Data Processing and Visualization: Given a dataset containing information about students' test scores, fetch the data from an API, calculate the average score, and create a bar chart to visualize the data.

In [1]:

```
import requests
from bs4 import BeautifulSoup
import csv
import pandas as pd
```

In [2]:

```
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, l
ike Gecko) Chrome/80.0.3987.149 Safari/537.36'
}

# Make GET request with custom headers
response = requests.get("https://roycekimmons.com/tools/generated_data/exams", headers=he
aders).text
#print(response)
```

In [3]:

```
soup = BeautifulSoup(response, 'html.parser')
table = soup.find('table')

# Initialize a list to store data
data_list = []

# Extract data from each row of the table
for row in table.find_all('tr')[1:]:  # Skip the header row
    columns = row.find_all('td')

    # Check if the row has enough columns
    if len(columns) >= 8:
        data = {
            "gender": columns[0].get_text(strip=True),
            "math_score": columns[5].get_text(strip=True),
            "reading_score": columns[6].get_text(strip=True),
            "writing_score": columns[7].get_text(strip=True),
        }
        data_list.append(data)

# Define the CSV file path
csv_file_path = "student_scores.csv"

# Write data to CSV file
with open(csv_file_path, mode='w', newline='') as file:
    fieldnames = ["gender", "math_score", "reading_score", "writing_score"]
    writer = csv.DictWriter(file, fieldnames=fieldnames)

    writer.writeheader()  # Write CSV header
    for data in data_list:
        writer.writerow(data)

print(f"CSV file '{csv_file_path}' created successfully.")
```

```
CSV file 'student_scores.csv' created successfully.
```

In [4]:

```
df = pd.read_csv('student_scores.csv')
df
```

Out[4]:

	gender	math_score	reading_score	writing_score
0	male	83	90	83
1	male	63	46	49
2	male	84	78	75
3	male	45	44	38
4	female	65	64	69
5	male	64	63	60
6	male	41	41	40
7	male	75	74	77
8	female	99	97	98
9	female	53	72	70

In [5]:

```
# Calculate average scores
average_math_score = sum(math_scores) / len(math_scores)
average_reading_score = sum(reading_scores) / len(reading_scores)
average_writing_score = sum(writing_scores) / len(writing_scores)

# Create bar chart
categories = ['Math', 'Reading', 'Writing']
average_scores = [average_math_score, average_reading_score, average_writing_score]

plt.bar(categories, average_scores, color=['blue', 'green', 'orange'])
plt.xlabel('Subject')
plt.ylabel('Average Score')
plt.title('Average Test Scores')
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[5], line 2
      1 # Calculate average scores
----> 2 average_math_score = sum(math_scores) / len(math_scores)
      3 average_reading_score = sum(reading_scores) / len(reading_scores)
      4 average_writing_score = sum(writing_scores) / len(writing_scores)

NameError: name 'math_scores' is not defined
```

In [ ]:

```
```