Stigler Center: 2022-23 Research Professional Task

George J. Stigler Center for Study of the Economy and the State*

Fall, 2022

Instructions

Please complete the following tasks and submit the required materials by the designated deadline. Ensure that you follow the directions carefully and submit all requested materials.

General

Both the code and the results that you submit should be legible to your peers and those who would be your supervisors. That is, ensuring that your code is well-documented and well-commented; and that your written reports and data visualizations clearly convey your analysis is of the utmost importance.

Additionally, writing manageable and reproducible code is critical for research. For each task, please write a `README` that describes each of the files that you have created/used and what each does. This should be written so that a user with a technical background but with no knowledge of the task itself can understand your code and reproduce all of the output products with minimal (if any!) changes to the code itself.

When submitting your materials please submit everything as a `.zip` file. The name and structure of that compressed file should be as follows:

```
[firstname]_[lastname].zip
+-- task1a
|       +-- [firstname]_[lastname]_task1a_README.md
|       +-- [firstname]_[lastname]_task1a.csv
|       +-- [firstname]_[lastname]_task1a.[py/ipynb]
+-- task1b
|       +-- [firstname]_[lastname]_task1B_README.text
|       +-- [firstname]_[lastname]_task1b.csv
|       +-- [firstname]_[lastname]_task1b.[py/ipynb]
+-- task2
        +-- [firstname]_[lastname]_task2.[py/ipynb]
```

*asdf

If examples are given, follow them closely. Good luck and have fun!

Disclaimers

The contents of this coding activity are **confidential**. Do NOT share this document or any of the attached material with anyone. Sharing of these instructions or any associated materials will result in your immediate disqualification and may preclude you from future opportunities with the University of Chicago.

1 Python and Web Scraping

1A *American Economic Review*:

Using Python, collect publication data from the *American Economic Review*. Using Python and your choice of packages (for example BeautifulSoup, Selenium, or re) and Pandas, collect the following information for all available issues:

- Volume and issue number;
- a hyperlink; and
- issue-date information.

Store the collected issue-level data in a `.csv` file. Your resultant spreadsheet should be structured follows:

Table 1: Example Task 1A Output Structure

volume	link	issue_date
Volume 112	https://www.aeaweb.org/issues/692	September 2022 (Vol. 112, No.9)
Volume 112	https://www.aeaweb.org/issues/689	August 2022 (Vol. 112, No.8)
Volume 112	https://www.aeaweb.org/issues/685	July 2022 (Vol. 112, No.7)
...	...	...

Inside a folder named `task1a` save your README (use any widely used text format you desire, such as markdown: `.md`), code, and spreadsheet as follows:

- `task1a_[firstname]_[lastname]_README.md`
- `task1a_[firstname]_[lastname].csv`
- `task1a_[firstname]_[lastname].[py/ipynb]`

1B *American Economic Review*: Redux

Note that each article published by the *American Economic Review* is labeled with JEL codes which describe to which field(s) and subfield(s) the article contributes. Each alphanumeric JEL code is associated with a human-readable description. Using the same packages as above, construct a *second* spreadsheet that contains the following bibliographic information:

- volume and issue number;
- issue-date information;
- the article's title;
- the article's author(s);
- the article's print page-numbers;
- a *permanent link* to the article;
- the article's JEL code(s); and
- the article's JEL code description(s).

Store the collected issue-JEL-code-level data in a `.csv` file. Your resultant spreadsheet should be structured as follows:
Inside a folder named `task1b` save your README (use any widely used text format you desire, such as plaintext: `.txt`), code, and spreadsheet as follows:

- `task1b_[firstname]_[lastname]_README.txt`
- `task1b_[firstname]_[lastname].csv`
- `task1b_[firstname]_[lastname].[py/ipynb]`

Table 2: Example Task 1B Output Structure

volume	issue_date	article_title	authors	page_numbers	article_link	jel_code	jel_description
Volume 112	September 2022 (Vol. 112, No.9)	Belief Elicitation and Behavioral Incentive Compatibility	David Danz; Lise Vesterlund; Alistair J. Wilson	pp. 2851-83	https://doi.org/10.1257/aer.20201248	D83	Search; Learning; Information and Knowledge; Communication; Belief; Unawareness
Volume 112	September 2022 (Vol. 112, No.9)	Belief Elicitation and Behavioral Incentive Compatibility	David Danz; Lise Vesterlund; Alistair J. Wilson	pp. 2851-83	https://doi.org/10.1257/aer.20201248	D91	Micro-Based Behavioral Economics: Role and Effects of Psychological, Emotional, Social, and Cognitive Factors on Decision Making
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	D22	Firm Behavior: Empirical Analysis
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	G31	Capital Budgeting; Fixed Investment and Inventory Studies; Capacity
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	G35	Payout Policy
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	H25	Business Taxes and Subsidies including sales and value-added (VAT)
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	H32	Fiscal Policies and Behavior of Economic Agents: Firm
...	...	...	...	...	...	...	...