

practical-exam-11-12

May 23, 2023

```
[1]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

1 Problem Statement 11 and 12

Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship.

11. Use the Seaborn library to see if we can find any patterns in the data. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.
12. Plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not. (Column names: 'sex' and 'age')

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

```
[24]: df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/exam_datasets/11-12.
↳titanic.csv')
```

```
[25]: 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
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), object(5)
memory usage: 83.7+ KB
```

```
[27]: df.isnull().sum()
```

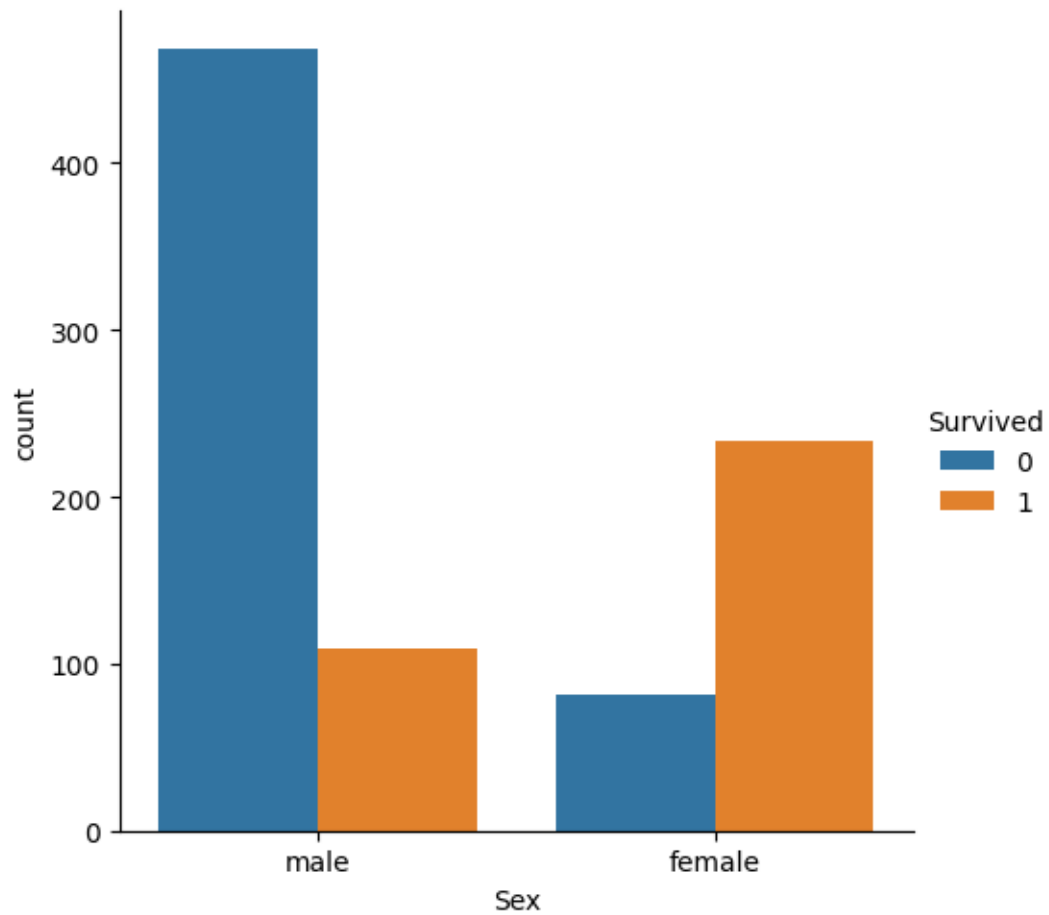
```
[27]: PassengerId      0
      Survived        0
      Pclass         0
      Name           0
      Sex            0
      Age          177
      SibSp          0
      Parch          0
      Ticket         0
      Fare           0
      Cabin         687
      Embarked       2
      dtype: int64
```

```
[28]: import seaborn as sns
      import matplotlib.pyplot as plt
```

```
[32]: # Survival VS Gender

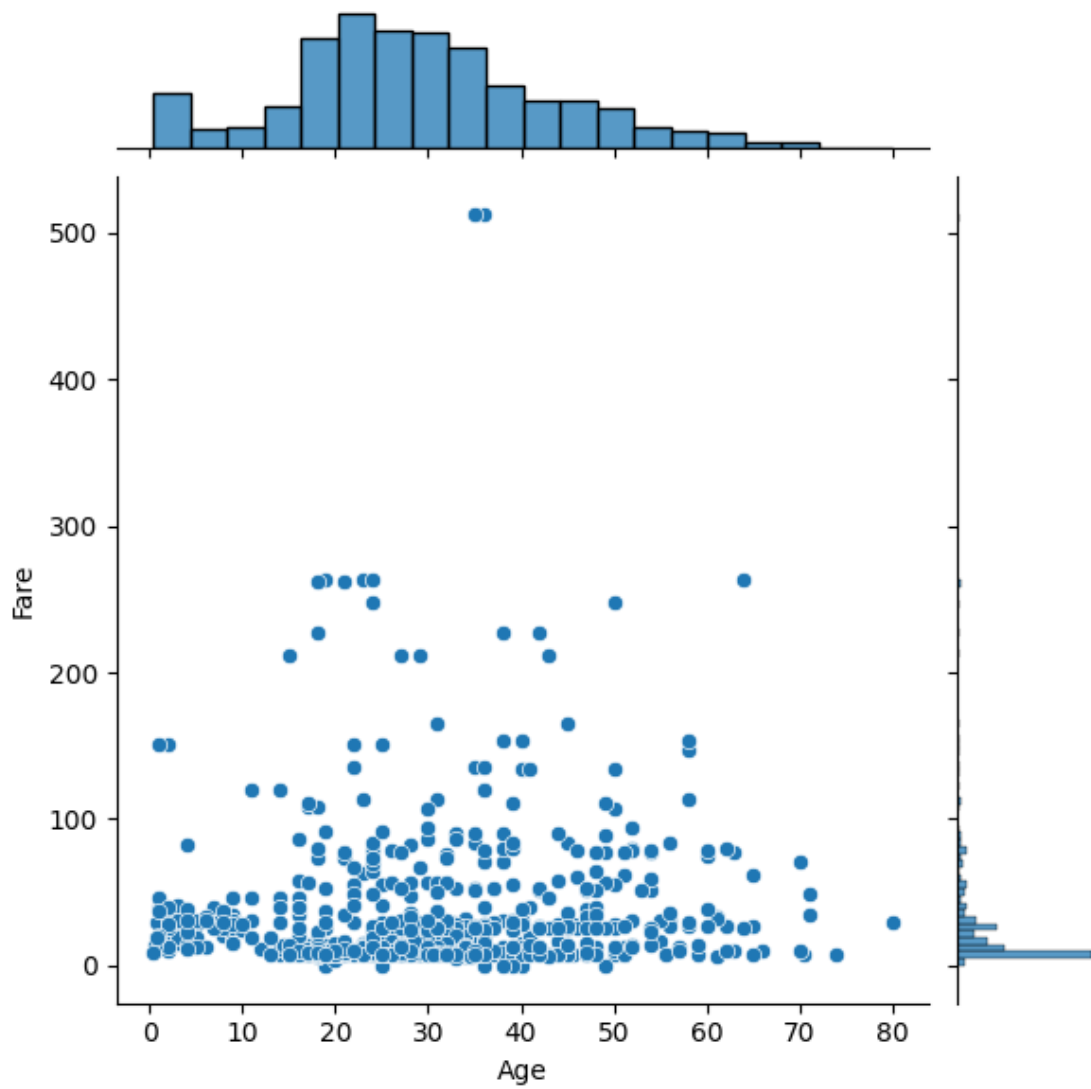
      sns.catplot(data=df, x='Sex', hue='Survived', kind='count')
```

```
[32]: <seaborn.axisgrid.FacetGrid at 0x7fd370e98040>
```



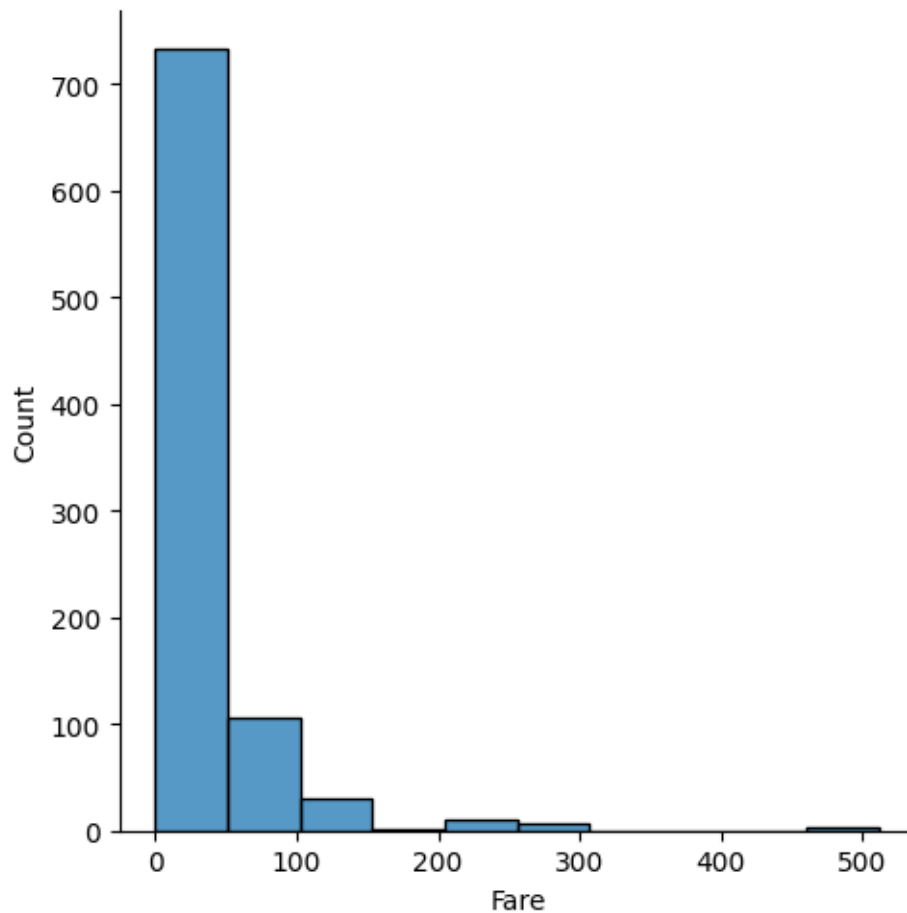
```
[34]: sns.jointplot(x='Age', y='Fare', data=df)
```

```
[34]: <seaborn.axisgrid.JointGrid at 0x7fd36c9bcbe0>
```



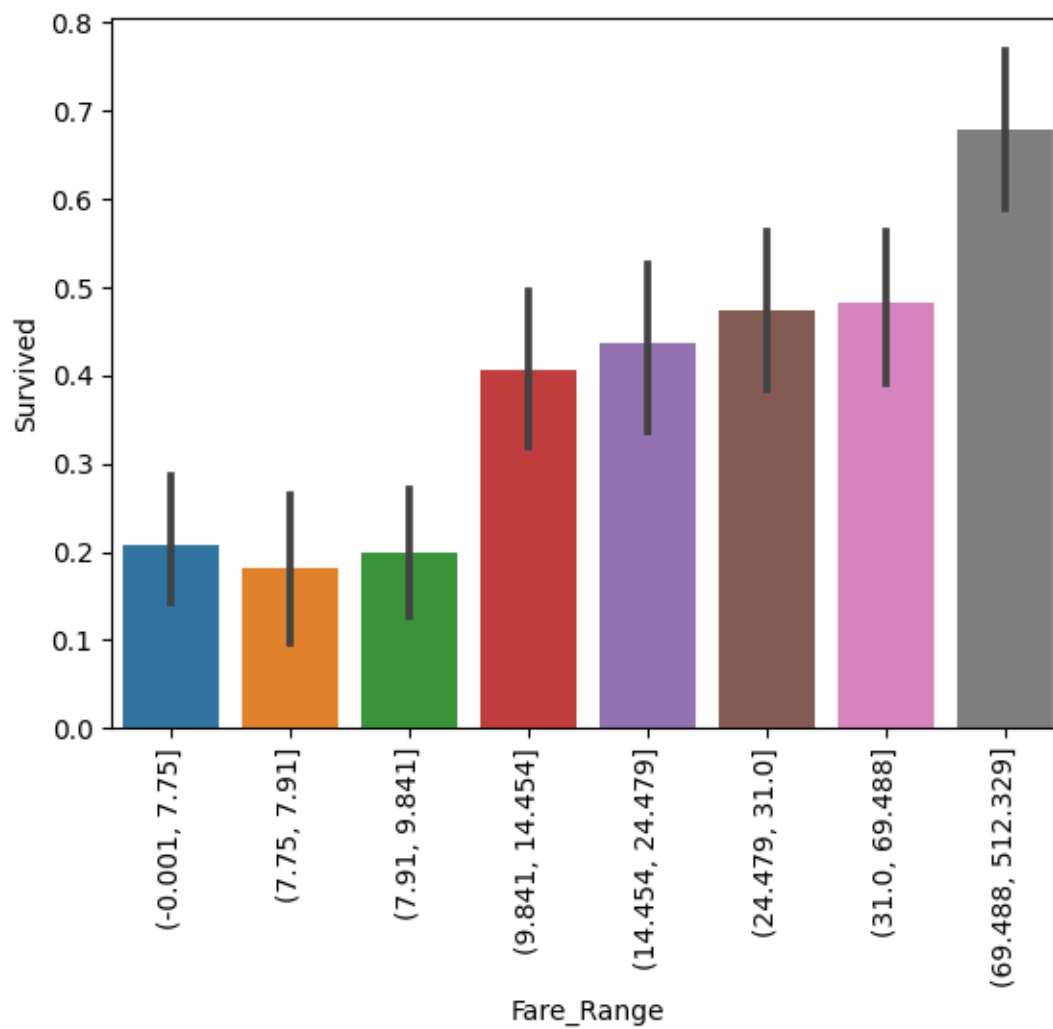
```
[38]: sns.displot(df['Fare'], kde=False, bins=10)
```

```
[38]: <seaborn.axisgrid.FacetGrid at 0x7fd367fefb20>
```



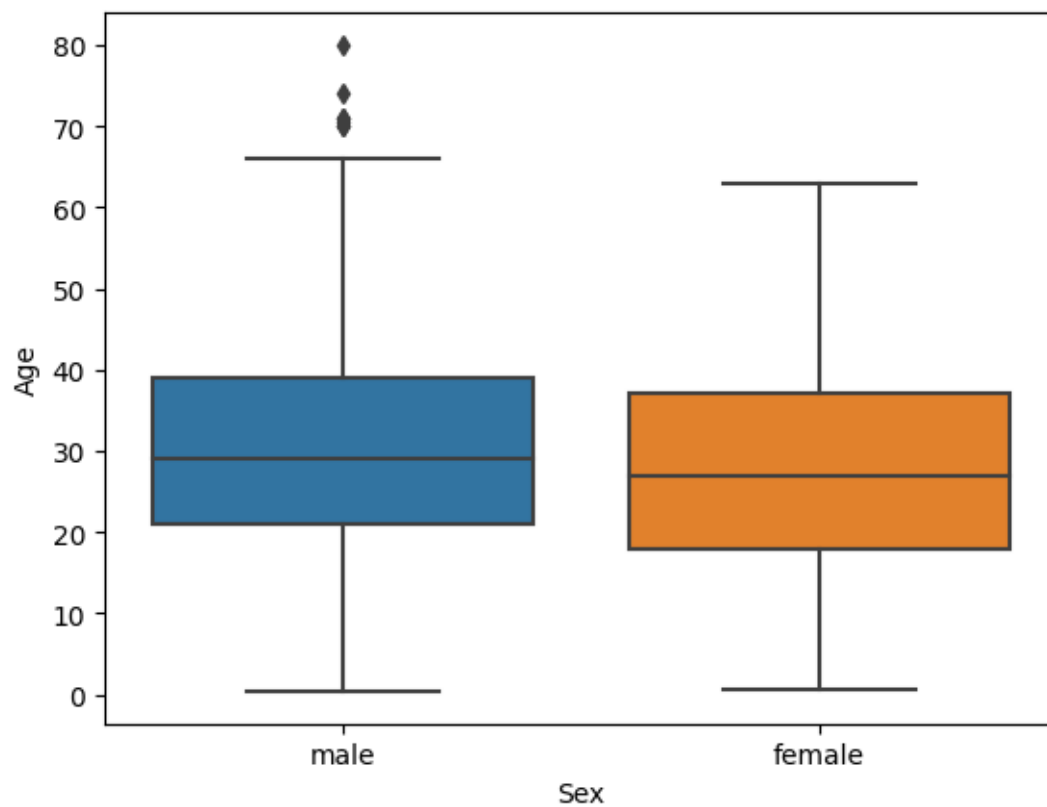
```
[42]: df['Fare_Range'] = pd.qcut(df['Fare'], 8, )
sns.barplot(x='Fare_Range', y='Survived', data=df)
plt.xticks(rotation=90)
```

```
[42]: (array([0, 1, 2, 3, 4, 5, 6, 7]),
      [Text(0, 0, '(-0.001, 7.75]'),
       Text(1, 0, '(7.75, 7.91]'),
       Text(2, 0, '(7.91, 9.841]'),
       Text(3, 0, '(9.841, 14.454]'),
       Text(4, 0, '(14.454, 24.479]'),
       Text(5, 0, '(24.479, 31.0]'),
       Text(6, 0, '(31.0, 69.488]'),
       Text(7, 0, '(69.488, 512.329]')])
```



```
[43]: sns.boxplot(x='Sex', y='Age', data=df)
```

```
[43]: <Axes: xlabel='Sex', ylabel='Age'>
```



```
[44]: sns.boxplot(x='Sex', y='Age', data=df, hue='Survived')
```

```
[44]: <Axes: xlabel='Sex', ylabel='Age'>
```

