DSCI 560 Lab 1

Vincent-Daniel Yun
ID: 4463771151
Email: yunjuyou@usc.edu
https://github.com/YUNBLAK/usc-dsci560-sp26
Youtube Link: https://youtu.be/sinnfKtAPZE

1. Installation and Setup

I registered at Broadcom and then downloaded and installed Vmware. And I downloaded Ubuntu ISO Image.

1.1. Register at Broadcom and then download and install VMware

1.2. Download Ubuntu ISO Image

1.3. Install Python on Linux
I opened a terminal in my ubuntu virtual machine. First of all, I updated the packages list using: **sudo apt update.**

I verified that Python 3 is installed successfully.

Moreover, I verified that pip is installed successfully.

2. Get Familiar with Linux and Python

2.1. Playing around with Linux Terminal

I opened the Linux terminal and created a new directory named daniel_4463771151. Inside this folder, there are two subdirectories named "data" and "scripts". I created an empty Python file inside the scripts folder named "task_1.py.

```
(base) daniel@daniel-AORUS-17H-BXF:~$ mkdir daniel_4463771151
(base) daniel@daniel-AORUS-17H-BXF:~$ cd daniel_4463771151/
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ mkdir data
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ mkdir scripts
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ touch scripts/task_1.py
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ ls
data  scripts
```

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ cd scripts/
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$ ls
task_1.py
```

I use the list command "ls" to view the created script file.

2.2. A basic Python Script

By using nano I opened the task_1.py Python file I'd created. I wrote a simple Python script that reads a name as input and says hello.

```
  GNU nano 6.2                                    task_1.py
name = input("Enter your name: ")
print(f"Hello, {name}!")
```

I saved and ran the Python code.

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$ nano task_1.py
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$ python3 task_1.py
Enter your name: Daniel Yun
Hello, Daniel Yun!
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$
```

2.3. Python Web-scraping Task

I created a new file named "web_scraper.py" in the scripts folder.

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$ touch web_scraper.py
```

Install the required libraries (Requests and BeautifulSoup4) using pip

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$ pip install requests beautifulsoup4
Requirement already satisfied: requests in /home/daniel/anaconda3/lib/python3.13/site-packages (2.32
.3)
Requirement already satisfied: beautifulsoup4 in /home/daniel/anaconda3/lib/python3.13/site-packages
 (4.12.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/daniel/anaconda3/lib/python3.13/sit
e-packages (from requests) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /home/daniel/anaconda3/lib/python3.13/site-packages (
from requests) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /home/daniel/anaconda3/lib/python3.13/site-pack
ages (from requests) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /home/daniel/anaconda3/lib/python3.13/site-pack
ages (from requests) (2025.4.26)
Requirement already satisfied: soupsieve>1.2 in /home/daniel/anaconda3/lib/python3.13/site-packages
(from beautifulsoup4) (2.5)
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/scripts$
```

I opened this website https://www.cnbc.com/world/?region=world. And I tried to analyze the HTML structure on the Page.

I created two new folders called "raw_data" and "processed_data" in "data" folder
I wrote a Python script that uses Requests and BeautifulSoup to collect data from the link.

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ ls
data  scripts
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ mkdir -p data/raw_data data/processed_data
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ ls
data  scripts
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ cd data
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/data$ ls
processed_data  raw_data
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/data$ pwd
/home/daniel/daniel_4463771151/data
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/data$
```

My script saves the collected data.

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ python3 scripts/web_scraper.py
Saved: data/raw_data/web_data.html
```

By using the terminal, I printed the first 10 lines of the created html file.

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ python3 scripts/web_scraper.py
Saved: data/raw_data/web_data.html
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ head -n 10 data/raw_data/web_data.html
<section class="MarketsBanner-container" data-analytics="HomePageInternational-marketsBanner-1-0" data-test="marke
tsBanner-0" id="HomePageInternational-MarketsBanner-1">
 <div class="MarketsBannerMenu-marketBannerMenuWrapper">
  <div class="CustomBannerAdder-container">
   <button aria-label="Add markets banner" class="CustomBannerAdder-addButton" data-testid="addButton">
    <span aria-hidden="true" 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 0.5469
71C8.27003 0.548995 8.00433 0.814694 8.0023 1.14043L7.96312 7.44462L1.46316 7.48502C1.13743 7.48704 0.871726 7.752
74 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 8.57769C1
5.7653 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>
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$
```

2.4. Data Filtering Task

I wrote a python script called "data_filter.py" reading the "web_data.html" file into a for extracting data.

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$ python3 scripts/data_filter.py
Reading HTML ...
Filtering fields ...
Storing Market data ...
CSV created: data/processed_data/market_data.csv rows= 5
Storing News data ...
CSV created: data/processed_data/news_data.csv rows= 30
Done.
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151$
```

My script stores the (marketCard_symbol, marketCard_stockPosition and marketCard-changePct) from the market banner and the (LatestNews-timestamp, title and link).

market_data.csv:

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/data/processed_data$ cat market_data.csv
symbol,stock_position,change
STOXX600*,614.38,-0.19
DAX*,25297.13,-55.26
FTSE*,10235.29,-3.65
CAC*,8258.94,-54.18
FTSE MIB*,45799.69,-50.08
```

news_data.csv

```
(base) daniel@daniel-AORUS-17H-BXF:~/daniel_4463771151/data/processed_data$ cat news_data.csv
timestamp,title,link
11 Hours Ago,Week in review: Stocks battled a flood of news and we booked some profits,https://www.cnbc.com/2026/01/17/week-in-review-stocks-battled-a-flood-
12 Hours Ago,Trump threatens to sue JPMorgan Chase for 'debanking' him,https://www.cnbc.com/2026/01/17/trump-jpmorgan-chase-debanking.html
14 Hours Ago,Trump: NATO members to face tariffs up to 25% until a Greenland deal is struck,https://www.cnbc.com/2026/01/17/trump-greenland-tariffs-nato.html
15 Hours Ago,"Led by Texas, states race to prove they can put bitcoin on public balance sheet",https://www.cnbc.com/2026/01/17/texas-us-states-budgets-bitco
17 Hours Ago,Unshaken: Why Brazilian stocks have looked past the Venezuela attack,https://www.cnbc.com/2026/01/17/unshaken-why-brazilian-stocks-have-looked-p
17 Hours Ago,Bestselling author: How to create better habits without relying on discipline,https://www.cnbc.com/2026/01/17/james-clear-how-to-create-better-h
17 Hours Ago,"Warren Buffett: To maximize your potential, ask yourself this question",https://www.cnbc.com/2026/01/17/warren-buffett-to-maximize-your-potenti
17 Hours Ago,"Buy these five stocks ahead of earnings, Bank of America says",https://www.cnbc.com/2026/01/17/stocks-to-buy-ahead-of-earnings-bank-of-america-
17 Hours Ago,This week's most overbought names include Darden Restaurants and Target,https://www.cnbc.com/2026/01/17/this-weeks-most-overbought-names-include
17 Hours Ago,"Buffett on parenting, giving up horse betting and why he stopped talking politics",https://www.cnbc.com/2026/01/17/warren-buffett-on-parenting-
itics.html
17 Hours Ago,"Unexpected expenses take 10% of retirees' income, on average, research shows",https://www.cnbc.com/2026/01/17/retirees-emergency-savings.html
18 Hours Ago,Disney dominated the 2025 box office. Here's how it could keep the crown in 2026,https://www.cnbc.com/2026/01/17/disney-dominated-2025-box-offic
21 Hours Ago,The founders of billion-dollar AI startups are getting younger — here's why,https://www.cnbc.com/2026/01/17/billion-dollar-ai-startup-founders-a
"January 16, 2026",Elon Musk's xAI faces tougher road building data centers after EPA rule update,https://www.cnbc.com/2026/01/16/musks-xai-faces-tougher-roa
ml
"January 16, 2026",Here's why Jim Cramer thinks chip stocks can go higher,https://www.cnbc.com/2026/01/16/heres-why-jim-cramer-thinks-chip-stocks-can-go-high
```

3. Team Formation

Done!

Name	Group	
Dylan Chen STUDENT dylanach@usc.edu	Group 1	✉
Angela Kang STUDENT kangange@usc.edu	Group 1	✉
Vincent-Daniel Yun STUDENT yunjuyou@usc.edu	Group 1	✉