

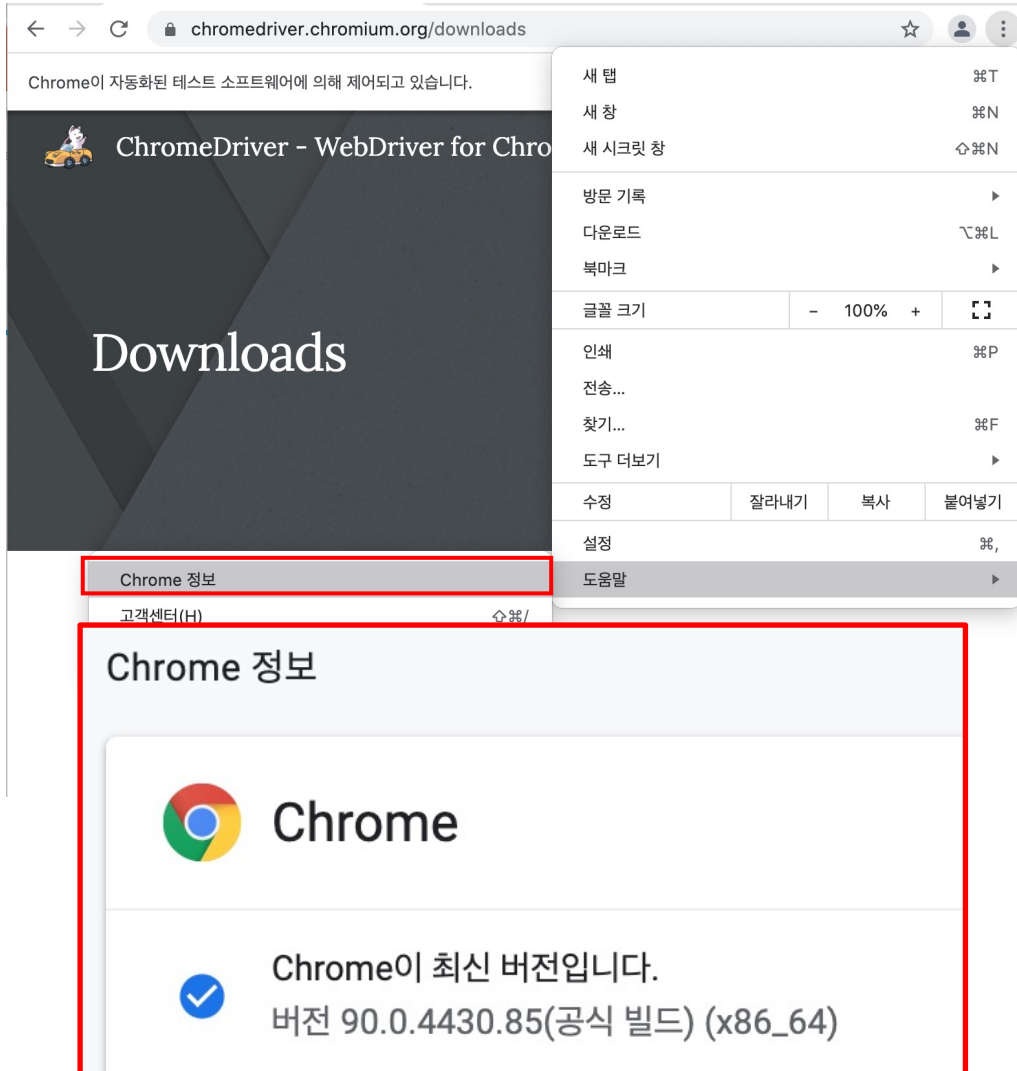
# 네이버 블로그 크롤링

네이버에서 “대전 맛집” 검색 후 블로그 글 크롤링

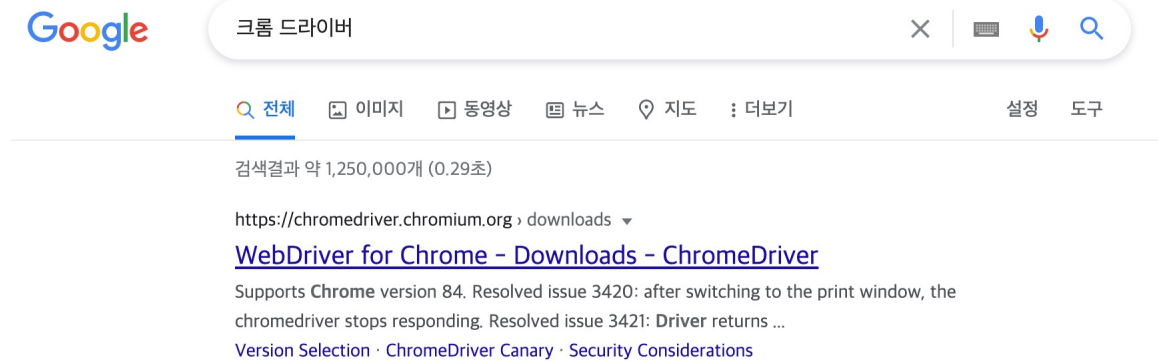
# 크롬 드라이버 다운로드

## chromedriver.exe 다운받기

### 1. 크롬 버전 확인 : 도움말 - 크롬 정보



### 2. 버전에 맞는 크롬 드라이버 다운로드



## Current Releases

- If you are using Chrome version 90, please download [ChromeDriver 91.0.4472.19](#)
- If you are using Chrome version 90, please download [ChromeDriver 90.0.4430.24](#)
- If you are using Chrome version 89, please download [ChromeDriver 89.0.4389.23](#)
- If you are using Chrome version 88, please download [ChromeDriver 88.0.4324.96](#)

원래는 웹 디자인에서 웹 페이지 상의 특정 요소를 선택하여 스타일을 적용하는 스타일 범위적용 기능을 수행하는 것  
우리는 웹 크롤링할 때 데이터 주소를 확인하는 용도로 사용합니다.

1. **태그명O** : 태그명이 O인 특정 태그를 선택합니다. O은 h, p, a 등등 모든 특정 태그를 포함합니다.

ex) `p { ~; ~; ~; }`

2. **.class** : 클래스 속성값이 특정 classname으로 지정된 경우에 해당 요소를 선택합니다.

ex) `.ft_accesstermd { ~; ~; ~; }`

3. **#id** : id 값이 특정 id으로 지정된 경우에 해당 요소를 선택합니다.

ex) `#ft_share_box { ~; ~; ~; }`

## 웹 페이지의 요소를 찾는 기본 공식

`driver.find_element_by_css_selector(#id or .class)`

```
driver.find_element_by_css_selector(.class or #id)
driver.find_element_by_link_text(텍스트)
driver.find_element_by_xpath(xpath)
driver.find_element_by_id(id)
driver.find_element_by_name(name)
```

# 네이버 블로그에서 “대전맛집”과 관련된 글들을 분석해본다면

N 대전맛집



통합 VIEW 이미지 지식iN 동영상 쇼핑 뉴스 어학사전 지도 책 ...

출처 • 전체 • 블로그 • 카페

정렬 • 관련도순 • 최신순

기간 • 전체 • 1시간 • 1일 • 1주 • 1개월 • 3개월 • 6개월 • 1년 • 직접입력 ▾

↻ 옵션 초기화

검색옵션 가이드 ?



Amy & BB | 6일 전

## 구수했던 대전 맛집 모음

이렇듯 대전 맛집은 건강과 맛 모두 잡을 수 있었던 곳이라 앞으로도 자주 찾을 계획입니다. 2. 대전 팡시온 불고기그릴버섯 오믈렛 15,000원...



29



Floating lamp | 2021.04.16.

## 강렬했던 # 대전 맛집 모음

강렬했던 대전 맛집 모음 얼마 전에는 친구들과 대전에 방문했다가 인기있는 대전 맛집을 찾았는데요! 마블링이 촘촘히 박혀있는 한우의 품질이 대단해서 먹는 내내...



27

## 크롤링 라이브러리 импорт

```
# 네이버에서 검색어 입력받아 검색 한 후 블로그 메뉴를 선택하고
# 오른쪽에 있는 검색옵션 버튼을 눌러서
# 정렬 방식과 기간을 입력하기

#Step 0. 필요한 모듈과 라이브러리를 로딩합니다.
import sys # 시스템
import os  # 시스템

import pandas as pd # 판다스 : 데이터분석 라이브러리
import numpy as np  # 넘파이 : 숫자, 행렬 데이터 라이브러리

from bs4 import BeautifulSoup # html 데이터를 전처리
from selenium import webdriver # 웹 브라우저 자동화
import time # 서버와 통신할 때 중간중간 시간 지연. 보통은 1초
from tqdm import tqdm_notebook # for문 돌릴 때 진행상황을 %게이지로 알려준다.
```

```
# 검색어, 검색 기간 정의
query_txt = input('1.크롤링할 키워드는 무엇입니까?: ')
# query_txt2 = input('2.제외할 첫번째 키워드를 입력하세요: ')
# query_txt3 = input('3.제외할 두번째 키워드를 입력하세요: ')

start_date = input('4.조회를 시작할 날짜를 입력하세요(예:20170101) :')
end_date = input('5.조회를 종료할 날짜를 입력하세요(예:20171231): ')

```

1.크롤링할 키워드는 무엇입니까?: 대전맛집  
4.조회를 시작할 날짜를 입력하세요(예:20170101) :20210101  
5.조회를 종료할 날짜를 입력하세요(예:20171231): 20210330



# 크롬 웹드라이버 실행

## 네이버 검색창에 대전맛집 검색 후 블로그 카테고리 선택

#Step 1. 크롬 웹브라우저 실행

```
path = "/Users/aiden/Desktop/chromedriver" # 윈도우는 "chromedriver.exe"
```

```
driver = webdriver.Chrome(path)
```

# 사이트 주소는 네이버

```
driver.get('http://www.naver.com')
```

```
time.sleep(2) # 2초간 정지
```

#Step 2. 네이버 검색창에 "검색어" 검색

```
element = driver.find_element_by_id("query")
element.send_keys(query_txt) # query_txt는 위에서 입력
element.submit()
time.sleep(1)
```

# 'VIEW' 클릭

```
driver.find_element_by_link_text("VIEW").click()
```

# '블로그' 클릭

```
driver.find_element_by_link_text("블로그").click()
time.sleep(1)
```

# '옵션' 클릭

```
driver.find_element_by_link_text("옵션").click()
time.sleep(1)
```

N 대전맛집

통합 VIEW 이미지 지식iN 동영상 쇼핑 뉴스 어학사전 지도 책 ...

출처 • 전체 • 블로그 • 카페

정렬 • 관련도순 • 최신순

기간 • 전체 • 1시간 • 1일 • 1주 • 1개월 • 3개월 • 6개월 • 1년 • 직접입력 ▾

⌂ 옵션 초기화

검색옵션 가이드 ?

Amy & BB 6일 전

구수했던 대전 맛집 모음

이렇듯 대전 맛집은 건강과 맛 모두 잡을 수 있었던 곳이라 앞으로도 자주 찾을 계획입니다. 2. 대전 팽시온 불고기그릴버섯 오믈렛 15,000원...



## 검색옵션 - 기간 '3개월'

```
# 검색옵션 확인
item_li = driver.find_elements_by_css_selector('.option .txt')

for i in range(0, len(item_li)):
    print(item_li[i].text)
```

전체  
블로그  
카페  
관련도순  
최신순  
전체  
1시간  
1일  
1주  
1개월  
3개월  
6개월  
1년  
직접입력옵션펼치기

```
# 검색기간 '3개월' 클릭
print(item_li[10].text)

item_li[10].click()
```

3개월



## 블로그 글 스크롤다운

스크롤다운 해서 블로그 글들을 크롤링 할 준비를 합니다.

```
# 스크롤 다운
# driver.execute_script("window.scrollTo(0, 500)")
# time.sleep(2)

# 스크롤을 밑으로 내려주는 함수
def scroll_down(driver):
    driver.execute_script("window.scrollTo(0, 99999999)")
    time.sleep(1)

# n: 스크롤할 횟수 설정
n = 5
i = 0
while i < 5: # 이 조건이 만족되는 동안 반복 실행
    scroll_down(driver) # 스크롤 다운
    i = i+1
```

## URL 수집 코드

```
# 블로그 글 url들 수집
url_list = []
title_list = []

# URL_raw 크롤링 시작
articles = ".api_txt_lines.total_tit"
article_raw = driver.find_elements_by_css_selector(articles)
article_raw[0]

<selenium.webdriver.remote.webelement.WebElement (session="0d0ce2e7444a3
b3e223f22314cfd2c59", element="89a65fae-7d52-4b08-ad73-6e127b9b7272")>
```

```
article_raw[0].get_attribute('href')

'https://blog.naver.com/jinmylove?Redirect=Log&logNo=222317287992'
```

```
# 크롤링한 url 정제 시작
for article in article_raw:
    url = article.get_attribute('href')
    url_list.append(url)

# 1초 쉬어 주고,
time.sleep(1)

# 제목 크롤링 시작
for article in article_raw:
    title = article.text
    title_list.append(title)

    print(title)

print("")
print('url갯수: ', len(url_list))
print('title갯수: ', len(title_list))
```

구수했던 대전 맛집 모음  
출중했던 대전 맛집 모음  
강렬했던 # 대전 맛집 모음  
꼭 가봐야할 대전 맛집 리스트  
군침 나온 대전 맛집 공유  
이제야 알아간 대전 맛집 목록  
식욕 폭발했던 대전 맛집 리스트  
[곰창맛집 & 대전맛집 - 군자대한곰창 대전엑스포점]  
인정했던 대전 맛집 리스트  
자꾸 생각나는 대전 맛집 모음  
정성이 깃든 대전맛집 공유  
임팩트 대단한 대전 맛집 리스트  
강렬했던 대전 맛집 소개  
단골이된 대전 맛집 모음  
대전 맛집 '현대식당', 대흥동 닭볶음탕 노포 맛집  
깊이를 느꼈던 대전 맛집 리스트  
초밥이 맛있는 대전맛집  
대전 맛집 애슐리퀸즈 집에서 즐기는 홈파티음식 방문포장  
대전맛집 대계요리맛집 크래버대계나라둔산점 가성비 짱  
대전맛집)서대전역맛집 소담회  
갈마동 족발 맛집, 레트로 분위기 물씬풍기는 대전맛집 대광죽  
(대전맛집) 대전 육사시미 맛집 '태평소국밥'  
대전맛집 대동오거리 신선칼국수 매장 후기 + 배달 후기  
대전 맛집 반찬이 푸짐한 월평동 태능돌삼겹  
명랑식당, 삼성동맛집, 대전맛집, 육개장맛집  
대전 맛집 @포트 키친(port kitchen) 신상 맛집 대만족♥  
선화동맛집, 대성집, since 1974 전통의 대전맛집  
대전 맛집 남도항아리 후기  
대전 맛집 | 광천식당 오징어 두루치기, 칼국수 ♥  
대전맛집/원조"바로그집"본점/아이스크림떡볶이/대전역우동국수  
현주민이 찾는 대전맛집 태평소국밥  
#대전맛집 탄방동 대만음식점<바오바빌러>  
대전가볼만한곳, 대전카페, 대전맛집, 대전데이트추천  
[대전맛집] 내돈내산/ 광천식당 두부두루치기&수육  
대전 맛집 - '서울촌놈' 추천 맛집 천복순대국밥  
대전 맛집 꼭 가야하는 식당 대선칼국수  
<대전맛집,치킨맛집> 대전 오류동 키로치킨 리뷰  
대전 맛집 예담추어정 추어탕 그대로 담았습니다. 유명 맛집...

## 수집된 url 저장

수집된 url은 엑셀 파일로 변환되어 저장됩니다.

```
df = pd.DataFrame({'url':url_list, 'title':title_list})
df
```

	url	title
0	https://blog.naver.com/jinmylove?Redirect=Log&...	구수했던 대전 맛집 모음
1	https://blog.naver.com/106904?Redirect=Log&log...	출중했던 대전 맛집 모음
2	https://blog.naver.com/jin3641?Redirect=Log&lo...	강렬했던 # 대전 맛집 모음
3	https://blog.naver.com/sooya1127?Redirect=Log&...	꼭 가봐야할 대전 맛집 리스트
4	https://blog.naver.com/coolj21104?Redirect=Log...	군침 나온 대전 맛집 공유
...	...	...
175	https://blog.naver.com/8mmm8?Redirect=Log&logN...	대전맛집 아메리칸 치즈버거 탄방동맛집추천 :: 고급스러운...
176	https://blog.naver.com/minacyworld?Redirect=Lo...	[대전_맛집] 대전숨은맛집 대전청국장맛집 양사싯골
177	https://blog.naver.com/legojoo?Redirect=Log&lo...	[대전맛집] #가수원동맛집 #완벽한소풍
178	https://blog.naver.com/i_gil?Redirect=Log&logN...	[대전 맛집] 탄방동 돈까스정식
179	https://blog.naver.com/shariel_s2?Redirect=Log...	대전맛집추천 쌀국수 잘알못.... 미스사이공에서 베트남쌀국수...

180 rows x 2 columns

```
# 저장하기
df.to_excel("blog_url.xlsx")
```

## 저장된 글들 url 로드

저장된 url 엑셀파일을 읽어서 url\_load변수에 저장해둡니다.  
이제 이 url 각각에 접속해서 글 내용을 크롤링할 것입니다.

```
# "url_list.csv" 불러오기
url_load = pd.read_excel("blog_url.xlsx") # 기본 모델

num_list = len(url_load)

print(num_list)
url_load
```

180

Unnamed: 0		url	title
0	0	https://blog.naver.com/jinmylove?Redirect=Log&...	구수했던 대전 맛집 모음
1	1	https://blog.naver.com/jin3641?Redirect=Log&lo...	강렬했던 # 대전 맛집 모음
2	2	https://blog.naver.com/sooya1127?Redirect=Log&...	꼭 가봐야할 대전 맛집 리스트
3	3	https://blog.naver.com/lovelyju311?Redirect=Lo...	유명할만한 대전 맛집
4	4	https://blog.naver.com/coolj21104?Redirect=Log...	군침 나온 대전 맛집 공유
...	...	...	...
175	175	https://blog.naver.com/6design?Redirect=Log&lo...	너무나도 알찼던 대전 맛집베스트10
176	176	https://blog.naver.com/twosevengirl?Redirect=L...	[대전 맛집/한남대 맛집] 김스돈까스 *
177	177	https://blog.naver.com/anjanjajaj?Redirect=Log...	대전맛집 성심당 테라스키친 배달 후기
178	178	https://blog.naver.com/ddoryddo?Redirect=Log&l...	대전맛집 가수원 찐 우동맛집 마코토

## 블로그 글 크롤링하는 코드

```
dict = {} # 전체 크롤링 데이터를 담을 그릇
```

```
# ★수집할 글 갯수 정하기
```

```
number = 30 # 30개만 크롤링해 보겠습니다.
```

```
for i in tqdm(range(0, number)):
```

```
    # 글 띄우기
```

```
    url = url_load['url'][i]
```

```
    driver = webdriver.Chrome("/Users/aiden/Desktop/chromedriver")
```

```
    driver.get(url) # 글 띄우기
```

저장된 url로 접속해서 크롬창을 띄울 겁니다.

한 글의 크롤링이 끝나면 다음 for문을 돌기 전에 크롬창을 닫아줍니다. -> driver.close()  
(안 그러면 총 100개의 크롬창이 떠있을 겁니다)

```
    # 크롤링
```

```
    try :
```

```
        # iframe 접근
```

```
        driver.switch_to.frame('mainFrame')
```

```
        target_info = {}
```

```
        # 제목 크롤링 시작
```

```
        overlays = ".se-module.se-module-text.se-title-text"
```

```
        tit = driver.find_element_by_css_selector(overlays)
```

```
        title = tit.text
```

글에서 크롤링할 것은,

1.제목 2.글쓴이 3.날짜 4.내용 입니다.

```
        # title
```

```
        # 글쓴이 크롤링 시작
```

```
        overlays = ".nick"
```

```
        nick = driver.find_element_by_css_selector(overlays)
```

```
        nickname = nick.text
```

```
        # nickname
```

```
        # 날짜 크롤링
```

```
        overlays = ".se_publishDate.pcol2"
```

```
        date = driver.find_element_by_css_selector(overlays)
```

```
        datetime = date.text
```

```
        # datetime
```

```
        # 내용 크롤링
```

```
        overlays = ".se-component.se-text.se-l-default"
```

```
        contents = driver.find_elements_by_css_selector(overlays)
```

```
        content_list = []
```

```
        for content in contents:
```

```
            content_list.append(content.text)
```

```
        content_str = ' '.join(content_list)
```

```
        # content_str
```



## target\_info와 dict 딕셔너리에 데이터 담기

각각 크롤링한 후에 미리 만들어둔 target\_info 딕셔너리에 저장합니다.  
그리고 전체 딕셔너리인 dict에 각각의 글을 차례대로 저장합니다.  
크롤링이 끝나면 그 크롬창은 종료합니다.

*# 글 하나는 target\_info라는 딕셔너리에 담기게 되고,*

```
target_info['title'] = title  
target_info['nickname'] = nickname  
target_info['datetime'] = datetime  
target_info['content'] = content_str
```

target\_info 딕셔너리에

title, nickname, datetime, content라는 key값으로 데이터들이 담깁니다.

*# 각각의 글은 dict라는 딕셔너리에 담기게 됩니다.*

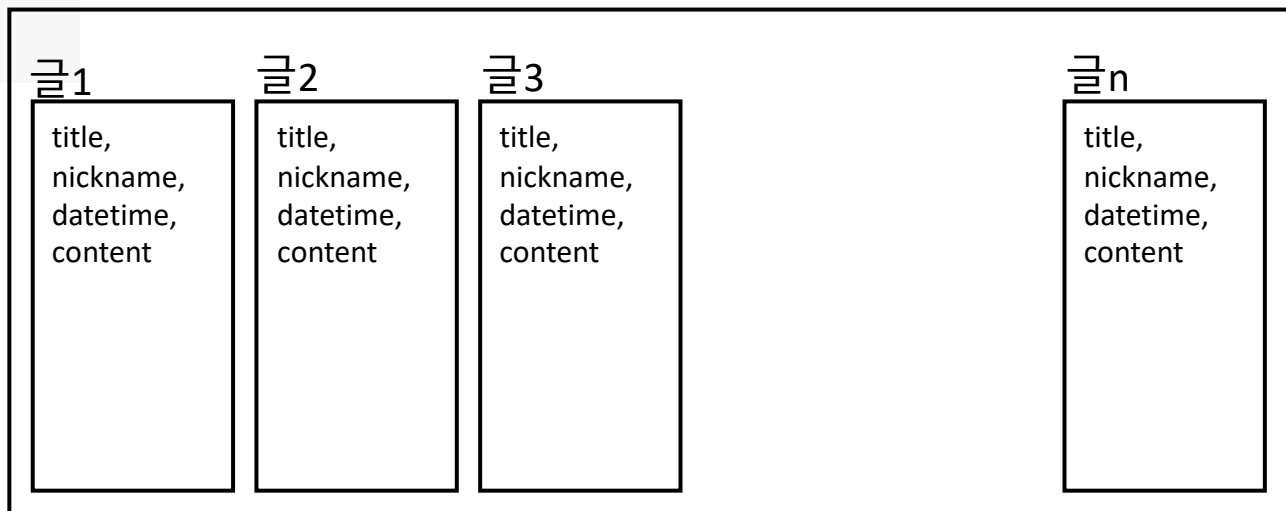
```
dict[i] = target_info  
time.sleep(1)
```

그리고 각각의 글인 target\_info는 dict에 차례대로 담깁니다.

```
print(i, title)
```

```
# 글 하나 크롤링 후 크롬 창 닫기  
driver.close()
```

dict





## 예외처리와 중간중간 저장하는 코드

항상 에러가 날 경우를 대비해야 합니다.

```
# 에러나면 현재 크롬창을 닫고 다음 글(i+1)로 이동합니다.
```

```
except:
```

```
    driver.close()
```

```
    time.sleep(1)
```

```
    continue
```

에러가 나면, 현재 크롬창을 끄고,  
다음 for문(글)으로 넘어가라

```
# 중간, 중간에 파일로 저장하기
```

```
if i == 30 or 50 or 80: 글 30번째, 50번째, 80번째 마다 엑셀로 저장하라
```

```
    # 판다스로 만들기
```

```
    import pandas as pd
```

```
    result_df = pd.DataFrame.from_dict(dict, 'index')
```

dict를 먼저 판다스로 변환한 후에

```
    # 저장하기
```

```
    result_df.to_excel("blog_content.xlsx")
```

중간 중간 엑셀로 저장해라

```
    time.sleep(3)
```

에러가 안 났다면, 최종적으로 엑셀로 저장

```
# 판다스로 만들기
```

```
import pandas as pd
```

```
result_df = pd.DataFrame.from_dict(dict, 'index')
```

```
# 저장하기
```

```
result_df.to_excel("blog_content.xlsx")
```

# 최종적으로 수집된 네이버 블로그 데이터

blog\_content.xlsx

	A	B	C	D	E
1		title	nickname	datetime	content
2	0	구수했던 대전 맛집 모음	에이미네(jinmylove)	2021. 4. 21. 10:10	지난 주말 가족들과 장동산림욕장으로 힐링을 하러 갔다가 미리 찾
3	1	강렬했던 # 대전 맛집 모음	날구	2021. 4. 16. 6:00	강렬했던 대전 맛집 모음 얼마 전에는 친구들과 대전에 방문했다가
4	2	꼭 가봐야할 대전 맛집 리스트	수야(sooya1127)	2021. 4. 16. 0:10	지난주 가족들과 외식을 하러 미리 알아봐둔 대전 맛집을 다녀왔어
5	3	유명할만한 대전 맛집	인절미	2020. 9. 29. 22:00	지난 주말에는 오랜만에 가족들과 대전 맛집에서 소고기를 먹고왔어
6	4	군침 나온 대전 맛집 공유	루루(coolj21104)	2021. 4. 23. 4:00	어제는 지인과 약속이 잡혀 전부터 가고자 했던 대전 맛집을 다녀왔
7	5	최고였던 대전 맛집	토로야(totoro_963)	2020. 10. 13. 12:50	얼마전 친구들과 모임이 생겨 미리 찾아둔 대전 맛집에 방문했답니
8	6	선택했던 대전 맛집 모음	보네르	2021. 4. 14. 19:41	며칠 전에 가족끼리 간만에 대전에 가서 외식을 하고 왔어요.퀄리티:
9	7	이제야 알아간 대전 맛집 목록	절대야연(lovegpop)	2021. 3. 22. 23:20	얼마 전 친구들과 친목도 다질겸 대전 맛집을 다녀왔어요. 내 취향
10	8	식욕 폭발했던 대전 맛집 리스트	뷰스타 루인(mjati)	2021. 2. 26. 19:30	얼마 전에는 대전에 사는 친구의 얼굴도 볼 겸 대전 맛집에서 선약
11	9	깔끔했던 대전 맛집 공유	마토단풍맘	2020. 11. 24. 23:57	얼마 전 지인들과 식사 자리가 있었는데요. 어떤 메뉴로 먹어볼까 고
12	10	입 소문난 대전 맛집 리스트!	코훗(5636in)	2021. 1. 15. 21:42	얼마 전 거래처 방문차 둔산동에 갔을때 미리 예약한 대전 맛집을 다
13	11	인정했던 대전 맛집 리스트	에이미	2021. 2. 22. 22:24	지난 주말 소고기 전문점이라 입소문 자자한 대전 맛집에서 모임이
14	12	자꾸 생각나는 대전 맛집 모음	달지우(idhsssh)	2021. 3. 27. 19:43	얼마 전 오랜만에 친구들과의 모임이 있었는데요.저번에 방송에서 노
15	14	99프로가 만족할 칼국수와 수육 대전맛집	제주왕자	2020. 7. 16. 23:22	잠시 카페에 앉아서 시간을 보내다가 저녁을 먹고 들어가기로 하면
16	15	든든했던 대전 맛집	바니커플	2020. 8. 20. 18:29	며칠전 친구들과 점심약속이 있어 대전 맛집을 다녀왔어요. 언제 먹
17	16	제대로 찾은 대전 맛집	망고탱이(flyfree6)	2020. 5. 23. 5:50	날씨가 더워지면서 슬슬 생각나는 음식이 바로 냉면이에요~지난주
18	17	정성이 깃든 대전맛집 공유	lonko	2021. 2. 27. 11:42	지난 달에는 친구들이랑 생면요리를 먹으려고 예전에 예약한 대전맛
19	18	내가 뽑은 대전 맛집	토로야(totoro_963)	2020. 9. 14. 11:13	얼마전에는 가족들과 제대로 된 곳에 가서 먹어보자고 얘기를 하다가
20	19	진정한 대전 맛집	퍼니	2020. 6. 20. 11:43	저번주 불일이 있어서 지인들과 터미널 근처 대전 맛집으로 유명세를
21	21	임팩트 대단한 대전 맛집 리스트	혜야	2021. 4. 15. 11:22	임팩트 묻어난 대전 맛집 리스트 저번주말에는 드라이브겸 꽃구경
22	22	대전 맛집 지금 가야해	미소짱 그녀	2020. 5. 26. 14:27	얼마전 친구들과 함께 모임이 있어서 장소를 물색하던 중대전 맛집
23	23	다시 생각나는 대전 맛집 리스트	은지새나	2020. 9. 19. 10:00	안녕하세요 은지새나입니다:) 지난 주말 친구들과 몸보신겸 한우를
24	24	강렬했던 대전 맛집 소개	코훗(5636in)	2021. 2. 6. 16:48	오랜만에 동료들과 거래처 용천동에 갔다가 풍미가 좋은 생면 칼국
25	25	단골이된 대전 맛집 모음	친절한용용씨	2021. 2. 24. 17:44	며칠전에 친구를 만나러 갔다왔어요.도착해서 시간을 보니 점심먹을
26	26	매료되었던 대전 맛집 리스트	루씨아(rostw92)	2020. 10. 28. 20:12	대전에 일이 있어 들렀다가 근처에 사는 지인과 대전 맛집에서 식사
27	27	초밥이 맛있는 대전맛집	칼군도마양(newosilent)	2021. 2. 22. 7:00	초밥이 맛있는 대전맛집 시장스시입 안 가득 내공이 느껴지는 잘 숙
28	28	대전맛집 대개요리맛집 크래버대게나라둔산점 가	유리호수(phrases11)	2021. 2. 22. 17:47	저번 주 남편 생일이라, 선물은 못 해주더라도평소 못 먹던 음식 먹
29	29	대전출장 #대전맛집 #광화문미진 #SRT #갤러리	A Myth	2020. 12. 12. 18:38	저 때만 해도 기차 안에 사람이 꽉 찼다. 몇 달을 기차도 타기 무서