k.likhith

2211cs010309

Titanic DataSet

Titanic Dataset Description

The dataset consists of 891 passenger records from the Titanic disaster, with 12 attributes describing passengers, their demographics, and survival status.

Key Columns:

PassengerId – Unique identifier for each passenger.

Survived – 1 if the passenger survived, 0 if they did not.

Pclass – Passenger class (1st, 2nd, or 3rd).

Name – Full name of the passenger.

Sex – Gender of the passenger.

Age – Age (some missing values).

SibSp & Parch – Number of siblings/spouses and parents/children aboard.

Ticket – Ticket number.

Fare – Ticket fare.

Cabin – Cabin number (many missing values).

Embarked – Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

Import Module

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings. filterwarnings( 'ignore')
%matplotlib inline

df=pd.read_csv('Titanic-Dataset.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]
```

Loading the dataset

```
train= pd.read_csv('train.csv')
test = pd.read_csv('test (1).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  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 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
```

```
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
```

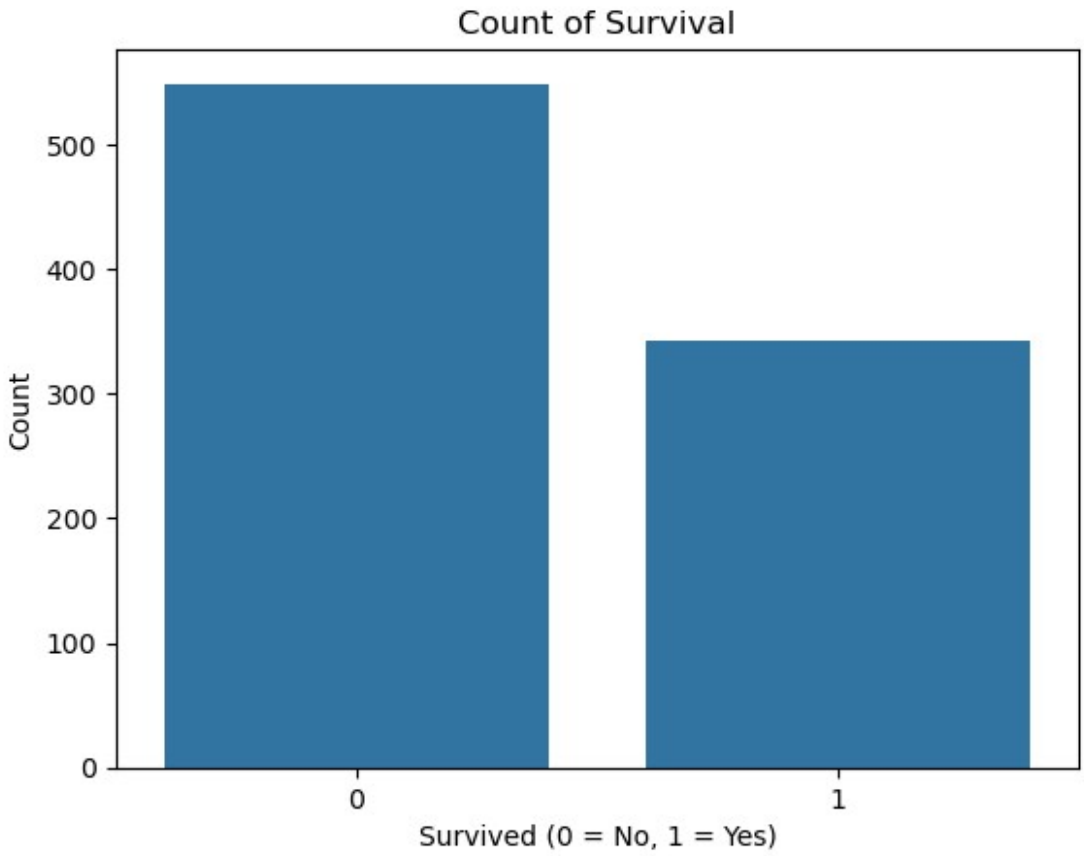
```
25%       0.000000    7.910400
50%       0.000000   14.454200
75%       0.000000   31.000000
max       6.000000  512.329200
```

statistical info

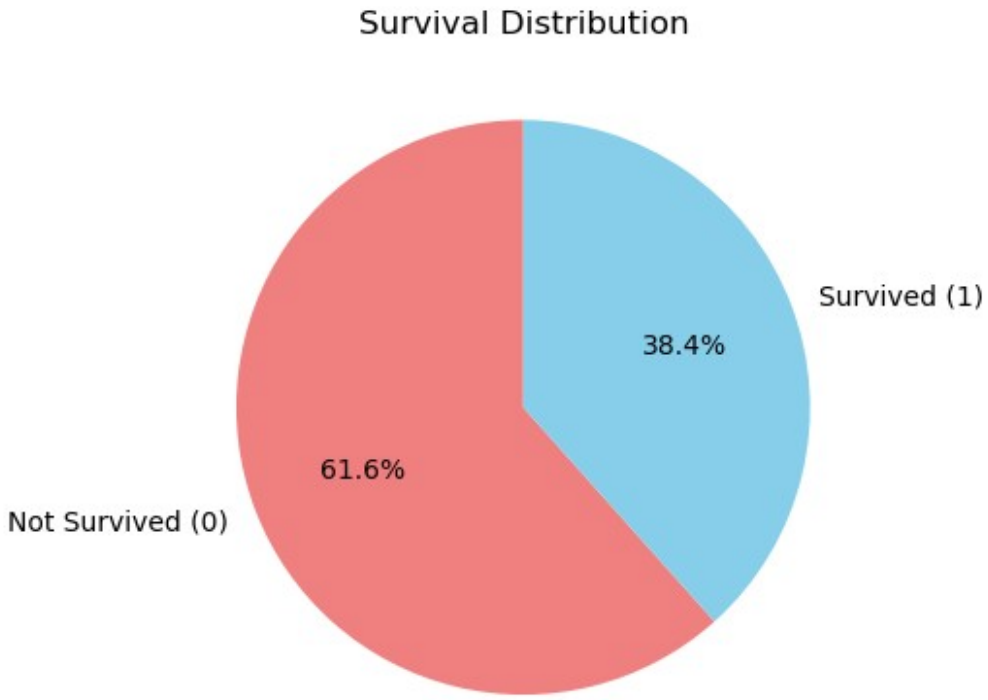
```
train.info()

<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

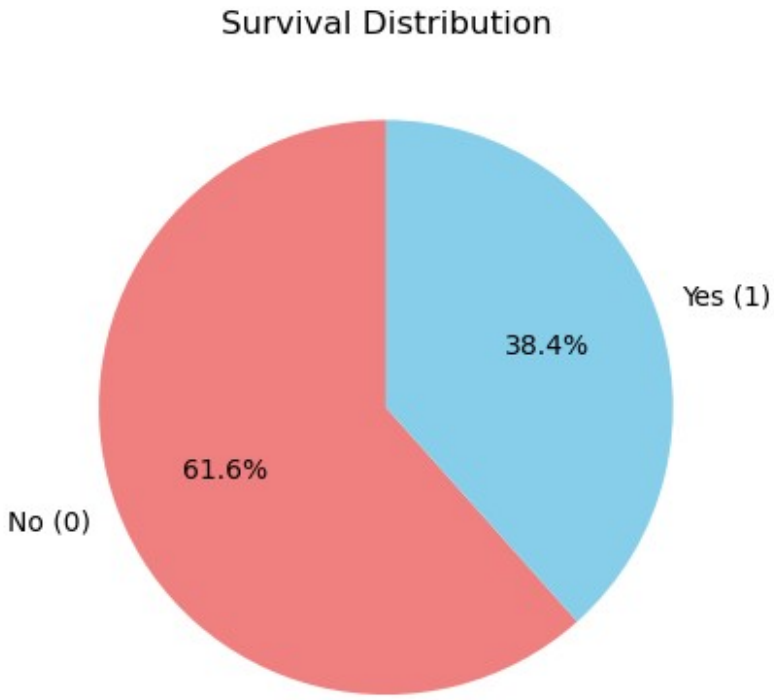
```
df=pd.read_csv('Titanic-Dataset.csv')
train= pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Survived', data=train)
plt.title("Count of Survival")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Count")
plt.show()
```



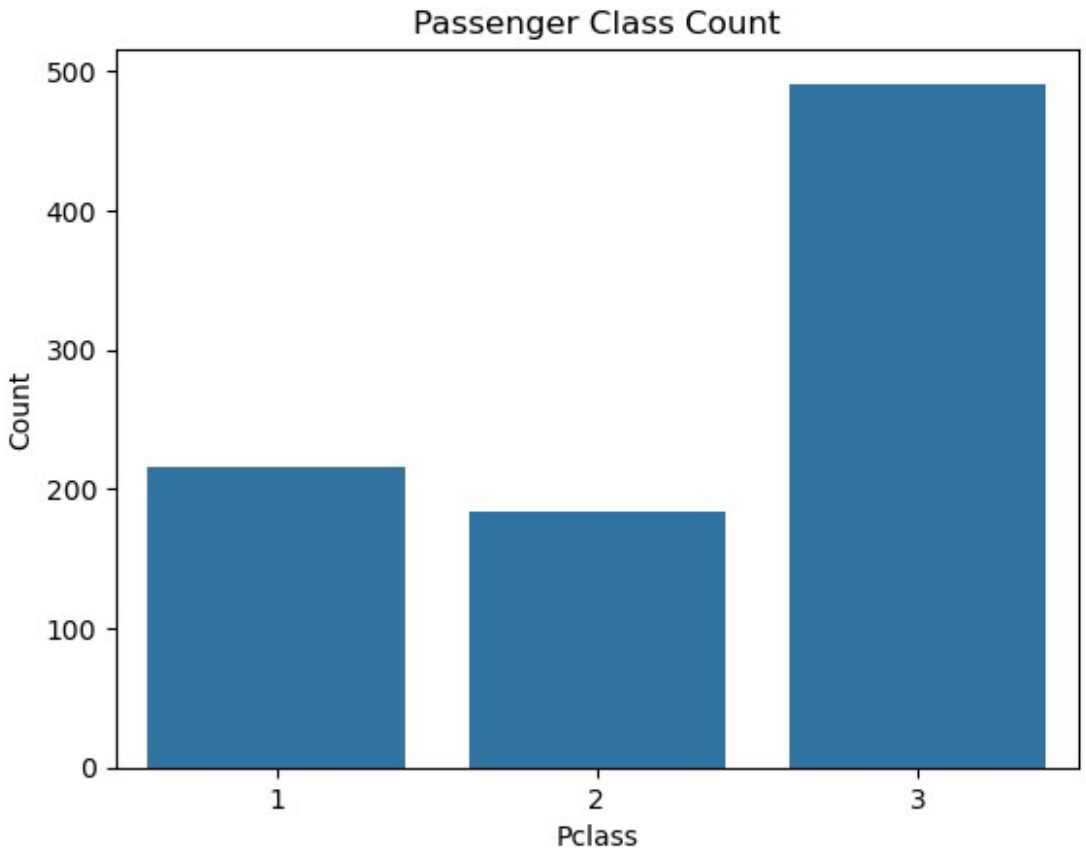

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Titanic-Dataset.csv')
train = pd.read_csv('Titanic-Dataset.csv')

survival_counts = train['Survived'].value_counts()

survival_counts.plot(
    kind='pie',
    autopct='%1.1f%%',
    startangle=90,
    labels=['Not Survived (0)', 'Survived (1)'],
    colors=['lightcoral', 'skyblue']
)
plt.title("Survival Distribution")
plt.ylabel('')
plt.show()
```



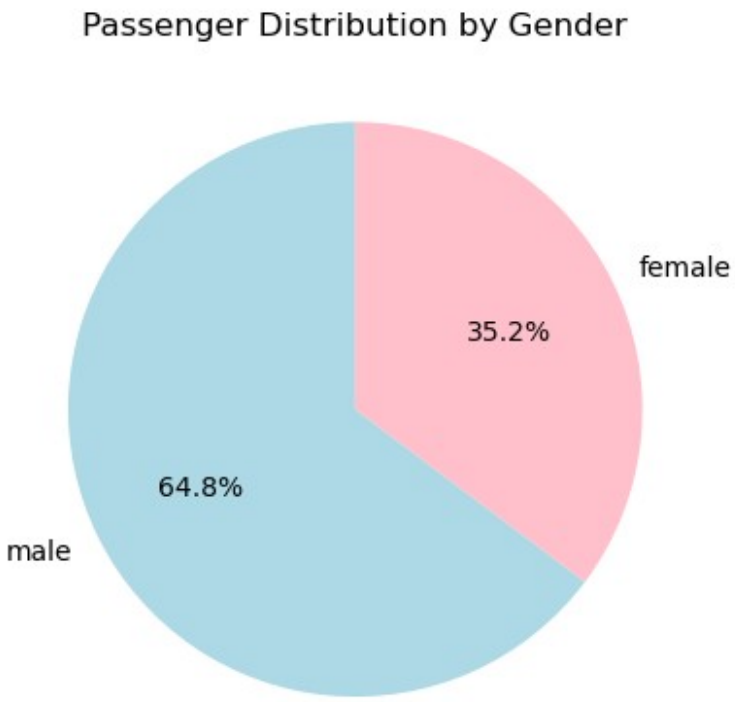

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Titanic-Dataset.csv')
train = pd.read_csv('Titanic-Dataset.csv')

survival_counts = train['Survived'].value_counts()

survival_counts.plot(kind='pie', autopct='%1.1f%%', startangle=90,
labels=['No (0)', 'Yes (1)'], colors=['lightcoral', 'skyblue'])
plt.title("Survival Distribution")
plt.ylabel('')
plt.show()
```



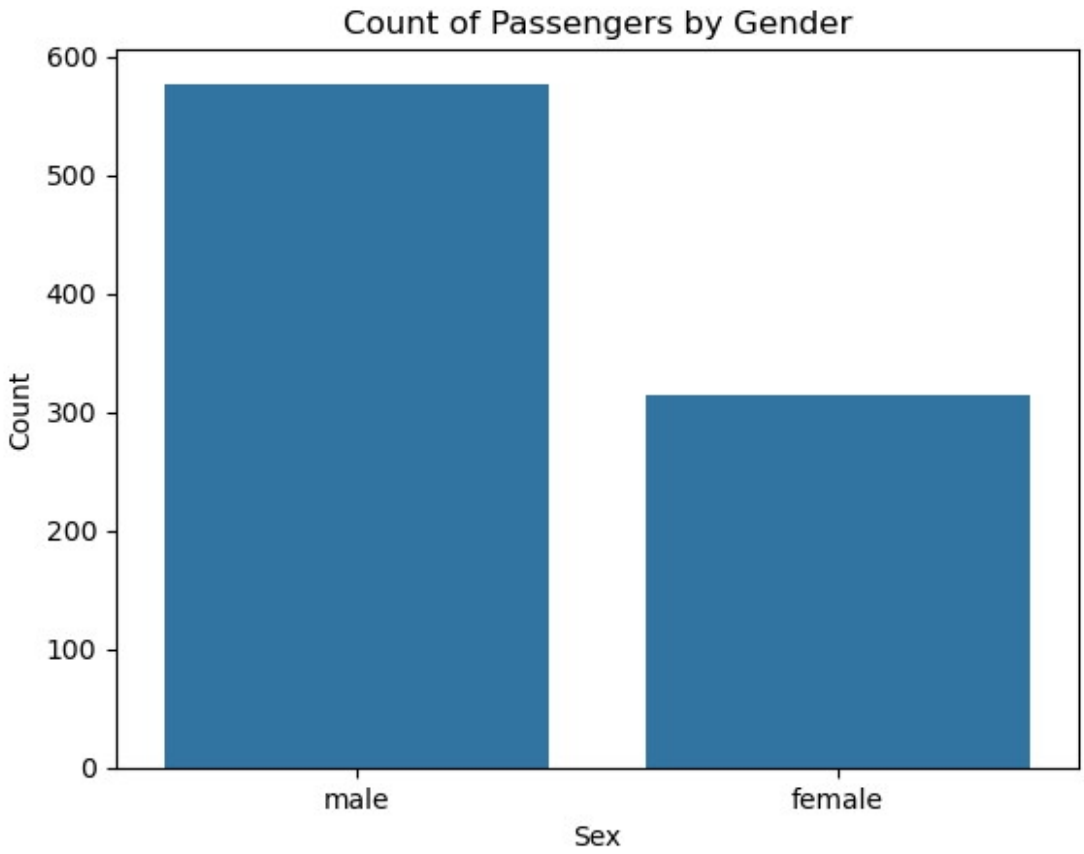

```
df=pd.read_csv('Titanic-Dataset.csv')
train= pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Pclass', data=train)
plt.title("Passenger Class Count")
plt.xlabel("Pclass")
plt.ylabel("Count")
plt.show()
```



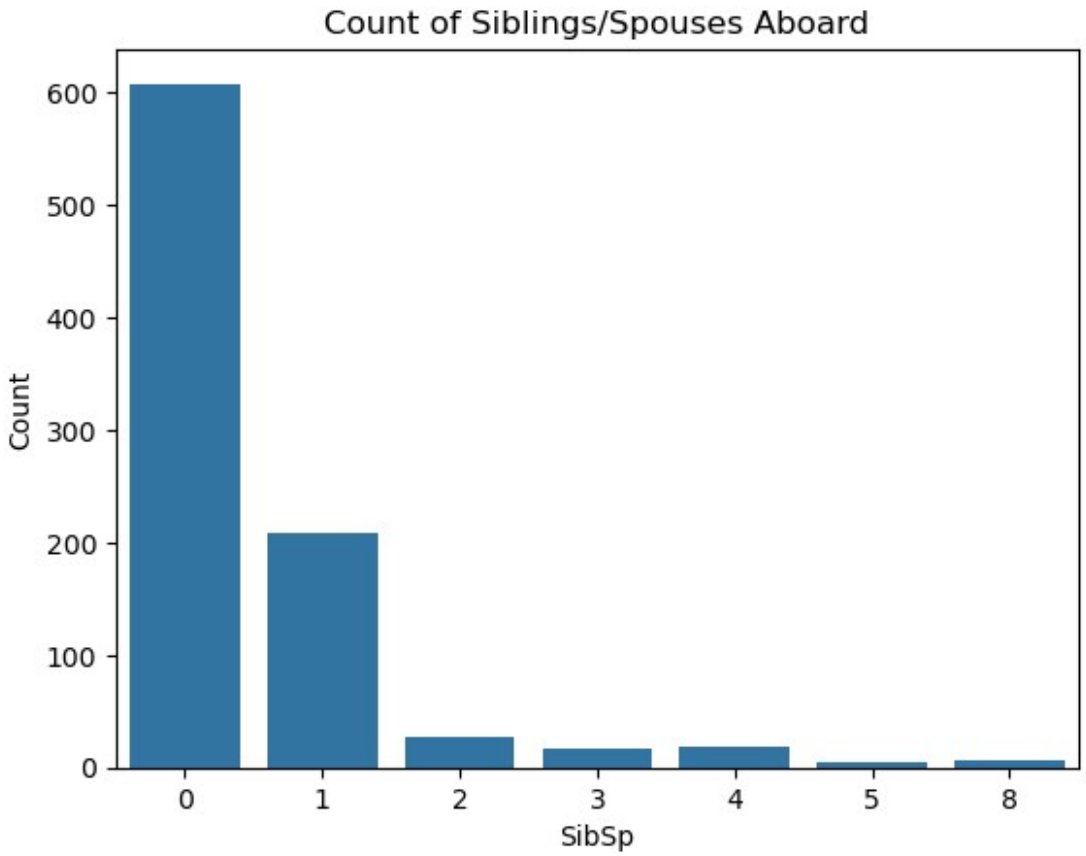

```
import pandas as pd
import matplotlib.pyplot as plt

train = pd.read_csv('Titanic-Dataset.csv')
gender_counts = train['Sex'].value_counts()
gender_counts.plot(kind='pie', autopct='%1.1f%%', startangle=90,
labels=gender_counts.index, colors=['lightblue', 'pink'])
plt.title("Passenger Distribution by Gender")
plt.ylabel('')
plt.show()
```



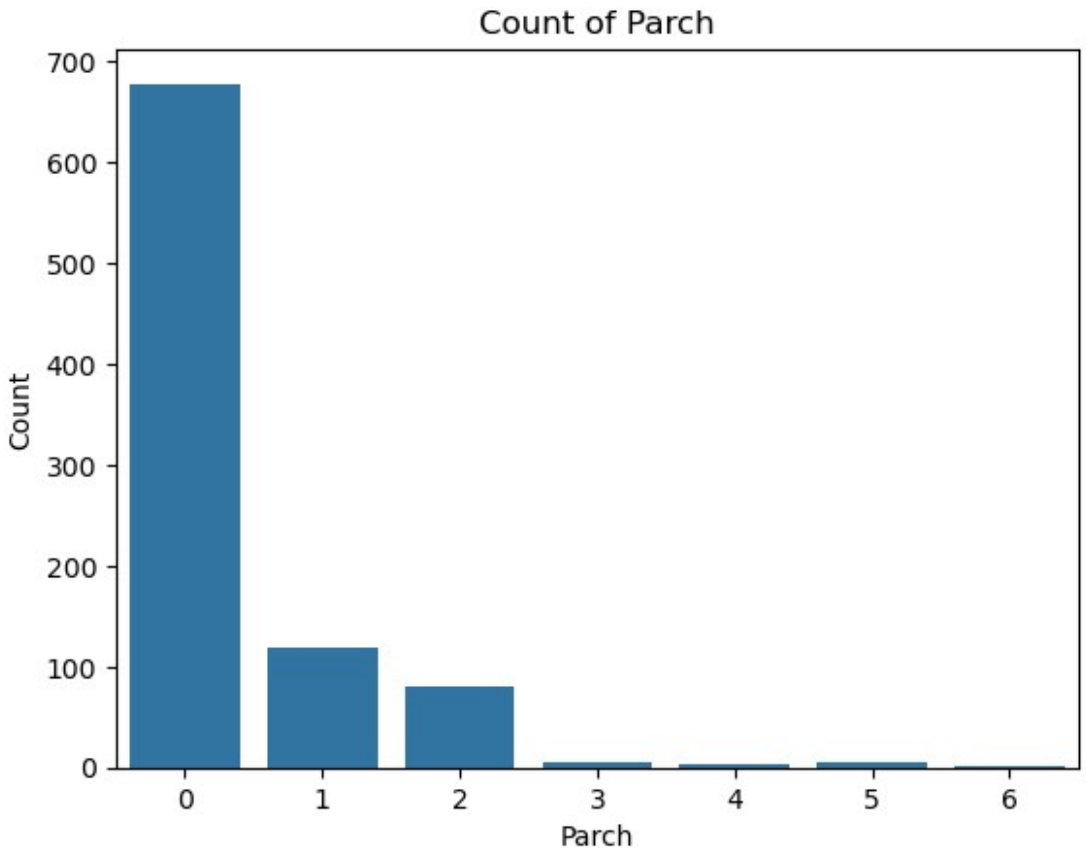

```
train = pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Sex', data=train)
plt.title("Count of Passengers by Gender")
plt.xlabel("Sex")
plt.ylabel("Count")
plt.show()
```



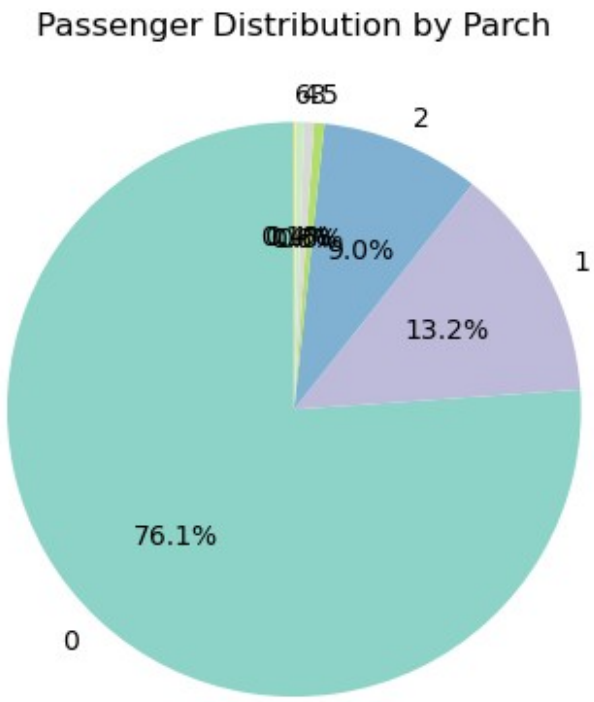

```
train = pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='SibSp', data=train)
plt.title("Count of Siblings/Spouses Aboard")
plt.xlabel("SibSp")
plt.ylabel("Count")
plt.show()
```



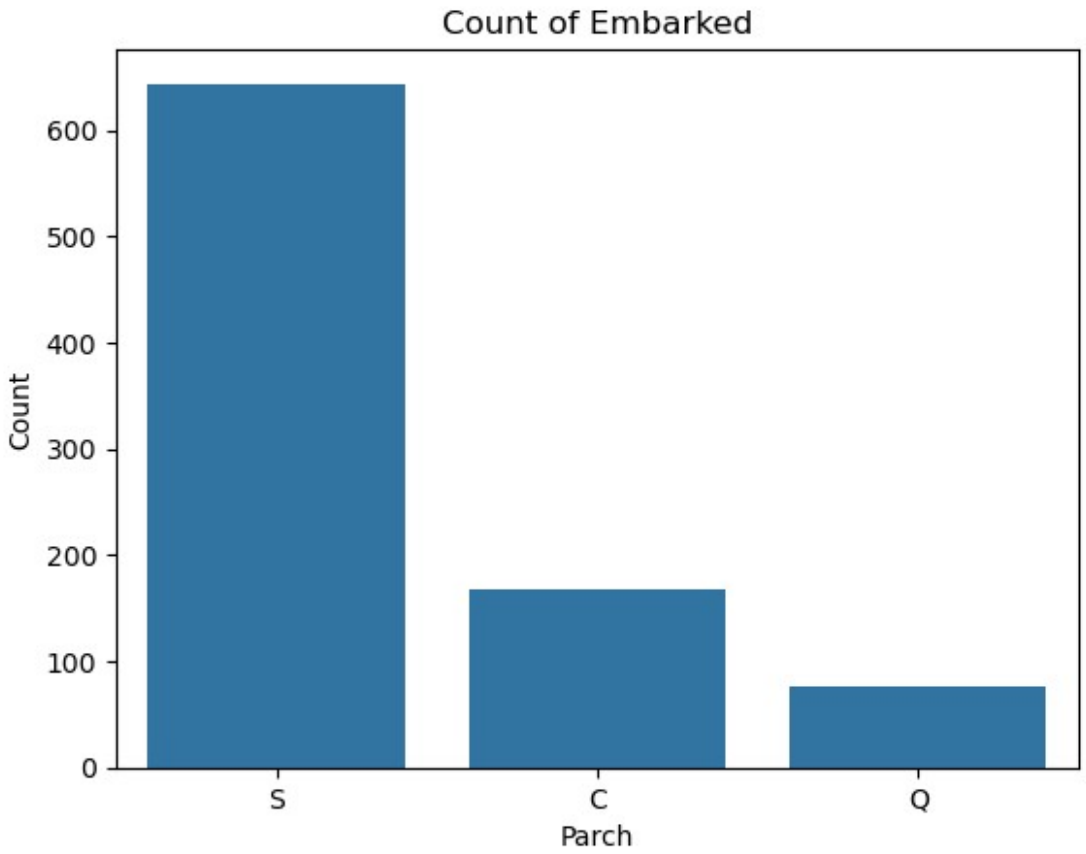

```
train = pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Parch', data=train)
plt.title("Count of Parch")
plt.xlabel("Parch")
plt.ylabel("Count")
plt.show()
```



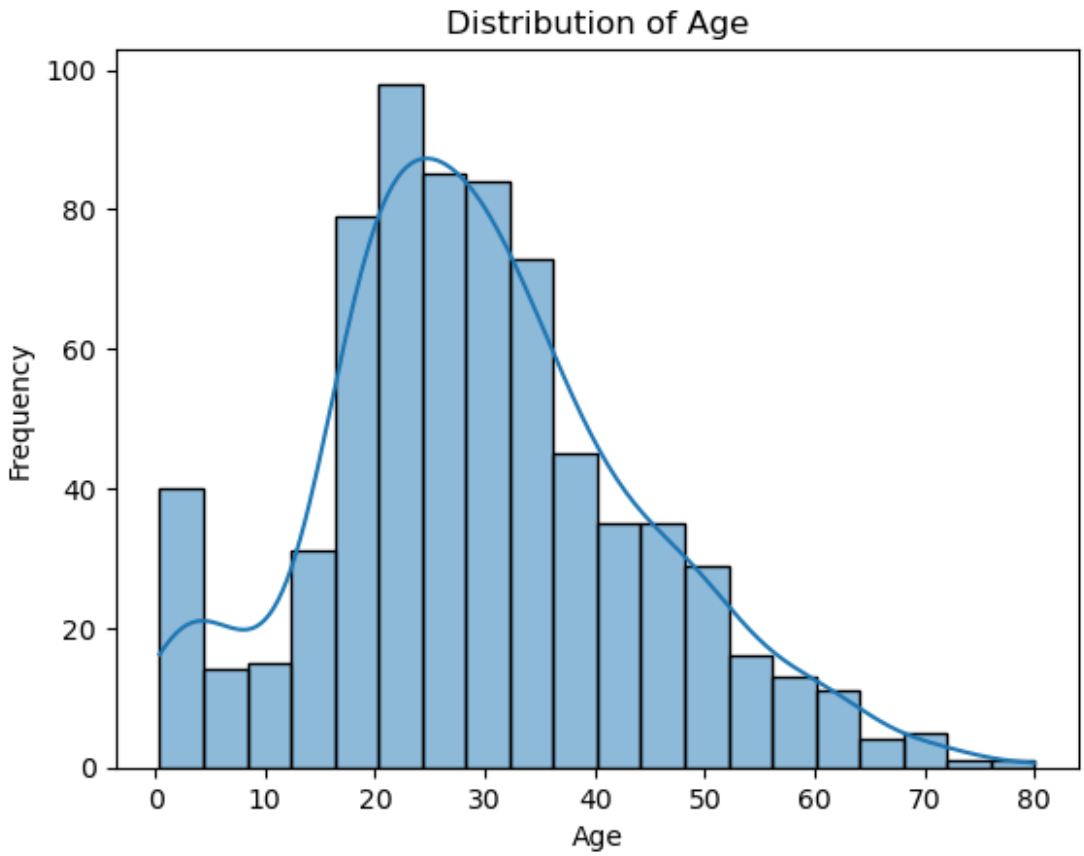

```
import pandas as pd
import matplotlib.pyplot as plt

train = pd.read_csv('Titanic-Dataset.csv')

parch_counts = train['Parch'].value_counts()

parch_counts.plot(
    kind='pie',
    autopct='%1.1f%%',
    startangle=90,
    labels=parch_counts.index,
    cmap='Set3'
)
plt.title("Passenger Distribution by Parch")
plt.ylabel('')
plt.show()
```



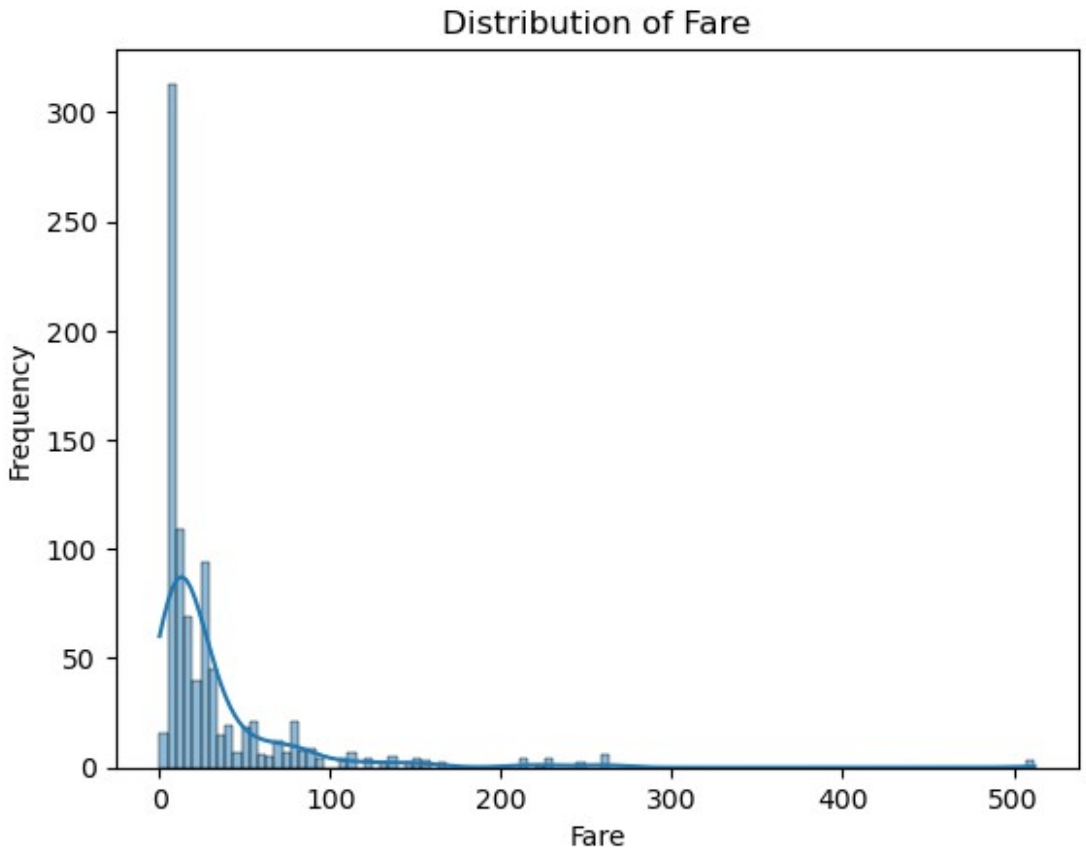

```
train = pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Embarked', data=train)
plt.title("Count of Embarked")
plt.xlabel("Parch")
plt.ylabel("Count")
plt.show()
```



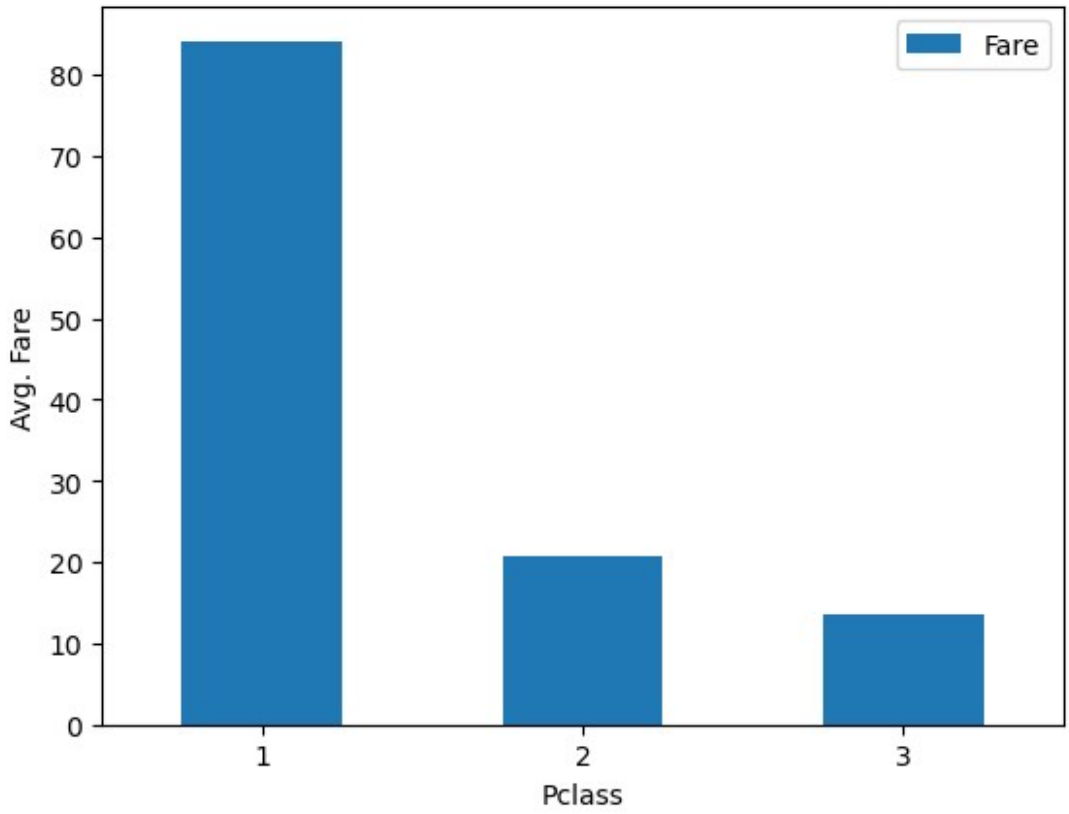

```
train = pd.read_csv('Titanic-Dataset.csv')
sns.histplot(train['Age'], kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



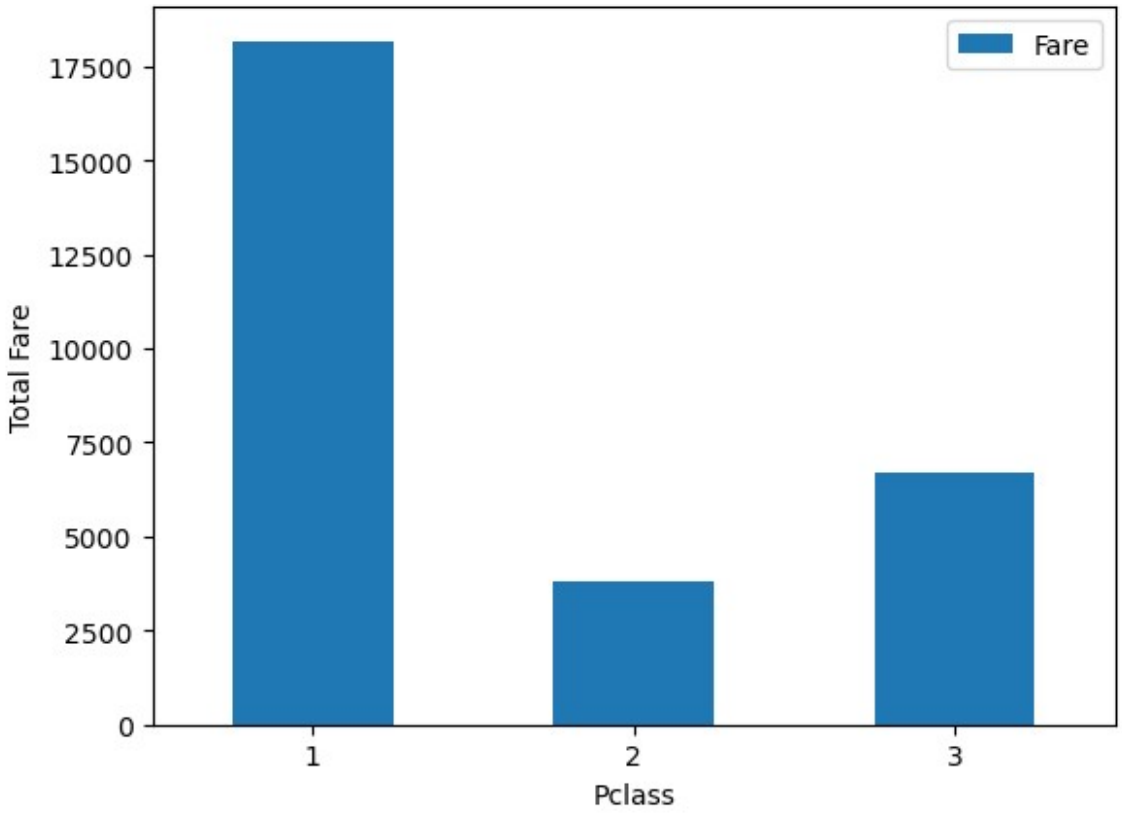

```
train = pd.read_csv('Titanic-Dataset.csv')
sns.histplot(train['Fare'], kde=True)
plt.title('Distribution of Fare')
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.show()
```




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




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



Data Preprocessing

```
train_len = len(train)
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

                                                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
```

```
   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
```

```
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
```

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

```
df['Age'].mean()
```

```
29.881137667304014
```

```
df['Age'] = df['Age'].fillna(df['Age'].mean())
df['Fare'] = df['Fare'].fillna(df['Fare'].mean())
```

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

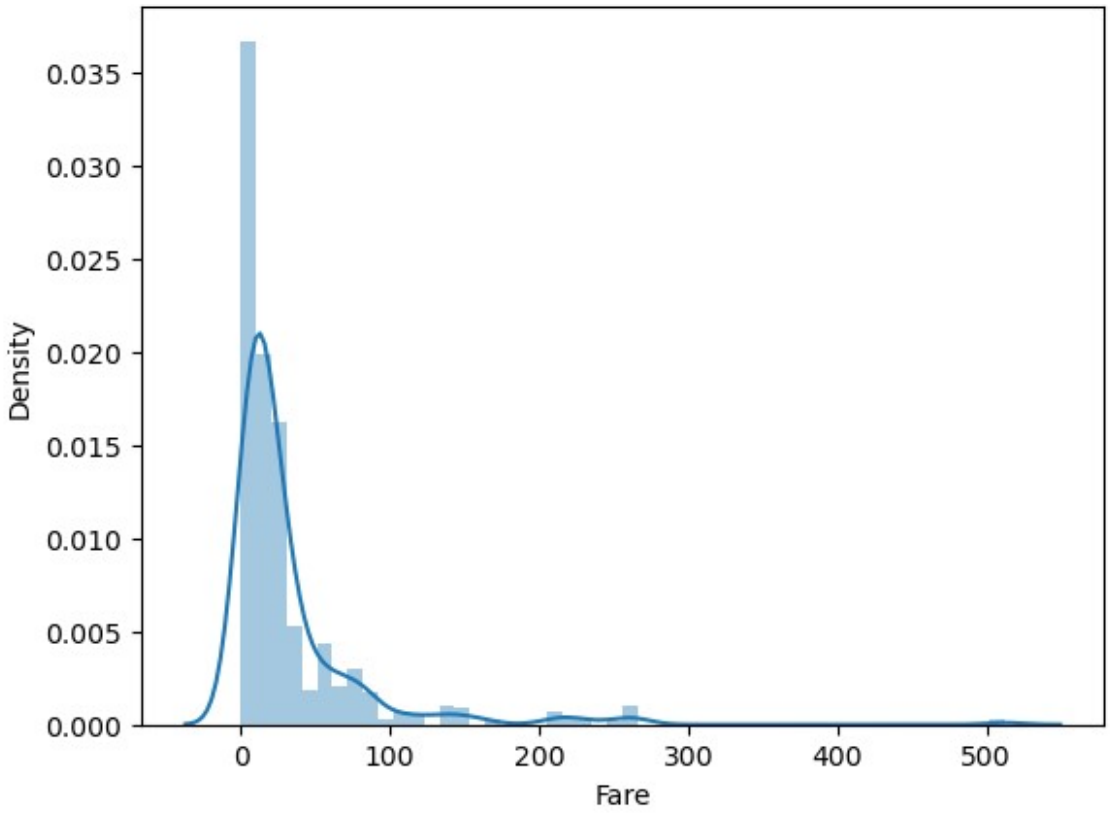
```
'S'
```

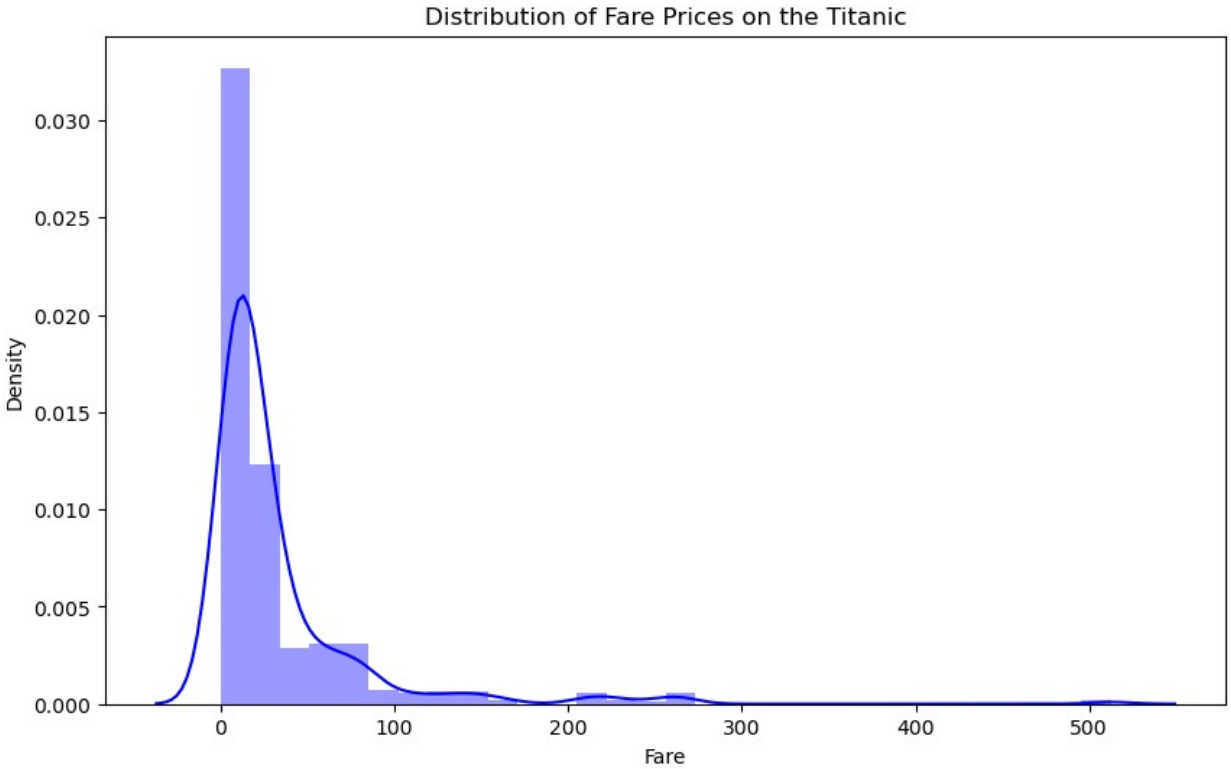
```
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])
```

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

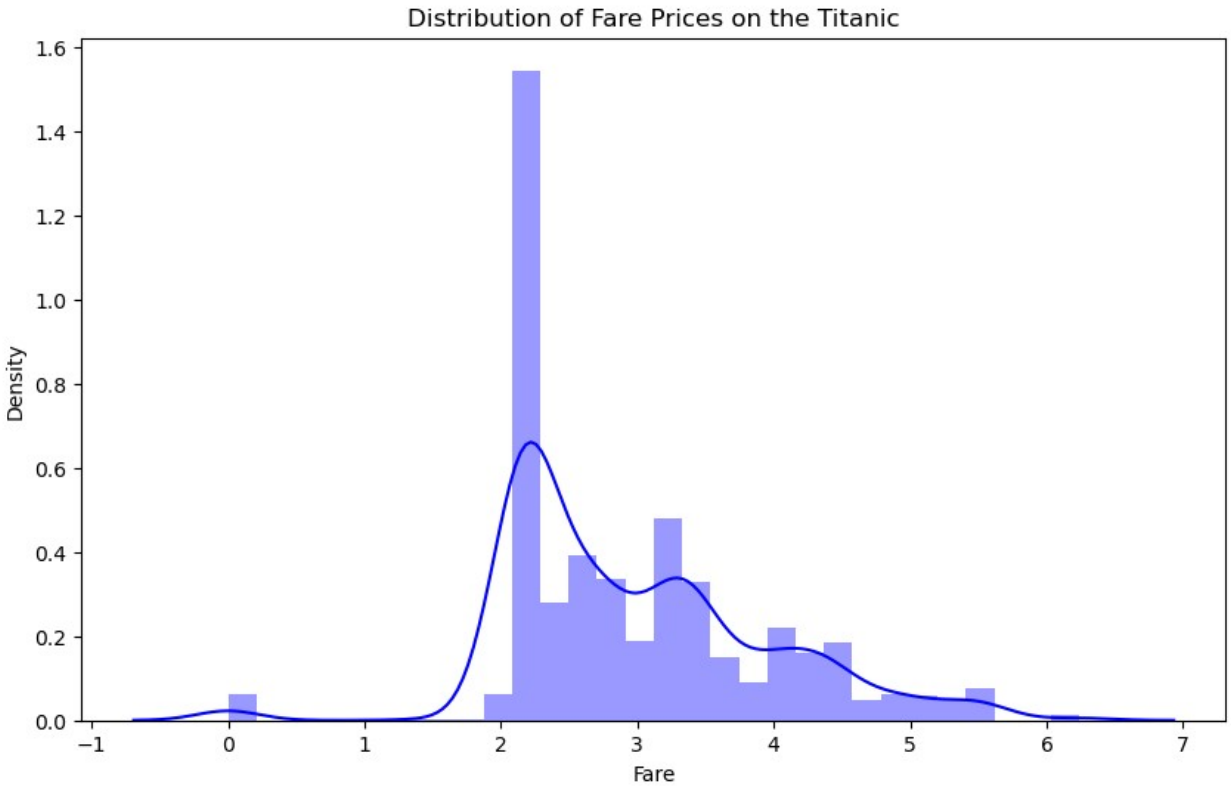
```
<Axes: xlabel='Fare', ylabel='Density'>
```

```
plt.figure(figsize=(10, 6))
sns.distplot(df['Fare'], bins=30, kde=True, color='blue')


plt.title('Distribution of Fare Prices on the Titanic')
plt.xlabel('Fare')
plt.ylabel('Density')
plt.show()
```


0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Density
0
100
200
300
400
500
Fare


Distribution of Fare Prices on the Titanic
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Density
0
100
200
300
400
500
Fare

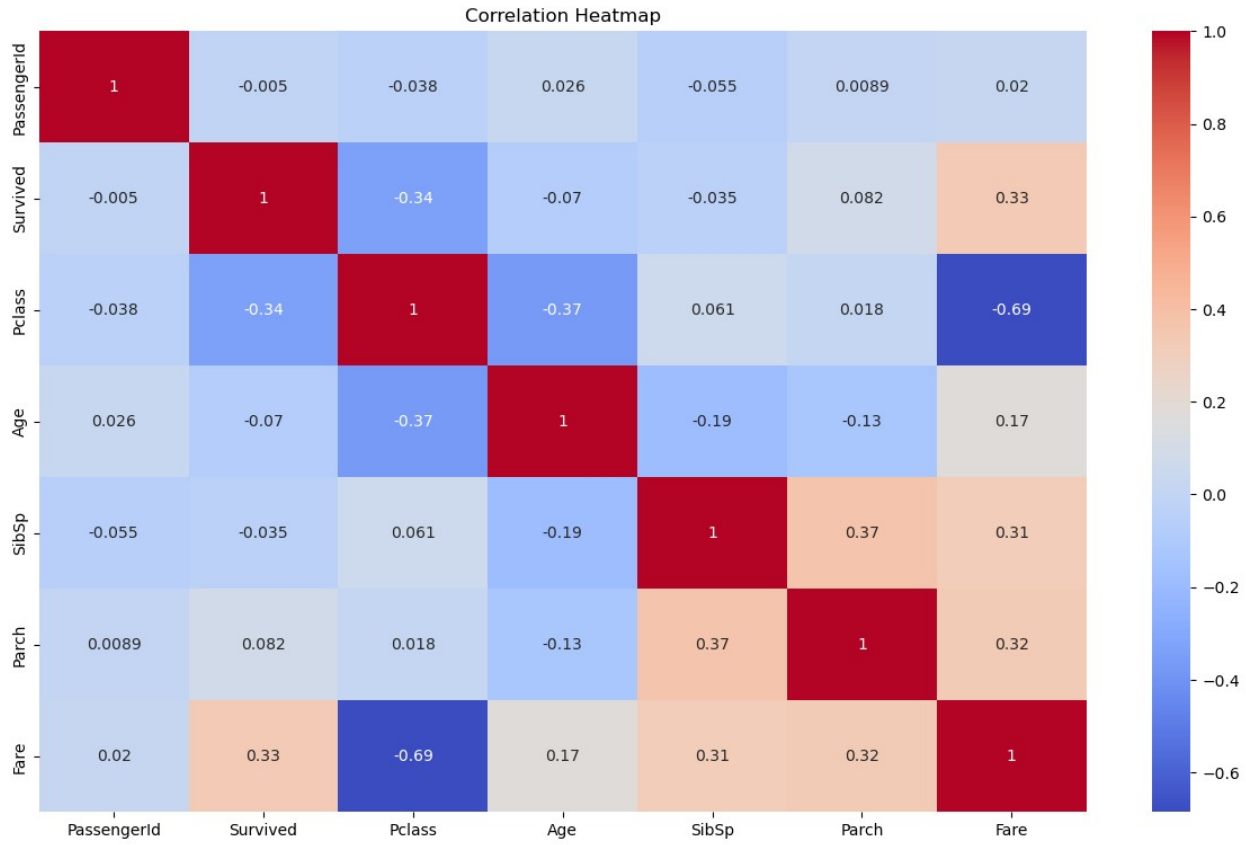
```
df['Fare'] = np.log(df['Fare']+1)

plt.figure(figsize=(10, 6))
sns.distplot(df['Fare'], bins=30, kde=True, color='blue')


plt.title('Distribution of Fare Prices on the Titanic')
plt.xlabel('Fare')
plt.ylabel('Density')
plt.show()
```



Correlation Matrix

```
non_numeric_cols = df.select_dtypes(exclude=['number']).columns
print("Non-numeric columns:", non_numeric_cols)
df_numeric = df.drop(columns=non_numeric_cols)
corr = df_numeric.corr()
plt.figure(figsize=(15, 9))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

```
Non-numeric columns: Index(['Name', 'Sex', 'Ticket', 'Embarked'],
dtype='object')
```




```
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

                                                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

   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 Coding

```
from sklearn.preprocessing import LabelEncoder
cols = ['Sex', 'Embarked']
le = LabelEncoder()

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
```

Train-Test Split

```
train_len = int(0.8 * len(df))
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  \
1047         1048       NaN       1    0  29.000000      0      0
5.406181
1048         1049       NaN       3    0  23.000000      0      0
2.180892
1049         1050       NaN       1    1  42.000000      0      0
3.316003
1050         1051       NaN       3    0  26.000000      0      2
2.692937
1051         1052       NaN       3    0  29.881138      0      0
2.167143

      Embarked
1047         2
1048         2
1049         2
1050         2
1051         1
```

```
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

```
import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.linear_model import LogisticRegression

df = pd.read_csv('Titanic-Dataset.csv')

df['Sex'] = df['Sex'].map({'male': 0, 'female': 1})
df['Embarked'] = df['Embarked'].map({'C': 0, 'Q': 1, 'S': 2})

df['Age'].fillna(df['Age'].median(), inplace=True)
df['Fare'].fillna(df['Fare'].median(), inplace=True)
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)

df['Survived'].dropna(inplace=True)

X = df[['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']]
y = df['Survived']

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)

    accuracy = model.score(X_test, y_test)
    print(f'Accuracy: {accuracy}')

    cv_scores = cross_val_score(model, X_train, y_train, cv=5)
    cv_score = cv_scores.mean()
    print(f'CV Score: {cv_score}')

model = LogisticRegression(max_iter=1000)


classify(model)


Accuracy: 0.8071748878923767
CV Score: 0.7978453596678262

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

```
Accuracy: 0.8071748878923767
CV Score: 0.7978453596678262

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

Accuracy: 0.7219730941704036
CV Score: 0.7575692963752665

from sklearn.ensemble import ExtraTreesClassifier
model = ExtraTreesClassifier()
classify(model)

Accuracy: 0.7892376681614349
CV Score: 0.7889686903826731

pip install lightgbm

from lightgbm import LGBMClassifier
model = LGBMClassifier()
classify(model)

!pip install catboost
from catboost import CatBoostClassifier
model = CatBoostClassifier(verbose=0)
classify(model)
```

Complete Model Training with Full Data

```
test.head()

      PassengerId  Survived  Pclass  Sex        Age  SibSp  Parch
Fare  \
1047         1048       NaN       1    0  29.000000      0      0
5.406181
1048         1049       NaN       3    0  23.000000      0      0
2.180892
1049         1050       NaN       1    1  42.000000      0      0
3.316003
1050         1051       NaN       3    0  26.000000      0      2
2.692937
1051         1052       NaN       3    0  29.881138      0      0
2.167143

      Embarked
1047         2
1048         2
1049         2
1050         2
1051         1
```

```
X_test = test.drop(columns=['PassengerId', 'Survived'], axis=1)

X_test.head()
```

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
1047	1	0	29.000000	0	0	5.406181	2
1048	3	0	23.000000	0	0	2.180892	2
1049	1	1	42.000000	0	0	3.316003	2
1050	3	0	26.000000	0	2	2.692937	2
1051	3	0	29.881138	0	0	2.167143	1

```
pred = model.predict(X_test)
pred
```

```
array([0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,
1,
       0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1,
0,
       0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0,
1,
       0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1,
1,
       0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1,
1,
       1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1,
1,
       1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1,
0,
       1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0,
1,
       1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1,
1,
       1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0,
0,
       0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0,
0,
       1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1],
      dtype=int64)
```

```
sub = pd.read_csv('gender_submission.csv')
sub.head()
```

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1

```
sub.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  418 non-null    int64
 1   Survived     418 non-null    int64
dtypes: int64(2)
memory usage: 6.7 KB
```

```
sub.head()
```

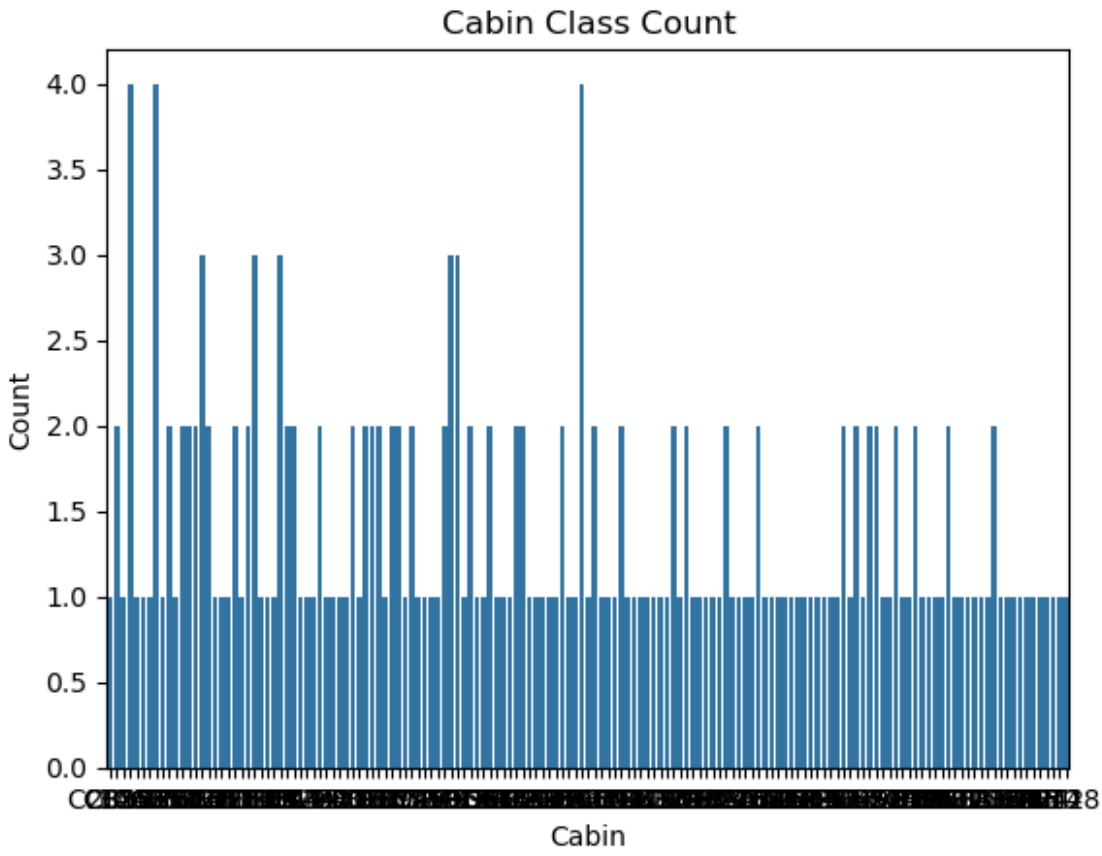
```
   PassengerId  Survived
0          892         0
1          893         1
2          894         0
3          895         0
4          896         1
```

```
sub.to_csv('submission.csv', index=False)
```

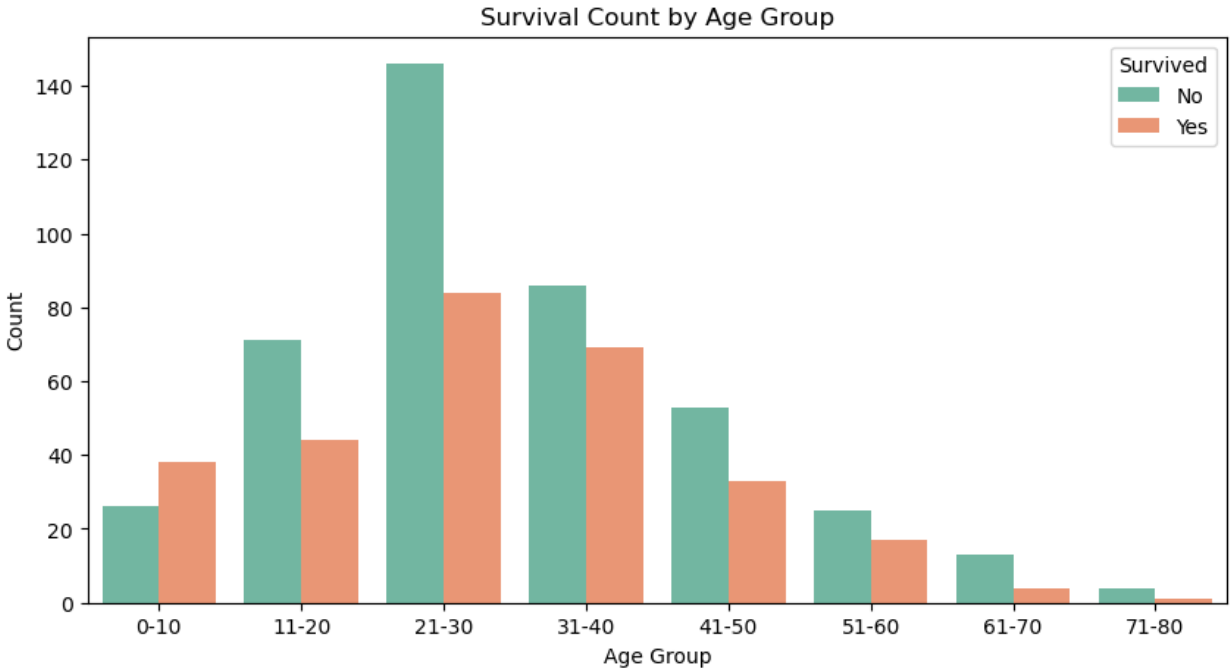
```
pip install xgboost
```

```
Requirement already satisfied: xgboost in c:\users\likhi\anaconda3\
lib\site-packages (2.1.3)
Requirement already satisfied: numpy in c:\users\likhi\anaconda3\lib\
site-packages (from xgboost) (1.26.4)
Requirement already satisfied: scipy in c:\users\likhi\anaconda3\lib\
site-packages (from xgboost) (1.13.1)
Note: you may need to restart the kernel to use updated packages.
```

```
df=pd.read_csv('Titanic-Dataset.csv')
train= pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Cabin', data=train)
plt.title("Cabin Class Count")
plt.xlabel("Cabin")
plt.ylabel("Count")
plt.show()
```




```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('Titanic-Dataset.csv')

bins = [0, 10, 20, 30, 40, 50, 60, 70, 80]
labels = ['0-10', '11-20', '21-30', '31-40', '41-50', '51-60', '61-
70', '71-80']
df['AgeGroup'] = pd.cut(df['Age'], bins=bins, labels=labels)

plt.figure(figsize=(10, 5))
sns.countplot(x='AgeGroup', hue='Survived', data=df, palette='Set2')
plt.title("Survival Count by Age Group")
plt.xlabel("Age Group")
plt.ylabel("Count")
plt.legend(title="Survived", labels=["No", "Yes"])
plt.show()
```




```
surviving_men_class1 = train[(train["Pclass"] == 1) & (train["Sex"] ==
"male") & (train["Survived"] == 1)]

print(surviving_men_class1.head())

num_surviving_men_class1 = surviving_men_class1.shape[0]
print(f"Number of surviving men in first class:
{num_surviving_men_class1}")
```

```
     PassengerId  Survived  Pclass  \
23            24         1       1
55            56         1       1
97            98         1       1
187          188         1       1
209          210         1       1

                                                 Name   Sex   Age  SibSp
Parch  \
23                      Sloper, Mr. William Thompson  male  28.0      0
0
55                                Woolner, Mr. Hugh  male   NaN      0
0
97                   Greenfield, Mr. William Bertram  male  23.0      0
1
187  Romaine, Mr. Charles Hallace ("Mr C Rolmane")  male  45.0      0
0
209                               Blank, Mr. Henry  male  40.0      0
0
```

```
       Ticket     Fare    Cabin Embarked
23     113788  35.5000       A6        S
55      19947  35.5000      C52        S
97   PC 17759  63.3583  D10 D12        C
187    111428  26.5500      NaN        S
209    112277  31.0000      A31        C
Number of surviving men in first class: 45

a=train[train['Survived']==1]
a

     PassengerId  Survived  Pclass  \
1              2         1       1
2              3         1       3
3              4         1       1
8              9         1       3
9             10         1       2
..           ...       ...     ...
875          876         1       3
879          880         1       1
880          881         1       2
887          888         1       1
889          890         1       1

                                                  Name     Sex   Age
SibSp  \
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
8    Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0
0
9                  Nasser, Mrs. Nicholas (Adele Achem)  female  14.0
1
..                                                 ...     ...   ...
...
875                   Najib, Miss. Adele Kiamie "Jane"  female  15.0
0
879      Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)  female  56.0
0
880       Shelley, Mrs. William (Imanita Parrish Hall)  female  25.0
0
887                         Graham, Miss. Margaret Edith  female  19.0
0
889                              Behr, Mr. Karl Howell    male  26.0
0

     Parch            Ticket     Fare Cabin Embarked
```

```
1         0          PC 17599  71.2833   C85        C
2         0  STON/O2. 3101282   7.9250   NaN        S
3         0            113803  53.1000  C123        S
8         2            347742  11.1333   NaN        S
9         0            237736  30.0708   NaN        C
..      ...               ...      ...   ...      ...
875       0              2667   7.2250   NaN        C
879       1             11767  83.1583   C50        C
880       1            230433  26.0000   NaN        S
887       0            112053  30.0000   B42        S
889       0            111369  30.0000  C148        C

[342 rows x 12 columns]
```

```
b=a[a['Pclass']==1]
b
```

```
     PassengerId  Survived  Pclass  \
1              2         1       1
3              4         1       1
11            12         1       1
23            24         1       1
31            32         1       1
..           ...       ...     ...
862          863         1       1
871          872         1       1
879          880         1       1
887          888         1       1
889          890         1       1

                                                  Name     Sex   Age
SibSp  \
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0
1
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0
1
11                            Bonnell, Miss. Elizabeth  female  58.0
0
23                         Sloper, Mr. William Thompson    male  28.0
0
31      Spencer, Mrs. William Augustus (Marie Eugenie)  female   NaN
1
..                                                 ...     ...   ...
...
862  Swift, Mrs. Frederick Joel (Margaret Welles Ba...  female  48.0
0
871   Beckwith, Mrs. Richard Leonard (Sallie Monypeny)  female  47.0
1
879      Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)  female  56.0
0
```

```
887                         Graham, Miss. Margaret Edith  female  19.0
0
889                                Behr, Mr. Karl Howell    male  26.0
0

     Parch    Ticket      Fare Cabin Embarked
1        0  PC 17599   71.2833   C85        C
3        0    113803   53.1000  C123        S
11       0    113783   26.5500  C103        S
23       0    113788   35.5000    A6        S
31       0  PC 17569  146.5208   B78        C
..     ...       ...       ...   ...      ...
862      0     17466   25.9292   D17        S
871      1     11751   52.5542   D35        S
879      1     11767   83.1583   C50        C
887      0    112053   30.0000   B42        S
889      0    111369   30.0000  C148        C

[136 rows x 12 columns]
```

```
b.Age.value_counts()
```

```
Age
35.00    9
36.00    7
24.00    5
48.00    5
30.00    5
38.00    4
39.00    4
49.00    4
22.00    4
27.00    3
52.00    3
33.00    3
42.00    3
16.00    3
18.00    3
17.00    3
58.00    3
23.00    3
31.00    3
40.00    3
19.00    3
26.00    2
45.00    2
56.00    2
28.00    2
54.00    2
51.00    2
```

```
44.00    2
50.00    2
25.00    2
60.00    2
32.00    2
21.00    2
29.00    1
15.00    1
62.00    1
11.00    1
80.00    1
43.00    1
41.00    1
53.00    1
34.00    1
4.00     1
14.00    1
0.92     1
63.00    1
37.00    1
47.00    1
Name: count, dtype: int64
```

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

train = pd.read_csv('Titanic-Dataset.csv')

female_deaths = train[(train['Sex'] == 'female') & (train['Survived']
== 0)].groupby('Pclass')['PassengerId'].count()

print(female_deaths)
```

```
Pclass
1     3
2     6
3    72
Name: PassengerId, dtype: int64
```

```
a=train[train['Survived']==0]
a
```

```
     PassengerId  Survived  Pclass
Name  \
0              1         0       3                     Braund, Mr. Owen
Harris
4              5         0       3                    Allen, Mr. William
Henry
5              6         0       3                            Moran, Mr.
```

```
James
6              7         0       1                    McCarthy, Mr.
Timothy J
7              8         0       3             Palsson, Master. Gosta
Leonard
..           ...       ...     ...
...
884          885         0       3                     Sutehall, Mr.
Henry Jr
885          886         0       3      Rice, Mrs. William (Margaret
Norton)
886          887         0       2                      Montvila, Rev.
Juozas
888          889         0       3  Johnston, Miss. Catherine Helen
"Carrie"
890          891         0       3                         Dooley, Mr.
Patrick

        Sex   Age  SibSp  Parch            Ticket     Fare Cabin
Embarked
0      male  22.0      1      0         A/5 21171   7.2500   NaN
S
4      male  35.0      0      0            373450   8.0500   NaN
S
5      male   NaN      0      0            330877   8.4583   NaN
Q
6      male  54.0      0      0             17463  51.8625   E46
S
7      male   2.0      3      1            349909  21.0750   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
888  female   NaN      1      2        W./C. 6607  23.4500   NaN
S
890    male  32.0      0      0            370376   7.7500   NaN
Q

[549 rows x 12 columns]
```

```
b=a[a['Pclass']==3]
b
```

```
     PassengerId  Survived  Pclass
Name  \
```

```
0              1         0       3                    Braund, Mr. Owen
Harris
4              5         0       3                   Allen, Mr. William
Henry
5              6         0       3                           Moran, Mr.
James
7              8         0       3             Palsson, Master. Gosta
Leonard
12            13         0       3             Saundercock, Mr. William
Henry
..           ...       ...     ...
...
882          883         0       3               Dahlberg, Miss. Gerda
Ulrika
884          885         0       3                    Sutehall, Mr.
Henry Jr
885          886         0       3      Rice, Mrs. William (Margaret
Norton)
888          889         0       3  Johnston, Miss. Catherine Helen
"Carrie"
890          891         0       3                         Dooley, Mr.
Patrick

        Sex   Age  SibSp  Parch           Ticket     Fare Cabin
Embarked
0      male  22.0      1      0        A/5 21171   7.2500   NaN
S
4      male  35.0      0      0           373450   8.0500   NaN
S
5      male   NaN      0      0           330877   8.4583   NaN
Q
7      male   2.0      3      1           349909  21.0750   NaN
S
12     male  20.0      0      0        A/5. 2151   8.0500   NaN
S
..      ...   ...    ...    ...              ...      ...   ...      .
..
882  female  22.0      0      0             7552  10.5167   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
888  female   NaN      1      2       W./C. 6607  23.4500   NaN
S
890    male  32.0      0      0           370376   7.7500   NaN
Q

[372 rows x 12 columns]
```

```
train_path = "/mnt/data/train.csv"
train= pd.read_csv('train.csv')
df['Deck'] = df['Cabin'].astype(str).str[0]
most_common_deck = df['Deck'].mode()[0]
df['Deck'].fillna(most_common_deck, inplace=True)

df.drop(columns=['Cabin'], inplace=True)

print(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  \
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 AgeGroup Deck
0      0         A/5 21171   7.2500        S    21-30    n
1      0          PC 17599  71.2833        C    31-40    C
2      0  STON/O2. 3101282   7.9250        S    21-30    n
3      0            113803  53.1000        S    31-40    C
4      0            373450   8.0500        S    31-40    n
```

```
# Titanic Dataset Descriptio
```