Product Review and Recommendation System

Final Project Part 2

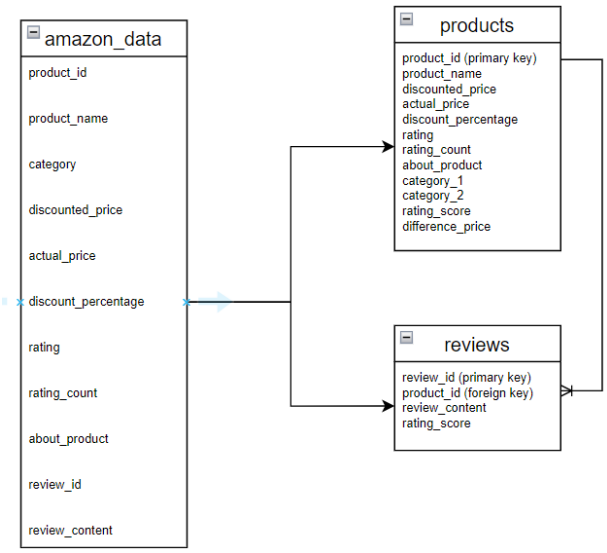
Team Name: ADT_Team_AAP

Team Members:

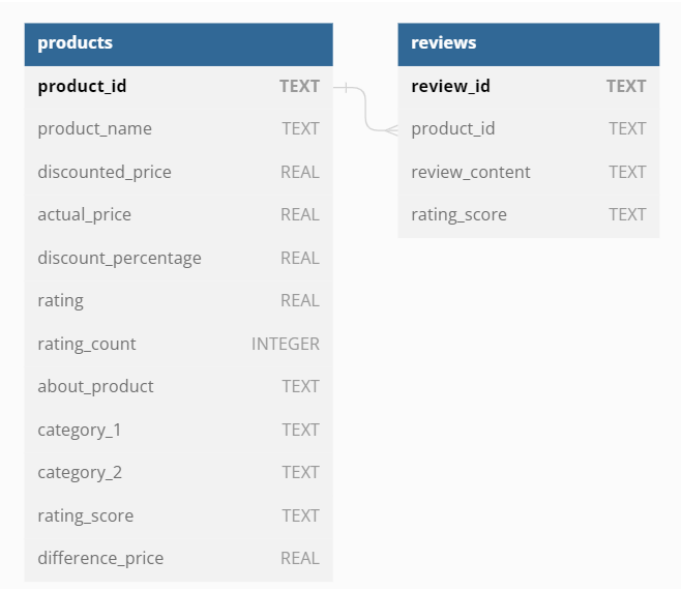
1. Priya Kumari (kumarip@iu.edu)
2. Ayesha Tajammul Ahmed Mulla (amulla@iu.edu)
3. Aazin Asif Shaikh (aazshaik@iu.edu)

Conceptual Diagram/Schema:

Conceptual Diagram:



Schema Diagram:



Explanation:

The conceptual diagram shows that we had a amazon_data table which was cleaned and normalized into two tables : products and reviews.

The schema diagram represents the two tables, "products" and "reviews", with their respective columns and relationships between them as follows:

1. Entity: "products" and "reviews" are entities in a our relational database. "products" table represents products and "reviews" table represents reviews.
2. Relationship Types:
   One-to-Many Relationship: The relationship between "products" and "reviews" is a one-to-many relationship. This means that each product in the "products" table can have multiple reviews in the "reviews" table, but each review in the "reviews" table is associated with only one product in the "products" table.
3. Objects:
   a. "products" table has the following objects:
      product_id (primary key): A unique identifier for each product.
      product_name: The name of the product.
      discounted_price: The discounted price of the product.
      actual_price: The actual/original price of the product.
      discount_percentage: The percentage of discount on the product.
      rating: The rating score of the product.
      rating_count: The number of ratings/reviews received by the product.
      about_product: Information about the product.
      category_1: The first category of the product.
      category_2: The second category of the product.
      rating_score: The rating score of the product (text type).
      difference_price: The difference between the discounted price and actual price of the product.

   b. "reviews" table has the following objects:

review_id (primary key): A unique identifier for each review.
product_id: The product ID to which the review belongs, acting as a foreign key that references the "product_id" column in the "products" table.
review_content: The content of the review.
rating_score: The rating score given in the review (text type).

Database:

Database Constraints:

The constraints which are present and defined for the tables in our project are as follows:

1. Uniqueness: Both tables have a primary key column defined as "product_id" in the "products" table and "review_id" in the "reviews" table. Primary keys are used to uniquely identify each row in a table and ensure that each row has a unique value for that column. There is a need to ensure the uniqueness of the data for maintaining data integrity and normalize our tables appropriately.
2. No Null Values: Columns other than primary keys in both tables are allowed to have null values, i.e., they can have missing or unknown values. While there is a need to handle null values appropriately in the application logic or database queries to avoid potential issues with data analysis or processing, our tables logically can have null values.
3. Data Types: Different data types are used to define columns, such as TEXT, REAL, and INTEGER.
    a. Fields with TEXT datatype:
        i. product_id (in products table)
        ii. product_name (in products table)
        iii. review_id (in reviews table)
        iv. review_content (in reviews table)
        v. rating_score (in products table and reviews table)
        vi. category_1 (in products table)
        vii. category_2 (in products table)
        viii. about_product (in products table)

    b. Fields with REAL datatype:
        i. discounted_price (in products table)
        ii. actual_price (in products table)
        iii. discount_percentage (in products table)
        iv. rating (in products table)
        v. difference_price (in products table)
    c. Fields with INTEGER datatype:
        i. rating_count (in products table)

    We have used the above mentioned datatypes to ensure that data being stored and the operations that will be performed on the data are consistent with its type.
4. Formatting: There is a ‘discount_percentage’ which has been kept upto two decimal places ie. as 0.64 to ensure that further calculations and count are performed with ease.
5. Relationships and foreign key: There is a relationship between the "products" and "reviews" tables established through the "product_id" column. The "product_id" column in the "reviews" table is a foreign key that references the "product_id" column in the "products" table. This establishes a one-to-many relationship where each product in the "products"

table can have multiple reviews in the "reviews" table. Proper use of foreign keys helps in maintaining data integrity, enforcing referential integrity constraints, and ensuring consistency in the relationships between tables.

Creating views, functions, and procedures:

We have created a few views in this part of the final project for each of our query such as:

1. product_count_by_category: Displays the number of products category and subcategory wise.
2. products_rating: Display the highest to lowest rated products.
3. product_discount: Displays the category of products giving highest to lowest discount percentage and corresponding rating
4. product_count_reviews: Displays the category of products having most reviews
5. rating_count_reviews: Displays the count and percentage of reviews recorded for each rating
6. category_discount_reviews: Displays the categorywise discount percentage offered and reviews written for corresponding category

In the subsequent milestones we plan to implement more such views and also incorporate functions and procedure for our project.

Assessment Table:

Priya Kumari	I'm satisfied with my performance and contribution to the team project. I actively participated in discussions and brainstorming sessions, contributing different ideas on database comparison selection,data cleaning,normalisation and queries to be used. I understood and shared resources and knowledge with my team members, which contributed to the timely completion of the project. However, I acknowledge the need to improve my skills in writing more advanced queries and understanding the data more thoroughly. Overall, I rate the task completion as a 9 out of 10, indicating a high level of satisfaction with the performance

Ayesha Tajammul Ahmed Mulla	Based on the project scope defined we have successfully been able to design the database within the stipulated amount of time and I am satisfied with my contribution towards the phase The collaborative teamwork helped in brainstorming ideas for the database design, meeting the deadlines through systematic and dynamic communication on regular basis using Zoom Meetings and In-person. I am satisfied with the task completion at a level of 9 out of 10 as we were able to deeply understand the data in the data cleaning process, performed normalization to get a better view

	of the different tables that will help in better visualization for our application. Few areas that could have been improvised were the data cleaning process that would lead to better normalization. Further, in the code section some additional queries can be incorporated for advanced data visualization. Overall, for this phase the requirements for database design have been successfully met with collaborative teamwork and individual efforts.

Aazin Asif Shaikh	We successfully designed the database within the given time frame, and I am content with my contribution towards the accomplishment of this project phase requirements. We worked together as a team to generate ideas and communicated regularly using Zoom and in-person meetings to meet the deadlines. I am mostly satisfied with our work as we were able to understand the data and perform normalization to visualize it better. Therefore, I am satisfied on a scale of 9 out of 10. However, we could have done better in the data cleaning process and in incorporating some advanced queries for data visualization in the code. Overall, we fulfilled the requirements for the database design with teamwork and individual efforts.

References:

1. https://learn.microsoft.com/en-us/sql/relational-databases/tables/primary-and-foreign-key-constraints?view=sql-server-ver16
2. https://www.sqlshack.com/
3. https://www.w3schools.com/sql/

Communication Channel: In-person meeting and Zoom Meeting

GitHub URL: https://github.iu.edu/kumarip/ADT_amulla_kumarip_aazshaik