732A76_report-chesh532

Chenjian Shi

2021/12/30

Abstract

.

Contents

1 Introduction **1**

2 Background **1**

3 Data Access **1**
3.1 Configs 2
3.2 Data downloads 2

4 Data Proccessing **2**
4.1 Config 4
4.2 Functions in the script 4

5 Result Anlysis **4**

6 Conclusion **5**

7 Appendix **5**

1 Introduction

2 Background

3 Data Access

The data is downloaded Folkhalsomyndigheten's daily excel spread sheet report concerning COVID19 in Sweden from Wayback machine. In order to access the data, here we used a python script[.python_script] which implemented by a package called selenium(Conservancy n.d.). The selenium we used in this project is the Selenium WebDriver. This is a kind of browser-based automation test suites, which allows python scripts run the automation tests like mouse clicking,etc. by locating the xpath elements of the website,etc..

In this paper, we would use the chrome driver(version 96.0.4664.45, based on Chrome 96.0.4664.110) which download from Chrome Driver's official website(Chromium and teams n.d.), which is the webdriver for the Chrome.

The total script contains the files below:

3.1 Configs

Before running the main script, we first need to do some settings in the config.py file.

```
# Config.py

# Location of chrome driver
chrome_driver = r"C:\Program Files (x86)\Google\Chrome\Application\chromedriver.exe"

# Url for Wayback Machine
Wayback_Machine_URL = 'https://web.archive.org/'

# Target URL
Target_URL = ('https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-
              'bekraftade-fall-i-sverige',
              'https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-
             'statistik-och-analyser/bekraftade-fall-i-sverige/')

# Define the start date and final date
start_date = 'APR 04 2020'
final_date = 'DEC 28 2021'

# Download directory
Download_directory = r'G:\LiU\732A76_Research_Project\data'
```

In this file, we need to set the location of chrome driver first, which helps python scripts to run the webdriver. Besides, we sets the url for the Wayback Machine and the url for the Folkhalsomyndigheten in order to the get the files from the past. Then, we sets the start and end dates of the files we need, which in form of 'MON DAY YEAR.' At last, we set the download directory for the files.

3.2 Data downloads

After finishing the setting of configs.py, we just run the 'main.py' file so that the download would start. The process of the downloads(in PyCharm IDE) would be like this

```
Jan 02 2022 - 23:25:14, Start to downloading files
Jan 02 2022 - 23:26:18, Finished downloading of the file in Sep 17 2021
Jan 02 2022 - 23:27:20, Finished downloading of the file in Sep 18 2021
```

If there is no data at that date or failed the downloads, the script would shows like this

```
Jan 02 2022 - 23:25:14, Start to downloading files
Jan 02 2022 - 23:26:18, failed to download the file in Sep 17 2021
Jan 02 2022 - 23:27:20, No file could be download in Sep 18 2021
```

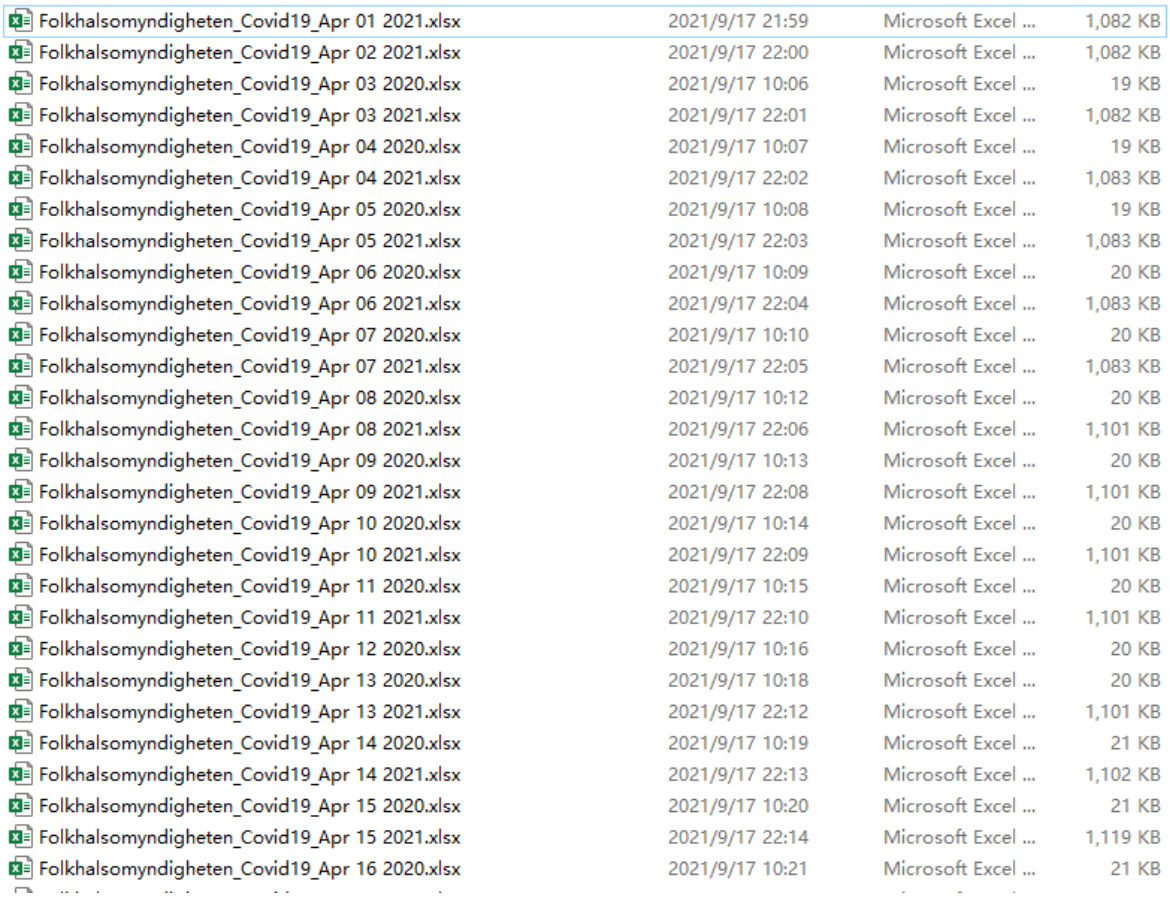
Each downloads may takes about 1 minitues, which depends on the network situation. The total time used in downloading the files from APR 03 2020 to DEC 28 2021 taks about 10 hours to download.

4 Data Proccessing

After we gets the data files like above, we start to do the process these data by R script[.script_for_processing].

The required packages for the script are below:

```
library(readxl)
library(tidyverse)
library(plotly)
```



Folkhalsomyndigheten_Covid19_Apr 01 2021.xlsx	2021/9/17 21:59	Microsoft Excel ...	1,082 KB
Folkhalsomyndigheten_Covid19_Apr 02 2021.xlsx	2021/9/17 22:00	Microsoft Excel ...	1,082 KB
Folkhalsomyndigheten_Covid19_Apr 03 2020.xlsx	2021/9/17 10:06	Microsoft Excel ...	19 KB
Folkhalsomyndigheten_Covid19_Apr 03 2021.xlsx	2021/9/17 22:01	Microsoft Excel ...	1,082 KB
Folkhalsomyndigheten_Covid19_Apr 04 2020.xlsx	2021/9/17 10:07	Microsoft Excel ...	19 KB
Folkhalsomyndigheten_Covid19_Apr 04 2021.xlsx	2021/9/17 22:02	Microsoft Excel ...	1,083 KB
Folkhalsomyndigheten_Covid19_Apr 05 2020.xlsx	2021/9/17 10:08	Microsoft Excel ...	19 KB
Folkhalsomyndigheten_Covid19_Apr 05 2021.xlsx	2021/9/17 22:03	Microsoft Excel ...	1,083 KB
Folkhalsomyndigheten_Covid19_Apr 06 2020.xlsx	2021/9/17 10:09	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 06 2021.xlsx	2021/9/17 22:04	Microsoft Excel ...	1,083 KB
Folkhalsomyndigheten_Covid19_Apr 07 2020.xlsx	2021/9/17 10:10	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 07 2021.xlsx	2021/9/17 22:05	Microsoft Excel ...	1,083 KB
Folkhalsomyndigheten_Covid19_Apr 08 2020.xlsx	2021/9/17 10:12	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 08 2021.xlsx	2021/9/17 22:06	Microsoft Excel ...	1,101 KB
Folkhalsomyndigheten_Covid19_Apr 09 2020.xlsx	2021/9/17 10:13	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 09 2021.xlsx	2021/9/17 22:08	Microsoft Excel ...	1,101 KB
Folkhalsomyndigheten_Covid19_Apr 10 2020.xlsx	2021/9/17 10:14	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 10 2021.xlsx	2021/9/17 22:09	Microsoft Excel ...	1,101 KB
Folkhalsomyndigheten_Covid19_Apr 11 2020.xlsx	2021/9/17 10:15	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 11 2021.xlsx	2021/9/17 22:10	Microsoft Excel ...	1,101 KB
Folkhalsomyndigheten_Covid19_Apr 12 2020.xlsx	2021/9/17 10:16	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 13 2020.xlsx	2021/9/17 10:18	Microsoft Excel ...	20 KB
Folkhalsomyndigheten_Covid19_Apr 13 2021.xlsx	2021/9/17 22:12	Microsoft Excel ...	1,101 KB
Folkhalsomyndigheten_Covid19_Apr 14 2020.xlsx	2021/9/17 10:19	Microsoft Excel ...	21 KB
Folkhalsomyndigheten_Covid19_Apr 14 2021.xlsx	2021/9/17 22:13	Microsoft Excel ...	1,102 KB
Folkhalsomyndigheten_Covid19_Apr 15 2020.xlsx	2021/9/17 10:20	Microsoft Excel ...	21 KB
Folkhalsomyndigheten_Covid19_Apr 15 2021.xlsx	2021/9/17 22:14	Microsoft Excel ...	1,119 KB
Folkhalsomyndigheten_Covid19_Apr 16 2020.xlsx	2021/9/17 10:21	Microsoft Excel ...	21 KB

Figure 1: Data files(parts)

```
library(config)
library(htmlwidgets)
```

4.1 Config

before start the process we need to do the config like we did before, but this time for the R script.

```
default:
  start_time: "2020/04/03"
  end_time: "2021/12/28"
  file_add_and_name: './data/Folkhalsomyndigheten_Covid19_'
  # Sheet type
  Antal per dag region: 1
  Antal avlidna per dag: 1
  Antal intensivvårdade per dag: 1
  Totalt antal per region: 2
  Totalt antal per kön: 2
  Totalt antal per åldersgrupp: 2
  Veckodata Region: 3
  Veckodata Kommun_stadsdel: 3
  Veckodata Riket: 3
  # Combined sheet
  Antal per dag region_on_kön: [4,5]
  Antal per dag Åldersgrupp_on_kön: [6,5]
  Antal per dag region_on_Åldersgrupp: [4,6]
```

4.2 Functions in the script

4.2.1 Data file reading

4.2.2 Visualing

4.2.3 Main script

4.2.4 Output

5 Result Anlysis

fig_Antal_avlidna_per_dag.html

fig_Antal_intensivvårdade_per_dag.html

fig_Antal_per_dag_Åldersgrupp_on_kön.html

fig_Antal_per_dag_region.html

fig_Antal_per_dag_region_on_Åldersgrupp.html

fig_Antal_per_dag_region_on_kön.html

fig_Totalt_antal_per_åldersgrupp.html

fig_Totalt_antal_per_kön.html

fig_Totalt_antal_per_region.html

6 Conclusion

7 Appendix

Code:

Code:

Chromium, and WebDriver teams. n.d. “ChromeDriver - WebDriver for Chrome.” Accessed December 30, 2021. https://chromedriver.chromium.org/.

Conservancy, Software Freedom. n.d. “Selenium.” Accessed December 30, 2021. https://www.selenium.dev/.