

「BDA 8기 학회원 최종 과제」 분석 보고서

과 제 명	학회원의 만족도에 영향을 끼치는 개인의 특성	성 명	송기웅
활용데이터	학회원 만족도 조사 데이터, 상점 데이터		

1. 주제 선정 및 분석 배경

- 주제 선정 및 분석 배경 등

어떤 학회원들이 BDA과정을 만족해하고 끝까지 수료를 하거나 다시 참여를 하는 지를 조사하기로 했다. 이를 통해, 끝 수료를 안 하는 학회원들이 느끼는 BDA의 문제점을 추가로 조사해볼 수 있고 다시 참여하는 학회원들의 특성을 확인해 볼 수 있어, 가진 문제점을 재고하고 해결방안을 도출할 수 있다.

2. 분석 결과 상세 내용

- 분석데이터, 전처리과정, 분석 툴, 알고리즘 및 방법론*

1. 개인의 특성 정의 및 전처리

개인의 특성과 만족도에 해당하는 변수 구분

1. 개인의 특성:

[하루공부시간, 직업, 최종학력, 교육희망직무, 희망직무, bda하는이유, 유입경로, 학점, 스터디경험, 참여경험, 프로젝트경험, 직무경험, 뉴스레터구독여부, 팔로우여부, 조별활동참여희망]

2. 만족도 특성:

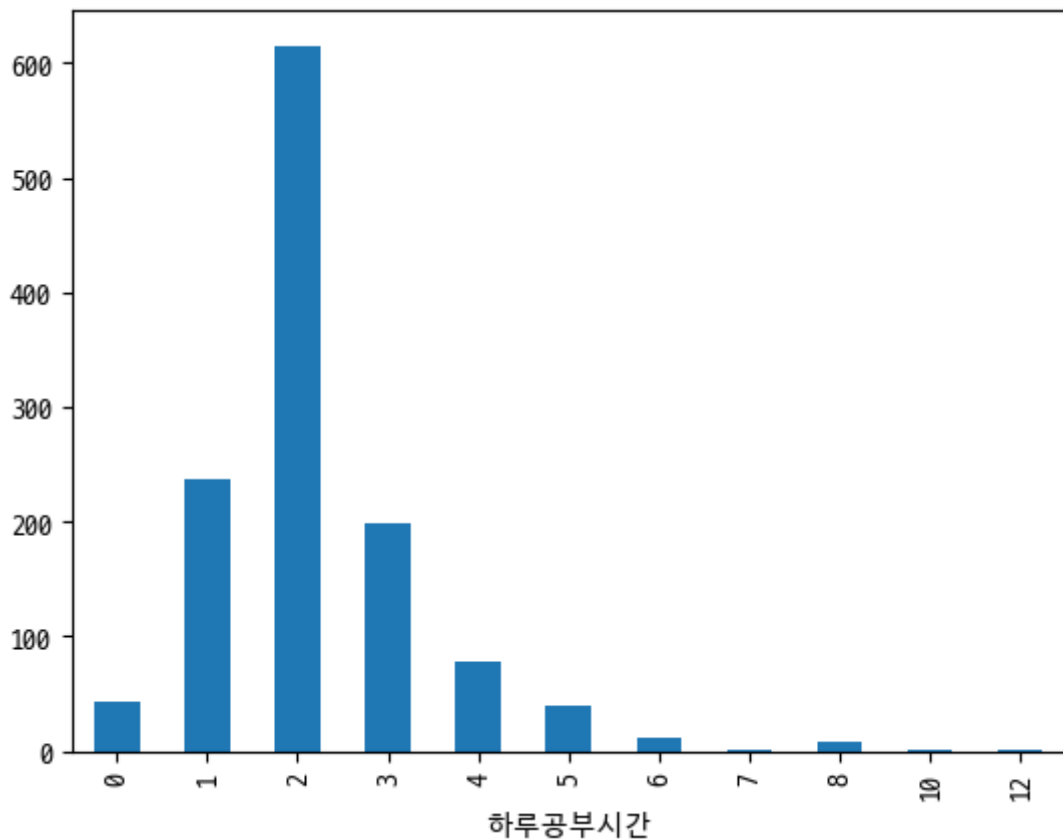
[상벌점, 7기 참여여부, 이전 참여여부]

우선적으로, 개인의 특성과 만족도를 정의할 필요가 있다. 개인의 특성은 데이터에서 나온 모든 변수를 특성으로 확인하기로 했다.

1-1. 하루공부시간

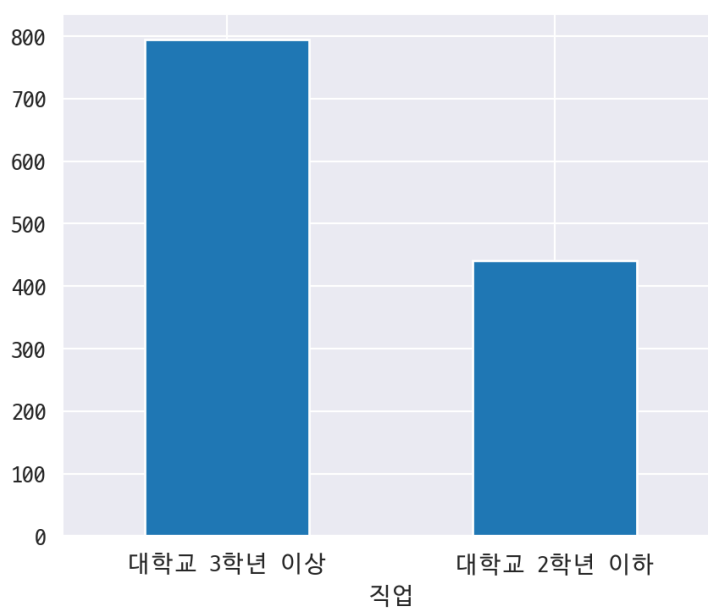
하루에 공부하는 시간을 개인이 데이터를 공부하려는 열정의 정도라고 생각하고 데이터에서 가장 우선적으로 나온 숫자로 설정했다. 변수마다 그 정도의 차이가 있긴

했으나 대부분 2시간 정도의 공부를 하는 것으로 나왔다.



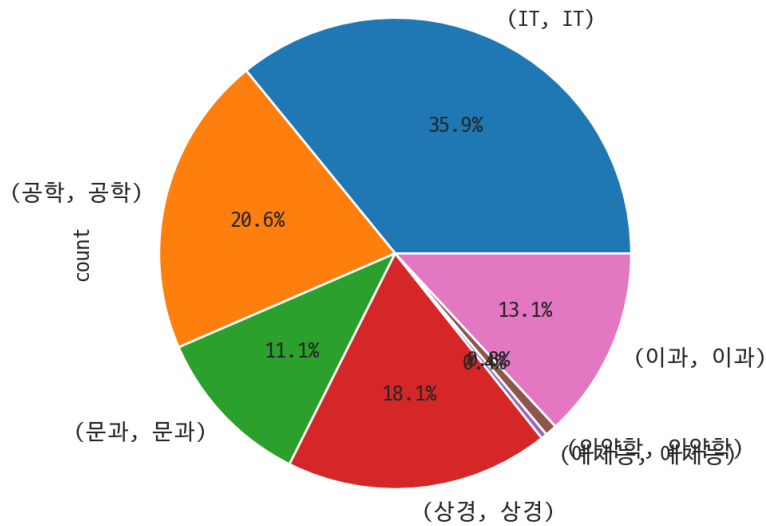
1-2. 직업

직업보다는 여기서 알 수 있는 나이 정보를 확인하려고 했다. 대학교 2학년 이하와 대학교 3학년 이상으로 구별해 나이가 많고 적음을 확인하려고 했다.



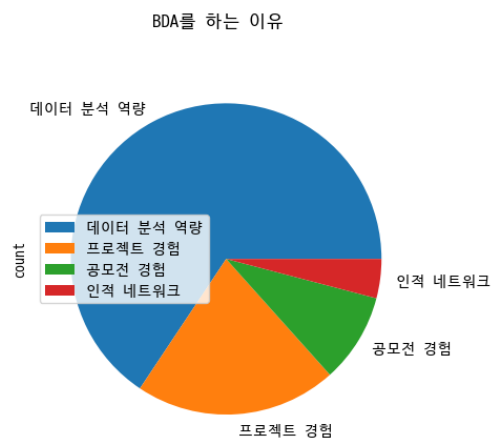
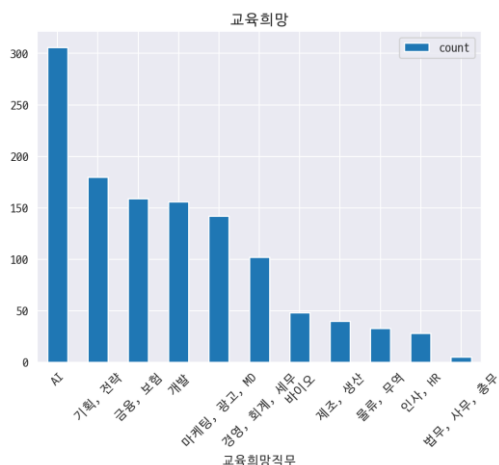
1-3. 최종학력

최종으로 가진 학과를 보면서 데이터나 코딩에 대해서 얼마나 밀접한 학과인지에 따라 IT, 공학, 이과, 상경, 문과로 구별했다.



1-4. 교육희망직무, BDA 하는 이유

이 둘이 그들이 BDA를 하는 직접적인 이유라고 생각했다. 만약, 만족하지 못한 학생들은 BDA 하는 이유나 교육희망직무에 대해서 제대로 된 공부가 되지 않아서 이탈할 확률이 높다. AI를 공부하려는 인원이 제일 많았고 데이터 분석 역량이 BDA를 하는 가장 큰 원인이 되었다.



1-5. 희망직무, 유입경로

이 2가지의 경우에는 파편화 되어 있었다. 그래서 하나씩 값들을 조정해줄 필요가 있

었다. 그래서 딕셔너리를 만들어 각 값들을 rename해줘 변수의 일정함을 갖도록 했다.

희망직무		유입경로	
DA	461	교내 플랫폼	841
DS	237	소셜 미디어	186
AI	85	기존 BDA 관련	149
마케팅	83	지인 추천	44
개발	66	검색/인터넷	12
DE	65	기타	3
연구	60		
기획	55		
기타	35		
미정	22		

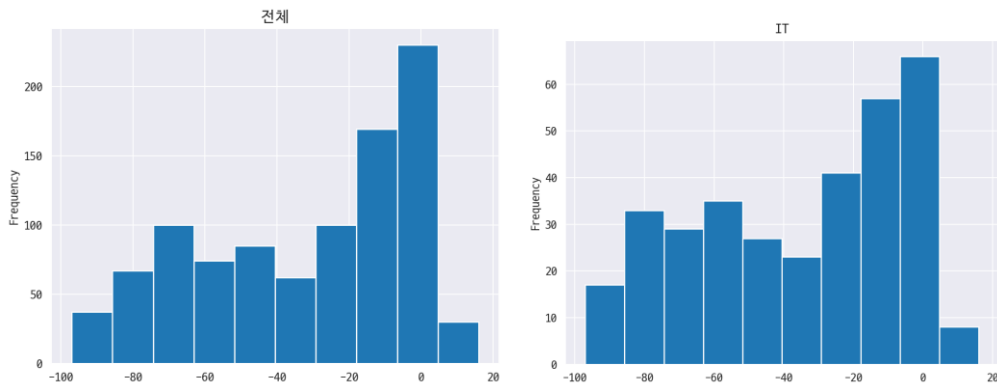
- 1-6. 프로젝트 경험, 직무경험, 스터디 경험, 팔로우 여부, 뉴스레터 구독 여부
있음과 없음, 예와 아니오 등으로 구분 된 데이터들은 1과 0으로만 바뀌었다.
직무경험의 경우 직무경험을 한 데이터가 너무 적어 제외하기로 했다.

참여경험		직무경험	
없음	993	없음	1197
있음	243	있음	38
		Name: count, dtype: int64	

2. 만족도 특성 전처리

2-1. 상별점

상별점 데이터에 관해서 -20점 이상의 학회원은 수료까지 하니 만족 카테고리에 넣기로 했다. 이탈하지 않았다는 것을 만족으로 넣기로 했다. 만족 컬럼을 만들어 만족이면 1 만족하지 않다는 0을 넣었다



2-2. 재참여

7기 참여, 이전 참여 여부와 함께 BDA를 치인에게 추천 정도가 0.7 이상인 경우에는 만족이라고 보고 만족 컬럼을 만들어 만족이면 1 만족하지 않다는 0을 넣었다.

3. 만족에 유의한 개인 특성 확인

3-1. 카이 제곱 검정

상별점 데이터, 재참여 데이터, 만족도 데이터 3개의 데이터가 있었으므로 각 데이터의 개인특성과 만족도에 어떤 영향을 끼치는지 확인하기로 했다. 영향과 원인을 대강 확인하기 위해 넣었으므로 유의확률은 0.1로 폭넓게 잡았다.

- 유의확률을 0.05로 봤을 때

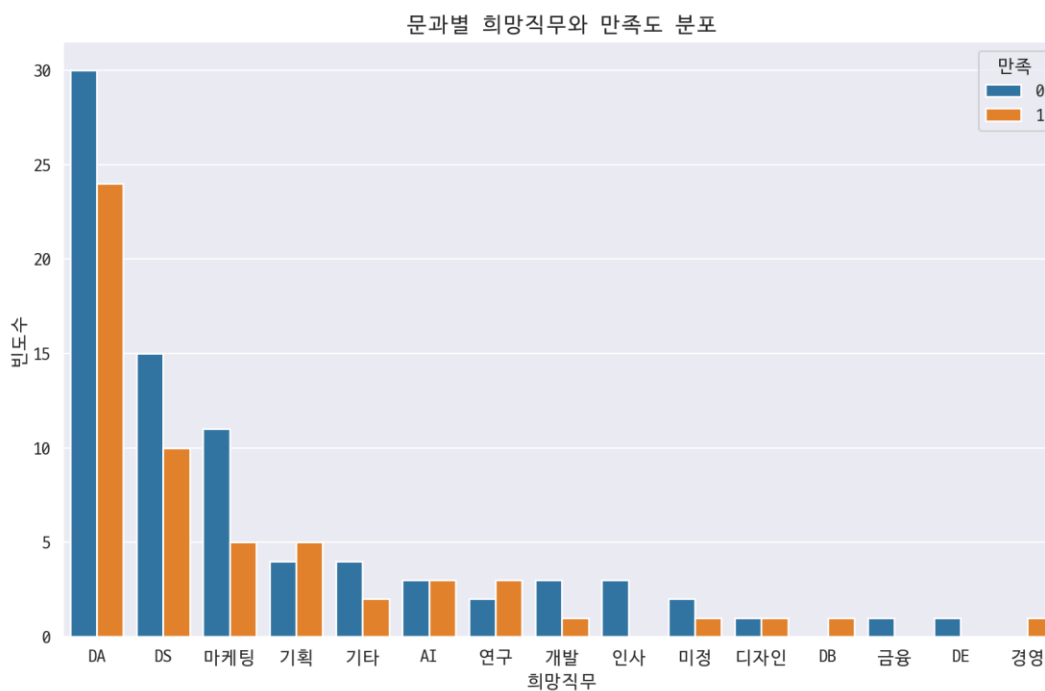
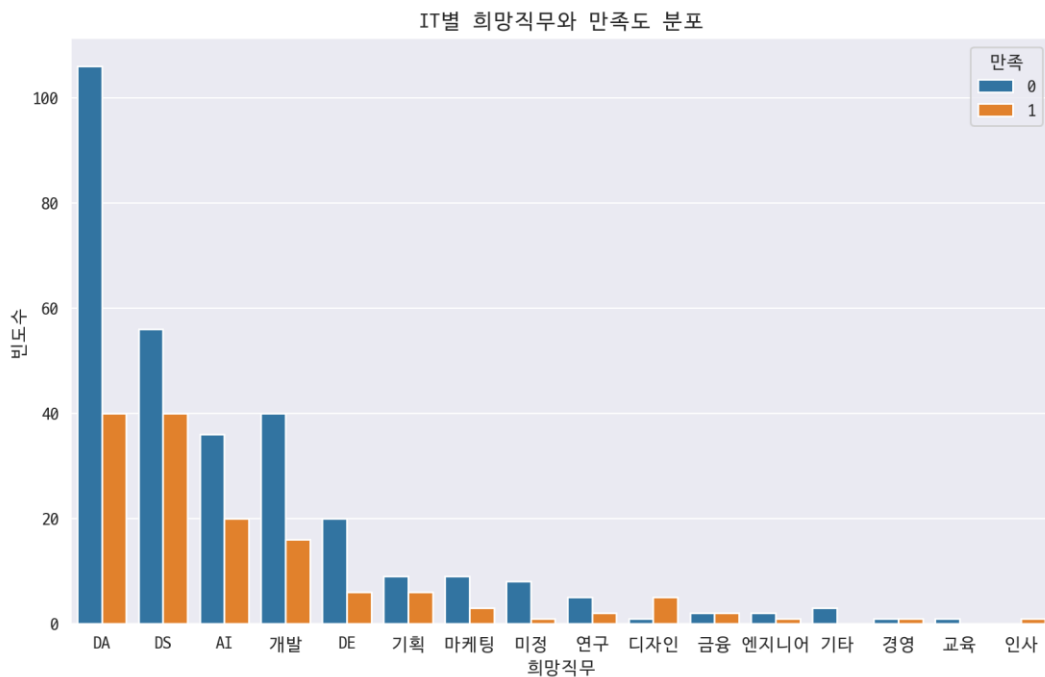
1. df_point : [희망직무]
2. df_re : [스터디 경험, 직무 경험]
3. df_satisfy : [졸업후 진로, 유입경로]

- 유의확률을 0.1로 봤을 때

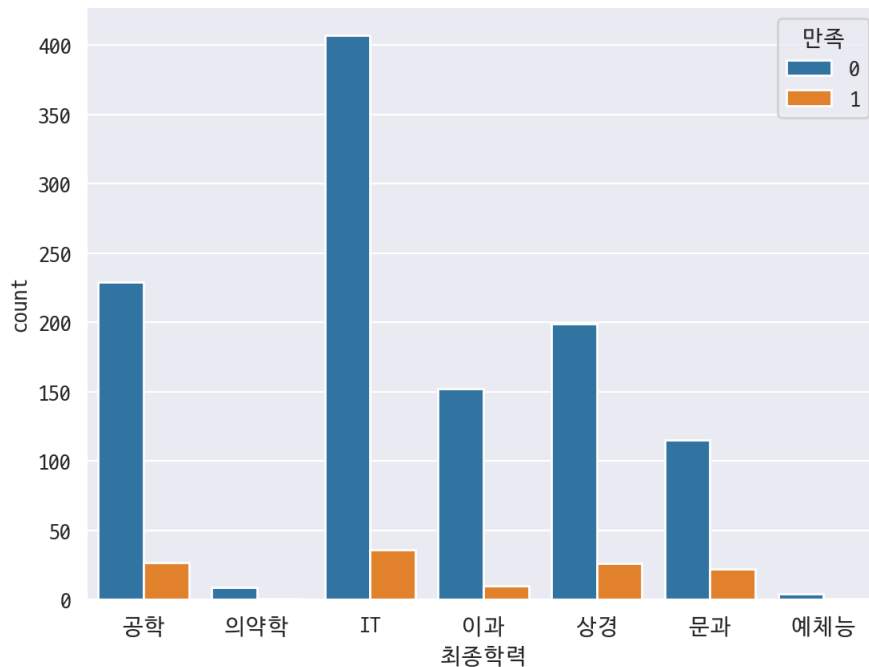
1. df_point : [최종학력, 희망직무]
2. df_re : [스터디 경험, 직무 경험]
3. df_satisfy : [최종학력, 졸업후 진로, 유입경로, 데이터역량필요]

3-2.

이를 통해 확인하니 전체 유료율은 48%로 확인되고 시각화하니 IT학과를 나온 사람들 중 데이터 분석가를 희망하는 사람들의 유료율은 39.2%로 약 10%p 낮은것을 확인할 수 있다 반면, 문과의 유료율은 52.1%로 전체 유료율보다 높은 것을 확인할 수 있다. 시각화로 확인하면 IT과보다 문과가 유료할 확률이 커 보인다.

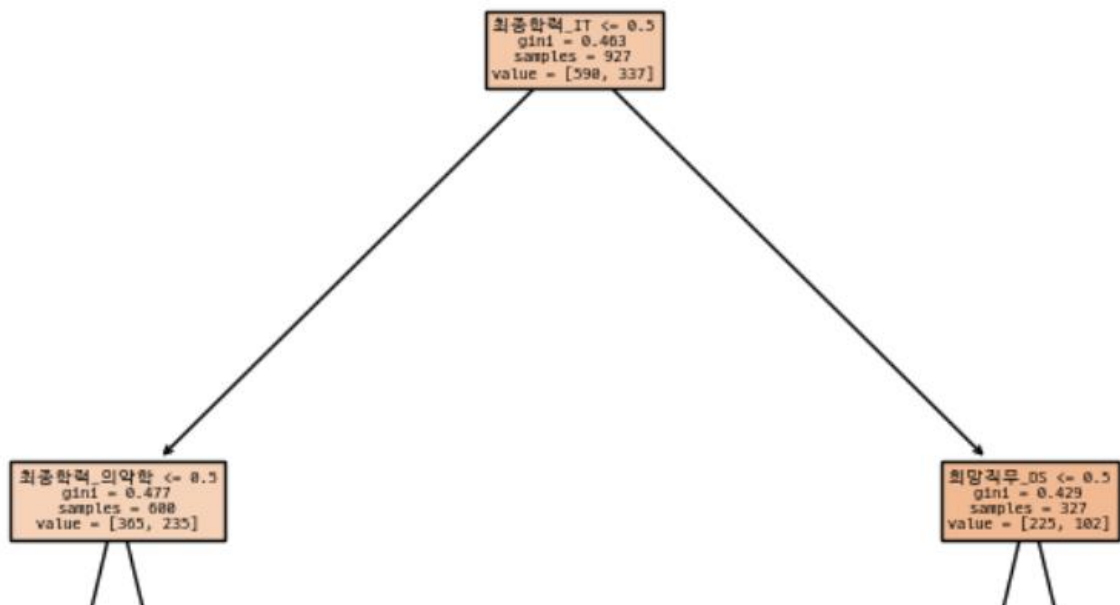


재참여에 있어서도 IT학과의 재참여율을 저조한 것을 확인할 수 있었다. 이는, 기초 수업 보다는 어려운 수업들을 듣기에 다시 들을 이유가 없거나 들어온 인원이 적은 것이라고 생각할 수 있다.



3-2. 의사결정나무

의사결정나무의 score 자체는 낮았으나 feature_importance와 tree plot을 통해 영향이 큰 변수들을 '확인하기로 했다. IT 학과가 가장 큰 영향도를 가지는 것을 확인할 수 있다.



```
{'최종학력_IT': 0.08497735337274806,
'최종학력_공학': 0.0438176530506728,
'최종학력_문과': 0.0202857609278138,
'최종학력_상경': 0.024779751203757687,
'최종학력_예체능': 0.06867337869124025,
'최종학력_의약학': 0.0337871360698655,
'최종학력_이과': 0.04059576420856861,
```

3-3 학과별 BDA 하는 이유

1. 문과

```
bda하는이유
데이터 분석 역량      98
프로젝트 경험          22
공모전 경험            11
인적 네트워크          5
공모전, 프로젝트 경험 둘 다  1
```

2. IT


```

bda하는이유
데이터 분석 역량          242
프로젝트 경험              135
공모전 경험                42
인적 네트워크             20
위 4항목 전부              1
할수 있는걸 모두 해보고 저에게 맞는것을 찾고 싶습니다      1
자격증 공부                1
파이썬 실력                1
Name: count, dtype: int64

```

3.

3. 결과 해석 및 아이디어 제시

- 1. 결과 해석

IT 관련 학과들의 수요율이 비교적 낮은 모습을 보인다. IT 학과들이 BDA 하는 이유는 프로젝트 경험을 쌓기 위해 학회를 참여한다. 가장 많은 인원들이 참여하지만 가장 큰 이탈율을 보이는 이유는 그들은 이론 보다 실전경험을 위해 BDA를 신청하기 때문이다.

-2. 아이디어 제시

실전 Kaggle 분석반이나 개인 프로젝트 포트폴리오반을 열어 실전경험을 쌓는 클래스를 만드는 게 좋을거 같다. ML반은 조원을 임의로 뽑는데 그것은 무작위 운에 기반한다. 어떤 조원들이 안 하거나 못하게 되면 나머지 조원들도 공모전을 못해 실전경험을 못 쌓게 되는 것이다. 그러니 오히려 타인의 영향을 받지 않고 개인 실전경험을 쌓을 수 있는 반을 만들면 좋을 것이라는 생각이 든다. 이 수업을 듣는 주된 목적의 대부분은 취업을 위함이다. 그러므로, 취업을 위해 공모전을 하고 실전 경험을 쌓길 바라는 것이기에 분야별 개인 프로젝트를 진행토록 하고 그 과정을 코칭해주는 클래스가 있으면 IT학과의 이탈율을 줄일 수 있을 거 같다.

4. 참고 문헌 출처 등

- 참고한 관련 문헌이 있을 경우 작성

