],
       [34.  ],
       [15.  ],
       [28.  ],
       [ 8.  ],
       [38.  ],
       [-1.  ],
       [19.  ],
       [-1.  ],
       [-1.  ],
       [40.  ],
       [-1.  ],
       [-1.  ],
       [66.  ],
       [28.  ],
       [42.  ],
       [-1.  ],
       [21.  ],
       [18.  ],
       [14.  ],
```

[14.],
[40.],
[27.],
[-1.],
[3.],
[19.],
[-1.],
[-1.],
[-1.],
[-1.],
[18.],
[7.],
[21.],
[49.],
[29.],
[65.],
[-1.],
[21.],
[28.5],
[5.],
[11.],
[22.],
[38.],
[45.],
[4.],
[-1.],
[-1.],
[29.],
[19.],
[17.],
[26.],
[32.],
[16.],
[21.],
[26.],
[32.],
[25.],
[-1.],
[-1.],
[0.83],
[30.],
[22.],
[29.],
[-1.],
[28.],
[17.],
[33.],
[16.],
[-1.],
[23.],
[24.],
[29.],
[20.],
[46.],
[26.],
[59.],
[-1.],
[71.],
[23.],
[34.],
[34.],
[28.],
[-1.],
[21.],
[33.],
[37.],
[28.],
[21.],
[-1.],
[38.],
[-1.],
[47.],
[14.5],

```
[14.5 ],
[22.  ],
[20.  ],
[17.  ],
[21.  ],
[70.5 ],
[29.  ],
[24.  ],
[ 2.  ],
[21.  ],
[-1.  ],
[32.5 ],
[32.5 ],
[54.  ],
[12.  ],
[-1.  ],
[24.  ],
[-1.  ],
[45.  ],
[33.  ],
[20.  ],
[47.  ],
[29.  ],
[25.  ],
[23.  ],
[19.  ],
[37.  ],
[16.  ],
[24.  ],
[-1.  ],
[22.  ],
[24.  ],
[19.  ],
[18.  ],
[19.  ],
[27.  ],
[ 9.  ],
[36.5 ],
[42.  ],
[51.  ],
[22.  ],
[55.5 ],
[40.5 ],
[-1.  ],
[51.  ],
[16.  ],
[30.  ],
[-1.  ],
[-1.  ],
[44.  ],
[40.  ],
[26.  ],
[17.  ],
[ 1.  ],
[ 9.  ],
[-1.  ],
[45.  ],
[-1.  ],
[28.  ],
[61.  ],
[ 4.  ],
[ 1.  ],
[21.  ],
[56.  ],
[18.  ],
[-1.  ],
[50.  ],
[30.  ],
[36.  ],
[-1.  ],
[-1.  ],
[ 9.  ],
```

[1.],
[4.],
[-1.],
[-1.],
[45.],
[40.],
[36.],
[32.],
[19.],
[19.],
[3.],
[44.],
[58.],
[-1.],
[42.],
[-1.],
[24.],
[28.],
[-1.],
[34.],
[45.5],
[18.],
[2.],
[32.],
[26.],
[16.],
[40.],
[24.],
[35.],
[22.],
[30.],
[-1.],
[31.],
[27.],
[42.],
[32.],
[30.],
[16.],
[27.],
[51.],
[-1.],
[38.],
[22.],
[19.],
[20.5],
[18.],
[-1.],
[35.],
[29.],
[59.],
[5.],
[24.],
[-1.],
[44.],
[8.],
[19.],
[33.],
[-1.],
[-1.],
[29.],
[22.],
[30.],
[44.],
[25.],
[24.],
[37.],
[54.],
[-1.],
[29.],
[62.],
[30.],
[41.],

[29.],
[-1.],
[30.],
[35.],
[50.],
[-1.],
[3.],
[52.],
[40.],
[-1.],
[36.],
[16.],
[25.],
[58.],
[35.],
[-1.],
[25.],
[41.],
[37.],
[-1.],
[63.],
[45.],
[-1.],
[7.],
[35.],
[65.],
[28.],
[16.],
[19.],
[-1.],
[33.],
[30.],
[22.],
[42.],
[22.],
[26.],
[19.],
[36.],
[24.],
[24.],
[-1.],
[23.5],
[2.],
[-1.],
[50.],
[-1.],
[-1.],
[19.],
[-1.],
[-1.],
[0.92],
[-1.],
[17.],
[30.],
[30.],
[24.],
[18.],
[26.],
[28.],
[43.],
[26.],
[24.],
[54.],
[31.],
[40.],
[22.],
[27.],
[30.],
[22.],
[-1.],
[36.],
[61.],

[36.],
[31.],
[16.],
[-1.],
[45.5],
[38.],
[16.],
[-1.],
[-1.],
[29.],
[41.],
[45.],
[45.],
[2.],
[24.],
[28.],
[25.],
[36.],
[24.],
[40.],
[-1.],
[3.],
[42.],
[23.],
[-1.],
[15.],
[25.],
[-1.],
[28.],
[22.],
[38.],
[-1.],
[-1.],
[40.],
[29.],
[45.],
[35.],
[-1.],
[30.],
[60.],
[-1.],
[-1.],
[24.],
[25.],
[18.],
[19.],
[22.],
[3.],
[-1.],
[22.],
[27.],
[20.],
[19.],
[42.],
[1.],
[32.],
[35.],
[-1.],
[18.],
[1.],
[36.],
[-1.],
[17.],
[36.],
[21.],
[28.],
[23.],
[24.],
[22.],
[31.],
[46.],
[23.],

[28.],
[39.],
[26.],
[21.],
[28.],
[20.],
[34.],
[51.],
[3.],
[21.],
[-1.],
[-1.],
[-1.],
[33.],
[-1.],
[44.],
[-1.],
[34.],
[18.],
[30.],
[10.],
[-1.],
[21.],
[29.],
[28.],
[18.],
[-1.],
[28.],
[19.],
[-1.],
[32.],
[28.],
[-1.],
[42.],
[17.],
[50.],
[14.],
[21.],
[24.],
[64.],
[31.],
[45.],
[20.],
[25.],
[28.],
[-1.],
[4.],
[13.],
[34.],
[5.],
[52.],
[36.],
[-1.],
[30.],
[49.],
[-1.],
[29.],
[65.],
[-1.],
[50.],
[-1.],
[48.],
[34.],
[47.],
[48.],
[-1.],
[38.],
[-1.],
[56.],
[-1.],
[0.75],
[-1.],

[38.],
[33.],
[23.],
[22.],
[-1.],
[34.],
[29.],
[22.],
[2.],
[9.],
[-1.],
[50.],
[63.],
[25.],
[-1.],
[35.],
[58.],
[30.],
[9.],
[-1.],
[21.],
[55.],
[71.],
[21.],
[-1.],
[54.],
[-1.],
[25.],
[24.],
[17.],
[21.],
[-1.],
[37.],
[16.],
[18.],
[33.],
[-1.],
[28.],
[26.],
[29.],
[-1.],
[36.],
[54.],
[24.],
[47.],
[34.],
[-1.],
[36.],
[32.],
[30.],
[22.],
[-1.],
[44.],
[-1.],
[40.5],
[50.],
[-1.],
[39.],
[23.],
[2.],
[-1.],
[17.],
[-1.],
[30.],
[7.],
[45.],
[30.],
[-1.],
[22.],
[36.],
[9.],
[11.],

[32.],
[50.],
[64.],
[19.],
[-1.],
[33.],
[8.],
[17.],
[27.],
[-1.],
[22.],
[22.],
[62.],
[48.],
[-1.],
[39.],
[36.],
[-1.],
[40.],
[28.],
[-1.],
[-1.],
[24.],
[19.],
[29.],
[-1.],
[32.],
[62.],
[53.],
[36.],
[-1.],
[16.],
[19.],
[34.],
[39.],
[-1.],
[32.],
[25.],
[39.],
[54.],
[36.],
[-1.],
[18.],
[47.],
[60.],
[22.],
[-1.],
[35.],
[52.],
[47.],
[-1.],
[37.],
[36.],
[-1.],
[49.],
[-1.],
[49.],
[24.],
[-1.],
[-1.],
[44.],
[35.],
[36.],
[30.],
[27.],
[22.],
[40.],
[39.],
[-1.],
[-1.],
[-1.],
[35.],

[24.],
[34.],
[26.],
[4.],
[26.],
[27.],
[42.],
[20.],
[21.],
[21.],
[61.],
[57.],
[21.],
[26.],
[-1.],
[80.],
[51.],
[32.],
[-1.],
[9.],
[28.],
[32.],
[31.],
[41.],
[-1.],
[20.],
[24.],
[2.],
[-1.],
[0.75],
[48.],
[19.],
[56.],
[-1.],
[23.],
[-1.],
[18.],
[21.],
[-1.],
[18.],
[24.],
[-1.],
[32.],
[23.],
[58.],
[50.],
[40.],
[47.],
[36.],
[20.],
[32.],
[25.],
[-1.],
[43.],
[-1.],
[40.],
[31.],
[70.],
[31.],
[-1.],
[18.],
[24.5],
[18.],
[43.],
[36.],
[-1.],
[27.],
[20.],
[14.],
[60.],
[25.],
[14.],

[19.],
[18.],
[15.],
[31.],
[4.],
[-1.],
[25.],
[60.],
[52.],
[44.],
[-1.],
[49.],
[42.],
[18.],
[35.],
[18.],
[25.],
[26.],
[39.],
[45.],
[42.],
[22.],
[-1.],
[24.],
[-1.],
[48.],
[29.],
[52.],
[19.],
[38.],
[27.],
[-1.],
[33.],
[6.],
[17.],
[34.],
[50.],
[27.],
[20.],
[30.],
[-1.],
[25.],
[25.],
[29.],
[11.],
[-1.],
[23.],
[23.],
[28.5],
[48.],
[35.],
[-1.],
[-1.],
[-1.],
[36.],
[21.],
[24.],
[31.],
[70.],
[16.],
[30.],
[19.],
[31.],
[4.],
[6.],
[33.],
[23.],
[48.],
[0.67],
[28.],
[18.],
[34.],

```
[33.  ],
[-1.  ],
[41.  ],
[20.  ],
[36.  ],
[16.  ],
[51.  ],
[-1.  ],
[30.5 ],
[-1.  ],
[32.  ],
[24.  ],
[48.  ],
[57.  ],
[-1.  ],
[54.  ],
[18.  ],
[-1.  ],
[ 5.  ],
[-1.  ],
[43.  ],
[13.  ],
[17.  ],
[29.  ],
[-1.  ],
[25.  ],
[25.  ],
[18.  ],
[ 8.  ],
[ 1.  ],
[46.  ],
[-1.  ],
[16.  ],
[-1.  ],
[-1.  ],
[25.  ],
[39.  ],
[49.  ],
[31.  ],
[30.  ],
[30.  ],
[34.  ],
[31.  ],
[11.  ],
[ 0.42],
[27.  ],
[31.  ],
[39.  ],
[18.  ],
[39.  ],
[33.  ],
[26.  ],
[39.  ],
[35.  ],
[ 6.  ],
[30.5 ],
[-1.  ],
[23.  ],
[31.  ],
[43.  ],
[10.  ],
[52.  ],
[27.  ],
[38.  ],
[27.  ],
[ 2.  ],
[-1.  ],
[-1.  ],
[ 1.  ],
[-1.  ],
[62.  ],
[15.  ],
```

```
[ 0.83],
[-1.  ],
[23.  ],
[18.  ],
[39.  ],
[21.  ],
[-1.  ],
[32.  ],
[-1.  ],
[20.  ],
[16.  ],
[30.  ],
[34.5 ],
[17.  ],
[42.  ],
[-1.  ],
[35.  ],
[28.  ],
[-1.  ],
[ 4.  ],
[74.  ],
[ 9.  ],
[16.  ],
[44.  ],
[18.  ],
[45.  ],
[51.  ],
[24.  ],
[-1.  ],
[41.  ],
[21.  ],
[48.  ],
[-1.  ],
[24.  ],
[42.  ],
[27.  ],
[31.  ],
[-1.  ],
[ 4.  ],
[26.  ],
[47.  ],
[33.  ],
[47.  ],
[28.  ],
[15.  ],
[20.  ],
[19.  ],
[-1.  ],
[56.  ],
[25.  ],
[33.  ],
[22.  ],
[28.  ],
[25.  ],
[39.  ],
[27.  ],
[19.  ],
[-1.  ],
[26.  ],
[32.  ]])
```

1. Standardizing Data

scaling

example: 20m, 30mm, 75cm, 1km lengths ---- Total length -- 20+30+75+1 =

single unit m or km or mm --- adding --- accurate output

weights... kg, g, mg, tones,

In [ ]:

```
```

In [43]:

```
adv.head()
```

Out[43]:

	TV	radio	newspaper	sales
0	**230.1**	**37.8**	**69.2**	**22.1**
1	**44.5**	**39.3**	**45.1**	**10.4**
2	**17.2**	**45.9**	**69.3**	**9.3**
3	**151.5**	**41.3**	**58.5**	**18.5**
4	**180.8**	**10.8**	**58.4**	**12.9**

In [44]:

```
adv.isnull().sum()
```

Out[44]:

```
TV           0
radio        0
newspaper    0
sales        0
dtype: int64
```

In [45]:

```
adv.mean()    # mean becomes 0
```

Out[45]:

```
TV           147.0425
radio         23.2640
newspaper     30.5540
sales         14.0225
dtype: float64
```

In [46]:

```
adv.std()   # std becomes 1     standed format
```

Out[46]:

```
TV           85.854236
radio        14.846809
newspaper    21.778621
sales         5.217457
dtype: float64
```

In [47]:

```
from sklearn.preprocessing import scale
```

In [48]:

```
adv_new = scale(adv)
adv_new
```

Out[48]:

```
array([[ 9.69852266e-01,  9.81522472e-01,  1.77894547e+00,
         1.55205313e+00],
       [-1.19737623e+00,  1.08280781e+00,  6.69578760e-01,
```

```
 -6.96046111e-01],
[-1.51615499e+00,  1.52846331e+00,  1.78354865e+00,
 -9.07405869e-01],
[ 5.20496822e-02,  1.21785493e+00,  1.28640506e+00,
  8.60330287e-01],
[ 3.94182198e-01, -8.41613655e-01,  1.28180188e+00,
 -2.15683025e-01],
[-1.61540845e+00,  1.73103399e+00,  2.04592999e+00,
 -1.31091086e+00],
[-1.04557682e+00,  6.43904671e-01, -3.24708413e-01,
 -4.27042783e-01],
[-3.13436589e-01, -2.47406325e-01, -8.72486994e-01,
 -1.58039455e-01],
[-1.61657614e+00, -1.42906863e+00, -1.36042422e+00,
 -1.77205942e+00],
[ 6.16042873e-01, -1.39530685e+00, -4.30581584e-01,
 -6.57617064e-01],
[-9.45155670e-01, -1.17923146e+00, -2.92486143e-01,
 -1.04190753e+00],
[ 7.90028350e-01,  4.96973404e-02, -1.22232878e+00,
  6.48970529e-01],
[-1.43908760e+00,  7.99208859e-01,  1.62704048e+00,
 -9.26620392e-01],
[-5.78501712e-01, -1.05768905e+00, -1.07502697e+00,
 -8.30547775e-01],
[ 6.66253447e-01,  6.50657027e-01,  7.11007392e-01,
  9.56402904e-01],
[ 5.64664612e-01,  1.65000572e+00,  1.02862691e+00,
  1.60969670e+00],
[-9.25304978e-01,  9.00494200e-01,  3.84117072e+00,
 -2.92541119e-01],
[ 1.56887609e+00,  1.10306488e+00,  1.16211917e+00,
  1.99398717e+00],
[-9.08957349e-01, -1.86635121e-01, -5.64073843e-01,
 -5.23115400e-01],
[ 3.00679600e-03,  4.29449843e-02, -5.27248393e-01,
  1.10963873e-01],
[ 8.33232798e-01,  2.99534513e-01,  1.05164281e+00,
  7.64257669e-01],
[ 1.05509347e+00, -1.22649795e+00, -3.24708413e-01,
 -2.92541119e-01],
[-1.56286250e+00, -4.97243498e-01,  8.76721921e-01,
 -1.61834324e+00],
[ 9.48833887e-01, -4.29719938e-01, -2.00422516e-01,
  2.83894584e-01],
[-9.89527805e-01, -7.20071247e-01, -5.64073843e-01,
 -8.30547775e-01],
[ 1.35285385e+00, -1.33453565e+00, -5.08835667e-01,
 -3.88613736e-01],
[-4.83714657e-02,  4.07572210e-01, -8.26455181e-01,
  1.87821967e-01],
[ 1.08662104e+00, -4.43224650e-01, -3.52327501e-01,
  3.60752677e-01],
[ 1.18820988e+00,  2.59020377e-01, -3.52327501e-01,
  9.37188380e-01],
[-8.92609721e-01, -4.90491142e-01,  4.71641962e-01,
 -6.76831588e-01],
[ 1.70316018e+00,  3.40048650e-01,  5.82118314e-01,
  1.41755147e+00],
[-3.98677796e-01, -3.95958157e-01,  3.70371972e-01,
 -4.07828260e-01],
[-5.82004775e-01, -1.46958277e+00, -2.55016247e-02,
 -8.49762299e-01],
[ 1.38438142e+00, -2.20396901e-01, -1.39264649e+00,
  6.48970529e-01],
[-5.99520091e-01, -1.47633512e+00, -1.06582061e+00,
 -8.68976822e-01],
[ 1.67747105e+00, -1.29402151e+00, -1.01518562e+00,
 -2.34897549e-01],
[ 1.39956136e+00,  1.38666383e+00, -1.17629696e+00,
  2.18613240e+00],
[-8.44734522e-01,  1.76479577e+00,  6.97197848e-01,
```

```
  1.30178396e-01],
[-1.21372386e+00,  2.32010953e-01,  2.09260624e-01,
 -7.53689682e-01],
[ 9.45330823e-01,  9.74770116e-01,  6.65620024e-02,
  1.43676599e+00],
[ 6.47570443e-01, -6.50927121e-02,  4.81492770e-02,
  4.95254341e-01],
[ 3.49810063e-01,  6.84418807e-01,  3.74975153e-01,
  5.91326959e-01],
[ 1.71133400e+00,  2.99534513e-01, -1.32359877e+00,
  1.28304980e+00],
[ 6.98948705e-01, -1.00367020e+00, -1.91216154e-01,
 -2.15683025e-01],
[-1.42390765e+00,  1.64487393e-01,  5.86721496e-01,
 -1.06112206e+00],
[ 3.27623995e-01, -5.15880000e-02,  4.35460956e-02,
  1.68607443e-01],
[-6.69581357e-01, -9.02384859e-01,  2.36879713e-01,
 -6.57617064e-01],
[ 1.08428567e+00,  1.23135965e+00, -5.54867481e-01,
  1.76341289e+00],
[ 9.35989321e-01, -5.03995854e-01,  8.90531465e-01,
  1.49392920e-01],
[-9.35814168e-01, -7.80842451e-01,  2.87514708e-01,
 -8.30547775e-01],
[ 6.16042873e-01, -1.36154507e+00,  1.86244718e-01,
 -5.03900877e-01],
[-5.44638766e-01, -9.22641928e-01, -1.24074150e+00,
 -6.38402541e-01],
[ 8.09879042e-01,  1.24486436e+00,  4.16403786e-01,
  1.64812575e+00],
[ 4.15200577e-01,  1.54872038e+00,  1.29561142e+00,
  1.37912242e+00],
[ 1.35051848e+00,  3.73810430e-01, -6.74550196e-01,
  1.18697718e+00],
[ 6.05533683e-01,  1.76479577e+00,  1.35545278e+00,
  1.85948550e+00],
[-1.63175608e+00,  3.26543937e-01,  4.99261050e-01,
 -1.63755776e+00],
[-1.26606546e-01, -2.74415749e-01, -6.42327927e-01,
 -1.58039455e-01],
[ 7.44488528e-01,  1.77830048e+00,  3.28943340e-01,
  1.87870003e+00],
[ 7.43320840e-01,  4.21076922e-01, -9.78360166e-01,
  8.41115763e-01],
[-1.09228433e+00, -1.43582099e+00, -4.21375221e-01,
 -1.13798015e+00],
[ 1.33417085e+00,  1.31238792e+00,  1.11148417e+00,
  1.95555812e+00],
[ 1.07727954e+00, -5.24252922e-01, -1.49787521e-01,
  3.22323631e-01],
[-5.17781948e-01,  4.27829278e-01, -1.01978880e+00,
 -4.32326777e-03],
[-1.86158622e-01,  1.31914027e+00, -7.61366196e-02,
  7.64257669e-01],
[-9.11292725e-01, -9.42898996e-01, -1.36502740e+00,
 -9.07405869e-01],
[-1.34917564e+00,  9.02114765e-02, -1.30518604e+00,
 -8.68976822e-01],
[-9.04082253e-02, -5.91776482e-01, -9.36931533e-01,
 -1.19610408e-01],
[ 1.05509347e+00,  2.86029801e-01, -9.00106083e-01,
  9.37188380e-01],
[ 8.14549794e-01,  1.39341619e+00, -1.54390703e-01,
  1.59048218e+00],
[ 6.07869059e-01,  4.95352838e-01,  3.74975153e-01,
  8.21901240e-01],
[-4.34876116e-01, -6.05281194e-01,  5.27524584e-02,
 -3.11755643e-01],
[-1.40405696e+00,  6.57409383e-01, -5.18042030e-01,
 -1.00347849e+00],
[-2.06009314e-01, -1.18598381e+00,  3.43397329e-02,
```

```
 -5.80758971e-01],
[ 7.74848409e-01,  9.02114765e-02, -8.03439274e-01,
  5.72112435e-01],
[-1.51965805e+00,  1.37991148e+00,  2.70878810e+00,
 -1.02269301e+00],
[-1.39588315e+00, -1.46283041e+00, -4.53597491e-01,
 -1.36855443e+00],
[-3.09933525e-01,  3.53553362e-01, -7.52804279e-01,
  3.41057791e-02],
[-1.65394214e+00,  4.48086346e-01, -9.73756984e-01,
 -1.67598681e+00],
[-3.62479475e-01, -1.05093669e+00, -3.43121138e-01,
 -5.80758971e-01],
[-8.24883830e-01,  2.32010953e-01, -3.79946589e-01,
 -4.27042783e-01],
[ 1.08311798e+00, -1.29402151e+00,  2.92117889e-01,
 -3.30970166e-01],
[-8.37728396e-01, -2.00139833e-01,  8.95779092e-02,
 -5.23115400e-01],
[-9.18298852e-01,  1.43393033e+00,  2.32276531e-01,
 -8.11813615e-02],
[ 7.76016097e-01,  1.33264499e+00,  1.49419267e-01,
  1.47519504e+00],
[ 5.38975481e-01, -3.28434597e-01,  1.61783412e+00,
  2.26251013e-01],
[-8.26051518e-01,  2.86029801e-01, -6.69947015e-01,
 -3.88613736e-01],
[-4.24366926e-01,  1.17058844e+00,  1.50275459e+00,
  3.79967201e-01],
[-6.85928986e-01,  1.50982681e-01,  1.97227908e+00,
 -2.15683025e-01],
[-4.34876116e-01,  1.65675807e+00,  9.59579186e-01,
  5.14468865e-01],
[-1.48792614e-01, -1.24000266e+00, -9.78360166e-01,
 -5.42329924e-01],
[-1.38303858e+00, -1.46958277e+00,  1.12593816e-01,
 -1.29169634e+00],
[ 8.25058983e-01,  6.91171163e-01,  1.30942097e+00,
  1.03326100e+00],
[ 1.21273132e+00,  8.93741844e-01,  1.92164409e+00,
  1.57126765e+00],
[-4.62900623e-01, -6.25538262e-01, -9.04709264e-01,
 -4.84686354e-01],
[ 1.89836839e-01,  5.62876398e-01,  1.02862691e+00,
  5.52897912e-01],
[ 5.90353742e-01, -1.33453565e+00, -1.13486833e+00,
 -4.46257307e-01],
[ 4.42057396e-01, -1.52873340e-01, -3.93756133e-01,
  2.83894584e-01],
[ 1.66579418e+00,  1.28537849e+00,  9.50372823e-01,
  2.18613240e+00],
[-1.38283424e-01,  1.24486436e+00,  7.06404211e-01,
  6.10541482e-01],
[ 8.79940308e-01, -1.28051680e+00,  8.85928284e-01,
 -4.46257307e-01],
[ 1.74402926e+00,  8.80237132e-01,  3.23815396e+00,
  1.87870003e+00],
[ 1.55486384e+00, -8.88880147e-01, -4.21375221e-01,
  1.49392920e-01],
[ 4.77088029e-01, -4.09462869e-01, -5.82486569e-01,
  1.30178396e-01],
[ 1.06443498e+00,  7.45190011e-01, -1.16248742e+00,
  1.28304980e+00],
[-1.06755854e-01,  1.56222509e+00,  1.30942097e+00,
  9.94831951e-01],
[-1.42507534e+00, -8.28108943e-01, -3.93111688e-02,
 -1.31091086e+00],
[-6.61407543e-01, -1.55061104e+00, -3.38517957e-01,
 -1.02269301e+00],
[-1.56403019e+00, -1.54385868e+00, -2.28041604e-01,
 -1.67598681e+00],
[ 1.26527727e+00,  2.45515665e-01, -1.15328106e+00,
```

```
  1.11011909e+00],
[ 9.19641692e-01, -1.01717491e+00,  1.19434143e+00,
 -1.19610408e-01],
[ 1.10530405e+00,  9.95027184e-01, -3.38517957e-01,
  1.49440956e+00],
[ 3.34630122e-01, -5.31005278e-01, -1.29597968e+00,
  1.48912557e-02],
[ 7.30476274e-01, -1.79882765e-01, -9.13915627e-01,
  3.60752677e-01],
[-8.03865450e-01,  1.58923451e+00,  1.81641536e-01,
  1.10963873e-01],
[-8.40063771e-01,  7.92456503e-01,  1.01942054e+00,
 -2.73326596e-01],
[-9.15759131e-02, -6.05281194e-01, -2.28041604e-01,
 -3.50184689e-01],
[-8.24883830e-01, -1.51684926e+00, -7.25185191e-01,
 -8.88191346e-01],
[-2.49213762e-01,  9.20751268e-01,  2.23926360e+00,
  3.60752677e-01],
[-1.49046586e+00, -4.90491142e-01, -3.79946589e-01,
 -1.42619800e+00],
[-6.70544700e-02,  2.38763309e-01,  7.20213755e-01,
  2.83894584e-01],
[-1.49747198e+00, -1.05606848e-01,  9.13547372e-01,
 -1.34933991e+00],
[ 8.98623313e-01, -1.40881156e+00, -6.88359740e-01,
 -4.65471830e-01],
[-2.79573643e-01,  7.65447079e-01, -8.35661544e-01,
  2.26251013e-01],
[ 9.62846140e-01,  6.10142891e-01,  2.00910454e+00,
  1.09090457e+00],
[-6.98773552e-01, -7.74090095e-01, -2.14232060e-01,
 -6.57617064e-01],
[-1.62591764e+00,  1.05579839e+00,  9.22753735e-01,
 -1.42619800e+00],
[-7.80511695e-01, -1.57086811e+00, -9.82963347e-01,
 -1.00347849e+00],
[ 8.55418865e-01,  1.73778635e+00, -1.25915423e+00,
  2.05163074e+00],
[-1.02105537e+00, -7.60585383e-01,  5.77515133e-01,
 -8.30547775e-01],
[-1.70882347e+00,  1.10306488e+00, -1.00597925e+00,
 -2.38692417e+00],
[ 1.37971067e+00, -1.37504978e+00,  5.72911952e-01,
 -2.54112072e-01],
[-1.61891151e+00,  2.65772733e-01, -1.30978922e+00,
 -1.59912871e+00],
[ 8.49580427e-01,  6.91171163e-01,  6.69578760e-01,
  1.07169004e+00],
[-1.28612050e+00,  1.03554132e+00,  1.61323094e+00,
 -6.19188018e-01],
[-1.15300409e+00,  1.60273923e+00, -1.01518562e+00,
 -4.65471830e-01],
[-1.41806922e+00,  1.06255074e+00, -9.78360166e-01,
 -8.68976822e-01],
[ 1.47896413e+00,  3.80562786e-01,  1.34164324e+00,
  1.30226433e+00],
[-1.21489154e+00,  1.77992105e-01, -4.62803854e-01,
 -8.49762299e-01],
[ 4.42057396e-01,  1.39341619e+00, -1.32820195e+00,
  1.28304980e+00],
[-8.59914463e-01, -4.22967582e-01, -8.12645637e-01,
 -5.99973494e-01],
[ 5.44813920e-01,  8.19465927e-01,  2.07354907e+00,
  9.94831951e-01],
[ 8.57754241e-01,  6.70914095e-01,  3.38149702e-01,
  1.16776266e+00],
[-4.95595880e-01, -1.18598381e+00,  1.77038355e-01,
 -6.96046111e-01],
[-5.93681653e-01, -5.71519414e-01,  3.84181516e-01,
 -5.03900877e-01],
[-7.87313476e-02, -1.44257334e+00, -9.92169710e-01,
```

```
 -7.15260635e-01],
[ 1.08662104e+00, -1.07794612e+00, -1.00597925e+00,
 -1.58039455e-01],
[ 1.12281936e+00,  1.73778635e+00,  6.32753309e-01,
  2.18613240e+00],
[-1.27327593e+00,  1.15033137e+00, -8.58677450e-01,
 -5.99973494e-01],
[-1.19504085e+00,  1.71239749e-01, -4.58200672e-01,
 -7.53689682e-01],
[ 1.56070228e+00, -6.32290618e-01,  2.96721070e-01,
  3.99181724e-01],
[-3.04095087e-01, -1.00367020e+00,  8.35293289e-01,
 -4.65471830e-01],
[ 5.90353742e-01,  2.43084817e-03, -7.52804279e-01,
  4.95254341e-01],
[ 2.83251860e-01,  1.10981724e+00,  3.28943340e-01,
  9.56402904e-01],
[ 4.75920341e-01, -1.46120984e-01, -9.69153803e-01,
  3.03109107e-01],
[-1.66912209e+00, -7.87594807e-01, -1.14407469e+00,
 -2.07949180e+00],
[-6.20538471e-01,  1.36640677e+00,  9.18150553e-01,
  2.45465537e-01],
[ 3.21989902e-02, -1.48308748e+00, -2.87882962e-01,
 -7.53689682e-01],
[-1.58037782e+00,  9.20751268e-01,  6.74181942e-01,
 -1.29169634e+00],
[-1.79152496e-01, -3.28434597e-01,  1.86244718e-01,
 -2.15683025e-01],
[ 2.97264113e-01, -3.48691665e-01,  6.72064478e-03,
  7.25348259e-02],
[-7.16288868e-01,  8.46475352e-01,  8.62912377e-01,
 -1.38824932e-01],
[ 4.82926468e-01, -3.48691665e-01, -2.28041604e-01,
  1.68607443e-01],
[ 1.92172214e-01,  9.13998912e-01, -1.06582061e+00,
  7.64257669e-01],
[-3.48467222e-01, -5.78271770e-01, -1.15788424e+00,
 -4.07828260e-01],
[ 1.02123053e+00, -1.34128800e+00,  2.49704176e+00,
 -4.07828260e-01],
[-1.50798117e+00,  9.68017760e-01, -4.12168859e-01,
 -1.15719467e+00],
[ 6.97781017e-01, -1.21974559e+00, -5.13438849e-01,
 -3.50184689e-01],
[ 7.98202165e-01,  2.26879163e-02,  1.24497643e+00,
  5.91326959e-01],
[ 1.60273904e+00, -8.55118367e-01, -1.11185242e+00,
  1.87821967e-01],
[-1.13315340e+00, -7.87594807e-01, -5.59470662e-01,
 -1.08033658e+00],
[ 2.03849092e-01, -1.59625696e-01,  7.75451931e-01,
  9.17493494e-02],
[-1.48813048e+00, -2.13644545e-01, -6.23915201e-01,
 -1.23405277e+00],
[ 2.49388915e-01, -1.09145083e+00, -8.17248818e-01,
 -4.46257307e-01],
[ 8.79940308e-01, -1.34128800e+00, -8.03439274e-01,
 -4.84686354e-01],
[ 1.51633014e+00,  1.73103399e+00,  5.17673775e-01,
  2.49356478e+00],
[ 1.18353913e+00,  4.68343414e-01, -4.72010216e-01,
  1.18697718e+00],
[ 2.70407294e-01, -1.04418434e+00,  2.13863806e-01,
 -4.46257307e-01],
[ 1.51399477e+00, -1.41556392e+00, -3.15502050e-01,
 -4.27042783e-01],
[ 2.16693657e-01, -8.95632503e-01, -5.96296113e-01,
 -2.73326596e-01],
[ 1.11601758e-01, -1.39530685e+00, -1.02439198e+00,
 -6.76831588e-01],
[ 8.34400486e-01, -1.20624088e+00, -1.45184340e-01,
```

```
        -3.50184689e-01],
       [-1.06075676e+00, -1.18598381e+00, -3.93111688e-02,
        -1.02269301e+00],
       [ 1.64127273e+00,  1.33264499e+00,  1.89862818e+00,
         2.33984859e+00],
       [ 1.24659427e+00, -1.32616272e-01, -2.55016247e-02,
         6.87399576e-01],
       [ 6.76762637e-01,  1.47444446e+00, -5.04232486e-01,
         1.64812575e+00],
       [-8.80728498e-02, -1.42906863e+00, -1.82009791e-01,
        -7.15260635e-01],
       [ 5.14454038e-01,  3.67058074e-01, -5.68677025e-01,
         6.29756005e-01],
       [ 1.62258973e+00, -6.32290618e-01, -1.23613832e+00,
         3.60752677e-01],
       [-1.49863967e+00, -7.53833027e-01, -3.29311594e-01,
        -1.40698348e+00],
       [-1.25576062e+00,  1.20435022e+00, -1.13947151e+00,
        -6.19188018e-01],
       [-8.35393020e-01, -8.41613655e-01, -1.13026515e+00,
        -7.92118728e-01],
       [-1.51615499e+00, -1.29402151e+00,  4.81492770e-02,
        -1.56069967e+00],
       [ 2.30705910e-01,  1.26512143e+00, -1.24074150e+00,
         1.07169004e+00],
       [ 3.10313024e-02,  8.32970639e-01, -1.13026515e+00,
         6.29756005e-01],
       [-1.27094056e+00, -1.32103093e+00, -7.71217005e-01,
        -1.23405277e+00],
       [-6.17035408e-01, -1.24000266e+00, -1.03359834e+00,
        -8.30547775e-01],
       [ 3.49810063e-01, -9.42898996e-01, -1.11185242e+00,
        -2.34897549e-01],
       [ 1.59456522e+00,  1.26512143e+00,  1.64085003e+00,
         2.20534693e+00],
       [ 9.93206022e-01, -9.90165488e-01, -1.00597925e+00,
        -1.19610408e-01]])
```

In [49]:

```
# 2D array into df
adv_new_df = pd.DataFrame(adv_new, columns = adv.columns)
adv_new_df.head()
```

Out[49]:

	TV	radio	newspaper	sales
0	**0.969852**	**0.981522**	**1.778945**	**1.552053**
1	**-1.197376**	**1.082808**	**0.669579**	**-0.696046**
2	**-1.516155**	**1.528463**	**1.783549**	**-0.907406**
3	**0.052050**	**1.217855**	**1.286405**	**0.860330**
4	**0.394182**	**-0.841614**	**1.281802**	**-0.215683**

In [51]:

```
adv_new_df.mean().round(2)
```

Out[51]:

```
TV           0.0
radio       -0.0
newspaper    0.0
sales       -0.0
dtype: float64
```

In [52]:

```
adv_new_df.std().round(2)
```

Out[52]:

```
TV           1.0
radio        1.0
newspaper    1.0
sales        1.0
dtype: float64
```

In [55]:

```
(adv["TV"][0] - adv["TV"].mean()) / (adv["TV"].std())
```

Out[55]:

```
0.9674245973763037
```

In [56]:

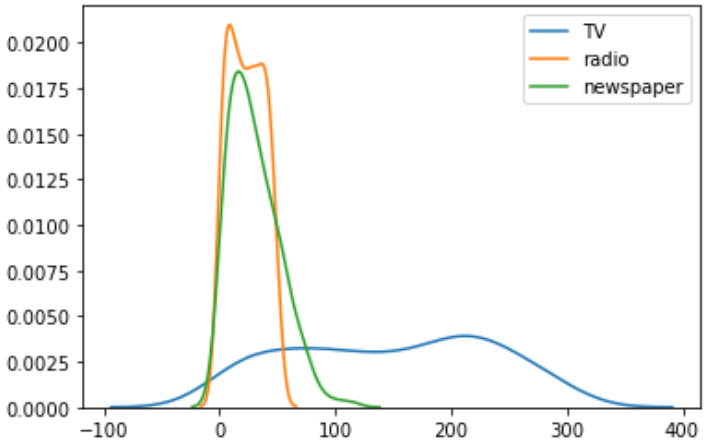
```
import seaborn as sns
```

In [57]:

```
sns.kdeplot(adv["TV"])
sns.kdeplot(adv["radio"])
sns.kdeplot(adv["newspaper"])
```

Out[57]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1939a9c96d0>
```

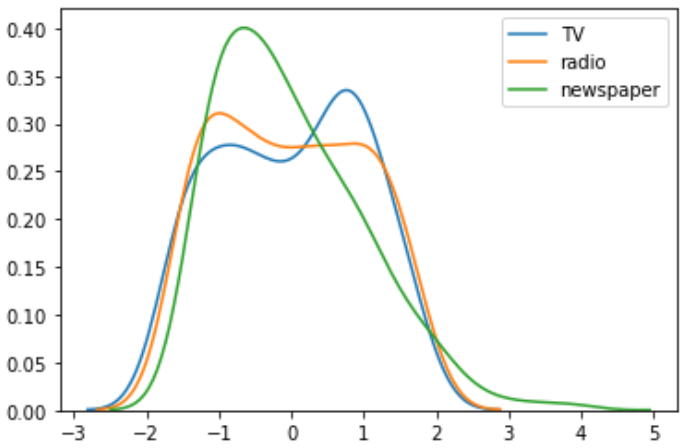


In [58]:

```
sns.kdeplot(adv_new_df["TV"])
sns.kdeplot(adv_new_df["radio"])
sns.kdeplot(adv_new_df["newspaper"])
```

Out[58]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1939aa54580>
```



2. Data Range

In [59]:

```
adv.head()
```

Out[59]:

	TV	radio	newspaper	sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

In [60]:

```
from sklearn.preprocessing import MinMaxScaler
```

In [61]:

```
mnscale = MinMaxScaler()
mnscale
```

Out[61]:

```
MinMaxScaler()
```

In [63]:

```
mnscale = mnscale.fit_transform(adv)
```

In [64]:

```
mnscaled_data = pd.DataFrame(mnscale, columns= adv.columns)
mnscaled_data
```

Out[64]:

	TV	radio	newspaper	sales
0	0.775786	0.762097	0.605981	0.807087
1	0.148123	0.792339	0.394019	0.346457
2	0.055800	0.925403	0.606860	0.303150
3	0.509976	0.832661	0.511873	0.665354
4	0.609063	0.217742	0.510994	0.444882
...	...	...	...	...
195	0.126818	0.074597	0.118734	0.236220
196	0.316199	0.098790	0.068602	0.318898
197	0.596212	0.187500	0.053650	0.440945
198	0.956713	0.846774	0.579595	0.940945
199	0.782550	0.173387	0.073879	0.464567

200 rows × 4 columns

In [67]:

```
adv.min(), mnscaled_data.min()
```

Out[67]:

```
(TV          0.7
```

```
 radio         0.0
 newspaper     0.3
 sales         1.6
 dtype: float64,
 TV            0.0
 radio         0.0
 newspaper     0.0
 sales         0.0
 dtype: float64)
```

In [68]:

```
adv.max(), mnscaled_data.max()
```

Out[68]:

```
(TV           296.4
 radio         49.6
 newspaper    114.0
 sales         27.0
 dtype: float64,
 TV            1.0
 radio         1.0
 newspaper     1.0
 sales         1.0
 dtype: float64)
```

In [70]:

```
(adv["TV"][0] - adv["TV"].min()) / (adv["TV"].max() - adv["TV"].min())
```

Out[70]:

```
0.7757862698681096
```

3. Normalizing Data

In [71]:

```
adv.head()
```

Out[71]:

	TV	radio	newspaper	sales
0	**230.1**	**37.8**	**69.2**	**22.1**
1	**44.5**	**39.3**	**45.1**	**10.4**
2	**17.2**	**45.9**	**69.3**	**9.3**
3	**151.5**	**41.3**	**58.5**	**18.5**
4	**180.8**	**10.8**	**58.4**	**12.9**

In [72]:

```
from sklearn.preprocessing import Normalizer
```

In [73]:

```
norm_data =  Normalizer()
```

In [75]:

```
norm_data = norm_data.fit_transform(adv)
```

In [76]:

```
# 2d into df
```

```
norm_data = pd.DataFrame(norm_data, columns = adv.columns)
norm_data
```

Out[76]:

	TV	radio	newspaper	sales
0	0.942116	0.154767	0.283331	0.090486
1	0.591135	0.522059	0.599106	0.138153
2	0.201426	0.537527	0.811561	0.108911
3	0.898632	0.244974	0.346997	0.109734
4	0.947881	0.056621	0.306174	0.067631
...	...	...	...	...
195	0.920781	0.089186	0.332638	0.183192
196	0.989807	0.051487	0.085111	0.101923
197	0.995383	0.052300	0.035991	0.071983
198	0.960250	0.142209	0.224149	0.086341
199	0.996958	0.036940	0.037370	0.057558

200 rows × 4 columns

4. Robust Scaling

In [77]:

```
adv.head()
```

Out[77]:

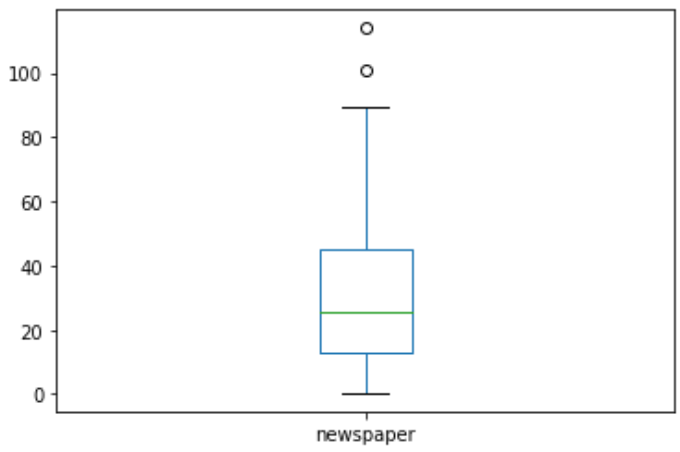
	TV	radio	newspaper	sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

In [78]:

```
adv["newspaper"].plot(kind = "box")
```

Out[78]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1939ac6ad90>
```



In [79]:

In [79]:

```
from sklearn.preprocessing import RobustScaler
```

In [80]:

```
rscale = RobustScaler()
rscale.fit(adv)
```

Out[80]:

```
RobustScaler()
```

In [82]:

```
rscale = rscale.transform(adv)
```

In [83]:

```
rscale = pd.DataFrame(rscale, columns = adv.columns)
rscale
```

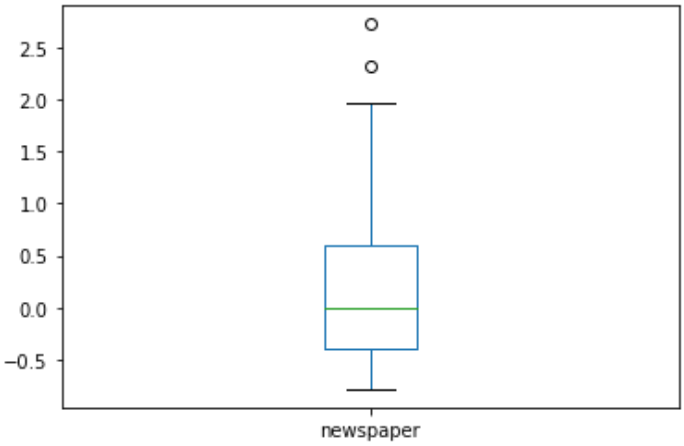
Out[83]:

	TV	radio	newspaper	sales
0	0.556248	0.561205	1.343122	1.309609
1	-0.728626	0.617702	0.598145	-0.355872
2	-0.917619	0.866290	1.346213	-0.512456
3	0.012115	0.693032	1.012365	0.797153
4	0.214953	-0.455744	1.009274	0.000000
...	...	...	...	...
195	-0.772240	-0.723164	-0.369397	-0.754448
196	-0.384562	-0.677966	-0.545595	-0.455516
197	0.188647	-0.512241	-0.598145	-0.014235
198	0.926618	0.719397	1.250386	1.793594
199	0.570093	-0.538606	-0.527048	0.071174

200 rows × 4 columns

In [85]:

```
rscale["newspaper"].plot(kind = "box")
plt.show()
```

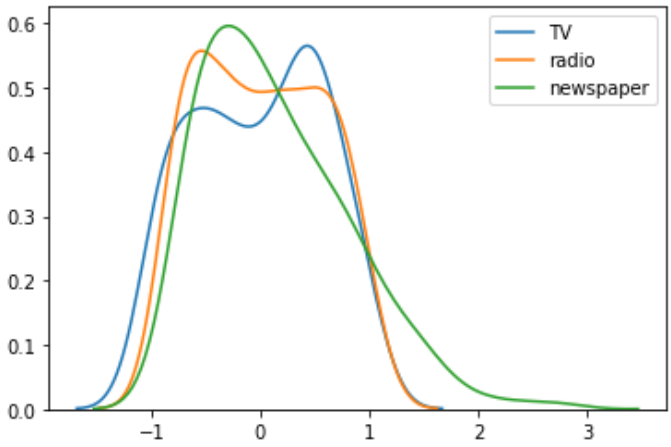


In [86]:

```
sns.kdeplot(rscale["TV"])
sns.kdeplot(rscale["radio"])
sns.kdeplot(rscale["newspaper"])
```

Out[86]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1939a6cf160>
```



Label Encoder

In [ ]:

```
# True and False
# 1 or 0
# used to convert categorial data into numeric type data
```

In [87]:

```
emp.head()
```

Out[87]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services

In [88]:

```
emp["Gender"].value_counts()
```

Out[88]:

```
Female    431
Male      424
Name: Gender, dtype: int64
```

In [89]:

```
from sklearn.preprocessing import LabelEncoder
```

In [90]:

```
Le_data = LabelEncoder()
```

In [92]:

```
Le_data.fit_transform(emp[["Gender"]])
```

```
C:\Users\lavan\anaconda3\lib\site-packages\sklearn\utils\validation.py:73: DataConversion
Warning: A column-vector y was passed when a 1d array was expected. Please change the sha
pe of y to (n_samples, ), for example using ravel().
  return f(**kwargs)
```

```
  return f(**kwargs)
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
~\anaconda3\lib\site-packages\sklearn\preprocessing\_label.py in _encode(values, uniques,
encode, check_unknown)
    112         try:
--> 113             res = _encode_python(values, uniques, encode)
    114         except TypeError:

~\anaconda3\lib\site-packages\sklearn\preprocessing\_label.py in _encode_python(values, u
niques, encode)
     60     if uniques is None:
---> 61         uniques = sorted(set(values))
     62         uniques = np.array(uniques, dtype=values.dtype)

TypeError: '<' not supported between instances of 'str' and 'float'

During handling of the above exception, another exception occurred:

TypeError                                 Traceback (most recent call last)
<ipython-input-92-c5b8c3f778d3> in <module>
----> 1 Le_data.fit_transform(emp[["Gender"]])

~\anaconda3\lib\site-packages\sklearn\preprocessing\_label.py in fit_transform(self, y)
    254         """
    255         y = column_or_1d(y, warn=True)
--> 256         self.classes_, y = _encode(y, encode=True)
    257         return y
    258

~\anaconda3\lib\site-packages\sklearn\preprocessing\_label.py in _encode(values, uniques,
encode, check_unknown)
    115             types = sorted(t.__qualname__
    116                            for t in set(type(v) for v in values))
--> 117             raise TypeError("Encoders require their input to be uniformly "
    118                             f"strings or numbers. Got {types}")
    119         return res

TypeError: Encoders require their input to be uniformly strings or numbers. Got ['float',
'str']
```

In [ ]:

```
# Got ['float', 'str']

# 100/100
```

In [93]:

```
emp["Gender"].isnull().sum()
# NAN -- not a number --- Special type of floating value
```

Out[93]:

```
145
```

In [94]:

```
emp.dropna(inplace = True)
```

In [95]:

```
emp.shape
```

Out[95]:

```
(764, 8)
```

In [96]:

```
emp.isnull().sum()
```

Out[96]:

```
First Name           0
Gender               0
Start Date           0
Last Login Time      0
Salary               0
Bonus %              0
Senior Management    0
Team                 0
dtype: int64
```

In [97]:

```
Le_data.fit_transform(emp[["Gender"]])
```

Out[97]:

```
array([1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1,
       0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0,
       0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0,
       0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0,
       1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0,
       1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1,
       1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0,
       1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1,
       0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0,
       0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1,
       0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0,
       0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0,
       0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0,
       0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0,
       0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0,
       1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1,
       0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1,
       0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0,
       1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0,
       1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1,
       0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0,
       1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0,
       0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0,
       0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0,
       0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0,
       0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1,
       1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0,
       1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0,
       1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1])
```

In [98]:

```
emp["Gender"]
```

Out[98]:

```
0        Male
2      Female
3        Male
4        Male
5        Male
        ...
994      Male
996      Male
997      Male
998      Male
999      Male
Name: Gender, Length: 764, dtype: object
```

In [ ]: