

```
!pip install kaggle
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.10/dist-packages (1.5.16)
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.10/dist-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from kaggle) (2023.7.22)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.31.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from kaggle) (4.66.1)
Requirement already satisfied: python-slugify in /usr/local/lib/python3.10/dist-packages (from kaggle) (8.0.1)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.0.4)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from kaggle) (6.0.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.10/dist-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.2.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.4)
```

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

```
Choose Files kaggle.json
• kaggle.json(application/json) - 75 bytes, last modified: 22/9/2023 - 100% done
Saving kaggle.json to kaggle.json
{'kaggle.json': b'{"username": "dhanyadeepikachukka", "key": "5335873ef4d5831cdebea367ba684842"}'}
```

```
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
```

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

```
Choose Files Titanic-Dataset.csv
• Titanic-Dataset.csv(text/csv) - 61194 bytes, last modified: 22/9/2023 - 100% done
Saving Titanic-Dataset.csv to Titanic-Dataset.csv
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
```

```
import pandas as pd
df = pd.read_csv('Titanic-Dataset.csv')
```

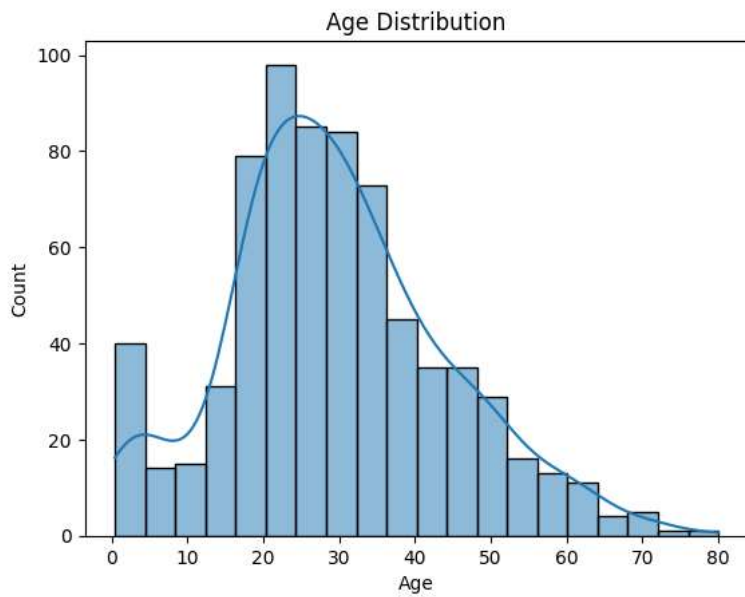
```
# Check for missing values
null_counts = df.isnull().sum()
print(null_counts)
```

```
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
```

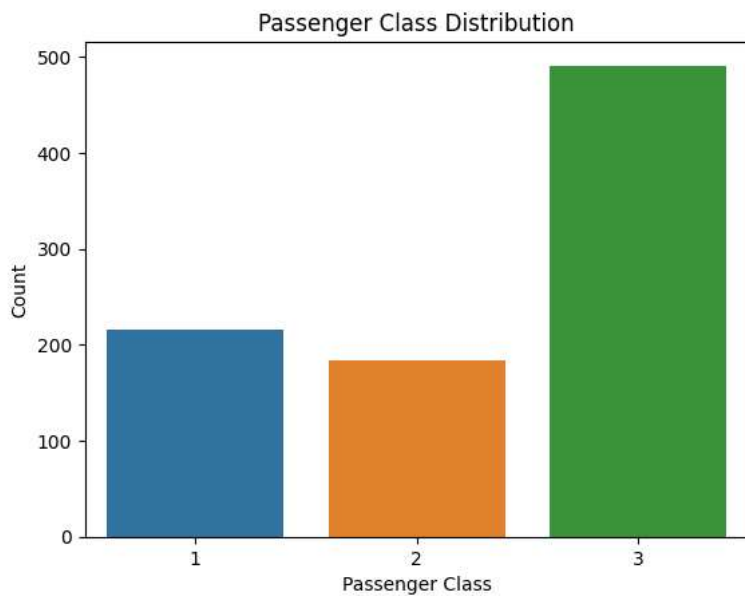
```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Plot a histogram for the 'Age' column
sns.histplot(data=df, x='Age', bins=20, kde=True)
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Age Distribution')
```

```
plt.show()
```

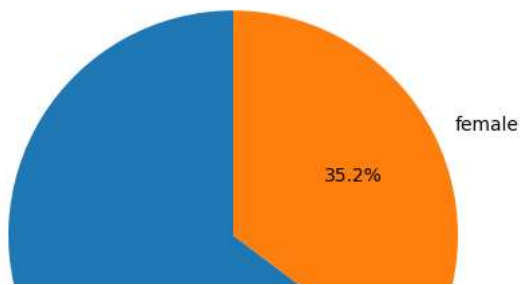


```
# Plot a bar chart for the 'Pclass' column
sns.countplot(data=df, x='Pclass')
plt.xlabel('Passenger Class')
plt.ylabel('Count')
plt.title('Passenger Class Distribution')
plt.show()
```

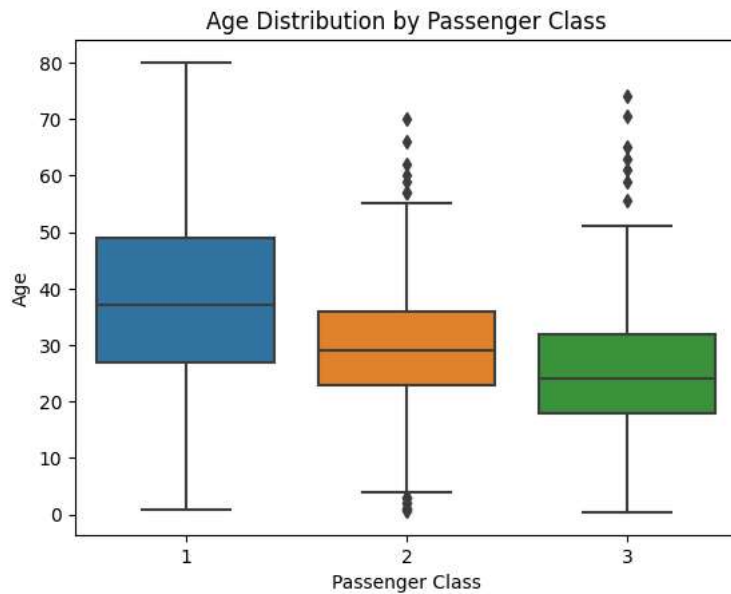


```
# Plot a pie chart for the 'Sex' column
sex_counts = df['Sex'].value_counts()
plt.pie(sex_counts, labels=sex_counts.index, autopct='%1.1f%%', startangle=90)
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.title('Gender Distribution')
plt.show()
```

Gender Distribution



```
# Plot a box plot for 'Age' by passenger class
sns.boxplot(data=df, x='Pclass', y='Age')
plt.xlabel('Passenger Class')
plt.ylabel('Age')
plt.title('Age Distribution by Passenger Class')
plt.show()
```



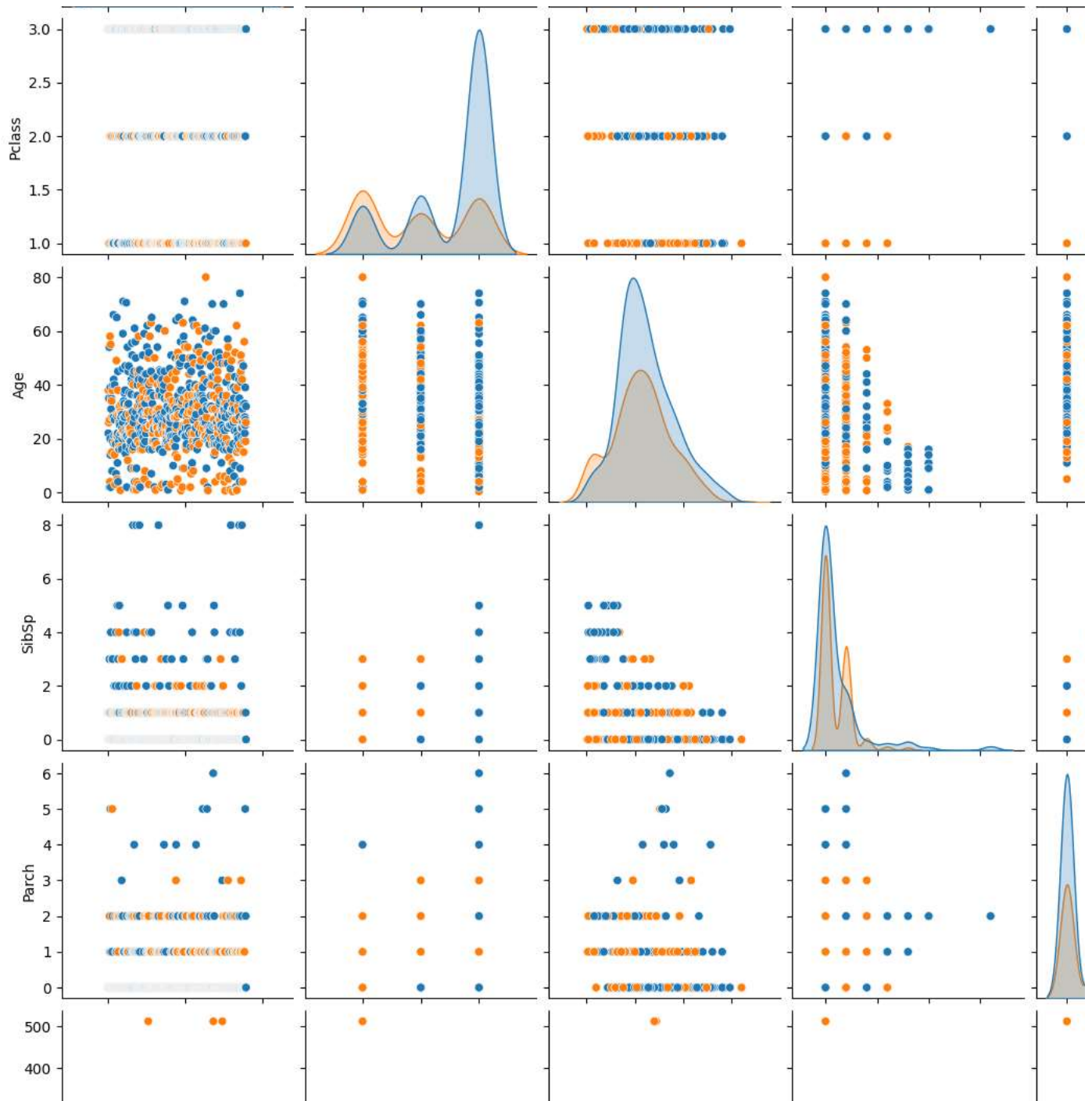
```
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()
```

```
<ipython-input-16-b3519a62c2f8>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version correlation_matrix = df.corr()
```

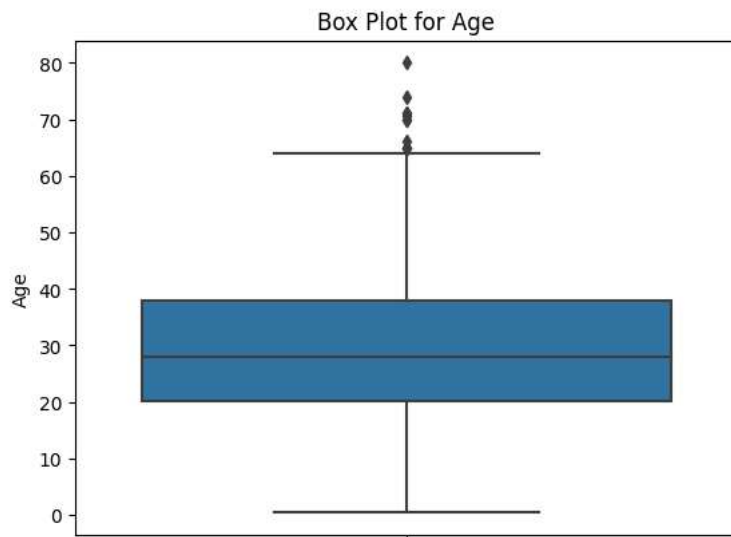
Correlation Heatmap



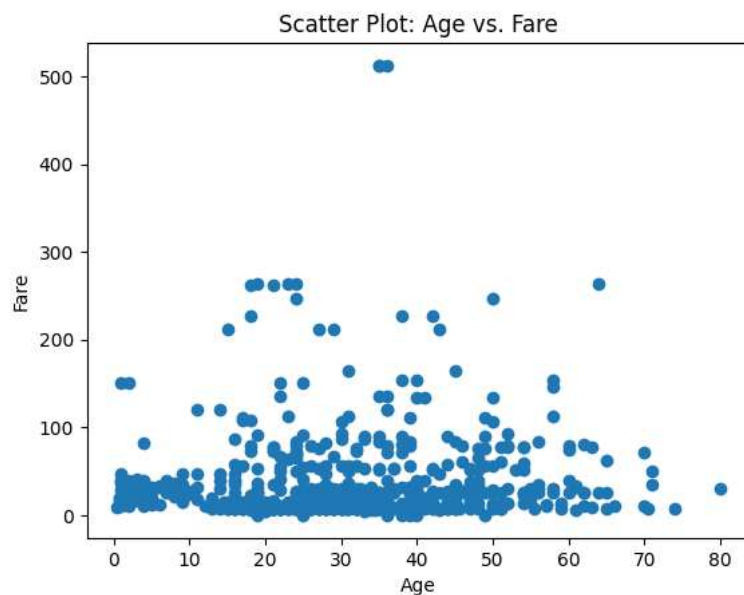
```
sns.pairplot(df, hue='Survived')  
plt.title('Pair Plot')  
plt.show()
```



```
# Plot a box plot for 'Age'
sns.boxplot(data=df, y='Age')
plt.ylabel('Age')
plt.title('Box Plot for Age')
plt.show()
```



```
# Plot a scatter plot for 'Age' vs. 'Fare'
plt.scatter(df['Age'], df['Fare'])
plt.xlabel('Age')
plt.ylabel('Fare')
plt.title('Scatter Plot: Age vs. Fare')
plt.show()
```



```
X = df.drop('Survived', axis=1) # Independent variables
y = df['Survived'] # Dependent variable
```

```
print(X.head()) # Display the first few rows of the independent variables
print(y.head()) # Display the first few rows of the dependent variable
```

PassengerId	Pclass	Name \
0	1	3 Braund, Mr. Owen Harris
1	2	1 Cumings, Mrs. John Bradley (Florence Briggs Th...
2	3	3 Heikkinen, Miss. Laina
3	4	1 Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	5	3 Allen, Mr. William Henry

Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
-----	-----	-------	-------	--------	------	-------	----------

```

0   male  22.0    1    0      A/5 21171  7.2500  NaN    S
1  female  38.0    1    0      PC 17599  71.2833  C85    C
2  female  26.0    0    0  STON/O2. 3101282  7.9250  NaN    S
3  female  35.0    1    0      113803  53.1000  C123   S
4   male  35.0    0    0      373450  8.0500  NaN    S
0    0
1    1
2    1
3    1
4    0
Name: Survived, dtype: int64

```

```

label_encoder = LabelEncoder()
X['Sex'] = label_encoder.fit_transform(X['Sex'])
print(X['Sex'].head())

```

```

0    1
1    0
2    0
3    0
4    1
Name: Sex, dtype: int64

```

```

scaler = StandardScaler()
X[['Age', 'Fare']] = scaler.fit_transform(X[['Age', 'Fare']])
print(X[['Age', 'Fare']].head())

```

```

      Age      Fare
0 -0.530377 -0.502445
1  0.571831  0.786845
2 -0.254825 -0.488854
3  0.365167  0.420730
4  0.365167 -0.486337

```

```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
print("X_train shape:", X_train.shape)
print("X_test shape:", X_test.shape)
print("y_train shape:", y_train.shape)
print("y_test shape:", y_test.shape)

```

```

X_train shape: (712, 11)
X_test shape: (179, 11)
y_train shape: (712,)
y_test shape: (179,)

```