In [4]:
```
#IMPORTING LIBRARIES
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:
```
go=pd.read_csv('train_and_test2.csv')
go.head(8)
```

Out[5]:

	Passengerid	Age	Fare	Sex	sibsp	zero	zero.1	zero.2	zero.3	zero.4	...	zero.12	zero.13
0	1	22.0	7.2500	0	1	0	0	0	0	0	...	0	0
1	2	38.0	71.2833	1	1	0	0	0	0	0	...	0	0
2	3	26.0	7.9250	1	0	0	0	0	0	0	...	0	0
3	4	35.0	53.1000	1	1	0	0	0	0	0	...	0	0
4	5	35.0	8.0500	0	0	0	0	0	0	0	...	0	0
5	6	28.0	8.4583	0	0	0	0	0	0	0	...	0	0
6	7	54.0	51.8625	0	0	0	0	0	0	0	...	0	0
7	8	2.0	21.0750	0	3	0	0	0	0	0	...	0	0

8 rows × 28 columns

In [6]:
```
go.isnull()
```

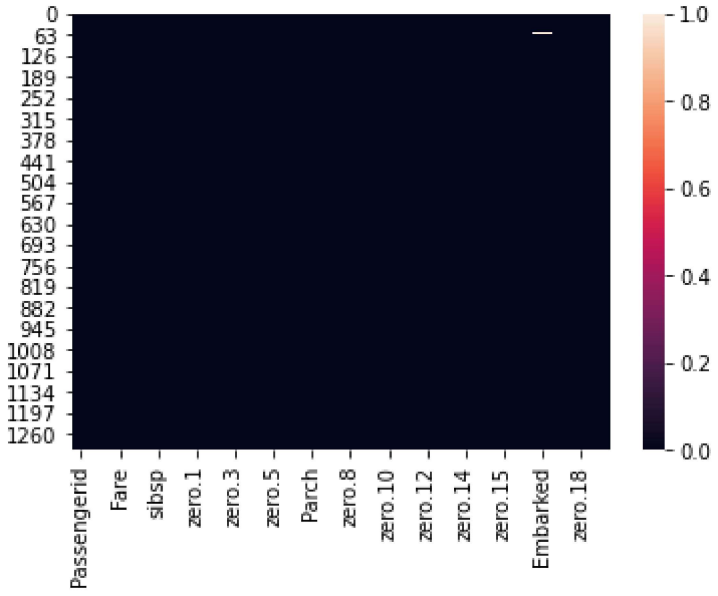
Out[6]:

	Passengerid	Age	Fare	Sex	sibsp	zero	zero.1	zero.2	zero.3	zero.4	...	zero.12	zero.
0	False	False	False	False	False	False	False	False	False	False	...	False	Fa
1	False	False	False	False	False	False	False	False	False	False	...	False	Fa
2	False	False	False	False	False	False	False	False	False	False	...	False	Fa
3	False	False	False	False	False	False	False	False	False	False	...	False	Fa
4	False	False	False	False	False	False	False	False	False	False	...	False	Fa
...	...	...	...	...	...	...	...	...	...	...	...	...	
1304	False	False	False	False	False	False	False	False	False	False	...	False	Fa
1305	False	False	False	False	False	False	False	False	False	False	...	False	Fa
1306	False	False	False	False	False	False	False	False	False	False	...	False	Fa
1307	False	False	False	False	False	False	False	False	False	False	...	False	Fa
1308	False	False	False	False	False	False	False	False	False	False	...	False	Fa

1309 rows × 28 columns

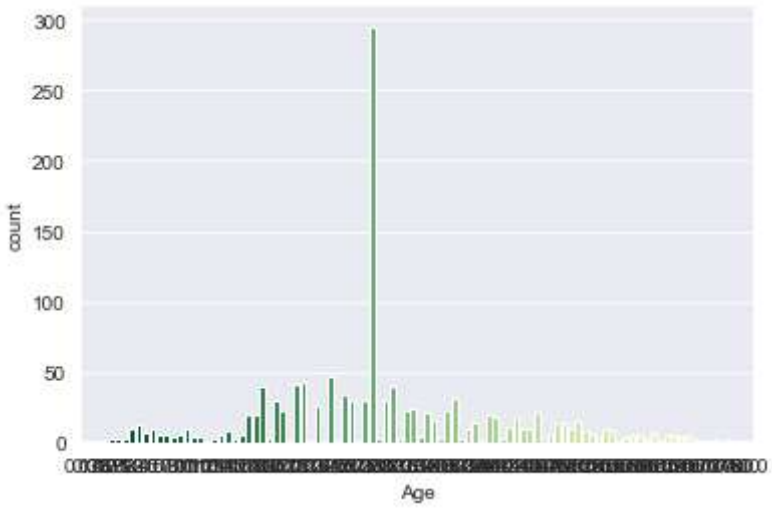
In [8]:
```
#HEAT MAPS
sns.heatmap(go.isnull())
```

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



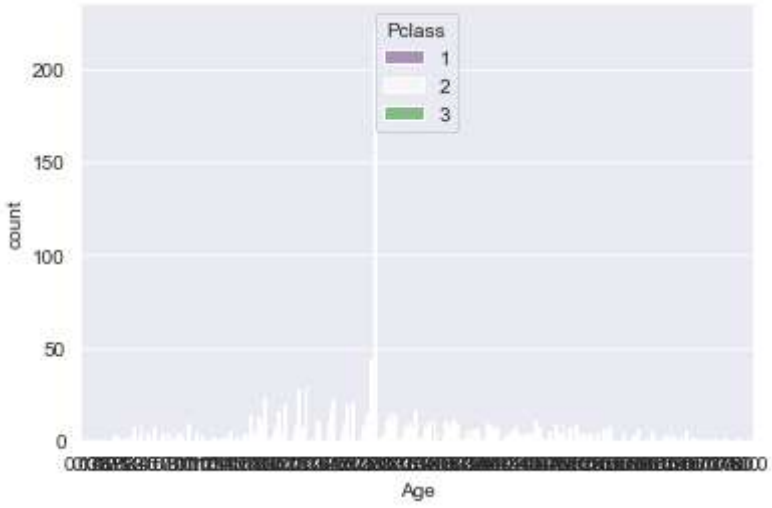
In [12]:
```
#COUNT PLOTS
sns.set_style('darkgrid')
sns.countplot(x='Age',data=go,palette='YlGn_r')
```

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



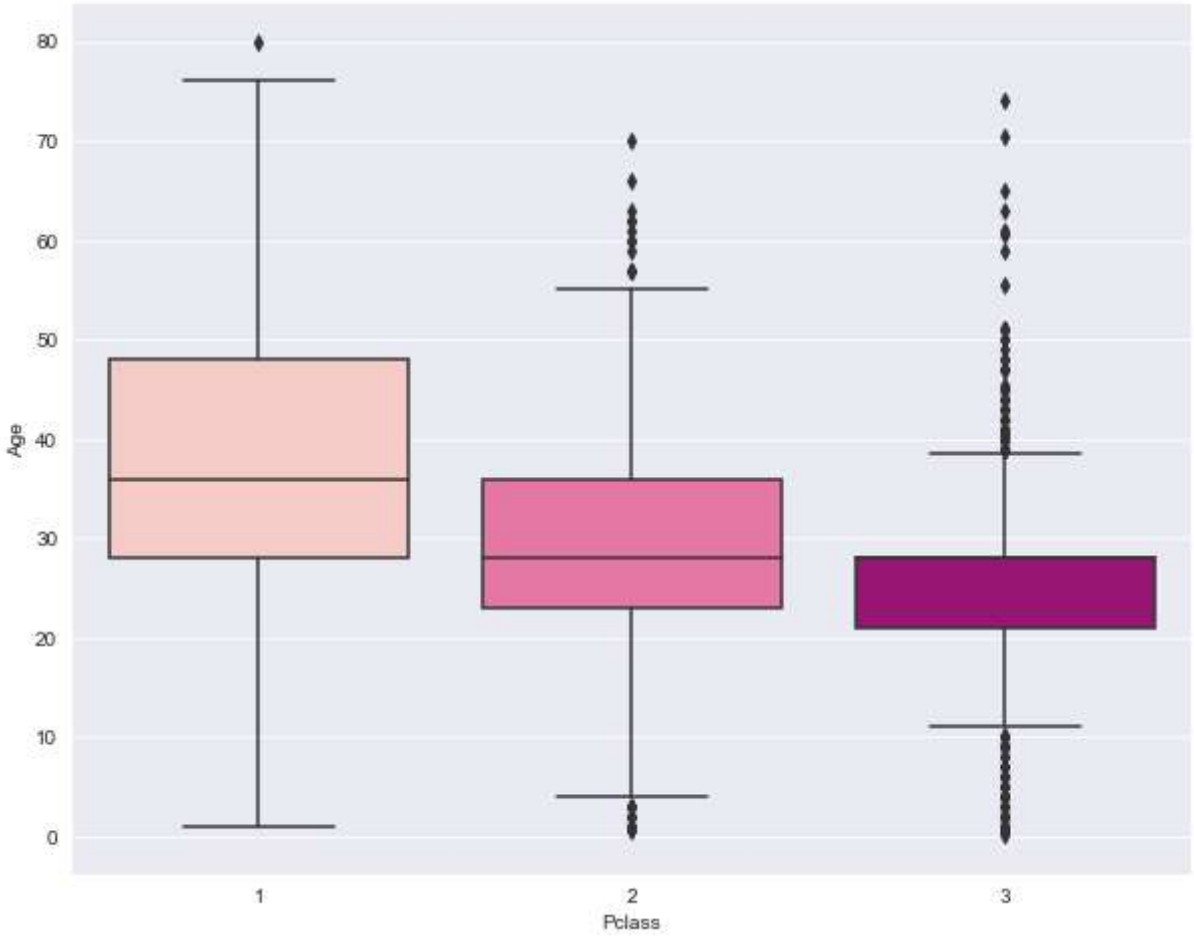
In [13]:
```
sns.set_style('darkgrid')
sns.countplot(x='Age',hue='Pclass',data=go,palette='PRGn')
```

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



In [14]:
```
#BOX PLOT
plt.figure(figsize=(10, 8))
sns.boxplot(x='Pclass',y='Age',data=go,palette='RdPu')
```

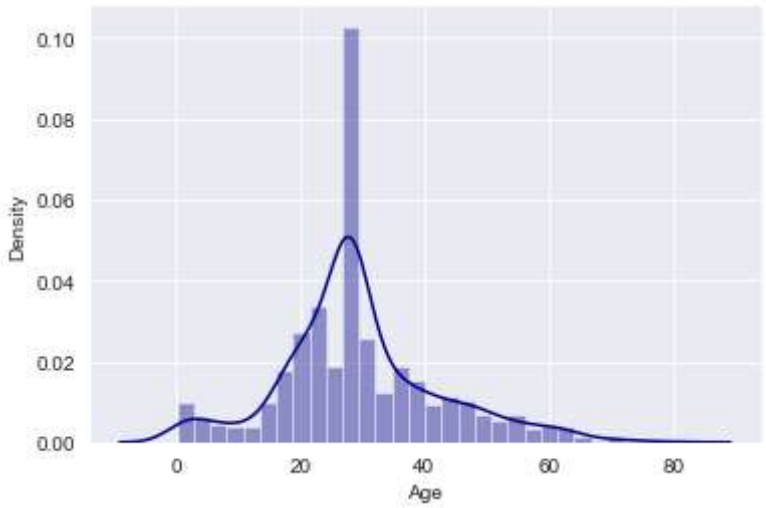
Out[14]: <AxesSubplot:xlabel='Pclass', ylabel='Age'>



In [15]:
```
#DIST PLOT
sns.distplot(go['Age'].dropna(),color='darkblue',bins=30)
```

```
C:\Users\panka\anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarn
ing: `distplot` is a deprecated function and will be removed in a future version. Pl
ease adapt your code to use either `displot` (a figure-level function with similar f
lexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

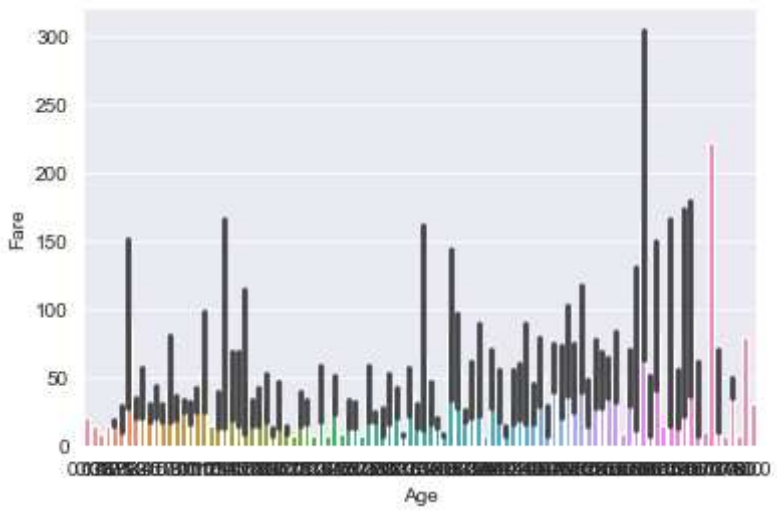
Out[15]: <AxesSubplot:xlabel='Age', ylabel='Density'>



In [18]:
```
#BAR PLOT
```

```
sns.barplot(data=go,x="Age",y="Fare")
```

Out[18]: <AxesSubplot:xlabel='Age', ylabel='Fare'>

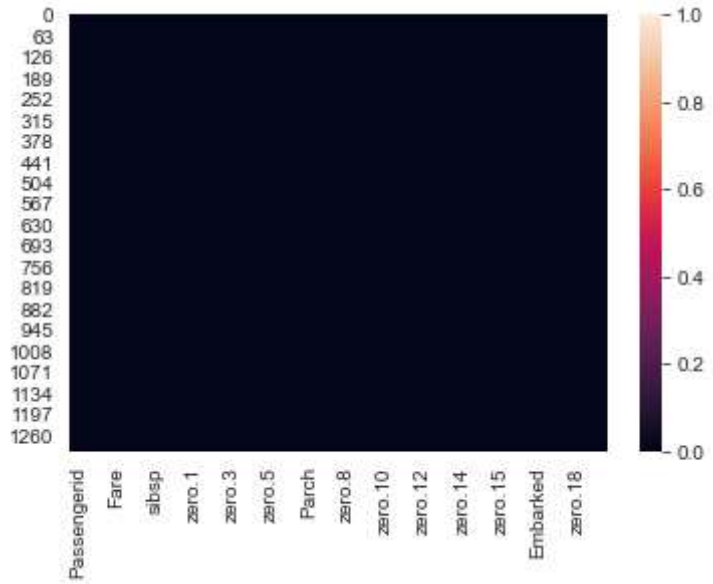


In [19]:
```
#HEATMAP(EXCLUDING NAN VALUES)
def null_age(cols):
    Age = cols[0]
    Pclass = cols[1]

    if pd.isnull(Age):

        if Pclass == 1:
            return 68

        elif Pclass == 2:
            return 46

        else:
            return 32

    else:
        return Age
```

In [20]:
```
go['Age'] = go[['Age','Pclass']].apply(null_age,axis=1)
```

In [21]:
```
sns.heatmap(go.isnull())
```

Out[21]: <AxesSubplot:>



In [22]:
```
#cleaning dataset
go.head(88)
```

Out[22]:

	Passengerid	Age	Fare	Sex	sibsp	zero	zero.1	zero.2	zero.3	zero.4	...	zero.12	zero.1
0	1	22.0	7.2500	0	1	0	0	0	0	0	...	0	
1	2	38.0	71.2833	1	1	0	0	0	0	0	...	0	
2	3	26.0	7.9250	1	0	0	0	0	0	0	...	0	
3	4	35.0	53.1000	1	1	0	0	0	0	0	...	0	
4	5	35.0	8.0500	0	0	0	0	0	0	0	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
83	84	28.0	47.1000	0	0	0	0	0	0	0	...	0	
84	85	17.0	10.5000	1	0	0	0	0	0	0	...	0	
85	86	33.0	15.8500	1	3	0	0	0	0	0	...	0	
86	87	16.0	34.3750	0	1	0	0	0	0	0	...	0	
87	88	28.0	8.0500	0	0	0	0	0	0	0	...	0	

88 rows × 28 columns