EDA experiment 1: Multivariete Analysis

Name: Soumyadeep Ganguly

Reg no: 24MDT0082

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

titanic=pd.read_csv("titanic.csv")
titanic.head()
```

Out[1]:

	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 [2]:
```
#percentage of women survived
women = titanic.loc[titanic.Sex == 'female']["Survived"]
rate_women = sum(women)/len(women)

#percentage of men survived
men = titanic.loc[titanic.Sex == 'male']["Survived"]
```

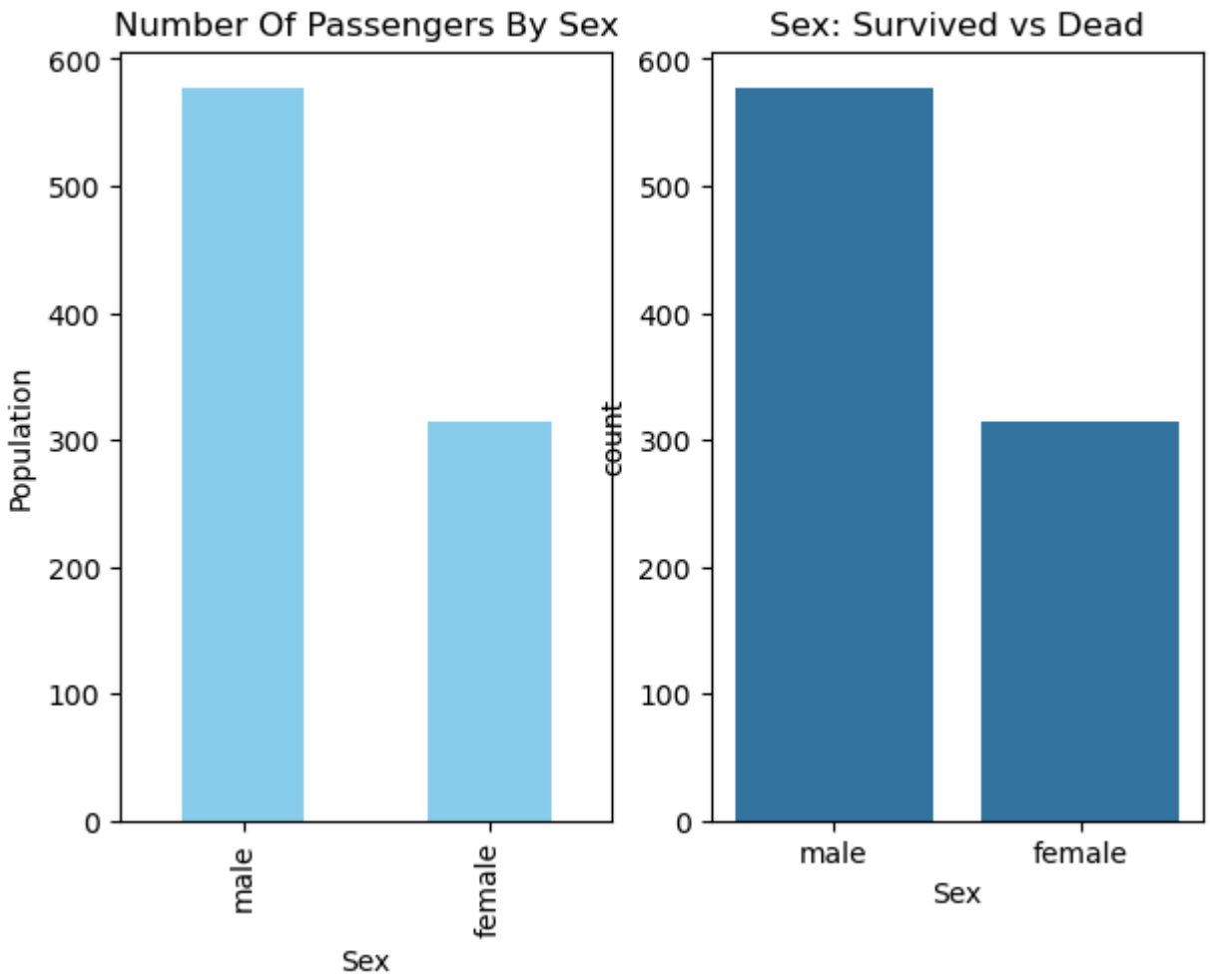
```
rate_men = sum(men)/len(men)

print(str(rate_women) +" % of women who survived." )
print(str(rate_men) + " % of men who survived." )
```

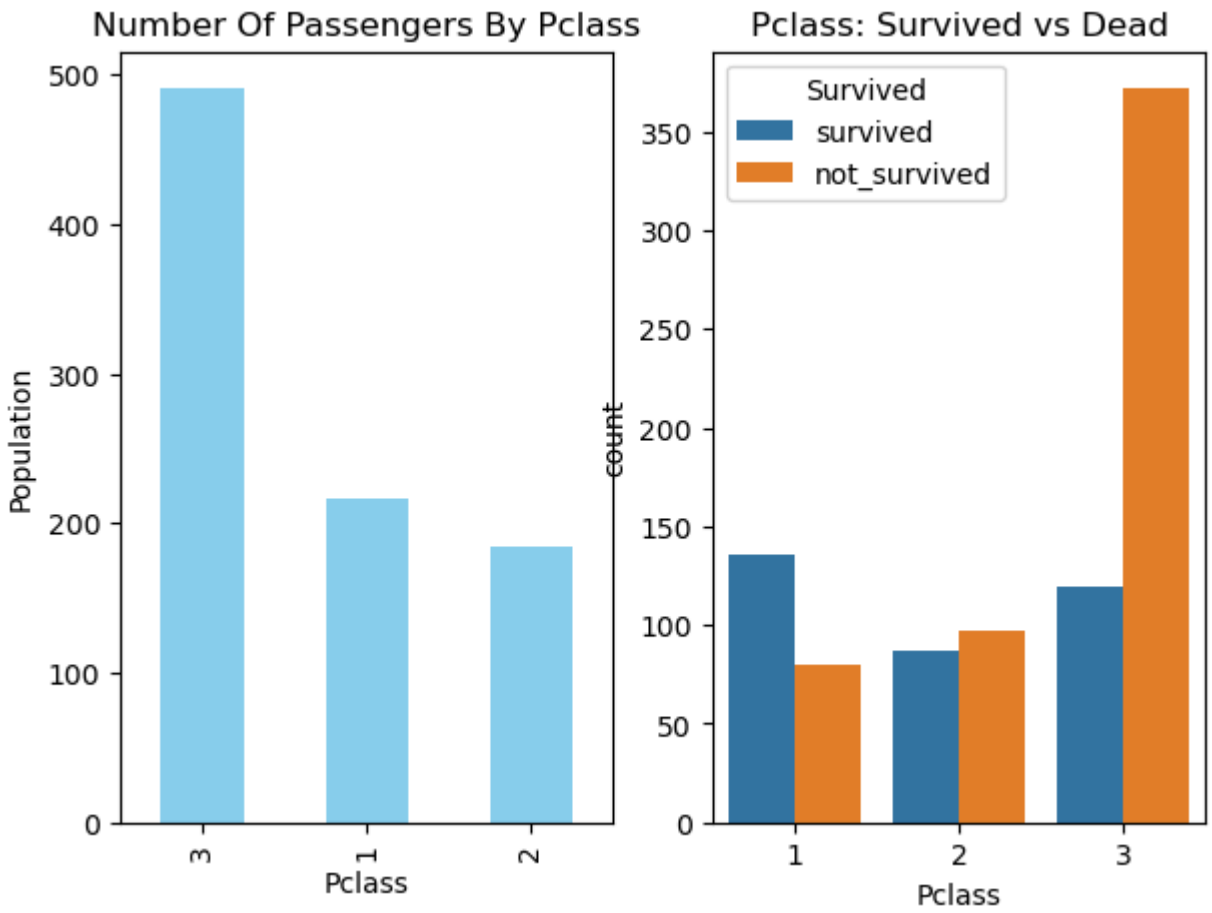
```
0.7420382165605095 % of women who survived.
0.18890814558058924 % of men who survived.
```

In [3]:
```
titanic['Survived'] = titanic['Survived'].map({0:"not_survived", 1:"survived"})

fig, ax = plt.subplots(1, 2, figsize = (7, 5))
titanic["Sex"].value_counts().plot.bar(color = "skyblue", ax = ax[0])
ax[0].set_title("Number Of Passengers By Sex")
ax[0].set_ylabel("Population")
sns.countplot(x="Sex", data=titanic, ax = ax[1])
ax[1].set_title("Sex: Survived vs Dead")
plt.show()
```



In [6]:
```
fig, ax = plt.subplots(1, 2, figsize = (7, 5))
titanic["Pclass"].value_counts().plot.bar(color = "skyblue", ax = ax[0])
ax[0].set_title("Number Of Passengers By Pclass")
ax[0].set_ylabel("Population")
sns.countplot(x="Pclass", hue = "Survived", data = titanic, ax = ax[1])
ax[1].set_title("Pclass: Survived vs Dead")
plt.show()
```



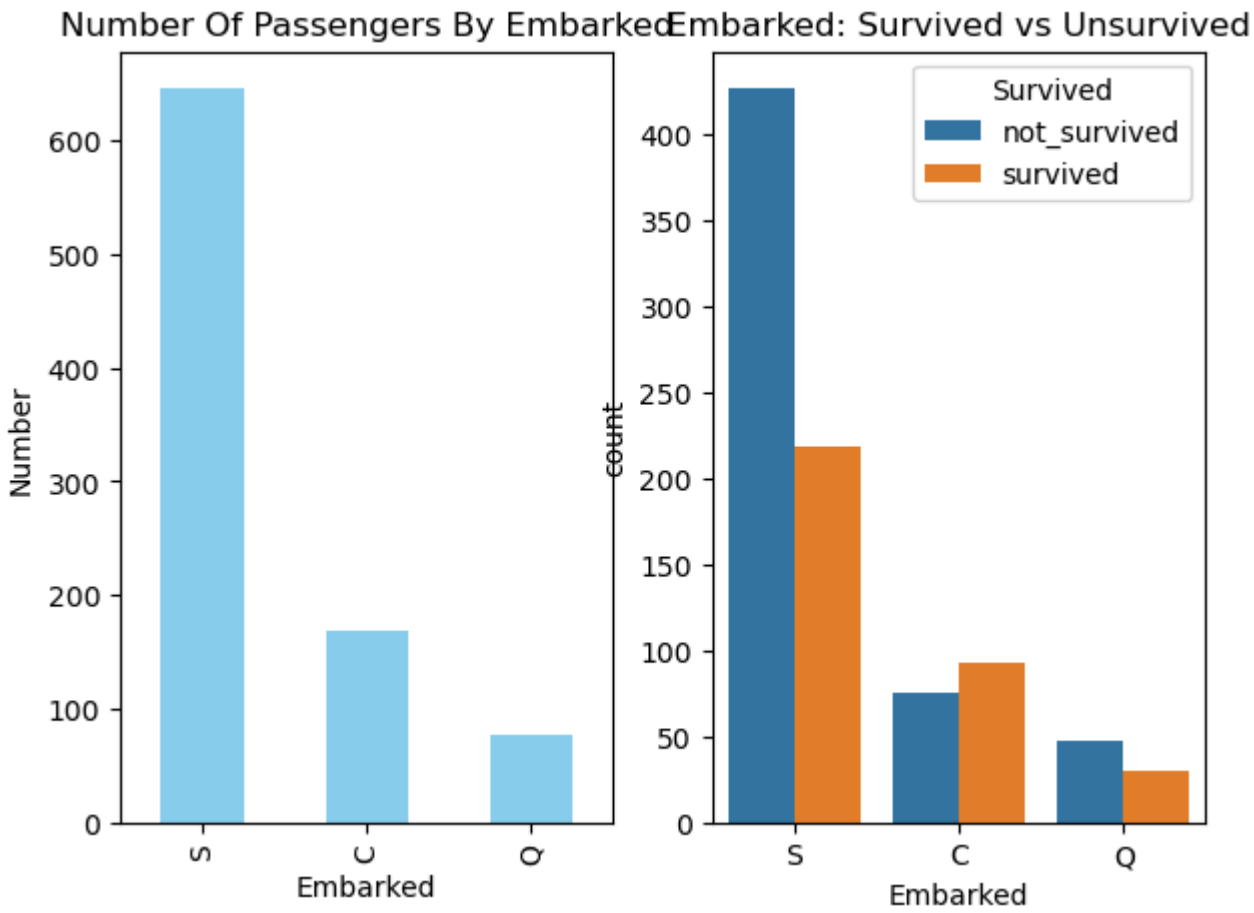
In [7]:
```
titanic["Embarked"] = titanic["Embarked"].fillna("S")
titanic
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	not_survived	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	survived	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	survived	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	survived	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	not_survived	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	not_survived	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	survived	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	not_survived	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	survived	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	not_survived	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

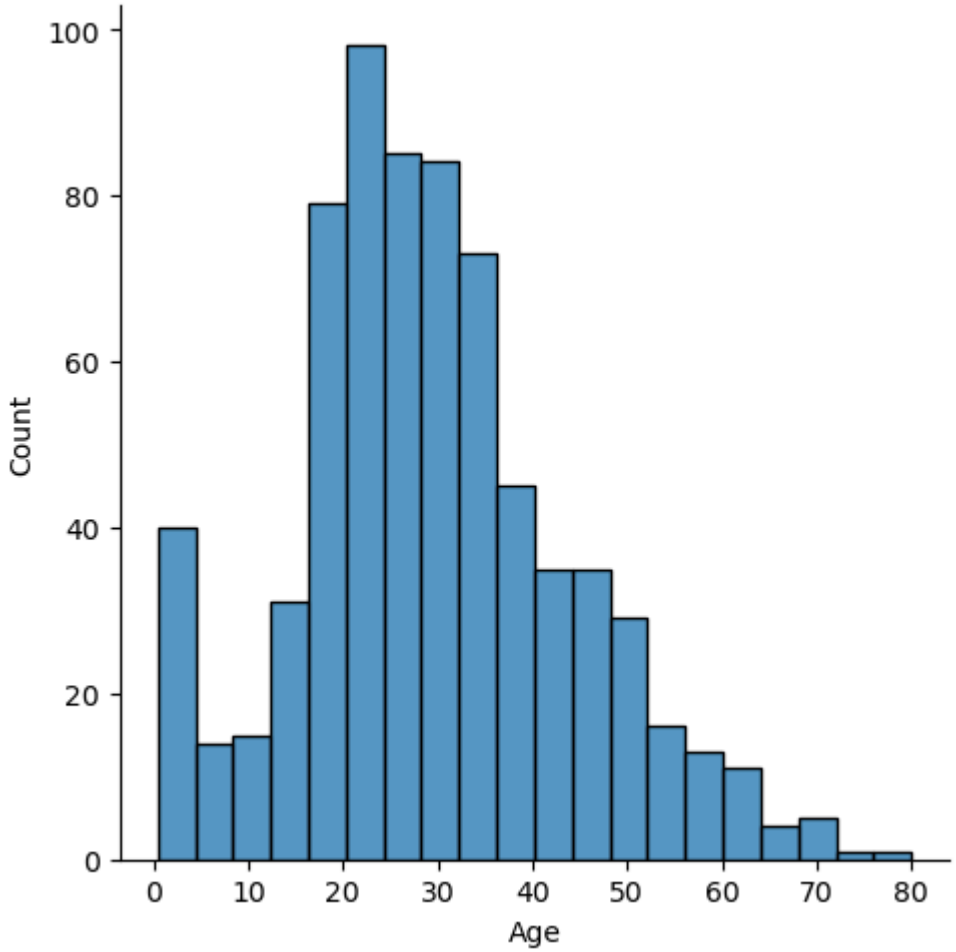
891 rows × 12 columns

In [9]:
```
fig, ax = plt.subplots(1, 2, figsize = (7, 5))
titanic["Embarked"].value_counts().plot.bar(color = "skyblue", ax = ax[0])
ax[0].set_title("Number Of Passengers By Embarked")
ax[0].set_ylabel("Number")
sns.countplot(x = "Embarked", hue = "Survived", data = titanic, ax = ax[1])
ax[1].set_title("Embarked: Survived vs Unsurvived")
plt.show()
```

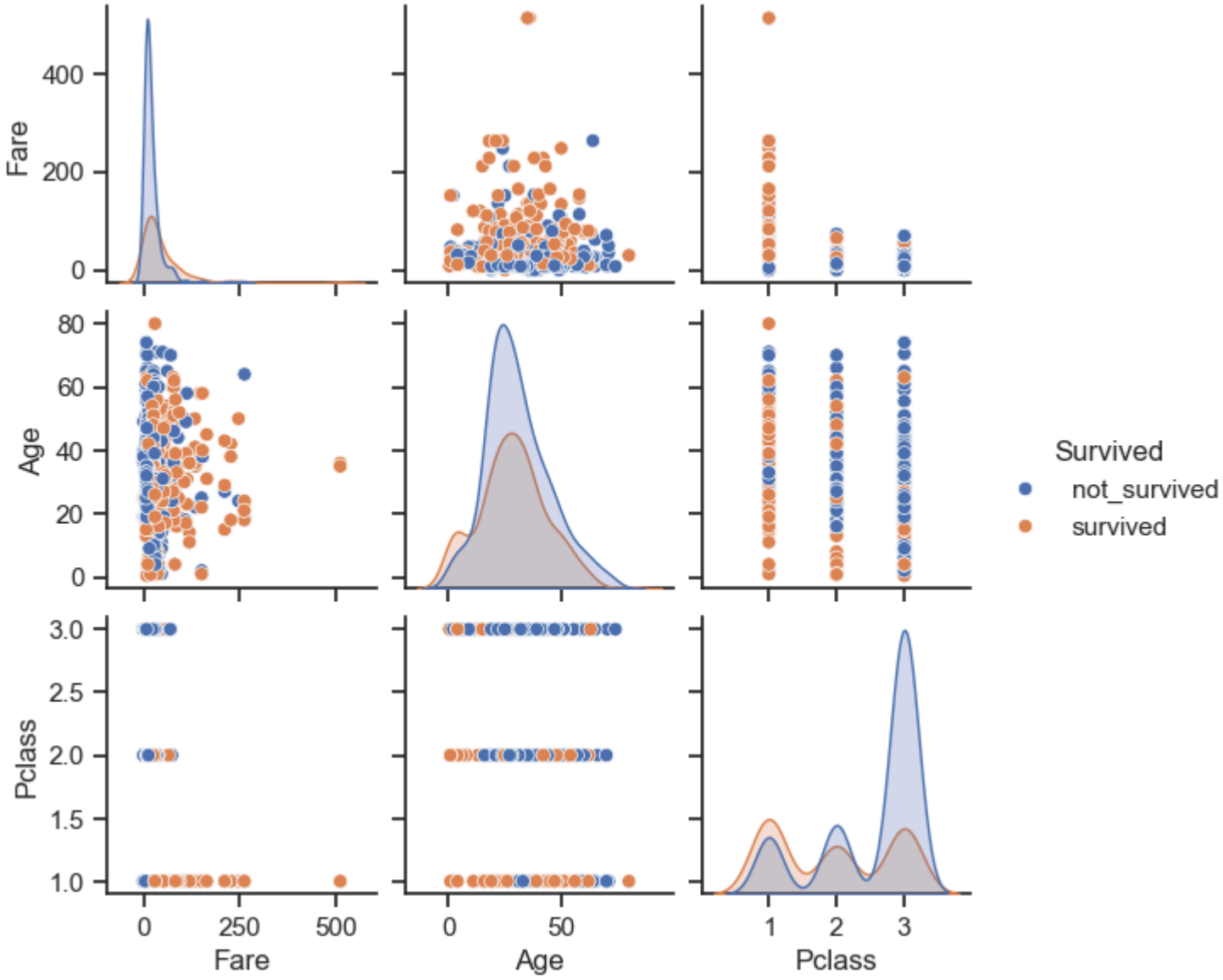


In [10]:
```
sns.displot(titanic['Age'].dropna())
```

Out[10]:
```
<seaborn.axisgrid.FacetGrid at 0x1c607d97f20>
```



In [11]:
```
sns.set(style="ticks", color_codes=True)
sns.pairplot(titanic,height=2,vars = [ 'Fare','Age','Pclass'], hue="Survived")
plt.show()
```



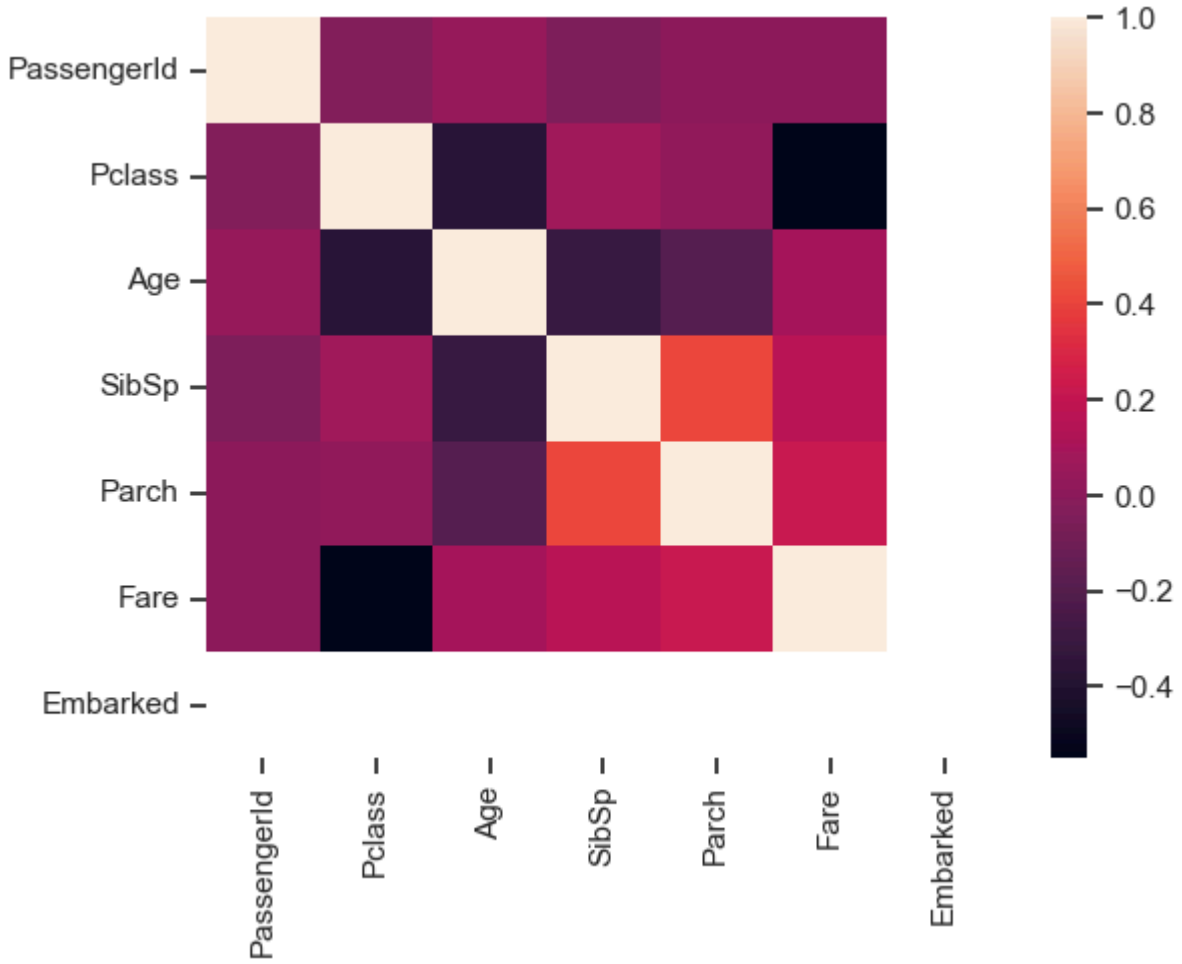
In [13]:
```
titanic['Embarked'] = titanic['Embarked'].map({"S":1, "C":2,"Q":2,"NaN":0})
Tcorrelation = titanic.corr(method='pearson', numeric_only=True)
Tcorrelation
```

Out[13]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Embarked
PassengerId	1.000000	-0.035144	0.036847	-0.057527	-0.001652	0.012658	NaN
Pclass	-0.035144	1.000000	-0.369226	0.083081	0.018443	-0.549500	NaN
Age	0.036847	-0.369226	1.000000	-0.308247	-0.189119	0.096067	NaN
SibSp	-0.057527	0.083081	-0.308247	1.000000	0.414838	0.159651	NaN
Parch	-0.001652	0.018443	-0.189119	0.414838	1.000000	0.216225	NaN
Fare	0.012658	-0.549500	0.096067	0.159651	0.216225	1.000000	NaN
Embarked	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [14]:
```
sns.heatmap(Tcorrelation,xticklabels=Tcorrelation.columns,
            yticklabels=Tcorrelation.columns)
```

Out[14]: <Axes: >


PassengerId
Pclass
Age
SibSp
Parch
Fare
Embarked
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

In [ ]: