Yifan Yang
8386626967

Data Science Professional Practicum (DSCI 560)
Laboratory Assignment 1

Summary of Lab 1

This experiment completed the initial setup and installation tasks of using Linux as the operating system and Python as the programming language, and learned some basic knowledge of the operating system and programming language. In the experiment, the Ubuntu Linux VM was successfully installed and the required Python software packages were installed. Afterwards, Python tasks such as data crawling were performed. Experiment 1 aims to let us learn to use virtual machines, clips, terminals, etc. to complete tasks such as data science, which meets the needs of the enterprise.

1. Installation and Setup

1.1. Install VirtualBox/VMware

Installing VirtualBox/VMware is a very simple step. VMware has been installed on my computer before, and it has a Chinese version interface. It should be noted that VMware requires a certificate.

1.2. Download Ubuntu ISO Image

Go directly to the official website to download the Ubuntu ISO image, then create a new virtual machine on the VM, install and set up the virtual machine to complete the entire operation. It should be noted that in the display settings, the 3D accelerator needs to be turned off. In addition, you need to install the tutorial to set up important information such as memory.

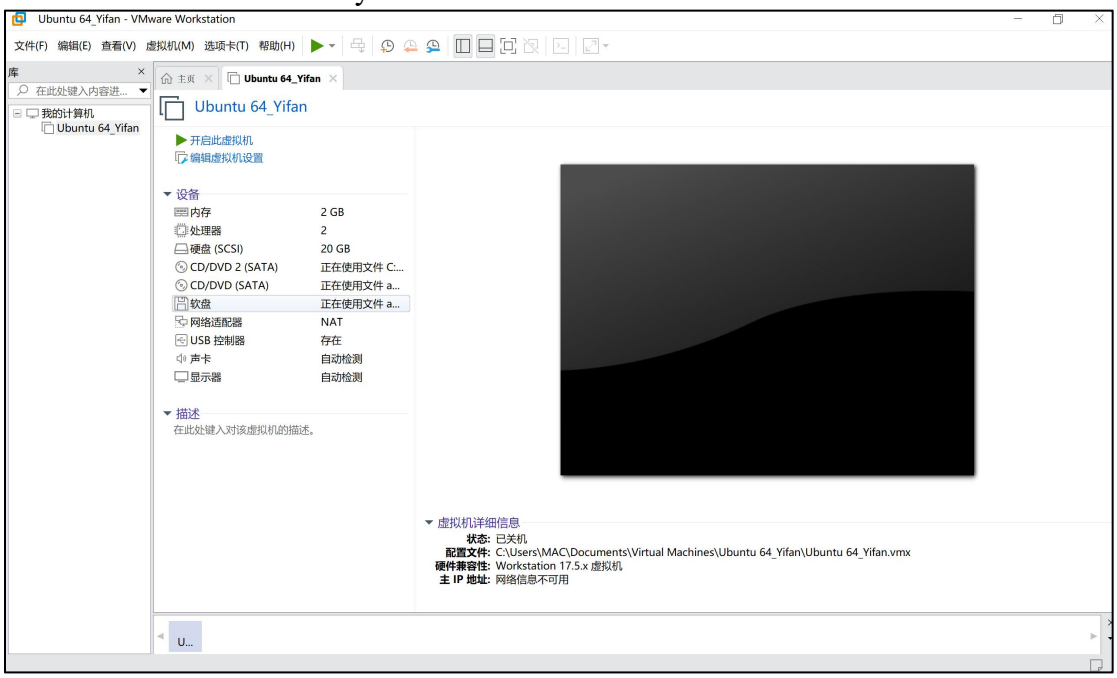


Image 1: Successfully installed VM and configured Ubuntu interface

1.3. Install Python on Linux

To install Python on Linux, you only need to follow the steps. Only the version number shown here means that we have successfully installed it.

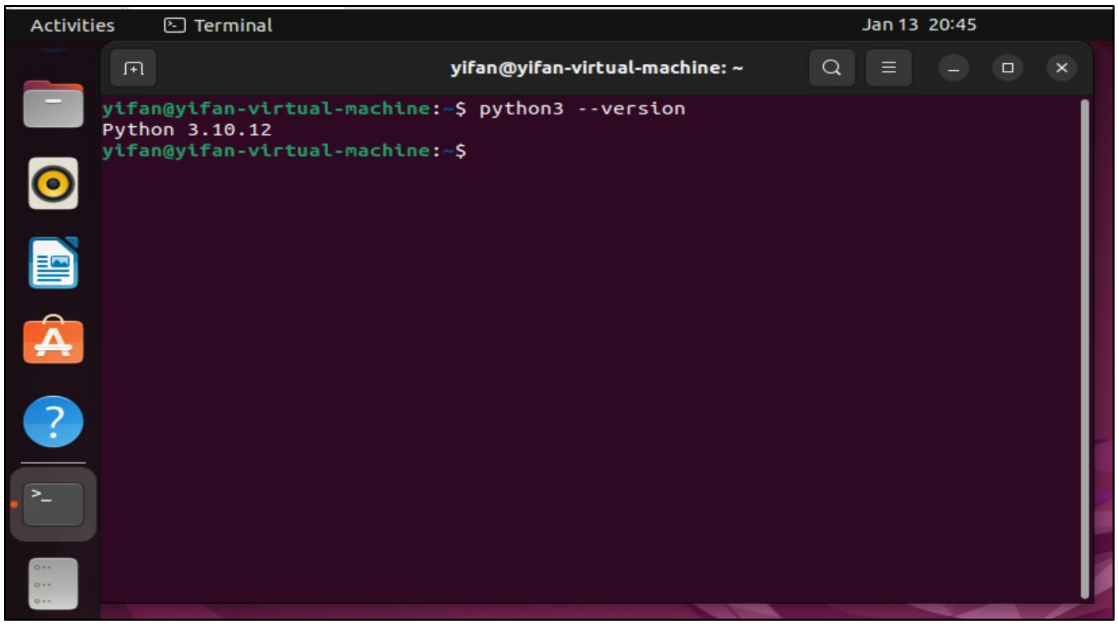


Image 2: Screenshot of successful python installation

1.4. Tutorials

I have systematically studied python and operating systems. Python is relatively simple and commonly used. Here is a summary of some common basic commands, mainly for Linux and Windows systems, because these two systems are the most widely used in virtual machines.

Basic Linux commands

ls - List files and folders in a directory.
cd [directory name] - Change the current directory.
pwd - displays the full path of the current directory.
mkdir [directory name] - Create a new directory.
rmdir [directory name] - delete an empty directory.
rm [filename] - Delete a file or directory.
cp [original file] [destination file] - Copies a file or directory.
mv [original file] [destination file] - Move or rename a file or directory.
touch [filename] - Creates an empty file or updates the file's timestamp.
cat [filename] - View file contents.
echo [text] - displays text.
grep [text] [filename] - Search a file for specified text.
chmod [permissions] [filename] - Change the permissions of a file or directory.
chown [user] [filename] - Change the owner of a file or directory.
top - Displays currently running processes and their resource usage.

Windows basic commands

dir - List files and folders in a directory.
cd [directory name] - Change the current directory.

md [directory name] - Create a new directory.
rd [directory name] - delete a directory.
del [filename] - Delete one or more files.
copy [original file] [destination file] - Copies one or more files.
move [original file] [destination file] - Moves one or more files.
rename [original file name] [new file name] - Rename a file.
type [filename] - displays the contents of a text file.
echo [text] - Display or write a message to a file.
find [text] [filename] - Search a file for a text string.
attrib [filename] - Display or change file attributes.
xcopy [original file] [destination file] - Copies files and directory trees.
chkdsk - Checks disks and displays status reports.
tasklist - displays all currently running processes.

2. Get Familiar with Linux and Python

2.1. Playing around with Linux Terminal

This step mainly uses the Linux terminal to create directories and create python files and script files. Just use the command above.

mkdir ~/Desktop/Yifan_Yang_8386626867/data
mkdir ~/Desktop/Yifan_Yang_8386626867/scripts
touch ~/Desktop/Yifan_Yang_8386626867/scripts/task_1.py
ls ~/Desktop/Yifan_Yang_8386626867/scripts

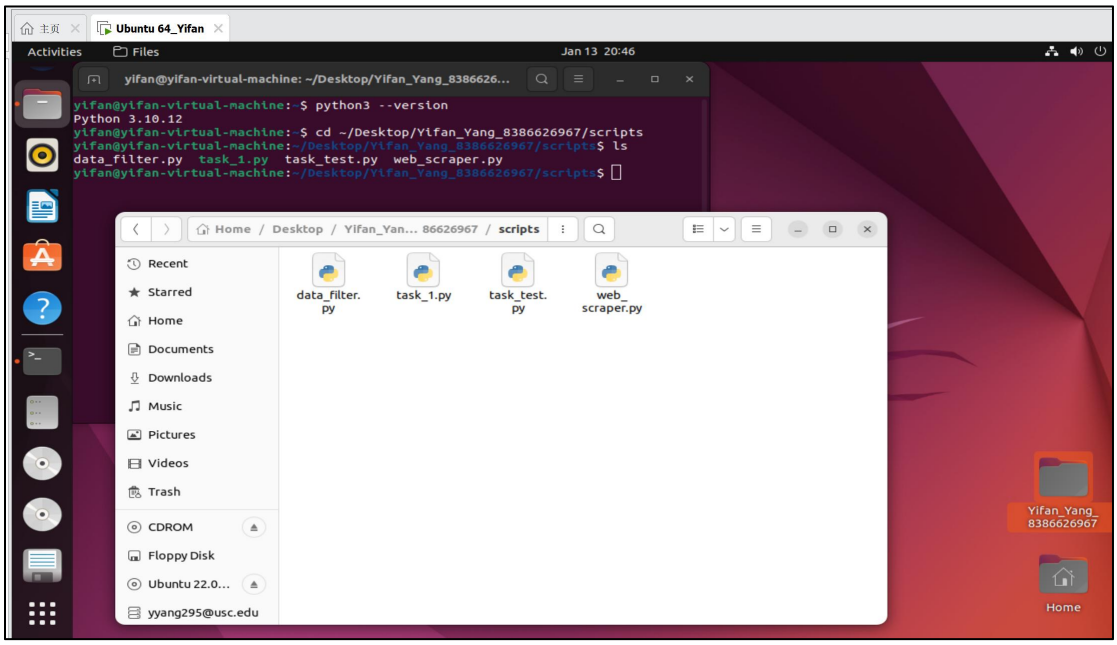


Image 3: Screenshot of directory and python successfully created

2.2. A basic Python Script

This step is mainly for python script encoding, which can be encoded in vim and nano. I chose nano here. I think nano creates task_1.py and it is simpler and more concise to write in python.

Write the Python Script: In vim or nano, type or paste the script provided above.
Save and Exit the Editor:
In vim: Press Esc to exit insert mode. Type :wq and then press Enter to save and exit.
In nano: Press Ctrl + O to write the file. Press Enter to confirm. Press Ctrl + X to exit.

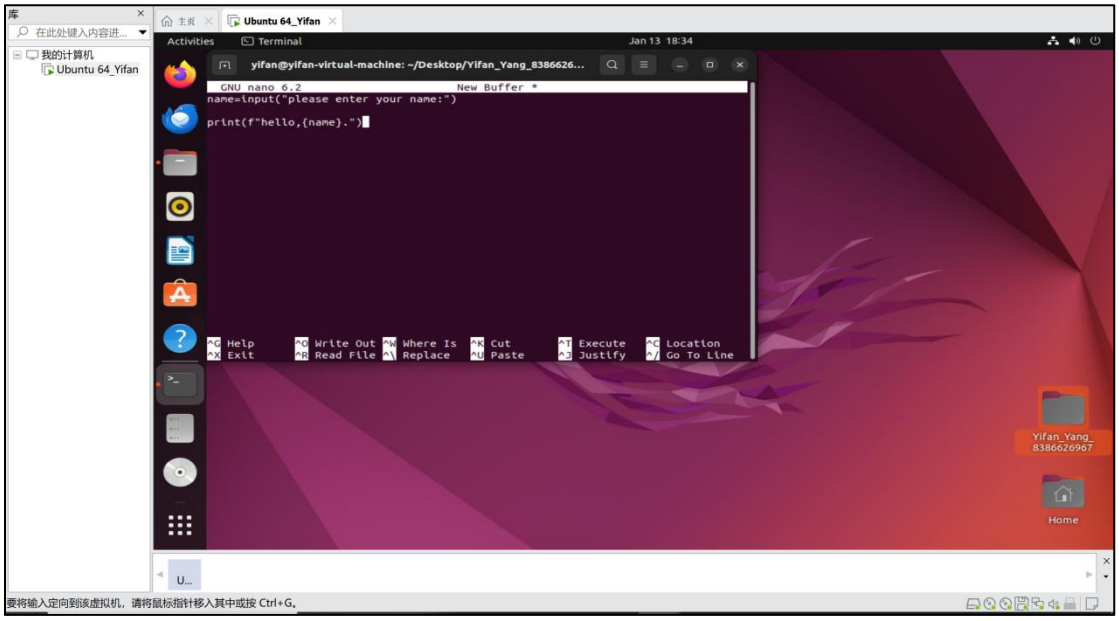


Image 4: nano coding screenshot

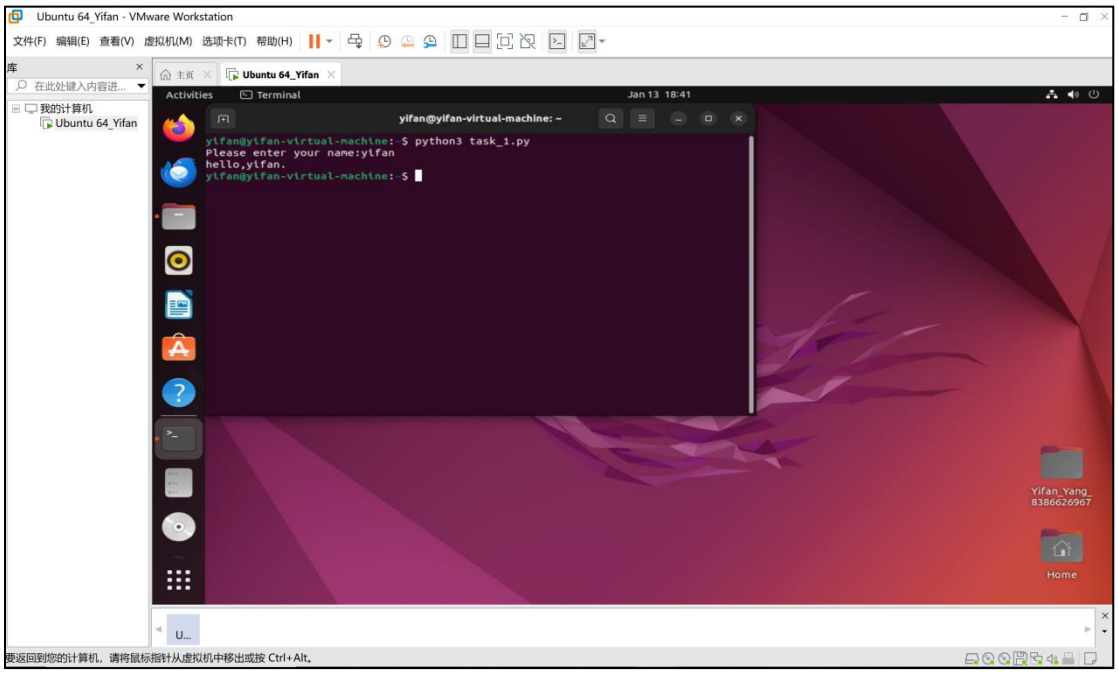


Image 5: Screenshot of successfully running task

2.3. Python Web-scraping Task

This step is to create a new file "web_scraper.py" in the script folder to complete the Python web scraping task. It should be noted that we need pip to install the required libraries. There are many libraries that need to be installed. The remaining steps are the encoding process, which is systematically studied in the DSCI510 course, and then stored in the corresponding folder as required and the first 10 lines of the created html file are printed on the terminal using the terminal.

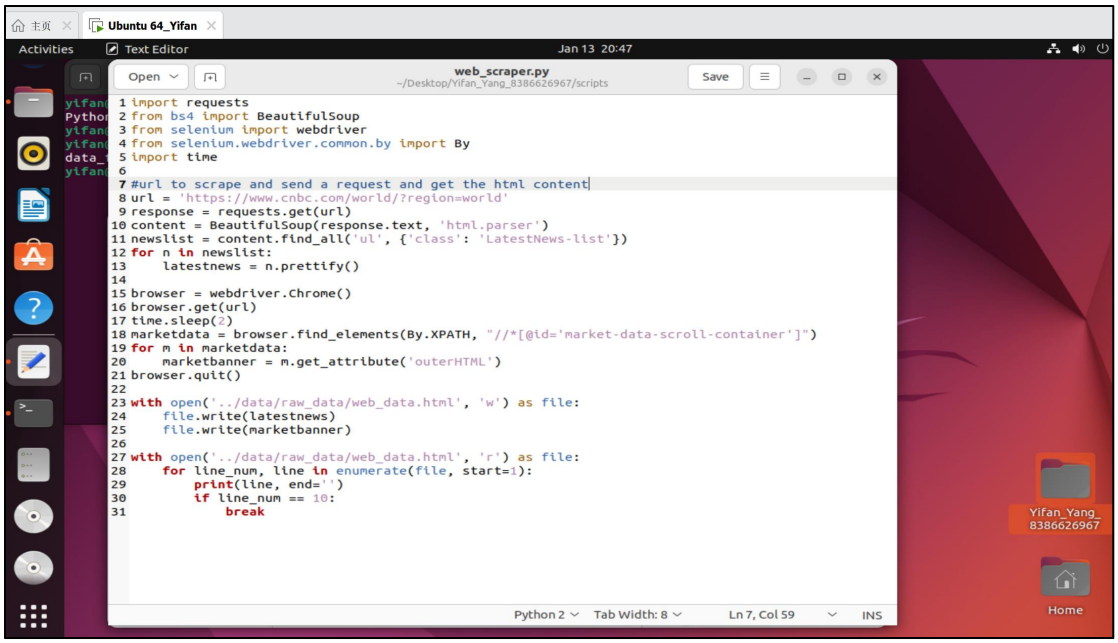


Image 6: Web scraping script code

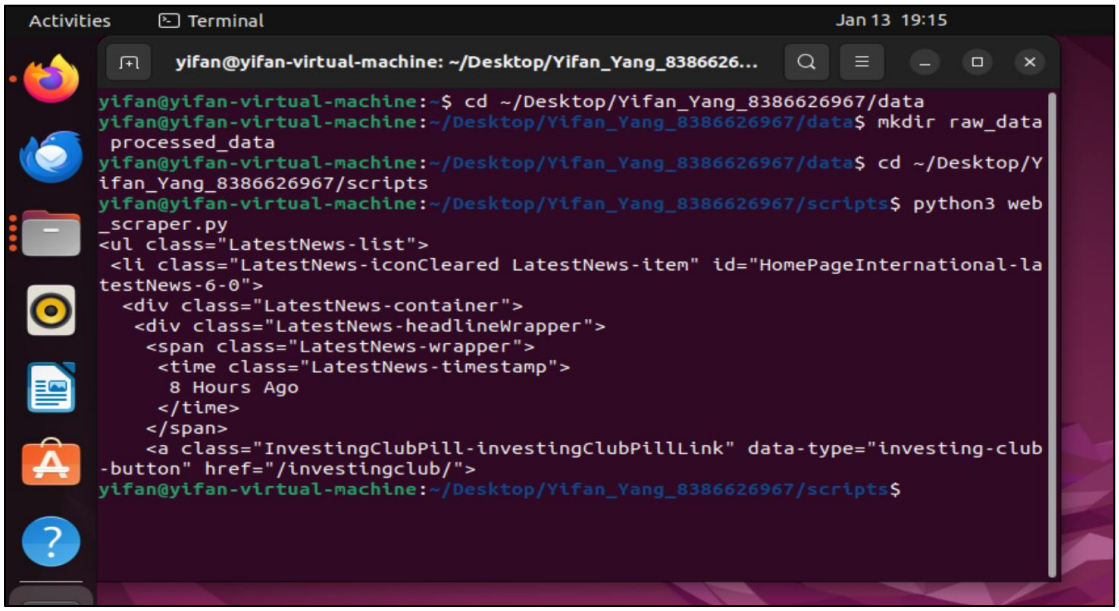


Image 7: Screenshot of successful operation

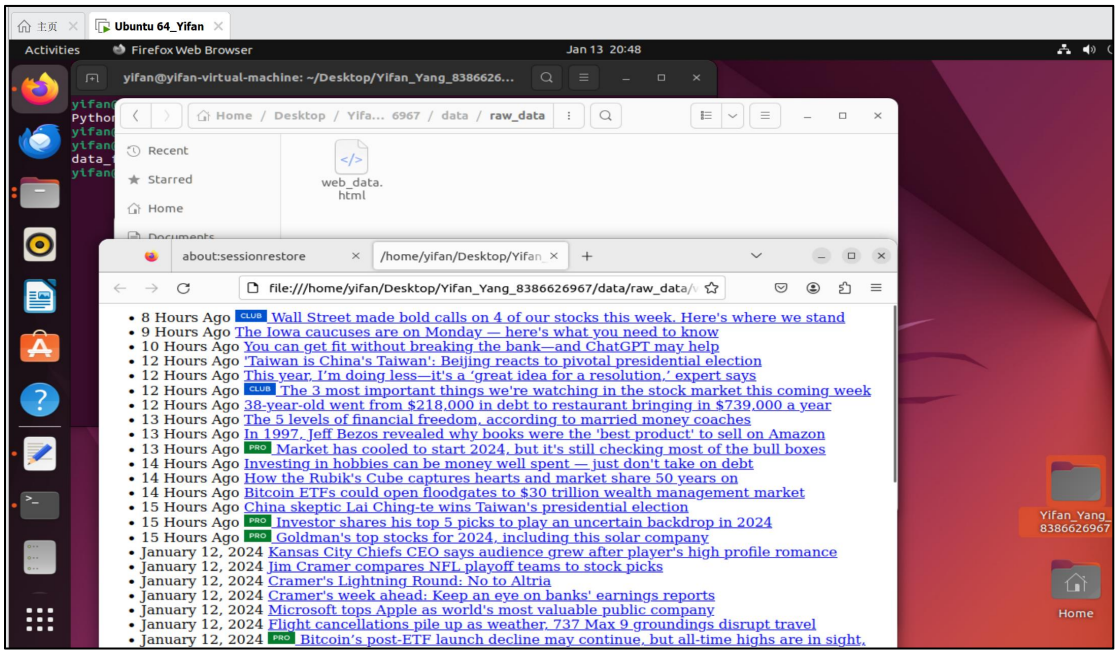


Image 8: Store data screenshot

2.4. Data Filtering Task

This task runs in a similar way to the previous task, requiring you to read the "web_data.html" file and then extract specific elements of interest from the data. These elements are then stored in a CSV named "market_data.csv" in the processed_data folder as required.

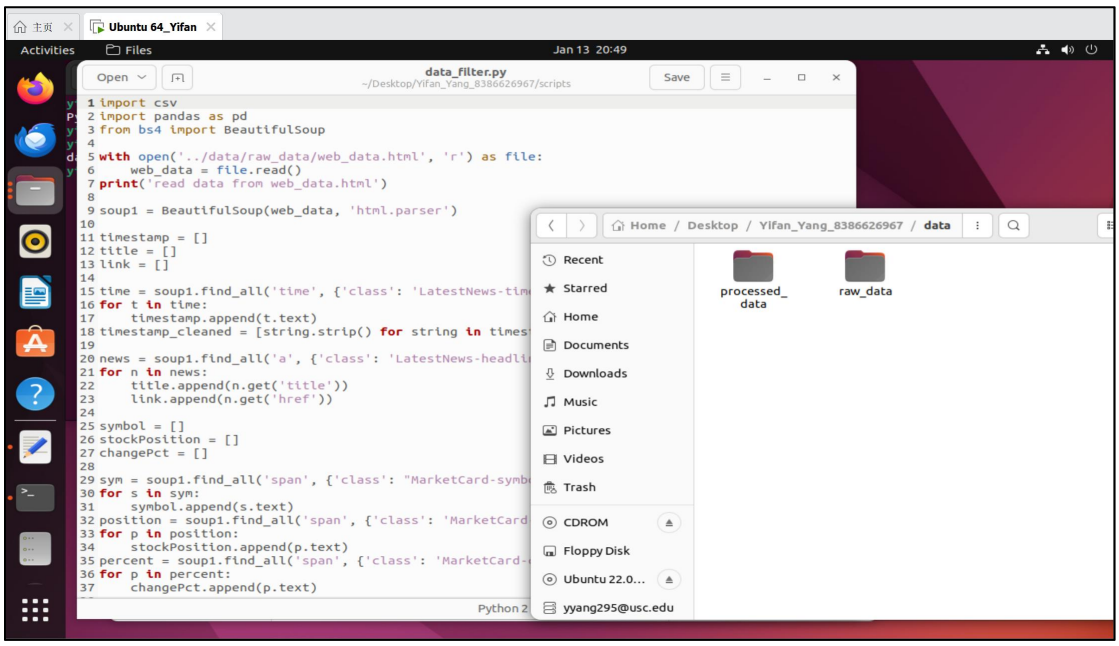


Image 9: Data Filtering script code

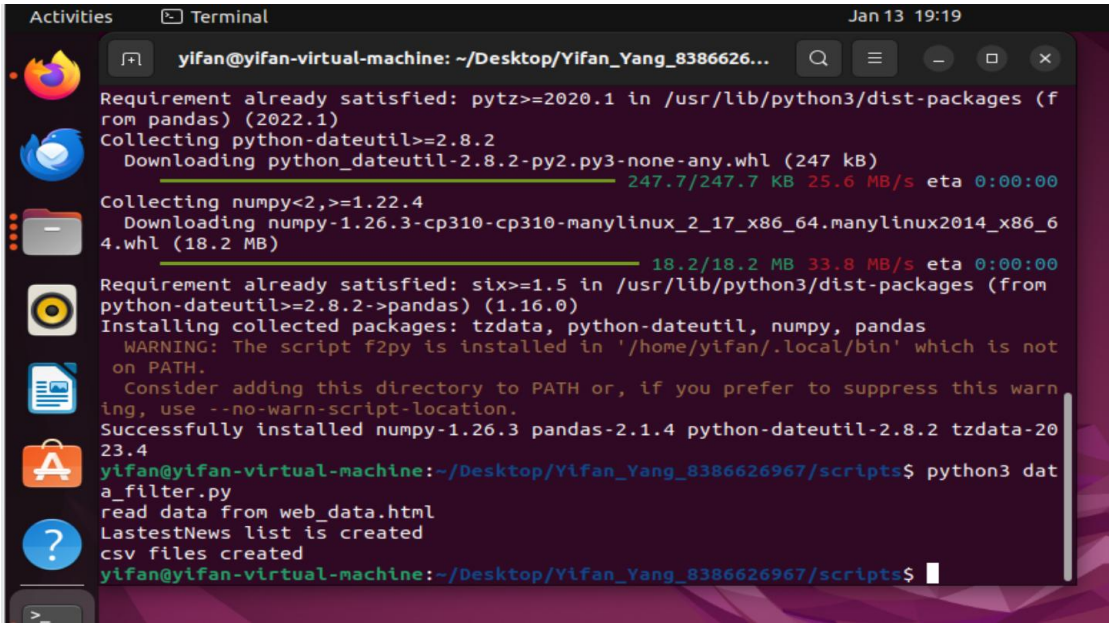


Image 10: Screenshot of successful operation

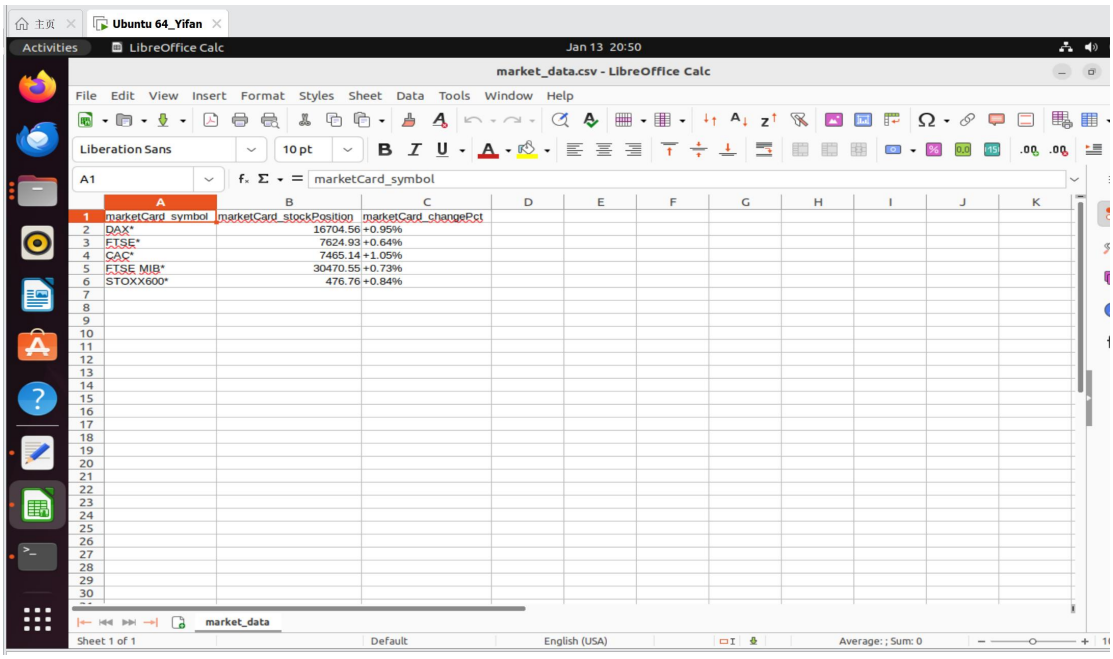
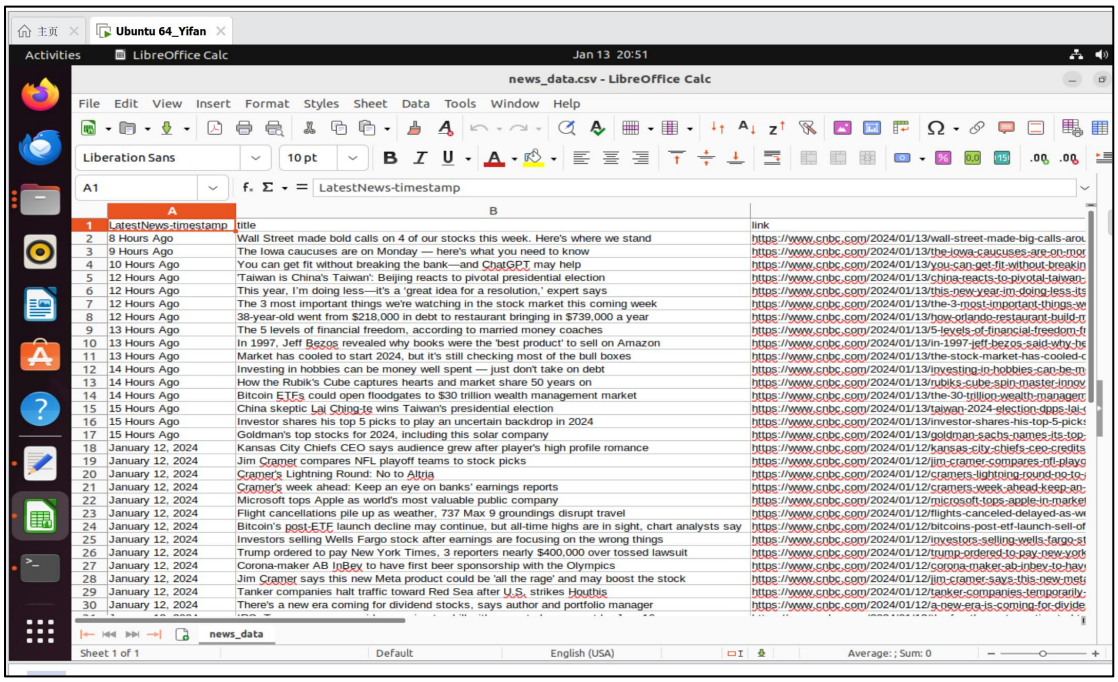


Image 11: Store data screenshot



	A	B	
1	LatestNews-timestamp	title	link
2	8 Hours Ago	Wall Street made bold calls on 4 of our stocks this week. Here's where we stand	https://www.cnbc.com/2024/01/13/wall-street-made-big-calls-arou
3	9 Hours Ago	The Iowa caucuses are on Monday — here's what you need to know	https://www.cnbc.com/2024/01/13/the-iowa-caucuses-are-on-mor
4	10 Hours Ago	You can get fit without breaking the bank—and ChatGPT may help	https://www.cnbc.com/2024/01/13/you-can-get-fit-without-breakin
5	12 Hours Ago	'Taiwan is China's Taiwan': Beijing reacts to pivotal presidential election	https://www.cnbc.com/2024/01/13/china-reacts-to-pivotal-taiwan-
6	12 Hours Ago	This year, I'm doing less—it's a 'great idea for a resolution,' expert says	https://www.cnbc.com/2024/01/13/this-new-year-im-doing-less-its
7	12 Hours Ago	The 3 most important things we're watching in the stock market this coming week	https://www.cnbc.com/2024/01/13/the-3-most-important-things-w
8	12 Hours Ago	38-year-old went from $218,000 in debt to restaurant bringing in $739,000 a year	https://www.cnbc.com/2024/01/13/how-orlando-restaurant-build-n
9	13 Hours Ago	The 5 levels of financial freedom, according to married money coaches	https://www.cnbc.com/2024/01/13/5-levels-of-financial-freedom-fr
10	13 Hours Ago	In 1997, Jeff Bezos revealed why books were the 'best product' to sell on Amazon	https://www.cnbc.com/2024/01/13/in-1997-jeff-bezos-said-why-bo
11	13 Hours Ago	Market has cooled to start 2024, but it's still checking most of the bull boxes	https://www.cnbc.com/2024/01/13/the-stock-market-has-cooled-c
12	14 Hours Ago	Investing in hobbies can be money well spent — just don't take on debt	https://www.cnbc.com/2024/01/13/investing-in-hobbies-can-be-m
13	14 Hours Ago	How the Rubik's Cube captures hearts and market share 50 years on	https://www.cnbc.com/2024/01/13/rubiks-cube-spin-master-innov
14	14 Hours Ago	Bitcoin ETFs could open floodgates to $30 trillion wealth management market	https://www.cnbc.com/2024/01/13/the-30-trillion-wealth-managem
15	15 Hours Ago	China skeptic Lai Ching-te wins Taiwan's presidential election	https://www.cnbc.com/2024/01/13/taiwan-2024-election-dpps-lai-c
16	15 Hours Ago	Investor shares his top 5 picks to play an uncertain backdrop in 2024	https://www.cnbc.com/2024/01/13/investor-shares-his-top-5-picks
17	15 Hours Ago	Goldman's top stocks for 2024, including this solar company	https://www.cnbc.com/2024/01/13/goldman-sachs-names-its-top-
18	January 12, 2024	Kansas City Chiefs CEO says audience grew after player's high profile romance	https://www.cnbc.com/2024/01/12/kansas-city-chiefs-ceo-credits
19	January 12, 2024	Jim Cramer compares NFL playoff teams to stock picks	https://www.cnbc.com/2024/01/12/jim-cramer-compares-nfl-playo
20	January 12, 2024	Cramer's Lightning Round: No to Altria	https://www.cnbc.com/2024/01/12/cramers-lightning-round-no-to-
21	January 12, 2024	Cramer's week ahead: Keep an eye on banks' earnings reports	https://www.cnbc.com/2024/01/12/cramers-week-ahead-keep-an-
22	January 12, 2024	Microsoft tops Apple as world's most valuable public company	https://www.cnbc.com/2024/01/12/microsoft-tops-apple-in-market
23	January 12, 2024	Flight cancellations pile up as weather, 737 Max 9 groundings disrupt travel	https://www.cnbc.com/2024/01/12/flights-canceled-delayed-as-we
24	January 12, 2024	Bitcoin's post-ETF launch decline may continue, but all-time highs are in sight, chart analysts say	https://www.cnbc.com/2024/01/12/bitcoins-post-etf-launch-sell-of
25	January 12, 2024	Investors selling Wells Fargo stock after earnings are focusing on the wrong things	https://www.cnbc.com/2024/01/12/investors-selling-wells-fargo-st
26	January 12, 2024	Trump ordered to pay New York Times, 3 reporters nearly $400,000 over tossed lawsuit	https://www.cnbc.com/2024/01/12/trump-ordered-to-pay-new-yor
27	January 12, 2024	Corona-maker AB InBev to have first beer sponsorship with the Olympics	https://www.cnbc.com/2024/01/12/corona-maker-ab-inbev-to-hav
28	January 12, 2024	Jim Cramer says this new Meta product could be 'all the rage' and may boost the stock	https://www.cnbc.com/2024/01/12/jim-cramer-says-this-new-meta
29	January 12, 2024	Tanker companies halt traffic toward Red Sea after U.S. strikes Houthis	https://www.cnbc.com/2024/01/12/tanker-companies-temporarily-
30	January 12, 2024	There's a new era coming for dividend stocks, says author and portfolio manager	https://www.cnbc.com/2024/01/12/a-new-era-is-coming-for-divide

Image 12: Store data screenshot

Yifan Yang
8386626967
01/13/2024