

수학 모의고사 성적 데이터 분석 포트폴리오

대형학원 사설 모의고사 분석

목차

1. 분석 시나리오

2. 데이터 수집

3. 데이터 구조

4. 데이터 전처리 과정

5. 데이터 분석 및 EDA

6. 결과 및 인사이트

분석 시나리오



모의고사 개요

- 대형학원 주관
- 2023학년도 9월 모의고사 대비
- 수학 사설 모의고사 실시
- 300명 학생 대상 시행

분석 목적

- 모의고사 난이도 평가
- 등급컷 설정
- 학생들의 성적분포 분석

데이터 수집 형태

Chat GPT 활용 데이터 생성

- 분석 시나리오 기반 chat GPT를 활용하여 가상의 데이터 생성
- 300명의 학생이 작성한 OMR을 바탕으로 데이터를 생성한다고 설정
- 실제 2023학년도 9월 모의고사 수학문제를 바탕으로 데이터 생성

데이터 구조

OMR 테이블 - 학생들이 작성한 OMR을 가지고 생성했다고 시나리오를 가정했다

	수험번호	선택과목	문제번호	대분류	중분류	소분류	학생이 쓴 답	정답	배점
0	1	공통	1	수학 I	지수와 로그	로그 함수의 성질	4	4	2
1	1	공통	2	수학 II	함수의 극한	함수의 극한값 계산	1	1	2
2	1	공통	3	수학 I	삼각함수	삼각함수의 성질	2	2	2
3	1	공통	4	수학 II	함수의 연속성	연속의 개념 및 조건	1	1	3
4	1	공통	5	수학 I	수열	등차수열의 일반항	3	3	3

데이터 전처리 과정

1. 수험번호 칼럼을 좀 더 현실적으로 반영하고 싶어서 20220001~ 20220300으로 변형했다.
2. 선택과목 칼럼은 학생들이 선택한 선택과목 (미적분/기하와 벡터/확률과 통계)인데 문항의 범위를 나타내어서 이를 수정하는 작업이 필요하다.
3. 학생들의 수험번호와 매칭되는 이름을 가진 테이블이 추가적으로 필요하다.

→ 전처리 후 MySQL에 적재

데이터 MySQL에 적재 결과

OMR 테이블

	수험번호	선택과목	문제번호	대분류	중분류	소분류	학생이 쓴 답	정답	배점
0	20220001	미적분	1	수학 I	지수와 로그	로그 함수의 성질	4	4	2
1	20220001	미적분	2	수학 II	함수의 극한	함수의 극한값 계산	1	1	2
2	20220001	미적분	3	수학 I	삼각함수	삼각함수의 성질	2	2	2
3	20220001	미적분	4	수학 II	함수의 연속성	연속의 개념 및 조건	1	1	3
4	20220001	미적분	5	수학 I	수열	등차수열의 일반항	3	3	3

학생 테이블

	수험번호	이름
0	20220001	강나래
1	20220002	강도윤
2	20220003	강도현
3	20220004	강민수
4	20220005	강민지

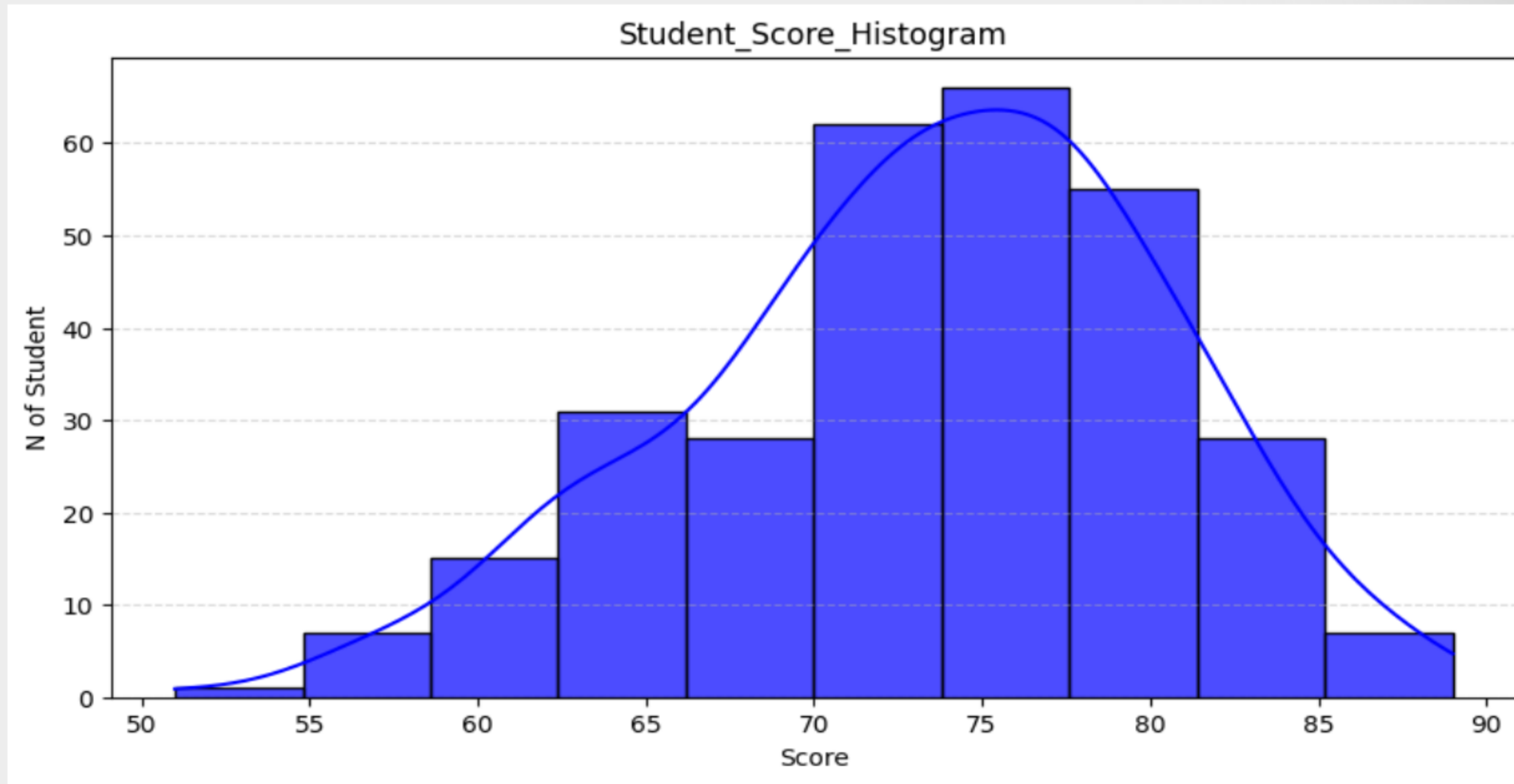
데이터 분석 및 EDA

1. MySQL에서 데이터를 불러와 학생들의 성적분포를 확인하기 위해 새로운 데이터프레임을 다음과 같이 생성

	수험번호	이름	점수	선택과목
0	20220001	강도윤	75	미적분
1	20220002	강도현	78	미적분
2	20220003	강민수	72	미적분
3	20220004	강민지	74	미적분
4	20220005	강서연	62	미적분

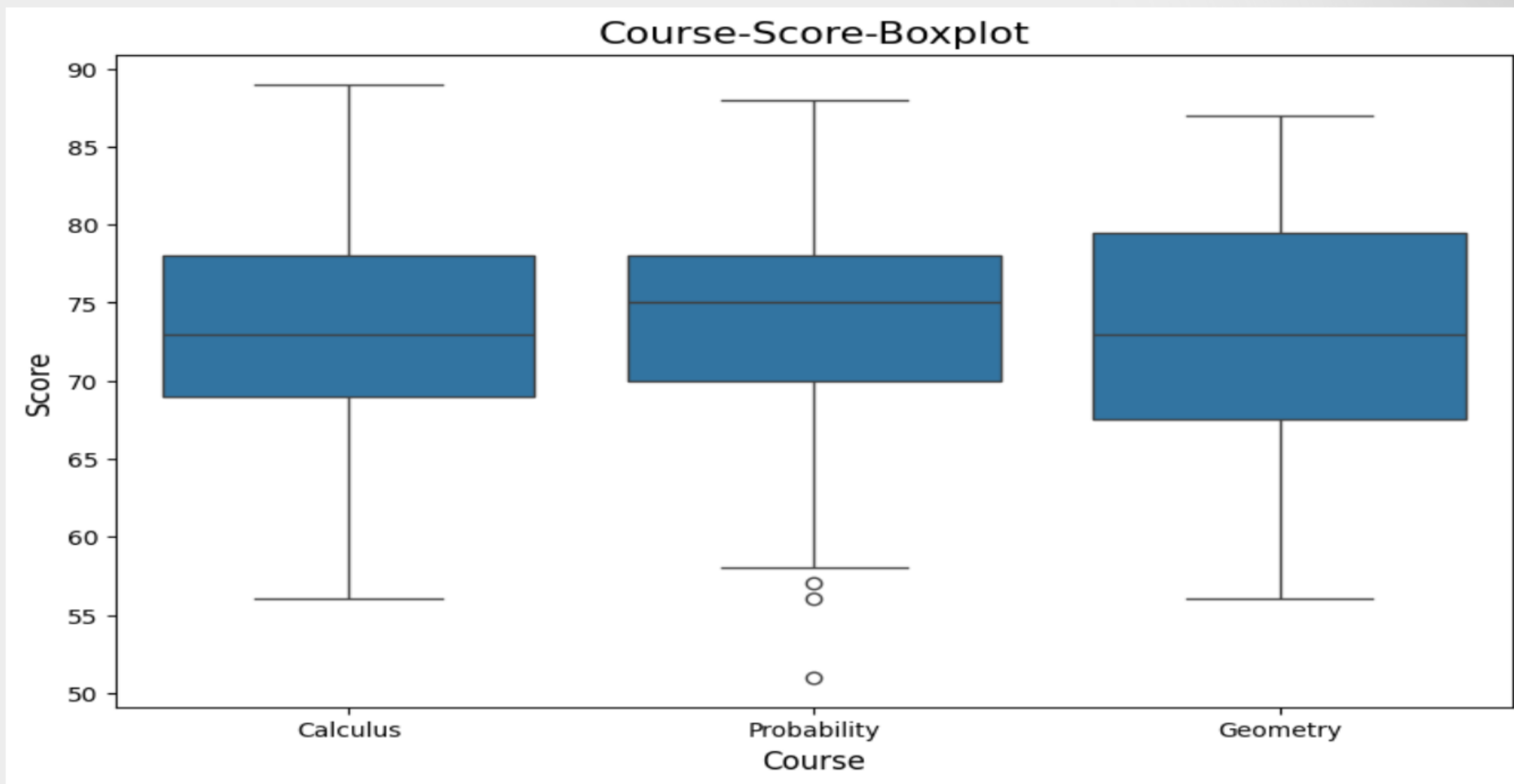
데이터 분석 및 EDA

2. 전체 학생들의 성적분포 (히스토그램)



데이터 분석 및 EDA

3. 선택과목별 학생들의 성적분포 (박스플롯)



데이터 분석 및 EDA

4. 모의고사 등급컷

선택 과목	1등급컷	2등급컷
미적분	85점	81점
기하와 벡터	87점	82점
확률과 통계	85점	82점

5. 최고 및 최저 점수

선택 과목	최고 점수	최저 점수
미적분	89점	56점
기하와 벡터	87점	56점
확률과 통계	88점	51점

데이터 분석 및 EDA

6. 문항별 난이도 TOP1, TOP2

◆ 공통과목 정답률이 낮은 문제:

✓ Top 1: 문제번호 21, 정답률 67.33%

✓ Top 2: 문제번호 22, 정답률 70.67%

◆ 선택과목: 기하와 벡터

✓ 문제번호 29, 정답률 0.00%

✓ 문제번호 30, 정답률 3.33%

◆ 선택과목: 미적분

✓ 문제번호 29, 정답률 6.00%

✓ 문제번호 30, 정답률 8.00%

◆ 선택과목: 확률과 통계

✓ 문제번호 29, 정답률 4.17%

✓ 문제번호 30, 정답률 11.67%

분석 결과 및 인사이트

실제 2023학년도 9월 모의평가 수학 1등급컷을 살펴보자면 미적분 85점, 기하와 벡터 86점, 확률과 통계 88점이다. 우리가 만든 사설 모의고사와 매우 유사한 난이도를 갖는다고 판단한다. 이후 수능대비 사설모의고사를 만들 때 현 난이도를 유지한 상태로 문제를 만들어야한다. 다만 기하와 벡터 29번처럼 정답률이 0%가 나온 문제가 있으므로 킬러문항에 대해서는 정답률이 너무 낮게 나온 문제들은 좀 더 쉽게 조절해서 출제할 필요가 있다.
