

```
import gdown

gdown.download('https://bit.ly/3q9SZix', '20s_best_book.json', quiet=False)

Downloading...
From: https://bit.ly/3q9SZix
To: /content/20s_best_book.json
100%|██████████| 92.9k/92.9k [00:00<00:00, 53.6MB/s]
'20s_best_book.json'
```

```
import pandas as pd

books_df = pd.read_json('20s_best_book.json')
books_df.head()
```



	no	ranking	bookname	authors	publisher	publication_year	isbn13	addi
0	1	1	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽	허블	2019	9791190090018	
1	2	2	달려구트 꿈 백화점. 이미예 장편소설	지은이: 이미예	팩토리나인	2020	9791165341909	
2	3	3	지구에서 한아뿐 :정세랑 장편소설	지은이: 정세랑	난다	2019	9791188862290	
3	4	4	시선으로 부터 :정세랑 장편소설	지은이: 정세랑	문학동네	2020	9788954672214	
4	5	5	아몬드 :손원평 장편소설	지은이: 손원평	창비	2017	9788936434267	

Next steps:

Generate code with books_df

View recommended plots

```
books = books_df[['no', 'ranking', 'bookname', 'authors', 'publisher', 'publication_year', 'isbn13']]
books.head()
```

	no	ranking	bookname	authors	publisher	publication_year	isbn13		
0	1	1	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽	허블	2019	9791190090018		
1	2	2	달려구트 꿈 백화점. 이미예 장편소설	지은이: 이미예	팩토리나인	2020	9791165341909		
2	3	3	지구에서 한아뿐 :정세랑 장편소설	지은이: 정세랑	난다	2019	9791188862290		

Next steps:

Generate code with books

View recommended plots

```
books_df.loc[[0,1], ['bookname', 'authors']]
```

	bookname	authors
0	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽
1	달려구트 꿈 백화점.이미에 장편소설	지은이: 이미에

```
books_df.loc[0:1, 'bookname':'authors']
```

	bookname	authors
0	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽
1	달려구트 꿈 백화점.이미에 장편소설	지은이: 이미에

```
books = books_df.loc[:, 'no':'isbn13']
books.head()
```

	no	ranking	bookname	authors	publisher	publication_year	isbn13
0	1	1	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽	허블	2019	9791190090018
1	2	2	달려구트 꿈 백화점.이미에 장편소설	지은이: 이미에	팩토리나인	2020	9791165341909
2	3	3	지구에서 한아뿐 :정세랑 장편소설	지은이: 정세랑	난다	2019	9791188862290

Next steps:

[Generate code with books](#)[View recommended plots](#)

```
books_df.loc[:, 'no':'isbn13'].head()
```

	no	ranking	bookname	authors	publisher	publication_year	isbn13
0	1	1	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽	허블	2019	9791190090018
2	3	3	지구에서 한아뿐 :정세랑 장편소설	지은이: 정세랑	난다	2019	9791188862290
4	5	5	아몬드 :손원평 장편소설	지은이: 손원평	창비	2017	9788936434267

```
import requests
```

```
isbn = 9791190090018 # '우리가 빛의 속도로 갈 수 없다면'의 ISBN
url = 'http://www.yes24.com/Product/Search?domain=B00K&query={}'
```

```
r = requests.get(url.format(isbn))
```

```
print(r.text)
```

```

</div>
<div class="yFootCon">
  <dl class="yFootAddr">
    <dt><a href="https://www.yes24.com/Main/default.aspx"><em class="fLogo bgFooter">&n
    <dd>
      <address>
        <span class="addrRow">대표 : 김석환, 최세라</span>
        <span class="addrRow">주소 : 서울시 영등포구 은행로 11, 5층~6층(여의도동, 일
        <span class="addrRow">사업자등록번호 : 229-81-37000 &nbsp;&nbsp;&nbsp; 통신판매업신고 :
        <span class="addrRow">이메일 : yes24help@yes24.com &nbsp;&nbsp;&nbsp; 호스팅 서비스사업
      <p class="txt_copyright">Copy right © YES24 Corp. All Rights Reserved.</p>
    </address>
  </dd>
</dl>
<ul class="yFootCsLi">
  <li class="csTel">
    <dl>
      <dt style="">고객문의</dt>
      <dd>
        <p><span class="item">도서/음반</span> <span class="yes_b"
        <p><span class="item w35">중고샵</span> <span class="yes_b"
        <p><span class="item">티켓에매</span> <span class="yes_b"
        <p><span class="item w35"><a href="https://www.yes24.com/Ma
      </dd>
    </dl>
  </li>
  <li class="csFaq clearfix">
    <a href="https://www.yes24.com/Mall/Help/Main" class="btnCs1 bgFooter">1:1
    <a href="https://www.yes24.com/Mall/Help/Faq" class="btnCs2 bgFooter">자주
    <a href="javascript:void(0);" class="btnCs3 bgFooter">상담시간 안내</a>
  </li>
</ul>
<div id="footerCsInfoPop" class="csInfoD" style="display:none">
  <div class="csInfoDCon">
    <strong class="bgFooter tit">상담시간 안내</strong>
    <ul class="infoDli">
      <li><strong class="item">도서/음반 문의</strong> <em class="telNum">1544-3800</em> 평일 9시 ~ 1
        <li><strong class="item">중고샵 문의</strong> <em class="telNum">15
        <li><strong class="item">티켓에매 문의</strong> <em class="telNum">
        <li><strong class="item">eBook 문의</strong> <a href="https://www.y
      </ul>
      <p class="csInfoCmt">점심시간 안내 - 티켓 12:00 ~ 13:00</p>
    <a href="javascript:void(0);" class="btnCloseCsInfo">상담시간 안내 레이어 닫기</a>
  </div>
</div>
<div class="yes24Glory">

```

```
from bs4 import BeautifulSoup
```

```
soup = BeautifulSoup(r.text, 'html.parser')
```

```
prd_link = soup.find('a', attrs={'class':'gd_name'})
```

```
print(prd_link)
```

```
<a class="gd_name" href="/Product/Goods/74261416" onclick="wiseLogV2('S', '101_005_003_001', ''); setGoodsClickExtraCodeHub('C
```

```
print(prd_link['href'])
```

```
/Product/Goods/74261416
```

```
# '우리가 빛의 속도로 갈 수 없다면'의 상세 페이지 가져오기
```

```
url = 'http://www.yes24.com'+prd_link['href']
```

```
r = requests.get(url)
```

```
print(r.text)
```

```

<script type="text/javascript" src="https://image.yes24.com/sysimage/Contents/Scripts/p/jquery/jquery.easing.1.3.mi
<script type="text/javascript" src="https://www.yes24.com/JavaScript/util.js?v=20230919"></script>
<script type="text/javascript" charset="euc-kr" src="https://www.yes24.com/JavaScript/Yes_header.js?v=20230620"></s
<script type="text/javascript" src="https://www.yes24.com/JavaScript/hiveSlide.js?v=20240321a"></script>
<script type="text/javascript" src="https://www.yes24.com/JavaScript/jqueryExtends.js?v=20240321a"></script>
<script type="text/javascript" src="https://www.yes24.com/JavaScript/recentviewgoods.js?v=20240307"></script>

<script type="text/javascript" src="https://www.yes24.com/JavaScript/Search/bulletSearch.js?v=20240228"></script>
<script type="text/javascript" src="https://www.yes24.com/24/Scripts/YES24.Common-1.6.js?v=20240226"></script>

<script type="text/javascript" src="https://www.yes24.com/JavaScript/redirectWebSite.js?v=20240223"></script>
<script type="text/javascript" src="https://image.yes24.com/sysimage/Contents/Scripts/d/scl/simplebar.js?v=20240222

<script type="text/javascript" src="https://image.yes24.com/sysimage/Contents/Scripts/p/jquery/jquery.lazyload.min.

<script type="text/javascript" src="https://image.yes24.com/sysimage/Contents/Scripts/p/ui/openinside.pack.js?v=202
<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/hack/Newstyle.css?v=20211122a"
<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/hack/renewHack.css?v=20210823"

<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/yesCom.css?v=20240119" />

<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/layout/yesHeaderN4.css?v=202402
<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/layout/yesFooterN4.css?v=202403
<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/layout/wideLayout.css?v=2024022
<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/layout/simplebar.css?v=20240222

<!--[if IE]>
<link rel="stylesheet" type="text/css" href="https://image.yes24.com/sysimage/yesUI/hack/renewHack_ie.css?v=2021082
<![endif]-->

<link href="https://image.yes24.com/sysimage/renew/gnb/favicon_n.ico" type="image/x-icon" rel="shortcut icon" />

<script type="text/javascript">
var YesData = new function() {
    var _HTTP_URL = "https://www.yes24.com";
    var _BLOG_URL = "https://blog.yes24.com/";
    var _IMG_HOST_IMG = "https://image.yes24.com";
    var _IsWelcomePage = "False";
    var _MEM_AGE = 0;
    var _PID = "";
    var _ParenerInfomagion = {};
    var _NOW_HTTP_URL = "https://www.yes24.com";
    var _IsdevServer = "False";
    var _IsHttps = "True";

    this.HTTP_URL = function(){return _HTTP_URL;}
    this.NOW_HTTP_URL = function(){return _NOW_HTTP_URL;}
    this.BLOG_URL = function(){return _BLOG_URL;}
    this.IMG_HOST_IMG = function(){return _IMG_HOST_IMG;}
    this.IsHttps = function(){return _IsHttps;}
    this.IsWelcomePage = function(){return _IsWelcomePage;}
    this.MEM_AGE = function(){return _MEM_AGE;}
    this.PID = function(){return _PID;}
    this.ParenerInfomagion = function(){return _ParenerInfomagion;}
    this.IsdevServer = function(){return _IsdevServer;}

```

```

soup = BeautifulSoup(r.text, 'html.parser')
prd_detail = soup.find('div', attrs={'id': 'infoSet_specific'})
print(prd_detail)

```

```

<div class="gd_infoSet infoSet_noLine" id="infoSet_specific">
<div class="tm_infoSet">
<h4 class="tit_txt">품목정보</h4>
</div>
<div class="infoSetCont_wrap">
<div class="yesTb">
<table class="tb_nor tb_vertical" summary="품목정보 국내도서, 외국도서" width="100%">
<caption>품목정보</caption>
<colgroup>
<col width="170"/>
<col width="*" />
</colgroup>
<tbody class="b_size">
<tr>
<th class="txt" scope="row">발행일</th>
<td class="txt lastCol">2019년 06월 24일</td>
</tr>
<tr>

```

```

<th class="txt" scope="row">판형</th>
<td class="txt lastCol">


양장



```

prd_tr_list = prd_detail.find_all('tr')
print(prd_tr_list)

```



```

[<tr>
<th class="txt" scope="row">발행일</th>
<td class="txt lastCol">2019년 06월 24일</td>
</tr>, <tr>
<th class="txt" scope="row">판형</th>
<td class="txt lastCol">

양장


```

for tr in prd_tr_list:
    if tr.find('th').get_text() == '쪽수, 무게, 크기':
        page_td = tr.find('td').get_text()
        break

```



```
print(page_td)
```



```
330쪽 | 496g | 130*198*30mm
```



```
print(page_td.split()[0])
```



```
330쪽
```


https://colab.research.google.com/drive/1_MQNrOfV/mzMqA-oIJGVH28kc_US8DxS0#scrollTo=cfR_ZdpLts1g&printMode=true

5/8


```


```

```
def get_page_cnt(isbn):
    # Yes24 도서 검색 페이지 URL
    url = 'http://www.yes24.com/Product/Search?domain=BOOK&query={}'
    # URL에 ISBN을 넣어 HTML 가져옵니다.
    r = requests.get(url.format(isbn))
    soup = BeautifulSoup(r.text, 'html.parser') # HTML 파싱
    # 검색 결과에서 해당 도서를 선택합니다.
    prd_info = soup.find('a', attrs={'class': 'gd_name'})
    if prd_info == None:
        return ''
    # 도서 상세 페이지를 가져옵니다.
    url = 'http://www.yes24.com'+prd_info['href']
    r = requests.get(url)
    soup = BeautifulSoup(r.text, 'html.parser')
    # 상품 상세정보 div를 선택합니다.
    prd_detail = soup.find('div', attrs={'id': 'infoSet_specific'})
    # 테이블에 있는 tr 태그를 가져옵니다.
    prd_tr_list = prd_detail.find_all('tr')
    # 쪽수가 들어 있는 th를 찾아 td에 담긴 값을 반환합니다.
    for tr in prd_tr_list:
        if tr.find('th').get_text() == '쪽수, 무게, 크기':
            return tr.find('td').get_text().split()[0]
    return ''
```

```
get_page_cnt(9791190090018)
```

```
'330쪽'
```

```
top10_books = books.head(10)
```

```
def get_page_cnt2(row):
    isbn = row['isbn13']
    return get_page_cnt(isbn)
```

```
page_count = top10_books.apply(get_page_cnt2, axis=1)
print(page_count)
```

```
0    330쪽
1    300쪽
2    228쪽
3    340쪽
4    264쪽
5    396쪽
6    272쪽
7
8    244쪽
9    296쪽
dtype: object
```

```
page_count = top10_books.apply(lambda row: get_page_cnt(row['isbn13']), axis=1)
```

```
page_count.name = 'page_count'
print(page_count)
```

```
0    330쪽
1    300쪽
2    228쪽
3    340쪽
4    264쪽
5    396쪽
6    272쪽
7
8    244쪽
9    296쪽
Name: page_count, dtype: object
```

```
top10_with_page_count = pd.merge(top10_books, page_count,
                                  left_index=True, right_index=True)
top10_with_page_count
```

	no	ranking	bookname	authors	publisher	publication_year	isbn13	page
0	1	1	우리가 빛의 속도로 갈 수 없다면 :김초엽 소설	지은이: 김초엽	허블	2019	9791190090018	
1	2	2	달리구트 꿈 백화점. 이미예 장편소설	지은이: 이미예	팩토리나인	2020	9791165341909	
2	3	3	지구에서 한아뿐 :정세랑 장편소설	지은이: 정세랑	난다	2019	9791188862290	
3	4	4	시선으로 부터, :정세랑 장편소설	지은이: 정세랑	문학동네	2020	9788954672214	
4	5	5	아몬드 :손원평 장편소설	지은이: 손원평	창비	2017	9788936434267	
5	6	6	피프티 피플 :정세랑 장편소설	지은이: 정세랑	창비	2016	9788936434243	

Next steps: [Generate code with top10_with_page_count](#) [View recommended plots](#)

```
df1 = pd.DataFrame({'col1': ['a','b','c'], 'col2': [1,2,3]})
df1
```

	col1	col2
0	a	1
1	b	2
2	c	3

Next steps: [Generate code with df1](#) [View recommended plots](#)

```
df2 = pd.DataFrame({'col1': ['a','b','d'], 'col3': [10,20,30]})
df2
```



	col1	col3
0	a	10
1	b	20
2	d	30

Next steps: [Generate code with df2](#) [View recommended plots](#)



```
pd.merge(df1,df2, on='col1')
```

	col1	col2	col3
0	a	1	10
1	b	2	20



```
pd.merge(df1,df2,how='left', on='col1')
```

	col1	col2	col3	
0	a	1	10.0	
1	b	2	20.0	
2	c	3	NaN	



```
pd.merge(df1,df2,how='outer',on='col1')
```

	col1	col2	col3	
0	a	1.0	10.0	
1	b	2.0	20.0	
2	c	3.0	NaN	
3	d	NaN	30.0	

```
pd.merge(df1,df2,left_on='col1',right_on='col1')
```

	col1	col2	col3	
0	a	1	10	
1	b	2	20	

```
pd.merge(df1,df2,left_on='col2',right_index=True)
```

	col1_x	col2	col1_y	col3	
0	a	1	b	20	
1	b	2	d	30	