The aim of this project to extract data (advertised jobs) from the website (staff.am).

```python
In [3]:  import sys
         sys.path.append('/usr/local/lib/python3.9/site-packages')
```

```python
In [4]:  import requests
```

```python
In [58]: import pandas as pd
```

```python
In [5]:  from bs4 import BeautifulSoup
```

The following function takes the url of website as an argument, accesses the website and gets the content. Afterwards, due to the BeautifulSoup function, the content then contains only HTML/valid website tags, from which the data about the job titles, company names, deadlines and locations are extracted.

The function returns a list containing dictionaries of the details mentioned above.

```python
In [6]:  def func_(url):
             response = requests.get(url)
             content = response.content
             soup = BeautifulSoup(content,"html.parser")
             names = soup.find_all("p",attrs={"class": "font_bold"})
             list_df=[]
             for i in range(len(names)-1):
                 dict_df={}
                 dict_df['Name'] = soup.find_all("p",attrs={"class": "font_bold"})[i].text.strip()
                 dict_df['Company Name'] = soup.find_all("p",attrs={"class": "job_list_company_title"})[i].text.s
                 dict_df['Deadline'] = soup.find_all("span",attrs={"class": "formatted_date"})[i].text.strip()
                 dict_df['Location'] = soup.find_all("p",attrs={"class": "job_location"})[i].text.strip()
                 list_df.append(dict_df)
             return list_df
```

For example, the jobs on the first page will be presented in the list1.

```python
In [15]: page1 = 'https://staff.am/en/jobs?JobsFilter%5Bkey_word%5D=&JobsFilter%5Bjob_candidate_level%5D=&JobsFil
```

```python
In [17]: list1 = func_(page1)
```

```python
In [18]: list1
```

```
Out[18]: [{'Name': 'System Administrator',
           'Company Name': 'Praemium RA LLC',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'},
          {'Name': 'Java Web Developer (Full Stack)',
           'Company Name': 'Praemium RA LLC',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'},
          {'Name': 'Full Stack .Net Developer',
           'Company Name': 'Praemium RA LLC',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'},
          {'Name': 'Mid C# / .NET developer',
           'Company Name': 'VECTO',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'},
          {'Name': 'SEO Specialist at Vbet',
           'Company Name': 'SoftConstruct',
           'Deadline': '31 July 2021',
           'Location': 'Yerevan'}
```

The following code gets the URLs of all the available pages on the website. Afterwards, for all the possible URLs, the func_ function is being executed and all the returned lists are joined to the main list.

```
In [21]: i = 1
         main_list = []
         while True:
             page_url = 'https://staff.am/en/jobs?JobsFilter%5Bkey_word%5D=&JobsFilter%5Bjob_candidate_level%5D=&
             list_dfs = func_(page_url)
             if list_dfs == []:
                 break
             main_list.extend(list_dfs)
             i +=1
```

```
In [22]: main_list
```

```
Out[22]: [{'Name': ' Օսկու գողում,հղկում,փայլեցում և 3D մոդելավորողներ',
           'Company Name': 'ADM Diamonds LLC',
           'Deadline': '19 August 2021',
           'Location': 'Abovyan'},
          {'Name': 'Revenue Assurance Unit Supervisor',
           'Company Name': 'Viva–MTS (MTS Armenia CJSC)',
           'Deadline': '27 July 2021',
           'Location': 'Yerevan'},
          {'Name': 'System Administrator',
           'Company Name': 'Praemium RA LLC',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'},
          {'Name': 'Java Web Developer (Full Stack)',
           'Company Name': 'Praemium RA LLC',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'},
          {'Name': 'Full Stack .Net Developer',
           'Company Name': 'Praemium RA LLC',
           'Deadline': '19 August 2021',
           'Location': 'Yerevan'}
```

```
In [63]: print('The website currently contains ' + str(len(main_list)) + ' job advertisments.')

         The website currently contains 1266 job advertisments.
```

```
In [53]: import pandas as pd
         df = pandas.DataFrame(main_list)
```

```
In [54]: df.head()
```

Out[54]:

	Name	Company Name	Deadline	Location
0	Օսկու գողում,հղկում,փայլեցում և 3D մոդելավորողներ	ADM Diamonds LLC	19 August 2021	Abovyan
1	Revenue Assurance Unit Supervisor	Viva-MTS (MTS Armenia CJSC)	27 July 2021	Yerevan
2	System Administrator	Praemium RA LLC	19 August 2021	Yerevan
3	Java Web Developer (Full Stack)	Praemium RA LLC	19 August 2021	Yerevan
4	Full Stack .Net Developer	Praemium RA LLC	19 August 2021	Yerevan

```
In [55]: df['Deadline'] = pd.to_datetime(df['Deadline'], format='%d %B %Y')
```

In [57]: `df.sort_values('Deadline')`

Out[57]:

	Name	Company Name	Deadline	Location
1265	Senior Backend Engineer, Enterprise (Node.js/J...	Picsart	2021-07-20	Yerevan
740	.NET Software Engineer - Mid 3 Level	HelpSystems Armenia	2021-07-20	Yerevan
739	.NET Software Engineer - Mid 1 Level	HelpSystems Armenia	2021-07-20	Yerevan
738	Senior .NET Engineer	HelpSystems Armenia	2021-07-20	Yerevan
721	Գլխավոր Հաշվապահ (ՎՌ Արմենիան Ֆրութ ՍՊԸ)	SoftConstruct	2021-07-20	Yerevan
...	...	...	...	...
40	Գնումների /տենդեր/ մասնագետ	Ռաֆ-Օջախ ՍՊԸ	2021-08-19	Yerevan
42	Wordpress developer	Inexxus	2021-08-19	Yerevan
43	Տեղագրական Գեոդեզիստ	AAB Construction	2021-08-19	Alaverdi
16	Ֆիրմայի ներկայացուցիչ	Bacon Product LLC	2021-08-19	Yerevan
0	Ոսկու զոդում,հղկում,փայլեցում և 3D մոդելավորողներ	ADM Diamonds LLC	2021-08-19	Abovyan

1266 rows × 4 columns

The top locations.

In [38]: `a = df.groupby('Location').count()`

In [66]: `a.sort_values('Name', ascending=False)`

Out[66]:

Location	Name	Company Name	Deadline
Yerevan	1194	1194	1194
Gyumri	12	12	12
Vanadzor	10	10	10
Dilijan	5	5	5
Remote	5	5	5
Abovyan	5	5	5
Armenia (All cities)	3	3	3
Hrazdan	3	3	3
Ashtarak	2	2	2
Vedi	2	2	2
London	2	2	2
Los Angeles	2	2	2
Vagharshapat	2	2	2
Toronto	2	2	2
Armavir	2	2	2
Stepanakert	2	2	2
Tashir	1	1	1
Yeghvard	1	1	1
Aparan	1	1	1
Tsaghkadzor	1	1	1
Nagorno-Karabakh	1	1	1
Stepanavan	1	1	1
Berd	1	1	1
Masis	1	1	1
Martuni (Artsakh)	1	1	1
Alaverdi	1	1	1
Ijevan	1	1	1
Artashat	1	1	1
Martuni	1	1	1

The top 20 companies.

In [44]: `companies = df.groupby('Company Name').count()`

```
In [50]: companies.sort_values('Name', ascending=False).head(20)
```

Out[50]:

Company Name	Name	Deadline	Location
SoftConstruct	53	53	53
ArctX LLC	30	30	30
Atenk Ltd	27	27	27
EPAM Systems	27	27	27
KINDDA	21	21	21
Synopsys Armenia	20	20	20
Digitain	19	19	19
TUMO Center for Creative Technologies	19	19	19
Webb Fontaine Holding LLC	19	19	19
Instigate Semiconductor	18	18	18
IMEX GROUP	17	17	17
Ardshinbank	16	16	16
Adjarabet.am	15	15	15
Krisp	15	15	15
ID Bank CJSC	15	15	15
ServiceTitan	13	13	13
Picsart	12	12	12
SADA Global Delivery Center	12	12	12
VOLO	11	11	11
Ucom	11	11	11

```
In [51]: df.to_csv("jobs.csv",index=False)
```

The data included in the project is presented on 20.07.21 at 16.00 pm.