**<포스트 코로나 데이터 시각화 경진대회>**

1829002 강다연

1829016 김지원

1829017 김하영

1. **참여한 공모전**

: 포스트 코로나 데이터 시각화 경진대회(KT | 유동인구 | 소비 | 포스트 코로나)



1. 세부내용

(1) 주제

* 포스트 코로나 시대 유망 품목 및 산업 발굴
* 코로나로 인한 국민의 생활/소비의 변화 파악

1. 목표

* 포스트 코로나 시대 도움되는 인사이트 발굴

1. 일정

1. 코드 및 PPT 제출 기간: 7.1 10:00 ~ 7.31 17:59

2. \*코드: 코드 공유 게시판에 게시

3. \*PPT: dacon@dacon.io로 제출

4. 팀 병합 마감: 7.20 23:59

5. 평가 기간: 8.1 00:00 ~ 8.6 23:59

6. 최종 순위 발표: 8.7 10:00

7. 시상식: 8.12 19:00

1. **주제 방향성**

: 코로나가 발생하기 전과 코로나가 발생한 이후 사람들의 소비 유형을 살펴보고 포스트 코로나 시대의 모습을 예측

1. **공모전 출품 내용**

1) 데이터 및 변수 설명

▶ fpopl.csv (1.6GB) : 행정동별 유동인구 데이터

|  |
| --- |
| - base\_ymd : 기준년월일  - tmzon\_se\_sode : 24시간대 구분코드  - sexdstn\_se\_code : 성별 구분코드 (M: 남성, F: 여성)  - agrde\_se\_code : 5세단위 연령대구분코드 (단, age\_00: 0세 ~ 9세, age\_70: 70세 이상)  - adstrd\_code : 행정동코드  - popltn\_cascnt : 인구수 |

▶ adstrd\_master.csv (256KB) : 8자리 행정동 코드 데이터

|  |
| --- |
| -adstd\_code : 행정동 코드  -adstrd\_nm : 행정동명  -brtc\_nm : 시도명  -signgu\_nm : 시군구명 |

▶ card.csv (0.2GB) : 업종 별 결제금액 데이터

|  |
| --- |
| -receipt\_dttm : 카드회사가 카드사용내역을 접수한일자  -adstrd\_code : 가맹점 위치 기준 행정동 코드  -adstrd\_nm : 가맹점 위치 기준 행정동명  -mrhst\_induty\_cl\_code : 가맹점 업종코드  -mrhst\_induty\_cl\_nm : 가맹점 업종명  -selng\_cascnt : 매출발생건수  -salamt : 매출발생금액 |

▶ delivery.csv (1.5GB) : 배달 호출 정보 데이터

|  |
| --- |
| - SERIAL\_NUMBER : 순번  - PROCESS\_DT : 처리일시  - DLVR\_RQESTER\_ID : 배달요청업체 ID  - DLVR\_REQUST\_STTUS\_VALUE : 배달요청상태값(1:완료, 2:취소, 3:사고, 4:문의)  - DLVR\_RCEPT\_CMPNY\_ID : 배달접수회사 ID  - DLVR\_STORE\_ID : 배달상점ID  - DLVR\_STORE\_INDUTY\_NM : 배달상점 업종이름  - DLVR\_STORE\_LEGALDONG\_CODE : 배달상점 주소 법정코드  - DLVR\_STORE\_SIDO : 배달상점 주소 법정동 시도명  - DLVR\_STORE\_SIGUNGU : 배달상점 주소 법정동 시군구명  - DLVR\_STORE\_DONG : 배달상점 주소 법정동 읍면동명  - DLVR\_STORE\_RI : 배달상점 주소 법정동 리명  - DLVR\_STORE\_ADSTRD\_CODE : 배달상점 주소 행정동 코드  - DLVR\_STORE\_RDNMADR\_CODE : 배달상점주소 도로명주소 코드  - DLVR\_STORE\_LEGALDONG\_CODE : 배달목적지 주소 법정동코드  - DLVR\_DSTN\_SIDO : 배달목적지 주소 법정동 시도명  - DLVR\_DSTN\_SIGUNGU : 배달목적지 주소 법정동 시군구명  - DLVR\_DSTN\_DONG : 배달목적지 주소 법정동 읍면동명  - DLVR\_DSTN\_RI : 배달목적지 주소 법정동 리명  - DLVR\_DSTN\_ADSTRD\_CODE : 배달목적지 주소 행정동 코드  - DLVR\_DSTN\_RDNMADR\_CODE : 배달목적지주소 도로명주소 코드  - DLVR\_MAN\_ID : 배달기사 ID  - DLVR\_AMOUNT : 배달비용  - CALL\_RLAY\_FEE\_AMOUNT : 호출중계수수료금액  - GOODS\_AMOUNT : 배달상품금액  - SETLE\_KND\_VALUE : 결제종류번호 (1:카드, 2:선불, 3:현금)  - SETLE\_CARD\_CN : 결제카드종류  - DLVR\_RCEPT\_TIME : 배달접수시간  - DLVR\_CARALC\_TIME : 배달배차시간  - DLVR\_COMPT\_TIME : 배달완료시간  - DLVR\_CANCL\_TIME : 배달취소시간 |

▶ index.csv (8MB) : 품목 별 소비지수 데이터

|  |
| --- |
| - period : 기준 월  -catl : 대분류  -catm : 중분류중분류  -age : 나이대  -gender : 성별  -sido : 지역  -sigungu : 세부지역  -cgi : 카테고리성장지수 (2018년 월평균 대비 매출 성장 비율, 100을 기준으로 이상이면 매출 상승, 이하면 하락) |

(데이터 출처: KT빅데이터플랫폼)

1. 가설

①코로나로 인해 온라인 결제가 증가했을 것이다.

②코로나로 인해 2019년도보다 2020년도의 소비량이 늘었다.

③20대 소비 변화가 다른 나이 대에 비해 매출에 가장 큰 영향을 미칠 것이다.

④ 코로나로 인해 건강/의료 용품의 소비가 증가했을 것이다.

⑤코로나로 인해 식품의 소비가 증가했을 것이다.

⑥코로나로 인해 화장품의 소비가 감소했을 것이다.

⑦ 2020년 1월부터 배달 건수가 꾸준히 증가할 것이다.

⑧ 월별 지역별 코로나 확진자 증가 추세는 크게 다르지 않을 것이다

⑨ 배달 음식의 직접 결제 건수가 현저히 줄었을 것이다.

⑩ 배달음식 결제수단으로 선불 결제가 가장 많을 것이다.

⑪지역별 배달 건수 추세는 비슷할 것이다.

⑫ 코로나 초반의 소비량이 현재의 소비량보다 많을 것이다.

1. 요일 별 소비량 변화와 코로나는 큰 관련이 없을 것이다.
2. 데이터 전처리

|  |
| --- |
| index  ## # A tibble: 127,525 x 8 ## period catl catm age gender sido sigungu cgi ## <dbl> <chr> <chr> <dbl> <chr> <chr> <chr> <dbl> ## 1 201901 건강/의료용품 건강관리용품 20 F 서울 관악구 115.  ## 2 201901 건강/의료용품 건강관리용품 20 F 서울 광진구 120.  ## 3 201901 건강/의료용품 건강관리용품 20 F 서울 도봉구 157.  ## 4 201901 건강/의료용품 건강관리용품 20 F 서울 동작구 58.3 ## 5 201901 건강/의료용품 건강관리용품 20 F 서울 마포구 145.  ## 6 201901 건강/의료용품 건강관리용품 20 F 서울 성북구 78.9 ## 7 201901 건강/의료용품 건강관리용품 20 F 서울 용산구 14.1 ## 8 201901 건강/의료용품 건강관리용품 20 M 서울 all 102.  ## 9 201901 건강/의료용품 건강관리용품 20 M 서울 강동구 126.  ## 10 201901 건강/의료용품 건강관리용품 20 M 서울 강서구 97.5 ## # ... with 127,515 more rows  #결측값 존재 확인 sum(is.na(index$age))  ## [1] 21343  index%>%filter(!age %in% c(20,30,40,50,60))  ## # A tibble: 21,343 x 8 ## period catl catm age gender sido sigungu cgi ## <dbl> <chr> <chr> <dbl> <chr> <chr> <chr> <dbl> ## 1 201901 건강/의료용품 건강관리용품 NA F 서울 all 91.9 ## 2 201901 건강/의료용품 건강관리용품 NA F 서울 관악구 69.3 ## 3 201901 건강/의료용품 건강관리용품 NA F 서울 마포구 144.  ## 4 201901 건강/의료용품 건강관리용품 NA F 서울 영등포구 108.  ## 5 201901 건강/의료용품 건강관리용품 NA M 서울 송파구 75.0 ## 6 201901 건강/의료용품 건강관리용품 NA M 서울 은평구 74.4 ## 7 201901 건강/의료용품 건강관리용품 NA M 서울 종로구 72.8 ## 8 201901 건강/의료용품 건강관리용품 NA M 서울 중구 134.  ## 9 201901 건강/의료용품 건강관리용품 NA all 서울 all 93.1 ## 10 201901 건강/의료용품 건강관리용품 NA all 서울 강북구 138.  ## # ... with 21,333 more rows  #age값이 all인 것이 결측값이므로 바꿔주었음  index$age[is.na(index$age)] <- 'all' index0<-index%>%filter(!age %in% c(20,30,40,50,60))  #앞으로 그림을 그리기 위해 period를 year과 month로 분리 index0<-index%>%mutate(year=period%/%100, month=period%%100) index0  ## # A tibble: 127,525 x 10 ## period catl catm age gender sido sigungu cgi year month ## <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> ## 1 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 관악구 115. 2019 1 ## 2 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 광진구 120. 2019 1 ## 3 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 도봉구 157. 2019 1 ## 4 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 동작구 58.3 2019 1 ## 5 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 마포구 145. 2019 1 ## 6 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 성북구 78.9 2019 1 ## 7 201901 건강/의료용품~ 건강관리용품~ 20 F 서울 용산구 14.1 2019 1 ## 8 201901 건강/의료용품~ 건강관리용품~ 20 M 서울 all 102. 2019 1 ## 9 201901 건강/의료용품~ 건강관리용품~ 20 M 서울 강동구 126. 2019 1 ## 10 201901 건강/의료용품~ 건강관리용품~ 20 M 서울 강서구 97.5 2019 1 ## # ... with 127,515 more rows |

|  |
| --- |
| summary(card)  ## receipt\_dttm adstrd\_code adstrd\_nm  ## Min. :20200104 Min. :1.111e+09 Length:3777647  ## 1st Qu.:20200229 1st Qu.:1.126e+09 Class :character  ## Median :20200406 Median :1.147e+09 Mode :character  ## Mean :20200377 Mean :1.145e+09  ## 3rd Qu.:20200512 3rd Qu.:1.162e+09  ## Max. :20200614 Max. :1.174e+09  ## mrhst\_induty\_cl\_code mrhst\_induty\_cl\_nm selng\_cascnt  ## Min. :1001 Length:3777647 Length:3777647  ## 1st Qu.:4201 Class :character Class :character  ## Median :7020 Mode :character Mode :character  ## Mean :6077  ## 3rd Qu.:8021  ## Max. :9998  ## salamt  ## Length:3777647  ## Class :character  ## Mode :character  # 원하는 형태의 자료형으로 바꿔주기  card$salamt <- as.numeric(card$salamt,na.rm=TRUE)  card$receipt\_dttm <- as.character(card$receipt\_dttm)  distinct(card,mrhst\_induty\_cl\_nm)  # 날짜별 소비량 확인 - 그래프  day\_by <- group\_by(card,receipt\_dttm) day\_by\_sum <- summarise(day\_by, sum=sum(salamt,na.rm=TRUE))  ggplot(day\_by\_sum,aes(x=receipt\_dttm,y=sum,group=1))+geom\_line()    소비량의 피크가 비슷한 주기로 찍힘. |

|  |
| --- |
| delivery<-delivery%>%separate(PROCESS\_DT,into =  c("year","month","day"))  # 연-월-일 형태로 저장되어 있는 PROCESS\_DT를 ‘year’, ‘month’,’day’로 분리 시킴  delivery$month = as.double(delivery$month)  delivery$day = as.double(delivery$day)  delivery1 = delivery      # 배달 도착지별 배달건수를 count한 결과 배달도착지명 중 NA 값이 많다는 것을 확인함. 17개의 행정구역 중 세종특별자치시가 누락된 것을 확인하고 NA 값을 ‘세종특별자치시’로 바꿈.  delivery1$DLVR\_DSTN\_SIDO<-replace(delivery$DLVR\_DSTN\_SIDO,is.na(delivery$DLVR\_DSTN\_SIDO), "세종특별자치시") |

1. 가설 검정
2. 코로나로 인해 온라인 결제가 증가했을 것이다.

|  |
| --- |
| library(gghighlight) ggplot(kinds,aes(x=receipt\_dttm,y=sum,color=mrhst\_induty\_cl\_nm))+ geom\_line(aes(group=mrhst\_induty\_cl\_nm))+ gghighlight(max(sum)>100000000000,max\_highlight=4,use\_direct\_label = TRUE) +theme\_minimal()  ## label\_key: mrhst\_induty\_cl\_nm    # 일별 카드결제건수를 살펴본 결과 인터넷P/G와 인터넷 Mall, 조세서비스 항목에서 결제건수가 가장 높게 나왔다.  **#코로나로 인해 인터넷 몰에서 마스크를 구입하고 외부 활동 감소로 온라인을 신선식품을 사는 등과 같은 인터넷 Mall에서의 카드결제가 증가했다고 추측할 수 있다. 또한 인터넷 상에서 금융 기관과 거래하는 인터넷 P/G도 앞선 인터넷 몰과 유사한 이유로 결제건수가 증가한 것으로 추측할 수 있다. 따라서 가설을 기각하지 않는다.** |

② 코로나로 인해 2019년도보다 2020년도의 소비량이 늘었다.

|  |
| --- |
| index1<-index0%>%filter(age=='all',gender=='all', sigungu=='all')%>%group\_by(year,month)%>% summarise(mean.cgi=mean(cgi)) index1  ## # A tibble: 17 x 3 ## # Groups: year [2] ## year month mean.cgi ## <dbl> <dbl> <dbl> ## 1 2019 1 102.  ## 2 2019 2 87.7 ## 3 2019 3 97.0 ## 4 2019 4 94.3 ## 5 2019 5 103.  ## 6 2019 6 103.  ## 7 2019 7 99.2 ## 8 2019 8 104.  ## 9 2019 9 97.8 ## 10 2019 10 84.1 ## 11 2019 11 84.5 ## 12 2019 12 94.8 ## 13 2020 1 94.5 ## 14 2020 2 80.3 ## 15 2020 3 79.9 ## 16 2020 4 78.5 ## 17 2020 5 73.6  ggplot(index1, aes(month,mean.cgi,group=year))+geom\_line(aes(color=factor(year)))+xlim(1,5)+geom\_hline(yintercept = 100)+ggtitle("cgi지수")  ## Warning: Removed 7 rows containing missing values (geom\_path).    # 나이, 성별과 지역은 구분하지 않고 2019년과 2020년 1~5월 cgi지수의 추세 그림을 그려보았다.  **# 1~2월 까지의 cgi지수는 2019년과 2020년이 비슷하지만 그 이후로는 2020년의 cgi 값이 감소하는 추세를 보인다. Cgi는 매출에 대한 지수이므로, 2020년의 소비량은 감소했다고 할 수 있다.** |

③ 20대 소비 변화가 다른 나이 대에 비해 매출에 가장 큰 영향을 미칠 것이다.

|  |
| --- |
| index2<-index0%>%filter(gender=='all', sigungu=='all')%>%group\_by(year,month, age)%>% summarise(mean.cgi=mean(cgi)) index2  ## # A tibble: 102 x 4 ## # Groups: year, month [17] ## year month age mean.cgi ## <dbl> <dbl> <chr> <dbl> ## 1 2019 1 20 108.  ## 2 2019 1 30 100.  ## 3 2019 1 40 101.  ## 4 2019 1 50 103.  ## 5 2019 1 60 98.4 ## 6 2019 1 all 102.  ## 7 2019 2 20 99.8 ## 8 2019 2 30 88.2 ## 9 2019 2 40 86.6 ## 10 2019 2 50 86.8 ## # ... with 92 more rows  ggplot(index2, aes(month,mean.cgi,group=year))+geom\_line(aes(color=factor(year)))+xlim(1,5)+facet\_wrap(~age)+geom\_hline(yintercept = 100)+ggtitle("나이에 따른 cgi지수")  ## Warning: Removed 7 rows containing missing values (geom\_path).    # 성별과 지역은 구분하지 않고 오로지 나이에 대해 구분한 후 2019년과 2020년 1~5월 cgi지수의 추세 그림을 그려보았다.  **# 20대를 제외한 나머지 나이대의 cgi지수는 2019년과 2020년 모두 기준 값인 100보다 전반적으로 아래에 위치해있지만, 20대의 경우는 2019년엔 100보다 위에 위치해있었던 cgi지수가 2020년이 되자 100 아래로 떨어졌다는 점에서 20대의 소비 변화가 매출 저하에 큰 타격을 미쳤다고 할 수 있다. 따라서 가설을 지지한다.** |

④ 코로나로 인해 건강/의료 용품의 소비가 증가했을 것이다.

|  |
| --- |
| index5<- index0%>%group\_by(year,month,catl)%>%summarise(mean.cgi=mean(cgi)) index5  ## # A tibble: 68 x 4 ## # Groups: year, month [17] ## year month catl mean.cgi ## <dbl> <dbl> <chr> <dbl> ## 1 2019 1 건강/의료용품 98.9 ## 2 2019 1 식품 107.  ## 3 2019 1 일용품 97.9 ## 4 2019 1 화장품 104.  ## 5 2019 2 건강/의료용품 101.  ## 6 2019 2 식품 92.5 ## 7 2019 2 일용품 90.7 ## 8 2019 2 화장품 85.9 ## 9 2019 3 건강/의료용품 107.  ## 10 2019 3 식품 103.  ## # ... with 58 more rows  ggplot(index5, aes(month,mean.cgi))+geom\_line(stat="identity",aes(color=factor(catl)))+facet\_wrap(year~catl,ncol=4)+ scale\_x\_continuous(breaks=c(1,2,3,4,5),limits=c(0,6))+geom\_hline(yintercept = 100)+ggtitle("카테고리별 cgi지수")    # 대카테고리 별로 구분한 후 2019년과 2020년 1~5월 cgi지수의 추세 그림을 그려보았다.  **# 2020년 1~2월 건강/의료 용품의 cgi지수는 실제로 증가했으나 그 이후엔 100 이하로 감소하였다는 점에서 가설을 기각한다.** |

⑤ 코로나로 인해 식품의 소비가 증가했을 것이다.

|  |
| --- |
| #가설 ④에서 그린 그림을 참고하고, 추가적으로 식품의 하위 카테고리 별 cgi 지수 추세선을 그려보았다.  index6<- index0%>%filter(catl=='식품')%>%group\_by(year,month,catm)%>%summarise(mean.cgi=mean(cgi)) index6  ## # A tibble: 170 x 4 ## # Groups: year, month [17] ## year month catm mean.cgi ## <dbl> <dbl> <chr> <dbl> ## 1 2019 1 가공식품 114.  ## 2 2019 1 기호식품 121.  ## 3 2019 1 담배 139.  ## 4 2019 1 빙과류 66.7 ## 5 2019 1 빵류 116.  ## 6 2019 1 신선식품 115.  ## 7 2019 1 유제품류 98.5 ## 8 2019 1 음료 95.0 ## 9 2019 1 제과류 98.3 ## 10 2019 1 주류 103.  ## # ... with 160 more rows  ggplot(index6, aes(month,mean.cgi))+geom\_line(stat="identity",aes(color=factor(catm)))+facet\_wrap(year~catm,ncol=10)+ scale\_x\_continuous(breaks=c(1,2,3,4,5),limits=c(0,6))+geom\_hline(yintercept = 100)+ggtitle("식품 카테고리별 cgi지수")    **# 가설 4에서 그린 그림을 보면 식품 카테고리의 cgi지수가 감소하므로 식품의 소비가 증가한다는 가설을 기각한다. 식품의 하위 카테고리 cgi지수 그래프를 보면 전반적으로 2019년에비해 2020년에 감소하였고, 담배, 유제품류, 주류 등은 전년도와 비슷하다는 것을 알 수 있다.** |

⑥ 코로나로 인해 화장품의 소비가 감소했을 것이다.

|  |
| --- |
| #가설 ④에서 그린 그림을 참고하고, 추가적으로 화장품의 하위 카테고리 별 cgi 지수 추세선을 그려보았다.  index8<- index0%>%filter(catl=='화장품')%>%group\_by(year,month,catm)%>%summarise(mean.cgi=mean(cgi)) index8  ## # A tibble: 66 x 4 ## # Groups: year, month [17] ## year month catm mean.cgi ## <dbl> <dbl> <chr> <dbl> ## 1 2019 1 기타화장품 59.5 ## 2 2019 1 바디/헤어용품 132.  ## 3 2019 1 뷰티소품 81.3 ## 4 2019 1 화장품 103.  ## 5 2019 2 기타화장품 77.4 ## 6 2019 2 바디/헤어용품 79.8 ## 7 2019 2 뷰티소품 81.2 ## 8 2019 2 화장품 97.5 ## 9 2019 3 기타화장품 69.4 ## 10 2019 3 바디/헤어용품 86.5 ## # ... with 56 more rows  ggplot(index8, aes(month,mean.cgi))+geom\_line(stat="identity",aes(color=factor(catm)))+facet\_wrap(year~catm,ncol=4)+ scale\_x\_continuous(breaks=c(1,2,3,4,5),limits=c(0,6))+geom\_hline(yintercept = 100)+ggtitle("화장품 카테고리별 cgi지수")    # 2019년과 2020년 1~5월 cgi지수의 추세 그림을 그려보았다.  **# 가설 4에서 그린 그림을 보면 화장품 카테고리의 cgi지수가 감소하므로 화장품의 소비가 증가한다는 가설을 지지한다. 화장품의 하위 카테고리 cgi지수 그래프를 보면 전반적으로 2019년에비해 2020년에 감소하였고, 특히 화장품이 2월 이후 전년도 대비 급격하게 감소하였다는 것을 알 수 있다. 이는 코로나의 영향을 받았을 것으로 추측할 수 있다.** |

⑦ 2020년 1월부터 배달 건수가 꾸준히 증가할 것이다.

|  |
| --- |
| # 월 별 날짜 수가 다르고 6월은 21일까지의 데이터만 있었기 때문에 각 월별 배달건수(n)을 날짜 수로 나누어 r을 구하여 bar 차트를 그려보았다.  **# 1월부터 5월까지는 꾸준히 증가하는 추세를 보이지만 6월에는 배달건수가 현격히 감소한 것을 확인할 수 있다. 따라서 ‘1월부터 배달건수는 꾸준히 증가할 것이다’의 가설을 기각한다.** |

⑧ 월별 지역별 코로나 확진자 증가 추세는 크게 다르지 않을 것이다

|  |
| --- |
| # 대구와 서울, 경기도처럼 확진자 수가 1000명이 넘어가는 지역 때문에 100명 이하의 확진자가 나오는 울산, 광주, 세종 등의 확진차 증가 추이를 확인하기 어려웠다. 그래서 facet\_wrap에 scale= ‘free’ 옵션을 사용하여 월별 지역별 확진자 추이를 확인해보았다.  **# 월별 지역별로 확진자가 증가하고는 있지만 서로 다른 양상을 띄며 증가하고 있다는 것을 확인할 수 있다. 예로 ‘서울’의 경우 3월 11일 구로구 콜센터에서 일어난 집단 감염으로 수도권의 확진자가 증가했으며 4월~5월 확진자가 10명 이내로 떨어지면서 증가추세가 주춤했다가 5월 2일 이태원 확진자가 발생하면서 확진자 수가 다시 증가하는 양상을 띈다. 이에 반해 ‘인천’은 2월부터 조금씩 확진자가 증가하다가 5월 2일 이태원 클럽발 코로나에 걸린 인천 학원강사로 인해 인천의 코로나 확진자가 증가했음을 알 수 있다. 이처럼 지역별 코로나 증가 추세 양상이 조금씩 다르기 때문에 가설을 기각한다.** |

⑨ 배달 음식의 직접 결제 건수가 현저히 줄었을 것이다.

|  |
| --- |
| # SETLE\_KND\_VALUE는 결제종류번호로 1은 카드, 2는 선불, 3은 현금을 나타낸다. 따라서 직접 결제는 1 + 3의 건수를 나타내므로 1과 3을 하나의 그룹으로 만들고 2를 다른 그룹으로 만들었다.  **# 위의 그림을 보면 직접 만나서 결제하는 건 수가 크게 줄지는 않아 가설을 기각한다.**  **하지만 직접 결제에 비해 1월부터 3월까지 선불결제가 눈에 띄게 증가했다는 것을 알 수 있다. 3월부터 6월까지 선불결제 건수가 감소하는 것을 보아 1월~3월까지는 대구 신천지발 코로나로 사람들이 경각심을 갖고 사람 간의 접촉을 최소화하는 생활을 했지만 코로나로 인한 사회적 거리두기가 장기화됨에 따라 코로나에 대한 경각심을 잃어가는 것으로 추측할 수 있다.** |

⑩ 배달음식 결제수단으로 선불 결제가 가장 많을 것이다.

|  |
| --- |
| d2<-delivery1%>%group\_by(month,SETLE\_KND\_VALUE)%>%count()%>%  mutate(r = if(month %in% c(1,3,5)) {r = n/31} else if(month == 2){r = n/29}   else if(month == 4){r = n/30} else{r= n/21}) ggplot(d2,aes(month,n,fill = factor(SETLE\_KND\_VALUE)))+geom\_bar(stat="identity")+facet\_wrap(~SETLE\_KND\_VALUE)    #1월부터 6월까지 월별 결제수단을 비교해본 결과 선불결제가 가장 많은 것을 알 수 있다. 그러므로 가설은 참이다. |

⑪ 지역별 배달 건수 추세는 비슷할 것이다.

|  |
| --- |
| # 울산광역시의 경우 4월만 존재하는 것으로 보아 자료가 시도별 또는 월별로 균일하게 수집되지 않은 것으로 추측할 수 있다.  **# 위의 결과로 모면 강원도,경기도,세종특별자치시,전라북도,충청남도는 배달건수가 1월부터 5월까지 증가하다가 6월에 감소하는 양상으로 나온다. 이에 반해 대구의 경우 1월과 2월 경의 배달건수 보다 3월 이후 배달건수가 더 낮은 것을 확인할 수 있다. 2월 18일 대구 신천지발 코로나 확진자가 급격히 증가함에 따라 2월까지는 배달건수가 많다가 3월부터 배달건수가 감소하는 것을 알 수 있다. 이렇듯 지역별로 배달건수의 증가추세가 다른 양상을 보이고 있으므로 가설을 기각한다.** |

⑫ 코로나 초반의 소비량이 현재의 소비량보다 많을 것이다.

|  |
| --- |
| # 날짜별 소비량 확인 - 그래프  day\_by <- group\_by(card,receipt\_dttm) day\_by\_sum <- summarise(day\_by, sum=sum(salamt,na.rm=TRUE))  ggplot(day\_by\_sum,aes(x=receipt\_dttm,y=sum,group=1))+geom\_line()    #소비량의 피크가 비슷한 주기로 찍힘.  **#소비량의 피크가 비슷한 주기로 찍힌다. 코로나가 처음 시작되고, 일시적으로 높아진 부분은 있지만 소비 패턴 주기와 양에는 크게 영향을 주지 않았다.** |

⑬ 요일 별 소비량 변화와 코로나는 큰 관련이 없을 것이다.

|  |
| --- |
| # 평일/주말로 나누어 살펴보기  my\_date\_list <- as.Date(parse\_datetime(day\_by\_sum$receipt\_dttm))  day\_by\_sum['week']=factor(weekdays(my\_date\_list)) weekend <- day\_by\_sum[(day\_by\_sum$week == '토요일') | (day\_by\_sum$week == "일요일"),] weekday <- day\_by\_sum[(day\_by\_sum$week == '월요일') | (day\_by\_sum$week == "화요일")| (day\_by\_sum$week == "수요일")| (day\_by\_sum$week == "목요일")| (day\_by\_sum$week == "금요일"),] week2345<-day\_by\_sum[(day\_by\_sum$week == "화요일")| (day\_by\_sum$week == "수요일")| (day\_by\_sum$week == "목요일")| (day\_by\_sum$week == "금요일"),]  day\_by\_sum$receipt\_dttm<-gsub("2020", "", day\_by\_sum$receipt\_dttm)  ggplot(day\_by\_sum,aes(x=receipt\_dttm,y=sum,fill=week))+geom\_point(shape=21)+geom\_line(aes(group=week,colour=week))+theme(axis.text.x = element\_text(angle = 90,size=6))    소비 패턴이 요일마다 다르게 나타남.  소비량이 많은 날들만 따로 빼내어 확인  what<-filter(day\_by\_sum,sum>4.0e+11) what  ## 1 20200129 837943967910 수요일 ## 2 20200401 491993822402 수요일 ## 3 20200428 461461418665 화요일 ## 4 20200505 471882056396 화요일 ## 5 20200512 446559451257 화요일 ## 6 20200528 768194217834 목요일  **#큰 특징과 연관성 없이 우연히 이러한 양상을 보인다는 점에서 가설을 기각할 수 없다.** |

5) 결론

코로나가 현대인의 생활양상과 사회에 어떤 영향을 끼쳤는지에 중점을 두고 분석을 해보았다. 인덱스 데이터를 통해 코로나가 발생하기 전(2019년)보다 코로나가 발생한 후(2020년) 월별 매출지수인 cgi지수가 감소한 것을 알 수 있었다. 특히 연령별로는 20대의 소비가 가장 많이 감소했으며 품목별로는 식품과 화장품 등이 감소됐다는 것을 확인할 수 있었다. 그리고 배달 데이터를 통해 6월을 제외한 1월에서 5월까지는 배달음식을 시킨 횟수가 증가했다는 것과 선불결제가 증가했다는 것을 확인할 수 있었다. 마지막으로 카드데이터를 통해 코로나 발생 이후 온라인 상의 결제와 쇼핑이 많아진 것을 확인했다.

데이터를 분석한 결과를 바탕으로 포스트 코로나 시대에 코로나와 같은 전염병이 다시 발생하게 된다면 기존 사업들이 받는 타격을 줄일 수 있는 혹은 사업을 키울 수 있는 방법에 대해 생각해보았다.

호흡기성 전염병의 경우 마스크 사용으로 색조화장품의 소비가 줄어들기 때문에 색조화장품의 공급량을 줄이는 방법으로 기업의 손실을 줄일 수 있다.

식품 중 기호식품과 빙과류의 소비가 많이 줄었기 때문에 이와 같은 식품류의 공급을 줄임으로써 전염병으로 인한 기업이 입을 수 있는 타격을 최소화할 수 있다.

배달 데이터를 분석해본 결과 1월부터 5월까지 배달건수가 꾸준히 증가하는 것을 확인했다. 포스트 코로나시대에 코로나와 같은 유행성이 강한 전염병이 다시 발생한다면, 요식업 종사자들은 배달산업에 주력하여 할인쿠폰 또는 브랜드 데이 등으로 가격을 낮추어 소비자들의 수요량을 증가시킴으로써 전염병으로 인한 타격을 피해갈 수 있을 것이다. 직접 가정에서 만들어 먹는 식사는 전염병으로 (코로나와 같은) 인해 식료품의 값이 변하더라도 수요량이 크게 변하지 않으므로(보통 식사를 집에서 만들어 먹는 사람들의 경우) 수요의 가격탄력성이 낮다고 볼 수 있다. 하지만 이에 비해 배달 음식의 경우 할인 또는 배달비 무료 쿠폰으로 인한 가격변동이 배달 건수에 영향을 많이 주기 때문에 수요의 가격탄력성이 높다고 볼 수 있다. 따라서 배달 음식의 가격을 낮추게 된다면 수요량이 늘어나므로 이에 맞춰 공급량을 늘려 사회적 최적 상태가 이루어져 효용이 증가하게 될 것이다.

선불 결제를 선호하는 사람이 많기 때문에 배달 업체 자체의 포인트를 충전하여 결제할 시 제공되는 혜택을 늘려 선불 결제를 늘리면 더 많은 이윤을 창출할 수 있을 것이다.

카드 데이터에서 확인했듯이 온라인 몰 상의 할인과 이벤트를 실시함으로써 소비를 촉진시켜 기업의 수입을 늘릴 수 있을 것이다.

앞서 언급한 방안들처럼 소비자와 기업 양쪽의 손실을 최소화함으로써 포스트 코로나 시대 유행성이 강한 질병으로 인해 사회가 받는 타격을 줄일 수 있는 방향으로 대안을 마련해야 할 것이다.