



제 5회 L.POINT COMPETITION

About Us

전 관 우

김 동 섭

정 현 용

L.POINT

CONTENTS

01

INTRODUCTION

목적 및 분석배경
기대효과

02

DATA
INTRODUCTION

데이터 전처리

03

DATA
ANALYSIS

데이터 분석

04

CONCLUSION

결과정리
활용

01

INTRODUCTION

목적 및 분석배경
기대효과

“ 온라인 행동 기반 트렌드 예측 ”

효율적 마케팅 가능 / 고객 충성도 상승 / 신규 고객 유입 증가 유도

선호 지수 분석
상품군별 수요 트렌드 예측 및 인사이트 도출
신규 서비스 제안



<고객>



<빅데이터>



<트렌드>



<예측>

02

DATA
INTRODUCTION

데이터 전처리

DATA INTRODUCTION

데이터 소개



데이터 내용

상품구매, 검색어, 회원, 세션, 상품분류
등 고객유형에 따른 다양한 변수 존재

총 5,024,906 개의 구매목록과
23개의 데이터 데이터 테이블



분석 프로세스

R을 이용한 데이터 분석

DATA PRETREATMENT

데이터 전처리

- ✓ 지출분야를 명확히 하기 위해 장르를 축소 및 통합
- ✓ CLNT_ID 별 소비기록데이터와 고객정보를 알기 위해 데이터를 통합 과정 필요
기존에 제공하는 자료 PRODUCT, MASTER, CUSTOMER, SESSION 파일을 통합
- ✓ 결측값 (NA) 제거 후 변수들의 데이터 형식 변경
- ✓ 파일 통합 및 전처리 완료 후의 데이터



지출분야를 명확히 하기 장르를 축소 및 통합

수정 전	수정 후
시즌스포츠, 아웃도어/레저, 헬스/피트니스, 구기/필드스포츠	스포츠 용품
계절가전, 냉장/세탁가전, 생활/주방가전, 영상/음향가전, 컴퓨터, 모바일	전자 기기
식기/조리기구, 주방잡화, 출산/육아용품, 패션잡화, 퍼스널케어, 청소/세탁/욕실용품, 세제/위생, 화장품/뷰티케어	잡화
건강식품, 과일, 냉동식품, 냉장식품, 음료, 축산물	식품
인테리어/조명, 침구/수예, 가구	가정
자동차용품, 원예/애완, 문구/사무용품, 완구, 상품권	기타

CLAC1_NM

이 외의 항목들은 변동사항 X



CLNT_ID 별 소비기록데이터와 고객정보를 알기 위해 데이터를 통합 과정 필요
기존에 제공하는 자료 PRODUCT, MASTER, CUSTOMER, SESSION 파일을 통합



Product

CLNT_ID 별 소비기록데이터와
고객정보를 알기 위해 데이터를 통합



Master

CLNT_ID별 상품군별 소비기록데이터를
알기 위해 데이터를 통합



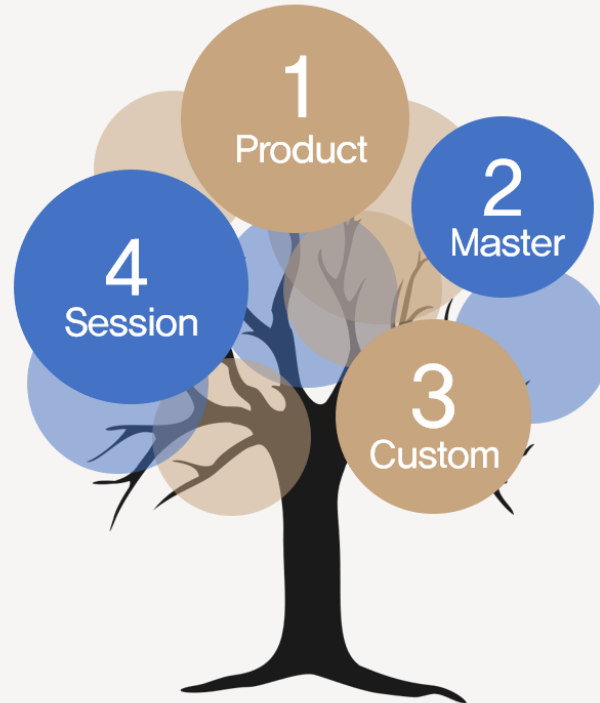
Session

CLNT_ID별 고객들의 거주지역,
구매날짜, 구매에 사용한
기기유형 등을 알기 위해 데이터를 통합



Custom

CLNT_ID별 고객들의 성별, 연령대를
알기 위해 데이터를 통합





CLNT_ID 별 소비기록데이터와 고객정보를 알기 위해 데이터를 통합 과정 필요
기존에 제공하는 자료 PRODUCT, MASTER , CUSTOMER, SESSION 파일을 통합



Product



Custom



통합

▶ product	5024906 obs. of 8 variables
▶ procust	3988587 obs. of 10 variables

Custom에 기록이 없는 비회원 고객 삭제

1,036,319개의 관찰값 삭제



결측값 (NA) 제거 후 변수들의 데이터 형식 변경

SESS_SEQ	SESS_DT	TOT_PAG_VIEW_CT	TOT_SESS_HR_V
47	2018-09-14	1	9
57	2018-09-26	1	9
253	2018-09-13	1	9
27	2018-09-12	1	9
625	2018-07-10	1	9
625	2018-07-10	1	9
3	2018-08-27	1	9
10	2018-06-05	1	9
23	2018-07-23	NA	
10	2018-05-11	NA	
35	2018-09-14	NA	
14	2018-04-30	NA	
3	2018-06-15	NA	
14	2018-06-03	NA	
137	2018-05-15	NA	
14	2018-06-03	NA	
123	2018-04-01	NA	

결측값 (NA) 제거

▶ proessscustma...	3988587 obs. of 21 variables
▶ proessscustma...	3988258 obs. of 21 variables

329개의 결측값 제거

```
> class(proessscustmaster$PD_BUY_AM)
[1] "character"
> class(proessscustmaster$PD_BUY_CT)
[1] "character"
> class(proessscustmaster$SESS_DT)
[1] "integer"
> class(proessscustmaster$CLNT_GENDER)
[1] "character"
> class(proessscustmaster$CLNT_AGE)
[1] "integer"
```



```
> class(proessscustmaster$PD_BUY_AM)
[1] "integer"
> class(proessscustmaster$PD_BUY_CT)
[1] "integer"
> class(proessscustmaster$SESS_DT)
[1] "Date"
> class(proessscustmaster$CLNT_GENDER)
[1] "factor"
> class(proessscustmaster$CLNT_AGE)
[1] "factor"
```

데이터 형식을 변경



파일 통합 및 전처리 완료 후의 데이터

prosesscustmaster	3988256 obs. of 21 variables
CLNT_ID	: int 4140076 4144914 4147643 4148106 4148129 4148527 4154576 4157355 4159738 4160219 ...
SESS_ID	: int 10189797 7935714 4050743 6537318 3942102 5718021 4253415 6349580 3278944 9232020 ...
HITS_SEQ	: int 13 12 66 21 2 19 19 34 42 32 ...
PD_C	: int 570603 554336 149995 570603 755148 608620 591229 554700 184776 368706 ...
PD_ADD_NM	: chr "색상:BLK0_(BLK0)BLACK 사이즈:120 / 2개" "색상:블랙 사이즈:160cm(12~13세) / 1개" "사이즈:옵션04.AN868_04
PD_BRA_NM	: chr "데상트" "아디다스 키즈" "앵글로매니아" "데상트" ...
PD_BUY_AM	: int 39000 39000 39000 39000 39000 39000 39000 39000 39000 39000 ...
PD_BUY_CT	: int 2 1 1 1 1 1 1 1 1 1 ...
CLNT_GENDER	: chr "F" "M" "F" "F" ...
CLNT_AGE	: int 40 40 30 30 20 30 40 30 40 30 ...
PD_NM	: chr "[데상트] 스포츠베이직 빅로고 티셔츠 - DMGR_DMELANGE / 115" "YG GU 3/4 TGT.A(BQ2873) - 블랙 / 110cm(6세)" "
CLAC1_NM	: chr "스포츠패션" "유아동의류" "잡화" "스포츠패션" ...
CLAC2_NM	: chr "남성일반스포츠의류" "여아의류하의" "안경/선글라스" "남성일반스포츠의류" ...
CLAC3_NM	: chr "남성스포츠티셔츠" "여아레깅스" "여성선글라스" "남성스포츠티셔츠" ...
SESS_SEQ	: int 172 24 18 58 3 16 410 218 488 105 ...
SESS_DT	: Date, format: "2018-04-13" "2018-05-21" "2018-07-24" "2018-06-13" ...
TOT_PAG_VIEW_CT	: int 18 13 73 20 1 28 62 33 145 94 ...
TOT_SESS_HR_V	: chr "113" "107" "3,850" "144" ...
DVC_CTG_NM	: chr "mobile" "mobile" "mobile" "mobile" ...
ZON_NM	: chr "Gyeonggi-do" "Seoul" "Seoul" "Gwangju" ...
CITY_NM	: chr "Icheon-si" "Seoul" "Seoul" "Gwangju" ...

METHOD OF ANALYSIS

분석 방법



Frequency analysis

연령대 별 인당 평균 사용 금액,
누적 구매 빈도를 분석

날짜별로 총 누적 구매 빈도를 분석



ANOVA

분산분석을 이용해 집단 간 어디에서
차이가 발생하는지 확인

Tukey 검정을 이용하여
모든 가능한 두 수준들의 평균 간의
차이가 있는지를 검정



Regression analysis

외부 데이터 기온을 이용
기간별로 소비자들의 구매액의 변화

구매빈도 변화의 원인을 파악하기 위해
예상되는 요인들과 구매액과의 연관성 파악

03

DATA
ANALYSIS

데이터 분석

Create Online Preference Index

온라인 선호 지수 생성

연령대에 초점을 두어, 상품 구매 빈도수와
평균 소비 금액을 분석



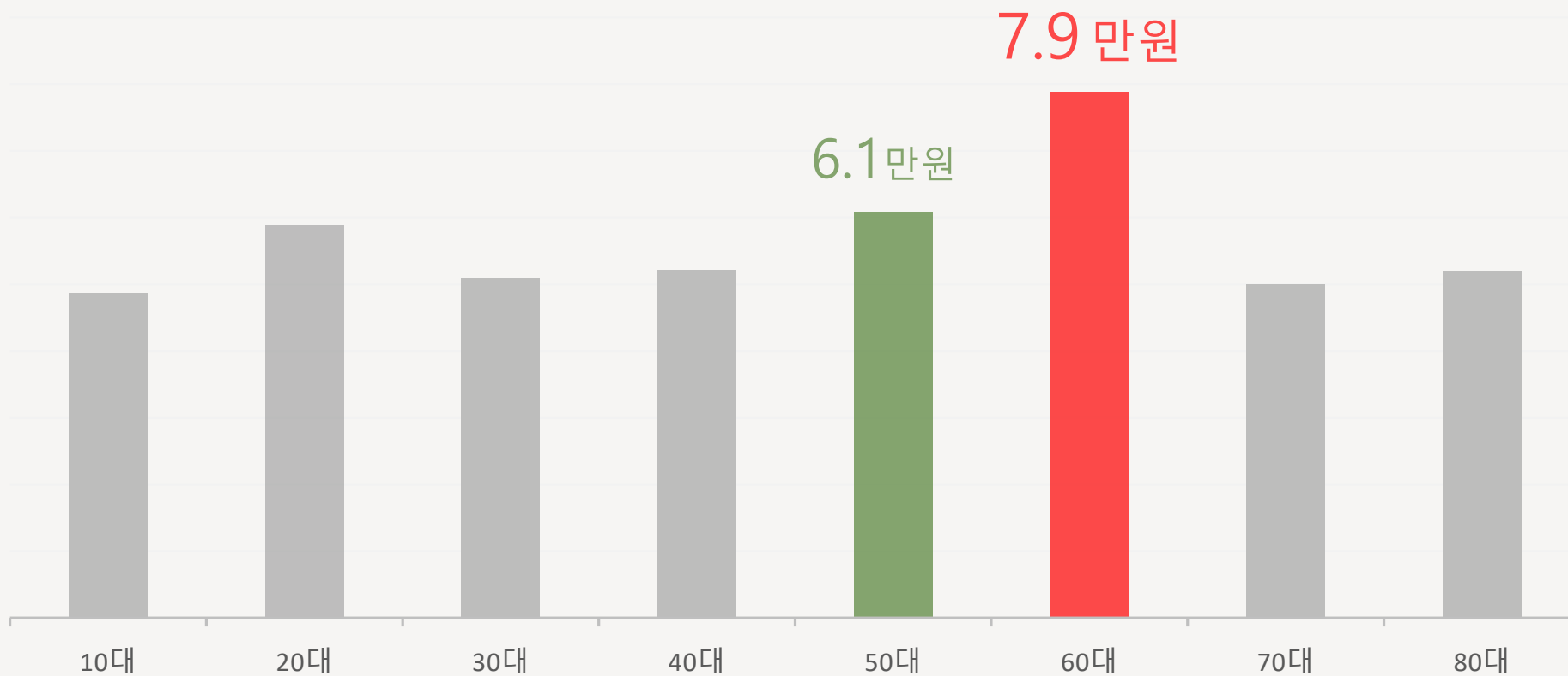
All purchase Frequency

연령별 상품 구매 빈도



Average Usage Amount

연령별 평균 사용 금액



ANOVA

분산분석

연령별 평균 사용 금액에
차이가 유의한지
분산분석과
TUKEY 검증을 사용한다.



Analysis of Variance

분산분석

```
> head(tdata2)
  age price count buyprice
1  40 39000     2   78000
2  40 39000     1   39000
3  30 39000     1   39000
4  30 39000     1   39000
5  20 39000     1   39000
6  30 39000     1   39000
```

일원분산분석

price(PD_BUY_AM) 와
count(PD_BUY_CT) 를 곱해서
새로운 변수 buyprice를 생성

```
> aov(buyprice~age,data=tdata2)
Call:
aov(formula = buyprice ~ age, data = tdata2)
```

```
Terms:
              age      Residuals
Sum of Squares 6.161536e+13 5.547666e+16
Deg. of Freedom              7      3988248

Residual standard error: 117940.8
Estimated effects may be unbalanced
```

```
> summary(aov(buyprice~age,data=tdata2))
              Df      Sum Sq   Mean Sq F value Pr(>F)
age              7 6.162e+13 8.802e+12   632.8 <2e-16 ***
Residuals    3988248 5.548e+16 1.391e+10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

0.05 이하

연령별 평균 구매금액에 차이가 존재한다.

Tukey의 HSD 검정

튜키검정

\$age	diff	lwr	upr	p adj
20-10	10178.6685	6879.7400	13477.597	0.0000000
30-10	2204.8870	-1050.1473	5459.921	0.4458554
40-10	3371.2275	114.9378	6627.517	0.0362699
50-10	12052.3638	8731.5408	15373.187	0.0000000
60-10	30118.0651	26383.4849	33852.645	0.0000000
70-10	1345.6839	-4474.6899	7166.058	0.9969977
80-10	3260.4703	-19357.8819	25878.823	0.9998638
30-20	-7973.7815	-8632.2551	-7315.308	0.0000000
40-20	-6807.4410	-7472.0921	-6142.790	0.0000000
50-20	1873.6953	942.9945	2804.396	0.0000000
60-20	19939.3966	17993.7789	21885.014	0.0000000
70-20	-8832.9846	-13702.7987	-3963.170	0.0000011
80-20	-6918.1982	-29310.7906	15474.394	0.9825406
40-30	1166.3404	773.8092	1558.872	0.0000000
50-30	9847.4768	9086.8692	10608.084	0.0000000
60-30	27913.1781	26042.9513	29783.405	0.0000000
70-30	-859.2031	-5699.3900	3980.984	0.9994551
80-30	1055.5832	-21330.5847	23441.751	0.9999999
50-40	8681.1363	7915.1747	9447.098	0.0000000
60-40	26746.8376	24874.4270	28619.248	0.0000000
70-40	-2025.5435	-6866.5747	2815.488	0.9106342
80-40	-110.7572	-22497.1077	22275.593	1.0000000
60-50	18065.7013	16083.1866	20048.216	0.0000000
70-50	-10706.6799	-15591.3524	-5822.007	0.0000000
80-50	-8791.8935	-31187.7220	13603.935	0.9350874
70-60	-28772.3812	-33947.2484	-23597.514	0.0000000
80-60	-26857.5948	-49318.5021	-4396.688	0.0070263
80-70	1914.7864	-20985.4761	24815.049	0.9999967

도시화

10 < 70 < 30 < 80 < 40 < 20 < 50 < 60

5060대는 모든 연령대와 차이가 존재하며,
10,30,70,80대는 서로 차이가 존재하지 않는다.

각 연령대별로 평균 구매금액 평균 차이를
검정한 결과.

5060

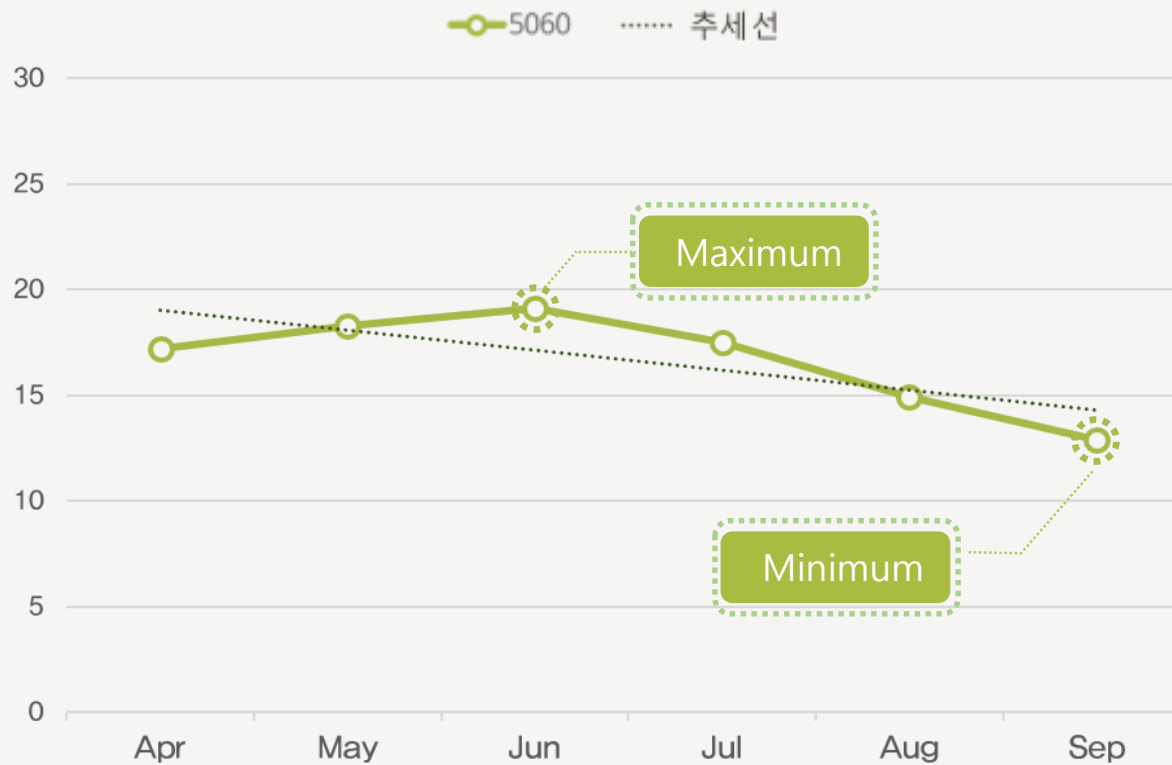
분석 결과, 5060대의 평균 구매 금액에
비해 상품 구매 빈도가 적은 것을 확인

5060대 신규고객 유입을 목표로 하고,
5060대 선호지수 분석



Monthly Purchase Percentage

5060대 월별 구매 비율



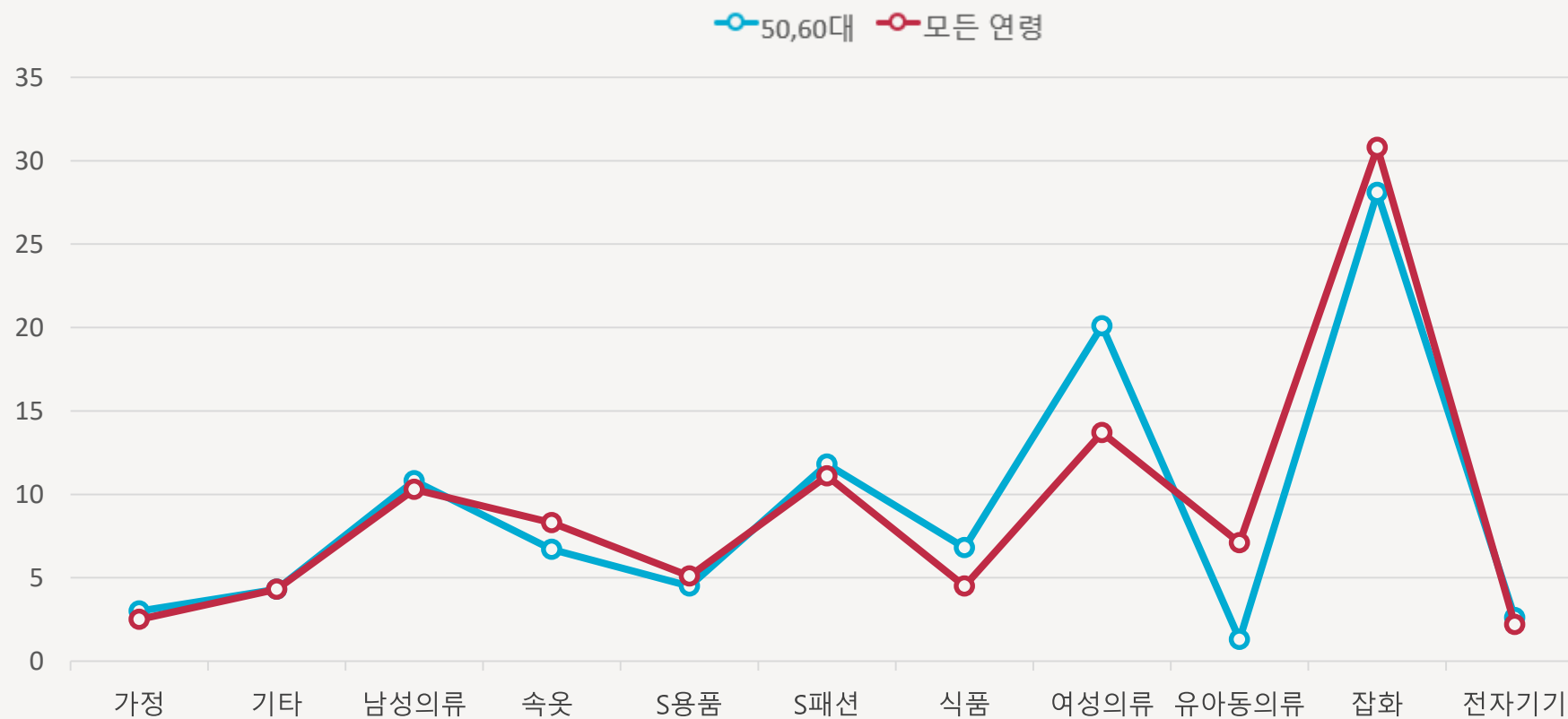
19.1%

6월에서 가장 많은 구매 빈도를 보이고 있으며, 두 번째로는 5월이 18.3%를 보이고 있다.

추세선을 보았을 때, 9월로 갈수록 구매빈도가 점차 낮아지는 모습을 볼 수 있다.

Purchase Percentage by Age

나이대별 상품군 구매 비율



28%

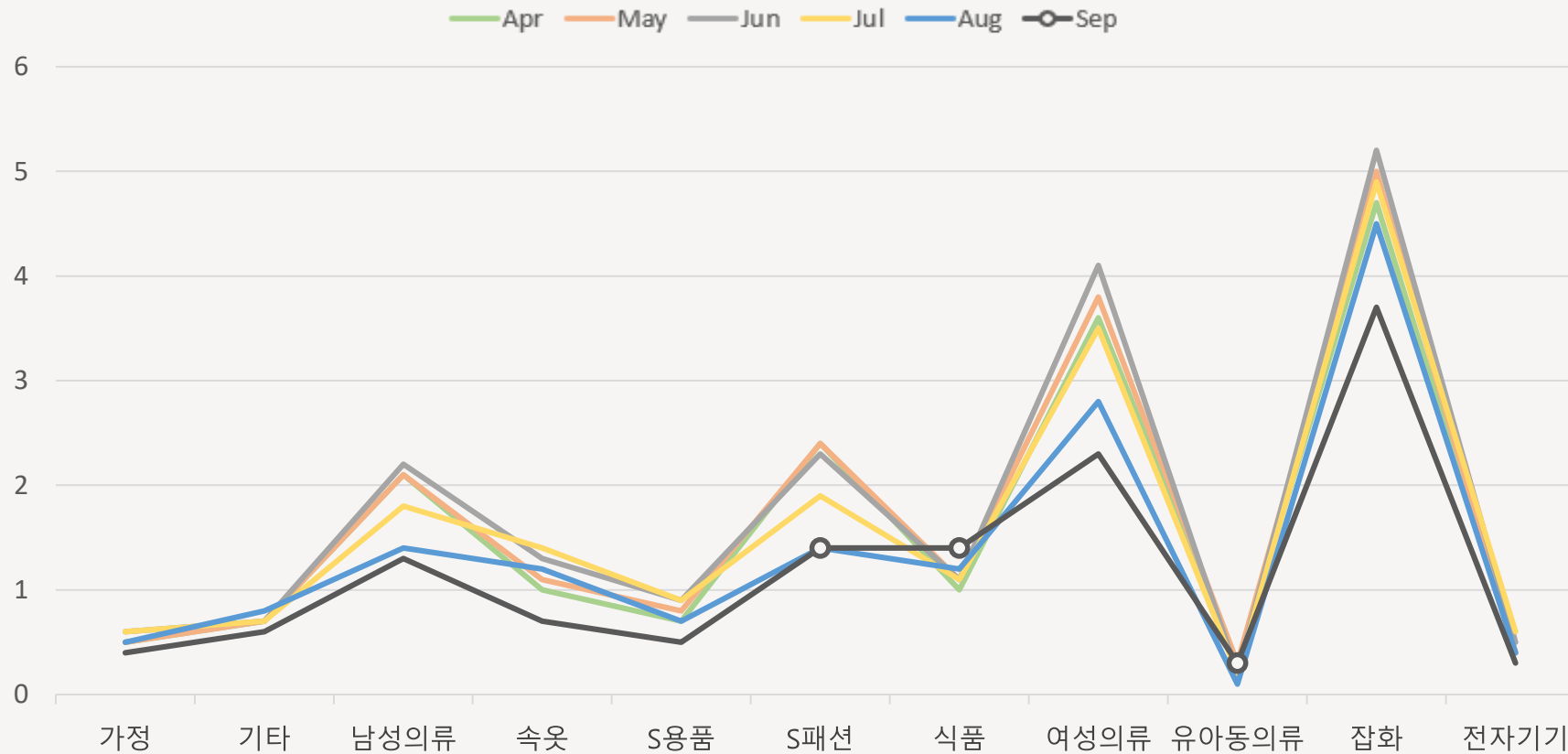
50,60대 사람들은 잡화를 가장 많이 구매하며, 그 다음은 여성의류가 20.1%를 차지한다.

30%

모든 연령대의 사람들은 잡화를 가장 많이 구매하며, 그 다음은 여성의류가 13.7%를 차지한다.

Product Purchase Percentage

5060대 월별 상품군 구매 비율



9월

월별 구매 빈도와 마찬가지로 9월로 갈수록 점차 감소하나, 9월에는 식품, 스포츠패션, 유아동의류가 소폭 상승하였다.

Regression Analysis

회귀분석

매출액에 가장 큰 영향을 주는 요인을
찾기 위해 분석 실시



Regression Analysis

회귀분석



Y : 일일 매출액

매출액을 증가시키는 영향력 있는 변수를
찾기 위해 종속변수로 설정



X1 : 일별 50,60대의 고객 수

50,60대 고객수가 매출액에 어느 정도의 영향력인지
알아보기 위해 독립변수로 선정



X2 : 일별 전국 평균 온도

월별 구매빈도의 차이를 보고 온도가 영향을 줄 것이라
판단되어 독립변수로 선정 (출처 : 기상자료개방포털)



X3 : 추석까지 남은 일수

추석에 가까워질수록 구매액에 영향을 줄 것으로
판단되어 독립변수로 선정 (출처 : 네이버 공휴일 검색)



X4 : 공휴일의 유무(질적 변수)

공휴일도 구매액에 영향을 줄 것으로 판단되어 공휴일 3일전과
3일후까지 1로 표시, 아닌 경우 0 (출처 : 네이버 공휴일 검색)



독립변수들을 단위정상법을 통해 데이터 표준화

```
#데이터 표준화
srdata=rdatatt
for(i in 2:5){
  srdata[,i]=scale(srdata[,i])
}(srdata)|
```

	buyprice	pview	fscust	chudday	temp	holicheck	avtemp
2018-09-09	1415927800	0.49461390	-0.17385717	-1.32146056	-0.172572566	0	20.4
2018-09-10	1462976410	0.59442516	-0.08190381	-1.34033857	-0.136664914	1	20.6
2018-09-11	1242763670	0.01500513	-1.12826964	-1.35921658	-0.316203172	1	19.6
2018-09-12	1356967800	0.07814046	-0.70338170	-1.37809459	-0.064849611	1	21.0
2018-09-13	1316964800	0.12034666	0.04492841	-1.39697259	-0.082803437	1	20.9
2018-09-14	1210351280	-0.52459428	-1.20753978	-1.41585060	0.042873343	1	21.6
2018-09-15	1415676210	0.17716429	0.04492841	-1.43472861	0.114688647	1	22.0
2018-09-16	1596174350	1.13118286	0.27322641	-1.45360662	0.240365427	1	22.7
2018-09-17	1593005540	0.59751492	0.05444083	-1.47248463	0.096734821	0	21.9
2018-09-18	1261463320	-0.10367598	-0.94436292	-1.49136263	-0.064849611	0	21.0
2018-09-19	961724830	-1.06313658	-1.87340895	-1.51024064	-0.136664914	0	20.6
2018-09-20	754100230	-2.06254266	-2.41878750	-1.52911865	-0.298249346	0	19.7
2018-09-21	600923720	-2.37997729	-3.12270634	-1.54799666	-0.370064649	0	19.3
2018-09-22	691756040	-1.70846707	-2.60586503	-1.56687467	-0.190526391	0	20.3
2018-09-23	588287270	-1.98530863	-2.91026236	-1.58575268	-0.441879952	0	18.9
2018-09-24	558885560	-2.06612685	-3.01172814	-1.60463068	-0.854817945	0	16.6
2018-09-25	846744200	-0.82542144	-1.13778206	-1.62350869	-0.872771770	0	16.5
2018-09-26	1298226850	1.06065607	-0.42752162	-1.64238670	-0.639372035	0	17.8
2018-09-27	1075025930	-0.52179419	-1.43583778	-1.66126471	-0.747094990	0	17.2
2018-09-28	1066987560	-0.49263457	-1.47071664	-1.68014272	-0.800956467	0	16.9
2018-09-29	1118601380	-0.21837169	-1.17583173	-1.69902072	-0.495741429	0	18.6
2018-09-30	1382744600	0.41650422	-0.29751859	-1.71789873	-0.513695255	0	18.5



회귀분석

```
#회귀분석
(reg1<-lm(buyprice~.,data=srdata))
summary(reg1)
```

```
> summary(reg1)

Call:
lm(formula = buyprice ~ ., data = srdata)

Residuals:
    Min       1Q   Median       3Q      Max
-277091015 -77938360 -25437177  79727022 431545399

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1152363414   10509591  109.649  < 2e-16 ***
fscust       221310910    11829696   18.708  < 2e-16 ***
chudday      -89225058    14422015   -6.187 4.10e-09 ***
temp        -89849066    12161600   -7.388 5.57e-12 ***
holicheck1    6011017     22612554    0.266  0.791
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

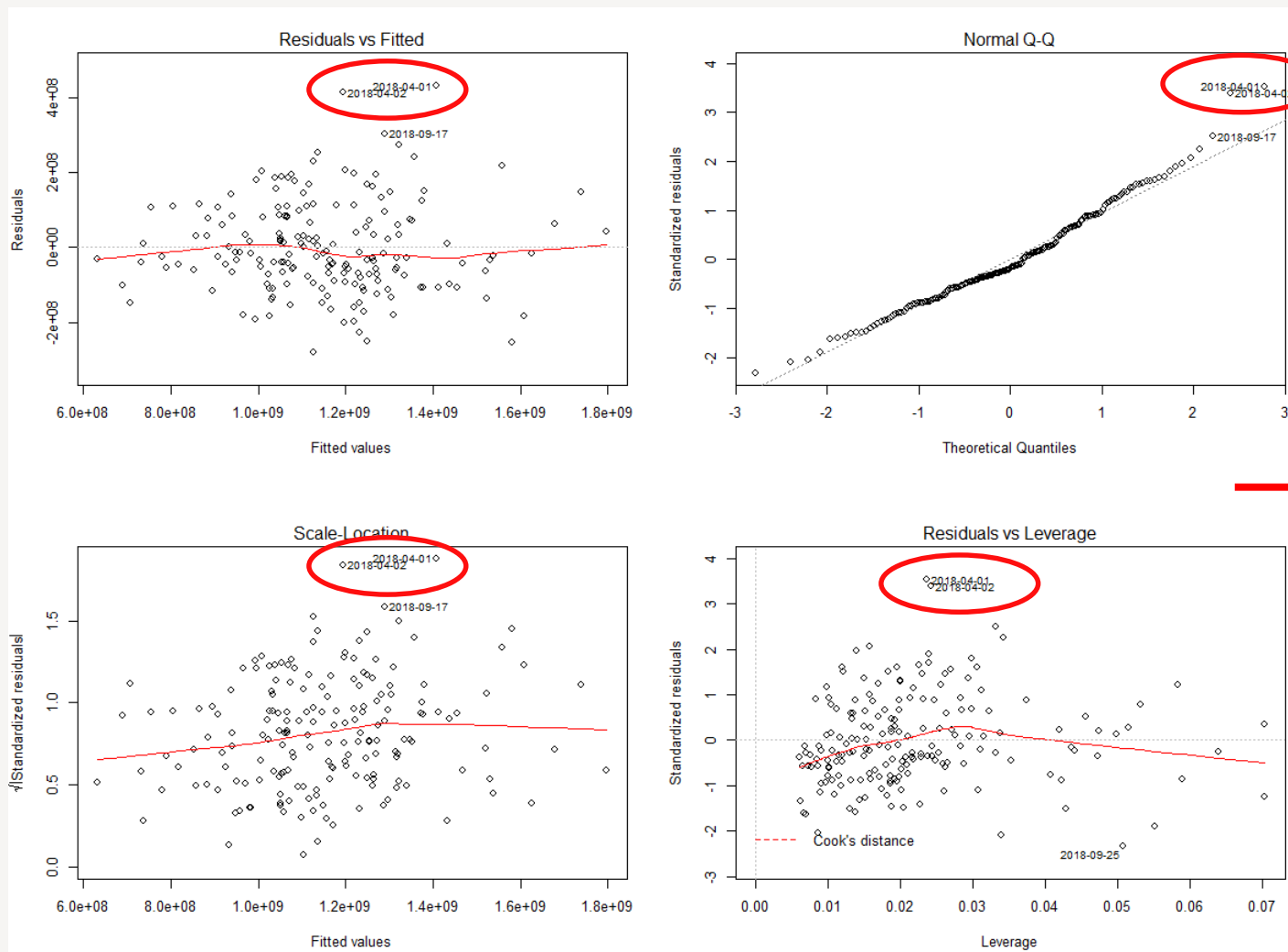
Residual standard error: 123600000 on 178 degrees of freedom
Multiple R-squared:  0.7262,    Adjusted R-squared:  0.72
F-statistic: 118 on 4 and 178 DF,  p-value: < 2.2e-16
```

0.05 이상

공휴일 유무 독립변수는
유의하지 않으므로 제거



이상값 확인



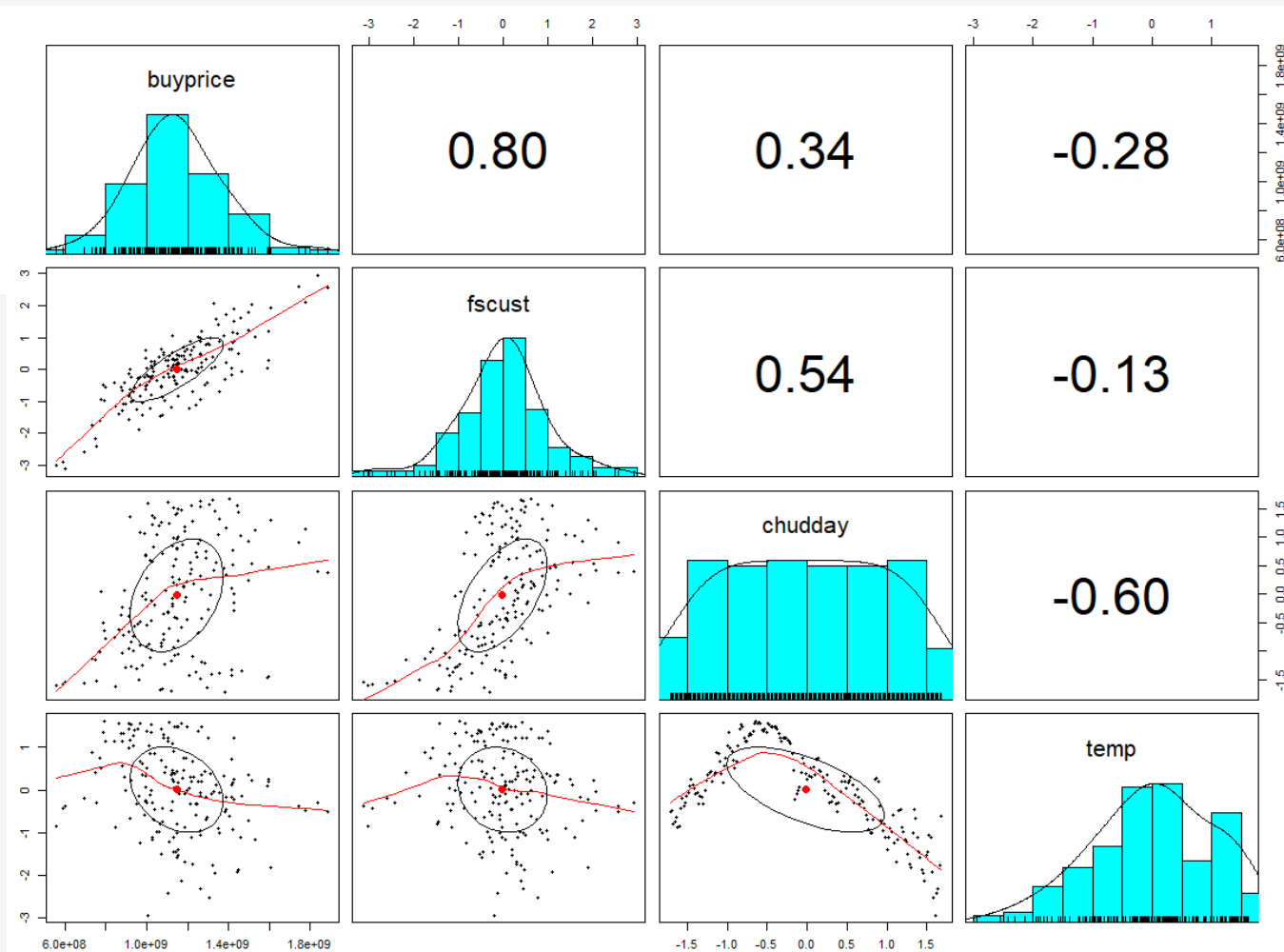
전체적인 그래프에서 4월 1일,
4월 2일 데이터가 이상값으로 보인다.

따라서, 제거하도록 한다.



유의하지 않은 공휴일 유무변수와 이상값 데이터 삭제 후 회귀분석

```
#유의하지 않았던 holichack 변수 삭제,  
#이상값인 4월1일 4월2일 데이터 삭제 후 회귀분석  
rregt<-srdata[,-c(5)]  
rregt<-rregt[-c(1:2),]  
pairs.panels(rregt)  
reg2<-lm(buyprice~.,data=rregt2)  
summary(reg2)  
par(mfrow=c(2,2))  
plot(reg2)  
par(mfrow=c(1,1))
```





회귀분석

```
> summary(reg2)
```

```
Call:
```

```
lm(formula = buyprice ~ ., data = rreg2)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-290604186 -69170060 -17769463  82155692 295139772
```

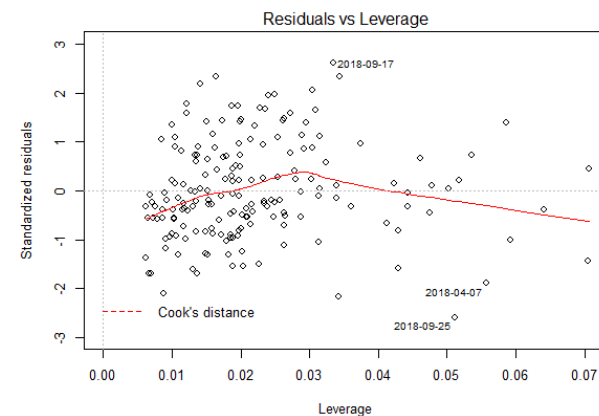
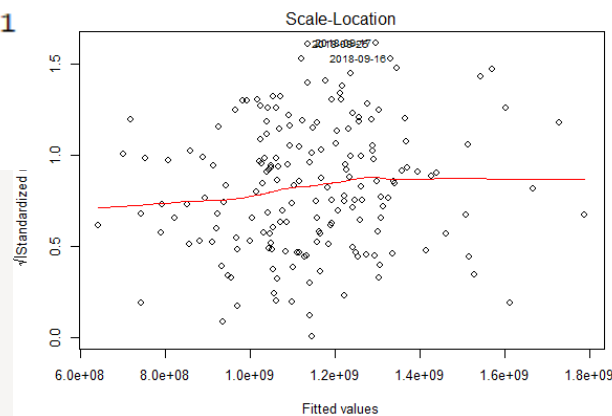
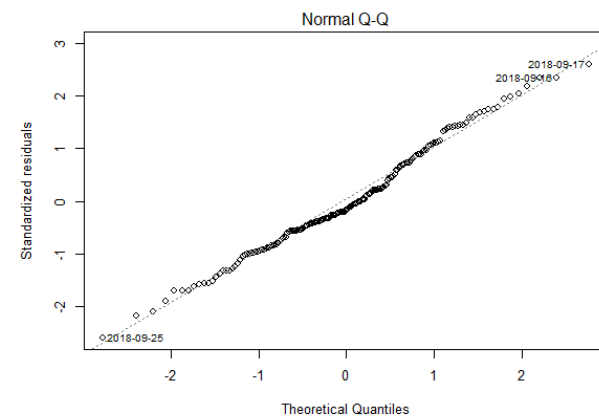
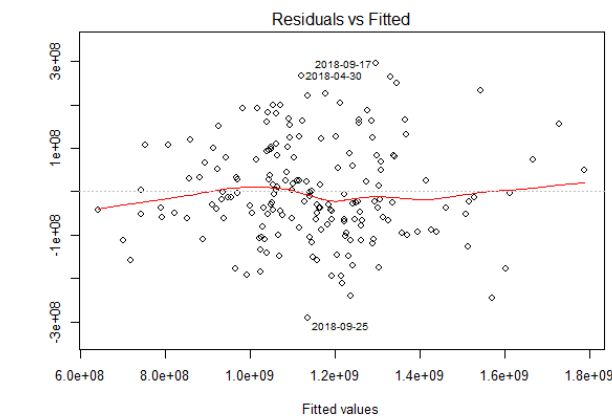
```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1148932969	8568073	134.095	< 2e-16 ***
fscust	221911367	10652584	20.832	< 2e-16 ***
chudday	-98977044	13369858	-7.403	5.19e-12 ***
temp	-91905010	11222619	-8.189	5.05e-14 ***

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

```
Residual standard error: 115200000 on 177 degrees of freedom
Multiple R-squared:  0.746,    Adjusted R-squared:  0.7417
F-statistic: 173.3 on 3 and 177 DF,  p-value: < 2.2e-16
```

모든 변수가 매우 유의함



✓ 회귀분석 결과

$$Y_i = 1148932969 + 221911367 * X_1 - 98977044 * X_2 - 91905010 * X_3$$

50,60대 고객수가 일매출액에 가장 큰 영향

```
> (lapply(regrd,mean))
$buyprice
[1] 1147468854

$fscust
[1] 1581.276

$chudday
[1] 73

$temp
[1] 21.4116
```

```
> predict(regr,data.frame("fscust"=3000,"chudday"=73,"temp"=21.3612),interval="confidence")
      fit      lwr      upr
1 2146568168 2050443230 2242693107
> predict(regr,data.frame("fscust"=3000,"chudday"=73,"temp"=21.3612),interval="prediction")
      fit      lwr      upr
1 2146568168 1899648458 2393487878
```

표준화전 회귀분석으로 매출액 예측

가장 큰 영향력을 가졌던 50,60대 고객수가 2배 증가하였다고
예상하고 매출액을 예측하였을때

95%신뢰구간, 예측구간에서 일매출이 2배가 증가함을 볼 수 있었다

Marketing Method

마케팅 방법

5060대의 구매 수단과 외부 데이터를
이용하여 마케팅 방법을 찾는다.

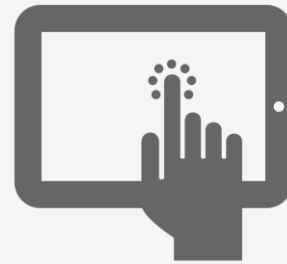


Type of device Purchased

5060대 상품 구매 기기의 비율



0.0 %



1.3 %



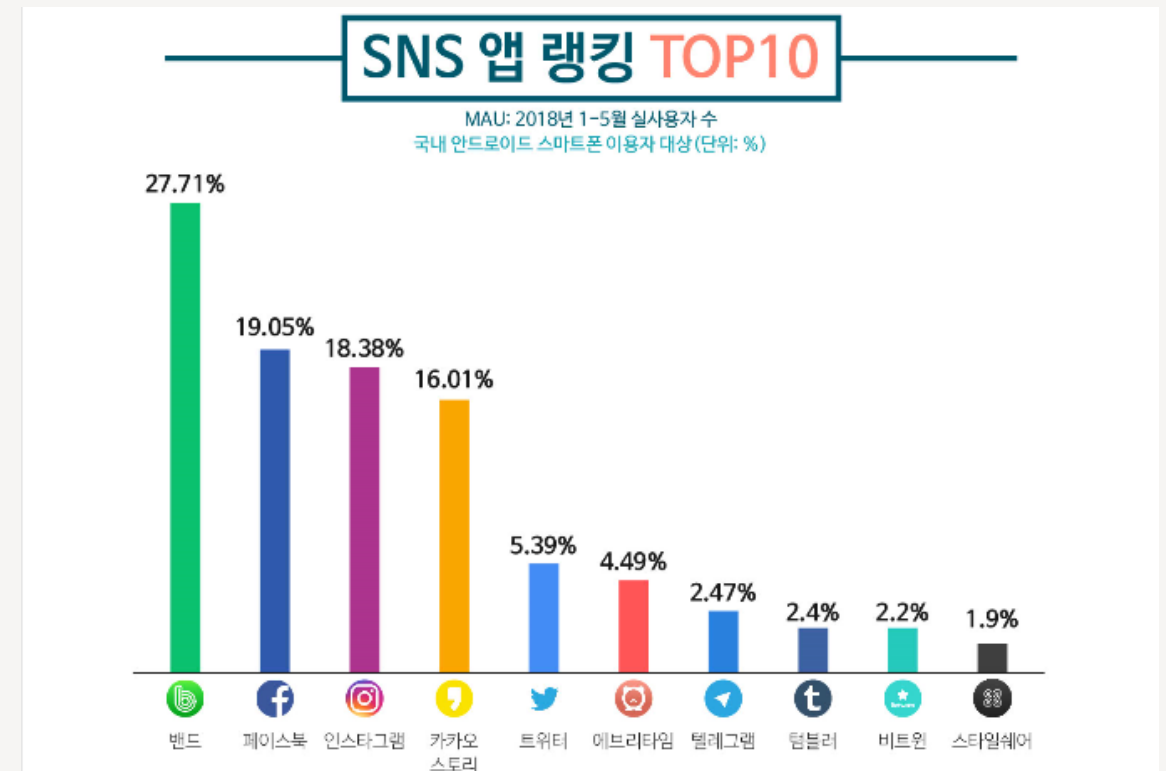
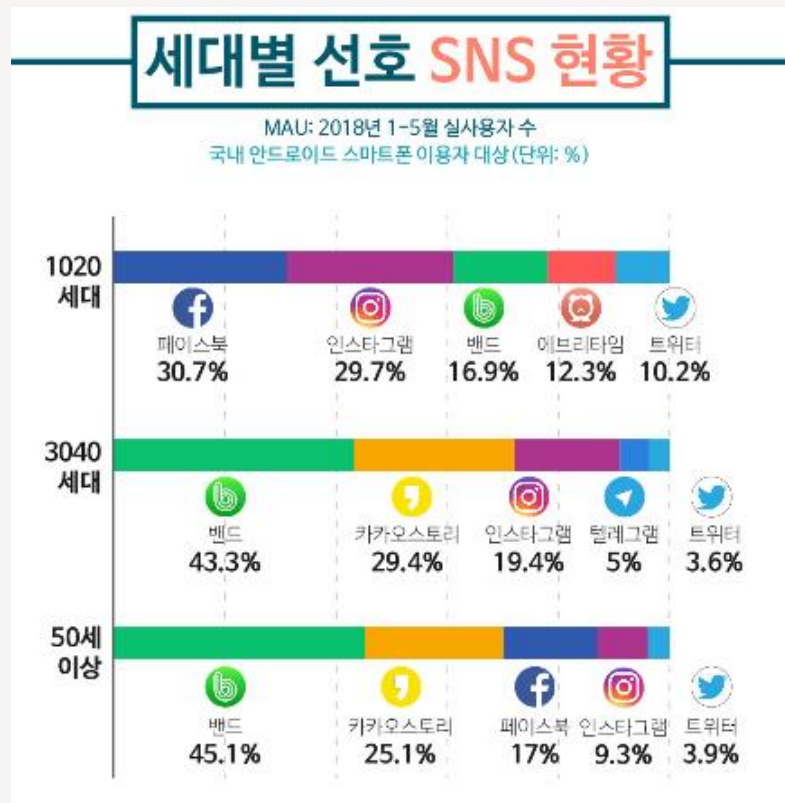
98.7 %

SNS preference Index

안드로이드 세대별 선호 SNS 현황

출처 : MOBIINSIDE

참고 : APP Ape (앱 분석 기관)의 데이터



04

CONCLUSION

결과정리
활용



28%
SUNDRIES

45%
BAND

95%
MOBILE

SNS MARKETING FOR 5060

NAVER BAND

앞선 분석 결과를 살펴보면, 5060대에서 구매 기기 유형이 98.7%로 대부분이 스마트폰을 이용하여 쇼핑을 하는 것을 알 수 있다.

또한 50대 이상 선호 SNS 앱을 살펴본 결과, 45.1%가 네이버 밴드를 선호한다.

따라서, 네이버 밴드를 이용한 홍보전략을 세우도록 한다.

BAND 밴드, 페이지, 게시물 검색

새글 피드 | 찾기

LOTTE

1,573명 구독 중

메시지 + 구독하기

홈
최신글
사진첩
연결 밴드 3

글 내용, #태그 검색

네이버 밴드에 롯데온라인 홍보 페이지 생성
롯데 홈쇼핑 페이지 구독 시 혜택 제공

이 페이지 공유하기

공지사항 7

[중요]

네이버 밴드 롯데 페이지 오픈 페스티벌 !!

롯데 페이지 구독 시 특별한 혜택 !!

롯데 페이지 구독하고 특별한 혜택 받아가자 !!

...더보기

최근 사진 더보기

내 손안에 TV홈쇼핑
바로 TV
롯데홈쇼핑

LOTTE
롯데홈쇼핑

연결된 밴드

롯데홈쇼핑 L.POINT

롯데페이지 구독

1. 롯데밴드 페이지 구독 시 혜택
2. 지인에게 구독 유도 혜택
3. 온라인 구매 시 혜택

밴드 신규 가입시 혜택

1. 밴드 신규 가입시 쿠폰 or 포인트 증정



월별 혜택

- 4월 ~ 6월 : 의류 홍보 및 이벤트
 4월 ~ 8월 : 잡화 홍보 및 이벤트
 9월 : 식품 홍보 및 이벤트

페이지 월별 혜택 상품을 밴드로 통해 구매시
 할인 쿠폰 or 페이백

롯데페이지와 롯데홈쇼핑 연동

롯데페이지와 롯데홈쇼핑을 오갈 수 있는
 루트 형성