

BY SEDRA BAIAN

# AMAZON PRODUCT REVIEWS ANALYSIS: DATA CLEANING & INSIGHTS

Welcome! In this project, I worked with Amazon product review data. My goal was to clean the dataset, ensure its integrity, and extract meaningful insights about product ratings and customer behavior.

# ABOUT THE DATASET

## CONTENT:

- Source: Kaggle – Amazon Product Reviews Dataset
- Columns: User ID, Product ID, Rating, Timestamp
- Total Records: 5000

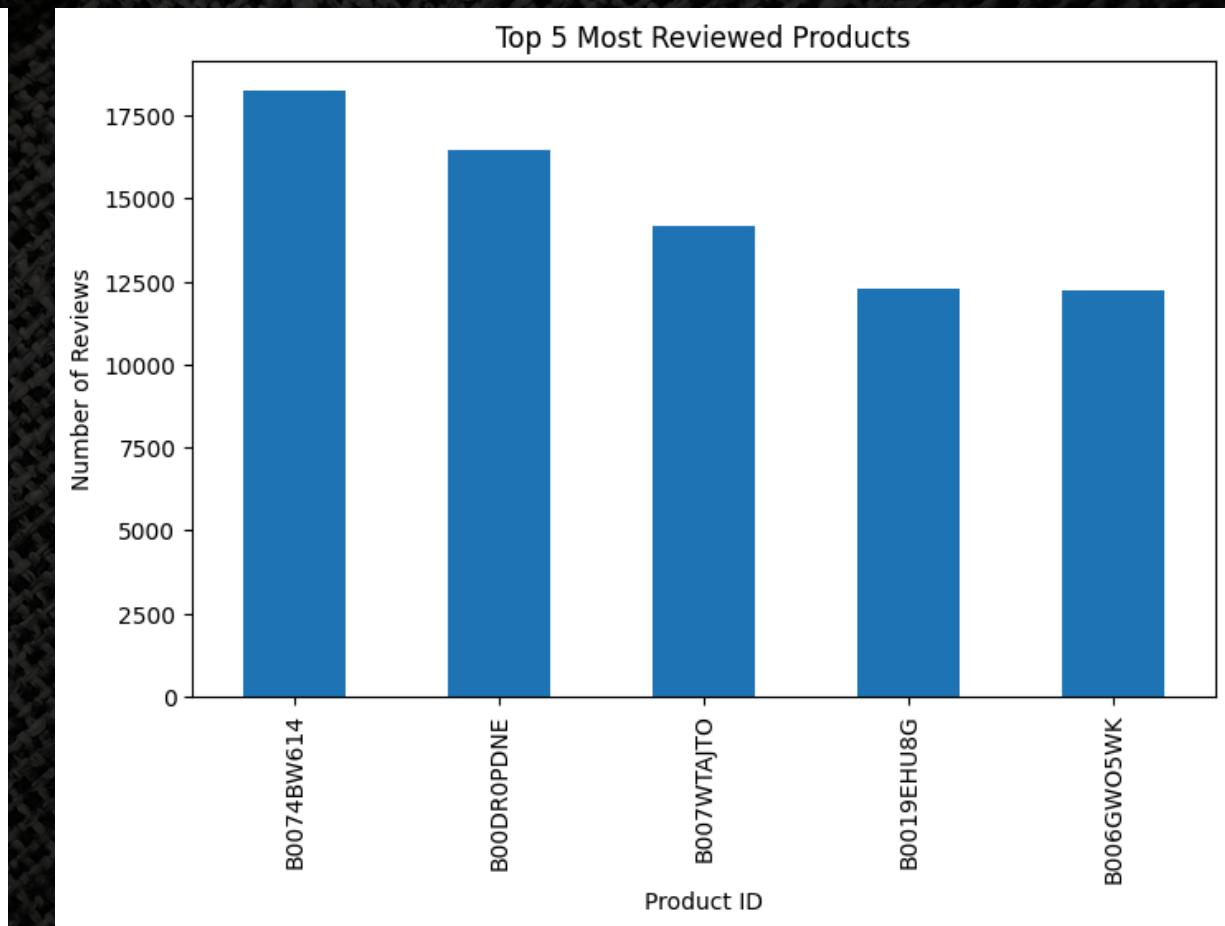
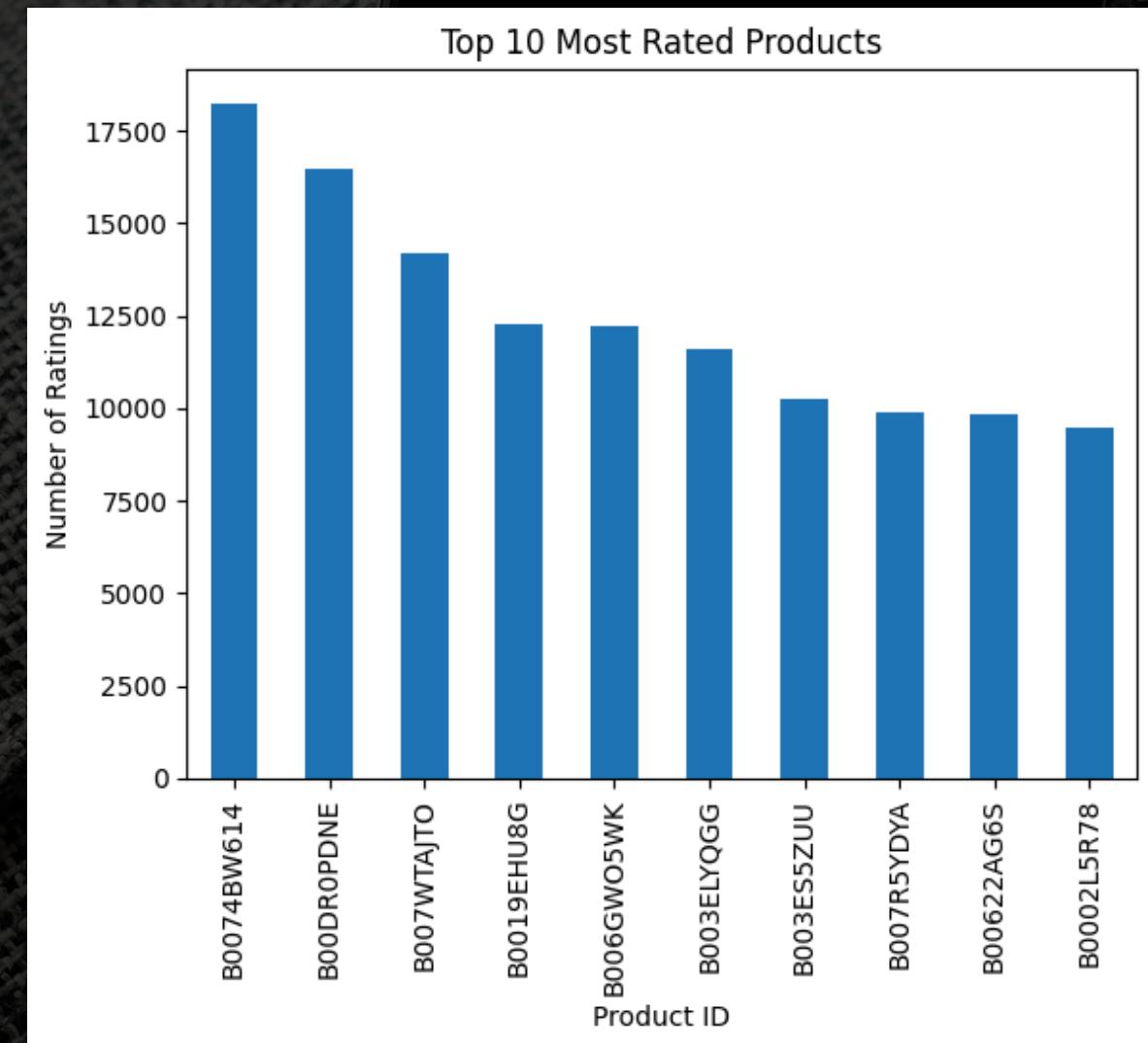
## CLEANING STEPS:

- Verified no missing values or duplicates
- Converted ratings to integer format
- Transformed UNIX timestamps to readable dates

I started with a large dataset containing Amazon product reviews. Each entry included the user, product, rating, and timestamp. Although the data was already quite clean, I validated it to confirm there were no missing values or duplicates. I also formatted the data for readability, converting ratings to integers and timestamps to actual dates.

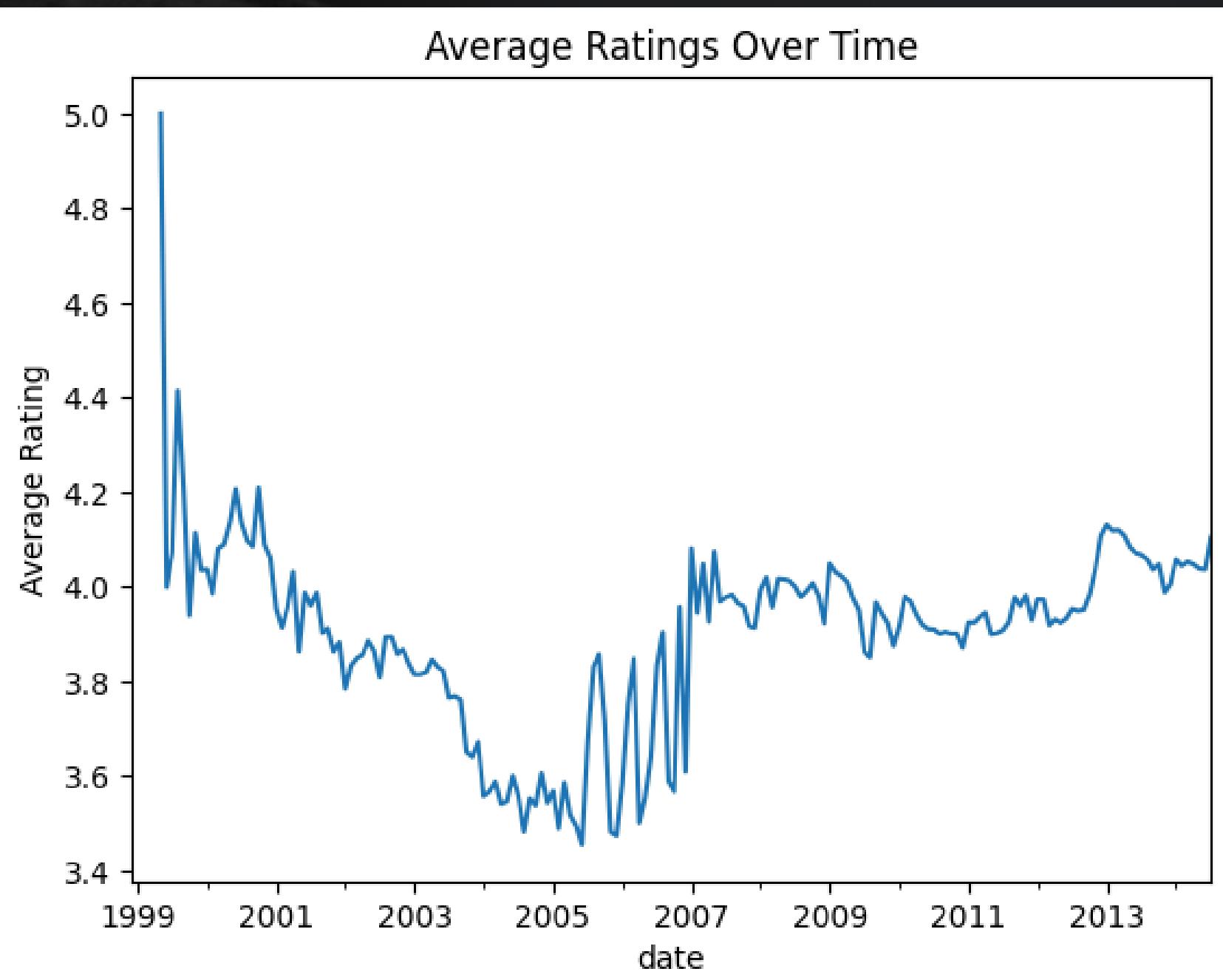
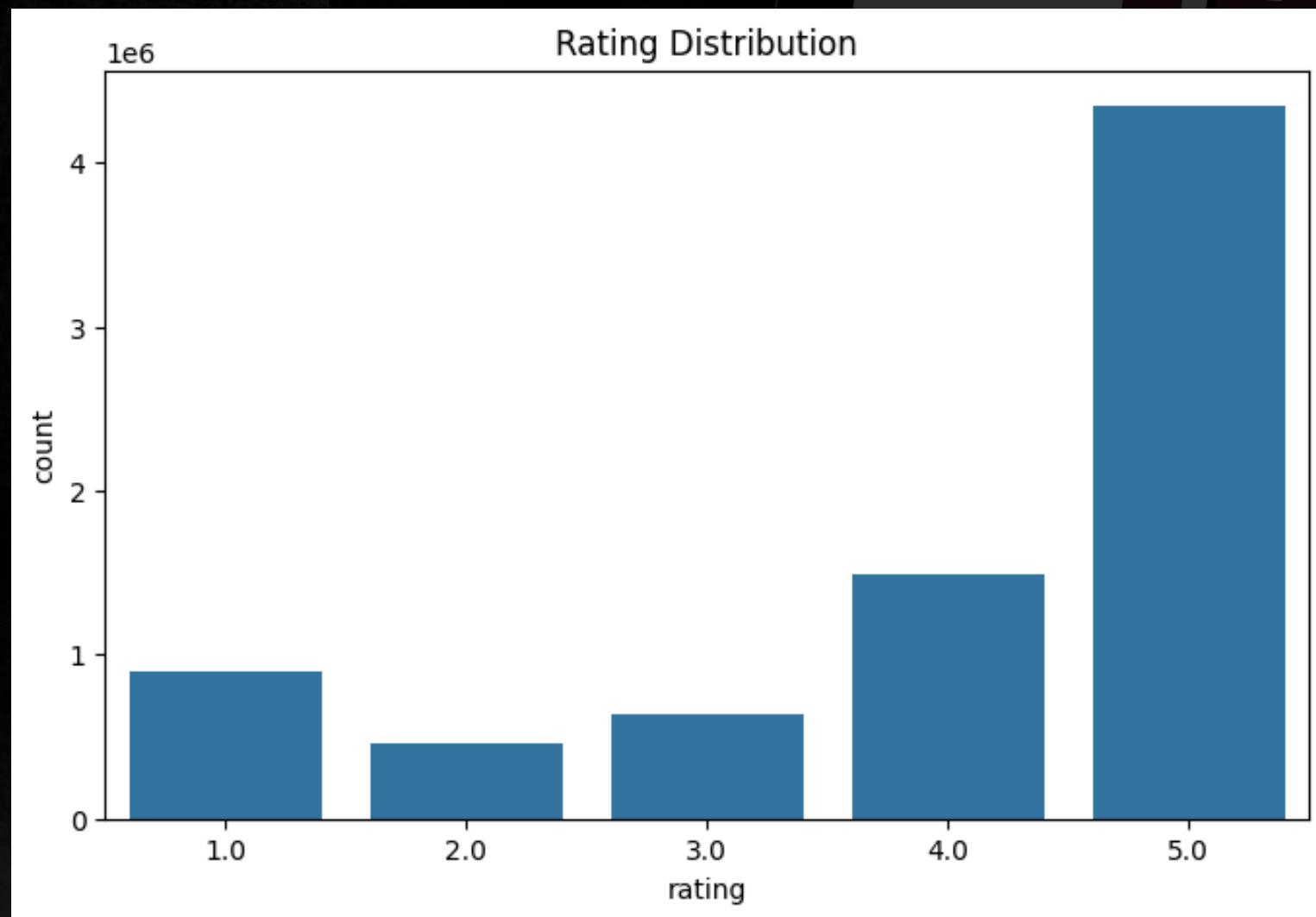
# MOST RATED & MOST REVIEWED PRODUCTS

- Here's a look at the most popular products in the dataset.
- On the left, we see the top 10 most rated products overall.
- On the right, I zoomed in to highlight the top 5 most reviewed products.
- Certain products appear consistently in both charts.
- This shows strong customer engagement and popularity for those products.



# RATING DISTRIBUTION & TRENDS

- On the left, the chart shows that 5-star ratings clearly dominate, reflecting strong overall customer satisfaction.
- On the right, the line graph illustrates how the average ratings have evolved over time.
- While there are some fluctuations, the trend remains consistently positive, with ratings stabilizing around 4 stars in recent years.



# KEY TAKEAWAYS & NEXT STEPS



INDUSTRY BACKGROUND

## KEY TAKEAWAYS:

- The dataset was already clean, but I validated it by checking for missing values and duplicates.
- The majority of product ratings are **highly positive**, with **5-star ratings** clearly dominating.
- Several products consistently appeared as the **most reviewed and most rated**, reflecting strong customer engagement.
- Average ratings remained **stable** around 4 stars over time, despite some fluctuations.

## NEXT STEPS & FUTURE IDEAS:

- Develop a product recommendation system based on customer ratings and review patterns.
- Apply **sentiment analysis** to customer review texts (if available) for deeper insights.
- Explore **time-based trends** further to identify potential seasonality or shifts in product popularity.

GROUND

THE INDUSTRY'S HISTORY

# THANK YOU

FOR YOUR ATTENTION

