Homework 5 – Data Wrangling and Analysis

In this homework, you will demonstrate greater proficiency in data processing and analysis in Python.

Notes: This homework purposefully contains less explicit instructions compared to previous homework assignments to allow you to demonstrate greater proficiency in the course material. You may only import the following libraries to complete the work for this homework: `numpy`, `pandas`, and, `re`.

Socially Responsible ETF Analysis

You are working at a financial services firm as a data analyst. They are interested in creating a new investment fund with an eye towards social responsibility. You have been asked to do some research to make recommendations on what they should pursue. To do this, you find a website that contains socially responsible exchange traded funds (ETFs): https://www.etf.com/channels/socially-responsible. To begin, you decide to scrape the data into an Excel workbook to share with executives, but now you need to perform some data analysis. You can review the website for additional context around the data set.

Create a Jupyter Notebook that performs the following:

1. Consider the data within the Excel spreadsheet:
    a. **(1 point)** Load the data from the file into a data frame.
    b. **(1 point)** Convert any missing values into NaN.
    c. **(1 point)** Print a data frame containing the first 20 funds, including all data known about them.

Processing

2. **(10 points)** You want to analyze the assets under management (AUM); however, after reviewing the data you notice the units are not the same for each value. Determine a method to address this issue and then modify the data frame.
3. **(10 points)** The Segment variable contains more than just the segment. It contains the asset class (e.g., equity, fixed income, etc.), market (e.g., U.S., global, emerging markets, etc.), and the segment itself (e.g., Total Market, Mid Cap Growth, etc.). Process this data into three new columns in the data frame.
4. **(5 points)** Modify the data frame such that it is indexed in multiple levels - first by the issuer and then by the ticker. Sort the data frame first by issuer and then by ticker. Display the first fifty records.
5. **(2 points)** Create an observation: What kinds of social issues are supported by the various ETFs? How might you group them together to provide categories of social issues?

Analysis

6. **(5 points)** To determine where you might want to consider research, determine how many socially responsible ETFs are issued by each issuer? From this analysis, which issuer(s) do you recommend studying further? Why?

7. **(5 points)** Review the total assets under management for each issuer. From this analysis, which issuer(s) do you recommend studying further? Why?
8. **(5 points)** Review the assets under management for each fund. From this analysis, which issuer(s) do you recommend studying further? Why?
9. **(5 points)** An expense ratio describes how much of a fund's assets are used for operating expenses, which will likely get passed onto investors; thus, a lower expense ratio is better as an investment. Analyze the expense ratios for the funds of each issuer. From this analysis, which issuer(s) do you recommend studying further? Why?
10. **(10 points)** Analyze frequency distributions of the asset classes, markets, and segments for the funds of each issuer. From this analysis, make a recommendation on the type of fund your company should pursue, factoring in an asset class, market, and segment. What do you recommend? Why?
11. **(5 points)** Another option to determine what the company should pursue is to look at the 3-month total return percentages. Determine the mean 3-month total return percentages factoring in a combination of asset class, market, and segment together. From this analysis, make a recommendation on the type of fund your company should pursue, factoring in an asset class, market, and segment. What do you recommend? Why?
12. **(5 points)** A senior executive at your company has said holding socially responsible investments is not profitable compared to holding investments that do not address socially responsible causes. Incorporate the analysis you have done in this assignment (and/or additional analysis if you would like) with a small amount of Internet research on current investments to prove or disprove the executive.

(5 points) Insert a cell at the top of your notebook, using markdown, filled out with your information:

```
Name: <Your Name>
Course: BUDT704
Section: <Your Section Number (make sure you have the right number)>
Date: <Date of Submission>
```

Insert a cell at the bottom of your notebook, using markdown, with the following statement. If you do not wish to include the pledge, include an explanation why in the cell. Make sure you include your name.

```
"I pledge on my honor that I have not given nor received any unauthorized
assistance on this assignment."
--<<include your name>>
```

Note that points are attached to this task. Failure to properly include the requested elements as prescribed will count as missing a business requirement.

(10 points) Your solution must include appropriate commenting describing the purpose of your code. Remember that you must include why you are writing the code and not just reiterate what the code is doing.

(10 points) Format the Jupyter Notebook using an appropriate level of markdown cells so that your analysis is fully explained at each step of what you have done. Some examples of formatted notebooks are as follows:

- https://anaconda.org/jbednar/plotting_pitfalls/notebook
- https://nbviewer.jupyter.org/github/jrjohansson/scientific-python-lectures/blob/master/Lecture-4-Matplotlib.ipynb

Note: These are just examples and not prescriptive for what your Notebook must look like. **Simply copying the instructions for your markdown cells is not appropriate**. Markdown cells must be your approach and analysis.

(5 points) Export your Jupyter Notebook as an .html file. Save the file, properly named with a file name of HW5_YourLastName_YourFirstName.html. Example: HW5_Bono_John.html. Note that points are attached to this task. Failure to properly name the file as prescribed will count as missing a business requirement.

Submission Files

Your submission to ELMS must contain **TWO** files: (to be eligible for full points)

- The .ipynb file (Jupyter Notebook) you created
- The .html file you exported from your Jupyter Notebook

WARNING! You must submit a runnable Jupyter Notebook file. Failure to submit a runnable Jupyter Notebook file will result in a grade of 0. There will be no exceptions, even for a first-time mistake. A non-running file cannot be graded. Please **triple check** you have submitted a runnable Jupyter Notebook to avoid issues.

Additional Assignment Requirements

- You may use any Python functions we have covered them in class, are in the assigned reading, or are present in one of the Python libraries permitted above. Importing any other library will result in a substantial grade deduction.
- Your solution may not use any functions or language constructs not covered during this semester's BUDT704 without prior authorization from me, even if you know other functions or language constructs. **Doing this may lead to a substantial grade penalty, a grade of zero, or a Code of Academic Integrity inquiry.** When in doubt, ask! Note: This is not to stifle your creativity, but rather to level the playing field across all students. If there is something you want to do, initiate a discussion with me first.
- Review the Syllabus guidelines about academic integrity. Your work will be checked using manual and automated means. If you have questions or concerns about academic integrity, ask before accidentally or intentionally putting yourself in a compromised situation. I want to help you avoid issues!
- Keep in mind the various best practices we have covered in class. Your solution must incorporate these best practices and avoid any programming/design pitfalls we discussed.