

Git History Cleanup Log

Git History Cleanup Process

You encountered an issue where a large file (`datacompr.csv`) was committed to your Git repository, exceeding GitHub's file size limit of 100 MB.

Here's a summary of the actions you performed to resolve the issue:

1. **Created a `.gitignore` file** to exclude `datacompr.csv` and other CSV files from being tracked.

- Used the command:

```
...
```

```
echo "CEE5735-GP_regression/datacompr.csv" >> .gitignore
```

```
...
```

2. **Attempted to remove the file from Git's tracking** using `git rm --cached`:

```
...
```

```
git rm --cached CEE5735-GP_regression/datacompr.csv
```

```
...
```

3. **Committed the changes** after untracking the file:

```
...
```

```
git commit -m "ignoring .csv data file"
```

```
...
```

4. **Tried to push the changes** but encountered GitHub's 100 MB file size limit:

```
...
```

Git History Cleanup Log

```
git push -u origin main
```

```
...
```

Result: Push was rejected due to the large file size.

5. **Cleared the Git cache** and re-added files to stage only what was needed:

```
...
```

```
git rm -r --cached .
```

```
git add .
```

```
git commit -m "Cleaned up notebook, removed starting and midterm materials"
```

```
...
```

6. **Next Steps Suggested**:

Since the file was already committed, you need to rewrite the Git history using a tool like `git filter-repo` or `BFG Repo-Cleaner`. The recommended steps involve:

- Installing `git filter-repo` via `pip` or `brew`.

- Running:

```
...
```

```
git filter-repo --path CEE5735-GP_regression/datacompr.csv --invert-paths
```

```
...
```

- Cleaning up Git references and forcing a push to update the remote repository.

Note: Force pushing is necessary after rewriting history, and collaborators need to re-clone the repository after the history rewrite.