Data Science with Python – Assignment #2

Assignment description

This assignment deals with data exploration and manipulation using the pandas library and some basic statistical analysis. The assignment includes two python jupyter notebooks, with multiple questions each, where each question is written under the "Question X" cell and should be followed by a solution cell named "Question X – solution".

Each solution (your computation and answer to the question) should be printed in the form of a DataFrame (a table), with meaningful phrases, as shown below. Make sure you follow this format throughout the assignment. Imagine that you present the results of your analysis to a data science team at work. Example question and the required answer format:

```
### how many purchases (rows in the table, not 'Quantity') were done by Male and Female customers?
```

Question 1 - solution: Example

```
m_amount = len(df[df['Gender'] == 'Male'])
f_amount = len(df[df['Gender'] == 'Female'])

# do not print any intermediate results
# print *only your final* result in the form of a data frame (for clarity)

result = [('purchases by male', m_amount), ('purchases by female', f_amount)]
pd.DataFrame(result)
```

	0	1
0	purchases by male	498
1	purchases by female	502

Wrong answer format:

```
m_amount = len(df[df['Gender'] == 'Male'])
f_amount = len(df[df['Gender'] == 'Female'])

# do not print any intermediate results
# print *only your final* result in the form of a data frame (for clarity)

m_amount, f_amount
```

```
(498, 502)
```

Comments:

(1) Note that there are multiple (valid) ways to solve the same question.

(2) Before solving the assignment, take few minutes to understand the data – make sure you know what each column represents, and what are the relations between the datasets (where exist).

(3) Although the assignment can be solved almost entirely using tools we learned, you can (and are encouraged) to use pandas functions not shown at the lecture, as long as you understand the functions' meaning and documentation.

(4) Before submission: (1) restart kernel (right click → restart kernel) and make sure your solution works as expected; (2) clear all outputs (right click → clear all outputs).

Submission

Submit a single zip file – assignment2_ xxxxxxxxx_xxxxxxxxx.zip , where "xxxxxxxxx" stands for a student id. Please specify two student ids (your and your partner's). It should include two files:

1. Your solution for notebook #1: `assignment2_task1.ipynb`
2. Your solution for notebook #2: `assignment2_task2.ipynb`

Grading criteria include: correctness, code design and effectiveness, readability and documentation.

Good Luck!