

Amazon 데이터셋 기반 추천 시스템 설계 프로젝트

🔔 서론: 온라인 쇼핑에서 고객은 보통 먼저 구매하고자 하는 카테고리를 탐색하고,

상품 정보를 비교하면서 평점·가격·할인율을 확인한 뒤, 마음에 드는 상품을 중심으로 유사 제품을 추가로 찾아보는 행동 패턴을 보인다. 이러한 실제 구매 흐름을 기반으로, Amazon 데이터셋을 활용하여 고객의 탐색 단계부터 개인화 단계까지 지원할 수 있는 3가지 추천 시스템을 설계하고 구현하고자 한다.

① 추천 시스템 설계 목표

단계	고객 행동 해석	추천 로직 정의
Step 1	카테고리 탐색	카테고리 상품 중 평점 + 가격할인율 (*해당 데이터셋은 리뷰 개수 확인 불가)
Step 2	특정 상품 집중 탐색	Co-Review 기반 함께 리뷰된 상품 추천
Step 3	반복 구매 및 취향 고도화	관심 카테고리 & 가격대별 상품 추천

컨셉안1) '이 카테고리에서 지금 가장 잘 팔리는 제품 TOP 10'

➡ 컨셉방향: 고객은 먼저 카테고리를 선택하고, 그 안에서 평점이 좋고(quality), 리뷰개수 및 내용 확인 후 가격확인. 해당 데이터 셋에서는 리뷰 개수 카운팅이 불가.

이에 관심 카테고리 안에서 1)평점이 4점 이상이고 2)할인률이 높은 순서대로 상품을 추천하고자 함.

*케이블을 구매하려는 고객을 가정하고 쿼리 구성

➡ 쿼리

```
WITH cleaned AS (
  SELECT
    product_id,
    product_name,
    category,
    -- rating: 문자열로 캐스팅 후 숫자/소수점만 남기고, 실수형으로 변환
    SAFE_CAST(
      REGEXP_REPLACE(CAST(rating AS STRING), r'^[0-9.]', '') AS FLOAT64
    ) AS rating_num,
    -- discount_percentage: 문자열로 캐스팅 후 숫자만 남기고, 실수형으로 변환
    SAFE_CAST(
      REGEXP_REPLACE(CAST(discount_percentage AS STRING), r'^[0-9.]', '') AS FLOAT64
    ) AS discount_pct
  FROM
    `my-project-2-480604.day4.amazon`
)
SELECT
  product_id,
  product_name,
  category,
  rating_num,
  discount_pct
FROM
  cleaned
WHERE
  rating_num IS NOT NULL      -- 숫자로 변환된 값만
  AND discount_pct IS NOT NULL
  AND rating_num >= 4.0        -- 평점 4점 이상
ORDER BY
  discount_pct DESC,           -- 할인율 높은 순
  rating_num DESC              -- 평점 높은 순 (보조 정렬)
LIMIT 10;
```

➡ 결과

SQL 핵심 로직 → `rating` 을 숫자로 변환하여 조건 필터링 및 정렬하여 아래와 같은 결과를 확인.

ranking	product_id	product_name	category	rating_num	discount_pct
1	B097C564GC	rts [2 Pack] Mini USB C Type C Adapter Plug, Type C Female to USB A Male Charger Charging Cable Adapter Converter compatible for iPhone, Samsung S20 ultra/S21/S10/S8/S9/MacBook Pro iPad Silver	Computers&Accessories Accessories&Peripherals Adapters USBtoUSBAdapters	4.3	0.94
2	B0BF4YBLPX	Fire-Bolt Ninja Call Pro Plus 1.83" Smart Watch with	Electronics WearableTechnology SmartWatches	4.2	0.91

		Bluetooth Calling, AI Voice Assistance, 100 Sports Modes IP67 Rating, 240*280 Pixel High Resolution			
3	B0BF57RN3K	Fire-Bolt Ninja Call Pro Plus 1.83" Smart Watch with Bluetooth Calling, AI Voice Assistance, 100 Sports Modes IP67 Rating, 240*280 Pixel High Resolution	Electronics WearableTechnology SmartWatches	4.2	0.91
4	B0BF563HB4	Fire-Bolt Ninja Call Pro Plus 1.83" Smart Watch with Bluetooth Calling, AI Voice Assistance, 100 Sports Modes IP67 Rating, 240*280 Pixel High Resolution	Electronics WearableTechnology SmartWatches	4.2	0.91
5	B0BF54LXW6	Fire-Bolt Ninja Call Pro Plus 1.83" Smart Watch with Bluetooth Calling, AI Voice Assistance, 100 Sports Modes IP67 Rating, 240*280 Pixel High Resolution	Electronics WearableTechnology SmartWatches	4.2	0.91
6	B0BF54972T	Fire-Bolt Ninja Call Pro Plus 1.83" Smart Watch with Bluetooth Calling, AI Voice Assistance, 100 Sports Modes IP67 Rating, 240*280 Pixel High Resolution	Electronics WearableTechnology SmartWatches	4.2	0.91
7	B0BBBW521YC	LAPSTER 12pcs Spiral Cable Protectors for Charger, Wires, Data Charger Cable Protector for Computers, Cell Phones etc.(Grey)	Electronics Mobiles&Accessories MobileAccessories Décor PhoneCharms	4.4	0.9
8	B08W56G1K9	LAPSTER Spiral Charger Spiral Charger Cable Protectors for Wires Data Cable Saver Charging Cord Protective Cable Cover Set of 3 (12 Pieces)	Computers&Accessories Accessories&Peripherals Cables&Accessories CableConnectionProtectors	4.1	0.9
9	B09SJ1FTYV	Sounce Protective Case Cover Compatible Boat Xtend Overall Protective Case TPU HD Clear Ultra-Thin Cover with Unbreakable Screen Guard	Electronics Mobiles&Accessories MobileAccessories Cases&Covers BasicCases	4	0.9
10	B085HY1DGR	Sounce Spiral Charger Cable Protector Data Cable Saver Charging Cord Protective Cable Cover Headphone MacBook Laptop Earphone Cell Phone Set of 3 (Cable Protector (12 Units))	Computers&Accessories Accessories&Peripherals Cables&Accessories CableConnectionProtectors	4	0.9

컨센안2) '이 상품 본 사람들은 함께 본 다른 상품'

➡ 컨센방향: 이 시스템은 특정 상품과 함께 많이 본 다른 상품을 추천하고자 한다. 즉 추천 기준 상품 ID를 직접 입력하여 해당 상품과 함께 본 비율이 높은 상품을 추천하는 방식이며 아래와 같이 구성하고자 한다.

▣ 쿼리

기준상품을 케이블 제품 중 '[B07JW9H4J1](#)'으로 지정하여 다음과 같은 쿼리 구성

```
-- [추천 시스템 2] 콘텐츠 기반 유사 상품 추천
-- 기준 상품: B07JW9H4J1

-- 1. rating, discounted_price 숫자 변환 처리
WITH cleaned AS (
SELECT
product_id,
product_name,
category,
SAFE_CAST(
    REGEXP_REPLACE(CAST(rating AS STRING), r'^[0-9.]', '') AS FLOAT64
) AS rating_num,
SAFE_CAST(
    REGEXP_REPLACE(CAST(discounted_price AS STRING), r'^[0-9.]', '') AS FLOAT64
) AS discounted_price_num
FROM
`my-project-2-480604.day4.amazon`
),

-- 2. 기준 상품 정보 추출
target AS (
```

```

SELECT
    product_id,
    category AS target_category,
    rating_num AS target_rating,
    discounted_price_num AS target_price
FROM
    cleaned
WHERE
    product_id = 'B07JW9H4J1'
),
-- 3. 동일 카테고리 & 평점 유사 & 가격대 유사 상품 탐색
similar AS (
SELECT
    c.product_id,
    c.product_name,
    c.category,
    c.rating_num,
    c.discounted_price_num,
    t.target_rating,
    t.target_price
FROM
    cleaned c
CROSS JOIN
    target t
WHERE
    c.product_id != t.product_id          -- 자기 자신 제외
    AND c.category = t.target_category    -- 같은 카테고리
    AND c.rating_num IS NOT NULL
    AND c.discounted_price_num IS NOT NULL
    AND ABS(c.rating_num - t.target_rating) <= 0.5      -- 평점 차이 허용 범위
    AND c.discounted_price_num BETWEEN t.target_price * 0.5 -- 가격대 유사
                                    AND t.target_price * 1.5
)
-- 4. 최종 추천 Top20
SELECT
    product_id,
    product_name,
    category,
    rating_num,
    discounted_price_num AS discounted_price
FROM
    similar
ORDER BY
    rating_num DESC,
    discounted_price DESC
LIMIT 20;

```

▶ 결과

기준 상품USB-C to Lightning Cable인 [B07JW9H4J1](#) 쿼리 실행 결과, 해당 상품과 동일한 카테고리인

Computers&Accessories → Accessories&Peripherals → Cables 영역에서 평점이 높은 유사 상품들이 추천되었다.

number	product_id	product_name	category	rating_num	discounted_price
1	B0B86CDHL1	oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22 S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	349
2	B0B86CDHL1	oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	349

		Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22 S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G			
3	B0B86CDHL1	oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22 S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	349
4	B0B86CDHL1	oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22 S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	349
5	B0B86CDHL1	oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22 S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	349
6	B0B86CDHL1	oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	349

		S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G			
7	B00NH13Q8W	AmazonBasics USB 2.0 Extension Cable for Personal Computer, Printer, 2-Pack - A-Male to A- Female - 3.3 Feet (1 Meter, Black)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	299
8	B00NH13Q8W	AmazonBasics USB 2.0 Extension Cable for Personal Computer, Printer, 2-Pack - A-Male to A- Female - 3.3 Feet (1 Meter, Black)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	299
9	B00NH13Q8W	AmazonBasics USB 2.0 Extension Cable for Personal Computer, Printer, 2-Pack - A-Male to A- Female - 3.3 Feet (1 Meter, Black)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	299
10	B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B- Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	209
11	B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B- Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	209
12	B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B- Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	209
13	B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B- Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	209
14	B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B- Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	209
15	B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B- Male - for	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5	209

		Personal Computer, Printer- 6 Feet (1.8 Meters), Black				
16	B083342NKJ	MI Braided USB Type-C Cable for Charging Adapter (Red)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.4	349	
17	B083342NKJ	MI Braided USB Type-C Cable for Charging Adapter (Red)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.4	349	
18	B083342NKJ	MI Braided USB Type-C Cable for Charging Adapter (Red)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.4	349	
19	B083342NKJ	MI Braided USB Type-C Cable for Charging Adapter (Red)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.4	349	
20	B083342NKJ	MI Braided USB Type-C Cable for Charging Adapter (Red)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.4	349	

컨셉안3) '나의 취향 저격 스타일'

▶ 컨셉방향:

사용자가 실제로 구매하고 리뷰를 남긴 이력을 바탕으로, 그 사용자의 '관심 카테고리'를 정의, 그리고 그 관심 분야 내에서 가장 품질이 겹중된(평점 4.0점 이상) 상품만을 선별하여 제품을 추천하고자 함.

- **사용자 취향 정의:** 타겟 사용자가 리뷰를 남긴 상품들의 `category` 필드를 분석하여, 가장 빈번하게 등장하는 카테고리 그룹(레벨 2)을 상위 3개 선정, 이것이 사용자의 '취향'으로 정의
- **추천 후보군 필터링:** 전체 상품 중에서 이 3개의 취향 카테고리에 속하는 상품만 골라냄.
- **최종 순위 결정:** 후보군 상품들 중 평균 평점 4.0 이상을 만족하는 상품선정 후 평점 순(내림차순)으로 정렬하여 추천.

▶ 쿼리

타겟 사용자 ID: [AECPPFYFQVRUWC3KGNLJIREFP5LQ](#)로 지정하여 분석

```
-- 1. rating 숫자 변환 + 기본 정보 정리
WITH cleaned AS (
SELECT
product_id,
product_name,
category,
user_id,
SAFE_CAST(
    REGEXP_REPLACE(CAST(rating AS STRING), r'^0-9.', '') AS FLOAT64
) AS rating_num
FROM
`my-project-2-480604.day4.amazon`
),

-- 2. (user_id, product_id) 관계 복원
exploded AS (
SELECT
product_id,
product_name,
category,
rating_num,
TRIM(u) AS user_id
FROM
cleaned,
UNNEST(SPLIT(CAST(user_id AS STRING), ',')) AS u
WHERE
TRIM(u) IS NOT NULL
AND TRIM(u) != ""
),

-- 3. 해당 사용자가 가장 많이 본(상호작용한) 카테고리 찾기
user_top_category AS (
SELECT
category,
COUNT(*) AS interaction_count
FROM
exploded
WHERE
user_id = 'AECPPFYFQVRUWC3KGNLJIREFP5LQ'
GROUP BY
category
ORDER BY
```

```

interaction_count DESC
LIMIT 1
)

-- 4. 그 카테고리 안에서 고평점 상품 추천(4점 이상)
SELECT
c.product_id,
c.product_name,
c.category,
c.rating_num
FROM
cleaned c
JOIN
user_top_category utc
ON
c.category = utc.category
WHERE
c.rating_num IS NOT NULL
AND c.rating_num >= 4.0      -- 고평점 기준
ORDER BY
c.rating_num DESC
LIMIT 10;

```

추천 시스템은 특정 사용자([AECPPFYFQVRUWC3KGNLJIOREFP5LQ](#))가 어떤 카테고리에서 가장 많은 상호작용을 했는지를 기반으로 취향 카테고리를 정의하고, 그 안에서 고평점 상품을 추천하도록 설계.

먼저, `UNNEST(SPLIT(user_id, ','))`를 사용, 한 행에 문자열로 묶어 있던 여러 사용자 ID를 분리해(`user_id, product_id`) 관계를 복원. 이후, 해당 사용자가 가장 많이 상호작용한 카테고리를 집계하여 개인의 주요 관심 카테고리를 추출. 마지막으로 이 카테고리 내에서 평점이 4.0 이상인 상품만 필터링하고, 평점이 높은 순으로 상위 10개 상품을 추천함으로써 개인화 시스템을 구성했다.

➡ 결과

product_id	product_name	category	rating_num
B0BP7XLX48	Syncwire LTG to USB Cable for Fast Charging Compatible with Phone 5/5C/5S/ 6/ 6S/ 7/8/ X/XR/XS Max/ 11/12/ 13 Series and Pad Air/Mini, Pod & Other Devices (1.1 Meter, White)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	5
B0BQRJ3C47	REDTECH USB-C to Lightning Cable 3.3FT, [Apple MFi Certified] Lightning to Type C Fast Charging Cord Compatible with iPhone 14/13/13 pro/Max/12/11/X/XS/XR/8, Supports Power Delivery - White	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	5
B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B00NH11PEY	AmazonBasics USB 2.0 - A-Male to A-Female Extension Cable for Personal Computer, Printer (Black, 9.8 Feet/3 Meters)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B084N18QZY	Belkin USB C to USB-C Fast Charging Type C Cable, 60W PD, 3.3 feet (1 meter) for Laptop, Personal Computer, Tablet, Smartphone - Black, USB-IF Certified	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B00NH11KIK	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B09C6HXFC1	Duracell USB Lightning Apple Certified (Mfi) Braided Sync & Charge Cable For Iphone, Ipad And Ipod. Fast Charging Lightning Cable, 3.9 Feet (1.2M) - Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B09C6HXFC1	Duracell USB Lightning Apple Certified (Mfi) Braided Sync & Charge Cable For Iphone, Ipad And Ipod. Fast Charging Lightning Cable, 3.9 Feet (1.2M) - Black	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B084MZBYTV	Belkin USB C to USB-C Fast Charging Type C Cable, 60W PD, 3.3 feet (1 meter) for Laptop, Personal Computer, Tablet, Smartphone - White, USB-IF Certified	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5
B00NH13Q8W	AmazonBasics USB 2.0 Extension Cable for Personal Computer, Printer, 2-Pack - A-Male to A-Female - 3.3 Feet (1 Meter, Black)	Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	4.5

⇒ 위의 시스템을 활용하여 해당 사용자([AECPPFYFQVRUWC3KGNLJIOREFP5LQ](#))의 관심카테고리 및 추천제품을 분석한 결과 아래와 같다.

| 관심 카테고리는 ⇒ Computers&Accessories | Accessories&Peripherals | Cables&Accessories | Cables | USBCables 등이다

즉 이 사용자는 컴퓨터 주변기기, USB 케이블류에 강한 관심을 보이고 있는 것을 확인했다.

🔔 결론 및 회고

본 프로젝트의 추천 시스템은 고객의 구매 여정을 단계별로 지원할 수 있도록, 평점·할인율·카테고리·사용자 취향 정보를 조합해 실제 서비스에서 활용 가능한 형태로 구성하는 것이었다.

다만, 데이터셋이 상품 요약 중심으로 이루어져 있어 리뷰 개수나 상세 사용자 로그를 기반으로 한 정교한 추천 알고리즘 적용에는 한계가 있었으며, 이에 따라 데이터 특성에 맞는 로직으로 조정하는 과정이 필요했다.

이번 프로젝트를 진행하며 쿠팡·W컨셉·오플라인 등 평소에 자주 활용하는 서비스들의 추천 노출 방식과

사용자 행동 데이터를 다시 생각해보는 계기가 되었다. 특히 카테고리 구조, 리뷰 기반 신뢰도, 사용자 행동 로그가 정확한 추천 시스템 구축의 핵심이라는 점을 직접 체감할 수 있는 유용한 경험이었다.