

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

❖ **Aim:** To perform Exploratory Data Analysis (EDA) on the Titanic dataset using Seaborn, with the goal of uncovering patterns, relationships, and insights related to passenger survival.

❖ **Software Used:** Python 3.12, Jupyter Notebook

❖ **Learning Objective:**

1. Understand the structure and contents of the Titanic dataset.
2. Perform univariate and bivariate analysis to explore the distribution of features.
3. Use Seaborn for visualizing relationships between features and survival rates.
4. Gain insights into factors influencing survival chances on the Titanic.

❖ **Learning Outcomes:**

After performing the experiment students will be able to-

Through the EDA of the Titanic dataset using Seaborn, students are able to uncover several key insights:

1. The survival rate varied significantly across different passenger classes, with higher classes generally having better survival rates.
2. Gender had a substantial impact on survival chances, with females having a higher survival rate compared to males.
3. The age and fare distributions provided additional context on survival, showing that both very young and older passengers, as well as those who paid higher fares, had different survival rates.

These insights can help in understanding the factors influencing survival on the Titanic and can serve as a basis for more advanced predictive modeling.

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

❖ Theory:

Exploratory Data Analysis (EDA) involves summarizing and visualizing datasets to understand their underlying patterns and relationships. The Titanic dataset provides information about passengers, including demographics, ticket details, and survival status.

Key techniques in EDA include:

- **Univariate Analysis:** Analyzing individual variables to understand their distribution and characteristics.
- **Bivariate Analysis:** Examining the relationship between two variables, particularly to understand how features affect survival rates.
- **Multivariate Analysis:** Investigating interactions between multiple variables to uncover more complex patterns.

❖ Code and Output:

Import Libraries and Load Dataset

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Load the Titanic dataset

```
df = pd.read_csv("D:\\DSV_Datadataset\\titanic\\train.csv")
```

##b. Data Overview

Display the first few rows of the dataset

```
print(df.head())
```

Summary of the dataset

```
print(df.info())
```

Statistical summary of numerical features

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

```
print(df.describe())
```

Output:

```
PassengerId Survived Pclass \
```

```
0      1      0      3
1      2      1      1
2      3      1      3
3      4      1      1
4      5      0      3
```

```
Name      Sex  Age SibSp \
```

```
0      Braund, Mr. Owen Harris  male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2      Heikkinen, Miss. Laina  female  26.0      0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4      Allen, Mr. William Henry  male  35.0      0
```

```
Parch      Ticket  Fare Cabin Embarked
0      0      A/5 21171  7.2500  NaN      S
1      0      PC 17599 71.2833  C85      C
2      0  STON/O2. 3101282  7.9250  NaN      S
3      0      113803 53.1000  C123      S
4      0      373450  8.0500  NaN      S
```

```
<class 'pandas.core.frame.DataFrame'>
```

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

RangeIndex: 891 entries, 0 to 890

Data columns (total 12 columns):

Column Non-Null Count Dtype

--- -----

0 PassengerId 891 non-null int64

1 Survived 891 non-null int64

2 Pclass 891 non-null int64

3 Name 891 non-null object

4 Sex 891 non-null object

5 Age 714 non-null float64

6 SibSp 891 non-null int64

7 Parch 891 non-null int64

8 Ticket 891 non-null object

9 Fare 891 non-null float64

10 Cabin 204 non-null object

11 Embarked 889 non-null object

dtypes: float64(2), int64(5), object(5)

memory usage: 83.7+ KB

None

	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

```
75%    668.500000    1.000000    3.000000    38.000000    1.000000
max     891.000000    1.000000    3.000000    80.000000    8.000000
```

```
      Parch      Fare
count  891.000000  891.000000
mean    0.381594   32.204208
std     0.806057   49.693429
min     0.000000    0.000000
25%     0.000000    7.910400
50%     0.000000   14.454200
75%     0.000000   31.000000
max      6.000000  512.329200
```

##Univariate Analysis

Distribution of Survival

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

plt.title('Survival Distribution')

plt.xlabel('Survived')

plt.ylabel('Count')

plt.show()
```

Distribution of Passenger Classes

```
sns.countplot(x='Pclass', data=df)

plt.title('Passenger Class Distribution')

plt.xlabel('Pclass')

plt.ylabel('Count')

plt.show()
```

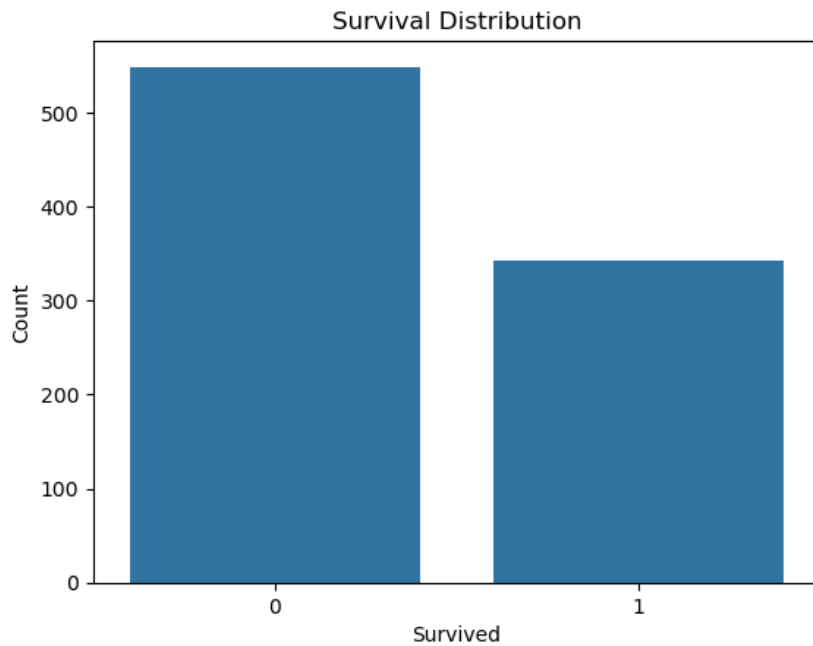
AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

Age Distribution

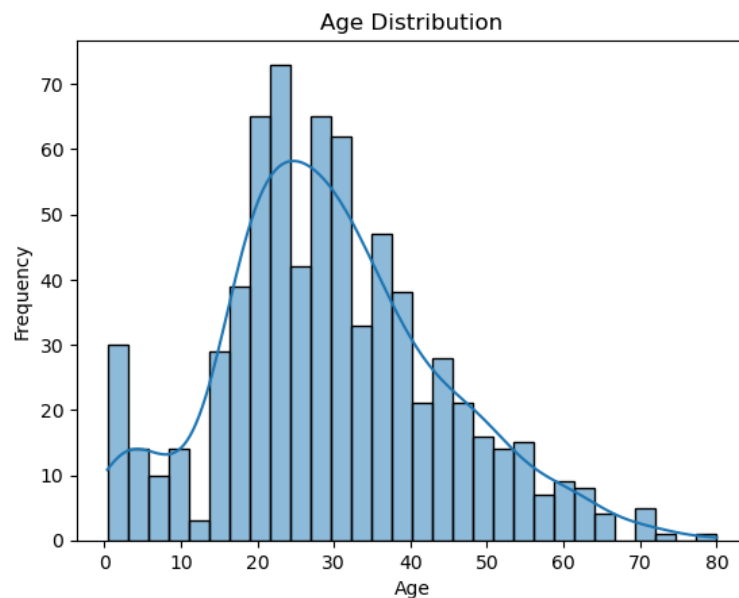
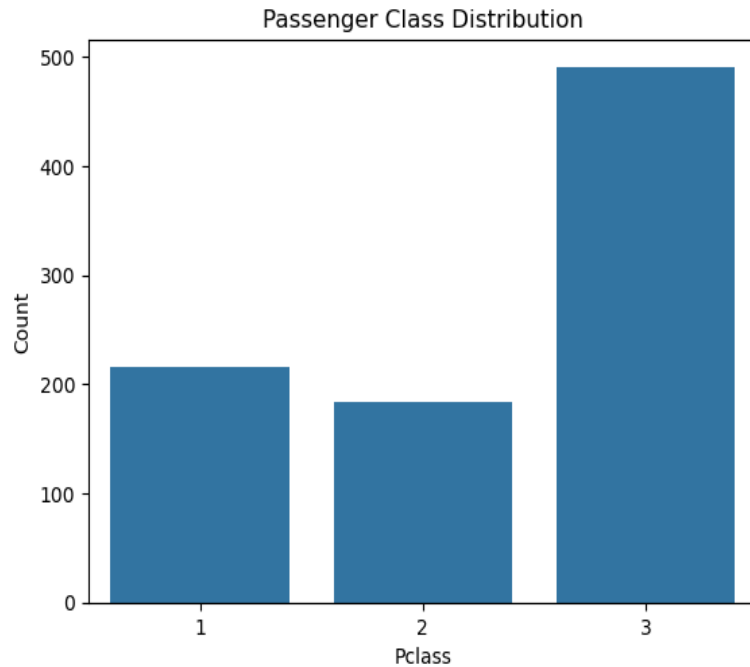
```
sns.histplot(df['Age'].dropna(), bins=30, kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

Output:



AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6



Bivariate Analysis

Survival Rate by Gender

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

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

```
plt.title('Survival Rate by Gender')
```

```
plt.xlabel('Sex')
```

```
plt.ylabel('Survival Rate')
```

```
plt.show()
```

Survival Rate by Passenger Class

```
sns.barplot(x='Pclass', y='Survived', data=df)
```

```
plt.title('Survival Rate by Passenger Class')
```

```
plt.xlabel('Pclass')
```

```
plt.ylabel('Survival Rate')
```

```
plt.show()
```

Survival Rate by Embarked

```
sns.barplot(x='Embarked', y='Survived', data=df)
```

```
plt.title('Survival Rate by Embarked')
```

```
plt.xlabel('Embarked')
```

```
plt.ylabel('Survival Rate')
```

```
plt.show()
```

Fare vs. Survival

```
sns.boxplot(x='Survived', y='Fare', data=df)
```

```
plt.title('Fare vs. Survival')
```

```
plt.xlabel('Survived')
```

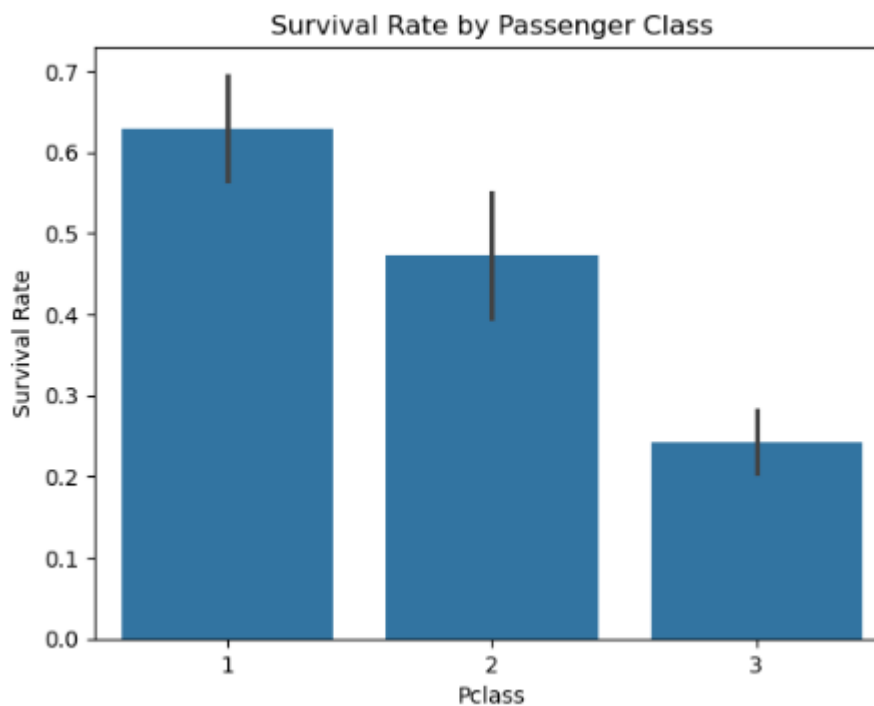
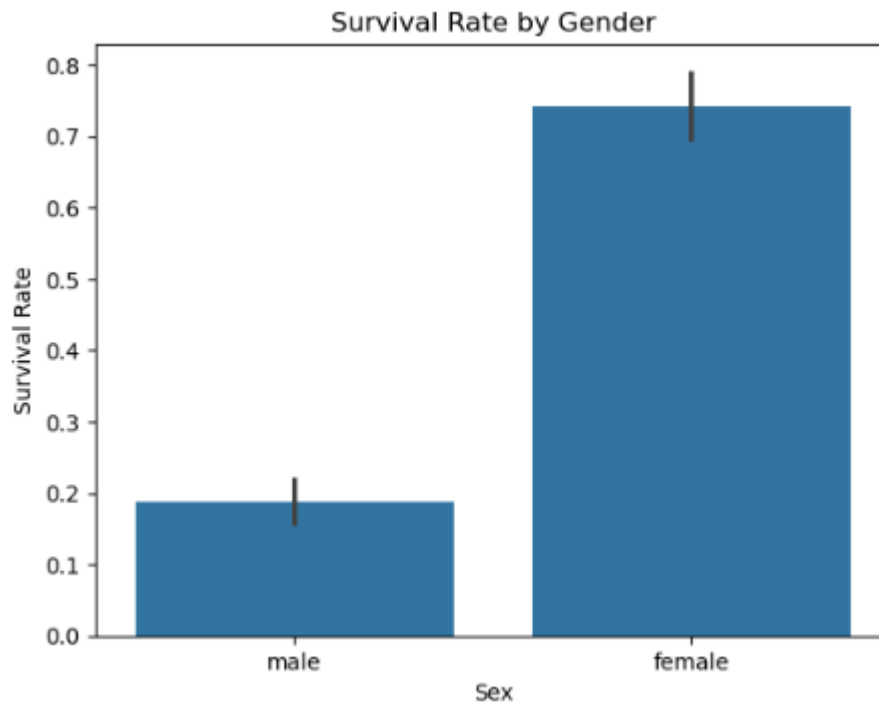
```
plt.ylabel('Fare')
```

```
plt.show()
```

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

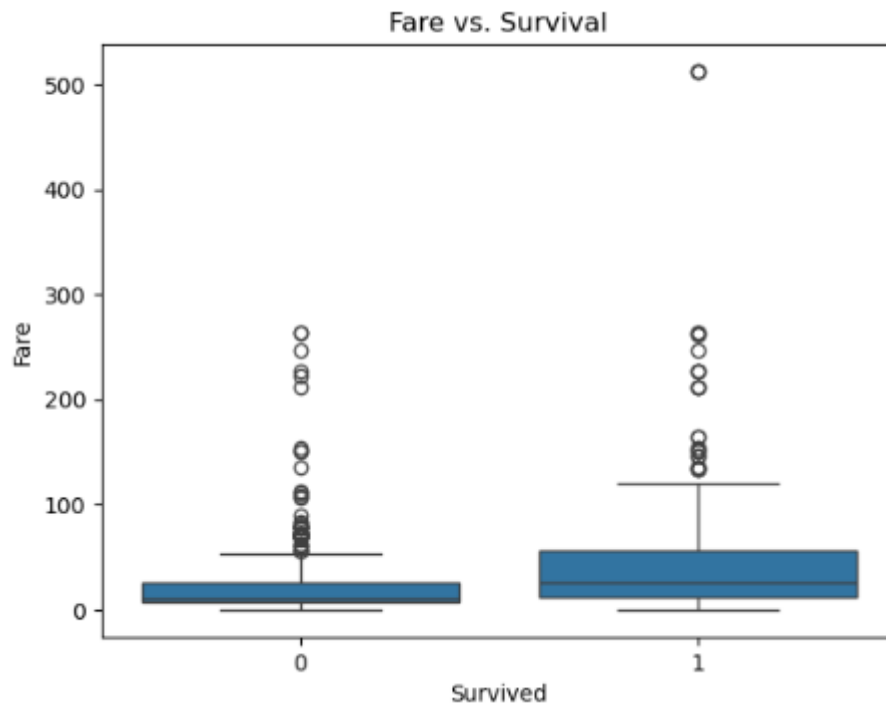
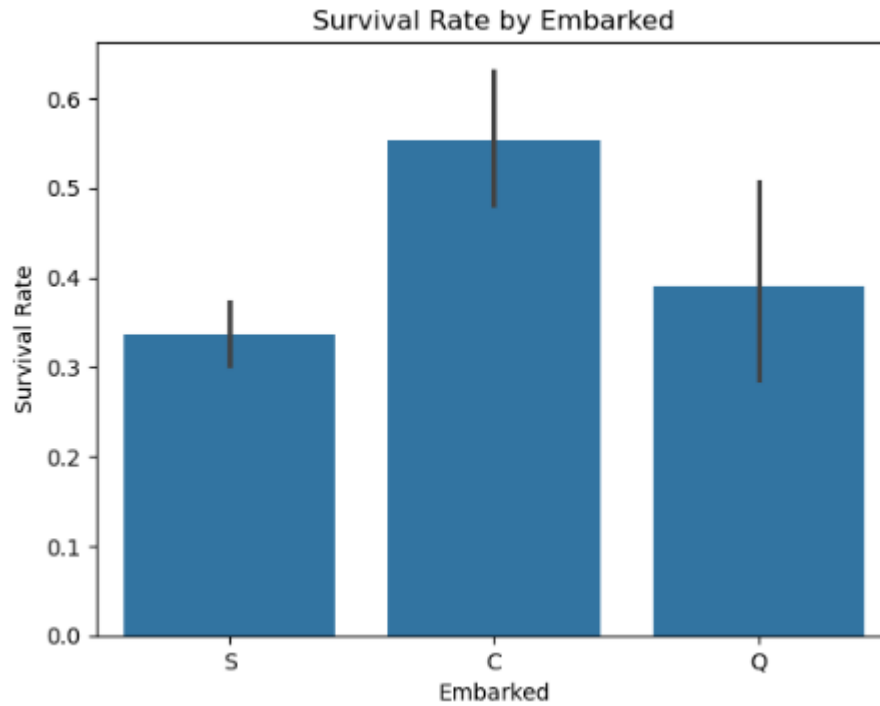
BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

Output:



AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6



AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

#Multivariate Analysis

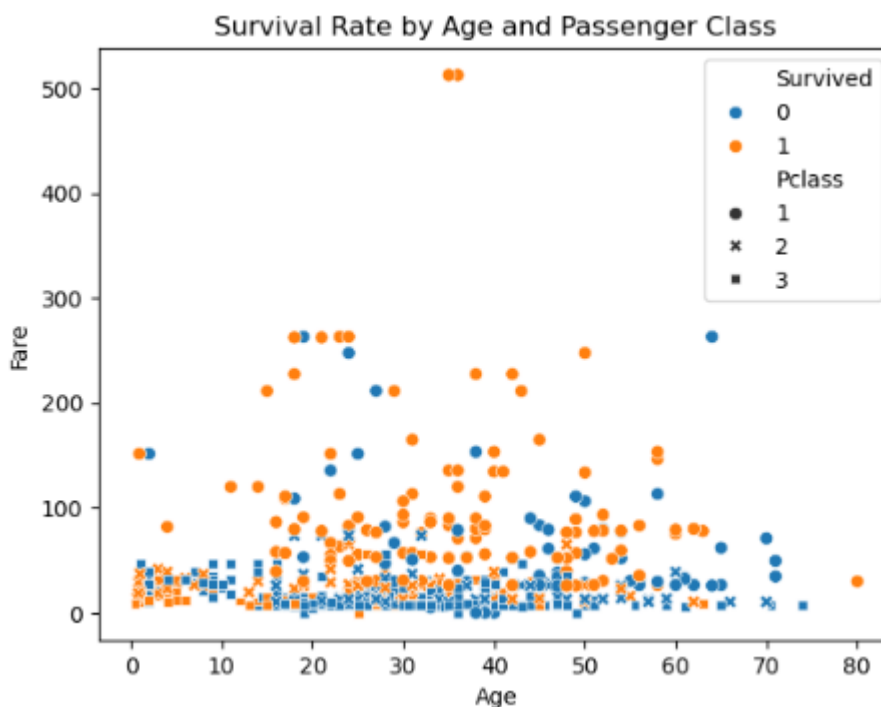
Survival Rate by Age and Passenger Class

```
sns.scatterplot(x='Age', y='Fare', hue='Survived', style='Pclass', data=df)
plt.title('Survival Rate by Age and Passenger Class')
plt.xlabel('Age')
plt.ylabel('Fare')
plt.show()
```

Pairplot for Selected Features

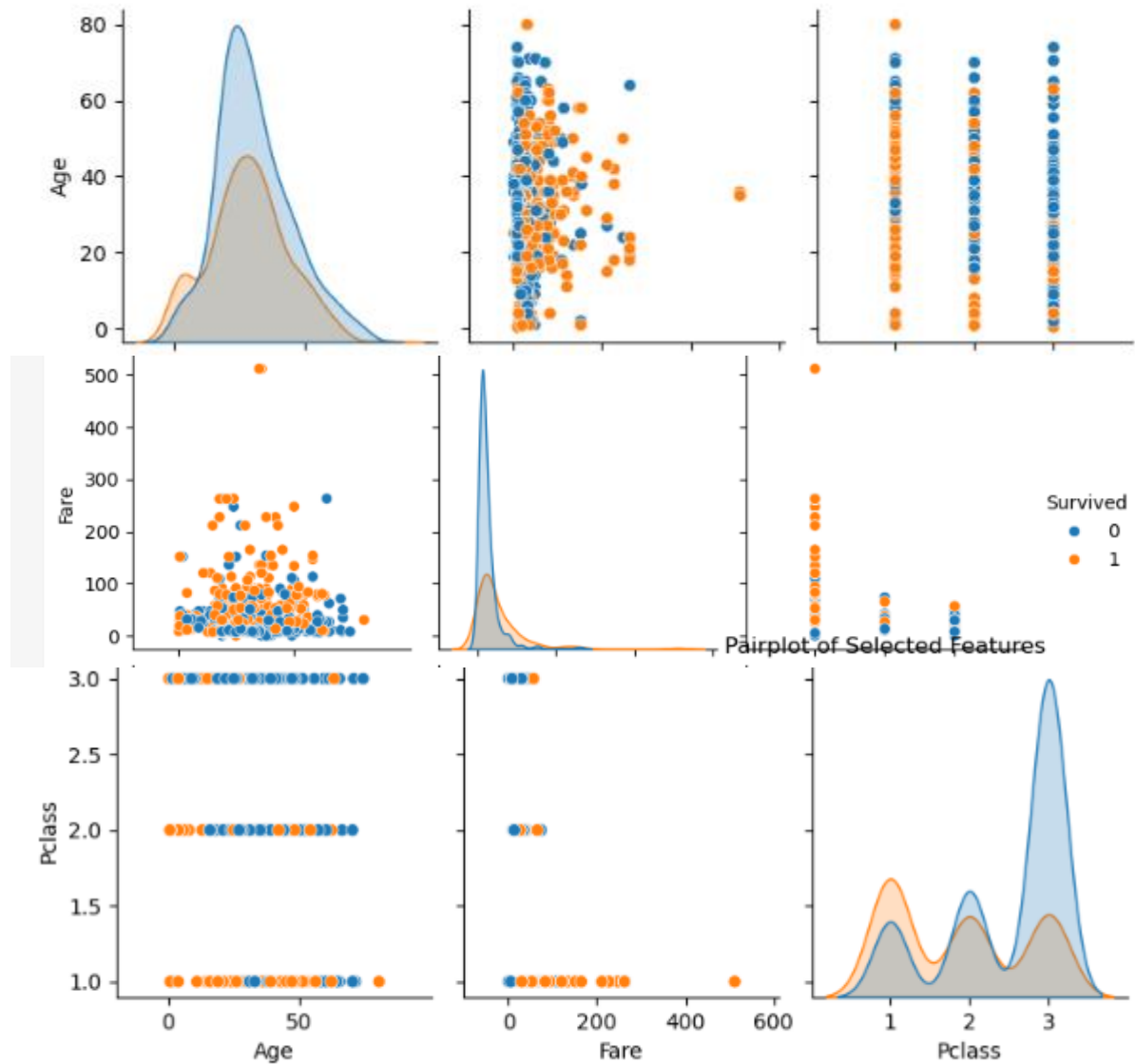
```
sns.pairplot(df[['Age', 'Fare', 'Pclass', 'Survived']].dropna(), hue='Survived')
plt.title('Pairplot of Selected Features')
plt.show()
```

Output:



AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6



AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.:05	Titanic Dataset Exploratory Data Analysis (EDA) using Seaborn.	Page: /6

❖ Conclusion:

❖ Questions:

1. What were the key demographic factors influencing the survival rate of passengers on the Titanic?
2. How did survival rates differ between passengers from different embarkation ports?
3. What patterns can be observed from the age distribution of passengers?
4. How did the fare paid by passengers relate to their survival chances?
5. What are the limitations of this EDA, and what additional analyses could be performed?