

# 날씨 빅데이터 콘테스트

날씨 데이터 분석을 통한  
랄라블라 재고관리와 유통망 개선

Weather Eyes Team

Registration Number  
202747

lala♥la

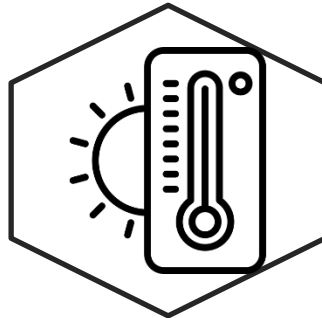
랄라블라  
beaut



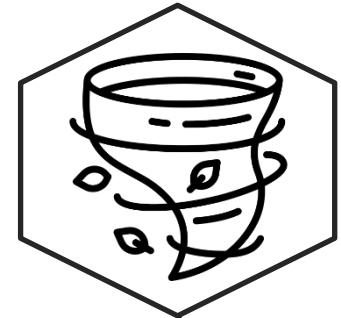
# CONTENTS



1. 공모 배경



2. 활용 데이터 정의



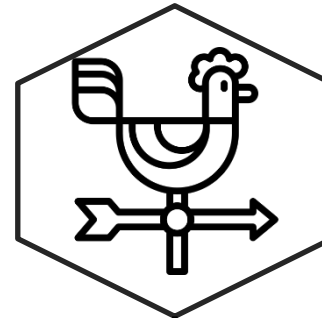
3. 분석 처리방안  
및  
활용 분석기법



4. 분석결과



5. 서비스 활용 방안



6. 서비스 기대효과



## 1.공모 배경

- (1) 문제 인식
- (2) 주제 선정

# 01. 공모 배경 (1) 문제 인식

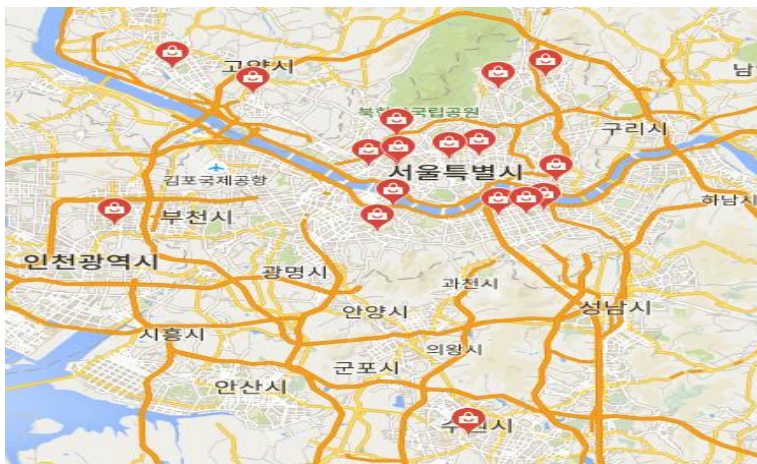
H&B(healthy & Beauty)업계 2위 랄라블라,  
영업적자 **Why?**

## 1. 서울과 수도권에 편중된 매장 수

✓ 유동인구가 중요한 H&B특성상 불가피

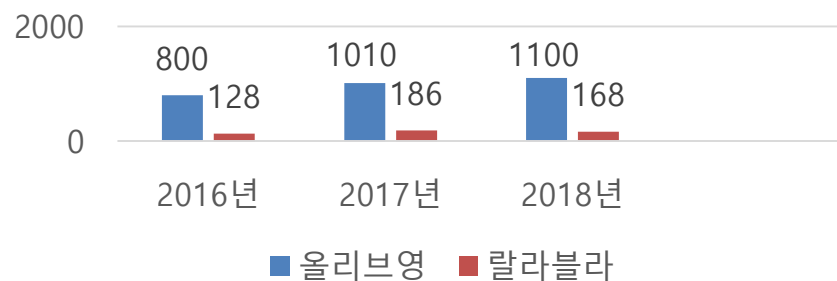
## 2. 업계 1위인 올리브 영과 큰 폭의 매장 수 차

✓ 올리브 영 매장 수 1100여개, 랄라블라와 10배



연도별 올리브 영 & 랄라블라 매장 수

(단위: 개)



# 01. 공모 배경 (2) 주제 선정

Watsons와 결별 후 추락하는 랄라블라



적자 재무구조에 벗어나기 위한 **공급망 개선**의 시급성 인식

- 전국으로 사업 영토를 확장 할 해결 방안이 필요함
  - 영업 적자와 매장 수를 줄이는 상황에서 신 매장 오픈은 불가능
- 경쟁사 올리브 영도 신 매장 부지 매입의 어려움으로  
점포 수 증가 폭이 주춤한 상황

2018년 점포 수 (단위: 개)

올리브 영	GS25	랄라블라
1,100	13,107	168

## 01. 공모 배경 (2) 주제 선정

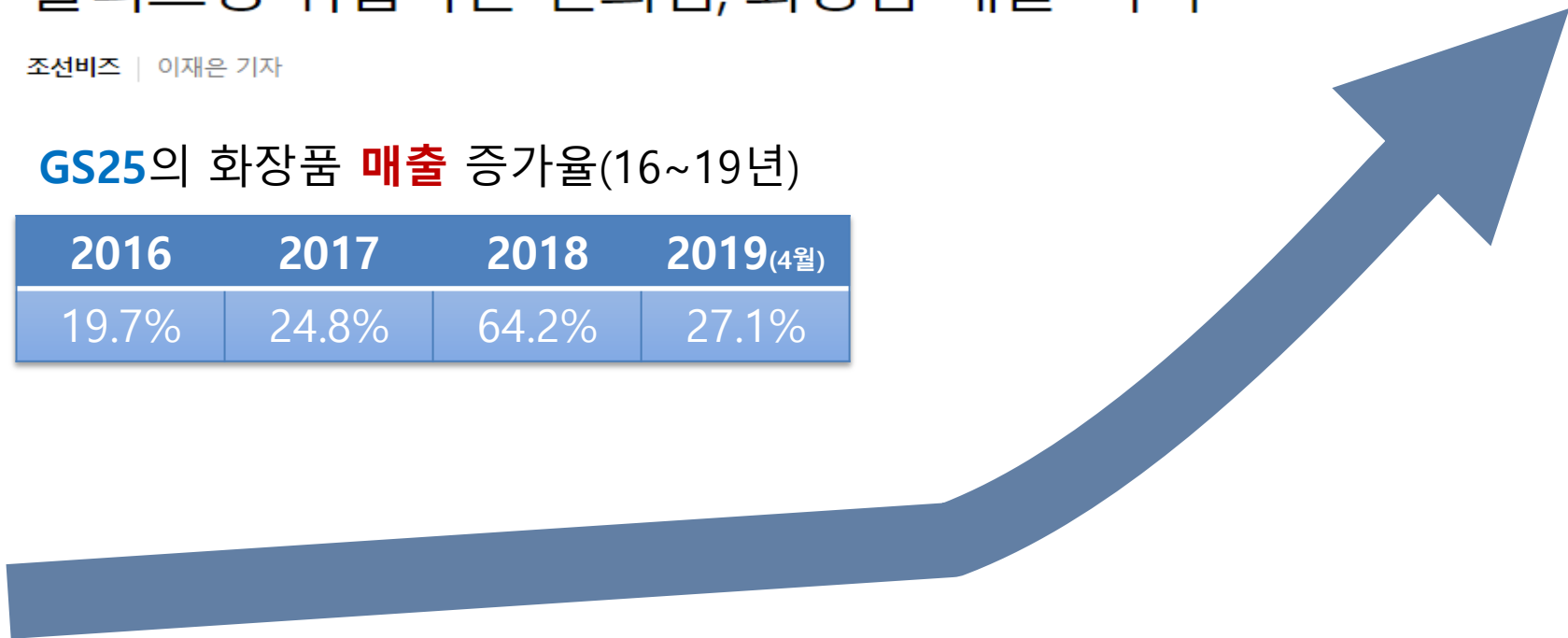
# GS25편의점의 화장품 매출이 최근 증가추세 올리브 영 위협할 수준

올리브영 위협하는 편의점, 화장품 매출 '쑥쑥'

조선비즈 | 이재은 기자

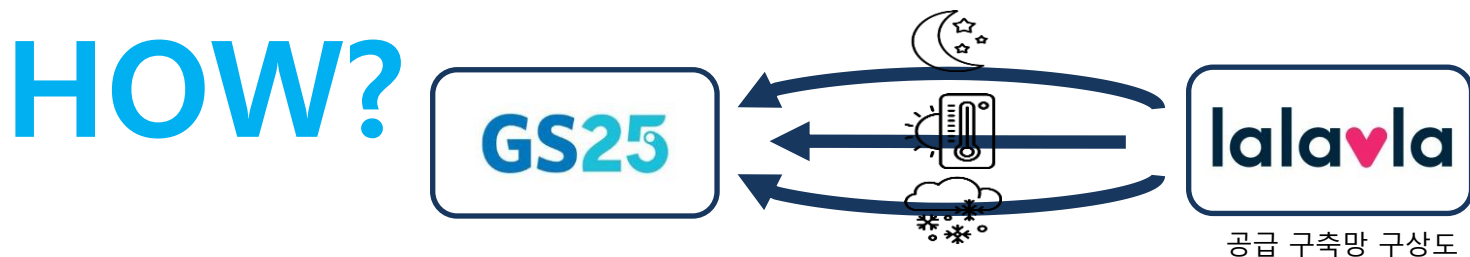
GS25의 화장품 **매출** 증가율(16~19년)

2016	2017	2018	2019 <sub>(4월)</sub>
19.7%	24.8%	64.2%	27.1%



## 01. 공모 배경 (2) 주제 선정

# 랄라블라의 상품을 GS25편의점에 진열하는 유통망 구축



## 화장품 수요 특성상 계절에 민감하게 반응

✓ 날씨 데이터와 화장품 수요의 연관성을 파악

→ GS25편의점 상품과 연계해 매장에 진열

→ 유통망 개선과 악성재고관리

→ 전국적 유통망 확보, 적자폭 감소

# 01. 공모 배경 (2) 주제 선정

## 연구가설 및 주제 요약

연구가설: 온도, 습도, 강수량, 일조시간 등  
날씨 변수가 화장품 판매량에 영향을 미칠 것이다

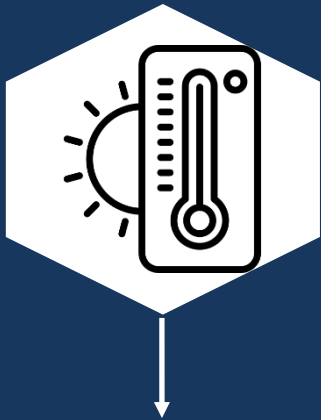
따라서 날씨 변화에 따른 GS25 판매량과 랄라블라 판매량의 상관관계를 분석하고  
랄라블라의 유통망 개선과 재고 관리의 Solution을 제안

- 👉 소주제 1: 날씨 변수들과 판매량의 변화율의 연관성을 회귀분석을 통해 파악,  
최적의 랄라블라 화장품을 GS25 매장에 배치 → **영업이익 증가**
- 👉 소주제 2: 날씨와 연관성이 높은 제품을 중심으로 서울시  
구별, 성별, 연령별 고객 Needs파악 → **제품관리의 효율성 증가**
- 👉 소주제 3: 연도별 판매량의 구매성향과 평균 판매량을 통해  
익년 적정 재고량 예측 → **악성 재고 관리**



# 데이터 처리 프로세스

---



## 2. 활용 데이터 정의

- (1) 날씨 데이터
- (2) 칼라블라 / GS25  
판매량 데이터
- (3) 소셜 데이터



## 3. 분석 처리방안 및 활용 분석기법



## 4. 분석 결과 및 분석의한계

## 02. 활용 데이터 정의 데이터 수집

### 날씨 데이터



- ✓ 기상자료개방포털에서 제공하는 날씨 데이터
- ✓ 관측지: 서울(108)
- ✓ 기간: 2016.01.01~2018.12.31

### 판매량 데이터



- ✓ 칼라블라에서 제공하는 판매량 데이터
- ✓ 관측지: 서울(108)
- ✓ 기간: 2016.01.01~2018.12.31



- ✓ 에어코리아에서 제공하는 미세먼지 데이터
- ✓ 관측지: 서울(108)
- ✓ 기간: 2016.03.01~2018.12.31



- ✓ GS25에서 제공하는 판매량 데이터
- ✓ 관측지: 서울(108)
- ✓ 기간: 2016.01.01~2018.12.31

## 02. 활용 데이터 정의 (1) 날씨 데이터

### 날씨 데이터 처리



Column 명	설명
Loc	지점번호
Date	날짜(yyyy-MM-dd)
Temp	일 평균온도
Rain	강수량(mm)
Wind	풍속(m/s)
Humid	평균 습도(%)
hpa	hpa
Sun_time	가조시간(hr)
Lgt_time	합계일조시간(hr)
Sun_energy	합계 일사(MJ/m2)
Snow	일 최심신적설(cm)
cloud	평균 전운량(1/10)

- ✓ year, month, day, weekday, 을 만들어 각 년도/월/일/요일 칼럼 생성
- ✓ 해당연도에 몇 번째 주 칼럼 생성 (nth\_week)
- ✓ '합계일사'에 해당하는 sun\_energy는 결측치가 제거
- ✓ 강수량/적설량의 결측치는 비/눈이 오지 않은 날이라 판단, 0으로 대체
- ✓ 이외의 값: 해당 월의 평균치로 대체

-> 실제 판매량에 영향을 미칠 것으로 예상 되는 변수를 10개의 독립변수로 추출

## 02. 활용 데이터 정의 (1) 날씨 데이터

### 대기오염 물질 데이터 전처리 및 추가



Column 명	설명
지역	도/시
측정소코드	측정소 번호
측정소명	측정소 위치 등
측정일시	날짜(yyyy-mm-dd)
SO2	ppm
CO	ppm
O3	ppm
NO2	ppm
PM10	( $\mu\text{g}/\text{m}^3$ )
PM25	( $\mu\text{g}/\text{m}^3$ )
주소	시/구

- ✓ 미세먼지 데이터가 화장품 판매량에 영향을 미칠 것으로 예상하여 데이터 추가
- ✓ 전처리:
  - 2016년~2018년도까지 분기단위로 나눈 데이터
  - > 연 단위로 합침
  - > 일 단위 평균치로 종합
  - > 결측치는 (연)평균으로 수정
- > 날씨데이터에 추가  
(기후정보 + 대기오염물질 정보)

\*\* 미세먼지 구성원소: SO2, CO, O3, NO2, PM10, PM25

## 02. 활용 데이터 정의 (2) 라라블라/ GS25 판매량 데이터

### 판매량 데이터 처리

	pvn_nm	bor_nm	sale_dt	gender	age_cd	category	qty
0	서울특별시	종로구	20160101	F	20~39	립컬러	73
1	서울특별시	종로구	20160101	F	20~39	립케어	219
2	서울특별시	종로구	20160101	F	20~39	마스크팩	657
3	서울특별시	종로구	20160101	F	20~39	바디로션	328
4	서울특별시	종로구	20160101	F	20~39	체중조절	109
5	서울특별시	종로구	20160101	F	20~39	크림로션	73
6	서울특별시	종로구	20160101	F	20~39	헤어샴푸렌저	182
7	서울특별시	종로구	20160101	F	60~99	마스크팩	146

	pvn_nm	bor_nm	sale_dt	gender	age_cd	category	qty
0	서울특별시	종로구	20160101	F	00~19	라면	7
1	서울특별시	종로구	20160101	F	20~39	과자	655
2	서울특별시	종로구	20160101	F	20~39	라면	282
3	서울특별시	종로구	20160101	F	20~39	마스크	35
4	서울특별시	종로구	20160101	F	20~39	맥주	183
5	서울특별시	종로구	20160101	F	20~39	생리대	78
6	서울특별시	종로구	20160101	F	20~39	생수	56
7	서울특별시	종로구	20160101	F	20~39	숙취해소제	14
8	서울특별시	종로구	20160101	F	20~39	스타킹	28
9	서울특별시	종로구	20160101	F	20~39	아이스크림	211

lala<sup>♥</sup>la GS25

- ✓ 일 별 판매량(sale\_dt)을 월 별, 연도 별로 그룹화하는 데이터 처리 요구  
-> year, month, day, weekday, 을 만들어 각 연도/월/일/요일/칼럼생성
- ✓ 해당연도에 몇 번째 주를 알려 주는 칼럼 생성(nth\_week)
- ✓ 각 제품의 판매량 합계를 종속변수로 규정
- ✓ 'korea\_hnb.pvn\_nm' 같은 칼럼 명들에서 'korea\_hnb.' 부분 제거 후 칼럼 정렬

## 02. 활용 데이터 정의 (3) 변수 제거 방법론

### 다중 공선성 (Multicollinearity)

- ✓ 독립변수들간에 강한 상관관계가 나타나는 문제

#### 해결책

☞ VIF(Variance Inflation Fator) 분산확대지수

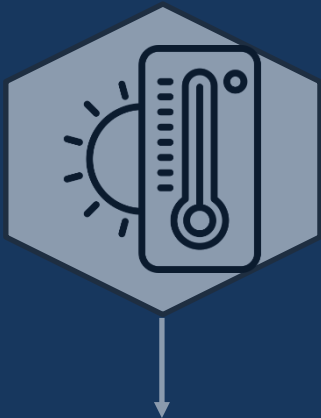
- ✓ 다중 공선성을 파악하기 위한 수치적 지표
- ✓ 0~ $\infty$ 까지 값을 가짐
- ✓ 수치가 10 이하인 변수들만 선택  
(snow, rain, PM10, wind, temp, cloud, lgt\_time  
변수를 최종적으로 채택)

CO, NO2, PM25, O3, humid, hpa, sun\_time,  
sun\_energe는 다중 공선성문제로 변수에서 제외

	VIF_Factor	features
0	1.239620	snow_or_not
1	2.479988	rain_or_not
2	3.114369	공기상태
3	4.013218	temp
4	6.725926	cloud
5	8.084050	lgt_time
6	9.334878	wind

# 데이터 처리 프로세스

---



## 2. 활용 데이터 정의

- (1) 날씨 데이터
- (2) 랄라블라  
및  
GS25 판매량 데이터
- (3) 소셜 데이터



## 3. 분석 처리방안 및 활용 분석기법



## 4. 분석 결과 및 분석의 한계

## 03. 분석 처리방안 및 활용 분석기법

### 데이터 처리 방안(서울중심으로 프로젝트시행)

- ✓ 전처리 된 데이터(결측치, 칼럼명, 데이터 통합 과정)  
(종속) 랄라블라 ,GS25 판매량 & (독립) 날씨변수

#### GS25

1. 과자
2. 라면
3. 마스크
4. 맥주
5. 면도기
6. 생리대
7. 생수
8. 숙취해소제
9. 스타킹
10. 아이스크림
11. 우산
12. 탄산음료

#### 날씨

1. snow
2. rain
3. cloud
4. temp
5. lgt\_time
6. wind
7. PM10

#### 랄라블라

1. 네일
2. 립컬러
3. 립케어
4. 마스크팩
5. 바디로션
6. 선케어
7. 제모제
8. 체중조절
9. 크림로션
10. 웨이셜 클렌저



## 03. 분석 처리방안 및 활용 분석기법

### 데이터 활용 분석기법 선정

#### 주요 분석기법 – 지도 학습 중 회귀분석 채택

##### 회귀분석 선정이유

날씨 변수 & GS25판매량 데이터, 날씨변수 & 알라블라 판매량 데이터의 연속적 선형 / 비선형의 함수를 도출하는데 유의한 방법론

##### 회귀분석?

훈련 데이터는 독립변수에 대한 속성을 벡터 형태로 포함하고 있으며 각각의 벡터에 대해 원하는 종속변수가 무엇인지 표시되어 있음, 연속적인 값을 출력하는 방법 즉, 독립변수에 대한 종속변수 값이 결정된 연속적 함수를 추출하는 방법

##### 지도학습?

훈련데이터(Training Data)로 부터 하나의 함수를 유추하는 기계학습 방법

## 03. 분석 처리방안 및 활용 분석기법

### 채택한 회귀 분석 방법론

#### 1. OLS (Ordinary least squares)

가장 일반적 회귀 분석법, 오차의 제곱의 합을 최소화,

예상하는  $\beta$ 값을 추정하는 최소 제곱법 중 하나, 오차  $e_i^2$ 의 합을 Minimize하는 방법론

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\sum \mathbf{x}_i \mathbf{x}_i^T)^{-1} (\sum \mathbf{x}_i y_i). \quad \sum e_i^2 = \sum (Y_i - \alpha - \beta X_i)^2$$

#### 2. Ridge Regression

최소제곱법과 매우 유사하나  $\lambda$ 값(Tuning parameter)을 조절해 최소 제곱 추정치를 구한다. RSS(Ridge regression)를 작게 만드는 계수추정치( $\beta$ )를 찾는 회귀분석법, Shrinkage penalty라고 불리는 두번째 항을 통하여 계수들이  $(\beta_1, \dots, \beta_p)$  0으로 가는 shrinkage 효과도 볼 수 있다.

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2,$$

#### 3. Lasso Regression

Ridge Regression과 매우 유사하나 Shrinkage penalty를 절대값으로 나타낸 방식

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$$

# 회귀분석 $R^2$ ?

## 결정계수 R-squared

회귀모형의 **설명력**을 표현(0~1의 값)  
값이 ↑ 설명력 ↑

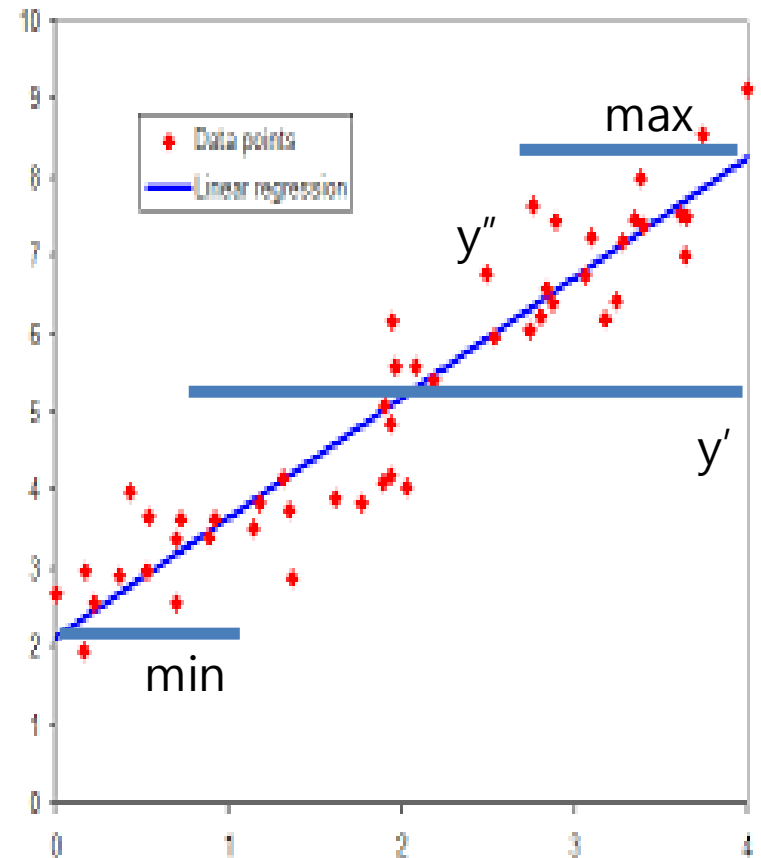
$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

**SST**와 **SSR**이 얼마나 비슷한지, **SST**와 **SSE**가 얼마나 다른  
지에 따라 **R-Squared**값이 달라짐

**SST(Sum of Square Total)** 편차 제곱의 합  
실제 값과 예측 값, 평균 값 사이에 발생한 차이 제곱 합  
$$SST = \sum_{i=1}^n (y_i - y''_i + y''_i - y')^2$$

**SSE(Sum of Square Error)** 회귀 식과 실제 값 차이  
 $y''$ 과 min, max간의 차이를 제곱하여 더한 것  
$$SSE = \sum (y - y'')^2$$

**SSR(Sum of Square Regression)** 회귀식과 평균값의 차이  
 $y''$ 과  $y'$ 간의 엇갈리며 차이가 발생함  
$$SSR = \sum (y'' - y')^2$$
  
즉  $y''$ 과 차이가날 수록 SSR값은 커짐  
 $y'$ 가 모든 데이터를 고루 설명하고 있는 것으로 해석  
SSR값은 **R-squared**값과 비례



## 03. 분석 처리방안 및 활용 분석기법

---

### GS25 판매량 & 날씨 관계 탐색

GS25 판매량과 날씨변수는 어떠한 관계를 가질까??

날씨 변수를 **x**로, GS25 제품 판매량을 **y**로 1차 모델링을 시도함

**x** : snow, rain, cloud, temp, lgt\_time, wind, PM10

**y** : 과자, 라면, 마스크, 맥주, 면도기, 생리대, 생수, 숙취해소제, 스타킹,  
아이스크림, 우산, 탄산음료

✓ Linear Regression, OLS, Ridge Regression, Lasso Regression을 실행

## 03. 분석 처리방안 및 활용 분석기법

### GS25 판매량 & 날씨 관계 탐색 - 1차분석 결과 시계열 분석 일 단위 분석

#### 1차 모델링 결과

Linear, Ridge, Lasso Regression 분석

$R^2$  값이 기준치보다 낮게 나온

GS25 제품은 날씨 변수들과

관계가 없다고 판단, 변수제거

GS 제품	$R^2$
과자	0.11
라면	0.17
면도기	0.1
생리대	0.25
숙취해소제	0.06
스타킹	0.17
우산	0.34

## 03. 분석 처리방안 및 활용 분석기법

### GS25 판매량 & 날씨 관계 탐색 - 1차분석 결과 시계열 분석 일 단위 분석

#### 1차 모델링 결과

GS 제품 마스크, 맥주, 생수, 탄산음료 ,  
아이스크림의 판매량 데이터  
&  
날씨 데이터간의 상관관계가  
가장 높게 나옴

→  $R^2$  값이 40%~80% 근사치의  
변수들만 고려

GS 제품	$R^2$
마스크	0.39
맥주	0.57
생수	0.68
탄산음료	0.40
아이스크림	0.79

### 03. 분석 처리방안 및 활용 분석기법

#### GS25 판매량 & 날씨 관계 탐색 - 2차분석 결과 시계열 분석 주 단위 분석

시계열 분석을  
일 단위 → 주 단위로 분석결과  
날씨변수와 상관관계가 있다고 판단했던  
마스크, 맥주, 생수, 아이스크림 변수의  
 $R^2$ 값이 0.4~0.8 범위 안에 들어옴

탄산음료는  $R^2$  값이 0.4이하라서  
두 변수를 제외

GS 제품	$R^2$
마스크	0.4
맥주	0.57
생수	0.68
탄산음료	0.34
아이스크림	0.79

## 03. 분석 처리방안 및 활용 분석기법

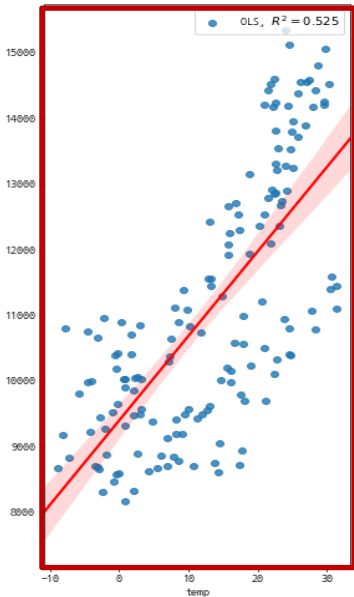
### GS25 판매량 & 날씨 R<sup>2</sup> 분석

GS 편의점	snow	rain	pm10	wind	temp	cloud	lgt_time
TOP 4							
맥주	0.049	0.087	0.074	0	0.525	0.142	0
마스크	0.017	0.034	0.386	0.001	0.103	0.016	0
아이스크림	0.077	0.067	0.052	0.01	0.762	0.096	0.044
생수	0.095	0.009	0.02	0.006	0.622	0.045	0.058
탄산음료	0.082	0.005	0	0	0.333	0.05	0.021
과자	0	0.004	0.056	0.312	0.001	0.009	0.001
라면	0.014	0	0.059	0.14	0.079	0.001	0.025
면도기	0.044	0.007	0.005	0.079	0.146	0.032	0.003
생리대	0.052	0.003	0.007	0.102	0.183	0.019	0.004
숙취해소제	0	0.015	0.077	0.045	0.002	0.002	0
스타킹	0.014	0.045	0.222	0.084	0.043	0.031	0.018
우산	0.01	0.002	0.033	0.027	0.002	0	0.004

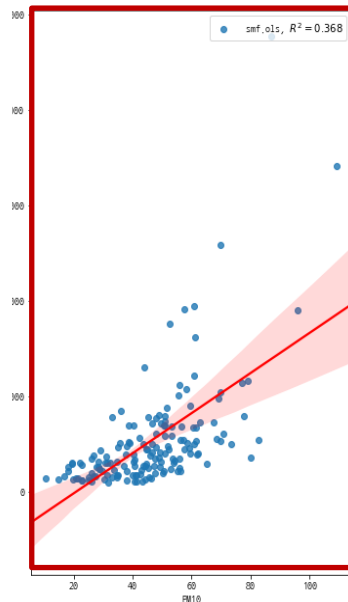


# 03. 분석 처리방안 및 활용 분석기법

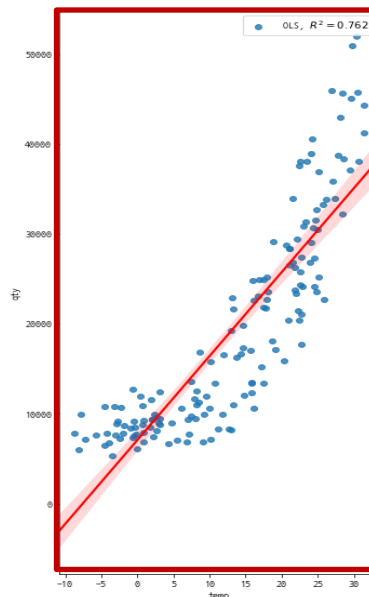
맥주 : Temp



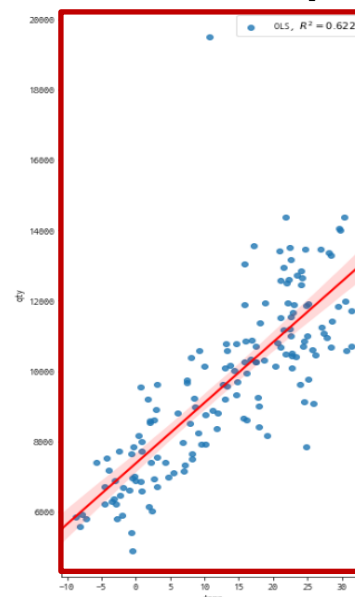
마스크 : PM10



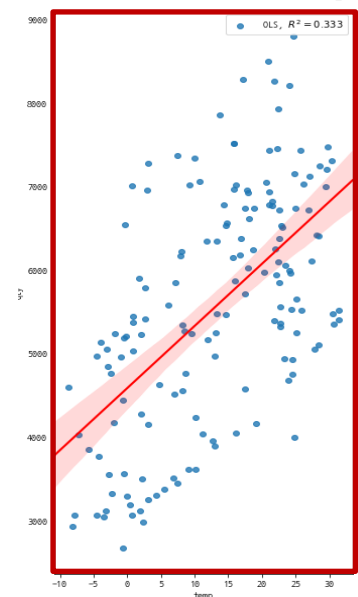
아이스크림 : Temp



생수 : Temp



탄산음료 : Temp



마스크는 미세먼지변수인 pm10과  $R^2$ 값이 **0.386**으로 가장 크게 나옴  
맥주, 아이스크림, 생수는 온도 변수인 temp와 각각 **0.525, 0.762, 0.622**로  
최고 값이 나옴

-> 따라서 마스크판매량은 미세먼지 수치와,  
맥주, 아이스크림, 생수는 온도 수치와  
가장 큰 상관관계가 있다는 결론을 도출

## 03. 분석 처리방안 및 활용 분석기법

---

### 랄라블라 판매량 & 날씨 관계 탐색

랄라블라 판매량과 날씨변수는 어떠한 관계를 가질까??

날씨 변수를 **x**로, 랄라블라 제품 판매량을 **y**로 1차 모델링을 시도함

**x**: snow, rain, cloud, temp, lgt\_time, wind, PM10

**y**: 네일, 립컬러, 립케어, 마스크팩, 바디로션, 선케어, 제모제, 체중조절, 크림로션  
웨이셜 클렌저

✓ Linear Regreesion, OLS, Ridge Regression, Lasso Regression을 실행

## 03. 분석 처리방안 및 활용 분석기법

랄라블라 판매량 & 날씨 관계 탐색

- 1차분석 결과

시계열 분석 일 단위 분석

### 1차 모델링 결과

Linear, Ridge, Lasso Regression 분석

$R^2$  값이 기준치보다 낮게 나온

랄라블라 제품은 날씨 변수들과

관계가 없다고 판단

랄라블라 제품명	$R^2$
립컬러	0.029
마스크팩	0.02
바디로션	0.017
체중조절	0.026
크림로션	0.033
웨이셜 크렌저	0.048

## 03. 분석 처리방안 및 활용 분석기법

랄라블라 판매량 & 날씨 관계 탐색

- 1차분석 결과

시계열 분석 일 단위 분석

### 1차 모델링 결과

랄라블라 제품 선케어, 제모제의  
판매량 데이터 & 날씨 데이터의  
상관관계가 가장 높게 나옴

→  $R^2$  값이 30%~50% 근사치의  
변수들만 고려

랄라블라 제품명	$R^2$
선케어	0.348
제모제	0.431
네일	0.320
립케어	0.293

## 03. 분석 처리방안 및 활용 분석기법

랄라블라 판매량 & 날씨 관계 탐색

- 2차분석 결과

시계열 분석 주 단위 분석

### 2차 모델링 결과

시계열 분석을 일 단위 → 주 단위로

선케어, 제모제의  
 $R^2$ 값이 0.5~0.6 범위 안에 들어옴

→ 선케어, 제모제는 날씨와  
상관관계가 있는 것으로 판단

랄라블라 제품	$R^2$
선케어	0.5
제모제	0.55

네일은 회귀분석결과 훈련데이터와  
검증데이터가 각각 0.53, -0.63이라서  
변수에서 제외

립케어는 회귀분석결과 훈련데이터와  
검증데이터가 각각 0.37, 0.28 이라서  
변수에서 제외

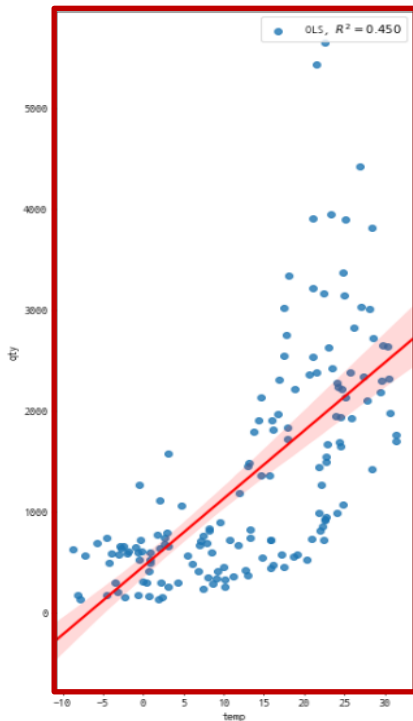
## 03. 분석 처리방안 및 활용 분석기법

### 랄라블라 판매량 & 날씨 R2 분석

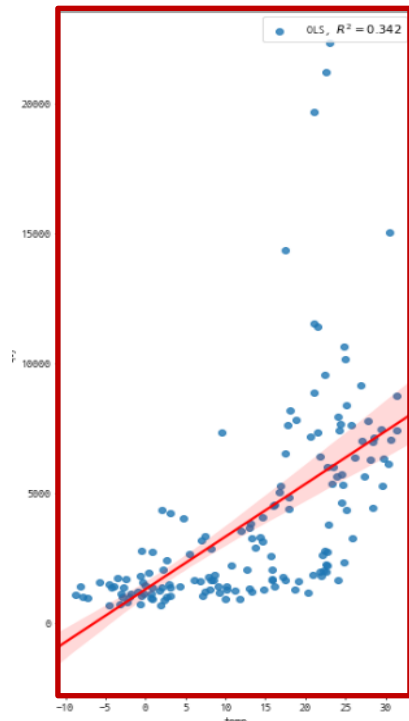
랄라블라	Snow	rain	pm10	wind	temp	cloud	lgt_time
TOP 2							
제모제	0.043	0.075	0.002	0.001	0.45	0.079	0.034
선풍기	0.037	0.012	0.003	0	0.342	0.009	0.139
네일	0.02	0.055	0.158	0.115	0.316	0.043	0.018
립케어	0.035	0.119	0.013	0.016	0.317	0.099	0.019
립컬러	0.003	0.019	0.012	0.032	0.004	0.011	0
마스크팩	0	0.007	0.002	0.008	0.009	0.001	0
바디로션	0.002	0.024	0.045	0.017	0.025	0.074	0.021
체중조절	0.007	0	0	0	0.047	0.002	0.008
크림로션	0.004	0.018	0.003	0.026	0.062	0.008	0.003
헤어샴푸 클렌저	0.008	0.007	0	0.017	0.033	0.001	0.007

## 03. 분석 처리방안 및 활용 분석기법

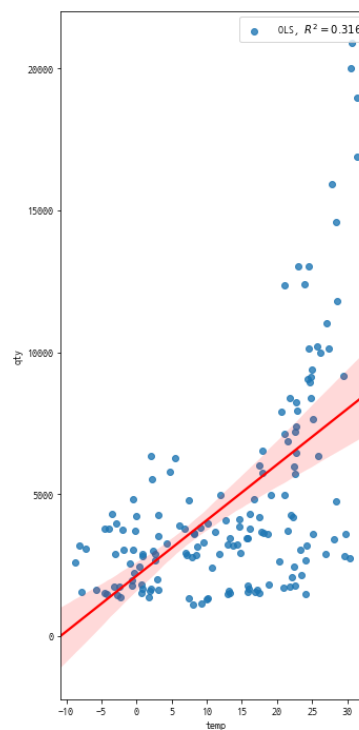
제모제 : Temp



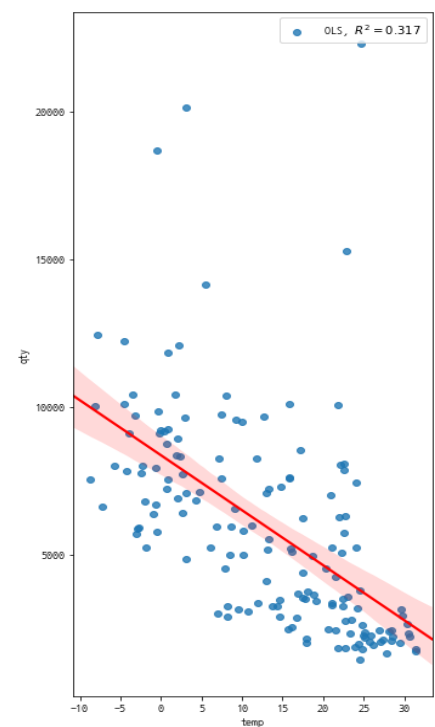
선케어 : Temp



네일 : Temp



립케어 : Temp



제모제와 선케어는 온도 변수인 temp와  $R^2$ 값이 **0.45, 0.342**로 최고 값이 나옴

-> 따라서 **제모제, 선케어 판매량**은 온도 수치와 가장 큰 상관관계가 있다는 결론을 도출

## 03. 분석 처리방안 및 활용 분석기법

### GS25, 랄라블라 판매량 & 날씨 관계 탐색 2차분석 결과 정리

랄라블라는 **선케어, 제모제**가 temp와 가장 큰  $R^2$  관계를 보임

랄라블라	
선케어	Temp
제모제	Temp

GS25는 마스크는 **미세먼지(PM10)**와 맥주, 생수, 아이스크림  
은 **temp**와 가장 큰  $R^2$ 보임

GS25	
마스크	미세먼지 (PM10)
아이스크림	Temp
맥주	Temp
생수	temp

선케어, 제모제, 마스크, 아이스크림, 맥주, 생수  
판매량을 **예측** 할 때,  
해당 기온 변수의 변화량으로 추정 가능!

판매량을 예측하기 전  
주요 **서울시 구별, 성별, 연령별** 판매량분석이 필요함  
공급처와, Targeting할 성별, 연령대의 구매량을 개량화  
최적의 공급처, 성별, 연령별을 선정하기 위함



## 03. 분석 처리방안 및 활용 분석기법

### GS25 판매량 분석 결과

서울시 주요 구별, 성별, 연령별을 중심으로

GS25판매량	아이스크림	맥주	마스크	생수	GS25판매량	아이스크림	맥주	마스크	생수
강남	1,010,603	440,636	28,619	415,578	강남	F, 2, 1	F, 2, 2	F, 2, 4	F, 2, 3
관악	728,236	493,911	23,354	294,069	관악	F, 2, 1	F, 2, 2	F, 2, 4	F, 2, 3
마포	768,265	342,207	25,911	401,673	마포	F, 2, 1	F, 2, 3	F, 2, 4	F, 2, 2
서초	449,657	192,098	16,909	283,812	서초	F, 2, 1	F, 2, 3	F, 2, 4	F, 2, 2
영등포	459,382	233,755	14,351	279,480	영등포	F, 2, 1	F, 2, 3	F, 2, 4	F, 2, 2
서대문	541,347	228,480	19,828	137,362	서대문	F, 2, 1	F, 2, 2	F, 2, 4	M, 2, 3

제품별판매량(단위: 개)

성별(F: female, M: male), 세대(2: 20~30대), 구별 판매순위)

제품 판매량이 가장 많은 TOP 6 구 선정

-> 강남, 관악, 마포, 서초, 영등포, 서대문

**GS25편의점** 제품(아이스크림, 맥주, 마스크, 생수)은 대체로 20~30대 여성이 가장 많이 구매하는 것으로 파악!

강남, 관악, 마포, 서초, 영등포, 서대문구를 중심으로  
20~30대 여성의 주요 고객 층으로 선정

## 03. 분석 처리방안 및 활용 분석기법

### 랄라블라 판매량 분석 결과

서울시 주요 구별, 성별, 연령별을 중심으로

랄라블라 판매량	제모제	선케어	랄라블라 판매량	제모제	선케어
강서	136,705	346,398	강서	F,2	F,2
마포	104,377	279,313	마포	F,2	F,2
광진	122,987	297,955	광진	F,2	F,2
관악	104,189	259,501	관악	F,2	F,2
동작	104,112	250,221	동작	F,2	F,2
서대문	95,023	275,987	서대문	F,2	F,2

제품 별 판매량(단위: 개)

성별(F: female), 세대(2: 20~30대)

제품 판매량이 가장 많은 **TOP 6** 구 선정

-> 강서, 마포, 광진, 관악, 동작, 서대문

**랄라블라** 제품(제모제, 선케어)은 대체로 **20~30대 여성**이  
가장 많이 구매하는 것으로 파악!

강서, 마포, 광진, 관악, 동작, 서대문 중심으로  
20~30대 여성의 주요 고객 층으로 선정



# 판매량 예측 모형

**DNN(Deep Neural Network)**

# 03. 분석 처리방안 및 활용 분석기법

## Deep Learning



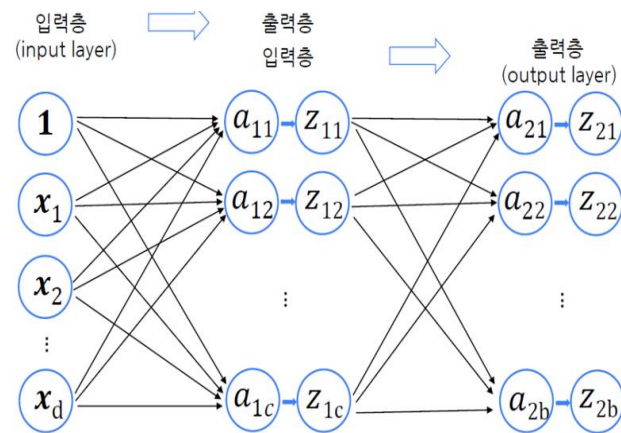
### 판매량 예측 모형 DNN

## What?

출력층이자 입력층이 되는 가운데 노드가 은닉층(Hidden Layer)  
은닉층의 수가 2개 이상인 것이 **DNN**  
은닉층이 학습을 반복하는 것이 **Deep Learning**

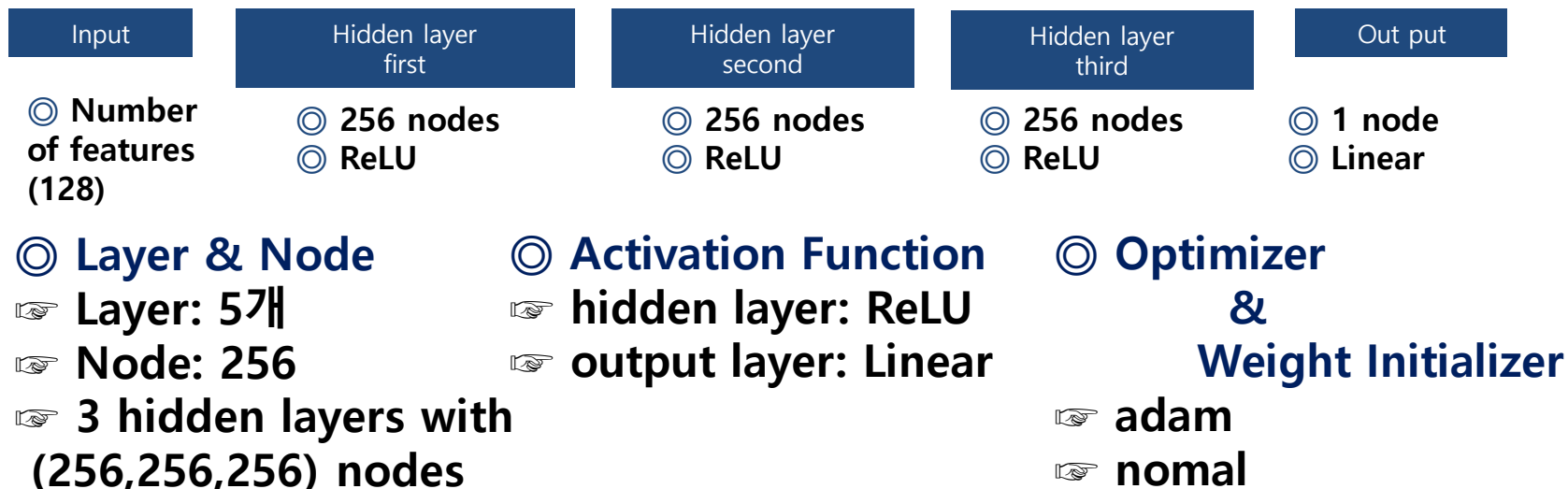
## Why?

1. Back propagation으로 학습 성능이 우수
2. 데이터에 따라 튜닝이 자유로움



# 03. 분석 처리방안 및 활용 분석기법

## How? 파라미터 튜닝



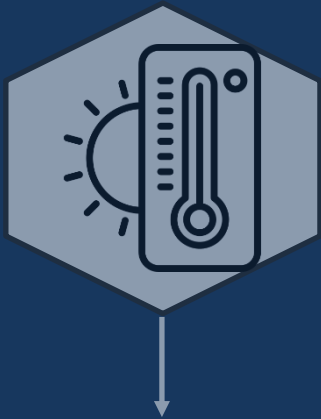
### Model: Sequential

- ✓ 학습 횟수  
Epochs: 1000회
- ✓ 기울기 업데이트할 샘플 개수  
Batch\_size: 32개

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 128)	1280
dense_2 (Dense)	(None, 256)	33024
dense_3 (Dense)	(None, 256)	65792
dense_4 (Dense)	(None, 256)	65792
dense_5 (Dense)	(None, 1)	257
Total params: 166,145		
Trainable params: 166,145		
Non-trainable params: 0		

# 데이터 처리 프로세스

---



## 2. 활용 데이터 정의

- (1) 날씨 데이터
- (2) 랄라블라  
및  
GS25 판매량 데이터
- (3) 소셜 데이터



## 3. 분석 처리방안 및 활용 분석기법

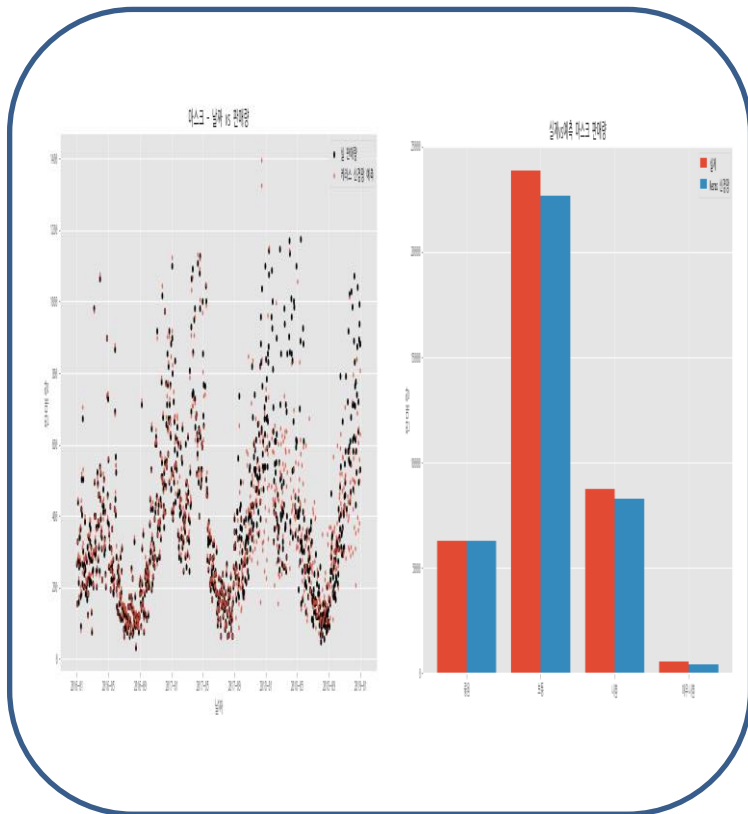


## 4. 분석 결과 및 분석의 한계

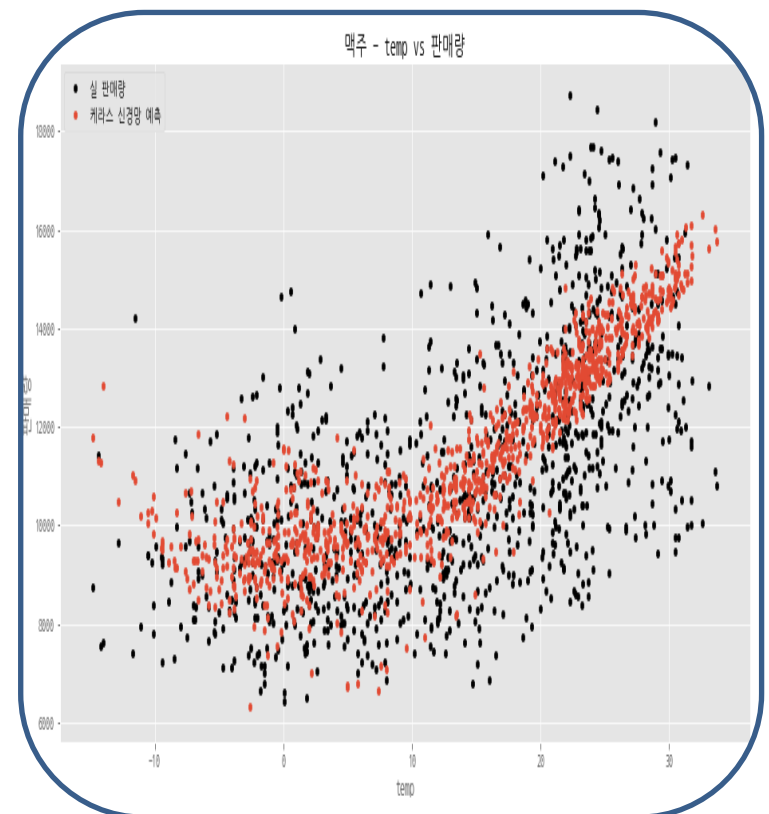
## 04. 분석 결과

# GS25 Result

Keras 신경망 예측  
마스크



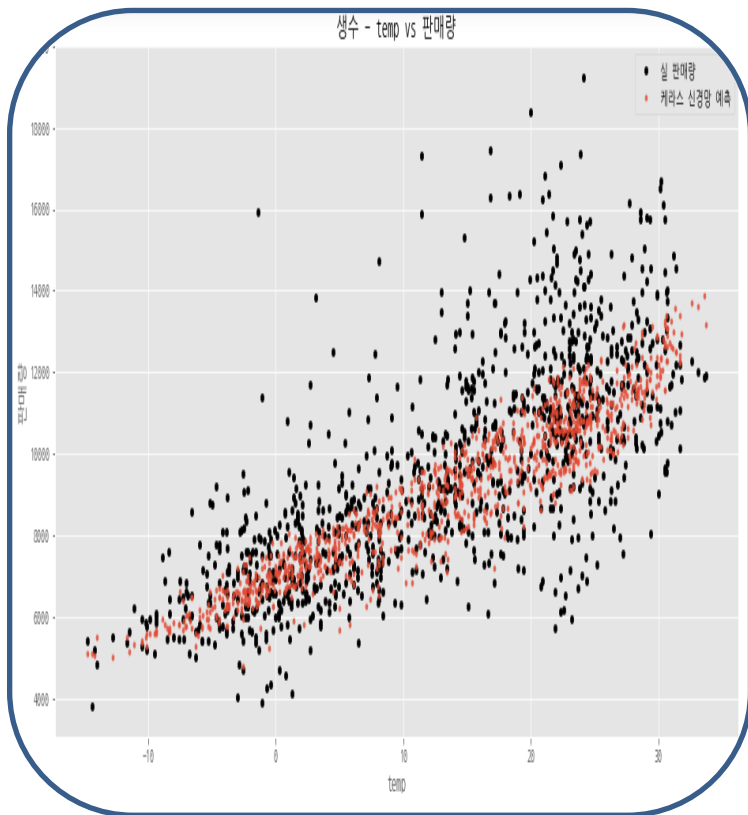
Keras 신경망 예측  
맥주



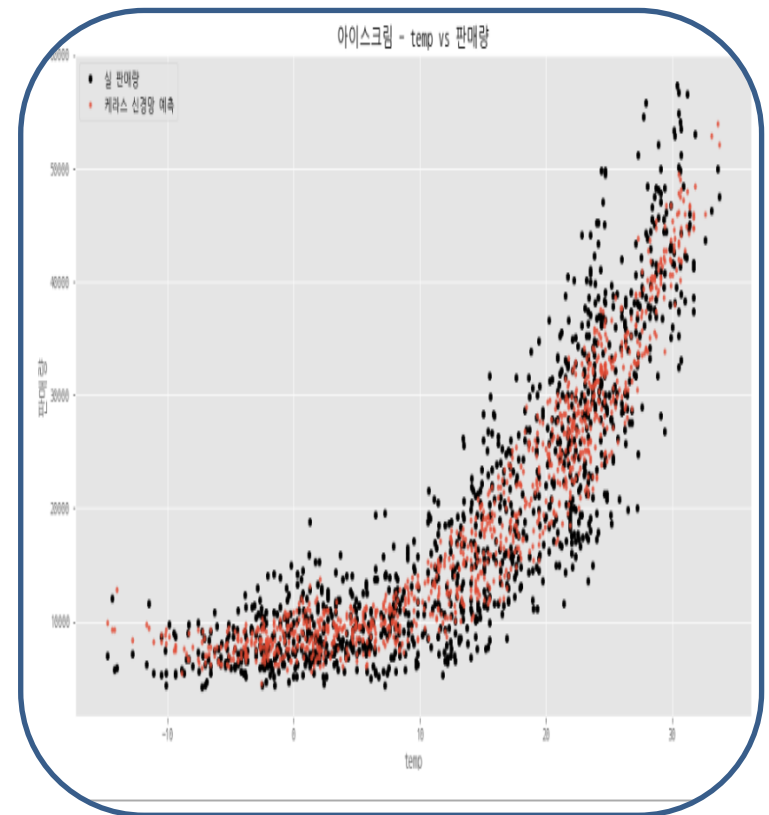
## 04. 분석 결과

# GS25 Result

Keras 신경망 예측  
생수



Keras 신경망 예측  
아이스크림

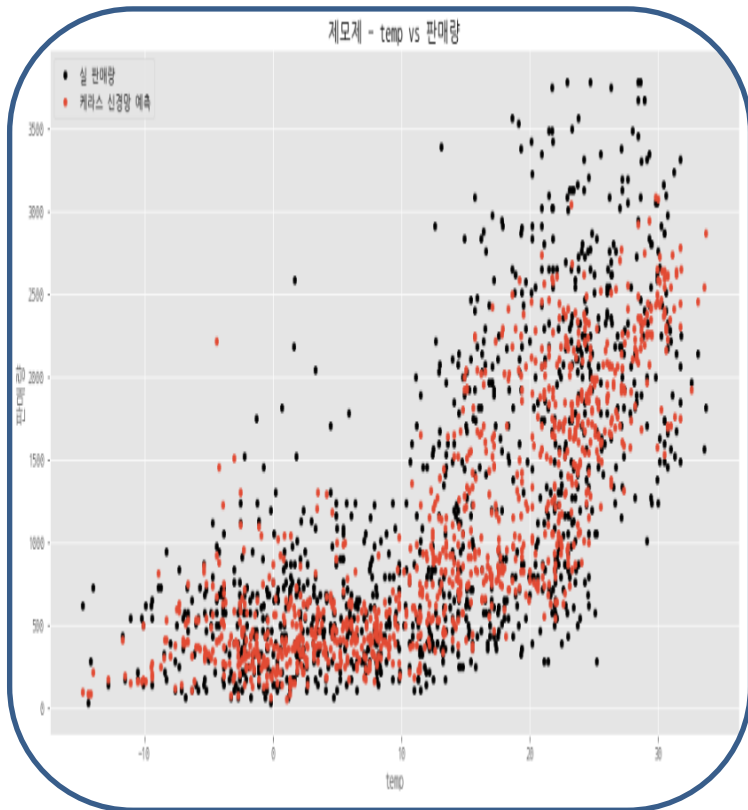




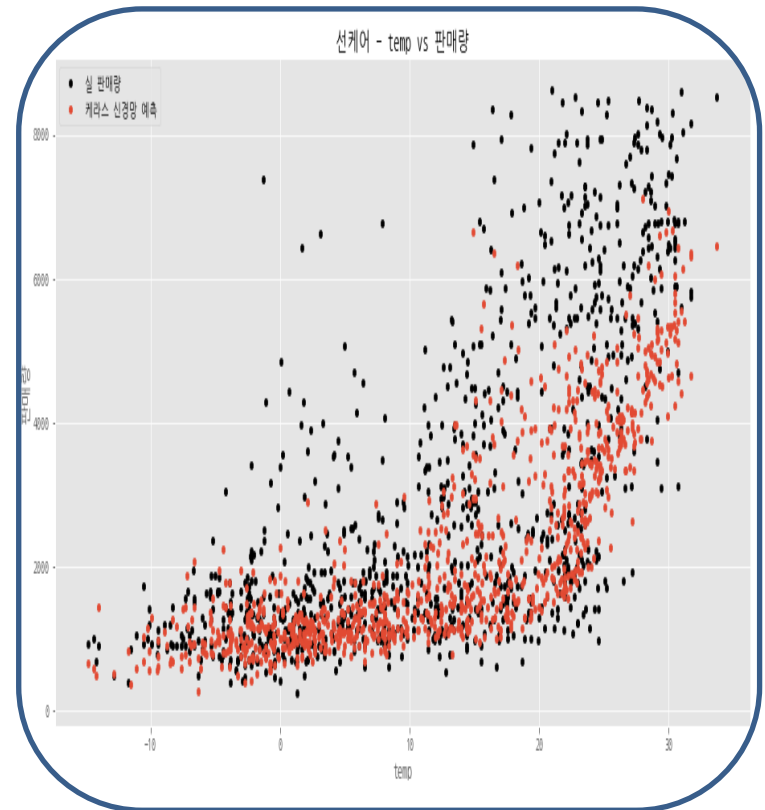
## 04. 분석 결과

# 랄라블라 Result

Keras 신경망 예측  
제모제



Keras 신경망 예측  
선케어



## 04. 분석 결과

### Result

판매량 예측 모형 평가  $M = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$   $A_t$ :실제값  
- 정확도 예측  $F_t$ :예측값

MAPE (Mean Absolute Percentage Error)

- 실제값과 예측값 간의 차이

MAPE	맥주	마스크	생수	아이스크림	제모제	선케어
Keras	0.15	0.37	0.14	0.23	0.49	0.36
Linear	0.17	0.41	0.15	0.38	0.89	0.59
Ridge	0.17	0.43	0.16	0.37	0.91	0.59
Lasso	0.17	0.41	0.15	0.38	0.89	0.59
OLS	0.16	0.41	0.16	0.27	0.83	0.52

Keras 모델의 MAPE 오차 값 < 다른 회귀분석의 오차 값

-> 모델의 정확도 ↑

## 04. 분석 한계

### 구매성향과 평균 판매량을 통해 익년도 적정 재고량 추정(Model)

	선케어(X)	제모제(Y)	총방문자수(Z)	구매성향 (P) (각상품별 판매 량/ 총판매량)	평균 구매량 (Q) (상품별 판매량)
2016	10	20	100	$P = 0.222$	$Q = 0.450$
2017	17	10	150	$P = 0.436$	$Q = 0.26$
예측 2018	14		125		

$$X \text{ 구매성향 } P = X / (X + Y)$$

$$X \text{ 평균 판매량 } Q = X / (Z * P)$$

2018년 Z = 2016, 2017년도 Z의 평균합계

2018년도 X제품을 구매할 고객 = 2018년 Z \* 2017년 P

2018 예측된 적정 판매량 = 2018년 X제품 구매할 고객 \* 2017년 Q

전제 조건 : 총 방문자 수는 중복없이 독립적임, 세대별 분류의 방문 수

모델의 한계 : GS25와 랄라블라의 수요 발생 원인의 불일치성이 존재,

☞ GS25 화장품 수요를 예측하기 어려움 GS25화장품 판매량 데이터가 부족

☞ 개별 방문자데이터가 요구

**But** 기존 랄라블라매장의 화장품 적정 재고량을 예측 가능

## 04. 분석 한계

**구매성향과 평균 판매량을 통해**

**익년도 적정 재고량 추정**

관악구

랄라블라	선케어(x) 단위:개	제모제(z) 단위:개	세대별 총방문자수 단위:명	세대별 구매성향 (각상품별 판매량 / 총판매량)	세대별 평균 구매량 (상품별 판매량) 단위:개
2016	139,963	52,129	1,457	X = 0.729	X = 131
2017	133,529	47,422	1,398	X = 0.737	X = 129
2018	124,,206	37,832	1,414	X = 0.76	X = 115
예측 2019	135,750		<b>1,423</b>	X = 0.763	X = 125



기존 랄라블라 매장의 화장품 적정 재고량 예측 가능

**관악구는 2019년 선케어를 약 135,750개 판매할 것으로 예상**

## 04. 분석 한계

### 구매성향과 평균 판매량을 통해 익년도 적정 재고량 추정

마포구

랄라블라	선케어(x) 단위:개	제모제(z) 단위:개	세대별 총방문자수 단위:명	세대별 구매성향 (각상품별 판매량 / 총판매량)	세대별 평균 구매량 (상품별 판매량) 단위:개
2016	138,622	55,743	1,459	X = 0.72%	X = 132
2017	147,224	51,797	1,463	X = 0.74%	X = 135
2018	167,262	35,278	1,419	X = 0.82%	X = 143
예측 2019	146,600		<b>1447</b>	X = 0.76%	X = 136



기존 랄라블라 매장의 화장품 적정 재고량 예측 가능

**마포구**는 2019년 선케어를 **약 146,600개** 판매할 것으로 예상

## 04. 분석 한계

**구매성향과 평균 판매량을 통해**

**익년도 적정 재고량 추정**

서대문구

랄라블라	선케어(x) 단위:개	제모제(z) 단위:개	세대별 총방문자수 단위:명	세대별 구매성향 (각상품별 판매량 / 총판매량)	세대별 평균 구매량 (상품별 판매량) 단위:개
2016	114,865	46,592	1,285	X = 0.71%	X = 125
2017	150,062	47,127	1,352	X = 0.76%	X = 146
2018	121,169	32,478	1,274	X = 0.79%	X = 120
예측 2019	127,010		<b>1,303</b>	X = 0.75%	X = 130



기존 랄라블라 매장의 화장품 적정 재고량을 예측 가능

**서대문구는 2019년 선케어가 약 127,010개 팔릴 것으로 예상**

## 05. 서비스 활용방안



랄라블라 영업점을 위한  
판매량 예측 홈페이지 개설

lalavla

랄라블라  
판매량



기상청  
날씨데이터



GS25  
판매량

√ 날씨데이터와 연간 판매량을 입력

마포, 관악, 서대문 구 랄라블라 영업점이  
제모제, 선크어 제품을 여성 20~30대를 주 고객층으로  
GS편의점에 아이스크림, 맥주 등과 함께 진열, 판매할 수 있다

## 05. 서비스 활용방안

---



랄라블라 영업점을 위한  
판매량 예측 홈페이지 개설

lalavla

랄라블라  
화장품 판매량



GS25 화장품  
판매량

- ✓ **DNN 예측모델과 구매성향, 평균구매량**을 통해 연간 총 방문자 수를 입력하면 **적정 재고량**을 파악할 수 있다



## 05. 서비스 기대효과

---

◎ 제모제, 선케어 등 랄라블라 제품 **공급망이 확대** 될 것으로 기대

√ GS25매장(13,107)에 제품을 공급,  
적자의 원인으로 지적했던 **랄라블라 영업점(168개) 수** 극복

☞ **수익 개선 !**

◎ DNN모델과 구매 성향, 평균 구매량으로  
**익년 재고 예측**이 가능해 질 것으로 기대

√ 랄라블라 매장의 **악성 재고 감소**

☞ **수익 개선 !**