

EDA/웹크롤링 과제 2회차

—

오늘 과제에 대해 드리는 말씀

오늘의 과제는

데이터 분석 주제를 선정하고, 데이터 수집, 데이터 전처리, 데이터 분석 결과 까지 진행하는 과제입니다 😊
가능하다면, 일정 시간 집중하여 하루 안에 끝내보시는 연습을 하시는 걸 권유드립니다!

그 동안 여러분이 수업을 들으셨던,

EDA/웹크롤링/파이썬 수업의 자료와 형식(+구글링)을 모두 참고하시어 진행하시면 됩니다!

그런데,

오늘의 과제를 주제로 한 파이썬 데이터 분석은 이미 공개된 영상과 코드, 블로그 등이 정말 많습니다.
추후에 복습하고 정리하실 때 참고해보시기를 바랍니다.

다만,

처음 과제를 진행하실 때는 최대한 기존 공개된 자료를 참고하지 마시고,
본인 스스로가 고민하고 진행하시며 작성하시는 것을 권장드립니다.

또한,

수강생분들간 의견 공유 또한 자제해주세요.

“안좋은 예) #질의응답 채널 또는 강사님께 DM으로, 이 홈페이지 크롤링 에러가 나는데 어떻게 해결해야 하나요?”
진행 중 발생하는 코드에러, 문제 상황 등에 대해 스스로 해결하시는 연습을 해주세요.

점수를 잘 받기 위한 것이 아닌, 나의 **성장**이 가장 중요하다는 점을 기억해주세요 🙏

EDA 과제 진행 안내

- 출제자: 민형기 강사님
- 시험범위: EDA 강의내용 전 범위
- 개인과제
- 반드시 jupyter notebook 파일로 작성해주세요
- **제일 중요!**
파일을 제출하실 때 반드시 아래와 같은 형식에 맞춰 제출해주세요.

[DS]eda2_honggildong.ipynb

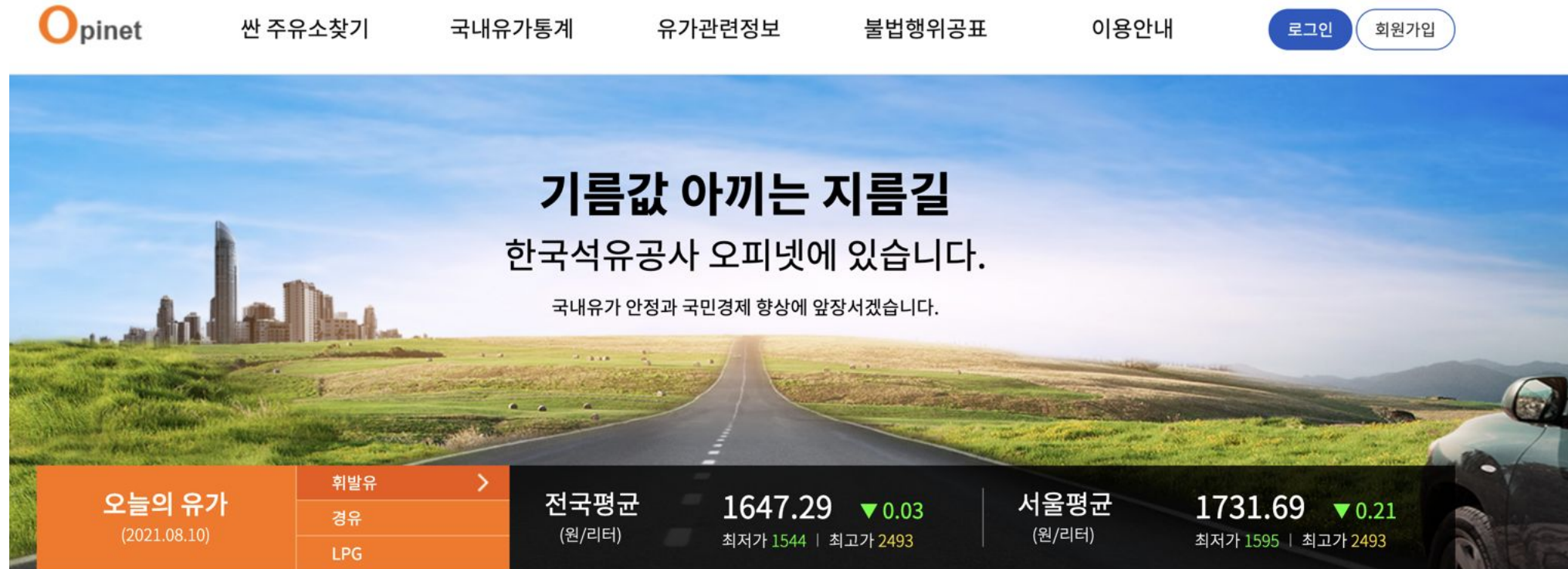
성함은 영문으로 성부터 이름 순서로 띄어쓰기 없이 소문자로 작성해주세요

꼼꼼하게 확인하시고, 꼭 지켜주세요!

- **또 중요!**
반드시 코드의 결과가 보이게끔 저장을 해주세요.
이는 추후에 저희가 채점을 진행해야하는데,
동작시킨 코드의 결과값을 저희가 볼 수 있어야 정확한 판단을 할 수 있기 때문입니다.

이제부터 과제 시작입니다!

저희 수업 중에는 셀프 주유소는 정말 저렴할까? 라는 주제의 프로젝트가 있었습니다.



The image shows the header and main banner of the Opinet website. The header includes the Opinet logo and navigation links: '싼 주유소찾기', '국내유가통계', '유가관련정보', '불법행위공표', and '이용안내'. There are also buttons for '로그인' (Login) and '회원가입' (Sign Up). The main banner features a scenic background of a road leading to a city skyline. The text on the banner reads: '기름값 아끼는 지름길' (Shortcuts to save on gas), '한국석유공사 오피넷에 있습니다.' (It is on the Korea Petroleum Corporation's Opinet), and '국내유가 안정과 국민경제 향상에 앞장서겠습니다.' (We will lead in stabilizing domestic oil prices and improving the national economy). Below the banner is a table showing today's oil prices.

오늘의 유가 (2021.08.10)	휘발유	전국평균 (원/리터)	1647.29 ▼0.03 최저가 1544 최고가 2493	서울평균 (원/리터)	1731.69 ▼0.21 최저가 1595 최고가 2493
	경유				
	LPG				


- <https://www.opinet.co.kr/user/main/mainView.do>
- 대한민국 주유 가격을 알아보는 사이트

싼 주유소 찾기에서 지역별 검색 탭으로 이동

싼 주유소찾기국내유가통계유가관련정보불법행위공표이용안내로그인회원가입

<ul style="list-style-type: none">· <u>지역별</u>· 경로별· 도로별· 면세유	<div>주유소/충전소</div> <ul style="list-style-type: none">· 주유소 ▾· 자동차충전소· 유가내려받기 <div>정유사/대리점/LPG</div> <ul style="list-style-type: none">· 정유사 ▾· 대리점· LPG용기충전소· LPG용기판매소· LPG집단공급	<ul style="list-style-type: none">· 국내유가동향· 국제유가 ▾· 전자상거래· 유류세· 단위환산표· 유가정보 API	<ul style="list-style-type: none">· 불법행위공표사항· 공표취소사항	<ul style="list-style-type: none">· 오피넷소개 ▾· 공지사항· 오피넷스타트업· 자주묻는질문· 사이트맵
--	---	--	---	--

그러면 이런 화면이 나타났습니다.



[싼 주유소찾기](#)
[국내유가통계](#)
[유가관련정보](#)
[불법행위공표](#)

지역별

경로별

도로별

면세유

주유소

충전소

지역

서울

성동구

읍/면/동

검색방법

지역

지도

상호

반경

형태

일반

셀프

불법

인증

전산

상표

전체

SK

GS

GS칼텍스

Hyundai

Gilbank

GS-OIL

인하

합동

PB

부가정보

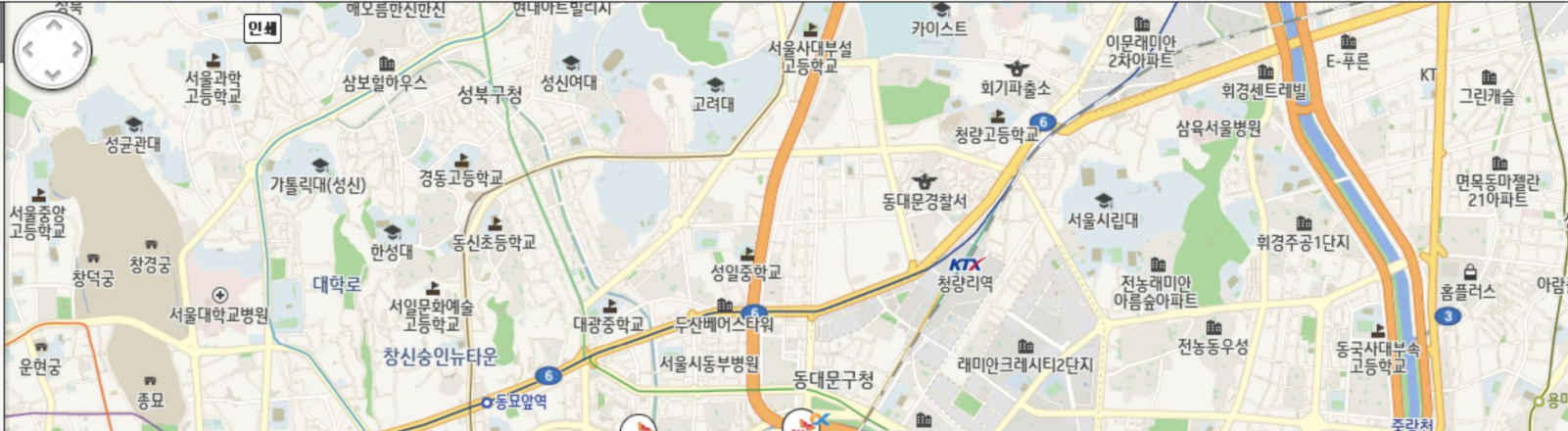
세차장

경정비

편의점

24시간

조회



아래와 같은 방법으로 데이터를 수집했었죠 ^^

주유소

충전소

지역

서울 ▼

성동구 ▼

읍/면/동 ▼

검색방법

☒ 지역

☐ 지도

☐ 상호

☐ 반경

형태

☒ 일반

☒ 셀프

☒ 불법

☒ 인증

☒ 전산

상표

전체

☒ SK 에너지

☒ GS칼텍스

☒ Hyundai Oilbank

☒ S-OIL

☒ 애플

☒ PB

•서울 => 각 구를 번갈아
선택

•엑셀파일을 다운로드 받음
•다운로드 받은 엑셀을
분석

그럼 이렇게 지역, 상호, 주소, 상표, 전화번호, 셀프여부, 고급휘발유, 휘발유, 경유, 실내등유 데이터가 담긴 엑셀 파일을 받아 왔습니다.

	A	B	C	D	E	F	G	H	I	J
1	지역_위치별(주유소)									
2										(단위 : 원/리터)
3	지역	상호	주소	상표	전화번호	셀프여부	고급휘발유	휘발유	경유	실내등유
4	서울특별시	수유동주유소	울 강북구 도봉로 395 (수유동)	GS칼텍스	02-902-9470	Y	-	1618	1418	-
5	서울특별시	㈜지에스이앤알 미아주유소	울 강북구 도봉로 122 (미아동)	GS칼텍스	02-987-5152	Y	1863	1618	1418	-
6	서울특별시	북서울고속주유소	울 강북구 삼양로 410 (수유동)	S-OIL	02-907-5182	Y	-	1622	1422	-
7	서울특별시	(주)석산에너지	울 강북구 삼양로 316 (수유동)	현대오일뱅크	02-980-1448	Y	-	1622	1422	900
8	서울특별시	KH에너지(주)직영 송정주유소	울 강북구 인수봉로 185 (수유동)	SK에너지	02-993-5189	N	-	1635	1429	-
9	서울특별시	SK에너지(주) 매일주유소	울 강북구 삼양로 217 (미아동)	SK에너지	02-982-8855	Y	-	1644	1444	-
10	서울특별시	덕릉로주유소	서울 강북구 덕릉로 158 (변동)	S-OIL	02-989-9806	N	-	1644	1435	-
11	서울특별시	현대오일뱅크(주)직영 새한셀프주유소	서울 강북구 도봉로 75	현대오일뱅크	02-985-4382	Y	1889	1648	1448	-
12	서울특별시	세원주유소	서울 강북구 한천로 918 (변동)	SK에너지	02-907-6890	Y	-	1648	1448	1100
13	서울특별시	오일뱅크(주)직영 변동셀프주유소	서울 강북구 한천로 963 (변동)	현대오일뱅크	02-992-3331	Y	1889	1648	1448	-
14	서울특별시	오일뱅크(주)직영 미아셀프주유소	울 강북구 도봉로 200 (미아동)	현대오일뱅크	02-945-7999	Y	1889	1648	1448	-
15	서울특별시	에덴주유소	서울 강북구 월계로 195 (변동)	SK에너지	02-987-8182	Y	-	1649	1437	-
16	서울특별시	대성석유(주)신광주유소	서울 강북구 삼양로 484	GS칼텍스	02-908-3320	N	-	1665	1455	1150
17										

그걸 이렇게,
데이터 프레임으로 만들어서 분석 했습니다 😊

	상호	주소	가격	셀프	상표	구
229	서남주유소	서울 중구 통일로 30	2511.0	N	SK에너지	중구
282	서계주유소	서울 용산구 청파로 367 (서계동)	2453.0	N	GS칼텍스	용산구
228	필동주유소	서울 중구 퇴계로 196 (필동2가)	2339.0	N	GS칼텍스	중구
227	SK에너지(주) 퇴계로주유소	서울 중구 퇴계로 228 (필동2가)	2305.0	N	SK에너지	중구
281	동자동주유소	서울 용산구 한강대로104길 6 (동자동)	2290.0	N	SK에너지	용산구
424	뉴서울(강남)	서울 강남구 언주로 716	2275.0	N	SK에너지	강남구
226	약수주유소	서울 중구 다산로 173	2218.0	N	GS칼텍스	중구
166	(주)대양씨앤씨 사직주유소	서울 종로구 사직로 65 (사직동)	2178.0	N	GS칼텍스	종로구
280	에너지비스	서울 용산구 한남대로 82 (한남동)	2178.0	N	SK에너지	용산구
279	한남지점	서울 용산구 한남대로21길 4 (한남동)	2178.0	N	SK에너지	용산구

여기서부터 집중!

그런데 이번에는, 부가정보 데이터도 한 번 가져와 보려고 합니다.

· 지역

· 검색방법 ☒ 지역 ☐ 지도 ☐ 상호 ☐ 반경

· 형태 ☒ 일반 ☒ 셀프 ☒ 불법 ☒ 인증 ☒ 전산

· 상표 ☒ SK ☒ GS칼텍스 ☒ Hyundai Oilbank ☒ S-OIL ☒ 알뜰 ☒ PB

· 부가정보 ☐ 세차장 ☐ 경정비 ☐ 편의점 ☐ 24시간

엑셀저장

·부가정보

- 세차장 유무
- 충전소 유무
- 경정비 유무
- 편의점 유무
- 24시간 운영 유무

단, 조건이 있습니다..

이번에는 모든 데이터를
엑셀저장이 아닌 코드를 통한 웹크롤링으로 데이터를
가져와주세요.

스타벅스 이디야 데이터 가져온 것 처럼^^

조건)

앞서 설명한 것 처럼,
저희는 서울시 구별 주유소 데이터 정보를 엑셀로 다운받았습니다.
이번에는 엑셀 파일 다운로드가 아닌, **크롤링 코드로 데이터를 수집해주세요.**
(조건은 똑같이 서울시 구별 주유소 데이터입니다)

문제 1)

수집한 데이터들을 **pandas** 데이터 프레임으로 정리해주세요.
부가 정보 데이터는 셀프 여부와 마찬가지로 **Y** 또는 **N** 으로 저장해주세요
최종적으로 데이터 프레임에 들어가야할 컬럼은 총 **14**개로 아래와 같습니다

주유소명, 주소, 브랜드, 휘발유 가격, 경유 가격,
셀프 여부, 세차장 여부, 충전소 여부, 경정비 여부, 편의점 여부, 24시간 운영 여부, 구, 위도, 경도

문제 2)

그리고 다시 한 번, 휘발유와 경유 가격이 셀프 주유소에서 정말 저렴한지 여러분의 분석 결과를 작성해주세요.
분석한 결과를 여러분의 **jupyter notebook**에 **markdown** 으로 설명해주시면 됩니다.

추가)

컬럼 이름은, 영문으로 저장해주세요
이 코드와 데이터는, 추후 **sql**과 태블로에서도 활용할 예정입니다. 꼭 잘 정리해두세요 ^^

EDA 과제 진행 안내

- 문제 1 60점
- 문제 2 40점
- 총점 100점
- 숙지사항
 - 문제1번은 정량 평가, 문제 3번은 정성 평가로 진행됨을 알려드립니다.
 - 작성하신 내용에 따라, 부분 점수 있습니다.
 - 제출하신 코드에 출력값이 반드시 있어야 합니다.
 - 결과 작성이 미흡하거나, 지나치게 간략할 시 감점입니다.
 - 예) 결론: 스타벅스와 이디야는 아무 관계가 없다
 - 과제 제출시, 압축하여 제출하시면 안됩니다. ipynb 파일 그대로 제출해주세요.
 - 기한이 지난 제출자는, 80점 만점으로 채점 진행합니다.