

데이터 크롤링과 정제

과제3: 네이버 증시 정보 크롤링

과제 #3

- 시가 총액 10위까지의 기업 정보를 크롤링
 - 네이버 금융 웹사이트
 - https://finance.naver.com/sise/sise_market_sum.naver
 - 크롤링 항목 7개 출력
 - 종목명, 종목코드, 현재가, 전일가, 시가, 고가, 저가

삼성전자		005930	코스피	📅 2022.07.22 기준 (장마감)	🕒 실시간	📄 기업개요 ▾
61,300	전일 61,800		고가 62,200 (상한가 80,300)		거래량 10,235,096	
	전일대비 ▼500 -0.81%		시가 61,800		저가 61,200 (하한가 43,300) 거래대금 630,265 백만	

- 네이버 웹사이트의 한글 깨짐 문제
 - requests 라이브러리 사용

```
import requests

html = requests.get(url)
soup = BeautifulSoup(html.text, 'html.parser')
```

- 왼쪽 “시가 총액” 메뉴에서 상위 10개 기업 정보 크롤링

- ```
<!-- global include -->
<html lang="ko" data-useragent="Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.5 Safari/605.1.15">
 <head></head>
 <body onload="getGNB();">
 <script type="text/javascript">var nclk_evt = 3; nclk_do();
 </script>
 <script type="text/javascript">...</script>
 <div id="u_skip">...</div>
 <div id="header">...</div>
 <div id="wrap">
 <div class="banner_smart">...</div>
 <!-- // global include -->
 <script language="JavaScript">...</script>
 <div id="newarea">
 <!-- leftmenu -->
 <div class="snb snb2">...</div>
 <!-- //leftmenu -->
 <!-- //leftmenu -->
 <div id="contentarea">
 <!-- 컨텐츠 -->
 <div id="contentarea_left" class="bnd_wid">...</div>
 <!-- right area -->
 <div id="contentarea_right"> </div>
 <!-- 종속 -->
 <div class="box_type_l">
 <div class="tab_style_1">...</div>
 <!-- [D] 활성화된 탭메뉴에 따라 blind text 변경해주세요 -->
 <h4 class="blind">코스피</h4>
 <table summary="코스피 시세정보를 선택한 항목에 따라 정보를 제공합니다." cellpadding="0" cellspacing="0" class="type_2">
 <caption>코스피</caption>
 <colgroup>...</colgroup>
 <thead>...</thead>
 <tbody>
 <tr>...</tr>
 <tr onmouseover="mouseOver(this)" onmouseout="mouseOut(this)" style="background-color: rgb(255, 255, 255);">...</tr>
 <tr onmouseover="mouseOver(this)" onmouseout="mouseOut(this)">...</tr>
 <tr onmouseover="mouseOver(this)" onmouseout="mouseOut(this)">...</tr>
 </tbody>
 </table>
 </div>
 </div>
 </div>
 </div>
```

# 네이버 주식: 시가 총액현황 #2

- 각 회사별 URL 및 이름을 리스트에 저장

`<a href="/item/main.naver?code=00590", class="title">삼성전자 </a>`

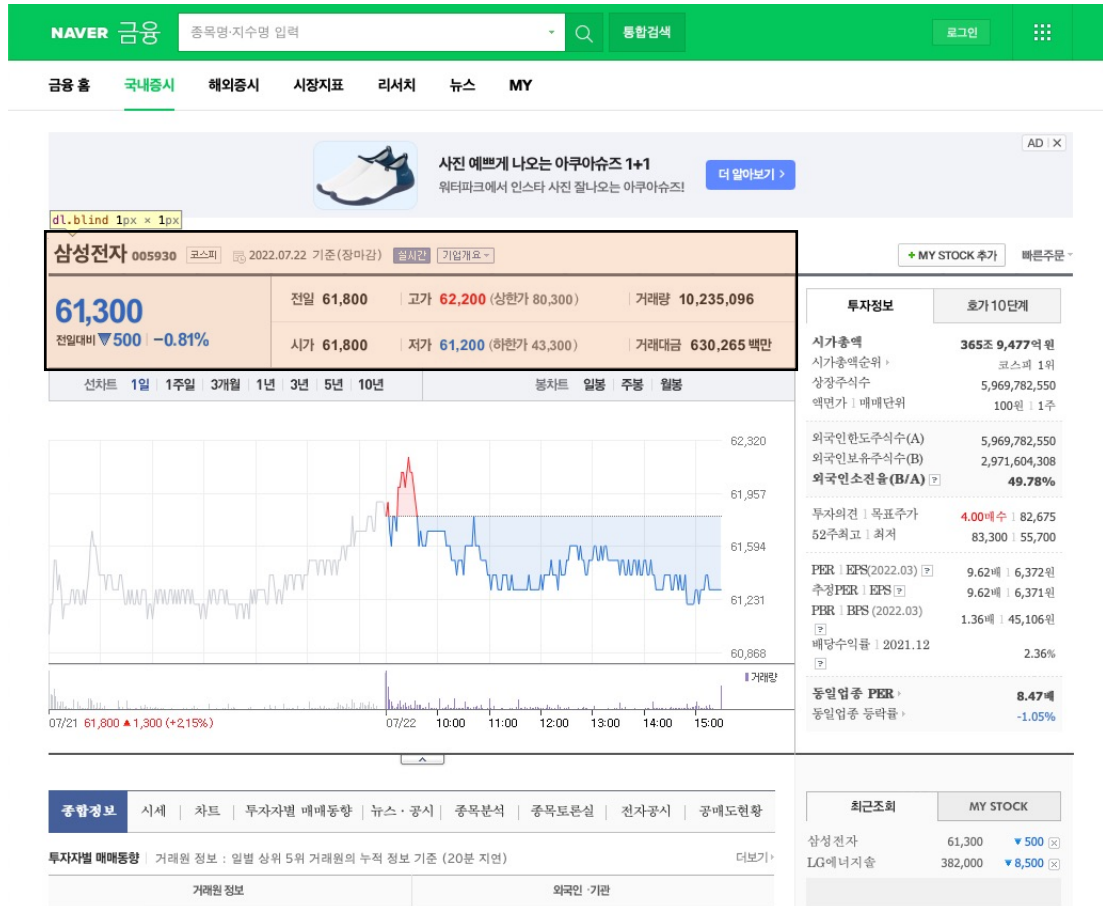
	코스피	코스닥								
N	가	전일비	등락률	액면가	시가총액	상장주식수	외국인비율	거래량		
1	삼성전자	61,300	▼ 500	-0.81%	100	3,659,477	5,969,783	49.78	10,235,096	
2	LG에너지솔루션	382,000	▼ 8,500	-2.18%	500	893,880	234,000	3.26	299,000	
3	SK하이닉스	100,000	▼ 2,500	-2.44%	5,000	728,002	728,002	49.91	3,101,000	
4	삼성바이오로직스	829,000	▲ 12,000	+1.47%	2,500	590,032	71,174	10.71	55,000	
5	삼성전자우	56,200	▼ 500	-0.88%	100	462,462	822,887	72.16	611,000	
6	NAVER	250,000	▼ 500	-0.20%	100	410,123	164,049	53.31	464,000	
7	현대차	191,000	▲ 2,000	+1.06%	5,000	408,106	213,668	26.93	1,371,000	
8	LG화학	561,000	▼ 8,000	-1.41%	5,000	396,023	70,592	47.85	160,000	
9	삼성SDI	568,000	▲ 1,000	+0.18%	5,000	390,583	68,765	42.84	142,000	
10	기아	80,900	▼ 1,100	-1.34%	5,000	327,939	405,363	36.20	2,159,000	
11	카카오	72,100	▼ 1,600	-2.17%	100	320,808	444,948	28.80	1,583,000	
12	셀트리온	179,500	▼ 3,000	-1.64%	1,000	252,678	140,768	20.74	213,000	
13	삼성물산	117,000	▼ 1,000	-0.85%	100	218,658	186,887	16.38	170,000	
14	현대모비스	218,500	▼ 1,000	-0.46%	5,000	206,642	94,573	34.17	186,000	
15	POSCO홀딩스	234,000	▼ 1,000	-0.43%	5,000	204,017	87,187	53.01	167,000	

```

<!-- leftmenu -->
<div class="snb snb2">...</div>
<!-- //leftmenu -->
<!-- //leftmenu -->
<div id="contentarea">
 <!-- 컨텐츠 -->
 <div id="contentarea_left" class="bnd_wid">...</div>
 <!-- right area -->
 <div id="contentarea_right"> </div>
 <!-- 종목 -->
 <div class="box_type_l">
 <div class="tab_style_1">...</div>
 <!-- [D] 활성화된 탭메뉴에 따라 blind text 변경해주세요 -->
 <h4 class="blind">코스피</h4>
 <table summary="코스피 시세정보를 선택한 항목에 따라 정보를 제공합니다." cellpadding="0" cellspacing="0" class="type_2">
 <caption>코스피</caption>
 <colgroup>...</colgroup>
 <thead>...</thead>
 <tbody>
 <tr>...</tr>
 <tr onmouseover="mouseOver(this)" onmouseout="mouseOut(this)" style="background-color: rgb(255, 255, 255);"> = $0
 <td class="no"></td>
 <td>
 삼성전자
 </td>
 <td class="number">61,300</td>
 <td class="number">...</td>
 <td class="number">...</td>
 <td class="number">100</td>
 <td class="number">3,659,477</td>
 <td class="number">5,969,783</td>
 <td class="number">49.78</td>
 <td class="number">10,235,096</td>
 </tr>
 </tbody>
 </table>
 </div>

```

# 삼성전자 주식 세부 현황 페이지



```
<html lang="ko" data-useragent="Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.5 Safari/605.1.15">
<head></head>
<body onload="getGNB();">
 <script type="text/javascript">var nclk_evt = 3;
 nclk_do();</script>
 <script type="text/javascript"></script>
 <div id="u_skip">

 </div>
 <div id="header">
 <div class="snb_area"></div>
 <div class="lnb_area"></div>
 <script type="text/javascript"></script>
 <iframe id="me_layers" name="test" title="네이버미 영역" width="0" height="0" scrolling="no" frameborder="0" style="display:block; top: 22px; right: 209px; position: absolute; z-index: 15;"></iframe>
 </div>
 <div id="wrap">
 <div class="banner_smart"></div>
 <script language="javascript" src="https://ssl.pstatic.net/imgstock/static.pc/20220719195125/js/ellipse.js"></script>
 <script type="text/javascript"></script>
 <script language="javascript" src="https://ssl.pstatic.net/imgstock/static.pc/20220719195125/js/recent_code.js"></script>
 <script language="javascript">addCode('005930');</script>
 <script language="JavaScript" src="https://ssl.pstatic.net/imgstock/static.pc/20220719195125/js/summary.js"></script>
 <script language="JavaScript">fchartStatus.setChartStatus('on');</script>
 <style>.hide_chart { position: absolute; top: 0; left: -9999px;}</style>
 </div>
 <div id="middle" class="new_totalinfo">
 <div class="blind">
 <dt>종목 상세 정보</dt>
 <dd>2022년 07월 22일 16시 11분 기준 정마감</dd>
 <dd>종목명 삼성전자</dd>
 <dd>종목코드 005930 코스피</dd>
 <dd>현재가 61,300 전일대비 하락 500 마이너스 0.81 퍼센트</dd>
 <dd>전일가 61,800</dd>
 <dd>시가 61,800</dd>
 <dd>고가 62,200</dd>
 <dd>상한가 80,300</dd>
 <dd>저가 61,200</dd>
 <dd>하한가 43,300</dd>
 <dd>거래량 10,235,096</dd>
 <dd>거래대금 630,265백만</dd>
 </div>
 </div>
 <div class="h_company"></div>
 <div class="content_wrap"></div>
 <div id="footer"></div>
 <script type="text/javascript"></script>
</div>
</body>
</html>
```

# 프로그램 동작 및 실행 결과

## ■ 프로그램 동작

- 코스피 상위 10개 기업의 URL 및 기업 이름을 리스트에 저장
- 메뉴에서 선택한 기업의 세부 주식 정보를 화면에 출력
- 사용자가 -1을 입력할 때 까지 계속 반복함

```

[네이버 코스피 상위 10대 기업 목록]

```

```
[1] 삼성전자
[2] LG에너지솔루션
[3] SK하이닉스
[4] 삼성바이오로직스
[5] 삼성전자우
[6] NAVER
[7] 현대차
[8] LG화학
[9] 삼성SDI
[10] 기아
```

```
주가를 검색할 기업의 번호를 입력하세요(-1: 종료): 1
```

```
https://finance.naver.com/item/main.naver?code=005930
```

```
종목명: 삼성전자
```

```
종목코드: 005930
```

```
현재가: 61,300
```

```
전일가: 61,800
```

```
시가: 61,800
```

```
고가: 62,200
```

```
저가: 61,200
```

# 실행 결과

[ 네이버 코스피 상위 10대 기업 목록 ]

- [ 1] 삼성전자
- [ 2] LG에너지솔루션
- [ 3] SK하이닉스
- [ 4] 삼성바이오로직스
- [ 5] 삼성전자우
- [ 6] NAVER
- [ 7] 현대차
- [ 8] LG화학
- [ 9] 삼성SDI
- [10] 기아

주가를 검색할 기업의 번호를 입력하세요(-1: 종료): 7

<https://finance.naver.com/item/main.naver?code=005380>

종목명: 현대차

종목코드: 005380

현재가: 191,000

전일가: 189,000

시가: 191,000

고가: 193,000

저가: 188,000

[ 네이버 코스피 상위 10대 기업 목록 ]

- [ 1] 삼성전자
- [ 2] LG에너지솔루션
- [ 3] SK하이닉스
- [ 4] 삼성바이오로직스
- [ 5] 삼성전자우
- [ 6] NAVER
- [ 7] 현대차
- [ 8] LG화학
- [ 9] 삼성SDI
- [10] 기아

주가를 검색할 기업의 번호를 입력하세요(-1: 종료): -1  
프로그램 종료



# Questions?