

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

```
In [2]: dataset = sns.load_dataset('titanic')
dataset.head()
```

Out[2]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	d
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	I
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	I
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	I

```
In [3]: dataset.shape
```

Out[3]: (891, 15)

```
In [4]: dataset.describe()
```

Out[4]:

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

```
In [5]: dataset.isnull().sum()
```

```
Out[5]: survived          0
pclass          0
sex             0
age            177
sibsp          0
parch          0
fare           0
embarked       2
class          0
who            0
adult_male     0
deck          688
embark_town    2
alive          0
alone          0
dtype: int64
```

```
In [6]: dataset.info()
```

```
<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
```

```
In [7]: dataset['age']=dataset['age'].fillna(np.mean(dataset['age']))
dataset['deck']=dataset['deck'].fillna('A')
dataset['embarked']=dataset['embarked'].fillna(dataset['embarked'].mode()[0])
dataset.isnull().sum()
```

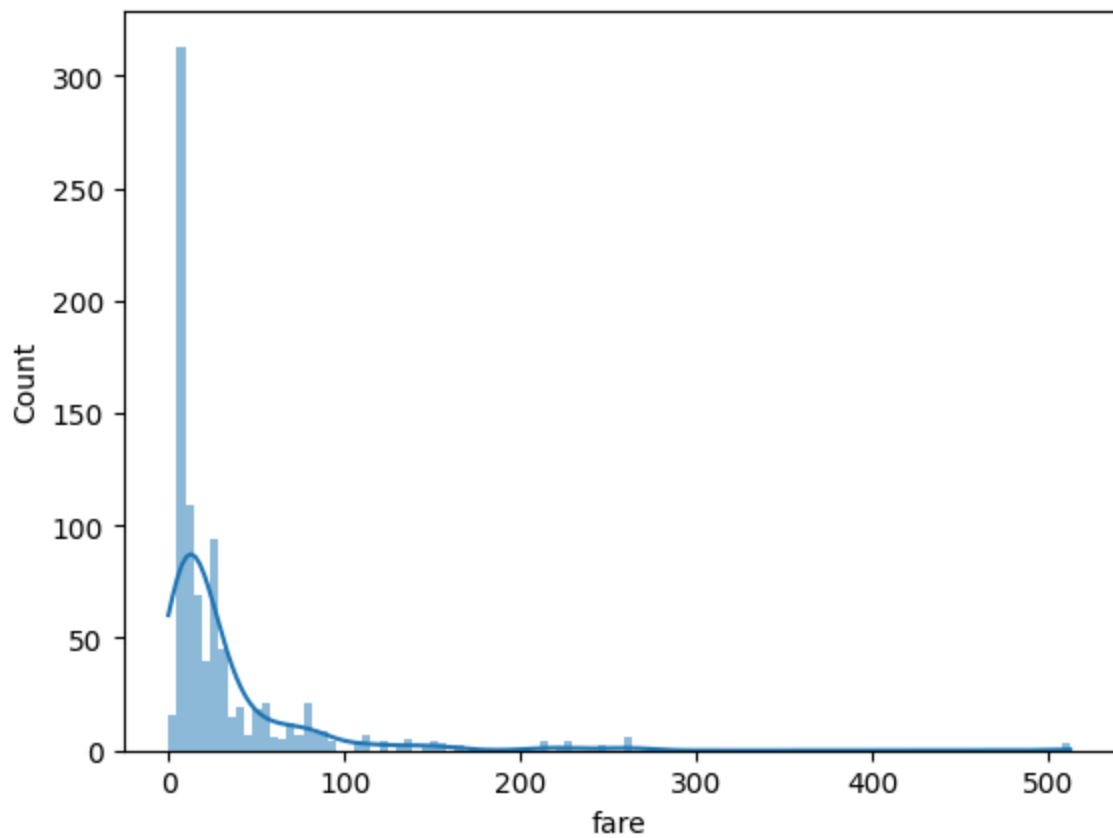
```
Out[7]: survived      0
pclass      0
sex         0
age         0
sibsp      0
parch      0
fare       0
embarked    0
class      0
who        0
adult_male  0
deck       0
embark_town 2
alive      0
alone      0
dtype: int64
```

```
In [8]: dataset['deck']
```

```
Out[8]: 0      A
1      C
2      A
3      C
4      A
..
886    A
887    B
888    A
889    C
890    A
Name: deck, Length: 891, dtype: category
Categories (7, object): ['A', 'B', 'C', 'D', 'E', 'F', 'G']
```

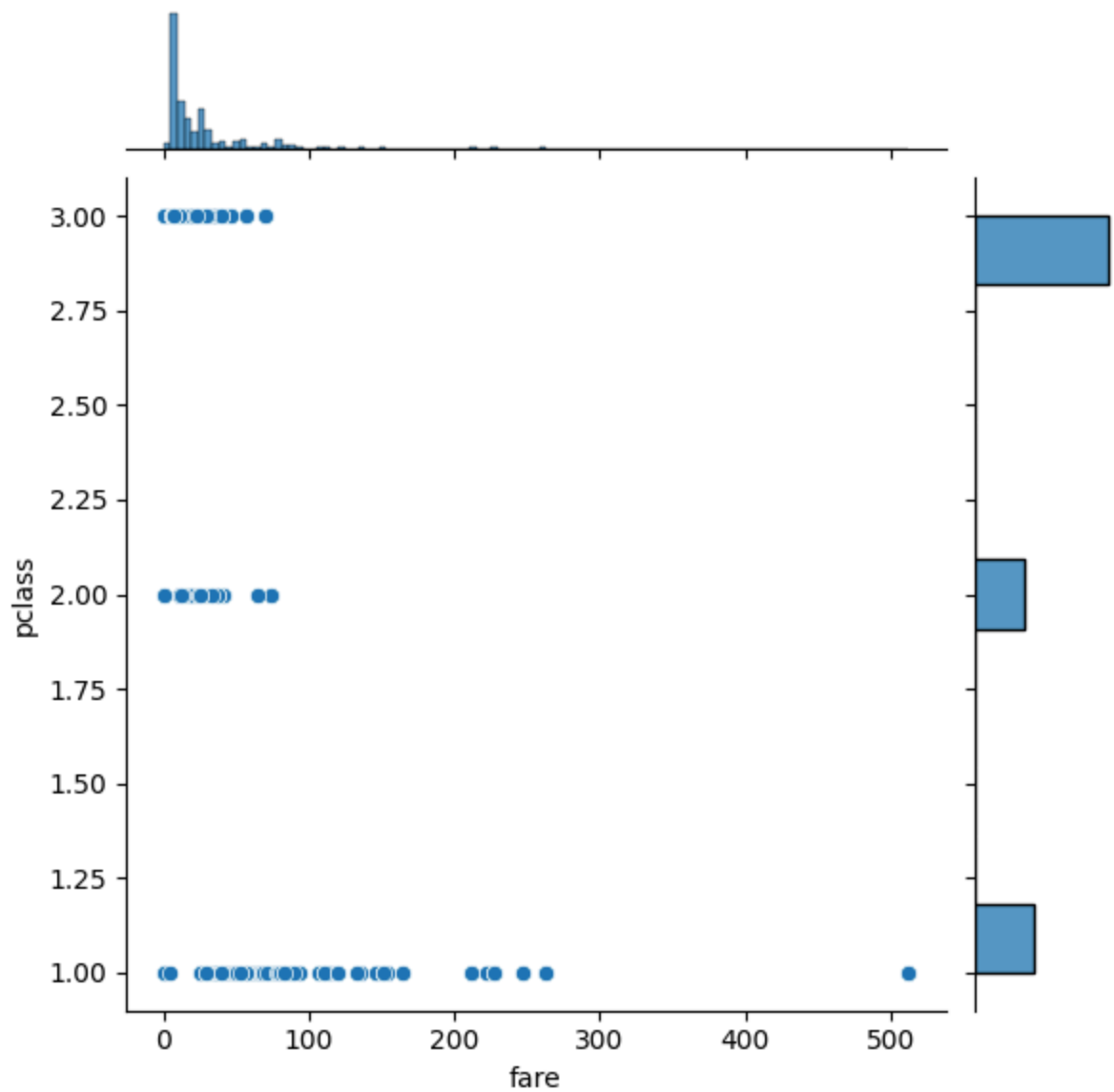
```
In [16]: sns.histplot(dataset['fare'], kde=True, linewidth=0)
```

```
Out[16]: <Axes: xlabel='fare', ylabel='Count'>
```



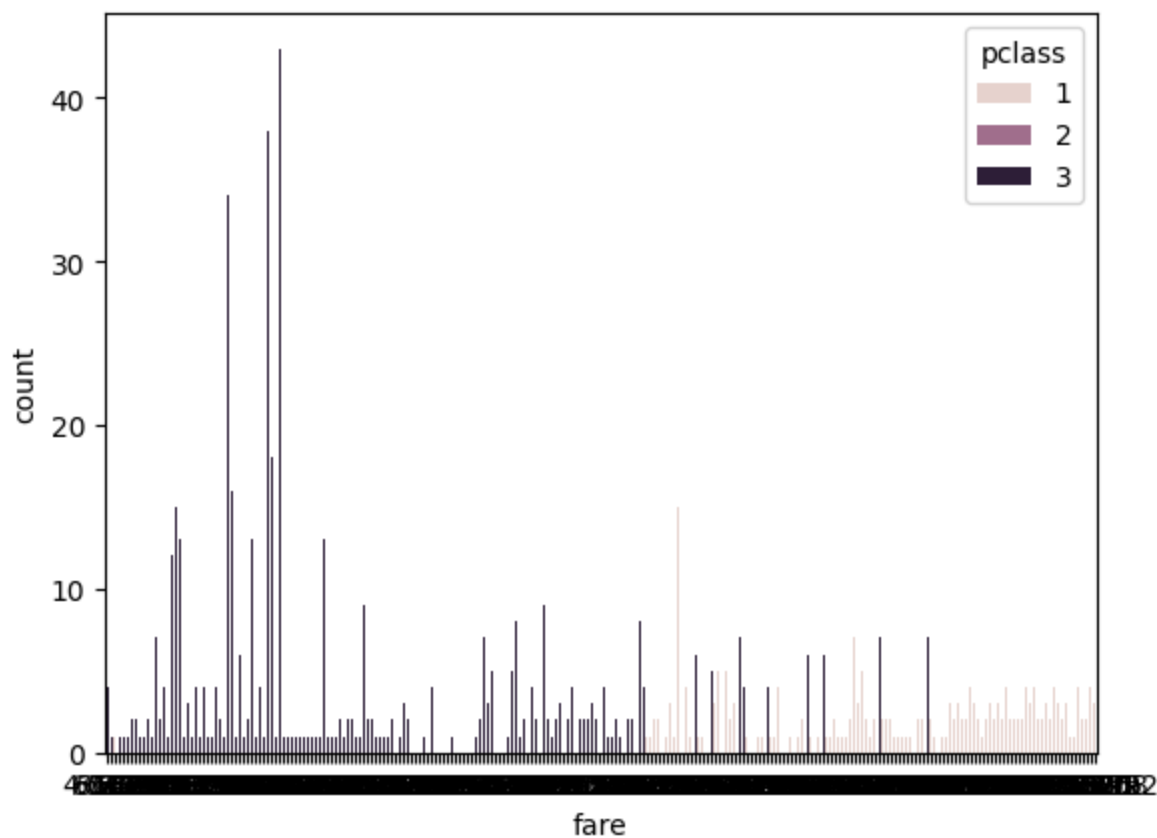
```
In [19]: sns.jointplot(x='fare', y='pclass', data=dataset)
```

```
Out[19]: <seaborn.axisgrid.JointGrid at 0x1a851b7be30>
```



```
In [20]: sns.countplot(x='fare', hue='pclass', data=dataset)
```

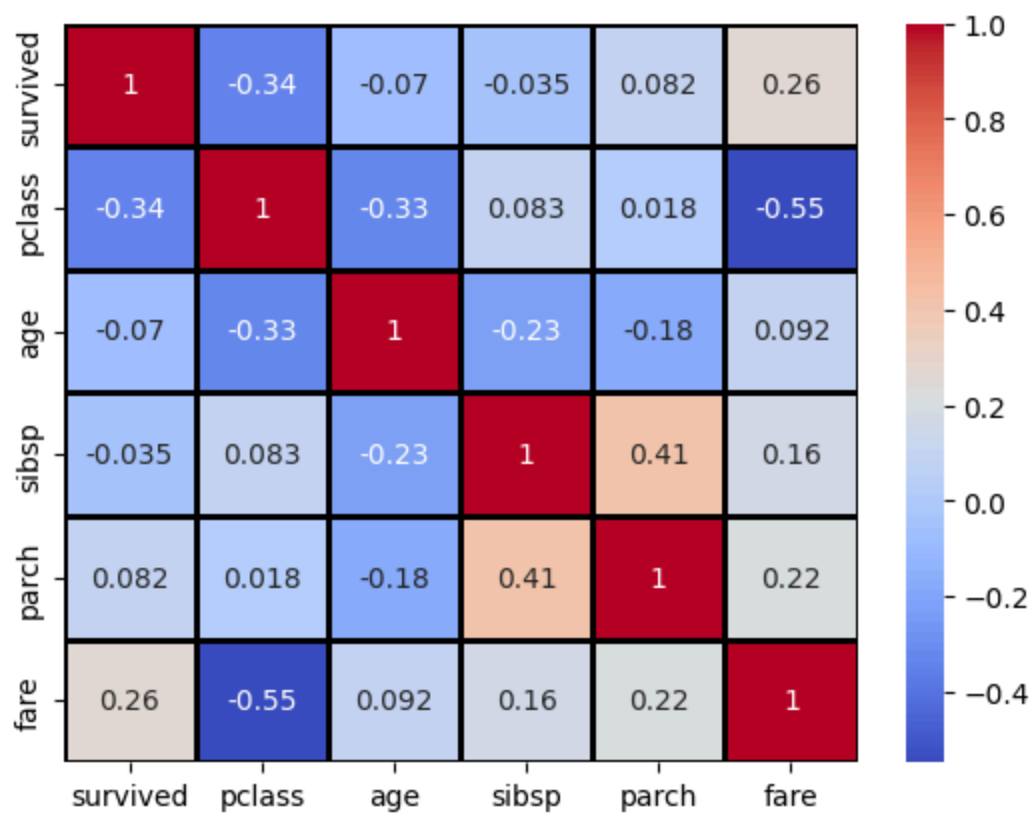
```
Out[20]: <Axes: xlabel='fare', ylabel='count'>
```



```
In [30]: numeric_dataset = dataset.select_dtypes(include=['number'])
```

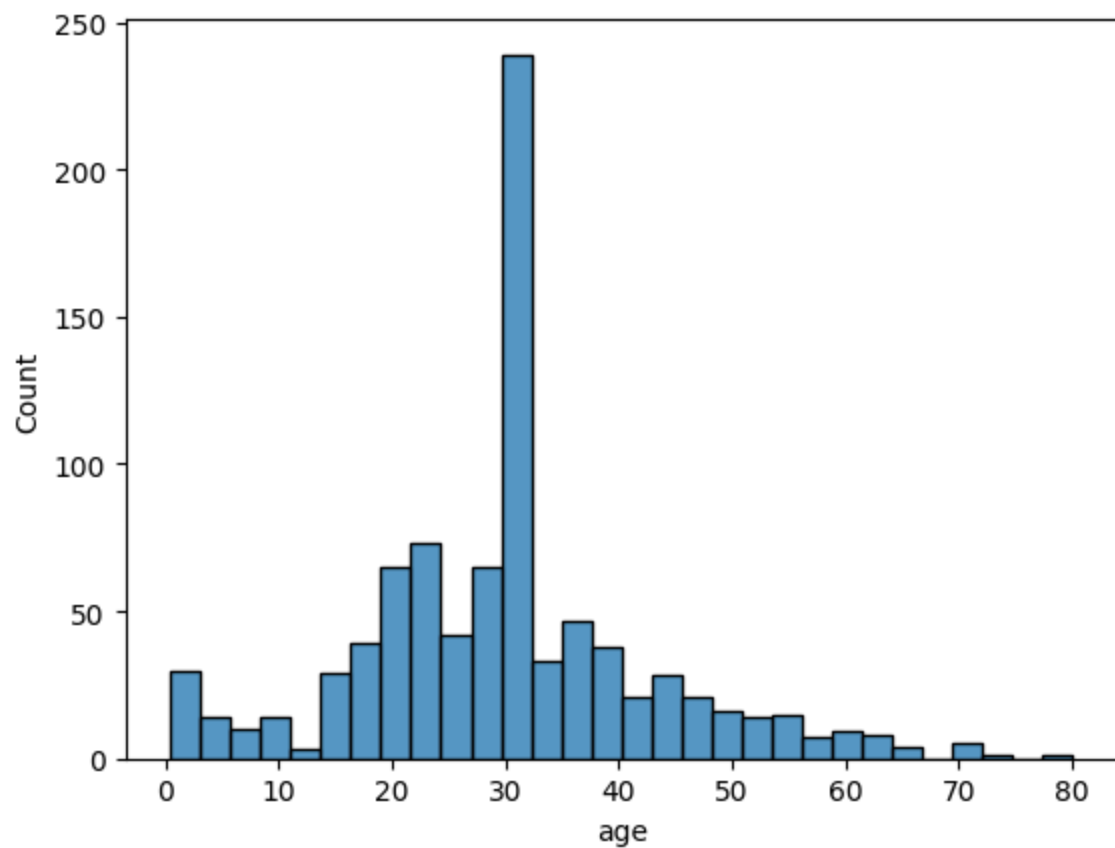
```
In [31]: corr_matrix = numeric_dataset.corr()
```

```
In [32]: sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', linewidths=1, linecolor='black',  
plt.show())
```



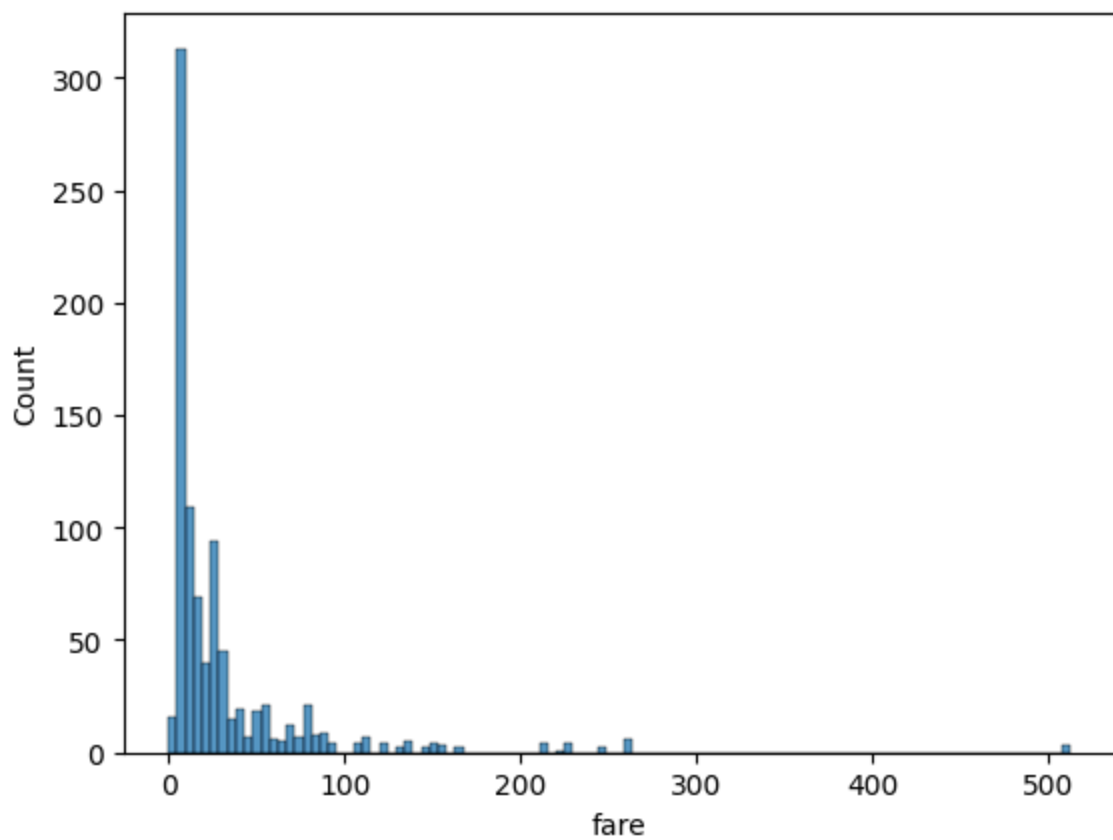
```
In [23]: sns.histplot(dataset['age'])
```

```
Out[23]: <Axes: xlabel='age', ylabel='Count'>
```



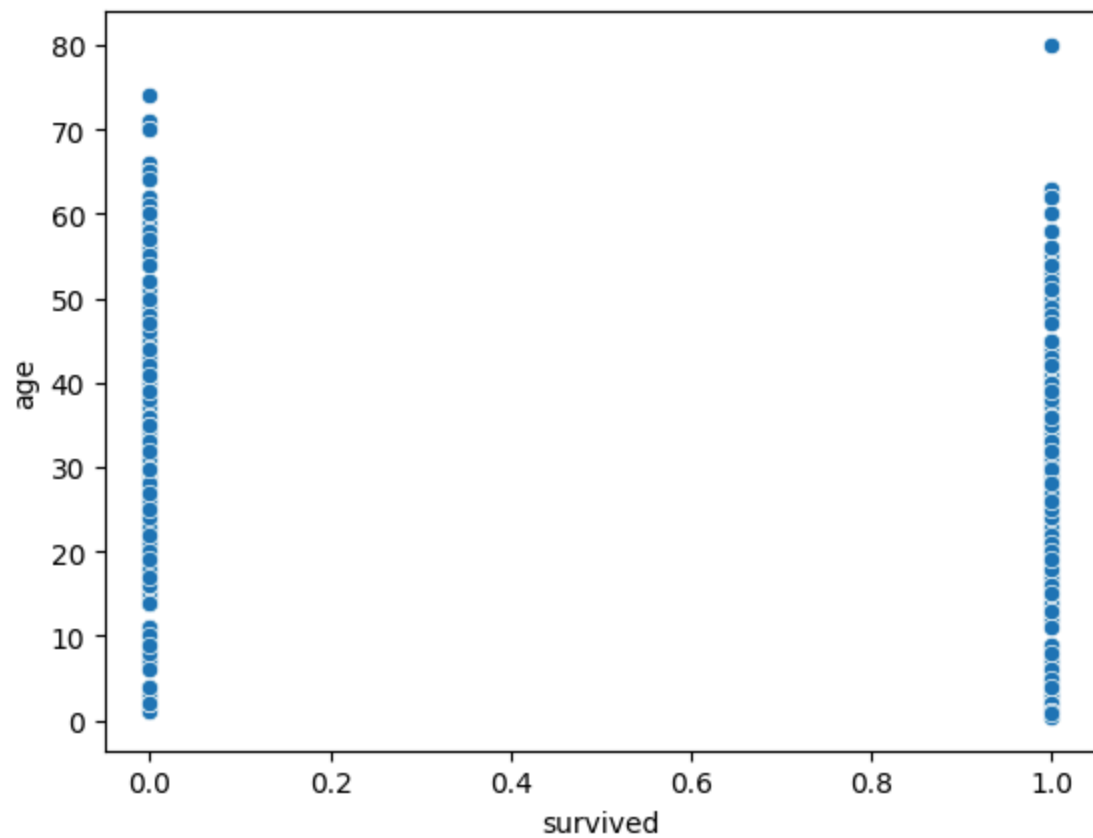

```
In [24]: sns.histplot(dataset['fare'])
```

```
Out[24]: <Axes: xlabel='fare', ylabel='Count'>
```



```
In [25]: sns.scatterplot(x='survived', y='age', data=dataset)
```

```
Out[25]: <Axes: xlabel='survived', ylabel='age'>
```



```
In [ ]:
```