WORKING ON PANDAS LIBRARY

In [1]:
```
# Importing Pandas Library's
import pandas as pd
```

In [2]:
```
# Upload Files through pandas library
df = pd.read_csv("titanic.csv")
```

In [3]:
```
# show data using print or variable through
df
```

Out[3]:

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

891 rows × 12 columns

WAp to print out the first 10 row along with all columns

In [11]:
```
df.head(10)
```

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

WAP to print out the last 10 rows along with all columns

In [12]:
```
df.tail(10)
```

Out[12]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	NaN	S
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

WAp to show all about the data in statistical wise

df.describe()

WAP to print out the count of row and column

In [14]:
```
df.shape
```

Out[14]:
```
(891, 12)
```

WAp to print out the name of columns

In [16]:
```
list(df.columns)
```

Out[16]:
```
['PassengerId',
 'Survived',
 'Pclass',
 'Name',
 'Sex',
 'Age',
 'SibSp',
 'Parch',
 'Ticket',
 'Fare',
 'Cabin',
 'Embarked']
```

WAP to print out the data types on data_sets

In [22]:
```
print(df.dtypes)
```

```
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
```

WAP to print out all about the data information

In [23]:
```
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
```

WAP to print out the String data type in data__sets

In [24]:
```
df.dtypes == 'int64'
```

Out[24]:
```
PassengerId     True
Survived        True
Pclass          True
Name           False
Sex            False
Age            False
SibSp           True
Parch           True
Ticket         False
Fare           False
Cabin          False
Embarked       False
dtype: bool
```

WAP to print out which column have float value

In [25]:
```
df.dtypes == 'float64'
```

Out[25]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare            True
Cabin          False
Embarked       False
dtype: bool
```

WAP to find out the which columns have string or object value

In [27]:
```
df.dtypes == 'object'
```

Out[27]:
```
PassengerId    False
Survived       False
Pclass         False
Name            True
Sex             True
Age            False
SibSp          False
Parch          False
Ticket          True
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

WAP to print out only integer value columns

In [28]:

```
df[df.dtypes[df.dtypes == 'int64'].index]
```

Out[28]:

	PassengerId	Survived	Pclass	SibSp	Parch
0	1	0	3	1	0
1	2	1	1	1	0
2	3	1	3	0	0
3	4	1	1	1	0
4	5	0	3	0	0
...	...	...	...	...	...
886	887	0	2	0	0
887	888	1	1	0	0
888	889	0	3	1	2
889	890	1	1	0	0
890	891	0	3	0	0

891 rows × 5 columns

WAP to print out only Float value columns

In [29]:

```
df[df.dtypes[df.dtypes == 'float64'].index]
```

Out[29]:

	Age	Fare
0	22.0	7.2500
1	38.0	71.2833
2	26.0	7.9250
3	35.0	53.1000
4	35.0	8.0500
...	...	...
886	27.0	13.0000
887	19.0	30.0000
888	NaN	23.4500
889	26.0	30.0000
890	32.0	7.7500

891 rows × 2 columns

WAP to print out the string or object value columns in dtaa_sets df

In [31]:

```
df[df.dtypes[df.dtypes == 'object'].index]
```

Out[31]:

	Name	Sex	Ticket	Cabin	Embarked
0	Braund, Mr. Owen Harris	male	A/5 21171	NaN	S
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	PC 17599	C85	C
2	Heikkinen, Miss. Laina	female	STON/O2. 3101282	NaN	S
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	113803	C123	S
4	Allen, Mr. William Henry	male	373450	NaN	S
...	...	...	...	...	...
886	Montvila, Rev. Juozas	male	211536	NaN	S
887	Graham, Miss. Margaret Edith	female	112053	B42	S
888	Johnston, Miss. Catherine Helen "Carrie"	female	W./C. 6607	NaN	S
889	Behr, Mr. Karl Howell	male	111369	C148	C
890	Dooley, Mr. Patrick	male	370376	NaN	Q

891 rows × 5 columns

WAP to print out only first row along with all columns

In [6]:

```
df.loc[0]
```

Out[6]:

```
PassengerId                          1
Survived                             0
Pclass                               3
Name           Braund, Mr. Owen Harris
Sex                               male
Age                               22.0
SibSp                                1
Parch                                0
Ticket                       A/5 21171
Fare                              7.25
Cabin                              NaN
Embarked                             S
Name: 0, dtype: object
```

WAp to print out PassengerId & Name column wise first 10 rows

In [12]:

```
columns_wise = df[['PassengerId', 'Name']]
print(columns_wise.head(10))
```

```
   PassengerId                                               Name
0            1                            Braund, Mr. Owen Harris
1            2  Cumings, Mrs. John Bradley (Florence Briggs Th...
2            3                             Heikkinen, Miss. Laina
3            4       Futrelle, Mrs. Jacques Heath (Lily May Peel)
4            5                           Allen, Mr. William Henry
5            6                                   Moran, Mr. James
6            7                            McCarthy, Mr. Timothy J
7            8                     Palsson, Master. Gosta Leonard
8            9  Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
9           10                Nasser, Mrs. Nicholas (Adele Achem)
```

WAP to print out the 0,99,100,500,699 Number row

In [14]:

```
row = df.loc[[0,98,99,499,698]]
```

In [15]:
```
row
```

Out[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
98	99	1	2	Doling, Mrs. John T (Ada Julia Bone)	female	34.0	0	1	231919	23.0000	NaN	S
99	100	0	2	Kantor, Mr. Sinai	male	34.0	1	0	244367	26.0000	NaN	S
499	500	0	3	Svensson, Mr. Olof	male	24.0	0	0	350035	7.7958	NaN	S
698	699	0	1	Thayer, Mr. John Borland	male	49.0	1	1	17421	110.8833	C68	C

WAp to print out 1 to 100 row only using dataset df

In [19]:
```
range_1 = list(range(100))
subset = df.loc[range_1]
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43,
44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85,
86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]
```

In [22]:
```
subset
```

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
95	96	0	3	Shorney, Mr. Charles Joseph	male	NaN	0	0	374910	8.0500	NaN	S
96	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	PC 17754	34.6542	A5	C
97	98	1	1	Greenfield, Mr. William Bertram	male	23.0	0	1	PC 17759	63.3583	D10 D12	C
98	99	1	2	Doling, Mrs. John T (Ada Julia Bone)	female	34.0	0	1	231919	23.0000	NaN	S
99	100	0	2	Kantor, Mr. Sinai	male	34.0	1	0	244367	26.0000	NaN	S

100 rows × 12 columns

WAP to print out 10, 20, 40 rows from name, sex, cabin columns

In [24]:
```
df.iloc[[9,19,39],[3,4,10]]
```

Out[24]:

	Name	Sex	Cabin
9	Nasser, Mrs. Nicholas (Adele Achem)	female	NaN
19	Masselmani, Mrs. Fatima	female	NaN
39	Nicola-Yarred, Miss. Jamila	female	NaN

WAP to print out the 1 to 30 row from Passenger ID, Name, Sex, Age

In [26]:
```
df.loc[0:29,['PassengerId','Name','Sex','Age']]
```

Out[26]:

	PassengerId	Name	Sex	Age
0	1	Braund, Mr. Owen Harris	male	22.0
1	2	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
2	3	Heikkinen, Miss. Laina	female	26.0
3	4	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0
4	5	Allen, Mr. William Henry	male	35.0
5	6	Moran, Mr. James	male	NaN
6	7	McCarthy, Mr. Timothy J	male	54.0
7	8	Palsson, Master. Gosta Leonard	male	2.0
8	9	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0
9	10	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0
10	11	Sandstrom, Miss. Marguerite Rut	female	4.0
11	12	Bonnell, Miss. Elizabeth	female	58.0
12	13	Saundercock, Mr. William Henry	male	20.0
13	14	Andersson, Mr. Anders Johan	male	39.0
14	15	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0
15	16	Hewlett, Mrs. (Mary D Kingcome)	female	55.0
16	17	Rice, Master. Eugene	male	2.0
17	18	Williams, Mr. Charles Eugene	male	NaN
18	19	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0
19	20	Masselmani, Mrs. Fatima	female	NaN
20	21	Fynney, Mr. Joseph J	male	35.0
21	22	Beesley, Mr. Lawrence	male	34.0
22	23	McGowan, Miss. Anna "Annie"	female	15.0
23	24	Sloper, Mr. William Thompson	male	28.0
24	25	Palsson, Miss. Torborg Danira	female	8.0
25	26	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38.0
26	27	Emir, Mr. Farred Chehab	male	NaN
27	28	Fortune, Mr. Charles Alexander	male	19.0
28	29	O'Dwyer, Miss. Ellen "Nellie"	female	NaN
29	30	Todoroff, Mr. Lalio	male	NaN

WAP to print out what is the gender or sex 23 number row have

In [27]:
```
df.loc[22,'Sex']
```

Out[27]:
```
'female'
```

WAP to change the name of columns in your ddataset

```
In [11]: data = df.rename(columns = {'PassengerId':'Passenger_Id','Sex':'Gender'})
```

WAP to print count of row columns wise

```
In [6]: df.count()
```

```
Out[6]: PassengerId    891
        Survived       891
        Pclass         891
        Name           891
        Sex            891
        Age            714
        SibSp          891
        Parch          891
        Ticket         891
        Fare           891
        Cabin          204
        Embarked       889
        dtype: int64
```

Using Pandas Library working on Data Cleaning Part

```
In [12]: data # we chnage the columns name using variable of data that's why we change the dataset name
```

Out[12]:

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

891 rows × 12 columns

WAP to print how many null value have dataset

```
In [13]: data.isnull().sum()
```

```
Out[13]: Passenger_Id      0
         Survived          0
         Pclass            0
         Name              0
         Gender            0
         Age             177
         SibSp             0
         Parch             0
         Ticket            0
         Fare              0
         Cabin           687
         Embarked          2
         dtype: int64
```

WAP to check how many duplicated data have in data set

```
In [19]: pf = data[data.duplicated()]
```

```
In [20]: pf.shape
```

```
Out[20]: (0, 12)
```

WAP to print out the how many null value have dataset in percent wise

```
In [21]: data.isnull().sum()/(len(data))*100
```

```
Out[21]: Passenger_Id     0.000000
         Survived         0.000000
         Pclass           0.000000
         Name             0.000000
         Gender           0.000000
         Age             19.865320
         SibSp            0.000000
         Parch            0.000000
         Ticket           0.000000
         Fare             0.000000
         Cabin           77.104377
         Embarked         0.224467
         dtype: float64
```

WAP to delete all duplicated value using dataset name data

In [38]:
```
data.drop_duplicates()
```

Out[38]:

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

891 rows × 12 columns

WAP to delete null value from dataset

In [39]:
```
pf = data.dropna()
```

In [40]:
```
pf
```

Out[40]:

	Passenger_Id	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
...	...	...	...	...	...	...	...	...	...	...	...	...
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	S
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C

183 rows × 12 columns

WAP to fill null value on age columns

In [56]:
```
data['Age'] = data['Age'].fillna(22)
data['Age'] = data['Age'].round()
```

In [57]:
```
data
```

Out[57]:

	Passenger_Id	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	30.0	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [69]:
```
data['Cabin'] = data['Cabin'].fillna('D22')
```

In [72]:
```
data
```

Out[72]:

	Passenger_Id	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	C78	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	D22	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	D22	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	D22	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	30.0	1	2	W./C. 6607	23.4500	D22	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	D22	Q

891 rows × 12 columns

WAp to change data types age columns

In [73]:
```
data['Age'] = df['Age'].astype('int64') # You can change float annd sttring data types also
```

In [74]:
```
data
```

Out[74]:

	Passenger_Id	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	C78	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	D22	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	D22	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27	0	0	211536	13.0000	D22	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	30	1	2	W./C. 6607	23.4500	D22	S
889	890	1	1	Behr, Mr. Karl Howell	male	26	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32	0	0	370376	7.7500	D22	Q

891 rows × 12 columns

WAP to change column name upper and Lower and Capitalize

In [80]:
```
data.columns = [col.capitalize().replace(' ','_') for col in data.columns] # you use your data types what you change
```

In [81]:
```
data
```

Out[81]:

	Passenger_id	Survived	Pclass	Name	Gender	Age	Sibsp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	C78	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	D22	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	D22	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27	0	0	211536	13.0000	D22	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	30	1	2	W./C. 6607	23.4500	D22	S
889	890	1	1	Behr, Mr. Karl Howell	male	26	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32	0	0	370376	7.7500	D22	Q

891 rows × 12 columns

WAP to delete unnessery columns

In [83]:
```
data.drop(columns = ['Pclass'])
```

Out[83]:

	Passenger_id	Survived	Name	Gender	Age	Sibsp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	C78	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C
2	3	1	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	D22	S
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
4	5	0	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	D22	S
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	Montvila, Rev. Juozas	male	27	0	0	211536	13.0000	D22	S
887	888	1	Graham, Miss. Margaret Edith	female	19	0	0	112053	30.0000	B42	S
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	30	1	2	W./C. 6607	23.4500	D22	S
889	890	1	Behr, Mr. Karl Howell	male	26	0	0	111369	30.0000	C148	C
890	891	0	Dooley, Mr. Patrick	male	32	0	0	370376	7.7500	D22	Q

891 rows × 11 columns

WAP to clean Unnessery whitespace, special characters, or inconsistentent formating

In [84]:
```
data['Name'] = data['Name'].str.strip().str.lower()
```

In [85]:
```
data
```

Out[85]:

	Passenger_id	Survived	Pclass	Name	Gender	Age	Sibsp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	braund, mr. owen harris	male	22	1	0	A/5 21171	7.2500	C78	S
1	2	1	1	cumings, mrs. john bradley (florence briggs th...	female	38	1	0	PC 17599	71.2833	C85	C
2	3	1	3	heikkinen, miss. laina	female	26	0	0	STON/O2. 3101282	7.9250	D22	S
3	4	1	1	futrelle, mrs. jacques heath (lily may peel)	female	35	1	0	113803	53.1000	C123	S
4	5	0	3	allen, mr. william henry	male	35	0	0	373450	8.0500	D22	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	montvila, rev. juozas	male	27	0	0	211536	13.0000	D22	S
887	888	1	1	graham, miss. margaret edith	female	19	0	0	112053	30.0000	B42	S
888	889	0	3	johnston, miss. catherine helen "carrie"	female	30	1	2	W./C. 6607	23.4500	D22	S
889	890	1	1	behr, mr. karl howell	male	26	0	0	111369	30.0000	C148	C
890	891	0	3	dooley, mr. patrick	male	32	0	0	370376	7.7500	D22	Q

891 rows × 12 columns

WAP to handel catagorical data

In [ ]:
```
data['Name'] = data['Name'].astype('Category').cat.codes
```

WAP to reset indexing and reset what you transforming data

In [88]:
```
data.reset_index(drop = True)
```

Out[88]:

	Passenger_id	Survived	Pclass	Name	Gender	Age	Sibsp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	braund, mr. owen harris	male	22	1	0	A/5 21171	7.2500	C78	S
1	2	1	1	cumings, mrs. john bradley (florence briggs th...	female	38	1	0	PC 17599	71.2833	C85	C
2	3	1	3	heikkinen, miss. laina	female	26	0	0	STON/O2. 3101282	7.9250	D22	S
3	4	1	1	futrelle, mrs. jacques heath (lily may peel)	female	35	1	0	113803	53.1000	C123	S
4	5	0	3	allen, mr. william henry	male	35	0	0	373450	8.0500	D22	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	montvila, rev. juozas	male	27	0	0	211536	13.0000	D22	S
887	888	1	1	graham, miss. margaret edith	female	19	0	0	112053	30.0000	B42	S
888	889	0	3	johnston, miss. catherine helen "carrie"	female	30	1	2	W./C. 6607	23.4500	D22	S
889	890	1	1	behr, mr. karl howell	male	26	0	0	111369	30.0000	C148	C
890	891	0	3	dooley, mr. patrick	male	32	0	0	370376	7.7500	D22	Q

891 rows × 12 columns

WAP to save clean data in your file using another name

In [ ]:
```
df.to_csv('cleaned_data.csv',index = False)
```