Lab 1 Solution

# Installation & Setup

* Development done locally in a local git repository on a windows machine.
* The code is synced with a remote git repository.

A screenshot of a computer

Description automatically generated

* A Linux Ubuntu instance is spin up using GitHub Codespace for the repository.

A screenshot of a computer

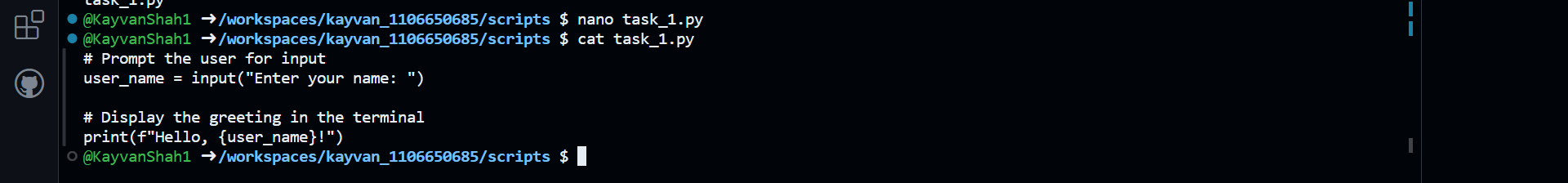
Description automatically generated

# Get Familiar with Linux and Python

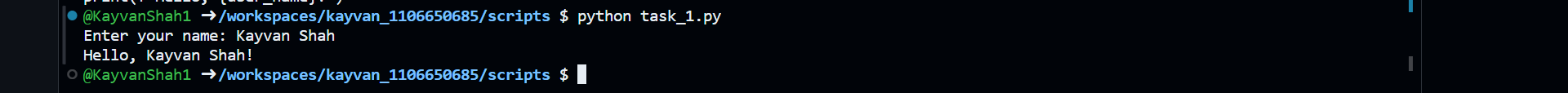
## Playing around with Linux Terminal

A screenshot of a computer

Description automatically generated



## A basic Python Script



## Web Scraping Task

* Create a virtual environment and install the dependencies using the ***requirements.txt***.
  + *venv* – Created a virtual environment using command
  + Linux has a built-in support for Chrome and using a library **chromedriver-py** we automatically manage the webdriver downloads and retrieval of the executable path while initializing the web driver.

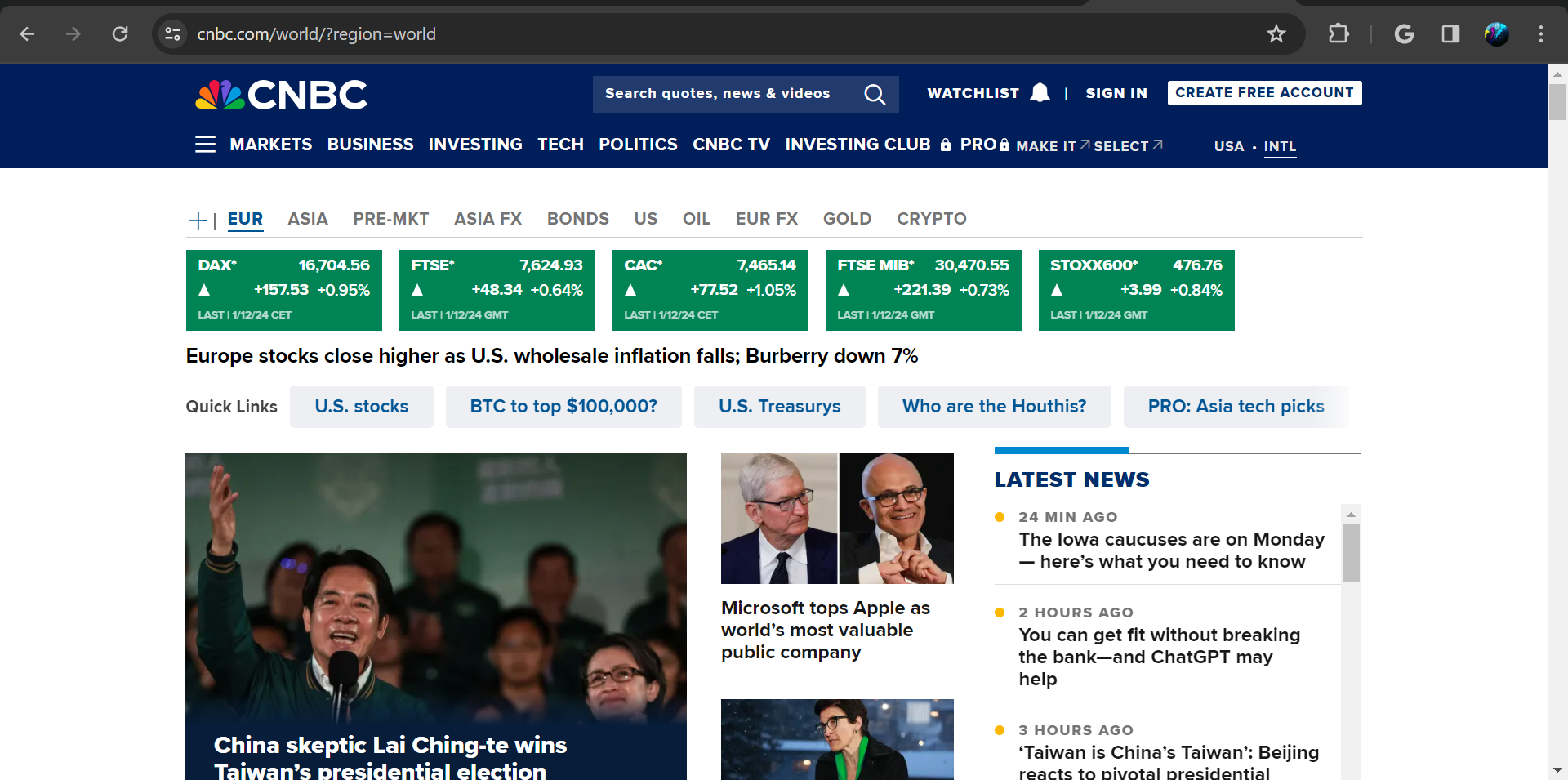
A screen shot of a computer

Description automatically generated

A computer screen shot of a program code

Description automatically generated

* A snapshot of the webpage from where the data is to be scraped.
  + The webpage loads a few fields dynamically. In this case the market banner is loaded using a JavaScript.
  + To capture the data from dynamically populated HTML elements we must use selenium with a web driver suited for the installed browser on the system.
  + While the latest news isn’t loaded dynamically hence, we can get the static snapshot of the HTML tree of the scrape that its section’s data.

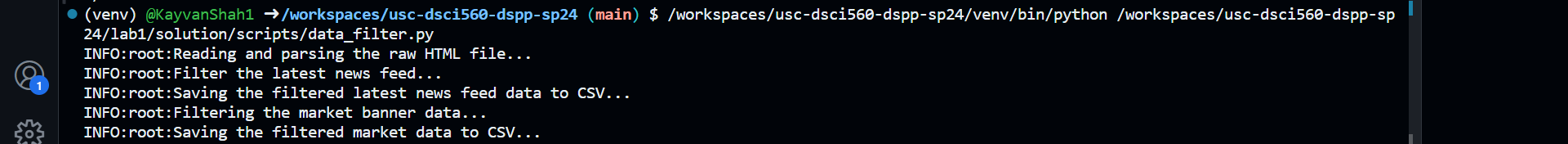


* Scraping the raw data

A black screen with white text

Description automatically generated

* Filtering the data to extract the fields of interest.



## Output Files

* Raw data output HTML file

A screenshot of a computer program

Description automatically generated

* Processed data CSV files
  + Market data

A screenshot of a computer program

Description automatically generated

* + Latest news

A screen shot of a computer

Description automatically generated