

## I. 참가자 정보

|       |                        |       |                |
|-------|------------------------|-------|----------------|
| 공 모 명 | 이용자 개인 맞춤형 식품 추천 서비스   |       |                |
| 성 명   | 팀명: 마, 이구미!   대표자: 최은선 | 지원 과제 | 데이터 분석<br>및 활용 |

## II. 세부 내용

### 1. 개요

#### 1-1. 데이터 분석 및 활용 방안의 핵심내용(요약)

- 데이터 분석 결과 및 활용 방안을 간략히 소개

저희 마, 이구미! 팀은 경상북도 농특산물 사이트 '사이소' 데이터 분석을 통해, 온라인 식품 쇼핑몰에서 판매하는 상품의 유형과 고객의 이용 패턴 데이터를 수집하였습니다.

그리고 이렇게 수집한 데이터를 분석하여, '이용자 개인 맞춤형 식품 추천 서비스' 모델링을 진행하였습니다.

구현한 개인 맞춤형 서비스를 기반으로 식품 쇼핑몰 사이트를 개별 고객마다 맞춤형으로 이용할 수 있도록, 다양한 각도에서 활용방안을 기획하였습니다.

#### 1-2. 주제 선정 및 배경(창의성)

- 주제 선정에 대해서 구체적으로 작성

(동기, 문제점, 필요성 등을 중심으로 설명)

(도표, 이미지, 영상 등 활용 가능)

##### ● 문제 현황 1

『'농산물 유통약자' 위한 대책 마련 시급하다』

'농산물 유통개혁'이 온-오프라인을 망라해 전체 농산물 거래 시장에서 진행되어야만 한다.

농민들이 생산한 농산물이 제 값을 받고 팔려 나가는 선순환 시스템을 하루 빨리 마련해야 한다.

[출처: 이병로, '농산물 유통약자' 위한 대책 마련 시급하다, 한국영농신문]

저희는 농특산물을 생산 및 제공하는 농업인 분들이 겪는 어려움을 깨닫고, 농산물 유통개혁이 그 분들에게 큰 도움이 될 수 있다는 것을 알게 되었습니다. 지역 농특산물 쇼핑몰의 활성화를 통해 농산물 유통약자들을 구제할 수 있을 것이라고 생각했기에, 추천 서비스 구현을 위해 노력하였습니다.

## ● 주제 설정 동기 1

『농업의 유통 판도 뒤바꿀 '온라인 유통 혁명'』

온라인 유통은 시간·공간적 한계를 초월해 저렴하고

편리한 가치를 소비자에게 전달하면서 최근 5년간 판매가 70% 증가했다.

온라인 유통 혁명의 시작은 여기에 있다.

**개별화된 소비자 요구를 파악한 효율적인 상품과 서비스 공급이 이뤄져야**

경쟁력을 확보할 수 있는 시대이기 때문이다.

특히 최근 코로나 사태 이후 비대면 거래가 폭발적으로 늘어나면서,

**온라인 유통혁명은 가속화될 것**이라는 게 업계 전망이다.

[출처: 석민정, 농업의 유통 판도 뒤바꿀 '온라인 유통 혁명'...스마트한 대응책 모색할 때, 농축유통신문]



위 기사에서 알 수 있듯, 최근 코로나 19로 인해 온라인 쇼핑몰에 대한 소비자의 관심이 눈에 띄게 증가하고 있음을 알 수 있습니다. 이처럼 치열해진 온라인 쇼핑몰 경쟁 속에서 쇼핑몰들은 소비자들의 요구와 선호를 적극적으로 파악할 필요가 있습니다. 저희 팀은 쇼핑몰 이용자 개인이 맞춤형으로 서비스를 이용할 수 있도록 함으로써 소비자들의 만족도를 높이고자 하였습니다.

## ● 주제 설정 동기 2

저희는 경상북도 농특산물 쇼핑몰 '사이소' 데이터를 분석하여, 기존의 지역 농특산물 쇼핑몰이 지닌 강점과 약점, 기회 요인과 위험 요인을 파악하였습니다.

### SWOT 분석이란?

**SWOT** (Strength, Weakness, Opportunity, Threat) **분석**은 내부 환경과 외부 환경 각각 요소를 바탕으로 현황을 분석하는 전략 수립 방법입니다.

SWOT 분석을 통해 비즈니스가 지닌 잠재적 가능성과 강점, 위험 요인 등을 발견할 수 있습니다.

#### STRENGTH

- 쇼핑몰이 지향하는 가치와 정체성이 뚜렷함
- 경상북도 특산물을 유통한다는 점에서 특수함을 지님
- 유통 과정을 단축하여, 농특산물 유통 과정 혁신에 기여함

#### WEAKNESS

- 판매 품목 유형 간 불균형
- 업체 간 수익 불균형
- 한정된 루틴의 마케팅 방식

#### OPPORTUNITY

- 코로나 19로 인한 온라인 식품 쇼핑몰의 주목도 증가
- 탄탄히 구축되어 있는 사이트 및 SNS
- 개인 맞춤형 추천 서비스 도입을 통한 수익 증대 및 이용자 만족도 향상 기대

#### THREAT

- 현재 사이트 운영 방식은 이용자들의 선택 범위를 한정 시킴
  - 이용자들의 흥미 감소
  - 사이트 이용 지속 기간 감소
  - 잠재 고객 확보 어려움
- 상품 종류, 가격 면에서 타 쇼핑몰에 비한 강점을 확보하기 어려운 경우, 사이트 지속 가능성 위협

SWOT 분석을 통해 사이소 사이트를 파악한 결과, 사이소는 높은 품질의 경상북도 농특산물을 보다 많은 사람들이 이용할 수 있도록 많은 노력을 하고 있음을 알 수 있었습니다. 온라인 쇼핑몰 사이트뿐만 아니라, 모바일 어플리케이션, SNS (Instagram), 네이버 쇼핑 LIVE (소셜 라이브 커머스), 인터파크와 같은 타 쇼핑몰과의 제휴 등을 통해 빠른 속도로 성장하고 있었습니다.

사이소는 쇼핑몰도, SNS 나 라이브 커머스 방송 시스템도 탄탄하게 구축되어 있는 쇼핑몰이라는 점에서 다른 쇼핑몰 사이트보다 확실한 강점을 가지고 있기에, 저희는 사이소 사이트 개선 방안을 도출하여 지역 농특산물 거래가 활성화되는 것을 지향하였습니다.

#### ● 문제 현황 2

SWOT 분석을 통해 사이소 사이트의 개선점을 찾을 수 있었습니다. 사이소 사이트의 SNS 를 살펴보면, 사이소의 마케팅 방식은 일정한 패턴을 가지고 있습니다. 특정 농특산물에 대한 라이브 커머스 방송 알림과 함께, 쿠폰과 특산물 증정을 통해 라이브 방송 시청 혜택을 제공하고 있습니다. 최근에는 '영천 딱딱이복숭아', '영주 승혜축산 한우', '청송 친환경 무농약 블루베리'를 소개하였습니다. 사이소 온라인 쇼핑몰 사이트에서는 월요일과 수요일 특가전, 그리고 신규 회원 가입 쿠폰 증정 혜택 등을 선보이고 있었습니다.

SNS에서는 타 게시물보다, 고객들이 적극적으로 참여하여 제품을 홍보할 수 있는 '체험단' 이벤트 게시물에 '좋아요 수'가 많은 것을 확인하여, 사이소가 진행하는 이벤트 중 고객 참여도가 높은 이벤트에 대한 수요가 있음을 파악하였습니다.

이처럼, 고객들의 소비 패턴은 '적극성'과 '자율성'을 지향하고 있습니다. 저희는 이러한 적극적이고 자유로운 소비를 지향하는데 현재 사이소가 진행하는 일정한 패턴의 마케팅 방식은 고객의 흥미와 쇼핑물을 지속적으로 이용하려는 동기는 충족시키기 어려울 것으로 판단하였고, 고객들이 쇼핑물을 계속해서 찾게 될 마케팅 방식으로 이용자 개인 맞춤형 식품 추천 서비스를 고안하게 되었습니다.

### 1-3. 기존 서비스와의 차별성 및 독창성(창의성)

- 기존 유사서비스와의 차별점에 대해서 구체적으로 작성

(기술적 우위, 소비자의 효용가치 등을 중심으로 설명)

- 분석 결과물의 창의적, 독창적인 기술, 서비스 등에 대해서 구체적으로 작성

(도표, 이미지, 영상 등 활용 가능)

#### ● 서비스 경쟁력 Check

저희는 사이소 쇼핑물의 경쟁 업체를 타지역 농특산물 쇼핑물로 타겟팅하였습니다.

전국 11개 시·도 지역 농특산물 쇼핑물과, 그외 경상북도 내 시·군·구 농특산물 쇼핑물 4곳, 타 지역 쇼핑물 17개의 리스트를 마련하여, 경쟁 쇼핑물 중 개인 맞춤형 식품 추천 서비스를 도입해 활용하고 있는 쇼핑물이 있는지 확인하였습니다. 그 결과, 경쟁 업체 중 여러 번의 시도는 있었으나, 아직 개인 맞춤형 식품 추천 서비스를 도입하여 활용 중인 쇼핑물은 존재하지 않음을 확인하였습니다.

#### ● 데이터 모델 소개

##### 데이터 모델링 소개

##### 콘텐츠 기반 필터링 (Content-based filtering)

상품 자체가 지닌 속성을 기반으로,  
특정 상품과 유사한 유형의 다른 상품을 추천하는 방식

##### 잠재요인 협업 필터링 (Latent Factor Collaborative filtering)

대규모 다차원 행렬을 SVD(차원 감소 기법)으로 분해하는 과정  
(Matrix Factorization)에서 잠재요인을 추출함  
사용자와 상품 평점 matrix 내에서 추출한 잠재요인을 기반으로  
상품을 추천하는 방식

##### 아이템 기반 협업 필터링 (Item-based Collaborative filtering)

최근접 이웃(Nearest Neighbor) 협업 필터링 중  
아이템 기반 협업 필터링 모델 구현

상품을 향한 고객의 평점, 관심도, 고객의 구매 이력 등  
사용자의 행동양식(User behavior)에 기반으로 함

사용자들이 상품에 대해 평가한 척도를 바탕으로,  
유사한 상품을 추천하는 방식

저희 서비스에서는 추천 성능을 높이고자, '콘텐츠 기반 필터링', '아이템 기반 협업 필터링', '잠재요인 협업 필터링'과 같이 여러 추천 모델을 결합하여, '하이브리드 추천 시스템'을 구현하였습니다.

## - 하이브리드 추천 시스템 (Hybrid Recommender Systems)만의 강점

### 1) 식품 쇼핑의 개인화 실현

협업 필터링을 이용하여 사용자 데이터를 철저하게 분석하고, 사용자 개인에게 딱 알맞은 상품을 추천해줄 수 있어 고객들의 만족도를 높일 수 있습니다.

### 2) 콜드 스타트 (Cold Start) 문제 해결

협업 필터링 추천 모델의 경우, 시스템이 신규 가입 고객에 대한 충분한 정보를 모으지 못했을 경우, 적절한 추천을 하지 못할 가능성이 큼니다. 따라서 신규 고객에게는 콘텐츠 기반 필터링 모델을 기반으로 추천 기능을 제공하는 것이

적절합니다.

### 3) 롱 테일(Long Tail) 법칙으로 인해 발생하는 수익 불균형 문제 해결

협업 필터링 모델의 경우, 사용자의 관심이 저조한 항목은

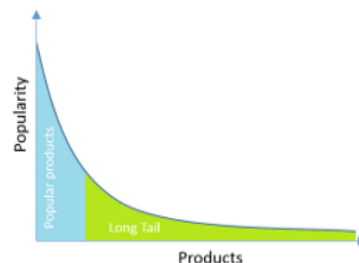
정보의 부족으로 추천 빈도수가 적어 고객의 관심으로부터 소외될 가능성이 있습니다.

따라서 콘텐츠 기반 필터링을 이용하여 보다 다양한 상품을 추천하여,

고객들의 시야를 확장할 수 있습니다.

\* 롱 테일 법칙: 전체 수익의 80%가 상위 20% 상품에서

기인함으로써 발생하는 수익 불균형 문제를 설명하는 법칙입니다.



하이브리드 추천 시스템은 여러 추천 시스템을 결합하였기에 확실한 강점을 가지고 있습니다.

저희는 여러 모델을 결합한 하이브리드 추천 시스템을 통해 개별 모델이 지닌 약점을 보완하여, 콜드 스타트와 롱 테일 문제를 해결하고, 식품 쇼핑의 개인화를 실현하고자 하였습니다.

## 2. 분석 방법

### 2-1. 활용 데이터

- 활용데이터를 요약적으로 제시
- 데이터 분석에 사용한 공공·민간데이터 출처 기재

저희 팀은 경상북도 공공데이터 경진대회 사이트(<http://www.gbdcon.kr/>)에서 제공된 경진대회용 메인데이터 '사이소 데이터'와 '사이소 분류정보'를 사용하였습니다.

(<http://www.gbdcon.kr/library/3>)

### 2-2. 데이터 분석(분석기법, 완성도)

- 데이터 분석 환경 및 도구, 전처리, 사용 기술 등 분석 과정에 대한 내용 작성

(도표, 이미지, 영상 등 활용 가능)

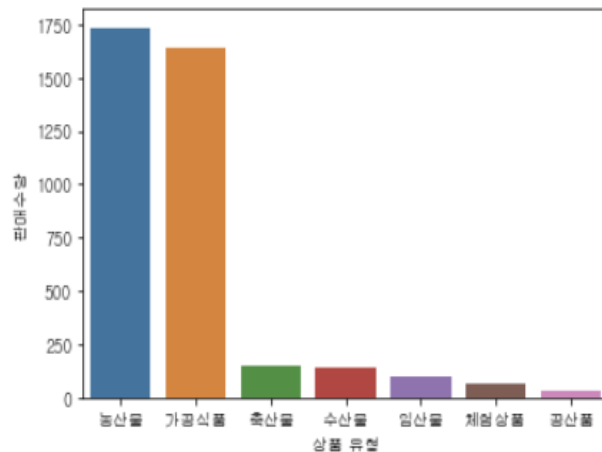
#### ● 데이터 분석 환경 및 도구

- 분석 환경: 구글 코랩 
- 분석 언어: 파이썬 
- 사용한 라이브러리: pandas, matplotlib.pyplot, numpy, sklearn, scipy

#### ● EDA

##### 1) 상품 유형 간의 불균형 발견

저희는 본격적인 데이터 분석에 앞서, 사이소의 상품 관리 테이블 EDA (탐색적 데이터 분석)를 진행하였습니다.



Python 의 Matplotlib 라이브러리를 이용해 상품 관리테이블의 '상품 유형'과 '판매수량' 간의 관계에 대하여 EDA 를 진행한 결과, 농산물, 가공식품 유형의 상품 판매수량이 다른 유형의 상품보다 판매수량이 월등히 높음을 알 수 있었습니다.

## 2) 업체 간의 불균형 발견

|     | 업체ID                                              | 판매수량 |
|-----|---------------------------------------------------|------|
| 503 | baffd9c61cf98685fa3508b7d77bcd087e7dfae0ce34ac... | 84   |
| 584 | dd10b7caf931dcd7bf81505d56531fedd94af61ee698ec... | 66   |
| 246 | 5ecdbdeca766d2bc85cd6c76c52221eb5f5c6082647ed8... | 65   |
| 604 | e970449a7fbb26c2239dc5152bf78ac93725a3ad5494af... | 62   |
| 249 | 5f90d34c68fb01f90df340c684d12fdb25346c06af0853... | 60   |

|     | 업체ID                                              | 판매수량 |
|-----|---------------------------------------------------|------|
| 505 | bbba8dc3d394f78b9e04f3fc908bef07683d6ac54189d8... | 1    |
| 227 | 56d7ad2b9fe1133fd91507e6510a99f81aa90dfaa0ec6f... | 1    |
| 508 | bc741afc77c27ca7748bb1d01e99eddb97aadce7818090... | 1    |
| 509 | bc741afc77c27ca7748bb1d01e99eddb97aadce7818090... | 1    |
| 656 | ff85e6c9fd60e14b4c2a82a307130075eeca9eb7a2da1d... | 1    |

또한, 업체 ID column 을 이용하여 전체 657 개의 업체 간의 판매수량을 비교한 결과, 상위 5% 업체와 하위 5% 업체의 판매수량 간의 차이가 확연히 드러났습니다.

## 3) 상품 관리테이블 EDA Insight

EDA 결과, 상품 유형과 업체 간의 판매 수량 불균형이 발견되었습니다.


사이소 사이트에서 판매 중인 전체 상품과, 전체 입점 업체 중 특정 유형의 상품과, 특정 업체에만 관심이 집중되고 있다는 것을 의미합니다. 이는 쇼핑몰의 지속 가능성을 저해하는 요소이므로, **보다 다양한 유형의 상품과 업체로** 고객의 관심도를 고르게 분포시킬 필요가 있습니다.

### ● 상품 관리테이블 상품 분류 정보 전처리

- 1) 특수한 상황이 아닌, 일반적인 상황에서 상품을 추천하는 환경을 구현하기 위하여 '기본분류', '2 차분류', '3 차분류' column 내 'zz' 값 (이벤트 상품)을 모두 제외하였습니다.

- 2) 상품 분류 정보를 parameter 로 이용하기 위해 '기본분류', '2 차분류', '3 차분류' column 을 '기본분류' column 으로 통합하였습니다.

| B    | C      | D    |
|------|--------|------|
| 기본분류 | 2차분류   | 3차분류 |
| zz   | 7020   |      |
| zz   | 201030 |      |
| zz   | 7020   |      |
| zz   | 3020   |      |
| zz   | 3020   |      |



| B      |
|--------|
| 기본분류   |
| 7020   |
| 201030 |
| 7020   |
| 3020   |
| 3020   |

<전처리 전>                      <전처리 후>

## ● Parameter 설정

### • 상품 관리테이블 column

콘텐츠 간 유사함의 기준이 되는, **상품 속성 parameter**: '기본분류', '분류명', '상품유형', '제조사'  
 개별 상품을 구분하기 위한 **고유 값 parameter**: '상품코드', '상품명', '업체ID'  
 고객의 선택을 기반으로 하는 **검색 필터링 parameter**: '판매가격', '기본배송비'

### • 회원 정보 테이블 column

회원 식별을 위한 **parameter**: '회원번호'

### • 장바구니 테이블 column

회원 구매 이력 추출을 위한 **parameter**: '회원ID', '상품코드', '상품명'

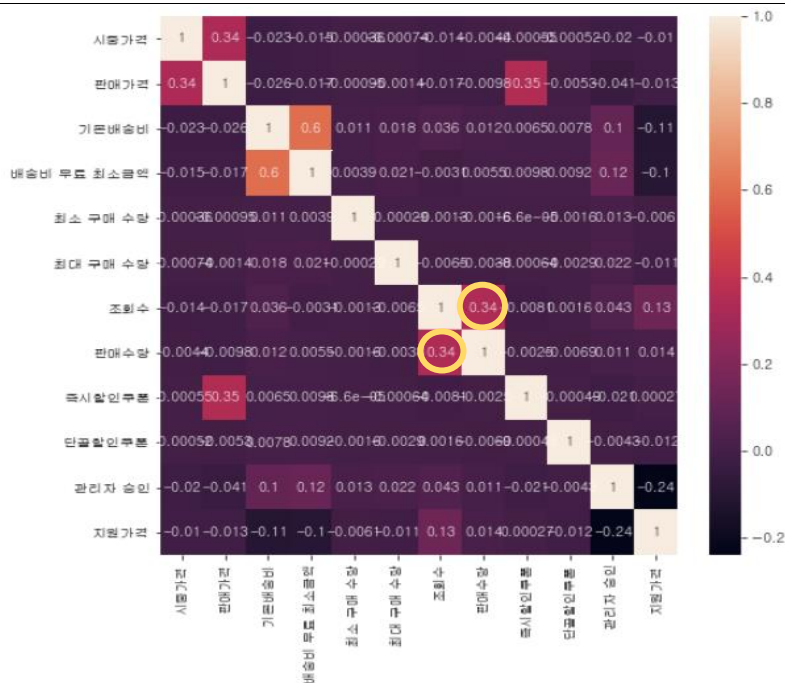
## ● 상품 선호도 테이블 구성

이용자 '개인 맞춤형' 식품 추천 서비스 구현을 위해서는, 상품을 향한 고객의 평점, 관심도 등의 사용자의 행동양식(User behavior) 데이터를 수집해야 합니다.

이 사용자의 행동양식을 파악하기 위하여, 해당 데이터 분석에서는 '선호도' 데이터를 얻고자 하였습니다.

heatmap chart 시각화를 통해 상품 관리테이블 parameter 간 상관관계를 분석하였습니다.





‘조회수’, ‘판매수량’ 간의 상관관계는 0.34 로, 두 parameter 간의 충분한 ‘상관관계가 있음’을 가정할 수 있었습니다.

‘조회수’, ‘판매수량’ parameter 를 이용하여 weighted rating 연산 과정을 수행하였고, 그 결과로 ‘선호도’ 데이터를 얻을 수 있었습니다.

$$\text{weighted rating (WR)} = (v \div (v+m)) \times R + (m \div (v+m)) \times C$$

- R: 개별 상품의 판매 수량 (Rating)
- v: 개별 상품에 대한 조회수 (Votes)
- m: 상위 500위 내에 포함되기 위한 최소 조회수 (minimum votes)
- C: 전체 상품의 평균 판매 수량

| 상품코드 | 기본분류       | 분류명    | 상품입장        | 상품명  | 제조사                                       | 판매가격     | 기본배송비 | 조회수  | 판매수량 | 업체ID | 선호도                                               |           |
|------|------------|--------|-------------|------|-------------------------------------------|----------|-------|------|------|------|---------------------------------------------------|-----------|
| 0    | 157092753  | 201030 | 꽃감/반건시/감말랭이 | NaN  | [상주전통꽃감] 전통꽃감 건시 4kg(65~70과)              | 상주전통꽃감   | 80000 | 0    | 0    | 0    | d3f908bbea5d2d7cac866508df9b2ffdf0088ab55d178...  | 22.961649 |
| 1    | 1578008979 | 201030 | 꽃감/반건시/감말랭이 | 농산물  | [구릿돌농원] 건시(3호 48구)                        | 구릿돌농원    | 55000 | 4000 | 130  | 0    | b4f5eabe0dd4a2502f409f9710682f2bbbc79de195121...  | 19.849017 |
| 2    | 1578031964 | 201030 | 꽃감/반건시/감말랭이 | 농산물  | [구릿돌농원] 건시(4호 36구)                        | 구릿돌농원    | 45000 | 4000 | 223  | 1    | b4f5eabe0dd4a2502f409f9710682f2bbbc79de195121...  | 18.306281 |
| 3    | 1578034429 | 201030 | 꽃감/반건시/감말랭이 | 농산물  | [구릿돌농원] 건시 낱개 포장 30구                      | 구릿돌농원    | 40000 | 4000 | 141  | 0    | b4f5eabe0dd4a2502f409f9710682f2bbbc79de195121...  | 19.623925 |
| 4    | 1578034546 | 201030 | 꽃감/반건시/감말랭이 | 농산물  | [구릿돌농원] 건시 낱개 포장1호                        | 구릿돌농원    | 30000 | 4000 | 136  | 0    | b4f5eabe0dd4a2502f409f9710682f2bbbc79de195121...  | 19.725603 |
| ...  | ...        | ...    | ...         | ...  | ...                                       | ...      | ...   | ...  | ...  | ...  | ...                                               | ...       |
| 3124 | 1604998092 | 80     | 김치/장류/참기름   | 가공식품 | 바론김장 1.8L                                 | 바론김장     | 24000 | 3000 | 186  | 1    | 926268d8bc66615558bacbd42caa33ac24a0dc7d032fd...  | 18.937150 |
| 3125 | 1606396480 | 80     | 김치/장류/참기름   | 가공식품 | 두레원무명청국장 1kg                              | 두레원식품    | 13000 | 3500 | 105  | 0    | a279427a10453ba64bc3ff604c06d63769647b4d7fa5d5... | 20.380307 |
| 3126 | 1606924853 | 80     | 김치/장류/참기름   | 가공식품 | 두레원무명청국장 200g                             | 두레원식품    | 3200  | 3500 | 194  | 45   | a279427a10453ba64bc3ff604c06d63769647b4d7fa5d5... | 27.140965 |
| 3127 | 1606196135 | b040   | 참기름         | 공산품  | [홍기연전 아이싱] 서각 큰 메밀배게                      | 홍기연전 아이싱 | 23000 | 0    | 3    | 0    | 1e5162a3dd797424426992b764688a7e7759b85a2597d9... | 22.878855 |
| 3128 | 1606296202 | e0     | 이전_농가채집     | NaN  | ★재550수요특가★ 갈고구마 상 10킬로 영주 농가 직전배출(하늘기둥농장) | 하늘기둥농장   | 43000 | 4000 | 3241 | 26   | 860422ba309816eacec97a4f09042e53fe1b2b1af39d8...  | 25.381132 |

3129 rows x 12 columns

필요한 parameter 들만 추출하여 구성한 상품 관리테이블 가장 오른쪽 column 에 각 상품에 대한 ‘선호도’ 데이터를 추가하였습니다.

### 3. 분석 결과(완성도)

- 분석 결과 제시

(도표, 이미지, 영상 등 활용 가능)

- **모델링 1 - 콘텐츠 기반 필터링 모델**

상품 간의 유사함을 측정할 수 있는 기준 parameter: '기본분류', '분류명', '상품유형', '제조사'

▶ 입력 받은 상품명을 기반으로, 상품과 유사한 다른 상품을 추천하는 함수

```
def get_recommend_merchandise_list(df, merc_name, top):  
  
    # 특정 상품과 비슷한 상품을 추천해야 하므로, '특정 상품' 정보 추출  
    target_merc_index = df[df["상품명"] == merc_name].index.values  
  
    # 코사인 유사도 중 비슷한 코사인 유사도를 가진 정보 추출  
    sim_index = category_c_sim[target_merc_index, :top].reshape(-1)  
    # 자기 자신을 제외  
    sim_index = sim_index[sim_index != target_merc_index]  
  
    # data frame으로 만들고, '선호도'로 정렬한 뒤 return  
    result = df.iloc[sim_index].sort_values('선호도', ascending=False)[:top]  
    return result
```

아래 예시는 '분류명'을 기준으로, '[구릿꿀농원] 건식 5 호 24 구' 상품과 유사한 상품 5 개를 추천한 것입니다.

분류명 간의 유사도는 코사인 유사도 (Cosine Similarity)를 통해 측정하였습니다.

```
# 임의의 상품명  
merc_name = "[구릿꿀농원] 건식5호 24구"  
top = 5  
  
get_recommend_merchandise_list(contents, merc_name, top)  
# 분류명을 기준으로, 위 상품과 유사한 상품 5개를 추천
```

|    | 상품코드       | 기본분류   | 분류명         | 상품유형 | 상품명                             | 제조사      | 판매가격  | 기본배송비 | 조회수 | 판매수량 | 업체ID                                              | 선호도       |
|----|------------|--------|-------------|------|---------------------------------|----------|-------|-------|-----|------|---------------------------------------------------|-----------|
| 0  | 157092753  | 201030 | 꽃감/썬건시/감말랭이 | NaN  | [상주전통꽃감] 전통꽃감 건식 4kg(65~70과)    | 상주전통꽃감   | 89000 | 0     | 0   | 0    | d3f908bbea5d2d7cac866508dff9b2ffdf0088ab55d178... | 22.961649 |
| 85 | 1600253088 | 201030 | 꽃감/썬건시/감말랭이 | NaN  | ★한가위함인판매★ 구릿꿀농원 썬건시2호(3.2kg40개) | 구릿꿀농원    | 50000 | 3000  | 130 | 0    | b4f5eabe0dd4a2502f409f9710682772bbcb79de195121... | 19.849017 |
| 83 | 1600043466 | 201030 | 꽃감/썬건시/감말랭이 | 농산물  | ★이웃사촌★ [선물용]서리꽃말랭이(1KG 20과~24과) | 내서 서리꽃마을 | 25500 | 4000  | 577 | 13   | 5f90d94c68fb01f90df940c684d12fdb25346c0af0853...  | 18.873547 |
| 84 | 1600129353 | 201030 | 꽃감/썬건시/감말랭이 | 농산물  | 프리미엄 개별건식 1호                    | 상주한사꽃감   | 50000 | 0     | 182 | 0    | 867288f154a3d1d3e3aa90236e341a3f82b7f69050d9fd... | 18.828098 |
| 73 | 1599056687 | 201030 | 꽃감/썬건시/감말랭이 | 농산물  | [상주아람꽃감]상주꽃감 오동선물세트(50g 48개)    | 상주아람꽃감농장 | 90000 | 0     | 248 | 0    | a37d03f03447c79d311ac6caa770bbdbcb75fc35d047b...  | 17.674287 |

분류명 외에도 상품의 다양한 속성을 기준으로 콘텐츠 기반 필터링을 이용한 추천 기능을 수행할 수 있다는 점에서 장점을 지닙니다.

## ● 모델링 2 - 아이템 기반 필터링 모델

### • 필요한 parameter 선정

회원 정보 테이블에서 '회원번호' 추출

상품 관리 테이블에서 '상품명' 추출

weighted rating 연산을 통해 얻은 '선호도' 추출

### - 아이템 기반 pivot 테이블 구성

index: 상품명 | columns: 회원번호 | data: 선호도

| 회원<br>번호                                                        | 1         | 60251 | 60603 | 60729 | 60846 | 61002 | 61062 | 61184 | 61484 | 61677 | 61772 | 61775 | 61891 | 61926 | 62191 | 62200 | 62201 |
|-----------------------------------------------------------------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 상품명                                                             |           |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| "임내<br>세요,<br>우리아<br>빠" 기<br>세용(곰<br>뽕이(곰<br>뽕이)진<br>액/액<br>기스] | 16.495395 | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   |
| (경북청<br>년농부<br>들) 민<br>서태버<br>섯농장<br>진술라<br>이스표<br>고버섯<br>150g  | 0.000000  |       |       |       |       | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   |
| (문경약<br>동화사)<br>문경약<br>동화지<br>LA식<br>갈비(냉<br>동)<br>1kg          | 0.000000  | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   |

아래 예시는 아이템 기반 테이블에서 상품 선호도 간의 유사도를 코사인 유사도 (Cosine Similarity)를 통해 측정하고, '피크닉사과 가정용 흙과 10kg' 상품과 유사한 상품 5 개를 추천한 것입니다.

```
# 임의의 상품명 입력
```

```
get_item_based_collabor('피크닉사과 가정용 흙과 10kg')
```

상품명

```

담운농장 큰머느리가 만든 현미누룽지, 흑임자[검정깨]누룽지,새싹보리누룽지,홍삼누룽지,천년초누룽지      0.169031
[하늘기둥농장] 간편 소포장 청국장 2킬로(200그램*10)                                       0.169031
명품만동소주 담금주 3600ml 30도                                                                0.169031
★이웃사촌★[달산마을]햇 알호두(30알)380g~400g                                              0.119523
[윤푸드] 이로우니 우리쌀 연근콘)까스 650g (130g*5개)                                         0.119523
[대우농장] 완숙 토마토 5kg (특품)                                                            0.119523

```

아이템 기반 필터링 모델에서 사용된 선호도는 '판매수량'과 '조회수'를 기반으로 계산되었으나, 선호도가 높은 상품만을 추천하는 것이 아니라, 사용자의 행동양식을 기반으로 유사한 선호도의 상품을 추천하므로 의미가 있습니다.

결과적으로 추천 시스템을 통해서 보다 다양한 상품을 이용자 개인 맞춤형으로 추천하는 것이 가능합니다.

### ● 모델링 3 – 구매이력 기반 잠재 요인 필터링 모델

고객의 구매 이력을 파악하기 위해 고객 – 상품 선호도 pivot 테이블 구성하였습니다.

| 상품코드   | 1577870016 | 1577933738 | 1577947433 | 1577948991 | 1577950662 | 1577971381 | 1578032329 | 1578043431 |
|--------|------------|------------|------------|------------|------------|------------|------------|------------|
| 회원번호   |            |            |            |            |            |            |            |            |
| 1      | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 60251  | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 60596  | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 60603  | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 60729  | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| ...    | ...        | ...        | ...        | ...        | ...        | ...        | ...        | ...        |
| 169279 | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 169280 | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 169282 | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 169286 | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |
| 169295 | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        | 0.0        |

8988 rows x 1590 columns

M 개의 상품에 대한 고객 N 명의 평균 선호도를 구한 뒤, 이 값을 위의 고객-상품 선호도 pivot 테이블 각각의 값에서 빼도록 합니다.

| 상품코드 | 1577870016 | 1577933738 | 1577947433 | 1577948991 | 1577950662 | 1577971381 | 1578032329 | 1578043431 |
|------|------------|------------|------------|------------|------------|------------|------------|------------|
| 0    | -0.273593  | -0.273593  | -0.273593  | -0.273593  | -0.273593  | -0.273593  | -0.273593  | -0.273593  |
| 1    | -0.071004  | -0.071004  | -0.071004  | -0.071004  | -0.071004  | -0.071004  | -0.071004  | -0.071004  |
| 2    | -0.088722  | -0.088722  | -0.088722  | -0.088722  | -0.088722  | -0.088722  | -0.088722  | -0.088722  |
| 3    | -0.023095  | -0.023095  | -0.023095  | -0.023095  | -0.023095  | -0.023095  | -0.023095  | -0.023095  |
| 4    | -0.083072  | -0.083072  | -0.083072  | -0.083072  | -0.083072  | -0.083072  | -0.083072  | -0.083072  |
| ...  | ...        | ...        | ...        | ...        | ...        | ...        | ...        | ...        |
| 8983 | -0.017102  | -0.017102  | -0.017102  | -0.017102  | -0.017102  | -0.017102  | -0.017102  | -0.017102  |
| 8984 | -0.018033  | -0.018033  | -0.018033  | -0.018033  | -0.018033  | -0.018033  | -0.018033  | -0.018033  |
| 8985 | -0.010583  | -0.010583  | -0.010583  | -0.010583  | -0.010583  | -0.010583  | -0.010583  | -0.010583  |
| 8986 | -0.009509  | -0.009509  | -0.009509  | -0.009509  | -0.009509  | -0.009509  | -0.009509  | -0.009509  |
| 8987 | -0.020094  | -0.020094  | -0.020094  | -0.020094  | -0.020094  | -0.020094  | -0.020094  | -0.020094  |

8988 rows x 1590 columns

Truncated SVD (Singular Value Decomposition, 특이값 분해)를 통한 행렬 분해 (Matrix Factorization) 과정을 거치고, 그 과정에서 추천의 기반이 되는 잠재요인을 추출합니다.

\* Truncated SVD:  $m \times n$  크기의 M 행렬을  $m \times m$  크기의 U 행렬,  $m \times n$  크기의 sigma 행렬,  $n \times n$  크기의 Vt 행렬로 분해한 뒤, 상위 p 개의 대각 성분을 추출하는 행렬 분해 방식

$$\begin{matrix} \text{Matrix} \\ M \end{matrix}
_{m \times n} =
\begin{matrix} \text{Matrix} \\ U \end{matrix}
_{m \times m}
\begin{matrix} \text{Matrix} \\ \Sigma \end{matrix}
_{m \times n}
\begin{matrix} \text{Matrix} \\ V^* \end{matrix}
_{n \times n}$$

생성한 U, sigma, Vt 행렬 shape 를 확인해 보았습니다.

```
[ ] from scipy.sparse.linalg import svds
    from scipy.linalg import svd

    U, sigma, Vt = svds(matrix_custom_mean, k=12)
    print(U.shape)
    print(sigma.shape)
    print(Vt.shape)

(8988, 12)
(12,)
(12, 1590)
```

U, sigma, Vt 행렬 내적을 수행하여 원본 행렬과 유사하게 복구한 뒤, 고객의 평균 선호도를 적용합니다.

```
[ ] svd_custom_pred_ratings = np.dot(np.dot(U, sigma), Vt) + custom_ratings_mean.reshape(-1,1)

[ ] df_svd_preds = pd.DataFrame(svd_custom_pred_ratings, columns = df_custom_merc_ratings.columns)
df_svd_preds
```

| 상품<br>코드 | 1577870016 | 1577933738 | 1577947433 | 1577948991 | 1577950662 | 1577971381 | 1578032329 | 157804... |
|----------|------------|------------|------------|------------|------------|------------|------------|-----------|
| 0        | 0.277008   | 0.277001   | 0.276990   | 0.277008   | 0.276682   | 0.277009   | 0.276836   | 0.27...   |
| 1        | 0.071255   | 0.071255   | 0.071256   | 0.071255   | 0.071274   | 0.071255   | 0.071258   | 0.07...   |
| 2        | 0.089616   | 0.089614   | 0.089611   | 0.089616   | 0.089859   | 0.089616   | 0.089554   | 0.08...   |
| 3        | 0.023322   | 0.023321   | 0.023321   | 0.023322   | 0.023323   | 0.023322   | 0.023319   | 0.02...   |
| 4        | 0.082882   | 0.082884   | 0.082888   | 0.082882   | 0.083058   | 0.082882   | 0.082940   | 0.08...   |
| ...      | ...        | ...        | ...        | ...        | ...        | ...        | ...        | ...       |

SVD 를 사용한 행렬 분해를 통해 데이터 변형을 완료하였으므로, 입력한 회원 번호에 해당하는 고객의 구매이력을 조회하고, 상품 선호도가 높은 데이터 순으로 정렬하여 추천 상품 리스트를 보여주는 함수를 생성합니다.

```
def recommend_merchandise(df_svd_preds, customer_id, ori_merc_df, ori_ratings_df, num_recommendations):

    # 최종적으로 만든 pred_df에서 고객 index에 따라 상품 데이터 정렬하기 -> 상품 선호도가 높은 순으로 정렬됨
    sorted_customer_pred=df_svd_preds.loc[df_svd_preds["회원번호"]==customer_id].set_index("회원번호").transpose().sort_values(by=customer_id,ascending=False).transpose()

    # 원본 선호도 데이터에서 회원번호에 해당하는 데이터 추출
    customer_data = df_ratings[df_ratings["회원번호"]==customer_id]

    # 추출한 회원 데이터와 원본 상품 데이터를 합침
    customer_history = customer_data.merge(df_merchandise, on='상품코드').sort_values(['선호도'], ascending=False)

    # 원본 상품 데이터에서 고객이 구매한 상품을 제외한 데이터를 추출
    recommendations = df_merchandise[~df_merchandise['상품코드'].isin(customer_history['상품코드'])]

    # 고객의 선호도가 높은 순으로 정렬된 데이터와, 위의 recommendation 데이터를 합침
    recommendations = recommendations.merge(pd.DataFrame(sorted_customer_pred.transpose()).reset_index(), on='상품코드')

    # 고객의 선호도가 높은 순으로 정렬된 데이터와, 위의 recommendation 데이터를 합침
    recommendations = recommendations.rename(columns={customer_id:"추천 예상"}).sort_values(by="추천 예상",ascending=False).head(num_recommendations)

    return customer_history, recommendations
```

아래 예시는 회원번호 '60729'인 고객님의 구매 이력과 상품 추천 리스트를 출력한 것입니다.

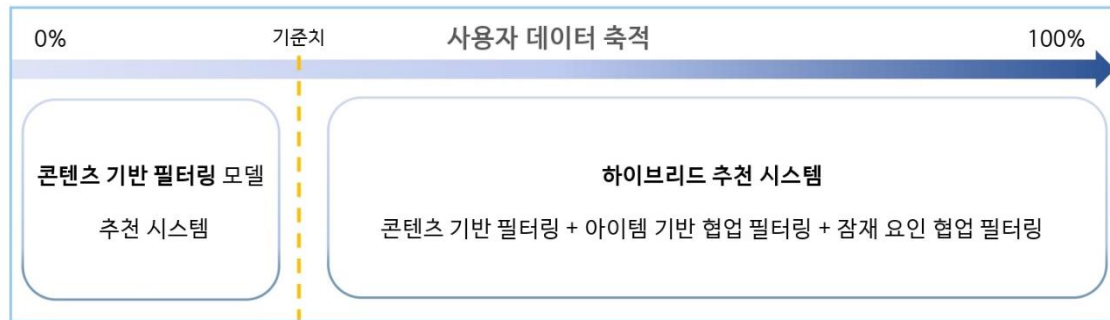
already\_rated (고객 구매이력 출력)

| 회원<br>번호 | 상품코드  | 선호도        | 상품명       | 주문수량                                    | 주문일자                | 배송일자                |
|----------|-------|------------|-----------|-----------------------------------------|---------------------|---------------------|
| 0        | 60729 | 1598427460 | 57.537618 | [마음드림사과] 2020 햇 사과, 가당 용 5kg 사과 20-23과나 | 2020-08-31 00:00:00 | 2020-09-01 00:00:00 |
| 37       | 60729 | 1598427460 | 57.537618 | [마음드림사과] 2020 햇 사과, 가당 용 5kg 사과 20-23과나 | 2020-08-05 00:00:00 | 2020-09-05 00:00:00 |
| 53       | 60729 | 1598427460 | 57.537618 | [마음드림사과] 2020 햇 사과, 가당 용 5kg 사과 20-23과나 | 2020-10-05 00:00:00 | 2020-10-06 00:00:00 |
| 60       | 60729 | 1598427460 | 57.537618 | [마음드림사과] 2020 햇 사과, 가당 용 5kg 사과 20-23과나 | 2020-09-31 00:00:00 | 2020-09-31 00:00:00 |

recommendation (구매 이력 기반의 상품 추천 리스트)

| 장바구니번호 | 상품코드    | 상품명                                                | 추천 예상    |
|--------|---------|----------------------------------------------------|----------|
| 10205  | 2649252 | 1.592374e+09 20년 속이 폭찬 유기농 양파 (중/대) 3kg, 5kg, 10kg | 0.100504 |
| 10153  | 2648439 | 1.592374e+09 20년 속이 폭찬 유기농 양파 (중/대) 3kg, 5kg, 10kg | 0.100504 |
| 10146  | 2646698 | 1.592374e+09 20년 속이 폭찬 유기농 양파 (중/대) 3kg, 5kg, 10kg | 0.100504 |
| 10147  | 2647097 | 1.592374e+09 20년 속이 폭찬 유기농 양파 (중/대) 3kg, 5kg, 10kg | 0.100504 |
| 10148  | 2648245 | 1.592374e+09 20년 속이 폭찬 유기농 양파 (중/대) 3kg, 5kg, 10kg | 0.100504 |

● 하이브리드 추천 시스템 모델 flow



#### 4. 활용 방안(시사성, 활용성)

- 분석 결과물의 사회 파급(기대)효과에 대해서 구체적으로 작성
  - 공공의 이익과 공동체 발전에 기여할 수 있는지 구체적으로 작성
  - 분석 결과의 활용방안 제시
  - 분석 결과물의 시장성 및 사업화 가능성에 대해서 구체적으로 작성
  - 분석 결과물의 구현, 실행, 매출 창출 구조 등 실현 가능성에 대해서 구체적으로 작성
- (도표, 이미지, 영상 등 활용 가능)
- 경상북도에 적용할 수 있는 방안에 대해서 작성

이용자 개인 맞춤형 식품 추천 서비스는 '식품 평점 데이터' 수집을 통해 성능을 개선할 수 있습니다.

해당 데이터 분석에서는 상품에 대한 고객의 '선호도' 데이터를 상품의 '조회수'와 '판매수량' parameter를 이용하여 구했습니다.

조회수와 판매수량 간의 상관관계를 살펴볼 때, 두 parameter 가 고객들의 선호도를 어느정도 뒷받침하고 있다고 볼 수 있으나, 각 식품에 대한 고객이 상품에 대한 후기와 평점을 남길 수 있는 기능을 추가적으로 구현한다면, 사용자 행동양식 데이터를 보다 다양한 관점에서 해석할 수 있게 될 것입니다.

따라서 상품 후기 및 평점 기능을 활성화하기 위해 쇼핑몰 모바일 어플리케이션에 후기와 평점 안내 팝업창을 띄우거나, 후기 및 평점을 남길 경우 쿠폰을 지급하는 이벤트를 진행하면 좋을 것 같습니다.



\* 평점 이벤트 팝업창 예시 이미지

또한, 신규 가입자로부터 몇 가지 데이터를 얻어 사용자 행동양식 데이터를 추가적으로 수집할 수 있습니다.

어떤 상품을 찾고 계신가요?

고객님의 관심사를 바탕으로 고객님의 맞춤형 상품을 추천해드립니다.

- |                                            |                                     |
|--------------------------------------------|-------------------------------------|
| <input type="checkbox"/> 쌀/잡곡              | <input type="checkbox"/> 차          |
| <input type="checkbox"/> 과일류               | <input type="checkbox"/> 김치         |
| <input type="checkbox"/> 채소류               | <input type="checkbox"/> 장류         |
| <input type="checkbox"/> 견과류               | - 된장/청국장, 고추장, 양념류, 참기름/들기름, 간장류    |
| <input type="checkbox"/> 축산물 - 한우, 돈육, 닭고기 | <input type="checkbox"/> 주류/전통주류/와인 |
| <input type="checkbox"/> 계란                | <input type="checkbox"/> 특산물        |
| <input type="checkbox"/> 수산물               | <input type="checkbox"/> 가공식품       |
| <input type="checkbox"/> 액젓/젓갈             | - 원액/청/음료, 과일말랭이/칩, 면류, 가루식품류       |
| <input type="checkbox"/> 꿀/홍삼              | <input type="checkbox"/> 기타         |

\* 신규 가입자 관심 분야 수집 페이지 예시 이미지

식품 추천 서비스를 기반으로 추천된 맞춤형 식품 외에도 상품 데이터를 추가적으로 활용하여 다양한 추천 카테고리를 개설할 수 있습니다.

이처럼, 여러 개의 카테고리로 구성된 추천 식품 리스트를 구성한다면, 평점 데이터뿐만 아니라 사용자 행동양식 데이터를 더욱 다양한 관점에서 해석할 수 있게 된다는 점에서 장점이 있습니다.



## # 핫 트렌드 실시간 인기상품



[두두네살이] 원주 물감외  
10kg 가정을 특가  
25,000원

[기원베이오] 미루담 청국장 발  
효자 3g X 30티백 ...  
9,000원

★이웃사랑★ 풍각마늘 반부레  
핀도복숭아 3.5kg(30...  
12,600원

[세연농원] 청송시과 껌실팩의  
는 얼음공 물 부사 가정을...  
38,000원

## # 수요일 특가전 할인 상품



[그랑농원] 직접재배 자연  
그대로 사과 100%...  
26,800원

물론도정삼지기  
삼문부지경이 300g...  
43,000원

문온 대지 닭고기 500g  
3,800원

래방류  
5,000원

## # '김철순' 님이 구매하신 [구릿뜰농원] 건시5호 24구 상품과 유사한 상품



[경원농원] 상주국감  
건시 1.5kg(4외)(36개)...  
60,000원

[아이스홍시] 청도영농조합  
아이스홍시개별포장...  
42,000원

반건시 1.5kg(개당50-  
60g 30개)  
42,500원

대봉 반건시 국감 1kg  
실속형  
19,900원

## # '김철순' 님을 위해 준비했어요! 스페셜 추천 상품



[통요로농장] 감시골  
당마음 2박스(100개)...  
45,000원

[정원농원] 100% 저온작업  
마루로도름 50개...  
35,000원

[술향기] 친환경 홍감자  
5kg (150g/개) 8개  
18,000원

[스파부드]  
서리대들깨가루 500g...  
15,400원

## # 코로나 19 유통 피해 농특산물 특별 판매전



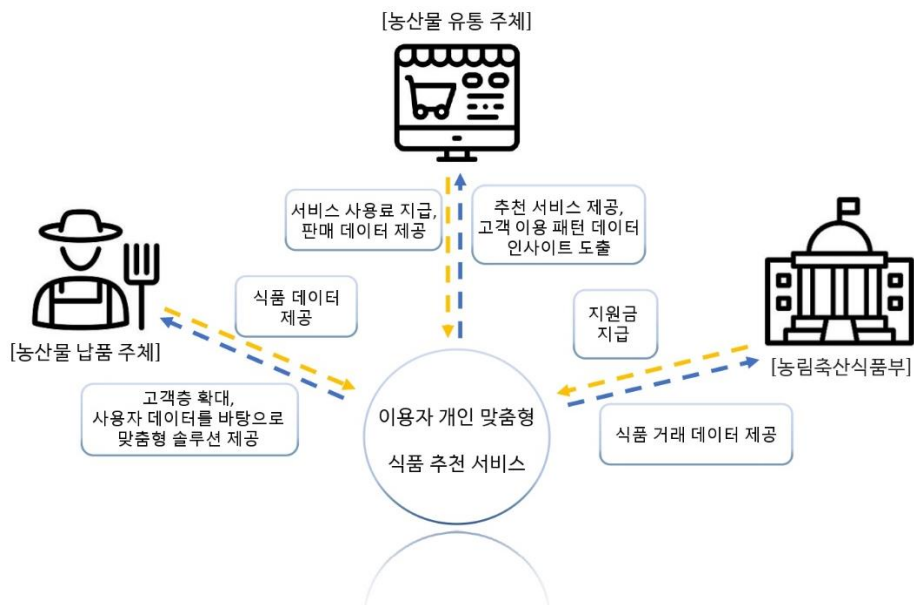
★유독피해특관전★물  
품도 터덕 산터덕(중)...  
34,800원

★유독피해 특관전★  
[재일소망] 울릉도 명이...  
47,400원

★유독피해 특관전★  
[별고울장외] 상주장외...  
19,800원

★유독피해  
특관전★[작관용이바]작  
관용이바용이바...  
11,900원

## ● 수익 구조





## ● 서비스 가치

- 맞춤형 식품 추천 서비스를 통해 보다 다양한 상품과 업체를 고객에게 추천할 수 있게 됨으로써, 농특산물 유통 약자를 보다 폭넓게 구제하는 것이 가능해집니다.
- 상품의 공급량과 유통 가격을 고려하면, 상품 유형과 가격 측면에서 타 지역 농특산물 사이트에 비한 강점을 확보하는 것에 한계가 있을 수 있습니다. 따라서 사이트 운영 방식에서 차별점을 주는 것이 효과적이라고 판단하였습니다. 국내 지자체에서 운영하는 농특산물 쇼핑몰에서는 아직 식품 추천 기능을 도입한 사례가 없으므로, 식품 추천 서비스는 확실한 강점이 될 것이라고 생각합니다.
- 맞춤형 식품 서비스를 통해 이용자의 구매 패턴을 보다 구체적으로 파악할 수 있으며, 다양한 관점에서 식품 수요를 파악하여, 쇼핑몰 내 판매하는 상품 유형의 다양화를 실현할 수 있습니다.
- 항상 한정된 패턴으로 쿠폰을 지급하는 것은 이용자들의 쿠폰 사용 동기를 감소시키며, 반복되는 패턴의 마케팅 방식은 쇼핑몰에 대한 이용자의 흥미를 감소시킬 수 있습니다. 따라서, 추천 서비스를 통해 고객들의 선호를 파악하여, 개별 고객에게 꼭 필요한 쿠폰을 지급한다면, 쿠폰 이용과 함께 상품 구매가 활성화될 것으로 기대됩니다.
- 뿐만 아니라 추천 서비스는 쇼핑몰에서 상품을 판매할 때, 다양한 선택을 가능하게 합니다. 특정 상품을 구매하는 고객 간의 유사성을 파악하고, 해당 고객 집단이 선호하는 유형의 상품을 파악하여 묶음 상품을 효율적으로 구성할 수 있게 됩니다.