



# 리뷰 데이터로 알아보는 “배달의 민족” 인식 변화 분석

2022.12

2010년, 국내에서 세계 최초의 배달 어플리케이션 ‘배달통’이 탄생했다.  
같은 해에 ‘배달의민족’과 2년 뒤 ‘요기요’ 등이 뛰어들면서  
음식 배달 산업의 시장은 엄청난 성장세를 보였다.

하지만 시장을 개척했던 선두주자들의 위치가 예전만 하지 못하다.

엄청난 자금력과 추진력을 갖춘 ‘쿠팡이츠’의 등장과 함께  
‘위메프오’, ‘티몬’ 등의 새로운 어플리케이션들이 상위권 자리를 위협하고 있다.

10년 동안 부동의 1위 자리를 지켜온 ‘배달의민족’을 위협하게 되는 요소가 무엇이 있는지

시장통계자료를 기반으로 시장의 흐름 변화를 파악한 후,  
구글 플레이스토어&앱 스토어에 작성된  
소비자들의 리뷰 데이터를 수집, 분석하여  
‘배달의민족’에 대한 인식 변화를 파악하려 한다.



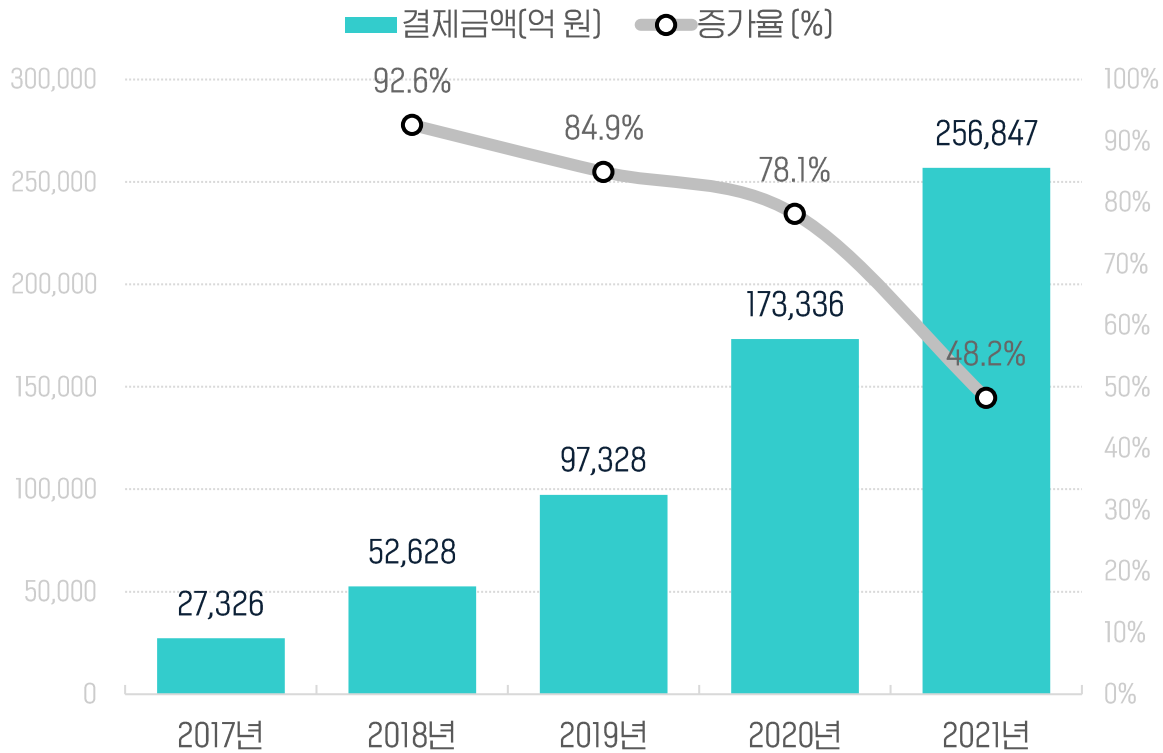
# CHAPTER.1

## 시장통계자료

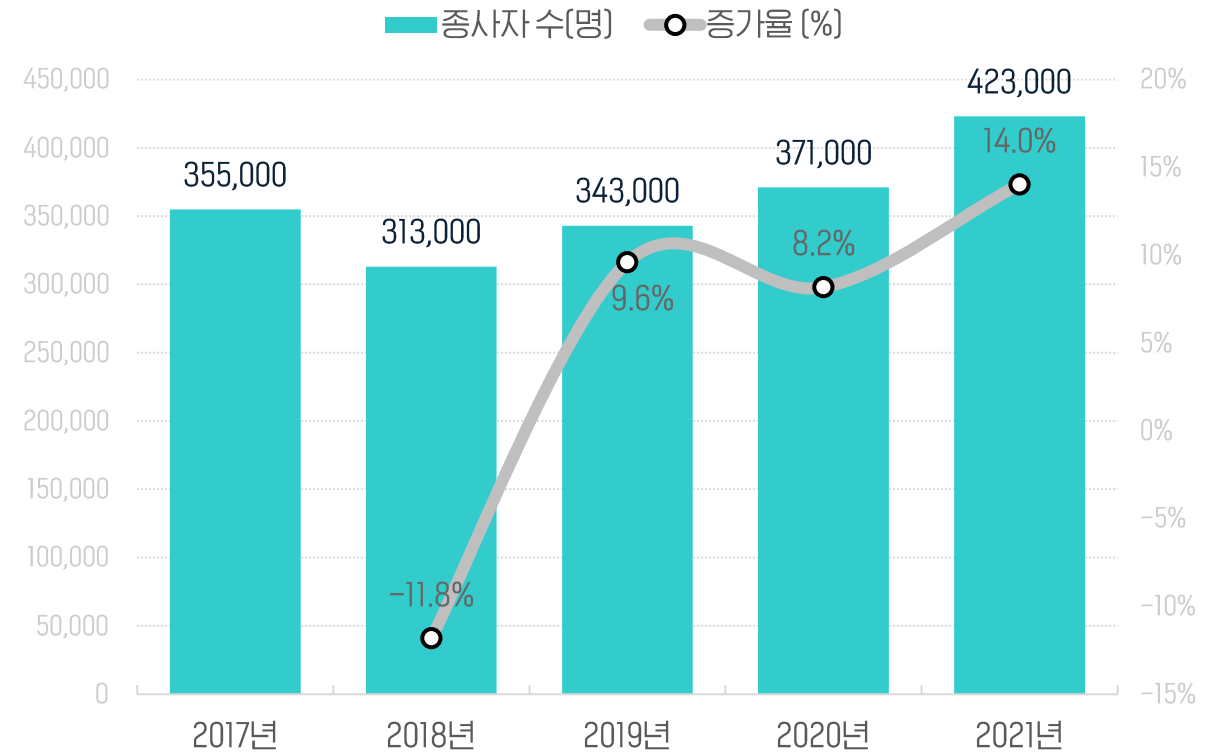
# 음식 배달 산업은 연평균 75.1%의 증가율을 기록 배달업 종사자는 40만명 이상을 돌파

- 연평균 75.1%의 증가율을 기록할 만큼 꾸준한 증가세를 보이며, 특히 코로나로 인하여 최근 2년간 연평균 증가율은 163.9%를 기록
- 배달업에 종사하는 수가 42만명을 넘어서며 최근 3년 사이에 35.1%가 증가

국내 음식 배달 산업 현황



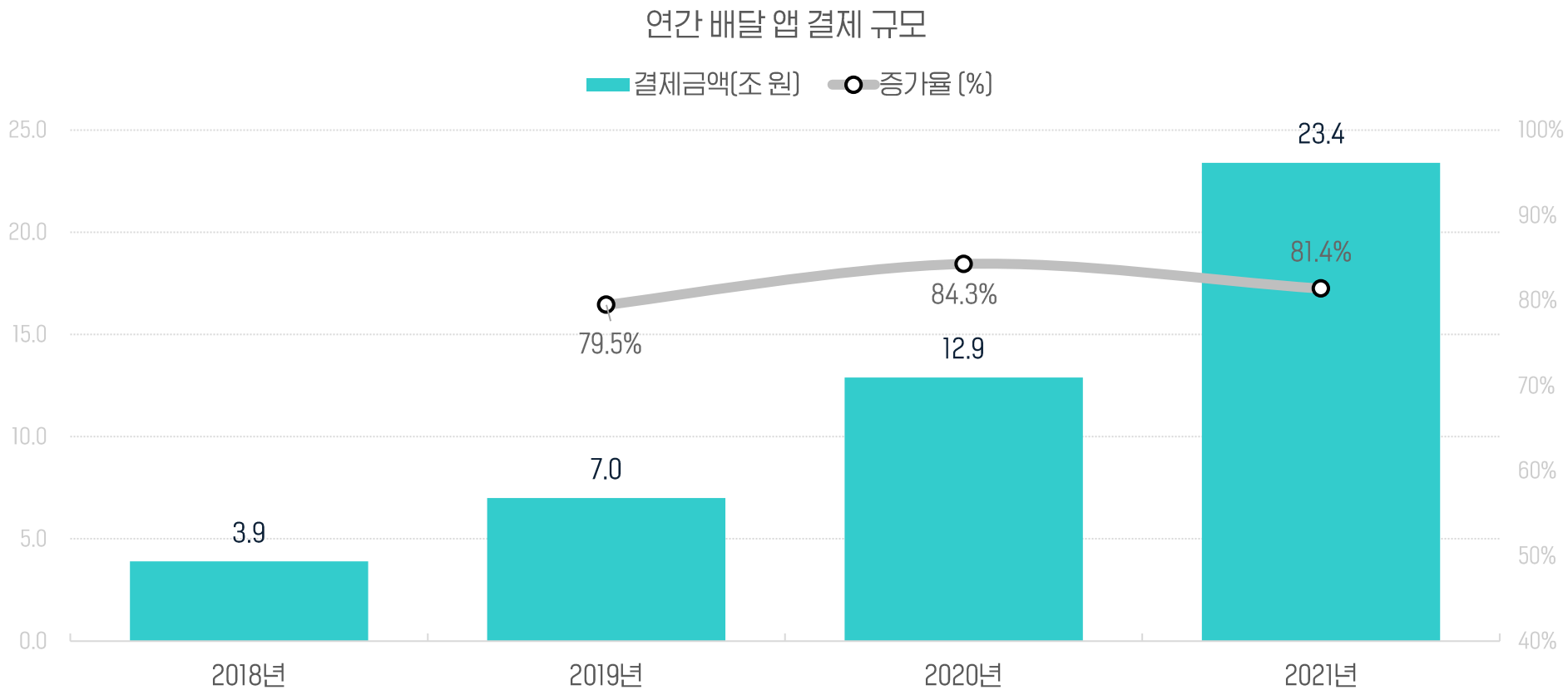
배달업 종사자 수 현황



\*출처 : 통계청 (<https://kostat.go.kr/>)

# 2021년 기준 결제 규모가 23조를 넘어서며 코로나 이후 80% 이상씩 증가 중

- 만 20세 이상의 개인이 결제한 금액의 통계 수치
- 10대의 결제금액, 법인카드결제, 현장결제, 간편결제는 제외한 수치이기 때문에 실제 배달 앱의 결제 규모는 더 클 것으로 예상

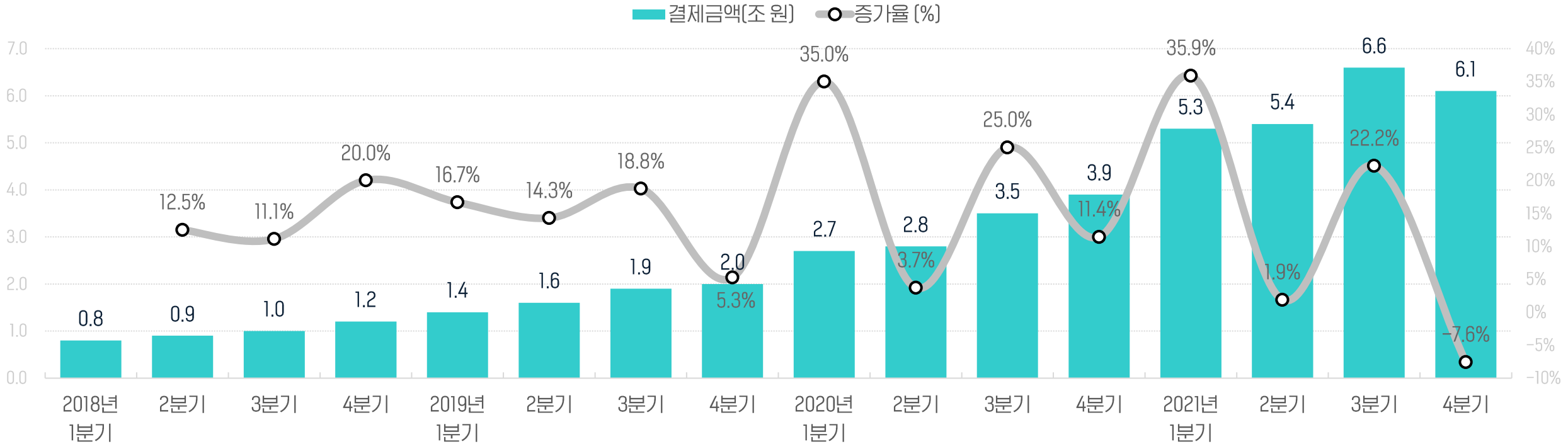


\*출처 : 와이즈앱 (<https://www.wiseapp.co.kr/>)

# 분기별 규모도 꾸준히 증가하는 추세이며 2021년 4분기 처음으로 감소

- 2021년 4분기, 워드코로나 선언으로 인해 꾸준히 증가 추이를 보이던 결제 규모가 처음으로 감소하는 현상이 발생

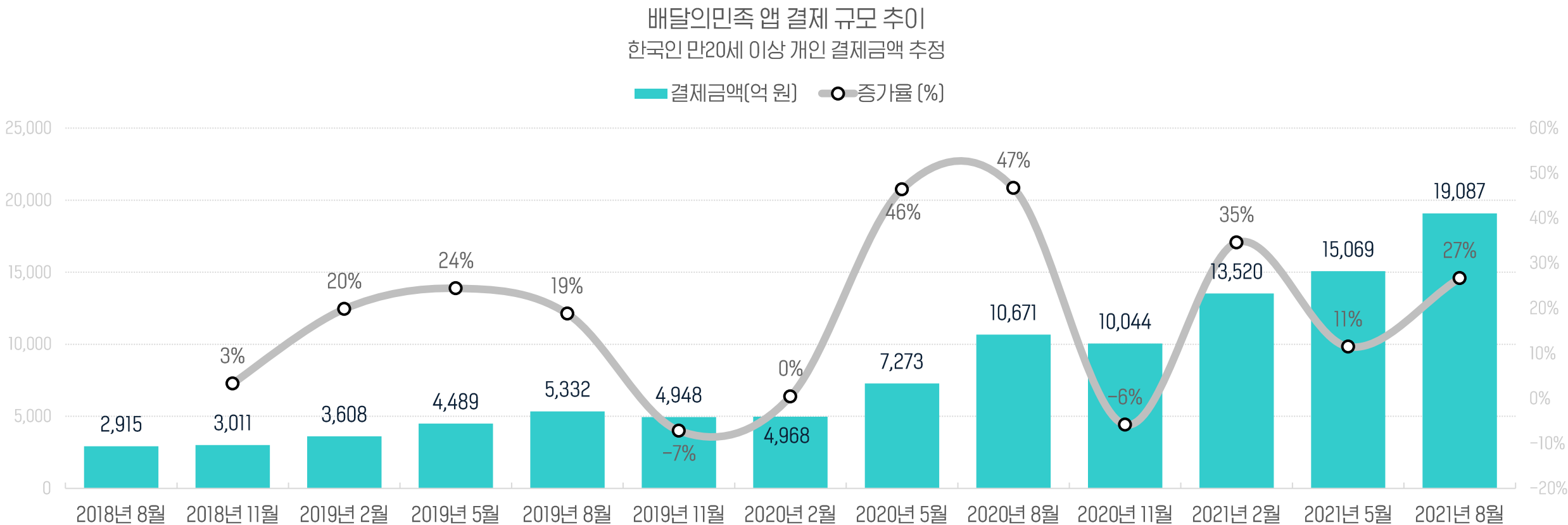
배달 앱 분기별 결제 규모  
한국인 만20세 이상 개인 결제금액(배달의민족+요기요+쿠팡이츠)



\*출처 : 와이즈앱 (<https://www.wiseapp.co.kr/>)

# 경쟁업체를 제외한 배달의 민족의 결제 규모는 지속적으로 증가하는 추세

- 경쟁업체를 제외하고 결제규모를 보더라도 꾸준히 증가하는 추세
- 코로나 발생으로 2020년 5~8월에 가장 큰 증가율을 보임
- 2021년 8월, 결제금액(1조 9,087억)과 결제자수(1,553만명)으로 역대 최고를 기록



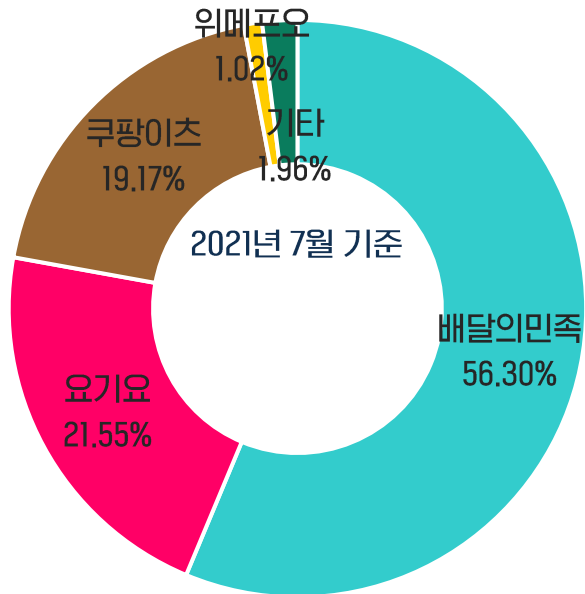
\*출처 : 와이즈앱 (<https://www.wiseapp.co.kr/>)

# 배달앱 시장은 배달의 민족이 압도적으로 점유 하지만 점유율 추이는 쿠팡이츠가 유일하게 증가

- 2021년 7월 기준으로 배달의 민족은 시장 점유율에 반 이상을 차지
- 1,2위를 다투는 배달앱이 시장 점유가 하락한 것에 비해 쿠팡이츠는 성장세  
2019년 출범한 이후 2년 만에 전년도 대비 점유율이 10배 성장

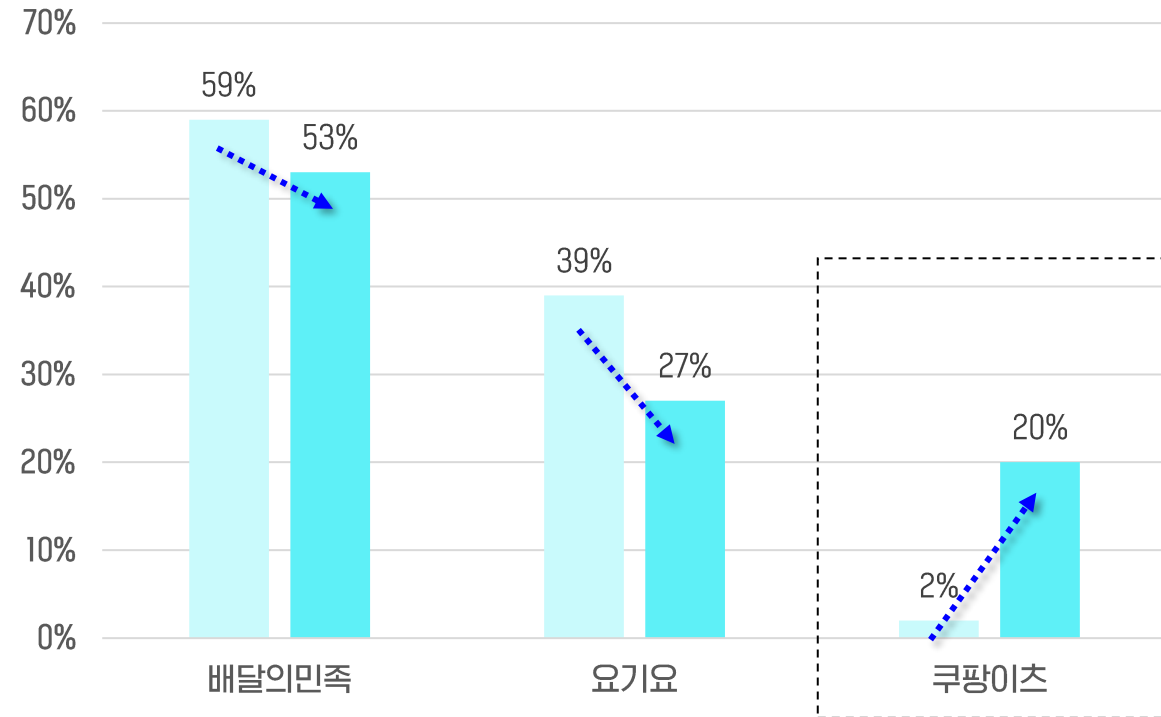
국내 주요 배달앱 시장 점유율

■ 배달의민족 ■ 요기요 ■ 쿠팡이츠 ■ 위메프오 ■ 기타



국내 주요 배달앱 시장 점유율 변화

■ 2020년 ■ 2021년



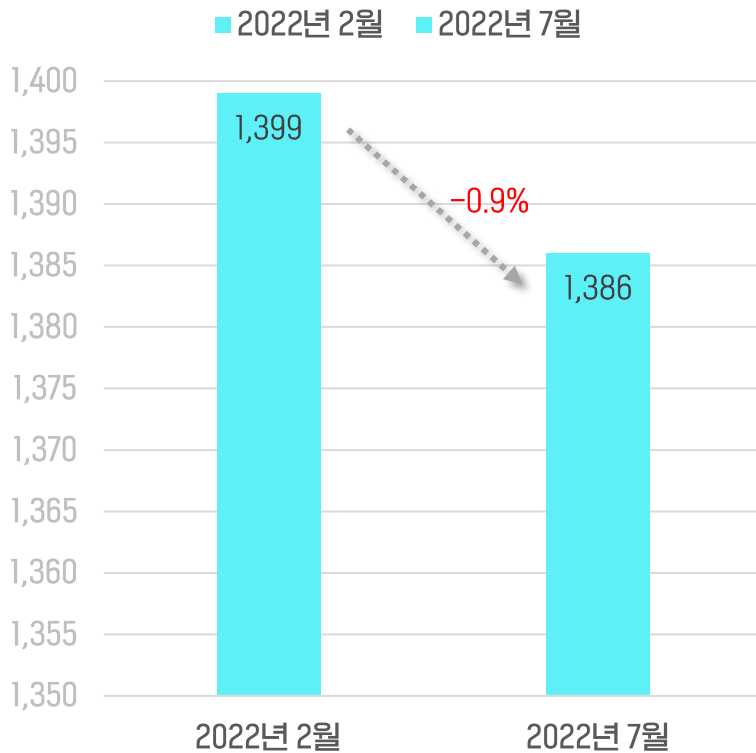
\*출처 : 글로벌빅데이터연구소 (<https://www.ciokorea.com/main/>)



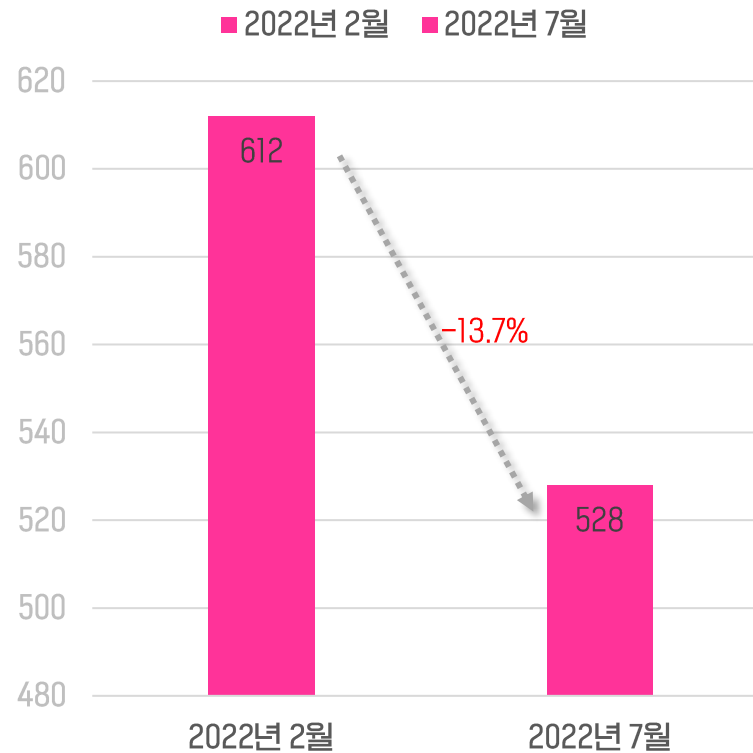
# 하지만 최근 월간 이용자 수를 확인해보면 주요 배달앱 모두 이용자 수가 감소하는 추이

- 배달앱 기업별로 각기 다른 감소율을 보이며, 배달의민족이 가장 낮은 감소율을 보임.
- 2021년, 시장 점유율이 전년도 대비 10배 가까이 성장한 쿠팡이츠의 경우, 이용자수의 감소 폭이 가장 큰 것으로 나타남.

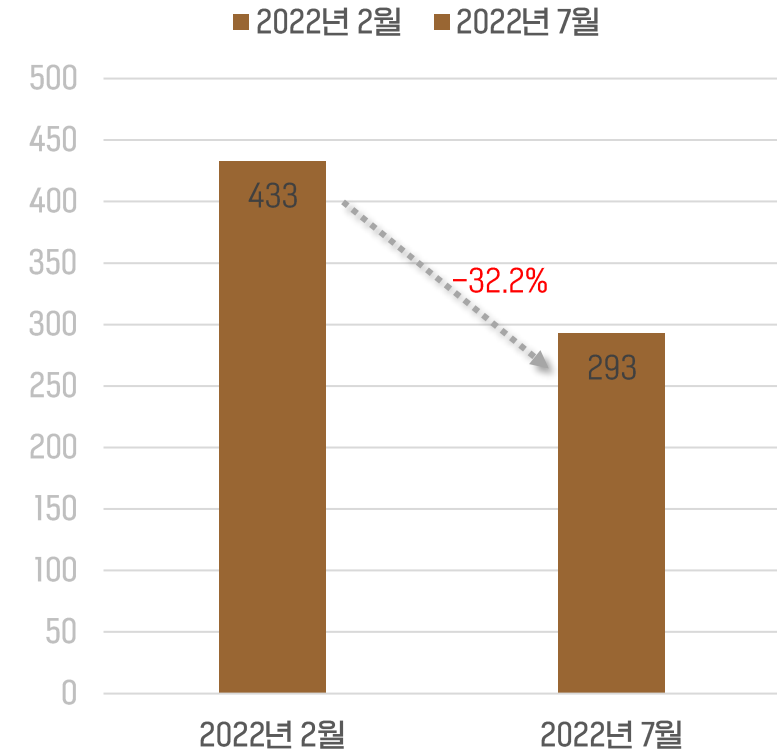
배달의민족 월간 이용자 수 추이(명)



요기요 월간 이용자 수 추이(명)



쿠팡이츠 월간 이용자 수 추이(명)

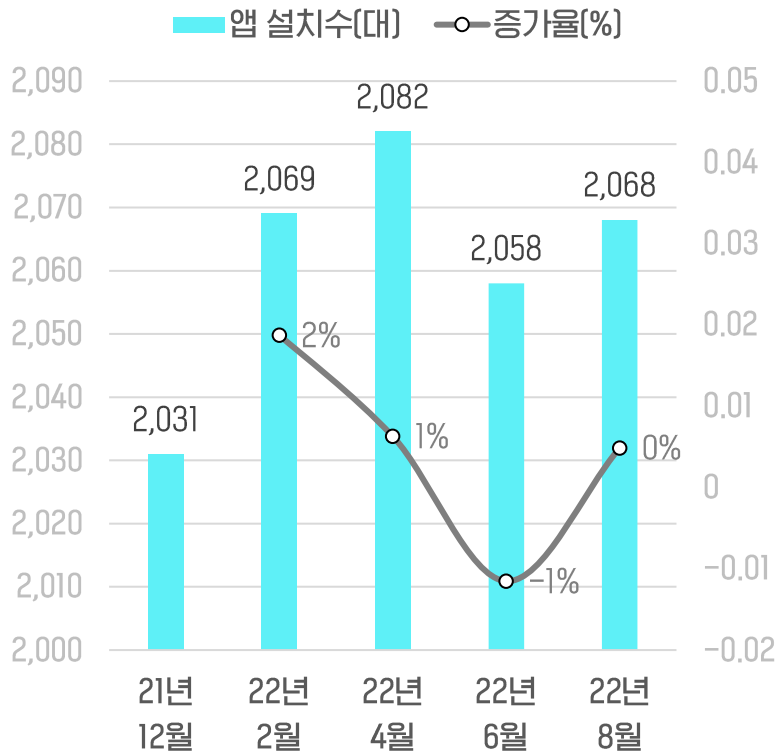


\*출처 : 모바일 인덱스(<https://www.mobileindex.com/home>)

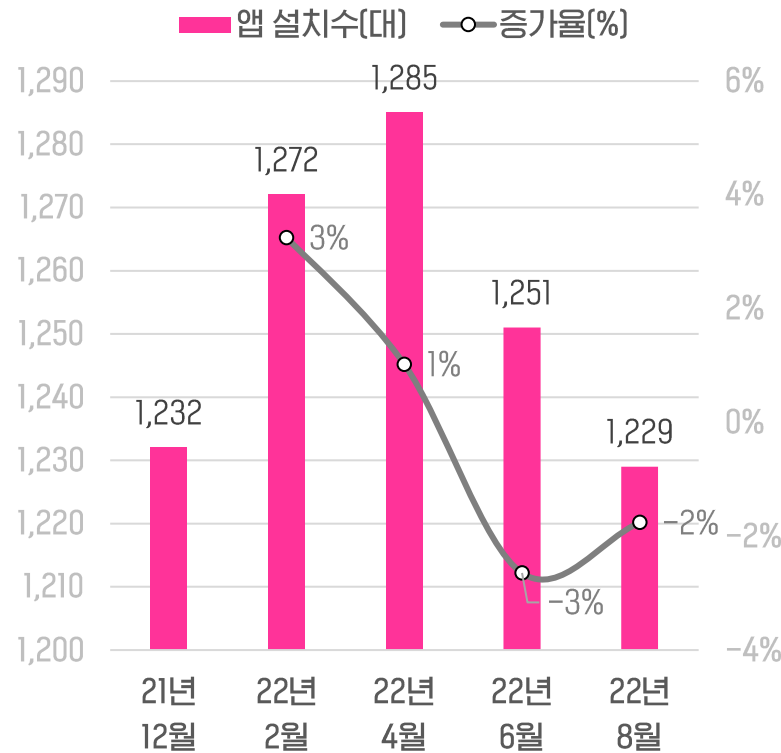
# 최근 앱 설치 수 추이를 확인해보도 성장이 다소 둔화된 양상

- 배달의민족의 경우 이미 사용자가 많은 앱인 만큼 증가 폭이 크지 않겠지만 사용자가 훨씬 적은 요기요, 쿠팡이츠의 앱 설치 수 증가율도 크게 다르지 않음.
- 쿠팡이츠가 다른 배달앱과 비교하여 증가와 감소의 폭이 상대적으로 큰 편

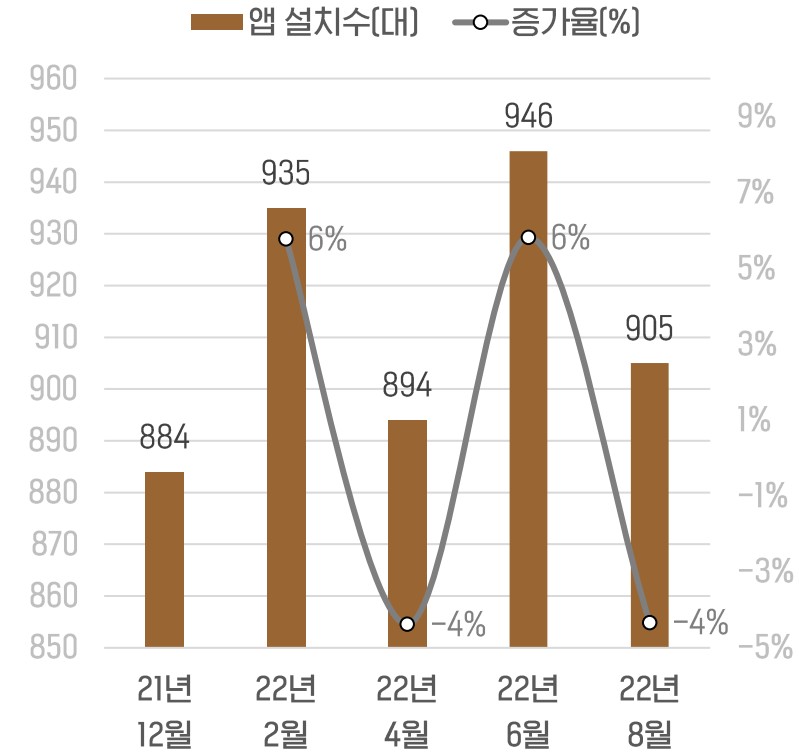
배달의민족 배달앱 설치 수 추이



요기요 배달앱 설치 수 추이



쿠팡이츠 배달앱 설치 수 추이



\*출처 : TDI 데이터 드래곤(<https://datadragon.co.kr/>)

배달 시장과 관련 어플리케이션은  
매년 괄목할 만한 성장 추이를 보이고 있었지만  
최근 들어 그 성장세가 다소 둔화된 시장 배경을 확인하였다.

꾸준히 성장해온 배달 시장과는 반대로  
최근 증가와 감소를 반복하는 배달 주요 앱 설치 수 추이와  
이용자 수의 감소는  
성장이 저하된 시장 배경을 증명해 준다.

배달의 민족의 시장 점유율은 여전히 1위를 달리고 있지만  
최근 점유율도 하락을 하는 추세에 직면해 있다.

시장 흐름 현황에 따른 소비자 인식은 어떻게 변화했는지  
데이터 측면에서 알아보려 한다.



# CHAPTER.2

## 소비자 인식 자료 분석

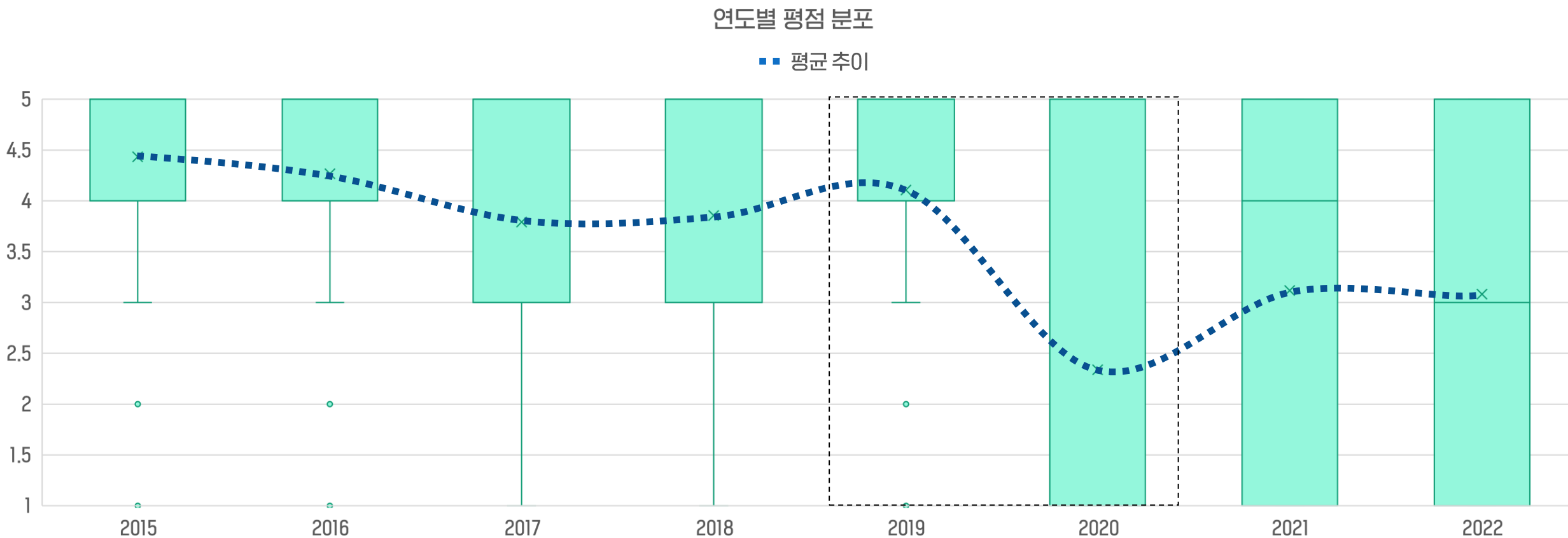
평점데이터를 통계적 기법으로 접근, 파악 후  
차이점이 생기는 부분에 대해 텍스트 분석으로 정밀하게 분석한다.

## 수치형 자료 (평점 데이터) 분석

# Box-plot\*을 활용한 연도별 평점 분포

## 평점의 분포와 평균이 2020년을 기점으로 하락

- 2019년까지 평점이 높은 점수를 유지하였지만 2020년을 기점으로 평점 분포가 큰 변화를 보임
- 평점의 평균 또한 2020년을 시작으로 하락



\*박스플롯(Box-plot): 많은 양의 데이터를 눈으로 확인하기 어려울 때 그림을 이용해 데이터 집합의 분포, 평균값, 중앙값 등을 빠르게 확인할 수 있는 목적으로 사용

## 변화가 생기는 지점에 대한 통계적 검증 진행



# T-test

## t-분포를 따르는 통계적 가설 검정법

두 집단 간의 평균의 차이를 비교하는 모수적 통계방법\*으로서  
모집단\*을 대표하는 표본\*으로부터 추정된 분산이나 표준편차를 이용하여  
검정하는 방법

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{X}$ : 두 집단 차이의 평균  
 $\mu$ : 모집단의 평균  
 $s$ : 두 집단 차이의 표준편차

\*모수적 통계방법: 관측값이 어느 특정한 확률분포를 따른다라는 전제 후에 검정을 실시하는 방법

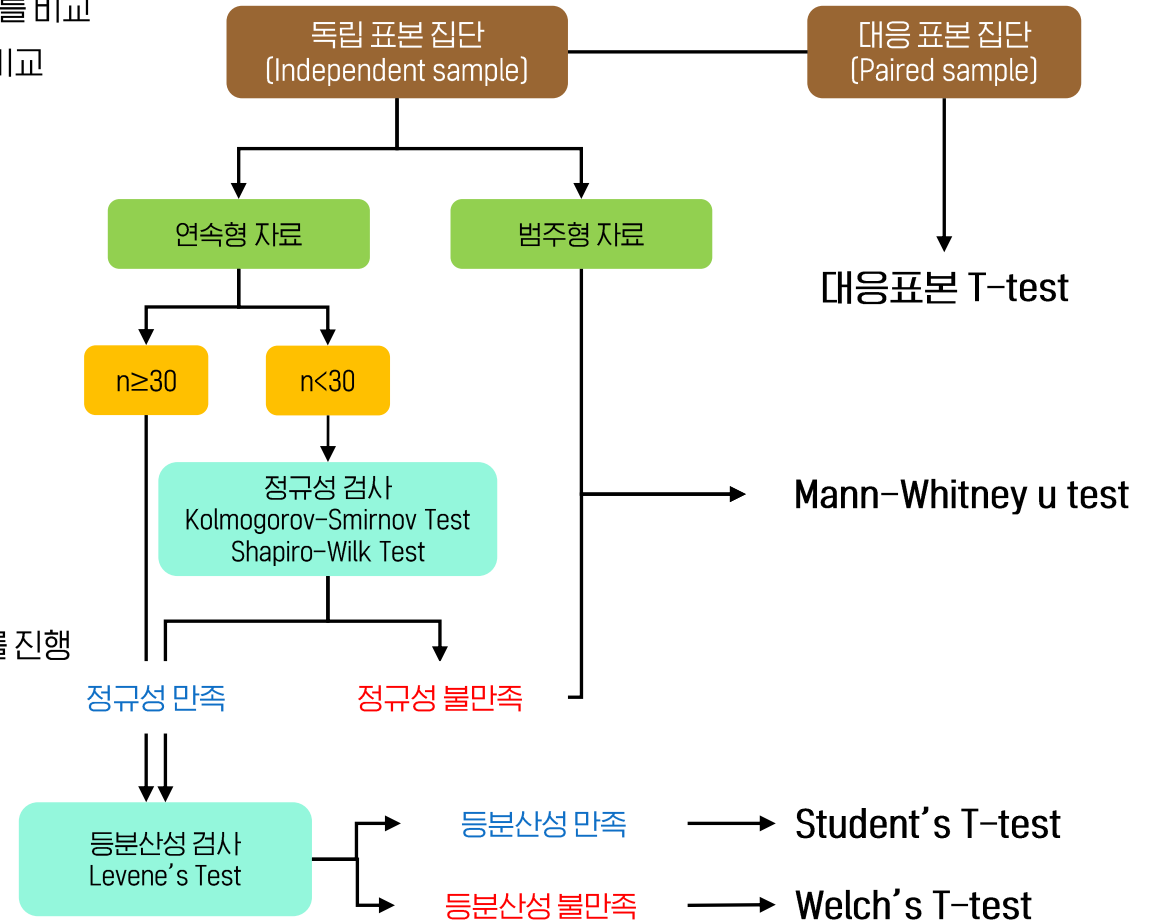
\*모집단: 정보를 얻고자 하는 대상이 되는 집단 단체

\*표본: 모집단의 부분집합으로, 여러 통계 자료를 포함하는 집단 속에서 그 일부를 뽑아내어 조사한 결과

# T-test 설명 및 절차

## 귀무가설과 대립가설\* 설정 후 절차에 따라 검증 진행

- 종류
  - 단일표본 t-test(One sample t-test): 하나의 집단에 대한 예상 평균치가 일치 하는지를 비교
  - 독립표본 t-test(Independent two sample t-test): 서로 다른 두개의 그룹 간의 평균 비교
  - 대응표본 t-test(Paired t-test): 하나의 집단에 대한 사전, 사후를 비교
- 가정
  - 독립성: 독립변수의 그룹 군은 서로 독립적이어야 한다.
  - 정규성: 독립변수에 따른 종속변수는 정규분포\*를 만족해야 한다.
  - 등분산성: 독립변수에 따른 종속변수 분포의 분산은 각 군마다 동일해야 한다.
- 방법
  - 독립성이 만족하지 못할 시, 독립표본 t-test가 아닌 대응표본 t-test로 진행
  - 관측치가 30개 이상일 시 정규성을 만족한다고 가정, 30개 이하일 시 Kolmogorov-Smirnov Test\*(혹은 Shapiro-Wilk test)로 정규성 검사를 진행
  - 정규성을 만족할 시, Levene's Test\*로 등분산성을 검증
    - 등분산성을 만족할 시, Student's t-test 진행
    - 등분산성을 만족하지 못할 시, Welch's t-test 진행



\*귀무가설과 대립가설: 연구는 검증해야 할 가설을 필요로 하는데, 일반적으로 연구에서 검증하는 가설을 귀무가설이라 하며, 귀무가설과 반대되는 가설을 대립가설이라고 일컫음

\*정규분포: 연속 확률 분포의 하나이며, 수집된 자료의 분포를 근사하는 데에 자주 사용

\*Kolmogorov-Smirnov Test: 표본을 참조 확률 분포와 비교하거나 두 표본을 비교하는 데 사용할 수 있는 연속 1차원 확률 분포의 동등성에 대한 비모수 검정

\*Levene's Test: 둘 이상의 그룹에 대해 계산된 변수에 대한 분산의 동등성을 평가하는 데 사용되는 추론 통계량

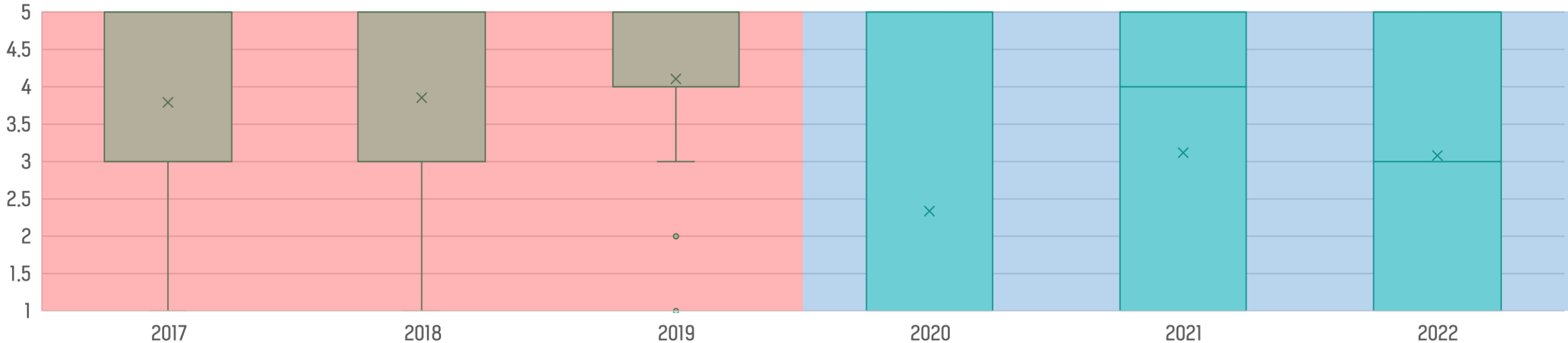
# 육안으로 식별한 연도별 평점 차이를 통계적 검증을 진행하여 증명

- 평점의 변화가 가장 컸던 2019년과 2020년을 기준으로 3개년도를 그룹 지어 검정 진행
- 귀무가설, 대립가설 설정에 따른 단계별 검정 진행

**귀무가설  $H_0$ : 2017~2019년의 평점 평균과 2020~2022년의 평점 평균은 차이가 없다.**

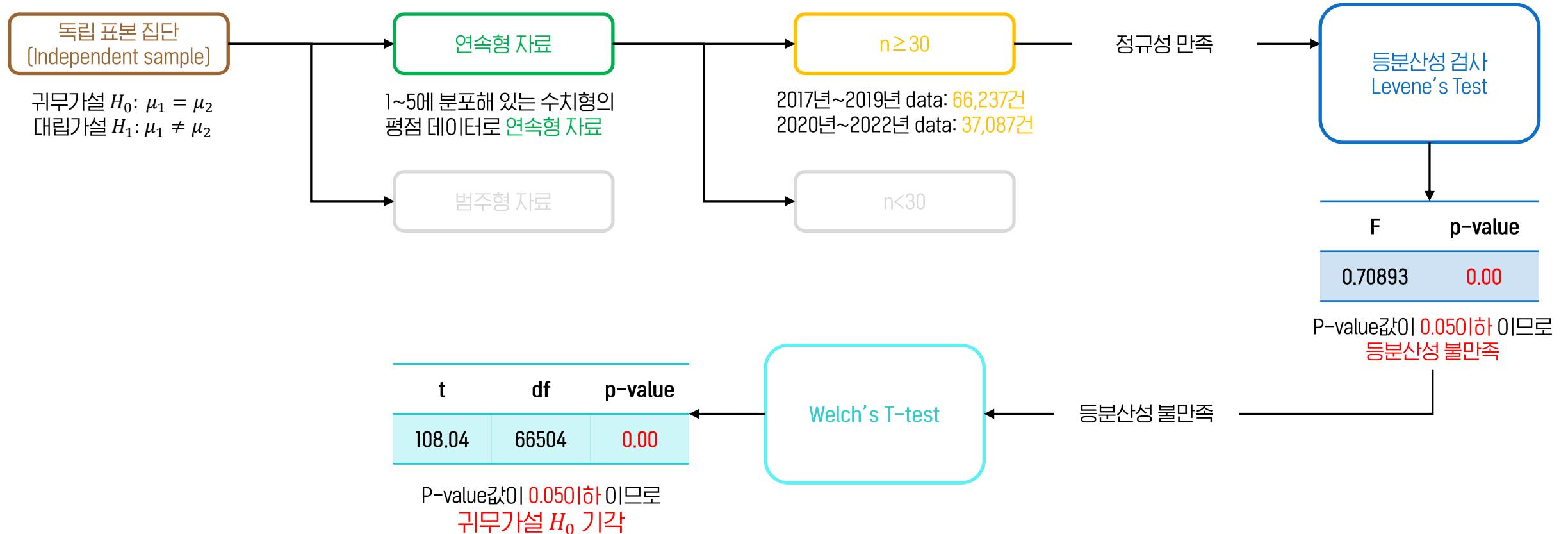
**대립가설  $H_1$ : 2017~2019년의 평점 평균과 2020~2022년의 평점 평균은 차이가 있다.**

연도별 평점 분포



# 독립된 두개의 집단이므로 독립표본 T-test 진행 유의수준 0.05에서 가설 검증

- 표본 집단의 종류 파악(결과: 독립 표본 집단)
- 자료의 특징 및 관측치의 개수로 정규성 검사 진행(결과: 정규성 만족)
- Levene's Test를 통해 등분산성 검사 진행(결과: 등분산성 불만족)
- 최종적으로 Welch's T-test 진행



# 독립표본 T-test 분석 결과

## 귀무가설 기각(대립가설 채택)

- 여러 검증 절차를 통하여 결론으로 나온 “비교 기간의 평점 평균은 차이가 있다” 라는 대립가설을 채택
- 해당 검증 내용을 근거로 비교 기간에 대한 텍스트 분석을 진행



## 텍스트 자료 (리뷰 데이터) 분석

# 연간 키워드 순위 목록

## 명사를 기준으로 추출

- 단순 명사 빈도수를 확인한 결과 2020년 이후에 부정적인 단어가 다수 형성, 특히 2020년은 '배달', '배민' 과 같은 일반적인 명사를 제외하고 '삭제' 키워드가 가장 높은 빈도를 기록
- '삭제' 라는 키워드 외에 '탈퇴', '취소' 등의 키워드가 새로 등장

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	어플	어플	어플	어플	배달	배달	배달	배달	배달	배달	배달	배달	배달
2	배달	배달	배달	배달	어플	어플	어플	결제	배민	배민	배민	배민	배민
3	유용	유용	배달의민족	배달의민족	배달의민족	배달의민족	배민	리뷰	어플	주문	삭제	주문	주문
4	메뉴	정말	정말	리뷰	민족	민족	결제	주문	주문	쿠폰	주문	리뷰	리뷰
5	전화	메뉴	유용	음식	결제	할인	주문	배민	리뷰	사용	민족	가게	가게
6	정말	편리	민족	바로	바로	결제	배달의민족	어플	결제	어플	리뷰	시간	쿠폰
7	무료	완전	메뉴	정말	주문	주문	리뷰	바로	이용	이벤트	수수료	음식	음식
8	디자인	필요	음식	메뉴	음식	배민	민족	전화	사용	결제	어플	쿠폰	시간
9	검색	최고	전화	민족	리뷰	바로	삭제	쿠폰	쿠폰	리뷰	탈퇴	고객	결제
10	완전	전단지	주년	편리	할인	리뷰	할인	이용	고객	이용	쿠폰	전화	어플
11	위치	배달의민족	편리	결제	정말	이용	이용	사용	개인정보	요기요	사용	결제	사용
12	최고	음식	리뷰	전화	이용	음식	음식	주소	삭제	편리	전화	사용	고객
13	사용	전화	축하	유용	사용	사용	이벤트	다른	음식	할인	가게	진짜	진짜
14	전단지	검색	전단지	사용	다른	다른	구매	삭제	전화	전화	다시	불편	메뉴
15	필요	사용	필요	전단지	최고	정말	바로	이벤트	업체	최고	음식	메뉴	다른
16	진짜	바로	사용	음식점	전화	최고	사용	할인	최고	고객	이용	어플	전화
17	등록	진짜	바로	이용	편리	자주	다른	음식	다른	음식	시간	취소	기능
18	편리	치킨	최고	주문	쿠폰	편리	부탁	불편	정말	시간	결제	다른	감사
19	음식점	음식점	완전	진짜	메뉴	전화	최고	업데이트	민족	다른	기업	기능	다시
20	치킨	정보	야식	맛집	자주	포인트	정말	최고	바로	정말	다른	감사	취소

# 연간 키워드 순위 목록

## 모든 품사를 기준으로 추출

- 모든 품사에 대한 빈도수를 확인한 결과 가장 눈에 띄는 점은 '좋아요' 키워드의 순위 변화
- 2020년을 기점으로 하락세가 눈에 띈

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	어플	어플	어플	어플	배달	배달	배달	배달	배달	배달	배달	배달	배달
2	배달	배달	배달	배달	어플	어플	어플	결제	좋아요	좋아요	삭제	주문	주문
3	좋아요	좋아요	좋아요	시켜	배달의민족	배달의민족	좋아요	좋아요	어플	주문	주문	리뷰	리뷰
4	유용	유용	배달의민족	배달의민족	결제	좋아요	결제	리뷰	주문	쿠폰	너무	너무	너무
5	메뉴	시켜	시켜	좋아요	바로	할인	주문	주문	리뷰	너무	리뷰	가게	가게
6	전화	정말	정말	리뷰	좋아요	결제	배달의민족	어플	결제	사용	수수료	시간	쿠폰
7	시켜	너무	너무	너무	시켜	주문	시켜	바로	이용	어플	어플	음식	음식
8	너무	메뉴	유용	음식	너무	시켜	리뷰	전화	너무	이벤트	합퇴	쿠폰	맛있
9	정말	^^	^^	바로	주문	바로	삭제	쿠폰	사용	결제	쿠폰	맛있	시간
10	^^	편리	메뉴	정말	음식	너무	너무	이용	쿠폰	리뷰	사용	고객	결제
11	디자인	완전	음식	메뉴	리뷰	리뷰	할인	사용	고객	이용	전화	전화	어플
12	무료	좋네요	전화	^^	할인	이용	이용	너무	개인정보	요기요	아니	아니	아니
13	검색	필요	주변	편리	정말	음식	음식	주소	삭제	편리	가게	결제	사용
14	좋네요	최고	편리	결제	이용	사용	이벤트	다른	음식	할인	좋아요	사용	고객
15	완전	전단지	리뷰	전화	사용	다른	구매	삭제	^^	^^	다시	진짜	진짜
16	위치	배달의민족	축하	좋은	다른	정말	바로	이벤트	전화	전화	음식	좋아요	좋아요
17	최고	음식	전단지	유용	최고	최고	사용	시켜	업체	최고	이용	불편	메뉴
18	좋은	전화	좋은	사용	전화	^^	다른	할인	시켜	고객	시간	메뉴	다른
19	전단지	검색	필요	전단지	좋은	많이	부탁	아니	최고	음식	결제	어플	시켜
20	사용	사용	사용	같아요	편리	좋은	최고	좋네요	다른	시간	기업	취소	전화



# 연간 키워드 순위 목록

## 모든 품사를 기준으로 추출

- 부정적인 서술어와 어울려 사용이 된다고 볼 수 있는 '너무' 키워드의 언급량 또한 시간이 지남에 따라 증가
- '아니' 키워드도 2020년 이후 처음 순위권에 자리잡음을 확인

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	어플	어플	어플	어플	배달	배달	배달	배달	배달	배달	배달	배달	배달
2	배달	배달	배달	배달	어플	어플	어플	결제	좋아요	좋아요	삭제	주문	주문
3	좋아요	좋아요	좋아요	시켜	배달의민족	배달의민족	좋아요	좋아요	어플	주문	주문	리뷰	리뷰
4	유용	유용	배달의민족	배달의민족	결제	좋아요	결제	리뷰	주문	주문	너무	너무	너무
5	메뉴	시켜	시켜	좋아요	바로	할인	주문	주문	리뷰	너무	리뷰	가게	가게
6	전화	정말	정말	리뷰	좋아요	결제	배달의민족	어플	결제	사용	수수료	시간	쿠폰
7	시켜	너무	너무	너무	시켜	주문	시켜	바로	이용	어플	어플	음식	음식
8	너무	메뉴	유용	음식	너무	시켜	리뷰	전화	너무	이벤트	탈퇴	쿠폰	맛있
9	정말	^^	^^	바로	주문	바로	삭제	쿠폰	사용	결제	쿠폰	맛있	시간
10	^^	편리	메뉴	정말	음식	너무	너무	이용	쿠폰	리뷰	사용	고객	결제
11	디자인	완전	음식	메뉴	리뷰	리뷰	할인	사용	고객	이용	전화	전화	어플
12	무료	좋네요	전화	^^	할인	이용	이용	너무	개인정보	요기요	아니	아니	아니
13	검색	필요	주변	편리	정말	음식	음식	주소	삭제	편리	가게	결제	사용
14	좋네요	최고	편리	결제	이용	사용	이벤트	다른	음식	할인	좋아요	사용	고객
15	완전	전단지	리뷰	전화	사용	다른	구매	삭제	^^	^^	다시	진짜	진짜
16	위치	배달의민족	축하	좋은	다른	정말	바로	이벤트	전화	전화	음식	좋아요	좋아요
17	최고	음식	전단지	유용	최고	최고	사용	시켜	업체	최고	이용	불편	메뉴
18	좋은	전화	좋은	사용	전화	^^	다른	할인	시켜	고객	시간	메뉴	다른
19	전단지	검색	필요	전단지	좋은	많이	부탁	아니	최고	음식	결제	어플	시켜
20	사용	사용	사용	같아요	편리	좋은	최고	좋네요	다른	시간	기업	취소	전화

## 리뷰 데이터 내용에 대해 Text-mining 기법 적용

# TF-IDF 키워드 중요도 분석

## Term Frequency-Inverse Document Frequency

데이터 내부에서 특정 키워드의 빈도를 이용하여  
키워드의 중요도를 나타내는 분석 기법

$$\begin{aligned}
 & \mathbf{TF - IDF}(t, d, D) \\
 & = \mathbf{TF}(t, d) \times \mathbf{IDF}(t, D)
 \end{aligned}$$

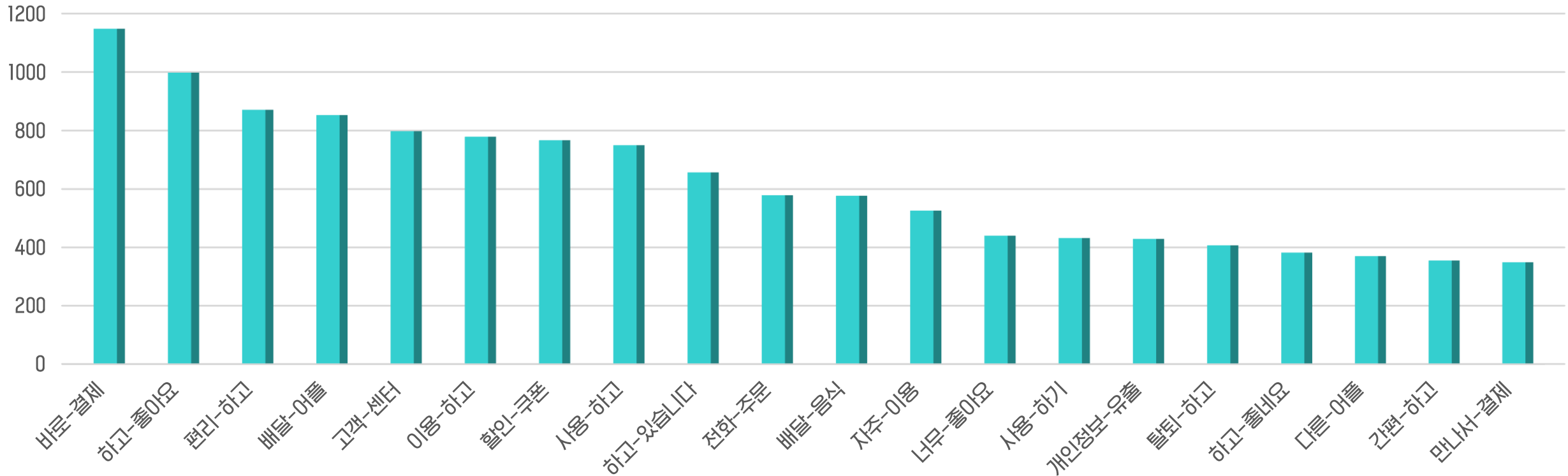
$$\mathbf{TF}(t, d) = \frac{\text{문서 } d \text{에서 특정 단어 } t \text{의 출현 빈도}}{\text{문서 } d \text{에서 총 단어의 수}}$$

$$\mathbf{IDF}(t, D) = \log \frac{\text{총 문서의 개수}}{\text{특정 단어 } t \text{를 포함하는 문서의 개수}}$$

# 2017~2019년 TF-IDF 분석 상위 20개 키워드 조합 결과

- 각 단어들의 중요한 정도를 가중치를 부여
- 다양한 키워드 조합을 통해 데이터의 주된 인식을 유추

2017~2019년 TF-IDF 분석

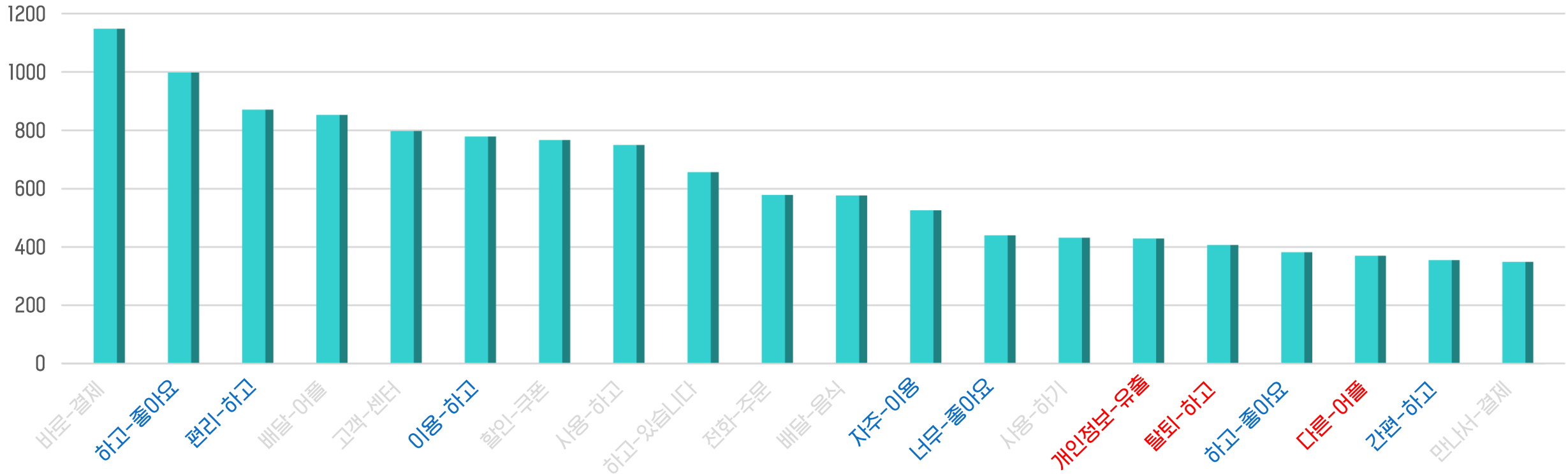


# 2017~2019년 TF-IDF 분석 키워드 조합 결과

## 긍정적인 인식이 주로 형성

- 개인정보 유출로 인하여 탈퇴 및 다른 앱을 사용한다는 인식이 형성
- 하지만 앱을 사용하는데 있어 편하고 좋으며, 자주 이용하고 있어 간편하다는 인식의 **가중치가 높게 형성**

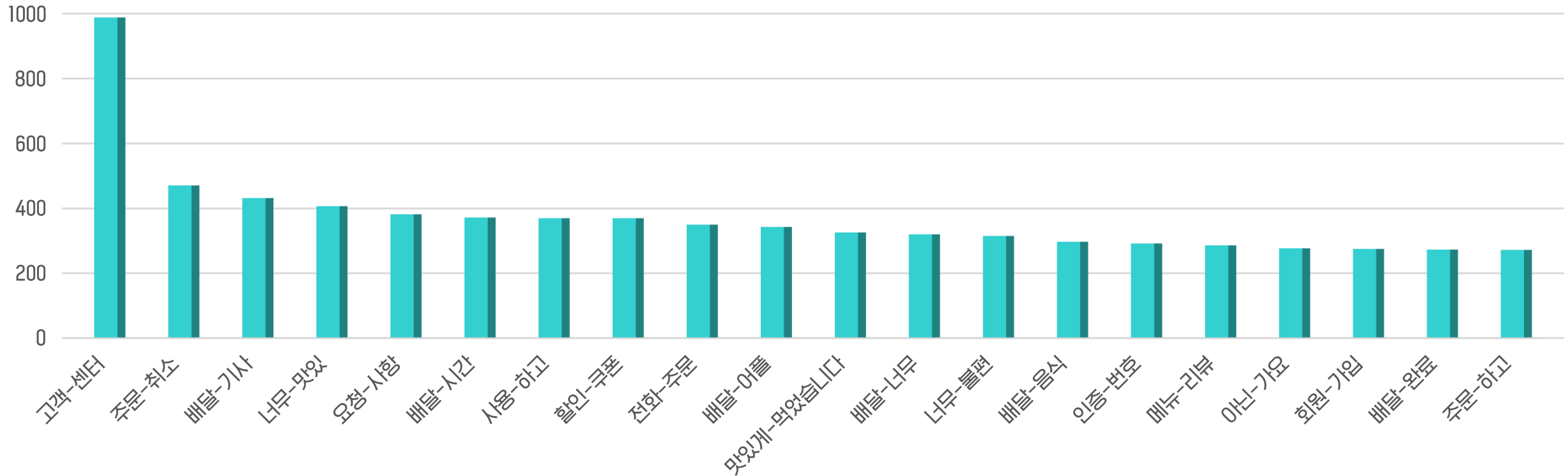
2017~2019년 TF-IDF 분석



# 2020~2022년 TF-IDF 분석 상위 20개 키워드 조합 결과

- 2017~2019년과 다르게 한 쌍의 키워드 조합에서 가중치가 압도적으로 높게 나옴

2020~2022년 TF-IDF 분석

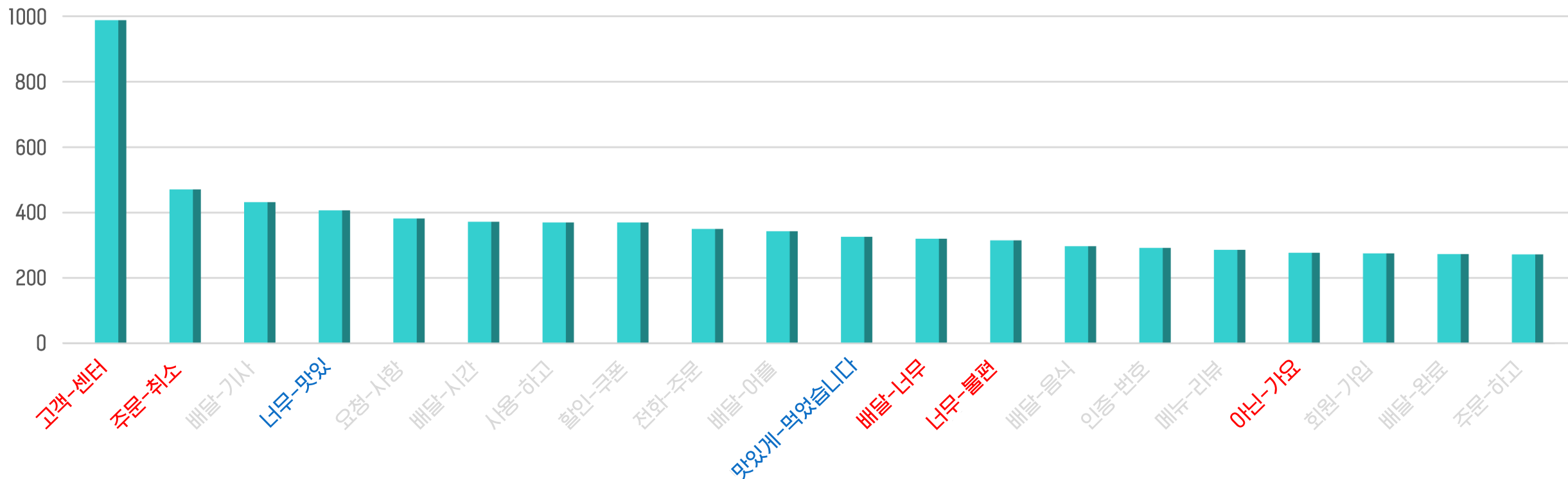


# 2020~2022년 TF-IDF 분석 키워드 조합 결과

## 부정적인 인식이 주로 형성

- 불편함을 호소하며 주문을 취소하고 고객센터에 문의가 많았음을 유추할 수 있음
- 맛있게 먹었다는 긍정적인 리뷰의 키워드 조합도 형성되었지만, 부정적인 인식 키워드의 가중치가 높게 형성

2020~2022년 TF-IDF 분석



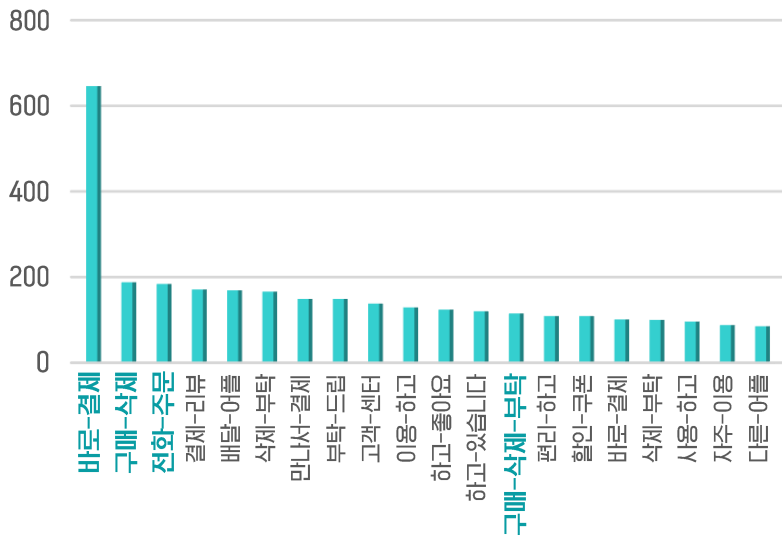
이를 연도별로 세부화하여  
분석해 보면,



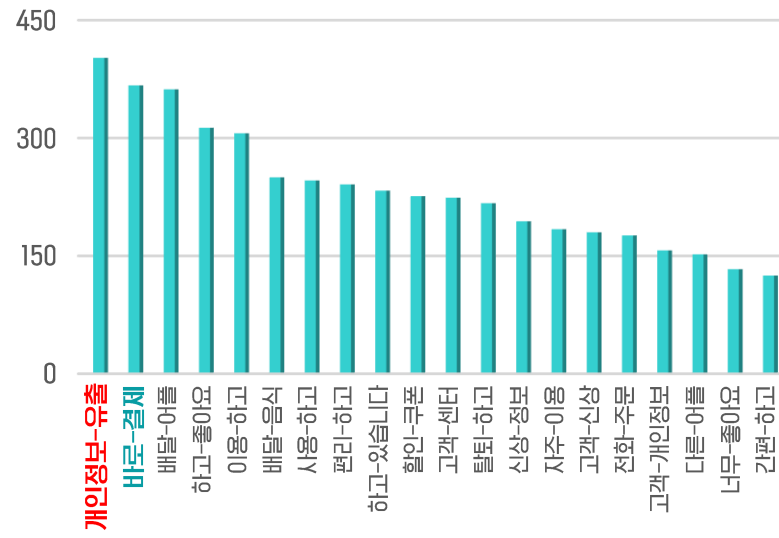
# 2017~2019년에 대해 연도별 분석을 진행 바로-결제, 개인정보 유출 등의 이슈 발견

- 2017년: 바로 결제에 대해서만 리뷰 작성이 가능해서 만나서 결제를 이용한 고객들의 리뷰를 달지 못하는 불만 형성, 고객들에게 구매 내역 삭제 권한이 없어 핸드폰 용량 부족에 따른 구매 내역 삭제 요청
- 2018년: 개인정보 유출 사건으로 인하여 배달의 민족의 방관에 대한 비판 글 형성
- 2019년: 할인 쿠폰 이벤트 시행 및 플랫폼 서비스에 대한 만족도가 전체적으로 높음을 알 수 있음

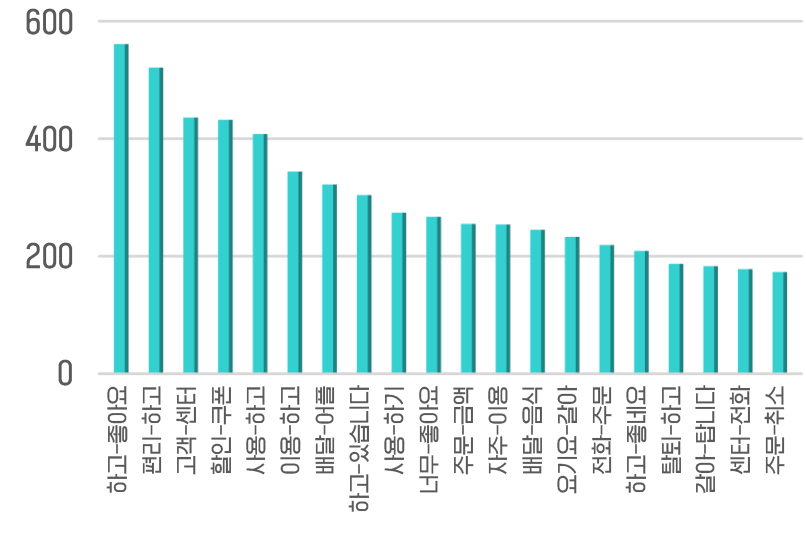
2017 TF-IDF 분석



2018 TF-IDF 분석



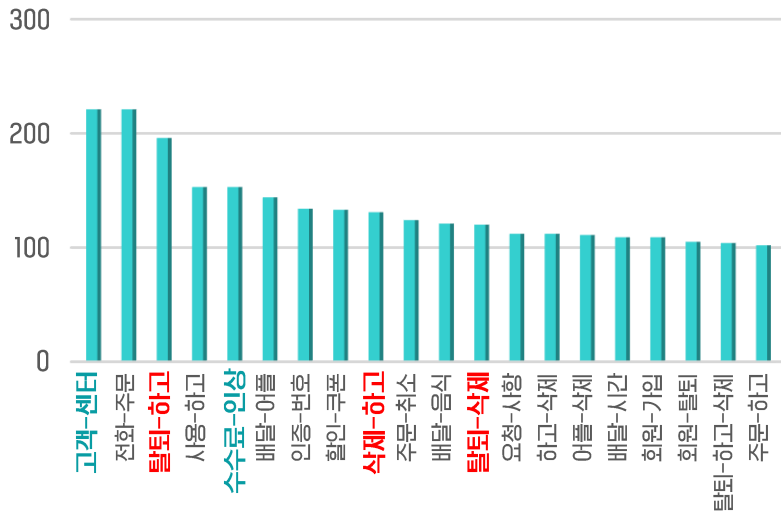
2019 TF-IDF 분석



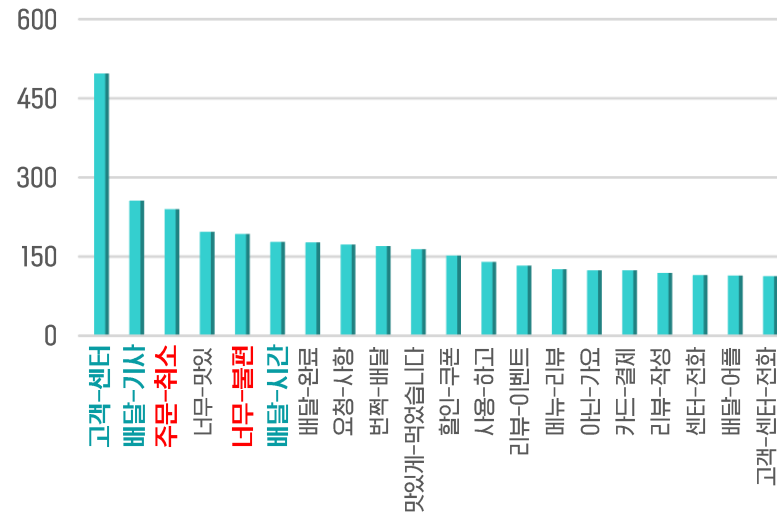
# 2020~2022년에 대해 연도별 분석을 진행 공통적으로 고객-센터 이슈가 가장 높게 형성

- 여러가지 사태로 인해 고객센터 문의 및 불만이 많았음을 유추 가능
- 2020년: 수수료인상에 따른 배달의 민족 어플을 탈퇴, 삭제하고 음식점에 직접 전화 주문을 하겠다는 여론이 발생
- 2021년: 배달기사에 대한 불만, 오래 걸리는 배달 시간으로 인해 불편함을 호소하며 주문 취소에 대한 고객센터 문의가 많음을 알 수 있음
- 2022년: 특정 이슈가 아닌 다양한 이슈로 고객센터에 대한 불만이 많았음을 알 수 있음

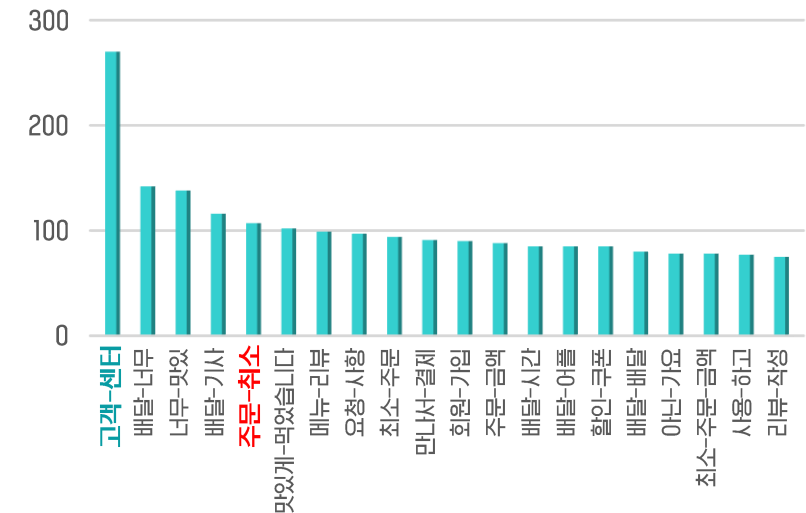
2020 TF-IDF 분석



2021 TF-IDF 분석



2022 TF-IDF 분석



2017 TF-IDF 분석

2018 TF-IDF 분석

2019 TF-IDF 분석

평점 분포를 box-plot으로 확인한 결과,  
2017~2019년은 대체적으로 좋은 평점인 것을 확인할 수 있었지만  
해당 분석에서는 해당 년도에 크고 작은 이슈들이 있었음을 확인할 수 있었다.  
전반적으로 만족하는 평점과 달리 불편함을 겪을 수 있는 이슈들이 발생했으며,  
평점분포나 키워드 중요도 분석을 통해 알 수 있듯이  
가장 만족도가 높았던 2019년에 비해

2020 TF-IDF 분석

2022 TF-IDF 분석

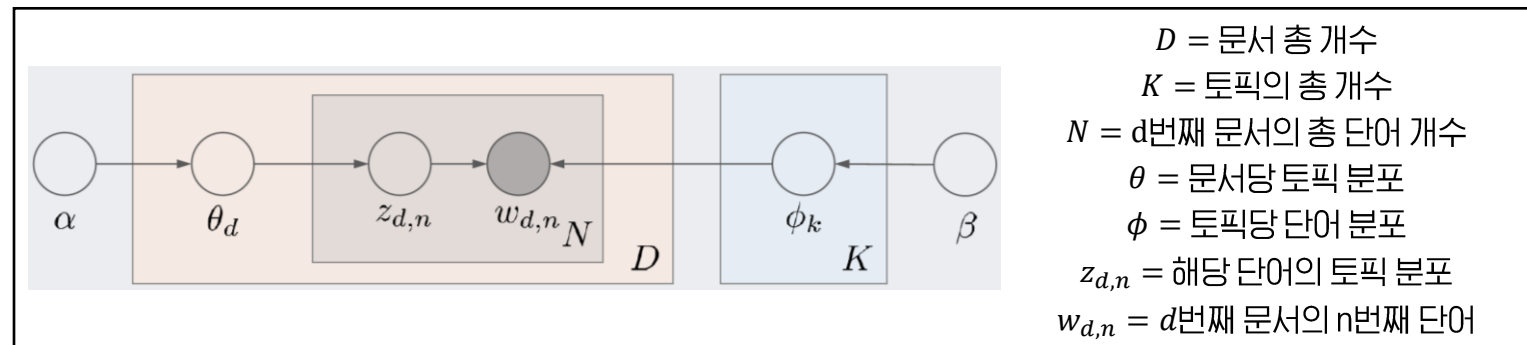
2020년 만족도에 큰 추락의 요인이 무엇인지 알 수 있었다.

이전보다 못한 서비스 및 어플 플랫폼에 대한 불만 또한 증가했으며,  
서비스가 좋았던 2019년과 상대적으로 비교가 많이 되어  
저조한 평점 및 다양한 불만이 발생했음을 발견하였다.

# LDA 잠재 디리클레 할당

## Latent Dirichlet Allocation

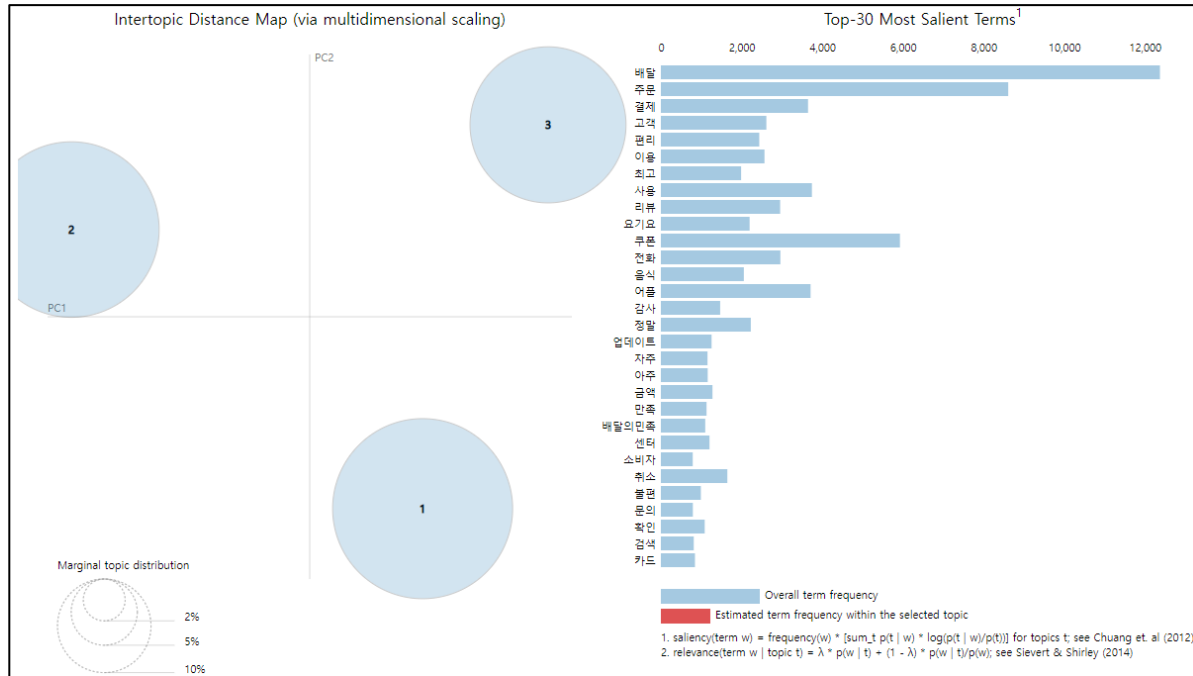
텍스트 기반의 문서 데이터에서 핵심 주제(topic)를 찾는 데이터 분석 방법론인 **토픽 모델링**의 가장 대표적인 알고리즘  
확률 기반의 모델링 기법을 통해 방대한 양의 문서 데이터를 분석



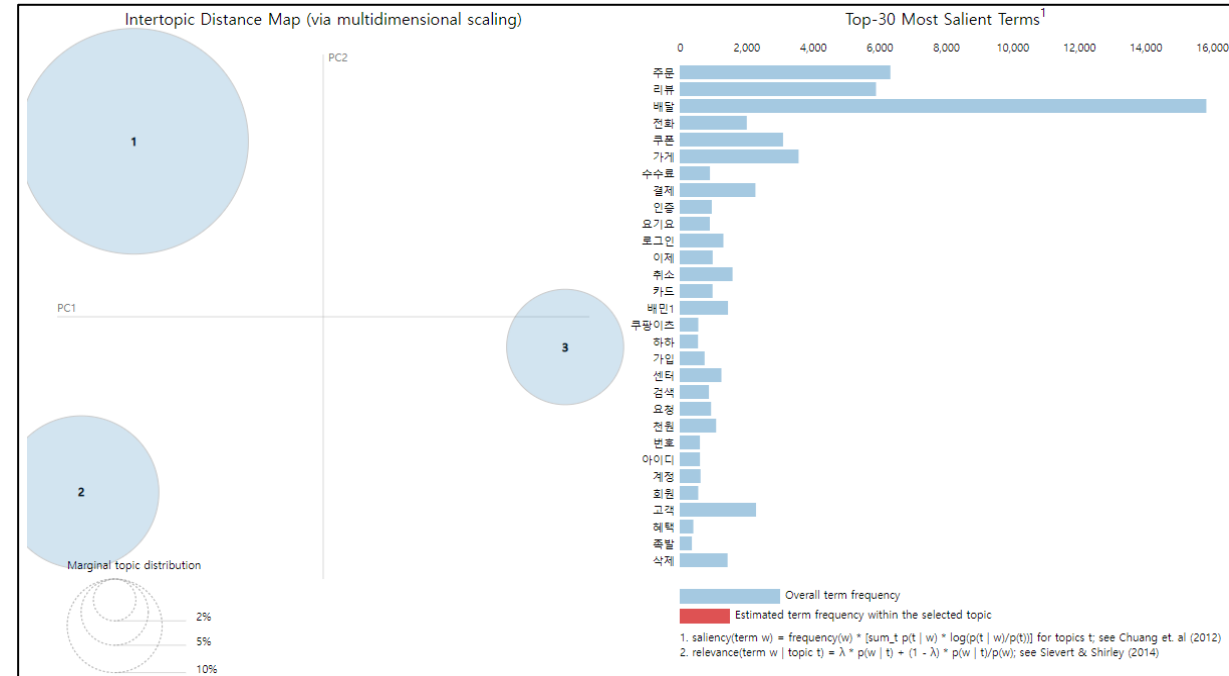
# LDA 토픽모델링 분석을 통한 기준년도에 대표 키워드 및 토픽 파악

- 토픽 별 비교분석을 위해 Gibbs sampling\* 방식을 이용한 토픽모델링 분석
- 기준년도 별로 토픽 수는 3개로 지정
- 토픽을 표현하는 원은 토픽 구성을 나타내며,  
원 사이의 거리가 멀수록 토픽 간의 연관성이 낮음을 의미

2017-2019 LDA 분석 결과



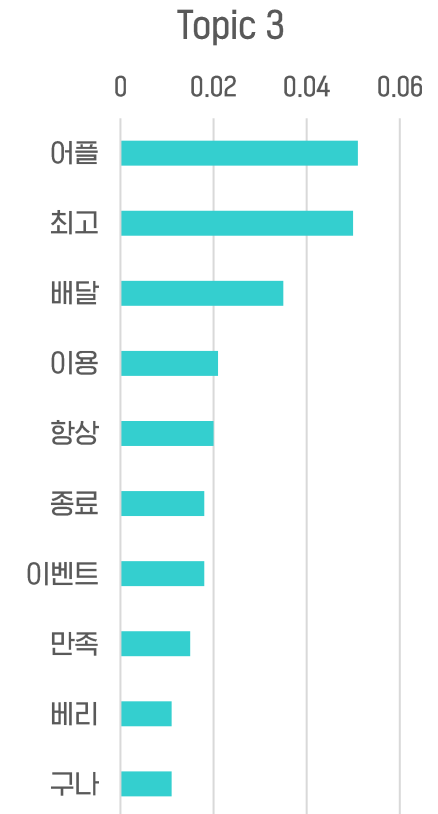
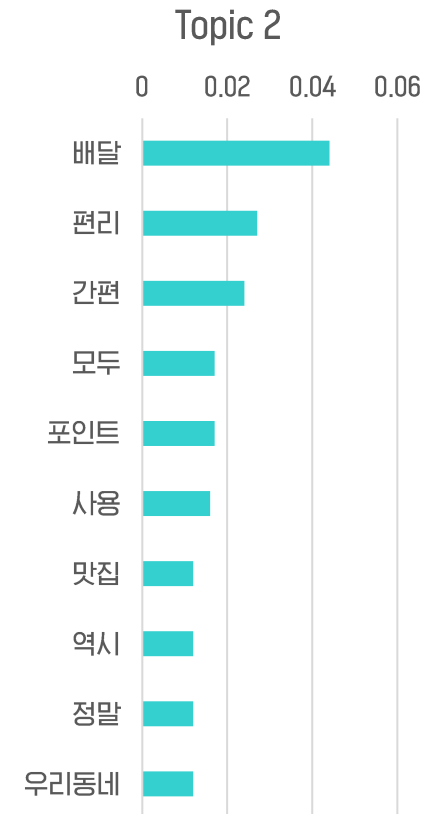
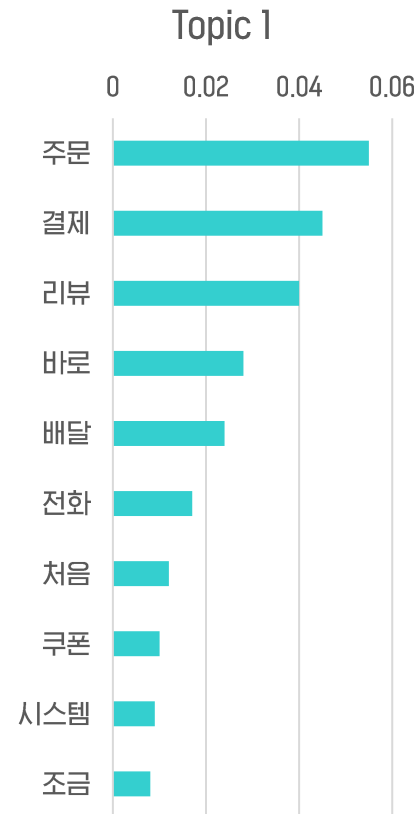
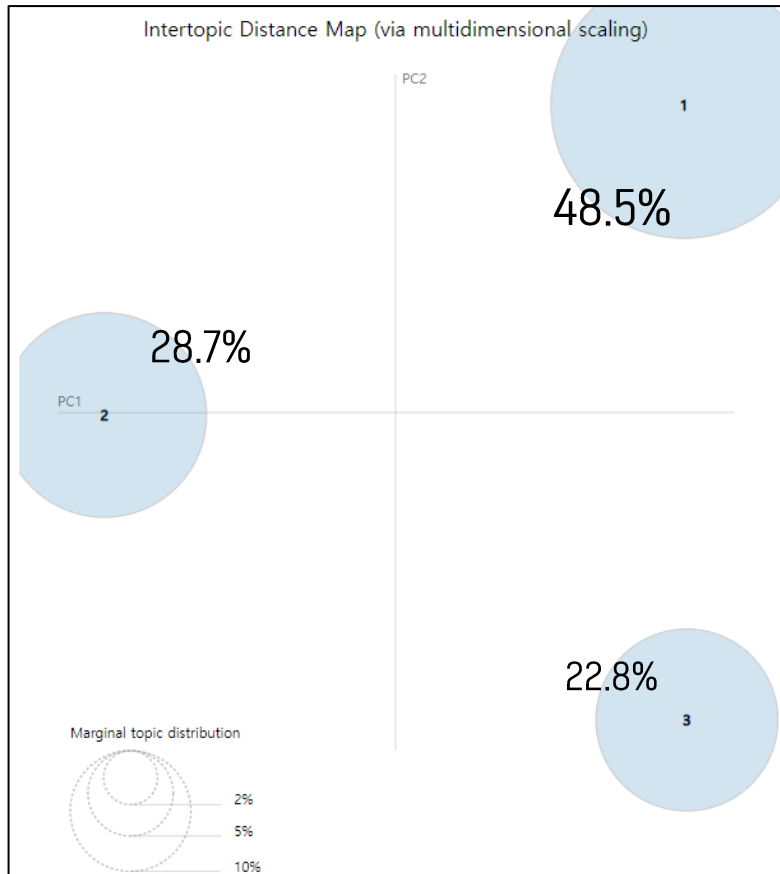
2020-2022 LDA 분석 결과



\*Gibbs sampling: 두개 이상의 확률 변수의 결합 확률 분포로부터 일련의 표본을 생성하는 확률적 알고리즘으로, 결합 확률 분포나 그에 관련된 확률 계산을 근사하기 위해 사용

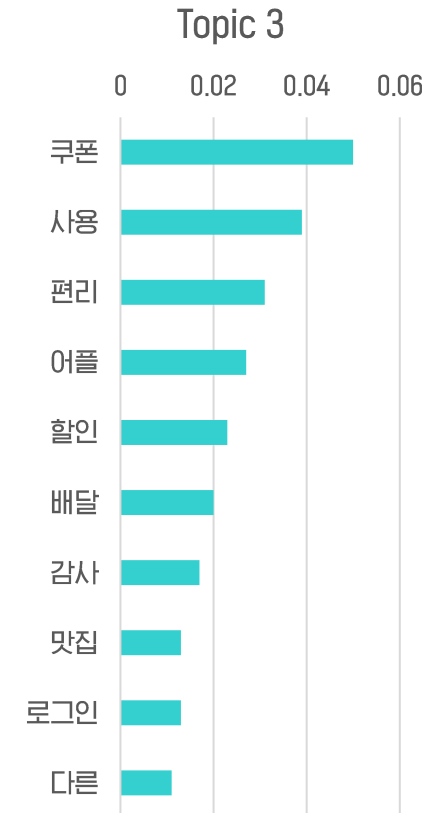
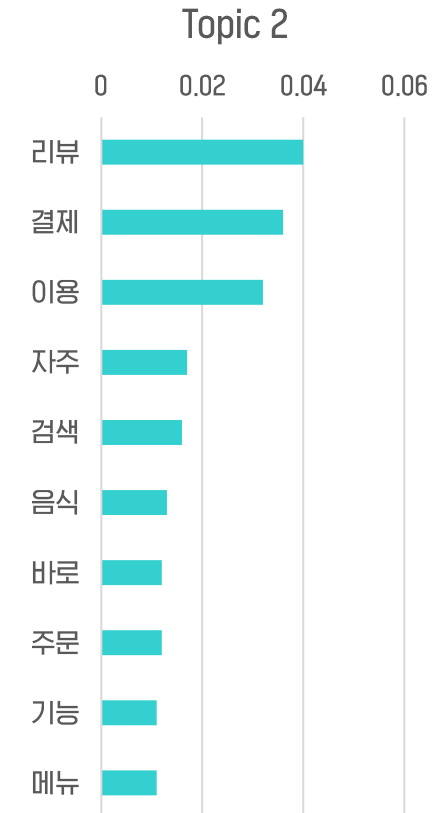
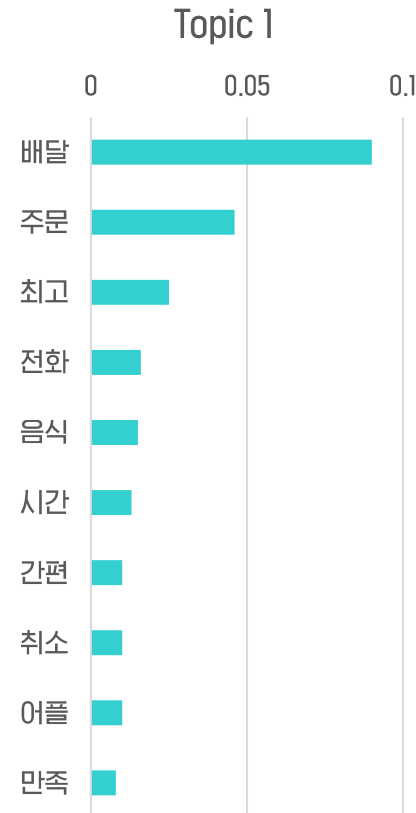
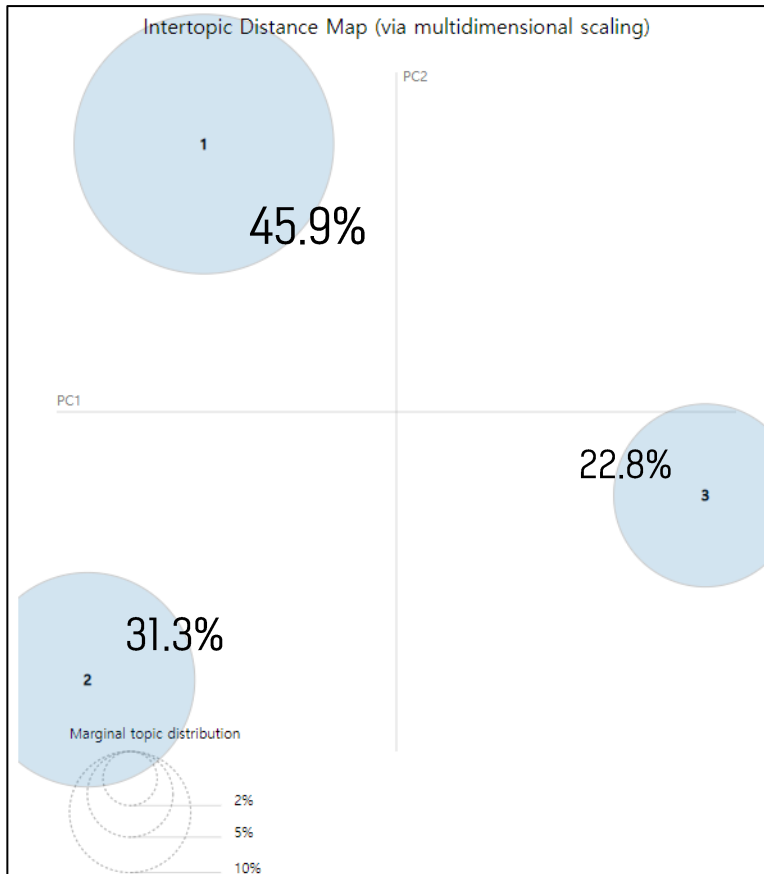
# 2017년

- Topic 1: 바로 결제 주문만 리뷰 작성이 가능한 것에 대한 불만 토로(48.5%)
- Topic 2: 배달 주문에 대한 편리, 간편함(28.7%)
- Topic 3: 최고의 배달 어플이라는 만족감(22.8%)



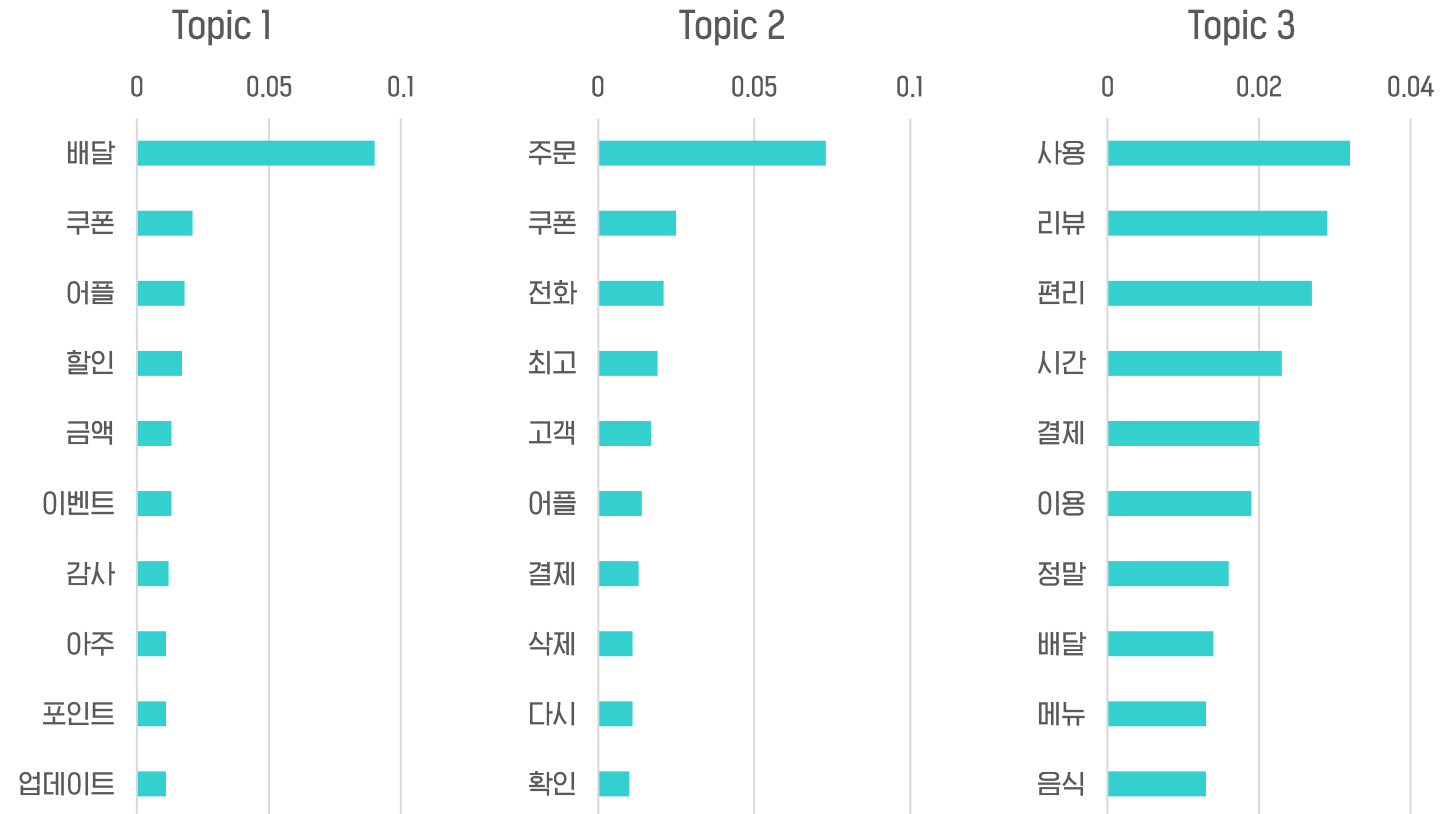
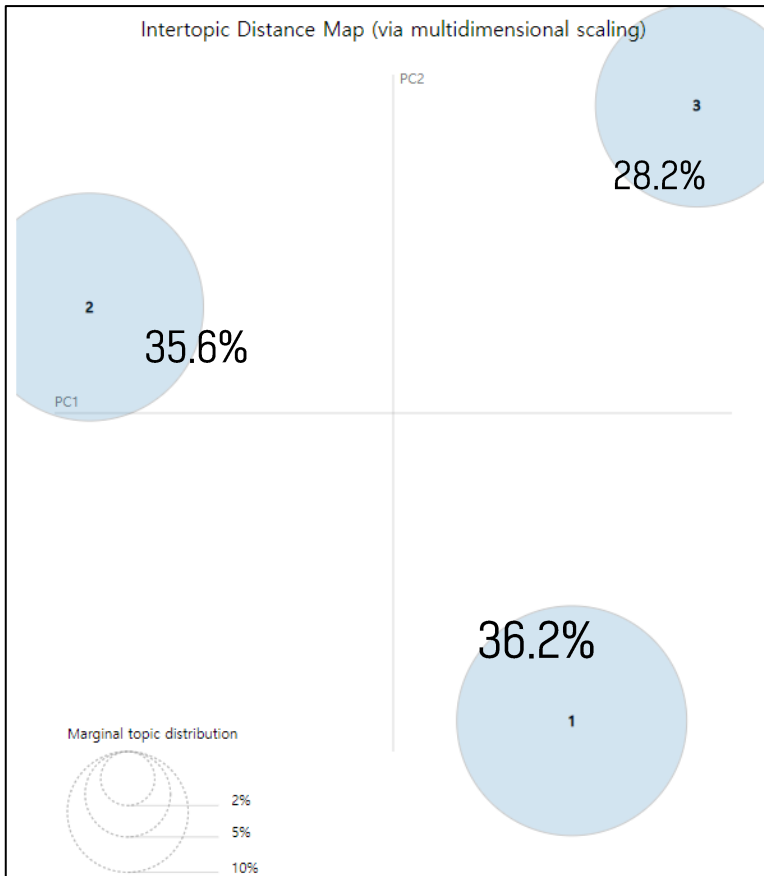
# 2018년

- Topic 1: 음식 배달 주문에 대한 대체적인 만족감(45.9%)
- Topic 2: 2017년에 이어 바로 결제 이슈에 대한 불만 토로(31.3%)
- Topic 3: 쿠폰 지급 및 사용에 대한 감사(22.8%)



# 2019년

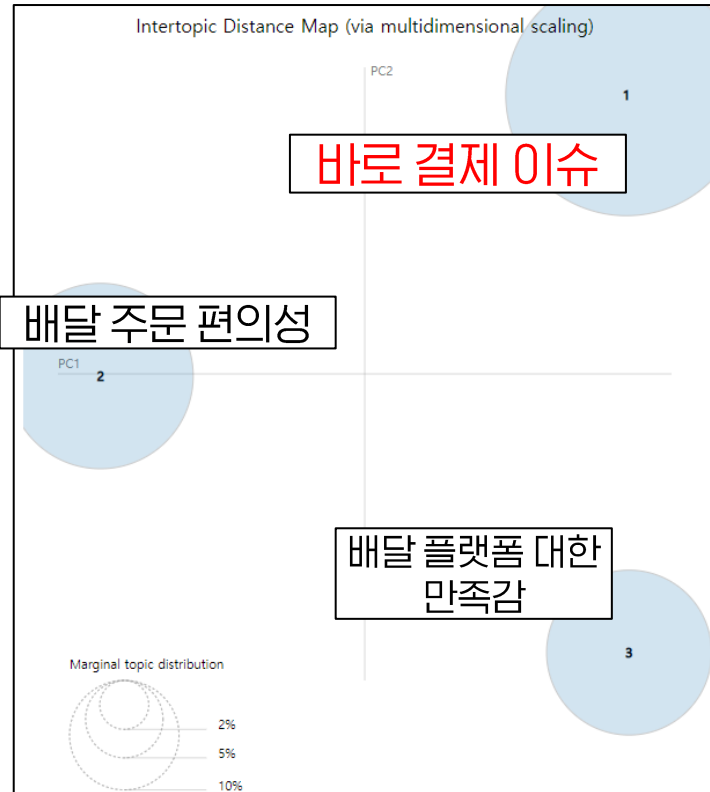
- Topic 1: 쿠폰이벤트로 할인 행사를 진행(36.2%)
- Topic 2: 첫 주문 쿠폰 이벤트 관련 쿠폰 사용이 안된다는 항의(35.6%)
- Topic 3: 배달 플랫폼을 사용함에 있어 많은 리뷰, 배달 시간 등에 대한 만족(28.2%)



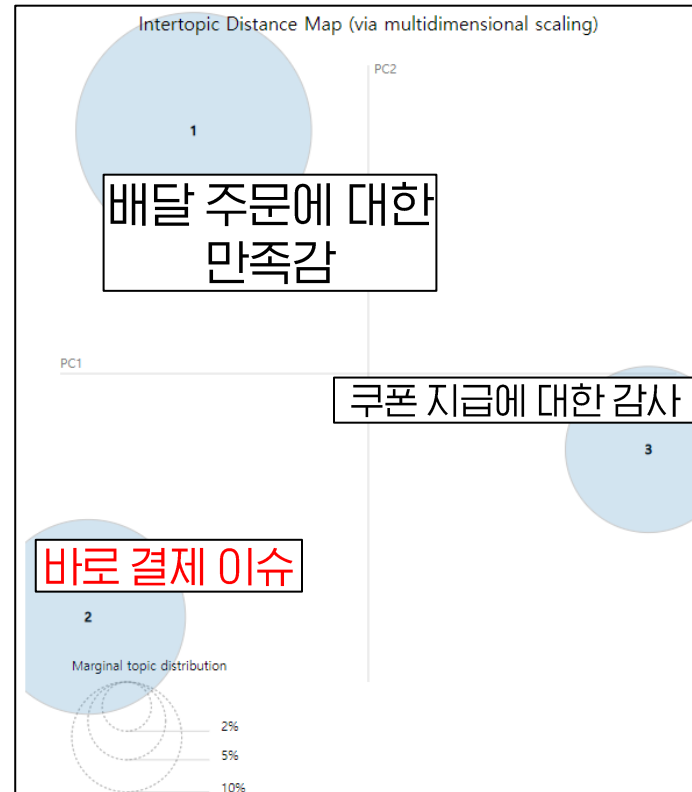


2017년에서 2019년까지,  
바로 결제를 진행한 고객에 대해서만 리뷰를 작성하게 한  
배달의 민족 플랫폼 정책을 비판하는 토픽 외에는  
전반적으로 만족감을 나타내는 토픽

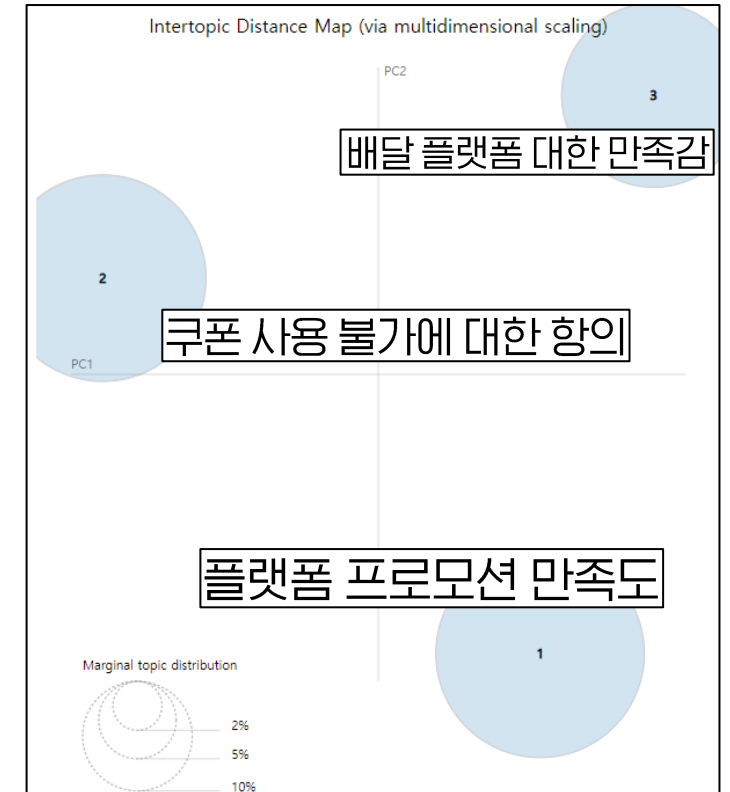
2017 LDA 분석 결과



2018 LDA 분석 결과



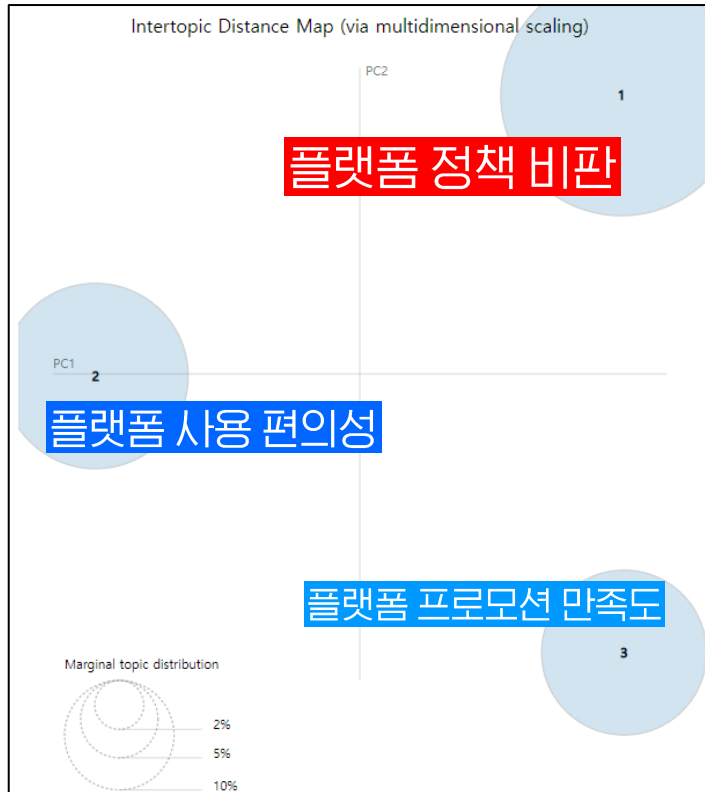
2019 LDA 분석 결과



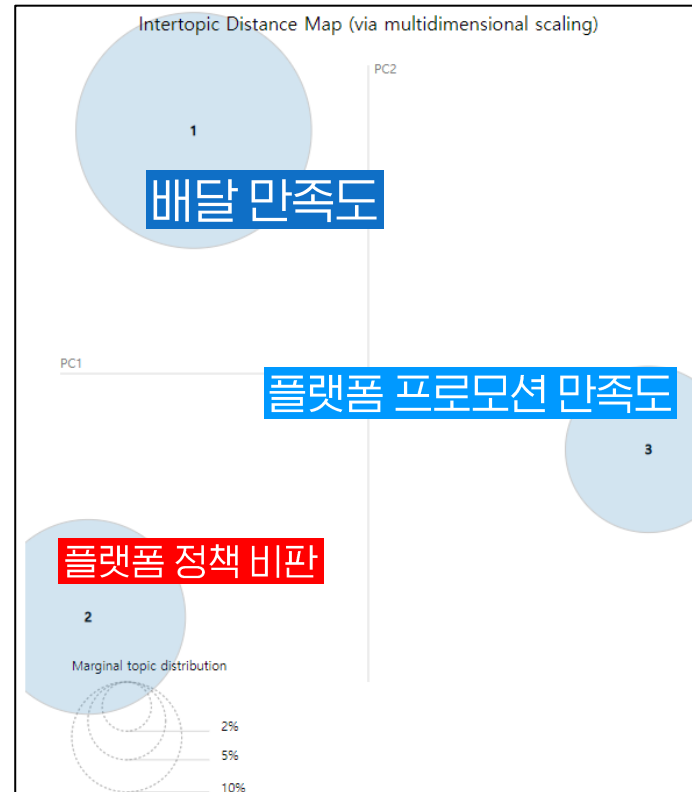
# 토픽을 유형별로 크게 4가지로 분류

1. 사용 편의성
2. 프로모션 만족도
3. 배달 만족도
4. 플랫폼 정책 비판

2017 LDA 분석 결과



2018 LDA 분석 결과



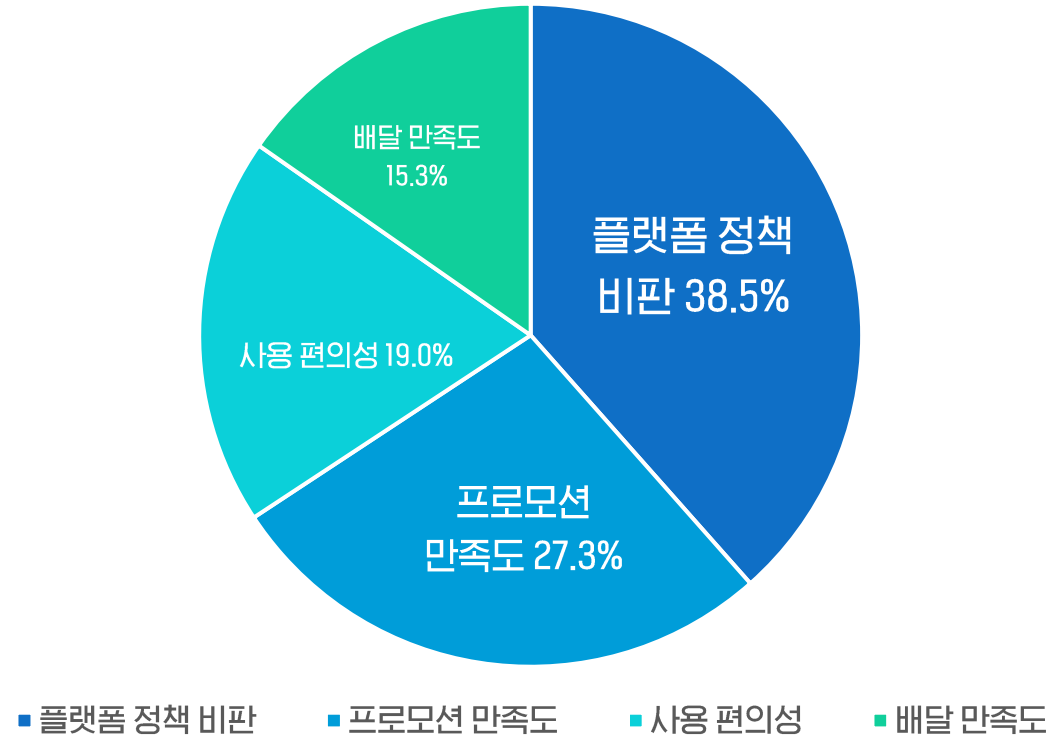
2019 LDA 분석 결과



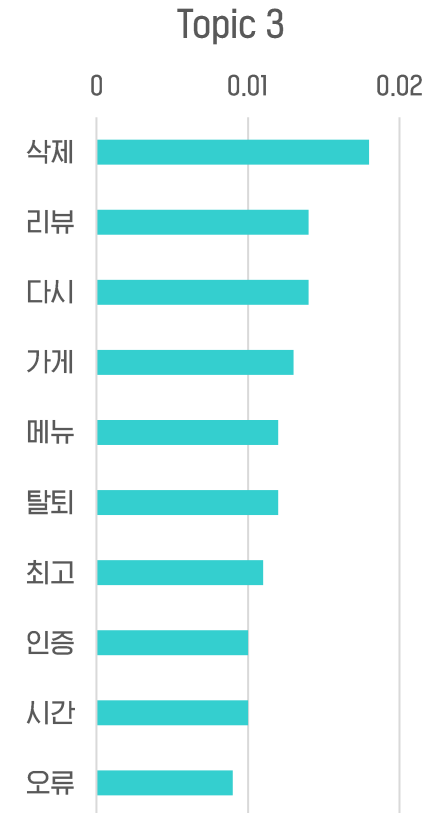
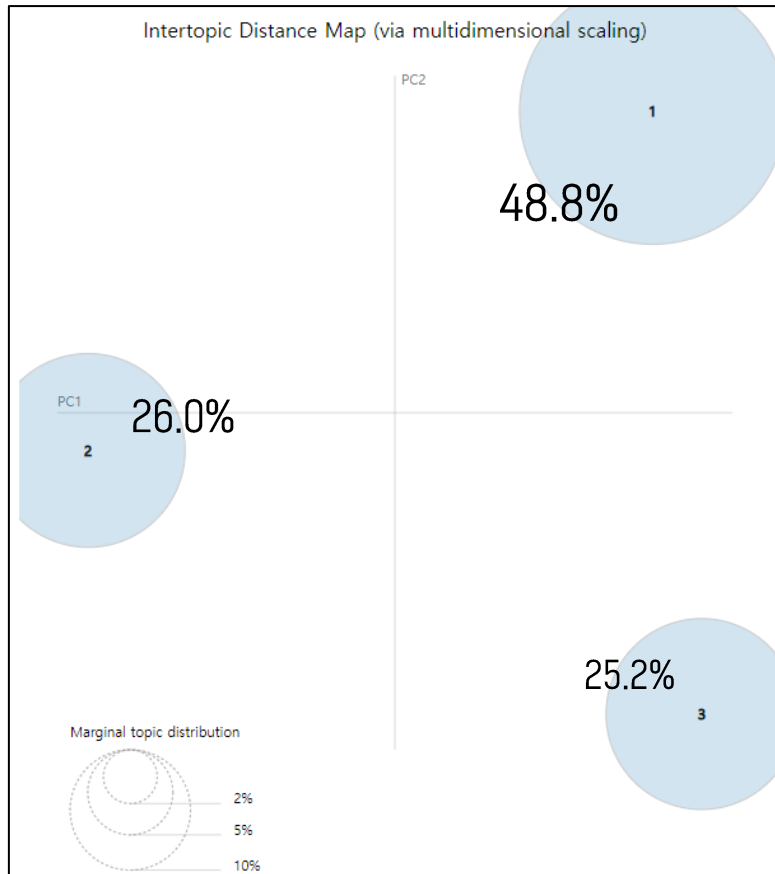
## 토픽의 유형별 분포를 확인한 결과, 비판 관련 토픽이 가장 많은 분포(38.5%)를 차지

- 긍정적인 반응과 관련된 토픽의 종류는 많았지만 플랫폼 정책 비판에 대한 토픽이 가장 많은 분포를 보임.
- 하지만 전반적인 만족을 나타내는 토픽이 61.5%를 차지

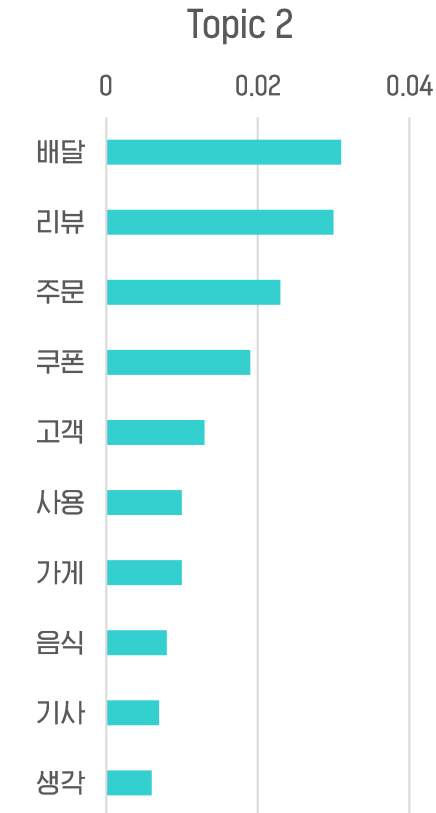
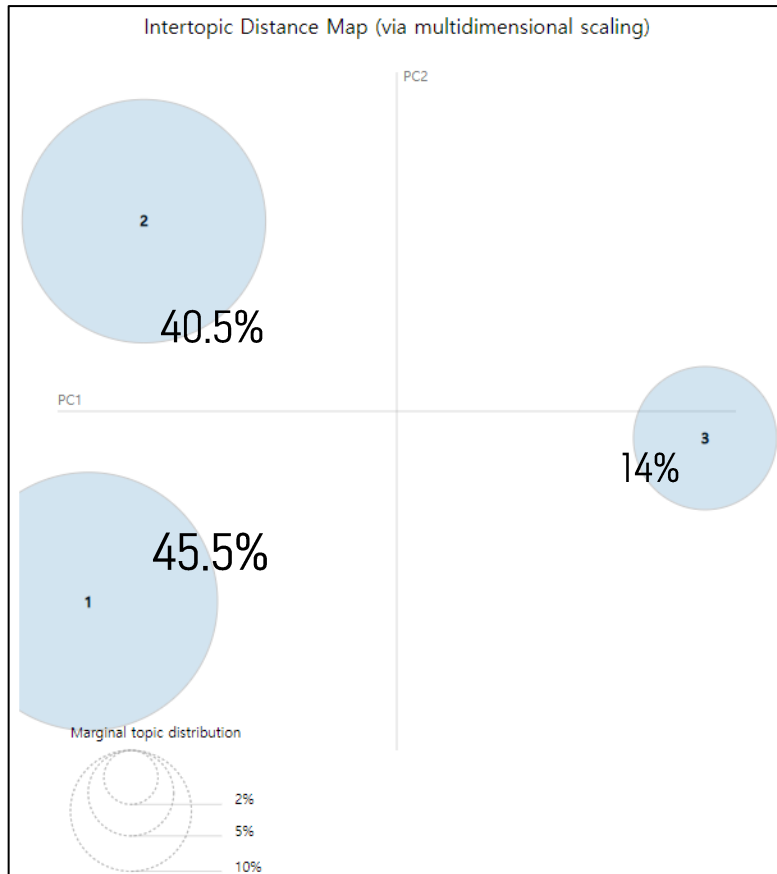
2017-2019 유형별 토픽 분포율



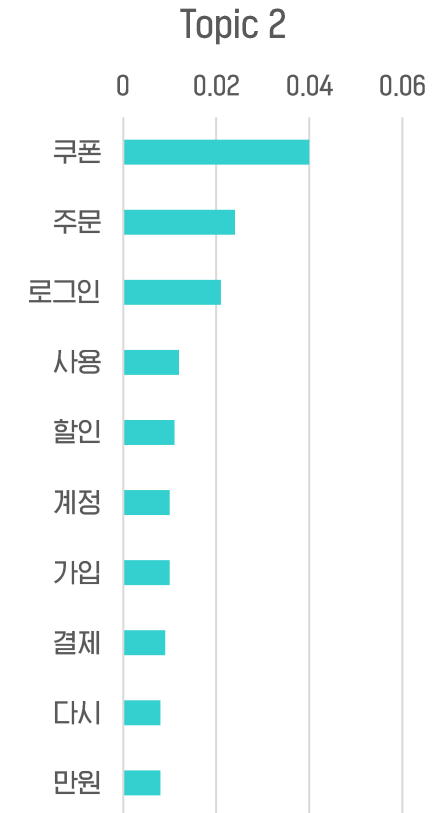
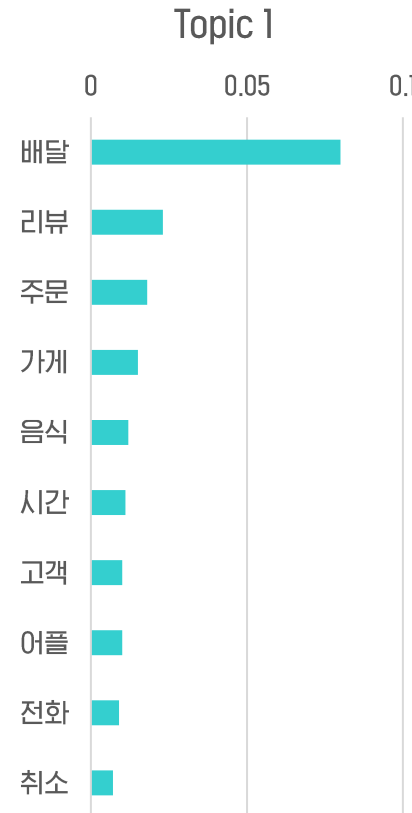
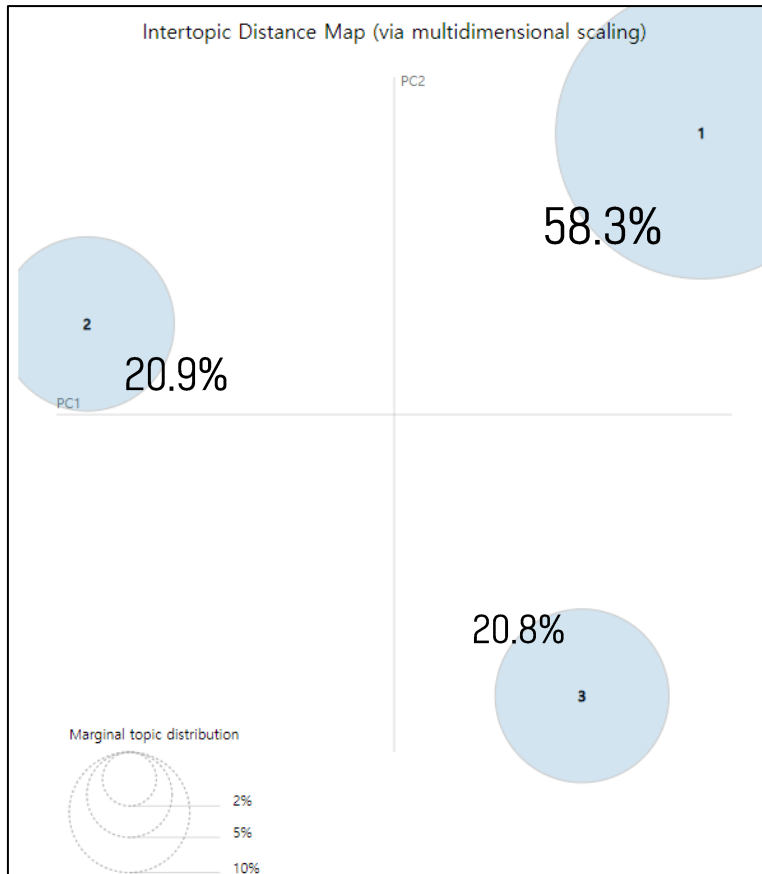
- Topic 1: 오래 걸리는 배달 시간에 대한 불만(48.8%)
- Topic 2: 외국기업 인수 및 독과점에 대한 비판(26.0%)
- Topic 3: 플랫폼에 대한 전반적인 불만(25.2%)



- Topic 1: 배민1의 출시와 함께 해당 플랫폼의 늦은 배달시간에 대한 불만(45.5%)
- Topic 2: 허위리뷰 및 리뷰삭제 이슈, 쿠폰 사용 오류 등의 불만(40.5%)
- Topic 3: 결제카드의 인증 오류에 대한 불만(14.0%)

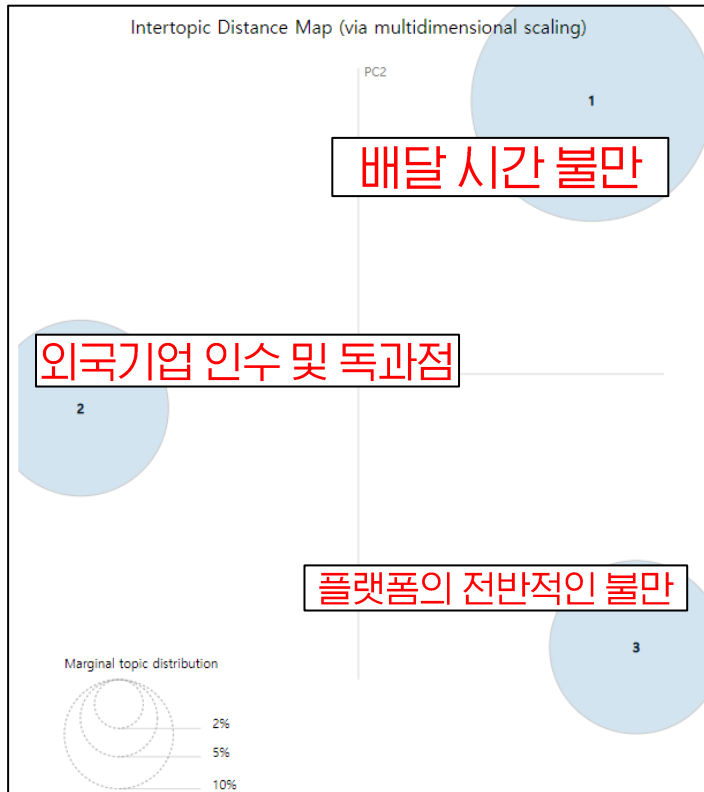


- Topic 1: 오래 걸리는 배달시간에 대한 불만(58.3%)
- Topic 2: 쿠폰 수령에 대한 로그인 문제(20.9%)
- Topic 3: 특정 메뉴에 대한 리뷰만 볼 수 있도록 기능 개선을 요구(20.8%)

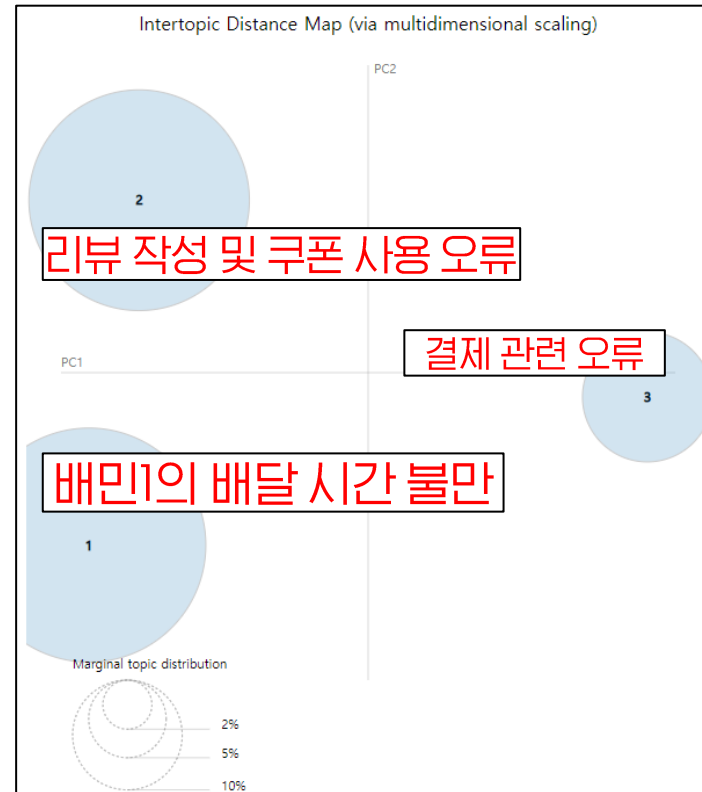


2020년에서 2022년까지,  
다양한 불만 관련 토픽이 결과로 나왔으며  
가장 눈의 띄는 것은 **배달시간에 대한 불만**  
그 외, 기업에 대한 비판과 시스템 오류에 대한 항의 등의 토픽이 형성

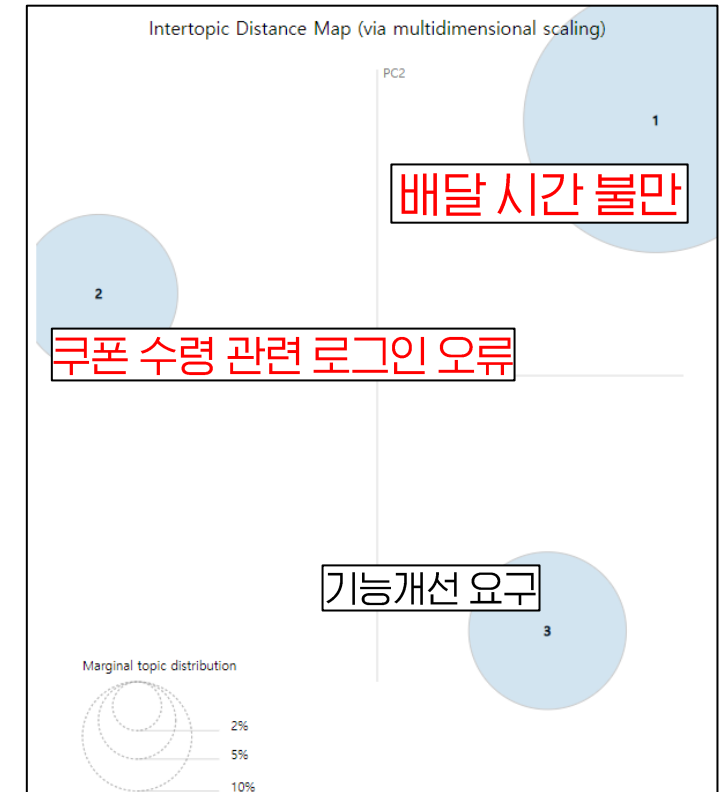
2020 LDA 분석 결과



2021 LDA 분석 결과



2022 LDA 분석 결과



# 토픽을 유형별로 크게 5가지로 분류

1. 배달 시간 불만
2. 시스템 오류 항의
3. 플랫폼 비판
4. 기업의 시장 비판
5. 기능개선 요구

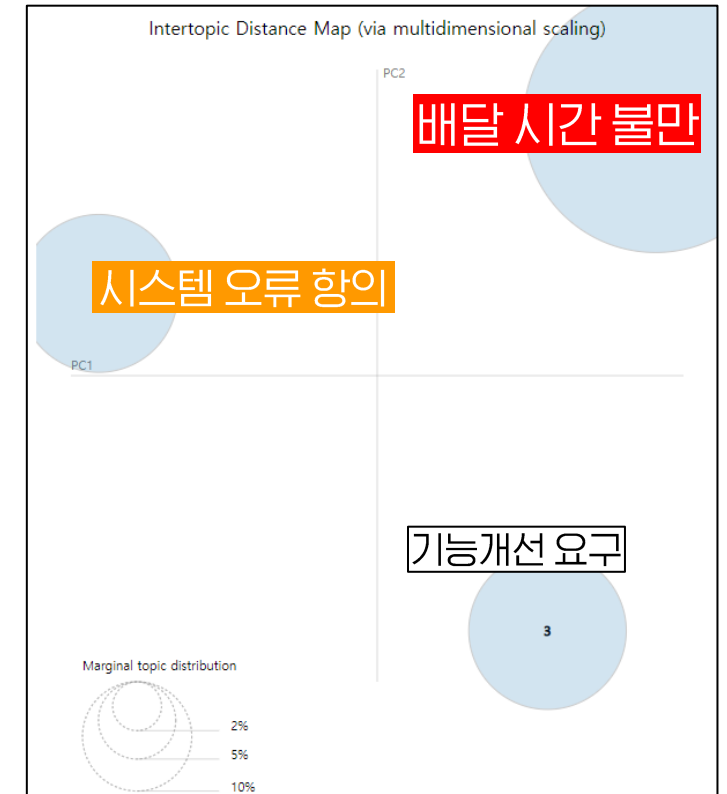
2020 LDA 분석 결과



2021 LDA 분석 결과



2022 LDA 분석 결과

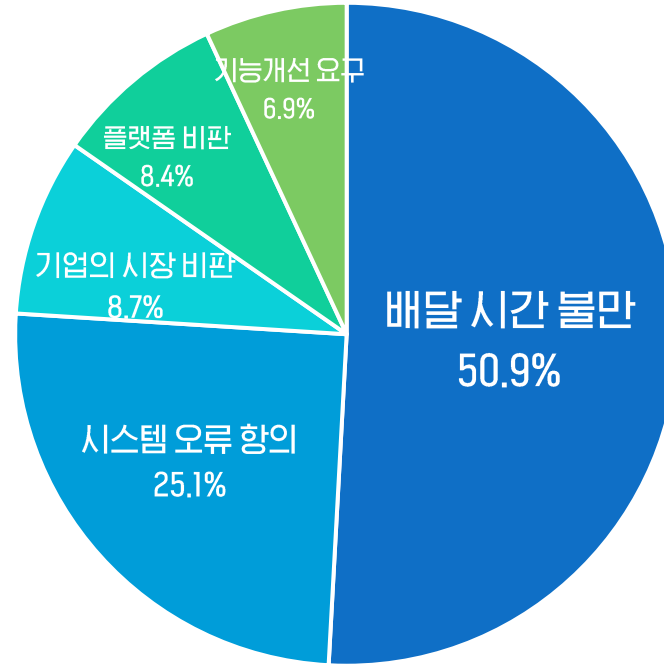




## 토픽의 유형별 분포를 확인한 결과, 배달 시간에 대한 불만이 가장 많은 분포(50.9%)를 차지

- 배달 시간 불만이 50%가 넘는 분포를 보였으며, 시스템 오류에 대한 항의도 약 25%의 분포를 차지
- 다양한 주제에 대한 비판이 해당 기간의 토픽을 형성

2020-2022 유형별 토픽 분포율



■ 배달 시간 불만 ■ 시스템 오류 항의 ■ 기업의 시장 비판 ■ 플랫폼 비판 ■ 기능개선 요구

LDA분석은 문서 데이터의 핵심 주제(토픽)를 파악하는 것에 목표를 두고 있다.  
키워드의 빈도 기반으로 분석이 진행되는 TF-IDF 분석과는 달리  
확률기반 분포를 통해 진행되는 LDA분석에서는  
또 다른 이슈 사항들을 파악할 수 있었다.

2017-2019 LDA 분석 결과 코로나 발생 이후 느린 배달 시간과 2020-2022 LDA 분석 결과

배달의 민족의 기업 인수합병 및 독과점 문제로 인한 2020년 평점의 하락,  
이후 배달의 민족의 시스템 오류와 전반적인 플랫폼에 대한 문제에 따른  
소비자들의 인식을 알 수 있었다.

또한 코로나 이전에는 플랫폼 정책 비판에 대한 토픽을 제외한  
대부분 인식의 만족도가 높은 토픽이 결과로 나왔지만  
코로나 발생 이후, 다양한 문제에 따른 소비자들의 불만과 비판이 확대되어  
분석 결과로 나온 모든 토픽이 비판과 관련된 토픽인 점도 눈여겨볼 결과이다.



# CHAPTER.3

## 결론

해당 보고서는 배달 시장의 동향 파악 및 인식의 변화를 분석하였으며 분석은 크게 **정형데이터(평점)**와 **비정형데이터(리뷰)**로 진행하였고 이는 크게 4단계로 구분할 수 있다.

1. Box-plot을 활용하여 평점데이터의 연도별 분포를 확인
2. 비교대상의 기간이 유의미한 차이가 있다라는 가설을 설정
3. 통계적 기법을 통해 해당 가설을 채택
4. Text-mining 진행, 유의미한 차이에 대해 정밀 분석

단순 평점으로 확인해보면, 코로나 전후에 대한 차이가 극명하게 났지만 평점과 함께 동반된 리뷰데이터를 분석한 결과, **연도별로 크고 작은 이슈들이 있었지만 기준을 세운 연도에 대한 인식 차이가 있다**는 것을 확인할 수 있었다.

해당 보고서의 분석의 목표는 이러한 결과를 얻기 위함이다.

*“평점이 좋았던 연도에는 무슨 이슈사항이 있었는지”*  
*“그에 따른 비판의 인식은 없었는지”*  
*“코로나 발생 이후에 어떤 문제가 낮은 평점의 원인이 되었는지”*

평점만으로 알 수 없는 사람들의 인식을 파악하여 '배달의 민족' 플랫폼의 이슈사항과 그에 따른 세부적인 인식의 변화를 알 수 있었다.

하지만 해당 보고서는 한계점 또한 당연히 존재한다.

Text-mining에 대한 기술과 성능은 지속적으로 발전하고 있기 때문에 해당 분석으로는 소비자 인식을 100% 결론을 내릴 수 없다.

토픽에 대한 감성분석, 알고리즘별 성능 비교 등 보다 정교하고 의미 있는 연구들이 충분히 존재하므로 향후 해당 보고서의 결과를 바탕으로 **감성분석과 다양한 연구 방법을 추가**하여 소비자들의 세부적인 감성 변화를 파악하고 세밀한 결론을 도출할 예정이다.

# E n d O f D o c u m e n t

## Contact Us

Website URL <http://bigdata.emforce.co.kr>

T 02. 6177. 1871 eMAIL [khbak@emforce.co.kr](mailto:khbak@emforce.co.kr)