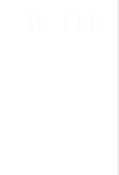
**-------------------------------------------------------------------------------------------------------------------------------------------**

**1 ) Use the inbuilt dataset 'titanic' as used in the above problem. 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')**

**Write observations on the inference from the above statistics.**

**-------------------------------------------------------------------------------------------------------------------------------------------**



**import** pandas **as** pd

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** warnings warnings**.**filterwarnings("ignore")

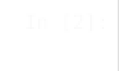
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **survived** | **pclass** | **sex** | **age** | **sibsp** | **parch** | **fare** | **embarked** | **class** | **who** | **adult\_male** | **deck** | **embark\_town** | **alive** | **alone** |
| **0** 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| **1** 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| **2** 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| **3** 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| **4** 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |



dataset**.**info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 891 entries, 0 to 890 Data columns (total 15 columns):

# Column Non-Null Count Dtype

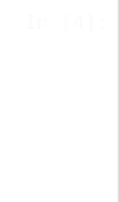


dataset **=** sns**.**load\_dataset('titanic') dataset**.**head()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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

1. **Use the inbuilt dataset 'titanic' as used in the above problem. 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')**

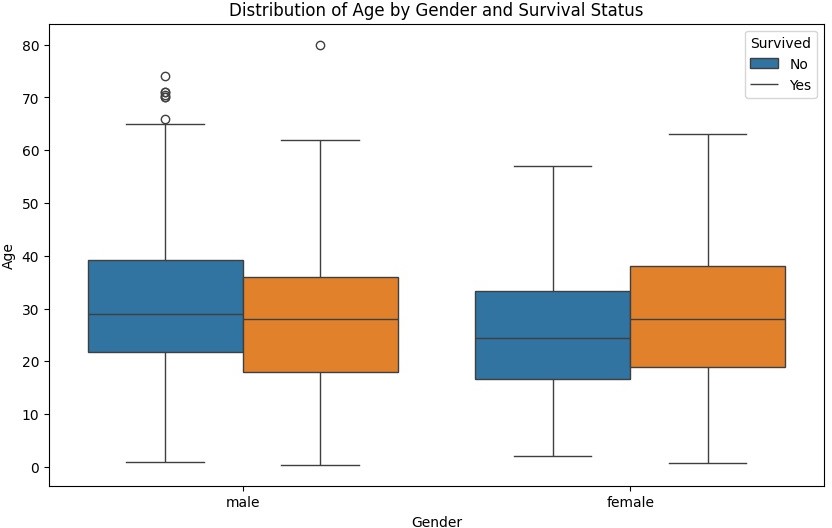


plt**.**figure(figsize**=**(10, 6))

sns**.**boxplot(x**=**'sex', y**=**'age', hue**=**'survived', data**=**dataset) plt**.**title('Distribution of Age by Gender and Survival Status') plt**.**xlabel('Gender')

plt**.**ylabel('Age')

plt**.**legend(title**=**'Survived', loc**=**'upper right', labels**=**['No', 'Yes']) plt**.**show()



1. **Write observations on the inference from the above statistics.**

If we want to see the box plots of forage of passengers of both genders, along with the information about whether or not they survived, we can pass the survived as value to the hue parameter.

We can also see the distribution of the passengers who survived. For instance, we can see that among the male passengers, on average more younger people survived as compared to the older ones. Similarly, we can see that the variation among the age of female passengers who did not survive is much greater than the age of the surviving female passengers.