In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_absolute_error
from sklearn.ensemble import RandomForestClassifier
```

In [3]:
```
df= pd.read_csv("Titanic-Dataset.csv")
```

In [4]:
```
df
```

Out[4]:

| | 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 |
| **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 [5]:
```
df.shape
```

Out[5]:
```
(891, 12)
```

In [6]:
```
df.head()
```

Out[6]:

| | 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 [7]:
```
df.tail()
```

Out[7]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13.00 | NaN | S |
| **887** | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 19.0 | 0 | 0 | 112053 | 30.00 | B42 | S |
| **888** | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" | female | NaN | 1 | 2 | W./C. 6607 | 23.45 | NaN | S |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30.00 | C148 | C |
| **890** | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 32.0 | 0 | 0 | 370376 | 7.75 | NaN | Q |

In [8]:
```
df.describe()
```

Out[8]:

| | 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 |

In [9]:
```
df.duplicated().sum()
```

Out[9]:
```
0
```

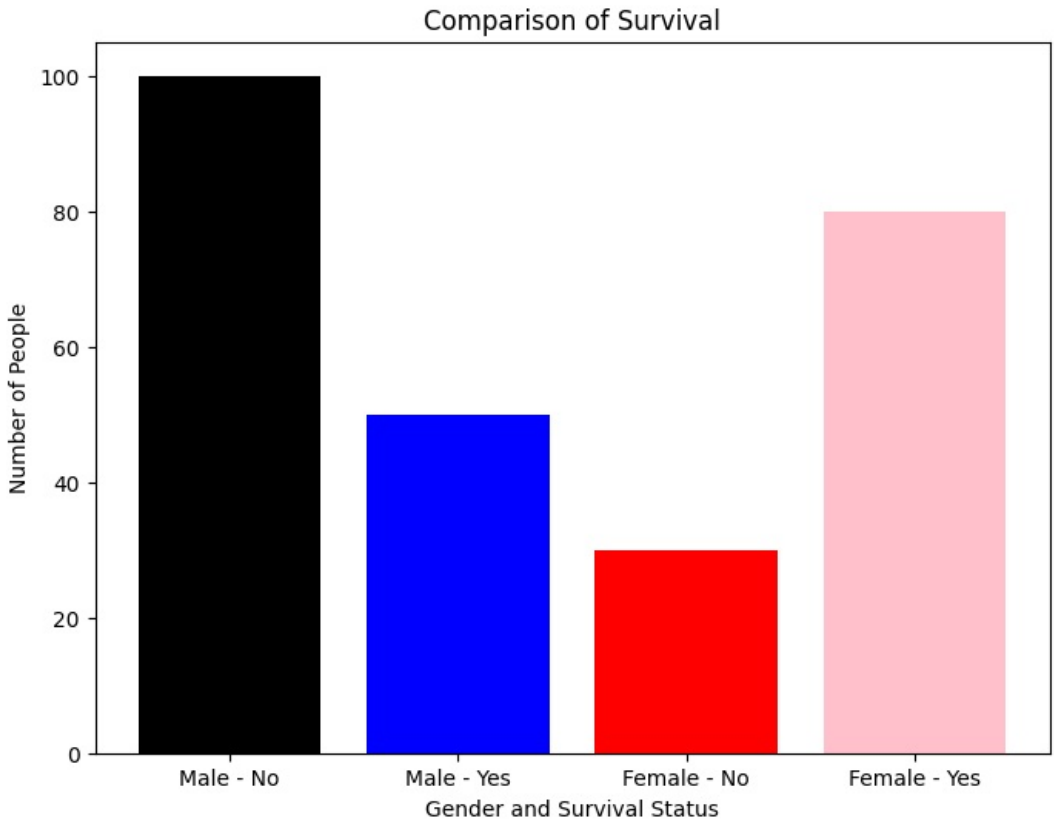
In [10]:
```
Survived = df['Survived'].value_counts().reset_index()
```

In [11]:
```
Survived
```

Out[11]:

| | Survived | count |
|---|---|---|
| **0** | 0 | 549 |
| **1** | 1 | 342 |

In [13]:
```
data = {'Survived': ['Male - No', 'Male - Yes', 'Female - No', 'Female - Yes'],
        'Counts': [100, 50, 30, 80]}  # replace with actual counts
Survived = pd.DataFrame(data)

plt.figure(figsize=(8, 6))
plt.bar(Survived['Survived'], Survived['Counts'],color=["Black","blue","red","pink"])
plt.xticks(Survived['Survived'])
plt.title('Comparison of Survival')
plt.xlabel('Gender and Survival Status')
plt.ylabel('Number of People')
plt.show()
```




In [14]:
```
df.drop(['PassengerId','Name','SibSp','Parch','Ticket','Cabin','Embarked'],axis='columns',inplace=True)
df.head()
```

Out[14]:

| | Survived | Pclass | Sex | Age | Fare |
|---|---|---|---|---|---|
| **0** | 0 | 3 | male | 22.0 | 7.2500 |
| **1** | 1 | 1 | female | 38.0 | 71.2833 |
| **2** | 1 | 3 | female | 26.0 | 7.9250 |
| **3** | 1 | 1 | female | 35.0 | 53.1000 |
| **4** | 0 | 3 | male | 35.0 | 8.0500 |

In [15]:
```
inputs = df.drop('Survived',axis='columns')
target = df['Survived']
```

In [16]:
```
target
```

Out[16]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

In [17]:
```
sex=pd.get_dummies(inputs.Sex)
sex.head()
```

Out[17]:

| | female | male |
|---|---|---|
| **0** | False | True |
| **1** | True | False |
| **2** | True | False |
| **3** | True | False |
| **4** | False | True |

In [18]:
```
inputs=pd.concat([inputs,sex],axis="columns")
inputs.head()
```

Out[18]:

| | Pclass | Sex | Age | Fare | female | male |
|---|---|---|---|---|---|---|
| **0** | 3 | male | 22.0 | 7.2500 | False | True |
| **1** | 1 | female | 38.0 | 71.2833 | True | False |
| **2** | 3 | female | 26.0 | 7.9250 | True | False |
| **3** | 1 | female | 35.0 | 53.1000 | True | False |
| **4** | 3 | male | 35.0 | 8.0500 | False | True |

In [19]:
```
inputs.drop(["Sex"],axis="columns",inplace=True)
```

In [20]:
```
inputs.head()
```

Out[20]:

| | Pclass | Age | Fare | female | male |
|---|---|---|---|---|---|
| **0** | 3 | 22.0 | 7.2500 | False | True |
| **1** | 1 | 38.0 | 71.2833 | True | False |
| **2** | 3 | 26.0 | 7.9250 | True | False |
| **3** | 1 | 35.0 | 53.1000 | True | False |
| **4** | 3 | 35.0 | 8.0500 | False | True |

In [21]:
```
inputs.isna().sum()
```

Out[21]:
```
Pclass      0
Age       177
Fare        0
female      0
male        0
dtype: int64
```

In [22]:
```
inputs.Age = inputs.Age.fillna(inputs.Age.mean())
```

```
inputs.head()
```

Out[22]:

| | Pclass | Age | Fare | female | male |
|---|---|---|---|---|---|
| **0** | 3 | 22.0 | 7.2500 | False | True |
| **1** | 1 | 38.0 | 71.2833 | True | False |
| **2** | 3 | 26.0 | 7.9250 | True | False |
| **3** | 1 | 35.0 | 53.1000 | True | False |
| **4** | 3 | 35.0 | 8.0500 | False | True |

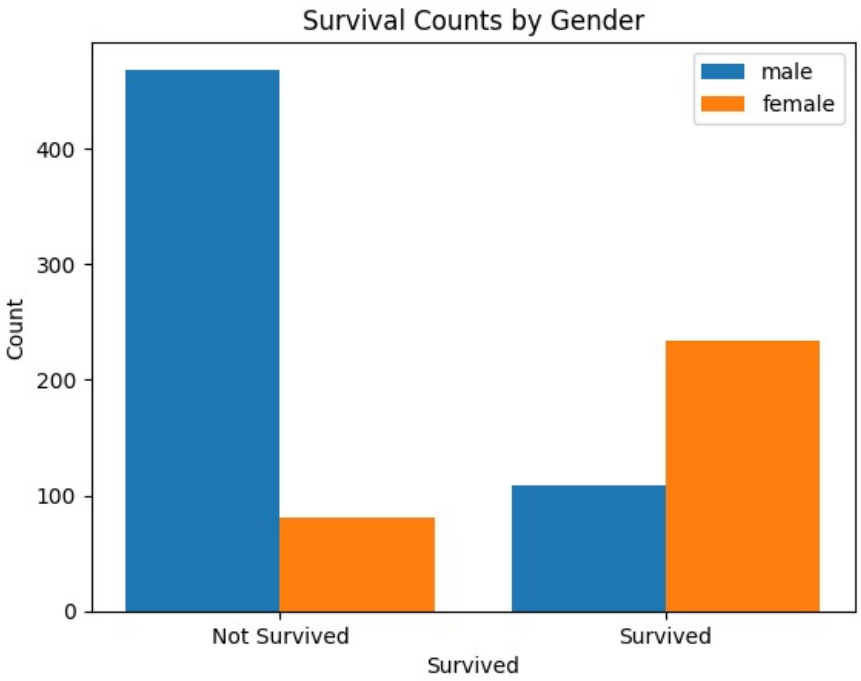
In [23]:
```
inputs.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Pclass  891 non-null    int64
 1   Age     891 non-null    float64
 2   Fare    891 non-null    float64
 3   female  891 non-null    bool
 4   male    891 non-null    bool
dtypes: bool(2), float64(2), int64(1)
memory usage: 22.8 KB
```

In [24]:
```
inputs.isna().sum()
```

Out[24]:
```
Pclass    0
Age       0
Fare      0
female    0
male      0
dtype: int64
```

In [25]:
```
counts = df.groupby(['Survived', 'Sex']).size().unstack().fillna(0)

# Define the bar width
bar_width = 0.40
index = counts.index

# Plotting
fig, ax = plt.subplots()

# Plot bars for each Sex
bar1 = ax.bar(index - bar_width/2, counts['male'], bar_width, label='male')
bar2 = ax.bar(index + bar_width/2, counts['female'], bar_width, label='female')

# Setting labels and title
ax.set_xlabel('Survived')
ax.set_ylabel('Count')
ax.set_title('Survival Counts by Gender')
ax.set_xticks(index)
ax.set_xticklabels(['Not Survived', 'Survived'])
ax.legend()

# Display the plot
plt.show()
```




In [26]:
```
X_train, X_test, y_train, y_test=train_test_split(inputs,target,test_size=0.2)
```

In [27]:
```
X_train
```

Out[27]:

| | Pclass | Age | Fare | female | male |
|---|---|---|---|---|---|
| **781** | 1 | 17.000000 | 57.0000 | True | False |
| **75** | 3 | 25.000000 | 7.6500 | False | True |
| **733** | 2 | 23.000000 | 13.0000 | False | True |
| **568** | 3 | 29.699118 | 7.2292 | False | True |
| **668** | 3 | 43.000000 | 8.0500 | False | True |
| **...** | ... | ... | ... | ... | ... |
| **328** | 3 | 31.000000 | 20.5250 | True | False |
| **625** | 1 | 61.000000 | 32.3208 | False | True |
| **25** | 3 | 38.000000 | 31.3875 | True | False |
| **572** | 1 | 36.000000 | 26.3875 | False | True |
| **30** | 1 | 40.000000 | 27.7208 | False | True |

712 rows × 5 columns

In [28]:
```
X_test
```

Out[28]:

| | Pclass | Age | Fare | female | male |
|---|---|---|---|---|---|
| **699** | 3 | 42.000000 | 7.6500 | False | True |
| **367** | 3 | 29.699118 | 7.2292 | True | False |
| **354** | 3 | 29.699118 | 7.2250 | False | True |
| **202** | 3 | 34.000000 | 6.4958 | False | True |
| **63** | 3 | 4.000000 | 27.9000 | False | True |
| **...** | ... | ... | ... | ... | ... |
| **417** | 2 | 18.000000 | 13.0000 | True | False |
| **406** | 3 | 51.000000 | 7.7500 | False | True |
| **108** | 3 | 38.000000 | 7.8958 | False | True |
| **523** | 1 | 44.000000 | 57.9792 | True | False |
| **737** | 1 | 35.000000 | 512.3292 | False | True |

179 rows × 5 columns

In [29]:
```
y_train
```

Out[29]:
```
781    1
75     0
733    0
568    0
668    0
      ..
328    1
625    0
25     1
572    1
30     0
Name: Survived, Length: 712, dtype: int64
```

In [30]:
```
y_test
```

Out[30]:
```
699    0
367    1
354    0
202    0
63     0
      ..
417    1
406    0
108    0
523    1
737    1
Name: Survived, Length: 179, dtype: int64
```

In [31]:
```
inputs.corr()
```

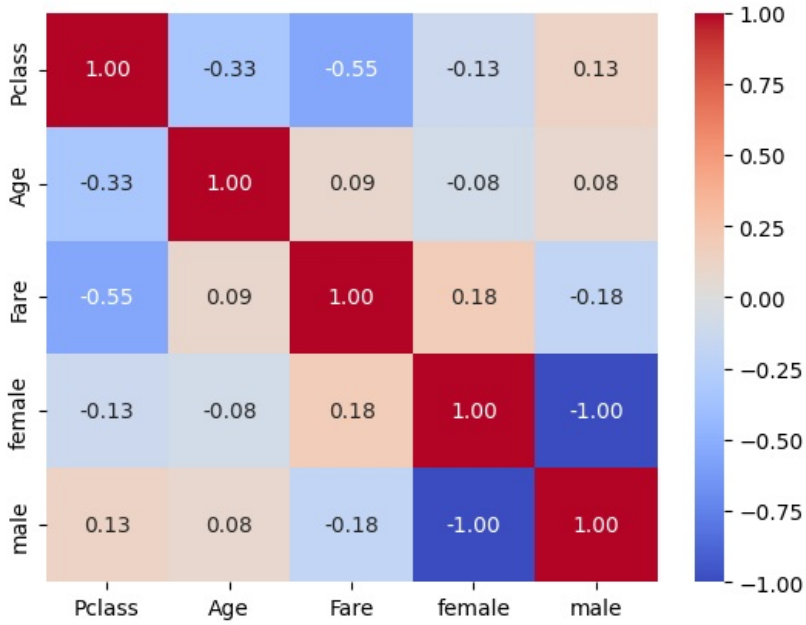
Out[31]:

| | Pclass | Age | Fare | female | male |
|---|---|---|---|---|---|
| **Pclass** | 1.000000 | -0.331339 | -0.549500 | -0.131900 | 0.131900 |
| **Age** | -0.331339 | 1.000000 | 0.091566 | -0.084153 | 0.084153 |
| **Fare** | -0.549500 | 0.091566 | 1.000000 | 0.182333 | -0.182333 |
| **female** | -0.131900 | -0.084153 | 0.182333 | 1.000000 | -1.000000 |
| **male** | 0.131900 | 0.084153 | -0.182333 | -1.000000 | 1.000000 |

In [32]:
```
import seaborn as sns
```

In [33]:
```
sns.heatmap(inputs.corr(), annot=True, cmap='coolwarm', fmt=".2f")
```

Out[33]:
```
<Axes: >
```




In [34]:
```
model=RandomForestClassifier()
```

In [35]:
```
model.fit(X_train,y_train)
```

Out[35]:
```
▾ RandomForestClassifier
RandomForestClassifier()
```

In [36]:
```
model.score(X_test,y_test)
```

Out[36]: 0.8156424581005587

In [37]:
```
pre=model.predict(X_test)
```

In [38]:
```
matrices=r2_score(pre,y_test)
matrices
```

Out[38]: 0.17935537649347066

In [ ]:

In [ ]: