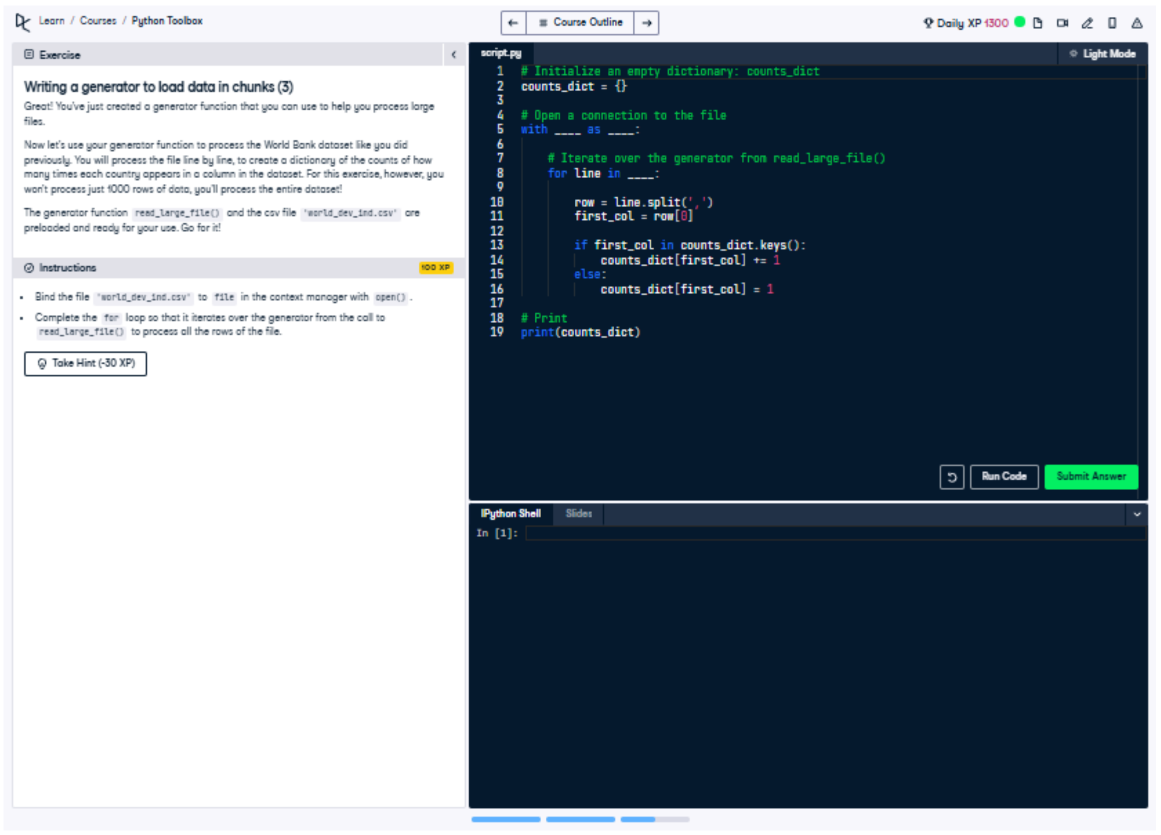

Learn / Courses / Python Toolbox

Exercise

Writing a generator to load data in chunks (3)

Great! You've just created a generator function that you can use to help you process large files.

Now let's use your generator function to process the World Bank dataset like you did previously. You will process the file line by line, to create a dictionary of the counts of how many times each country appears in a column in the dataset. For this exercise, however, you won't process just 1000 rows of data, you'll process the entire dataset!

The generator function `read_large_file()` and the csv file `'world_dev_ind.csv'` are preloaded and ready for your use. Go for it!

Instructions 100 XP

- Bind the file `'world_dev_ind.csv'` to `file` in the context manager with `open()`.
- Complete the `for` loop so that it iterates over the generator from the call to `read_large_file()` to process all the rows of the file.

script.py

```
# Initialize an empty dictionary: counts_dict
counts_dict = {}

# Open a connection to the file
with ____ as ____:

    # Iterate over the generator from read_large_file()
    for line in ____:

        row = line.split(',')
        first_col = row[0]

        if first_col in counts_dict.keys():
            counts_dict[first_col] += 1
        else:
            counts_dict[first_col] = 1

# Print
print(counts_dict)
```

```
In [1]:
```

Question: Writing a generator to load data in chunks (3)

Correct Answer and Explanation:

Code Implementation:

```
# Initialize an empty dictionary: counts_dict
counts_dict = {}

# Open a connection to the file
with open('world_dev_ind.csv') as file:

    # Iterate over the generator from read_large_file()
    for line in read_large_file(file):
        # Split the current line into a list: row
        row = line.split(',')

        # Get the value for the first column: first_col
        first_col = row[0]

        # If the column value is in the dict, increment its value
        if first_col in counts_dict.keys():
            counts_dict[first_col] += 1
        else:
            # Else, add to the dict and set value to 1
            counts_dict[first_col] = 1

# Print the resulting dictionary
print(counts_dict)
```

Explanation:
1. `counts_dict = {}`:
   - Initializes an empty dictionary to store counts of the first column values.

2. `with open('world_dev_ind.csv') as file:`:
   - Opens the file `world_dev_ind.csv` for reading.

3. `for line in read_large_file(file):`:
   - Iterates over each line of the file using the generator function `read_large_file()`.

4. `row = line.split(',')`:
   - Splits the current line into a list of values using a comma as the delimiter.

5. `first_col = row[0]`:
   - Extracts the first value of the row, corresponding to the first column.

6. `if first_col in counts_dict.keys():`:
   - Checks if the value of `first_col` is already in the dictionary. If true, increments its count.

7. `else:`:
   - If the value is not present in the dictionary, adds it with an initial count of 1.

8. `print(counts_dict)`:
   - Prints the resulting dictionary containing counts of values in the first column.