```
practical no.1
1. Introduction to Dataset
2. Python Libraries for Data Science
3. Description of Dataset
4. Panda Dataframe functions for load the dataset
5. Panda functions for Data Preprocessing
6. Panda functions for Data Formatting and Normalisation
7. Panda Functions for handling categorical variables
```

In [3]:
```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
```

In [8]:
```
df=sns.get_dataset_names()
print(data_set_name)
```

```
['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes', 'diamonds', 'dots', 'dowjon
es', 'exercise', 'flights', 'fmri', 'geyser', 'glue', 'healthexp', 'iris', 'mpg', 'penguins', 'pla
nets', 'seaice', 'taxis', 'tips', 'titanic', 'anagrams', 'anagrams', 'anscombe', 'anscombe', 'atte
ntion', 'attention', 'brain_networks', 'brain_networks', 'car_crashes', 'car_crashes', 'diamonds',
'diamonds', 'dots', 'dots', 'dowjones', 'dowjones', 'exercise', 'exercise', 'flights', 'flights',
'fmri', 'fmri', 'geyser', 'geyser', 'glue', 'glue', 'healthexp', 'healthexp', 'iris', 'iris', 'mp
g', 'mpg', 'penguins', 'penguins', 'planets', 'planets', 'seaice', 'seaice', 'taxis', 'taxis', 'ti
ps', 'tips', 'titanic', 'titanic', 'anagrams', 'anscombe', 'attention', 'brain_networks', 'car_cra
shes', 'diamonds', 'dots', 'dowjones', 'exercise', 'flights', 'fmri', 'geyser', 'glue', 'healthex
p', 'iris', 'mpg', 'penguins', 'planets', 'seaice', 'taxis', 'tips', 'titanic']
```

In [12]:
```
df=sns.load_dataset('titanic')
df
```

Out[12]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	al
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	y
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	y
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	y
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Southampton	
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	y
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Southampton	
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	y
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Queenstown	

891 rows × 15 columns

In [13]:
```
data1=df.head()
data1

```

Out[13]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no

In [14]:
```
data2=df.tail()
data2
```

Out[14]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive
886	0	2	male	27.0	0	0	13.00	S	Second	man	True	NaN	Southampton	no
887	1	1	female	19.0	0	0	30.00	S	First	woman	False	B	Southampton	yes
888	0	3	female	NaN	1	2	23.45	S	Third	woman	False	NaN	Southampton	no
889	1	1	male	26.0	0	0	30.00	C	First	man	True	C	Cherbourg	yes
890	0	3	male	32.0	0	0	7.75	Q	Third	man	True	NaN	Queenstown	no

In [16]:
```
data3=df.info()
data3
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```

In [17]:
```
data4=df['sex'].value_counts(normalize=True)
data4
```

Out[17]:
```
sex
male      0.647587
female    0.352413
Name: proportion, dtype: float64
```

In [18]:
```
data5=df.describe()
data5
```

Out[18]:

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [19]:
```
data6=df["deck"].value_counts(normalize=True)
data6
```

Out[19]:
```
deck
C    0.290640
B    0.231527
D    0.162562
E    0.157635
A    0.073892
F    0.064039
G    0.019704
Name: proportion, dtype: float64
```

In [20]:
```
data7=df.drop(["deck"], axis=1)
data7
```

Out[20]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	embark_town	alive	al
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	Southampton	no	F
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	Cherbourg	yes	F
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	Southampton	yes	T
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	Southampton	yes	F
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	Southampton	no	T
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	Southampton	no	T
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	Southampton	yes	T
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	Southampton	no	F
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	Cherbourg	yes	T
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	Queenstown	no	T

891 rows × 14 columns

In [24]:
```
 data8=df.drop(["embarked","class","who","adult_male","deck","embark_town","alone"],axis=1)
data8
```

Out[24]:

	survived	pclass	sex	age	sibsp	parch	fare	alive
0	0	3	male	22.0	1	0	7.2500	no
1	1	1	female	38.0	1	0	71.2833	yes
2	1	3	female	26.0	0	0	7.9250	yes
3	1	1	female	35.0	1	0	53.1000	yes
4	0	3	male	35.0	0	0	8.0500	no
...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	no
887	1	1	female	19.0	0	0	30.0000	yes
888	0	3	female	NaN	1	2	23.4500	no
889	1	1	male	26.0	0	0	30.0000	yes
890	0	3	male	32.0	0	0	7.7500	no

891 rows × 8 columns

In [26]:
```
data9=df['sex'].mode()[0]
data9
```

Out[26]:
```
'male'
```

In [35]:
```
data10=df['age'].mode
data10
```

Out[35]:
```
<bound method Series.mode of 0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888     NaN
889    26.0
890    32.0
Name: age, Length: 891, dtype: float64>
```

In [28]:
```
data11=df['age'].mean
data11
```

Out[28]:
```
<bound method Series.mean of 0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888     NaN
889    26.0
890    32.0
Name: age, Length: 891, dtype: float64>
```

In [29]:
```
data12=df.loc[:,"sex"].mode()
data12
```

Out[29]:
```
0    male
Name: sex, dtype: object
```

In [ ]:
```

```

In [50]:
```
bool_series = pd.notnull(df["sex"])
bool_series

```

Out[50]:
```
0      True
1      True
2      True
3      True
4      True
       ...
886    True
887    True
888    True
889    True
890    True
Name: sex, Length: 891, dtype: bool
```

In [55]:
```
df['age'].fillna(df['age'].mean(), inplace=True)

```

In [56]:

```
data15=df.info()
data15
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          891 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     891 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
```

```
name- Suhas Nidgundi
rollno. - 13249
```