In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import math
% matplotlib in line
```

```
UsageError: Line magic function `%` not found.
```

In [2]:
```python
titanic= pd.read_csv("train.csv")
titanic.head()
```

Out[2]:

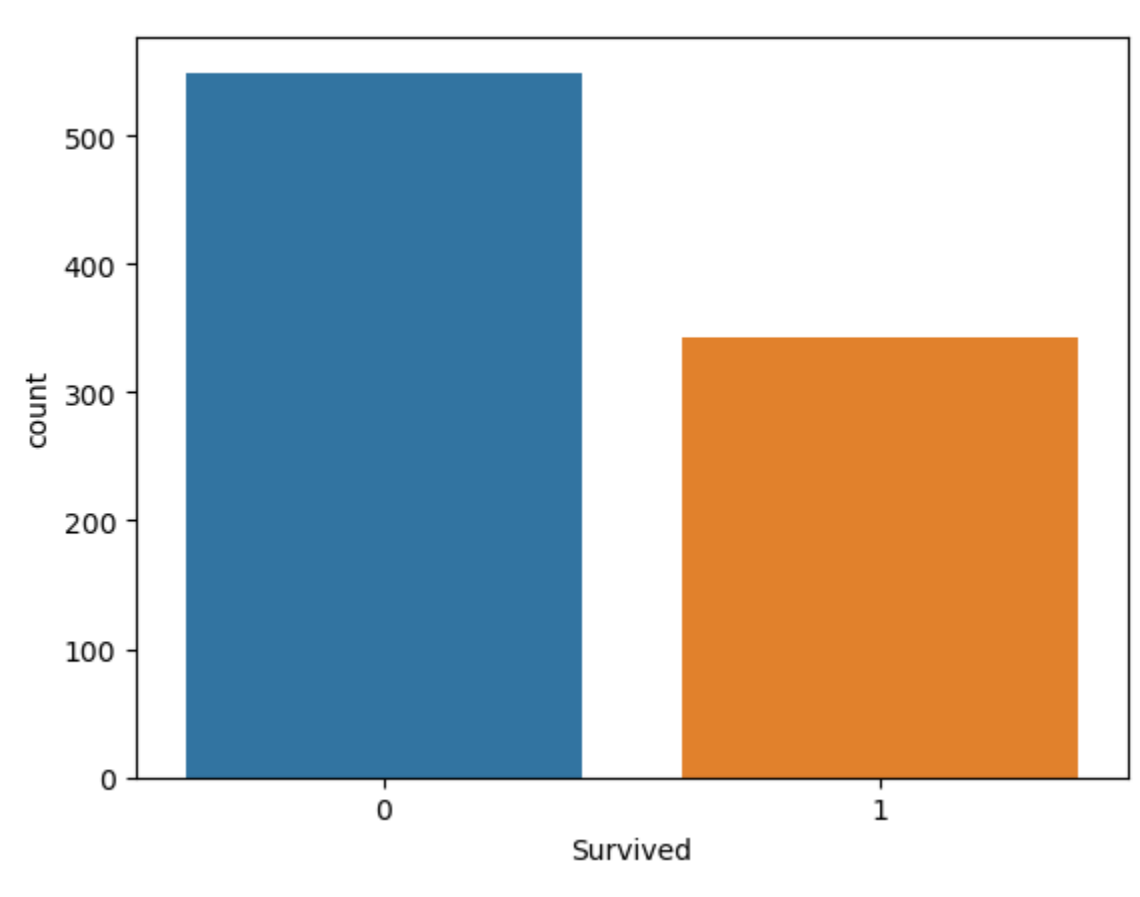
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 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 [3]:
```python
titanic.shape                # to obtain the no of rows and column
```

Out[3]: (891, 12)

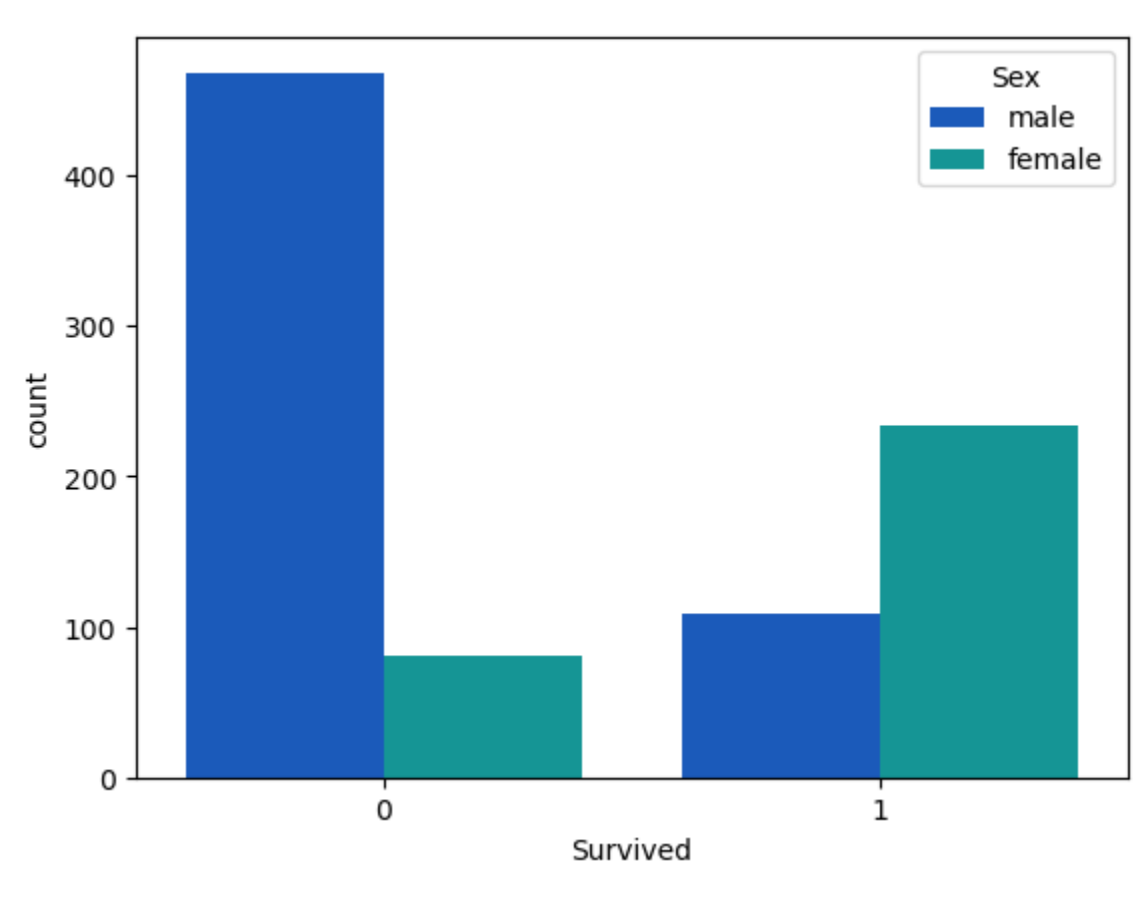
In [4]:
```python
sns.countplot(x= 'Survived', data=titanic)
```

Out[4]: <Axes: xlabel='Survived', ylabel='count'>



In [5]:
```python
sns.countplot(x='Survived', hue='Sex', data=titanic, palette='winter')
```

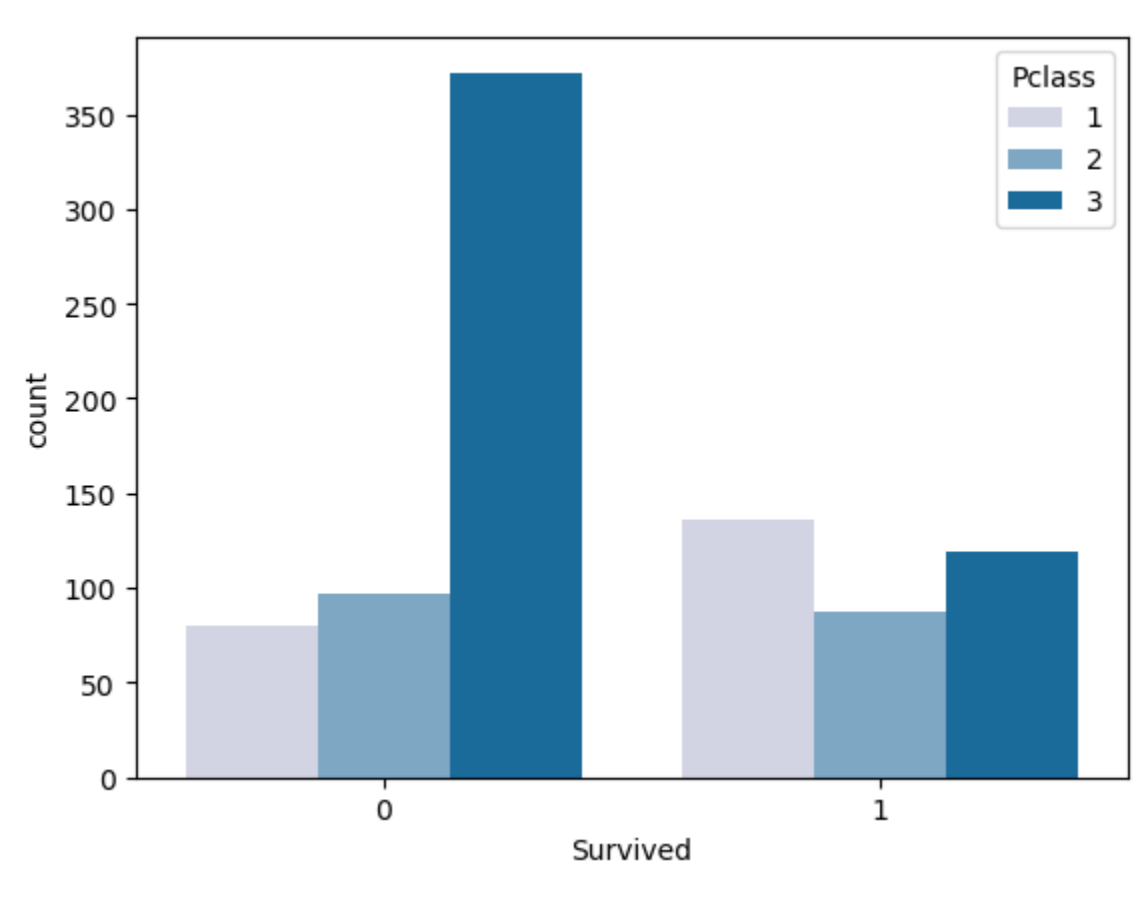
Out[5]: <Axes: xlabel='Survived', ylabel='count'>



In [ ]:
```python
## Analysis: 0 represent not survived and 1 is survived
```

In [6]:
```python
sns.countplot(x='Survived', hue='Pclass', data=titanic, palette='PuBu')
```

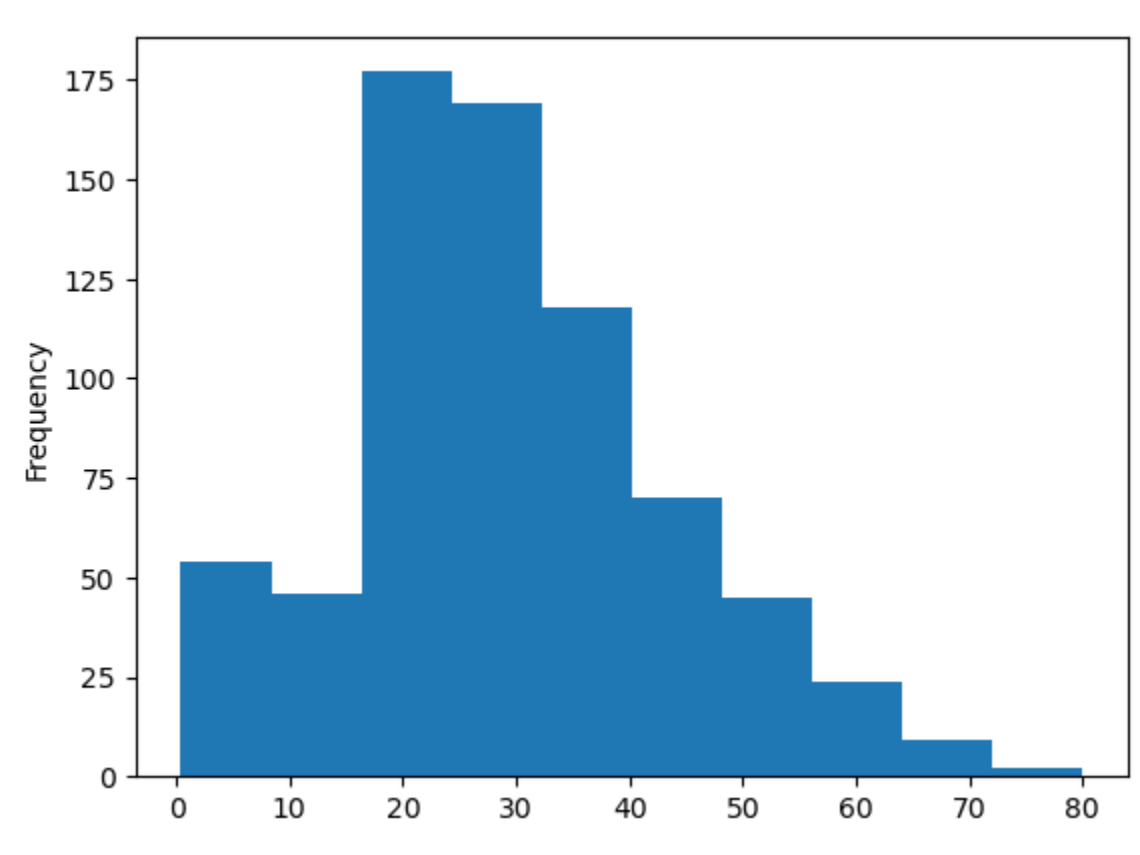
Out[6]: <Axes: xlabel='Survived', ylabel='count'>



In [ ]:
```python
## Analysis: the passenger who did not survived belong to the 3 class
# 1 class passenger are likely to survive
```

In [7]:
```python
titanic['Age'].plot.hist()
```

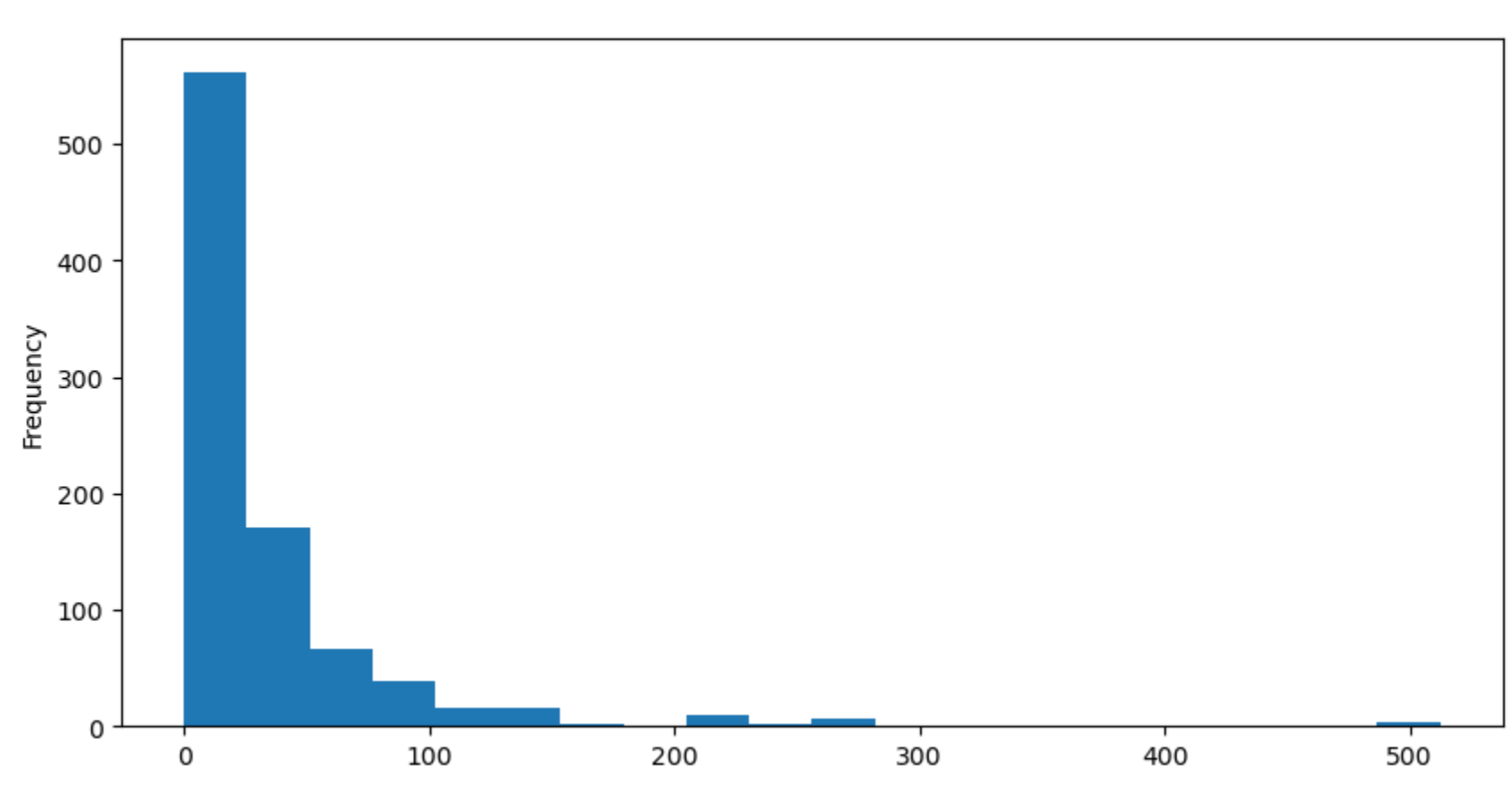
Out[7]: <Axes: ylabel='Frequency'>



In [ ]:
```python
# we notice that highest group travelling are among the age between 20-40
# very few passenger are in the group 70-80
```

In [8]:
```python
titanic['Fare'].plot.hist(bins=20, figsize=(10,5))
```

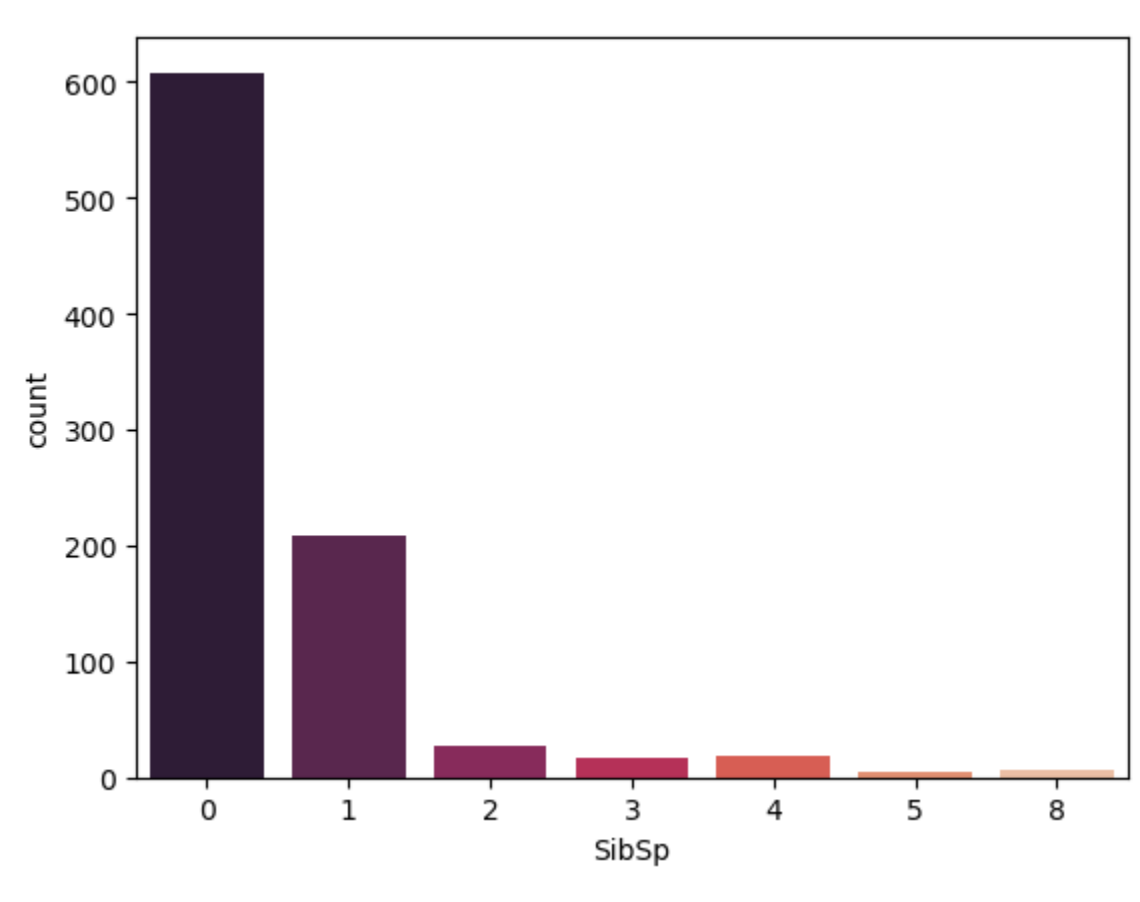
Out[8]: <Axes: ylabel='Frequency'>



In [ ]:
```python
# we observe that most of the tickets are bought under fare 100
# and very few are on the higher side of fare i.e. 220-500 range
```

In [9]:
```python
sns.countplot(x='SibSp', data=titanic, palette='rocket')
```

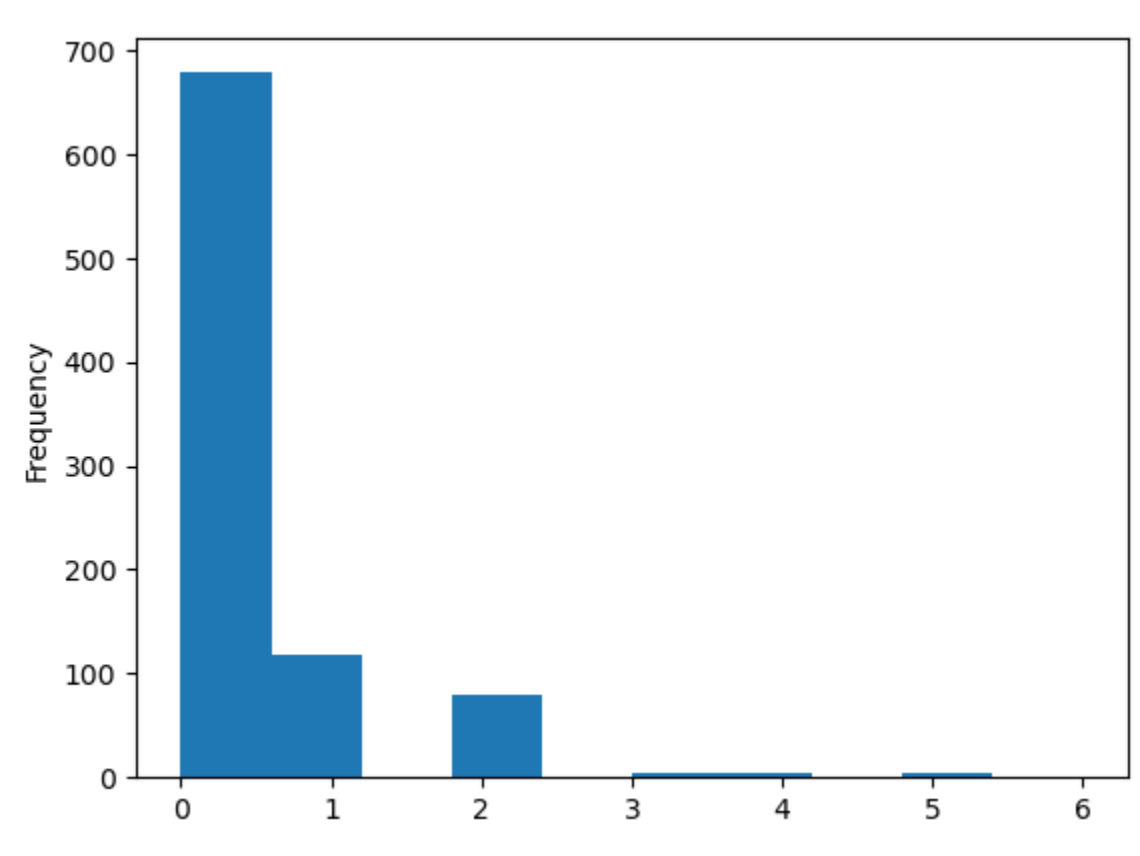
Out[9]: <Axes: xlabel='SibSp', ylabel='count'>



In [ ]:
```python
# we notice that most of the passenger donot have their siblings aboard.
```

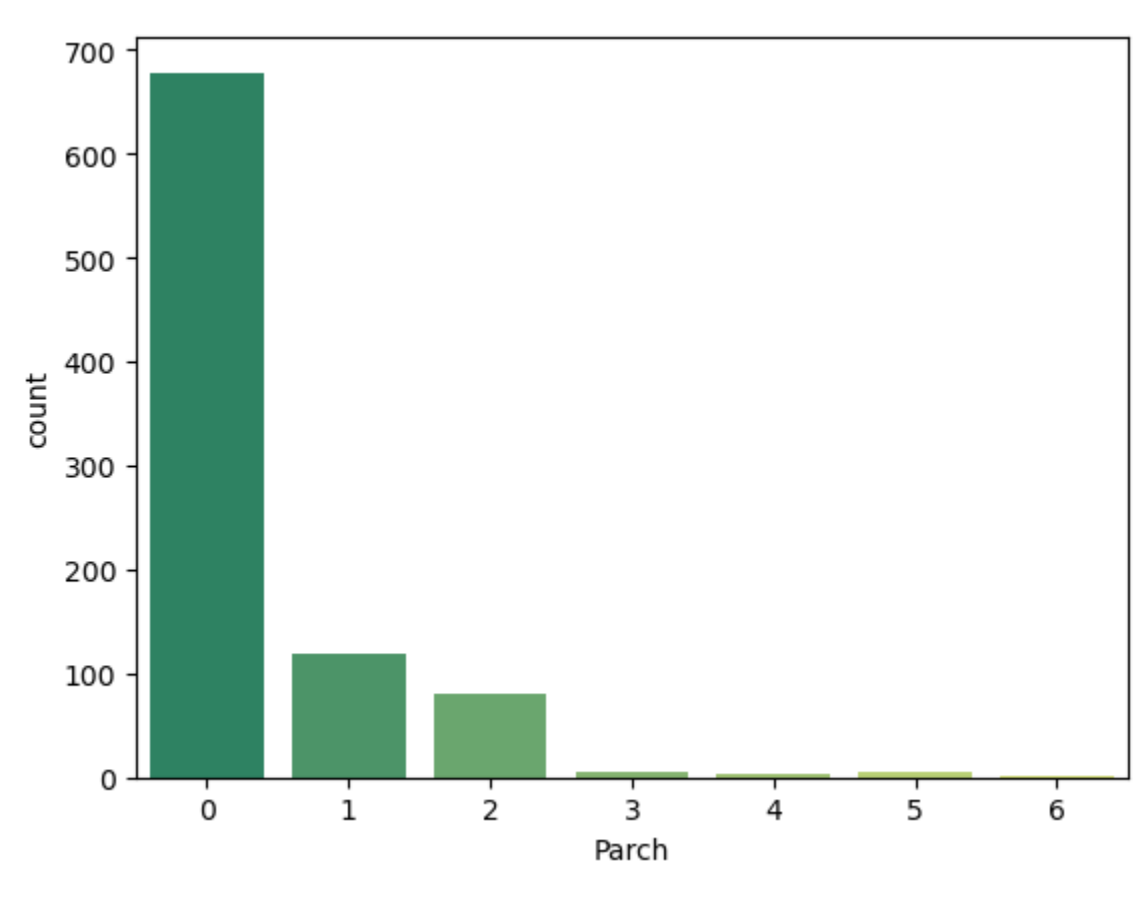
In [10]:
```python
titanic['Parch'].plot.hist()
```

Out[10]: <Axes: ylabel='Frequency'>



In [11]:
```python
sns.countplot(x='Parch', data=titanic, palette='summer')
```

Out[11]: <Axes: xlabel='Parch', ylabel='count'>



In [ ]:
```python
# data wrangling means cleaning the data, removing the null values
## deopping unwanted columns, adding new ones if needed.
```

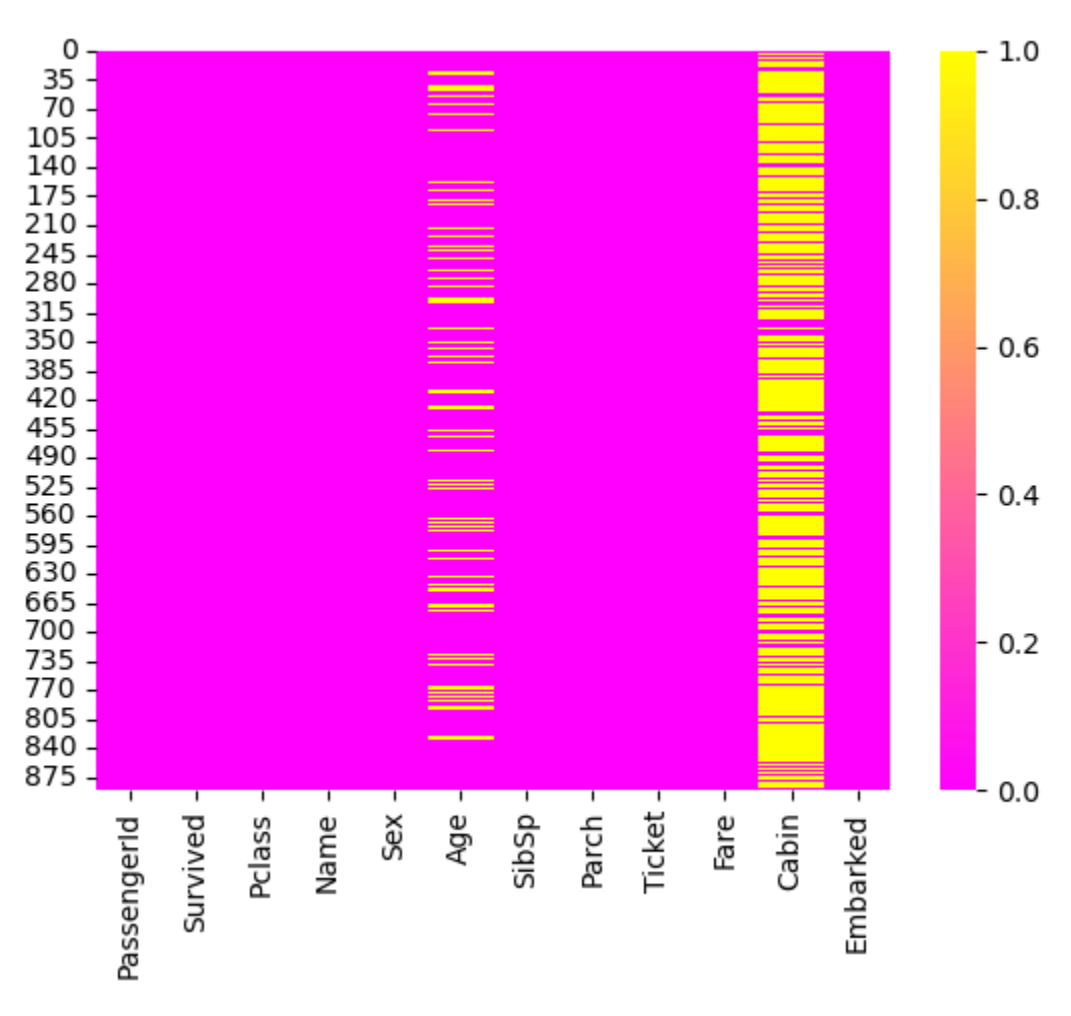
In [12]:
```python
titanic.isnull().sum()
```

Out[12]:
```
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 [ ]:
```python
## age and cabin has most null values and embarked too has null values
## we can plot it on heat map
```

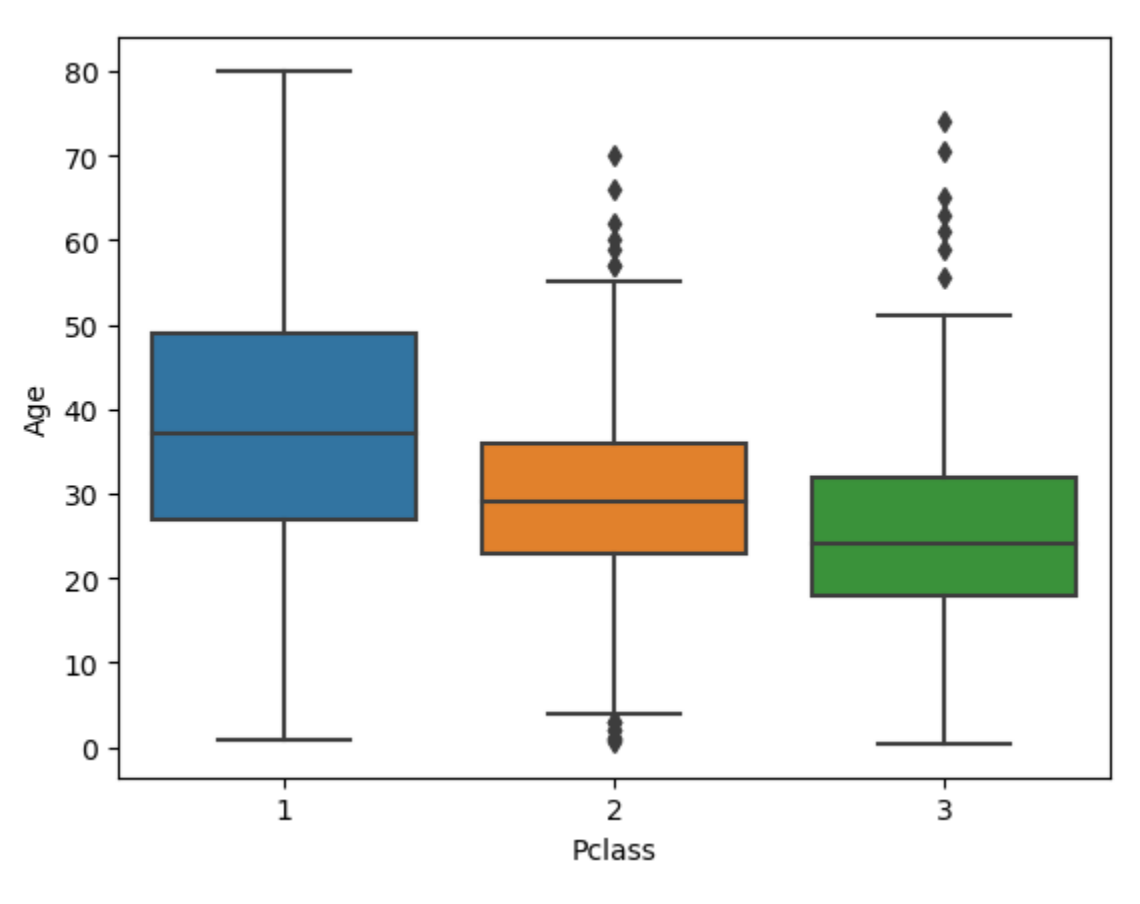
In [13]:
```python
sns.heatmap(titanic.isnull() , cmap='spring')
```

Out[13]: <Axes: >



In [15]:
```python
sns.boxplot(x='Pclass', y='Age', data=titanic)
```

Out[15]: <Axes: xlabel='Pclass', ylabel='Age'>



In [16]:
```python
titanic.head()
```

Out[16]:

	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

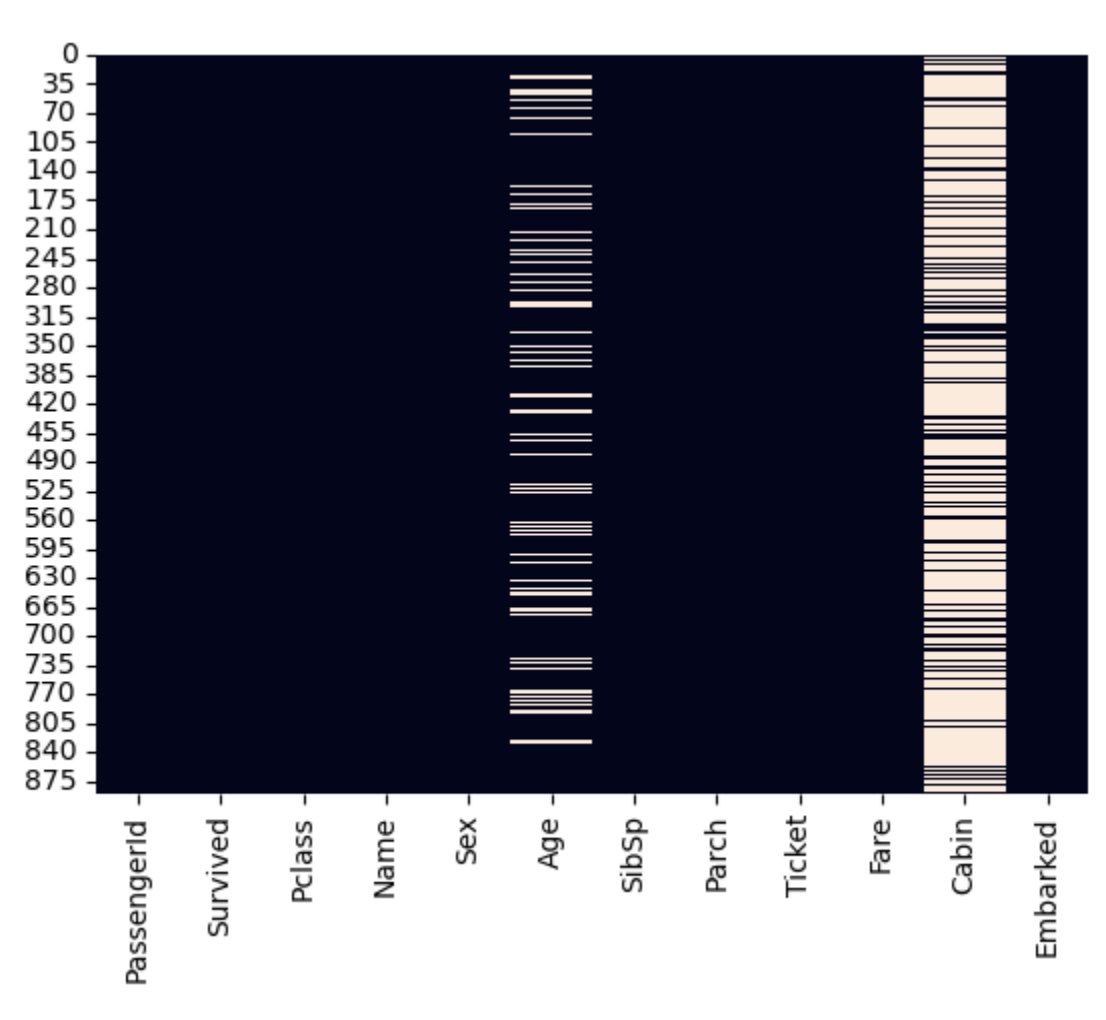
In [17]:
```python
titanic.head(3)
```

Out[17]:

	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

In [18]:
```python
sns.heatmap(titanic.isnull(), cbar=False)
```

Out[18]: <Axes: >



In [19]:
```python
titanic.isnull().sum()
```

Out[19]:
```
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 [20]:
```python
titanic.head(2)
```

Out[20]:

	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 [21]:
```python
pd.get_dummies(titanic['Sex']).head()
```

Out[21]:

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

In [24]:
```python
sex=pd.get_dummies(titanic['Sex'], drop_first=True)
sex.head(3)
```

Out[24]:

	male
0	1
1	0
2	0

In [26]:
```python
embark=pd.get_dummies(titanic['Embarked'])
```

In [27]:
```python
embark.head(3)
```

Out[27]:

	C	Q	S
0	0	0	1
1	1	0	0
2	0	0	1

In [28]:
```python
embark=pd.get_dummies(titanic['Embarked'], drop_first=True)
```

In [29]:
```python
embark.head(3)
```

Out[29]:

	Q	S
0	0	1
1	0	0
2	0	1

In [30]:
```python
Pcl=pd.get_dummies(titanic['Pclass'], drop_first=True)
Pcl.head(3)
```

Out[30]:

	2	3
0	0	1
1	0	0
2	0	1

In [ ]:
```python
# our data is converted into categorial data
```

In [31]:
```python
titanic=pd.concat([titanic, sex, embark,Pcl], axis=1)
```

In [32]:
```python
titanic.head(3)
```

Out[32]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	male	Q	S	2	3
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	1	0	1	0	1
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	0	0	0	0	0
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	0	0	1	0	1

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: