

# 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.

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.

---

\*asdf

# 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;
- issue-date information.
- a hyperlink; and

Store the collected issue-level data in a `.csv` file. Your resultant spreadsheet should look as follows: *Inside* a folder named `task1a` save your README (use any widely use text format your desire), code, and spreadsheet as follows:

- `task1a_[firstname]_[lastname]_README`
- `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;
- the article's print page-numbers;
- issue-date information;
- a *permanent link* to the article;
- the article's title;
- the article's JEL code(s); and
- the article's author(s);
- the article's JEL code description(s).