

# # SQL 프로젝트(이택원)

## # 제출 형식

### 제출 템플릿

추천 시스템	내용
추천 시스템 1	<p>1. 추천 시스템 이름 → 파레토 프론티어 픽(Pareto Frontier: 가격↓·품질↑)추천시스템</p> <p>2. 추천 시스템의 테마: 추천 시스템의 고유 컨셉에 대한 설명→ 같은 리프 카테고리 내에서 더 싸고(=할인가 낮음) 또는 더 좋고(=평점 높음) 한쪽이라도 열세면 제외. 즉, 누구에게도 지지 않는(비지배) 제품만 고르는 다목적 최적화 추천합니다.</p> <p>3. 구현 로직: SQL 쿼리 설명 및 주요 로직 설명→ SQL 코드 첨부</p> <p>구현 로직 -leaf_cat = category 의 마지막 토큰 -(price, rating) 기준으로 지배되는 제품 제거( NOT EXISTS 자기조인) -신뢰도 가중: rating_count 하한</p> <p>--파레토 프론티어 픽(Pareto Frontier: 가격↓·품질↑) #standardSQL WITH typed AS ( SELECT product_id, product_name, ARRAY_REVERSE(SPLIT(CAST(category AS STRING), ' '))[SAFE_OFFSET(0)] AS leaf_cat, SAFE_CAST(REGEXP_REPLACE(CAST(discounted_price AS STRING), r'[*0-9.];','') AS FLOAT64) AS price, SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating, SAFE_CAST(REGEXP_REPLACE(CAST(rating_count AS STRING), r',;','') AS INT64) AS rating_count FROM smiling-mark-468901-i5.recommendation_system.amazon ) , filt AS ( SELECT * FROM typed WHERE leaf_cat IS NOT NULL AND price IS NOT NULL AND price &gt; 0 AND rating IS NOT NULL AND rating_count IS NOT NULL AND rating_count &gt;= 50 ) , pareto AS ( SELECT a.* FROM filt a WHERE NOT EXISTS ( SELECT 1 FROM filt b WHERE b.leaf_cat = a.leaf_cat AND b.price &lt;= a.price AND b.rating &gt;= a.rating AND (b.price &lt; a.price OR b.rating &gt; a.rating) -- 엄격 지배 ) ) SELECT leaf_cat, product_id, product_name, rating, rating_count, price FROM pareto ORDER BY leaf_cat, rating DESC, price ASC, rating_count DESC LIMIT 50;</p> <p>4. 결과: 추천 결과의 샘플을 출력하고, 이에 대해서 간단하게 설명해주세요. → 빅쿼리 화면을 캡처하여 이미지를 제출하고, 이에 대해서 간단히 설명해주세요. 결과 해석</p> <p>카테고리별로 가격·품질 효율 최상단(프론티어)에 선 제품만 남아. 캡처 시 leaf_cat 별 상위 몇 개를 보여주면 설득력. 파레토 프론티어: “카테고리별로 가격·품질 균형 최상위만 남긴 결과입니다. ‘이보다 싸고 더 좋은’ 제품이 같은 카테고리에 존재하지 않습니다.”</p> <p>파레토 프론티어 픽 (가격↓·품질↑, 누구에게도 지지 않는 제품) (주요 컬럼)</p>

추천 시스템	내용
	<p><code>leaf_cat</code> : 가장 하위 카테고리(예: "USBCables")</p> <p><code>product_id</code>, <code>product_name</code></p> <p><code>rating</code>, <code>rating_count</code></p> <p><code>price</code> : 할인가</p> <p>(예시)</p> <p><b>USBCables   B08...   "케이블 X"   평점 4.6   리뷰 3,200   ₩19,900</b></p> <p>같은 <code>leaf_cat</code> 안에서 <b>나보다 더 싸면서 평점도 더 높은</b> 제품이 <b>하나도 없기 때문에</b> 이 제품은 "프론티어"에 서 있어요. 즉, 가격-품질 면에서 <b>비지배(우수한 균형)</b> 상태입니다.</p> <p>왜 고객에게 가치가 있나요?</p> <p>"가성비 좋고 평도 좋은 것"만 걸러 보여줍니다.<b>카테고리 안에서 최적 효율</b> 제품을 즉시 파악할 수 있습니다.</p> <p>검증/주의점 (간단 체크)</p> <p><b>검증:</b> 같은 카테고리에서 이 제품보다 "가격↓ &amp; 평점↑"인 제품이 실제로 없는지 2~3개 샘플을 수작업으로 확인해보세요.<b>주의:</b> 리뷰 수가 너무 적은 제품은 프론티어에 잘못 올라갈 수 있으므로, 쿼리에서 <code>rating_count &gt;= 50</code> 같은 <b>최소 표본 필터</b>가 이미 들어가 있습니다. 필요하다면 100으로 조정해도 좋습니다.</p>

쿼리 결과

결과 저장

작업 정보결과시각화JSON실행 세부정보실행 그래프

명	leaf_cat	product_id	product_name	rating	rating_count	price
1	Condenser	B076865508	Boya BY-M1 Auxiliary Omnidirectional Lavalier Condenser Microphone with 20ft Audio Cable (Black)	4.0	68664	798.0
2	Condenser	B07JF9B592	MAGNO AU-400 Lavalier Auxiliary Omnidirectional Microphone (Black)	3.8	20218	478.0
3	CompositionNotebooks	B086Q3QMF5	Classmate Drawing Book - Unruled, 40 Pages, 210 mm x 297 mm - Pack Of 4	4.5	4951	120.0
4	CompositionNotebooks	B00P93X2H6	Classmate Pulse Spiral Notebook - 5x8 1/2 mm x 11 1/2	4.1	1269	67.0

페이지당 결과 수: 501 ~ 50 (전체 50명)

페이지당 결과 수: 50 1 - 50 (전체 50명)

추천 시스템 2	<p><b>1. 추천 시스템 이름 →</b></p> <p><b>롱테일 브랜드 젬(Long-tail Brand Gems) 추천시스템</b></p> <p><b>2. 추천 시스템의 테마: 추천 시스템의 고유 컨셉에 대한 설명→</b></p> <p>유명 대형 브랜드가 아닌, 상품 수가 적지만 평균 평점이 높은 숨은 브랜드를 발굴하고 그 중 대표작을 추천합니다.</p> <p><b>3. 구현 로직: SQL 쿼리 설명 및 주요 로직 설명→ SQL 코드 첨부</b></p> <p>구현 로직</p> <p>-브랜드 = <code>product_name</code> 의 첫 단어(간단 규칙)</p> <p>-브랜드별 아이템 수-평균 평점 집계</p> <p>-소수 브랜드(예: 2~10개) &amp; 평균 평점 상위만 남기고, 각 브랜드 대표작을 점수 ( <code>rating*log(1+rating_count)</code> )로 선정</p> <p>#standardSQL</p> <p>WITH typed AS (</p> <p>SELECT</p> <p>product_id, product_name,</p> <p>SPLIT(TRIM(CAST(product_name AS STRING)), ' ')[SAFE_OFFSET(0)] AS brand,</p> <p>SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,</p> <p>SAFE_CAST(REGEXP_REPLACE(CAST(rating_count AS STRING), r'[0-9]', '')) AS INT64) AS rating_count,</p> <p>SAFE_CAST(REGEXP_REPLACE(CAST(discounted_price AS STRING), r'[0-9]', '')) AS FLOAT64) AS price</p> <p>FROM <code>smiling-mark-468901-i5.recommendation_system.amazon</code></p> <p>),</p> <p>brand_stats AS (</p> <p>SELECT brand, COUNT() AS n_items, AVG(rating) AS avg_rating</p> <p>FROM typed</p> <p>WHERE brand IS NOT NULL AND rating IS NOT NULL</p> <p>GROUP BY brand</p> <p>HAVING COUNT() BETWEEN 2 AND 10 AND AVG(rating) &gt;= 4.2</p> <p>),</p> <p>candidates AS (</p> <p>SELECT</p> <p>t.*,</p> <p>(t.rating * LN(1 + t.rating_count)) / NULLIF(LN(1 + t.price), 0) AS brand_gem_score</p> <p>FROM typed t</p>
----------	---

```

JOIN brand_stats s USING (brand)
WHERE t.rating IS NOT NULL AND t.rating_count IS NOT NULL AND t.rating_count >= 50
AND t.price IS NOT NULL AND t.price > 0
)
SELECT
brand, product_id, product_name, rating, rating_count, price,
ROUND(brand_gem_score, 4) AS brand_gem_score
FROM candidates
ORDER BY brand_gem_score DESC, rating_count DESC
LIMIT 30;

```

4. 결과: 추천 결과의 샘플을 출력하고, 이에 대해서 간단하게 설명해주세요.

→ 빅쿼리 화면을 캡처하여 이미지를 제출하고, 이에 대해서 간단히 설명해주세요.

결과 해석

작지만 강한 브랜드의 에이스 제품이 상단. “브랜드 다양화·발굴” 관점의 큐레이션에 유용.

통테일 브랜드 점: “대형 브랜드 외에도 평판 좋은 소수 브랜드의 에이스를 찾아 추천합니다(평점×리뷰·가격 보정).”

통테일 브랜드 점 (작지만 강한 브랜드의 대표작)

(주요 컬럼)

brand : product\_name 의 첫 단어로 추정(간단 규칙)

product\_id , product\_name

rating , rating\_count , price

brand\_gem\_score : (평점 × ln(1+리뷰수)) / ln(1+가격)

(예시)

BrandZ | B09... | “무선 이어폰 Z” | 평점 4.5 | 리뷰 820 | ₩49,000 | 점수 12.34

이 브랜드는 전체 상품 수가 적지만(통테일), 브랜드 평균 평점이 높아 “숨은 강자”로 분류되고, 그중에서도 이 제품이 리뷰 수와 평점 대비 가격 효율이 가장 뛰어나서 대표로 뽑혔습니다.

왜 고객에게 가치가 있나요?

유명 대형 브랜드 외에도 \*\*품질 좋은 ‘숨은 보석’\*\*을 발견시켜 줍니다. 가격 대비 만족도가 높은 새로운 선택지를 제공합니다.

검증/주의점 (간단 체크)

검증: 결과에서 특정 브랜드를 클릭해 동일 브랜드의 다른 제품 평점을 2~3개 비교해보면, 브랜드 평균 이 실제로 높은지 확인됩니다. 주의: 브랜드 추출을 첫 단어로 했기 때문에, “New”, “2024” 같은 접두어가 브랜드로 잡힐 수 있어요. 필요하면 REGEXP\_REPLACE 로 전처리 추가 가능합니다.

쿼리 결과

결과

rank	brand	product_id	product_name	rating	rating_count	price	brand_gem_score
1	AmazonBasics	B0733ZM876	AmazonBasics Micro USB Fast Charging Cable For Android Smartphone/Personal Computer/Printer With Gold Plated Connectors (3 Feet)	4.2	92595	199.0	9.0654
2	AmazonBasics	B0733ZM876	AmazonBasics Micro USB Fast Charging Cable For Android Smartphone/Personal Computer/Printer With Gold Plated Connectors (3 Feet)	4.2	92595	199.0	9.0654
3	Redgear	B01J1C9OM	Redgear RGB15 Speed Type Gaming Mouse (Black)	4.6	33434	299.0	8.4014
4	Panasonic	B08UMFA3A	Panasonic CR2032 3V Lithium Coin Cell Battery (10 Pack)	4.4	26056	225.0	8.2691
5	ELV	B07C4QY1DL	ELV Aluminum Adjustable Mobile Phone Foldable	4.5	28978	249.0	8.3585

페이지당 결과 수: 50 1 - 30 (전체 30행)

### 추천 시스템 3

#### 1. 추천 시스템 이름 →

스펙-대비-가격 헌터(Spec-per-Price: mAh/W) 추천시스템

#### 2. 추천 시스템의 테마: 추천 시스템의 고유 컨셉에 대한 설명 →

설명/제목에서 정량 스펙(배터리 mAh 또는 충전기 W)을 뽑아 가격 대비 성능으로 랭킹을 추천합니다. (텍스트에 스펙이 없으면 자동 제외)

#### 3. 구현 로직: SQL 쿼리 설명 및 주요 로직 설명 → SQL 코드 첨부

구현 로직

- 텍스트에서  $\text{(\d{4,6})mAh}$  ,  $\text{(\d{2,3})W}$  추출

-  $\text{perf} = \text{spec\_value} / \text{price}$

- 스펙 타입별('mAh', 'W') 평균·표준편차로 정규화 z-score

- 점수 =  $\text{z} * \text{rating} * \log(1 + \text{rating\_count})$

#standardSQL

WITH base AS (

SELECT

product\_id, product\_name,

LOWER(CONCAT(IFNULL(CAST(product\_name AS STRING),''),

```

',IFNULL(CAST(about_product AS STRING),''),' ',IFNULL(CAST(review_content AS
STRING),'')) AS txt,
SAFE_CAST(REGEXP_REPLACE(CAST(discounted_price AS STRING), r'[*0-9.]','') AS
FLOAT64) AS price,
SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,
SAFE_CAST(REGEXP_REPLACE(CAST(rating_count AS STRING), r',')) AS INT64) AS
rating_count
FROM smiling-mark-468901-i5.recommendation_system.amazon
),
specs AS (
SELECT product_id, product_name, price, rating, rating_count, 'mAh' AS spec_type,
SAFE_CAST(REGEXP_EXTRACT(txt, r'(\d{4,6})\smAh') AS INT64) AS spec_value
FROM base
UNION ALL
SELECT product_id, product_name, price, rating, rating_count, 'W' AS spec_type,
SAFE_CAST(REGEXP_EXTRACT(txt, r'(\d{2,3})\sw(?:att)?') AS INT64) AS spec_value
FROM base
),
perf AS (
SELECT *, SAFE_DIVIDE(spec_value, NULLIF(price,0)) AS spec_per_price
FROM specs
WHERE spec_value IS NOT NULL AND price IS NOT NULL AND price > 0
AND rating IS NOT NULL AND rating_count IS NOT NULL AND rating_count >= 30
),
norm AS (
SELECT spec_type, AVG(spec_per_price) AS mu, STDDEV_SAMP(spec_per_price) AS sd
FROM perf GROUP BY spec_type
)
SELECT
p.spec_type, p.product_id, p.product_name,
p.rating, p.rating_count, p.price, p.spec_value,
ROUND(SAFE_DIVIDE(p.spec_per_price - n.mu, NULLIF(n.sd,0)), 3) AS z_perf,
ROUND(SAFE_DIVIDE(p.spec_per_price - n.mu, NULLIF(n.sd,0)) * p.rating * LN(1 +
p.rating_count), 4) AS spec_value_score
FROM perf p
JOIN norm n USING (spec_type)
ORDER BY spec_value_score DESC
LIMIT 30;

```

**4. 결과:** 추천 결과의 샘플을 출력하고, 이에 대해서 간단하게 설명해주세요. → 빅쿼리 화면을 캡처하여 이미  
지를 제출하고, 이에 대해서 간단히 설명해주세요.

결과 해석

**mAh/원** 또는 **W/원** 효율이 뛰어나고, 평판·리뷰가 받쳐주는 제품이 상단.

**스펙-대비-가격:** “텍스트에서 **mAh/W** 스펙을 추출해 전성비를 계산했습니다. 평균 대비 z-score가 큰 제  
품이 상위에 위치합니다.”

스펙-대비-가격 헌터 (mAh/원, W/원으로 성능/가격 비율)

(주요 컬럼)

**spec\_type** : 'mAh' 또는 'W' (텍스트에서 추출된 스펙 단위)

**spec\_value** : 숫자 스펙(예: 20000 mAh, 65 W)

**price** , **rating** , **rating\_count**

**z\_perf** : 카테고리 전체 대비 성능/가격(**z-score**)

**spec\_value\_score** : 정규화 성능 × 평점 ×  $\ln(1+\text{리뷰수})$

(예시)

**mAh | B07...** | “보조배터리 A” | **20000 mAh | ₩24,900 | 평점 4.4 | 리뷰 2,100 | z\_perf 1.8 | 점수 15.7**

동일 스펙군에서 “**mAh/원** 효율”이 평균보다 **1.8 표준편차**나 높고, 평점·리뷰도 탄탄하니 **가격 대비 성능**이  
매우 뛰어납니다.

왜 고객에게 가치가 있나요?

“**전성비**”가 좋은 제품을 빠르게 찾습니다. 숫자로 명확해 초보자도 이해하기 쉬운 **객관 지표**입니다.

검증/주의점 (간단 체크)

**검증:** 결과 상위 2~3개 제품의 상세 페이지에서 \*\*스펙 표기(예: 20000mAh/65W)\*\*가 실제로 존재하는  
지 확인하세요. **주의:** 텍스트에서 **정규표현식으로 스펙을 추출**하므로, 표기 방식이 다른 경우 놓칠 수 있습니  
다. 필요하면 패턴을 `(\d{1,2})s?w` → `(\d+)\s?w(?:att)?` 처럼 보강할 수 있습니다.

쿼리 결과										
작업 정보	결과	시각화	JSON	상행 세부정보	상행 그래프					
명	spec_type	product_id	product_name	rating	rating_count	price	spec_value	z_perf	spec_value_score	
1	W	B00N4H1K0K	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer-6 Feet (1.8 Meters), Black	4.5	107687	209.0	700	9.25	482.2965	
2	W	B00N4H1K0K	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer-6 Feet (1.8 Meters), Black	4.5	107686	209.0	700	9.25	482.2965	
3	W	B097AQ1J50	Zeltronics ZES-300-B USB Hi-J, 4-Port, Pocket Sized, Plug & Play, for Laptop & Computers	3.4	9385	179.0	719	11.204	348.433	
4	W	B0789LZTCJ	boAt Rugged v8 Extra Tough Unbreakable Braided Micro	4.2	94364	299.0	399	2.764	132.985	

페이지당 결과 수: 50 1 ~ 30 (전체 30행)

#### 추천 시스템 4

##### 1. 추천 시스템 이름➔

**호환성 범용 픽(Compatibility Breadth)추천시스템**

##### 2. 추천 시스템의 테마: 추천 시스템의 고유 컨셉에 대한 설명➔

설명/후기에서 여러 기기군(iphone, android, ipad, samsung, pixel, ps5, switch, macbook, windows)을 동시에 지원하는지 확인해 **호환성** 폭이 넓은 제품을 추천합니다.

##### 3. 구현 로직: SQL 쿼리 설명 및 주요 로직 설명➔ SQL 코드 첨부

구현 로직

-기기군 키워드별 매칭 ➔ **고유 매칭 개수(breadth)**

-점수 =  $\text{breadth} * \text{rating} * \log(1 + \text{rating\_count}) / \log(1 + \text{price})$

-하한:  $\text{breadth} \geq 2$ ,  $\text{rating} \geq 4.0$ ,  $\text{rating\_count} \geq 50$

#standardSQL

WITH base AS (

SELECT

product\_id, product\_name,

LOWER(CONCAT(IFNULL(CAST(about\_product AS STRING),''),

'IFNULL(CAST(review\_content AS STRING),''), 'IFNULL(CAST(product\_name AS

STRING),'')) AS txt,

SAFE\_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,

SAFE\_CAST(REGEXP\_REPLACE(CAST(rating\_count AS STRING), r'[', ''] AS INT64) AS

rating\_count,

SAFE\_CAST(REGEXP\_REPLACE(CAST(discounted\_price AS STRING), r'[^\d-0-9.]\*', '') AS

FLOAT64) AS price

FROM smiling-mark-468901-i5.recommendation\_system.amazon

),

flags AS (

SELECT

product\_id, product\_name, rating, rating\_count, price,

ARRAY(

SELECT DISTINCT lab FROM UNNEST([

STRUCT('iphone' AS lab, REGEXP\_CONTAINS(txt, r'\biphone\b') AS hit),

('android', REGEXP\_CONTAINS(txt, r'\bandroid\b')),

('ipad', REGEXP\_CONTAINS(txt, r'\bipad\b')),

('samsung', REGEXP\_CONTAINS(txt, r'\bsamsung\b')),

('pixel', REGEXP\_CONTAINS(txt, r'\bpixel\b')),

('ps5', REGEXP\_CONTAINS(txt, r'\bps5\b')),

('switch', REGEXP\_CONTAINS(txt, r'\bswitch\b')),

('macbook', REGEXP\_CONTAINS(txt, r'\bmacbook\b')),

('windows', REGEXP\_CONTAINS(txt, r'\bwindows\b'))

) WHERE hit

) AS hits

FROM base

)

SELECT

product\_id, product\_name, rating, rating\_count, price,

ARRAY\_LENGTH(hits) AS breadth,

ROUND(ARRAY\_LENGTH(hits) \* rating \* LN(1 + rating\_count) / NULLIF(LN(1 + price), 0), 4)

AS compat\_breadth\_score,

hits

FROM flags

WHERE rating IS NOT NULL AND rating\_count IS NOT NULL AND price IS NOT NULL AND

price > 0

AND ARRAY\_LENGTH(hits) >= 2 AND rating >= 4.0 AND rating\_count >= 50

ORDER BY compat\_breadth\_score DESC

LIMIT 30;

##### 4. 결과: 추천 결과의 샘플을 출력하고, 이에 대해서 간단하게 설명해주세요. ➔ 빅쿼리 화면을 캡처하여 이

미지를 제출하고, 이에 대해서 간단히 설명해주세요.

결과 해석

여러 생태계에 동시에 맞는 ‘범용’ 제품이 상단. 실제 서비스에서 호환 질문이 많은 카테고리에 유용.

호환성 범용: “여러 생태계를 동시에 지원하는 제품일수록 점수가 높습니다. 선물/기관 구매에 적합합니다.”

호환성 범용 픽 (여러 생태계 지원 제품)

(주요 컬럼)

hits : 감지된 기기 생태계 키워드 목록(예: ["iphone","android","windows"] )

breadth : 매칭된 생태계 개수

rating , rating\_count , price

compat\_breadth\_score : breadth × 평점 × ln(1+리뷰수) / ln(1+가격)

(예시)B08... | “허브/어댑터 Y” | hits=[iphone, android, windows] | breadth=3 | 평점 4.3 | 리뷰 980 | ₩34,900 | 점수 9.8

아이폰·안드로이드·윈도우 모두에서 쓸 수 있는 범용성 높은 제품이고, 평판/리뷰도 좋아서 호환성 걱정이 적은 선택입니다.

왜 고객에게 가치가 있나요?

내가 어떤 기기를 쓰든 잘 맞는 제품을 고를 수 있어요. 선물/기업구매처럼 사용 환경이 섞인 상황에 특히 유용합니다.

검증/주의점 (간단 체크)

검증: 상위 결과 2~3개에서 제품 설명에 실제 호환 문구(예: “for iPhone/Android/Windows”)가 있는지 확인하세요. 주의: 키워드 세트는 예시입니다. “Lightning”, “Type-C”, “MagSafe” 등 추가 키워드를 넣으면 탐자가 더 정확해집니다.

쿼리 결과									
작업 정보	결과	시각화	JSON	설명 세부정보	설명 그래프				
원	product_id	product_name	rating	rating_count	price	breadth	compat_breadth...	hits	
1	B07QKH3691	STRIPP FPD-01 Multi Angle Mobile/Tablet Tabletop Stand, Phone Holder for iPhone, Android, Samsung, OnePlus, Xiaomi. Portable, Foldable Cell	4.3	42641	99.0	5	49.7708	iphone	
								android	
								ipad	
								samsung	
								pixel	
2	B07Y1P880V	Weyone Nylon Braided Lish Type C 3Ft 1M 5A Fast Charger Cable For Samsung Galaxy S9 S8 (WC0301, Black)	4.2	10576	325.0	5	40.3522	iphone	
								ipad	
								samsung	
								pixel	
								switch	
								macbook	
3	B09Y14LFP3	STRIPP UPK02W Multi Angle Tablet/Mobile Stand, Holder for iPhone, Android, Samsung, OnePlus, Xiaomi. Portable/Foldable	4.1	2451	99.0	5	34.7426	iphone	
								android	
								ipad	
								samsung	
								pixel	
4	B09VZ8G1N	STRIPP Multi Angle Tablet/Mobile Stand, Holder	4.1	2451	99.0	5	34.7426	iphone	

페이지당 결과 수: 50 1 ~ 30 (전체 30행)

## 추천 시스템 5

### 1. 추천 시스템 이름→

정직한 할인 지수(Honest Deal Index) 추천 시스템

### 2. 추천 시스템의 테마: 추천 시스템의 고유 컨셉에 대한 설명→

카테고리 내 \*\*정가 뺀튀기(가짜 세일)\*\*를 걸러내고, 진짜로 유리한 가격/품질 조합만 추천합니다.

### 3. 구현 로직: SQL 쿼리 설명 및 주요 로직 설명→ SQL 코드 첨부

구현 로직

-leaf\_cat별 정가/할인가 중앙값

- markup\_index = actual\_price / med\_actual\_price (너무 높으면 제외)

-점수 = discount\_pct \* (med\_actual\_price/actual\_price) \* rating \* log(1+rating\_count)

#standardSQL

WITH typed AS (

SELECT

product\_id, product\_name,

ARRAY\_REVERSE(SPLIT(CAST(category AS STRING), '|'))[SAFE\_OFFSET(0)] AS leaf\_cat,

SAFE\_CAST(REGEXP\_REPLACE(CAST(actual\_price AS STRING), r'[^\d.]\*')) AS FLOAT64) AS

actual\_price,

SAFE\_CAST(REGEXP\_REPLACE(CAST(discounted\_price AS STRING), r'[^\d.]\*')) AS FLOAT64)

AS price,

CASE

WHEN REGEXP\_CONTAINS(CAST(discount\_percentage AS STRING), r'%')

THEN SAFE\_CAST(REPLACE(CAST(discount\_percentage AS STRING), '%','')) AS FLOAT64)/100

ELSE SAFE\_CAST(CAST(discount\_percentage AS STRING) AS FLOAT64)

```

END AS discount_pct,
SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,
SAFE_CAST(REGEXP_REPLACE(CAST(rating_count AS STRING), r',','') AS INT64) AS
rating_count
FROM smiling-mark-468901-i5.recommendation_system.amazon
),
leaf_med AS (
SELECT
leaf_cat,
APPROX_QUANTILES(actual_price, 100)[OFFSET(50)] AS med_actual,
APPROX_QUANTILES(price, 100)[OFFSET(50)] AS med_price
FROM typed
WHERE leaf_cat IS NOT NULL AND actual_price IS NOT NULL AND price IS NOT NULL
GROUP BY leaf_cat
),
joined AS (
SELECT
t.*, m.med_actual, m.med_price,
SAFE_DIVIDE(t.actual_price, NULLIF(m.med_actual,0)) AS markup_index
FROM typed t JOIN leaf_med m USING (leaf_cat)
)
SELECT
product_id, product_name, leaf_cat, rating, rating_count, price, discount_pct,
ROUND(markup_index, 3) AS markup_index,
ROUND(discount_pct * SAFE_DIVIDE(med_actual, NULLIF(actual_price,0)) * rating * LN(1 +
rating_count), 4) AS honest_deal_score
FROM joined
WHERE price IS NOT NULL AND price > 0 AND rating IS NOT NULL AND rating_count IS NOT
NULL
AND discount_pct IS NOT NULL AND discount_pct >= 0.05
AND markup_index <= 2.5
AND price <= med_price
ORDER BY honest_deal_score DESC, rating_count DESC
LIMIT 30;

```

4. 결과: 추천 결과의 샘플을 출력하고, 이에 대해서 간단하게 설명해주세요. ➡ 빅쿼리 화면을 캡처하여 이미지를 제출하고, 이에 대해서 간단히 설명해주세요.

#### 결과 해석

정가 뺀 뒤가 위험이 낮고, 실제로 카테고리 기준 대비 유리한 가격·평판을 가진 상품만 남습니다.

“카테고리 중앙값을 영커로 삼아 정가 부풀림을 배제, 진짜 유리한 가격/품질을 보장합니다.”

정직한 할인 지수 (가짜 세일 걸러내고 진짜 할인)

(주요 컬럼)

leaf\_cat : 하위 카테고리

price (할인가), discount\_pct , rating , rating\_count

markup\_index : 내 정가 / 카테고리 정가 중앙값

(값이 너무 높으면 “정가 부풀리기” 의심)

honest\_deal\_score : 할인을  $\times$  (카테고리 정가중앙/내 정가)  $\times$  평점  $\times \ln(1+\text{리뷰수})$

조건: price <= 카테고리 할인가 중앙값 , markup\_index <= 2.5

(예시)

**USBCables | B0A... | “케이블 Z” | price ₩7,900 | 할인 40% | 평점 4.4 | 리뷰 3,400 | markup\_index 1.2 | 점수 10.2**

정가가 \*\*카테고리 중앙값 근처(1.2배)\*\*라 과한 부풀림이 아니고, 현재 할인가도 중앙값 이하, 평판/리뷰도 좋아서 진짜 득템입니다.

왜 고객에게 가치가 있나요?

“정가를 부풀려 크게 할인하는 척”을 필터링합니다.카테고리 기준으로 진짜 유리한 가격을 가진 제품만 보여줍니다.

검증/주의점 (간단 체크)

검증: 상위 2~3개 제품의 정가/할인가를 비슷한 카테고리 제품과 비교해 보세요. 실제로 중앙값보다 유리한지 눈으로도 확인됩니다.주의: 특정 카테고리의 표본 수가 너무 적으면 중앙값이 불안정합니다. 쿼리에서 표본 하한(예: HAVING COUNT(\*) >= 20 )을 걸어둔 이유입니다.

쿼리 결과

결과 저장

작업 정보

결과

시각화

JSON

실행 세부정보

실행 그래프

명	product_id	product_name	leaf_cat	rating	rating_count	price	discount_pct	markup_index	honest_deal_score
1	B0BK4PSZ3V	Tukzer Capacitive Stylus Pen for Touch Screens Devices, Fine Point, Lightweight Metal Body with Magnetism Cover Cap for	StylusPens	3.8	16557	349.0	0.65	0.167	144.0908
2	B0BK4RDQ71	Tukzer Capacitive Stylus Pen for Touch Screens Devices, Fine Point, Lightweight Metal Body with Magnetism Cover Cap for	StylusPens	3.8	16557	349.0	0.65	0.167	144.0908
3	B0BK4PSZ3V	Tukzer Capacitive Stylus Pen for Touch Screens Devices, Fine Point, Lightweight Metal Body with Magnetism Cover Cap for	StylusPens	3.8	16557	349.0	0.65	0.167	144.0908
4	B0B29BD54H	Prolet Classic Bumper Case Cover for Samsung Galaxy Watch 4 44mm TPU Plated Full Screen Protector (Black)	SmartWatches	3.7	465	265.0	0.73	0.125	192.7903
5	B0B9MQXNFB	FLIX (Beetel Flow USB to Micro USB PVC Data Sync & 12W(2.4A) Fast Charging Cable,Made in India,480Mbps Data Sync,Solid Cable,1 Meter	USBCables	4.0	9378	57.89	0.71	0.236	110.2971

페이지당 결과 수:

50

1 ~ 30 (전체 30명)