



# COVID19의 출현 이후 변화한 업종과 업종의 전망 예상

-K-Shape Clustering을 중심으로-

담당교수 김 규 성

서울시립대학교 통계학과 2018580008

문혜현

2020년 12월

## 목차

1. 서론 .....	2
1.1 연구 목적 .....	2
1.2 문헌 연구 .....	2
1.3 데이터 설명 .....	3
1.4 분석 방법 .....	3
1.5 결과 활용 및 기대 효과 .....	5
2. 본론 .....	6
2.1 분석 과정 소개 .....	6
2.2 데이터 분석 및 결과 설명 .....	6
2.2.1 이동평균을 이용한 추세 추출 .....	6
2.2.2 K-shape clustering .....	7
2.2.3 결과 설명 .....	8
2.3 분석의 타당성 설명 .....	15
2.3.1 유클리드 거리측정과 DTW 거리측정 .....	15
2.3.2 계층 군집분석과 비교 .....	16
2.3.3 k-means 군집분석과 비교 .....	17
3. 결론 .....	19
3.1 분석 결과 요약 .....	19
3.2 분석의 장점 및 한계점 설명 .....	21
4. 참고자료 .....	22

# 1. 서론

## 1.1 연구 목적

2020년은 COVID-19의 출현으로 많은 변화가 있었다. 이에 코로나 종식 이후 경제 및 생활에 다가올 변화 및 시기를 뜻하는 “포스트 코로나”라는 단어도 생겨났다. 많은 기업들이 코로나로 인해 변화한 시대에 맞추어 미래를 내다보고 있다. 그렇다면 우리는 코로나가 무엇을 변화시켰는지 그 변화가 얼마나 지속될지 등을 정확히 알아야 할 것이다. 따라서, 장기적인 관점에서 코로나가 여러 산업 군의 업종들을 어떻게 변화시켰는지 살펴보고 이에 근거하여 앞으로의 각 업종의 전망을 예상해본다.

## 1.2 문헌 연구

삼성KPMG경제연구원(2020)은 코로나19의 영향으로 국내 주요 산업이 어떻게 변화했는지를 폭넓게 다루고 있다. 이 보고서에 따르면 금융, 자동차, 휴대전화, 디스플레이, 항공, 해운, 호텔, 패션 및 의류 부분은 코로나로 인해 부정적 영향을 받았다고 한다. 반도체, 건설, 정유 및 석유화학, 철강, 유통, 화장품, 식음료, 영화 및 영상미디어는 일부 부정적이며 게임이 일부 긍정적이라고 평가했다.<sup>1</sup> 하지만, 보고서는 3월까지의 지표들을 이용하여 분석한 결과이며 코로나가 장기화된 지금 각 산업이 회복했는지, 오히려 성장했는지에 대한 후속 연구가 필요하다.

마찬가지로 이제까지 나온 연구들은 2, 3월의 코로나가 막 출현하고 위험이 커졌을 때 어떤 업종이 코로나로 인해 성장했고, 부진했는지에 대해 논하고 있다. 이번 연구는 앞서 나온 연구들을 참고해서 보다 장기적인 코로나 및 코로나로 인해 생긴 여러 환경요인과 업종의 관계를 살펴볼 것이다. 이를 통해 앞으로의 코로나 상황에 따른 업종의 미래를 미리 예측한다는 점에서 의의가 있다.

---

<sup>1</sup> 삼성KPMG경제연구원, “코로나19에 따른 산업별 영향 분석”

### 1.3 데이터 설명

KT 빅데이터플랫폼의 업종 별 카드 결제금액 데이터를 사용한다. 3713952 개의 행과 7 개의 변수를 가지고 있다. 변수는 아래와 같다.

- Receipt\_dttm : 카드회사가 카드사용내역을 접수한 일자
- Abstrd\_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

실제 분석에 사용하기 위해 2020 년 1 월 4 일부터 2020 년 6 월 14 일까지 중 결측치가 다수 포함되어 있는 1 월 4 일부터 2 월 5 일까지를 제거하였다. 1 월 20 일 첫 확진자가 발생해 2 월 5 일까지 누적 확진자가 16 명으로 코로나의 영향이 크지 않았던 점을 고려하여 연구에 영향을 주지 않는다. 또한, 결측치로 처리된 업종과 해석이 모호할 수 있는 기타대인서비스, 기타수리서비스 등의 기타제품들을 제외하고 117 개의 업종에 대해서 분석을 하였다.

### 1.4 분석 방법

#### (1) 단순 이동평균(Simple Moving Average)

이동평균은 전체 데이터 집합의 여러 부분 집합에 대한 일련의 평균을 만들어 데이터 요소를 분석하는 계산법이다. 일정 크기의 부분 집합을 이동시키며 연속적인 평균값을 산출함으로써 일정 기간 혹은 데이터 구간의 평균의 흐름을 알 수 있게 해준다. 이 흐름을 이은 선을 이동평균선이라 하며 추세 파악과 변화를 파악하는 데 용이하다. 아래는 단순 이동평균의 공식이다. ( $n$ 은 부분집합의 크기이며  $P_{t-i}$ 는 데이터 값을 의미한다.)

$$MA = \frac{\sum_{i=0}^{n-1} P_{t-i}}{n}$$

## (2) 군집분석(Cluster Analysis)

군집분석(Cluster analysis)은 개별 군집에 대한 사전 정의가 없는 상황에서 데이터객체들을 유사한 그룹으로 구분해 가는 탐색적 통계분석법을 말한다. 비지도학습 기법으로 유사한 데이터들은 동질이거나 연관성이 높은 그룹으로 배치되며, 결과적으로 하나의 군집에 속하는 데이터객체 간에는 서로 다른 군집 내의 객체들과는 구분되는 유사성이 존재하게 된다. 군집 분석은 집단 구조를 파악하기 위한 것으로 난해한 데이터의 복잡한 관계를 쉽게 이해하는 데 도움이 된다. 다시 말해, 개별의 관측치 또는 데이터객체들을 개별적으로 파악하기보다 객체들을 군집화함으로써 각 데이터객체를 잘 대표하는 경향이 나 특성을 파악할 수 있다는 데 의미를 가진다.

군집의 구조적 종류에 따라 크게 나누면 계층적 군집화(hierarchical clustering)와 비계층적 군집화(non-hierarchical clustering)로 구분할 수 있다. 먼저, 계층적 군집은 가장 유사한 개체를 묶어 나가는 과정을 반복하여 계층적으로 군집을 형성하는 방법이다. 사전에 군집 수에 대한 예상이 필요하지 않는다는 점에서 비계층적 기법과 차이를 보인다.

비계층적 군집화 기법은 사전에 군집의 크기와 수 등 여러 기준을 설정한 후, 그 기준에 따라 개체를 적합한 군집으로 재배치한다. 비계층적 군집화의 대표적인 형태인 분할적(partitional) 알고리즘에서는 사전에 예상된 군집 수에 따라 최적화된 형태로 데이터를 나누고, 분할된 각 군집의 중심이 되는 데이터 값을 설정하여 그것과 유사성이 높은 개체들을 각 군집에 배정하는 과정을 따라 군집화가 이루어진다.

## (3) 유클리드 거리

유클리드 거리(Euclidean distance)는 두 점 사이의 거리를 계산할 때 쓰는 방법이다. 직교 좌표계로 나타낸 점  $p = (p_1, p_2, \dots, p_n)$  와  $q = (q_1, q_2, \dots, q_n)$ 가 있을 때, 두 유클리드 노름을 이용하여 두 점 p, q의 거리를 계산하면 다음과 같다.

$$\|p - q\| = \sqrt{(p - q) * (p - q)}$$

## (4) DTW

DTW(Dynamic time warping)는 데이터 마이닝 영역에서 유클리드 거리와 관련된 한계를 극복하기 위해 논의된 아이디어로, 두 개의 시계열을 비교하여 단조성(monotonicity)과 같은 특정 제약 조건에 따른 최적의 뒤틀림(warping) 경로를 찾기 위해 개발된 동적 프

로그래밍 알고리즘이다. DTW 알고리즘은 두 시계열  $x, y$  사이에서 정렬을 최소화하는 경로를 찾고, 그 비용을 합산하는 과정을 통해 거리를 계산한다. 두 시계열 자료  $X, Y$ 와 가능한 모든 정렬(alignments)의 집합  $A(x, y)$ 에 대하여 기존의 DTW 거리는 다음과 같이 정의된다(이석준, 2019)<sup>2</sup>. 결과적으로 각 지점과 가장 유사한 위상의 점을 찾게 된다.

$$S(\pi) = \sum_{i=1}^{|\pi|} \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})$$

$\varphi : \text{arbitrary conditionally positive - definite kernel}$

## (5) K-shape Clustering

K-shape Clustering은 Paparrizos and Gravano (2015)에 의해 개발된 알고리즘이다. 비계층적 군집화 기법으로 DTW의 변형인 SBD(Shape-based distance)와 shape extraction centroid 함수를 사용한다. R의 dtwclust 패키지의 tsclust 함수로 구현할 수 있으며 표준 정규화가 이후에 이루어진다. 더 자세한 사항은 Paparrizos and Gravano (2015)를 참고한다(Alexis Sarda-Espinosa, 2019)<sup>3</sup>.

## 1.5 결과 활용 및 기대 효과

영국의 철학자 프랜시스 베이컨이 “아는 것이 힘이다”라고 말했듯이 급변하는 21세기 현대사회 속에서 아는 것, 정보력은 매우 중요한 자산이다. 특히 미래를 아는 것은 계속해서 변화하는 사회에서 살아남기 위한 가장 중요한 무기이다.

본 연구에서 COVID19로 인한 업종의 변화추세를 성장, 회복, 쇠퇴, 기타 네 가지로 분류한 결과를 통해 미래의 코로나의 계속된 횡행 또는 종식 이후의 유망한 업종과 회복 및 쇠퇴할 업종들을 예측할 수 있다. 또한, COVID19와 관련된 환경요인이 업종에 주는 영향을 파악한 결과가 예측을 더 정밀하게 만들것이다.

개인은 창업, 주식 투자, 직업 등을 계획하는 데 본 연구에서 얻은 정보를 참고할 수

---

<sup>2</sup> 이석준, “데이터마이닝을 통한 주택 하위시장 구분 및 주택가격 예측”

<sup>3</sup> Alexis Sarda-Espinosa, “Comparing Time-Series Clustering algorithms in R Using the dtwclust Package

있다. 비단 개인뿐만 아니라 본 연구를 이용해 기업은 미래의 사업 계획, 마케팅 방향 등을 결정하고, 정부는 법안 발의, 정책 방향성 등을 도모하는 데 사용될 수 있다.

## 2. 본론

### 2.1 분석 과정 소개

우선, 각 업종 별 일별 매출액에서 이동평균을 이용한 추세 요소를 추출합니다. 추세 요소를 추출하는 이유는 시기에 따라서 업종이 어떤 추이를 보이는지 더 확실히 알기 위해서입니다.

두 번째로 시계열 데이터의 특성을 고려하여 k-shape clustering을 사용합니다. 일별 매출액의 추이가 비슷한 업종끼리 분류합니다.

분류한 결과를 이용해 각 군집이 어떤 추이를 가지는지와 코로나와 관련된 중심사건을 기준으로 어떤 연관이 있는지 분석합니다. 그리고 코로나가 장기화되면서 매출 추이가 원상복귀 되었는지 아니면 더 높아지거나 더 낮아지는지 추이를 통해 군집을 성장업종, 회복업종, 쇠퇴업종 세 가지로 분류합니다. 이를 통해 포스트 코로나 시대의 각 업종의 성장성을 구별할 수 있습니다.

### 2.2 데이터 분석 및 결과 설명

#### 2.2.1 이동평균을 이용한 추세 추출

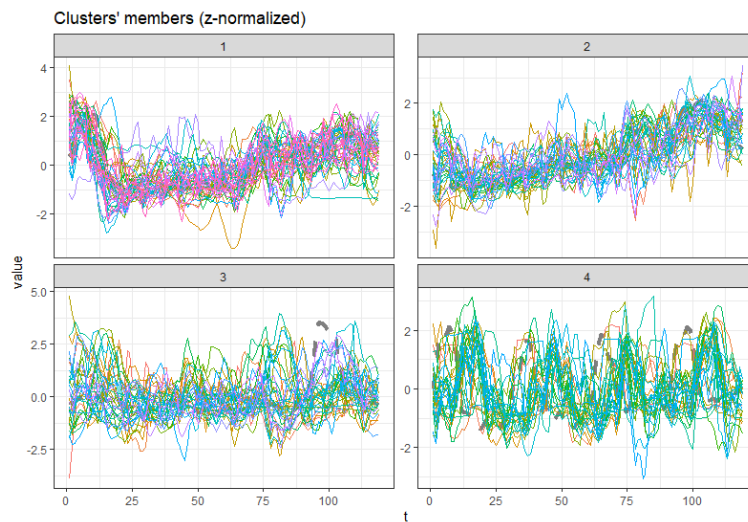
코로나 확진자는 집단 감염 사태가 발생하는 시점을 기준으로 급격히 높아지는 경향이 있다. 따라서 단기적인 변동이 심하므로 코로나 심각성이 업종에 주는 영향을 더 잘 관측하기 위해서 일별 단위로 업종 매출금액을 합하였다. 또한, 일별 단위이기 때문에 추가적인 조치 없이는 불규칙성이 많아 업종의 매출추이를 파악하기 어렵다.

따라서 데이터의 무작위성을 제거하고 추세를 추출하기 위해 이동평균을 사용하였다. 일주일 단위로 주기성을 보이는 데이터들이 많음과 최신 데이터에 가중을 두기 위해 짝수 차수 6의 중심화된 이동평균을 사용하였다. 2\*6 MA를 사용하였고, 이를 구현하기 위해 R의 MA함수를 이용하였다.

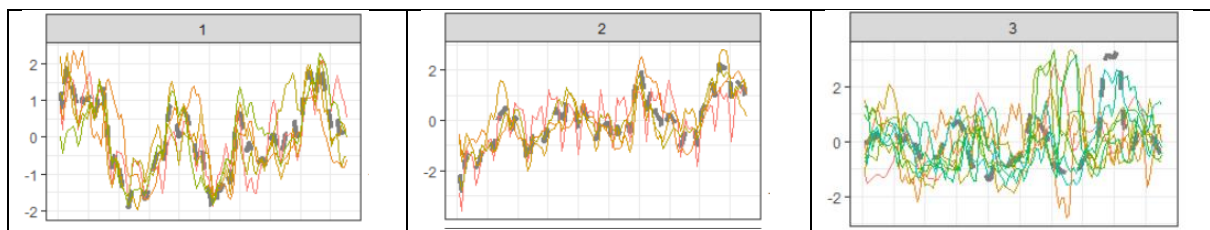
### 2.2.2 K-shape clustering

업종을 성장업종, 회복업종, 쇠퇴업종, 기타업종으로 크게 네 가지로 분류한다. 성장업종은 코로나 시기를 거치면서도 꾸준히 매출성장을 이룬 업종이다. 회복업종은 코로나가 심각할 때는 매출이 정체되었으나 그 이후 코로나가 완화되며 매출이 다시 상승하는 업종이다. 쇠퇴업종은 코로나 출현 이후 지속적인 매출하락세를 보이거나 매출의 회복속도가 느린 업종이다. 기타업종은 코로나와 매출이 관련이 없거나 추세가 뚜렷하지 않은 업종을 이른다.

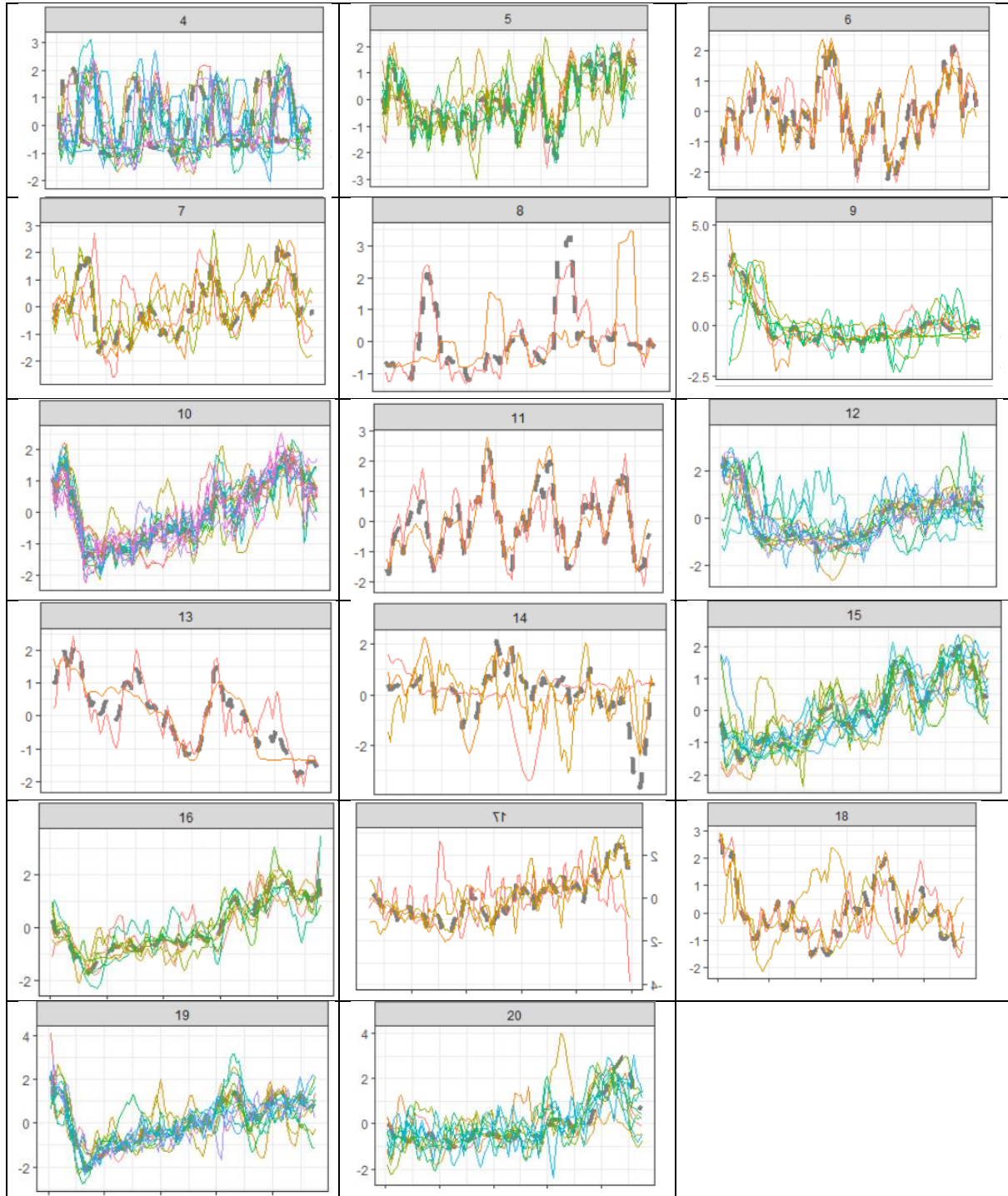
업종을 4가지로 분류함에 따라 K=4로 하여 업종을 분류할 경우 업종의 매우 많은 패턴을 과도하게 일반화해 전혀 다른 추이의 업종들이 하나의 군집으로 묶이는 것을 확인했다.



따라서 K=20으로 하여 추이는 같더라도 코로나의 심각성, 재난지원금 등의 환경적 영향에 따라서 세세하게 추이가 변경되는 것이 반영되도록 하였다.





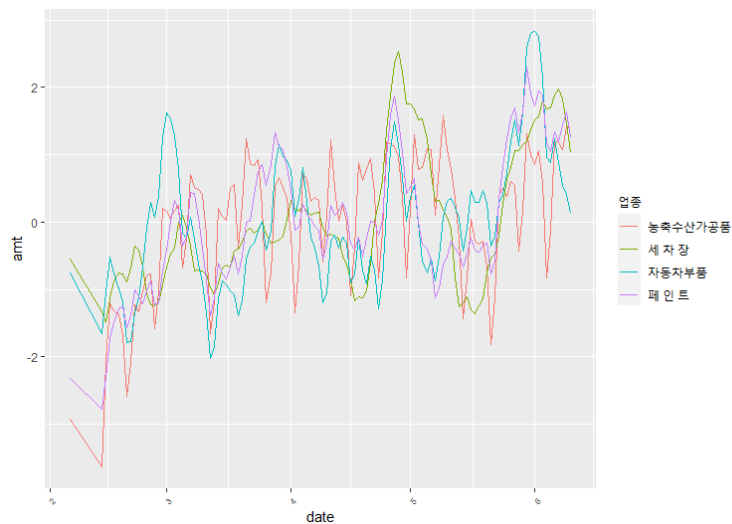


### 2.2.3 결과 설명

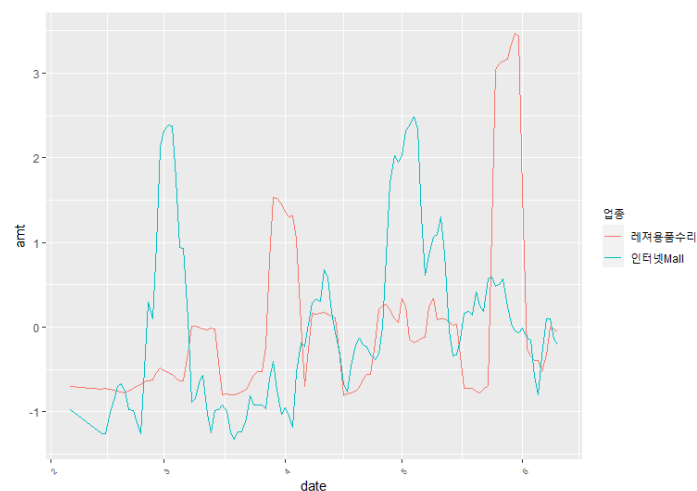
K=20으로 k-shape clustering 한 결과 중 성장추이를 보이는 군집이 2, 8, 15, 16이고, 회복추이를 보이는 군집이 5, 10, 19이다. 그리고 회복되지 못하거나 계속해서 매출금액이 낮아지는 군집은 9, 12, 13, 17이다. 군집별로 살펴보며 어떤 특성이 있는지 살펴보도록 한다.

### (1) 성장 업종

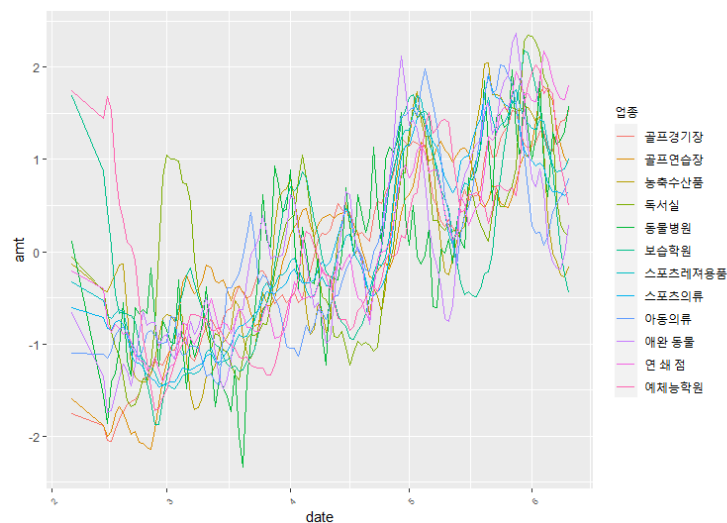
2번 군집은 세차장, 자동차부품, 페인트, 농축수산가공품을 포함한다. 2월 신천지 집단 감염과 5월 초에 발생한 이태원 집단감염 사건으로 확진자가 급격히 늘어났을 때, 매출이 줄어들지만 그 이후 빠르게 회복이 되며 오히려 이전보다 더 매출이 늘어나는 것을 확인할 수 있다. 이는 코로나 감염을 막기 위해 개인 차를 이용하는 것을 권장한 것이 영향을 주었다고 추측할 수 있다.



8번 군집은 레저용품수리와 인터넷MALL을 포함한다. 인터넷MALL은 코로나의 영향을 크게 받으면서 계속해서 성장하는 모양을 보이고 있다. 2월 말의 신천지 집단감염과 5월 이태원 집단감염의 시기와 비슷하게 급격히 매출금액이 높아짐을 볼 수 있다. 레저용품수리는 코로나가 완화되었을 때, 그래프가 높아지는 것을 통해서 사람들이 야외활동에 대한 욕구를 충족시키기 위해 구매를 하는 것으로 생각할 수도 있지만 6월 이후 다시 매출이 정상화되는 것을 통해 코로나의 영향이 아닌 단순히 날씨가 따뜻해지면서 야외활동을 하는 경우도 생각해 보아야 한다.



15번 군집은 골프경기장, 골프연습장, 농축수산물, 독서실, 동물병원, 보습학원, 스포츠레저용품, 스포츠의류, 아동의류, 애완동물, 인쇄점, 예체능학원을 포함한다. 한 달 또는 일주일의 주기성이 뚜렷하게 나타난다는 점을 감안하고 보았을 때, 매출금액이 2월부터 점점 성장하는 모습이 보인다. 특히, 8번 군집에서와 비슷하게 스포츠레저용품, 스포츠의류의 판매가 늘어났고, 집단감염의 우려에서 비교적 자유로울 수 있는 골프경기장과 골프연습장의 매출이 늘어나는 점이 두드러진다.



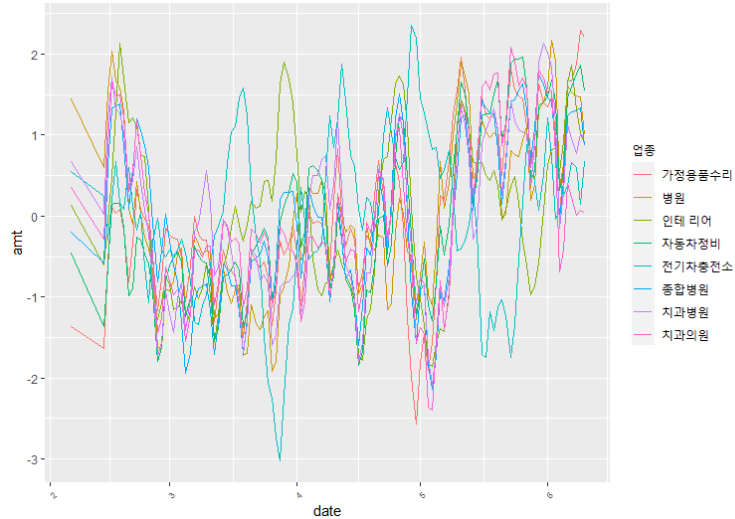
16번 군집은 가전제품 내의판매점, 스크린골프, 신발, 이용원, 일반 가구, 정장, 캐주얼 의류를 포함한다. 15번 군집에서와 같이 개인이 즐길 수 있는 스크린골프의 매출이 증가한 것이 눈에 띈다. 또한, 2월의 매출과 비교하였을 때, 가장 매출이 높아진 제품으로 가전제품과 일반가구를 확인할 수 있다. 집에서 계속 생활하며 집의 인테리어에 관심을 가지게 된 사람들이 많아졌음을 예상할 수 있다.

내의판매점, 신발, 정장, 캐주얼 등의 의류 역시 꾸준히 매출이 증가하지만 겨울에서 봄과 여름으로 넘어가는 시즌이기 때문에 이러한 형태를 띄는 것인지 아니면 비교적 인터넷시장이 잘 형성되어 있는 의류계에 소비가 쏠린 것인지 추가적인 조사가 필요하다.

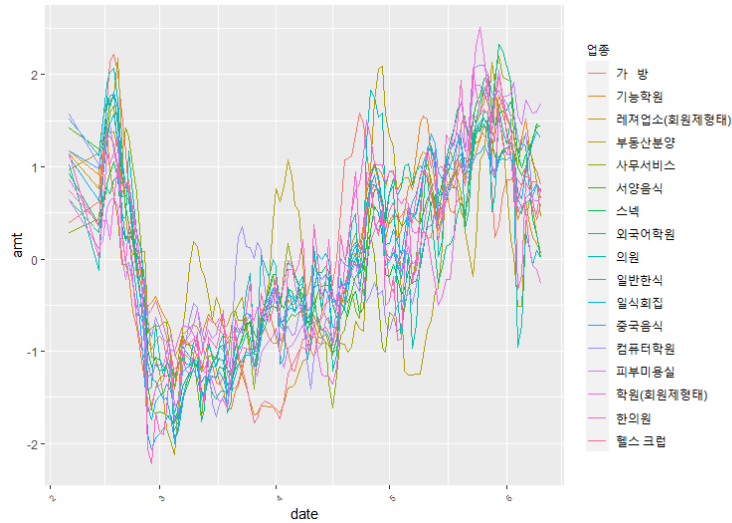


## (2) 회복 업종

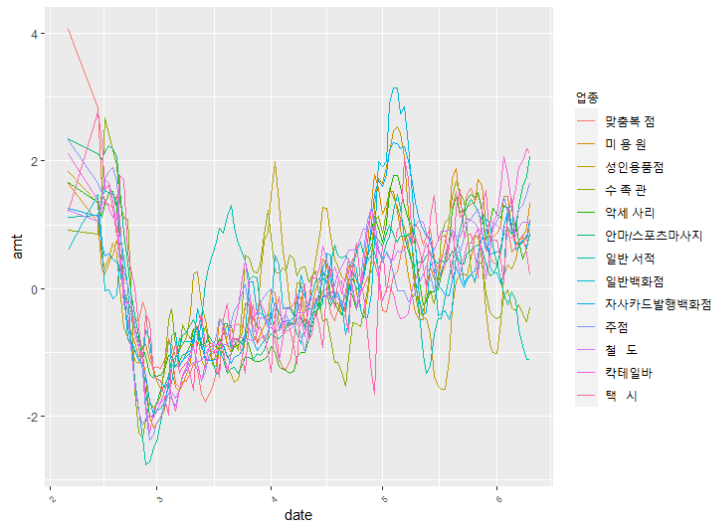
5번 군집은 가정용품수리, 병원, 인테리어, 자동차정비, 전기차충전소, 종합병원, 치과병원, 치과의를원을 포함한다. 전체적으로 2월 말의 신천지 집단감염과 5월 이태원 집단감염 시기에 공통적으로 매출이 하락했지만 금세 매출을 회복하는 것으로 보인다.



10번 군집은 가방, 기능학원, 레저업소(회원제형태), 부동산분양, 사무서비스, 서양음식, 스낵, 외국어학원, 의원, 일반한식, 일식회집, 중국음식, 컴퓨터학원, 피부미용실, 학원(회원제형태), 한의원 헬스클럽을 포함한다. 다양한 업종을 포함하는데 요약하면 학원, 음식점, 병원, 뷰티 및 의류, 헬스이다. 2월 말 코로나 확진자가 급증할 때 큰 폭으로 매출이 하락했다가 세세한 차이는 있지만 다시 정상적인 매출의 궤도로 서서히 올라오는 것을 볼 수 있다.



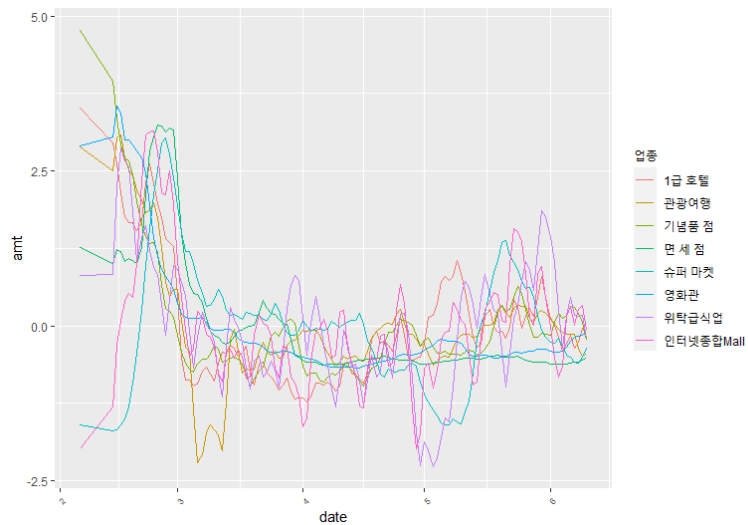
19번 군집은 맞춤복점, 미용원, 성인용품점, 수족관, 액세서리, 안마/스포츠마사지, 일반 백화점, 자사카드발행백화점, 주점, 철도, 칵테일바, 택시를 포함한다. 요약하면 뷰티 및 의류, 주점, 개인적으로 이동할 수 있는 택시로 볼 수 있다. 2월 확진자의 급증으로 큰 폭으로 매출이 하락했지만 서서히 매출이 올라가는 추세이며 다만, 10번 군집에 비해 회복 속도가 느리다. 5월에 거의 모든 업종의 매출이 잠깐 올라가는데 5월에 발행된 재난지원금의 영향으로 보인다.



### (3) 쇠퇴 업종

9번 군집은 1급 호텔, 관광여행, 기념품점, 면세점, 슈퍼 마켓, 영화관, 위탁급식업, 인터넷종합Mall을 포함한다. 코로나로 인해 여러 항공편이 막히고 여행에 제약을 주는 자가 격리기간이 생겨나면서 급격한 해외 관광객의 하락과 집 밖의 생활을 자제시키는 캠페인으로 국내여행 관광객의 하락이 영향을 주었을 것이다. 또한, 학교의 원격수업과 밀폐공간인 영화관에서 매출이 낮아진 것도 당연하다. 다만, 추이가 다른 두 가지 업종이

있다. 슈퍼마켓과 인터넷종합몰로 2월 집단감염 사태 이후 오히려 올랐으며 그 이후로도 2월 초에 비해 높은 수준을 유지하고 있다.



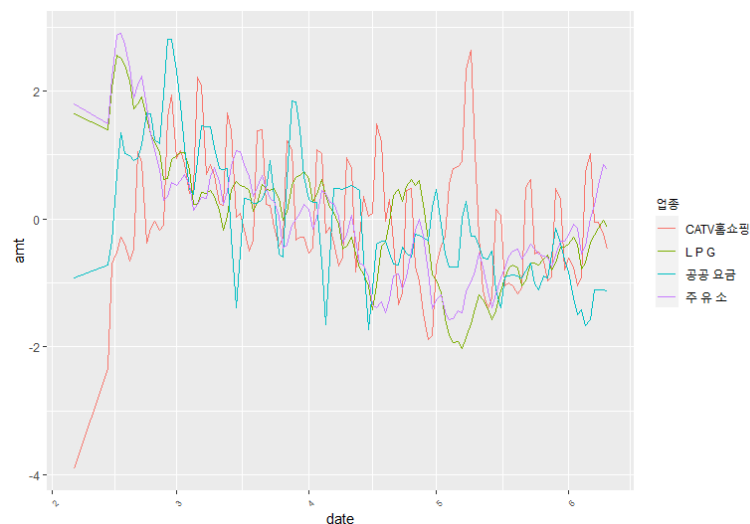
12번 군집은 2급 호텔, 고속버스, 노래방, 문화취미/기타, 사진관, 수영장, 유리, 종합레저타운, 철제 가구, 특급 호텔, 티켓, 한방병원, 한정식, 항공사를 포함한다. 주로 감염되기 쉬운 공간이거나 관광과 관련되어 있는 항목들이다. 전체적으로 2월 이후 급격한 매출하락이 생기고, 점차 회복하는 모습이 보이지만 완전히 회복하기에는 많은 시간이 걸릴 것으로 예상된다. 다만, 이 군집에서도 전체적인 추이와 다른 양상을 띄는 추이들이 4개 발견되는데 사진관, 유리, 철제 가구, 종합레저타운이 이에 해당하므로 해석에 주의해야 한다.



13번 군집은 유류판매와 유휴주점이다. 유류는 전 세계 석유제품 수요가 가파르게 감소하면서 정제 이후 제품 가격이 원재료 가격보다 낮게 거래되는 경향이 반영되어 있다. 유휴주점은 감염이 쉽게 일어날 수 있는 공간이며 5월의 재난지원금이 발행되었을 때 한시적으로 매출이 늘어난 것으로 보아 코로나로 인한 경제 악화와도 연관이 있어 보인다.



17번 군집은 CATV 홈쇼핑, LPG, 공공요금, 주유소를 포함한다. LPG와 주유소는 코로나의 영향보다는 유류의 하락으로 인한 영향이 큰 것으로 보인다. CATV홈쇼핑과 공공요금은 다른 두 업종과 상이한 결과를 가져옴을 알 수 있다. 코로나 출현 이후 오히려 매출이 성장했으며 특히 CATV 홈쇼핑이 집에서 생활하는 시간이 많아지고 온라인과 같이 배달로 상품을 살 수 있다는 점에서 코로나가 장기화되면서도 꾸준히 매출이 유지 또는 상승세이다.

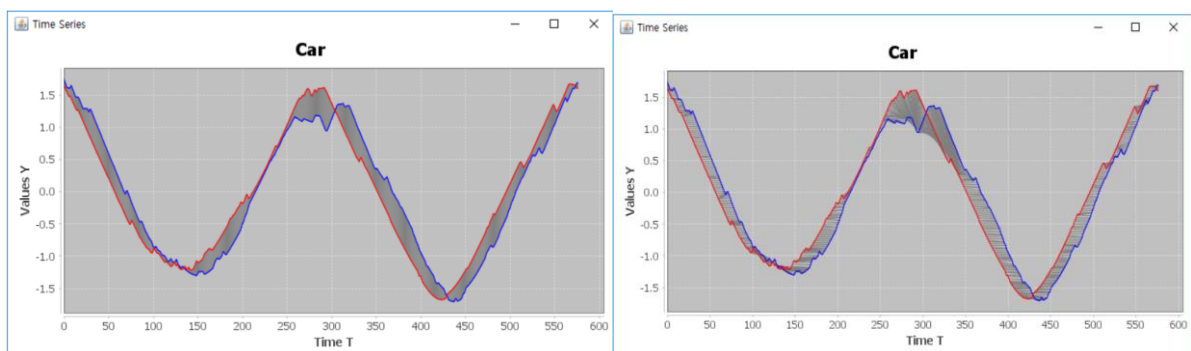


## 2.3분석의 타당성 설명

군집분석은 다양한 유형과 다양한 조건들을 이용해 군집을 만들고 분석할 수 있다. 또한, 회귀분석과 같이 어떤 방법이 더 낫고 그른지를 판단할 수 있는 지표가 명확하지 않다. 따라서 본 연구에서 사용한 k-shape 군집분석이 “다변량통계학” 강의에서 배운 계층적 군집분석과 k-means 군집분석에 비해 연구목적과 데이터에 더 적합하다는 것을 보임을 통해 분석의 타당성을 입증하려 한다.

### 2.3.1 유클리드 거리측정과 DTW 거리측정

시계열 데이터에서 서로 다른 추세를 비교하기 위한 유사도 측정으로 다양한 방식을 사용하고 있지만 여기서는 유클리드 거리측정과 DTW 거리측정을 비교하겠다. 두 방식의 거리측정방식의 차이는 다음과 같다. 시계열 데이터에서의 유클리드 거리측정은 각 시점에서 두 값들의 차이를 계산한다. 하지만, DTW 거리측정은 비슷한 위상에 있는 점들 간의 차이를 이용해 계산한다. 두 개념에 대한 자세한 설명은 1.4를 참고하길 바란다. 그렇다면 두 측정방식 중 어느 것이 어떤 시계열 데이터에 좀 더 적합할까? 아래 그림의 예를 보면 서로 어느 정도 비슷한 시계열 사이에서는 유클리드 거리측정을 적용할 때 보다 DTW 거리측정을 적용했을 경우가 어느 정도의 왜곡에 대해 수용해주는 것을 볼 수 있다. 이처럼 DTW는 시간적으로 어느 정도 밀려 있거나 앞서갈 때도 적절한 위치를 매칭시켜 비교해줌으로써 두 추세 간의 유사도를 유클리드 거리측정보다 좀 더 유연하게 비교해 줄 수 있을 것이다.



왼쪽 그림 : 유클리드 거리측정방식, 오른쪽 그림 : DTW 거리측정방식

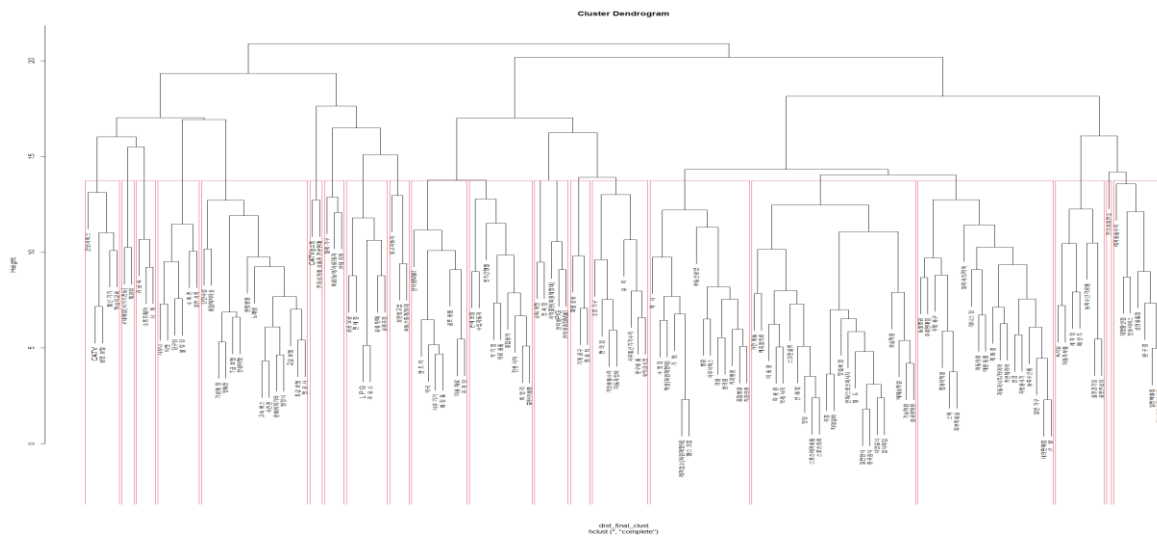
본 연구에서 사용한 업종별 카드 데이터는 시계열이며 업종에 따라 그 추세가 모두 다르다. 업종마다 영향을 받는 환경요인, 영향이 나타나고 사라지는 기간 등 비슷하거나 매우 다를 수도 있다. 이에 업종의 변화를 확진자 급증, 사회적 거리두기, 재난지원금 등과 같은 환경요인에 비추어 살펴보는 것이 본 연구의 목적이므로 환경요인으로부터 영향이



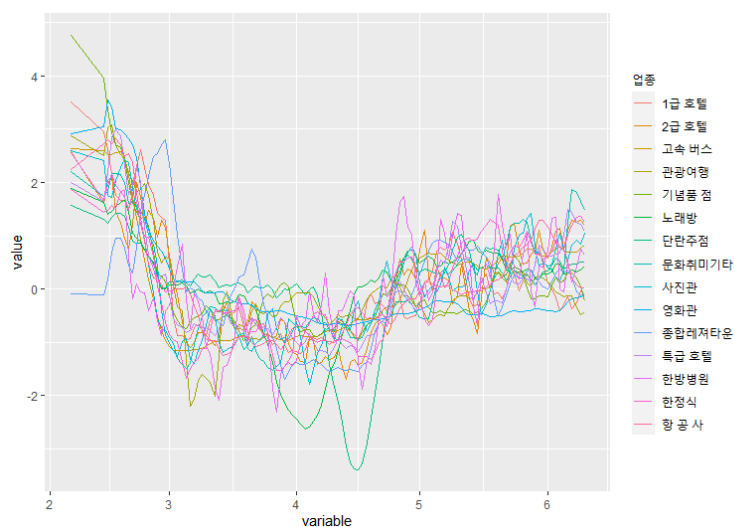
나타나기까지의 기간이 업종마다 차이가 날 수 있다는 점을 보완할 수 있는 DTW 거리 측정방식을 채택한다.

### 2.3.2 계층 군집분석과 비교

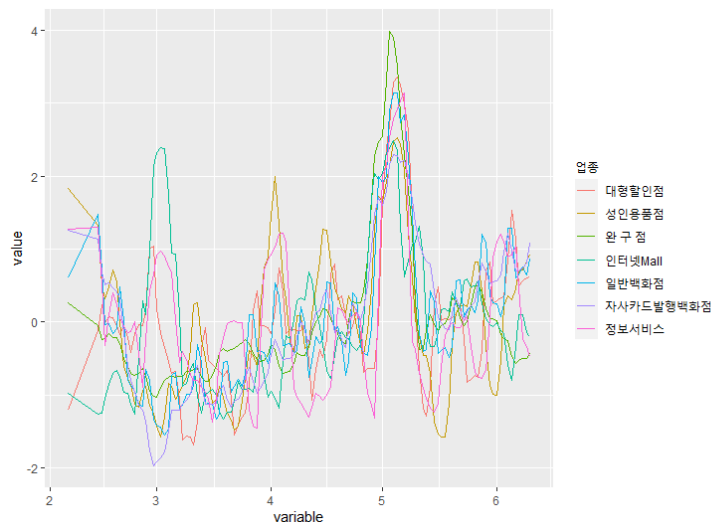
유클리드(Euclidean) 거리방식을 사용해 최장연결법(Complete linkage)으로 군집화를 한다. K-shape clustering과 마찬가지로 k=20으로 하여 군집을 생성한다.



계층 군집분석을 통해 나온 아래 군집은 2월에 급격한 매출 하락 이후 회복이 되지 않는 추이가 보인다. 또한, k-shape 군집분석에서 쇠퇴업종으로 분류된 9번과 12번의 업종들과 다수 겹치며 업종도 잘 분류된 것으로 판단된다.



하지만, 아래 그림을 보면 전체적으로 회복 추세이지만, 국지적으로 보면 다른 해석을 할 수 있는 그래프들이 함께 모인 것을 확인할 수 있다. 예를 들어, 인터넷Mall의 경우 코로나의 영향을 크게 받으며 계속해서 성장하지만, 다른 업종들은 코로나로 인해 급격한 하락양상을 보인다. 2월 말에 코로나로 인해 매출이 증가한 업종들도 있는 반면 매출이 하락한 업종들이 있어 구분이 필요하다.



### 2.3.3 k-means 군집분석과 비교

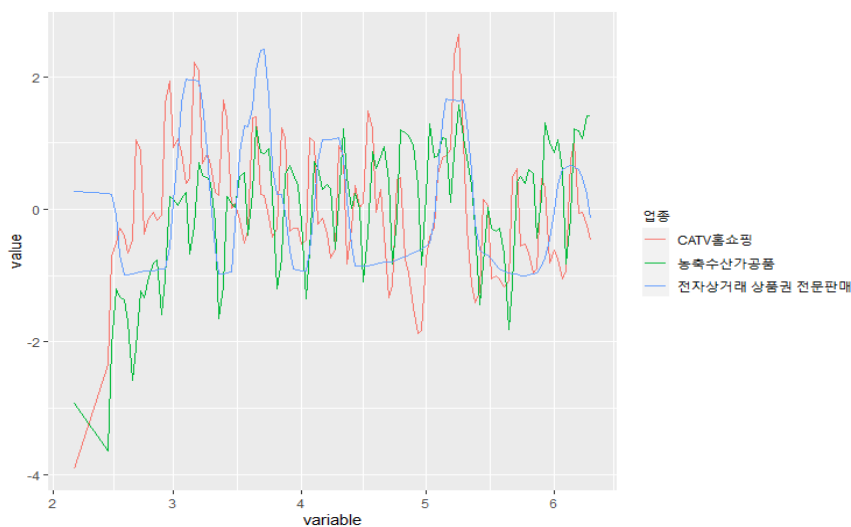
유클리드(Euclidean) 거리방식을 사용해 K-shape clustering과 마찬가지로  $k=20$ 으로 하여 군집을 생성한다.

k-means 군집분석을 통해 나온 아래 군집은 k-shape 군집분석에서 쇠퇴업종을 다수 포함한 군집과 계층 군집분석에서의 쇠퇴업종을 다수 포함한 군집과 비교했을 때 대부분의 업종이 비슷하지만 다른 것을 확인할 수 있다. 따라서 세 군집방식 모두 비교적 추이를 잘 분류하지만 거리측정 방식과 군집생성 방식에 따라 세세하게 다른 군집이 만들어지는 것을 알 수 있다.



K-shape 군집보다 k-means 군집에서는 명확한 추세를 볼 수 있을 것으로 기대된다. K-shape 군집은 시차를 허용하고 위상을 중심으로 분류하기 때문에 다소 복잡해 보일 수 있다. 하지만 k-means 군집은 군집의 중심과 유클리드 거리가 최소가 되도록 군집이 형성되어 각 시점에서 높은 매출은 높은 매출끼리 낮은 매출은 낮은 매출끼리 묶일 것이기 때문이다.

하지만, 아래 그림을 보면 유클리드 기반 k-means 방식의 문제점이 보인다. 전자상거래 상품권 전문판매 업종은 다른 업종과 전체적인 추이는 비슷하고 국지적으로 다른 양상을 띄고 있어 유클리드 거리상으로는 비슷하기 때문에 같은 군집으로 묶였다. 국지적으로 보면 전자상거래 상품권 전문판매 업종은 코로나와 큰 연관이 없어 보이지만 CATV 홈쇼핑과 농축수산물가공품은 2월의 확진자 증가와 5월의 재난지원금과 연관이 커 보인다.



### 3. 결론

#### 3.1 분석 결과 요약

구조가 복잡하고 변수가 많은 대용량 데이터, 시계열 데이터라는 특성과 코로나의 심각성, 재난지원금 등과 같은 환경요인에 따른 업종의 변화 및 추이 파악이란 연구목적에 고려하여 일반적인 유클리드 거리측정을 사용한 군집분석이 아닌 DTW 거리측정을 이용한 비계층적 방식인 k-shape 군집분석을 사용하였다.

k-shape 군집들을 성장 업종, 회복 업종, 쇠퇴 업종, 기타 업종 총 4가지로 분류하였고, 결과를 종합하면 다음과 같다. 성장업종은 2월 말의 확진자 급증 이후 매출이 하락했지만 이를 웃도는 성장이 있거나 계속해서 성장한 업종이다. 자동차 관련 제품, 가구, 인테리어, 의류, 인터넷 Mall/CATV 홈쇼핑 골프, 레저가 있다. 코로나 감염예방의 캠페인의 일환이었던 자차이용, 야외활동 자제, 밀집공간 피하기와 사회적 거리두기의 영향으로 성장한 업종으로 볼 수 있다. 현재 2020년 12월 서울시의 사회적 거리두기 2.5단계로 코로나가 장기화된 만큼 위에서 언급된 업종들은 계속해서 성장할 것이다. 특히 코로나가 심각해질 때 매출이 급성장하며 그 이외의 시기에도 성장을 유지하는 인터넷 Mall과 CATV 홈쇼핑의 전망이 매우 좋을 것으로 기대된다. 하지만, 의류와 레저는 겨울에서 봄, 여름으로 계절이 바뀌는 것에 대한 영향이 있을 수 있다.

회복업종은 코로나 확진자 급증 이후 매출이 급격히 하락했다가 다시 정상 궤도로 돌아온 업종을 이른다. 병원, 학원, 음식점, 뷰티, 헬스장, 택시가 이에 해당한다. 병원의 경우 매출의 하락이 있었으나 다른 업종에 비해 크지 않았다. 학원, 음식점의 회복이 빠른 것에 비해 뷰티, 택시는 회복이 비교적 느렸다. 코로나가 장기화될수록 큰 영향이 없을 것으로 예상되지만 재난지원금이 발급된 5월 잠깐 증가한 것을 보았을 때, 경제악화로 인한 사치품에 해당하는 뷰티와 택시 업종에 대해서 소비자들의 지출이 적어질 것을 우려해야 한다.

쇠퇴업종은 2월 말 이후 매출이 하락된 채로 유지되거나 정상 궤도로 회복되지 않은 업종을 말한다. 숙박, 관광, 면세점, 항공, 고속버스, 노래방, 수영장, 영화관을 포함한다. 코로나로 인해 많은 항공편이 닫히고 자가격리 의무시행으로 해외 관광객의 감소와 밀폐 공간에 대한 감염우려, 4월에 시행된 유흥업소집합금지 명령으로 인한 매출 하락으로 보인다. 코로나가 장기화됨에 따라 고속버스, 노래방, 수영장 영화관 업종은 회복추세를 이어나갈 것 같지만 관광과 관련된 업종들은 회복하기까지 많은 시간이 걸릴 것으로 예상된다.

[성장 업종]		
군집종류	업종	비고
2번 군집	<b>세차장, 자동차부품, 페인트, 농축수산물가공품</b>	코로나 감염을 막기 위해 개인 차를 이용하는 것을 권장한다.
8번 군집	<b>레저용품수리, 인터넷MALL</b>	코로나영향 크다.
15번 군집	<b>골프경기장, 골프연습장, 농축수산물, 독서실, 동물병원, 보습학원, 스포츠레저용품, 스포츠의류, 아동의류, 애완동물, 인쇄점, 예체능학원</b>	골프, 레저는 감염에 비교적 자유롭다.
16번 군집	<b>가전제품 내의판매점, 스크린골프, 신발, 이용원, 일반 가구, 정장, 캐주얼의류</b>	의류, 인테리어, 스크린골프 매출이 증가한다.

[회복 업종]		
군집종류	업종	비고
5번 군집	<b>가정용품수리, 병원, 인테리어, 자동차정비, 전기차충전소, 종합병원, 치과병원, 치과의원</b>	공통적으로 매출이 하락했지만 금세 매출을 회복한다.
10번 군집	<b>가방, 기능학원, 레저업소(회원제형태), 부동산분양, 사무서비스, 서양음식, 스넥, 외국어학원, 의원, 일반한식, 일식회집, 중국음식, 컴퓨터학원, 피부미용실, 학원(회원제형태), 한의원, 헬스클럽</b>	학원, 음식점, 병원, 뷰티 및 의류, 헬스 다시 정상적인 매출의 궤도로 서서히 올라온다.
19번 군집	<b>맞춤복점, 미용원, 성인용품점, 수족관, 액세서리, 안마/스포츠마사지, 일반백화점, 자사카드발행백화점, 주점, (철도), 칵테일바, 택시</b>	뷰티 및 의류, 주점, 택시 서서히 매출이 올라가는 추세이며 다만, 10번 군집에 비해 회복속도가 느리다.

[쇠퇴 업종]		
군집종류	업종	비고
9번	<b>1급 호텔, 관광여행, 기념품점, 면세점, 슈퍼마켓, 영화관, 위탁급식업, 인터넷종합 Mall</b>	외국관광객감소, 학교의 원격수업, 밀폐공간 단, 슈퍼마켓과 인터넷종합몰은 증가한다.

12번	<b>2급 호텔, 고속버스, 노래방, 문화취미/기타, 사진관, 수영장, 유리, 종합레저타운, 철제 가구, 특급 호텔, 티켓, 한방병원, 한정식, 항공사</b>	외국관광객 감소, 밀폐공간 점차 회복하지만 느리다. 단, 사진관, 유리, 철제 가구, 종합레저타운은 아니다.
13번	<b>유류판매, 유흥주점</b>	전 세계 석유제품 수요가 가파르게 감소한다.
17번	<b>CATV 홈쇼핑, LPG, 공공요금, 주유소</b>	코로나의 영향보다는 유류의 하락 영향이 크다. 단, CATV홈쇼핑은 매출성장이다.

### 3.2 분석의 장점 및 한계점 설명

많은 업종들에 대해 군집분석을 사용함으로 각 군집 속 업종들의 특성을 파악하고 데이터 전체에 대한 이해를 도울 수 있었다. 또한,  $K=20$ 으로 군집을 세분화시켜 같은 추이를 가지더라도 어떤 환경적 요인에 영향을 어떻게 받는지를 살피면서 앞으로의 환경 변화에 따른 업종의 미래를 예상하는 데 용이하게 하였다.

이동평균을 사용하여 추세선을 추출하였는데 주 단위의 주기성과 한 달 단위의 주기성이 추세선에 제거되지 않은 채 내포되어 있어 군집을 만들 때 추세만이 아닌 주기성도 포함되어 계산이 되었다.

2월 초부터 데이터가 존재하여 그 이전의 데이터와 2020년의 데이터를 비교할 수 없어 레저 또는 의류의 경우 봄, 여름과 같은 계절상의 이유로 인한 매출증가인지 아니면 코로나에 의한 매출증가인지를 확실히 알 수 없었다.

비계층 군집분석의 한계로서 군집분석의 개수가 정해지지 않았기 때문에 본 연구의 목적에 맞게 군집의 개수를 20으로 설정하였다. 하지만 본 연구와 다른 군집의 개수를 결정해 더 나은 분석 결과를 도출할 가능성이 있다.

## 4. 참고자료

데이터 출처 : KT빅데이터플랫폼

삼성KPMG경제연구원, "코로나19에 따른 산업별 영향 분석"

이석준, "데이터마이닝을 통한 주택 하위시장 구분 및 주택가격 예측"

Alexis Sarda-Espinosa, "Comparing Time-Series Clustering algorithms in R Using the dtwclust Package"

<https://m.blog.naver.com/PostView.nhn?blogId=rkdwnsdud555&logNo=221155705904&proxyReferer=https:%2F%2Fwww.google.com%2F>