

Shops

Shops 는 , 스크래핑 기술을 통해 온라인 쇼핑몰들에 판매되는 브랜드의 모든 상품들에 대한 판매정보와 리뷰를 분석해 주는 서비스입니다.

- 아모레 퍼시픽 자사몰, 지에스샵, 글로우픽에서 판매되는 [이니스프리] 상품에 대한 정보를 한 곳에서 확인이 가능합니다.
- 각 쇼핑몰 별 전체 리뷰 수 확인이 가능합니다.
- 상품별 리뷰수와 판매량 등의 결과를 한 눈에 확인 할 수 있습니다.
- KoNLPy를 통한 리뷰 분석이 가능합니다.

문희수 : heesoo3458@gmail.com

박대현 : lieunmere@gmail.com

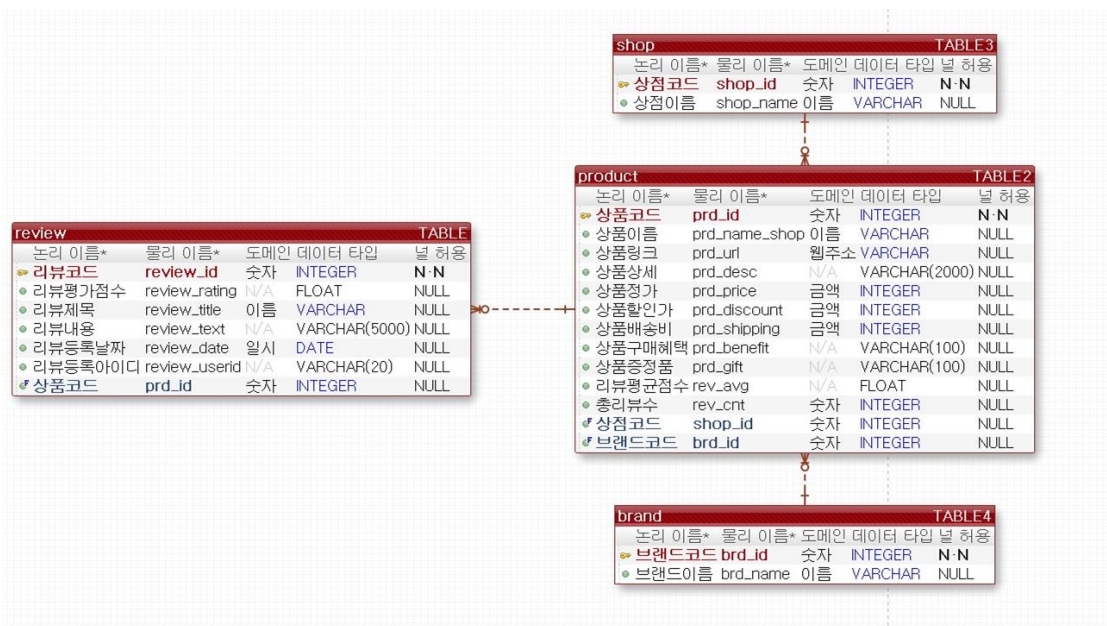
최형인 : guddls0101@gmail.com

조미래 : zzomill0427@gmail.com

신주현 : jshin@gmail.com

사용 기술





테이블

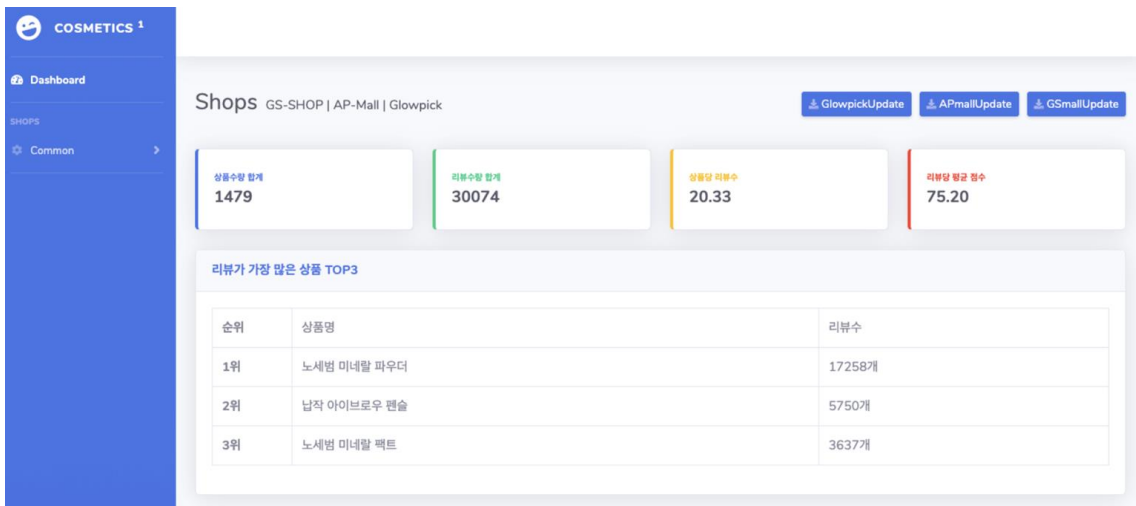
1. BRAND(브랜드)
2. SHOP(상점)
3. PRODUCT(상점상품)
4. REVIEW(상품리뷰)

기본키

1. REVIEW_id
2. SHOP_id
3. PRD_id
4. BRD_id

외래키

1. REVIEW.prd_id
2. PRODUCT.shop_id
3. PRODUCT.brd_id



1. Glowpick, APmall, GSshop Update

각각의 사이트에서 새로 등록된 상품 및 기존 상품에서 추가된 리뷰만 크롤링 후 DB저장

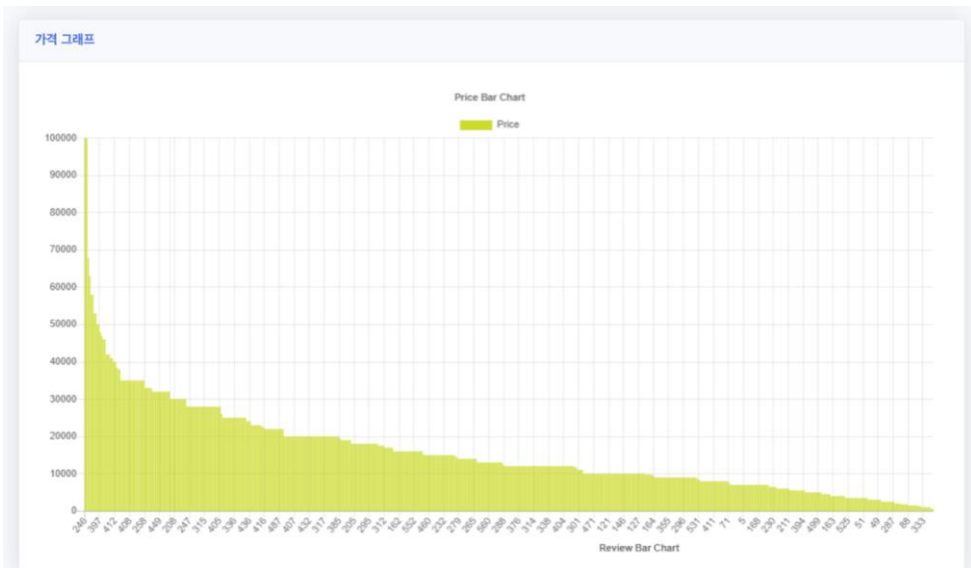
1. 상품수량, 리뷰수량, 리뷰 수, 평균 점수

DB에 저장된 데이터를 기반으로 추출

1. 리뷰가 가장 많은 상품 TOP3

DB에 저장된 데이터를 기반으로 리뷰가 많은 순위 추출

가격 그래프



Whole Price bar Chart (per product ID)

- x : 상품 아이디(prd_id)
- y : 상품가격(prd_price)

DB에 저장된 데이터기반으로 브랜드의
각 제품별 가격추이 확인 가능 : 브랜드 정보

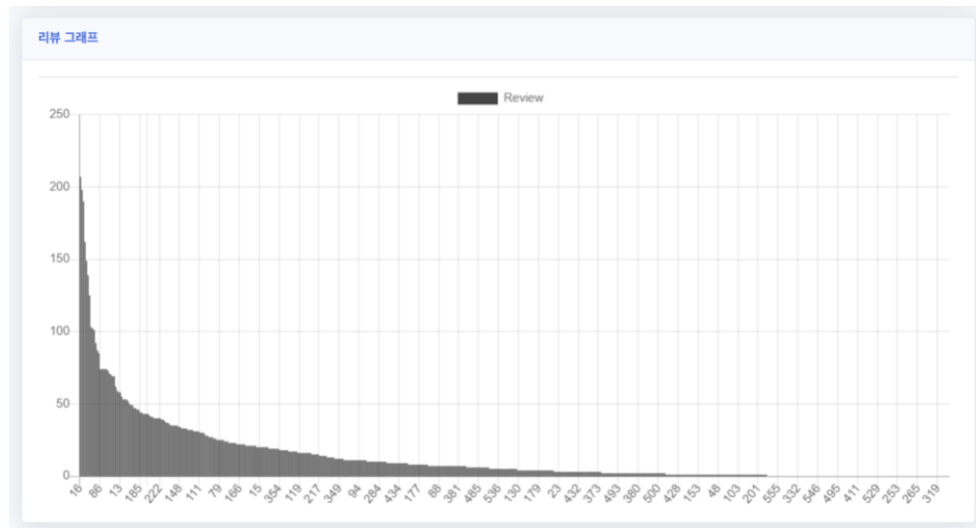
- chart.js에서 제공하는 bar graph 사용
- view : index 페이지 안에서 ajax로 노출

전 상품군대상, 가격이 높은 순서대로 orderby 적용

innisfree 브랜드의 [가격분포, 최저/최대상품가격, 평균 가격대]를 확인할 수 있다.

>>추후 보완 사항 : 각 브랜드 별 제품가격분포도
한눈에 비교할 수 있도록 그래프 추가

리뷰 그래프



Whole Reviews bar Chart (per reviews ID)

- x : 제품 아이디(prd_id)
- y : 리뷰 갯수(rev_cnt)

DB에 저장된 데이터기반으로 전체 제품 별 리뷰 갯수 확인 : 제품 인기도 측정

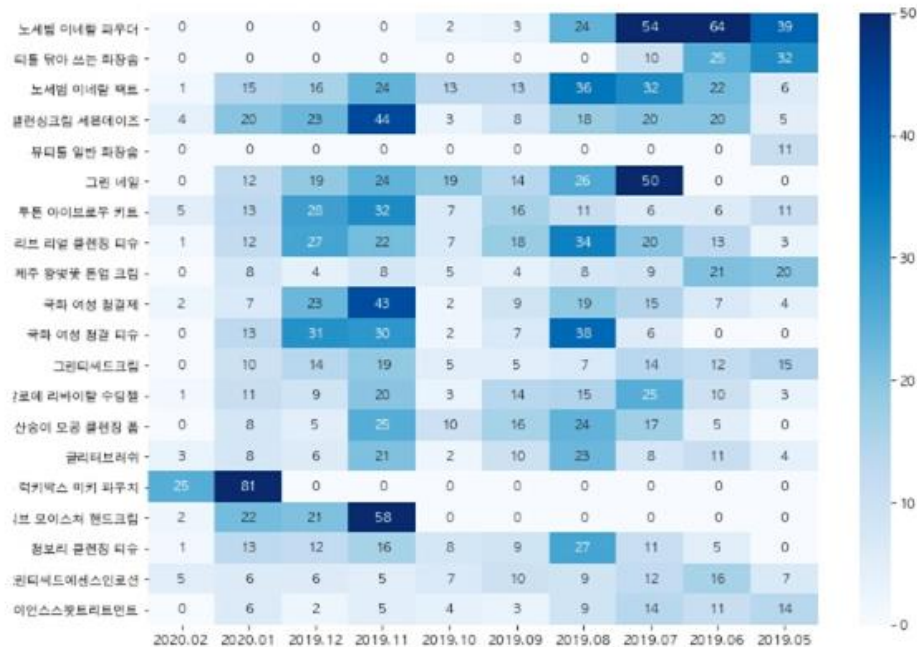
- chart.js에서 제공하는 bar graph 사용
- view : index 페이지 안에서 ajax로 호출

전 상품군대상, 리뷰 수가 많은 순서대로
orderby 적용

innisfree 상품 내 [인기제품, 리뷰 활성화도, 최저/
최대리뷰 갯수]를 알아낼 수 있다.

>>추후 보완 사항 : Review테이블의 rev_date를
사용하여 원하는 구간 일정에서의 리뷰갯수
count확인

TOP20 제품의 월별 리뷰수 표현



1. 전체 리뷰 수가 높은 제품 20개를 추출, 정규표현식을 활용하여 불필요한 단어 제거 (용량, 이벤트 등)

1. 제품별 전체 리뷰를 월별로 분리

1. x축을 시간, y축을 top20 제품으로 설정하여 리뷰 수 표현. 리뷰 수가 많을수록 색이 짙음

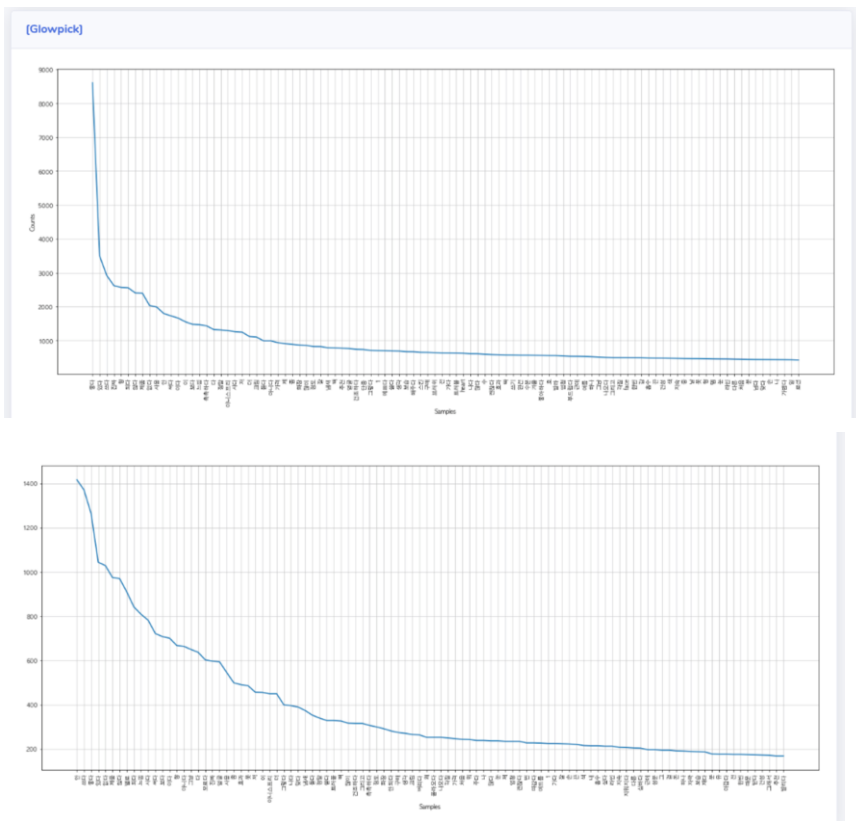
(노세범 미네랄 파우더 기준 2019.06월 리뷰수 64개, 2019.10월 리뷰수 2개)

리뷰 빈출 단어로 워드클라우드



1. KoNLPy를 활용하여 품사를 태깅하고, 필요 없는 단어 제거 (형태소가 Punctuation, Josa, Modifier, Koreanparticle 등)
1. `most_common()`을 활용하여 최빈 단어 순으로 정렬, 이후 하나의 string으로 연결
1. `wordcloud` 모듈로 워드클라우드를 만들고, `matplotlib`를 통해 원하는 형태로 출력

긍 / 부정 리뷰 빈출 단어 출력



1. 리뷰 전체를 긍정과 부정으로 분류 (별점 기준)
2. KoNLPy를 활용하여 품사를 태깅하고, 그 중 명사 동사 형용사만 새로운 리스트에 추가
3. '하다', '있다' 등 긍정 부정에 영향을 미치지 않는 단어들 제거
4. collections 모듈을 활용하여 리스트 안 단어 개수를 count하고, count가 높은 순으로 단어 정렬
5. matplotlib로 최빈 단어 100개를 그 개수와 함께 그래프로 그려줌
6. 최빈 단어 (긍정: '좋다' 부정: '안')