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

In [2]:
```
# Importing DataSet
df = pd.read_csv("C:\\Users\HP\\OneDrive\\Desktop\\Internships\\Prodigy InfoTech\\Task-2\\titanic\\train.csv")
```

In [3]:
```
# Sample of the DataSet
df.head()
```

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

In [4]:
```
# Dealing with missing values
df.isnull().sum()
```

Out[4]:
```
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
```

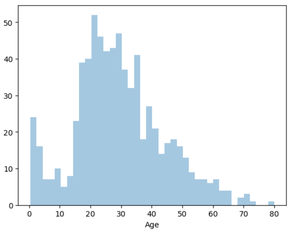
In [5]:
```
# visualize the missing values
sns.heatmap(df.isnull(),yticklabels=False,cbar=False)
plt.title("Columns With Missing Values")
```

Out[5]:
```
Text(0.5, 1.0, 'Columns With MIssing Values')
```

In [6]:
```
# Handling Missing Values from Age column
sns.distplot(df.Age,kde=False,bins=40)
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_22924\2347417904.py:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df.Age,kde=False,bins=40)
```

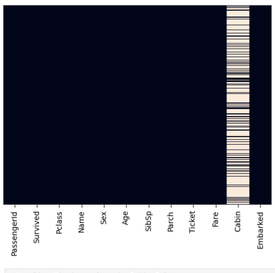
Out[6]:
```
<Axes: xlabel='Age'>
```



In [7]:
```
# Replacing the missing values with median in Age column
df["Age"].fillna(df["Age"].median(),inplace=True)
```

In [8]:
```
# Checking null values in Age column
sns.heatmap(df.isnull(),yticklabels=False,cbar=False)
```

Out[8]:
```
<Axes: >
```



In [9]:
```
# Handling Missing values in Cabin Colum n
df.drop("Cabin",axis=1,inplace=True)
```

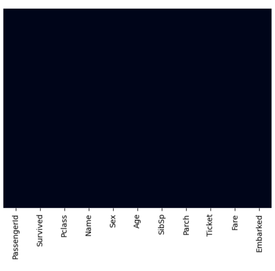
In [10]:
```
df.head()
```

Out[10]:

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

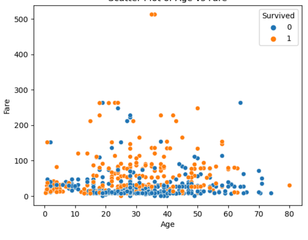
In [11]:
```
sns.heatmap(df.isnull(),yticklabels=False,cbar=False)
```

Out[11]:
```
<Axes: >
```



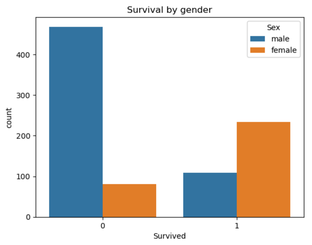
In [19]:
```
# Identifying patterns and trends in the data
sns.scatterplot(data=df,x="Age",y="Fare",hue="Survived")
plt.title("Scatter Plot of Age vs Fare")
```

Out[19]:
```
Text(0.5, 1.0, 'Scatter Plot of Age vs Fare')
```



In [12]:
```
sns.countplot(x="Survived",hue="Sex",data=df)
plt.title("Survival by gender")
```

Out[12]:
```
Text(0.5, 1.0, 'Survival by gender')
```


Survival by gender
Sex
male
female
count
0
100
200
300
400
0
1
Survived

In [ ]: