

Git-OSS-um - A Tool for Newcomers to Open Source

Stephen White, Gary Baker, Van Steinbrenner

School of Informatics, Computing, and Cyber Systems

Motivation

Dr. Igor Steinmacher is an Assistant Professor at NAU. His focus is aiding newcomers in making contributions to open source software projects.

Newcomers face **3 major obstacles** when trying to contribute:

1. **Information overload**
2. **Inability to gauge repository activity**
3. **General discouragement**



With **1.1 billion open source contributions per year**, a 10% decline would greatly affect the wealth of innovation the world currently experiences.

Solution Overview

A web application that helps a newcomer to Open Source find a project by displaying easy to read graphics, gauging the activity of a repository, and to show the acceptance rate of submitted contributions.

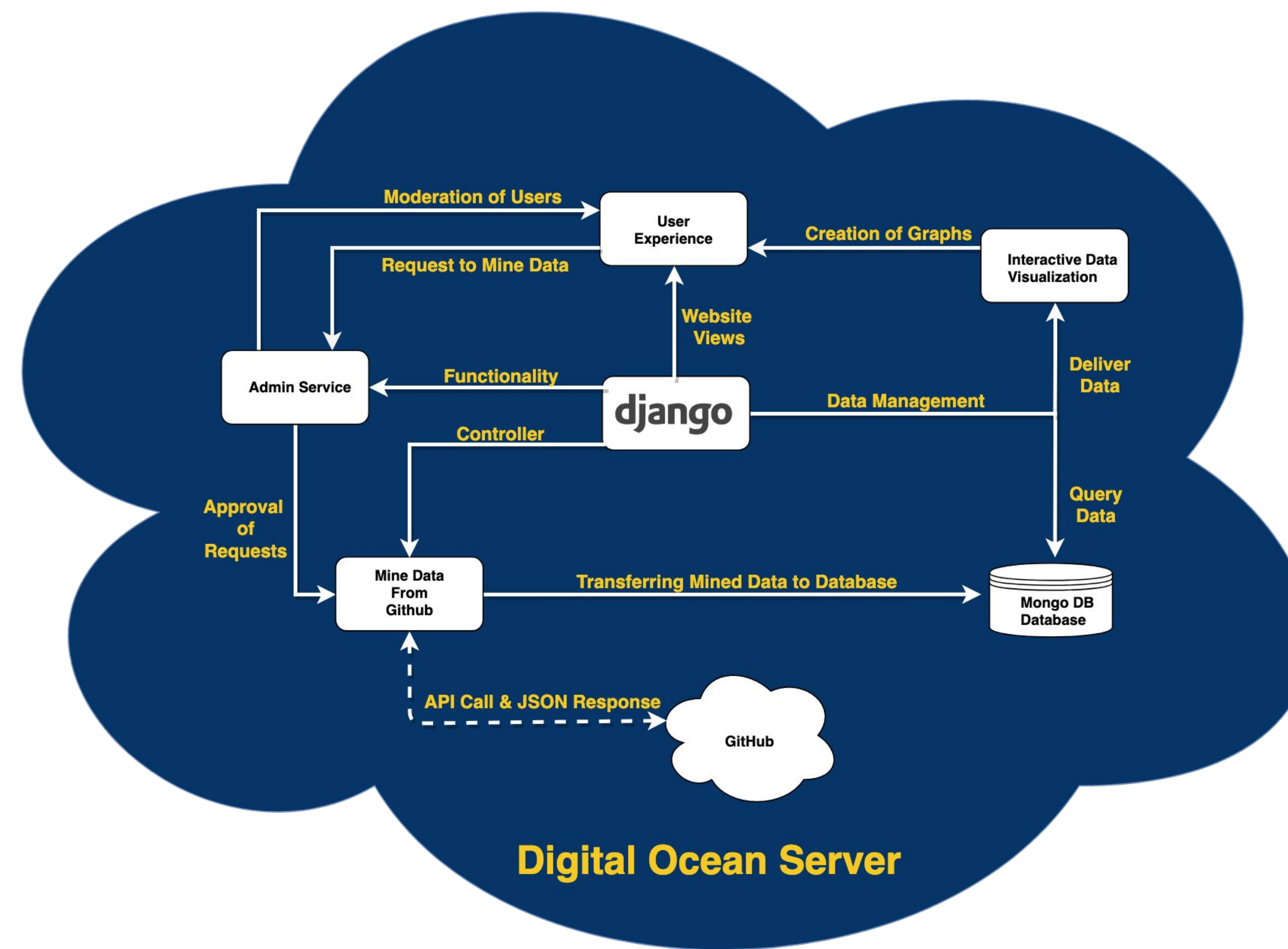
Mined Repos

The screenshot shows a list of mined repositories on the left, including 'torvalds/linux', 'google/jumbo-parser', 'microsoft/calculator', 'jabref/fabref', 'qiime2/qiime2', 'blockchain/blockchain-wallet-v4-frontend', 'tidyverse/dplyr', 'trekhleb/javascript-algorithms', 'tuvtran/project-based-learning', 'valloric/youcomplete', 'kriadmin/30-seconds-of-python-code', 'django/django', 'rails/rails', and 'elixir-lang/elixir'. On the right, there is a search filter sidebar with options for languages (C, C++, Elixir, HTML, Java, JavaScript, Python, R, Ruby), min/max pull requests, and whether to include repositories with a wiki page.

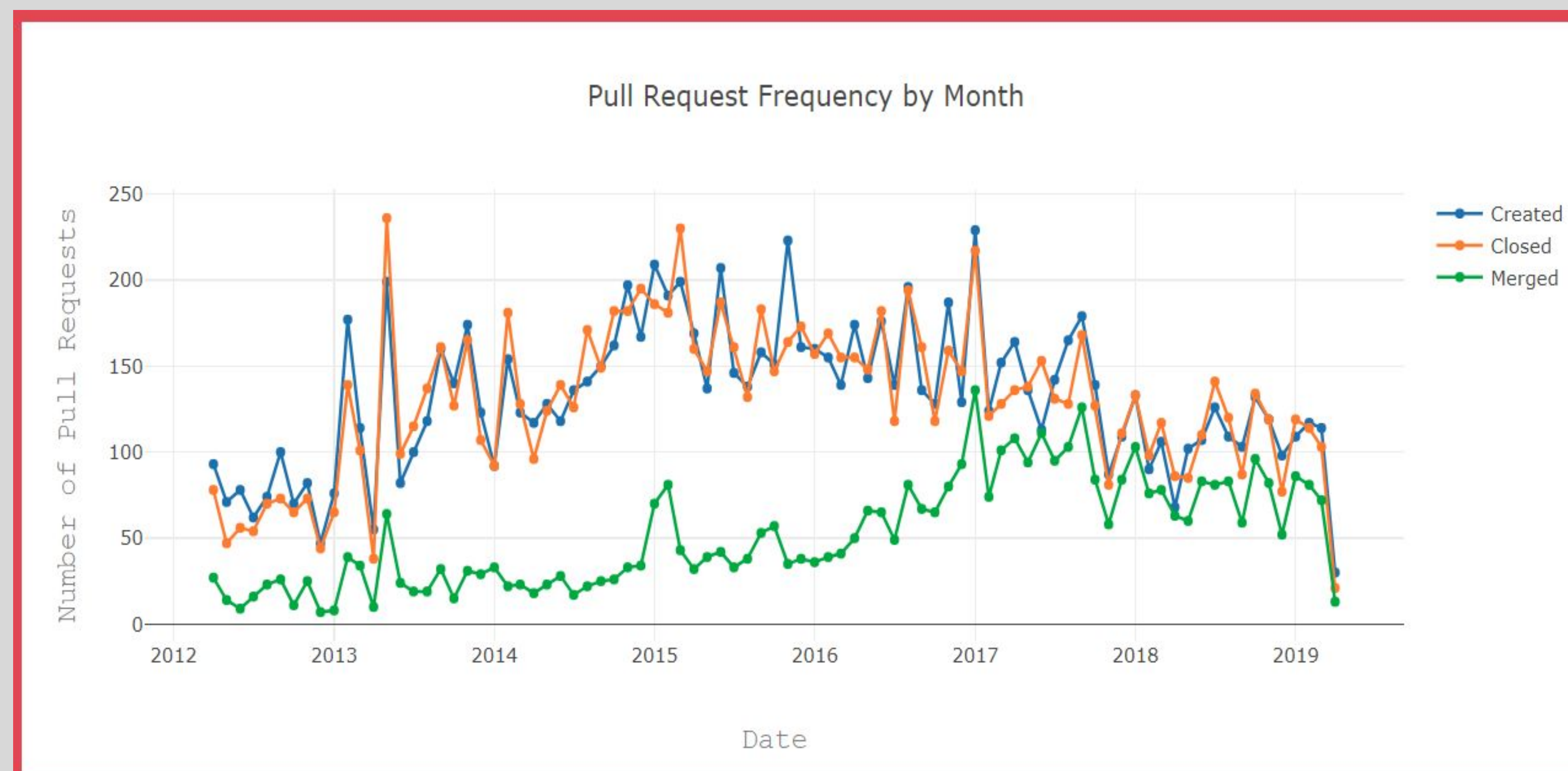
Testing

- **100 Unit Tests** comprising **7 Test Suites**
- **Integration Tests** verifying the integrity of information presented to our users on the application
- Observing the **Usability** of our software through an observation study

Architecture



- Git-OSS-um revolves around the **Django Web Framework**.
- From Django, our **Python** mining script is called to retrieve GitHub information and store it in our **MongoDB** database.
- Interactive graphics are generated using the open source library **Plotly** with data we have gathered and stored from **GitHub**.



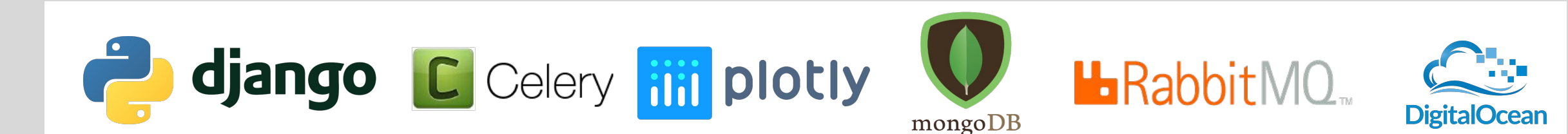
Outcomes

We have given power back to the newcomer by providing them with a safe place to explore interesting repositories, and ease them into the process of contributing to an open source project.

Challenges

1. **Mine Data in Parallel:** Mine thousands of pull requests efficiently
2. **Filtering Data:** Use set theory to apply many filters at once.
3. **Scheduling Future Tasks:** Learn a completely new technology
4. **Visualizing Data:** Apply data science obtain correct information
5. **Optimizing Response Time:** Generate interactive graphics quickly

Technologies

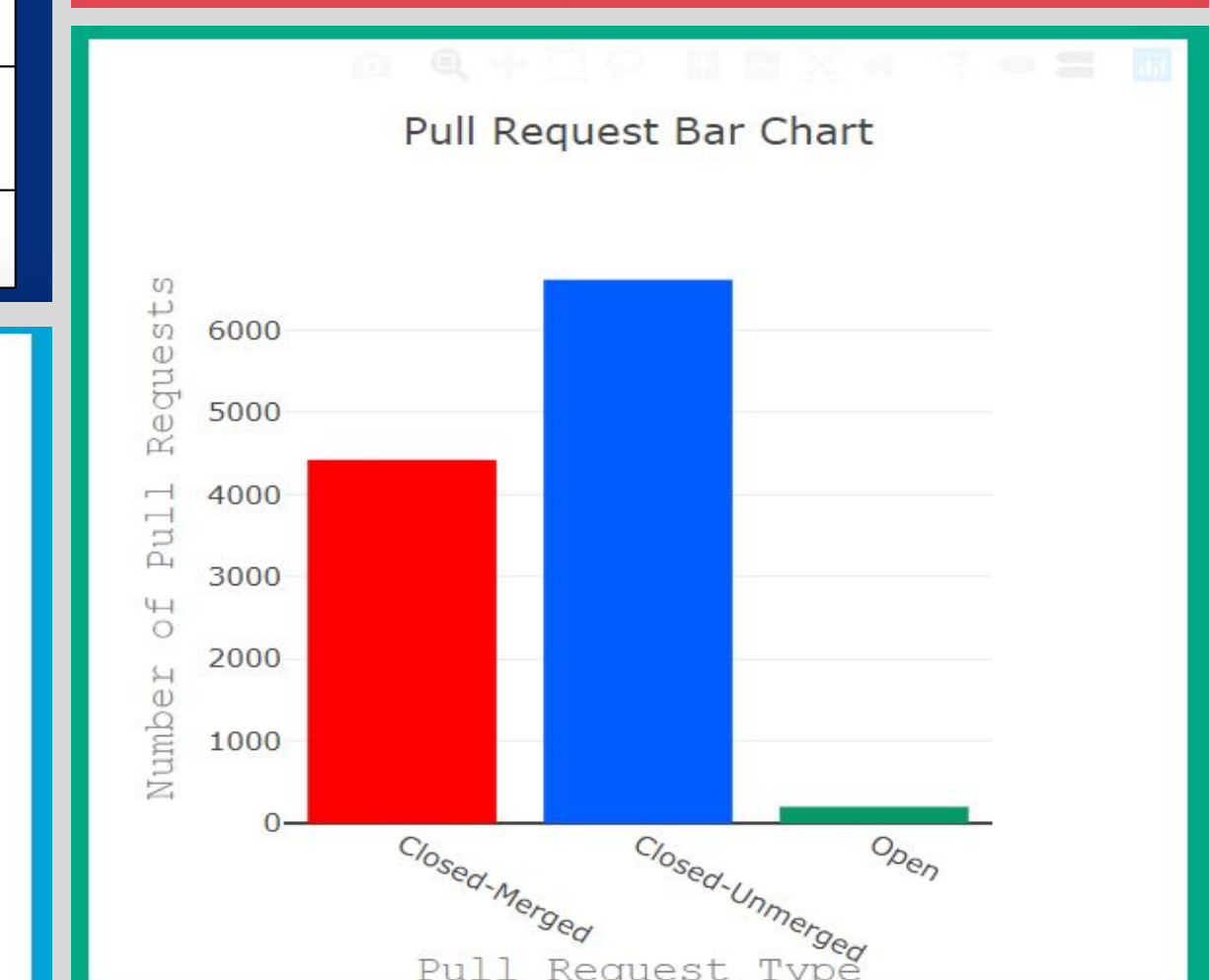
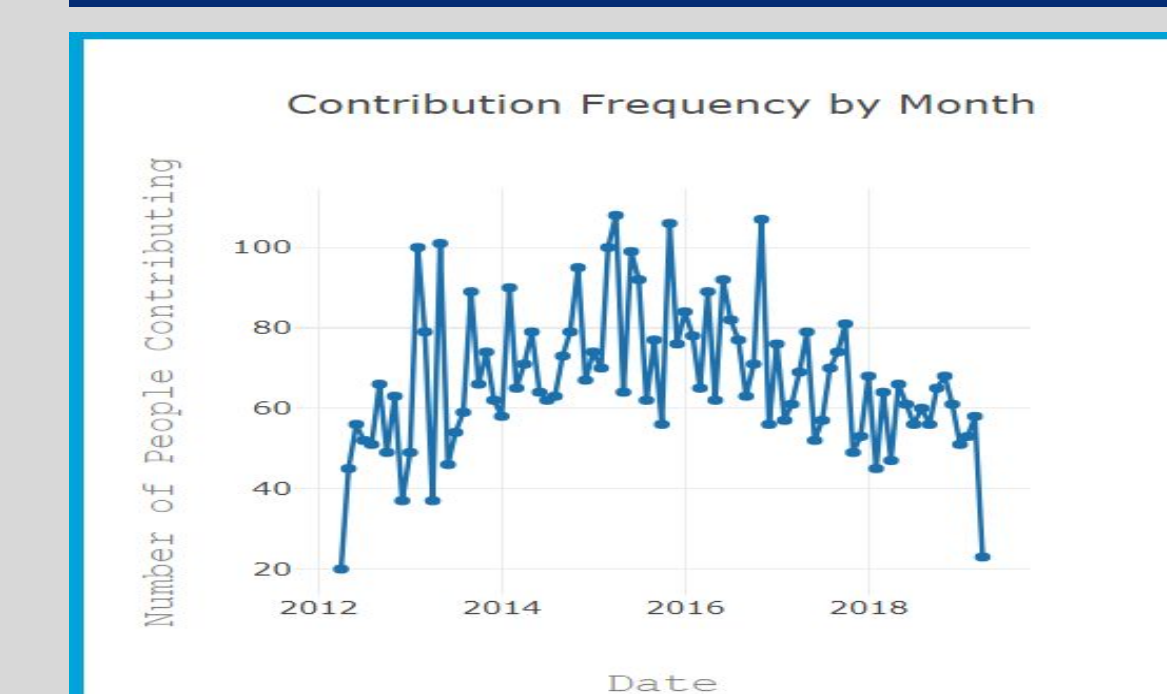


Key Features

- **Filter System:** Filter repositories based on things like Language, number of pull requests, and search queries.
- **Comparison System:** Users can compare up to three repositories side by side.
- **Request-Based Information Gathering:** Users can submit a form to request a specific repository and get updated if their request was accepted or declined.
- **Intuitive Graphics:** Simple, easy to understand graphics that assist the user in grasping information regarding a repository.

Comparison Overview			
Item	torvalds/linux	jabref/fabref	django/django
Date Created	Sept. 4, 2011, 3:48 p.m.	March 11, 2014, 7:48 a.m.	April 27, 2012, 7:47 p.m.
Date Last Updated	April 17, 2019, 9 a.m.	April 17, 2019, 6:08 a.m.	April 17, 2019, 9:12 a.m.
Date Last Mined	April 17, 2019, 9:14 a.m.	April 17, 2019, 9:16 a.m.	April 17, 2019, 9:27 a.m.
Clone URL	https://github.com/torvalds/linux.git	https://github.com/jabref/fabref.git	https://github.com/django/django.git
Repository Homepage	None	https://www.jabref.org	https://www.djangoproject.com/
Number of Stargazers	73616	1281	40917
Programming Language	C	Java	Python
Has a Wiki Page	False	True	False
Development License	Other	MIT License	Other
Number of Open Issues	262	244	223
Number of Forks	26789	901	17622
Number of Watchers	6707	94	2164
Total Pull Requests	640	2730	11233
Number of Open Pull Requests	257	21	201
Number of Accepted Pull Requests	10	2462	4421
Number of Denied Pull Requests	372	267	6811

Item	Content
Date Created	April 27, 2012, 7:47 p.m.
Date Last Updated	April 17, 2019, 9:12 a.m.
Date Last Mined	April 17, 2019, 9:27 a.m.
Clone URL	https://github.com/django/django.git
Repository Homepage	https://www.djangoproject.com/
Number of Stargazers	40917
Programming Language	Python
Has a Wiki Page	False
Development License	Other
Number of Open Issues	223
Number of Forks	17622
Number of Watchers	2164



Future Work

- Sentiment analysis of pull request comments
- Analysis of commit patterns
- User-base tracking
- Repository activity projection using machine learning