핸드크림 리뷰 분석을 통한 소비 트렌드 파악

목차

1.주제 선정 배경

2.분석 과정 설명

- 크롤링
- EDA
- 리뷰 분석
- 원료 스코어링

3. 결과 해석

4. 한계점 및 의의





주제 선정 배경

- 다양한 신생브랜드, 직영점 및 가맹점의 빠른 증가로 인한 가격 경쟁 심화
 - → 화장품 제조사의 수익성 악화

(산업동향 보고서, 화장품 산업 동향 및 K-뷰티 글로벌 경쟁력 분석)

• SNS 데이터 등 빅데이터를 활용한 최신 트렌드 분석 기법 적용 시도

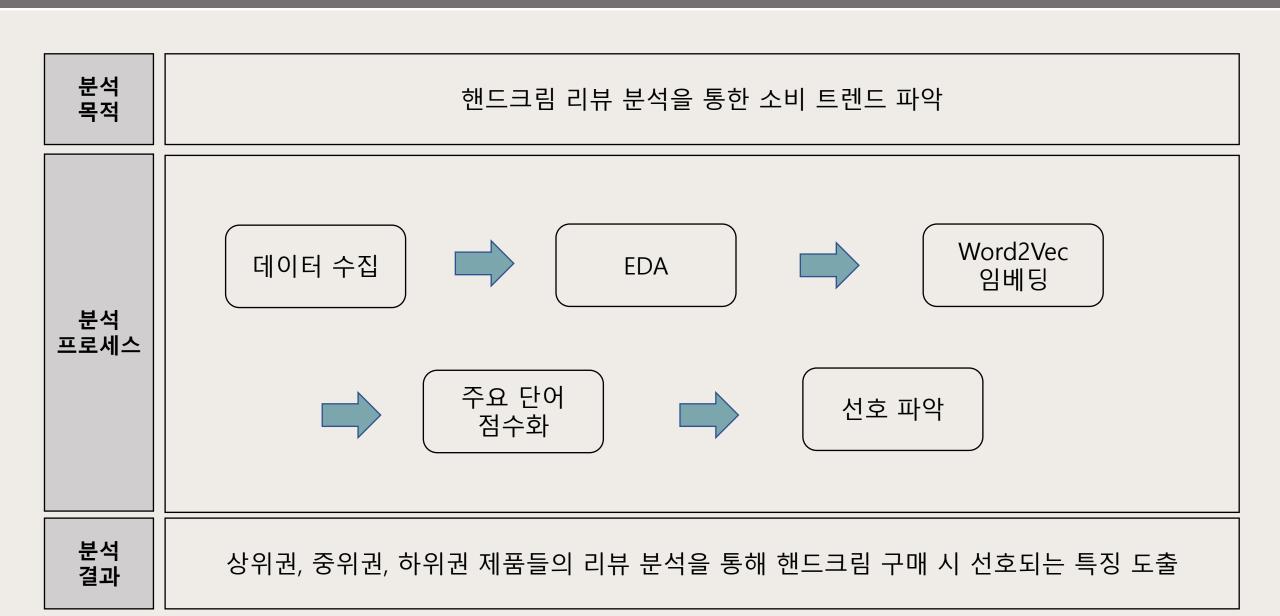
(2018 글로벌 화장품 시장 동향분석 세미나)

• 스킨케어 제품 중 아시아의 점유율은 약 41%로 압도적인 경쟁력

(화장품 회사의 빅데이터 분석을 통한 브랜드 컨셉 개발 사례 분석)

스킨케어의 대표 제품 "핸드크림"에 대한 리뷰 분석을 통한 소비 트렌드 파악

분석 개요



크롤링

• 화장품 리뷰 사이트 내 핸드크림 제품(상위권, 중위권, 하위권) 리뷰 크롤링

■ 상위권: 1위 ~ 3위 제품

■ 중위권: 50위 ~ 52위 제품

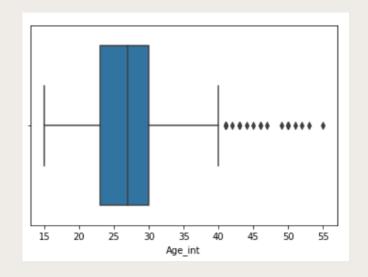
■ 하위권: 98위 ~ 100위 제품

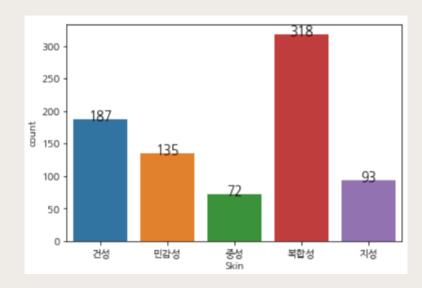
• 제품에 대한 평점과 함께 리뷰어의 나이, 성별, 피부타입, 리뷰 작성 날짜 크롤링

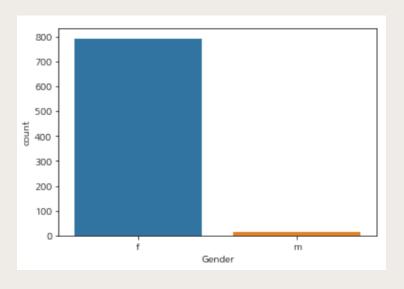
전체 EDA

상위권, 중위권, 하위권 모든 제품에 대한 데이터 탐색

- 1. 리뷰어 특성 파악
 - 나이대: 15세 ~ 55세 까지 연령대 다양
 - 피부 타입: 복합성이 가장 많았으며, 건성, 민감성, 지성, 중성 순으로 많음
 - 성별: 여성 리뷰어가 압도적으로 많음.



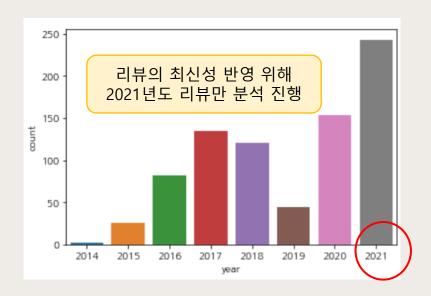


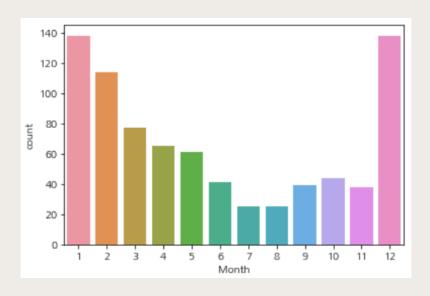


전체 EDA

상위권, 중위권, 하위권 모든 제품에 대한 데이터 탐색

- 2. 리뷰 작성 날짜 파악
 - 년도 비교: 가장 오래된 년도는 2014년이었으며, 가장 최근의 2021년도 리뷰가 가장 많음
 - 작성 월 비교: 동절기인 12월 ~ 2월 작성된 리뷰의 개수가 가장 많음
 - → 핸드크림 소비 패턴 시기와 일치





전체 EDA

상위권, 중위권, 하위권 모든 제품에 대한 데이터 탐색

- 3. 리뷰어 특성에 따른 리뷰 워드 클라우드 생성
 - 리뷰어 특성 (나이, 성별, 피부타입)에 따른 리뷰에 등장하는 단어의 빈도가 매우 유사함



20대 여성 중 건성 피부를 가진 리뷰어의 워드 클라우드



30대 남성 중 복합성 피부를 가진 리뷰어의 워드 클라우드

나이대와 성별, 피부 타입에 관계없이 사용자들은 향, 보습, 흡수, 가격 등을 고려해 핸드크림 구매함

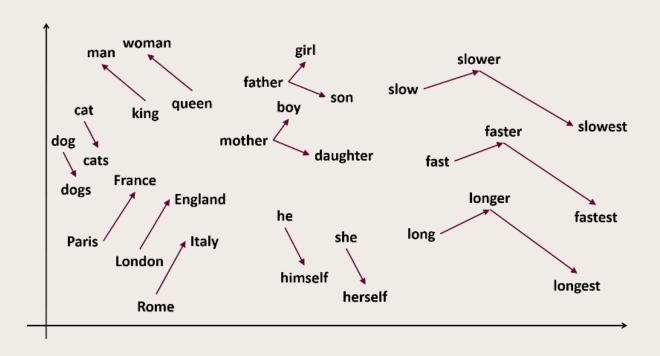
리뷰 분석 방향

- EDA를 통해 얻은 중심단어 (향, 보습, 가격, 흡수) 기준으로 함께 쓰인 형용사 분석 (ex. 좋다, 미끌거린다)
- 자연어 처리 분석 기법의 임베딩 방법론 중, Neural network 기반한 Word2Vec 방법 사용
- 중심단어가 주어졌을 때, 주변단어가 나타날 확률 계산하면서 Word2Vec 학습해 단어 임베딩

리뷰 분석 방향

- 임베딩: 사람이 쓰는 자연어를 기계가 이해할 수 있는 벡터로 바꾼 결과
- 중심단어(향, 보습, 흡수, 가격)를 기준으로 코사인 유사도가 높은 형용사 추출

•
$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i * B_i}{\sqrt{\sum_{i=1}^{n} (A_i^2)} * \sqrt{\sum_{i=1}^{n} (B_i^2)}}$$



상위권 제품 리뷰 분석

1위 세라마이딘 핸드크림

2위 세라마이드 무향 핸드크림

3위 카밀 핸드크림 튜브







	1위	2위	3위
향	(좋다) 0.69	(좋다) 0.18	(좋다) 0.83
보습	(좋다) 0.70	(건조하다) 0.18	(좋다) 0.80
	(미끌거리다) 0.68	(미끌거리다) 0.20	-
흡수	(좋다) 0.70	(없다) 0.19	(좋다) 0.79
가격	(좋다) 0.56	(않다) 0.11	(좋다) 0.71

상위권 제품 원료 분석 -향료 중심으로

1위 세라마이딘 핸드크림



베르가모트 열매오일 센티드 제라늄 꽃 오일 세이지 오일 광곽향 오일 울금뿌리 추출물

2위 세라마이드 무향 핸드크림



부틸렌글라이콜 글리세린 토코페롤

3위 카밀 핸드크림 튜브



캐모마일 꽃 추출물 비사볼을 벤질살리 실레이트 쿠마린 <mark>헥실신남알</mark> 리모넨 유제놀

상위권 분석 결과, 구매자들이 선호하는 향은 **달달**하면서 **플로럴 향** 선호, 무향은 선호하지 않음.

상위권 제품 리뷰 분석 결론

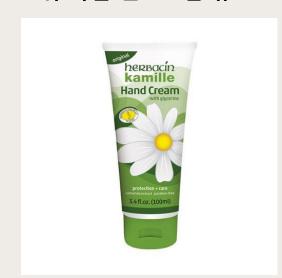
1위 세라마이딘 핸드크림



2위 세라마이드 무향 핸드크림



3위 카밀 핸드크림 튜브



- 향: 달달하면서 플로럴 향 선호
- 보습: 상위권에 위치한 핸드크림들의 보습력에 만족
- 흡수: 상위권에 위치한 핸드크림들의 흡수력에 만족
- 가격: 10,000원 ~ 12,000 사이 가격대에 만족

중위권 제품 리뷰 분석

50위 핑크 그레이프후룻 핸드크림



*리뷰의 수가 5개로 매우 적어 다른 키워드 분석 진행 불가능

51위 페퓸드 핸드크림- 피치앤 아프리콧



*리뷰의 수가 1개 모델링 불가능

52위 라벤더 핸드크림



*리뷰의 수가 10개로 적어 다른 키워드 분석 진행 불가능

	50위	51위	52위
향	(좋다) 0.01	-	(좋다) 0.04
보습	-	-	(좋다) 0.02

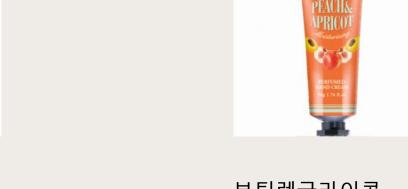
중위권 제품 원료 분석 -향료 중심으로

50위 핑크 그레이프후룻 핸드크림

51위 퍼퓸드 핸드크림- 피치앤 아프리콧

52위 라벤더 핸드크림







카프릴릭 스위트 아몬드 오일 오렌지 씨 오일 콩오일 리모넨 리날룰 헥실신남알 시트랄 부틸렌글라이콜 자몽 추출물 페녹시에탄올

중위권 제품은 <mark>플로럴 향</mark>료 많이 사용됨 시트러스 향료 많음 잇꽃씨 오일 캐모마일 꽃 추출물 카렌둘라꽃 추출물 알로에베라 잎 추출물 리날룰 리모넨 쿠마린 제라니올

중위권 제품 리뷰 분석 결론

50위 핑크 그레이프후룻 핸드크림

51위 퍼퓸드 핸드크림- 피치앤 아프리콧

52위 라벤더 핸드크림







- 향: 중위권 향 선호하지 않음 (복합적인 플로럴향, 시트러스 향)
- 보습: 중위권 핸드 크림의 보습력에 만족하지 않음

하위권 제품 리뷰 분석

98위 마이-라임오렌지 핸드크림



*리뷰의 수가 1개 모델링 불가능

99위 토탈 모이스처 헬씨 핸드 앤 스트롱거 네일 로션



*리뷰의 수가 9개로 매우 적어 다른 키워드 분석 진행 불가능

100위 슈가딜라이트 너리싱 핸드크림



*리뷰의 수가 2개로 매우 적어 다른 키워드 분석 진행 불가능

	98위	99위	100위
향	(좋다) 0.13	-	(좋다) 0.009
흡수	(좋다) 0.02	-	-

하위권 제품 원료 분석 - 향료 중심으로

98위 마이-라임오렌지 핸드크림



올리브 오일 글리세릴스테아레이트 카렌둘라 꽃 추출물 개박하추출물 붉은토끼풀꽃 추출물 폴리소르베이트60 베르가모트 열매오일 라임오일 시트로넬올 리모넨 제라니올 리날룰

99위 토탈 모이스처 헬씨 핸드 앤 스트롱거 네일 로션



올리브오일 벤질살리실레이트 제라니올 부틸페닐메칠프로피오날 리날룰 시트로넬올 헥실신남알 리모넨 알파-이소메칠이오논

100위 슈가딜라이트 너리싱 핸드크림



다시마 추출물 시트랄 <mark>제라니올</mark> 부틸페닐메칠프로피오날 리날룰 시트로넬올

하위권 제품은 여러가지 플로럴 향료 사용 인공 합성향을 내는 향료 사용됨

하위권 제품 리뷰 분석 결론

98위 마이-라임오렌지 핸드크림



99위 토탈 모이스처 헬씨 핸드 앤 스트롱 거 네일 로션



100위 슈가딜라이트 너리싱 핸드크림



- 향: 하위권 향 선호하지 않음 (복합적인 플로럴향, 인공 합성향)
- 흡수: 하위권 핸드 크림의 흡수력에 불만족

결론

- 리뷰 중, 많이 등장하는 단어는 **향, 보습, 흡수, 가격** 순으로 등장하였고 이는 소비자들이 핸드크림을 구매할 때 중요하게 생각하는 순서로 파악할 수 있다.
- 소비자들이 선호하는 **향**은 플로럴 향과 달달한 향이 섞여있는 향이지만, 너무 많은 플로럴 향이 느껴지는 것은 선호하지 않는다. 또한 무향과 인공 합성향도 선호하지 않는 것으로 나타났다.
- 소비자들이 원하는 보습 정도는 3위, 1위, 52위였고, 2위의 경우 보습에 대해 건조하다는 리뷰 존재
- 핸드크림 흡수의 측면에서 가장 선호하는 순위는 3위, 1위, 99위 였고, 2위는 "없다"는 리뷰 존재
- 소비자들이 원하는 가격대는 3위, 1위 순으로 가격에 만족함을 보임.

분석 의의

- 기존 분석과의 **차별점** 존재
 - → 리뷰에 대한 궁/부정 분석에서 더 나아가 궁/부정적으로 생각하는 원인을 향, 가격,보습, 흡수 측면 에서 다각도로 분석
- 구축된 분석 파이프 라인을 통해 시간에 따른 소비자들의 니즈를 빠르게 따라잡을 수 있음.
- 신제품을 위한 시장조사에 드는 제조비용을 절약할 수 있음

한계점

- 리뷰 개수 한계
 - → 중, 하위권 제품에 대해서는 리뷰의 개수가 적어 분석 결과 **일반화 불가능**
- 2위 제품 리뷰 신뢰도 (데이터 오염)
 - → 2위 제품은 일반적인 소비자들의 리뷰와 "평가단"으로 분류되는 소비자들의 리뷰가 섞여 있음
 - → 평가단의 리뷰는 일반적인 소비자들의 평점보다 더 낮게 주는 경향이 있으므로 리뷰의 신뢰도 하락
- 글로우픽 사이트 내부 순위에 대한 타당성
 - → 글로우픽에서 **자체 알고리즘**을 통해 화장품 순위를 매기는데 그 기준을 알 수 없음

참고 문헌

- 도서
 - 한국어 임베딩 (이기창 저)
- 논문 및 보고서
 - 화장품 산업 동향 및 K-뷰티 글로벌 경쟁력 분석(산업 동향보고서)
 - 음식의 재료들을 고밀도 벡터 공간상 화학적 조합으로 임베딩하기 위한 방법론 (박동현 외 4인)
 - 화장품 회사의 빅데이터 분석을 통한 브랜드 컨셉 개발 사례 분석 (이주민, 방정혜)
 - Efficient Estimation of Word Representations in Vector Space (Tomas Mikolov 외 2인)

감사합니다.