

November 1, 2023

1 Missing Value Treatment

```
[1]: #Name: Vedant Wankhade  
#Sec : B  
#Roll no: 74  
#Dt: 16/09/2023
```

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

```
[3]: import os
```

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

```
[4]: 'C:\\Users\\HP'
```

```
[5]: os.chdir('C:\\Users\\HP\\Desktop')
```

```
[6]: data=pd.read_csv('titanic_train.csv')
```

```
[7]: data.head()
```

```
[7]: PassengerId  Survived  Pclass  \  
0             1         0       3  
1             2         1       1  
2             3         1       3  
3             4         1       1  
4             5         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
```

```
    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	

```
[8]: data.tail()
```

```
[8]:
```

	PassengerId	Survived	Pclass	Name \
886	887	0	2	Montvila, Rev. Juozas
887	888	1	1	Graham, Miss. Margaret Edith
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"
889	890	1	1	Behr, Mr. Karl Howell
890	891	0	3	Dooley, Mr. Patrick

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	male	27.0	0	0	211536	13.00	NaN	S
887	female	19.0	0	0	112053	30.00	B42	S
888	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	male	26.0	0	0	111369	30.00	C148	C
890	male	32.0	0	0	370376	7.75	NaN	Q

```
[9]: 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
```

```
[10]: data.describe()
```

```
[10]:
```

	PassengerId	Survived	Pclass	Age	SibSp
count	891.000000	891.000000	891.000000	714.000000	891.000000

mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

```
[11]: data.isna()
```

```
[11]:
```

	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	
..	
886	False	False	False	False	False	False	False	False	False	
887	False	False	False	False	False	False	False	False	False	
888	False	False	False	False	False	True	False	False	False	
889	False	False	False	False	False	False	False	False	False	
890	False	False	False	False	False	False	False	False	False	

	Fare	Cabin	Embarked
0	False	True	False
1	False	False	False
2	False	True	False
3	False	False	False
4	False	True	False
..
886	False	True	False
887	False	False	False
888	False	True	False
889	False	False	False
890	False	True	False

```
[891 rows x 12 columns]
```

```
[12]: data.isna().any()
```

```
[12]: 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
```

```
[13]: data.shape
```

```
[13]: (891, 12)
```

```
[14]: data.isna().sum()
```

```
[14]: 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
```

```
[15]: data["Age"].fillna(29.699118)
```

```
[15]: 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
```

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

```
[16]: 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
```

```
[17]: df=data.fillna(data.mean())
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_7788\151685739.py:1: FutureWarning:
Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None')
is deprecated; in a future version this will raise TypeError.  Select only valid
columns before calling the reduction.
  df=data.fillna(data.mean())
```

```
[18]: df.isna().sum()
```

```
[18]: PassengerId      0
      Survived        0
      Pclass         0
      Name           0
      Sex            0
      Age           0
      SibSp          0
      Parch          0
      Ticket         0
      Fare           0
      Cabin         687
      Embarked       2
      dtype: int64
```

```
[20]: data.isnull()
```

```

[20]:
      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
..           ...      ...    ...    ...    ...    ...    ...    ...
886           False      False   False  False  False  False  False  False  False
887           False      False   False  False  False  False  False  False  False
888           False      False   False  False  False  True   False  False  False
889           False      False   False  False  False  False  False  False  False
890           False      False   False  False  False  False  False  False  False

      Fare  Cabin  Embarked
0      False  True   False
1      False  False  False
2      False  True   False
3      False  False  False
4      False  True   False
..           ...    ...
886  False  True   False
887  False  False  False
888  False  True   False
889  False  False  False
890  False  True   False

[891 rows x 12 columns]

```