

```
In [1]: #importing libraries
from selenium import webdriver
from bs4 import BeautifulSoup as s
import pandas as pd
from webdriver_manager.chrome import ChromeDriverManager
```

```
In [2]: #web driver
driver = webdriver.Chrome('chromedriver')
```

```
In [4]: #Load all pages and extract data
data=[]
b=[]
data1=[]
driver.get("https://www.payscale.com/research/IN/Job")
soup = s(driver.page_source, 'html.parser')
for i in soup.find_all('a',class_='related-content-card'):
    if i.has_attr('href'):
        a=""
        a="https://www.payscale.com"+i['href']
        data.append(a)
for i in data:
    driver.get(i)
    soup=s(driver.page_source, 'html.parser')
    for j in soup.find_all('a',class_='alpha-nav__link'):
        a=""
        a=i+"/"+j.text.strip()
        b.append(a)
for i in b:
    driver.get(i)
    soup=s(driver.page_source, 'html.parser')
    for j in soup.find_all('a',class_='subcats__links__item')]:
        data1.append(j.text.strip())
```

```
In [5]: #extracted profiles
data1
'Quality Control (QC) Specialist',
'Quality Control (QC) Technician',
'Quality Engineer',
'Quality Engineer, Electronics',
'Quality Engineer, Medical Devices',
'Quality Engineer, Metal Machining',
'Radio Frequency (RF) Engineer',
'Radio Frequency (RF) Optimization Engineer',
'Reliability Engineer',
'Research Engineer',
'Research and Development Engineer',
'Reservoir Engineer',
'Robotics Engineer',
'Safety Coordinator',
'Safety Engineer',
'Safety Officer',
'Senior ASIC Design Engineer',
'Senior ASIC Engineer',
'Senior Application Engineer, Electrical / Power Control',
'Senior Automation Engineer',
```

```
In [6]: #number of profiles extracted
len(data1)
```

Out[6]: 4894

```
In [7]: #saving data as csv file
d={'profile':data1}
df=pd.DataFrame(d)
```

```
In [10]: df.index+=1
df
```

Out[10]:

| | profile |
|------|----------------------------|
| 1 | Accountant |
| 2 | Accounting Analyst |
| 3 | Accounting Consultant |
| 4 | Accounting Director |
| 5 | Accounting Manager |
| ... | ... |
| 4890 | Warehouse Associate |
| 4891 | Warehouse Coordinator |
| 4892 | Warehouse Manager |
| 4893 | Warehouse Material Handler |
| 4894 | Warehouse Worker |

4894 rows × 1 columns

```
In [11]: df.to_csv('file.csv')
```

```
In [ ]:
```