

37. 인터넷에 있는 데이터를 수집하는 방법



1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링

인터넷에 있는 데이터를 가져오는 방법

1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링

인터넷은 큰 틀로 보면, HTML이라는 언어로 구성되어 있다.

따라서 Python으로 HTML을 다루어야 하며, 페이지 요청을 할 수 있어야 한다.

1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링

라이브러리

requests : 페이지 요청 라이브러리

bs4의 BeautifulSoup : Python으로 HTML을 다루는 라이브러리

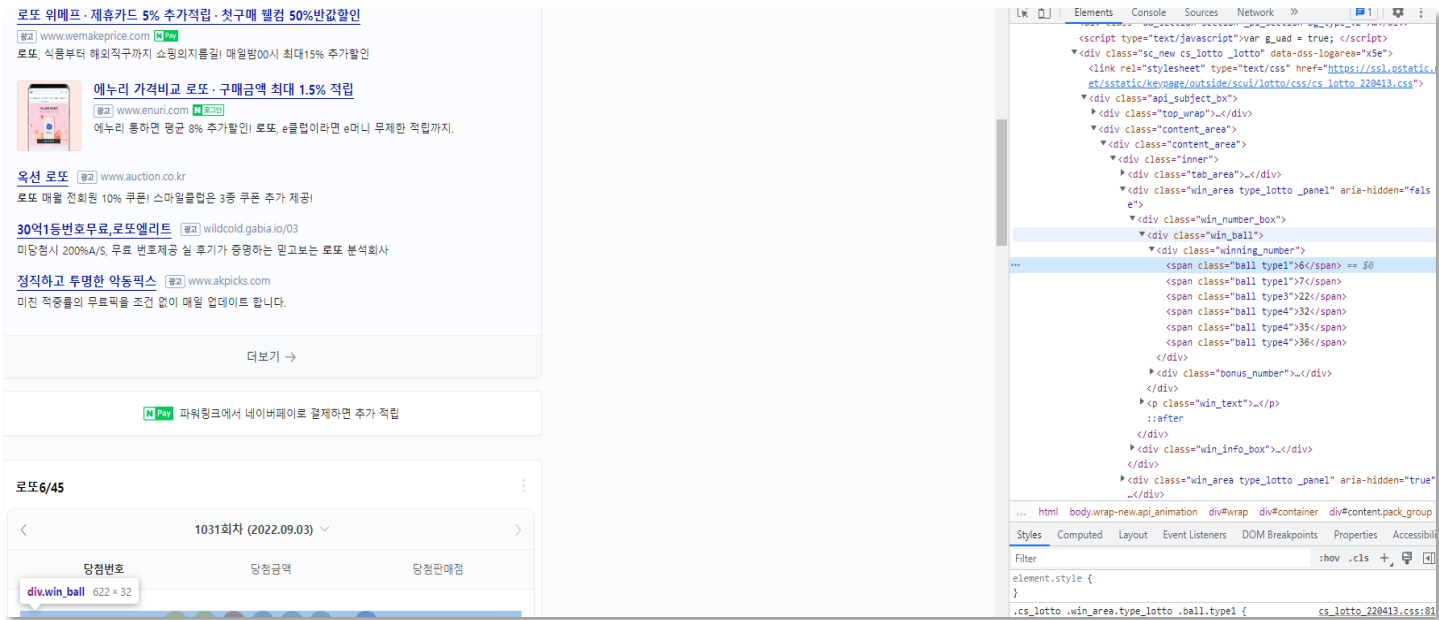
time : 안전하게 수집하는 방법

pandas : 수집한 데이터 엑셀화

37. 인터넷에 있는 데이터를 수집하는 방법

1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링



1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링

태그명

```
<div class="win_number_box">
  <div class="win_ball">
    <div class="winning_number">
      <span class="ball type1">6</span> == 8
      <span class="ball type1">7</span>
      <span class="ball type3">22</span>
      <span class="ball type4">32</span>
      <span class="ball type4">35</span>
      <span class="ball type4">36</span>
    </div>
```

속성값

1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링

requests 활용

requests.get(링크) 를 하면 페이지를 요청할 수 있음

1. 인터넷에 있는 데이터를 수집하는 방법

데이터 크롤링

BeautifulSoup 활용

BeautifulSoup(url.text) 를 하면 수집한 데이터를 HTML로 변환할 수 있음

find : 해당 HTML을 찾아서 HTML로 반환

find_all : 해당 HTML을 모두 찾아서 리스트로 반환