practical-9-piyusha-supe

April 12, 2025

1 Piyusha Supe 23CO315

Data Visualization II 1. Use the inbuilt dataset 'titanic' as used in the above problem. Plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not. (Column names : 'sex' and 'age') 2. Write observations on the inference from the above statistics.

[1]:

```
#import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:

```
titanic = sns.load_dataset("titanic")
```

[3]:

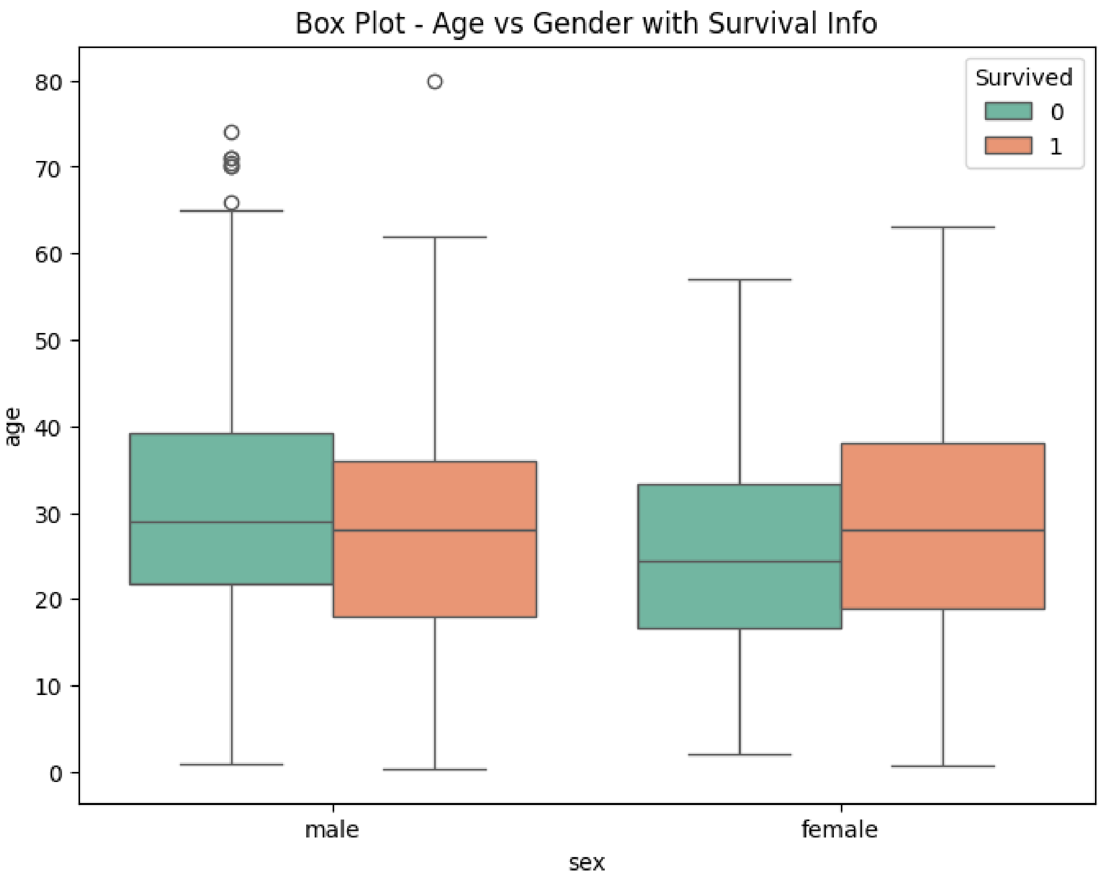
```
print(titanic.head())
print(titanic.tail())
print(titanic.info())
print(titanic.describe(include="all"))
print(titanic.shape)
print(titanic.size)
print(titanic.ndim)
print(titanic.columns)
```

```
   survived  pclass     sex   age  sibsp  parch     fare embarked  class  \
0         0       3    male  22.0      1      0   7.2500        S  Third
1         1       1  female  38.0      1      0  71.2833        C  First
2         1       3  female  26.0      0      0   7.9250        S  Third
3         1       1  female  35.0      1      0  53.1000        S  First
4         0       3    male  35.0      0      0   8.0500        S  Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True
3  woman       False    C  Southampton   yes  False
4    man        True  NaN  Southampton    no   True
     survived  pclass     sex   age  sibsp  parch   fare embarked   class  \
886         0       2    male  27.0      0      0  13.00        S  Second
887         1       1  female  19.0      0      0  30.00        S   First
888         0       3  female   NaN      1      2  23.45        S   Third
889         1       1    male  26.0      0      0  30.00        C   First
890         0       3    male  32.0      0      0   7.75        Q   Third

       who  adult_male deck  embark_town alive  alone
886    man        True  NaN  Southampton    no   True
887  woman       False    B  Southampton   yes   True
888  woman       False  NaN  Southampton    no  False
889    man        True    C    Cherbourg   yes   True
890    man        True  NaN   Queenstown    no   True
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
None
          survived      pclass  sex         age       sibsp       parch  \
count   891.000000  891.000000  891  714.000000  891.000000  891.000000
unique         NaN         NaN    2         NaN         NaN         NaN
top            NaN         NaN  male         NaN         NaN         NaN
freq           NaN         NaN   577         NaN         NaN         NaN
mean      0.383838    2.308642   NaN   29.699118    0.523008    0.381594
std       0.486592    0.836071   NaN   14.526497    1.102743    0.806057
min       0.000000    1.000000   NaN    0.420000    0.000000    0.000000
25%       0.000000    2.000000   NaN   20.125000    0.000000    0.000000
50%       0.000000    3.000000   NaN   28.000000    0.000000    0.000000
75%       1.000000    3.000000   NaN   38.000000    1.000000    0.000000
max       1.000000    3.000000   NaN   80.000000    8.000000    6.000000
```

```
              fare embarked  class  who adult_male deck  embark_town alive  \
count   891.000000      889    891  891        891  203          889   891   
unique         NaN        3      3    3          2    7            3     2   
top            NaN        S  Third  man       True    C  Southampton    no   
freq           NaN      644    491  537        537   59          644   549   
mean     32.204208      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
std      49.693429      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
min       0.000000      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
25%       7.910400      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
50%      14.454200      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
75%      31.000000      NaN    NaN  NaN        NaN  NaN          NaN   NaN   
max     512.329200      NaN    NaN  NaN        NaN  NaN          NaN   NaN   

       alone  
count    891  
unique     2  
top     True  
freq     537  
mean     NaN  
std      NaN  
min      NaN  
25%      NaN  
50%      NaN  
75%      NaN  
max      NaN  
(891, 15)
13365
2
Index(['survived', 'pclass', 'sex', 'age', 'sibsp', 'parch', 'fare',
       'embarked', 'class', 'who', 'adult_male', 'deck', 'embark_town',
       'alive', 'alone'],
      dtype='object')
```

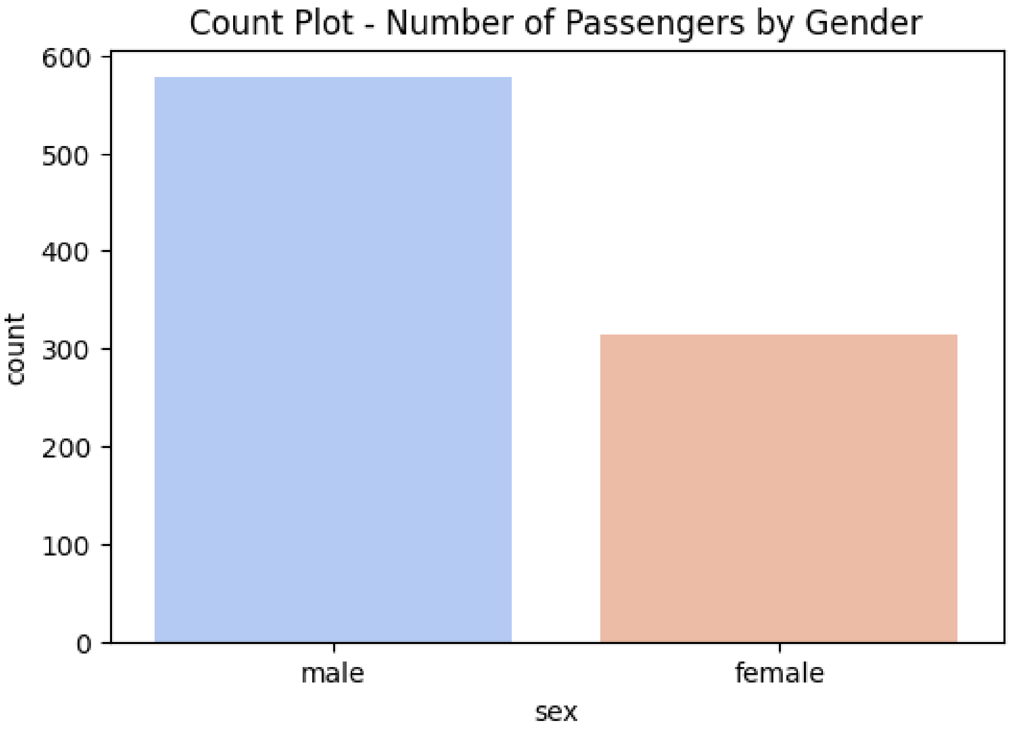
[4]:

```python
# 1. Box Plot: Age Distribution by Sex and Survival
plt.figure(figsize=(8, 6))
sns.boxplot(x='sex', y='age', hue='survived', data=titanic, palette='Set2')
plt.title("Box Plot - Age vs Gender with Survival Info")
plt.legend(title="Survived", loc="upper right")
plt.show()
```



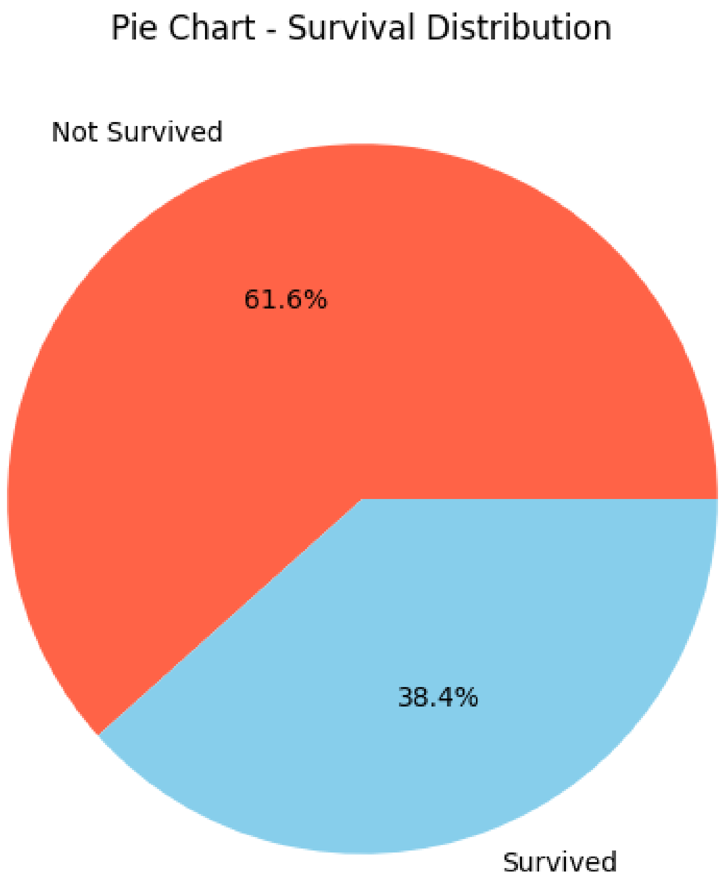
```
[5]: # 2. Count Plot: Number of Passengers by Gender
     plt.figure(figsize=(6, 4))
     sns.countplot(x='sex', data=titanic, palette='coolwarm')
     plt.title("Count Plot - Number of Passengers by Gender")
     plt.show()
```

```
<ipython-input-5-ba1750916024>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(x='sex', data=titanic, palette='coolwarm')
```

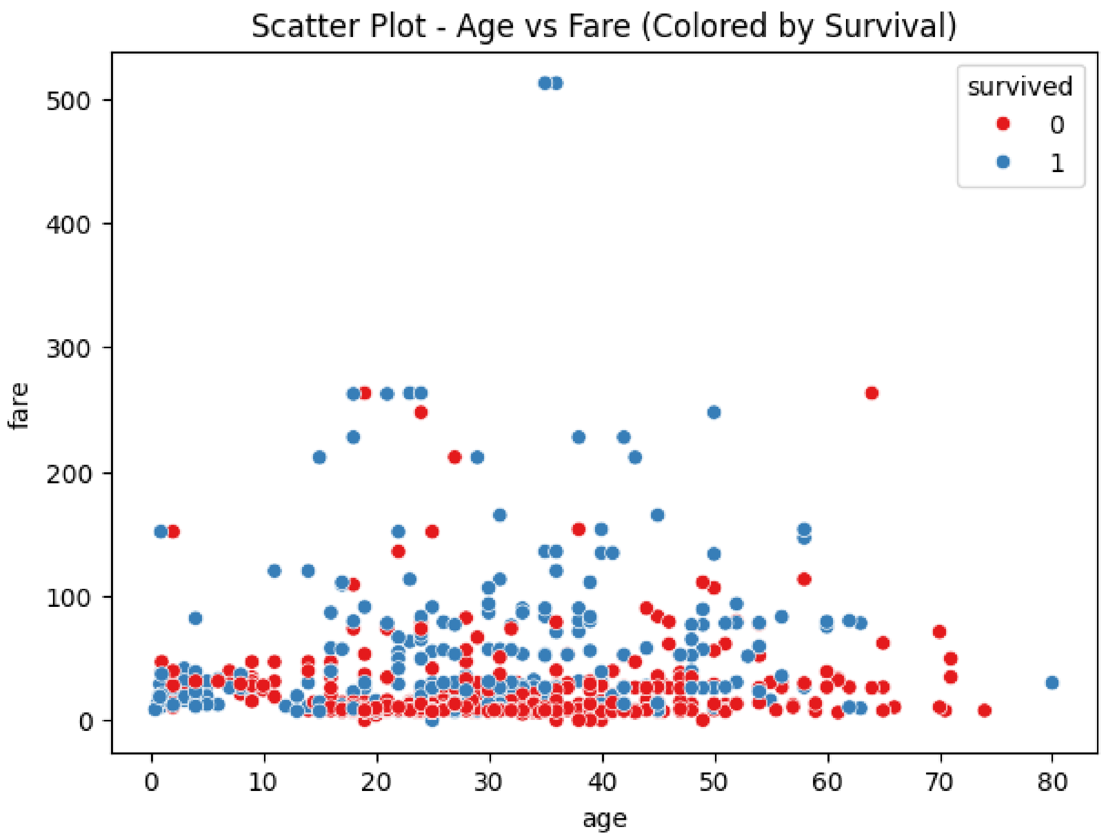


[6]:

```
# 3. Pie Chart: Survival Ratio
survived_counts = titanic['survived'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(survived_counts, labels=['Not Survived', 'Survived'], autopct='%1.
 ↪1f%%', colors=['tomato', 'skyblue'])
plt.title("Pie Chart - Survival Distribution")
plt.show()
```

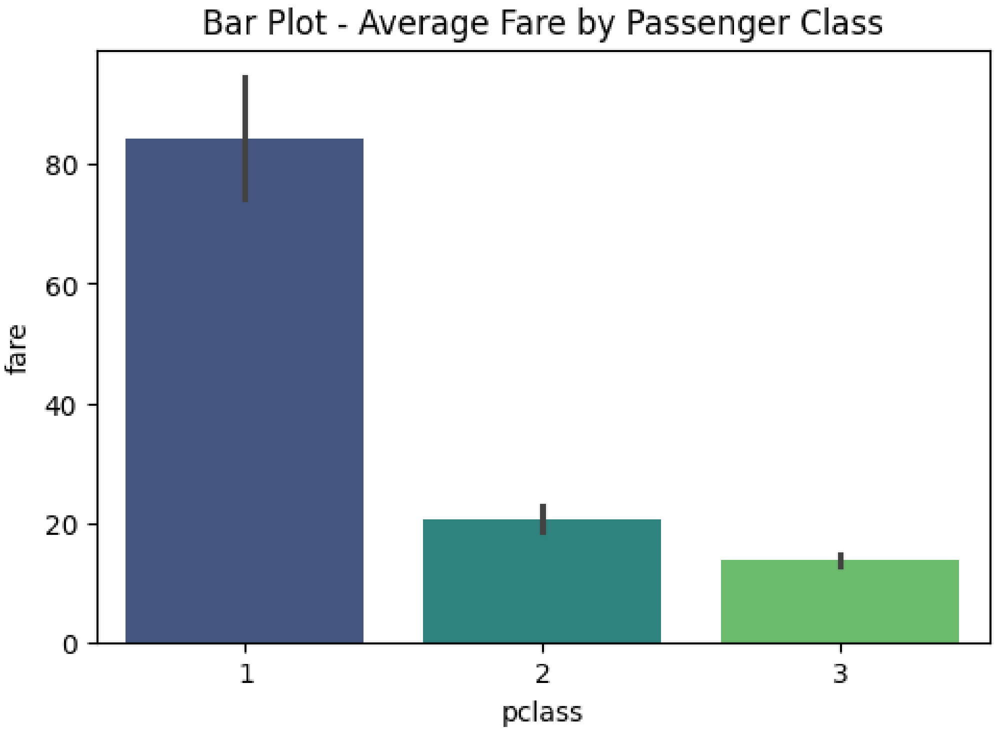


[7]:
```
# 4. Scatter Plot: Age vs Fare by Survival
plt.figure(figsize=(7, 5))
sns.scatterplot(x='age', y='fare', hue='survived', data=titanic, palette='Set1')
plt.title("Scatter Plot - Age vs Fare (Colored by Survival)")
plt.show()
```

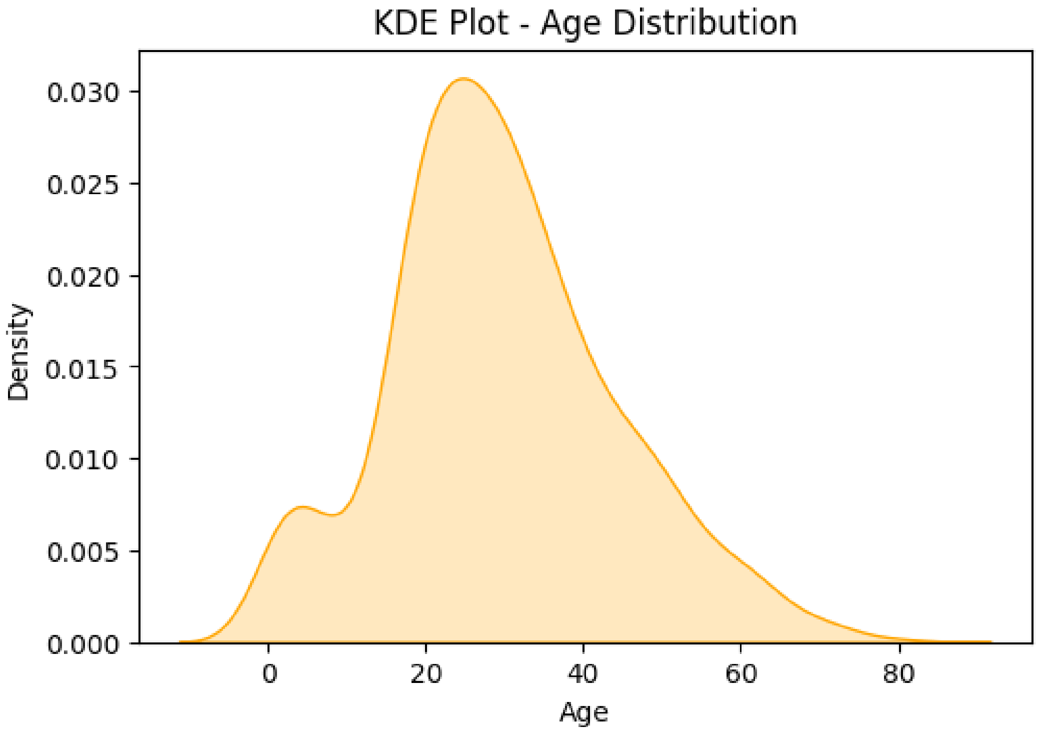


[8]:
```python
# 5. Bar Plot: Average Fare by Passenger Class
plt.figure(figsize=(6, 4))
sns.barplot(x='pclass', y='fare', data=titanic, palette='viridis')
plt.title("Bar Plot - Average Fare by Passenger Class")
plt.show()
```

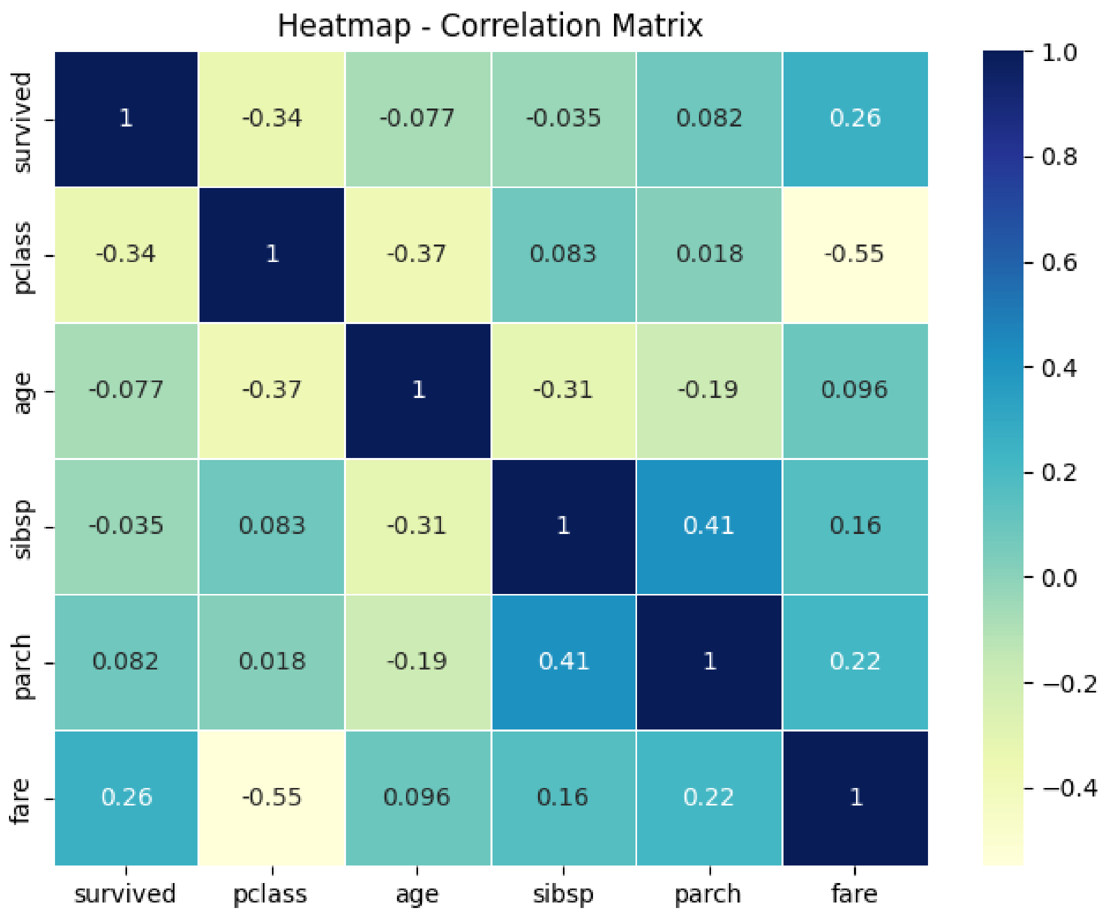
```
<ipython-input-8-23b140553b58>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(x='pclass', y='fare', data=titanic, palette='viridis')
```



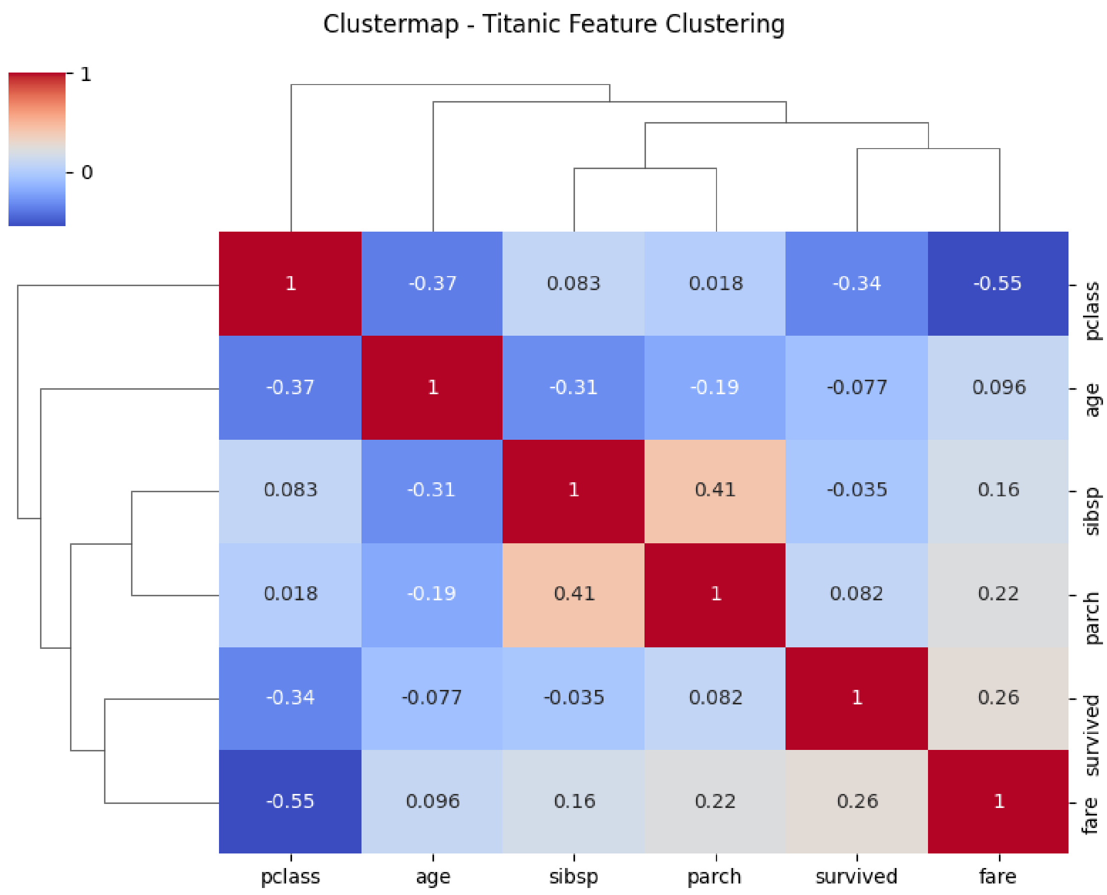
[9]:
```
# 6. Distplot (KDE only): Age Distribution
plt.figure(figsize=(6, 4))
sns.kdeplot(titanic['age'].dropna(), fill=True, color='orange')
plt.title("KDE Plot - Age Distribution")
plt.xlabel("Age")
plt.show()
```



[10]:
```
# 7. Heatmap: Correlation Matrix
plt.figure(figsize=(8, 6))
corr = titanic[['survived', 'pclass', 'age', 'sibsp', 'parch', 'fare']].corr()
sns.heatmap(corr, annot=True, cmap='YlGnBu', linewidths=0.5)
plt.title("Heatmap - Correlation Matrix")
plt.show()
```



[11]:
```
# 8. Clustermap: Clustering of Numerical Features
sns.clustermap(corr, annot=True, cmap='coolwarm', figsize=(8, 6))
plt.suptitle("Clustermap - Titanic Feature Clustering", y=1.05)
plt.show()
```



Clustermap - Titanic Feature Clustering

2 PIYUSHA SUPE 23CO315