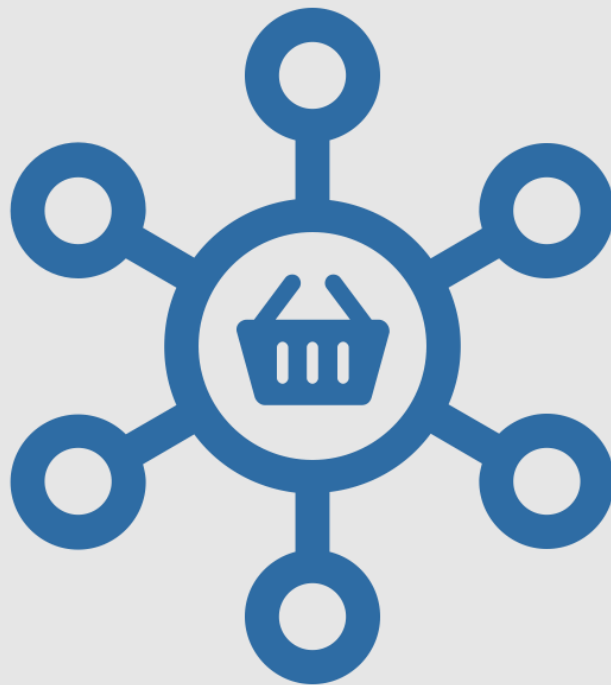


2019 날씨 빅데이터 콘테스트

날씨를 활용한 랄라블라 수요 예측 및 판매 전략

202937 임밴저스

김종윤 손해서 심혜진 정동훈



Contents

- 1 주제 정의
| 공모배경 / 분석목적 및 과정요약
- 2 활용 데이터 정의
| 데이터 설명 / 변수 생성 및 설명
- 3 데이터 준비과정
| 데이터 탐색 / 데이터 정제
- 4 통계적 분석을 통한 판매 예측 모형 개발
| 목표 모델 정의 / 통계 분석 및 모델링 / 모델 비교 및 평가
- 5 분석 결과 활용방안
| 서비스 활용방안 및 기대 효과

1. 주제 정의

공모 배경 / 분석 목적 및 과정 요약

[날씨경영 스토리] 유통업계 실적 좌우하는 '날씨 경영'

하늘을 봐야 돈을 벌지... 기업들 날씨 마케팅 넘어 '날씨 경영'

불황 극복 비결은 날씨마케팅 적극 도입

[트렌드]유통업계 최고 영업사원은 '날씨'

[이브닝 비즈뉴스] 비 오는 날이면, 이색 '장마 마케팅'

유통 ▾

"날씨가 甲" 폭염·장마에 희비 엇갈리는 유통가(종합)

[트렌드 바꾸는 기후변화]날씨경영 따라 엇갈리는 기업 희비...날씨 빅데이터 활용
하면 재고 ↓ 매출 ↑



날씨데이터 분석으로 도출된 “판매 예측 모델”은 산업의 효율성 & 사회·개인적 효용을 증대 ↑

*** 전 직원 폭로..."에어컨 설치비용 때문에 화장품 방치"

머니투데이 이호길 인턴기자 | 입력 : 2019.07.09 10:47

전 직원 "에어컨 설치에 수천만원 든다며 미뤘다"

*** 전 직원이 ***에서 판매한 화장품에 대해 폭로했다.

MBC 시사교양프로그램인 '당신이 믿었던 페이크 시즌2'는 지난 8일 방송을 통해 SNS(사회관계망서비스) 인플루언서 *** 논란을 다뤘다.

이날 방송에서 *** 전 직원 A씨는 ***의 화장품은 냉방시설이 제대로 완비되지 않은 열악한 환경에서 보관됐다고 주장했다. 그는 보관 환경에 대해 "한여름에 밖보다 안이 더 더웠다"고 설명했다.

이어 "화장품이 녹을 것 같아 에어컨을 요청했는데 노후건물이라 설치에 수천만원이 든다며 미뤘다"며 "며 "이번 사건이 터지면서 제조일자로 확인하고 온도도 측정하기 시작한 것"이라고 말했다.

일부 소비자들은 *** 화장품을 쓰고 난 뒤 피부질환이 발생했다며 집단 손해배상 청구에 나서기도 했다. 소비자들을 대리하는 강용석 법무법인 넥스트로 변호사는 지난달 18일 ***의 모회사인 *****를 상대로 1인당 1000만원씩 총 3억7000만원의 손해배상 청구 소송을 제기했다.

분석 목적 및 과정 요약



주제1 날씨 데이터를 활용한 수요 예측



주제2 날씨와 관련된 사전 마케팅 시스템 구축

“판매 예측 모형”



(재고 관리 방안 & 맞춤형 활용방안)

✓ 분석 목적

날씨 데이터를 활용한 판매 예측 모형 개발

- [개인] 원하는 물품을 원하는 시기에 구입
- [기업] 효율적인 마케팅 전략 수립
- [국가] 불필요한 자원 낭비 최소화

Overview

- 01.활용 데이터의 정의
- 02.목표 모델 정의
- 03.통계 분석 및 모델링
- 04.판매 예측 모델 개발

2. 활용 데이터 정의

데이터 설명 / 새로운 변수 정의

데이터 설명 — 기상 데이터 및 칼라블라 판매 데이터

✓ 수집 기간 : 2016년 1월 1일 ~ 2018년 12월 31일 (데이터 단위 : 일 단위)

✓ 기상 데이터



bigcon_weather



	tm	stn_id	pvn_nm	bor_nm	max_ta	max_ws	min_ta	avg_ta	avg_rhm	avg_ws	sum_rn
1	20160101	98	경기도	동두천시	6.5	3.8	-6.4	-0.2	74	0.9	0.0
2	20160102	98	경기도	동두천시	10.1	3.2	1.9	5.4	70	1.3	0.0
3	20160103	98	경기도	동두천시	9.0	2.4	2.1	4.7	82	0.7	0.0
4	20160104	98	경기도	동두천시	8.0	5.3	-3.0	1.9	55	2.2	0.0
5	20160105	98	경기도	동두천시	3.1	5.6	-6.4	-2.8	36	1.8	0.0

❖ Table info.

변수 : tm ~ sum_rn (11개)
행 : 관측일 날씨 정보 (59,113개)

- tm : 관측일 (YYYYMMDD)
- stn_id : 관측 지점 번호
- pvn_nm : 법정동 코드(도)
- bor_nm : 법정동 코드(시,군)
- max_ta : 최고 기온(°C)
- max_ws : 최대 풍속(m/s)
- min_ta : 최소 기온(°C)
- avg_ta : 평균 기온(°C)
- avg_rhm : 평균 상대습도(%)
- avg_ws : 평균 풍속(m/s)
- sum_rn : 합계 강수량(mm)

✓ 유통_칼라블라 판매 데이터



korea_hnb



	sals_dt	pvn_nm	bor_nm	gen_cd	age_cd	category	qty
1	20160101	서울특별시	종로구	F	20~39	립컬러	73
2	20160101	서울특별시	종로구	F	20~39	립케어	219
3	20160101	서울특별시	종로구	F	20~39	마스크팩	657
4	20160101	서울특별시	종로구	F	20~39	바디로션	328
5	20160101	서울특별시	종로구	F	20~39	체중조절	109

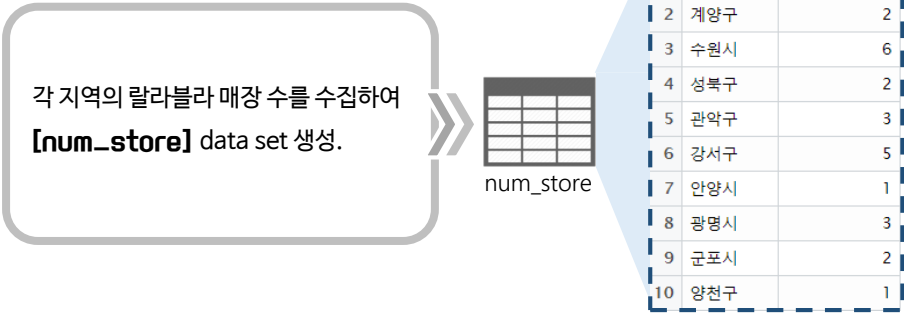
❖ Table info.

변수 : sals_dt ~ qty (7개)
행 : 관측일 판매 정보 (820,274개)

- sals_dt : 관측일
- pvn_nm : 법정동 코드(도)
- bor_nm : 법정동 코드(시,군)
- gen_cd : 성별
- age_cd : 연령
- category : 판매 품목 카테고리
- qty : 수량

새로운 변수 정의

✓ 생성 데이터 정의



✓ 생성 변수 정의

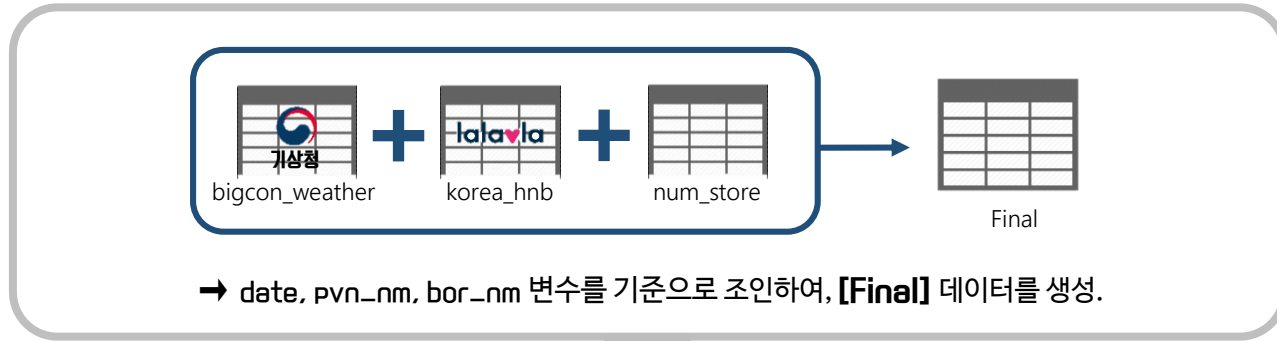
변수명	변수설명	변수타입
month	월(1 ~ 12)	factor
day	일(1 ~ 31)	numeric
weekday	요일(월요일 ~ 일요일)	factor
season	계절(1 = 봄, 2 = 여름, 3 = 가을, 4 = 겨울)	factor
num_store	지역별 점포 수	numeric
each_qty	점포 당 판매량 (= qty ÷ num_store)	numeric
yesterday_qty	전일 판매량	numeric

✓ 결과 테이블

	date	month	day	weekday	season	qty	num_store	each_qty	yesterday_qty
1	20160101	1	1	금요일	4	73	4	18.25000	73
2	20160101	1	1	금요일	4	255	4	63.75000	255
3	20160101	1	1	금요일	4	146	4	36.50000	146
4	20160101	1	1	금요일	4	36	4	9.00000	36
5	20160101	1	1	금요일	4	109	4	27.25000	109
6	20160101	1	1	금요일	4	255	4	63.75000	255
7	20160101	1	1	금요일	4	182	4	45.50000	182
8	20160101	1	1	금요일	4	36	4	9.00000	36
9	20160101	1	1	금요일	4	657	4	164.25000	657
10	20160101	1	1	금요일	4	73	4	18.25000	73
11	20160101	1	1	금요일	4	36	4	9.00000	36
12	20160101	1	1	금요일	4	182	4	45.50000	182
13	20160101	1	1	금요일	4	219	4	54.75000	219
14	20160101	1	1	금요일	4	73	4	18.25000	73
15	20160101	1	1	금요일	4	146	4	36.50000	146
...									
820260	20181231	12	31	월요일	4	219	2	109.50000	438
820261	20181231	12	31	월요일	4	109	2	54.50000	109
820262	20181231	12	31	월요일	4	73	2	36.50000	73
820263	20181231	12	31	월요일	4	73	2	36.50000	36
820264	20181231	12	31	월요일	4	219	2	109.50000	36
820265	20181231	12	31	월요일	4	146	2	73.00000	146
820266	20181231	12	31	월요일	4	438	2	219.00000	803
820267	20181231	12	31	월요일	4	36	2	18.00000	109
820268	20181231	12	31	월요일	4	2408	2	1204.00000	949
820269	20181231	12	31	월요일	4	36	2	18.00000	36
820270	20181231	12	31	월요일	4	365	2	182.50000	109
820271	20181231	12	31	월요일	4	511	2	255.50000	255
820272	20181231	12	31	월요일	4	146	2	73.00000	36
820273	20181231	12	31	월요일	4	36	2	18.00000	73
820274	20181231	12	31	월요일	4	292	2	146.00000	73

데이터 획득 _메인 데이터[Final]

✓ [Final] 생성



✓ 테이블 결과

	date	month	day	season	weekday	pvn_nm	bor_nm	gen_cd	age_cd	category	qty	num_store	each_qty	yesterday_qty	max_ta	max_ws	min_ta	avg_ta	avg_rhm	avg_ws	sum_rn
1	20160101	1	1	4	금요일	경기도	고양시	F	2	네일	73	4	18.25000	73	5.2	2.5	-5.0	1.2	73	0.6	0
2	20160101	1	1	4	금요일	경기도	고양시	F	2	헤어샬롱렌저	255	4	63.75000	255	5.2	2.5	-5.0	1.2	73	0.6	0
3	20160101	1	1	4	금요일	경기도	고양시	F	3	헤어샬롱렌저	146	4	36.50000	255	5.2	2.5	-5.0	1.2	73	0.6	0
4	20160101	1	1	4	금요일	경기도	고양시	M	3	헤어샬롱렌저	36	4	9.00000	146	5.2	2.5	-5.0	1.2	73	0.6	0
5	20160101	1	1	4	금요일	경기도	고양시	M	2	헤어샬롱렌저	109	4	27.25000	36	5.2	2.5	-5.0	1.2	73	0.6	0
6	20160101	1	1	4	금요일	경기도	고양시	F	3	립컬러	255	4	63.75000	255	5.2	2.5	-5.0	1.2	73	0.6	0
7	20160101	1	1	4	금요일	경기도	고양시	F	3	마스크팩	182	4	45.50000	182	5.2	2.5	-5.0	1.2	73	0.6	0
8	20160101	1	1	4	금요일	경기도	고양시	F	1	마스크팩	36	4	9.00000	182	5.2	2.5	-5.0	1.2	73	0.6	0
9	20160101	1	1	4	금요일	경기도	고양시	F	2	마스크팩	657	4	164.25000	36	5.2	2.5	-5.0	1.2	73	0.6	0
10	20160101	1	1	4	금요일	경기도	고양시	F	3	네일	73	4	18.25000	73	5.2	2.5	-5.0	1.2	73	0.6	0

✓ Table info.

칼럼 : date ~ sum_rn (21개)

행 : 당일 날씨와 라라블라 매출 정보(820,274개)

3. 데이터 준비 과정

데이터 탐색 / 데이터 정제

데이터 탐색 - 구매고객 분포

✓ 성별 분포

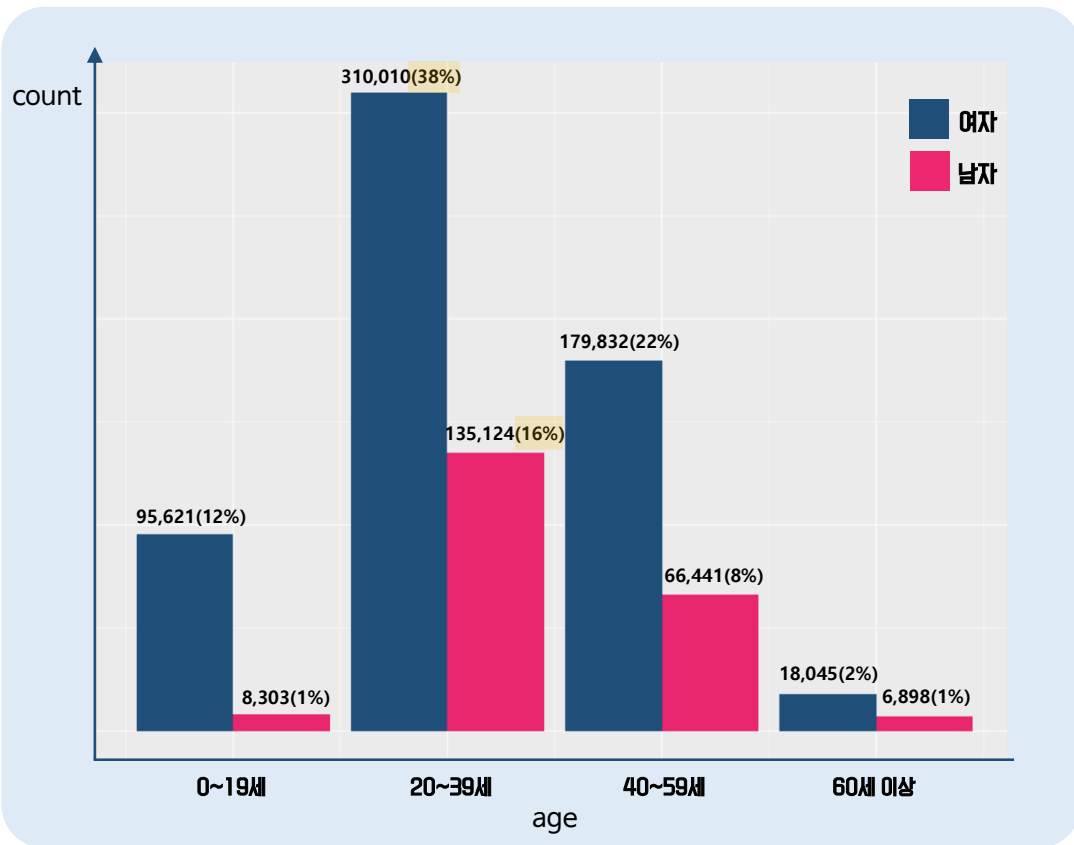
- 표본 820,274명의 남녀 성별 비율은 3:7
- 남성 = 216,766명 (26%)
- 여성 = 603,508명 (74%)

→ “여성 고객”의 구매량이 월등히 크다.

✓ 나이 분포

- 표본으로 추출된 사람들의 나이 분포는
- 0~19세 = 103,992명 (13%)
- 20~39세 = 445,134명 (54%)
- 40~59세 = 246,273명 (30%)
- 60세 이상 = 24,943명이다. (3%)

→ “20~39세”의 구매량이 가장 많다.

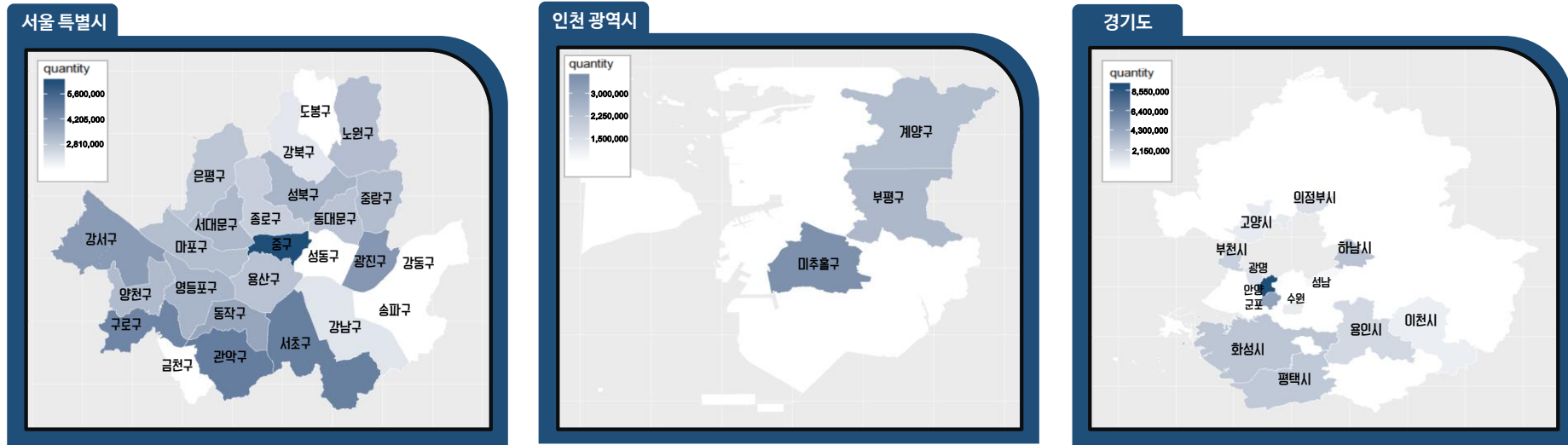


→ 랠라블라의 주요 고객층은 “20 · 30대 여성(38%)”

→ (남성의 경우 20·30대 구매율이 전체에서 3번째로 총16%를 차지했지만, 남성 고객 한에서는 62%로 가장 높다.)

데이터 탐색_지역별 판매량 확인

→ 지역별 판매량에 차이가 있는지 확인하기 위해 점포당 판매량을 지도로 시각화 해보았다.



2016년부터 2018년까지 서울시, 인천시, 경기도에서 총 36지역의 판매량 차이를 분석한 결과,

- ✓ 전체 36지역 중 판매량이 가장 많은 **“안양시”**는 1000만개가 넘는 판매량으로 전체 판매량의 10%를 차지한다.
- ✓ “중구”는 약 700만개로 판매량 2위(7%)이고, 그 다음으로 “군포시(5%) · 관악구(4.7%) · 서초구(4.6%)” 순으로 판매량이 높다.
- ✓ 전체 36지역 중 판매량이 가장 작은 **“성남시”**는 전체 판매량의 0.3%를 차지한다.

데이터 탐색_타임 트렌트 확인

→ 계절성 차이가 존재하는지 확인하고자 2016년부터 2018년까지 각 제품의 월별 판매량 히스토그램을 그려 보았다.

월별 판매량



네일 · 선크어 · 제모제

- ✓ 동일한 계절성을 갖고 6, 7월에 판매량이 증가하는 경향을 보였다.

립케어 · 바디로션

- ✓ 동일한 계절성을 갖고 12, 1월에 판매량이 증가하는 경향을 보였다.

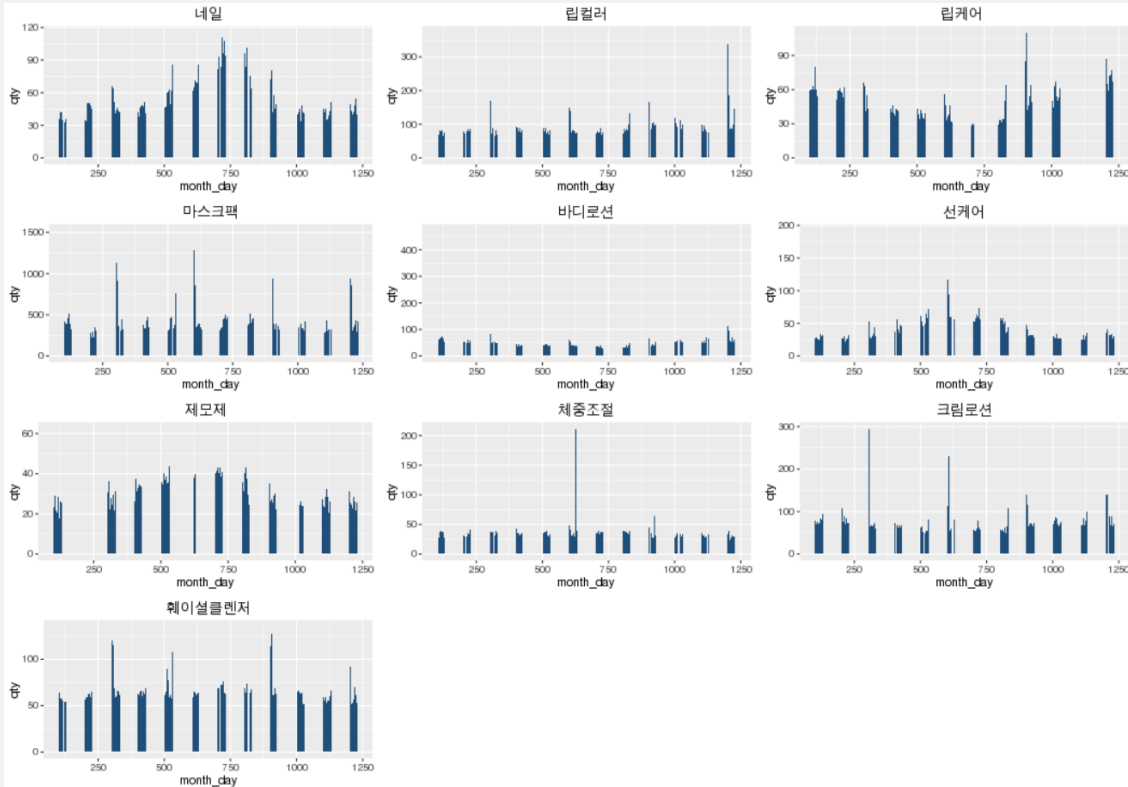
- ✓ 립컬러 : 12월 판매량이 가장 높다.
- ✓ 마스크팩 : 분기말마다 판매량이 높다.
- ✓ 체중조절 : 6월 판매량이 가장 높다.
- ✓ 크림로션 : 3월과 12월 판매량이 가장 높다.
- ✓ 웨이셜클렌저 : 월별 판매량 차이가 크게 나타나지 않는, 꾸준하게 판매되는 상품이다.

→ 제품별로 월별 판매량 분포가 다르다.

데이터 탐색 -타임 트렌드 확인

→ 시기별 판매량 차이가 존재하는지 확인하고자 2016년부터 2018년까지 각 제품의 판매량을 확인해 보았다.

일별 · 월별 판매량



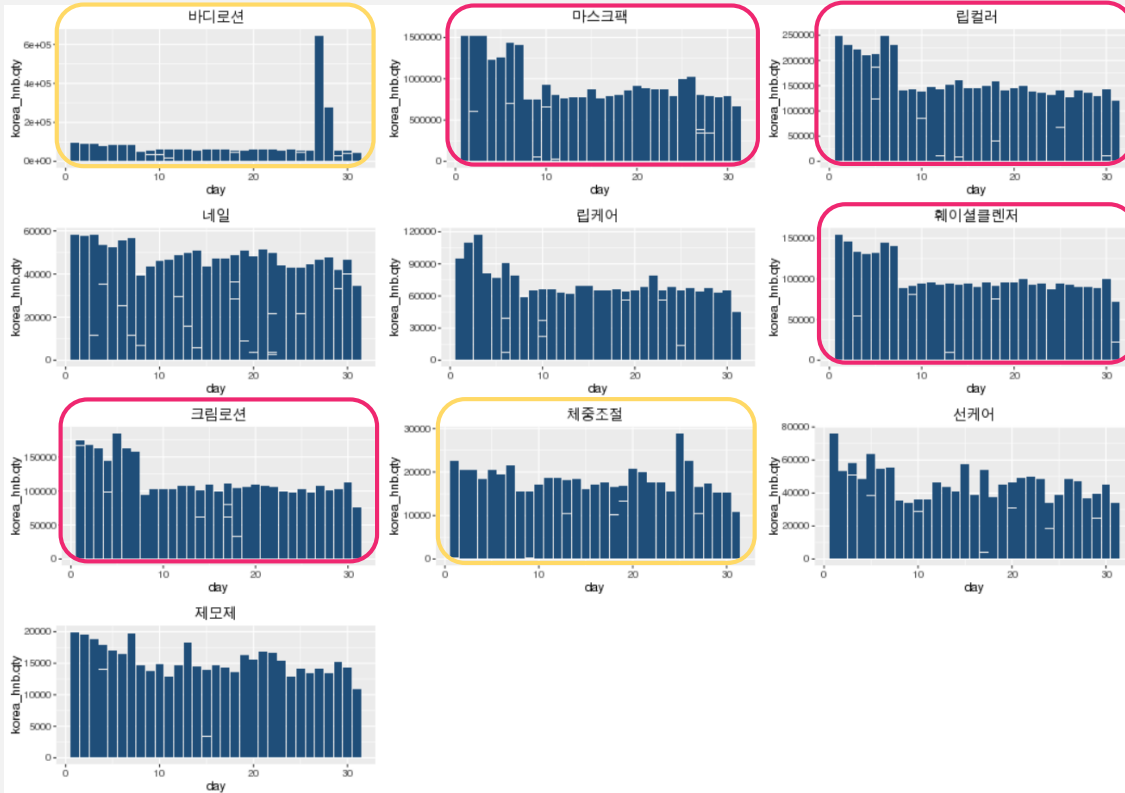
- ✓ 1월 1일부터 12월 31일까지 일자별 3개년 평균 판매량을 확인하는 막대그래프이다.
- ✓ 월별 특징도 동시에 확인하기 위해, 월과 월 사이에 간격을 두고 그래프를 그렸다.
- ✓ 각 월의 그래프는 U자형을 그리고 있으므로, **대다수의 품목이 월초나 월말에 많이 팔리는 경향이 있다고 할 수 있다.**

→ 월초, 월말의 판매량 추세를 정확히 보기 위해, 다음 페이지에서 Day로 구성된 막대그래프를 다시 그려보았다.

데이터 탐색 -타임 트렌트 확인

➔ 월초 & 월말의 판매량 추세를 확인하고자 2016년부터 2018년까지 각 제품의 일별 판매량 히스토그램을 그려 보았다.

일일 판매량



바디로션 · 체중조절

- ✓ 해당 상품은 월말에 많이 팔리는 경향을 보이고 있다.

마스크팩 · 립컬러 · 웨이셜클렌저 · 크림로션

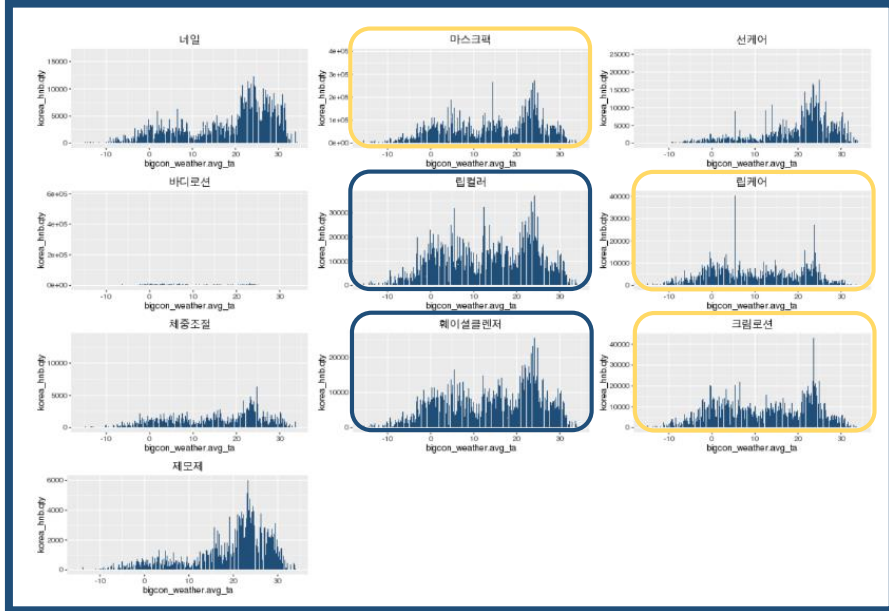
- ✓ 해당 상품은 월초에 많이 팔리는 경향을 보이고 있다.
- ✓ 매달 월초에 진행하는 세일정책에 영향을 받은 것으로 보인다.

립케어 · 네일 · 선크어 · 제모제

- ✓ 일별 판매량 큰 차이가 없다.
- ✓ 세일에 영향을 받지 않는 제품으로 보인다.

데이터 탐색_기상에 따른 판매량

1. 평균 기온



- ✓ 대다수의 품목은 20°C ~ 27°C에서 판매량이 **증가**한다.
- ✓ 그러나, 겨울과 관련된 제품과 꾸준한 수요가 있는 생필품의 경우,

〔크림로션 · 립케어 · 마스크팩〕 & 〔립컬러 · 웨이셜클렌저〕

→ 0°C 근처에서도 판매량이 **증가**한다.

2. 함께 강수량



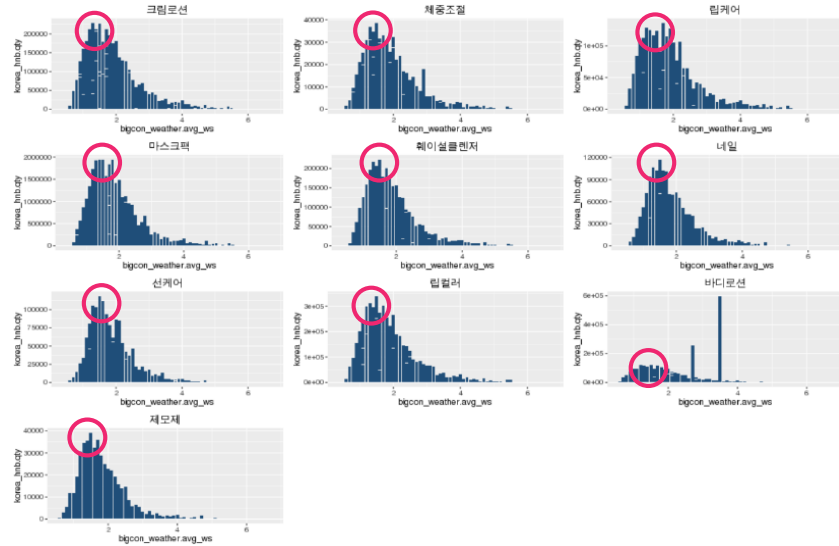
- ✓ 대다수의 품목은 강수량이 적을 수록 판매량이 **증가**한다.

✓ 〔네일 · 웨이셜클렌저 · 체중조절 · 제모제〕의 경우,

→ 강수량에 상관없이 **꾸준한 수요**가 있는 제품이다.

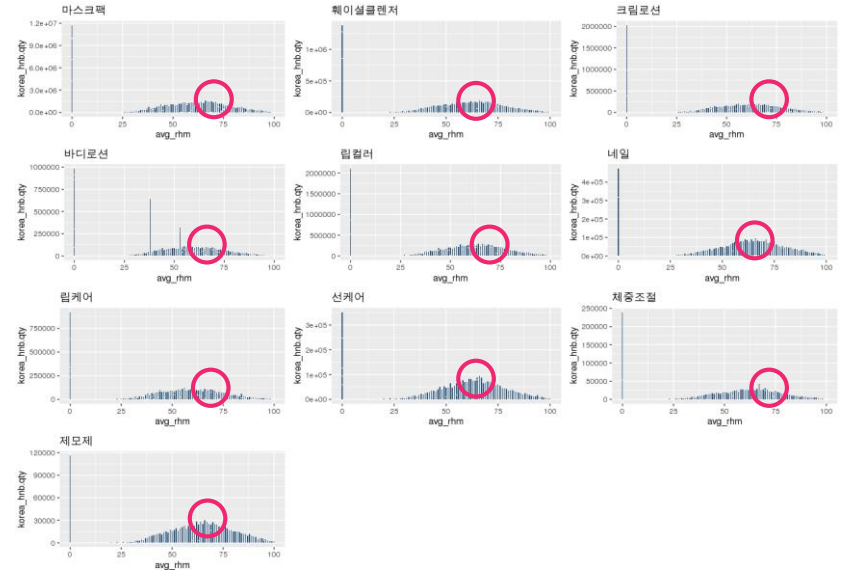
데이터 탐색_기상에 따른 판매량

3. 평균 풍속



- ✓ 몇몇의 특이값을 제외하고는, 모든 품목에 있어서 풍속이 약 1.5m/s 근처일 때 수요가 높다.

4. 평균 상대습도



- ✓ 평균 상대습도 데이터에 결측치가 많아, 대체된 0%의 값이 높게 나타난다.
- ✓ 일반적으로 60%~75% 정도의 상대습도일 때 모든 품목이 수요가 높다.

데이터 정제

→ RAW DATA가 지닌 정보를 극대화 하기 위해 다음과 같이 잡음을 제거 및 변환하였다.

결측치 처리

기상 데이터 **[bigcon_weather]** 에서 7개의 기상변수에 대한 결측치는 총 **16,745개** 임을 확인 했다.

- ✓ 확인된 결측치는 관측일의 월과 일이 일치하는 다른 연도 관측값의 평균값으로 대체하였다.
- ✓ 만약 모든 연도의 데이터가 결측값이라면 그대로 결측값으로 두었다.

→ 결측치 처리 결과 : 모든 연도의 관측값이 결측치인, 평균 상대습도 (avg_rhm)에서만 총 12,189개의 결측치가 남았다.

처리 전

Weather_NA	bor_nm	max_ta	max_ws	min_ta	avg_ta	avg_rhm	avg_ws	sum_rn
1	가평군	8	8	8	9	11	8	0
2	강남구	1	1	1	1	1	1	0
3	강동구	6	6	6	7	7	6	2
4	강북구	2	3	2	2	2	3	0
51	파주시	6	6	6	3	3	4	2
52	평택시	2	2	2	2	2	2	0
53	포천시	9	9	9	9	7	9	0
54	화성시	3	3	3	3	2	3	1

합계	287	335	287	299	15,200	285	54	16,745
----	-----	-----	-----	-----	--------	-----	----	--------

처리 후

Weather_NA	bor_nm	max_ta	max_ws	min_ta	avg_ta	avg_rhm	avg_ws	sum_rn
1	가평군	0	0	0	0	0	0	0
2	강남구	0	0	0	0	0	0	0
3	강동구	0	0	0	0	0	0	0
4	강북구	0	0	0	0	0	0	0
51	파주시	0	0	0	0	0	0	0
52	평택시	0	0	0	0	0	0	0
53	포천시	0	0	0	0	0	0	0
54	화성시	0	0	0	0	0	0	0

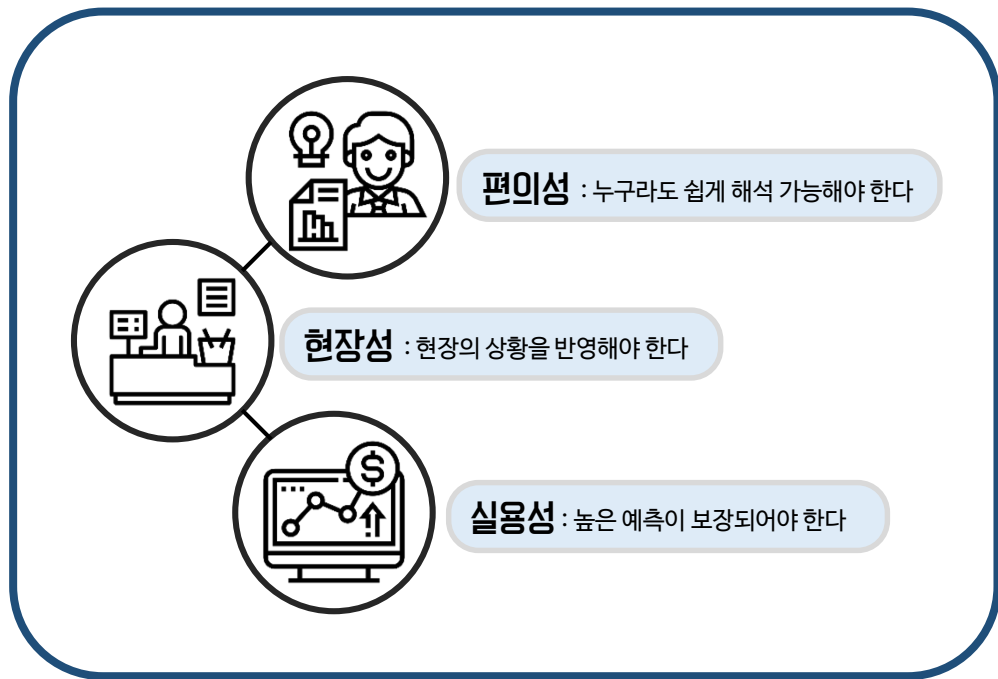
합계	0	0	0	0	12,189	0	0	12,189
----	---	---	---	---	--------	---	---	--------

4. 통계적 분석을 통한 판매 예측 모형 개발

목표 모델 정의 / 통계 분석 및 모델링 / 모델 비교 및 평가

목표 모델 정의

✓ 목표 모델

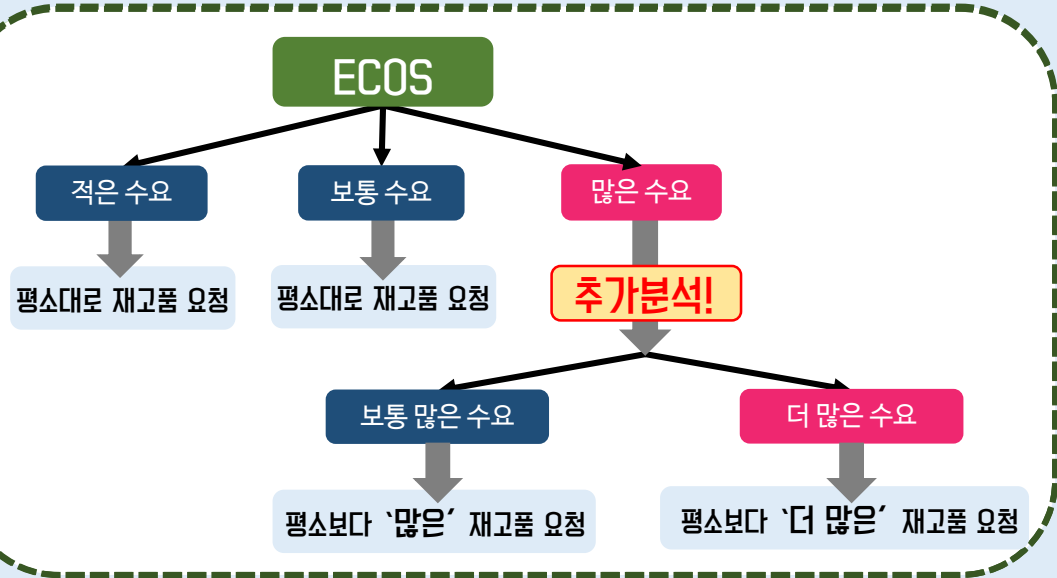


목표 모델 정의

① 편의성

알아보기 편리한 Classification을 주된 타겟 모델로 설정하여 매장 직원 누구라도 해석 가능하게 함.

예시)



목표 모델 정의

② 현장성

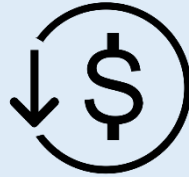
현업에서 실질적으로 필요하고 적용 가능한 데이터를 타겟 : **예상 판매량 정도를 예측**

예시)

✓ 적은 수요 예측

불필요한

재고 관리비 절약



✓ 많은 수요 예측

판매 가능한 상품을 보유하지 못하는

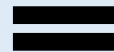
재고부족 위험 제거



③ 실용성

적합도 (prediction rate)가 가장 높은 모델을 판매 예측 모형으로 선정

예시)



통계분석 및 모델링

- ✓ Regression 방법으로 모델링
- ✓ 관심 변수 : 카테고리별 판매량
- ✓ 설명 변수 : max_ta, max_ws, min_ta, avg_ta, avg_ws, sum_rn

ARDL Time-Series

- ✓ 초기 사전분포는 각 분류에 1/4, 1/4, 1/2 로 설정하여 Classification 방법으로 모델링
- ✓ 관심 변수 : 카테고리별 판매량
- ✓ 설명 변수 : korea_hnb, age_cd, num_store, avg_ta, sum_rn, season, month, weekday, day, yesterday_qty, qty

LDA

Naive Bayes

- ✓ 초기 사전확률분포를 균일하게 설정하여 Classification 방법으로 모델링
- 분류 1 : each_qty ≤ 17인 값으로, 전체 중 ¼의 비중 차지
- 분류 2 : 17 ≤ each_qty ≤ 34 인 값으로, 전체 중 1/4의 비중 차지
- 분류 3 : 34 ≤ each_qty 인 값으로, 전체 중 2/4의 비중 차지
- ✓ 관심 변수 : 카테고리별 판매량
- ✓ 설명 변수 : bor_nm, category, age_cd, num_store, season, month, weekday, day, yesterday_qty, qty, avg_ta, sum_rn

Random Forest

- ✓ Split당 사용된 변수 16개, 나무개수 500으로 설정하여 Classification 방법으로 모델링
- 분류 1 : each_qty ≤ 17인 값으로, 전체 중 ¼의 비중 차지
- 분류 2 : 17 ≤ each_qty ≤ 34 인 값으로, 전체 중 1/4의 비중 차지
- 분류 3 : 34 ≤ each_qty 인 값으로, 전체 중 2/4의 비중 차지
- ✓ 관심 변수 : 카테고리별 판매량
- ✓ 설명 변수 : bor_nm, category, age_cd, gen_cd, season, month, weekday, day, yesterday_qty, qty, max_ws, max_ta, min_ta, avg_ws, avg_ta, sum_rn

모델 비교 및 평가

분석 결과 및 모델 비교

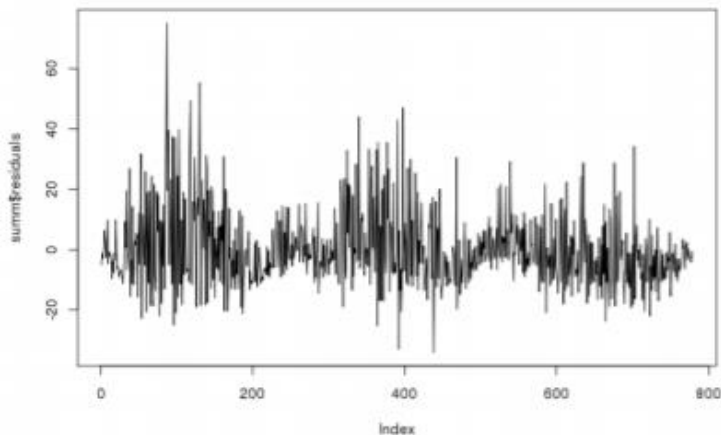
modeling	Prediction Rate	R-squared	Optimal Variable
LDA	61.40%	-	웨이셜클렌저 점포당 판매량
Naive Bayes	73%	0.145	웨이셜클렌저 점포당 판매량
Random Forest	78% (88%)	-	웨이셜클렌저 점포당 판매량
ARDL Time-Series	-	0.3284	마포구 · 제모제 점포당 판매량



Prediction rate 과 **R-Squared** 가 **높은 모델을** 최종 모델로 **선택!**

모델 비교 및 평가

잔차 가정 확인



▪ Durbin-Watson statistic = 2.13

잔차 가정
 $e_t \sim (0, \sigma^2)$

시간에 따른 추세변화와 자기상관성이 없어야 하며,
등분산성을 충족해야 한다.

- ✓ 잔차 plot으로 보아, 시간에 따른 추세변화와 자기상관성은 없었다. 그러나, ARDL Time-series는 이분산의 특징이 있었다.
 - ✓ 이에 해당 모델의 기본적인 가정을 충족 할 수 없어, 결과 또한 신뢰할 수 없다.
- 그러므로 Prediction rate이 가장 높은 Random Forest가 판매량 예측에 가장 적합한 모델이라고 판단하였다.

Random Forest 모델을 통한 판매 예측으로 미래 수요를 대비

모델 비교 및 평가

예시 1

웨이셜클렌저 수요 예측

✓ Random Forest 분석 결과

```
Call:
  randomForest(formula = quantity ~ ., data = cleanser, mtry = 16,      ntree = 500, subset = train)
  Type of random forest: classification
    Number of trees: 500
No. of variables tried at each split: 16

  OOB estimate of error rate: 21.35%

Confusion matrix:
      1      2      3 class.error
1 6915  932   540  0.1755097
2 2084 4038  2382  0.5251646
3  720 2251 21866  0.1196199
```

100% - 21.35% = 78.65%

100% - 11.96199% ≈ 88%

✓ **전체 수요**에 대하여 **78.65%**의 높은 예측률 !

✓ 관심 타겟인 '**많은 수요**'에 대하여는 **약 88%**로 월등한 예측률을 보임

모델 비교 및 평가

예시 2

웨이셜클렌저 수요 심화 예측 [random forest 2회 적용]

✓ Random Forest 분석 결과

```
> cleanser2.500

Call:
  randomForest(formula = quantity ~ ., data = cleanser2, mtry = 16,      ntree = 500, subset = train)
  Type of random forest: classification
  Number of trees: 500
  No. of variables tried at each split: 16

  OOB estimate of error rate: 24.71%

Confusion matrix:
      1      2 class.error
1 5001 3538  0.4143342
2 2588 13663 0.1592517
```

100% - 24.71% = 75.29%

100% - 15.92517% ≈ 84%

- ✓ 상대적으로 적은 데이터에도 불구하고, **전체 수요**에 대하여 **예측률 75%**
- ✓ '**더 많은 수요**'에 대하여는 **약 84%**의 높은 예측률을 보임.

모델 비교 및 평가

Overview

- ① 목표 모델 도출을 위한 통계적 분석 및 통계적 모형 소개 .
- ② Random Forest를 목표 모델로 선정.
- ③ Random Forest를 통하여, 현장성 · 편의성 · 실용성을 확보함.

편의성 : 3가지 수준으로 직관적인 수요의 정도 예측

현장성 : 저장 공간이 부족한 상황을 반영하고 적용 수월

실용성 : 높은 예측률 78% (88%)



판매 예측 모형
“ECOS” 개발

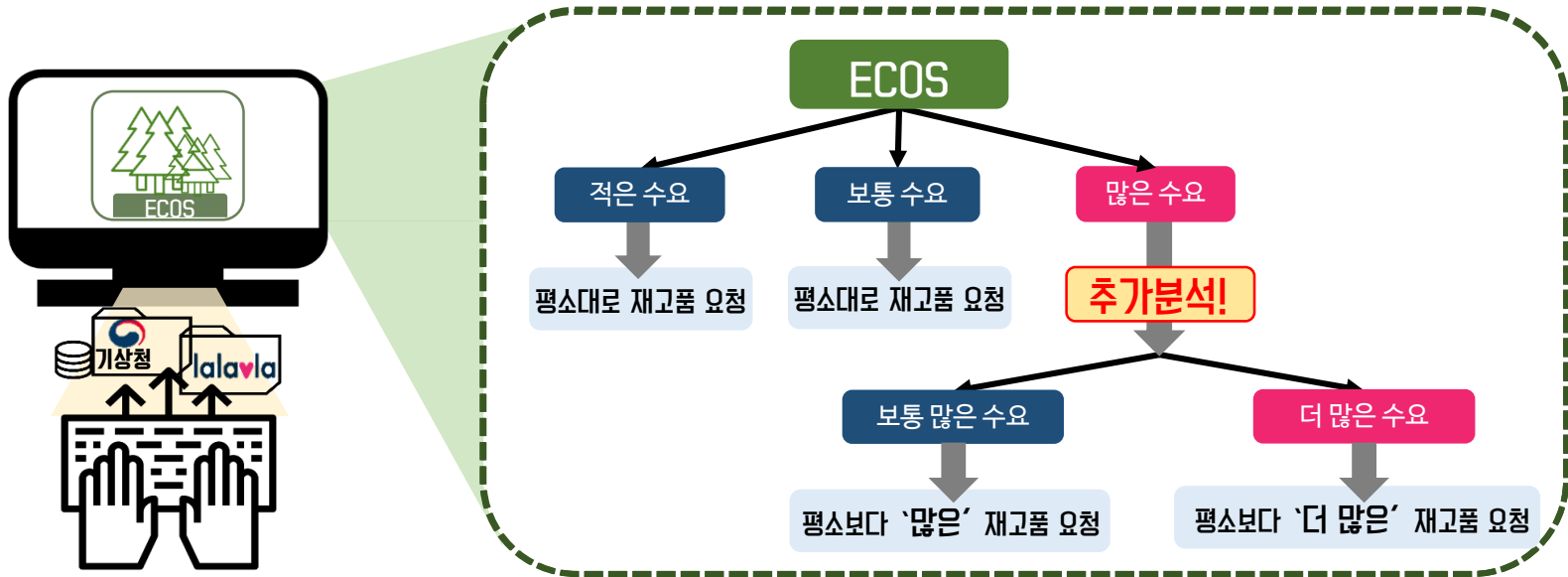
* ECOS : Easy Check Of Stock

5. 분석결과 활용방안

서비스 활용방안 / 기대 효과

서비스 활용방안

- ✓ 목적 : 기상청의 날씨 예보 데이터와 칼라블라의 최신 유통 데이터를 결합하여 미래의 판매량 예측 및 대응
- ✓ 방법 : “ 예측판매량의 숲”에서 **예상 판매량 조회**



마케팅 활용방안

실제 활용 예시 1

강서구의 한 사업장의 임종순 매니저는 본사로 웨이셜클렌저 재고물품을 요청해야 하는데 걱정이 이만저만이 아니다. 평소처럼만 팔리면 매대에 둔 것으로 충분하지만, 그게 아니라면 추가 재고 요청을 해야 하는데, 한 여름인 탓에 채종조절, 선풍기 등 극서기를 위한 물품으로 이미 재고창고가 가득 찬 탓이다 .

그래서 임종순 매니저는 팔라블라 데이터광장의 “예측판매량의 숲”(인트라넷 내 랜덤포레스트 기능이 탑재 된 데이터베이스)에 로그인하였다. (“예측판매량의 숲”에는 기상청의 과거와 예측된 날씨데이터가 자동으로 업로드 되고 있다.)

그리고 예측하고 싶은 날짜인 8월 1일을 선택하여 기상청의 예상 날씨를 확인한 후, 강서구 지점을 선택하여 결과보기를 눌렀다.

▶ 그 결과, 다행히도 **8월 1일에 웨이셜클렌저가 적게 판매될 것으로 예상**되었다.

왜냐하면 “예측판매량의 숲”이 이제까지 강서구 매장의 데이터와

예상된 날씨 데이터를 독립변수로 분석한 결과 “적은 수요일”값 을 예상하였기 때문이다.

덕분에 임종순 매니저는 작년과 달리 웨이셜클렌저 재고를 창고 밖 매장에 보관할 필요가 없었다.

✓ 입력 데이터

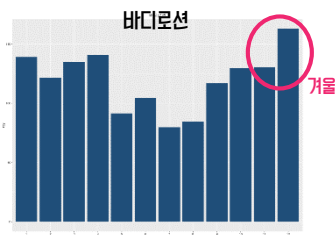
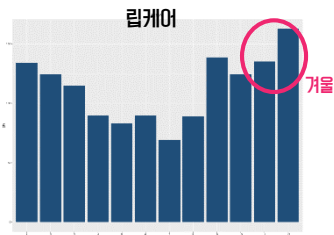
날짜	0801
지점	강서구
카테고리	웨이셜클렌저
가격	12000
계절	2(여름)
요일	수요일
최근 판매량	328
예상 최고기온	38.1
예상 최저기온	26.8
예상 평균기온	32.2
예상 최고풍속	3.9
예상 평균풍속	1.5
예상 강수량합계	0



마케팅 활용방안

✓ 실제 활용 예시 2

마포구의 한 사업장의 임종숙 매니저는 본사로 바디로션 제품의 인기로 인해 추가적으로 재고물품을 요청해야 하는데, 고민이 하나 있다. 립케어와 같은 제품도 지난 겨울에 잘 팔렸던 기억이 있어, 바디로션의 인기를 이유로 추가요청 하기엔 재고창고가 비좁기 때문이다.



그래서 임종숙 매니저는 라라블라 데이터광장의 “예측판매량의 숲”(인트라넷 내 랜덤포레스트 기능이 탑재된 데이터베이스)에 로그인하였다. (“예측판매량의 숲”에는 기상청의 과거와 예측된 날씨데이터가 자동으로 업로드 되고 있다.) 그리고 예측하고 싶은 날짜인 1월 31일을 선택하여 기상청의 예상 날씨를 확인한 후, 마포구 지점을 선택하여 결과보기를 눌렀다

그 결과, “예측판매량의 숲”은 1월 31일에 바디로션의 ‘많은 수요’를 예상하였고, 추가적인 분석으로 바디로션은 ‘더 많은 수요’로, 립케어는 ‘보통 수요’로 예상되었다.

덕분에 임종숙 매니저는 바디로션이 상대적으로 부족할 것을 깨닫고 립케어 대비 바디로션을 추가 재고를 요청하였으며, 고객들이 바디로션 재고가 없어 구매를 못하는 일을 사전에 막을 수 있었다.



임종숙 매니저

기대 효과

판매 예측 모형 “예측판매량의 숲” 활용한 서비스

개인적 측면

- 필요한 물건을 필요한 때에 소비할 수 있습니다.



기업적 측면

- 소비자의 소비패턴을 미리 파악하여 대응할 수 있습니다.
- 잘못된 소비패턴 예측에 수반하는 비용(보관비 등)을 아낄 수 있습니다.



국가적 측면

- 불필요한 자원적인 낭비를 최소화하여 사회 전체의 효용을 증대할 수 있습니다.



Q & A

감사합니다

✓ 참고문헌

Linear Discriminant Analysis Achieves High Classification Accuracy for the BOLD fMRI Response to Naturalistic Movie Stimuli

Rényi, Alfred. "On the amount of information in a random variable concerning an event." *J. Math. Sci* 1 (1966): 30-33.

Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model

Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." International Conference on Intelligent Data Engineering and Automated Learning. Springer, Berlin, Heidelberg, 2013.

Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees

Bharathidasan, S., and C. Jothi Venkataeswaran. "Improving classification accuracy based on random forest model with uncorrelated high performing trees." *Int. J. Comput. Appl* 101.13 (2014): 26-30.