

Step-by-Step Installation Instructions

Following are step-by-step instructions for run scraping script on windows 10.

1. Download and install latest Python on windows
Follow this Python installation instruction from video: <https://www.youtube.com/watch?v=dX2-V2BocqQ>
2. Install Python PIP on Windows 10
Follow Python PIP installation instruction from this video:
<https://www.youtube.com/watch?v=AVCcFyYynQY>
3. Setting Environment Variables and Running the Script from Command Prompt
Follow installation instruction from this video: <https://www.youtube.com/watch?v=wIRzg6fD2Gs>
4. Install requests python library

Open command prompt and run command “pip install requests”

```
--cert <path>          Path to alternate CA bundle.
--client-cert <path>   Path to SSL client certificate, a single file containing the private key and the
                        certificate in PEM format.
--cache-dir <dir>      Store the cache data in <dir>.
--no-cache-dir          Disable the cache.
--disable-pip-version-check
                        Don't periodically check PyPI to determine whether a new version of pip is available for
                        download. Implied with --no-index.
--no-color              Suppress colored output
--no-python-version-warning
                        Silence deprecation warnings for upcoming unsupported Python's.

C:\Users\PSI-PC>pip install requests
Collecting requests
  Downloading requests-2.23.0-py2.py3-none-any.whl (58 kB)
    | 58 kB 466 kB/s
Collecting chardet<4,>=3.0.2
  Downloading chardet-3.0.4-py2.py3-none-any.whl (133 kB)
    | 133 kB 819 kB/s
Collecting idna<3,>=2.5
  Downloading idna-2.9-py2.py3-none-any.whl (58 kB)
    | 58 kB 880 kB/s
Collecting certifi>=2017.4.17
  Downloading certifi-2020.4.5.1-py2.py3-none-any.whl (157 kB)
    | 157 kB 939 kB/s
Collecting urllib3[=1.25.0,!=1.25.1,<1.26,>=1.21.1]
  Downloading urllib3-1.25.9-py2.py3-none-any.whl (126 kB)
    | 126 kB 930 kB/s
Installing collected packages: chardet, idna, certifi, urllib3, requests
Successfully installed certifi-2020.4.5.1 chardet-3.0.4 idna-2.9 requests-2.23.0 urllib3-1.25.9

C:\Users\PSI-PC>
```

5. Install beautifulsoup4 python library

Open command prompt and run command “pip install beautifulsoup4”

```
C:\Users\PSI-PC>pip install beautifulsoup4
Collecting beautifulsoup4
  Downloading beautifulsoup4-4.9.0-py3-none-any.whl (109 kB)
    | 109 kB 1.3 MB/s
Collecting soupsieve>1.2
  Downloading soupsieve-2.0-py2.py3-none-any.whl (32 kB)
Installing collected packages: soupsieve, beautifulsoup4
Successfully installed beautifulsoup4-4.9.0 soupsieve-2.0

C:\Users\PSI-PC>
```

6. Install pandas python library

Open command prompt and run command “pip install pandas”

```
C:\Users\PSI-PC>pip install pandas
Collecting pandas
  Downloading pandas-1.0.3-cp38-cp38-win_amd64.whl (8.9 MB)
    | 8.9 MB 3.2 MB/s
Collecting numpy>=1.13.3
  Downloading numpy-1.18.4-cp38-cp38-win_amd64.whl (12.8 MB)
    | 12.8 MB 3.3 MB/s
Collecting pytz>=2017.2
  Downloading pytz-2020.1-py2.py3-none-any.whl (510 kB)
    | 510 kB 2.2 MB/s
Collecting python-dateutil>=2.6.1
  Downloading python_dateutil-2.8.1-py2.py3-none-any.whl (227 kB)
    | 227 kB 2.2 MB/s
Collecting six>=1.5
  Downloading six-1.14.0-py2.py3-none-any.whl (10 kB)
Installing collected packages: numpy, pytz, six, python-dateutil, pandas
Successfully installed numpy-1.18.4 pandas-1.0.3 python-dateutil-2.8.1 pytz-2020.1 six-1.14.0
C:\Users\PSI-PC>
```

7. Install openpyxl python library

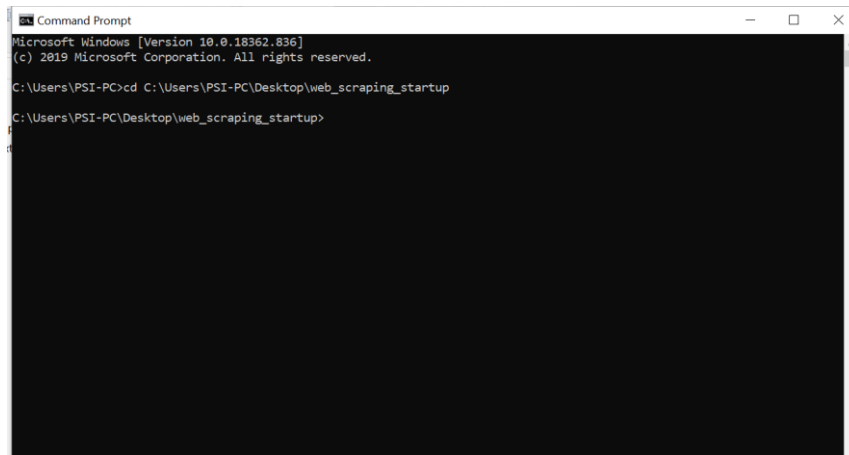
Open command prompt and run command “pip install openpyxl”

```
Command Prompt
Collecting six>=1.5
  Downloading six-1.14.0-py2.py3-none-any.whl (10 kB)
Installing collected packages: numpy, pytz, six, python-dateutil, pandas
Successfully installed numpy-1.18.4 pandas-1.0.3 python-dateutil-2.8.1 pytz-2020.1 six-1.14.0

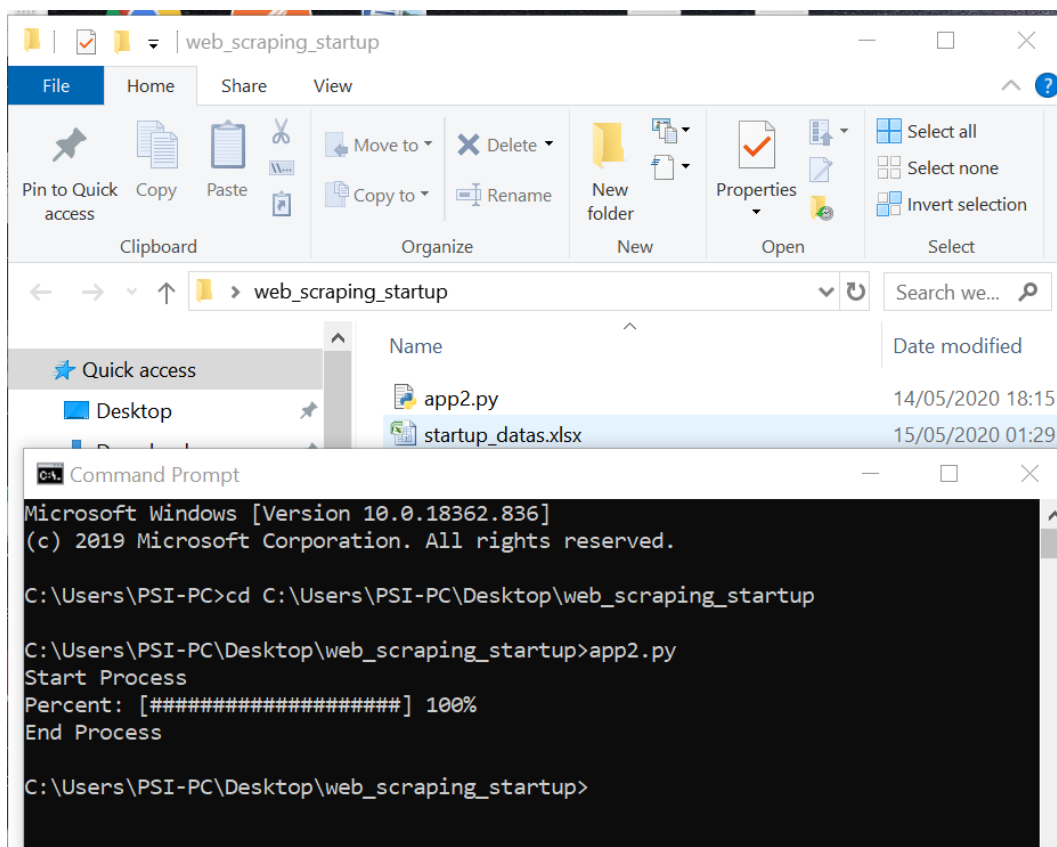
C:\Users\PSI-PC>pip install openpyxl
Collecting openpyxl
  Downloading openpyxl-3.0.3.tar.gz (172 kB)
    | 172 kB 1.3 MB/s
Collecting jdcal
  Downloading jdcal-1.4.1-py2.py3-none-any.whl (9.5 kB)
Collecting et_xmlfile
  Downloading et_xmlfile-1.0.1.tar.gz (8.4 kB)
Building wheels for collected packages: openpyxl, et-xmlfile
Building wheel for openpyxl (setup.py) ... done
Created wheel for openpyxl: filename=openpyxl-3.0.3-py2.py3-none-any.whl size=241264 sha256=a0f8f2169b8303242a7c57e12884139f50fafdd1baefc7099ba6c3666990c9d
Stored in directory: c:\users\psi-pc\appdata\local\pip\cache\wheels\34\21\39\70d175b8eb9f953ff614f6881346cf556226951d79e4163a32
Building wheel for et-xmlfile (setup.py) ... done
Created wheel for et-xmlfile: filename=et_xmlfile-1.0.1-py3-none-any.whl size=8920 sha256=05d7a3dd476d0613d383959ab4b8197b2af2c929fa07cebc4f2d4f0a05a8effc
Stored in directory: c:\users\psi-pc\appdata\local\pip\cache\wheels\6e\df\38\abda47b884e3e25f9b6430e5ce44c47670758a50c0c51759
Successfully built openpyxl et-xmlfile
Installing collected packages: jdcal, et-xmlfile, openpyxl
Successfully installed et-xmlfile-1.0.1 jdcal-1.4.1 openpyxl-3.0.3
C:\Users\PSI-PC>
```

8. Run script python for scraping

Open command prompt in folder “web_scraping_startup”



Run python script app2.py on command prompt



Wait until the process becomes 100% and the data will appear in Excel format

Name								
	Name	Description	Product	Creators	Domain	Email	Phone	Category
1	Ibat	La start-up Solutions	Mélanie Li	https://m.contact@i01 77 45 5	Logiciels	https://www.usine-digitale.fr/annuaire-start-up/ibat,942496		
2	Madbox	Madbox e Jeux sur n	Jean-Nicol	https://m.contact@r01 42 61 4	Jeux vidéo	https://www.usine-digitale.fr/annuaire-start-up/madbox,942491		
3	GreenBig	La start-up b:bot : ma	Benoît Pa	https://m.info@gre No	Phone Ecoconce	https://www.usine-digitale.fr/annuaire-start-up/greenbig,963886		
4	Loutsas	Loutsas est Café en gr	Guillaume	https://m.contact@i04 28 29 7	Commerce	https://www.usine-digitale.fr/annuaire-start-up/loutsas,942481		
5	noCRM.io	La start-up Logiciel de	Sunny Par	https://m.contact@y No	Phone Logiciels	https://www.usine-digitale.fr/annuaire-start-up/nocrm-io,963431		
6	ChargeGu	La start-up Solutions	Gilles Gon	http://m.contact@r01 88 24 6	Mobilité	https://www.usine-digitale.fr/annuaire-start-up/chargeguru,963381		
7	Quicktext	La start-up Solution d	Daniel Do	https://m.hello@qui01 85 54 0	Voyage, h	https://www.usine-digitale.fr/annuaire-start-up/quicktext,963366		
8	Calixar	Calixar est Technolog	Emmanue	https://m.ca.contact@r04 81 07 6	Biotechno	https://www.usine-digitale.fr/annuaire-start-up/calixar,942476		
9	Linguifami	La start-up Plateform	Agnès Bar	https://m.contact@i06 87 14 1	Education	https://www.usine-digitale.fr/annuaire-start-up/linguifamily,963121		
10	Tacylate	Créée en : Mower Cl	Larbi Nah	https://m.contact@i09 84 21 5	Commerce	https://www.usine-digitale.fr/annuaire-start-up/tacylate,963051		
11	Label Exp	Créée en : Agence de	Sophie Da	https://m.lal.audrey@l No	Phone Webmark	https://www.usine-digitale.fr/annuaire-start-up/label-experience,963026		
12	Pixpay	Pixpay prc Carte de p	Benoît Gr	https://m.hello@pix09 54 77 0	Financem	https://www.usine-digitale.fr/annuaire-start-up/pixpay,942421		
13	ActivCorn	ActivCorn Plateform	Yacine Chi	https://m.hello@act09 54 77 0	Services a	https://www.usine-digitale.fr/annuaire-start-up/activcorner,941456		
14	Telegrafik	La start-up Solutions	Carole Zis	https://m.contact@i05 82 95 5	E-santé	https://www.usine-digitale.fr/annuaire-start-up/telegrafik,941441		
15	Ausha by i	Créée en : Plateform	Charles de	https://m.fr.hello@aur No	Phone Logiciels	https://www.usine-digitale.fr/annuaire-start-up/ausha-by-icreo,941436		
16	GamersOrigin	Implantée Agence sp	Guillaume	http://m.ganttechnique07 78 04 0	Jeux vidéo	https://www.usine-digitale.fr/annuaire-start-up/gamersorigin,962251		
17	Fabulabor	Créée en : Jouets d	Alexandra	https://m.contact@i04 74 72 4	Commerce	https://www.usine-digitale.fr/annuaire-start-up/fabulabor,962231		
18	Nyctale	La start-up Outils de	Blaise Cav	https://m.ny.contact@l No	Phone Logiciels	https://www.usine-digitale.fr/annuaire-start-up/nyctale,962226		
19	Lizee	La start-up Solution p	Tanguy Fri	http://m.contact@l No	Phone Start-up E	https://www.usine-digitale.fr/annuaire-start-up/lizee,962221		
20	Isybot	La start-up Robots de	Yvan Mea	https://m.contact@l No	Phone Electroniq	https://www.usine-digitale.fr/annuaire-start-up/isybot,961726		
21	Sarus Tec	Implantée Solution d	Maxime A	https://m.sa.contact@: No	Phone Logiciels	https://www.usine-digitale.fr/annuaire-start-up/sarus-technologies,961716		
22								