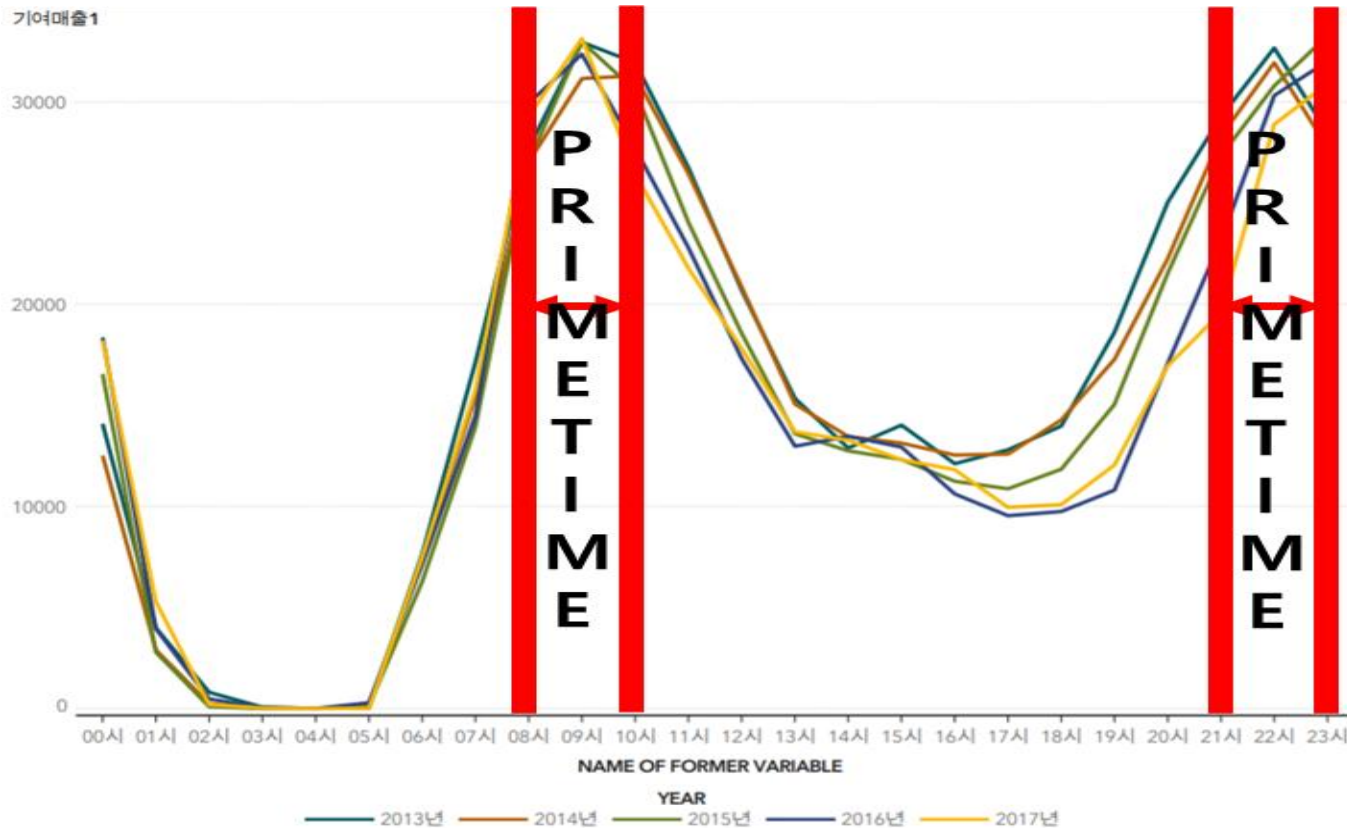


SAS분석챔피언십 최종 보고서

SA1001

요인 탐색 및 시각화 (EDA)

시간대(M)별 매출액



일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR_SALE_AMT PUR_SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7195335.05	7195335.05	2922.67	<.0001
Error	29075	71579877.39	2461.90		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR_SALE_AMT Mean
0.091340	78.28406	49.61758	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
PRIMETIME	1	7195335.054	7195335.054	2922.67	<.0001

Generated by the SAS System ('SASApp', Linux) on 2019년 08월 30일 at 11:05:15 PM

페이지 나누기

일원분산분석

결과

The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
PRIMETIME	1	2630.9	<.0001

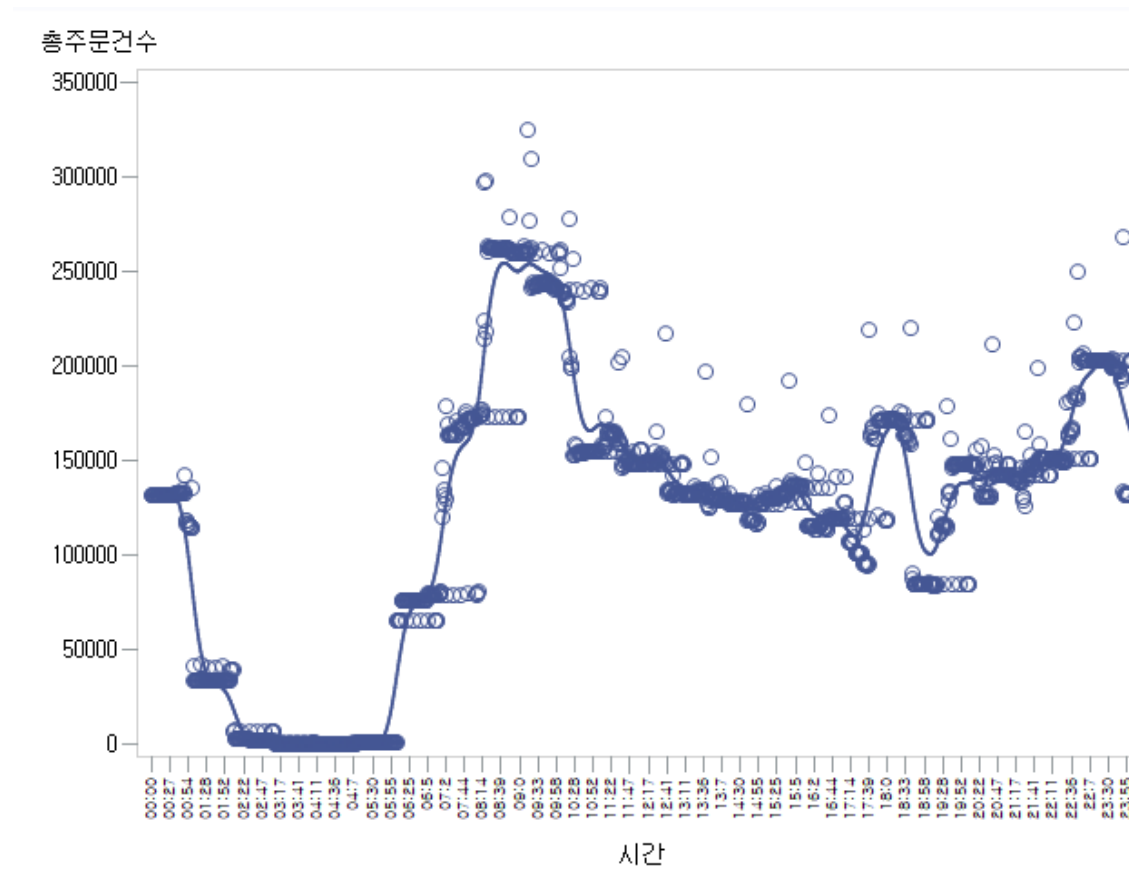
Welch's ANOVA for PUR_SALE_AMT

Source	DF	F Value	Pr > F
PRIMETIME	1.0000	2435.72	<.0001
Error	17343.9		

Generated by the SAS System ('SASApp', Linux) on 2019년 08월 30일 at 11:05:15 PM

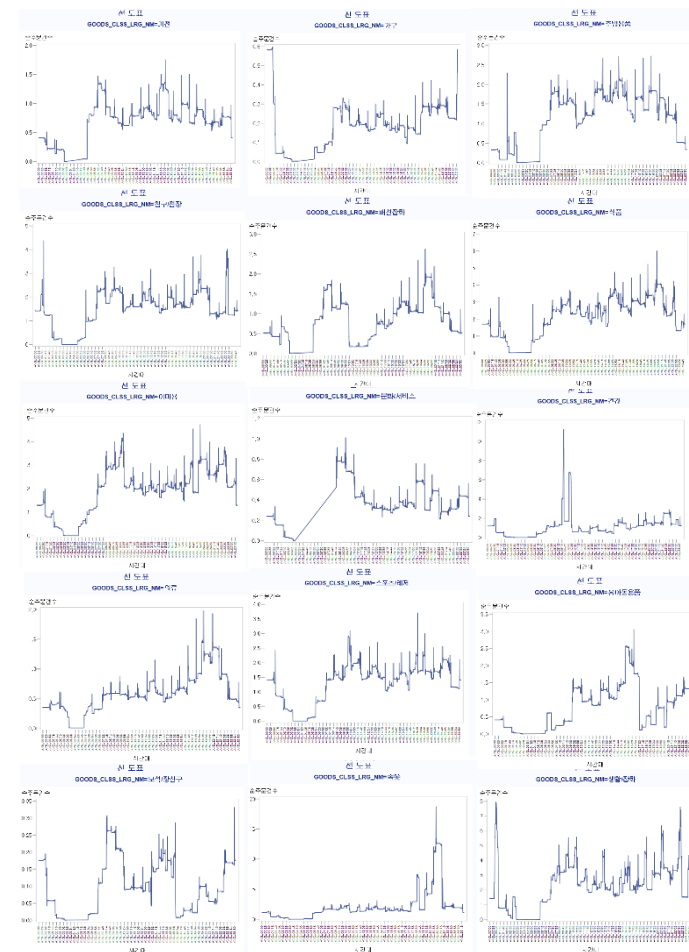
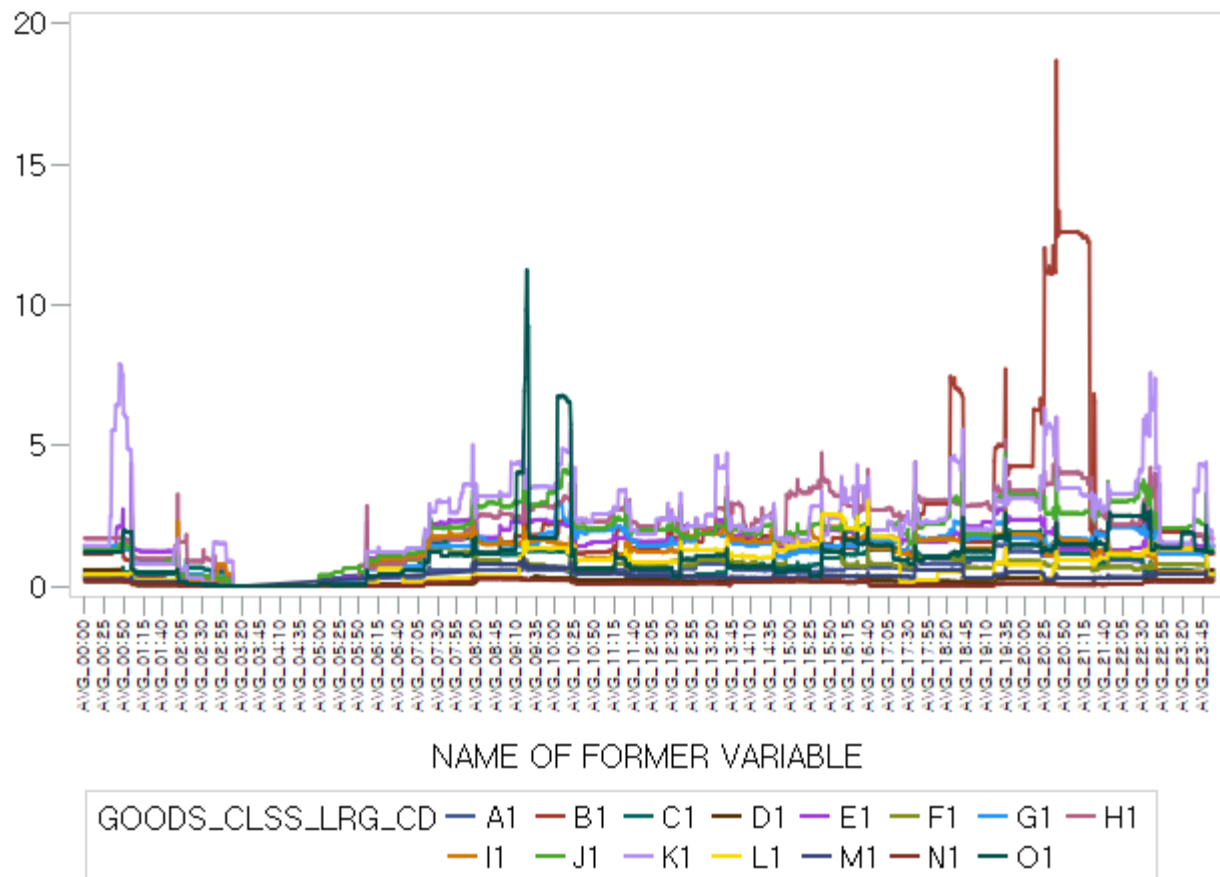
시간대(M)별 매출액의 차이 및 PRIME TIME 존재

시간대(단위:분)별 순주문건수



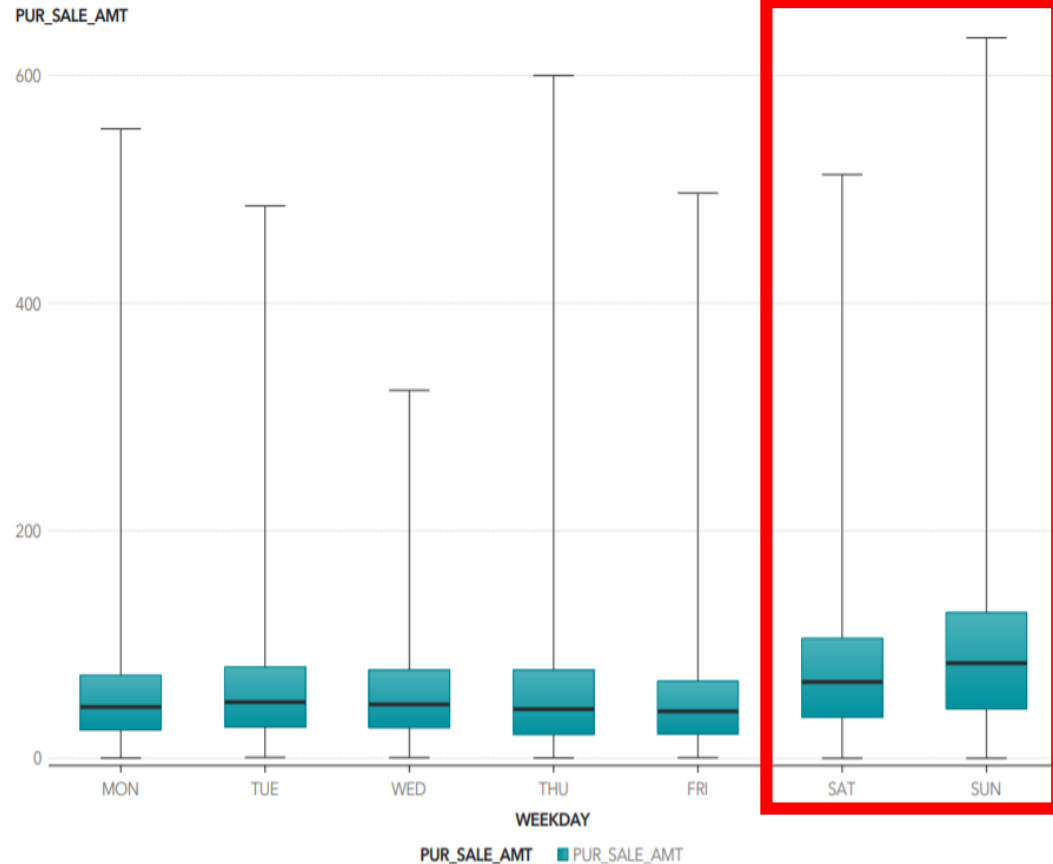
시간대(M)별 순주문건수의 차이 및 PRIME TIME 존재

대분류의 시간대(단위:분)별 순주문건수



대분류마다 시간대(M)별 순주문건수의 차이 및 PRIME TIME 존재

요일(WEEKDAY)별 매출액 차이



일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR SALE_AMT PUR SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	5077822.21	846303.70	333.83	<.0001
Error	29070	73697390.23	2535.17		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR SALE_AMT Mean
0.064460	79.44037	50.35047	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
WEEKDAY	6	5077822.210	846303.702	333.83	<.0001

생성 환경: SAS 시스템('SASApp', Linux), 생성 일시: 2019년 08월 12일, 오후 06:41:57

페이지 나누기

일원분산분석

결과

The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
WEEKDAY	6	1876.0	<.0001

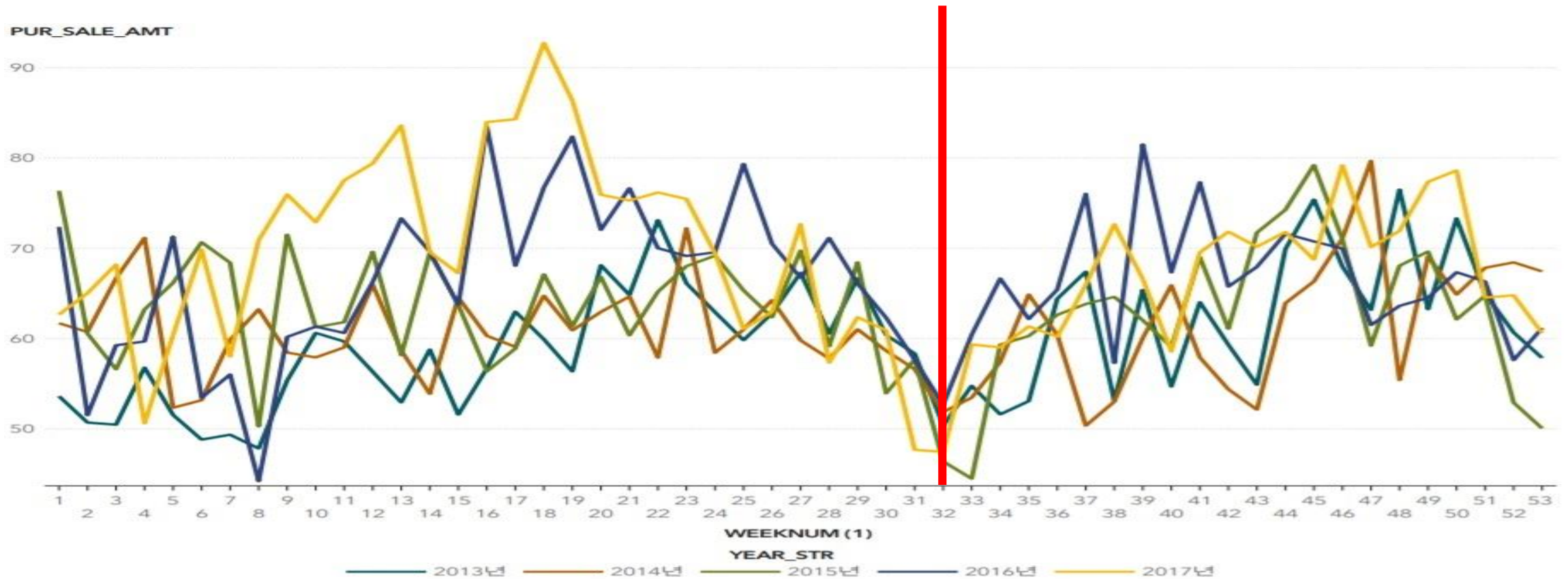
Welch's ANOVA for PUR SALE_AMT

Source	DF	F Value	Pr > F
WEEKDAY	6.0000	278.31	<.0001
Error	12243.5		

생성 환경: SAS 시스템('SASApp', Linux), 생성 일시: 2019년 08월 12일, 오후 06:41:57

요일(WEEKDAY) 별 매출액 차이가 존재

주차(WEEKNUM)별 매출액 차이



주차(WEEKNUM)별 매출액의 차이가 및 트렌드 존재

중(MID), 소(SML), 세(THN) 별 매출액 차이

중분류(GOODS_CLSS_MID_CD)	소분류(GOODS_CLSS_SML_CD)	세분류(GOODS_CLSS_THN_CD)
일원분산분석 결과 The ANOVA Procedure	일원분산분석 결과 The ANOVA Procedure	일원분산분석 결과 The ANOVA Procedure
Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance	Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance	Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance
Source	Source	Source
GOODS_CLSS_MID_CD	GOODS_CLSS_SML_CD	GOODS_CLSS_THN_CD
DF	DF	DF
Chi-Square	Chi-Square	Chi-Square
Pr > ChiSq	Pr > ChiSq	Pr > ChiSq
62	188	498
5399.6	7352.5	9289.4
<.0001	<.0001	<.0001
Welch's ANOVA for PUR_SALE_AMT	Welch's ANOVA for PUR_SALE_AMT	Welch's ANOVA for PUR_SALE_AMT
Source	Source	Source
GOODS_CLSS_MID_CD	GOODS_CLSS_SML_CD	GOODS_CLSS_THN_CD
DF	DF	DF
F Value	F Value	F Value
Pr > F	Pr > F	Pr > F
62.0000	188.0	498.0
207.24	120.75	59.06
<.0001	<.0001	<.0001
Error	Error	Error
283.7	551.1	833.9

PROBLEM) 범주형 변수의 LEVEL이 많아 과 적합 문제 발생



SOLUTION) 파생변수 생성 시 매출액을 기준으로 군집분석으로 새로운 기준 마련

이상치 처리

비영리 목적 방송 이상치 제거

[Re] 데이터 관련 질문 드립니다.

관리자

8/2/2019

51

URL http://www.sas-analytics.co.kr/qna/qa_view.asp?list_gotopage=17&Sid_key=377&wtnum=&sas_sort=0

COPY

[본문내용]

안녕하세요!데이터 확인 중 의문점이 발생해서 질문 드립니다.?PRGM_ACRS 데이터셋에서 방송프로그램 코드가?230898인 튜플은매출액이 4.5로 값이 있음에도 불구하고 순주문건수와 순주문수량가 0입니다.매출액이 (순매출액이 아니라) 주문건수와 수량 외 변수의 영향을 받기도 하나요?????

주문 수량과 주문건수가 0인 데이터는 일반적인 (영리 목적의) 방송 상품이 아닙니다.

분석가의 판단에 따라 분석 진행해주시기 바랍니다.?

비영리 목적의 방송(순주문건수 = 0, 순주문수량 =0, 매출액 >0) 레코드 제거

데이터 변환

날짜 데이터 변환

BDCT_DT
2013-01-01

INPUT(SUBSTR(BDCT_DT, 1, 4), 4.)

INPUT(SUBSTR(BDCT_DT, 6, 2), 2.)

INPUT(SUBSTR(BDCT_DT, 9, 2), 2.)

YEAR	MONTH	DAY
2013	1	1

MDY(MONTH, DAY, YEAR)

BDCT_DT_MDY
19359

시간대 데이터 변환

BDCT_STRT_DTIME	BDCT_END_DTIME
2013-01-01 01:00:00	2013-01-01 02:10:00


INPUT(SUBSTR(BDCT_STRT/END_DTIME, 12, 2), 2.)
INPUT(SUBSTR(BDCT_STRT/END_DTIME, 15, 2), 2.)

BDCT_STRT_DTIME		BDCT_END_DTIME	
HOUR	MINUTE	HOUR	MINUTE
1	0	2	10


HOUR + MINUTE/60

STRT_TIME_REALNUM	END_TIME_REALNUM
1	2.1666666667

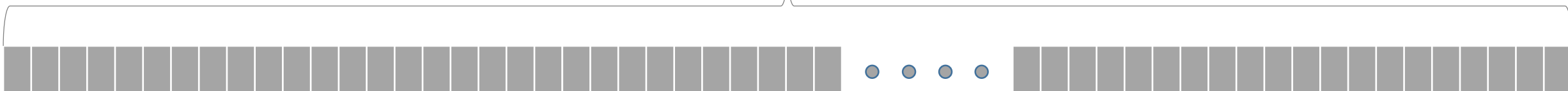
파생변수 생성

시간대 변수 생성(1)

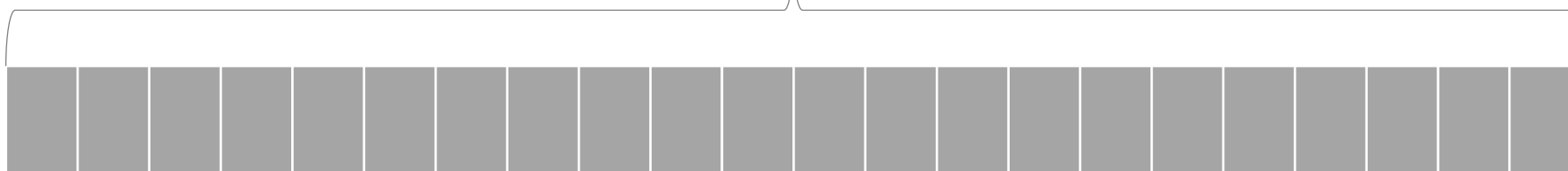
STRT_TIME_REALNUM	END_TIME_REALNUM
1	2.1666666667



1440분

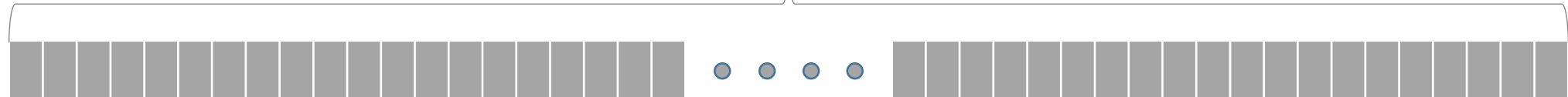


24시간



시간대 변수 생성(2)

1440분



```
%MACRO TEMP;
```

```
DATA TIMETABLE;  
set WORK.TIME_REALNUM;
```

```
%DO h=0 %TO 230 %by 10;
```

```
  %DO m=0 %TO 59;
```

```
    if END_DAY = START_DAY
```

```
      then if ( (&h/10+&m/60 <= START_TIME_REALNUM) AND (START_TIME_REALNUM < &h/10+(&m+1)/60) ) OR
```

```
              ( (&h/10+&m/60 <= END_TIME_REALNUM) AND (END_TIME_REALNUM < &h/10+(&m+1)/60) ) OR
```

```
              ( (START_TIME_REALNUM <= &h/10+&m/60) AND (&h/10+(&m+1)/60 <= END_TIME_REALNUM) ) then V&h&m = 1;
```

```
      else V&h&m = 0;
```

```
    else
```

```
      if ( (&h/10+&m/60 <= START_TIME_REALNUM) AND (START_TIME_REALNUM < &h/10+(&m+1)/60) ) OR
```

```
          ( (&h/10+&m/60 <= END_TIME_REALNUM) AND (END_TIME_REALNUM < &h/10+(&m+1)/60) ) OR
```

```
          ( (START_TIME_REALNUM <= &h/10+&m/60) AND (&h/10+(&m+1)/60 <= 24) ) OR
```

```
          ( (0 <= &h/10+&m/60) AND (&h/10+(&m+1)/60 <= END_TIME_REALNUM) ) then V&h&m = 1;
```

```
      else V&h&m = 0;
```

```
  %END;
```

```
%END;
```

```
output;
```

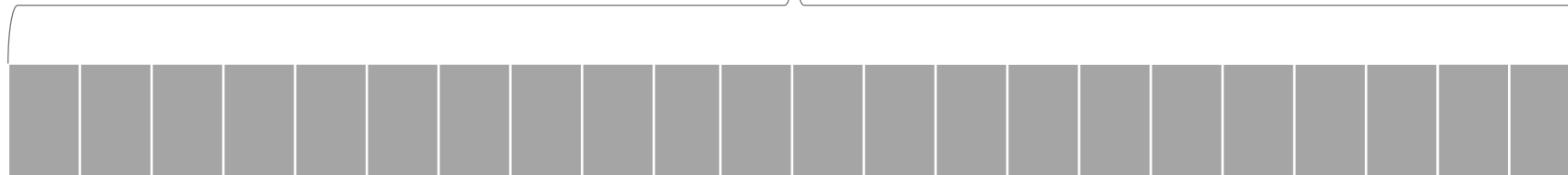
```
run;
```

```
%MEND TEMP;
```

```
%TEMP;
```


시간대 변수 생성(3)

24시간



BDCT_STRT_DTIME	BDCT_END_DTIME
2013-01-01 01:00:00	2013-01-01 02:10:00



M10	M20
1	0.18333333333

```
%MACRO MOD;
```

```
DATA TIMETABLE_2;  
set WORK.TIMETABLE;
```

```
%DO h=0 %TO 230 %by 10;  
  M&h = 0;  
  %DO m=0 %TO 59;  
    M&h = M&h + v&h&m;  
  %END;  
  M&h = M&h / 60;
```

```
%END;  
output;
```

```
%DO h=0 %TO 230 %by 10;  
  %DO m=0 %TO 59;  
    drop v&h&m;  
  %END;
```

```
%END;  
run;
```

```
%MEND MOD;
```

```
%MOD;
```

TERM_MIN(방송기간) 변수 생성(1)

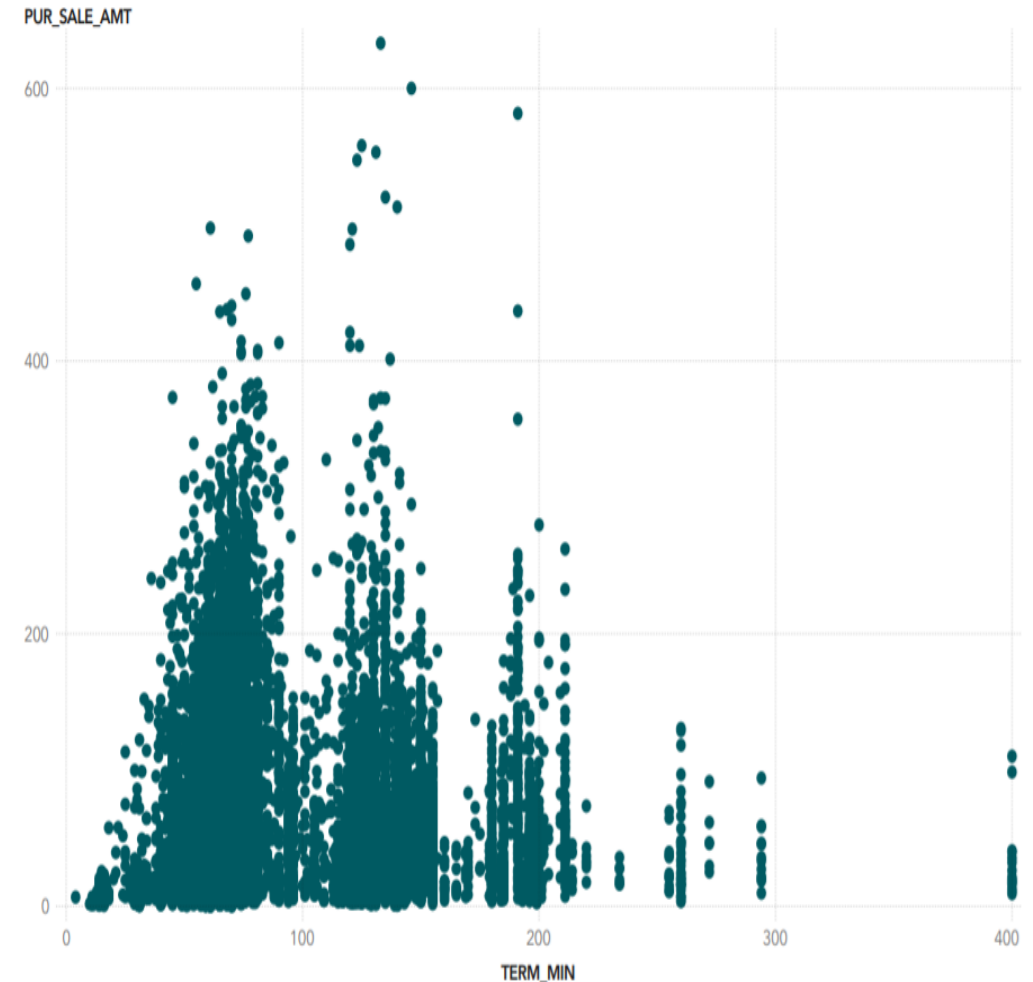
BDCT_STRT_DTIME	BDCT_END_DTIME
2013-01-01 01:00:00	2013-01-01 02:10:00

DHMS(date, hour, minute, second)

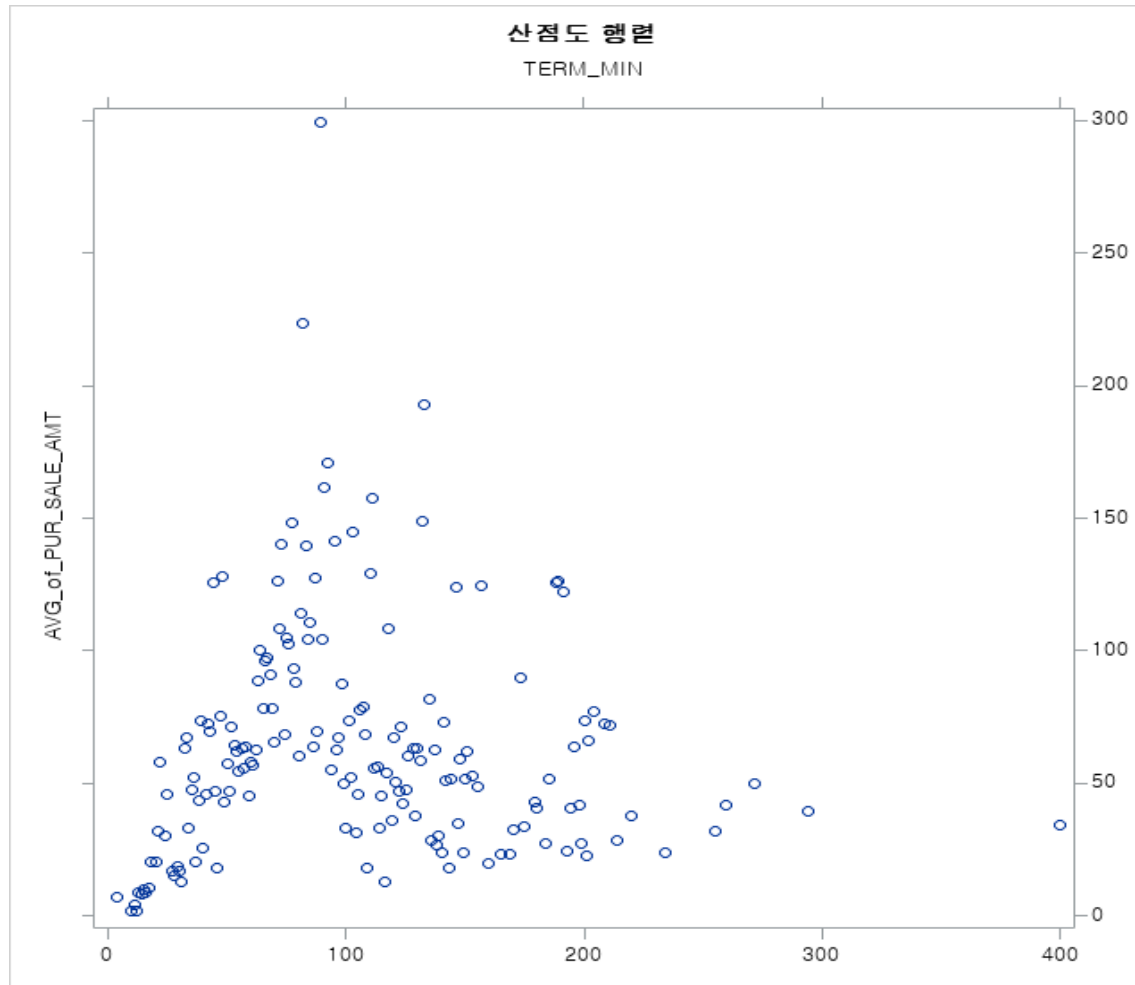
BDCT_STRT_DTIME (SAS날짜 시간 값)	BDCT_END_DTIME (SAS날짜 시간 값)
1672621200	1672625400

$(\text{BDCT_END_DTIME} - \text{BDCT_STRT_DTIME}) / 60$

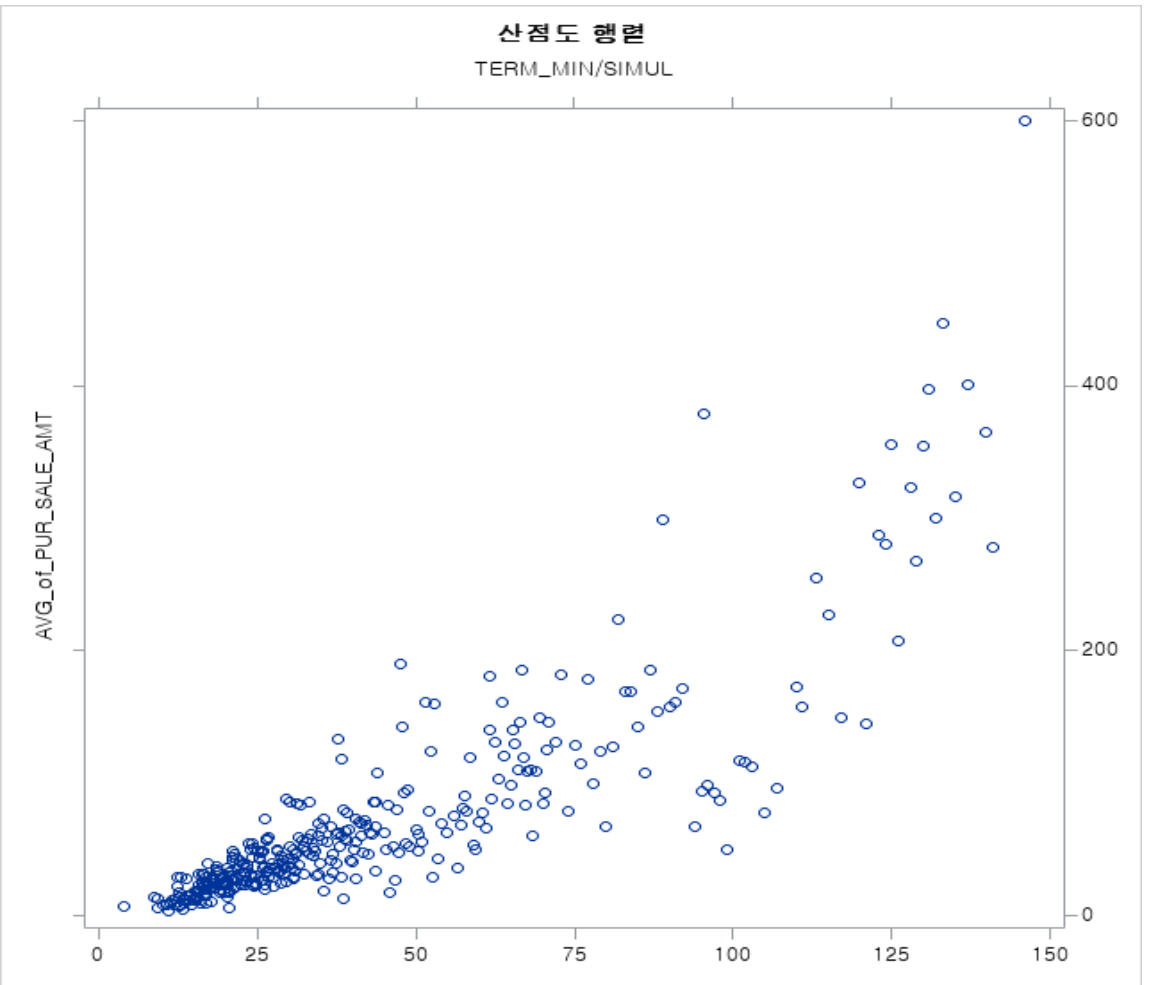
TERM_MIN
70(분)



TERM_MIN(방송기간) 변수 생성(2)

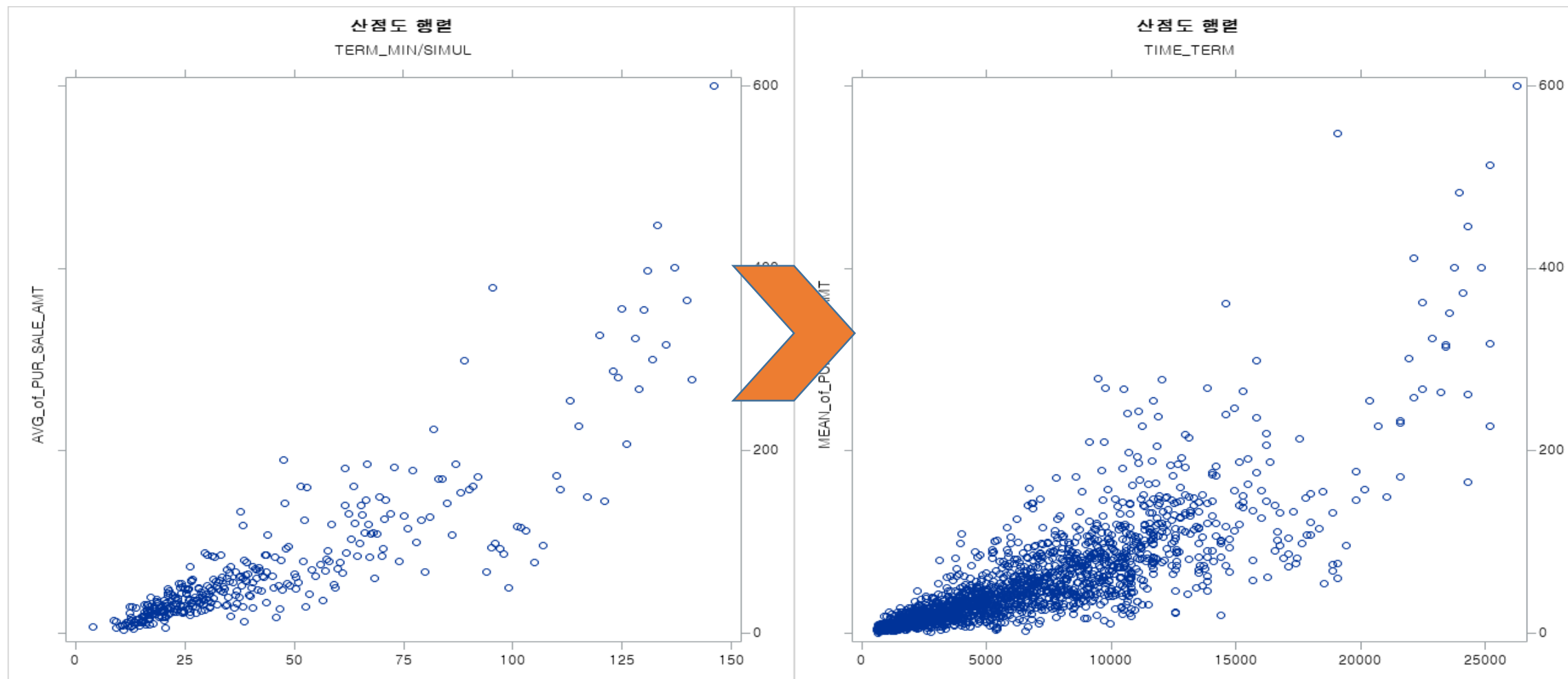


TERM_MIN(상관계수 : 0.007)



TERM_MIN/SIMUL(상관계수 : 0.8597)

TERM_MIN(방송기간) 변수 생성(3)



TERM_MIN/SIMUL(상관계수 : 0.8597)

TIME_TERM(상관계수 : 0.8081)

SIMUL(동시편성 여부) 변수 생성(1)

TV_GOODS_GRP_CD	BDCT_PGM_ID	SIMUL(동시편성)
00101	189993	4
00100	189993	4
00098	189993	4
00099	189993	4
00102	189994	0
00103	189995	0

동일 BDCT_PGM_ID(방송프로그램ID)내에
TV_GOODS_GRP_CD(TV상품그룹코드)가
다른 것이 존재한다면 동시편성으로 봄

SIMUL(동시편성 여부) 변수 생성(2)

일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR_SALE_AMT PUR_SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	10890508.42	907542.37	388.55	<.0001
Error	29064	67884704.02	2335.70		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR_SALE_AMT Mean
0.138248	76.25108	48.32905	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
COUNT_of_TV_GOODS_GRP	12	10890508.42	907542.37	388.55	<.0001

환경: SAS 시스템('SASApp', Linux), 생성 일시: 2019년 08월 27일, 오후 01:3

페이지 나누기

일원분산분석

결과

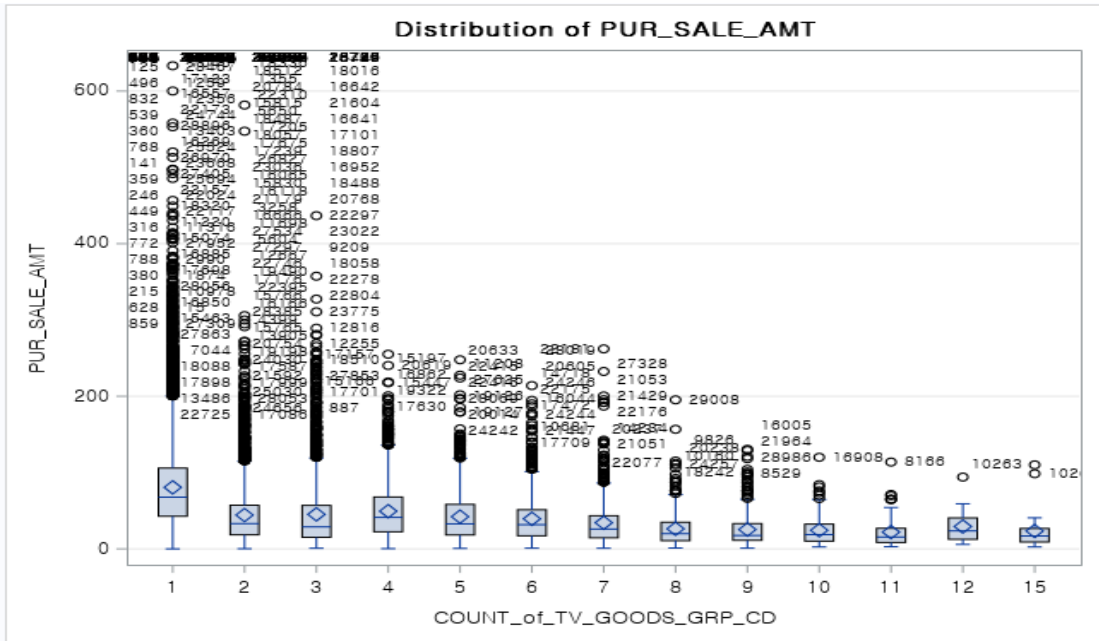
The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
COUNT_of_TV_GOODS_GRP	12	3041.3	<.0001

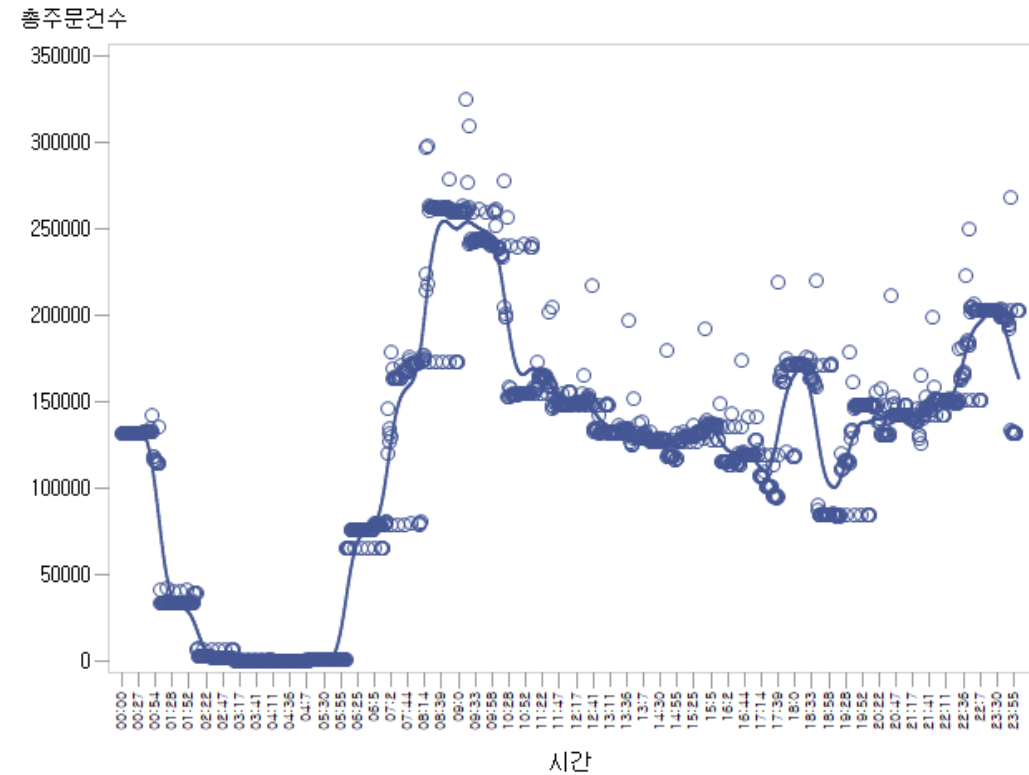
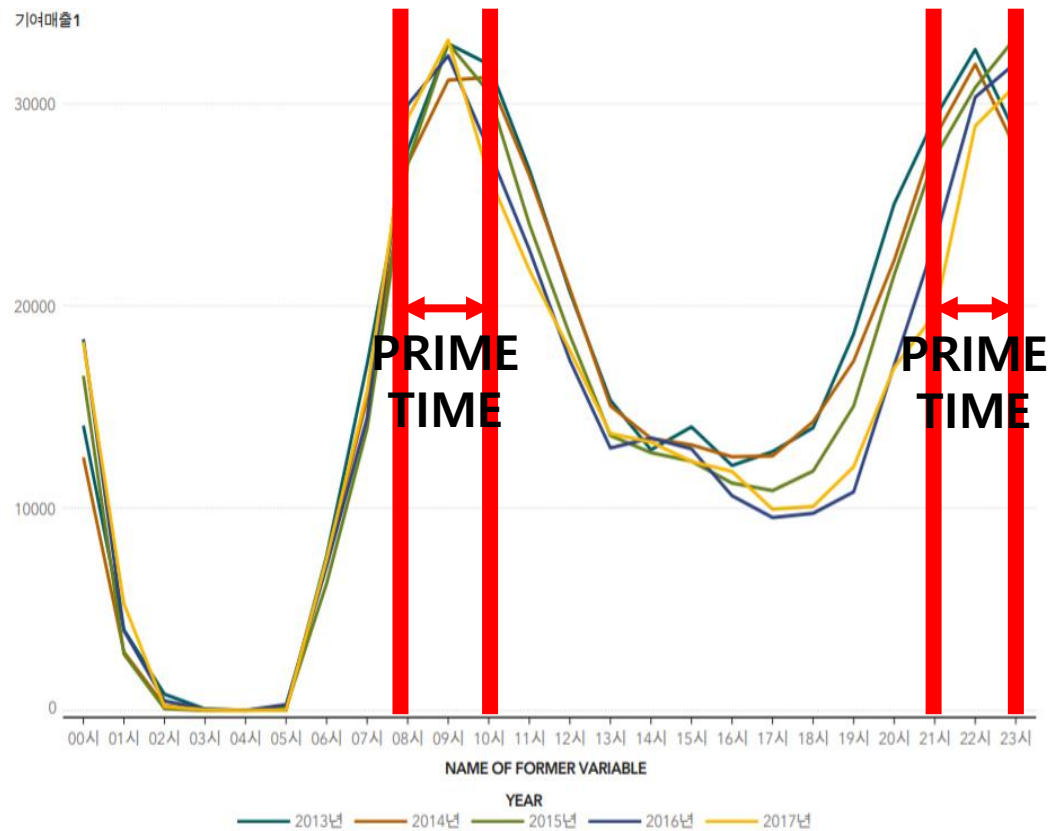
Welch's ANOVA for PUR_SALE_AMT

Source	DF	F Value	Pr > F
COUNT_of_TV_GOODS_GRP	12.0000	536.93	<.0001
Error	532.9		



Level of COUNT_of_TV_GOODS_GRP_CD	N	PUR_SALE_AMT Mean	PUR_SALE_AMT Std Dev
1	15899	80.5819171	55.7013935
2	4594	44.2823901	39.4198419
3	2184	45.1876374	47.5769921
4	2276	49.3593146	35.1136227
5	1285	42.2263813	33.5076380
6	1014	39.3743590	31.4866084
7	770	34.2427273	30.6616945
8	432	26.2960648	22.4651906
9	342	25.2692982	22.7437664
10	150	24.3506667	19.1523554
11	77	21.9753247	19.9621296
12	24	29.3041667	21.0168975
15	30	23.4233333	24.4991581

PRIMETIME 변수 생성



매출액과 순주문건수 기록을 토대로 핵심 시간대를 탐색 (08시~10시, 21시~23시)

WEEKDAY(요일)와 GOODS_CLSS_LRG_NM(대분류) 별 교호작용 변수 생성

⚠ Dependent	⚠ Parameter	⑫ Estimate	⑫ tValue	⑫ Probt
PUR_SALE_AMT	WEEKDAY*GOODS_CLSS_L FRI 가구	58,5546386	12,83	<,0001
PUR_SALE_AMT	WEEKDAY*GOODS_CLSS_L FRI 가전	28,2612515	7,33	<,0001
PUR_SALE_AMT	WEEKDAY*GOODS_CLSS_L FRI 보석/장신구	15,9330330	3,36	0,0008



WEEKDAY	GOODS_CLSS_LRG_NM	FRIX가구	FRIX가전	FRIX보석장신구
FRI	가구	1	0	0
FRI	가전	0	1	0
FRI	보석/장신구	0	0	1






HOLIDAY(휴일)와 GOODS_CLSS_LRG_NM(대분류) 별 교호작용 변수 생성

⚠ Dependent	⚠ Parameter	⑫ Estimate	⑫ tValue	⑫ Probt
PUR_SALE_AMT	WEEKDAY*SIMUL FRI 1	30,29367605	2,52	0,0117
PUR_SALE_AMT	WEEKDAY*SIMUL MON 1	32,29205870	2,69	0,0072
PUR_SALE_AMT	WEEKDAY*SIMUL SAT 1	64,67696534	5,38	<,0001



HOLIDAY	GOODS_CLSS_LRG_NM	HOLIDAY0X가구	HOLIDAY0X가전	HOLIDAY0X문화서비스
0	가구	1	0	0
0	가전	0	1	0
0	문화/서비스	0	0	1






WEEKDAY(요일)와 SIMUL(동시편성) 별 교호작용 변수 생성

 Dependent	 Parameter	 Estimate	 tValue	 Probt
PUR_SALE_AMT	WEEKDAY*SIMUL FRI 1	30,29367605	2,52	0,0117
PUR_SALE_AMT	WEEKDAY*SIMUL MON 1	32,29205870	2,69	0,0072
PUR_SALE_AMT	WEEKDAY*SIMUL SAT 1	64,67696534	5,38	<,0001



WEEKDAY	SIMUL	FRIXSIMUL1	MONXSIMUL1	SATXSIMUL1
FRI	1	1	0	0
MON	1	0	1	0
SAT	1	0	0	1

HOLIDAY(휴일)와 SIMUL(동시편성) 별 교호작용 변수 생성

 Dependent	 Parameter	 Estimate	 tValue	 Probt
PUR_SALE_AMT	HOLIDAY*SIMUL 0 1	54,24340088	3,84	0,0001
PUR_SALE_AMT	HOLIDAY*SIMUL 0 4	31,80838758	2,24	0,0249
PUR_SALE_AMT	HOLIDAY*SIMUL 1 1	87,20449856	6,17	<,0001



HOLIDAY	SIMUL	HOLIDAY0XSIMUL1	HOLIDAY0XSIMUL4	HOLIDAY0XSIMUL1
0	1	1	0	0
0	4	0	1	0
1	1	0	0	1

CHOI_SHOW(최유라 쇼) 프로그램 여부 변수 생성

최유라 대표 프로그램

★LUCKY 7_최유라쇼
 ★넘버원 토요일_최유라
 ★더드림박람회_Kitchen_최유라
 ★딱이틀만! 2번쓴다_최유라쇼
 ★설마중 복이평평_최유라쇼
 ★최유라쇼_설마음한상
 ★최유라쇼_설마중
 최유라 쇼
 최유라쇼_16대브랜드
 최유라쇼_1등상품대전
 최유라쇼_50주년성원감사전
 최유라쇼_SUPERWEEK
 최유라쇼_★추석마음한상★
 최유라쇼_내고향상품전
 최유라쇼_16대브랜드
 최유라쇼_리빙쇼 2018
 최유라쇼_브랜드결산
 최유라쇼_슈퍼워크
 최유라쇼_추석_내고향상품전
 투게더 리그전_최유라쇼

일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR_SALE_AMT PUR_SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7936614.71	7936614.71	3257.50	<.0001
Error	29075	70838597.73	2436.41		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR_SALE_AMT Mean
0.100750	77.87765	49.36000	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
CHOI_SHOW	1	7936614.710	7936614.710	3257.50	<.0001

환경: SAS 시스템('SASApp', Linux), 생성 일시: 2019년 08월 28일, 오후 07

페이지 나누기

일원분산분석

결과

The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
CHOI_SHOW	1	565.3	<.0001

Welch's ANOVA for PUR_SALE_AMT

Source	DF	F Value	Pr > F
CHOI_SHOW	1.0000	1389.68	<.0001
Error	1043.3		

HOT_CHOICE 프로그램 여부 변수 생성

Hot Choice
프로그램

Hot Choice

일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR_SALE_AMT PUR_SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1354855.57	1354855.57	508.81	<.0001
Error	29075	77420356.87	2662.78		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR_SALE_AMT Mean
0.017199	81.41519	51.60214	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
HOTCHOICE	1	1354855.567	1354855.567	508.81	<.0001

rated by the SAS System ('SASApp', Linux) on 2019년 08월 30일 at 11:06:1

페이지 나누기

일원분산분석

결과

The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
HOTCHOICE	1	1825.7	<.0001

Welch's ANOVA for PUR_SALE_AMT

Source	DF	F Value	Pr > F
HOTCHOICE	1.0000	23230.5	<.0001
Error	5051.2		

컨템포러리 프로그램 여부 변수 생성

컨템포러리
프로그램

컨템포러리

일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR_SALE_AMT PUR_SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1753489.64	1753489.64	661.93	<.0001
Error	29075	77021722.80	2649.07		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR_SALE_AMT Mean
0.022259	81.20532	51.46912	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
컨템포러리	1	1753489.644	1753489.644	661.93	<.0001

Generated by the SAS System ('SASApp', Linux) on 2019년 08월 30일 at 11:06:5

페이지 나누기

일원분산분석

결과

The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
컨템포러리	1	1161.0	<.0001

Welch's ANOVA for PUR_SALE_AMT			
Source	DF	F Value	Pr > F
컨템포러리	1.0000	1836.85	<.0001
Error	3713.1		

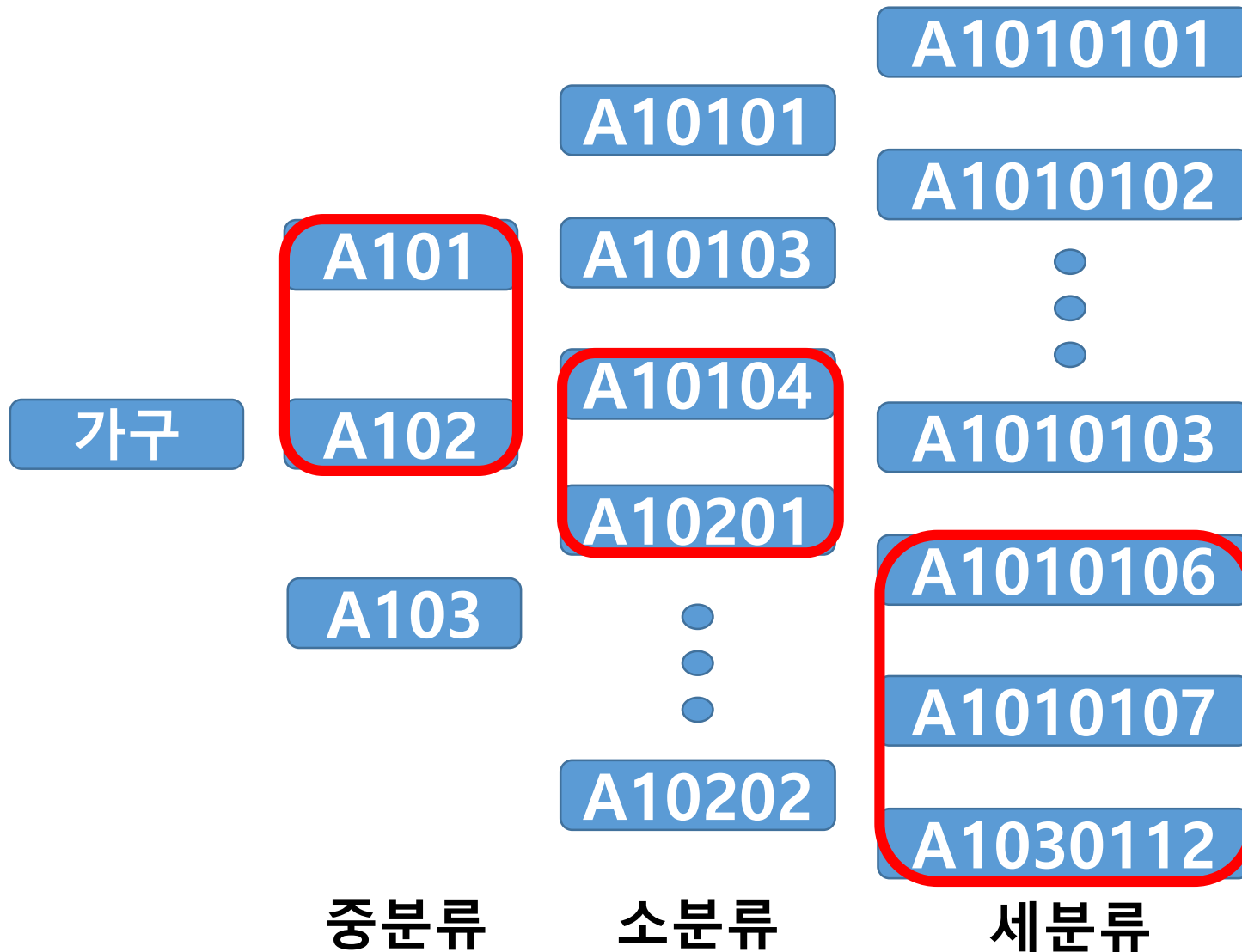
대분류 별 대, 중, 소(NEW_MID, NEW_SML, NEW_THN) 분류기준 마련

GOODS_CLSS_MID_CD	빈도
K101	24
K102	10
K103	79
K104	33
K105	56
K107	150
K108	455
K109	79

LEVEL 65개
빈도 30개 미만 15개

매출액 기준
군집분석 실시

대분류 별 대, 중, 소(NEW_MID, NEW_SML, NEW_THN) 분류기준 마련



매출액을 기준으로 군집분석을 실시하여 중분류 8개, 소분류 17개, 세분류 19개 군집으로 묶음

TVGGRP_SALE_PRC(TV상품그룹판매가) 예측

예측 기준	1	TV_GOODS_GRP_CD	BRND_MID_CD	GOODS_CLSS_THN_CD
	2		BRND_MID_CD	GOODS_CLSS_THN_CD
	3			GOODS_CLSS_THN_CD
	4			GOODS_CLSS_SML_CD
	5			GOODS_CLSS_MID_CD
	6			GOODS_CLSS_LRG_CD

같은 기준을 가지고 있는 상품의 평균과 중간값의 평균으로 가격 예측

TVGGRP_SALE_PRC(TV상품그룹판매가) 예측

2013~2017	TV_GOODS_GRP_CD	BRND_MID_CD	GOODS_CLSS_THN_CD	TVGGRP_SALE_PRC (P)
	00001	100890	N1060101	1880000
	00001	100890	N1060101	1520000
	00001	100890	N1060101	1520000
	00001	100890	N1060101	1880000
				AVG(AVG(P),MEDIEN(P))
2018	TV_GOODS_GRP_CD	BRND_MID_CD	GOODS_CLSS_THN_CD	TVGGRP_SALE_PRC (P)
	00001	100890	N1060101	
	00001	100890	N1060101	

TVGGRP_SALE_PRC(TV상품그룹판매가) 2017 검증 과정



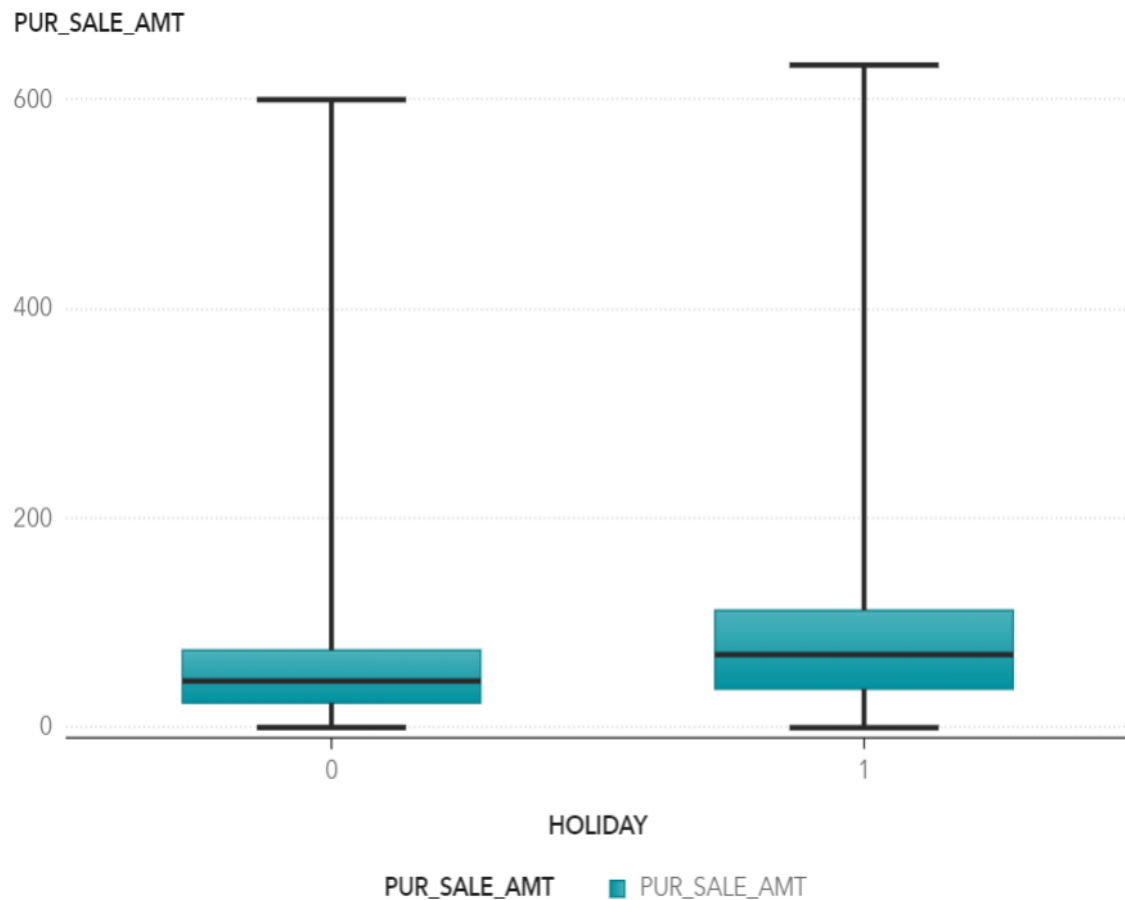
이러한 기준을 갖고 계산했을 때 ACC(정확도)약 70%가 나옴

AVG_of_BEST_PRED_ACC
0.7039461766

외부변수 생성

휴일(HOLIDAY) 변수 생성

가정 : 휴일(주말, 공휴일 포함) 여부에 따른 매출액의 차이가 난다.



일원분산분석

결과

The ANOVA Procedure

Dependent Variable: PUR_SALE_AMT PUR_SALE_AMT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3892969.90	3892969.90	1511.55	<.0001
Error	29075	74882242.55	2575.49		
Corrected Total	29076	78775212.44			

R-Square	Coeff Var	Root MSE	PUR_SALE_AMT Mean
0.049419	80.06953	50.74924	63.38147

Source	DF	Anova SS	Mean Square	F Value	Pr > F
HOLIDAY	1	3892969.896	3892969.896	1511.55	<.0001

환경 : SAS 시스템('SASApp', Linux), 생성 일시 : 2019년 08월 26일, 오후 02

페이지 나누기

일원분산분석

결과

The ANOVA Procedure

Bartlett's Test for Homogeneity of PUR_SALE_AMT Variance

Source	DF	Chi-Square	Pr > ChiSq
HOLIDAY	1	829.1	<.0001

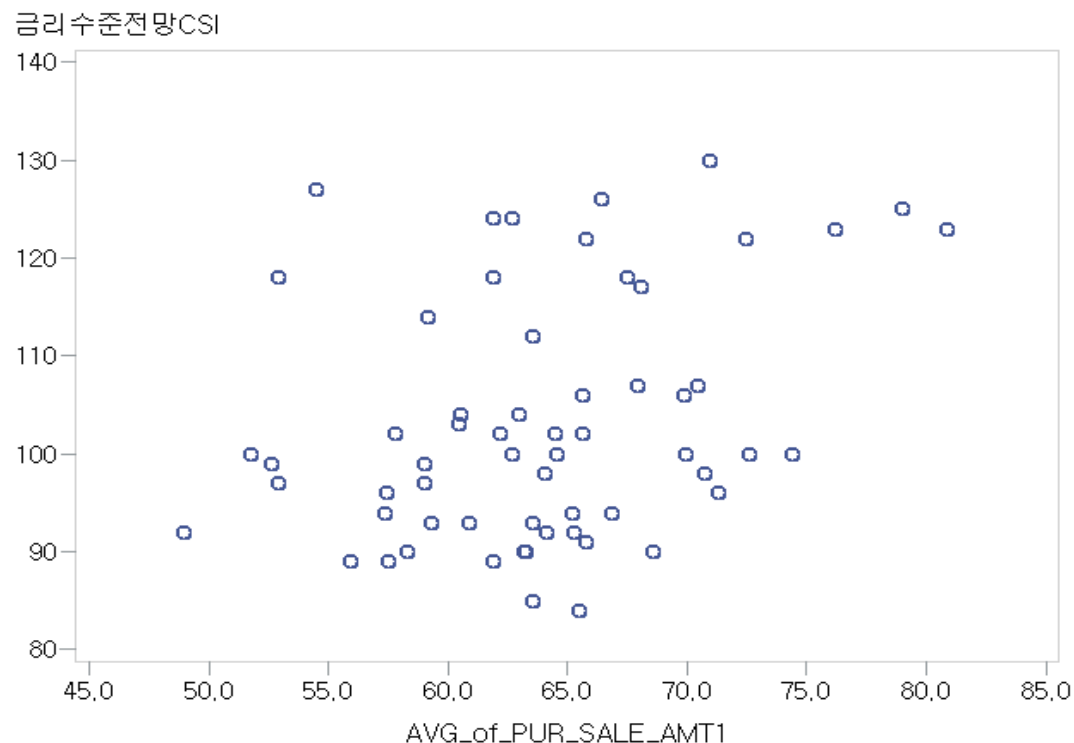
Welch's ANOVA for PUR_SALE_AMT

Source	DF	F Value	Pr > F
HOLIDAY	1.0000	1230.37	<.0001
Error	13024.0		

금리수준전망CSI 변수 생성

- 가정 : 1) 경기 전망을 긍정적으로 평가하면 한달 후 매출액에 영향을 미친다.
2) 선행지표이기 때문에 따로 예측하지 않고 한달 전 지표를 사용

산점도



상관분석

CORR 프로시저

1 개의 With 변수:	금리수준전망CSI
1 개의 변수:	AVG_of_PUR_SALE_AMT1

단순 통계량

변수	N	평균	표준편차	합	최솟값	최댓값
금리수준전망CSI	34237	106.50296	14.52830	3646342	84.00000	135.00000
AVG_of_PUR_SALE_AMT1	29077	63.38147	6.46655	1842943	48.93681	80.82796

피어슨 상관 계수
H0: $\rho=0$ 가정하에서 $\text{Prob} > |r|$
관측값 개수

	AVG_of_PUR_SALE_AMT1
금리수준전망CSI	0.32641
	<.0001
	29077

Generated by the SAS System ('SASApp', Linux) on 2019년 08월 30일 at 12:15:22 PM

날씨변수 생성

가정 : 기후에 따른 계절성을 지닌 상품의 매출액에 영향을 받을 것이다.

속성

일별 최저/최고온도 및 한파/폭염특보 파생변수 생성

기준

2013년부터 2017년까지 실측치를 통해 모델을 형성

예측시기인 2018에 대해서는 기상청의 중기예보 데이터를 활용

한파/폭염특보 파생변수는 기상청의 기상특보 발표기준을 참고

날씨변수 생성(2013년~2017년)

가정 : 기후에 따른 계절성을 지닌 상품의 매출액에 영향을 받을 것이다.

▪ 자료구분 ▪ 자료형태 ▪ 기간 ~
 ▪ 지역/지점

날짜	지점	평균기온(°C)	최저기온(°C)	최고기온(°C)
2016-01-06	전국	0.7	-4.2	5.4
2016-01-06	서울경기	-1.1	-4.5	2.5
2016-01-06	강원영동	1.2	-2.6	6.1
2016-01-06	강원영서	-3.7	-9.1	2.3
2016-01-06	충북	-1.7	-7.7	3.3
2016-01-06	충남	-0.1	-5.8	4.4
2016-01-06	경북	1	-3.6	5.5
2016-01-06	경남	2.6	-2.5	8
2016-01-06	전북	1.2	-3.9	5.8
2016-01-06	전남	3.6	-1	7.7
2016-01-06	제주	8.3	6.3	10.6

지점별 기온 실측자료 병합결과[2016-01-06기준]

지역/지점은 전국 및 기타 쏘지역(서울경기, 강원영동, 강원영서, 충북, 충남, 경북, 경남, 전북, 전남, 제주)의 자료를 모두 다운로드하여 병합

날씨변수 생성(2018년)

가정 : 기후에 따른 계절성을 지닌 상품의 매출액에 영향을 받을 것이다.

최저/최고기온(℃)

- 그래프를 클릭하면 최저/최고기온과 변동성 정보를 확인하실 수 있습니다.

지역	도시	02일 (금)	03일 (토)	04일 (일)	05일 (월)	06일 (화)	07일 (수)	08일 (목)	09일 (금)
서울 - 인천 - 경기도	서울	24 / 31	25 / 33	25 / 33	25 / 34	25 / 34	24 / 33	25 / 33	25 / 33
	인천	24 / 29	25 / 31	25 / 31	25 / 32	25 / 32	25 / 31	24 / 31	24 / 31
	수원	24 / 31	24 / 33	24 / 33	25 / 33	25 / 33	25 / 33	25 / 33	25 / 33
	파주	24 / 31	24 / 33	24 / 33	24 / 34	24 / 34	24 / 33	24 / 32	24 / 32
	이천	23 / 31	23 / 33	23 / 33	24 / 34	24 / 34	24 / 33	24 / 32	24 / 32
	평택	24 / 31	24 / 33	24 / 33	25 / 34	25 / 34	25 / 33	25 / 33	25 / 33

기상청 날씨 누리 중기예보 [2019-07-30 오전 10시 기준]

예측하고자 하는 시기(2018년)에 대해서는 기상청의 2018년 최저/최고기온 중기예보 자료로 최종 예측을 산출

주의보/경보 변수 생성

가정 : 기후에 따라 계절성을 지닌 상품의 매출액에 영향을 받을 것이다.

종류	주의보	경보
한파	10월~4월에 다음 중 하나에 해당하는 경우 ① 아침 최저기온이 전날보다 10℃ 이상 하강하여 3℃ 이하이고 평년값보다 3℃가 낮을 것으로 예상될 때 ② 아침 최저기온이 -12℃ 이하가 2일 이상 지속될 것이 예상될 때 ③ 급격한 저온현상으로 중대한 피해가 예상될 때	10월~4월에 다음 중 하나에 해당하는 경우 ① 아침 최저기온이 전날보다 15℃ 이상 하강하여 3℃ 이하이고 평년값보다 3℃가 낮을 것으로 예상될 때 ② 아침 최저기온이 -15℃ 이하가 2일 이상 지속될 것이 예상될 때 ③ 급격한 저온현상으로 광범위한 지역에서 중대한 피해가 예상될 때
폭염	일최고기온이 33℃ 이상인 상태가 2일 이상 지속될 것으로 예상될 때	일최고기온이 35℃ 이상인 상태가 2일 이상 지속될 것으로 예상될 때

기상청 기상특보 발표기준

중기예보로 아침 최저기온을 예측하지 못하므로 아침 최저기온을 해당 날짜의 최저기온으로 봄. 첫번째 한파특보 발령기준(①)을 고려하지 않음.

홈쇼핑 매출액 예측을 위한 입력변수 설정(1)

변수이름	유형	레벨	비고
NEW_MID	숫자	Nominal	
NEW_SML	숫자	Nominal	
NEW_THN	숫자	Nominal	
WEEKNUM_NUM	숫자	Interval	
TERM_MIN (TERM_MIN/SIMUL)	숫자	Interval	
WEEKDAY_NUM	숫자	Nominal	
SIMUL	숫자	Nominal	
PRIMETIME	숫자	Binary	
TVGRP_SALE_PRC	숫자	Interval	2018 예측

홈쇼핑 매출액 예측을 위한 입력변수 설정(2)

변수이름	유형	레벨	비고
M0	숫자	Interval	
M10	숫자	Interval	
M20	숫자	Interval	
M30	숫자	Interval	
M40	숫자	Interval	
M50	숫자	Interval	
M60	숫자	Interval	
M70	숫자	Interval	
M80	숫자	Interval	
M90	숫자	Interval	
M100	숫자	Interval	
M110	숫자	Interval	
M120	숫자	Interval	

홈쇼핑 매출액 예측을 위한 입력변수 설정(3)

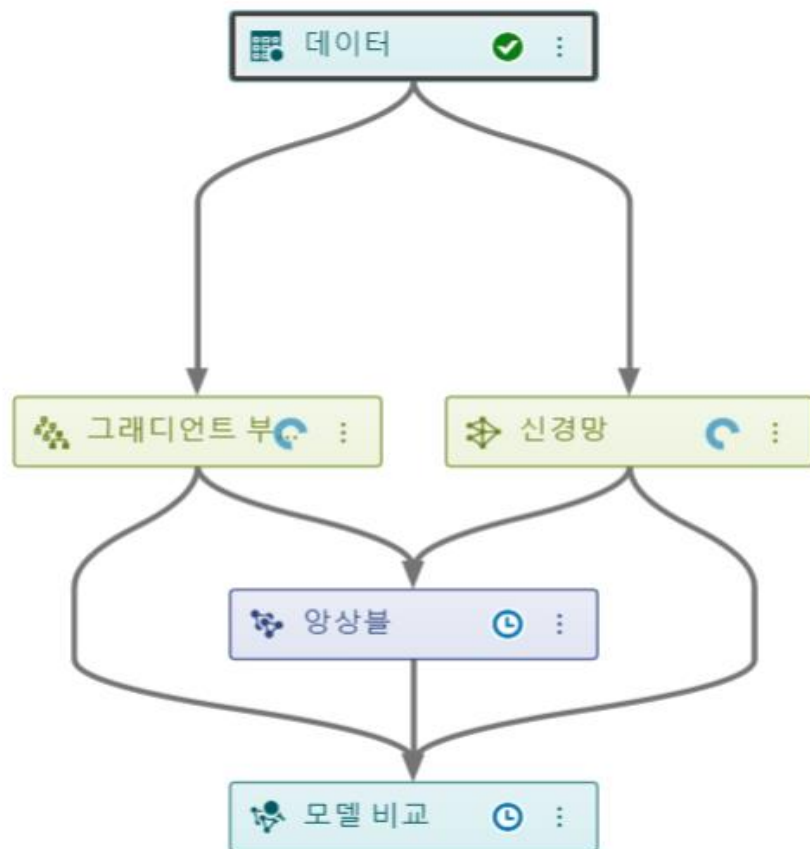
변수이름	유형	레벨	비고
M130	숫자	Interval	
M140	숫자	Interval	
M150	숫자	Interval	
M160	숫자	Interval	
M170	숫자	Interval	
M180	숫자	Interval	
M190	숫자	Interval	
M200	숫자	Interval	
M210	숫자	Interval	
M220	숫자	Interval	
M230	숫자	Interval	

홈쇼핑 매출액 예측을 위한 입력변수 설정(4)

변수이름	유형	레벨	비고
HOLIDAY	숫자	Binary	
금리수준전망CSI	숫자	Interval	
최저기온	숫자	Interval	
최고기온	숫자	Interval	
폭염경보	숫자	Interval	
한파주의보	숫자	Interval	
한파경보	숫자	Interval	
대분류XWEEKDAY	숫자	Binary	교호작용
대분류XHOLIDAY	숫자	Binary	교호작용
SIMULXWEEKDAY	숫자	Binary	교호작용
SIMULXHOLIDAY	숫자	Binary	교호작용

모델링

예측 모델링



그래디언트 부스팅 모델과 신경망
모델 각각의 단점을 보완할 수 있는 앙
상블 모델을 예측모델로 채택

활용방안 및 경영전략

최유라 쇼





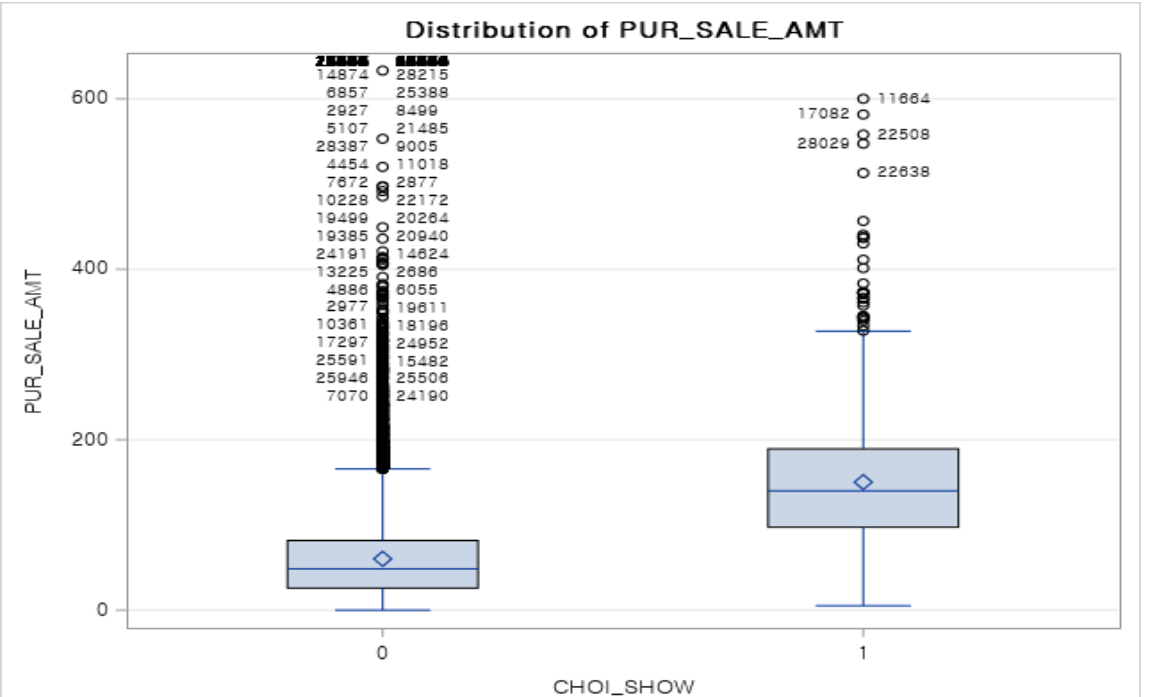
최유라쇼

최유라가 제안하는
건강한 라이프 스타일

목요일 저녁 8:40 | 토요일 아침 8:20

🔔 알림신청

📷 인스타그램



CHOI_SHOW	PUR_SALE_AMT의 평균
0	60.2
1	150.3

제 2의 최유라



제 2의 최유라를 어떻게 찾아야 할까?

제 2의 최유라를 어떻게 찾아야 할까?

롯데 홈쇼핑에게 우량한 쇼호스트는 누구인가?



쇼 호스트

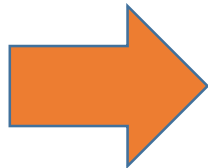
RECENCY
(최근성)

FREQUENCY
(빈도)

MONETARY
(매출액)

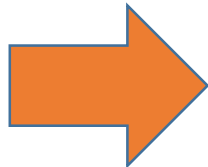
쇼 호스트 RFM 분석

RECENCY



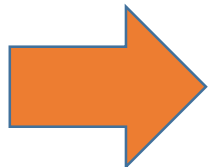
쇼 호스트's 최근 방송 날짜

FREQUENCY



쇼 호스트's 방송 빈도

MONETARY



쇼 호스트's 평균매출액

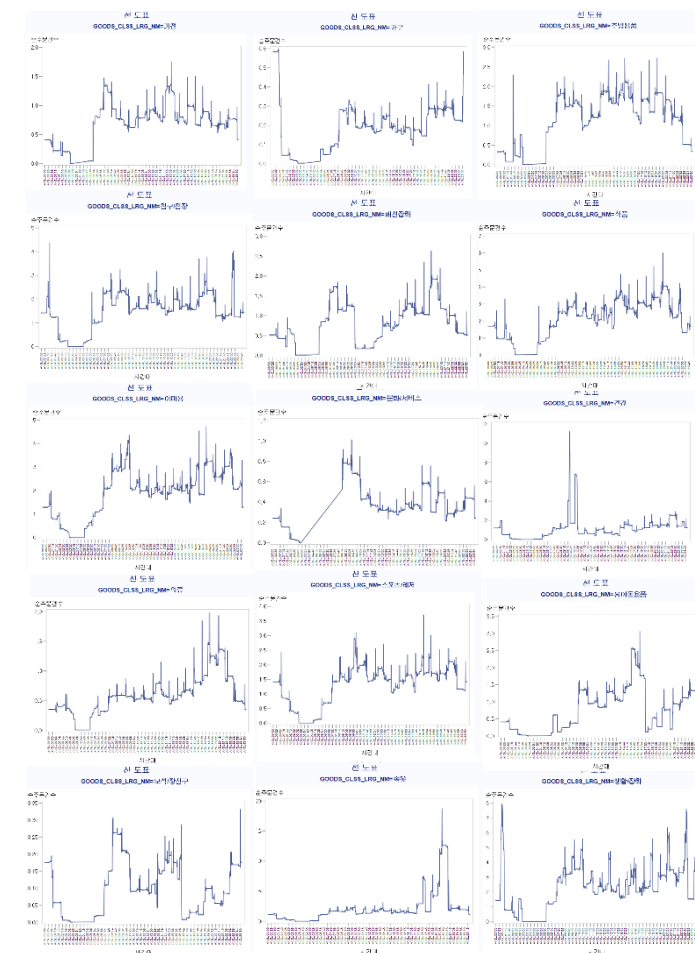
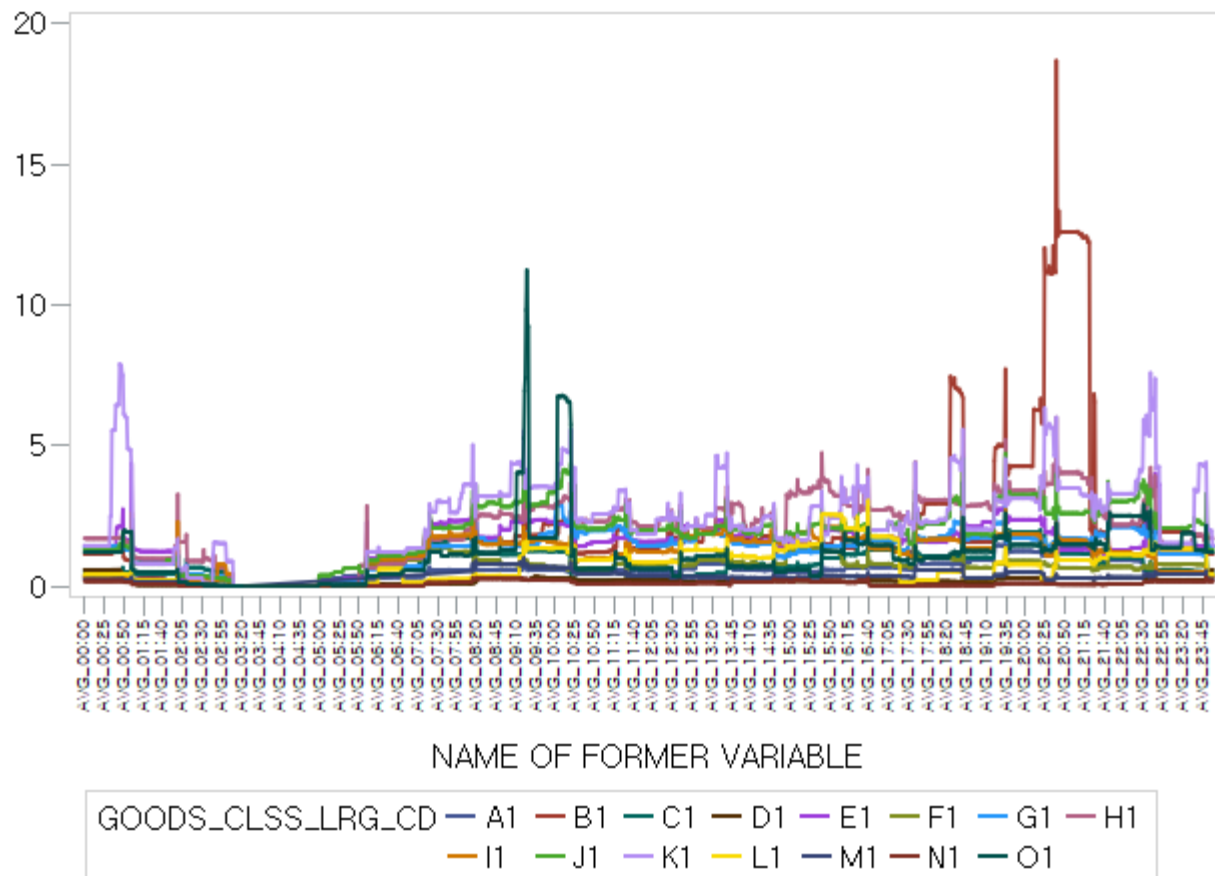
우량한 쇼 호스트 네임드 쇼 호스트로 전환

SHOW HOST	프로그램명
F00027	컨템포러리
F00027	언더웨어
F00027	영캐주얼
F00027	슈즈
F00027	캐릭터
F00027	핸드백/잡화
F00027	레디고!바캉스
F00027	교육문화서비스

네임드
쇼 호스트
전환

SHOW HOST	프로그램명
F00027	김현학 쇼

실패하지 않고도 성공할 수는 없을까?



대분류의 시간대별 순주문건수 예측에 집중하여
특정 그룹의 신규 편성시 참고하자!

감사합니다.