DAY 6 (Learning AI and ML) Mini project (Titanic Survial Predicition in python (Machine Learning project)

Required Libraries

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 1: Data frames Loading and importing

```python
df=pd.read_csv("C:/Users/DELL/Desktop/A/day 6(project)/train.csv")
```

```python
titanic_data = pd.read_csv("C:/Users/DELL/Desktop/A/day 6(project)/train.csv")
```

```python
titanic_data
```

Out[7]:

	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

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
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

In [16... `titanic_data.head()`

Out[16...

	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 [9]: `titanic_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
```

In [10... `titanic_data.describe()`

	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

shape of data sheet

```
In [11... titanic_data.shape
```

```
Out[11...  (891, 12)
```

```
In [12... titanic_data.columns
```

```
Out[12...  Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
                'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
              dtype='object')
```

Checking for Missing values

Step 2: Data Cleaning

```
In [13... titanic_data.isnull().sum()
```

```
Out[13...   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
```

```python
In [14...   print(titanic_data.columns)
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

```python
In [17...   titanic_data = pd.read_csv("C:/Users/DELL/Desktop/A/day 6(project)/train.csv")

            # Now safely drop columns
            titanic_data.drop(['Ticket', 'Cabin', 'Name'], axis=1, inplace=True, errors='ignore')
```

```python
In [21...   import matplotlib.pyplot as plt
            import seaborn as sns

            f, ax = plt.subplots(1, 2, figsize=(12, 4))

            titanic_data['Survived'].value_counts().plot.pie(
                explode=[0, 0.1], autopct='%1.1f%%', ax=ax[0], shadow=False)

            ax[0].set_title('Survivors (1) and the Dead (0)')
            ax[0].set_ylabel('')

            sns.countplot(x='Survived', data=titanic_data, ax=ax[1])
```
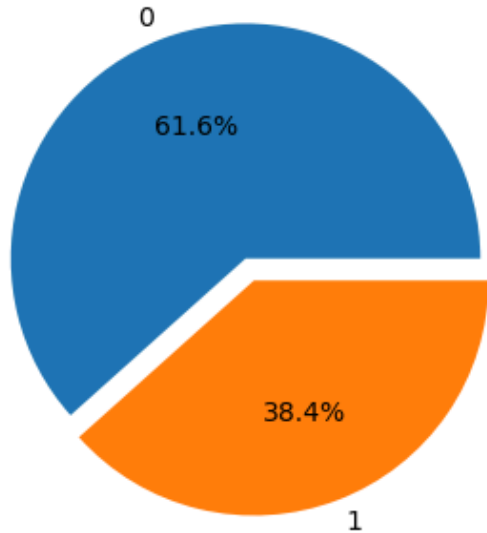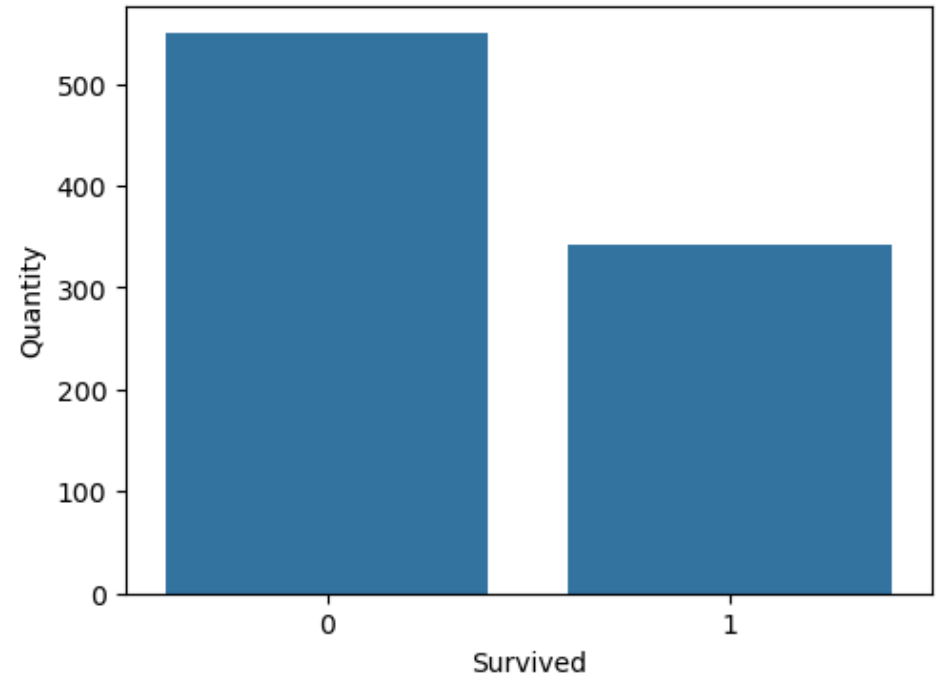
```
ax[1].set_ylabel('Quantity')
ax[1].set_title('Survivors (1) and the Dead (0)')

plt.show()
```
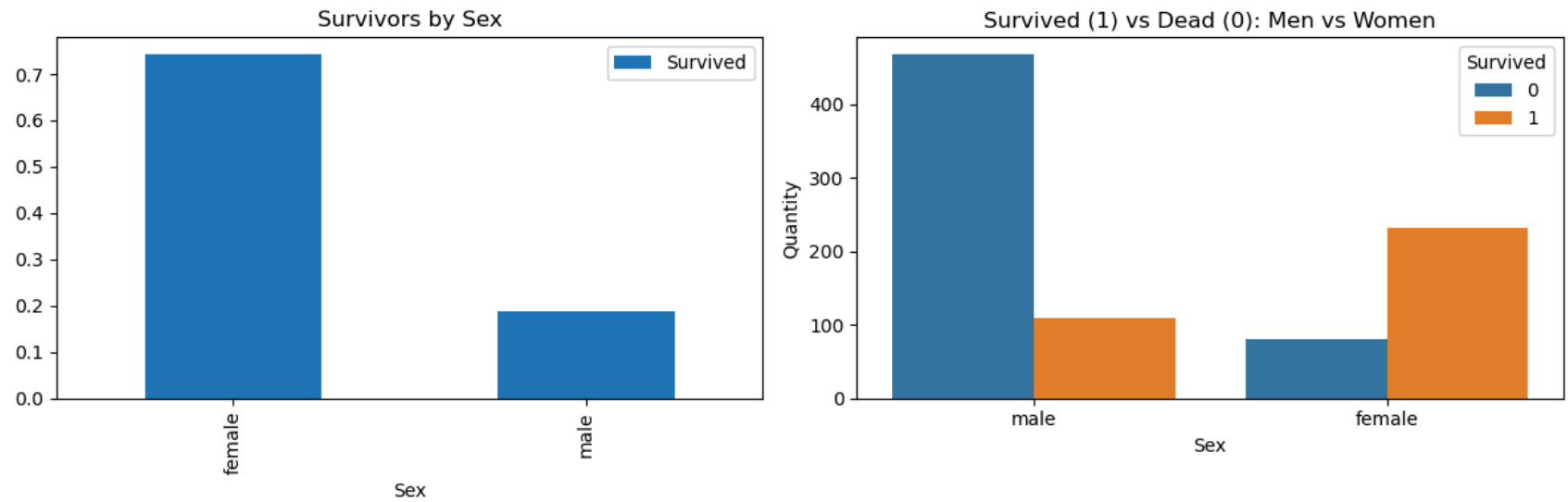


```
In [23…   f, ax = plt.subplots(1, 2, figsize=(12, 4))

          # Plot mean survival rate by gender
          titanic_data[['Sex', 'Survived']].groupby(['Sex']).mean().plot.bar(ax=ax[0])
          ax[0].set_title('Survivors by Sex')

          # Countplot with hue for Survived
          sns.countplot(x='Sex', hue='Survived', data=titanic_data, ax=ax[1])
          ax[1].set_ylabel('Quantity')
          ax[1].set_title('Survived (1) vs Dead (0): Men vs Women')
```

```
plt.tight_layout()
plt.show()
```


Step 3 : Feature Engineering (Optimizing Data for Model Training) next step