Roll No: 31440

DSBDAL Assignment-9 ¶

Importing libraries

In [1]:

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

Dataset Loading ,observation and preprocessing

In [2]:

```
df=pd.read_csv("titanic.csv")
df
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
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 [3]:

```
df.isnull().sum()
```

Out[3]:

```
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
```

In [4]:

```
df['Age'].value_counts()
```

Out[4]:

```
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 [5]:

```
df['Age'].mode()
```

Out[5]:

```
0    24.0
dtype: float64
```

In [6]:

```
df['Embarked'].value_counts()
```

Out[6]:

```
S    644
C    168
Q     77
Name: Embarked, dtype: int64
```

In [7]:

```
df["Age"].fillna(df["Age"].mean(),inplace=True)
df["Cabin"].fillna(0, inplace=True)
df["Embarked"].fillna("S", inplace=True)
df
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	211536	13
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0	112053	30
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	2	W./C. 6607	23
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0	111369	30
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	370376	7

891 rows × 12 columns

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

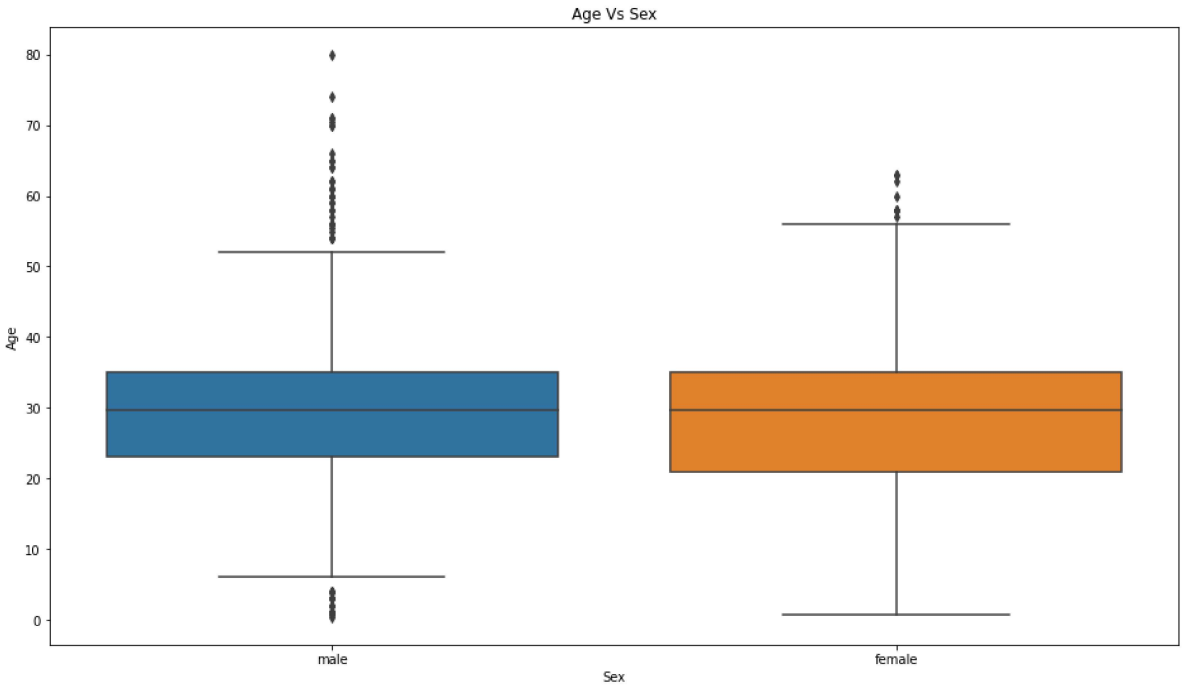
BoxPlot

In [9]:

```
plt.figure(figsize=(16,9))
sns.boxplot(y=df["Age"],x=df["Sex"]).set(title="Age Vs Sex")
```

Out[9]:

```
[Text(0.5, 1.0, 'Age Vs Sex')]
```

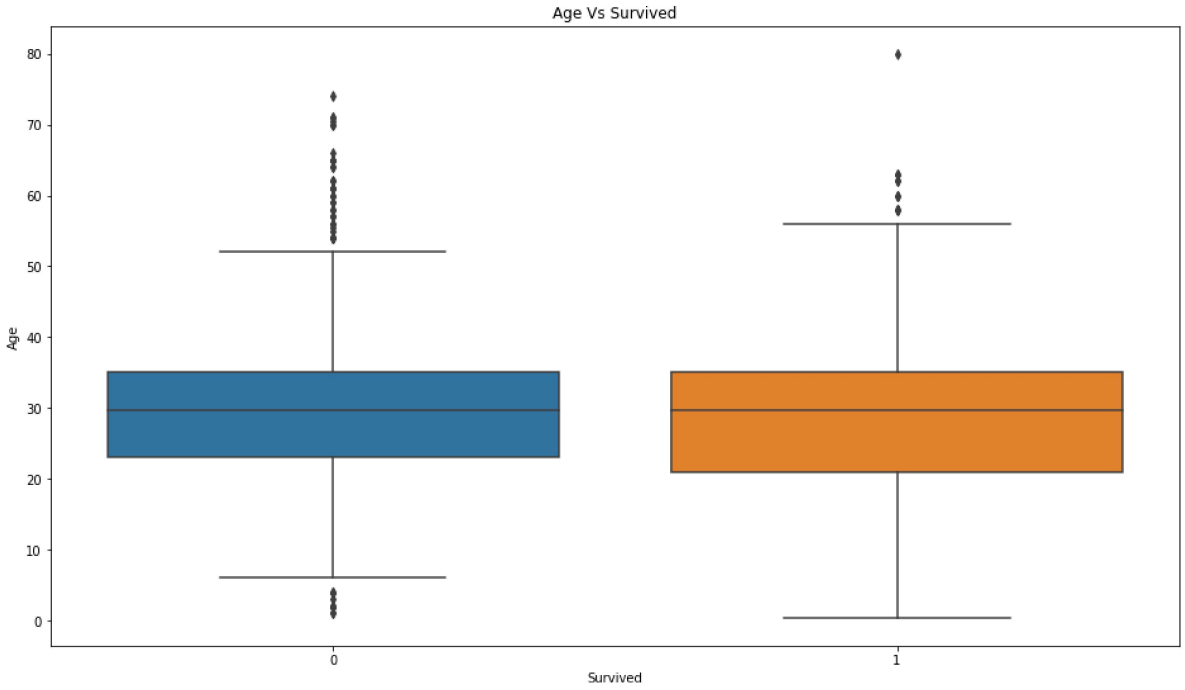


In [10]:

```
plt.figure(figsize=(16,9))
sns.boxplot(y=df["Age"],x=df["Survived"]).set(title="Age Vs Survived")
```

Out[10]:

```
[Text(0.5, 1.0, 'Age Vs Survived')]
```



In [11]:

```
plt.figure(figsize=(16,9))
sns.boxplot(y=df["Age"],x=df["Sex"],hue=df["Survived"]).set(title="Age Vs Sex with Survived
```

Out[11]:

```
[Text(0.5, 1.0, 'Age Vs Sex with Survived')]
```