DHRUV GARG 16 BCE 1190

DATA VISUALIZATION

LAB 5

Aim: To make visualizations on the Titanic dataset using Python.

Importing libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import csv
```

Reading the .csv file and getting the dimensions

```
df = pd.read_csv('titanic_data.csv')
```

```
df.shape
```

Output : (891, 12)

```
df.head()
```

Output:

	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

list(df)

Output:

['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked']

Created a function, which given two parameters, creates a pivot table and a bar chart to show their relationship

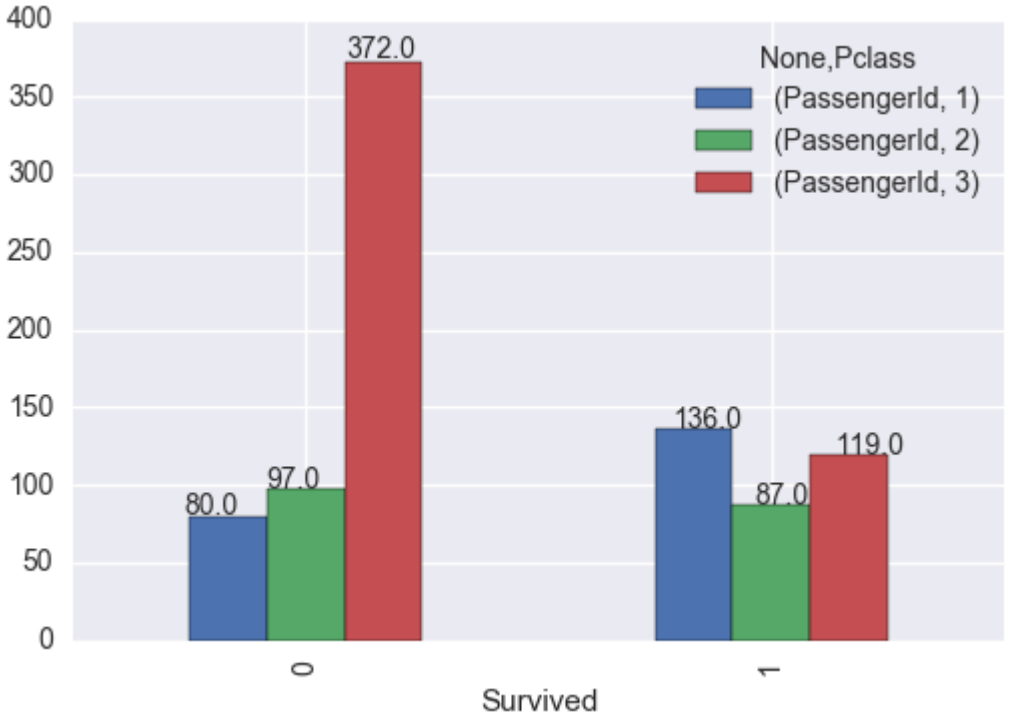
```
def make_pivot (param1, param2):
    df_slice = df[[param1, param2, 'PassengerId']]
    slice_pivot = df_slice.pivot_table(index=[param1], columns=[param2],aggfunc=np.size, fill_value=0)

    p_chart = slice_pivot.plot.bar()
    for p in p_chart.patches:
        p_chart.annotate(str(p.get_height()), (p.get_x() * 1.05, p.get_height() * 1.01))

    return slice_pivot
    return p_chart
```

1. **Relation between passengers' survival and booking class**

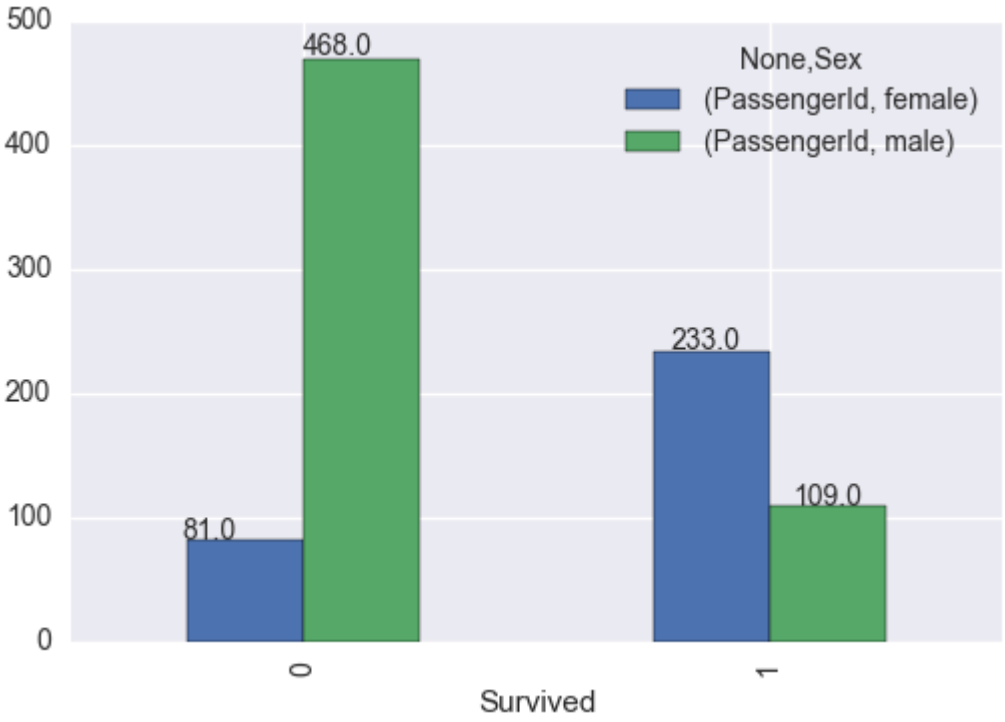
   make_pivot ('Survived','Pclass')



	PassengerId		
Pclass	**1**	**2**	**3**
Survived			
0	80	97	372
1	136	87	119

2. Relation between passengers' survival and their sex

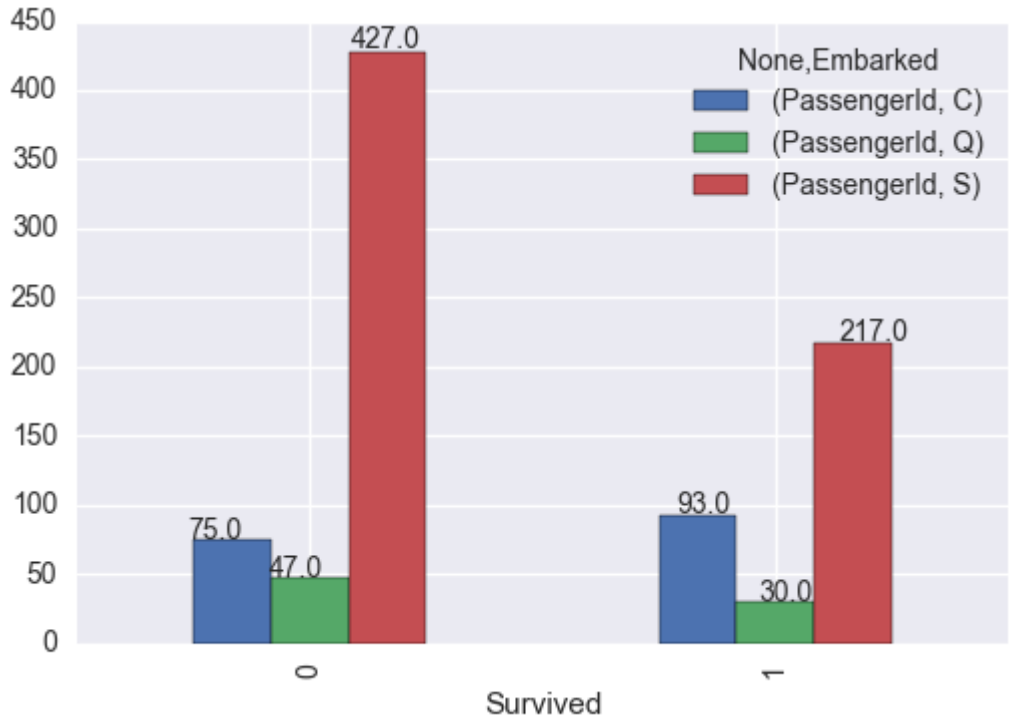
make_pivot ('Survived','Sex')



	PassengerId	
Sex	**female**	**male**
Survived		
0	81	468
1	233	109

3. Relation between passengers' survival and port of embarkation

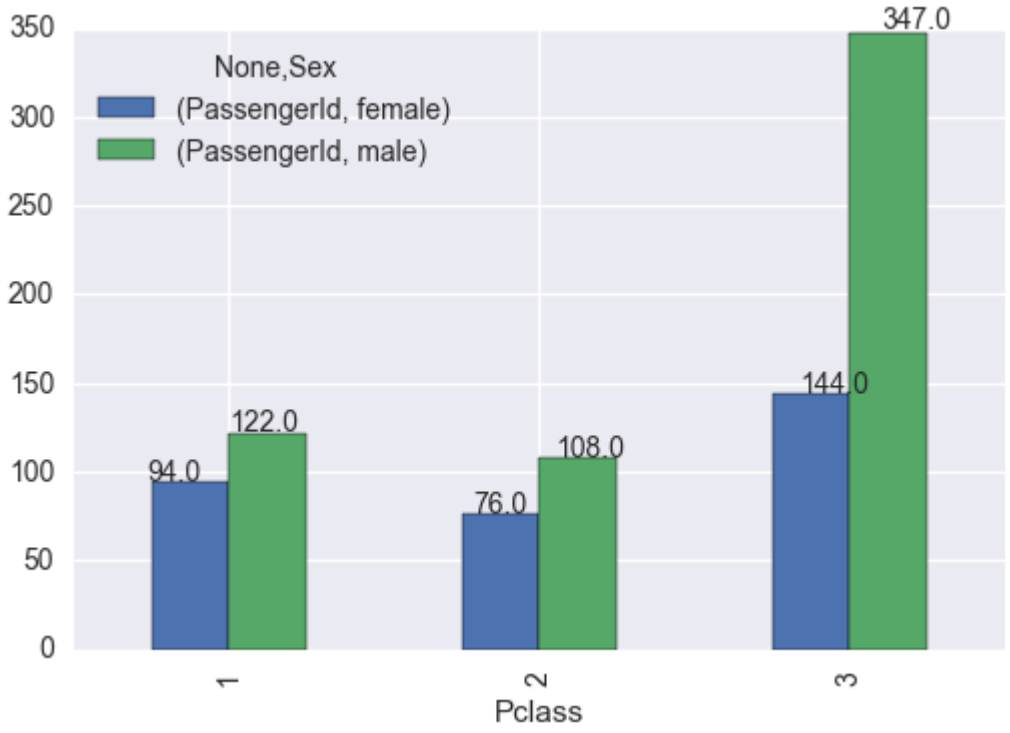
make_pivot ('Survived','Embarked')



	PassengerId		
Embarked	**C**	**Q**	**S**
Survived			
0	75	47	427
1	93	30	217

4. Relation between passengers' booking class and their sex

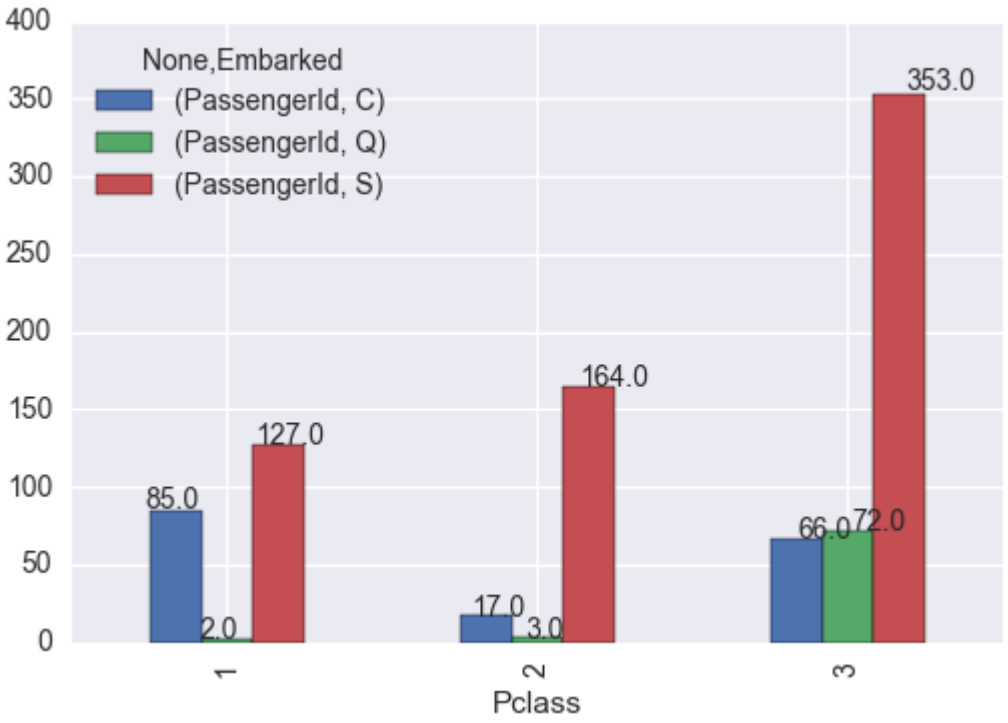
make_pivot ('Pclass','Sex')



	PassengerId	
Sex	**female**	**male**
Pclass		
1	94	122
2	76	108
3	144	347

5. Relation between passengers' booking class and port of embarkation

make_pivot ('Pclass','Embarked')



	PassengerId		
Embarked	**C**	**Q**	**S**
Pclass			
1	85	2	127
2	17	3	164
3	66	72	353