

CS 241: Project

Project Due: Monday, December 13th at 11 AM

1 Overview

Throughout CS 241: Data Structures we have discussed techniques used to create efficient and effective computer systems.

The project is a chance to apply this new knowledge regarding computing with respect to both technology and context, and present this knowledge to your peers for feedback. It will allow you to apply the knowledge learned in this course collaboratively and create a portfolio piece to demonstrate your knowledge to the broader society.

To do so, form a group of 1-2 students (or seek instructor approval) and organize around an existing data set. For example, you may consider the Federal Election Commission's data sets on individual political contributions from citizens listing Willamette University as their/his/her employer. You can see this data set at the following url:

https://www.fec.gov/data/receipts/individual-contributions/?contributor_employer=Willamette+University&two_year_transaction_period=2022&min_date=01%2F01%2F2021&max_date=12%2F31%2F2022

or see the downloaded csv file at

https://github.com/cd-public/xfer/blob/master/sample_dataset.csv.

2 Project Requirements

From the provided or sample data set, develop Python scripts to load the data set into internal Python data structures (insert), check for the presence of certain data (contains), and organize the data in some way (sort). For example, with political contributions, you may wish to add new contributions as made since downloading the initial data set, check for contributions to a certain PAC or candidate, or sort contributions by size.

For each of these operations, you may use data structures and algorithms presented in this class.

3 Project Submission

Your project submission should contain three parts, either in a Github repository (good to show potential future employers) or a zip file (good to email). If you need help with either, ask the instructor or your peers.

The first part should be your data set.

The second part should be your Python scripts, containing at least a notion of "insert", "contains", and "sort". This could be one or multiple Python scripts.

The third and final part should be a write up, in text, pdf, or .doc* format, describing

- How to use your Python scripts to understand the data set, such as by running "script.py" in the same folder.
- Why you chose to implement "insert", "contains", and "sort" the way you did.
- What you learned about the data set from this project. What did you expect? What surprised you?