

pr9

May 4, 2024

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

```
[3]: df=pd.read_csv("C:\\Users\\nayan\\Downloads\\Titanic-Dataset.csv")
df
```

```
[3]:
```

	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]

```
[4]: df.head()
```

```
[4]: PassengerId  Survived  Pclass  \
0             1         0         3
1             2         1         1
2             3         1         3
3             4         1         1
4             5         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	

	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

```
[5]: df.describe()
```

```
[5]: PassengerId  Survived  Pclass  Age  SibSp  \
count  891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000
```

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

```
[6]: df.shape
```

```
[6]: (891, 12)
```

```
[7]: df.size
```

```
[7]: 10692
```

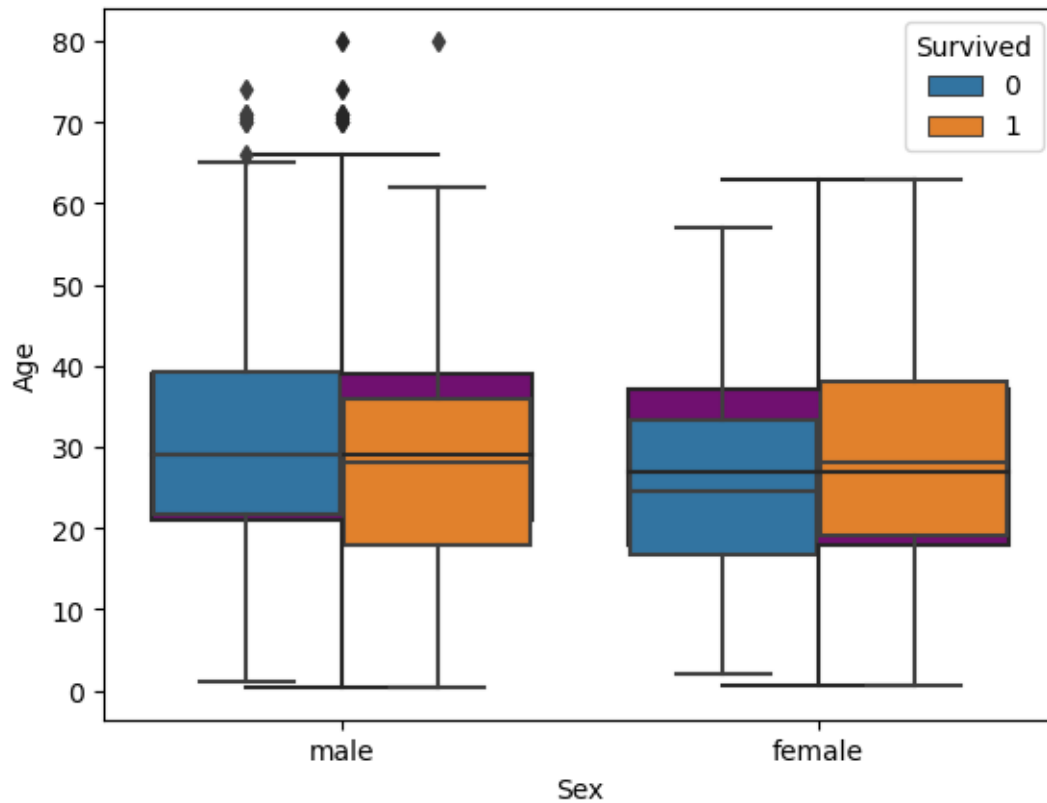
```
[9]: df.columns
```

```
[9]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
         'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
         dtype='object')
```

```
[13]: df_cleaned=df.drop(['Pclass', 'Embarked', 'Fare'],axis=1)
```

```
[15]: sns.boxplot(data=df,x="Sex",y="Age", color="Purple")
      sns.boxplot(data=df,x="Sex",y="Age", hue="Survived")
```

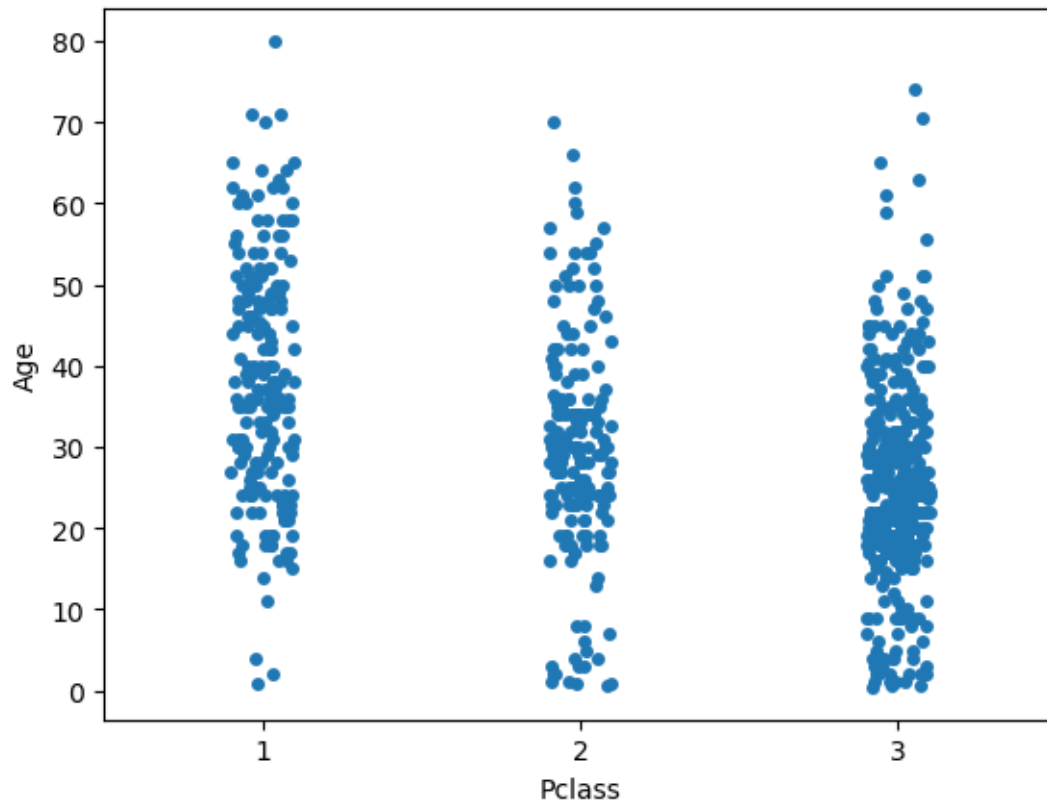
```
[15]: <Axes: xlabel='Sex', ylabel='Age'>
```



```
[19]: #It's a plot between a continuous variable and a categorical variable. It plots
      ↪ as a scatter plot but supplementarily uses
      #categorical encodings of the categorical variable. The stripplot() function is
      ↪ used to plot the violin plot. Like the box plot,
      #the first parameter is the categorical column, the second parameter is the
      ↪ numeric column while the third parameter is the dataset.

      sns.stripplot(y=df['Age'], x=df['Pclass'])
```

```
[19]: <Axes: xlabel='Pclass', ylabel='Age'>
```



```
[20]: #It is the combination of a strip plot and a violinplot.
```

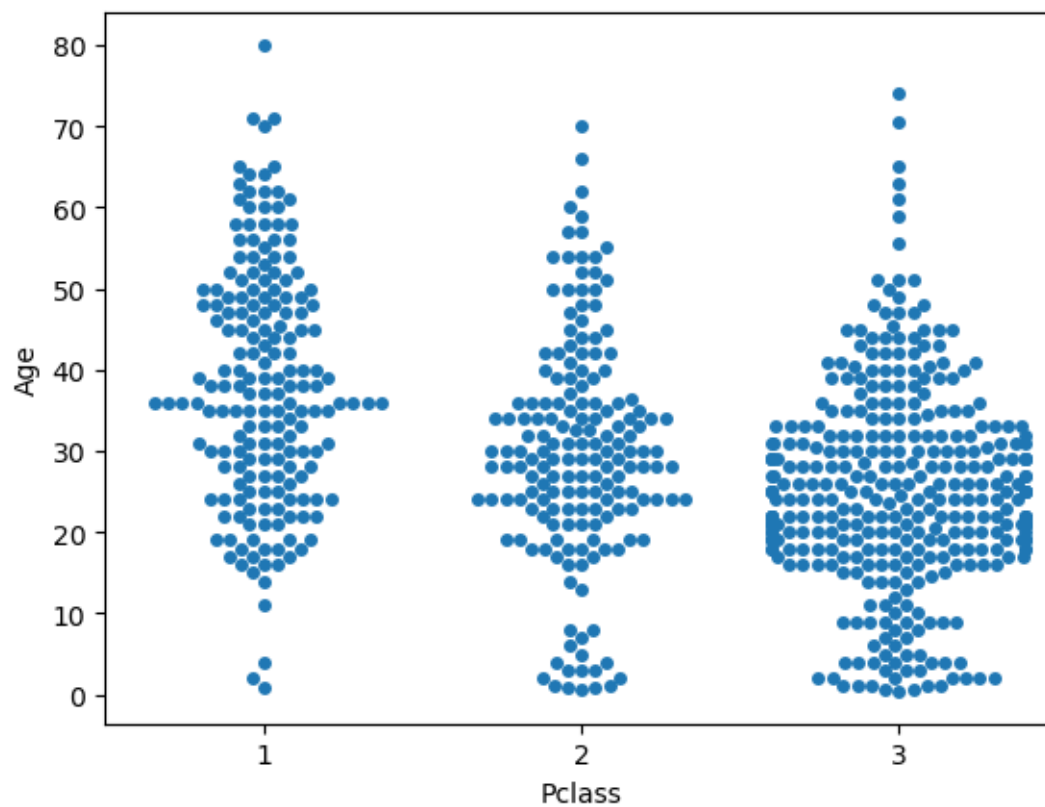
```
sns.swarmplot(y=df['Age'],x=df['Pclass'])
```

```
[20]: <Axes: xlabel='Pclass', ylabel='Age'>
```

C:\Users\nayan\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:

UserWarning: 15.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.

```
warnings.warn(msg, UserWarning)
```



[ ]:

[ ]: