Question 1

In [1]:
```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get('https://www.linkedin.com/jobs/search/?currentJobId=3532460530'
time.sleep(5)
i=1
roles=[]
try:
    while 1:
        roles.append(driver.find_element(By.XPATH, '//*[@id="main-content"
        i=i+1
except:
    print("List of roles")
    for i in roles:
        print(i)
driver.close()
```

```
List of roles
Production Operator- Nokian Tyres
Clinical Research Associate - Remote
Laboratory Assistant
Warehouse Worker
Press Helper
Food and Beverage Supervisor
Guest Service Attendant
Shipping & Receiving Professional
Junior Camp Counselor - South City YMCA
General Manager
Dishwasher
Assistant Superintendent
General Laborer
Core Technician
CAD Drafter - Entry Level
Multi-Store Supervisor - #655 - Evenings/Overnight
Picker/Packer 1st Shift
Microbiology Lab Assistant
Greenskeeper
Remote Data Entry Clerk No Experience
Greenskeeper
Patient Sitter
Janitor/Day Porter/Cleaner
Food Runner (FT/PT/Seasonal)
Museum Technician.
```

In [8]:

```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get('https://www.linkedin.com/jobs/search/?currentJobId=3532460530'
time.sleep(5)
i=1
company=[]
try:
    while 1:
        company.append(driver.find_element(By.XPATH, '//*[@id="main-conten
        i=i+1
except:
    print("List of Companies")
    for i in company:
        print(i)
driver.close()
```

```
List of Companies
Alabama Department of Corrections
Alabama Department of Corrections
Northwell Health
Tao Group Hospitality
Cigna Healthcare
Northwell Health
Vivian Health
Vivian Health
Lifecare
Cresa
Vivian Health
Cresa
Cresa
Sunstates Security
ARO Liquidation Inc
Cresa
United States Postal Service
University of Houston
The Brothers that just do Gutters
Kittitas Valley Healthcare
Cresa
WorkFello
Northwell Health
Northwell Health
```

In [3]:
```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get('https://www.linkedin.com/jobs/search/?currentJobId=3532460530'
time.sleep(5)
i=1
location=[]
try:
    while 1:
        location.append(driver.find_element(By.XPATH, '//*[@id="main-conte
        i=i+1
except:
    print("List of locations")
    for i in location:
        print(i)
driver.close()
```

```
List of locations
United States
Lubbock, TX
Baton Rouge, LA
Bozeman, MT
Kent, WA
Princeton, NJ
Palm Desert, CA
Little Rock, AR
Estero, FL
Washington, DC
Rancho Santa Margarita, CA
Shreve, OH
Sonora, CA
Bridgewater, NJ
Jurupa Valley, CA
Bend, OR
New York, NY
Dallas, TX
Southampton, PA
Las Vegas, NV
Yonkers, NY
Bradenton, FL
Goodlettsville, TN
Estero, FL
Dover, DE
```

In [4]:
```
# Ranking of the company is not available
```

In [5]:

```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get('https://www.linkedin.com/jobs/search/?currentJobId=3532460530'
time.sleep(5)
i=1
time=[]
try:
    while 1:
        time.append(driver.find_element(By.XPATH, '//*[@id="main-content"]
        i=i+1
except:
    print("List of time")
    for i in time:
        print(i)
driver.close()
```

```
List of time
2023-04-09
2023-04-09
2023-03-31
2023-04-07
2023-04-09
2023-03-21
2023-04-07
2023-04-09
2023-03-15
2023-04-09
2023-03-15
2023-04-07
2023-03-15
2023-03-20
2023-04-09
2023-03-24
2023-04-09
2023-04-09
2023-04-09
2023-04-09
2023-04-09
2023-03-15
2023-04-09
2023-03-20
2023-04-02
```

In [6]:
```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get('https://www.linkedin.com/jobs/search/?currentJobId=3532460530'
time.sleep(5)
i=1
time=[]
location=[]
company=[]
roles=[]
rows=[]
try:
    while 1:
        location.append(driver.find_element(By.XPATH, '//*[@id="main-conte
        time.append(driver.find_element(By.XPATH, '//*[@id="main-content"]
        company.append(driver.find_element(By.XPATH, '//*[@id="main-conten
        roles.append(driver.find_element(By.XPATH, '//*[@id="main-content"
        i=i+1
except:
#     print("List of time")
    for i in range(len(time)):
        rows.append([roles[i],company[i],location[i],"",time[i]])
driver.close()
```

In [8]:
```
fields=["Title","Company","Location","Rating","Date Posted"]
import csv
with open('job_1.csv', 'w') as f:
    write = csv.writer(f)
    write.writerow(fields)
    write.writerows(rows)
f.close()
```

Question 2

In [14]:

```
import scrapy
from scrapy.crawler import CrawlerRunner
from crochet import setup, wait_for
setup()
class JobItem(scrapy.Item):
    title = scrapy.Field()
    company = scrapy.Field()
    location = scrapy.Field()


class PythonDocumentationSpider(scrapy.Spider):
    name = 'pydoc_bot'
    allowed_domans=['linkedin.com']
    start_urls = ['https://www.linkedin.com/jobs/search/?currentJobId=3532
    custom_settings = {
                'FEEDS': {
            'job_2.csv': {
                'format': 'csv',
                'overwrite': True
            }
        }
    }

    def parse(self, response):
        i=1

        try:
            while 1:
                section= JobItem()
                section['title'] = response.xpath('//*[@id="main-content"]
                section['company'] = response.xpath('//*[@id="main-content
                section['location']= response.xpath('//*[@id="main-content
#                 print(section['company'])
                yield(section)
                i=i+1
        except:
            i=i+1


@wait_for(10)
def run_spider():
    crawler = CrawlerRunner()
    d = crawler.crawl(PythonDocumentationSpider)
    return d
run_spider()
```

Question 3

In [20]:

```
import scrapy
from scrapy.crawler import CrawlerRunner
from crochet import setup, wait_for
setup()

class JobItem(scrapy.Item):
    title = scrapy.Field()
    company = scrapy.Field()
    location = scrapy.Field()


class PythonDocumentationFollowingSpider(scrapy.Spider):
    name = 'pydoc_bot'
    allowed_domans=['linkedin.com']
    start_urls = ['https://www.linkedin.com/jobs/search/?currentJobId=3532
    custom_settings = {
                'FEEDS': {
            'job_3.csv': {
                'format': 'csv',
                'overwrite': True
            }
        }
    }
    def parse(self, response):

        follow_url = self.start_urls[0]
        yield scrapy.Request(follow_url, callback=self.parse_page_title)

    def parse_page_title(self, response):
        i=1

        try:
            while 1:
                section= JobItem()
                section['title'] = response.xpath('//*[@id="main-content"]
                section['company'] = response.xpath('//*[@id="main-content
                section['location']= response.xpath('//*[@id="main-content
                yield(section)
                i=i+1
        except:
            i=i+1
@wait_for(10)
def run_spider():
    crawler = CrawlerRunner()
    d = crawler.crawl(PythonDocumentationFollowingSpider)
    return d
run_spider()
```

In [ ]: