안녕하세요 패스트캠퍼스입니다.

파이널 프로젝트를 수행해주세요.

**※ 파이널 프로젝트 주제 : 스타벅스 이벤트 관련 고객 설문 데이터 EDA**

**※ 프로젝트 수행방법**

아래 드라이브 링크에서 과제의 내용을 확인합니다.

<https://drive.google.com/drive/folders/11lxbNC9KoG1r3ncIlewG7yeoQ4sWuPqF?usp=sharing>

해당 데이터는 pandas, seaborn을 중점으로 사용하여, kaggle survey와 같은 분석을 연습하는 최종 프로젝트입니다.

프로젝트에서 수행해야 할 사항은 다음과 같습니다.

1. 주어진 3개의 데이터 (profile, transcript, portfolio)를 활용하여 설문에 참여한 고객들이 이벤트에 어떤 반응을 했는지 조사해보세요.

2. profile 테이블에 있는 각 column에 대해서 데이터의 통계량을 측정하는 그래프를 그려보고, 그래프에 대한 해석을 적어주세요.

3. Transcript 테이블에 있는 event, time에 대한 통계량을 측정하는 그래프를 그려보고, 그래프에 대한 해석을 적어주세요.

4. 결측치가 존재하는 데이터에 대해선, 결측치를 처리해주세요.

**※ 프로젝트 수행방법 (상세)**

EDA PJT를 수행할 때는, 분석 목적을 명확하게 하고 목적을 달성하기 위한 여러가지 지표를 만드는 것을 목표로 합니다. 지표를 만들기 위해서는 여러가지 통계량을 계산하고, 시각화하는 것이 중요합니다.

앞에 강의에서 배웠던 라이브러리들을 이용하여 문제에 대한 근거가 되는 수치들을 찾으세요. 이 수치값들은 데이터에 기반한 논지를 펼칠 때 도움이 됩니다.

아래 질문들은 해당 데이터를 통해서 생각해볼 수 있는 분석 목표들입니다. 각 질문들에 대한 적절한 논지 전개를 위해 EDA를 수행해보세요.

—

Q. 이벤트에 참여한 고객들의 성별 비율은 어떻게 되나요? o

Q. 참여한 고객들의 성별에 따른 수입의 차이가 있나요? o

Q. 참여한 고객들의 평균 소득은 얼마이며, 가장 소득이 높은 나이 구간은 언제인가요? o

Q. 참여한 고객들이 가장 많이 가입한 기간은 언제인가요? 어떤 계절에 가입이 많은가요? o

Q. 이벤트에 실제로 참여하여 결제한 고객의 비율은 얼마나 되나요?o

Q. 이벤트 소식을 받고, 실제로 결제까지 이루어진 고객들은 평균적으로 어느정도의 시간이 소요가 되었나요? o

Q. 위에서 찾은 특징을 기반하여, 어떤 고객에게 해당 이벤트를 강력 제안하면 좋을까요?

—

pandas의 indexing 기법들과 pivot\_table, corr 함수를 이용하여 적절한 통계량을 찾아보세요.

pyplot과 seaborn을 이용하여 데이터의 분포와 흐름을 찾아보세요.

**\*코랩을 처음 시작할때 환경설정 부분에 셀출력 생략이 기본으로 체크되어있는데,  
 이 부분을 해제하고 저장해야 출력결과가 저장됩니다! 꼭 해체하고 저장해주세요!**

**※ 프로젝트 제출방법**

1. 프로젝트 결과물은 **아래 링크**를 통해 제출해주세요.

<https://docs.google.com/forms/d/e/1FAIpQLSdH7M2qotOluQZ3vF956KrMdqpOMs3MEijbhHfiQg1HE-Bf3w/viewform?usp=sf_link>

2. 성함과 메일주소 입력 시 오탈자가 발생하지 않도록 주의해 주세요.

오탈자 발생으로 수강생 정보확인이 어려워 프로젝트 리뷰가 불가할 수 있습니다.

**※ 주의사항**

1. 제출기한 : 학습 종료일 23:59:59 까지

2. 제출 시 반드시 ZIP 파일로 제출해주세요. 파일 오류로 평가가 불가능할 수 있습니다.

3. 제출기한 이후에 제출한 과제는 피드백이 제공되지 않으니 제출 기한을 꼭 지켜주시길 바랍니다.

프로젝트 관련하여 문의사항이 있다면, 슬랙으로 자유롭게 남겨주세요 :D