

Github link: <https://github.com/Howard-Pai/DSCI-560/tree/main/Lab1>

```
# raw_data ls
web_data.html
# raw_data head -n 10 web_data.html
<!DOCTYPE html>
<html itemscope="" itemtype="https://schema.org/WebPage" lang="en" prefix="og:https://ogp.me/ns#"\>
<head>
<meta content="website" property="og:type"\>
<meta content="International: Top News And Analysis" property="og:title"\>
<meta content="CNBC International is the world leader for news on business, technology, China, trade, oil prices, the Middle East and markets." property="og:description"\>
<meta content="https://www.cnbc.com/world/" property="og:url"\>
<meta content="CNBC" property="og:site_name"\>
<meta content="max-image-preview:large" name="robots"\>
<meta content="telephone=no" name="format-detection"\>
# raw_data cd ..
# data cd ..
# ChihYunPai_6777867067 cd s
cd: no such file or directory: s
# ChihYunPai_6777867067 cd scripts
# scripts python3 data_filter.py
Traceback (most recent call last):
  File "/home/shiro/ChihYunPai_6777867067/scripts/data_filter.py", line 3, in <module>
    from bs4 import BeautifulSoup
ModuleNotFoundError: No module named 'bs4'
# scripts source ..venv/bin/activate
source: no such file or directory: ..venv/bin/activate
# scripts ls
data_filter.py task_1.py venv web_scraper.py
# scripts source venv/bin/activate
(venv) # scripts python3 data_filter.py
Reading raw HTML file...
Filtering Market banner fields...
No Market data found!
Filtering Latest News fields...
Storing Latest News data (30 entries)...
CSV created: ../data/processed_data/news_data.csv
```

Not sure what “record of your commits and detailed commit description” means in the assignment description. But basically I run the code either on my end (for complex one like code) or on ubuntu(for others). The commits on GitHub are at the time that everything works perfectly.