

Name: Rongzi Xie

USC ID:9100096985

Github link and push records:

[https://github.com/Bananarya/DSCI560\\_lab1](https://github.com/Bananarya/DSCI560_lab1)

## Activity

All branches

All activity

All users

All time

Showing most recent first

update web\_scraper so it can scrape complete news data

Bananarya pushed 1 commit to main • 2d11595...acb8bf8 • 53 seconds ago

update web scraper, filter and data files

Bananarya created main • 2d11595 • 39 minutes ago

## 1.3 Install Python on Linux

```
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ sudo apt install python3
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
python3 is already the newest version (3.12.3-0ubuntu2).
python3 set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 21 not upgraded.
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ python3 --version
python3 --version: command not found
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ python3 --version
Python 3.12.3
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ mkdir rongzixie_9100096985
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ cd rongzixie_9100096985
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ mkdir data
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ mkdir scripts
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ cd scripts
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$ bbbbbbbbbb
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$
```

Above I installed python3 in ubuntu

## 2.2 Playing around with Linux Terminal

```
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ mkdir rongzixie_9100096985
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ cd rongzixie_9100096985
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ mkdir data
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ mkdir scripts
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ cd scripts
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$ bbbbbbbbbb
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$
```

```
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ cd Desktop
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop$ cd rongzixie_9100096985
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ ls
data  myenv  scripts
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985$ cd data
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/data$ ls
processed_data  raw_data
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/data$
```

Above screenshots showing my folder structure.

```

bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$ ls
data_filter.py  task_1.py  web_scraper.py
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$

```

Above showing the ls command and its results.

## 2.2 A basic Python Script

```

bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$ python3 task_1.py
please provide you name
Arya
Hello, Arya!

```

Above showing the input and result of my python script

## 2.3 Python Web-scraping Task

```

bananarya@bananarya-VMware-Virtual-Platform: ~/Desktop/rongzixie_91...
bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/data/raw_data$ head -n 10 web_data.html
<section class="MarketsBanner-container" data-analytics="HomePageInternational-marketsBanner-1-0" data-test="marketsBanner-0" id="HomePageInternational-MarketsBanner-1">
  <div class="MarketsBannerMenu-marketBannerMenuWrapper">
    <div class="CustomBannerAdder-container">
      <button class="CustomBannerAdder-addButton" data-testid="addButton">
        <span class="CustomBannerAdder-plusButton" data-testid="plusButton">
          <svg data-analytic-id="plus-icon" fill="none" height="17" viewBox="0 0 17 17" width="17" xmlns="http://www.w3.org/2000/svg">
            <path clip-rule="evenodd" d="M9.1427 7.43729L9.18188 1.1331C9.18391 0.807363 8.92149 0.544946 8.59576 0.546971C8.27003 0.548995 8.00433 0.814694 8.0023 1.14043L7.96312 7.44462L1.46316 7.48502C1.13743 7.48704 0.871726 7.75274 0.869702 8.07847C0.867677 8.40421 1.13009 8.66662 1.45583 8.6646L7.95579 8.6242L7.91539 15.1242C7.91337 15.4499 8.17579 15.7123 8.50152 15.7103C8.82725 15.7083 9.09295 15.4426 9.09497 15.1168L9.13537 8.61687L15.4396 8.57769C15.7653 8.57566 16.031 8.30997 16.033 7.98423C16.035 7.6585 15.7726 7.39608 15.4469 7.39811L9.1427 7.43729Z" fill="#005594" fill-rule="evenodd">
            </path>
          </svg>
        </span>
      </button>
    </div>
  </div>

```

Above is the first ten line of the scraped data

## 2.4 Data Filtering Task

```

(myenv) bananarya@bananarya-VMware-Virtual-Platform:~/Desktop/rongzixie_9100096985/scripts$ python3 data_filter.py
reading data from html
filtering market data
storing market data
filtering news data
storing news_data
success!!!

```

Terminal output lines

```
Open  ~
market_data.csv
~/Desktop/frongzixie_9100096985/data/processed_data

market_data.csv
news_data.csv

,symbol,stockPosition,changesPct
0,DJIA,"45,636.90",+0.16%
1,S&P 500,"6,501.86",+0.32%
2,NASDAQ,"21,705.158",+0.53%
3,RUSS 2K*, "2,378.413",+0.19%
4,VIX,14.43,-2.83%
```

markets data scraped and filtered.

```
Open  ~
news_data.csv
~/Desktop/DSCI560_lab1/frongzixie_9100096985/data/processed_data

,titles,timestamp,link
0,Judge: Google gets to keep Chrome but is barred from exclusive search deals,38 Min Ago,https://www.cnbc.com/2025/09/02/google-antitrust-search-ruling.html
1,"Santoli's Tuesday market wrap-up: Fresh challenges to prevailing, bullish beliefs",47 Min Ago,https://www.cnbc.com/2025/09/02/santolis-tuesday-market-wrap-up-challenges-to-prevailing-bullish-beliefs.html
2,Trump will ask Supreme Court for 'expedited ruling' on tariffs appeal,2 Hours Ago,https://www.cnbc.com/2025/09/02/trump-trade-supreme-court-tariffs-appeal.html
3,"Investors can grab double-digit yields with these buy-rated funds, BofA says",2 Hours Ago,https://www.cnbc.com/2025/09/02/grab-double-digit-yields-with-these-buy-rated-funds-bank-of-america-says.html
4,Airplane leasing world shrinks with $7.4 billion takeover of Air Lease,2 Hours Ago,https://www.cnbc.com/2025/09/02/airplane-leasing-air-lease-takeover.html
5,"OpenAI acquires Statsig for $1.1 billion, brings on CEO as applications executive",2 Hours Ago,https://www.cnbc.com/2025/09/02/openai-buys-statsig-for-1point1-billion-hires-ceo-as-applications-exec.html
6,"Harvard expert shares his 1 tip for living a happier, more successful life",2 Hours Ago,https://www.cnbc.com/2025/09/02/harvard-expert-arthur-brooks-top-tip-for-happier-more-successful-life.html
7,"The best and worst in Tuesday's down stock market – plus, Nvidia quashes rumors",2 Hours Ago,https://www.cnbc.com/2025/09/02/the-best-and-worst-in-tuesdays-down-stock-market-plus-nvidia-quashes-rumors.html
8,Goldman adds Walmart to September 'conviction list.' Here's who else made it,3 Hours Ago,https://www.cnbc.com/2025/09/02/goldman-adds-walmart-to-september-conviction-list-who-else-made-it.html
9,The hidden gem of Home Depot's over $22 billion of deals to boost its pro sales,4 Hours Ago,https://www.cnbc.com/2025/09/02/the-hidden-gem-of-home-depots-over-22-billion-of-deals-to-boost-its-pro-sales.html
10,What Trump court loss means for billions in tariffs paid to government,4 Hours Ago,https://www.cnbc.com/2025/09/02/what-trump-court-loss-means-for-billions-in-tariffs-paid-to-government.html
11,"Stocks making biggest moves midday: Frontier Group, Kraft Heinz, Biogen & more",4 Hours Ago,https://www.cnbc.com/2025/09/02/stocks-making-the-biggest-midday-moves-ulcc-khc-pep-biib.html
```

newsdata scraped and filtered