

## Group Assignment 1

### Stock Price Prediction Data Pipeline

#### **Groups:**

<https://docs.google.com/spreadsheets/d/1eIMCII0xN9YbgT7fqi01Efblb2pvq6BkJ0Pd6Kesns/edit?usp=sharing>

#### **Learning Objectives:**

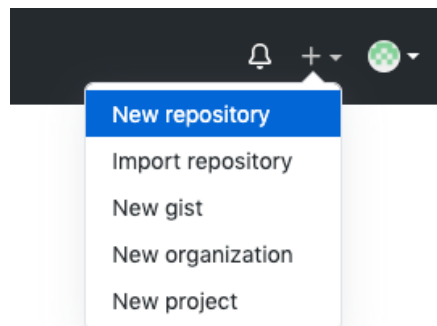
1. Collaborate using Git and GitHub
2. Building data pipelines in Python

#### **Background**

If we can accurately predict stock price movement, we can trade stocks based on these predictions. This is a hard problem, and there are many ways to accomplish this task. The goal of this assignment is not to try out several algorithms, or learn about algorithms. Instead, the goal is to pick your favorite algorithm (see <https://scikit-learn.org/stable/>), and prepare the data (described below) so that you can estimate a model that predicts stock price, or stock price direction with the goal of trading that stock. Grading of your code in this assignment is based on creativity, quality, and effort. Given the many ways to solve this problem, the open-ended nature of this assignment, and the dependency on the algorithm that you use, an instructor solution will not be provided to this assignment.

#### **Deliverables**

1. In your initial team meeting assign a Repo Owner, who will review pull requests and will merge code from other branches into the main branch. The Repo Owner will also create a new private repo per the screenshots below. In the screenshots, replace mballings with the Repo Owners account.



## Create a new repository

Owner \* mballings / Repository name \* assignment1 ✓

Great repository names are short and memorable. Need inspiration? How about [glowing-chainsaw](#)?

Description (optional)

☐ Public  
Anyone on the internet can see this repository. You choose who can commit.

☒ Private  
You choose who can see and commit to this repository.

Initialize this repository with:  
Skip this step if you're importing an existing repository.

☒ Add a README file  
This is where you can write a long description for your project. [Learn more.](#)

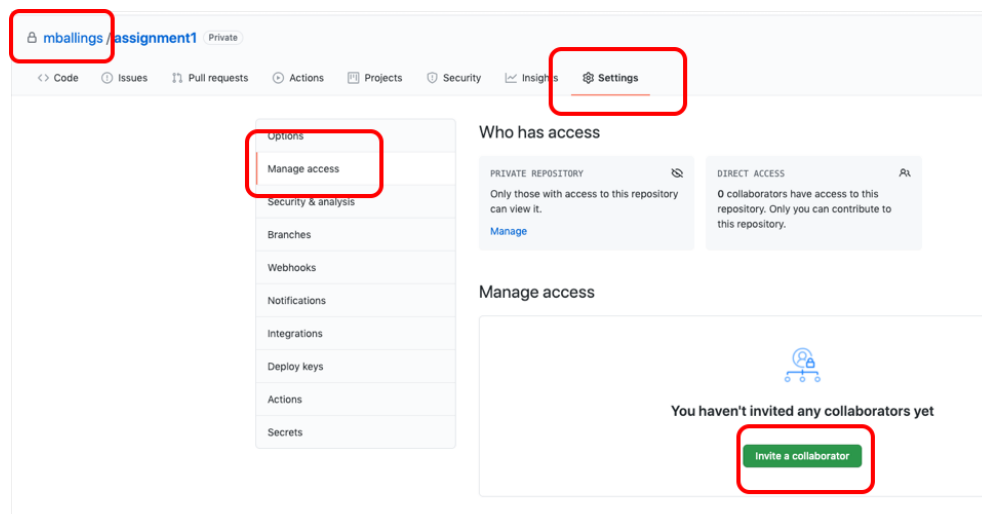
☐ Add .gitignore  
Choose which files not to track from a list of templates. [Learn more.](#)

☐ Choose a license  
A license tells others what they can and can't do with your code. [Learn more.](#)

This will set `main` as the default branch. Change the default name in your [settings](#).

Create repository

Once the Repo Owner has created the private repo, s/he will invite her/his team members and ballingsclass.



Once the repo has been created, the goal is to develop high-quality working code and store it in the main branch. When code is in the main branch it is a signal to collaborators that the author wanted to share this code with the team, and that the code is vetted by the Repo Owner. This means that the Repo Owner will run the code that is the subject of the pull request and see if it works, before performing the merge.

The owner and collaborators should never push directly to the main branch, as this may destroy working code. While this can always be retrieved, it will make you lose valuable time. Always push to topic branches (never the main branch), and create pull requests

when you want to merge your branch with the main branch. The Repo Owner then has to review the request and perform the merge or request changes. To avoid naming conflicts with topic branches, it is good to append your first name to the branch name. For example, if you are writing code to read data, you could create a branch called `readdata_john`, if your first name is John. If Alice is also working on the same feature independently then she would create a branch called `readdata_alice`. This would avoid unintended consequences.

2. A .py file with working Python code to prepare your stock price prediction data for analysis. Also include the code that you have written for the presentation and your model code.
3. A presentation (PDF or PowerPoint) with details on data preparation and model performance.

## **Data**

Five-second bars: a bar is the open, high, low, close, and average price of a stock, and trading volume, in a five-second period.

Variable name	Description	When record inserted?	Summary
tickerid	The identifier of the stock	epochtime + 5s	{0,1,2,3,4}
epochtime	The number of seconds since Jan 1, 1970 and the start of the five-second bars	epochtime + 5s	[1611864635, 1612472920]
open	The price of the stock at the beginning of the 5s period.	epochtime + 5s	[18.39, 865.9]
high	The maximum price of the stock in the 5s period.	epochtime + 5s	[18.4, 866.3]
low	The minimum price of the stock in the 5s period.	epochtime + 5s	[18.39, 865.9]
close	The price of the stock at the end of the 5s period.	epochtime + 5s	[18.39, 866.3]
volume	The number of stocks traded in the 5s period.	epochtime + 5s	[0, 11334]
weightedavgprice	The volume weighted average price (wap) of	epochtime + 5s	[18.4, 866]

	<p>the stock in the 5s period:</p> $wap = \frac{\sum_j P_j Q_j}{\sum_j Q_j}$ <p>P<sub>j</sub>: price of trade <i>j</i> Q<sub>j</sub>: quantity of trade <i>j</i></p>		
--	--	--	--

### **Grading**

Can be different for members of the same group and is based on

- Number and quality of commits in the main branch
- Python code
- Presentation

### **Deadline**

3/8 – 2:50pm