In [77]:
```
#importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [78]:
```
#load titanic
df=pd.read_csv("titanic.csv")
```

In [79]:
```
df.head(3)
```

Out[79]:

	PassengerId	Name	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
0	1	Braund, Mr. Owen Harris	3	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	Heikkinen, Miss. Laina	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	

In [80]:
```
df.tail(3)
```

Out[80]:

	PassengerId	Name	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
888	889	Johnston, Miss. Catherine Helen "Carrie"	3	female	NaN	1	2	W./C. 6607	23.45	NaN	
889	890	Behr, Mr. Karl Howell	1	male	26.0	0	0	111369	30.00	C148	
890	891	Dooley, Mr. Patrick	3	male	32.0	0	0	370376	7.75	NaN	

In [81]:
```
#basic info
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   Name         891 non-null    object 
 2   Pclass       891 non-null    int64  
 3   Sex          891 non-null    object 
 4   Age          714 non-null    float64
 5   SibSp        891 non-null    int64  
 6   Parch        891 non-null    int64  
 7   Ticket       891 non-null    object 
 8   Fare         891 non-null    float64
 9   Cabin        204 non-null    object 
 10  Embarked     889 non-null    object 
 11  Survived     891 non-null    int64  
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [82]:
```
#description
df.describe()
```

Out[82]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Survivec
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	2.308642	29.699118	0.523008	0.381594	32.204208	0.383838
std	257.353842	0.836071	14.526497	1.102743	0.806057	49.693429	0.486592
min	1.000000	1.000000	0.420000	0.000000	0.000000	0.000000	0.000000
25%	223.500000	2.000000	20.125000	0.000000	0.000000	7.910400	0.000000
50%	446.000000	3.000000	28.000000	0.000000	0.000000	14.454200	0.000000
75%	668.500000	3.000000	38.000000	1.000000	0.000000	31.000000	1.000000
max	891.000000	3.000000	80.000000	8.000000	6.000000	512.329200	1.000000

In [83]:
```
#Checking Missing values
df.isnull().sum()
```

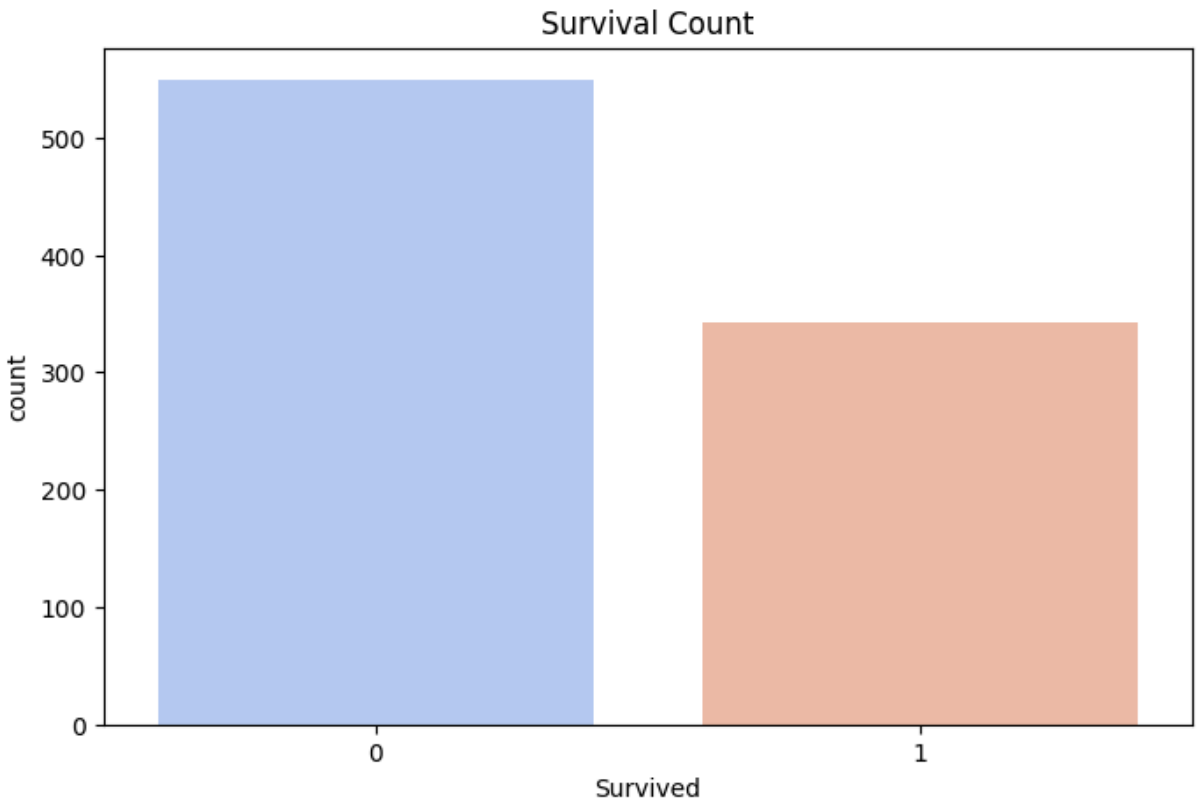
Out[83]:
```
PassengerId      0
Name             0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
Survived         0
dtype: int64
```

In [84]:
```
#Handle Missing Values
df['Age'].fillna(df['Age'].median(), inplace=True)  # Fill missing age with median
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)  # Fill missing Embar
df.drop(columns=['Cabin'], inplace=True)  # Drop 'Cabin' due to too many missing va
```

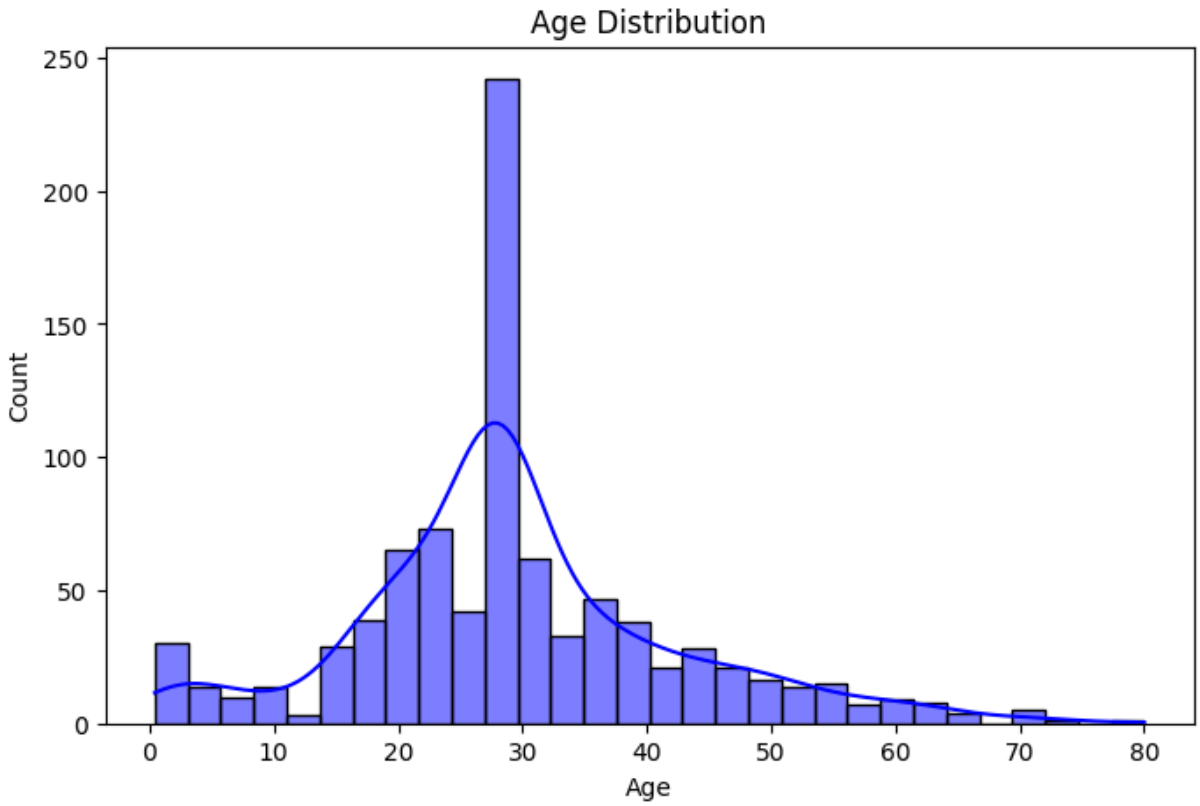
```
C:\Users\Hp\AppData\Local\Temp\ipykernel_7080\3438732544.py:2: FutureWarning: A valu
e is trying to be set on a copy of a DataFrame or Series through chained assignment
using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because
the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method
({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform
the operation inplace on the original object.


  df['Age'].fillna(df['Age'].median(), inplace=True)  # Fill missing age with median
C:\Users\Hp\AppData\Local\Temp\ipykernel_7080\3438732544.py:3: FutureWarning: A valu
e is trying to be set on a copy of a DataFrame or Series through chained assignment
using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because
the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method
({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform
the operation inplace on the original object.


  df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)  # Fill missing Emba
rked with mode
```

In [85]:
```
#Data Visualization
plt.figure(figsize=(8,5))
sns.countplot(data=df, x='Survived', palette='coolwarm')
# Survival count
plt.title('Survival Count')
plt.show()
```

```
C:\Users\Hp\AppData\Local\Temp\ipykernel_7080\1607887851.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1
4.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(data=df, x='Survived', palette='coolwarm')
```

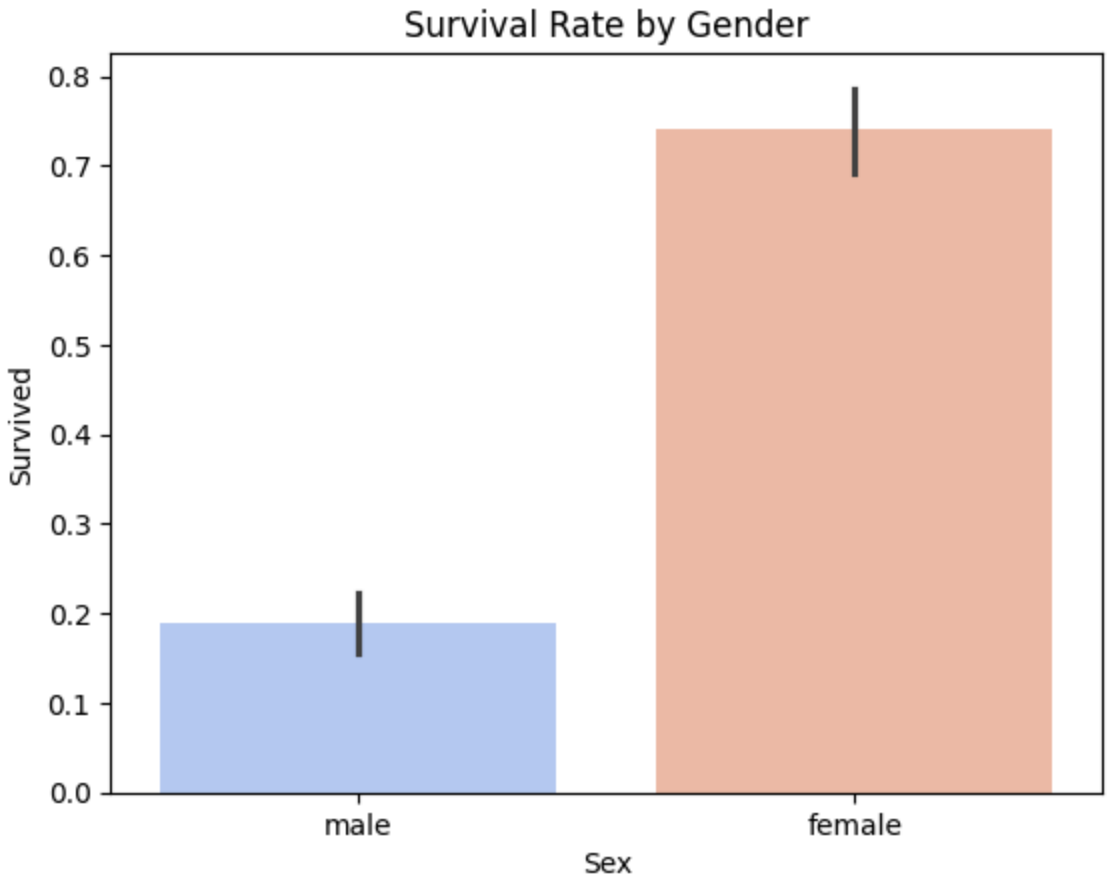


In [86]:
```
#we use hisplot to visualise by dsitributing  the value eg age , fare
plt.figure(figsize=(8,5))
sns.histplot(df['Age'], bins=30, kde=True, color='blue')  # Age distribution
plt.title('Age Distribution')
plt.show()
```



In [87]:
```
#barplot :here we will do camparision between sex & survivid for calculating the su
sns.barplot(x='Sex', y='Survived', data=df, palette='coolwarm')
plt.title('Survival Rate by Gender')
plt.show()
```

```
C:\Users\Hp\AppData\Local\Temp\ipykernel_7080\2271513014.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1
4.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x='Sex', y='Survived', data=df, palette='coolwarm')
```



In [88]:
```
# Step 4: Feature Engineering
# Convert categorical variables into numerical
df['Sex'] = df['Sex'].map({'male': 0, 'female': 1})
df = pd.get_dummies(df, columns=['Embarked'], drop_first=True)
```

In [89]:
```
# Final look at the cleaned dataset
print(df.head())
```

```
   PassengerId                                               Name  Pclass  \
0            1                            Braund, Mr. Owen Harris       3
1            2  Cumings, Mrs. John Bradley (Florence Briggs Th...       1
2            3                             Heikkinen, Miss. Laina       3
3            4       Futrelle, Mrs. Jacques Heath (Lily May Peel)       1
4            5                           Allen, Mr. William Henry       3

   Sex   Age  SibSp  Parch            Ticket     Fare  Survived  Embarked_Q  \
0    0  22.0      1      0         A/5 21171   7.2500         0       False
1    1  38.0      1      0          PC 17599  71.2833         1       False
2    1  26.0      0      0  STON/O2. 3101282   7.9250         1       False
3    1  35.0      1      0            113803  53.1000         1       False
4    0  35.0      0      0            373450   8.0500         0       False

   Embarked_S
0        True
1       False
2        True
3        True
4        True
```

In [90]:
```
#thnku
```