Data Analytics Internship Project:

Name: Mohammad Altamas

Task 2: Exploratory Data Analysis

Introduction:

This project analyzes student performance data using Python.

Tools Used:

- Python
- Pandas
- Matplotlib
- Seaborn

Results:

(Insert screenshots here)

Conclusion:

EDA and visualization successfully performed.

Task 2(EDA):

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

# ----------------------------
# Create Sample Dataset
# ---------------------------- data
= {
    'Gender': ['Male', 'Female', 'Male', 'Female',
'Male', 'Female'],
    'Math': [78, 66, 44, 88, 55, 95],
    'Science': [80, 89, 92, 85, 90, 98],
    'English': [66, 42, 85, 90, 56, 96]
}  df =
pd.DataFrame(data)

# ----------------------------
# Add Total Column
# ---------------------------- df['Total']
= df['Math'] + df['Science'] +
df['English']

# ----------------------------
# Display Dataset
# ---------------------------- print("\n=====
DATASET PREVIEW =====") print(df)

# ----------------------------
# Average Marks (Only Numeric Columns)
# ----------------------------
print("\n===== AVERAGE MARKS =====")
```

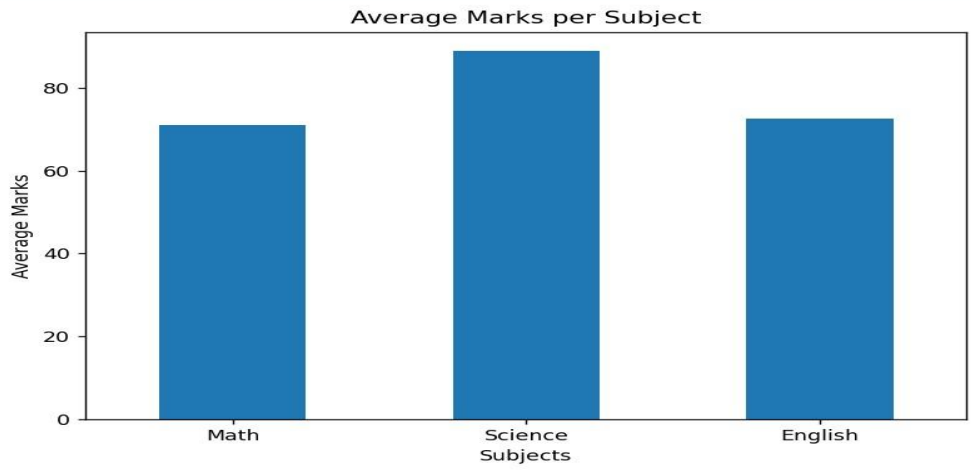
```
print(df.mean(numeric_only=True))

# ----------------------------
# Summary Statistics
# ---------------------------- print("\n=====
SUMMARY STATISTICS =====")
print(df.describe())

# ----------------------------
# VISUALIZATION SECTION
# ----------------------------

# 1 Bar Chart - Subject Average plt.figure()
df[['Math', 'Science',
'English']].mean().plot(kind='bar')
plt.title("Average Marks per Subject")
plt.ylabel("Average Marks")
plt.xlabel("Subjects")
plt.xticks(rotation=0)
plt.tight_layout() plt.show()

# 2 Gender-wise Total Score plt.figure()
sns.barplot(x='Gender', y='Total',
data=df) plt.title("Total Score by
Gender") plt.tight_layout() plt.show()

# 3 Correlation Heatmap (Only Numeric)
plt.figure() sns.heatmap(df.corr(numeric_only=True),
annot=True) plt.title("Correlation Heatmap")
```

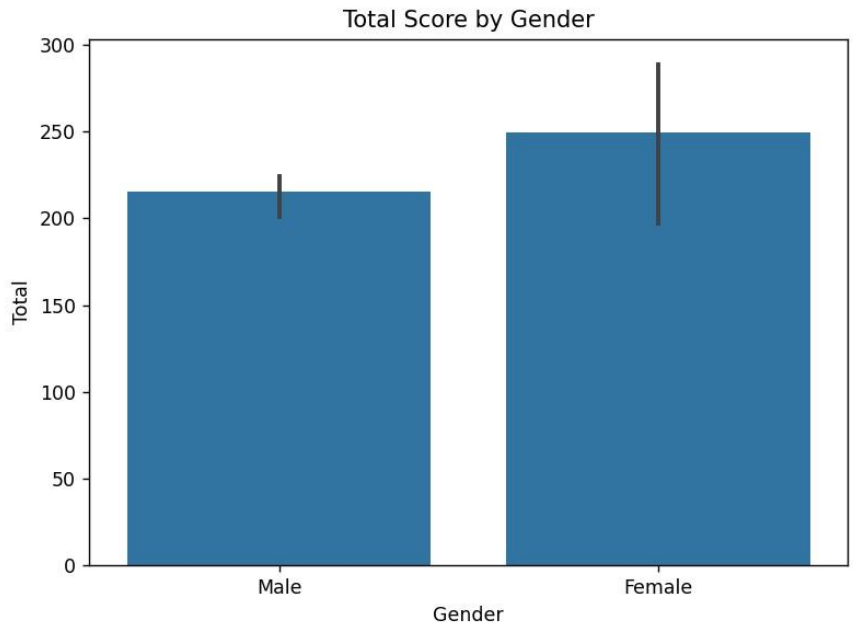
```
plt.tight_layout() plt.show()
 print("\n☑ Project Executed
Successfully!")
```

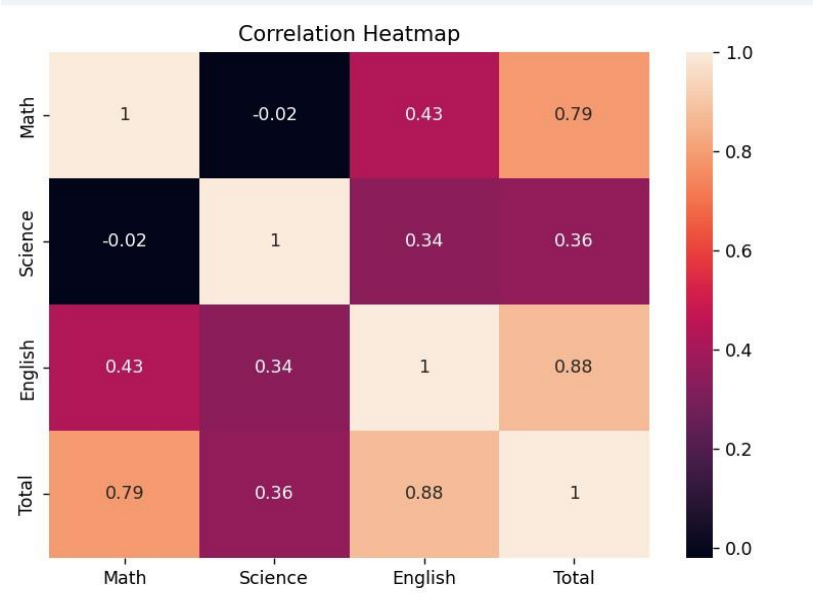
Output:

```
PS C:\Users\Mohd Ayan\OneDrive\Desktop\New folder> python -u "c:\Users\Mohd Ayan\OneDrive\Desktop\New folder\codealpha_(EDA).py"

===== DATASET PREVIEW =====
   Gender  Math  Science  English  Total
0    Male    78       80       66    224
1  Female    66       89       42    197
2    Male    44       92       85    221
3  Female    88       85       90    263
4    Male    55       90       56    201
5  Female    95       98       96    289

===== AVERAGE MARKS =====
Math       71.0
Science    89.0
English    72.5
Total     232.5
dtype: float64

===== SUMMARY STATISTICS =====
            Math    Science    English       Total
count   6.00000   6.000000   6.00000    6.000000
mean   71.00000  89.000000  72.50000  232.500000
std    19.61632   6.131884  21.25794   36.275336
min    44.00000  80.000000  42.00000  197.000000
25%    57.75000  86.000000  58.50000  206.000000
50%    72.00000  89.500000  75.50000  222.500000
75%    85.50000  91.500000  88.75000  253.250000
max    95.00000  98.000000  96.00000  289.000000
```

Graphs:







Conclusion:

The project successfully analyzed and visualized student performance data using Python. The use of Pandas, Matplotlib, and Seaborn helped in identifying patterns, trends, and correlations effectively.

Project Completed Successfully

Tools Used: Python, Pandas, Matplotlib, Seaborn