In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:
```
titanic_data = pd.read_csv(r"C:\Users\SAMEEP GUPTA\OneDrive\Desktop\DAV practice files\Titanic_data.csv"
```

In [6]:
```
titanic_data.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]:
```
column_missing_values = titanic_data.isnull().sum().idxmax()
titanic_cleaned = titanic_data.drop(column_missing_values, axis=1)
```

In [9]:
```
column_missing_values
```

Out[9]:
```
'Cabin'
```

In [10]:
```
titanic_cleaned.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 [13]:
```
pass_30 = titanic_cleaned[titanic_cleaned.Age > 30]
```

In [14]:
```
pass_30.head()
```

Out[14]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C |
| **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 |
| **6** | 7 | 0 | 1 | McCarthy, Mr. Timothy J | male | 54.0 | 0 | 0 | 17463 | 51.8625 | S |
| **11** | 12 | 1 | 1 | Bonnell, Miss. Elizabeth | female | 58.0 | 0 | 0 | 113783 | 26.5500 | S |

In [15]:
```
tot_pass_30 = len(pass_30)
```

In [16]:
```
tot_pass_30
```

Out[16]:
```
305
```

In [22]:
```
tot_2_class = titanic_cleaned[titanic_cleaned['Pclass']=='2']['Fare'].sum()
```

In [23]:
```
tot_2_class
```

Out[23]:
```
0.0
```

In [31]:
```
#number of survivors of each passenger by class
survivor_class = titanic_cleaned.groupby('Pclass')['Survived']
```

In [32]:
```
survivor_class
```

Out[32]:
```
<pandas.core.groupby.generic.SeriesGroupBy object at 0x0000014A8818D350>
```

In [33]:
```
survivor_class.sum()
```

Out[33]:
```
Pclass
1    136
2     87
3    119
Name: Survived, dtype: int64
```

In [40]:
```
age_gen_stats=titanic_cleaned.groupby('Age')['Sex']
```

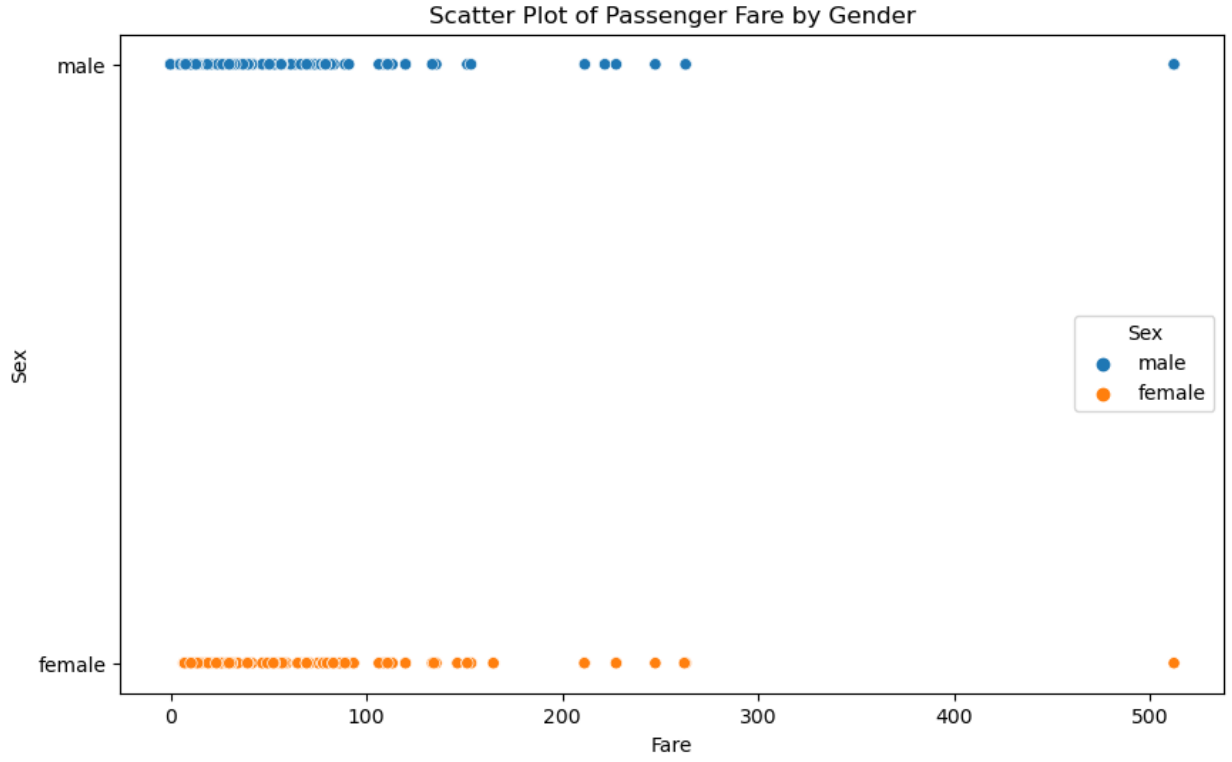
In [41]:
```
age_gen_stats.describe()
```

Out[41]:

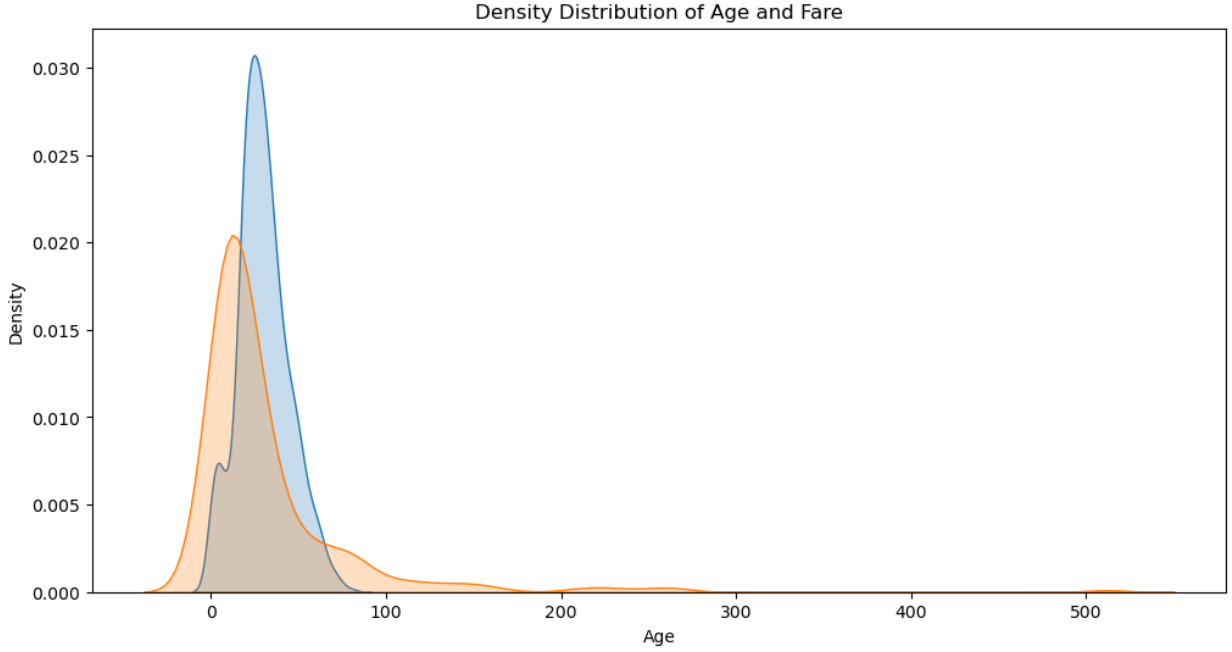
| | count | unique | top | freq |
|---|---|---|---|---|
| **Age** | | | | |
| **0.42** | 1 | 1 | male | 1 |
| **0.67** | 1 | 1 | male | 1 |
| **0.75** | 2 | 1 | female | 2 |
| **0.83** | 2 | 1 | male | 2 |
| **0.92** | 1 | 1 | male | 1 |
| **...** | ... | ... | ... | ... |
| **70.00** | 2 | 1 | male | 2 |
| **70.50** | 1 | 1 | male | 1 |
| **71.00** | 2 | 1 | male | 2 |
| **74.00** | 1 | 1 | male | 1 |
| **80.00** | 1 | 1 | male | 1 |

88 rows × 4 columns

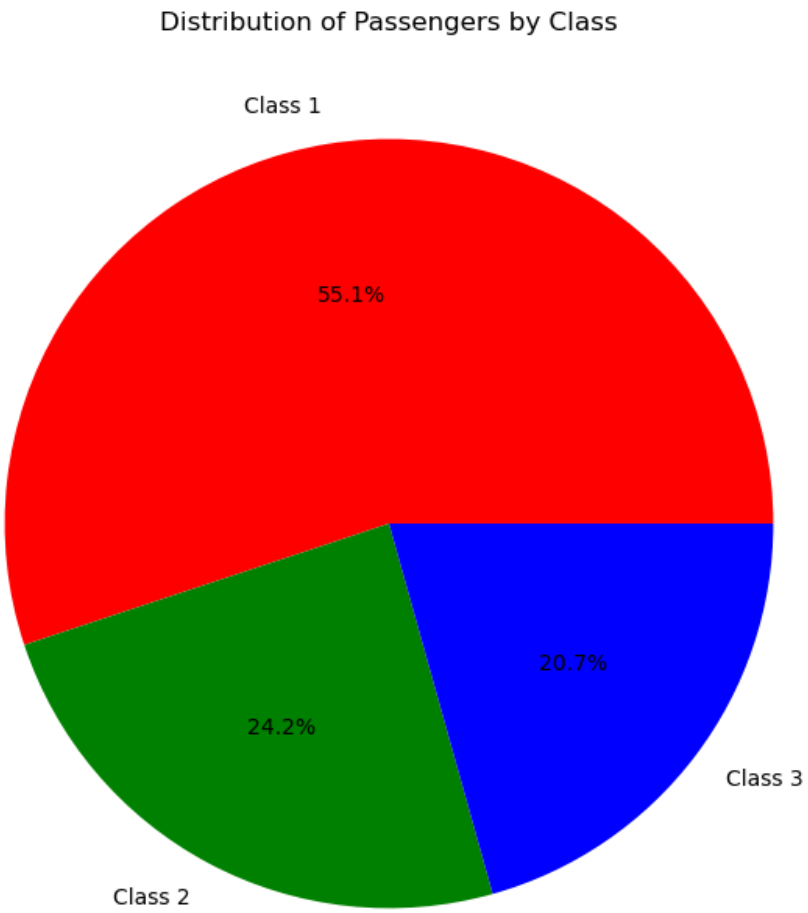
In [43]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Fare', y='Sex', data=titanic_cleaned, hue='Sex')
plt.title('Scatter Plot of Passenger Fare by Gender')
plt.show()
```




In [44]:
```
plt.figure(figsize=(12, 6))
sns.kdeplot(data=titanic_cleaned, x='Age', label='Age', fill=True)
sns.kdeplot(data=titanic_cleaned, x='Fare', label='Fare', fill=True)
plt.title('Density Distribution of Age and Fare')
plt.show()
```



In [45]:
```
class_counts = titanic_cleaned['Pclass'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(class_counts, labels=['Class 1', 'Class 2', 'Class 3'], autopct='%1.1f%%', colors=['red', 'green
plt.title('Distribution of Passengers by Class')
plt.show()
```




In [47]:
```
survival_percentage_by_class = titanic_cleaned.groupby('Pclass')['Survived'].mean() * 100
```

In [49]:
```
survival_percentage_by_class # Yes, class played a great role in survival of the passengers
```

Out[49]:
```
Pclass
1    62.962963
2    47.282609
3    24.236253
Name: Survived, dtype: float64
```