Finding Outliers Using IQR - Step 4/4

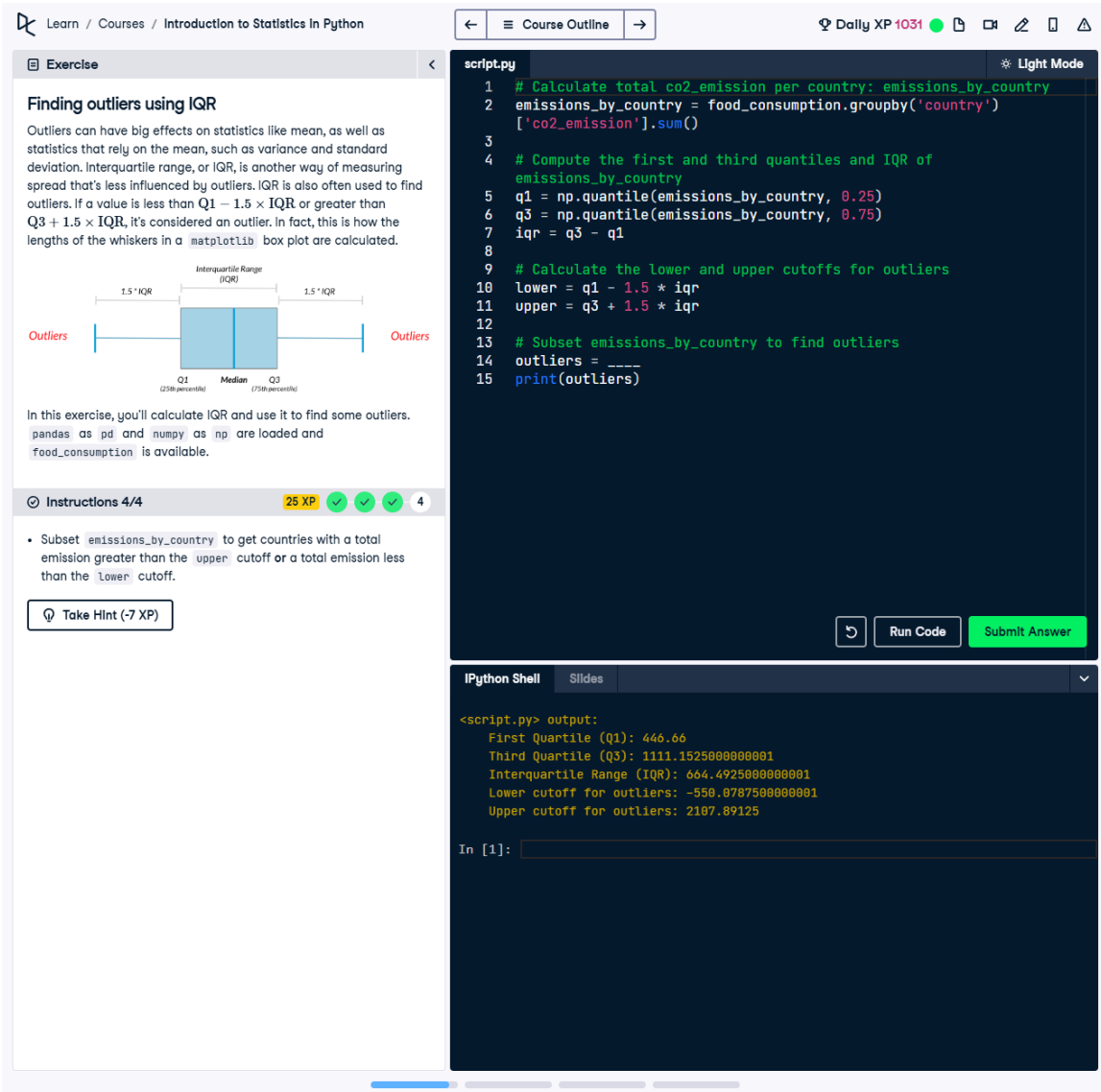

Figure 4: Screenshot showing the subsetting of data to identify countries with outlier CO2 emissions.

Question

Outliers can have big effects on statistical measures such as mean, variance, and standard deviation. Interquartile range (IQR) is another way of measuring spread that is less influenced by outliers. In this final step, you'll subset `emissions_by_country` to identify countries with CO2 emissions that are outliers based on the calculated `lower` and `upper` cutoffs.

Instructions:
1. Subset `emissions_by_country` to include countries with a total emission greater than the `upper` cutoff or less than the `lower` cutoff.
2. Print the resulting subset of countries with outlier emissions.

Corrected Code Solution

```
import pandas as pd
import numpy as np

# Calculate total CO2 emissions per country
emissions_by_country = food_consumption.groupby('country')
['co2_emission'].sum()

# Compute Q1 and Q3
q1 = np.quantile(emissions_by_country, 0.25)
q3 = np.quantile(emissions_by_country, 0.75)
iqr = q3 - q1

# Calculate lower and upper cutoffs for outliers
lower = q1 - 1.5 * iqr
upper = q3 + 1.5 * iqr

# Subset emissions_by_country to find outliers
outliers = emissions_by_country[(emissions_by_country < lower) |
(emissions_by_country > upper)]
print(f"Lower cutoff for outliers: {lower}")
print(f"Upper cutoff for outliers: {upper}")
print("Countries with outlier CO2 emissions:")
print(outliers)
```

Answer Explanation

1. **Subsetting for Outliers:** The `emissions_by_country` DataFrame is filtered using the calculated `lower` and `upper` bounds to include only countries with total CO2 emissions outside these thresholds.

2. **Lower and Upper Cutoffs:** These cutoffs, derived from Q1, Q3, and IQR, define the range of "normal" emissions. Countries below the lower cutoff or above the upper cutoff are considered outliers.

3. **Output Analysis:** The printed subset includes the list of countries with

outlier CO2 emissions. These outliers may represent exceptional cases, errors, or important patterns in the data.