Amazon Sales Data Analysis

Using SQL

Introduction

The Amazon Sales Data Analysis project aims to derive meaningful insights from Amazon's product listing data. This project utilizes SQL queries to analyze key aspects such as product pricing, discount trends, customer reviews, and ratings. The analysis helps in understanding product demand, competitive pricing strategies, and customer sentiments, providing valuable insights for sellers and businesses.

Dataset Overview

The dataset, amazon_data, consists of various attributes that describe product details, pricing, and customer reviews. The key fields include:

- product_id: Unique identifier for each product.
- product_name: Name of the product.
- category: Category under which the product is listed.
- discounted_price: Price after discount.
- actual_price: Original price before discount.
- discount_percentage: Percentage of discount applied.
- rating: Average rating given by users.
- rating_count: Total number of ratings received.
- about_product: Additional product information.
- user_id: Unique identifier for each user.
- review_id: Unique identifier for each review.
- review_title: Title of the review.
- review_content: Content of the review.
- img_link: URL of the product image.
- product_link: URL of the product page on Amazon.

SQL Queries and Analysis

The following SQL queries were executed to analyze different aspects of Amazon sales data.

Basic Queries

1. List all products with a discounted price below ₹500

```sql
SELECT *
FROM amazon_data
WHERE discounted_price < 500;
```

This query identifies budget-friendly products available on Amazon, helping customers find affordable deals.

2. Find products with a discount percentage of 50% or more

```sql
SELECT *
FROM amazon_data
WHERE discount_percentage >= 0.5;
```

Identifies highly discounted products, which can be useful for tracking promotional trends.

3. Retrieve all products where the name contains the word "Cable."

```sql
SELECT *
FROM amazon_data
WHERE product_name LIKE '%Cable%';
```

This query extracts all products related to cables, useful for niche market analysis.

4. Display the difference between the average actual price and the discounted price for each product

```sql
SELECT product_id, product_name,
       (AVG(actual_price) - AVG(discounted_price)) AS avg_price_difference
FROM amazon_data
GROUP BY product_id, product_name;
```

5. Query reviews that mention "fast charging" in their content

```sql
SELECT *
FROM amazon_data
WHERE review_content LIKE '%fast charging%';
```

Analyzes user reviews to understand customer sentiment regarding fast-charging products.

6. Identify products with a discount percentage between 20% and 40%

```sql
SELECT *
FROM amazon_data
WHERE discount_percentage BETWEEN 0.2 AND 0.4;
```

Extracts products that offer moderate discounts.

7. Find products with an actual price above ₹1,000 and a rating of 4 stars or above

```sql
SELECT product_id, product_name, actual_price, rating
FROM amazon_data
WHERE actual_price > 1000 AND rating >= 4;
```

This query identifies premium, highly rated products.

8. Find products where the discounted price ends with a 9

```sql
SELECT *
FROM amazon_data
WHERE discounted_price LIKE '%9';
```

Extracts products that follow psychological pricing strategies (e.g., ₹999).

9. Display review contents containing words like "worst," "waste," "poor," or "not good"

```sql
SELECT product_name, review_content
FROM amazon_data
WHERE review_content LIKE '%worst%'
   OR review_content LIKE '%waste%'
   OR review_content LIKE '%poor%'
   OR review_content LIKE '%not good%';
```

Identifies negative reviews to help businesses improve their products.

10. List all products where the category includes "Accessories."

```sql
SELECT *
FROM amazon_data
WHERE category LIKE '%Accessories%';
```

Finds all accessories available in the dataset.

Advanced Queries

11. Identify the Top 5 Most Discounted Products in Each Category

```sql
WITH ProductDiscounts AS (
    SELECT
        product_id,
        category,
        discount_percentage,
        RANK() OVER (PARTITION BY category ORDER BY discount_percentage DESC) AS discount_rank
    FROM
        amazon_data
)
SELECT
    product_id,
    category,
    discount_percentage
FROM ProductDiscounts
WHERE discount_rank <= 5
ORDER BY category, discount_percentage DESC;
```

This query identifies the most heavily discounted products within each category.

12. Calculate the Percentage of 5-Star Reviews for Each Product Relative to All Reviews

```sql
WITH FiveStarReviews AS (
    SELECT
        product_id,
        COUNT(CASE WHEN rating = 5 THEN 1 END) AS five_star_count,
        COUNT(rating) AS total_review_count
    FROM amazon_data
    GROUP BY product_id
)
SELECT
    product_id,
    five_star_count,
    total_review_count,
    ROUND((five_star_count * 100.0 / total_review_count), 2) AS five_star_percentage
FROM FiveStarReviews
ORDER BY five_star_percentage DESC;
```

Calculates the proportion of 5-star reviews for each product, which helps in identifying highly rated items.

13. Calculate the Sales Contribution of Each Product to Its Category

```sql
WITH CategorySales AS (
    SELECT
        category,
        SUM(discounted_price) AS total_category_sales
    FROM amazon_data
    GROUP BY category
)
SELECT
    a.product_id,
    a.category,
    a.discounted_price AS product_sales,
    ROUND((a.discounted_price * 100.0 / cs.total_category_sales), 2) AS sales_contribution_percent
FROM amazon_data as a
JOIN CategorySales cs ON a.category = cs.category
ORDER BY
a.category, sales_contribution_percentage DESC;
```

Conclusion

This project provided a comprehensive analysis of Amazon sales data using SQL. The queries focused on extracting insights related to pricing strategies, discount trends, customer sentiments, and sales contributions. Advanced SQL techniques such as window functions, CTEs, and ranking were utilized to derive deeper insights.

These findings can help Amazon sellers optimize their pricing strategies, identify popular products, and enhance customer satisfaction based on review analysis.