

A comparative analysis of travelers' online reviews  
among China, South Korea, and USA  
using sentiment analysis in the era of the COVID-19 pandemic

2021.06.02

부산대학교 경영대학 경영학과  
홍준우, 원종관, 홍태호



## 목 차

1. 서론
2. 선행연구
3. 연구 프레임워크
4. 실험 및 결과
  - ① 데이터 수집 및 분석
  - ② LDA를 통한 토픽 추출 및 비교
  - ③ 감성분석을 통한 국가별 비교
5. 결론

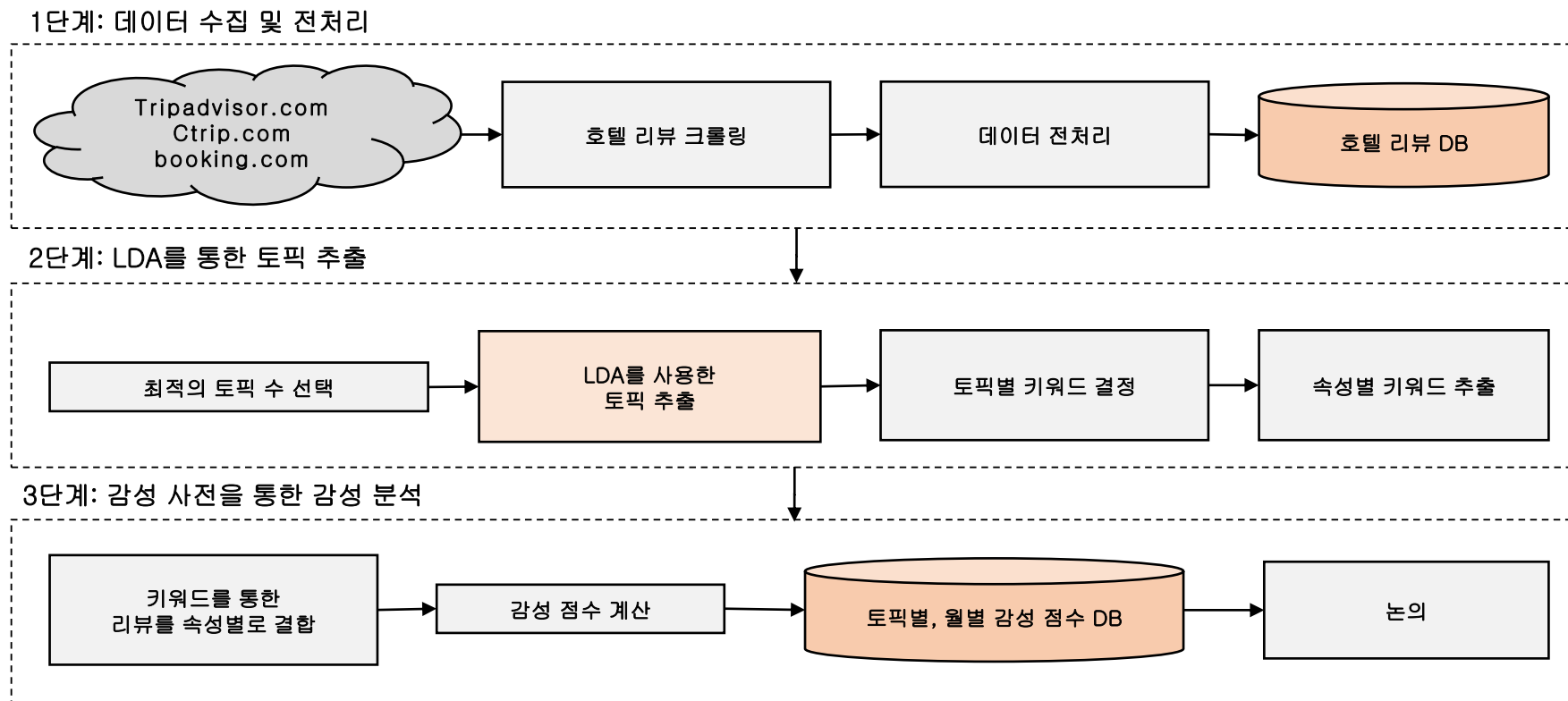
## 1. 서론

- 다양한 플랫폼의 등장으로 전자상거래 시장은 성장했으며, ‘코로나19’의 확산으로 성장은 가속화되었다(AI-maaitah et al., 2021).
- 경험을 근거로 작성된 온라인 리뷰는 플랫폼을 통해 공유되며, 소비자의 잠재적인 정보탐색과 구매에 영향을 미친다(Yao et al., 2020).
- 온라인 리뷰의 적절한 활용은 입장별 이점을 발생시키며, ‘코로나19’의 확산으로 더욱더 중요해졌다(Gretzel et al., 2020).
  - 고객의 경우, 상품의 탐색, 구매 등 의사결정에 이점을 얻는다(Cheng and Ho., 2015)
  - 기업의 경우, 온라인 리뷰 분석을 통해 매출 증가 등의 이점을 얻을 수 있다(Li et al., 2013).
- 한편, ‘코로나19’ 확진자 수는 국가별로 다르게 발생하며, 서로 다른 영향을 미치고 있다(Fanelli and Piazza., 2020).
- ‘코로나19’의 확산과 국가별 여행제한 정책 등으로 인해, 관광산업(호텔, 레스토랑, 쇼핑몰 등)은 “이미 붕괴에 직면” 하거나 “생존을 위한 투쟁 중”인 상태이다(Jiang and Wen., 2020).
- 따라서, 본 연구에서는 ‘코로나19’ 시대 온라인 리뷰를 통한 국가별 호텔 리뷰를 비교하는 연구를 진행하였다.

## 2. 선행연구

- LDA (Latent Dirichlet allocation)
  - 토픽 수(k) 결정 방법
    - ✓ 문장에서 단어들의 관련성을 토픽별로 분류하는 기술로써, 추출할 토픽 수(k)를 지정한다.
    - ✓ 최적의 토픽 수를 결정하기 위해 교차 검증(cross-validation)과 혼잡도(perplex)를 활용하였다(Blei et al., 2003, Hong et al., 2018)
  - LDA기법을 통한 특정 관점별 토픽 추출 연구
    - ✓ 적은 수의 데이터를 LDA기법에 적용하였다(Tong and Zhang., 2016).
    - ✓ LDA기법을 적용하여 호텔 산업의 3가지 관점('위치', '서비스', '가치')을 추출했으며(Poria et al., 2016), 호텔 리뷰의 5가지 관점('청결', '편안함', '서비스', '음식', '위치')에 적용하여 주제를 추출하였다(Priyantina & Sarno., 2019).
- 감성 분석(Sentiment Analysis)
  - 온라인 리뷰를 활용하여 감성 점수를 계산하는 방법으로 '어휘기반', '기계학습 기반', '하이브리드 방법' 등이 있으며(Alamoodi et al., 2020), 이러한 방법을 통해, 사람의 감성과 생각을 추출할 수 있음을 입증하였다(Medhat et al., 2014).
  - 감성 분석을 통해 긍정 혹은 부정의 의견을 표명하는지 여부를 정량화 하였다(Li et al., 2010).
  - 어휘 기반 감성 분석을 통해 관광객의 리뷰와 평점을 분석했으며, 이를 통해, 관광지 추천 시스템을 개발하였다(Shafqat & Byun., 2020).

### 3. 연구 프레임워크



< 그림 1>. 연구 프레임워크

## 4. 실험 및 결과: ①. 데이터 수집 및 분석

- 데이터 수집:
  - 중국: 4, 5성급 호텔 + Ctrip.com
  - 미국: 4, 5성급 호텔 + Tripadvisor.com + booking.com

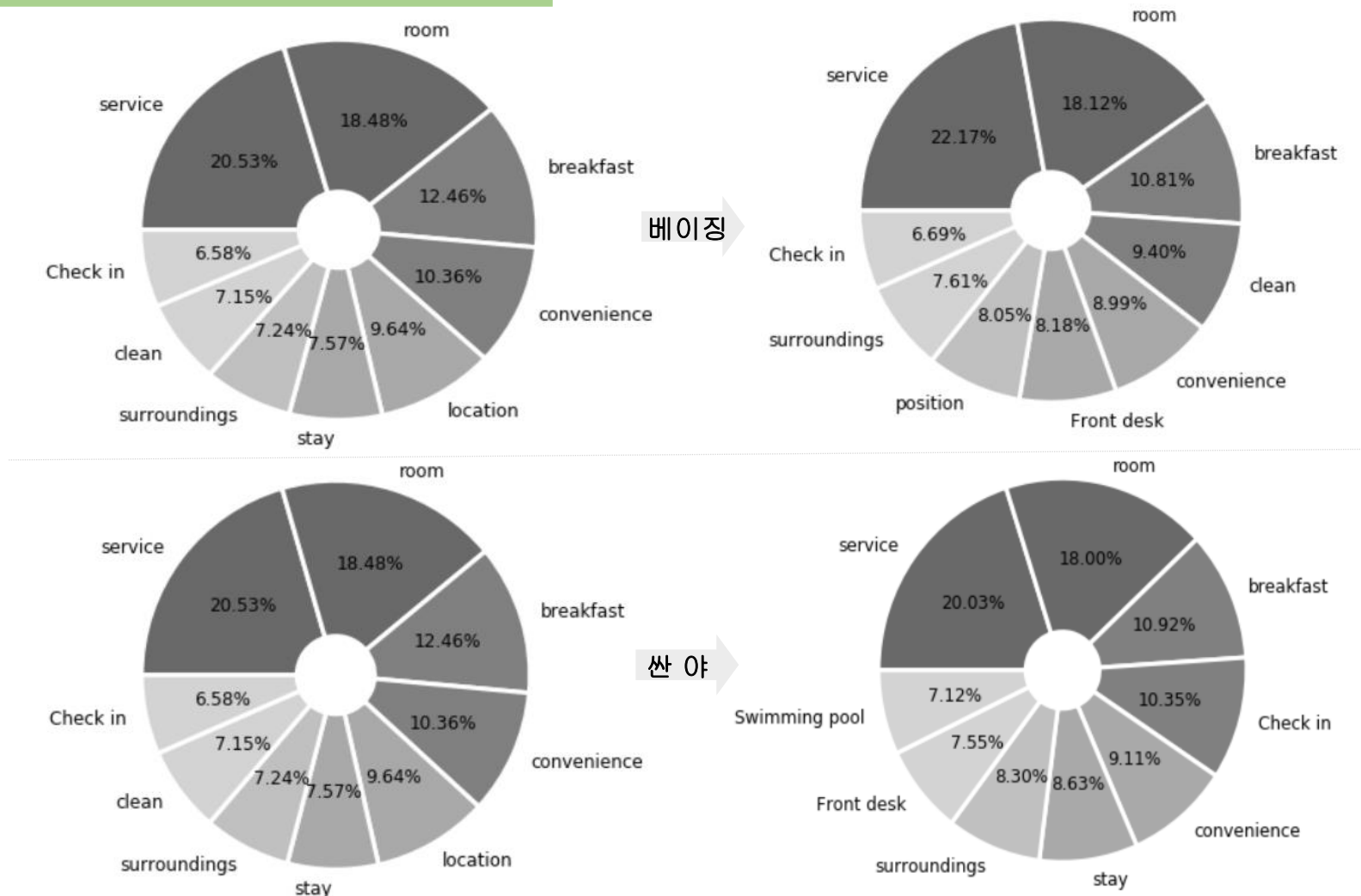
- 데이터 분석

국가	중국				미국			
도시	베이징		싼야		뉴욕		하와이	
기간 (코로나19)	이전	이후	이전	이후	이전	이후	이전	이후
평균점수	4.74	4.76	4.71	4.68	4.58	4.34	4.19	4.31
리뷰 수	37,941	19,515	84,165	80,382	9,661	1,691	4,201	572
총 리뷰 수	57,456		164,563		11,352		4,773	

- 이전 : 2019.01.01 - 2019.11.31
- 이후 : 2020.04.01 - 2021.02.28

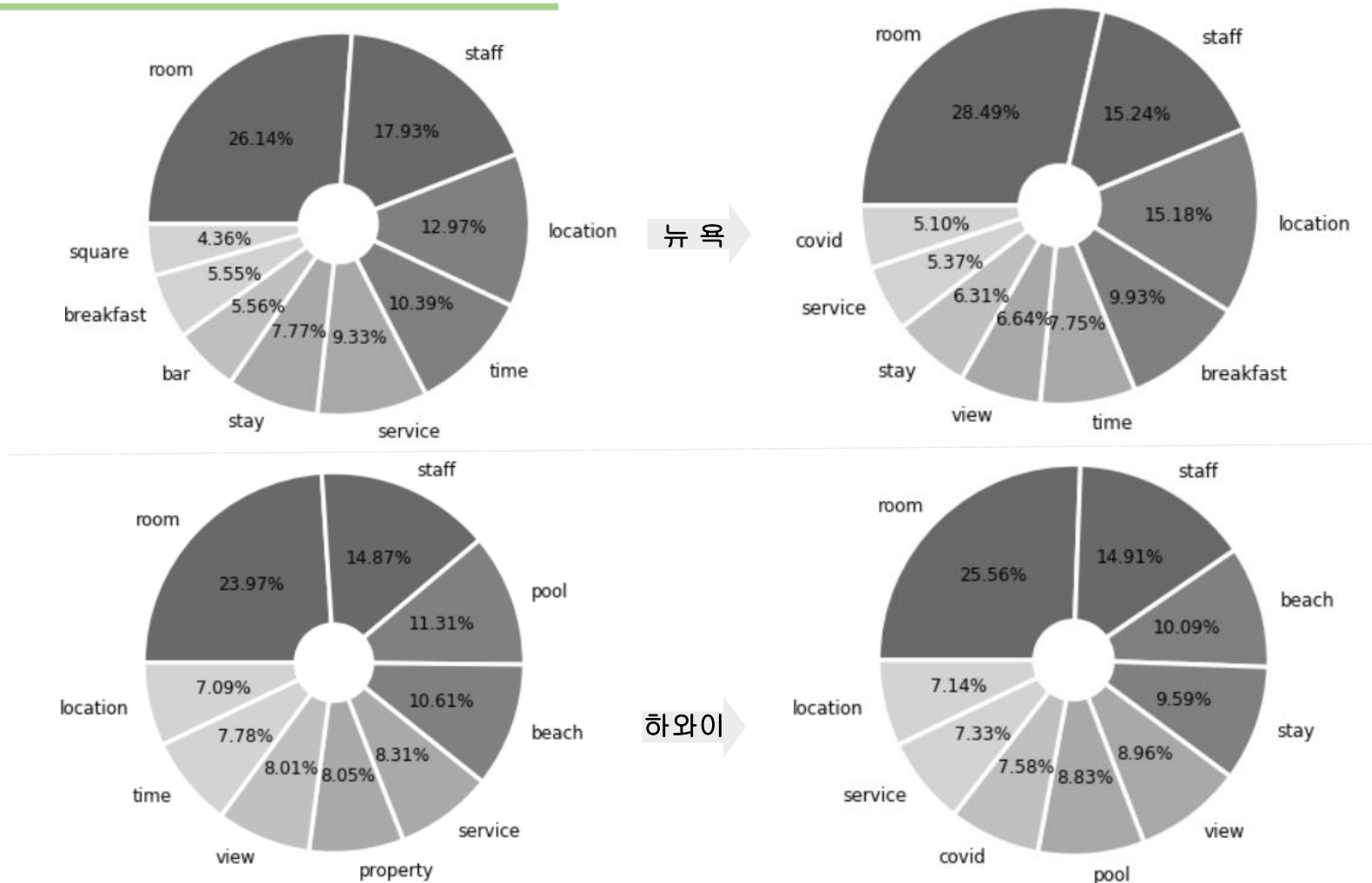
< 표 1>. 데이터 세트

#### 4. 실험 및 결과: ①. 데이터 수집 및 분석



<그림 3>. 코로나19 전/후 별 빈도분석 (중국-베이징, 싼야)

#### 4. 실험 및 결과: ①. 데이터 수집 및 분석



<그림 3>. 코로나19 전/후 별 빈도분석 (미국-뉴욕,하와이)

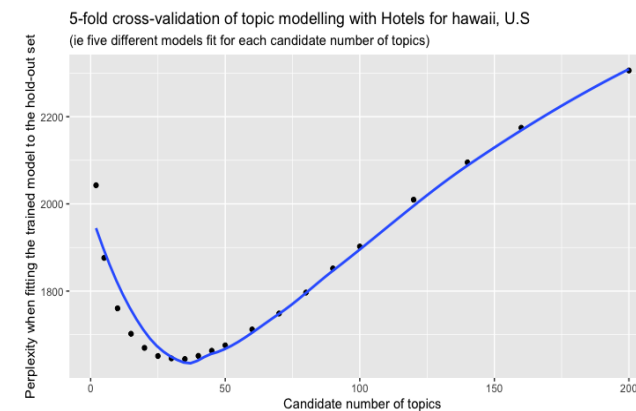
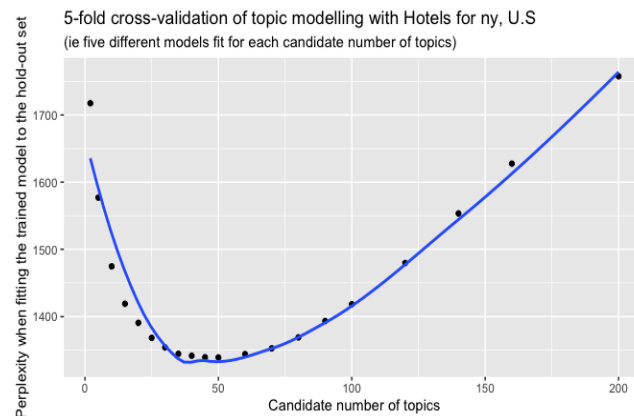


## 4. 실험 및 결과: ②. LDA를 통한 토픽 추출 및 비교

- LDA를 통한 토픽 추출을 위해 토픽 수(k)를 결정했으며, 본 연구에서는 5겹 교차 검증(5-fold cross-validation)과 혼잡도(perplexity)를 활용한다.

국가	미국		중국	
도시	뉴욕	하와이	베이징	싼야
토픽 수	45	30	40	50

<표 2>. 도시별 토픽 수



<그림 2>. 토픽 수에 따른 적합도 비 (미국- 뉴욕, 하와이)

#### 4. 실험 및 결과: ②. LDA를 통한 토픽 추출 및 비교

토픽ID	토픽	단어	상위 10개 키워드
1	서비스	meal	breakfast, free, buffet, bar, drinks, restaurant, food, dinner, live
2	객실	room	room(s), balcony, view(s), door, outdated, bed, pool, bedroom
...	...	...	...
29	수면의 질	noise	relax(ing), hear, loud, quiet, close, air, partial, noise, night
30	위치	accessibility	parking, walk, easy, drive, far, bay, short, main, town, street

<표 3>. 토픽 및 단어 추출 결과 (미국-하와이)

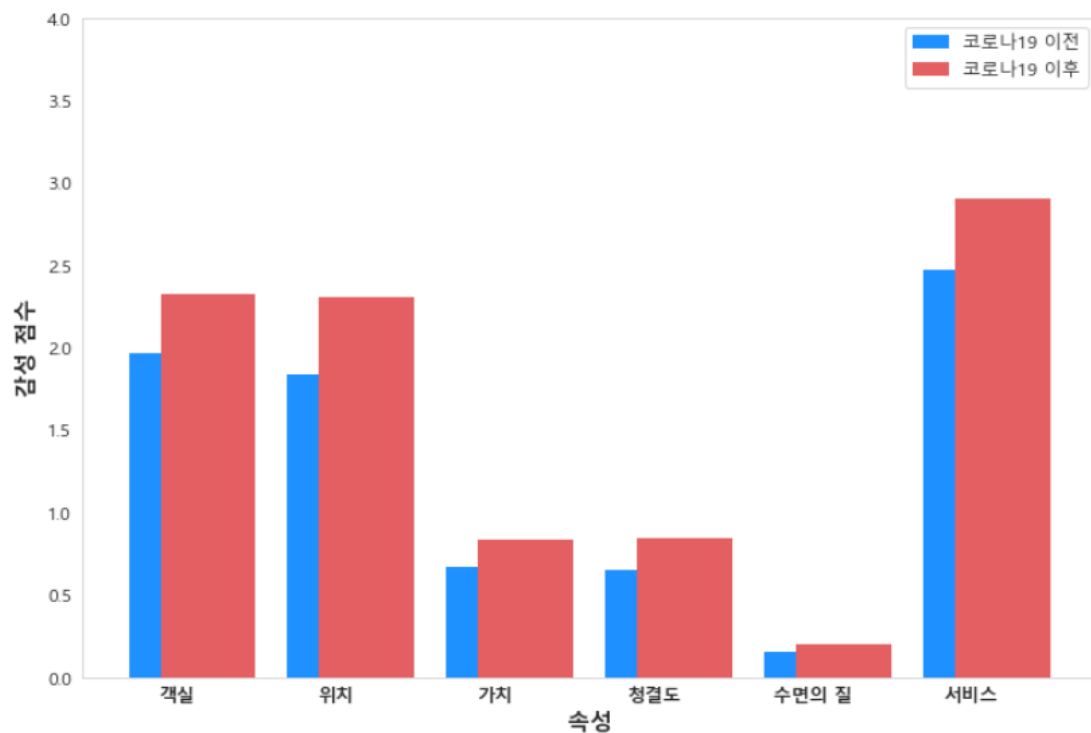
도시	속성	키워드
뉴욕	가치	cheap, value, expensive, fee(s), worth, bar, parking fee, valet, price
	객실	room, desk, bed, floor, place, bathroom, view, small, large, big, air, Conditioner
	위치	walking, location, time, night, square, central, park, times, front, close, city, entrance
	청결도	clean, dirty, covid, smell, trash, bathroom, detergent, area, poor, concern
	수면의 질	noise, noisy, quiet, sound, atmosphere, comfortable, deep, condition, experience, sleep
	서비스	staff, helpful, friendly, check, lobby, comfortable, time, desk, service, breakfast, bar, front, experience, food, menu, trip, service, wine, coffee, beverages, lobby, snack, furniture
하와이	가치	free, easy, worth, fee(s), value, price, pay, extra, rental
	객실	sunset(s), warm, view(s), room(s), small, space, pool(s), chair(s), facilities, outdated, suite, old, options, main, private, space, daily
	위치	time, spent, location, close, park, easy, access, oceanfront, rays, bay, sunsets, forward, construction, looking, seaside
	청결도	covid, clean, dirty, negative, bad, cleaning, evening, dirt, issue, full
	수면의 질	relax(ing), hear, loud, quiet, close, air, partial, noise, night
	서비스	lobby, reviews, buffet, bar, drinks, happy, restaurant(s), food, dinner, live, breakfast, spa, service(s), amenities, accommodations, options, front desk, people, help, care, housekeeping, shuttle, check, welcoming, professional, housekeeping, concierge, attentive, busy, enjoy, kids, children, families, available, taken, working, quick, job

<표 4>. 6가지 토픽별 키워드 (미국-뉴욕, 하와이)

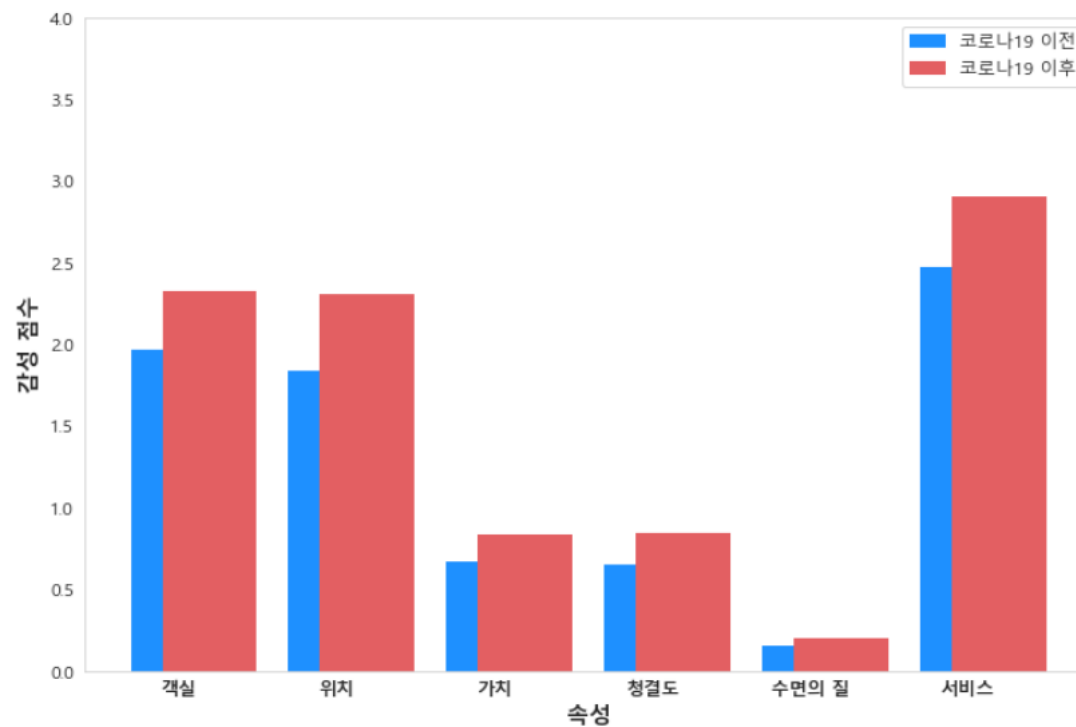
#### 4. 실험 및 결과: ③. 감성분석을 통한 국가별 비교

## 중 국

베이징



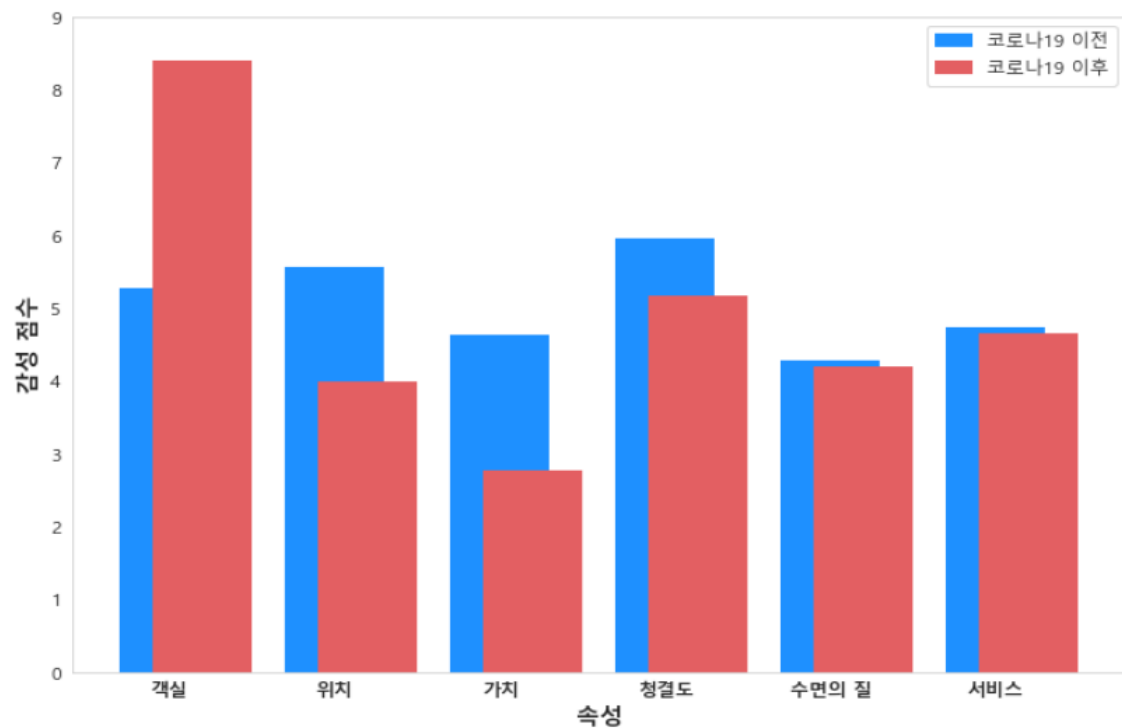
싼야



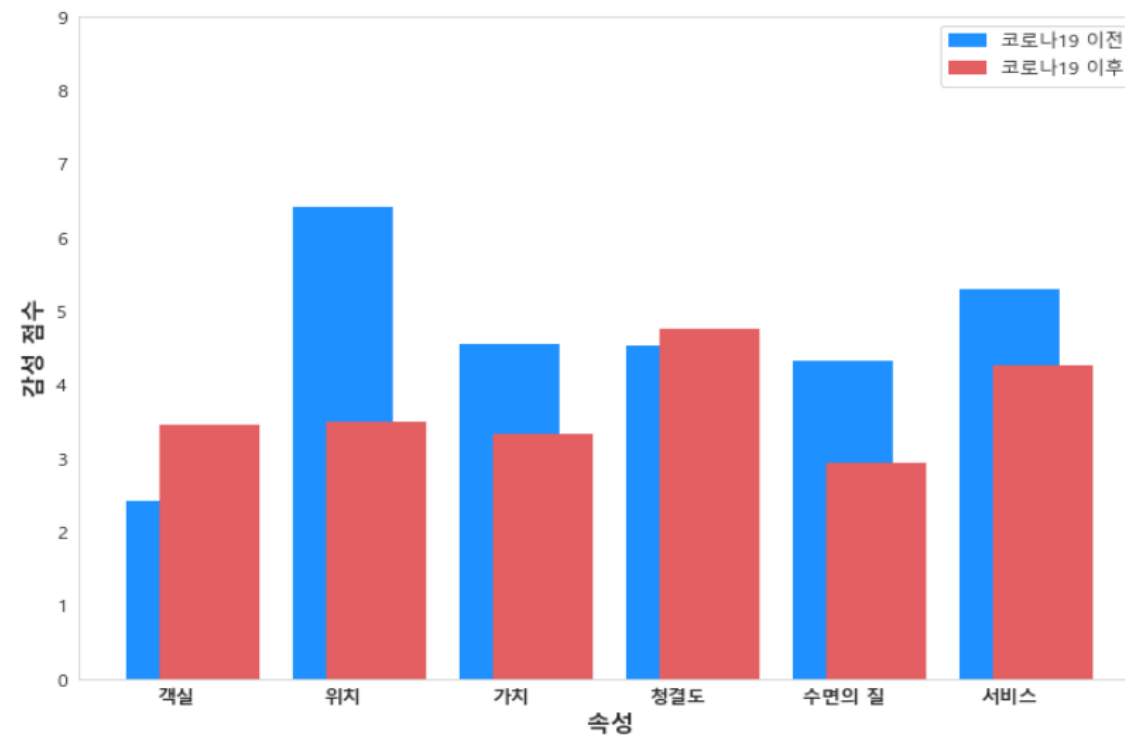
#### 4. 실험 및 결과: ③. 감성분석을 통한 국가별 비교

## 미 국

뉴욕



하와이



## 5. 결 론

- 결론
  - 국가별 도시의 온라인 리뷰를 토픽모델링을 통해 6가지 속성으로 분류하였고, 이를 코로나19 전과 후로 분류하여 6가지 속성을 관점으로 비교 분석하였다.
  - 코로나19 전/후를 국가별 도시 비교:
    - ✓ 중국의 베이징, 쑤아: 베이징의 “객실”, “위치”, “서비스”, “청결도”의 측면은 만족도가 감소, 나머지는 속성 및 도시는 전반적으로 증가
    - ✓ 미국의 뉴욕, 하와이: 뉴욕의 “객실”, 하와이의 “객실”, “청결도 ” 측면은 만족도가 증가, 나머지는 전반적으로 감소
  - 국가별 도시의 만족도 비교를 통해, 베이징과 뉴욕은 만족도가 감소했으며, 쑤아와 하와이는 만족도가 증가하였다.  
이를 통해, 코로나19 이후 도심지역보다 섬지역의 도시가 만족도가 증가하였고, 이는 내국인에 의해 관광이 이루어져 긍정적인 만족도를 낳았을 것으로 예상한다.
- 한계점 및 시사점
  - 미국과 중국의 대표적인 두 곳의 도시의 온라인 호텔 리뷰를 활용하여 연구를 진행하였다. 하지만 온라인 호텔 리뷰를 불균형하게 수집하였으며, 향후 연구에서는 다양한 웹사이트에서 추가적으로 데이터를 수집할 필요가 있다.
  - 감성 점수를 계산할 때, 언어별 감성사전을 활용하였으며, 향후 연구에서는 딥러닝 기법 등의 감성 점수를 계산하는 새로운 접근법이 필요하다.
  - 미국과 중국의 대표적 관광도시를 대상으로 본 연구를 진행했으며, 향후 연구에서는 한국 등 다양한 국가를 추가한 연구가 필요하다.

# Q & A

감사합니다!