Importing required Libraries

In [1]:
```
import pandas as pd
import seaborn as sb
```

In [2]:
```
df_train=pd.read_csv('train.csv')
df_test=pd.read_csv('test.csv')
```

In [3]:
```
1 df_train
```

Out[3]:

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

891 rows × 12 columns

In [4]:
```
df_train.info()
df_train.head()
```

```
<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
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cab
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
1	2	1	1	Cumings, 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	N
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N

In [5]:
```
1 df_test.info()
2 df_test.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Pclass       418 non-null    int64  
 2   Name         418 non-null    object 
 3   Sex          418 non-null    object 
 4   Age          332 non-null    float64
 5   SibSp        418 non-null    int64  
 6   Parch        418 non-null    int64  
 7   Ticket       418 non-null    object 
 8   Fare         417 non-null    float64
 9   Cabin        91 non-null     object 
 10  Embarked     418 non-null    object 
dtypes: float64(2), int64(4), object(5)
memory usage: 36.1+ KB
```

Out[5]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarke
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	

Data Cleaning

In [6]:
```
df_train['Age_Cat'] = df_train['Age'].apply(lambda Age:'Child' if Age<18 e
df_train['Age_Cat']
```

Out[6]:
```
0      Non_Child
1      Non_Child
2      Non_Child
3      Non_Child
4      Non_Child
         ...
886    Non_Child
887    Non_Child
888    Non_Child
889    Non_Child
890    Non_Child
Name: Age_Cat, Length: 891, dtype: object
```

In [7]:
```
v=df_train['Age'].mean()
df_train['Age'].fillna(v,inplace=True)
df_test['Age'].fillna(v,inplace=True)


v=df_train['Fare'].mean()
df_train['Fare'].fillna(v,inplace=True)
df_test['Fare'].fillna(v, inplace=True)
```

In [8]:
```
y_train=df_train['Survived']
X_train=df_train[['Pclass','SibSp','Parch','Age','Fare']]
```

In [9]:
```
from sklearn.ensemble import RandomForestClassifier
model=RandomForestClassifier()
```

In [10]:
```
model.fit(X_train,y_train)
```

Out[10]:
```
RandomForestClassifier
RandomForestClassifier()
```

In [11]:
```
model.score(X_train,y_train)
```

Out[11]:
```
0.957351290684624
```

In [12]:
```
# Make predictions
```

In [13]:
```
X_test=df_test[['Pclass','SibSp','Parch','Age','Fare']]
yp=model.predict(X_test)
```

In [14]:
```
1 X_test
```

Out[14]:

	Pclass	SibSp	Parch	Age	Fare
0	3	0	0	34.500000	7.8292
1	3	1	0	47.000000	7.0000
2	2	0	0	62.000000	9.6875
3	3	0	0	27.000000	8.6625
4	3	1	1	22.000000	12.2875
...	...	...	...	...	...
413	3	0	0	29.699118	8.0500
414	1	0	0	39.000000	108.9000
415	3	0	0	38.500000	7.2500
416	3	0	0	29.699118	8.0500
417	3	1	1	29.699118	22.3583

418 rows × 5 columns

In [15]:
```
1 # Making sub2 file
```

In [16]:
```
1 sub2=pd.DataFrame({'PassengerId':df_test['PassengerId'],
2                    'Survived':yp})
```

In [17]:
```
1 sub2.to_csv('sub2.csv',index=False)
```

In [18]:
```
1 import pickle
2 pickle.dump(model, open('titanic.pkl','wb'))
3 load_model=pickle.load(open ('titanic.pkl','rb'))
```

In [ ]:
```
1
```