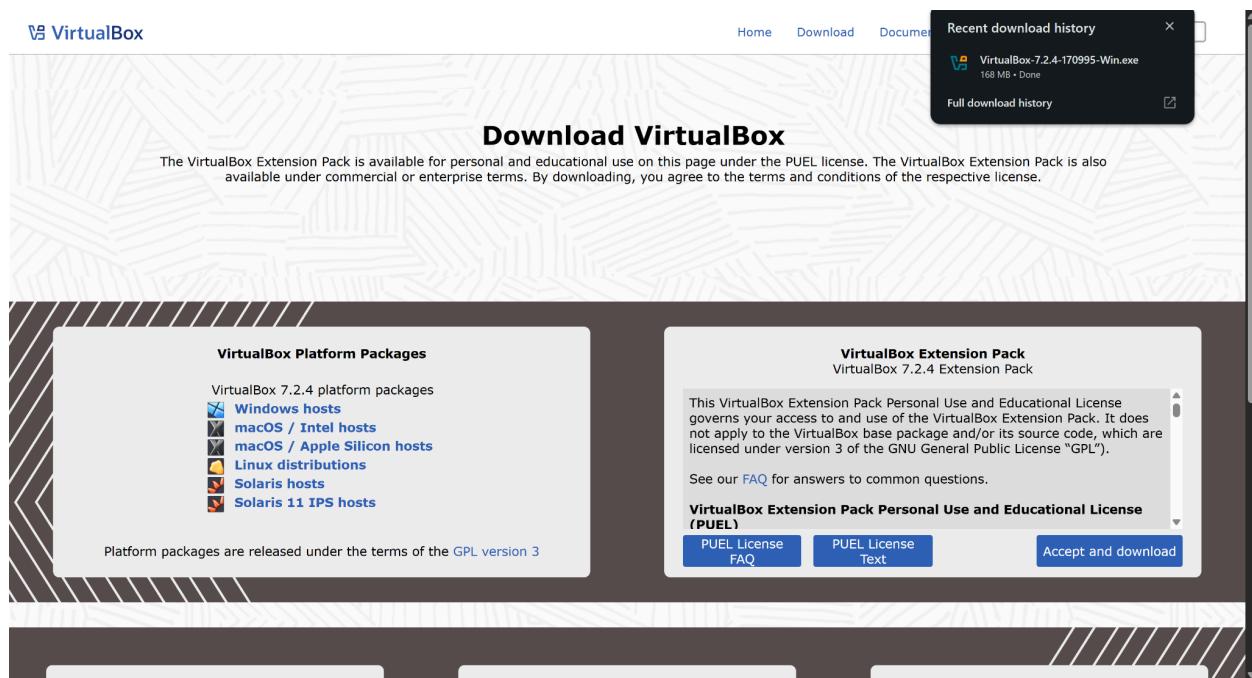
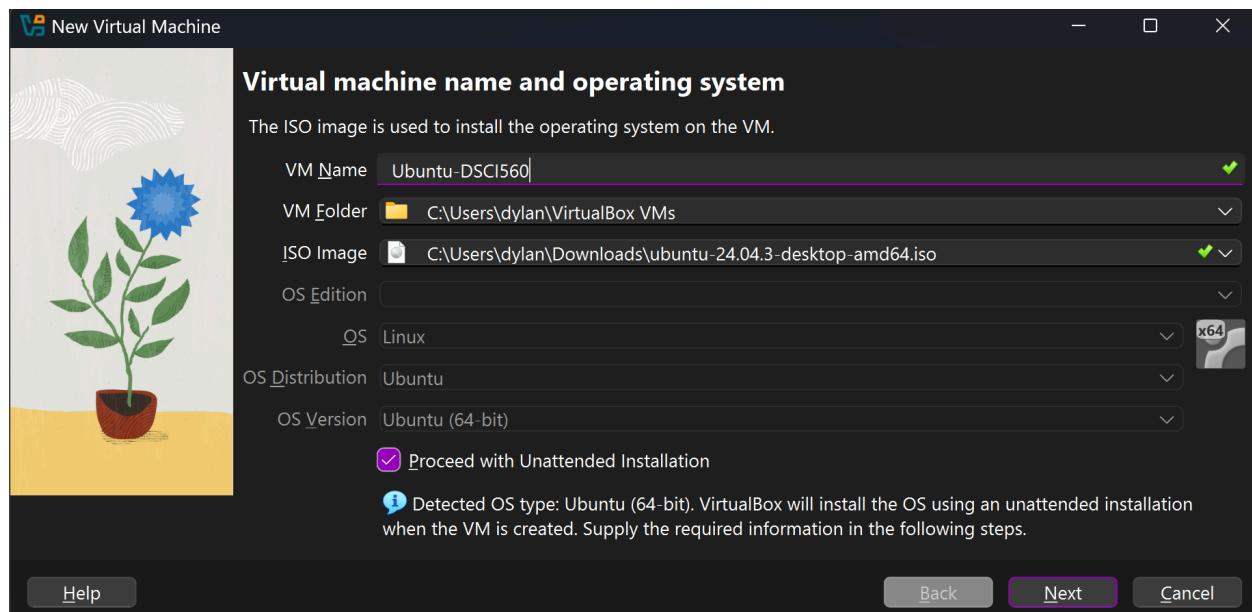


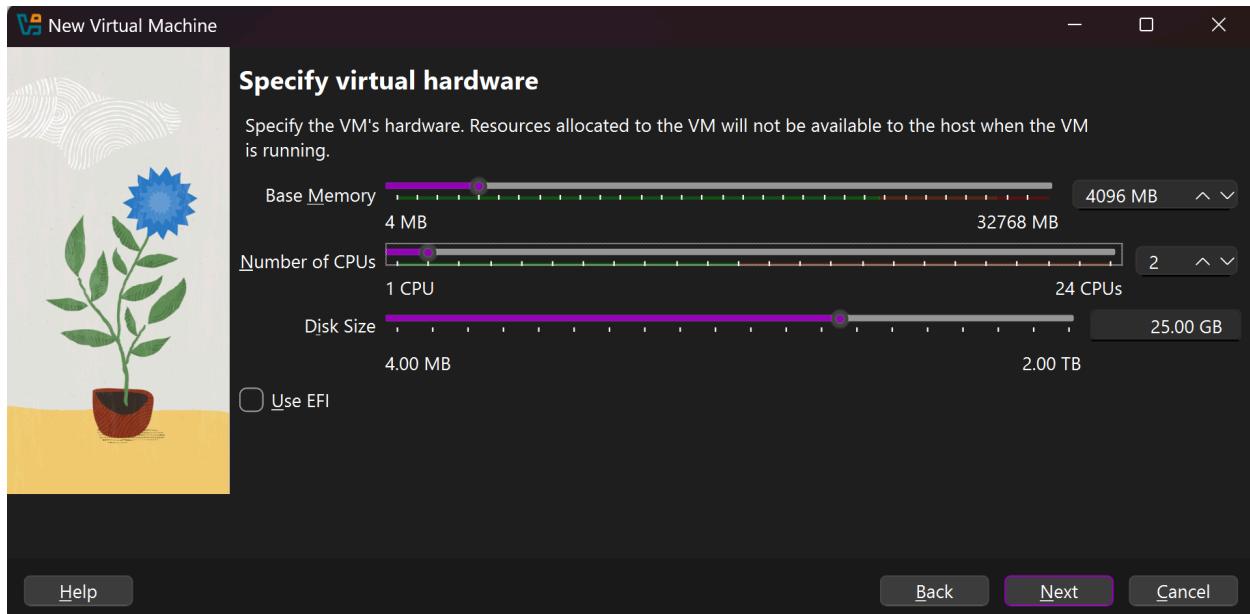
Name: Dylan Chen
ID: 6984540266

1. Installation and Setup

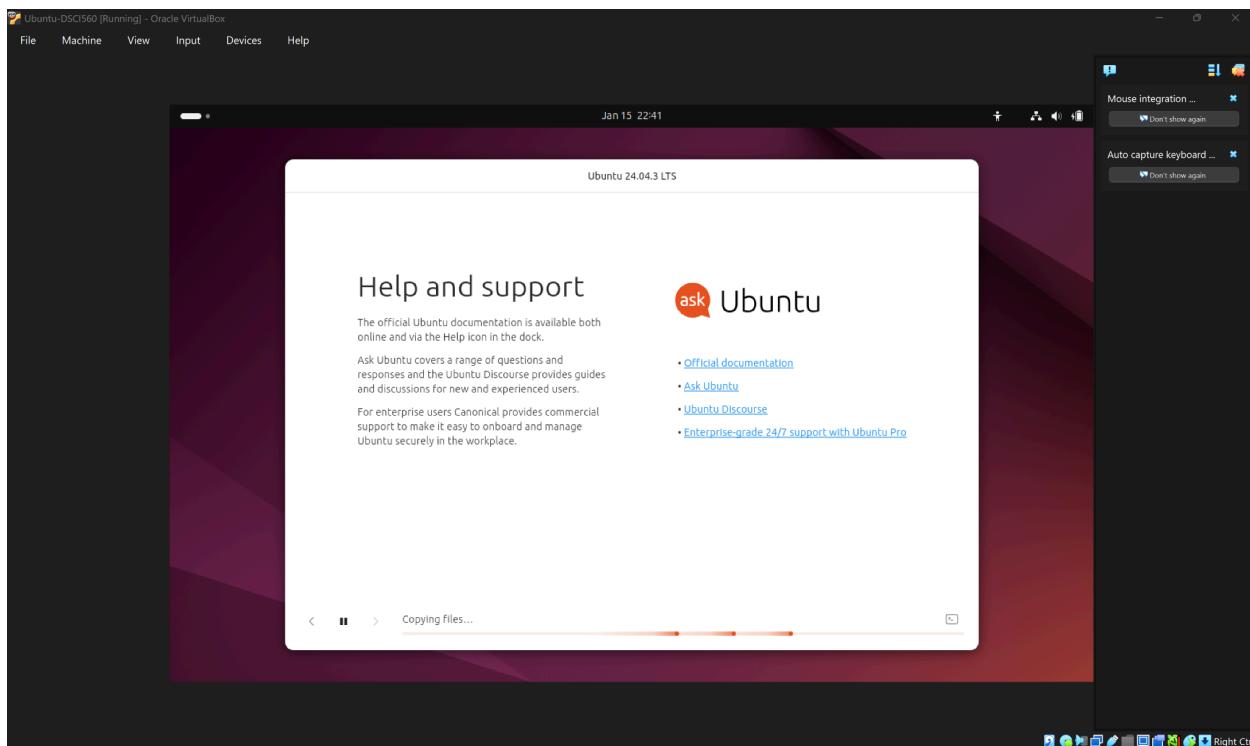


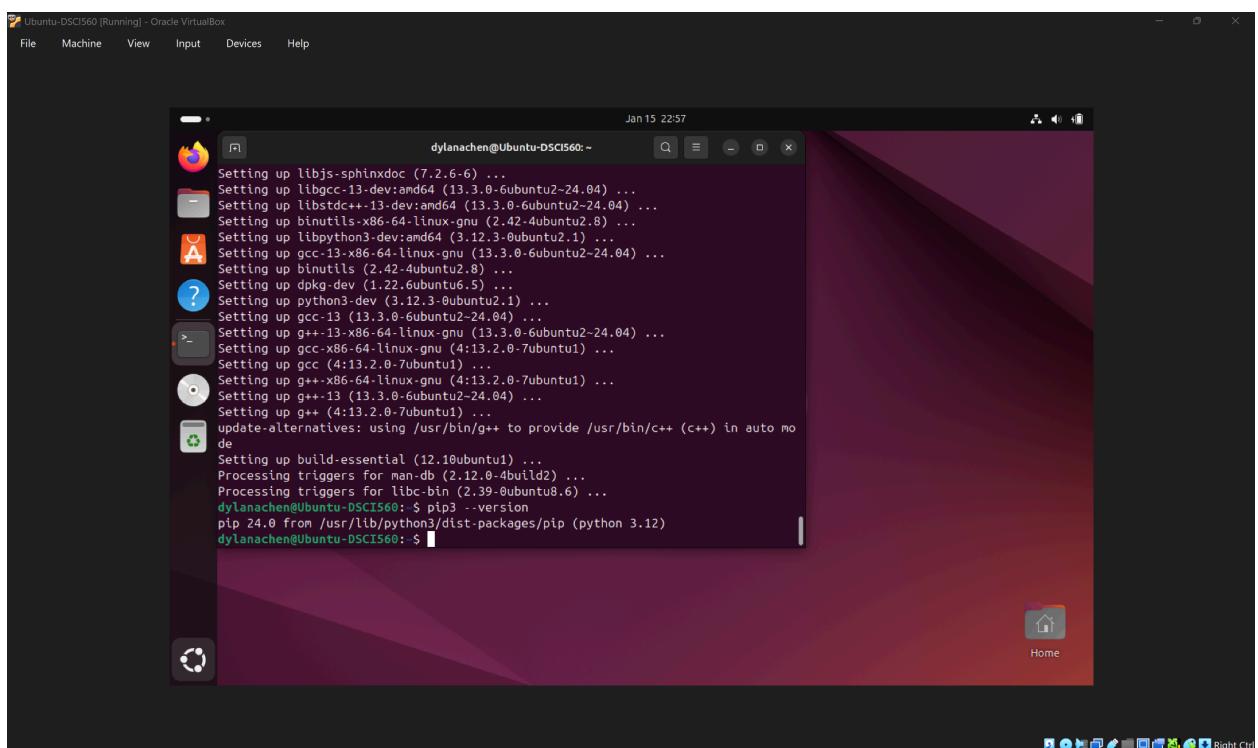
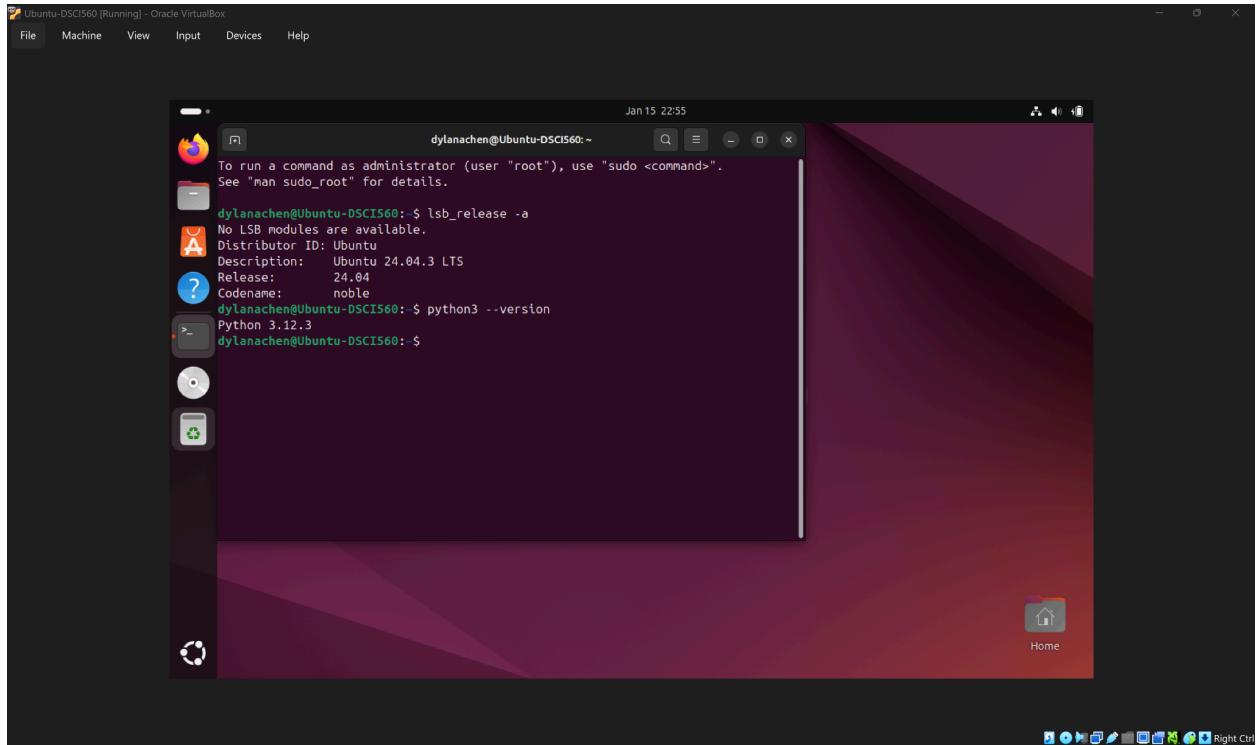
Pictured above is downloading VirtualBox due to issues accessing the Site ID required for the appropriate VMWare Workstation Pro download for Windows (as opposed to VMWare Fusion for Mac).





Creating a new Linux virtual machine using the Ubuntu Desktop ISO image (24.04.3 LTS for Intel or AMD 64-bit architecture). This includes setting up the VM with 4096 MB of RAM and 2 processors with the default 25 GB of disk size.





Setting up and downloading Ubuntu on the virtual machine. Then, downloading python3 and pip through the Linux terminal with sudo commands. Finally, verifying downloads and versions through the –version command.

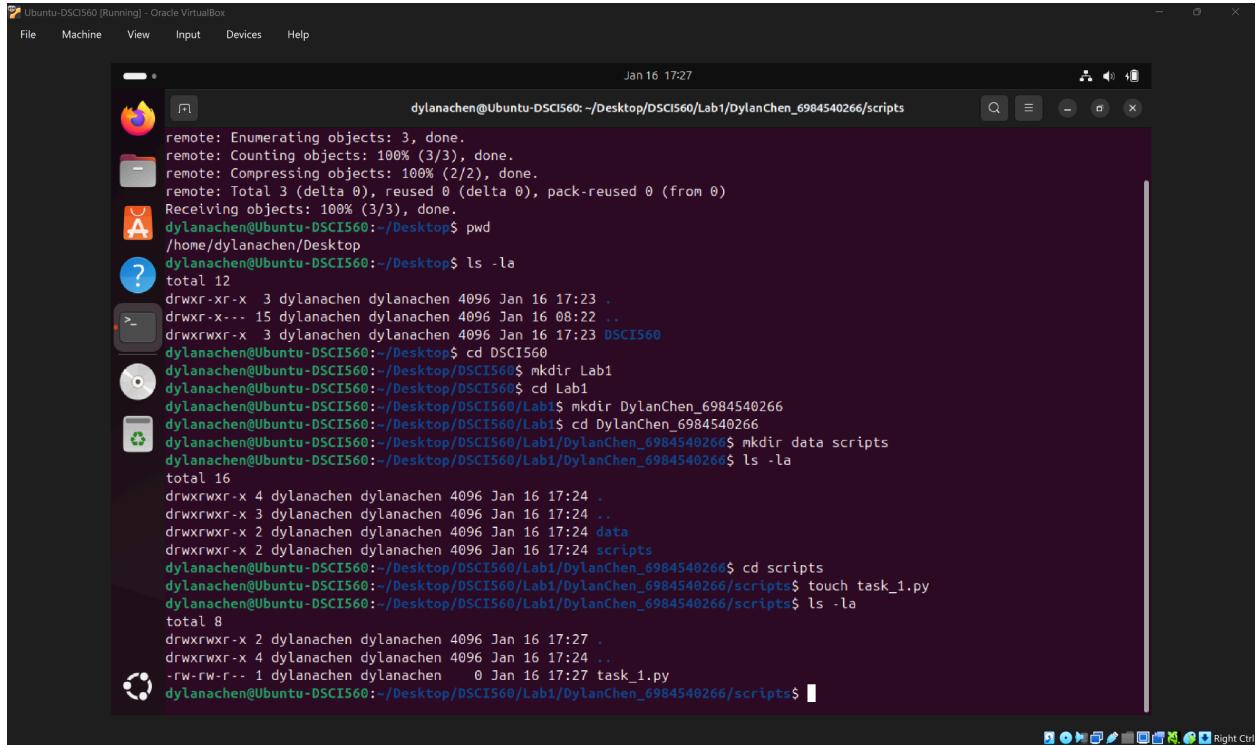
The image shows a dual-screen setup. The top screen displays a terminal window on an Ubuntu desktop within Oracle VirtualBox. The terminal output is as follows:

```
git-man liberror-perl
Suggested packages:
  git-daemon-run | git-daemon-sysvinit git-doc git-email git-gui gitk gitweb
  git-cvs git-mediawiki git-svn
The following NEW packages will be installed:
  git git-man liberror-perl
0 upgraded, 3 newly installed, 0 to remove and 148 not upgraded.
Need to get 4,886 kB of archives.
After this operation, 24.5 MB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://us.archive.ubuntu.com/ubuntu noble/main amd64 liberror-perl all 0.17029-2 [25.6 kB]
Get:2 http://us.archive.ubuntu.com/ubuntu noble-updates/main amd64 git-man all 1:2.43.0-1ubuntu7.3 [1,100 kB]
Get:3 http://us.archive.ubuntu.com/ubuntu noble-updates/main amd64 git amd64 1:2.43.0-1ubuntu7.3 [3,680 kB]
Fetched 4,886 kB in 1s (4,384 kB/s)
Selecting previously unselected package liberror-perl.
(Reading database ... 154004 files and directories currently installed.)
Preparing to unpack .../liberror-perl_0.17029-2_all.deb ...
Unpacking liberror-perl (0.17029-2) ...
Selecting previously unselected package git-man.
Preparing to unpack .../git-man_1X3a2.43.0-1ubuntu7.3_all.deb ...
Unpacking git-man (1:2.43.0-1ubuntu7.3) ...
Selecting previously unselected package git.
Preparing to unpack .../git_1X3a2.43.0-1ubuntu7.3_amd64.deb ...
Unpacking git (1:2.43.0-1ubuntu7.3) ...
Setting up liberror-perl (0.17029-2) ...
Setting up git-man (1:2.43.0-1ubuntu7.3) ...
Setting up git (1:2.43.0-1ubuntu7.3) ...
Processing triggers for man-db (2.12.0-4build2) ...
dylanachen@Ubuntu-DSCL560:~$ git --version
git version 2.43.0
dylanachen@Ubuntu-DSCL560:~$ git config --global user.email "dylanchen05@gmail.com"
dylanachen@Ubuntu-DSCL560:~$
```

The bottom screen shows a GitHub repository page for "dylanachen / DSCL560". The repository has 1 branch and 0 tags. It contains a single commit by dylanachen titled "Initial commit" made 1 minute ago. The commit message is "Initial commit". The repository description is "Lab assignments for DSCI 560 - Data Science Professional Practicum". The sidebar shows 0 stars, 0 forks, and 0 releases.

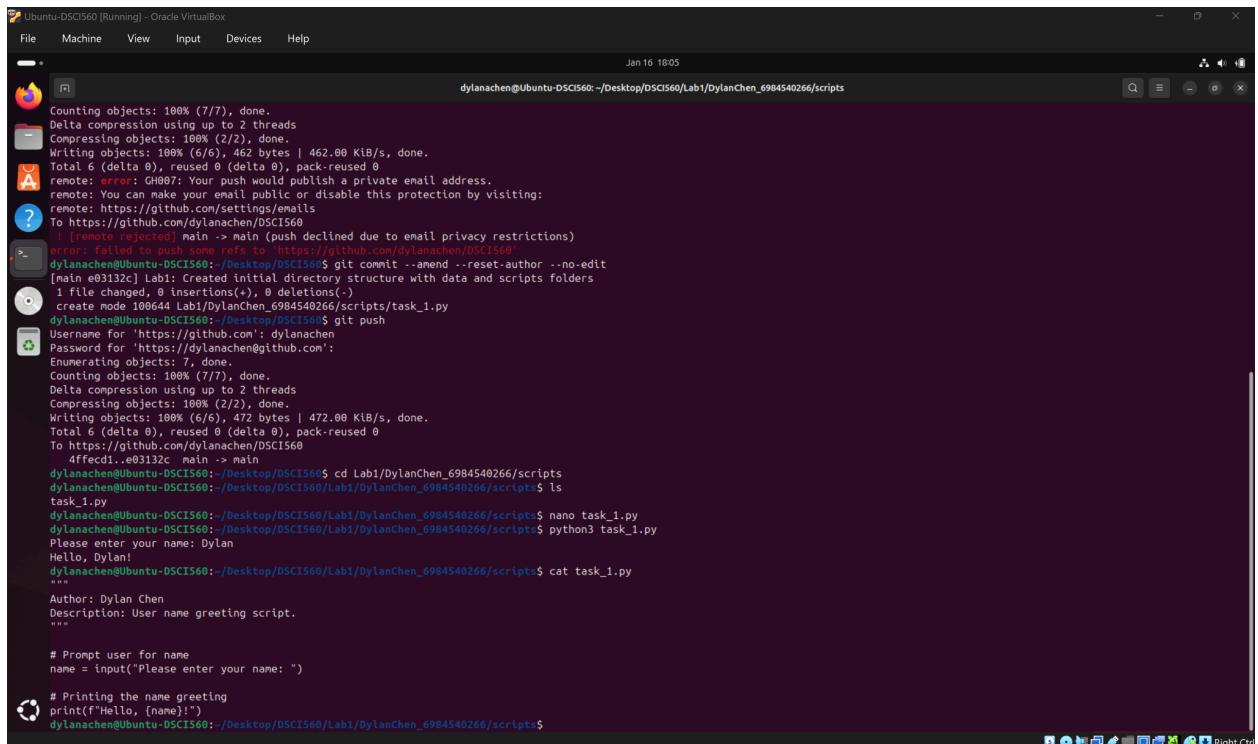
Downloaded git and set up basic configuration information as well as created a GitHub repository for all DSCI 560 labs. This will be later cloned to the VM Ubuntu desktop wherein files can be created.

2. Get Familiar with Linux and Python



```
remote: Enumerating objects: 3, done.
remote: Counting objects: 100% (3/3), done.
remote: Compressing objects: 100% (2/2), done.
remote: Total 3 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (3/3), done.
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560/Lab1/DylanChen_6984540266/scripts$ pwd
/home/dylanachen/Desktop
dylanachen@Ubuntu-DSCI560:~/Desktop$ ls -la
total 12
drwxr-xr-x 3 dylanachen dylanachen 4096 Jan 16 17:23 .
drwxr-xr-x 15 dylanachen dylanachen 4096 Jan 16 08:22 ..
drwxrwxr-x 3 dylanachen dylanachen 4096 Jan 16 17:23 DSCI560
dylanachen@Ubuntu-DSCI560:~/Desktop$ cd DSCI560
dylanachen@Ubuntu-DSCI560:~/DSCI560$ mkdir Lab1
dylanachen@Ubuntu-DSCI560:~/DSCI560$ cd Lab1
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1$ mkdir DylanChen_6984540266
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1$ cd DylanChen_6984540266
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1/DylanChen_6984540266$ mkdir data scripts
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1/DylanChen_6984540266$ ls -la
total 16
drwxrwxr-x 4 dylanachen dylanachen 4096 Jan 16 17:24 .
drwxrwxr-x 3 dylanachen dylanachen 4096 Jan 16 17:24 ..
drwxrwxr-x 2 dylanachen dylanachen 4096 Jan 16 17:24 data
drwxrwxr-x 2 dylanachen dylanachen 4096 Jan 16 17:24 scripts
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1/DylanChen_6984540266$ cd scripts
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1/DylanChen_6984540266$ touch task_1.py
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1/DylanChen_6984540266$ ls -la
total 8
drwxrwxr-x 2 dylanachen dylanachen 4096 Jan 16 17:27 .
drwxrwxr-x 4 dylanachen dylanachen 4096 Jan 16 17:24 ..
-rw-rw-r-- 1 dylanachen dylanachen 0 Jan 16 17:27 task_1.py
dylanachen@Ubuntu-DSCI560:~/DSCI560/Lab1/DylanChen_6984540266$ script$
```

2.1 Cloned the GitHub repo onto the VM desktop and created a directory to contain Lab1. Then, created the DylanChen_6984540266 directory and data and scripts subdirectories with scripts/task_1.py. Used ls to view with l and a tags for more information.



```
Counting objects: 100% (7/7), done.
Delta compression using up to 2 threads
Compressing objects: 100% (2/2), done.
Writing objects: 100% (6/6), 462 bytes | 462.00 KiB/s, done.
Total 6 (delta 0), reused 0 (delta 0), pack-reused 0
remote: error: GH007: Your push would publish a private email address.
remote: You can make your email public or disable this protection by visiting:
remote: https://github.com/settings/emails
To https://github.com/dylanachen/DSCI560
! [remote rejected] main -> main (push declined due to email privacy restrictions)
error: failed to push some refs to 'https://github.com/dylanachen/DSCI560'
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560$ git commit --amend --reset-author --no-edit
[main e03132c] Lab1: Created initial directory structure with data and scripts folders
 1 file changed, 0 insertions(+), 0 deletions(-)
 create mode 100640 Lab1/DylanChen_6984540266/scripts/task_1.py
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560$ git push
Username for 'https://github.com': dylanachen
Password for 'https://dylanachen@github.com':
Enumerating objects: 7, done.
Counting objects: 100% (7/7), done.
Delta compression using up to 2 threads
Compressing objects: 100% (2/2), done.
Writing objects: 100% (6/6), 472 bytes | 472.00 KiB/s, done.
Total 6 (delta 0), reused 0 (delta 0), pack-reused 0
To https://github.com/dylanachen/DSCI560
4ffecd1..e03132c main -> main
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560$ cd Lab1/DylanChen_6984540266/scripts
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560/Lab1/DylanChen_6984540266$ ls
task_1.py
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560/Lab1/DylanChen_6984540266$ nano task_1.py
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560/Lab1/DylanChen_6984540266$ python3 task_1.py
Please enter your name: Dylan
Hello, Dylan!
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560/Lab1/DylanChen_6984540266$ cat task_1.py
"""
Author: Dylan Chen
Description: User name greeting script.

# Prompt user for name
name = input("Please enter your name: ")

# Printing the name greeting
print("Hello, [name]!")
dylanachen@Ubuntu-DSCI560: ~/Desktop/DSCI560/Lab1/DylanChen_6984540266$
```

```
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ git remote set-url origin https://github.com/dylanachen/DSCI560-Lab1.git
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ git remote -v
origin  https://github.com/dylanachen/DSCI560-Lab1.git (fetch)
origin  https://github.com/dylanachen/DSCI560-Lab1.git (push)
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ mv Lab1/DylanChen_USCID/* .
mv: cannot stat 'Lab1/DylanChen_USCID/*': No such file or directory
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ mv Lab1/DylanChen_6984540266/* .
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ rm -rf Lab1/
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ ls -la
total 24
drwxrwxr-x 5 dylanachen dylanachen 4096 Jan 18 18:14
drwxr-xr-x 3 dylanachen dylanachen 4096 Jan 18 17:23 ..
drwxrwxr-x 2 dylanachen dylanachen 4096 Jan 18 17:24 data
drwxrwxr-x 8 dylanachen dylanachen 4096 Jan 18 18:13 git
-rw-rw-r-- 1 dylanachen dylanachen 77 Jan 18 17:23 README.md
drwxrwxr-x 2 dylanachen dylanachen 4096 Jan 18 18:03 scripts
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ git add .
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ git commit -m "Reorganized repo structure to be individual for labs"
[main 1c9c936] Reorganized repo structure to be individual for labs
 1 file changed, 0 insertions(+), 0 deletions(-)
 rename {Lab1/DylanChen_6984540266/scripts => scripts}/task_1.py (100%)
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560$ cd ..
dylanachen@Ubuntu-DSCI560:~/Desktop$ mv DSCI560 DSCI560-Lab1
dylanachen@Ubuntu-DSCI560:~/Desktop$ ls
DSCI560-Lab1
dylanachen@Ubuntu-DSCI560:~/Desktop$ cd DSCI560-Lab1
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1$
```

2.2 Using nano to write a basic name-input greeting python script in task_1.py and then test running it in the terminal. Using cat to display the file contents. Then, I renamed the GitHub repo and reorganized the VM desktop files to match after receiving confirmation from Tanya to create an individual repo per lab.

CNBC

MARKETS BUSINESS INVESTING TECH POLITICS VIDEO INVESTING CLUB PRO LIVESTREAM

Search quotes, news & videos

WATCHLIST SIGN IN CREATE FREE ACCOUNT

MAKE IT SELECT USA - INTL

div#market-data-scroll-contain er.MarketsBanner-marketData 1290 x 76 OIL GOLD FX CRYPTO

STOXX600*	614.38	-0.19 -0.03%	DAX*	25,297.13	-55.26 -0.22%	FTSE*	10,235.29	-3.65 -0.04%	CAC*	8,258.94	-54.18 -0.65%	FTSE MIB*	45,799.69	-50.08 -0.11%
LAST 15 MINUTE PM UNIT			LAST 15 MINUTE PM UNIT			LAST 15 MINUTE PM UNIT			LAST 15 MINUTE PM UNIT			LAST 15 MINUTE PM UNIT		

European stocks finish lower as Greenland remains in focus; gold and silver prices dip

LEARN MORE

Elements Console Sources Network Performance Memory Application Privacy and security Lighthouse Recorder

Styles Computed Layout Event Listeners >

div.LatestNews-isHomePage.LatestNews-isIntlHomepage 299.99 x 517.68

LATEST NEWS

- 20 MIN AGO These stocks reporting earnings next week have a history of beating expectations
- 21 MIN AGO Biggest offshore wind project in U.S. to resume construction after judge lifts Trump suspension
- 28 MIN AGO OpenAI to begin testing ads on ChatGPT in the U.S.
- 32 MIN AGO Micron stock climbs as CEO

Trump floats new tariffs in push to

div.sidebarArticle.sidebar.PageBuilder-sidebar data-test="SidebarArticle" data-kind="parent" style="height: calc(100% - 324.963px);"

div.sidebarArticle-noSticky

headerLatestNews-header

h2Latest News

ulLatestNews-list

buttonLatestNews-button

View More

Elements Console Sources Network Performance Memory Application Privacy and security Lighthouse Recorder

Styles Computed Layout Event Listeners >

```

dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1$ cd DylanChen_6984540266/scripts
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ nano web_scraper.py
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ cd ..
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266$ cd data
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/data$ mkdir raw_data processed_data
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/data$ cd ..
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266$ cd scripts
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ nano web_scraper.py
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ python3 web_scraper.py
Web Scraping Target URL: https://www.cnbc.com/world/?region=world
Retrieved page (Status Code: 200)
HTML content parsed successfully
HTML data saved to: /home/dylanachen/Desktop/DSCI560-Lab1/DylanChen_6984540266/data/raw_data/web_data.html
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ ls ..data/raw_data/
web_data.html
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ head -n 10 ..data/raw_data/web_data.html
<!DOCTYPE html>
<html itemscope="" itemtype="https://schema.org/WebPage" lang="en" prefix="og:https://ogp.me/ns#>
<head>
<meta content="website" property="og:type"/>
<meta content="International: Top News And Analysis" property="og:title"/>
<meta content="CNBC International is the world leader for news on business, technology, China, trade, oil prices, the Middle East and markets." property="og:description"/>
<meta content="https://www.cnbc.com/world/" property="og:url"/>
<meta content="CNBC" property="og:site_name"/>
<meta content="max-image-preview:large" name="robots"/>
<meta content="telephone=no" name="format-detection"/>
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ 

```

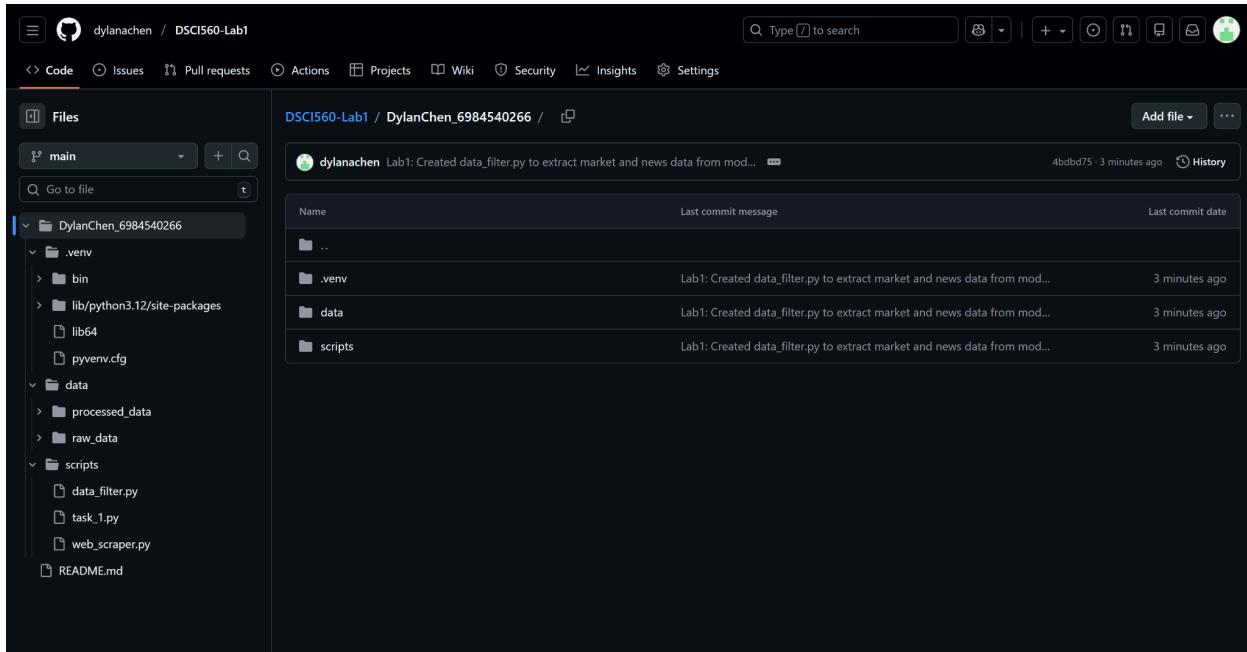
2.3 Inspecting CNBC world news site for the tags for the Markets banner and the Latest News section. Creating the new data folders and using nano to write a web_scraper.py script that uses bs4 and requests to scrape raw HTML data into data/raw_data/web_data.html. Then, printing the first 10 lines in the terminal.

```

Ubuntu-DSCI560 [Running] - Oracle VirtualBox
File Machine View Input Devices Help
Jan 16 22:59
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts

</span>
<span class="MarketCard-changesPct">
-0.11%
</span>
</div>
</div>
<div class="MarketCard-row">
<span class="MarketCard-lastTime">
LAST | 4:35:00 PM GMT
</span>
</div>
</div>
Found 5 total cards
Symbol: STOXX600*
Position: 614.38
Change PCT: -0.03%
Symbol: DAX*
Position: 25,297.13
Change PCT: -0.22%
Symbol: FTSE*
Position: 10,235.29
Change PCT: -0.04%
Symbol: CAC*
Position: 8,258.94
Change PCT: -0.65%
Symbol: FTSE MIB*
Position: 45,799.69
Change PCT: -0.11%
Fields Filtered
Fields saved to CSV
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ cat ..data/processed_data/market_data.csv
symbol,stock_position,change_pct
STOXX600*,614.38,-0.03%
DAX*,25,297.13,-0.22%
FTSE*,10,235.29,-0.04%
CAC*,8,258.94,-0.65%
FTSE MIB*,45,799.69,-0.11%
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ head -5 ..data/processed_data/news_data.csv
timestamp,title,link
1 Hour Ago,More employers worry about workers' financial wellbeing, research shows.,https://www.cnbc.com/2026/01/16/employers-focusing-more-on-employee-financial-wellbeing-study-shows.html
2 Hours Ago,Republicans want to end the 'marriage penalty' for this childcare tax credit.,https://www.cnbc.com/2026/01/16/child-and-dependent-care-tax-credit.html
2 Hours Ago,Labor Dept. accused of echoing Nazi slogan in social media post.,https://www.cnbc.com/2026/01/16/trump-labor-nazi-slogan-social-media.html
2 Hours Ago,Education Department to delay collections on defaulted student loans.,https://www.cnbc.com/2026/01/16/student-loan-collections-paused.html
dylanachen@Ubuntu-DSCI560:~/Desktop/DSCI560-Lab1/DylanChen_6984540266/scripts$ 

```



Created the data_filter.py file to extract stock, position, and change percentage from the market banner and news information from the latest news links. However, because the market banner was loaded dynamically with JavaScript, I had to edit the web_scraper.py file from 2.3 to use Selenium with a headless browser to capture those final tags. Then, all was committed and pushed to GitHub. I had committed once after each section (e.g., 2.1, 2.2, 2.3 ...)