In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
#Reading the CSV file
titanic1 = pd.read_csv("TITANIC.csv")
```

In [3]:
```
#1
median_Fare = titanic1['Fare'].median()
print("Median fare: ${:.2f} ".format(median_Fare))
```

```
Median fare: $14.45
```

In [4]:
```
#2
mean_Age_Of_Male = titanic1.loc[titanic1['Sex'] == 'male', 'Age'].mean()
print("Mean age: {:.2f} years".format(mean_Age_Of_Male))
```

```
Mean age: 30.27 years
```

In [5]:
```
#3
mode_No_Of_Siblings = titanic1['SibSp'].mode()[0]
print("Mode of the no. of siblings/spouses aboard: {}".format(mode_No_Of_Siblings))
```

```
Mode of the no. of siblings/spouses aboard: 0
```

In [6]:
```
#4
ticket_Price_Range = titanic1['Fare'].max() - titanic1['Fare'].min()
print("Range of ticket prices was: ${:.2f}".format(ticket_Price_Range))
```

```
Range of ticket prices was: $512.33
```

In [7]:
```
#5
cheapest_Ticket_Cost = titanic1['Fare'].min()
print("Cost of cheapest ticket: ${:.2f}".format(cheapest_Ticket_Cost))
```

```
Cost of cheapest ticket: $0.00
```

In [8]:
```
#6
correlation_Sex_Survival = titanic1['Sex'].astype('category').cat.codes.corr(titanic1['Survived'])
print("Correlation between Sex and Survival is: {:.2f}".format(correlation_Sex_Survival))
```

```
Correlation between Sex and Survival is: -1.00
```

In [9]:
```
#7
variance_Passenger_Class = titanic1['Pclass'].var()
std_Dev_Passenger_Class = titanic1['Pclass'].std()
print("Variance: {:.2f}".format(variance_Passenger_Class))
print("Standard Deviation: {:.2f}".format(std_Dev_Passenger_Class))
```

```
Variance: 0.71
Standard Deviation: 0.84
```

8

1. Clearly displays the average age of male passengers

2. Demonstrates the most frequently occuring number of siblings/spouses aboard

3. Displays the spread between the cheapest and the most expensive tickets

4. This indicates the least ticket price

5. This demonstrates whether there is a connection between gender and survival

6. This measures the degree of the passenger class values' dispersion. More spread is indicated by a bigger variance

In [15]:
```
#9
#Filling missing values with the mean of the column
titanic1['Age'].fillna(titanic1['Age'].mean())
```

Out[15]:
```
0      34.50000
1      47.00000
2      62.00000
3      27.00000
4      22.00000
         ...
413    30.27259
414    39.00000
415    38.50000
416    30.27259
417    30.27259
Name: Age, Length: 418, dtype: float64
```

In [16]:
```
#9
#Deleting rows with missing data
titanic1.dropna(axis=0)
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
12	904	1	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	female	23.0	1	0	21228	82.2667	B45	S
14	906	1	1	Chaffee, Mrs. Herbert Fuller (Carrie Constance...	female	47.0	1	0	W.E.P. 5734	61.1750	E31	S
24	916	1	1	Ryerson, Mrs. Arthur Larned (Emily Maria Borie)	female	48.0	1	3	PC 17608	262.3750	B57 B59 B63 B66	C
26	918	1	1	Ostby, Miss. Helene Ragnhild	female	22.0	0	1	113509	61.9792	B36	C
28	920	0	1	Brady, Mr. John Bertram	male	41.0	0	0	113054	30.5000	A21	S
...	...	...	...	...	...	...	...	...	...	...	...	...
404	1296	0	1	Frauenthal, Mr. Isaac Gerald	male	43.0	1	0	17765	27.7208	D40	C
405	1297	0	2	Nourney, Mr. Alfred (Baron von Drachstedt")"	male	20.0	0	0	SC/PARIS 2166	13.8625	D38	C
407	1299	0	1	Widener, Mr. George Dunton	male	50.0	1	1	113503	211.5000	C80	C
411	1303	1	1	Minahan, Mrs. William Edward (Lillian E Thorpe)	female	37.0	1	0	19928	90.0000	C78	Q
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	C

87 rows × 12 columns

In [10]:
```
#11
#Identifying outliers in the dataset
Q1 = titanic1.quantile(0.25)
Q3 = titanic1.quantile(0.75)

#computing interquartile range
interquartileRange = Q3 - Q1

#Defining lower bounds and upper bounds to identify outliers
lowerBound = Q1 - 1.5 * interquartileRange
upperBound = Q3 + 1.5 * interquartileRange

#Identifying outliers
outliers = ((titanic1 < lowerBound) | (titanic1 > upperBound)).any(axis=1)

#Displaying rows containing outliers
print(titanic1[outliers])
```

```
     PassengerId  Survived  Pclass  \
4            896         1       3
7            899         0       2
12           904         1       1
21           913         0       3
23           915         0       1
..           ...       ...     ...
407         1299         0       1
409         1301         1       3
411         1303         1       1
414         1306         1       1
417         1309         0       3

                                              Name     Sex   Age  SibSp  \
4        Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female  22.0      1
7                         Caldwell, Mr. Albert Francis    male  26.0      1
12       Snyder, Mrs. John Pillsbury (Nelle Stevenson)  female  23.0      1
21                            Olsen, Master. Artur Karl    male   9.0      0
23                       Williams, Mr. Richard Norris II    male  21.0      0
..                                                ...     ...   ...    ...
407                          Widener, Mr. George Dunton    male  50.0      1
409                          Peacock, Miss. Treasteall  female   3.0      1
411  Minahan, Mrs. William Edward (Lillian E Thorpe)  female  37.0      1
414                        Oliva y Ocana, Dona. Fermina  female  39.0      0
417                           Peter, Master. Michael J    male   NaN      1

     Parch              Ticket      Fare Cabin Embarked
4        1             3101298   12.2875   NaN        S
7        1              248738   29.0000   NaN        S
12       0               21228   82.2667   B45        S
21       1             C 17368    3.1708   NaN        S
23       1            PC 17597   61.3792   NaN        C
..     ...                 ...       ...   ...      ...
407      1              113503  211.5000   C80        C
409      1  SOTON/O.Q. 3101315   13.7750   NaN        S
411      0               19928   90.0000   C78        Q
414      0            PC 17758  108.9000  C105        C
417      1                2668   22.3583   NaN        C

[126 rows x 12 columns]
C:\Users\Samuel\AppData\Local\Temp\ipykernel_19136\2375327813.py:2: FutureWarning: The default value of numeric_only in DataFrame.quantile is depre
cated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  Q1 = titanic1.quantile(0.25)
C:\Users\Samuel\AppData\Local\Temp\ipykernel_19136\2375327813.py:3: FutureWarning: The default value of numeric_only in DataFrame.quantile is depre
cated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  Q3 = titanic1.quantile(0.75)
C:\Users\Samuel\AppData\Local\Temp\ipykernel_19136\2375327813.py:13: FutureWarning: Automatic reindexing on DataFrame vs Series comparisons is d
eprecated and will raise ValueError in a future version. Do `left, right = left.align(right, axis=1, copy=False)` before e.g. `left == right`
  outliers = ((titanic1 < lowerBound) | (titanic1 > upperBound)).any(axis=1)
```

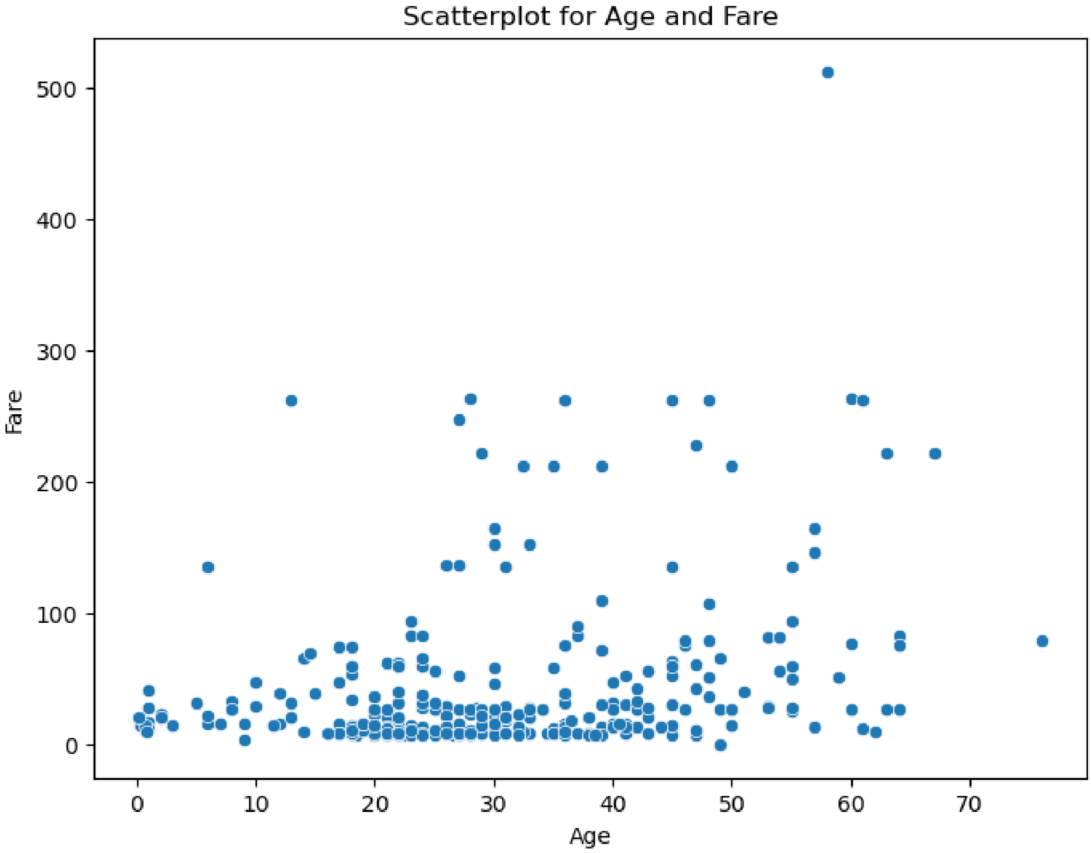
In [11]:

```
#10
#Checking for missing values
missing_values = titanic1.isnull().sum()
print("Missing values are:\n", missing_values)
```

```
Missing values are:
 PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

In [12]:

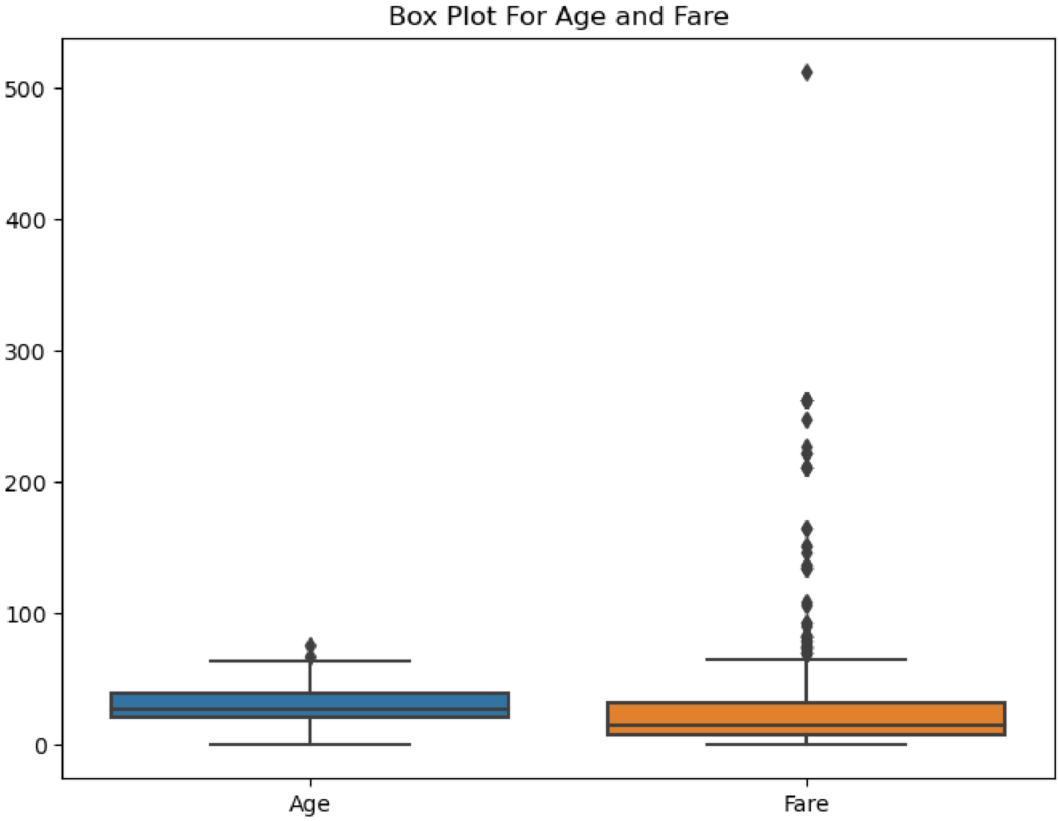
```
#11
#Using scatter plot to visualize the data
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Age', y='Fare', data=titanic1)
plt.title("Scatterplot for Age and Fare")
plt.show()
```



Scatter plots highlight the relationships between continuous variables, identifying outliers as data points deviationg significantly from the entire pattern, which assists in the provision of valuable insight into correlation between variables. Based on where they are in relation to other data points, scatter plots can be used to identify outliers. By observing points that are far from the primary cluster, it is possible to find outliers in the range of "age" to "fare".

In [13]:

```
#11
#Using boxplot to visualize the data
plt.figure(figsize=(8, 6))
sns.boxplot(data=titanic1[['Age', 'Fare']])
plt.title("Box Plot For Age and Fare")
plt.show()
```

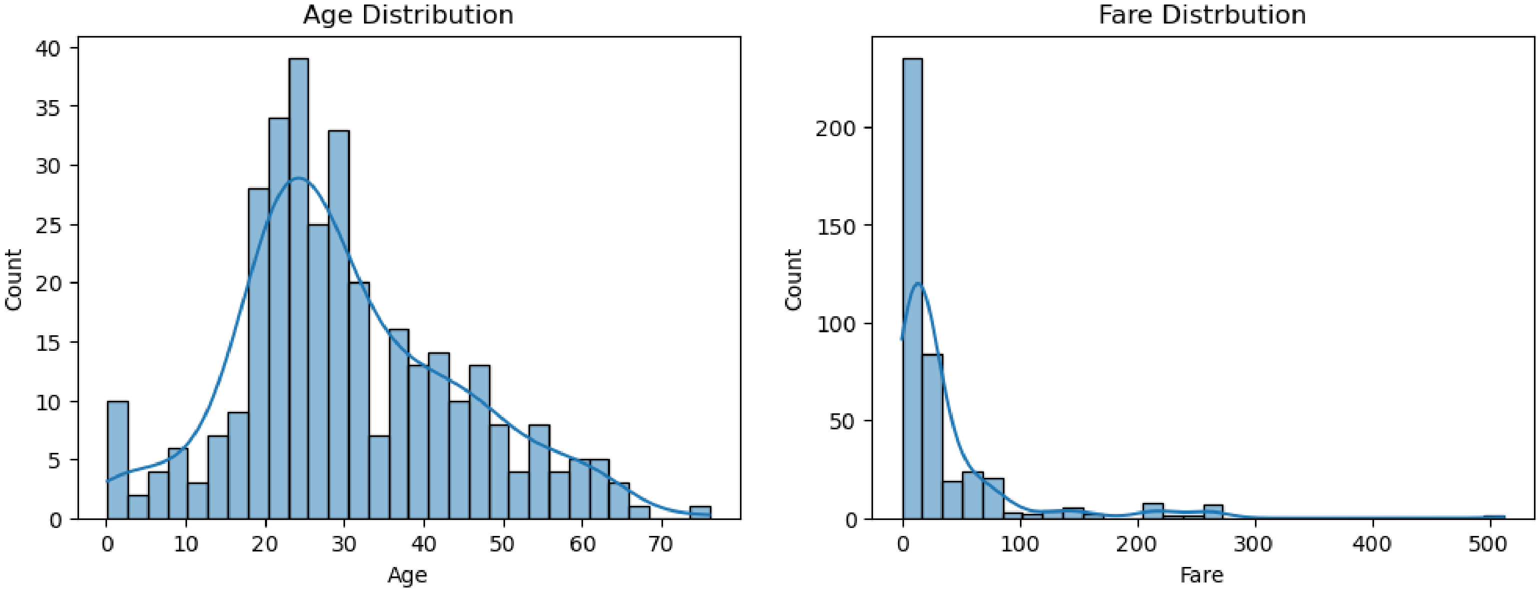


Box plots examine the spread of features and locate outliers in each category to depict the data distribution and identify outliers within the Interquartile Range. They give a clear picture of the distribution of the data, making it easy to spot outliers and compare features such as 'age' and 'fare'.

In [14]:

```
#11
#Using histogram to visualize the data
plt.figure(figsize=(12, 4))
plt.subplot(1, 2, 1)
sns.histplot(titanic1['Age'], bins=30, kde=True)
```

```
plt.title("Age Distribution")
plt.subplot(1, 2, 2)
sns.histplot(titanic1['Fare'], bins=30, kde=True)
plt.title("Fare Distrbution")
plt.show()
```



Histograms display the continuous random variable visually, thus they can be used to spot outliers and areas with unusually high data densities by looking for odd peaks and gaps. By depicting age and fare, the presence of outliers can be identified.