In [70]:

```python
import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt
import math
%matplotlib in line
```

```
UsageError: unrecognized arguments: line
```

In [14]:

```python
titanic=pd.read_csv(r"C:\Users\Burhan\OneDrive\Desktop\train.csv")
```

In [15]:

```python
titanic.head()
```

Out[15]:

	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

In [16]:

```python
titanic.shape
```

Out[16]:

```
(891, 12)
```

In [ ]:

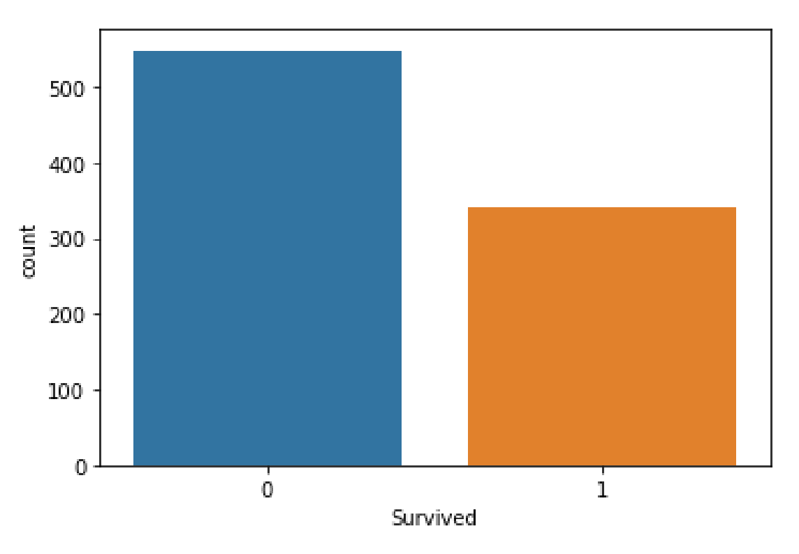
```python
#ANALYSING DATA
```

In [29]:

```python
sb.countplot(x="Survived",data=titanic)
```

Out[29]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

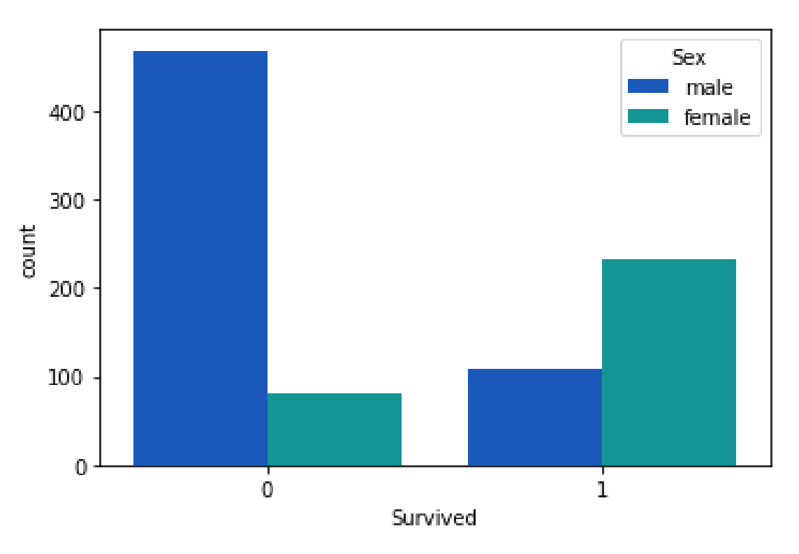


In [30]:

```python
sb.countplot(x="Survived",hue="Sex",data=titanic,palette="winter")
```

Out[30]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

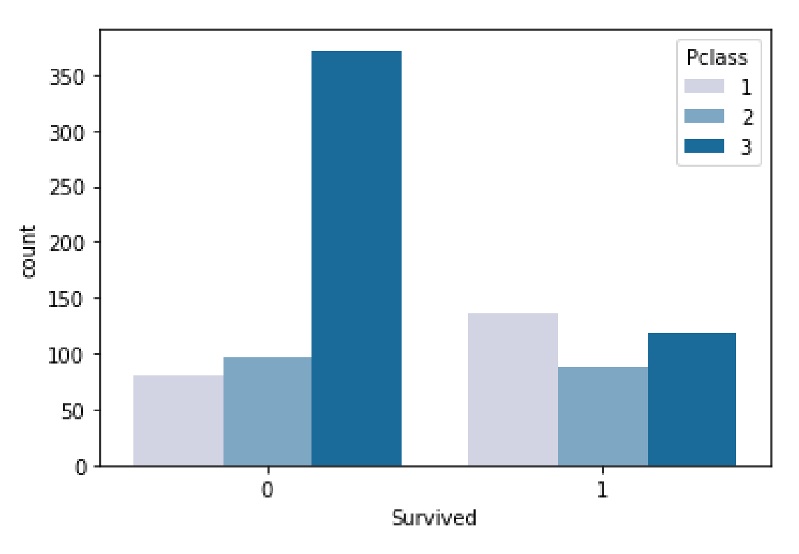


In [32]:

```python
sns.countplot(x="Survived",hue="Pclass",data=titanic,palette="PuBu")
```

Out[32]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

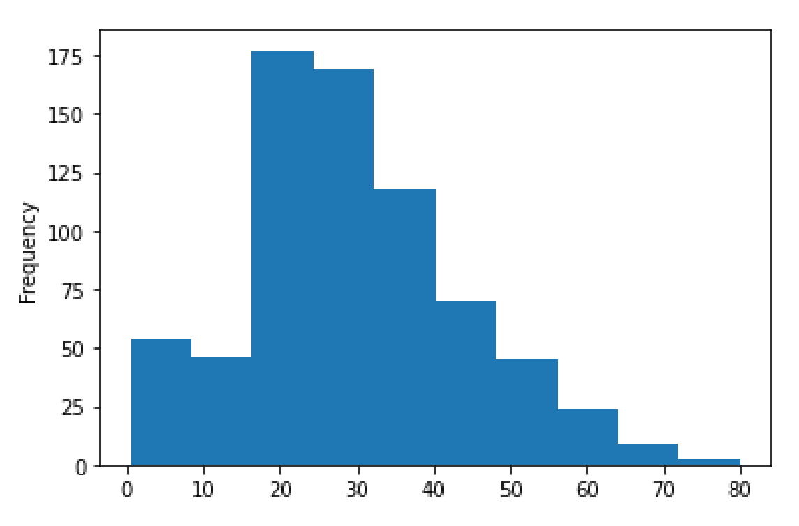


In [33]:

```python
titanic["Age"].plot.hist()
```

Out[33]:

```
<AxesSubplot:ylabel='Frequency'>
```

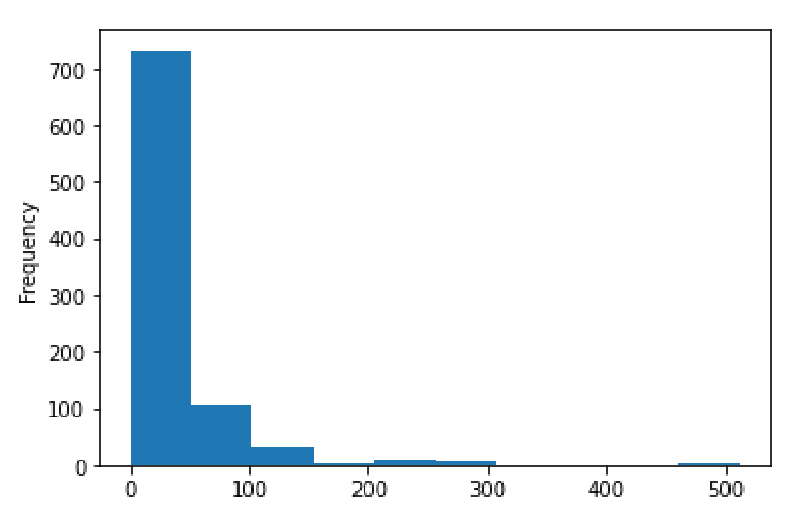


In [34]:

```python
titanic["Fare"].plot.hist()
```

Out[34]:

```
<AxesSubplot:ylabel='Frequency'>
```



In [ ]:

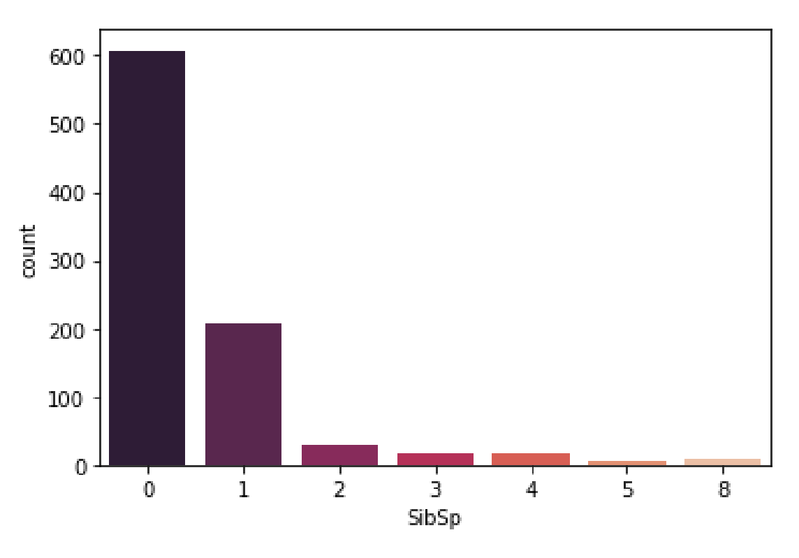
```python
#countplot for siblings
```

In [37]:

```python
sb.countplot(x="SibSp",data=titanic,palette="rocket")
```

Out[37]:

```
<AxesSubplot:xlabel='SibSp', ylabel='count'>
```

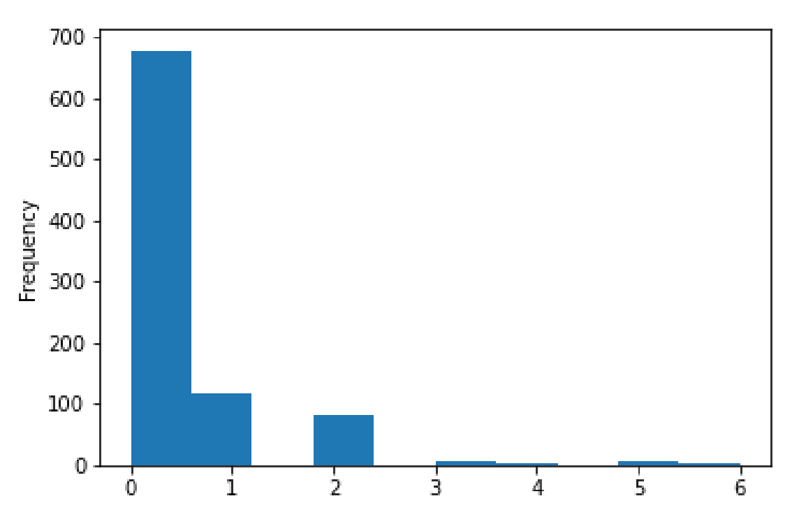


In [38]:

```python
titanic["Parch"].plot.hist()
```

Out[38]:

```
<AxesSubplot:ylabel='Frequency'>
```



In [39]:

```python
titanic.isnull().sum()
```

Out[39]:

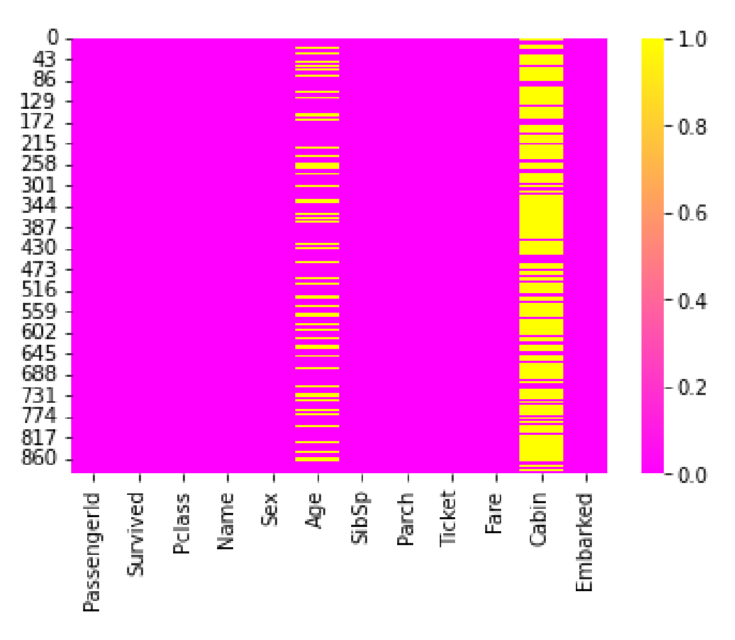
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [40]:

```python
sb.heatmap(titanic.isnull(),cmap="spring")#to determine null values
```

Out[40]:

```
<AxesSubplot:>
```

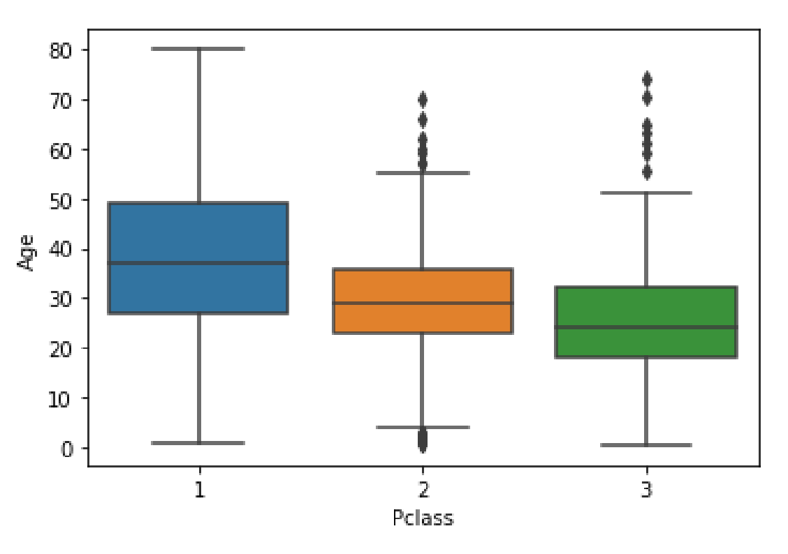


In [41]:

```python
sb.boxplot(x="Pclass",y="Age",data=titanic)
```

Out[41]:

```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



In [42]:

```python
titanic.head()
```

Out[42]:

	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

In [44]:

```python
titanic.drop("Cabin",axis=1,inplace=True)#drop cabin column
```

In [45]:

```python
titanic.head()
```

Out[45]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S

In [46]:

```python
titanic.dropna(inplace=True)
```

In [47]:

```python
titanic.head()
```

Out[47]:

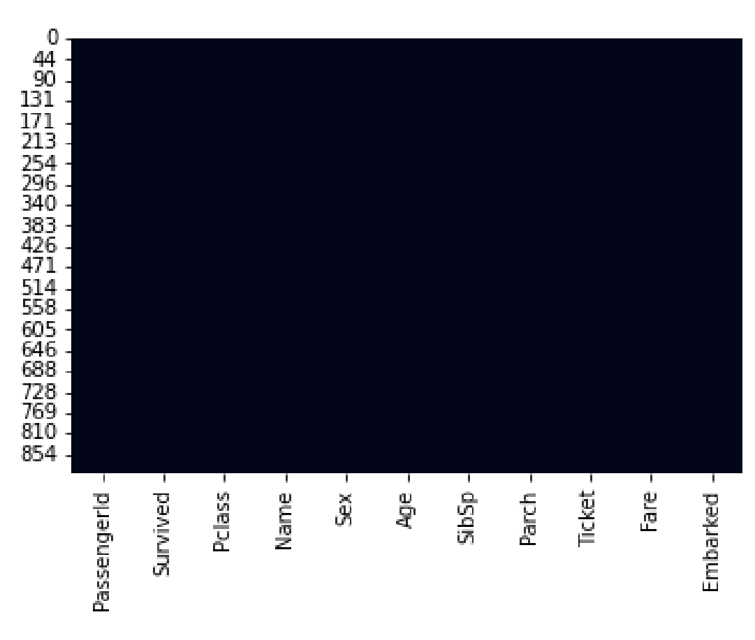
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S

In [48]:

```python
sb.heatmap(titanic.isnull(),cbar=False)
```

Out[48]:

```
<AxesSubplot:>
```



In [49]:

```python
titanic.isnull().sum()
```

Out[49]:

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [50]:

```python
titanic.head(2)
```

Out[50]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C

In [ ]:

```python
#we will convert some column (string)into categorical data
```

In [51]:

```python
pd.get_dummies(titanic["Sex"]).head()
```

Out[51]:

	female	male
0	0	1
1	1	0
2	1	0
3	1	0
4	0	1

In [54]:

```python
sex=pd.get_dummies(titanic["Sex"],drop_first=True)
```

In [55]:

```python
sex.head(3)
```

Out[55]:

	male
0	1
1	0
2	0

In [56]:

```python
embark=pd.get_dummies(titanic["Embarked"]).head()
```

In [57]:

```python
embark.head(2)
```

Out[57]:

	C	Q	S
0	0	0	1
1	1	0	0

In [58]:

```python
embark=pd.get_dummies(titanic["Embarked"],drop_first=True)
```

In [59]:

```python
embark.head(2)
```

Out[59]:

	Q	S
0	0	1
1	0	0

In [60]:

```python
pclass=pd.get_dummies(titanic["Pclass"]).head()
```

In [61]:

```python
pclass.head(3)
```

Out[61]:

	1	2	3
0	0	0	1
1	1	0	0
2	0	0	1

In [62]:

```python
pclass=pd.get_dummies(titanic["Pclass"],drop_first=True)
```

In [63]:

```python
pclass.head(3)
```

Out[63]:

	2	3
0	0	1
1	0	0
2	0	1

In [ ]:

```python
#our data is converted into categorical data.
```

In [ ]:

```python
#DELETING UNWANTED COLUMNS.
```

In [64]:

```python
titanic.drop(["Name","PassengerId","Pclass","Ticket","Sex","Embarked"],axis=1,inplace=True)
```

In [65]:

```python
titanic.head()
```

Out[65]:

	Survived	Age	SibSp	Parch	Fare
0	0	22.0	1	0	7.2500
1	1	38.0	1	0	71.2833
2	1	26.0	0	0	7.9250
3	1	35.0	1	0	53.1000
4	0	35.0	0	0	8.0500

In [ ]:

```python
#TRAIN DATA
```

In [68]:

```python
x=titanic.drop("Survived",axis=1)
y=titanic["Survived"]
```

In [69]:

```python
x=titanic.drop("Survived",axis=1)
y=titanic["Survived"]
```

In [ ]:

```python

```