**< 1차 아이데이션 작성 - 오프라인 1조>**

1. **프로젝트 주제**

패스트캠퍼스 마케팅을 위한 데이터 분석(EDA)

1. **"어떻게" 새로운 인사이트를 도출할 계획인가요?**

01 콘텐츠 기획 분석 : 사람들은 어떤 콘텐츠를 많이 구매할까? 마케팅 효과가 있는지, 특정 그룹의 상품이 효과가 있는지 등을 확인

02 시기별 분석 : 어떤 년도, 월에 강의 콘텐츠 구매가 높을까? 어떤 시기에 구매율이 높고 왜 높은지 생각(코로나, 재택근무, 마케팅 증가 등)

03 금액별 분석 : 판매 되는 가격과 실제 결제 가격의 차이가 얼마나 될까? 쿠폰할인, PG사 포인트할인 등을 사용해 얼마나 할인을 받아 구매가 이루어지는지, 쿠폰을 안쓰는 사람들은 얼마나 있는지

04 강의별 분석 : 많이 구매된 강의는 어떤 특징을 가질까? 판매 시점부터 얼마나 지나서 구매가 많이 이루어질지 생각, 환불율이 높은 강의는 어떤 강의인지 분석, 금액이 높은강의가 잘 팔릴지, 낮은 강의가 잘 팔릴지

1. **프로젝트 진행 계획은 어떻게 되나요?**

저희는 state가 completed, type이 payment인 사람들을 위주로 분석할

예정입니다. 패스트캠퍼스 강의를 듣는 사람들이 주로 어떤 콘텐츠를

구매할지, 어떤 시기에 구매가 많아지는지, 많이 구매된 강의는 어떤 특징을

가지는지 pandas, matplotlib, seaborn 등 다양한 시각화 라이브러리를

활용해 분석해볼 예정입니다.

각자 분석해보고 싶은 파트를 중심으로 분석해보고 분석한 데이터를

바탕으로 다른데이터도 분석해보려 파트를 나눴습니다.

* 최윤석: 금액, 시기
* 김주호: 강의, 기획
* 안희수: 금액, 기획
* 허재호: 기획, 강의,(시기)
* 송희연: 강의, 시기
* 조세이: 시기, 강의

1. **진행하면서 어려웠던 점은 무엇인가요?**

허재호 : 데이터의 column별 정보가 없고 데이터가 정리가 안되있어서 전처리를 어떻게 해야하는지 분석을 어떻게 해야하는지 어려웠습니다. 시각화 할때의 그래프 사용이나 pandas등 여러가지 코드를 작성하는데 어려움이 있습니다

안희수 : 데이터의 결측치가 많고 일관성이 적어서 전처리가 굉장히 힘듦.현재 course\_title,category\_title, format의 결측치는 해당 행을 drop 했고, 'coupon\_title'과 'coupon\_discount\_amount'는 각각 해당 없음과 0.0으로 채우고, subcategory\_title의 결측치를 course\_title에 맞게 채우는 중이다.

김주호 : 데이터를 분류하고 싶은데 해당 데이터의 컬럼이 결측치이거나 비어있는 경우가 있어서 분류 기준을 선정하기 힘들었음. / 데이터 타입이 datetime64인 데이터에서 날짜를 제외하고 시간만 가져오려고 하는데 구글링해도 방법이 잘 나오지 않음. / 시각화할 때 그래프를 다양하게 사용하고 싶은데 아직 시각화 라이브러리에 대한 이해가 부족해서 단순한 그래프만 출력하고 있음.

송희연 : 일단 제가 파이썬에 익숙하지 않아서 어렵고, 결제 칼럼에서 각 value가 무엇을 의미하는지 정확한 의미 파악이 어려움. 프로모션이 어떤시기에 어떻게 이루어졌는지 알면 프로모션의 효과성 여부를 파악하기 좋을 것 입니다.

조세이 : 데이터가 분석을 위한 데이터가 아니기 때문에 기존의 캐글에서 본 데이터들에 비해서 분석에 적합하다고 보기 어려웠다. 데이터의 결측치도 많고, 보안상 의미를 알 수 없는 부분이 많았다. 하지만 실제 기업에서도 분석을 위한 데이터가 정제되어서 존재하지 않기 때문에 그런 점에서는 대비하기 괜찮은 데이터같다.

최윤석: 도출해보고 싶은 아이디어를 코드로 실현하는 부분에서 실질적으로 가능한지, 맞게 하고 있는 것인지 아닌지에 대해서 확신이 들지 않는 부분이 있었고, 데이터 분석 후 dataframe을 활용하여 표로 나타낸 코드를 시각화를 통해서 출력해보고 싶은데, 이 부분이 쉽지 않아서 어렵다고 느끼고 있습니다.

**Feedback**

고민을 많이 하셨고, 많이 진행된 것 같습니다.

올바른 방향으로 프로젝트를 진행하고 있는것 같고 세부적인 계획 또한 잘 진행되고 있는 것 같습니다.

지금 전처리를 진행중이신데, 잘 진행되고 있으실 것 같지만 겹치는 데이터 예를들면 금액, 기획, 시기 등 서로 분석하고 싶은 데이터를 전처리 중인것으로 판단됩니다.

이에대해 데이터를 전처리함에 있어 일관성을 두고 분석하게 된다면 추후에 시각화에 있어서 좀 더 일관적인 결과를 얻어 볼 수 있고 시간도 절약될 것으로 생각됩니다 :)

지금 하시는 기획처럼 계속 구체적으로 진행해주시면 좋을 것 같아요~