In [6]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [7]:
```
titanic_train = pd.read_csv('titanic_train.csv')
titanic_test = pd.read_csv('titanic_test.csv')
```

In [8]:
```
titanic_train.head()
```

Out[8]:

	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

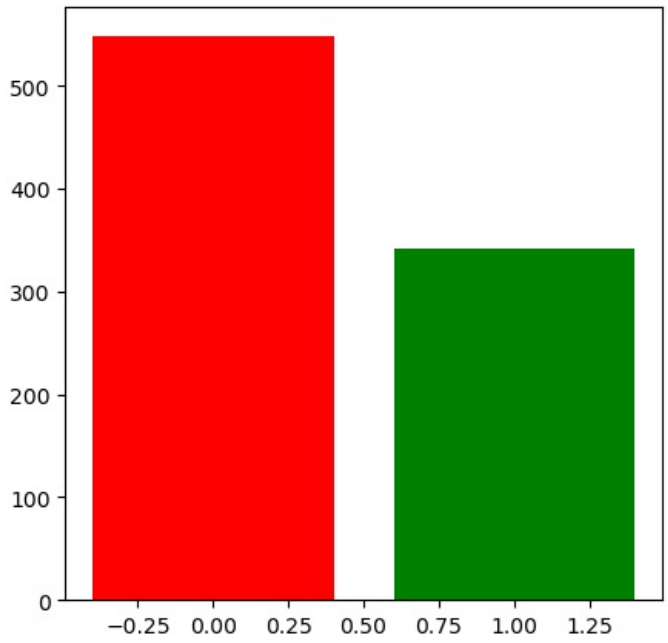
In [9]:
```
titanic_train.shape
```

Out[9]:
```
(891, 12)
```

In [10]:
```
titanic_train['Survived'].value_counts()
```

Out[10]:
```
0    549
1    342
Name: Survived, dtype: int64
```

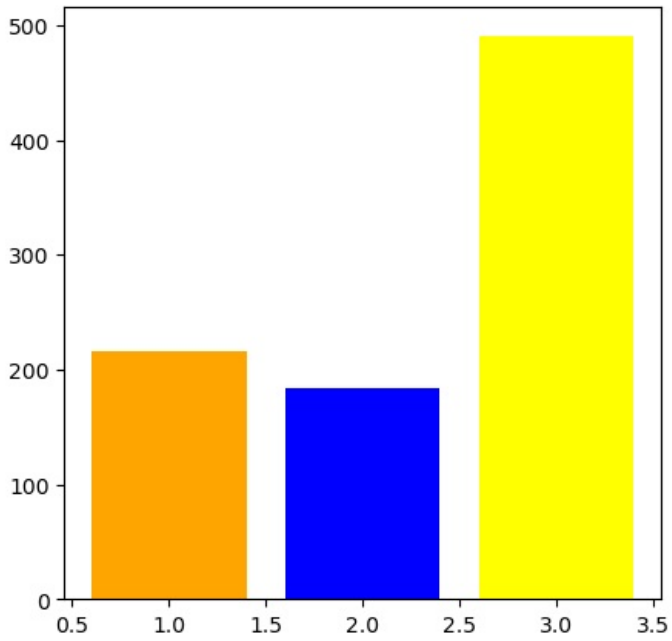
In [11]:
```
plt.figure(figsize=(5,5))
plt.bar(list(titanic_train['Survived'].value_counts().keys()),list(titanic_train['Survived'].value_counts()),co
plt.show()
```



In [12]:
```
titanic_train['Pclass'].value_counts()
```

Out[12]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

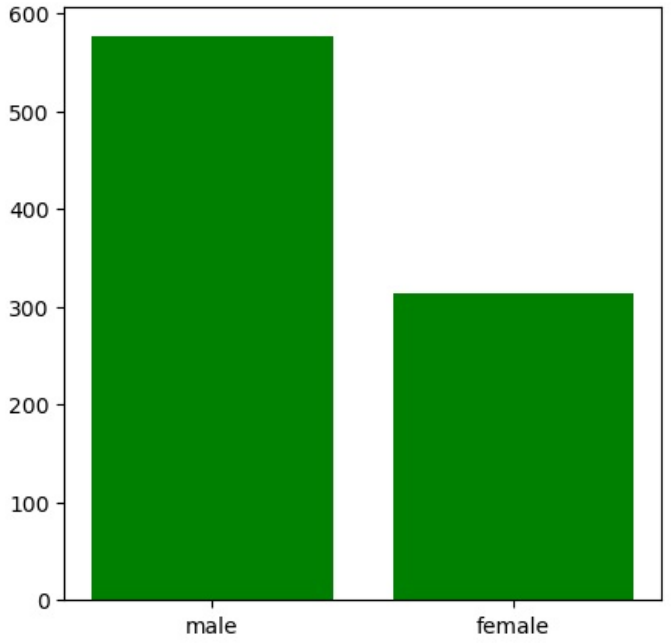
In [13]:
```
plt.figure(figsize=(5,5))
plt.bar(list(titanic_train['Pclass'].value_counts().keys()),list(titanic_train['Pclass'].value_counts()),color=
plt.show()
```



In [14]:
```
titanic_train['Sex'].value_counts()
```

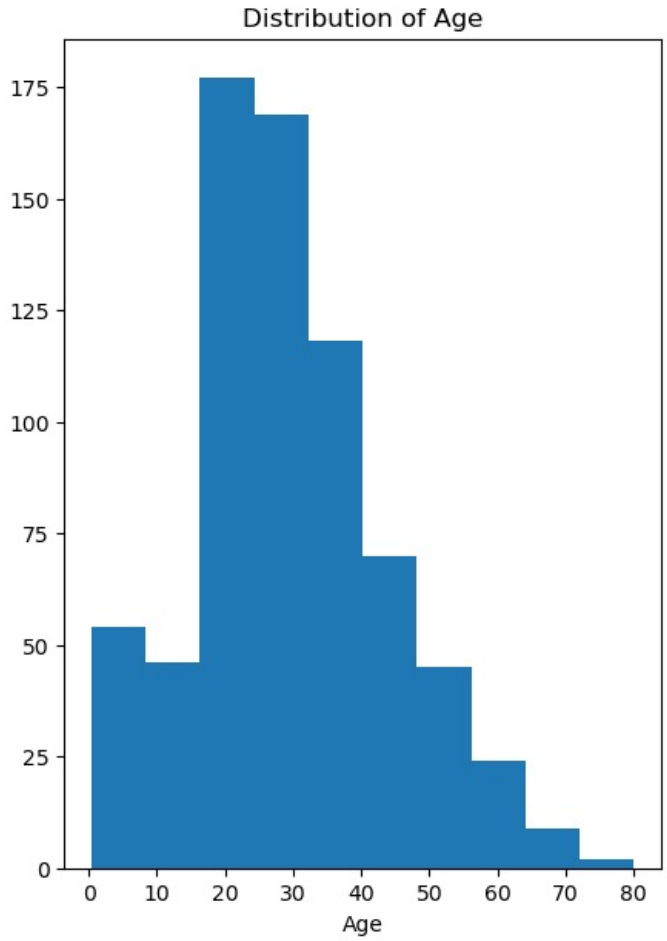
Out[14]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [15]:
```
plt.figure(figsize=(5,5))
plt.bar(list(titanic_train['Sex'].value_counts().keys()),list(titanic_train['Sex'].value_counts()),color="Green
plt.show()
```



In [16]:
```
import matplotlib.pyplot as plt

plt.figure(figsize=(5, 7))
plt.hist(titanic_train['Age'])
plt.title("Distribution of Age")
plt.xlabel("Age")  # Corrected method name
plt.show()
```



In [17]:
```
titanic_train['Survived'].isnull()
```

Out[17]:
```
0      False
1      False
2      False
3      False
4      False
       ...
886    False
887    False
888    False
889    False
890    False
Name: Survived, Length: 891, dtype: bool
```

In [18]:
```
sum(titanic_train['Survived'].isnull())
```

Out[18]:
```
0
```

In [19]:
```
titanic_train['Age'].isnull()
```

Out[19]:
```
0      False
1      False
2      False
3      False
4      False
       ...
886    False
887    False
888     True
889    False
890    False
Name: Age, Length: 891, dtype: bool
```

In [20]:
```
sum(titanic_train['Age'].isnull())
```

Out[20]:
```
177
```

In [21]:
```
titanic_train = titanic_train.dropna()
```

In [22]:
```
sum(titanic_train['Survived'].isnull())
```

Out[22]:
```
0
```

In [23]:
```
sum(titanic_train['Age'].isnull())
```

Out[23]:
```
0
```

In [28]:
```
x_train = titanic_train [['Age']]
y_train = titanic_train [['Survived']]
```

In [29]:
```
from sklearn.tree import DecisionTreeClassifier
```

In [30]:
```
dtc = DecisionTreeClassifier()
```

In [31]:
```
dtc.fit(x_train, y_train)
```

Out[31]:
```
DecisionTreeClassifier()
```

In [32]:
```
sum(titanic_test['Age'].isnull())
```

Out[32]:
```
86
```

In [33]:
```
titanic_test = titanic_test.dropna()
```

In [34]:
```
sum(titanic_test['Age'].isnull())
```

Out[34]:
```
0
```

In [35]:
```
x_test = titanic_test[['Age']]
```

In [36]:
```
y_pred = dtc.predict(x_test)
```

In [37]:
```
y_pred
```

Out[37]:
```
array([1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0,
       1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1,
       1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1],
      dtype=int64)
```

In [ ]: