In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:
```
import pandas as pd
df = pd.read_csv('train.csv')
print(df)
```

```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           ...       ...     ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex   Age  SibSp  \
0                              Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                               Heikkinen, Miss. Laina  female  26.0      0
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                             Allen, Mr. William Henry    male  35.0      0
..                                                 ...     ...   ...    ...
886                              Montvila, Rev. Juozas    male  27.0      0
887                       Graham, Miss. Margaret Edith  female  19.0      0
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                              Behr, Mr. Karl Howell    male  26.0      0
890                                Dooley, Mr. Patrick    male  32.0      0

     Parch            Ticket     Fare Cabin Embarked
0        0         A/5 21171   7.2500   NaN        S
1        0          PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250   NaN        S
3        0            113803  53.1000  C123        S
4        0            373450   8.0500   NaN        S
..     ...               ...      ...   ...      ...
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

In [3]:
```
df.head()
```

Out[3]:

	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 [4]:
```
df.tail()
```

Out[4]:

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

In [5]:
```
df.shape
```

Out[5]:
```
(891, 12)
```

In [6]:
```
df.info()
```

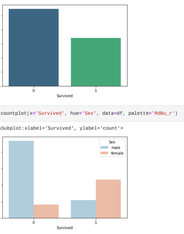
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 66.2+ KB
```

In [7]:
```
sns.heatmap(df.isnull(), yticklabels=False, cbar=False, cmap="viridis")
```

Out[7]:
```
<AxesSubplot:>
```

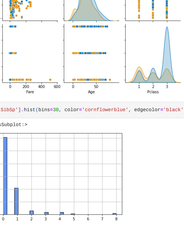
In [8]:
```
#correlation matrix--the darker columns are more related to the output column
nfsdf.select_dtypes(include=[np.number])
corrmat.corr()
sns.heatmap(corr, vmax=.8, annot_kws={'size':10}, annot=True)
```

Out[9]:
```
<AxesSubplot:>
```



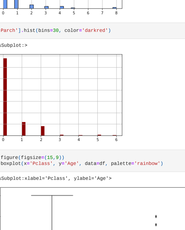
In [10]:
```
sns.countplot(x='Survived', hue='Pclass', data=df)
```

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



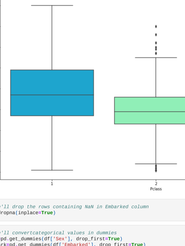
In [11]:
```
df['Age'].hist(bins=30, color='lightseagreen', edgecolor='green')
```

Out[11]:
```
<AxesSubplot:>
```



In [12]:
```
sns.countplot(x='Survived', data=df, palette='viridis')
```

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



In [13]:
```
sns.countplot(x='Survived', hue='Sex', data=df, palette='RdBu_r')
```

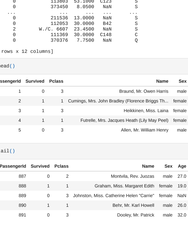
Out[13]:
```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

In [14]:
```
plt.figure(figsize=(12,10))
sns.pairplot(df, vars=['Fare', 'Age', 'Pclass'], hue='Survived', palette='colorblind')
```

Out[14]:
```
<seaborn.axisgrid.PairGrid at 0xaa88580>
```


In [15]:
```
df['SibSp'].hist(bins=30, color='cornflowerblue', edgecolor='black')
```

Out[15]:
```
<AxesSubplot:>
```



In [16]:
```
df['Parch'].hist(bins=30, color='darkred')
```

Out[16]:
```
<AxesSubplot:>
```

In [17]:
```
plt.figure(figsize=(15,9))
sns.boxplot(x='Pclass', y='Age', data=df, palette='rainbow')
```

Out[17]:
```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```


In [18]:
```
# we'll drop the rows containing NaN in Embarked column
df.dropna(inplace=True)
```

In [19]:
```
# we'll convert categorical values in dummies
sex=pd.get_dummies(df['Sex'], drop_first=True)
embark=pd.get_dummies(df['Embarked'], drop_first=True)
```

In [20]:
```
# we'll drop foll columns
df.drop(['Sex', 'Embarked', 'Name', 'Ticket'], axis=1, inplace=True)
df=pd.concat([df,sex,embark], axis=1)
```

In [21]:
```
df.head()
```

Out[21]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	male	Q	S
1	2	1	1	38.0	1	0	71.2833	C85	0	0	0
3	4	1	1	35.0	1	0	53.1000	C123	0	0	1
6	7	0	1	54.0	0	0	51.8625	E46	1	0	1
10	11	1	3	4.0	1	1	16.7000	G6	0	0	1
11	12	1	1	58.0	0	0	26.5500	C103	0	0	1

In [ ]:
```

```