

Please follow this rules to scrap the data:

1) Docker and scrapy_splash should be installed

2) As per the settings, please run docker like that:

\$ docker run -it -p 8050:8050 scrapinghub/splash --max-timeout 300

(More details: <https://splash.readthedocs.io/en/stable/api.html>)

3) Be sure that scrape is working:

\$ scrapy

4) You should go to the main folder scrapy/WS_project:

\$ cd ...

Please follow this order to run the scripts (once you are in the correct folder):

1) It should be run the spider: link_country:

\$ scrapy crawl link_country -o Data/Links/link_country.csv

2) It should be run the spider: pages:

\$ scrapy crawl pages -o Data/Links/pages.csv

3) It should be run the spider: pollution:

\$ scrapy crawl pollution -o Data/pollution.csv

4) It should be run the python script: Analysis.py:

\$ python3 Analysis.py