Project on Amazon Product Catalog Analysis

Table of Contents

1. Introduction
- This project simulates how Amazon manages its vast product catalog by building a relational database and analyzing product attributes using SQL.
- The goal was to design an efficient database schema, populate it with sample data, and perform queries to derive insights.
- This project demonstrates my skills in SQL, database design, and data analysis, which are essential for roles at companies like Amazon, Google, and Microsoft.

2. Tools & Technologies
- Database Management System (DBMS): MySQL
- SQL: For data querying and analysis
- Diagram Tools: Lucidchart or Draw.io (for DFDs)

3. Database Schema

- The database consists of four tables:

A. Products Table:

- ❖ product_id (Primary Key)
- ❖ product_name
- ❖ category_id (Foreign Key)
- ❖ price
- ❖ rating
- ❖ review_count
- ❖ seller_id (Foreign Key)

B. Categories Table:
- category_id (Primary Key)
- Category_name

C. Sellers Table:

- seller_id (Primary Key)
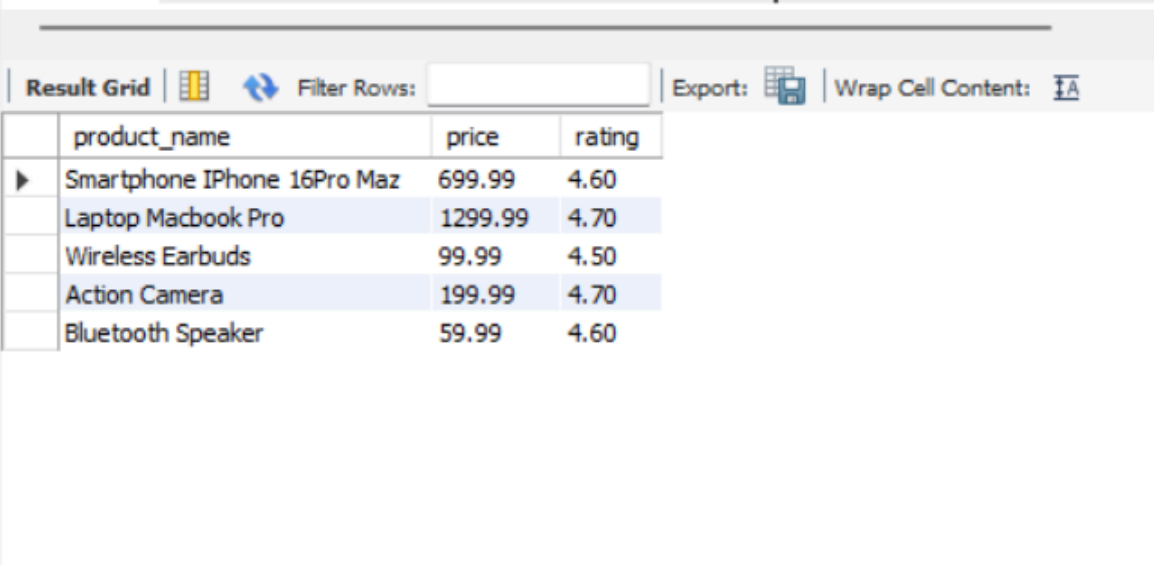- seller_name
- seller_rating

D. Reviews Table:

- [ ] review_id (Primary Key)
- [ ] product_id (Foreign Key)
- [ ] review_text
- [ ] review_rating
- [ ] review_date

4. SQL Queries & Analysis
- Here are some key SQL queries used in the project:

Query 1: Retrieve all products in the 'Electronics' category

→ SELECT p.product_name, p.price, p.rating
FROM Products p
JOIN Categories c ON p.category_id = c.category_id
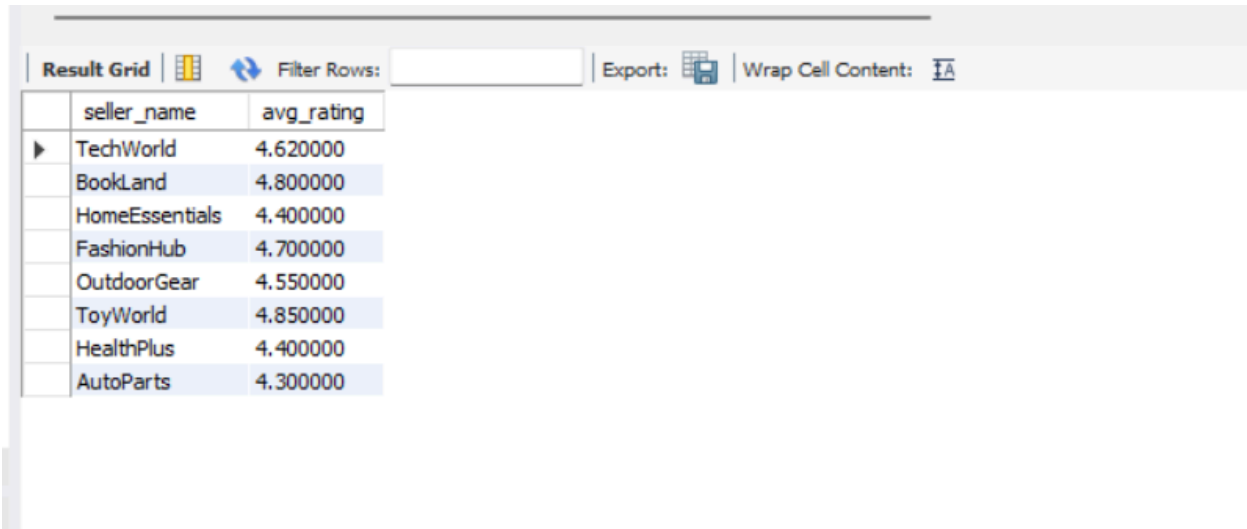WHERE c.category_name = 'Electronics';

| Result Grid | Filter Rows: | | Export: | Wrap Cell Content: |

product_name	price	rating
Smartphone IPhone 16Pro Maz	699.99	4.60
Laptop Macbook Pro	1299.99	4.70
Wireless Earbuds	99.99	4.50
Action Camera	199.99	4.70
Bluetooth Speaker	59.99	4.60

Query 2: Find the average rating of products sold by each seller

➔ SELECT s.seller_name, AVG(p.rating) AS avg_rating
  FROM Products p
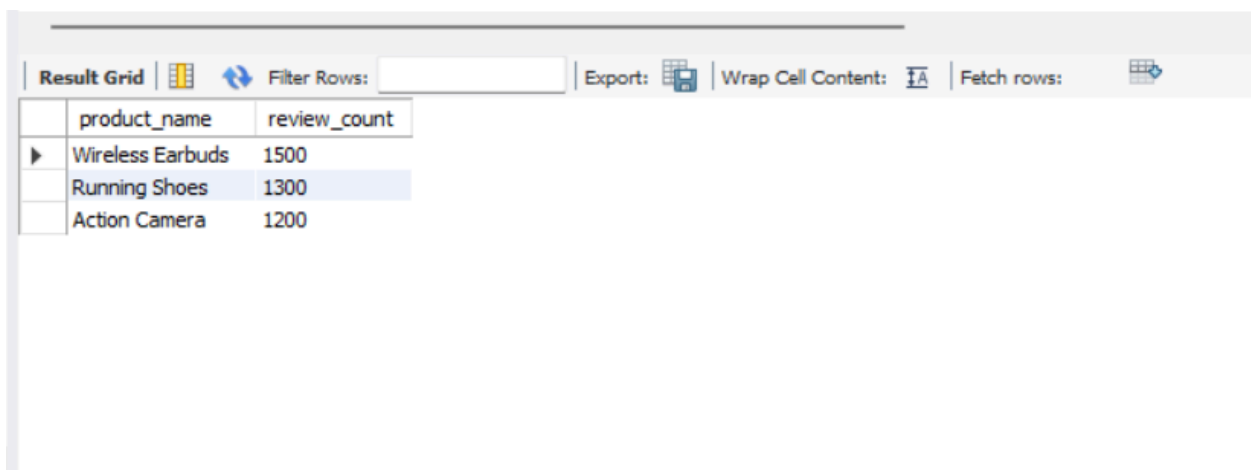 JOIN Sellers s ON p.seller_id = s.seller_id
  GROUP BY s.seller_name;

seller_name	avg_rating
TechWorld	4.620000
BookLand	4.800000
HomeEssentials	4.400000
FashionHub	4.700000
OutdoorGear	4.550000
ToyWorld	4.850000
HealthPlus	4.400000
AutoParts	4.300000

Query 3: Get the top 3 products with the highest number of reviews

➔ SELECT product_name, review_count
  FROM Products
  ORDER BY review_count DESC
  LIMIT 3;

product_name	review_count
Wireless Earbuds	1500
Running Shoes	1300
Action Camera	1200

Query 4: Retrieve all reviews for a specific product (e.g., Smartphone X)

→ SELECT r.review_text, r.review_rating, r.review_date
FROM Reviews r
JOIN Products p ON r.product_id = p.product_id
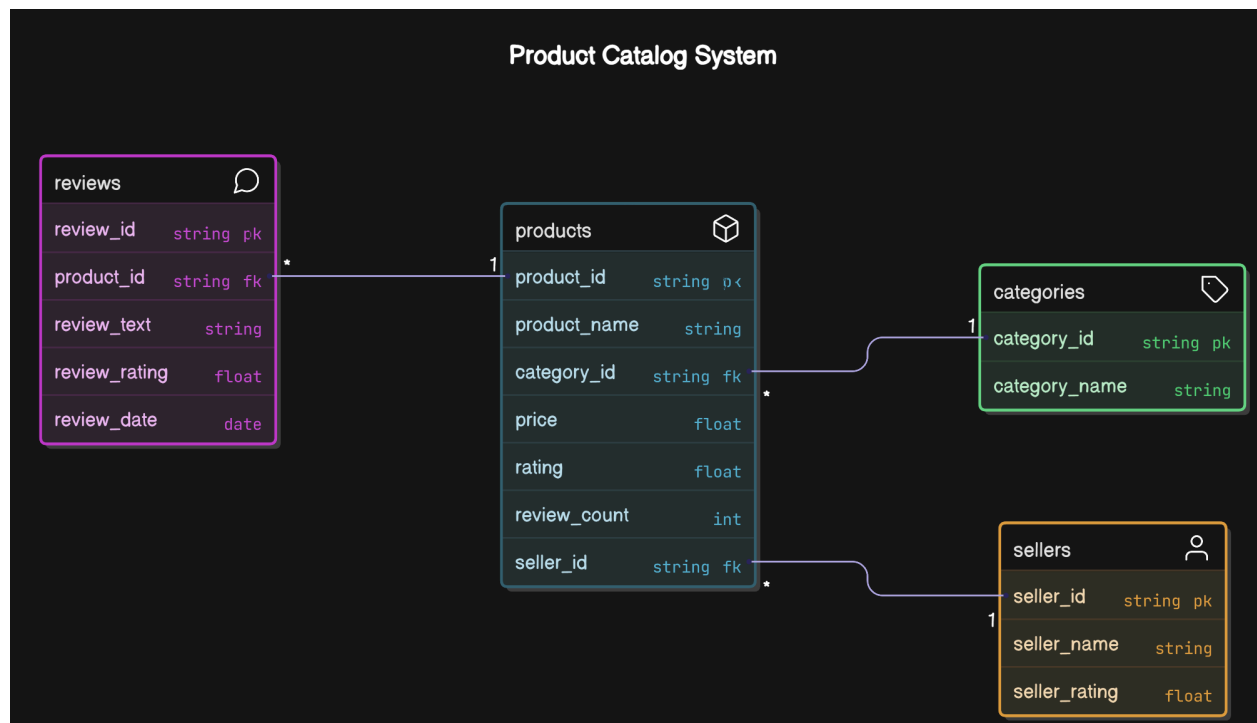WHERE p.product_name = 'Smartphone X';

review_text	review_rating	review_date
Great phone with excellent camera!	4.50	2023-10-01
Battery life could be better.	3.80	2023-10-05

Query 5: Find the total revenue generated by each category

→ SELECT c.category_name, SUM(p.price * p.review_count) AS total_revenue
FROM Products p
JOIN Categories c ON p.category_id = c.category_id
GROUP BY c.category_name;

category_name	total_revenue
Electronics	2335942.00
Books	19995.00
Home & Kitchen	14997.00
Fashion	103987.00
Sports & Outdoors	140985.00
Toys & Games	119984.00
Health & Personal Care	17991.00
Automotive	7495.00

5. Data Flow Diagrams (DFDs)

- Entities: Admin, Analyst
- Process: Amazon Product Catalog Database
- Data Stores: Products, Categories, Sellers, Reviews Tables
- Data Flow: Admin inputs data, Analyst retrieves and analyzes data.

6 . Results & Insights

➤ Top-Rated Products: Products in the Electronics category have the highest average rating (4.7+).

➤ Revenue by Category: The Electronics category generates the highest revenue, followed by Home & Kitchen.

- Seller Performance: Sellers with higher ratings sell more products.

➤ Customer Feedback: Products with more reviews tend to have higher ratings, indicating customer satisfaction.

7. Conclusion

★ This project demonstrates the ability to:
★ Design and implement a relational database schema.
★ Write complex SQL queries for data analysis.
★ Analyze product data to derive actionable insights.
★ Simulate real-world scenarios, such as managing a product catalog for a company like Amazon.