

포스트 코로나 데이터 시각화 경진대회



다각적 통계 분석을 통한 유망산업 발굴

:재정의된 '포스트 코로나'를 중심으로

Team 피너트리

김상태 박서영 박정현



Table of Contents



I. 서론

1. 문제 정의
2. COVID-19 기본 분석
3. COVID-19 시기 재정의

II. 본론

1. Delivery data로 보는 유망 산업
2. Index data로 보는 유망 산업
3. Card data로 보는 유망 산업

III. 결론

Table of Contents



I. 서론

1. 문제 정의
2. COVID-19 기본 분석
 - 실확진자 추이 분석
 - 누적확진자 분석
 - 유동인구 추이 분석
3. COVID-19 시기 재정의
 - 코로나 시기 시각화 및 정리

II. 본론

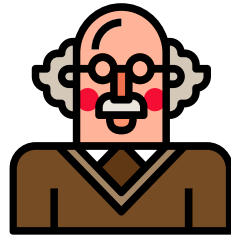
III. 결론



문제 정의

포스트 코로나에 대한 기존 논의가 불명확하다는 문제점을 짚고 방향, 시기, 형태 등을 재정립하여 분석 목적을 명확히 함

'포스트 코로나' 를 대표하는 키워드?

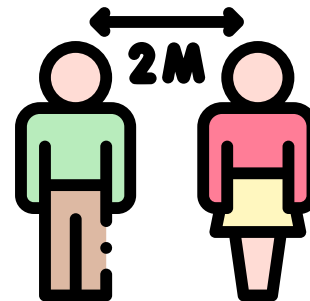


Thomas Frey, an American futurist and celebrity speaker

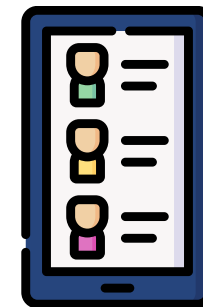
"포스트 코로나는 한 단어로 '**리부트(Reboot)**',
현재는 재시동을 앞둔 '**일시 중지(pause)**' 단계다."



세계화의 종말
(The End of
Globalization)



언택트
(Untact)



디지털 콘택트
(Digital contact)

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의

문제 정의

포스트 코로나에 대한 기존 논의가 불명확하다는 문제점을 짚고 방향, 시기, 형태 등을 재정립하여 분석 목적을 명확히 함

'포스트 코로나'에 대한 기존의 논의

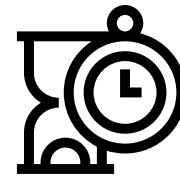
방향성과 시기, 형태 등이 구체적으로 정의되지 않은 아노미 상태

→ 포스트 코로나에 대한 **표준 재정의** 필요!

" *POST-COVID19* "



IN WHAT
DIRECTION?



WHEN?



HOW?

UNCLEAR

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



문제 정의

포스트 코로나에 대한 기존 논의가 불명확하다는 문제점을 짚고 방향, 시기, 형태 등을 재정립하여 분석 목적을 명확히 함

분석 배경과 목적

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



IN WHAT DIRECTION?

기존에 불분명했던 코로나바이러스 등장 이후의 시기를
'가-승-전-결'의 형태로 재정의



WHEN?

재정의한 시기를 바탕으로 **'포스트 코로나'**를 정의



HOW?

대중의 **소비 패턴** 변화 파악, **유망 산업** 발굴



문제 정의

본 조사에서 사용한 기본 데이터와 추가적으로 사용한 데이터를 밝힘

활용 데이터 설명



유동인구 데이터
(fpopl)



업종 별 결제 금액 데이터
(card)



배달 정보 데이터
(delivery)



품목별 매출 성장 지수 데이터
(index)



COVID-19 관련 데이터
(Time, TimeProvince,
Timeage, Timegender)



아모레퍼시픽 산업보고서



국토정보지리원 연속수치지형도
행정경계 데이터

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



실확진자 추이 분석

'Time.csv' 에서 실확진자 수를 구하기 위해 데이터를 전처리함

전처리한 데이터

date	time	test	negative	confirmed (누적확진자수)	released	deceaesd
2020-01-20	16	1	0	1	0	1
			.			
			.			
			.			
2020-06-30	0	1273766	1240157	12800	11537	282

date	실확진자수
2020-01-20	1
.	.
2020-06-30	43

"기준 누적 확진자 수 - 전날 누적 확진자 수"

< 0일 경우는 0명으로

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



실확진자 추이 분석

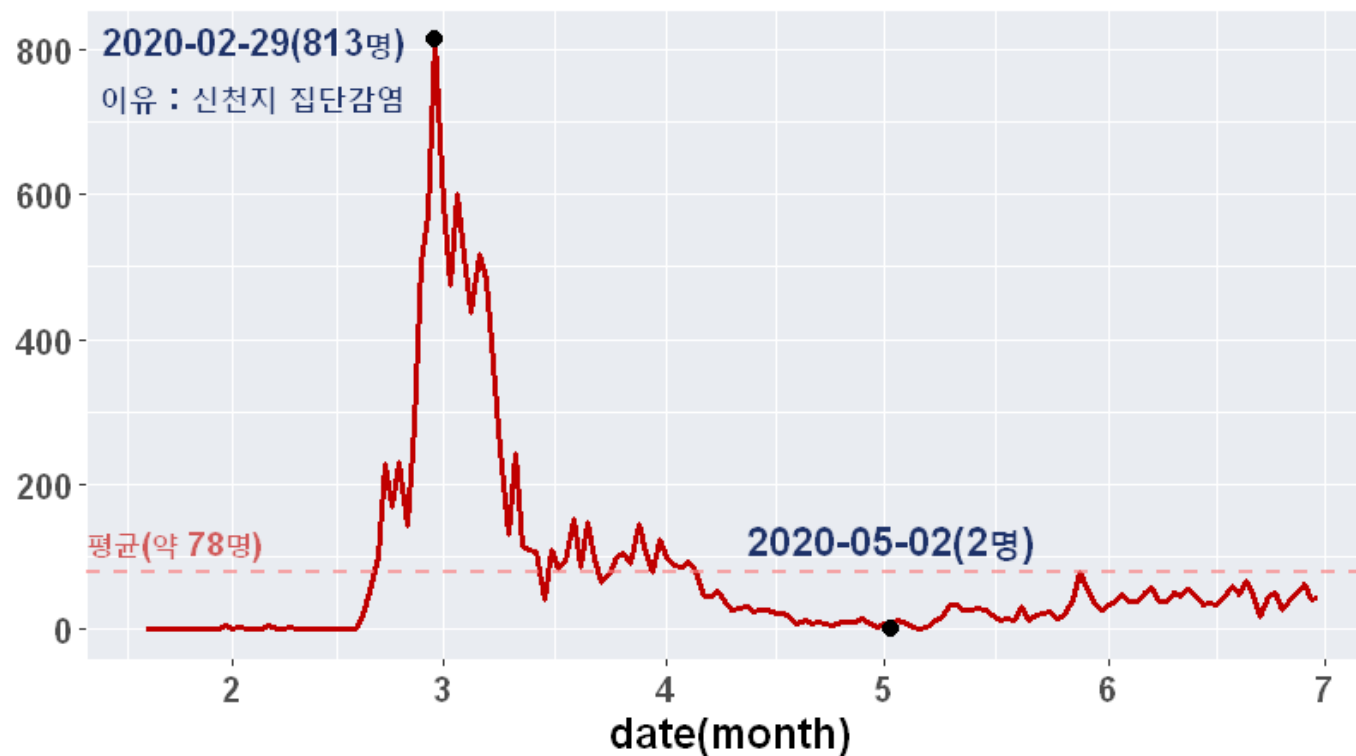
신천지 집단감염 사례 이후로 줄어가는 추세이지만, 종종 집단감염 및 해외유입으로 인해 확진자 수가 늘어나기도 하는 상황

시각화 결과

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의





누적 확진자 분석

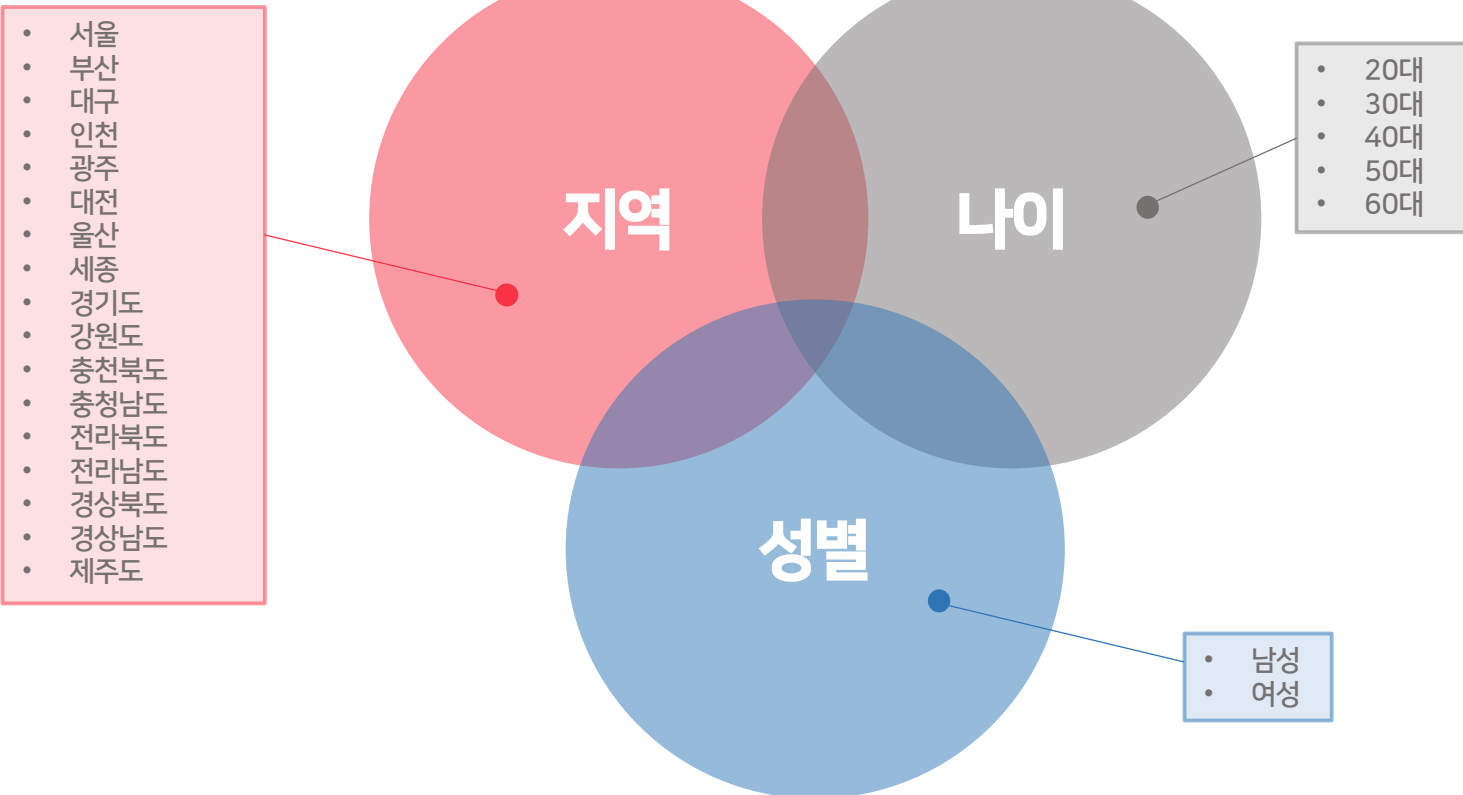
지역 · 나이대 · 성별 누적확진자 분석을 통해 국내 COVID-19 확진자들의 특징을 파악함

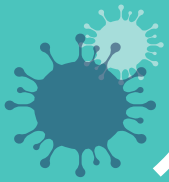
COVID- 19 국내 확진자

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의





서론

누적 확진자 분석:지역별

'TimeProvince.csv' 의 2020-06-30일 기준 지역 별 누적확진자 수 데이터를 지도 데이터의 해당하는 지역에 병합함
(국토정보지리원 연속수치지형도 행정경계 데이터 이용)

전처리한 데이터

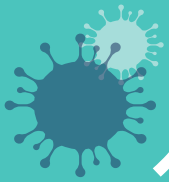


province	confirmed
서울	1312
부산	154
대구	6906
인천	341
광주	44
대전	117
울산	55
세종	50
경기도	1207
강원도	65
충청북도	65
충청남도	167
전라북도	27
전라남도	24
경상북도	1389
경상남도	134
제주도	19

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의

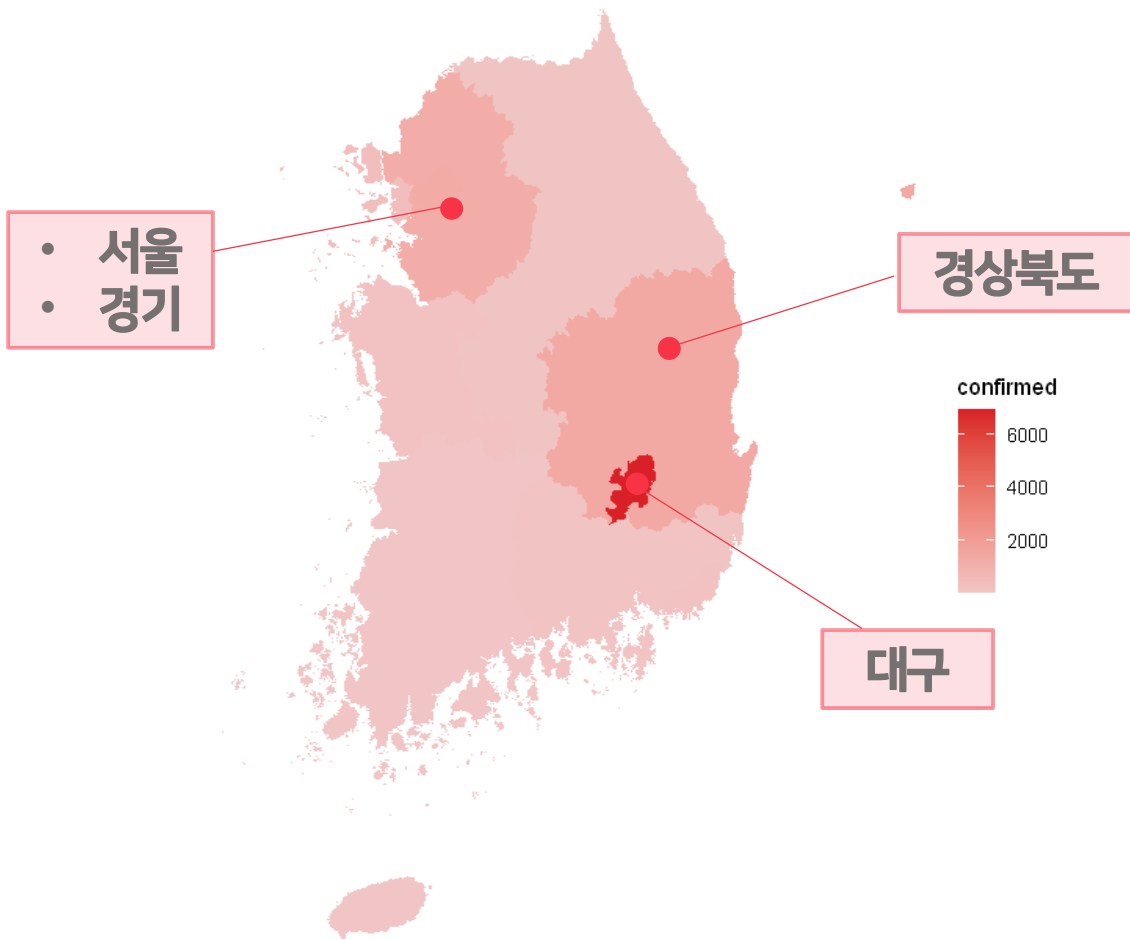


서론

누적 확진자 분석: 지역별

심각도에 따라 진한 색으로 나타낸 결과, 서울 · 경기 · 경북 · 대구가 심각 지역인 것으로 보임

시각화 결과



문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



누적 확진자 분석: 나이대 & 성별

'TimeAge.csv' 와 'TimeGender.csv'의 2020-06-30일 기준 나이대별·성별 누적확진자 수를 추림

전처리한 데이터

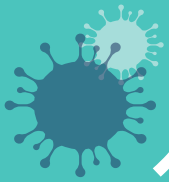
date	time	age	confirmed	deceased
2020-06-30	0	20s	3362	0
.	.	30s	1496	2
.	.	40s	1681	3
.	.	50s	2286	15
2020-06-30	0	60s	1668	41
date	time	sex	confirmed	deceased
2020-06-30	0	male	5495	151
2020-06-30	0	female	7305	131

“ 나이대는 *index.csv*와 동일하게
20대에서 60대만 확인 ”

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



서론

누적 확진자 분석: 나이대 & 성별

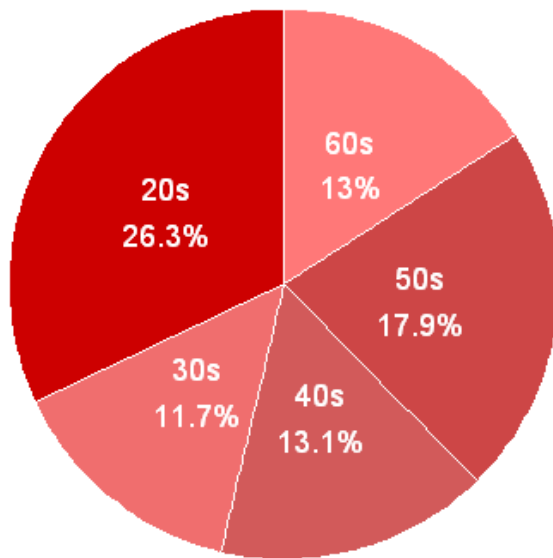
나이대별은 20, 50, 40, 60, 30 대 순으로 많음 / 성별은 확진자 비율은 비슷하나, 여성이 좀 더 많음

시각화 결과

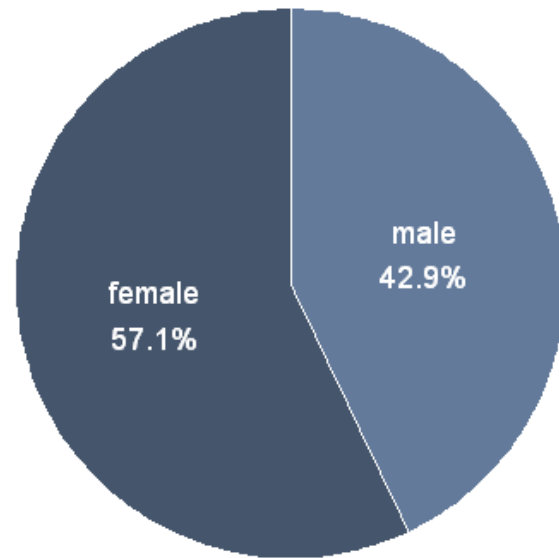
문제 정의

COVID-19 기본 분석

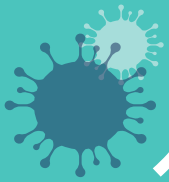
COVID-19 시기 재정의



[Age]



[Gender]



서론

유동인구 추이 분석

COVID-19 확산과 국민의 오프라인 활동 간의 관계를 보기 위해, 'fpopl.csv'를 이용해 주별 유동인구 추이와 실확진자 추이를 비교함

전처리한 데이터

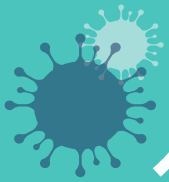
week	주별 실확진자 평균	주별 유동인구 평균
3	0.50	204.02
4	0.43	178.47
	.	
	.	
23	44.43	201.10
24	43.29	200.08

“ 일 단위를 주 단위로 바꿔
두 데이터의 **공통 주**를 확인 ”

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의

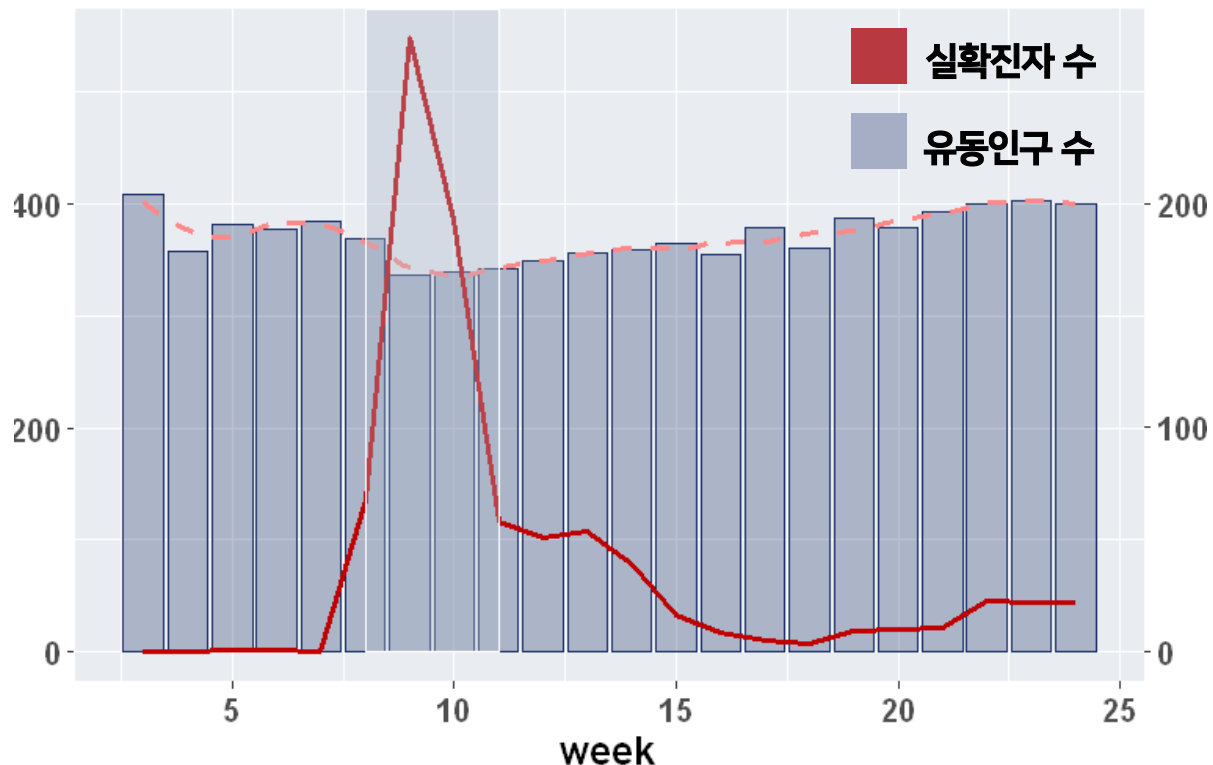


서론

유동인구 추이 분석

실확진자 수가 가장 많았던 시기에 유동인구가 줄었고, 그 이후에는 유동인구가 점점 증가함.
즉, 오프라인 활동과 코로나 확진자 수는 상관관계가 있음

시각화 결과

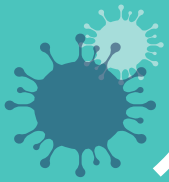


“확진자가 상대적으로 감소한 현재는 다소 **경각심이 해이해진** 경향을 보임”

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의



서론

COVID-19 시기 재정의

COVID-19 출현 이후의 시기를 관련 정책과 데이터를 바탕으로 재정의함

COVID-19의 기-승-전-결

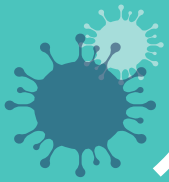


문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의

코로나 시기 정의	설명	
기 (01/04-02/22)	특징	• 코로나바이러스 출현
	정책 및 데이터	• 감염병 위기경보 단계 ' 경계 '
승 (02/23-03/08)	특징	• 신천지 사태 로 확진자 수 급증
	정책 및 데이터	• 감염병 위기경보 단계 ' 심각 ' 단계로 격상 (02/23)
전-1 (03/09-05/06)	특징	• 유동인구 추이 다시 상승
	정책 및 데이터	• 마스크 5부제 시행(03/09)
전-2 (05/07~)	특징	• 대중의 경각심이 해이해지기 시작
	정책 및 데이터	• 신천지 사태 후 최저 확진자 (05/06,2명)



서론

COVID-19 시기 재정의

COVID-19 출현 이후의 시기를 관련 정책과 데이터를 바탕으로 재정의함

문제 정의

COVID-19 기본 분석

COVID-19 시기 재정의

COVID-19의 기-승-전-결

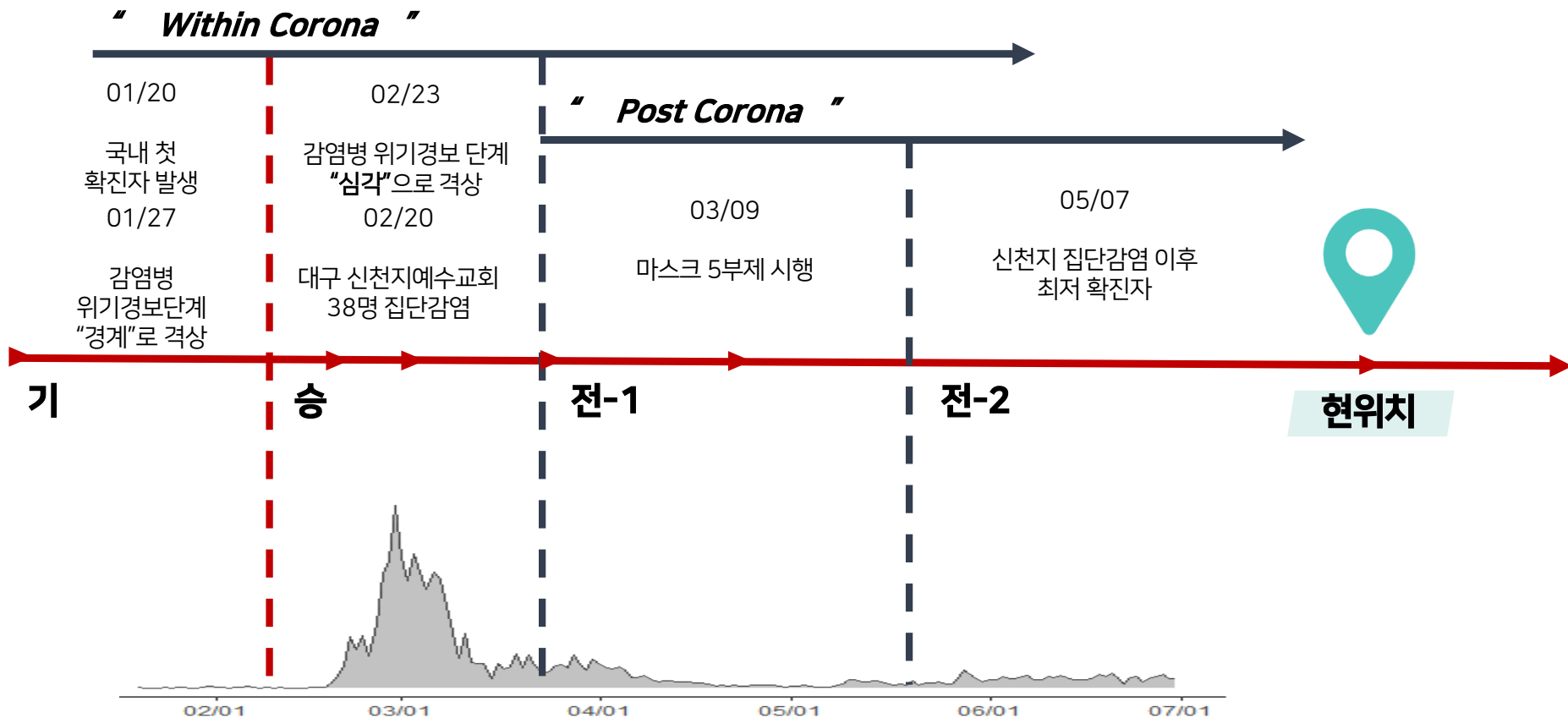


Table of Contents



I. 서론

II. 본론

1. Delivery data로 보는 유망 산업
2. Index data로 보는 유망 산업
 - Factor Analysis
3. Card data로 보는 유망 산업
 - Clustering(K-means)
 - 시계열 Factor Analysis

III. 결론

Delivery Data : 배달 호출 정보

Delivery 데이터의 특징을 살펴보고, 그에 따라 중점적으로 살펴볼 세가지 Key Point를 정리

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

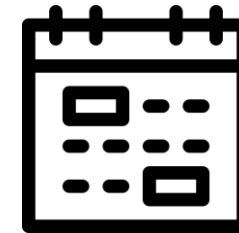
Delivery Data 특징



배달호출 정보 데이터



전국구 데이터



일별 데이터
2020.01.01~2020.06.21

Key Point

- 코로나로 인한 배달 업종의 매출 및 배달 건수는 어떻게 변했을까?
- 지역별 코로나 심각정도에 따라 배달 호출 건수와 배달 품목이 달라질까?
- 코로나에 가장 큰 영향을 받은 업종과 그 특징은 무엇일까?



Delivery Data : 배달 호출 정보

코로나 여파로 일별 배달 매출 및 건수는 꾸준히 증가하는 추세를 보임

각 데이터로 보는 유망 산업

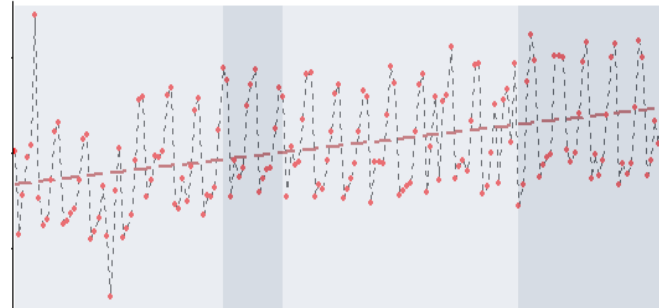
Delivery data

Index data

Card data

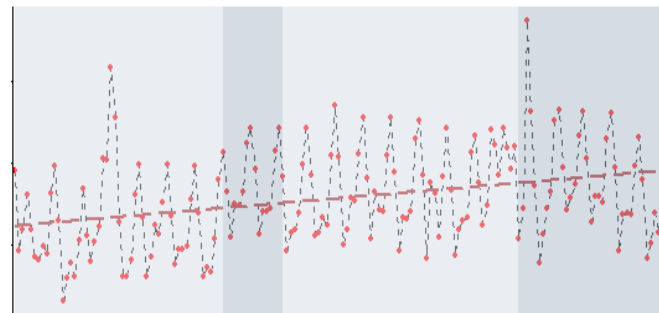
일별 매출 및 건수 변화

일별 평균 배달 건수



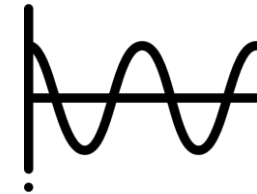
일자

일별 평균 배달 매출



일자

- 기
- 승
- 전-1
- 전-2



☑ 7일간격으로 주기성이 보임



☑ 배달 건수 & 매출 모두 증가하는 추세



각 데이터로 보는 유망 산업

Delivery data

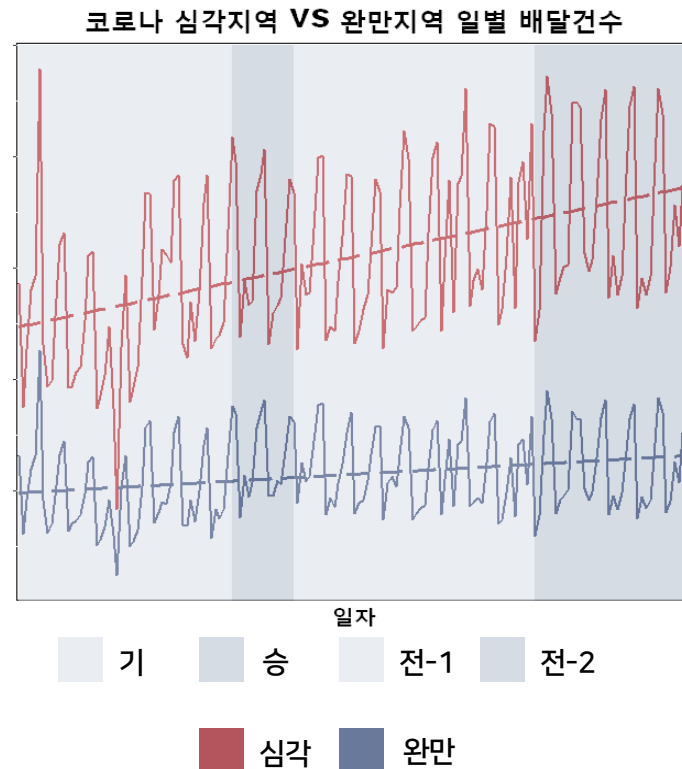
Index data

Card data

Delivery Data : 배달 호출 정보

코로나 심각 수준 지역(서론에서 상위 4개 지역)이 그 외 지역들보다 가파른 증가 추세를 보임

코로나 심각 수준에 따른 배달 증가 추세



코로나 심각 수준

심각 : 서울, 경기, 대구, 경북

완만 : 그 외 지역



코로나 심각 지역이 보다
가파르게 증가하는 추세를 보임

단순 양이 아닌 추세를 보는 이유 : 서울, 경기권이 배달 건수 상위지역이므로 증가 추세(기울기)를 봄

Delivery Data : 배달 호출 정보

다른 배달 품목들에 비해 심부름, 도시락 품목이 보다 가파른 매출 증가 추세를 보임

각 데이터로 보는 유망 산업

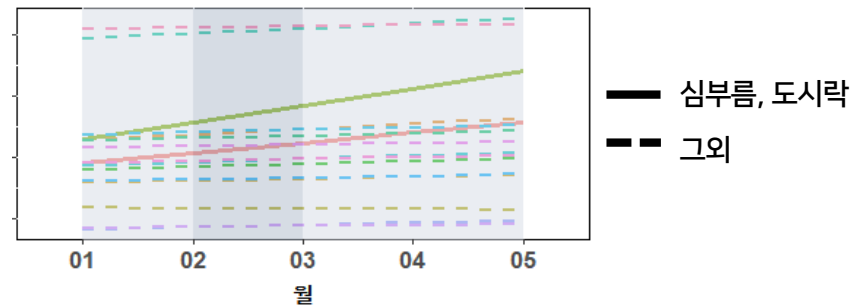
Delivery data

Index data

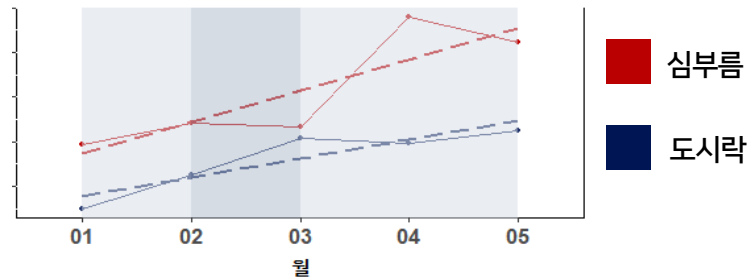
Card data

매출 증가 품목: 도시락&심부름

배달품목별 월별 매출 변화 추세



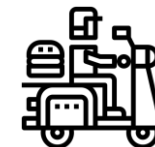
배달품목별 월별 매출 변화 추세



기 승 전-1 전-2

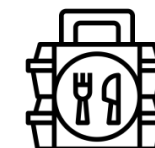


도시락, 심부름이 다른 품목들에 비해 급격하게 매출 증가



심부름(배달대행)

전-1에서 급격히 증가



도시락

기~승에서 급격히 증가

6월의 경우 21일까지 밖에 없으므로 온전히 있는 1~5월 데이터로 시각화

Delivery Data : 배달 호출 정보

도시락&심부름 품목이 그외 품목들에 비해 매출 성장은 급격하지만, 배달 건수는 오히려 일정하게 보임

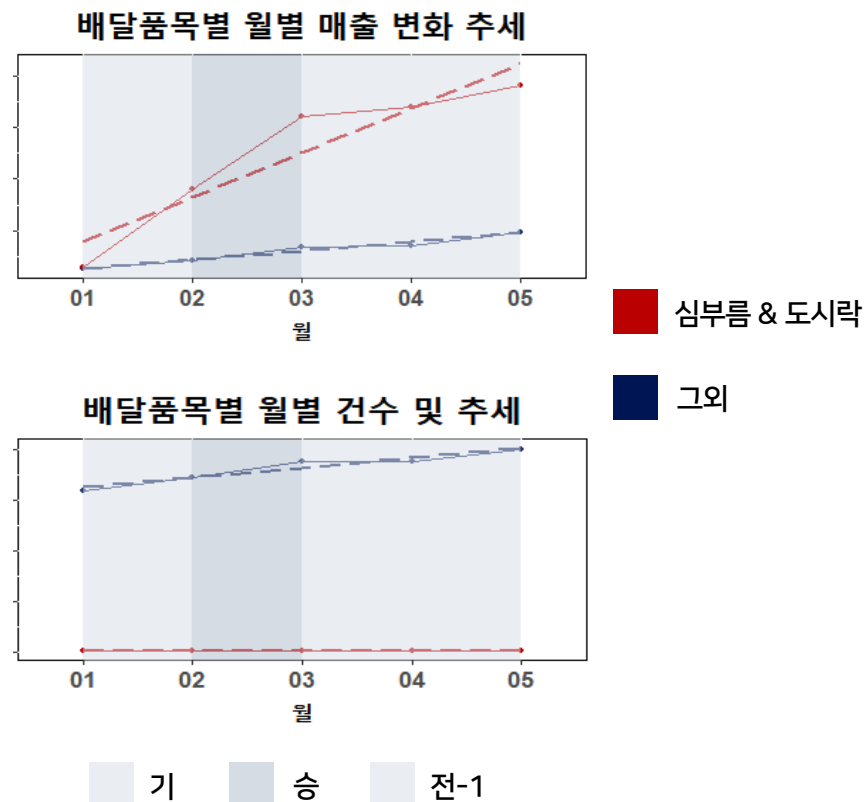
각 데이터로 보는 유망 산업

Delivery data

Index data

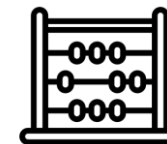
Card data

도시락&심부름 vs 그 외 품목 추세 비교



매출

심부름&도시락이
더 가파른 증가 추세를 보임



배달 건수

그외 품목 : 증가 추세
심부름&도시락 : 일정



Scale 문제로 보이므로
심부름&도시락만 배달 건수 시각화

Delivery Data : 배달 호출 정보

도시락&심부름 품목의 월별 배달 건수가 증가했다가 감소하는 오목함수(Concave)형태를 보임

각 데이터로 보는 유망 산업

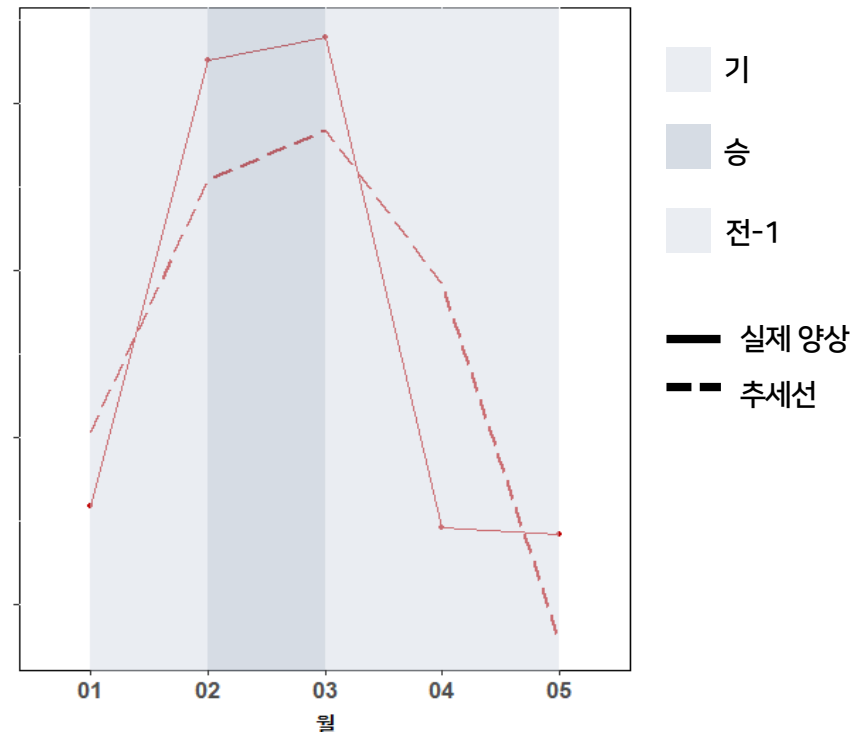
Delivery data

Index data

Card data

도시락&심부름 vs 그 외 품목 추세 비교

도시락,심부름 : 월별 건수 변화 및 추세



도시락&심부름 배달 건수 양상은 오목함수(Concave) 형태



도시락&심부름 배달 매출 양상은 가파른 증가 추세 형태



한번에 대량 구매 양상 유추 가능

Delivery Data : 배달 호출 정보

전국구 데이터인 Delivery 데이터를 통해 사람들의 소비 패턴이 본인이 속한 지역의 코로나 심각 정도에 따라 달라지는지 확인

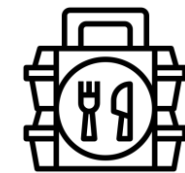
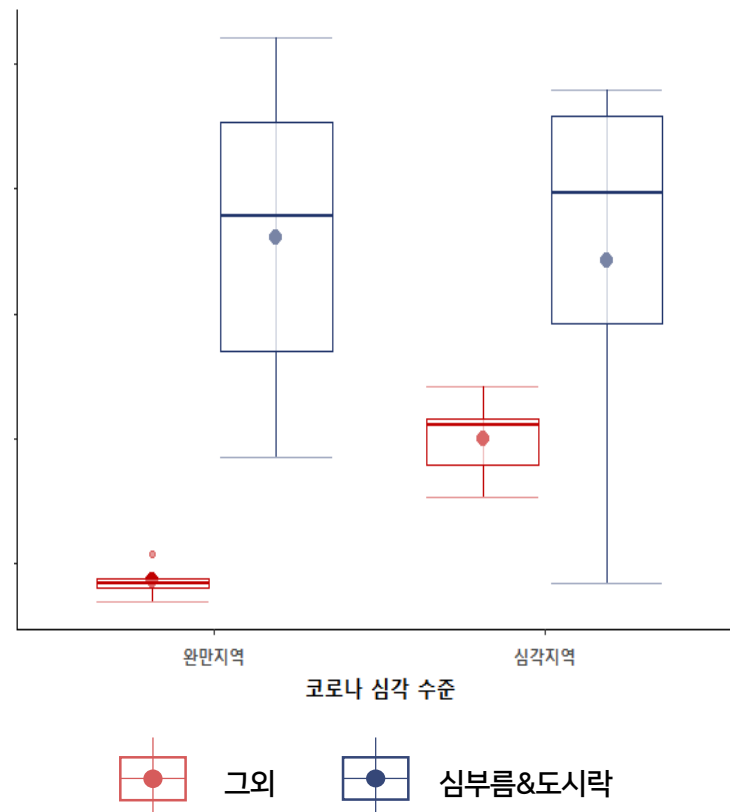
각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

코로나 심각도에 따른 소비 패턴



심부름&도시락

코로나 심각정도에 상관없이
비슷한 매출 양상



그외

코로나 심각 지역이 보다
큰 매출



코로나 심각정도에 상관없이
간편한 언택트 서비스(심부름&도시락)
이용 경향



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

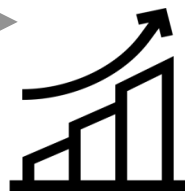
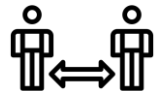
Delivery Data : 배달 호출 정보

배달 건수와 매출은 코로나 여파 이후 꾸준히 증가하는 추세를 보임.
코로나 심각 지역이 완만 지역에 비해 배달 건수 증가 추세가 보다 크게 나타남

배달 호출 정보 데이터 분석 요약



배달 건수와 매출 모두 코로나 여파 이후 꾸준히 증가하는 추세
→ 언택트(Untact) 소비 경향



코로나 심각지역이 완만지역에 비해 증가하는 추세가 더 가파름
→ 코로나의 영향으로 배달량이 더 빠르게 증가





각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

Delivery Data : 배달 호출 정보

도시락&심부름 품목이 다른 품목들에 비해 매출이 급증하는 경향을 보임.
반면에 배달 건수는 오목함수(Concave)형태로 한번의 배달 주문에 대량으로 구매하는 양상 유추 가능

배달 호출 정보 데이터 분석 요약



도시락&심부름 품목이 다른 품목들에 비해 매출이 급증,
but 배달 건수는 오목함수 형태



→ 코로나 사태 이후 직장인들의 점심 대용을 위한 대량 주문으로 유추 가능



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

Delivery Data : 배달 호출 정보

본인의 지역 코로나 심각 수준에 상관없이 코로나와 상관관계가 높은 품목(심부름&도시락)을 소비하는 양상 확인

배달 호출 정보 데이터 분석 요약



도시락&심부름 품목의 코로나 심각 여부에 따른 소비 양상 확인
:지역에 따른 차이가 크지 않았음

→ 본인 지역의 코로나 심각정도가 낮아도 **코로나에 대한 경각심** ↑





각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

Index Data : 품목별 매출 성장 지수

Index 데이터의 특징을 살펴보고, 그에 따라 중점적으로 살펴볼 두가지 Key Point를 정리

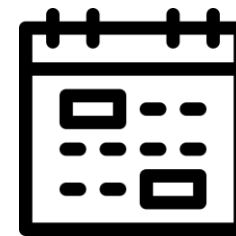
Index Data 특징



2018년 동월 대비 매출 성장 비율



나이, 성별, 지역 별 수치



월별 데이터
2019.01~2020.05

Key Point

- 어떤 소비자 유형을 타겟팅하면 좋을까? (Factor Analysis 사용)
- 유망 품목의 매출을 상승시키기 위한 구체적 솔루션은 무엇일까?

Index Data : 품목별 매출 성장 지수

Factor Analysis의 정의 및 예시 설명

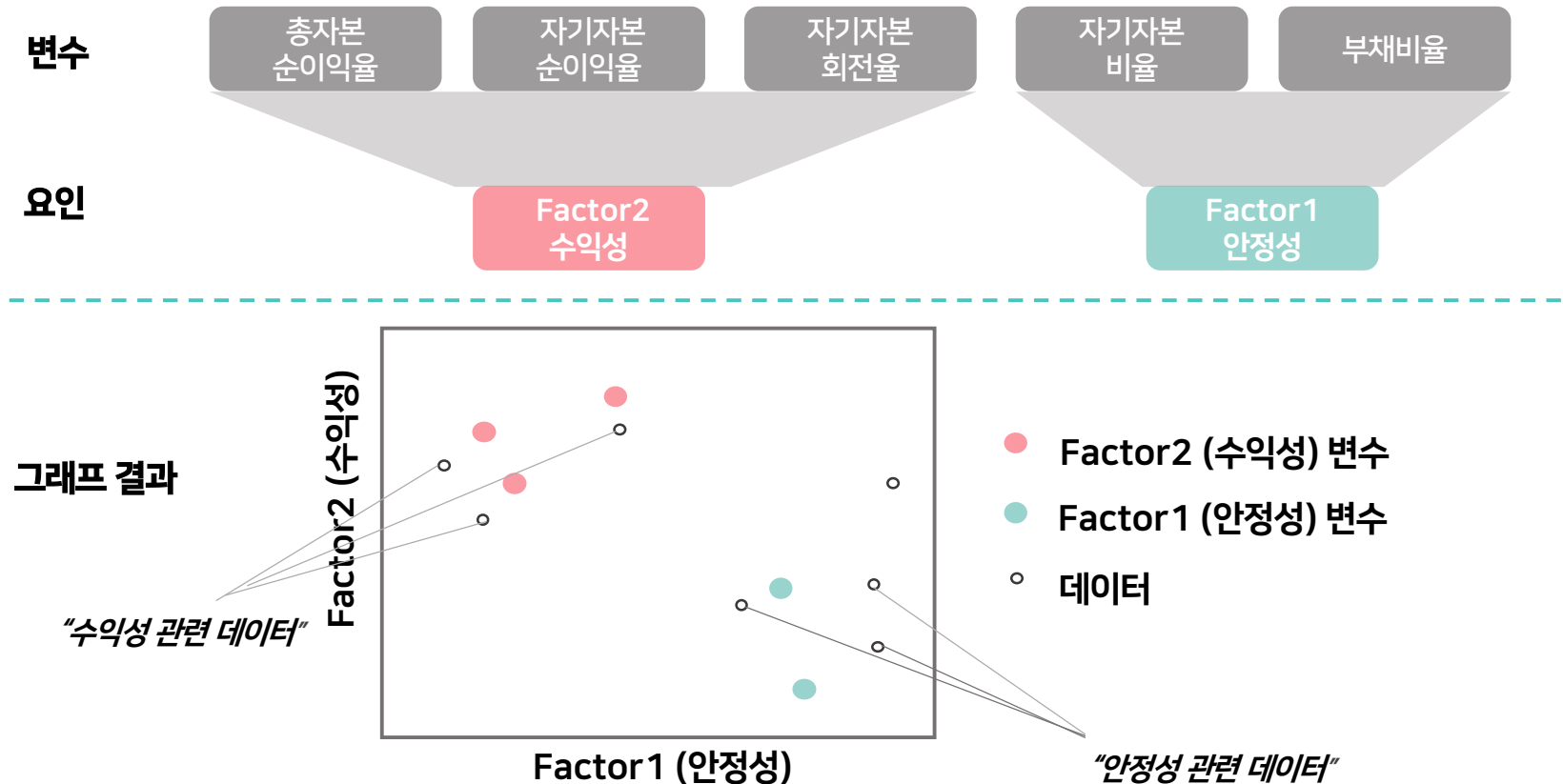
각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

What is Factor Analysis? 변수들을 요인으로 묶고 이와 근접한 거리의 데이터들을 같은 요인으로 보는 방법





각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

Index Data : 품목별 매출 성장 지수

'Index.csv'를 소비자 유형(타겟) 별 3-5월의 품목별 성장 비율 변화율로 전처리함

전처리: 나이대별/성별 타겟팅

age	gender	가공식품		화장품
20	all	-0.232		-0.221
20	F	-0.246		-0.087
20	M	-0.186	-0.465
.				
.				
all	M	-0.192		-0.173

“20-60대 & 성별로 총 18개의 타겟
(모든 나이대, 성별 포함) ”

“변수: 총 17개 품목별 3-5월 변화율
(승-전으로 넘어가는 양상으로
포스트 코로나 예측 목적) ”



각 데이터로 보는 유망 산업

Delivery data

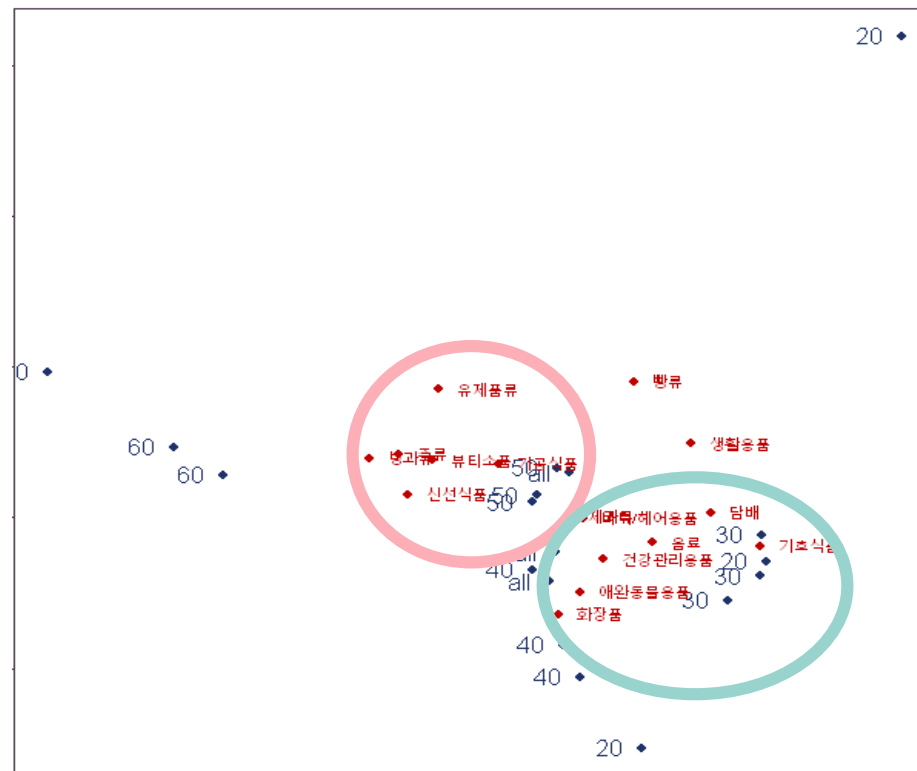
Index data

Card data

Index Data : 품목별 매출 성장 지수

전처리한 데이터를 바탕으로 Factor Analysis를 함

Factor Analysis 시각화



Factor1

건강관리용품, 기호식품,
애완동물용품, 화장품,
담배 등



Factor2

유제품류, 신선식품,
주류, 빙과류



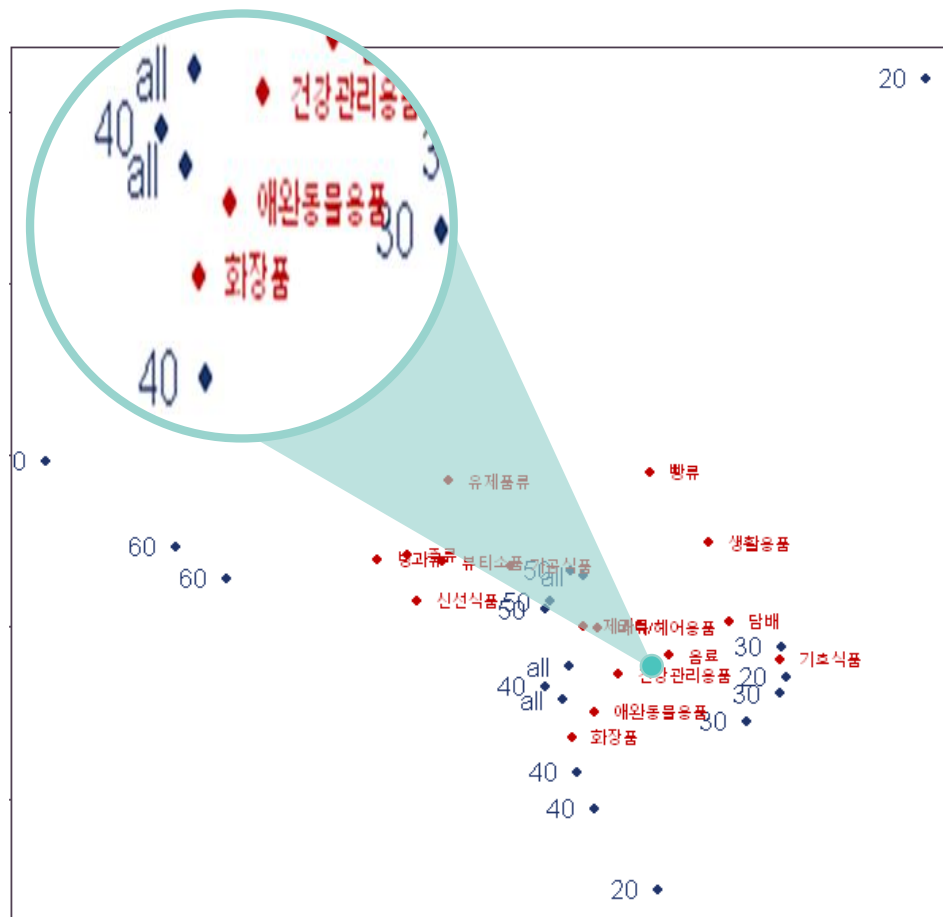
전처리한 데이터를 바탕으로 Factor Analysis를 함

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data



'화장품, 애완동물용품, 건강관리용품'
이 같이 묶이고, 주 타겟이 40대임

→세 품목이 상승세라면,
40대로 타겟팅했을 때
유망할 가능성이 있음

Index Data : 품목별 매출 성장 지수

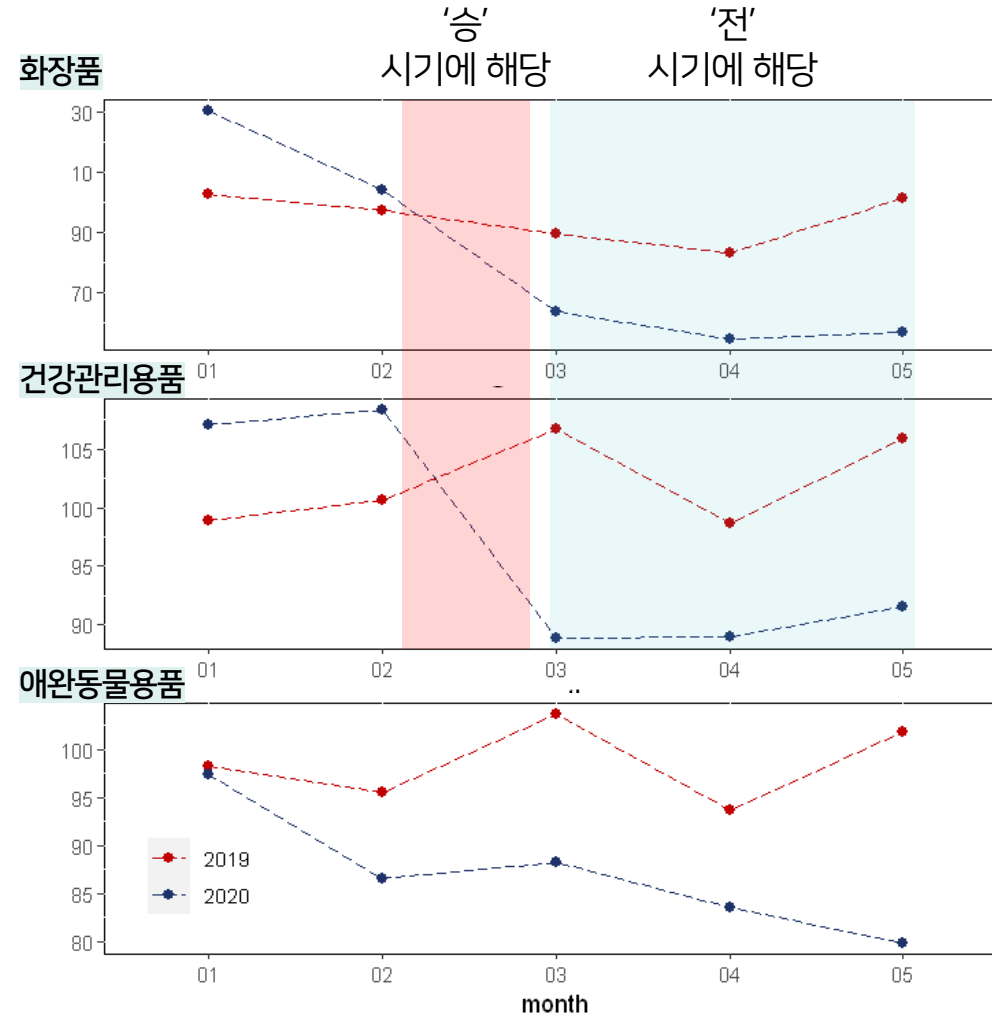
같이 묶인 '화장품, 애완동물용품, 건강관리용품'의 2019년과 2020년의 매출 성장 비율의 차이 및 추이를 봄

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data



'화장품과 건강관리용품'

:'승'시기에 매출 감소했으나
'전'시기에 상대적으로 매출 호전

→ 화장품과 건강관리용품을
40대로 타겟팅 했을 때
유망할 가능성이 있음

Index Data : 품목별 매출 성장 지수

국내 3대 화장품 회사인 '아모레퍼시픽 산업보고서' 를 통해 '화장품' 품목의 매출 상승 솔루션을 구체화함

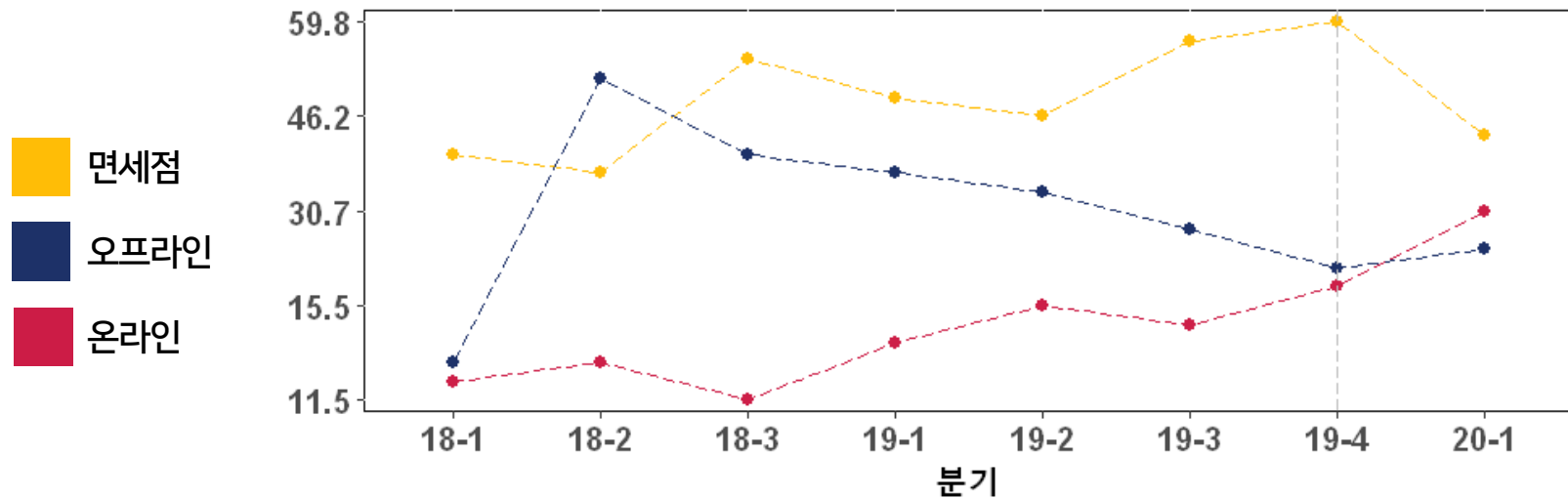
각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

솔루션 구체화



'승' 에서 매출 감소 요인 : 중국인
입국금지로 면세점 매출 비율 하락

→입국금지 풀렸으므로
면세점 매출로 인한 매출 호전 예상



오프라인매장 매출 감소와
온라인 매장 매출 증가

→온라인매장 적극 활용 시
매출 증가 예상



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 업종 별 결제금액

업종 별 결제금액 데이터의 특징과 K-MEANS 클러스터링, 시계열 요인분석 두가지 분석 방법 정리

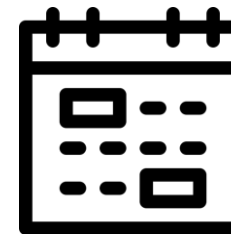
Card Data 특징



업종 별 결제금액 데이터



서울시 내 자치구 ·
행정동별 데이터



일별 데이터
2020.01.01~2020.06.14

Key Point

- 시기 별 평균 판매량·매출·가격에 따라 클러스터링
- 업종 별 평균 거래량의 시계열 변화 양상에 따라 시계열 요인 분석

Card Data : 업종 별 결제금액

K-MEANS 알고리즘의 원리와 특징 설명

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

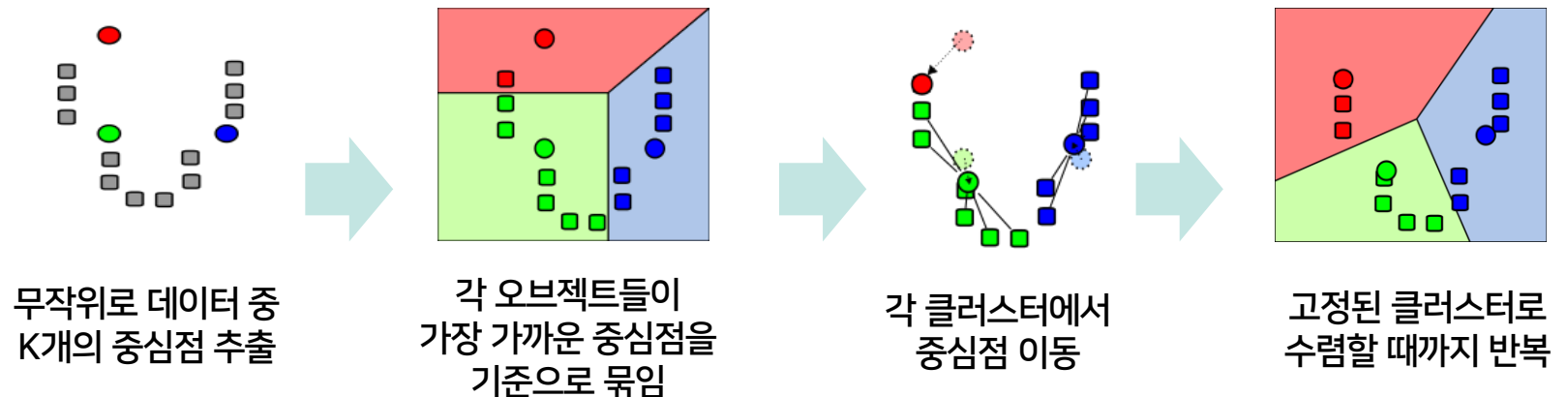
- K-MEANS 클러스터링
- Time series FA

Why K-MEANS?

출처: Wikipedia

K-MEANS 알고리즘

주어진 데이터를 k개의 클러스터로 묶는 알고리즘으로,
각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다.





각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 업종 별 결제금액

K-MEANS는 클러스터의 중심점을 이동할 때 이상치의 영향을 크게 받으므로 클러스터링 결과에서 이상치 발견 쉬움

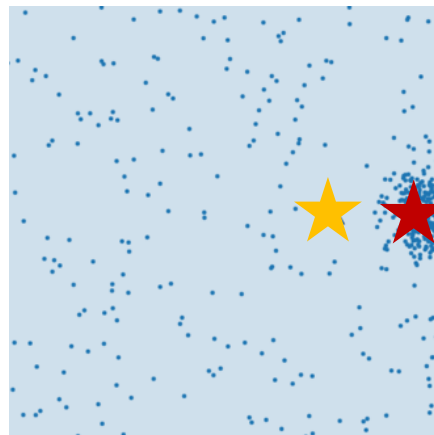
Why K-MEANS?

출처: Wikipedia

K-MEANS 알고리즘

클러스터의 중심점을 이동할 때 이상치의 영향을 크게 받는다.

→ 클러스터링 결과에서 이상치 발견이 쉬움



★ : K-Means로 수렴한 중심점

★ : 실제 데이터의 중심점

→ 이상치의 영향을 받아 클러스터의 중심점이 실제 중심점보다 떨어지게 됨



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 업종 별 결제금액

이상치(유망 산업) 발견이 쉽다는 장점을 이용하여 인사이트를 더 쉽게 발굴하고자 K-MEANS 알고리즘 적용

Why K-MEANS?

출처: Wikipedia

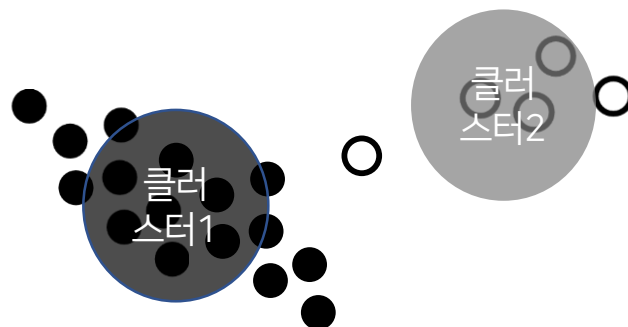
K-MEANS 알고리즘

클러스터링 결과에서 이상치 발견이 쉬움

→ 우리가 발굴하고자 하는 '유망 산업' 일 가능성 ↑

∴ 이상치를 반영하여 유망산업에 대한 인사이트를 더 쉽게 발굴

이상치를 제거하지 않을 경우



이상치를 제거할 경우





각 데이터로 보는 유망 산업

Delivery data

Index data

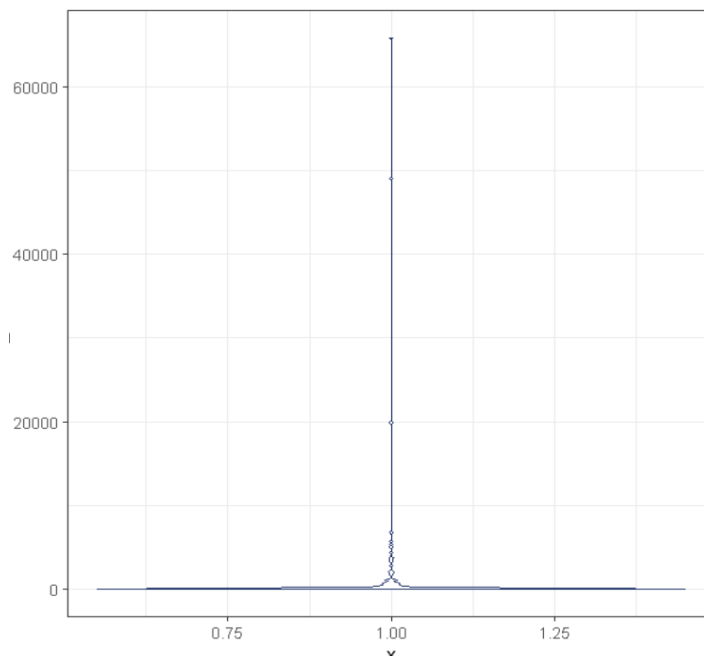
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 업종 별 결제금액

업종 별 결제금액 데이터를 시기별 평균 판매량, 평균 매출, 평균 가격으로 K-MEANS 클러스터링하여 유망 산업 발굴

업종별 데이터의 전체 기간 평균 판매량 분포: Violinplot



- 거의 모든 데이터가 아래쪽에 쏠려 있음
- 전체 기간 평균 판매량 기준,
하위 25%인 업종은 제거

→ 절대적인 판매량 자체가 낮을 경우,
유망 산업이라고 보기 어려우므로!

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

클러스터링에 필요한 시기별 각 업종의 평균 판매량, 평균 매출 데이터 생성

클러스터링에 필요한 데이터 생성



재정의된 시기별 각 업종의 평균 판매량, 평균 매출 데이터



데이터를 시기별/ 업종별로 group_by



판매량과 매출의 평균을 구함



<품목별 평균 판매량>

	기	승	전-1	전-2
품목1				
품목2				
품목3				
.				
.				

<품목별 평균 매출>

	기	승	전-1	전-2
품목1				
품목2				
품목3				
.				
.				

COMPLETE

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

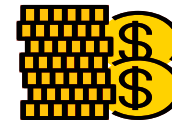
Card Data : K-MEANS Clustering

클러스터링에 필요한 시기별 각 업종의 평균 가격 데이터 생성

클러스터링에 필요한 데이터 생성



재정의된 시기별 각 품목의 평균 가격



데이터를 시기별/ 업종별로 group_by



매출의 합을 판매량의 합으로 나눔



<품목별 평균 가격>

	기	승	전-1	전-2
품목1				
품목2				
품목3				
.				
.				

COMPLETE



각 데이터로 보는 유망 산업

Delivery data

Index data

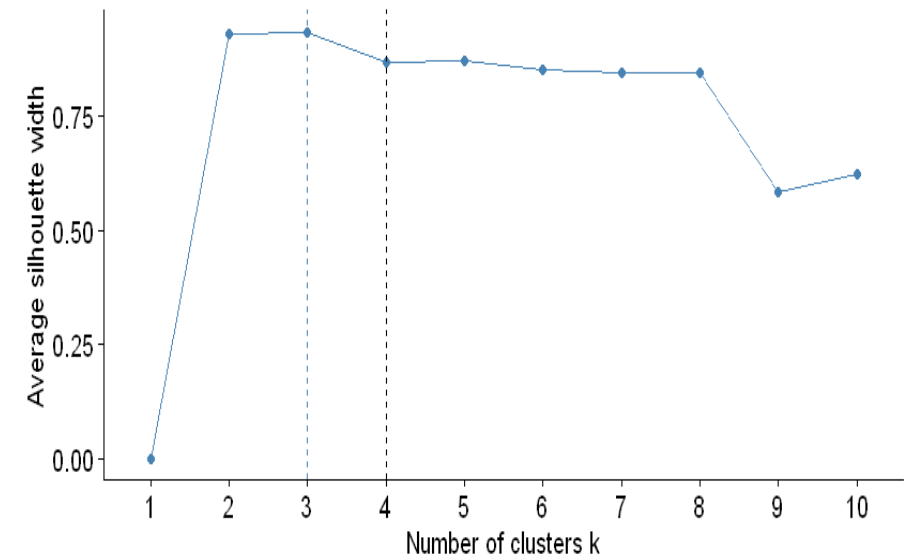
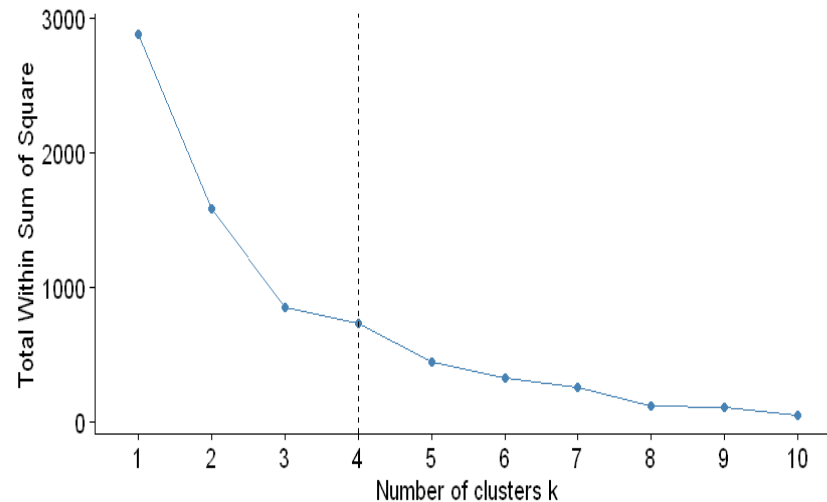
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

K-MEANS 클러스터링에 적절한 K값 찾기: Elbow와 silhouette 값 모두를 고려하여 k=4로 결정함

적절한 K값(클러스터의 수) 찾기



Elbow 값과 silhouette 값 모두를 고려하여 K=4 로 설정함!



각 데이터로 보는 유망 산업

Delivery data

Index data

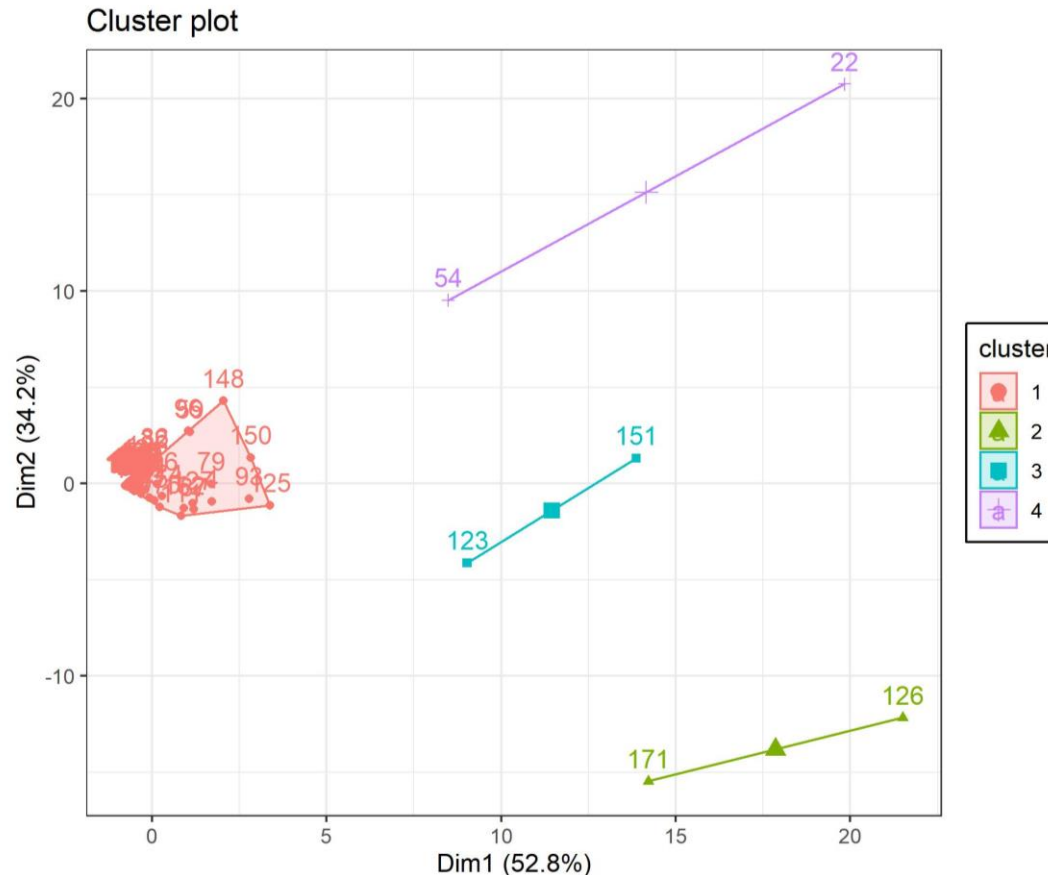
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

데이터들을 네 개의 군집으로 분류한 결과 cluster plot

클러스터링 결과 : Cluster plot



Card Data : K-MEANS Clustering

네개의 클러스터별로 시기별 평균 매출, 평균 판매량 추이를 알아봄.

각 데이터로 보는 유망 산업

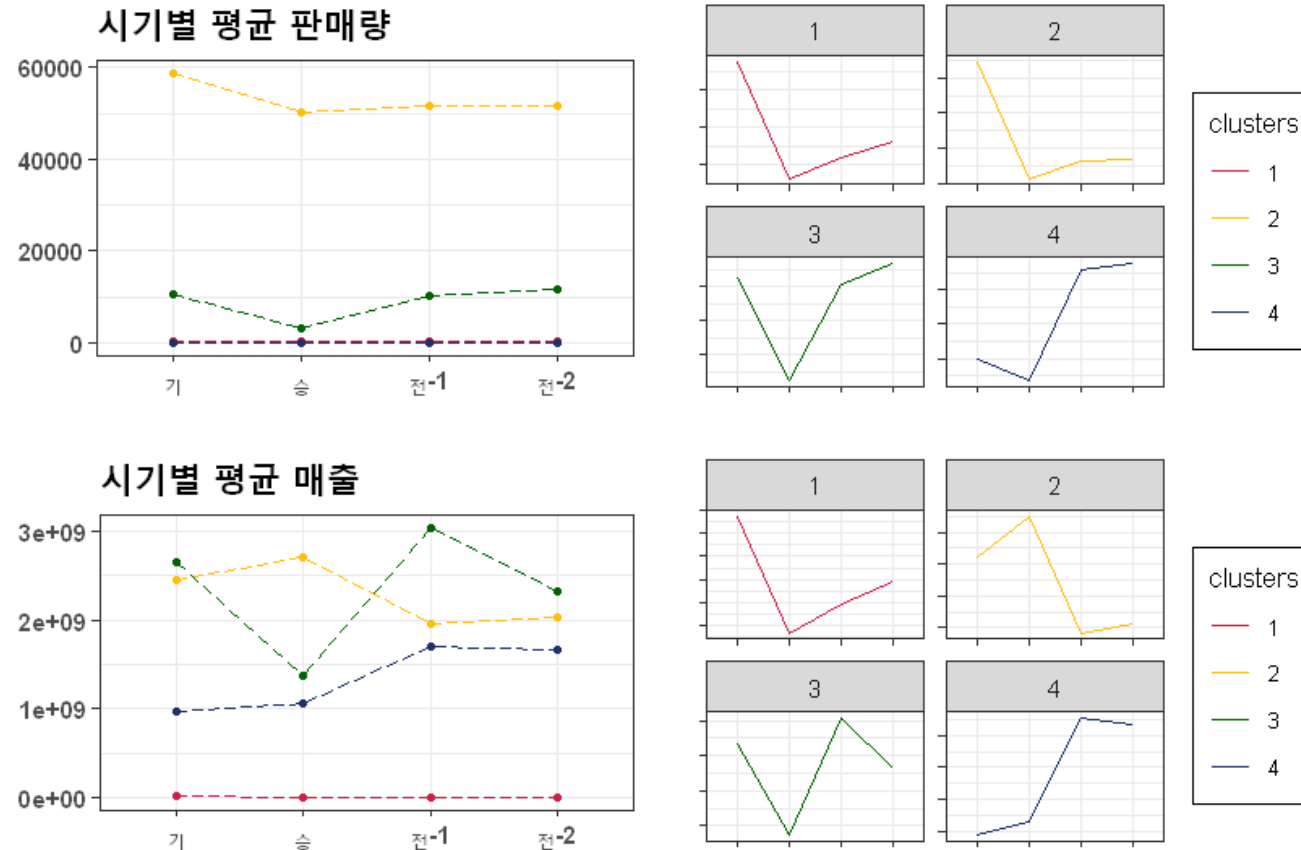
Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

클러스터링 결과 : 시기별 평균 판매량, 시기별 평균 매출



Card Data : K-MEANS Clustering

네개의 클러스터별로 시기별 평균 가격을 알아봄. 가격대를 파악하기 위한 것이므로 절대적인 값이 더 중요

각 데이터로 보는 유망 산업

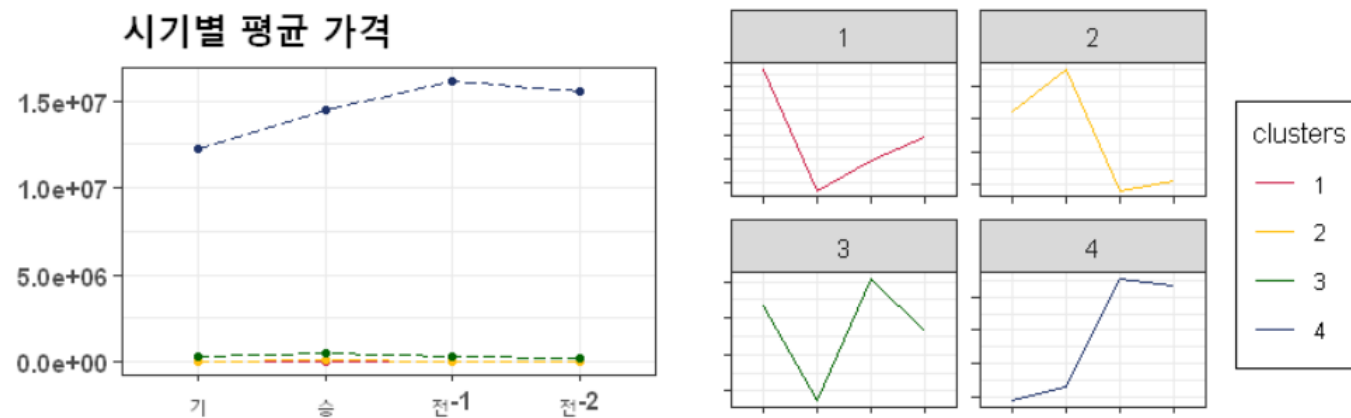
Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

클러스터링 결과 : 시기별 평균 가격



- 평균 가격의 경우 클러스터별 추세보다는 절대적인 값에 더 비중을 두어야 함!
- 클러스터 4에 속한 항목들의 물가가 다른 클러스터보다 월등히 높음



각 데이터로 보는 유망 산업

Delivery data

Index data

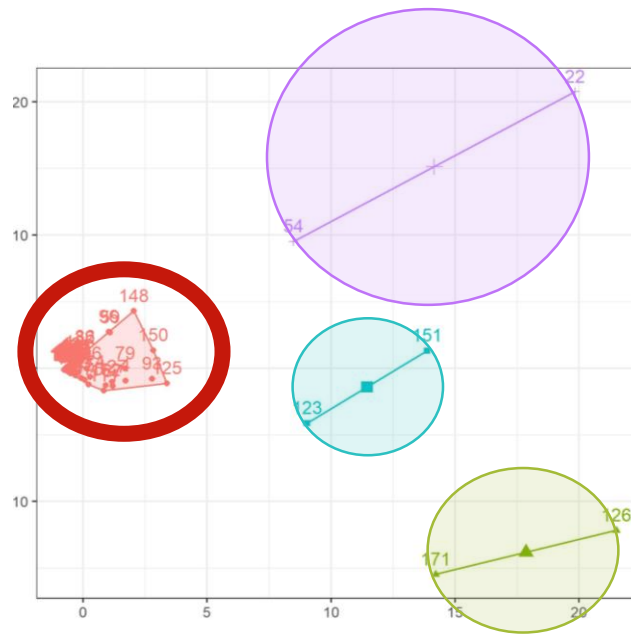
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

클러스터 1에 묶인 품목들은 경제가 침체됨과 함께 평균 판매량과 매출이 감소하고 이후 자연스레 회복되는 것을 보아, 생활 기저 산업일 것으로 예상됨.

클러스터1 : 생활 기저 산업



항목	형태
시기별 평균 판매량	감소했다가 회복하는 형태
시기별 평균 매출	감소했다가 회복하는 형태
평균 가격	상대적으로 낮음

- 가장 많은 품목이 묶였음.
- 다른 클러스터와 단위가 크게 차이남.
- 평균 판매량과 매출 모두 '승'시기에 급격히 감소하였다가 다시 원래의 판매량과 매출을 회복



각 데이터로 보는 유망 산업

Delivery data

Index data

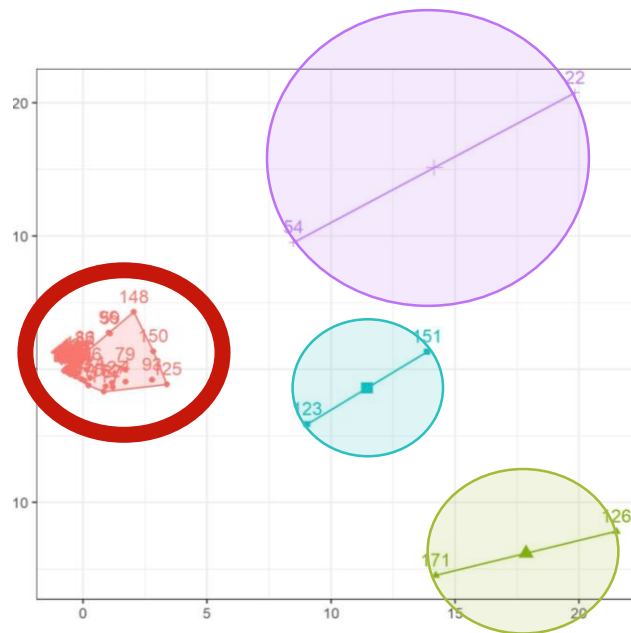
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

클러스터 1에 묶인 품목들은 생활 기저 산업이므로, 포스트 코로나의 변화하는 소비 패턴과 관련이 있다고 보기 어려움.

클러스터1 : 생활 기저 산업



- ☑ 생활에 필수적이거나, 대부분 유흥과 관련된 품목
- ☑ 코로나 사태가 심각해지고, 경제가 침체되면서 판매량과 매출 감소
- ☑ 포스트 코로나의 소비 패턴 변화와 관련이 있다고 보기 **어려움**

Card Data : K-MEANS Clustering

클러스터 2는 다윗수 소액 결제를 기반으로 한 언택트 소비 경향을 보여줌. 승 시기를 기준으로 평균 판매량과 매출의 증감이 뒤집힘

각 데이터로 보는 유망 산업

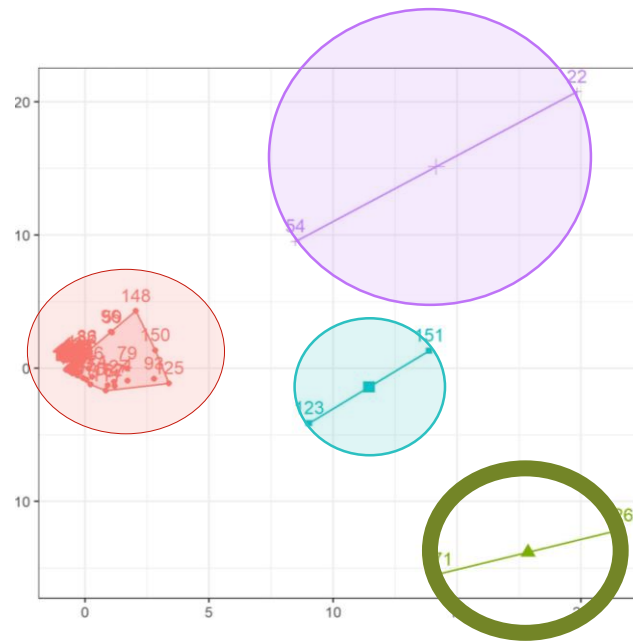
Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

클러스터2 : 다윗수 소액 결제를 기반으로 한 언택트(Untact)소비



항목	형태
시기별 평균 판매량	감소했다가 증가하는 형태
시기별 평균 매출	증가했다가 감소하는 형태
평균 가격	가장 낮음

- 묶인 품목: 인터넷 P/G(인터넷 결제대행), 택시
- '승' 시기를 기점으로 평균 판매량과 매출의 증감이 뒤집힘



각 데이터로 보는 유망 산업

Delivery data

Index data

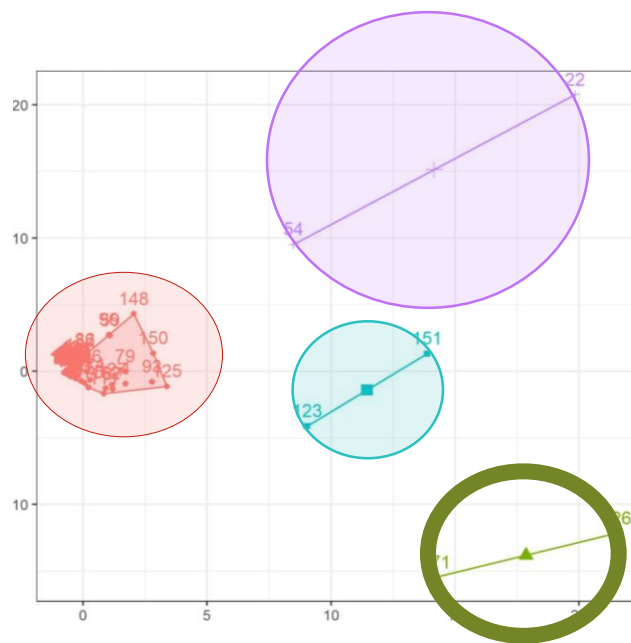
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

포스트 코로나 시기에 한 번에 큰 수요를 해결하기보다, 작은 수요를 그때그때 간편한 서비스를 이용해 해결하는 패턴 보임

클러스터2 : 다횟수 소액 결제를 기반으로 한 언택트(Untact)소비



'승' 시기 이전 : 인터넷을 이용해 한 번에 필요한 생활 용품을 많이 구매하거나, 대중교통 이용을 자제하여 이동시 대중교통보다 택시를 이용하게 되었음.



'승' 시기 이후 : 한 번에 큰 수요를 해결하기보다, 작은 수요를 그때그때 인터넷이나 택시와 같은 간편한 서비스를 이용해 해결하고자 함



각 데이터로 보는 유망 산업

Delivery data

Index data

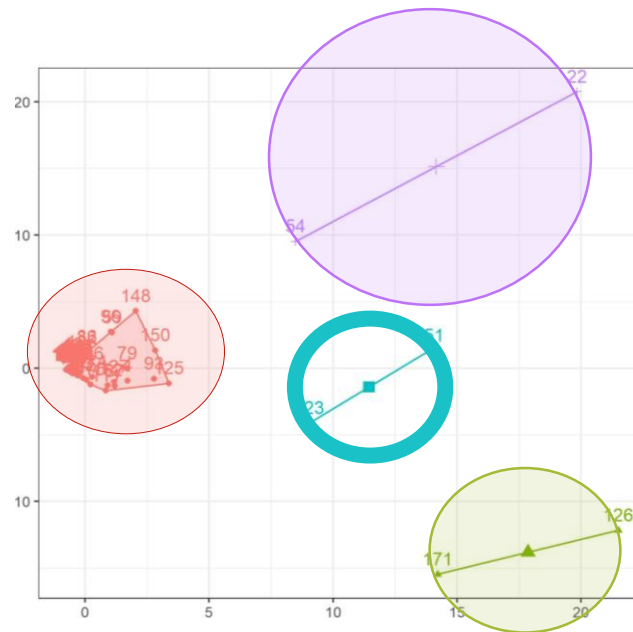
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

클러스터 3은 디지털 콘택트와 정부 개입의 효율성을 보여줌. 평균 판매량과 매출 모두 감소하였다가 이전의 평균 매출, 판매량 이상으로 증가하는 모습을 보임

클러스터3 : 디지털 콘택트(Digital Contact)와 정부 개입의 효율성



항목	형태
시기별 평균 판매량	감소했다가 증가하는 형태
시기별 평균 매출	감소했다가 증가하는 형태
평균 가격	상대적으로 낮음

- 묶인 품목: 이동통신요금, 조세서비스(특약)
- 평균 판매량과 매출 모두 '승'시기까지는 감소, '전'시기에 이전의 평균 매출, 평균 판매량 이상으로 증가함



각 데이터로 보는 유망 산업

Delivery data

Index data

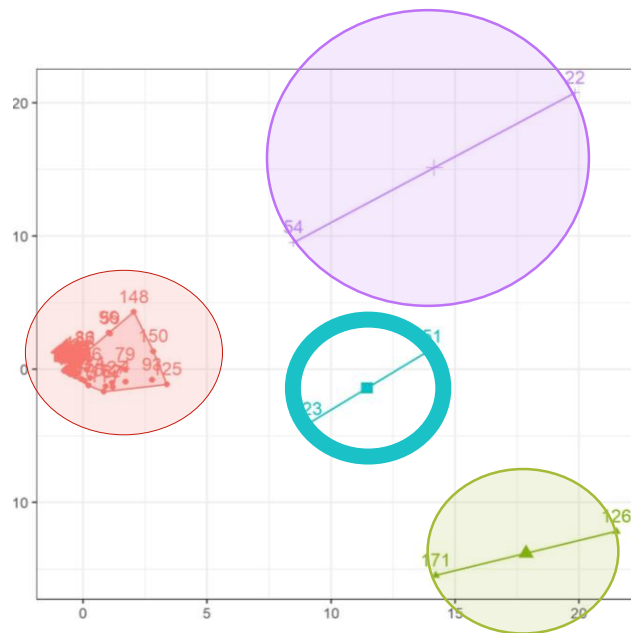
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

이동통신산업의 경우 디지털 콘택트로 대표되는 포스트 코로나 시기의 대표적인 유망 산업 중 하나임.

클러스터3 : 디지털 콘택트(Digital Contact)와 정부 개입의 효율성



이동통신산업의 경우 '디지털 콘택트'로 대표되는 포스트 코로나 시기의 대표적인 유망 산업 중 하나



유동인구가 침체된 '승' 시기에 평균 판매량과 매출이 감소, 유동인구가 풀리는 '전' 시기에 급격하게 증가



각 데이터로 보는 유망 산업

Delivery data

Index data

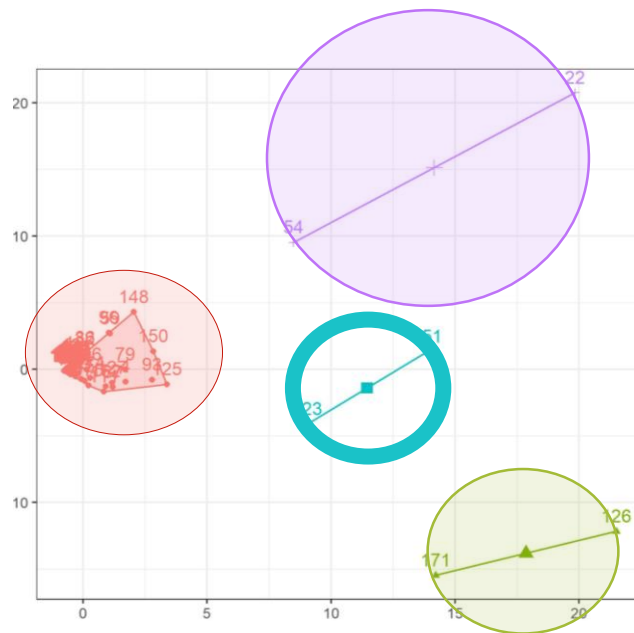
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

정부가 지방세 납부에 대한 지역 주민의 부담을 덜고자 지원 정책을 시행하자 평균 납부 금액과 납부 횟수 증가:정부 개입의 효율성

클러스터3 : 디지털 콘택트(Digital Contact)와 정부 개입의 효율성



- ✓ 조세서비스(특약): 각종 혜택이 적용된 지방세 납부
- ✓ '승'시기에 코로나19 확진자가 급격히 증가하고 경제가 침체되자 지방세 납부가 어려워짐
- ✓ 정부의 지역 주민의 부담을 덜고자 지방세 납부 기간을 연장하고 징수를 유예하는 등 정책 시행 → 평균 납부 금액과 평균 납부 횟수가 '전' 시기 이후 증가

Card Data : K-MEANS Clustering

클러스터 4는 포스트 코로나 시기 내수 시장이 활성화된 모습을 대변함.
한 번에 결제하게 되는 금액이 크며, '승' 시기 이후 평균 판매량과 매출이 모두 증가하였음

각 데이터로 보는 유망 산업

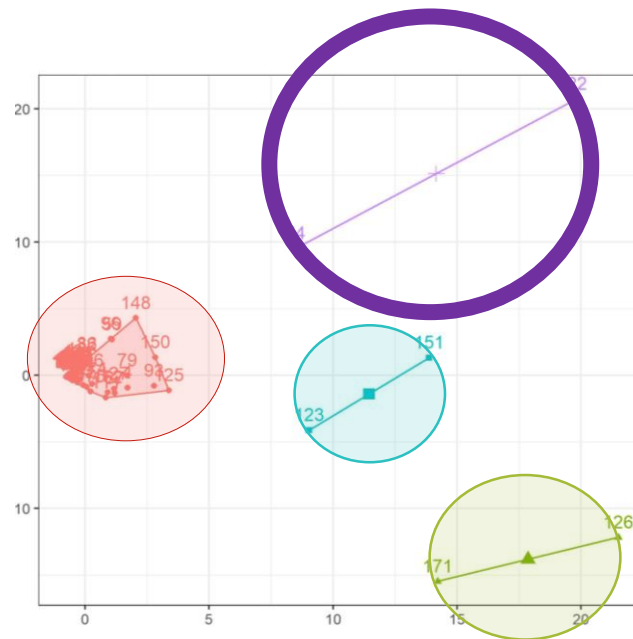
Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

클러스터4 : 내수 시장 활성화



항목	형태
시기별 평균 판매량	감소했다가 급격히 증가하는 형태
시기별 평균 매출	증가하는 형태
평균 가격	매우 높음

- 묶인 품목: 국산 신차, 농축수산물가공품
- 상대적으로 한 번에 결제하게 되는 금액이 큼.
- '승'시기 이후 평균 판매량과 매출 모두 증가



각 데이터로 보는 유망 산업

Delivery data

Index data

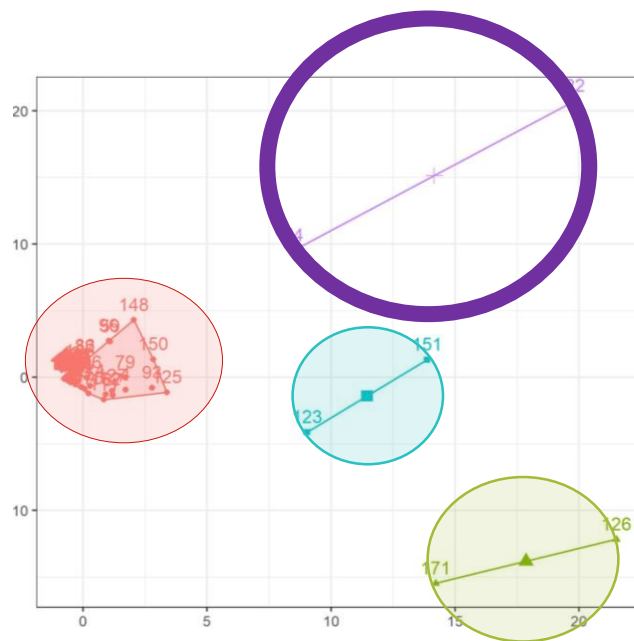
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

국산 신차의 판매량과 매출이 증가하는 경향을 띠며, 정부가 개소세 인하 정책을 도입하자 전년 동기 대비 국산차 판매량이 증가함

클러스터4 : 내수 시장 활성화



국산 신차의 판매량, 매출 증가



2월 말 정부가 코로나 19사태 조기 진압을 위해
개소세 인하 정책 도입
→ 전년 동기 대비 국내 완성차 내수 판매량 증가



같은 기간 5개사의 해외 판매량은 급감



각 데이터로 보는 유망 산업

Delivery data

Index data

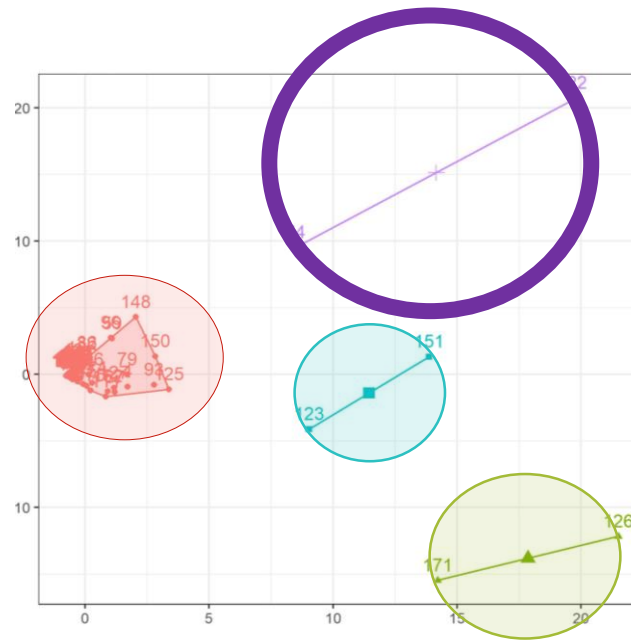
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : K-MEANS Clustering

코로나19사태로 국산 농산물에 대한 선호도가 증가함에 따라 농축수산가공품의 판매량과 매출이 증가함

클러스터4 : 내수 시장 활성화



- ✓ 농축수산가공품의 판매량, 매출 증가
- ✓ 코로나 19 사태로 국산 농산물에 대한 선호 ↑
- ✓ '집밥족'들의 증가 → 국산 농식품 구매 증가



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 시계열 요인 분석

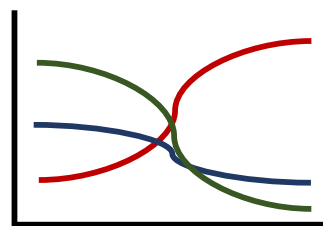
시계열 요인분석의 개념과 원리에 대해 설명

Why Time Series FA?

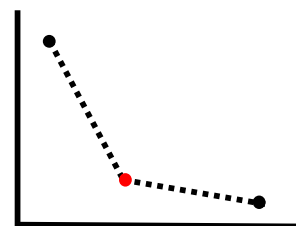


Time Series FA

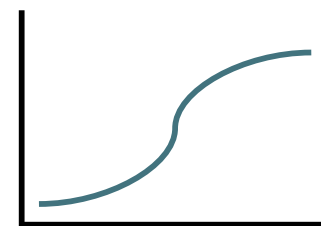
다수의 시계열 데이터 안의 잠재 요인을 찾는 기법으로,
데이터 간의 상관관계를 분석하여 공통 잠재 요인으로 축약하는 기법



다수의 시계열 데이터



고유값 확인



잠재요인 1



잠재요인 2



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 시계열 요인 분석

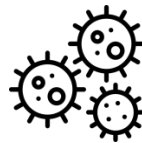
시계열 예측 모델의 경우 업종별로 모든 변수를 각각 고려하기 힘들

Why Time Series FA?

Time Series FA

시계열 예측 모델의 경우 코로나 이외의 업종별로 판매량 증감에 영향을 주는 변수를 모두 고려하기 힘들

예시



코로나



유가



환율



기온

영향



항공사 거래량



Card Data : 시계열 요인 분석

시계열 요인 분석을 통해 증가하는 잠재 요인에 어떠한 업종들이 반영되어있는지 확인함으로써 유망업종 파악 가능

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

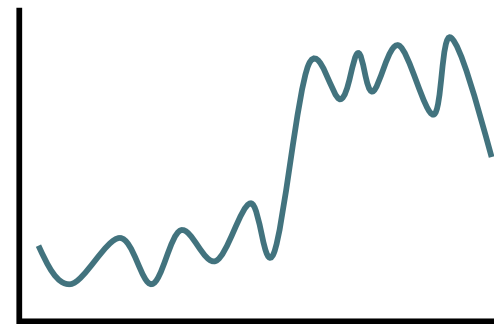
- K-MEANS 클러스터링
- Time series FA

Why Time Series FA?

Time Series FA

산출된 잠재 요인 중 증가하는 추세를 가진 잠재요인에
어떠한 업종들이 반영되어 있는지 확인함으로써 유망 업종 파악 가능

예시



잠재요인 1



택시



인터넷 쇼핑



가전 제품



핸드폰 결제



각 데이터로 보는 유망 산업

Delivery data

Index data

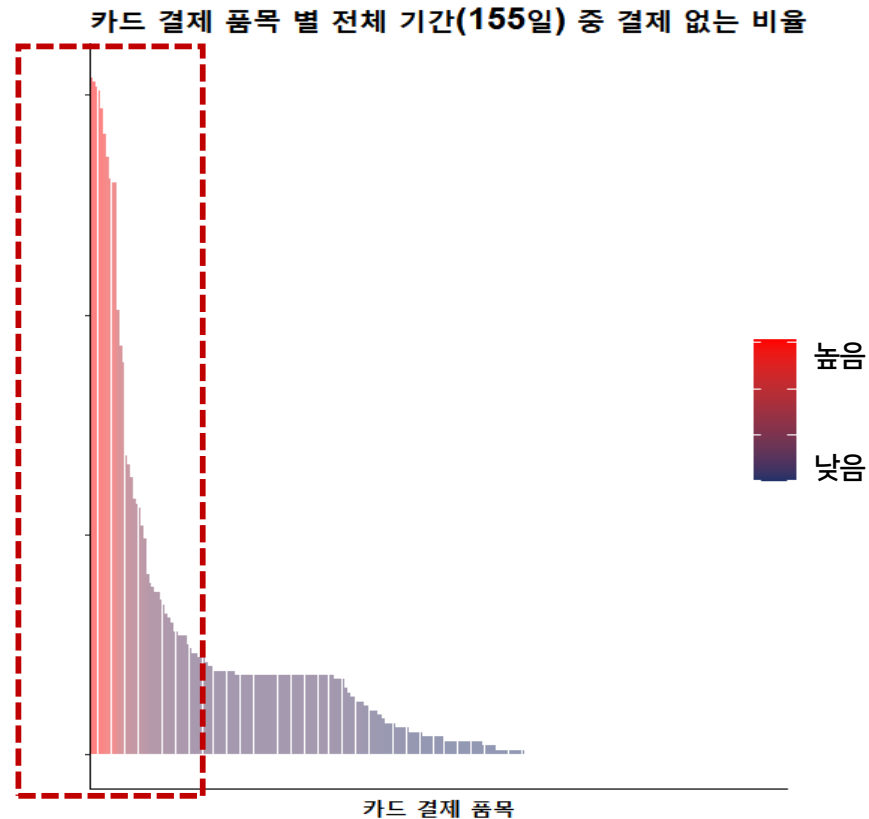
Card data

- K-MEANS 클러스터링
- Time series FA

Card Data : 시계열 요인 분석

전체 기간 중 결제가 발생하지 않은 비중이 큰 품목에 경우 분석에 노이즈(Noise)로 작용할 수 있으므로 제거

품목 별 전체 기간 중 결제가 발생하지 않은 횟수



결제가 발생하지 않은 횟수가 큰 품목



유망업종이라 보기 힘들므로
제거하고 진행

결제가 없는 품목이 노이즈(Noise)로 작용해 분석결과가 왜곡될 수 있음



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

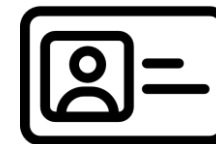
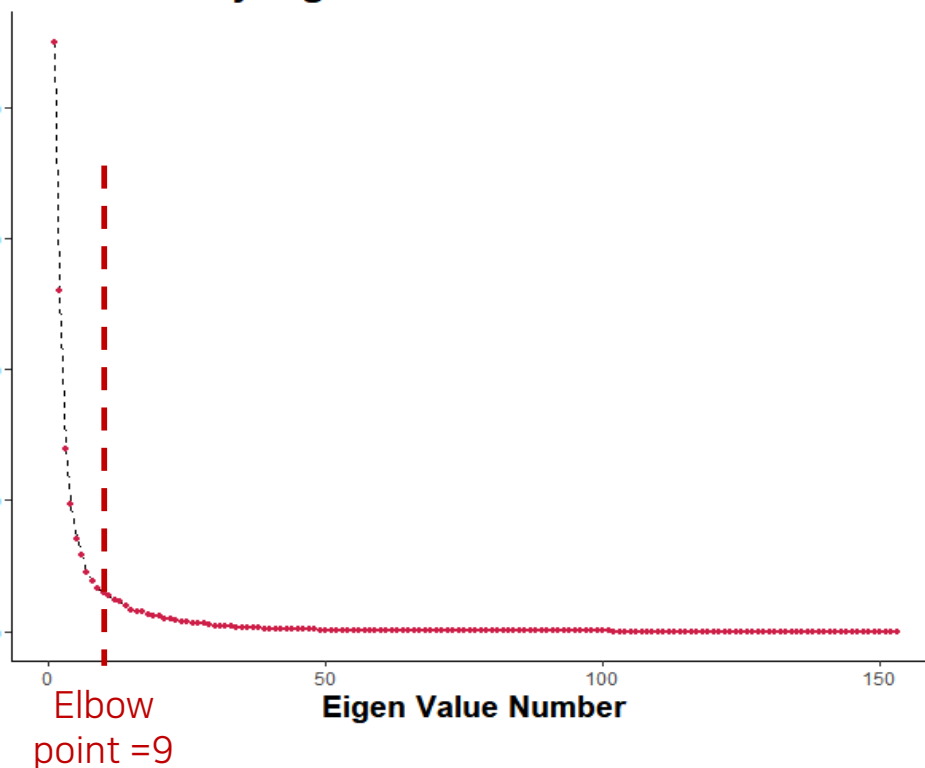
- K-MEANS 클러스터링
- Time series FA

Card Data : 시계열 요인 분석

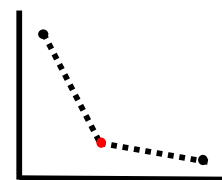
고유값(Eigen value) Scree plot을 통해 시계열 요인 분석 결과로 나올 요인 개수 결정

고유값(Eigen value) Scree plot

Scree Plot by Eigen Value



요인 개수 결정을 위해 고유값 확인



고유값이 확 떨어지는 지점인
9개로 요인 개수 결정(elbow point)

Card Data : 시계열 요인 분석

시계열 요인 분석을 통해 9가지 잠재요인 산출. 잠재요인 6과 잠재요인7에서 증가하는 추세 확인

각 데이터로 보는 유망 산업

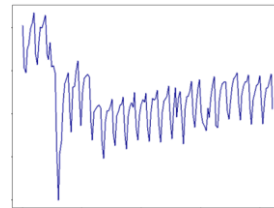
Delivery data

Index data

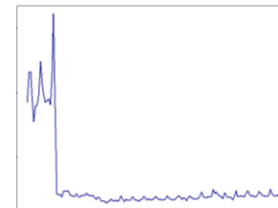
Card data

- K-MEANS 클러스터링
- Time series FA

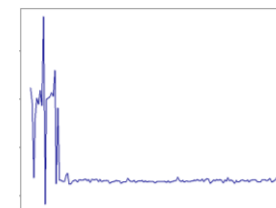
산출된 9가지 시계열 잠재 요인(factor)



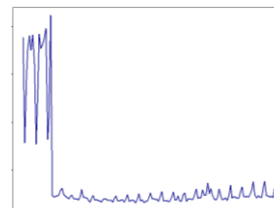
잠재요인 1



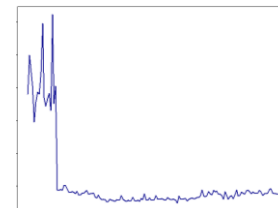
잠재요인 2



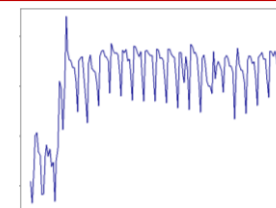
잠재요인 3



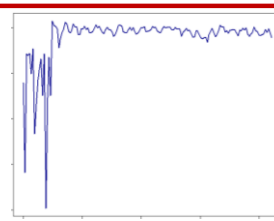
잠재요인 4



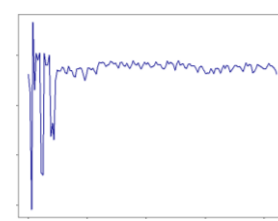
잠재요인 5



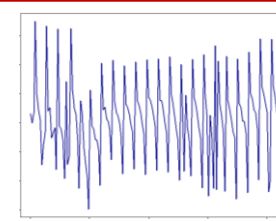
잠재요인 6



잠재요인 7



잠재요인 8



잠재요인 9

 증가 추세



각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

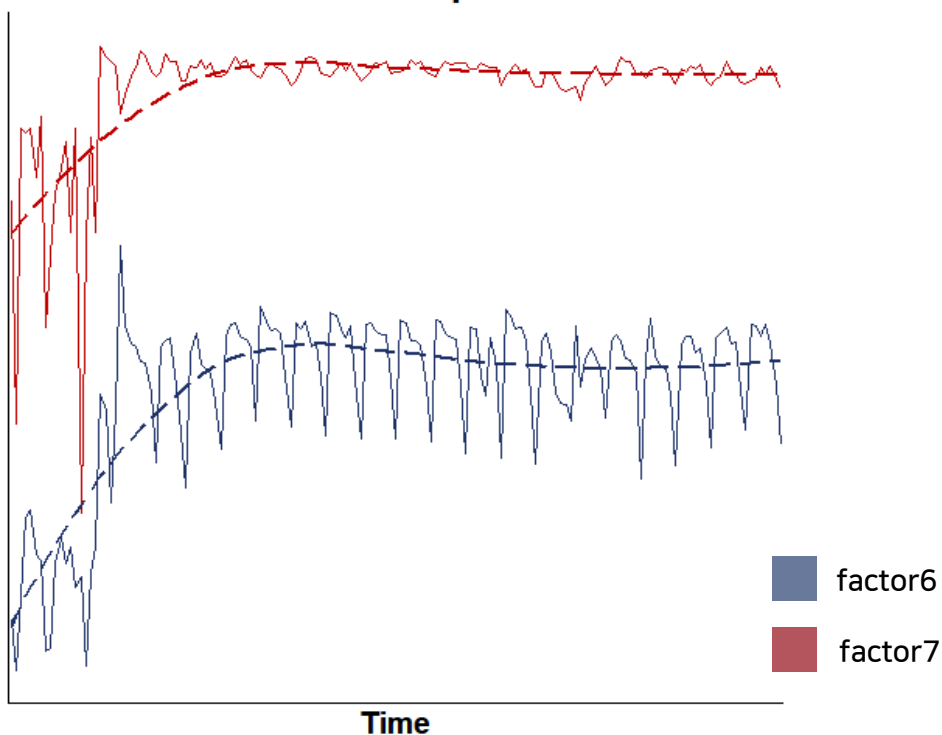
- K-MEANS 클러스터링
- Time series FA

Card Data : 시계열 요인 분석

잠재요인(factor) 6과 7에서 모두 증가하는 추세를 확인. 잠재요인(factor) 6이 보다 가파른 증가 추세를 보임

잠재요인(factor) 6 & 7 시계열 그래프 및 추세

Factor 6 & 7 Time Series plot



잠재요인 6 & 7 모두 증가하는 추세



잠재요인 6이 보다
가파른 증가 추세를 보임

Card Data : 시계열 요인 분석

가파른 증가 추세를 보인 잠재요인(factor) 6에 택시와 인터넷 결제와 홈쇼핑같이 Untact를 반영해주는 품목들이 크게 반영됨

각 데이터로 보는 유망 산업

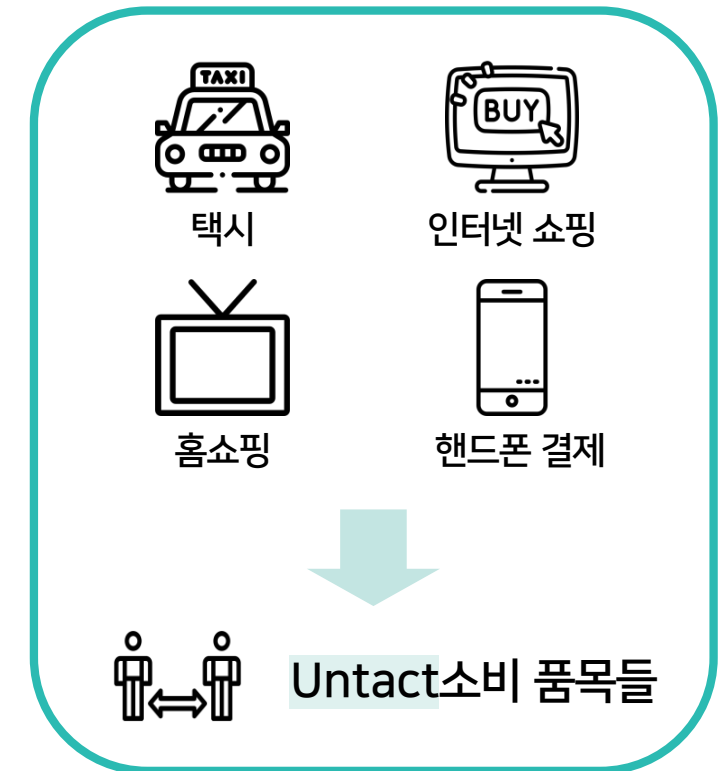
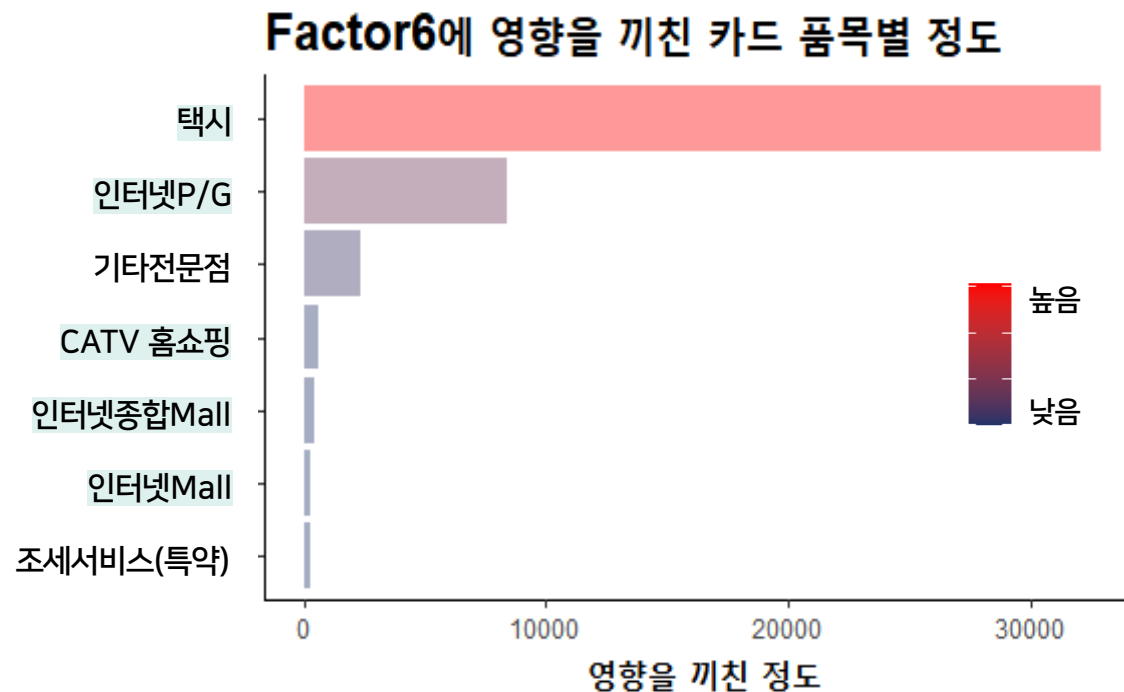
Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

잠재요인(factor) 6에 크게 반영된 품목들



기타전문점: 해당 업종 코드 확인 결과, 기타 품목에 해당하는 것을 확인. 기타라는 범주가 포괄적이라 나타난 결과라고 해석

Card Data : 시계열 요인 분석

증가 추세를 보인 잠재요인(factor) 7에 택시와 인터넷 결제와 홈쇼핑같이 Untact를 반영해주는 품목과 가전제품이 크게 반영됨

각 데이터로 보는 유망 산업

Delivery data

Index data

Card data

- K-MEANS 클러스터링
- Time series FA

잠재요인(factor) 7에 크게 반영된 품목들

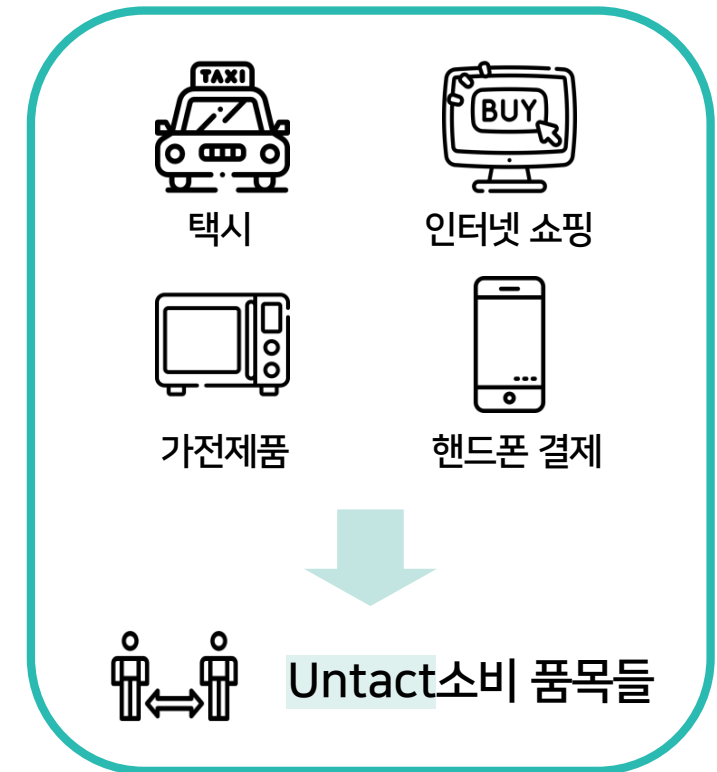
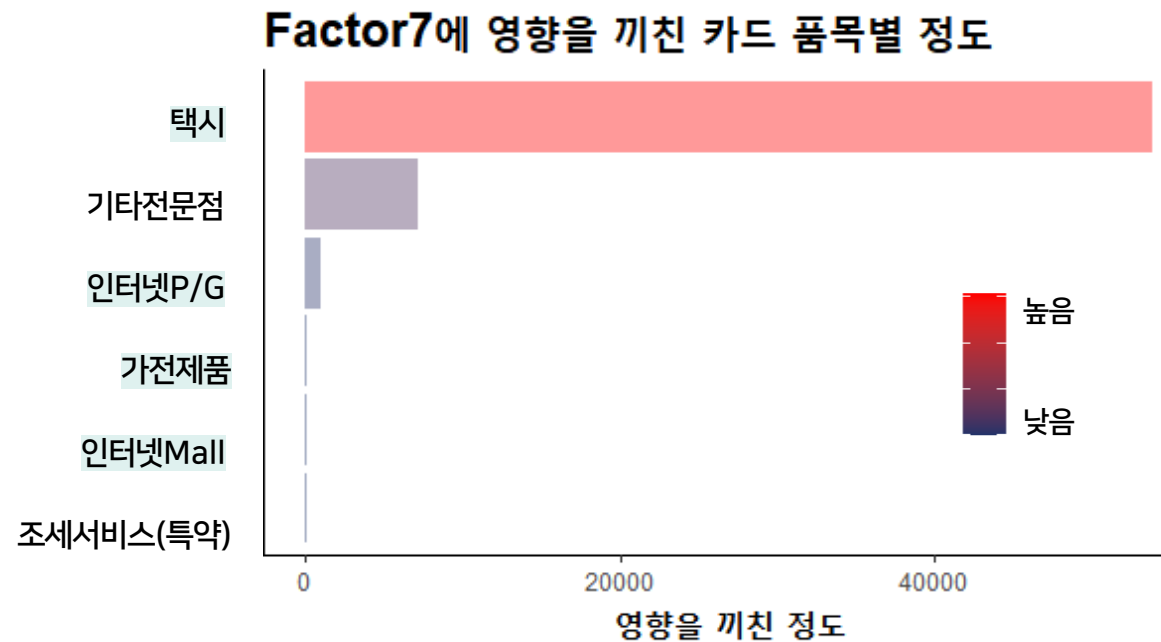


Table of Contents



I. 서론

II. 본론

III. 결론

1. 인사이트 정리

2. 한계 및 시사점



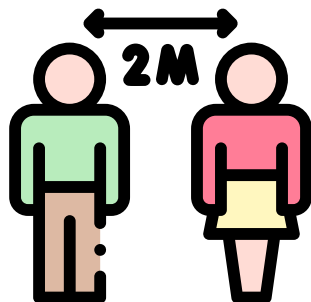
인사이트 정리

배달 호출 정보 데이터(delivery), 품목 별 소비지수 데이터(Index), 업종 별 결제금액 데이터(card)에서 도출한 소비 인사이트 정리

포스트 코로나의 소비 패턴은?

인사이트 정리

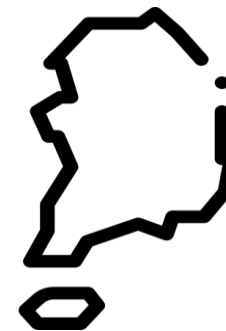
한계 및 시사점



'언택트' 소비
(UNTACT)



즉각적인 수요 해결
(IMMEDIATE)



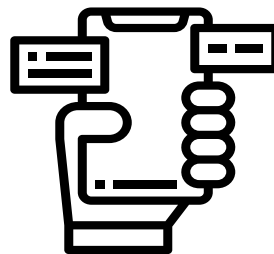
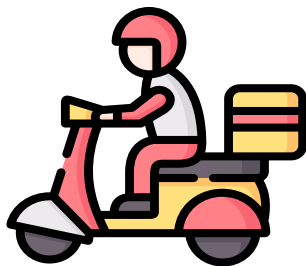
내수 시장 활성화
(DOMESTIC)



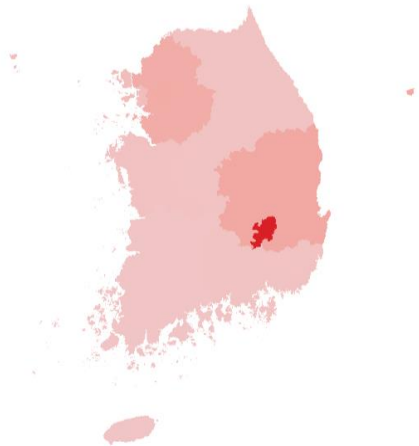
인사이트 정리

언택트(Untact)소비 경향 ①: 코로나 사태 이후 전체적으로 배달량이 증가하였고, 심각 지역에서 증가율이 더 컸음

언택트 (Untact) 소비 경향 : 배달량의 증가



코로나 완만 지역 vs. 코로나 심각 지역(서울, 경기도, 경상북도, 대구)



“ 코로나 사태 이후 전체적으로 배달량 증가
심각 지역에서 배달 건수의 증가 추세가 눈에 띄게 가파름 ”

인사이트 정리

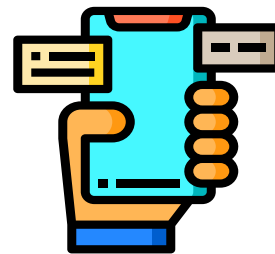
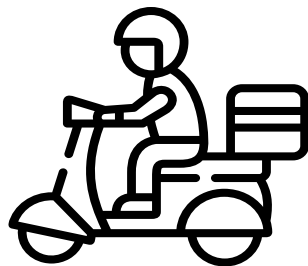
한계 및 시사점



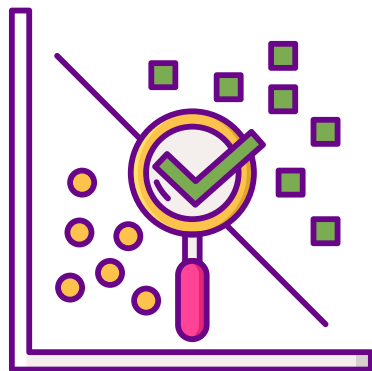
인사이트 정리

언택트(Untact)소비 경향 ②: 온라인 매장을 기반으로 한 화장품 산업, 디지털 통신기반 산업과 이동통신업 등이 유망

언택트 (Untact) 소비 경향 : '디지털 콘택트(Digital Contact)'추구



업종 클러스터링과 품목별 잠재 요인 분석 결과



“ 온라인 매장을 기반으로 한 화장품 산업 활성화,
디지털 통신(인터넷, TV)기반 산업과 이동통신 산업 유망 ”

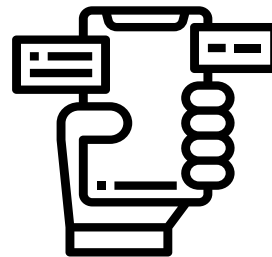
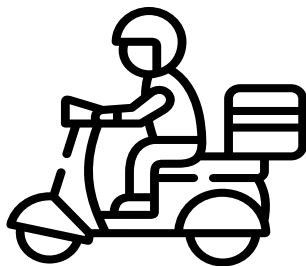
인사이트 정리

한계 및 시사점

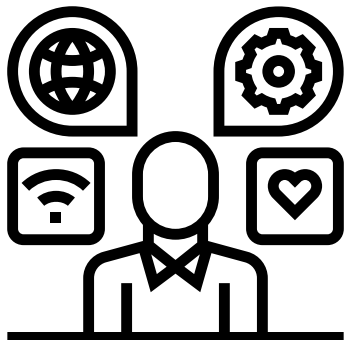
인사이트 정리

언택트(Untact)소비 경향 ③: 다수와의 접촉을 최소화하는 개인화된 서비스의 이용 증가함, 대표적인 예가 택시

언택트 (Untact) 소비 경향 : 개인화된 서비스(택시) 이용 증가



업종 클러스터링과 시계열 잠재 요인 분석 결과



“ 대중교통보다 다수와의 접촉이 적은 택시 이용 증가,
이동 서비스뿐 아니라 개인화된(Personalized) 서비스 유망 ”

인사이트 정리

한계 및 시사점



인사이트 정리

즉각적인 수요(IMMEDIATE) 해결 ①: 그때그때의 작은 수요를 간편 시스템 통해 해결하고자 하는 다횃수 소액 결제 선호 경향성

즉각적인(IMMEDIATE) 수요 해결: 다횃수 소액 결제 선호



업종 클러스터링 결과



“ 한번에 큰 수요를 해결하기보다, 그때그때의 작은 수요
인터넷 쇼핑이나 택시와 같은 간편 시스템 통해 해결하고자 함 ”

인사이트 정리

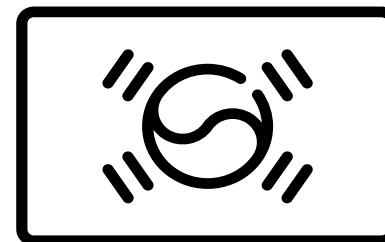
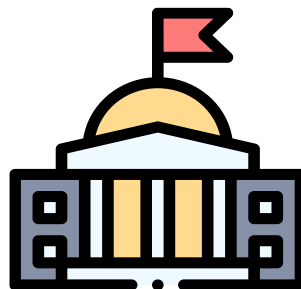
한계 및 시사점



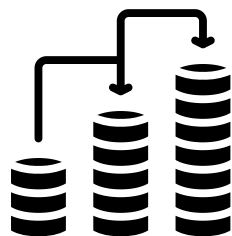
인사이트 정리

내수 시장 활성화 경향 ①: 포스트 코로나 시기에 정부 개입 효과가 극대화됨

내수 시장 활성화 (DOMESTIC): 정부 개입 효과 극대화



업종 클러스터링 결과



“ 정부가 소비 시장에 개입하여(개소세 인하, 지방세 납입 지원 등)
경제를 활성화하고자 할 경우, 그 효과가 극대화됨 ”

인사이트 정리

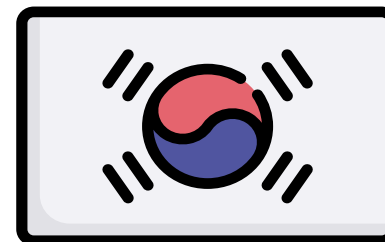
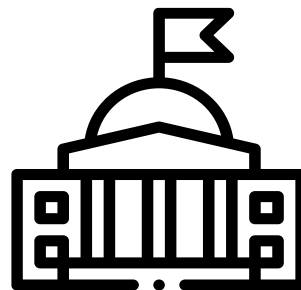
한계 및 시사점



인사이트 정리

내수 시장 활성화 경향 ②: 수입 품목보다 국산 품목에 대한 선호도가 증가함

내수 시장 활성화 (DOMESTIC): 국산 품목에 대한 선호 증가



업종 클러스터링 결과



“

코로나 19사태로 가정 내 조리가 증가하면서,
국산 농산물에 대한 선호도가 높아짐

”

인사이트 정리

한계 및 시사점

한계 및 시사점

한계 ①: 데이터별로 날짜 범위와 특징이 달라, 데이터 간 병합이 어려워 분석을 따로따로 진행하였음

본 조사의 한계

데이터별로 날짜 범위와 특징이 달라, 데이터 간 병합이 어려웠음
→ 분석 따로따로 진행

인사이트 정리

한계 및 시사점





한계 및 시사점

한계 ②: 데이터별로 존재하는 품목이 달라, 데이터 간 병합이 어려워 분석을 따로따로 진행하였음

본 조사의 한계

↓
데이터별로 품목이 달라, 어느 한쪽으로 병합할 수 없었음
→ 분석 따로따로 진행

품목별 소비지수



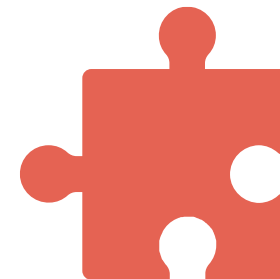
대형마트 내에서 파는
품목만 존재

업종 별 결제금액



결제 가능한
모든 항목 존재

배달 호출 정보



식품만 존재

인사이트 정리

한계 및 시사점



한계 및 시사점

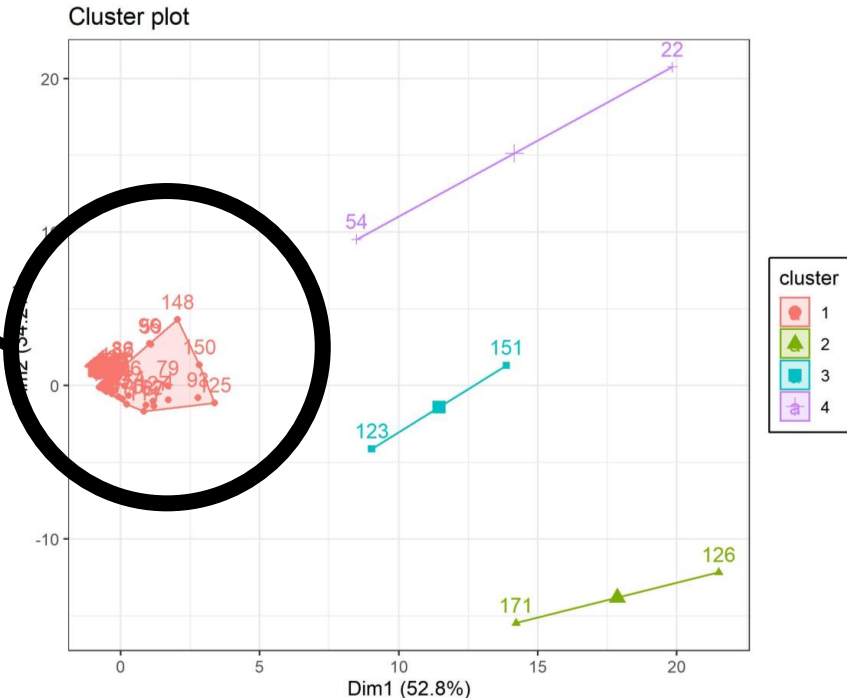
한계 ③: 클러스터링에서 한쪽 클러스터에 너무 많은 품목이 묶이게 되는 불균형 문제를 해결하지 못함

본 조사의 한계



클러스터링에서, 한쪽 클러스터에 너무 많은 품목이 묶이게 되는 불균형(Unbalanced) 문제를 해결하지 못함.

80% 이상의
업종이 묶임



인사이트 정리

한계 및 시사점

한계 및 시사점

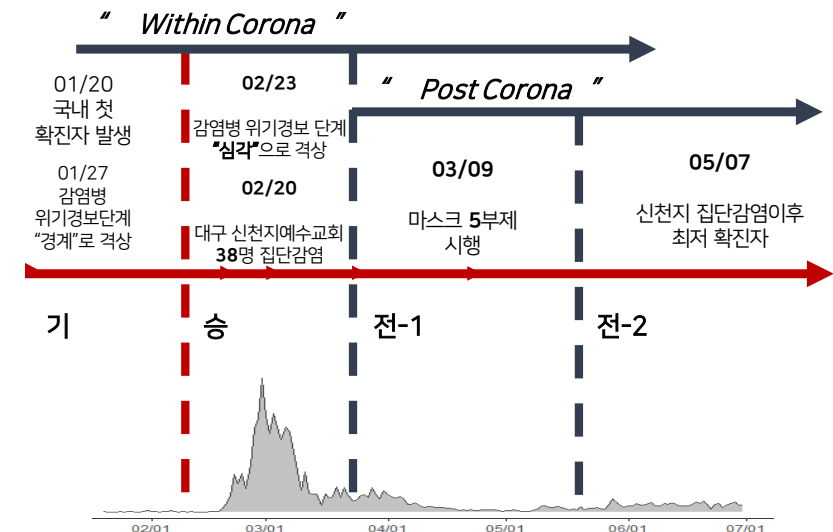
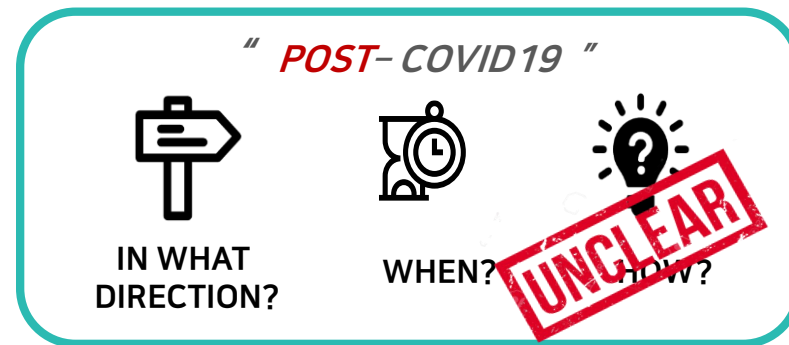
시사점①: 기존에 명확히 정의되지 않았던 포스트 코로나의 시기를 재정의하고 방향을 제시함

본 조사의 시사점

기존에 명확히 구분되지 않았던 코로나19 사태 이후의 시기를
정책과 확진자 데이터를 기반으로 기-승-전-결의 형태로 재정의함.

인사이트 정리

한계 및 시사점





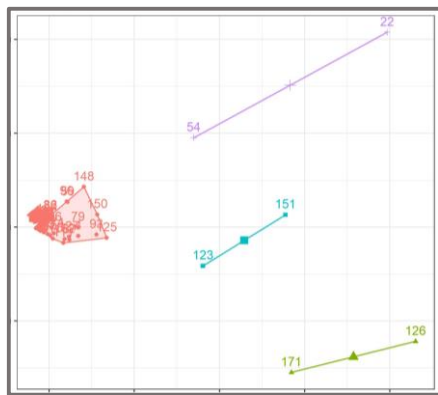
한계 및 시사점

시사점②: 다양한 비지도학습 기법을 활용하여 분석을 진행함

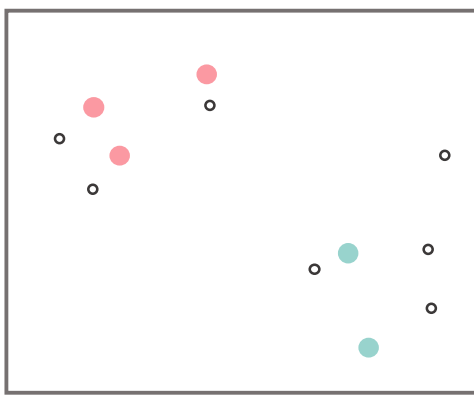
본 조사의 시사점



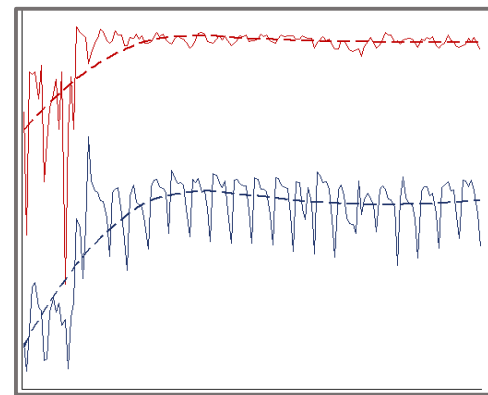
K-MEANS 클러스터링, FA(Factor Analysis), 시계열 FA 등 다양한 비지도학습 기법을 사용하여 분석을 진행함.



K-MEANS 클러스터링



FA(Factor Analysis)



Time Series FA

인사이트 정리

한계 및 시사점

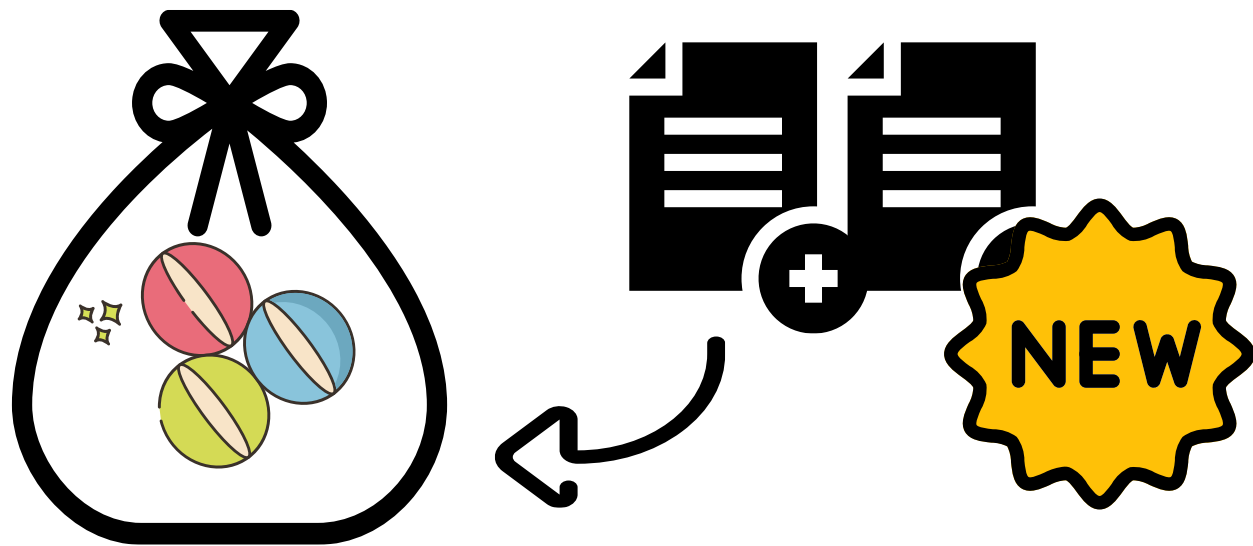


한계 및 시사점

시사점③: 기본 제공된 데이터 뿐 아니라, 아모레퍼시픽 산업보고서 등 다양한 데이터를 활용하여 분석을 진행함

본 조사의 시사점

↳ 기본 제공된 데이터 뿐 아니라, 아모레퍼시픽 산업보고서 등 다양한 데이터를 사용하여 분석을 진행함.



인사이트 정리

한계 및 시사점

감사합니다
THANK YOU

