

```
In [15]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
```

```
In [16]: df=pd.read_csv('Titanic-Dataset.csv')
df
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599 7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803 5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450
...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536 1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053 3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607 2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369 3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376

891 rows × 12 columns



```
In [17]: train= pd.read_csv('train.csv')
test = pd.read_csv('test.csv')
train.head()
```

Out[17]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05

In [18]: train.describe()

Out[18]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.200000
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693000
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.450000
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.320000

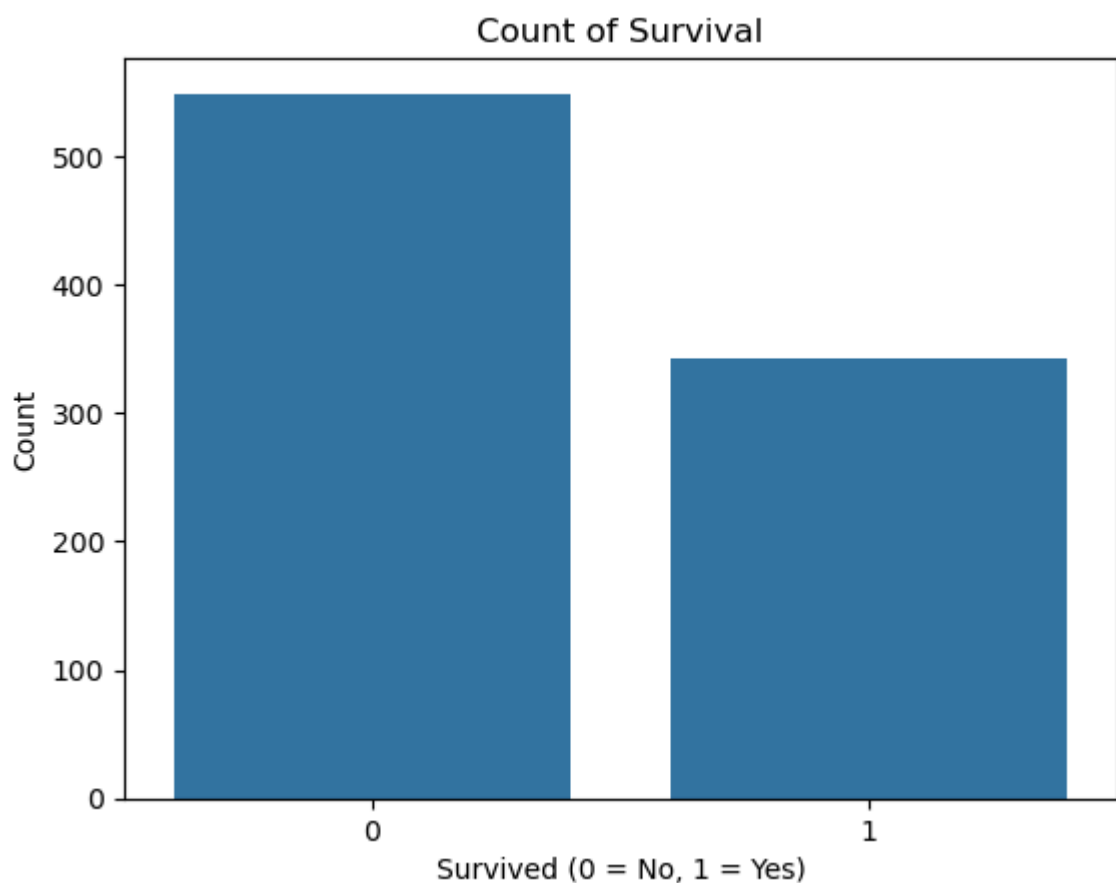
# statistical info

In [19]: train.info()

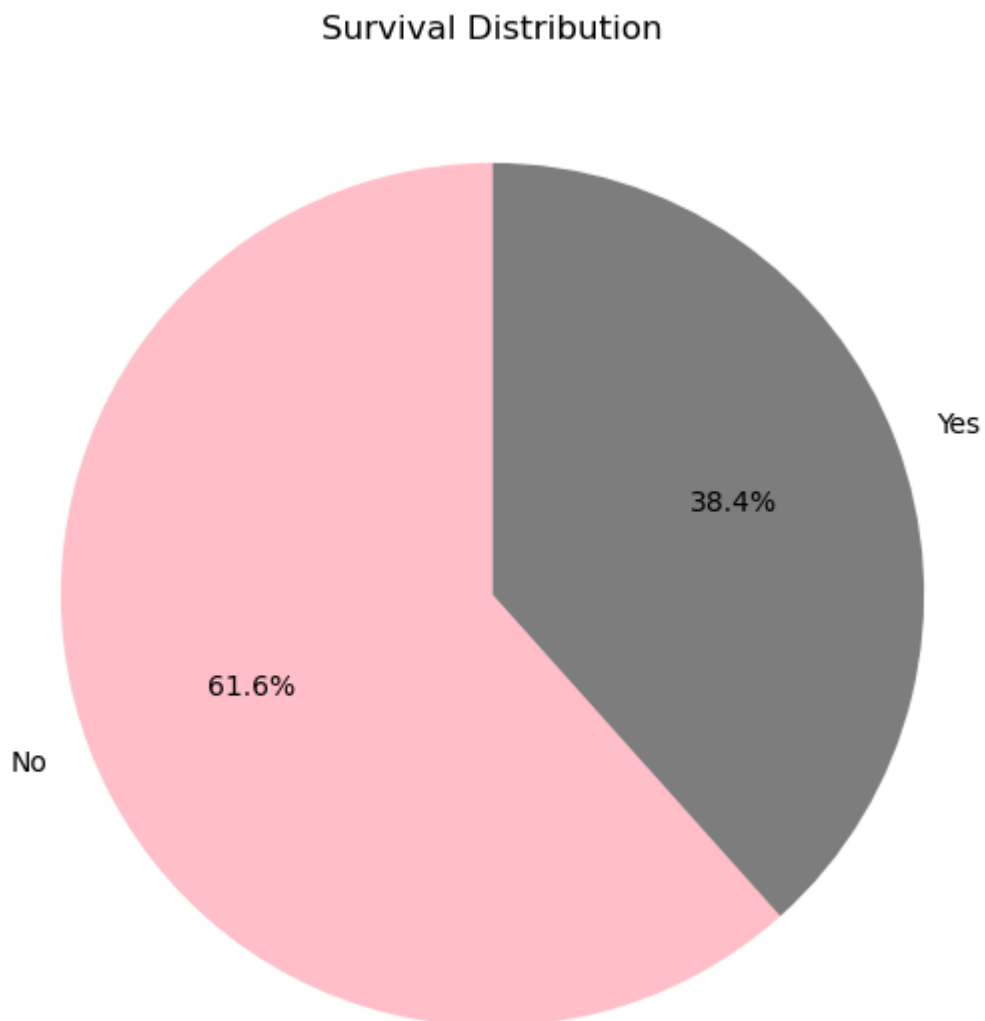
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  891 non-null    int64
1   Survived     891 non-null    int64
2   Pclass       891 non-null    int64
3   Name         891 non-null    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

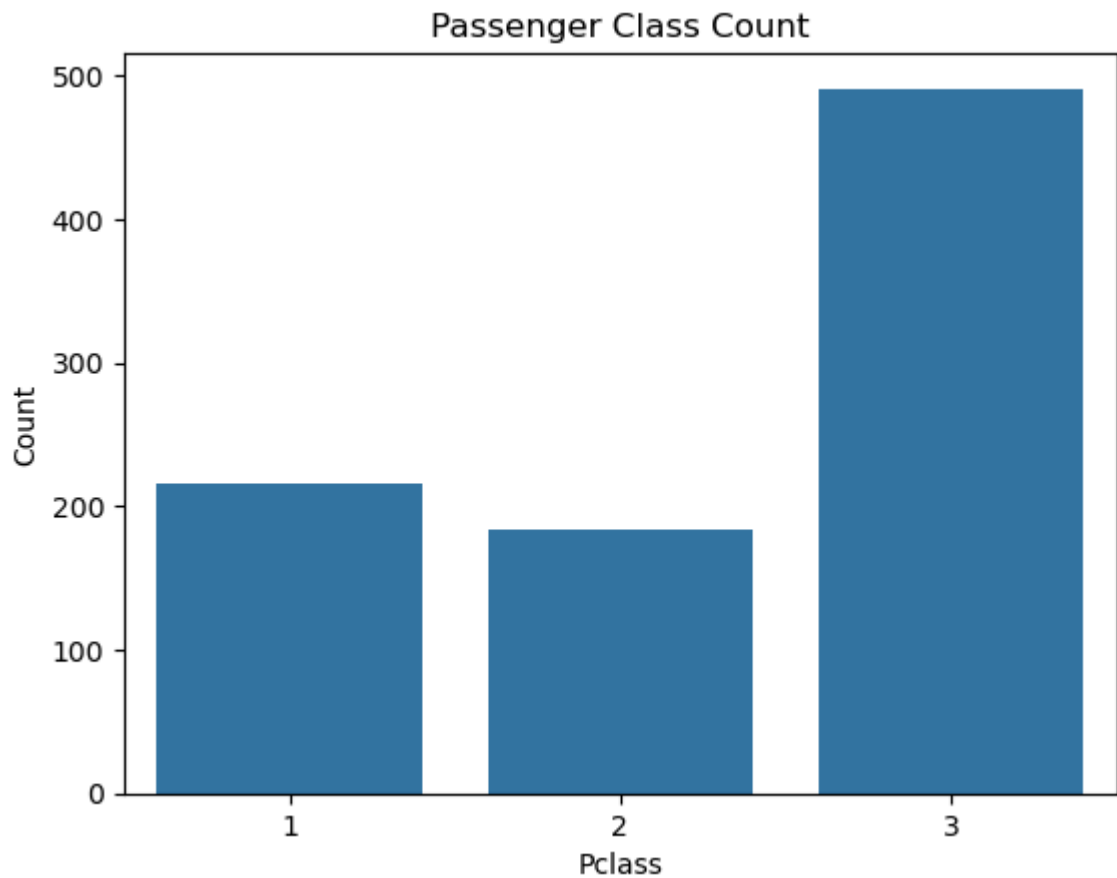
```
In [20]: df=pd.read_csv('Titanic-Dataset.csv')
train= pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Survived', data=train)
plt.title("Count of Survival")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Count")
plt.show()
```



```
In [21]: train = pd.read_csv('Titanic-Dataset.csv')
survival_counts = train['Survived'].value_counts()
plt.figure(figsize=(7, 7))
plt.pie(survival_counts, labels=['No', 'Yes'], autopct='%1.1f%%', startangle=90,
plt.title("Survival Distribution")
plt.show()
```



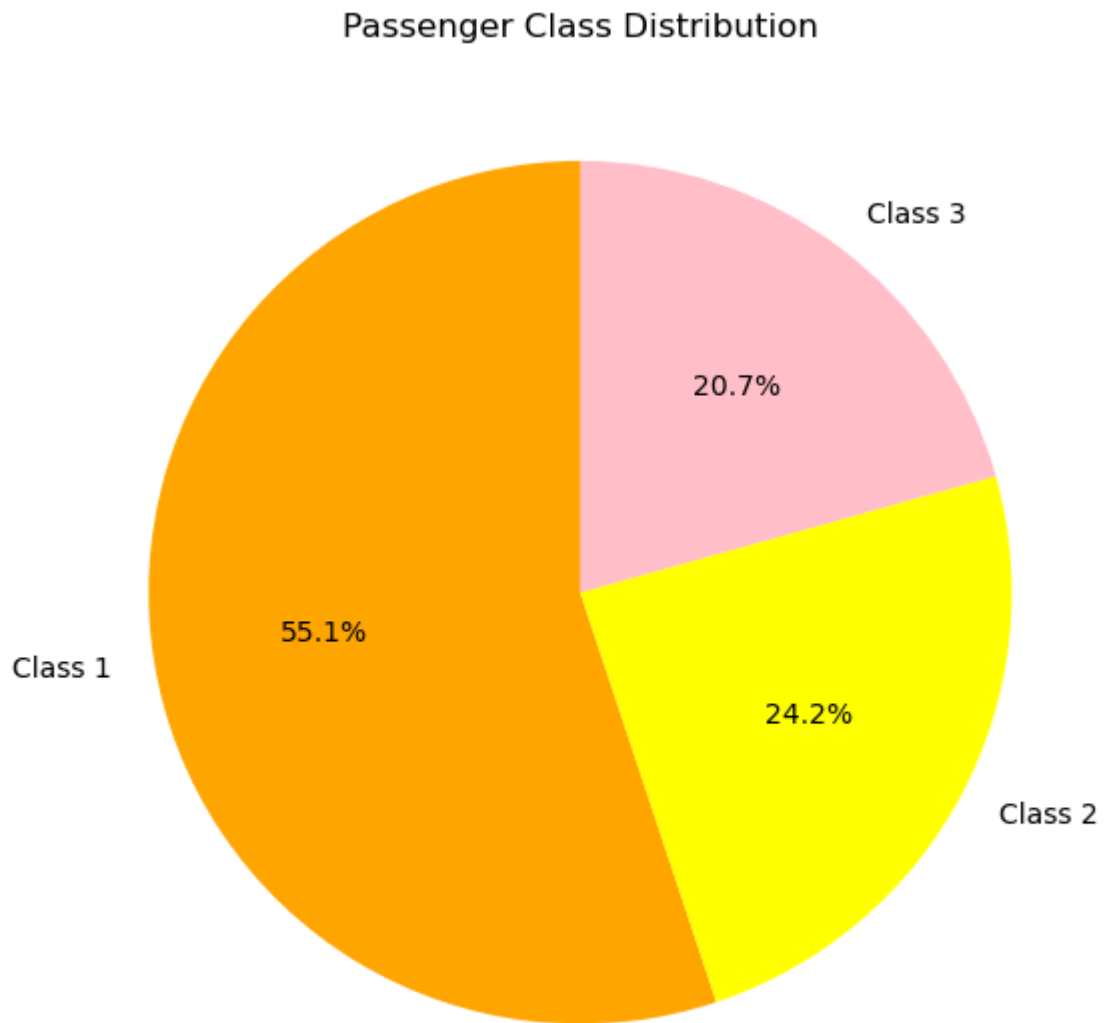
```
In [22]: df=pd.read_csv('Titanic-Dataset.csv')
train= pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Pclass', data=train)
plt.title("Passenger Class Count")
plt.xlabel("Pclass")
plt.ylabel("Count")
plt.show()
```



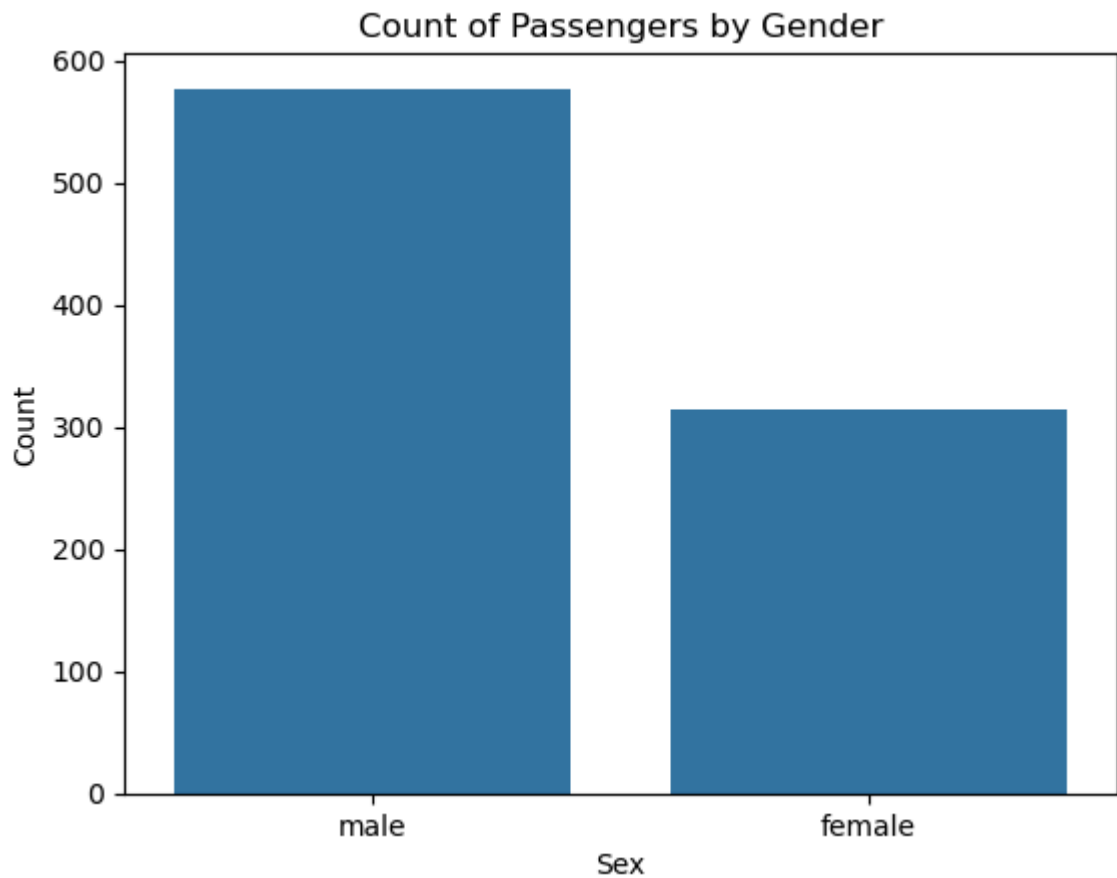
```
In [23]: train = pd.read_csv('Titanic-Dataset.csv')

pclass_counts = train['Pclass'].value_counts()

plt.figure(figsize=(7, 7))
plt.pie(pclass_counts, labels=['Class 1', 'Class 2', 'Class 3'], autopct='%1.1f%%')
plt.title("Passenger Class Distribution")
plt.show()
```



```
In [24]: 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()
```



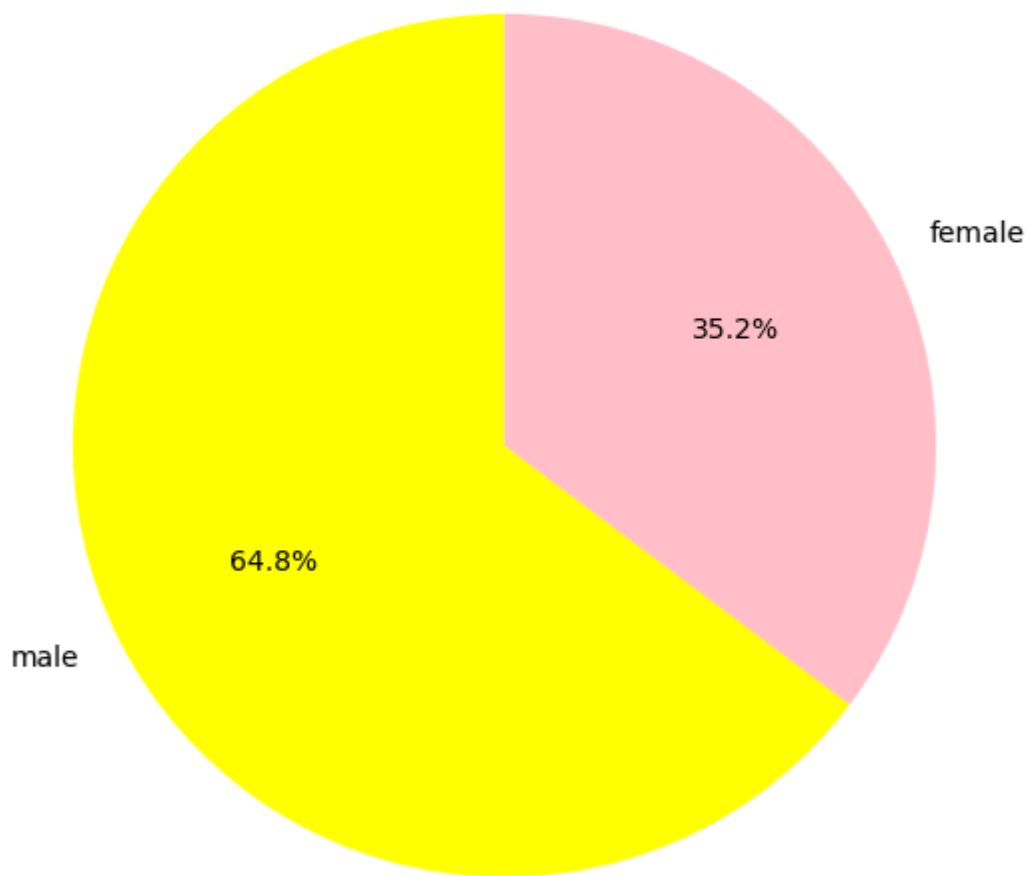
```
In [25]: train = pd.read_csv('Titanic-Dataset.csv')

sex_counts = train['Sex'].value_counts()

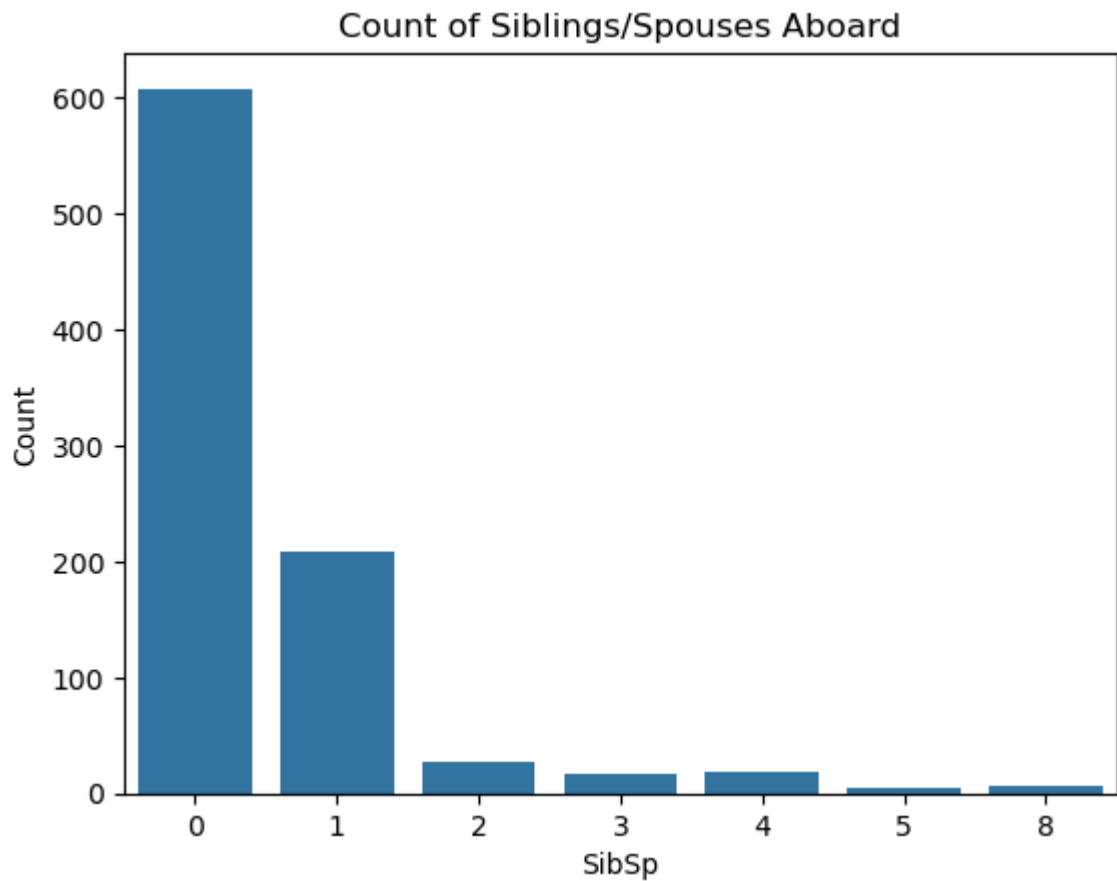
plt.figure(figsize=(7, 7))
plt.pie(sex_counts, labels=sex_counts.index, autopct='%1.1f%%', startangle=90, c
plt.title("Gender Distribustion of Passenger")
plt.show()
```



## Gender Distribution of Passenger



```
In [26]: 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()
```

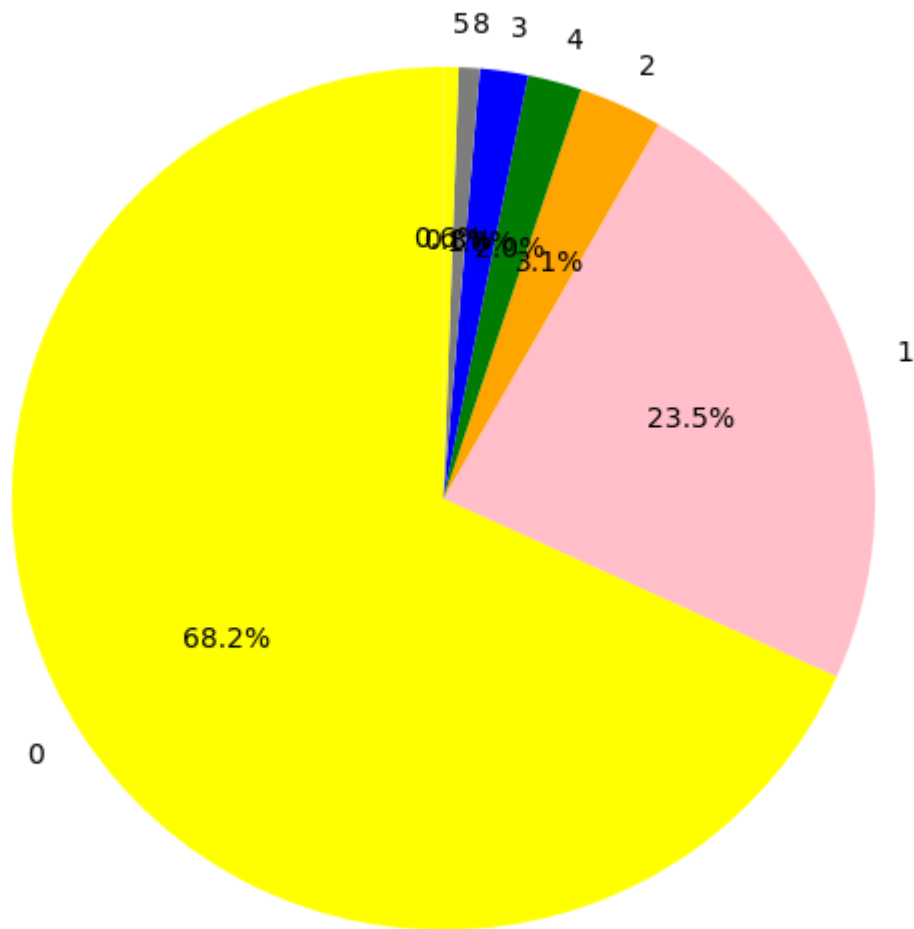


```
In [27]: train = pd.read_csv('Titanic-Dataset.csv')

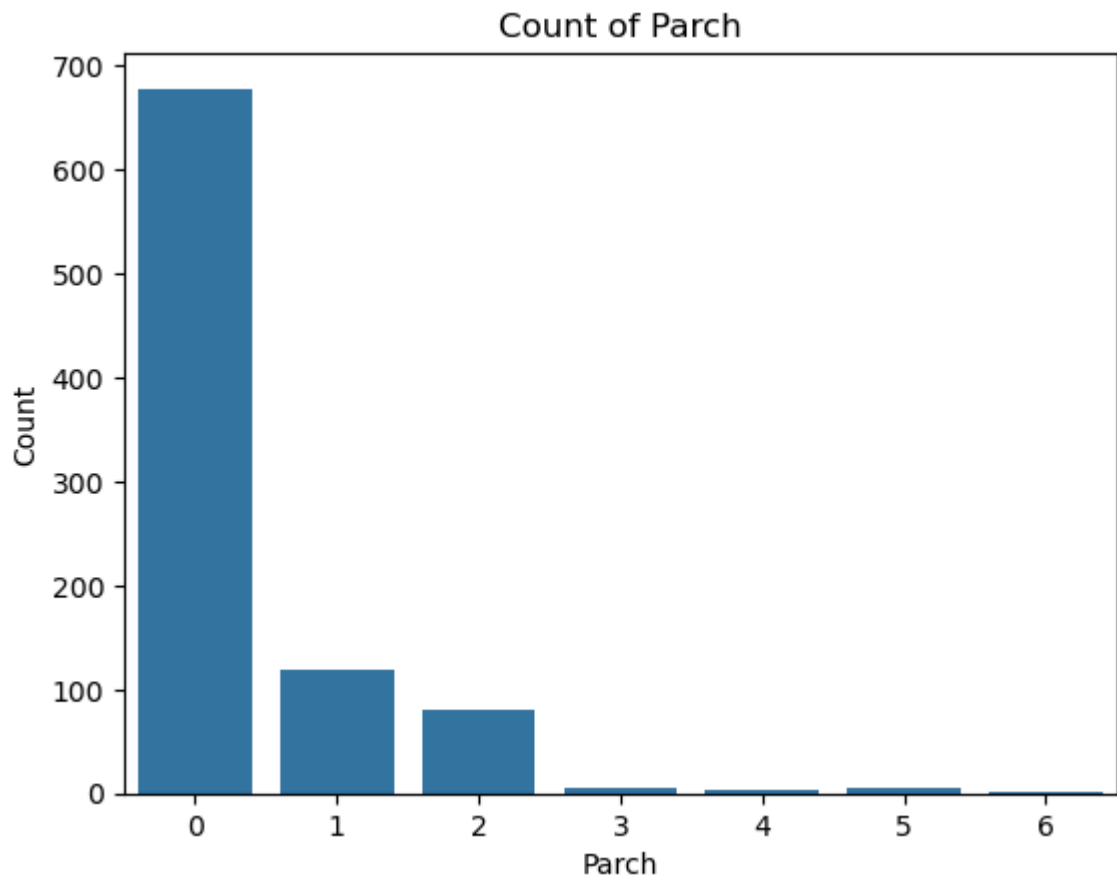
sibsp_counts = train['SibSp'].value_counts()

plt.figure(figsize=(7, 7))
plt.pie(sibsp_counts, labels=sibsp_counts.index, autopct='%1.1f%%', startangle=90)
plt.title("Distribution of Siblings/Spouses Aboard")
plt.show()
```

## Distribution of Siblings/Spouses Aboard

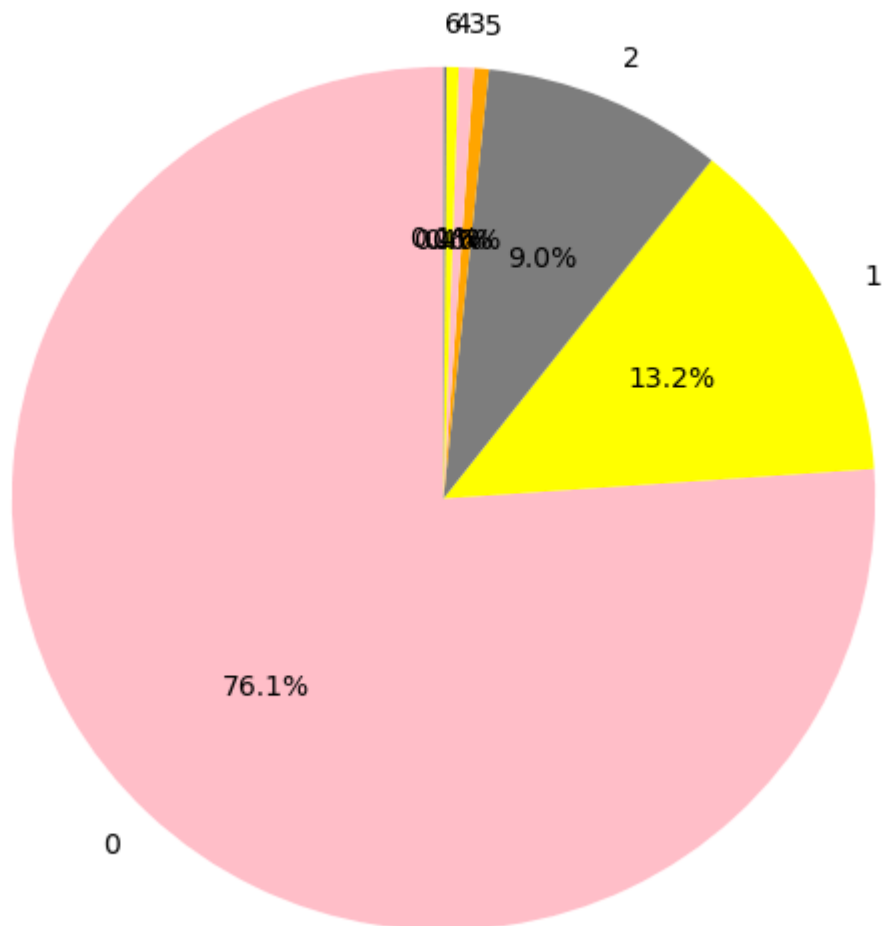


```
In [28]: train = pd.read_csv('Titanic-Dataset.csv')
sns.countplot(x='Parch', data=train)
plt.title("Count of Parch")
plt.xlabel("Parch")
plt.ylabel("Count")
plt.show()
```

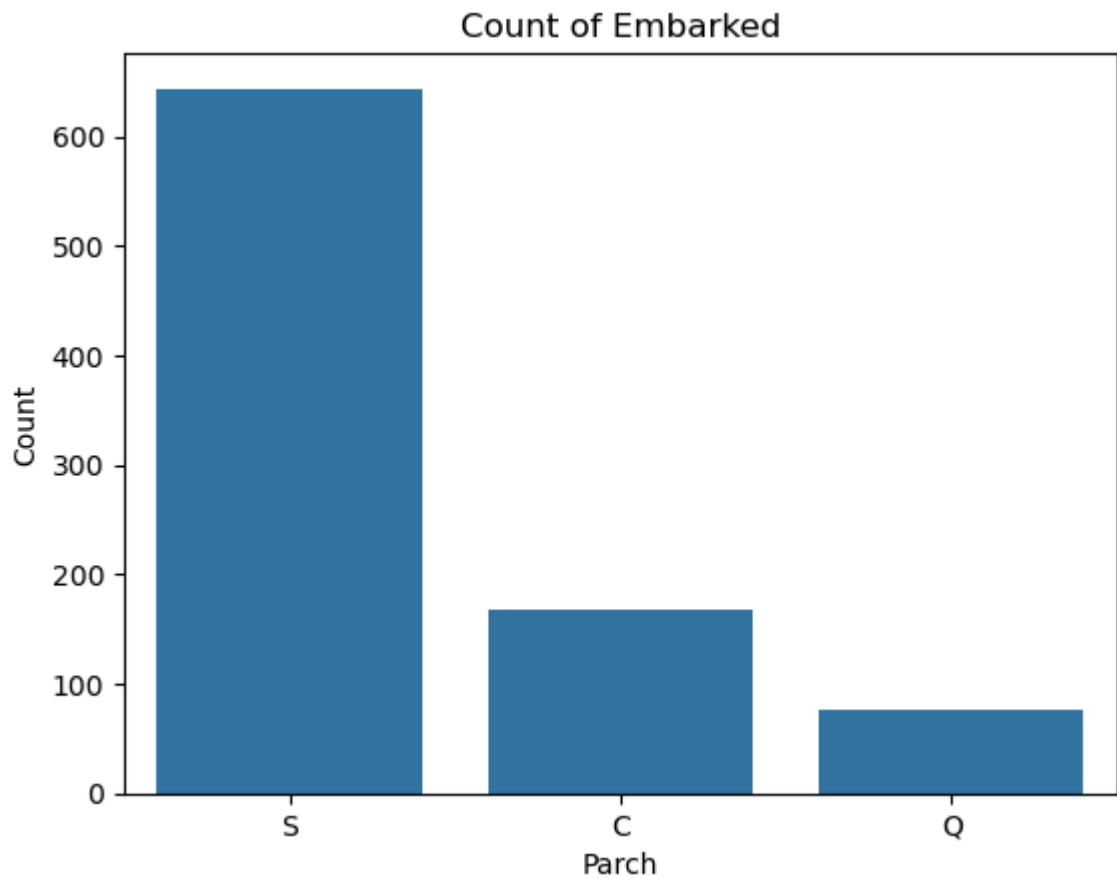


```
In [29]: train = pd.read_csv('Titanic-Dataset.csv')
parch_counts = train['Parch'].value_counts()
plt.figure(figsize=(7, 7))
plt.pie(parch_counts, labels=parch_counts.index, autopct='%1.1f%%', startangle=90)
plt.title("Distribution of Parents/Children Aboard")
plt.show()
```

## Distribution of Parents/Children Aboard



```
In [30]: 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()
```

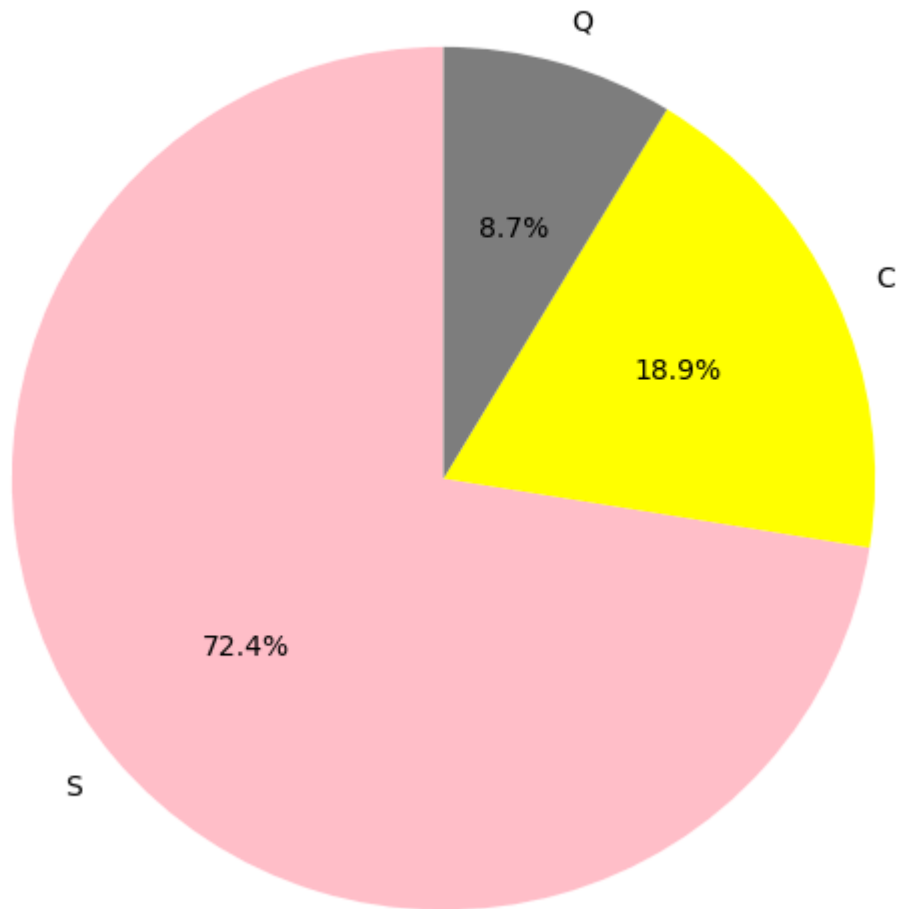


```
In [31]: train = pd.read_csv('Titanic-Dataset.csv')

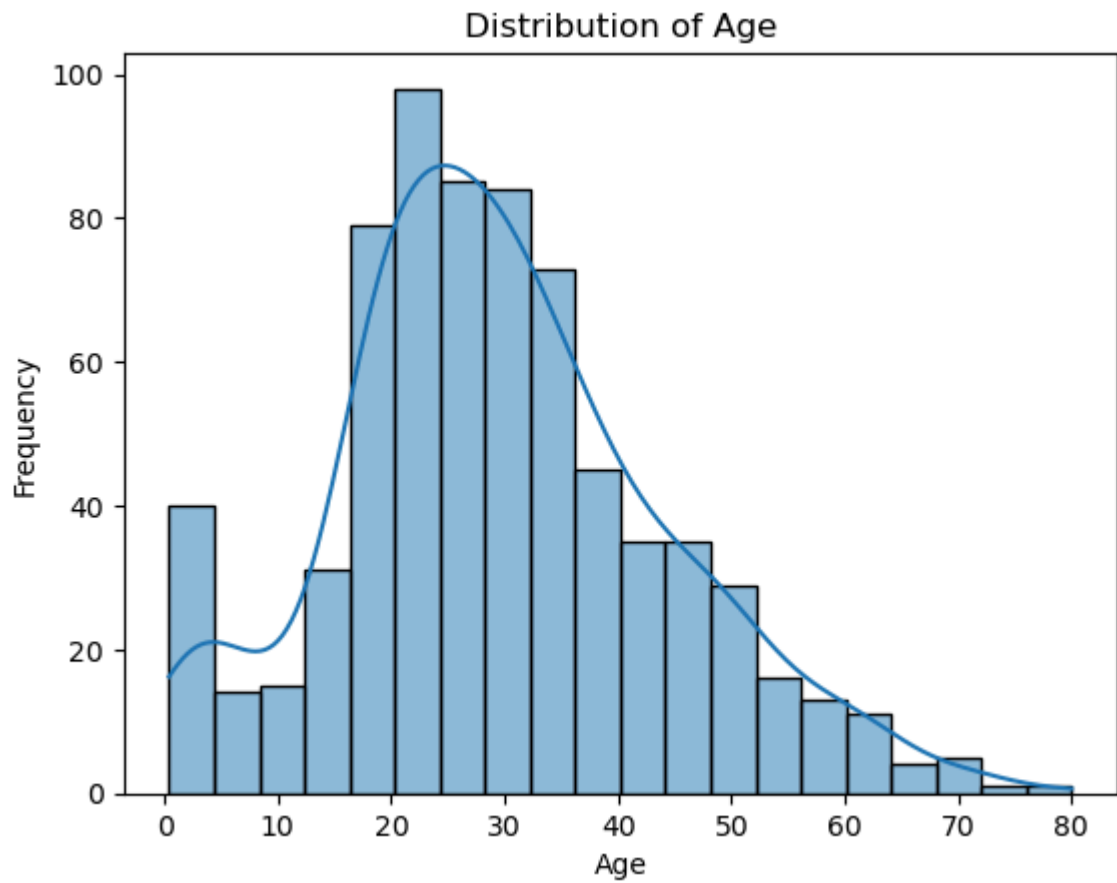
embarked_counts = train['Embarked'].value_counts()

plt.figure(figsize=(7, 7))
plt.pie(embarked_counts, labels=embarked_counts.index, autopct='%1.1f%%', startangle=90)
plt.title("Distribution of Embarked Ports")
plt.show()
```

## Distribution of Embarked Ports

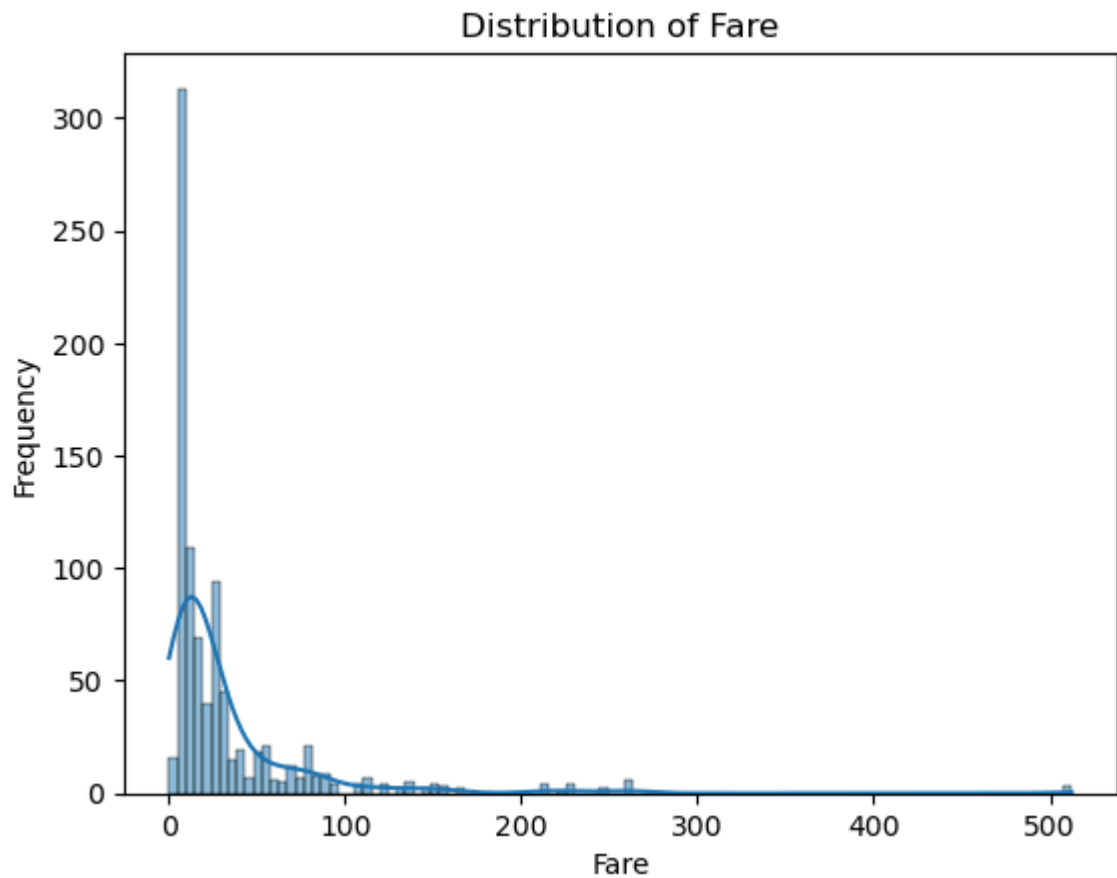


```
In [32]: train = pd.read_csv('Titanic-Dataset.csv')
sns.histplot(train['Age'], kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

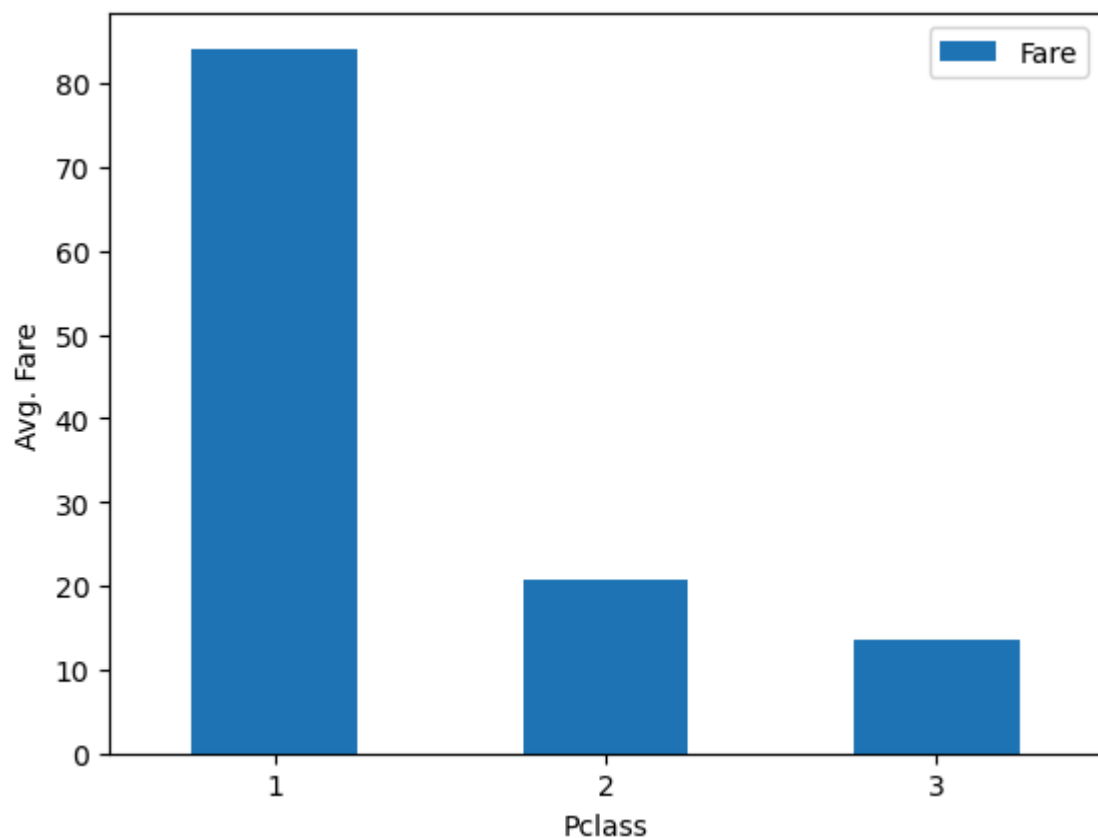


```
In [33]: train = pd.read_csv('Titanic-Dataset.csv')
sns.histplot(train['Fare'], kde=True)
plt.title('Distribution of Fare')
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.show()
```

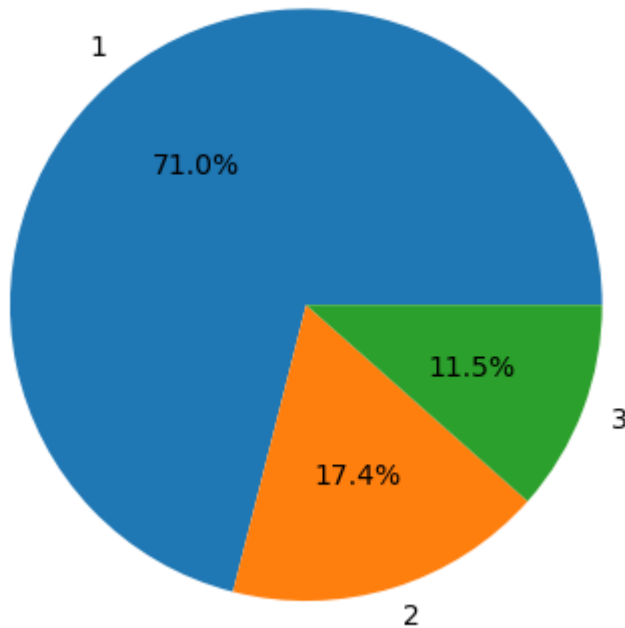




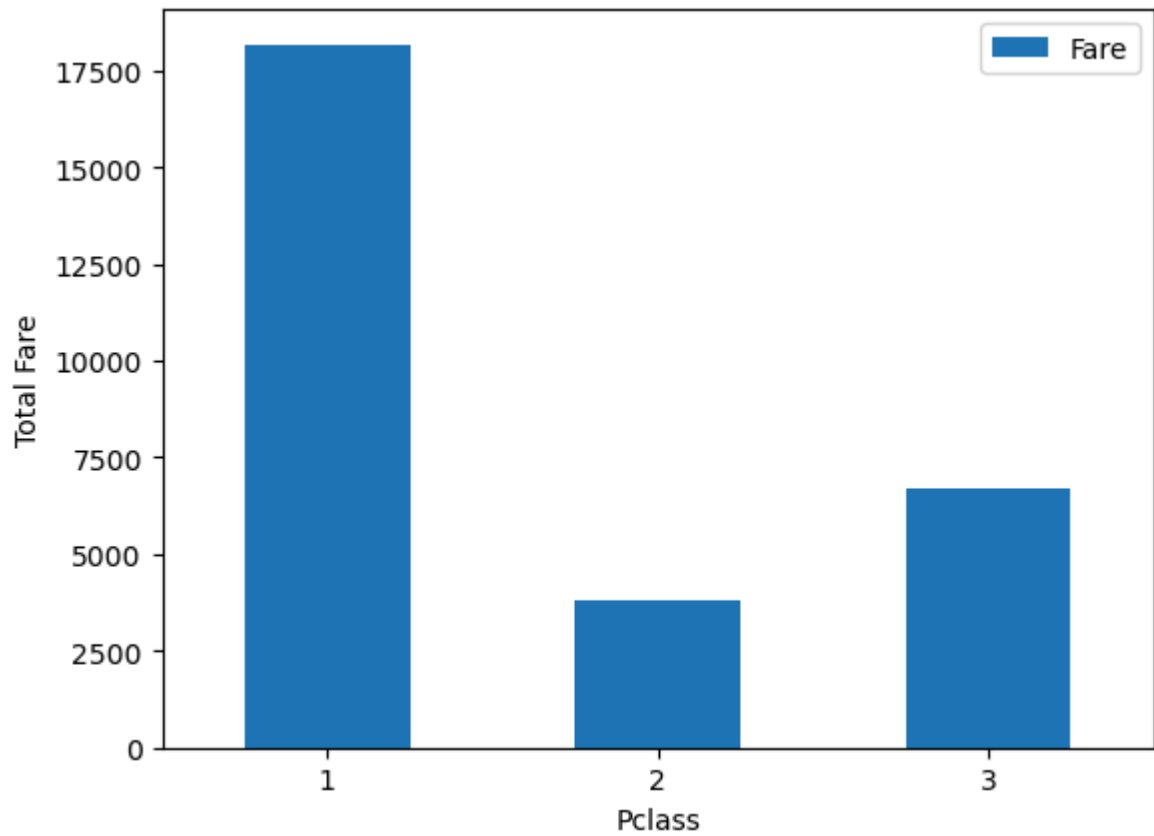
```
In [34]: class_fare = train.pivot_table(index='Pclass', values='Fare')
class_fare.plot(kind='bar')
plt.xlabel('Pclass')
plt.ylabel('Avg. Fare')
plt.xticks(rotation=0)
plt.show()
```



```
In [35]: class_fare = train.pivot_table(index='Pclass', values='Fare')
class_fare.plot(kind='pie', y='Fare', autopct='%1.1f%%', legend=False)
plt.ylabel('')
plt.show()
```



```
In [36]: 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

```
In [37]: train_len = len(train)
df = pd.concat([train, test], axis=0)
df = df.reset_index(drop=True)
df.head()
```

Out[37]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0

In [38]: df.tail()

Out[38]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
1304	1305	NaN	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236
1305	1306	NaN	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758
1306	1307	NaN	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262
1307	1308	NaN	3	Ware, Mr. Frederick	male	NaN	0	0	359309
1308	1309	NaN	3	Peter, Master. Michael J	male	NaN	1	1	2668

In [39]: df.isnull().sum()

```
Out[39]: 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
```

```
In [40]: df = df.drop(columns=['Cabin'], axis=1)
```

```
In [41]: df['Age'].mean()
```

```
Out[41]: 29.881137667304014
```

```
In [42]: df['Age'] = df['Age'].fillna(df['Age'].mean())
         df['Fare'] = df['Fare'].fillna(df['Fare'].mean())
```

```
In [43]: df['Embarked'].mode()[0]
```

```
Out[43]: 'S'
```

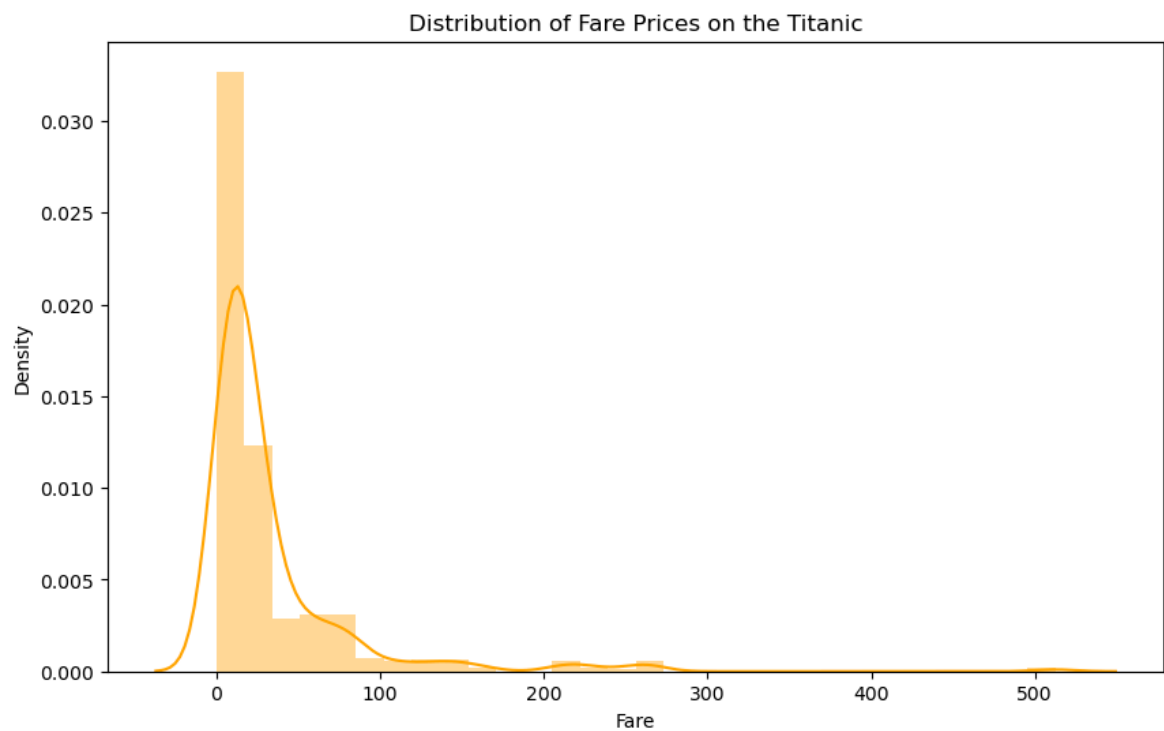
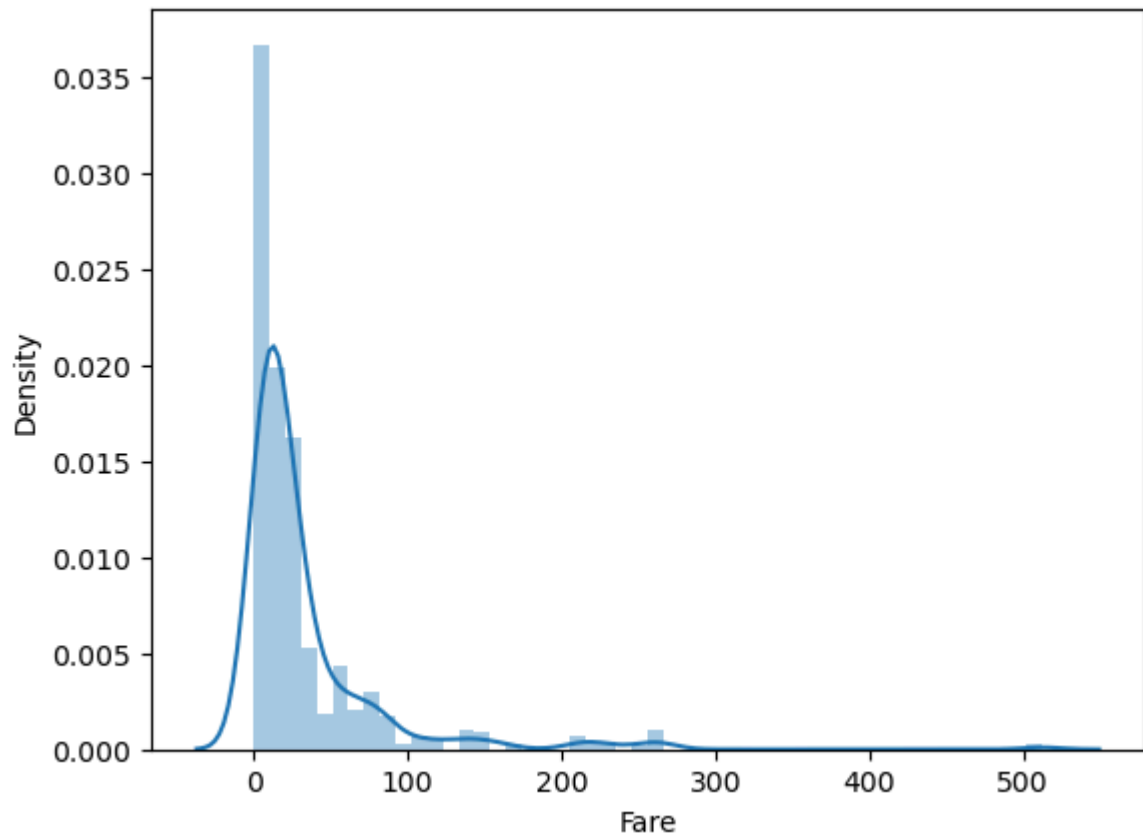
```
In [44]: df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])
```

```
In [45]: sns.distplot(df['Fare'])
```

```
Out[45]: <Axes: xlabel='Fare', ylabel='Density'>
```

```
In [46]: plt.figure(figsize=(10, 6))
         sns.distplot(df['Fare'], bins=30, kde=True, color='orange')

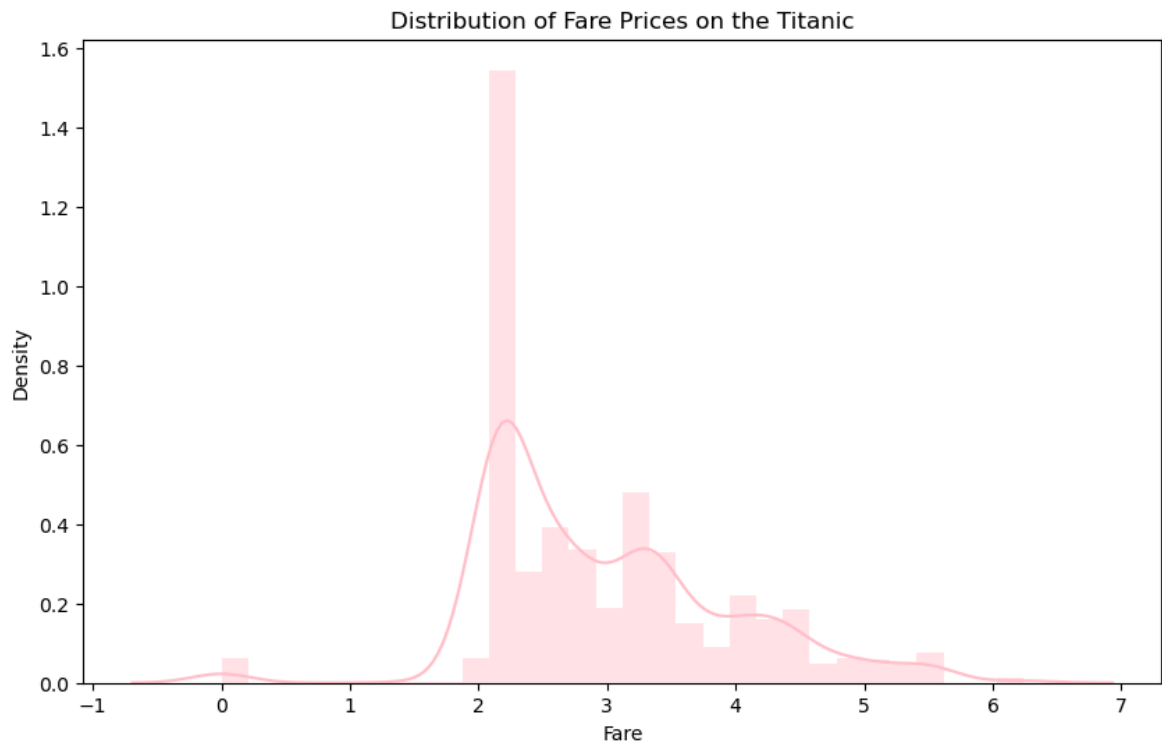
         plt.title('Distribution of Fare Prices on the Titanic')
         plt.xlabel('Fare')
         plt.ylabel('Density')
         plt.show()
```



```
In [47]: df['Fare'] = np.log(df['Fare']+1)
```

```
In [48]: plt.figure(figsize=(10, 6))
sns.distplot(df['Fare'], bins=30, kde=True, color='pink')

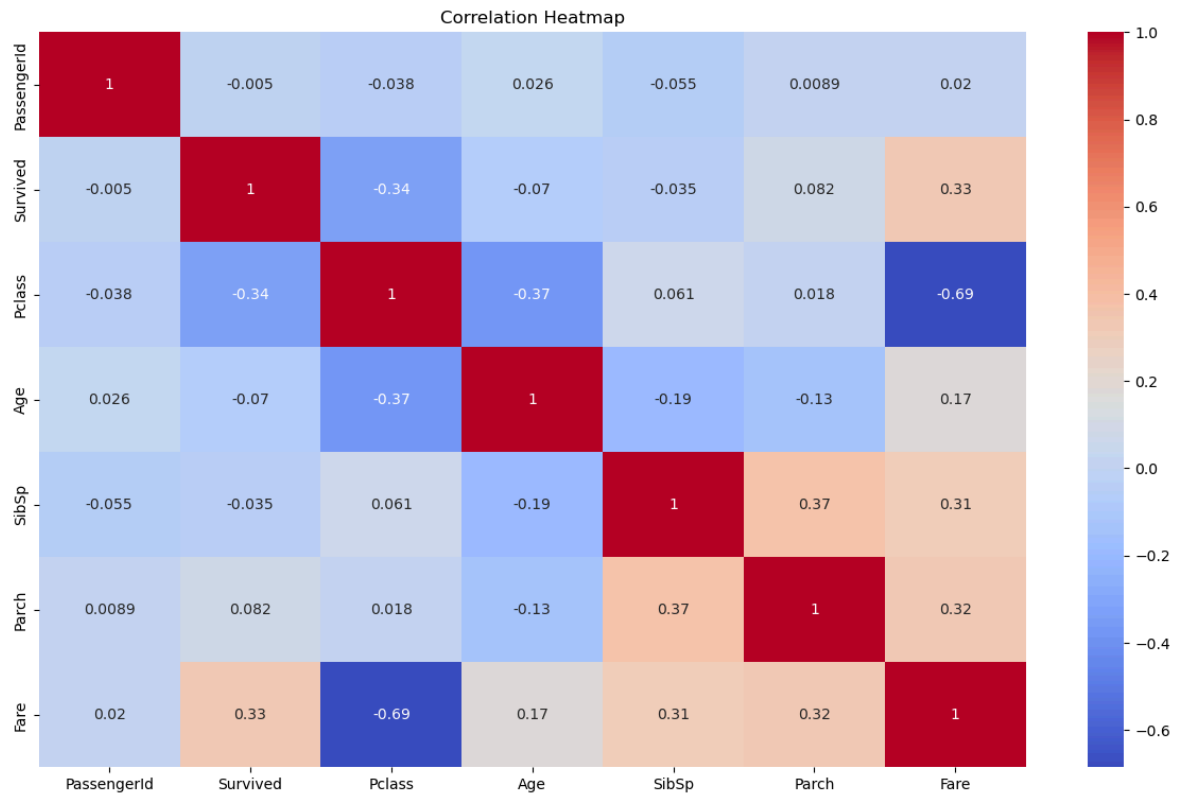
plt.title('Distribution of Fare Prices on the Titanic')
plt.xlabel('Fare')
plt.ylabel('Density')
plt.show()
```



## Correlation Matrix

```
In [49]: 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')



```
In [50]: df.head()
```

```
Out[50]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	2.11
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	4.28
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	2.16
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	3.99
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	2.20

```
In [51]: df = df.drop(columns=['Name', 'Ticket'], axis=1)
df.head()
```



Out[51]:

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

## Label Coding

```
In [52]: from sklearn.preprocessing import LabelEncoder
cols = ['Sex', 'Embarked']
le = LabelEncoder()

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

Out[52]:

	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

```
In [53]: train_len = int(0.8 * len(df))
train = df.iloc[:train_len, :]
test = df.iloc[train_len:, :]
```

In [54]: train.head()


Out[54]:

	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

In [55]: test.head()

Out[55]:

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



In [56]: `X = train.drop(columns=['PassengerId', 'Survived'], axis=1)`  
`y = train['Survived']`

In [57]: `X.head()`

Out[57]:

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

## Model Training

In [58]:

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

# Load and preprocess the Titanic dataset
df = pd.read_csv('Titanic-Dataset.csv')

# Encode categorical variables to numeric values
df['Sex'] = df['Sex'].map({'male': 0, 'female': 1}) # Map 'male' to 0 and 'female' to 1
df['Embarked'] = df['Embarked'].map({'C': 0, 'Q': 1, 'S': 2}) # Map 'C', 'Q', 'S' to 0, 1, 2

# Fill missing values in features with the median or mode
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)

# Handle missing target values (Survived)
df['Survived'].dropna(inplace=True) # Drop rows where 'Survived' is NaN

# Define features (X) and target (y)
X = df[['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']]
y = df['Survived']

# Define the classify function
def classify(model):
```

```

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, ra

# Train the model
model.fit(X_train, y_train)

# Compute accuracy on the test set
accuracy = model.score(X_test, y_test)
print(f'Accuracy: {accuracy}')
```

```

# Perform cross-validation
cv_scores = cross_val_score(model, X_train, y_train, cv=5) # 5-fold cross-v
cv_score = cv_scores.mean() # Mean of CV scores
print(f'CV Score: {cv_score}')
```

```

# Initialize a model, e.g., Logistic Regression
model = LogisticRegression(max_iter=1000)

# Call the classify function with the model
classify(model)
```

Accuracy: 0.8071748878923767  
CV Score: 0.7978453596678262

```
In [59]: from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
classify(model)
```

Accuracy: 0.8071748878923767  
CV Score: 0.7978453596678262

```
In [60]: from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
classify(model)
```

Accuracy: 0.726457399103139  
CV Score: 0.7590730557737627

```
In [61]: from sklearn.ensemble import ExtraTreesClassifier
model = ExtraTreesClassifier()
classify(model)
```

Accuracy: 0.7802690582959642  
CV Score: 0.7889799124677366

```
In [62]: pip install xgboost
```

Requirement already satisfied: xgboost in c:\users\saddi\anaconda3\lib\site-packa  
ges (2.1.3)Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: numpy in c:\users\saddi\anaconda3\lib\site-package  
s (from xgboost) (1.26.4)  
Requirement already satisfied: scipy in c:\users\saddi\anaconda3\lib\site-package  
s (from xgboost) (1.13.1)

```
In [63]: from xgboost import XGBClassifier
model = XGBClassifier()
classify(model)
```

Accuracy: 0.7892376681614349  
CV Score: 0.79947256200202

In [64]: `pip install lightgbm`

Requirement already satisfied: lightgbm in c:\users\saddi\anaconda3\lib\site-packages (4.5.0)  
Requirement already satisfied: numpy>=1.17.0 in c:\users\saddi\anaconda3\lib\site-packages (from lightgbm) (1.26.4)  
Requirement already satisfied: scipy in c:\users\saddi\anaconda3\lib\site-packages (from lightgbm) (1.13.1)  
Note: you may need to restart the kernel to use updated packages.

In [65]: `from lightgbm import LGBMClassifier  
model = LGBMClassifier()  
classify(model)`

File "C:\Users\saddi\anaconda3\Lib\site-packages\joblib\externals\loky\backend\context.py", line 282, in \_count\_physical\_cores  
raise ValueError(f"found {cpu\_count\_physical} physical cores < 1")

[illegible]

30/43

31/43

32/43



file:///C:/Users/saddi/Downloads/Titanic\_337\_(1).html

[illegible]

file:///C:/Users/saddi/Downloads/Titanic\_337\_(1).html

[illegible]

file:///C:/Users/saddi/Downloads/Titanic\_337\_(1).html

38/43

39/43

[LightGBM] [Warning] No further splits with positive gain, best gain: -inf  
CV Score: 0.8248905846706318

```
In [66]: !pip install catboost
         from catboost import CatBoostClassifier
         model = CatBoostClassifier(verbose=0)
         classify(model)
```

Requirement already satisfied: catboost in c:\users\saddi\anaconda3\lib\site-packages (1.2.7)  
Requirement already satisfied: graphviz in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (0.20.3)  
Requirement already satisfied: matplotlib in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (3.9.2)  
Requirement already satisfied: numpy<2.0,>=1.16.0 in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (1.26.4)  
Requirement already satisfied: pandas>=0.24 in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (2.2.2)  
Requirement already satisfied: scipy in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (1.13.1)  
Requirement already satisfied: plotly in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (5.24.1)  
Requirement already satisfied: six in c:\users\saddi\anaconda3\lib\site-packages (from catboost) (1.16.0)  
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\saddi\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in c:\users\saddi\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2024.1)  
Requirement already satisfied: tzdata>=2022.7 in c:\users\saddi\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2023.3)  
Requirement already satisfied: contourpy>=1.0.1 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (1.2.0)  
Requirement already satisfied: cyclor>=0.10 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (0.11.0)  
Requirement already satisfied: fonttools>=4.22.0 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (4.51.0)  
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (1.4.4)  
Requirement already satisfied: packaging>=20.0 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (24.1)  
Requirement already satisfied: pillow>=8 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (10.4.0)  
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\saddi\anaconda3\lib\site-packages (from matplotlib->catboost) (3.1.2)  
Requirement already satisfied: tenacity>=6.2.0 in c:\users\saddi\anaconda3\lib\site-packages (from plotly->catboost) (8.2.3)  
Accuracy: 0.8340807174887892  
CV Score: 0.8203905285602063

## Complete Model Training with Full Data

```
In [77]: model = LGBMClassifier()
         model.fit(X, y)
```



```
[LightGBM] [Info] Number of positive: 342, number of negative: 549
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing
was 0.000262 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 221
[LightGBM] [Info] Number of data points in the train set: 891, number of used fea
tures: 7
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.383838 -> initscore=-0.473288
[LightGBM] [Info] Start training from score -0.473288
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

Out[77]:

▼ LGBMClassifier ⓘ

LGBMClassifier()

In [78]: test.head()

Out[78]:

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

In [79]: X\_test = test.drop(columns=['PassengerId', 'Survived'], axis=1)

In [80]: X\_test.head()

Out[80]:

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

In [81]: pred = model.predict(X\_test)  
pred

```
Out[81]: array([0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1,
                0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0,
                0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0,
                0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0,
                0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0,
                1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1,
                0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0,
                0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1,
                0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1,
                1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,
                0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0,
                1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0],
                dtype=int64)
```

```
In [82]: sub = pd.read_csv('gender_submission.csv')
sub.head()
```

```
Out[82]:
```

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

```
In [83]: 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
```

```
In [84]: sub.head()
```

```
Out[84]:
```

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

```
In [85]: sub.to_csv('submission.csv', index=False)
```

```
In [88]: y_pred = voting_model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
```

```
print(f" Final Accuracy: {accuracy * 100:.2f}%")
```

Final Accuracy: 88.18%