**HW 12 Part 2**

**Learning Experience**

**Part 1:**

The homework of this week 12 was a great introduction to Web Scrapping and manipulation HTML files. We use the Beautiful Soup library to use BeautifulSoup for parsing the page and searching for specific elements and retrieve its data. HTML is a most common file used for storing the structure of a webpage which can also include text, links to images and videos. Since I have gotten some experience using HTML files. I was able to manipulate them with using the Beautiful Soup and extract some data. For the first enhancement, I chose Wikipedia website to look for the titles of the webpage, which talk about Python programming language.

Text

Description automatically generated

However, for the second enhancement, I decided to use another website “What is Python? The popular, scalable programming language, explained,” from <https://www.businessinsider.com/what-is-python> and try to extract some data. I chose to extract the title and subtitles of this webpage. Since the title of a webpage is store in the <h1> and <h2>, I use ‘h1’ and ‘h2’ in the function find\_all to get all possible titles and subtitles from this webpage. The following code shows:

For titles:

Text

Description automatically generated

The output is:



For subtitles:

Text

Description automatically generated

The output is:

Text

Description automatically generated with medium confidence

Overall, the assignment helped me out a lot to understand how to access HTML files from websites and how to web scrap them to access specific data and extract them.