

## data cleaning\_missing value treatment

```
In [ ]: 1 #Name : Achal Gajanan Ghorad  
2 #Roll no. :39  
3 #Section :3A  
4 #Date:03/08/2024
```

```
In [1]: 1 #Aim: to find out data cleaning ,missing value
```

```
In [2]: 1 import pandas as pd
```

```
In [3]: 1 import os
```

```
In [4]: 1 os.getcwd()
```

```
Out[4]: 'C:\\Users\\ACHAL '
```

```
In [6]: 1 os.chdir("C:\\Users\\ACHAL\\OneDrive\\Desktop")
```

```
In [7]: 1 data=pd.read_csv("archive (5).zip")
```

In [8]:

1 data

Out[8]:

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

418 rows × 12 columns



In [9]: ▶ 1 data.head(25)

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298
5	897	0	3	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538
6	898	1	3	Connolly, Miss. Kate	female	30.0	0	0	330972
7	899	0	2	Caldwell, Mr. Albert Francis	male	26.0	1	1	248738
8	900	1	3	Abraham, Mrs. Joseph (Sophie Halaut Easu)	female	18.0	0	0	2657
9	901	0	3	Davies, Mr. John Samuel	male	21.0	2	0	A/4 48871
10	902	0	3	Ilieff, Mr. Yljo	male	NaN	0	0	349220
11	903	0	1	Jones, Mr. Charles Cresson	male	46.0	0	0	694
12	904	1	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	female	23.0	1	0	21228
13	905	0	2	Howard, Mr. Benjamin	male	63.0	1	0	24065
14	906	1	1	Chaffee, Mrs. Herbert Fuller (Carrie Constance...)	female	47.0	1	0	W.E.P. 5734
15	907	1	2	del Carlo, Mrs. Sebastiano (Argenia Genovesi)	female	24.0	1	0	SC/PARIS 2167
16	908	0	2	Keane, Mr. Daniel	male	35.0	0	0	233734

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
17	909	0	3	Assaf, Mr. Gerios	male	21.0	0	0	2692
18	910	1	3	Ilmakangas, Miss. Ida Livija	female	27.0	1	0	STON/O2. 3101270
19	911	1	3	Assaf Khalil, Mrs. Mariana (Miriam)"	female	45.0	0	0	2696
20	912	0	1	Rothschild, Mr. Martin	male	55.0	1	0	PC 17603
21	913	0	3	Olsen, Master. Artur Karl	male	9.0	0	1	C 17368
22	914	1	1	Flegenheim, Mrs. Alfred (Antoinette)	female	NaN	0	0	PC 17598
23	915	0	1	Williams, Mr. Richard Norris II	male	21.0	0	1	PC 17597
24	916	1	1	Ryerson, Mrs. Arthur Larned (Emily Maria Borie)	female	48.0	1	3	PC 17608

In [10]: 1 data.info

```
Out[10]: <bound method DataFrame.info of
\
0      892      0      3
1      893      1      3
2      894      0      2
3      895      0      3
4      896      1      3
..      ...      ...      ...
413    1305      0      3
414    1306      1      1
415    1307      0      3
416    1308      0      3
417    1309      0      3

      Name      Sex      Age      Si
bSp  Parch  \
0      Kelly, Mr. James      male      34.5
0      0
1      Wilkes, Mrs. James (Ellen Needs)      female      47.0
1      1
```

In [11]: 1 data.describe()

Out[11]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	418.000000	418.000000	418.000000	332.000000	418.000000	418.000000	417.00
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	0.392344	35.62
std	120.810458	0.481622	0.841838	14.181209	0.896760	0.981429	55.90
min	892.000000	0.000000	1.000000	0.170000	0.000000	0.000000	0.00
25%	996.250000	0.000000	1.000000	21.000000	0.000000	0.000000	7.89
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	0.000000	14.45
75%	1204.750000	1.000000	3.000000	39.000000	1.000000	0.000000	31.50
max	1309.000000	1.000000	3.000000	76.000000	8.000000	9.000000	512.32

In [12]: 1 data.shape

Out[12]: (418, 12)

In [13]: 1 data.size

Out[13]: 5016

In [14]: 1 data.ndim

Out[14]: 2

In [15]: 1 data.isna()

Out[15]:

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

418 rows × 12 columns

In [16]: 1 data.isna().any()

Out[16]: PassengerId False  
Survived False  
Pclass False  
Name False  
Sex False  
Age True  
SibSp False  
Parch False  
Ticket False  
Fare True  
Cabin True  
Embarked False  
dtype: bool

In [17]: 1 data.isna().sum()

Out[17]: PassengerId 0  
Survived 0  
Pclass 0  
Name 0  
Sex 0  
Age 86  
SibSp 0  
Parch 0  
Ticket 0  
Fare 1  
Cabin 327  
Embarked 0  
dtype: int64

In [19]: 1 data=data.dropna()

In [20]:

1 data

Out[20]:

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

87 rows × 12 columns





In [21]: 1 data["Age"].fillna(30.272590)

Out[21]:

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 [22]: 1 data.isna().sum()

Out[22]:

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 [ ]: 1