

# 포스트 코로나 시대에 따른 소비자와 산업의 변화

TeamFPB

중앙대학교 응용통계학과

강민석, 김준영, 배형준, 송희성

## 목 차

1. 주제 선정 및 분석 개요.....	2
2. 국내 코로나 확진자 현황 및 포스트 코로나 시대 정의.....	3
3. 코로나 시대 전후의 소비자 심리 변화.....	4
3.1 소비자 심리지수 추이.....	4
3.2 소득수준별 전망 비교.....	6
3.3 연령별 전망 비교.....	9
4. 상품 소분류별 판매지수.....	13
4.1 담배 품목의 성장.....	13
4.2 20 대의 주류 소비량 증가.....	14
5. 소매 업태별 판매지수.....	16
5.1 소매 업태별 현황.....	16
5.2 무점포 소매와 면세점 분석.....	18
6. 코로나 시대의 업종별 추이.....	20
6.1 군집화 방법.....	20
6.2 코로나 이전의 업종별 군집화 결과.....	21
6.3 코로나 이후의 업종별 군집화 결과.....	25
7. 결론.....	29

## 1. 주제 선정 및 분석 개요

유례없는 전염병인 코로나 바이러스 감염증(COVID-19)이 대한민국을 덮쳤습니다. 정치, 경제, 사회, 문화, 의료, 교육 등 다양한 분야의 종사자들이 전염병의 확산을 막기 위해 노력하였고 그로 인해 우리 사회는 많은 변화를 겪고 있습니다. 초등학교부터 대학교까지 학기를 온라인으로 진행하였고 많은 기업들이 재택 근무를 시행하는 등 일상의 비대면화가 빠르게 진행중입니다. 사람들이 많이 모이는 공연, 행사는 대부분 취소되었고 불가피하게 모이는 일이 있으면 소독과 방역이 뒤따랐습니다.

다양한 변화 중 저희 팀은 경제적인 측면에 주목하였습니다. 경제 부문은 일상생활에 가장 밀접한 영향을 끼치기 때문입니다. 코로나 시대에 접어들면서 바뀐 일상을 통해 소비자들의 심리 상태와 소비 패턴이 코로나 이전과는 많이 다를 것이라 예상하였습니다. 또한 코로나에 타격을 받은 산업과 그렇지 않은 산업, 코로나 속에서도 성장하는 산업을 관찰함으로써 흥미로운 사실들을 발견할 수 있을 것이라고 생각하였습니다.

이에 저희는 **“포스트 코로나 시대에 따른 소비자와 산업의 변화”**를 주제로 분석을 진행하였습니다. 분석 목적과 사용한 데이터는 다음과 같습니다.

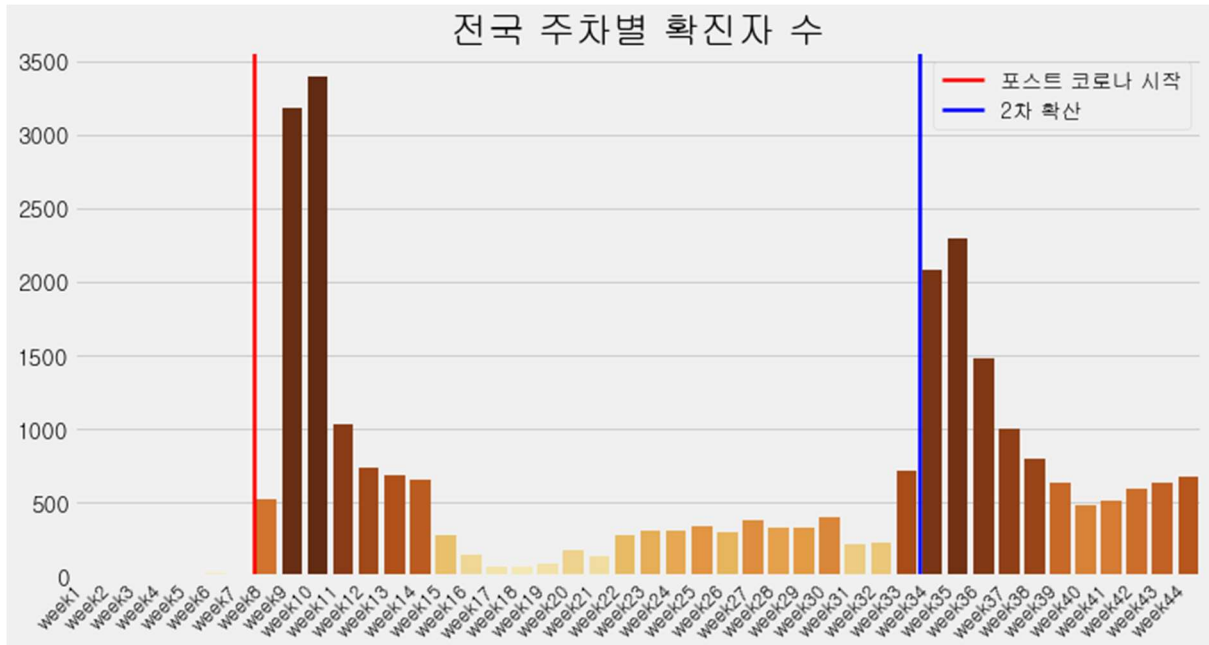
### ● 분석 목적

- 1) 포스트 코로나 시대 정의
- 2) 코로나 전후 소비자의 심리 변화 관찰
- 3) 상품별, 소매업체별 판매지수에 대한 코로나의 영향 분석
- 4) 업종별 판매지수 분석을 통한 산업 전망 분석

### ● 사용 데이터

- 1) 보건복지부 코로나19 감염 현황
- 2) 품목별 소비지수 데이터
- 3) 월별 소비자 동향조사
- 4) 소매업체별 월별, 분기별 판매액 계절조정지수
- 5) 온·오프라인 유통업체 식품군 월간 전년동기대비 성장률
- 6) 기간별 해외 입국자 수
- 7) 업종별 매출 실적

## 2. 국내 코로나 확진자 현황 및 포스트 코로나 시대 정의



코로나 전후의 소비자와 산업의 변화를 관찰하기 위해선 포스트 코로나 시대의 시작 시점을 명확하게 정의할 필요가 있습니다. 위의 주차별 확진자 수 그래프를 보면, 8주차에 확진자 수가 갑자기 500명 이상으로 증가한 것과 34주차에 확진자 수가 전주대비 다시 한 번 크게 증가한 것을 확인할 수 있었습니다. 이를 통해 8주차(2월 17일~2월 23일)를 포스트 코로나 시대 시작 시점으로, 34주차(8월 16일~8월 23일)를 2차 확산 시점으로 정의하였습니다.

따라서 이후 분석에서는 기간이 1년 이하인 데이터에 대해선 2월(8주차 포함)과 8월(34주차 포함)을 기준으로 전후를 구분하여 비교하였습니다. 기간이 1년 이상인 데이터에 대해선 필요에 따라 20년 1월 또는 2월을 기준으로 전후를 구분하여 비교하였습니다. 1월을 기준으로 사용한 경우는 1월에서 2월로 변화하는 양상을 관찰하기 위함입니다. 코로나 확산에 큰 영향을 미친 주요 사건들은 다음과 같습니다.

### ● 주요 사건

4주차 (1월 20일) : 국내 첫 코로나 확진자 발생

8주차 (2월 20일) : 신천지 대구 교회에서 38명의 대규모 감염 발생

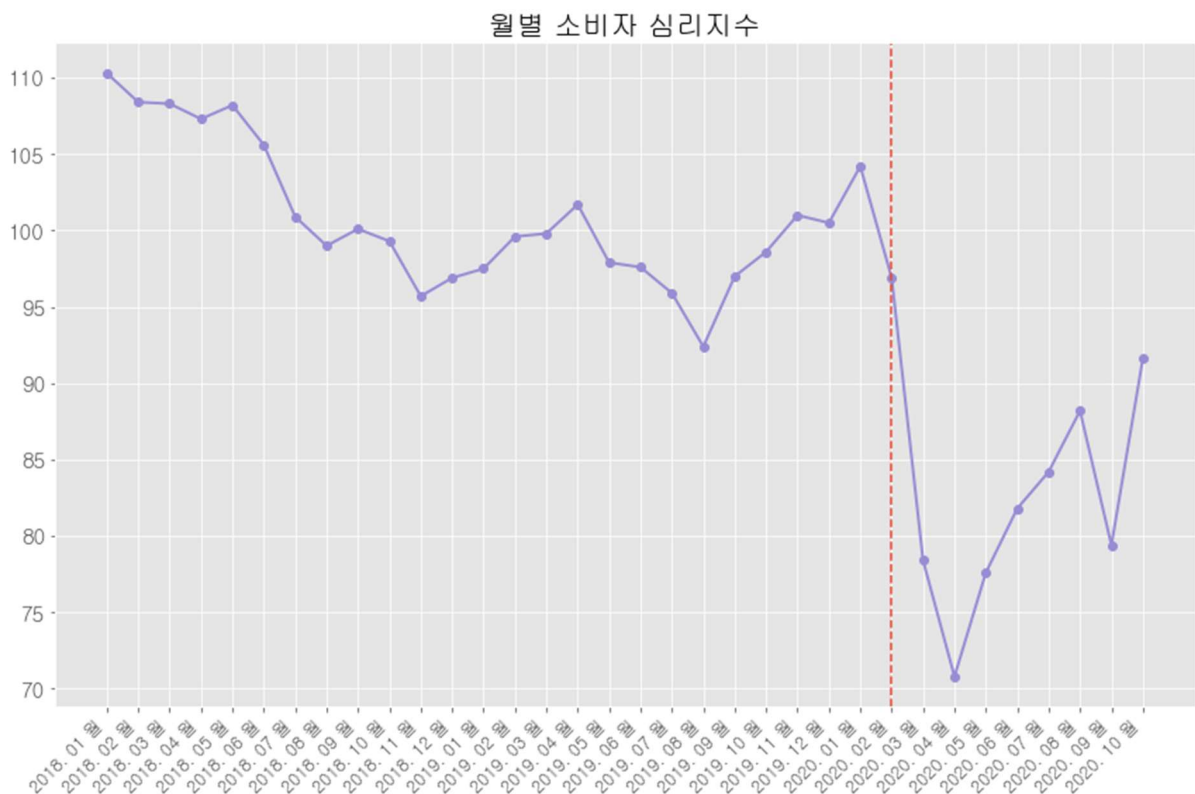
34~35주차 (8월 16일~8월 30일): 광복절 집회 이후 수도권을 중심으로 대유행 감염

36~38주차 (8월 30일~9월 20일) : 수도권 사회적 거리두기 2.5단계 시행

### 3. 코로나 시대 전후의 소비자 심리 변화

산업을 분석하기에 앞서 소비자에 대한 분석이 선행되어야 한다고 생각했습니다. 같은 산업 분석 결과라도 소비자의 심리에 따라 다른 해석이 나올 수 있기 때문입니다. 국가 통계포털의 월별 소비자동향 지수 데이터를 활용하여 코로나 이전과 이후의 전반적인 소비자 심리지수의 차이를 확인하고, 세부적으로 연령별 및 소득수준별 소비 전망과 현황에 대해 살펴보았습니다. 18년 1월~20년 10월까지의 데이터에서 18년 1월~20년 1월을 코로나 이전 시대로, 20년 2월~20년 10월을 코로나 이후 시대로 데이터를 분리하였습니다.

#### 3.1 소비자 심리지수 추이



최근 3년간 소비자 심리지수의 전체적인 추세는 다음과 같습니다. 18년 1월부터 19년 8월까지 감소하는 추세로 해당 기간동안 소비 심리가 점점 얼어붙었으나 19년 9월을 기점으로 반등하여 20년 1월까지 증가 추세를 보여줍니다. 하지만 20년 2월 코로나의 영향으로 20년 4월에 심리지수가 약 70까지 감소한 것을 확인할 수 있습니다. 코로나 전후의 심리지수 차이를 통계적으로 검증하기 위하여 20년 2월을 기준으로 데이터를 나누는 뒤 등분산검정 및 평균 차이 검정을 실행하였습니다.

```
> var.test(CCSI_bf$value,CCSI_af$value)

      F test to compare two variances

data:  CCSI_bf$value and CCSI_af$value
F = 0.34676, num df = 24, denom df = 8, p-value = 0.04228
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.08784894 0.96369032
sample estimates:
ratio of variances
 0.3467591

> t.test(CCSI_bf$value,CCSI_af$value, var.equal = FALSE, alternative = "greater")

      welch Two sample t-test

data:  CCSI_bf$value and CCSI_af$value
t = 6.291, df = 10.07, p-value = 4.371e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 12.63044      Inf
sample estimates:
mean of x mean of y
100.94800  83.21111
```

- 등분산검정

$H_0$  : 두 집단의 분산은 같다.

$H_1$  : 두 집단의 분산은 같지 않다.

유의수준 0.05 하에서 검정통계량 F의 값은 0.34676으로 유의확률은 0.04228입니다. 유의확률이 유의수준보다 작으므로 두 집단의 분산이 같다는 영가설을 기각합니다. 두 집단의 분산이 같지 않으니 var.equal=FALSE인자를 사용하여 Welch T test를 시행하도록 하겠습니다.

- 평균 차이 검정

$H_0$  : 두 집단의 평균은 같다.

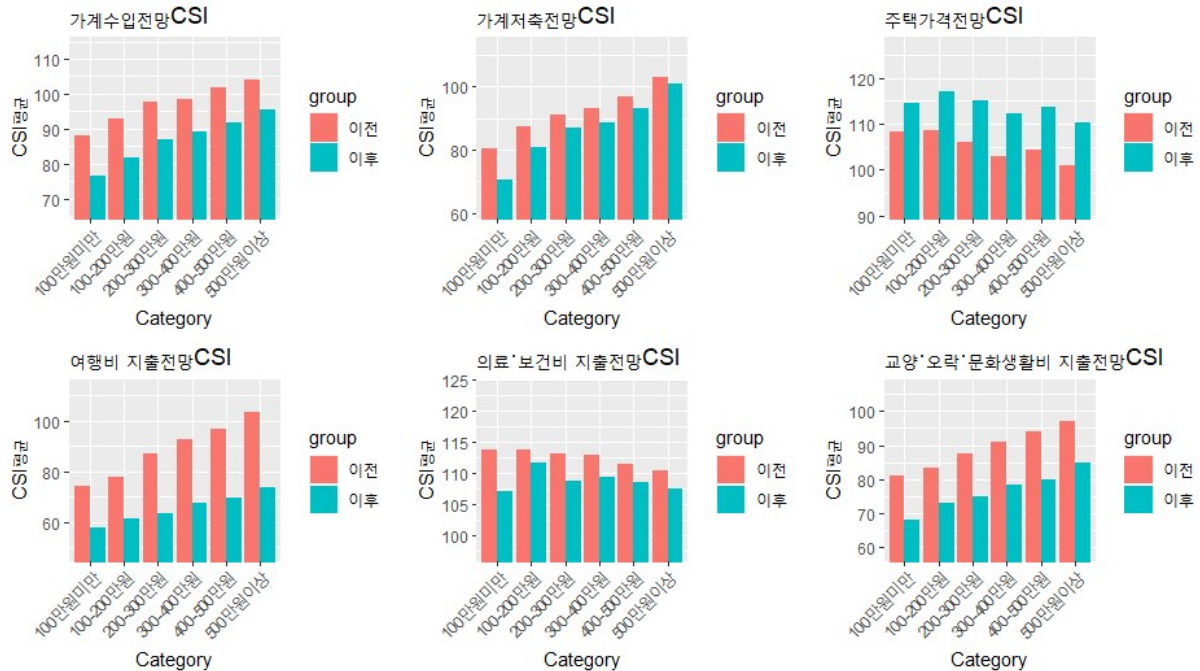
$H_1$  : 두 집단의 평균은 같지 않다.

유의수준 0.05 하에서 검정통계량 T의 값은 6.291으로 유의확률은  $4.371 \times 10^{-5}$ 입니다. 유의확률이 유의수준보다 작으므로 두 집단의 평균이 같다는 영가설을 기각합니다. 따라서 코로나 이전의 소비자 심리지수 평균이 코로나 이후의 소비자 심리지수 평균보다 크다고 볼 수 있습니다.

이후 분석에서는 소비자 심리지수뿐만 아니라 가계수입전망, 가계지출전망, 현재생활형편 등 다양한 CSI를 소득수준별, 연령별로 나누어 전망을 비교해보도록 하겠습니다.

## 3.2 소득수준별 전망 비교

모든 지수코드를 대상으로, CSI평균을 이용하여 소득수준별로 Bar-graph를 그려보았습니다. 그 이후 소득수준별로 특징이 뚜렷한 CSI지수 6개를 선별하여 해석을 해보았습니다.



- 1) 전반적으로 코로나의 발생이 모든 소득 별 소비자들의 소비심리에 큰 영향을 미친 것으로 확인되었습니다.
- 2) 대표적으로 위 그래프의 가계수입전망, 가계저축전망을 확인해보았을 때 코로나로 인해 떨어지는 폭이 저소득층의 폭이 고소득층의 폭보다 크게 확인되는 것으로 보아 코로나의 발생이 저소득층에게 치명적으로 다가온다는 것을 알 수 있었습니다.
- 3) 주택가격전망CSI를 확인해 보았을 때 저소득층이 코로나 이전과 이후 모두 고소득층에 비해 주택가격의 상승을 높게 전망하는 것을 확인할 수 있었습니다.
- 4) 여행 산업에서의 고소득층 소비전망의 큰 폭의 하락은 코로나 이후 여행이나 숙박업에 큰 불황을 가져올 것이라고 예상됩니다.

실제로 소득층별로 나누어 전망을 해석하는 것이 타당한지 확인하기 위해 각 CSI 지수별로 분산분석을 진행하였습니다. 모든 가설 검정들에서 '소득 수준 별 CSI평균이 모두 같다'는 영가설이 기각되었고, 이는 소득 계층별 CSI전망이 각각 다르다는 것을 의미합니다. 조금 더 자세하게 알아보기 위해 Tukey 사후검정을 진행해 보았습니다.

가계수입전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value	groups
분류코드별	5	4206	841.2	197.037	< 2e-16 ***	500만원이상	103.92	a
variable	24	826	34.4	8.066	1.33e-15 ***	400-500만원	101.88	b
Residuals	120	512	4.3			300-400만원	98.60	c
---						200-300만원	97.72	c
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						100-200만원	92.88	d
						100만원미만	88.24	e
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value	groups
분류코드별	5	2159.4	431.9	104.02	< 2e-16 ***	500만원이상	95.55556	a
variable	8	886.6	110.8	26.69	9.92e-14 ***	400-500만원	91.66667	b
Residuals	40	166.1	4.2			300-400만원	89.00000	bc
---						200-300만원	87.00000	c
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						100-200만원	81.55556	d
						100만원미만	76.44444	e

가계저축전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value	groups
분류코드별	5	7472	1494.3	334.372	< 2e-16 ***	500만원이상	102.96	a
variable	24	271	11.3	2.528	0.000515 ***	400-500만원	96.80	b
Residuals	120	536	4.5			300-400만원	93.28	c
---						200-300만원	91.04	d
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						100-200만원	87.48	e
						100만원미만	80.48	f
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value	groups
분류코드별	5	4749	949.9	207.62	< 2e-16 ***	500만원이상	100.66667	a
variable	8	385	48.1	10.52	8.58e-08 ***	400-500만원	92.88889	b
Residuals	40	183	4.6			300-400만원	88.55556	c
---						200-300만원	87.00000	c
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						100-200만원	80.88889	d
						100만원미만	70.66667	e



주택가격전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	5	1147	229.5	35.17	<2e-16 ***	100-200만원	108.72 a
variable	24	19672	819.7	125.62	<2e-16 ***	100만원미만	108.32 a
Residuals	120	783	6.5			200-300만원	106.04 b
---						400-500만원	104.24 bc
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						300-400만원	103.00 cd
						500만원이상	101.04 d
							value groups
						100-200만원	117.1111 a
						200-300만원	115.1111 ab
						100만원미만	114.5556 ab
						400-500만원	113.5556 abc
						300-400만원	112.3333 bc
						500만원이상	110.2222 c
							value groups
						100-200만원	117.1111 a
						200-300만원	115.1111 ab
						100만원미만	114.5556 ab
						400-500만원	113.5556 abc
						300-400만원	112.3333 bc
						500만원이상	110.2222 c

여행비 지출전망CSI 에 대한 분산분석 결과입니다.

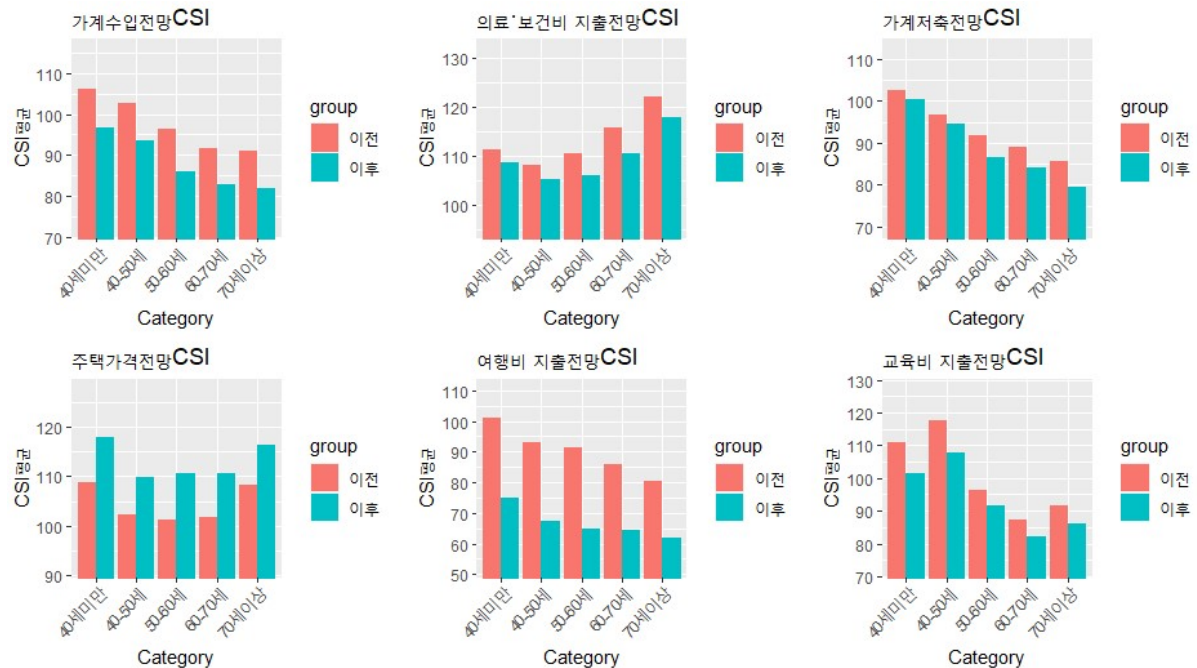
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	5	15536	3107.2	501.30	<2e-16 ***	500만원이상	103.20 a
variable	24	1570	65.4	10.56	<2e-16 ***	400-500만원	96.88 b
Residuals	120	744	6.2			300-400만원	92.40 c
---						200-300만원	87.04 d
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						100-200만원	77.80 e
						100만원미만	74.24 f
							value groups
						500만원이상	73.55556 a
						400-500만원	69.55556 b
						300-400만원	67.33333 b
						200-300만원	63.33333 c
						100-200만원	61.33333 c
						100만원미만	57.66667 d

각 검정들에서 위쪽은 코로나 이전, 아래는 코로나 이후 데이터들 검정한 것입니다. Tukey 사후검정 결과 대부분의 CSI 지수에서 임금이 높을수록 CSI 평균 지수가 높다는 것을 확인하였습니다. 또한 소득이 200 만원 미만인 소비자들과 400 만원 이상인 소비자들은 모든 CSI 지수에서 다른 그룹으로 분류된다는 것을 알 수 있습니다.



### 3.3 연령별 전망 비교

모든 지수코드를 대상으로, CSI평균을 이용하여 소득수준별로 Bar-graph를 그려보았습니다. 그 이후 연령별로 특징이 뚜렷한 CSI지수 6개를 선별하여 해석을 해보았습니다.



- 1) 코로나의 발생이 연령별 소비지수에서도 큰 영향을 미친 것을 알 수 있습니다.
- 2) 주택가격전망을 보면, 40 세 미만과 70 세 이상 연령층이 상대적으로 40 세~70 세 연령층에 비해 주택 가격에 대해 높게 전망하고 있는 것으로 나타났습니다.
- 3) 교육비 지출전망을 보면 코로나 이전에는 양육을 하는 40 세 이상과 50 세 미만 연령층이 교육비 지출전망을 높게 보았고, 그 외 연령대는 지출전망이 낮게 관측되었습니다. 코로나 이후 역시 40 세 이상과 50 세 미만 연령층의 경우 10 정도의 감소폭을 보였으나, 그 외의 연령대에서는 5 정도의 감소를 보였습니다.
- 4) 가계저축전망에 대해서는 전반적으로 소비가 줄어들어 코로나 이후에 더 저축전망이 올라갈 것으로 기대하였으나 기대와는 다르게 코로나로 인해 소폭 감소를 보였다는 점과 각 연령별로 큰 차이가 없었다는 점이 의외였습니다. 코로나로 인한 재택근무나 무급휴가, 또는 중지된 산업으로 인해 가계수입전망이 떨어진 것이 저축전망에도 영향을 미쳤습니다.

- 5) 의료보건비 지출전망을 보면 전반적으로 100 이상임을 확인할 수 있었습니다. 특히 60 세 이상에서의 지출전망이 높는데, 의료보건비에서 60 세 이상이 지출 부담을 크게 느끼는 것이 아닐까 하는 생각이 듭니다.

소득별로 진행했던 것과 같은 방식으로 해당 CSI 지수별로 분산분석을 진행하였고, 모든 평균이 모두 같다는 영가설이 기각되었고, 이 의미는 연령별 CSI전망이 각각 다르다는 것을 의미합니다. 조금 더 자세하게 알아보기 위해 Tukey 사후검정을 진행해 보았습니다. 각 검정들에서 위쪽은 코로나 이전, 아래는 코로나 이후 데이터들 검정한 것입니다.

가계수입전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	4480	1119.9	352.164	< 2e-16 ***	40세미만	106.24 a
variable	24	610	25.4	7.993	4.68e-14 ***	40-50세	102.56 b
Residuals	96	305	3.2			50-60세	96.28 c
---						60-70세	91.76 d
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						70세이상	90.96 d

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	1565.6	391.4	124.81	< 2e-16 ***	40세미만	96.66667 a
variable	8	671.2	83.9	26.75	4.4e-12 ***	40-50세	93.55556 b
Residuals	32	100.4	3.1			50-60세	86.11111 c
---						60-70세	82.77778 d
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						70세이상	81.88889 d

가계저축전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	4437	1109.3	267.056	< 2e-16 ***	40세미만	102.68 a
variable	24	223	9.3	2.233	0.00319 **	40-50세	96.72 b
Residuals	96	399	4.2			50-60세	91.92 c
---						60-70세	89.00 d
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						70세이상	85.72 e

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	2540.8	635.2	216.13	< 2e-16 ***	40세미만	100.33333 a
variable	8	246.2	30.8	10.47	4.56e-07 ***	40-50세	94.66667 b
Residuals	32	94.0	2.9			50-60세	86.66667 c
---						60-70세	84.00000 d
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						70세이상	79.44444 e

가계수입전망, 가계저축전망에선 나이가 어릴수록 CSI 의 평균 값이 높다는 것을 확인하였습니다. 코로나 이후에도 순위의 변동은 없었습니다. 60 세 이상, 70 세 이상 외에는 모두 다른 그룹으로 묶였습니다.

주택가격전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	1451	362.8	50.04	<2e-16 ***	40세미만	108.92 a
variable	24	18032	751.3	103.62	<2e-16 ***	70세이상	108.40 a
Residuals	96	696	7.3			40-50세	102.32 b
---						60-70세	101.64 b
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						50-60세	101.28 b
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	510	127.4	24.14	2.82e-09 ***	40세미만	118.0000 a
variable	8	5068	633.5	120.06	< 2e-16 ***	70세이상	116.3333 a
Residuals	32	169	5.3			50-60세	110.6667 b
---						60-70세	110.6667 b
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						40-50세	109.8889 b

주택가격전망에선 40 세미만 과 70 세 이상이 한 그룹으로 묶이고, 그 외 연령층이 묶이는 것을 확인하였습니다.

의료·보건비 지출전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	2988.4	747.1	336.983	< 2e-16 ***	70세이상	122.00 a
variable	24	109.2	4.5	2.052	0.00752 **	60-70세	115.80 b
Residuals	96	212.8	2.2			40세미만	111.44 c
---						50-60세	110.60 c
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						40-50세	108.08 d
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	907.0	226.76	57.690	3.52e-14 ***	70세이상	117.8889 a
variable	8	162.0	20.25	5.152	0.000352 ***	60-70세	110.5556 b
Residuals	32	125.8	3.93			40세미만	108.7778 bc
---						50-60세	106.2222 cd
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						40-50세	105.2222 d

여행비 지출전망CSI 에 대한 분산분석 결과입니다.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	6004	1501.0	200.427	< 2e-16 ***	40세미만	100.92 a
variable	24	1279	53.3	7.116	1.17e-12 ***	40-50세	93.28 b
Residuals	96	719	7.5			50-60세	91.52 b
---						60-70세	85.76 c
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						70세이상	80.48 d
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		value groups
분류코드별	4	926.9	231.7	61.02	1.61e-14 ***	40세미만	75.11111 a
variable	8	1957.4	244.7	64.44	< 2e-16 ***	40-50세	67.44444 b
Residuals	32	121.5	3.8			50-60세	65.00000 bc
---						60-70세	64.55556 c
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						70세이상	61.77778 d

교육비 지출전망CSI 에 대한 분산분석 결과입니다.

							value groups	
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		40-50세	117.60 a
분류코드별	4	16652	4163	964.646	< 2e-16 ***		40세미만	111.08 b
variable	24	432	18	4.175	2.74e-07 ***		50-60세	96.56 c
Residuals	96	414	4				70세이상	91.76 d
---							60-70세	87.44 e
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1								
							value groups	
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		40-50세	107.88889 a
분류코드별	4	4115	1029	347.1	< 2e-16 ***		40세미만	101.66667 b
variable	8	384	48	16.2	2.93e-09 ***		50-60세	91.66667 c
Residuals	32	95	3				70세이상	86.00000 d
---							60-70세	82.33333 e
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1								

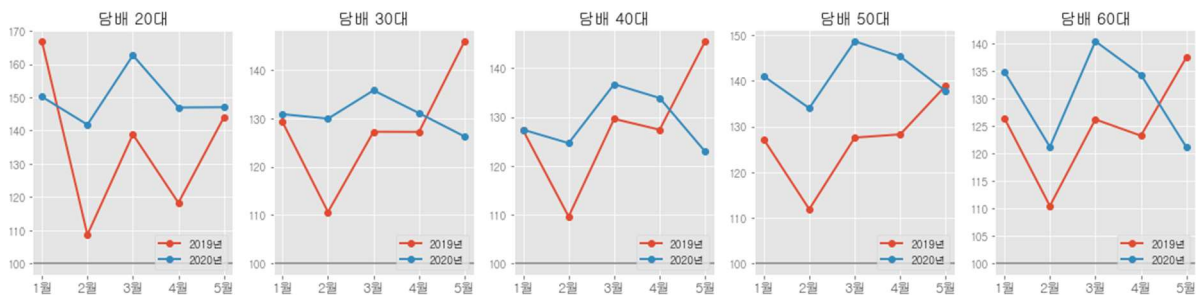
나머지 지출전망들에 대한 Tukey 사후검정 결과 70 세 이상과 40 세 미만은 같은 그룹으로 분류된 적이 없습니다.



## 4. 상품 소분류별 판매지수

전반적으로 경기가 좋지 않음을 쉽게 얘기할 수 있으나, 직접 체감할 수 있는 항공업, 숙박업 등을 제외하고는 구체적으로 어떤 부분이 얼마나 타격을 받았는지는 알기 어렵습니다. 이를 알아보기 위해 품목별 소비지수 데이터를 통해 어떤 양상을 띠는지 확인해보고자 합니다. 2018년 평균 판매지수를 100으로 본 데이터입니다. 주어진 데이터의 지역이 서울에만 한정되어 있고, 기간이 2019년 01월부터 2020년 5월까지인 점에서 한계가 뚜렷하지만, 코로나 시기 전후의 차이를 잘 보여줍니다. 자세한 차이를 보기 위해 연령별, 성별로 세분화해서 분석해보았습니다. 카테고리 성장지수의 평균을 기준으로 코로나 이전 상위 3개 품목과 하위 3개 품목, 코로나 이후 상위 3개 품목과 하위 3개 품목은 크게 변화가 없습니다. 양쪽 모두 상위 3개 품목은 담배, 빵, 가공식품이고, 하위 3개 품목은 뷰티소품, 화장품, 애완동물용품으로 절대적인 수치 자체는 감소하지만, 품목의 변화는 거의 없었습니다.

### 4.1 담배 품목의 성장



```
1 index_summarise = masking(index_data, age=['all'], gender='all')
2
3 index_pivot_bf = index_summarise.iloc[1:6, :]['담배']
4 index_pivot_af = index_summarise.iloc[-5:, :]['담배']
```

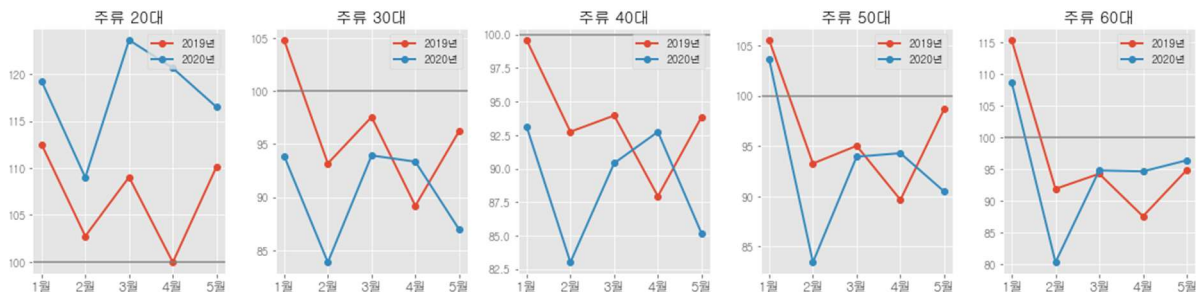
```
1 mannwhitneyu(index_pivot_bf, index_pivot_af)
```

MannwhitneyuResult(statistic=7.0, pvalue=0.1481349357421432)

대부분의 품목이 작년 동기대비 하락하는 모습을 보이는 것으로 보아 코로나의 여파는 상당히 의미하다고 할 수 있습니다. 그럼에도 불구하고 담배 품목은 작년 동기대비 상승하는 점을 볼 수 있습니다. 샘플 사이즈가 각각 5개이기 때문에 모수 검정이 적합하지 않다고 생각하여 비모수 검정인 Mann-Whitney 검정을 하였습니다. 귀무가설과 대립가설을 각각 "코로나 전,후 두 집단의 담배 판매 지수에는 차이가 없다.", "코로나 전,후 두 집단의 담배 판매 지수에는 차이가 있다."로 설정한 two-sided 통계 검정 결과 유의수준 0.05에서 p-value가 0.148로 0.05보다 크기 때문에 귀무가설을 기각할 수 없는, 즉 두 집

단 간 차이가 없음을 확인하였습니다. 그럼에도 불구하고 전 연령에서 담배 소비가 증가한 점에서 어느 정도 힘든 시기를 겪고 있음을 짐작할 수 있습니다.

## 4.2 20대의 주류 소비량 증가



연령별 비교에서 가장 주목할 부분은 주류 부분입니다. 20대를 제외한 전 연령대에서 주류는 기준점인 100보다 작은 모습을 보여줍니다. 그러나 20대의 경우 코로나 이전에도 100이상의 지수를 보이지만, 코로나 이후에는 더 증가한 모습을 보여줍니다. 샘플 사이즈가 각각 5개이기 때문에 모수 검정이 적합하지 않다고 생각하여 비모수 검정인 Mann-Whitney 검정을 시행하였습니다. 귀무가설과 대립가설을 각각 “코로나 전, 후 두 집단의 20대 주류 판매 지수에는 차이가 없다.”, “코로나 전, 후 두 집단의 20대 주류 판매 지수에는 차이가 있다.”로 설정한 two-sided 통계 검정 결과 유의수준 0.05에서 p-value가 0.03으로 0.05보다 작기 때문에 귀무가설을 기각, 즉 두 집단 간 차이가 존재함을 확인하였습니다.

```
1 index_summarise = masking(index_data, age=['20'], gender='all')
2
3 index_pivot_bf = index_summarise.iloc[1:6, :]['주류']
4 index_pivot_af = index_summarise.iloc[-5:, :]['주류']
```

```
1 mannwhitneyu(index_pivot_bf, index_pivot_af)
```

MannwhitneyuResult(statistic=3.0, pvalue=0.030051402969433157)

다른 연령대와 비교하기 위해 ANOVA 검정을 시행하였습니다. 귀무가설과 대립가설을 각각 ‘연령별 집단 간 주류 판매지수에는 차이가 없다.’, ‘연령별 집단 간 주류 판매지수에는 차이가 있다.’로 설정하였습니다.

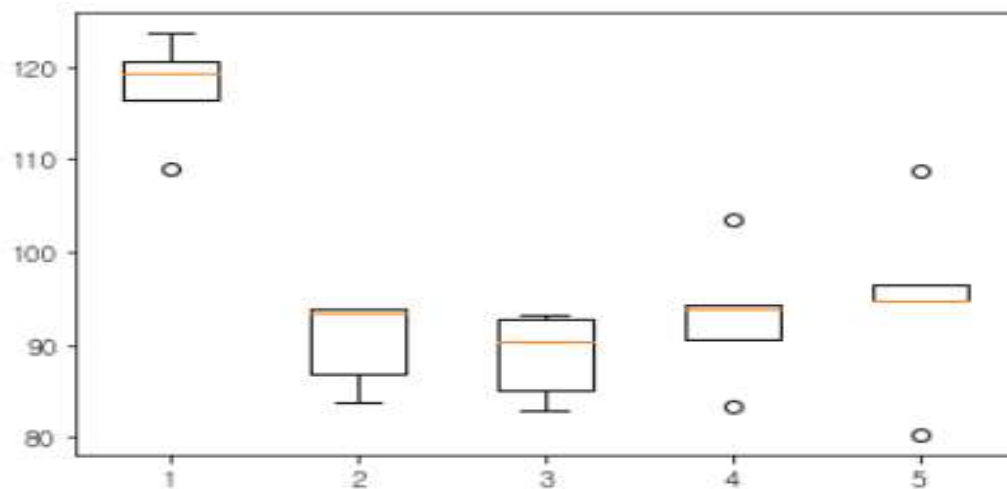
```
1 kruskal(df_20, df_30, df_40, df_50, df_60)
```

KruskalResult(statistic=14.658461538461552, pvalue=0.005464625635224572)

통계 검정 결과 유의수준 0.05에서 p-value가 0.00546으로 유의수준 0.05 하에서 p-value가 유의수준보다 작기 때문에 귀무가설을 기각하였습니다. 따라서 연령별 집단간



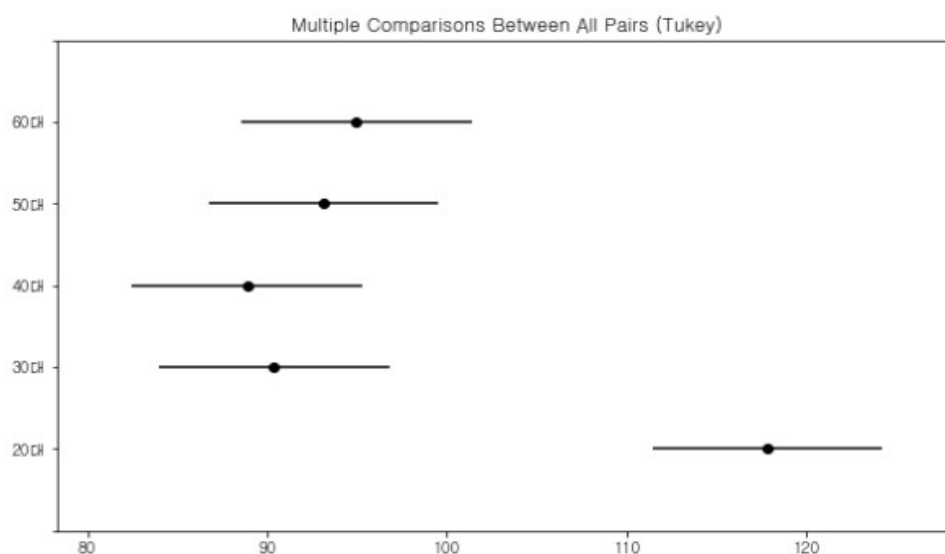
주류 판매지수에는 차이가 존재합니다.



Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
20CH	30CH	-27.4034	0.001	-40.2018	-14.605	True
20CH	40CH	-28.9197	0.001	-41.7181	-16.1213	True
20CH	50CH	-24.6504	0.001	-37.4488	-11.852	True
20CH	60CH	-22.8241	0.001	-35.6225	-10.0257	True
30CH	40CH	-1.5163	0.9	-14.3147	11.2821	False
30CH	50CH	2.7531	0.9	-10.0453	15.5515	False
30CH	60CH	4.5794	0.7981	-8.219	17.3778	False
40CH	50CH	4.2694	0.8372	-8.529	17.0678	False
40CH	60CH	6.0956	0.6069	-6.7028	18.894	False
50CH	60CH	1.8263	0.9	-10.9721	14.6247	False

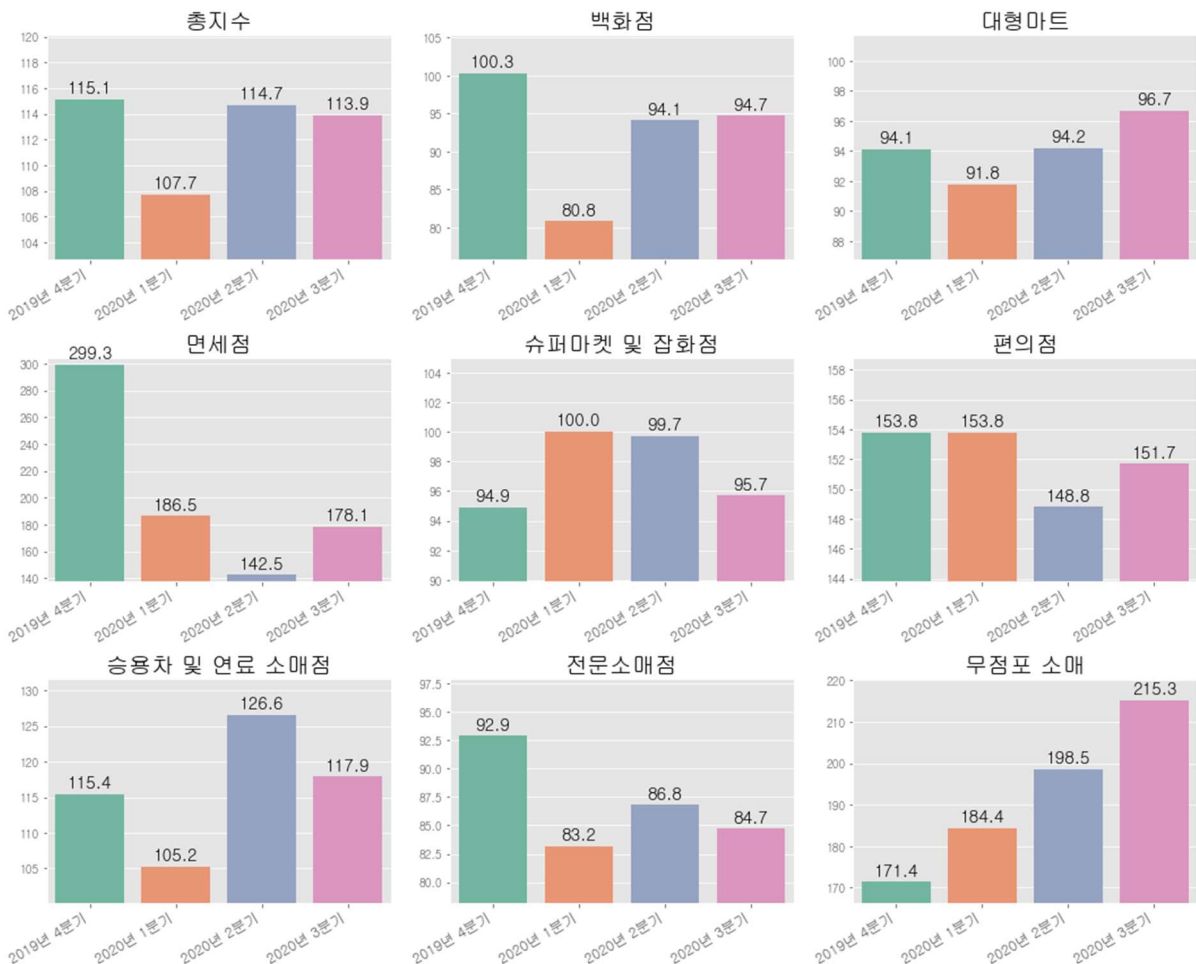
사후 분석으로 Tukey 검정 결과, 20대 외의 연령대에서는 차이가 나지 않지만, 20대는 다른 연령대와 차이가 나는 결과를 보이기 때문에 20대의 주류 소비량이 다른 연령대 집단들에 비해 눈에 띄게 많다는 결론이 나옵니다.



## 5. 소매 업체별 판매지수

코로나에 가장 많은 영향을 받은 기업은 대면을 기반으로 다수의 소비자와 자주 접촉하는 소매업체라고 생각하였습니다. 그래서 소매업체별 월별 (18년 1월~20년 9월), 분기별 (19년 4분기~20년 3분기) 판매지수 데이터를 통해 소매 업체별로 어떤 양상을 띠고 있는지 확인해보려고 합니다. 추세 비교를 위해 비대면 기반인 무점포 소매도 포함하였습니다. 일반적인 판매지수는 계절성이 강해 추세를 확인하기 힘들다고 판단하여 계절조정지수를 이용하였습니다. (2015년 분기별 데이터에선 분기 평균 판매액이 100, 2015년 월별 데이터에선 월 평균 판매액이 100입니다.)

### 5.1 소매 업체별 현황



총 지수부터 살펴보겠습니다. 19년 4분기에 115이던 지수가 20년 1분기에 코로나의 여파로 107로 감소하였지만 20년 2분기 114로 금세 회복한 것을 확인할 수 있습니다. 이 사실을 통해 코로나로 인해 개별 소매업체별로 등락이 존재하지만 전체적으로 봤을 땐 코로나 이전만큼 소비되고 있다고 볼 수 있습니다. 즉, 소비하는 방법이 변화했을 뿐 전

체 소비 금액은 비슷한 수준으로 이뤄지고 있다고 해석할 수 있습니다.

가장 눈에 띄는 변화는 면세점의 급락과 무점포 소매의 급등입니다. 면세점의 경우 국내 소비자보다는 해외 관광객에 의해 매출이 결정되는 편입니다. 코로나로 인해 해외 관광객이 급락했기 때문에 매출 감소를 직격탄으로 맞았습니다. 추세를 봤을 때 19년 4분기만큼의 실적으로 회복되기에는 많은 시간이 필요할 것으로 보입니다. 반면 홈쇼핑, 인터넷 쇼핑, 배달 소매점 등을 포함하고 있는 무점포 소매는 코로나로 인해 지속적인 성장을 경험하고 있습니다. 이는 대면 소비에서 비대면 소비로의 전환을 보여주는 대표적인 사례라고 생각합니다. 두 소매업태에 대해선 뒤에서 한 번 더 다루도록 하겠습니다.



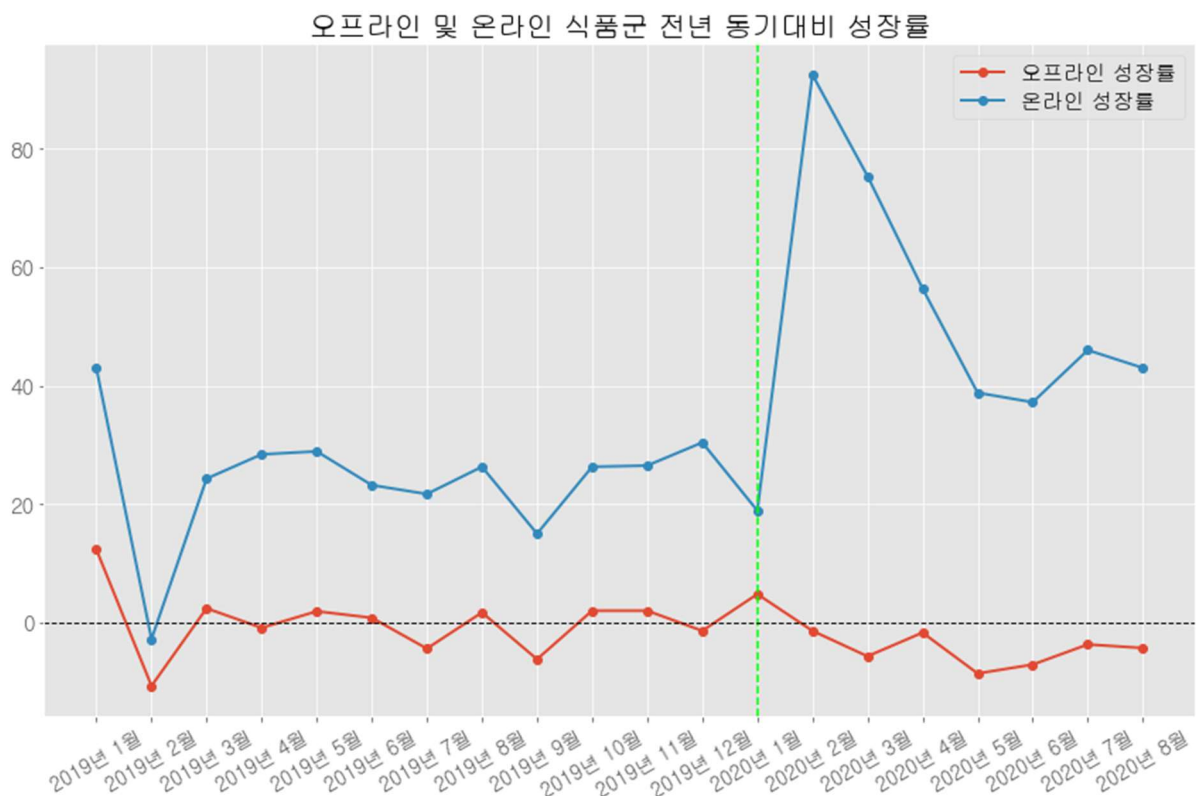
백화점, 대형마트, 슈퍼마켓 및 잡화점, 편의점의 3년치 추이는 위와 같습니다. 백화점은 18-19년도에 현상 유지를 지속하다가 코로나 쇼크로 20년 2월에 1월 대비 판매지수가 22.6 하락하였습니다. 동시기에 판매지수가 11.9 증가한 슈퍼마켓 및 잡화점과는 대비되는 현상입니다. 소비자들이 전염병을 피하기 위해 사람들이 많이 모이는 백화점보다는 집 주변의 슈퍼마켓 및 잡화점을 많이 이용하고 있는 상황으로 보입니다. 하지만 2차 확산인 8월에는 다른 양상을 보여줍니다. 백화점의 경우 7월 대비 8월의 판매지수가 3.6 정도만 하락하였고 9월에는 오히려 상승하였습니다. 이는 소비자들이 코로나에 대한 공포심이 약해져서 이전만큼 대면 쇼핑을 자제하지 않는 것으로 보입니다.

대형마트의 경우 지난 2년간은 약간의 감소 추세를 보이다가 코로나 이후에는 판매지

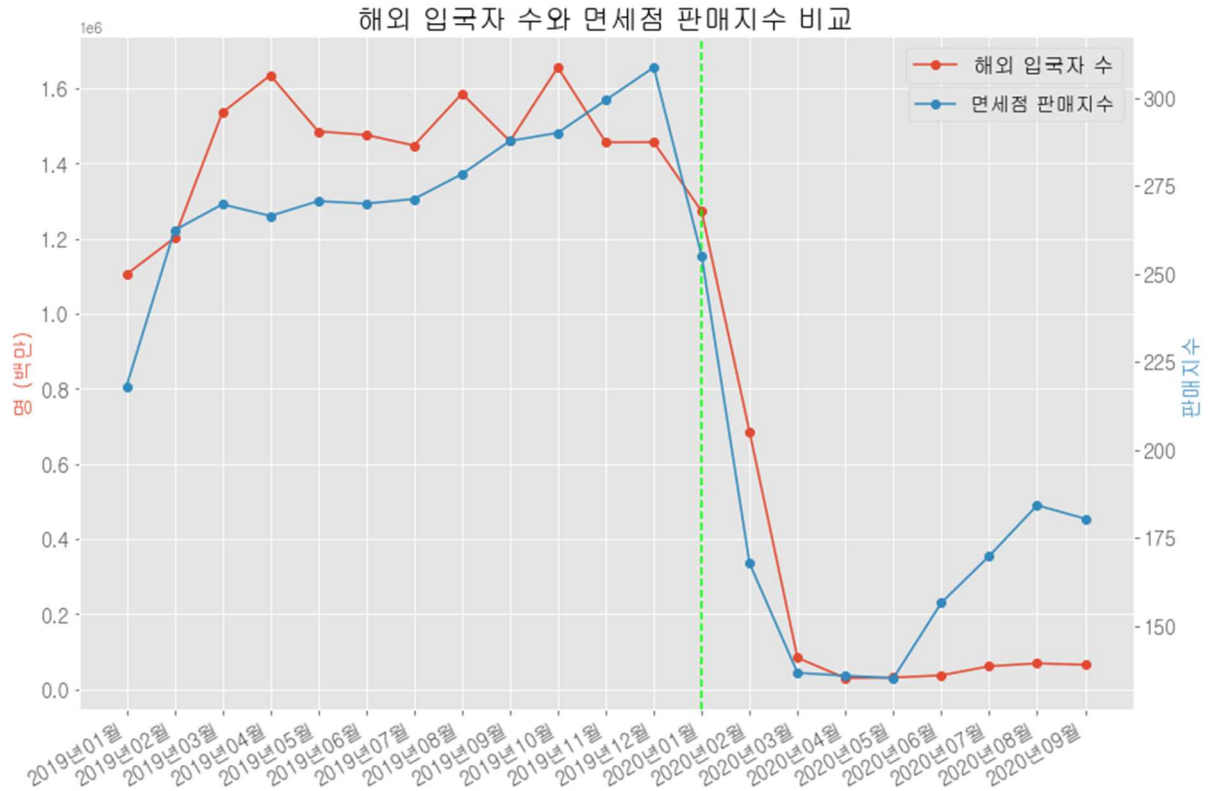
수 90에서 100 사이를 왔다갔다하는 모습을 보여줍니다. 9월 매출에는 추석을 '언택트 명절'로 보내면서 고가 선물 판매가 늘어난 게 실적에 반영된 것으로 보입니다. 백화점처럼 빠지지 않는 것으로 보아 소비자들이 대형마트까지는 코로나를 무릅쓰고 많이 이용하는 것으로 보입니다.

가장 의외인 소매업태는 편의점이었습니다. 주거지와 가장 밀접하게 위치해 있어서 코로나때에도 증가할 것으로 예상했지만 20년 2~3월 전체적인 유동인구 감소로 인해 대학가, 유흥가 지점들의 매출이 급락하여 판매지수가 1월 대비 17.7정도 하락한 모습을 보여줍니다.

## 5.2 무점포 소매와 면세점 분석



온라인 쇼핑의 등장은 1996년부터 시작하였습니다. 특히 최근 들어 급격한 성장을 보입니다. 무점포 소매의 가파른 성장엔 다양한 요인이 있겠지만 저희 팀은 식품군에서의 성장에 주목하였습니다. 포스트 코로나 시대의 시작인 20년 2월~3월에 온라인 식품군 판매량이 가장 큰 성장률을 기록한데 비해 오프라인 식품군의 성장률은 마이너스 값을 보여줍니다. 이는 온, 오프라인 시장 간의 차이가 점점 벌어지고 있는 것으로 해석됩니다.



국내 면세점의 매출은 해외 의존도가 높은 편입니다. 해외 입국자 수 증가에 맞춰 면세점 판매지수도 19년 내내 증가 추세였지만 포스트 코로나 시대가 열리면서 급격한 하락세를 보입니다. 19년도에는 월 해외 입국자 수가 최대 160만 명 정도였지만 코로나 이후에는 10만 명 정도로 대폭 감소하였고 면세점의 매출 역시 비슷한 추이로 감소하였습니다. 다만 면세점은 20년 6월부터 장기 재고 상품 판매, 내국인 전용 제주국제공항 내 면세점 매출 증가 등으로 조금 회복한 반면 해외 입국자 수는 여전히 회복할 기미가 보이지 않습니다..

## 6. 코로나 시대의 업종별 추이

코로나 여파가 각 업종 별로 어떤 영향을 주었는지 알아보기 위해 업종별 월별 매출실적 데이터(18년 1월~20년 10월)를 분석하였습니다. 고려한 업종은 총 14개로 다음과 같습니다.

식료품, 섬유, 의복모피, 의료물질·의약품, 전자·영상·통신장비, 자동차, 가구, 건설업, 숙박업, 정보통신업, 부동산업, 사업시설·사업지원·임대업, 예술·스포츠·여가, 서비스업

코로나로 인해 어떤 업종들이 타격을 크게 받았고 어떤 업종들이 빨리 회복하였는지 분류하기 위해 군집분석(clustering)을 사용하였습니다.

### 6.1 군집화 방법

우선 계절성을 없애기 위해 업종별로 1년 전 동월 값으로 차분한 값을 이용하였습니다. 그 후 노이즈(noise)를 제거하기 위해 4개월 이동평균으로 추세를 계산하여 군집화를 진행하였습니다.

군집화 알고리즘은 K-means 방법을 기반으로 한 시계열 군집화를 사용하였습니다. 시계열 데이터 간의 유사성을 계산하기 위해 시계열 데이터들 간의 거리를 이용하였습니다. 일반적으로 시계열 데이터 사이의 거리는 각 시점 값의 차이의 제곱 합인 유클리디언 거리(Euclidean Distance, ED)를 사용합니다. 하지만 이 방법은 그래프끼리 교차되더라도 같은 군집으로 묶인다는 점과 동일한 패턴을 보여도 속도나 시작점이 다르면 다르게 인식한다는 점에서 저희 데이터에는 적합하지 않았습니다.

따라서 단순히 유클리디언 거리를 계산하기보단 DTW(Dynamic Time Warping) 통계량을 사용하는 것이 맞다고 판단하였습니다. DTW 통계량은 거리가 가장 짧은 시점(유사도가 높은 지점)을 찾아 계산하므로 ED(Euclidean Distance) 방법의 단점을 보완합니다.



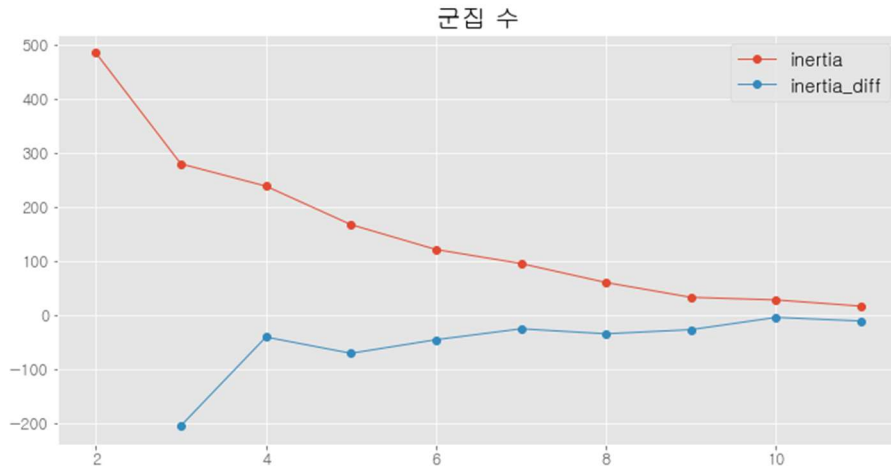
**Fig. 1.** Note that, while the two time series have an overall similar shape, they are not aligned in the time axis. Euclidean distance, which assumes the  $i^{\text{th}}$  point in one sequence is aligned with the  $i^{\text{th}}$  point in the other, will produce a pessimistic dissimilarity measure. The nonlinear dynamic time warped alignment allows a more intuitive distance measure to be calculated

이렇게 구한 DTW 통계량으로 K-means 클러스터링을 진행합니다. 코로나 전후의 군집을 비교하기 위해 20년 1월을 기준으로 데이터를 분리시켜 군집화를 진행하였습니다.

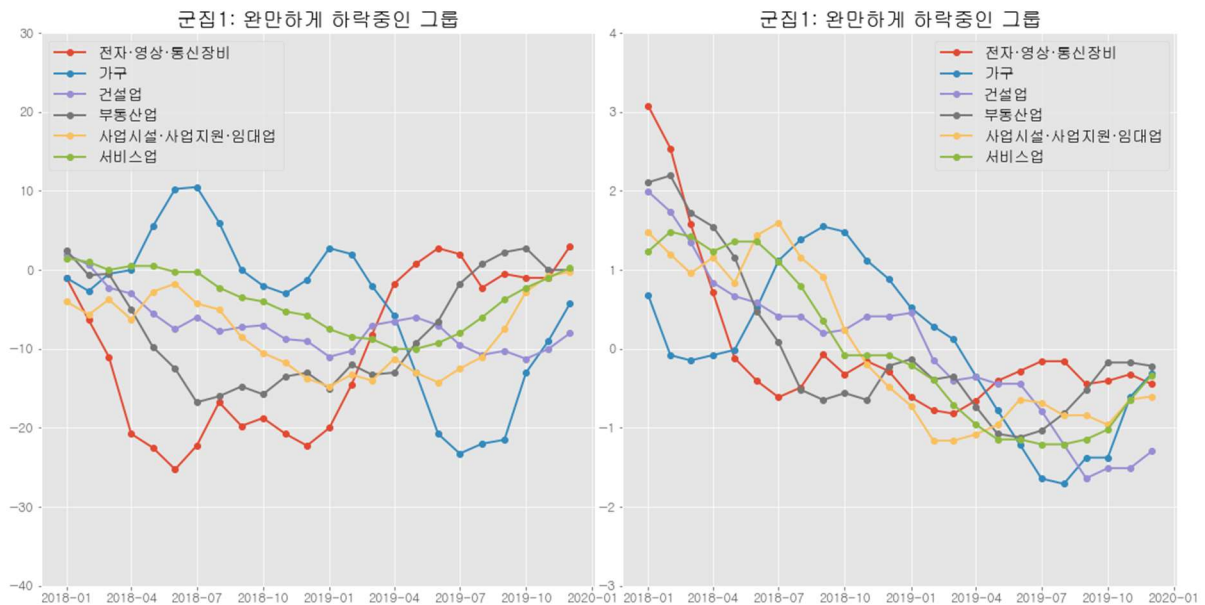


## 6.2 코로나 이전의 업종별 군집화 결과

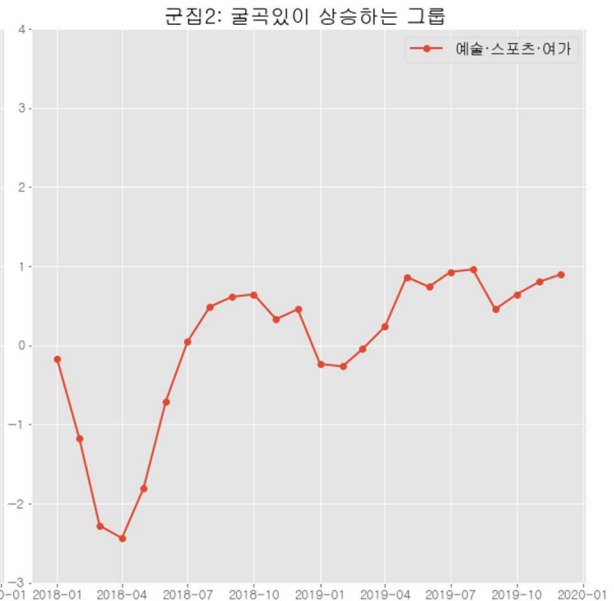
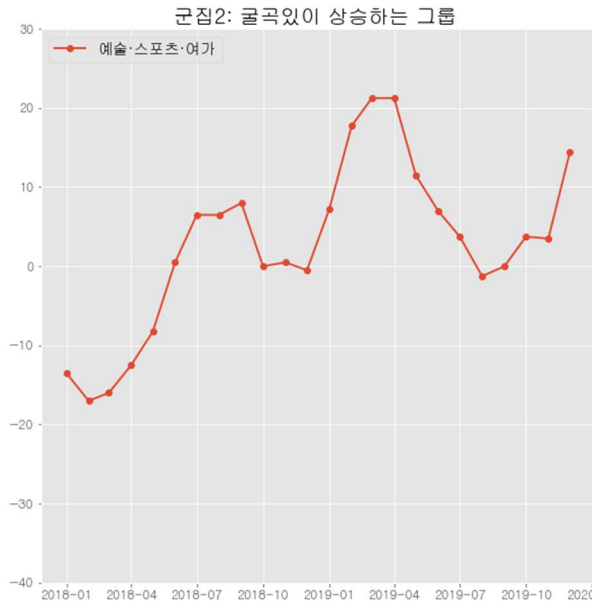
먼저 코로나 이전(18년 1월~19년 12월)의 업종별 군집화 결과입니다. 적절한 군집 수를 구하기 위해 Inertia 값을 사용하였습니다. 그 결과 elbow point인 6개로 군집을 나누는 것이 적절하다고 판단하였습니다.



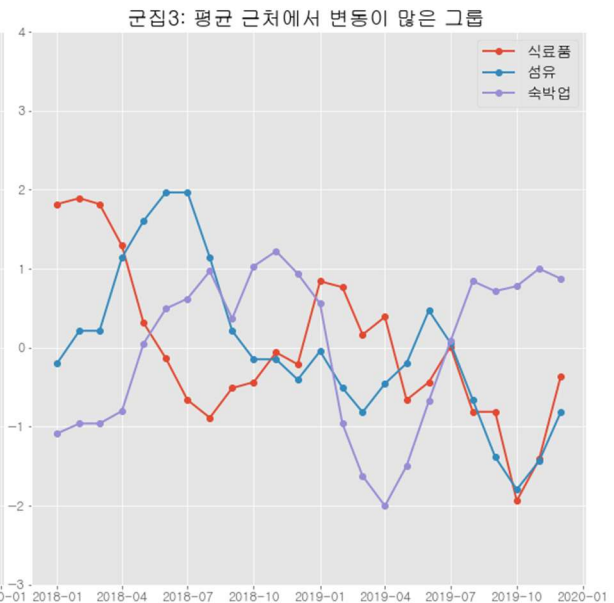
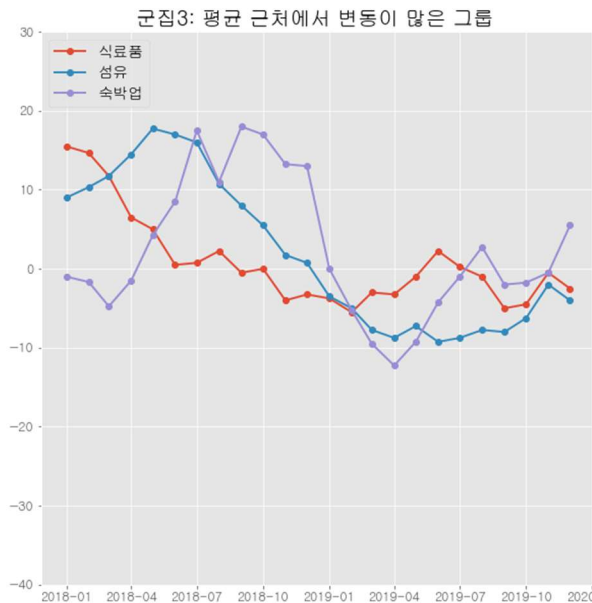
다음은 군집 분석 결과입니다. 왼쪽의 그래프는 전년 동월 값을 차분한 후 이동평균을 나타낸 데이터이고 오른쪽 그래프는 차분을 하지 않고 이동평균만 계산하여 표준화한 데이터입니다.



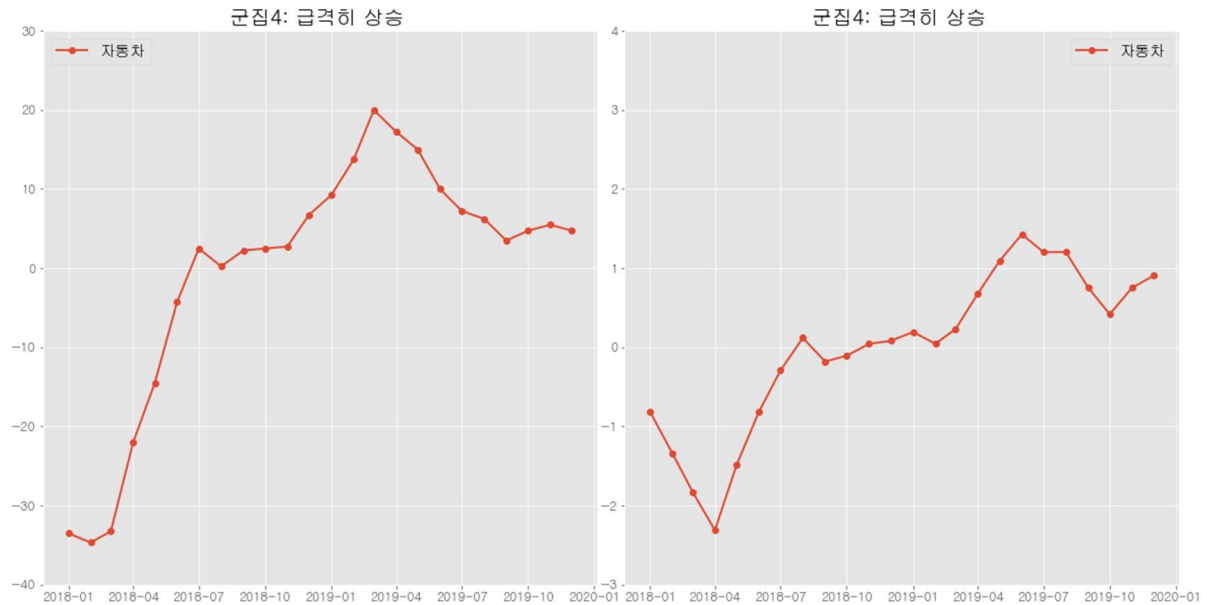
첫 번째 군집은 판매 실적이 지속적으로 완만하게 하락하는 업종들입니다. 전자·영상·통신장비, 가구, 건설업, 부동산업, 사업시설·사업지원·임대업, 서비스업이 해당합니다. 가장 많은 업종들이 포함된 군집으로 대한민국 전반적인 경기 흐름을 따라가는 업종들이 포함되어 있습니다.



두 번째 군집은 굴곡은 있지만 꾸준히 상승하는 군집입니다. 예술·스포츠·여가 업종이 여기에 해당됩니다. 시간이 지날수록 사람들이 문화 생활, 운동 등에 관심이 많아지는 것을 확인할 수 있었습니다.



세 번째 군집은 평균 근처에서 변동이 큰 업종들입니다. 식료품, 섬유, 숙박업이 여기에 속합니다. 이들은 시기에 따라 변동이 많지만 전반적으로 평균 근처에 머무르고 있습니다.



네 번째 군집은 2018년 초에는 불황이었지만 급격한 상승을 이룬 업종입니다. 자동차 산업이 여기 해당됩니다. 신차의 출시, 2019년 유가의 하락 등이 영향을 미친 것으로 판단됩니다.



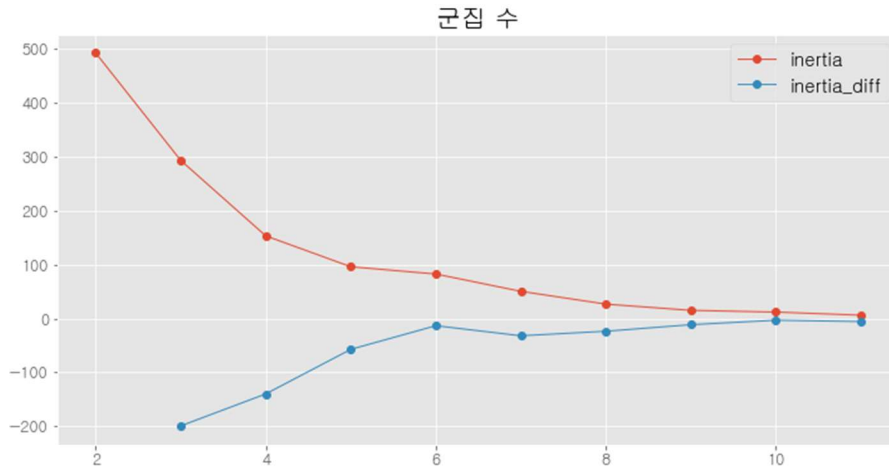
다섯 번째 군집은 2018년까지는 상승했다가 2019년 들어 급격히 하락한 군집으로 의복모피가 해당됩니다. 의복 시장은 작년 하반기에 겨울의 이상 고온 현상으로 겨울 상품 판매 실적이 부진하였습니다. 또한 불매운동으로 인해 유니클로, 무인양품 등 일본 의류 브랜드에 대한 소비가 감소하여 전체적인 판매 지수가 떨어진 것으로 보입니다.



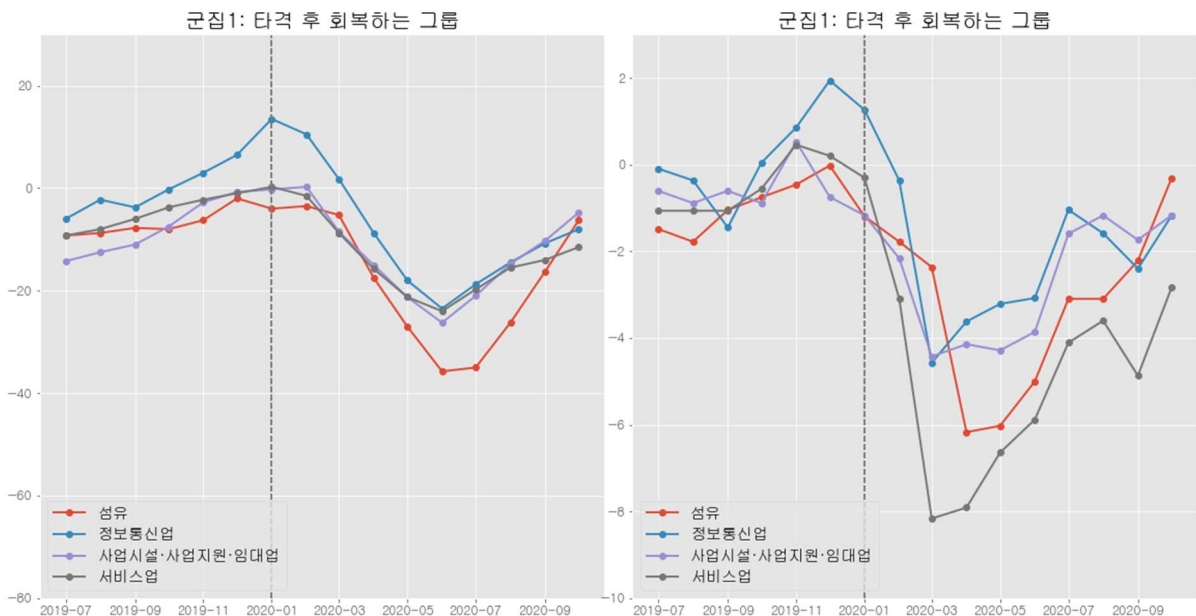
여섯 번째 군집은 2018년에 하락세를 보이다가 2019년에 상승한 업종들입니다. 의료물질·의약품, 정보통신업이 해당합니다. 의료물질·의약품은 첨단기술과 헬스케어의 융합이 현실화되고 세포 및 유전자 치료제가 시장에 출시되어 상승세를 보인 것으로 판단됩니다. 정보통신업은 ICT 기술의 발전과 모바일 시장의 확대, 특히 유튜브, 넷플릭스, 왓챠 등 콘텐츠 소비의 급격한 증가가 영향을 준 것으로 보입니다.

### 6.3 코로나 이후의 업종별 군집화 결과

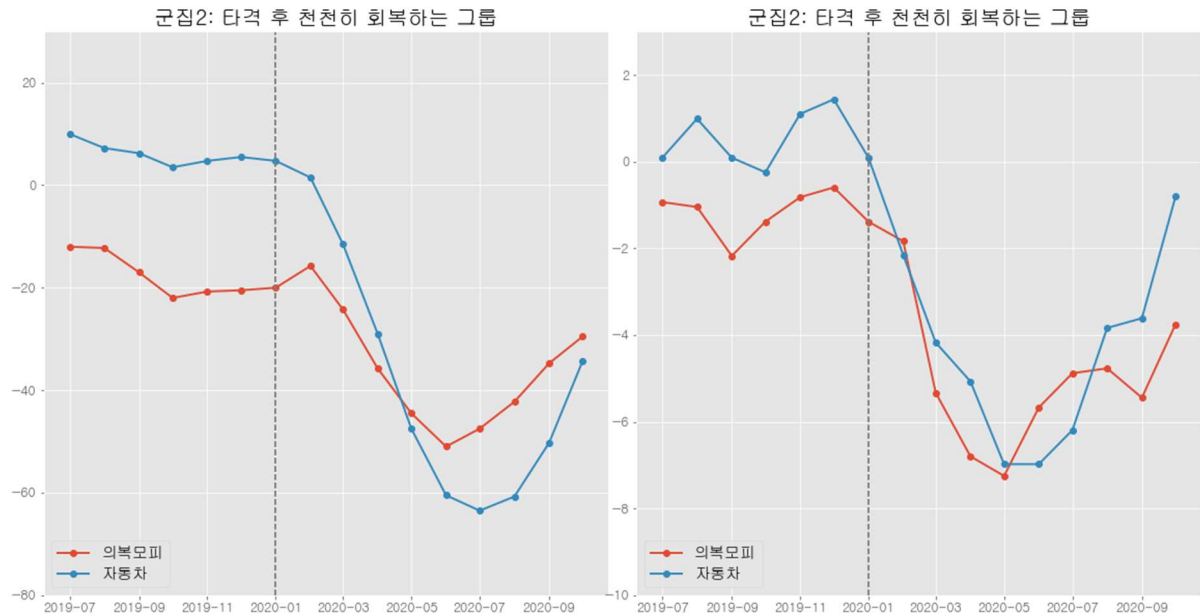
코로나가 각 업종에 얼마나 영향을 미쳤는지 보기 위해 20년 1월~20년 10월 데이터를 이용하여 군집화를 진행하였습니다. Inertia 값의 elbow point를 확인한 결과 5개의 군집으로 나누는 것이 적절하다고 판단하였습니다.



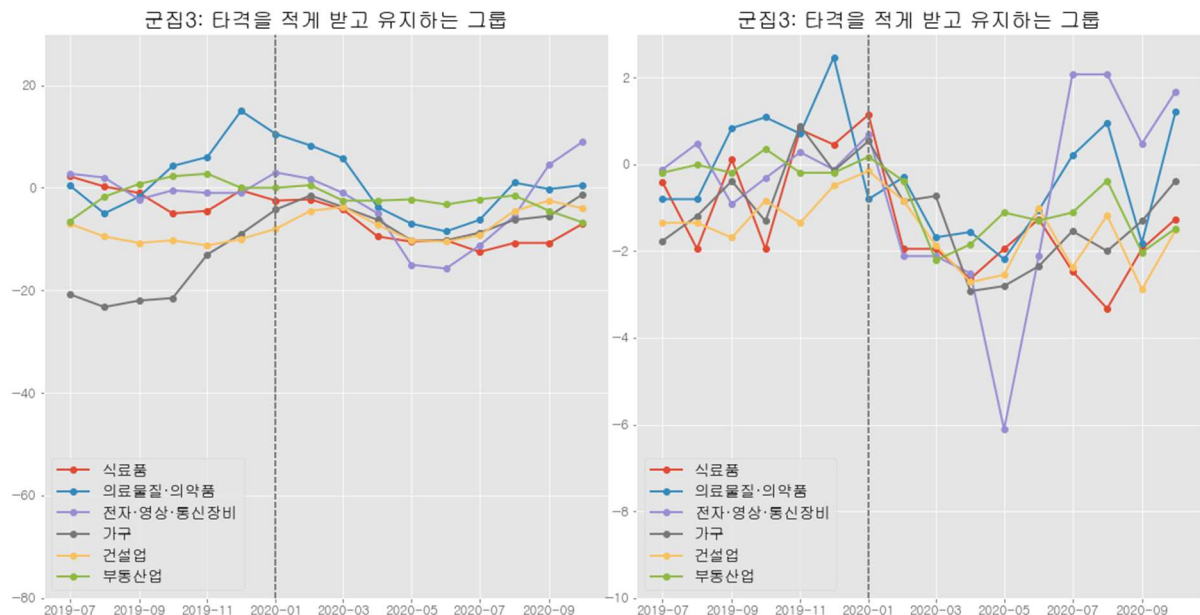
다음은 군집 분석 결과입니다. 왼쪽의 그래프는 전년 동월 값을 차분한 후 이동평균을 나타낸 데이터이고 오른쪽 그래프는 차분을 하지 않고 이동평균만 계산하여 표준화한 데이터입니다. 회색 선 이전은 19년 7월부터 19년 12월 데이터이고 회색 선 이후는 20년 1월부터 20년 10월 데이터입니다.



첫 번째 군집은 코로나로 인해 타격을 받았지만 회복에 성공한 군집입니다. 섬유, 정보통신업, 사업시설·사업지원·임대업, 서비스업이 해당됩니다. 이들은 일상 생활에 필수적인 산업들이라 무난히 회복에 성공했다고 추정됩니다.



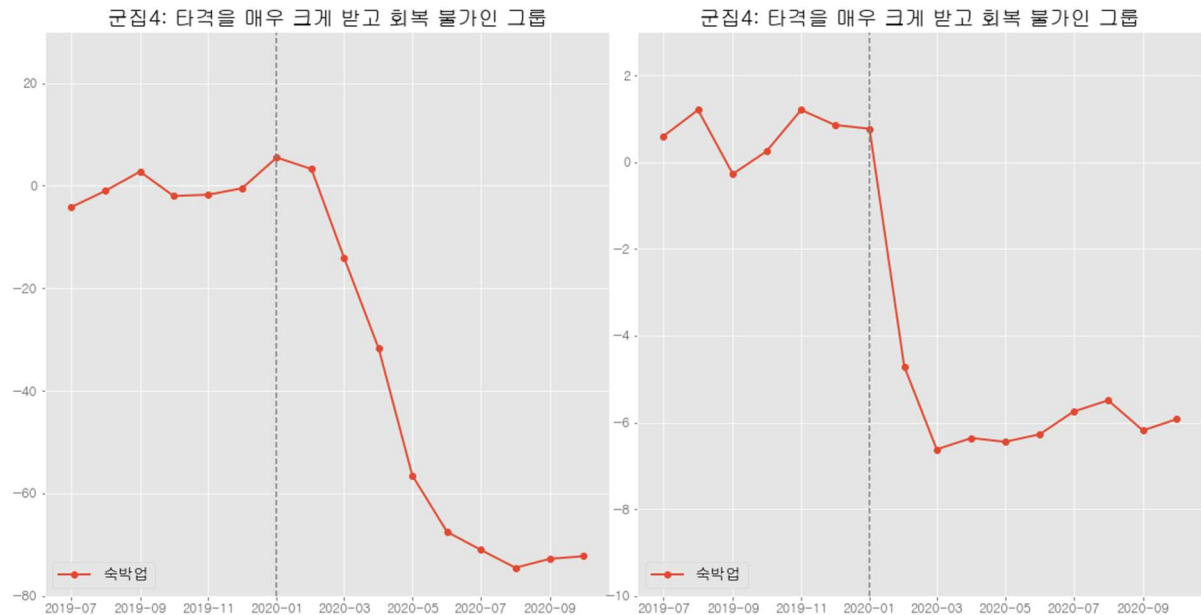
두 번째 군집은 타격 후 천천히 회복하는 그룹입니다. 의복모피, 자동차가 해당됩니다. 의복모피의 경우 코로나로 인한 외출 자제 현상으로 매출이 하락하였습니다. 하지만 시간이 지남에 따라 유동인구 수가 늘어나면서 매출 실적이 올랐습니다. 자동차 역시 코로나 직후에는 판매량이 줄었으나 유동인구 수 증가, 자차 선호 현상 등으로 회복 중에 있습니다.



세 번째 군집은 타격을 적게 받고 유지하는 그룹입니다. 식료품, 의료물질·의약품, 전자·영상·통신장비, 가구, 건설업, 부동산업이 해당됩니다. 식료품과 가구는 사람들이 집에 있는 시간이 증가하여 매출에 큰 영향을 받지 않은 것으로 보입니다. 건설업, 부동산업은 실물 자산에 관련된 업종들로 다른 업종들보다 코로나로 인한 경제 변동에 상대적으로



덜 민감한 것으로 추정됩니다. 의료물질·의약품은 다른 업종들과 달리 코로나 초기에 판매실적이 증가하였습니다. 건강에 대한 관심이 늘어났기 때문입니다. 전자·영상·통신장비는 코로나보다는 신제품의 출시가 더 큰 영향을 주는 것으로 보입니다.



네 번째 군집은 타격을 매우 크게 받고 회복 불가인 군집입니다. 숙박업이 해당됩니다. 유동인구의 감소, 해외 관광객의 급감 등으로 인해 다른 업종들과 달리 아직까지도 회복하지 못하고 있습니다. 코로나가 완전히 종식되기 전까지는 회복하기 힘들 것으로 판단되는 업종입니다. 하지만 코로나가 끝나면 가장 빠르게 매출이 오를 것으로 예상됩니다.



마지막 군집은 타격을 크게 받고 회복하려고 했으니 실패한 그룹입니다. 예술·스포츠·여가 산업이 해당됩니다. 2-3월 유동인구의 감소로 큰 타격을 받은 이후 조금 회복하는 듯 보였으나 2차 유행으로 인해 다시 감소 추세입니다. 숙박업보다는 상황이 낫지만 코로나 이전으로 회복하기에는 많은 시간이 필요한 것으로 보입니다.

## 7. 결론

코로나의 영향을 소비자의 시각과 산업별 추이를 통해 분석해 보았습니다. 그 결과 코로나는 전반적으로 경제에 타격을 주어 경제 활동을 위축시켰고, 소비자와 산업에 큰 변화를 야기하여 코로나 발생 이전과는 다른 모습을 보여주었습니다. 저희 팀의 분석 내용은 아래와 같습니다.

1. 소비자들의 소비 심리는 큰 폭으로 감소한 뒤 아직 회복되지 않았습니다.
2. 임금이 낮은 집단이 임금이 높은 집단에 비해 여러 지표에서 더 큰 감소를 보였습니다.
3. 전반적으로 30~60대의 소비패턴은 유사한 양상을 보여주었으나 20대의 소비 패턴은 다른 연령대와 차이를 보였습니다.
4. 온라인 시장이 매출 감소의 돌파구가 될 수 있음을 보았습니다.
5. 야외활동이나 여행에 관련된 업종들이 심각한 타격을 입은 뒤 회복하지 못하는 모습을 보여주고 있습니다.
6. 실생활에 관련된 산업은 덜 타격 받았을 뿐 아니라 안정적인 모습을 보입니다.

※ 데이터 출처

1) 보건복지부 코로나19 감염 현황

출처 : 공공데이터포털

링크 : <https://www.data.go.kr/tcs/dss/selectApiDataDetailView.do?publicDataPk=15043376>

2) 품목별 소비지수 데이터

출처 : 데이콘 포스트 코로나 데이터 시각화 경진대회

링크 : <https://www.dacon.io/competitions/official/235618/data/>

3) 월별 소비자 동향조사

출처 : KOSIS 국가통계포털

링크 : [https://kosis.kr/statHtml/statHtml.do?orgId=301&tblId=DT\\_040Y002&checkFlag=N](https://kosis.kr/statHtml/statHtml.do?orgId=301&tblId=DT_040Y002&checkFlag=N)

4) 소매업태별 월별, 분기별 판매액 계절조정지수

출처 : KOSIS 국가통계포털

링크 : [https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\\_1K31013](https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1K31013)

5) 온오프라인 유통업체 식품군 월간 전년동기대비 성장률

출처 : 공공데이터포털

링크 : <https://www.data.go.kr/data/15061361/fileData.do>

6) 기간별 해외 입국자 수

출처 : 관광지식정보시스템

링크 : <https://know.tour.go.kr/stat/exitEntryTourStatDis.do?enterCode=1>

7) 업종별 매출 실적

출처 : ECOS 한국은행 경제통계시스템

링크 : <https://ecos.bok.or.kr/>

---

<sup>i</sup> 사진 출처: <http://practicalquant.blogspot.com/2012/10/mining-time-series-with-trillions-of.html>