

EDA/웹크롤링 과제 1회차

—

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]eda1_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.
- 코드의 결과가 보이지 않거나, 요구하는 조건의 형태가 아니면 감점으로 간주합니다.

이제부터 과제 시작입니다!

이디야커피는 가끔 스타벅스 커피 매장이 위치하는 곳에 매장을 위치시키는 것이 아니냐는 의심을 받곤 합니다. 공식적인 인터뷰에서 이디야커피 회장은 이 사실을 부인했습니다.

여러분들이 이 사실을 확인해 주면 좋겠습니다.

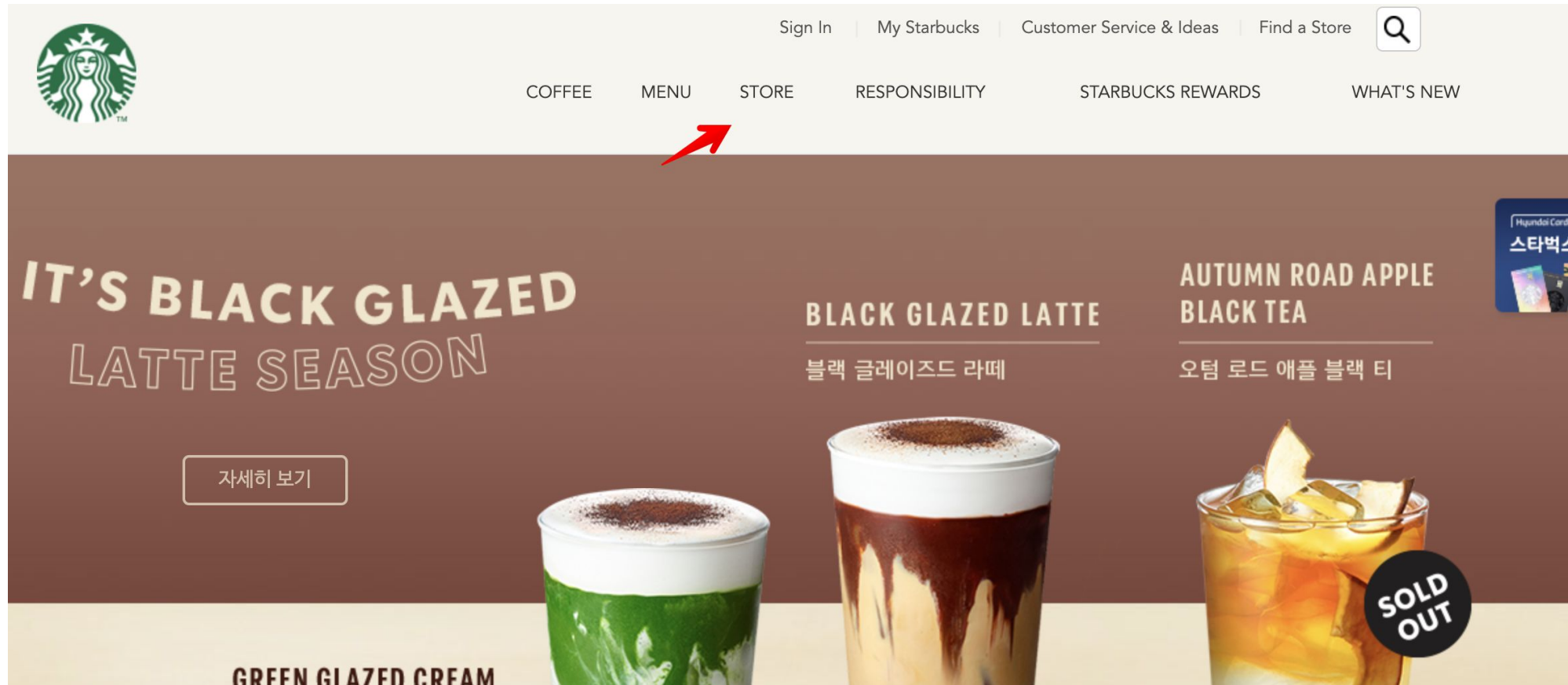
여러분들이 이제 막 크롤링을 배웠음을 알고 있습니다. 그러나 인터넷에는 많은 자료가 있고, 저는 여러분들이 충분히 이번 미션을 수행할 수 있는 능력이 있다는 것을 알고 있습니다.

다음 페이지부터 작은 단계들을 나눠 놓았습니다. 그것을 따라가며 이 미션을 완료해주길 바랍니다.

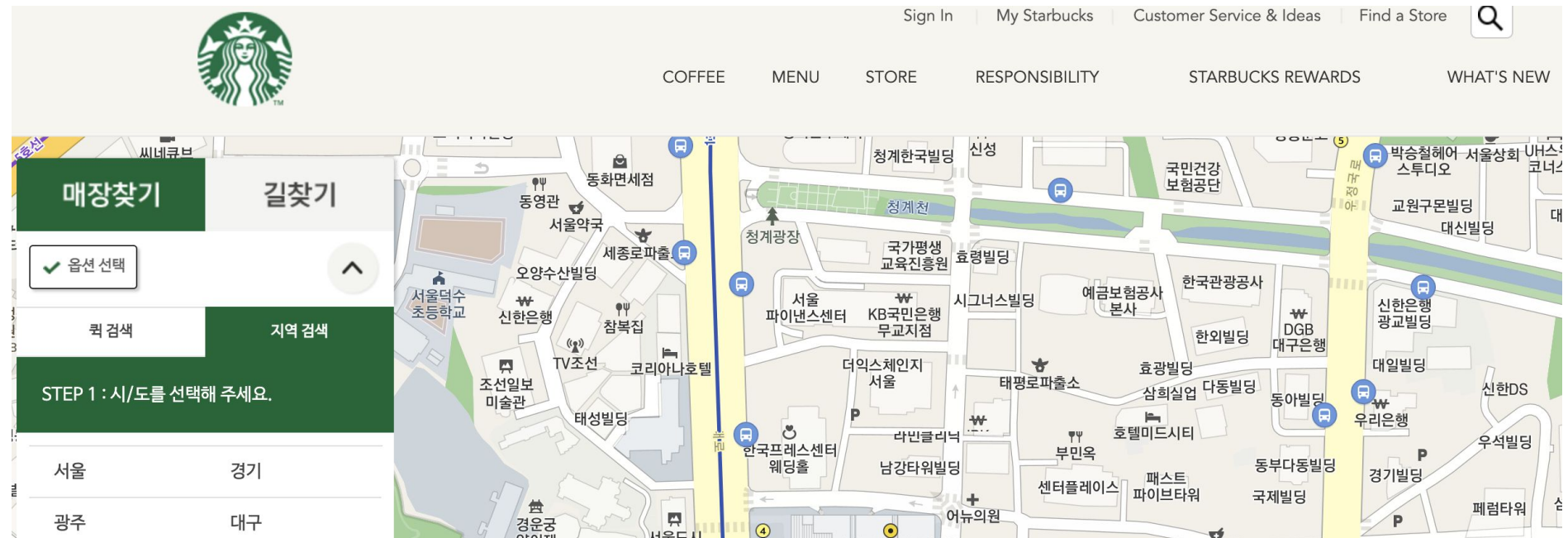
코드를 작성한 파일인 ipynb 파일을 여러분은 제출하면 됩니다. 그런데, 채점자는 모든 코드를 다 실행해 볼 수는 없습니다. 여러분들이 문제를 풀었는지 아닌지를 알기 위해 코드의 실행결과가 항상 out session에 잘 나타나도록 저장해 주세요.

이번 과제는 정답 코드가 없습니다.

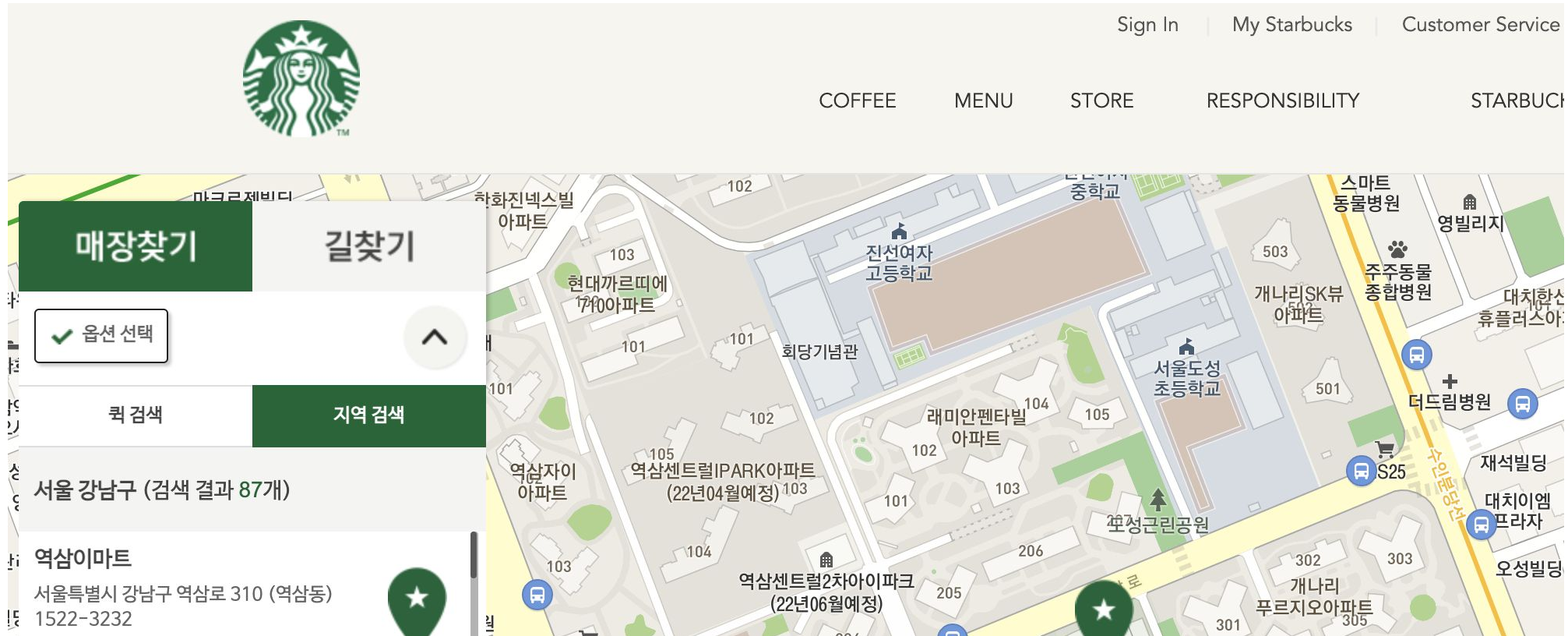
스타벅스 홈페이지에는 매장의 위치를 찾을 수 있는 메뉴가 있습니다.



여기서 매장찾기 지역검색으로 들어가면 아래의 그림의 화면을 만날 것입니다.



우리의 가정은 서울시로 한정하겠습니다. 서울시에서 강남구를 찾으면 이렇게 화면이 나타납니다. 아마 여러분은 이 과정을 서울시 구별로 동작하기 위해 셀레니움을 이용하고 그 다음 단계에서는 beautiful soup을 이용해야 할 지도 모릅니다.



문제1)

지금까지의 상황을 본다면 아마도 서울시의 구를 선택하는 화면까지는 셀레니움으로 진행을 해야할 지도 모릅니다. 그리고 난 후 구별 매장의 위치는 **beautiful soup**으로 가지고 올 수 있을 것 같습니다.

여러분들은 서울시의 스타벅스 매장의 이름과 주소, 구 이름을 **pandas data frame**으로 정리해 두세요.

이디야 커피 홈페이지에서도 매장찾기 메뉴가 있습니다.

브랜드 사이트 | 기업 사이트 | 창업안내 | 채용안내

로그인 | 매장찾기 | 고객의 소리 | 이디야 공식물



Korean



EDIYA COFFEE

메뉴

이디야멤버스

기프트카드

이디야커피랩

이디야컬처랩

이디야 소식

EDIYA COFFEE × PAYCOIN

페이코인 결제 시 전메뉴

국민지원금
사용안내

특 · 광역시는 광역단체 내,
도는 기초단체 (시 · 군) 내
이디야커피 가맹점에서 사용 가능합니다.

그러나 이디야 커피 홈페이지에서는 주소찾기가 스타벅스와는 조금 다릅니다.



HOME > 고객센터 > 매장검색



이렇게 주소입력란에 직접 입력을 해야합니다.

소비자중심경영(CCM)	매장검색	고객의 소리	FAQ	칭찬매장	제휴/제안
--------------	------	--------	-----	------	-------



아래 그림처럼 이디야 커피는 서울시 각 구의 이름을 주소에 입력해서 매장의 리스트를 찾아야 할 지도 모릅니다.

이렇게 서울시의 각 구 이름을 아래 그림의 위치에 각각 입력해서 나타난 결과를 수집해야 합니다. 아마도 구 이름을 입력하는 단계까지는 셀레니움을 사용하고 그 후는 **beautiful soup**을 사용해야 할 지도 모릅니다.



문제2)

여러분들은 서울시의 이디야커피 매장의 이름과 주소, 구 이름을 pandas data frame으로 정리해 두세요.

문제3)

문제 1과 2의 결과를 가지고 이제 이디야 커피는 스타벅스 커피 매장 근처에 있는지를 분석해보세요.
이 과정은 여러분의 몫입니다.

예를 들어 모든 커피 매장의 주소에서 위도/경도 정보를 가져와서 물리적인 거리를 측정하려고 노력하는 것도 있을 수 있겠죠. 혹은 도로명 주소를 가지고 단순히 유추하는 방법도, 또 혹은 folium으로 시각화한 후 육안으로 확인하는 방법도 있습니다.

방법에는 정답이 없습니다.

여러분들이 수행해야할 미션은 이디야커피는 과연 스타벅스 커피 주변에 위치해 있는가를 검증하는 것입니다.

같은 데이터를 놓고 판단하는 사람에 따라 그 분석 결과는 다를 수 있습니다. 그것도 상관없습니다.

여러분의 분석결과(예를 들어 이디야는 스타벅스 매장 근처에 위치한 것이 전략적이라고 볼 수 없다)는 여러분만의 논리적 근거가 있으면 됩니다.

그래서 그렇게 분석한 결과를 여러분의 jupyter notebook에 markdown으로 설명해 두세요.

EDA 과제 진행 안내

- 문제 1 30점
- 문제 2 30점
- 문제 3 40점
- 총점 100점
- 숙지사항
 - 문제1, 2번은 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.

정리가 다 되셨으면,
과제 제출 양식에 맞춰 제출해주세요^^