NAME: DESMOND ELORM HONU

COURSE: DATA SCIENCE

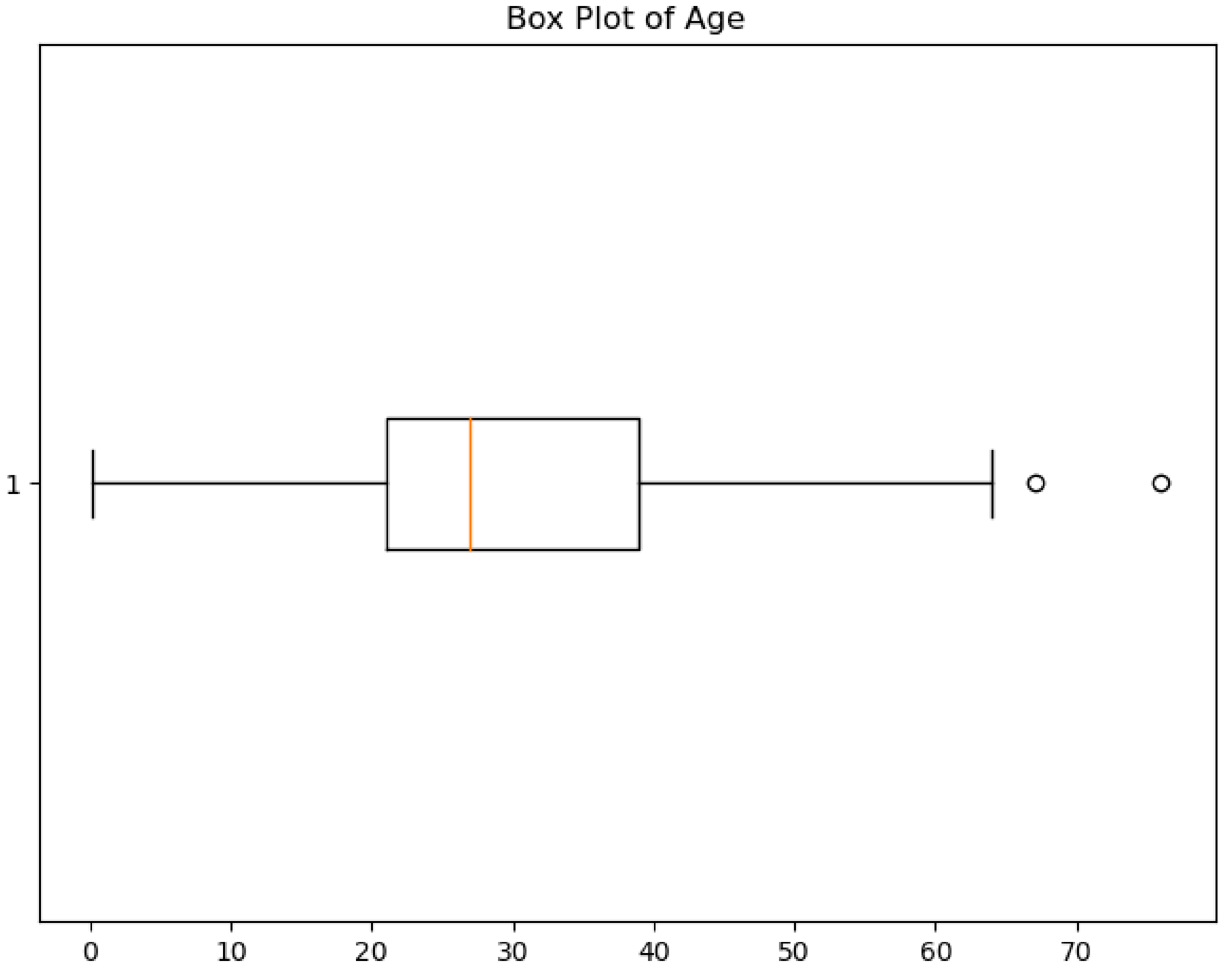
ROLL NUMBER: 10211100281

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Load the Titanic dataset (assuming the dataset is in a CSV file)
titanic_df = pd.read_csv('Desktop/TITANIC.csv')

# 1.median fare paid
median_fare = titanic_df['Fare'].median()
print("Median Fare Paid:", median_fare)

# 2.age of male passengers
mean_age_male = titanic_df[titanic_df['Sex'] == 'male']['Age'].mean()
print("Mean Age of Male Passengers:", mean_age_male)

# 3.number of siblings/spouses on the Titanic
sibsp_mode = titanic_df['SibSp'].mode()[0]
print("Mode of Siblings/Spouses Aboard:", sibsp_mode)

# 4.ticket prices
ticket_price_range = titanic_df['Fare'].max() -
titanic_df['Fare'].min()
print("Range of Ticket Prices:", ticket_price_range)

# 5.cheapest ticket
cheapest_ticket_cost = titanic_df['Fare'].min()
print("Cheapest Ticket Cost:", cheapest_ticket_cost)

# 6. Sex and Survival.
correlation_sex_survival = titanic_df[['Sex',
'Survived']].corr().iloc[0, 0]
print("Correlation between Sex and Survival:",
correlation_sex_survival)

# 7.standard deviation of the passenger class.
variance_passenger_class = titanic_df['Pclass'].var()
std_deviation_passenger_class = titanic_df['Pclass'].std()
print("Variance of Passenger Class:", variance_passenger_class)
print("Standard Deviation of Passenger Class:",
std_deviation_passenger_class)


# Method 1: Removing rows with missing data
titanic_df_cleaned1 = titanic_df.dropna()

# Method 2: Filling missing data with mean or median
titanic_df_cleaned2 = titanic_df.fillna({'Age':
titanic_df['Age'].median()})

missing_values = titanic_df_cleaned1.isna().sum()
```

```
print("Missing Values After Cleaning (Method 1):\n", missing_values)

missing_values = titanic_df_cleaned2.isna().sum()
print("Missing Values After Cleaning (Method 2):\n", missing_values)




# Box plot for Age
plt.figure(figsize=(8, 6))
plt.boxplot(titanic_df['Age'].dropna(), vert=False)
plt.title("Box Plot of Age")
plt.show()
```

```
Median Fare Paid: 14.4542
Mean Age of Male Passengers: 30.27273170731707
Mode of Siblings/Spouses Aboard: 0
Range of Ticket Prices: 512.3292
Cheapest Ticket Cost: 0.0
Correlation between Sex and Survival: 1.0
Variance of Passenger Class: 0.7086904638968277
Standard Deviation of Passenger Class: 0.8418375519640519
Missing Values After Cleaning (Method 1):
 PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
Missing Values After Cleaning (Method 2):
 PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

```
C:\Users\MY PC\AppData\Local\Temp\ipykernel_6644\1362693881.py:29:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only
valid columns or specify the value of numeric_only to silence this
warning.
  correlation_sex_survival = titanic_df[['Sex',
'Survived']].corr().iloc[0, 0]
```



```
8. Explain your findings in questions (1-7):

The median fare paid by passengers was determined in question 1.
The mean age of male passengers was calculated in question 2.
The mode of the number of siblings/spouses aboard was found in question 3
.
The range of ticket prices was computed in question 4.
The cost of the cheapest ticket was determined in question 5.
The correlation between Sex and Survival was calculated in question 6.
The variance and standard deviation of the passenger class were found in
question 7.
```