

Analysis & Visualization of Product Performance on Amazon Platform

Katarzyna Kopania, Malwina Schonhofer, Juliusz Szymajda

November 29, 2024

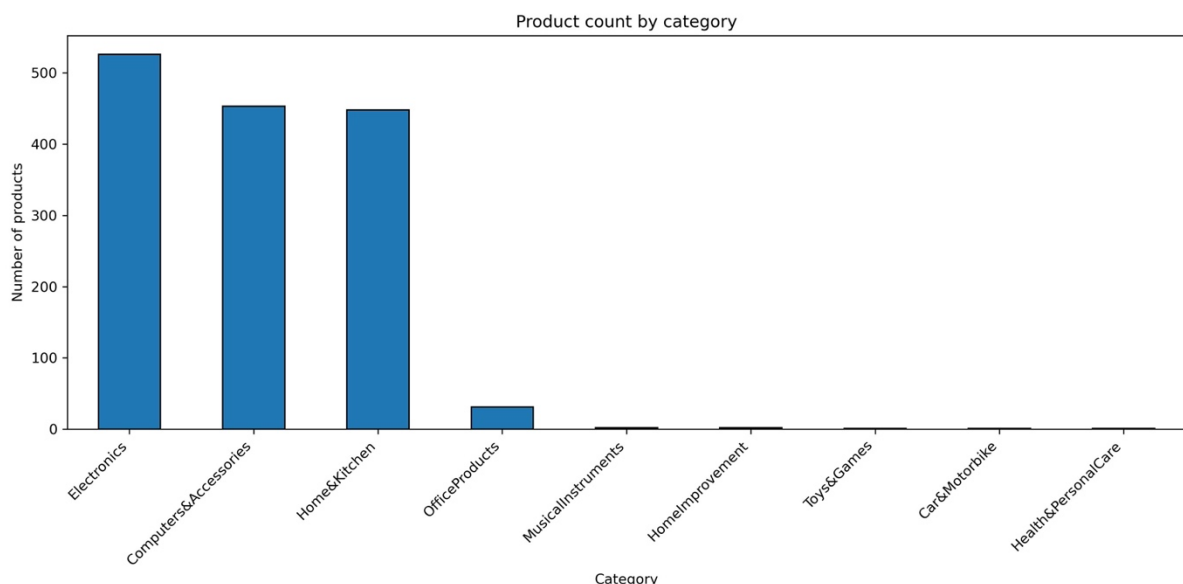
1) Introduction

The core objective of this analysis is to visualize and evaluate sales data of product performance on Amazon platform. We aim to investigate provided dataset on various angles including: general product information, price fluctuations, rating fluctuations, correlations and top ranking products.

2) Methodology

2.1) Data Preparation & Product Count

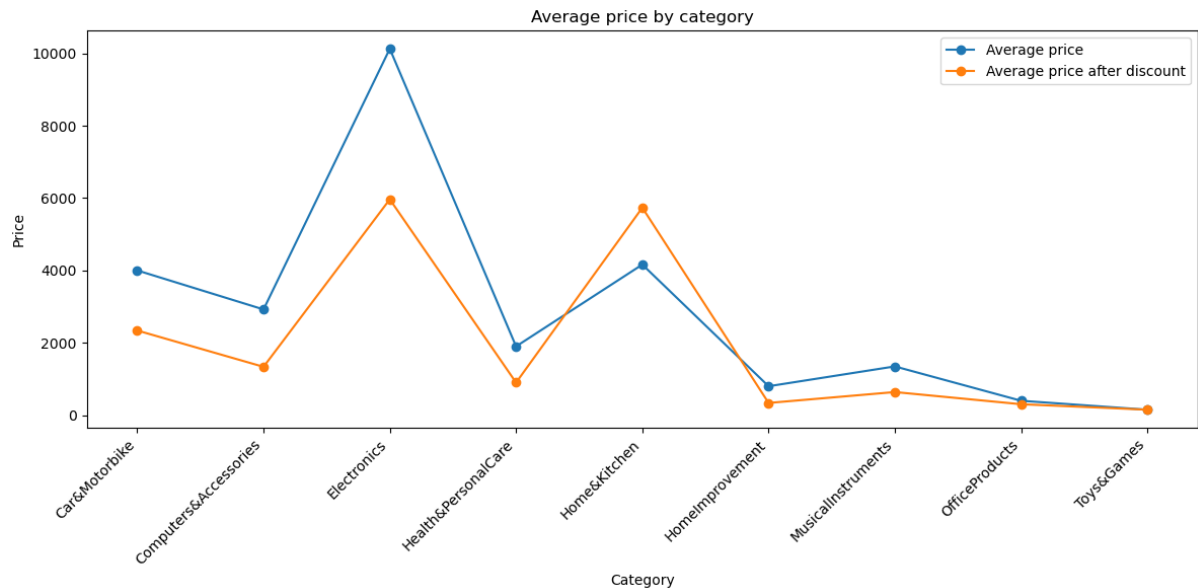
The provided dataset has been sorted by 'category' feature into five following groups according to the distribution of product count. Next, all missing, null or unrelated data to our analysis features were dropped.



According to the distribution of product count into categories, there are three prominent categories including: Electronics, Computer&Accessories, Home&Kitchen which all are oscillating above number of 400 products. As another separate category, there is

OfficeProducts with data from around 30 products. The remaining categories have been grouped into Others, due to the smaller number of products.

3) Price



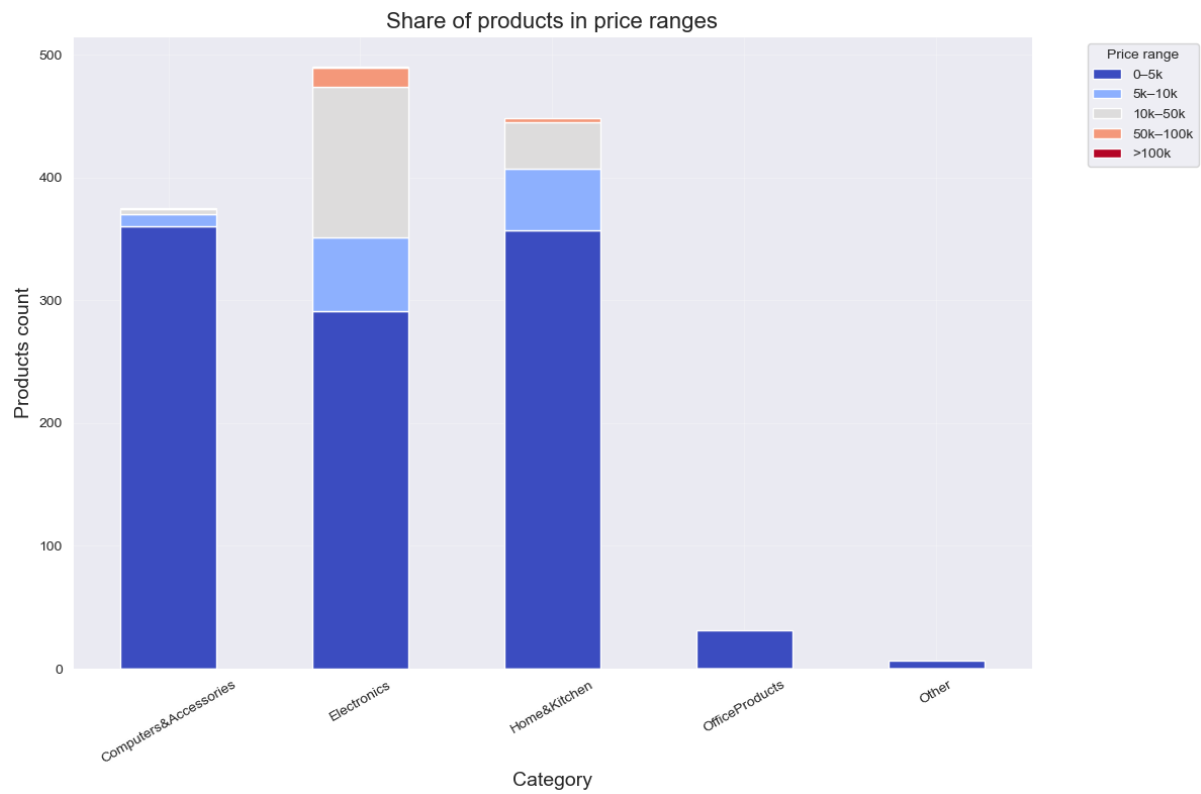
Firstly, we looked at the average prices in each category. **The highest average price is notably in Electronics category**, then slightly lower followed by Home&Kitchen, while the average price of other products is relatively low.



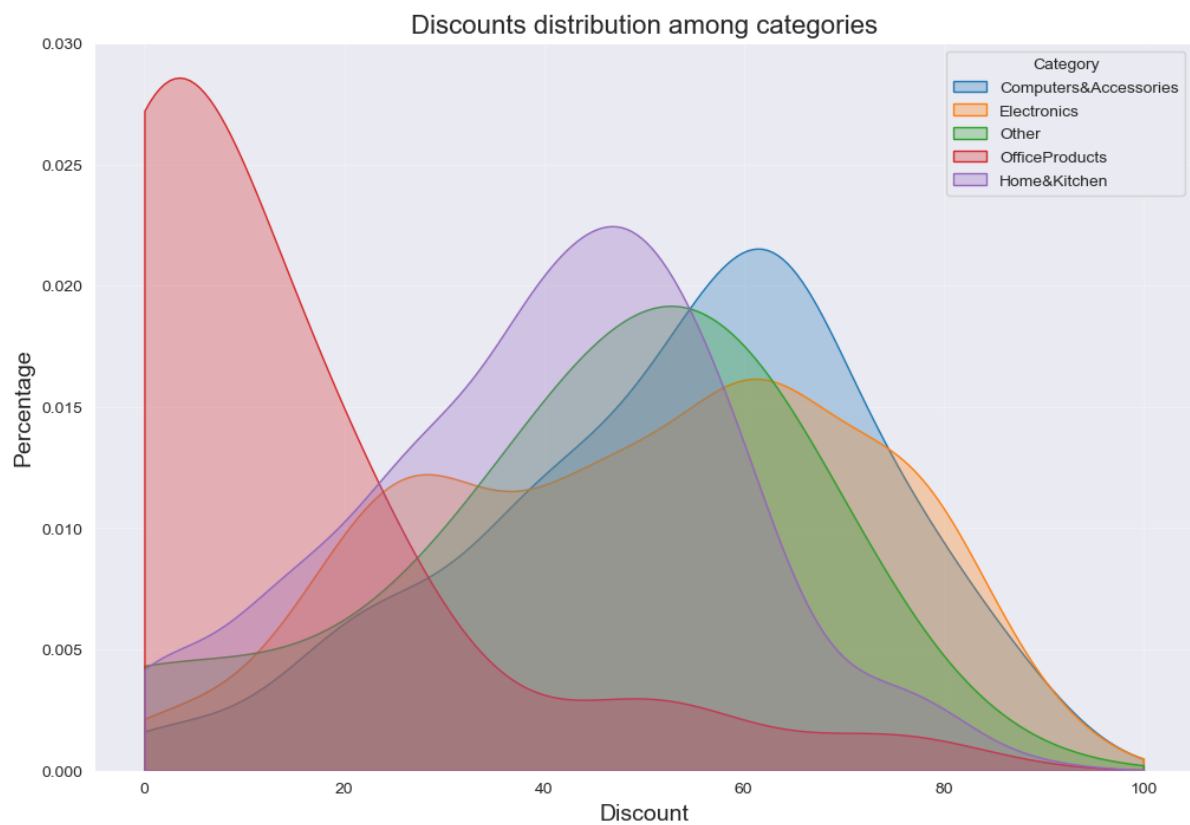
Secondly, we decided to provide a clear breakdown of the average discount in different price ranges. As noted on the graph the distribution of discount percentage is oscillating around similar range, with highest average discounts being put on products in **5-10k category (52%)** and the lowest average noted in **10k-50k category (39%)**.



The histogram of prices before and after discount shows the impact of price discount by shifting the peak frequency of price from **3.0 (before discount – by stabilizing data with log10)** to **2.5 (after discount - by stabilizing data with log10)**. This histogram also suggest that discounts reduce the spread of prices as it concentrates more products in the lower price range.



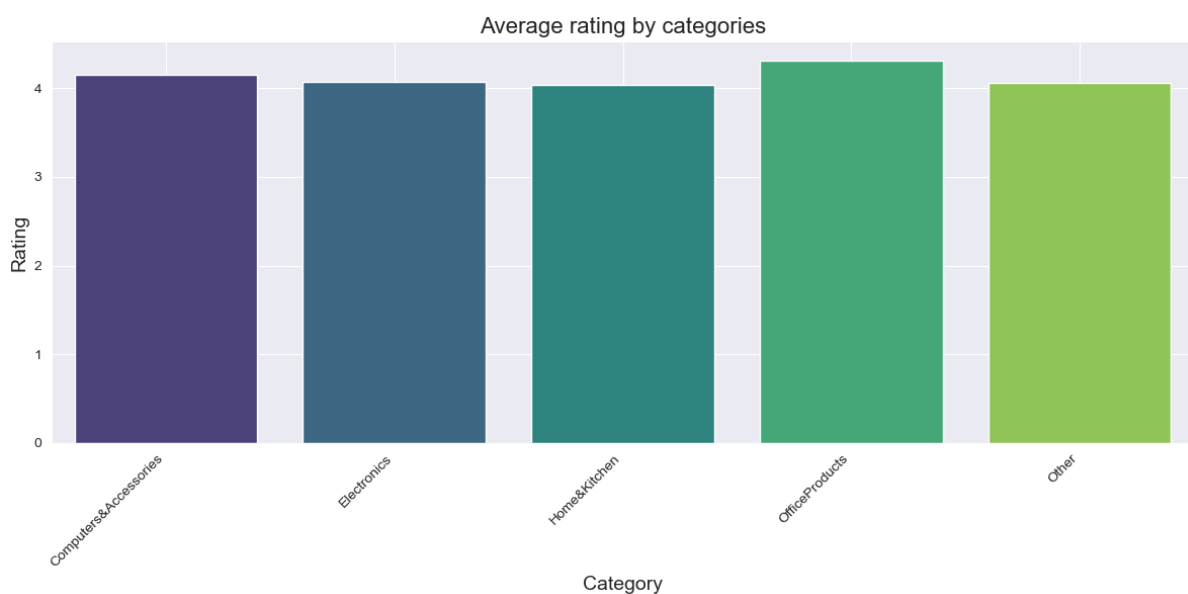
As demonstrated on the chart above, in each of 5 categories, the biggest share of products lands in the price range of **0-5k**.



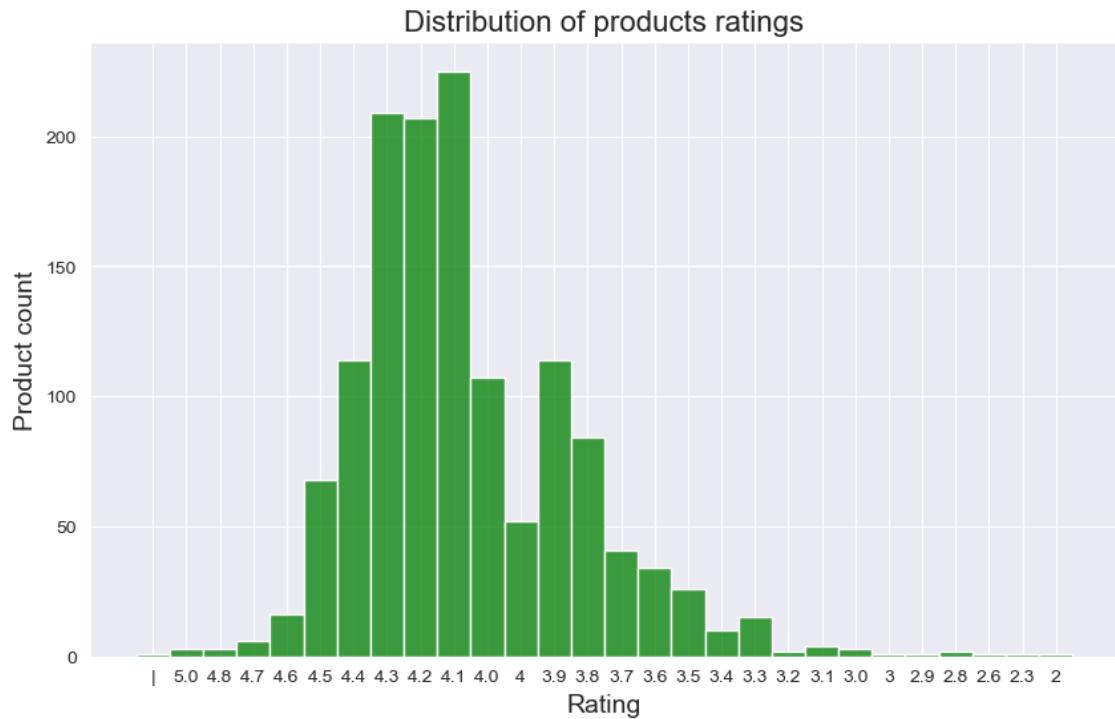
For the distribution of discounts across categories, as we can see each category holds its own discount pattern. Categories like "Office Products" tend to have smaller discounts, while "Electronics" and "Computers & Accessories" have higher discounts. However, it is important to notice that the density curves overlap in some regions, **indicating that certain discount ranges are common across multiple categories.**

- Office Products (Red) - peak at around **10%**
- Home&Kitchen (Purple) – peak at around **40%**
- Other (Green) - peak at around **50%**
- Computer& Accessories (Blue) - peak at around **60%**
- Electronics (Orange) – peak at around **60%**

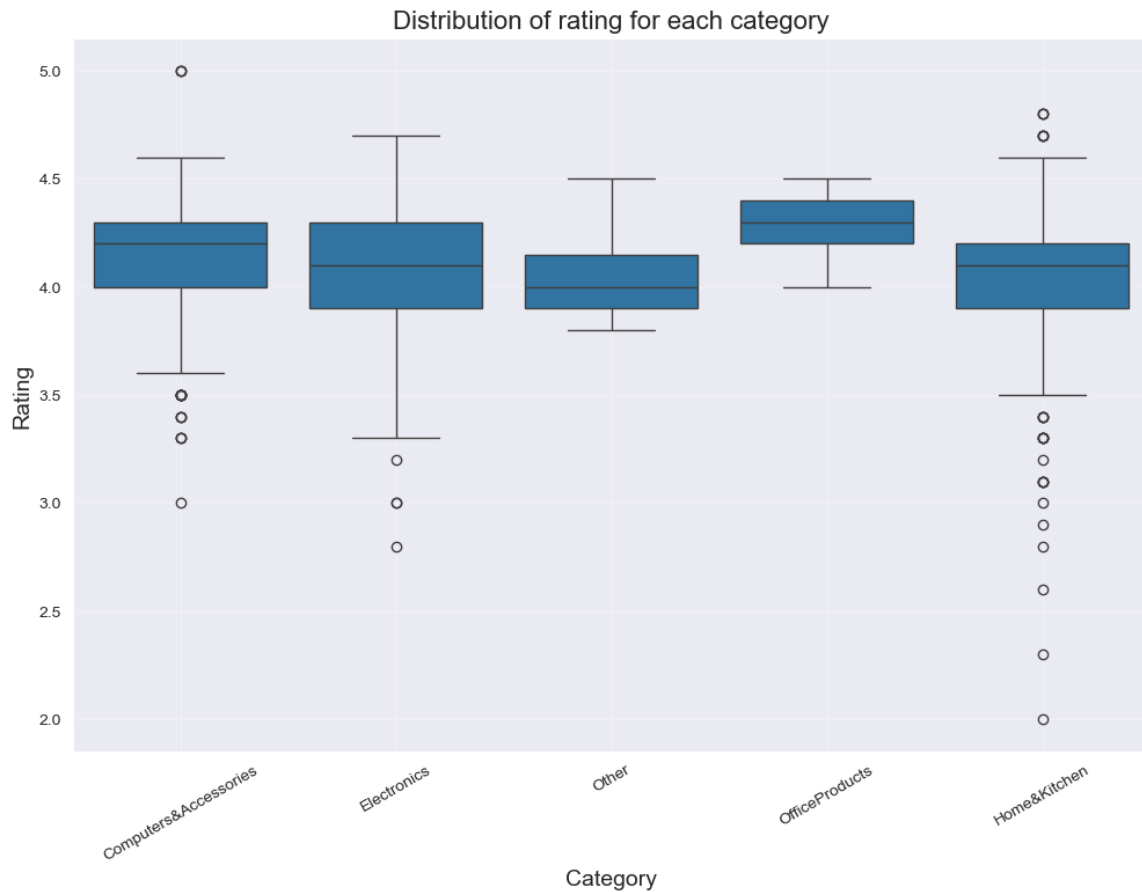
4) Rating



The average ratings range from, **slightly below 4 to slightly above 4.5** in all categories, however the differences between the categories are relatively small, **indicating overall positive feedback across all categories.**



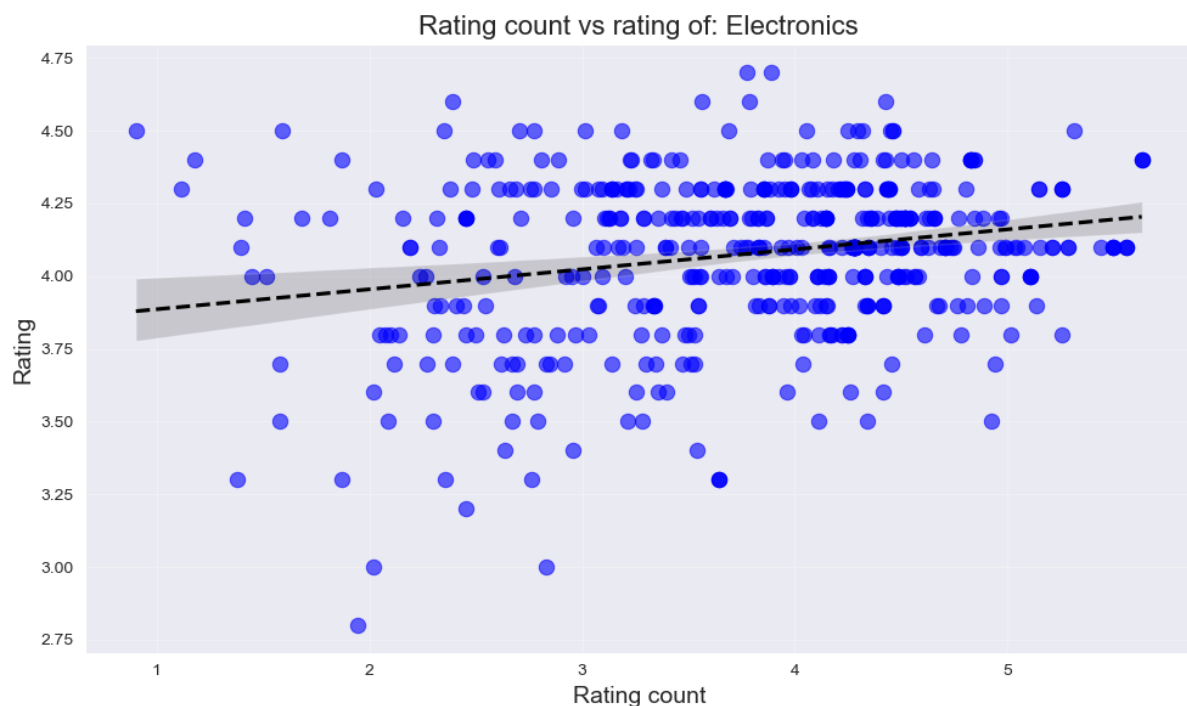
The highest product count is observed at ratings around **4.1-4.3 with number of products exceeding 200**, indicating that more people tend to do leave favourable reviews. Rating of 2.0 has the lowest number of products.



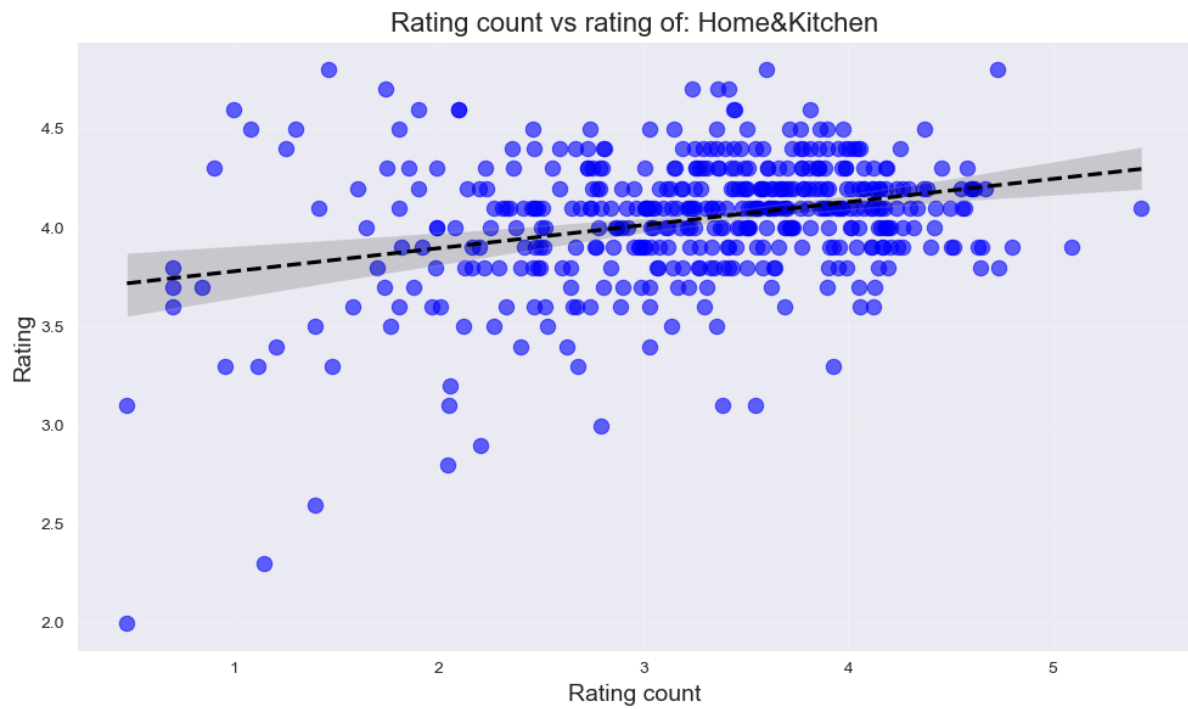
The box plot shows the distribution of ratings across categories. **Computers & Accessories** and **Electronics** have similar medians around 4.0-4.2 and moderate variability, with a few outliers below 3.5. **Other** has a slightly lower median, with some outliers below 3.5. **Office Products** has a tight distribution, with a narrow IQR around 4.1-4.3, no visible outliers, and consistent ratings. **Home & Kitchen** shows the widest variability, with numerous outliers below 3.5, some as low as 2.0, suggesting a broader spread in ratings. Overall, categories vary in consistency and the presence of outliers.

Rating Count vs. Rating Across Categories

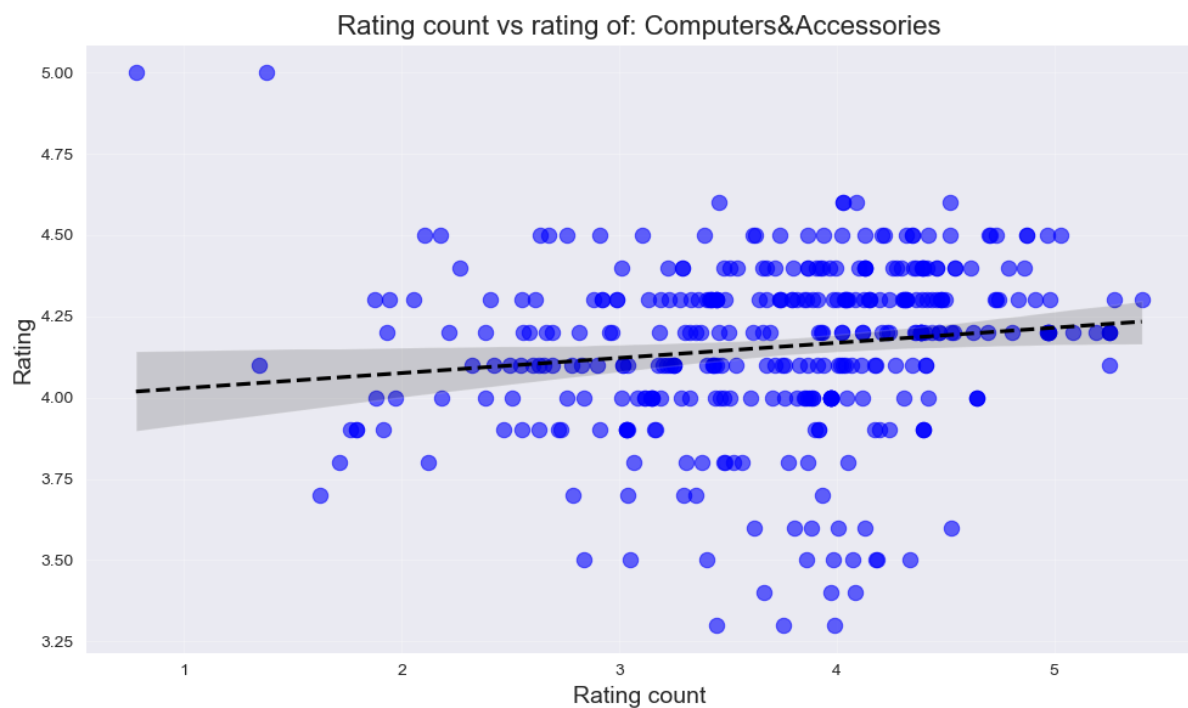
The next visualizations depict the relationship between the **rating count** and the **average rating** for products in five categories: **Electronics**, **Home & Kitchen**, **Computers & Accessories**, **Office Products**, and **Other**.



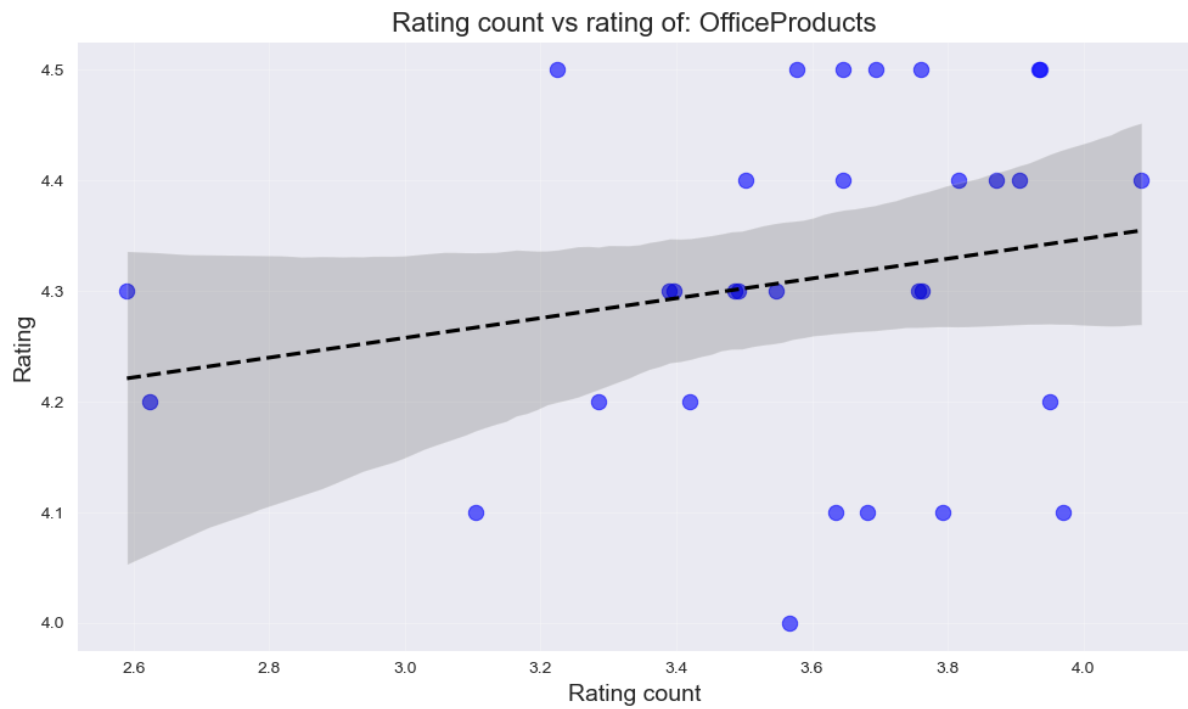
Electronics - a weak positive correlation is observed. The ratings cluster around 4.0–4.5 and higher rating counts slightly correspond to marginally better ratings.



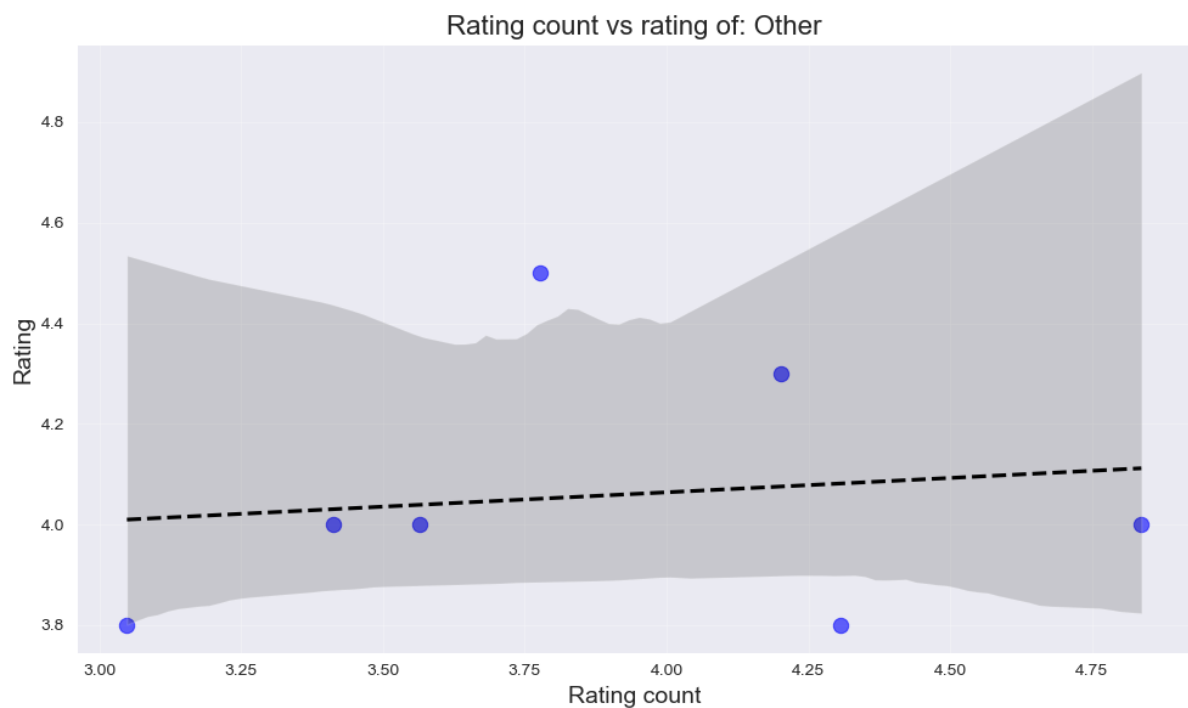
Home & Kitchen - a trend similar to Electronics, with a slight increase in rating as rating count grows. Ratings are more tightly distributed, indicating consistency in user reviews.



Computers & Accessories- a wider spread of data is evident, especially at lower rating counts. Despite some variability, the average rating still shows a positive trend with increasing rating count.



Office Products - sparse data with fewer products and reviews, a mild positive trend exists, but the uncertainty (shaded region) suggests low statistical confidence in the correlation.

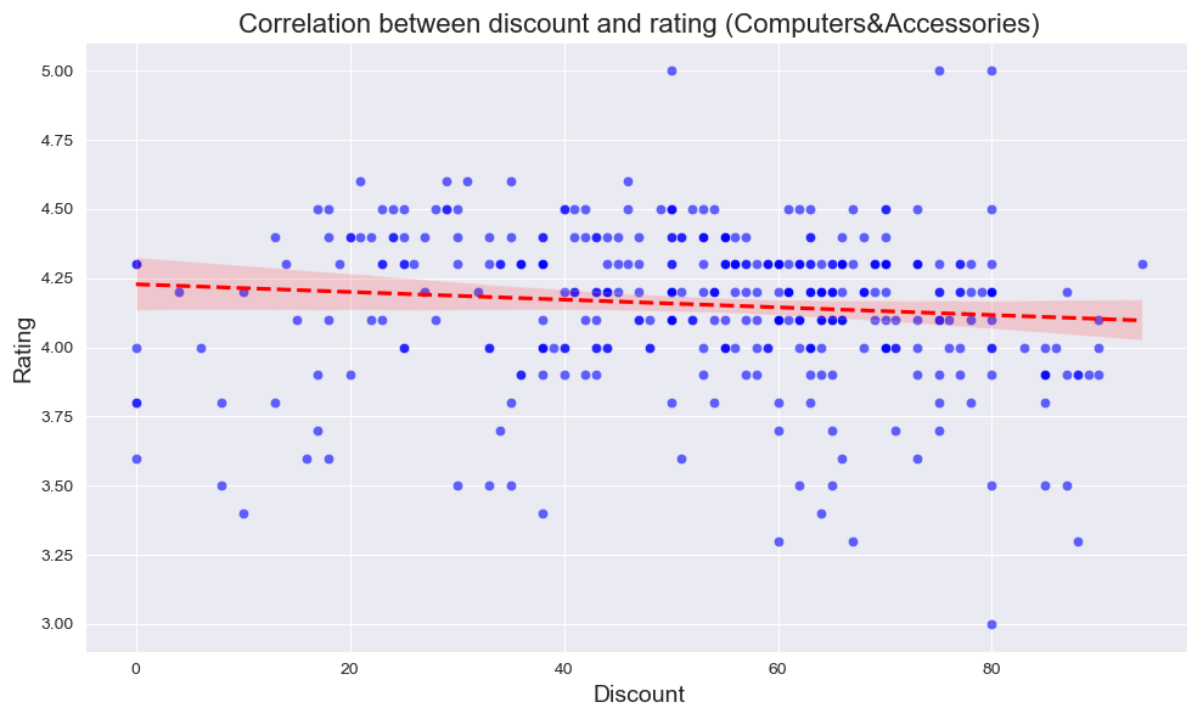


Other - limited data points result in a less defined trend, a positive slope is visible, but the wide shaded area indicates high variability and less reliability in this observation.

4) Correlation

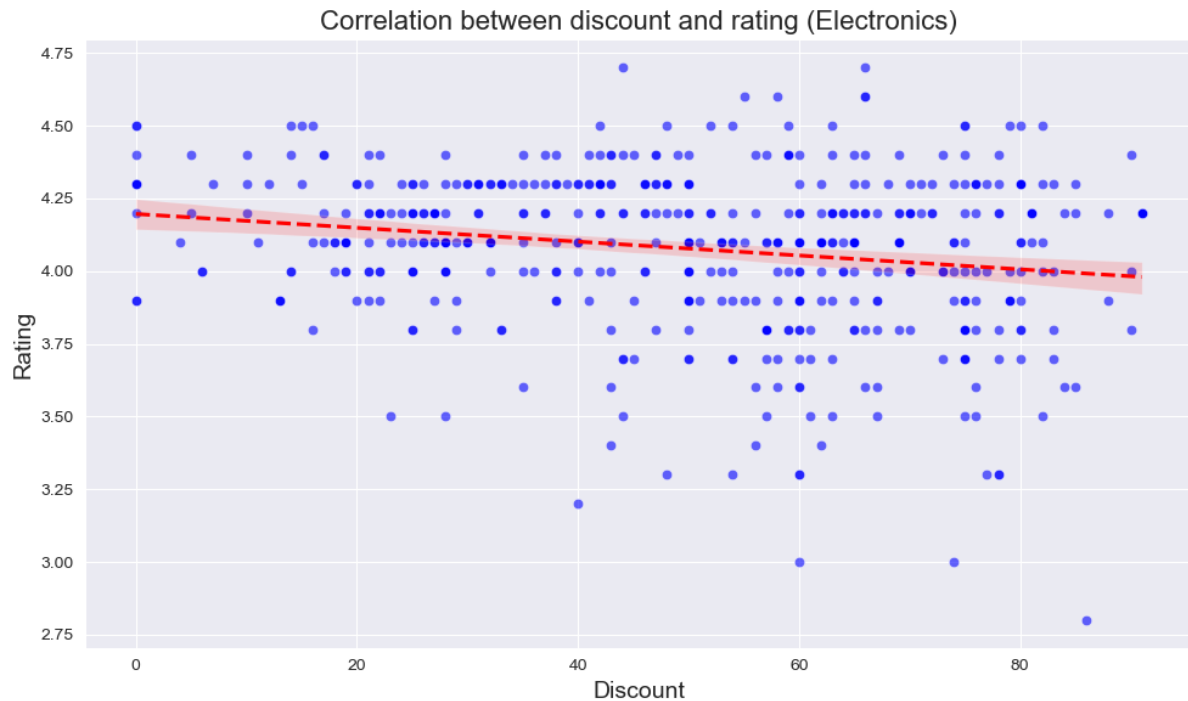
This analysis examines the relationship between discount percentage and customer ratings of Amazon products across different categories. Each scatter plot visualizes the data for a specific category and includes a regression trendline to show the overall trend. The red-shaded area around the trendline represents the **confidence interval**, which provides an estimate of the reliability of the observed trend. A narrower confidence interval indicates more consistent data and greater confidence in the trend, while a wider confidence interval suggests greater variability or limited data.

4.1. Computers & Accessories



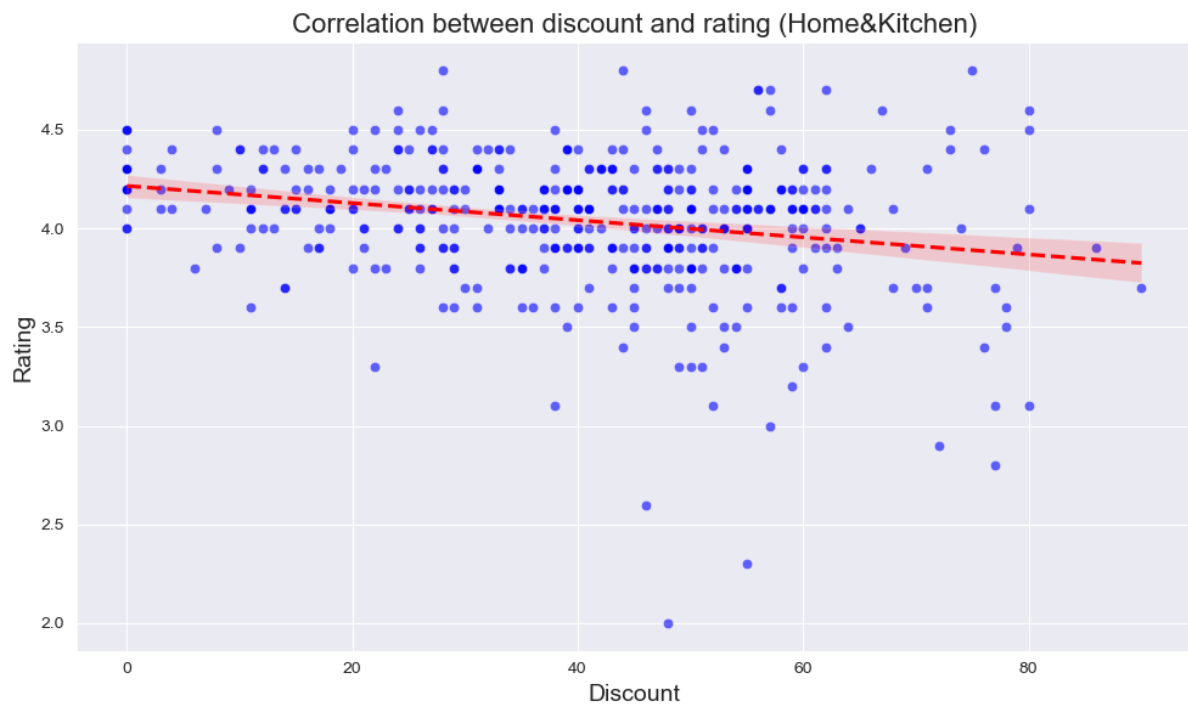
- **Trendline:** Slightly negative correlation between discount percentage and ratings.
- **Observation:** Products with lower discounts tend to have marginally higher ratings, though the correlation is weak.
- **Confidence Interval:** Narrow, indicating that the trendline is relatively reliable and the data is consistent.

4.2. Electronics



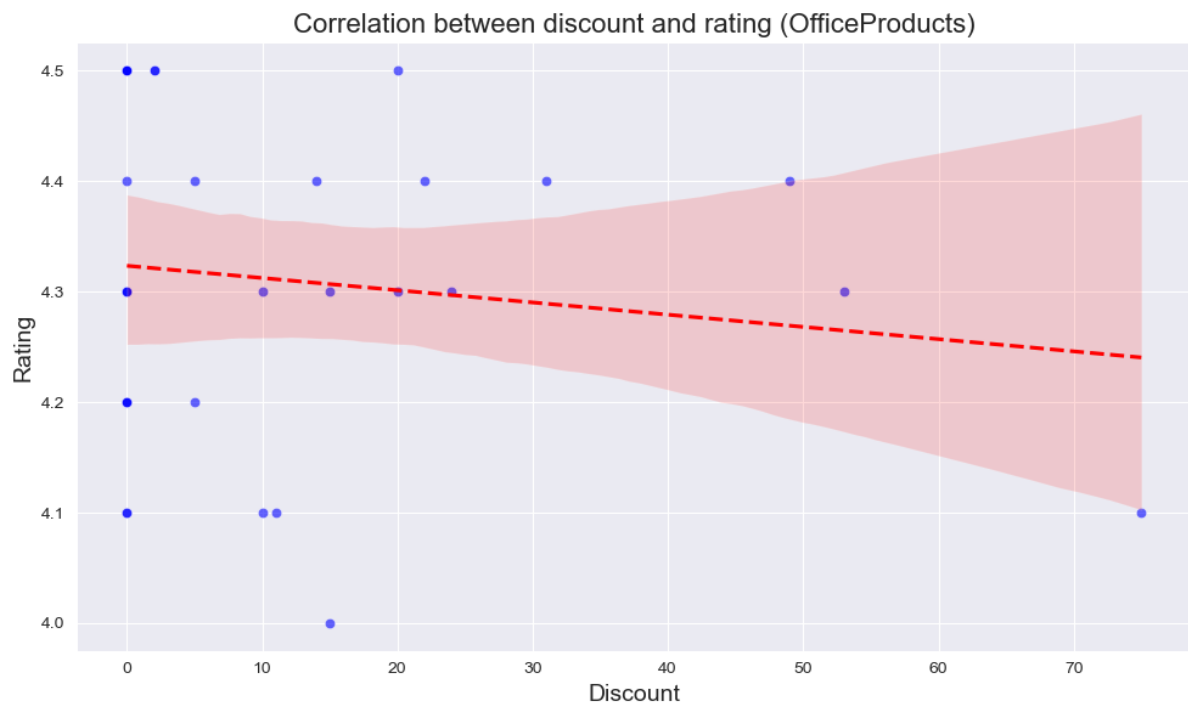
- **Trendline:** Mild negative correlation between discount percentage and ratings.
- **Observation:** Higher discounts are associated with slightly lower ratings, though there is significant scatter in the data.
- **Confidence Interval:** Narrow but slightly wider than Computers & Accessories, indicating moderate consistency in the data.

4.3. Home & Kitchen



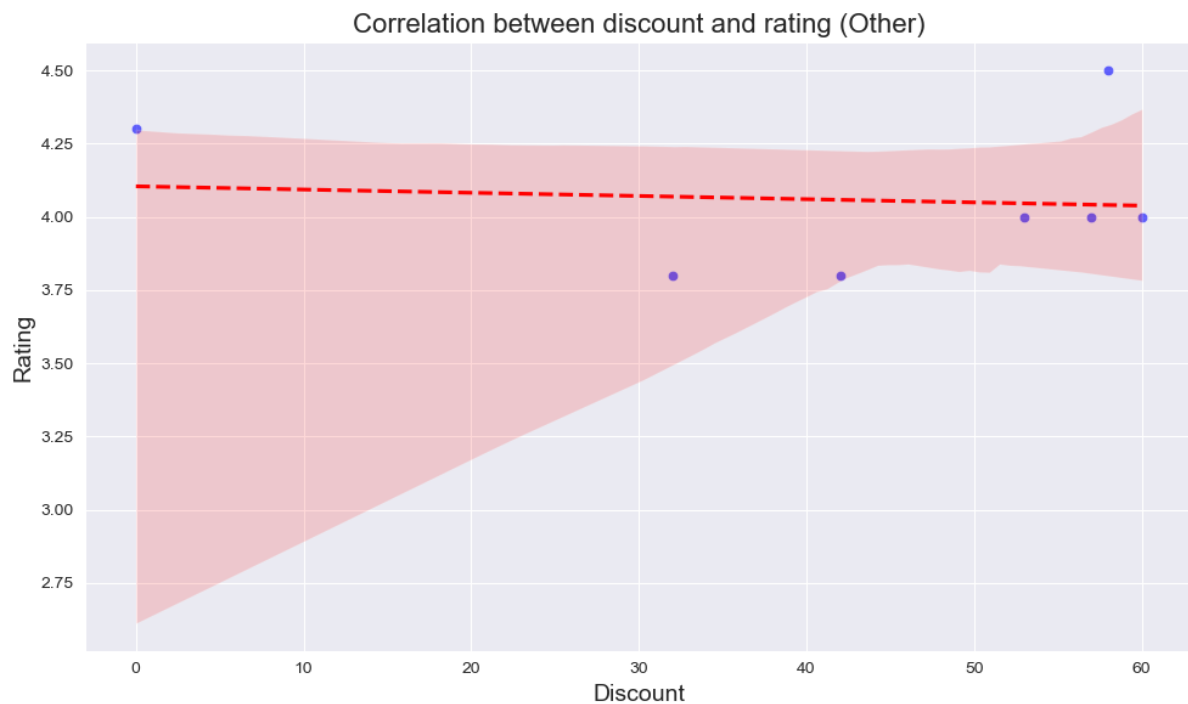
- **Trendline:** Consistent negative correlation between discount percentage and ratings.
- **Observation:** Products with higher discounts tend to have lower ratings, though the difference is small.
- **Confidence Interval:** Narrow, indicating a reliable trendline and consistent data.

4.4. Office Products



- **Trendline:** Weak negative correlation between discount percentage and ratings.
- **Observation:** Ratings show a small decline as discounts increase, but the range of ratings is relatively narrow.
- **Confidence Interval:** Moderate width, showing some variability in the data, but still allowing for a general downward trend.

4.5. Other



- **Trendline:** Nearly flat, showing little to no correlation between discount percentage and ratings.
- **Observation:** The data does not indicate any clear pattern or relationship. The variability in the data suggests that discounts have minimal impact on ratings in this category.
- **Confidence Interval:** Very wide, indicating high uncertainty due to limited data points or diverse product types.

4.6. Discount vs rating count

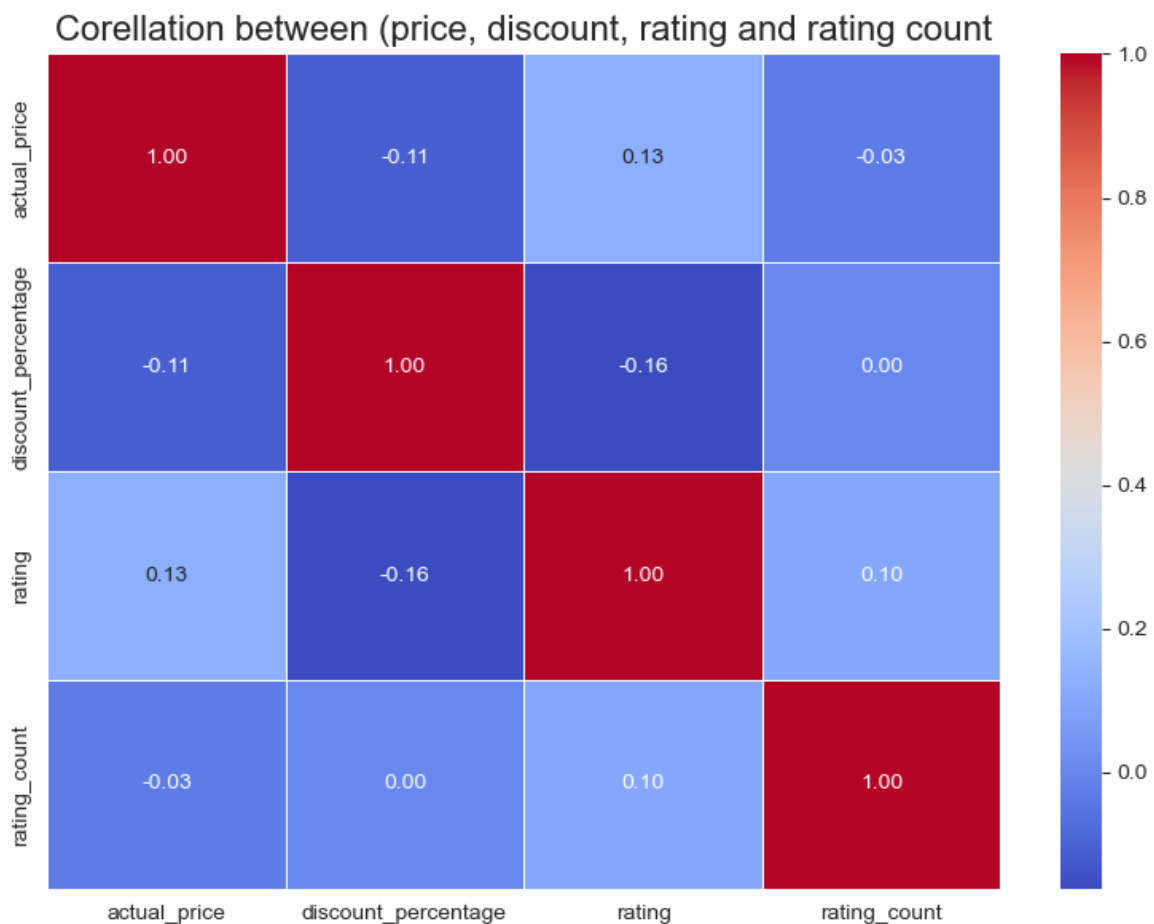
This set of scatter plots explores the relationship between discount percentages and the rating count (popularity) of products in specified categories. Rating count indicates the number of reviews a product has received, reflecting its overall popularity rather than its average rating score.



Observations:

- **No Strong Correlation:** Across all categories, there does not appear to be a strong correlation between discount percentage and rating count. Popularity (rating count) seems to be independent of discounts.
- **Outliers:** In each category, a small number of products receive significantly higher rating counts, which may be due to other factors like product quality, brand reputation, or marketing.
- Categories like "Home & Kitchen" and "Electronics" have more high-rating count outliers compared to "Other" or "Office Products."

4.7. Correlations heatmap



1. Actual Price vs. Other Variables

- **Discount Percentage (-0.11):** Slight negative correlation, indicating that higher-priced products tend to have slightly lower discounts.
- **Rating (0.13):** Weak positive correlation, suggesting that higher-priced products may slightly tend to have better ratings.
- **Rating Count (-0.03):** Virtually no correlation, implying that the price of a product does not significantly influence its popularity.

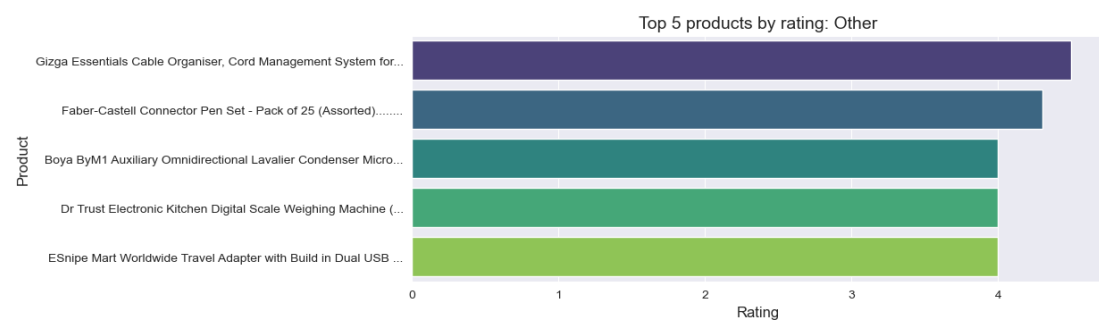
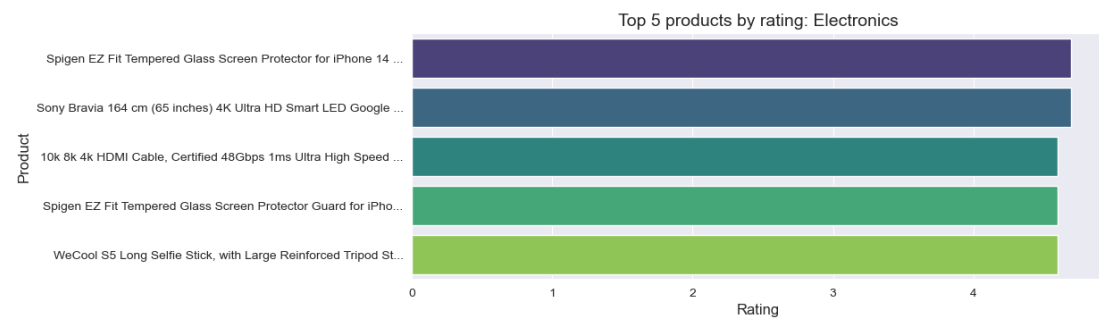
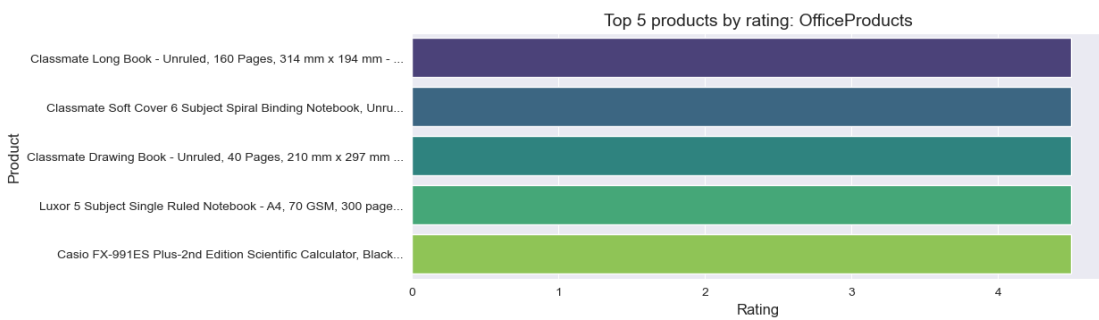
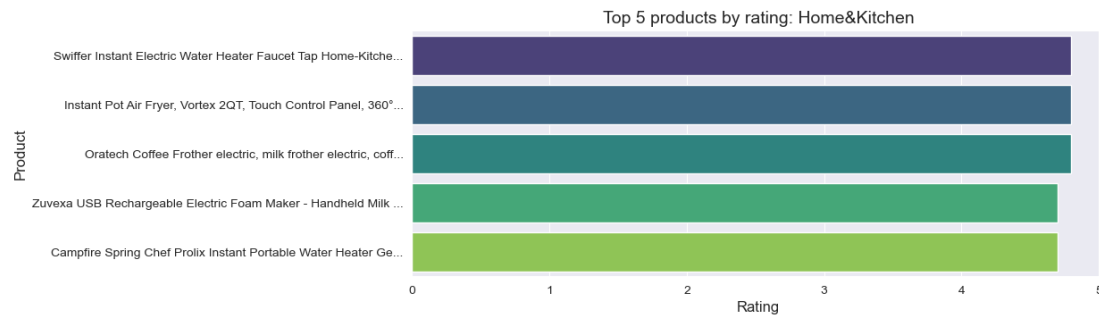
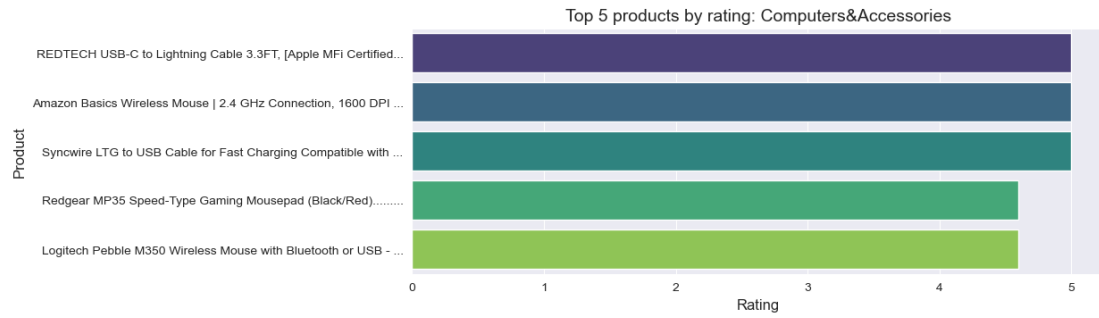
2. Discount Percentage vs. Other Variables

- **Rating (-0.16):** Weak negative correlation, indicating that products with higher discounts may slightly have lower ratings.
- **Rating Count (0.00):** No correlation, suggesting that the level of discount does not directly impact the number of reviews a product receives.

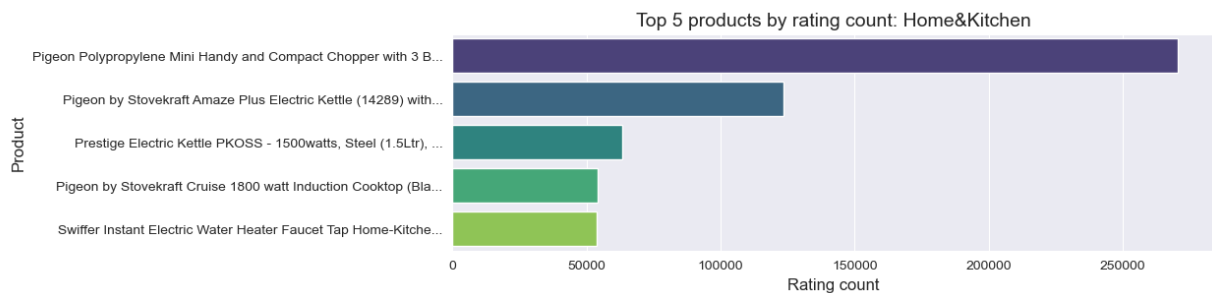
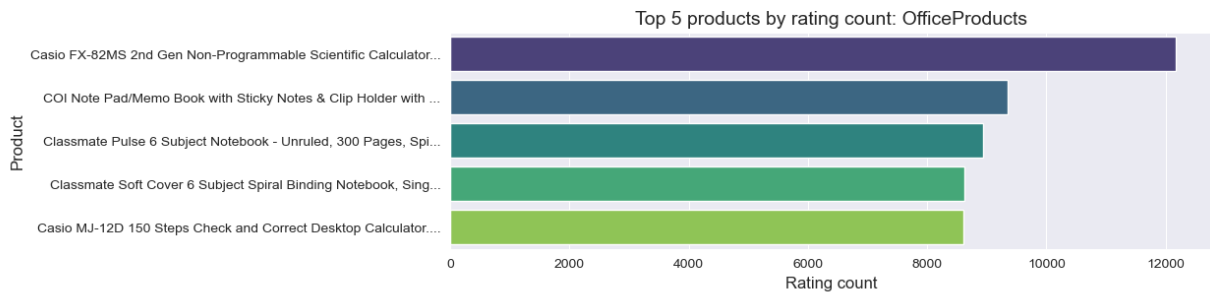
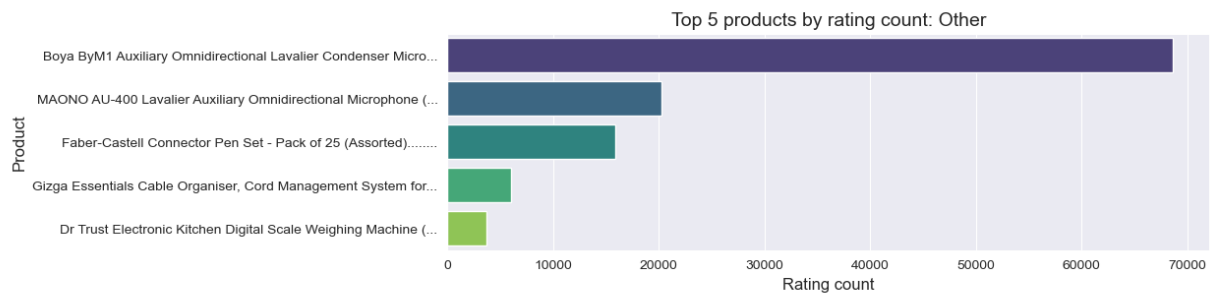
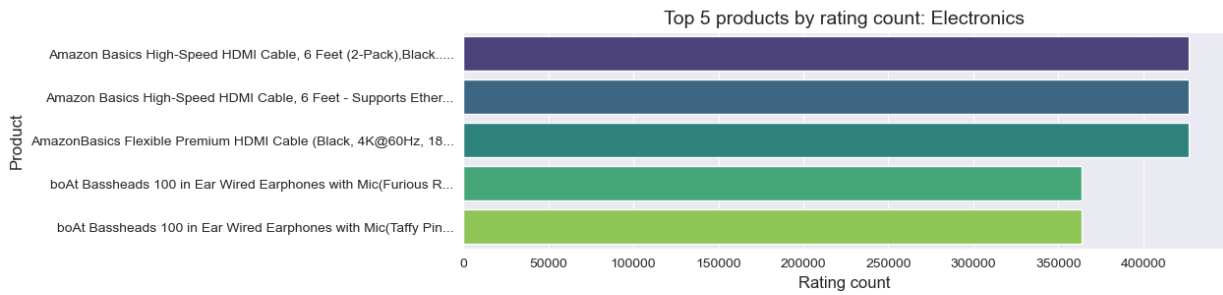
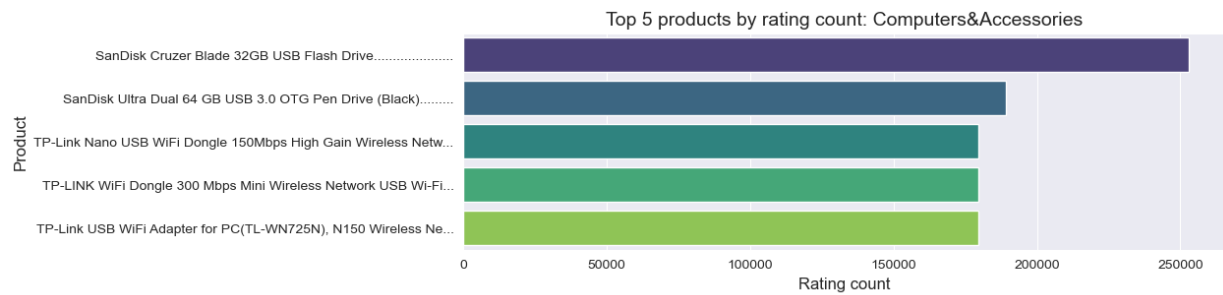
3. Rating vs. Rating Count

- **(0.10):** Weak positive correlation, suggesting that products with higher ratings may slightly attract more reviews.

5) Top Products Rankings

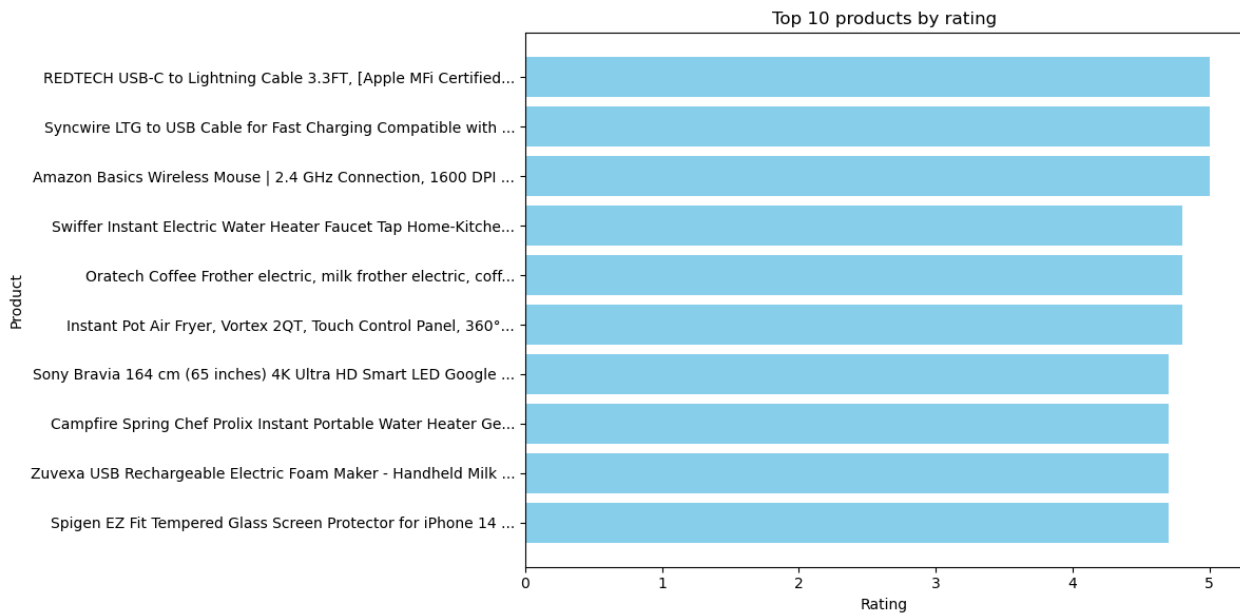


Top products in Computer categories are all landing in the concept as accessories such as cables, mouses, mousepads etc. In Home&Kitchen, there isn't a common denominator for all products. In Office Products, the top five are stationary tech accessories. In Electronic and Other there isn't a common denominator for top 5 products by rating products.



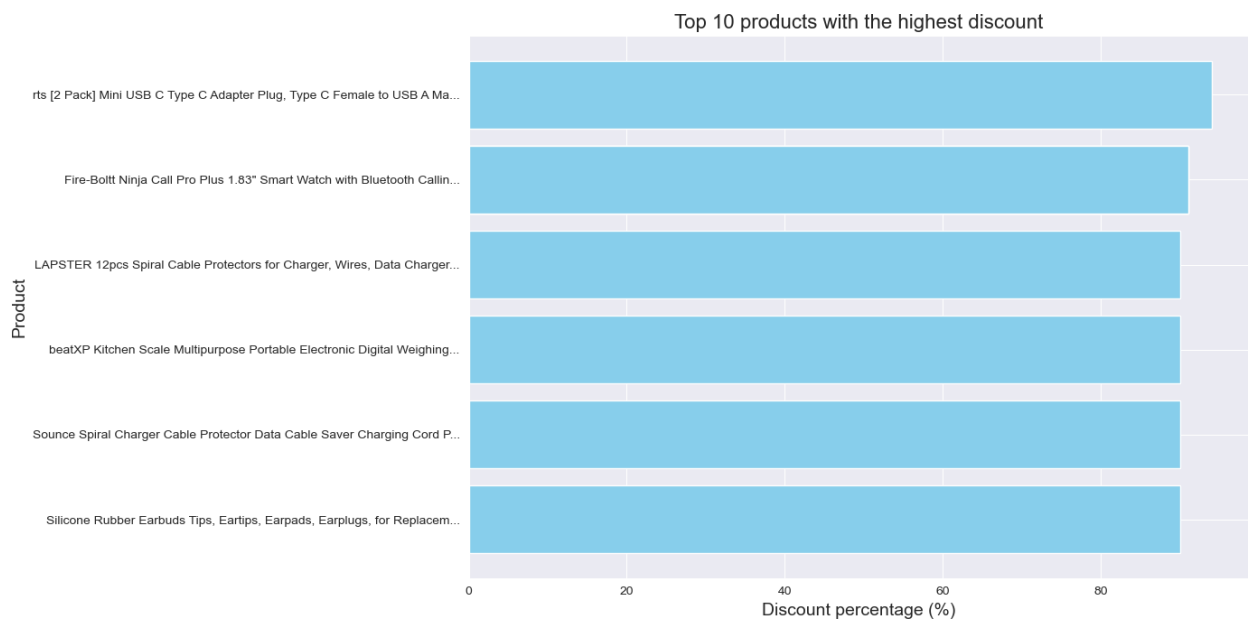
Among all categories, there are various differences between rating count of top 5 products, with the biggest difference landing in both Other and Home&Kitchen categories.

5.3. Top 10 Products by Rating



The bar chart displays the **Top 10 Products by Rating**, with the x-axis representing ratings from 0 to 5 and the y-axis listing product names. All products on the graph have ratings close to the maximum (5), indicating a high level of customer satisfaction. The items include various categories, such as electronics, home appliances and computers & accessories. This chart effectively highlights the most highly rated products, emphasizing their popularity and positive feedback from users.

5.3. Top 10 Products by Discount



The bar chart illustrates the **Top 10 Products with the highest discount**. The x-axis represents the discount percentage ranging from 0 to 100%, while the y-axis contains the product names. Each bar corresponds to a product, with its length indicating the percentage of the discount offered. All products displayed have significant discounts, ranging close to 80% or more, across various categories, including electronics, computer & accessories and home essentials. This visualization highlights the top discounted products, showcasing their attractiveness to price-conscious consumers.