Missing Value Treatment

```
In [1]:   #Aim : To Find out the Missing Value, Data Cleaning
```

```
In [30]:  #Name: Shruti Anil Dhote
          #RolL no. :72
          #Sub : ET1
          #section:C
          #Date:16/08/2024
```

```
In [5]:   import pandas as pd
```

```
In [6]:   import os
```

```
In [7]:   os.getcwd()
```

```
Out[7]:   'C:\\Users\\SURUTI DHOTE\\Desktop'
```

```
In [8]:   os.chdir("C:\\Users\\SURUTI DHOTE\\Desktop")
```

```
In [9]:   data=pd.read_csv("Titanic.csv")
```

```
In [10]:  data
```

Out[10]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875
...	...	...	...	...	...	...	...	...	...	...
413	1305	0	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500
416	1308	0	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500
417	1309	0	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583

418 rows × 12 columns

In [11]:
```python
data.head()
```

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN

In [13]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  418 non-null    int64
 1   Survived     418 non-null    int64
 2   Pclass       418 non-null    int64
 3   Name         418 non-null    object
 4   Sex          418 non-null    object
 5   Age          332 non-null    float64
 6   SibSp        418 non-null    int64
 7   Parch        418 non-null    int64
 8   Ticket       418 non-null    object
 9   Fare         417 non-null    float64
 10  Cabin        91 non-null     object
 11  Embarked     418 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

In [14]:
```python
data.describe()
```

Out[14]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	418.000000	332.000000	418.000000	418.000000	417.000000
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	0.392344	35.627188
std	120.810458	0.481622	0.841838	14.181209	0.896760	0.981429	55.907576
min	892.000000	0.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	996.250000	0.000000	1.000000	21.000000	0.000000	0.000000	7.895800
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	0.000000	14.454200
75%	1204.750000	1.000000	3.000000	39.000000	1.000000	0.000000	31.500000
max	1309.000000	1.000000	3.000000	76.000000	8.000000	9.000000	512.329200

In [15]: `data.shape`

Out[15]: `(418, 12)`

In [16]: `data.size`

Out[16]: 5016

In [17]: `data.ndim`

Out[17]: 2

In [18]: `data.isna()`

Out[18]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embar
0	False	False	False	False	False	False	False	False	False	False	True	F
1	False	False	False	False	False	False	False	False	False	False	True	F
2	False	False	False	False	False	False	False	False	False	False	True	F
3	False	False	False	False	False	False	False	False	False	False	True	F
4	False	False	False	False	False	False	False	False	False	False	True	F
...	...	...	...	...	...	...	...	...	...	...	...	
413	False	False	False	False	False	True	False	False	False	False	True	F
414	False	False	False	False	False	False	False	False	False	False	False	F
415	False	False	False	False	False	False	False	False	False	False	True	F
416	False	False	False	False	False	True	False	False	False	False	True	F
417	False	False	False	False	False	True	False	False	False	False	True	F

418 rows × 12 columns

In [24]: `data.isna().any`

Out[24]:   `<bound method NDFrame._add_numeric_operations.<locals>.any of          PassengerId  Su`
`rvived  Pclass    Name    Sex    Age  SibSp  Parch  Ticket  \`

12	False	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False	False	False
..	...	...	...	...	...	...	...	...	...
404	False	False	False	False	False	False	False	False	False
405	False	False	False	False	False	False	False	False	False
407	False	False	False	False	False	False	False	False	False
411	False	False	False	False	False	False	False	False	False
414	False	False	False	False	False	False	False	False	False

	Fare	Cabin	Embarked
12	False	False	False
14	False	False	False
24	False	False	False
26	False	False	False
28	False	False	False
..	...	...	...
404	False	False	False
405	False	False	False
407	False	False	False
411	False	False	False
414	False	False	False

`[87 rows x 12 columns]>`

In [25]:
```python
data.isna().sum()
```

Out[25]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

In [26]:
```python
data=data.dropna()
```

In [27]:
```python
data
```

Out[27]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
12	904	1	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	female	23.0	1	0	21228	82.2667
14	906	1	1	Chaffee, Mrs. Herbert Fuller (Carrie Constance...	female	47.0	1	0	W.E.P. 5734	61.1750
24	916	1	1	Ryerson, Mrs. Arthur Larned (Emily Maria Borie)	female	48.0	1	3	PC 17608	262.3750
26	918	1	1	Ostby, Miss. Helene Ragnhild	female	22.0	0	1	113509	61.9792
28	920	0	1	Brady, Mr. John Bertram	male	41.0	0	0	113054	30.5000
...	...	...	...	...	...	...	...	...	...	...
404	1296	0	1	Frauenthal, Mr. Isaac Gerald	male	43.0	1	0	17765	27.7208
405	1297	0	2	Nourney, Mr. Alfred (Baron von Drachstedt")"	male	20.0	0	0	SC/PARIS 2166	13.8625
407	1299	0	1	Widener, Mr. George Dunton	male	50.0	1	1	113503	211.5000
411	1303	1	1	Minahan, Mrs. William Edward (Lillian E Thorpe)	female	37.0	1	0	19928	90.0000
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000

87 rows × 12 columns

In [28]:
```python
data["Age"].fillna(30.272590)
```

Out[28]:
```
12      23.0
14      47.0
24      48.0
26      22.0
28      41.0
         ...
404     43.0
405     20.0
407     50.0
411     37.0
414     39.0
Name: Age, Length: 87, dtype: float64
```

In [29]: `data.isna().sum()`

Out[29]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

In [ ]: