



```
In [11]: # Calculating the Null values in AGE Columns
df['Age'].isnull().sum()
```

```
Out[11]: 177
```

```
In [12]: # Calculating the Null values in Cabin Columns
df['Cabin'].isnull().sum()
```

```
Out[12]: 687
```

```
In [13]: # get some initial statistics
df.describe()
```

```
Out[13]:
```

	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 [14]: # getting some information about dataset
df.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 [15]: # finding data types
df.dtypes
```

```
Out[15]: PassengerId    int64
Survived              int64
Pclass                int64
Name                  object
Sex                   object
Age                   float64
SibSp                 int64
Parch                 int64
Ticket                object
Fare                  float64
Cabin                 object
Embarked              object
dtype: object
```

```
In [16]: # finding dimensions of the data frame
df.shape
```

```
Out[16]: (891, 12)
```

```
In [20]: def impute_age(cols):
Age = cols[0]
Pclass = cols[1]

if pd.isnull(Age):

    if Pclass == 1:
        return 37

    elif Pclass == 2:
        return 29

    else:
        return 24

else:
    return Age
```

```
In [21]: # applying the function
df['Age'] = df[['Age', 'Pclass']].apply(impute_age,axis=1)
```

```
In [22]: # dropping cabin column
df.drop('Cabin',axis=1,inplace=True)
```

```
In [23]: df.dropna(inplace=True)
```

In [24]: df.head()

```
Out[24]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S

In [25]: df.tail()

```
Out[25]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	24.0	1	2	W./C. 6607	23.45	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	Q

In [26]: df.isnull().sum()

```
Out[26]: PassengerId    0
Survived              0
Pclass                0
Name                  0
Sex                   0
Age                   0
SibSp                 0
Parch                 0
Ticket                0
Fare                  0
Embarked              0
dtype: int64
```

In [27]: df.dtypes

```
Out[27]: PassengerId    int64
Survived              int64
Pclass                int64
Name                  object
Sex                   object
Age                   float64
SibSp                 int64
Parch                 int64
Ticket                object
Fare                  float64
Embarked              object
dtype: object
```

In [28]: *# data type conversion*
df['Age'] = df['Age'].astype('int')

In [29]: df.dtypes

```
Out[29]: PassengerId    int64
Survived              int64
Pclass                int64
Name                  object
Sex                   object
Age                   int32
SibSp                 int64
Parch                 int64
Ticket                object
Fare                  float64
Embarked              object
dtype: object
```

In [30]: *# data types conversion*
df['Age'] = df['Age'].round(0).astype('int')

```
In [31]: df.dtypes
```

```
Out[31]: PassengerId    int64
Survived              int64
Pclass               int64
Name                 object
Sex                  object
Age                 int32
SibSp               int64
Parch              int64
Ticket              object
Fare               float64
Embarked            object
dtype: object
```

```
In [32]: df.head()
```

```
Out[32]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	S

```
In [33]: # Converting Categorical Variables to Quantitative Variables
cat = pd.get_dummies(df, columns=['Sex'])
```

```
In [34]: cat.head()
```

```
Out[34]:
```

	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	Embarked	Sex_female	Sex_male
0	1	0	3	Braund, Mr. Owen Harris	22	1	0	A/5 21171	7.2500	S	False	True
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38	1	0	PC 17599	71.2833	C	True	False
2	3	1	3	Heikkinen, Miss. Laina	26	0	0	STON/O2. 3101282	7.9250	S	True	False
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35	1	0	113803	53.1000	S	True	False
4	5	0	3	Allen, Mr. William Henry	35	0	0	373450	8.0500	S	False	True

```
In [36]: # Female = 0
cat["Sex_female"]
```

```
Out[36]: 0    False
1     True
2     True
3     True
4    False
...
886   False
887    True
888    True
889   False
890   False
Name: Sex_female, Length: 889, dtype: bool
```

```
In [37]: # Male = 1
cat["Sex_male"]
```

```
Out[37]: 0     True
1    False
2    False
3    False
4     True
...
886    True
887   False
888   False
889    True
890    True
Name: Sex_male, Length: 889, dtype: bool
```