

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

Matplotlib is building the font cache; this may take a moment.

```
In [2]: #READ THE DATASET('TITANIC')
sc = pd.read_csv('train.csv')
```

```
In [3]: sc.tail()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Er
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	



```
In [4]: sc.head()
```

Out[4]:

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

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	
3	4	1	1	Futrelle,	female	35.0	1	0	113803	53.1000	C123
				Mrs.							
				Jacques							
				Heath							
				(Lily May							
				Peel)							
4	5	0	3	Allen, Mr.	male	35.0	0	0	373450	8.0500	NaN
				William							
				Henry							



In [5]:

sc.shape

Out[5]: (891, 12)

In [6]:

sc.isnull().sum()

Out[6]: 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 [7]:

sc.isnull()

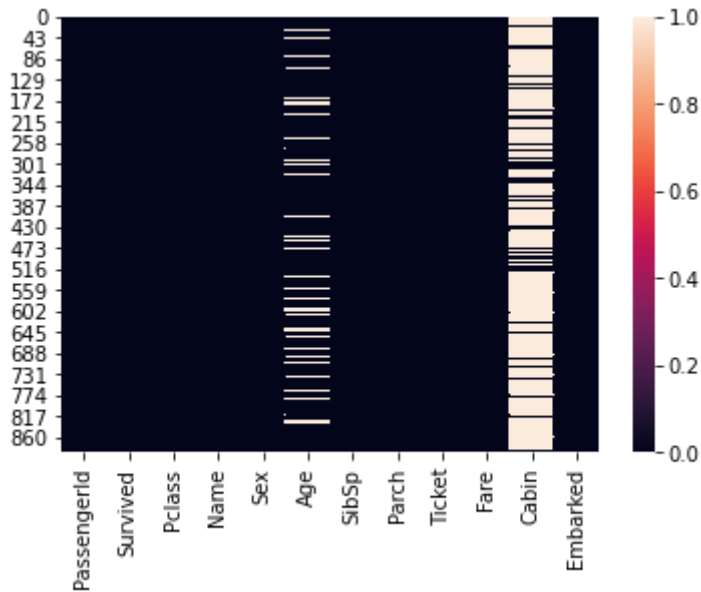
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	True	Fal
1	False	False	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	True	Fal
3	False	False	False	False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

891 rows × 12 columns



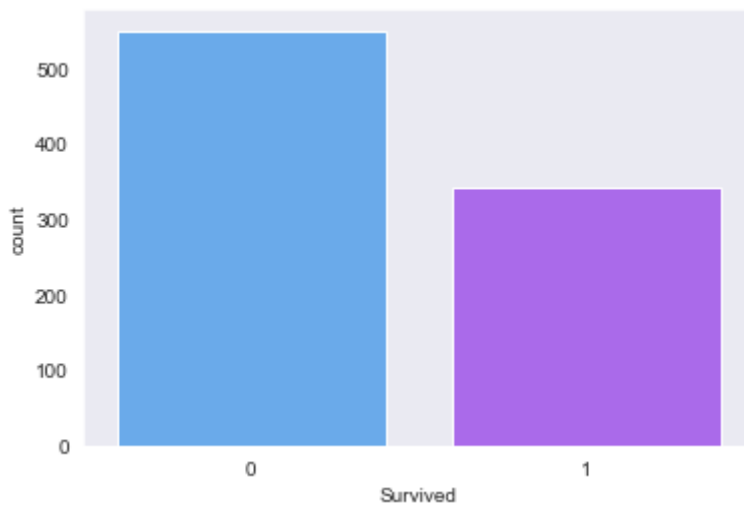
```
In [8]: sns.heatmap(sc.isnull())
```

```
Out[8]: <AxesSubplot:>
```



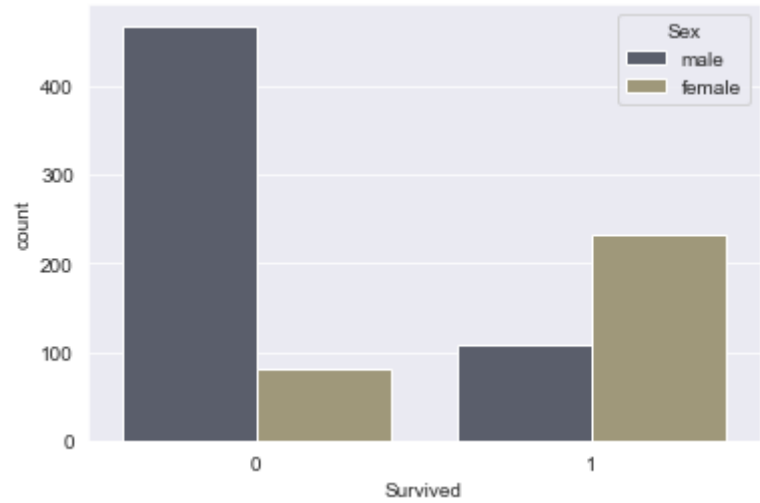
```
In [9]: #DATA VISUALIZATION
sns.set_style('dark')
sns.countplot(x='Survived', data=sc, palette='cool')
```

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



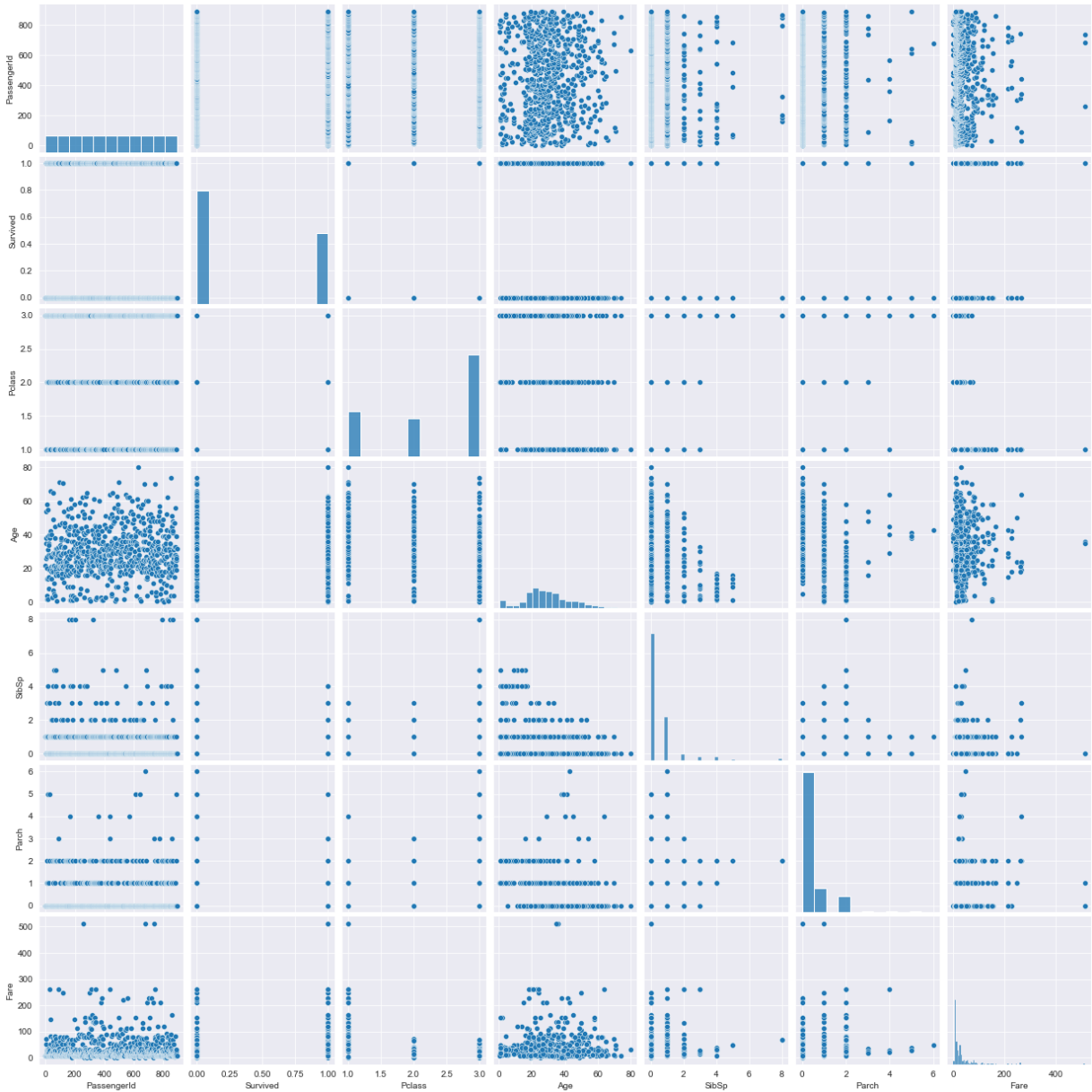
```
In [10]: #COUNT PLOT
sns.set_style('darkgrid')
sns.countplot(x='Survived', hue='Sex', data=sc, palette='cividis')
```

```
Out[10]: <AxesSubplot:xlabel='Survived', ylabel='count'>
```



```
In [11]: #PAIR PLOT
sns.pairplot(sc)
```

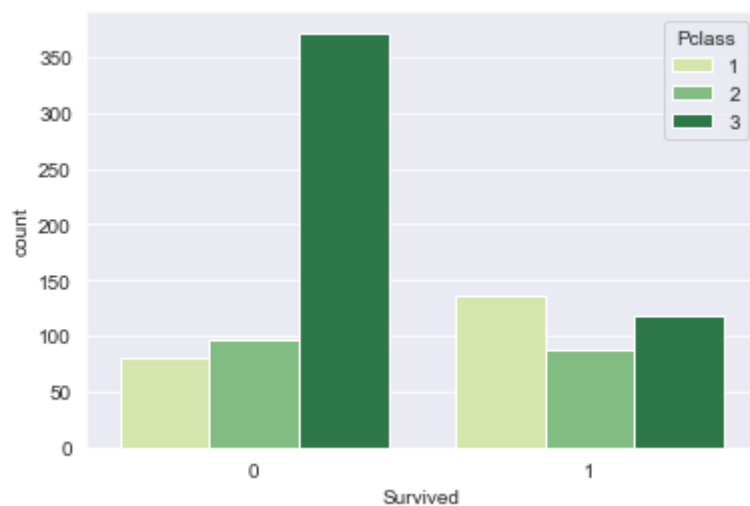
Out[11]: <seaborn.axisgrid.PairGrid at 0x25c34f4eb50>



```
In [12]: #COUNT PLOT
sns.set_style('darkgrid')
```

```
sns.countplot(x='Survived',hue='Pclass',data=sc,palette='YlGn')
```

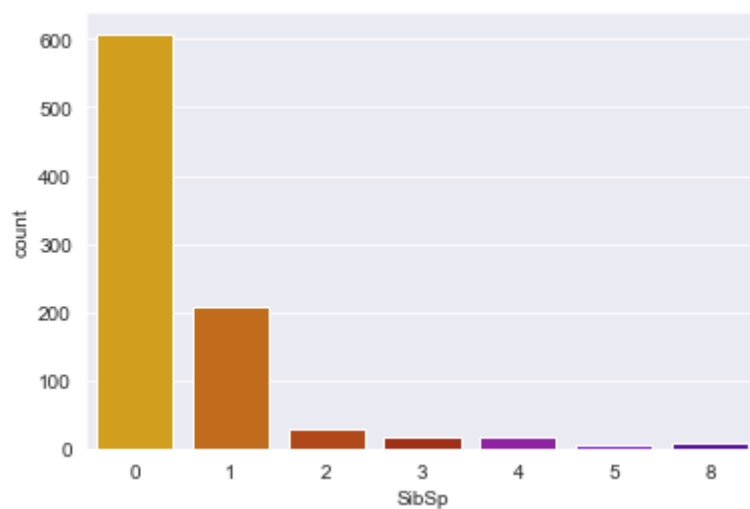
Out[12]: <AxesSubplot:xlabel='Survived', ylabel='count'>



In [13]:

```
#COUNT PLOT  
sns.countplot(x='SibSp',data=sc,palette='gnuplot_r')
```

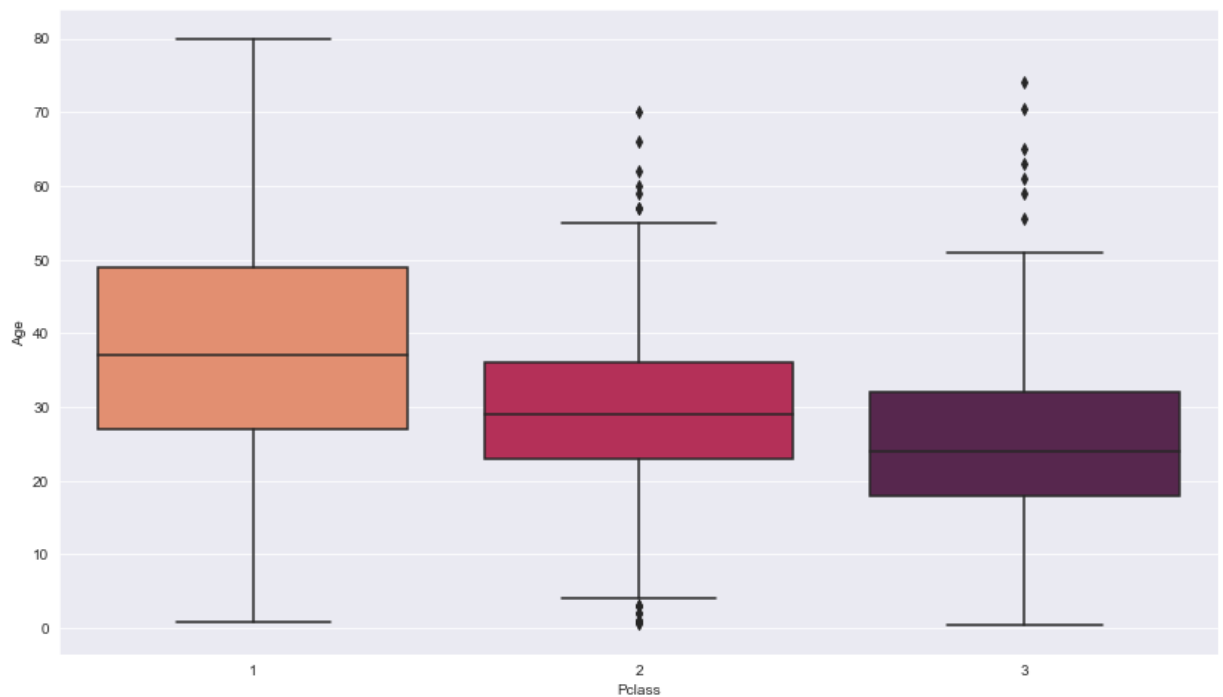
Out[13]: <AxesSubplot:xlabel='SibSp', ylabel='count'>



In [14]:

```
#DATA CLEANING  
#BOX PLOT  
plt.figure(figsize=(14,8))  
sns.boxplot(x='Pclass',y='Age',data=sc,palette='rocket_r')
```

Out[14]: <AxesSubplot:xlabel='Pclass', ylabel='Age'>

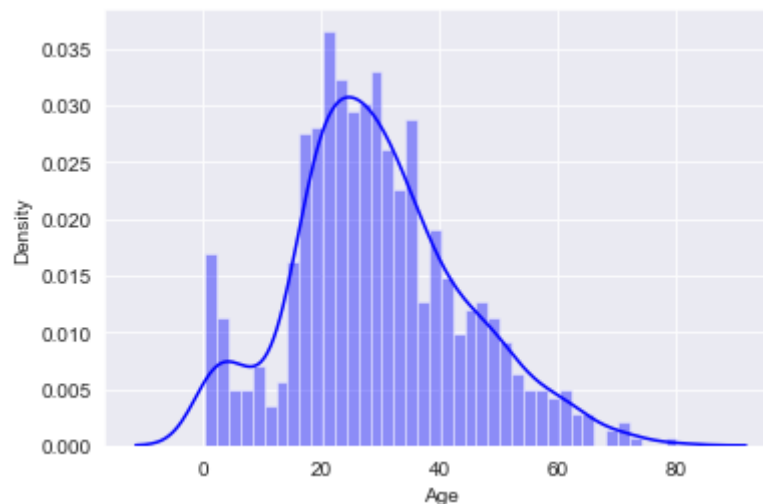


```
In [15]: #DIST PLOT
sns.distplot(sc['Age'].dropna(),color='blue',bins=40)
```

d:\Users\Micro\anaconda3\lib\site-packages\seaborn\distributions.py:2557: 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[15]: <AxesSubplot:xlabel='Age', ylabel='Density'>
```



```
In [16]: sc.drop('Cabin',axis=1,inplace=True)
```

```
In [17]: #FUNCTION
def null_age(cols):
    Age = cols[0]
    Pclass = cols[1]
    if pd.isnull(Age):
        if Pclass == 1:
            return 37
        elif Pclass == 2:
            return 29
        else:
```

```

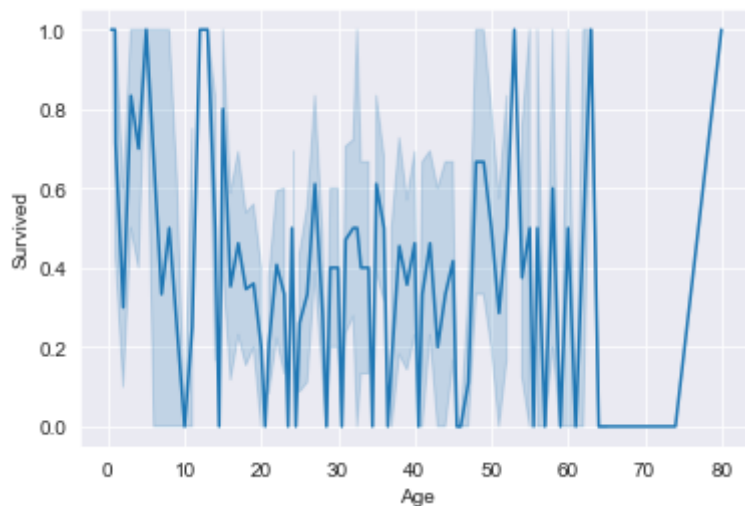
        return 24
    else:
        return Age

```

```

In [18]: #LINE PLOT
sns.lineplot(data=sc,x="Age",y="Survived")
plt.show()

```



```

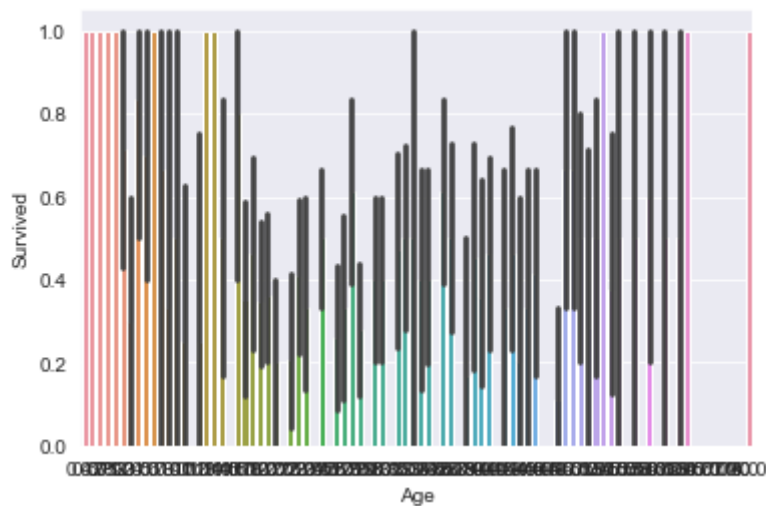
In [19]: #BAR PLOT
sns.barplot(data=sc,x="Age",y="Survived")

```

```

Out[19]: <AxesSubplot:xlabel='Age', ylabel='Survived'>

```



```

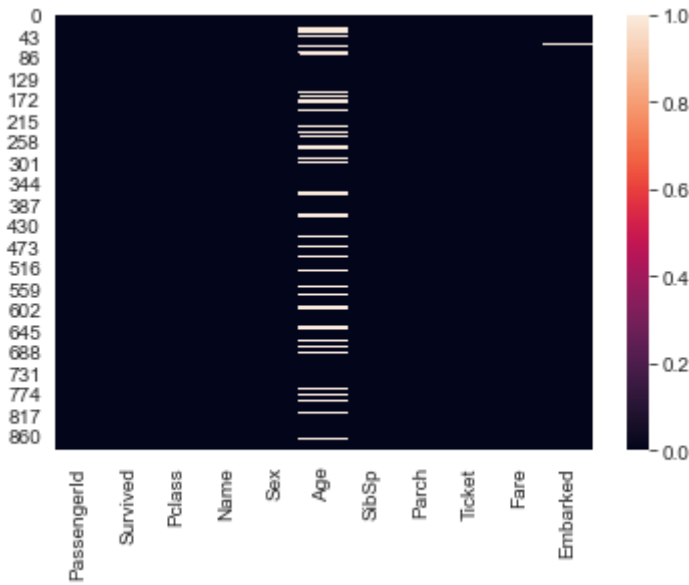
In [20]: #HEAT MAP (NULL VALUES)
sns.heatmap(sc.isnull())

```

```

Out[20]: <AxesSubplot:>

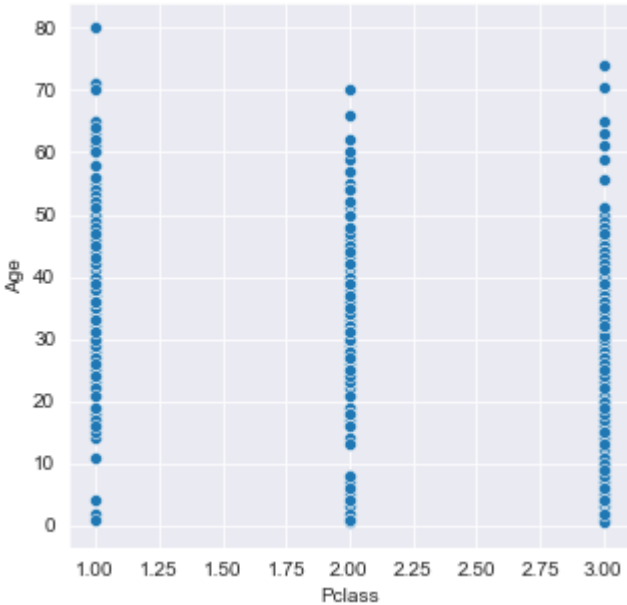
```



```
In [ ]:
```

```
In [21]: #SCATTER PLOT
plt.figure(figsize=(5,5))
sns.scatterplot(x='Pclass',y='Age',data=sc,palette='Dark2_r')
```

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



```
In [ ]:
```