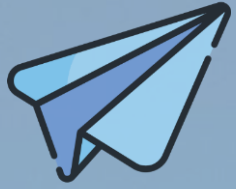


날씨 데이터를 활용한 모바일 전통시장 ‘또바기’



언제나 한결같이,
이제 당신의 곁으로





Contents



공모배경



데이터 분석



활용방안 및 기대효과



데이터
설명 및 전처리



서비스 활용 방안





공모배경



공모배경



코로나로 인한 전통시장 매출 하락

〈코로나19 1년, 서울 골목상권 59% 매출 하락...
총매출 19.6% 감소〉

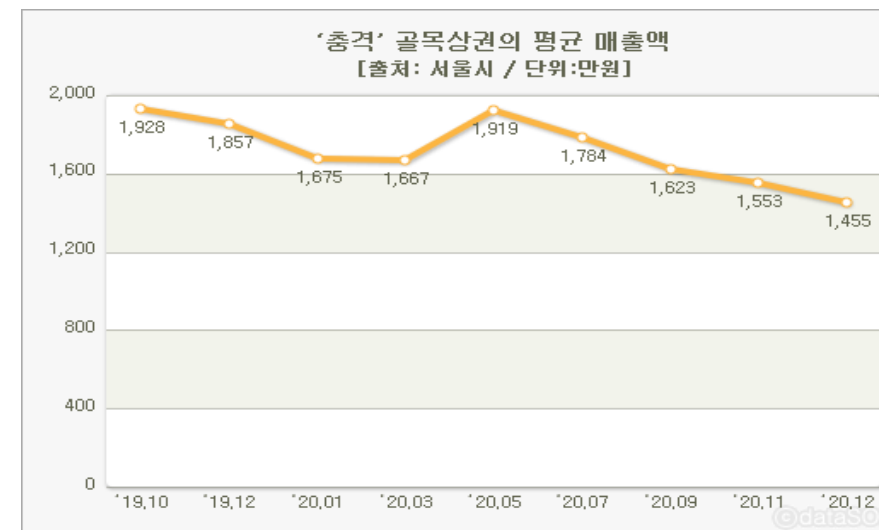
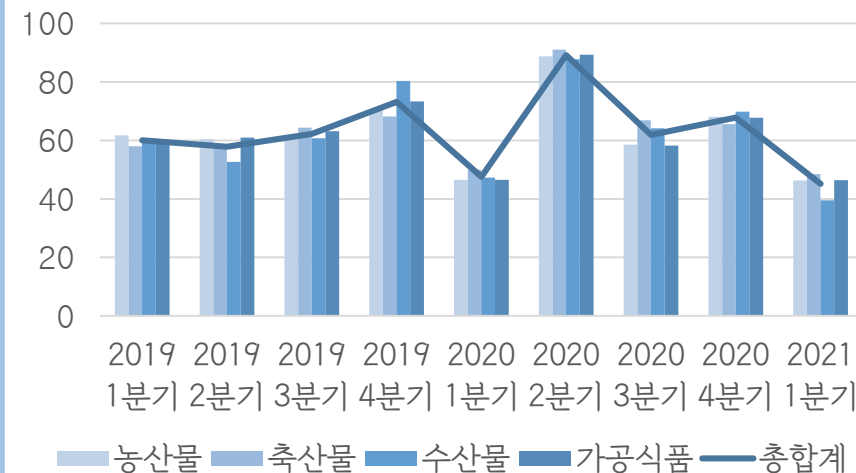
분석 결과 코로나 이전 대비 골목상권 총매출은 약 2조원에서 1.6조
원으로, 월 평균 점포당 매출도 1천9백만 원에서 1천7백만 원으로 각
각 19.6%, 13.8% 감소했다.

월별로 보면 월간 총매출액이 코로나19 1차 지역 확산 국면이던 작
년 1~3월까지 감소하다 재난지원금이 지급된 4~5월엔 일부 반등했
다. 이후 작년 하반기 지역확산 시기 때마다 매출액이 감소한 것으로
나타났다.

<http://www.datasom.co.kr/news/articleView.html?idxno=112942>

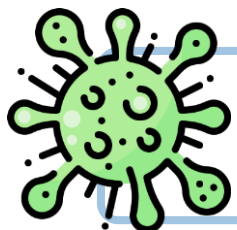
KOSIS 국가통계포털
Korean Statistical Information Service

〈전통시장 체감지수〉



2020-2분기는 5월 재난지원금 지급으로 인한 매출 증가로 예상됨

공모배경



포스트 코로나 시대의 온라인 소비

코로나19는 반강제로 우리 생활을 비대면으로 전환시켰으며, 오프라인 쇼핑을 더 즐기던 50·60세대도 온라인 쇼핑을 경험하는 계기가 됨

- 그동안 온라인 쇼핑은 디지털 환경과 언택트 소비문화에 익숙한 MZ세대*의 전유물이었으나 50·60대가 온라인을 체험하면서 부정적 인식이 바뀌는 계기가 됨

온라인 구매가 증가하면서 구매 연령과 품목에도 변화가 일어나고 있어, 향후 소비패턴은 기존과 다른 형태를 보일 것으로 예상됨

- 50대 이상 국내 소비자의 온라인 구매 증가율은 2020년 2월부터 3월 초까지 한 달간 생필품과 생활용품, 식품에서 가장 큰 폭 증가
- 미국에서 식료품 대리구매 및 배달, 픽업 서비스를 제공하는 업체인 인스타카트는 미국 전체 가구의 20%가 5년 안에 온라인 식료품 구매를 할 것으로 예상¹⁵⁾

출처 - 포스트코로나 시대 언택트 소비로 인한 소매공간 수요변화와 시사점



코로나 이후에도 온라인 소비의 규모가 커질 것으로 예상됨



공모배경



온라인 유통시장 규모 매년 증가

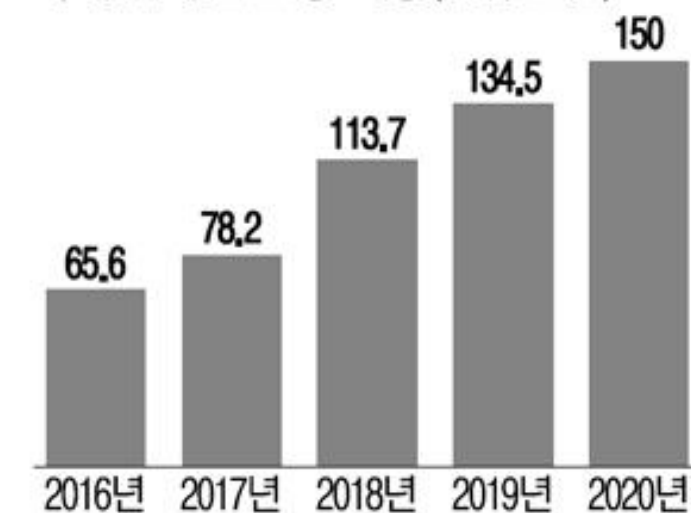
〈코로나 특수 맞은 '온라인 식품시장'... 올해 40조원 돌파〉

올해 코로나바이러스 감염증(코로나19)으로 비대면 소비가 활성화되자 온라인 식품 시장이 크게 성장한 것으로 나타났다.

28일 통계청에 따르면 지난 1~10월 온라인 식품시장 거래액은 34조6000억원으로, 작년 동기 대비 60.3% 증가했다. 이는 음·식료품, 농·축·수산물, 음식 서비스(피자·치킨 등 배달 서비스)의 인터넷·모바일쇼핑 거래액 기준이다.

https://biz.chosun.com/site/data/html_dir/2020/12/28/2020122800943.html

<https://www.mk.co.kr/news/business/view/2020/11/1178514/>
국내 온라인 쇼핑 시장 (단위=조원)



*올해는 업계 추정, 자료=통계청

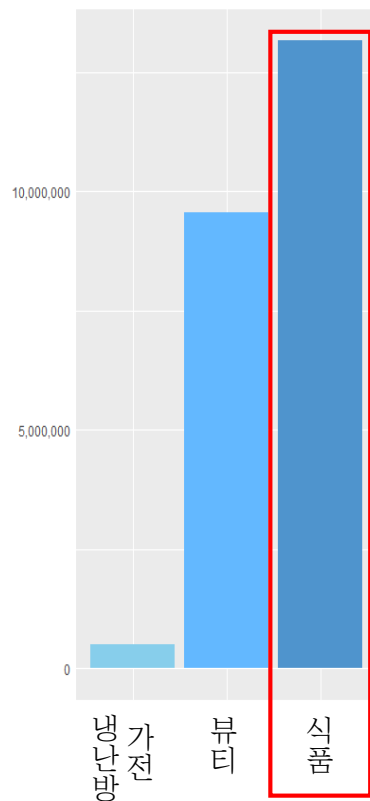


〈주제〉

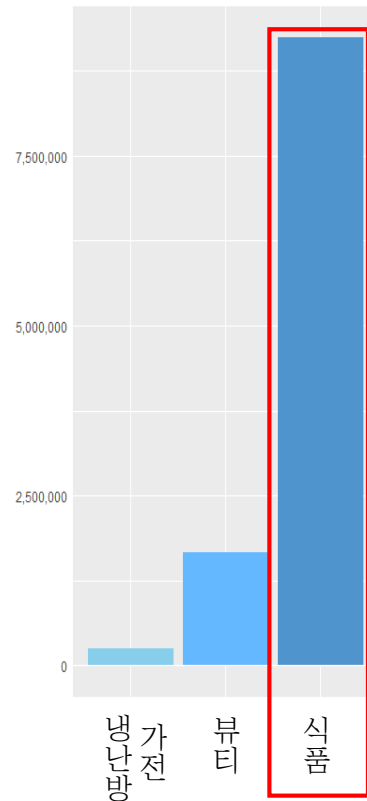
전통시장 애플리케이션 개선을 통한 매출 확대와 소상공인 지원

공모 방향 및 목표

구매건수



SNS 언급건수



온라인 구매 이력 데이터, 소셜데이터를 대분류 별로 확인,
두 데이터 모두 “식품”이 압도적

➔ 식품 관련 방향 설정.

〈 목표 〉

기상정보를 통한 제품의 판매량 예측

- 데이터 : 기상관측 데이터, 온라인 구매이력 데이터
- 분석기법 : 로지스틱 회귀분석, **Randomforest**
XGboost, Adaboost, RNN

온라인 데이터를 활용한 연관 상품 추천

- 데이터 : 온라인 크롤링 데이터
- 분석기법 : **KoNLPy** 형태소분석, 불용어 제거, **Word cloud**,
연관어분석, 토픽모델링



데이터 설명 및 전처리



활용 데이터



〈 SNS데이터 〉

- 언급일자 : 2018년 1월 1일 ~ 2019년 12월 31일
- 소분류 : 212개 데이터
- 언급건수



〈 구매이력 데이터 〉

- 구매일자 : 2018년 1월 1일 ~ 2019년 12월 31일
- 성별 : F, M
- 연령대 : 20대, 30대, 40대, 50대, 60대
- 소분류 : 212개 데이터
- 구매건수

기상자료개방포털



〈 미세먼지(PM10) 〉

- 일자 : 2018년 1월 1일 ~ 2019년 12월 31일
- 미세먼지 농도 : Min. $7.0\mu\text{g}/\text{m}^3$
Max. $193.0\mu\text{g}/\text{m}^3$
Mean $40.9\mu\text{g}/\text{m}^3$



〈 종관기상관측(ASOS) 〉

- 일자 : 2018년 1월 1일 ~ 2019년 12월 31일
- 기온 : Min. -14.7°C Max. 33.7°C Mean 13.3°C
- 습도 : Min. 22.3% Max. 95.9% Mean 57.1%
- 강수량 : Min. 0.0mm Max. 96.5mm Mean 3.0mm
- 풍속 : Min. 0.7m/s Max. 5.9m/s Mean 1.8m/s

데이터 전처리



변수 제거



buy2018_1
buy2018_2
buy2019_1
buy2019_2



buydata



▶ buy_20	232858 obs. of 6 variables
▶ buy_30	280650 obs. of 6 variables
▶ buy_40	280541 obs. of 6 variables
▶ buy_50	251166 obs. of 6 variables
▶ buy_60	184604 obs. of 6 variables

〈 연령대별 데이터 추출 결과 〉

연령대 데이터 간의 개수 차이 존재



특히, 60대의 데이터의 차이가 심함.

60대 데이터 제외.





데이터 전처리



변수 추가

중분류

중분류	소분류						총 개수
건강식품	감마리놀렌산 영양제	건강즙	건강즙/녹용	...	홍삼절편/홍삼정과	환자식	35
건해산물	건어물 건새우	건어물 노가리	건어물 마른오징어	...	해조류 미역	해초류	10
견과/견과	감말랭이	견대추	견망고	...	반건조고구마	혼합견과	17
⋮	⋮						
차	곡물차	과실차	녹차	...	허브차	홍차	22
채소	감자	고추/피망/파프리카	기타 농산물	...	햇개/가시오가피	호박	28

소분류 상위 카테고리 필요 →

대형마트의 카테고리를 참고하여

중분류를 데이터에 추가



데이터 전처리



결측치 제거



2018년 일별 기상데이터
2019년 일별 기상데이터



일자	기온	강수량	풍속	습도
2018-01-01	-1.3	NA	1.4	39.1
2018-01-02	-1.8	NA	1.8	42.0
2018-01-03	-4.7	NA	2.2	42.3
2018-01-04	-4.7	NA	1.4	43.0
2018-01-05	-3.0	NA	1.7	48.4
2018-01-06	-2.8	NA	1.5	45.8
2018-01-07	-0.8	NA	0.8	42.1
2018-01-08	1.3	0.9	1.8	51.9

〈 일자별 데이터 추출 결과 〉

강수량 데이터에 결측치 존재



기상청 예보 확인

일요일	월요일	화요일	수요일	목요일	금요일	토요일
	1일	2일	3일	4일	5일	6일
	평균기온:-1.3℃ 최고기온:3.8℃ 최저기온:-5.1℃ 평균운량:1.0 일강수량:-	평균기온:-1.8℃ 최고기온:1.8℃ 최저기온:-4.3℃ 평균운량:2.3 일강수량:-	평균기온:-4.7℃ 최고기온:-0.4℃ 최저기온:-7.1℃ 평균운량:3.6 일강수량:-	평균기온:-4.7℃ 최고기온:-0.7℃ 최저기온:-8.7℃ 평균운량:6.5 일강수량:-	평균기온:-3.0℃ 최고기온:1.6℃ 최저기온:-5.6℃ 평균운량:3.6 일강수량:-	평균기온:-2.8℃ 최고기온:2.9℃ 최저기온:-6.8℃ 평균운량:0.0 일강수량:-
7일	8일	9일	10일	11일	12일	13일
평균기온:-0.8℃ 최고기온:2.8℃ 최저기온:-5.7℃ 평균운량:7.5 일강수량:-	평균기온:1.3℃ 최고기온:4.0℃ 최저기온:-1.2℃ 평균운량:9.0 일강수량:0.9mm	평균기온:-4.2℃ 최고기온:-1.2℃ 최저기온:-6.1℃ 평균운량:3.5 일강수량:0.5mm	평균기온:-7.5℃ 최고기온:-4.8℃ 최저기온:-10.3℃ 평균운량:0.9 일강수량:0.3mm	평균기온:-11.1℃ 최고기온:-7.4℃ 최저기온:-13.6℃ 평균운량:0.4 일강수량:-	평균기온:-10.2℃ 최고기온:-5.4℃ 최저기온:-15.3℃ 평균운량:2.3 일강수량:0.0mm	평균기온:-4.4℃ 최고기온:-1.2℃ 최저기온:-6.6℃ 평균운량:8.3 일강수량:0.4mm

〈 해당 일자의 기상청의 예보 〉

NA = 무강수임을 확인



결측치 = 0으로 처리



데이터 전처리



결측치 제거



2018_미세먼지
2019_미세먼지

일자	미세먼지 농도
2018-02-23	NA
2018-02-28	NA
2018-03-05	NA
2018-03-09	NA
2018-03-16	NA
2018-03-25	NA
2018-03-26	NA
2018-04-04	NA

결측치 행 출력

일자	미세먼지 농도
2018-02-23	41
2018-02-28	55
2018-03-05	17
2018-03-09	17
2018-03-16	17
2018-03-25	38
2018-03-26	23
2018-04-04	41

결측치 대체

[PM10 : 1시간 평균값] 서울(108) 2018년 02월 28일															
날짜	01 시	02 시	03 시	04 시	05 시	06 시	07 시	08 시	09 시	10 시	11 시	12 시	13 시	14 시	15 시
2018.02.28.	89	103	95		.										
2018.02.27.	70	75	80	64	60	65	67	69	78	80	90	97	93	68	61
2018.02.26.	18	34	36	39	41	34	36	33	35	72	90	91	83	74	70
2018.02.25.	31	40	45	38	42	42	43	39	26	39	63	79	62	32	35
2018.02.24.	68	93	94	96	93	86	81	86	84	76	75	74	73	64	48
2018.02.23.	.								.	51	81	103	120	137	130
2018.02.22.	21	27	27	37	40	41	30	37	39	54	63	78	68	59	45

〈 해당 일자의 기상청 미세먼지 측정값 〉

미세먼지 데이터에
결측치 존재



“Random Forest” 를 통해
결측치 처리



데이터 전처리



dummy 변수화

성별

성별
F
M



여성	남성
1	0
0	1

연령대

연령대
20
30
40
50



20대	30대	40대	50대
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

범주형 데이터인 성별, 연령대를
양적 데이터로 변환



다중 공선성의 문제 발생



남성, 50대 열 삭제





최종 데이터



구매이력 데이터



종관기상관측 데이터
(ASOS)



미세먼지 데이터
(PM 10)



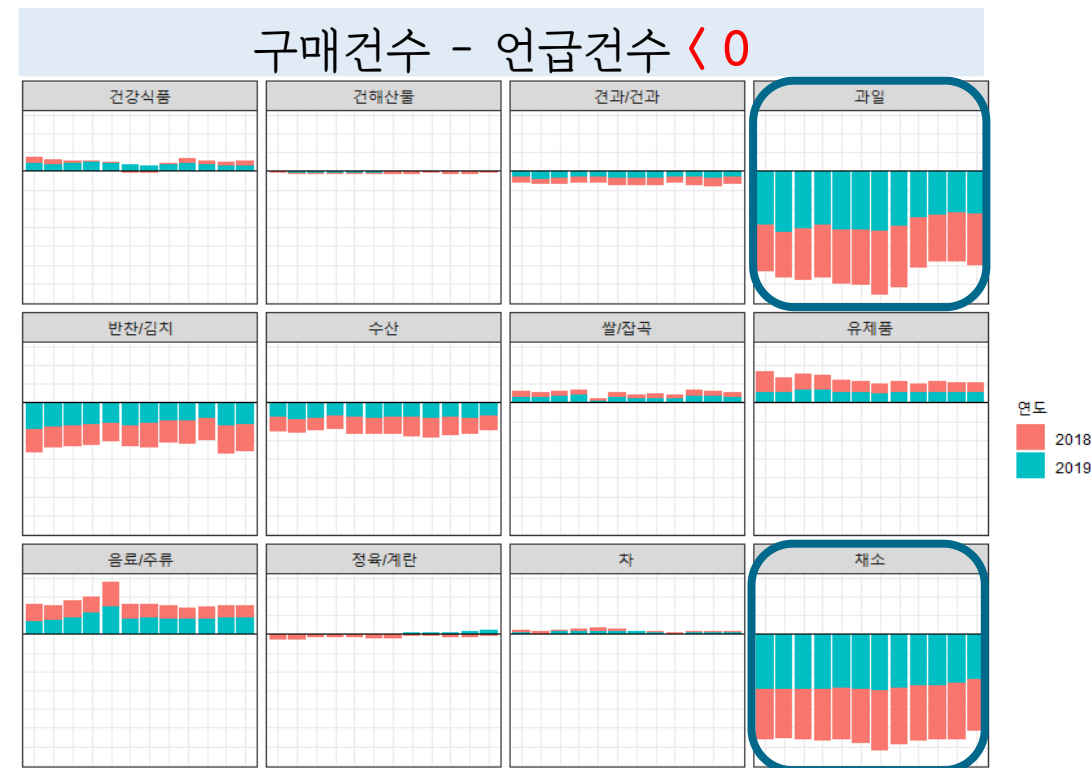
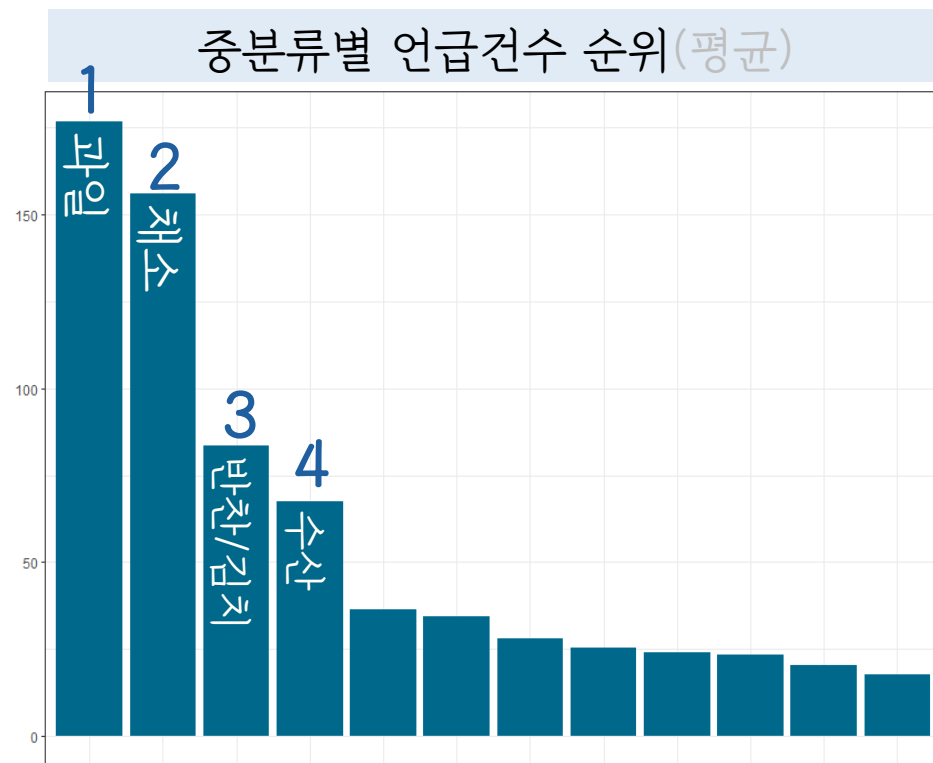
buy_total_data

buy_total_data

중분류	소분류를 구분하는 상위 카테고리
소분류	구매건수를 구분하는 카테고리
구매건수	어떤 제품을 구매한 건수의 일 합계
여성	성별이 여성인 경우 1로 표시됨
남성	성별이 남성인 경우 0으로 표시됨
20대	연령이 20~29세의 구매자 여부
30대	연령이 30~39세의 구매자 여부
40대	연령이 40~49세의 구매자 여부
50대	연령이 50~59세의 구매자 여부
기온	평균 기온
습도	상대 습도
강수량	일일 강수량
미세먼지농도	일 평균 미세먼지 농도
풍속	일 평균 풍속

EDA with Plot

< SNS 데이터 >



언급건수 데이터가 온라인 구매내역에 직결되지 않는다고 판단,

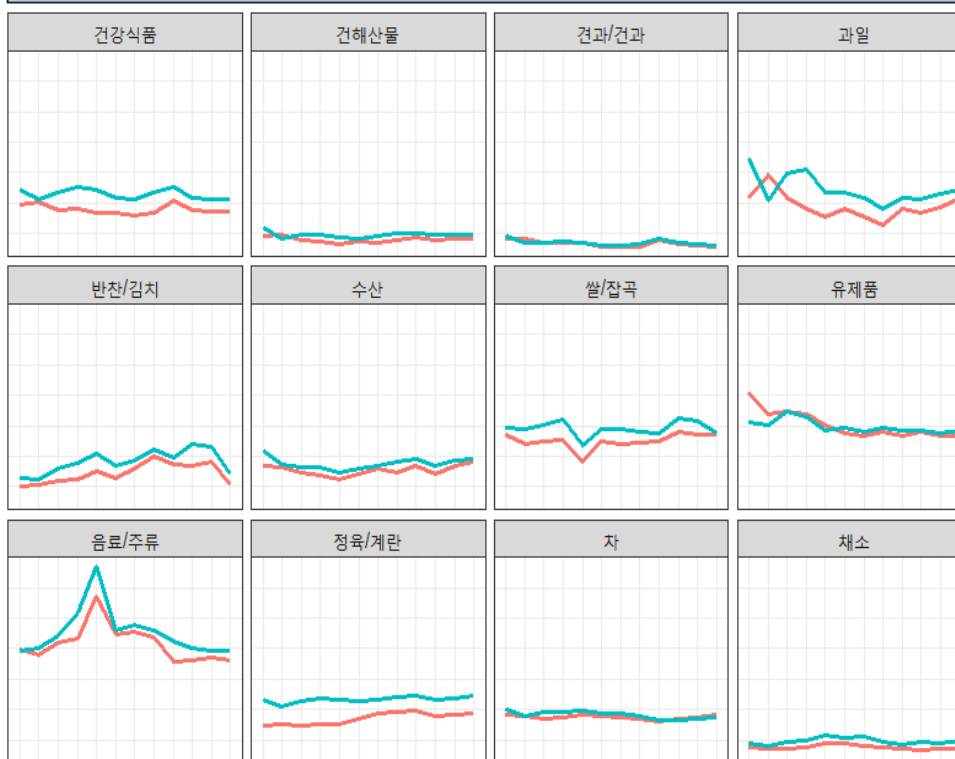
SNS데이터 제외

EDA with Plot

< 식품 구매이력 데이터 >

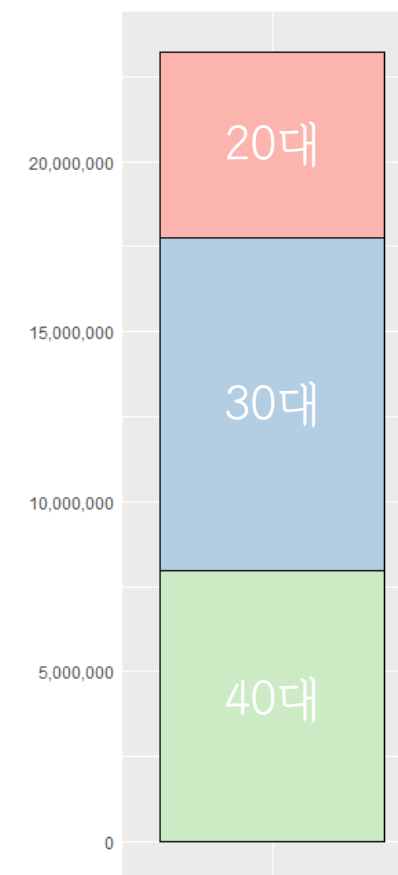


월별 각 중분류별 평균 구매건수



연도
— 2018
— 2019

여성의 식품 구매건수
연령대 분포





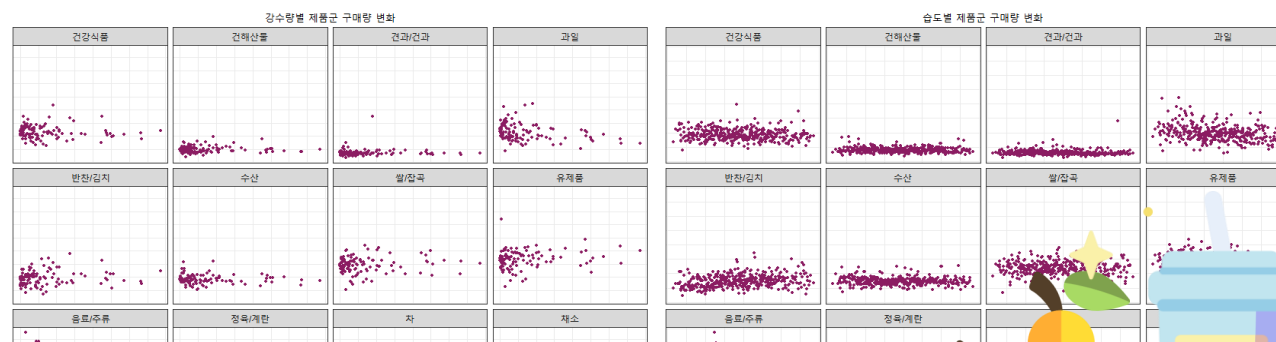
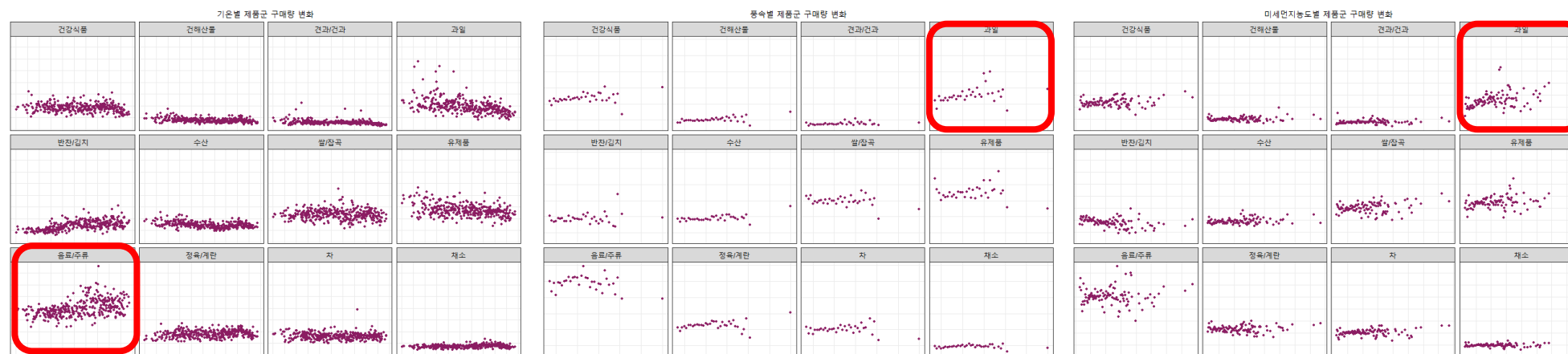
데이터 분석



중분류 선택

판매량 예측 시스템

구매이력데이터 + 기상데이터



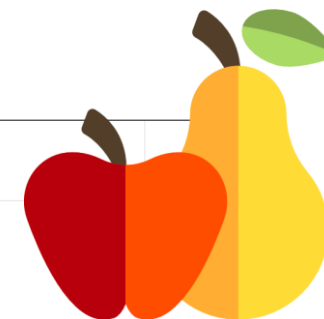
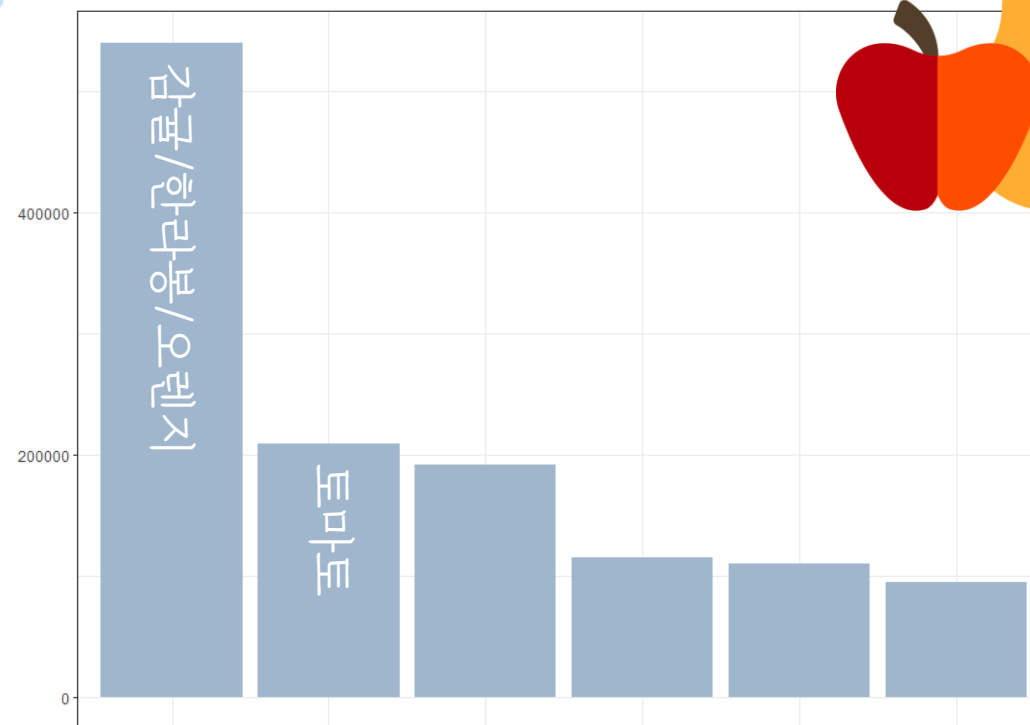
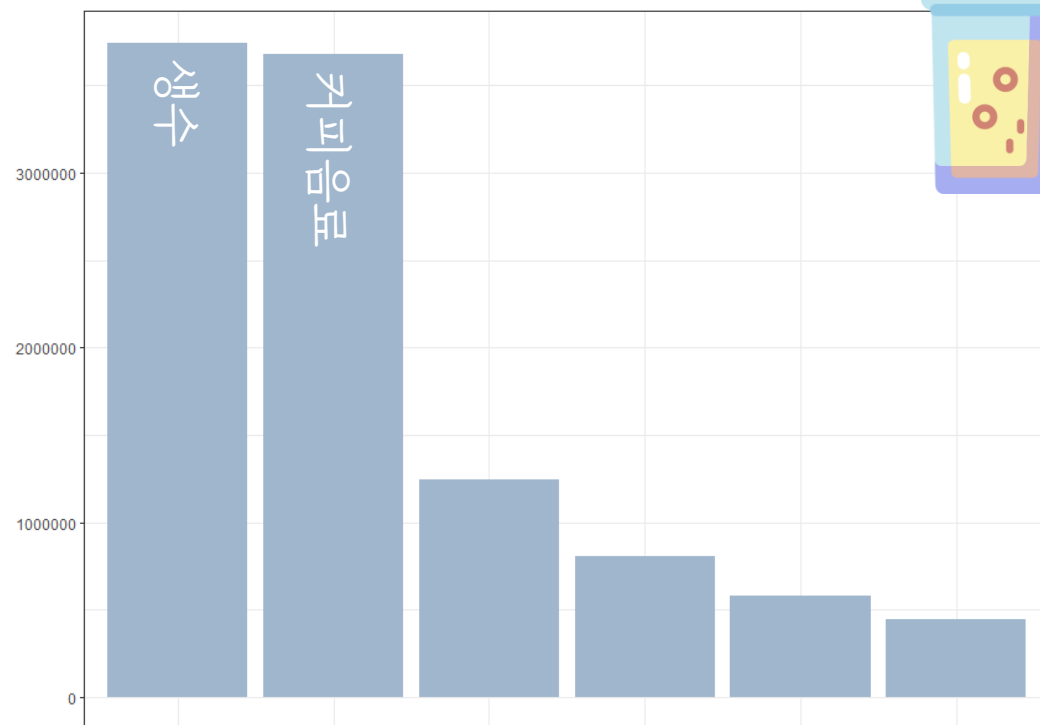
음료/주류, 과일 선택



소분류 선택

판매량 예측 시스템

구매이력데이터 + 기상데이터



음료/주류, 과일 선택

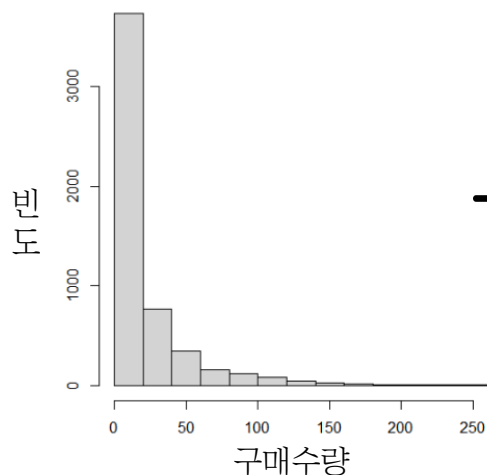


변수 선택

판매량 예측 시스템

구매이력데이터 + 기상데이터

histogram



소분류 < 수박 > 예시

번호	수량
0	5개 미만
1	5개 이상 10개 미만
2	10개 이상 15개 미만
3	15개 이상 20개 미만
4	20개 이상 25개 미만
.	.
.	.
.	.

< 구매 수량을 카테고리화 >

summary

Min	1st Qu.	Median	Mean	3rd Qu.	Max
1	3	8	20.84	26	257

<로지스틱 회귀분석>

1. 종속변수(구매 수량) 구간화 진행
2. 종속변수 범주형 → factor화
3. Library(MASS)

```
step(glm(구매건수~.,family = binomial, data)
```

```
  여성      x20대      x30대      x40대
1.208660 -0.940718  0.667128  1.607282
```

```
  습도      강수량  미세먼지농도      풍속
-0.011494  0.013356  0.007149      0.351805
```

```
  기온
0.154923
```

→ 9개의 변수가 선택 됨.

변수 선택 결과



머신러닝을 통한 판매량 예측 시스템

❖ 생수	→	여성 미세먼지	20대 농도	30대 강수량	40대	기온	풍속	8개
❖ 커피음료	→	여성	20대	30대	40대	기온	풍속	6개
❖ 오렌지	→	여성 강수량	20대	30대	40대	기온	풍속	7개
❖ 토마토	→	여성 미세먼지	20대 농도	30대	40대	기온	풍속	7개
❖ 수박	→	여성 미세먼지	20대 농도	30대 풍속	40대 강수량	기온	습도	9개



예측모델 생성



머신러닝을 통한 판매량 예측 시스템

1

모델 검증을 위한 Testset : Trainset = 9 : 1 분리

2

❖ Random Forest

기본값 사용

❖ XGboost

기본값 사용

❖ Adaboost

n_estimators = 50

learning_rate = 0.7

❖ RNN

epochs = 100

batch_size = 20



예측모델 생성



머신러닝을 통한 판매량 예측 시스템

< 최종 모형 선택 >

→ RMSE는 낮을수록, 결정계수는 높을수록 좋음

소분류	Random Forest		XGboost		Adaboost		RNN	
	RMSE	Adj.R2	RMSE	Adj.R2	RMSE	Adj.R2	RMSE	Adj.R2
생수	2.326	0.335	2.180	0.415	2.104	0.455	2.036	0.490
커피음료	2.590	0.738	2.377	0.779	3.047	0.637	2.366	0.781
오렌지	2.224	0.544	1.976	0.640	1.933	0.656	2.044	0.615
토마토	3.949	0.259	3.210	0.510	3.307	0.480	3.299	0.483
수박	2.478	0.479	1.872	0.702	2.951	0.260	2.263	0.565

예측모델 생성



최종모형 결과 예시

❖ 생수 : RNN

```
array([4, 8, 7, 2, 1, 5, 4, 1])
```

20대 여성 155-187개

30대 여성 312-416개

40대 여성 252-312개

50대 여성 101-133개

20대 남성 75-101개

30대 남성 187-217개

40대 남성 155-187개

50대 남성 101-133개



❖ 토마토 : XGboost

```
array([5,13, 8, 5, 5,10, 5,5])
```

20대 여성 12-15개

30대 여성 50-60개

40대 여성 22개-25개

50대 여성 12-15개

20대 남성 12-15개

30대 남성 30-35개

40대 남성 12-15개

50대 남성 12-15개



❖ 수박 : XGboost

```
array([1,1, 3, 0, 0, 0, 0, 0])
```

20대 여자 5개-10개

30대 여자 5개-10개

40대 여자 15개-20개

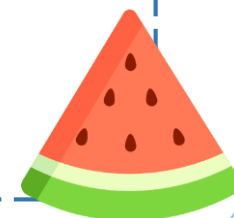
50대 여자 5개 미만

20대 남자 5개 미만

30대 남자 5개 미만

40대 남자 5개 미만

50대 남자 5개 미만



개요



보완재 분석을 통한 추천 시스템

보완재



- 두 제품을 독립적으로 소비하였을 때보다 함께 사용하였을 때 효용이 증가하는 제품

추천 시스템



- 온라인 상품 추천 서비스는 소비자의 쇼핑을 도와주고 추가적인 매출을 유도할 수 있는 마케팅 도구
 - 안전 재고 소진 : 재고를 파악하여 연관 상품 추천
 - 추천 시스템을 경험한 사용자는 높은 확률로 충성 고객이 될 수 있음
 - 고객이 한 개의 상품이라도 더 클릭하도록 개선하는 것이 중요함.
이는 누적 클릭 횟수에 따라 고객의 구매전환율이 비례해서 증가하기 때문.

분석



보완재 분석을 통한 추천 시스템

- 1 소셜 데이터(sns_data) 소분류 중 "곱창" , "오이" 카테고리 선택
- 2 네이버 블로그 크롤링 데이터 수집
 - python selenium webdriver를 이용한 크롤링
 - konlpy를 사용하여 한국어 형태소 분석 후 명사 추출

블로그 내용 수집

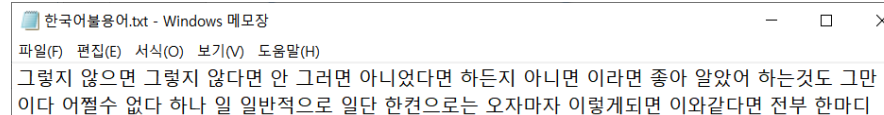
〈 ‘곱창’ 의 크롤링 데이터 〉

텍스트

0	공유가 해답~^^ \n친구끼리 단톡방이 있는데 이런저런 말을 마니 하지만\n가장 많...
1	인천 핫플 인천대공원 맛집 '광명대창집'\n맛집이란 말을 하도 듣고 가는 길이어서 ...
2	인천 소곱창 맛집에서는 의외의 아늑함이 있다? ㅎㅎ\n곱창을 파는 곳이라면 인식이 ...
3	특히 창가 자리는 best!! \n울드한 컨셉으로 옛날 식당온듯한 문구들이 창에 적...
4	WHY~?? 인천 핫플인지는 메뉴판만 봐도 알 수 있다 \n첫째 종류가 많다 구이류...
...	...

konlpy 형태소 분석

Konlpy를 통해 형태소 분석 후 명사만 추출

 불용어 제거

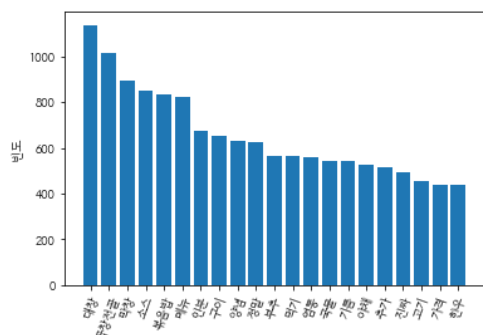
분석

3

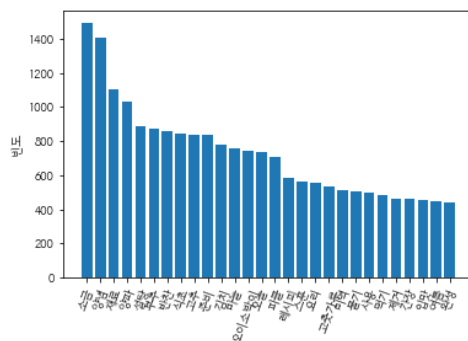
- ### 보완재 분석
- 방법 1. 단순 빈도수 계산하여 막대그래프, wordcloud 작성
 - 방법 2. 연관어 분석
 - 방법 3. 토픽 모델링

1. 단순 언급 단어 빈도 분석

< ‘곱창’의 막대그래프와 wordcloud >



< ‘오이’의 막대그래프와 wordcloud >



2. 연관어 분석

Word2Vec - `wv.most_similar()` 사용

size = 100, window = 5, min_count = 5,
workers = 4, sg = 0

< ‘곱창’의 연관어 분석 결과 >



한우, 돼지, 구이, 특양, 초벌, 부위, 막창, 모듬, 갈비살, 우모, 곱, 홍창, 직화, 대흥동, 문집, 대창, 안지, 염통, 늑간살, 막창구이 등

< ‘오이’의 연관어 분석 결과 >



도라지, 깍두기, 야채, 양배추, 배추, 백오이, 소세지, 미나리, 진미채깍두기, 등

분석

3

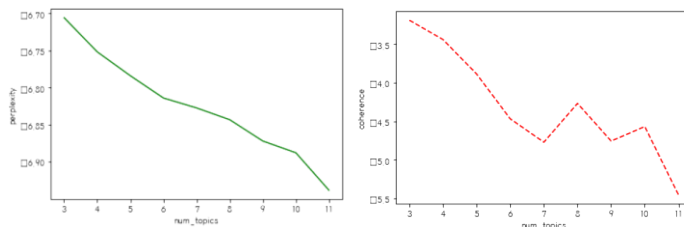
보완재 분석

3. 토픽 모델링

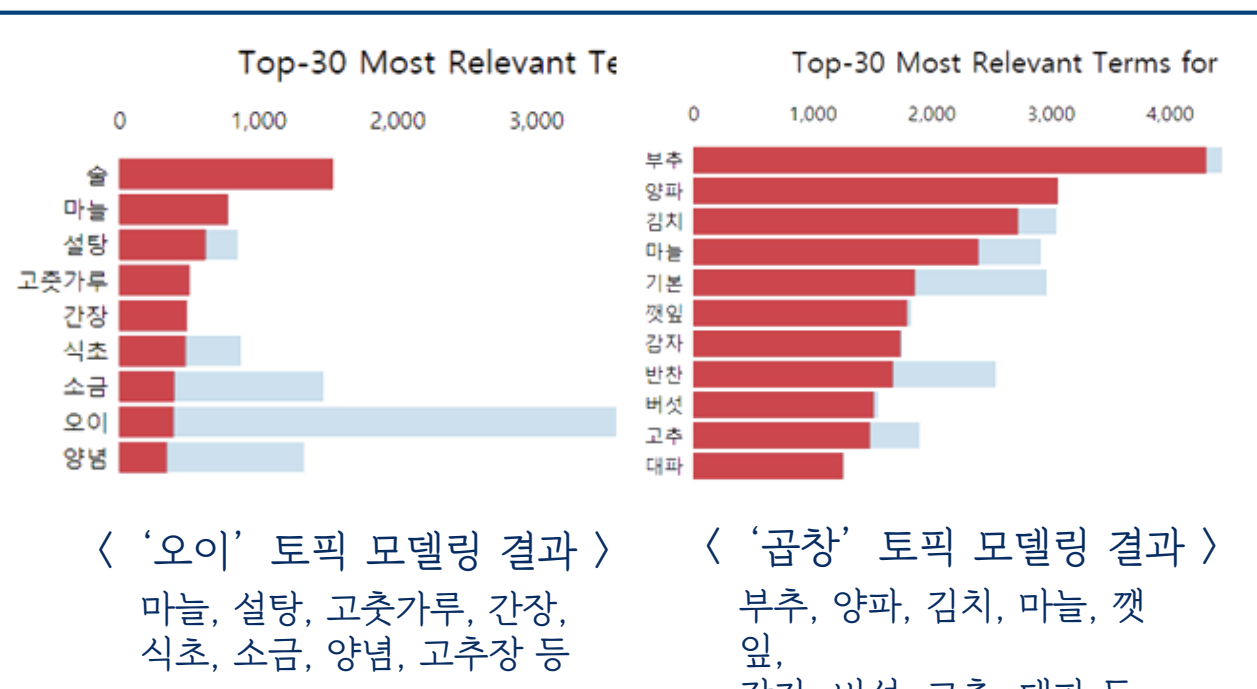
< LDA 토픽 모델링 >

1. Perplexity와 Coherence를 통해
최적의 토픽 수를 찾음 ➡ 오이 : 9개
곱창 : 10개

< ‘곱창’ 의 Perplexity와 Coherence >



2. pyLDAvis를 통해 시각화



< ‘오이’ 토픽 모델링 결과 >

마늘, 설탕, 고춧가루, 간장,
식초, 소금, 양념, 고추장 등

< ‘곱창’ 토픽 모델링 결과 >

부추, 양파, 김치, 마늘, 찹
쌀,
감자, 버섯, 고추, 대파 등

➡ 1 ~ 3 중 보완재를 가장 잘 나타내는 토픽 모델링 기법 선택



서비스 활용 방안



활용 예시



애플리케이션 < 토바기 > 예시

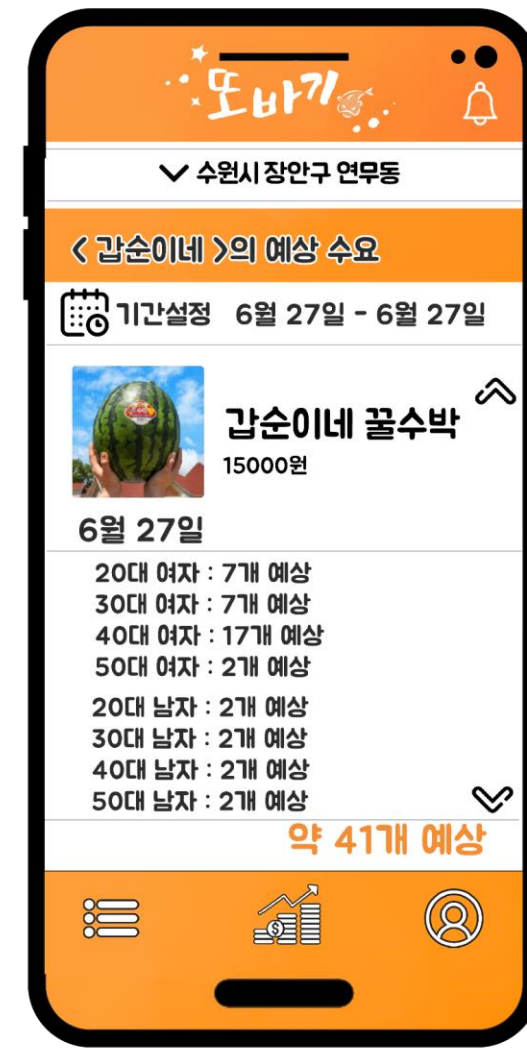


토바기 : 언제나 한결같이 꼭 그렇게

- 구매처로 언제나 전통시장을 고려해줬으면 좋겠다는 의미
- 전통시장의 전통을 한결같이 이어가고 싶다는 의미



<소비자 메인 화면>

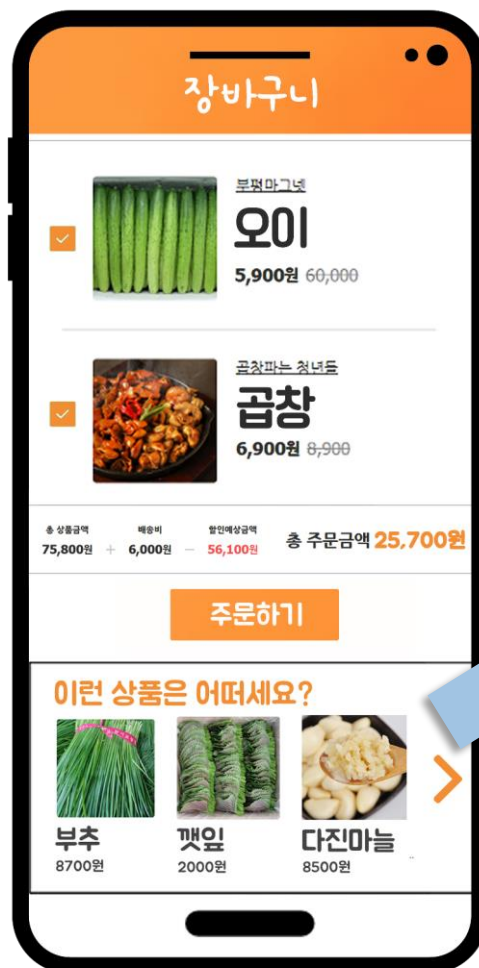


<판매자 판매량 예측 화면>

활용 예시



보완재 분석을 통한 추천 시스템

장바구니 화면
추천 시스템 예시

이런 상품은 어떠세요?

부추
8700원갯잎
2000원다진마늘
8500원

‘오이’의 보완재인 마늘,
‘곱창’의 보완재인 부추, 갯잎 추천

활용 예시

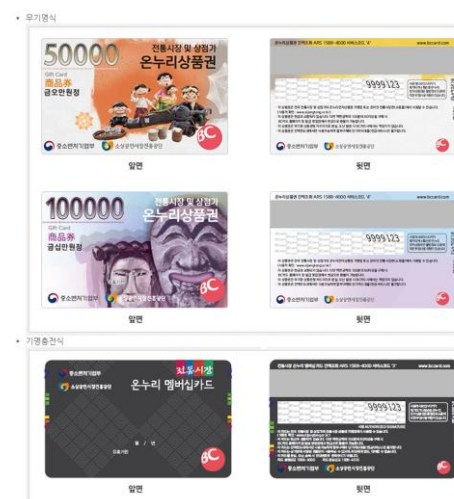
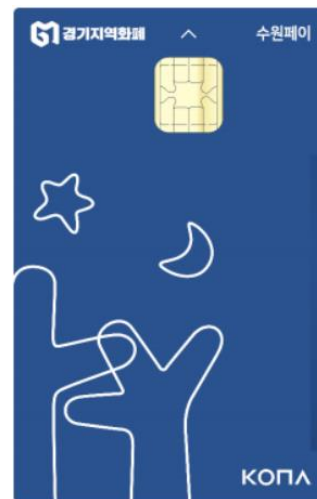


지역 화폐 및 온누리 상품권 활용



경기지역화폐 결제 가능!

각종 재난지원금, 온누리 상품권도 사용 가능



전통시장 이용객 증대



결론



예측모델 생성



기대효과

① 전통시장의 단점 보완

- 1) 접근성이 떨어지는 전통시장을 온라인 매장으로 보완
- 2) 재고 관리의 어려움을 예측 시스템으로 보완

② 전통시장만의 경쟁력 증가

- 1) 주된 판매 제품이 국내 농수산물인 점을 통해 대형마트보다 신선한 제품 판매 가능
- 2) 대형마트에서는 사용하지 못하는 지역화폐, 온누리 상품권의 사용 편의성 증가
- 3) 보완재 추천 시스템으로 인하여 전통시장 내 상점들 간의 상호 보완 작용의 발생으로 시장 활성화

③ 전통시장의 활성화에 따른 농/수산업의 활성화 기대 가능



마무리



소비자



비대면 거래 방식으로 판매량 증가



지역 주민들의 지역화폐 사용 장려



전통시장 상인들과의 거래량 증가

또바기를 통한 여러 측면에서의 활성화 기대

참고자료

- 코로나19 1년, 서울 골목상권 59% 매출 하락...총매출 19.6% 감소
<http://www.datasom.co.kr/news/articleView.html?idxno=112942>
- 코로나 특수 맞은 '온라인 식품시장'... 올해 40조원 돌파
https://biz.chosun.com/site/data/html_dir/2020/12/28/2020122800943.html
- 네이버·쿠팡·SK '온라인쇼핑 삼국지'
<https://www.mk.co.kr/news/business/view/2020/11/1178514/>
- 문세하,(2020) 포스트코로나 시대 언택트 소비로 인한 소매공간 수요변화와 시사점, 국토연구원
- 로지스틱 회귀분석 : <https://bioinformaticsandme.tistory.com/296>
- 이미지 : <https://www.flaticon.com/kr/>
<https://www.crowdpic.net/>
<https://www.logoyogo.com/>
<https://kr.lovepik.com/>





THE END

