

코로나 19 로 인한 소비변화에 대한 분석과 마케팅 전략

2019150445 통계학과 신백록

2020/12/05

1. Overview

i) 탐구동기와 목표

몇 달 전에 삼성카드에서 카드 데이터를 이용해 마케팅 전략을 수립하는 공모전을 열었다. 그 당시에 관심이 있어서 공모전에 참여하려고 하였지만, 코딩 능력, 분석 능력 등 여러 방면에서 많이 부족하다는 것을 느끼고 그만 뒀었다. 이번에 데이터 분석 프로젝트 주제를 선정할 때 가장 먼저 이 데이터가 떠올랐고, 내가 전에 비해서 발전이 있었는지 알아보고 싶어서 카드 데이터셋을 이용해 주제를 선정하였다. 또한 현재 상황에 가장 알맞은 데이터셋이라 생각했고, 현실에서도 충분히 적용할 수 있을 것이라고 생각해서 탐구 주제를 이와 같이 선정하였다.

카드 데이터 셋을 이용하여 코로나 시대에 소비 변화가 있었는지, 있었다면 어느 분야가 성장하였고, 어느 분야가 떨어졌는지를 알아볼 것이다. 또한 어느 성별대, 어느 연령대에서 그러한 변화가 나타났는지 분석해볼 것이고, 마지막엔 이 분석 결과를 이용해서

ii) Overview

데이터 분석을 위해 삼성카드에서 제공한 카드이용 데이터셋을 활용하였다. 데이터셋에 대한 설명과 함께 glimpse 를 읽어보면, 각 observation 은 각각의 사람에 대한 나이, 연령대, 기혼 여부, 유아자녀 여부 등의 정보와 어느 업종에서 소비를 했는지를 나타내고 있다. 주어진 카드 데이터의 소비 년/월은 2019 년 4 월, 2019 년 5 월, 2020 년 4 월, 2020 년 5 월이다. 이는 실제 카드 이용 내역 데이터로 민감한 부분이 많기에 성별은 남/녀가 아닌 0/1 로, 연령대는 A/B/C/D/E/F/G/H 로 구분해 성별과 연령대를 특징지을 수 없게 주어졌다. 또한 카드 소비 내역에 대한 정보로 어떤 자녀가 있는지, 전업 주부인지, 결혼은 했는지 등을 카드사에서 추측해 스코어로 주어졌다.

```

#Load dataset
x<-read.csv('C:/trend_w_demo.csv')
x %>% glimpse()

## Rows: 452,038
## Columns: 10
## $ YM                <int> 202005, 202005, 202005, 202005, 202005, 202005, ...
## $ Category          <chr> "할인점", "취미", "오픈마켓/소셜", "뷰티", "오픈마켓/소셜",
, "디저트", ...
## $ 성별구분          <int> 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0...
## $ 연령대            <chr> "F", "B", "D", "D", "G", "E", "B", "D", "F", "D", "...
## $ 기혼스코어        <chr> "high", "high", "mid", "mid", "high", "mid", "high", ...
## $ 유아자녀스코어    <chr> "low", "low", "mid", "mid", "low", "low", "low", "high".
..
## $ 초등학생자녀스코어 <chr> "high", "mid", "mid", "mid", "mid", "low", "high", "low"
,...
## $ 중고생자녀스코어  <chr> "mid", "mid", "mid", "mid", "mid", "low", "mid", "mid",
...
## $ 대학생자녀스코어  <chr> "low", "low", "low", "low", "mid", "low", "low", "low",
...
## $ 전업주부스코어    <chr> "low", "low", "mid", "low", "low", "low", "low",
"mid",...

```

다음은 각각의 카테고리에 대한 설명이다.

디저트: 커피 카페, 제과점

호텔/숙박: 호텔, 펜션 등

항공/여행상품: 항공사, 여행사, 렌터카 등

취미: 스튜디오, 문구점, 전시회, 낚시 등

할인점: 오프라인 마트, 오프라인 아울렛 등

뷰티: 화장품 샵, 피부관리, 헤어 등

면세점: 면세점

종합몰: 온라인 쇼핑몰(백화점, 마트 등)

오픈마켓/소셜: 온라인 쇼핑몰

전문물: 온라인 쇼핑몰(서적, 교육, 특화 스포츠 등)

iii) Overview-continued

#몇몇 변수들이 한글로 되어있기 때문에 분석의 편리함을 위해 영어로 rename 하였다.

```
x<-x %>% rename(Gender=성별구분, Age_Group=연령대,
                Marriage=기혼스코어, Infant_Child=유아자녀스코어,
                Elementary_Child=초등학생자녀스코어,
                MiddleHigh_Child=중고생자녀스코어,
                University_Child=대학생자녀스코어,
                House_Wife=전업주부스코어)
```

```
x %>% group_by(YM) %>% summarize(n=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 4 x 2
##   YM      n
##   <int> <int>
## 1 201904 120119
## 2 201905 121451
## 3 202004 101374
## 4 202005 109094
```

#year, month 별 observation 수가 상이하기 때문에 비율로 분석할 것이다.

#각 연월 당 카테고리별 차지비율 YM_Category 로 저장하였다.

```
YM_Category<-x %>% group_by(YM ,Category) %>%
  summarize(n=n()) %>% mutate(prop=n/sum(n))
```

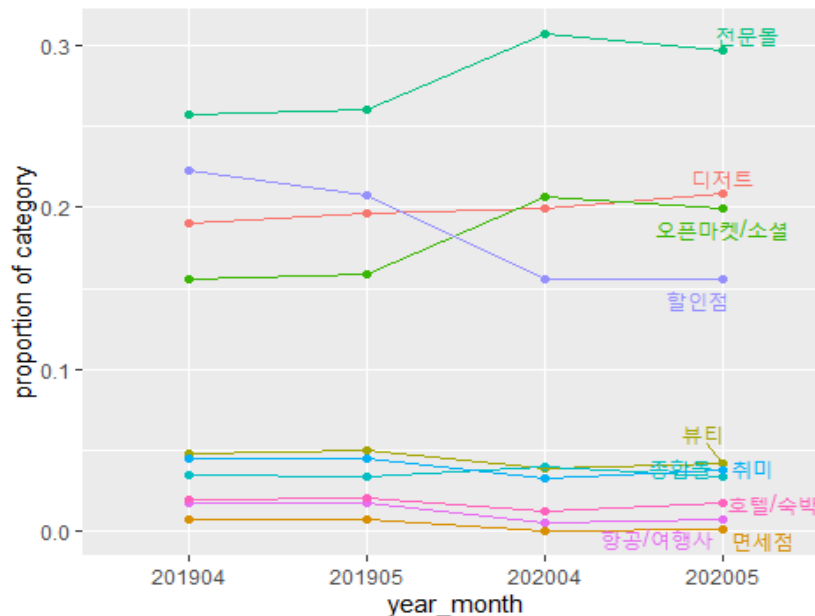
```
## `summarise()` regrouping output by 'YM' (override with `.groups` argument)
```

```
YM_Category
```

```
## # A tibble: 40 x 4
## # Groups:   YM [4]
##   YM Category      n    prop
##   <int> <chr>    <int>  <dbl>
## 1 201904 디저트   22918  0.191
## 2 201904 면세점    947  0.00788
## 3 201904 뷰티    5734  0.0477
## 4 201904 오픈마켓/소셜 18766  0.156
## 5 201904 전문물   30911  0.257
## 6 201904 종합물   4189  0.0349
```

```
## 7 201904 취미          5469 0.0455
## 8 201904 할인점        26721 0.222
## 9 201904 항공/여행사   2132 0.0177
## 10 201904 호텔/숙박    2332 0.0194
## # ... with 30 more rows
```

```
YM_Category %>% ungroup %>% ggplot(aes(x=as.character(YM),y=prop,col=Category,group=Category))+
  geom_point()+geom_line()+geom_text_repel(data=subset(YM_Category,YM=='202005'),aes(
    label=Category))+
  xlab('year_month')+ylab('proportion of category')+ theme(legend.position = 'none')
```

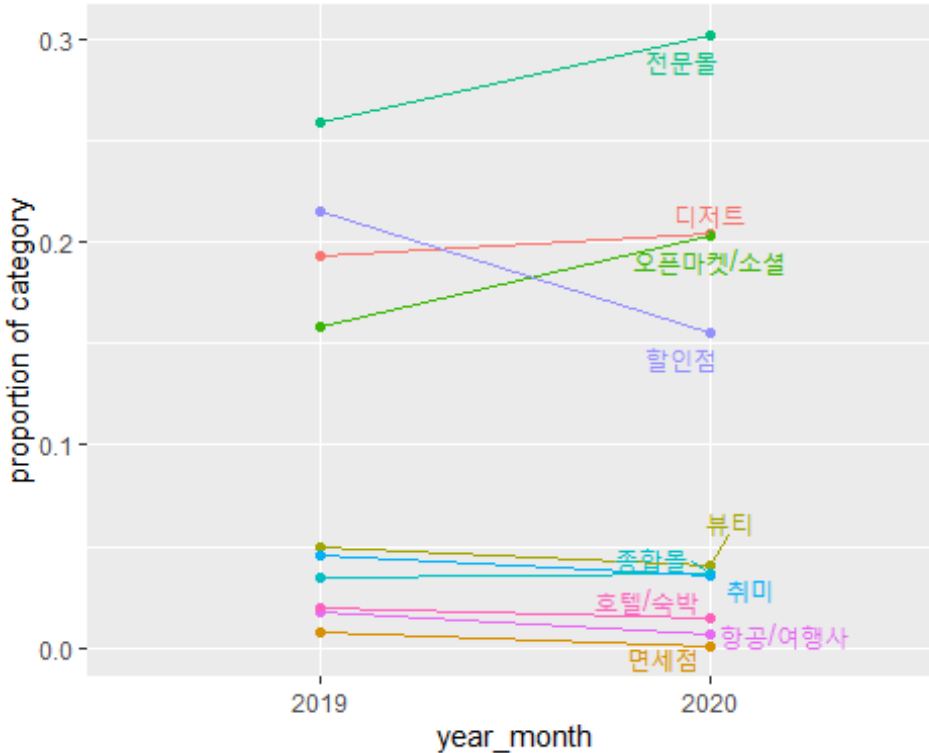


간단하게 그래프를 살펴보면, 오픈마켓/소셜 카테고리가 이번 년도 들어서 전체 소비에서 차지하는 비율이 대폭 상승하였고, 전문몰도 마찬가지로 상승하였다. 종합몰도 큰 폭은 아니지만 우상향 직선을 나타내고 있다.

반면에, 할인점, 취미, 항공/여행사, 호텔/숙박 등은 전체에서 차지하는 비율이 작년에 비해 감소하였음을 알 수 있다. 하지만 우리가 원하는 것은 코로나 19 가 소비에 영향을 미쳤는지를 알아보기 위한 그래프이기 때문에 4 월과 5 월의 데이터를 하나로 합쳐서 연별로 분석하는 것이 더 바람직해 보인다.

```
Y_Category<- x%>% mutate(year=ifelse(YM %in% c(201904,201905),2019,2020)) %>%
  group_by(year,Category) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))
## `summarise()` regrouping output by 'year' (override with `.groups` argument)
```

```
Y_Category %>% ggplot(aes(as.character(year),prop,col=Category,group=Category))+geom_
point()+geom_line()+
geom_text_repel(data=subset(Y_Category, year=='2020'),aes(label=Category))+
xlab('year_month')+ylab('proportion of category')+ theme(legend.position = 'none')
```



위에서 봤듯이, 오픈마켓/소셜, 전문물, 디저트 카테고리는 우상향 그래프를 그리고 있고, 반면에 할인점, 항공/여행사, 호텔/숙박, 면세점, 취미, 뷰티 카테고리는 우하향 그래프를 나타내고 있다. 위의 업종분류표를 보면 오픈마켓/소셜, 전문물은 모두 온라인 쇼핑몰인 것을 알 수 있고, 두 업종 모두 이번 년도 들어서 코로나 19 의 여파로 전체 소비에서 차지하는 비중이 대폭 상승하였다고 볼 수 있다.

반면에 호텔/숙박, 면세점, 취미, 할인점 등은 오프라인 매장에 거점을 두고 운영하고 있기에 여행을 기피하고 집 밖으로 잘 나가지 않는 현재의 상황에 비추어보면 우하향 그래프를 가지고 있는 것이 꽤나 자연스러운 결과로 보인다.

하나 의문스러운 결과는 디저트 카테고리이다. 이 역시 오프라인 매장을 중심으로 돌아가는 업종인데, 2020 년 들어서 오히려 소비가 증가한 것을 볼 수 있다. 앞의 월별 그래프를 봐도 디저트 카테고리는 계속 조금씩 증가하는 추세를 가지고 있다. 4 월 5 월에도 사회적 거리두기 정책이 시행되었지만 그 적용 범위가 넓지 않았고, 현재 정책과는 달리 카페, 제과점 등의 영업에는 아무런 제약이 걸리지 않았었다. 따라서 매출이 감소하는 경향은 크게 보이지 않았을 것이고, 이는 비율에 대한 그래프이기에 다른 업종에서의 소비가 큰 폭으로 감소함에 따라 카페에서의 소비가 소폭 상승한 것처럼 보인 것으로 생각할 수 있을 것이다. 아마 4, 5 월에 정부에서 (자발적) 사회적 거리두기 정책을 시행하였지만 카페에서의 확진자가 잇달아 나옴에도

불구하고 카페에서의 소비가 줄지 않은 것을 확인하고 지금 단계의 강제적 사회적 거리두기로 카페 영업 제한을 걸은 것으로 보인다.

하지만 정말 이것이 코로나 19 의 영향으로 증감 추세를 보이는지 아니면 우연에 의한 것인지 의문이 든다. 물론 다른 요인에 의한 것은 통제하지 못하지만, 2019 년과 2020 년 사이의 카테고리 별 소비에 유의미한 변화가 있었는지만은 오즈비로 구해볼 수 있다.

iv) Odds ratio

#카테고리별 odd ratio 계산

```
Category2019 <-Y_Category %>% ungroup %>% slice(1:10)
Category2020 <-Y_Category %>% ungroup %>% slice(11:20)

Category2019<-Category2019 %>% select(-year,-prop) %>% rename(n2019=n)
Category2020<-Category2020 %>% select(-year,-prop) %>% rename(n2020=n)

Category_or<-Category2019 %>% left_join(Category2020)

## Joining, by = "Category"

Category_or<-Category_or %>% mutate(odds_ratio=(n2019/(sum(n2019)-n2019))/
                                   (n2020/(sum(n2020)-n2020))) %>%
  arrange(-odds_ratio)
Category_or

## # A tibble: 10 x 4
##   Category      n2019 n2020 odds_ratio
##   <chr>      <int> <int>     <dbl>
## 1 면세점        1873   195      8.43
## 2 항공/여행사   4297  1464      2.59
## 3 할인점       51937 32699      1.49
## 4 호텔/숙박     4846  3159      1.34
## 5 취미         10976  7473      1.29
## 6 뷰티         11864  8589      1.21
## 7 종합물        8364  7755      0.938
## 8 디저트       46738 43005      0.934
## 9 전문물       62566 63496      0.809
## 10 오픈마켓/소셜 38109 42633      0.737

#These differences may occur just due to chance. So for each Category, we can compute confidence interval.
log_or<-Category_or %>% mutate(log_or=log(odds_ratio)) %>%
  mutate(se=sqrt(1/n2019+1/(sum(n2019)-n2019)+
```

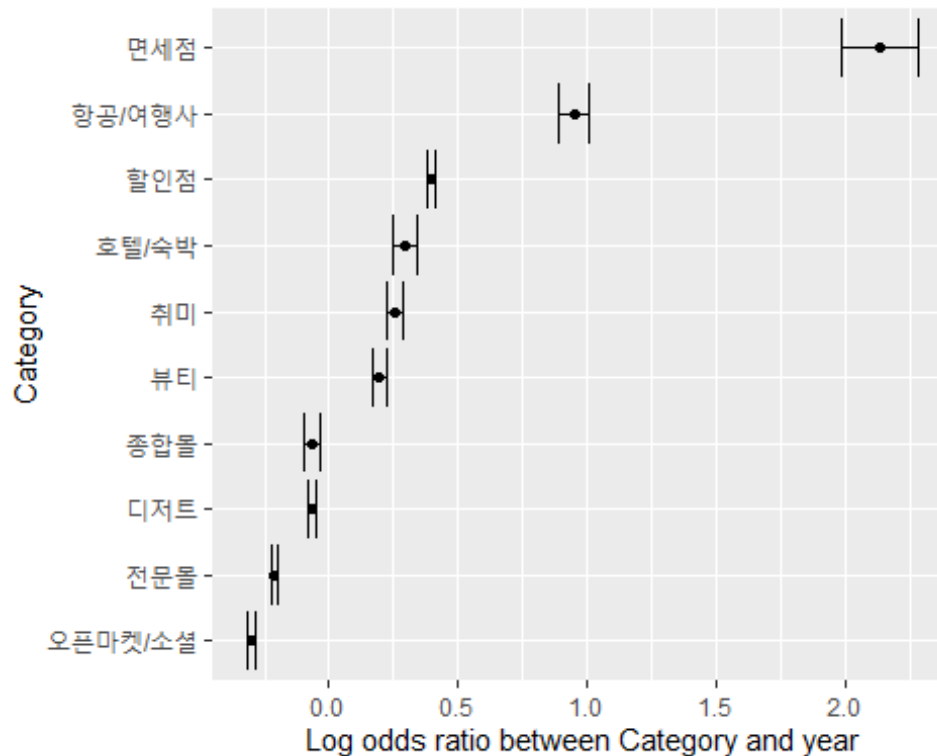
```

1/n2020+1/(sum(n2020)-n2020))) %>%
mutate(conf.low=log_or-qnorm(0.975)*se,
        conf.high=log_or+qnorm(0.975)*se)
log_or

## # A tibble: 10 x 8
##   Category      n2019 n2020 odds_ratio log_or      se conf.low conf.high
##   <chr>      <int> <int>      <dbl>  <dbl>    <dbl>    <dbl>    <dbl>
## 1 면세점        1873   195      8.43   2.13    0.0753     1.98     2.28
## 2 항공/여행사   4297  1464      2.59   0.950   0.0304     0.890    1.01
## 3 할인점       51937 32699      1.49   0.398   0.00779    0.383    0.413
## 4 호텔/숙박     4846  3159      1.34   0.295   0.0231     0.250    0.340
## 5 취미          10976 7473      1.29   0.257   0.0153     0.227    0.287
## 6 뷰티          11864 8589      1.21   0.194   0.0145     0.165    0.222
## 7 종합물        8364  7755      0.938 -0.0645  0.0161    -0.0960   -0.0331
## 8 디저트       46738 43005      0.934 -0.0681  0.00747   -0.0828   -0.0535
## 9 전문물       62566 63496      0.809 -0.212   0.00664   -0.225   -0.199
## 10 오픈마켓/소셜 38109 42633      0.737 -0.305   0.00778   -0.320   -0.289

log_or %>%
mutate(Category=reorder(Category,log_or)) %>%
ggplot(aes(x=Category, ymin=conf.low,ymax=conf.high))+
geom_errorbar()+
geom_point(aes(Category,log_or)) + coord_flip()+
ylab('Log odds ratio between Category and year')

```

#All Categories are significant since error bars does not capture 0.

위 그래프를 살펴보면, 모든 카테고리의 error bar 가 0 을 포함하고 있지 않기에 2019 년과 2020 년에 유의미한 소비 차이가 있었다고 볼 수 있다. 특히 면세점과 항공/여행사, 할인점, 호텔/숙박은 2019 년에 비해 2020 년에 유의미하게 감소하였고, 오픈마켓/소셜은 2019 년에 비해 2020 년에 유의미하게 증가한 것을 알 수 있다. 면세점, 항공/여행사, 호텔/숙박이 가장 큰 비율로 감소한 것으로 보아 관광업계에 큰 타격이 있었던 것으로 보이고, 오프라인 중심의 시장보다 온라인 중심의 시장이 커지고 있는 것을 확인할 수 있다. 물론 코로나 19 가 없었어도 오프라인 시장에서 온라인 시장으로의 전환은 이미 일어나고 있었지만, 코로나 19 가 이 변화를 더욱 촉진시킨 것 같다.

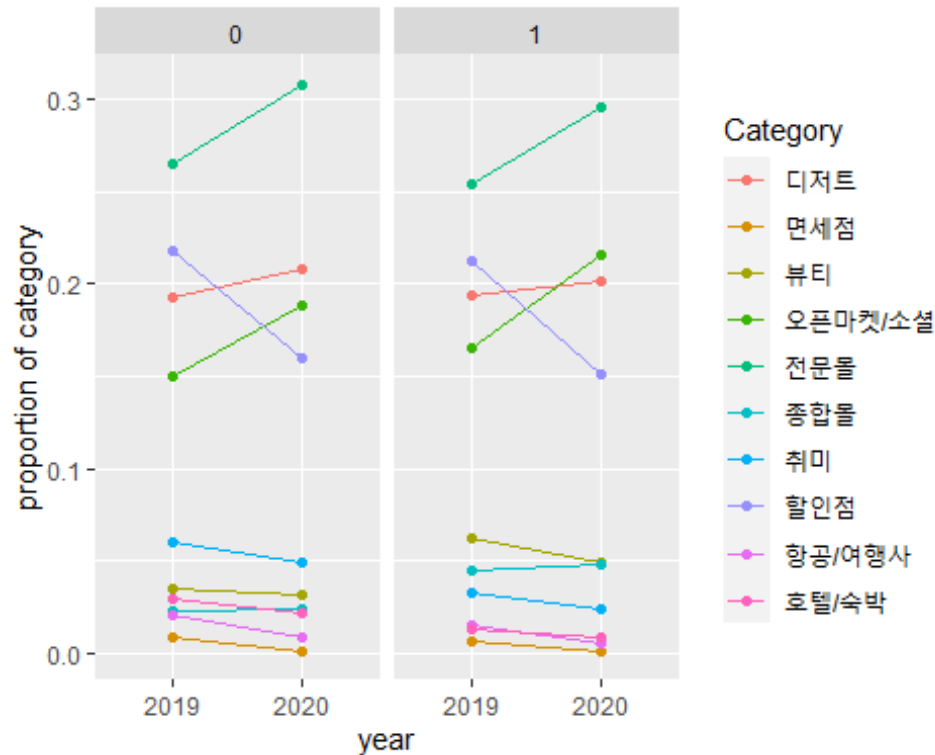
2. Gender 별 소비 변화 분석

1) Overview

```
Y_Category_Gender<-x %>% mutate(year=ifelse(YM %in% c(201904,201905),2019,2020)) %>%
  group_by(Gender,year, Category) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))
```

```
## `summarise()` regrouping output by 'Gender', 'year' (override with `.groups` argument)

Y_Category_Gender %>% ungroup %>% ggplot(aes(factor(year),prop,col=Category, group=Category)) +
  geom_line()+xlab('year')+ylab('proportion of category')+
  facet_grid(.~Gender)+geom_point()
```



아쉽게도 이 데이터셋의 성별은 남/녀가 아닌 0 과 1 로 나와있다. Gender 별 그래프를 살펴보면 대부분의 카테고리가 비슷한 형태를 띄고 있다. 눈에 띄는 것은 성별 1 의 뷰티와 종합물 카테고리가 성별 0 의 그것보다 많이 높게 있다는 것이다. 대신 성별 0 은 취미 카테고리가 더 높게 나타난다. 이를 가지고 조심스럽게 추측해보건데, 아마 성별 0 의 남자이고, 성별 1 이 여자이지 않을까 싶다. 이 외에는 이 그래프에서 딱히 알아낼 수 있는 것은 없어 보인다. 성별 0 과 성별 1 은 대부분의 카테고리에서 증감을 같이하는 것처럼 보이기 때문이다. 이 변화도 연도별로 유의미한 것인지 알아보기 위해 오즈비를 계산해보자.

2) Odds ratio

```
#odds ratio
x_gen0<-x %>% filter(Gender==0)
x_gen1<-x %>% filter(Gender==1)
```

```

x_gen0<-x_gen0 %>% mutate(year=ifelse(YM %in% c(201904,201905),2019,2020)) %>%
  group_by(year,Category) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))

## `summarise()` regrouping output by 'year' (override with `.groups` argument)

x_gen1<-x_gen1 %>% mutate(year=ifelse(YM %in% c(201904,201905),2019,2020)) %>%
  group_by(year,Category) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))

## `summarise()` regrouping output by 'year' (override with `.groups` argument)

#gen0
x_gen0_2019<-x_gen0 %>% ungroup %>% slice(1:10)
x_gen0_2020<-x_gen0 %>% ungroup %>% slice(11:20)

x_gen0_2019 <-x_gen0_2019 %>% select(-year,-prop) %>% rename(n2019=n)
x_gen0_2020 <-x_gen0_2020 %>% select(-year,-prop) %>% rename(n2020=n)
x_gen0_or<- x_gen0_2019 %>% left_join(x_gen0_2020)

## Joining, by = "Category"

gen0_log_or<-x_gen0_or %>% mutate(log_or=log((n2019/(sum(n2019)-n2019))/
                                           (n2020/(sum(n2020)-n2020)))) %>%
  mutate(se=sqrt(1/n2019+1/(sum(n2019)-n2019)+
                1/n2020+1/(sum(n2020)-n2020))) %>%
  mutate(conf.low=log_or-qnorm(0.975)*se,
         conf.high=log_or+qnorm(0.975)*se) %>%
  select(Category,log_or,conf.low,conf.high) %>%
  arrange(-log_or)
gen0_log_or

## # A tibble: 10 x 4
##   Category      log_or conf.low conf.high
##   <chr>         <dbl>   <dbl>   <dbl>
## 1 면세점      2.33      2.11     2.55
## 2 항공/여행사 0.847     0.769    0.926
## 3 할인점      0.381     0.359    0.403
## 4 호텔/숙박   0.272     0.217    0.326
## 5 취미        0.211     0.173    0.249
## 6 뷰티        0.110     0.0617   0.158
## 7 종합몰     -0.0260   -0.0824   0.0304
## 8 디저트     -0.0973   -0.119    -0.0760
## 9 전문몰     -0.214    -0.233    -0.195
## 10 오픈마켓/소셜 -0.276    -0.299    -0.254

#gen1
x_gen1_2019<-x_gen1 %>% ungroup %>% slice(1:10)
x_gen1_2020<-x_gen1 %>% ungroup %>% slice(11:20)

x_gen1_2019 <-x_gen1_2019 %>% select(-year,-prop) %>% rename(n2019=n)

```

```

x_gen1_2020 <-x_gen1_2020 %>% select(-year,-prop) %>% rename(n2020=n)

x_gen1_or<- x_gen1_2019 %>% left_join(x_gen1_2020)

## Joining, by = "Category"

gen1_log_or<-x_gen1_or %>% mutate(log_or=log((n2019/(sum(n2019)-n2019))/
                                             (n2020/(sum(n2020)-n2020)))) %>%
  mutate(se=sqrt(1/n2019+1/(sum(n2019)-n2019)+
                1/n2020+1/(sum(n2020)-n2020))) %>%
  mutate(conf.low=log_or-qnorm(0.975)*se,
         conf.high=log_or+qnorm(0.975)*se) %>%
  select(Category,log_or,conf.low,conf.high) %>%
  arrange(-log_or)
gen1_log_or

## # A tibble: 10 x 4
##   Category      log_or conf.low conf.high
##   <chr>         <dbl>   <dbl>   <dbl>
## 1 면세점         1.94     1.74     2.14
## 2 항공/여행사    1.09     0.994    1.18
## 3 할인점         0.414     0.393    0.435
## 4 호텔/숙박      0.357     0.277    0.437
## 5 취미           0.343     0.294    0.393
## 6 뷰티           0.237     0.202    0.272
## 7 디저트        -0.0420   -0.0622   -0.0219
## 8 종합물        -0.0855   -0.123    -0.0474
## 9 전문물        -0.210    -0.228    -0.192
## 10 오픈마켓/소셜 -0.329    -0.349    -0.308

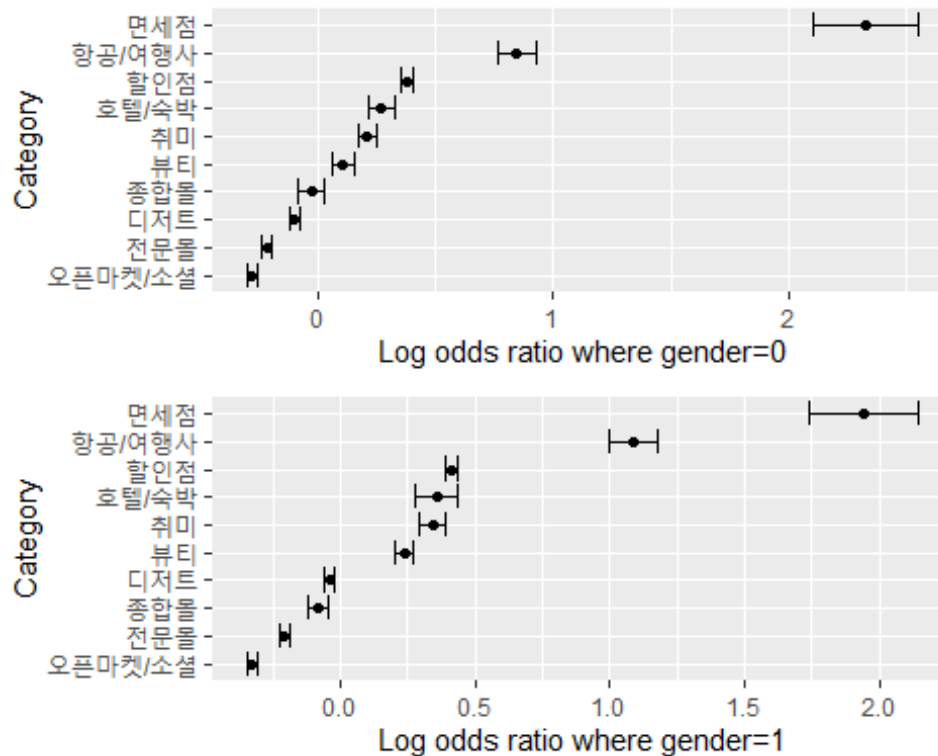
#plot

plot0<-gen0_log_or %>%
  mutate(Category=reorder(Category,log_or)) %>%
  ggplot(aes(x=Category, ymin=conf.low,ymax=conf.high))+
  geom_errorbar()+
  geom_point(aes(Category,log_or)) + coord_flip()+
  ylab('Log odds ratio where gender=0')

plot1<-gen1_log_or %>%
  mutate(Category=reorder(Category,log_or)) %>%
  ggplot(aes(x=Category, ymin=conf.low,ymax=conf.high))+
  geom_errorbar()+
  geom_point(aes(Category,log_or)) + coord_flip()+
  ylab('Log odds ratio where gender=1')

grid.arrange(plot0,plot1,nrow=2)

```



Gender 별 오즈비를 살펴보면, 위의 Gender 별 대략적인 그래프의 경향이 비슷하다. 성별 1 이 성별 0 에 비해 그래프가 더 퍼져있다는 것을 제외하면 말이다. 즉 성별 1 이 성별 0 에 비해 2020 년의 소비와 2019 년의 소비가 유의미하게 다르다는 것이다. 성별 1 이 전체적으로 트렌드, 유행에 영향을 더 많이 받는다고 볼 수 있을 것이다.

종합몰 카테고리를 보면, 성별 0 의 종합몰 카테고리는 0 이 error bar 에 들어온다. 다시 말해, 성별 0 은 2019 년과 2020 년에 종합몰 카테고리에서 소비 차이가 크게 나지 않았다는 것이다. 실제 성별 0 의 종합몰 카테고리의 log odds ratio 값은 -0.0855 로 0 에 매우 가까운 음의 값을 갖는다. 반면에 성별 1 의 종합몰 카테고리의 오즈비가 유의미하다고 볼 수 있고, 2019 년에 비해 2020 년에 유의미하게 종합몰 카테고리의 소비 비율이 증가했다는 것을 알 수 있다. 요즘 기사를 보면 코로나 19 로 인해 백화점을 포함한 유통업체들은 오프라인 매장을 정리하고 온라인 매장으로 이를 대신하는 분위기이다. 하지만 이 그래프를 본다면 과연 온라인 매장으로 종합몰을 만드는 것이 맞는 선택인지 다시 생각해볼게 될 것이다. 주 타겟이 성별 1 인 브랜드에겐 나쁘지 않은 선택이겠지만 성별 0 이 주 타겟인 브랜드에겐 좀처럼 괜찮은 선택이 되기 힘들 것처럼 보이기 때문이다.

만약 오프라인 중심의 브랜드들이 오픈마켓 혹은 소셜커머스에 발을 들여놓는다면 좋은 선택이 될 것이다. 어느 성별이든 간에 오픈마켓/소셜커머스 카테고리는 2020 년에 유의미하게 성장했기 때문이다. 성별 0 이든 성별 1 이든 간에 상관없지만 특히 성별 1 이 주 타겟인 브랜드들이면 말이다.

면세점은 어느 성별이든 간에 소비가 심각하게 많이 줄어든 것을 볼 수 있다. 이에 면세점은 코로나 19로 인한 타격이 상당히 클 것으로 예상된다. 그러면 면세점은 마케팅 전략을 어떻게 수립해야 살아남을 수 있을까.

3. 면세점의 마케팅 전략과 결론

```
#Duty free
Duty_free<-x %>% mutate(year= ifelse(YM %in% c(201904,201905),2019,2020)) %>% group_by(Gender, Age_Group,year) %>% filter(Category=='면세점') %>% summarize(n=n()) %>% spread(year, n) %>% rename(n2019='2019',n2020='2020') %>% mutate(prop_20202019=n2020/n2019) %>% arrange(-prop_20202019)

## `summarise()` regrouping output by 'Gender', 'Age_Group' (override with `.groups` argument)

Duty_free[is.na(tax_free)]<-0
Duty_free

## # A tibble: 16 x 5
## # Groups:   Gender, Age_Group [16]
##   Gender Age_Group n2019 n2020 prop_20202019
##   <int> <chr>      <int> <int>      <dbl>
## 1      0 H          67    15      0.224
## 2      1 E         159    34      0.214
## 3      1 H          94    16      0.170
## 4      0 E          83    12      0.145
## 5      1 F         103    14      0.136
## 6      0 C         149    20      0.134
## 7      1 G         115    15      0.130
## 8      0 B          91    11      0.121
## 9      1 D          93    11      0.118
## 10     1 C         121    11      0.0909
## 11     1 A          83     7      0.0843
## 12     0 D         195    12      0.0615
## 13     0 A         154     8      0.0519
## 14     0 F         160     6      0.0375
## 15     0 G         109     3      0.0275
## 16     1 B          97     0         0
```

데이터를 보면 상위 4 개 비율의 연령대가 E 와 H 이다. 비율을 보면 면세점은 어느 성별, 어느 연령대에서나 2020 년에 전년에 비해 소비가 감소하였지만, 피해를 최소화하기 위해서는 그나마 소비가 덜

줄어든 연령대 H, 혹은 연령대 E 를 타겟으로 하는 상품들을 주력상품으로 내세워야 할 것이다. 특히 성별 1 & 연령대 B 는 2020 년에 들어서 면세점에서의 소비가 아예 없고, 데이터의 아래에 위치한 성별과 연령대의 제품들과 그에 맞는 마케팅들은 잠시 넣어두는 것도 나쁘지 않아 보인다.

이런 식으로 면세점 외에 다른 카테고리들에 대해서도 여러 인사이트를 가지고 많은 분석들을 할 수 있을 것이고, 데이터를 기반으로 한 마케팅 전략을 수립할 수 있을 것이다.