Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 49 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

Open in... Python 3 (ipykernel)

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

[3]:
```
df = pd.read_csv("train.csv")
```

[4]:
```
df.head()
```

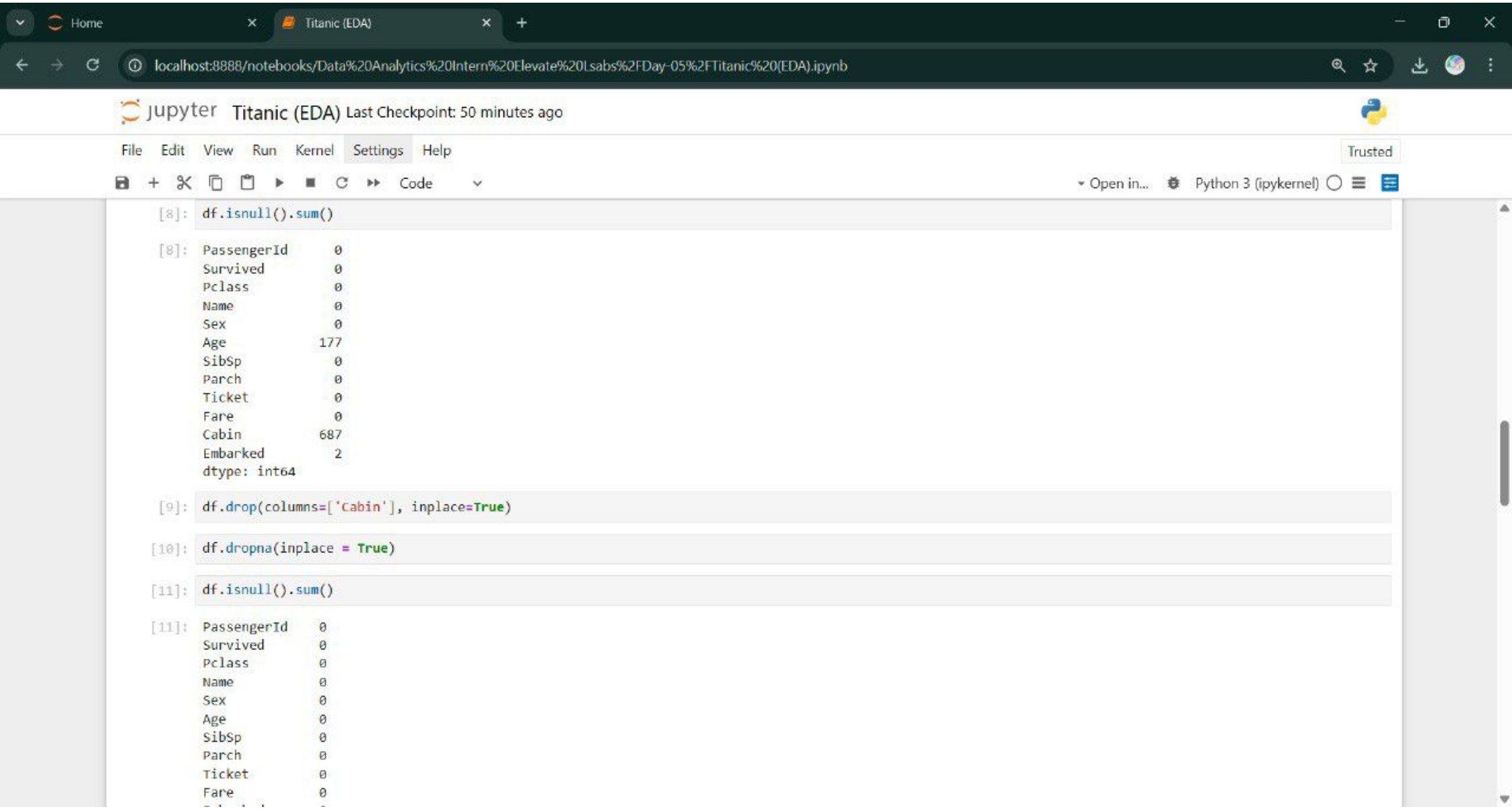
[4]:

	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

[5]:
```
df.tail()
```

[5]:

	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


Home
Titanic (EDA)
localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb
jupyter Titanic (EDA) Last Checkpoint: 50 minutes ago
File Edit View Run Kernel Settings Help
Trusted
Code
Open in...
Python 3 (ipykernel)
[8]: df.isnull().sum()
[8]: PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 177
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 687
Embarked 2
dtype: int64
[9]: df.drop(columns=['Cabin'], inplace=True)
[10]: df.dropna(inplace = True)
[11]: df.isnull().sum()
[11]: PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 0
SibSp 0
Parch 0
Ticket 0
Fare 0

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 49 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

Open in... Python 3 (ipykernel)

[6]:
```
df.describe()
```

[6]:

	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

[7]:
```
df.info()
```

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

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 50 minutes ago

File Edit View Run Kernel Settings Help

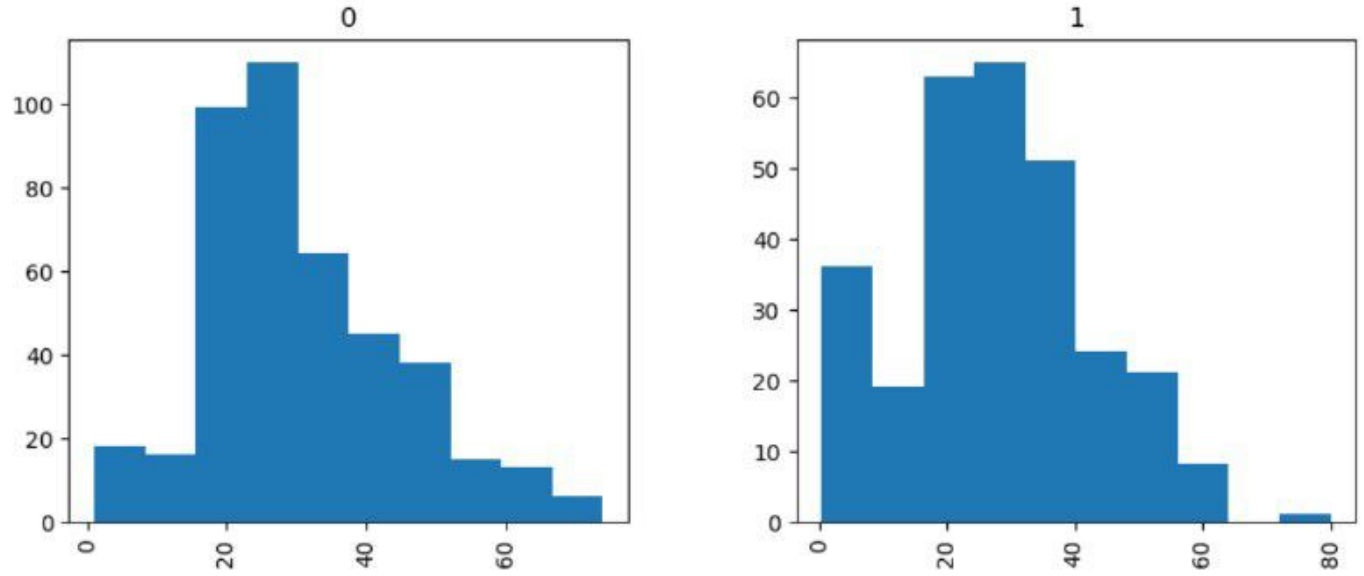
Trusted

Code

Open in... Python 3 (ipykernel)

[14]:
```
df['Age'].hist(by=df['Survived'], figsize=(10,4))
```

[14]:
```
array([<Axes: title={'center': '0'}>, <Axes: title={'center': '1'}>],
      dtype=object)
```



Summary: Children had higher survival rates than middle-aged adults.

[16]:
```
sns.boxplot(x='Survived', y='Fare', data=df)
```

[16]:
```
<Axes: xlabel='Survived', ylabel='Fare'>
```

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 50 minutes ago

File Edit View Run Kernel Settings Help

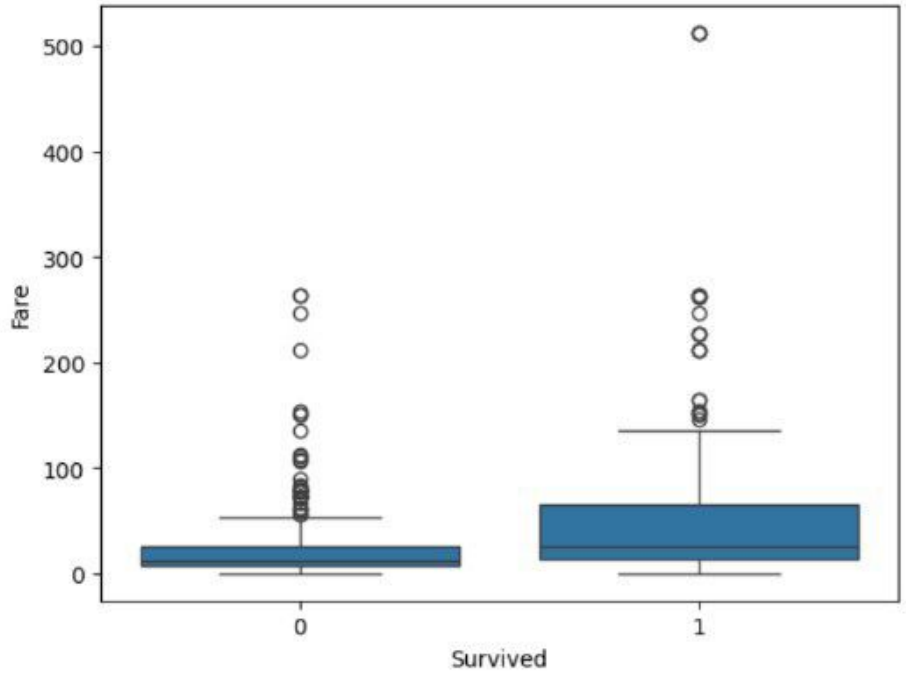
Trusted

Code

Open in... Python 3 (ipykernel)

[16]:
```
sns.boxplot(x='Survived', y='Fare', data=df)
```

[16]:
```
<Axes: xlabel='Survived', ylabel='Fare'>
```



Summary: Survivors paid higher fares on average.

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 50 minutes ago

File Edit View Run Kernel Settings Help

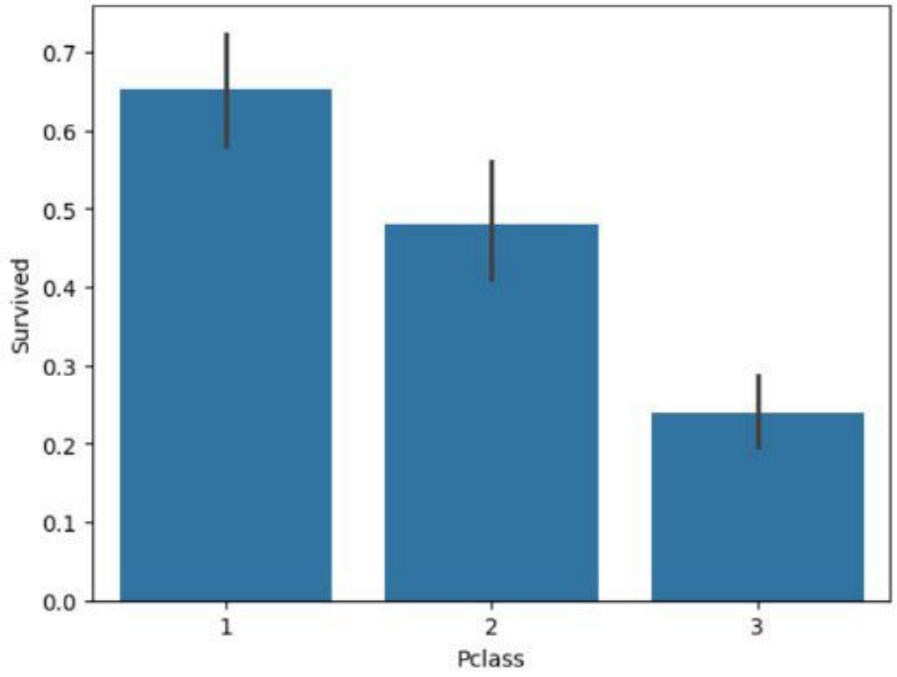
Trusted

Code

Open in... Python 3 (ipykernel)

[18]:
```
sns.barplot(x='Pclass', y='Survived', data=df)
```

[18]:
```
<Axes: xlabel='Pclass', ylabel='Survived'>
```



Summary: Females had a significantly higher survival rate than males.

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 51 minutes ago

File Edit View Run Kernel Settings Help

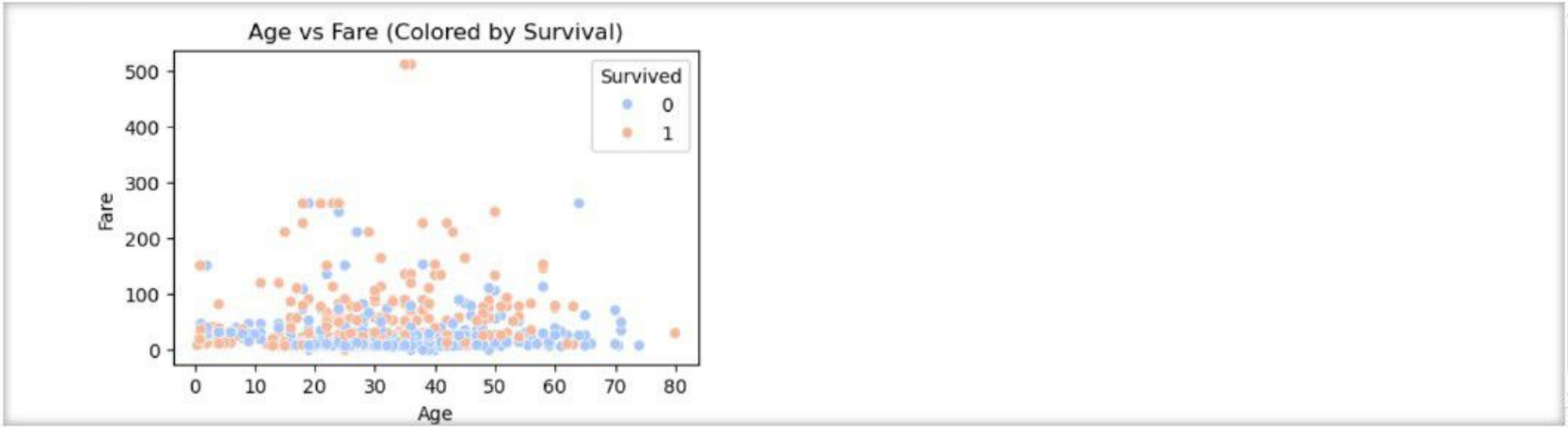
Trusted

Code

Open in... Python 3 (ipykernel)

dtype: int64

[21]:
```
plt.figure(figsize=(5,3))
sns.scatterplot(data=df, x='Age', y='Fare', hue='Survived', palette='coolwarm')
plt.title('Age vs Fare (Colored by Survival)')
plt.show()
```



Summary: Many survivors paid higher fares; younger survivors also noticeable.

[23]:
```
sns.countplot(x='Pclass', hue='Sex', data=df[df['Survived'] == 1])
plt.title("Survived Passengers by Class and Gender")
plt.show()
```

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 51 minutes ago

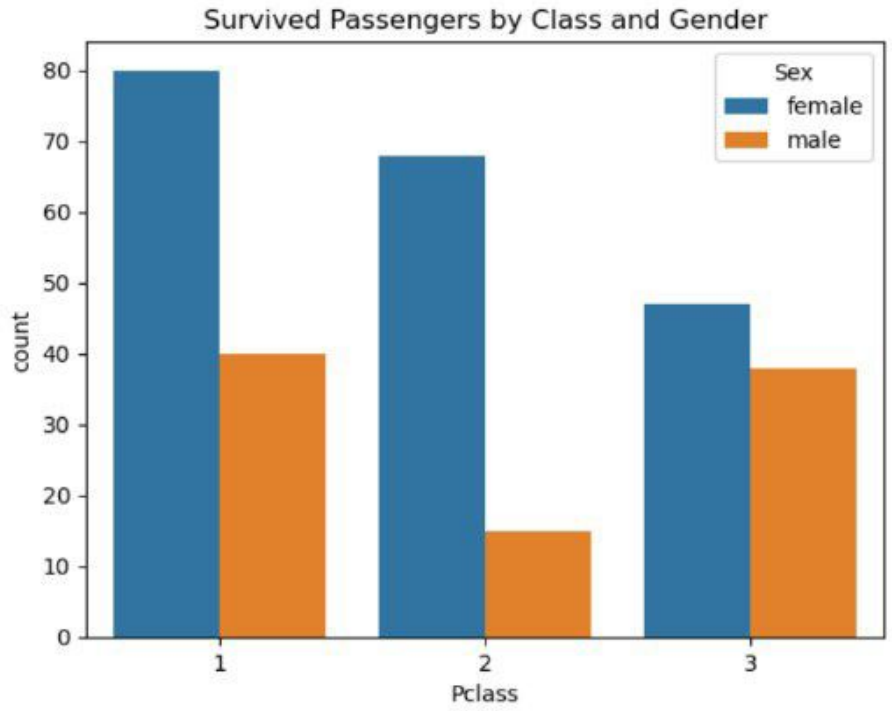
File Edit View Run Kernel Settings Help

Trusted

Code

Open in... Python 3 (ipykernel)

[23]:
```
sns.countplot(x='Pclass', hue='Sex', data=df[df['Survived'] == 1])
plt.title("Survived Passengers by Class and Gender")
plt.show()
```



Summary: Most female survivors were from 1st and 2nd class; most deaths were 3rd class males.

Home | Titanic (EDA)

localhost:8888/notebooks/Data%20Analytics%20Intern%20Elevate%20Lsabs%2FDay-05%2FTitanic%20(EDA).ipynb

jupyter Titanic (EDA) Last Checkpoint: 51 minutes ago

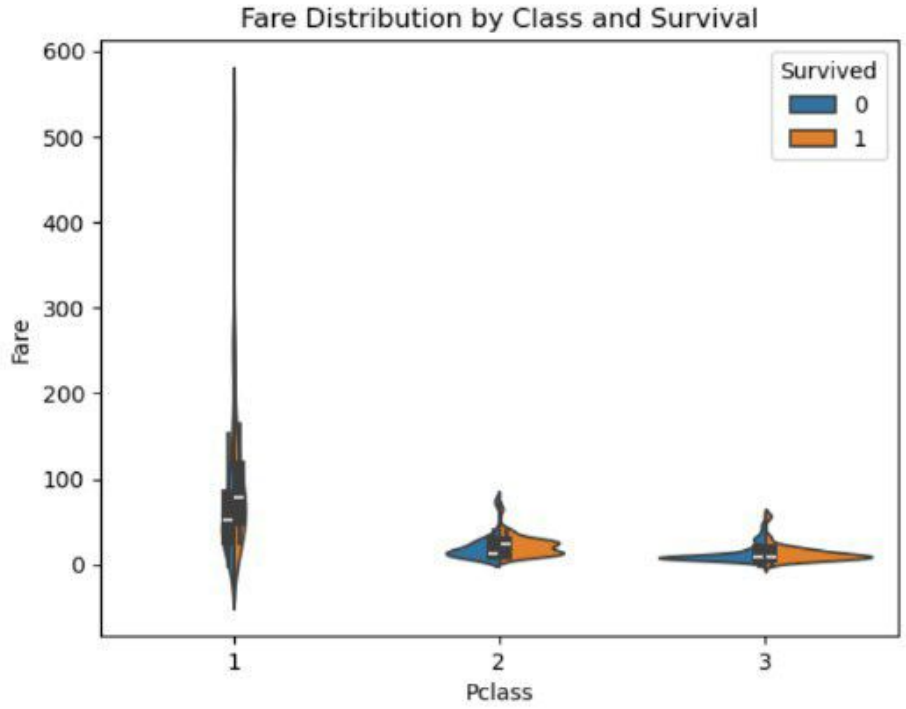
File Edit View Run Kernel Settings Help

Trusted

Code

Open in... Python 3 (ipykernel)

[25]:
```
sns.violinplot(x='Pclass', y='Fare', hue='Survived', data=df, split=True)
plt.title("Fare Distribution by Class and Survival")
plt.show()
```



Summary: Survivors in 1st class paid significantly higher fares.