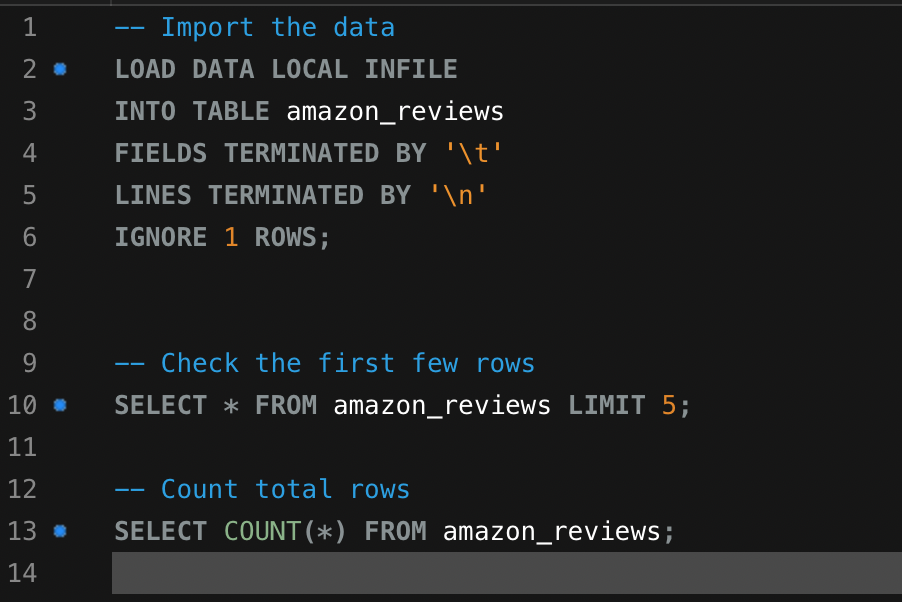
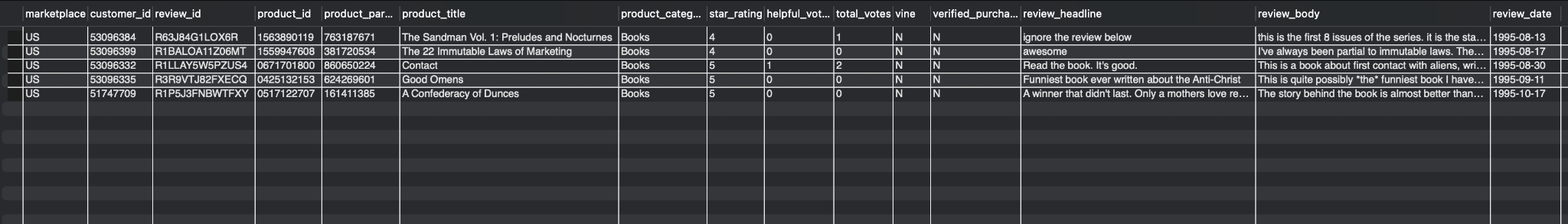
The dataset is imported into a table called amazon\_reviews using the LOAD DATA LOCAL INFILE command.

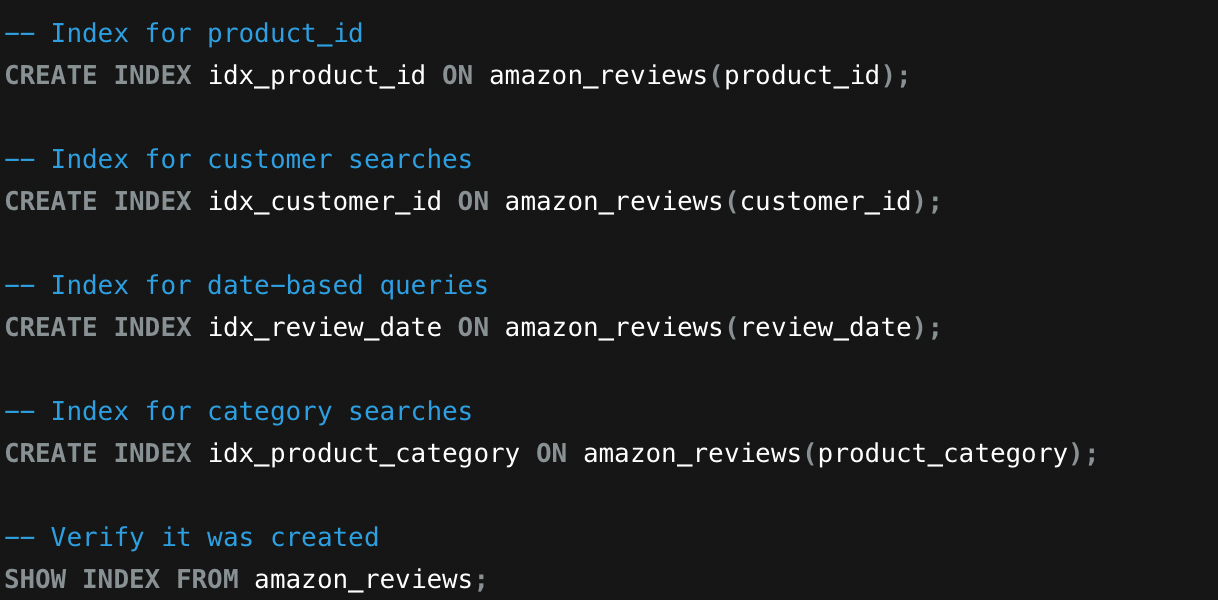


After importing the dataset, the table looks like this:



Indexes are created to optimize query performance on commonly used columns. For this dataset, the following indexes were added:

* product\_id for product-specific queries.
* customer\_id for customer-specific searches.
* review\_date for time-based queries.
* product\_category for category-based searches.

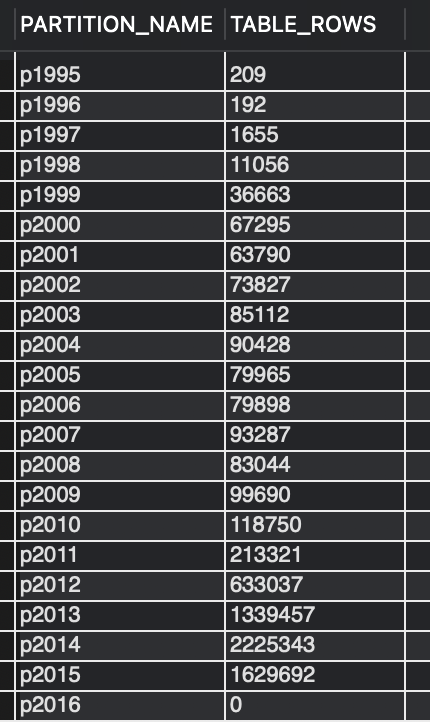


Indexes are added to make searching and retrieving data faster, especially for frequently used columns.

The table is also partitioned by year using the review\_date column. Each partition contains data for reviews from a specific year, ensuring efficient querying and storage. A range partitioning scheme is implemented to create partitions for every year from 1995 to 2023, with an additional partition for future dates.

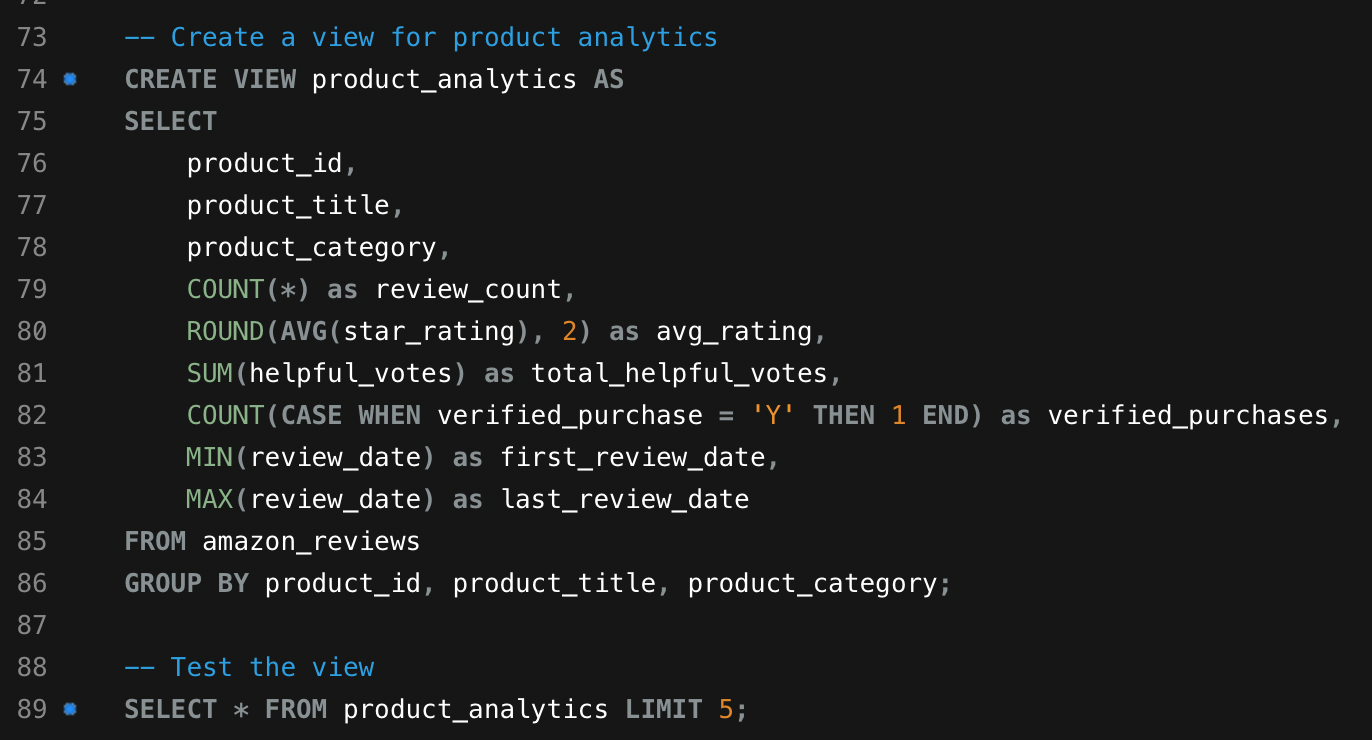


The table is split into smaller parts by year to improve query speed and manage large amounts of data efficiently.

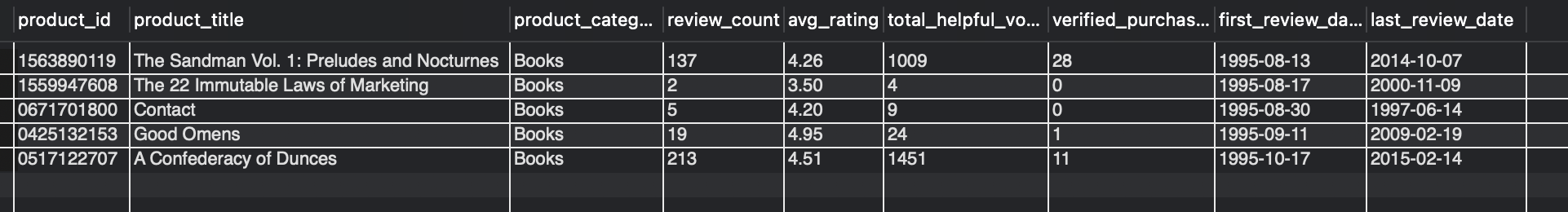


A view named product\_analytics is also created to provide aggregated insights into products. This view includes:

* Total number of reviews (review\_count).
* Average star rating (avg\_rating).
* Total helpful votes (total\_helpful\_votes).
* Number of verified purchases (verified\_purchases).
* Earliest and latest review dates (first\_review\_date, last\_review\_date).
* This view simplifies the analysis of product-level data.

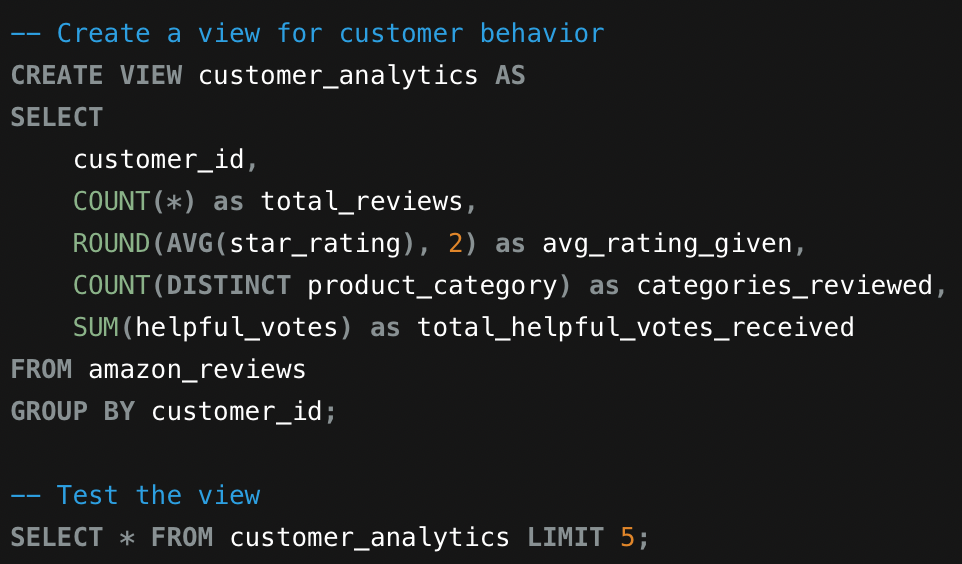


We create a summary of product data to easily analyze things like average ratings, total reviews, and helpful votes for each product.

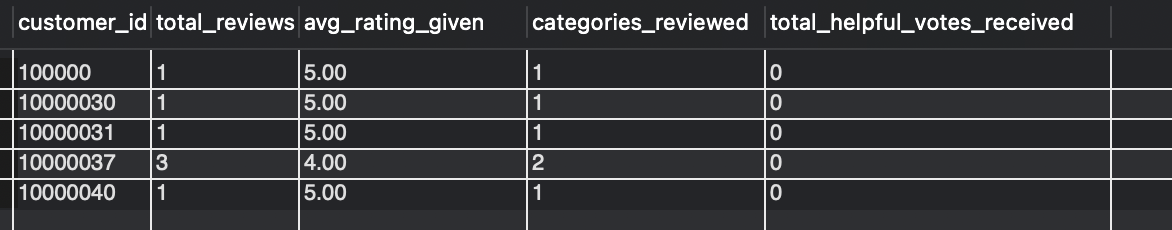


A view named customer\_analytics is also created to analyze customer activity. This view includes:

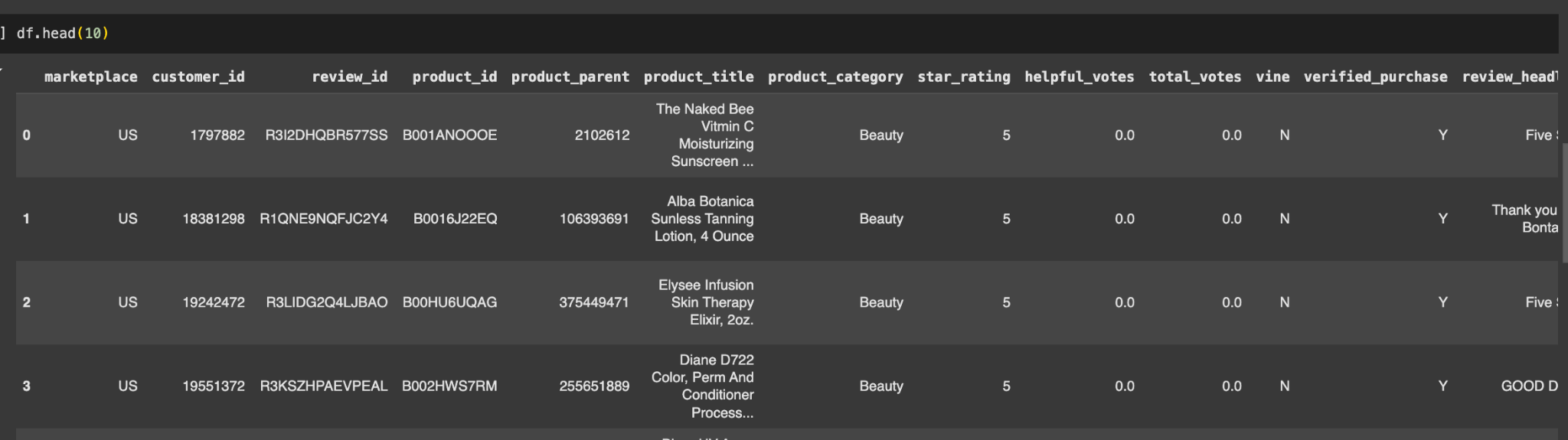
* Total reviews written by each customer (total\_reviews).
* Average rating given by each customer (avg\_rating\_given).
* Number of product categories reviewed (categories\_reviewed).
* Total helpful votes received for their reviews (total\_helpful\_votes\_received).

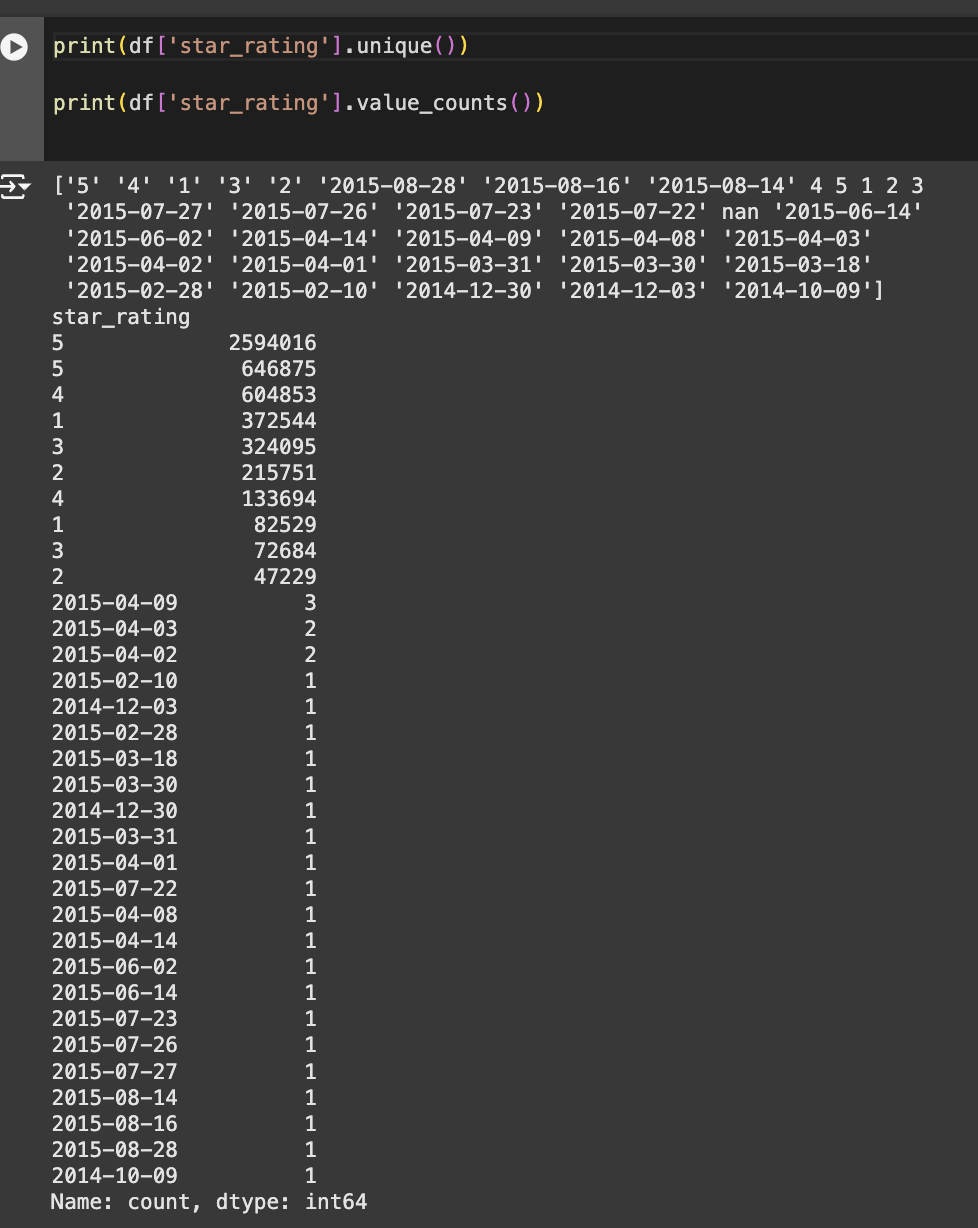


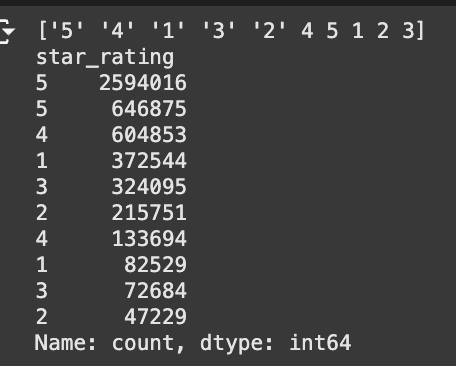
This provides insights into individual customer behavior. We summarize customer activity to understand their review habits, such as how many reviews they’ve written and their average ratings.



preprocess the data







### 

### **Recommended System Output Output**

The recommendation system was successfully implemented and tested using the SVD (Singular Value Decomposition) algorithm on a sampled subset of our extensive product review dataset. After processing the data and training the model, recommendations were generated for a selected user based on their historical rating data and the predictive modeling of user preferences.

