CS 301 HW 03

Total possible score: 20 points

NOTE: Your submission must be in a .ipynb (Jupyter Notebook) file, **and** you must upload it to your GitHub account, **and** it must include the following comments in the first cell:

```
# Author: Firstname (Middlename) Lastname
# Date: YYYYMMDD
# CS301-006, Professor Watson
# HW## Solution
# Brief description of the assignment / project / work that is done
# link to the git repo (ex: https://github.com/cww5/web_scraper)
# link-to-the-relevant-git-commit (ex: https://github.com/cww5/web_scraper/tree/60e907c00ecbe21bc6a543621c9683ebf7f9693a)
# name-of-the-branch (ex: master)
```

Problem 1) 5 test cases, 1 point for each passed test case. (5 points total)
Write a function called *data_reader* which takes as input:

- data_path (string) - a path to a file in your file explorer

The function *data_reader* should **try** to return a DataFrame containing the data in the file. You may assume that the input file is a .csv file. If the file was read successfully, print the shape of the DataFrame and return the DataFrame, **except** when there is an error, in which you should return None and print a message to the user saying there was an error when reading the file. You will need Exception handling. Using the drinks.csv file, here is an example:

```
my = data_reader('C:\\Users\\watson\\Documents\\CS301_Planning\\drinks.csv')
```

```
The shape is: (193, 6)
```

```
my = data_reader('i_dont_exist.csv')
```

```
Error during reading, please try again.
```

Problem 2) 5 test cases, 1 point for each passed test case. (5 points total)
Write a function called *get_num_outliers* which takes as input:

- df (DataFrame) - a DataFrame object
- col (string) - the name of a column in the DataFrame

The function *get_num_outliers* should return the total number of outliers in that column of the DataFrame. You can assume that the value entered in the *col* parameter exists as one of the columns in the DataFrame (meaning, you do not need to do any exception handling here). Using the drinks.csv file, here is an example:

```
get_num_outliers(my, 'wine_servings')
```

```
26
```

Problem 3) 5 test cases, 1 point for each passed test case. (5 points total)
Write a function called *get_num_deviants* which takes as input:

- df (DataFrame) - a DataFrame object
- col (string) - the name of a column in the DataFrame
- multiplier (string) - the number of standard deviations away from the mean

The function *get_num_deviants* should return the total number of values in that column of the DataFrame which exist past *multiplier* deviants away from the mean. You can assume that the value entered in the col parameter exists as one of the columns in the DataFrame (meaning, you do not need to do any exception handling here). Using the drinks.csv file, here is are two examples:

```
get_num_deviants(my, 'wine_servings', 1)
```

```
28
```

```
get_num_deviants(my, 'wine_servings', 2)
```

```
15
```

Problem 4) 5 points for completing the survey
Please complete the weekly survey available here.
This is the link: https://forms.gle/sS76YiGEnxWAoCMP9

This survey is for you to submit your opinions about the topics covered each week. It is also feedback that I will use to improve the course moving forward.
NOTE: Remember the topics covered this week were lectures 2,3.