**Lab 1 Report**

Yunlin Peng UID: 2396710607

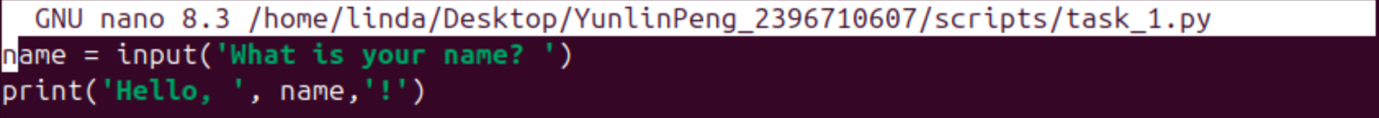
**2.1 Playing around with Linux Terminal**

A screen shot of a computer

AI-generated content may be incorrect.

The command *mkdir* creates a new directory, so by using it, I create a new folder named YunlinPeng\_2396710607 and two subfolders named data and scripts inside it. I also use the command *touch* to create a file named task\_1.py inside the scripts folder. The *ls* command lists out all directories under the scripts folder, and it shows that the task\_1.py file is created successfully.

**2.2 A basic Python Script**

A computer screen with white text

AI-generated content may be incorrect.

I first use command *nano* to open the task\_1.py file, and then I use input function to ask the user to enter the name and then print out the name. Finally, I run this task\_1.py file in the terminal, and it asks me to enter my name and returns “Hello, Linda!”.

**2.3 Python Web-scraping task**

The website we need to collect data from in this task is CNBC webpages, and there are dynamic blocks on the webpage that are scripted by JavaScript. In order to get dynamic data from the webpage, I install a package named selenium in Linux, which works well for getting dynamic data from the Chrome browser. Hence, in my script, I first set up the driver for this, then collect dynamic data from the webpage using selenium, and collect static data from the webpage using packages requests and beautifulSoup, and finally save the data to an HTML file and print out first 10 lines of it. In order to install selenium, I also create a virtual environment in Linux and install and run all scripts inside this virtual environment.

A screen shot of a computer screen

AI-generated content may be incorrect.

The above snapshot is the HTML format of the web\_data.html. The top line contains dynamic data from the webpage which is associated with the market banner section on the webpage. Other lines are related with the latest news section on the webpage.

A computer screen with white text

AI-generated content may be incorrect.

The above snapshot shows the first 10 lines printed on the terminal in Linux.

**2.4 Data Filtering Task**

In this task, I need to extract market banner related information and latest news related information from web\_data.html. First, I read web\_data.html as a lists of items, and then I use empty space to join these lines, and convert it to beautifulSoup format so that I can use the beautifulSoup package to directly extract specific information. Finally, I save market related data and latest news related data in two separated CSV files.

A screen shot of a computer

AI-generated content may be incorrect.

The snapshot above shows the process information I print out during this extraction program.

A black background with white text

AI-generated content may be incorrect.

The snapshot above shows the market CSV file I created which contains the market symbol, stock position, and its changePct.

A screen shot of a computer screen

AI-generated content may be incorrect.

The snapshot above shows the latest news CSV file I created which contains the latest news timestamp, title, and link.