

MISSING VALUE TREATMENT

#Aim : To Find out the Missing Value, Data Cleaning

#Name: Achal Chandure

#Roll no : 08

#Sub : ET1

#section:C

#Date:16/08/2024

```
import pandas as pd
```

```
import os
```

```
os.getcwd()
```

```
'C:\\Users\\HP'
```

```
os.chdir("C:\\Users\\HP\\Desktop")
```

```
data=pd.read_csv("Titanic.csv")
```

```
data
```

```
PassengerId  Survived  Pclass  \
```

Parch	\	Name	Sex	Age	SibSp
0		Kelly, Mr. James	male	34.5	0
0					
1		Wilkes, Mrs. James (Ellen Needs)	female	47.0	1
0					
2		Myles, Mr. Thomas Francis	male	62.0	0
0					
3		Wirz, Mr. Albert	male	27.0	0
0					
4		Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1
1					

```

..          ...      ...      ...      ...
...
413          Spector, Mr. Woolf      male      NaN      0
0
414          Oliva y Ocana, Dona. Fermina      female      39.0      0
0
415          Saether, Mr. Simon Sivertsen      male      38.5      0
0
416          Ware, Mr. Frederick      male      NaN      0
0
417          Peter, Master. Michael J      male      NaN      1
1

```

```

Ticket      Fare Cabin Embarked
0      330911      7.8292      NaN      Q
1      363272      7.0000      NaN      S
2      240276      9.6875      NaN      Q
3      315154      8.6625      NaN      S
4      3101298      12.2875      NaN      S
..          ...      ...      ...      ...
413      A.5. 3236      8.0500      NaN      S
414      PC 17758      108.9000      C105      C
415      SOTON/O.Q. 3101262      7.2500      NaN      S
416      359309      8.0500      NaN      S
417      2668      22.3583      NaN      C

```

[418 rows x 12 columns]

```
data.head()
```

```

PassengerId  Survived  Pclass  \
0      892      0      3
1      893      1      3
2      894      0      2
3      895      0      3
4      896      1      3

```

```

Name      Sex  Age  SibSp
Parch  \
0      Kelly, Mr. James      male      34.5      0
0
1      Wilkes, Mrs. James (Ellen Needs)      female      47.0      1
0
2      Myles, Mr. Thomas Francis      male      62.0      0
0
3      Wirz, Mr. Albert      male      27.0      0
0
4      Hirvonen, Mrs. Alexander (Helga E Lindqvist)      female      22.0      1
1

```

	Ticket	Fare	Cabin	Embarked
0	330911	7.8292	NaN	Q
1	363272	7.0000	NaN	S
2	240276	9.6875	NaN	Q
3	315154	8.6625	NaN	S
4	3101298	12.2875	NaN	S

```
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
```

```
data.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	\
count	418.000000	418.000000	418.000000	332.000000	418.000000	
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	
std	120.810458	0.481622	0.841838	14.181209	0.896760	
min	892.000000	0.000000	1.000000	0.170000	0.000000	
25%	996.250000	0.000000	1.000000	21.000000	0.000000	
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	
75%	1204.750000	1.000000	3.000000	39.000000	1.000000	
max	1309.000000	1.000000	3.000000	76.000000	8.000000	

	Parch	Fare
count	418.000000	417.000000
mean	0.392344	35.627188
std	0.981429	55.907576
min	0.000000	0.000000
25%	0.000000	7.895800
50%	0.000000	14.454200
75%	0.000000	31.500000
max	9.000000	512.329200

```
data.shape
```

```
(418, 12)
```

```
data.size
```

```
5016
```

```
data.ndim
```

```
2
```

```
data.isna()
```

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

	Fare	Cabin	Embarked
0	False	True	False
1	False	True	False
2	False	True	False
3	False	True	False
4	False	True	False
...
413	False	True	False
414	False	False	False
415	False	True	False
416	False	True	False
417	False	True	False

```
[418 rows x 12 columns]
```

```
data.isna().any
```

```
<bound method NDFrame._add_numeric_operations.<locals>.any of  
PassengerId  Survived  Pclass    Name      Sex      Age SibSp  Parch  
Ticket  \
```

0		False	False	False	False	False	False	False	False
False									
1		False	False	False	False	False	False	False	False
False									
2		False	False	False	False	False	False	False	False
False									
3		False	False	False	False	False	False	False	False
False									
4	False	False	False	False	False	False	False	False	
False									
..	
...									
413		False	False	False	False	False	True	False	False414
False									
False		False	False	False	False	False	False	False415	False
False									
False	False	False	False	False	False	False	False416	False	False
False									
False	False	False	True	False	False417	False	False	False	False
False									
False	True	False	False						
False									

	Fare	Cabin	Embarked
0	False	True	False
1	False	True	False
2	False	True	False
3	False	True	False
4	False	True	False
..
413	False	True	False
414	False	False	False
415	False	True	False
416	False	True	False
417	False	True	False

```
[418 rows x 12 columns]>
```

```
data.isna().sum()
```

```
-
```

```
Pclass      0
Name        0
Sex         0
Age        86
SibSp       0
Parch       0
Ticket      0
Fare        1
Cabin      327
Embarked    0
dtype: int64
```

```
data=data.dropna()
```

```
data
```

```

      PassengerId  Survived  Pclass  \
12             904         1       1
14             906         1       1
24             916         1       1
26             918         1       1
28             920         0       1
..            ...         ...     ...
404            1296         0       1
405            1297         0       2
407            1299         0       1
411            1303         1       1
414            1306         1       1

```

```

                                     Name      Sex  Age
SibSp  \
12      Snyder, Mrs. John Pillsbury (Nelle Stevenson)  female  23.0
1
14      Chaffee, Mrs. Herbert Fuller (Carrie Constance...  female  47.0
1
24      Ryerson, Mrs. Arthur Larned (Emily Maria Borie)   female  48.0
1
26                                     Ostby, Miss. Helene Ragnhild  female  22.0
0
28                                     Brady, Mr. John Bertram      male  41.0
0
..                                     ...         ...     ...
...
404                                     Frauenthal, Mr. Isaac Gerald      male  43.0
1
405      Nourney, Mr. Alfred (Baron von Drachstedt)" "      male  20.0
0
407                                     Widener, Mr. George Dunton      male  50.0
1

```

```
411      Minahan, Mrs. William Edward (Lillian E Thorpe)  female  37.0
```

```
1
414                                Oliva y Ocana, Dona. Fermina  female  39.0
0
```

	Parch	Ticket	Fare	Cabin	Embarked
12	0	21228	82.2667	B45	S
14	0	W.E.P. 5734	61.1750	E31	S
24	3	PC 17608	262.3750	B57 B59 B63 B66	C
26	1	113509	61.9792	B36	C
28	0	113054	30.5000	A21	S
...
404	0	17765	27.7208	D40	C
405	0	SC/PARIS 2166	13.8625	D38	C
407	1	113503	211.5000	C80	C
411	0	19928	90.0000	C78	Q
414	0	PC 17758	108.9000	C105	C

```
[87 rows x 12 columns]
```

```
data["Age"].fillna(30.272590)
```

```
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
```

```
data.isna().sum()
```

```
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
```