

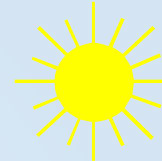


기상데이터를 통한 랄라블라 실적개선

Lalavla X Weather

접수번호 202841

팀명 ho-hyun!



SUNNY



CLOUDY



RAINY



CONTENTS

공모배경

데이터
정의

데이터
EDA 및
모델링

결과

활용방안

기대효과



1. 공모배경



01 공모배경
문제도출



CLOUDY



RAINY

OLIVE  YOUNG

lala♥la

LOHB 
LOVE HEALTH & BEAUTY

랄라블라?

올리브영, 롭스와 함께
국내 3대 H&B 스토어!

그러나..





01 공모배경 문제도출



위기의 lalavla 전략

1) 타 H&B스토어와의 **차별성** 필요

+

➔ 2) **실적 개선** 필요

"GS리테일이 랄라블라의
시장 점유율을 포기하고
실적 개선에 적극 나서고 있다"

-gs 리테일 유통 관계자

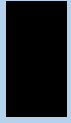
✓ 지속적인 적자로 인한 매장 수 감소, 점유율 순위 하락 위기



CLOUDY



RAINY



01 공모배경 공모주제

주제: 기상데이터를 통한 랄라블라 실적개선

✧ 판매예측 ✧ 날씨 마케팅 ✧ 효율적 재고관리

GPS로 고객위치 파악해 '맞춤형 피부 예보' 제공

동아일보 | 입력 2011-11-26 03:00 | 수정 2011-11-26 05:46

■ 날씨 마케팅의 진화 판매예측에도 활용, 재고 줄여

변덕스러운 날씨를 활용한 기업들의 마케팅 활동이 소비자들의 호응을 얻고 있다.

가장 적극적인 날씨 마케팅을 펼치는 곳은 화장품 업체다. 이들은 스마트폰에 '피부 예보' 애플리케이션(앱)을 선보이며 매출 상승 효과를 톡톡히 보고 있다.

미국 화장품 회사인 크리니크는 위성위치확인시스템(GPS)을 활용해 현재 소비자가 있는 곳의 자외선 지수, 습도 등과 그에 맞는 피부 관리 요령을 알려주는 '피부 예보' 앱을 선보인 뒤 자외선 차단제 매출이 급증했다. 올해 7, 8월 두 달 매출은 1년 전보다 300%가량 늘어났다.

아모레퍼시픽이 6년 동안 수집한 기후 통계자료 정보를 바탕으로 개발한 앱 '피부에 보++'도 날씨에 맞는 피부관리법에 관심이 많은 소비자들이 출시 3주 만에 1만 건 넘게 내려받았다.

랄라블라, 건조한 날씨 지속으로 보습제품 매출 전년 동기대비 매출 21.5% 신장

이준호 | 승인 2019.07.18 09:41 | 댓글 0



- 전년에 이어 마른 장마가 이어지며 여름철임에도 불구하고 단계별 보습케어 실시하는 고객 크게 늘어나
- 랄라블라, 18일부터 고객들의 알뜰한 쇼핑 돕는 역(逆)시즌 및 시즌 특가 행사 진행
- 멤버십 3만원 이상 구매 고객에게 6천원 할인혜택 제공하며 최대 60% 할인으로 고객 만족 높일 예정

1 날씨에 따른 '피부예보' 제공으로
매출을 급증시킨 화장품 회사 크리니크의 선례

2 날씨에 따라 특정 카테고리의 매출의 급증을 노릴 수 있고
이를 예측할 수도 있음.

화장품 업계 실적개선, 답은 '날씨'다.



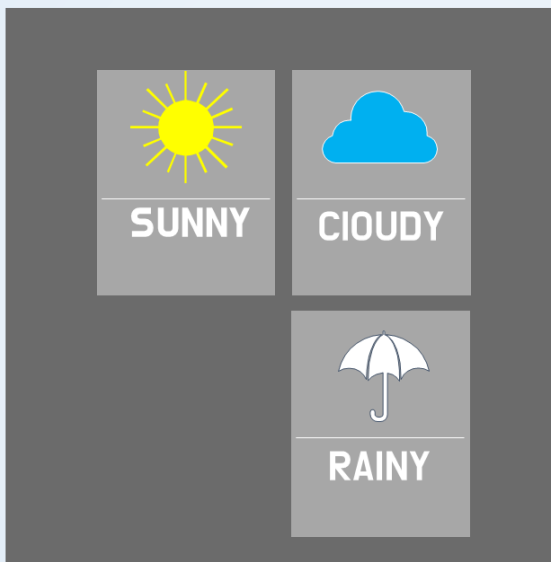
2.데이터 정의



02 데이터 정의 판매량 예측 알고리즘



랄라블라 데이터



날씨 데이터



모델링



‘카테고리별’ 판매량 예측
-> 실적개선

출
처

GS 리테일

출
처

날씨마루
서울 열린데이터광장
기상자료개방포털



02 데이터소개

Input 전처리



칼라블라 데이터

날짜

(201601~20181231)

연령대

(10대/2,30대/4,50대/60대)

성별

(남, 여)

지역별

(서울시 자치구 20개)

카테고리 별

(10개 종류)

판매량 데이터

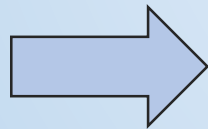
| 날짜 | | 화장품 카테고리 | | | | | | | | | |
|----------|------|----------|---------|----------|------------|---------|------------|--------|-------------|--------|--|
| Sale_dt | nail | lipcolor | lipcare | maskpack | bodylotion | suncare | depilatory | weight | creamlotion | facial | |
| 20160101 | 1491 | 6260 | 6192 | 35566 | 6444 | 689 | 72 | 1307 | 5531 | 6592 | |
| 20160102 | 1853 | 7065 | 8344 | 76466 | 8378 | 831 | 145 | 1635 | 12502 | 6481 | |
| 20160103 | 1673 | 6845 | 7648 | 66757 | 8123 | 615 | 108 | 2112 | 8927 | 7542 | |
| 20160104 | 1783 | 7137 | 9032 | 76978 | 9251 | 688 | 290 | 1417 | 10424 | 8054 | |
| 20160105 | 1238 | 6883 | 7356 | 74894 | 8819 | 651 | 108 | 1563 | 10603 | 7906 | |
| 20160106 | 1637 | 7103 | 9434 | 65303 | 8266 | 687 | 72 | 1054 | 9253 | 7977 | |
| 20160107 | 2002 | 7210 | 9401 | 65451 | 8707 | 579 | 180 | 1416 | 10785 | 7905 | |

↓

201601~20181231 네일, 립 색조, 립 케어, 마스크팩, 바디로션, 선 케어, 제모제, 체중조절, 크림로션, 웨이셜 **판매량**

날짜별로만 10개의 카테고리별로 **판매량**을 종합

[1096 rows X 11 columns(날짜+카테고리10개)] table 생성



연령대

성별

지역:

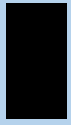
: 모든 품목에서 2,30
대-4,50대-10대-60
대 이상 순으로 판매
량이 많음

: 모든 품목에서 여성
판매량이 압도적으로
많음

해당 지역의 칼라블라
매장 수와 판매량이
단순 비례함

->수요예측을 위한 주요 지표로 적절하지 않음

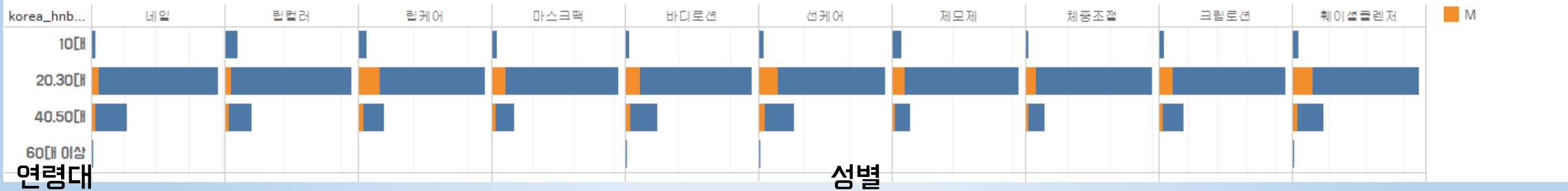
연령대, 성별, 지역, 날짜 중 '날짜' 만 수요 예측을 위한 지표로 선정.



02 데이터소개

Input 전처리

카테고리별 연령&성별 판매량



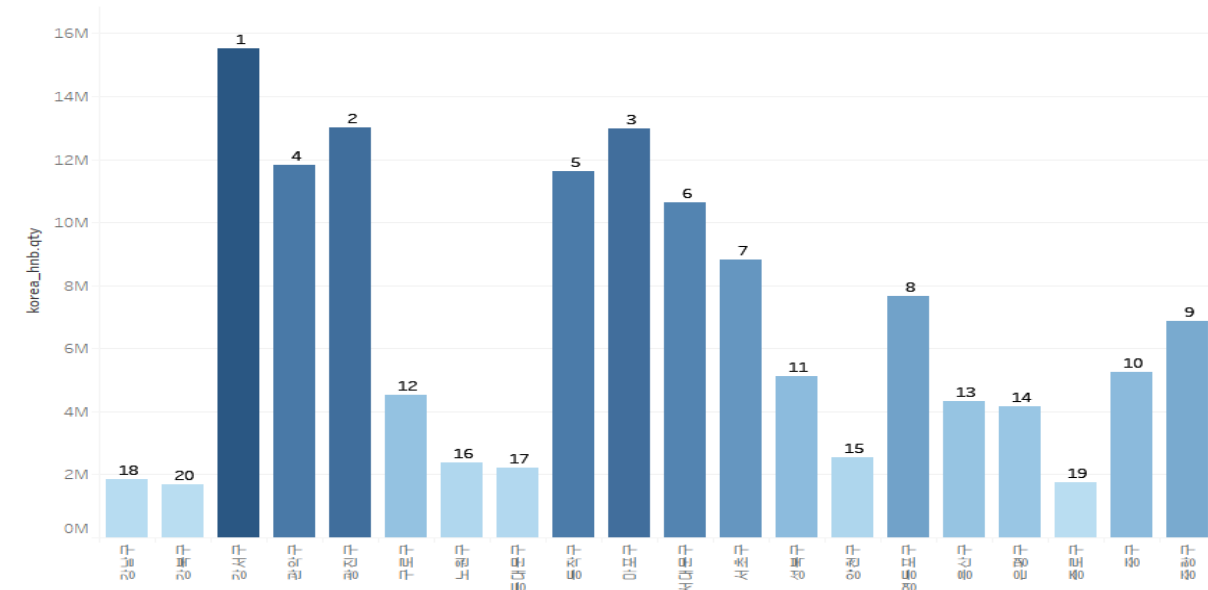
연령대

: 모든 품목에서 2,30대-4,50대-10대-60대 이상 순으로 판매량이 많음

성별

: 모든 품목에서 여성 판매량이 압도적으로 많음

지역별 판매량



지역:

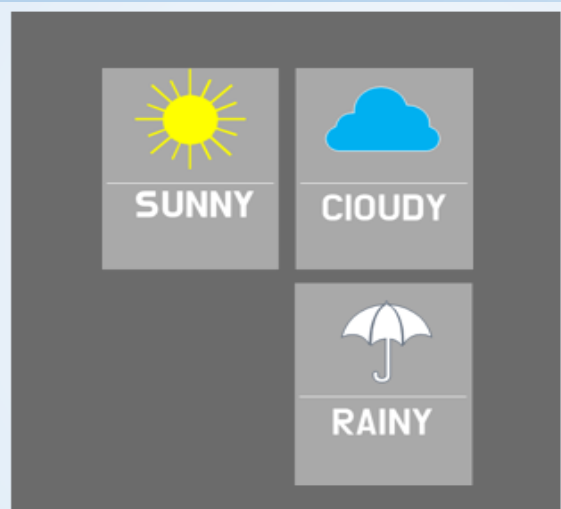
해당 지역의 칼라블라 매장 수와 판매량이 단순 비례함

판매량 1등 강서구 -> 강서구 내 매장 수 6개

판매량 꼴등 강북구 -> 강북구 내 매장 수 2개



02 데이터소개 Input 전처리



날씨 데이터

총 19가지 날씨변수

- 기온**
Avg_ta(평균 온도)
Min_ta(최저 온도)
Max_ta(최고 온도)
Avg_sufta(평균 지면온도)
Avg_dewp(평균 이슬점 온도)
- 강수량**
Cont_rn(강수 지속 시간)
Sum_rn(강수량)
- 습도**
Avg_rhm(평균 상대습도)
Min_rhm(최저 상대습도)
- 일조량**
Sum_sunt(일조 합계시간)
Max_sunh(최대 일조시간)
Sum_sun(일조량)
- 미세먼지**
Avg_pm10(미세먼지)
Avg_pm2.5(초미세먼지)
- 풍속**
Maxi_ws(순간 최대 풍속)
Max_ws(최대 풍속)
Avg_ws(평균 풍속)
- 적설량**
Sum_snow(적설량)
Sum_snow3h(3시간 적설량)

19가지의 날씨 변수를 7가지의 대분류 (기온, 강수량, 미세먼지, 풍속, 습도, 일조량, 적설량)로 분류

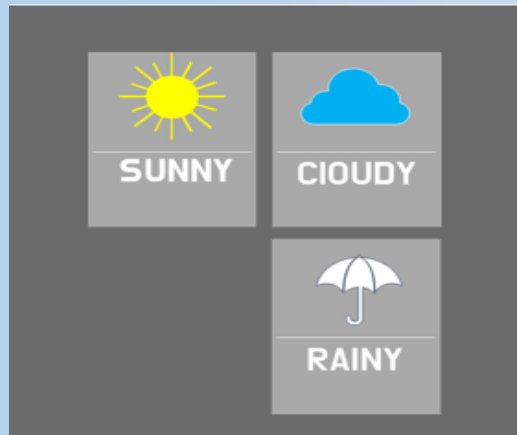


02 데이터소개

Input 전처리

날짜

날씨 변수



| Sale_dt | avg_ta | min_ta | max_ta | cont_rn | sum_rn | avg_pm10 | avg_pm2.5 | maxi_ws | max_ws | avg_ws | avg_dewp | min_rhm | avg_rhm | sum_sunt | max_sunh | sum_sun | sum_snow | sum_snow3h | avg_sufta | ta+rhm |
|----------|---------|--------|--------|---------|--------|----------|-----------|---------|--------|--------|----------|---------|---------|----------|----------|---------|----------|------------|-----------|--------|
| 20160101 | 1.2-3.3 | | 4 | 0 | 0 | 70.73684 | 44.18421 | 5.6 | 3.5 | 1.6 | -3.3 | 56 | 73 | 2.1 | 0.9 | 4.61 | 0 | 0 | 0.2 | 74.2 |
| 20160102 | 5.7 | 1 | 9.5 | 0 | 0 | 53.78947 | 31.21053 | 6.9 | 4.5 | 2 | 1.9 | 62 | 76.9 | 3.6 | 0.76 | 4.5 | 0 | 0 | 2.8 | 82.6 |
| 20160103 | 6.5 | 5.1 | 9.4 | 0 | 0 | 96.07895 | 57.89474 | 5.7 | 4 | 1.8 | 3.2 | 61 | 80.6 | 2.2 | 0.89 | 4.85 | 0 | 0 | 4.8 | 87.1 |
| 20160104 | 2 | -2.5 | 5.3 | 0 | 0 | 76.20513 | 45.07692 | 8.6 | 5.1 | 3.1 | -6.9 | 28 | 54.4 | 8.6 | 1.51 | 9.03 | 0 | 0 | 3 | 56.4 |

날짜
(201601~20181231)
시간
(24시간)
날씨
(19개 지표)
측정값 데이터

201601~20181231



이상치 : 날씨 카테고리별 3σ 이상의 값을 이상치라고 판단, 중앙값으로 변경
결측치 : 중앙값으로 변경
시간 : 평균값으로 날씨 통합

날짜별 19개의 카테고리별로 측정값 전처리 결과
[1096 rows X 20 columns(날짜+날씨 19개)] table 생성



02 데이터소개

Input 데이터 join

| Sale_dt | nail | lipcolor | lipcare | maskpack | bodylotion | suncare | depilatory | weight | creamlotion | facial | avg_ta | min_ta | max_ta | cont_rn | sum_rn | avg_pm10 | avg_pm2.5 | maxi_ws | max_ws | avg_ws | avg_dewp | min_rhm | avg_rhm | sum_sunt | max_sunh | sum_sun | sum_snow | sum_snow_3h | avg_sufta | ta+rhm |
|----------|------|----------|---------|----------|------------|---------|------------|--------|-------------|--------|--------|--------|--------|---------|--------|------------|-----------|---------|--------|--------|----------|---------|---------|----------|----------|---------|----------|-------------|-----------|---------|
| 20160101 | 1491 | 6260 | 6192 | 35566 | 6444 | 689 | 72 | 1307 | 5531 | 6592 | 1.2 | -3.3 | 4 | 0 | 0 | 70.73684 | 44.18421 | 5.6 | 3.5 | 1.6 | -3.3 | 56 | 73 | 2.1 | 0.9 | 4.61 | 0 | 0 | 0.2742 | 274.2 |
| 20160102 | 1853 | 7065 | 8344 | 76466 | 8378 | 831 | 145 | 1635 | 12502 | 6481 | 5.7 | 1 | 9.5 | 0 | 0 | 53.78947 | 31.21053 | 6.9 | 4.5 | 2 | 1.9 | 62 | 76.9 | 3.6 | 0.76 | 4.5 | 0 | 0 | 2.8826 | 288.26 |
| 20160103 | 1673 | 6845 | 7648 | 66757 | 8123 | 615 | 108 | 2112 | 8927 | 7542 | 6.5 | 5.1 | 9.4 | 0 | 0 | 96.07895 | 57.89474 | 5.7 | 4 | 1.8 | 3.2 | 61 | 80.6 | 2.2 | 0.89 | 4.85 | 0 | 0 | 4.8871 | 488.71 |
| 20160104 | 1783 | 7137 | 9032 | 76978 | 9251 | 688 | 290 | 1417 | 10424 | 8054 | 2 | -2.5 | 5.3 | 0 | 0 | 76.20513 | 45.07692 | 8.6 | 5.1 | 3.1 | -6.9 | 28 | 54.4 | 8.6 | 1.51 | 9.03 | 0 | 0 | 356.4 | 356.4 |
| 20160105 | 1238 | 6883 | 7356 | 74894 | 8819 | 651 | 108 | 1563 | 10603 | 7906 | -2.7 | -4.8 | 1.5 | 0 | 0 | 34.33333 | 16.66667 | 7.3 | 4.6 | 2.3 | -15.2 | 22 | 39.4 | 7.9 | 1.56 | 8.22 | 0 | 0 | 0.1367 | 136.7 |
| 20160106 | 1637 | 7103 | 9434 | 65303 | 8266 | 687 | 72 | 1054 | 9253 | 7977 | -1.7 | -4.9 | 1.7 | 0 | 0 | 58.2953846 | 20.71795 | 6.9 | 3.8 | 1.8 | -9.7 | 42 | 54.3 | 2.2 | 0.98 | 4.69 | 0 | 0 | -0.5526 | -55.26 |
| 20160107 | 2002 | 7210 | 9401 | 65451 | 8707 | 579 | 180 | 1416 | 10785 | 7905 | -3.4 | -5.9 | 1.4 | 0 | 0 | 41.79487 | 20.71795 | 9 | 5.2 | 2.5 | -12.4 | 25 | 51.8 | 9.1 | 1.65 | 9.76 | 0 | 0 | -148.4 | -148.4 |
| 20160108 | 1636 | 8489 | 9252 | 78253 | 9621 | 724 | 217 | 1126 | 11516 | 8600 | -3.3 | -6.9 | 1 | 0 | 0 | 37.02564 | 18.48718 | 9.4 | 5.9 | 2 | -13 | 24 | 49.8 | 9.2 | 1.66 | 9.83 | 0 | 0 | -1.5465 | -154.65 |
| 20160109 | 1891 | 8200 | 9691 | 74895 | 10274 | 798 | 181 | 1271 | 14727 | 8050 | -2.1 | -6.2 | 2.4 | 0 | 0 | 39.17949 | 19.89744 | 9 | 5.3 | 2.1 | -9.6 | 32 | 57.1 | 8 | 1.64 | 9.28 | 0 | 0 | -155 | -155 |

날짜 기준

알라블라 10개의 카테고리 판매량데이터

날씨 19가지 변수 데이터

Join

[1096 rows X 30columns(날짜+카테고리10개+날씨 19개) table 생성



3. 데이터 EDA 및 모델링



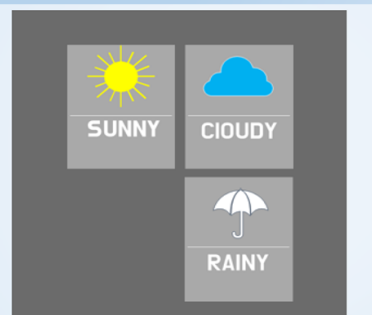
03 데이터 EDA 및 모델링

데이터 EDA

모델링을 위한 전처리



02 데이터소개 Input 전처리



날씨 데이터
총 19가지 날씨변수

- 기온**
Avg_ta(평균 온도)
Min_ta(최저 온도)
Max_ta(최고 온도)
Avg_sufta(평균 지면온도)
Avg_dewp(평균 이슬점 온도)
- 강수량**
Cont_rn(강수 지속 시간)
Sum_rn(강수량)
- 습도**
Avg_rhm(평균 상대습도)
Min_rhm(최저 상대습도)
- 일조량**
Sum_sunt(일조 합계시간)
Max_sunh(최대 일조시간)
Sum_sun(일조량)
- 미세먼지**
Avg_pm10(미세먼지)
Avg_pm2.5(초미세먼지)
- 풍속**
Maxi_ws(순간 최대 풍속)
Max_ws(최대 풍속)
Avg_ws(평균 풍속)
- 적설량**
Sum_snow(적설량)
Sum_snow3h(3시간 적설량)

1) 19가지의 모든 날씨변수가 판매량에 영향을 미치는 것은 아니다.

⇒ 모델에 영향을 미치는 중요한 대분류를 살펴본다.



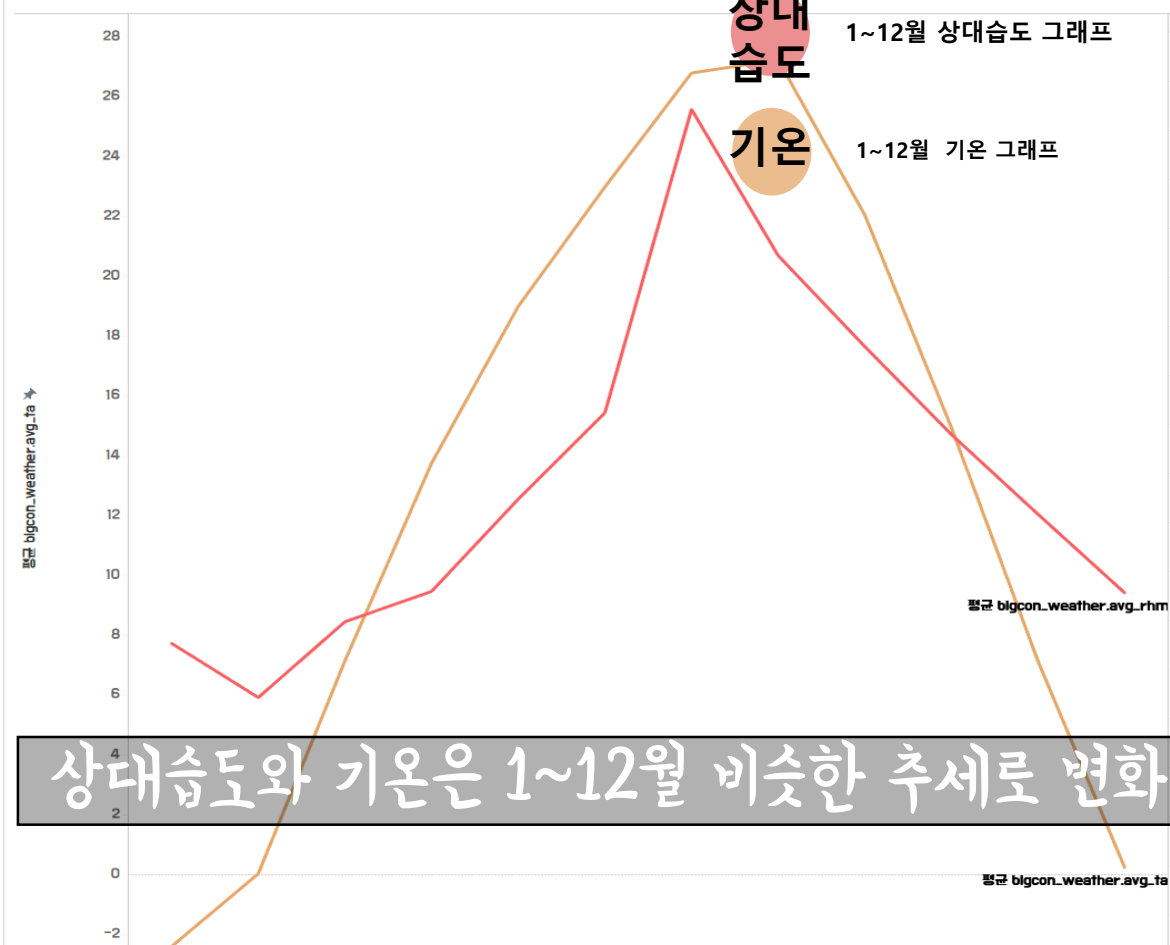
03 데이터 EDA 및 모델링

데이터 시각화

시각화> 날씨 데이터 월별 시각화

-데이터시각화 tool **tableau** 이용

월별 평균 상대습도와 평균온도



월별 평균 상대습도와 강수량



상대습도, 기온, 강수량 = 유의미한 대분류



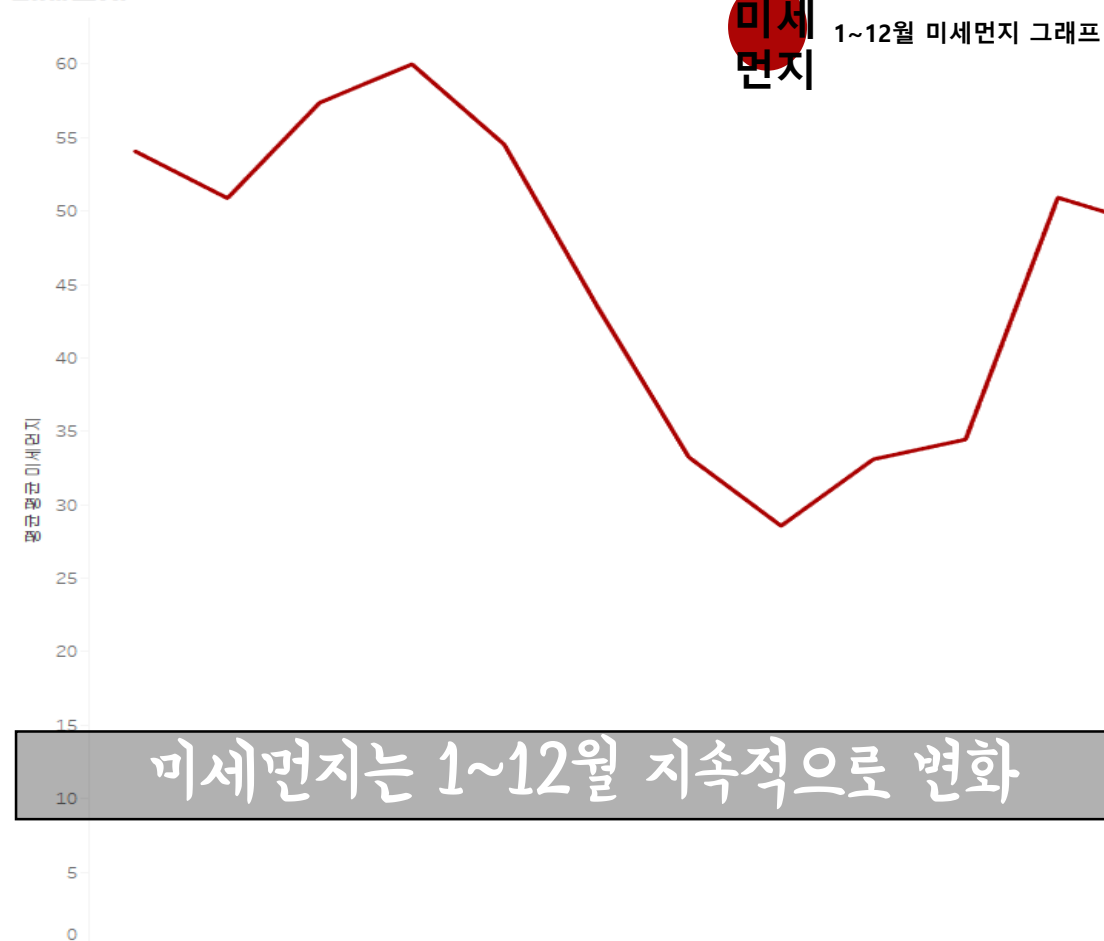
03 데이터 EDA 및 모델링

데이터 시각화

시각화> 날씨 데이터 월별 시각화

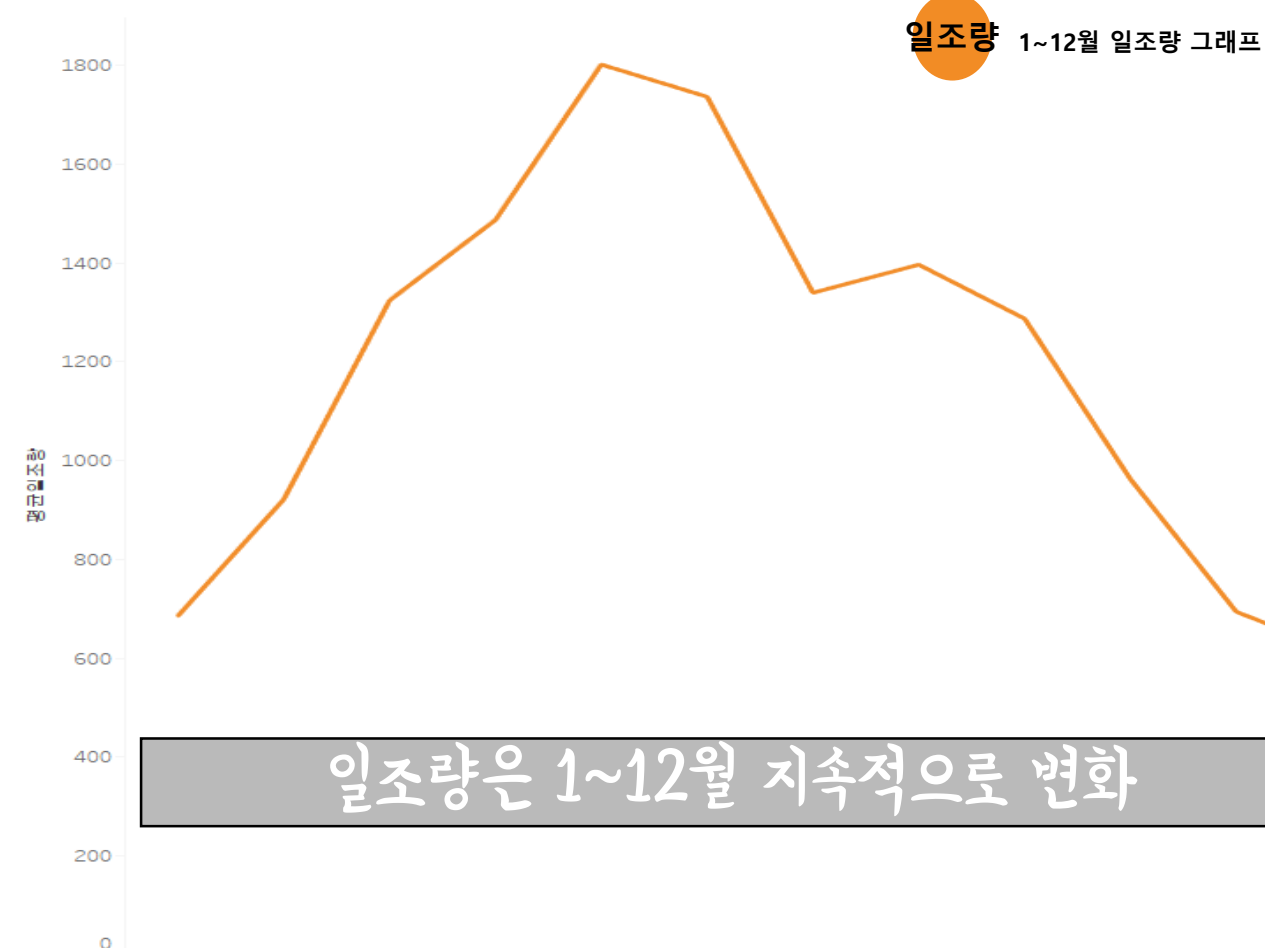
-데이터시각화 tool **tableau** 이용

미세먼지



미세먼지는 1~12월 지속적으로 변화

일조량



일조량은 1~12월 지속적으로 변화

미세먼지, 일조량 = 유의미한 대분류



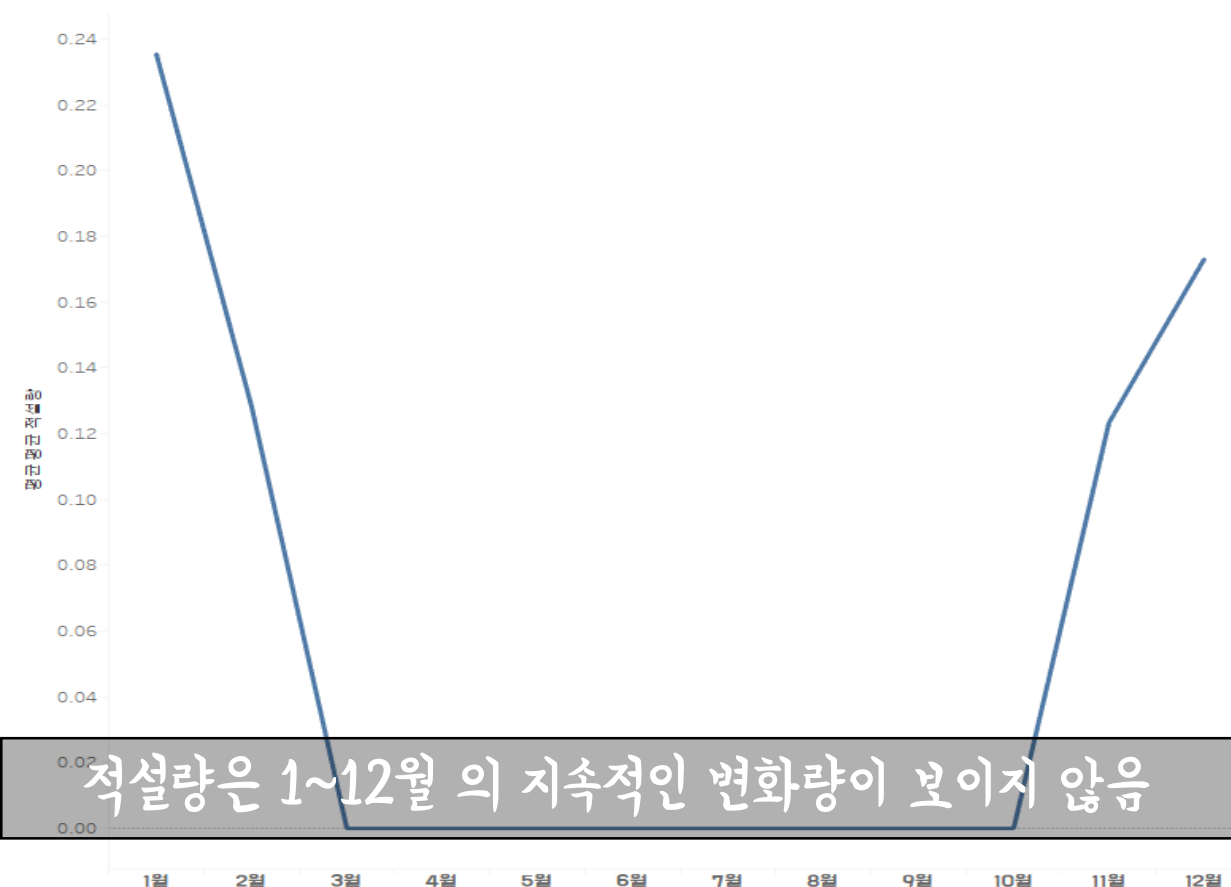
03 데이터 EDA 및 모델링

데이터 시각화

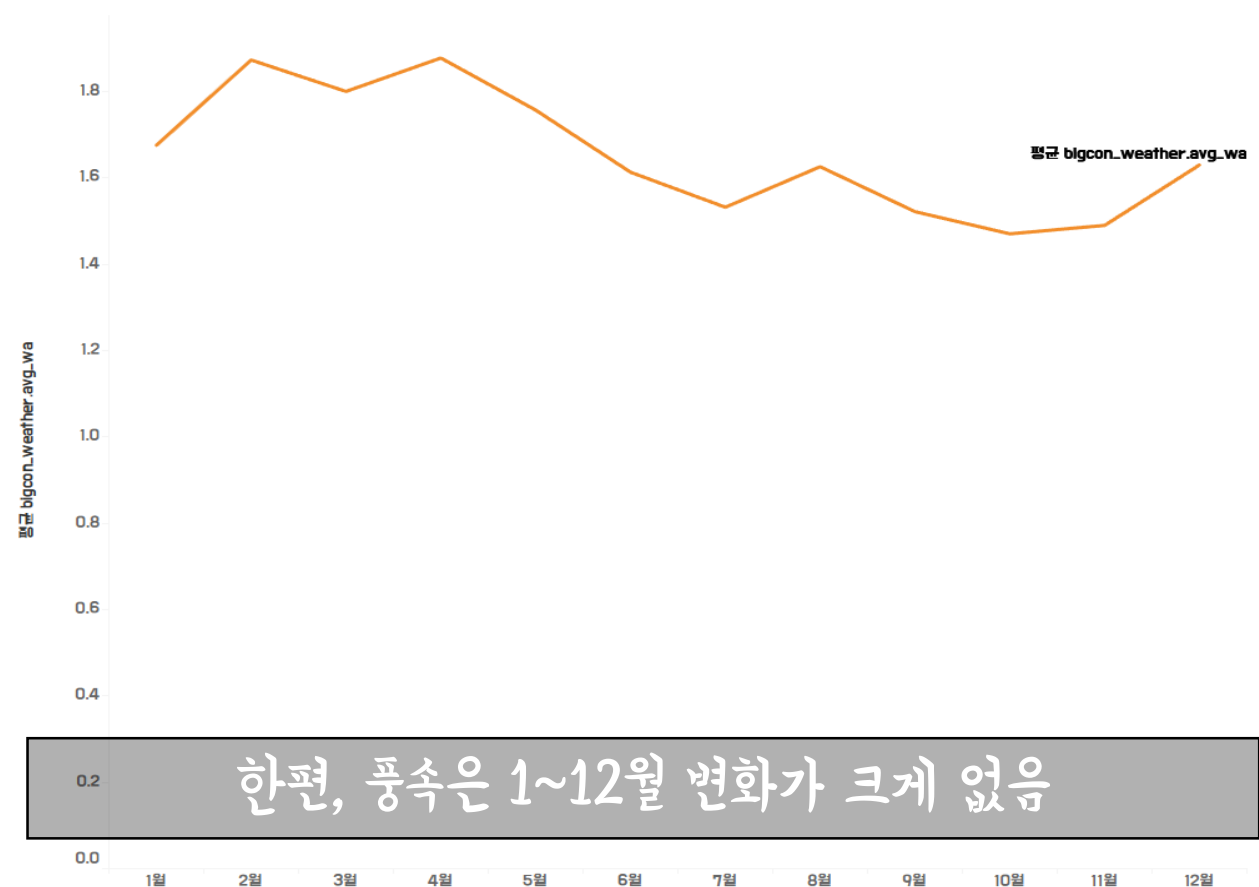
시각화> 날씨 데이터 월별 시각화

-데이터시각화 tool **tableau** 이용

적설량



월별 평균 풍속



적설량, 풍속 = 무의미한 대분류



03 데이터 EDA 및 모델링

데이터 시각화

시각화> 날씨 데이터 월별 시각화

-데이터시각화 tool **tableau** 이용

1) 19가지의 모든 날씨변수가 판매량에 영향을 미치는 것은 아니다.

⇒ 모델에 영향을 미치는 중요한 지표를 뽑아서 살펴본다

월별로 변화량이 큰 날씨변수가 있고,
연중 비슷한 날씨변수가 있음.

**변화량이 큰 날씨변수를 따로 뽑고 이를
모델에 영향을 미치는 중요한 지표로 설정.**

=>중요지표: 기온, 일조량, 상대습도(강수량), 미세먼지



03 데이터 EDA 및 모델링

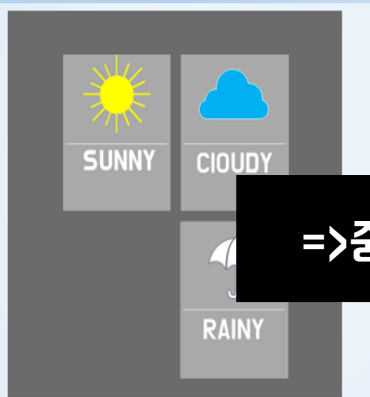
데이터 EDA

EDA> 탐색적 분석

-데이터시각화 tool **tableau** 이용



02 데이터소개 Input 전처리



날씨 데이터

총 19가지 날씨변수

1 기온

Avg_ta(평균 온도)
Min_ta(최저 온도)
Max_ta(최고 온도)

3 습도

Avg_rhm(평균 상대습도)
Min_rhm(최저 상대습도)

5 미세먼지

Avg_pm10(미세먼지)
Avg_pm2.5(초미세먼지)

2 강수량

Cont_rn(강수 지속 시간)
Sum_rn(강수량)

Sum_sunh(일조 시간)
Max_sunh(최대 일조시간)
Sum_sun(일조량)

7 적설량

Sum_snow(적설량)
Sum_snow3h(3시간 적설량)

=>중요지표: 기온, 일조량, 상대습도(강수량), 미세먼지

중요지표별로

판매량에 영향을 받
는 화장품 품목을

'민감품목'

영향을 받지 않는
화장품 품목을

'둔감품목'

으로 분석하고자 함.



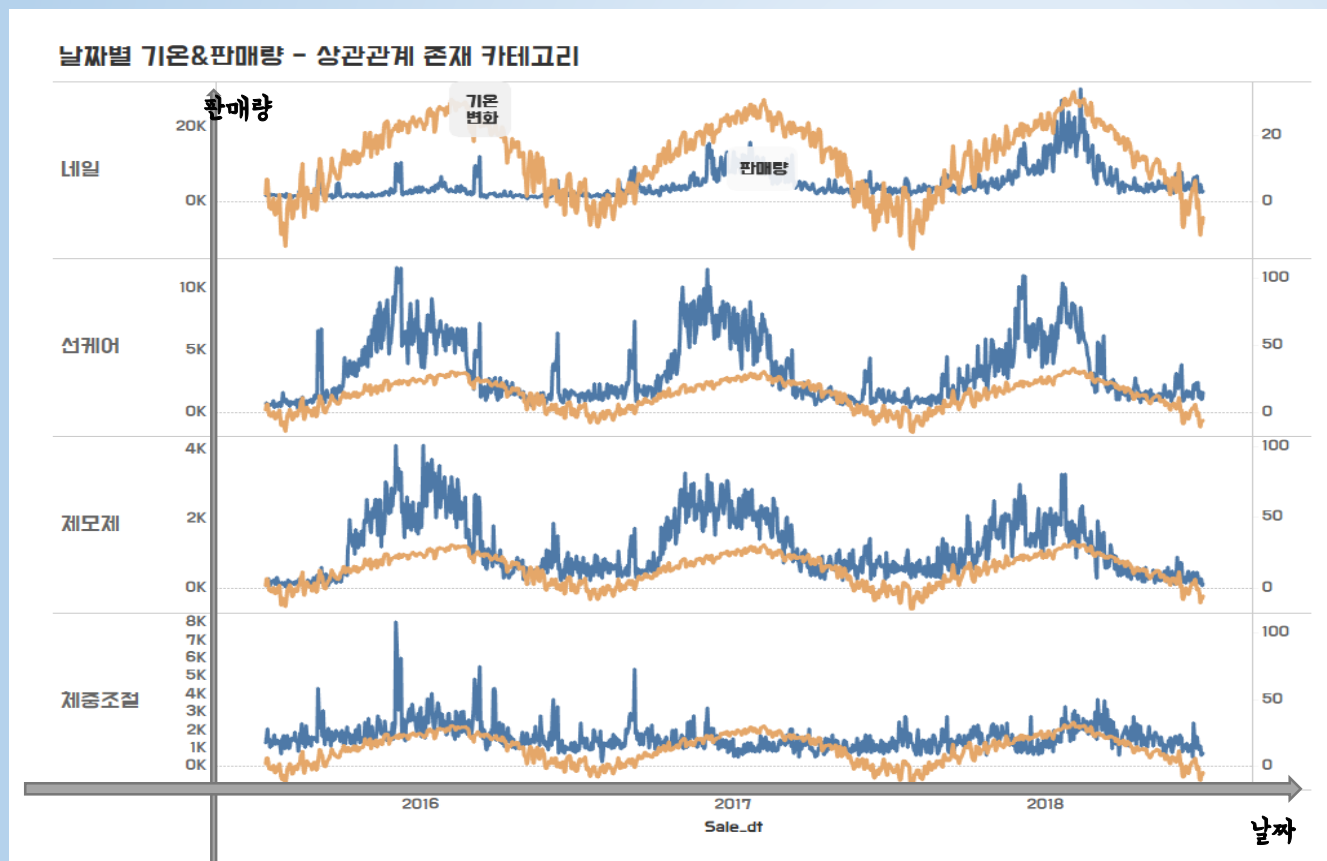
03 데이터 EDA 및 모델링

데이터 EDA

〈날씨 요소별 랄라블라 판매 데이터 EDA〉

-데이터시각화 tool tableau 이용

1) 기온별



기온

2016~2018 기온변화 그래프

판매량

2016~2018 판매량변화 그래프

네일, 선크림,
제모제, 제중조절

-> 기온의 변화와
양(+)의 상관관계를
가지고 있다

-기온민감품목



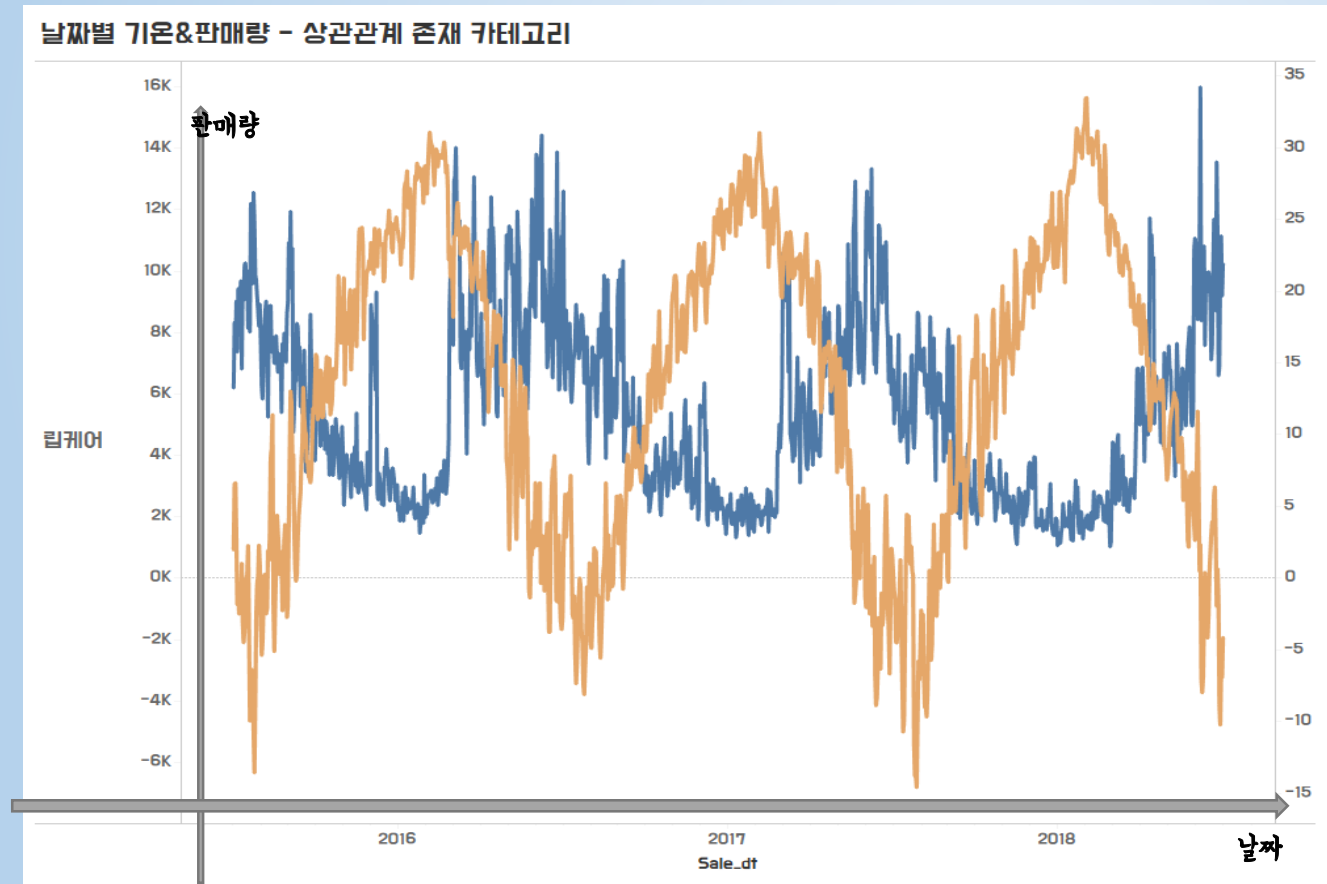
03 데이터 EDA 및 모델링

데이터 EDA

〈날씨 요소별 랄라블라 판매 데이터 EDA〉

-데이터시각화 tool tableau 이용

1) 기온별



기온 2016~2018 기온변화 그래프

판매량 2016~2018 판매량변화 그래프

링크어

-> 기온의 변화와
음(-)의 상관관계를
가지고 있다

-기온민감품목



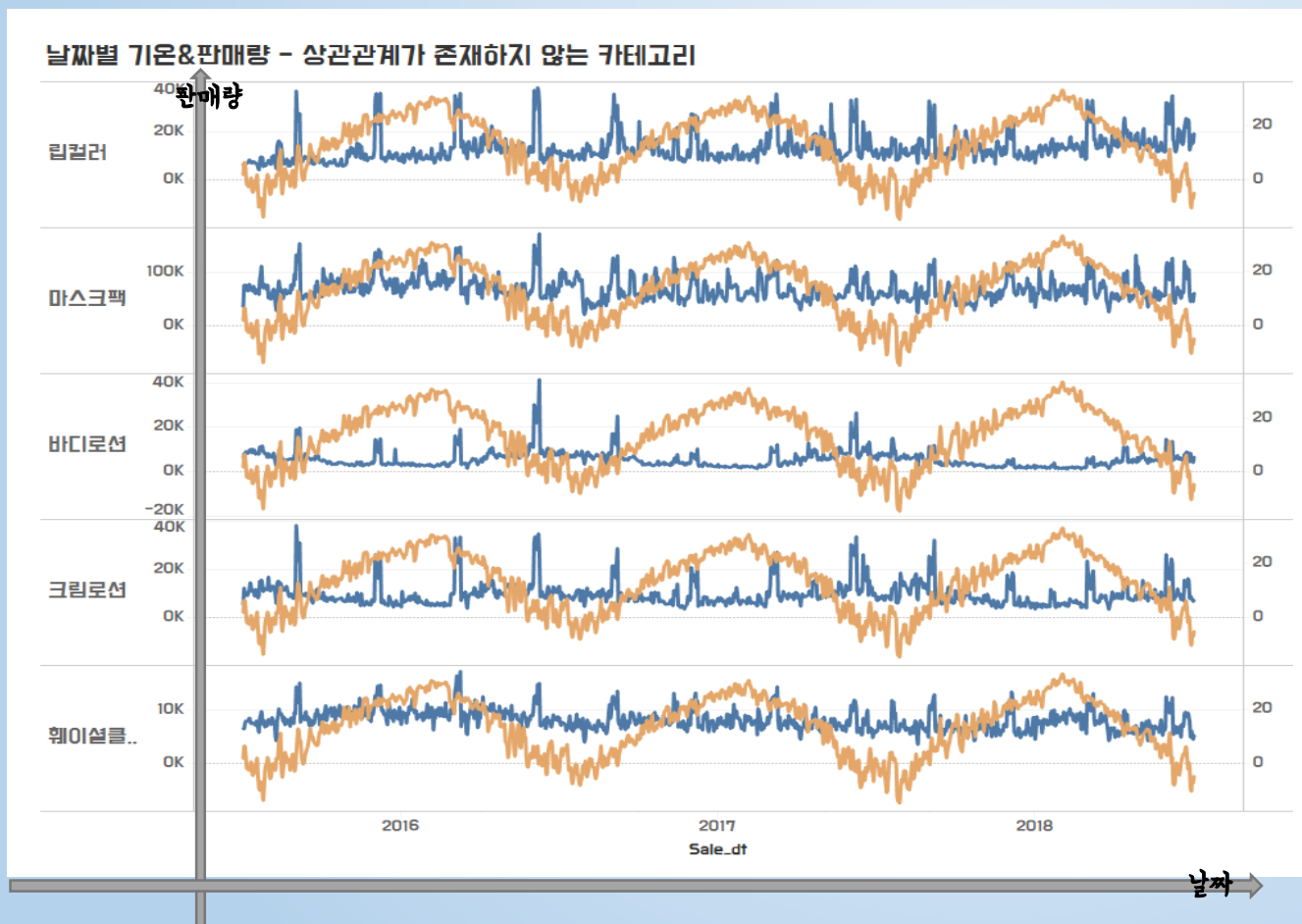
03 데이터 EDA 및 모델링

데이터 EDA

〈날씨 요소별 란라블라 판매 데이터 EDA〉

-데이터시각화 tool tableau 이용

1) 기온별



기온 2016~2018 기온변화 그래프

판매량 2016~2018 판매량변화 그래프

립컬러, 마스크팩,
바디로션, 크림로션, 웨이셜클렌저

-> 기온의 변화와
상관관계가 없다.

-기온둔감품목



03 데이터 EDA 및 모델링

데이터 EDA

-기온-판매량 추세모델설명

* 개별 추세선

*추세선 모델

-기온 평균에 주어진 판매량에 대한 선형 추세 모델

*유효 모델구간 :

$p \leq 0.05$

*분산 분석:

| 필드 | DF | SSE | MSE | F | p-값 |
|----|----|-----------|---------|---------|----------|
| 기온 | 4 | 171904.99 | 42976.2 | 140.289 | < 0.0001 |

*모델링된 관측값 수: 153

*필터링된 관측값 수:0

*모델 자유도:6

*잔차 자유도(DF):147

*SSE(오차제곱합):45032

*MSE(평균 제곱 오차):306.34

*R-제곱:0.865753

*표준 오차:17.5026

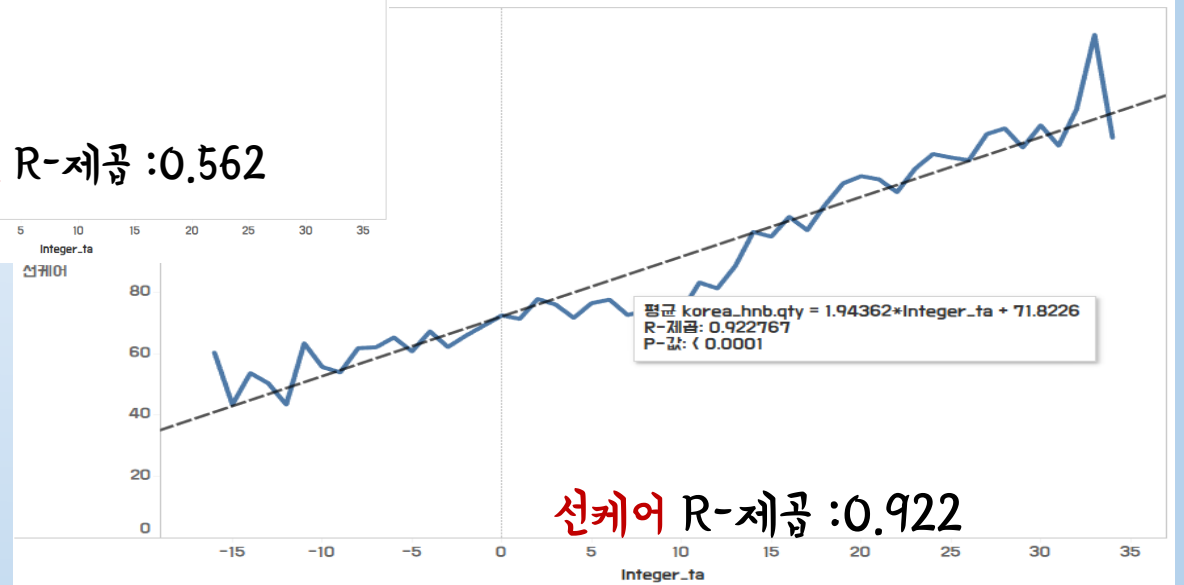
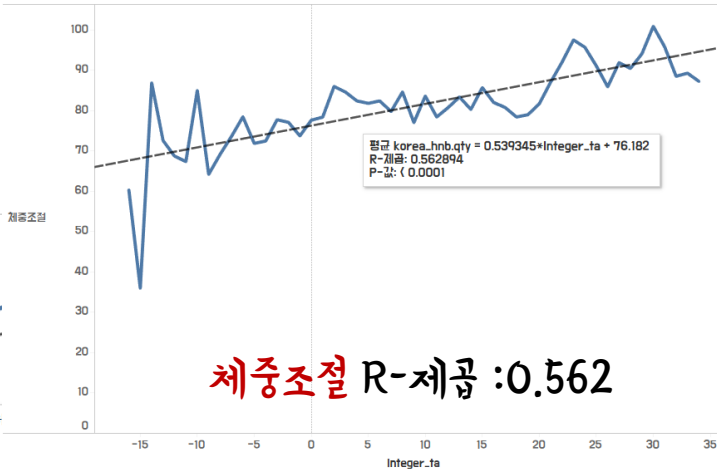
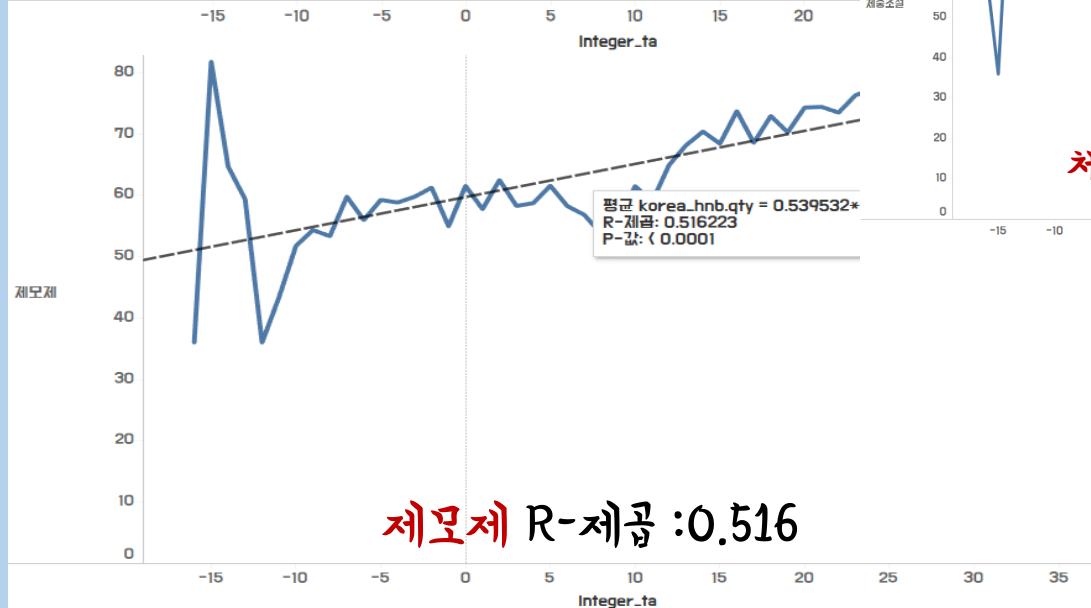
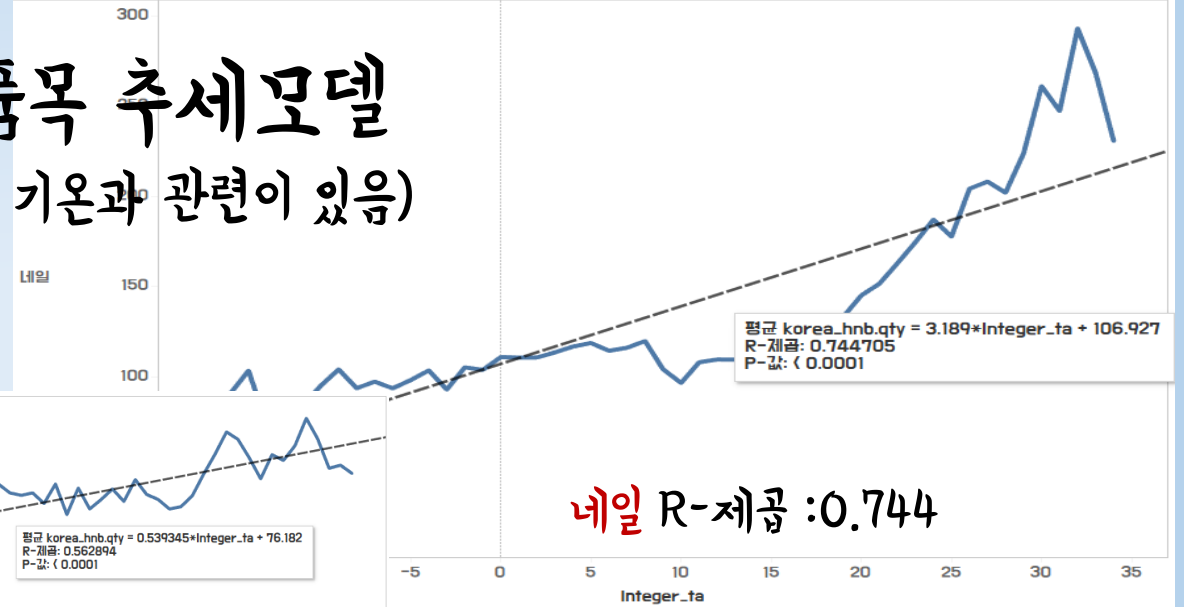
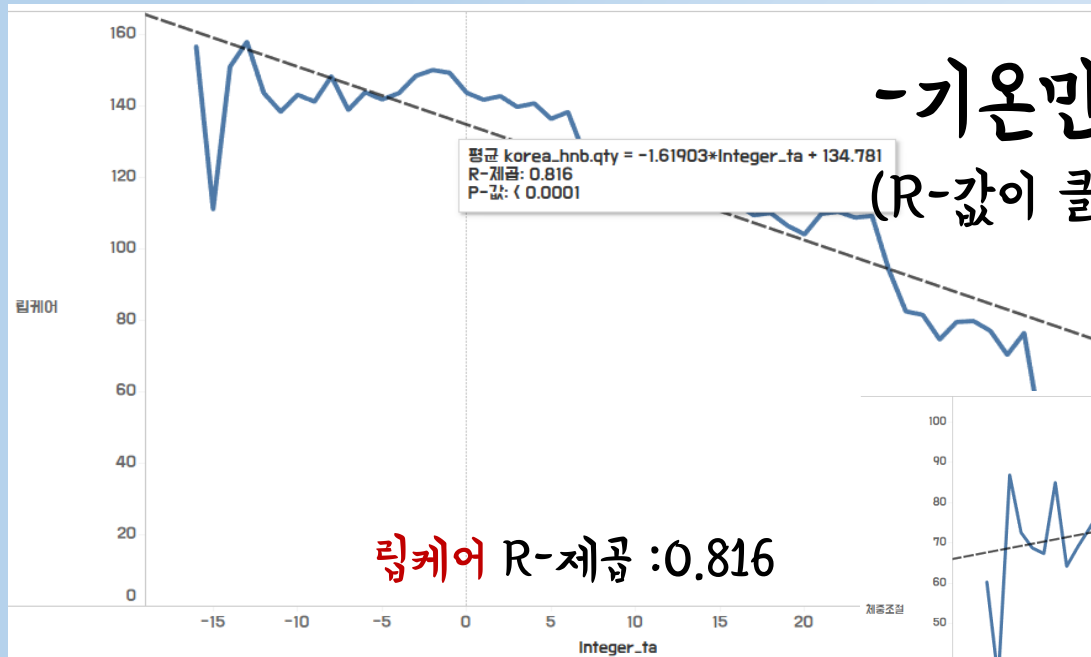
*p-값(유의):< 0.0001

| 패널 행 | 열 | 라인 p-값 | DF | 계수 | 값 | 표준 오차 | t-값 | p-값 |
|---------|------------|-----------|----|----|-----------|-----------|----------|-----------|
| 네일 | Integer_ta | < 0.0001 | 49 | | 3.189 | 0.266738 | 11.9555 | < 0.0001 |
| | | | | | 106.927 | 4.60204 | 23.2346 | < 0.0001 |
| 립컬러 | Integer_ta | 0.0424471 | 49 | | 0.490984 | 0.235657 | 2.08347 | 0.0424471 |
| | | | | | 205.506 | 4.0658 | 50.5452 | < 0.0001 |
| 립케어 | Integer_ta | < 0.0001 | 49 | | -1.61903 | 0.10983 | -14.7413 | < 0.0001 |
| | | | | | 134.781 | 1.8949 | 71.1286 | < 0.0001 |
| 마스크팩 | Integer_ta | 0.0058061 | 49 | | 2.59944 | 0.901075 | 2.88482 | 0.0058061 |
| | | | | | 911.011 | 15.5463 | 58.6 | < 0.0001 |
| 바디로션 | Integer_ta | < 0.0001 | 49 | | -1.38424 | 0.209102 | -6.61995 | < 0.0001 |
| | | | | | 140.109 | 3.60763 | 38.8367 | < 0.0001 |
| 선크어 | Integer_ta | < 0.0001 | 49 | | 1.94362 | 0.0803281 | 24.196 | < 0.0001 |
| | | | | | 71.8226 | 1.3859 | 51.8237 | < 0.0001 |
| 제모제 | Integer_ta | < 0.0001 | 49 | | 0.539532 | 0.0746145 | 7.23093 | < 0.0001 |
| | | | | | 59.6018 | 1.28732 | 46.299 | < 0.0001 |
| 체중조절 | Integer_ta | < 0.0001 | 49 | | 0.539345 | 0.0678967 | 7.94361 | < 0.0001 |
| | | | | | 76.182 | 1.17142 | 65.0337 | < 0.0001 |
| 크림로션 | Integer_ta | 0.0006862 | 49 | | -0.857002 | 0.236405 | -3.62514 | 0.0006862 |
| | | | | | 183.271 | 4.0787 | 44.9336 | < 0.0001 |
| 웨이셜클렌저 | Integer_ta | < 0.0001 | 49 | | 0.604075 | 0.116387 | 5.19024 | < 0.0001 |
| | | | | | 138.034 | 2.00802 | 68.7411 | < 0.0001 |

03 데이터 EDA 및 모델링

데이터 EDA

-기온민감품목 추세모델 (R-값이 클 수록 기온과 관련이 있음)

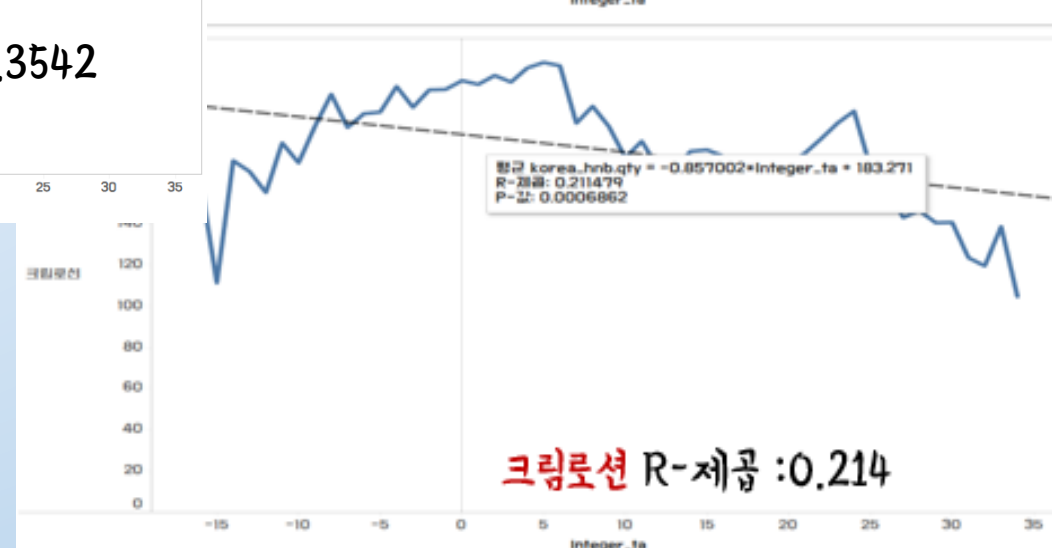
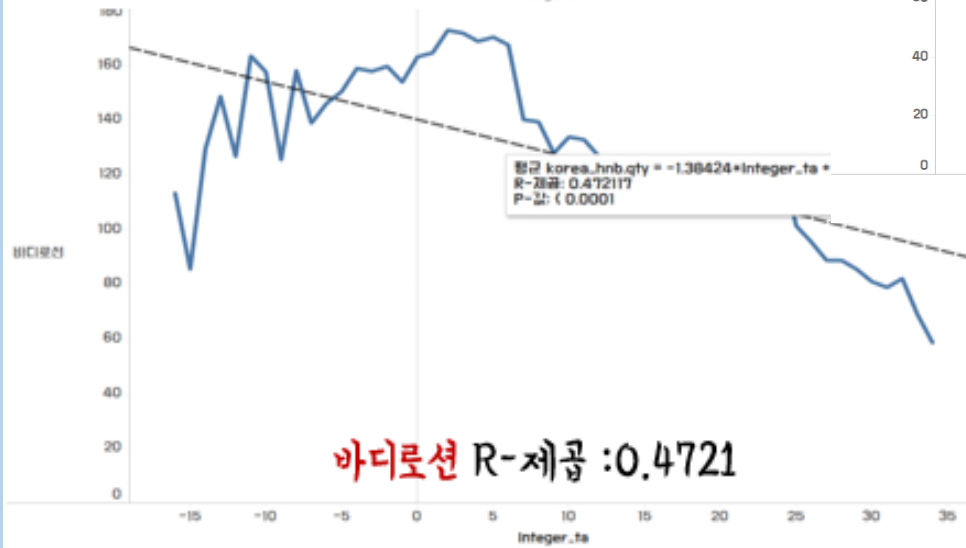
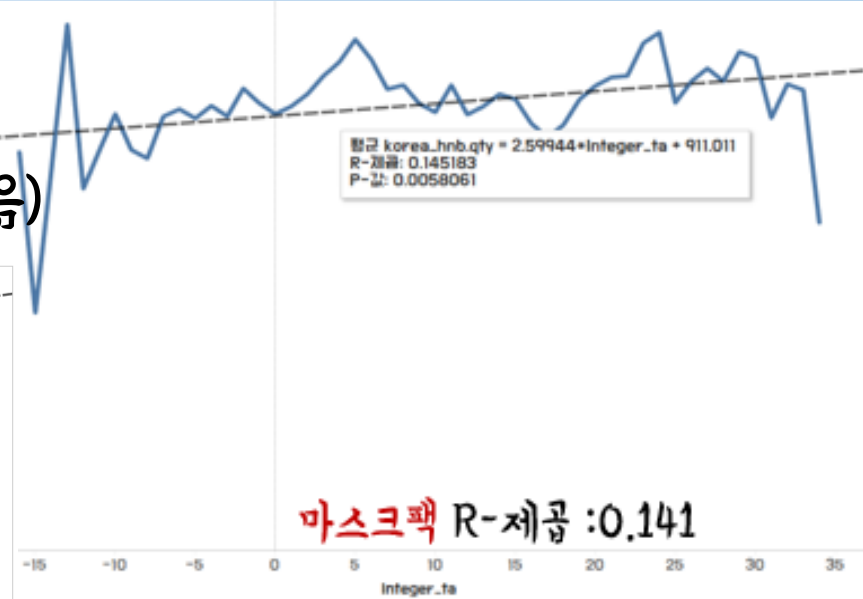
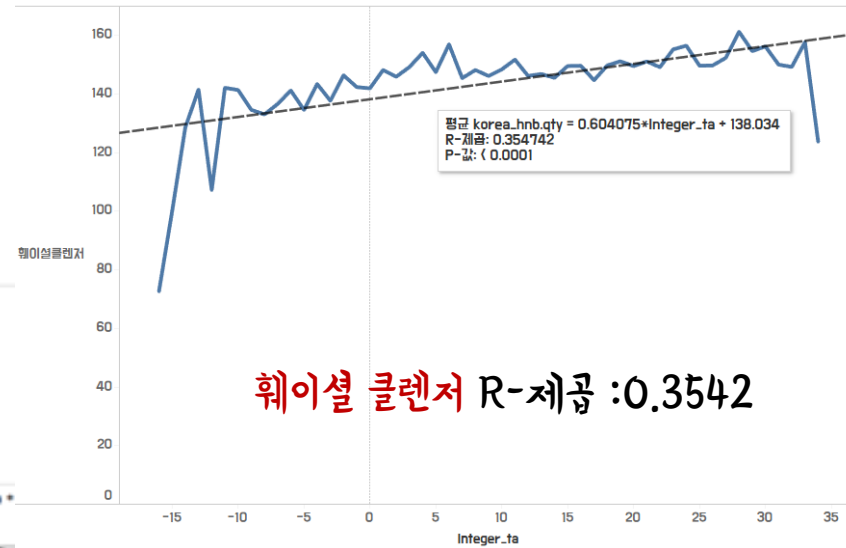
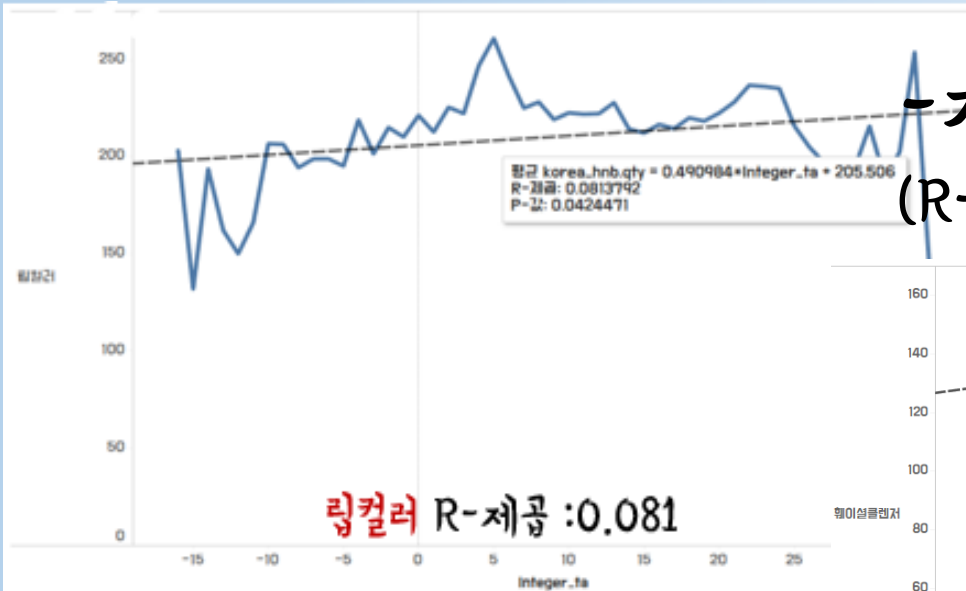




03 데이터 EDA 및 모델링

데이터 EDA

-기온둔감품목 추세 모델
(R-값이 작을 수록 기온과 관련이 없음)



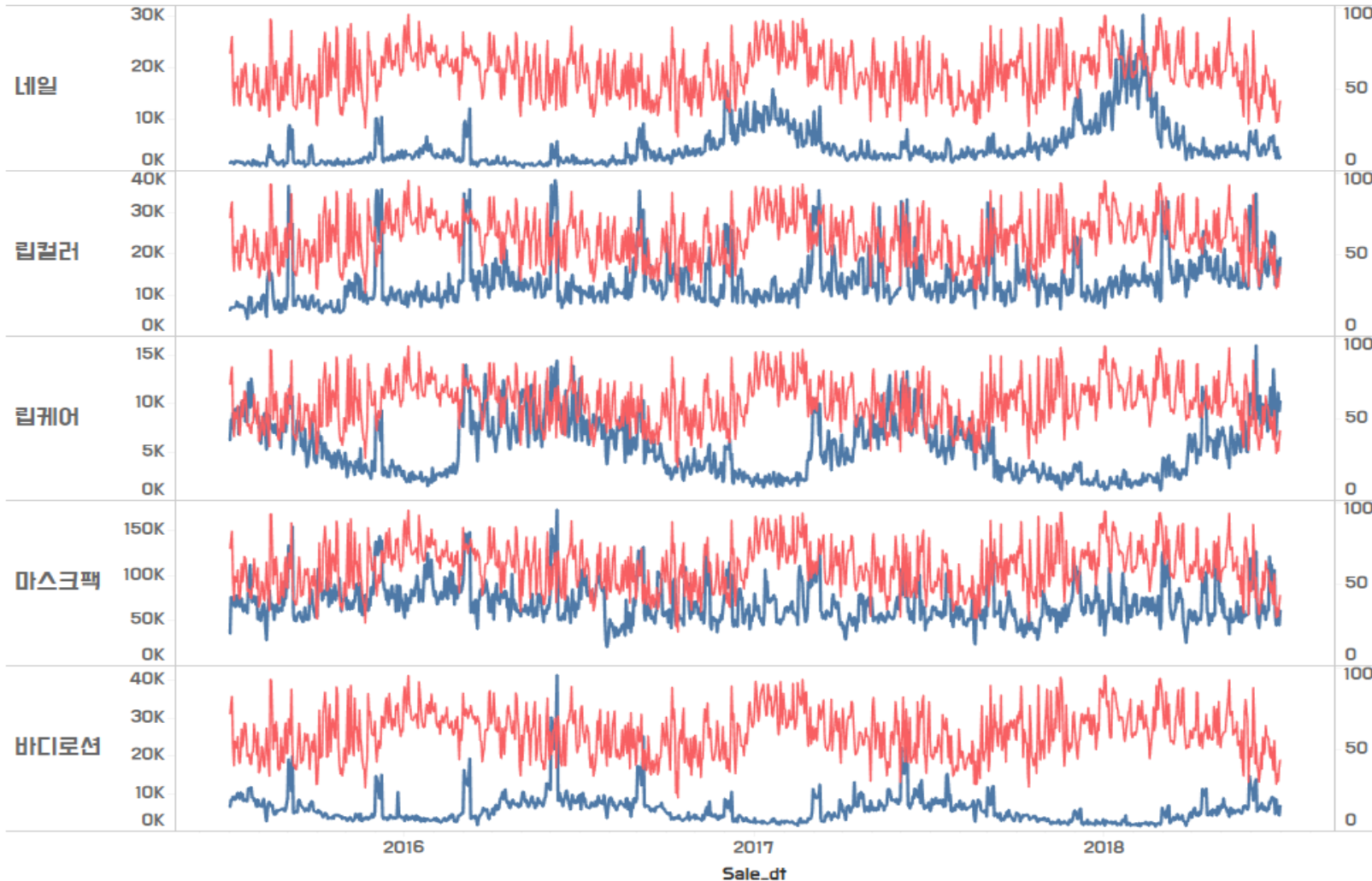


03 데이터 EDA 및 모델링

데이터 EDA

2) 상대습도별

날짜별 상대습도 & 판매량 - 상관관계가 존재하지 않는 카테고리



상대
습도

2016~2018 상대습도 그래프

판매량

2016~2018 판매량변화 그래프

네일, 립컬러, 립케어,
마스크팩, 바디로션

-> 상대습도의 변화와
관계가 없다.

-습도 둔감 품목

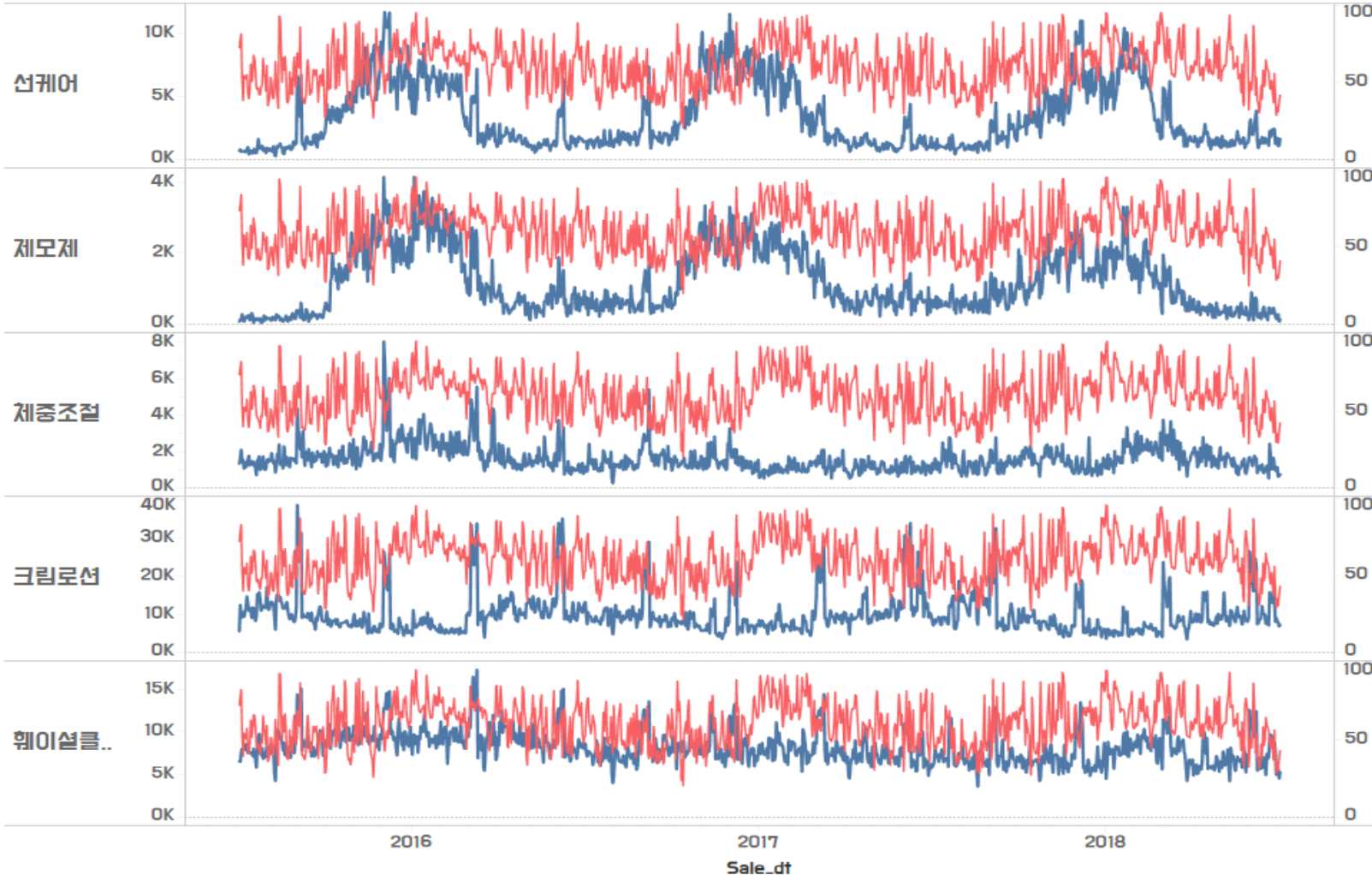


03 데이터 EDA 및 모델링

데이터 EDA

2) 상대습도별

날짜별 상대습도 & 판매량 - 상관관계가 존재하지 않는 카테고리



상대
습도

2016~2018 상대습도 그래프

판매량

2016~2018 판매량변화 그래프

선케어, 제모제
체중조절, 크림로션, 웨이셜클렌저

-> 상대습도의 변화와
관계가 없다.

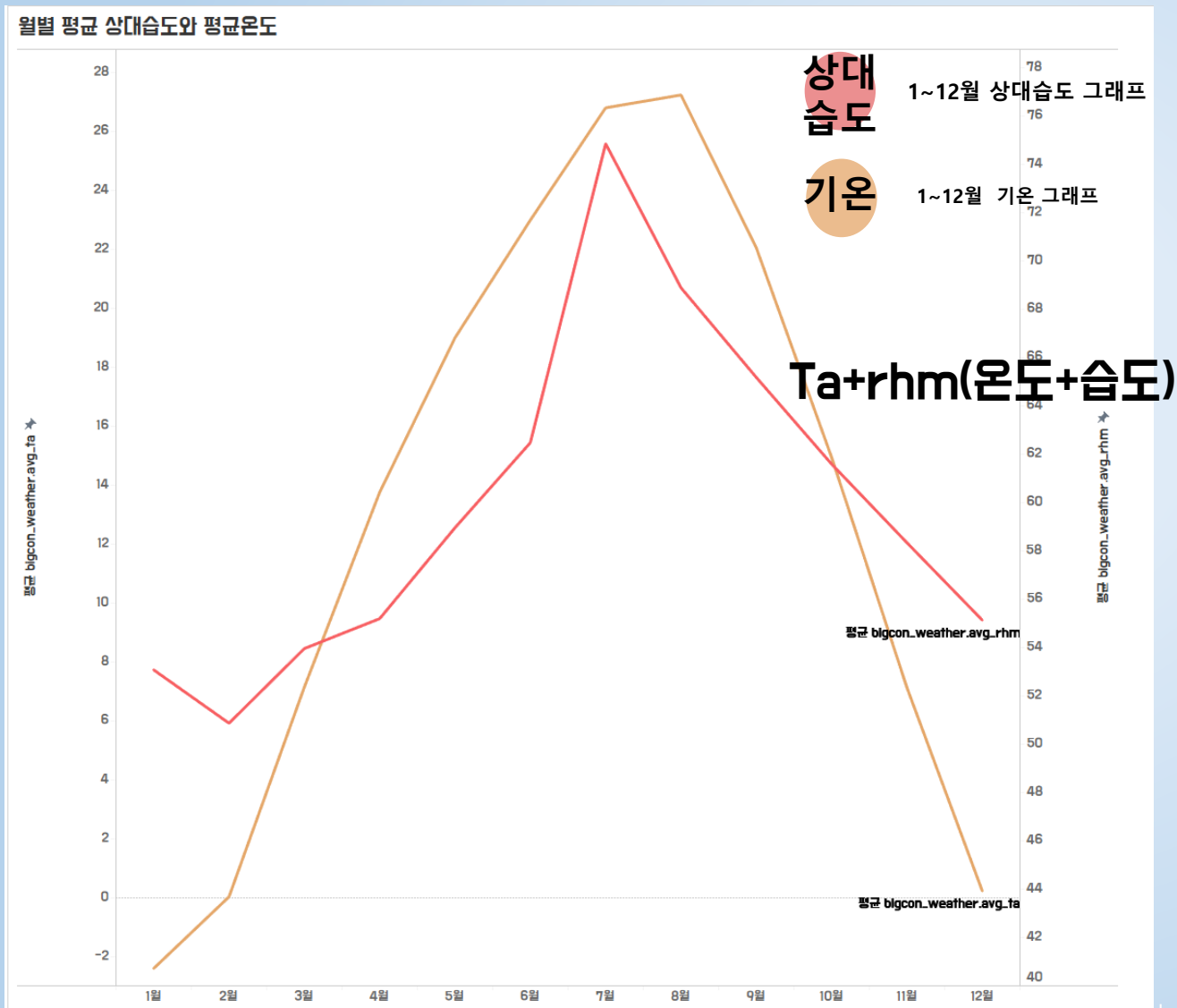
-습도 둔감 품목



03 데이터 EDA 및 모델링

데이터 EDA

2) 상대습도별



그런데,
기온 과 상대습도는
비슷한 추세를 보인다.

그렇다면 모든 품목은
상대습도에 둔감한 것일까?

습도와 기온을 모두 고려해보자.

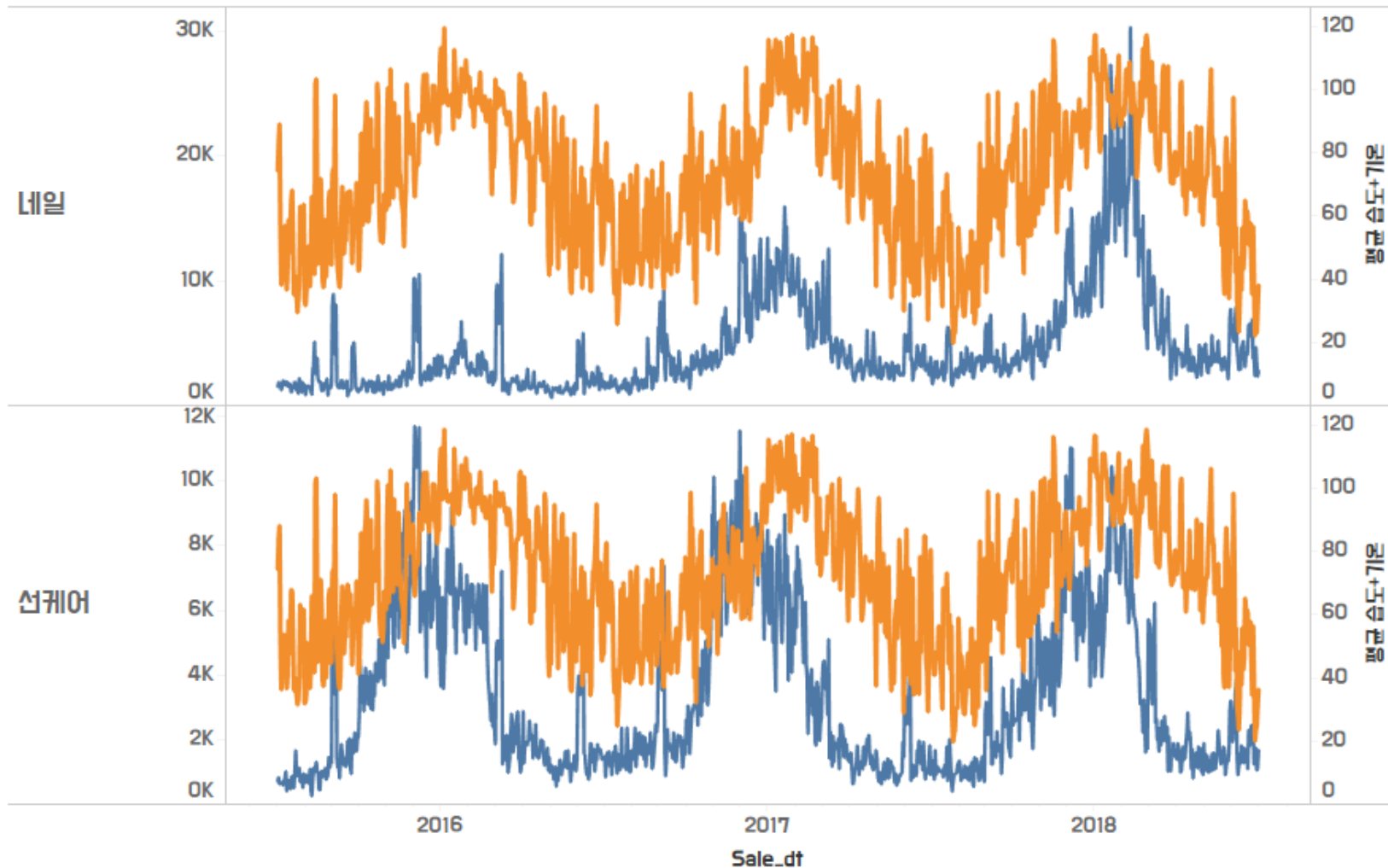


03 데이터 EDA 및 모델링

데이터 EDA

3) 상대습도+ 기온별

날짜별 상대습도+온도 & 판매량 - 상관관계가 존재하는 카테고리



습도
+
온도

2016~2018 상대습도+온도
변화 그래프

판매량

2016~2018 판매량변화 그래프

네일, 선케어

-> 상대습도+온도 의 변화와
양(+)의
상관관계를 갖고 있다.

-습도+온도민감품목

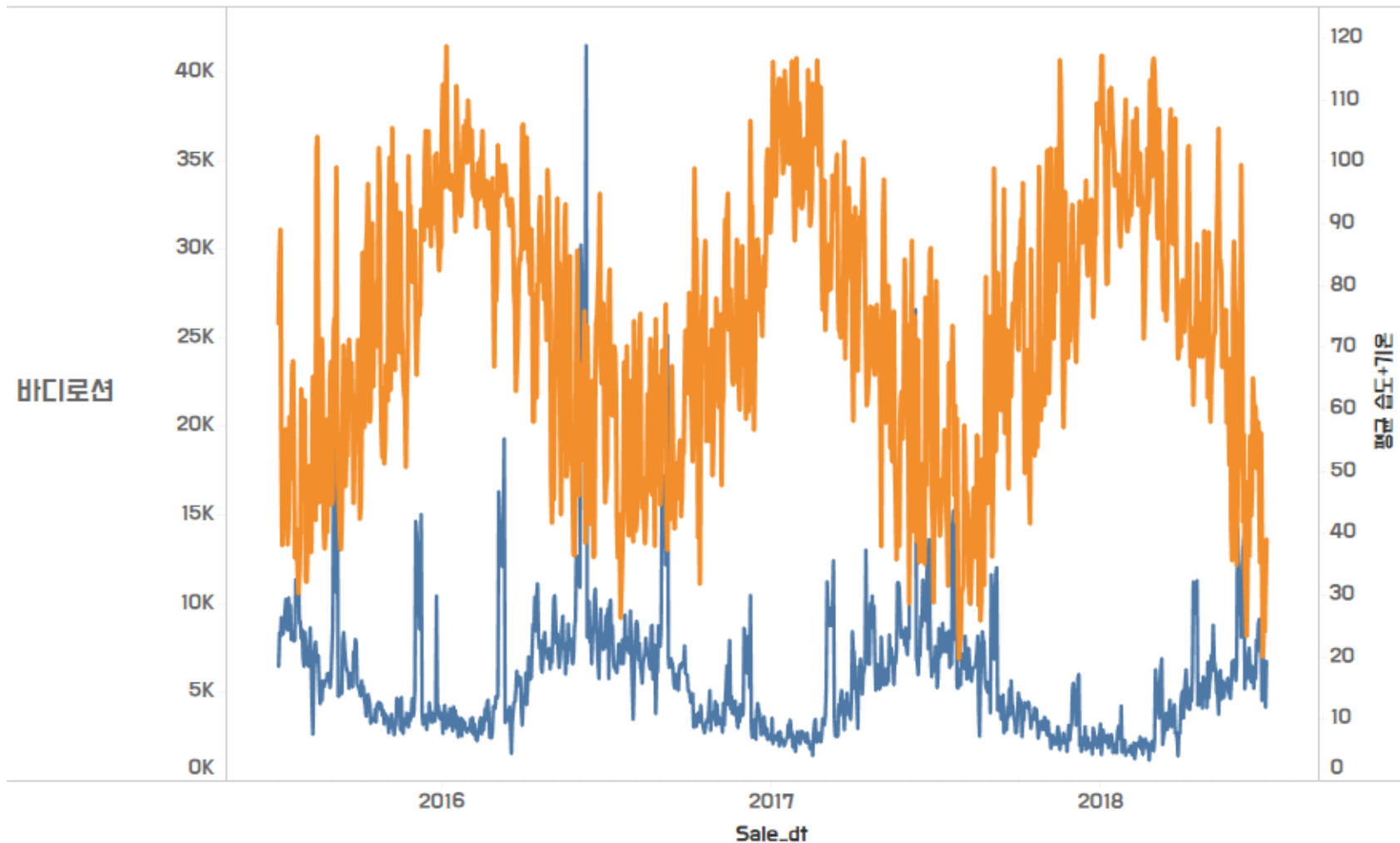


03 데이터 EDA 및 모델링

데이터 EDA

3) 상대습도+ 기온별

날짜별 상대습도+온도 & 판매량 - 상관관계가 존재하는 카테고리



습도
+
온도

2016~2018 상대습도+온도
변화 그래프

판매량

2016~2018 판매량변화 그래프

바디로션

-> 상대습도+온도 의 변화와
음(-)의
상관관계를 갖고 있다.

-습도+온도민감품목

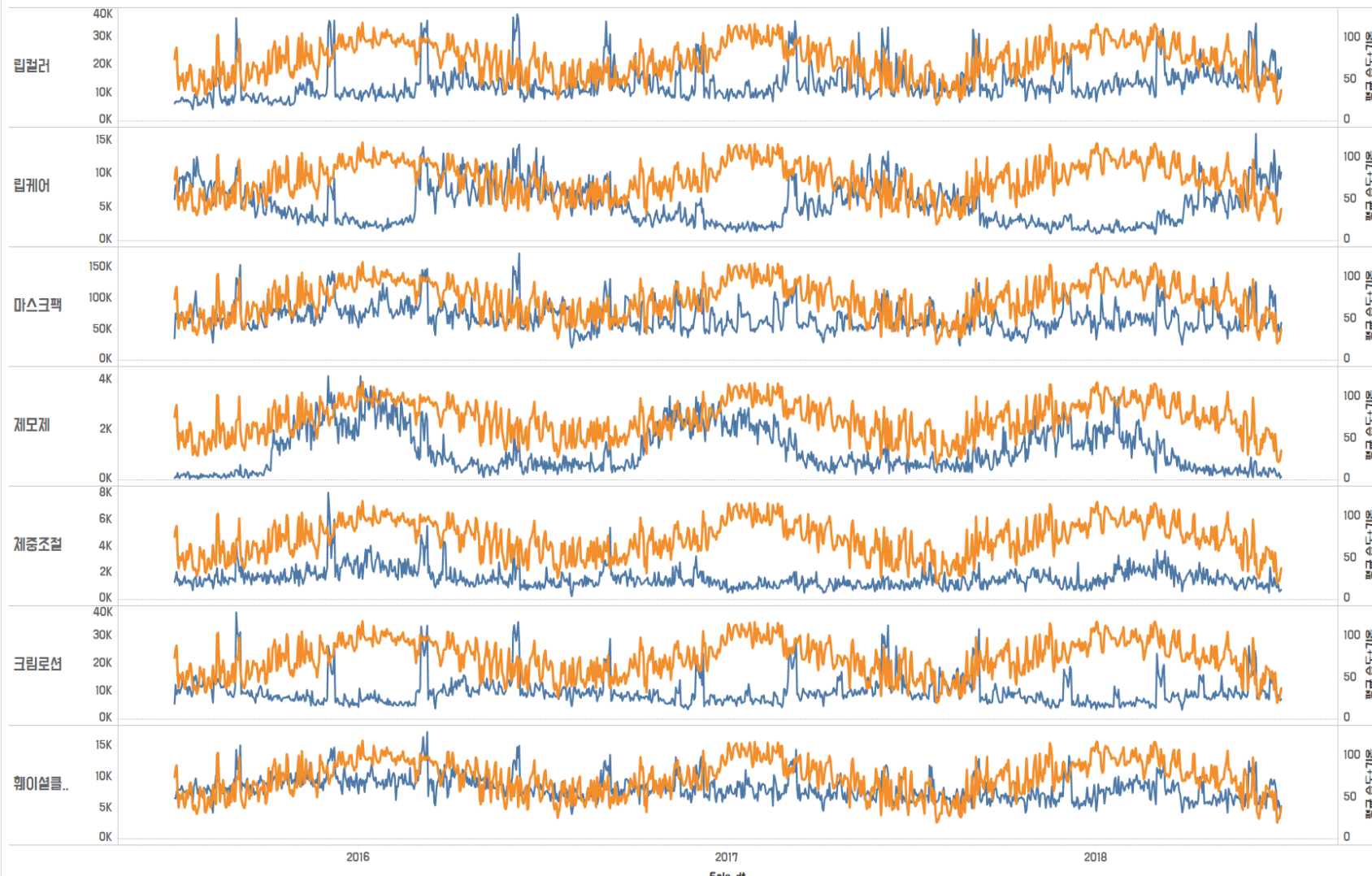


03 데이터 EDA 및 모델링

데이터 EDA

3) 상대습도+ 기온별

날짜별 상대습도+온도 & 판매량 - 상관관계가 존재하지 않는 카테고리



습도
+
온도

2016~2018 상대습도+온도
변화 그래프

판매량

2016~2018 판매량변화 그래프

립컬러, 립케어, 마스크팩
제모제, 체중조절, 크림로션,
웨이셜클렌저
-> 상대습도의 변화와
관계가 없다.

-습도+온도둔감품목

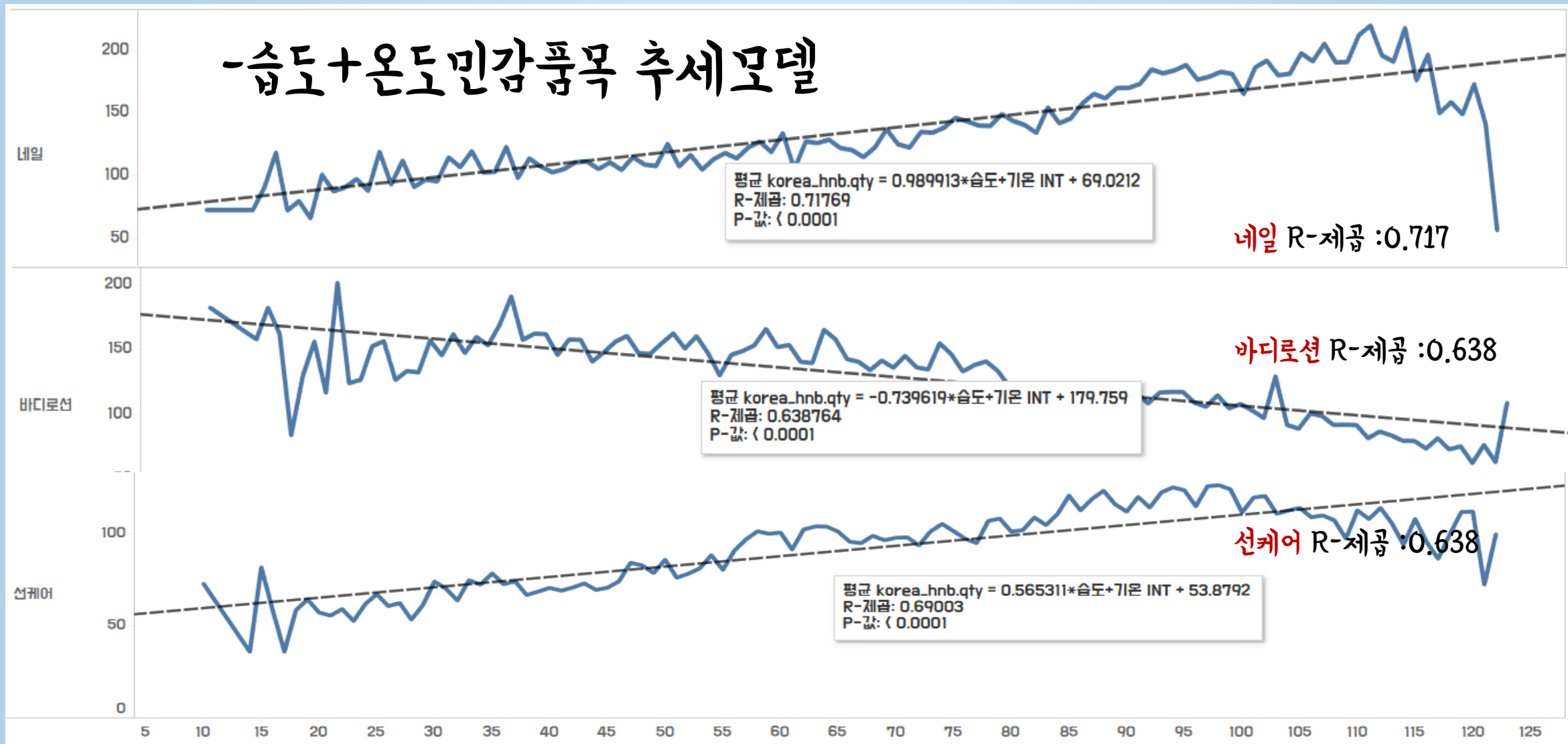


03 데이터 EDA 및 모델링

데이터 EDA

3) 상대습도+ 기온별

-습도+온도민감품목 추세모델





03 데이터 EDA 및 모델링

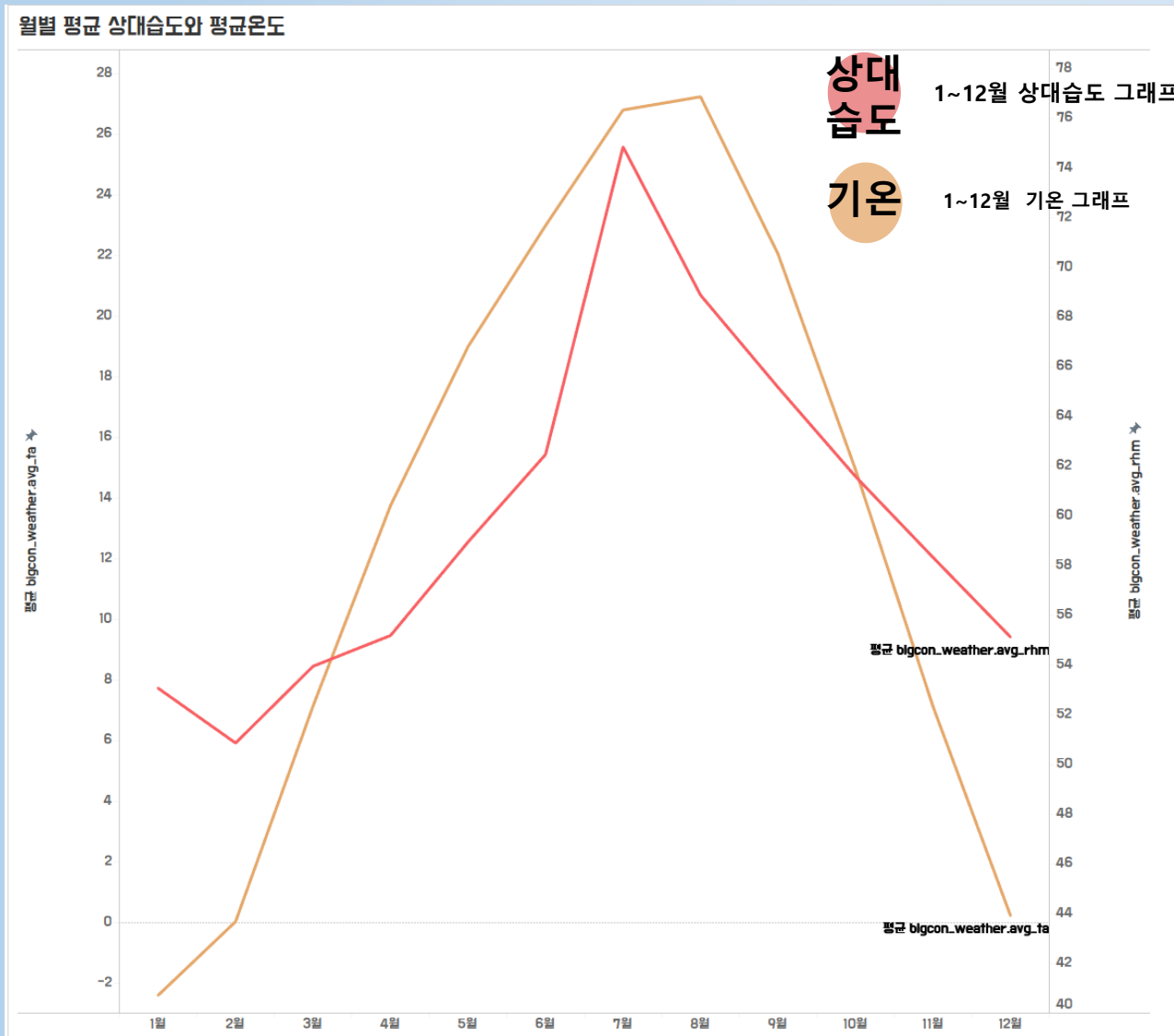
데이터 EDA



New 지표 생성



Ta+rhm(온도+습도)



1~3을 정리하자면

오직 '기온'만 관련 있는
상품이 존재하고

'상대습도' 단독으로는
관련도가 없으며,

'**습도+온도**' 복합지표는 관련 있는
상품이 존재한다.



03 데이터 EDA 및 모델링

기온

Avg_ta(평균 온도)
Min_ta(최저 온도)
Max_ta(최고 온도)
Avg_sufta(평균 지면온도)
Avg_dewp(평균 이슬점 온도)

강수량

Cont_rn(강수 지속 시간)
Sum_rn(강수량)

미세먼지

Avg_pm10(미세먼지)
Avg_pm2.5(초미세먼지)

주요대분류

습도

Avg_rhm(평균 상대습도)
Min_rhm(최저 상대습도)

일조량

Sum_sunt(일조 합계시간)
Max_sunh(최대 일조시간)
Sum_sun(일조량)

적설량

Sum_snow(적설량)
Sum_snow3h(3시간 적설량)

풍속

Maxi_ws(순간 최대 풍속)
Max_ws(최대 풍속)
Avg_ws(평균 풍속)

총 19가지 날씨변수
+1개의 복합변수
Ta+rhm(온도+습도)

X=>20가지 (feature)



Y => 10개의 카테고리별 판매량



SUNNY



CLOUDY



RAINY



모델링 > 머신러닝 기법을 사용하여 판매량 예측

- | | |
|----------------------|------------------|
| 1. Linear Regression | 4. Random Forest |
| 2. Ridge Regression | 5. XGBoost |
| 3. Lasso Regression | 6. LightGBM |



03 데이터 EDA 및 모델링

모델링 - 회귀모델

1. 네일

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.4037 | 1834.9772 |
| Ridge | 0.4039 | 1834.8035 |
| Lasso | 0.4055 | 1832.4267 |

2. 립컬러

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.0279 | 4942.8514 |
| Ridge | 0.0281 | 4942.4096 |
| Lasso | 0.0282 | 4942.1009 |

5. 바디로션

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.3524 | 2700.8345 |
| Ridge | 0.3526 | 2700.3865 |
| Lasso | 0.3510 | 2705.5520 |

6. 섀케어

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.6538 | 1721.0461 |
| Ridge | 0.6538 | 1720.9635 |
| Lasso | 0.6541 | 1720.2799 |

3. 립케어

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.4756 | 2141.1572 |
| Ridge | 0.4757 | 2141.0577 |
| Lasso | 0.4721 | 2148.3772 |

4. 마스크팩

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.0201 | 20727.742 |
| Ridge | 0.0201 | 20727.115 |
| Lasso | 0.0197 | 20731.500 |

7. 제모제

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.5764 | 531.0459 |
| Ridge | 0.5767 | 530.8845 |
| Lasso | 0.5795 | 529.0971 |

8. 체중조절

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.1035 | 619.0302 |
| Ridge | 0.1037 | 618.9535 |
| Lasso | 0.1091 | 617.0896 |

9. 크림로션

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.1359 | 4433.0318 |
| Ridge | 0.1362 | 4432.2491 |
| Lasso | 0.1360 | 4432.8113 |

10. 웨이셜클렌저

| | R-square value | RMSE value |
|--------|----------------|------------|
| Linear | 0.1029 | 1830.1034 |
| Ridge | 0.1029 | 1830.0617 |
| Lasso | 0.1040 | 1828.9178 |

10개의 카테고리 별 rmse value가 가장 낮은 선형회귀 선정



03 데이터 EDA 및 모델링

모델링 – Random Forest와 비교

1. 네일

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Lasso | 0.4055 | 1832.4267 |
| Random Forest | 0.4891 | 1698.6053 |

2. 립컬러

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Lasso | 0.0282 | 4942.1009 |
| Random Forest | -0.0041 | 5023.8114 |

5. 바디로션

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Ridge | 0.3526 | 2700.3865 |
| Random Forest | 0.3167 | 2774.2830 |

6. 선크어

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Lasso | 0.6541 | 1720.2799 |
| Random Forest | 0.7410 | 1488.5180 |

3. 립케어

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Ridge | 0.4757 | 2141.0577 |
| Random Forest | 0.5128 | 2063.9664 |

4. 마스크팩

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Ridge | 0.0201 | 20727.115 |
| Random Forest | 0.0610 | 20290.551 |

7. 제모제

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Lasso | 0.5795 | 529.0971 |
| Random Forest | 0.7049 | 443.2472 |

8. 체중조절

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Lasso | 0.1091 | 617.0896 |
| Random Forest | 0.1377 | 607.0774 |

9. 크림로션

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Ridge | 0.1362 | 4432.2491 |
| Random Forest | 0.0974 | 4530.7938 |

10. 웨이셜클렌저

| | R-square value | RMSE value |
|---------------|----------------|------------|
| Lasso | 0.1040 | 1828.9178 |
| Random Forest | 0.1594 | 1771.5264 |

카테고리별 가장 성능이 좋은 선형회귀와 random forest 성능 비교

대부분의 카테고리에서 **Random Forest** 모델이 가장 좋은 성능을 보임

한편, R-square value가 0.5 이상 나온 카테고리 : '립케어', '선크어', '제모제' => 추가 모델링

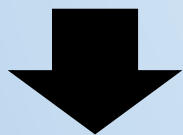


03 데이터 EDA 및 모델링

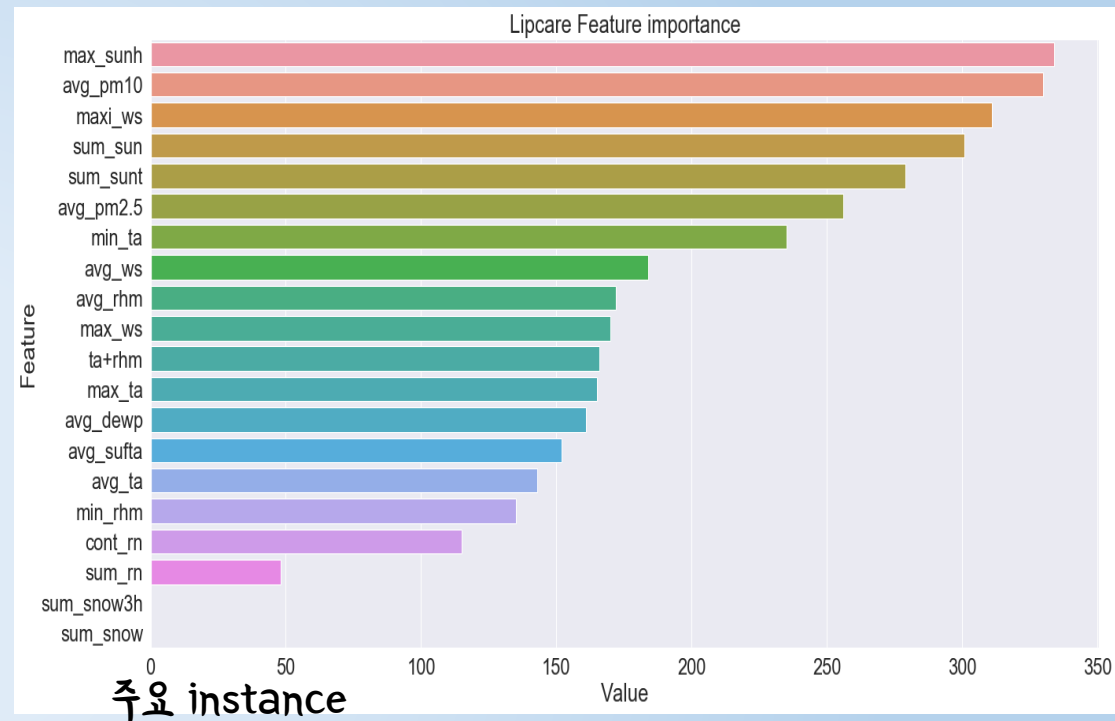
모델링 - Boost모델과 비교

3. 립케어

| | R-squre value | MAE | RMSE |
|------------------|------------------|---------|---------|
| Random Forest | 0.5144 | 1555.66 | 2063.96 |
| XGBoost | 0.5250 | 1506.02 | 2037.88 |
| LightGBM | 0.5400 | 1475.08 | 2004.23 |



Ensanble(0.3*XGB + 0.7*LGB) : 1996.65



- 최대 일조시간
- 평균 미세먼지
- 합계 일사량
- 합계 일조 시간
- 평균기온

일조량
기온
미세먼지

부스팅 모델과 이를 앙상블한 모델로 성능 향상, Feature importance 추출

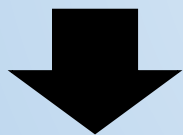


03 데이터 EDA 및 모델링

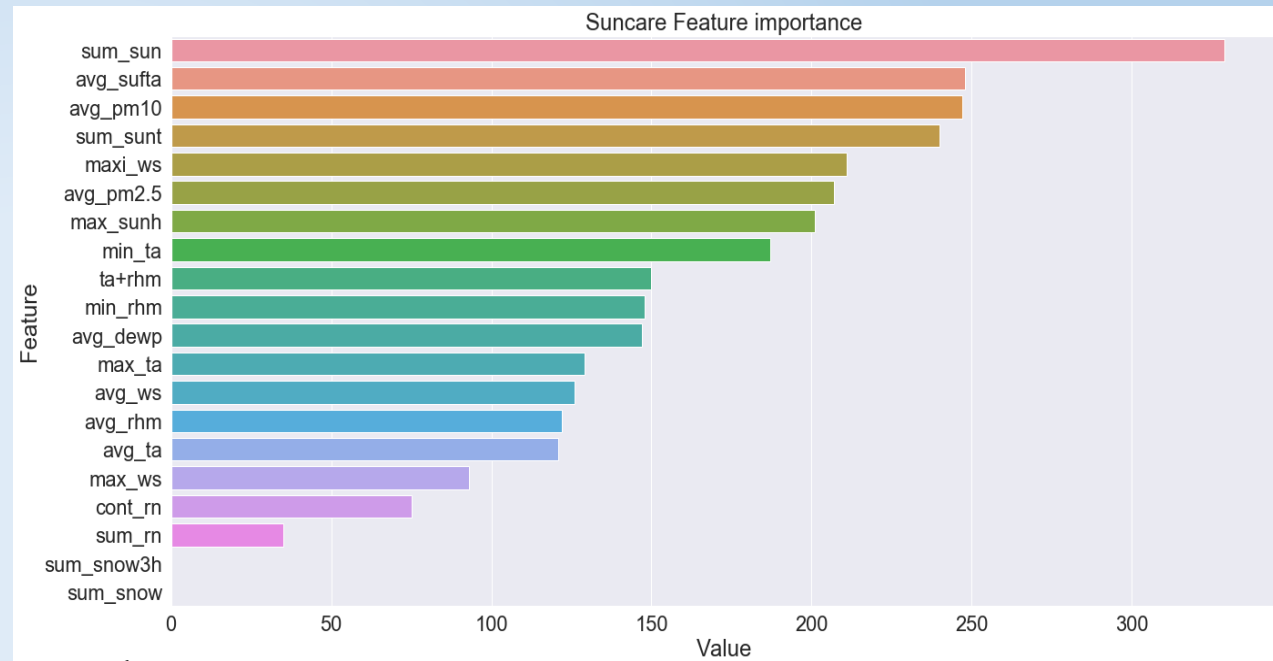
모델링 - Boost모델과 비교

6. 선택어

| | R-squre value | MAE | RMSE |
|------------------|------------------|--------|---------|
| Random Forest | 0.7441 | 987.39 | 1479.74 |
| XGBoost | 0.7483 | 960.94 | 1467.42 |
| LightGBM | 0.7607 | 961.77 | 1431.12 |



Ensanble(0.2*XGB + 0.8*LGB) : 1427.29



주요 instance

- 합계 일사량
- 평균 지면온도
- 평균 미세먼지
- 합계 일조 시간

일조량
기온
미세먼지

부스팅 모델과 이를 앙상블한 모델로 성능 향상, Feature importance 추출

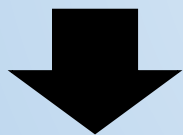


03 데이터 EDA 및 모델링

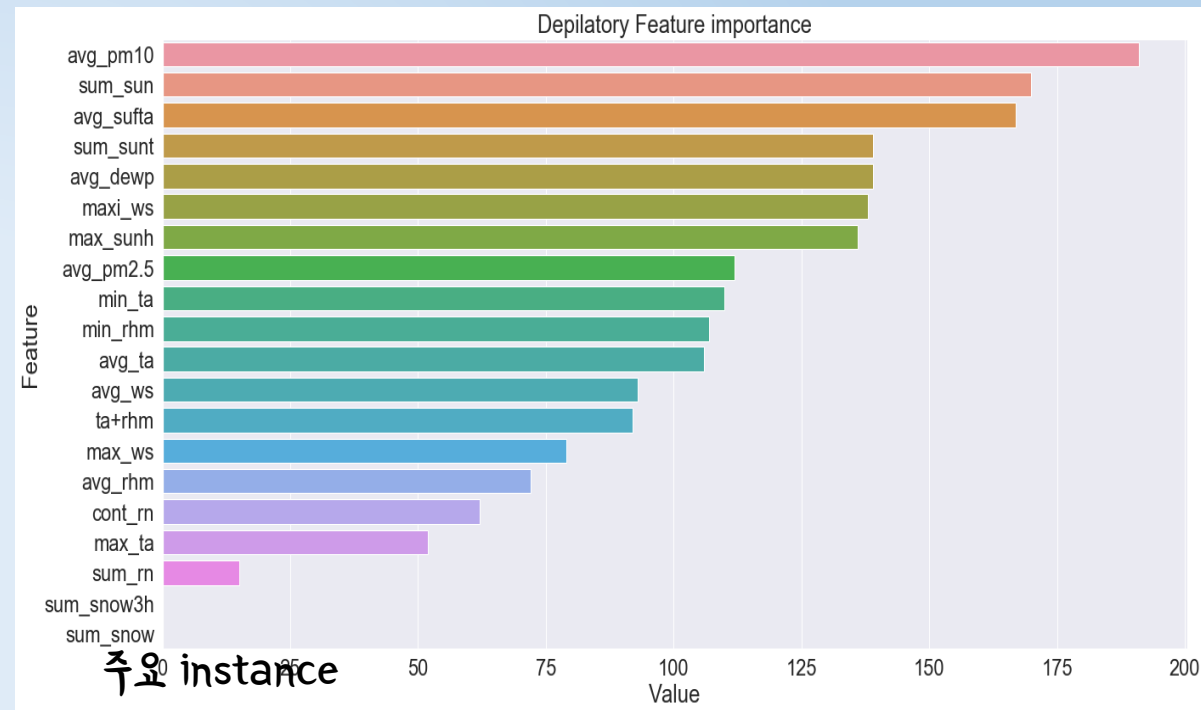
모델링 - Boost모델과 비교

7. 제모제

| | R-squre value | MAE | RMSE |
|------------------|------------------|--------|--------|
| Random Forest | 0.7104 | 340.01 | 439.08 |
| XGBoost | 0.7037 | 339.48 | 444.18 |
| LightGBM | 0.7268 | 328.09 | 426.47 |



Ensanble(0.1*XGB + 0.9*LGB) : 426.05



- 평균 미세먼지
- 합계 일사량
- 평균 지면온도
- 합계 일조시간
- 평균 이슬점 온도

일조량
기온
미세먼지

부스팅 모델과 이를 앙상블한 모델로 성능 향상, Feature importance 추출



4. 결과



'기온민감품목'

'습도+온도 민감품목'

제모제,
체중조절
립케어

네일, 선크어

바디로션

오직 기온과 관련이 있다.

기온, 습도 모두 관련 있지만
습도보다는 기온에 민감하다.

습도와 온도 두 지표를
동시에 보아야 한다.
즉, 단순히 온도만 내려
간다고 판매량이 증가
하는 것이 아니라,
온도와 습도가 동시에
내려야 유의미하게
판매량이 증가한다.



그렇다면 이 모든 상품을 '민감 상품' 이라고 할 수 있을까?

NO! 모델링을 통해 R-square value가 높았던 카테고리

'립케어', '선케어', '제모제'

품목에 집중하여 판매를 예측해야한다.

립케어) '최대 일조시간', '평균 온도'와, '평균 미세먼지'와 가장 크게 (-)음의 상관관계를 가짐

선케어) '합계 일사량', '평균 미세먼지', '평균 지면온도' 과 가장 크게 (+) 양의 상관관계를 가짐

재모제) '합계 일사량', '평균 미세먼지', '평균지면온도' 와 가장 크게 (+) 양의 상관관계를 가짐



5. 활용방안



05 활용방안 판매예측



기온



먼지



습도

월별 우세품목 약세품목 예측

예)

5월: 대한민국 기온 상승

->기온과 양(+)의 상관관계를 가진 선풍기

= 우세품목

->음(-)의 상관관계를 가진 립케어 = 약세품목

10월 : 대한민국 상대습도와 기온이 내려가는 월

-> 상대습도+기온과 음의 상관관계를 가진

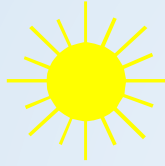
바디로션 = 약세품목

효율적 재고관리를
통한 실적개선





05 활용방안 판매예측



기온



먼지



습도

품목별 중요 변수를 통해
예측의 '가중치'를 부여하여 세분화

판매 예측의
'확률'제공

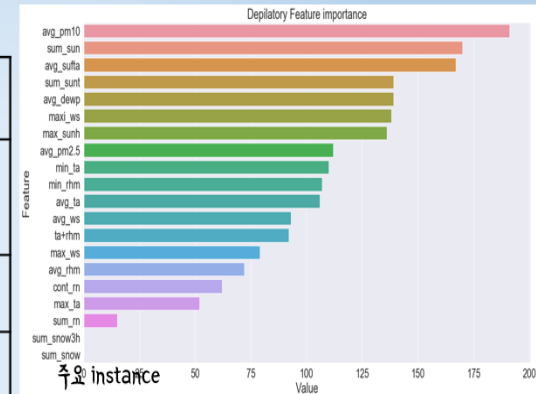
예) 제모제는 기온과 가장 밀접하다.
그런데 기온을 나타내는 지표는 다양한데,
이중, '평균 지면온도'와
'평균 이슬점 온도에 가장 영향을
많이 받으므로

해당 지표가 가장 크게 변화하는 날에
판매량이 변화 할 확률이 높을 것이라고
가중치를 부여한다.

7. 제모제

| | RMSE value |
|---------------|------------|
| Random Forest | 443.2472 |
| XGBoost | 444.18 |
| LightGBM | 426.47 |

Ensanble(0.1*XGB + 0.9*LGB) : 426.05



- 평균 미세먼지
- 하계 일사량
- 평균 지면온도
- 하계 일조시간
- 평균 이슬점 온도

일조량
기온
미세먼지



날씨화장품 알리기



**칼라블라 고객에게
예측되는 날씨와 해당 날씨와 관련된 품목을
함께 추천하는 문자를 발송하여 홍보**

**예) 10월 첫째주 대한민국 가~장 상대습도
낮아지는 건조할 때! 다른 고객들은 립케어로
입술 보호중!**

**예2) 11월 첫째주! 날로 갈수록 추워지는 온
도에 가뭄으로 인한 건조함까지! 다른 고객들
은 바디로션으로 이 날씨를 이겨냈다! 지금
사야 할 때!**



6. 7기대효과



1> 실적개선

세분화된 판매예측으로
인한, 재고관리

-> 실적 개선

2> 타 업체와의 차별화

타 업체보다
고객에게 세분화된 날씨와 관
련된 품목 추천 홍보 가능

-> 차별화

3> 브랜드 이미지, 건강한 아름다움

칼라블라이모토,
건강한 아름다움
'날씨'에 대비하여 자신의 피부건강
을 위해 화장품을 준비하게 하는 홍보

-> 브랜드이미지 확충