

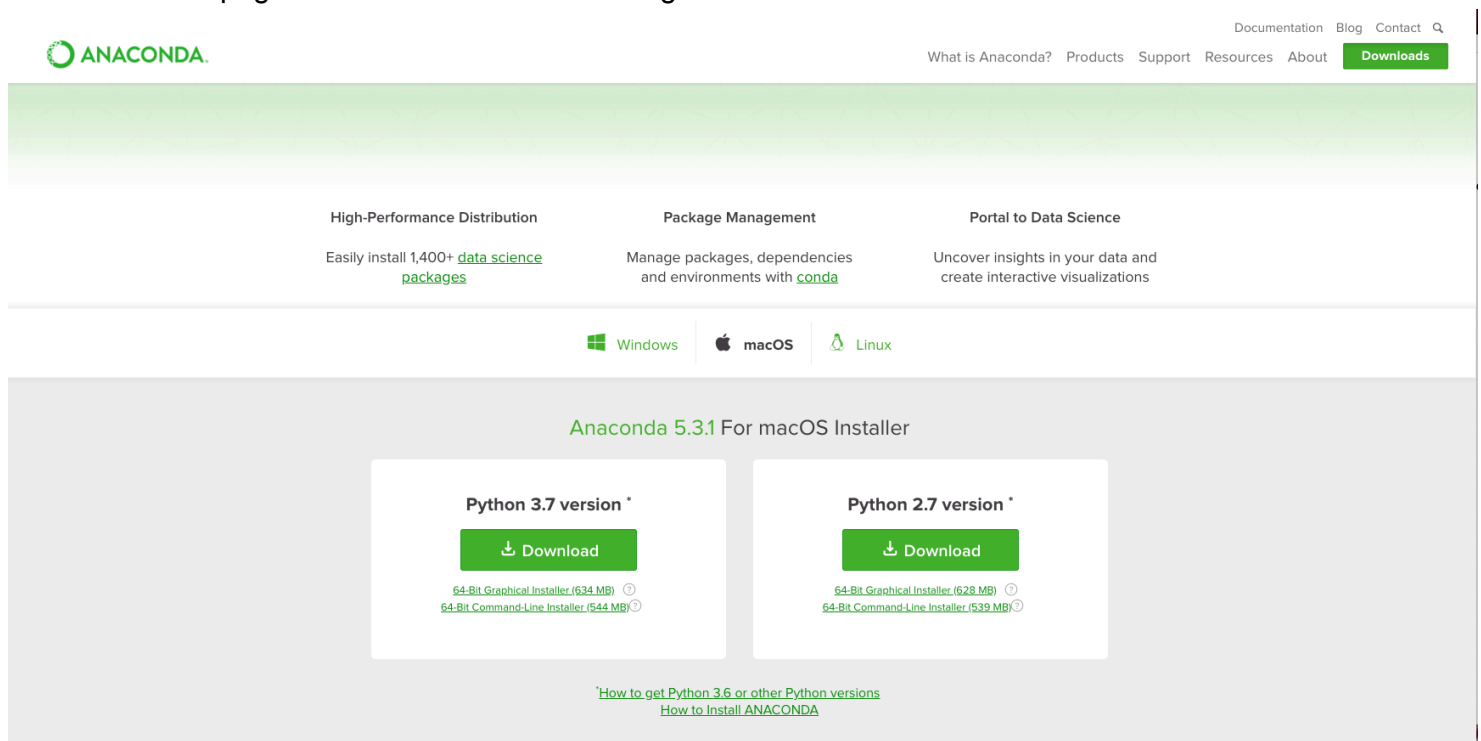
Instruction Manual

Step 1. Install Anaconda

Anaconda is a software that allows you to run the Jupyter Notebook on your computer. Please refer to this link for download:

<https://www.anaconda.com/download/>

Scroll down the page a little bit to find the following section:

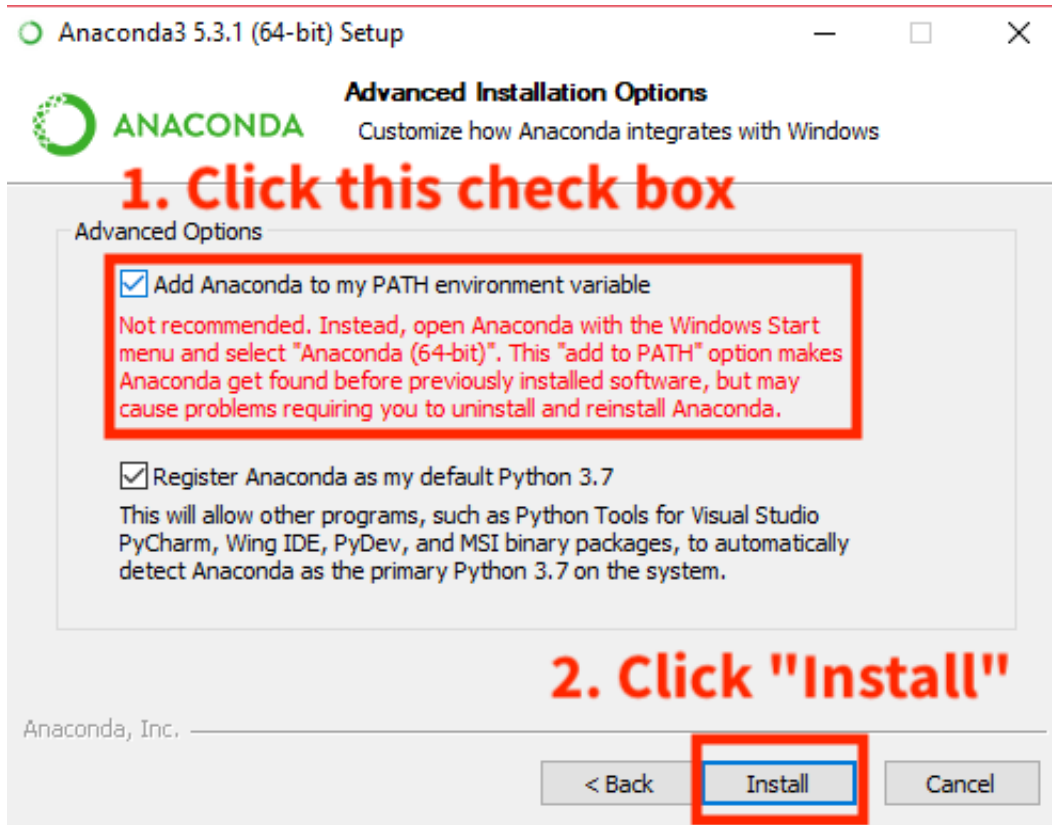


Please download the **Python 3.7 version** of Anaconda

Then please install Anaconda using the downloaded executable.

Windows User

Please execute the installer and keep click on '**Next**' without changing any settings until you see this page:



Make sure you check the option "**Add Anaconda to my PATH environment variable**" before you click "**Install**".

If the installer is asking you to install "MS Visual Studio Code", please ignore it.

Step 2. Install Required Python Libraries

We are now installing the required Python libraries to allow the program to execute correctly.

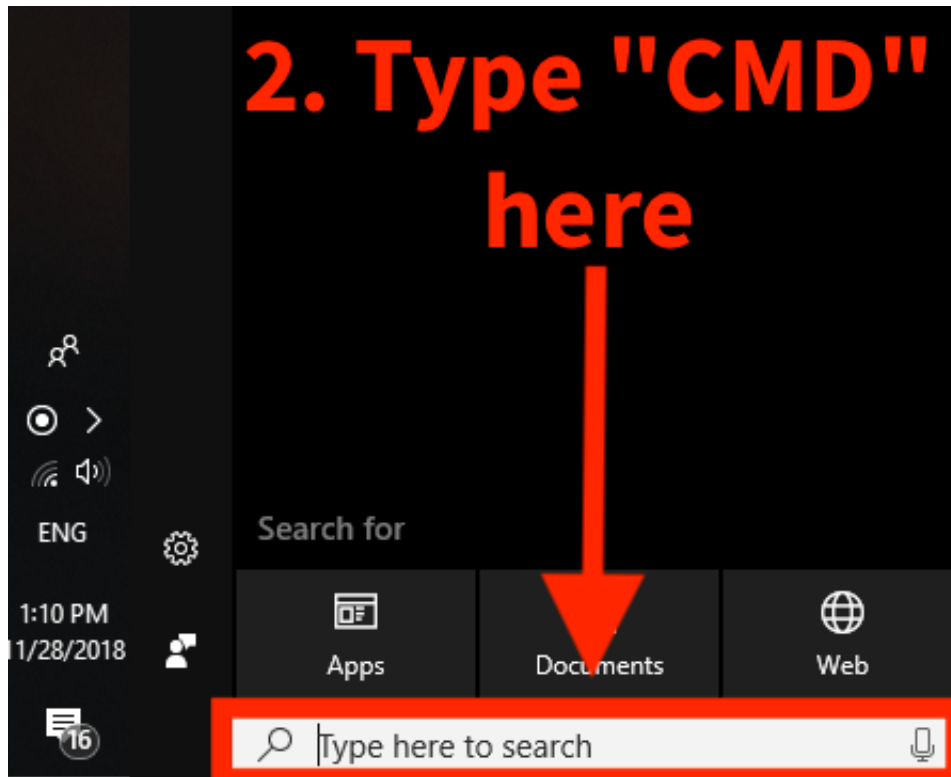
Windows User

1. Please open **CMD** program with "Administrator Privilege"

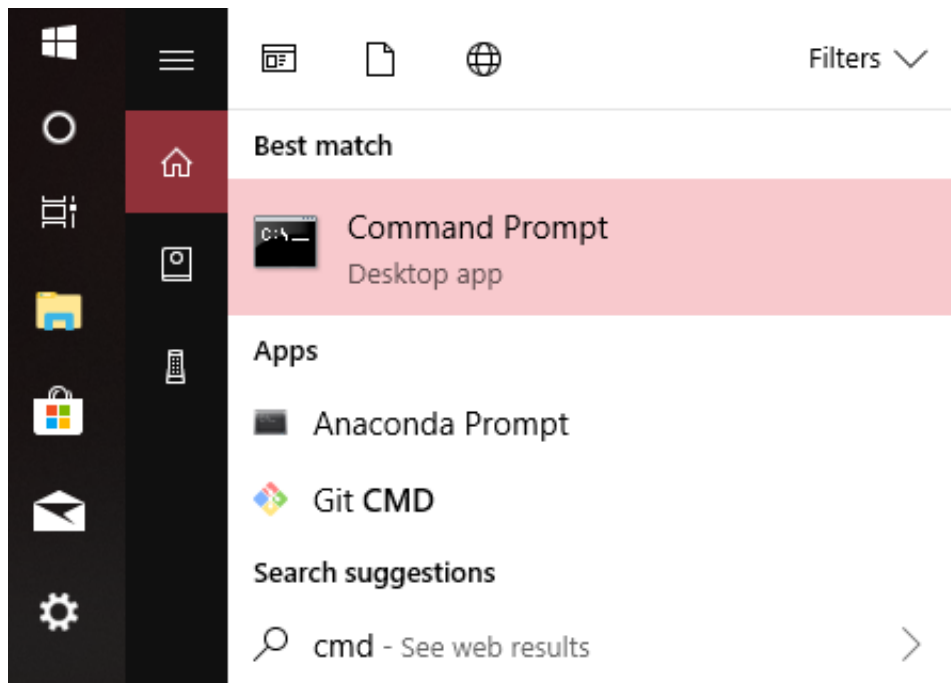
On the Windows taskbar, find a circle icon and click it



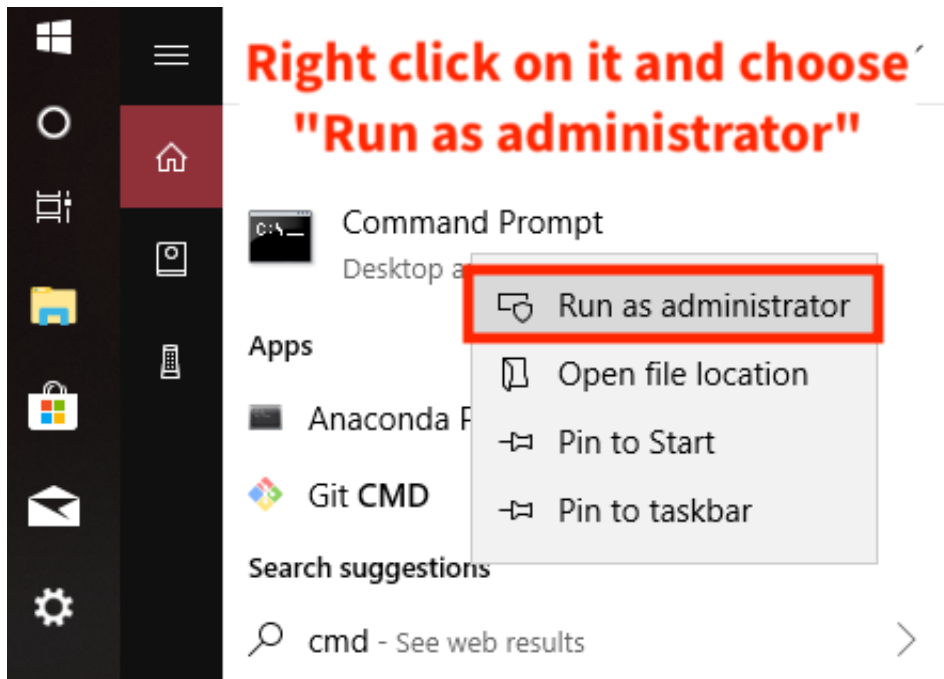
You will see a input field at the button, type "**cmd**"



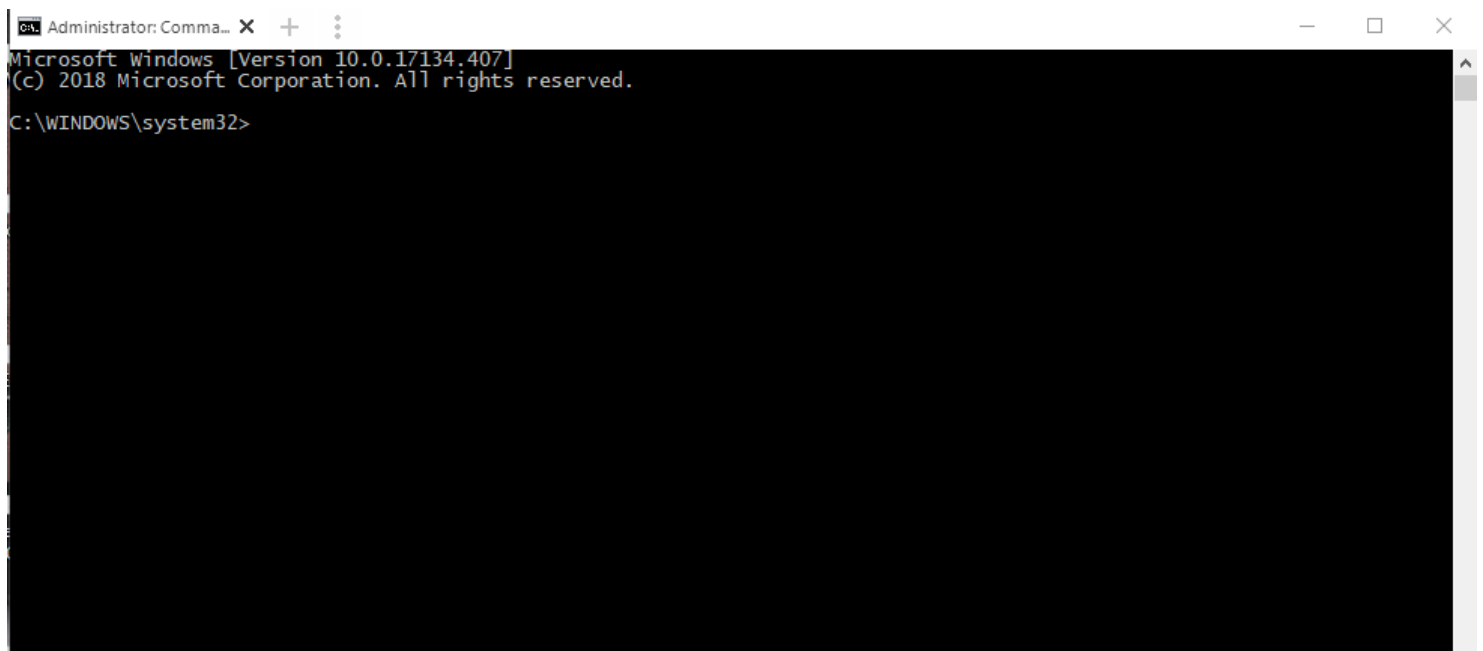
The cmd program will show up in the window.



Please right click on it and choose "Run as administrator"



Here is the CMD window:



2. Insert the command to install the libraries:

This is the command for installation:

```
conda install -c conda-forge selenium pandas numpy
```

Copy & paste this line to the CMD window and press "Enter" key.

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.17134.407]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>conda install -c conda-forge selenium pandas numpy

Copy & Paste the command here and press "Enter" key
```

Then, you will see a message like the following, type 'y' and press "Enter" key.

```
The following NEW packages will be INSTALLED:

  libflang: 5.0.0-vc14_20180208 conda-forge [vc14]
  llvm-meta: 5.0.0-0 conda-forge
  openblas: 0.3.3-h535eed3_1001 conda-forge
  openmp: 5.0.0-vc14_1 conda-forge [vc14]

The following packages will be UPDATED:

  blas: 1.0-mkl --> 1.1-openblas conda-forge
  numpy: 1.15.1-py37ha559c80_0 --> 1.15.4-py37_blas_openblash442142e_1000 conda-forge [blas_openblas]
  pandas: 0.23.4-py37h830ac7b_0 --> 0.23.4-py37h830ac7b_1000 conda-forge

Proceed ([y]/n)? y Type 'y' and press "Enter" key
```

Wait until the progress complete.

Step 3. Download additional Lib. file and the Program

Download the program files:

Please refer to this link: <https://github.com/hippoandy/UNWeb scrapingWORKANA/>

Download the program .zip file as:

Search or jump to... Pull requests Issues Marketplace Explore

hippoandy / UN_Web scraping_WORKANA

Watch 0 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

This is the project aims to find information of job market among Latin America and Caribbean countries. Edit

Manage topics

40 commits 1 branch 0 forks

Branch: master New pull request

Create new file Upload files Find file Clone or download

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

https://github.com/hippoandy/UN_Webscrap

Open in Desktop Download ZIP

1. Click "Clone or download"

2. Click "Download ZIP"

Open the downloaded .zip file and extract the content:

UN_Web scraping_WO... x +

File Home Share View Extract

Documents Pictures iCloud Drive Desktop Music

Extract all

Click "Extract all"

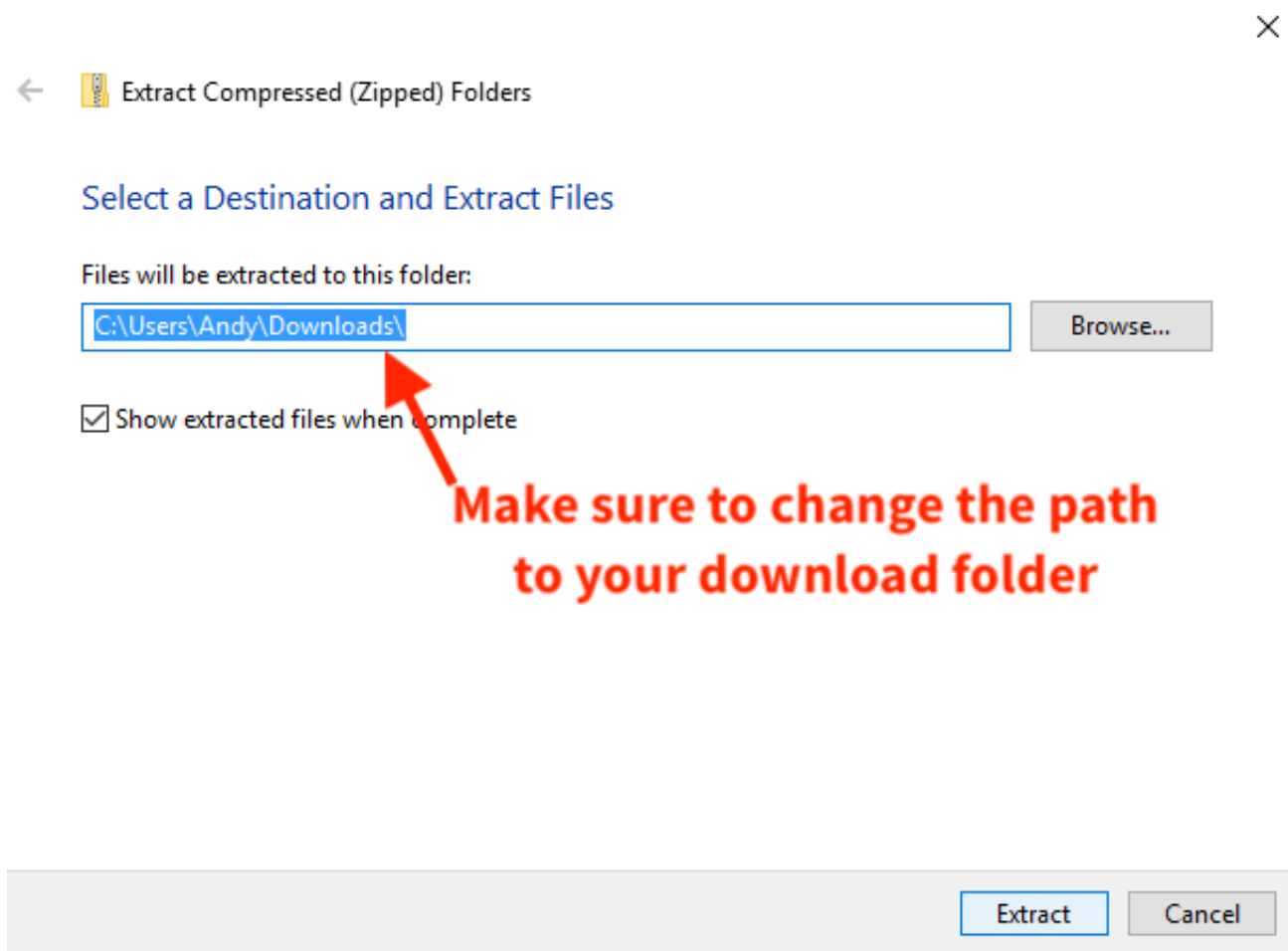
Extract all

Extract all items in this folder.

Name type Compressed size Password ... Size Ratio Date modified

UN_Web scraping_WORKANA-mas... File folder 11/27/2018 8:44 PM

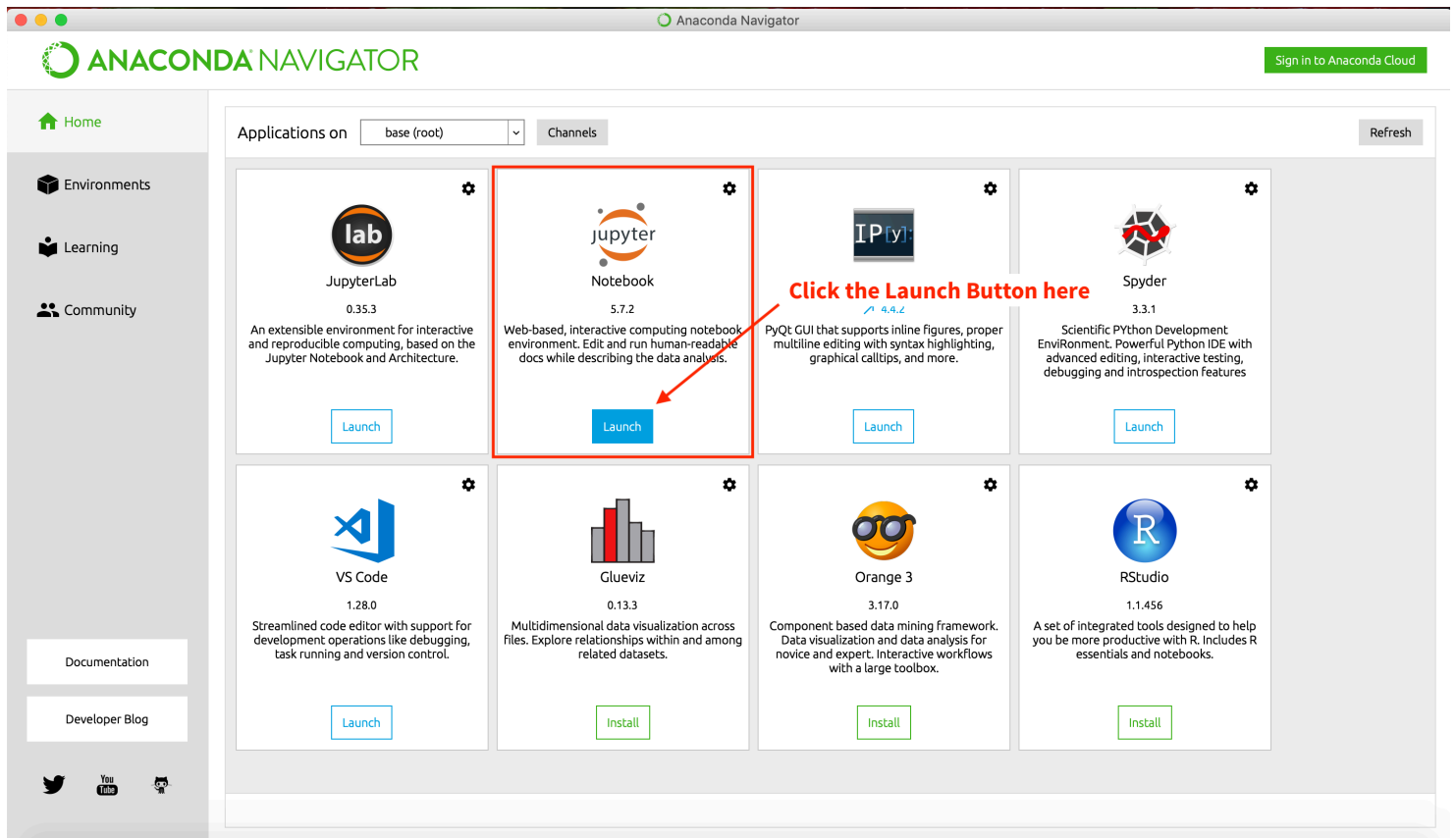
Make sure to change to destination path to your **Download** folder:



Step 4. Open Jupyter Notebook Application

1. Open Anaconda then Jupyter

Open you installed Anaconda Program, you should see a window like this:



To launch the Jupyter Notebook, click the "**Launch**" button of it.

Then, you should see you browser opens up and shows a page like this:

Home x +

localhost:8888/tree

Apps LeetCode iCloud Google Cloud Scholar Translate YouTube Facebook Messenger LinkedIn Paper Overleaf Other bookmarks

jupyter Quit Logout

Files Running Clusters

Select items to perform actions on them. Upload New ↺

| 0 | / | Name ↓ | Last Modified | File size |
|--------------------------|-------------|--------|---------------|-----------|
| <input type="checkbox"/> | 3D Objects | | 13 days ago | |
| <input type="checkbox"/> | Anaconda3 | | an hour ago | |
| <input type="checkbox"/> | Contacts | | 13 days ago | |
| <input type="checkbox"/> | Desktop | | 4 days ago | |
| <input type="checkbox"/> | Documents | | 2 hours ago | |
| <input type="checkbox"/> | Downloads | | seconds ago | |
| <input type="checkbox"/> | Dropbox | | 13 days ago | |
| <input type="checkbox"/> | Favorites | | 11 days ago | |
| <input type="checkbox"/> | iCloudDrive | | 18 hours ago | |
| <input type="checkbox"/> | Links | | 13 days ago | |
| <input type="checkbox"/> | Music | | 13 days ago | |
| <input type="checkbox"/> | OneDrive | | 18 hours ago | |
| <input type="checkbox"/> | Pictures | | 12 days ago | |
| <input type="checkbox"/> | Roaming | | 14 days ago | |
| <input type="checkbox"/> | Saved Games | | 13 days ago | |
| <input type="checkbox"/> | Searches | | 13 days ago | |
| <input type="checkbox"/> | Videos | | 12 days ago | |

Click it to navigate to your Download folder

Navigate to your Download folder, and you should see this:

jupyter Quit Logout

Files Running Clusters

Select items to perform actions on them. Upload New ↺

| 0 | / Downloads | Name ↓ | Last Modified | File size |
|--------------------------|--------------------------------|--------|---------------|-----------|
| <input type="checkbox"/> | .. | | seconds ago | |
| <input type="checkbox"/> | UN_Webscrapping_WORKANA-master | | 9 minutes ago | |
| <input type="checkbox"/> | chromedriver.exe | | an hour ago | 6.7 MB |

Continue to navigate to the program folder

Here you will see the downloaded files instructed in the previous section. Continue to navigate to the program folder:

UN_Webscrapping_WORKANA-master > scrpaing > selenium

Then, you should see this page:

Files Running Clusters

Select items to perform actions on them.

Here is the path, make sure it is the same as yours

Upload

New



| | | | | | | |
|---|--|--|--|------|----------------|-----------|
| 0 | / Downloads / UN_Web scraping_WORKANA-master / scraping / selenium | | | Name | Last Modified | File size |
| | .. | | | | seconds ago | |
| | chrome-driver | | | | 11 minutes ago | |
| | WORKANA_Freelancers_Scraping_using_Selenium.ipynb | | | | 11 minutes ago | 15.4 kB |
| | result.csv | | | | 11 minutes ago | 2 kB |
| | scrap.py | | | | 11 minutes ago | 5.4 kB |
| | tutorial.py | | | | 11 minutes ago | 2.11 kB |

Open this Jupyter Notebook

Click on the Jupyter Notebook to open the program.

Here is the Jupyter Notebook program

Run the notebook through this button

```
In [1]: '''
This program scraps the information from the freelancers section of the WORKANA website.
This program utilizes the Selenium library for web scraping, which enables the program to change conditions or
filters the website provided to show different result

Author: Yu-Chang (Andy) Ho
Date: 2018/11/27
'''

Out[1]: '\nThis program scraps the information from the freelancers section of the WORKANA website.\nThis program utilizes the Selenium
library for web scraping, which enables the program to change conditions or\nfilters the website provided to show different res
ult\n\nAuthor: Yu-Chang (Andy) Ho\nDate: 2018/11/27\n'
```

```
In [2]: # import the required libraries
from selenium import webdriver
# the library to control program execution at run time
import time
import pandas as pd
import numpy as np

print( 'This section import the required libraries.' )

This section import the required libraries.
```

```
In [3]: # parameter settings -----

# the keyword which to perform a search
# as an example, we are scraping the freelancers who is familiar with SQL Language
KEY_search = 'SQL'

# the target URL
URL_target = 'https://www.workana.com/freelancers'

# time to wait for the page to refresh
# to make sure the changes applied to the website and allow the page to reload
TIME_pending = 2

# path of the output file
PATH_output = './result.csv'

# the search results will be seperated into multiple subpages
# this is a large number to cover all the page number
limit = 3

# ----- parameter settings

print( 'Here are the parameters that able to be modified.' )

Here are the parameters that able to be modified.
```

```
In [4]: # self-defined functions -----

# return the string 'N/A'
def invalid_val(): return 'N/A'

# make sure there is no special char in a value
def clear_str( text ):
    text = str(text).replace( '\n', '' ).replace( '\r', '' ).replace( '\t', '' )
    return text
```

Please use the run button to execute the notebook.

2. Run the Notebook

If you see this window while running the program, please click "**Allow**".



Windows Defender Firewall has blocked some features of this app

Windows Defender Firewall has blocked some features of Python on all public and private networks.



Name: Python
Publisher: Python Software Foundation
Path: C:\users\andy\anaconda3\python.exe

Allow Python to communicate on these networks:

- ☐ Private networks, such as my home or work network
- ☒ Public networks, such as those in airports and coffee shops (not recommended because these networks often have little or no security)

[What are the risks of allowing an app through a firewall?](#)



Allow access

Cancel