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

In [3]:
```
titanic=pd.read_csv('titanic_train.csv')
```

In [5]:
```
titanic.head(2)
```

Out[5]:

	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

In [8]:
```
sns.countplot(x='Survived' , data=titanic)

#more then 500 people died adn around 300 survived
```

Out[8]:
```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```


In [9]:
```
sns.countplot(x='Survived' , hue='Sex', data=titanic)

#more male has died as compare to female and more female survived also
```

Out[9]:
```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

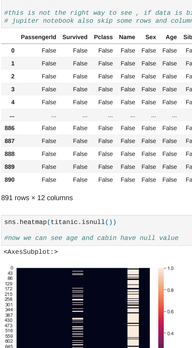

In [11]:
```
sns.countplot(x='Survived' , hue='Pclass', data=titanic)
sns.set_style('whitegrid')

#we can pclass 3 max people died
```

In [ ]:
```
#from both below graph we can see max people where of age btwn 20-40
```

In [12]:
```
sns.distplot(titanic['Age'])
```

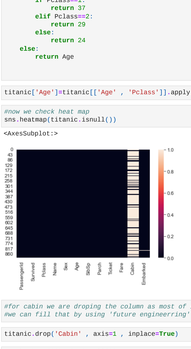
```
C:\Users\india\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[12]:
```
<AxesSubplot:xlabel='Age', ylabel='Density'>
```


In [15]:
```
plt.hist(titanic['Age'])
plt.show()
```

In [16]:
```
sns.countplot(x='SibSp' , data=titanic)

#sibsp -sibling or spouse , max people were single in ship
```

Out[16]:
```
<AxesSubplot:xlabel='SibSp', ylabel='count'>
```


In [ ]:
```
###removing null values
```

In [38]:
```
titanic.isnull()

#this is not the right way to see , if data is big we can not see everything
# jupiter notebook also skip some rows and column and
```

Out[38]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False	False	False	True	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	False	False	False	False	False	True	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True	False

891 rows × 12 columns

In [18]:
```
sns.heatmap(titanic.isnull())

#now we can see age and cabin have null value
```

Out[18]:
```
<AxesSubplot:>
```


In [ ]:
```
#first we will solve for age
#from data we can see there is relation btwn age and pclass
#we try to see that from box pot
```

In [21]:
```
sns.boxplot( x='Pclass',y='Age'  , data=titanic)

#from here we get avg age for pclass 1 is 37
#avg age for pclass 2 is 29 and for pclass 3 is 24
```

Out[21]:
```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```


In [35]:
```
#creating func to fill Age null value

def inpute_age(cols):
    Age=cols[0]
    Pclass=cols[1]

    if pd.isnull(Age):

        if Pclass==1:
            return 37
        elif Pclass==2:
            return 29
        else:
            return 24
    else:
        return Age
```

In [36]:
```
titanic['Age']=titanic[['Age' , 'Pclass']].apply(inpute_age , axis=1)
```

In [37]:
```
#now we check heat map
sns.heatmap(titanic.isnull())
```

Out[37]:
```
<AxesSubplot:>
```


In [ ]:
```
#for cabin we are droping the column as most of it parameter has null value
#we can fill that by using 'future engineerring' but it will take lot of time
```

In [41]:
```
titanic.drop('Cabin' , axis=1 , inplace=True)
```

In [42]:
```
sns.heatmap(titanic.isnull())
```

Out[42]:
```
<AxesSubplot:>
```


In [ ]:
```
#now we can see we have clear null values
```

In [ ]:
```
```