Collectively, the files took 800+ MB of storage, we then used Git LFS to upload the large files.

**Steps:**

1. Clone the repo

2. Change the directory to the cloned repo, using Git Bash

3. Install Git LFS using "git install lfs"

4. Track the file you want to commit using "git lfs track filename.filetype"

    - e.g git lfs track cars.csv

5. Stage the file using "git add filename.filetype"

6. Commit with a message using "git commit -m "added xxx file"

7. Push to the GitHub repo using "git push"

The above steps worked.

**Some problems that we encountered:**

**PROBLEM #1**:

The only challenge we encountered when uploading the files was that one of the team members who tried to upload the files at first was almost out of storage space and would not be able to use Git LFS without paying for additional storage.

**SOLUTION**:

Another member of the team uploaded the files using their account, which had enough storage.

From there, the team could then pull from the main branch and update their branches

**PROBLEM #2**

Forked the Streamlit repo to the local machine and followed the instructions from the template repo.

* When we had to run "conda install -c conda-forge scikit-surprise" on git bash, we received an error saying "bash: conda: command not found".

**Attempts at solving:**

(1) Opened Anaconda Prompt, Pasted the code "**conda install -c conda-forge scikit-surprise**" and it downloaded and installed **successfully**.

* We then went back to git bash, navigated to the streamlit repo, and ran the .py file to launch the web server again, the server **launched** but we received an error "**ModuleNotFoundError: No module named 'surprise'** "

(2) Went to Jupyter notebook and tried installing surprise package with **"!pip install surprise**", and it said requirement already satisfied and that it installed successfully.

* Launched the server again, and received the same error as in (1) above.

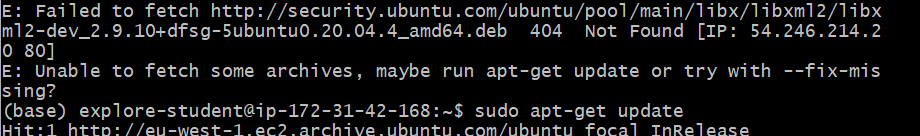
(3) We proceeded by using another team member’s local machine, and did not get errors.

**PROBLEM #3 :**

While following the instructions from <https://gist.github.com/JonGerrand/01c3b3e7e377fcd3aa55114631d2cb5e>, after running the first line, i.e,

“sudo apt-get install automake autotools-dev fuse g++ git libcurl4-gnutls-dev libfuse-dev libssl-dev libxml2-dev make pkg-config -y”

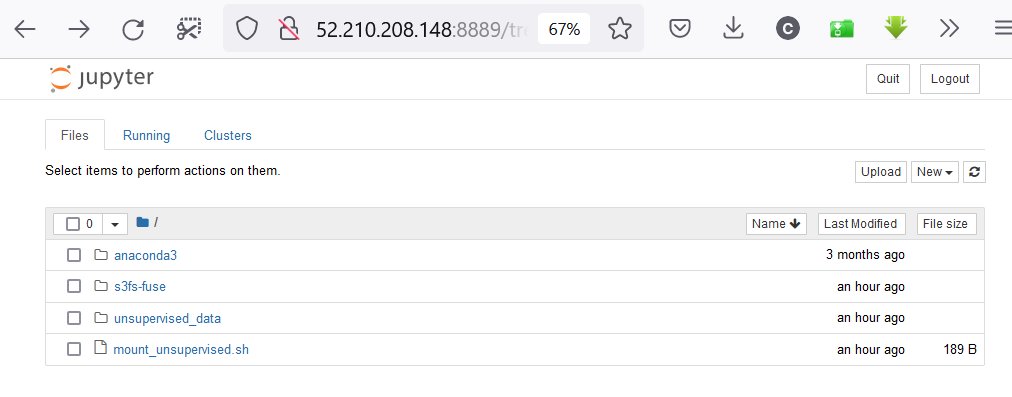
* We received the error below:



This was then solved by running “sudo apt-get update”. We then ran the above first code again and it worked. The rest of the instructions did not give any problems.

**NB:** We only have read access to the files, and this was intentional. “You’ll notice that **your mounted directory has *read-only permissions****.* This is intentional, as all students are accessing the same S3 bucket, and implies that **you cannot write or save work within this directory** - becareful of this fact!”

Just for observation, we then opened jupyter notebooks from the EC2 instance to see the files and this is what was observed:



Graphical user interface, text, application, email

Description automatically generated

**Graphical user interface

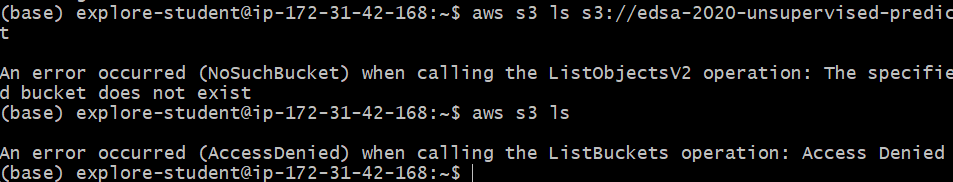
Description automatically generated**

**QUESTION**: Where or how do we access the actual data files? Or would we have to add them to the S3 bucket ourselves?

**Solution we’ve tried**

1. Logged into the EC2 instance and ran “sudo apt install awscli” to be able to use the awscli.

* Tried to access the bucket using “aws s3 ls s3://edsa-2020-unsupervised-predict”
* Also tried checking the list of buckets.
* And received the following error:

****