```python
from google.colab import files
from google.colab import auth
from google.colab import drive
auth.authenticate_user()
drive.mount('/content/drive')
```

```
    Mounted at /content/drive
```

```python
!pip install pip

import pip

for each in ["requests","beautifulsoup4"]:
    pip.main(['install', each])

import requests
from bs4 import BeautifulSoup
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: pip in /usr/local/lib/python3.8/dist-packages (21.1.3)
WARNING: pip is being invoked by an old script wrapper. This will fail in a future version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the underlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running pip directly.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: requests in /usr/local/lib/python3.8/dist-packages (2.23.0)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.8/dist-packages (from requests) (1
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.8/dist-packages (from requests) (2022.9.24)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.8/dist-packages (from requests) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.8/dist-packages (from requests) (3.0.4)
WARNING: pip is being invoked by an old script wrapper. This will fail in a future version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the underlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running pip directly.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.8/dist-packages (4.6.3)
```

```python
#scraping indeed for positions according to candidate profile
import datetime, time
!pip install wget
import wget

import csv, re
import logging

import pandas as pd
import sqlite3
import json
from pytz import timezone

from requests.auth import HTTPBasicAuth
import requests
from requests.exceptions import HTTPError
import traceback
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting wget
  Downloading wget-3.2.zip (10 kB)
Building wheels for collected packages: wget
  Building wheel for wget (setup.py) ... done
  Created wheel for wget: filename=wget-3.2-py3-none-any.whl size=9674 sha256=a9e31dbfa5af116e92b5058b1c74b8ca3722c9bb922671efe0948
  Stored in directory: /root/.cache/pip/wheels/bd/a8/c3/3cf2c14a1837a4e04bd98631724e81f33f462d86a1d895fae0
Successfully built wget
Installing collected packages: wget
Successfully installed wget-3.2
NumExpr defaulting to 2 threads.
```

```python
client_key = "77zk7frrida7k3"
client_secret = "CCISHQyTUjP3xqvZ"


job_positions = ["data%20scientist%20intern","software%20engineering%20intern","Devops%20intern","Software%20Engineer","Data%20Analyst",'
headers = {'Accept': 'application/json'}
auth = HTTPBasicAuth(client_key, client_secret)
job_url_template = "https://www.linkedin.com/jobs/view/"
job_df = pd.DataFrame(columns=["job_title","industry","company","description","job_locality","job_region","skills","job_post_expiry","edu

list_of_links = []
try:
```

```
        s = requests.Session()

    for position in job_positions:
        for i in range(0,500,25):
        #for i in range(0,1):
            linkedin_url = 'https://www.linkedin.com/jobs-guest/jobs/api/seeMoreJobPostings/search?keywords='+position+\
                            '&geoId=103644278&trk=public_jobs_jobs-search-bar_search-submit&position=8&pageNum=0&currentJobId=32578844338

            time.sleep(1)
            html_text = s.get(linkedin_url, headers=headers).text
            soup = BeautifulSoup(html_text, "html.parser")
            count=0
            for meat in soup.find_all('a'):
                if (meat.get('href') != None):
                    if(job_url_template in meat.get('href')):
                        list_of_links.append(meat.get('href'))
                        count+=1
    print(len(list_of_links))
except HTTPError as http_err:
    print(f'HTTP error occurred: {http_err}')
except Exception as err:
    print(f'Other error occurred: {err}')

    2997


job_data = []
print(len(list_of_links))
for i in range(0,len(list_of_links)):
    job_req = requests.Session()
    job_url=list_of_links[i]
    try:
        job_text = job_req.get(job_url, headers=headers, auth=auth).text
        job_soup = BeautifulSoup(job_text, "html.parser")
        job_desc_soup = BeautifulSoup(job_text, "lxml")
        job_details_content = job_soup.find('script',  {"type": "application/ld+json"})
        job_desc = job_desc_soup.find('div', class_='description__text')

        if(job_details_content and job_details_content.text):
            job_json = json.loads(job_details_content.text)
            job_details = [job_json.get('title'), job_json.get('industry')]

            if(job_json.get('identifier')):
                job_details.append(job_json.get('identifier').get('name'))
            else:
                job_details.append(None)

            if(job_desc):
                job_details.append(job_desc.text.replace("Show more"," ").replace("Show less", " "))
            else:
                job_details.append(None)

            if('jobLocation' in job_json and 'address' in job_json.get('jobLocation')):
                job_details.extend( [job_json.get('jobLocation').get('address').get('addressLocality'),
                                    job_json.get('jobLocation').get('address').get('addressRegion')])
            else:
                job_details.extend([None,None])

            job_details.extend( [job_json.get('skills'), job_json.get('validThrough')])

            if('educationRequirements' in job_json):
                job_details.append(job_json.get('educationRequirements').get('credentialCategory'))
            else:
                job_details.append(None)

            if('baseSalary' in job_json and 'value' in job_json.get('baseSalary')):
                job_details.extend([job_json.get('baseSalary').get('value').get('minValue'),
                        job_json.get('baseSalary').get('value').get('maxValue'), job_json.get('baseSalary').get('currency')])
            else:
                job_details.extend([None,None,None])
            job_details.append(job_url)

            job_data.append(job_details)
            time.sleep(1.25)
        else:
            print(job_url)
    except Exception as e:
        logging.error(traceback.format_exc())

    2997
    https://www.linkedin.com/jobs/view/data-scientist-at-pomeroy-3384652177?refId=rB2AOc0hyuwFgg3vO5QRPg%3D%3D&trackingId=0hQTD%2B2F
    https://www.linkedin.com/jobs/view/data-scientist-full-time-at-bardess-group-ltd-3389875542?refId=BV5TqViF4IDtWAggaunTqw%3D%3D&tr
    https://www.linkedin.com/jobs/view/data-scientist-at-r-d-partners-3384656781?refId=BV5TqViF4IDtWAggaunTqw%3D%3D&trackingId=Waevn1
```

https://www.linkedin.com/jobs/view/data-scientist-at-hireresources-3384151389?refId=BV5TqViF4IDtWAggaunTqw%3D%3D&trackingId=0Sn6O
https://www.linkedin.com/jobs/view/python-data-scientist-at-brains-workgroup-3395566644?refId=P85ov0M%2B4Tnt8o8jAr4Hkg%3D%3D&trac
https://www.linkedin.com/jobs/view/machine-vision-engineer-at-w-h-leary-co-inc-3395534646?refId=LG1B5qge00RMJ%2BIvm525ew%3D%3D&tr
https://www.linkedin.com/jobs/view/data-scientist-cost-analyst-at-quest-consulting-inc-3382643925?refId=Qczps%2BwTp6REK0TrY%2F%2F
https://www.linkedin.com/jobs/view/software-engineer-summer-intern-2023-at-staples-3385013751?refId=cMp0u4wqS53ilCG8%2BjfFVA%3D3
https://www.linkedin.com/jobs/view/software-engineer-internship-jan-march-at-arivo-acceptance-3368897330?refId=OpGQZkdVeLssSbYGSu
https://www.linkedin.com/jobs/view/software-engineer-intern-at-cubrc-3361176547?refId=EcV%2BkC%2FypOgXJTiZf0xjzA%3D%3D&trackingId
https://www.linkedin.com/jobs/view/aws-devops-admin-at-avance-consulting-3385121117?refId=pe60mf5Q54iZc0B5lFBxyg%3D%3D&trackingId
https://www.linkedin.com/jobs/view/cloud-engineer-at-finlocker-3395490904?refId=pe60mf5Q54iZc0B5lFBxyg%3D%3D&trackingId=%2FXJhPts
https://www.linkedin.com/jobs/view/sql-developer-intern-at-custom-computer-specialists-3394347972?refId=pe60mf5Q54iZc0B5lFBxyg%3D
https://www.linkedin.com/jobs/view/cloud-operations-engineer-at-veterans-sourcing-group-llc-3395533898?refId=JxdTA84TWvjwFB%2FYdO
https://www.linkedin.com/jobs/view/cloud-security-engineer-at-trapp-technology-3394077063?refId=jo1ISWKv1p8iLgbSofGRIQ%3D%3D&trac
https://www.linkedin.com/jobs/view/it-applications-engineer-at-globe-life-3395492896?refId=jo1ISWKv1p8iLgbSofGRIQ%3D%3D&trackingI
https://www.linkedin.com/jobs/view/it-applications-engineer-at-globe-life-3395492896?refId=q5H6dRPmBqWqF3aZ62RbsQ%3D%3D&trackingI
https://www.linkedin.com/jobs/view/cloud-infrastructure-engineer-at-opspro-llc-3395456417?refId=L2Czj2oSXte0IEkQdoEH%2Fw%3D%3D&tr
https://www.linkedin.com/jobs/view/cloud-infrastructure-engineer-at-opspro-llc-3395456417?refId=ad0tevDvRjyzHOKUBk0QVw%3D%3D&trac
https://www.linkedin.com/jobs/view/systems-engineer-at-sja-solutions-3395494595?refId=gJlBMWFF9oG0xkhbyqGxpw%3D%3D&trackingId=hB2
https://www.linkedin.com/jobs/view/junior-software-developer-at-cagenix-3395457398?refId=mo0%2F6yPLnH4ZG1nVXOt2ag%3D%3D&trackingId
https://www.linkedin.com/jobs/view/associate-cloud-engineer-at-umass-chan-medical-school-3395594737?refId=ENdgciUMngyQSiiN3YaVmw%
https://www.linkedin.com/jobs/view/jr-software-engineer-at-tellabs-3382702753?refId=vuvnGNNCkmIbUdtpKqghpw%3D%3D&trackingId=BIA6B
https://www.linkedin.com/jobs/view/software-engineer-at-foxconn-industrial-internet-3394335327?refId=S8Jfn%2FpX4iWTqhGCFuDicQ%3D%
https://www.linkedin.com/jobs/view/software-engineer-at-productive-robotics-llc-3385802893?refId=S8Jfn%2FpX4iWTqhGCFuDicQ%3D%3D&t
https://www.linkedin.com/jobs/view/software-engineer-at-lightriver-companies-3382651454?refId=4mrE2QBT0mPd%2B13CO90LvA%3D%3D&trac
https://www.linkedin.com/jobs/view/software-engineer-at-tribalco-3395458254?refId=4mrE2QBT0mPd%2B13CO90LvA%3D%3D&trackingId=TSWWY
https://www.linkedin.com/jobs/view/software-engineer-at-nagrastar-3384330203?refId=4mrE2QBT0mPd%2B13CO90LvA%3D%3D&trackingId=GvnL
https://www.linkedin.com/jobs/view/software-engineer-at-impact-consulting-solutions-inc-3382908427?refId=4mrE2QBT0mPd%2B13CO90LvA
https://www.linkedin.com/jobs/view/software-engineer-c%23-net-at-darbytek-3389892452?refId=4mrE2QBT0mPd%2B13CO90LvA%3D%3D&trackin
https://www.linkedin.com/jobs/view/software-engineer-at-rsi-visuals-3394061760?refId=3JgqkNSezkdaMS9v0Djcsw%3D%3D&trackingId=u5h%
https://www.linkedin.com/jobs/view/software-engineer-at-enhance-it-3390256067?refId=3JgqkNSezkdaMS9v0Djcsw%3D%3D&trackingId=oZSz6
https://www.linkedin.com/jobs/view/software-engineer-at-omnitech-inc-3395460187?refId=3JgqkNSezkdaMS9v0Djcsw%3D%3D&trackingId=wiO
https://www.linkedin.com/jobs/view/software-engineer-at-omnitech-inc-3395460187?refId=8VkvWOwgRwPAzZdW2ukIkA%3D%3D&trackingId=x0P
https://www.linkedin.com/jobs/view/software-developer-at-information-technology-3395496313?refId=8VkvWOwgRwPAzZdW2ukIkA%3D%3D&tra
https://www.linkedin.com/jobs/view/software-developer-backend-at-certedrive-3395491498?refId=FarYanI9SHtQs6y6yWAz9A%3D%3D&trackin
https://www.linkedin.com/jobs/view/back-end-software-engineer-at-healium-3394334470?refId=hSxhJs91qJ%2FgKhZ7ALhOoQ%3D%3D&tracking
https://www.linkedin.com/jobs/view/software-engineer-backend-application-engineer-at-ashael-tek-solutions-llc-3395451531?refId=Wa
https://www.linkedin.com/jobs/view/software-application-engineer-hybrid-at-devonway-3394406604?refId=X6G0Lnya1cyW6RigrtoN8A%3D%3D
https://www.linkedin.com/jobs/view/full-stack-software-engineer-at-avirtek-3390024233?refId=y8eWg4vLp87YoH3sJx8Z6A%3D%3D&tracking
https://www.linkedin.com/jobs/view/software-engineer-at-staff-experts-llc-3387515528?refId=%2BYRhVyXvOscyqLY073yU1Q%3D%3D&trackin
https://www.linkedin.com/jobs/view/software-engineer-at-staff-experts-llc-3387515528?refId=BFKPsVUl5SraRzM46DT6Tw%3D%3D&trackingI
https://www.linkedin.com/jobs/view/java-developer-software-engineer-at-enhance-it-3388668654?refId=IN4W7cVHIB61AslSljJptg%3D%3D&t
https://www.linkedin.com/jobs/view/software-development-engineer-at-the-job-connection-inc-3395538518?refId=IN4W7cVHIB61AslSljJpt
https://www.linkedin.com/jobs/view/software-developer-at-quality-manufacturing-systems-inc-qmsi-3395513146?refId=4bSZqyV7iGfeex3g
https://www.linkedin.com/jobs/view/c%23-software-engineer-at-black-swan-search-3385772333?refId=4bSZqyV7iGfeex3g2jBPaA%3D%3D&trac
https://www.linkedin.com/jobs/view/software-engineer-front-end-at-ashael-tek-solutions-llc-3395455101?refId=Z9gJzrghI0vJjpj0KKKcS
https://www.linkedin.com/jobs/view/data-analyst-at-surge-staffing-3395492666?refId=DkWbLn8n%2BJjxXNA9FwZy6Q%3D%3D&trackingId=dx1r
https://www.linkedin.com/jobs/view/data-analyst-at-bluewave-resource-partners-3385812728?refId=I77hvCsPIvaa%2FyldXa5iWQ%3D%3D&tra
https://www.linkedin.com/jobs/view/data-analyst-entry-level-at-flexon-technologies-inc-3387396862?refId=I77hvCsPIvaa%2FyldXa5iWQ%
https://www.linkedin.com/jobs/view/entry-level-data-analyst-at-asta-crs-3395541204?refId=I77hvCsPIvaa%2FyldXa5iWQ%3D%3D&track
https://www.linkedin.com/jobs/view/data-analyst-at-paradigm-3394407665?refId=ISVef0cGc0sDcnTeIxAzYw%3D%3D&trackingId=Q95kX4r4NOIV
https://www.linkedin.com/jobs/view/data-analyst-i-at-tech-providers-inc-3388709415?refId=GHeIi8J4W6VxY4UMDAJ80g%3D%3D&trackingId=
https://www.linkedin.com/jobs/view/data-analyst-remote-at-aston-carter-3395514800?refId=H7HCbsVzQc5JGwM%2FsaI78A%3D%3D&trackingId

```
job_df = pd.DataFrame(job_data,
                      columns=["job_title","industry","company","description", "job_locality","job_region","skills","job_post_expiry","e

job_df
```

|   | job_title | industry | company | description |
|---|---|---|---|---|
| 0 | Data Scientist Internship - OR (US) | Insurance | Asurion | \n\n\n 2022-75202Asurion Data Science I... |
| 1 | Intern - Data Scientist | Medical Equipment Manufacturing | Dexcom | \n\n\nAbout DexcomDexcom, Inc. empowers people... |
| 2 | Data Science, Graduate Internship | Retail | Lowe's Companies, Inc. | \n\n\nLowe's Summer Internship Program Overvie... |
| 3 | Data Scientist Internship - ML (US) | Insurance | Asurion | \n\n\n 2022-75205Asurion Data Science I... |

```
job_df['description']
```

```
0       \n\n\n        2022-75202Asurion Data Science I...
1       \n\n\nAbout DexcomDexcom, Inc. empowers people...
2       \n\n\nLowe's Summer Internship Program Overvie...
3       \n\n\n        2022-75205Asurion Data Science I...
4       \n\n\n        We're a leader in the pet care i...
                              ...
2930    \n\n\nDescriptionData EngineerWe are currently...
2931    \n\n\nJob DescriptionPerform Data Engineering ...
2932    \n\n\nJob DescriptionData model development an...
2933    \n\n\nAbout NeudesicPassion for technology dri...
2934    \n\n\n        Evaluating existing data systems...
Name: description, Length: 2935, dtype: object
```

```
`  `  `  `  `  `  `
```

```
print(job_data)
```

```
IOPub data rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_data_rate_limit`.

Current values:
NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
NotebookApp.rate_limit_window=3.0 (secs)
```

```
job_df = pd.DataFrame(job_data,
                      columns=["job_title","industry","company","description", "job_locality","job_region","skills","job_post_expiry","e
outfile="jobs_data"
outfile=outfile+'.csv'
print(outfile)
job_df.to_csv(outfile, sep=',', encoding='utf-8')
```

```
job_df
```

jobs_data.csv

|   | job_title | industry | company | description |
|---|---|---|---|---|
| 0 | Data Scientist Internship - OR (US) | Insurance | Asurion | \n\n\n 2022-75202Asurion Data Science I... |
| 1 | Intern - Data Scientist | Medical Equipment Manufacturing | Dexcom | \n\n\nAbout DexcomDexcom, Inc. empowers people... |
| 2 | Data Science, Graduate Internship | Retail | Lowe's Companies, Inc. | \n\n\nLowe's Summer Internship Program Overvie... |
| 3 | Data Scientist Internship - ML (US) | Insurance | Asurion | \n\n\n 2022-75205Asurion Data Science I... |
| 4 | Data Science | Manufacturing | Nestlé Purina North | \n\n\n We're a leader |

```
!cp "/content/jobs_data.csv" "/content/drive/MyDrive/jobs_data.csv"
print("Uploaded")
```

Uploaded

|   | job_title | industry | company | description |
|---|---|---|---|---|
| 2930 | Engineer (AWS, Python, SQL) | Staffing and Recruiting | Cambay Healthcare, LLC | \n\n\nDescriptionData EngineerWe are currently... |
| 2931 | Data Engineer | Software Development | Diverse Lynx | \n\n\nJob DescriptionPerform Data Engineering ... |
|  | Data | Software | | \n\n\nJob DescriptionData |

Colab paid products  -  Cancel contracts here