

```
In [1]: #exp no: 4
```

```
In [2]: #Name:Vedant M.Padole
#Roll no:42
#Sec:C
#Subject:ET1
#Date:
```

```
In [12]: import pandas as pd
```

```
In [13]: import os
```

```
In [ ]:
```

```
In [14]: os.getcwd()
```

```
Out[14]: 'C:\\\\Users\\\\DELL\\\\desktop'
```

```
In [15]: os.chdir('C:\\\\Users\\\\DELL\\\\desktop')
```

```
In [16]: data=pd.read_csv("titanic.csv")
```

```
In [18]: data.head(20)
```

```
Out[18]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
<b>0</b>	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Na
<b>1</b>	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina Futrelle, Mrs. Jacques	female	38.0	1	0	PC 17599	71.2833	C8
<b>2</b>	3	1	3	Allen, Mr. William Henry	female	26.0	0	0	STON/O2. 3101282	7.9250	Na
<b>3</b>	4	1	1	Heath (Lily May Peel) Allen, Mr.	female	35.0	1	0	113803	53.1000	C12
<b>4</b>	5	0	3	WIlliam Henry Moran, Mr.	male	35.0	0	0	373450	8.0500	Na
<b>5</b>	6	0	3	James McCarthy, Mr. Timothy J	male	NaN	0	0	330877	8.4583	Na
<b>6</b>	7	0	1	Palsson, Master. Gosta Leonard Johnson, Mrs. Oscar W (Elisabeth	male	54.0	0	0	17463	51.8625	E4
<b>7</b>	8	0	3	Johnson, Mrs. Oscar W (Elisabeth	male	2.0	3	1	349909	21.0750	Na
<b>8</b>	9	1	3	W (Elisabeth	female	27.0	0	2	347742	11.1333	Na

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
9	10	1	Vilhelmina Berg)							
10	11	1	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	Na
11	12	1	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G
12	13	0	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C10
13	14	0	Saundercock, Mr. William	male	20.0	0	0	A/5. 2151	8.0500	Na
14	15	0	Andersson, Mr. Anders	male	39.0	1	5	347082	31.2750	Na
15	16	1	Johan Vestrom, Miss. Hulda	female	14.0	0	0	350406	7.8542	Na
16	17	0	Amanda Adolfina Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0000	Na
17	18	1	Rice, Master. Eugene	male	2.0	4	1	382652	29.1250	Na
18	19	0	Williams, Mr. Charles	male	Nan	0	0	244373	13.0000	Na
19	20	1	Eugene Vander	female	31.0	1	0	345763	18.0000	Na
			Planke, Mrs. Julius (Emelia							
			Maria Vande...	female						
			Masselmani, Mrs. Fatima	female						

In [19]: `data.tail(20)`

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cab
871	872	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D
872	873	0	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cab
873	874	0	Vander Cruyssen, Mr. Victor	male	47.0	0	0	345765	9.0000	N
874	875	1	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1	0	P/PP 3381	24.0000	N
875	876	1	Najib, Miss. Adele Kiamie "Jane"	female	15.0	0	0	2667	7.2250	N
876	877	0	Gustafsson, Mr. Alfred Ossian	male	20.0	0	0	7534	9.8458	N
877	878	0	Petroff, Mr. Nedelio	male	19.0	0	0	349212	7.8958	N
878	879	0	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	N
879	880	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C
880	881	1	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	N
881	882	0	Markun, Mr. Johann Dahlberg,	male	33.0	0	0	349257	7.8958	N
882	883	0	Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	N
883	884	0	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	N
884	885	0	Suttehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	N
885	886	0	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	N
886	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	N
887	888	1	Graham, Miss.	female	19.0	0	0	112053	30.0000	B

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cab
888	889	0	Margaret Edith Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	N
889	890	1	Behr, Mr. Karl Howell Dooley,	male	26.0	0	0	111369	30.0000	C1
890	891	0	Mr. Patrick	male	32.0	0	0	370376	7.7500	N

In [22]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column      Non-Null Count Dtype  
 --- 
 0   PassengerId 891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object  
 4   Sex          891 non-null    object  
 5   Age          714 non-null   float64 
 6   SibSp        891 non-null   int64  
 7   Parch        891 non-null   int64  
 8   Ticket       891 non-null   object  
 9   Fare         549 non-null   float64 
 10  Cabin        204 non-null   object  
 11  Embarked     891 non-null   object  
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [23]: `data.shape`

Out[23]: (891, 12)

In [24]: `data.size`

Out[24]: 10692

In [25]: `data.size`

Out[25]: 10692

In [29]: `data.describe()`

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400

	<b>PassengerId</b>	<b>Survived</b>	<b>Pclass</b>	<b>Age</b>	<b>SibSp</b>	<b>Parch</b>	<b>Fare</b>
<b>50%</b>	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
<b>75%</b>	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
<b>max</b>	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [32]: `data`

	<b>PassengerId</b>	<b>Survived</b>	<b>Pclass</b>	<b>Name</b>	<b>Sex</b>	<b>Age</b>	<b>SibSp</b>	<b>Parch</b>	<b>Ticket</b>	<b>Fare</b>	<b>Cabin</b>
<b>0</b>	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
<b>1</b>	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen,	female	38.0	1	0	PC 17599	71.2833	C85
<b>2</b>	3	1	3	Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
<b>3</b>	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel) Allen, Mr.	female	35.0	1	0	113803	53.1000	C123
<b>4</b>	5	0	3	William Henry	male	35.0	0	0	373450	8.0500	NaN
...	...	...	...	...	...	...	...	...	...	...	...
<b>886</b>	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN
<b>887</b>	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42
<b>888</b>	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN
<b>889</b>	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148
<b>890</b>	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN

891 rows × 12 columns

```
In [33]: data.isna()
```

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

891 rows × 12 columns

```
In [34]: data.isna().sum()
```

```
Out[34]: 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 [36]: data.describe()
```

```
Out[36]:    PassengerId  Survived  Pclass   Age     SibSp  Parch     Fare
count  891.000000  891.000000  891.000000  714.000000  891.000000  891.000000  891.000000
mean   446.000000  0.383838  2.308642  29.699118  0.523008  0.381594  32.204208
std    257.353842  0.486592  0.836071  14.526497  1.102743  0.806057  49.693429
min    1.000000  0.000000  1.000000  0.420000  0.000000  0.000000  0.000000
25%   223.500000  0.000000  2.000000  20.125000  0.000000  0.000000  7.910400
50%   446.000000  0.000000  3.000000  28.000000  0.000000  0.000000  14.454200
75%   668.500000  1.000000  3.000000  38.000000  1.000000  0.000000  31.000000
max   891.000000  1.000000  3.000000  80.000000  8.000000  6.000000  512.329200
```

```
In [38]: data["Age"].fillna(32.204208)
```

```
Out[38]: 0    22.000000
2    38.000000
3    26.000000
4    35.000000
      ...
886   ...
887   27.000000
888   19.000000
889   32.204208
890   26.000000
      32.000000
Name: Age, Length: 891, dtype: float64
```

```
In [39]: data.isna().sum()
```

```
Out[39]: 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 [40]: data.dropna()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103
...	...	...	...	...	...	...	...	...	...	...	...
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson) Graham, Miss. Margaret Edith Behr, Mr.	female	56.0	0	1	11767	83.1583	C50
887	888	1	1	Karl Howell	female	19.0	0	0	112053	30.0000	B42
889	890	1	1		male	26.0	0	0	111369	30.0000	C148

183 rows × 12 columns

In [42]: `data.isna()`

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

891 rows × 12 columns

## Conclusion :

Missing value treatment is a crucial preprocessing step that ensures data quality and model accuracy. Proper handling of missing data minimizes bias and improves the reliability of statistical analysis and machine learning outcomes.

In [ ]:

