

Missing Value Treatment

```
In [2]: #Name: Yash Pravin Gadbail  
#Roll no. : 35  
#Sec: 3rd A  
#Sub : ET 1  
#Date:03/08/2024
```

```
In [4]: #Aim : To Find out the Missing Value, Data Cleaning
```

```
In [6]: import pandas as pd
```

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

```
In [2]: os.getcwd()
```

```
Out[2]: 'C:\\Users\\RAG'
```

```
In [12]: os.chdir("C:\\Users\\OneDrive\\Desktop")
```

```
In [14]: data=pd.read_csv("Titanic.csv")
```

In [16]: data

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75

891 rows × 12 columns



In [18]: `data.head(25)`

Out[18]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	I
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5
12	13	0	3	Saunderscock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0
16	17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29.1

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	I	
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18.0
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2
20	21	0	2	Fynney, Mr. Joseph J	male	35.0	0	0	239865	26.0
21	22	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13.0
22	23	1	3	McGowan, Miss. Anna "Annie"	female	15.0	0	0	330923	8.0
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.5
24	25	0	3	Palsson, Miss. Torborg Danira	female	8.0	3	1	349909	21.0

In [20]: 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            891 non-null    float64
10  Cabin           204 non-null    object
11  Embarked        889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [22]: data.describe()

Out[22]:

	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 [24]: data.shape

Out[24]: (891, 12)

In [28]: data.size

Out[28]: 10692

In [30]: data.ndim

Out[30]: 2

In [32]: data.isna()

Out[32]:

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

891 rows × 12 columns

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

```
Out[34]: 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 [36]: data.isna().sum()
```

```
Out[36]: 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 [38]: data=data.dropna()
```

In [40]: data

Out[40]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500
...
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000

183 rows × 12 columns




```
In [42]: data["Age"].fillna(30.272590)
```

```
Out[42]: 1      38.0
          3      35.0
          6      54.0
          10     4.0
          11     58.0
          ...
          871    47.0
          872    33.0
          879    56.0
          887    19.0
          889    26.0
          Name: Age, Length: 183, dtype: float64
```

```
In [44]: data.isna().sum()
```

```
Out[44]: 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 [ ]:
```