

titanic-survival-prediction-task-2

February 16, 2024

TITANIC SURVIVAL PREDICTION

Importing required libraries

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

Read the data set

```
[ ]: df = pd.read_csv("/content/sample_data/titanic data set.csv")
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]

```
[ ]: df.head(6)
```

```
[ ]: PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
5              6         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	
5	Moran, Mr. James	male	NaN	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
5	0	330877	8.4583	NaN	Q

```
[ ]: df.tail(10)
```

```
[ ]:      PassengerId  Survived  Pclass                                Name \
881          882         0         3                        Markun, Mr. Johann
882          883         0         3          Dahlberg, Miss. Gerda Ulrika
883          884         0         2      Banfield, Mr. Frederick James
884          885         0         3          Sutehall, Mr. Henry Jr
885          886         0         3      Rice, Mrs. William (Margaret Norton)
886          887         0         2          Montvila, Rev. Juozas
887          888         1         1          Graham, Miss. Margaret Edith
888          889         0         3  Johnston, Miss. Catherine Helen "Carrie"
889          890         1         1          Behr, Mr. Karl Howell
890          891         0         3          Dooley, Mr. Patrick
```

```
      Sex  Age  SibSp  Parch            Ticket     Fare Cabin Embarked
881  male  33.0     0     0          349257     7.8958   NaN        S
882  female  22.0     0     0             7552    10.5167   NaN        S
883  male  28.0     0     0  C.A./SOTON 34068    10.5000   NaN        S
884  male  25.0     0     0  SOTON/OQ 392076     7.0500   NaN        S
885  female  39.0     0     5          382652    29.1250   NaN        Q
886  male  27.0     0     0          211536    13.0000   NaN        S
887  female  19.0     0     0          112053    30.0000   B42        S
888  female   NaN     1     2      W./C. 6607    23.4500   NaN        S
889  male  26.0     0     0          111369    30.0000  C148        C
890  male  32.0     0     0          370376     7.7500   NaN        Q
```

Getting information about the data

```
[ ]: df.shape
```

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

```
[ ]: 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: 83.7+ KB

```

```
[ ]: df.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

```

```
[ ]: df.mode()
```

```

[ ]:
      PassengerId  Survived  Pclass                                Name  \
0                1         0.0      3.0                      Abbing, Mr. Anthony
1                2         NaN      NaN                      Abbott, Mr. Rossmore Edward
2                3         NaN      NaN          Abbott, Mrs. Stanton (Rosa Hunt)
3                4         NaN      NaN                      Abelson, Mr. Samuel
4                5         NaN      NaN  Abelson, Mrs. Samuel (Hannah Wozosky)
..              ...         ...      ...                                ...
886             887         NaN      NaN                      de Mulder, Mr. Theodore
887             888         NaN      NaN                      de Pelsmaeker, Mr. Alfons
888             889         NaN      NaN                      del Carlo, Mr. Sebastiano
889             890         NaN      NaN          van Billiard, Mr. Austin Blyler
890             891         NaN      NaN          van Melkebeke, Mr. Philemon

      Sex  SibSp  Parch  Ticket  Fare  Cabin Embarked
0    1.0    0.0    0.0    1601   8.05    B96 B98      S
1    NaN    NaN    NaN    347082   NaN  C23 C25 C27    NaN
2    NaN    NaN    NaN    CA. 2343   NaN      G6    NaN
3    NaN    NaN    NaN      NaN    NaN      NaN    NaN
4    NaN    NaN    NaN      NaN    NaN      NaN    NaN

```

```

..   ...   ...   ...   ...   ...   ...   ...
886  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
887  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
888  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
889  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
890  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN

```

[891 rows x 11 columns]

```
[ ]: df.std()
```

```

[ ]: PassengerId    257.353842
     Survived       0.486592
     Pclass         0.836071
     Sex            0.477990
     SibSp          1.102743
     Parch          0.806057
     Fare           49.693429
     dtype: float64

```

Checking missing values

```
[ ]: df.isnull().sum()
```

```

[ ]: 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

```

```
[ ]: df.isna().sum()
```

```

[ ]: PassengerId    0
     Survived       0
     Pclass         0
     Name           0
     Sex            0
     SibSp          0
     Parch          0

```

```
Ticket      0
Fare        0
Cabin      687
Embarked    2
dtype: int64
```

```
[ ]: df['Survived'].value_counts()
```

```
[ ]: 0    549
     1    342
     Name: Survived, dtype: int64
```

```
[ ]: df.value_counts("Parch")
```

```
[ ]: Parch
     0    678
     1    118
     2     80
     3      5
     5      5
     4      4
     6      1
     dtype: int64
```

```
[ ]: df.value_counts("Sex")
```

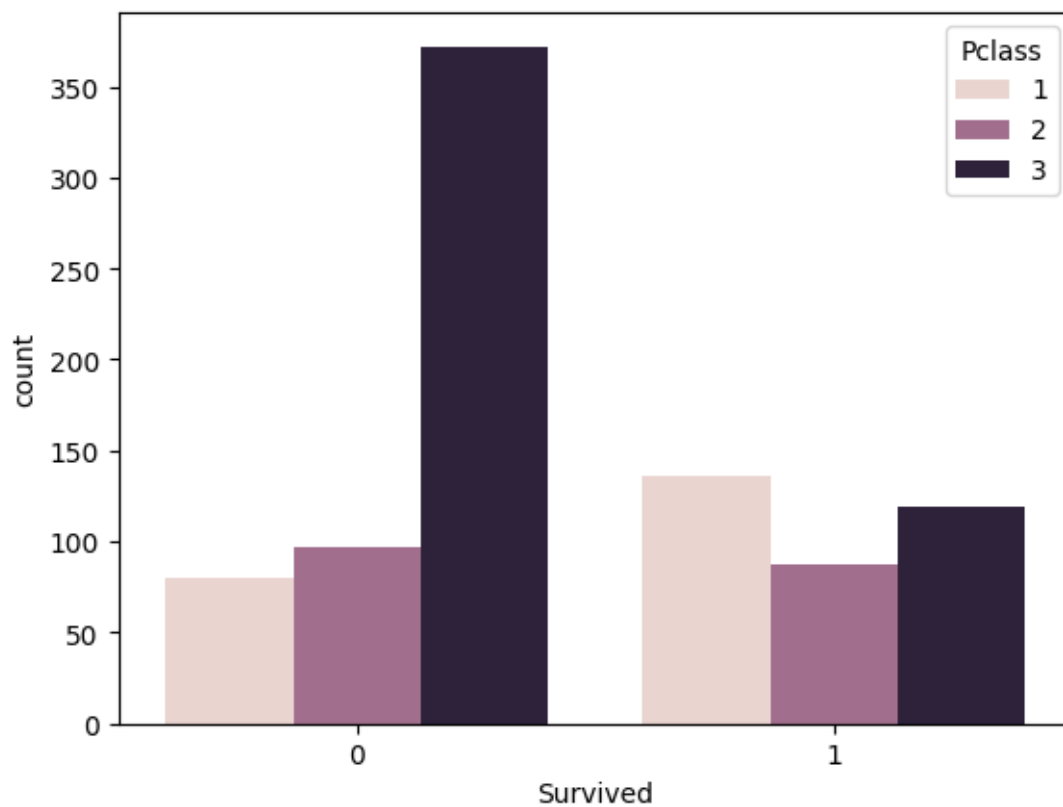
```
[ ]: Sex
     1    577
     0    314
     dtype: int64
```

visualization

count plot

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

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

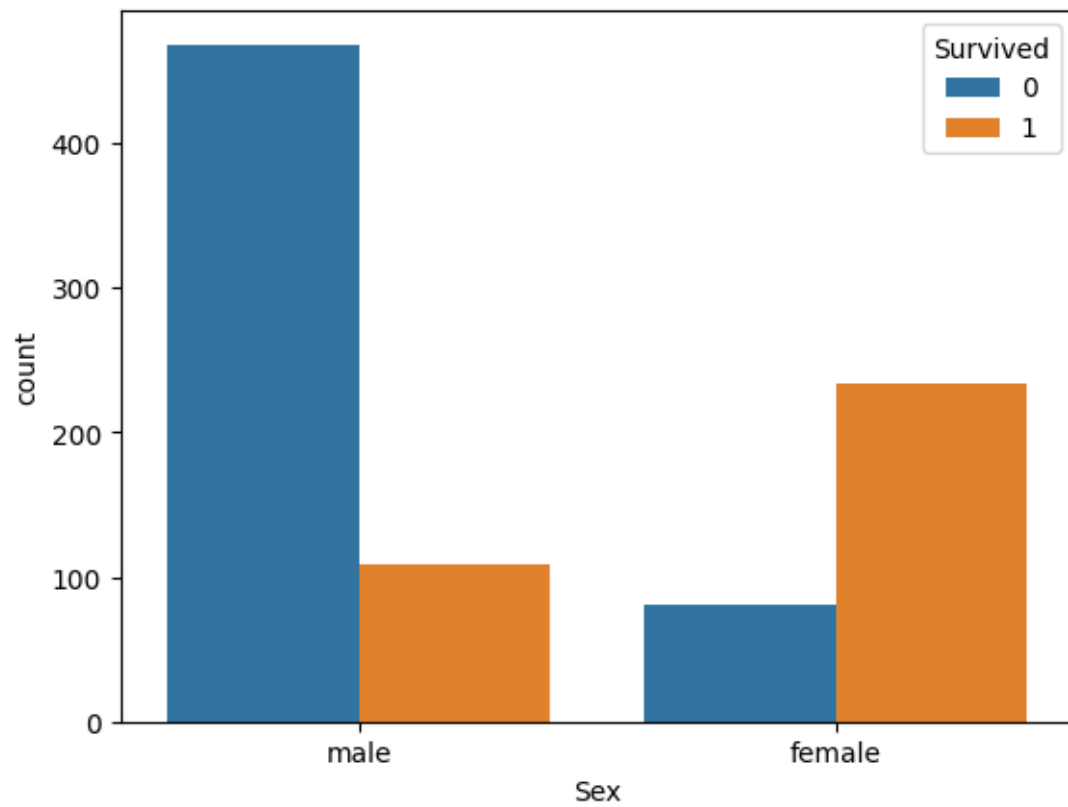


```
[ ]: df["Sex"]
```

```
[ ]: 0      male
      1      female
      2      female
      3      female
      4      male
      ...
      886     male
      887     female
      888     female
      889     male
      890     male
      Name: Sex, Length: 891, dtype: object
```

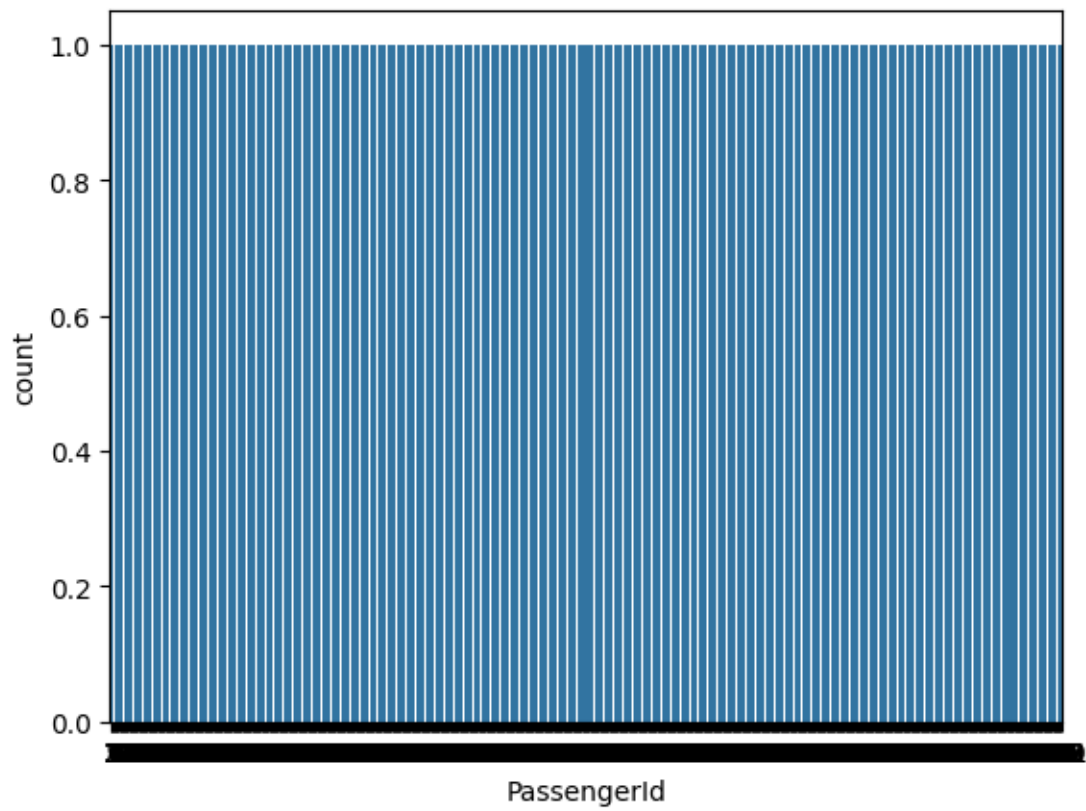
```
[ ]: sns.countplot(x=df['Sex'],hue=df['Survived'])
```

```
[ ]: <Axes: xlabel='Sex', ylabel='count'>
```

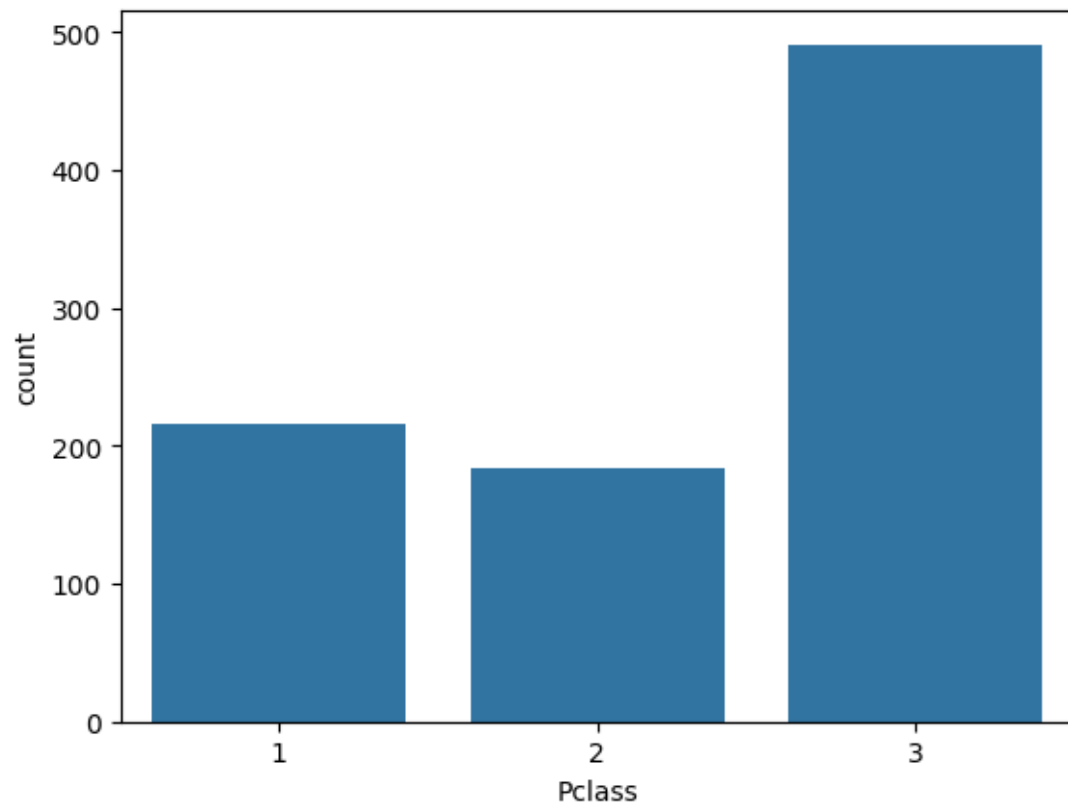


```
[ ]: sns.countplot(x='PassengerId',data=df)
```

```
[ ]: <Axes: xlabel='PassengerId', ylabel='count'>
```

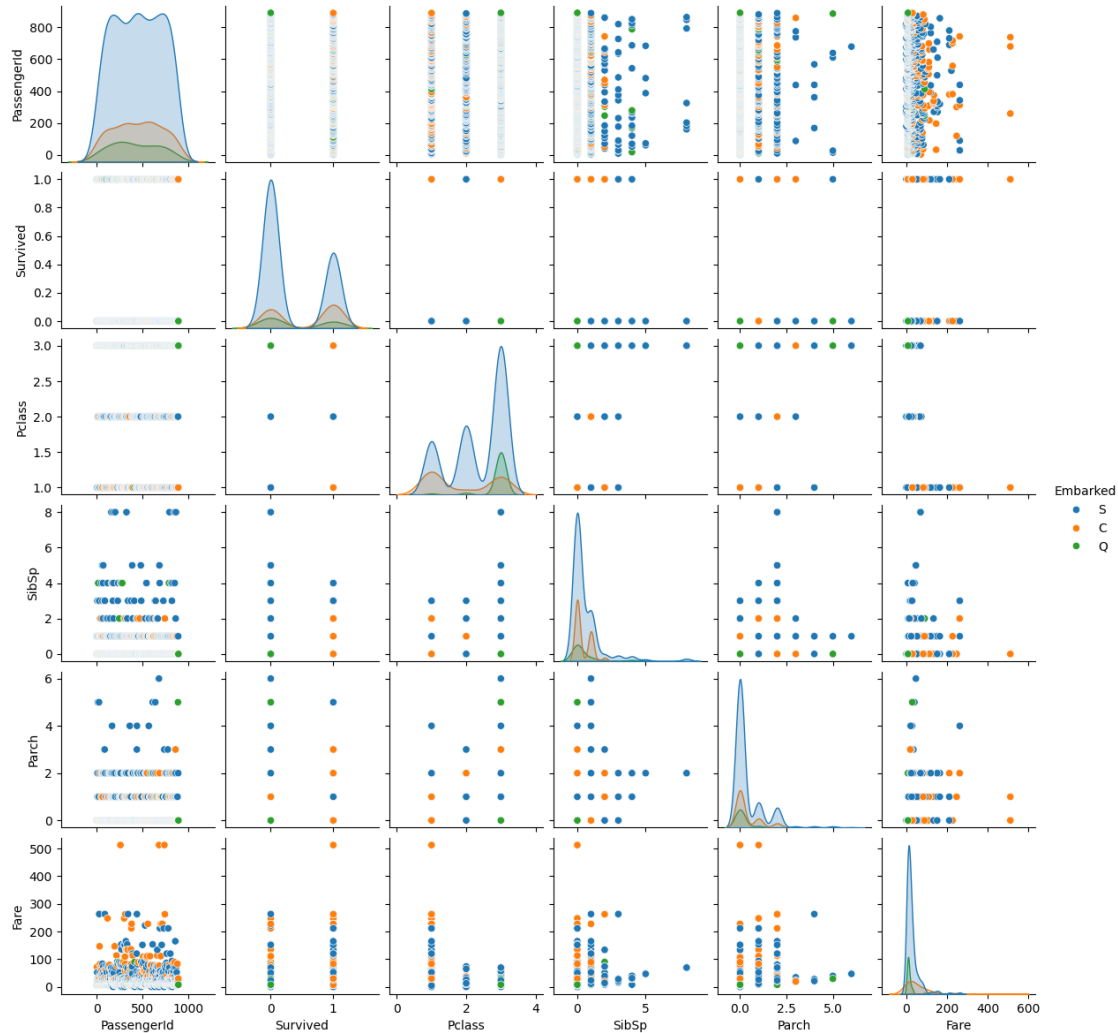
```
[ ]: sns.countplot(x='Pclass',data=df)
plt.show()
```



Pair plot

```
[ ]: sns.pairplot(df.drop(['Sex'], axis = 1), hue="Embarked", height=2)
```

```
[ ]: <seaborn.axisgrid.PairGrid at 0x7dd2b7f3d900>
```



```
[ ]: df.groupby('Sex')[['Survived']].mean()
```

```
[ ]:      Survived
Sex
female  0.742038
male    0.188908
```

<google.colab._quickchart_helpers.SectionTitle at 0x7dd2b470f040>

```
from matplotlib import pyplot as plt
_df_2['Survived'].plot(kind='hist', bins=20, title='Survived')
plt.gca().spines[['top', 'right']].set_visible(False)
```

<google.colab._quickchart_helpers.SectionTitle at 0x7dd2b470f190>

```
from matplotlib import pyplot as plt
_df_3['Survived'].plot(kind='line', figsize=(8, 4), title='Survived')
```

```
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
[ ]: df['Sex'].unique()
```

```
[ ]: array(['male', 'female'], dtype=object)
```

```
[ ]: from sklearn.preprocessing import LabelEncoder
labelencoder = LabelEncoder()
df['Sex']=labelencoder.fit_transform(df['Sex'])
df.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  Parch  \
0                        Braund, Mr. Owen Harris    1  22.0    1    0
1  Cumings, Mrs. John Bradley (Florence Briggs Th...    0  38.0    1    0
2                        Heikkinen, Miss. Laina    0  26.0    0    0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    0  35.0    1    0
4                        Allen, Mr. William Henry    1  35.0    0    0
```

```
                                Ticket    Fare  Cabin  Embarked
0          A/5 21171    7.2500   NaN      S
1          PC 17599   71.2833   C85      C
2  STON/O2. 3101282    7.9250   NaN      S
3          113803   53.1000  C123      S
4          373450    8.0500   NaN      S
```

```
[ ]: df['Sex'],df['Survived']
```

```
[ ]: (0      1
      1      0
      2      0
      3      0
      4      1
      ..
      886    1
      887    0
      888    0
      889    1
      890    1
      Name: Sex, Length: 891, dtype: int64,
      0      0
```

```

1      1
2      1
3      1
4      0
..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64)

```

```
[ ]: df=df.drop(['Age'],axis=1)
```

```
[ ]: df_final= df
df_final.head(10)
```

```
[ ]:
 PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3
5            6         0       3
6            7         0       1
7            8         0       3
8            9         1       3
9           10         1       2

```

```

                                Name  Sex  SibSp  Parch  \
0                        Braund, Mr. Owen Harris    1     1     0
1  Cumings, Mrs. John Bradley (Florence Briggs Th...    0     1     0
2                        Heikkinen, Miss. Laina    0     0     0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    0     1     0
4                        Allen, Mr. William Henry    1     0     0
5                        Moran, Mr. James          1     0     0
6                        McCarthy, Mr. Timothy J    1     0     0
7                        Palsson, Master. Gosta Leonard    1     3     1
8  Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)    0     0     2
9                        Nasser, Mrs. Nicholas (Adele Achem)    0     1     0

```

```

      Ticket      Fare Cabin Embarked
0    A/5 21171   7.2500   NaN        S
1    PC 17599  71.2833   C85        C
2  STON/O2. 3101282   7.9250   NaN        S
3      113803  53.1000  C123        S
4      373450   8.0500   NaN        S

```

5	330877	8.4583	NaN	Q
6	17463	51.8625	E46	S
7	349909	21.0750	NaN	S
8	347742	11.1333	NaN	S
9	237736	30.0708	NaN	C

Handling correlation

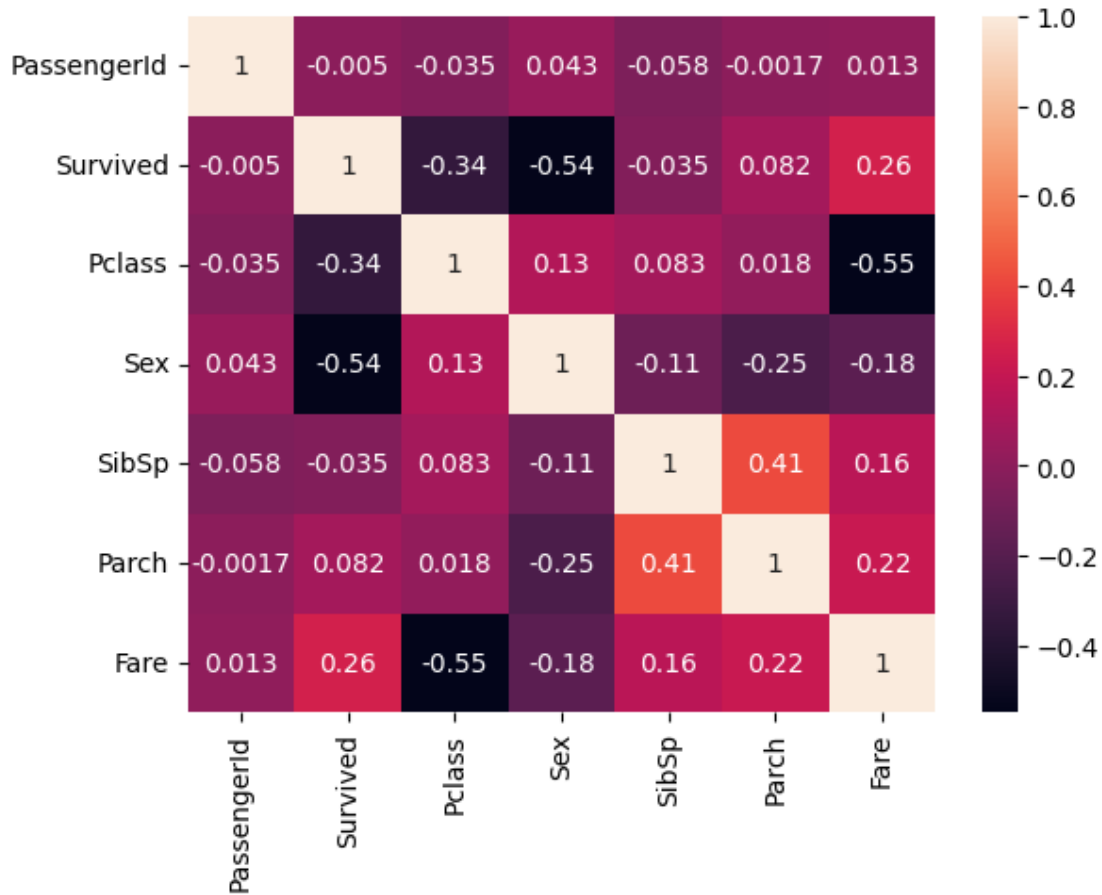
```
[ ]: df.corr(method='pearson')
```

```
[ ]:
      PassengerId  Survived  Pclass    Sex  SibSp  Parch  \
PassengerId      1.000000 -0.005007 -0.035144  0.042939 -0.057527 -0.001652
Survived         -0.005007  1.000000 -0.338481 -0.543351 -0.035322  0.081629
Pclass           -0.035144 -0.338481  1.000000  0.131900  0.083081  0.018443
Sex               0.042939 -0.543351  0.131900  1.000000 -0.114631 -0.245489
SibSp            -0.057527 -0.035322  0.083081 -0.114631  1.000000  0.414838
Parch            -0.001652  0.081629  0.018443 -0.245489  0.414838  1.000000
Fare              0.012658  0.257307 -0.549500 -0.182333  0.159651  0.216225
```

	Fare
PassengerId	0.012658
Survived	0.257307
Pclass	-0.549500
Sex	-0.182333
SibSp	0.159651
Parch	0.216225
Fare	1.000000

Heat map

```
[ ]: sns.heatmap(df.corr(method='pearson').drop([], axis=1).drop([], axis=0),
                annot = True);
plt.show()
```



Model trianing

```
[ ]: X=df[['Pclass','Sex']]
      Y=df['Survived']
      from sklearn.model_selection import train_test_split
      X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=0)
```

```
[ ]: from sklearn.linear_model import LogisticRegression
      log=LogisticRegression(random_state=0)
      log.fit(X_train,Y_train)
```

```
[ ]: LogisticRegression(random_state=0)
```

MODEL PREDICTION

```
[ ]: pred = print(log.predict(X_test))
```

```
[0 0 0 1 1 0 1 1 0 1 0 1 0 1 1 1 0 0 0 0 0 1 0 0 1 1 0 1 1 1 0 1 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 1 1 0 1 0 1 0 1 1 1 0 0 0
0 1 0 0 0 0 0 0 1 0 0 1 1 1 1 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 1 1 1 1 0 1 0
```

```

1 0 1 0 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 1 1 1 0 1
1 0 0 1 1 0 1 0 1 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0]

```

```
[ ]: print(Y_test)
```

```

495    0
648    0
278    0
31     1
255    1
..
780    1
837    0
215    1
833    0
372    0
Name: Survived, Length: 179, dtype: int64

```

```
[ ]: import warnings
warnings.filterwarnings("ignore")
res=log.predict([[2,0]])
if(res==0):
    print("So Sorry ! Not Survived ")
else:
    print("Survived!")
```

Survived!

```
[ ]: import warnings
warnings.filterwarnings("ignore")
res=log.predict([[2,2]])
if(res==0):
    print("So Sorry ! Not Survived ")
else:
    print("Survived!")
```

So Sorry ! Not Survived