Data Science Internship Assignment

#

Task1: TITANIC SURVIVAL PREDICTION

In [39]:
```
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import MinMaxScaler
from sklearn import preprocessing
```

In [4]:
```
df= pd.read_csv('tested.csv')
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          332 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         417 non-null    float64
 10  Cabin        91 non-null     object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

In [7]:
```
df.head()
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

In [8]:
```
df.tail()
```

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
413	1305	0	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	
416	1308	0	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	
417	1309	0	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	

In [9]:
```
df.describe()
```

Out[9]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	418.000000	332.000000	418.000000	418.000000	417.000000
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	0.392344	35.627188
std	120.810458	0.481622	0.841838	14.181209	0.896760	0.981429	55.907576
min	892.000000	0.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	996.250000	0.000000	1.000000	21.000000	0.000000	0.000000	7.895800
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	0.000000	14.454200
75%	1204.750000	1.000000	3.000000	39.000000	1.000000	0.000000	31.500000
max	1309.000000	1.000000	3.000000	76.000000	8.000000	9.000000	512.329200

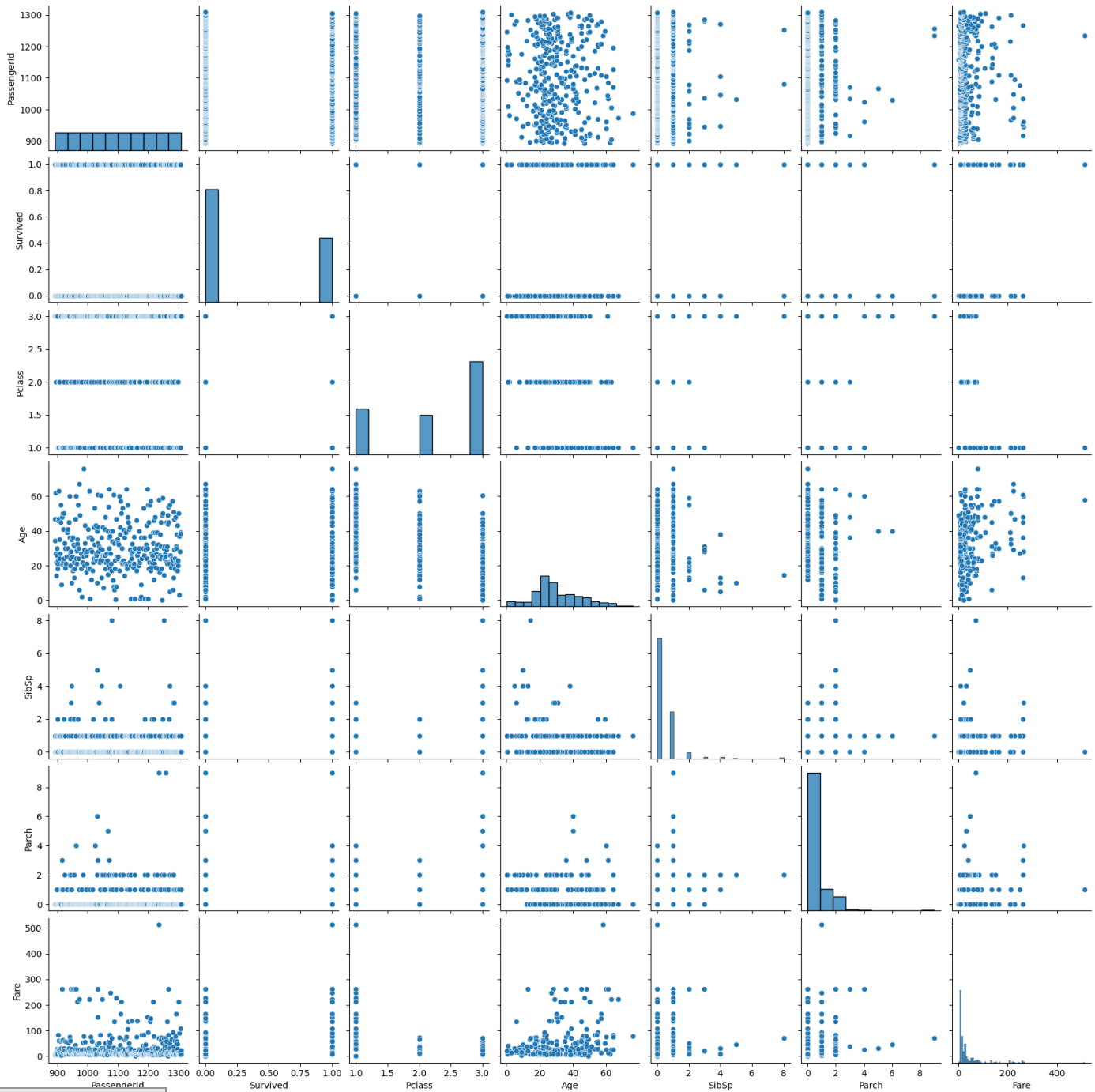
In [10]:
```
list(df)
```

Out[10]:
```
['PassengerId',
 'Survived',
 'Pclass',
 'Name',
 'Sex',
 'Age',
 'SibSp',
 'Parch',
 'Ticket',
 'Fare',
 'Cabin',
 'Embarked']
```

In [ ]:
```
```

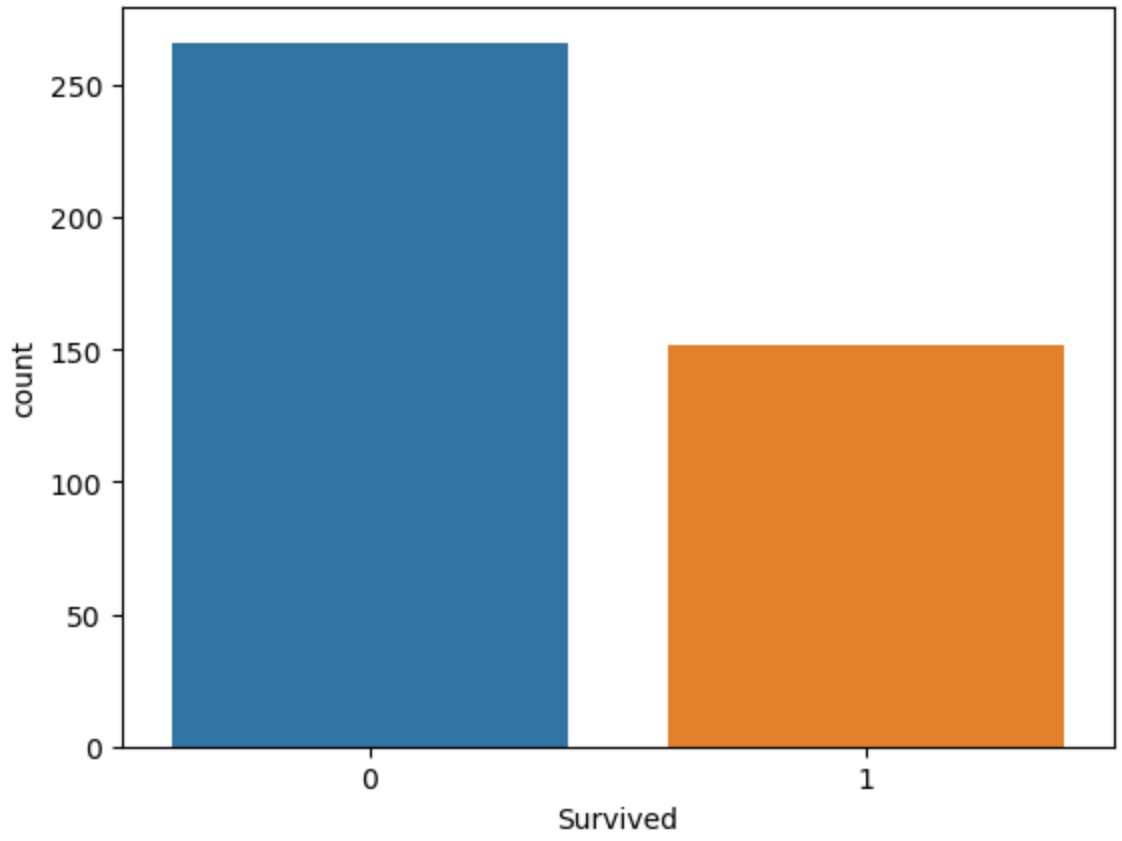
In [6]:
```
sns.pairplot(df)
```

Out[6]:
```
<seaborn.axisgrid.PairGrid at 0x22bcf962c80>
```



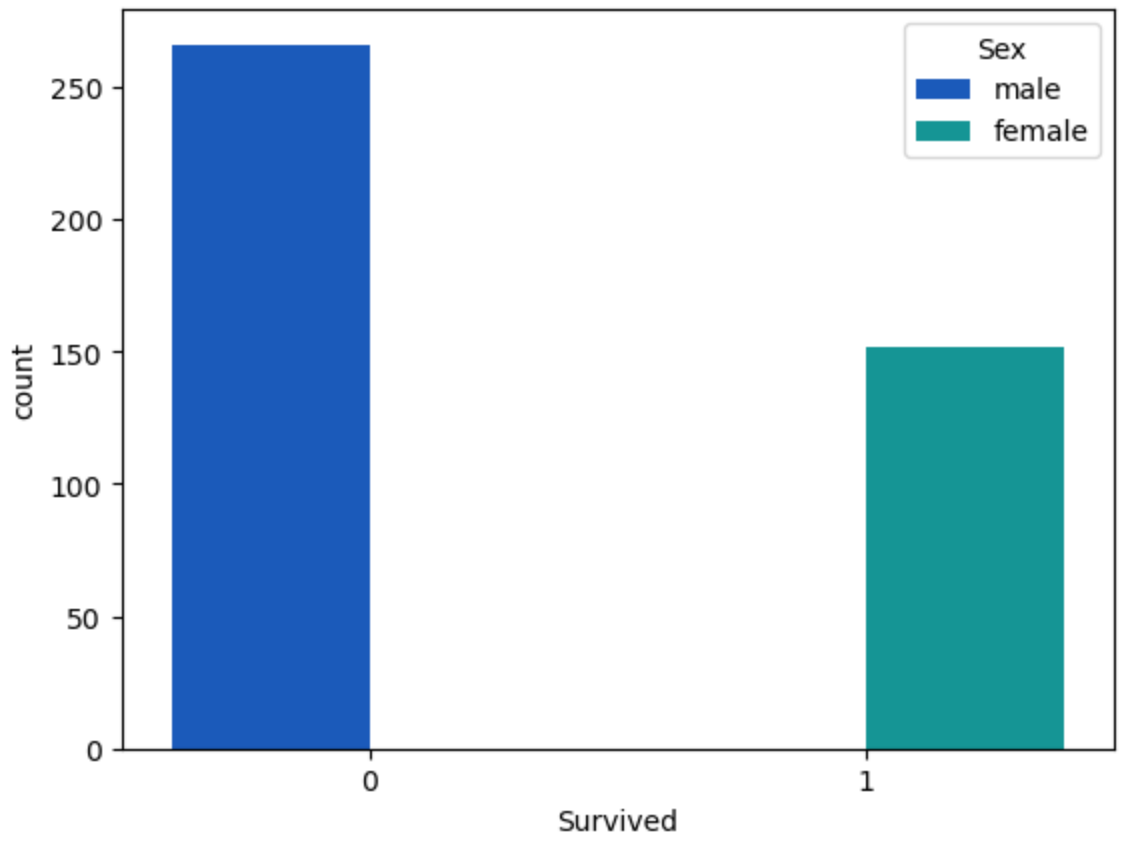
In [17]:
```
sns.countplot(x='Survived',data=df)
```

Out[17]:
```
<Axes: xlabel='Survived', ylabel='count'>
```



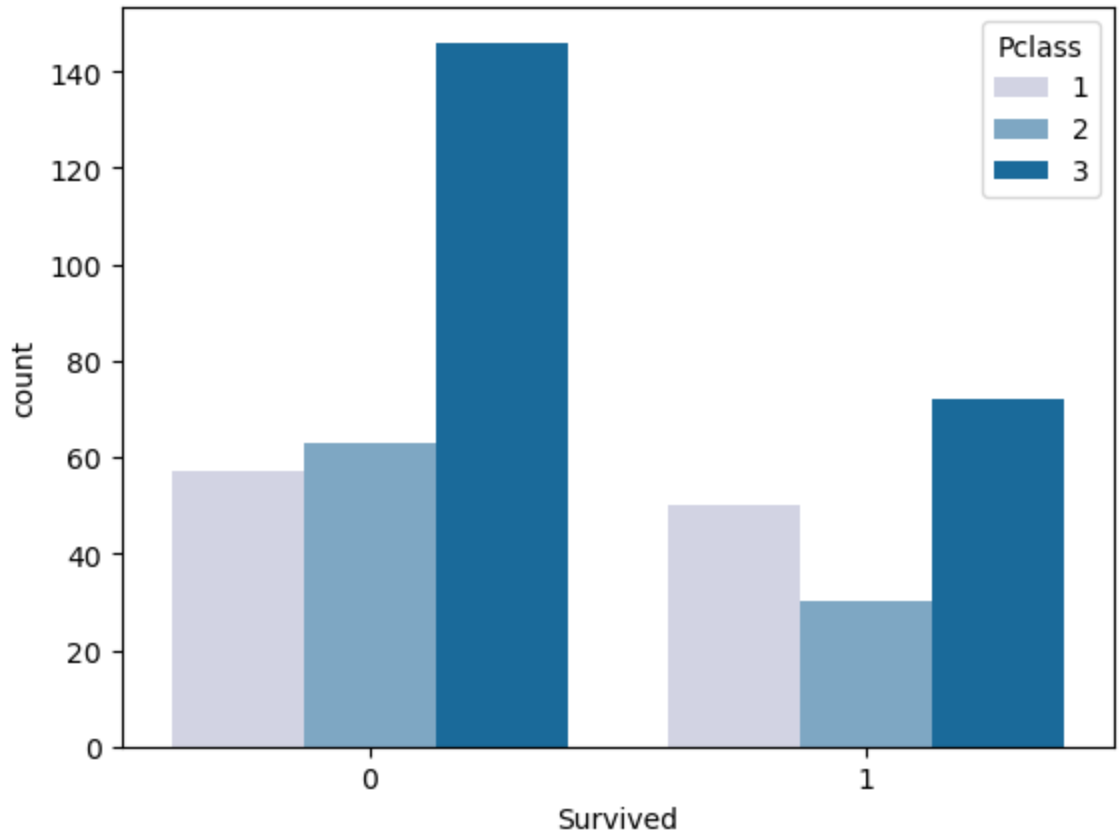
In [18]:
```
sns.countplot(x='Survived', hue= 'Sex' , data =df, palette='winter')
```

Out[18]:
```
<Axes: xlabel='Survived', ylabel='count'>
```



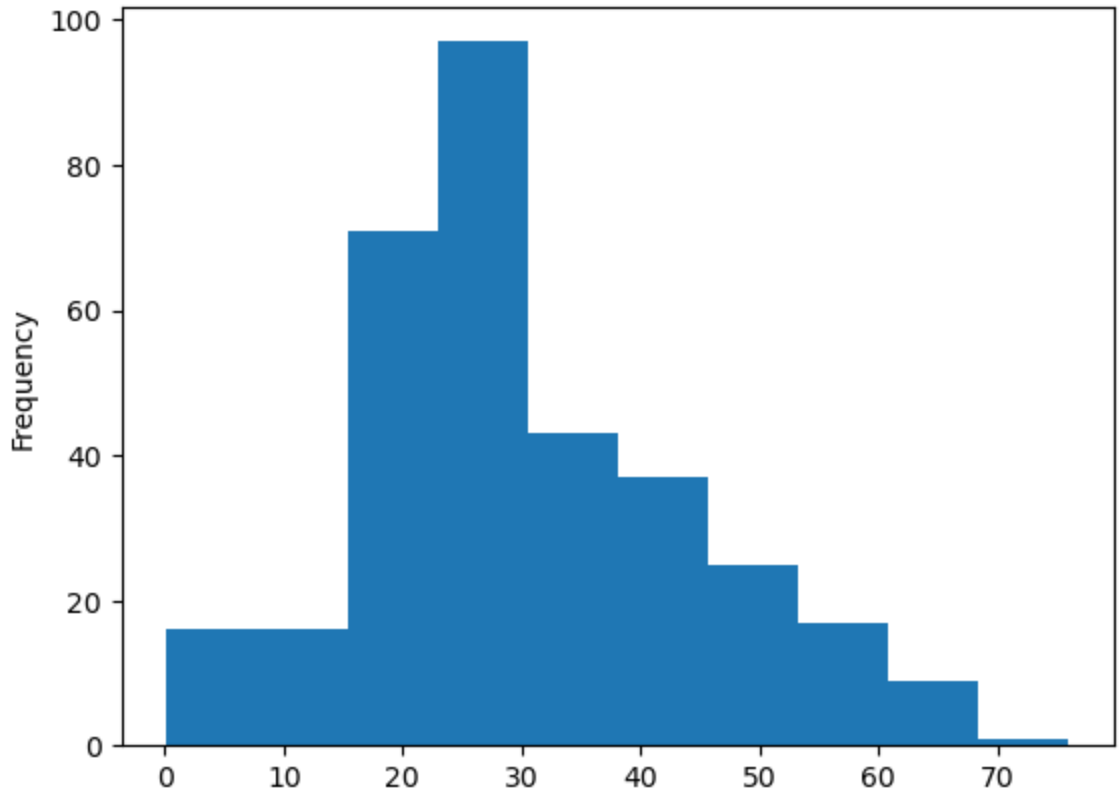
In [20]:
```
sns.countplot(x='Survived', hue= 'Pclass' , data =df, palette='PuBu')
```

Out[20]: <Axes: xlabel='Survived', ylabel='count'>



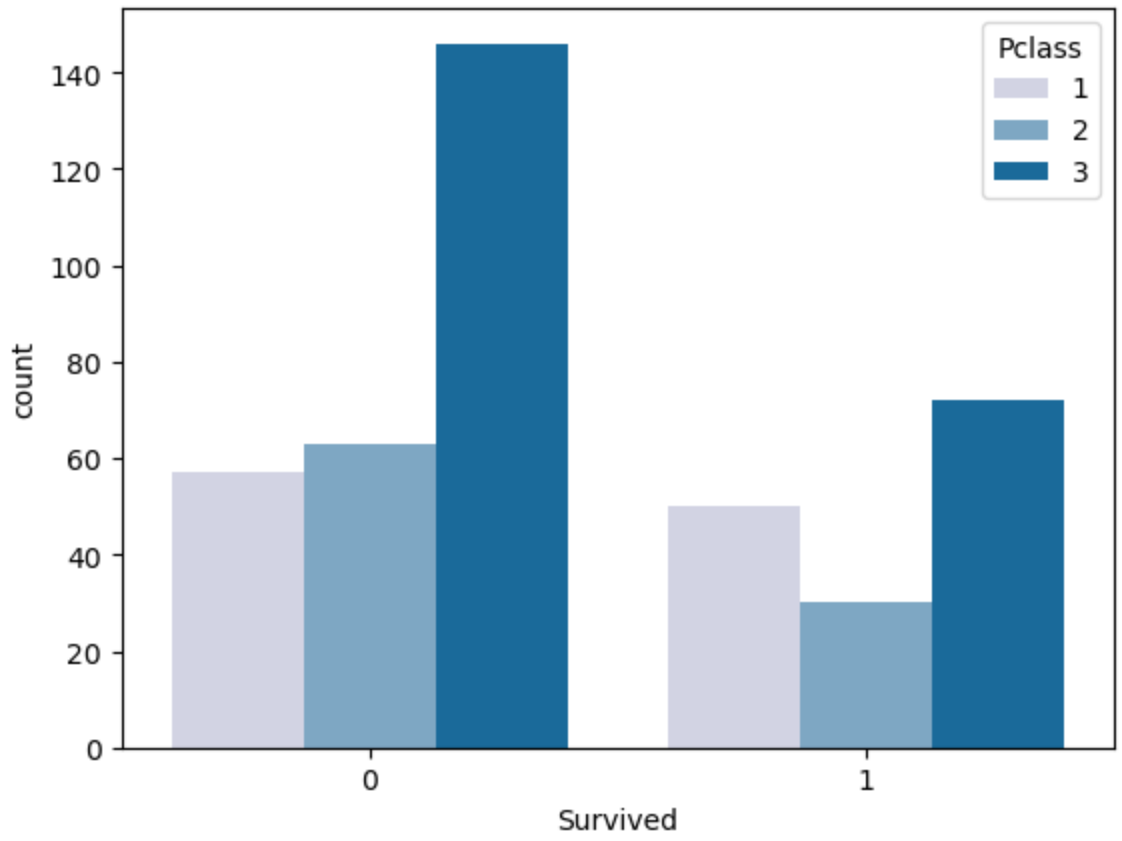
In [23]:
```
df['Age'].plot.hist()
```

Out[23]: <Axes: ylabel='Frequency'>



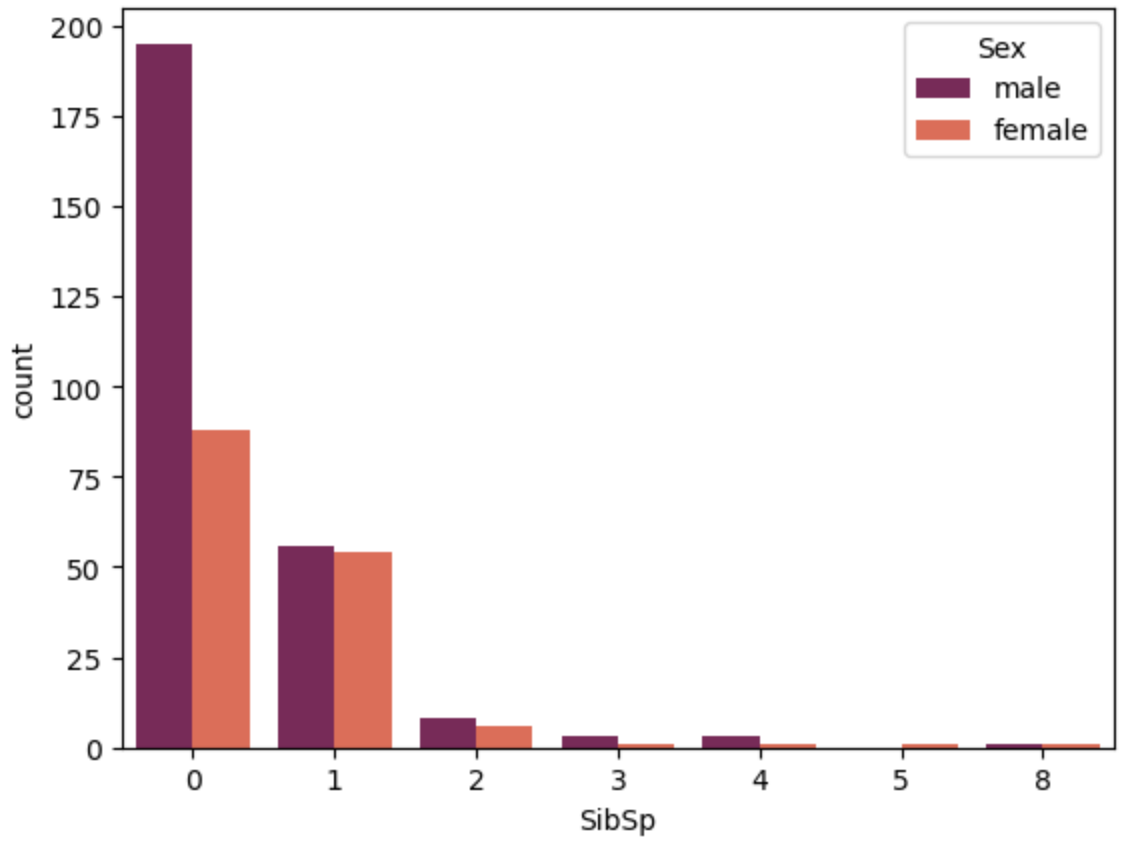
In [24]:
```
sns.countplot(x='Survived', hue= 'Pclass' , data =df, palette='PuBu')
```

Out[24]: <Axes: xlabel='Survived', ylabel='count'>



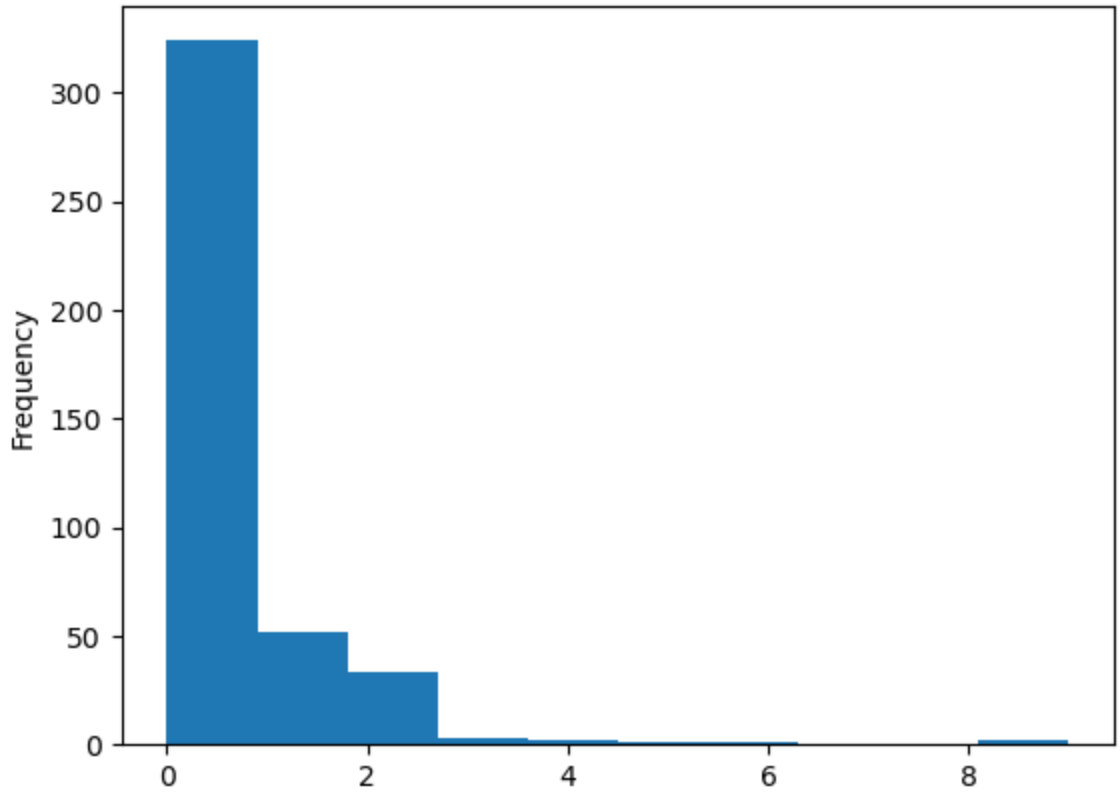
In [27]:
```
sns.countplot(x='SibSp', hue= 'Sex' ,data=df , palette='rocket')
```

Out[27]:
```
<Axes: xlabel='SibSp', ylabel='count'>
```



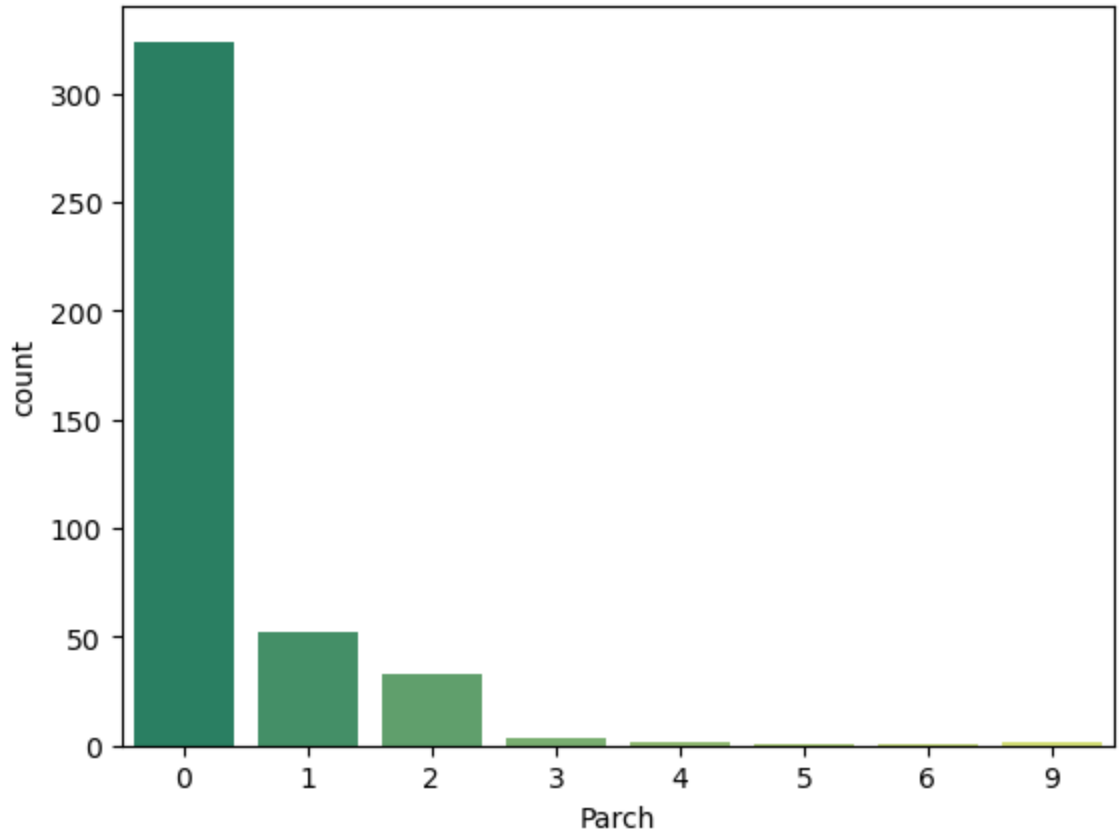
In [31]:
```
df['Parch'].plot.hist()
```

Out[31]:
```
<Axes: ylabel='Frequency'>
```



In [33]:
```
sns.countplot(x='Parch', data=df , palette='summer')
```

Out[33]:
```
<Axes: xlabel='Parch', ylabel='count'>
```



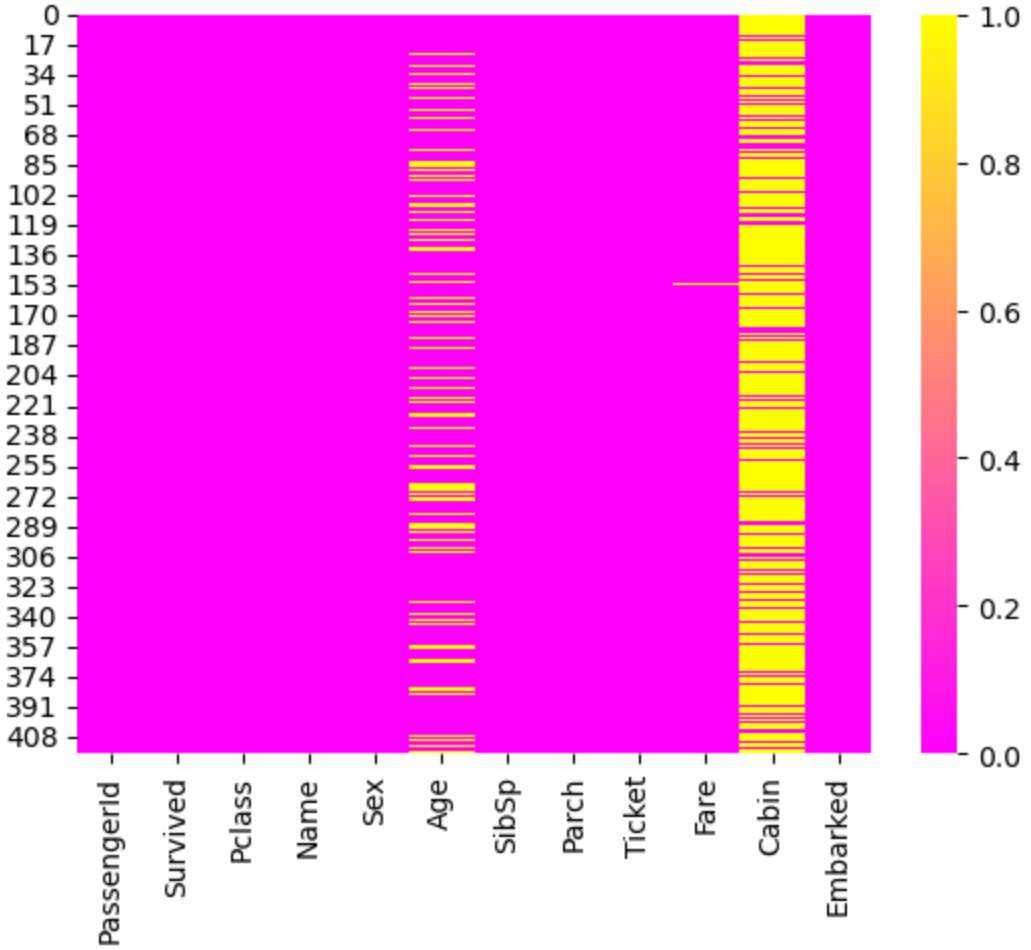
In [34]:
```
# cleaning the data
```

In [35]:
```
df.isnull().sum()
```

Out[35]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

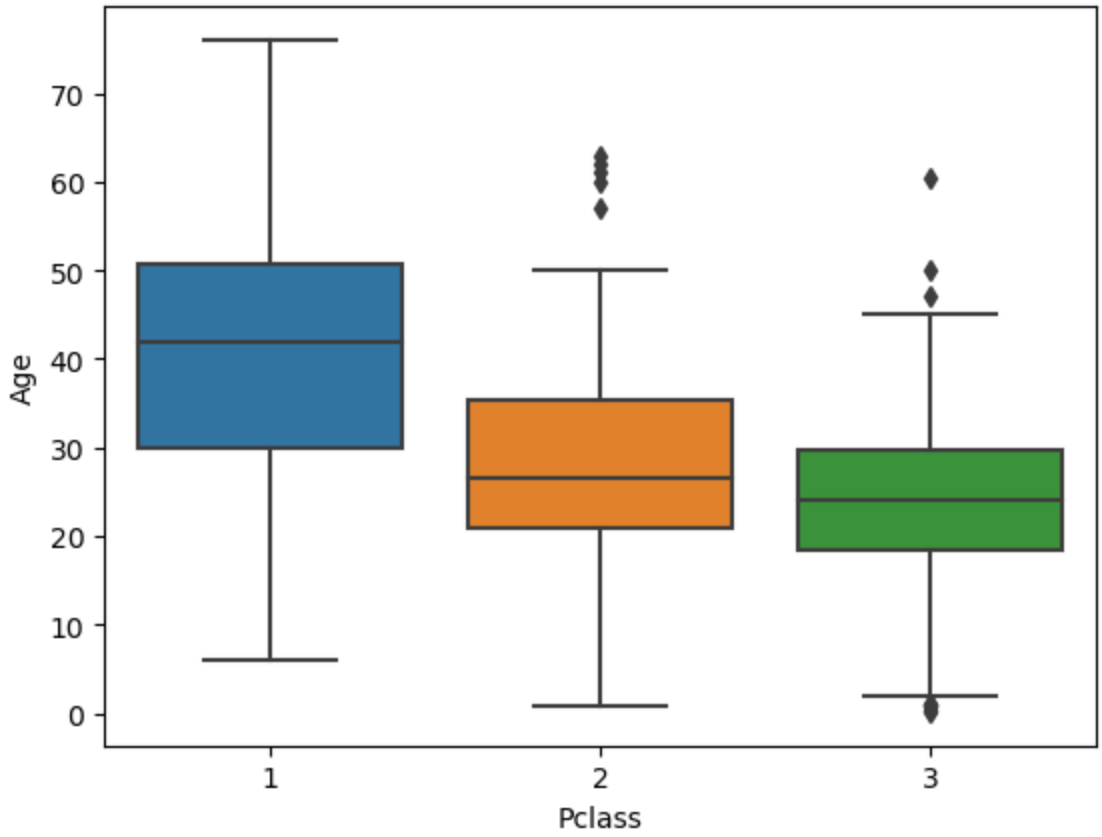
In [37]:
```
sns.heatmap(df.isnull() , cmap='spring')
```

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



In [38]:
```
sns.boxplot(x='Pclass' ,y='Age', data=df)
```

Out[38]:
```
<Axes: xlabel='Pclass', ylabel='Age'>
```



In [42]:
```
df.drop('Cabin', axis=1, inplace=True)
```

In [44]:
```
df.head(20)
```

Out[44]:

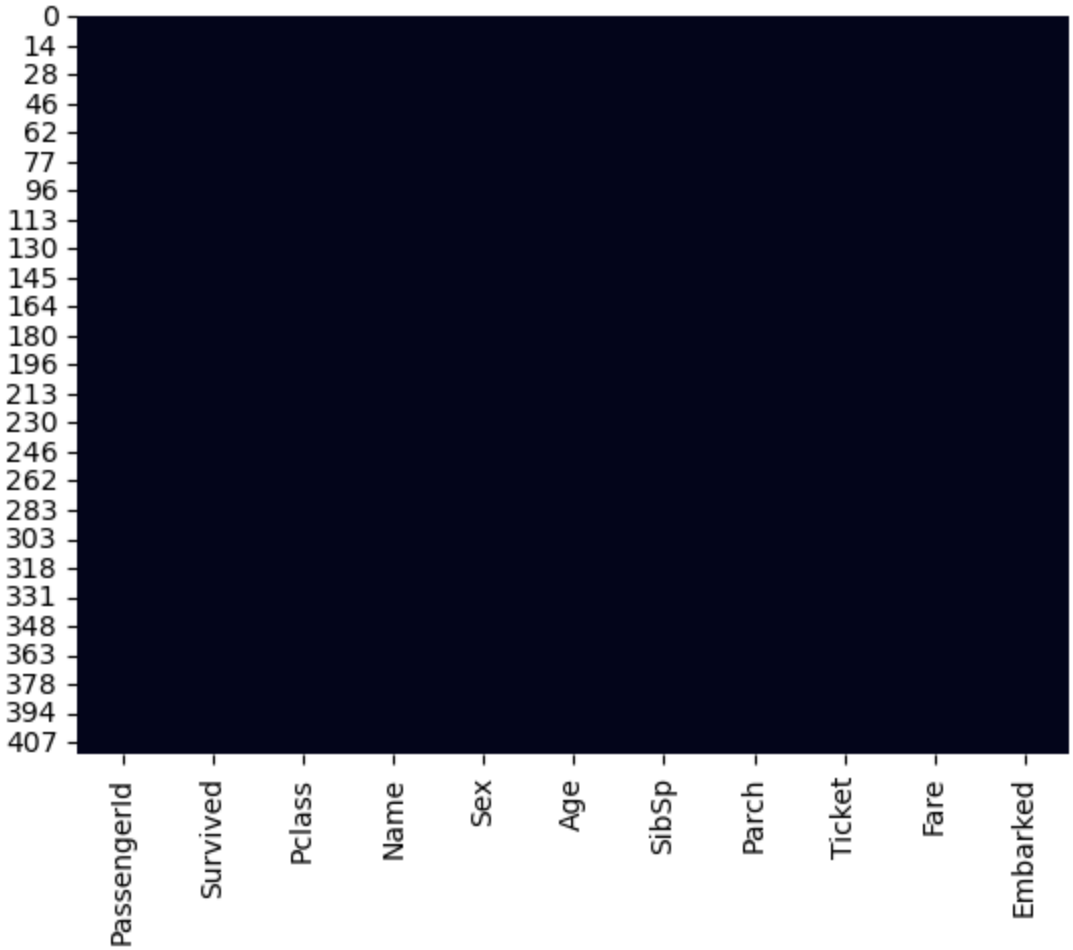
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	S
5	897	0	3	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538	9.2250	S
6	898	1	3	Connolly, Miss. Kate	female	30.0	0	0	330972	7.6292	Q
7	899	0	2	Caldwell, Mr. Albert Francis	male	26.0	1	1	248738	29.0000	S
8	900	1	3	Abrahim, Mrs. Joseph (Sophie Halaut Easu)	female	18.0	0	0	2657	7.2292	C
9	901	0	3	Davies, Mr. John Samuel	male	21.0	2	0	A/4 48871	24.1500	S
10	902	0	3	Ilieff, Mr. Ylio	male	NaN	0	0	349220	7.8958	S
11	903	0	1	Jones, Mr. Charles Cresson	male	46.0	0	0	694	26.0000	S
12	904	1	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	female	23.0	1	0	21228	82.2667	S
13	905	0	2	Howard, Mr. Benjamin	male	63.0	1	0	24065	26.0000	S
14	906	1	1	Chaffee, Mrs. Herbert Fuller (Carrie Constance...	female	47.0	1	0	W.E.P. 5734	61.1750	S
15	907	1	2	del Carlo, Mrs. Sebastiano (Argenia Genovesi)	female	24.0	1	0	SC/PARIS 2167	27.7208	C
16	908	0	2	Keane, Mr. Daniel	male	35.0	0	0	233734	12.3500	Q
17	909	0	3	Assaf, Mr. Gerios	male	21.0	0	0	2692	7.2250	C
18	910	1	3	Ilmakangas, Miss. Ida Livija	female	27.0	1	0	STON/O2. 3101270	7.9250	S

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
19	911	1	3	Assaf Khalil, Mrs. Mariana (Miriam")"	female	45.0	0	0	2696	7.2250	C

In [45]:
```
df.dropna(inplace=True)
```

In [47]:
```
sns.heatmap(df.isnull() , cbar=False)
```

Out[47]:
```
<Axes: >
```



In [51]:
```
#-------------------------------------THE END-----------------------------------------
```

In [ ]:
```

```