```
INPUT

# TASK 1

#IMPORT MODULES

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

# Load data
data =
pd.read_csv("/content/test.csv")

# Display first and last rows of the
dataframe
print("HEAD DATA")
print(data.head())
print("*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*")

print("TAIL DATA")
print(data.tail())
print("*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*")
```

```
# DATA DESCRIPTION
print(" DISPLAY BASIC STATISTICS AND
INFO")
print(data.describe())
print(data.info())
print("*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*")

# HANDLING MISSING VALUES
print(" CHECK FOR MISSING VALUES")
print(data.isnull().sum())
print("*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*")

# VERIFY MISSING VALUES
print(" VERIFY THAT THERE ARE NO MORE
MISSING VALUES")
print(data.isnull().sum())

# HANDLING DUPLICATE VALUES
print(" CHECK FOR DUPLICATE ROWS")
print(data.duplicated().sum())

# CONSTRUCT HISTOGRAM
print("PLOT AGE DISTRIBUTION")
```

```
plt.figure(figsize=(6, 3))
sns.histplot(data["Age"], kde=True)
plt.title("AGE DISTRIBUTION")
plt.xlabel("AGE")
plt.ylabel("COUNT")

plt.show()

#CONSTRUCT HISTOGRAM 2
print("SURVIVAL BY GENDER")
plt.figure(figsize=(6, 3))
sns.histplot(data=data,x="Sex",hue="Se
x")
plt.title("SURVIVAL BY GENDER")
plt.xlabel("Gender")
plt.ylabel("Count")
plt.show()

#CONSTRUCT SCATTER PLOT
print("SCATTER PLOT")
plt.figure(figsize=(6, 3))
sns.histplot(data=data,x="Age",y="Fare
",hue="Sex")
plt.title("SCATTERPLOT OF AGE AND
FARE")
plt.xlabel("AGE")
plt.ylabel("FARE")
```

```
plt.show()
```

OUTPUT

```
HEAD DATA
   PassengerId  Pclass
Name     Sex  \
0          892       3
Kelly, Mr. James    male
1          893       3                    Wilkes,
Mrs. James (Ellen Needs)  female
2          894       2
Myles, Mr. Thomas Francis    male
3          895       3
Wirz, Mr. Albert    male
4          896       3  Hirvonen, Mrs.
Alexander (Helga E Lindqvist)  female
```

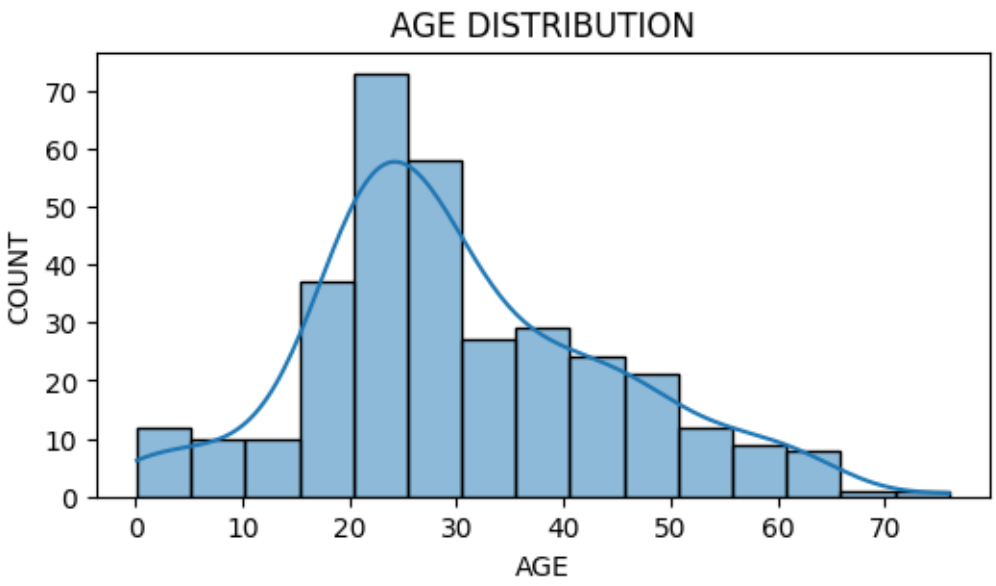
```
    Age  SibSp  Parch   Ticket     Fare
Cabin Embarked
0  34.5      0      0   330911   7.8292
NaN        Q
1  47.0      1      0   363272   7.0000
NaN        S
2  62.0      0      0   240276   9.6875
NaN        Q
3  27.0      0      0   315154   8.6625
NaN        S
4  22.0      1      1  3101298  12.2875
NaN        S
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*
TAIL DATA
     PassengerId  Pclass
Name     Sex   Age  SibSp  \
413         1305       3
Spector, Mr. Woolf    male   NaN      0
414         1306       1  Oliva y Ocana,
Dona. Fermina  female  39.0      0
415         1307       3  Saether, Mr.
Simon Sivertsen    male  38.5      0
416         1308       3           Ware,
Mr. Frederick    male   NaN      0
417         1309       3      Peter,
Master. Michael J    male   NaN      1

     Parch              Ticket      Fare
Cabin Embarked
```

```
413          0            A.5. 3236    8.0500
NaN          S
414          0             PC 17758  108.9000
C105         C
415          0  SOTON/O.Q. 3101262    7.2500
NaN          S
416          0              359309    8.0500
NaN          S
417          1                2668   22.3583
NaN          C
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*
 DISPLAY BASIC STATISTICS AND INFO
       PassengerId      Pclass         Age
SibSp        Parch         Fare
count   418.000000  418.000000  332.000000
418.000000  418.000000  417.000000
mean   1100.500000    2.265550   30.272590
0.447368    0.392344   35.627188
std     120.810458    0.841838   14.181209
0.896760    0.981429   55.907576
min     892.000000    1.000000    0.170000
0.000000    0.000000    0.000000
25%     996.250000    1.000000   21.000000
0.000000    0.000000    7.895800
50%    1100.500000    3.000000   27.000000
0.000000    0.000000   14.454200
75%    1204.750000    3.000000   39.000000
1.000000    0.000000   31.500000
max    1309.000000    3.000000   76.000000
8.000000    9.000000  512.329200
```

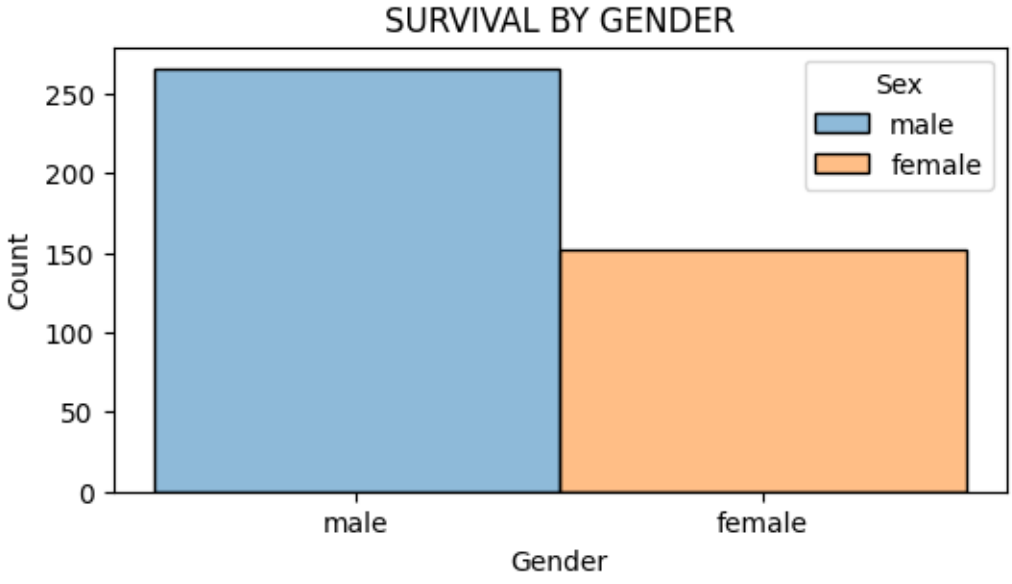
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Pclass       418 non-null    int64  
 2   Name         418 non-null    object 
 3   Sex          418 non-null    object 
 4   Age          332 non-null    float64
 5   SibSp        418 non-null    int64  
 6   Parch        418 non-null    int64  
 7   Ticket       418 non-null    object 
 8   Fare         417 non-null    float64
 9   Cabin        91 non-null     object 
 10  Embarked     418 non-null    object 
dtypes: float64(2), int64(4), object(5)
memory usage: 36.0+ KB
None
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*
 CHECK FOR MISSING VALUES
PassengerId      0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
```

```
Cabin            327
Embarked           0
dtype: int64
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-
*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*
 VERIFY THAT THERE ARE NO MORE MISSING
VALUES
PassengerId        0
Pclass             0
Name               0
Sex                0
Age               86
SibSp              0
Parch              0
Ticket             0
Fare               1
Cabin            327
Embarked           0
dtype: int64
 CHECK FOR DUPLICATE ROWS
0
```

PLOT AGE DISTRIBUTION



SURVIVAL BY GENDER



SCATTER PLOT