2025

Amazon Dataset-Based Recommender System Design Project

Project Overview

This project **involves designing and implementing a personalized recommendation system based on** Amazon datasets. Amazon is one of the largest e-commerce platforms globally, with vast customer reviews and product data. In this project, the goal is to analyze this data using SQL alone and design the basis for a recommendation system that can improve the customer's buying experience.

INDEPTH ANALYSIS OF AMAZON DATA
JI HYUN MOON

1. **Over view Data**
2. **Refine and clean raw data**
3. **Easy to enhance level of EDA plans**
4. **In-depth EDA and insights**
5. **Conclusion**

추천 시스템 설정	**추천 시스템의 테마 설정**
리뷰를 많이 남긴 상품들이에요	많은 리뷰 수 + 높은 평점을 반영하여 신뢰할 수 있는 제품 추천
최대 할인율 기반 추천 시스템	할인율을 기준으로 추천 점수에 반영하여 가성비 좋은 제품 추천
주요 타깃 추천 시스템	키워드 기반의 맞춤형 제품 추천
특정 타깃 / 아마존에 특화 카테고리 내에서	중심 트랜드 키워드를 기반으로 사용자 대상 추천
정제된 중복 제거 추천 시스템	추천 제품 중복 제거 후 상위 제품만 정제하여 사용자에게 노출

Tier	Recommendation Type	Description	Method
🟢 **Easy**	Top Rated + High Reviewed	Find products with rating > 4.3, high rating_count	Simple SQL / Pandas
🟢 **Easy**	Top Discounts	Products with discount > 50%	Filter and sort
🟢 **Easy**	Top Products per Category	groupby category → max rating/sales	Aggregate
🟡 **Medium**	Price Buckets	Group into Low, Mid, High for targeting	Conditional binning
🟡 **Medium**	Trending Subcategories	Subcategory frequency over time	Count + plot
🟡 **Medium**	Gender/Age Segments	Keyword search: "women", "kids", etc.	NLP string match
🟠 **Hard**	Feature Similar Products	Based on product name/text similarity	TF-IDF + Cosine
🟠 **Hard**	Clustering by Price-Rating	KMeans or scatter insights	Scikit-learn
🔴 **Advanced**	Reorder / Basket Analysis	Users who reviewed X also reviewed Y	Need session data
🔴 **Advanced**	Review Sentiment/NLP	Extract topics, sentiment	TextBlob, LDA

From easy to difficult planner

1. Data review : '재료' 이해하기 (데이터셋 살펴보기)

Column	Description	Notes
product_id, product_name	Product identity	✅ Use for grouping/sales
category	Hierarchical category path	🧠 Needs parsing (split by `

discounted_price, actual_price, discount_percentage	Price features	🔄 Must convert to numeric (remove ₹ and %)
rating, rating_count	Review metrics	🔁 Clean commas and cast to float/int
about_product	Product description	🧠 NLP opportunities
user_id, user_name, review_id	Identity fields	Optional, but can be used for behavior analysis
review_title, review_content	Customer feedback	🧠 Sentiment, keyword clustering
img_link, product_link	URLs	Not relevant for EDA

Task	Details
Price, Discount → numeric	Strip ₹ and % symbols, cast to float
Rating & Count → numeric	Convert rating and remove comma from rating_count
Category Levels	Split category column by `
Optional	Lowercase, strip spaces in product names/descriptions

Logic	How to Implement
"Similar products"	Text vector similarity on product_name or about_product
"In your price range"	Filter by price bucket
"Highly rated alternatives"	Combine rating + price in filter
"Based on what others said…"	NLP topic modeling on review_content

데이터 살펴보고 정제하기

Category value 정제

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	product_id	product_name	category	category_value
1	B0BPJBTB3F	Khaitan ORFin Fan heater for Ho...	Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters	Home&Kitchen
2	B0BPJBTB3F	Khaitan ORFin Fan heater for Ho...	Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters	Heating,Cooling&AirQuality
3	B0BPJBTB3F	Khaitan ORFin Fan heater for Ho...	Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters	RoomHeaters
4	B0BPJBTB3F	Khaitan ORFin Fan heater for Ho...	Home&Kitchen|Heating,Coolin	FanHeaters

Price /Review/Rate data clean

```
SELECT
  product_id,
  product_name,

  -- 가격 문자열 처리 후 float 변환
  SAFE_CAST(REPLACE(REPLACE(CAST(discounted_price AS STRING), '₹', ''), ',', '') AS FLOAT64) AS discounted_price_clean,
  SAFE_CAST(REPLACE(REPLACE(CAST(actual_price AS STRING), '₹', ''), ',', '') AS FLOAT64) AS actual_price_clean,

  -- 할인율 처리 (이미 float 형임을 확인했지만, 이 부분을 수행하니 나중에 오류발생, 즉 SAFE_CAST(STRING)에서만 작용함.
  -- SAFE_CAST(REPLACE(CAST(discount_percentage AS STRING), '%', '') AS FLOAT64) AS discount_percentage_clean,

  -- 평점
```

```
    SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating_clean,

    -- 리뷰 수 처리
    SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',', '') AS INT64) AS rating_count_clean

  FROM
    `shining-bearing-465502-n1.Mission.Amazon`
  LIMIT 10;

```

Job information | **Results** | Chart | JSON | Execution details | Execution graph

Row	product_id	product_name	discounted_price...	actual_price_clean	discount_percent...	rating_clean	rating_count_clean
1	B0BPJBTB3F	Khaitan ORFin Fan heater for Ho...	1299.0	2495.0	0.48	2.0	2
2	B0BFBNXS94	Personal Size Blender, Portable Blender, Battery Powered USB Blender, with Four Blades, Mini Blender Travel Bottle for Juice, Shakes,	669.0	1499.0	0.55	2.3	13
3	B0B7L86YCB	Green Tales Heat Seal Mini Food Sealer-Impulse Machine for Sealing Plastic Bags Packaging	161.0	300.0	0.46	2.6	24

Data review: 최하, 최저, 최고, 최다 값 확인

After read data types, characteristics, let's find max, min value from key-point of features for sales.

```
WITH cleaned_data AS (
SELECT
  product_id,
  product_name,

  -- 가격 처리
  SAFE_CAST(REPLACE(REPLACE(CAST(discounted_price AS STRING), '₹', ''), ',', '') AS FLOAT64) AS discounted_price,
  SAFE_CAST(REPLACE(REPLACE(CAST(actual_price AS STRING), '₹', ''), ',', '') AS FLOAT64) AS actual_price,

~~  -- 할인율 처리~~
~~  SAFE_CAST(REPLACE(CAST(discount_percentage AS STRING), '%', '') AS FLOAT64) AS discount_percentage,~~

  -- 평점
  SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,

  -- 리뷰 수
```

```
    SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',', '') AS INT64) AS rating_count,

    -- 카테고리 분할
    SPLIT(category, '|')[OFFSET(0)] AS category_main,
    SPLIT(category, '|')[OFFSET(1)] AS category_sub,
    SPLIT(category, '|')[OFFSET(2)] AS category_micro


FROM
  `shining-bearing-465502-n1.Mission.Amazon`
)
-- sub query 로 저장한 후, 데이터 확인
SELECT
  MAX(discount_percentage) AS max_discount,
  MIN(discount_percentage) AS min_discount,
  MAX(rating) AS max_rating,
  MIN(rating) AS min_rating,
  MAX(rating_count) AS max_reviews,
  MIN(rating_count) AS min_reviews
FROM cleaned_data
```

Job information | **Results** | Chart | JSON | Execution details | Execution graph

Row	max_discount	min_discount	max_rating	min_rating	max_reviews	min_reviews
1	0.94	0.0	5.0	2.0	426973	2

QUANTILES : 고득점 리뷰와 평가수를 어떤 기준으로 마련할까?

분위수기준득점 고려

rating_quartiles ..... review_count_quartiles

"[2.0,4.0,4.1,4.3,5.0]"..... "[2,1191,5057,17413,426973]"

```
-- 고평점, 다수 리뷰 상품 찾기
--  기준에 대한 근거 확립 : 분위수 기준 확인
SELECT
  APPROX_QUANTILES(SAFE_CAST(rating AS FLOAT64),4) AS rating_quartiles,
  APPROX_QUANTILES(SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',','') AS INT64), 4) AS
review_count_quartiles
FROM `shining-bearing-465502-n1.Mission.Amazon`
```

uery completed

uery results Save results Open in

o information | Results | Chart | JSON | Execution details | Execution graph

	rating_quartiles	review_count_quartiles
1	2.0	2
	4.0	1191
	4.1	5057
	4.3	17413
	5.0	426973

리뷰 수	실무적 해석
1~100	리뷰 조작 가능성 높음 (신제품, 광고성)
500~1000	일부 사용자에게만 노출
5000 이상	**대중적 확산 + 리뷰 통계적 신뢰 확보**
10,000 이상	베스트셀러 수준

고득점 평가과 최대 리뷰 상품 찾기.

```
-- 고평점+다수 리뷰 상품 찾기.
WITH cleaned_data AS (
  SELECT
    product_id,
    product_name,
    SAFE_CAST(REPLACE(REPLACE(CAST(discounted_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS discounted_price,
    SAFE_CAST(REPLACE(REPLACE(CAST(actual_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS actual_price,
    SAFE_CAST(REPLACE(CAST(discount_percentage AS STRING), '%', '') AS FLOAT64) AS
discount_percentage,
    SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,
```

```
    SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',', '') AS INT64) AS
rating_count,
    SPLIT(category, '|')[OFFSET(0)] AS category_main
  FROM `shining-bearing-465502-n1.Mission.Amazon`
)

SELECT
  product_id,
  product_name,
  rating,
  rating_count,
  discount_percentage,
  discounted_price,
  actual_price,
  category_main
FROM cleaned_data
WHERE
  rating >= 4.3
  AND rating_count >= 5000
ORDER BY rating DESC, rating_count DESC
LIMIT 30;
```

: Rate 가장 높은 상품은
'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 / Home*Kitchen

최다리뷰 상품은: SanDisk Extreme SD UHS I 64GB Card for 4K Video for DSLR and Mirrorless Cameras 170MB/s Read & 80MB/s Write: 205052 / 리뷰 : 4.5 /Electronics.

product_id	product_name	rating	rating_count	discount_percent	discounted_price	actual_price	category_main
B0BR4F878Q	Swiffer Instant El	4.8	53803	0.28	1439	1999	Home&Kitchen
B0B23LW7NV	Spigen EZ Fit Te	4.7	7779	0.66	999	2899	Electronics
B09WN3SRC7	Sony Bravia 164	4.7	5935	0.44	77990	139900	Electronics
B01J1CFO5I	Redgear MP35 S	4.6	33434	0.46	299	550	Computers&Accessories
B095RTJH1M	Spigen EZ Fit Te	4.6	26603	0.66	999	2899	Electronics
B01MQ2A86A	Logitech M331 S	4.6	12375	0.21	1295	1645	Computers&Accessories
B00NFD0ETQ	Logitech G402 H	4.6	10760	0.31	1995	2895	Computers&Accessories
B07X2L5Z8C	Logitech Pebble	4.6	10652	0.35	1490	2295	Computers&Accessories
B00K57MR22	Sujata Dynamix I	4.6	6550	0.28	6120	8478	Home&Kitchen
B0B244R4KB	Spigen EZ Fit Te	4.6	6129	0.66	999	2899	Electronics
B09X7DY7Q4	SanDisk Extreme	4.5	205052	0.48	939	1800	Electronics
B00NH11KIK	AmazonBasics U	4.5	107687	0.7	209	695	Computers&Accessories
B00NH11KIK	AmazonBasics U	4.5	107686	0.7	209	695	Computers&Accessories
B07G3YNLJB	Crucial BX500 2	4.5	92925	0.41	1815	3100	Computers&Accessories
B00NH13Q8W	AmazonBasics U	4.5	74977	0.63	299	800	Computers&Accessories
B00NH11PEY	AmazonBasics U	4.5	74976	0.73	199	750	Computers&Accessories
B00NH11PEY	AmazonBasics U	4.5	74976	0.73	199	750	Computers&Accessories
B004IO5BMQ	Logitech M235 W	4.5	54405	0.3	699	995	Computers&Accessories
B08ZJDWTJ1	Seagate Expansi	4.5	50810	0.18	4098	4999	Computers&Accessories
B094QZLJQ6	Seagate One Tou	4.5	50273	0.28	5799	7999	Computers&Accessories
B01HJI0FS2	Dell MS116 1000	4.5	33176	0.54	299	650	Computers&Accessories
B088ZFJY82	Elv Aluminium Ac	4.5	28978	0.79	314	1499	Electronics
B07Q4QV1DL	ELV Aluminum A	4.5	28978	0.82	269	1499	Electronics
B014SZO90Y	Duracell Ultra Alk	4.5	28030	0.16	266	315	Electronics
B08C4Z69LN	Crucial RAM 8GE	4.5	26194	0.49	1792	3500	Computers&Accessories
B083GKDRKR	Havells Aqua Plu	4.5	23484	0.46	1625	2995	Home&Kitchen
B00V4BGDKU	TP-Link UE300 U	4.5	22420	0.42	1099	1899	Computers&Accessories
B08FYB5HHK	TP-Link UE300C	4.5	22420	0.4	1199	1999	Computers&Accessories
B00NNQMYNE	AirCase Rugged	4.5	21010	0.4	299	499	Computers&Accessories
B0083T231O	Belkin Essential	4.5	20668	0.14	1289	1499	Electronics

```sql
-- 최고 평점 상품 TOP10
WITH cleaned_data AS (
  SELECT
    product_id,
    product_name,
    SAFE_CAST(REPLACE(REPLACE(CAST(discounted_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS discounted_price,
    SAFE_CAST(REPLACE(REPLACE(CAST(actual_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS actual_price,
    SAFE_CAST(REPLACE(CAST(discount_percentage AS STRING), '%', '') AS FLOAT64) AS
discount_percentage,
    SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,
    SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',', '') AS INT64) AS
rating_count,
    SPLIT(category, '|')[OFFSET(0)] AS category_main
  FROM `shining-bearing-465502-n1.Mission.Amazon`
)

SELECT
  product_id,
  product_name,
  rating,
  rating_count,
  category_main
```

```
FROM cleaned_data
WHERE rating IS NOT NULL
ORDER BY rating DESC, rating_count DESC
LIMIT 10;
```

product_ic	product_name	rating	rating_count	category_main
B09ZHCJDP1	Amazon Basics Wireless Mouse | 2.4 GHz Connection, 1600 DPI | Type - C Adapter | Upto 12 Months of Battery Life | Ambidext	5	23	Computers&Accessories
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 a	5	5	Computers&Accessories
B0BQRJ3C47	REDTECH USB-C to Lightning Cable 3.3FT, [Apple MFi Certified] Lightning to Type C Fast Charging Cord Compatible with iPhon	5		Computers&Accessories
B0BR4F878Q	Swiffer Instant Electric Water Heater Faucet Tap Home-Kitchen Instantaneous Water Heater Tank less for Tap, LED Electric He	4.8	53803	Home&Kitchen
B0B53DS4TF	Instant Pot Air Fryer, Vortex 2QT, Touch Control Panel, 360° EvenCrisp™ Technology, Uses 95 % less Oil, 4-in-1 Appliance: Air Fr	4.8	3964	Home&Kitchen
B0BQ3K23Y1	Oratech Coffee Frother electric, milk frother electric, coffee beater, cappuccino maker, Coffee Foamer, Mocktail Mixer, Coffee F	4.8	28	Home&Kitchen
B0B23LW7NV	Spigen EZ Fit Tempered Glass Screen Protector for iPhone 14 Pro Max - 2 Pack (Sensor Protection)	4.7	7779	Electronics
B09WN3SRC7	Sony Bravia 164 cm (65 inches) 4K Ultra HD Smart LED Google TV KD-65X74K (Black)	4.7	5935	Electronics
B0BP89YBC1	Campfire Spring Chef Prolix Instant Portable Water Heater Geyser 1Ltr. for Use Home Stainless Steel Baking Rack | Restaurant	4.7	2591	Home&Kitchen
B0BM9H2NY9	Multifunctional 2 in 1 Electric Egg Boiling Steamer Egg Frying Pan Egg Boiler Electric Automatic Off with Egg Boiler Machine N	4.7	2300	Home&Kitchen

*average rate, average rate_count 의미있는 분석이 아님.

그렇다면 평점/리뷰 top 10 카테고리를 확인해보자. 너무 적은 평균 리뷰 수에 고득점 상품군이 존재.

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	category_main	product_count	avg_rating	avg_review_count
1	MusicalInstruments	2	3.9	44441.0
2	Electronics	526	4.08	29998.0
3	Computers&Accessories	451	4.16	17137.0
4	Toys&Games	1	4.3	15867.0
5	Home&Kitchen	447	4.04	6689.0
6	OfficeProducts	31	4.31	4828.0
7	HomeImprovement	2	4.25	4283.0
8	Health&PersonalCare	1	4.0	3663.0

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	product_id	product_name	rating	rating_count	category_main
1	B09ZHCJDP1	Amazon Basics Wireless Mouse | 2.4 GHz Connection, 1600 DPI | Type - C Adapter | Upto 12 Months of Battery Life | Ambidextrous Design |	5.0	23	Computers&Accessories
2	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 &	5.0	5	Computers&Accessories
3	B0BQRJ3C47	REDTECH USB-C to Lightning Cable 3.3FT, [Apple MFi Certified] Lightning to Type C Fast Charging Cord Compatible with iPhone	5.0	null	Computers&Accessories
4	B0BR4F878Q	Swiffer Instant Electric Water Heater Faucet Tap Home-Kitchen Instantaneous Water Heater Tank less for Tap, LED Electric Head Water	4.8	53803	Home&Kitchen
5	B0B53DS4TF	Instant Pot Air Fryer, Vortex 2QT, Touch Control Panel, 360° EvenCrisp™ Technology, Uses	4.8	3964	Home&Kitchen

```
--  카테고리별 평균 평점과 리뷰 수 분석
WITH cleaned_data AS (
  SELECT
    product_id,
    product_name,
    SAFE_CAST(REPLACE(REPLACE(CAST(discounted_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS discounted_price,
    SAFE_CAST(REPLACE(REPLACE(CAST(actual_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS actual_price,
    SAFE_CAST(REPLACE(CAST(discount_percentage AS STRING), '%', '') AS FLOAT64) AS
discount_percentage,
    SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,
```

```
    SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',', '') AS INT64) AS
rating_count,
    SPLIT(category, '|')[OFFSET(0)] AS category_main
  FROM `shining-bearing-465502-n1.Mission.Amazon`
)

SELECT
  category_main,
  COUNT(*) AS product_count,
  ROUND(AVG(rating), 2) AS avg_rating,
  ROUND(AVG(rating_count), 0) AS avg_review_count
FROM cleaned_data
WHERE rating IS NOT NULL AND rating_count IS NOT NULL
GROUP BY category_main
ORDER BY avg_rating DESC, avg_review_count DESC
```

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	category_main	product_count	avg_rating	avg_review_count
1	OfficeProducts	31	4.31	4828.0
2	Toys&Games	1	4.3	15867.0
3	HomeImprovement	2	4.25	4283.0
4	Computers&Accessories	451	4.16	17137.0
5	Electronics	526	4.08	29998.0
6	Home&Kitchen	447	4.04	6689.0
7	Health&PersonalCare	1	4.0	3663.0
8	MusicalInstruments	2	3.9	44441.0
9	Car&Motorbike	1	3.8	1118.0

아래 질문 사항은 GPT 를 통해서 확인된 부분. "그냥 평점 최고점을 구할때 VS 내가 세운 척도기준으로 적용할 경우 "

질문	정리된 답변
왜 평점 5.0 이 나오는가?	리뷰 수가 적은 제품에서 평점 5.0 은 흔함
과거에는 4.8 이 최고였는데 왜 차이가?	과거 기준은 **"고평점 + 다수 리뷰"라는 필터**를 사용했을 가능성 높음

질문	정리된 답변
해결 방법	rating_count >= 1000 이상 필터와 함께 평점 비교 진행해야 신뢰 확보

5 점 고득+리뷰수도 높은 상품 (>1000 rating_count) 찾아보기 : 0 / 너무 높은 준거기준이 됨. 조정이 필요.

There is no data to display.

--극히 소수의 제품에 대한 리뷰가 존재하는 신제품이나 존재할 수 있음.
--실제로 믿을 만한 제품은 4.5~4.8 대에 집중됨.

카테고리별 평균 평점과 리뷰 수 분석

목표: 어떤 카테고리(category_main)에 좋은 평점과 높은 리뷰 수 제품이 몰려 있는지 파악 → 추천 제품군 정의

: Musicalinstruments 가 review count 평균이 높았지만 product count 상품 수2로 신빙성 있는 결과가 아니다. 이는 새로나온 제품이거나 제품 종류가 적은것에 리뷰가 몰린것으로 판단이 된다.

제품수의 수도 많고, 리뷰 평점도 좋고, 리뷰 수도 좋은 카테고리는 Electronics군으로 보인다.

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	category_main	product_count	avg_rating	avg_review_count
1	MusicalInstruments	2	3.9	44441.0
2	Electronics	526	4.08	29998.0
3	Computers&Accessories	451	4.16	17137.0
4	Toys&Games	1	4.3	15867.0
5	Home&Kitchen	447	4.04	6689.0
6	OfficeProducts	31	4.31	4828.0
7	HomeImprovement	2	4.25	4283.0
8	Health&PersonalCare	1	4.0	3663.0
9	Car&Motorbike	1	3.8	1118.0

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	category_main	product_count	avg_rating	avg_review_count
1	OfficeProducts	31	4.31	4828.0
2	Toys&Games	1	4.3	15867.0
3	HomeImprovement	2	4.25	4283.0
4	Computers&Accessories	451	4.16	17137.0
5	Electronics	526	4.08	29998.0
6	Home&Kitchen	447	4.04	6689.0
7	Health&PersonalCare	1	4.0	3663.0
8	MusicalInstruments	2	3.9	44441.0
9	Car&Motorbike	1	3.8	1118.0

할인율 + 평점 조합 분석 (가성비 제품 찾기) :

목표: 높은 할인율과 높은 평점을 동시에 만족하는 "가성비 상품"을 찾아 추천

데이터 기반 준거: 할인율 분위수 확인

우리는 할인율(discount_percentage)의 분위수를 확인하여 "상대적 기준"을 설정

```
-- 데이터 기반 준거/할인율 분위수 확인
-- discount_percentage의 분위수를 확인하여 상대적 기준을 설정해본다.
select
approx_quantiles(
  safe_cast(replace(cast(discount_percentage AS STRING), '%', '') AS FLOAT64),4
) AS discount_quartiles
FROM `shining-bearing-465502-n1.Mission.Amazon`
```

Query completed

Query results | Save results | Open in

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	discount_quartiles
1	0.0
	0.32
	0.5
	0.63
	0.94

50% discount_quartile 확인

지금까지 신뢰할 수 있는 분위수 확인 결과 아래와 같다.

조건 조합	타당성 강화
rating >= 4.3	신뢰할 수 있는 제품

rating_count >= 1000	대표성 있는 피드백
discount_percentage >= 0.5	실질적 가격 메리트

: 가성비 상품 결과 :

```
-- 할인율+평점 조합 분석 / 가성비 제품 찾아보기
WITH cleaned_data AS (
  SELECT
    product_id,
    product_name,
    SAFE_CAST(REPLACE(REPLACE(CAST(discounted_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS discounted_price,
    SAFE_CAST(REPLACE(REPLACE(CAST(actual_price AS STRING), '₹', ''), ',', '') AS
FLOAT64) AS actual_price,
    SAFE_CAST(REPLACE(CAST(discount_percentage AS STRING), '%', '') AS FLOAT64) AS
discount_percentage,
    SAFE_CAST(CAST(rating AS STRING) AS FLOAT64) AS rating,
    SAFE_CAST(REPLACE(CAST(rating_count AS STRING), ',', '') AS INT64) AS
rating_count,
    SPLIT(category, '|')[OFFSET(0)] AS category_main
  FROM `shining-bearing-465502-n1.Mission.Amazon`
)
SELECT
  product_id,
  product_name,
  rating,
  discount_percentage,
  discounted_price,
  actual_price,
  ROUND(actual_price-discounted_price, 0) AS price_saved
  FROM cleaned_data
  WHERE rating >=4.3
   AND discount_percentage >=0.5
  ORDER BY rating DESC,
  discount_percentage DESC
  LIMIT 20;
```

prc	product_name	rating	discount_percentag	discounte	actual_pri	price_save
B0BP7	Syncwire LTG to USB Cable for Fast Charging Compatible with Phone 5/ 5C/ 5S/ 6/ 6S/ 7/8/ X/XR/XS M	5	0.8	399	1999	1600
B0BQR	REDTECH USB-C to Lightning Cable 3.3FT, [Apple MFi Certified] Lightning to Type C Fast Charging Cord	5	0.75	249	999	750
B09ZH	Amazon Basics Wireless Mouse | 2.4 GHz Connection, 1600 DPI | Type - C Adapter | Upto 12 Months of	5	0.5	499	1000	501
B0B53	Instant Pot Air Fryer, Vortex 2QT, Touch Control Panel, 360° EvenCrisp™ Technology, Uses 95 % less Oil,	4.8	0.75	4995	20049	15054
B0B23	Spigen EZ Fit Tempered Glass Screen Protector for iPhone 14 Pro Max - 2 Pack (Sensor Protection)	4.7	0.66	999	2899	1900
B0BLC	Zuvexa USB Rechargeable Electric Foam Maker - Handheld Milk Wand Mixer Frother for Hot Milk, Hand	4.7	0.62	499	1299	800
B0BP8	Campfire Spring Chef Prolix Instant Portable Water Heater Geyser 1Ltr. for Use Home Stainless Steel B	4.7	0.57	1499	3500	2001
B0BM9	Multifunctional 2 in 1 Electric Egg Boiling Steamer Egg Frying Pan Egg Boiler Electric Automatic Off wit	4.7	0.56	699	1599	900
B0BM4	FIGMENT Handheld Milk Frother Rechargeable, 3-Speed Electric Frother for Coffee with 2 Whisks and C	4.7	0.56	699	1599	900
B0BJ96	Aquadpure Copper + Mineral RO+UV+UF 10 to 12 Liter RO + UV + TDS ADJUSTER Water Purifier with C	4.6	0.8	4999	24999	20000
B09P1	Melbon VM-905 2000-Watt Room Heater (ISI Certified, White Color) Ideal Electric Fan Heater for Small	4.6	0.67	998	2999	2001
B0B244	Spigen EZ Fit Tempered Glass Screen Protector for iPhone 14 Pro - 2 Pack (Sensor Protection)	4.6	0.66	999	2899	1900
B095RT	Spigen EZ Fit Tempered Glass Screen Protector Guard for iPhone 14/13/13 Pro - 2 Pack	4.6	0.66	999	2899	1900
B08TG	10k 8k 4k HDMI Cable, Certified 48Gbps 1ms Ultra High Speed HDMI 2.1 Cable 4k 120Hz 144Hz 2k 16	4.6	0.58	999	2399	1400
B09JN	Lint Remover Woolen Clothes Lint Extractor Battery Lint Removing Machine Bhur Remover	4.6	0.57	319	749	430
B0B9B	WeCool S5 Long Selfie Stick, with Large Reinforced Tripod Stand up to 61 Inch / 156 Cms, Ultra Long M	4.6	0.55	1799	3999	2200
B0B9B	WeCool S5 Long Selfie Stick, with Large Reinforced Tripod Stand up to 61 Inch / 156 Cms, Ultra Long M	4.6	0.55	1799	3999	2200
B0BN6	VRPRIME Lint Roller Lint Remover for Clothes, Pet | 360 Sheets Reusable Sticky Easy-Tear Sheet Brush	4.6	0.5	499	999	500
B07Q4	ELV Aluminum Adjustable Mobile Phone Foldable Tabletop Stand Dock Mount for All Smartphones, Tab	4.5	0.82	269	1499	1230
B0BMX	Lapster 65W compatible for OnePlus Dash Warp Charge Cable , type c to c cable fast charging Data Sy	4.5	0.8	199	999	800

B0B23LW7NV	Spigen EZ Fit Ter	4.7	7779	0.66	999	2899
B0BP89YBC1	Campfire Spring	4.7	2591	0.57	1499	3500
B0BM9H2NY9	Multifunctional 2	4.7	2300	0.56	699	1599
B0BM4KTNL1	FIGMENT Handh	4.7	1729	0.56	699	1599
B0B244R4KB	Spigen EZ Fit Ter	4.6	6129	0.66	999	2899
B095RTJH1M	Spigen EZ Fit Ter	4.6	26603	0.66	999	2899
B08TGG316Z	10k 8k 4k HDMI (	4.6	3664	0.58	999	2399
B07Q4QV1DL	ELV Aluminum A	4.5	28978	0.82	269	1499
B088ZFJY82	Elv Aluminium A	4.5	28978	0.79	314	1499
B00NH11PEY	AmazonBasics U	4.5	74976	0.73	199	750
B00NH11PEY	AmazonBasics U	4.5	74976	0.73	199	750
B00NH11KIK	AmazonBasics U	4.5	107686	0.7	209	695
B08WJ86PV2	RPM Euro Games	4.5	2453	0.7	299	990
B00NH11KIK	AmazonBasics U	4.5	107687	0.7	209	695
B0765B3TH7	Gizga Essentials	4.5	13568	0.67	199	599
B00NH13Q8W	AmazonBasics U	4.5	74977	0.63	299	800
B0B9BD2YL4	KINGONE Wirele	4.5	1526	0.63	2599	6999
B0B9BD2YL4	KINGONE Wirele	4.5	1526	0.63	2599	6999
B07RY2X9MP	AmazonBasics 1	4.5	1029	0.59	609	1500

가성비 준거 기준 합당성 입증.

이를 토대로 좋은 제품 = 추천 상품의 기준 틀을 성립할 수 있다.

"본 분석은 제품 평점, 리뷰 수, 할인율을 기준으로 가중치를 부여하여 추천 상품을 도출했습니다.
특히 평점 4.3 이상, 리뷰 수 상위 25%, 할인율 30% 이상인 제품을 '추천 제품군'으로 정의하였으며, 이는 고객 만족도와 가성비를 균형 있게 반영한 결과입니다.

좋은 제품이란 어떤 걸까 ?
이를 위한 준거 기준이 세워질 수 있을까?

요소	설명	판단 기준 (예시)
고평점 (Quality)	사용자가 만족한 정도	rating ≥ 4.0 이상
리뷰 수 (Reliability)	얼마나 많은 사용자 의견이 누적됐는가	rating_count ≥ 500 또는 사분위 기준
할인율 (Value)	얼마나 저렴하게 구매 가능한가	discount_percentage ≥ 0.3 또는 상위 30%
카테고리 신뢰도	해당 카테고리 전체가 평균적으로 우수한가	카테고리별 평균 분석
신상품 여부	최근 등록된 제품인가 (옵션)	등록일 있을 경우
일관성	리뷰 편차 없이 고르게 평가되었는가	표준편차 활용 (추후 분석 가능)

지표기준확인 작업

Job information | Results | Chart | JSON | Execution detail

Row	rating_quartiles	review_quartiles	discount_deciles
1	2	2	0.0
	4	1191	0.18
	4.1	5057	0.28
	4.3	17413	0.37
	|	426973	0.43
			0.5
			0.55
			0.6
			0.66
			0.76
			0.94

지표	분포 요약	Q2 (중간값)	Q3 (상위 25%) 기준
평점 (rating)	[2.0, 4.0, 4.1, 4.3, ?]	4.1	4.3
리뷰 수 (rating_count)	[2, 1191, 5057, 17413, 426,973]	5057	17,413
할인율 (discount_percentage)	[0.0, ..., 0.94]	0.43~0.5 (중간구간)	0.5 이상 (상위 30%)

새로운 준거 기준 틀 작성

등급

추천 제품

→ rating ≥ 4.3 (Q3 이상)

→ rating_count ≥ 17,413 (Q3 이상)

→ discount_percentage ≥ 0.5 (상위 30%)

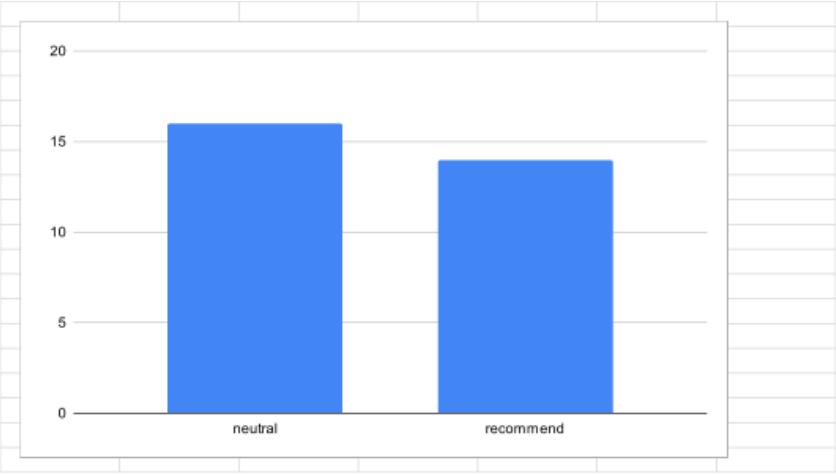
중간 제품

→ 조건 2 개 이상 충족 시

→ 예: 평점 높고 리뷰 많지만 할인 낮은 제품 등

비 추천 제품

→ 조건 1 개 이하 충족 (또는 평점 < 4.0 등)



B014I8SX4Y	Amazon Basics I	Electronics	4.4	426973	0.78	28.67	recommend
B07KSMBL2H	AmazonBasics F	Electronics	4.4	426973	0.69	28.22	recommend
B07KSMBL2H	AmazonBasics F	Electronics	4.4	426972	0.69	28.22	recommend
B014I8SSD0	Amazon Basics I	Electronics	4.4	426973	0.35	26.52	neutral
B09X7DY7Q4	SanDisk Extreme	Electronics	4.5	205052	0.48	26.3	neutral
B00NH11KIK	AmazonBasics U	Computers&Accessories	4.5	107687	0.7	26.14	recommend
B00NH11KIK	AmazonBasics U	Computers&Accessories	4.5	107686	0.7	26.14	recommend
B07GPXXNNG	boAt Bassheads	Electronics	4.1	363713	0.65	26.05	neutral
B005FYNT3G	SanDisk Cruzer B	Computers&Accessories	4.3	253105	0.56	26.03	recommend
B071Z8M4KX	boAt BassHeads	Electronics	4.1	363711	0.63	25.95	neutral
B07GQD4K6L	boAt Bassheads	Electronics	4.1	363713	0.62	25.9	neutral
B09MT84WV5	Samsung EVO Pl	Electronics	4.3	140035	0.74	25.83	recommend
B09MT84WV5	Samsung EVO Pl	Electronics	4.3	140036	0.71	25.68	recommend
B01N6LU1VF	SanDisk Ultra Du	Computers&Accessories	4.3	189104	0.59	25.64	recommend
B00NH11PEY	AmazonBasics U	Computers&Accessories	4.5	74976	0.73	25.59	recommend
B00NH11PEY	AmazonBasics U	Computers&Accessories	4.5	74976	0.73	25.59	recommend
B09MT6XSFW	Samsung EVO Pl	Electronics	4.3	140036	0.68	25.53	recommend
B01LWYDEQ7	Pigeon Polyprop	Home&Kitchen	4.1	270563	0.6	25.27	neutral
B00NH13Q8W	AmazonBasics U	Computers&Accessories	4.5	74977	0.63	25.09	recommend
B08HVL8QN3	Mi 10000mAH Li	Electronics	4.3	178912	0.48	24.99	neutral
B08HVJCW95	MI 10000mAh 3i	Electronics	4.3	178912	0.48	24.99	neutral
B08L5FM4JC	SanDisk Ultra mi	Electronics	4.4	67260	0.73	24.89	recommend
B08H9Z3XQW	boAt Bassheads	Electronics	4.1	161677	0.69	24.81	neutral
B07DC4RZPY	Amazon Basics U	Computers&Accessories	4.1	178817	0.65	24.78	neutral
B0088TKTY2	TP-LINK WiFi Do	Computers&Accessories	4.2	179691	0.54	24.77	neutral
B008IFXQFU	TP-Link USB WiF	Computers&Accessories	4.2	179691	0.5	24.57	neutral
B07G3YNLJB	Crucial BX500 24	Computers&Accessories	4.5	92925	0.41	24.41	neutral
B01DF26V7A	JBL C100SI Wire	Electronics	4.1	192589	0.54	24.37	neutral
B08TV2P1N8	boAt Rockerz 25	Electronics	4.1	141841	0.65	24.37	neutral
B07S9S86BF	boAt Bassheads	Electronics	4.1	161679	0.6	24.36	neutral

이 결과를 기반으로

각 대표 카테고리(category_main)별로 얼마나 많은 제품이 추천, 중간, 비추천 등급에 속하는지 파악

Row	category_main	total_products	recommended_co...	neutral_count	not_recommende...	recommended_ra...
1	Computers&Accessories	451	32	202	217	7.1
2	Electronics	526	20	228	278	3.8
3	Home&Kitchen	448	0	120	328	0.0
4	MusicalInstruments	2	0	0	2	0.0
5	Car&Motorbike	1	0	0	1	0.0
6	Health&PersonalCare	1	0	0	1	0.0
7	HomeImprovement	2	0	1	1	0.0
8	OfficeProducts	31	0	13	18	0.0
9	Toys&Games	1	0	1	0	0.0

→ 신뢰할 수 있는 추천군이 몰린 카테고리 도출: Computer*Accessories, Electronics 신뢰 제품을 다수 보유함. 전반적인 만족도가 높은 것으로 home*kitchen 도 살펴 볼 수 있었다.

Tech & Electronics 카테고리 내/ 추천 점수 상위 제품 20 개 조회

```
WITH cleaned AS (
  SELECT
    *,

    -- 안전한 타입 변환: 문자열을 실수(float64)로
```

```sql
    SAFE_CAST(rating AS FLOAT64) AS rating_f,
    SAFE_CAST(rating_count AS FLOAT64) AS rating_count_f,

    -- CASE 문을 통해 메인 카테고리를 재그룹핑
    CASE
      WHEN LOWER(category) LIKE '%electronic%' OR LOWER(category) LIKE '%computer%' OR
LOWER(category) LIKE '%accessorie%' THEN 'Tech & Electronics'
      WHEN LOWER(category) LIKE '%home%' OR LOWER(category) LIKE '%kitchen%' THEN
'Home & Living'
      WHEN LOWER(category) LIKE '%music%' THEN 'Music & Instruments'
      WHEN LOWER(category) LIKE '%office%' THEN 'Office Supplies'
      WHEN LOWER(category) LIKE '%toy%' OR LOWER(category) LIKE '%game%' THEN 'Kids &
Toys'
      WHEN LOWER(category) LIKE '%health%' OR LOWER(category) LIKE '%personal%' THEN
'Health & Wellness'
      WHEN LOWER(category) LIKE '%car%' OR LOWER(category) LIKE '%motor%' THEN 'Auto &
Outdoors'
      ELSE 'Other'
    END AS category_group

  FROM `shining-bearing-465502-n1.Mission.Amazon`
)

-- 상위 추천 제품 조회
SELECT
  product_name,
  category_group,
  rating_f AS rating,
  rating_count_f AS rating_count,
  discount_percentage,
  ROUND(rating_f * LOG10(rating_count_f + 1) + discount_percentage * 5, 2) AS
recommendation_score
FROM cleaned
WHERE category_group = 'Tech & Electronics'
  AND rating_f IS NOT NULL
  AND rating_count_f IS NOT NULL
  AND discount_percentage IS NOT NULL
ORDER BY recommendation_score DESC
```

Row	product_name	category_group	rating	rating_count	discount_percent...	recommendation...
1	Amazon Basics High-Speed HD...	Tech & Electronics	4.4	426973.0	0.78	28.67
2	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	Tech & Electronics	4.4	426973.0	0.69	28.22
3	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	Tech & Electronics	4.4	426972.0	0.69	28.22
4	Amazon Basics High-Speed HDMI Cable, 6 Feet - Supports Ethernet, 3D, 4K video,Black	Tech & Electronics	4.4	426973.0	0.35	26.52
5	SanDisk Extreme SD UHS I 64GB Card for 4K Video for DSLR and Mirrorless Cameras 170MB/s Read & 80MB/s Write	Tech & Electronics	4.5	205052.0	0.48	26.3
6	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Tech & Electronics	4.5	107687.0	0.7	26.14
7	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6	Tech & Electronics	4.5	107686.0	0.7	26.14

추천 카테고리 그룹의 정교화를 통해서 특정 타깃을 세워 집중 전략을 세워보기

사용자군	특징	주요 키워드	추천 기준 아이디어
게이머 (Gamer)	고성능/게임 관련 제품에 관심	gaming, headset, monitor, keyboard, mouse	평점 4.1↑, 리뷰 수 많고 할인 제품 우대
재택근무 직장인	사무/영상회의 장비, 무선	wireless, webcam, laptop, speaker	리뷰 많고 무선 기능 필수
육아/가정 중심	실속 제품, 안정성, 평판	baby, toys, safety, monitor	할인율 높고 평점 4.3↑
음향 애호가	고품질 사운드, 블루투스	headphone, bluetooth, noise cancelling	리뷰 수 5 천 이상, 할인 있으면 가산점

"게이머(Gamer)" 타깃 전략 세우기

기존 추첨 기준 척도를 가지고

키워드 필터: "gaming", "monitor", "headset", "keyboard", "mouse"

[추천 조건 설정]

평점 ≥ 4.1

리뷰 수 ≥ 5000
할인율 ≥ 30%

boAt Xtend Smar	4.2	69622	0.71	23.89
boAt Xtend Smar	4.2	69619	0.71	23.89
HP X1000 Wired	4.3	54315	0.59	23.31
Redgear MP35 S	4.6	33434	0.46	23.11
Fire-Boltt Phoeni:	4.3	27696	0.8	23.1
Fire-Boltt Phoeni:	4.3	27704	0.8	23.1
Fire-Boltt Phoeni:	4.3	27696	0.8	23.1
Fire-Boltt Phoeni:	4.3	27709	0.8	23.1
TP-Link USB Blue	4.3	95116	0.33	23.06
Dell MS116 1000	4.5	33176	0.54	23.04

게이머를 타깃으로 했지만, 스마트워치와 마우스, 게이밍 장패드가 상위.

인사이트

"gaming", "mouse", "headset" 등의 키워드 기반으로 추천 점수가 높은 제품은 많지 않지만,

마우스/장패드와 스마트워치는 할인율이 높고 리뷰 수가 많아 상위권 진입

이 결과는 Amazon 에서 게이머 관련 제품에 대한 실질적 리뷰 집중도를 반영한 것으로 보여짐.

FINDINGS

할인이 강력한 제품 + 리뷰 수 많은 제품이 상위 점수로 이어짐

- boAt, Fire-Boltt, TP-Link, HP, Dell 등 브랜드가 신뢰 기반
- 마우스, 마우스패드, 스마트워치, 블루투스 동글 등 **게이밍/컴퓨터 주변기기** 집중

그렇다면, 집중된 키워드에 적합한 적중 타깃으로 음향 시스템을 중시하는 고객을 상대로 추천 시스템을 만들어보자.

카테고리: Tech & Electronics

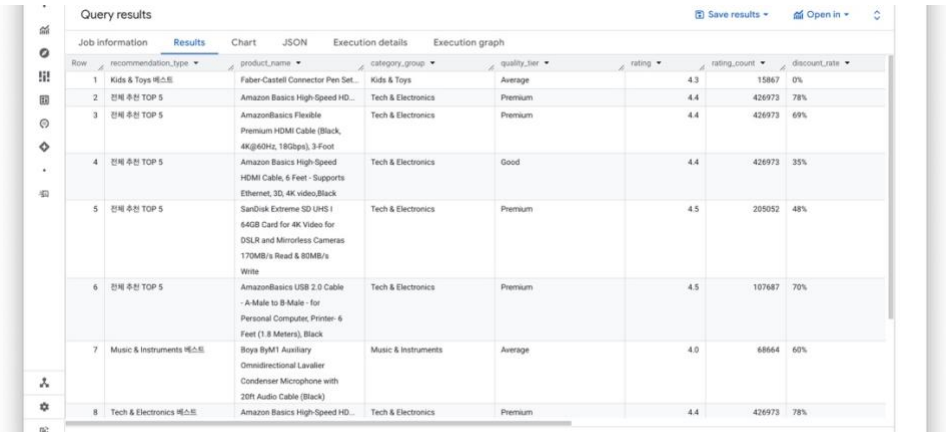
rating ≥ 4.3, **rating_count ≥ 5000**, **discount ≥ 50%**

제품명 키워드: "headphone", "bluetooth", "earphone", "noise cancelling" 등

추천 점수: 평점 × log10(리뷰 수 + 1) + 할인률 × 5

Row	product_name	rating	rating_count	discount_percent...	recommendation...
1	Gizga Essentials Hard Drive Case Shell, 6.35cm/2.5-inch, Portable Storage Organizer Bag for Earphone USB Cable Power Bank Mobile Charger	4.5	13568.0	0.67	21.95
2	Fire-Boltt Phoenix Smart Watch with Bluetooth Calling 1.3",120+ Sports Modes, 240*240 PX High Res with SpO2, Heart Rate Monitoring &	4.3	27696.0	0.8	23.1
3	Fire-Boltt Phoenix Smart Watch with Bluetooth Calling 1.3",120+ Sports Modes, 240*240 PX High Res with SpO2, Heart Rate Monitoring &	4.3	27696.0	0.8	23.1
4	Fire-Boltt Phoenix Smart Watch with Bluetooth Calling 1.3",120+ Sports Modes, 240*240 PX High Res with SpO2, Heart Rate Monitoring &	4.3	27704.0	0.8	23.1
5	Fire-Boltt Phoenix Smart Watch with Bluetooth Calling 1.3",120+ Sports Modes,	4.3	27709.0	0.8	23.1

추가 분석: 중복 제거 효과 확인



Query results

Save results Open in

Job information | Results | Chart | JSON | Execution details | Execution graph

Row	recommendation_type	product_name	category_group	quality_tier	rating	rating_count	discount_rate
1	Kids & Toys 베스트	Faber-Castell Connector Pen Set...	Kids & Toys	Average	4.3	15867	0%
2	전체 추천 TOP 5	Amazon Basics High-Speed HD...	Tech & Electronics	Premium	4.4	426973	78%
3	전체 추천 TOP 5	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot	Tech & Electronics	Premium	4.4	426973	69%
4	전체 추천 TOP 5	Amazon Basics High-Speed HDMI Cable, 6 Feet - Supports Ethernet, 3D, 4K video,Black	Tech & Electronics	Good	4.4	426973	35%
5	전체 추천 TOP 5	SanDisk Extreme SD UHS I 64GB Card for 4K Video for DSLR and Mirrorless Cameras 170MB/s Read & 80MB/s Write	Tech & Electronics	Premium	4.5	205052	48%
6	전체 추천 TOP 5	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black	Tech & Electronics	Premium	4.5	107687	70%
7	Music & Instruments 베스트	Boya ByM1 Auxiliary Omnidirectional Lavalier Condenser Microphone with 20ft Audio Cable (Black)	Music & Instruments	Average	4.0	68664	60%
8	Tech & Electronics 베스트	Amazon Basics High-Speed HD...	Tech & Electronics	Premium	4.4	426973	78%

-- Amazon 스토어 데이터 중복 제거 추천 시스템

-- 목표: 동일 제품명의 중복을 제거하고 추천 점수 기반 상위 5 개 제품 추출

-카테고리 재/그룹핑

```
-- 카테고리 그룹핑
    CASE
      WHEN LOWER(category) LIKE '%electronic%' OR LOWER(category) LIKE '%computer%' OR
LOWER(category) LIKE '%accessorie%' THEN 'Tech & Electronics'
      WHEN LOWER(category) LIKE '%home%' OR LOWER(category) LIKE '%kitchen%' THEN
'Home & Living'
      WHEN LOWER(category) LIKE '%music%' THEN 'Music & Instruments'
      WHEN LOWER(category) LIKE '%office%' THEN 'Office Supplies'
      WHEN LOWER(category) LIKE '%toy%' OR LOWER(category) LIKE '%game%' THEN 'Kids &
Toys'
      WHEN LOWER(category) LIKE '%health%' OR LOWER(category) LIKE '%personal%' THEN
'Health & Wellness'
      WHEN LOWER(category) LIKE '%car%' OR LOWER(category) LIKE '%motor%' THEN 'Auto &
Outdoors'
      ELSE 'Other'
    END AS category_group

  FROM `shining-bearing-465502-n1.Mission.Amazon`
  WHERE rating IS NOT NULL
    AND rating_count IS NOT NULL
    AND discount_percentage IS NOT NULL
),
```

recommendation_type	product_name
Kids & Toys 베스트	Faber-Castell Connector Pen Set - Pack of 25 (Assorted)
전체 추천 TOP 5	Amazon Basics High-Speed HDMI Cable, 6 Feet (2-Pack),Black
전체 추천 TOP 5	AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot
전체 추천 TOP 5	Amazon Basics High-Speed HDMI Cable, 6 Feet - Supports Ethernet, 3D, 4K video,Black
전체 추천 TOP 5	SanDisk Extreme SD UHS I 64GB Card for 4K Video for DSLR and Mirrorless Cameras 170MB/s Read & 80MB/s Write
전체 추천 TOP 5	AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black
Music & Instruments 베스트	Boya ByM1 Auxiliary Omnidirectional Lavalier Condenser Microphone with 20ft Audio Cable (Black)
Tech & Electronics 베스트	Amazon Basics High-Speed HDMI Cable, 6 Feet (2-Pack),Black
Office Supplies 베스트	COI Note Pad/Memo Book with Sticky Notes & Clip Holder with Pen for Gifting
Home & Living 베스트	Pigeon Polypropylene Mini Handy and Compact Chopper with 3 Blades for Effortlessly Chopping Vegetables and Fruits for Your Kitchen (12420, Green, 400 ml)

1~5 추천 시스템 정리 표

추천 시스템 이름	추천 시스템의 테마	구현 로직	결과 요약
리뷰를 많이 남긴 상품들이에요	많은 리뷰 수 + 높은 평점을 반영하여 신뢰할 수 있는 제품 추천	SQL 에서 rating_count >= 5000, rating >= 4.3 이상 조건 필터 후 추천 점수 계산	HP 마우스, boAt 스마트워치, Redgear 마우스패드 등 리뷰 수 3 만~9 만, 평점 4.2 이상 제품들이 추출됨
최대 할인율 기반 추천 시스템	할인율을 기준으로 추천 점수에 반영하여 가성비 좋은 제품 추천	추천 점수 = 평점 * log10(리뷰수 + 1) + 할인율 * 5 식으로 계산. discount_percentage ≥ 0.3 조건 포함	80% 할인 스마트워치, 71% 할인 이어폰 등 높은 할인율 제품 다수 추천됨
게이머 타깃 추천 시스템	'gaming', 'mouse', 'keyboard' 키워드 기반의 게이머 맞춤형 제품 추천	카테고리 'Tech & Electronics', 키워드 포함, 평점 ≥ 4.1, 리뷰수 ≥ 5000 조건 적용	HP 게이밍 마우스, Redgear 마우스패드 등 실사용 많은 제품이 추출됨
음향 애호가 추천 시스템	블루투스, 이어폰, 노이즈캔슬링 키워드를 기반으로 사운드 중심 사용자 대상 추천	'headphone', 'bluetooth' 키워드 필터 + 리뷰 수, 할인율, 평점 기반 점수 계산 및 정렬	boAt 헤드폰, TP-Link 블루투스 동글 등 리뷰 많은 음향 관련 제품 도출

정제된 중복 제거 추천 시스템	추천 제품 중복 제거 후 상위 제품만 정제하여 사용자에게 노출	drop_duplicates(subset='product_name') 적용 후 추천 점수 상위 5 개 추출	동일 제품 반복 제거 → 정확한 top-N 리스트 제공, 예: HP 마우스 단일화