

```
In [1]: #EXP_4
```

```
In [2]: #Aim:To learn Data Preprocessing And Missing Values
```

```
In [1]: # Name:Dev Sanjay Vaidya
# Roll no.:69
# Sec : B
# Subject:ET1
# Date:14/08/25
```

```
In [4]: import pandas as pd
```

```
In [5]: import os
```

```
In [6]: os.getcwd()
```

```
Out[6]: 'C:\\\\Users\\\\LENOVO'
```

```
In [7]: os.chdir("C:\\\\Users\\\\LENOVO\\\\Desktop")
```

```
In [8]: data=pd.read_csv("C:\\\\Users\\\\LENOVO\\\\Desktop\\\\titanic.csv")
```

```
In [9]: data.head()
```

```
Out[9]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tic
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	21
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/ 3101
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373

```
In [10]: data.tail()
```

Out[10]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tic
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211.
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112.
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W 6
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111.
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370.

In [11]: `data.shape`

Out[11]: (891, 12)

In [12]: `data.ndim`

Out[12]: 2

In [13]: `data.columns`

Out[13]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
 dtype='object')

In [14]: `data.describe`

```
Out[14]: <bound method NDFrame.describe of
          PassengerId  Survived  Pclass \
0              1      0      3
1              2      1      1
2              3      1      3
3              4      1      1
4              5      0      3
...
886            887      0      2
887            888      1      1
888            889      0      3
889            890      1      1
890            891      0      3

                                                Name     Sex   Age  SibSp
\
0           Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th... female  38.0      1
2           Heikkinen, Miss. Laina female  26.0      0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35.0      1
4           Allen, Mr. William Henry    male  35.0      0
...
886            Montvila, Rev. Juozas    male  27.0      0
887            Graham, Miss. Margaret Edith female  19.0      0
888            Johnston, Miss. Catherine Helen "Carrie" female   NaN      1
889            Behr, Mr. Karl Howell    male  26.0      0
890            Dooley, Mr. Patrick    male  32.0      0

   Parch      Ticket     Fare Cabin Embarked
0      0       A/5 21171  7.2500   NaN      S
1      0        PC 17599  71.2833   C85      C
2      0  STON/O2. 3101282  7.9250   NaN      S
3      0       113803  53.1000  C123      S
4      0       373450  8.0500   NaN      S
...
886      0       211536 13.0000   NaN      S
887      0       112053 30.0000   B42      S
888      2       W./C. 6607 23.4500   NaN      S
889      0       111369 30.0000  C148      C
890      0       370376  7.7500   NaN      Q

[891 rows x 12 columns]>
```

```
In [15]: data.info
```

```
Out[15]: <bound method DataFrame.info of
0          1          0          3
1          2          1          1
2          3          1          3
3          4          1          1
4          5          0          3
...
886      887          0          2
887      888          1          1
888      889          0          3
889      890          1          1
890      891          0          3
```

		Name	Sex	Age	SibSp
0		Braund, Mr. Owen Harris	male	22.0	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th... 2	Heikkinen, Miss. Laina	female	38.0	1
3	Futrelle, Mrs. Jacques Heath (Lily May Peel) 4	Allen, Mr. William Henry	female	26.0	0
4			male	35.0	1
886		Montvila, Rev. Juozas 887	male	27.0	0
888	Graham, Miss. Margaret Edith 889	Johnston, Miss. Catherine Helen "Carrie"	female	19.0	0
889		Behr, Mr. Karl Howell	male	NaN	1
890		Dooley, Mr. Patrick	male	26.0	0
890			male	32.0	0

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
888	2	W./C. 6607	23.4500	NaN	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

[891 rows x 12 columns]>

```
In [16]: data.isna()
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0		False	False	False	False	False	False	False	False
1		False	False	False	False	False	False	False	False
2		False	False	False	False	False	False	False	False
3		False	False	False	False	False	False	False	False
4		False	False	False	False	False	False	False	False
...
886		False	False	False	False	False	False	False	False
887		False	False	False	False	False	False	False	False
888		False	False	False	False	False	True	False	False
889		False	False	False	False	False	False	False	False
890		False	False	False	False	False	False	False	False

891 rows × 12 columns

In [17]: `data.isna().any()`

Out[17]:

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

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

Out[18]:

PassengerId	0
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 [19]: data
```

```
Out[19]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	...
0		1	0	Braund, Mr. Owen Harris	male	22.0	1	0	
1		2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC
2		3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STC 31
3		4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	1
4		5	0	Allen, Mr. William Henry	male	35.0	0	0	3
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	2
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	1
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NAN	1	2	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	1
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	3

891 rows × 12 columns

Missing Values Treatment

```
In [20]: #Check the missing value by record  
data.isna()
```

```
Out[20]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0		False	False	False	False	False	False	False	False
1		False	False	False	False	False	False	False	False
2		False	False	False	False	False	False	False	False
3		False	False	False	False	False	False	False	False
4		False	False	False	False	False	False	False	False
...
886		False	False	False	False	False	False	False	False
887		False	False	False	False	False	False	False	False
888		False	False	False	False	False	True	False	False
889		False	False	False	False	False	False	False	False
890		False	False	False	False	False	False	False	False

891 rows × 12 columns

```
In [21]: data.isna().any()
```

```
Out[21]: PassengerId    False  
Survived      False  
Pclass        False  
Name          False  
Sex           False  
Age           True  
SibSp         False  
Parch         False  
Ticket        False  
Fare          False  
Cabin         True  
Embarked      True  
dtype: bool
```

```
In [22]: df1=data.dropna()
```

```
In [23]: df1
```

Out[23]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	T
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Futrelle, Mrs. Jacques Heath (Lily May Peel) McCarthy, Mr. Timothy J Sandstrom, Miss. Marguerite Rut Bonnell, Miss. Elizabeth Beckwith, Mrs. Richard Leonard (Sallie Monypenny) Carlsson, Mr. Frans Olof Potter, Mrs. Thomas Jr (Lily Alexenia Wilson) Graham, Miss. Margaret Edith Behr, Mr. Karl Howell	female female male female female female female female female female female male female female female female male	38.0 35.0 54.0 4.0 58.0 47.0 33.0 56.0 19.0 26.0		1 1 0 1 0 1 0 0 0 1 1 1 1 1 0 0	1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 11
3	4	1	1						
6	7	0	1						
10	11	1	3						
11	12	1	1						
...
871	872	1	1						
872	873	0	1						
879	880	1	1						
887	888	1	1						
889	890	1	1						

183 rows × 12 columns

In [24]: df1.isna().sum()

```
Out[24]: 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 [25]: data[ "Age" ].fillna( 29.699118 )
```

```
Out[25]: 0      22.000000  
1      38.000000  
2      26.000000  
3      35.000000  
4      35.000000  
...  
886    27.000000  
887    19.000000  
888    29.699118  
889    26.000000  
890    32.000000  
Name: Age, Length: 891, dtype: float64
```

```
In [ ]:
```