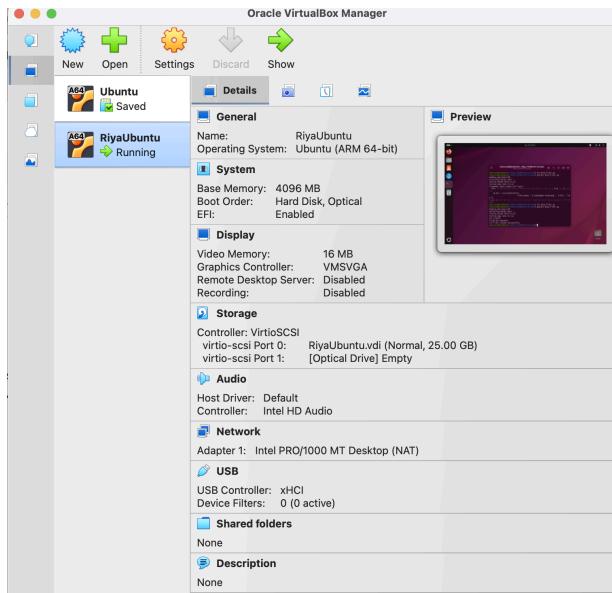


Name: Riya Berry
USC ID: 7828-6427-21
Github Link: https://github.com/18rberry/DSCI_560_Lab1

Task 1:

I used Ubuntu ISO image 25.10 because I needed a version that was compatible with my Mac (MacOs Tahoe 26.2). As a result, I also had to use VirtualBox rather than VMWare Fusion. Below is a screenshot of the Oracle VirtualBox Manager, with all of my settings.



I then installed Python and some packages with Linux (see below):

```
vboxuser@RiyaUbuntu: ~/Riya_7828642721/scripts ... ~ /Riya_7828642721/scripts
Listing...
accountsservice/questing,now 23.13.9-7ubuntu1 arm64 [installed,automatic]
acl/questing,now 2.3.2-2 arm64 [installed,automatic]
adduser/questing,now 3.152ubuntu1 all [installed,automatic]
adwaita-icon-theme/questing,now 49.0-1 all [installed,automatic]
alsa-base/questing,now 1.0.25+dfsg-0ubuntu8 all [installed,automatic]
alsa-topology-conf/questing,now 1.2.5.1-3 all [installed,automatic]
alsa-ucm-conf/questing,now 1.2.12-1ubuntu2 all [installed,upgradable to: 1.2.12-1ubuntu2.1]
alsa-utils/questing,now 1.2.14-1ubuntu1 arm64 [installed,automatic]
anacron/questing,now 2.3-43ubuntu1 arm64 [installed,automatic]
aparmor/questing,now 5.0.0-alpha1-0ubuntu8 arm64 [installed,upgradable to: 5.0.0-alpha1-0ubuntu8.3]
apport-core-dump-handler/questing,now 2.33.1-0ubuntu3 all [installed,automatic]
apport-gtk/questing,now 2.33.1-0ubuntu3 all [installed,automatic]
apport-symptoms/questing,now 0.25 all [installed,automatic]
apport/questing,now 2.33.1-0ubuntu3 all [installed,automatic]
appstream/questing,now 1.0.6-2 arm64 [installed,automatic]
apt-config-icons-hidpi/questing,now 1.0.6-2 all [installed,automatic]
apt-config-icons/questing,now 1.0.6-2 all [installed,automatic]
apt/questing,now 3.1.6ubuntu2 arm64 [installed,automatic]
aptdaemon-data/questing,now 2.0.2 all [installed,automatic]
aptdaemon/questing,now 2.0.2 all [installed,automatic]
```

Task 2:

First, I created my directory and then within it – two subdirectories named “data” and “scripts.”

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$ cd ..  
vboxuser@RiyaUbuntu:~/Riya_7828642721$ ls  
data scripts  
vboxuser@RiyaUbuntu:~/Riya_7828642721$
```

Using VirtualBox’s linux terminal, I created a basic Python script (**task_1.py**) that reads the user’s name (when asked) and greets them politely!

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts ... ~/Riya_7828642721/scripts  
  
# ask the user to enter their name  
name = input("Please enter your name: ")  
  
# greet the user  
print(f"Hello, {name}!")  
  
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$ vim task_1.py  
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$ python3 task_1.py  
Please enter your name: Riya Berry  
Hello, Riya Berry!  
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$
```

Then, I created a Python script (**web_scraping.py**) to scrape data data from this source: <https://www.cnbc.com/world/?region=world>. I also used my Desktop Chrome to inspect the HTML structure/tags of this source. Using terminal, I printed the first 10 lines of the html file (here is the end of the printed output):

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$ head -n 10 ../data/raw_data/web_d  
ata.html
```

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts ... ~/Riya_7828642721/scripts  
  
(min-width:760px){.RiverThumbnail-breakerImage{margin-bottom:none}}.RiverThumbnail  
il-breakerImage img{height:auto;width:100%}.RiverThumbnail-imageThumbnail{positi  
on:relative}.RiverThumbnail-videoFooter[background:linear-gradient(180deg,#e5e579  
900 0,#000);bottom:0;filter:progid:dximagetransform.microsoft.gradient(startColo  
rstr="#001e5799",endColorstr="#000000",GradientType=0);height:40px;padding-left:  
10px;padding-top:15px;position:absolute;width:100%}.RiverThumbnail-videoLabel[co  
lor:#fcb700;font-size:10px;font-weight:800;letter-spacing:1.67px;line-height:12p  
x]@media (min-width:760px){.RiverThumbnail-videoLabel{font-size:12px;letter-spac  
ing:2px;line-height:15px}}.RiverThumbnail-videoTime[color:#fff;font-weight:600;  
letter-spacing:1.2px;padding-left:10px]@media (min-width:760px){.RiverThumbnail  
-videoTime[letter-spacing:1.44px]}.RiverHeadline-headline{font-size:16px;font-wei  
ght:600;letter-spacing:.03px;line-height:18px;padding-bottom:24px}@media (min-wi  
dth:760px){.RiverHeadline-headline{font-size:20px;line-height:24px;padding-right  
:32px}}@media (min-width:1020px){.RiverHeadline-headline{font-size:17px;line-he  
ight:19px;padding-bottom:32px;padding-right:24px}}@media (min-width:1340px){.Rive  
rHeadline-headline{font-size:20px;line-height:24px;padding-right:56px}}.RiverHea  
dline-headline a:focus,.RiverHeadline-headline a:hover{color:#171717;text-decorati  
on:underline}.RiverHeadline-hasThumbnail[padding-bottom:0;padding-right:0]@media (min-width:760px){.RiverHeadline-hasThumbnail[padding-bott  
om:16px]}@media (min-width:1340px){.RiverHeadline-hasThumbnail[padding-bottom:8p  
x]}.RiverPlusBreaker-container[border-bottom:1px solid #dedede;display:flex;just  
ify-content:space-around].RiverPlus-riverPlusContainer[padding-bottom:30px]  
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$
```

Additionally, here is the output of running **web_scaper.py** on my Terminal.

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$ python3 web_scaper.py
Fetching data from CNBC
Data was successfully fetched!
Data saved to ../data/raw_data/web_data.html
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$
```

And here is first 5 rows of the resulting HTML file (**web_data.html**), which I saw by running the command: `head -n 5 web_data.html`

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/data/raw...
~/Riya_7828642721/data/raw_data

Wrapper-description[max-width:520px]}.CNBCHomepageProWrapper-description a[color :#002f6c].CNBCHomepageProWrapper-description a:hover{color:#fcfc700}.CNBCHomepage ProWrapper-iframeContainer[border:none;height:470px;width:100%].CNBCHomepageProW rapper-proLiveCollapsibleContainer[align-items:center;background:linear-gradient(181.49deg,#171717 1.27%,#17171700 427.68%);border-top-left-radius:4px;border-top-right-radius:4px;box-shadow:5px 5px 20px 0 rgba(0,0,0,.251);display:flex}@media (min-width:1020px){.CNBCHomepageProWrapper-proLiveCollapsibleContainer{float:left;height:52px;width:calc(45% - 15px)}}@media (min-width:1340px){.CNBCHomepageProWrapper-proLiveCollapsibleContainer{width:calc(50% - 15px)}}.CNBCHomepageProWr apper-proLiveCollapsedContainer[border-radius:4px].CNBCHomepageProWrapper-liveIndicat orContainer[color:#ffff;display:flex;font-size:14px;font-weight:700;left:20px;line-height:16px;position:relative].CNBCHomepageProWrapper-liveIndicator[background:#e80101;border-radius:50%;content:"";height:8px;margin-right:6px;opacity:1;position:relative;top:4px;transition:opacity .15s cubic-bezier(0,.25,.25,1);width:8px].CNBCHomepageProWrapper-proLivePlayerCloseOrExpand[cursor:pointer;margin-left:auto;position:relative;right:20px;transition:all .1s linear].CNBCHomepagePr oWrapper-dismissButton[height:16px;pointer-events:none;width:16px].CNBCHomepagePr oWrapper-proLivePlayerClosed[transform:rotate(45deg)].LegacyPlayerContainer-leg acyPlayerContainer[background-color:#ff7f7f7;margin-bottom:30px;padding:40px;widt h:100vw]@media (min-width:1340px){.LegacyPlayerContainer-legacyPlayerContainer{max-width:none}}@media (min-width:760px){.LegacyPlayerContainer-legacyPlayerContainer{padding:40px 20px}}@media (min-width:1020px){.LegacyPlayerContainer-legacyPlayerContainer{margin-bottom:40px}}
```

Lastly, I created **data_filter.py** script to extract market banner data into a CSV named “**market_data.csv**”, and to extract latest news data into a csv named “**news_data.csv**.” I also printed relevant messages to the console to help users understand what was happening in the back end (filtering fields, creating csv, extracting data, etc..)

```
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$ python3 data_filter.py
Reading web_data.html
Extracting market data ...
Storing market data to csv
Storing news data to csv
CSV created
Filtering complete!
CSV files created successfully.
vboxuser@RiyaUbuntu:~/Riya_7828642721/scripts$
```

Task 3:

I will be working in a team of 3 this semester, with Ben Lu and Ryan T.