

Tanya Charan

USC ID: 6316469892

YOUTUBE LINK: <https://youtu.be/SS-X1uPgndg>

GITHUB LINK: <https://github.com/tanyacharan/dsci560-lab1>

Overview:

The assignment required us to:

1. Setup the environment: Ubuntu VM, Python, pip.
2. Go over Linux and Python essential basics: terminal navigation, understanding a simple script.
3. Do web scraping: fetch CNBC World News webpage.
4. Do Data Filtering: extract specific fields into CSVs.
5. Update Github: commit/push daily with meaningful commit messages.

This is how I set up my files in Ubuntu:

TanyaCharan_6316469892/

```
|— data/
|   |— raw_data/
|   |   |— web_data.html
|   |   |— processed_data/
|   |       |— market_data.csv
|   |       |— news_data.csv
|— scripts/
|   |— task_1.py
|   |— web_scraper.py
|   |— data_filter.py
|—
```

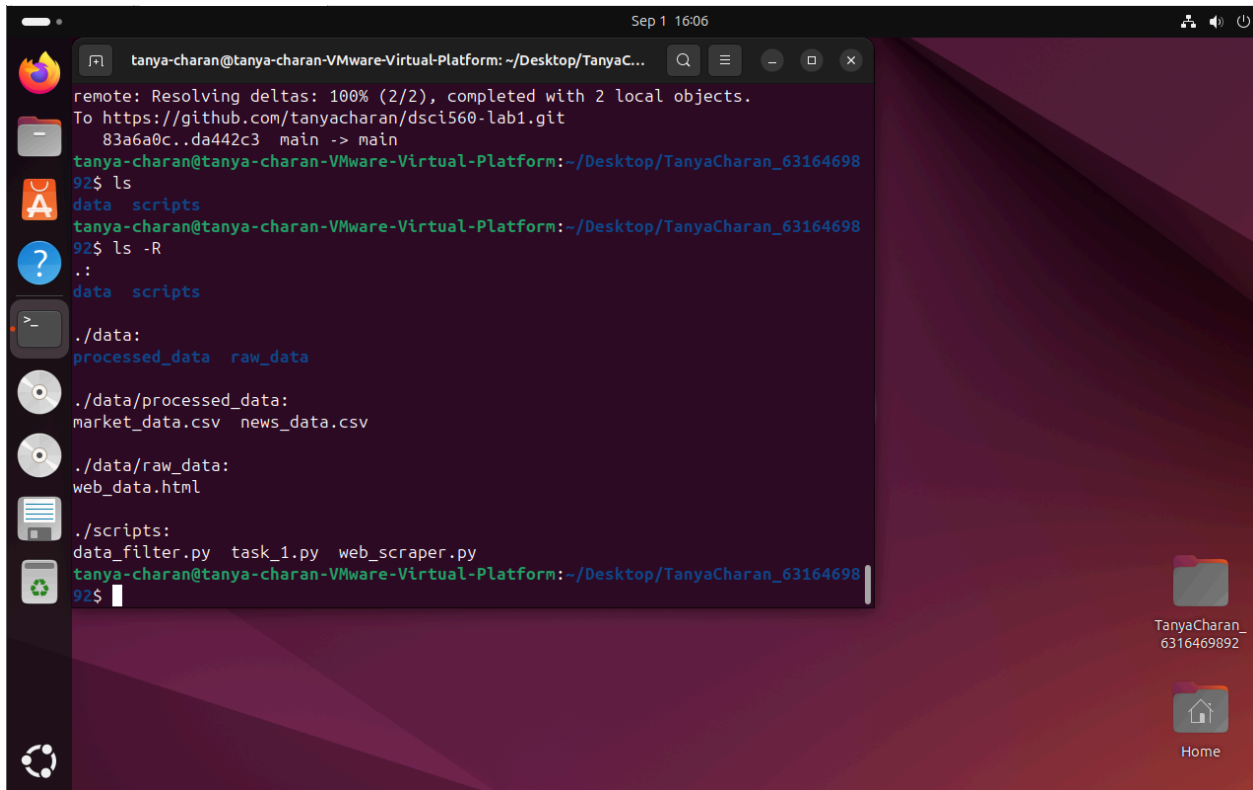
For the basic python script I was able to prompt a user to put in a name and then give out the greeting.

For the web scraping task I was able to use web_scraper.py to implement requests and BeautifulSoup to collect raw HTML from the CNBC website and then save that to an html file (web_data.html). Then I looked at the file contents to make sure there was information loaded in the html file.

For the data filtering task I had to create `data_filter.py`. For Latest News I saved the information to `news_data.csv`. I ran into an issue with the Market Banner. Though I wrote the code, I noticed that the website uses JavaScript for this information so the script I wrote does not extract the information how I wanted it to. To help debug and understand this issue in the script I included console messages and print statements in between each step.

Screenshots of my lab:

Here is the file hierarchy as shown in the terminal:

A screenshot of a terminal window on a Linux desktop. The terminal shows the output of several commands. At the top, it says 'remote: Resolving deltas: 100% (2/2), completed with 2 local objects.' followed by 'To https://github.com/tanyacharan/dsci560-lab1.git' and a commit hash. Then, the user runs 'ls' and 'ls -R'. The output shows a directory structure with 'data' and 'scripts' folders. The 'data' folder contains 'processed_data' and 'raw_data' subfolders. 'processed_data' contains 'market_data.csv' and 'news_data.csv'. 'raw_data' contains 'web_data.html'. The 'scripts' folder contains 'data_filter.py', 'task_1.py', and 'web_scraper.py'. The terminal window is titled 'tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaC...'. The desktop background is a dark purple gradient. On the left, there is a dock with icons for Firefox, a file manager, an application store, a help icon, a terminal, a CD icon, a folder icon, and a trash icon. On the right, there are desktop icons for a folder named 'TanyaCharan_6316469892' and a 'Home' icon.

```
remote: Resolving deltas: 100% (2/2), completed with 2 local objects.
To https://github.com/tanyacharan/dsci560-lab1.git
 83a6a0c..da442c3  main -> main
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92$ ls
data  scripts
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92$ ls -R
.:
data  scripts

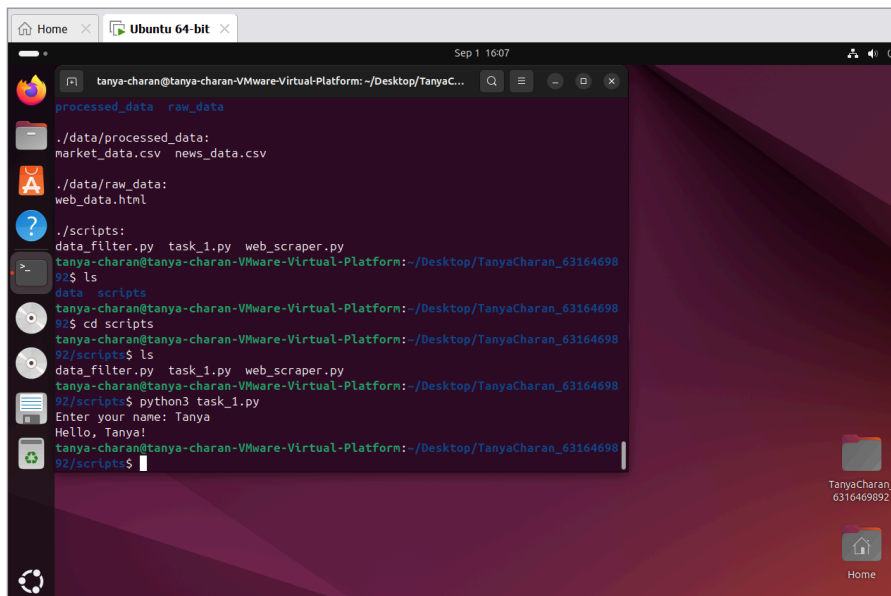
./data:
processed_data  raw_data

./data/processed_data:
market_data.csv  news_data.csv

./data/raw_data:
web_data.html

./scripts:
data_filter.py  task_1.py  web_scraper.py
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92$
```

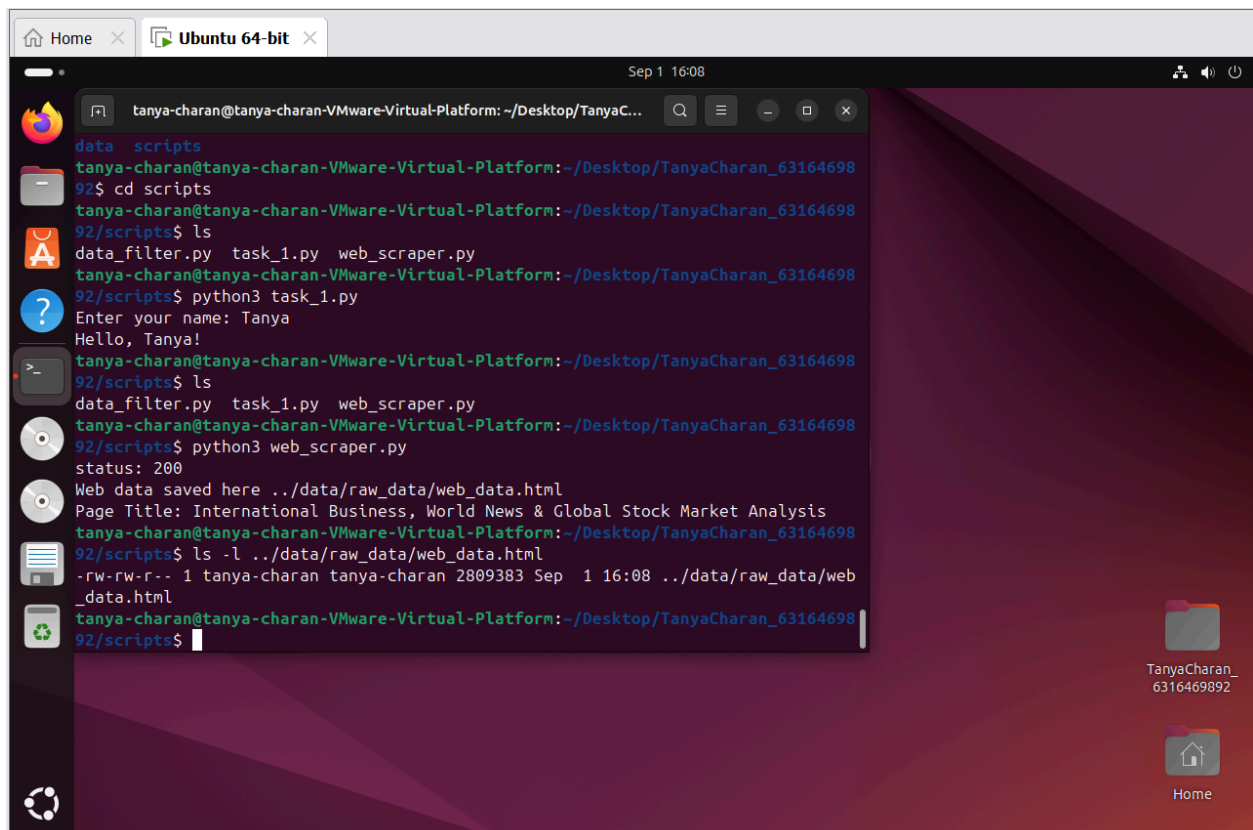
Here is an example of task_1.py running in the terminal:



A terminal window titled 'tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaC...' is shown. The terminal displays the following commands and output:

```
processed_data raw_data
./data/processed_data:
market_data.csv news_data.csv
./data/raw_data:
web_data.html
./scripts:
data_filter.py task_1.py web_scraper.py
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92$ ls
data scripts
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92$ cd scripts
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ ls
data_filter.py task_1.py web_scraper.py
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ python3 task_1.py
Enter your name: Tanya
Hello, Tanya!
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$
```

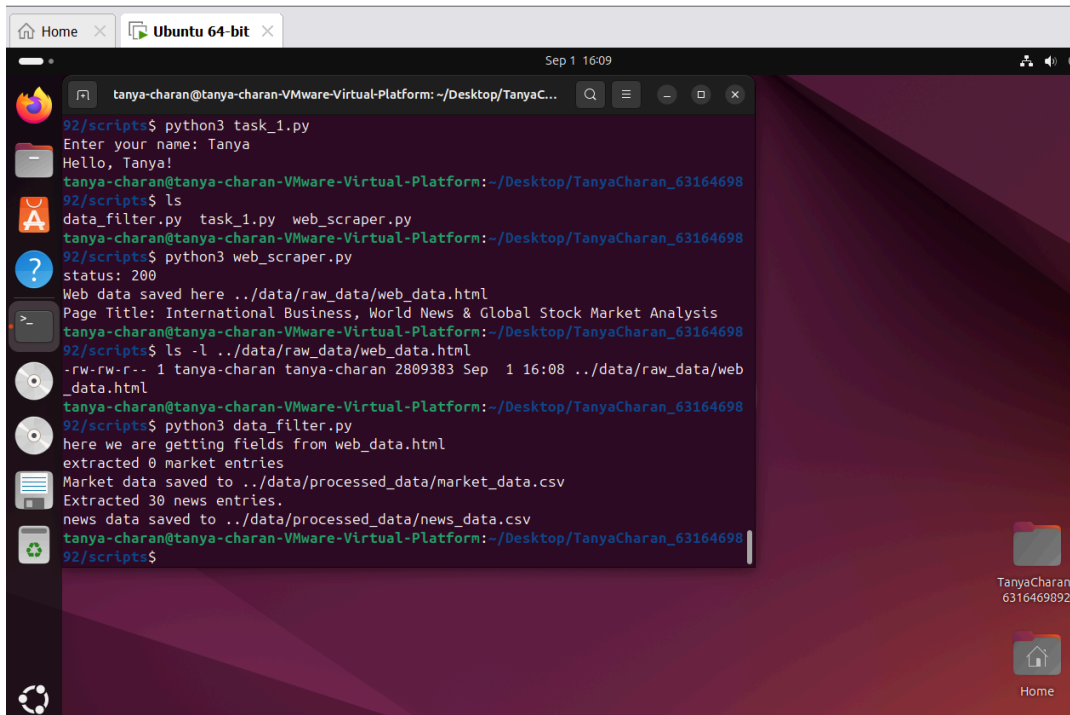
Here is an example of what running web_scraper.py looks like:



A terminal window titled 'tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaC...' is shown. The terminal displays the following commands and output:

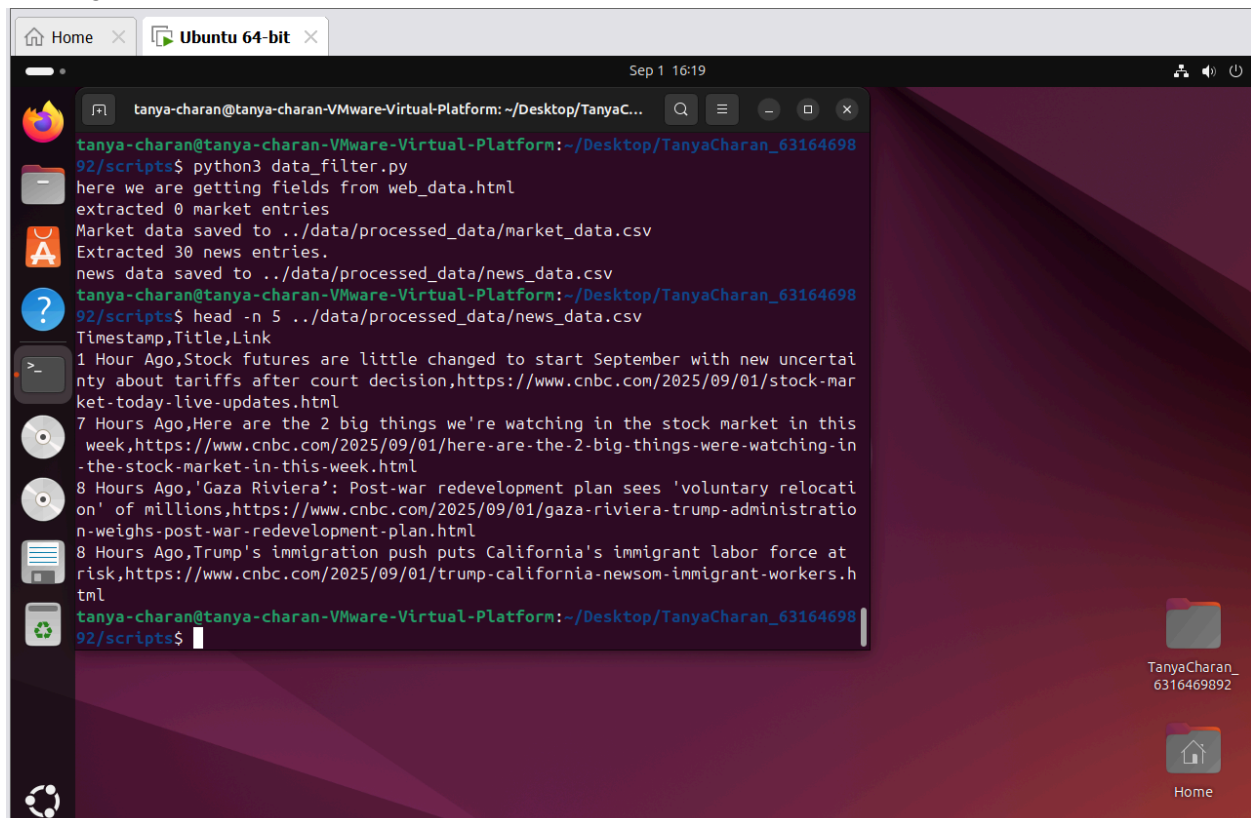
```
data scripts
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92$ cd scripts
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ ls
data_filter.py task_1.py web_scraper.py
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ python3 task_1.py
Enter your name: Tanya
Hello, Tanya!
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ ls
data_filter.py task_1.py web_scraper.py
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ python3 web_scraper.py
status: 200
Web data saved here ../data/raw_data/web_data.html
Page Title: International Business, World News & Global Stock Market Analysis
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$ ls -l ../data/raw_data/web_data.html
-rw-rw-r-- 1 tanya-charan tanya-charan 2809383 Sep  1 16:08 ../data/raw_data/web
_data.html
tanya-charan@tanya-charan-VMware-Virtual-Platform:~/Desktop/TanyaCharan_63164698
92/scripts$
```

Here I am running data_filter.py:



```
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaC...
92/scripts$ python3 task_1.py
Enter your name: Tanya
Hello, Tanya!
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$ ls
data_filter.py task_1.py web_scraper.py
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$ python3 web_scraper.py
status: 200
Web data saved here ../data/raw_data/web_data.html
Page Title: International Business, World News & Global Stock Market Analysis
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$ ls -l ../data/raw_data/web_data.html
-rw-rw-r-- 1 tanya-charan tanya-charan 2809383 Sep  1 16:08 ../data/raw_data/web
_data.html
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$ python3 data_filter.py
here we are getting fields from web_data.html
extracted 0 market entries
Market data saved to ../data/processed_data/market_data.csv
Extracted 30 news entries.
news data saved to ../data/processed_data/news_data.csv
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$
```

Looking at news_data.csv:



```
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaC...
92/scripts$ python3 data_filter.py
here we are getting fields from web_data.html
extracted 0 market entries
Market data saved to ../data/processed_data/market_data.csv
Extracted 30 news entries.
news data saved to ../data/processed_data/news_data.csv
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$ head -n 5 ../data/processed_data/news_data.csv
Timestamp,Title,Link
1 Hour Ago,Stock futures are little changed to start September with new uncertai
nty about tariffs after court decision,https://www.cnbc.com/2025/09/01/stock-mar
ket-today-live-updates.html
7 Hours Ago,Here are the 2 big things we're watching in the stock market in this
week,https://www.cnbc.com/2025/09/01/here-are-the-2-big-things-were-watching-in
-the-stock-market-in-this-week.html
8 Hours Ago,'Gaza Riviera': Post-war redevelopment plan sees 'voluntary relocati
on' of millions,https://www.cnbc.com/2025/09/01/gaza-riviera-trump-administratio
n-weighs-post-war-redevelopment-plan.html
8 Hours Ago,Trump's immigration push puts California's immigrant labor force at
risk,https://www.cnbc.com/2025/09/01/trump-california-newsom-immigrant-workers.h
tml
tanya-charan@tanya-charan-VMware-Virtual-Platform: ~/Desktop/TanyaCharan_63164698
92/scripts$
```