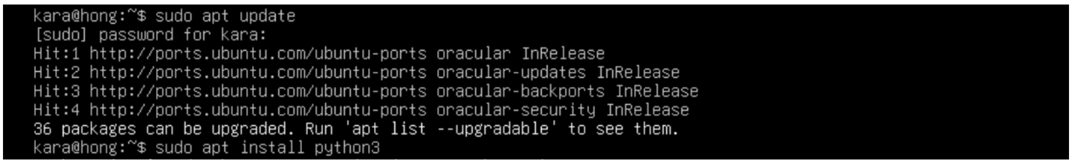
Lab 1
Yiqing Hong
USC ID: 4395913002

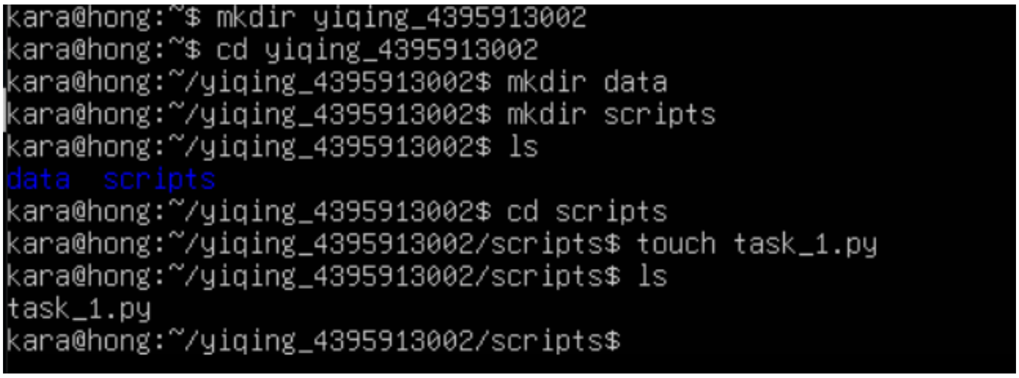
Installation and Setup



```
kara@hong:~$ sudo apt update
[sudo] password for kara:
Hit:1 http://ports.ubuntu.com/ubuntu-ports oracular InRelease
Hit:2 http://ports.ubuntu.com/ubuntu-ports oracular-updates InRelease
Hit:3 http://ports.ubuntu.com/ubuntu-ports oracular-backports InRelease
Hit:4 http://ports.ubuntu.com/ubuntu-ports oracular-security InRelease
36 packages can be upgraded. Run 'apt list --upgradable' to see them.
kara@hong:~$ sudo apt install python3
```

I installed VMware and set up Ubuntu terminal. I installed and updated python packages using “sudo apt update”, “sudo apt install python3”, “sudo apt install python3-pip”.

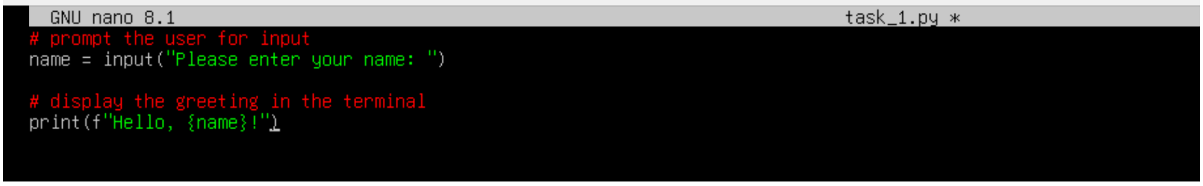
Playing around with Linux Terminal



```
kara@hong:~$ mkdir yiqing_4395913002
kara@hong:~$ cd yiqing_4395913002
kara@hong:~/yiqing_4395913002$ mkdir data
kara@hong:~/yiqing_4395913002$ mkdir scripts
kara@hong:~/yiqing_4395913002$ ls
data  scripts
kara@hong:~/yiqing_4395913002$ cd scripts
kara@hong:~/yiqing_4395913002/scripts$ touch task_1.py
kara@hong:~/yiqing_4395913002/scripts$ ls
task_1.py
kara@hong:~/yiqing_4395913002/scripts$
```

I created a directory by “mkdir yiqing_4395913002”. Inside the folder, I created two subdirectories by “mkdir data” and “mkdir scripts”. Inside scripts, I created a Python file by “touch task_1.py”. I used “ls” to see the created file and directories.

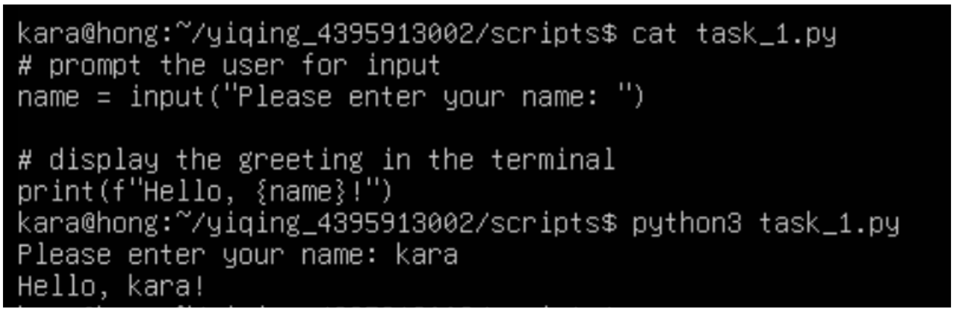
A basic Python Script



```
# prompt the user for input
name = input("Please enter your name: ")

# display the greeting in the terminal
print(f"Hello, {name}!")
```

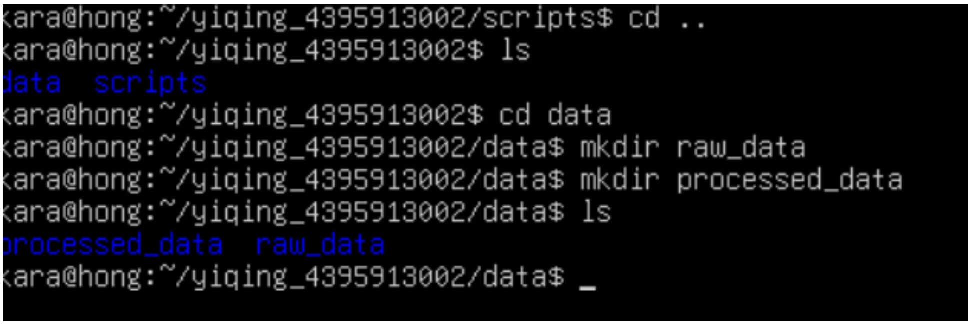
I used nano to edit the Python file and execute the code.



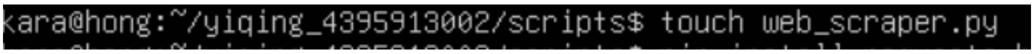
```
kara@hong:~/yiqing_4395913002/scripts$ cat task_1.py
# prompt the user for input
name = input("Please enter your name: ")

# display the greeting in the terminal
print(f"Hello, {name}!")
kara@hong:~/yiqing_4395913002/scripts$ python3 task_1.py
Please enter your name: kara
Hello, kara!
```

Python Web-scraping Task

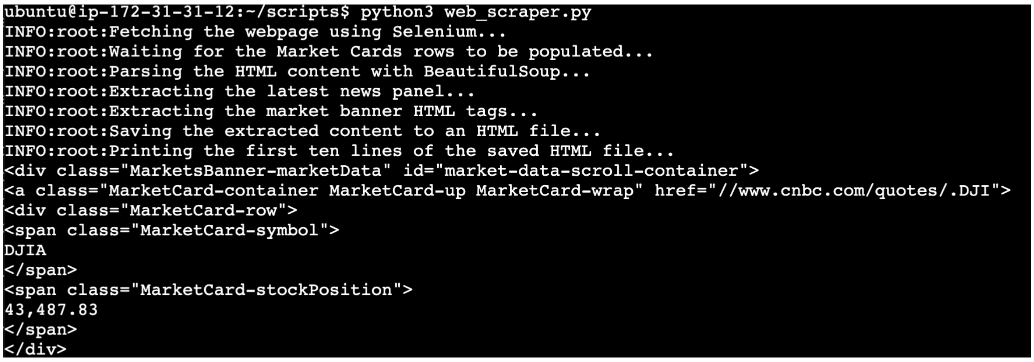


```
kara@hong:~/yiqing_4395913002/scripts$ cd ..
kara@hong:~/yiqing_4395913002$ ls
data  scripts
kara@hong:~/yiqing_4395913002$ cd data
kara@hong:~/yiqing_4395913002/data$ mkdir raw_data
kara@hong:~/yiqing_4395913002/data$ mkdir processed_data
kara@hong:~/yiqing_4395913002/data$ ls
processed_data  raw_data
kara@hong:~/yiqing_4395913002/data$ _
```



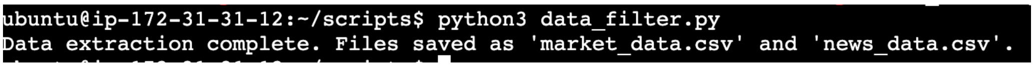
```
kara@hong:~/yiqing_4395913002/scripts$ touch web_scraper.py
```

I created a new file web_scraper.py in the scripts folder. And I used “sudo apt install python3-requests python3-beautifulsoup4” to install packages. I also created files like raw_data and processed_data. I wrote the scraper and ran it on AWS EC2. I printed out first 10 lines of the created html file.



```
ubuntu@ip-172-31-31-12:~/scripts$ python3 web_scraper.py
INFO:root:Fetching the webpage using Selenium...
INFO:root:Waiting for the Market Cards rows to be populated...
INFO:root:Parsing the HTML content with BeautifulSoup...
INFO:root:Extracting the latest news panel...
INFO:root:Extracting the market banner HTML tags...
INFO:root:Saving the extracted content to an HTML file...
INFO:root:Printing the first ten lines of the saved HTML file...
<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>
```

Data Filtering Task



```
ubuntu@ip-172-31-31-12:~/scripts$ python3 data_filter.py
Data extraction complete. Files saved as 'market_data.csv' and 'news_data.csv'.
```

I also wrote the data_filter.py to processed the html file I got by scraper and stored the processed data as csv files. The Python files and processed data files were uploaded on GitHub.