1-a

List the libraries or framework you have used creating a python web crawler. In 200 words or less, describe the working flow.

Libraries/Frameworks: requests, urllib, BeautifulSoup, Parsel, Scrapy, Selenium, lxml, MechanicalSoup, mechanize, aiohttp, html5lib, MarkupSafe, xhtml2pdf, Bleach, sanitize, fake-useragent, cssutils.

I investigate the target source of the scraping first. For instance, if the target we need to scrape is a website I will analyze the website structure and try to understand how it is organized. Then I will look for possible scraping traps and hurdles, namely anti-scraping mechanisms and IP restrictions. The info that we want to gather is the main asset, so I need to be sure that is accessible to us, first of all. Then I will put together a small scraping prototype to test my assumptions and later on start building the main scraper itself.

1-b

Explain difference between python list and tuple.

While a list is mutable, a tuple is immutable. Tuples are heterogeneous data structures, while lists are homogeneous sequences. Tuples have structure, lists have order.