```
import pandas as pd
import  numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import io
```

```
from google.colab import files
uploaded = files.upload()
```

Choose Files titanic.csv

- **titanic.csv**(text/csv) - 29474 bytes, last modified: 12/2/2021 - 100% done

```
Saving titanic.csv to titanic.csv
```

```
data = pd.read_csv('titanic.csv')
```

```
data.sample(5)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
358	1250	0	3	O'Keefe, Mr. Patrick	male	NaN	0	0	368402	7.7500	NaN	Q
128	1020	0	2	Bowenur, Mr. Solomon	male	42.0	0	0	211535	13.0000	NaN	S
171	1063	0	3	Zakarian, Mr. Ortin	male	27.0	0	0	2670	7.2250	NaN	C
166	1058	0	1	Brandeis, Mr. Emil	male	48.0	0	0	PC 17591	50.4958	B10	C

```
data.info(), data.isna().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          332 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         417 non-null    float64
 10  Cabin        91 non-null     object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
(None,
 PassengerId      0
 Survived         0
 Pclass           0
 Name             0
 Sex              0
 Age             86
 SibSp            0
 Parch            0
 Ticket           0
 Fare             1
 Cabin          327
 Embarked         0
 dtype: int64)
```

```
data['Age'] = data['Age'].fillna(np.mean(data['Age']))
```

```
data['Cabin'] = data['Cabin'].fillna(data['Cabin'].mode()[0])
```

```
data['Embarked'] = data['Embarked'].fillna(data['Embarked'].mode()[0])
```

```
data.isnull().sum()
```

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
```

```
Parch          0
Ticket         0
Fare           1
Cabin          0
Embarked       0
dtype: int64
```

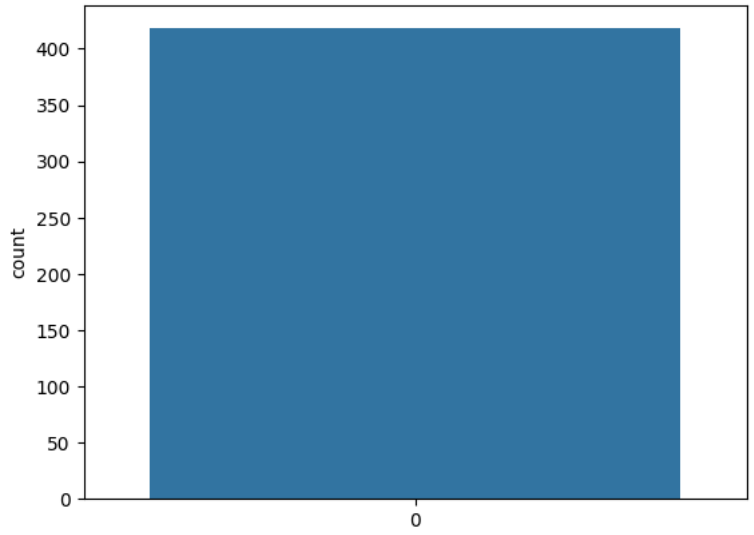
```
data['Fare'] = data['Fare'].fillna(data['Fare'].mode()[0])
```

```
data.isnull().sum()
```

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

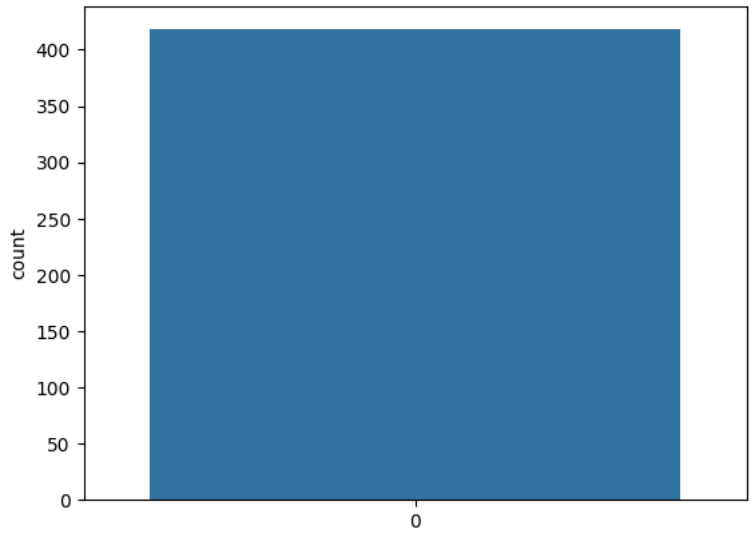
```
sns.countplot(data['Survived'])
```

```
<Axes: ylabel='count'>
```



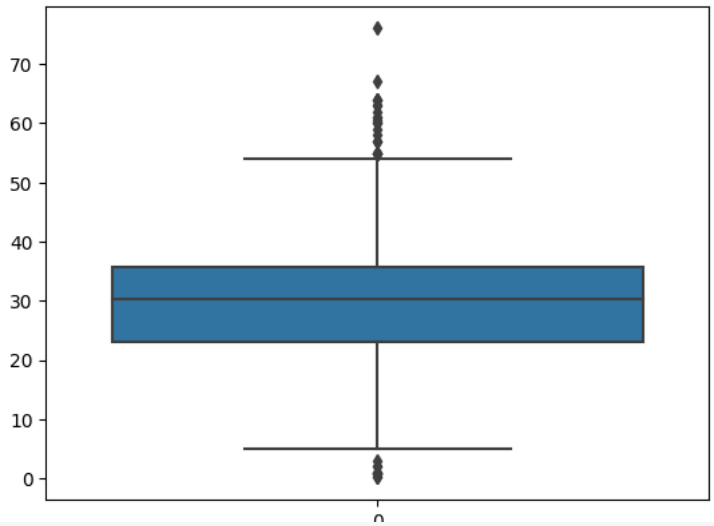

```
sns.countplot(data['Pclass'])
```

```
<Axes: ylabel='count'>
```



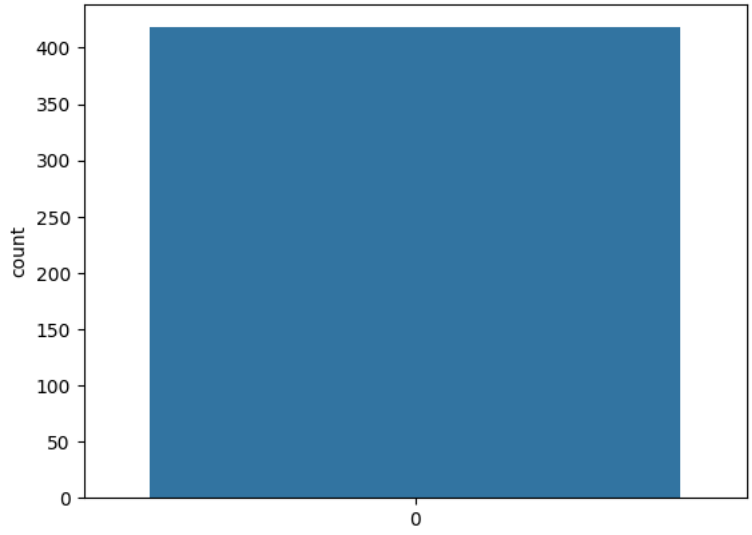

```
sns.boxplot(data['Age'])
```

```
<Axes: >
```



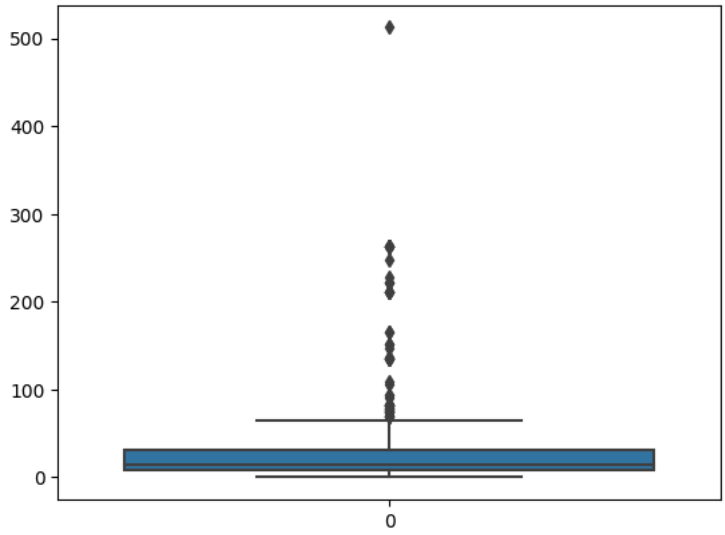

```
sns.countplot(data['Fare'])
```

```
<Axes: ylabel='count'>
```



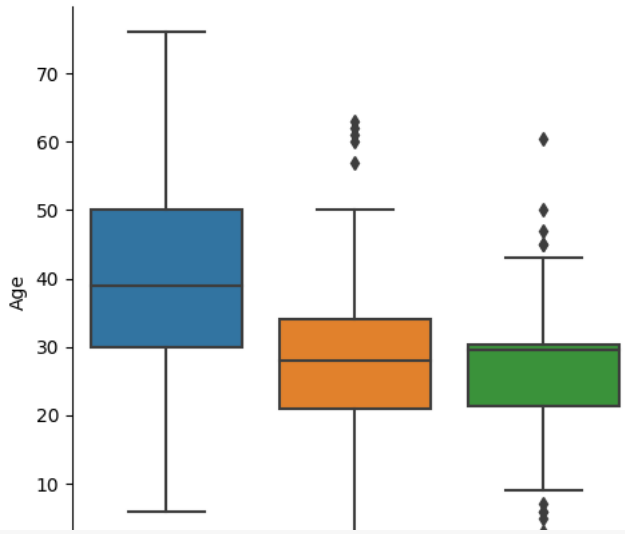

```
sns.boxplot(data['Fare'])
```

```
<Axes: >
```



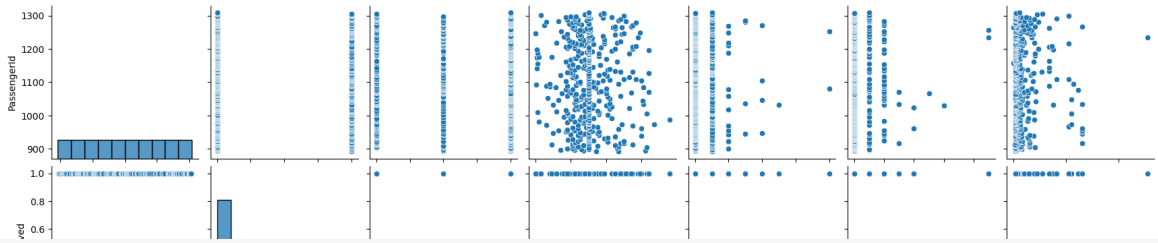

```
sns.catplot(x= 'Pclass', y = 'Age', data=data, kind = 'box')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f9cb053fa30>
```



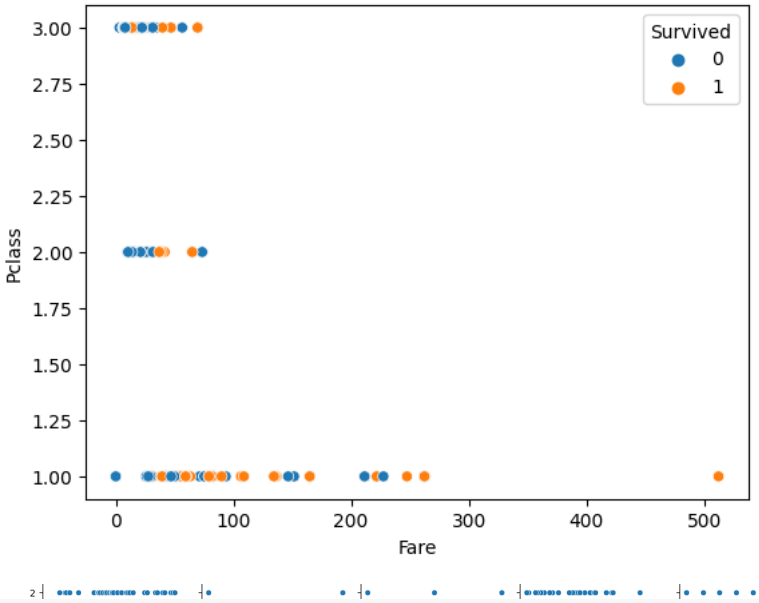

```
sns.pairplot(data)
```

```
<seaborn.axisgrid.PairGrid at 0x7f9cb053f790>
```



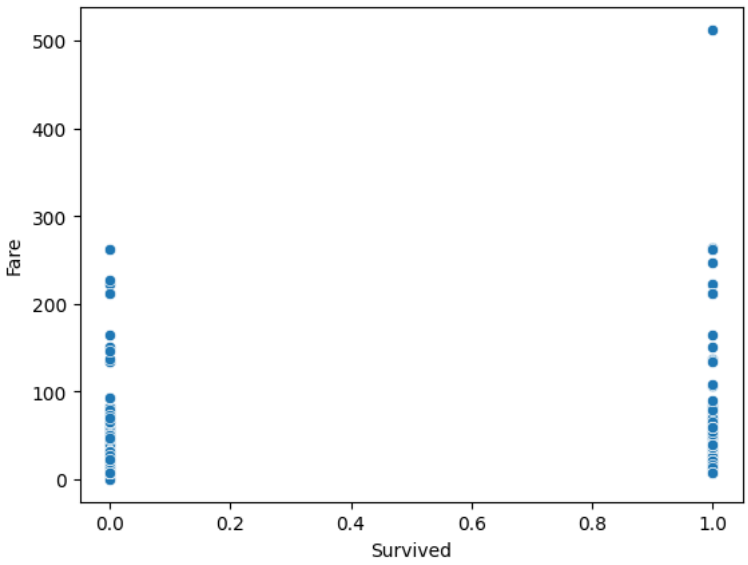

```
sns.scatterplot(x = 'Fare', y = 'Pclass', hue = 'Survived', data = data)
```

```
<Axes: xlabel='Fare', ylabel='Pclass'>
```



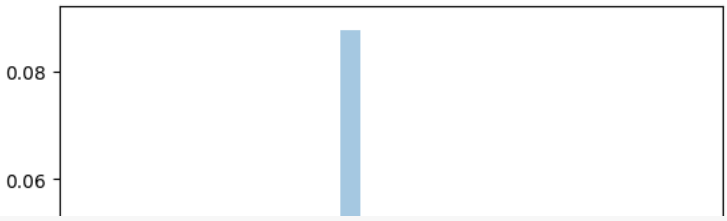

```
sns.scatterplot(x = 'Survived', y = 'Fare', data = data)
```

```
<Axes: xlabel='Survived', ylabel='Fare'>
```



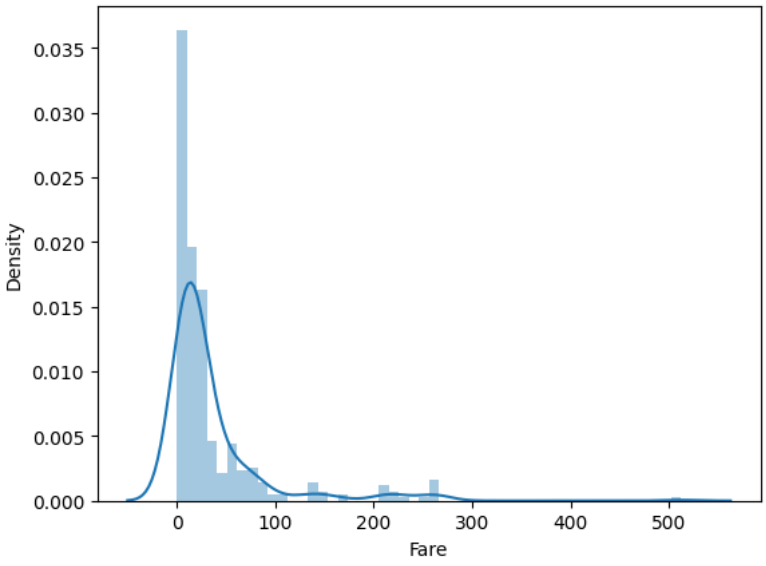

```
sns.distplot(data['Age'])
```

```
<ipython-input-24-ec135ce2bd5e>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(data['Age'])
<Axes: xlabel='Age', ylabel='Density'>
```



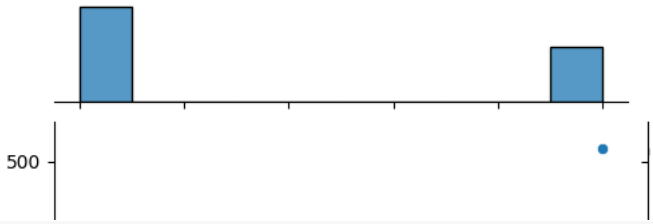

```
sns.distplot(data['Fare'])
```

```
<ipython-input-25-1c4203d5582e>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(data['Fare'])
<Axes: xlabel='Fare', ylabel='Density'>
```



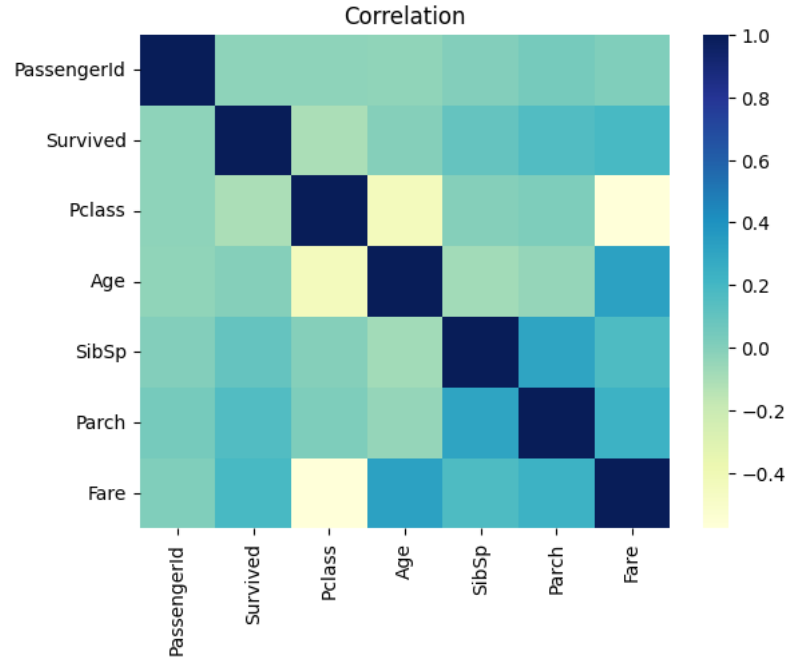

```
sns.jointplot(x = "Survived", y = "Fare", kind = "scatter", data = data)
```

```
<seaborn.axisgrid.JointGrid at 0x7f9ca998ab60>
```



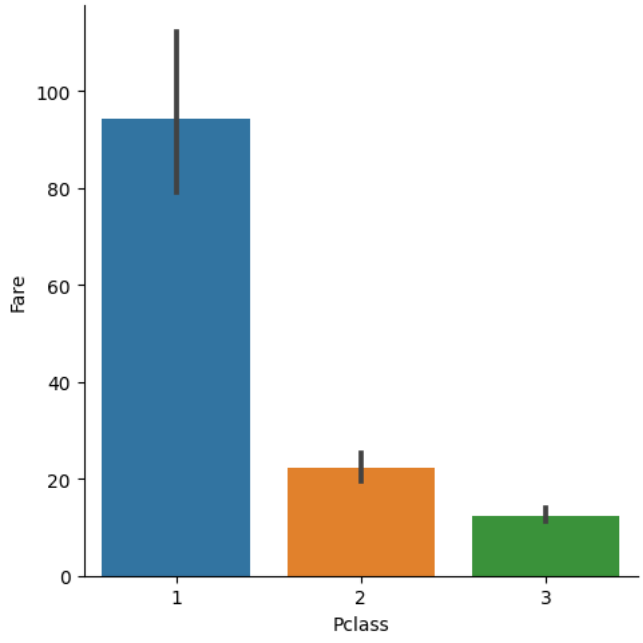

```
tc = data.corr()
sns.heatmap(tc, cmap="YlGnBu")
plt.title('Correlation')
```

```
<ipython-input-27-54deec0cf45b>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecat
  tc = data.corr()
Text(0.5, 1.0, 'Correlation')
```




```
#price of ticket for each pasenger is distributed
sns.catplot(x='Pclass', y='Fare', data=data, kind='bar')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f9cae48d480>
```



✓ 0s completed at 3:26 PM