

Mining Software Repositories Assignment 1 (MSR1)

Submit Assignment

Due Oct 16 by 11:59pm **Points** 15 **Submitting** a file upload **File Types** pdf and ipynb

This is the first of two mining repositories assignments that will teach you how to investigate data within real-world project repositories. In the assignment, we will focus on basic data collection from GitHub and how to answer direct data-driven questions from that mined data. The second mining repositories assignment will focus more on quantitative analysis and using graphical data representations to explore and present data-driven results.

This is an individual assignment, and each student is expected to develop their own methods for answering the questions. This type of analysis can be accomplished in multiple ways, so please try to develop an intuition for your chosen methodology.

Problems:

FILE: [MSR1.ipynb](#)

The Jupyter notebook file included above should be your starting point for this assignment. All questions outlined below are also included in the file. The assignment has two parts, each with their own targets from the GitHub API:

Part 1: discourse/discourse repo

For this part, we will investigate the [discourse/discourse](https://github.com/discourse/discourse) project.

- Question 1: What is the total number of unique contributors for this project? (Contributions include commits and pull requests)
- Question 2: Which user made the most contributions to the project?
- Question 3: Which user made the most commits to the [discourse/app/models/badge.rb](https://github.com/discourse/discourse/blob/master/app/models/badge.rb) file?

Part 2: Dan Abramov (gaearon)

For this part, we will investigate [Dan Abramov](https://github.com/gaearon) (prominent developer of [React](https://reactjs.org/), co-author of [Redux](https://redux.js.org/) and [Create React App](https://create-react-app.dev/)).

- Question 1: Which of Dan's projects did he commit to most often in the past three years? (From June 1, 2017 to June 1, 2020)
- Question 2: The [React](https://reactjs.org/) project was founded in 2013. When did Dan first make a release in the [facebook/react](https://github.com/facebook/react) project? And what was the version number of that release?

What to hand in:

Report (filename: `[onid-username]-msr1.pdf`)

- Detail your methodology and results for each question, and include citations for any references/resources used.
- Write in an academic tone similar to the research papers assigned in the course. Be precise and write only as much as needed to convey your results.
- Any format is allowed, but if possible follow the [ACM](https://www.acm.org/publications/proceedings-template) or [IEEE](https://www.ieee.org/conferences/publishing/templates.html) formats since it will provide you with practice.
- Do not simply include code to demonstrate your methods (those implementation details belong in the notebook file).

Jupyter Notebook (filename: `[onid-username]-msr1.ipynb`)

- Include all data extraction and analysis code used to answer each question.
- Scrub any personal or confidential information (e.g. authentication tokens) from the file.
- Do not hard-code data into your notebook, but instead include your commands for interacting with the GitHub API.