

고등학생 식단 점검 및 추천 시스템

— 2019.12.14 —

2017204002 김영서
2017204016 이지환
2017204068 김동영



1 주제 소개



일주일치 식단(급식+간식),
몸무게, 키, 나이, 성별,
영양소 권장량, 알레르기



하루 식단 점검,
일주일 식단 총평,
식단 추천

**Highschool student diet evaluation and
recommendation-system**

1 주제 소개



고등학생을 대상으로 선정

-> 성장을 위한 영양소 섭취가
중요한 나이

인체 치수 데이터.csv



급식 식단을 데이터로 선정

-> 실제 식단 데이터를
얻기가 어려워, 얻기 쉬운
급식 식단 데이터 선정
-> 영양소 정보가 잘 정리 되어 있음

급식식단 데이터.csv



간식 데이터 추가

-> 급식 이외에도
영양소 섭취를 하기 때문에
현실성 고려

간식 데이터.csv

1) 인체 치수 데이터

- 국가통계 마이크로데이터 <https://mdis.kostat.go.kr/index.do>

인구·가구 고용·임금 물가·가계 보건·복지 사회 지자체사회조사 농림어업 광공업·에너지 건설·주택·토지 교통·정보통신 도소매·서비스 경기·기업경영 교육·문화 과학·환경 교육실습용	>	급성심장정지조사 녹색생활조사 보육실태조사 사망원인보안조사 사망원인통계 사회서비스수요·공급실태조사 산후조리실태조사 여주시복지실태조사 영덕군노인실태조사 의료기기제조유통조사 의료서비스경험조사 장애인생활체육실태조사 한국인인체치수조사 한부모가족실태조사 한의약산업실태조사	>	총괄(제공)	<input checked="" type="checkbox"/> 2015 <input type="checkbox"/> 2010
-----------------------------------------------------------------------------------------------------------------------------------------------------	---	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---	--------	---------------------------------------------------------------------------

보건/복지 항목에 '한국인인체치수조사' 2015년 데이터 다운

2 구성요소

1) 인체 치수 데이터

- 국가통계 마이크로데이터 <https://mdis.kostat.go.kr/index.do>

	A	B	C	D	E	F	G	H	I
1	성별	골격근량	체지방량	BMI	체지방율	복부지방율	몸무게	나이	키
2	1	21.7	6	16.95419	13.2	0.7	45.6	17	164
3	1	22.1	7	16.83558	14.7	0.7	47.8	17	168.5
4	1	22.3	6.7	17.79882	14.3	0.8	47	17	162.5
5	1	22.4	5.9	18.05486	12.7	0.8	46.8	17	161
6	1	22.5	6.2	16.5024	13.2	0.7	47.3	17	169.3
7	1	22.6	8	17.89867	16.3	0.8	49.5	17	166.3
8	1	22.7	8.6	18.66659	17.2	0.8	49.9	17	163.5
9	1	22.8	3.5	17.15937	7.8	0.8	45.2	17	162.3
10	1	22.9	6.9	16.70823	14.1	0.7	49.2	17	171.6

1405 row x 9 column

고등학생(나이 17세~19세)인 데이터만 선택하여
'고등학생 인체치수.csv'로 저장

2-1) 급식 식단 데이터

- 나이스 교육정보 개방 포털 <https://open.neis.go.kr/portal/mainPage.do>

검색하기 ▲

· 시도교육청코드 서울특별시교육청 ▼

· 학교명 고등학교

· 식사명

· 급식일자 ~

검색

* *최대 1 만건 까지만 화면에 조회되고 다운로드 됩니다* 검색조건에 따른 총 검색건수 (93,196) 건

· 파일변환저장 XLS CSV JSON XML TXT

번호	시도교육청코드	시도교육청명	표준학교코드	학교명	식사코드	식사명	급식일자	급식인원수	요리명
1	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20200205	570	토마토스파게티
2	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20200204	570	칼슘기장밥k L
3	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20200203	570	콩나물무밥&달
4	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191230	570	칼슘찐쌀밥k L
5	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191227	570	칼슘흑미밥k L
6	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191226	570	칼슘기장밥k L
7	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191224	570	맹초고기주먹밥
8	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191223	570	나물비빔밥k5
9	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191220	561	칼슘현미밥k L
10	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191219	561	칼슘15곡혼합
11	B10	서울특별시교육청	7010057	가락고등학교	2	중식	20191218	561	추가밥k L

ChromeDriver, Selenium 모듈을 이용하여
각 시도교육청의 고등학교 '급식식단정보' 데이터 다운

2 구성요소

2-1) 급식 식단 데이터

- 나이스 교육정보 개방 포털 <https://open.neis.go.kr/portal/mainPage.do>

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		시도교육청	시도교육청	표준학교코드	학교명	식사코드	식사명	급식일자	급식인원수	요리명	원산지정보	칼로리정보	영양정보	적재일시
2	0	C10	부산광역시	7150119	가야고등학교	2	중식	20200205	358	숙주나물냉면	쌀 : 국내산	98.7 Kcal	탄수화물(g)	2.02E+13
3	1	C10	부산광역시	7150119	가야고등학교	2	중식	20200204	358	참치야채밥	쌀 : 국내산	1120.1 Kcal	탄수화물(g)	2.02E+13
4	2	C10	부산광역시	7150119	가야고등학교	2	중식	20200203	358	김가루볶음	쌀 : 국내산	1415.7 Kcal	탄수화물(g)	2.02E+13
5	3	C10	부산광역시	7150119	가야고등학교	2	중식	20191226	358	햄김치볶음	쌀 : 국내산	1745.6 Kcal	탄수화물(g)	2.02E+13
6	4	C10	부산광역시	7150119	가야고등학교	2	중식	20191224	358	뽕초고기주	쌀 : 국내산	1835.0 Kcal	탄수화물(g)	2.02E+13
7	5	C10	부산광역시	7150119	가야고등학교	2	중식	20191223	358	현미밥(가)	쌀 : 국내산	1318.4 Kcal	탄수화물(g)	2.02E+13
8	6	C10	부산광역시	7150119	가야고등학교	2	중식	20191220	358	나물비빔밥	쌀 : 국내산	1427.4 Kcal	탄수화물(g)	2.02E+13
9	7	C10	부산광역시	7150119	가야고등학교	2	중식	20191219	358	흑미밥(가)	쌀 : 국내산	1153.8 Kcal	탄수화물(g)	2.02E+13
10	8	C10	부산광역시	7150119	가야고등학교	2	중식	20191218	358	김치참치김	쌀 : 국내산	1580.5 Kcal	탄수화물(g)	2.02E+13

167540 row x 14 column

총 17개 각각의 시도교육청 급식식단 csv파일을
하나로 합쳐 '급식식단 데이터.csv'로 저장

2-2) 병원 식단 데이터

- 공공데이터포털 <https://www.data.go.kr/dataset/15017299/fileData.do>

국립소록도병원 영양급식 정보 20181231			
업데이트 주기	연간	차기등록예정일	2019-12-21
비용부과유무	무료	비용부과기준 및 단위	없음
다운로드 횟수	33		
등록일	2019-04-30	수정일	2019-05-09
이용허락범위	이용허락범위 제한 없음		
제공형태	공공데이터포털에서 다운로드(원문파일등록)		
URL	https://www.data.go.kr/dataset/fileDownload.do?atchFileId=FILE_000000001520785&fileDetailSn=1		
설명	국립소록도병원 입원한센인 영양급식(식단)_조식,중식,석식,일자등 정보 제공		
관리부서명	서무과	관리부서 전화번호	061-840-0528

보건복지부에서 운영하는 2018년 국립소록도병원 식단 다운
-> 후에 식단 추천 시 사용할 데이터임

2 구성요소

2-2) 병원 식단 데이터

- 공공데이터포털 <https://www.data.go.kr/dataset/15017299/fileData.do>

	A	B	C	D
1	일자	구분	메뉴명	
2	2018-01-02	조식	오징어무국	
3	2018-01-02	조식	스크램블에그	
4	2018-01-02	조식	베이컨감자채볶음	
5	2018-01-02	조식	숙갓나물	
6	2018-01-02	조식	배추김치	
7	2018-01-02	중식	툇밥	
8	2018-01-02	중식	추어탕	
9	2018-01-02	중식	돈육계란장조림	
10	2018-01-02	중식	해물파전	
11	2018-01-02	중식	깻잎지	
12	2018-01-02	중식	알타리김치	

5627 row x 3 column
'hospital.csv'로 저장

3) 간식 데이터

- 식품안전나라 <https://www.foodsafetykorea.go.kr/main.do>

간편검색

DB정보 공유가이드

식품

음식

가공식품

가공식품군

전체

가공식품명

검색

자료다운

Total: 13,497 Items, 1/675 Pages

번호	가공식품군	1회제공량 (g)	열량 (kcal)	탄수화물 (g)	단백질 (g)	지방 (g)
	가공식품명	당류 (g)	나트륨 (mg)	콜레스테롤 (mg)	포화지방산 (g)	트랜스지방산 (g)
13497	과자류	30	375	43.75	6.25	18.75
	150누네피네, 2000누네피네, 3600누네피네	1.5	312.5	0	12.5	0

식품영양성분DB 메뉴의 가공식품 메뉴에 들어가
고등학생들이 즐겨먹는 가공식품(간식) 데이터 다운

2 구성요소

3) 간식 데이터

- 식품안전나라 <https://www.foodsafetykorea.go.kr/main.do>

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		번호	식품군	식품이름	1회제공량	열량 (kcal)	탄수화물 (g)	단백질 (g)	지방 (g)	당류 (g)	나트륨 (mg)	콜레스테롤 (mg)	포화지방산 (g)	트랜스지방 (g)	연도
2	0	1	빵 또는 떡	1,000원의	70	202.54	36.92	6.98	3	2.69	326.35	1.51	1.09	0	2013
3	1	2	빵 또는 떡	1,000원의	70	141.78	25.84	4.88	2.1	2.45	228.45	1.06	0.76	0	2017
4	2	3	빵 또는 떡	100% 우리	70	242.31	31.23	4.31	11.85	6.46	280	0	5.38	0	2012
5	3	4	빵 또는 떡	100% 우리	70	260.95	33.63	4.64	12.76	4.52	301.54	0	5.8	0	2017
6	4	5	빵 또는 떡	100% 우리	70	215.38	31.23	3.23	8.62	5.38	258.46	0	4.85	0	2012
7	5	6	빵 또는 떡	100% 우리	70	239.32	34.7	3.59	9.57	3.77	287.18	0	5.38	0	2017
8	6	7	빵 또는 떡	100% 우리	70	231.54	32.31	3.23	10.77	7.54	247.69	0	5.38	0	2012
9	7	8	빵 또는 떡	100% 우리	70	249.35	34.79	3.48	11.6	5.28	266.75	0	5.8	0	2017
10	8	9	빵 또는 떡	100%우리	70	236.92	32.31	4.31	11.85	6.46	258.46	0	6.46	0	2012

9430 row x 15 column

총 8개 각각의 가공식품군 csv파일을
하나로 합쳐 '간식 데이터.csv'로 저장

3 인체 데이터 전처리

1) '권장섭취량', '탄수화물', '단백질', '지방' column 추가

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	성별	골격근량	체지방량	BMI	체지방율	복부지방율	몸무게	나이	키	권장섭취량	탄수화물	단백질	지방
2	1	21.7	6	16.95419	13.2	0.7	45.6	17	164	2700	209.7825	104.8913	15.53944
3	1	22.1	7	16.83558	14.7	0.7	47.8	17	168.5	2700	217.695	108.8475	16.12556
4	1	22.3	6.7	17.79882	14.3	0.8	47	17	162.5	2700	211.545	105.7725	15.67
5	1	22.4	5.9	18.05486	12.7	0.8	46.8	17	161	2700	210.0075	105.0038	15.55611
6	1	22.5	6.2	16.5024	13.2	0.7	47.3	17	169.3	2700	217.2638	108.6319	16.09361
7	1	22.6	8	17.89867	16.3	0.8	49.5	17	166.3	2700	219.5513	109.7756	16.26306
8	1	22.7	8.6	18.66659	17.2	0.8	49.9	17	163.5	2700	218.2763	109.1381	16.16861
9	1	22.8	3.5	17.15937	7.8	0.8	45.2	17	162.3	2700	207.6825	103.8413	15.38389
10	1	22.9	6.9	16.70823	14.1	0.7	49.2	17	171.6	2700	222.9075	111.4538	16.51167

‘나이, 키, 체중’ 값을 가지고 각 행에
권장섭취량(kcal), 탄수화물(g), 단백질(g), 지방(g)
값을 계산하여 column에 추가
(* 코드 없이 엑셀 기능 사용)

3 인체 데이터 전처리

2) '비타민', '티아민', '리보플라빈', '칼슘', '철분' column 추가



비타민A	티아민	리보플라빈	비타민C	칼슘	철분	알레르기
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	18
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	10
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	4

음식 식단 데이터에 기록되어 있는 추가 영양소의
권장 섭취량 값을 나이, 성별에 따라 column에 추가
(* 2010 한국인 영양섭취기준 문서 참고)

3 인체 데이터 전처리

3) '알레르기' 보유 여부를 확률로 설정하여 알레르기 번호 추가



비타민A	티아민	리보플라빈	비타민C	칼슘	철분	알레르기
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	18
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	10
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	
850	1.3	1.7	110	900	15	4

1.난류, 2.우유, 3.메밀, 4.땅콩, 5.대두, 6.밀, 7.고등어, 8.게, 9.새우, 10.돼지고기, 11.복숭아, 12.토마토, 13.아황산염, 14.호두, 15.닭고기, 16.쇠고기, 17.오징어, 18.조개류(굴,전복,홍합 등)

알레르기 0개 – 80%

알레르기 1개 – 15%

알레르기 2개 – 5%

3 인체 데이터 전처리

4) 전처리 결과를 '인체치수 최종본.csv'로 저장

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		성별	골격근량	체지방량	BMI	체지방율	복부지방량	몸무게	나이	키	기초대사량	탄수화물	단백질	지방	비타민A	티아민	리보플라빈	비타민C	칼슘	철분	알레르기
2	0	1	21.7	6	16.9542	13.2	0.7	45.6	17	164	1398.55	209.783	104.891	15.5394	850	1.3	1.7	110	900	15	
3	1	1	22.1	7	16.8356	14.7	0.7	47.8	17	168.5	1451.3	217.695	108.848	16.1256	850	1.3	1.7	110	900	15	
4	2	1	22.3	6.7	17.7988	14.3	0.8	47	17	162.5	1410.3	211.545	105.773	15.67	850	1.3	1.7	110	900	15	
5	3	1	22.4	5.9	18.0549	12.7	0.8	46.8	17	161	1400.05	210.008	105.004	15.5561	850	1.3	1.7	110	900	15	
6	4	1	22.5	6.2	16.5024	13.2	0.7	47.3	17	169.3	1448.43	217.264	108.632	16.0936	850	1.3	1.7	110	900	15	18
7	5	1	22.6	8	17.8987	16.3	0.8	49.5	17	166.3	1463.68	219.551	109.776	16.2631	850	1.3	1.7	110	900	15	
8	6	1	22.7	8.6	18.6666	17.2	0.8	49.9	17	163.5	1455.18	218.276	109.138	16.1686	850	1.3	1.7	110	900	15	10
9	7	1	22.8	3.5	17.1594	7.8	0.8	45.2	17	162.3	1384.55	207.683	103.841	15.3839	850	1.3	1.7	110	900	15	
10	8	1	22.9	6.9	16.7082	14.1	0.7	49.2	17	171.6	1486.05	222.908	111.454	16.5117	850	1.3	1.7	110	900	15	
11	9	1	23.4	2.9	16.7493	6.3	0.7	45.6	17	165	1403.55	210.533	105.266	15.595	850	1.3	1.7	110	900	15	
12	10	1	23.5	5.2	17.4628	11	0.7	47.6	17	165.1	1431.55	214.733	107.366	15.9061	850	1.3	1.7	110	900	15	
13	11	1	23.5	9	19.5649	17.3	0.8	51.6	17	162.4	1473.05	220.958	110.479	16.3672	850	1.3	1.7	110	900	15	4
14	12	1	23.8	9.2	19.2371	17.5	0.8	52.5	17	165.2	1499.43	224.914	112.457	16.6603	850	1.3	1.7	110	900	15	
15	13	1	23	5.9	15.8575	12.2	0.7	47.9	17	173.8	1479.18	221.876	110.938	16.4353	850	1.3	1.7	110	900	15	
16	14	1	23	8.5	17.7794	16.8	0.8	50.9	17	169.2	1497.43	224.614	112.307	16.6381	850	1.3	1.7	110	900	15	
17	15	1	24.1	7.1	17.5706	14	0.8	50.6	17	169.7	1495.8	224.37	112.185	16.62	850	1.3	1.7	110	900	15	
18	16	1	24.2	5.7	18.5633	11.6	0.8	49.2	17	162.8	1442.05	216.308	108.154	16.0228	850	1.3	1.7	110	900	15	18
19	17	1	24.3	5.8	18.0072	11.7	0.8	50.1	17	166.8	1474.43	221.164	110.582	16.3825	850	1.3	1.7	110	900	15	
20	18	1	24.3	7.2	19.6545	14.1	0.8	51.2	17	161.4	1462.55	219.383	109.691	16.2506	850	1.3	1.7	110	900	15	

1405 row x 21 column

3 급식식단 데이터 전처리

1) 알레르기 column 추가

‘요리 이름’ column 값

참치야채비빔밥(가)5.6.10.13.
콩가루배추국(가)5.6.9.13.
명엽채볶음
(가)5.6.13.
지파이(가)1.5.6.12.13.15.
깍두기(가)9.13.



알레르기 번호 중복없이 나열

1, 5, 6, 9, 10, 12, 13, 15



1.난류	2.우유	3.메밀	4.땅콩	5.대두	6.밀	7.고등어	8.게	9.새우	10.돼지고	11.복숭아	12.토마토	13.아황산	14.호두	15.닭고기	16.쇠고기	17.오징어	18.조개류(굴,전복,홍합 등)		
1	0	0	0	1	1	0	0	1	1	0	1	1	0	1	0	0	0		

해당하는 알레르기가 포함되면 1
포함되지 않으면 0

3 급식식단 데이터 전처리

2) '요리 이름' 문자열 변환

'요리 이름' column 값

참치야채비빔밥(가)5.6.10.13.
콩가루배추국(가)5.6.9.13.
명엽채볶음
(가)5.6.13.
지파이(가)1.5.6.12.13.15.
깍두기(가)9.13.



필요 없는 문자열 제거하여 요리명만 남기기

참치야채비빔밥,콩가루배추국,명엽채볶음,지파이,깍두기

3 급식식단 데이터 전처리

3) '칼로리정보', '영양정보' column 값 변환

'칼로리정보' column 값
1120.1 Kcal

'영양정보' column 값

탄수화물(g) : 148.7
단백질(g) : 58.4
지방(g) : 29.8
비타민A(R.E) : 401.1
티아민(mg) : 0.5
리보플라빈(mg) : 0.4
비타민C(mg) : 34.3
칼슘(mg) : 341.4
철분(mg) : 11.1



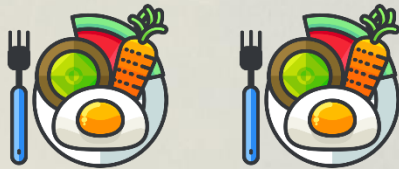
탄수화물(g)	단백질(g)	지방(g)	비타민A(R.E)	티아민(mg)	리보플라빈(mg)	비타민C(mg)	칼슘(mg)	철분(mg)	칼로리(Kcal)
148.7	58.4	29.8	401.1	0.5	0.4	34.3	341.4	11.1	1120.1

각 정보에 해당하는 column을 만들어 깔끔하게 보이도록 정리

4) '한끼, 두끼 식사 / 세끼 식사' 로 나누기



or



'일이식이.csv'

54694 row x 30 column



'삼식이.csv'

55374 row x 30 column

한끼 급식만 제공(중식)하는 날은 한끼
 두끼 급식을 제공(중식, 석식)하는 날은 두끼
 세끼 급식을 제공(조식, 중식, 석식)하는 날은 세끼
 로 묶어 하루치 식단으로 합침

3 간식 데이터 전처리

1) 식품이름이 중복인 행 제거

식품이름	1회제공량	열량 (kcal)	탄수화물 (g)	단백질 (g)	지방 (g)	당류 (g)	나트륨 (mg)	콜레스테롤 (mg)	포화지방산 (g)	트랜스지방산 (g)	연도
1,000원의 행복 토스트 식빵	70	202.54	36.92	6.98	3	2.69	326.35	1.51	1.09	0	2013
1,000원의 행복 토스트 식빵	70	141.78	25.84	4.88	2.1	2.45	228.45	1.06	0.76	0	2017

식품이름이 같은 행이 두개 씩 존재 (2013, 2017 데이터)
-> 2013년 데이터는 지우기

2) 식품군이 '면류'인 행 중 식품이름에 '분유', '이유식'이 들어간 행 제거

- 고등학생이 분유와 이유식 식품을 먹을 리가 없으므로!

면류	분유,전지분유, 찬물에잘녹는전지분유(미)
면류	분유,전지분유, 찬물에잘녹는전지분유(미)
면류	분유,조제분유, 건양밀(일동후디스)
면류	분유,조제분유, 건양밀(일동후디스)
면류	분유,조제분유, 골드뉴로히트1(파스퇴르)
면류	분유,조제분유, 골드뉴로히트1(파스퇴르)

면류	이유식,이유식, 맘마밀-1(매일유업)
면류	이유식,이유식, 맘마밀-1(매일유업)
면류	이유식,이유식, 맘마밀-2(매일유업)
면류	이유식,이유식, 맘마밀-2(매일유업)
면류	이유식,이유식, 맘마밀-3(매일유업)
면류	이유식,이유식, 맘마밀-3(매일유업)

3 간식 데이터 전처리

3) NaN 값이 포함된 행 제거

```
# NaN 값이 있는지 확인  
snack_data.isnull().sum()
```

번호	0
식품군	0
식품이름	0
1회제공량 (g)	0
열량 (kcal)	0
탄수화물 (g)	0
단백질 (g)	3
지방 (g)	0
당류 (g)	189
나트륨 (mg)	1
콜레스테롤 (mg)	437
포화지방산 (g)	436
트랜스지방산 (g)	437
년도	0
dtype:	int64



데이터 분석 시 문제가 되므로 제거

4) 필요 없는 열 지우기



‘번호’
‘콜레스테롤(mg)’
‘1회제공량(g)’
‘트랜스지방산(g)’
‘년도’

3 간식 데이터 전처리

4) 전처리 결과를 '간식 데이터 최종본.csv'로 저장

	A	B	C	D	E	F	G	H	I	J	
1		식품군	식품이름	열량 (kcal)	탄수화물 (g)	단백질 (g)	지방 (g)	당류 (g)	나트륨 (mg)	포화지방산 (g)	
2	0	빵 또는 떡	1,000원의 행복 토스트 식빵	141.78	25.84	4.88	2.1	2.45	228.45	0.76	
3	1	빵 또는 떡	100% 우리밀로 만든 미니땅콩샌드	260.95	33.63	4.64	12.76	4.52	301.54	5.8	
4	2	빵 또는 떡	100% 우리밀로 만든 미니복숭아샌드	239.32	34.7	3.59	9.57	3.77	287.18	5.38	
5	3	빵 또는 떡	100% 우리밀로 만든 미니옥수수샌드	249.35	34.79	3.48	11.6	5.28	266.75	5.8	
6	4	빵 또는 떡	100%우리밀로 만든 미니딸기샌드	255.15	34.79	4.64	12.76	4.52	278.34	6.96	
7	5	빵 또는 떡	100%우리밀로 만든 미니생크림샌드	255.15	34.79	3.48	11.6	6.79	289.94	5.8	
8	6	빵 또는 떡	100%우리밀로 만든 생크림 샌드	254.87	34.7	3.59	11.97	6.79	299.14	5.98	
9	7	빵 또는 떡	1964정통 크림빵	269.37	39.2	4.02	11.06	8.29	231.18	4.83	
10	8	빵 또는 떡	56시간 부드러운熟숙	159.49	28	4.24	3.14	1.94	229.05	2.04	
11	9	빵 또는 떡	56시간 저온숙성 숙	135.17	23.61	4.38	2.58	1.75	230.02	1.45	
12	10	빵 또는 떡	BGF통식빵	139.65	25.48	4.9	1.91	1.96	200.9	0.98	

4095 row x 10 column

4 준비 과정

일주일치 식단 만들기

- 학생 한 명 당 일주일치 식단 transaction을 배치함



요일	하루 먹은 식단	영양소
월	조식, 중식, 석식, 간식	탄, 단, 지 등등
화	조식, 중식, 석식, 간식	탄, 단, 지 등등
수	조식, 중식, 석식, 간식	탄, 단, 지 등등
목	조식, 중식, 석식, 간식	탄, 단, 지 등등
금	조식, 중식, 석식, 간식	탄, 단, 지 등등
토	중식, 석식, 간식	탄, 단, 지 등등
일	석식, 간식	탄, 단, 지 등등

균형, 불균형 식단으로 나눈 이유는?

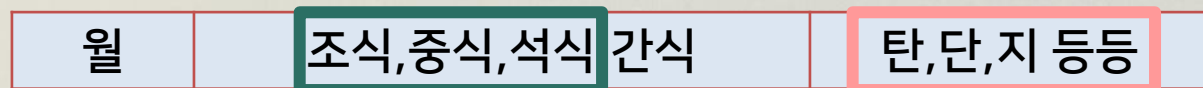
- 균형적인 식단으로만 일주일치 식단을 짜는 것은 비현실적이고, 분석 결과에 특이점이 없을 것이므로. 불균형적인 식단은 outlier라 생각할 수 있다.

균형적인 식단 (3끼 꼬박꼬박 먹음)

불균형적인 식단 (1끼 또는 2끼만)

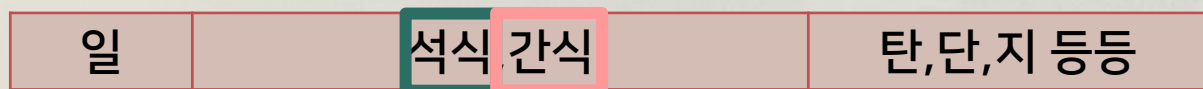
4 준비 과정

일주일치 식단의 column 값 설명



↓
'삼식이.csv' 데이터에서 랜덤으로
한 개 행 선택해서 넣음

↓
전체 식단의 영양성분을
모두 합해서 저장



↓
'일이식이.csv' 데이터에서 랜덤으로
한 개 행 선택해서 넣음

↓
'간식 데이터.csv' 에서 랜덤으로
두개 행 선택해서 넣음
(간식을 하루에 2개씩 먹는다 가정)

4 준비 과정

일주일치 식단을 두개로 나누기

class	요일	식단
1	월~일	하루 식단 x 7일
2	월~일	하루 식단 x 7일
3	월~일	하루 식단 x 7일
4	월~일	하루 식단 x 7일
5	월~일	하루 식단 x 7일
6	월~일	하루 식단 x 7일
7	월~일	하루 식단 x 7일

하나의 행에 일주일치 식단이
모두 통합된 데이터



일주일 식단을 총평 할 때 사용

class	요일	하루 먹은 식단
1	월	조식,중식,석식,간식
1	화	조식,중식,석식,간식
1	수	조식,중식,석식,간식
1	목	조식,중식,석식,간식
1	금	조식,중식,석식,간식
1	토	중식,석식,간식
1	일	석식,간식

하나의 행에 각 요일마다 하루치의
식단만 포함한 데이터



하루 식단 점검 할 때 사용

4 식단 평가 방법

하루 식단 점검 방법

요일	하루 먹은 식단	영양소, 알레르기 여부
월	조식,중식,석식,간식	총 영양소, 알레르기 정보
화	조식,중식,석식,간식	총 영양소, 알레르기 정보
수	조식,중식,석식,간식	총 영양소, 알레르기 정보
목	조식,중식,석식,간식	총 영양소, 알레르기 정보
금	조식,중식,석식,간식	총 영양소, 알레르기 정보
토	중식,석식,간식	총 영양소, 알레르기 정보
일	석식,간식	총 영양소, 알레르기 정보



각종 영양분 권장량
알레르기 여부

하루치 식단에 들어있는 각종 영양소와
각 학생 개인의 영양분 권장량을 비교하여 문제가 있는 영양분 지적
(+알레르기가 있는 음식을 먹었을 때도 알려줌)

4 식단 평가 방법

일주일 식단 총평 방법

요일	식단	일주일 섭취한 총 영양분
월~일	하루 식단 x 7일	총 영양소



각종 일일 영양분 권장량
X 7

일주일치 식단에 들어있는 총 영양분과
각 학생 개인의 (일일 영양분 권장량 X 7)을 비교하여 모든 영양소 총평

4 식단 평가 방법

영양소 섭취 평가 기준

영양소	평가 기준
열량(kcal)	권장 섭취량의 5% 이내 양을 섭취하지 못했으면 지적
탄수화물(g)	
단백질(g)	
지방(g)	
각종 비타민(mg)	
칼슘(mg)	
철분(mg)	
나트륨(mg)	300mg 이상 먹었을 시 지적
당류(g)	20mg 이상 먹었을 시 지적
포화지방산(g)	12mg 이상 먹었을 시 지적

4 식단 평가 결과

하루 식단 점검 결과 예시

아침

흰밥, 시나몬토스트, 쇠고기무국, 닭가슴살장조림, 두부매운조림, 배추김치, 우유, 옥수수콘 버터구이

점심

현미밥, 콩나물수제비국, 등갈비찜, 풋마늘골뱅이초회, 마늘돈까스, 양배추샐러드, 오렌지

저녁

기장밥, 쇠고기콩나물국, 참나물오이무침, 닭날개 당호박조림, 김치떡볶음, 배추김치, 얼라이브쥬스

간식

망고맛 푸딩, 무기후아바닐라맛웨이퍼



'권장섭취량은(는) 이 식습관 유지하세요',
'탄수화물을 1038.1g 만큼을 덜 드세요',
'단백질을 100.85g 만큼을 덜 드세요',
'지방을 373.0g 만큼을 덜 드세요',
'비타민A을 475.3mg 만큼을 덜 드세요',
'티아민을 15.7mg 만큼을 덜 드세요',
'리보플라빈을 9.4mg 만큼을 덜 드세요',
'비타민C은(는) 이 식습관 유지하세요',
'칼슘을 1432.7mg 만큼을 덜 드세요',
'철분을 30.8mg 만큼을 덜 드세요'

4 식단 평가 결과

일주일 식단 총평 결과 예시

일주일 조식 또는 중식 또는 석식

친환경쌀발아현미밥,콩나물국저염식,한우버섯장조림,수제소시지구이,유기농배추김치,파스타발사믹샐러드,파인애플과일,우리밀크린베리쿠키빵//친환경쌀수수밥,냉이콩국저염식,통삼겹살구이,보쌈김치,파절이생채나물,양배추숙회,물미역숙회나물,수제머핀//친환경쌀낙지고추장비빔밥,봄동된장국저염식,연두부,섭산적샐러드,잡채,유기농배추김치,파인애플쥬스 ... 생략

일주일 간식

큰컵간썬빵, 롤스토리 골든 롤 초코@@초코팡, 하이랜드스톰샤워(크라운제과)@@오리지널단팥크림빵,로아커웨하스 나폴리타너@@인디 라운드 나초칩,매운꿀과배기(농심)@@맥시 카사바 칩(사워크림),청정원우리밀찰진국수 (대정농산㈜)@@진라면 컵매운맛(오뚜기라면), 치오 스틱-감자맛@@세일럼 클래식 슈가 쿠키, 초이스엘 나초칩

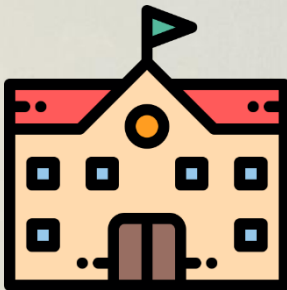


'권장섭취량을 1676.95kcal 만큼을 더 드세요',
'탄수화물을 890.17g 만큼을 덜 드세요',
'단백질을 50.47g 만큼을 더 드세요',
'지방을 351.7g 만큼을 덜 드세요',
'비타민A을 635.7mg 만큼을 덜 드세요',
'티아민을 4.4mg 만큼을 덜 드세요',
'리보플라빈을 1.5mg 만큼을 더 드세요',
'비타민C은(는) 이 식습관 유지하세요',
'칼슘을 1495.0mg 만큼을 더 드세요',
'철분을 23.1mg 만큼을 덜 드세요'
'나트륨(mg)을 2968.21 만큼 덜 드세요'

5 식단 추천 방법



일주일동안 **올바른 식단**
을 먹은 학생



현식습관을 유지하도록
일주일치 세끼 **급식 식단 추천**



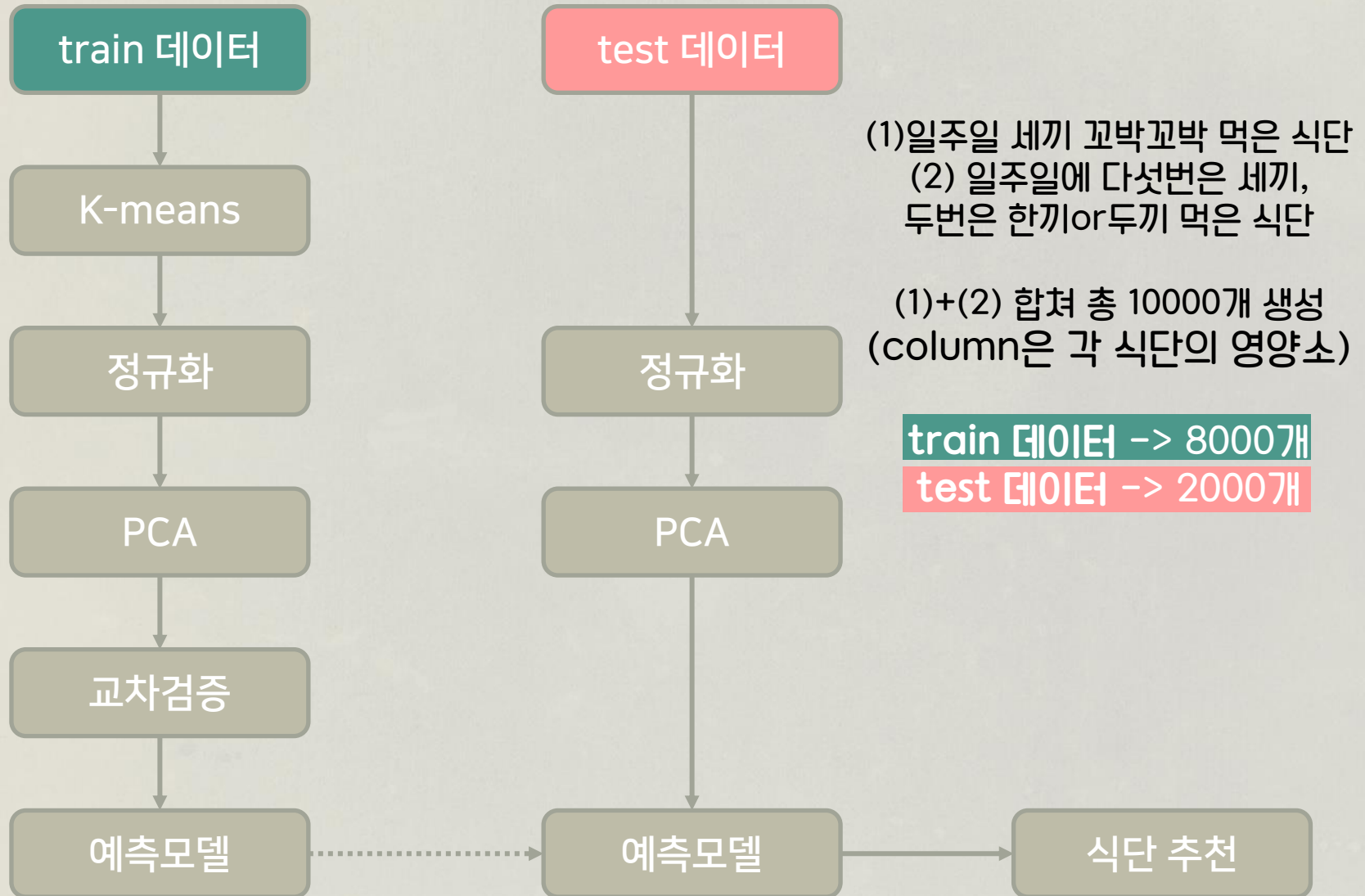
일주일동안 **올바르지 않은 식단**
을 먹은 학생



급식 식단보다 더 건강하게 짜여진
일주일치 세끼 **병원 식단 추천**

올바른 식단과
올바르지 않은 식단으로
나눌 필요가 있음!

5 식단 추천 방법



5 식단 추천 방법

k-means clustering – train 데이터를 라벨링

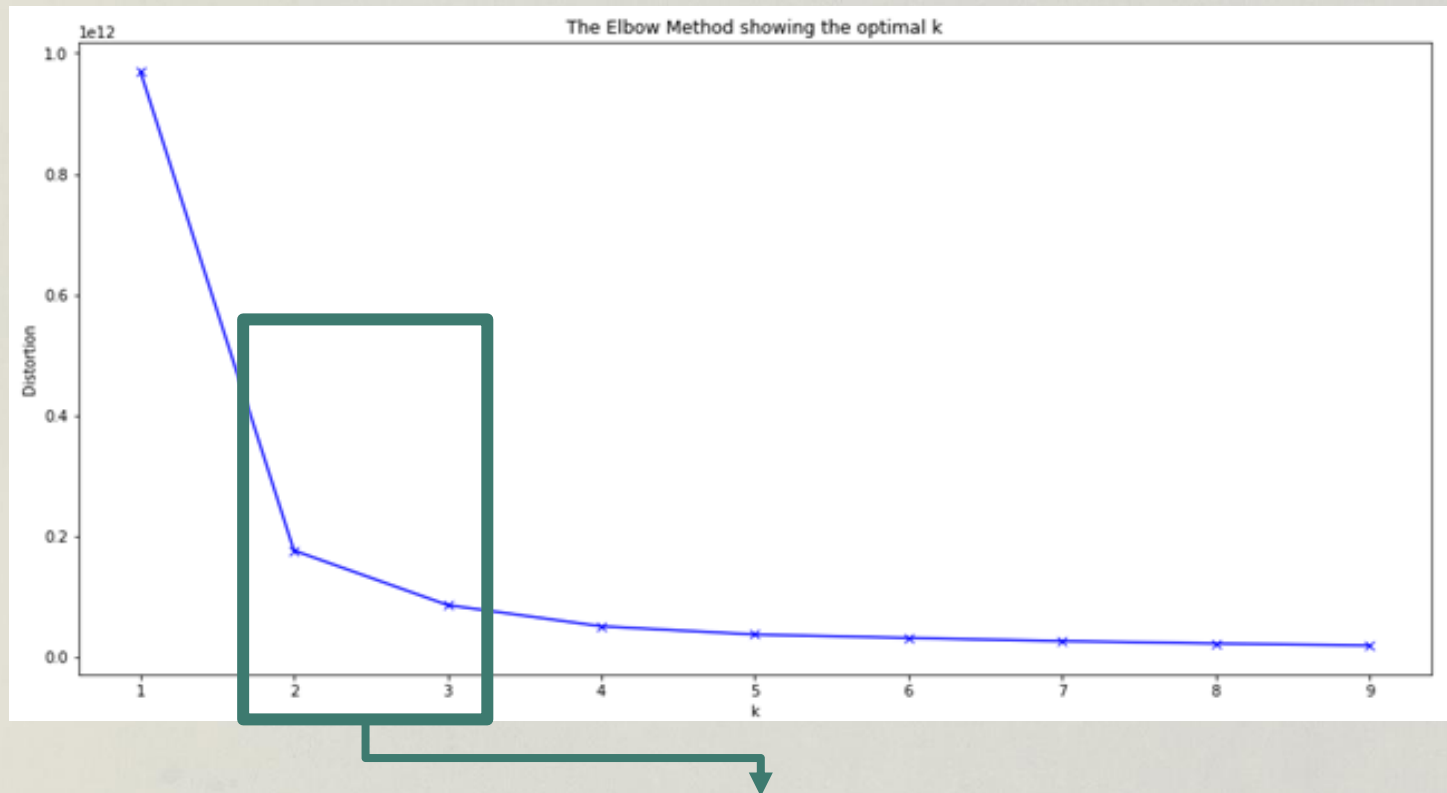
index	각 9개 영양소 col	label
1	일주일치 총합	?
2	일주일치 총합	?
3	일주일치 총합	?
4	일주일치 총합	?
5	일주일치 총합	?
6	일주일치 총합	?
7	일주일치 총합	?

label을 나누는 이유는?

- 일주일치 식단을 군집화하여
test 데이터를 올바르게 올바르지 않은 식단,
올바른 식단으로 나누기 위하여

5 식단 추천 방법

k-means clustering → label을 몇 개로 나눌 것인가?



Elbow 메서드를 이용하여 최적의 label 개수 탐색
→ 2, 3 중 홀수인 3개로 선택

5 식단 추천 방법

9개의 영양소 column 값 정규화

‘칼로리’, ‘탄수화물’, ‘단백질’, ‘지방’, ‘비타민A’, ‘티아민’, ‘리보플라빈’, ‘비타민C’, ‘칼슘’, ‘철분’

```
from sklearn.preprocessing import StandardScaler  
scaler = StandardScaler()  
scaler.fit(value)  
value_transform=scaler.transform(value)
```



	칼로리(Kcal)	탄수화물(g)	단백질(g)	지방(g)	비타민A(mg)	티아민(mg)	비타민C(mg)	리보플라빈(mg)	칼슘(mg)	철분(mg)
0	-0.234927	-0.901105	-1.198286	-0.032841	-0.312273	0.790511	-0.637530	-0.366321	-1.317727	-0.543711
1	-0.149315	0.097261	-0.515773	-0.132695	0.087358	-0.448208	0.040284	-0.264731	0.666130	-0.053567
2	-0.123845	-0.230331	-0.336077	-0.034330	-0.213755	-0.442903	-0.680212	-0.357855	-0.572266	-0.000196
3	-0.197394	-0.681058	-0.915264	-0.029034	-0.152285	-0.493300	-1.205135	-0.346567	-1.264799	-0.147239
4	-0.019752	-0.264617	0.009173	-0.086904	-0.007463	3.037182	-0.085851	3.341712	0.119358	0.489948
...
7995	-0.051822	-0.119705	0.314490	-0.043092	0.271467	-0.347413	1.509942	-0.225224	-0.375244	-0.112384
7996	0.094170	0.883161	0.543270	-0.033482	0.518152	-0.387200	1.206709	-0.225224	1.040413	-0.049210
7997	-0.200605	-1.003133	-0.584656	0.014363	-0.442683	1.021279	-1.213417	-0.010756	-0.490868	-0.414095
7998	-0.085196	-0.207295	0.490027	-0.093688	0.836544	0.058420	-0.315186	-0.225224	0.113224	0.163185
7999	-0.111021	-0.571584	0.144777	-0.036497	0.044827	2.565037	0.023083	2.876092	0.059842	-0.165756

5 식단 추천 방법

주성분 분석(PCA) – 9개의 영양소 feature 값을 2개의 feature로 변환

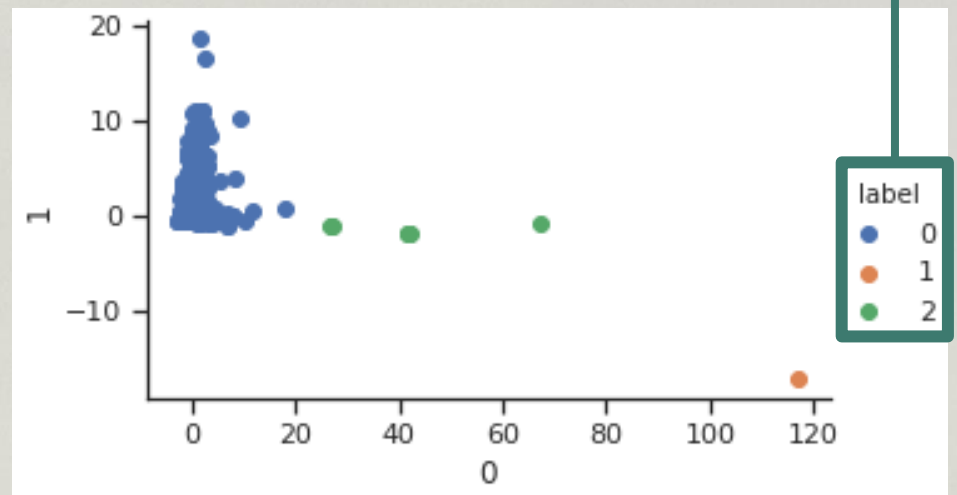
Why?

총 9개의 feature를 한꺼번에 고려할 수 없으므로
이를 대표하는 2개의 feature로 축약하기 위해서

2개의 feature (0, 1)

	0	1
0	-1.785358	0.360137
1	-0.057914	-0.448880
2	-0.787412	-0.561652
3	-1.599988	-0.584167
4	0.444868	4.456463
...
7995	0.287034	-0.329894
7996	1.376286	-0.413287
7997	-1.332372	0.721748
7998	0.398626	-0.149083
7999	0.190439	3.825443

3개의 label (0, 1, 2)



라벨링과 PCA 분석을 모두 마친 데이터를 시각화

5 식단 추천 방법

k-NN 알고리즘 -> k 값을 정하기 위해 교차검증(k-fold) 실시

k값에 4, 5, 6을 넣어
교차검증의 평균 값을 비교

```
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score

model = KNeighborsClassifier(n_neighbors=5) # 모델 객체 생성
kfold = KFold(n_splits=5)
scores = cross_val_score(model, value, label, cv=5)
print('cross validation score: {}'.format(scores))
print("교차검증의 평균 : %.3f" %(scores.mean()))
```

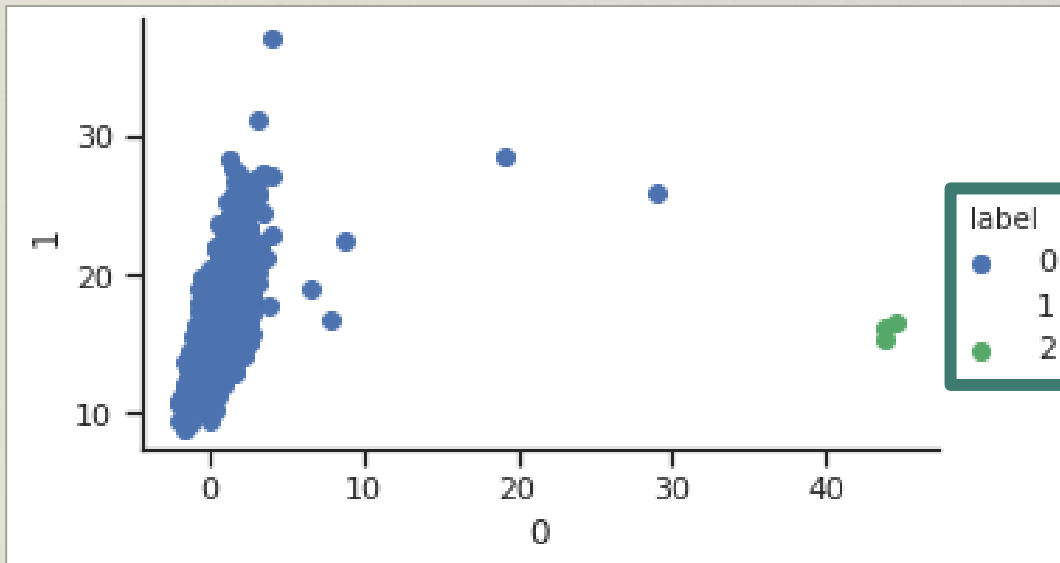
k=4, k=5, k=6

모두 같은 값이 나와서 그 중 가운데 값인 5로 k값을 결정

-> 이후, train 데이터에 k-NN 알고리즘을 적용하여 예측모델 생성

5 식단 추천 방법

test 데이터에 예측모델을 적용



Label 0
올바른 식단을 먹은 학생
-> 급식 식단 추천!

Label 2
올바르지 않은 식단을
먹은 학생
-> 병원 식단 추천!

Label 0이 많은 이유

임의로 test데이터와 train데이터를 서로 비슷한 급식데이터로 나뉘었기 때문

-> 더 차이가 뚜렷한 데이터들을 집어넣으면 라벨마다 데이터의 분포가 다양해질 것이다

Label 0으로 분류된 학생

월	조식	찰현미밥, 설렁탕, 돈육볶음, 깍두기, 호두과자, 우유
	중식	차수수밥, 콩나물쫄면, 팥이버섯된장국, 도토리묵 상추무침, 치킨바겐, 배추김치
	석식	치볶음밥, 콩나물쫄면, 미소된장국, 단무지무침, 불고기함박스테이크, 요구르트
화	조식	잡쌀밥, 북어무국, 떡갈비견과류조림, 청경채나물, 배추김치, 바나나
	중식	찰보리밥, 부대찌개, 오이고추무침, 오리불고기, 감자전, 배추겉절이
	석식	잡곡밥, 오징어국, 훈제오리&머스터드소스, 두부구이, 깍두기, 부추겉절이
수	조식	케찹, 누룽지탕, 눈꽃치즈토마토, 배추김치, 불고기 타코, 초코에몽
	중식	강황쌀밥, 우리밀삼색수제비국, 콩나물파채무침, 통삼겹구이, 배추김치, 수박, 상추
	석식	후랑크소시지도시락, 근대된장국, 깍두기, 회오리감자, 청포도에이드

급식 식단 데이터들 중에 랜덤으로 하나를 가져와 일주일치를 추천해줌

Label 2으로 분류된 학생

월	조식	삼색수제비국, 삼치카레구이, 진미채볶음, 냉이 된장무침, 배추김치
	중식	오징어무국, 스크램블에그, 베이컨감자채볶음, 썩갠나물, 배추김치
	석식	톳밥, 추어탕, 돈육계란장조림, 해물파전, 깻잎지, 알타리김치
화	조식	동태알탕, 닭감자조림, 톳두부무침, 애호박새우젓볶음, 배추김치
	중식	육개장, 두부전&양념장, 코다리순살무조림, 유채나물, 배추김치
	석식	매생이굴국, 돼지갈비떡찜, 숙주맛살무침, 참나물무침, 배추김치
수	조식	쇠고기떡국, 참치야채볶음, 감자메추리알조림, 유채나물, 배추김치
	중식	누룽지, 쇠고기계란장조림, 표고버섯새우살볶음, 어묵고추장볶음, 깍두기
	석식	찰현미밥, 두부호박청국장국, 삼치카레구이, 돌김&양념장, 오이소박이, 파김치

병원 식단 데이터들 중에 랜덤으로 하나를 가져와 일주일치를 추천해줌

1) 기대효과

1. 식습관 검증 및 식습관 개선 방안에 대해 알 수 있음
2. 부족한 영양소가 무엇인지 알고, 보충 할 수 있음
3. 현재 실시하고 있는 급식 식단의 영양소가 잘 짜여 있는지 알 수 있음
4. 나에게 맞는 올바른 식단을 추천 받을 수 있음

2) 추후연구

1. 고등학생 대상에서 전체 연령으로 대상 확대
2. 실제 식단 데이터를 input 받아 연구 실시
3. 보유한 알레르기 위험이 있는 식단은 제외하고 식단 추천



6 한계점

1. 간식 데이터의 함량 영양소 값에 오류가 있는 경우가 있음
-> 데이터 제공 기관의 측정 오류이므로 해결할 방도가 없었음
2. 실제 섭취한 식단을 input 하지 않고 급식 식단으로 대체함
-> 실제 섭취한 special한 식단 데이터를 얻기가 어려워,
데이터를 얻기 쉬운 general한 급식 식단을 이용함
3. 각 개인의 알레르기 정보 데이터를 얻지 못함
-> 실제 데이터가 아닌 랜덤으로 알레르기 정보를 새로 생성함



The background consists of horizontal stripes. At the top and bottom are teal-colored bands. Between them are grey bands, and in the center is a wide tan-colored band. The stripes are separated by thin white lines, and the teal bands at the top and bottom have a decorative, serrated edge.

Thank you