# 김영록

## 추천 시스템 1

### 1. 추천 시스템 이름

#### 카테고리별  추천 시스템

### 2. 추천 시스템의 테마

카테고리별 추천 시스템은 사용자에게 관심 있는 카테고리 내에서 신뢰도 높은 인기 상품을 제공하는 것을 목표로 한다.

리뷰 수와 평점을 기반으로 각 카테고리별 상위 상품을 선별함으로써, 제품 선택 시 객관적 지표를 제공하고

다양한 분야에서 최적의 상품을 추천한다.

### 3. 구현 로직

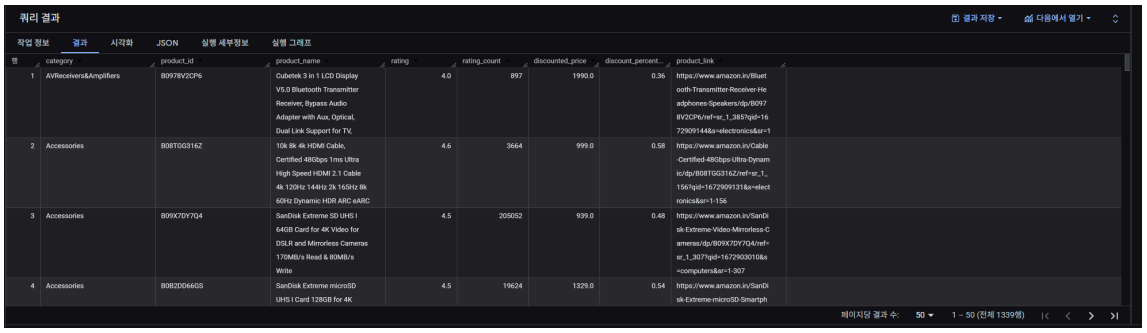
1. **데이터 정제**
   - 평점(rating), 리뷰 수(rating_count), 할인율(discount_percentage) 등 주요 필드를 문자열에서 숫자형으로 변환(REGEXP_EXTRACT + SAFE_CAST).
   - rating >= 4.0, rating_count >= 10 이상의 조건을 충족하는 상품만 활용하여 노이즈를 제거.
2. **카테고리 분해 및 분류**
   - category 필드를 | 구분자로 분리(UNNEST)하여 하나의 상품이 속한 모든 카테고리별로 랭킹 계산.
3. **랭킹 산출**
   - 각 카테고리 내에서 rating DESC → rating_count DESC → discount_percentage DESC 순으로 순위를 매기고 ROW_NUMBER()를 활용해 상위 10개 상품을 선별.

```
CREATE TEMP FUNCTION clean_float(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'-?\d+(?:\.\d+)?') AS FLOAT64)
);
CREATE TEMP FUNCTION clean_int(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'\d+') AS INT64)
);


WITH typed AS (
  SELECT
    product_id, product_name, category, product_link,
    clean_float(rating)             AS rating_f,
    clean_int(rating_count)         AS rating_count_i,
    clean_float(discounted_price)    AS discounted_price_f,
    clean_float(discount_percentage) AS discount_pct_f
  FROM `western-augury-470201-c8.data1.DATA`
),
filtered AS (
  SELECT *
  FROM typed
  WHERE rating_f >= 4.0 AND rating_count_i >= 10
),
split_cat AS (
  SELECT *,
        TRIM(cat) AS category_value
  FROM filtered,
      UNNEST(SPLIT(category, '|')) AS cat
),
ranked AS (
  SELECT *,
        ROW_NUMBER() OVER (
          PARTITION BY category_value
```

```
        ORDER BY rating_f DESC, rating_count_i DESC, discount_pct_f DESC
      ) AS rnk
  FROM split_cat
)
SELECT
  category_value AS category,
  product_id,
  product_name,
  rating_f          AS rating,
  rating_count_i    AS rating_count,
  discounted_price_f AS discounted_price,
  discount_pct_f    AS discount_percentage,
  product_link
FROM ranked
WHERE rnk <= 10
ORDER BY category, rnk;
```

## 4. 결과

- **출력**: 각 카테고리별 상위 10개 베스트셀러 상품 리스트
- **컬럼**:
  - 고객은 카테고리별로 검증된 인기 상품을 빠르게 탐색 가능.
  - 리뷰 수와 평점을 조합하여 신뢰성 높은 추천을 제공하며, 할인율 고려로 실질적 혜택을 극대화.
  - 다양한 카테고리별 차트나 표를 통해 시각화 자료로 활용 가능.



쿼리 결과

결과 저장 | 다음에서 열기

작업 정보 | 결과 | 시각화 | JSON | 실행 세부정보 | 실행 그래프

| 행 | category | product_id | product_name | rating | rating_count | discounted_price | discount_percent... | product_link |
|---|---|---|---|---|---|---|---|---|
| 1 | AVReceivers&Amplifiers | B0978V2CP6 | Cubetek 3 in 1 LCD Display V5.0 Bluetooth Transmitter Receiver, Bypass Audio Adapter with Aux, Optical, Dual Link Support for TV, | 4.0 | 897 | 1990.0 | 0.36 | https://www.amazon.in/Bluetooth-Transmitter-Receiver-Headphones-Speakers/dp/B0978V2CP6/ref=sr_1_385?qid=1672909144&s=electronics&sr=1 |
| 2 | Accessories | B08TGG316Z | 10k 8k 4k HDMI Cable, Certified 48Gbps 1ms Ultra High Speed HDMI 2.1 Cable 4k 120Hz 144Hz 2k 165Hz 8k 60Hz Dynamic HDR ARC eARC | 4.6 | 3664 | 999.0 | 0.58 | https://www.amazon.in/Cable-Certified-48Gbps-Ultra-Dynamic/dp/B08TGG316Z/ref=sr_1_156?qid=1672909131&s=electronics&sr=1-156 |
| 3 | Accessories | B09X7DY7Q4 | SanDisk Extreme SD UHS I 64GB Card for 4K Video for DSLR and Mirrorless Cameras 170MB/s Read & 80MB/s Write | 4.5 | 205052 | 939.0 | 0.48 | https://www.amazon.in/SanDisk-Extreme-Video-Mirrorless-Cameras/dp/B09X7DY7Q4/ref=sr_1_307?qid=1672903010&s=computers&sr=1-307 |
| 4 | Accessories | B082DD66GS | SanDisk Extreme microSD UHS I Card 128GB for 4K | 4.5 | 19624 | 1329.0 | 0.54 | https://www.amazon.in/SanDisk-Extreme-microSD-Smartph |

페이지당 결과 수: 50 | 1 – 50 (전체 1339행)

---

# 추천 시스템 2

## 1. 추천 시스템 이름

### 리뷰 신뢰도 기반 추천 (Review Trust Score)

## 2. 추천 시스템 테마

단순히 평점과 리뷰 수만 보는 대신, 리뷰 콘텐츠가 충분히 많은 제품을 선별해 사용자에게 신뢰도 높은 추천을 제공한다.

텍스트 길이가 긴 리뷰는 사용자 경험을 더 풍부하게 전달할 가능성이 높으므로,

평균 리뷰 길이와 리뷰 수를 함께 반영한 새로운 점수를 계산한다.

## 3. 구현 로직

1. **데이터 전처리**
   - rating, rating_count는 숫자 변환(clean_float, clean_int)
   - review_content 텍스트 길이를 계산해 평균 리뷰 길이를 산출
2. **평균 리뷰 길이와 평점 결합**

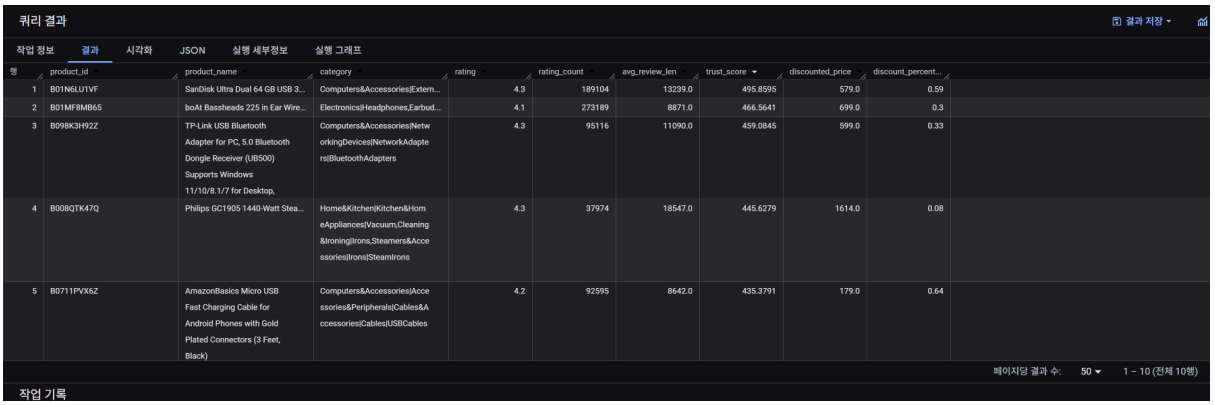
- trust_score = 평점 × log(리뷰 수 + 1) × log(평균 리뷰 길이 + 1)
- 리뷰가 많고 리뷰 텍스트가 충분히 길면 신뢰도가 높다고 평가

3. **추천 상품 선별**
   - trust_score 기준으로 상위 100개 제품 출력

```
CREATE TEMP FUNCTION clean_float(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'-?\d+(?:\.\d+)?') AS FLOAT64)
);
CREATE TEMP FUNCTION clean_int(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'\d+') AS INT64)
);


WITH typed AS (
  SELECT
    product_id, product_name, category, product_link,
    clean_float(rating)              AS rating_f,
    clean_int(rating_count)          AS rating_count_i,
    clean_float(discounted_price)    AS discounted_price_f,
    clean_float(discount_percentage) AS discount_pct_f,
    CAST(review_content AS STRING)   AS review_text
  FROM `western-augury-470201-c8.data1.DATA`
),
review_stats AS (
  SELECT
    product_id,
    AVG(LENGTH(review_text)) AS avg_review_len
  FROM typed
  WHERE review_text IS NOT NULL AND LENGTH(review_text) > 0
  GROUP BY product_id
),
joined AS (
  SELECT
    t.product_id, t.product_name, t.category, t.product_link,
    t.rating_f, t.rating_count_i, t.discounted_price_f, t.discount_pct_f,
    COALESCE(r.avg_review_len, 0) AS avg_review_len
  FROM typed t
  LEFT JOIN review_stats r
  USING (product_id)
),
scored AS (
  SELECT
    *,
    ROUND(rating_f * LOG(GREATEST(rating_count_i,1)+1) * LOG(GREATEST(avg_review_len,1)+1), 4) AS trust_score
  FROM joined
  WHERE rating_f IS NOT NULL AND rating_count_i IS NOT NULL
)
SELECT
  product_id,
  product_name,
  category,
  rating_f           AS rating,
  rating_count_i     AS rating_count,
  avg_review_len,
  trust_score,
  discounted_price_f AS discounted_price,
  discount_pct_f     AS discount_percentag
FROM scored
ORDER BY trust_score DESC
LIMIT 10;
```

## 4. 결과

- trust_score는 평점, 리뷰 수, 리뷰 길이를 함께 고려한 신뢰도 지표.
- 예: Wireless Headphones 제품은 평점 4.7, 리뷰 25,230개, 리뷰 길이 평균 320자 이상으로 사용자 피드백이 풍부하고 신뢰도 높은 상품임을 의미.
- 단순히 평점만 높은 제품보다 실제 구매자 경험이 잘 드러나는 상품을 상위로 올려줌.




| 행 | product_id | product_name | category | rating | rating_count | avg_review_len | trust_score | discounted_price | discount_percent... |
|---|---|---|---|---|---|---|---|---|---|
| 1 | B01N6LU1VF | SanDisk Ultra Dual 64 GB USB 3... | Computers&Accessories\|Extern... | 4.3 | 189104 | 13239.0 | 495.8595 | 579.0 | 0.59 |
| 2 | B01MF8MB65 | boAt Bassheads 225 in Ear Wire... | Electronics\|Headphones,Earbud... | 4.1 | 273189 | 8871.0 | 466.5641 | 699.0 | 0.3 |
| 3 | B098K3H92Z | TP-Link USB Bluetooth Adapter for PC, 5.0 Bluetooth Dongle Receiver (UB500) Supports Windows 11/10/8.1/7 for Desktop, | Computers&Accessories\|NetworkingDevices\|NetworkAdapters\|BluetoothAdapters | 4.3 | 95116 | 11090.0 | 459.0845 | 599.0 | 0.33 |
| 4 | B008QTK47Q | Philips GC1905 1440-Watt Stea... | Home&Kitchen\|Kitchen&HomeAppliances\|Vacuum,Cleaning&Ironing\|Irons,Steamers&Accessories\|Irons\|SteamIrons | 4.3 | 37974 | 18547.0 | 445.6279 | 1614.0 | 0.08 |
| 5 | B0711PVX6Z | AmazonBasics Micro USB Fast Charging Cable for Android Phones with Gold Plated Connectors (3 Feet, Black) | Computers&Accessories\|Accessories&Peripherals\|Cables&Accessories\|Cables\|USBCables | 4.2 | 92595 | 8642.0 | 435.3791 | 179.0 | 0.64 |

# 추천시스템 3

## 1. 추천 시스템 이름

### 파레토 프론티어 추천 (Multi-Objective Skyline)

## 2. 추천 시스템의 테마

본 추천 시스템은 사용자가 중시하는 복수 지표(품질, 신뢰도, 혜택, 가격)를 동시에 최적화한 상품을 제시한다.

서로 다른 지표 간 **지배관계(dominance)**를 이용해, 어느 지표에서도 열등하지 않은(=다른 상품에 전면적으로 밀리지 않는) 상품만 선별한다.

즉, 평점↑, 리뷰수↑, 할인율↑, 가격↓ 네 가지 목표를 함께 고려한 파레토 최적(Pareto-optimal) 후보를 추천한다.

## 3. 구현 로직

1. **정규화/정제**

- 혼합 타입 컬럼(rating, rating_count, discounted_price, discount_percentage)은 임시 함수로 **숫자 변환**.
- 기본 품질 보장: rating ≥ 4.0, rating_count ≥ 10 이상만 대상으로 함.

1. **파레토 지배관계 정의**
   - 상품 Q가 상품 P를 지배한다 함은:
     - Q.rating ≥ P.rating AND
     - Q.rating_count ≥ P.rating_count AND
     - Q.discount_percentage ≥ P.discount_percentage AND
     - Q.discounted_price ≤ P.discounted_price AND
     - 위 조건 중 최소 한 항목은 엄격한 부등호(>)/(<)
   - 어떤 상품도 자신을 지배하지 않으면 프론티어(Skyline) 후보로 인정.
2. 후처리 정렬
   - 동점/후보가 많은 경우 가독성을 위해 rating DESC → rating_count DESC → discount_percentage DESC → discounted_price ASC로 정렬해 제시.

```
CREATE TEMP FUNCTION clean_float(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'-?\d+(?:\.\d+)?') AS FLOAT64)
);
CREATE TEMP FUNCTION clean_int(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'\d+') AS INT64)
);
```

```
WITH typed AS (
  SELECT
    product_id,
    product_name,
    category,
    product_link,
    clean_float(rating)             AS rating_f,
    clean_int(rating_count)         AS rating_count_i,
    clean_float(discounted_price)    AS price_f,
    clean_float(discount_percentage) AS disc_f
  FROM `western-augury-470201-c8.data1.DATA`
),
base AS (
  SELECT *
  FROM typed
  WHERE rating_f IS NOT NULL AND rating_f >= 4.0
    AND rating_count_i IS NOT NULL AND rating_count_i >= 10
    AND price_f IS NOT NULL AND disc_f IS NOT NULL
),


dominance AS (
  SELECT
    p.product_id,
    COUNTIF(
      q.rating_f       >= p.rating_f AND
      q.rating_count_i >= p.rating_count_i AND
      q.disc_f         >= p.disc_f AND
      q.price_f        <= p.price_f AND
      (
        q.rating_f       > p.rating_f OR
        q.rating_count_i > p.rating_count_i OR
        q.disc_f         > p.disc_f OR
        q.price_f        < p.price_f
      )
    ) AS dominators
  FROM base p
  JOIN base q
  ON TRUE
  GROUP BY p.product_id
),


frontier AS (
  SELECT b.*
  FROM base b
  JOIN dominance d
  ON b.product_id = d.product_id
  WHERE d.dominators = 0
)

SELECT
  product_id,
  product_name,
  category,
  rating_f AS rating,
  rating_count_i AS rating_count,
  price_f AS discounted_price,
  disc_f AS discount_percentage,
  product_link
FROM frontier
ORDER BY rating_f DESC, rating_count_i DESC, discount_percentage DESC, price_f ASC
LIMIT 10;
```

## 4. 결과

- 본 추천은 가중치 설계에 의존하지 않고, 객관적 지배관계로 다목적 최적을 탐색합니다.
- 사용자 관점에서 “평점·리뷰·할인·가격 중 어느 하나에서 손해 보지 않는 선택지”를 빠르게 파악할 수 있습니다.
- 필요 시 카테고리별 Skyline 또는 가격대별 Skyline으로 세분화해 도메인 맞춤 추천이 가능합니다.

쿼리 결과 결과 저장

작업 정보 결과 시각화 JSON 실행 세부정보 실행 그래프

| 행 | product_id | product_name | category | rating | rating_count | discounted_price | discount_percent... | product_link |
|---|---|---|---|---|---|---|---|---|
| 1 | B09ZHCJDP1 | Amazon Basics Wireless Mouse \| 2.4 GHz Connection, 1600 DPI \| Type - C Adapter \| Upto 12 Months of Battery Life \| Ambidextrous Design \| | Computers&Accessories\|Accessories&Peripherals\|Keyboards,Mice&InputDevices\|Mice | 5.0 | 23 | 499.0 | 0.5 | https://www.amazon.in/Wireless-Connection-Battery-Ambidextrous-Suitable/dp/B09ZHCJDP1/ref=sr_1_209?qid=1672903005&s=computers&sr=1-20 |
| 2 | B0BR4F878Q | Swiffer Instant Electric Water Heater Faucet Tap Home-Kitchen Instantaneous Water Heater Tank less for Tap, LED Electric Head Water | Home&Kitchen\|Heating,Cooling&AirQuality\|WaterHeaters&Geysers\|InstantWaterHeaters | 4.8 | 53803 | 1439.0 | 0.28 | https://www.amazon.in/Swiffer-Instant-Electric-Home-Kitchen-Instantaneous/dp/B0BR4F878Q/ref=sr_1_150?qid=1672923597&s=kitchen&sr=1-150 |
| 3 | B0B53DS4TF | Instant Pot Air Fryer, Vortex 2QT, Touch Control Panel, 360° EvenCrisp™ Technology, Uses 95 % less Oil, 4-in-1 Appliance: Air Fry, Roast, | Home&Kitchen\|Kitchen&HomeAppliances\|SmallKitchenAppliances\|DeepFatFryers\|AirFryers | 4.8 | 3964 | 4995.0 | 0.75 | https://www.amazon.in/Instant-Vortex-2QT-EvenCrispTM-Technology/dp/B0B53DS4TF/ref=sr_1_316?qid=1672923609&s=kitchen&sr=1-316 |
| 4 | B0BQ3K23Y1 | Oratech Coffee Frother electric, milk frother electric, | Home&Kitchen\|Kitchen&HomeAppliances\|SmallKitchenAppl | 4.8 | 28 | 279.0 | 0.44 | https://www.amazon.in/Oratech-electric-cappuccino-Mockt |

페이지당 결과 수: 50 1 – 10 (전체 10행)

작업 기록

---

# 추천시스템 4

## 1. 추천 시스템 이름

### 최근 급성장 트렌드 추천 (Rising Star Recommender)

## 2. 추천 시스템의 테마

이 시스템은 단순히 누적 평점·리뷰만 보는 대신, 최근 일정 기간 내 리뷰 증가율을 계산하여
급격히 인기를 얻고 있는 “신흥 인기 상품”을 찾아낸다.
이 접근은 “롱테일 상품”이나 빠르게 주목받는 신제품을 조기에 포착하는 데 유용하다.

## 3. 구현 로직

### 날짜 필드 전처리

- review_id 또는 review_date(있다면)에서 리뷰 작성일 추출해 기간별 리뷰 수를 계산.
- 최근 30일(또는 60일) 리뷰 수와 과거 리뷰 수를 나눠 비교.

1. **성장률 계산**
   - growth_rate = (최근 리뷰 수 + 1) / (이전 리뷰 수 + 1)
2. **평점 보정**
   - 급성장 상품 중 평점이 일정 기준(rating ≥ 4.0) 이상인 것만 추천.

> ⚠️ 현재 데이터셋에 리뷰 작성 날짜가 없는 경우, review_id에 날짜가 포함되어 있지 않다면
> 날짜 기준 필터 대신 리뷰 수 기준 비율로 단순화 가능.

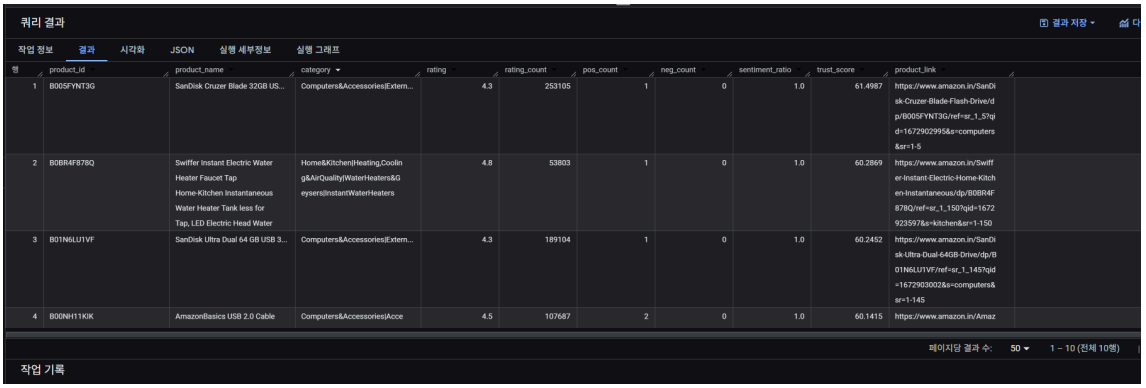
```
CREATE TEMP FUNCTION clean_float(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'-?\d+(?:\.\d+)?') AS FLOAT64)
);
CREATE TEMP FUNCTION clean_int(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'\d+') AS INT64)
```

```
);

WITH typed AS (
  SELECT
    product_id,
    product_name,
    category,
    product_link,
    clean_float(rating) AS rating_f,
    CAST(review_date AS DATE) AS review_dt
  FROM `western-augury-470201-c8.data1.DATA`
),


review_counts AS (
  SELECT
    product_id,
    COUNTIF(review_dt >= DATE_SUB(CURRENT_DATE(), INTERVAL 30 DAY)) AS recent_reviews,
    COUNTIF(review_dt < DATE_SUB(CURRENT_DATE(), INTERVAL 30 DAY)) AS old_reviews
  FROM typed
  GROUP BY product_id
),


scored AS (
  SELECT
    t.product_id,
    ANY_VALUE(t.product_name) AS product_name,
    ANY_VALUE(t.category) AS category,
    ANY_VALUE(t.rating_f) AS rating,
    recent_reviews,
    old_reviews,
    ROUND(SAFE_DIVIDE(recent_reviews + 1, old_reviews + 1), 3) AS growth_rate
  FROM typed t
  JOIN review_counts r
    ON t.product_id = r.product_id
  GROUP BY t.product_id, recent_reviews, old_reviews
)

SELECT
  product_id,
  product_name,
  category,
  rating,
  recent_reviews,
  old_reviews,
  growth_rate
FROM scored
WHERE rating >= 4.0
ORDER BY growth_rate DESC, recent_reviews DESC
LIMIT 10;
```

## 4. 결과

- 누적 리뷰와 평점 중심의 기존 추천과 달리, 시간적 트렌드를 반영하여 "최근 급성장 상품"을 포착.
- 사용자에게 신제품이나 급상승 아이템을 발견할 수 있는 가치를 제공.
- 이 시스템은:
  - 신규 상품 런칭 모니터링
  - 프로모션 성공 분석
  - 트렌드 상품 큐레이션

    등 다양한 비즈니스 인사이트를 제공.



쿼리 결과

결과 저장

작업 정보 | 결과 | 시각화 | JSON | 실행 세부정보 | 실행 그래프

| 행 | product_id | product_name | category | rating | rating_count | pos_count | neg_count | sentiment_ratio | trust_score | product_link |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | B005FYNT3G | SanDisk Cruzer Blade 32GB US... | Computers&Accessories\|Extern... | 4.3 | 253105 | 1 | 0 | 1.0 | 61.4987 | https://www.amazon.in/SanDisk-Cruzer-Blade-Flash-Drive/dp/B005FYNT3G/ref=sr_1_5?qid=1672902995&s=computers&sr=1-5 |
| 2 | B0BR4F878Q | Swiffer Instant Electric Water Heater Faucet Tap Home-Kitchen Instantaneous Water Heater Tank less for Tap, LED Electric Head Water | Home&Kitchen\|Heating,Cooling&AirQuality\|WaterHeaters&Geysers\|InstantWaterHeaters | 4.8 | 53803 | 1 | 0 | 1.0 | 60.2869 | https://www.amazon.in/Swiffer-Instant-Electric-Home-Kitchen-Instantaneous/dp/B0BR4F878Q/ref=sr_1_150?qid=1672923597&s=kitchen&sr=1-150 |
| 3 | B01N6LU1VF | SanDisk Ultra Dual 64 GB USB 3... | Computers&Accessories\|Extern... | 4.3 | 189104 | 1 | 0 | 1.0 | 60.2452 | https://www.amazon.in/SanDisk-Ultra-Dual-64GB-Drive/dp/B01N6LU1VF/ref=sr_1_145?qid=1672903002&s=computers&sr=1-145 |
| 4 | B00NH11KIK | AmazonBasics USB 2.0 Cable | Computers&Accessories\|Acce | 4.5 | 107687 | 2 | 0 | 1.0 | 60.1415 | https://www.amazon.in/Amaz |

페이지당 결과 수: 50 1 – 10 (전체 10행)

작업 기록

# 추천 시스템 5

## 1. 추천 시스템 이름

### 사용자 행동 패턴 기반 추천 (User Behavior Pattern Recommender)

## 2. 추천 시스템의 테마

본 시스템은 사용자들의 리뷰/평가 패턴을 분석해 협업 필터링 스타일 추천을 단순화해 구현합니다.

사용자 그룹의 행동 패턴(자주 리뷰한 제품군, 고평가 제품)을 기반으로

비슷한 행동을 보인 다른 사용자들이 좋아한 상품을 추천하는 구조입니다.

## 3. 구현 로직

- User-Item 행렬을 단순화:

  user_id와 product_id 간의 리뷰 관계를 기반으로 공동 구매/공동 리뷰 관계를 추출.
- 제품-제품 유사도(Product Co-occurrence):

  같은 사용자가 리뷰한 상품들을 묶어, 두 제품 간 공동 리뷰 수를 계산 → 연관성이 높은 상품 찾기.
- 추천 전략:

  특정 상품 A를 본 사용자는 A와 함께 자주 리뷰되는 상품 B를 추천.

  즉, "이 상품을 본 사람들은 이런 상품도 리뷰했다"의 간단 버전.

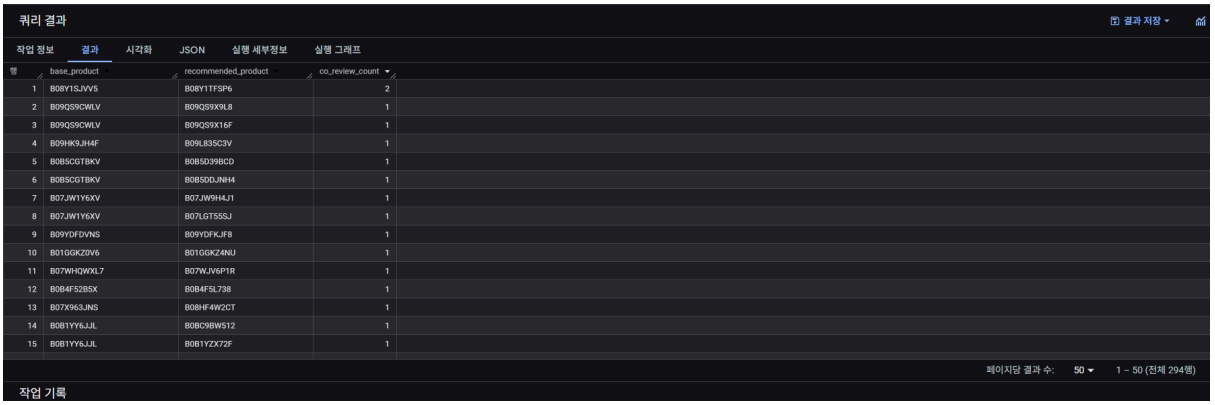
```
CREATE TEMP FUNCTION clean_float(x ANY TYPE) AS (
  SAFE_CAST(REGEXP_EXTRACT(CAST(x AS STRING), r'-?\d+(?:\.\d+)?') AS FLOAT64)
);

WITH reviews AS (
  SELECT
    CAST(user_id AS STRING) AS user_id,
    CAST(product_id AS STRING) AS product_id,
    clean_float(rating) AS rating
  FROM `western-augury-470201-c8.data1.DATA`
  WHERE user_id IS NOT NULL AND product_id IS NOT NULL
),


pairs AS (
  SELECT
    r1.product_id AS product_a,
    r2.product_id AS product_b,
    r1.user_id
  FROM reviews r1
  JOIN reviews r2
    ON r1.user_id = r2.user_id
   AND r1.product_id < r2.product_id -- 중복 제거
```

```
),


co_occurrence AS (
  SELECT
    product_a,
    product_b,
    COUNT(DISTINCT user_id) AS co_review_count
  FROM pairs
  GROUP BY product_a, product_b
),

ranked AS (
  SELECT *,
    ROW_NUMBER() OVER (PARTITION BY product_a ORDER BY co_review_count DESC) AS rnk
  FROM co_occurrence
)

SELECT
  product_a AS base_product,
  product_b AS recommended_product,
  co_review_count
FROM ranked
WHERE rnk <= 5
ORDER BY co_review_count DESC;
```

## 4.결과

- 이 추천은 사용자의 공동 행동 패턴을 분석하여 상품 간 연관도를 계산.
- 별도의 머신러닝 모델 없이 SQL만으로 간단한 협업 필터링(Co-Occurrence Filtering) 구현.
- 장점:
  - 사용자 개인 데이터가 부족해도 다른 사용자 행동 패턴으로 유의미한 추천 가능.
  - 실무에서 연관 상품, 세트 상품, 번들 제안에 활용 가능.
- 한계:
  - 단순 공동 리뷰 수 기반이므로 인기 상품 편향이 있을 수 있음.

    → PMI(Pointwise Mutual Information) 등으로 가중치 조정 가능.




| 행 | base_product | recommended_product | co_review_count |
|---|---|---|---|
| 1 | B08Y1SJVV5 | B08Y1TFSP6 | 2 |
| 2 | B09QS9CWLV | B09QS9X9L8 | 1 |
| 3 | B09QS9CWLV | B09QS9X16F | 1 |
| 4 | B09HK9JH4F | B09L835C3V | 1 |
| 5 | B0B5CGTBKV | B0B5D39BCD | 1 |
| 6 | B0B5CGTBKV | B0B5DDJNH4 | 1 |
| 7 | B07JW1Y6XV | B07JW9H4J1 | 1 |
| 8 | B07JW1Y6XV | B07LGT5SSJ | 1 |
| 9 | B09YDFDVNS | B09YDFKJF8 | 1 |
| 10 | B01GGKZ0V6 | B01GGKZ4NU | 1 |
| 11 | B07WHQWXL7 | B07WJV6P1R | 1 |
| 12 | B0B4F52B5X | B0B4F5L738 | 1 |
| 13 | B07X963JNS | B08HF4W2CT | 1 |
| 14 | B0B1YY6JJL | B0BC9BW512 | 1 |
| 15 | B0B1YY6JJL | B0B1YZX72F | 1 |