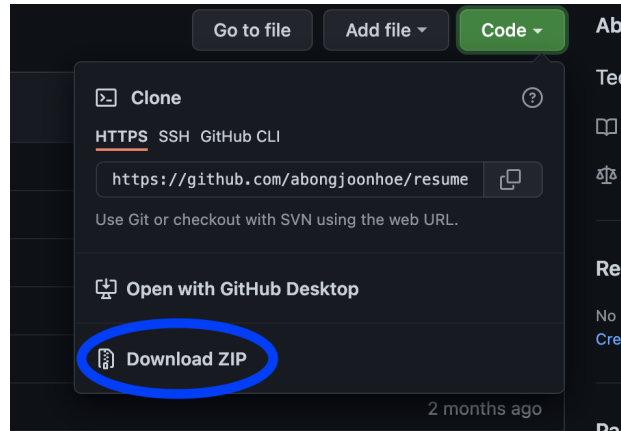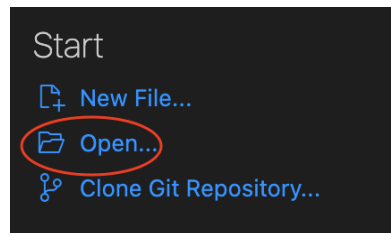# How to Clean and Migrate Check-In Data

1. Setting up Workspace
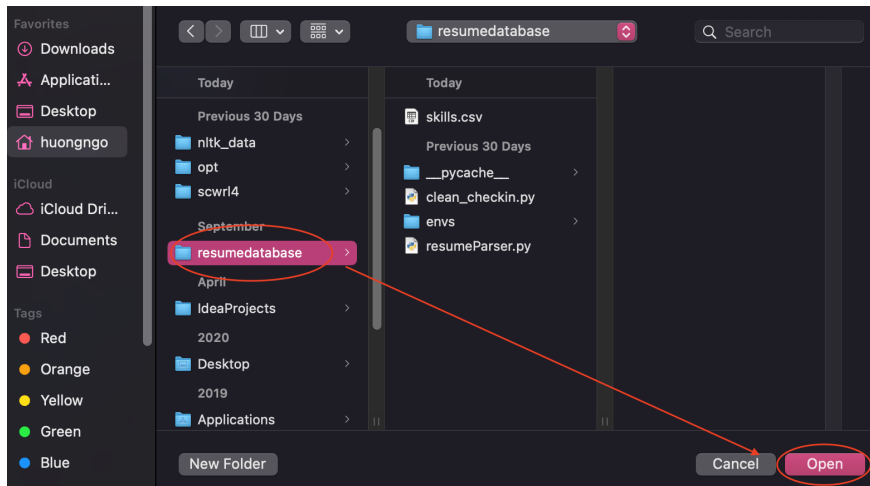   a. Download **resumedatabase** repository from
       https://github.com/abongjoonhoe/resumedatabase

       i.
       ii. Take note of where this downloaded folder is located on your computer. If this folder is in your Downloads directory, please move it to your **Home** directory.
    b. Download Visual Studio Code at https://code.visualstudio.com/
        i. You'll need to use this to run the scripts inside the folder you just downloaded.
    c. Open the file that you downloaded. This will be your workspace.

       i.
       ii. Once you click open, it will prompt you to a page that will allow you to look through your computer to find the folder that you want to open.
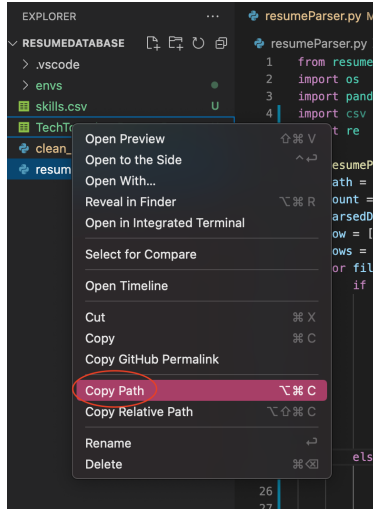   iii. You'll need to find the folder that you downloaded and choose to open that.

1.

d. Once you've set up your workspace, you need to set up an **environment**.

    i. Download **Anaconda** at https://www.anaconda.com/products/individual and install it

    ii. Go to **Terminal** and type in this command to change your current working directory to the directory of the downloaded folder. This command will only work if your downloaded folder is in your **Home** directory

```
1. cd ~/resumedatabase
```

    iii. To create your environment, type in this command

```
1. conda create --prefix ./envs python=3.8
```

    iv. Activate your environment in your workspace in Visual Studio Code in the Terminal console using this command

```
1. conda activate [path of your downloaded
   folder]/envs
```

e. Download these packages to use the packages that are required for the scripts to work

    i. Follow the instructions under **Installation** in this repository https://github.com/kbrajwani/resume_parser

    ii. Paste all the code chunks into the Terminal console

    iii. Messages should pop up showing that it's being installed

2. Making the Final Resume Book

a. Download the check-in spreadsheet as a CSV file. Move this file into the folder you downloaded

b. In clean_checkin.py, you'll see this line at the top.

    i.

    ii. You'll need to right-click on the check-in dataset to get the path. The files are listed in the column on the right side.

1.

iii. Copy the path and paste it in between the quotation marks

c. In the same file, you'll see this line at the bottom

```
finalCSV = finalDf.to_csv('NAME OF THE FINAL CHAPTER-SPECIFIC RESUME BOOK')
```
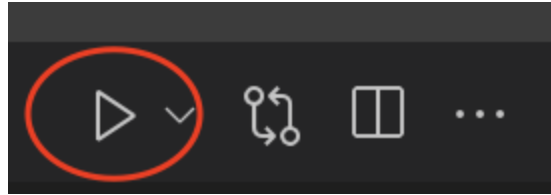
i.

ii. The final output will be the chapter-specific resume book, with not only the parsed resume information, but also the cleaned check-in data. Thus, you should appropriately name it as something like *final-ttb-resumebook.csv*.

iii. Take note that when you are making file names, you must include the extension. Here it is .csv

d. In resumeParser.py, you'll see this line at the top

```
resumeParser(fileList):
    path = "PATH OF WHERE THE ENTIRE WORKSPACE WILL BE"
```
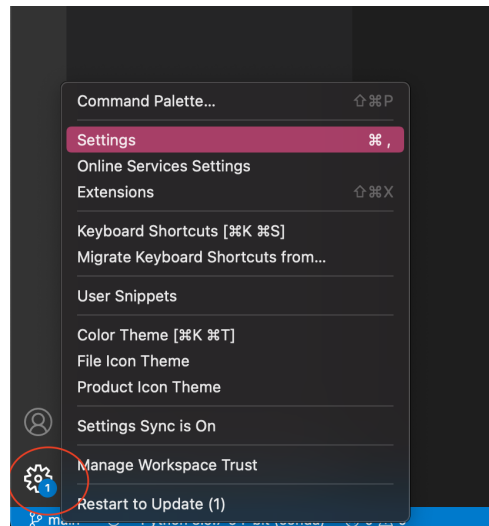
i.

1. Here, you'll need to copy the path of your workspace, which is the folder you downloaded at the beginning.

2. You can get the path from the path of the check-in dataset you copy and pasted in Step 2b. It should just be /Users/[name of your Home directory]/resumedatabase

3. Copy the path between the quotation marks

e. In the same file, you'll also see these lines

```
parsedCSV = parsedDf.to_csv('NAME OF CSV FILE WITH PARSED RESUME DATA', header=['Degre
finalResDf = cleanData('NAME OF CSV FILE WITH PARSED RESUME DATA (SAME WITH PREVIOUS L
finalResCSV = finalResDf.to_csv('NAME OF CSV FILE WITH PARSED RESUME DATA (SAME WITH P
```
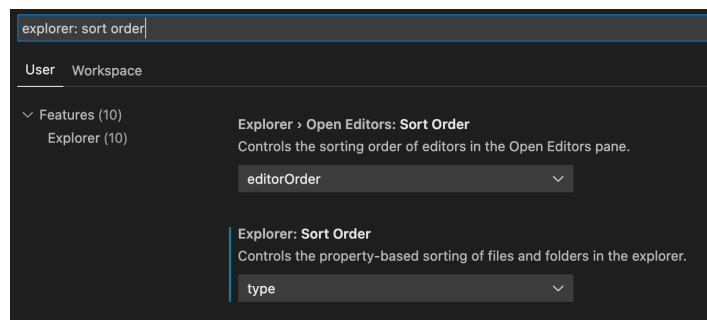
i.

1. Here, you'll need to replace what is inside the quotation marks with an appropriate name of the CSV file with the parsed resume data.

2. It is recommended that you name it like this *ttb-resumes.csv*

f. Now, all you have to do is run the clean_checkin.py script

i. Press this triangle button that is located on the top left hand corner of VS Code

    ii.

    iii.   This should initiate a console to pop up from the bottom and there will be messages printing on the console to show that the script is running.

  g. There will be messages indicating which checkpoint the check-in file has gone thorugh and when the entire process is completed.

  h. Keep in mind that when the resumes are being downloaded, they'll be downloaded inside the workspace. Moreover, the file outputs will also be in the workspace after the script is done running.

  i. Because there will be a lot of file, it is better to have it sorted by type. Go into the Settings of VS Code, search *explorer: sort order* in the search bar at the top and **change the sort order from default to type**.

    i.

    ii.

3. Merging Resume Book and Resume Database
   a. Import the CSV file of the final resume book into your Google Drive and open it up as a spreadsheet. Open up the National Resume Database spreadsheet as well.
   b. Copy the entire final resume book spreadsheet and paste it after the last entry of the Data spreadsheet in the National Resume Database.

c.   You will have to get rid of the first row as it will contain the names of the columns and is not needed.
d.   Make sure to label the resume book with the name of the chapter and the season year of the hackathon event. Follow the format that is used for the other resume books in the Data spreadsheet.
e.   You can then select all the rows of the new final resume book and move it up to the top. Make sure to select the entry with the label too.

4.   Check the front-end of the database to make sure the new resume book data has been updated on that end too.

For any issues, please contact Huong/Zoe Ngo on the TechTogether Slack or at zoengo2002@gmail.com