titanic_8

May 11, 2022

0.0.1 Data Visualization I

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to see if we can find any patterns in the data.
2. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

[2]:
```
#import inbuilt Dataset and libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = sns.load_dataset('titanic')
df.head(10)
```

[2]:

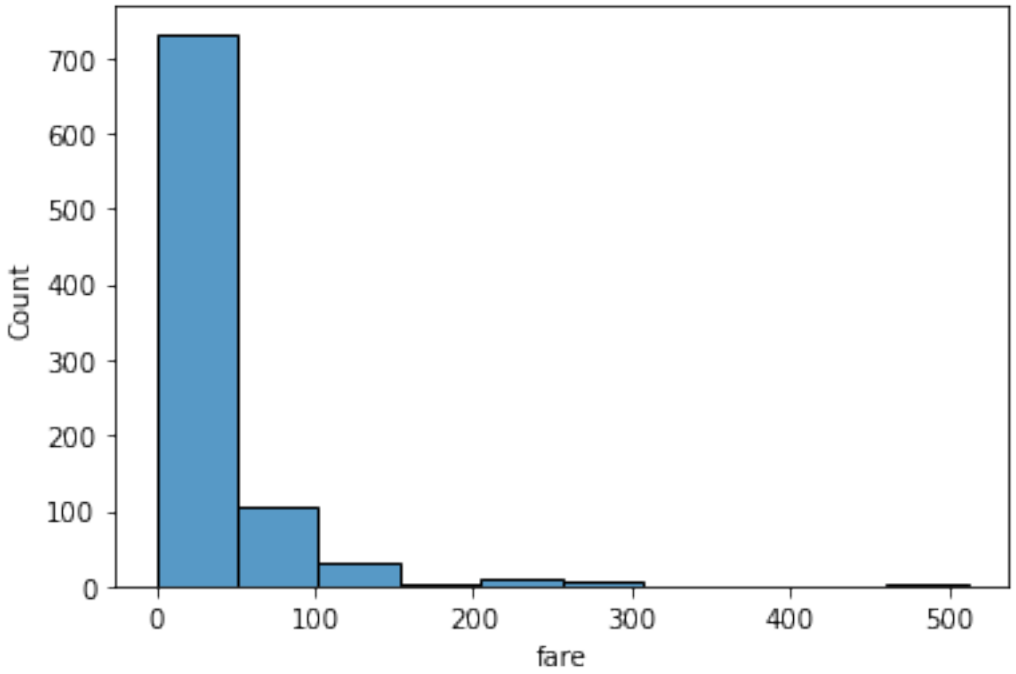
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	\
0	0	3	male	22.0	1	0	7.2500	S	Third	
1	1	1	female	38.0	1	0	71.2833	C	First	
2	1	3	female	26.0	0	0	7.9250	S	Third	
3	1	1	female	35.0	1	0	53.1000	S	First	
4	0	3	male	35.0	0	0	8.0500	S	Third	
5	0	3	male	NaN	0	0	8.4583	Q	Third	
6	0	1	male	54.0	0	0	51.8625	S	First	
7	0	3	male	2.0	3	1	21.0750	S	Third	
8	1	3	female	27.0	0	2	11.1333	S	Third	
9	1	2	female	14.0	1	0	30.0708	C	Second	

	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
5	man	True	NaN	Queenstown	no	True
6	man	True	E	Southampton	no	True
7	child	False	NaN	Southampton	no	False
8	woman	False	NaN	Southampton	yes	False

```
9  child       False  NaN    Cherbourg   yes  False
```

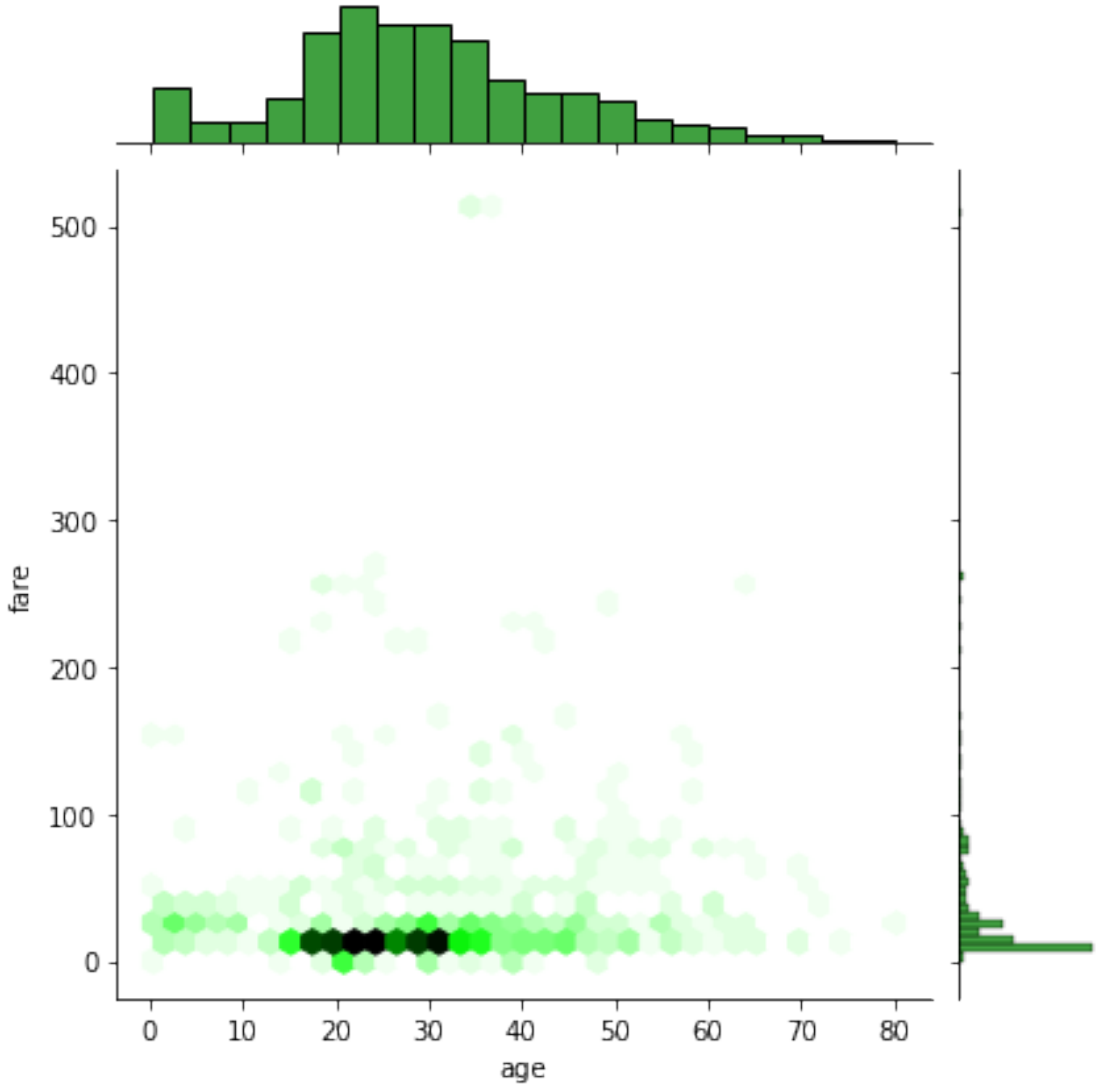
[6]:
```
sns.histplot(df["fare"],bins=10)
#from the plot we can say that more than 700 passengerss had a ticket in the␣
 ↪range 0-50 dollars
```

[6]:
```
<AxesSubplot:xlabel='fare', ylabel='Count'>
```



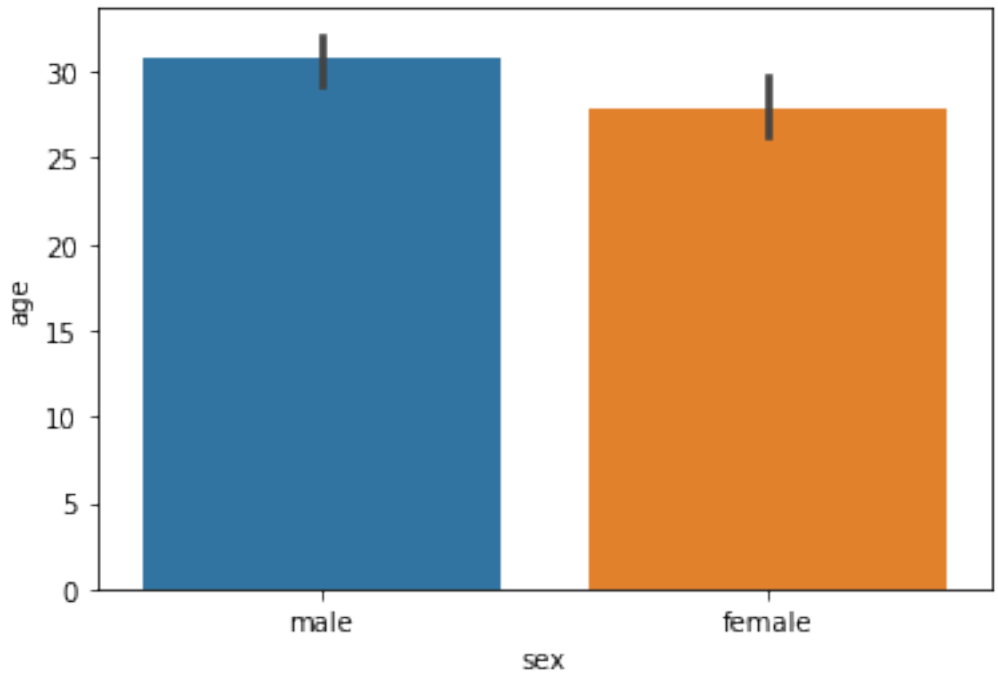
[25]:
```
sns.jointplot(x='age', y='fare', data=df,kind="hex" ,color="green")
#from this jointplot we can see that most of the passengers are  aged 20-30 and␣
 ↪they fare was under 50 dollars
```

[25]:
```
<seaborn.axisgrid.JointGrid at 0x7fa9c19e0550>
```



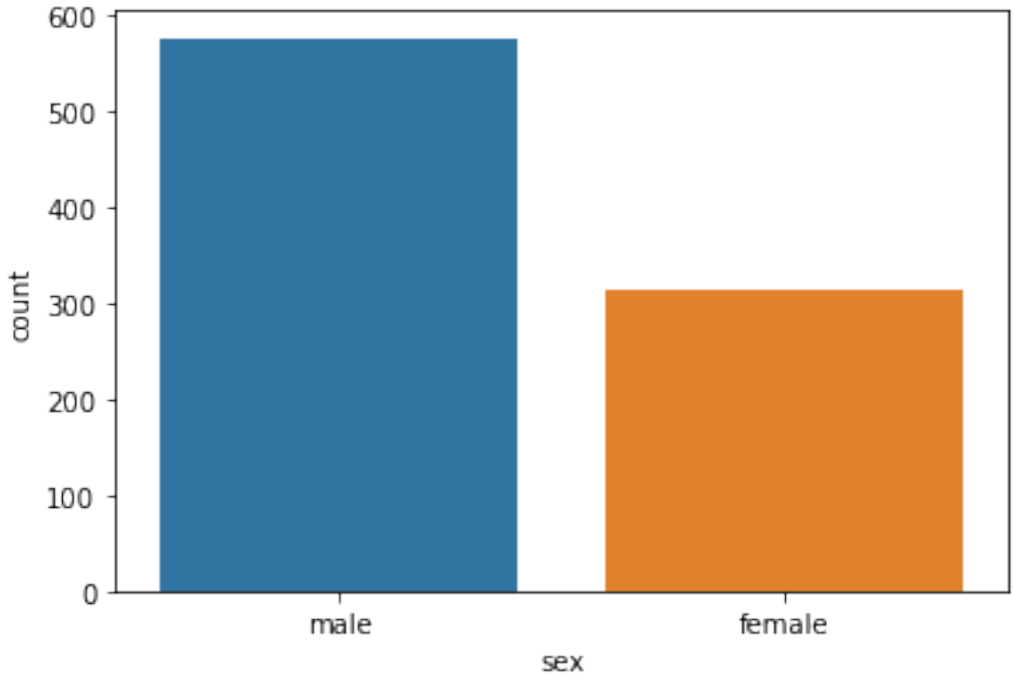
[26]:
```
sns.barplot(x='sex', y='age', data=dataset)
# average age of male passenger is greater than female
```

[26]: <AxesSubplot:xlabel='sex', ylabel='age'>



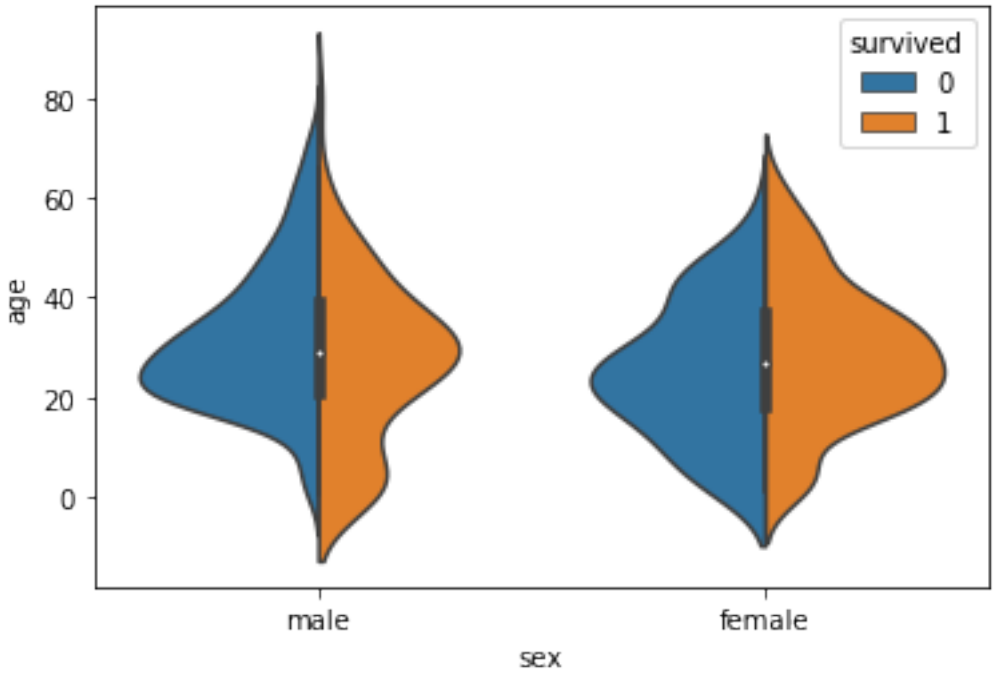
[27]:
```
sns.countplot(x='sex', data=dataset)
#there are more male boarding the titanic
```

[27]: <AxesSubplot:xlabel='sex', ylabel='count'>



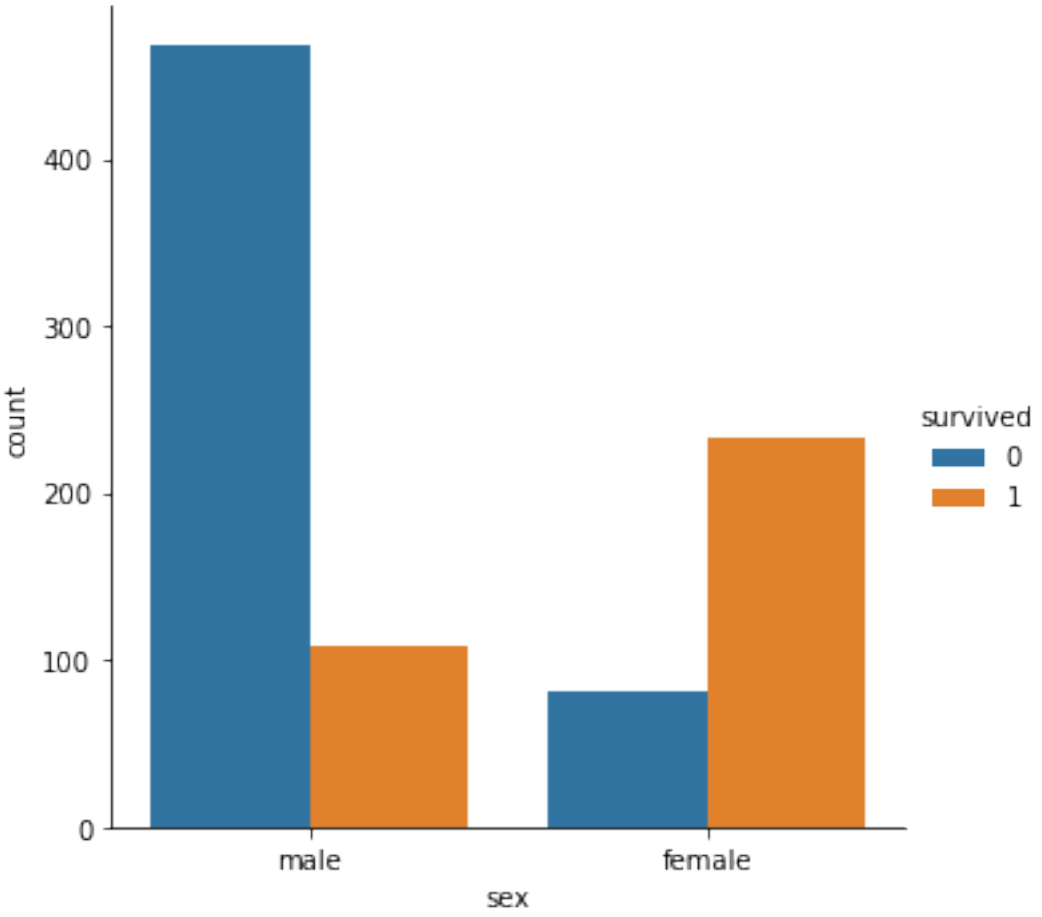
[28]:
```
sns.violinplot(x='sex', y='age', data=df, hue='survived', split=True)
#In case of male we can see that younger male survived more than the no of␣
 ↪young male who did not survived.
```

[28]: <AxesSubplot:xlabel='sex', ylabel='age'>



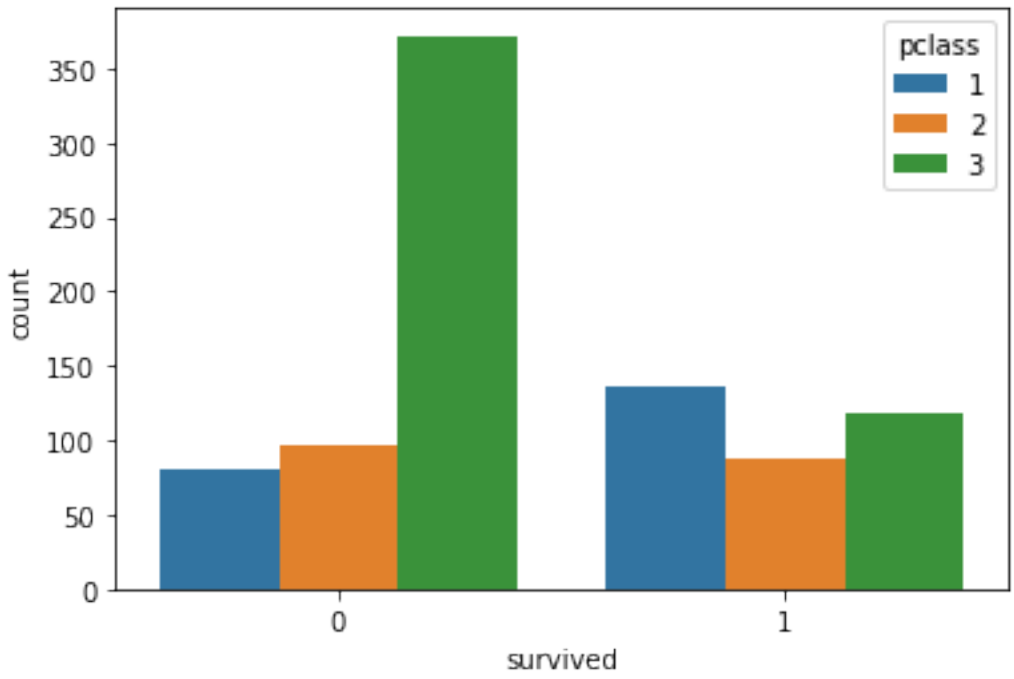
[30]:
```
sns.catplot(x ="sex", hue ="survived",
kind ="count", data = df)
```

[30]: <seaborn.axisgrid.FacetGrid at 0x7fa9d4e50e80>



[32]:
```
sns.countplot(x="survived",hue="pclass",data=df)
#class 1 people survived more than the other
```

[32]: <AxesSubplot:xlabel='survived', ylabel='count'>


pclass
1
2
3
350
300
250
200
150
100
50
0
count
0
1
survived