SQL로 추천 시스템 설계하기

0. 데이터 준비

원본 Kaggle 데이터셋: **Amazon Sales Dataset (Amazon India, ₹ 단위)**

1단계: 원본 테이블 로드 (amazon.csv)

- **스키마**: 자동 감지 ON

- **건너뛸 헤더 행**: 1 (첫 줄은 컬럼명이니까)

- **따옴표 줄바꿈 허용**: 체크 (리뷰 본문에 줄바꿈 들어갈 수 있음)

- **허용 오류 레코드 수**: 10~20 (기호나 잘못된 데이터가 있어도 무시하고 로드)

2단계: 정제 테이블 생성 (amazon_clean)

- `discounted_price` / `actual_price` / `discount_percentage` → FLOAT

- `rating_count` → INTEGER

- rating → STRING으로 들어와 있으므로 FLOAT 변환 필요

```
CREATE OR REPLACE TABLE cobalt-chalice-470802-i8.day4.amazon_clean AS
SELECT
product_id,
product_name,
category,
discounted_price,
actual_price,
discount_percentage,
SAFE_CAST(rating AS FLOAT64) AS rating,   -- 변환
rating_count,
about_product,
user_id,
user_name,
review_id,
review_title,
review_content,
img_link,
product_link
FROM cobalt-chalice-470802-i8.day4.amazon;
```

```
-- 숫자 범위 확인
SELECT
  MIN(discounted_price) AS min_disc, MAX(discounted_price) AS max_disc,
  MIN(actual_price) AS min_actual, MAX(actual_price) AS max_actual,
  MIN(discount_percentage) AS min_pct, MAX(discount_percentage) AS max_pct,
  MIN(rating) AS min_rating, MAX(rating) AS max_rating,
  MIN(rating_count) AS min_rc, MAX(rating_count) AS max_rc
FROM `cobalt-chalice-470802-i8.day4.amazon_clean`;
```

행	min_disc	max_disc	min_actual	max_actual	min_pct	max_pct	min_rating	max_rating	min_rc	max_rc
1	39.0	77990.0	39.0	139900.0	0.0	0.94	2.0	5.0	2	426973

```
-- 혹시 숫자 컬럼에 문자 섞였는지(있으면 1 이상 나옴)
SELECT
  SUM(CASE WHEN REGEXP_CONTAINS(CAST(discounted_price AS STRING),
r'[^0-9.\-]') THEN 1 ELSE 0 END) AS bad_disc_price,
  SUM(CASE WHEN REGEXP_CONTAINS(CAST(actual_price AS STRING), r'[^0-
9.\-]') THEN 1 ELSE 0 END) AS bad_actual_price,
  SUM(CASE WHEN REGEXP_CONTAINS(CAST(discount_percentage AS STRIN
G), r'[^0-9.\-]') THEN 1 ELSE 0 END) AS bad_pct
FROM `cobalt-chalice-470802-i8.day4.amazon_clean`;
```

행	bad_disc_price ▼	bad_actual_price ▼	bad_pct ▼
1	0	0	0

👉 **숫자 데이터 점검 결과**

구분	컬럼명	최소값	최대값	비고
가격	discounted_price	39	77,990	(할인 적용가)
	actual_price	39	139,900	(정가)
할인율	discount_percentage	0.00	0.94	최대 94% 할인
평점	rating	2.0	5.0	2점 이하 없음
리뷰 수	rating_count	2	426,973	분포 폭이 큼

1. 추천 지표 정의

- **PopQ (Popular × Quality)**
    - PopQ = rating × LOG10(1 + rating_count)
    - **정의:** 평점과 리뷰 수를 결합한 지표
    - **의도:** 평점이 높아도 리뷰가 1~2개뿐이면 신뢰하기 어려움.
      리뷰 수를 로그로 반영해, 리뷰가 많을수록 점수가 높아지도록 설계.
    - **해석:** "사람들이 많이 평가했는데도 평점이 높은 상품 = 신뢰할 수 있는 인기 상품"

- **QPR (Quality–Price Ratio)**
    - QPR = rating ÷ LOG10(1 + discounted_price)

- 　　○ **정의:** 가격 대비 품질을 나타내는 지표

- 　　○ **의도:** 같은 평점이라면 더 저렴한 상품이 더 매력적임.

　　　　비싼 상품에는 log(가격)으로 완만한 페널티를 주어 공정하게 비교.

- 　　○ **해석:** "비슷한 만족도를 더 저렴하게 누릴 수 있는 상품 = 가성비 좋은 상품"

- **베이즈 점수 (Bayesian Average)**

  - ○ `Bayes = (C*m + rating*rating_count) ÷ (m + rating_count)`

    (C = 전체 평균 평점, m = 50)

  - ○ **정의:** 리뷰 수가 적을수록 전체 평균에 수렴하도록 보정

  - ○ **의도:** 리뷰 1~2개로 5.0점을 받은 상품은 과대평가될 수 있음.

    가상의 리뷰(m=50)를 추가한 것처럼 계산해 신뢰도 확보.

  - ○ **해석:** "리뷰 수가 많아도 여전히 좋은 점수를 유지하는 상품 = 믿고 사는 스테디셀러"

---

2. 추천 시스템 설계 및 구현

1) 예산대별 베스트

"내 지갑 사정에 맞춘 찐템 리스트 💸"

- **컨셉:** 사용자의 예산 구간 안에서 **가장 신뢰할 수 있는 베스트 상품**을 추천

- **핵심 질문:** "같은 예산 안에서, 가장 많은 사람들에게 좋은 평가를 받은 상품은 무엇일까?"

- **SQL 로직:**

  - ○ 가격을 구간별로 나눔 (`price_band`)

  - ○ PopQ = `rating × LOG10(1+rating_count)` 계산

  - ○ 구간별 `DENSE_RANK()` 로 Top-N 상품 선정

```
-- 가격대를 구간으로 나누고, 각 구간에서 PopQ 점수가 높은 상품 추천
WITH base AS (
SELECT
product_id, product_name, rating, rating_count, discounted_price,
-- 가격 구간(루피 단위)으로 구분
CASE
WHEN discounted_price < 500 THEN '₹0–499'
WHEN discounted_price < 2000 THEN '₹500–1,999'
WHEN discounted_price < 5000 THEN '₹2,000–4,999'
```

```
WHEN discounted_price < 15000 THEN '₹5,000-14,999'
ELSE '₹15,000+'
END AS price_band,
-- PopQ = 평점 × log(리뷰수)
rating * LOG10(1 + rating_count) AS popq
FROM cobalt-chalice-470802-i8.day4.amazon_clean
WHERE rating IS NOT NULL AND rating_count IS NOT NULL
AND discounted_price > 0
AND rating >= 3.8      -- 최소 평점 조건
AND rating_count >= 10  -- 최소 리뷰 수 조건
),
ranked AS (
SELECT base.*,
-- 구간별 순위: PopQ 점수 기준
DENSE_RANK() OVER (PARTITION BY price_band ORDER BY popq DESC) AS rk
FROM base
)
SELECT price_band, rk, product_id, product_name,
rating, rating_count, discounted_price, ROUND(popq,3) AS popq
FROM ranked
WHERE rk <= 2   -- 구간별 Top 2
ORDER BY
-- 구간 순서를 가격대별로 정렬
CASE price_band
WHEN '₹0-499' THEN 1
WHEN '₹500-1,999' THEN 2
WHEN '₹2,000-4,999' THEN 3
WHEN '₹5,000-14,999' THEN 4
ELSE 5
END,
rk;
```

결과 (샘플)

행	price_band	rk	product_id	product_name	rating	rating_count	discounted_price	popq
1	₹0–499	1	B014I8SX4Y	Amazon Basics High-Speed HDMI Cable, 6 Feet (2-Pack),Black	4.4	426973	309.0	24.774
2	₹0–499	1	B014I8SSD0	Amazon Basics High-Speed HDMI Cable, 6 Feet - Supports Ethernet, 3D, 4K video,Black	4.4	426973	309.0	24.774
3	₹0–499	1	B07KSMBL2H	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	4.4	426973	219.0	24.774
4	₹0–499	2	B07KSMBL2H	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	4.4	426972	219.0	24.774
5	₹500–1,999	1	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	23.903
6	₹500–1,999	2	B0BR4F878Q	Swiffer Instant Electric Water Heater Faucet Tap Home-Kitchen Instantaneous Water Heater Tank less for Tap, LED Electric Head Water Heaters Tail Gallon Comfort(3000W) ((Pack of 1))	4.8	53803	1439.0	22.708
7	₹2,000–4,999	1	B08HV83HL3	MI Power Bank 3i 20000mAh Lithium Polymer 18W Fast Power Delivery Charging	Input- Type C	Micro USB| Triple Output	Sandstone Black	4.3	178912	2049.0	22.586
8	₹2,000–4,999	2	B08ZJDWTJ1	Seagate Expansion 1TB External HDD - USB 3.0 for Windows and Mac with 3 yr Data Recovery Services, Portable Hard Drive (STKM1000400)	4.5	50810	4098.0	21.177
9	₹5,000–14,999	1	B09GFPVD9Y	Redmi 9 Activ (Carbon Black, 4GB RAM, 64GB Storage)	Octa-core Helio G35	5000 mAh Battery	4.1	313836	8499.0	22.536
10	₹5,000–14,999	1	B09GFLXVH9	Redmi 9A Sport (Coral Green, 2GB RAM, 32GB Storage)	2GHz Octa-core Helio G25 Processor	5000 mAh Battery	4.1	313836	6499.0	22.536
11	₹5,000–14,999	2	B09GFM8CGS	Redmi 9A Sport (Carbon Black, 2GB RAM, 32GB Storage)	2GHz Octa-core Helio G25 Processor	5000 mAh Battery	4.1	313832	6499.0	22.536
12	₹5,000–14,999	2	B09GFPN6TP	Redmi 9A Sport (Coral Green, 3GB RAM, 32GB Storage)	2GHz Octa-core Helio G25 Processor	5000 mAh Battery	4.1	313832	7499.0	22.536
13	₹15,000+	1	B09RFC46VP	Redmi 108 cm (43 inches) 4K Ultra HD Android Smart LED TV X43	L43R7-7AIN (Black)	4.2	45238	26999.0	19.553
14	₹15,000+	1	B08Y55LPBF	Redmi 126 cm (50 inches) 4K Ultra HD Android Smart LED TV X50	L50M6-RA (Black)	4.2	45238	32999.0	19.553
15	₹15,000+	2	B09Q5P2MT3	OnePlus 108 cm (43 inches) Y Series Full HD Smart Android LED TV 43 Y1S (Black)	4.2	34899	24999.0	19.08
16	₹15,000+	2	B09Q5SWVBJ	OnePlus 80 cm (32 inches) Y Series HD Ready Smart Android LED TV 32 Y1S (Black)	4.2	34899	15999.0	19.08

결과 해석:

- 0–499 구간에서는 **AmazonBasics HDMI 케이블 시리즈**가 평점과 리뷰 수 모두 압도적이라 Top1~2를 독식.

- 2000–4999 구간에서는 **MI Power Bank**와 **Seagate 외장하드**가 각각 휴대용 충전기와 스토리지 부문에서 최고 PopQ 점수로 선정.

- 15000+ 구간에서는 **Redmi 스마트폰 시리즈**가 상위권을 차지, 높은 리뷰 수와 준수한 평점으로 신뢰도 확보.

2) 품질-가격 지수(QPR)

"돈값 제대로 하는 인생템 찾기 💎"

- **컨셉:** 같은 평점이라면 **더 저렴한 상품**이 더 효율적임 → 가성비 높은 상품을 우선 추천

- **핵심 질문:** "비슷한 만족도를 더 적은 비용으로 줄 수 있는 상품은 무엇일까?"

- **SQL 로직:**

    - QPR = `rating ÷ LOG10(1+discounted_price)`

    - 최소 평점·리뷰 수 기준으로 기본 품질 보장

```
-- QPR = 평점 ÷ log(가격) → 가격 대비 품질이 좋은 상품 추천
SELECT
  product_id, product_name,
  rating, rating_count, discounted_price,
  ROUND(rating / LOG10(1 + discounted_price), 3) AS qpr
FROM `cobalt-chalice-470802-i8.day4.amazon_clean`
WHERE rating IS NOT NULL AND rating_count IS NOT NULL
```

```
    AND discounted_price > 0
    AND rating >= 4.2        -- 기본 품질 가드
    AND rating_count >= 20    -- 리뷰 수 최소 조건
ORDER BY qpr DESC
LIMIT 10;
```

결과 (샘플)

행	product_id	product_name	rating	rating_count	discounted_price	qpr
1	B07KCMR8D6	Classmate Octane Neon- Blue Gel Pens(Pack of 5)|Smooth Writing Pen|Attractive body colour for Boys & Girls|Waterproof ink for smudge free writing|Preferred by Students for Exam|Study at home essential	4.3	5792	50.0	2.518
2	B09RF2QXGX	Gizga Essentials Webcam Cover, Privacy Protector Webcam Cover Slide, Compatible with Laptop, Desktop, PC, Smartphone, Protect Your Privacy and Security, Strong Adhesive, Set of 3, Black	4.3	255	69.0	2.33
3	B0116MIKKC	Goldmedal Curve Plus 202042 Plastic Spice 3-Pin 240V Universal Travel Adaptor (White)	4.5	11339	99.0	2.25
4	B00N1U7JXM	3M Post-it Sticky Note Cube, 200 Sheets (4 Colors x 50 Sheets)	3" x 3" Size	For notes, reminders, study, school and organizing	4.4	7429	90.0	2.246
5	B07JB2Y4SR	Classmate Octane Colour Burst-Multicolour Gel Pens (Pack of 10)	Gold & Silver Glitter Sparkle Pens|10 colour ink shades for art lovers and kids|Fun at home essentials	4.4	10718	90.0	2.246
6	B0B8ZWNR5T	STRIFF 12 Pieces Highly Flexible Silicone Micro USB Protector, Mouse Cable Protector, Suit for All Cell Phones, Computers and Chargers (Black)	4.2	1949	79.0	2.207
7	B0BBW521YC	LAPSTER 12pcs Spiral Cable Protectors for Charger, Wires, Data Charger Cable Protector for Computers, Cell Phones etc.(Grey)	4.4	305	99.0	2.2
8	B0926V9CTV	Elv Mobile Phone Mount Tabletop Holder for Phones and Tablets - Black	4.3	2351	89.0	2.2
9	B00LM4X3XE	Parker Quink Ink Bottle (Black)	4.3	3061	90.0	2.195
10	B086Q3QMFS	Classmate Drawing Book - Unruled, 40 Pages, 210 mm x 297 mm - Pack Of 4	4.5	4951	120.0	2.161

결과 해석:

- QPR 상위권은 대부분 **₹50~₹120 구간 초저가 문구류/액세서리(볼펜, 포스트잇, 젤펜, 케이블 프로텍터 등)**.

- 낮은 가격에도 불구하고 평점이 4.3~4.5로 높아, **'저렴하지만 만족도가 높은 숨은 보석 템'**임을 확인할 수 있음.

- 결론: **실용적이고 저렴한 생활필수품**이 가성비 Best로 추천됨.

3) 카테고리 균형 피드

"익숙한 취향에서 한 발 더: 카테고리별 베스트 탐색 🌈"

- **컨셉:** 특정 분야에만 치우치지 않고, **모든 카테고리에서 고르게 Top 상품**을 추천

- **핵심 질문:** "내가 늘상 보던 카테고리 외에, 다른 카테고리에서도 많은 사람들이 선택했고 평점도 높은 상품은 무엇일까?"

- **SQL 로직:**
  - `category` 에서 leaf 카테고리 추출
  - PopQ = `rating × LOG10(1+rating_count)` 계산
  - 카테고리별 `DENSE_RANK()` 로 순위 매겨 Top-N 상품 선정

```sql
-- 카테고리별로 균형 있게 추천 (leaf 기준, Top3)
WITH base AS (
  SELECT
    product_id, product_name, rating, rating_count,
    -- leaf 카테고리 추출: 'A|B|C' → 'C'
    SPLIT(category, '|')[OFFSET(ARRAY_LENGTH(SPLIT(category, '|')) - 1)] AS category_leaf,
    -- PopQ = 품질 × 인기도(로그 보정)
    rating * LOG10(1 + rating_count) AS popq
  FROM `cobalt-chalice-470802-i8.day4.amazon_clean`
  WHERE category IS NOT NULL
    AND rating IS NOT NULL
    AND rating_count IS NOT NULL
    AND rating >= 3.8              -- 최소 평점 조건
    AND rating_count >= 10         -- 최소 리뷰 수 조건
),
cat_size AS (
  -- 너무 작은 카테고리(샘플 부족) 제외하기 위한 기준 계산
  SELECT category_leaf, COUNT(*) AS n_items
  FROM base
  GROUP BY category_leaf
),
ranked AS (
  SELECT
    b.*,
    -- 카테고리별 순위(높은 PopQ 우선)
    DENSE_RANK() OVER (
      PARTITION BY b.category_leaf
      ORDER BY b.popq DESC, b.rating_count DESC, b.product_id
    ) AS rank_in_category
  FROM base AS b
  JOIN cat_size AS s USING (category_leaf)
  WHERE s.n_items >= 20            -- 샘플이 너무 적은 카테고리는 제외
)
```

```
SELECT
  category_leaf, rank_in_category, product_id, product_name,
  rating, rating_count, ROUND(popq, 3) AS popq
FROM ranked
WHERE rank_in_category <= 3      -- 카테고리별 Top 3
ORDER BY
  category_leaf,
  rank_in_category,
  rating_count DESC,              -- 동점일 때 리뷰 많은 상품 우선
  product_id                      -- 그래도 동점이면 ID 순으로 고정
```

결과 (샘플)

행	category_leaf	rank_in_category	product_id	product_name	rating	rating_count	popq
1	DryIrons	1	B008YW8M0G	Bajaj DX-7 1000W Dry Iron with Advance Soleplate and Anti-bacterial German Coating Technology, White	4.2	46647	19.609
2	DryIrons	2	B008YW3CYM	USHA EI 1602 1000 W Lightweight Dry Iron with Non-Stick Soleplate (Multi-colour)	4.1	37126	18.736
3	DryIrons	3	B01C8P29T4	Bajaj Majesty DX-11 1000W Dry Iron with Advance Soleplate and Anti-bacterial German Coating Technology, White and Blue	4.2	24247	18.416
4	HDMICables	1	B014I8SSD0	Amazon Basics High-Speed HDMI Cable, 6 Feet - Supports Ethernet, 3D, 4K video,Black	4.4	426973	24.774
5	HDMICables	2	B014I8SX4Y	Amazon Basics High-Speed HDMI Cable, 6 Feet (2-Pack),...	4.4	426973	24.774
6	HDMICables	3	B07KSMBL2H	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	4.4	426973	24.774
7	In-Ear	1	B07GPXXNNG	boAt Bassheads 100 in Ear Wired Earphones with Mic(Taffy Pink)	4.1	363713	22.799
8	In-Ear	2	B07GQD4K6L	boAt Bassheads 100 in Ear Wired Earphones with Mic(Furious Red)	4.1	363713	22.799
9	In-Ear	3	B071Z8M4KX	boAt BassHeads 100 in-Ear Wired Headphones with Mic (...	4.1	363711	22.799
10	Mice	1	B004IO5BMQ	Logitech M235 Wireless Mouse, 1000 DPI Optical Tracking, 12 Month Life Battery, Compatible with Windows, Mac, Chromebook/PC/Laptop	4.5	54405	21.31

결과 해석:

- **DryIrons** 카테고리에서는 Bajaj와 USHA 제품이 Top3를 형성하며, 가전 필수 아이템으로 신뢰성 확보.

- **HDMICables**는 AmazonBasics 제품군이 압도적으로 상위권을 차지.

- **In-Ear 이어폰**은 boAt 브랜드가 1~3위를 모두 차지해 시장 점유율이 두드러짐.

- 따라서 특정 카테고리에만 노출이 치우치지 않고, 사용자는 다양한 분야의 베스트셀러를 접할 수 있음.

4) 신뢰도 챔피언 (베이즈 Top-N)

"후기 많은 게 국룰 👍 믿고 사는 스테디셀러"

- **컨셉:** 리뷰 수가 적으면 과대평가될 수 있음 → **리뷰가 많아도 여전히 높은 평점**을 유지하는 상품만 추천

- **핵심 질문:** "많은 이용자들이 평가했는데도 변함없이 높은 점수를 받은 상품은 무엇일까?"

- **SQL 로직:**

  - 전역 평균 C 계산

  - 베이즈 점수 `(C*m + rating*count) ÷ (m+count)` 적용 (m=50)

  - 리뷰 적은 상품은 평균 쪽으로 보정, 리뷰 많은 상품은 원래 평점에 가깝게 유지

```
-- 베이즈 평균으로 리뷰 적은 상품은 보정, 리뷰 많은 스테디셀러 추천
WITH stats AS (
  -- 전체 평균 평점 C 계산
  SELECT AVG(rating) AS C
  FROM `cobalt-chalice-470802-i8.day4.amazon_clean`
  WHERE rating IS NOT NULL
),
scored AS (
  SELECT
  p.product_id, p.product_name,
  p.rating, p.rating_count,
  -- 베이즈 점수 = (C*m + 평점×리뷰수) ÷ (m+리뷰수)
  (s.C * 50 + p.rating * p.rating_count) / (50 + p.rating_count) AS bayes_score
  FROM `cobalt-chalice-470802-i8.day4.amazon_clean` p
  CROSS JOIN stats s
  WHERE p.rating IS NOT NULL AND p.rating_count IS NOT NULL
)
SELECT product_id, product_name,
       rating, rating_count,
       ROUND(bayes_score, 3) AS bayes_score
FROM scored
WHERE rating_count >= 5        -- 리뷰 5개 미만은 제외
ORDER BY bayes_score DESC
LIMIT 10;
```

결과 (샘플)

행	product_id	product_name	rating	rating_count	bayes_score
1	B0BR4F878Q	Swiffer Instant Electric Water Heater Faucet Tap Home-Kitchen Instantaneous Water Heater Tank less for Tap, LED Electric Head Water Heaters Tail Gallon Comfort(3000W) ((Pack of 1))	4.8	53803	4.799
2	B0B53DS4TF	Instant Pot Air Fryer, Vortex 2QT, Touch Control Panel, 360° EvenCrisp™ Technology, Uses 95 % less Oil, 4-in-1 Appliance: Air Fry, Roast, Bake, Reheat (Vortex 1.97Litre, Black)	4.8	3964	4.791
3	B0B23LW7NV	Spigen EZ Fit Tempered Glass Screen Protector for iPhone 14 Pro Max - 2 Pack (Sensor Protection)	4.7	7779	4.696
4	B09WN3SRC7	Sony Bravia 164 cm (65 inches) 4K Ultra HD Smart LED Google TV KD-65X74K (Black)	4.7	5935	4.695
5	B0BP89YBC1	Campfire Spring Chef Prolix Instant Portable Water Heater Geyser 1Ltr. for Use Home Stainless Steel Baking Rack	Restaurant	Office	Labs	Clinics	Saloon	with Installation Kit (With MCB)	4.7	2591	4.689
6	B0BM9H2NY9	Multifunctional 2 in 1 Electric Egg Boiling Steamer Egg Frying Pan Egg Boiler Electric Automatic Off with Egg Boiler Machine Non-Stick Electric Egg Frying Pan-Tiger Woods (Multy)	4.7	2300	4.687
7	B0BM4KTNL1	FIGMENT Handheld Milk Frother Rechargeable, 3-Speed Electric Frother for Coffee with 2 Whisks and Coffee Decoration Tool, Coffee Frother Mixer, CRESCENT ENTERPRISES VRW0.50BK (A1)	4.7	1729	4.683
8	B095RTJH1M	Spigen EZ Fit Tempered Glass Screen Protector Guard for iPhone 14/13/13 Pro - 2 Pack	4.6	26603	4.599
9	B01J1CFO5I	Redgear MP35 Speed-Type Gaming Mousepad (Black/Red)	4.6	33434	4.599
10	B01MQ2A86A	Logitech M331 Silent Plus Wireless Mouse, 2.4GHz with USB Nano Receiver, 1000 DPI Optical Tracking, 3 Buttons, 24 Month Life Battery, PC/Mac/Laptop - Black	4.6	12375	4.598

결과 해석:

- 베이즈 점수는 리뷰가 많은 상품일수록 안정적으로 높은 순위에 오름.

- 리뷰 30만 건 이상에도 4.3~4.4점을 유지한 상품들이 상위권 → **대규모 고객 검증을 통과한 스테디셀러**임을 의미.

- 결론: 단순히 별점이 아닌 **"시간이 지나도, 많은 사람에게도" 꾸준히 사랑받는 상품**을 보여 줌.

5) 위험딜 대체 추천

"과대 할인 저품질 대신, 안전하게 득템 💥"

- **컨셉:** 할인율은 높지만 평점이 낮은 상품은 위험 → 같은 카테고리에서 **더 믿을 수 있는 대체 재** 추천

- **핵심 질문:** "후회할 가능성 큰 위험 딜 대신, 어떤 대체 상품을 선택하면 안전할까?"

- **SQL 로직:**

  - 위험 후보 = `discount ≥ 60% AND rating < 3.8`

  - 대체 후보 = 같은 카테고리, ±20% 가격대, 평점·리뷰 기준 충족

  - 대체 후보 중 PopQ 점수 상위 1개 매칭

```sql
-- 위험 상품(큰 할인+낮은 평점) → 같은 카테고리/근접 가격 대체재 추천
WITH leaf AS (
  SELECT *,
      -- leaf 카테고리 추출
      SPLIT(category, '|')[OFFSET(ARRAY_LENGTH(SPLIT(category, '|')) - 1)] AS
category_leaf
      FROM `cobalt-chalice-470802-i8.day4.amazon_clean`
),
risky AS (
  -- 위험 후보: 큰 할인 + 낮은 평점
  SELECT
    product_id AS risky_id, product_name AS risky_name,
    category_leaf, discounted_price AS risky_price,
    rating AS risky_rating, discount_percentage
  FROM leaf
  WHERE discount_percentage >= 0.6
  AND rating IS NOT NULL AND rating < 3.8
  AND discounted_price > 0
),
cands AS (
  -- 대체 후보: 같은 카테고리, 평점/리뷰 기준 충족
  SELECT
    product_id, product_name, category_leaf,
    discounted_price, rating, rating_count,
    rating * LOG10(1 + rating_count) AS popq
  FROM leaf
  WHERE rating >= 4.2 AND rating_count >= 20
    AND discounted_price > 0
),
match AS (
  -- 위험 상품별로 가장 PopQ 높은 대체재 1개 매칭
  SELECT
    r.risky_id, r.risky_name, r.risky_rating, r.discount_percentage, r.risky_price,
    c.product_id AS alt_id, c.product_name AS alt_name,
    c.discounted_price AS alt_price, c.rating AS alt_rating, c.rating_count AS alt_r
ating_count,
    ROW_NUMBER() OVER (PARTITION BY r.risky_id ORDER BY c.popq DESC) A
S rn
  FROM risky r
  JOIN cands c
```

```
    ON c.category_leaf = r.category_leaf
    AND ABS(c.discounted_price - r.risky_price) <= 0.2 * r.risky_price    -- ±20%
가격 범위
)
SELECT
  risky_id, risky_name, risky_rating, discount_percentage, risky_price,
  alt_id, alt_name, alt_rating, alt_rating_count, alt_price
FROM match
WHERE rn = 1
ORDER BY discount_percentage DESC, risky_rating
LIMIT 10;      -- 캡처용
```

결과 (샘플)

행	risky_id	risky_name	risky_rating	discount_percentage	risky_price	alt_id	alt_name	alt_rating	alt_rating_count	alt_price
1	B0BVKRP7B	SHREENOVA ID116 Plus Bluetooth Fitness Smart Watch for Men Women and Kids Activity Tracker (Black)	2.8	0.86	281.0	B0BMM7R92G	Noise_Colorfit Smart Watch Charger 2 Pin USB Fast Charger Magnetic Charging Cable Adapter (Smart Watch Charger 2 pin)	4.5	38	249.0
2	B09ND94ZRG	Boult Audio Airbass Propods X TWS Bluetooth Truly Wireless in Ear Earbuds with Mic, 32H Playtime, Fast Charging Type-C, Ipx5 Water Resistant, Touch Controls and Voice Assistant (Red)	3.5	0.82	1099.0	B01F262EUU	Samsung Original EHS64 Wired in Ear Earphones with Mic, Black	4.2	31539	949.0
3	B0B94JPY2N	Amazon Brand - Solimo 65W Fast Charging Braided Type C to C Data Cable | Suitable For All Supported Mobile Phones (1 Meter, Black)	3.0	0.8	199.0	B00NH11KIK	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
4	B0B53QFZPY	PTron Newly Launched Force X10 Bluetooth Calling Smartwatch with 1.7" Full Touch Color Display, Real Heart Rate Monitor, SpO2, Watch Faces, 5 Days Runtime, Fitness Trackers & IP68 Waterproof (Pink)	3.3	0.78	1299.0	B09YV42QHZ	Fire-Boltt Ninja 3 Smartwatch Full Touch 1.69" & 60 Sports Modes with IP68, Sp02 Tracking, Over 100 Cloud based watch faces (Green)	4.2	22638	1499.0
5	B0B53QLB9H	PTron Newly Launched Force X10 Bluetooth Calling Smartwatch with 1.7" Full Touch Color Display, Real Heart Rate Monitor, SpO2, Watch Faces, 5 Days Runtime, Fitness Trackers & IP68 Waterproof (Blue)	3.3	0.78	1299.0	B09YV42QHZ	Fire-Boltt Ninja 3 Smartwatch Full Touch 1.69" & 60 Sports Modes with IP68, Sp02 Tracking, Over 100 Cloud based watch faces (Green)	4.2	22638	1499.0
6	B07B275VN9	Airtel DigitalTV DTH Television, Setup Box Remote Compatible for SD and HD Recording (Black)	3.7	0.78	179.0	B09L835C3V	Smashtronics® - Case for Firetv Remote, Fire Stick Remote Cover Case, Silicone Cover for TV Firestick 4K/TV 2nd Gen(3rd Gen) Remote Control - Light Weight/Anti Slip/Shockproof (Black)	4.2	1335	199.0
7	B0B53NXFFR	PTron Newly Launched Force X10 Bluetooth Calling Smartwatch with 1.7" Full Touch Display, Real Heart Rate Monitor, SpO2, Watch Faces, 5 Days Runtime, Health/Fitness Trackers & IP68 Waterproof (Black)	3.3	0.77	1399.0	B09YV42QHZ	Fire-Boltt Ninja 3 Smartwatch Full Touch 1.69" & 60 Sports Modes with IP68, Sp02 Tracking, Over 100 Cloud based watch faces (Green)	4.2	22638	1499.0
8	B0B2931FCV	ZEBRONICS Zeb-Sound Bomb N1 True Wireless in Ear Earbuds with Mic ENC, Gaming Mode (up to 50ms), up to 18H Playback, BT V5.2, Fidget Case, Voice Assistant, Splash Proof, Type C (Midnight Black)	3.5	0.76	999.0	B01F262EUU	Samsung Original EHS64 Wired in Ear Earphones with Mic, Black	4.2	31539	949.0
9	B07V5YF4ND	LOHAYA LCD/LED Remote Compatible for Sony Bravia Smart LCD LED UHD OLED QLED 4K Ultra HD TV Remote Control with YouTube & Netflix Function [Compatible for Sony Tv Remote Control]	3.5	0.75	299.0	B09BW334ML	Dealfreez Case Compatible with Fire TV Stick 3rd Gen 2021 Full Wrap Silicone Remote Cover Anti-Lost with Loop (D-Black)	4.3	4145	349.0
10	B08GJNM9N7	LOHAYA Television Remote Compatible for VU LED LCD HD Tv Remote Control Model No :- EN2B27V	3.7	0.75	299.0	B09BW334ML	Dealfreez Case Compatible with Fire TV Stick 3rd Gen 2021 Full Wrap Silicone Remote Cover Anti-Lost with Loop (D-Black)	4.3	4145	349.0

결과 해석:

- 위험 상품(할인 70~80% 이상, 평점 3점대 초반)은 대부분 **저가 웨어러블·액세서리** 제품이 차지.

- 대체 추천으로는 같은 카테고리 내 **삼성 정품 이어폰, Fire-Boltt 스마트워치, AmazonBasics 케이블** 등 평점 4.2 이상 검증된 상품이 매칭됨.

- 결론: 사용자는 **낮은 품질 리스크를 피하면서도 합리적인 가격대의 대체재**를 선택 가능.