

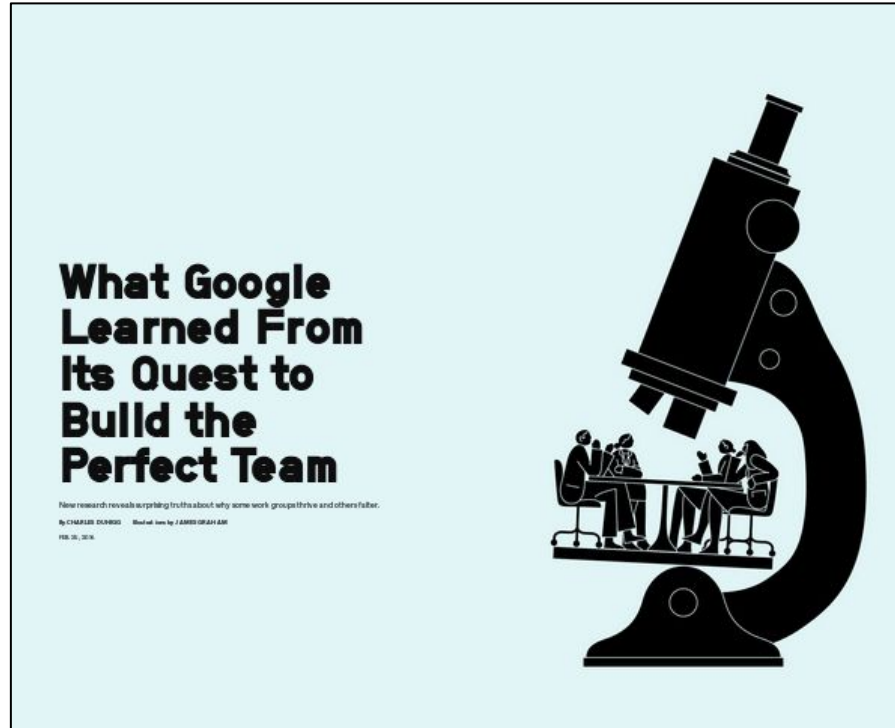


Case Study Prelude

Objectives

1. Congratulations on sticking with it (really).
2. What team characteristics lead to success?
3. How do you collaborate, as a team, with Git and Github?

What teams succeed



<https://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html>

“Who” didn’t seem to matter

‘We had lots of data, but there was nothing showing that a mix of specific personality types or skills or backgrounds made any difference. The “who” part of the equation didn’t seem to matter.’

Goal: Collective intelligence

'As long as everyone got a chance to talk, the team did well. But If only one person or a small group spoke all the time, the collective intelligence declined.'

Goal: Collective intelligence

'As long as everyone got a chance to talk, the team did well. But If only one person or a small group spoke all the time, the collective Intelligence declined.'

Please take the time to listen to each other. You'll learn something, your team will likely do better, and you'll like each other more.

Collaborating on Github


Goal: Team members will work on their own parts of the project and then combine all these parts into one repo on Github. At the end of the project, all team members will have an up-to-date fork of the case-study repo.




Case study high-level workflow suggestion

- 1) One team member should fork the case study. This will be called the upstream repo.
- 2) All other team members should fork the upstream repo.
- 3) Everyone clones their own forked repos down to their local machines.
- 4) On your local machine, create and checkout a branch to work on. **No one works on the master branch, even the upstream owner.**
- 5) Do your work.
- 6) Push your branch to your fork.
- 7) Issue a pull request to merge your fork with the upstream repo. The owner of the upstream repo will accept your pull request and merge it into the upstream master branch, then delete your branch.

In this process, everything will eventually be merged into the master branch in the upstream repo. This will be the “production” code that everyone will have a copy of in the end..

1. One person creates upstream

 **Frank-W-B / regression-case-study** Private
forked from zipfian/regression-case-study

 Watch ▾ 16  Star 0  Fork 223

[Code](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Pulse](#) [Graphs](#) [Settings](#)

Case study for regression (week 4). [Edit](#)

[Add topics](#)

16 commits


1 branch

0 releases









6 contributors


Branch: master ▾ [New pull request](#) [Create new file](#) [Upload files](#) [Find file](#) [Clone or download ▾](#)

This branch is 2 commits ahead of zipfian:master. [Pull request](#) [Compare](#)

 **Frank-W-B** Add fb solution files

Latest commit 7f39063 9 hours ago

 data	Add solution files	3 months ago
 images	minor changes	2 years ago
 soln_fb	Add fb solution files	9 hours ago
 src	Add solution files	3 months ago
 .gitignore	create .gitignore so we don't track Train.csv	6 months ago
 README.md	Updating Scary Wording about submission	5 months ago
 config.yaml	remove default channel/username in config.yaml	8 months ago
 score_model.py	Initial commit.	2 years ago


 **README.md**

1. Upstream owner adds collaborators


- Settings
- Collaborators & Teams
- Scroll down to bottom
- Add username(s) of collaborators
- Send invitation (permission level Write)
- Person invited needs to check their email associated with GH account to accept invitation.
- When they accept invitation will look like this in upstream repo:

The screenshot shows the GitHub repository settings for 'Frank-W-B / regression-case-study', which is a private repository forked from 'zipfian/regression-case-study'. The repository has 16 watchers, 0 stars, and 223 forks. The 'Settings' tab is selected, and the 'Collaborators & teams' section is active in the left sidebar. The 'Teams' section lists three teams: 'zipfian-instructors' (81 members, Admin permission), 'seattle-1' (Seattle Cohort 1 (July 2015), 13 members, Read permission), and 'cohort-9' (San Francisco - September 2015, 23 members, Read permission). The 'Collaborators' section below shows a message: 'This repository doesn't have any collaborators yet. Use the form below to add a collaborator.' The form includes a search bar with the text 'fwbtest71' and an 'Add collaborator' button. Below the main screenshot, a smaller inset shows the 'Collaborators' section with a single collaborator, 'fwbtest71', with a 'Write' permission level.


2. Collaborator forks upstream repo

 **fwbtest71** / **regression-case-study** Private


forked from Frank-W-B/regression-case-study

 Unwatch ▾

1


 Star

0


 Fork

224


<> Code


 Pull requests


0


 Projects

0

 Wiki

 Pulse

 Graphs

 Settings

Case study for regression (week 4).

Edit

Add topics

📄 16 commits

🌿 1 branch

📦 0 releases

👥 6 contributors

Branch: master ▾

New pull request


Create new file


Upload files


Find file

Clone or download ▾









This branch is even with Frank-W-B:master.


 Pull request

 Compare

 **Frank-W-B** Add fb solution files

Latest commit 7f39063 9 hours ago

 data	Add solution files	3 months ago
 images	minor changes	2 years ago
 soln_fb	Add fb solution files	9 hours ago
 src	Add solution files	3 months ago
 .gitignore	create .gitignore so we don't track Train.csv	6 months ago
 README.md	Updating Scary Wording about submission	5 months ago
 config.yaml	remove default channel/username in config.yaml	8 months ago
 score_model.py	Initial commit.	2 years ago

 [README.md](#)

11

3. Everyone clones their repo (and more)

All should `$ cd` into their cloned directory and type `$ git remote -v` to see the remote, called origin, associated with the Github repo they cloned the repo from. It should show origin and nothing else. **The collaborators need to add the upstream remote, too, so that they can pull down new files and versions of files when they are added to the master branch of the upstream repo.**

```
[mbp:~ frank.burkholder$  
[mbp:~ frank.burkholder$ git clone https://github.com/fwbtest71/regression-case-study.git  
Cloning into 'regression-case-study'...  
remote: Counting objects: 67, done.  
remote: Total 67 (delta 0), reused 0 (delta 0), pack-reused 67  
Unpacking objects: 100% (67/67), done.  
[mbp:~ frank.burkholder$ cd regression-case-study/  
✓ ~/regression-case-study [master {origin/master}] ✓  
[00:24 $ git remote -v  
origin https://github.com/fwbtest71/regression-case-study.git (fetch)  
origin https://github.com/fwbtest71/regression-case-study.git (push)  
✓ ~/regression-case-study [master {origin/master}] ✓  
[00:24 $ git remote add upstream https://github.com/Frank-W-B/regression-case-study.git  
✓ ~/regression-case-study [master {origin/master}] ✓  
[00:27 $ git remote -v  
origin https://github.com/fwbtest71/regression-case-study.git (fetch)  
origin https://github.com/fwbtest71/regression-case-study.git (push)  
upstream https://github.com/Frank-W-B/regression-case-study.git (fetch)  
upstream https://github.com/Frank-W-B/regression-case-study.git (push)
```

4. Everyone makes their own branch to work in

```
[00:34] $  
✓ ~/regression-case-study [master {origin/master}|✓]  
[00:34] $ git branch  
* master  
✓ ~/regression-case-study [master {origin/master}|✓]  
[00:34] $ git branch frank  
✓ ~/regression-case-study [master {origin/master}|✓]  
[00:34] $ git branch  
frank  
* master  
✓ ~/regression-case-study [master {origin/master}|✓]  
[00:34] $ git checkout frank  
Switched to branch 'frank'  
✓ ~/regression-case-study [frank L|✓]  
[00:34] $ git branch  
* frank  
master
```

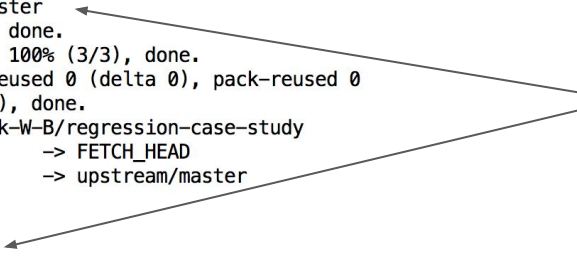
5. Do you work!

```
1 import numpy as np
2
3 from sklearn.datasets import load_boston
4 from sklearn.ensemble import RandomForestRegressor
5 from sklearn.pipeline import Pipeline
6 from sklearn.preprocessing import Imputer
7 from sklearn.model_selection import cross_val_score
8
9 rng = np.random.RandomState(0)
10
11 dataset = load_boston()
12 X_full, y_full = dataset.data, dataset.target
13 n_samples = X_full.shape[0]
14 n_features = X_full.shape[1]
15
16 # Estimate the score on the entire dataset, with no missing values
17 estimator = RandomForestRegressor(random_state=0, n_estimators=100)
18 score = cross_val_score(estimator, X_full, y_full).mean()
19 print("Score with the entire dataset = %.2f" % score)
20
21 # Add missing values in 75% of the lines
22 missing_rate = 0.75
23 n_missing_samples = np.floor(n_samples * missing_rate)
24 missing_samples = np.hstack((np.zeros(n_samples - n_missing_samples, dtype=np.bool),
25                               np.ones(n_missing_samples, dtype=np.bool)))
26 rng.shuffle(missing_samples)
27 missing_features = rng.randint(0, n_features, n_missing_samples)
28
29 # Estimate the score without the lines containing missing values
30 X_filtered = X_full[~missing_samples, :]
31 y_filtered = y_full[~missing_samples]
32 estimator = RandomForestRegressor(random_state=0, n_estimators=100)
33 score = cross_val_score(estimator, X_filtered, y_filtered).mean()
34 print("Score without the samples containing missing values = %.2f" % score)
35
36 # Estimate the score after imputation of the missing values
37 X_missing = X_full.copy()
38 X_missing[np.where(missing_samples)[0], missing_features] = 0
39 y_missing = y_full.copy()
40 estimator = Pipeline([("imputer", Imputer(missing_values=0, strategy="mean", axis=0)),
41                       ("forest", RandomForestRegressor(random_state=0, n_estimators=100))])
42 score = cross_val_score(estimator, X_missing, y_missing).mean()
43 print("Score after imputation of the missing values = %.2f" % score)
~
~
"imputing_values.py" [New] 43L, 1863C written
```

6. Push your branch to your fork, but WAIT!

Before you add, commit, and push you should really make sure your repo is current with the upstream master branch!

```
✓ ~/regression-case-study [frank L|...1]
01:45 $ ls
README.md      config.yaml    data           images          imputing_values.py  score_model.py  soln_fb        src
✓ ~/regression-case-study [frank L|...1]
01:45 $ git pull upstream master
remote: Counting objects: 3, done.
remote: Compressing objects: 100% (3/3), done.
remote: Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (3/3), done.
From https://github.com/Frank-W-B/regression-case-study
 * branch          master      -> FETCH_HEAD
 * [new branch]    master      -> upstream/master
Updating 7f39063..9028a12
Fast-forward
 data_cleaning.py | 4 ++++
 1 file changed, 4 insertions(+)
 create mode 100644 data_cleaning.py
✓ ~/regression-case-study [frank L|...1]
01:45 $ ls
README.md      data           data_cleaning.py  images          imputing_values.py  score_model.py  soln_fb        src
config.yaml
```



When pulled upstream master branch it downloaded a `data_cleaning` file that the upstream owner merged onto master branch

6. Ok now push your branch to your fork

```
✓ ~/regression-case-study [frank L|...1]
```

```
[01:53 $ git status
```

```
On branch frank
```

```
Untracked files:
```

```
(use "git add <file>..." to include in what will be committed)
```

```
    imputing_values.py
```

```
nothing added to commit but untracked files present (use "git add" to track)
```

```
✓ ~/regression-case-study [frank L|...1]
```

```
[01:53 $ git add imputing_values.py
```

```
✓ ~/regression-case-study [frank L|●1]
```

```
[01:53 $ git commit -m "Add imputing values file"
```

```
[frank 8d622f7] Add imputing values file
```

```
1 file changed, 43 insertions(+)
```

```
create mode 100644 imputing_values.py
```

```
✓ ~/regression-case-study [frank L|✓]
```

```
[01:53 $ git push origin frank
```

```
Counting objects: 3, done.
```

```
Delta compression using up to 4 threads.
```

```
Compressing objects: 100% (3/3), done.
```

```
Writing objects: 100% (3/3), 924 bytes | 0 bytes/s, done.
```

```
Total 3 (delta 1), reused 0 (delta 0)
```

```
remote: Resolving deltas: 100% (1/1), completed with 1 local object.
```

```
To https://github.com/fwbttest71/regression-case-study.git
```

```
9028a12..8d622f7 frank -> frank
```

```
✓ ~/regression-case-study [frank L|✓]
```

```
01:54 $ █
```

Add, commit like usual.

Push the branch you made to origin



6. See that your branch exists on your fork

fwbtest71 / regression-case-study Private
forked from Frank-W-B/regression-case-study

Unwatch 1 Star 0 Fork 224

Code Pull requests 0 Projects 0 Wiki Pulse Graphs Settings

Case study for regression (week 4).

Edit

[Add topics](#)

New branch on GH



18 commits 2 branches 0 releases 6 contributors

Branch: frank New pull request

Create new file Upload files Find file Clone or download

This branch is 1 commit ahead of Frank-W-B:master. Pull request Compare

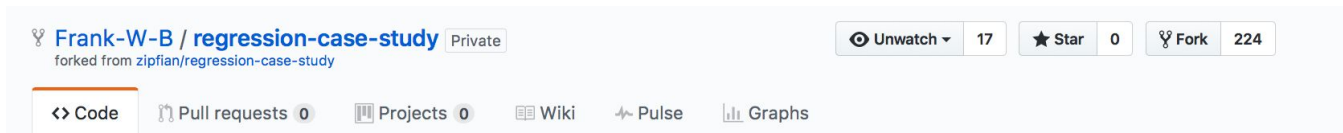
Frank-W-B Add imputing values file Latest commit 8d622f7 a minute ago

data	Add solution files	3 months ago
images	minor changes	2 years ago
soln_fb	Add fb solution files	11 hours ago
src	Add solution files	3 months ago
.gitignore	create .gitignore so we don't track Train.csv	6 months ago
README.md	Updating Scary Wording about submission	5 months ago
config.yaml	remove default channel/username in config.yaml	8 months ago
data_cleaning.py	Add data cleaning file	an hour ago
imputing_values.py	Add imputing values file	a minute ago
score_model.py	Initial commit.	2 years ago

New files



7. Collaborator issues a pull request to upstream



Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also [compare across forks](#).

base fork: Frank-W-B/regression-case-... base: master ... head fork: fwbttest71/regression-case-st... compare: frank

✓ Able to merge. These branches can be automatically merged.

Add imputing values file

Write Preview

Leave a comment

Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.

☒ Allow edits from maintainers. [Learn more](#)

Create pull request

Reviewers

No reviews—request one

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Upstream

Collaborator

7. Upstream sees the pull request...

Frank-W-B / **regression-case-study** Private
forked from zipflan/regression-case-study

Watch 17 Star 0 Fork 224

Code Pull requests 1 Projects 0 Wiki Pulse Graphs Settings

Filters is:pr is:open Labels Milestones New pull request

☐ 1 Open ✓ 0 Closed Author Labels Projects Milestones Reviews Assignee Sort

☐ **Add imputing values file**
#1 opened 3 minutes ago by fwbtst71

ProTip! Type on any issue or pull request to go back to the issue listing page.

7. And chooses to merge it

The screenshot shows a GitHub pull request for the repository "Frank-W-B / regression-case-study". The pull request is titled "Add imputing values file #1" and is from the "fwbtest71" branch to the "Frank-W-B:master" branch. The interface includes a navigation bar with tabs for Code, Pull requests (1), Projects (0), Wiki, Pulse, Graphs, and Settings. The pull request details show a conversation with a comment from "fwbtest71" stating "No description provided." and a commit titled "Add imputing values file" with commit hash "8d622f7". A green box highlights the message: "This branch has no conflicts with the base branch. Merging can be performed automatically." Below this message is a green button labeled "Merge pull request" and a link to "view command line instructions." The right sidebar contains sections for Reviewers, Assignees, Labels, Projects, and Milestones, all of which are currently empty.

Frank-W-B / regression-case-study Private
forked from ziplan/regression-case-study

Watch 17 Star 0 Fork 224

Code Pull requests 1 Projects 0 Wiki Pulse Graphs Settings

Add imputing values file #1

Edit

Open fwbtest71 wants to merge 1 commit into Frank-W-B:master from fwbtest71:frank

Conversation 0 Commits 1 Files changed 1 +43 -0

fwbtest71 commented 4 minutes ago

No description provided.

Add imputing values file 8d622f7

Add more commits by pushing to the **frank** branch on **fwbtest71/regression-case-study**.

This branch has no conflicts with the base branch
Merging can be performed automatically.

Merge pull request or view [command line instructions](#).

Reviewers: No reviews—request one

Assignees: No one—assign yourself

Labels: None yet

Projects: None yet

Milestone:

7. See confirmation - and delete the branch

The screenshot shows a GitHub pull request interface for the repository 'Frank-W-B / regression-case-study'. The pull request is titled 'Add imputing values file #1' and is in the 'Merged' state. It shows that 'Frank-W-B' merged 1 commit into 'Frank-W-B:master' from 'fwbtest71:frank' 16 seconds ago. Below the merge information, there is a section for 'Add imputing values file' with a commit hash '8d622f7'. A notification banner at the bottom states 'Pull request successfully merged and closed' and 'You're all set—the fwbtest71:frank branch can be safely deleted.' A button labeled 'Delete branch' is visible in the bottom right corner of the notification banner, with an arrow pointing to it.

Frank-W-B / regression-case-study Private Watch 17 Star

Code Pull requests 1 Projects 0 Wiki Pulse Graphs Settings

Add imputing values file #1

Merged Frank-W-B merged 1 commit into Frank-W-B:master from fwbtest71:frank 16 seconds ago

Conversation 0 Commits 1 Files changed 1

fwbtest71 commented 7 minutes ago

No description provided.

Add imputing values file 8d622f7

Frank-W-B merged commit da14edf into Frank-W-B:master 16 seconds ago Revert

Avoid bugs by automatically running your tests.
Continuous integration can help catch bugs by running your tests automatically.
Merge your code with confidence using one of our continuous integration providers. [Learn more](#)

Pull request successfully merged and closed
You're all set—the fwbtest71:frank branch can be safely deleted. [Delete branch](#)

7. Now, for collaborator to have local and remote repos up-to-date with upstream:

```
$ git pull upstream master  
$ git push origin master
```

For more info:

How to collaborate on git:

<https://code.tutsplus.com/tutorials/how-to-collaborate-on-github--net-34267>

The difference between origin and upstream on github:

<http://stackoverflow.com/questions/9257533/what-is-the-difference-between-origin-and-upstream-on-github>

Resolving merge conflicts using the command line:

<https://help.github.com/articles/resolving-a-merge-conflict-using-the-command-line/>

Resolving merge conflicts on Github:

<https://help.github.com/articles/resolving-a-merge-conflict-on-github/>

