Assignment 3: Data Privacy

In this assignment we will work on collecting data through web scraping, how to identify individuals from improperly anonymized datasets, and how to properly anonymize data using the Safe Harbour method. Note that the datasets provided are synthetic data.

Tasks / Learning Goals

- Explore collecting data through web scraping
- Explore how individuals can be identified from improperly anonymized datasets using small amount of external data, and simple matching procedures
- Gain experience with properly anonymizing datasets using the Safe Harbour method

Due Date

11:59 pm on Sunday February 11th, submitted on TritonED.

Submitting Assignments

You will submit a Jupyter notebook file (.ipynb) to TritonED. Make sure that the file you submit has the following filename (filled in with your course ID number - first letter of your last name, followed by the last 4 numbers of your student ID number):

'A3_$####.ipynb'

Grading Rubric

This assignment is worth 12% of your grade (12 points).

There are 2 parts to this assignment, with the following point values:

Part 1: Web Scraping	2.5 points
Part 2: Identifying Individuals	4.5 points
Part 3: Anonymizing Data	5 points

Questions with detailed instructions are all available directly in the assignment notebook.

Associated Data Files

The following data files are provided to you, and should be used to complete the assignment:
- anon_user_dat.json
- employee_info.json
- user_dat.json
- zip_pop.csv