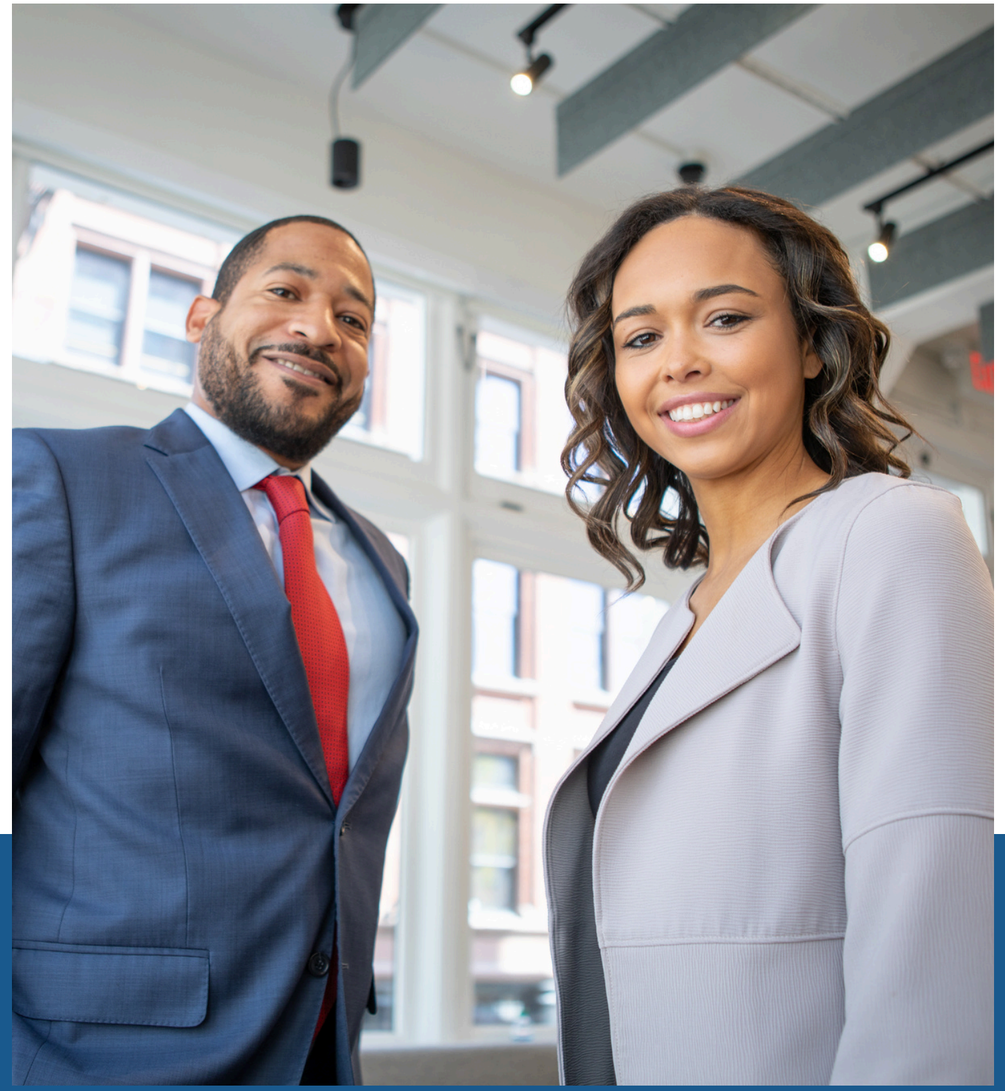
Amazon Sales Data Analysis

Subtitle: Insights Derived from Structured Queries

Presented By: Aditi Sharma

Date: 2024/12/12

Presented By : Aditi Sharma



Overview

Objective: Analyze Amazon sales data using SQL queries.

- Data Structure: The dataset contains the following columns:
- Product information: product_id, product_name, category, discounted_price, actual_price, discount_percentage, rating, rating_count, about_product
- Review details: user_id, user_name, review_id, review_title, review_content
- Links: img_link, product_link

Total Number of Products

- Query: SELECT COUNT(Product_id) AS total_products FROM amazon_products;
- Insight: Provides the total count of products in the dataset.

List All Products with Details

- Query: SELECT product_id, product_name, category FROM amazon_products;
- Insight: Lists all products with their ID, name, and category.

High-Rating Products

- Query: SELECT * FROM amazon_products WHERE CAST(rating AS DECIMAL(3,2)) >= 4.0;
- Insight: Displays all products with ratings of 4.0 and above.

Products in Specific Category

- Query: SELECT * FROM amazon_products WHERE category = 'Computers&Accessories';
- Insight: Lists all products in the "Computers & Accessories" category.

Products with Keyword in Description

- Query: SELECT * FROM amazon_products WHERE about_product LIKE '%durable%';
- Insight: Finds products where the description contains the word "durable."

Average Rating Query:

- Query: SELECT AVG(CAST(rating AS DECIMAL(3,2))) AS average_rating FROM amazon_products;
- Insight: Calculates the average rating of all products.

Top 5 Highest-Rated Products

- Query: SELECT product_id, product_name, rating FROM amazon_products ORDER BY CAST(rating AS DECIMAL(3,2)) DESC LIMIT 5;
- Insight: Identifies the top 5 products with the highest ratings.

Number of Reviews per Product

- Query: SELECT product_id, product_name, COUNT(review_id) AS review_count
  FROM amazon_products
  GROUP BY product_id, product_name;
- Insight: Displays the number of reviews for each product.

Products with Same Rating in Same Category

- Query: SELECT p1.product_id, p1.product_name, p1.category, p1.rating
  FROM amazon_products p1
  JOIN amazon_products p2
    ON p1.category = p2.category AND p1.rating = p2.rating
  AND p1.product_id != p2.product_id;

- Insight: Finds products in the same category with identical ratings.

Categorizing Products by Rating

- Query: SELECT product_id, product_name, rating,
  CASE
    WHEN CAST(rating AS DECIMAL(3,2)) >= 4.5 THEN 'Excellent'
    WHEN CAST(rating AS DECIMAL(3,2)) >= 4.0 THEN 'Good'
    ELSE 'Average'
  END AS rating_category
FROM amazon_products;

- Insight: Categorizes products into "Excellent," "Good," and "Average" based on ratings.

Add Column for Discount Amount

- Query: ALTER TABLE amazon_products ADD discount_amount DECIMAL(10, 2);
  UPDATE amazon_products
  SET discount_amount = CAST(actual_price AS DECIMAL(10,2)) - CAST(discounted_price AS DECIMAL(10,2));

- Insight: Adds a new column to calculate the difference between actual and discounted prices.

Category with Highest Average Rating

- Query: SELECT category, AVG_rating
FROM (
  SELECT category, AVG(CAST(rating AS DECIMAL(3,2))) AS AVG_rating
  FROM amazon_products
  GROUP BY category
) AS category_avg
ORDER BY AVG_rating DESC
LIMIT 1;
- Insight: Identifies the category with the highest average product rating.

Product Ranking Within Categories

- Query: SELECT product_id, product_name, category, rating,
  RANK() OVER (PARTITION BY category ORDER BY CAST(rating AS DECIMAL(3,2)) DESC) AS rating_rank
  FROM amazon_products;

- Insight: Ranks products within each category based on their rating.

Conclusion

- Key Insights:
  - Top-rated products and categories identified.
  - Products categorized by ratings.
  - Added new metrics for analysis, such as discount amount and cumulative counts.
- Tools Used: SQL queries, aggregate functions, and window functions.
- Next Steps:
  - Deeper analysis of review sentiments.
  - Explore correlations between discounts and ratings.

Thank You