## Naver Café Crawling



Park byeong jo





- Setting
- Source Code
- Result
- Reference





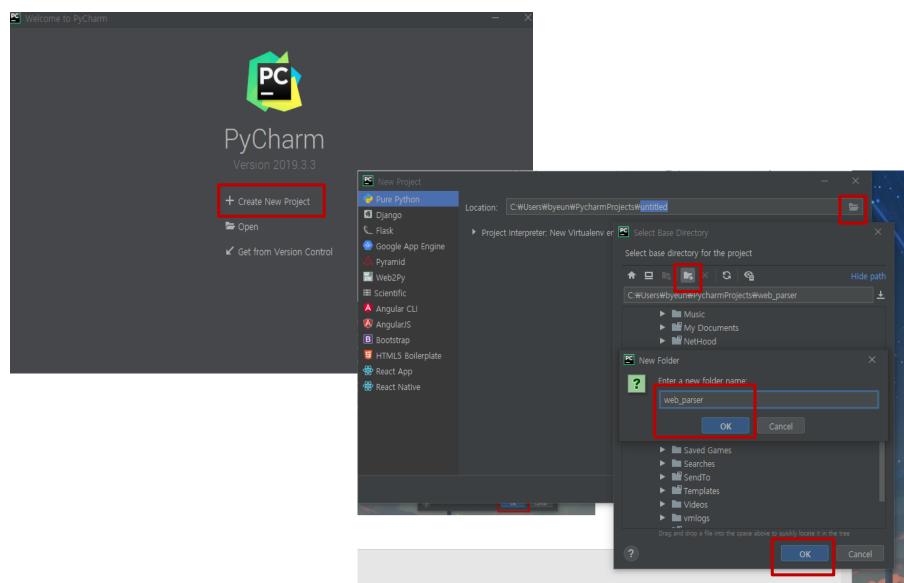
# Setting Setting

- ◆ Tool
  - Pycharm
- Python version: 3.8
- Chromedriver
  - Chrome > Chrome://settings/help > version check
  - https://chromedriver.chromium.org/downloads > now version download
- Install Packages
  - Selenium
  - Django
  - **❖** Bs4







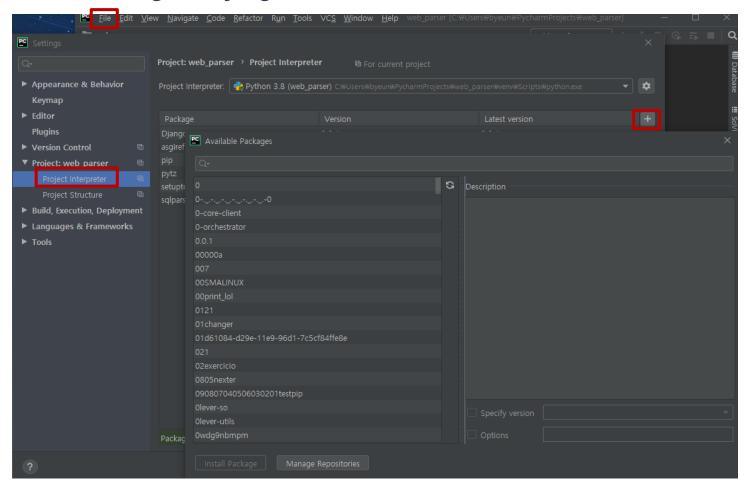








- ♦ File > Settengs > Project:name > Project interperter
- ♦ Install Packeges : Django, selenium, bs4

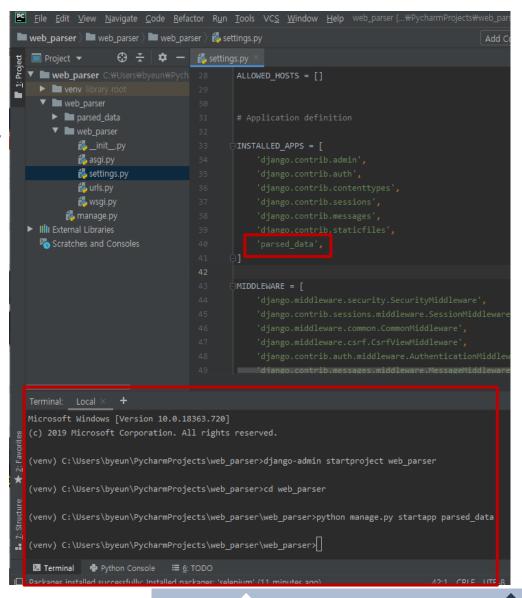




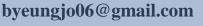




- Terminal setting
  - Django-admin startproject name\_1
  - ❖ cd name\_1
  - python manage.py startapp name\_2
- Setting.py
  - ❖ Add name\_2 to INSTALLD\_APPS

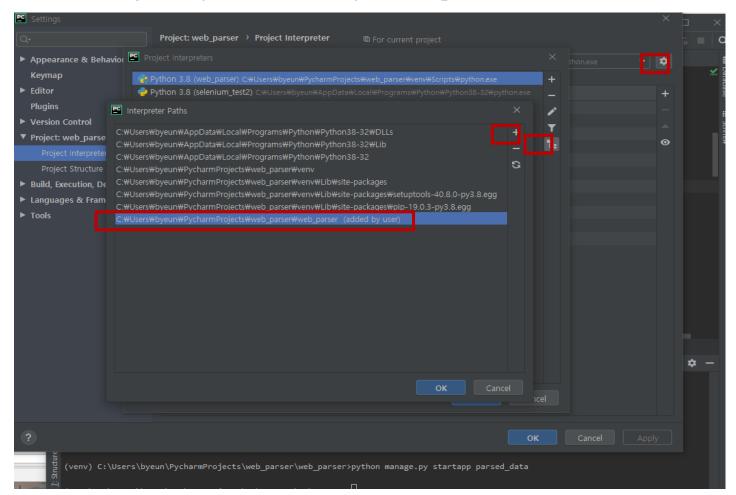








- Add Path
  - File > Setting > Project:name > Project Interpreters

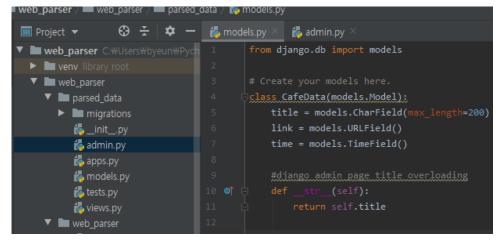


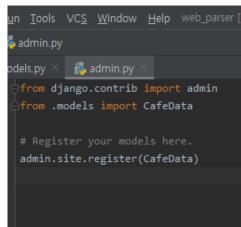






- DB Setting
  - models.py
  - admin.py





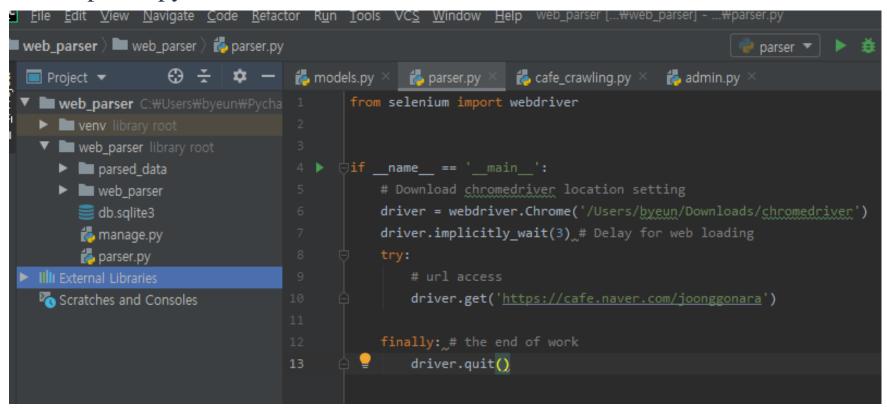
- ❖ DB create : Terminal
  - ◆ python manage.py migrate
  - python manage.py makemigrations parsed\_data
  - python manage.py migrate parsed\_data
- Django Admin
  - ◆ python manage.py createsupseruser
  - ◆ Don not need to Email







#### Create parser.py









#### Create modules > café\_crawling.py

```
File Edit View Navigate Code Refactor Run Tools VCS Window Help web_parser [...\web_parser] - ...\cafe_crawling.py
■ web_parser 〉 ■ web_parser 〉 ■ parsed_data 〉 ■ modules 〉 & cafe_crawling.py
                                                                                                                                                                                                                                                                                                                                                      parser *
          ■ Project ▼

    — $\begin{align*} \text{models.py} \times \begin{align*} \begin{align*} \text{depth} \text{parser.py} \times \begin{align*} \begin{align*} \text{depth} \text{cafe_crawling.py} \text{\text{o}} \\
\text{depth} \text{models.py} \times \begin{align*} \begin{align*} \text{depth} \text{parser.py} \times \begin{align*} \begin{align*} \text{depth} \text{cafe_crawling.py} \text{\text{o}} \\
\text{depth} \text{depth} \text{depth} \text{depth} \text{depth} \\
\text{depth} \text{depth} \text{depth} \text{depth} \\
\text{depth} \text{depth} \text{depth} \\
\text{depth} \text{depth} \text{depth} \\
\text{depth} \text{depth} \\
\text{depth} \text{depth} \\
\text{d
                                                                                                                                                                                                                                                                                                              ઢ admin.py ×
                web_parser C:\Users\byeun\Pychar 1
                                                                                                                                                                from bs4 import BeautifulSoup
                 venv library root
                                                                                                                                                          ▼ web_parser library root
                          ▼ parsed_data
                                  migrations
                                                                                                                                                           dclass cafe data:...
                                   ▼ ■ modules
                                                     afe_crawling.py
                                                                                                                                                                # cafe-menu search and page change
                                            🛵 __init__.py
                                                                                                                                                               def parse cafe(driver, query, data time):...
                                            🖧 admin.py
                                            🖧 apps.py
                                            models.py
                                            tests.py
                                                                                                                                                               def page cafe(driver, data, data time):...
                                           💤 views.py
                          web_parser
                                  db.sqlite3
                                                                                                                                                               def crawling page(driver, data, data time):...
                                   the manage.py
                                   aparser.py
        ▶ III External Libraries
                                                                                                                                                            def time check(cafe time, data time):...
                Scratches and Consoles
```







#### café\_crawling.py

```
web_parser \ \bigcap web_parser \ \bigcap parsed_data \ \bigcap modules \ \cdot \ \bigcap cafe_crawling.py
■ Project ▼
                                     nodels.py × 🐉 parser.py × 📸 cafe_crawling.py
                                                                                            🖧 admin.py
  web_parser C:\Users\byeun\Pychar
  ▶ iii venv library root
  ▼ web_parser library root
                                              ⇒class cafe data:
     ▼ ■ parsed_data
                                                    def __init__(self, title, link, time):
                                                        self.title = title
       ▶ migrations
                                                        self.link = link
        ▼ Immodules
             afe_crawling.py
                                                        self.time = time
          init_.py
          admin.py
                                              def parse cafe(driver, query, data time):
           apps.py
          models.py
          tests.py
                                                   driver.get("https://cafe.naver.com/joonggonara")

<sup>™</sup> views.py

                                                   time.sleep(1)
     ▶ ■ web_parser
       db.sqlite3
                                                   driver.find_element_by_partial_link_text(query).click()
        the manage.py
        aparser.py
                                                    time.sleep(1)
                                                   driver.switch_to.frame('cafe_main')
  IllI External Libraries
  Scratches and Consoles
                                                   data = []
                                                        t = page_cafe(driver, data, data_time)
                                                    return data
```





#### café\_crawling.py

```
🛮 web_parser 🕽 🖿 web_parser 🥽 🖿 parsed_data 🕽 🖿 modules 🕽 💪 cafe_crawling.py
                                                                                                                                                                                                                                                                                                                                        🌍 parser 🔻 🕨 🇯 👣 🚱
   ■ Project ▼

    parser.py 
    ★ cafe_crawling.py

                                                                                                                                                                                                                                                admin.py
  ▼ web_parser C:\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\Users\User
        ▶ wenv library root
                                                                                                                          def page cafe(driver, data, data time):
        ▼ I web_parser library root
                                                                                                                                        soup = BeautifulSoup(driver.page_source, 'html.parser')
                ▼ parsed_data
                       ▶ ■ migrations
                                                                                                                                       count = 15
                       ▼ I modules
                                                                                                                                       for t in hf:
                                      🖧 cafe_crawling.py
                              🛵 __init__.py
                              🛵 admin.py
                              apps.py
                                                                                                                                                              driver.find element by link text(t.text).click()
                              the models.py
                              tests.py
                                                                                                                                                   elif t.text == '이전':
                              📥 views.py
               ▶ web parser
                       db.salite3
                                                                                                                                                             driver.find_element_by_link_text(t.text).click()
                        the manage.py
                                                                                                                                                             count = crawling_page(driver, data, data_time)
                        aparser.py
        IllI External Libraries
        Scratches and Consoles
                                                                                                                            def crawling page(driver, data, data time):
                                                                                                                                       soup = BeautifulSoup(driver.page_source, 'html.parser')
                                                                                                                                       cafe title = soup.select('div:nth-child(6) > table > tbody > tr > td > div > div > a.article')
                                                                                                                                       cafe time = soup.select('div:nth-child(6) > table > tbody > tr > td.td date')
                                                                                                                                       count = 0
                                                                                                                                       for t in cafe title:
                                                                                                                                                  if ti is True:
                                                                                                                                                             data.append(cafe data(t.text.strip(), t.get('href'), cafe time[count].text))
                                                                                                                                                             count += 1
```





### café\_crawling.py

```
web_parser > web_parser > parsed_data > modules > 💪 cafe_crawling.py
■ Project ▼
                             🌣 💳 🐉 models.py × 🚜 parser.py × 🚜 cafe_crawling.py ×
                                                                                        🖧 admin.py 🗵
  web_parser C:\Users\byeun\Pychar 1
                                             from bs4 import BeautifulSoup
  ▶ wenv library root
  ▼ web_parser library root
                                             import time
    ▼ ■ parsed data
       ▶ ■ migrations
       ▼ ■ modules
                                             class cafe data:...
             afe_crawling.py
          🛵 __init__.py
          admin.py
                                             def parse cafe(driver, query, data time):...
          apps.py
          nodels.py
                                             def page cafe(driver, data, data time):...
          tests.py
          a views.py
    ▶ ■ web_parser
                                             def crawling page(driver, data, data time):...
       db.sqlite3
       the manage.py
       aparser.py
                                             def time check(cafe time, data time):
  Illi External Libraries
                                                 if cafe_time.text.find(':') == -1 or cafe time.text < data time:</pre>
  Scratches and Consoles
```





13



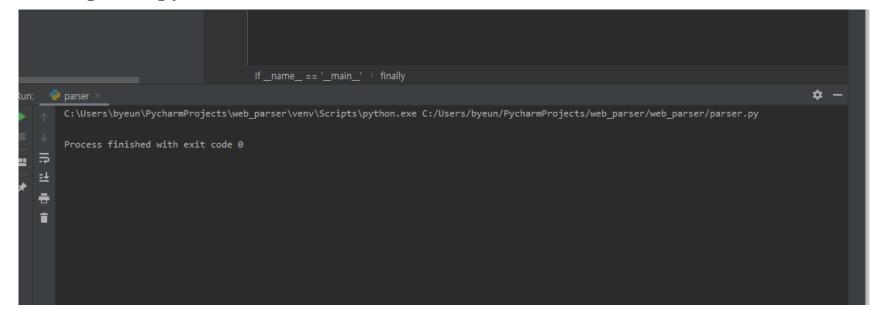
#### parser.py

```
ile <u>Edit View Navigate Code Refactor Run Tools VCS Window Help</u> web_parser [C:\Users\byeun\PycharmProjects\web_parser] -
eb parser 🕽 🖿 web parser 🕽 ち parser.py
                              ─ ♣ models.py × ♣ parser.py
                                                                 🛵 admin.py 🗡
                                                                                ち settings.py 🗡
                                                                                               👗 manage.py
                                          from selenium import webdriver
 web_parser C:\Users\byeun\Pychar 1
                                           from parsed data.modules.cafe crawling import parse cafe
  venv library root
  web_parser library root
                                           import os
                                          import django
   ▼ ■ parsed data
     migrations
                                           os.environ.setdefault("DJANGO_SETTINGS_MODULE", "web_parser.settings")
     ▼ ■ modules
                                           django.setup()
                                           from parsed data.models import CafeData
           afe_crawling.py
        _init_.py
                                          if name == ' main ':
        🛵 admin.py
        🛵 apps.py
                                               driver = webdriver.Chrome('/Users/byeun/Downloads/chromedriver')
        the models.py
        tests.py
                                               driver.implicitly_wait(3)_# Delay for web loading
        to views.py
   ▼ web_parser
                                                   driver.get('https://cafe.naver.com/joonggonara')
        _init_.py
                                                   data = parse cafe(driver, '컴퓨터', '07:00')
        📥 asgi.py
        settings.py
        ち urls.py
                                                   for t in data:
        🛵 wsgi.py
                                                       CafeData(title=t.title, link=t.link, time=t.time).save()
     db.sqlite3
     the manage.py
     aparser.py
                                                   driver.quit()
IIII External Libraries
  Scratches and Consoles
```





### Run parser.py









### ◆ Terminal > python manage.py runserver

```
If _name_ == '_main_' > finally

Terminal: _Local × +

(venv) C:\Users\byeun\PycharmProjects\web_parser\web_parser>python manage.py runserver

Watching for file changes with StatReloader

Performing system checks...

System check identified no issues (0 silenced).

March 17, 2020 - 15:49:58

Django version 3.0.4, using settings 'web_parser.settings'

Starting development server at <a href="http://127.0.0.1:8000/">http://127.0.0.1:8000/</a>

Quit the server with CTRL-BREAK.

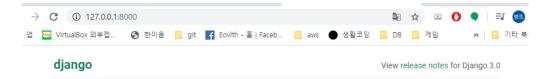
▶ 4: Run | □ 6: TODO □ Terminal ◆ Python Console
```







http://127.0.0.1:8000/





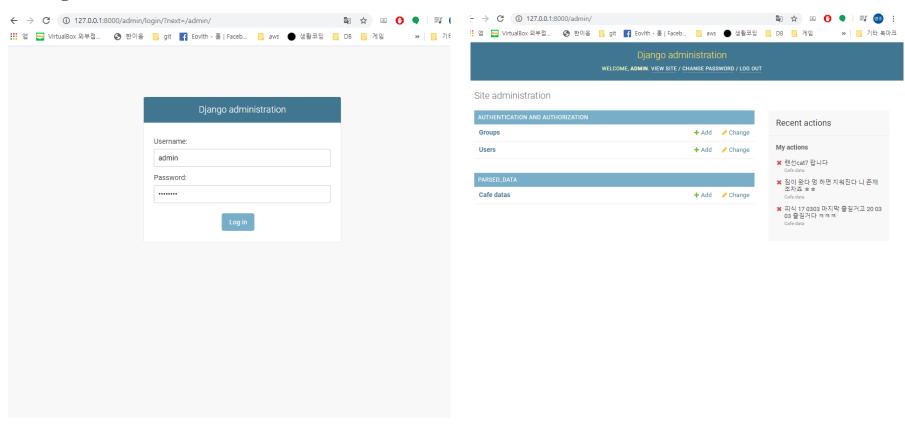
#### The install worked successfully! Congratulations!

You are seeing this page because DEBUG=True is in your settings file and you have not configured any URLs.





#### • http://127.0.0.1:8000/admin

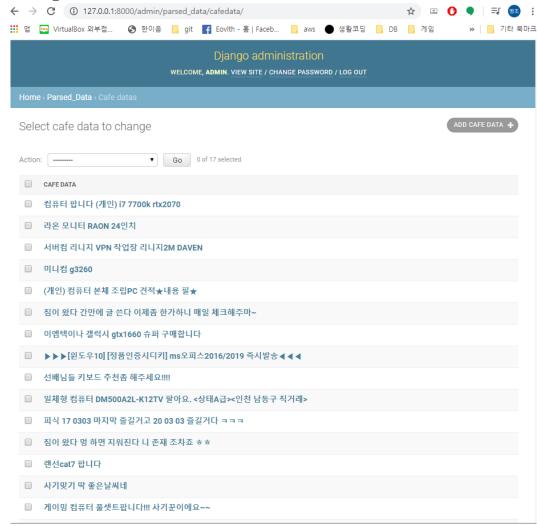








#### http://127.0.0.1:8000/admin









- https://docs.djangoproject.com/ko/3.0/intro/
- https://beomi.github.io/gb-crawling/

