

```
!pip install matplotlib
```

Requirement already satisfied: matplotlib in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (0.1.9)

Requirement already satisfied: pyloco>=0.0.134 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (0.0.139)

Requirement already satisfied: matplotlib>=3.1.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (3.7.3)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (1.1.0)

Requirement already satisfied: cyclopy>=0.10 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (0.11.0)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (4.22.1)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (1.4.5)

Requirement already satisfied: numpy<2,>=1.20 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (1.25.2)

Requirement already satisfied: packaging>=20.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (23.1)

Requirement already satisfied: pillow>=6.2.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (10.0.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (3.1.1)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib>=3.1.1->matplotlib) (2.8.2)

Requirement already satisfied: ushlex in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pyloco>=0.0.134->matplotlib) (0.99.1)

Requirement already satisfied: websocket-client in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pyloco>=0.0.134->matplotlib) (1.6.2)

Requirement already satisfied: twine in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pyloco>=0.0.134->matplotlib) (4.0.2)

Requirement already satisfied: typing in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pyloco>=0.0.134->matplotlib) (3.7.4.3)

Requirement already satisfied: SimpleWebSocketServer in c:\users\

ilove\appdata\local\programs\python\python311\lib\site-packages (from pyloco>=0.0.134->matplotlib) (0.1.2)  
Requirement already satisfied: six>=1.5 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.7->matplotlib>=3.1.1->matplotlib) (1.16.0)  
Requirement already satisfied: pkginfo>=1.8.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (1.9.6)  
Requirement already satisfied: readme-renderer>=35.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (42.0)  
Requirement already satisfied: requests>=2.20 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (2.31.0)  
Requirement already satisfied: requests-toolbelt!=0.9.0,>=0.8.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (1.0.0)  
Requirement already satisfied: urllib3>=1.26.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (2.0.4)  
Requirement already satisfied: importlib-metadata>=3.6 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (6.8.0)  
Requirement already satisfied: keyring>=15.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (24.2.0)  
Requirement already satisfied: rfc3986>=1.4.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (2.0.0)  
Requirement already satisfied: rich>=12.0.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from twine->pyloco>=0.0.134->matplotlib) (13.5.2)  
Requirement already satisfied: zipp>=0.5 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from importlib-metadata>=3.6->twine->pyloco>=0.0.134->matplotlib) (3.16.2)  
Requirement already satisfied: jaraco.classes in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from keyring>=15.1->twine->pyloco>=0.0.134->matplotlib) (3.3.0)  
Requirement already satisfied: pywin32-ctypes>=0.2.0 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from keyring>=15.1->twine->pyloco>=0.0.134->matplotlib) (0.2.2)  
Requirement already satisfied: nh3>=0.2.14 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from readme-renderer>=35.0->twine->pyloco>=0.0.134->matplotlib) (0.2.14)  
Requirement already satisfied: docutils>=0.13.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from readme-renderer>=35.0->twine->pyloco>=0.0.134->matplotlib) (0.20.1)  
Requirement already satisfied: Pygments>=2.5.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from

```
readme-renderer>=35.0->twine->pyloco>=0.0.134->matplotlib) (2.16.1)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\
ilove\appdata\local\programs\python\python311\lib\site-packages (from
requests>=2.20->twine->pyloco>=0.0.134->matplotlib) (3.2.0)
Requirement already satisfied: idna<4,>=2.5 in c:\users\ilove\appdata\
local\programs\python\python311\lib\site-packages (from
requests>=2.20->twine->pyloco>=0.0.134->matplotlib) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
requests>=2.20->twine->pyloco>=0.0.134->matplotlib) (2023.7.22)
Requirement already satisfied: markdown-it-py>=2.2.0 in c:\users\
ilove\appdata\local\programs\python\python311\lib\site-packages (from
rich>=12.0.0->twine->pyloco>=0.0.134->matplotlib) (3.0.0)
Requirement already satisfied: mdurl~=0.1 in c:\users\ilove\appdata\
local\programs\python\python311\lib\site-packages (from markdown-it-
py>=2.2.0->rich>=12.0.0->twine->pyloco>=0.0.134->matplotlib) (0.1.2)
Requirement already satisfied: more-itertools in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
jaraco.classes->keyring>=15.1->twine->pyloco>=0.0.134->matplotlib)
(10.1.0)
```

```
!pip install scikit-learn
```

```
Collecting scikit-learn
```

```
Obtaining dependency information for scikit-learn from
https://files.pythonhosted.org/packages/77/85/bff3a1e818ec6aa3dd466ff4
f4b0a727db9fdb41f2e849747ad902ddbe95/scikit_learn-1.3.0-cp311-cp311-
win_amd64.whl.metadata
```

```
Downloading scikit_learn-1.3.0-cp311-cp311-win_amd64.whl.metadata
(11 kB)
```

```
Requirement already satisfied: numpy>=1.17.3 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
scikit-learn) (1.25.2)
```

```
Requirement already satisfied: scipy>=1.5.0 in c:\users\ilove\appdata\
local\programs\python\python311\lib\site-packages (from scikit-learn)
(1.11.2)
```

```
Requirement already satisfied: joblib>=1.1.1 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
scikit-learn) (1.1.1)
```

```
Collecting threadpoolctl>=2.0.0 (from scikit-learn)
```

```
Obtaining dependency information for threadpoolctl>=2.0.0 from
https://files.pythonhosted.org/packages/81/12/fd4dea011af9d69e1cad05c7
5f3f7202cdcbeac9b712eea58ca779a72865/threadpoolctl-3.2.0-py3-none-
any.whl.metadata
```

```
Downloading threadpoolctl-3.2.0-py3-none-any.whl.metadata (10.0 kB)
```

```
Downloading scikit_learn-1.3.0-cp311-cp311-win_amd64.whl (9.2 MB)
```

```
----- 0.0/9.2 MB ? eta -:-:--
----- 0.2/9.2 MB 3.5 MB/s eta
```

```
0:00:03
```

```
-- ----- 0.5/9.2 MB 5.7 MB/s eta
```

```

0:00:02
----- 0.9/9.2 MB 6.6 MB/s eta
0:00:02
----- 1.6/9.2 MB 8.4 MB/s eta
0:00:01
----- 2.0/9.2 MB 8.3 MB/s eta
0:00:01
----- 2.5/9.2 MB 8.9 MB/s eta
0:00:01
----- 3.1/9.2 MB 9.3 MB/s eta
0:00:01
----- 3.3/9.2 MB 8.9 MB/s eta
0:00:01
----- 4.0/9.2 MB 9.0 MB/s eta
0:00:01
----- 4.4/9.2 MB 9.1 MB/s eta
0:00:01
----- 4.8/9.2 MB 9.3 MB/s eta
0:00:01
----- 5.2/9.2 MB 8.9 MB/s eta
0:00:01
----- 5.6/9.2 MB 9.1 MB/s eta
0:00:01
----- 6.0/9.2 MB 8.9 MB/s eta
0:00:01
----- 6.5/9.2 MB 9.2 MB/s eta
0:00:01
----- 6.9/9.2 MB 9.2 MB/s eta
0:00:01
----- 7.5/9.2 MB 9.2 MB/s eta
0:00:01
----- 7.8/9.2 MB 9.1 MB/s eta
0:00:01
----- 8.3/9.2 MB 9.1 MB/s eta
0:00:01
----- 8.8/9.2 MB 9.3 MB/s eta
0:00:01
----- 9.0/9.2 MB 9.2 MB/s eta
0:00:01
----- 9.2/9.2 MB 9.0 MB/s eta
0:00:01
----- 9.2/9.2 MB 8.4 MB/s eta
0:00:00

```

```

Downloading threadpoolctl-3.2.0-py3-none-any.whl (15 kB)
Installing collected packages: threadpoolctl, scikit-learn
Successfully installed scikit-learn-1.3.0 threadpoolctl-3.2.0

```

```

import pandas as pd
import numpy as np
import seaborn as sns

```

```
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
```

## loading the dataset

```
train = pd.read_csv('train.csv')
```

```
test = pd.read_csv('test.csv')
```

```
train.head()
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

		Name	Sex	Age
SibSp	\			
0		Braund, Mr. Owen Harris	male	22.0
1				
1	Cummings, 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				

	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

```
# statistical info
train.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	\
count	891.000000	891.000000	891.000000	714.000000	891.000000	
mean	446.000000	0.383838	2.308642	29.699118	0.523008	
std	257.353842	0.486592	0.836071	14.526497	1.102743	
min	1.000000	0.000000	1.000000	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	20.125000	0.000000	

50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

```
# datatype info
```

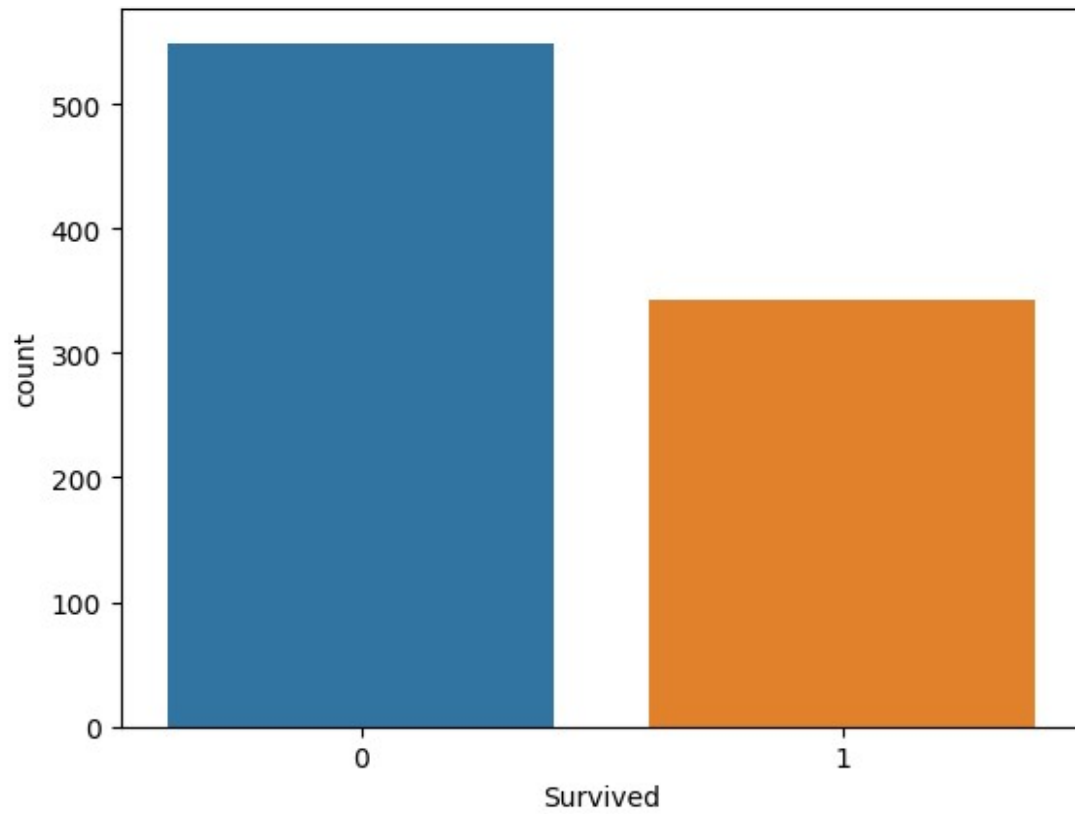
```
train.info()
```

```
<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: 83.7+ KB
```

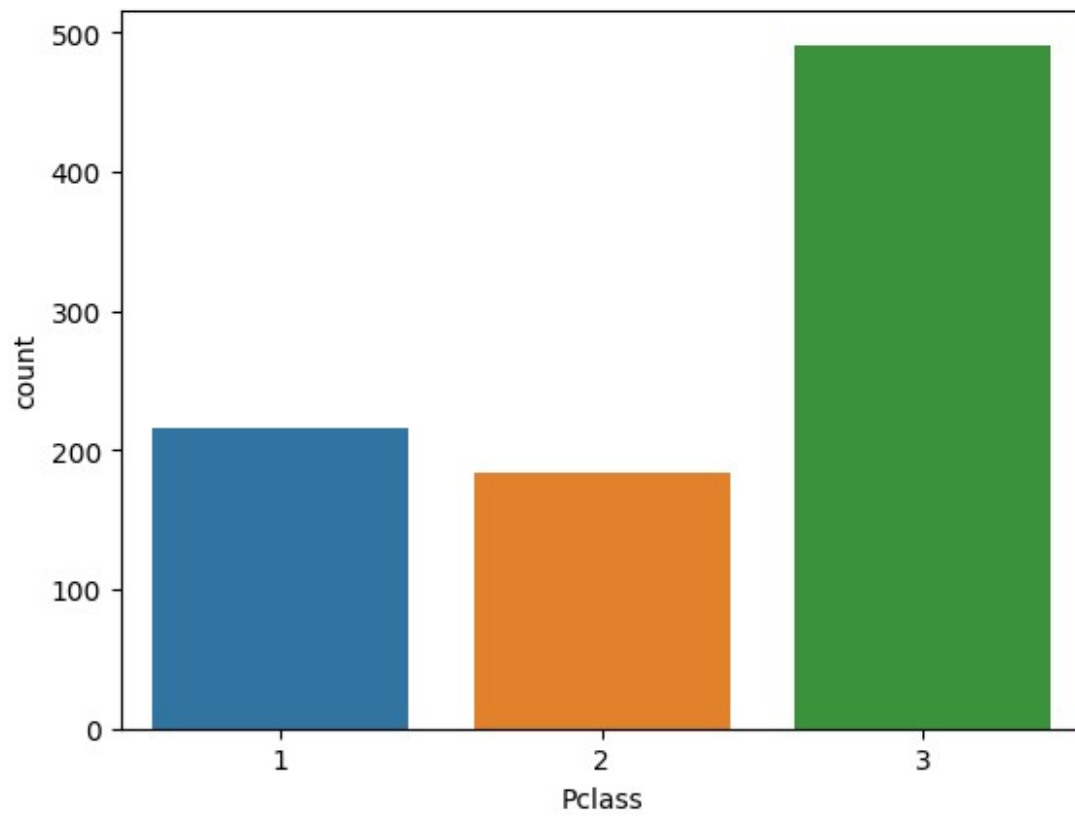
## exploratory data analysis

```
sns.countplot(x = 'Survived', data = train)
```

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

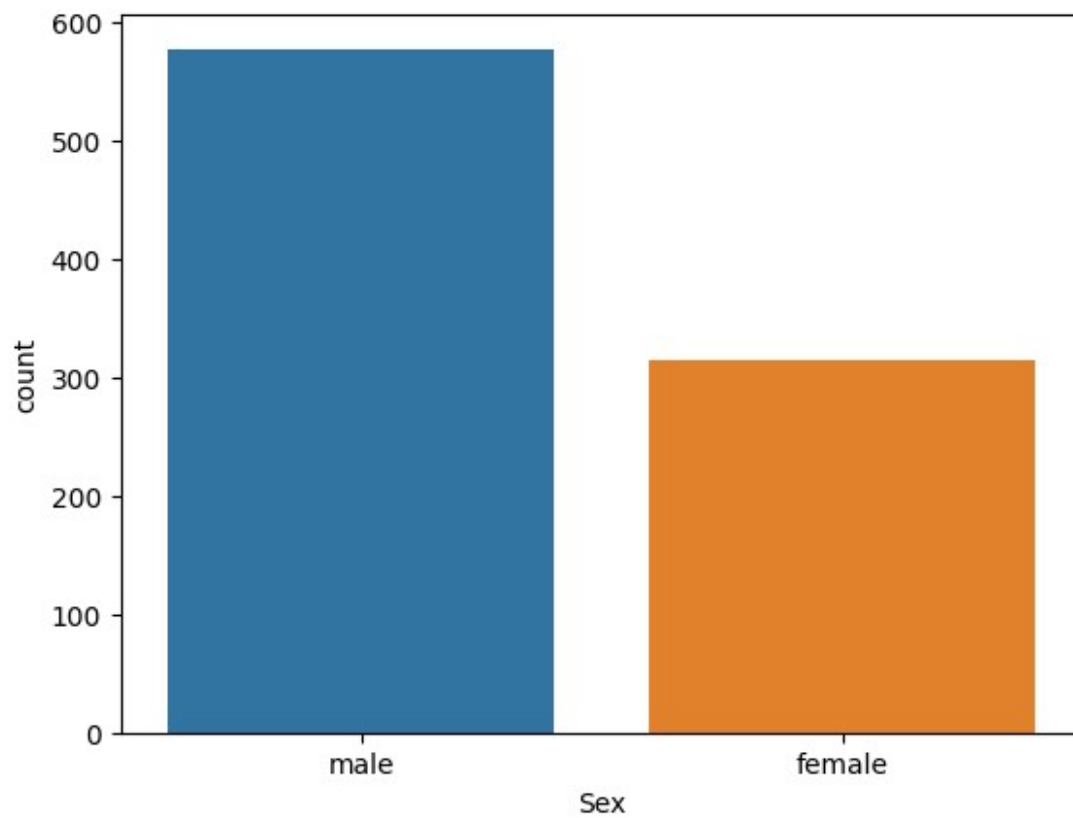


```
sns.countplot(x = 'Pclass', data = train)  
<Axes: xlabel='Pclass', ylabel='count'>
```

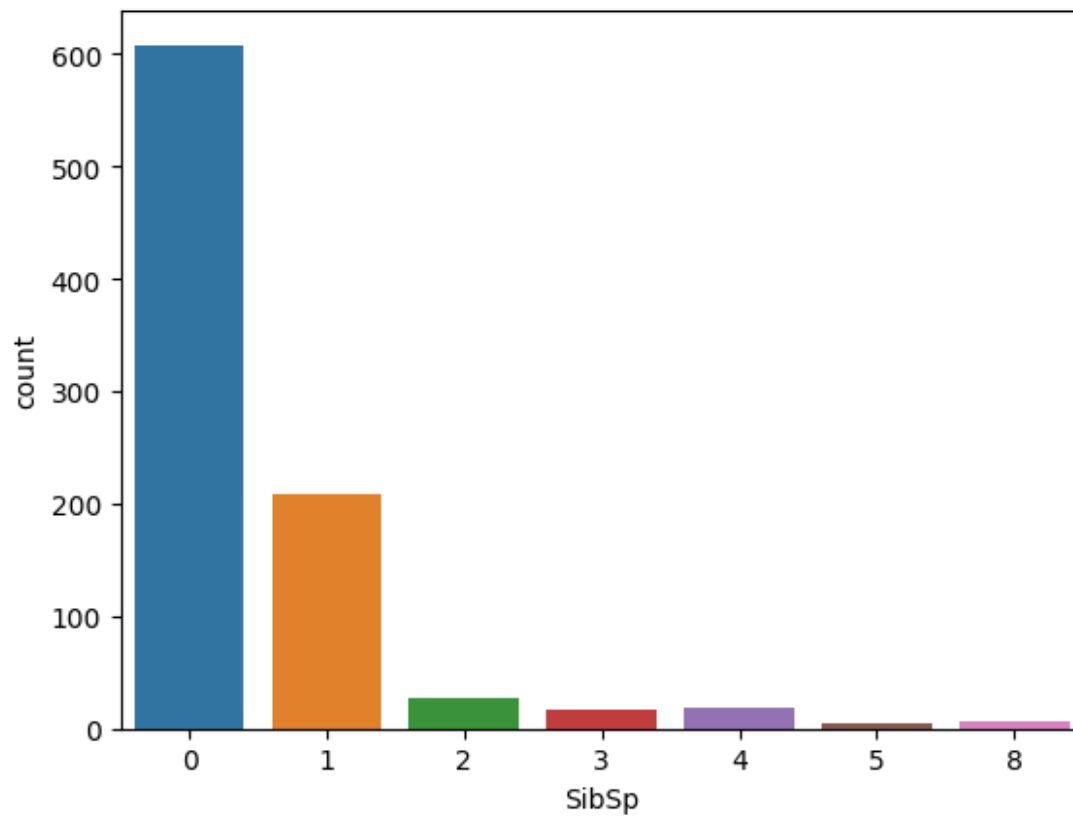


```
sns.countplot(x = 'Sex', data = train)  
<Axes: xlabel='Sex', ylabel='count'>
```

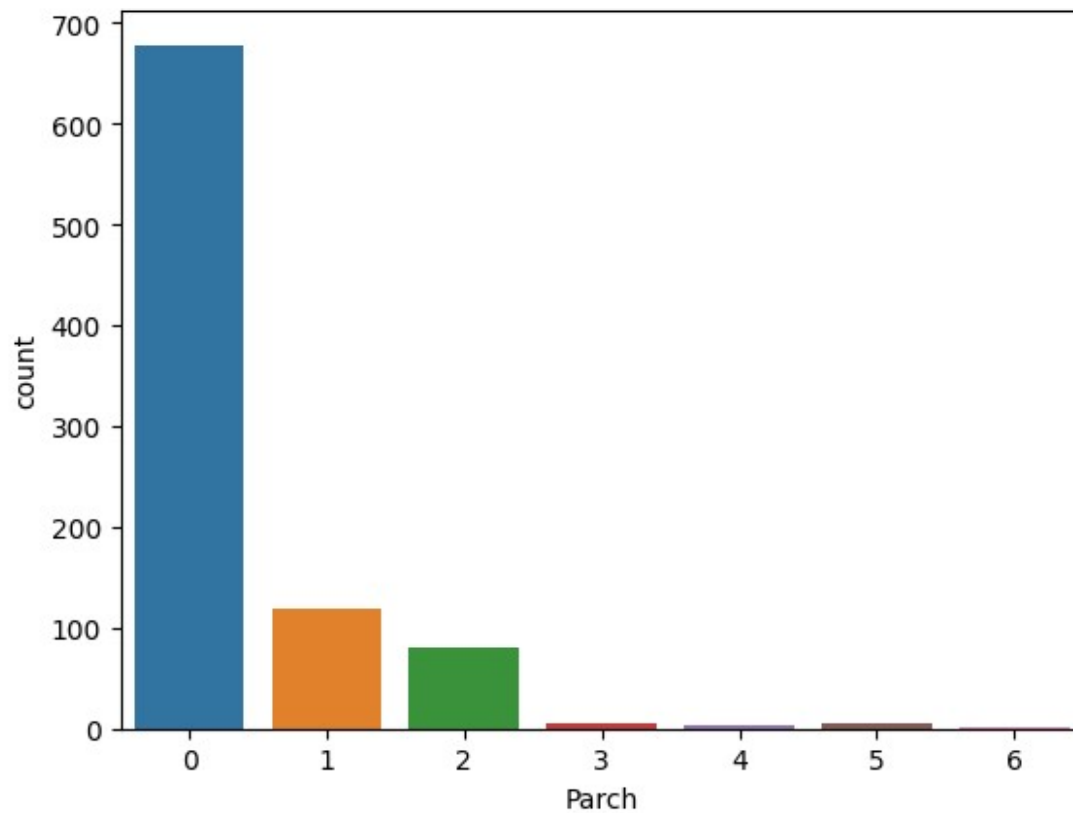




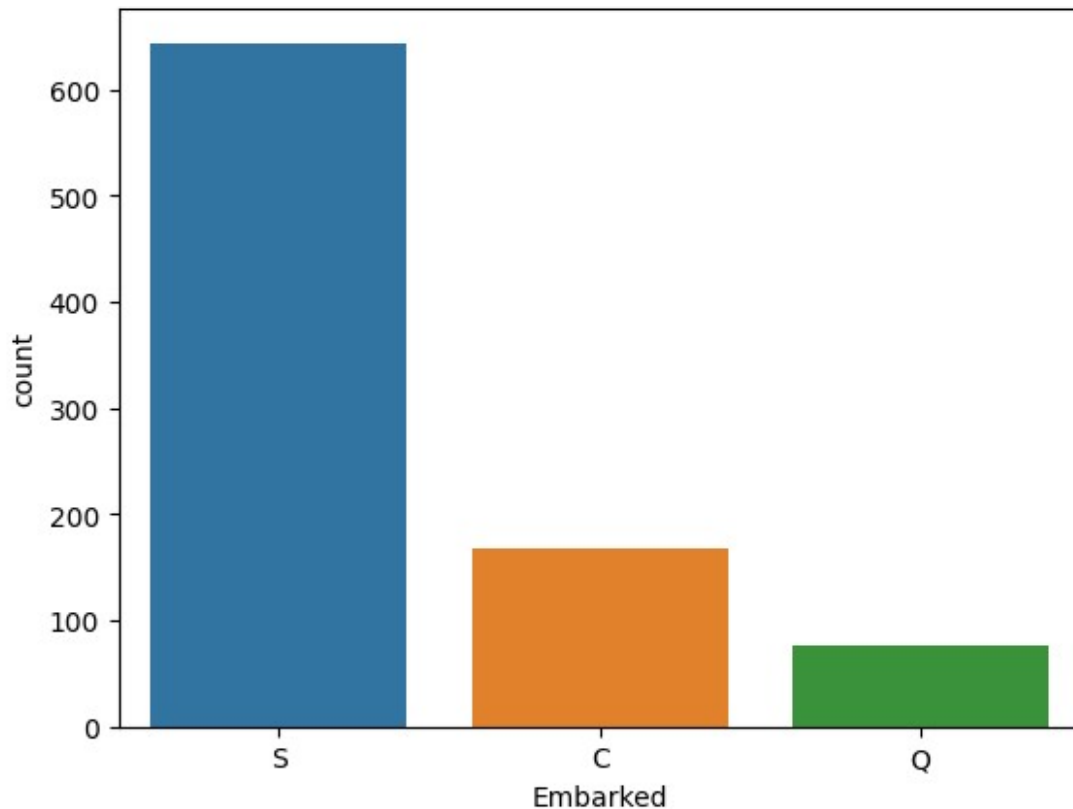
```
sns.countplot(x = 'SibSp', data = train)  
<Axes: xlabel='SibSp', ylabel='count'>
```



```
sns.countplot(x = 'Parch', data = train)  
<Axes: xlabel='Parch', ylabel='count'>
```



```
sns.countplot(x = 'Embarked', data = train)  
<Axes: xlabel='Embarked', ylabel='count'>
```



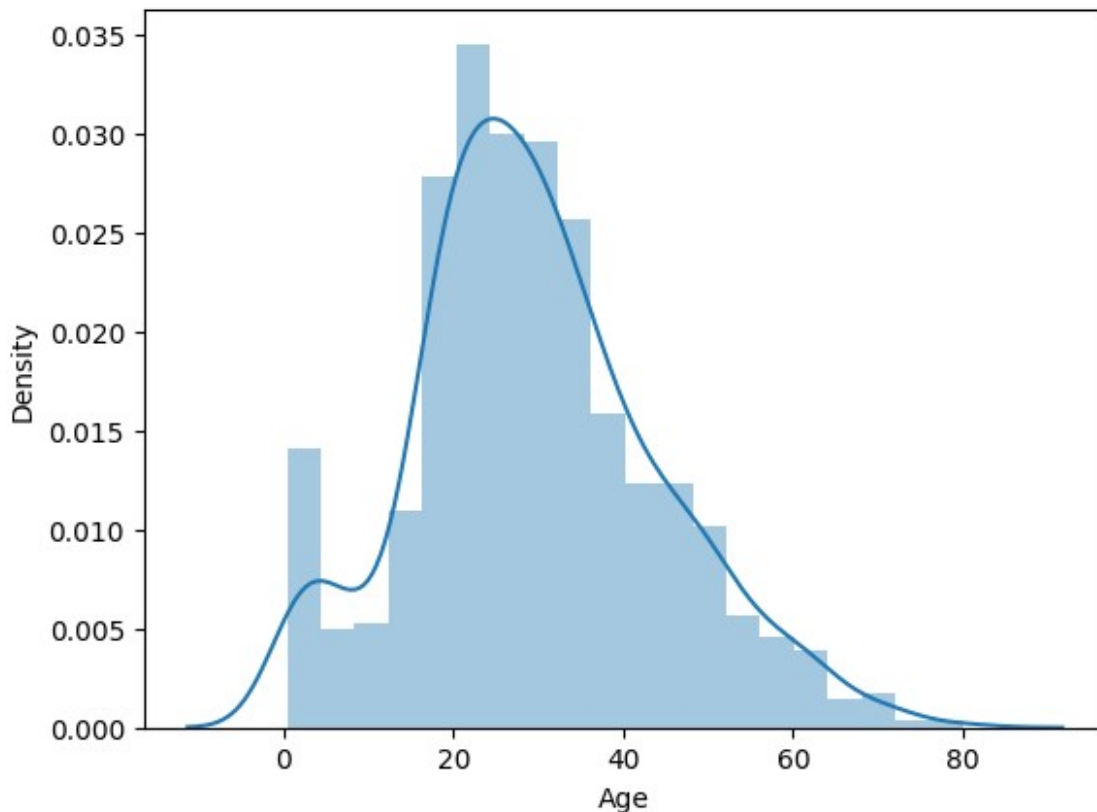
## numerical attributes

```
'''  
In summary, the main difference between countplot and distplot is the  
type of data they  
are designed to visualize:  
  
countplot is for categorical data and shows the count of unique  
categories.  
  
distplot is for continuous numerical data and shows the distribution  
of values along with  
an estimation of the probability density function.  
  
You should choose the appropriate plot based on the nature of your  
data and the specific  
insights you want to gain from your visualization.  
'''  
  
'\nIn summary, the main difference between countplot and distplot is  
the type of data they \nare designed to visualize:\n\ncountplot is for  
categorical data and shows the count of unique categories.\n\ndistplot  
is for continuous numerical data and shows the distribution of values
```

along with \nan estimation of the probability density function.\n\nYou should choose the appropriate plot based on the nature of your data and the specific \ninsights you want to gain from your visualization.\n'

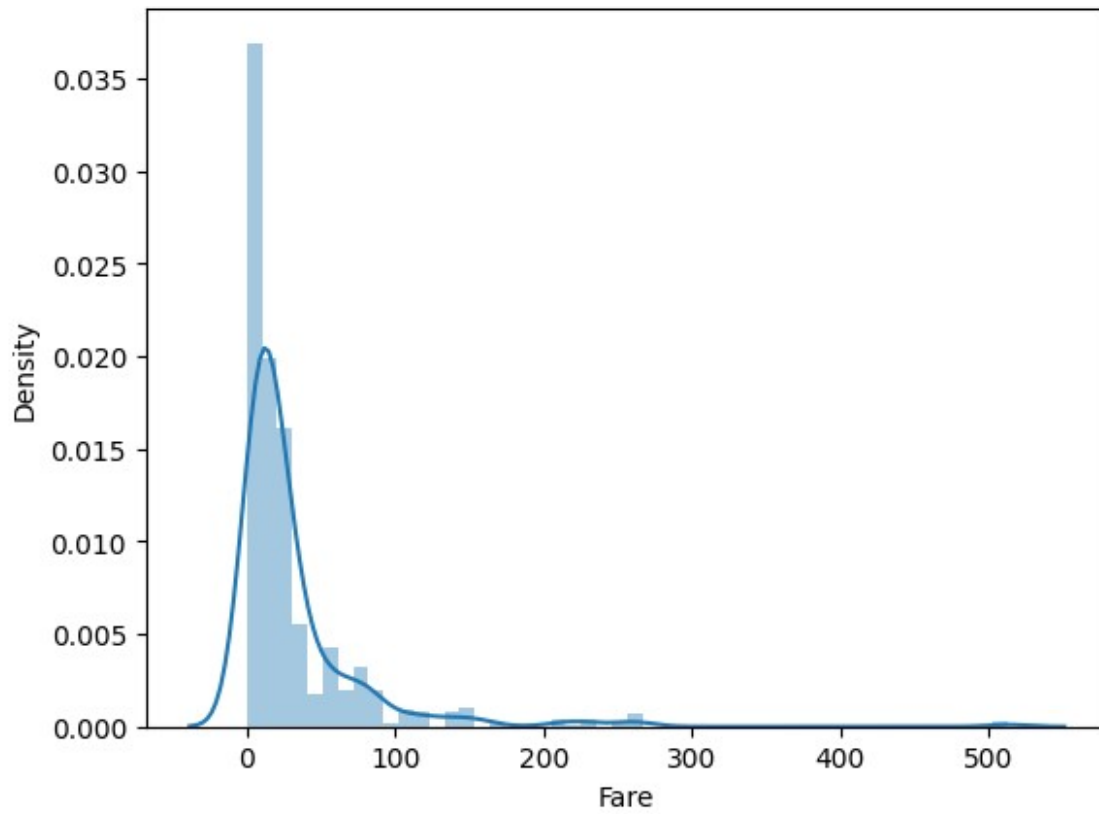
```
sns.distplot(train['Age'])
```

```
<Axes: xlabel='Age', ylabel='Density'>
```

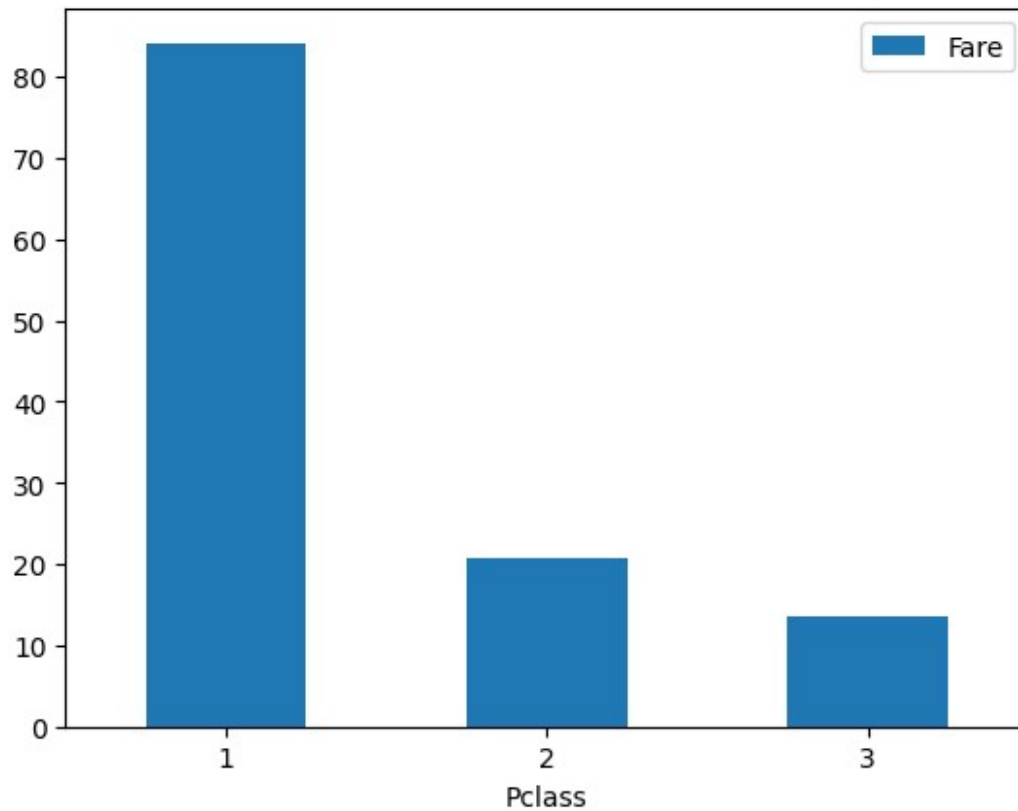


```
sns.distplot(train['Fare'])
```

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

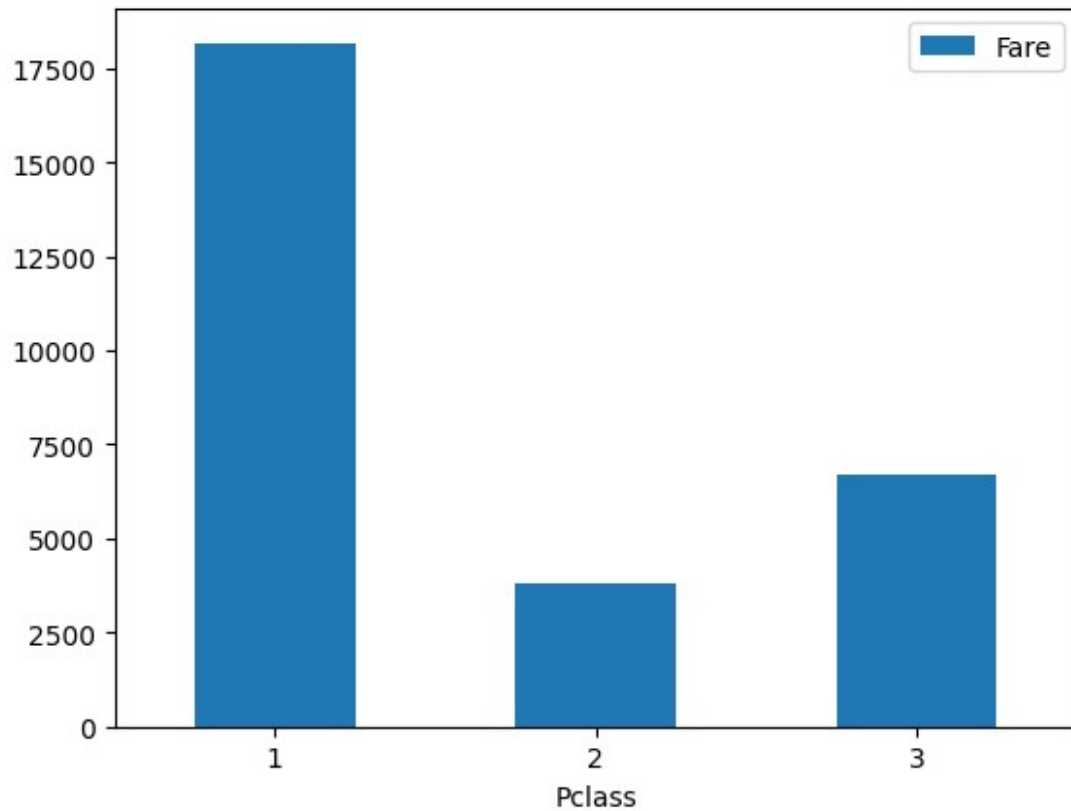


```
class_fare = train.pivot_table(index = 'Pclass', values = 'Fare')
class_fare.plot(kind = 'bar')
plt.xlabel('Pclass')
plt.xticks(rotation = 0)
plt.show()
```



*# this is for the total fare*

```
class_fare = train.pivot_table(index = 'Pclass', values = 'Fare',  
aggfunc = np.sum)  
class_fare.plot(kind = 'bar')  
plt.xlabel('Pclass')  
plt.xticks(rotation = 0)  
plt.show()
```

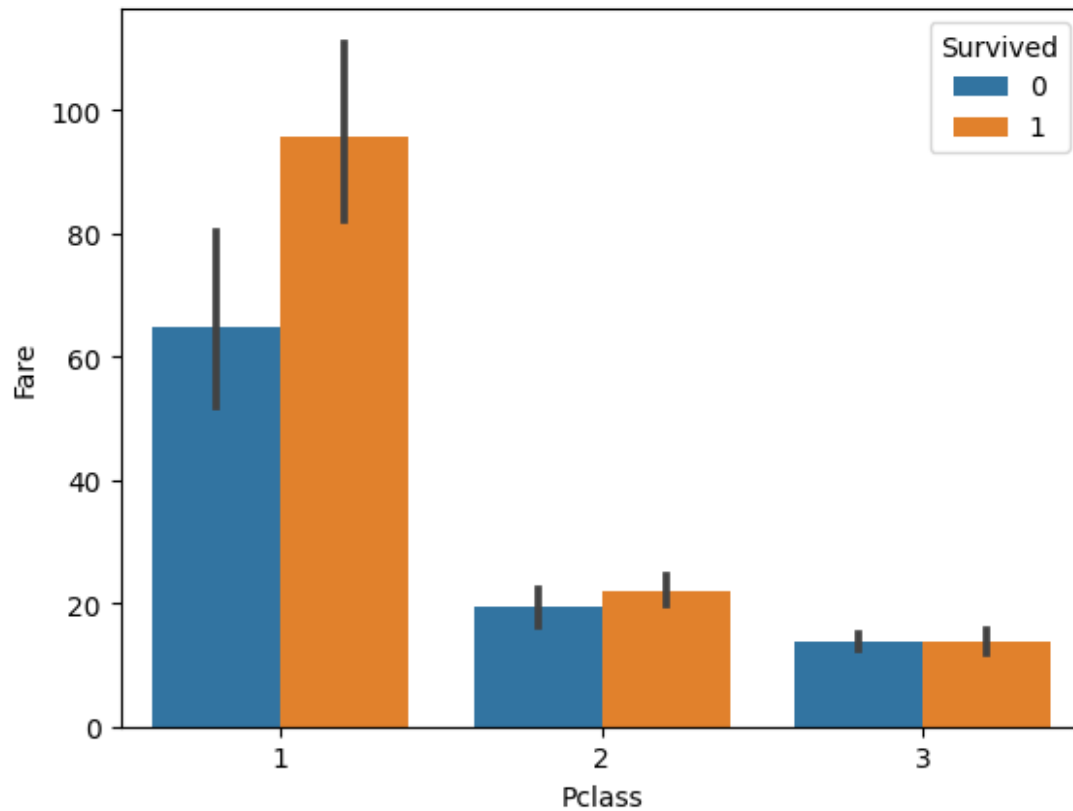


```
sns.barplot(data = train, x = 'Pclass', y = 'Fare', hue = 'Survived')
```

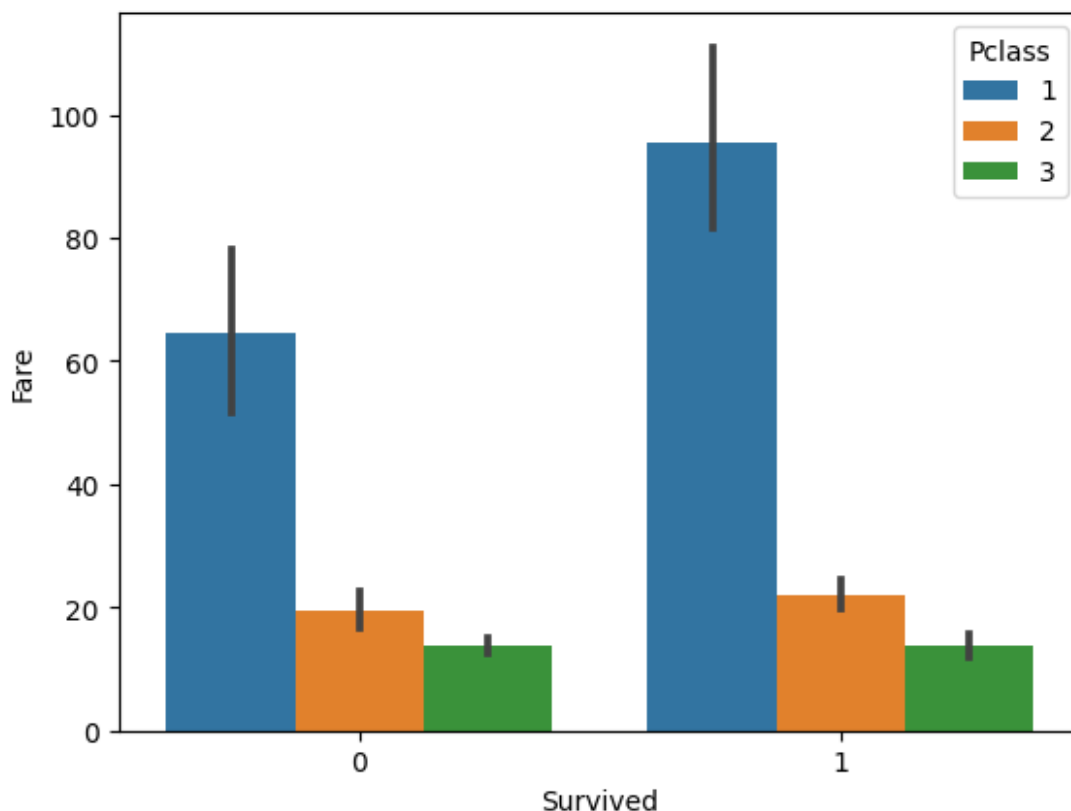
```
# this is for making the barplot  
# hue here is used to color the barplot
```

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





```
sns.barplot(data = train, x = 'Survived', y = 'Fare', hue = 'Pclass')  
  
# what this barplot tells us is that the number of people who survived  
# and whom did not  
# survive based on their class and fare  
  
<Axes: xlabel='Survived', ylabel='Fare'>
```



## data pre processing

```
train_len = len(train)
# combine two data frames
df = pd.concat([train, test], axis = 0)
df = df.reset_index(drop = True )
df.head()
```

	PassengerId	Survived	Pclass	\
0	1	0.0	3	
1	2	1.0	1	
2	3	1.0	3	
3	4	1.0	1	
4	5	0.0	3	

SibSp	\	Name	Sex	Age
0		Braund, Mr. Owen Harris	male	22.0
1		Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
1		Heikkinen, Miss. Laina	female	26.0
2				
0				

```

3      Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0
1
4      Allen, Mr. William Henry      male  35.0
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

```

```
df.tail()
```

```

      PassengerId  Survived  Pclass      Name
Sex \
1304      1305      NaN      3      Spector, Mr. Woolf
male
1305      1306      NaN      1  Oliva y Ocana, Dona. Fermina
female
1306      1307      NaN      3  Saether, Mr. Simon Sivertsen
male
1307      1308      NaN      3      Ware, Mr. Frederick
male
1308      1309      NaN      3      Peter, Master. Michael J
male

```

```

      Age  SibSp  Parch      Ticket      Fare Cabin Embarked
1304  NaN      0      0      A.5. 3236    8.0500  NaN      S
1305  39.0      0      0      PC 17758  108.9000  C105      C
1306  38.5      0      0  SOTON/O.Q. 3101262    7.2500  NaN      S
1307  NaN      0      0      359309    8.0500  NaN      S
1308  NaN      1      1      2668    22.3583  NaN      C

```

```
# find the null values
```

```
train_len
```

```
891
```

```
df.isnull().sum()
```

```

PassengerId      0
Survived          418
Pclass            0
Name              0

```

```

Sex          0
Age         263
SibSp        0
Parch        0
Ticket       0
Fare         1
Cabin       1014
Embarked     2
dtype: int64

# dropping the coloumn cabin
df = df.drop(columns = ['Cabin'], axis = 1)

df['Age'].mean()

29.881137667304014

df['Embarked'].mode()[0]

'S'

# filling the missing values using mean of that column
df['Age'] = df['Age'].fillna(df['Age'].mean())
df['Fare'] = df['Fare'].fillna(df['Fare'].mean())
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])

```

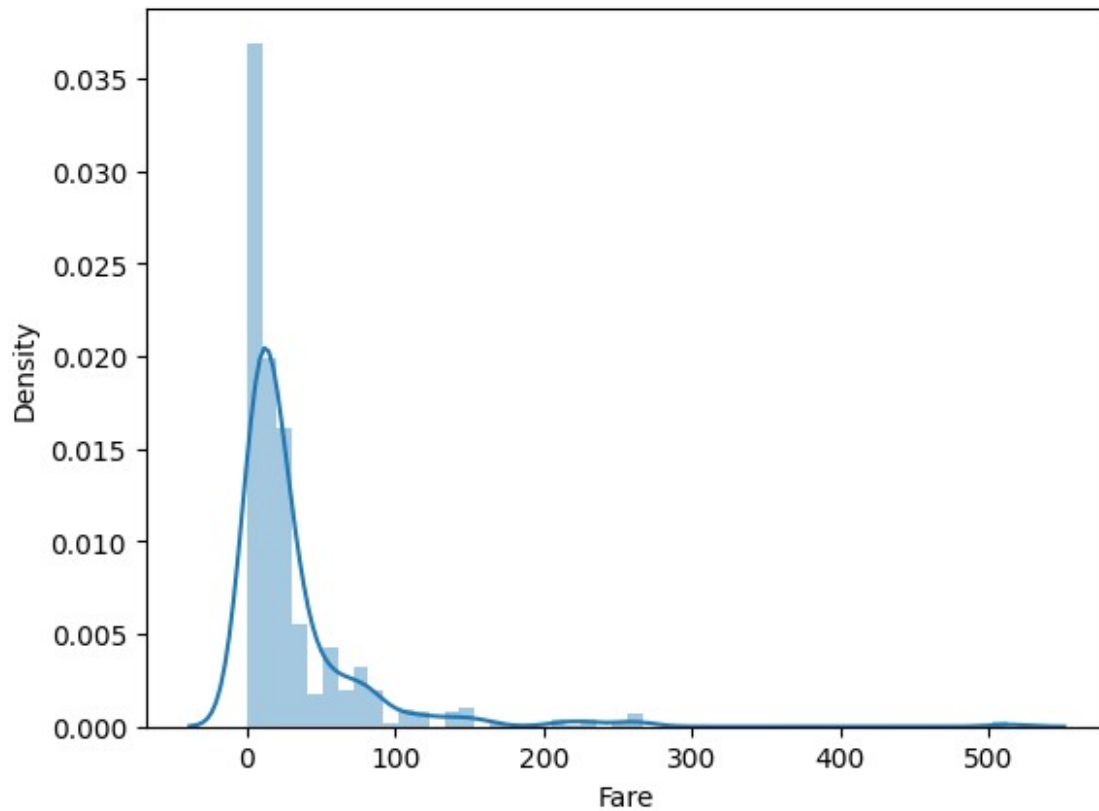
## log transformation for uniform data distribution

```

sns.distplot(train['Fare'])

<Axes: xlabel='Fare', ylabel='Density'>

```

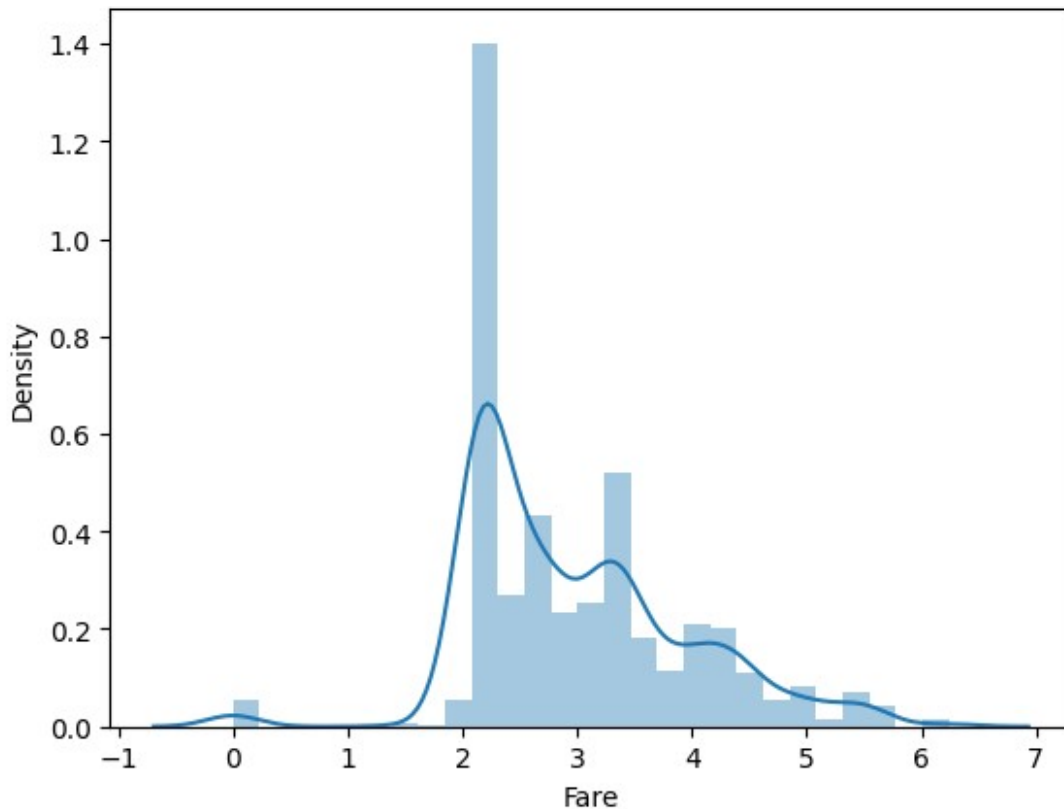


```
df['Fare'] = np.log(df['Fare'] + 1)

# what we are doing here is that we are replacing the null values with
mean

sns.distplot(df['Fare'])

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



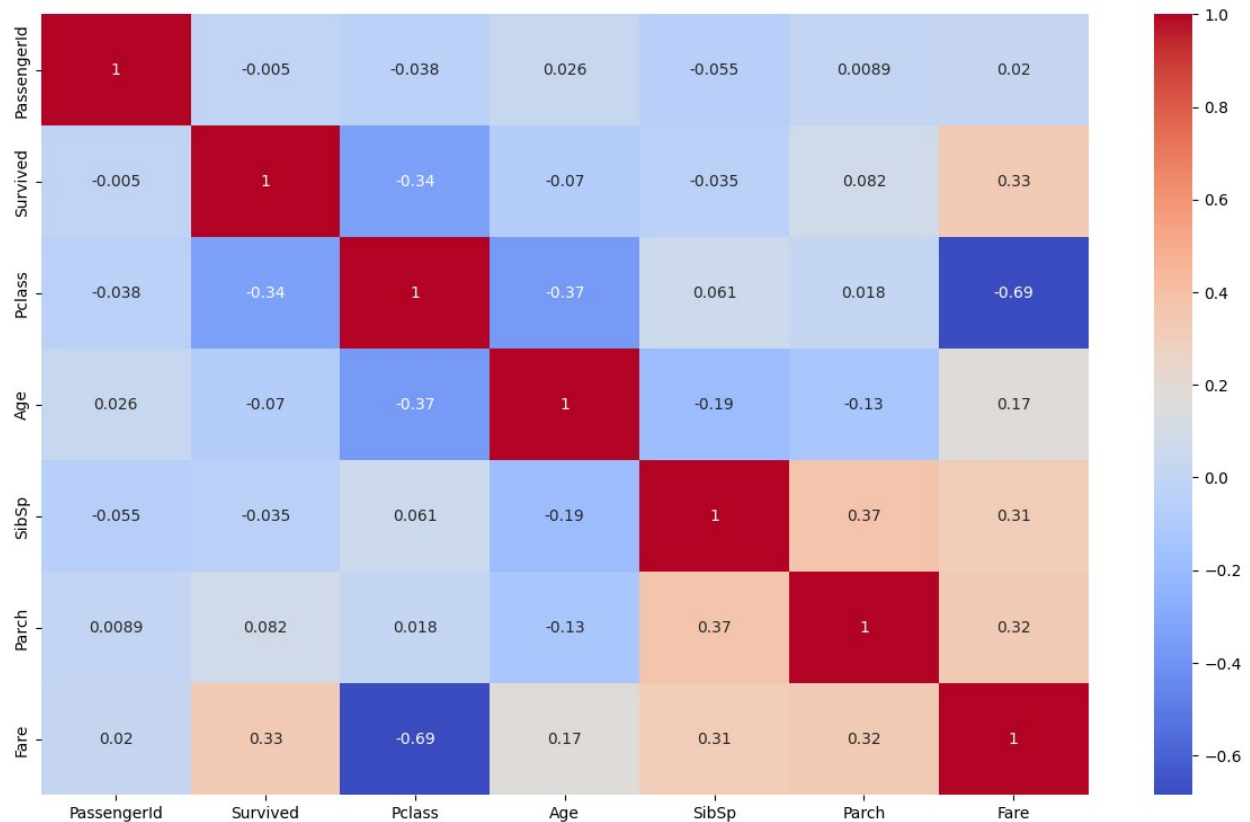
## co relation matrix

```
# this will only include the numeric column and remove the non-numeric
ones
numeric_df = df.select_dtypes(include=['float64', 'int64'])

# co relation for the numeric column
corr = numeric_df.corr()

# Create the heatmap
plt.figure(figsize=(15, 9))
sns.heatmap(corr, annot=True, cmap='coolwarm')

<Axes: >
```



*# small example for understanding heatmap*

```
data = [[1, 0.7, 0.3],
        [0.7, 1, 0.5],
        [0.3, 0.5, 1]]
```

*# Create the heatmap*

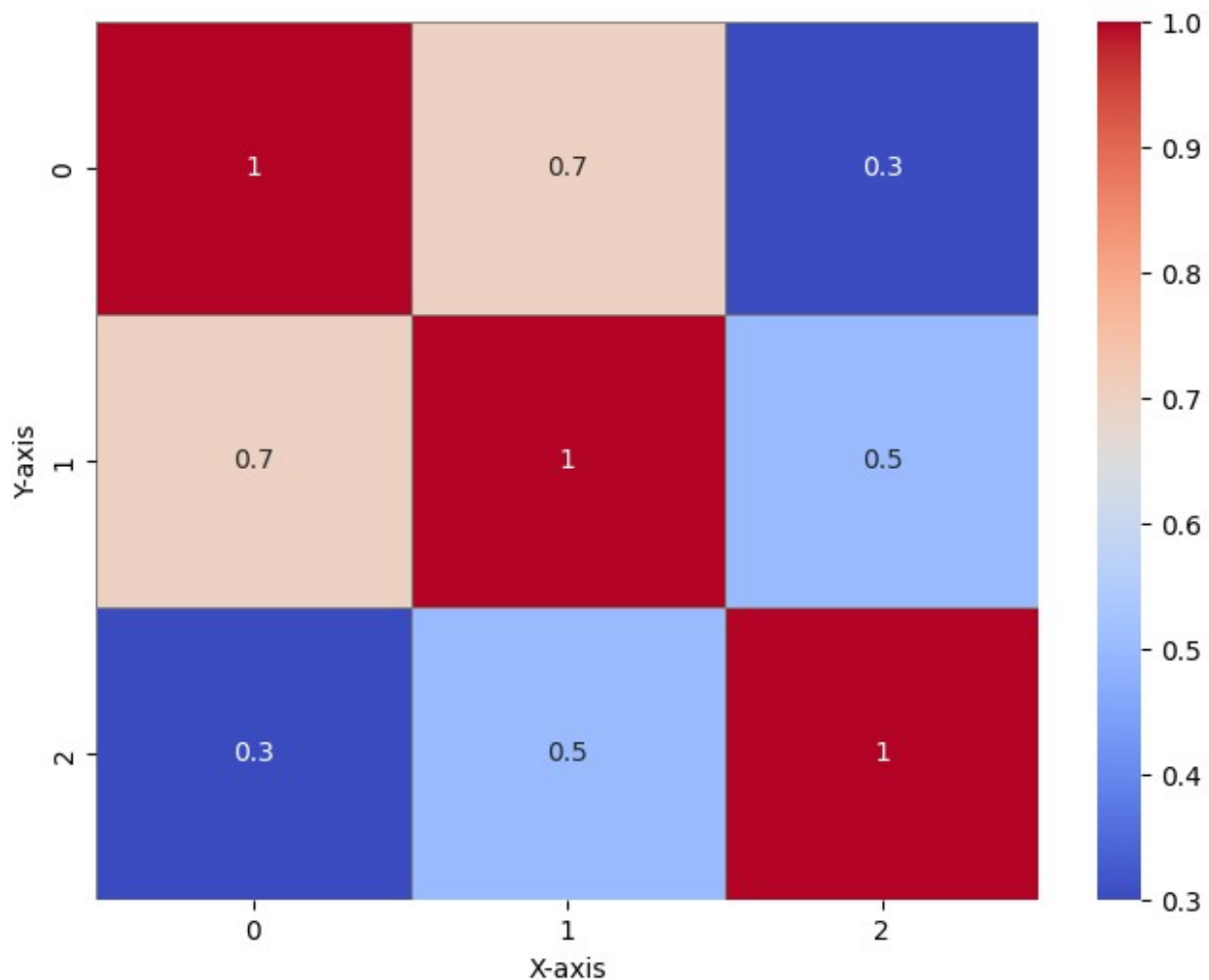
```
plt.figure(figsize=(8, 6)) # Adjust the figure size if needed
sns.heatmap(data, annot=True, cmap='coolwarm', linewidths=0.5,
            linecolor='gray')
```

*# Add labels to the axes (optional)*

```
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
```

*# Show the heatmap*

```
plt.show()
```



```
df.head()
```

	PassengerId	Survived	Pclass	\
0	1	0.0	3	
1	2	1.0	1	
2	3	1.0	3	
3	4	1.0	1	
4	5	0.0	3	

	SibSp	\	Name	Sex	Age
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			

	Parch	Ticket	Fare	Embarked
0	0	A/5 21171	2.110213	S
1	0	PC 17599	4.280593	C
2	0	STON/O2. 3101282	2.188856	S
3	0	113803	3.990834	S
4	0	373450	2.202765	S

```
df = df.drop(columns = ['Name', 'Ticket'], axis = 1)
```

```
df.head()
```

\	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare
0	1	0.0	3	male	22.0	1	0	2.110213
1	2	1.0	1	female	38.0	1	0	4.280593
2	3	1.0	3	female	26.0	0	0	2.188856
3	4	1.0	1	female	35.0	1	0	3.990834
4	5	0.0	3	male	35.0	0	0	2.202765

	Embarked
0	S
1	C
2	S
3	S
4	S

## label encoding

```
from sklearn.preprocessing import LabelEncoder

cols = ['Sex', 'Embarked']
le = LabelEncoder()
# what we are doing here is that we are creating an instance of the
label encoder class

for col in cols:
    df[col] = le.fit_transform(df[col])
df.head()
```

PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare
Embarked							

0	1	0.0	3	1	22.0	1	0	2.110213
2								
1	2	1.0	1	0	38.0	1	0	4.280593
0								
2	3	1.0	3	0	26.0	0	0	2.188856
2								
3	4	1.0	1	0	35.0	1	0	3.990834
2								
4	5	0.0	3	1	35.0	0	0	2.202765
2								

*# now here what we can see is that the values of sex and embarked which were strings  
# before have been converted to numerical values which can now be used to make and visuali  
# properly*

## train test split

```
train = df.iloc[:train_len, :]
test = df.iloc[train_len:, :]
```

```
train.head()
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0.0	3	1	22.0	1	0	2.110213	
2									
1	2	1.0	1	0	38.0	1	0	4.280593	
0									
2	3	1.0	3	0	26.0	0	0	2.188856	
2									
3	4	1.0	1	0	35.0	1	0	3.990834	
2									
4	5	0.0	3	1	35.0	0	0	2.202765	
2									

```
test.head()
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare
891	892	NaN	3	1	34.5	0	0	2.178064
892	893	NaN	3	0	47.0	1	0	2.079442
893	894	NaN	2	1	62.0	0	0	2.369075
894	895	NaN	3	1	27.0	0	0	2.268252

895	896	NaN	3	0	22.0	1	1	2.586824
-----	-----	-----	---	---	------	---	---	----------

	Embarked
891	1
892	2
893	1
894	2
895	2

```
X = train.drop(columns = ['PassengerId', 'Survived'], axis = 1)
y = train['Survived']
```

```
X.head()
```

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	3	1	22.0	1	0	2.110213	2
1	1	0	38.0	1	0	4.280593	0
2	3	0	26.0	0	0	2.188856	2
3	1	0	35.0	1	0	3.990834	2
4	3	1	35.0	0	0	2.202765	2

## model training

```
from sklearn.model_selection import train_test_split, cross_val_score
# classify column
```

```
def classify(model):
    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size = 0.25, random_state = 42)
    model.fit(X_train, y_train)
    print('Accuracy: ', model.score(X_test, y_test))

    score = cross_val_score(model, X, y, cv = 5)
    print('CV Score: ', np.mean(score))
```

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
classify(model)
```

```
Accuracy: 0.8071748878923767
CV Score: 0.7833971502102819
```

```
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
classify(model)
```

```
Accuracy: 0.726457399103139
CV Score: 0.7677044755508129
```

```
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
classify(model)
```

Accuracy: 0.8026905829596412

CV Score: 0.815956311593748

```
!pip install xgboost
```

Collecting xgboost

Obtaining dependency information for xgboost from  
[https://files.pythonhosted.org/packages/32/10/4689bda37403f7dd029d550c4446e0097c2f33b8ae877b235e76d5c49bc2/xgboost-2.0.0-py3-none-win\\_amd64.whl.metadata](https://files.pythonhosted.org/packages/32/10/4689bda37403f7dd029d550c4446e0097c2f33b8ae877b235e76d5c49bc2/xgboost-2.0.0-py3-none-win_amd64.whl.metadata)

Downloading xgboost-2.0.0-py3-none-win\_amd64.whl.metadata (2.0 kB)  
Requirement already satisfied: numpy in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from xgboost) (1.25.2)

Requirement already satisfied: scipy in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from xgboost) (1.11.2)

Downloading xgboost-2.0.0-py3-none-win\_amd64.whl (99.7 MB)

```
----- 0.0/99.7 MB ? eta -:-:--
----- 0.0/99.7 MB ? eta -:-:--
----- 0.1/99.7 MB 656.4 kB/s eta
0:02:32
----- 0.5/99.7 MB 3.8 MB/s eta
0:00:27
----- 0.9/99.7 MB 5.3 MB/s eta
0:00:19
----- 1.5/99.7 MB 6.9 MB/s eta
0:00:15
----- 2.1/99.7 MB 8.0 MB/s eta
0:00:13
----- 2.7/99.7 MB 8.6 MB/s eta
0:00:12
----- 3.4/99.7 MB 9.3 MB/s eta
0:00:11
----- 4.0/99.7 MB 9.9 MB/s eta
0:00:10
----- 4.6/99.7 MB 10.1 MB/s eta
0:00:10
----- 5.1/99.7 MB 10.1 MB/s eta
0:00:10
----- 5.6/99.7 MB 10.2 MB/s eta
0:00:10
----- 6.2/99.7 MB 10.4 MB/s eta
0:00:09
----- 6.8/99.7 MB 10.5 MB/s eta
0:00:09
----- 7.3/99.7 MB 10.4 MB/s eta
0:00:09
```

0:00:09	-----	8.0/99.7 MB	10.8 MB/s	eta
0:00:09	-----	8.1/99.7 MB	10.7 MB/s	eta
0:00:10	-----	8.3/99.7 MB	10.0 MB/s	eta
0:00:10	-----	8.8/99.7 MB	9.8 MB/s	eta
0:00:09	-----	9.4/99.7 MB	10.2 MB/s	eta
0:00:09	-----	10.0/99.7 MB	10.3 MB/s	eta
0:00:09	-----	10.6/99.7 MB	11.1 MB/s	eta
0:00:08	-----	11.1/99.7 MB	11.3 MB/s	eta
0:00:08	-----	11.6/99.7 MB	11.3 MB/s	eta
0:00:08	-----	12.2/99.7 MB	11.1 MB/s	eta
0:00:08	-----	12.7/99.7 MB	11.3 MB/s	eta
0:00:08	-----	13.3/99.7 MB	11.1 MB/s	eta
0:00:08	-----	14.0/99.7 MB	10.9 MB/s	eta
0:00:08	-----	14.3/99.7 MB	11.1 MB/s	eta
0:00:08	-----	15.1/99.7 MB	11.1 MB/s	eta
0:00:08	-----	15.7/99.7 MB	11.1 MB/s	eta
0:00:08	-----	16.2/99.7 MB	11.3 MB/s	eta
0:00:08	-----	16.8/99.7 MB	11.3 MB/s	eta
0:00:08	-----	17.4/99.7 MB	11.3 MB/s	eta
0:00:08	-----	18.0/99.7 MB	11.3 MB/s	eta
0:00:07	-----	18.5/99.7 MB	11.9 MB/s	eta
0:00:07	-----	19.1/99.7 MB	12.1 MB/s	eta
0:00:07	-----	19.7/99.7 MB	11.9 MB/s	eta
0:00:07	-----	20.3/99.7 MB	11.9 MB/s	eta
	-----	20.8/99.7 MB	11.9 MB/s	eta

0:00:07	-----	21.4/99.7 MB 11.9 MB/s eta
0:00:07	-----	22.0/99.7 MB 11.9 MB/s eta
0:00:07	-----	22.5/99.7 MB 11.9 MB/s eta
0:00:07	-----	23.1/99.7 MB 12.4 MB/s eta
0:00:07	-----	23.7/99.7 MB 12.1 MB/s eta
0:00:07	-----	24.3/99.7 MB 12.1 MB/s eta
0:00:07	-----	24.8/99.7 MB 12.4 MB/s eta
0:00:07	-----	25.4/99.7 MB 11.9 MB/s eta
0:00:07	-----	26.0/99.7 MB 12.1 MB/s eta
0:00:07	-----	26.6/99.7 MB 12.1 MB/s eta
0:00:07	-----	27.1/99.7 MB 12.1 MB/s eta
0:00:06	-----	27.7/99.7 MB 12.1 MB/s eta
0:00:06	-----	28.2/99.7 MB 12.1 MB/s eta
0:00:06	-----	28.9/99.7 MB 11.9 MB/s eta
0:00:06	-----	29.3/99.7 MB 11.9 MB/s eta
0:00:06	-----	29.9/99.7 MB 12.1 MB/s eta
0:00:06	-----	30.5/99.7 MB 12.1 MB/s eta
0:00:06	-----	31.0/99.7 MB 12.1 MB/s eta
0:00:06	-----	31.6/99.7 MB 11.9 MB/s eta
0:00:06	-----	32.2/99.7 MB 11.9 MB/s eta
0:00:06	-----	32.8/99.7 MB 11.9 MB/s eta
0:00:06	-----	33.4/99.7 MB 11.9 MB/s eta
0:00:06	-----	34.0/99.7 MB 11.9 MB/s eta
0:00:06	-----	34.5/99.7 MB 11.9 MB/s eta

-----	35.1/99.7 MB	11.9 MB/s	eta
0:00:06			
-----	35.6/99.7 MB	11.9 MB/s	eta
0:00:06			
-----	36.2/99.7 MB	12.1 MB/s	eta
0:00:06			
-----	36.8/99.7 MB	12.1 MB/s	eta
0:00:06			
-----	37.4/99.7 MB	12.1 MB/s	eta
0:00:06			
-----	38.0/99.7 MB	12.1 MB/s	eta
0:00:06			
-----	38.7/99.7 MB	12.1 MB/s	eta
0:00:06			
-----	39.2/99.7 MB	12.1 MB/s	eta
0:00:05			
-----	39.8/99.7 MB	12.4 MB/s	eta
0:00:05			
-----	40.4/99.7 MB	12.4 MB/s	eta
0:00:05			
-----	40.9/99.7 MB	12.1 MB/s	eta
0:00:05			
-----	41.6/99.7 MB	12.1 MB/s	eta
0:00:05			
-----	42.1/99.7 MB	12.4 MB/s	eta
0:00:05			
-----	42.7/99.7 MB	12.1 MB/s	eta
0:00:05			
-----	43.2/99.7 MB	12.1 MB/s	eta
0:00:05			
-----	43.3/99.7 MB	11.9 MB/s	eta
0:00:05			
-----	44.1/99.7 MB	11.7 MB/s	eta
0:00:05			
-----	44.6/99.7 MB	11.7 MB/s	eta
0:00:05			
-----	45.1/99.7 MB	11.7 MB/s	eta
0:00:05			
-----	45.6/99.7 MB	11.7 MB/s	eta
0:00:05			
-----	46.0/99.7 MB	11.3 MB/s	eta
0:00:05			
-----	46.3/99.7 MB	11.1 MB/s	eta
0:00:05			
-----	46.7/99.7 MB	10.7 MB/s	eta
0:00:05			
-----	47.1/99.7 MB	10.7 MB/s	eta
0:00:05			
-----	47.7/99.7 MB	10.7 MB/s	eta

0:00:05	-----	48.1/99.7 MB 10.6 MB/s eta
0:00:05	-----	48.7/99.7 MB 10.7 MB/s eta
0:00:05	-----	49.2/99.7 MB 10.6 MB/s eta
0:00:05	-----	49.8/99.7 MB 10.6 MB/s eta
0:00:05	-----	50.4/99.7 MB 10.6 MB/s eta
0:00:05	-----	50.9/99.7 MB 10.6 MB/s eta
0:00:05	-----	51.5/99.7 MB 10.4 MB/s eta
0:00:05	-----	52.0/99.7 MB 10.6 MB/s eta
0:00:05	-----	52.6/99.7 MB 10.6 MB/s eta
0:00:05	-----	53.1/99.7 MB 10.6 MB/s eta
0:00:05	-----	53.5/99.7 MB 10.9 MB/s eta
0:00:05	-----	54.1/99.7 MB 10.6 MB/s eta
0:00:05	-----	54.6/99.7 MB 10.4 MB/s eta
0:00:05	-----	55.1/99.7 MB 10.6 MB/s eta
0:00:05	-----	55.7/99.7 MB 10.6 MB/s eta
0:00:05	-----	56.2/99.7 MB 10.9 MB/s eta
0:00:04	-----	56.8/99.7 MB 11.3 MB/s eta
0:00:04	-----	57.3/99.7 MB 11.3 MB/s eta
0:00:04	-----	57.8/99.7 MB 11.1 MB/s eta
0:00:04	-----	58.3/99.7 MB 11.3 MB/s eta
0:00:04	-----	58.8/99.7 MB 11.3 MB/s eta
0:00:04	-----	59.3/99.7 MB 11.3 MB/s eta
0:00:04	-----	59.6/99.7 MB 11.1 MB/s eta
0:00:04	-----	60.4/99.7 MB 11.1 MB/s eta



-----	61.0/99.7 MB 11.3 MB/s eta
0:00:04	
-----	61.4/99.7 MB 11.1 MB/s eta
0:00:04	
-----	62.0/99.7 MB 11.1 MB/s eta
0:00:04	
-----	62.5/99.7 MB 11.1 MB/s eta
0:00:04	
-----	63.1/99.7 MB 11.1 MB/s eta
0:00:04	
-----	63.6/99.7 MB 10.9 MB/s eta
0:00:04	
-----	64.2/99.7 MB 11.3 MB/s eta
0:00:04	
-----	64.7/99.7 MB 11.1 MB/s eta
0:00:04	
-----	65.2/99.7 MB 11.1 MB/s eta
0:00:04	
-----	65.8/99.7 MB 11.3 MB/s eta
0:00:04	
-----	66.4/99.7 MB 11.3 MB/s eta
0:00:03	
-----	66.9/99.7 MB 11.1 MB/s eta
0:00:03	
-----	67.5/99.7 MB 11.5 MB/s eta
0:00:03	
-----	68.1/99.7 MB 11.3 MB/s eta
0:00:03	
-----	68.7/99.7 MB 11.5 MB/s eta
0:00:03	
-----	69.3/99.7 MB 11.5 MB/s eta
0:00:03	
-----	69.8/99.7 MB 11.5 MB/s eta
0:00:03	
-----	70.4/99.7 MB 11.7 MB/s eta
0:00:03	
-----	70.9/99.7 MB 11.7 MB/s eta
0:00:03	
-----	71.4/99.7 MB 11.7 MB/s eta
0:00:03	
-----	72.0/99.7 MB 11.5 MB/s eta
0:00:03	
-----	72.6/99.7 MB 11.7 MB/s eta
0:00:03	
-----	73.1/99.7 MB 11.7 MB/s eta
0:00:03	
-----	73.7/99.7 MB 11.9 MB/s eta
0:00:03	
-----	74.3/99.7 MB 11.7 MB/s eta

0:00:03	-----	74.8/99.7 MB	11.7 MB/s	eta
0:00:03	-----	75.4/99.7 MB	11.7 MB/s	eta
0:00:03	-----	75.8/99.7 MB	11.5 MB/s	eta
0:00:03	-----	76.4/99.7 MB	11.5 MB/s	eta
0:00:03	-----	77.0/99.7 MB	11.5 MB/s	eta
0:00:02	-----	77.5/99.7 MB	11.5 MB/s	eta
0:00:02	-----	78.1/99.7 MB	11.7 MB/s	eta
0:00:02	-----	78.6/99.7 MB	11.5 MB/s	eta
0:00:02	-----	79.3/99.7 MB	11.7 MB/s	eta
0:00:02	-----	79.7/99.7 MB	11.5 MB/s	eta
0:00:02	-----	80.2/99.7 MB	11.5 MB/s	eta
0:00:02	-----	80.3/99.7 MB	10.9 MB/s	eta
0:00:02	-----	81.2/99.7 MB	11.3 MB/s	eta
0:00:02	-----	81.8/99.7 MB	11.3 MB/s	eta
0:00:02	-----	82.4/99.7 MB	11.7 MB/s	eta
0:00:02	-----	83.0/99.7 MB	11.3 MB/s	eta
0:00:02	-----	83.5/99.7 MB	11.5 MB/s	eta
0:00:02	-----	84.1/99.7 MB	11.7 MB/s	eta
0:00:02	-----	84.6/99.7 MB	11.5 MB/s	eta
0:00:02	-----	85.2/99.7 MB	11.5 MB/s	eta
0:00:02	-----	85.7/99.7 MB	11.7 MB/s	eta
0:00:02	-----	86.3/99.7 MB	11.7 MB/s	eta
0:00:02	-----	86.9/99.7 MB	11.7 MB/s	eta
0:00:02	-----	87.4/99.7 MB	11.7 MB/s	eta

```
----- 88.0/99.7 MB 11.5 MB/s eta
0:00:02
----- 88.4/99.7 MB 11.5 MB/s eta
0:00:01
----- 88.9/99.7 MB 11.5 MB/s eta
0:00:01
----- 89.5/99.7 MB 11.5 MB/s eta
0:00:01
----- 90.1/99.7 MB 11.5 MB/s eta
0:00:01
----- 90.6/99.7 MB 12.1 MB/s eta
0:00:01
----- 91.0/99.7 MB 11.5 MB/s eta
0:00:01
----- 91.3/99.7 MB 11.3 MB/s eta
0:00:01
----- 91.7/99.7 MB 10.9 MB/s eta
0:00:01
----- 92.2/99.7 MB 10.9 MB/s eta
0:00:01
----- 92.7/99.7 MB 10.9 MB/s eta
0:00:01
----- 93.1/99.7 MB 10.7 MB/s eta
0:00:01
----- 93.7/99.7 MB 10.7 MB/s eta
0:00:01
----- 94.2/99.7 MB 10.7 MB/s eta
0:00:01
----- 94.2/99.7 MB 10.7 MB/s eta
0:00:01
----- 94.2/99.7 MB 10.7 MB/s eta
0:00:01
----- 94.3/99.7 MB 9.5 MB/s eta
0:00:01
----- 94.9/99.7 MB 9.4 MB/s eta
0:00:01
----- 95.3/99.7 MB 9.2 MB/s eta
0:00:01
----- 95.9/99.7 MB 9.4 MB/s eta
0:00:01
----- 96.5/99.7 MB 9.2 MB/s eta
0:00:01
----- 97.1/99.7 MB 9.2 MB/s eta
0:00:01
----- 97.6/99.7 MB 9.5 MB/s eta
0:00:01
----- 98.2/99.7 MB 9.2 MB/s eta
0:00:01
----- 98.7/99.7 MB 9.4 MB/s eta
```

```

0:00:01 ----- 99.3/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 9.5 MB/s eta
0:00:01 ----- 99.7/99.7 MB 6.6 MB/s eta
0:00:00

```

Installing collected packages: xgboost  
Successfully installed xgboost-2.0.0

```

from sklearn.svm import SVC
model = SVC()
classify(model)

```

```

Accuracy: 0.6188340807174888
CV Score: 0.6835289686774214

```

```

from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier()
classify(model)

```

```

Accuracy: 0.7802690582959642
CV Score: 0.7744397715146569

```

```

import xgboost as xgb

```

```

from xgboost import XGBClassifier
model = XGBClassifier()
classify(model)

```

```

Accuracy: 0.7847533632286996
CV Score: 0.8148327160881301

```

```

!pip install lightgbm catboost

```

```

Collecting lightgbm
  Obtaining dependency information for lightgbm from

```

[https://files.pythonhosted.org/packages/b3/f8/ee33e36194eb03a76eccf3adac3fba51f0e56fbd20609bb531659d48d3cb/lightgbm-4.1.0-py3-none-win\\_amd64.whl.metadata](https://files.pythonhosted.org/packages/b3/f8/ee33e36194eb03a76eccf3adac3fba51f0e56fbd20609bb531659d48d3cb/lightgbm-4.1.0-py3-none-win_amd64.whl.metadata)

Downloading lightgbm-4.1.0-py3-none-win\_amd64.whl.metadata (19 kB)

Collecting catboost

Obtaining dependency information for catboost from

[https://files.pythonhosted.org/packages/e2/a7/aed402760cde64ca28b7787ad2a63a3f1d2066c2f058b2d974ce5479e32e/catboost-1.2.1.1-cp311-cp311-win\\_amd64.whl.metadata](https://files.pythonhosted.org/packages/e2/a7/aed402760cde64ca28b7787ad2a63a3f1d2066c2f058b2d974ce5479e32e/catboost-1.2.1.1-cp311-cp311-win_amd64.whl.metadata)

Downloading catboost-1.2.1.1-cp311-cp311-win\_amd64.whl.metadata (1.2 kB)

Requirement already satisfied: numpy in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from lightgbm) (1.25.2)

Requirement already satisfied: scipy in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from lightgbm) (1.11.2)

Collecting graphviz (from catboost)

Downloading graphviz-0.20.1-py3-none-any.whl (47 kB)

----- 0.0/47.0 kB ? eta

--:--:--

----- 41.0/47.0 kB 991.0 kB/s

eta 0:00:01

----- 47.0/47.0 kB 782.9 kB/s

eta 0:00:00

Requirement already satisfied: matplotlib in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from catboost) (3.7.3)

Requirement already satisfied: pandas>=0.24 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from catboost) (2.1.0)

Collecting plotly (from catboost)

Obtaining dependency information for plotly from

<https://files.pythonhosted.org/packages/df/79/c80174d711ee26ee5da55a9cc3e248f1ec7a0188b5e4d6bbbbc09b974b0/plotly-5.17.0-py2.py3-none-any.whl.metadata>

Downloading plotly-5.17.0-py2.py3-none-any.whl.metadata (7.0 kB)

Requirement already satisfied: six in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from catboost) (1.16.0)

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pandas>=0.24->catboost) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pandas>=0.24->catboost) (2023.3)

Requirement already satisfied: tzdata>=2022.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from pandas>=0.24->catboost) (2023.3)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\ilove\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->catboost) (1.1.0)

```

Requirement already satisfied: cycycler>=0.10 in c:\users\ilove\appdata\
local\programs\python\python311\lib\site-packages (from matplotlib-
>catboost) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
matplotlib->catboost) (4.22.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
matplotlib->catboost) (1.4.5)
Requirement already satisfied: packaging>=20.0 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
matplotlib->catboost) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
matplotlib->catboost) (10.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\ilove\
appdata\local\programs\python\python311\lib\site-packages (from
matplotlib->catboost) (3.1.1)
Collecting tenacity>=6.2.0 (from plotly->catboost)
  Obtaining dependency information for tenacity>=6.2.0 from
  https://files.pythonhosted.org/packages/f4/f1/990741d5bb2487d529d20a43
  3210ffa136a367751e454214013b441c4575/tenacity-8.2.3-py3-none-
  any.whl.metadata
  Downloading tenacity-8.2.3-py3-none-any.whl.metadata (1.0 kB)
Downloading lightgbm-4.1.0-py3-none-win_amd64.whl (1.3 MB)
----- 0.0/1.3 MB ? eta -:--:--
----- 0.3/1.3 MB 10.2 MB/s eta
0:00:01
----- 0.9/1.3 MB 9.3 MB/s eta
0:00:01
----- 1.3/1.3 MB 10.4 MB/s eta
0:00:01
----- 1.3/1.3 MB 10.4 MB/s eta
0:00:00
Downloading catboost-1.2.1.1-cp311-cp311-win_amd64.whl (101.0 MB)
----- 0.0/101.0 MB ? eta -:--:--
----- 0.5/101.0 MB 15.2 MB/s eta
0:00:07
----- 1.1/101.0 MB 13.9 MB/s eta
0:00:08
----- 1.6/101.0 MB 12.4 MB/s eta
0:00:09
----- 2.2/101.0 MB 12.6 MB/s eta
0:00:08
----- 2.6/101.0 MB 12.9 MB/s eta
0:00:08
----- 3.2/101.0 MB 12.7 MB/s eta
0:00:08
----- 3.8/101.0 MB 12.6 MB/s eta

```

```
0:00:08
- ----- 4.1/101.0 MB 12.0 MB/s eta
0:00:09
- ----- 4.6/101.0 MB 12.4 MB/s eta
0:00:08
- ----- 5.1/101.0 MB 12.0 MB/s eta
0:00:08
- ----- 5.7/101.0 MB 12.1 MB/s eta
0:00:08
- ----- 6.1/101.0 MB 12.2 MB/s eta
0:00:08
- ----- 6.6/101.0 MB 12.0 MB/s eta
0:00:08
- ----- 6.6/101.0 MB 11.4 MB/s eta
0:00:09
- ----- 6.7/101.0 MB 10.5 MB/s eta
0:00:10
- ----- 6.9/101.0 MB 9.9 MB/s eta
0:00:10
- ----- 7.6/101.0 MB 10.3 MB/s eta
0:00:10
- ----- 8.0/101.0 MB 10.3 MB/s eta
0:00:10
- ----- 8.6/101.0 MB 10.4 MB/s eta
0:00:09
- ----- 9.1/101.0 MB 10.5 MB/s eta
0:00:09
- ----- 9.5/101.0 MB 10.7 MB/s eta
0:00:09
- ----- 10.0/101.0 MB 10.7 MB/s
eta 0:00:09
- ----- 10.5/101.0 MB 10.6 MB/s
eta 0:00:09
- ----- 10.9/101.0 MB 10.4 MB/s
eta 0:00:09
- ----- 11.5/101.0 MB 10.4 MB/s
eta 0:00:09
- ----- 12.1/101.0 MB 10.6 MB/s
eta 0:00:09
- ----- 12.6/101.0 MB 10.6 MB/s
eta 0:00:09
- ----- 13.2/101.0 MB 10.6 MB/s
eta 0:00:09
- ----- 13.6/101.0 MB 10.6 MB/s
eta 0:00:09
- ----- 14.2/101.0 MB 10.7 MB/s
eta 0:00:09
- ----- 15.3/101.0 MB 10.9 MB/s
eta 0:00:08
```

-----	15.9/101.0 MB 10.7 MB/s
eta 0:00:08	
-----	16.4/101.0 MB 10.7 MB/s
eta 0:00:08	
-----	16.9/101.0 MB 11.9 MB/s
eta 0:00:08	
-----	17.4/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	18.0/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	18.5/101.0 MB 12.4 MB/s
eta 0:00:07	
-----	18.9/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	19.9/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	20.4/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	21.2/101.0 MB 12.3 MB/s
eta 0:00:07	
-----	21.7/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	22.2/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	22.7/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	23.3/101.0 MB 11.9 MB/s
eta 0:00:07	
-----	23.7/101.0 MB 11.9 MB/s
eta 0:00:07	
-----	24.2/101.0 MB 12.1 MB/s
eta 0:00:07	
-----	24.8/101.0 MB 11.9 MB/s
eta 0:00:07	
-----	25.4/101.0 MB 11.7 MB/s
eta 0:00:07	
-----	25.9/101.0 MB 11.9 MB/s
eta 0:00:07	
-----	26.3/101.0 MB 11.7 MB/s
eta 0:00:07	
-----	26.8/101.0 MB 11.7 MB/s
eta 0:00:07	
-----	27.3/101.0 MB 11.5 MB/s
eta 0:00:07	
-----	27.8/101.0 MB 11.5 MB/s
eta 0:00:07	
-----	28.4/101.0 MB 11.7 MB/s
eta 0:00:07	
-----	28.8/101.0 MB 11.5 MB/s



eta 0:00:07	-----	29.3/101.0 MB 11.7 MB/s
eta 0:00:07	-----	29.9/101.0 MB 11.9 MB/s
eta 0:00:06	-----	30.4/101.0 MB 11.7 MB/s
eta 0:00:07	-----	30.8/101.0 MB 11.7 MB/s
eta 0:00:07	-----	31.4/101.0 MB 11.5 MB/s
eta 0:00:07	-----	31.9/101.0 MB 11.5 MB/s
eta 0:00:07	-----	32.5/101.0 MB 11.7 MB/s
eta 0:00:06	-----	33.0/101.0 MB 11.5 MB/s
eta 0:00:06	-----	33.6/101.0 MB 11.7 MB/s
eta 0:00:06	-----	34.0/101.0 MB 11.5 MB/s
eta 0:00:06	-----	34.7/101.0 MB 11.5 MB/s
eta 0:00:06	-----	35.2/101.0 MB 11.7 MB/s
eta 0:00:06	-----	35.7/101.0 MB 11.5 MB/s
eta 0:00:06	-----	36.3/101.0 MB 11.5 MB/s
eta 0:00:06	-----	36.9/101.0 MB 11.7 MB/s
eta 0:00:06	-----	37.5/101.0 MB 11.9 MB/s
eta 0:00:06	-----	38.0/101.0 MB 11.7 MB/s
eta 0:00:06	-----	38.5/101.0 MB 11.7 MB/s
eta 0:00:06	-----	39.0/101.0 MB 11.5 MB/s
eta 0:00:06	-----	39.6/101.0 MB 11.7 MB/s
eta 0:00:06	-----	40.1/101.0 MB 11.5 MB/s
eta 0:00:06	-----	40.6/101.0 MB 11.7 MB/s
eta 0:00:06	-----	41.1/101.0 MB 11.7 MB/s
eta 0:00:06	-----	41.5/101.0 MB 11.7 MB/s
eta 0:00:06		

-----	42.0/101.0 MB	11.7 MB/s	
eta 0:00:06			
-----	42.3/101.0 MB	11.3 MB/s	
eta 0:00:06			
-----	43.0/101.0 MB	11.7 MB/s	
eta 0:00:05			
-----	43.4/101.0 MB	11.5 MB/s	
eta 0:00:06			
-----	43.9/101.0 MB	11.3 MB/s	
eta 0:00:06			
-----	44.4/101.0 MB	11.5 MB/s	
eta 0:00:05			
-----	44.9/101.0 MB	11.5 MB/s	
eta 0:00:05			
-----	45.5/101.0 MB	11.5 MB/s	
eta 0:00:05			
-----	46.0/101.0 MB	11.7 MB/s	
eta 0:00:05			
-----	46.5/101.0 MB	11.7 MB/s	
eta 0:00:05			
-----	47.0/101.0 MB	11.7 MB/s	
eta 0:00:05			
-----	47.6/101.0 MB	11.5 MB/s	
eta 0:00:05			
-----	47.8/101.0 MB	11.5 MB/s	
eta 0:00:05			
-----	47.9/101.0 MB	10.7 MB/s	
eta 0:00:05			
-----	48.1/101.0 MB	10.6 MB/s	
eta 0:00:06			
-----	48.4/101.0 MB	10.2 MB/s	
eta 0:00:06			
-----	48.5/101.0 MB	9.9 MB/s	eta
0:00:06			
-----	48.7/101.0 MB	9.6 MB/s	eta
0:00:06			
-----	49.2/101.0 MB	9.8 MB/s	eta
0:00:06			
-----	49.7/101.0 MB	9.5 MB/s	eta
0:00:06			
-----	50.3/101.0 MB	9.8 MB/s	eta
0:00:06			
-----	50.8/101.0 MB	9.6 MB/s	eta
0:00:06			
-----	51.3/101.0 MB	9.6 MB/s	eta
0:00:06			
-----	51.8/101.0 MB	9.8 MB/s	eta
0:00:06			
-----	52.3/101.0 MB	9.8 MB/s	eta

0:00:05	-----	52.8/101.0 MB 9.8 MB/s eta
0:00:05	-----	53.3/101.0 MB 9.8 MB/s eta
0:00:05	-----	53.6/101.0 MB 9.6 MB/s eta
0:00:05	-----	53.8/101.0 MB 9.5 MB/s eta
0:00:05	-----	54.1/101.0 MB 9.2 MB/s eta
0:00:06	-----	54.8/101.0 MB 9.2 MB/s eta
0:00:06	-----	55.2/101.0 MB 9.2 MB/s eta
0:00:05	-----	55.7/101.0 MB 9.2 MB/s eta
0:00:05	-----	55.9/101.0 MB 9.0 MB/s eta
0:00:06	-----	56.1/101.0 MB 8.8 MB/s eta
0:00:06	-----	56.5/101.0 MB 8.7 MB/s eta
0:00:06	-----	56.9/101.0 MB 8.6 MB/s eta
0:00:06	-----	57.5/101.0 MB 8.7 MB/s eta
0:00:05	-----	58.1/101.0 MB 8.8 MB/s eta
0:00:05	-----	58.6/101.0 MB 9.8 MB/s eta
0:00:05	-----	59.2/101.0 MB 10.1 MB/s
eta 0:00:05	-----	59.7/101.0 MB 10.1 MB/s
eta 0:00:05	-----	60.3/101.0 MB 10.1 MB/s
eta 0:00:05	-----	60.8/101.0 MB 10.1 MB/s
eta 0:00:04	-----	61.2/101.0 MB 10.1 MB/s
eta 0:00:04	-----	61.7/101.0 MB 10.1 MB/s
eta 0:00:04	-----	62.0/101.0 MB 9.8 MB/s eta
0:00:04	-----	62.5/101.0 MB 9.9 MB/s eta
0:00:04	-----	63.1/101.0 MB 9.8 MB/s eta
0:00:04		

-----	63.4/101.0 MB 9.8 MB/s eta
0:00:04	
-----	64.1/101.0 MB 10.6 MB/s
eta 0:00:04	
-----	64.5/101.0 MB 10.4 MB/s
eta 0:00:04	
-----	65.0/101.0 MB 10.6 MB/s
eta 0:00:04	
-----	65.5/101.0 MB 10.4 MB/s
eta 0:00:04	
-----	66.0/101.0 MB 10.6 MB/s
eta 0:00:04	
-----	66.6/101.0 MB 11.3 MB/s
eta 0:00:04	
-----	67.0/101.0 MB 11.1 MB/s
eta 0:00:04	
-----	67.5/101.0 MB 11.1 MB/s
eta 0:00:04	
-----	68.1/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	68.7/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	69.2/101.0 MB 11.3 MB/s
eta 0:00:03	
-----	69.7/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	70.2/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	70.5/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	71.2/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	71.6/101.0 MB 11.3 MB/s
eta 0:00:03	
-----	72.2/101.0 MB 11.5 MB/s
eta 0:00:03	
-----	72.7/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	73.2/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	73.8/101.0 MB 11.3 MB/s
eta 0:00:03	
-----	74.2/101.0 MB 11.3 MB/s
eta 0:00:03	
-----	74.7/101.0 MB 11.1 MB/s
eta 0:00:03	
-----	75.3/101.0 MB 11.3 MB/s
eta 0:00:03	
-----	75.8/101.0 MB 11.3 MB/s

eta 0:00:03	-----	76.3/101.0 MB 11.3 MB/s
eta 0:00:03	-----	76.8/101.0 MB 11.1 MB/s
eta 0:00:03	-----	77.3/101.0 MB 11.3 MB/s
eta 0:00:03	-----	77.7/101.0 MB 11.1 MB/s
eta 0:00:03	-----	78.3/101.0 MB 11.1 MB/s
eta 0:00:03	-----	78.9/101.0 MB 11.1 MB/s
eta 0:00:02	-----	79.3/101.0 MB 11.3 MB/s
eta 0:00:02	-----	79.8/101.0 MB 11.1 MB/s
eta 0:00:02	-----	80.4/101.0 MB 11.1 MB/s
eta 0:00:02	-----	80.9/101.0 MB 11.5 MB/s
eta 0:00:02	-----	81.4/101.0 MB 11.1 MB/s
eta 0:00:02	-----	81.9/101.0 MB 11.3 MB/s
eta 0:00:02	-----	82.5/101.0 MB 11.3 MB/s
eta 0:00:02	-----	83.0/101.0 MB 11.1 MB/s
eta 0:00:02	-----	83.5/101.0 MB 11.3 MB/s
eta 0:00:02	-----	83.8/101.0 MB 10.9 MB/s
eta 0:00:02	-----	84.3/101.0 MB 10.7 MB/s
eta 0:00:02	-----	84.8/101.0 MB 11.1 MB/s
eta 0:00:02	-----	85.4/101.0 MB 11.1 MB/s
eta 0:00:02	-----	85.7/101.0 MB 10.9 MB/s
eta 0:00:02	-----	86.3/101.0 MB 10.9 MB/s
eta 0:00:02	-----	86.7/101.0 MB 10.9 MB/s
eta 0:00:02	-----	87.2/101.0 MB 10.9 MB/s
eta 0:00:02	-----	87.7/101.0 MB 10.9 MB/s
eta 0:00:02		

-----	88.3/101.0 MB	10.9 MB/s
eta 0:00:02		
-----	88.7/101.0 MB	11.1 MB/s
eta 0:00:02		
-----	89.3/101.0 MB	10.9 MB/s
eta 0:00:02		
-----	89.8/101.0 MB	11.1 MB/s
eta 0:00:02		
-----	90.1/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	90.6/101.0 MB	11.1 MB/s
eta 0:00:01		
-----	91.1/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	91.6/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	92.1/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	92.5/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	92.9/101.0 MB	10.6 MB/s
eta 0:00:01		
-----	93.4/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	93.9/101.0 MB	10.6 MB/s
eta 0:00:01		
-----	94.3/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	94.4/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	95.2/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	95.7/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	96.2/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	96.6/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	97.2/101.0 MB	10.9 MB/s
eta 0:00:01		
-----	97.6/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	98.1/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	98.5/101.0 MB	10.7 MB/s
eta 0:00:01		
-----	98.8/101.0 MB	10.4 MB/s
eta 0:00:01		
-----	99.3/101.0 MB	10.6 MB/s
eta 0:00:01		

```

----- 99.7/101.0 MB 10.2 MB/s
eta 0:00:01
----- 100.1/101.0 MB 10.2 MB/s
eta 0:00:01
----- 100.4/101.0 MB 10.2 MB/s
eta 0:00:01
----- 100.7/101.0 MB 9.9 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 9.6 MB/s
eta 0:00:01
----- 101.0/101.0 MB 6.7 MB/s
eta 0:00:00
Downloading plotly-5.17.0-py2.py3-none-any.whl (15.6 MB)
----- 0.0/15.6 MB ? eta -:--:--
----- 0.2/15.6 MB 6.3 MB/s eta
0:00:03
----- 0.6/15.6 MB 6.5 MB/s eta
0:00:03
----- 1.0/15.6 MB 7.3 MB/s eta
0:00:02
----- 1.4/15.6 MB 7.4 MB/s eta
0:00:02
----- 1.9/15.6 MB 8.3 MB/s eta
0:00:02
----- 2.3/15.6 MB 8.1 MB/s eta
0:00:02
----- 2.8/15.6 MB 8.6 MB/s eta
0:00:02
----- 3.3/15.6 MB 9.1 MB/s eta
0:00:02
----- 3.7/15.6 MB 9.0 MB/s eta
0:00:02
----- 4.2/15.6 MB 9.3 MB/s eta

```

```
0:00:02
----- 4.7/15.6 MB 9.4 MB/s eta
0:00:02
----- 5.3/15.6 MB 9.6 MB/s eta
0:00:02
----- 5.7/15.6 MB 9.6 MB/s eta
0:00:02
----- 6.2/15.6 MB 9.6 MB/s eta
0:00:01
----- 6.6/15.6 MB 9.5 MB/s eta
0:00:01
----- 7.0/15.6 MB 9.7 MB/s eta
0:00:01
----- 7.2/15.6 MB 9.8 MB/s eta
0:00:01
----- 7.9/15.6 MB 9.7 MB/s eta
0:00:01
----- 8.4/15.6 MB 9.6 MB/s eta
0:00:01
----- 8.9/15.6 MB 9.9 MB/s eta
0:00:01
----- 9.5/15.6 MB 9.7 MB/s eta
0:00:01
----- 9.9/15.6 MB 9.8 MB/s eta
0:00:01
----- 10.5/15.6 MB 10.1 MB/s eta
0:00:01
----- 11.0/15.6 MB 10.2 MB/s eta
0:00:01
----- 11.4/15.6 MB 10.4 MB/s eta
0:00:01
----- 12.0/15.6 MB 10.4 MB/s eta
0:00:01
----- 12.3/15.6 MB 10.4 MB/s eta
0:00:01
----- 12.8/15.6 MB 10.2 MB/s eta
0:00:01
----- 13.2/15.6 MB 10.2 MB/s eta
0:00:01
----- 13.7/15.6 MB 10.1 MB/s eta
0:00:01
----- 14.2/15.6 MB 10.1 MB/s eta
0:00:01
----- 14.6/15.6 MB 9.9 MB/s eta
0:00:01
----- 15.1/15.6 MB 10.1 MB/s eta
0:00:01
----- 15.6/15.6 MB 10.2 MB/s eta
0:00:01
```



```
----- 15.6/15.6 MB 9.6 MB/s eta
0:00:00
Downloading tenacity-8.2.3-py3-none-any.whl (24 kB)
Installing collected packages: tenacity, graphviz, plotly, lightgbm,
catboost
Successfully installed catboost-1.2.1.1 graphviz-0.20.1 lightgbm-4.1.0
plotly-5.17.0 tenacity-8.2.3

from lightgbm import LGBMClassifier
from catboost import CatBoostClassifier

from lightgbm import LGBMClassifier
model = LGBMClassifier(verbose = -1)
classify(model)

Accuracy:  0.8116591928251121
CV Score:  0.8238277572029377

from catboost import CatBoostClassifier
model = CatBoostClassifier(verbose = 0)
classify(model)

Accuracy:  0.8295964125560538
CV Score:  0.8226790534178645
```