# ****Analysis of Product Ratings: Trends and Insights****

### ****1. Introduction****

#### ****Project Overview****

In this project, I conducted an in-depth analysis of product ratings to uncover key trends and insights. The primary objective was to analyze user ratings over time, identify patterns in user behavior, and explore seasonal trends. This project was executed as a portfolio piece, demonstrating my ability to work with large datasets, perform advanced SQL queries, and create insightful visualizations in R.

#### ****Dataset Description****

The dataset used in this analysis consists of 4.2 million records with four key fields:

* **userId:** Unique identifier for each user.
* **productId:** Unique identifier for each product.
* **Rating:** Numeric rating provided by the user.
* **timestamp:** Unix timestamp representing when the rating was given.

### ****2. Data Preparation****

#### ****Data Cleaning****

Before diving into the analysis, I ensured the data was clean and ready for processing. Key steps included:

* **Handling Missing Values:** Checked for and addressed any missing data in the dataset.
* **Timestamp Conversion:** Converted Unix timestamps to readable date formats to facilitate time-based analysis.
* **Filtering Data:** Removed any outliers or irrelevant records to ensure the integrity of the analysis.

#### ****Data Transformation****

To enhance the analysis, I performed several data transformations:

* **Feature Engineering:** Created new features such as the number of ratings per product and the time interval between ratings.

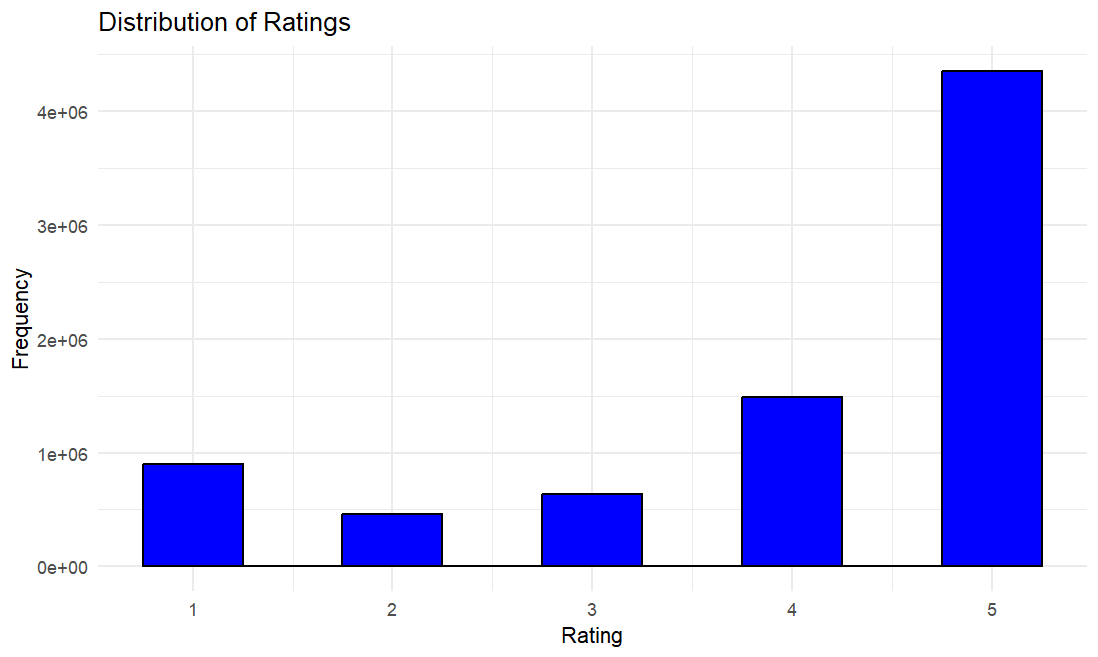
### ****3. Exploratory Data Analysis (EDA)****

#### ****Descriptive Statistics****

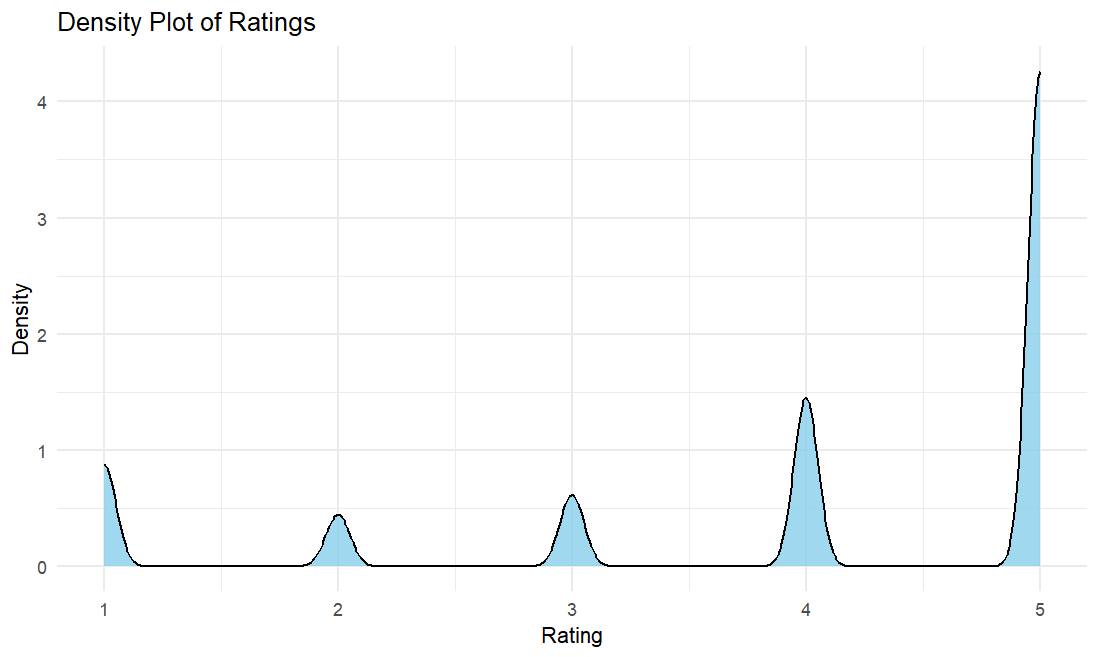
* **Average Rating:** The overall average rating across all products is 4.012.
* **Highest and Lowest Average Ratings:**
  + Highest: 5.00
  + Lowest: 1.00
* **Ratings per Product:** The dataset includes products with as few as 1 rating to those with over 18,244 ratings.

#### ****Visualization****

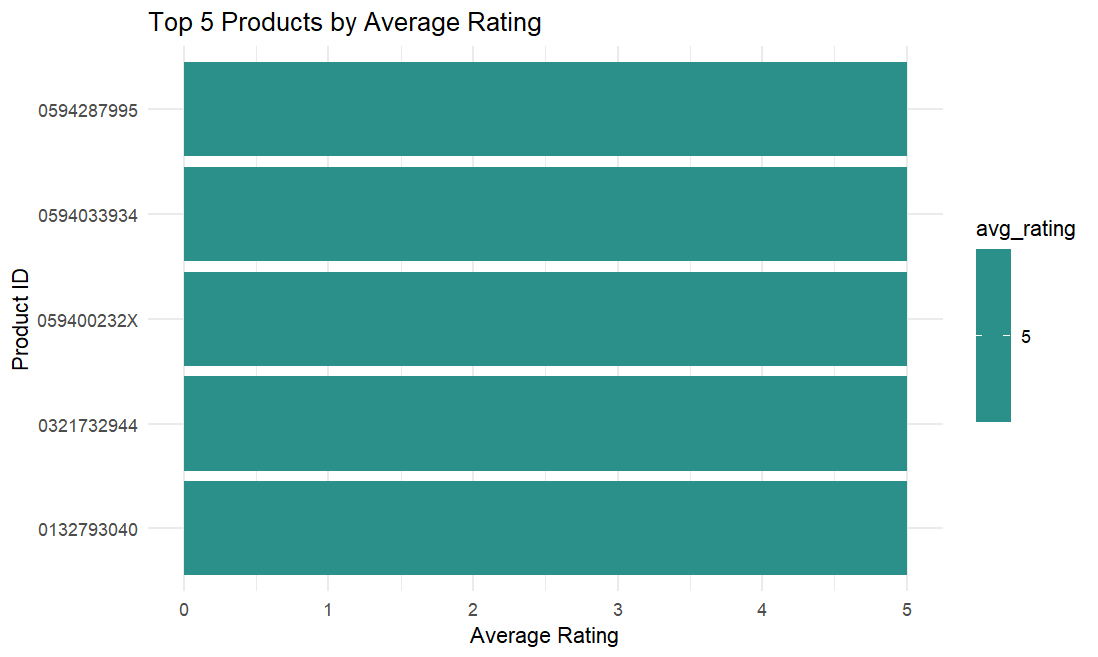
* **Histogram of Ratings:**



* **Density Plot:**



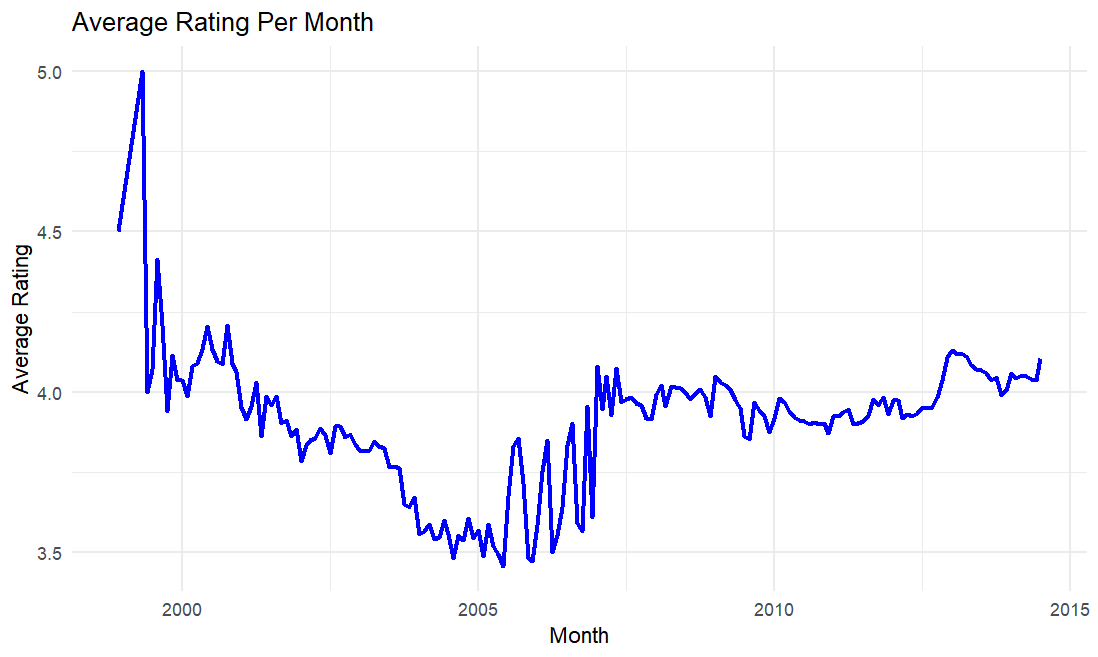
* **Bar Chart of Top-Rated Products:**



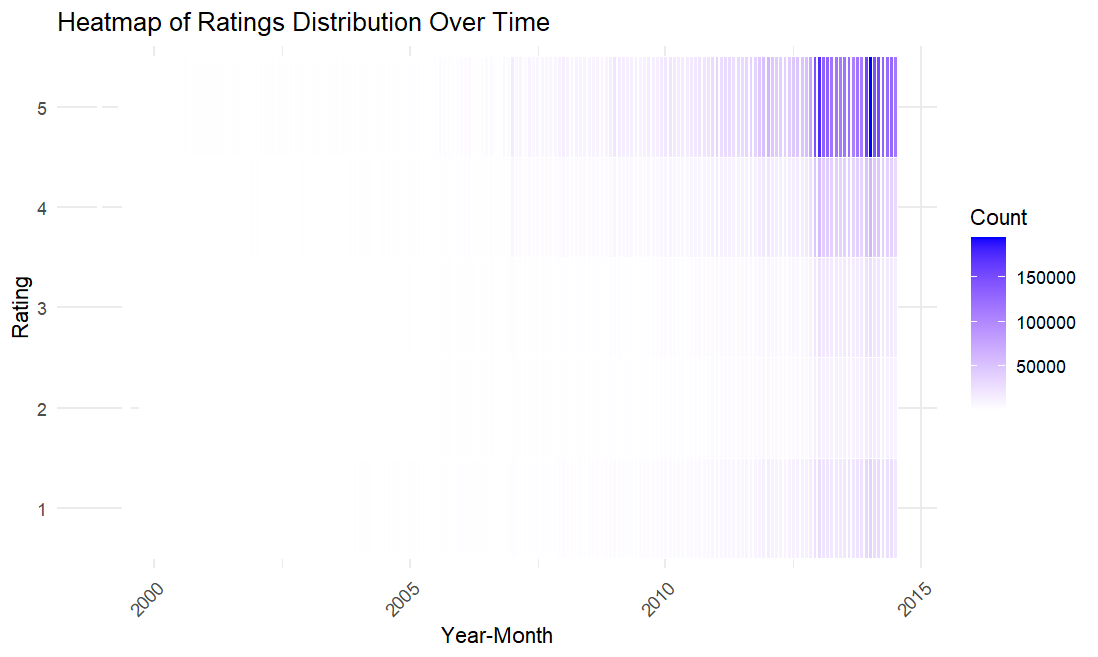
### ****4. Advanced Data Analysis****

#### ****Trend Analysis****

* **Average Ratings Over Time for Products:**

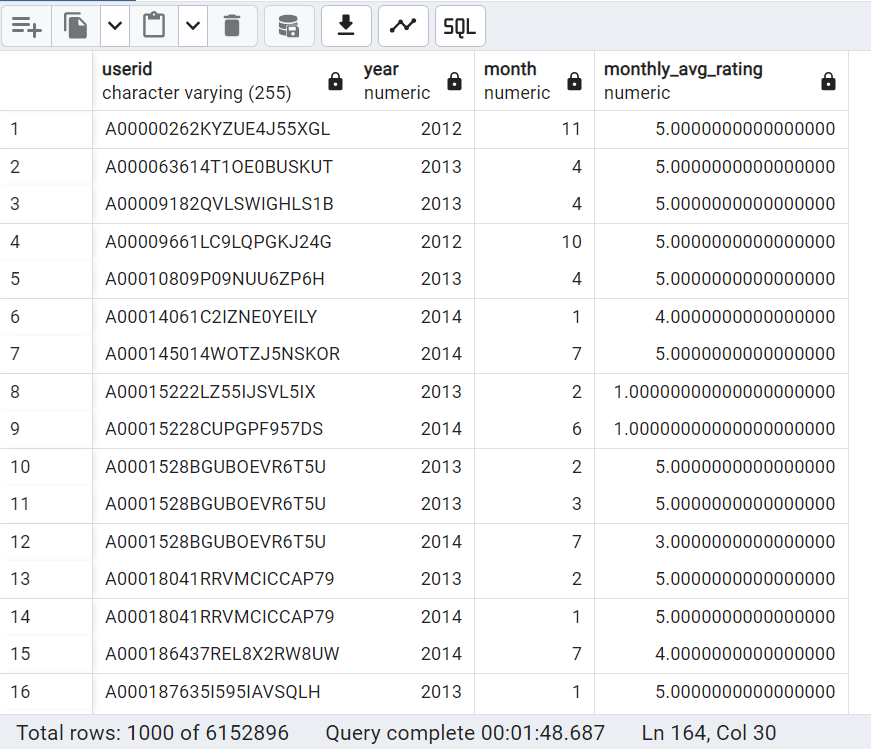


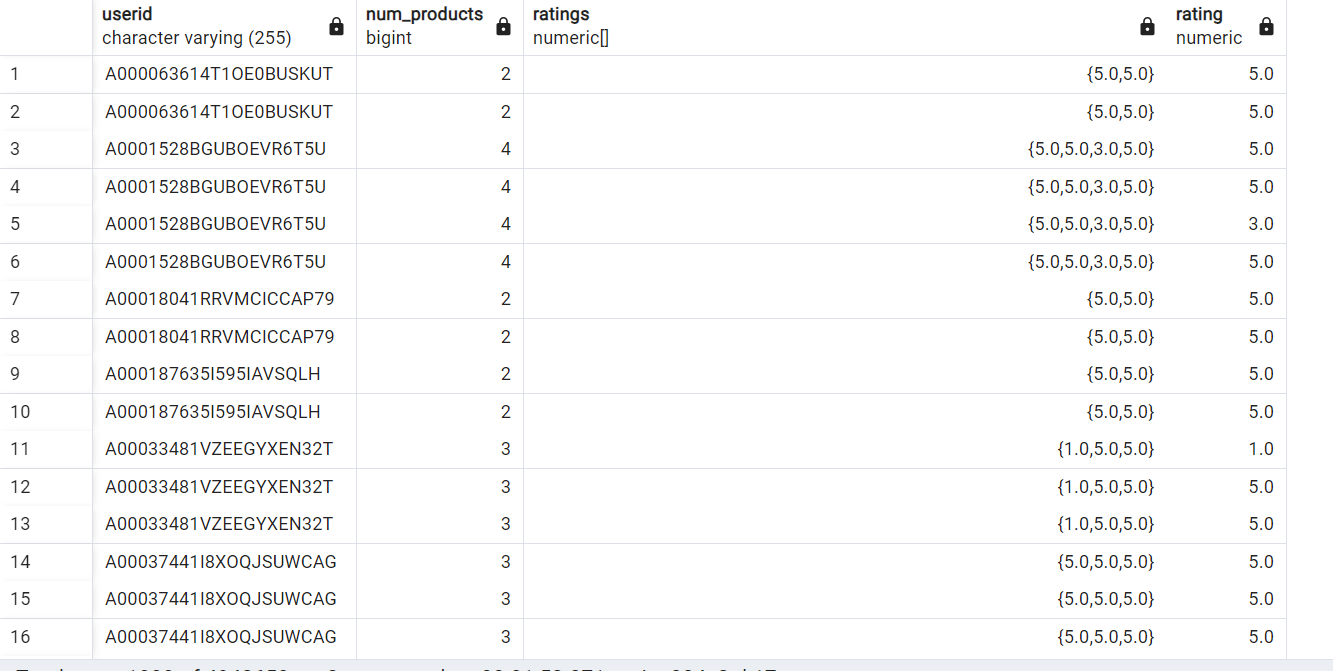
* **Seasonal Analysis**



#### ****User Behavior****

* **User Rating Patterns:** Analyzed changes in user ratings over time.

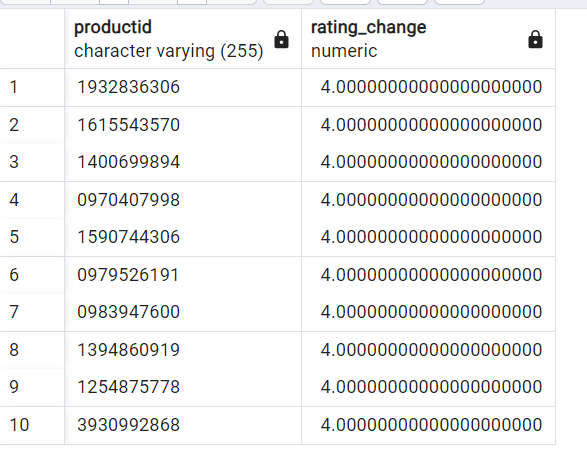


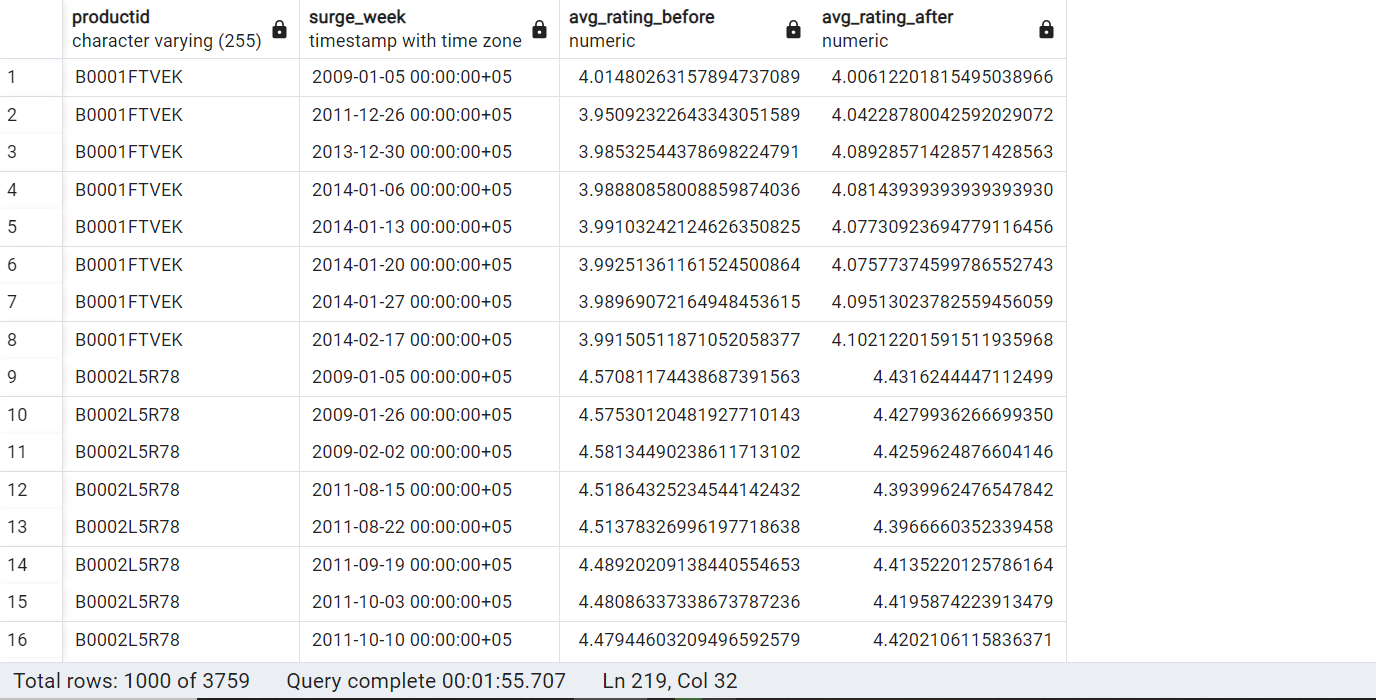


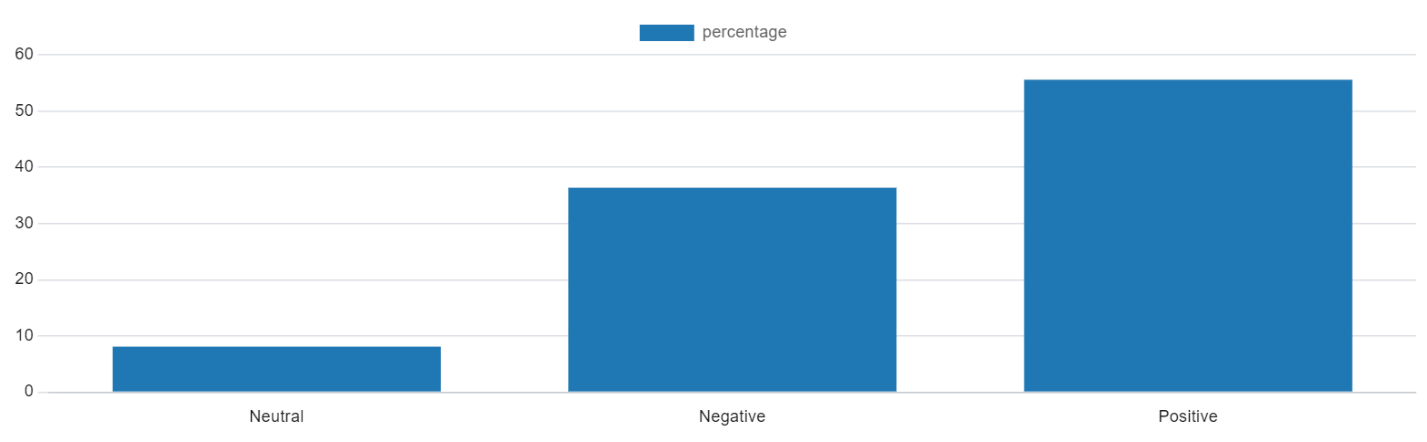
### ****5. Key Findings****

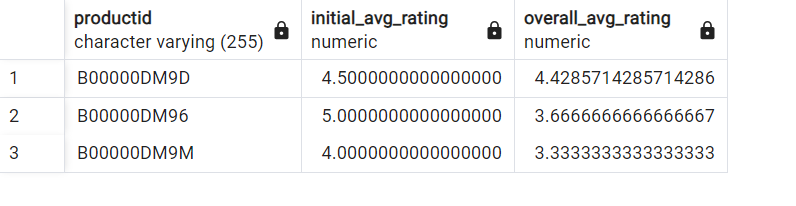
#### ****Rating Trends****

* **Product Popularity:** Products with consistently high ratings over time tend to attract more users and maintain their rating.







* **Impact of Early Ratings:** Early ratings significantly influence a product's long-term average rating.

#### ****Seasonality Insights****

* Clear seasonal patterns were not observed.

### ****6. Challenges and Solutions****

#### ****Challenges Encountered****

* **Large Dataset Handling:** The dataset's size posed a challenge in processing and analyzing it efficiently. This was addressed by optimizing SQL queries and using efficient data structures in R.
* **Timestamp Conversion:** Converting Unix timestamps into readable formats required careful handling to ensure accuracy in time-based analysis.

#### ****Solutions Implemented****

* **Optimized Queries:** Complex SQL queries were optimized for performance to handle the large dataset without compromising speed or accuracy.
* **Visualization Enhancements:** The visualizations were refined to effectively communicate the findings, using clear and concise charts that highlight key trends.

### ****7. Conclusion****

#### ****Summary of Insights****

This project provided valuable insights into how product ratings evolve over time, the impact of early ratings, and seasonal trends. The findings suggest that early positive ratings can significantly influence a product's long-term success, and there are discernible patterns in how users rate products over time.

#### ****Future Work****

For future analysis, incorporating additional data fields like product categories or reviews would allow for more granular insights. Additionally, applying machine learning techniques could help predict product success based on early ratings.

#### ****Personal Reflection****

This project enhanced my skills in handling large datasets, performing advanced SQL queries, and creating meaningful visualizations in R. It demonstrated my ability to derive actionable insights from data, which is essential for making data-driven decisions in a business context.

### ****8. Appendices****

#### ****R for Visualizations****

#### ****TIME SERIES OF AVERAGE WEEKLY RATINGS****

#### 

#### ****Trends for Top-Products****

#### 

#### AVERAGE RATINGS OVER TIME

