

「BDA 10기 학회원 최종 과제」 분석 보고서

*제출시 파일 이름은 ‘이름_2자리 숫자_최종과제_제출’입니다.

과제명	조별활동 활성화 방안 제시	성명	김혜원95, 박찬우46, 변해민27
활용데이터	member.csv, group_study – group_study.csv, group_point – point.csv		

1. 주제 선정 및 분석 배경

본 보고서는 학회원들의 수료율을 제고하기 위한 효과적인 방안을 모색하고자 작성되었다. 9기 최종 과제 분석 결과, 조별활동 이탈은 수료율과 부정적인 상관관계를 보였으며, 이는 조별활동이 학회 참여의 지속성과 밀접하게 연결되어 있음을 의미한다. 따라서 그룹 활동의 활성화가 수료율 증대에 실질적인 기여를 할 수 있음을 확인할 수 있다. 이에 따라 본 분석은 조별활동의 효과를 검토하고, 참여를 유도하며 이탈을 방지할 수 있는 개선 방안을 마련하는 것을 목적으로 한다. 이를 위해 조별활동 참여자와 미참여자의 특성, 전공 유형, 점수 및 잔류율을 비교분석하였으며, member, study, point 데이터를 활용하였다. 세 개의 데이터를 통해 조별활동이 수료율에 미치는 영향을 체계적으로 파악하고, 향후 학회 운영 전략 수립에 필요한 인사이트를 도출하고자 한다.

2. 분석 결과 상세 내용

1. 데이터 전처리

1-1. 데이터 정제 과정

1-1-1. 잘못된 member_id 및 leave_point 정제

먼저, study의 member_id 컬럼에서 "기말 휴회기간 이후", "2023060268"처럼 조와 관계없는 잘못된 ID를 제거했다. 또한 leave_point 컬럼에는 "기말 휴회기간 이", "기말 휴회기간 이휴" 등 오탈자와 유사표현이 많아, "기말 휴회기간 이후" 같은 통일된 표기로 일괄 치환하였다.

1-1-2. IT/Non-IT 전공 분류 처리

major1_1, major1_2, major1_3 세 컬럼 중 하나라도 IT 관련 전공이라면 'IT 계열'로 분류하도록 설계되었다. 전공 명은 it_majors라는 사전 정의 리스트를 기준으로 판단된다.

1-1-3. leave_point의 결측 처리 및 '확인 불가' 치환

leave_point는 공백 제거 및 nan, None 값을 모두 "해당없음"으로 통일한 후, 이 중 이탈자(group_leave == 1)인데 leave_point가 "해당없음"인 경우에는 실질적 이탈 사유가 확인되지 않으므로 "확인 불가"로 재정의하였다.

1-1-4. 유효 그룹 판별 및 고아 그룹 제거

그룹을 식별하기 위해 세 가지 키(generation, class_code, group_number)를 기준으로 study와 point에 존재하는 그룹들을 비교하였다.

- study에는 있지만 point에는 없는 그룹 = orphans_in_study
- point에는 있지만 study에는 없는 그룹 = orphans_in_point

이러한 그룹들을 고아 그룹(orphan group)이라 하며, 이들을 모두 제거하고 양쪽에 모두 존재하는 그룹만 유효 그룹으로 남겼다. 고아 그룹에 속한 member_id들은 이후 모든 데이터에서 제거했다.

1-1-5. study_ok ↔ member 간 키 정합성 확인

study_ok에 존재하지만 member에는 없는 (member_id, generation) 쌍이 존재할 수 있으므로, member_keys 기준

으로 inner-merge하여 최종적으로 study_ok에도 member 정보가 존재하는 사람만 남겼다.

1-2. 중복 참여자 처리 방식

분석 전처리에서는 주로 (member_id, generation)을 키로 사용하고 있으며, 이를 기준으로 중복 제거(drop_duplicates)를 수행하기 때문에 동일 기수에서 여러 조에 참여한 사람도 1명으로만 취급된다. 이는 개인 특성 비교 분석에서 중복 참여로 인한 왜곡을 방지하는 데 효과적이다.

1-3. 최종 산출물

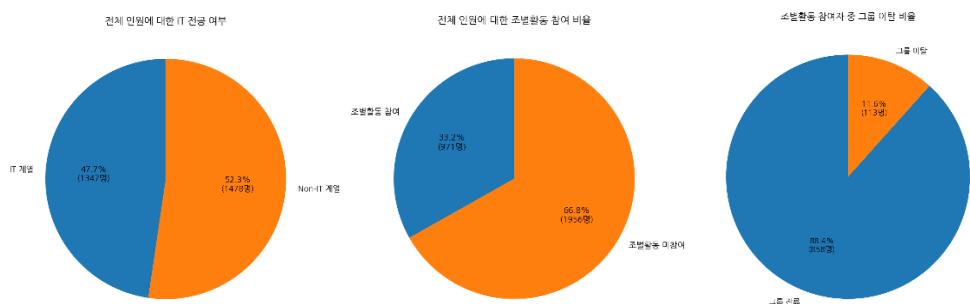
본 전처리 과정을 통해 다음과 같은 분석용 테이블들이 구축되었다:

- **total**: 전체 멤버 기반, 참여 여부 포함(전체 비교용)
- **group_yes**: 조별활동 참여자 전용 테이블
- **group_yes_point**: 점수 정보까지 포함된 참여자 테이블
- **group_no**: 조별활동 미참여자 전용 테이블

이러한 정제된 데이터를 바탕으로 조별 활동의 효과성과 참여자 특성 간의 관계를 체계적으로 분석할 수 있는 기반을 마련하였다.

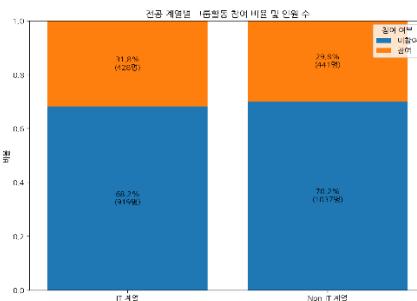
2. 개인 특성별 프로파일링

2-1. 전체 현황 개요



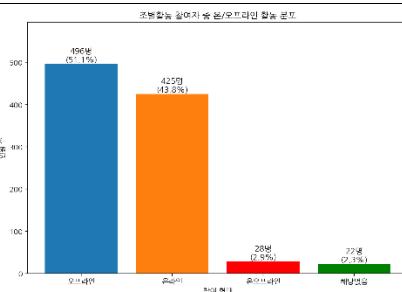
전체 인원 중 IT 전공자는 47.7%(1,347명), 비IT 전공자는 52.3%(1,478명)으로, 전공 배경이 고르게 분포되어 있음을 확인할 수 있다. 또한, 전체 인원 중 약 33.2%(971명)이 조별활동에 참여하였으며, 동일한 기수에서 복수 조에 참여한 인원은 중복 없이 1명으로 집계하였다. 이로써 전체 구성원의 약 3명 중 1명이 실질적으로 조별활동에 참여하고 있는 것으로 나타났다. 조별활동 참여자를 대상으로 중도 이탈률을 분석한 결과, 11.6%(113명)이 활동 도중 조를 이탈한 것으로 확인되었다.

2-2. 전공별 참여 현황



조별활동 참여 양상을 전공별로 세분화하여 분석한 결과, IT 전공 여부와 관계없이 조별활동 참여율은 약 30% 수준으로 유사한 경향을 보였다. 이는 IT 계열과 비IT 계열 모두에서 조별활동 참여율에 뚜렷한 차이가 나타나지 않았다.

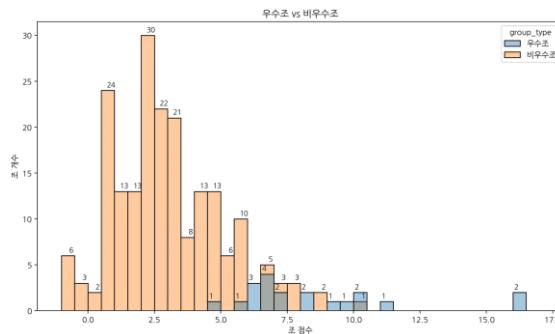
2-3. 참여자 활동 특성



조별활동 참여자를 대상으로 진행한 설문 조사에서 활동 참여 방식(온라인·오프라인 여부)에 대한 응답 결과, 전체 응답자 중 오프라인으로 참여한 비율이 가장 높았으며 51.1%에 해당하는 496명이었다. 그 뒤를 이어 온라인 참여자는 43.8%로 425명, 온·오프라인을 병행한 응답자는 2.9%로 28명으로 나타났다.

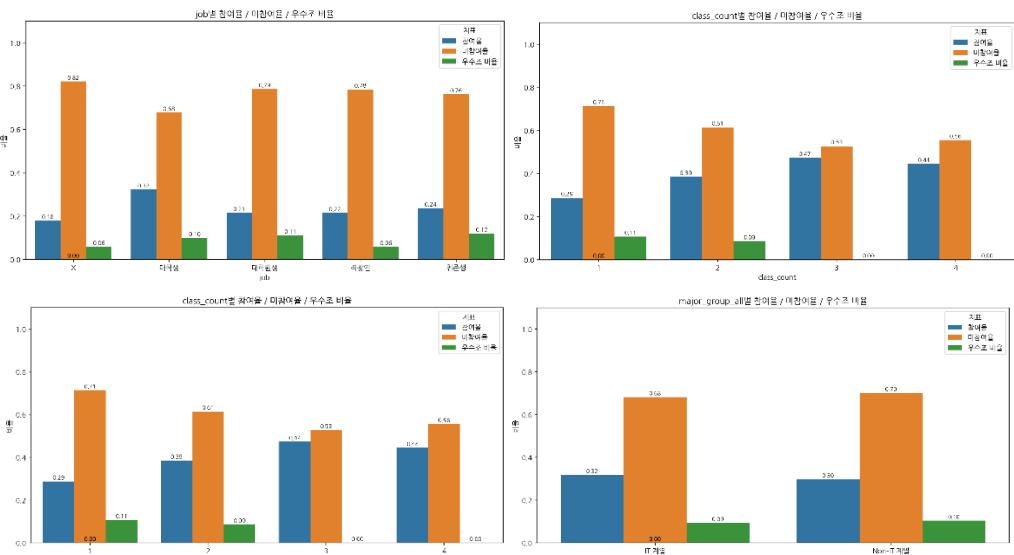
3. 우수조 프로파일링

3-1. 우수조·일반조 점수 분포 현황



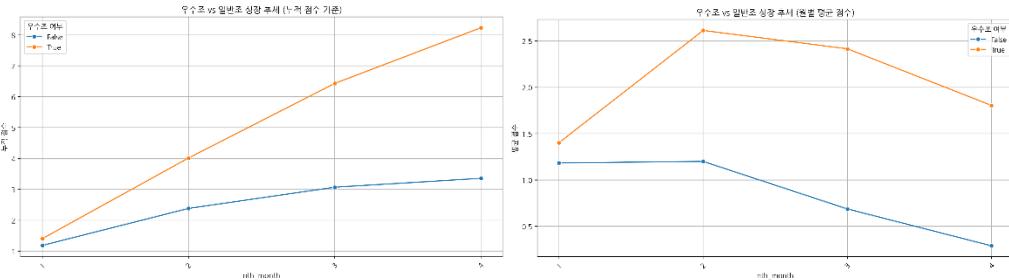
우수조는 주로 6.0점 이상의 고득점 구간에 분포하며, 7.0점 이상에서는 대부분 우수조만 존재한다. 비우수조는 2.0~5.5점 사이에 집중되어 있으며, 특히 2.5~3.0점 구간에 조가 가장 많다. 3.0~6.0점 구간에서는 두 그룹이 함께 분포하지만, 1.0점 이하의 저득점 영역은 비우수조만 존재한다. 전체적으로 우수조는 점수 기준과 밀접하게 연결되어 있는 분포를 보인다.

3-2. 개인특성별 우수조·일반조·미참여 비교 분석



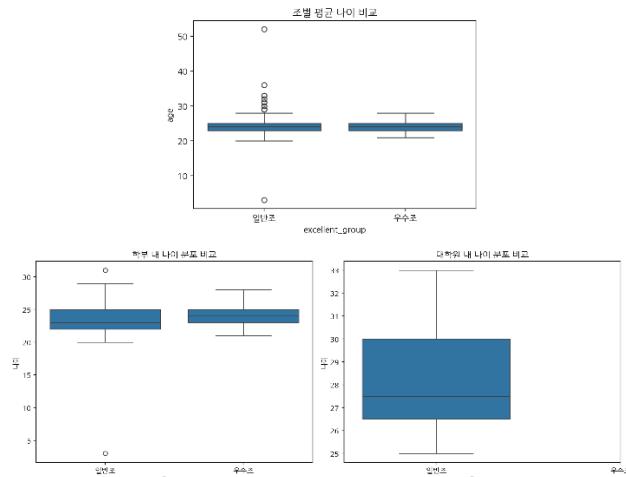
참여율과 우수조 비율을 분석 결과, 직업별로는 대학생들이 가장 많이 조별활동에 참여했으며, 우수조 비율은 대학원생, 대학생, 취업준비생 순으로 높게 나타났다. 수업 수가 많은 학회원일수록 참여율은 높지만, 실제 우수조에 포함되는 학회원은 수업을 1~2개 수강한 그룹에 집중되는 경향을 보였다. 전공 계열별로는 IT 계열 학회원의 참여율이 비IT 계열보다 높았으나, 우수조 비율에는 큰 차이가 없었다. 이러한 결과는 특정 변수에서 참여율이 높다고 해서 반드시 우수조가 될 확률이 높아지는 것이 아니며 우수조를 결정하는 명확한 기준이나 패턴을 찾기 어렵다는 것을 보여준다.

3-3. 우수조·일반조 점수 성장 추세 비교



우수조는 초기부터 누적 점수가 빠르게 상승하여 일반조보다 4개월 차 기준 거의 2배 높은 수준을 기록하였으며 월별 평균 점수도 초기 최고점 도달 후 소폭 하락했으나 여전히 일반조보다 꾸준히 높게 나타났다. 반면 일반조는 초기 소폭 상승 후 점수가 감소하는 경향을 보여, 우수조는 높은 참여도와 학습 효과로 안정적인 성장세를 보인 반면 일반조는 성장 유지에 어려움이 있는 것으로 해석할 수 있다.

3-4. 우수조·일반조 나이 분포 비교 분석



조별 평균 나이를 비교한 결과, 일반조와 우수조 모두 중앙값은 23~25세 전후로 유사하였다. 그러나 일반조에서는 극단적으로 나이가 낮거나 높은 일부 이상치가 확인되어 전체 나이 범위가 더 넓게 나타났다. 학부생 기준으로 볼 때 두 그룹의 중앙값과 분포 범위는 크게 차이가 없었으나, 대학원생은 일반조에서만 확인되었다. 이러한 결과는 우수조가 상대적으로 연령 편차가 적고 동질적 구성을 이루어 협업과 소통이 원활했을 가능성이 높다.

4. 조별활동 이탈자 프로파일링

4-1. 탈퇴 시점



탈퇴 시점별 분포를 살펴본 결과 중간 휴학기간 이후 탈퇴가 39.8%로 가장 높게 나타났으며, 그 다음은 중간 휴학 이전(27.4%), 기말 휴학 이후(25.7%) 순으로 나타났다. 이는 학기 중후반부에서 탈퇴가 집중되는 경향을 보여주며, 학업 부담이나 활동 피로 누적 등이 주요 원인으로 작용했을 가능성을 시사한다. 또한 일부 데이터(7.1%)는 확인이 불가하여 탈퇴 시점 정보의 정확성을 보완할 필요가 있다.

5. 통계분석 (group_leave 대상)

5-1. 상관관계 분석

```

group_leave_numeric 1.000000    --- job ---
age 0.031990      job
school1_proc 0.013799      징장인 0.300000
school2_proc 0.002452      대학생 0.117849
class_count -0.081986      취준생 0.062500
excellent_group -0.108743      X 0.056556
before_exist NaN      대학원생 0.000000
Name: group_leave_numeric, dtype: float64
Name: group_leave_numeric, dtype: float64

--- major_group_all ---
major_group_all IT 계열 0.127389
Non-IT 계열 0.106000
Name: group_leave_numeric, dtype: float64
Name: group_leave_numeric, dtype: float64

--- group_master ---
group_master 0 0.125846
1 0.086207
Name: group_leave_numeric, dtype: float64
Name: group_leave_numeric, dtype: float64

--- on_offline ---
on_offline
해당없음 1.000000
온오프라인 0.142857
온라인 0.122353
오프라인 0.070565
Name: group_leave_numeric, dtype: float64

```

그룹 이탈 유무와 다른 변수들 간의 상관관계 분석 결과, 탈퇴율과 가장 밀접한 요인은 직업, 참여 방식, 조장 여부로 확인되었다. 직업별로는 직장인이 가장 높은 탈퇴율을 보였으며, 대학생과 취업준비생은 중간 수준, 대학원생은 상대적으로 낮았다. 이는 직업군 특성이 활동 지속성에 직접적인 영향을 줄 수 있음을 보여준다. 또한 온·오프라인 참여 방식 역시 탈퇴율과 유의미한 양의 상관관계를 보여, 온라인 또는 혼합 참여자는 오프라인 참여자보다 탈퇴 가능성이 높았다. 이는 활동 방식의 일관성이 안정성 유지에 중요한 역할을 함을 시사한다. 전공(IT/Non-IT 계열)은 탈퇴율과의 상관계수가 약 +0.10~+0.12로 크지는 않지만 일정한 영향을 보였다. 이는 개인 단위의 전공 특성이 직접적으로 탈퇴를 결정짓는 수준은 아니더라도, 구성의 조합이나 학습 맥락에서는 일정 부분 작용할 수 있음을 의미한다.

5-2. 카이제곱 연관성 분석

	group_leave	0	1	
generation	→ p-value: 0.0000	Y	0.274501 -0.756394	
school1	→ p-value: 0.0004	X	-0.046361 -0.127148	group_leave
major_group_all	→ p-value: 0.3479	대학생	0.410602 -1.191425	0
salary_1	→ p-value: 0.9999	대학원생	0.392602 -0.100430	1
salary_2	→ p-value: 0.9963	취준생	0.390705 -1.054451	
job	→ p-value: 0.0400	company	0 11 12	0.553656 -1.525114
job_domain	→ p-value: 0.0000	IT 기업	0.128001 -0.341138	-0.291321 0.775188
cooperate	→ p-value: 0.0115	비IT 기업	0.128001 -0.341138	0.436559 -1.208461
major_data	→ p-value: 0.9996	X	0.106897 -0.294559	0 15 0.150008 -0.4149573
class_count	→ p-value: 0.0740	전공	-0.948001 2.555059	0 16 0.362008 -0.000005
class1	→ p-value: 0.0375	전공1	0.128001 -0.341138	0.429312 1.182390
class2	→ p-value: 0.0673	전공2	0.128001 -0.341138	0 -0.160659 0.443279
class3	→ p-value: 0.0000	전공3	0.128001 -0.341138	5 -0.287453 0.792083
class4	→ p-value: 0.0591	전공4	0.128001 -0.341138	6 -0.223023 0.614544
class5	→ p-value: 0.0000	전공5	0.128001 -0.341138	7 0.17397 -0.000005
before_exist	→ p-value: 1.0000	대기업	0.128001 -0.341138	0 8 0.356008 -0.988237
class_code	→ p-value: 0.0780	중소기업	0.128001 -0.341138	0.955957 2.739385
group_master	→ p-value: 0.1272	IT 기업	-0.577149 1.590349	0 1 일반학회원
on_offline	→ p-value: 0.0000	비IT 기업	-0.577149 -0.341138	0 0 0.247603 -0.682275
excellent_group	→ p-value: 0.0013	X	0.128001 -0.341138	0 -0.331800 0.914283
excellent_group_all	→ p-value: 0.0507	연구원	0.128001 -0.341138	0 1 1.107312 -3.051228
age	→ p-value: 0.1200	교수	0.128001 -0.341138	
school1_proc	→ p-value: 0.7925	기술자	0.128001 -0.341138	
school2_proc	→ p-value: 1.0000	비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128001 -0.341138	
		연구원	0.128001 -0.341138	
		교수	0.128001 -0.341138	
		기술자	0.128001 -0.341138	
		비즈니스	0.128	

로 판단되어 이 피처가 가장 중요도가 높게 나왔지만 실제 영향력이 크다고 보기 어렵다. 나이 피처 또한 대부분의 회원이 23~25세에 분포해 있어 영향력이 크다고 보기 어렵다.

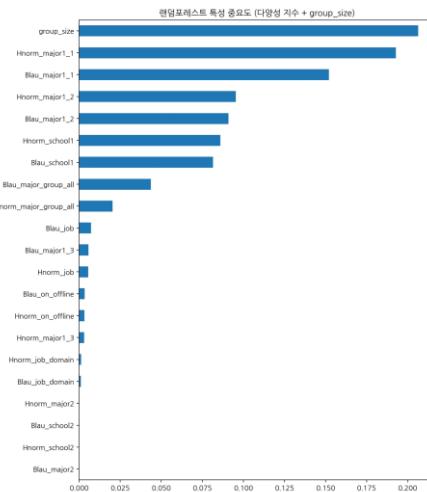
6. 다중 회귀분석 (leave_point 대상)

Class 0:	major1_1_IT(컴퓨터 공학 포함): 2.20 major1_1_경영학: 1.89 major1_1_교육학: 1.62 major1_1_인문학: 0.55 major1_1_환경공학: 1.05 job_domain_R_보건업 및 사회복지 서비스업: 1.80 company_연구보조원: 1.80 on_offline_온프라인: 1.69 on_offline_해당없음: 0.61
Class 1:	major1_1_X: 1.69 major1_1_환경공학: 1.91 major1_1_인문학: 1.65 major1_1_경영학: 1.65 major1_1_전자공학: 2.05 major1_1_환경통계학과: 1.54 major1_1_컴퓨터소프트웨어학부: 2.05 major1_1_한문학과: 2.03 on_offline_온라인: 0.41 on_offline_해당없음: 0.61
Class 2:	major1_1_경계생물학: 1.95 major1_1_사회복지학: 1.62 major1_1_예체능: 0.67 major1_1_의약학: 1.97 on_offline_온라인: 2.01 on_offline_해당없음: 0.46
Class 3:	major1_1_시용한학부: 0.64 major1_1_거버넌스경영학부: 0.54 major1_1_경제통상학부: 0.60 major1_1_국제사무학과: 0.67 major1_1_소프트웨어학과: 0.45 major1_1_전자통계학과: 0.52 major1_1_환경통계학과: 0.65 major1_1_경영보건신공학부: 0.52 major1_1_환경분야학부: 0.50 job_domain_R_보건업 및 사회복지 서비스업: 0.66 company_연구보조원: 0.66 on_offline_온프라인: 0.53 on_offline_해당없음: 0.11
Class 4:	major1_1_환경공학: 1.95 major1_1_사회복지학: 1.62 major1_1_예체능: 0.67 major1_1_의약학: 1.97 on_offline_온라인: 2.01 on_offline_해당없음: 0.46
클래스 설명	
Class 0 : 기말 휴회기간 이후	excellent_group: 3.78 excellent_group_master: 2.87
Class 1 : 중간 휴회기간 이전	class_count: 1.53 before_exist: 2.47
Class 2 : 중간 휴회기간 이후	generation: 0.64 on_offline: 0.34
Class 3 : 해당없음	excellent_group: 0.52 excellent_group_master: 0.66
Class 4 : 확인 불가	school1_proc: 0.80 school2_proc: 2.06 generation: 0.26

마스킹 처리되어 있는 학교 데이터와 수업 데이터, 또는 데이터가 지나치게 다양한 학과 데이터는 분석 대상에서 제외하였다. Class 0~2는 조별활동 중 이탈한 학회원들을 의미한다. 분석 결과, 온/오프라인 참여 방식이 온라인인 경우, 중간 휴회기간 이전에는 조별활동 이탈 가능성을 낮추는 효과가 나타났으나, 중간 휴회기간 이후에는 이탈 가능성을 높이는 경향을 보였다. 또한, 중간 휴회기간 동안 연구보조원 직무를 수행한 학회원과 온/오프라인 참여방식이 온라인인 경우에는 기말 휴회기간 이후 조별활동을 이탈할 가능성이 증가하는 것으로 나타났다. Class 3은 조별활동 중 이탈하지 않은 학회원들을 의미한다. 분석 결과, 직장인, 특히 연구보조원 직무를 가진 학회원과 온/오프라인 참여방식을 혼합으로 선택한 학회원은 조별활동 이탈 가능성이 상대적으로 높은 것으로 확인되었다.

7. 다양성 지수

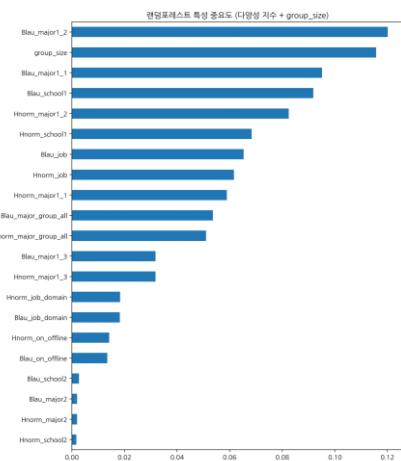
7-1. 조별활동 지속 요인 다양성 분석



조별활동의 잔류율에 영향을 미치는 요인을 파악하기 위해, 그룹 단위의 다양성 지수 및 조 규모와 잔류율 간의 상관관계를 분석하였다. Blau_major1_3 +0.190, Hnorm_major1_3 +0.188, Blau_major1_1 +0.119, Hnorm_major1_1 +0.111, Blau_school2 +0.060, Hnorm_school2 +0.060으로, 학력 배경의 다양성이 잔류율과 양의 상관관계를 보이는 것으로 확인되었다. Hnorm_on_offline -0.263, Blau_on_offline -0.259, group_size -0.058로, 온·오프라인 혼합 구성이나 조 규모가 클수록 잔류율이 음의 상관관계를 보이는 것으로 확인되었다. 이는 전공 및 최종 학력의 다양성은 오히려 조직의 안정성 유지에 긍정적인 역할을 하며, 활동 방식이 일관되지 않고 혼합된(on/offline) 조는

안정성이 떨어질 가능성이 높다.

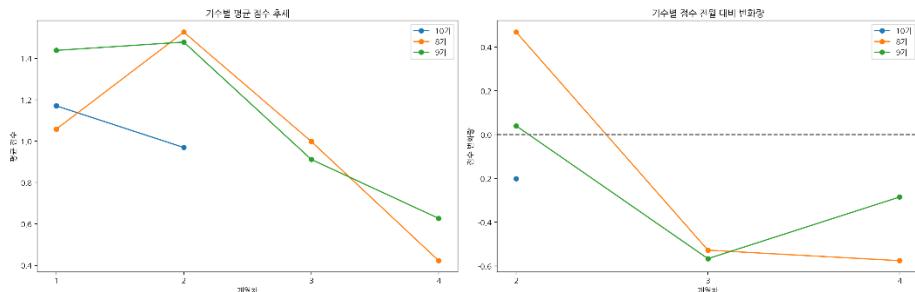
7-2. 조별 성과 향상 요인 다양성 분석



조별활동에서 평균 점수에 영향을 미치는 요인을 파악하기 위해, 그룹 단위의 다양성 지수 및 조 규모와 조 평균 점수(point_mean) 간의 상관관계를 분석하였다. Blau_major1_1 +0.136, Hnorm_major1_1 +0.104, group_size +0.073, Hnorm_major_group_all +0.067, Blau_school1 +0.063으로, 주전공의 구성 다양성과 출신학교, 전공 계열 혼합 정도, 조 규모가 평균 점수와 양의 상관관계를 보이는 것으로 확인되었다. 반면, Hnorm_major1_3 -0.084, Blau_major1_3 -0.090으로, 세 번째 부전공의 존재는 평균 점수와 음의 상관관계를 보이는 경향을 나타냈다. 이는 주전공의 다양성은 조별활동의 학습성과에 긍정적인 영향을 미칠 수 있으며, 학교의 다양성 역시 점수 향상과 일정 부분 연관이 있을 수 있다. 그리고 세 번째 부전공의 존재는 부정적인 영향을 미칠 수 있다.

8. 점수 추세 분석

8-1. 기수별 점수 추세 및 변화량 분석



기수별 평균 점수 추세를 살펴본 결과, 9기와 8기는 초반 점수가 비교적 높았으나 2개월 차 이후부터 활동이 줄어들며 3개월부터 점수가 급격히 하락하는 공통적인 패턴을 보였다. 이는 모든 기수에서 중반 이후 활동 저하가 일관되게 나타난다는 점을 보여주며, 특히 2개월 차 이후 활동 관리의 필요성을 시사한다.

전월 대비 점수 변화율을 기수별로 살펴보면, 8기는 2개월 차까지 상승세를 보였으나 이후부터 급격히 하락하였다. 9기 또한 2개월 차까지는 안정적인 흐름을 유지했으나, 이후 점수가 점차 감소하여 3개월 차에는 뚜렷한 하락세를 보였다. 이러한 결과는 2개월 차 이후부터 활동 저하가 본격화된다는 공통점을 확인시켜준다.

8-2. 9기 특화 분석 기준 및 평가 체계

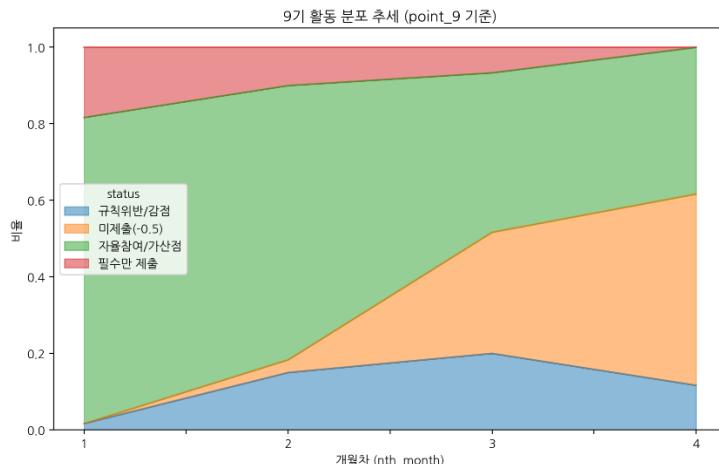
8-2-1. 분석 대상 선정 배경

9기는 4개월치 데이터가 모두 존재하며, 활동 전 기간에 걸친 추세 분석이 가능하다. 또한, 평가 기준이 일관되게 적용되어, 활동 결과를 시점별로 비교·분석하는 데 적합하다. 반면, 8기의 경우 달마다 세부 기준이 변동되었고(예: 가산점 부여 방식 변화), 미제출 점수 또한 9기 및 10기와 달라 직접적인 비교가 어렵다. 10기는 데이터가 2개월치에 불과하여 장기적인 추세나 활동 지속성 분석에 한계가 있다. 따라서 본 보고서는 데이터 기간이 충분하고, 평가 체계가 가장 안정적으로 적용된 9기 데이터를 중심으로 분석을 진행하였다.

8-2-2. 9기 평가 체계

9기 평가 기준을 정리하면, 이를 기반으로 본 분석에서는 점수를 네 가지 상태로 단순화하여 분류하였다. **미제출(-0.5)**: 활동 자체를 제출하지 않은 경우, **규칙위반/감점(<1점)**: 자각·캠 미온·필수 일부 미제출 등으로 감점된 경우, **필수만 제출(=1점)**: 모든 필수 활동만 이행한 경우, **자율참여/가산점(>1점)**: 자율 활동이나 가산점 활동까지 참여한 경우로 정의하였다.

8-3. 9기 활동 분포 추세



9기의 활동 분포 추세를 세부적으로 분석한 결과, 1개월 차에는 자율참여 및 가산점 활동이 높은 비중을 차지하며 참여 열기가 강했음을 알 수 있다. 그러나 시간이 지남에 따라 자율참여의 상대적 비중은 줄어들고, 대신 미제출과 규칙위반/감점 활동의 비율이 꾸준히 증가하는 양상을 보였다. 특히 2개월 차 이후 활동부터 미제출 비율이 크게 확대되어 4개월차에는 전체 활동의 절반 이상을 차지하였다. 반면, 필수만 제출 활동은 초반에 일정 비중이 있었으나 후반부로 갈수록 점차 감소하여 4개월 차에는 거의 사라졌다.

3. 결과 해석 및 아이디어 제시

1. 조 구성 다양성

1-1. 분석 결과

조 평균 점수는 주전공과 출신학교의 다양성, 전공 계열 혼합 정도, 조 규모와 양의 상관관계를 보였으며, 세 번째 부전공의 다양성은 음의 상관관계를 나타냈다. 잔류율은 전공 및 학력 배경의 다양성과 양의 상관관계를 보였으나, 온·오프라인 혼합 구성과 조 규모 확대는 음의 상관관계를 보였다.

1-2. 개선 방향

조별활동의 성과와 안정성을 동시에 높이기 위해서는 주전공과 출신학교의 다양성을 확대하는 방향으로 조를 구성하는 것이 바람직하다. 다만, 세 번째 부전공까지 과도하게 분산되는 경우에는 학습 성과 저하가 우려되므로 배정 시 신중함이 필요하다. 또한, 온·오프라인 활동 방식을 명확히 구분하여 운영하고, 조 규모는 적정 수준으로 유지하는 것이 중요하다.

2. 학회원 지속성 강화

2-1. 분석 결과

활동 2개월 차 점수 자체는 높은 수준을 보였으나, 진행 시점인 '중간 휴회기간 이후'에 다수의 학회원이 조별활동을 이탈하는 현상이 관찰되었다. 이로 인해 이 시점 이후부터 참여 인원이 줄어들면서 자율참여/가산점 활동이 감소하고, 미제출 비율이 높아져 3개월 차부터 조별 점수가 하락하는 경향이 확인되었다.

2-2. 개선 방향

조별활동의 지속성을 강화하기 위해 중후반기 집중 관리 프로그램을 운영할 필요가 있다. 첫째, 2개월 차 활동 결과가 발표된 직후 중간 성과 공유 세션을 운영하여 각 조가 활동 성과를 공유하고, 이를 통해 학회원들이 단기 목표를 설정하고 동기를 유지할 수 있도록 한다. 이 과정에서 중간 우수조 발표 제도를 도입해 참여 의욕을

높인다. 둘째, 개별 피드백 제공을 의무화하여 공모전뿐만 아니라 프로젝트 등 다양한 자율 활동에도 체계적인 피드백을 제공한다. 이를 통해 공모전에 참여하기 어려운 학회원들에게도 실질적인 성장 기회를 보장하고 자율 활동 참여를 꾸준히 이어가게 하는 환경을 조성한다. 셋째, 성과 기반 인센티브 제도를 도입하여 3개월 차까지 성실히 활동하고 일정 기준 이상의 자율활동 점수를 달성한 학회원에게는 Knock 서비스 Q&A 무료 이용권을 제공한다. 이를 통해 학회원들이 장기적인 관점에서 활동을 지속할 수 있도록 유도한다.

2-3. 기대 효과

이러한 제도 도입을 통해 다음과 같은 개선 효과가 기대된다. 중간 우수조 발표는 학회원들에게 단기 목표를 제공하여 중간 휴회기간 이후의 이탈률을 감소시킬 것으로 예상된다. 확대된 피드백 시스템은 학회원들의 다양한 활동에 대한 방향성을 제시하여 참여 만족도를 높이고, 학회원들의 지속적인 성장을 지원할 것이다. 또한 성과 기반 인센티브 제도는 우수 학회원의 장기적 참여를 유도하는 동시에 다른 학회원들에게도 긍정적 자극 부여함으로써 전체적인 학회 활동의 질적 향상과 참여율 증대에 기여할 것으로 기대된다.

3. 오프라인 활동 활성화

3-1. 분석 결과

다중 회귀 분석을 통해 조 이탈 시점을 분석한 결과, 온라인으로만 참여하는 학회원들이 중간 휴회기간 이후와 기말 휴회기간 이후에 조 이탈할 가능성이 높은 것으로 나타났다. 이는 온라인 참여 방식이 학회원들의 소속감과 지속적 참여 의지에 부정적 영향을 미치는 것으로 해석된다.

3-2. 개선 방향

조별활동에서 오프라인 참여를 유도하기 위해 다음과 같은 방안을 제시한다. 첫째, 조별활동 보고서 평가 시 오프라인으로 진행된 활동에 대해 가산점을 부여한다. 둘째, 자율활동 중 오프라인으로 진행되는 프로젝트에 대해서는 추가적인 가산점을 제공하여 적극적인 오프라인 참여를 장려한다.

3-3. 기대 효과

오프라인 활동 가산점 제도를 통해 학회원들의 대면 소통이 증가하고, 이는 조별 결속력 강화와 소속감 향상으로 이어질 것으로 예상된다. 결과적으로 휴회기간 이후 이탈률을 감소시키고 학회원들의 지속적인 참여를 유도할 수 있을 것으로 기대된다.

4. 이탈 원인 분석 체계화

4-1. 분석 결과

제공된 데이터로는 조 이탈 유무와 이탈 시점만 파악할 수 있어, 조 이탈의 근본적 원인에 대한 분석이 제한적인 것으로 나타났다. 학회 운영의 지속적 개선을 위해서는 이탈 원인에 대한 세부적인 정보 수집이 필요한 것으로 판단된다.

4-2. 개선 방향

조 이탈 원인의 체계적 분석을 위해 다음과 같은 방안을 제시한다. 첫째, 조 이탈 시 의무적으로 이탈 사유 작성 폼을 제출하도록 한다. 둘째, 이탈 사유를 작성하지 않고 탈퇴하는 경우 BDA 영구 제재 대상으로 분류하여 향후 재가입을 제한한다.

4-3. 기대 효과

이탈 사유 수집 체계화를 통해 학회원 이탈의 주요 원인을 파악할 수 있으며, 이는 향후 학회 운영 개선 방향 수립에 핵심적인 데이터로 활용될 것이다. 또한 의무적 사유 작성 제도는 학회원들로 하여금 이탈 결정을 신중하게 고려하도록 유도하여 충동적 이탈을 방지하는 부차적 효과도 기대된다.

5. 분석의 한계점

이번 분석 과정에서는 데이터 구조 및 품질과 관련하여 몇 가지 중요한 한계점이 확인되었다.

5-1. 선제적 마스킹의 한계

첫째, 데이터 보호를 위한 마스킹 기준이 일관되지 않아 분석에 상당한 제약이 발생했다. 예를 들어, class1,

class2 등의 항목은 BDA 수업 참여 여부를 판단할 수 있는 핵심 정보임에도 불구하고 과도한 마스킹으로 인해 어떤 수업을 의미하는지 확인할 수 없었다. 반면 개인정보로 분류되는 생년월일은 마스킹되지 않은 상태로 제공되어, 데이터 보호 기준이 일관되지 않다는 문제가 드러났다. 이러한 불일치는 분석의 정확성을 저해하는 동시에 개인정보 보호 측면에서도 우려를 야기한다.

5-2. 데이터 정합성 문제

둘째, 설문조사 형식의 비표준화로 인해 심각한 데이터 정합성 문제가 발생하였다. Member 테이블의 birthday, major1_1, major1_2, major1_3 등 주요 컬럼에 대해 정규화된 형식 없이 다양한 형태의 입력이 존재되어 있었다. 구체적으로 생년월일의 경우 "2021", "2022", "420", "응답 없음" 등으로 입력된 사례가 존재했으며, 전공 정보에는 오타와 비정형 분류가 다수 포함되어 IT 전공 여부 판단이나 전공 간 우선순위 식별이 어려웠다. 이러한 데이터 품질 문제는 정확한 통계 분석을 수행하는 데 근본적인 장애요소로 작용했다.

5-3. 데이터 오류 및 테이블 간 불일치

셋째, 여러 테이블 간 참조 무결성 문제와 평가 기준의 불명확성이 확인되었다. Member 테이블에 있는 ID가 Study 테이블에 존재하지 않거나, Study 테이블에는 조 정보가 있으나 해당 조가 Point 테이블에는 없는 경우 등이 빈번하게 발견되었다. 특히 일부 조는 같은 구성원이 포함되어 있음에도 불구하고 excellent_group 컬럼 기준으로 우수조와 비우수조로 존재되어 분류된 사례도 확인되었다. 또한 Point 값이 어떤 기준으로 부여되었는지 명확하지 않았으며, 확인 결과 8기, 9기, 10기마다 평가 기준이 달랐던 것으로 보인다.

5-4. 미흡한 데이터 정의 및 품질

넷째, 제공된 컬럼 정의서의 컬럼 설명이 불충분하여 데이터 해석에 어려움이 있었다. 실제 테이블 내에 존재하는 일부 컬럼이 컬럼정의서에 누락되어 있어 해당 컬럼의 의미를 파악할 수 없었다. 또한 9기의 학회원들의 수료율이 다른 기수들에 비해 비정상적으로 높게 나타나 데이터의 신뢰성에 의문이 제기되었다. 더불어 온/오프라인 참여 방식을 나타내는 컬럼에서 '해당없음'으로 분류된 데이터가 구체적으로 무엇을 의미하는지 명확하지 않아 참여 형태 분석에 제약이 따랐다.

이와 같은 오류 및 불일치 현상으로 인해 대부분의 데이터를 주관적 판단에 따라 변환하거나 정제해야 했다. 따라서 이번 분석 결과는 정확한 데이터 분석이라기보다는 오류를 최대한 보정한 경험적 분석에 가깝다는 한계를 갖는다. 이러한 문제들은 향후 데이터 수집과 관리 체계의 전면적인 개선이 필요함을 시사하며, 데이터 품질이 개선된다면 보다 신뢰도 높은 분석과 구체적인 전략 제안이 가능할 것으로 판단된다.

4. 참고 문헌 출처 등

없음

※ 5~10장 내외로 작성