

L.POINT

Digital Trend Analyzer

Final Project

- 상품군별 온라인 선호지수 생성 및 전략도출



📁 훈련 과정 명 : 데이터베이스기반 빅데이터 분석가 양성과정

📁 훈련 기간 : 2019-11-26 ~ 2020-04-29 (800시간/100일)

📁 팀 명 : NOAH 조

📁 | 팀 원 : 권나*, 김지*, 김현*, 박윤*, 이승*, 최승*, 황윤*



1. 프로젝트 개요



2. 분석



3. 소비자 선호지수



4. 서비스 제안



5. 결론



6. 부록



1. 프로젝트 개요

📁 1. 분석개요

📁 2. 분석시스템아키텍처

📁 3. 분석시스템환경구성도

📁 4. 주요변수기술통계

📁 5. 데이터분포

1. 프로젝트 개요 - 1) 분석개요 (1/2)

- Be the L.BA 데이터 설명서 기반으로 작성됨

분석 데이터

- 롯데그룹 온라인 계열사의 구매 및 이용 이력

데이터 수집 채널

- 롯데그룹 온라인 계열사(롯데닷컴, 롯데홈쇼핑 등)
※ 임의 선정, 본 데이터는 가공된 샘플 데이터로 실제 시장 데이터와 차이가 있음

데이터 수집 기간

- 총 6개월 (2018년 04월 ~ 2018년 09월)

제공 데이터 종류

- 온라인 행동 데이터(Product, Search1, Search2, Custom, Session), 분류(Master)

분석 방법

- 통계기반 분석 기법(기초 통계 분석, 회귀 분석 등)을 통해 빅데이터 분석 진행

분석 내용

- L.POINT 멤버십 가입자 소비동향 및 디지털 트렌드 Key Needs & Key Insight 도출

1. 프로젝트 개요 - 1) 분석개요 (2/2)

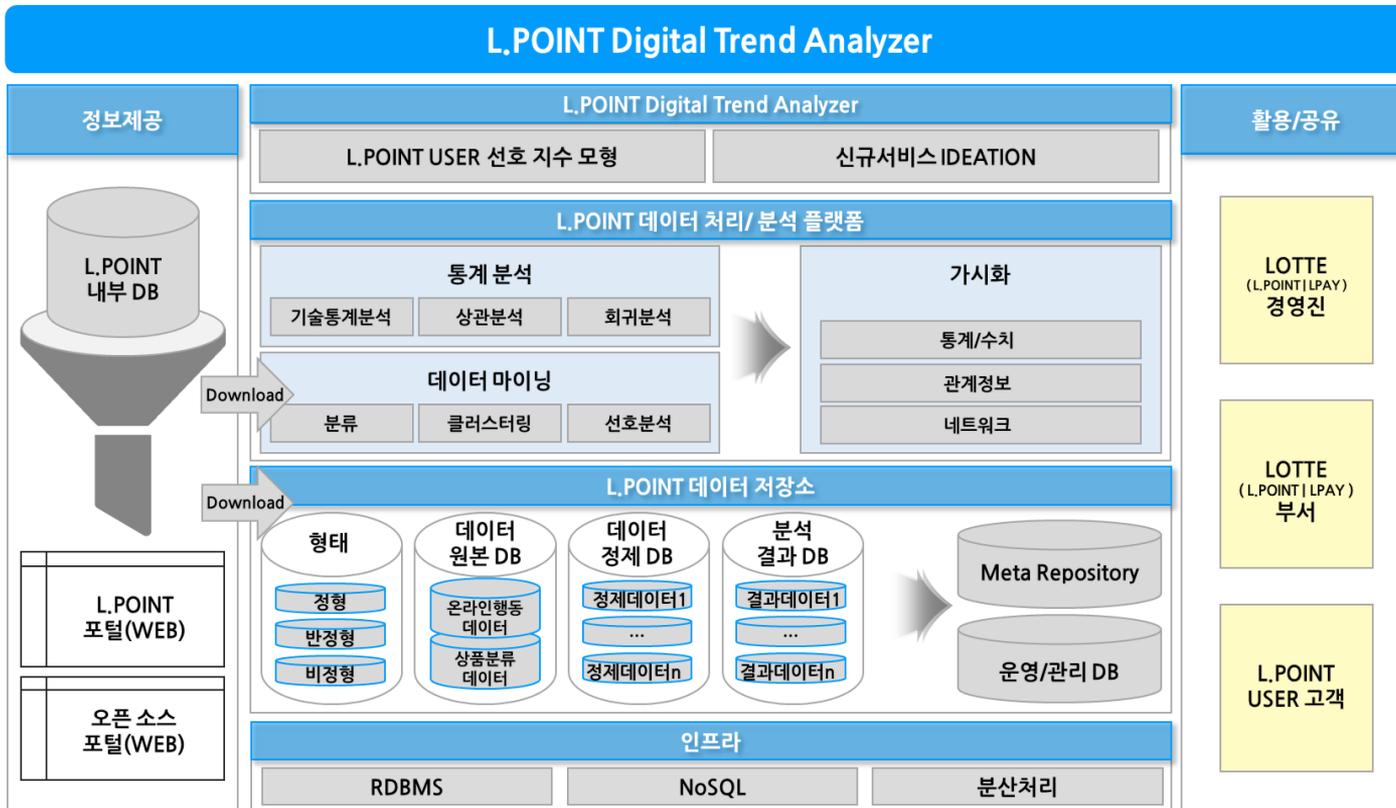
분석 주제

L.POINT USER 온라인 행동기반 트렌드 분석

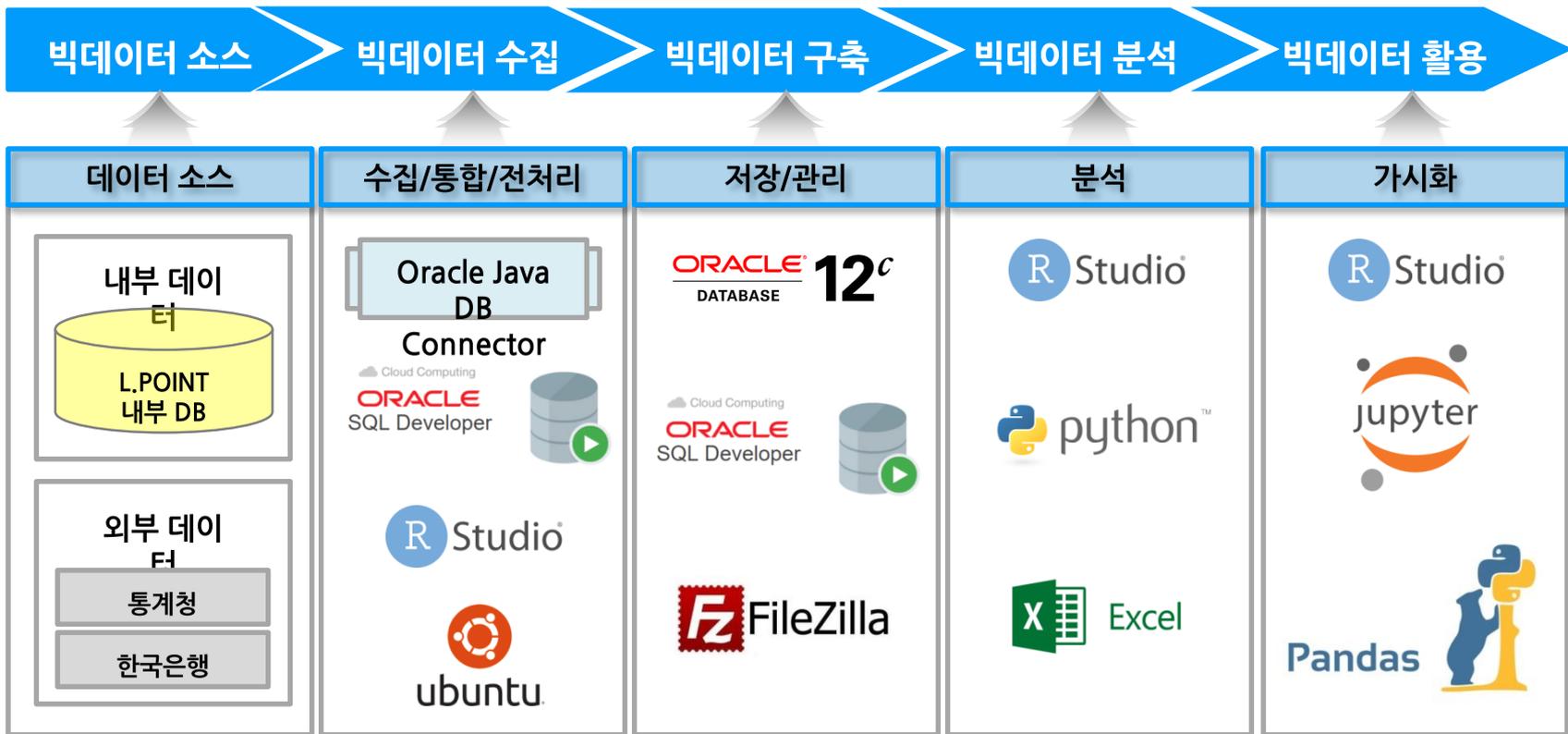
분석 목적

- 주요 상품군별 온라인 선호지수 생성
- 온라인 선호지수 활용한 인사이트 도출
- 1), 2)를 활용한 신규 서비스 제안

1. 프로젝트 개요 - 2) 분석 시스템 아키텍처



1. 프로젝트 개요 - 3) 분석 시스템 환경 구성도



1. 프로젝트 개요 - 4) 주요변수 기술통계 (1/2)

상품구매 (PRODUCT)

Moments			
N	3970214	Sum Weights	3970214
Mean	88.4976644	Sum Observations	351354666
Std Deviation	89.3879947	Variance	7990.21359
Skewness	1.93188052	Kurtosis	3.87441131
Uncorrected SS	6.28169E10	Corrected SS	3.17228E10
Coeff Variation	101.00605	Std Error Mean	0.04486134

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.170844	Pr > D	<0.0100

Quantiles (Definition 5)	
Level	Quantile
100% Max	500
75% Q3	115
50% Median	56
25% Q1	28
0% Min	2

HIT_SEQ

상품구매 (PRODUCT)

Moments			
N	3970214	Sum Weights	3970214
Mean	51135.0857	Sum Observations	2.03017E11
Std Deviation	115098.783	Variance	1.32477E10
Skewness	15.070338	Kurtosis	556.966908
Uncorrected SS	6.29776E16	Corrected SS	5.25963E16
Coeff Variation	225.087689	Std Error Mean	57.7648662

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.328732	Pr > D	<0.0100

Quantiles (Definition 5)	
Level	Quantile
100% Max	17779000
75% Q3	50000
50% Median	28900
25% Q1	13000
0% Min	90

PD_BUY_AM

상품구매 (PRODUCT)

Moments			
N	3970214	Sum Weights	3970214
Mean	1.15036217	Sum Observations	4567184
Std Deviation	2.76194176	Variance	7.62832228
Skewness	186.189229	Kurtosis	59080.5032
Uncorrected SS	35539980	Corrected SS	30286064.3
Coeff Variation	240.093236	Std Error Mean	0.00138614

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.478292	Pr > D	<0.0100

Quantiles (Definition 5)	
Level	Quantile
100% Max	1440
75% Q3	1
50% Median	1
25% Q1	1
0% Min	1

PD_BUY_CT

1. 프로젝트 개요 - 4) 주요변수 기술통계 (2/2)

상품구매 (PRODUCT)

Moments			
N	3970214	Sum Weights	3970214
Mean	52907.9665	Sum Observations	2.10056E11
Std Deviation	117981.374	Variance	1.39196E10
Skewness	15.7932164	Kurtosis	634.455931
Uncorrected SS	6.63774E16	Corrected SS	5.52638E16
Coeff Variation	222.99359	Std Error Mean	59.2115584

Tests for Normality				
Test	Statistic	p Value		
Kolmogorov-Smirnov	D	0.327904	Pr > D	<0.0100

Quantiles (Definition 5)	
Level	Quantile
100% Max	17779000
75% Q3	53000
50% Median	29000
25% Q1	14400
0% Min	90

PD_BUY_TOT

상품구매 (PRODUCT)

Moments			
N	3970214	Sum Weights	3970214
Mean	111.690703	Sum Observations	443435993
Std Deviation	105.447245	Variance	11119.1215
Skewness	1.69090858	Kurtosis	2.63333709
Uncorrected SS	9.3673E10	Corrected SS	4.41453E10
Coeff Variation	94.4100468	Std Error Mean	0.05292103

Tests for Normality				
Test	Statistic	p Value		
Kolmogorov-Smirnov	D	0.157769	Pr > D	<0.0100

Quantiles (Definition 5)	
Level	Quantile
100% Max	499
75% Q3	148
50% Median	75
25% Q1	37
0% Min	1

TOT_PAG_VIEW_CT

상품구매 (PRODUCT)

Moments			
N	3970214	Sum Weights	3970214
Mean	1683.589	Sum Observations	6684208614
Std Deviation	1700.30297	Variance	2891030.19
Skewness	3.19590191	Kurtosis	29.5718377
Uncorrected SS	2.27315E13	Corrected SS	1.1478E13
Coeff Variation	100.992758	Std Error Mean	0.8533346

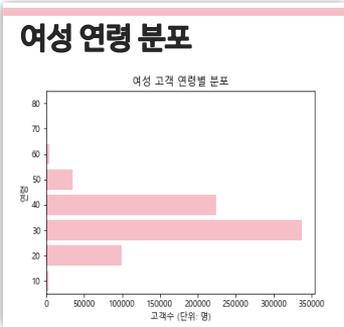
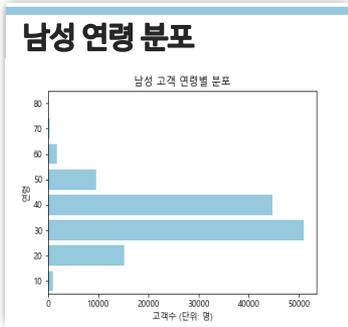
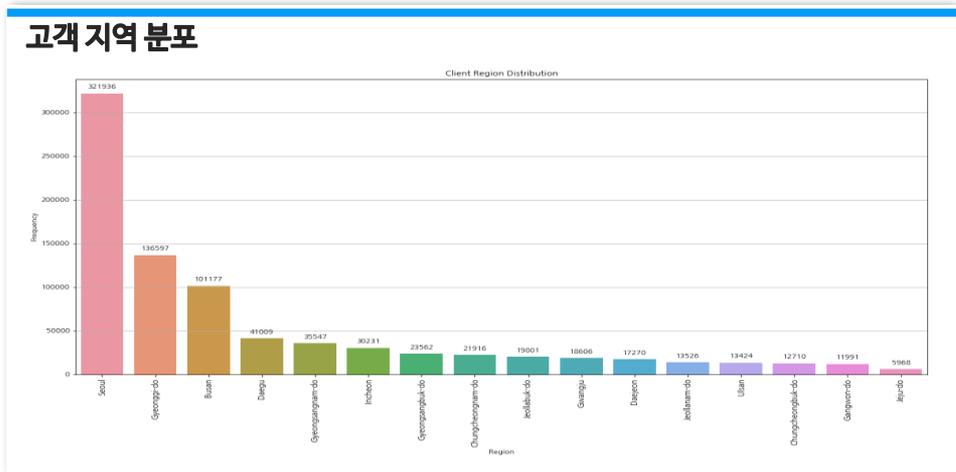
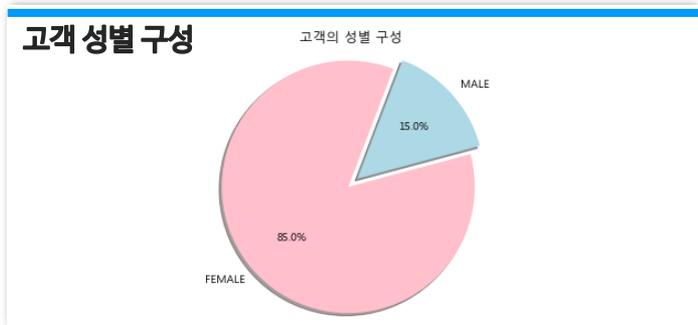
Tests for Normality				
Test	Statistic	p Value		
Kolmogorov-Smirnov	D	0.164238	Pr > D	<0.0100

Quantiles (Definition 5)	
Level	Quantile
100% Max	39329
75% Q3	2318
50% Median	1163
25% Q1	506
0% Min	1

TOT_SESS_HR_V

1. 프로젝트 개요 - 5) 데이터 분포 _ 고객데이터 (1/2)

▪ LPOINT User Profile



✓ LPOINT User **여자** × **30대** × **서울**

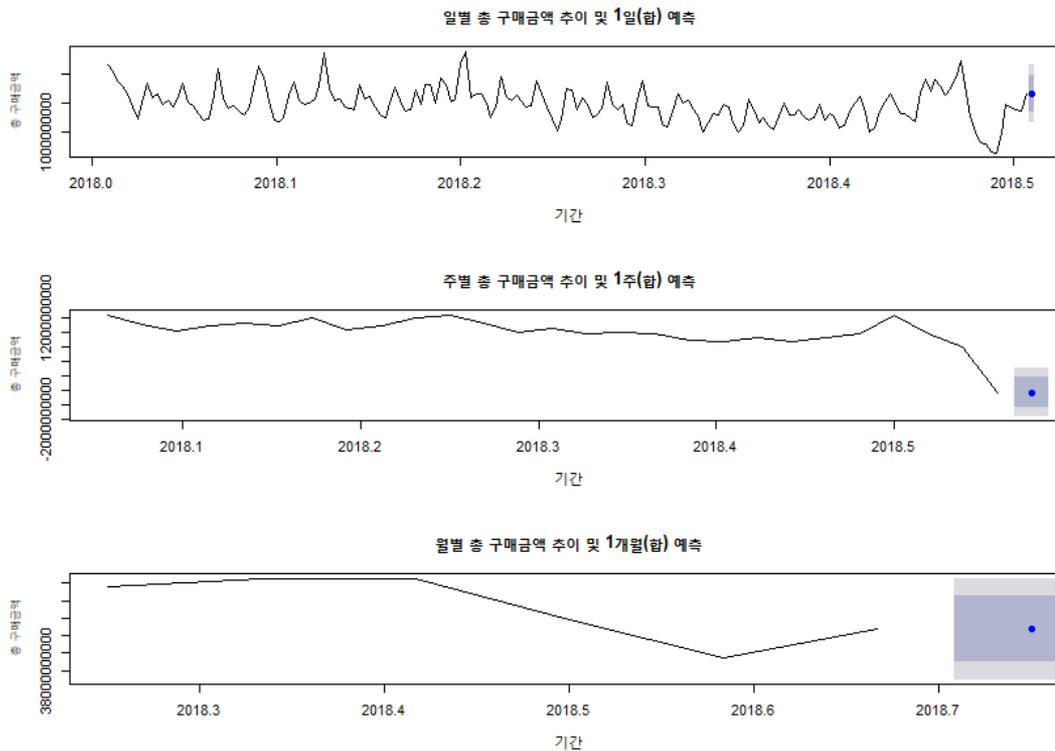


2. 분석

- 1. 시계열 분석
- 2. t - test
- 3. 분산분석

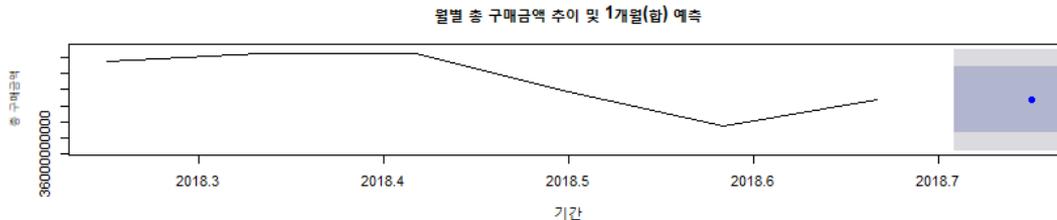
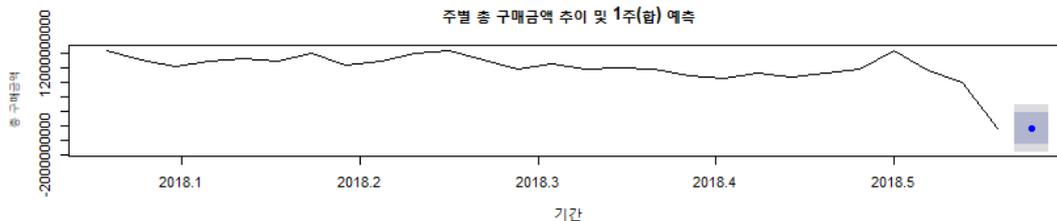
2. 분석 - 1) 시계열 분석

단순 지수 예측 모형 (일, 주, 월)



2. 분석 - 1) 시계열 분석

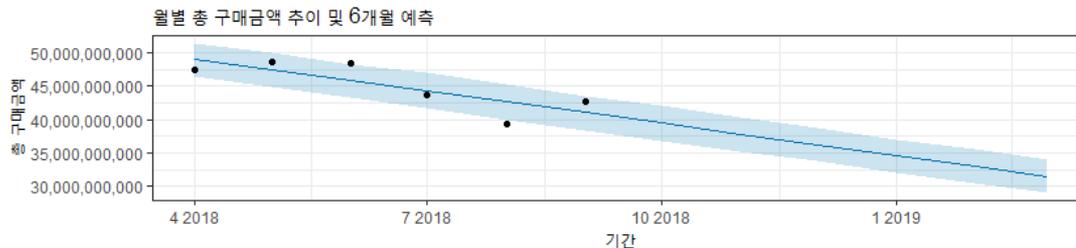
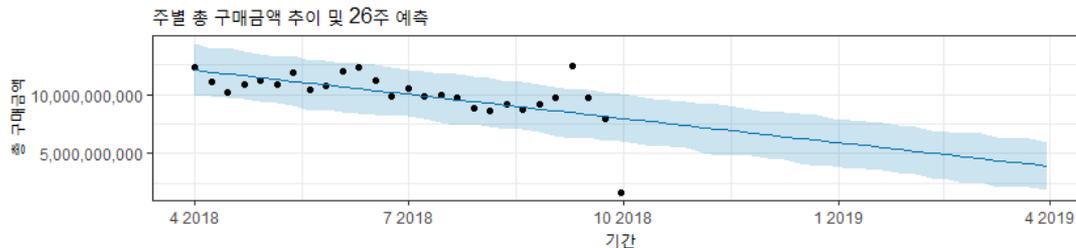
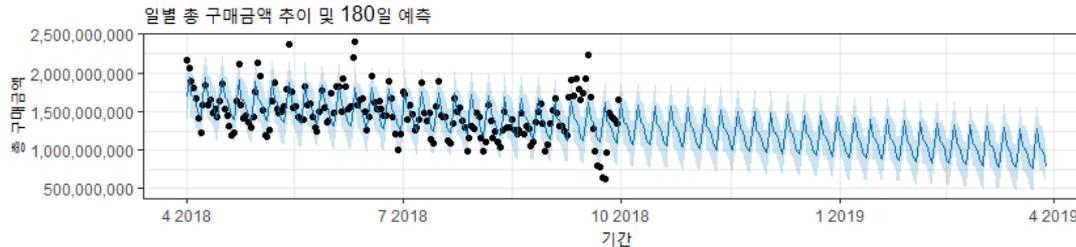
ARIMA 모형 (일, 주, 월)



2. 분석 - 1) 시계열 분석

Prophet 패키지

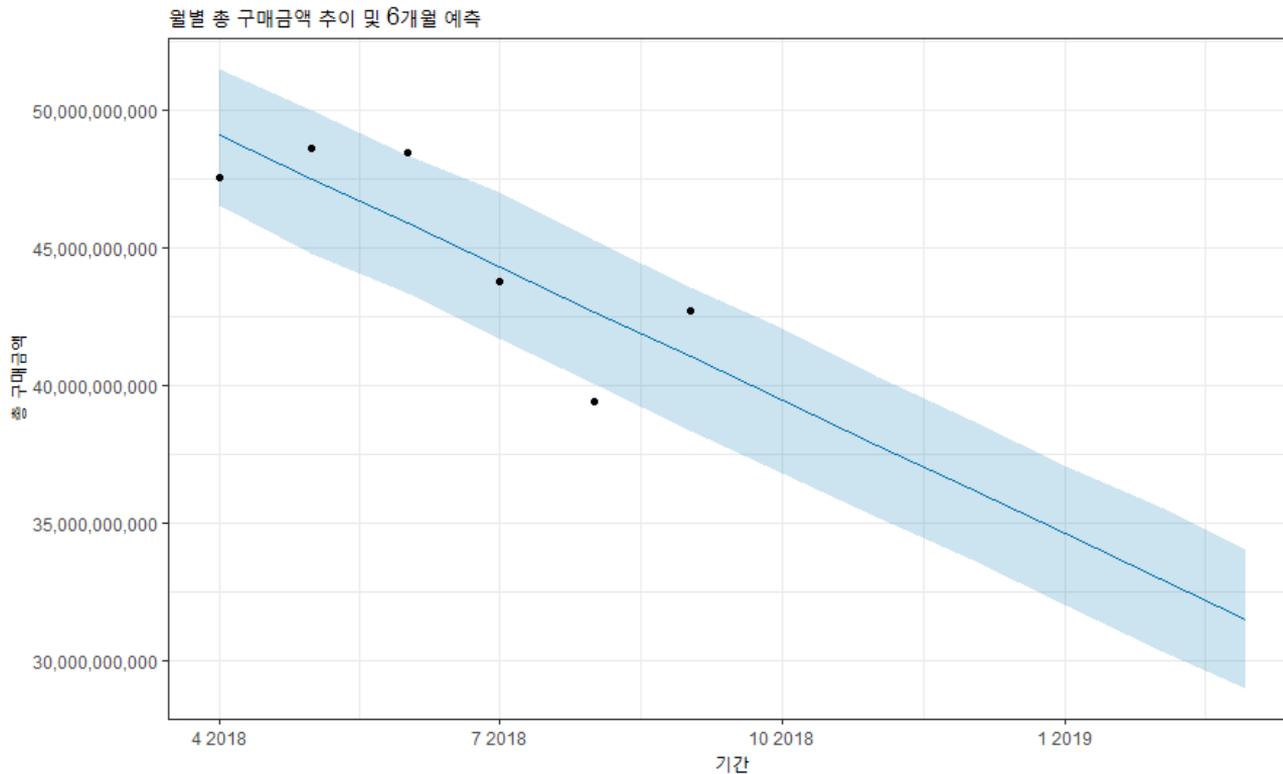
(일, 주, 월)



2. 분석 - 1) 시계열 분석

Prophet 패키지

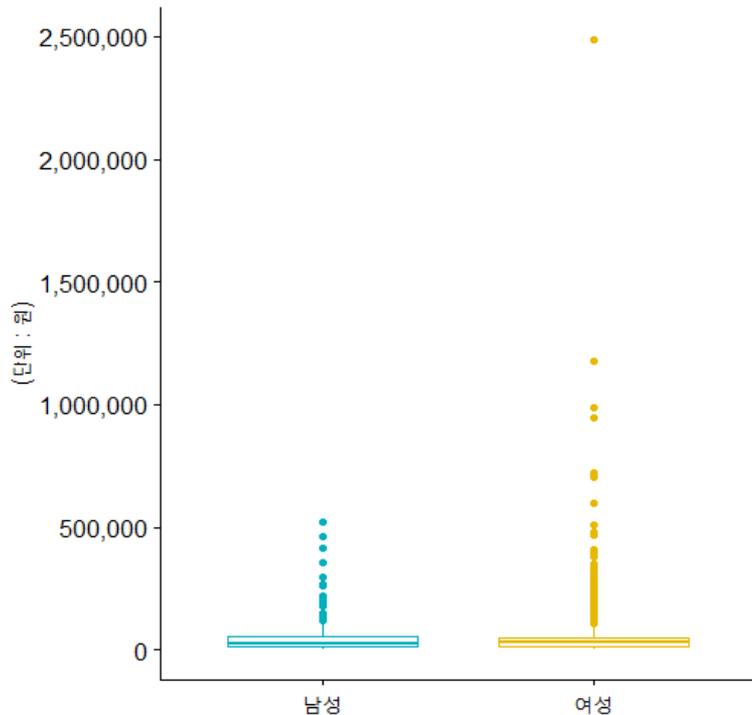
(월별)



2. 분석 - 2) t - test

성별

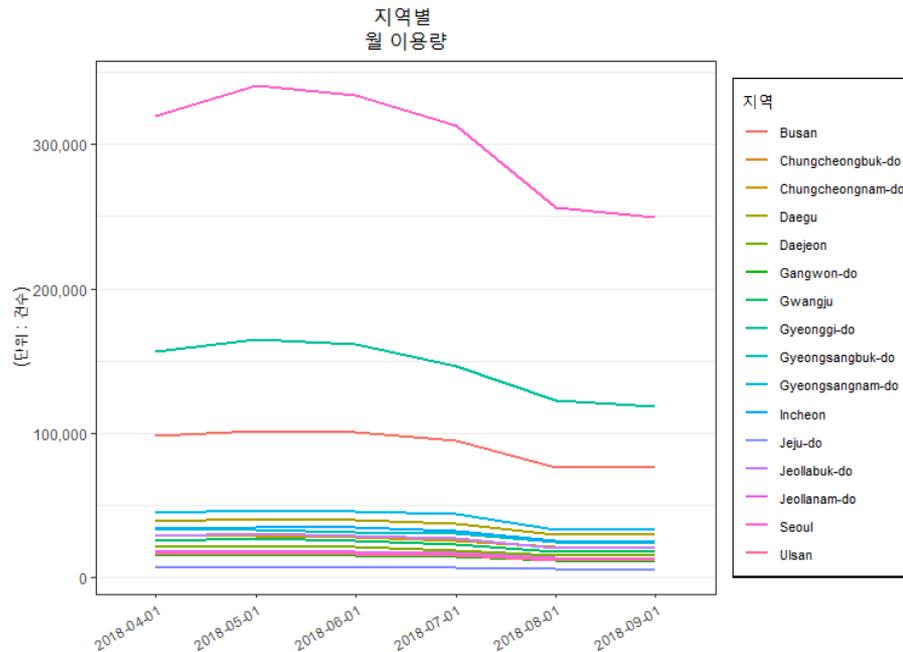
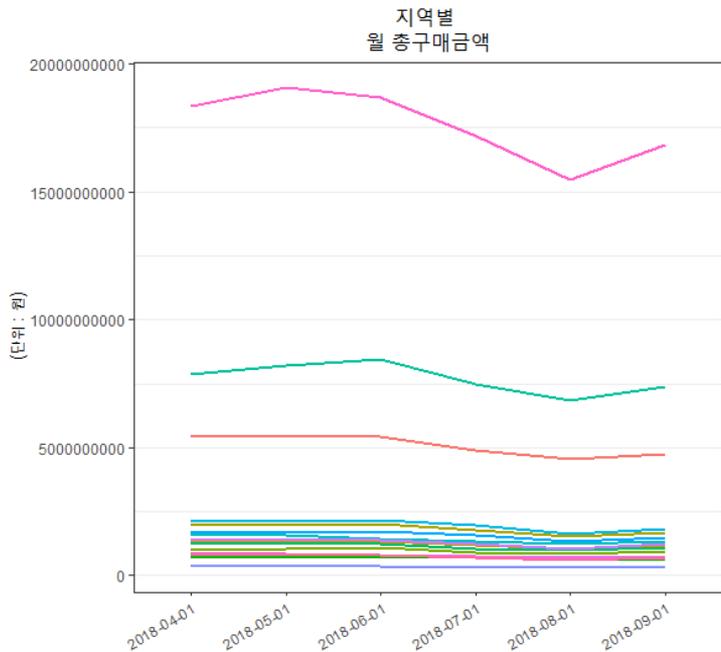
성별 총 구매금액



남성과 여성
t - test로 두 집단간의 차이를
Box-plot을 통해 시각화

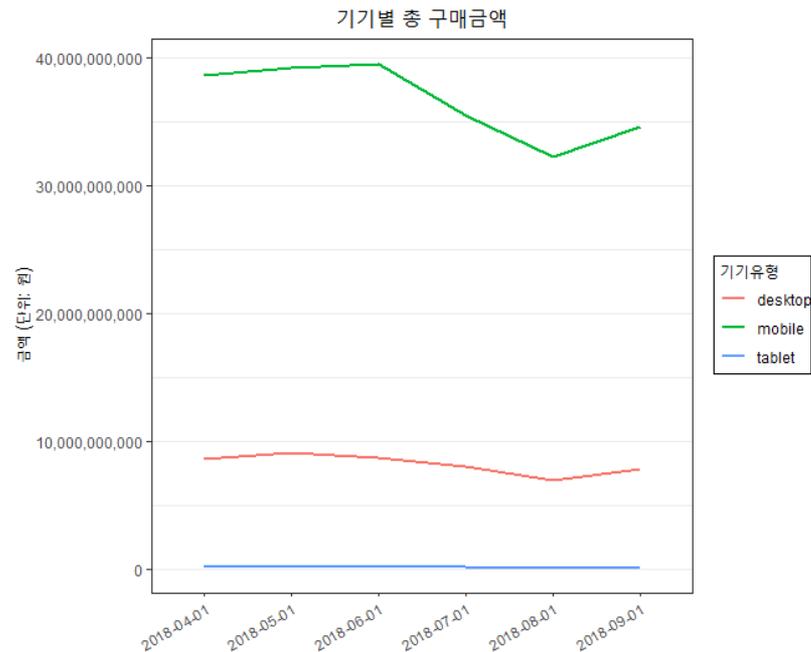
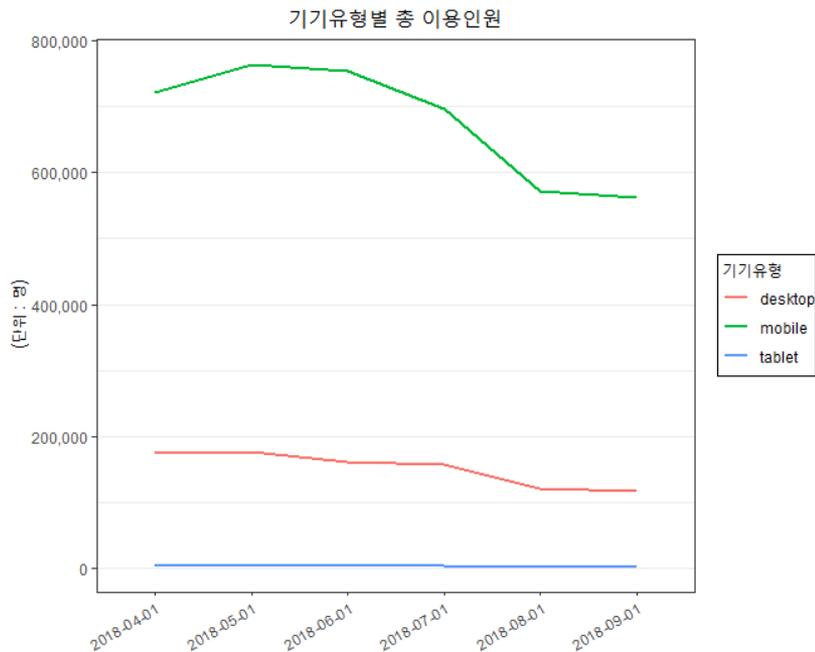
2. 분석 - 3) 분산분석

지역 분류별



2. 분석 - 3) 분산분석

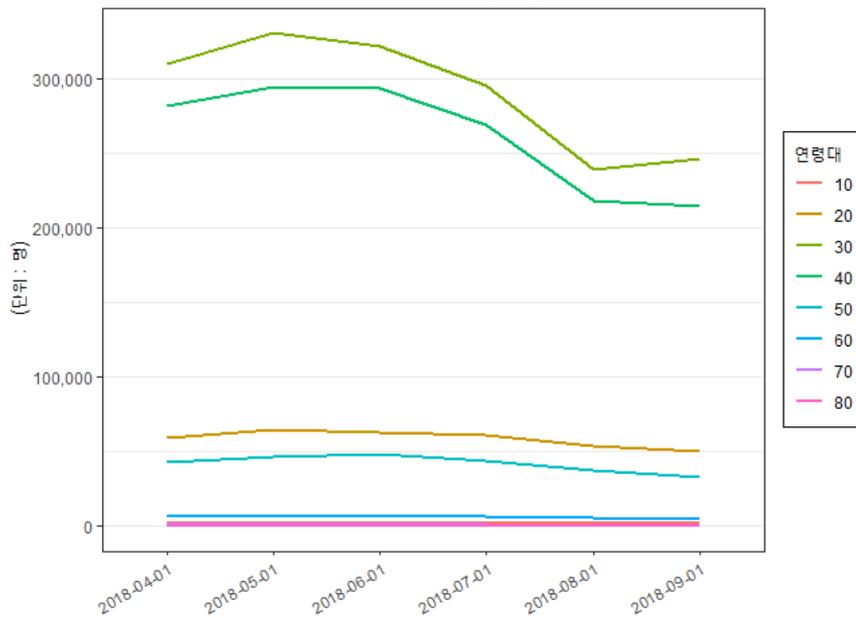
기기 유형별



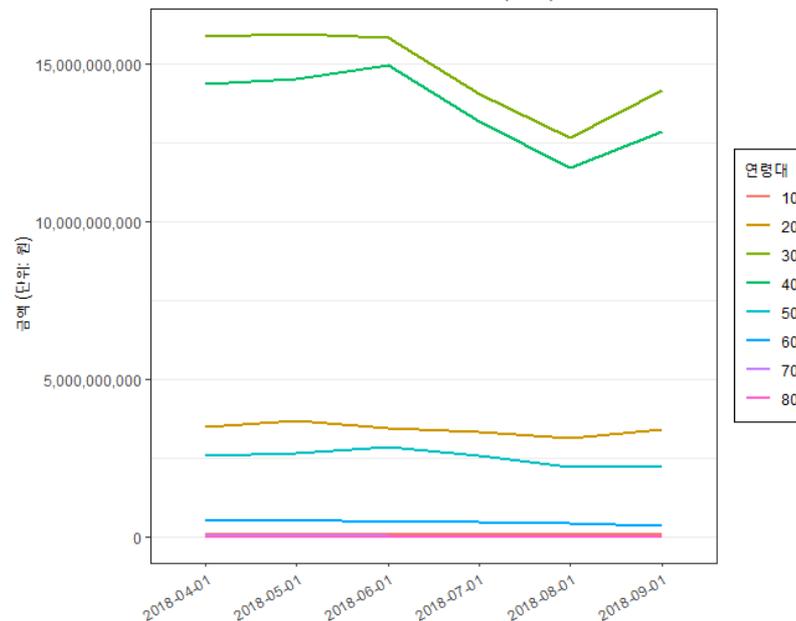
2. 분석 - 3) 분산분석

나이대별

나이대별 이용자(월별)



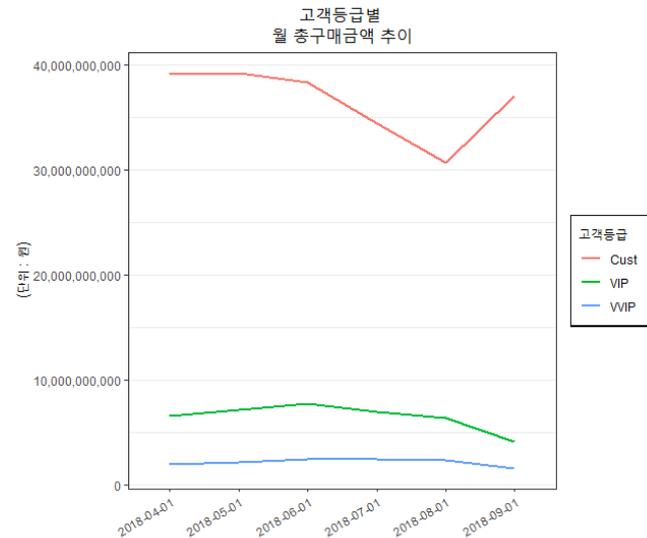
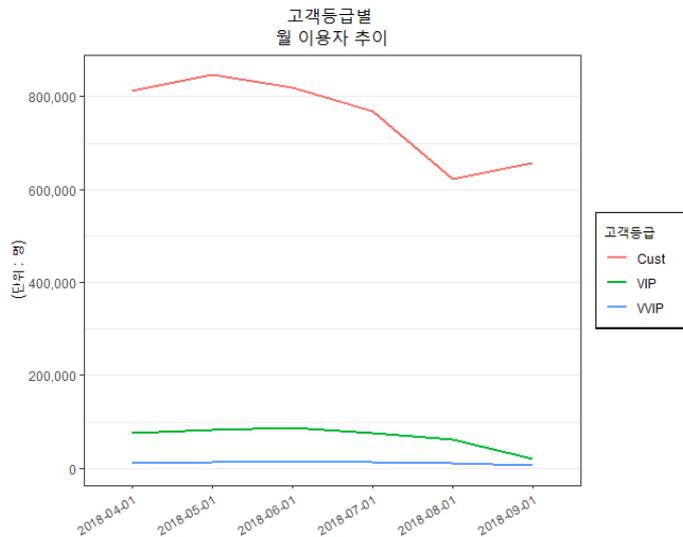
나이대별 총 구매금액(월별)



2. 분석 - 3) 분산분석

고객 등급별

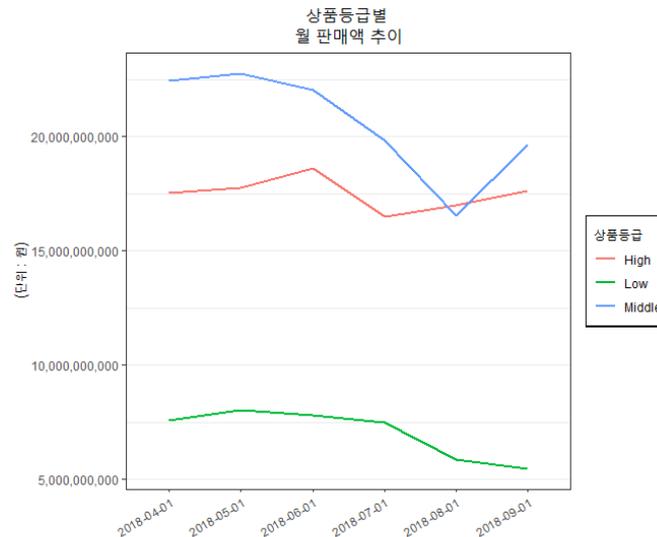
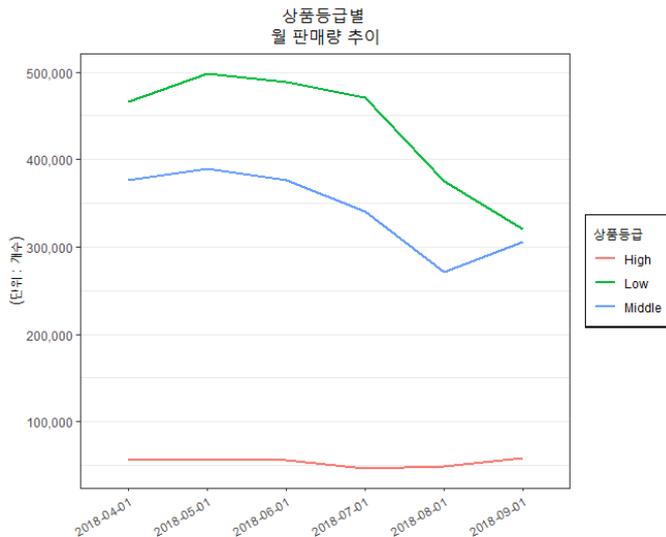
- 롯데 하이마트와 백화점의 분류기준 참고하여 새로운 고객등급 생성
- 총 구매액 기준 VVIP > 10,000,000 / VIP > 2,500,000 / 그 외는 Cust



2. 분석 - 3) 분산분석

상품 등급별

- 전체 상품 단가의 분포와 box-plot의 IQR을 기준으로 새로운 상품등급 생성
- 상품 단가 기준 High > 135,150 / Middle > 29,000 / 그 외는 Low





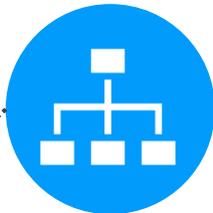
3. 소비자 선호지수

- 1. 소비자 선호지수 식
- 2. 도출과정
- 3. 소비자 선호지수 활용

3. 소비자 선호지수 - 1) 소비자 선호지수 식



상품별 소비자 선호지수
Consumer Preference Index (CP Index)



“상품 데이터”

CLAC3_NM : 소분류명
Count : 소분류별 판매된 빈도 수



“구매 데이터”

PD_BUY_TOT : 총 구매액



“온라인 행동데이터”

HIT_SEQ : 클릭 횟수
TOT_PAG_VIEW_CT : 시청 페이지 갯수
TOT_SESS_HR_V : 세션에 머문 시간

3. 소비자 선호지수 - 1) 소비자 선호지수 식

- 지수의 단위를 낮추기 위해서 전체 선호지수의 평균으로 나눈 값을 사용함

최종 소비자 선호지수 도출식

$$\text{NEW CP INDEX} = \frac{0.04X_1 + 0.4X_2 + X_3 - 0.55X_4}{\text{Mean}(\text{Total CP INDEX}) * 5}$$

X₁: PD_BUY_TOT 총합

X₂: HIT_SEQ 총합

X₃: SESS_HR_V 총합

X₄: PAG_VIEW_CT 총합

3. 소비자 선호지수 - 2) 도출 과정

소분류별 그룹화 & 표준화

STEP 1.

STEP 2.

STEP 3.

STEP 4.

```
mall_01 <- mall_01 %>%
  group_by(CLAC3_NM) %>%
  summarise(sum_PD_BUY_TOT = sum(PD_BUY_TOT),
            sum_PD_BUY_CT = sum(PD_BUY_CT),
            sum_PD_BUY_AM = sum(PD_BUY_AM),
            sum_HITS_SEQ = sum(HITS_SEQ),
            sum_TOT_PAG_VIEW_CT = sum(TOT_PAG_VIEW_CT),
            sum_TOT_SESS_HR_V = sum(TOT_SESS_HR_V),
            count = n())

# 표준화
mall_02 <- mall_01[-1]

mall_02 <- stdize(as.matrix(mall_02))
mall_02 <- as.data.frame(mall_02)
```

1. 소분류(CLAC3_NM) 기준으로 연속형 변수 sum()
2. 소분류별 판매 빈도를 나타내는 새 변수 count 생성
3. 표준화

3. 소비자 선호지수 - 2) 도출 과정

회귀식 적합

STEP 1.

STEP 2.

STEP 3.

STEP 4.

```
fit_01 <- lm(count ~ sum_PD_BUY_TOT +  
              sum_HITS_SEQ + sum_TOT_PAG_VIEW_CT +  
              sum_TOT_SESS_HR_V, data = mall_02)  
summary(fit_01)
```

1. '소비자 선호' 는 소분류별 판매 빈도 'count' 로 정의함
2. 'count' 변수와 가장 관련이 높다고 판단한 변수를 넣어 회귀식 피팅

→ 회귀식을 통해 소비자 선호지수 식의 가중치를 얻기 위함

3. 소비자 선호지수 - 2) 도출 과정

회귀식 적합

STEP 1.

STEP 2.

STEP 3.

STEP 4.

```
Call:
lm(formula = count ~ sum_PD_BUY_TOT + sum_HITS_SEQ + sum_TOT_PAG_VIEW_CT +
    sum_TOT_SESS_HR_V, data = mall_02)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.04990 -0.02140 -0.01727  0.00081  0.82349
```

```
Coefficients:
                Estimate      Std. Error t value      Pr(>|t|)
(Intercept) -0.00000000000001233  0.0031970913142504048    0.000    1.00000
sum_PD_BUY_TOT  0.0501272349530799549  0.0079667055886525635    6.292  0.00000000049 ***
sum_HITS_SEQ    0.4490293336809165758  0.1499479708639510356    2.995    0.00282 **
sum_TOT_PAG_VIEW_CT -0.6271499609954591659  0.1507169135170686458   -4.161  0.00003473159 ***
sum_TOT_SESS_HR_V  1.1339695784996883798  0.0519486396449044679   21.829 < 0.00000000000000002 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.09581 on 893 degrees of freedom
Multiple R-squared:  0.9909,    Adjusted R-squared:  0.9908
F-statistic: 2.421e+04 on 4 and 893 DF,  p-value: < 0.00000000000000022
```

3. 소비자 선호지수 - 2) 도출 과정

가중치 생성

STEP 1.

STEP 2.

STEP 3.

STEP 4.

회귀 계수가 가장 높은
“sum_TOT_SESS_HR_V” 기준 비율,
가중치 생성



- sum_TOT_SESS_HR_V : 1
- sum_PD_BUY_TOT : 0.04
- sum_HIT_SEQ : 0.4
- sum_TOT_PAG_VIEW_CT : - 0.55

“ NEW WEIGHT ”

3. 소비자 선호지수 - 2) 도출 과정

최종 CP INDEX 도출

STEP 1.

STEP 2.

STEP 3.

STEP 4.

$$CP\ INDEX = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

생성한 가중치 적용



최종 소비자 선호지수 도출식

$$NEW\ CP\ INDEX = \frac{0.04X_1 + 0.4X_2 + X_3 - 0.55X_4}{Mean(Total\ CP\ INDEX) * 5}$$

- 지수의 단위를 낮추기 위해서 전체 선호지수의 평균값에 5를 곱하여 나눔

3. 소비자 선호지수 - 3) 소비자 선호지수 활용

최종 소비자 선호지수 도출식

$$\text{NEW CP INDEX} = \frac{0.04X_1 + 0.4X_2 + X_3 - 0.55X_4}{\text{Mean}(\text{Total CP INDEX}) * 5}$$

✓ 선호지수 활용한
선호경향의 정량적 파악

3. 소비자 선호지수 - 3) 소비자 선호지수 활용

- 선호지수를 활용하여 선호경향을 정량적으로 파악 가능

	상품 소분류	선호지수
1	남성티셔츠	843.0677
2	여성원피스	631.4557
3	여성남방셔츠	423.5394
4	여성티셔츠/탑	402.6201
5	스킨케어세트	373.4655
6	남성캐주얼바지	294.4451
7	여성술더백	273.8663
8	여성바지	265.436
9	BB/파운데이션/컴팩트류	212.0384
10	홍삼액	192.8416
11	블러셔/웨이딩/하이라이터	189.6135
12	남성런닝/트레이닝화	187.9776
13	남성스포츠티셔츠	180.9974
14	여성토트백	169.4179
15	여성런닝/트레이닝화	168.8413

[종합 선호지수 상위 15개]

	상품 소분류	선호지수
1	DIY 완구	0.002652
2	부채	0.003209
3	미용거울	0.003251
4	수도용품	0.003415
5	방한모	0.004214
6	닭근위	0.004361
7	과실주병	0.005543
8	냉동디저트	0.005781
9	살구	0.005930
10	발	0.007221
11	석류	0.008001
12	아이젠	0.009162
13	닭아랫날개(wing)	0.010168
14	닭위날개(bone)	0.010289
15	기저귀크림/파우더	0.011046

[종합 선호지수 하위 15개]

3. 소비자 선호지수 - 3) 소비자 선호지수 활용

▪ 월별 선호지수의 변화

	상품 소분류	선호지수
1	남성티셔츠	912.0276
2	여성원피스	504.9956
3	여성티셔츠/탑	406.573
4	여성남방셔츠	376.2952
5	스킨케어세트	292.6114
6	남성런닝/트레이닝화	252.5714
7	여성숄더백	250.3394
8	남성캐주얼바지	237.8968
9	여성재킷	234.4434
10	여성바지	225.942
11	여성런닝/트레이닝화	207.2684
12	영유아티셔츠/탑	202.2192
13	BB/파운데이션/컴팩트류	187.8334
14	여성크로스백	187.0473
15	남성스포츠티셔츠	182.3626

[종합 선호지수 4월]

	상품 소분류	선호지수
1	남성티셔츠	883.689
2	여성원피스	606.1454
3	여성티셔츠/탑	449.7104
4	여성남방셔츠	407.0986
5	남성캐주얼바지	283.7178
6	스킨케어세트	267.4522
7	여성바지	259.6792
8	여성숄더백	256.254
9	블러셔/쉐이딩/하이라이터	233.7494
10	BB/파운데이션/컴팩트류	223.1048
11	여성재킷	202.3618
12	영유아티셔츠/탑	192.4918
13	여아티셔츠/탑	183.487
14	여성토티백	183.1458
15	여성런닝/트레이닝화	173.4316

[종합 선호지수 5월]

	상품 소분류	선호지수
1	여성원피스	723.6912
2	남성티셔츠	583.2434
3	여성남방셔츠	379.8556
4	여성티셔츠/탑	339.6414
5	스킨케어세트	330.685
6	여성숄더백	321.52
7	BB/파운데이션/컴팩트류	255.7514
8	블러셔/쉐이딩/하이라이터	237.9704
9	여성토티백	235.862
10	남성캐주얼바지	227.3988
11	여성바지	221.5852
12	여성코트	174.2953
13	남성런닝/트레이닝화	171.1338
14	여성속옷세트	158.495
15	여성스커트	150.5106

[종합 선호지수 8월]

3. 소비자 선호지수 - 3) 소비자 선호지수 활용

▪ 매출에서 가장 높은 비율을 차지하던 30대/여성/서울 고객의 월별 선호도 비교

	상품 소분류	선호지수
1	여성원피스	625.3804
2	남성티셔츠	613.3656
3	여성남방셔츠	354.9906
4	스킨케어세트	335.1474
5	여성티셔츠/탑	322.8186
6	BB/파운데이션/컴팩트류	235.5472
7	여성재킷	233.554
8	여성숄더백	230.2448
9	여성런닝/트레이닝화	223.8606
10	여성바지	216.4486
11	블러셔/쉐이딩/하이라이터	209.9792
12	여성스커트	203.3288
13	영유아티셔츠/탑	195.8908
14	유아동내의	195.6481
15	남성런닝/트레이닝화	185.3199

[서울 30대 여성 종합 선호지수 4월]

	상품 소분류	선호지수
1	여성원피스	660.0904
2	남성티셔츠	630.9488
3	여성티셔츠/탑	359.7842
4	여성남방셔츠	351.8304
5	스킨케어세트	314.9682
6	BB/파운데이션/컴팩트류	293.2614
7	블러셔/쉐이딩/하이라이터	285.0878
8	여성바지	222.2132
9	유아동내의	213.6864
10	영유아티셔츠/탑	210.7594
11	여성숄더백	209.3876
12	여성재킷	202.6288
13	여성샌들	199.6467
14	여성스커트	192.0258
15	여아티셔츠/탑	175.0346

[서울 30대 여성 종합 선호지수 5월]

	상품 소분류	선호지수
1	여성원피스	752.042
2	남성티셔츠	401.078
3	스킨케어세트	382.1358
4	여성숄더백	327.9388
5	여성남방셔츠	323.8214
6	BB/파운데이션/컴팩트류	323.2856
7	블러셔/쉐이딩/하이라이터	311.5686
8	여성티셔츠/탑	254.5046
9	여성바지	189.049
10	메이크업세트	170.6261
11	여성스커트	168.2228
12	남성캐주얼바지	164.8835
13	여성코트	158.3282
14	여성토트백	147.0108
15	여성속옷세트	134.665

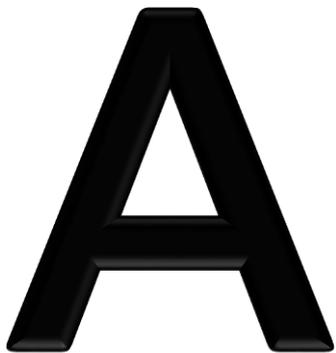
[서울 30대 여성 종합 선호지수 8월]



4. 서비스 제안

- 1. 역시즌 프로모션
- 2. 역타겟팅

4. 서비스제안



Turn A 전략

All 을 뒤집어
완전히 다른,
모두를 위한,
새로운 전략

4. 서비스제안 - 1) 역시즌 프로모션

- 5월 ~ 8월 선호지수 상승한 상품 상위 15개

	상품 소분류	5월 선호지수	8월 선호지수	선호지수 격차
1	여성숄더백	209.3876	327.9388	118.5512
2	여성코트	41.41444	158.3282	116.9137
3	남성등산패딩	20.90496	123.1796	102.2746
4	여성원피스	660.0904	752.042	91.95166
5	남성수영복	29.33062	102.7651	73.43444
6	스킨케어세트	314.9682	382.1358	67.16758
7	여성패딩	16.36869	80.22886	63.86018
8	기타일반문구/사무용품	44.92548	106.5824	61.6569
9	핸디형청소기	36.88898	92.59248	55.7035
10	여성비치웨어	41.70092	96.93474	55.2338
11	여성등산패딩	4.356172	55.31534	50.95916
12	아이새도우	56.6371	106.6968	50.05972
13	여성토트백	98.8598	147.0108	48.15102
14	멀티형에어컨	50.59842	92.73814	42.13972
15	메이크업세트	128.5712	170.6261	42.0549

- 5월 ~ 8월 선호지수가 상승한 상품 상위 15개
- 선호지수 격차 = (8월 선호지수 - 5월 선호지수)

4. 서비스제안 - 1) 역시즌 프로모션

- 5월 ~ 8월 선호지수 상승한 상품 상위 15개

	상품 소분류	5월 선호지수	8월 선호지수	선호지수 격차
1	여성숄더백	209.3876	327.9388	118.5512
2	여성코트	41.41444	158.3282	116.9137
3	남성등산패딩	20.90496	123.1796	102.2746
4	여성원피스	660.0904	752.042	91.95166
5	남성수영복	29.33062	102.7651	73.43444
6	스킨케어세트	314.9682	382.1358	67.16758
7	여성패딩	16.36869	80.22886	63.86018
8	기타일반문구/사무용품	44.92548	106.5824	61.6569
9	핸디형청소기	36.88898	92.59248	55.7035
10	여성비치웨어	41.70092	96.93474	55.2338
11	여성등산패딩	4.356172	55.31534	50.95916
12	아이새도우	56.6371	106.6968	50.05972
13	여성토트백	98.8598	147.0108	48.15102
14	멀티형에어컨	50.59842	92.73814	42.13972
15	메이크업세트	128.5712	170.6261	42.0549

- 5월 ~ 8월 선호지수가 상승한 상품 상위 15개
- 선호지수 격차 = (8월 선호지수 - 5월 선호지수)
- 5월 대비 여름인 8월에 겨울 상품인 패딩의 선호지수 증가폭이 큰 것으로 나타남

4. 서비스제안 - 1) 역시즌 프로모션



4. 서비스제안 - 1) 역시즌 프로모션



역시즌 프로모션 제안

역시즌 겨울 상품 특가

시즌되면
늦으리

여름에만 볼 수 있는 가격

FW 역시즌 & 시즌오프 상품을 가장 저렴하게!



9월 본시즌에서는 만날 수 없는 핫 PRICE
선물같은 시원한 특가로, 역시즌 상품을 지금 득템하세요~!

4. 서비스제안 - 1) 역시즌 프로모션



4. 서비스제안 - 2) NEW 타겟팅

[30대/여성/서울 고객의 종합 선호지수 상위 30개]

	상품 소분류	선호지수		상품 소분류	선호지수
1	여성원피스	771.2878	16	여성런닝/트레이닝화	156.2127
2	남성티셔츠	599.4832	17	메이크업세트	143.5675
3	여성남방셔츠	382.4264	18	여성재킷	141.5689
4	스킨케어세트	369.0396	19	여아티셔츠/탑	140.0369
5	여성티셔츠/탑	333.5052	20	여성토트백	139.6153
6	BB/파운데이션/کم팩트류	288.285	21	남성런닝/트레이닝화	126.1778
7	여성술더백	263.473	22	여성속옷세트	124.5697
8	블러셔/셰이딩/하이라이터	261.0342	23	유아동스포츠샌들/슬리퍼	121.0978
9	여성바지	232.2796	24	크림/밤/오일	117.5315
10	여성스커트	210.4632	25	남아바지	117.3614
11	여성샌들	204.6992	26	여성코트	115.6114
12	남성캐주얼바지	194.4412	27	여성스니커즈	113.4587
13	영유아티셔츠/탑	185.7149	28	아동수영복	109.7027
14	유아동내의	171.7973	29	남성스포츠샌들/슬리퍼	105.6703
15	홍삼액	159.9542	30	남성스포츠티셔츠	104.0394

- 30대 여성의 종합 선호지수 상위 30개에서 남성용품이 상위권에 위치해 있음

4. 서비스제안 - 2) 역타겟팅

특정 남성제품의 실 구매자는 “여성”인 경우가 많다는 것



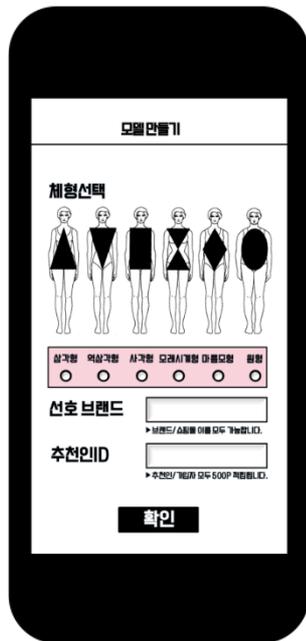
여성에게 선호도가 높은 남성제품은 “여성” 으로

“역타겟팅” 마케팅 전략 필요

4. 서비스제안 - 2) 역타겟팅



< 모델 만들기 화면1 >



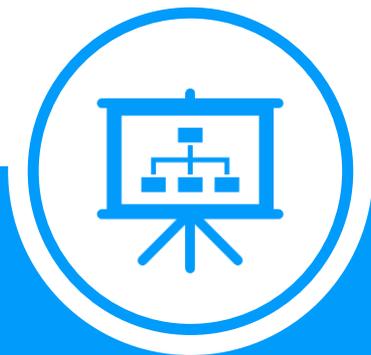
< 모델 만들기 화면2 >



< 모델 생성 화면 >



< 옷입히기 화면1 >



5. 결론

5. 결론

LOTTE 분석의 결론?!

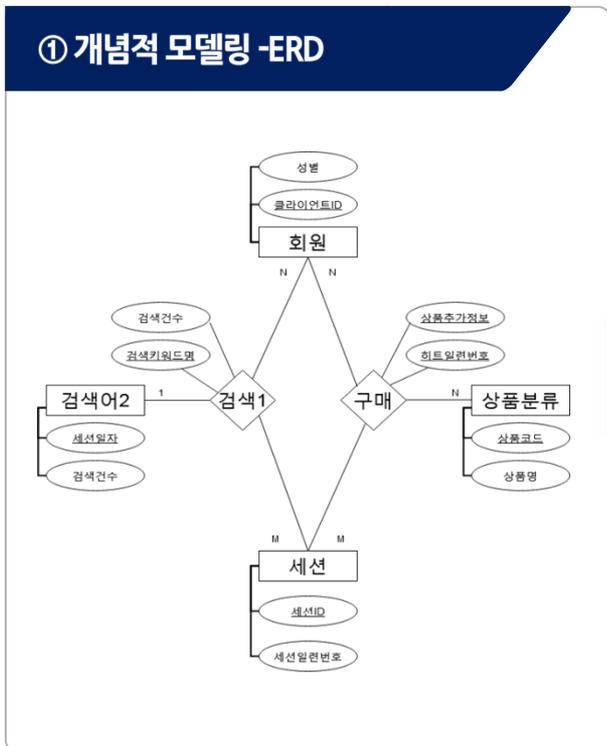




6. 부록

6. 부록 - 데이터베이스 모델링

- 훈련기간동안 수행한 데이터베이스 모델링 절차를 적용해 테이블간 관계를 추출함



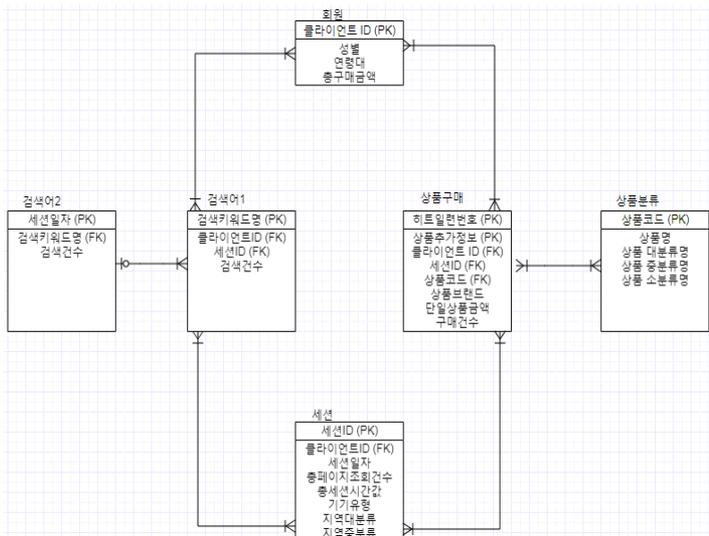
② 논리적 모델링 -Mapping

단계	타입	구분	사상
1단계	개체타입	강한 개체타입	회원(클라이언트ID(PK), 성별, 연령대) 상품분류(상품코드(PK), 상품명) 검색어2(세션일자(PK), 검색어키워드명) 세션(세션ID(PK), 세션일련번호)
2단계		약한 개체타입	없음
3단계		이진 1:1 관계 타입	없음
4단계	관계타입	이진 1:N 관계 타입	회원(클라이언트ID(PK), 성별, 연령대) 상품분류(상품코드(PK), 상품명) 회원(클라이언트ID(PK), 성별, 연령대) 검색어2(세션일자(PK), 검색어키워드명) 회원(클라이언트ID(PK), 성별, 연령대) 세션(세션ID(PK), 세션일련번호)
5단계		이진 N:M 관계 타입	검색어1(클라이언트ID(PK), 세션ID(PK), 검색어키워드명) 회원(클라이언트ID(PK), 성별, 연령대) 세션(세션ID(PK), 세션일련번호) 구매(클라이언트ID(PK), 세션ID(PK), 히트일련번호)
6단계	관계타입	N진 관계 타입	회원(클라이언트ID(PK), 성별, 연령대) 상품분류(상품코드(PK), 상품명) 세션(세션ID(PK), 세션일련번호) 구매(클라이언트ID(PK), 세션ID(PK), 상품코드(PK), 상품브랜드) 회원(클라이언트ID(PK), 성별, 연령대) 상품분류(상품코드(PK), 상품명) 검색어2(세션일자(PK), 검색어키워드명) 검색어1(클라이언트ID(PK), 세션ID(PK), 세션일자(PK), 검색어키워드명)
7단계		속성	다중 값 속성

6. 부록 - 데이터베이스 모델링

- 훈련기간동안 수행한 데이터베이스 모델링 절차를 적용해 테이블간 관계를 추출함

③ 물리적 모델링



```

CREATE TABLE CUSTOMER(
  CLNT_ID CHAR PRIMARY KEY(),          -VARCHAR2(10)
  CLNT_GENDER CHAR(),                 -VARCHAR2(1)
  CLNT_AGE INT
);
    
```

```

CREATE TABLE SEARCH1(
  CLNT_ID CHAR PRIMARY KEY(),          -VARCHAR2(10)
  SESS_ID CHAR PRIMARY KEY(),         -VARCHAR2(10)
  KWVD_NM CHAR,                       -VARCHAR2(1000)
  SEARCH_CNT INT
);
    
```

```

CREATE TABLE PRODUCT(
  CLNT_ID CHAR PRIMARY KEY(),          -VARCHAR2(10)
  SESS_ID CHAR PRIMARY KEY(),         -VARCHAR2(10)
  HITS_SEQ INT PRIMARY KEY,
  PD_C CHAR PRIMARY KEY(),            -VARCHAR2(1000)
  PD_ADD_NM CHAR PRIMARY KEY(),       -VARCHAR2(1000)
  PD_BRA_NM CHAR,                     -VARCHAR2(1000)
  PD_BUY_AM INT,
  PD_BUY_CT INT,
);
    
```

```

CREATE TABLE SEARCH2(
  SESS_DT DATE PRIMARY KEY,           -VARCHAR2(1000)
  KWVD_NM CHAR PRIMARY KEY(),         -VARCHAR2(1000)
  SEARCH_CNT INT
);
    
```

```

CREATE TABLE MASTER(
  PD_C CHAR PRIMARY KEY(),            -VARCHAR2(1000)
  PD_NM CHAR(),                       -VARCHAR2(1000)
  CLAC1_NM CHAR(),                    -VARCHAR2(1000)
  CLAC2_NM CHAR(),                    -VARCHAR2(1000)
  CLAC3_NM CHAR(),                    -VARCHAR2(1000)
);
    
```

```

CREATE TABLE SESSION(
  CLNT_ID CHAR PRIMARY KEY(),         -VARCHAR2(10)
  SESS_ID CHAR PRIMARY KEY(),         -VARCHAR2(10)
  SESS_SEQ INT,
  SESS_DT DATE,
  TOT_PAG_VIEW_CT INT,
  TOT_SESS_HR_V INT,
  DVC_CTG_NM CHAR(),                 -VARCHAR2(1000)
  ZON_NM CHAR(),                     -VARCHAR2(1000)
  CITY_NM CHAR(),                     -VARCHAR2(1000)
);
    
```


6. 부록 - 데이터 전처리

구분	Data frame		Cleansing Rule	전체 건수	클린징 건수	클린징비율	분석대상 건수	
이행 데이터	회원 (Custom)	CLNT_ID	클라이언트ID	CLNT_ID IS NULL	671,679	-		
		CLNT_GENDER	성별	CLNT_AGE IS NULL	671,679	-		
		CLNT_AGE	연령대	CLNT_GENDER IS NOT IN ('M','F')	671,679	-	671,679	
	세션 (Session)	CLNT_ID	클라이언트id	CUST원장에 CLIENT_ID가 없음	2,698,696	477,574	17.60%	
		SESS_ID	세션ID	PRODUCT원장에 SESS_ID가 없음	2,221,122	1,043	0.04%	2,220,079
		SESS_SEQ	세션일련번호					
		SESS_DT	세션일자					
		TOT_PAG_VIEW_CT	총 페이지 조회건수	TOT_PAG_VIEW_CT IS NULL TOT_PAG_VIEW_CT = 0	2,712,907 2,698,696	274 -	0.01%	2,698,696
		TOT_SESS_HR_V	총세션시간값	TOT_SESS_HR_V IS NULL TOT_SESS_HR_V = 0	2,712,633 2,698,696	3,937 -	0.51%	2,698,696
		DVC_CTG_NM	기기유형					
ZON_NM	지역대분류							
CITY_NM	지역중분류							
제품	상품분류 (Master)	PD_C	상품코드	PD_C IS NULL	847,652	-		
		PD_NM	상품명	PD_NM IS NULL	847,652	-		
		CLAC1_NM	상품 대분류명	CLAC1_NM IS NULL	847,652	-		
		CLAC2_NM	상품 중분류명	CLAC2_NM IS NULL	847,652	-		
		CLAC3_NM	상품 소분류명	CLAC3_NM IS NULL	847,652	-	847,652	

Q & A

감사합니다!

