

SQL Project Report – Amazon Analysis

Author: SAHIL SHAIKH

Date: AUGUST 2025

1. Project Overview

The objective of this project is to analyze Amazon product data to identify top-performing brands, their ratings, and sales trends across different categories.

This analysis also focuses on determining the highest and lowest priced products in the dataset and studying patterns in discounts, ratings, and product popularity.

The project demonstrates data cleaning, transformation, and advanced SQL queries to extract meaningful insights that can help businesses in pricing strategies, category performance evaluation, and customer preference analysis. Dataset Details:

- **Source:** <https://www.kaggle.com/datasets/karkavelrajaj/amazon-sales-dataset>
 - **Rows:** 339
 - **Columns:** 19
 - **Key Fields:** brand, product_name, rating, rating_count, actual_price_num, discounted_price_num
-

2. Tools & Technologies Used

- Oracle SQL Developer
 - SQL (SELECT, GROUP BY, ORDER BY, UNION, Aggregate Functions)
 - CSV dataset
-

3. Queries & Insights

Converting actual price column from varchar to number and create new column brand

-- creating new column for converting varchar columns to num columns

```
ALTER TABLE amazon ADD actual_price_num NUMBER;
```

```
ALTER TABLE amazon ADD discounted_price_num NUMBER;
```

```
UPDATE amazon
```

```
SET actual_price_num = TO_NUMBER(REPLACE(REPLACE(actual_price, 'â', ''), ',', ''))
```

```
WHERE REGEXP_LIKE(actual_price, '^[0-9â,1\.\.]+$', 'i');
```

```
UPDATE amazon
```

```
SET discounted_price_num = TO_NUMBER(REPLACE(REPLACE(discounted_price, 'â', ''), ', ', ''));
```

```
WHERE REGEXP_LIKE(discounted_price, '^[0-9â,\.]+\d$');
```

```
SELECT actual_price, actual_price_num, discounted_price, discounted_price_num
```

```
FROM amazon
```

```
WHERE ROWNUM <= 10;
```

```
--
```

```
-- separating first word using regex to get separate brand name column for ease --
```

```
ALTER TABLE amazon ADD (brand VARCHAR2(100));
```

```
UPDATE amazon
```

```
SET brand = REGEXP_SUBSTR(product_name, '^[^ ]+');
```

```
SELECT product_name, brand
```

```
FROM amazon
```

```
WHERE ROWNUM <= 20;
```

--Result Screenshot:

```
select * from amazon;

-- creating new column for converting varchar columns to num columns
ALTER TABLE amazon ADD actual_price_num NUMBER;
ALTER TABLE amazon ADD discounted_price_num NUMBER;

UPDATE amazon
SET actual_price_num = TO_NUMBER(REPLACE(REPLACE(actual_price, 'â', ''), ', ', ''))
WHERE REGEXP_LIKE(actual_price, '^[0-9â,\.]+\d$');

UPDATE amazon
SET discounted_price_num = TO_NUMBER(REPLACE(REPLACE(discounted_price, 'â', ''), ', ', ''))
WHERE REGEXP_LIKE(discounted_price, '^[0-9â,\.]+\d$');

SELECT actual_price, actual_price_num, discounted_price, discounted_price_num
FROM amazon
WHERE ROWNUM <= 10;

--
-- separating first word using regex to get separate brand name column for ease --
ALTER TABLE amazon ADD (brand VARCHAR2(100));
UPDATE amazon
SET brand = REGEXP_SUBSTR(product_name, '^[^ ]+');
SELECT product_name, brand
FROM amazon
WHERE ROWNUM <= 20;

-- Top 10 Product by Product Rating --
SELECT brand, rating, RATING_COUNT, actual_price_num
FROM amazon
WHERE RATING >= 4.5
AND RATING_COUNT >= 5000
ORDER BY rating DESC, RATING_COUNT DESC
FETCH FIRST 10 ROWS ONLY;
```

Insight:

This is done to ensure the arithmetic operation such as discount should be calculated and creating a new column brand to get top brand

Query 1 – Top 10 Brand by Product Rating

SELECT brand, rating, RATING_COUNT, actual_price_num

FROM amazon

WHERE RATING >= 4.5

AND RATING_COUNT >= 5000

ORDER BY rating DESC, RATING_COUNT DESC

FETCH FIRST 10 ROWS ONLY;

Result Screenshot:

The screenshot shows an SQL Worksheet interface. The top section is the 'Query Builder' tab, which contains the following SQL query:

```
-- Top 10 Product by Product Rating --  
SELECT brand, rating, RATING_COUNT, actual_price_num  
FROM amazon  
WHERE RATING >= 4.5  
AND RATING_COUNT >= 5000  
ORDER BY rating DESC, RATING_COUNT DESC  
FETCH FIRST 10 ROWS ONLY;  
  
-- Products with highest discount %  
SELECT product_name,  
       actual_price_num,
```

The bottom section is the 'Query Result' tab, which displays the results of the first query. It shows a table with 4 columns: BRAND, RATING, RATING_COUNT, and ACTUAL_PRICE_NUM. The results are sorted by rating in descending order, and then by rating count in descending order. The first 10 rows are displayed.

	BRAND	RATING	RATING_COUNT	ACTUAL_PRICE_NUM
1	Spigen	4.7	7779	2899
2	Logitech	4.6	10652	2295
3	Sujata	4.6	6550	8478
4	Spigen	4.6	6129	2899
5	SanDisk	4.5	205052	1800
6	AmazonBasics	4.5	107686	695
7	AmazonBasics	4.5	74976	750
8	SanDisk	4.5	19624	2900
9	Gigga	4.5	13568	599
10	Philips	4.5	7949	2095

Insight:

Top 10 Brand with product rating, their price and rating count

Query 2 – Products with highest discount %

SELECT PRODUCT_NAME,

actual_price_num,

discounted_price_num,

ROUND(((actual_price_num - discounted_price_num) * 100) / actual_price_num, 2) AS
discount_pct

FROM amazon

WHERE actual_price_num > 0

AND discounted_price_num IS NOT NULL

ORDER BY discount_pct DESC

FETCH FIRST 10 ROWS ONLY;ORDER BY distance DESC;

Result Screenshot:

The screenshot shows a SQL Worksheet interface with a query editor and a results pane. The query is as follows:

```
-- Products with highest discount %
SELECT PRODUCT_NAME,
       actual_price_num,
       discounted_price_num,
       ROUND(((actual_price_num - discounted_price_num) * 100) / actual_price_num, 2) AS discount_pct
FROM amazon
WHERE actual_price_num > 0
AND discounted_price_num IS NOT NULL
ORDER BY discount_pct DESC
FETCH FIRST 10 ROWS ONLY;
```

The results pane displays the following data:

PRODUCT_NAME	ACTUAL_PRICE_NUM	DISCOUNTED_PRICE_NUM	DISCOUNT_PCT
1 LAFSTER Spiral Charger Spiral Charger Cable Protectors for Wires Data Cable Saver Charging Cord Protective Cable Cover Set of 3 (12 Pieces)	999	99	90.09
2 Sounce Fast Phone Charging Cable & Data Sync USB Cable Compatible for iPhone 13, 12,11, X, 8, 7, 6, 5, iPad Air, Pro, Mini & iOS Devices	1899	199	89.52
3 Sounce Gold Plated 3.5 mm Headphone Splitter for Computer 2 Male to 1 Female 3.5mm Headphone Mic Audio Y Splitter Cable Smartphone Headset to PC Adapter 86" (Black,20cm)	999	120	87.99
4 PTron Solero T241 2.4A Type-C Data & Charging USB Cable, Made in India, 480Mbps Data Sync, Durable 1-Meter Long USB Cable for Type-C USB Devices for Charging Adapter (Black)	800	99	87.63
5 STRIFF Mpad Mouse Mat 230X190X3mm Gaming Mouse Pad, Non-Slip Rubber Base, Waterproof Surface, Premium-Textured, Compatible with Laser and Optical Mice(Universe Black)	999	129	87.09
6 Macmillan Aquafresh 5 Micron PS-05 10" in PP Spun Filter Candle Set for All Type RO Water Purifier 10 inch (4)	1499	215	85.66
7 PTron Solero TB301 3A Type-C Data and Fast Charging Cable, Made in India, 480Mbps Data Sync, Strong and Durable 1.5-Meter Nylon Braided USB Cable for Type-C Devices for Charging Adapter (Black)	1000	149	85.1
8 PTron Solero TB301 3A Type-C Data and Fast Charging Cable, Made in India, 480Mbps Data Sync, Strong and Durable 1.5-Meter Nylon Braided USB Cable for Type-C Devices for Charging Adapter (Black)	1000	149	85.1
9 Lapster 5 pin mini usb cable, usb b cable,camera cable usb2.0 for External HDDS/Card Readers/Camera etc.	999	149	85.09
10 Caldipree Silicone Case Cover Compatible for 2022 Samsung Smart TV Remote QLED TV BN68-13897A TM2280E (2022-BLACK)	2999	547	81.76

Insight:

query is used to get the most discounted products on amazon with actual price and discounted price

Query 3 – Highest Price vs Lowest Price of product range

SELECT

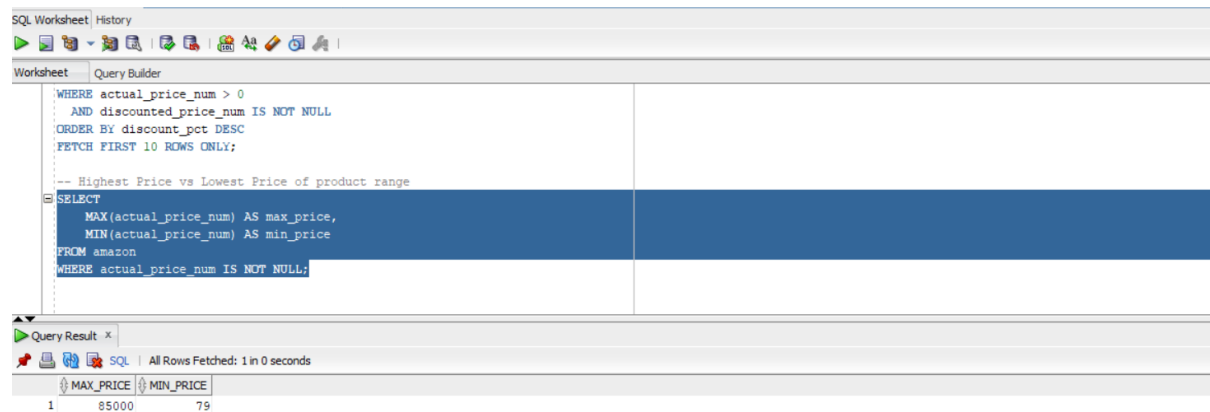
MAX(actual_price_num) AS max_price,

MIN(actual_price_num) AS min_price

FROM amazon

WHERE actual_price_num IS NOT NULL;

Result Screenshot:



The screenshot shows an SQL Worksheet interface. The 'Query Builder' tab is active, displaying a SQL query. The query is as follows:

```
WHERE actual_price_num > 0  
AND discounted_price_num IS NOT NULL  
ORDER BY discount_pct DESC  
FETCH FIRST 10 ROWS ONLY;  
  
-- Highest Price vs Lowest Price of product range  
SELECT  
MAX(actual_price_num) AS max_price,  
MIN(actual_price_num) AS min_price  
FROM amazon  
WHERE actual_price_num IS NOT NULL;
```

Below the query editor, the 'Query Result' tab is active, showing the results of the query. The results are displayed in a table with two columns: MAX_PRICE and MIN_PRICE. The first row shows a MAX_PRICE of 85000 and a MIN_PRICE of 79.

	MAX_PRICE	MIN_PRICE
1	85000	79

Insight:

This Query is used to get highest price and lowest price range in amazon

Query 4 – Top 10 Expensive product which has highest rating

SELECT product_name, actual_price_num, rating, 'High Price' AS flag

FROM amazon

WHERE actual_price_num > 2000

UNION

SELECT product_name, actual_price_num, rating, 'High Rating' AS flag

FROM amazon

WHERE rating >= 4.5

FETCH FIRST 10 ROWS ONLY;

Result Screenshot:

The screenshot shows a SQL Worksheet with a query in the Query Builder and its results in the Query Result pane. The query is a UNION of two SELECT statements from the 'amazon' table. The first SELECT statement filters for products with an actual price greater than 2000, and the second filters for products with a rating greater than or equal to 4.5. The results are sorted by product name.

PRODUCT_NAME
1 !!1000 Watt/2000-Watt Room Heater!! Fan Heater!!Pure White!!HN-2500!!Made in India!!
2 7SEVEN Compatible with Fire Tv Stick Remote with Voice Command Feature Suitable for Second Generation Amazon Fire Tv Stick Remote Only - Pairing Must
3 AGARO 33398 Rapid 1000-Watt, 10-Litre Wet & Dry Vacuum Cleaner, with Blower Function (Red & Black)
4 AGARO Glory Cool Mist Ultrasonic Humidifier, 4.5Litres, For Large Area, Room, Home, Office, Adjustable Mist Output, Ceramic Ball Filter, Ultra Quiet, 360° Rotatable Nozzle, A
5 AGARO Regal Electric Rice Cooker, 3L Ceramic Inner Bowl, Cooks Up to 600 Gms Raw Rice, SS Steamer, Preset Cooking Functions, Preset Timer, Keep Warm Function, LED Display, Bla
6 AGARO Supreme High Pressure Washer, 1800 Watts, 120 Bars, 6.5L/Min Flow Rate, 8 Meters Outlet Hose, Portable, for Car,Bike and Home Cleaning Purpose, Black and Orange
7 Active Easy Mix Nutri Mixer Grinder 500 Watt Long Lasting Shock Proof ABS Body Heavy Duty Motor With Nano - Grinding Technology
8 AmazonBasics USB 2.0 - A-Male to A-Female Extension Cable for Personal Computer, Printer (Black, 9.8 Feet/3 Meters)
9 AmazonBasics USB 2.0 Cable - A-Male to B-Male - for Personal Computer, Printer- 6 Feet (1.8 Meters), Black
10 Ant Esports GM320 RGB Optical Wired Gaming Mouse 8 Programmable Buttons 12800 DPI

The screenshot shows the same SQL Worksheet as above, but with the columns 'ACTUAL_PRICE_NUM', 'RATING', and 'FLAG' highlighted in the Query Result pane. The results are sorted by product name.

	ACTUAL_PRICE_NUM	RATING	FLAG
1.te!!HN-2500!!Made in India!!	1599	4.5	High Rating
2) Command Feature Suitable for Second Generation Amazon Fire Tv Stick Remote Only - Pairing Must	2999	3.3	High Price
3)leaner, with Blower Function (Red & Black)	5999	4	High Price
4) For Large Area, Room, Home, Office, Adjustable Mist Output, Ceramic Ball Filter, Ultra Quiet, 360° Rotatable Nozzle, Auto Shut Off, Grey	5799	4.3	High Price
5) Cooks Up to 600 Gms Raw Rice, SS Steamer, Preset Cooking Functions, Preset Timer, Keep Warm Function, LED Display, Black	5495	4.1	High Price
6), 6.5L/Min Flow Rate, 8 Meters Outlet Hose, Portable, for Car,Bike and Home Cleaning Purpose, Black and Orange	8990	4.3	High Price
7)ing Shock Proof ABS Body Heavy Duty Motor With Nano - Grinding Technology	2990	3.8	High Price
8) e for Personal Computer, Printer (Black, 9.8 Feet/3 Meters)	750	4.5	High Rating
9)onal Computer, Printer- 6 Feet (1.8 Meters), Black	695	4.5	High Rating
10)grammable Buttons 12800 DPI	2799	4.2	High Price

Insight:

Checking the most expensive products and their price and rating to see whether they are worth it or not.

Query 5 – Brand with highest rating and rating count on their products

SELECT brand,

SUM(rating_count) AS total_ratings,

ROUND(AVG(rating),2) AS avg_rating

FROM amazon

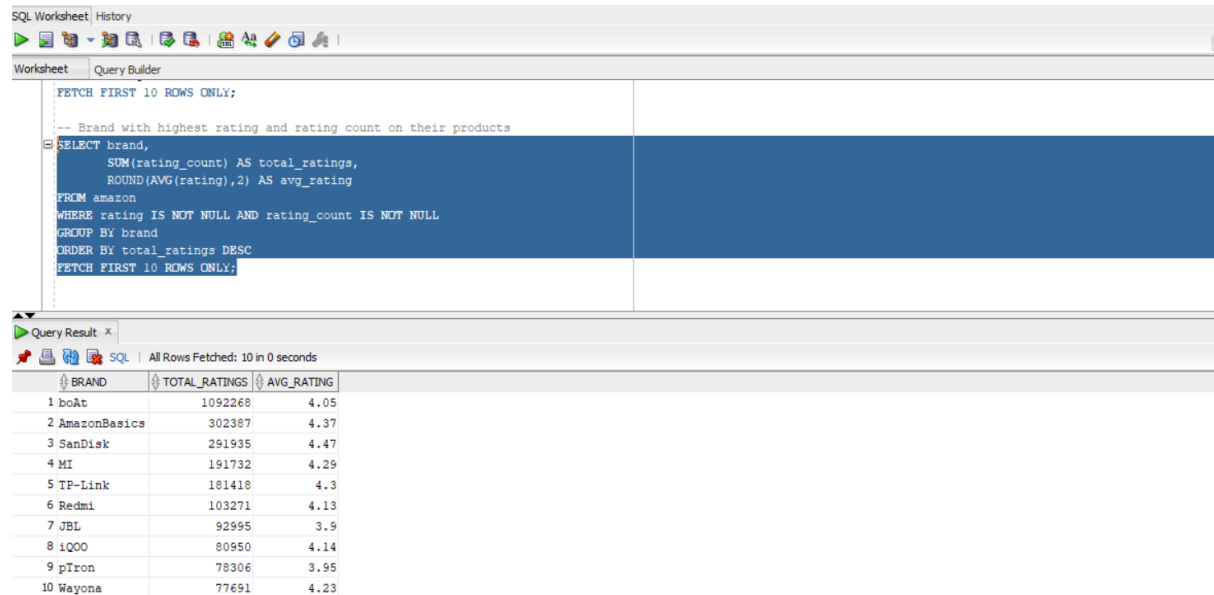
WHERE rating IS NOT NULL AND rating_count IS NOT NULL

GROUP BY brand

ORDER BY total_ratings DESC

FETCH FIRST 10 ROWS ONLY;

Result Screenshots:-



The screenshot shows an SQL Worksheet interface. The top section is the 'Query Builder' with a text area containing the following SQL query:

```
FETCH FIRST 10 ROWS ONLY;

-- Brand with highest rating and rating count on their products
SELECT brand,
       SUM(rating_count) AS total_ratings,
       ROUND(AVG(rating),2) AS avg_rating
FROM amazon
WHERE rating IS NOT NULL AND rating_count IS NOT NULL
GROUP BY brand
ORDER BY total_ratings DESC
FETCH FIRST 10 ROWS ONLY;
```

The bottom section is the 'Query Result' table, which displays the results of the query. It has 3 columns: BRAND, TOTAL_RATINGS, and AVG_RATING. The results are as follows:

BRAND	TOTAL_RATINGS	AVG_RATING
1 boAt	1092268	4.05
2 AmazonBasics	302387	4.37
3 SanDisk	291935	4.47
4 MI	191732	4.29
5 TP-Link	181418	4.3
6 Redmi	103271	4.13
7 JBL	92995	3.9
8 iQOO	80950	4.14
9 pTron	78306	3.95
10 Wayona	77691	4.23

Insight:

Top 10 Brands with highest number of Rating counts and avg number of rating on their every products

Query 6 – Category with most rating count

SELECT CATEGORY,

SUM(rating_count) AS total_ratings,

ROUND(AVG(rating),2) AS avg_rating

FROM amazon

WHERE rating IS NOT NULL AND rating_count IS NOT NULL

GROUP BY CATEGORY

ORDER BY total_ratings DESC

FETCH FIRST 10 ROWS ONLY;

Result Screenshots:-

SQL Worksheet History		
Worksheet Query Builder		
<pre> -- Category with most rating count, note rating count act as pseudo sales column as dataset dont have sales column SELECT CATEGORY, SUM(rating_count) AS total_ratings, ROUND(AVG(rating),2) AS avg_rating FROM amazon WHERE rating IS NOT NULL AND rating_count IS NOT NULL GROUP BY CATEGORY ORDER BY total_ratings DESC FETCH FIRST 10 ROWS ONLY; </pre>		
Query Result x		
All Rows Fetched: 10 in 0 seconds		
CATEGORY	TOTAL_RATINGS	AVG_RATING
1 Electronics Headphones,Earbuds Accessories Headphones In-Ear	933241	3.99
2 Computers Accessories Accessories Peripherals Cables Accessories Cables USBCables	809566	4.14
3 Electronics Accessories MemoryCards MicroSD	291935	4.47
4 Electronics HomeTheater,TV&Video Televisions SmartTelevisions	185697	4.19
5 Computers Accessories NetworkingDevices Repeaters Extenders	156638	4.2
6 Electronics Mobiles Accessories Smartphones BasicMobiles Smartphones	152252	4.13
7 Electronics WearableTechnology SmartWatches	121709	3.87
8 Electronics HomeTheater,TV&Video Accessories Cables HDMICables	109145	4.27
9 Electronics HomeTheater,TV&Video Accessories Cables RCCACables	73834	4.3
10 Electronics Mobiles Accessories MobileAccessories Chargers PowerBanks	72768	4

Insight:

Analysing which category type of products get the most rating count and average rating, note rating count act as pseudo sales column as dataset dont have sales column

Query 7 – Most expensive vs cheapest products in Electronics,Headphones Category

SELECT

MAX(actual_price_num) AS max_price,

MIN(actual_price_num) AS min_price

FROM amazon

WHERE actual_price_num IS NOT NULL

AND CATEGORY LIKE '%Electronics|Headphones%';

Result Screenshots:-

The screenshot shows a SQL Worksheet interface with a 'Query Builder' tab. The query is as follows:

```
GROUP BY CATEGORY
ORDER BY total_ratings DESC
FETCH FIRST 10 ROWS ONLY;

-- Most expensive vs cheapest products in Electronics,Headphones Category
SELECT
  MAX(actual_price_num) AS max_price,
  MIN(actual_price_num) AS min_price
FROM amazon
WHERE actual_price_num IS NOT NULL
AND CATEGORY LIKE '%Electronics|Headphones%';
```

Below the query, the 'Query Result' tab shows the results of the query. The results are displayed in a table with two columns: MAX_PRICE and MIN_PRICE. The first row shows a MAX_PRICE of 4499 and a MIN_PRICE of 499.

	MAX_PRICE	MIN_PRICE
1	4499	499

Insight:

after analysing top category we analyze its most expensive to least expensive price range

Query 7 – Most Expensive vs Cheapest products in Top 1 Brand in Amazon according

```
SELECT product_name, brand, actual_price_num, rating, rating_count
```

```
FROM amazon
```

```
WHERE brand = (
```

```
  SELECT brand
```

```
  FROM amazon
```

```
  WHERE rating IS NOT NULL
```

```
    AND rating_count IS NOT NULL
```

```
  GROUP BY brand
```

```
  ORDER BY SUM(rating_count) DESC
```

```
  FETCH FIRST 1 ROW ONLY
```

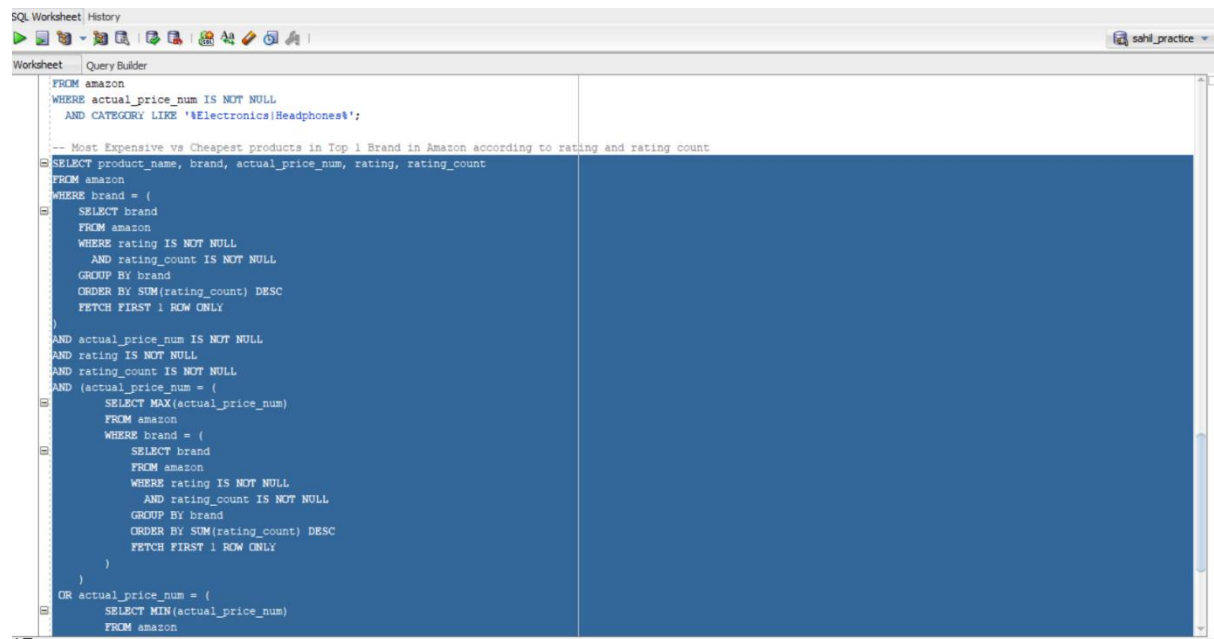
```
)
```

```
AND actual_price_num IS NOT NULL
```

```
AND rating IS NOT NULL
```

```
AND rating_count IS NOT NULL
AND (actual_price_num = (
    SELECT MAX(actual_price_num)
    FROM amazon
    WHERE brand = (
        SELECT brand
        FROM amazon
        WHERE rating IS NOT NULL
        AND rating_count IS NOT NULL
        GROUP BY brand
        ORDER BY SUM(rating_count) DESC
        FETCH FIRST 1 ROW ONLY
    )
)
OR actual_price_num = (
    SELECT MIN(actual_price_num)
    FROM amazon
    WHERE brand = (
        SELECT brand
        FROM amazon
        WHERE rating IS NOT NULL
        AND rating_count IS NOT NULL
        GROUP BY brand
        ORDER BY SUM(rating_count) DESC
        FETCH FIRST 1 ROW ONLY
    )
))
ORDER BY actual_price_num DESC;
```

Result Screenshots:-

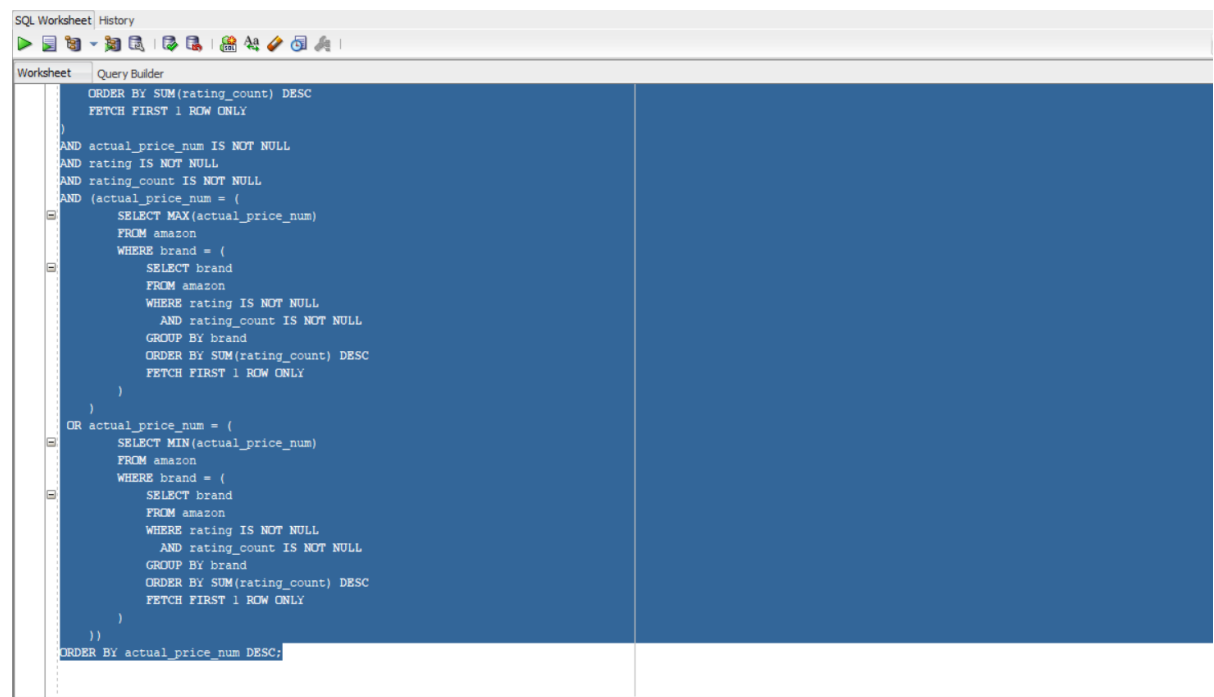


SQL Worksheet History

Worksheet Query Builder

```
FROM amazon
WHERE actual_price_num IS NOT NULL
AND CATEGORY LIKE '%Electronics|Headphones%';

-- Most Expensive vs Cheapest products in Top 1 Brand in Amazon according to rating and rating count
SELECT product_name, brand, actual_price_num, rating, rating_count
FROM amazon
WHERE brand = (
    SELECT brand
    FROM amazon
    WHERE rating IS NOT NULL
    AND rating_count IS NOT NULL
    GROUP BY brand
    ORDER BY SUM(rating_count) DESC
    FETCH FIRST 1 ROW ONLY
)
AND actual_price_num IS NOT NULL
AND rating IS NOT NULL
AND rating_count IS NOT NULL
AND (actual_price_num = (
    SELECT MAX(actual_price_num)
    FROM amazon
    WHERE brand = (
        SELECT brand
        FROM amazon
        WHERE rating IS NOT NULL
        AND rating_count IS NOT NULL
        GROUP BY brand
        ORDER BY SUM(rating_count) DESC
        FETCH FIRST 1 ROW ONLY
    )
)
OR actual_price_num = (
    SELECT MIN(actual_price_num)
    FROM amazon
```



SQL Worksheet History

Worksheet Query Builder

```
ORDER BY SUM(rating_count) DESC
FETCH FIRST 1 ROW ONLY
)
AND actual_price_num IS NOT NULL
AND rating IS NOT NULL
AND rating_count IS NOT NULL
AND (actual_price_num = (
    SELECT MAX(actual_price_num)
    FROM amazon
    WHERE brand = (
        SELECT brand
        FROM amazon
        WHERE rating IS NOT NULL
        AND rating_count IS NOT NULL
        GROUP BY brand
        ORDER BY SUM(rating_count) DESC
        FETCH FIRST 1 ROW ONLY
    )
)
OR actual_price_num = (
    SELECT MIN(actual_price_num)
    FROM amazon
    WHERE brand = (
        SELECT brand
        FROM amazon
        WHERE rating IS NOT NULL
        AND rating_count IS NOT NULL
        GROUP BY brand
        ORDER BY SUM(rating_count) DESC
        FETCH FIRST 1 ROW ONLY
    )
))
ORDER BY actual_price_num DESC;
```

The screenshot shows an SQL Worksheet interface with a query in the Query Builder. The query is as follows:

```


SELECT brand
FROM amazon
WHERE rating IS NOT NULL
AND rating_count IS NOT NULL
GROUP BY brand
ORDER BY SUM(rating_count) DESC
FETCH FIRST 1 ROW ONLY
)
OR actual_price_num = (
SELECT MIN(actual_price_num)
FROM amazon
WHERE brand = :1
)

```

The Query Result pane shows the following data:

PRODUCT_NAME
1 boAt Wave Call Smart Watch, Smart Talk with Advanced Dedicated Bluetooth Calling Chip, 1.698" HD Display with 550 NITS & 70% Color Gamut, 150+ Watch Faces, Multi-Sport Modes,
2 boAt Newly Launched Wave Electra with 1.81" HD Display, Smart Calling with Ultra-Seamless BT Calling Chip, 20 Built-In Watch Faces, 100+ Sports Modes, Menu Personalization, In-Bu
3 boAt Micro USB 55 Tangle-free, Sturdy Micro USB Cable with 3A Fast Charging & 480Mbps Data Transmission (Black)


SQL Worksheet | History



Worksheet | Query Builder

```
SELECT brand
FROM amazon
WHERE rating IS NOT NULL
AND rating_count IS NOT NULL
GROUP BY brand
ORDER BY SUM(rating_count) DESC
FETCH FIRST 1 ROW ONLY
)
)
OR actual_price_num = (
SELECT MIN(actual_price_num)
FROM amazon
WHERE brand = (
```

Query Result x

 SQL | All Rows Fetched: 3 in 0.004 seconds

	BRAND	ACTUAL_PRICE_NUM	RATING	RATING_COUNT	
1	Calling Chip, 1.698" HD Display with 550 NITS & 70% Color Gamut, 150+ Watch Faces, Multi-Sport Modes, HR, SpO2 (Caribbean Green)	boAt	7990	3.8	17833
2	ch Ultra-Seamless BT Calling Chip, 20 Built-In Watch Faces, 100+ Sports Modes, Menu Personalization, In-Built Games (Charcoal Black)	boAt	7990	4.1	154
3	ing & 480Mbps Data Transmission (Black)	boAt	499	4.1	15188

Insight:

Analysing top 1 brand i.e Boat with its most expensive product to least expensive product according to actual price and rating_count

4. Key Findings

Data Cleaning and Transformation

- Converted price columns from VARCHAR2 to numeric (NUMBER) for analysis.
- Extracted **brand names** from product titles using regex for grouping insights.

Top Performing Products

- Identified products with **ratings ≥ 4.5** and **5,000+ reviews**, highlighting the most trusted items.

Highest Discounts

- Found products offering **>50% discount**, useful for price strategy evaluations.

Price Distribution

- Calculated **minimum and maximum product prices** to understand the marketplace spread.
- Highlighted **high-priced products with high ratings** to show premium items well received by customers.

Brand Insights

- Determined **top brands by total rating count and average rating**, identifying leaders in customer engagement.

Category Insights

- Used **rating count as a proxy for sales** to determine **categories with the highest customer activity**.

Special Focus Queries

- Analyzed **price range within the Electronics & Headphones category**.
- Found **most expensive vs cheapest products for top brand by rating count** (executed as an advanced nested query).

5. Conclusion

This project demonstrates **end-to-end SQL skills**:

- **Schema transformation** (altering tables, adding numeric columns).
- **Data parsing using regex** to derive brand information.
- **Analytical SQL queries** with filtering, grouping, and subqueries.

The insights are directly **actionable for e-commerce companies**:

- Identify **best-performing brands** and **products worth promoting**.
- Understand **customer behavior via rating count and discount impact**.
- Spot **premium vs budget product positioning** to optimize sales strategies.

This project also shows you can handle **real-world messy datasets** and convert them into **clean, structured insights** for decision-making.