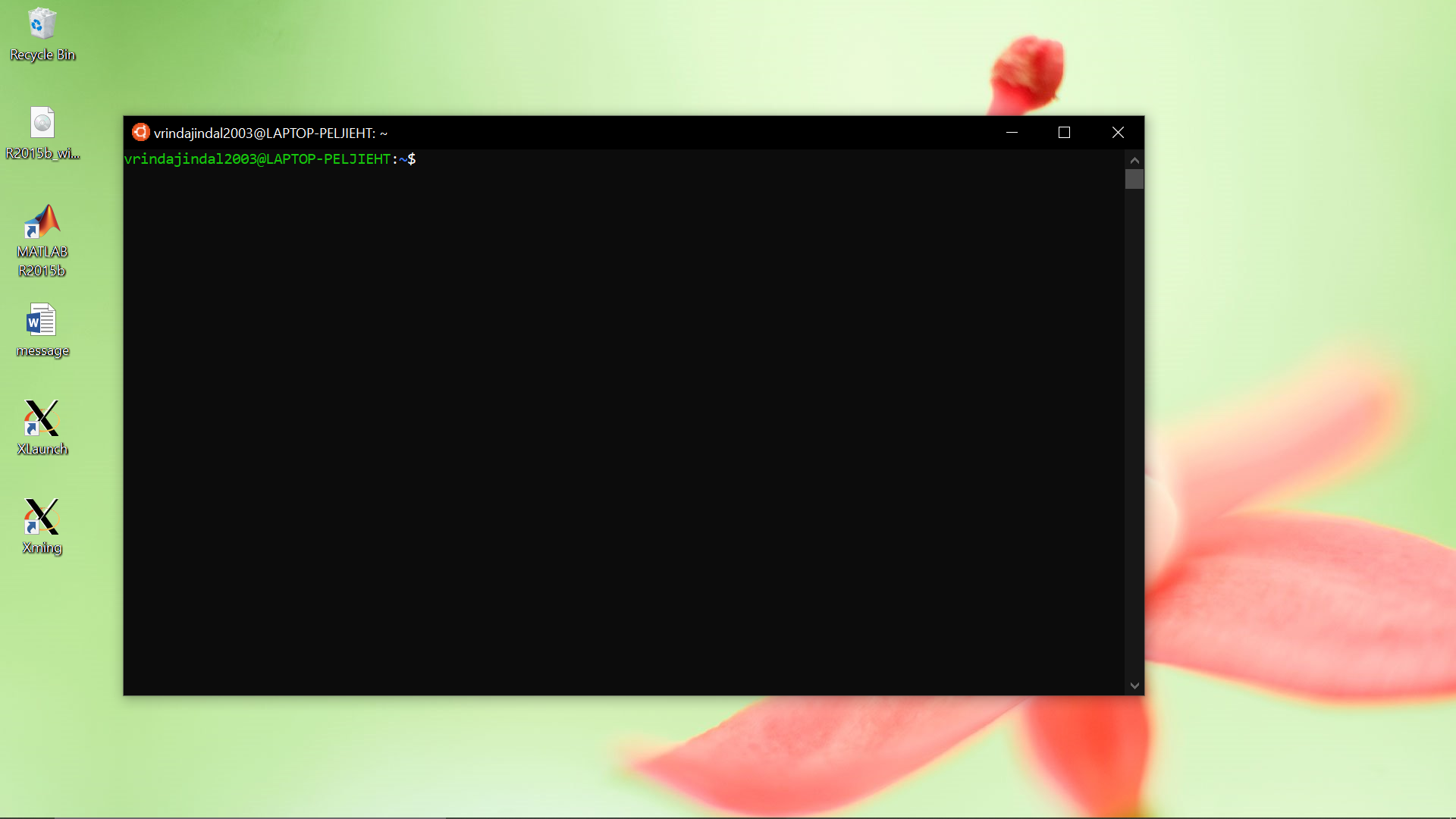
SET UP:

* + - Shalini told me that windows10 is most commonly used in the lab.
    - I used python and bash to complete the task at hand. Follow the instructions to get your set up
    - I use Windows Subsytem for linux (WSL), follow the tutorial below to get WSL on your systems –

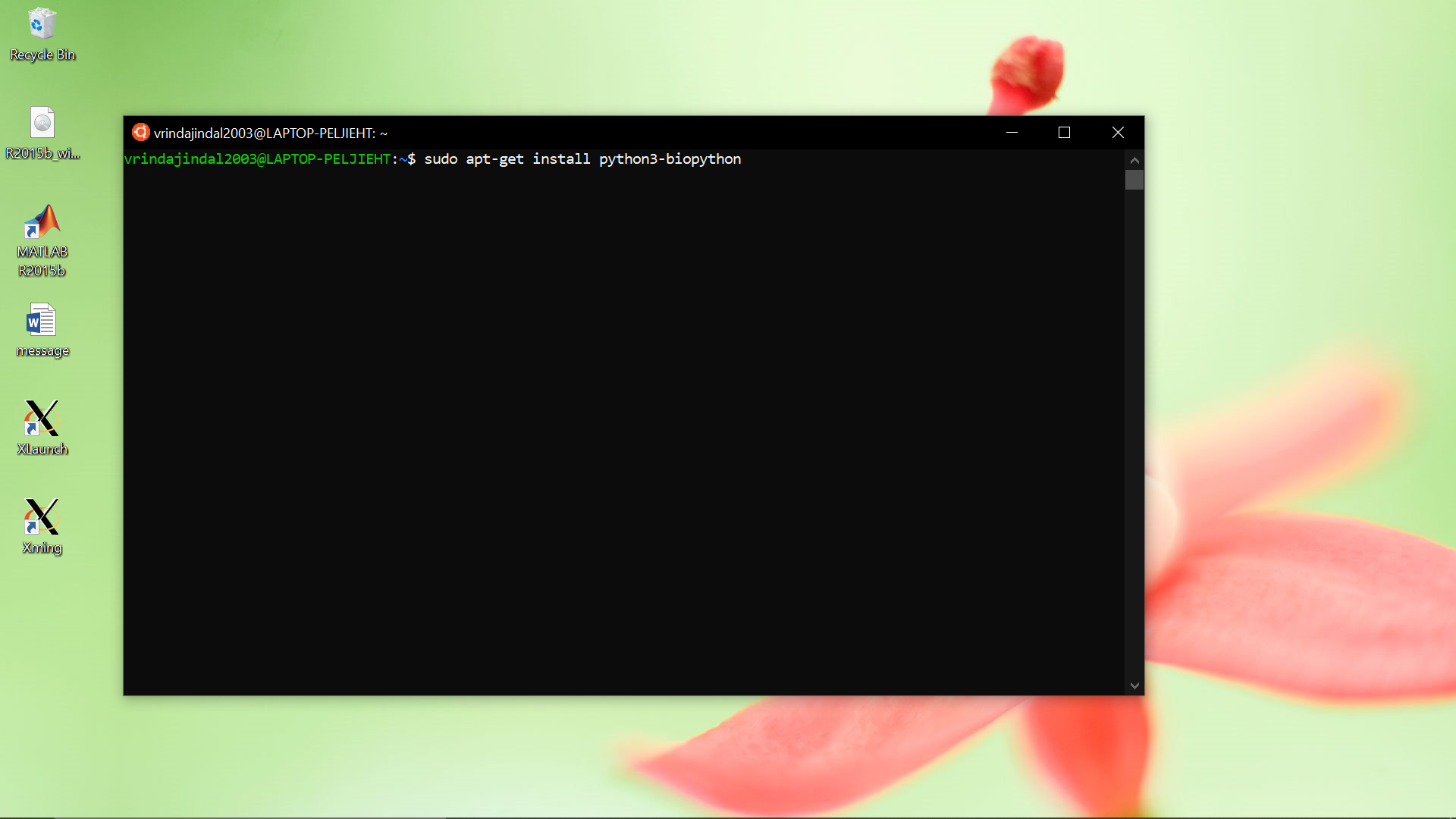
<https://devtidbits.com/2017/11/09/ubuntu-linux-on-windows-10-how-to/>



You should see something like this once you follow the tutorial

Next we will install the biopython module that is required to run the program

Write the below mentioned command on the terminal window



Press Enter

It will ask for the password you set up in the beginning. When you type the password it wont be visible. After typing the password press enter. Let it run and finish.

Congratulations, you have successfully completed the set up now you can start to run your program.

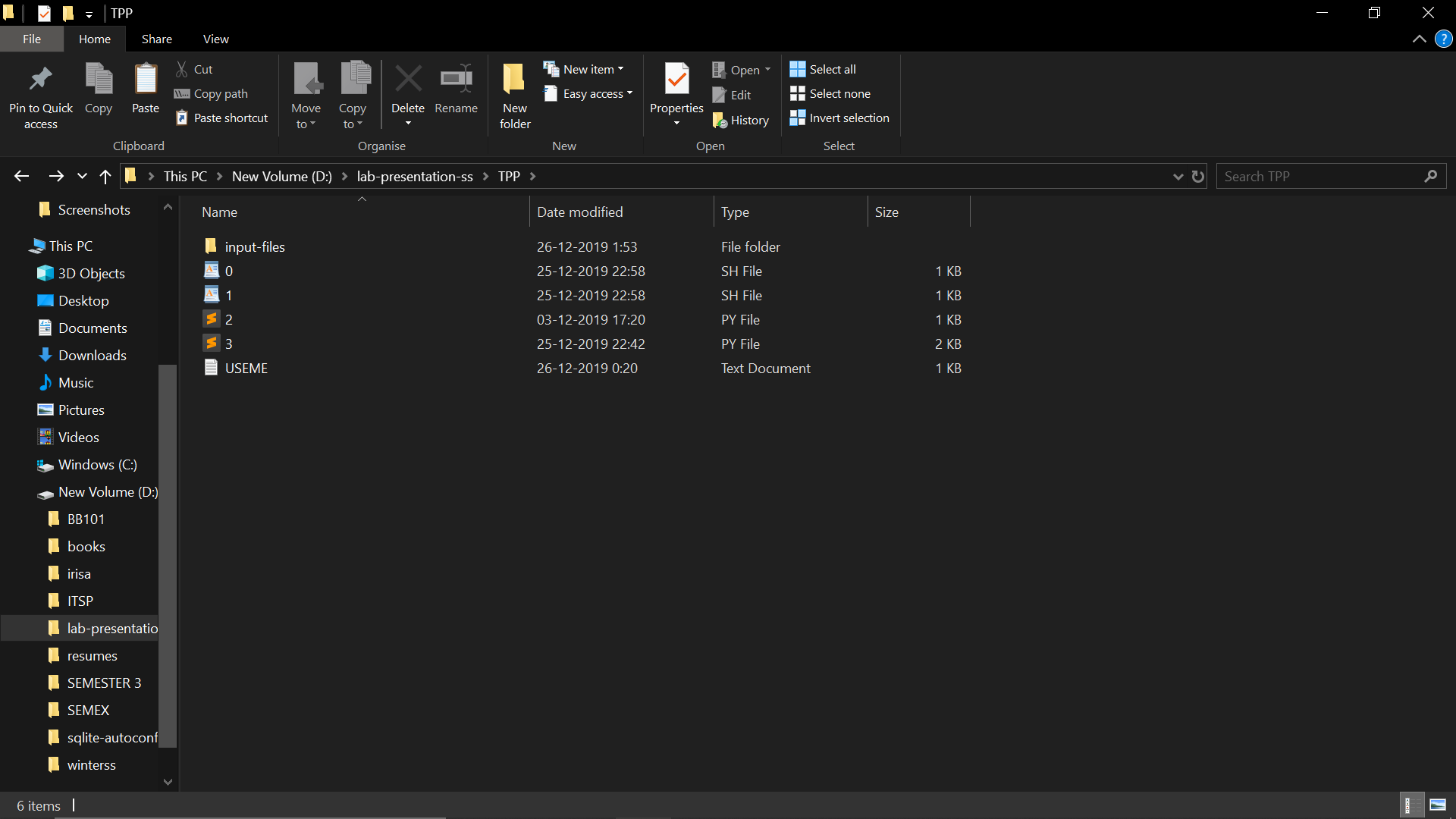
TPP FILE FORMAT:

Download the contents of the folder provided. (TPP)

(if you download the entire folder you will have to unzip it, via the terminal. To keep things easy download the contents in a separate folder (eg named TPP)

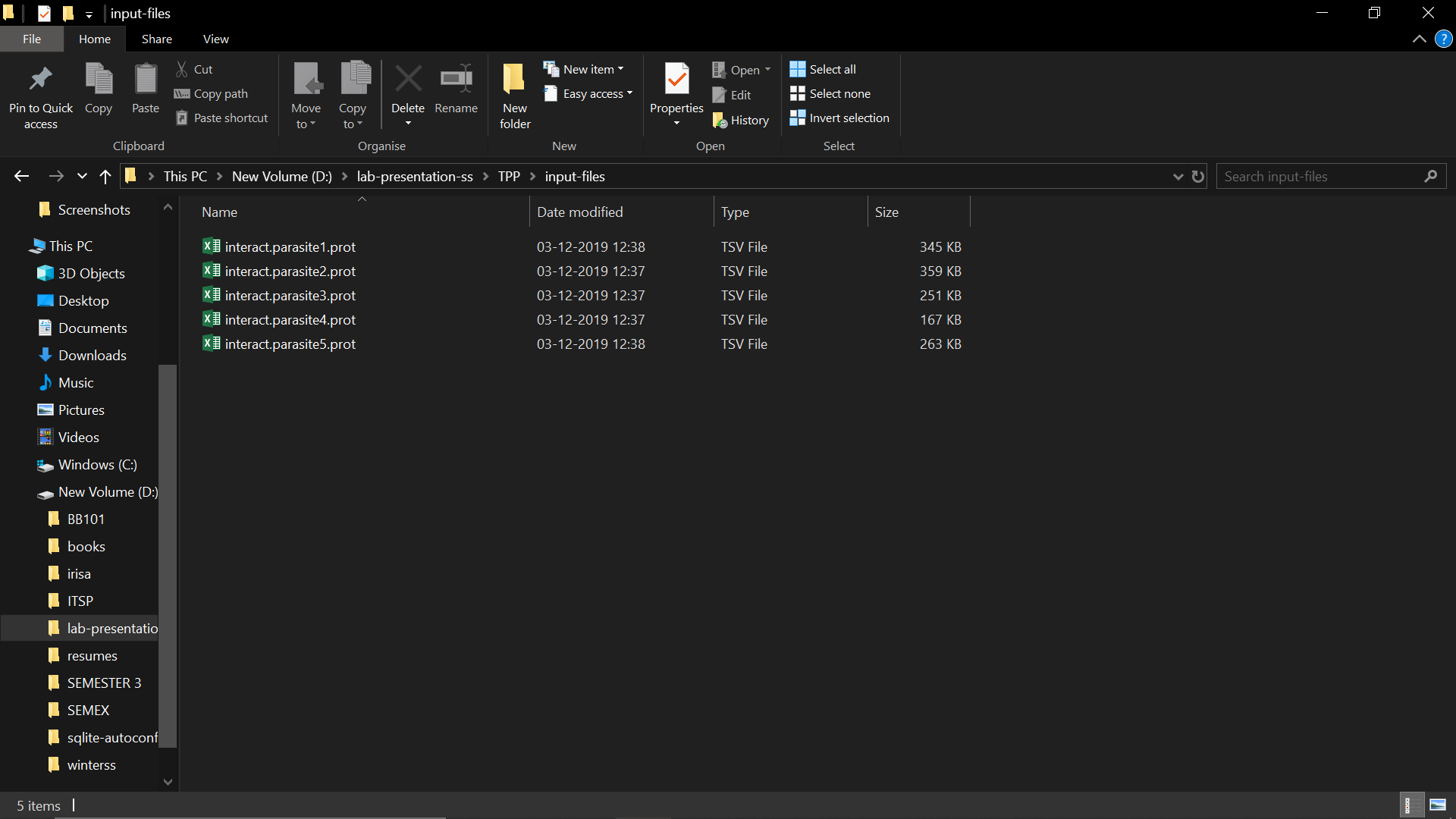
I recommend downloading all the contents in a separate folder at some place in the D/C/E drive of your system.

The downloaded folder will look like this:



STEP1: copy-paste all the files to be processed in the folder input-files (initially the folder is empty).

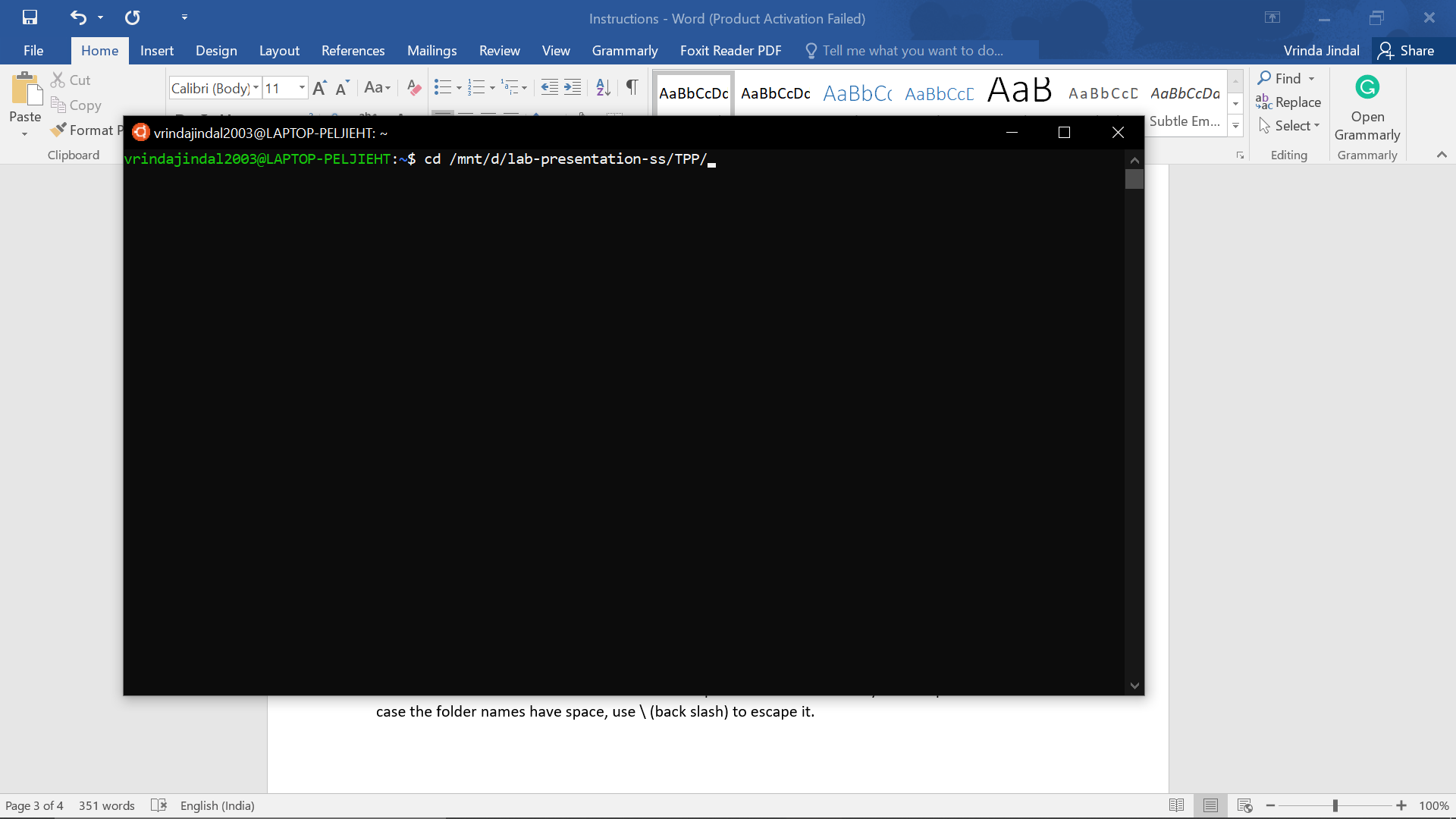
Copy-paste the database(fasta format) in to the main folder. Rename it database.

For example, I want to process the following 5 files:

STEP2: open the terminal and navigate to the folder using the following command.

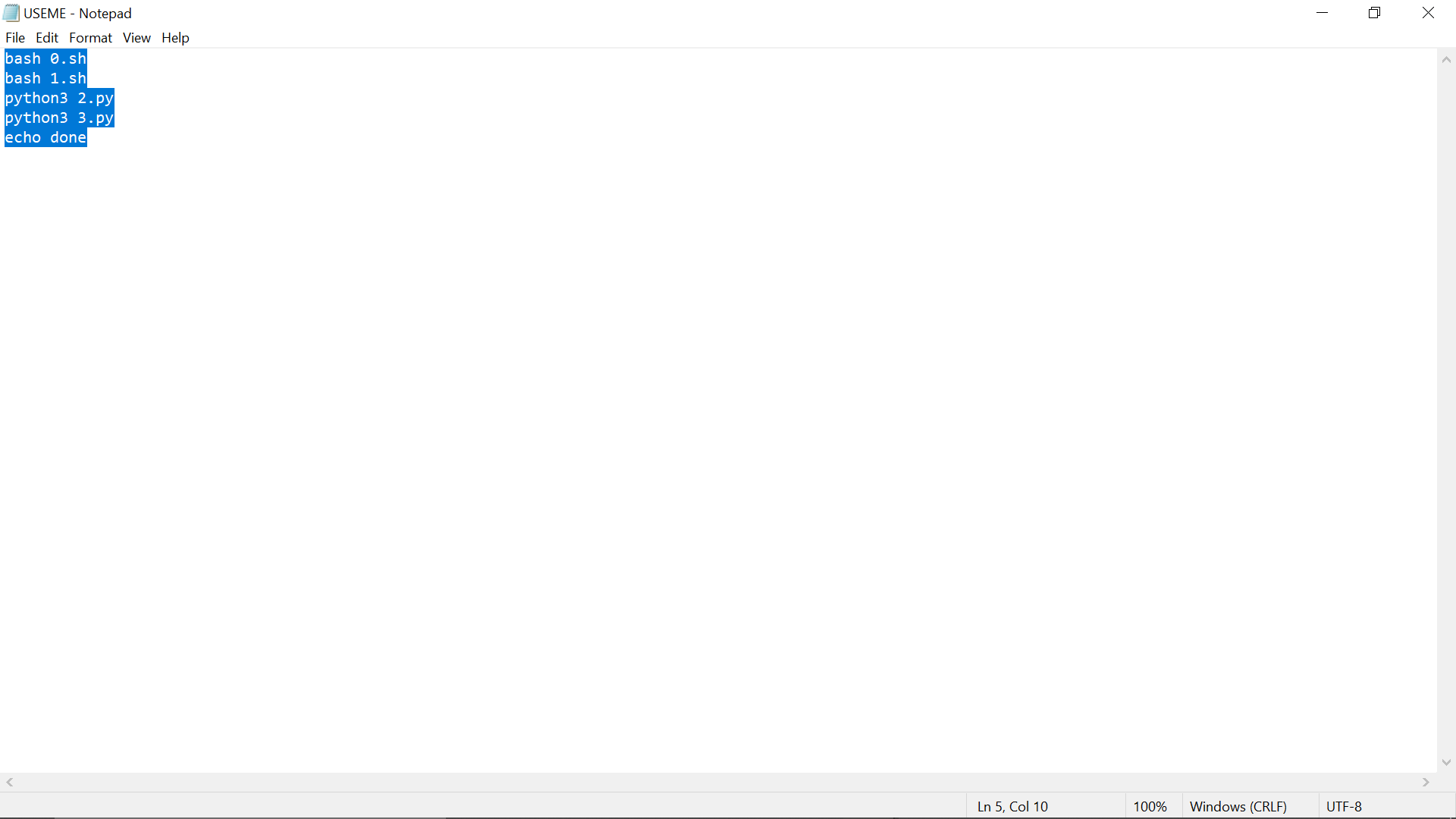
For example, if you have saved the folder in d drive: cd /mnt/d/<folder name1>/<folder name2>/..

Where folder name1 and 2 and so on are the sequential folders in which you have placed the file. In case the folder names have space, use \ (back slash) to escape it.



STEP3: open the USEME.txt file present in the folder. Copy its entire contents. (I recommend using ctrl+A followed by ctrl+C).

The useme file looks like this:



STEP4: go back to the terminal. And paste what you have copied. In most systems, tapping on the trackpad with two fingers will work. Or ctrl+v will work. Or ctrl+shift+v will work.

A lot of things will start appearing on the screen. These are the things I wanted to print out. Final after a few seconds/minutes/hours (depending on the size of the input files), it will stop. You can minimize the screen while the program is running.

Finally you will get the following files/folder in the main folder:

Final.tsv : the final output file

db.tsv : the database format that was used

a folder o-files: which contains one file for each of the input file, and has the sorted protein-peptide list. If input file is file1, the corresponding file in o-file is ofile1.

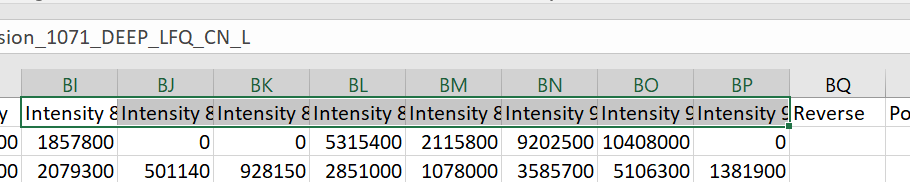
DONE!

To use this again, delete the o-files folder, db.tsv, final.tsv. and then repeat the steps :P

MAXQUANT:

Maxquant files have the samples written with the file, so the relevant columns that point out to the relevant columns have to be manually checked.

For example, in the file I processed, there were 8 samples and were present in column 61-68.

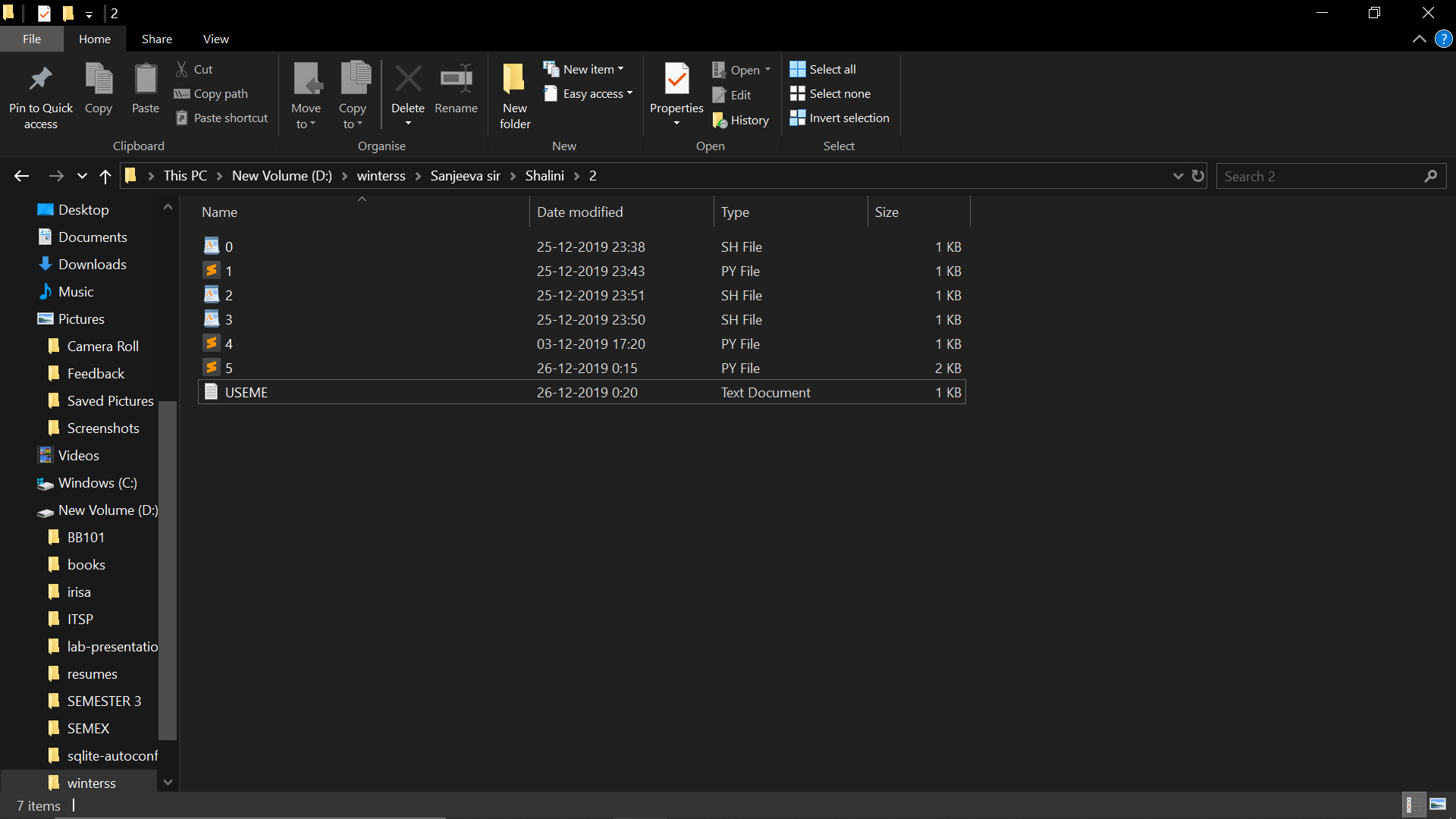


Download the contents of the folder provided. (MAXQUANT)

(if you download the entire folder you will have to unzip it, via the terminal. To keep things easy download the contents in a separate folder (eg named MAXQUANT)

I recommend downloading all the contents in a separate folder at some place in the D/C/E drive of your system.

The downloaded folder will look like this:

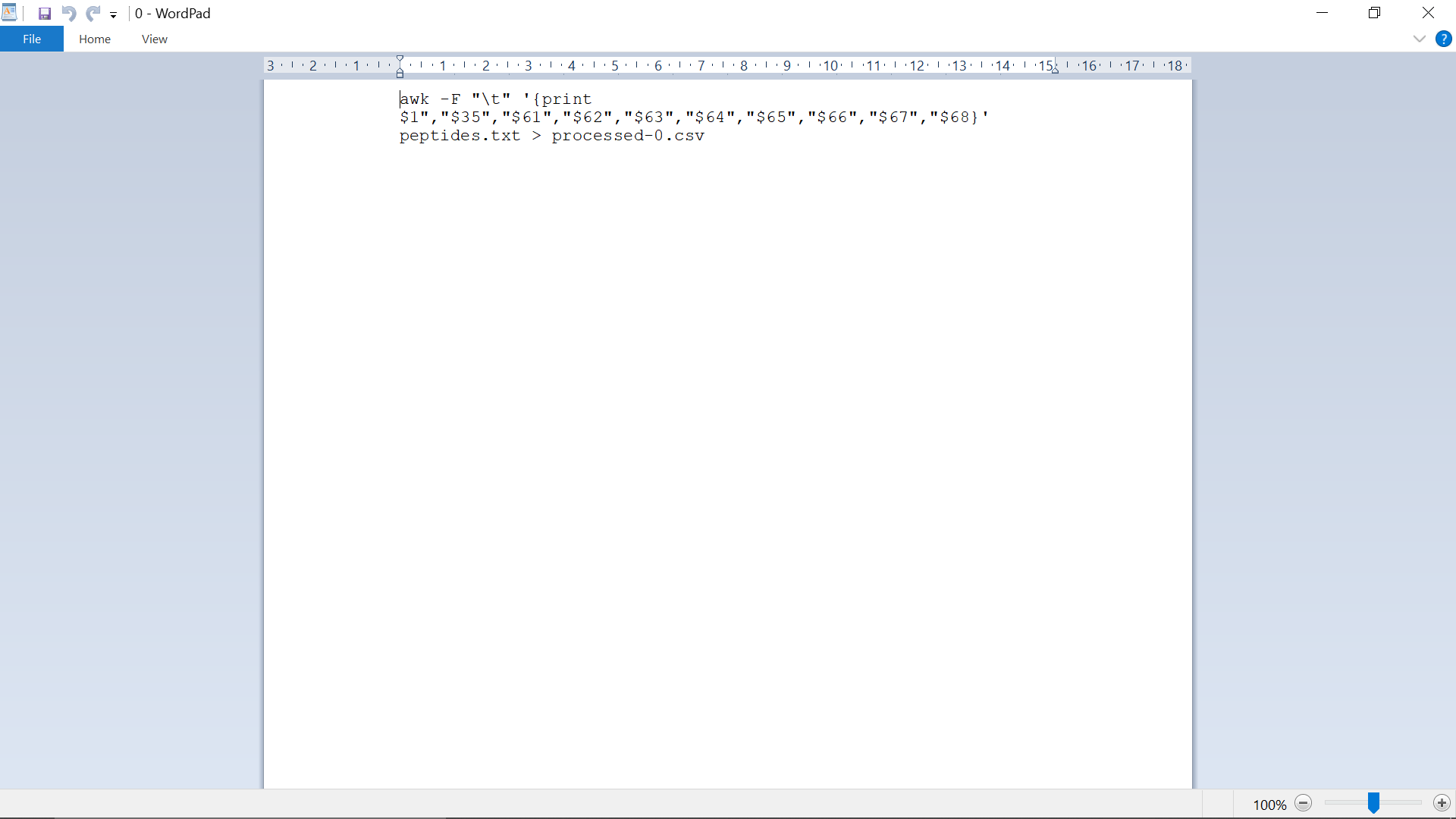


STEP1: copy-paste the input file into the main folder. Rename is peptides

Copy-paste the fasta-database into the main folder. Rename it database

STEP2: open the file 0.sh in notepad/word-pad/any other text-editor. Change the number “$61”,”$62”,…. To the numbers that denote your column numbers ie the column number that contain the intensities.

For example in the below picture, the intensities are from column 61 to 68. (the first two numbers correspond to the peptides and proteins respectively).



STEP3: follow step2, 3, 4 for TPP.

You will get a lot of files. Final.tsv is the required file. The other processed-x files are those formed intermediately during the processing as mentioned in the presentation.

I can easily write code to remove these intermediate files, if you want. But I kept them to give a taste of what actually is happening as mentioned in the presentation.

PD file:

Place the input files in the input-files folder

Place the database in the main folder

Run $ python3 pd.py