

Advanced Python

Prof. Groner

Assignment 1 – S&P500 Review with pandas DataFrame

THE ASSIGNMENT

The file `SP500_Constituents_20230101.csv` has data for the S&P 500 securities, as of 2023-01-01. Use Jupyter Notebook or Google Colab to answer the following questions.

- 1) Using `read_csv()`, load the `SP500_Constituents_20230101.csv` file into a `DataFrame`. Use the data in the first column (Symbol) as the index of the data.
- 2) Display the first row of the data.
- 3) Determine how many rows are in the data.
- 4) Which columns have missing values?
- 5) Use the `loc[]` method to display the information about Apple (symbol=AAPL).
- 6) What are the distinct values within the column 'GICS Sector'?
- 7) How many stocks are there in each of the GICS Sectors?
- 8) Use a filter expression to display the rows where the stock is in the Energy sector.
- 9) Use `sort_values()` to order the data for Energy stocks, sorting based on the 'Date first added'. What company was added first, what company was added most recently?
- 10) For the full set of stocks, using `isna()` and column selection, display the Symbol, Security, and Founded for those stocks that are missing data for Date first added.

SUBMISSION DIRECTIONS

Use Jupyter Notebook or Google Colab to answer the above questions. Submit one file named like:

<Lastname><Firstname>Asn1.ipynb

for example:

SmithJohnAsn1.ipynb

The first cell of your notebook should have a comment like:

John Smith – Assignment 1

The first line of each cell of your notebook should have a comment.