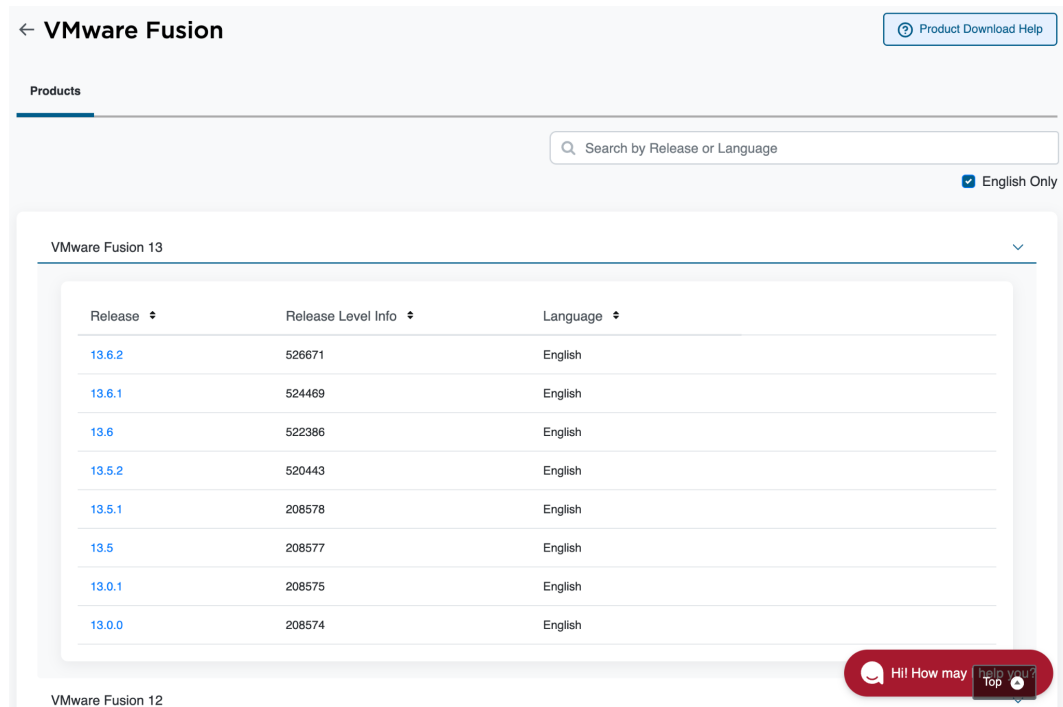
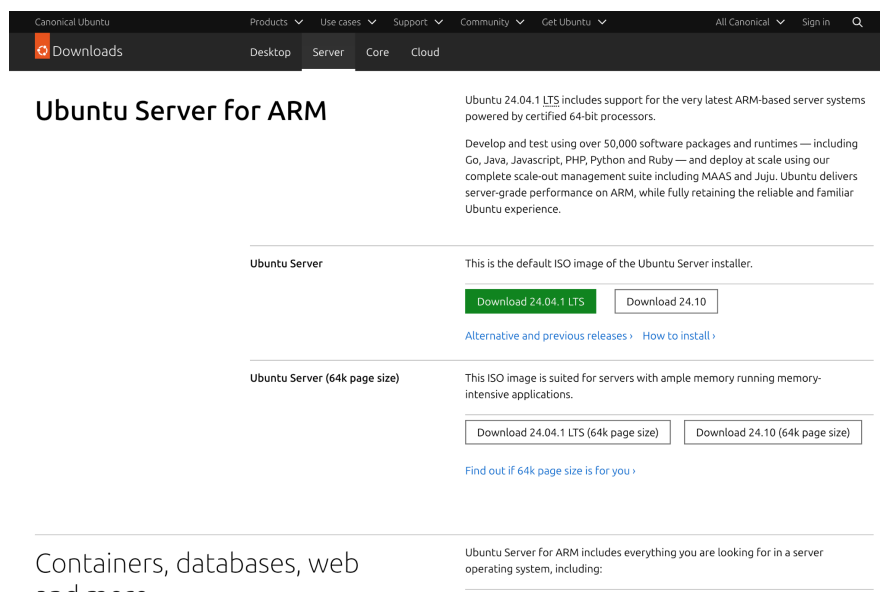


Name: Jiaying (Ashley) Wang
USCID: 3018-1829-67

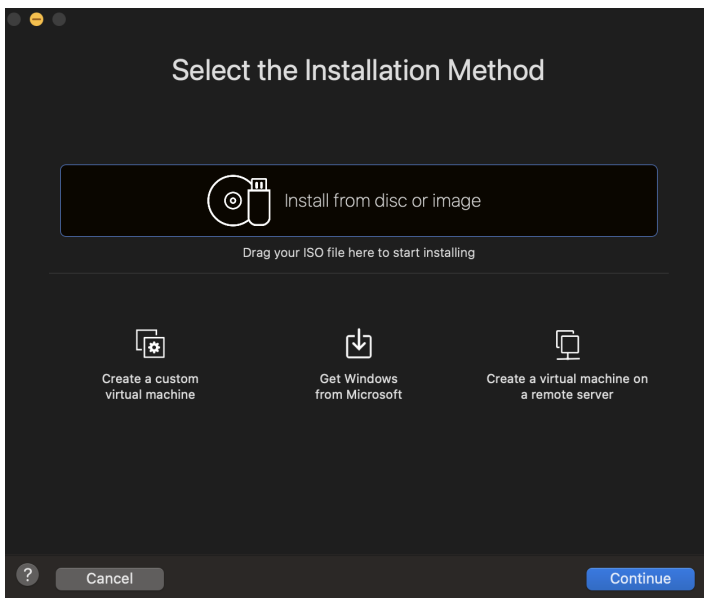
1.1 installed **VMware Fusion** instead of Workstation because it's designed for macOS. personal use isn't explicitly labeled in the dropdown, and since the personal-use and commercial-use editions are identical, I downloaded the latest version, **13.6.2**.



1.2
Downloaded Ubuntu designed for ARM architecture from <https://ubuntu.com/download/server/arm> since I have M3 chip.



Dragged the **Ubuntu ISO** file I downloaded earlier into the box.



```
Ubuntu 24.04.1 LTS ashley-server tty1
ashley-server login: ashley
Password:
Welcome to Ubuntu 24.04.1 LTS (GNU/Linux 6.8.0-51-generic aarch64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/pro

System information as of Fri Jan 17 11:21:32 PM UTC 2025

System load:  0.0           Temperature:   11758.9 C
Usage of /:   42.9% of 9.75GB Processes:      202
Memory usage: 5%           Users logged in: 0
Swap usage:   0%           IPv4 address for ens160: 192.168.86.129

Expanded Security Maintenance for Applications is not enabled.

77 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The programs included with the Ubuntu system are free software:
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ashley@ashley-server:~$
```

1.3 Python 3 and pip installed on virtual machine

```
ashley@ashley-server:~$ python3 --version
Python 3.12.3
ashley@ashley-server:~$ pip3 --version
pip 24.0 from /usr/lib/python3/dist-packages/pip (python 3.12)
```

1.4

2.1

Created a directory called Desktop, Create a new directory named "<your name>_<your USC ID>" on the desktop. Inside the folder, create two subdirectories named "data" and "scripts" Create an empty Python file inside the scripts folder named "task_1.py"

```
ashley@ashley-server:~$ pwd
/home/ashley
ashley@ashley-server:~$ mkdir Desktop
ashley@ashley-server:~$ cd Desktop
ashley@ashley-server:~/Desktop$ mkdir ashley_3018182967
ashley@ashley-server:~/Desktop$ cd ashley_3018182967/
ashley@ashley-server:~/Desktop/ashley_3018182967$ mkdir data scripts
ashley@ashley-server:~/Desktop/ashley_3018182967$ cd scripts
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ touch task_1.py
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ ls
task_1.py
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ cd ~/Desktop/ashley_3018182967/
ashley@ashley-server:~/Desktop/ashley_3018182967$ ls
data  scripts
```

2.2

Wrote task_1.py that reads user's name as input and greets the user with "Hello, [name]!".

```
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ python3 task_1.py
Enter your name: Ashley
Hello, Ashley!
```

2.3

Pip install encountered "Python's environment is "externally managed error".

To resolve this issue, I Installed the Packages System-Wide Using apt

```
sudo apt install python3-requests python3-bs4
```

```
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ pip install requests beautifulsoup4
error: externally-managed-environment

× This environment is externally managed
╰─> To install Python packages system-wide, try apt install
    python3-xyz, where xyz is the package you are trying to
    install.

    If you wish to install a non-Debian-packaged Python package,
    create a virtual environment using python3 -m venv path/to/venv.
    Then use path/to/venv/bin/python and path/to/venv/bin/pip. Make
    sure you have python3-full installed.

    If you wish to install a non-Debian packaged Python application,
    it may be easiest to use pipx install xyz, which will manage a
    virtual environment for you. Make sure you have pipx installed.

    See /usr/share/doc/python3.12/README.venv for more information.

note: If you believe this is a mistake, please contact your Python installation or OS distrib
Python installation or OS, by passing --break-system-packages.
hint: See PEP 668 for the detailed specification.
```

Web scraper.py uses Selenium to load the webpage, waits for market banner to become visible, extracts relevant data sections with BeautifulSoup (e.g., market banner and latest news), and saves them into a structured HTML file for further use.

```
GNU nano 7.2 web_scraper.py *
from bs4 import BeautifulSoup
# from requests.html import HTMLSession
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import requests

chrome_options = Options()
chrome_options.add_argument("--headless")
chrome_options.add_argument("--no-sandbox")
chrome_options.add_argument("--disable-dev-shm-usage")
chrome_options.add_argument("--user-data-dir=/tmp/selenium_user_data")

service = Service('/usr/bin/chromedriver')
driver = webdriver.Chrome(service=service, options=chrome_options)
# session = HTMLSession()
url = "https://www.cnbc.com/world?region=world"
# session.browser_args = {'executablePath': '/usr/bin/chromium-browser'}
# r = session.get(url)
# r.html.render()
# html = r.html.raw_html
# soup = BeautifulSoup(html, "html.parser")
driver.get(url)

WebDriverWait(driver, 30).until(EC.visibility_of_element_located((By.CLASS_NAME, "MarketCard-row")))
soup = BeautifulSoup(driver.page_source, "html.parser")
driver.quit()

market_banner = soup.find("div", class_="MarketsBanner-marketData").prettify()

resp = requests.get(url)
latest_soup = BeautifulSoup(resp.text, "html.parser")
latest_news = latest_soup.find("ul", class_="LatestNews-list").prettify()

raw_data_path = "../data/raw_data/web_data.html"
with open(raw_data_path, "w", encoding='utf-8') as file:
    # file.write(soup.prettify())
    file.write(market_banner)
    file.write("\n\n")
    file.write(latest_news)
```

Created two new folders in the “data” folder called “raw_data” and “processed_data”.

```
(myenv) ashley@ashley-server:~/Desktop/ashley_3018182967/data$ ls
processed_data  raw_data
```

Saved the collected data in the “raw_data” folder to a file named “web_data.html”. Used the terminal print the first 10 lines of the created html file on the terminal.

```
(myenv) ashley@ashley-server:~/Desktop/ashley_3018182967/data/raw_data$ ls
web_data.html
(myenv) ashley@ashley-server:~/Desktop/ashley_3018182967/data/raw_data$ head -n 10 web_data.html
<div class="MarketsBanner-marketData" id="market-data-scroll-container">
  <a class="MarketCard-container MarketCard-up MarketCard-wrap" href="//www.cnbc.com/quotes/.DJI">
    <div class="MarketCard-row">
      <span class="MarketCard-symbol">
        DJIA
      </span>
      <span class="MarketCard-stockPosition">
        43,487.83
      </span>
    </div>
```

2.4

data_filter.py reads the saved HTML file, uses BeautifulSoup to parse and extract market data and news data, and then writes the processed information into two separate CSV files for further analysis or storage.

```
GNU nano 7.2 data_filter.py
from bs4 import BeautifulSoup
import csv

with open("../data/raw_data/web_data.html", "r") as file:
    soup = BeautifulSoup(file, "html.parser")

print("Filtering market banner data...")
market_data = []
market_banner = soup.find('div', class_ = "MarketsBanner-marketData")
market_cards = market_banner.find_all("a", class_ = "MarketCard-container")
#market_cards = soup.select('a.MarketCard-container')
# print(market_cards)

for card in market_cards:
    symbol = card.find("span", class_ = "MarketCard-symbol").text
    stock_position = card.find("span", class_ = "MarketCard-stockPosition").text
    change_pct = card.find("div", class_ = "MarketCard-changeData").find("span", class_ = "MarketCard-changesPct").text
    market_data.append([symbol, stock_position, change_pct])
    # print(market_data)

print("Filtering latest news data...")
latest_news_data = []
latest_news_list = soup.find("ul", class_ = "LatestNews-list")
news_items = latest_news_list.find_all("li")

for item in news_items:
    timestamp = item.find("time", class_ = "LatestNews-timestamp").text.strip()
    title_tag = item.find("a", class_ = "LatestNews-headline")
    title = title_tag.text.strip()
    link = title_tag["href"]
    latest_news_data.append([timestamp, title, link])

print("Storing Market Data...")
with open("../data/processed_data/market_data.csv", "w", newline = "") as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(["marketCard_symbol", "marketCard_stockPosition", "marketCard_changePct"])
    writer.writerows(market_data)
print("Market data CSV created")

print("Storing Latest News Data...")
with open("../data/processed_data/news_data.csv", "w", newline = "") as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(["LatestNews-timestamp", "Title", "Link"])
    writer.writerows(latest_news_data)
```

Print appropriate messages to the console

```
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ ls
data_filter.py task_1.py web_scraper.py
ashley@ashley-server:~/Desktop/ashley_3018182967/scripts$ python3 data_filter.py
Filtering market banner data...
Filtering latest news data...
Storing Market Data...
Market data CSV created
Storing Latest News Data...
Latest News Data CSV created
```

Wrote to 2 csv files.

DSCI-560-Lab1 / data / processed_data / **news_data.csv**

jiaingwang117 scraped 70b65a0 · 28 minutes ago History

Preview Code Blame 31 lines (31 loc) · 5.61 KB Code 55% faster with GitHub Copilot Raw Copy Download Edit

Search this file

	LatestNews-timestamp	Title	Link
1	7 Hours Ago	Perplexity AI makes a bid to merge with TikTok U.S.	https://www.cnbc.com/2025/01/18/perplexity-ai-makes-a-bid-to-me
2	10 Hours Ago	Solana surges 12% on launch of Trump-themed meme coin, ether falls	https://www.cnbc.com/2025/01/18/crypto-market-today.html
3	11 Hours Ago	What to expect from travel prices in 2025, and which spots have the best deals	https://www.cnbc.com/2025/01/18/what-to-expect-from-travel-pric
4	12 Hours Ago	Consumer protection agencies at risk in Trump's second term: What it means for you	https://www.cnbc.com/2025/01/18/how-trumps-second-term-could
5	13 Hours Ago	Why the gold boom is causing a surge in illegal mining	https://www.cnbc.com/2025/01/18/why-the-gold-boom-is-causing-
6	13 Hours Ago	Google Maps is turning 20 — mapping more countries and adding AI capabilities	https://www.cnbc.com/2025/01/18/google-maps-turns-20-adds-ai-
7	14 Hours Ago	Trump inauguration trades: The sectors that could win in the early days	https://www.cnbc.com/2025/01/18/trump-inauguration-trades-the-s
8	14 Hours Ago	'Physical AI' is the next frontier for the tech cycle. How to play it	https://www.cnbc.com/2025/01/18/is-physical-ai-the-next-growth-p
9	14 Hours Ago	GLP-1s and Brian Thompson's killing loom large at top health-care conference	https://www.cnbc.com/2025/01/18/glp-1s-brian-thomson-killing-loo
10	14 Hours Ago	Bank of America says these stocks are table-pounding buys ahead of earnings	https://www.cnbc.com/2025/01/18/stocks-with-upside-bank-of-ame
11	15 Hours Ago	TikTok creators post farewell videos to their fans ahead of expected U.S. ban	https://www.cnbc.com/2025/01/18/tiktok-creators-post-farewell-vid
12	January 17, 2025	TikTok says it will go dark on Sunday unless Biden intervenes; White House calls it a 'stunt'	https://www.cnbc.com/2025/01/17/tiktok-says-it-will-go-dark-on-su
13	January 17, 2025	Cramer's Lightning Round: Adobe is a buy	https://www.cnbc.com/2025/01/17/cramers-lightning-round-adobe-
14	January 17, 2025	Jim Cramer parses big bank earnings, says stocks have 'quite a bit more upside'	https://www.cnbc.com/2025/01/17/jim-cramer-parses-bank-earning
15	January 17, 2025	Cramer's week ahead: Inauguration, Procter & Gamble, American Express earnings	https://www.cnbc.com/2025/01/17/cramers-week-ahead-inauguratic
16	January 17, 2025	DOJ sues Walgreens, alleging it filled millions of unlawful prescriptions	https://www.cnbc.com/2025/01/17/doj-sues-walgreens-prescription
17			

DSCI-560-Lab1 / data / processed_data / **market_data.csv**

jiaingwang117 scraped

Preview Code Blame 36 lines (36 loc) · 357 Bytes Code 55% faster with GitHub Copilot

Search this file

	marketCard_symbol	marketCard_stockPosition	marketCard_changePct
1	DJIA	43,487.83	+0.78%
2	S&P 500	5,996.66	+1.00%
3	NASDAQ	19,630.20	+1.51%
4	RUSS 2K*	2,275.88	+0.40%
5	VIX	15.97	-3.80%
6			

Push to GitHub workflow

```
git status
```

```
git add .
```

```
git commit -m ""
```

```
git push origin main
```