

수험생이 좋아하는 최고의 강사는!?

목차

- 1 주제 소개
- 2 파이프라인
- 3 데이터 수집 및 정제
- 4 데이터베이스설계 및 저장
- 5 데이터 분석/시각화





주제 선정 이유



코로나19로 인한 온라인 수업 및 인터넷 강의 수요 증가

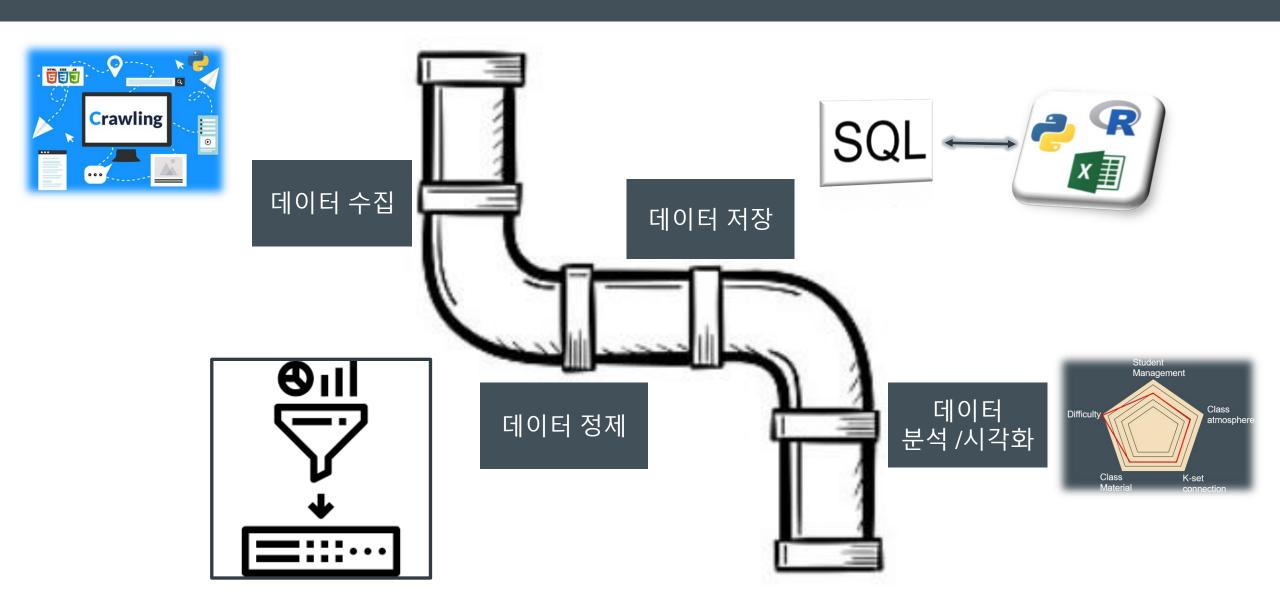
주제 소개



수험생들을 위한 인터넷 강의 선생님 추천 서비스

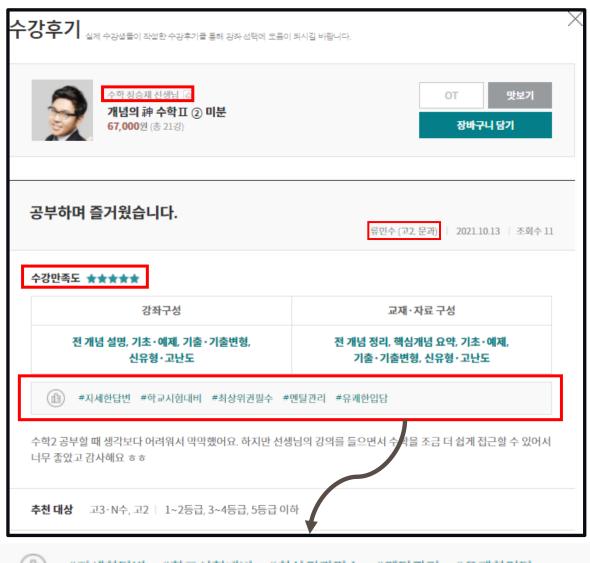


파이프라인





데이터 수집 방법 – 크롤링



데이터 수집 – 크롤링

```
wb = openpyxl.Workbook()
        sheet = wb.active
        sheet.append(["과목", "선생님이름", "학생이름", "학생학년", "문이과", "별점", "키워드(해쉬태그)"])
        print("엑셀로 변환 시작하겠습니다!!!")
        print("총 " + str(len(LINK LIST)) + "개의 데이터가 삽입될 예정입니다.")
        count = 1
        for link in LINK LIST:
            print(f"{count}번째 데이터 삽입 시작합니다.")
            html = requests.get(link)
            soup = BeautifulSoup(html.content.decode('euc-kr', 'replace'))
            lect name = soup.find(attrs={'class': 'lect name'}).text
            lect and teacher = lect name.split(" ")[0].replace(u'\xa0', u' ')
            try:
                lecture = lect_and_teacher.split(" ")[0]
            except:
                lecture = "null"
110
            try:
                teacher = lect and teacher.split(" ")[1]
```

db_etoos.py

start.py

크롤링 결과

06		: X	√ fx											
	Α	В	C	D	E	F	G	Н		J	К	L	М	N
1	과목	선생님이름		학생학년	문이과	<u>'</u> 별점	키워드(해수		'	,	IX.		141	11
2	국어	김민정	김민지	고3	문과		#자세한답	,	│ ·한슸자류 #	치상위궈픽	소 수 #깍끔하	│ ├판서 #생생	하현장감	
3	국어	김민정	이단비	고3	예체능계		#최상위권	_,		-101122	1," 202	L 1,,, 00		
4	국어	김민정	임보림	고1	계열 선택		#최상위권	,		· 하입담				
5	국어	김민정	이윤서	고3	문과		#풍부한학		_ ,		 장감 #수현	 생핔수		
6	국어	김민정	박소정	고3	예체능계		#예·복습관			,			담	
7	국어	김민정	김유빈	고3	문과								니 상위권필수	
8	국어	김민정	이소명	고3	문과			_,	,				생한현장감	
9	국어	김민정	이채린	고3	예체능계		#자세한답				_,		02202	
10	국어	김민정	문소현	고3	문과	5	#빠른답변	_, 속도.#EBS	, 완벽대비,#학	· 학교시험대!	_ 비,#최상위·	⊥ 권필수,#유:	래한입담	
11	국어	신영균	김진우	재수 이상	문과		#출제1순위		,		.,			
12	국어	김민정	노지윤	고3	 예체능계		#풍부한학	,	,					
13	국어	김민정	이의인	재수 이상	이과	5	#예·복습관	_ · , _ }리,#출제1÷	 순위정리,#3	최상위권필:	 수	_		
14	국어	김민정	박주영	재수 이상	예체능계	5	#예·복습관	· · · - · · ·리,#출제1·		최상위권필	· 수,#멘탈관	_ 리,#생생한	현장감	
15	국어	김민정	강한별	고3	이과	5	#출제1순위	위정리,#최성		#멘탈관리,	#유쾌한입[담,#생생한학	현장감	
16	국어	김민정	김솔내	null	null	5	null							
17	국어	김민정	양우빈	고3	문과	5	#출제1순위	위정리,#최성	상위권필수,	#멘탈관리, :	#유쾌한입[_ 담,#생생한학	현장감	
18	국어	김민정	문주은	재수 이상	문과	5	#빠른답변	- 속도,#출제	1순위정리,	#최상위권됨	필수,#멘탈	 관리,#생생점	한현장감	
19	국어	김민정	오채영	고3	문과	5	#자세한답	변,#예·복습	관리,#최싱	위권필수,#	·멘탈관리,#	#생생한현 정	감	
20	국어	김민정	유연진	재수 이상	문과	5	#출제1순위	위정리,#생성	생한현장감					
21	국어	김민정	신진호	고2	문과	5	#출제1순위	위정리,#최성	상위권필수,	#멘탈관리, :	#유쾌한입[담,#생생한학	현장감	
22	국어	김민정	임소영	고1	계열 선택?	5	#예·복습관	·리,#학교시	험대비,#낄	끔한판서				
23	국어	김민정	곽은진	재수 이상	문과	5	#유쾌한입	담,#생생한	현장감					
2/	구어	기미전	전이지	.13	에체느계	5	#치산의귀	핀수 #메탄	과리 #으긞	하인단 #새	새하혀짓기	+		



데이터 정제

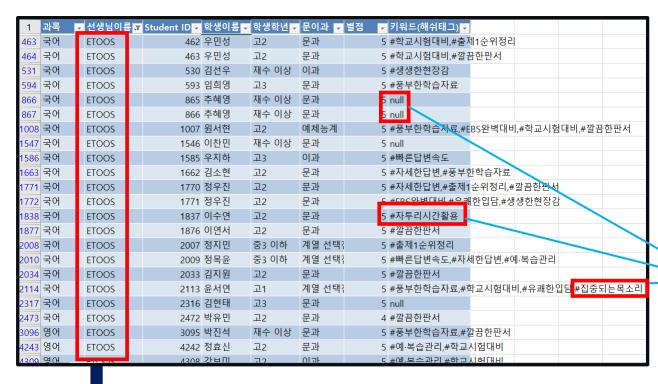


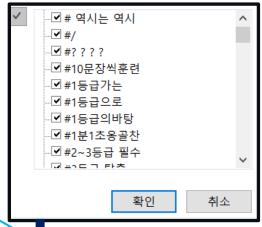


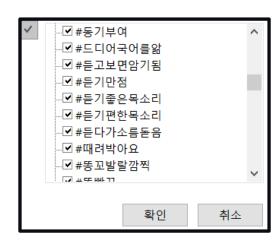
1	과목 ▼	선생님이름	Student ID -	학생이름▼	학생학년 -	문이과 .	별점	~
16	국어	김민정	15	김솔내				5
34	국어	김민정	33	전다빈				5
60	국어	김민정	59	이상윤				5
84	국어	김민정	83	이수빈				5
202	국어	신영균	201	jaeha643				5

null 로 입력된 데이터 <mark>빈칸</mark>으로 변경

데이터 정제



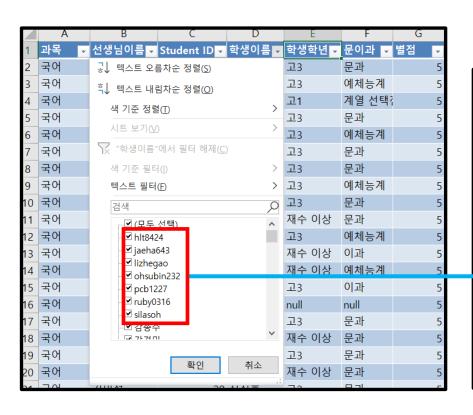


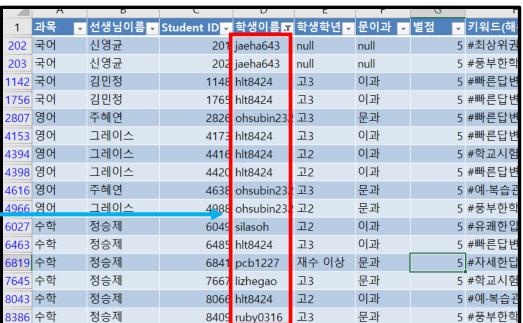


· <u>지정한 키워드 11개에서 벗어나는 키워드들과</u> null 로 입력된 값은 빈칸(null)으로 변경

선생님 이름이 ETOOS로 정의된 데이터 제외

데이터 정제





학생이름이 아이디로 입력된 데이터 제외



관계 설정하기

원시 데이터 과목, 선생님 이름, 학생이름, 학생학년, 문이과, 별점, 키워드 1,2,3,4,5

	A	В	С	D	E	F	G	Н		J	K	L	
1	student_id			student_na ▼	student_g ~	student_major 🔻			▼ keyoword2	▼ keyword3	▼ keyword4	▼ keyword5	_
2		1 국어	김민정	김민지	고3	문과		5 #자세한답변	#풍부한학습자료	#최상위권필수	#깔끔한판서	#생생한현장감	
3		2 국어	김민정	이단비	고3	예체능계		5 #최상위권필수	#생생한현장감				
4	;	국어	김민정	임보림	고1	계열 선택전(고1)		5 #최상위권필수	#멘탈관리	#유쾌한입담			
5		4 국어	김민정	박소정	⊒3	예체능계		5 #예·복습관리	#EBS완벽대비	#최상위권필수	#멘탈관리	#유쾌한입담	
6		국어	김민정	김유빈	⊒3	문과		5 #자세한답변	#예·복습관리	#풍부한학습자료	#출제1순위정리	#최상위권필수	
7	(5 국어	김민정	이소명	⊒3	문과		5 #빠른답변속도	#자세한답변	#풍부한학습자료	#최상위권필수	#생생한현장감	
8		7 국어	김민정	이채린	⊒3	예체능계		5 #자세한답변	#예·복습관리	#유쾌한입담			
9		3 국어	신영균	김진우	재수 이상	문과		5 #출제1순위정리	#멘탈관리	#유쾌한입담	#깔끔한판서	#생생한현장감	
10		국어	김민정	노지윤	⊒3	예체능계		5 #풍부한학습자료	#멘탈관리	#유쾌한입담	#생생한현장감		
11	10	국어	김민정	이의인	재수 이상	이과		5 #예·복습관리	#출제1순위정리	#최상위권필수			
12	11	1 국어	김민정	박주영	재수 이상	예체능계		5 #예·복습관리	#출제1순위정리	#최상위권필수	#멘탈관리	#생생한현장감	
13	12	2 국어	김민정	강한별	고3	이과		5 #출제1순위정리	#최상위권필수	#멘탈관리	#유쾌한입담	#생생한현장감	
14	13	국어	김민정	김솔내				5					
15	14	4 국어	김민정	양우빈	⊒3	문과		5 #출제1순위정리	#최상위권필수	#멘탈관리	#유쾌한입담	#생생한현장감	
16	1:	국어	김민정	문주은	재수 이상	문과		5 #빠른답변속도	#출제1순위정리	#최상위권필수	#멘탈관리	#생생한현장감	
17	10	5 국어	김민정	오채영	고3	문과		5 #자세한답변	#예·복습관리	#최상위권필수	#멘탈관리	#생생한현장감	
18	1	7 국어	김민정	유연진	재수 이상	문과		5 #출제1순위정리	#생생한현장감				
19	18	3 국어	김민정	신진호	고2	문과		5 #출제1순위정리	#최상위권필수	#멘탈관리	#유쾌한입담	#생생한현장감	
20	19	국어	김민정	곽은진	재수 이상	문과		5 #유쾌한입담	#생생한현장감				
21	20	국어	김민정	정의진	고3	예체능계		5 #최상위권필수	#멘탈관리	#유쾌한입담	#생생한현장감		

Redundancy를 최소화하기 위해서 테이블 분리 → 관계설정 ※ 엑셀을 사용하여 테이블 분리

테이블 분리

Student 테이블

Student_id (PK), student_name, keyword_id (FK), teacher_name (FK), info_i(FK)

	Α	В	С	D	E
1	student_id 🔻	teacher_nai 🕶	student_na 🔻	keyword_i(info_id 🔽
2	1	김민정	김민지	1	1
3	2	김민정	이단비	2	2
4	3	김민정	임보림	3	3
5	4	김민정	박소정	4	2
6	5	김민정	김유빈	5	1
7	6	김민정	이소명	6	1
8	7	김민정	이채린	7	2
9	8	신영균	김진우	8	4
10	9	김민정	노지윤	9	2
11	10	김민정	이의인	10	5
12	11	김민정	박주영	11	6
13	12	김민정	강한별	12	7

Student_info 테이블

info_id (PK), 학생 학년, 문·이과

	Α	В	С
1	info_id 🔽	student_g ▼	student_major_
3	1	고3	문과
	2	고3	예체능계
4	3	고1	계열 선택전(고1)
5	4	재수 이상	문과
6 7	5	재수 이상	이과
	6	재수 이상	예체능계
8	7	고3	이과
9	8		
10	9	고2	문과
11	10	고2	이과
12	11	재수 이상	전문계
13	12	고1	이과

테이블 분리

Teacher 테이블

teacher_name (PK), subject



Keyword 테이블

keyword_id (PK), keyword1, keyword2, keyword3, keyword4, keyword5, rating

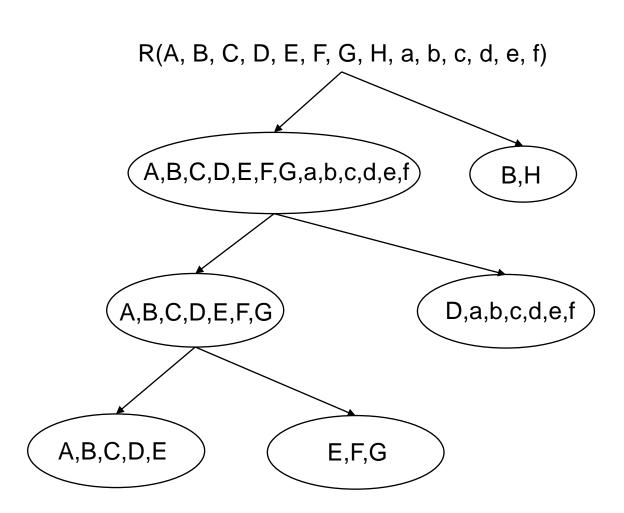
	Α	В	С	D	E	F	G	
1	keyword_id 🔽	keyword1	keyword2	keyword3	keyword4	keyword5 🔻	rating	*
2	1	#자세한답변	#풍부한학습자료	#최상위권필수	#깔끔한판서	#생생한현장감		5
3	2	#최상위권필수	#생생한현장감					5
4	3	#최상위권필수	#멘탈관리	#유쾌한입담				5
5	4	#예·복습관리	#EBS완벽대비	#최상위권필수	#멘탈관리	#유쾌한입담		5
6	5	#자세한답변	#예·복습관리	#풍부한학습자료	#출제1순위정리	#최상위권필수		5
7	6	#빠른답변속도	#자세한답변	#풍부한학습자료	#최상위권필수	#생생한현장감		5
8	7	#자세한답변	#예·복습관리	#유쾌한입담				5
9	8	#출제1순위정리	#멘탈관리	#유쾌한입담	#깔끔한판서	#생생한현장감		5
10	9	#풍부한학습자료	#멘탈관리	#유쾌한입담	#생생한현장감			5
11	10	#예·복습관리	#출제1순위정리	#최상위권필수				5
12	11	#예·복습관리	#출제1순위정리	#최상위권필수	#멘탈관리	#생생한현장감		5
13	12	#출제1순위정리	#최상위권필수	#멘탈관리	#유쾌한입담	#생생한현장감		5

- ① {Student_id} → {name}
- ② {Student_info_id} → {grade, major}
- ③ {Teacher_name} → {subject}
- (4) {keyword_id} → {keyword1, keyword2, keyword3, keyword4, keyword5, rating}

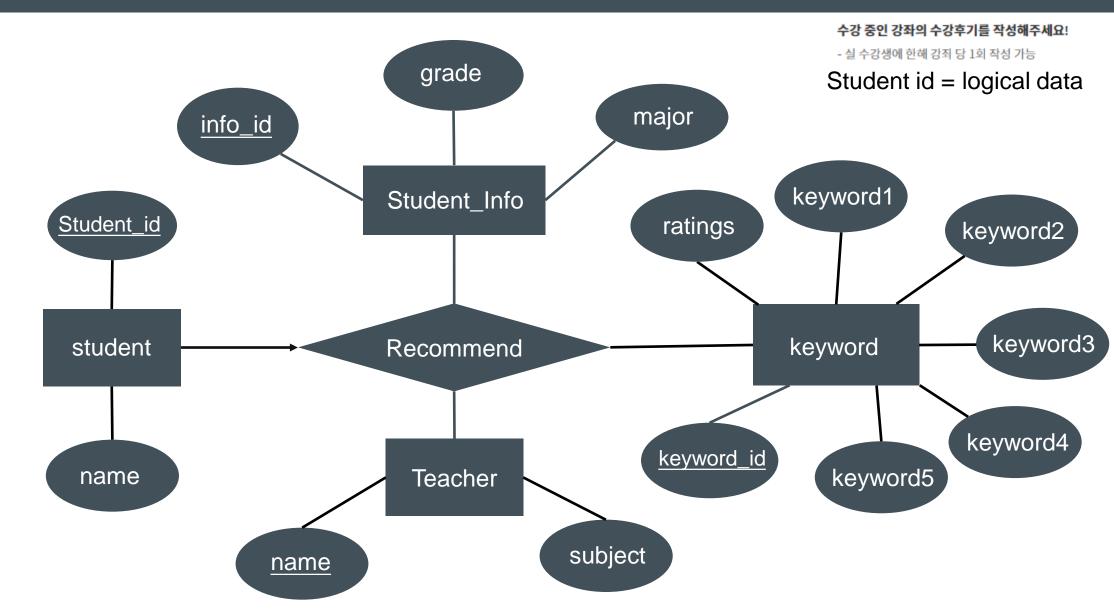
normalization

```
student_id = A
teacher_name = B
student_name = C
keyword_id = D
info id = E
student_grade = F
student_major = G
subject = H
keyword1 = a
keyword2 = b
keyword3 = c
keyword4 = d
keyword5 = e
rating = f
```

R(A, B, C, D, E, F, G, H, a, b, c, d, e, f)
A → B, C, D, E, F, G, H, a, b, c, d, e, f
B → H
D → a, b, c, d, e, f
E → F, G



ER model



Relational schema

ETOOS _Teacher

subject
-
-

ETOOS _Student

student_ID	student_name	Info_ID	Keyword_ID	Teacher_name
-	-	-	-	
-	-	-	-	

ETOOS_Keyword

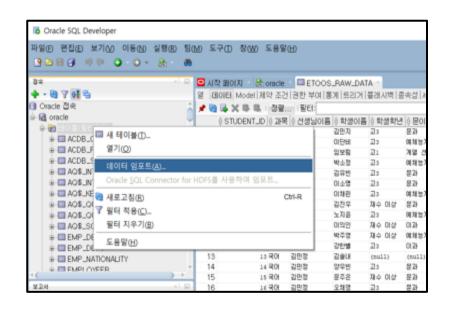
keyword_ID	Keyword1	Keyword2	keyword3	keyword4	keyword5	ratings
-	-	-	-	-	-	-
-	-	-	-	-	-	-

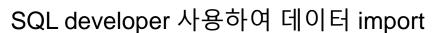
ETOOS _Student_Info

<u>Info_ID</u>	Student_grade	student_major
-	-	
-	-	



데이터 저장





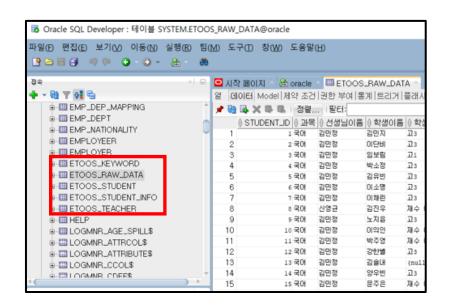
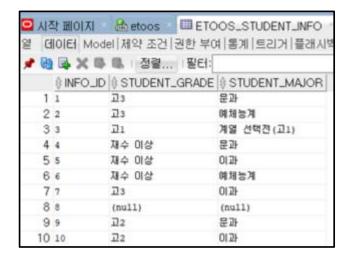


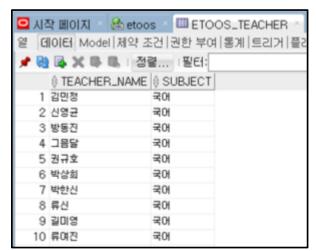
Table 4개 생성

데이터 저장

0	시작	막 페이지 🏻 🦀	etoos × ETOOS	S_STUDENT ×		
열	G	IOI터 Model 제	약 조건 권한 부여 등	통계 트리거 플래시백	[종속성 세부정5	리분할 영역
×	1	BX BB	정렬 밀터:			
		♦ STUDENT_ID	⊕ TEACHER_NAME			
	1	1	김민정	김민지	1	1
	2	2	김민정	이단비	2	2
	3	3	김민정	임보림	3	3
	4	4	김민정	박소정	4	2
	5	5	김민정	김유빈	5	1
2	6	6	김민정	이소명	6	1
	7	7	김민정	이채린	7	2
V.	8	8	신영균	김진우	8	4
	9	9	김민정	노지윤	9	2
	10	10	김민정	이익인	10	5
			310174	WITO		-





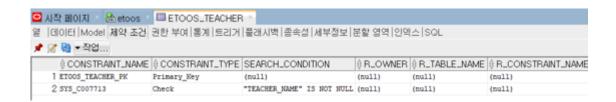


데이터 저장



☑ 시작 페이지 🏻 🔠 etoos 🦜	ETOOS_STUDENT	LINFO *							
열 [데이터 Model 제약 조건]	볼 데이터 Model 제약 조건 권한 부여 통계 트리거 플래시백 종속성 세부정보 분할 영역 인덱스 SQL								
★ 37 HB ▼작업									
	⊕ CONSTRAINT_TYPE	SEARCH_CONDITION	() R_OWNER	() R_TABLE_NAME	⊕ R_CONSTRAINT_NAME				
1 ETOOS_STUDENT_INFO_PK	Primary_Key	(null)	(null)	(null)	(null)				
2 SYS_C007715	Check	"INFO_ID" IS NOT NULL	(null)	(null)	(null)				





Student 테이블에 Foreign key로 Info_id, Keyword_id, Teacher_name를 각각의 테이블에서 참조

FD 검증

FD(1)

SQL> select student_id 2 from etoos_raw_data 3 group by student_id 4 having count(학생이름)>1; 선택된 레코드가 없습니다. FD(2)

```
SQL> select info_id
2 from etoos_student_info
3 group by info_id
4 having count(학년)>1 and count(전공)>1;
선택된 레코드가 없습니다.
```

FD 검증

FD(3)

```
SQL> select 선생님이름
2 from etoos_teacher
3 group by 선생님이름
4 having count(과목)>1;
선택된 레코드가 없습니다.
```

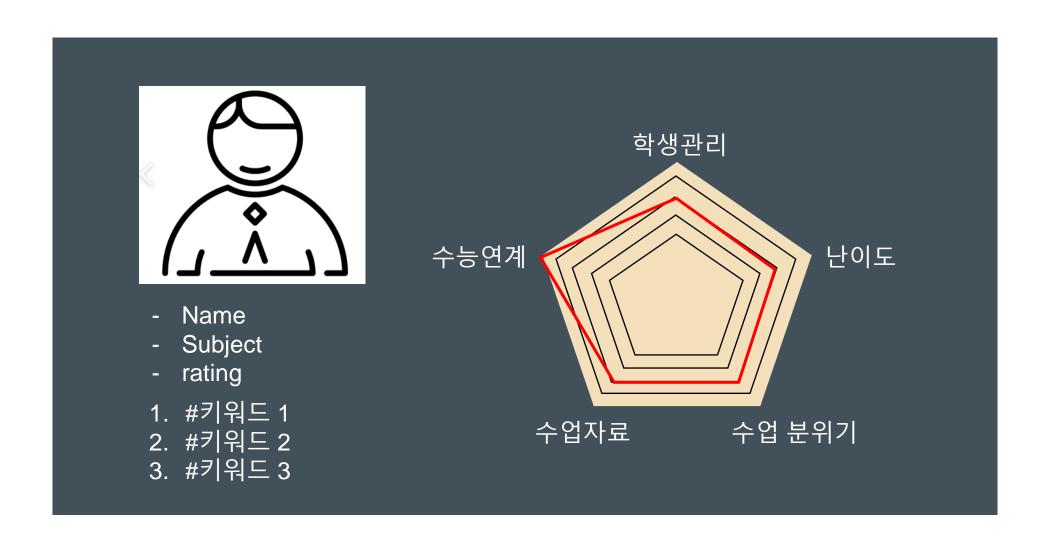
FD(4)

```
SQL> select keyword_id
2 from etoos_keyword
3 group by keyword_id
4 having count(keyword1)>1
5 and count(keyword2)>1
6 and count(keyword3)>1
7 and count(keyword4)>1
8 and count(rating)>1;
선택된 레코드가 없습니다.
```

Part 7 데이터 분석



데이터분석



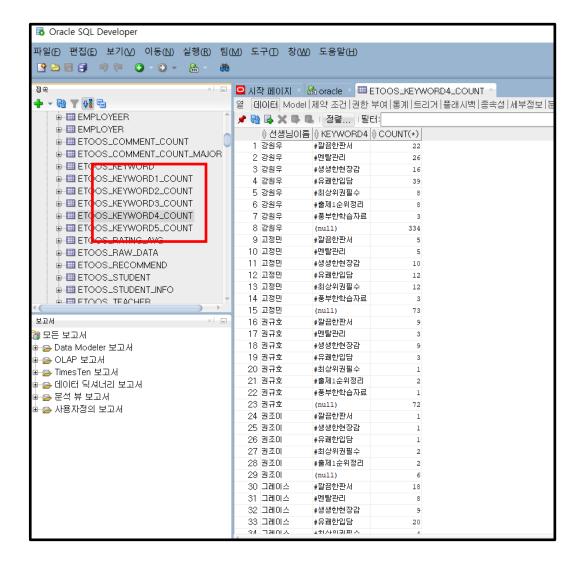
학생 관리	#빠른 답변속도, #자세한답변, #예 복습관리, #멘탈 관리
난이도	#최상위권필수
수업 분위기	#유쾌한 입담, #생생한 현장감, #깔끔한 판서
수업자료	#풍부한 학습자료
수능연계	#EBS완벽대비, #출제 1순위 정리

데이터 분석

```
오각형 수치화 기준
기준 : 구간[0,1]
student management = value / (4 x 전체댓글 수)
difficulty = value / (1 x 전체댓글 수)
class atmosphere = value / (3 x 전체 댓글 수)
class material = value / (1 x 전체 댓글 수)
K-set connection = value / (2 x 전체 댓글 수)
```

데이터 분석

```
SQL> create table etoos_keyword1_count as select * from(
2 select s.선생님이름,k.keyword1,count(*)
3 from etoos_student s ,etoos_keyword k
4 where s.keyword_id = k.keyword_id
5 group by k.keyword1,s.선생님이름
6 order by 선생님이름);
테이블이 생성되었습니다.
```



데이터분석

		Α	В	С
	1	선생님이론	KEYWORD	COUNT(*)
	2	강원우	#EBS완벽[11
	3	강원우	#깔끔한판	8
	4	강원우	#멘탈관리	15
	5	강원우	#빠른답변	54
	6	강원우	#생생한현	11
	7	강원우	#예·복습관	95
	8	강원우	#유쾌한입	10
	9	강원우	#자세한답	49
	10	강원우	#최상위권	17
	11	강원우	#출제1순위	38
	12	강원우	#풍부한학	54
	13	강원우		94
	14	고정민	#깔끔한판	2
	15	고정민	#빠른답변	28
	16	고정민	#생생한현	2
	17	고정민	#예·복습관	23
	18	고정민	#유쾌한입	1
	19	고정민	#자세한답	17
	20	고정민	#최상위권	8
	21	고정민	#출제1순위	11
	22	고정민	#풍부한학	15
	23	고정민		13
L	24	권규ㅎ	#EBC라벼[2

	А	В	С	
1	선생님이론	KEYWORD	COUNT(*)	
2	강원우	#EBS완벽[5	
3	강원우	#깔끔한판	21	
4	강원우	#멘탈관리	21	
5	강원우	#생생한현	20	
6	강원우	#예·복습관	37	
7	강원우	#유쾌한입	40	
8	강원우	#자세한답	21	
9	강원우	#최상위권	34	
10	강원우	#출제1순위	53	
11	강원우	#풍부한학	47	
12	강원우		157	
13	고정민	#EBS완벽[6	
14	고정민	#깔끔한판	2	
15	고정민	#멘탈관리	1	
16	고정민	#생생한현	4	
17	고정민	#예·복습관	15	
18	고정민	#유쾌한입	7	
19	고정민	#자세한답	10	
20	고정민	#최상위권	15	
21	고정민	#출제1순위	17	
22	고정민	#풍부한학	18	
23	고정민		25	

	Α	В	С	
1	선생님이를	KEYWORD	COUNT(*)	
2	강원우	#EBS완벽[6	
3	강원우	#깔끔한판	13	
4	강원우	#멘탈관리	34	
5	강원우	#빠른속도	1	
6	강원우	#생생한현	36	
7	강원우	#예·복습관	9	
8	강원우	#유쾌한입	32	
9	강원우	#최상위권	40	
10	강원우	#출제1순위	26	
11	강원우	#풍부한학	14	
12	강원우		245	
13	고정민	#깔끔한판	8	
14	고정민	#멘탈관리	2	
15	고정민	#생생한현	4	
16	고정민	#예·복습관	4	
17	고정민	#유쾌한입	13	
18	고정민	#최상위권	22	
19	고정민	#출제1순위	16	
20	고정민	#풍부한학	6	
21	고정민		45	
22	권규호	#EBS완벽[3	
23	권규호	#깔끔한판	11	

	Α	В	С	
1	선생님이론	KEYWORD	COUNT(*)	
2	강원우	#깔끔한판	22	
3	강원우	#멘탈관리	26	
4	강원우	#생생한현	16	
5	강원우	#유쾌한입	39	
6	강원우	#최상위권	8	
7	강원우	#출제1순위	8	
8	강원우	#풍부한학	3	
9	강원우		334	
10	고정민	#깔끔한판	5	
11	고정민	#멘탈관리	5	
12	고정민	#생생한현	10	
13	고정민	#유쾌한입	12	
14	고정민	#최상위권	12	
15	고정민	#풍부한학	3	
16	고정민		73	
17	권규호	#깔끔한판	9	
18	권규호	#멘탈관리	3	
19	권규호	#생생한현	9	
20	권규호	#유쾌한입	3	
21	권규호	#최상위권	1	
22	권규호	#출제1순위	2	
23	권규호	#풍부한학	1	
~ .	귀그ㅎ		70	

	Α	В	С	
1	선생님이론	KEYWORD	COUNT(*)	
2	강원우	#깔끔한판	13	
3	강원우	#멘탈관리	4	
4	강원우	#생생한현	35	
5	강원우	#유쾌한입	11	
6	강원우	#최상위권	2	
7	강원우	#출제1순위	1	
8	강원우		390	
9	고정민	#깔끔한판	9	
10	고정민	#멘탈관리	1	
11	고정민	#생생한현	7	
12	고정민	#유쾌한입	5	
13	고정민	#최상위권	3	
14	고정민		95	
15	권규호	#깔끔한판	1	
16	권규호	#생생한현	8	
17	권규호	#유쾌한입	2	
18	권규호	#최상위권	2	
19	권규호	#출제1순위	1	
20	권규호		86	
21	권조이	#깔끔한판	1	
22	권조이	#생생한현	2	
23	권조이	#최상위권	2	

keyword1 keyword2 keyword3 keyword4 keyword5

데이터분석

```
SQL> select b.선생님이름,a.과목,count(*)
2 from etoos_teacher a , etoos_student b
3 where a.선생님이름 = b.선생님이름
4 group by b.선생님이름,a.과목
5 order by a.과목,b.선생님이름;
 선생님이름
과목
                                                                                              COUNT(*)
 선생님이름
과목
                                                                                             COUNT(*)
                                                                                                    1448
 선생님이름
과목
                                                                                              COUNT(*)
```

데이터분석

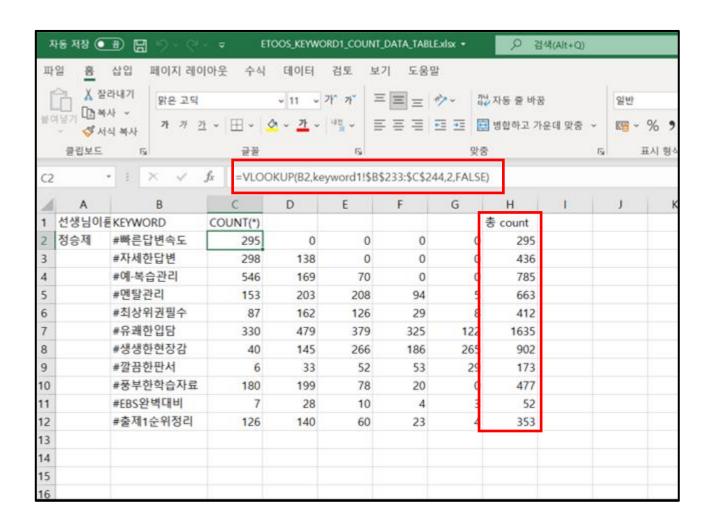
SQL> select b.선생님이름,a.과목,count(*) 2 from etoos_teacher a , etoos_student b 3 where a.선생님이름 = b.선생님이름 4 group by b.선생님이름,a.과목 5 having count(*)>50 6 order by a.과목,b.선생님이름;	
선생님이름	
과목	COUNT(*)
권규호 국어	100
그믐달 국어	156
김민정 국어	1448
선생님이름	
과목	COUNT(*)
박상희 국어	159
방동진 국어	111
신영균 국어	162
선생님이름	
과목	COUNT(*)
고정민 수학	120
이미지 수학	62
이하영 수학	132

이하영 수학	132
선생님이름	
 과목	COUNT(*)
	2515
차현우 수학	69
한정윤 수학	60
선생님이름	
	COUNT(*)
- 강원우 영어	456
그레이스 영어	261
윤훈관 영어	104
선생님이름 	
과목	COUNT(*)
전 홍철 영어	493
주혜연 영어	494
17 행이 선택되었습니다.	

Having count(*)>50

→ 댓글 50개 이하인 선생님은 제외되도록 추출

데이터분석



데이터분석

	Α	В	C	D	E	F	G	H	1
	선생님이	€ KEYWORD	COUNT(*)					총 count	
	주혜연	#빠른답변속도	52	0	0	0	0	52	
		#자세한답변	77	32	0	0	0	109	
		#예·복습관리	115	35	16	0	0	166	
		#멘탈관리	38	62	79	40	6	225	
		#최상위권필수	13	26	-11	7	2	59	
		#유쾌한입담	9	35	41	46	11	142	
		#생생한현장감	0	2	5	7	14	28	
		#깔끔한판서	13	45	65	55	64	242	
		#풍부한학습자료	53	60	19	2	0	134	
		#EBS완벽대비	23	25	13	4	0	65	
		#출제1순위정리	18	33	22	7	0	80	
			학생 관리	난이도	수업 분위기	수업 자료	수능 연계		댓글 수
		빈도수 총 합	552	59	412	134	145		494
		정규화된 값	0.279	0.119	0.278	0.271	0.147		
i							1		
			학생 관리	난이도	수업 분위기	수업 자료	수능 연계		
			0.279	0.119	0.278	0.271	0.147		

E1	5 \	$ \cdot \times \sqrt{f_x} = \E	\$14/(\$I\$14*3)						
	Α	В	С	D	Е	F	G	Н	1
1	선생님이	E KEYWORD	COUNT(*)					총 count	
2	주혜연	#빠른답변속도	52	0	0	0	0	52	
3		#자세한답변	77	32	0	0	0	109	
4		#예·복습관리	115	35	16	0	0	166	
5		#멘탈관리	38	62	79	40	6	225	
6		#최상위권필수	13	26	11	7	2	59	
7		#유쾌한입담	9	35	41	46	11	142	
8		#생생한현장감	0	2	5	7	14	28	
9		#깔끔한판서	13	45	65	55	64	242	
0		#풍부한학습자료	53	60	19	2	0	134	
1		#EBS완벽대비	23	25	13	4	0	65	
2		#출제1순위정리	18	33	22	7	0	80	
3			학생 관리	난이도	수업 분위기	수업 자료	수능 연계		댓글 수
4		빈도수 총 합	552	59	412	134	145		49
5		정규화된 값	0.279	0.11	0.278	0.271	0.147	1	
6									
7			학생 관리	난이도	수업 분위기	수업 자료	수능 연계		
8			0.279	0.119	0.278	0.271	0.147		

	Α	В	С	D	E	F	G	H	1
1	선생님이	E KEYWORD	COUNT(*)					총 count	
2	주혜연	#빠른답변속도	52		0 0	0	0	52	
3		#자세한답변	77		32 0	0	0	109	
4		#예·복습관리	115		35 16	0	0	166	
5		#멘탈관리	38		52 79	40	6	225	
6		#최상위권필수	13		26 11	7	2	59	
7		#유쾌한입담	9		35 41	46	11	142	
В		#생생한현장감	0		2 5	7	14	28	
9		#깔끔한판서	13		15 65	55	64	242	
0		#풍부한학습자료	53		50 19	2	0	134	
1		#EBS완벽대비	23		25 13	4	0	65	
2		#출제1순위정리	18		33 22	7	0	80	
3			학생 관리	난이도	수업 분위기	수업 자료	수능 연계		댓글 수
4		빈도수 총 합	552		9 412	134	145		49
5		정규화된 값	0.27	0.1	0.278	0.271	0.147		
6									
7			학생 관리	난이도	수업 분위기	수업 자료	수능 연계		
8			0.279	0.1	19 0.278	0.271	0.147		

분석과정

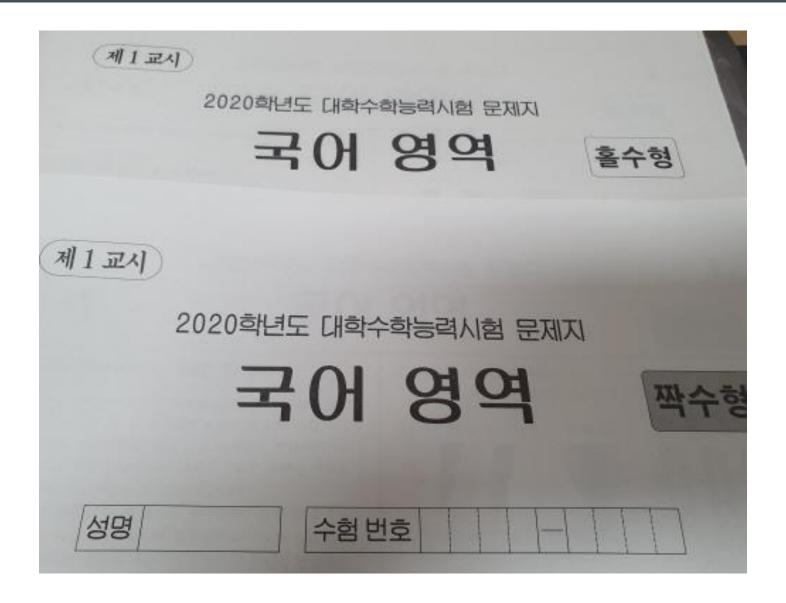
SQL> select a.선생님이름,avg(rating) 2 from etoos_student a , etoos_keyword b 3 where a.keyword_id = b.keyword_id 4 group by a.선생님이름;

선생님 평균 별점 데이터 추출

선생님이름	AVG(RATING)
길미영 강원우 정지안 윤훈관 류십	4.94117647 4.98026316 4.97727273 4.97115385 5
서현아 이하영 김민정 김예림	5 4.99242424 4.98895028 5
천홍철 최원규	4.97971602 5
선생님이름	AVG(RATING)
정승제 한정윤 고정민 김보건 방통지	4.98926441 5 5 5 5 4.97297297
원유호 권조이 이세호	4.97
^{지에도} 최지수 정현경 남궁율	5 5 5 5
선생님이름	AVG(RATING)
신영균 그름달	4.97530864 5
그레이스 유영태	4.96551724 4 5
막하다 이주영 <u>의</u> 밀지	5 4.9516129
사연주 박상희 박한신 류여진	5 4.99371069 5 5
선생님이름	AVG(RATING)
이정훈 주혜연 이현로 이장규	4.36363636 4.99190283 4.97058824 4.8
37 행이 선택되었습니다.	l l



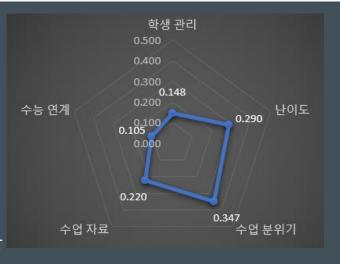
데이터시각화





권규호 / 국어 ★ 4.97

- l. #생생한 현장감
- 2. #깔끔한 판서
- 3. #유쾌한 입담





김민정 / 국어 ★ 4.99

- 1. #유쾌한 입담
- 2. #최상위권 필수
- 3. #생생한 현장감





그믐달 / 국어 ★ 5

- 1. #풍부한 학습자료
- 2. #최상위권 필수
- 3. #유쾌한 입담





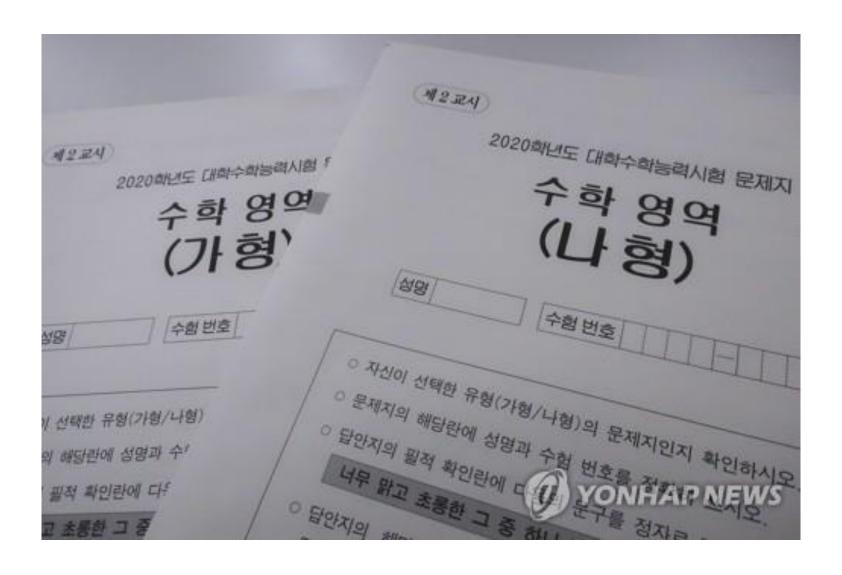
방동진 / 국어 ★ 4.97

- 1. #깔끔한 판서
- 2. # 예·복습관리
- 3. #유쾌한 입담





데이터시각화



#예·복습관리





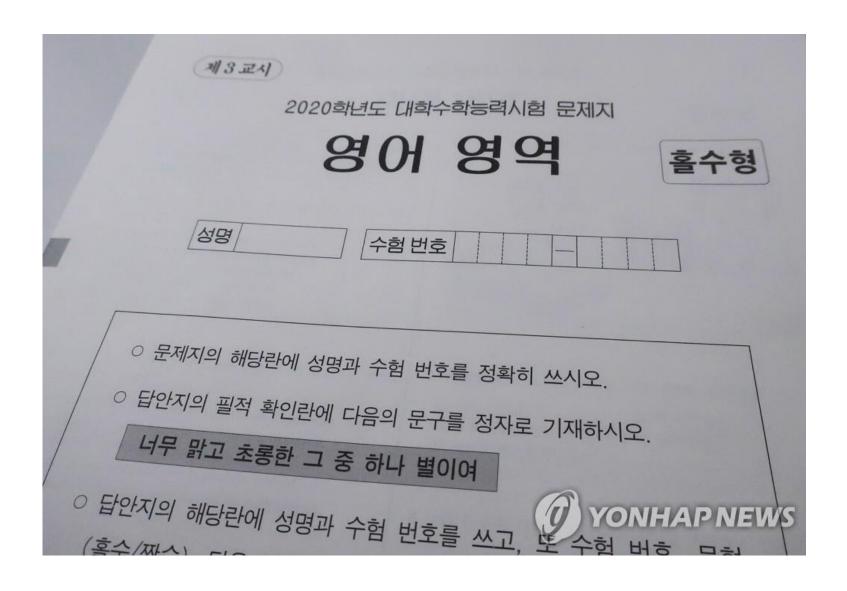








데이터 시각화





강원우 / 영어 ★ 4.98

- 1. #예·복습관리
- 2. #유쾌한 입담
- 3. #출제1순위정리





윤훈관 / 영어 ★ 4.97

- 1. #깔끔한 판서
- 2. #유쾌한 입담
- 3. #최상위권 필수









팀 프로젝트를 마무리하며...

이투스 뿐 아닌 다른 인강사이트와의 데이터를 연결시켜 새로운 정보를 도출해낼 수 있지 않을까 했습니다.

추출한 데이터 와 다른 고도의 분석기법을 통해 다른 유의미한 정보를 얻고 싶었지만 한계가 있었습니다.

저희가 저장한 테이블을 사용하여, SQL을 통해 시각화 하는 과정 중 선생님 별 키워드 빈도 수를 뽑아내는 과정이 힘들었습니다. Keyword 1,2,3,4,5 별로 빈도수를 먼저 구하고, 엑셀에서 합치는 과정에서 번거로움을 느껴. 처음부터 데이터 분석에 용이한 테이블을 만들었다면 더 편리하지 않았을까 하는 아쉬움이 있었습니다.

선생님 추천카드에 포함되는 별점을 통해 선생님들 간의 차이를 확인하고 싶었지만, 분석 해본 결과 모두 5점에 가까운 점<u>수를 받아 유의미한 결과를 얻어내는데 실패해 아쉬웠습니다.</u>