```
# Name:Atharva Kangralkar
# Roll no : 54
# CS - AIML - A
# Colab link:- https://colab.research.google.com/drive/1AR_JgPSIOc_-HN9vlyI6153KcMOCtOKR?usp=sharing
# Lab Assignment 8


# import the necessary libraries for data analysis and visualization.
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt



# Load Dataset into df object.
# (Use pd.read_csv )
df = pd.read_csv('Titanic-Dataset.csv')


# Prints information about the Data Frame.
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
```

```
# Displaying the first few rows of the Data Frame df.
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
# Displaying the last few rows of the Data Frame df.
df.tail()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

```
# Calculating the sum of missing values in each column of the Data Frame df.
df.isna().sum()
```

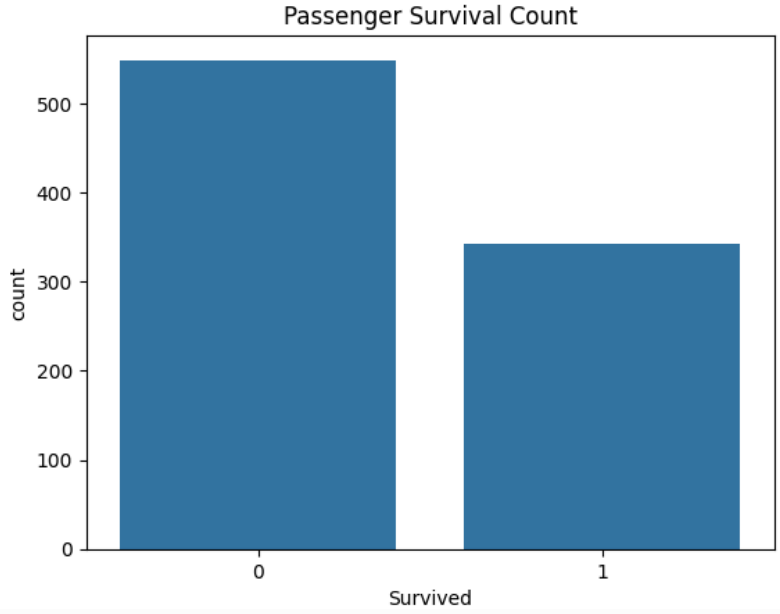
	0
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

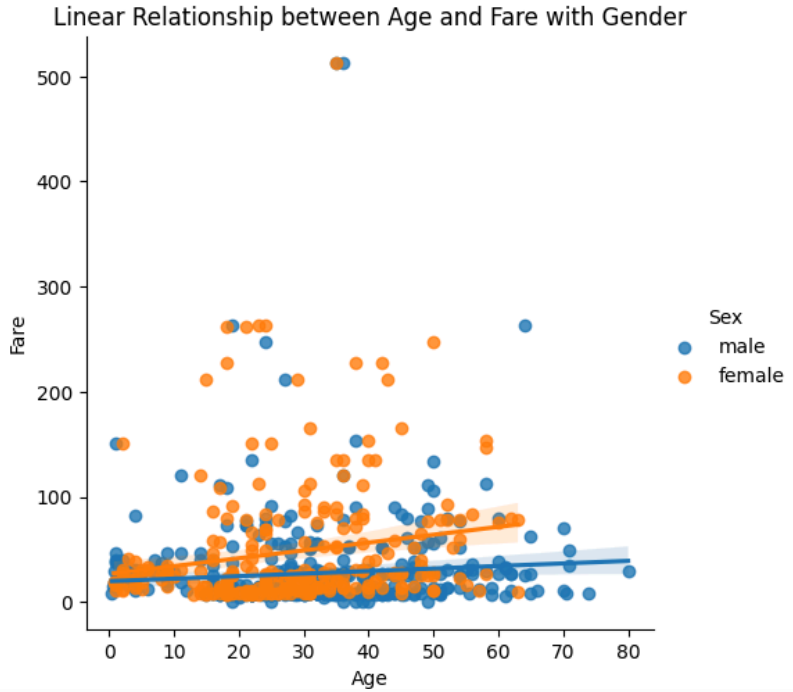
```
# Generating descriptive statistics summary for the Data Frame df.
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
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	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
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	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
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	512.329200

```
# Countplot to visualize the count of passengers who survived or not.
# (Use  sns.countplot )
sns.countplot(x='Survived', data=df)
plt.title('Passenger Survival Count')
plt.show()
```




```
# Visualize the linear relationship between age and fare, with gender as a hue.
# (Use  sns.lmplot)
sns.lmplot(x='Age', y='Fare', hue='Sex', data=df)
plt.title('Linear Relationship between Age and Fare with Gender')
plt.show()
```

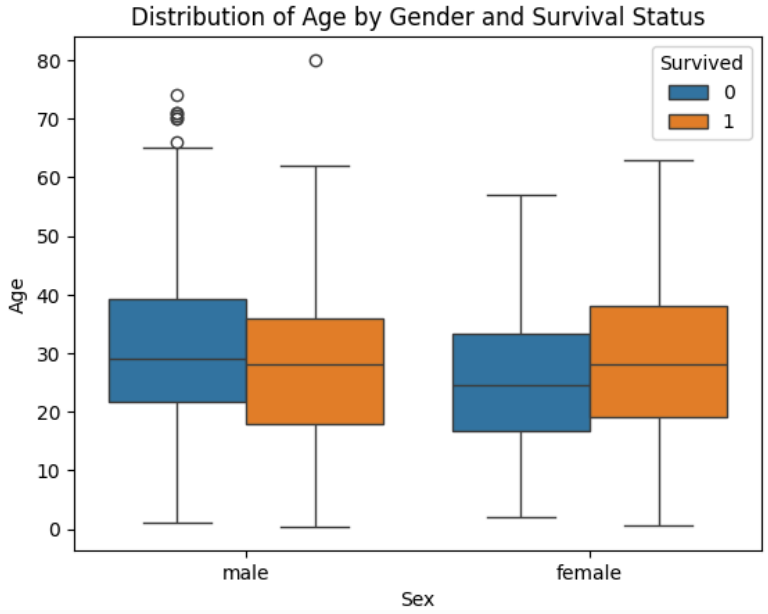



```
# sns.load_dataset("/content/Titanic-Dataset.csv")
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-10-b8ca54dfee76> in <cell line: 0>()
----> 1 sns.load_dataset("/content/Titanic-Dataset.csv")

/usr/local/lib/python3.11/dist-packages/seaborn/utils.py in load_dataset(name, cache, data_home, **kws)
    571         if not os.path.exists(cache_path):
    572             if name not in get_dataset_names():
--> 573                 raise ValueError(f"'{name}' is not one of the example datasets.")
    574             urlretrieve(url, cache_path)
    575         full_path = cache_path

ValueError: '/content/Titanic-Dataset.csv' is not one of the example datasets.
```

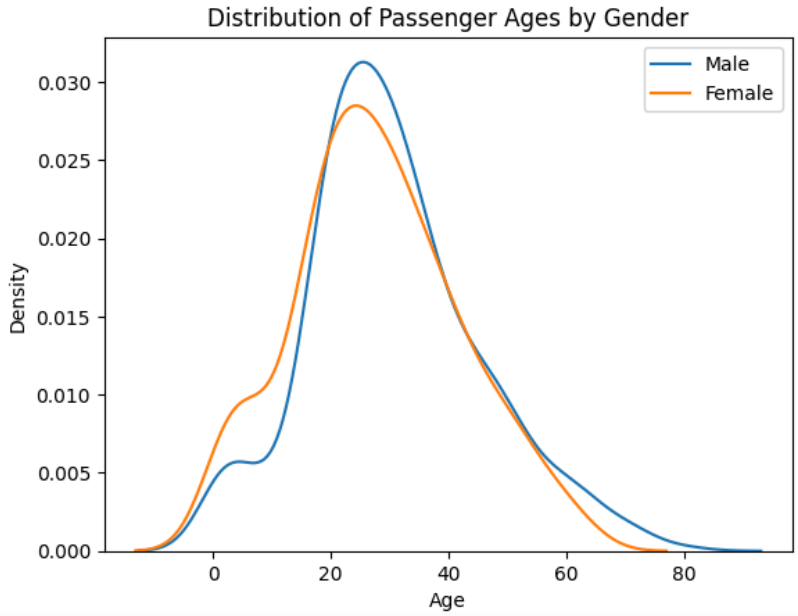
Next steps: Explain error

```
# Plotting box plot for distribution of age with respect to each gender and survival status.

# (Use sns.boxplot )
sns.boxplot(x='Sex', y='Age', hue='Survived', data=df)
plt.title('Distribution of Age by Gender and Survival Status')
plt.show()
```



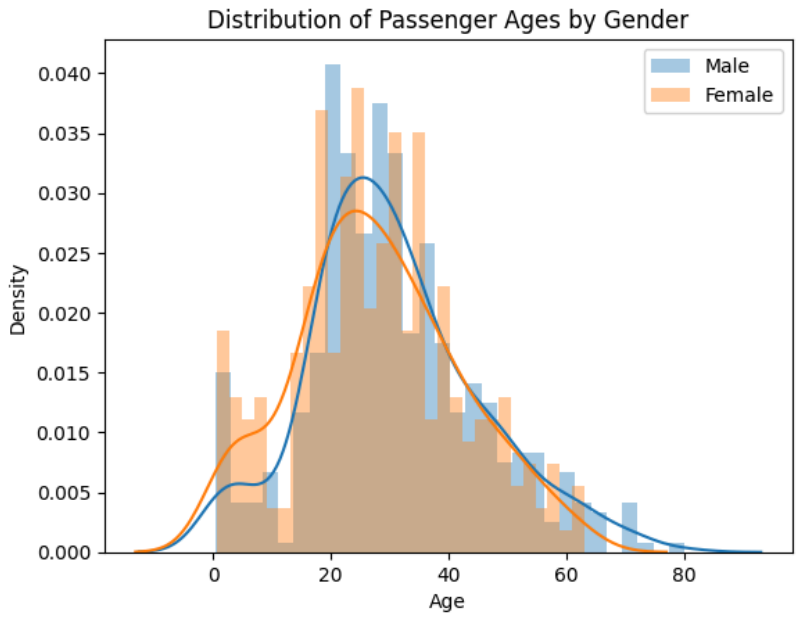

```
# Visualize the distribution of passenger ages by gender.

# (Use sns.distplot )

sns.distplot(df[df['Sex'] == 'male']['Age'], kde=True, hist=False, label='Male')
sns.distplot(df[df['Sex'] == 'female']['Age'], kde=True, hist=False, label='Female')

plt.title('Distribution of Passenger Ages by Gender')
plt.xlabel('Age')
plt.ylabel('Density')
plt.legend()

plt.show()
```

```
<ipython-input-12-3346a3b9111e>:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `kdeplot` (an axes-level function for kernel density plots).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df['Sex'] == 'male']['Age'], kde=True, hist=False, label='Male')
<ipython-input-12-3346a3b9111e>:6: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `kdeplot` (an axes-level function for kernel density plots).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df['Sex'] == 'female']['Age'], kde=True, hist=False, label='Female')
```




```
# Visualize the distribution of passenger ages by gender with histogram.
# ( Use sns.distplot ,hist=True)
sns.distplot(df[df['Sex'] == 'male']['Age'], kde=True, hist=True, label='Male', bins=30)
sns.distplot(df[df['Sex'] == 'female']['Age'], kde=True, hist=True, label='Female', bins=30

plt.title('Distribution of Passenger Ages by Gender')
plt.xlabel('Age')
plt.ylabel('Density')
plt.legend()
```

```
<ipython-input-14-aacafb94d6dd>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df['Sex'] == 'male']['Age'], kde=True, hist=True, label='Male', bins=30)
<ipython-input-14-aacafb94d6dd>:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df[df['Sex'] == 'female']['Age'], kde=True, hist=True, label='Female', bins=30)
```




```
# according to the countplot, many people did not survive
# The dataset showed that the majority of passengers were male (577 out of 891), indicating a gender imbalance in the passenger group
# Most people were aged between 20 and 40 for both male and female
```