

```
[2]: import pandas as pd
```

```
[4]: train=pd.read_csv('train.csv')
train
```

	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
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows x 12 columns

```
[6]: train=train.head()
```

```
[8]: train
```

```
[6]: train=train.head()
```

```
[8]: train
```

[8]:	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

```
[10]: train=train.describe(include='all')
      train
```

[10]:	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	5.000000	5.000000	5.000000	5	5	5.00000	5.000000	5.0	5	5.000000	2	5
unique	NaN	NaN	NaN	5	2	NaN	NaN	NaN	5	NaN	2	2
top	NaN	NaN	NaN	Braund, Mr. Owen Harris	female	NaN	NaN	NaN	A/5 21171	NaN	C85	S
freq	NaN	NaN	NaN	1	3	NaN	NaN	NaN	1	NaN	1	4
mean	3.000000	0.600000	2.200000	NaN	NaN	31.20000	0.600000	0.0	NaN	29.521660	NaN	NaN
std	1.581139	0.547723	1.095445	NaN	NaN	6.83374	0.547723	0.0	NaN	30.510029	NaN	NaN
min	1.000000	0.000000	1.000000	NaN	NaN	22.00000	0.000000	0.0	NaN	7.250000	NaN	NaN
25%	2.000000	0.000000	1.000000	NaN	NaN	26.00000	0.000000	0.0	NaN	7.925000	NaN	NaN
50%	3.000000	1.000000	3.000000	NaN	NaN	35.00000	1.000000	0.0	NaN	8.050000	NaN	NaN
75%	4.000000	1.000000	3.000000	NaN	NaN	35.00000	1.000000	0.0	NaN	53.100000	NaN	NaN
max	5.000000	1.000000	3.000000	NaN	NaN	38.00000	1.000000	0.0	NaN	71.283300	NaN	NaN



```
[12]: train=train.info()
train

<class 'pandas.core.frame.DataFrame'>
Index: 11 entries, count to max
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  8 non-null      float64
1   Survived     8 non-null      float64
2   Pclass       8 non-null      float64
3   Name         4 non-null      object
4   Sex          4 non-null      object
5   Age          8 non-null      float64
6   SibSp        8 non-null      float64
7   Parch        8 non-null      float64
8   Ticket       4 non-null      object
9   Fare         8 non-null      float64
10  Cabin        4 non-null      object
11  Embarked     4 non-null      object
dtypes: float64(7), object(5)
memory usage: 1.1+ KB
```

```
[28]: import pandas as pd
import seaborn as sns
```

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

[30]:	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

```
10 Cabin      4 non-null object
11 Embarked   4 non-null object
dtypes: float64(7), object(5)
memory usage: 1.1+ KB
```

```
[28]: import pandas as pd
import seaborn as sns
```

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

```
[30]:
```

	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
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

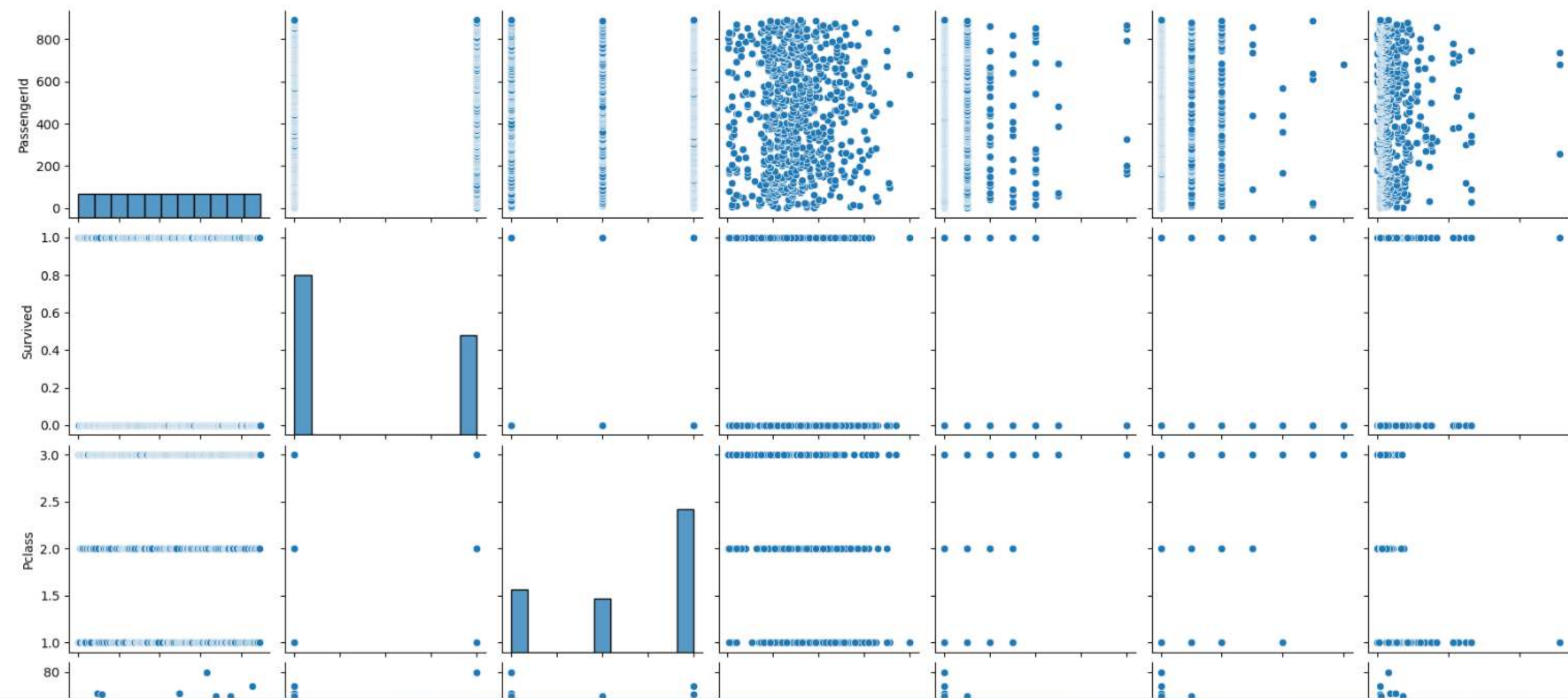
```
[32]: import seaborn as sns
```





[36]: sns.pairplot(train)

[36]: <seaborn.axisgrid.PairGrid at 0x2703a5e7950>



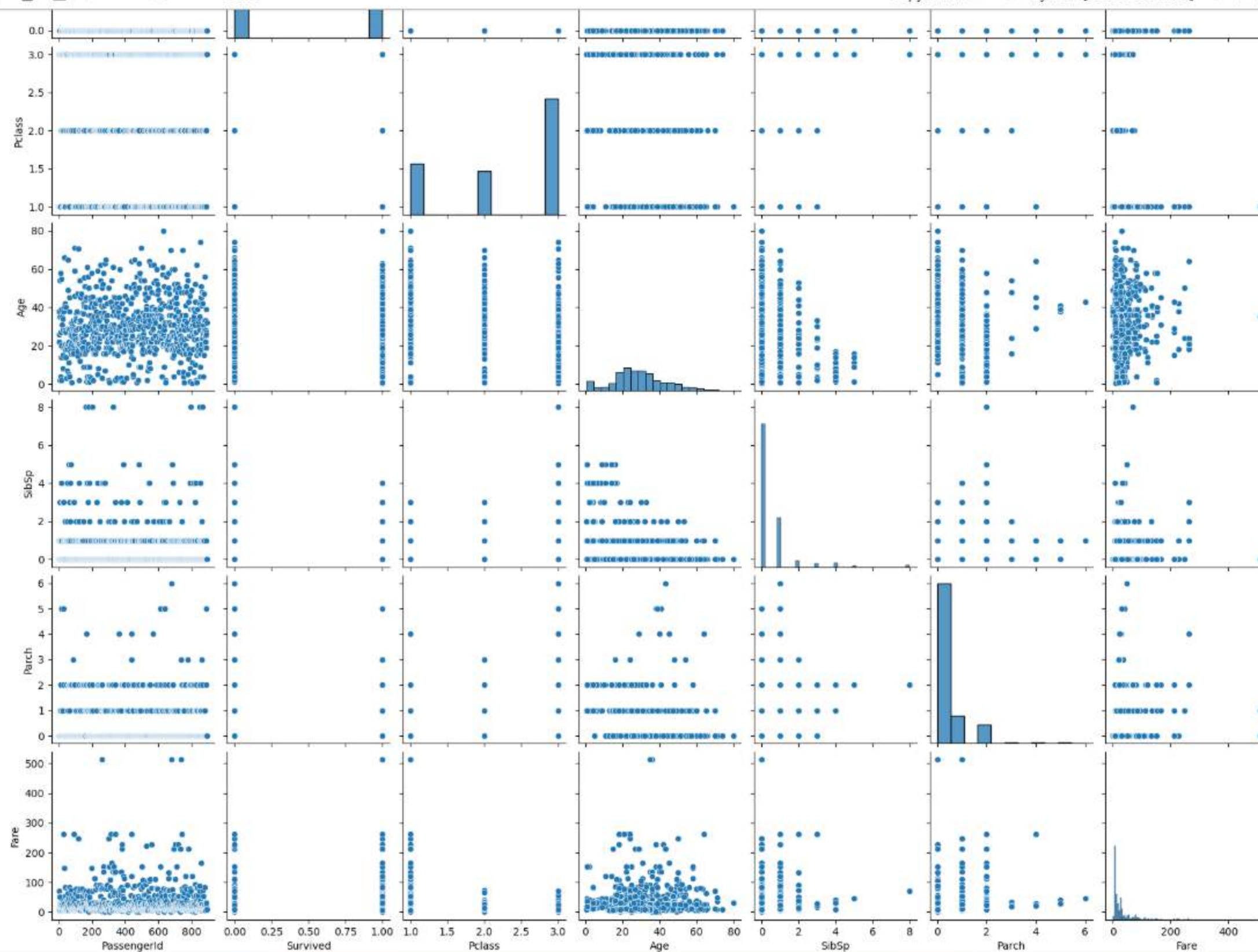
jupyter Untitled Last Checkpoint: 10 hours ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python [conda env:base]





```
[38]: print(train)
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
..	...	...	...	
886	887	0	2	
887	888	1	1	
888	889	0	3	
889	890	1	1	
890	891	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	
..	...	...	...	...	
886	Montvila, Rev. Juozas	male	27.0	0	
887	Graham, Miss. Margaret Edith	female	19.0	0	
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	
889	Behr, Mr. Karl Howell	male	26.0	0	
890	Dooley, Mr. Patrick	male	32.0	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
..	...	...	...	...	...
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
888	2	W./C. 6607	23.4500	NaN	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

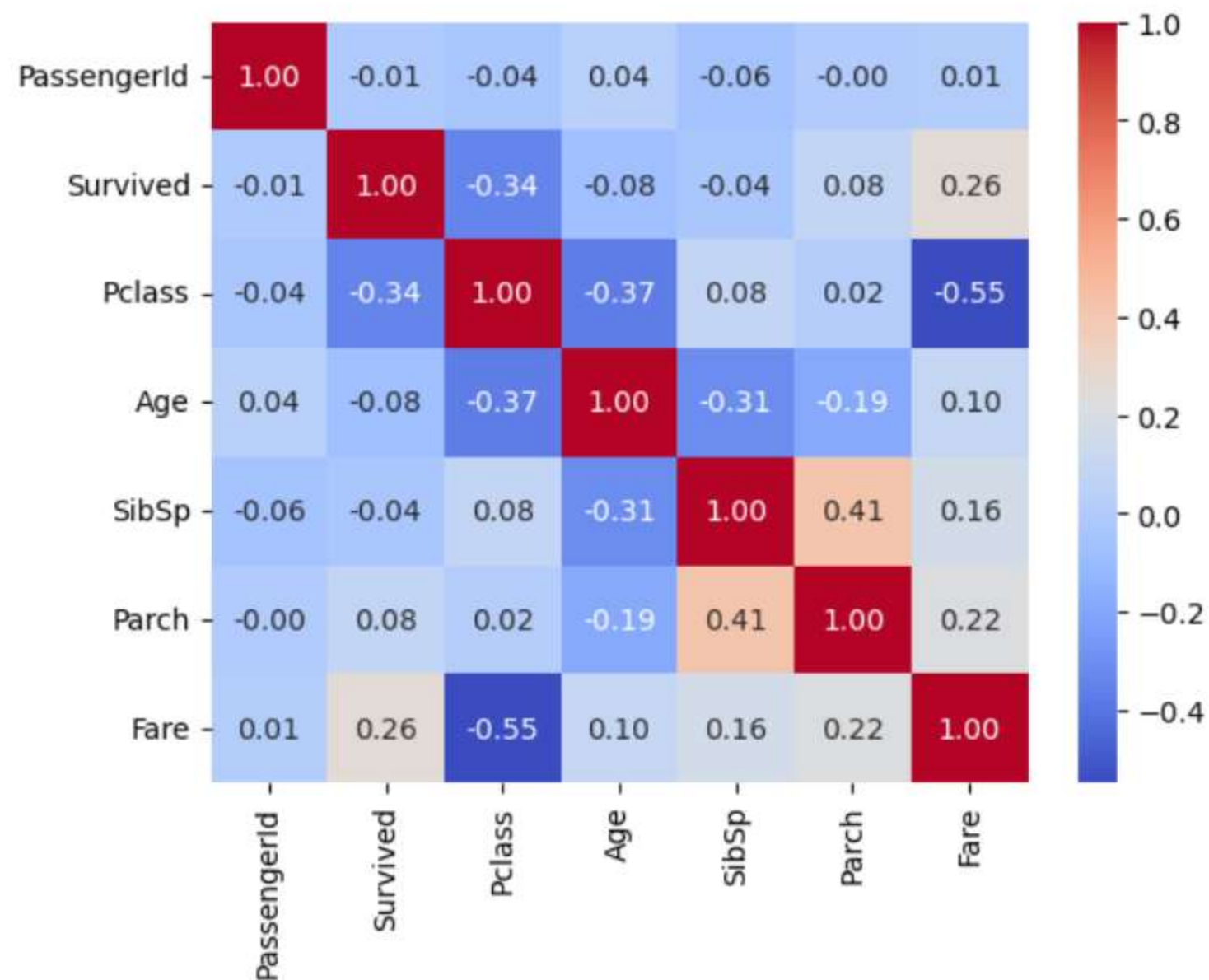
[891 rows x 12 columns]

[891 rows x 12 columns]

```
[76]: corr_matrix = train.select_dtypes(include=['number']).corr()
```

```
[78]: sns.heatmap(corr_matrix, annot=True, cmap="coolwarm", fmt=".2f")
```

[78]: <Axes: >





```
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), int8(1), object(4)
memory usage: 77.6+ KB
None
```

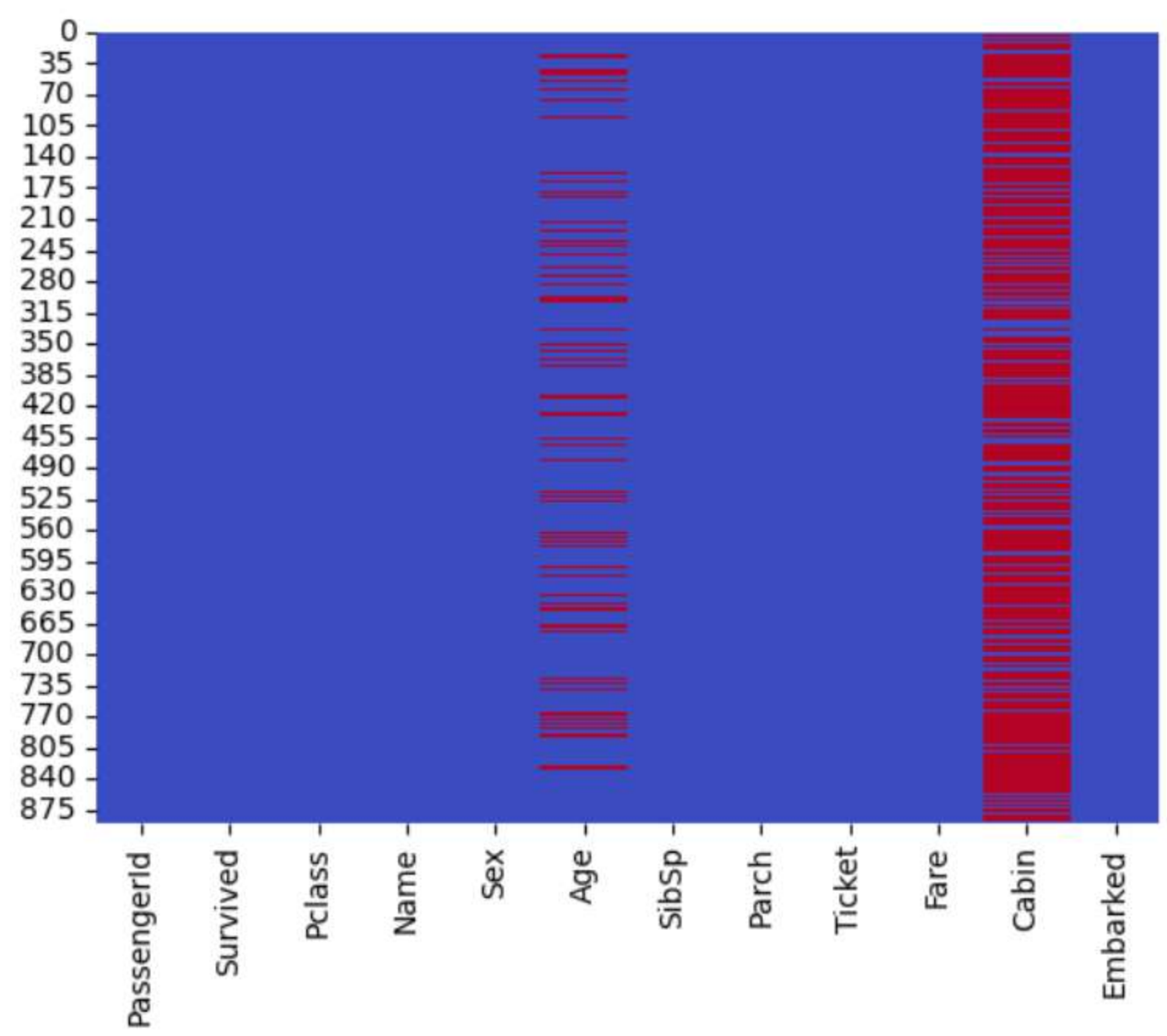
```
[106]: # Summary statistics
print(train.describe())
```

	PassengerId	Survived	Pclass	Sex	Age
count	891.000000	891.000000	891.000000	891.000000	714.000000
mean	446.000000	0.383838	2.308642	0.647587	29.699118
std	257.353842	0.486592	0.836071	0.477990	14.526497
min	1.000000	0.000000	1.000000	0.000000	0.420000
25%	223.500000	0.000000	2.000000	0.000000	20.125000
50%	446.000000	0.000000	3.000000	1.000000	28.000000
75%	668.500000	1.000000	3.000000	1.000000	38.000000
max	891.000000	1.000000	3.000000	1.000000	80.000000

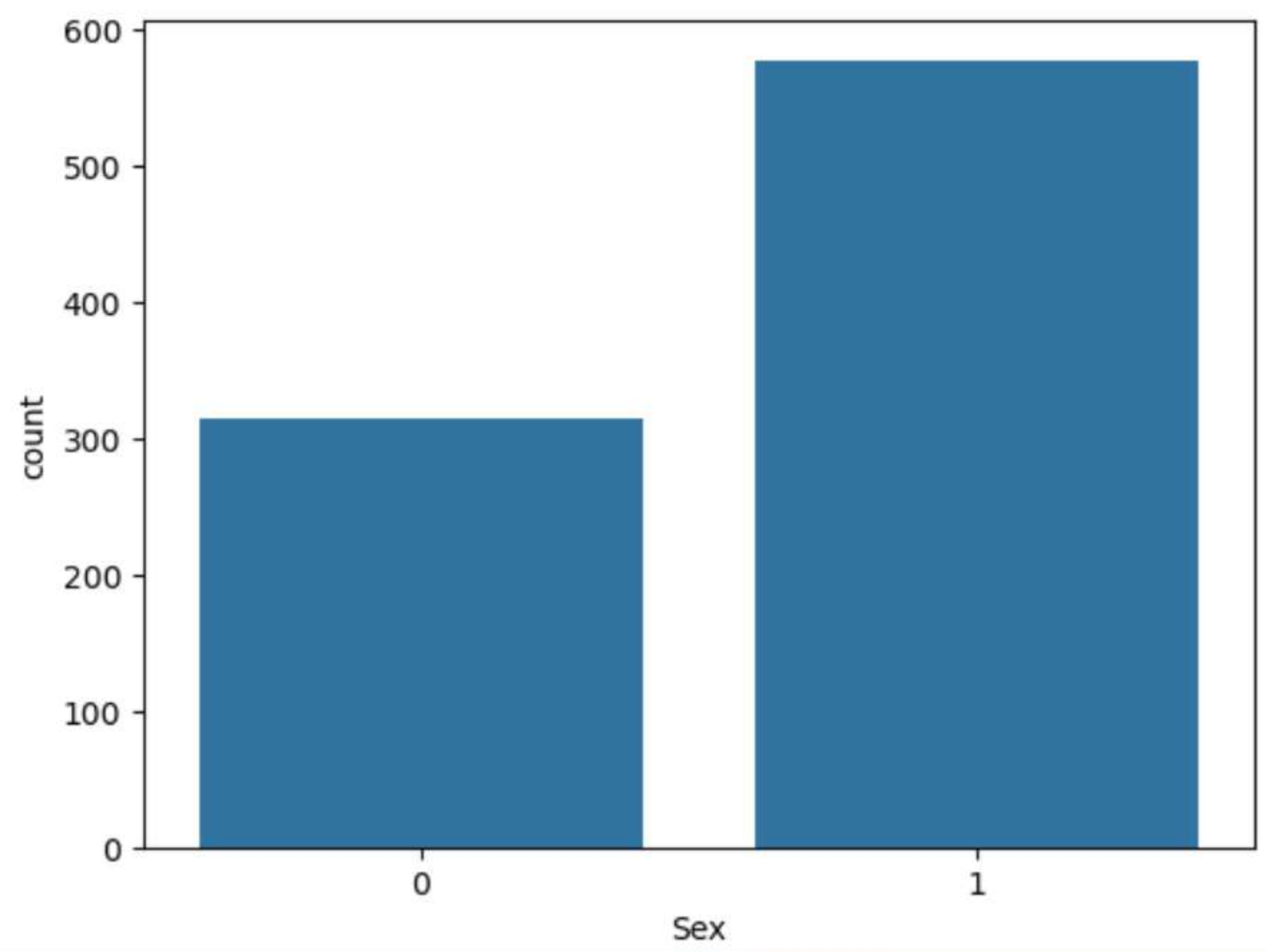
	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000
mean	0.523008	0.381594	32.204208
std	1.102743	0.806057	49.693429
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	7.910400
50%	0.000000	0.000000	14.454200
75%	1.000000	0.000000	31.000000
max	8.000000	6.000000	512.329200

```
[108]: sns.heatmap(train.isnull(), cmap="coolwarm", cbar=False)
plt.show()
```

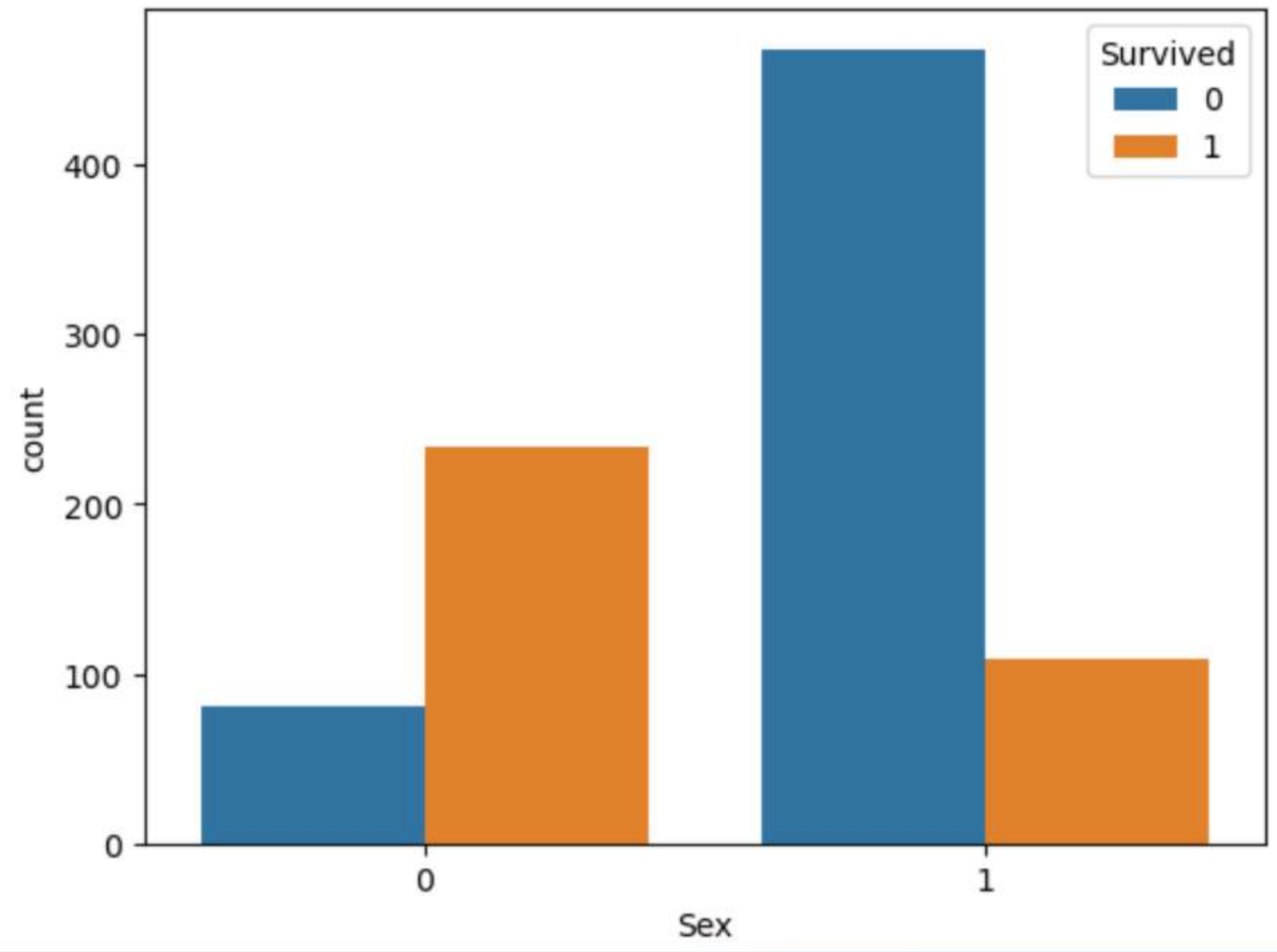




```
[114]: sns.countplot(x='Sex', data=train)
plt.show()
```



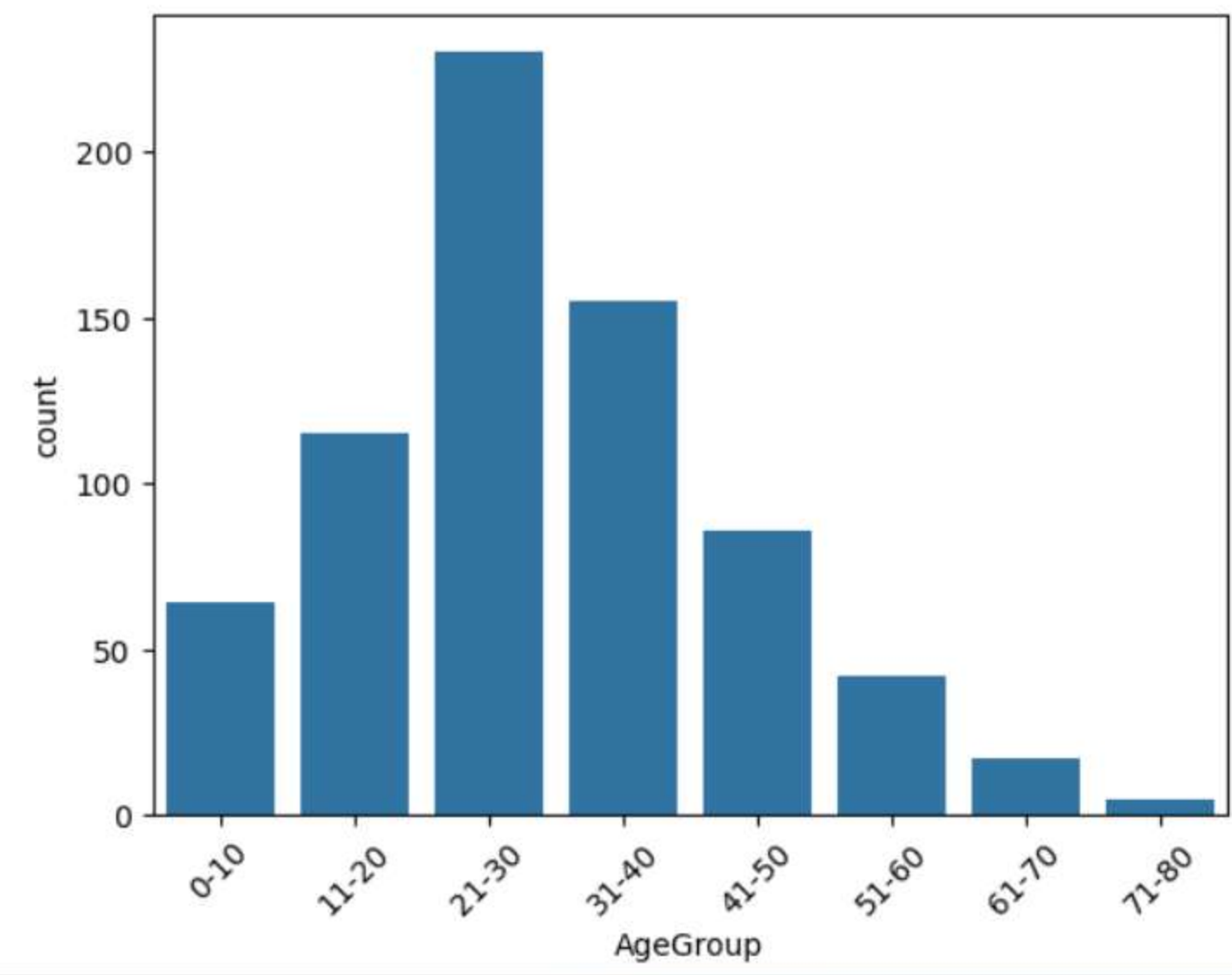
```
[118]: sns.countplot(x='Sex', hue='Survived', data=train)
plt.show()
```





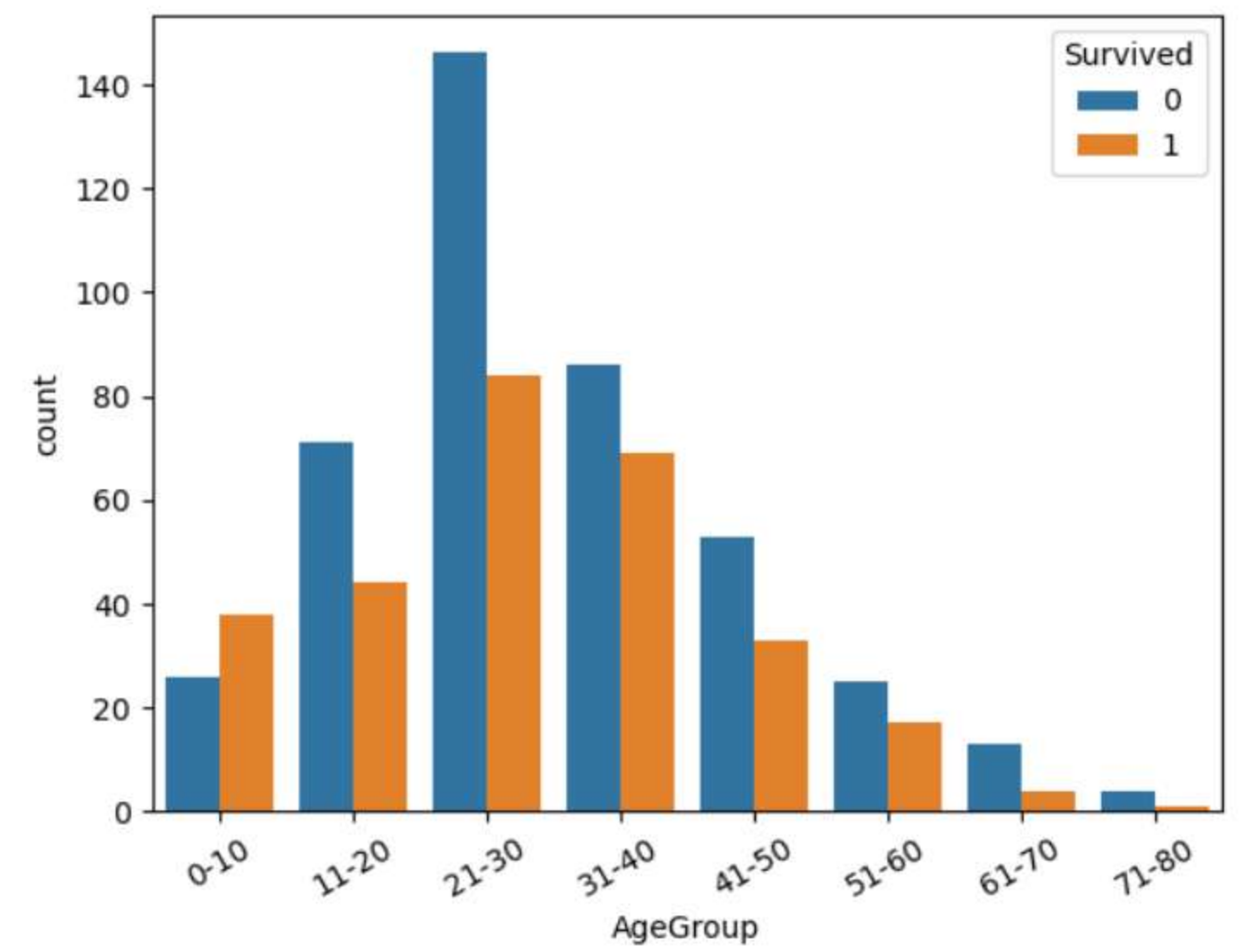
```
[122]: # Define age bins
train['AgeGroup'] = pd.cut(train['Age'], bins=[0, 10, 20, 30, 40, 50, 60, 70, 80],
                           labels=["0-10", "11-20", "21-30", "31-40", "41-50", "51-60", "61-70", "71-80"])

# Plot passenger count by age group
sns.countplot(x='AgeGroup', data=train)
plt.xticks(rotation=45) # Rotate labels for better readability
plt.show()
```

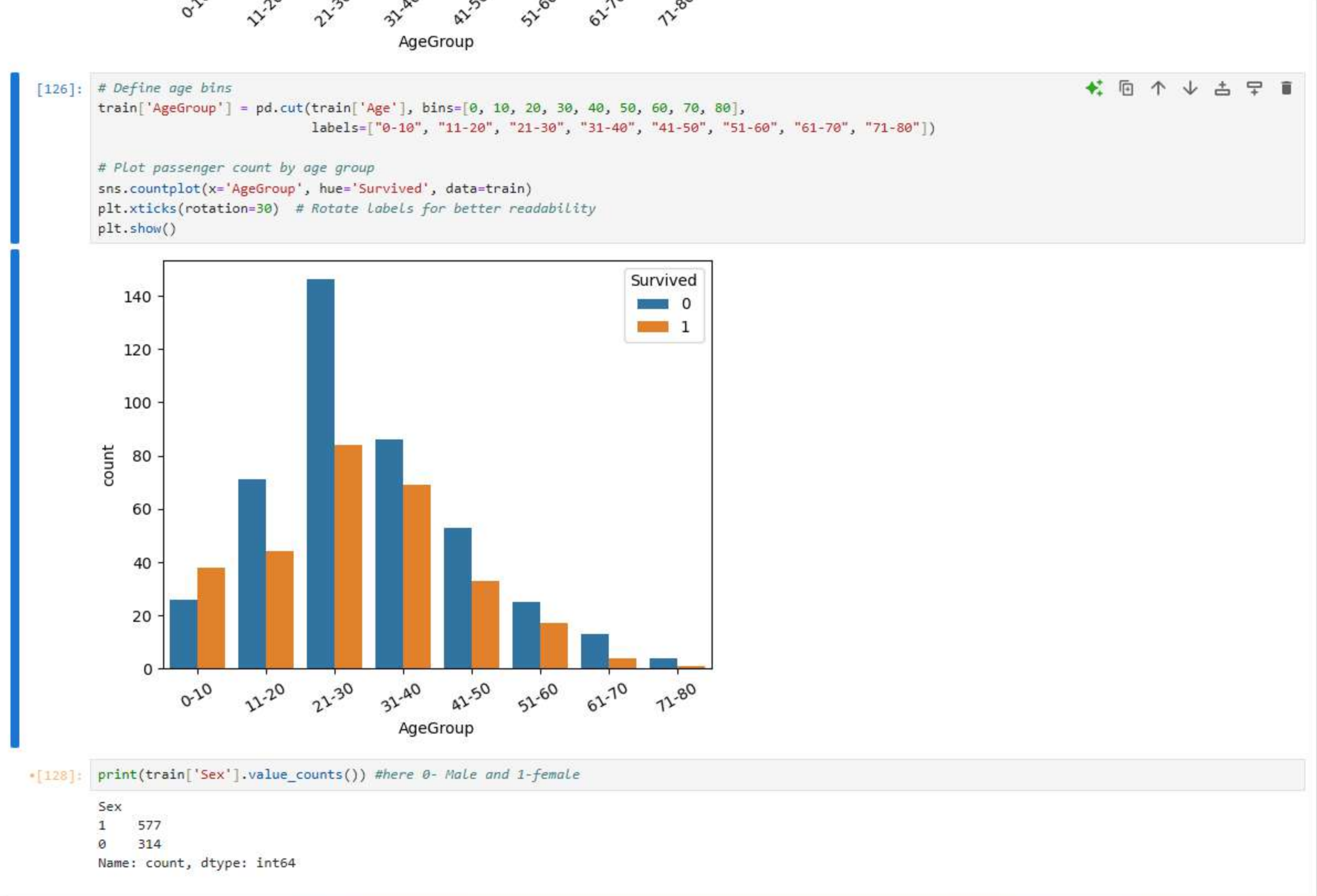


```
[126]: # Define age bins
train['AgeGroup'] = pd.cut(train['Age'], bins=[0, 10, 20, 30, 40, 50, 60, 70, 80],
                           labels=["0-10", "11-20", "21-30", "31-40", "41-50", "51-60", "61-70", "71-80"])

# Plot passenger count by age group
sns.countplot(x='AgeGroup', hue='Survived', data=train)
plt.xticks(rotation=30) # Rotate labels for better readability
plt.show()
```







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

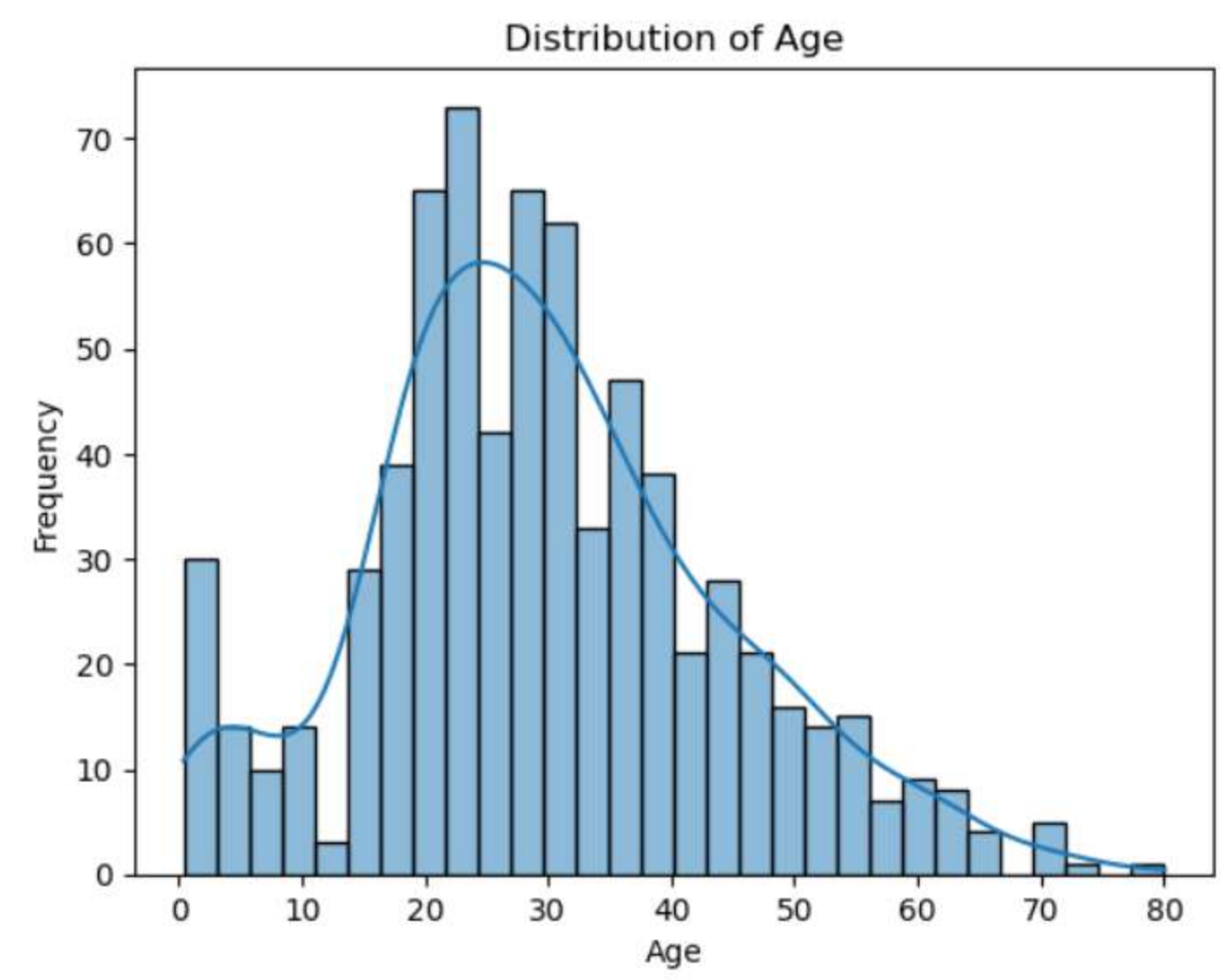
```
[12]: train=pd.read_csv('train.csv')
      train
```

[12]:	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
	...	...	...	...	...	...	...	...	...	...	...	...	...
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

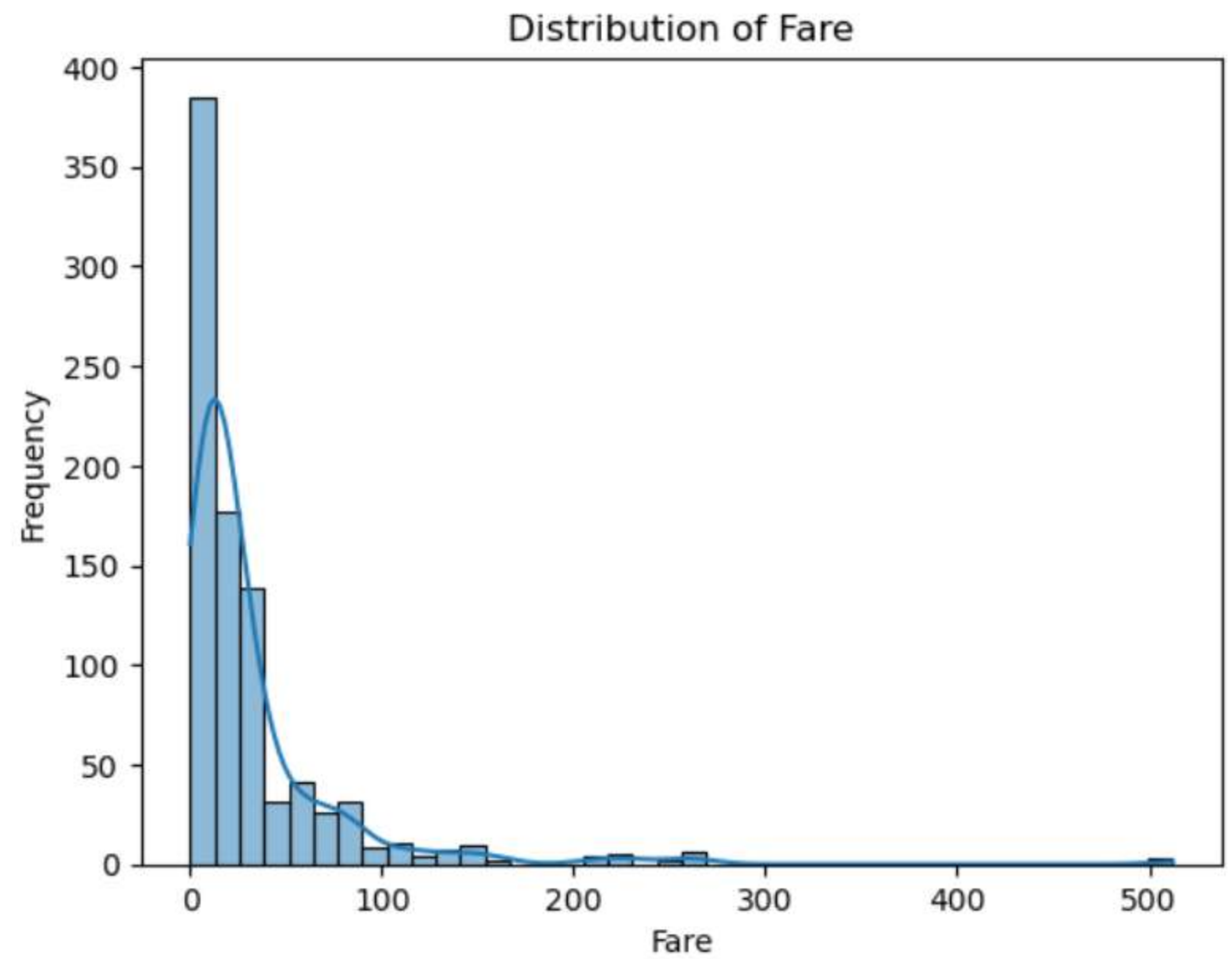
891 rows x 12 columns



```
[16]: # Histogram
sns.histplot(train['Age'], bins=30, kde=True) #bins=30: Defines the number of intervals in the histogram
#kde=True: Adds a smooth density curve to highlight patterns.
plt.title("Distribution of Age")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()
```

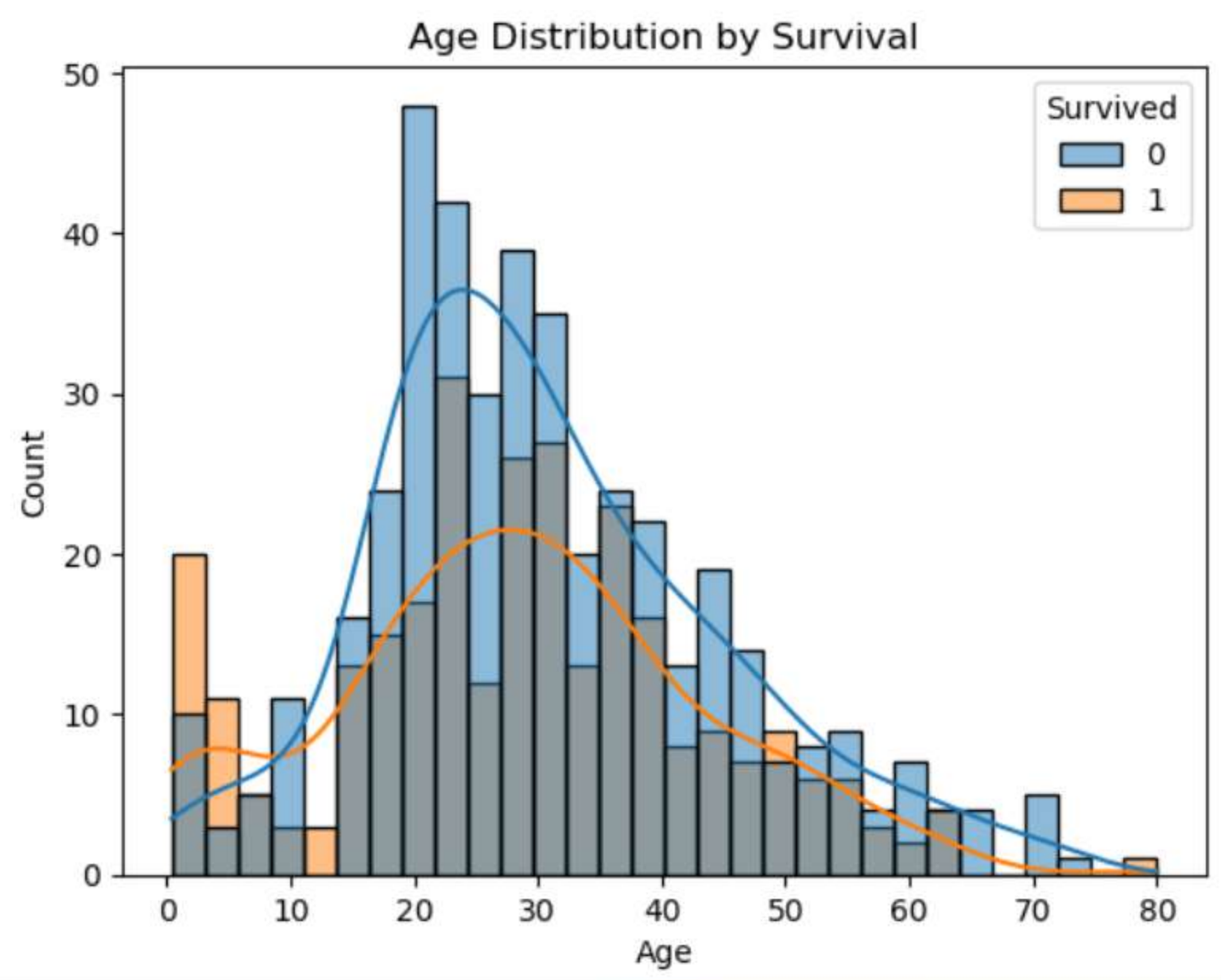


```
[34]: sns.histplot(train['Fare'], bins=40, kde=True)
plt.title("Distribution of Fare")
plt.xlabel("Fare")
plt.ylabel("Frequency")
plt.show()
```

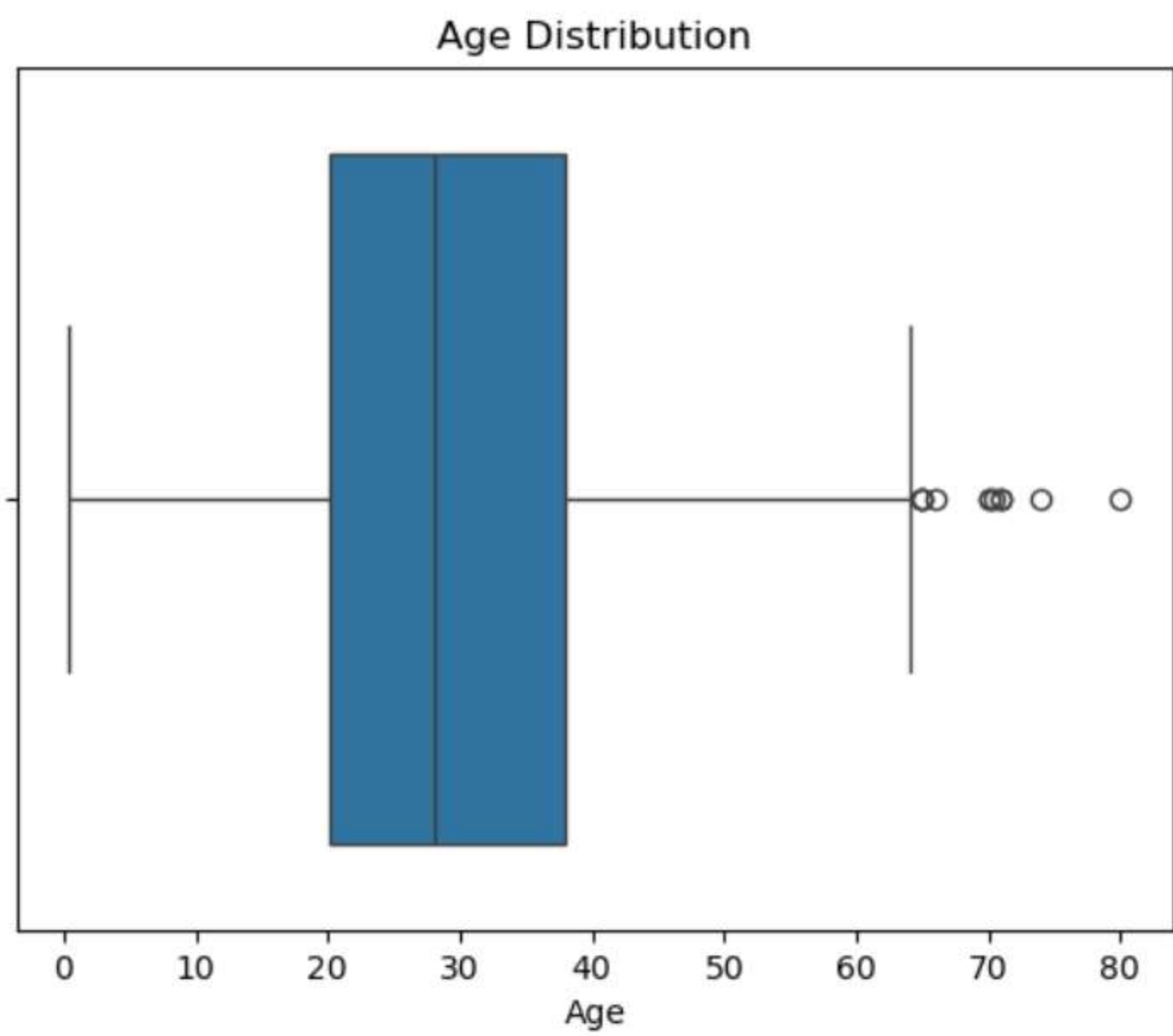




```
[38]: sns.histplot(train, x="Age", hue="Survived", bins=30, kde=True)
plt.title("Age Distribution by Survival")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```

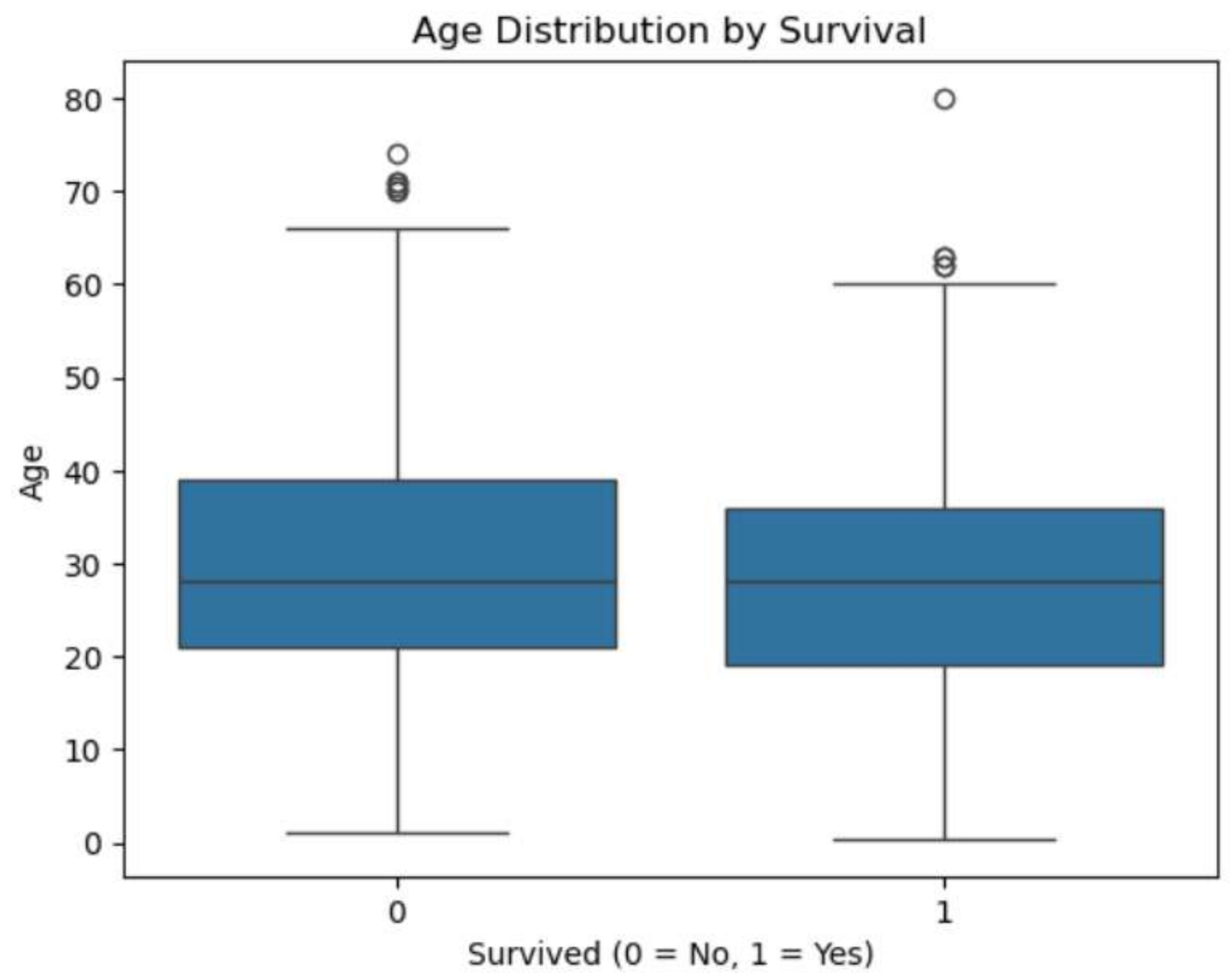


```
[32]: # Boxplot
sns.boxplot(x=train['Age'])
plt.title("Age Distribution")
plt.xlabel("Age")
plt.show()
```

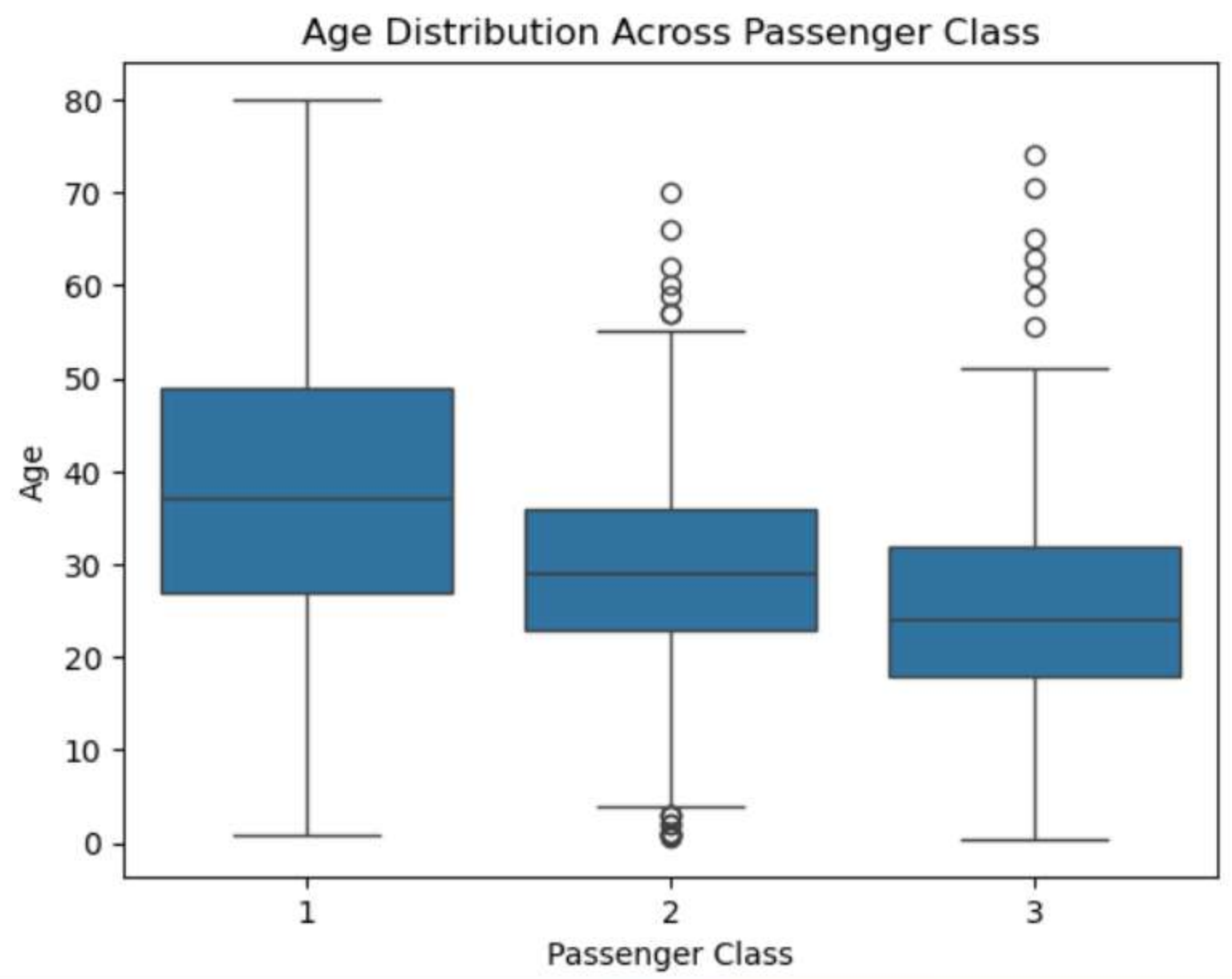




```
[18]: sns.boxplot(x='Survived', y='Age', data=train)
plt.title("Age Distribution by Survival")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Age")
plt.show()
```

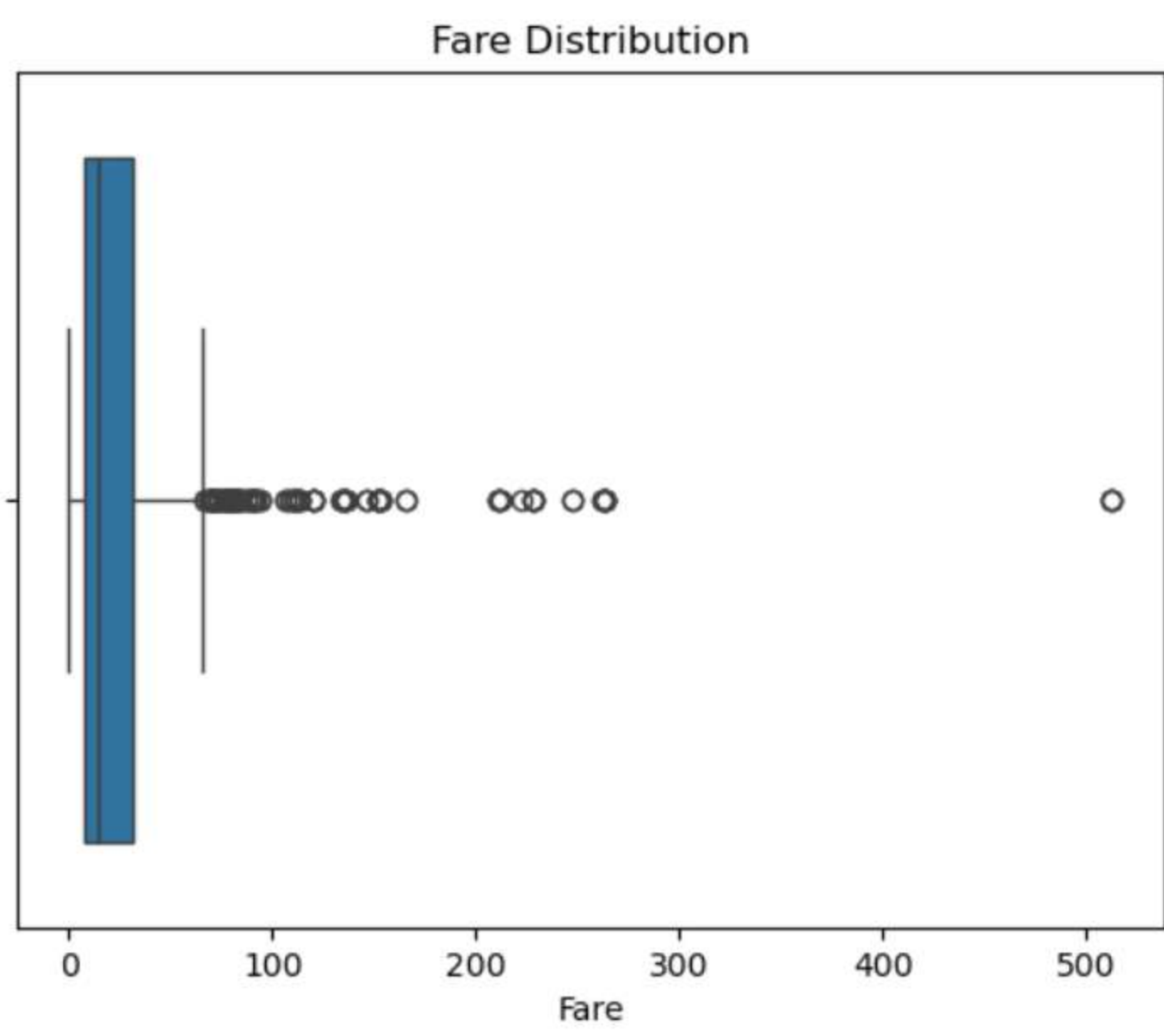


```
[26]: sns.boxplot(x='Pclass', y='Age', data=train)
plt.title("Age Distribution Across Passenger Class")
plt.xlabel("Passenger Class")
plt.ylabel("Age")
plt.show()
```

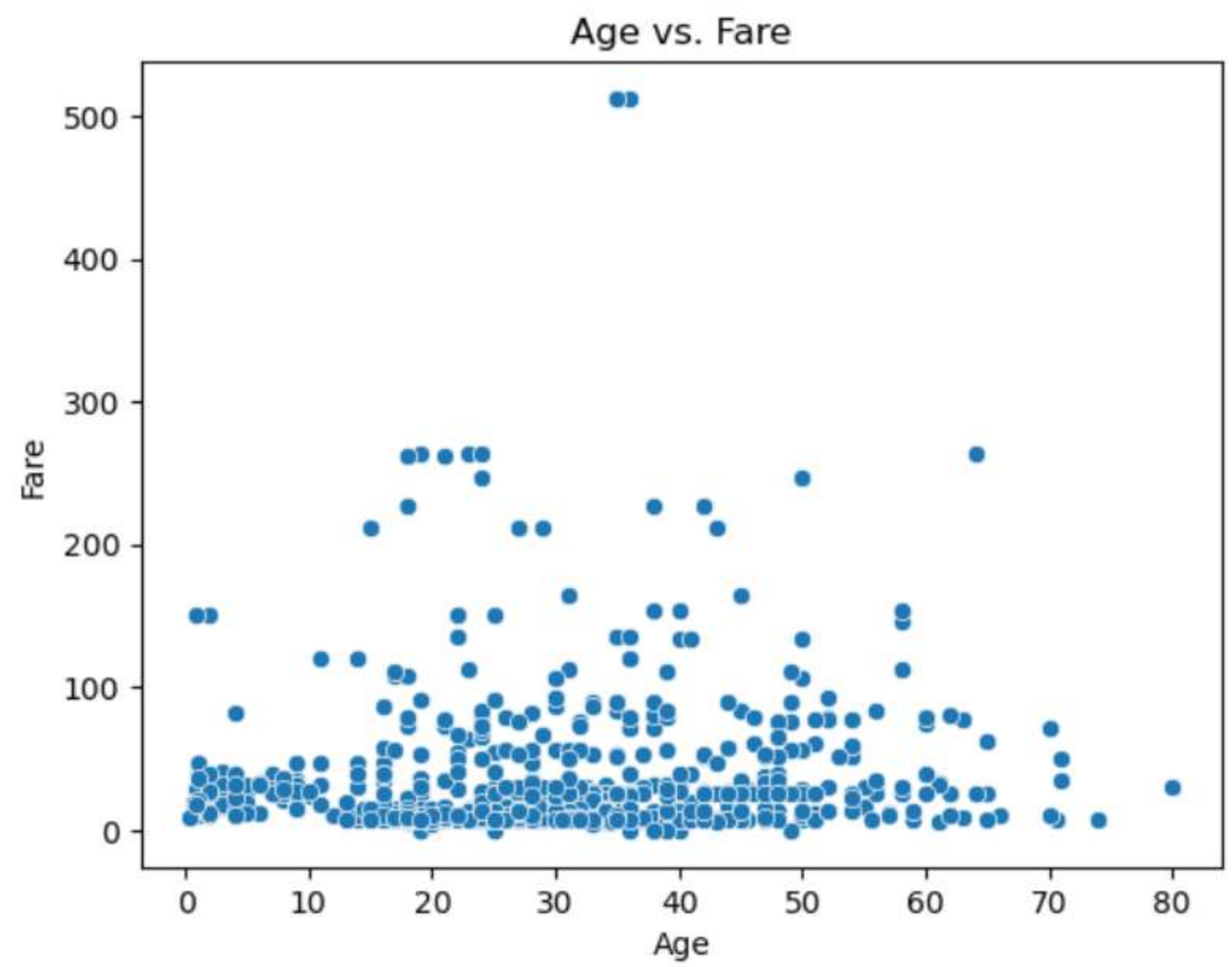




```
[28]: sns.boxplot(x=train['Fare'])  
plt.title("Fare Distribution")  
plt.xlabel("Fare")  
plt.show()
```

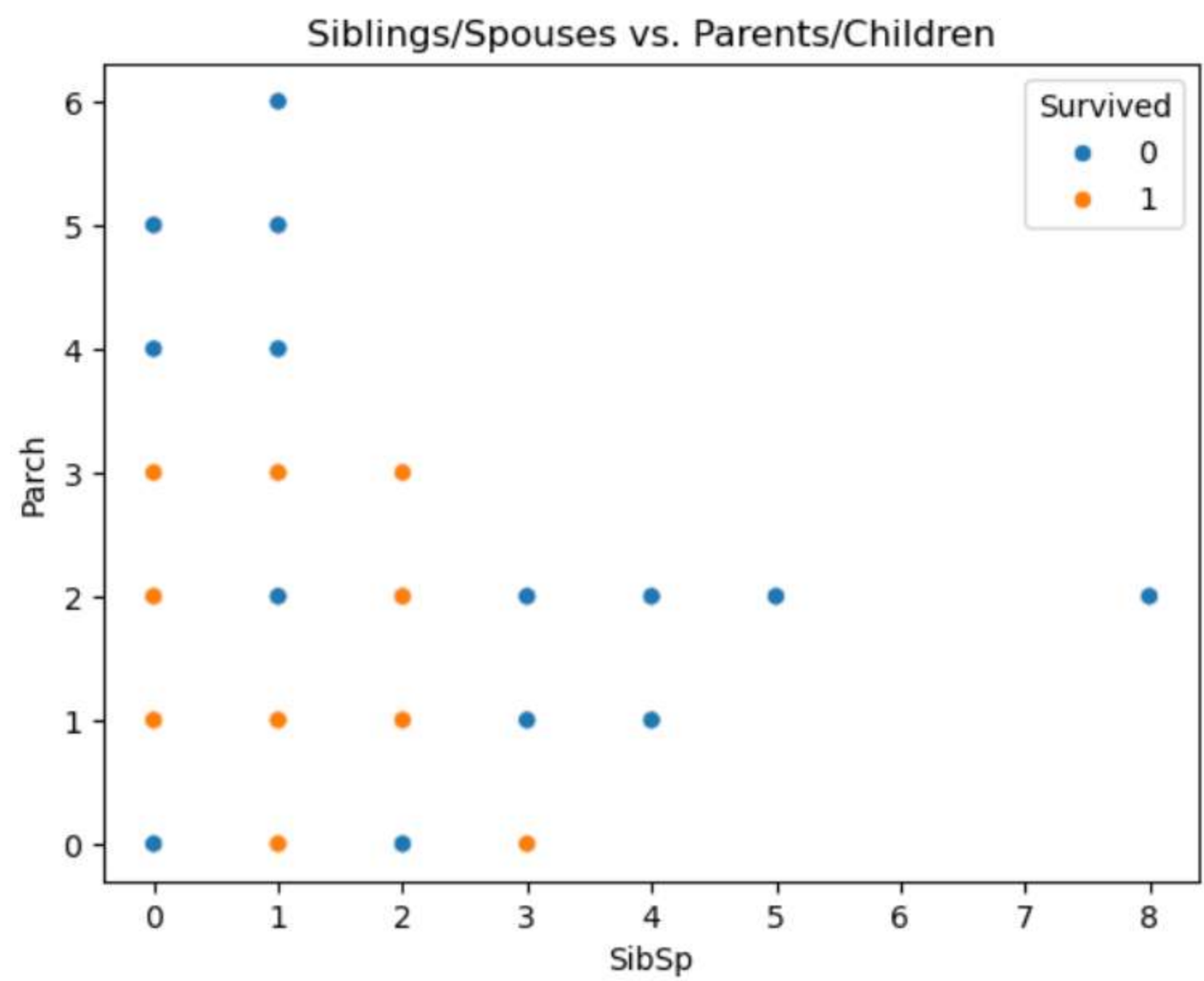


```
[20]: sns.scatterplot(x='Age', y='Fare', data=train)
plt.title("Age vs. Fare")
plt.xlabel("Age")
plt.ylabel("Fare")
plt.show()
```





```
[22]: sns.scatterplot(x='SibSp', y='Parch', data=train, hue='Survived')
plt.title("Siblings/Spouses vs. Parents/Children")
plt.xlabel("SibSp")
plt.ylabel("Parch")
plt.show()
```



```
[24]: sns.scatterplot(x='Age', y='Fare', hue='Survived', data=train)
plt.title("Fare vs. Age Colored by Survival")
plt.xlabel("Age")
plt.ylabel("Fare")
plt.show()
```

