



고객 상품구매 주기패턴을 고려한 고객맞춤형 쿠폰제시 방안

경영학과 박규리  
통계학과 방채연



- L.POINT현황
- 문제해결을 위한 분석과정 도출
- R을 이용한 데이터 분석 과정
- 데이터분석 결과를 이용한 시사점 및 서비스 제안

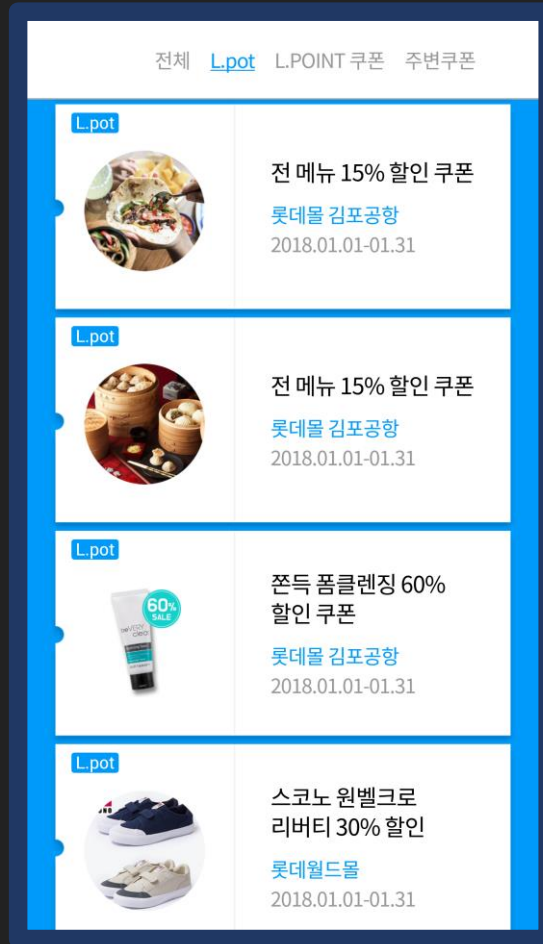
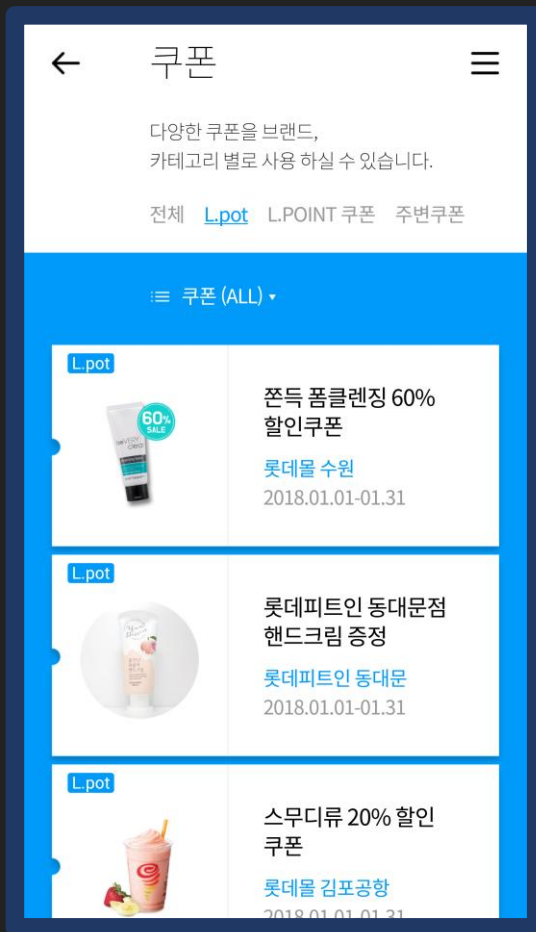
# L.POINT 문제인식

L.POINT의 현재 쿠폰제시현황은 어떤 문제를 가지고 있는가



# L.POINT 문제인식

## ① 고객 맞춤형 쿠폰 제시가 아닌 획일화된 쿠폰 제시



고객의 특징(구매빈도, 구매주기, 구매패턴 등)을 고려하여 쿠폰을 진열하지 않는 L.POINT 쿠폰 카테고리 현황



매우 비효율적인 시스템화 되어 있는 L.POINT 쿠폰 카테고리



## L.POINT 문제인식

### ② 쿠폰의 목록화 된 제공으로 겪는 사용자들의 불편함

쿠폰이 카테고리 안에 목록으로 제공되어,  
고객들이 직접 쿠폰을 번거롭게 찾아야하는  
L.POINT 쿠폰 제공 현황



고객들이 매우 **불편하게 찾아야하는**  
L.POINT 쿠폰 제공 시스템



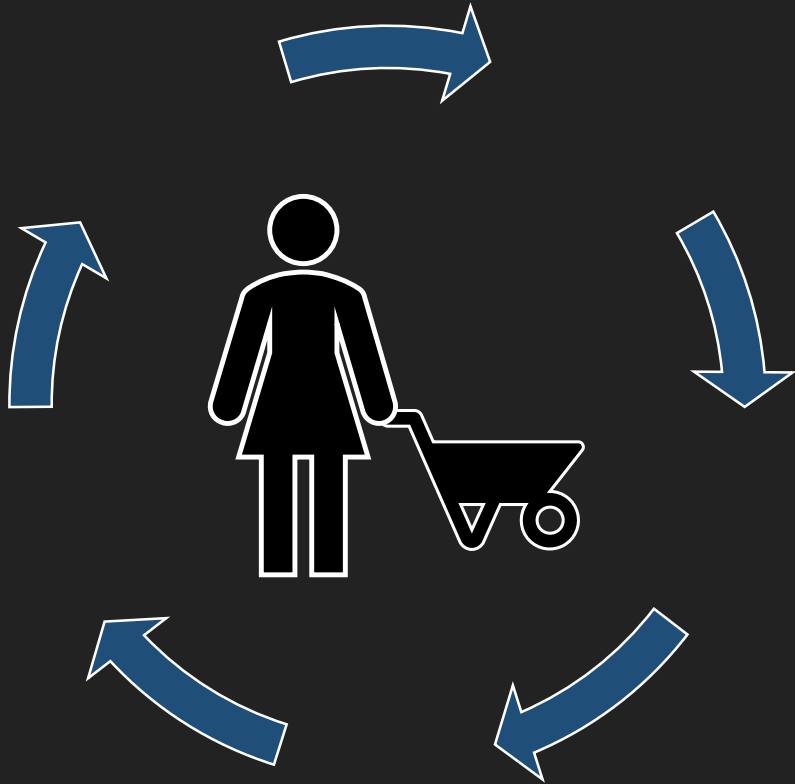
# 문제해결을 위한 분석과정 도출

L.POINT의 이 현황을 개선하기 위해서 무엇을 분석해야 하는가



## 문제해결을 위한 분석과정 도출

### ① 고객 각각의 구매주기 계산



같은 상품이라도 고객마다 구매하는 주기가 다르다

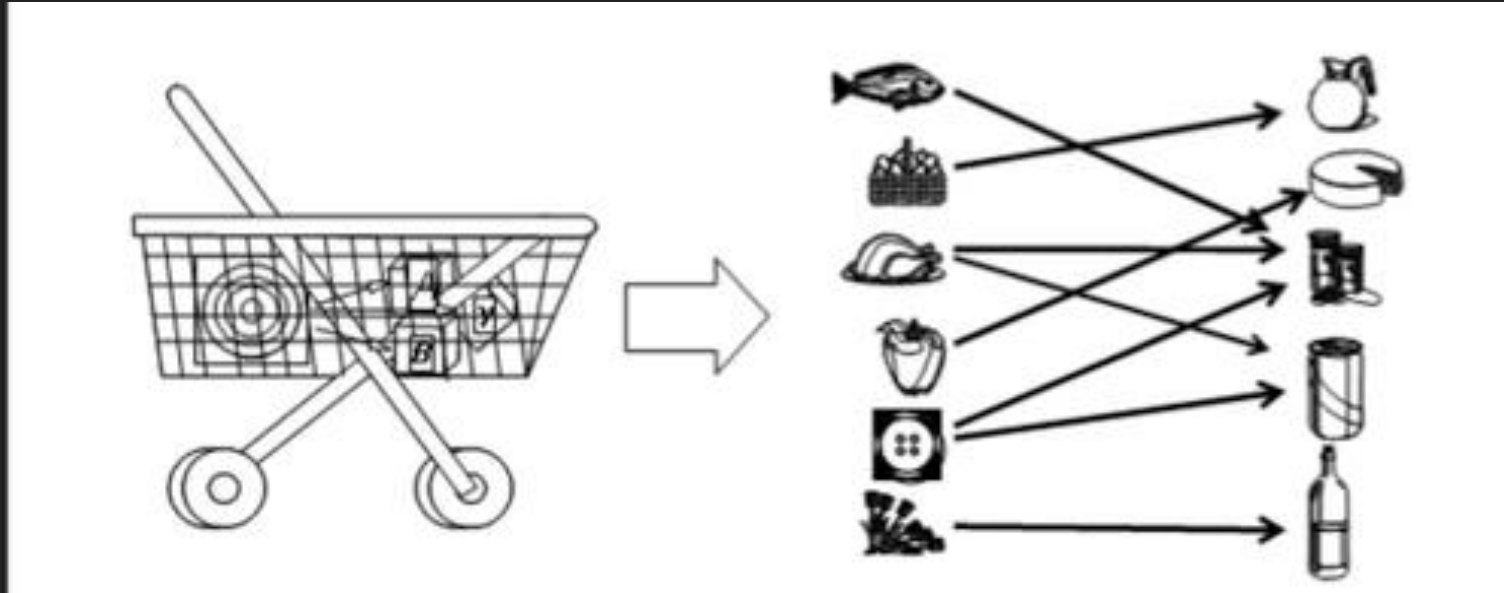


그러므로 각각의 고객마다  
**‘중복구매상품의 구매주기’**를 계산하자



## 문제해결을 위한 분석과정 도출

### ② 구매하는 상품별로 연관분석



고객들은 마트에서 쇼핑을 할 때 여러 상품을 동시에 담는다  
그러므로 상품들을 통째로 묶어 **‘상품끼리 연관분석’**을 실시하자





## 문제해결을 위한 분석과정 도출

### ② 분석데이터를 대형마트에 한정



【 주기적인 구매 패턴을 보이고, 한번에 다양한 상품을 구매하는  
가장 적합한 업종이 '대형마트'이기 때문에 대형마트만 분석 】

# R을 이용한 데이터 분석 과정

데이터 분석에서 직접 이용된 R코드



# R을 이용한 데이터 분석 과정

## ① 고객의 상품구매 주기계산 R코드

```
#대형마트 구매 데이터 불러오기
shop<-read.table("purshopping.txt",header=T,sep=",")
shop<-shop[,c(1,3,4,6,9)]
shop<-subset(shop,BIZ_UNIT=='A02')
head(shop)
summary(shop)
```

```
#중복구매 이력이 있는 회원만 골라내어 별도 파일에 저장
shoplist<-shop[,c(1,3)]
head(shoplist)
a<-shop[which(duplicated(shoplist) | duplicated(shoplist, fromLast=T) ),]
dupshop<-a[c(order(a$ID,a$PD_S_C,a$DE_DT)),]
head(dupshop)
rownames(dupshop)<-NULL
write.table(dupshop,'C:/Users/ATIV/Documents/dupshop.txt',sep=",",row.names=FALSE)
```



# R을 이용한 데이터 분석 과정

## ① 고객의 상품구매 주기계산 R코드

```
#구매날짜열을 날짜변수로 변환 후 날짜차이값 계산을 위해 df열 추가
dupshop$DE_DT<-strptime(dupshop$DE_DT,"%Y%m%d")
head(dupshop)
df<-rep(0,nrow(dupshop))
dupshop<-cbind(dupshop,df)

#한 사람이 제품 구매 후 동일한 제품을 중복구매하기까지 걸리는 시간 계산하여 df열에 입력
for (i in 2:nrow(dupshop)){
  if (dupshop[i,3]==dupshop[i-1,3]){

    dupshop[i,'df']<-dupshop[i,4]-dupshop[i-1,4]
  }
}
head(dupshop,20)
write.table(dupshop,'C:/Users/ATIV/Documents/dupshop2.txt',sep=" ",row.names=FALSE)
```



# R을 이용한 데이터 분석 과정

## ① 고객의 상품구매 주기계산 R코드

```
#의미있는 주기만을 보기위해 동일한 제품을 5회 이상 구매한 이력만 골리내고 df=0인 데이터는 제거
library(dplyr)
dupshop<-read.table("dupshop2.txt",header=T,sep=",")
head(dupshop,15)
dupshop.sum<-dupshop %>% group_by(ID,PD_S_C) %>% summarise(sumbuy=sum(BUY_CT))
dupshopa<-merge(dupshop.sum,dupshop,by.x=c('ID','PD_S_C'),by.y=c('ID','PD_S_C'))
dupshopb<-dupshopa[c(order(dupshopa$ID,dupshopa$PD_S_C,dupshopa$DE_DT)),]
dupshopc<-subset(dupshopb,sumbuy>=5)
dupshopd<-subset(dupshopc,df!=0)

#df의 평균과 중앙값 각각 계산
df.mean<-dupshopd %>% group_by(ID,PD_S_C) %>% summarise(df_mean=mean(df))
df.median<-dupshopd %>% group_by(ID,PD_S_C) %>% summarise(df_median=median(df))
head(df.mean,20)
head(df.median,20)
```



# R을 이용한 데이터 분석 과정

## ① 고객의 상품구매 주기계산 R코드

#중앙값으로 선택 예시

```
subset(df.mean,ID==14296&PD_S_C==1444)
```

```
subset(df.median,ID==14296&PD_S_NM=='떠먹는요구르트')
```

```
subset(dupshop,ID==14296&PD_S_C==1444)
```

#중앙값으로 선택 후 음식물봉투, 재사용봉투, 종량제봉투 제거

```
df.median<-merge(df.median,dupshop.sum,by.x=c('ID','PD_S_C'),by.y=c('ID','PD_S_C'))
```

```
df.median<-subset(df.median,PD_S_C!=678&PD_S_C!=679&PD_S_C!=680)
```

```
df.mean<-subset(df.mean,PD_S_C!=678&PD_S_C!=679&PD_S_C!=680)
```

```
rownames(df.mean)<-NULL
```

#쇼핑업종분류 데이터 불러오기

```
shopclass<-read.csv('shopclass.csv',header=T)
```

```
head(shopclass)
```

```
shopclass<-subset(shopclass,BIZ_UNIT=='A02')
```

```
shopclass<-shopclass[,2:3]
```



# R을 이용한 데이터 분석 과정

## ① 고객의 상품구매 주기계산 R코드

```
#가독성을 위해 상품코드를 상품명으로 대체 후 별도 파일로 저장
df.median.mer<-merge(df.median,shopclass,by.x='PD_S_C',by.y='PD_S_C')
df.median.mer<-df.median.mer[order(df.median.mer$ID),]
df.median<-df.median.mer[,c('ID','PD_S_NM','sumbuy','df_median')]
head(df.median,5)
write.table(df.median,'C:/Users/ATIV/Documents/df_median.txt',sep="," ,row.names=FALSE)
```



# R을 이용한 데이터 분석 과정

## ② 상품 연관분석 R코드

```
#데이터 삽입
b<-read.table('b.txt',sep=',',header=T)
f<-read.csv('f.csv',sep=',',header=T)
#점포코드, 상품소분류코드,소분류명 추출
f_so<-f[c(1,2,3)]
#고객 식별번호, 점포코드, 상품소분류코드, 구매가 발생한 일자 추출
b<-b[,c(1,3,4,6)]
#쇼핑업종상품구매정보에서 대형마트의 정보 추출
b_A02<-subset(b, BIZ_UNIT=='A02')
#쇼핑업종 상품분류정보에서 대형마트의 정보 추출
f_so_A02<-subset(f_so,BIZ_UNIT=='A02')
#대형마트 정보만을 기반으로 쇼핑업종상품구매정보와 쇼핑업종 상품분류정보 합침
mer_A02<-merge(b_A02,f_so_A02,by.x='PD_S_C',by.y='PD_S_C')
#저장
write.csv(mer_A02,'A02.csv'))
```





# R을 이용한 데이터 분석 과정

## ② 상품 연관분석 R코드

```
#ID와 날짜별로 정렬
mer_A02_order<-mer_A02[order(mer_A02$ID,mer_A02$DE_DT),]
#ID와 날짜, 상품소분류코드 추출
mer_A02_order<-mer_A02_order[c('ID','DE_DT','PD_S_NM')]
#상품소분류코드 자료형 캐릭터로 바꾸기
mer_A02_order[,3]<-as.character(mer_A02_order[,3])
#첫줄 넣어주기
ss<-mer_A02_order[1,3]
#매트릭스로 바꾸기
ss<-as.matrix(ss)
#행이 몇개인지 파악
nrow(mer_A02_order)
```



# R을 이용한 데이터 분석 과정

## ② 상품 연관분석 R코드

```
# ID와 날짜가 같으면 콤마로 한줄로 붙이고, 다르면 다음행으로 붙이는 함수 돌리기
for (i in 1:1727091){
  if (mer_A02_order[i,1]==mer_A02_order[i+1,1]&mer_A02_order[i,2]==mer_A02_order[i+1,2]){
    ss[nrow(ss),1]<-paste(ss[nrow(ss),1],mer_A02_order[i+1,3],sep=',')}
    else {
      ss<-rbind(ss,mer_A02_order[i+1,3])
    }
  }
}

#변수이름 바꾸기
final<-ss
```



# R을 이용한 데이터 분석 과정

## ② 상품 연관분석 R코드

```
#재사용봉투 없애기  
final1<-str_replace_all(final,"재사용봉투","")
```

```
#재사용 봉투 있는 채로 콤마로 분리  
q<-lapply(final, strsplit, split=",")  
#재사용 봉투 없는 채로 콤마로 분리  
q_1<-lapply(final1, strsplit, split=",")
```

```
#list 풀기  
q<-lapply(q, unlist)  
#list 풀기  
q_1<-lapply(q_1, unlist)
```



# R을 이용한 데이터 분석 과정

## ② 상품 연관분석 R코드

```
#재사용봉투를 포함한 채로 연관분석
rule<-apriori(q, parameter =list(support=0.003, confidence = 0.2, minlen =2))
#재사용봉투 포함하지 않은 채로 연관분석
rule2<-apriori(q_1, parameter =list(support=0.003, confidence = 0.2, minlen=2))
#연관분석 한 것 보기
inspect(rule)
inspect(rule2)
#연관분석 한 것 데이터프레임으로 변환
rule3<-as.data.frame(inspect(rule2))
rule4<-as.data.frame(inspect(rule))
#연관분석한 것 저장
write.csv(rule3,'rule.csv')
write.csv(rule4,'rule2.csv')
#연관분석 시각화
plot(rule2, method="graph", control=list(type="items"))
```

# 데이터분석 결과를 이용한 시사점 및 서비스 제안

데이터분석 결과는 어떻게 해석하고 어떻게 활용해야 하는가



# 시사점 및 서비스 제안

## ① 분석 단계의 특이점1 – 평균주기는 중간값으로 계산

```
> subset(df.mean,ID==14296&PD_S_C==1444)
  ID PD_S_C df_mean
82659 14296  1444 31.625
> subset(df.median,ID==14296&PD_S_NM=='떠먹는요구르트')
  ID    PD_S_NM sumbuy df_median
86013 14296 떠먹는요구르트    9    19.5
> subset(dupshop,ID==14296&PD_S_C==1444)
  ID BIZ_UNIT PD_S_C  DE_DT BUY_CT  df
1037380 14296  A02  1444 2015-01-12   1   0
1037381 14296  A02  1444 2015-02-17   1  36
1037382 14296  A02  1444 2015-03-12   1  23
1037383 14296  A02  1444 2015-03-27   1  15
1037384 14296  A02  1444 2015-04-07   1  11
1037385 14296  A02  1444 2015-05-01   1  24
1037386 14296  A02  1444 2015-05-17   1  16
1037387 14296  A02  1444 2015-09-09   1 115
1037388 14296  A02  1444 2015-09-22   1  13
```

떠먹는 요구르트의 경우  
대부분 날짜차이가 15-20일 근방  
그러나 115일이라는 이상치로 인해  
평균값 왜곡



따라서 주기 계산에서 **평균이 아닌**  
이상치에 영향을 받지 않는  
**중앙값을 선정**

[ 고객들의 구매주기를 계산할 때 '평균 주기'는 '중간값'으로 계산 ]



## 시사점 및 서비스 제안

### ① 분석 단계의 특이점2 – 연관분석 과정에서 ‘재사용봉투’ 제외

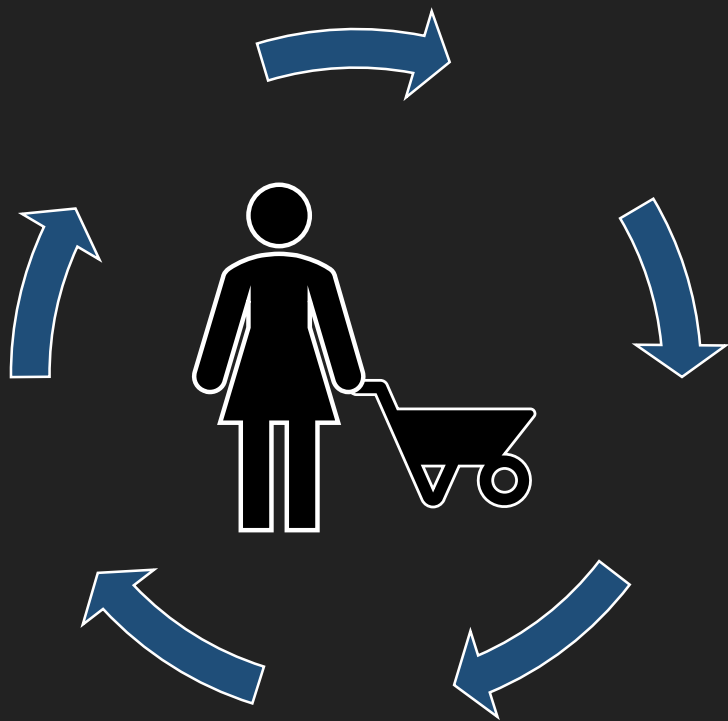
{절임반찬}	=>	{재사용봉투}
{찌개두부}	=>	{재사용봉투}
{설거지용품}	=>	{랩/호일}
{설거지용품}	=>	{다이소}
{설거지용품}	=>	{재사용봉투}
{칫솔}	=>	{다이소}
{칫솔}	=>	{재사용봉투}
{비엔나소시지}	=>	{일반흰우유}
{비엔나소시지}	=>	{붕지라면}
{비엔나소시지}	=>	{재사용봉투}
{가정용화장지}	=>	{붕지라면}
{검용두부}	=>	{일반흰우유}

연관분석을 할 때 ‘재사용봉투’가 다수 나오지만 고객들의 구매패턴 분석에  
의미없는 결과값이라고 판단하여 **연관분석 과정에서 ‘재사용봉투’ 제외**



## 시사점 및 서비스 제안

① 분석 단계의 특이점3 – 주기계산 과정에서 5회 미만 중복구매 이력 제외



**5회 이상**

[ 5회 미만 구매한 이력을 주기적인 구매 패턴으로 보기 힘들다고 판단하여  
**5회 이상 중복구매 데이터만 사용** ]





## 시사점 및 서비스 제안

### ① 분석 단계의 특이점4 – 연관분석 과정에서 신뢰도와 지지도

신뢰도  
**20%**

물품A를 구매한 사람 중에서 물품B도  
함께 구매한 비율이 최소한 **20% 이상**  
이어야 연관성이 있다고 생각

지지도  
**0.3%**

전체 거래 중 물품A와 물품B를 함께  
구매한 비율이 최소한 **0.3% 이상**  
이어야 연관성이 있다고 생각



## 시사점 및 서비스 제안

### ② 분석결과를 바탕으로 한 시사점 - 상품 연관분석

{구이진미}	=>	{훈제}
{훈제}	=>	{구이진미}
{스틱과자}	=>	{일반스낵}
{쌀스낵}	=>	{일반스낵}
{일반시리얼}	=>	{일반흰우유}
{여성화장품}	=>	{일반화장품}
{3분요리류}	=>	{봉지라면}
{수입면류}	=>	{소스}
{일반물티슈}	=>	{다이소}
{PB바나나}	=>	{일반흰우유}

[ 상품 연관분석 결과 다양하고 활용할 수 있는 결과 상당수 도출 ]



## 시사점 및 서비스 제안

### ② 분석결과를 바탕으로 한 시사점 - 상품 연관분석

{샌드}	=>	{일반스낵}
{샌드}	=>	{봉지라면}
{크래커}	=>	{일반스낵}
{크래커}	=>	{봉지라면}
{당근}	=>	{양파}
{차별화계란}	=>	{일반흰우유}
{차별화계란}	=>	{봉지라면}
{랩/호일}	=>	{다이소}
{컵라면}	=>	{일반스낵}
{컵라면}	=>	{봉지라면}
{소스}	=>	{봉지라면}

[ 상품 연관분석 결과 다양하고 활용할 수 있는 결과 상당수 도출 ]



## 연관분석 시각화

## 상품 연관분석 결과

많은 상품들이 서로서로

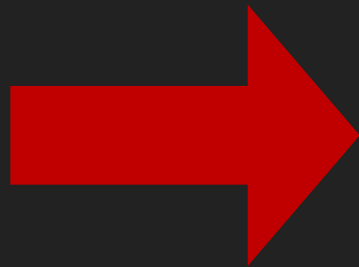
## 연관 되어 있음을 확인



## 시사점 및 서비스 제안

### ③ L.point 서비스제안1 – 구매주기마다 할인쿠폰 제공

한 고객당 동일한 상품을 다섯번이상 사면



구매주기마다 할인쿠폰 제공



\*새롭게 구매시 정보 업데이트



## 시사점 및 서비스 제안

### ③ L.point 서비스제안1 – 구매주기마다 할인쿠폰 제공

구매주기 : 한 고객이 동일한 하나의 상품을 중복 구매한 경우  
다음 구매까지 걸리는 평균적인 기간





## 시사점 및 서비스 제안

### ③ L.point 서비스제안1 – 쿠폰제공 시 다른 상품도 같이 쿠폰 제공



【고객들이 동시에 많이 구매한 상품의 쿠폰을 보내줄 시  
상품을 같이 구매할 확률 상승】



## 시사점 및 서비스 제안

### ③ L.point 서비스제안1 – 쿠폰제공 시 다른 상품도 같이 쿠폰 제공

연관분석 결과를 기반으로 나온 연관된 상품은 고객들이 동시에 구매할 확률이 높으므로  
**연관이 되어있는 상품도 같이 쿠폰 제공**







## 시사점 및 서비스 제안

### ③ L.point 서비스제안2 – 총 구매수량에 따라 할인율을 다르게 적용

〈예시〉

10번 이상 중복구매 – 5%  
20번 이상 중복구매 – 7%  
30번 이상 중복구매 – 10%

.  
. .  
.

[ 고객들이 반복주문 할 수록  
다시 구매할 수 있는 요인이 커지도록 쿠폰 할인율 적용 ]



## 시사점 및 서비스 제안

### ③ L.point 서비스제안3 – 주기적인 PUSH 알람 서비스

〈PUSH 알람 예시1〉

‘○○님, □상품을 구매할 때가 되지 않으셨나요?  
이번주 내에 구매하시면 5% 할인!’

〈PUSH 알람 예시2〉

○○님, 현재까지 □상품 8번 구매하셨어요  
2번만 더 구매하시면 5% 할인!

【 고객들이 주기적으로 구매하는 **상품들을 상기**시키고  
**쿠폰으로 유인**할 수 있도록 PUSH알람 실행 】



## 시사점 및 서비스 제안

### ③ L.point 서비스제안3 – 구매자가 오랜 주기동안 미구매시 쿠폰 지급

〈PUSH알람 예시〉

‘○○님, □ 상품을 구매하신지 오래되지 않으셨나요?  
이번주 내에 구매하시면 5% 할인!’

【 각각의 고객의 구매주기가 3배 이상 지날 시 쿠폰을 지급 하면  
오랜기간 구매하지 않은 상품으로 고객 유인 가능 】



THANK YOU FOR YOUR ATTENTION