

# DS EatSight 6주차 발표

## 데이터 수집

김예준, 오유성, 오현진, 조영서

- 해당 주차는 Selenium으로 **립틴트** 올리브영 3제품의 리뷰 데이터를 수집.

**립틴트**

|    |            |     |      |    |      |
|----|------------|-----|------|----|------|
| 전체 | <b>립틴트</b> | 립스틱 | 립라이너 | 립밤 | 립글로스 |
|----|------------|-----|------|----|------|

- 대상 제품은 리뷰데이터가 많은 것으로 선정. (총 리뷰 최소 2만건 이상)



롬앤  
[NEW/단독기획] 롬앤 더 주시 래스팅  
틴트 (기획/단품)  
13,000원- **11,050원~**  
쿠폰 **오늘드림**



페리페라  
[4월올영픽][+오히키링] 페리페라 잉크  
무드 글로이 틴트  
11,000원- **8,800원~**  
세일 증정 **오늘드림**



클리오  
[해리포터 콜라보] 클리오 크리스탈  
글램 틴트 단품/기획  
18,000원- **12,600원~**  
세일 **오늘드림**

롬앤 169,225 / 페리페라 87,999 / 클리오 26,397

- 웹에선 **필터 당 최대 100pg** 까지 리뷰를 제공. (1000건)  
날짜별 필터링 항목이 존재하지 않아 **평점 별로 수집**.

<<
 91
92
93
94
95
96
97
98
99
100

- 이로 제품 당 약 5000개 이하의 데이터를 수집 (1~5점 각 약 1000건)

**평점**  
전체
5점
4점
3점
2점
1점

☐ 나의 피부 컨디션 정보 불러오기

**피부타입**

지성
건성
복합성
민감성
약건성
트러블성

중성

**피부톤**

쿨톤
웜톤
봄웜톤
여름쿨톤
가을웜톤

겨울쿨톤

**피부 고민**

잡티
미백
주름
각질
트러블
블랙헤드

피지과다
민감성
모공
탄력
홍조
아토피

다크서클

**연령대**

10대
20대
30대
40대 이상

**성별**

여성
남성



나익히

TOP 887

건성 · 가을웜톤 · 각질 · 다크서클

한달이상사용

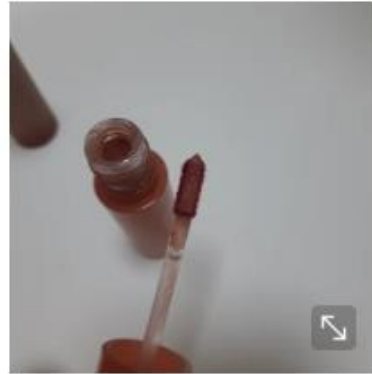


2023.01.12 · 매장

[옵션] [단품] 11 극락좌표

**발색력** 다소아쉬워요 | **지속력** 지속이오래돼요 | **발림성** 아주만족해요 | **수분감** 아주촉촉해요

손등발색했을 때 색상이랑 입술에 올렸을때랑 너무 달라요. 입술에서 너무 어둡게 착색돼서 하... 아쉽다. 유리알광택 마음에 드는데..



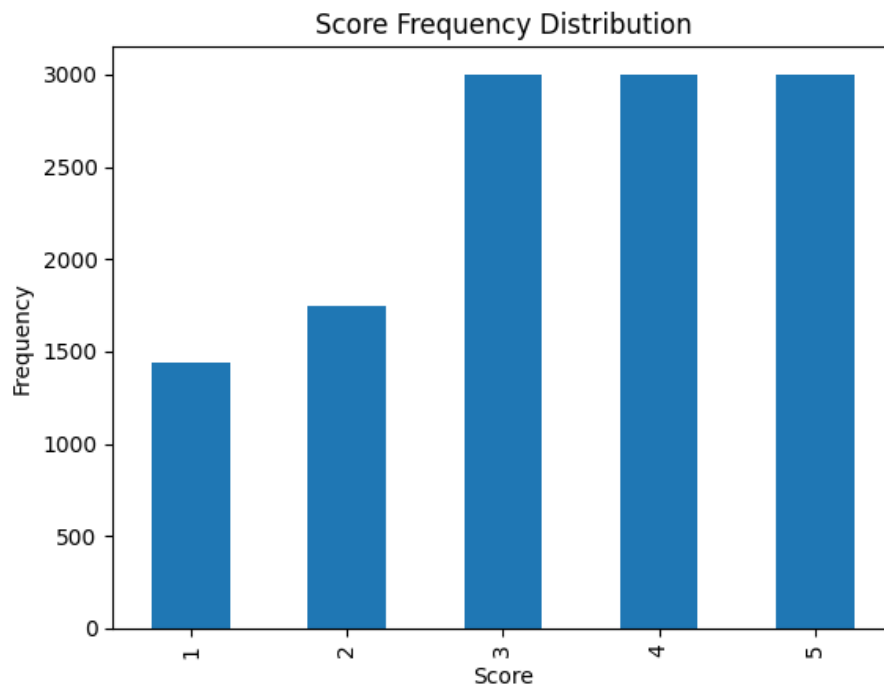
이 리뷰가 도움이 돼요!

4

신고하기

| date       | score | user_tags          | user_rank | usage_duration | item_option       | text_review                                       | recommendation_score | 발색력     | 지속력      | 발림성     | 수분감     | skin_tone | product_name      |
|------------|-------|--------------------|-----------|----------------|-------------------|---|----------------------|---------|----------|---------|---------|-----------|-------------------|
| 2023.01.12 | 1     | 건성, 가을웜톤, 각질, 다크서클 | 887.0     | 한달이상사용         | [옵션] [단품] 11 극락좌표 | 손등발색했을 때 색상이랑 입술에 올렸을때랑 너무 달라요. 입술에서 너무 어둡게 착색... | 4                    | 다소 아쉬워요 | 지속이 오래돼요 | 아주 만족해요 | 아주 촉촉해요 | 가을웜톤      | 페리페라 잉크 무드 글로이 틴트 |

- 총 데이터는 12,182개 수집. **1~2점 데이터가 부족.**
- 피부톤 데이터가 있는 경우는 **반정도.** (6742개는 데이터가 없음)



## 향후 계획

1. 데이터 추가 수집(1, 2점 위주 & 톤) 및 전처리
2. 분석 시작 (톤별 토픽 모델링 & 점수 예측 머신러닝)



감사합니다.