1. BUSINESS PROBLEM UNDERSTANDING

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.simplefilter('ignore')
```

In [2]:
```
df=pd.read_csv('Titanic_train.csv')
df.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

2. DATA UNDERSTANDING

In [3]:
```
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 [4]:
```
df.shape
```

Out[4]:
```
(891, 12)
```

In [5]:
```
df['Survived'].value_counts()
```

Out[5]:
```
0    549
1    342
Name: Survived, dtype: int64
```

In [6]:
```
df['Pclass'].value_counts()
```

Out[6]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [7]:
```
df['Embarked'].value_counts()
```

Out[7]:
```
S    644
C    168
Q     77
Name: Embarked, dtype: int64
```

In [8]:
```
df['SibSp'].value_counts()
```

Out[8]:
```
0    608
1    209
2     28
4     18
3     16
8      7
5      5
Name: SibSp, dtype: int64
```

In [9]:
```
df['Age'].value_counts()
```

Out[9]:
```
24.00    30
22.00    27
18.00    26
19.00    25
28.00    25
         ..
36.50     1
55.50     1
0.92      1
23.50     1
74.00     1
Name: Age, Length: 88, dtype: int64
```

In [10]:
```
df['Parch'].value_counts()
```

Out[10]:
```
0    678
1    118
2     80
5      5
3      5
4      4
6      1
Name: Parch, dtype: int64
```

In [11]:
```
df['Sex'].value_counts()
```

Out[11]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [12]:
```
df = df.drop(['Name','Cabin','Ticket'],axis=1)
df.head()
```

Out[12]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0	3	male	22.0	1	0	7.2500	S
1	2	1	1	female	38.0	1	0	71.2833	C
2	3	1	3	female	26.0	0	0	7.9250	S
3	4	1	1	female	35.0	1	0	53.1000	S
4	5	0	3	male	35.0	0	0	8.0500	S

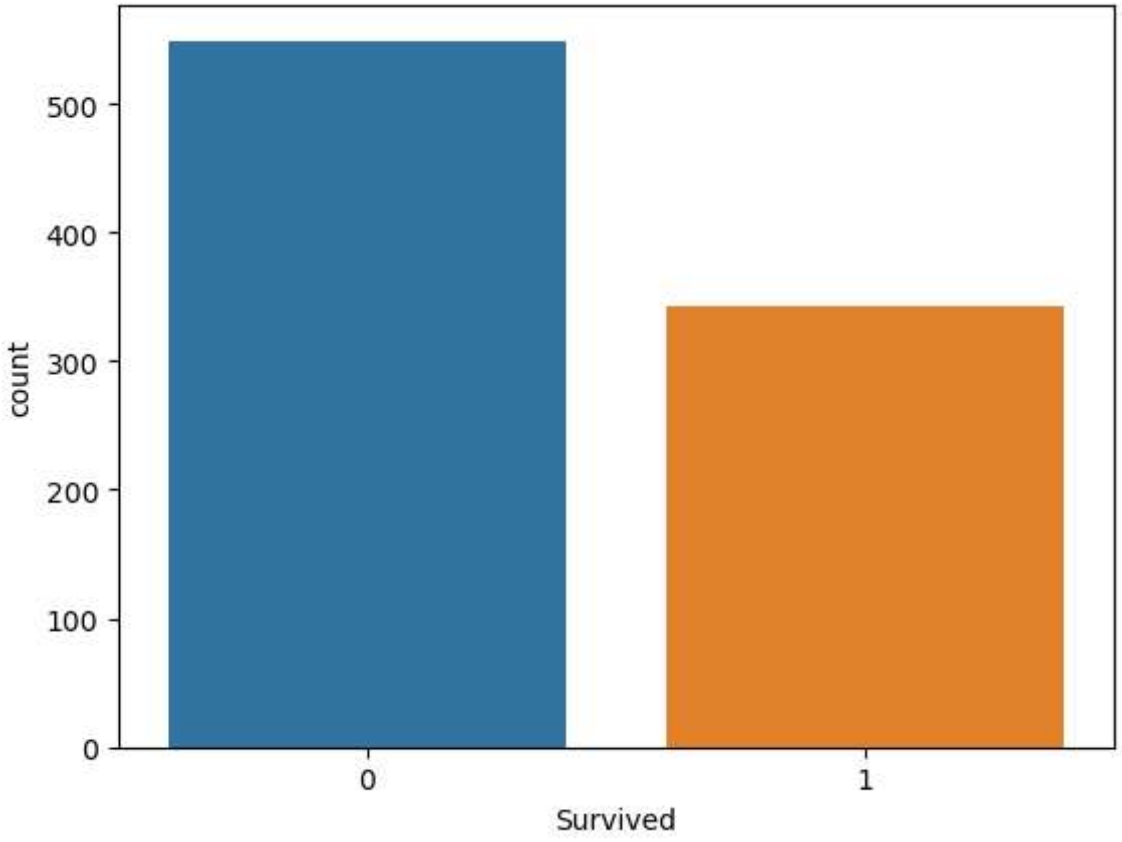
In [13]:
```
df.isnull().sum()
```

Out[13]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Fare             0
Embarked         2
dtype: int64
```

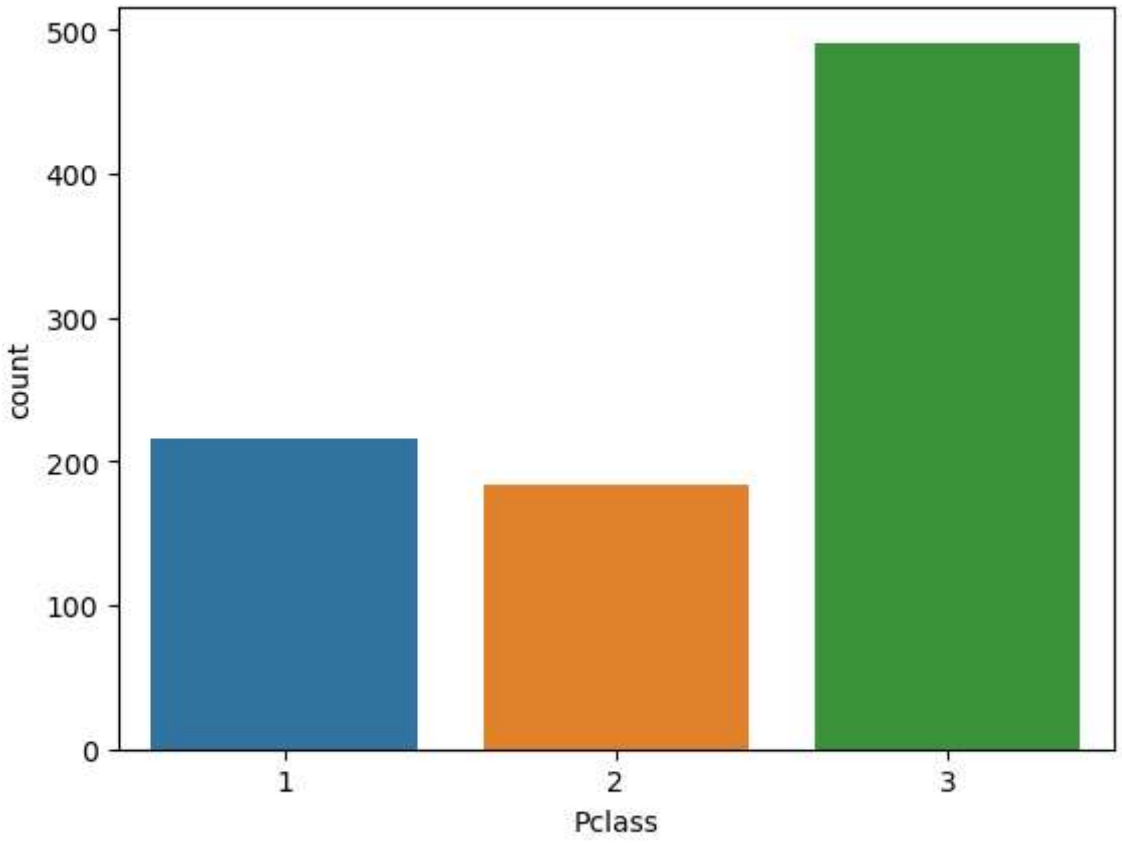
3. DATA PREPROCESSING

In [14]:
```
sns.countplot(x='Survived',data=df)
```

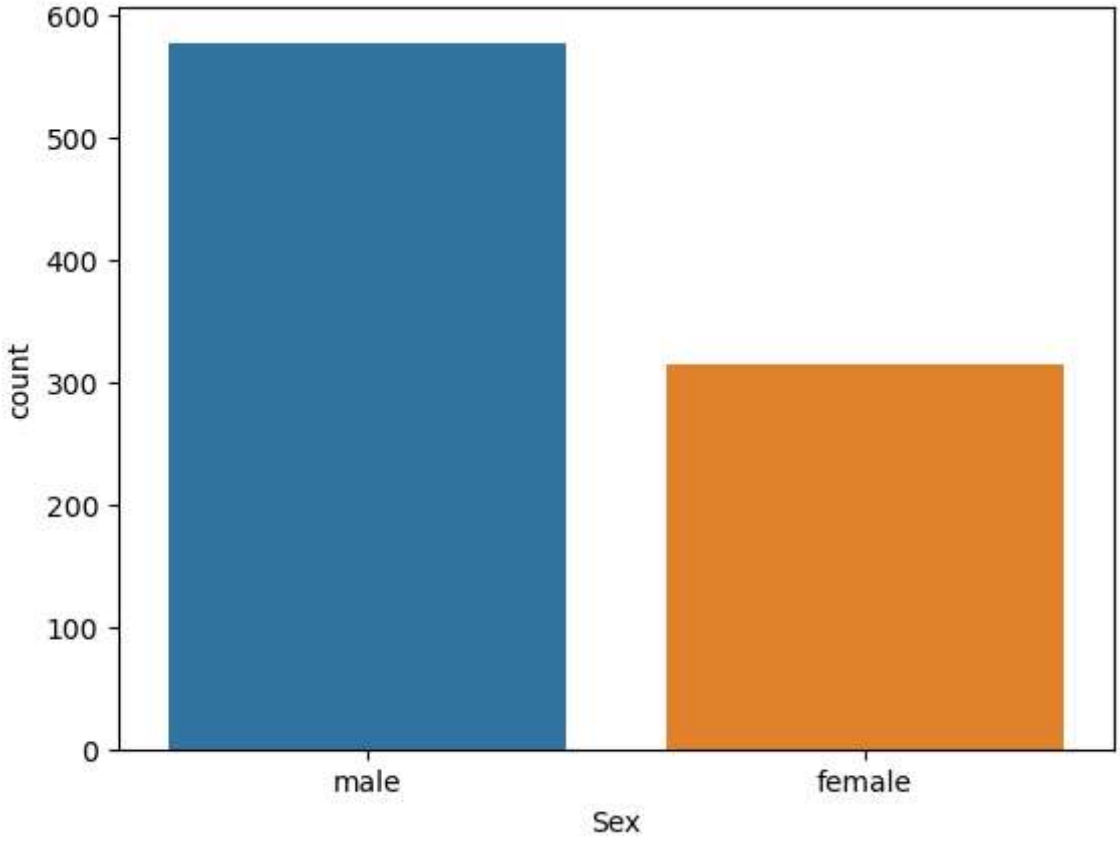
Out[14]: <Axes: xlabel='Survived', ylabel='count'>



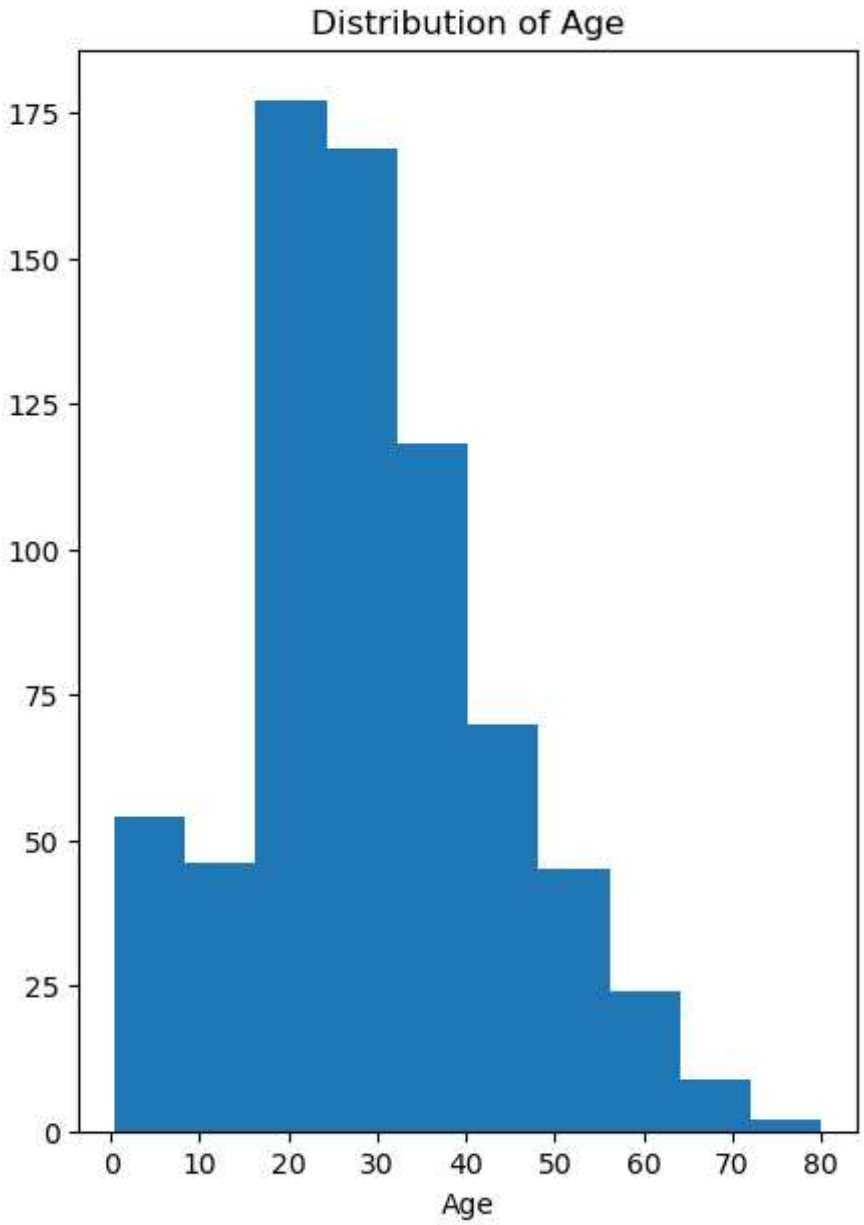
In [15]:
```
sns.countplot(x='Pclass',data=df)
plt.show()
```



In [16]:
```
sns.countplot(x='Sex',data=df)
plt.show()
```



In [17]:
```
plt.figure(figsize=(5,7))
plt.hist(df['Age'])
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.show()
```



4. DATA CLEANING

In [18]:
```
df.isnull().sum()
```

Out[18]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Fare             0
Embarked         2
dtype: int64
```

In [31]:
```
from sklearn.impute import SimpleImputer
median_imputer=SimpleImputer(strategy='median')

df['Age']=median_imputer.fit_transform(df[['Age']])
df.head()
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[31], line 1
----> 1 from sklearn.impute import SimpleImputer
      2 median_imputer=SimpleImputer(strategy='median')
      4 df['Age']=median_imputer.fit_transform(df[['Age']])

ModuleNotFoundError: No module named 'sklearn'
```

In [20]:
```
mode_imputer = SimpleImputer(strategy='most_frequent')

df['Embarked']=mode_imputer.fit_transform(df[['Embarked']])

df.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[20], line 1
----> 1 mode_imputer = SimpleImputer(strategy='most_frequent')
      3 df['Embarked']=mode_imputer.fit_transform(df[['Embarked']])
      5 df.head()

NameError: name 'SimpleImputer' is not defined
```

In [21]:
```
df.isnull().sum()
```

Out[21]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Fare             0
Embarked         2
dtype: int64
```

In [ ]:
```

```

In [22]:
```
df.Age=(df.Age-min(df.Age))/(max(df.Age)-min(df.Age))
df.Fare=(df.Fare-min(df.Fare))/(max(df.Fare)-min(df.Fare))

df.head()
```

Out[22]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0	3	male	0.271174	1	0	0.014151	S
1	2	1	1	female	0.472229	1	0	0.139136	C
2	3	1	3	female	0.321438	0	0	0.015469	S
3	4	1	1	female	0.434531	1	0	0.103644	S
4	5	0	3	male	0.434531	0	0	0.015713	S

In [23]:
```
df['Pclass'].value_counts()
```

Out[23]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [24]:
```
df=pd.get_dummies(df,drop_first=True)
df.head()
```

Out[24]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Emba
0	1	0	3	0.271174	1	0	0.014151	1	0	
1	2	1	1	0.472229	1	0	0.139136	0	0	
2	3	1	3	0.321438	0	0	0.015469	0	0	
3	4	1	1	0.434531	1	0	0.103644	0	0	
4	5	0	3	0.434531	0	0	0.015713	1	0	

In [25]:
```
x=df.drop('Survived',axis=1)
y=df['Survived']
```

In [26]:
```
x.head()
```

Out[26]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Embarked_S
0	1	3	0.271174	1	0	0.014151	1	0	1
1	2	1	0.472229	1	0	0.139136	0	0	0
2	3	3	0.321438	0	0	0.015469	0	0	1
3	4	1	0.434531	1	0	0.103644	0	0	1
4	5	3	0.434531	0	0	0.015713	1	0	1

In [27]:
```
y.head()
```

Out[27]:
```
0    0
1    1
2    1
3    1
4    0
Name: Survived, dtype: int64
```

In [28]:
```
df['Age'].skew()
```

Out[28]:
```
0.38910778230082615
```

In [29]:
```
df.skew()
```

Out[29]:
```
PassengerId    0.000000
Survived       0.478523
Pclass        -0.630548
Age            0.389108
SibSp          3.695352
Parch          2.749117
Fare           4.787317
Sex_male      -0.618921
Embarked_Q     2.948778
Embarked_S    -0.997083
dtype: float64
```

In [30]:
```
import statsmodels.formula.api as smf
model1=smf.ols('y~x',data=df).fit()
```

```
model1.summary()
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[30], line 1
----> 1 import statsmodels.formula.api as smf
      2 model1=smf.ols('y~x',data=df).fit()
      3 model1.summary()

ModuleNotFoundError: No module named 'statsmodels'
```

In [ ]:

5. MODELLING

In [105…

```
from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

In [ ]:

In [ ]:

APPLYING LOGISTIC REGRESSION

In [32]:

```
from sklearn.linear_model import LogisticRegression
LR_model=LogisticRegression()
LR_model.fit(x_train,y_train)
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[32], line 1
----> 1 from sklearn.linear_model import LogisticRegression
      2 LR_model=LogisticRegression()
      3 LR_model.fit(x_train,y_train)

ModuleNotFoundError: No module named 'sklearn'
```

In [107…

```
train_pred=LR_model.predict(x_train)
test_pred=LR_model.predict(x_test)
```

In [108…

```
print('Train accuracy:',LR_model.score(x_train,y_train))
print('Test accuracy:',LR_model.score(x_test,y_test))


from sklearn.model_selection import cross_val_score
scores=cross_val_score(LR_model,x,y,cv=5)
print('Cv score:',scores.mean())
```

```
Train accuracy: 0.800561797752809
Test accuracy: 0.7988826815642458
Cv score: 0.7890276818780994
```

In [ ]:

DECISION TREE

In [109…
```
from sklearn.tree import DecisionTreeClassifier
dt_default=DecisionTreeClassifier()
dt_default.fit(x_train,y_train)


pred_train = dt_default.predict(x_train)
base_pred = dt_default.predict(x_test)


from sklearn.metrics import accuracy_score
print('Train accuracy:',accuracy_score(pred_train,y_train))
print('Test_accuracy:',accuracy_score(base_pred,y_test))

from sklearn.model_selection import cross_val_score

scores= cross_val_score(dt_default,x,y,cv=5)
print('cross validation score:',scores.mean())
```

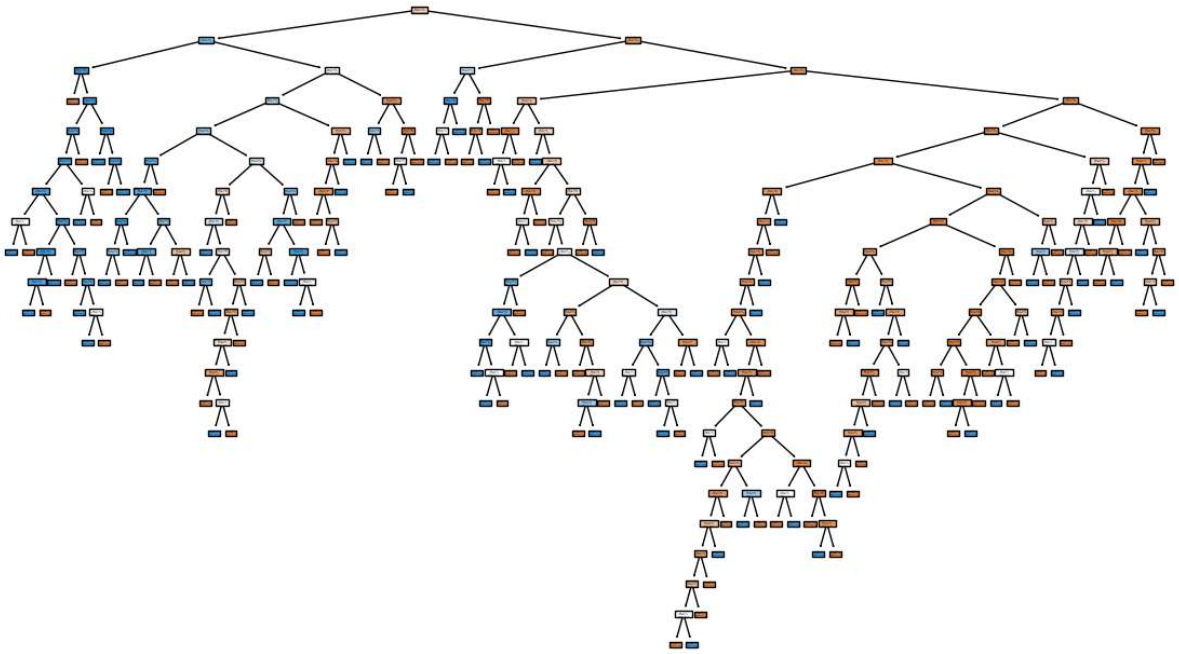
```
Train accuracy: 1.0
Test_accuracy: 0.8100558659217877
cross validation score: 0.756619170171364
```

In [110…
```
x.head()
```

Out[110]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Embarked_S
0	1	3	0.271174	1	0	0.014151	1	0	1
1	2	1	0.472229	1	0	0.139136	0	0	0
2	3	3	0.321438	0	0	0.015469	0	0	1
3	4	1	0.434531	1	0	0.103644	0	0	1
4	5	3	0.434531	0	0	0.015713	1	0	1

In [111…
```
from sklearn.tree import plot_tree
plt.figure(figsize=(14,8),dpi=100)
plot_tree(dt_default,filled=True,feature_names=x.columns)
plt.show()
```



In [112…
```
dt_default.feature_importances_
```

Out[112]:
```
array([0.21450982, 0.09221184, 0.15163065, 0.05588783, 0.01637632,
       0.15964223, 0.29364526, 0.00247538, 0.01362067])
```

In [113…
```
pd.DataFrame(index=x.columns,data=dt_default.feature_importances_)
```

Out[113]:

	0
PassengerId	0.214510
Pclass	0.092212
Age	0.151631
SibSp	0.055888
Parch	0.016376
Fare	0.159642
Sex_male	0.293645
Embarked_Q	0.002475
Embarked_S	0.013621

Hyper parameter Tuning

In [114…
```
from sklearn.model_selection import GridSearchCV

estimator=DecisionTreeClassifier(random_state=0)

param_grid={'criterion':['gini','entropy'],
            'max_depth':[1,2,3,4]}

grid=GridSearchCV(estimator,param_grid,scoring='accuracy',cv=5)

grid.fit(x_train,y_train)

grid.best_params_
```

Out[114]:
```
{'criterion': 'gini', 'max_depth': 4}
```

In [115…
```
from sklearn.tree import DecisionTreeClassifier
dt_bhp=DecisionTreeClassifier(criterion='gini',max_depth=4,random_state=0)
dt_bhp.fit(x_train,y_train)

y_pred_train = dt_bhp.predict(x_train)

y_pred_test = dt_bhp.predict(x_test)

print('Train accuracy:',accuracy_score(y_pred_train,y_train))
print('Test_accuracy:',accuracy_score(y_pred_test,y_test))

scores= cross_val_score(dt_bhp,x,y,cv=5)
print('cross validation score:',scores.mean())
```

```
Train accuracy: 0.8426966292134831
Test_accuracy: 0.8156424581005587
cross validation score: 0.7779423764986505
```

In [116…
```
dt_bhp.score(x_test,y_test)
```

Out[116]:
```
0.8156424581005587
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

TEST DATA

In [117…
```
df1=pd.read_csv('Titanic_test.csv')
df1.head()
```

Out[117]:

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

In [118…
```
df1.isnull().sum()
```

Out[118]:
```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

In [119…
```
from sklearn.impute import SimpleImputer
median_imputer=SimpleImputer(strategy='median')
```

```
df1['Age']=median_imputer.fit_transform(df1[['Age']])
df1.head()
```

Out[119]:

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

In [120…

```
df1=df1.drop(['Name','Ticket','Cabin'],axis=1)
df1.head()
```

Out[120]:

	PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	892	3	male	34.5	0	0	7.8292	Q
1	893	3	female	47.0	1	0	7.0000	S
2	894	2	male	62.0	0	0	9.6875	Q
3	895	3	male	27.0	0	0	8.6625	S
4	896	3	female	22.0	1	1	12.2875	S

In [121…

```
from sklearn.impute import SimpleImputer
median_imputer=SimpleImputer(strategy='median')

df1['Fare']=median_imputer.fit_transform(df1[['Fare']])
df1.head()
```

Out[121]:

	PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	892	3	male	34.5	0	0	7.8292	Q
1	893	3	female	47.0	1	0	7.0000	S
2	894	2	male	62.0	0	0	9.6875	Q
3	895	3	male	27.0	0	0	8.6625	S
4	896	3	female	22.0	1	1	12.2875	S

In [ ]:

In [122…
```
df1.isnull().sum()
```

Out[122]:
```
PassengerId    0
Pclass         0
Sex            0
Age            0
SibSp          0
Parch          0
Fare           0
Embarked       0
dtype: int64
```

In [ ]:

In [123…
```
df1.Age=(df1.Age-min(df1.Age))/(max(df1.Age)-min(df1.Age))
df1.Fare=(df1.Fare-min(df1.Fare))/(max(df1.Fare)-min(df1.Fare))

df1.describe()
```

Out[123]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	418.000000	418.000000	418.000000	418.000000
mean	1100.500000	2.265550	0.388096	0.447368	0.392344	0.069441
std	120.810458	0.841838	0.167530	0.896760	0.981429	0.109012
min	892.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	996.250000	1.000000	0.301068	0.000000	0.000000	0.015412
50%	1100.500000	3.000000	0.353818	0.000000	0.000000	0.028213
75%	1204.750000	3.000000	0.469207	1.000000	0.000000	0.061429
max	1309.000000	3.000000	1.000000	8.000000	9.000000	1.000000

In [124…
```
df1=pd.get_dummies(df1,drop_first=True)
df1.head()
```

Out[124]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Embarked_S
0	892	3	0.452723	0	0	0.015282	1	1	0
1	893	3	0.617566	1	0	0.013663	0	0	1
2	894	2	0.815377	0	0	0.018909	1	1	0
3	895	3	0.353818	0	0	0.016908	1	0	1
4	896	3	0.287881	1	1	0.023984	0	0	1

In [ ]:

In [125…
```
df1.head()
```

Out[125]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Embarked_S
0	892	3	0.452723	0	0	0.015282	1	1	0
1	893	3	0.617566	1	0	0.013663	0	0	1
2	894	2	0.815377	0	0	0.018909	1	1	0
3	895	3	0.353818	0	0	0.016908	1	0	1
4	896	3	0.287881	1	1	0.023984	0	0	1

In [ ]:

In [126…
```
Pred_LR= LR_model.predict(df1)
Pred_LR
```

Out[126]:
```
array([0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0,
       1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1,
       1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1,
       1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1,
       0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0,
       1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
       0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
       0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0,
       0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0,
       1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1,
       0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0],
      dtype=int64)
```

In [127…
```
LR_sub = pd.read_csv('gender_submission.csv')
LR_sub.head()
```

Out[127]:

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1

In [129…
```
LR_sub['Survived'] = Pred_LR
LR_sub.head()
```

Out[129]:

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1

In [130…
```
LR_sub.to_csv('submission_13.csv',index = False)
```

In [131…
```
#0.75358
```

In [ ]:

In [ ]:

In [134…
```
Pred_DT = dt_bhp.predict(df1)
Pred_DT
```

Out[134]:
```
array([0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1,
       1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1,
       1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1,
       1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1,
       0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0,
       1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
       0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1,
       0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0,
       0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0,
       1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1,
       0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0],
      dtype=int64)
```

In [135…
```
DT_sub = pd.read_csv('gender_submission.csv')
DT_sub.head()
```

Out[135]:

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1

In [136…
```
DT_sub['Survived'] = Pred_DT
DT_sub.head()
```

Out[136]:

	PassengerId	Survived
0	892	0
1	893	0
2	894	0
3	895	0
4	896	1

In [1]:
```
DT_sub.to_csv('submission_14.csv',index = False)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_12628\3138850983.py in <module>
----> 1 DT_sub.to_csv('submission_14.csv',index = False)

NameError: name 'DT_sub' is not defined
```

In [ ]:
```
#0.78229
```

In [ ]:
```

```