EDA PROJECT REPORT

(Project Term August- November 2023)

"AMAZON PRODUCT SALES ANALYSIS"

Submitted by –

Name : Satwik Uppada

Registration Number : 12111298

Course Code : INT353 (EDA Project)

Under the guidance of -

Shivangini Guptha Mam

School Of Computer Science and Engineering

CA-1

Introduction:

Amazon is the world's largest online retailer, and it offers a vast selection of products for sale. This makes it a great source of data for product sales analysis. In this report, I will analyze the sales data for all Amazon products to identify trends and insights that can be used to improve marketing and pricing strategies.

DOMAIN KNOWLEDGE:

Amazon: From Online Bookstore to E-commerce Giant

Introduction:

- Founded by Jeff Bezos in 1994.
- Launched as an online bookstore in 1995.
- Vision: Leverage the internet for an expansive product selection and exceptional customer experience.

Early Days (1994-2000):

- Jeff Bezos founded Amazon in 1994.
- Jeff Bezos takes charge as CEO in 1995.
- 1997: Amazon's Initial Public Offering (IPO).
- Diversification into books, music, and videos.
- Introduction of Amazon Marketplace.

Dot-Com Boom and Surviving the Bust (2000-2001):

- Adapting to the dot-com bubble burst.
- Expansion of Amazon Marketplace with third-party sellers.

Diversification and Innovation (2002-2010):

- Expansion into electronics, clothing, and more.
- Launch of Amazon Web Services (AWS) in 2006.
- Introduction of the Kindle e-reader in 2007.
- International expansion.

Prime, Fulfillment Centers, and Streaming (2011-2020):

- Introduction of Amazon Prime in 2005.
- Rapid expansion of fulfillment centers.
- Entry into streaming with Amazon Prime Video.
- COVID-19 pandemic impact: Surge in e-commerce demand.

AI Innovations:

- *Alexa and Echo Devices:* Voice-activated virtual assistant with natural language processing (NLP).
- *Recommendation Systems:* Personalized product recommendations.
- *Amazon Go Stores:* Cashier-less shopping with computer vision and AI.
- *Amazon Web Services (AWS):* AI and machine learning services.
- *Amazon Rekognition:* Image and video analysis.

- *Amazon Polly*: Text-to-speech service.
- *Amazon Comprehend:* Natural language processing.
- *Amazon Robotics:* Warehouse automation with AI.
- *Amazon Forecast:* Time-series forecasting.
- *AI in Healthcare:* Initiatives in pharmacy and healthcare AI.

Present and Future (2022 and Beyond):

- Jeff Bezos steps down as CEO in 2021, succeeded by Andy Jassy.
- Global e-commerce dominance.
- Innovation in healthcare and grocery.
- Focus on artificial intelligence and sustainability.

Conclusion:

- Amazon's transformation from a bookstore to a technology and e-commerce giant.
- A testament to visionary leadership, innovation, and customer-centricity.

DATA UNDERSTANDING:

The data that i will be using for this analysis is from Kaggle. This data includes information on product names, product sales, pricing, rating, and reviews. I will use this data to track the sales trends for all Amazon products, identify the most popular products, and assess the competitive landscape.

Data understanding of each column:

name: This column contains the names or titles of the products. Data type: Object.

main category: This column represents the primary category to which the product belongs. Data type: Categorical .

sub category: This column represents a more specific sub-category or type within the main category. Data type: Categorical .

actual price: This column contains the actual/original price of the products. Data type: Numerical (Float).

discount price: This column represents the discounted price of the products. Data type: Numerical (Float).

ratings: This column contains the ratings or customer reviews for the products. Data type: Numerical (Float).

no of ratings: This column contains the number of ratings or reviews received for each product. Data type: Numerical (Float).

BASIC QUESTIONS:

1. **How many rows are there in the dataset?**

- There are 551,585 rows in the dataset.

2. **How many columns are in the dataset?**

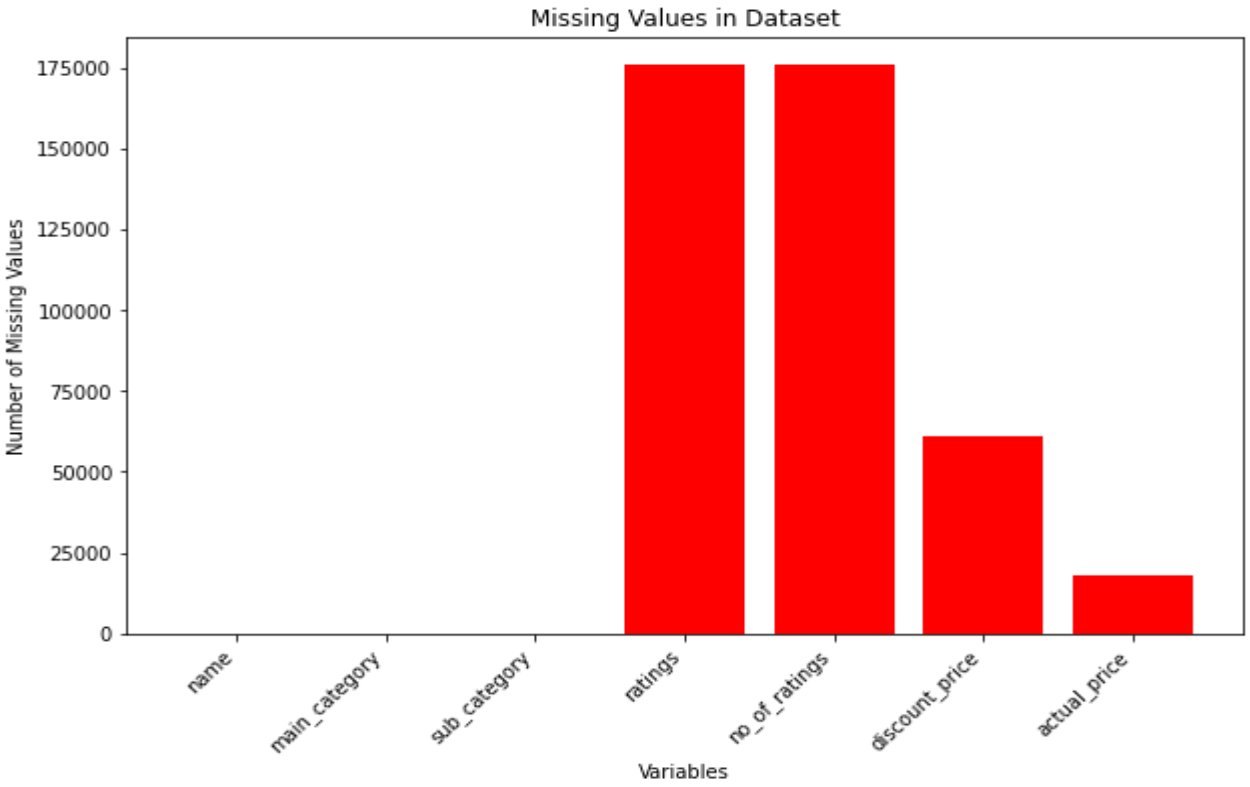
- There are 10 columns in the dataset.

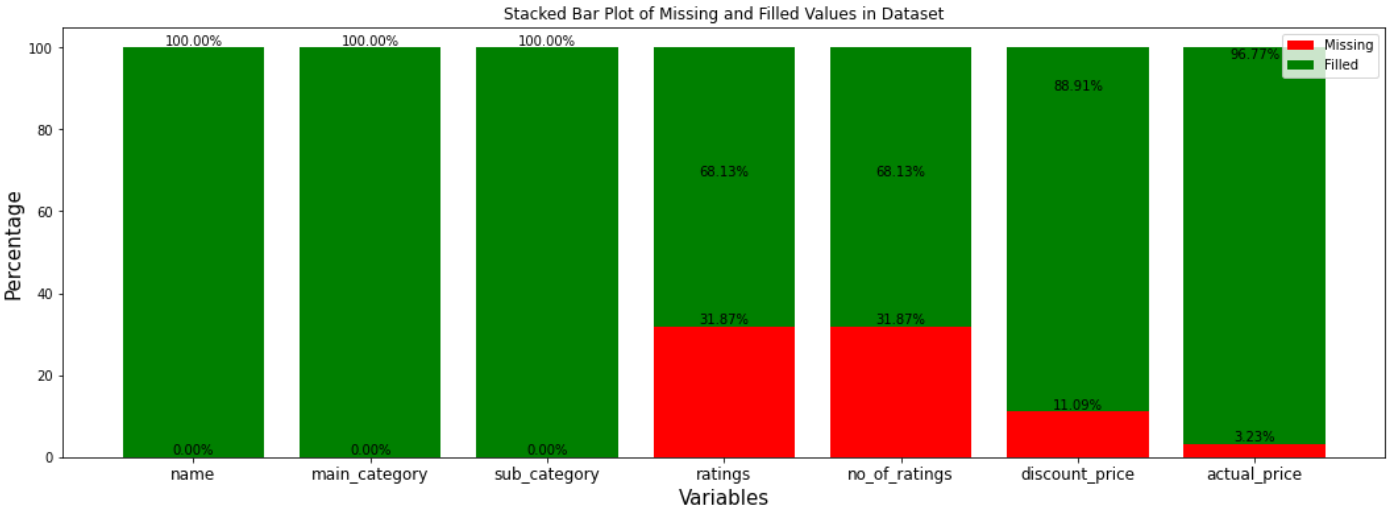
3. **What are the data types of each column in the dataset?**

- 'Unnamed': Numerical(int)
- 'name': Text/String
- 'main_category': Categorical (Text/String)
- 'sub_category': Categorical (Text/String)
- ‘image’ : Categorical (Text/String)
- `link`: Categorical (Text/String)
- 'actual_price': (Text/String)
- 'discount_price': (Text/String)
- 'ratings': Numerical Categorical (Text/String)
- 'no_of_ratings': (Text/String)

4. **Are there any missing or null values in the dataset?**

- No, there are no null values in the dataset. All columns have non-null entries.
- But we have missing values in ratings, no.of ratings, discount_price, actual_price





`ratings` has 175794 (31.9%) missing values	Missing
`no_of_ratings` has 175794 (31.9%) missing values	Missing
`discount_price` has 61163 (11.1%) missing values	Missing
`actual_price` has 17813 (3.2%) missing values	Missing

5. **What is the range or distribution of values for specific columns?**

- To analyze the range and distribution of values for specific columns, you can create histograms, box plots, or summary statistics for those columns. If you have specific columns in mind, please let me know, and I can provide further details.

6. **Are there any categorical columns, and what are the unique categories within them?**

Yes, there are categorical columns:

- 'main_category'
- 'sub_category'
- 'Rating_level'
- To find the unique categories within these columns, you can use the **unique()** function or perform value counts for each categorical column.

7. **What is the overall structure of the dataset?**

- The dataset has a structured tabular format with rows and columns. Each row represents a unique product, and each column represents a specific attribute or feature related to the products.

8. **Are there any initial observations or patterns in the data?**

- Without specific analysis or questions, it's challenging to provide initial observations or patterns. Exploratory Data Analysis (EDA) techniques can help uncover insights and patterns in the data based on your specific research goals and questions.

Dataset Statistics:

Dataset statistics

Number of variables	10
Number of observations	551585
Missing cells	430564
Missing cells (%)	7.8%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	42.1 MiB
Average record size in memory	80.0 B

Variable types

Numeric	1
Text	7
Categorical	2

Reasons for Choosing the Dataset: I chose this dataset because it is a comprehensive and reliable source of information on Amazon product sales. It covers all product categories on Amazon. This makes it a valuable resource for product sales analysis.

Real-World Complexity: The dataset reflects the real-world complexity of e-commerce operations, including variations in product offerings, customer behavior, and market dynamics. Analyzing such complex data prepares us for real-world challenges in the field.

Practical Application: The insights gained from this analysis can have practical applications for businesses operating in the e-commerce domain. Recommendations derived from our findings can be used to optimize sales strategies and enhance customer experiences.

Questions for Analysis:

The following are some of the questions that we will be answering in this report:

A. What are the top 10 brands with respective to their product counts?
B. What are the least 10 brands with respective to their product counts?
C. What are the main categories of the products in terms of their counts?
D. Which products are the most popular on Amazon?
E. What are the top 10 most popular products by ratings?
F. What is the distribution of the number of ratings for the products in the dataset, and what insights can we gain from this distribution?
G. What is the distribution of ratings given to the products in the dataset, and what can we infer from this rating distribution?
H. Explain the pricing strategies you observe among top sellers?
I. Explain the relationship between ratings and sales performance?
J. Explain which main categories have high average ratings?
K. Explain how ratings impact the sales performance of products?
L. Explain which main categories are the most popular based on frequency?
M. Explain which main categories are the most competitive based on the number of unique products?
N. Explain the relationship between average ratings and main categories?
O. Explain the top 5 main categories with the highest sales_Performance?
P. Explain the least 5 main categories with the lowest sales performance?
Q. Explain the products with the highest and lowest profit margins?
R. Explain the relationship between discount percentage and sales_performace?
S. Explain how actual prices compare to discounted prices by category?

CA-2

Libraries Used:

- Pandas – used for Data Manuplation and Handling
- Numpy – used for Numerical Operations
- Matplotlib – used for plotting graphs
- Seaborn – used for better visualization
- Statistics – used for Statistical Calculations

Steps for Exploratory Data Analysis(EDA):

Data Collection

- Identify and collect relevant data sources.
- Ensure that the data is clean and complete.
- Convert the data into a format that is compatible with your analysis tools.

Data Inspection

- Understand the shape, dtypes, and content of your data.
- Identify and address missing values, outliers, and other data quality issues.

Data Cleaning

1. **Handling Missing values:**
    - Remove rows or columns with greater than 70% missing values, if necessary.(if it is a crucial pillar of analysis try different approach)
    - Fill null values with 0, the mean, median, mode, or other appropriate methods.(In my project I fill with median)
2. **Remove Duplicates:**
    - Identify and Remove duplicate rows.

Data Transformation

1. **Convert Data Types:**
    - Convert data types to ensure that each column has the correct data type.
2. **Feature Scalling:**
    - Standardize and normalize numerical features to ensure they have similar scales.
3. **Feature Engineering:**
    - Create new features that may be more informative for the analysis or modeling.
    - Transform existing features as needed.

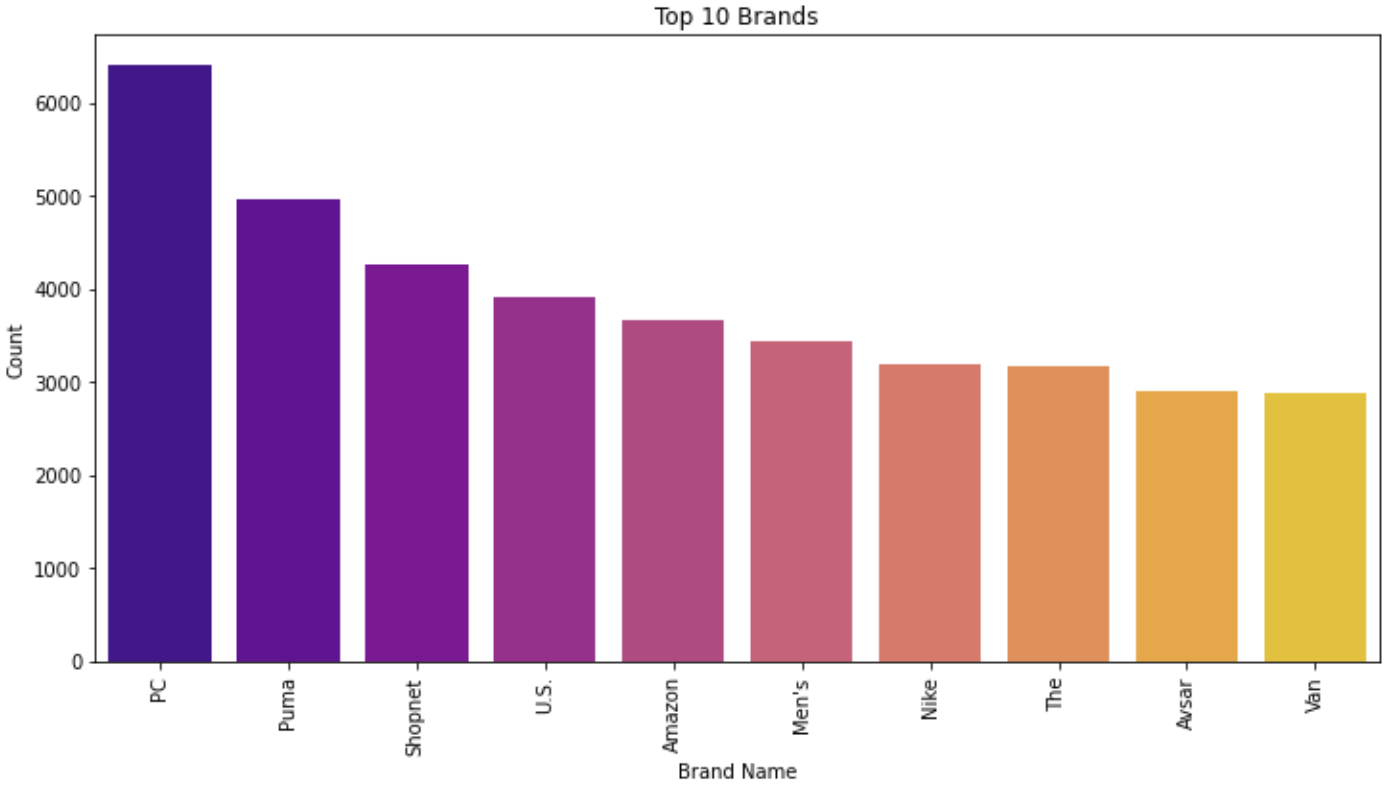
Data Visualization

EDA:

- Perform uni-variate, bi-variate, and multi-variate analysis to identify patterns and trends in the data.
- Detect outliers and identify influential points.
- Calculate summary statistics to describe the central tendency and dispersion of your data.
- Create visualizations to communicate findings to others.
- Generate insights from EDA and use them to guide next steps.

1. What are the top 10 brands with respective to their product counts?

A. After plotting barplot, we can conclude that brands like PC,PUMA, SHOPNET, U.S, AMAZON, MEN'S, NIKE, AVSAR, VAN.

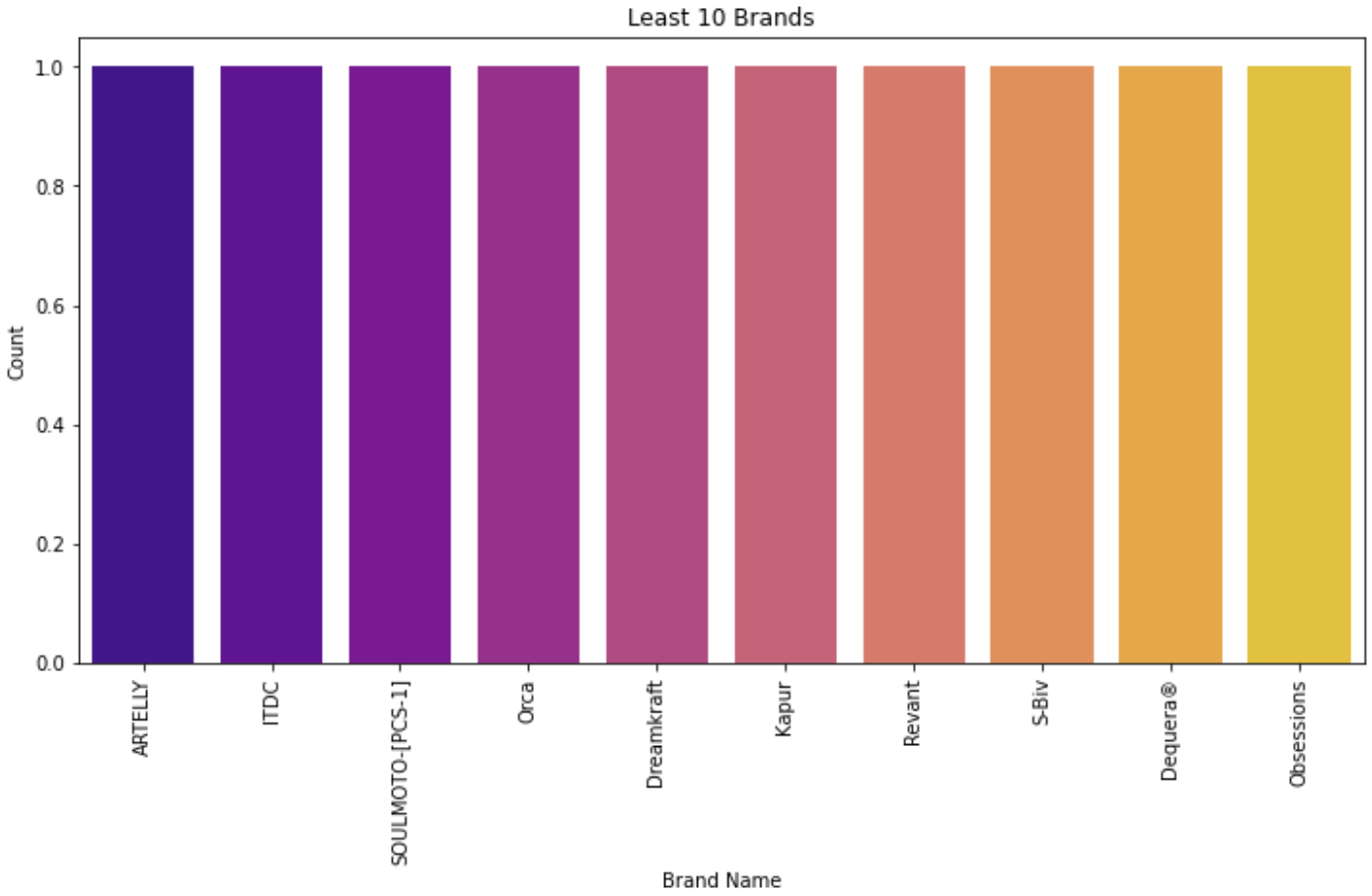


Insight:

From above, we can conclude that the most selling products are from PC brand.

2. What are the least 10 brands with respective to their product counts?

A. After plotting barplot, we can conclude that brands like 'ARTELLY', 'ITDC', 'SOULMOTO-[PCS-1]', 'Orca', 'Dreamkraft', 'Kapur', 'Revant', 'S-Biv', 'Dequera®', 'Obsessions'.

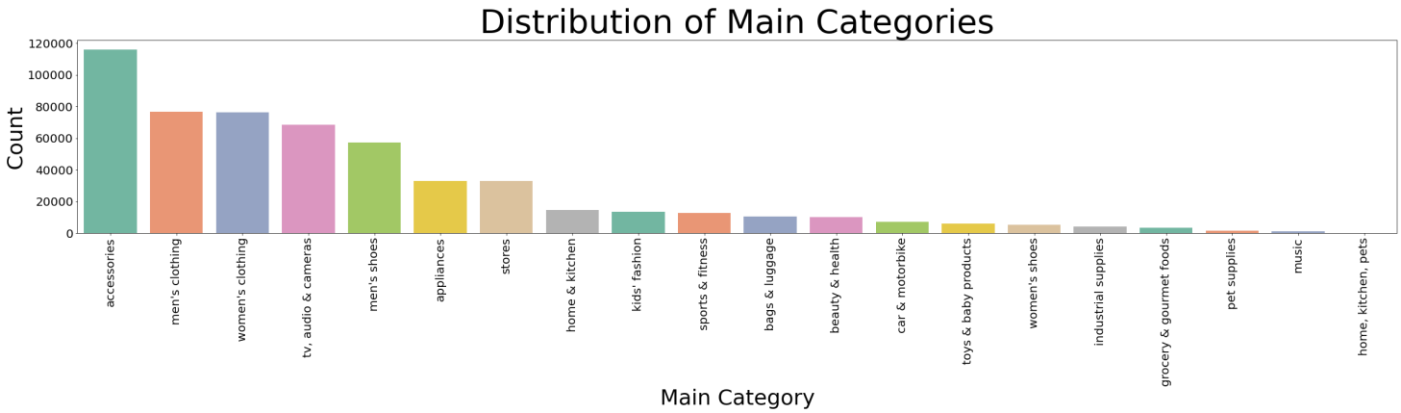


Insight:

From above, we can conclude that the least selling products are from 'ARTELLY', 'ITDC', 'SOULMOTO-[PCS-1]', 'Orca', 'Dreamkraft', 'Kapur', 'Revant', 'S-Biv', 'Dequera®', 'Obsessions' *brands.*

3. What are the main categories of the products in terms of their counts?

A. After plotting barplot, we can conclude that categories like 'appliances', 'car & motorbike', 'tv, audio & cameras','sports & fitness', 'grocery & gourmet foods', 'home & kitchen', 'pet supplies', 'stores', 'toys & baby products', "kids' fashion",'bags & luggage', 'accessories', "women's shoes",'beauty & health', "men's shoes", "women's clothing", 'industrial supplies', "men's clothing", 'music', 'home, kitchen, pets'

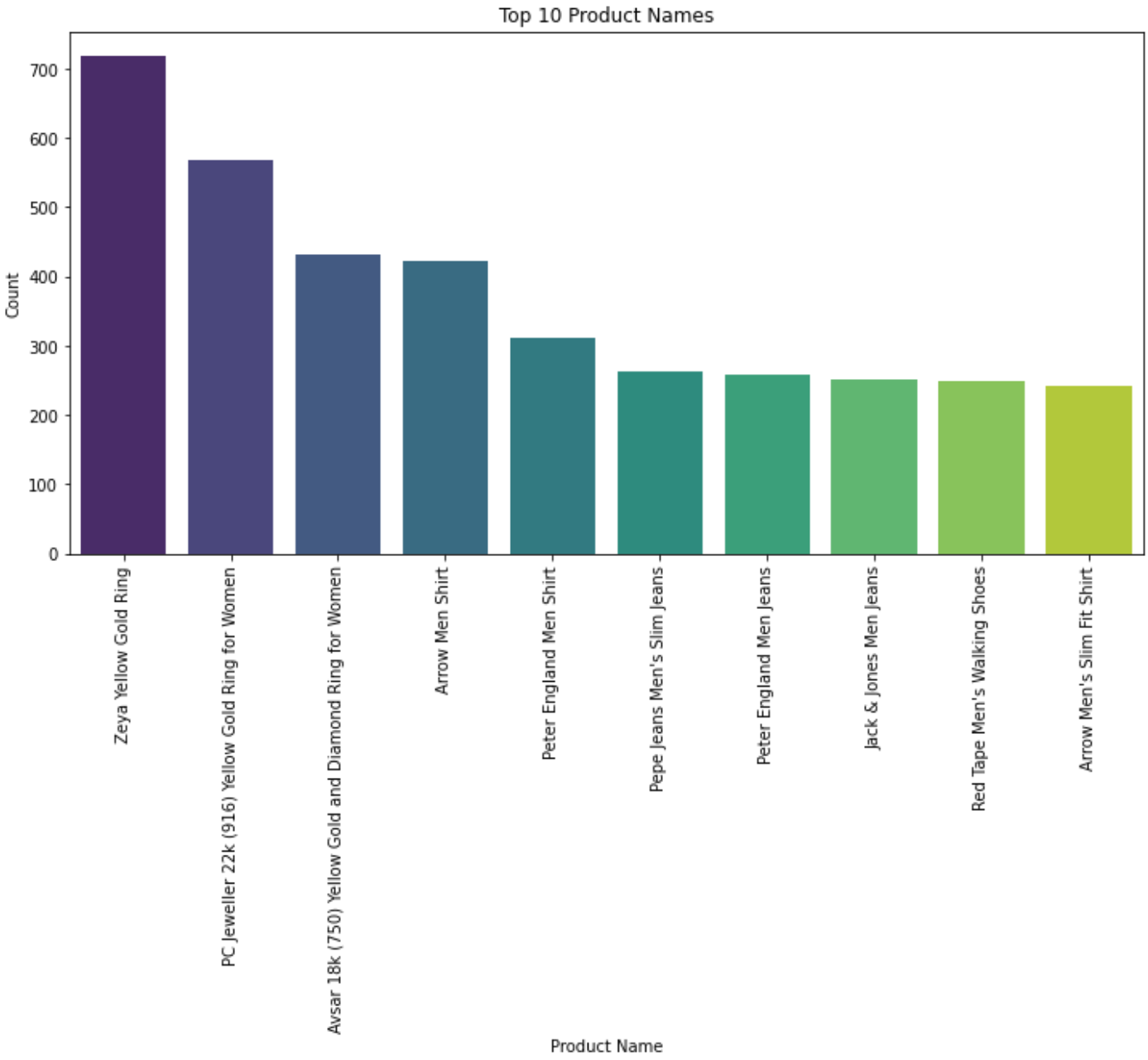


Insight:

From above, we can conclude that the most selling main category is Accessories.

4. Which products are the most popular on Amazon?

A. Most popular products are 'Zeya Yellow Gold Ring', 'PC Jeweller 22k (916) Yellow Gold Ring for Women', 'Avsar 18k (750) Yellow Gold and Diamond Ring for Women', 'Arrow Men Shirt', 'Peter England Men Shirt', 'Pepe Jeans Men's Slim Jeans', 'Peter England Men Jeans', 'Jack & Jones Men Jeans', 'Red Tape Men's Walking Shoes', 'Arrow Men's Slim Fit Shirt'

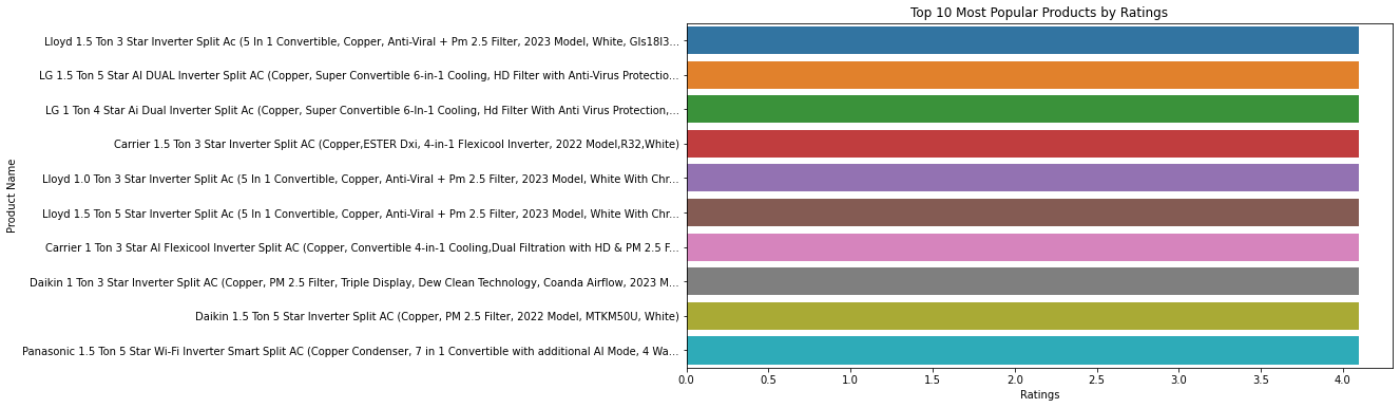


Insight:

'Zeya Yellow Gold Ring' is most popular on amazon

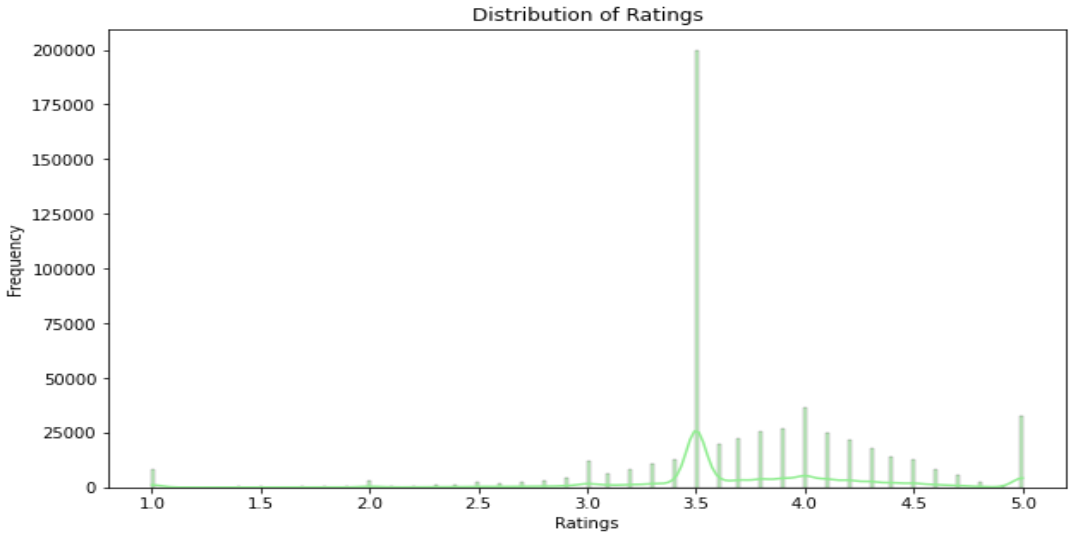
5. What are the top 10 most popular products by ratings?

A. The top 10 most popular products by ratings are 'Hitachi Split Ac - 1.5 Ton Kiyora 5200Fx I Fresh Inverter - R32 - RSRG518FFEO (Gold)', 'Panasonic 2 Ton 5 Star Wi-Fi Inverter Smart Split AC (Copper, 7 in 1 Convertible with additional AI Mode, Twin Cool, PM 0....', 1.5 Ton 3 Star Hot and Cold Wi-Fi Inverter Smart Split AC (Copper, 7 in 1 Convertible with additional AI Mode, T...', 'Voltas 2 Ton 5 Star Inverter Split AC (Copper Condenser, 245V EAZS, White)', 'Voltas 1 Ton 5 Star, Inverter Split AC(Copper, 4-in-1 Adjustable Mode, Anti-dust Filter, 2023 Model, 125V Vectra Elite, Wh...', 'ONIDA 1.5 Ton 3 Star Inverter Split AC (Copper, 2 Way Swing, Dust Filter, 2023 Model, IR183DAS, White)', 'AC Stand/Heavy Duty Air Conditioner Outdoor Unit Mounting Bracket', 'Carrier 1.5 Ton 5 Star Smart AI Flexicool Hybridjet,Wi-Fi, Inverter Split AC (Copper, Convertible 6-in-1 with Anti-Viral G...', 'Daikin 1.8 Ton 5 Star Inverter Split AC (Copper,2022,FTKM60U,White)', 'O-General 1 Ton 5 Star EFFICIENT & TROPICAL Inverter Split AC (Copper Condenser, ASGG12CGTB-B, White)'



6. What is the distribution of the ratings for the products in the dataset, and what insights can we gain from this distribution?

A. In Ratings column we have these unique values : 4.2, 4. , 4.1, 4.3, 3.9, 3.8, 3.5, 4.6, 3.3, 3.4, 3.7, 2.9, 5. ,4.4, 3.6, 2.7, 4.5, 3. , 3.1, 3.2, 4.8, 4.7, 2.5, 1. , 2.6, 2.8, 2.3, 1.7, 1.8, 2.4, 4.9, 2.2, 1.6, 1.9, 2. , 1.4, 2.1, 1.2, 1.3, 1.5, 1.1

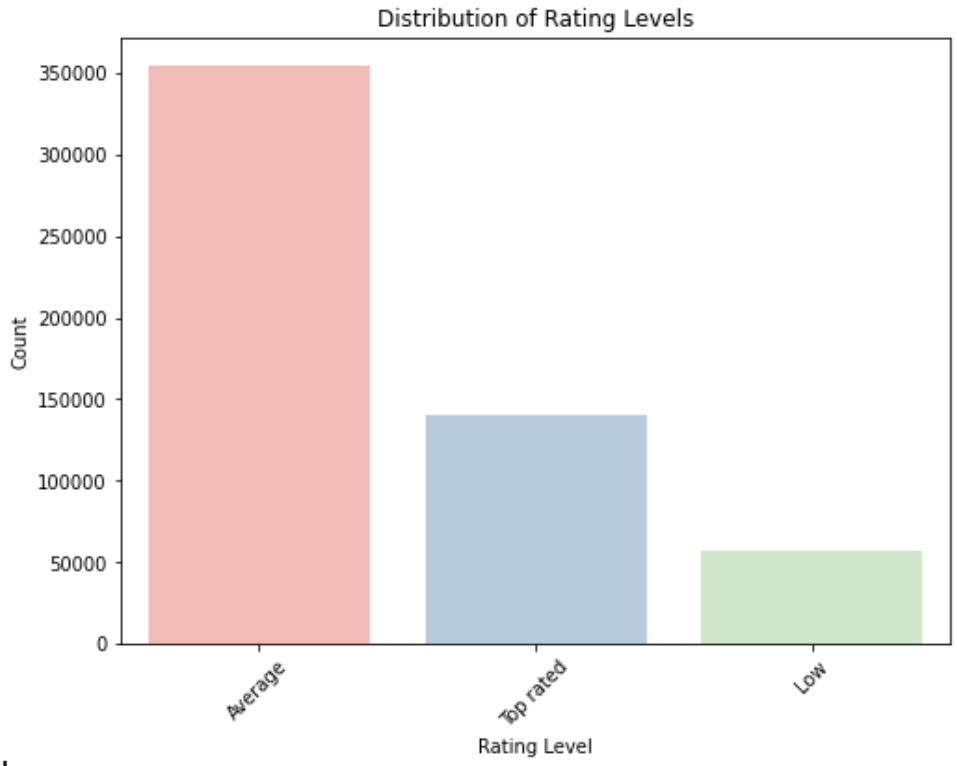


Insight:

From graph, we can conclude that we have more number of ratings as 3.5.

7. What is the distribution of ratings given to the products in the dataset, and what can we infer from this rating distribution?

A. We have three levels in ratings like Top-rated, Average, Low rated.



Insight:

From Graph, we can conclude that Average rating products are high.

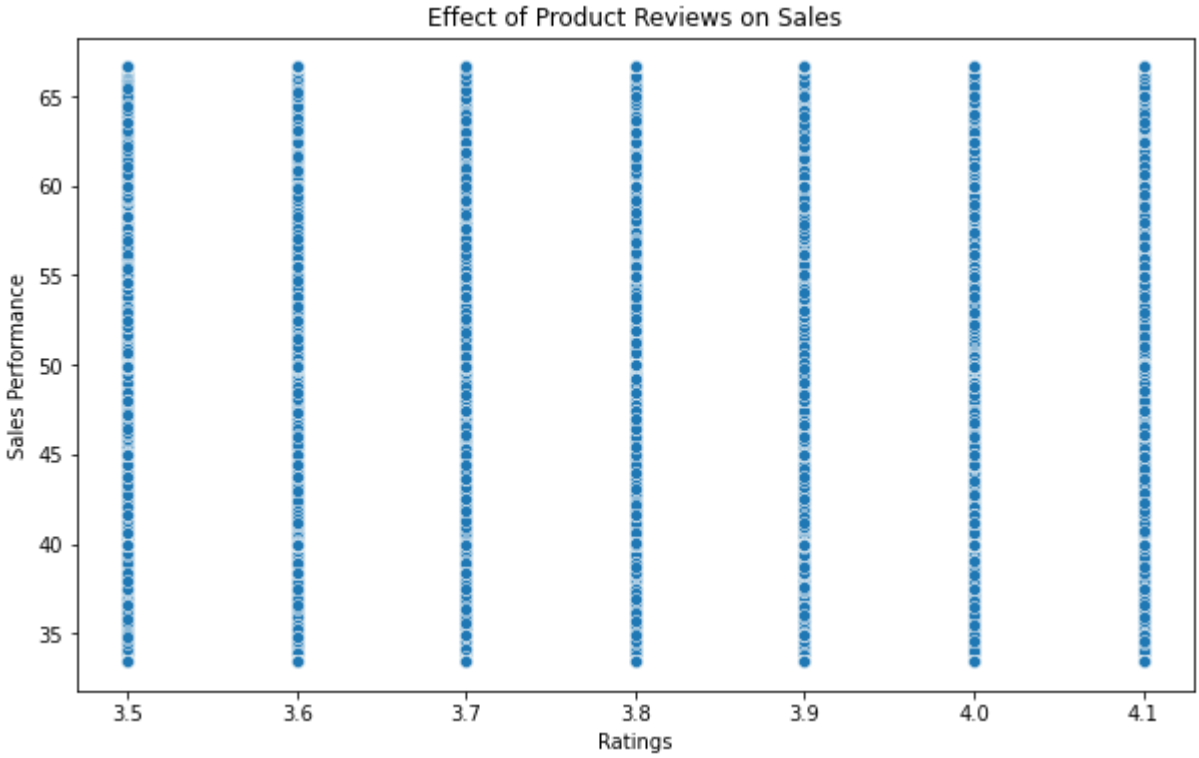
8. Explain the pricing strategies you observe among top sellers?

A. Insight:

The pricing strategy observed among top sellers appears to follow a linear trend. This suggests that top-selling products tend to maintain a consistent pricing structure that aligns with the perceived value of the product in the eyes of consumers.

9. Explain the relationship between ratings and sales performance?

A.

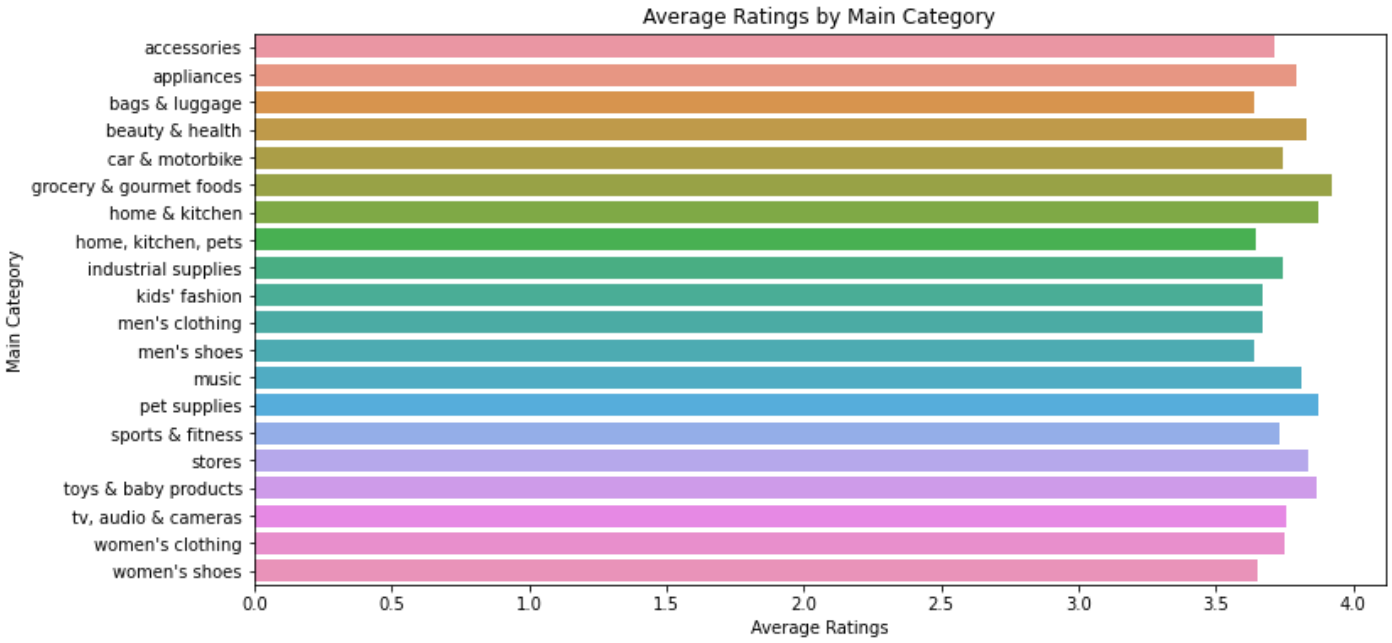


Insight:

Specifically, we can observe that a significant concentration of products falls within the rating range of 3.5 to 4.1, and this concentration is represented by a dense cluster of data points in the scatter plot.

10. Explain which main categories have high average ratings?

A. We can observe that the "Grocery & Gourmet Foods" category has a notably high average rating, while the "Home, Kitchen & Pets" category has a comparatively lower average rating.

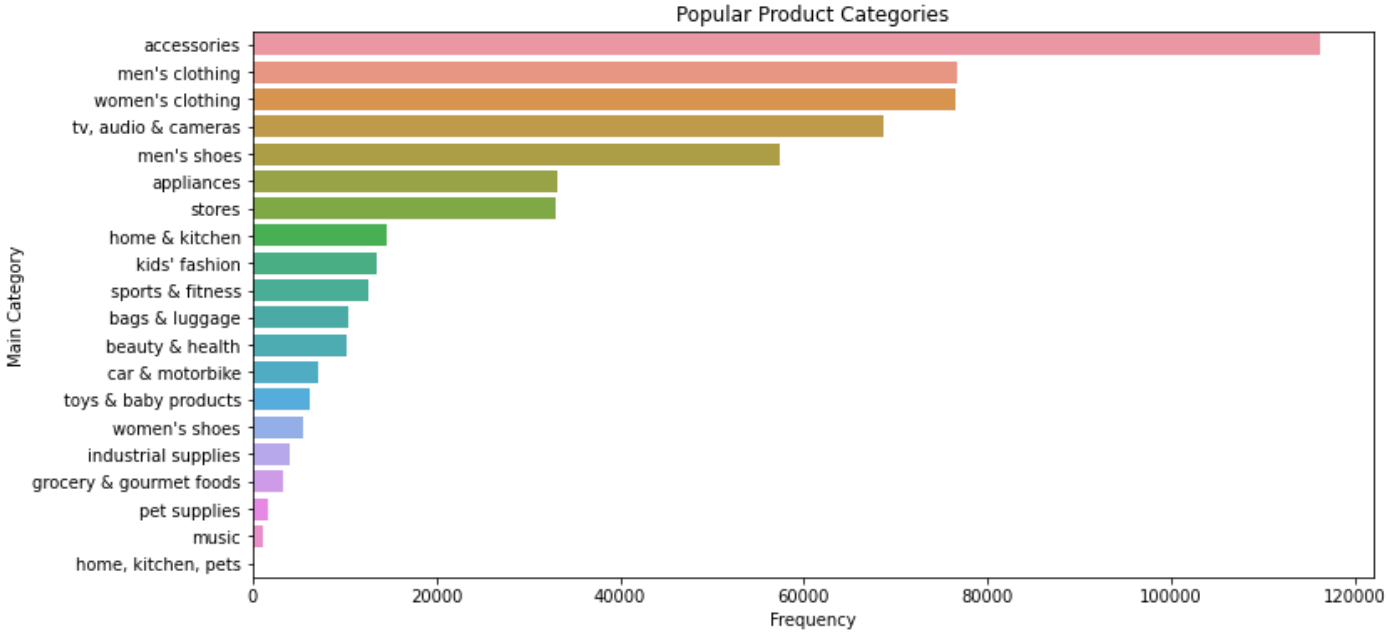


11. Explain how ratings impact the sales performance of products?

A. We can observe that products with ratings ranging from 3.0 to 4.6 tend to exhibit high sales performance.

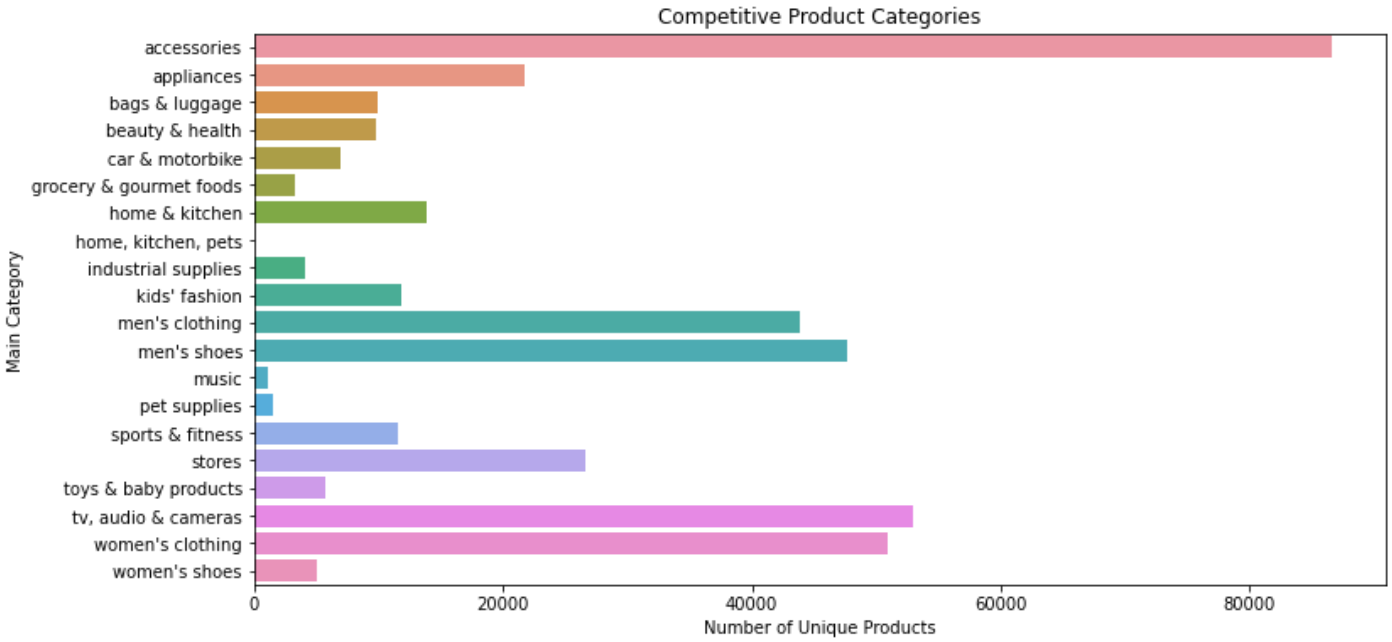
12. Explain which main categories are the most popular based on frequency?

A. The analysis of main categories based on frequency reveals insights into the popularity of different product categories in the dataset. Specifically, we observe that the "Accessories" category has a higher frequency, indicating its popularity, while the "Home, Kitchen & Pets" category has a lower frequency.



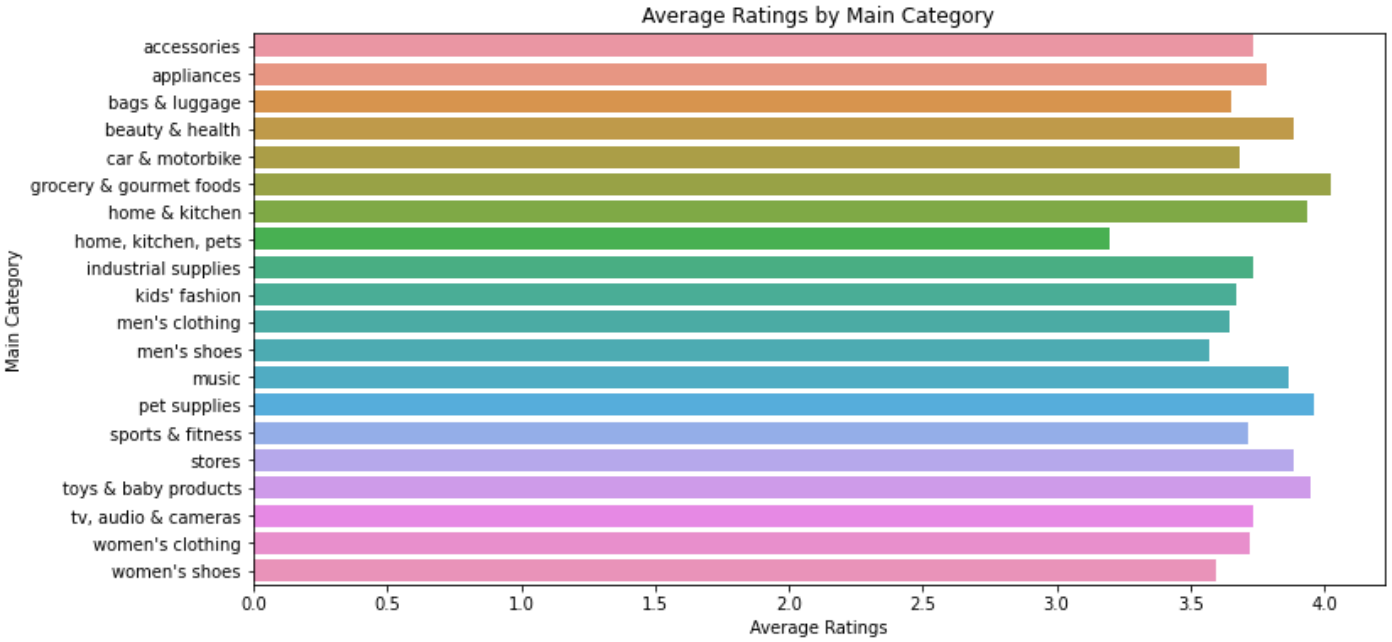
13. Explain which main categories are the most competitive based on the number of unique products?

A. The analysis of main categories based on the number of unique products reveals insights into the level of competition within each category. Specifically, we observe that the "Accessories" category has a higher number of unique products, indicating a high level of competition, while the "Home, Kitchen & Pets" category has a lower number of unique products, suggesting lower competition.



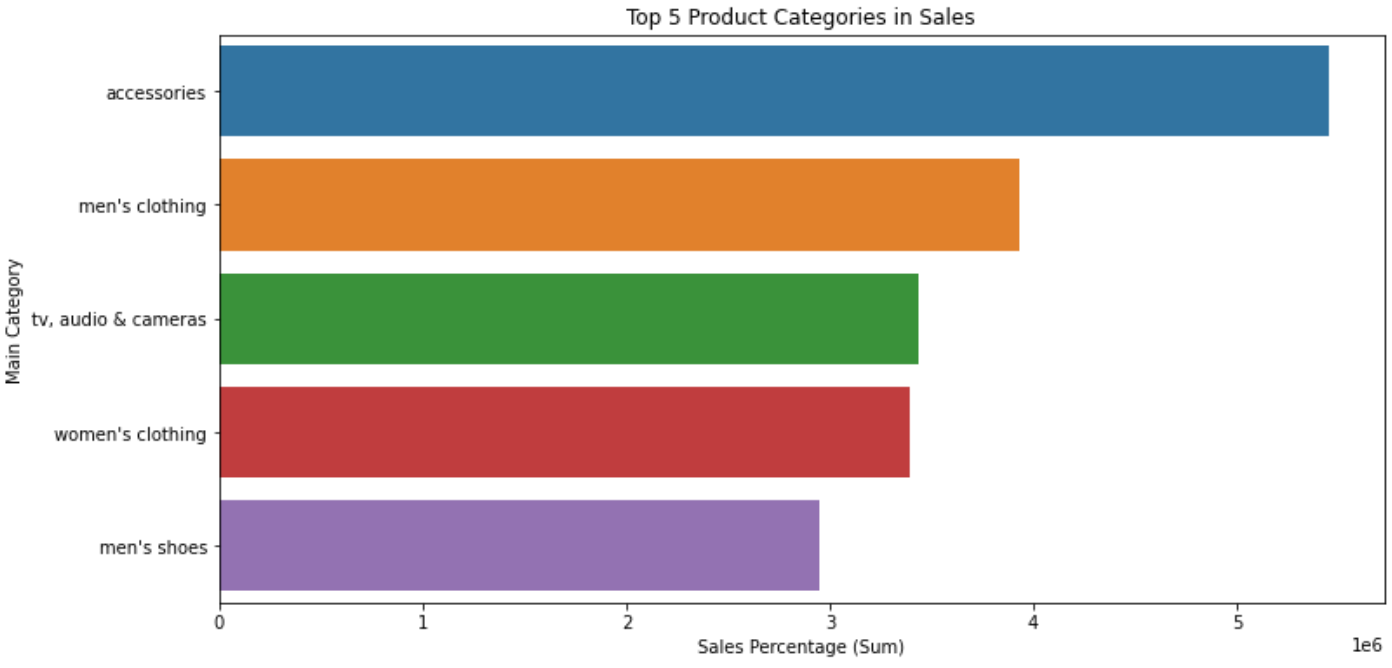
14. Explain the relationship between average ratings and main categories?

A. The analysis of the relationship between average ratings and main categories provides insights into how different product categories are perceived by customers. Specifically, we observe that the "Grocery & Gourmet Foods" category tends to have higher average ratings, while the "Home, Kitchen & Pets" category has lower average ratings.



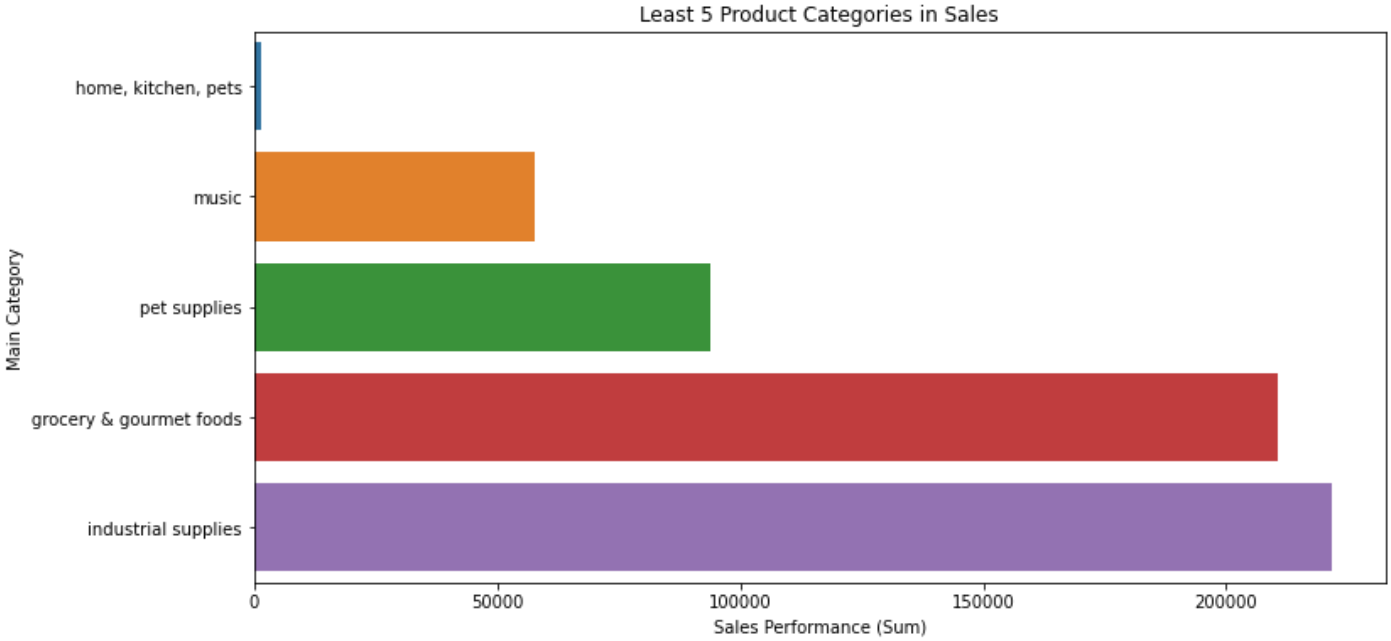
15. Explain the top 5 main categories with the highest sales_Performance?

A. The analysis of the top 5 main categories with the highest sales performance provides insights into which product categories generate the most sales. Specifically, we observe that the "Men's Clothing" category has the highest sales performance, while the "TV, Audio & Cameras" category has lower sales performance.

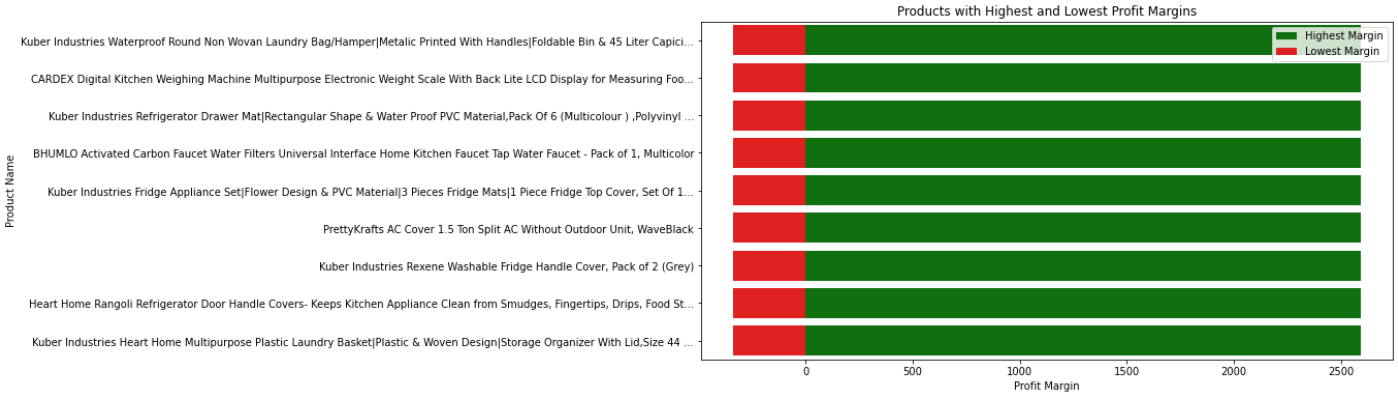


16. Explain the least 5 main categories with the lowest sales performance?

A. The analysis of the least 5 main categories with the lowest sales performance reveals insights into which product categories generate the least sales. Specifically, we observe that the "Women's Shoes" category has relatively higher sales performance among the lowest performers, while the "Home, Kitchen & Pets" category exhibits lower sales performance.

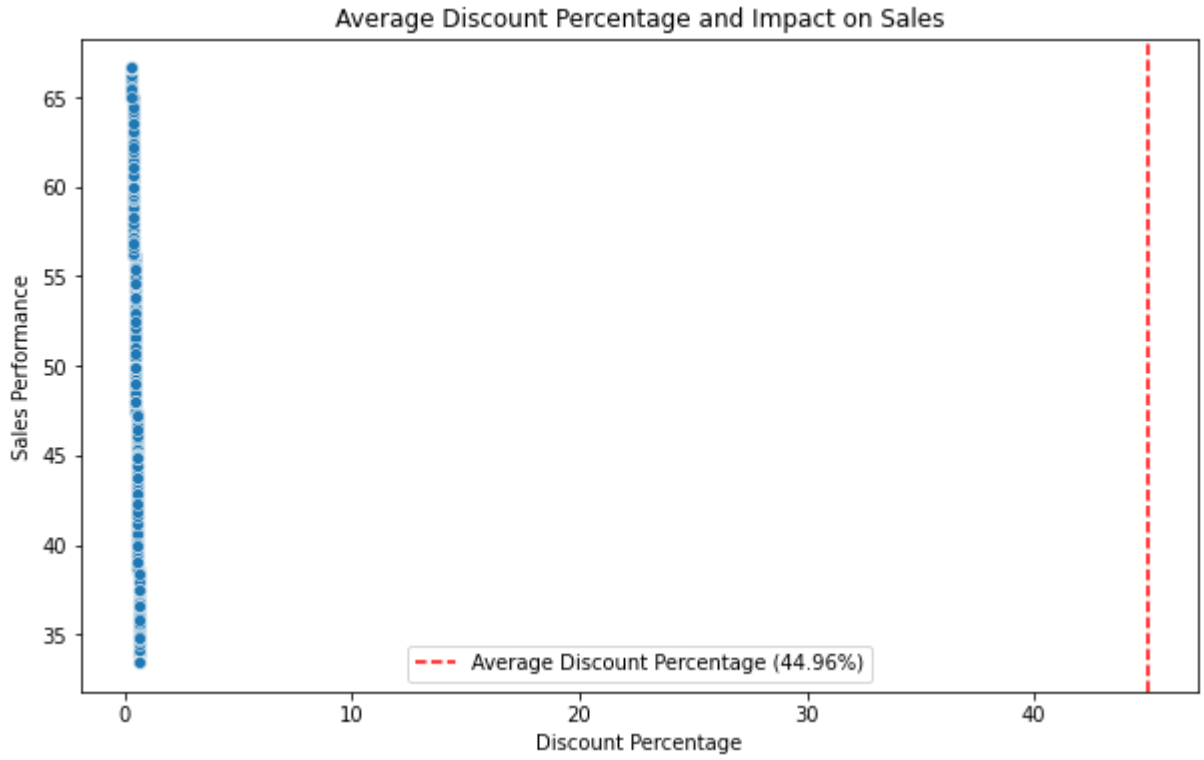


17. Explain the products with the highest and lowest profit margins?



18. Explain the relationship between discount percentage and sales_performace?

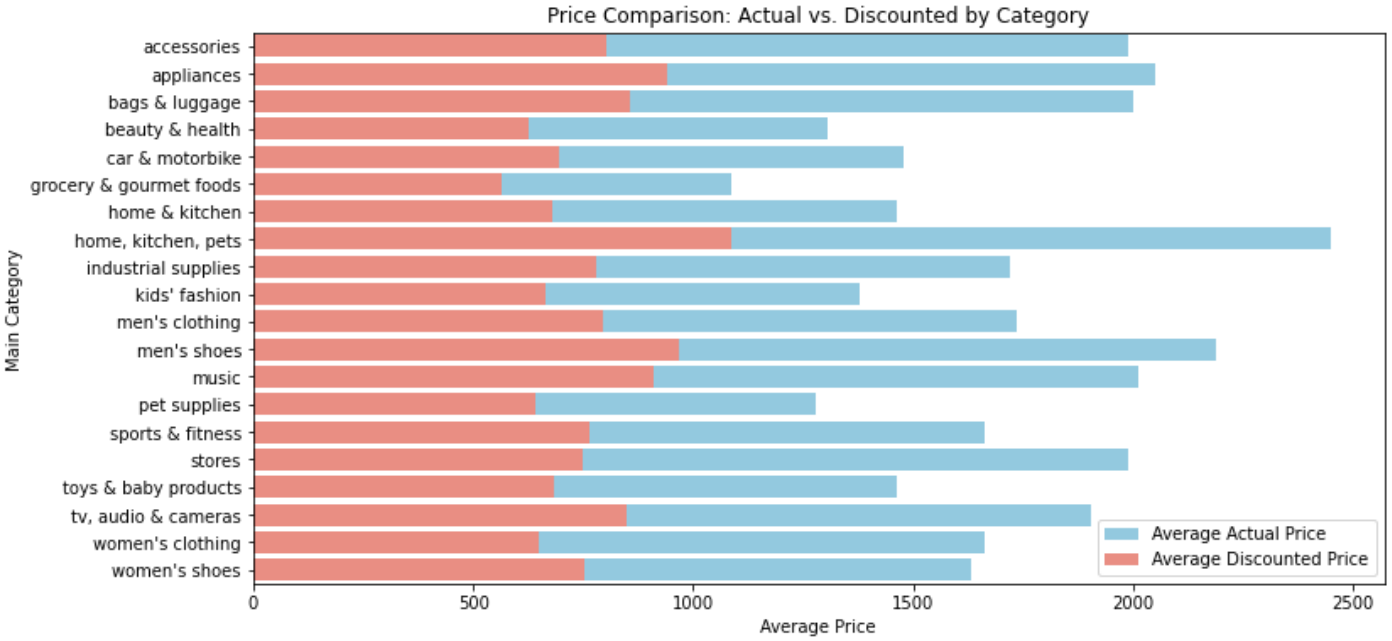
A. The average discount percentage is 44.96%, indicating heavy reliance on discounts by retailers to boost sales. While discounts can increase short-term sales, they can also reduce profit margins and create customer expectations.

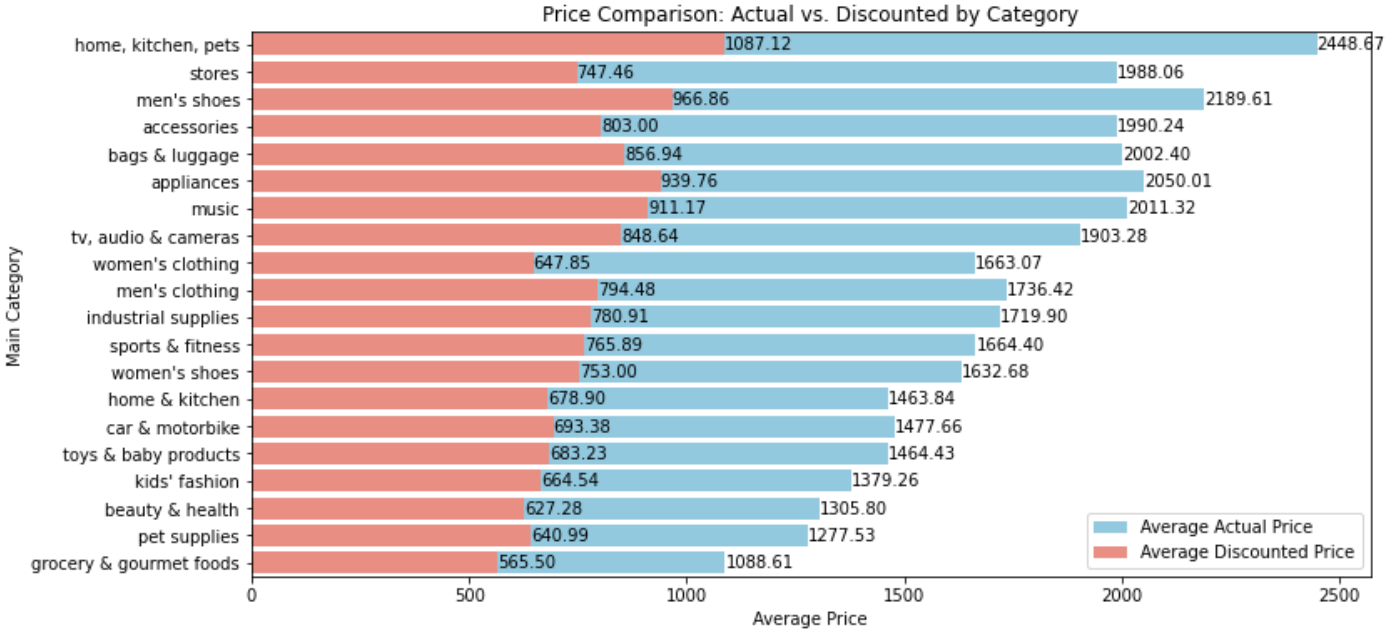


20.Explain how actual prices compare to discounted prices by category?

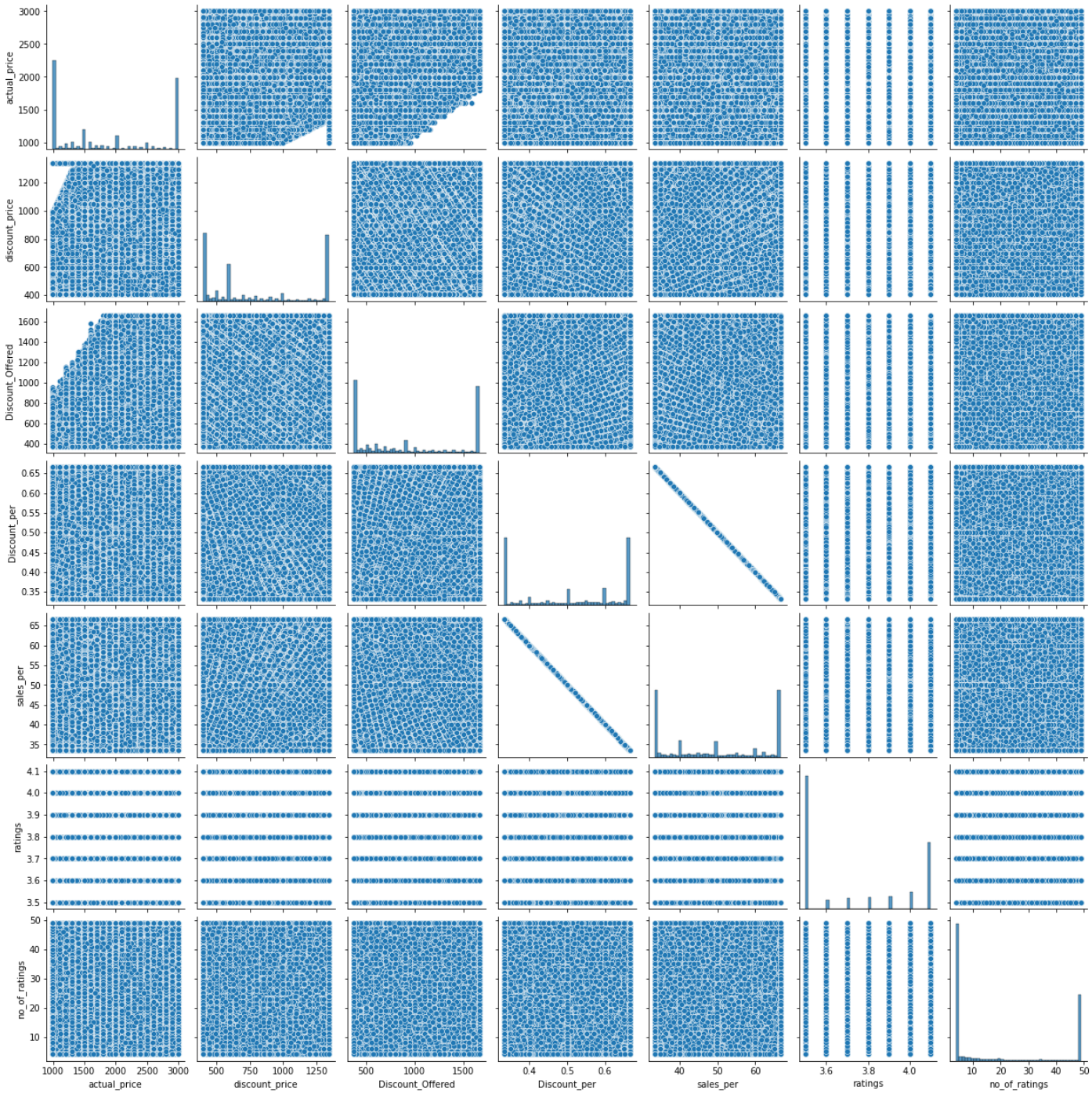
A. The analysis of how actual prices compare to discounted prices by category provides insights into the pricing strategies employed within different product categories. Specifically, we observe variations in discount levels across categories, with some categories offering high discounts and others offering lower discounts.

B. High discount in Home,Kitchen,pets and low discount in grocery & gournment foods

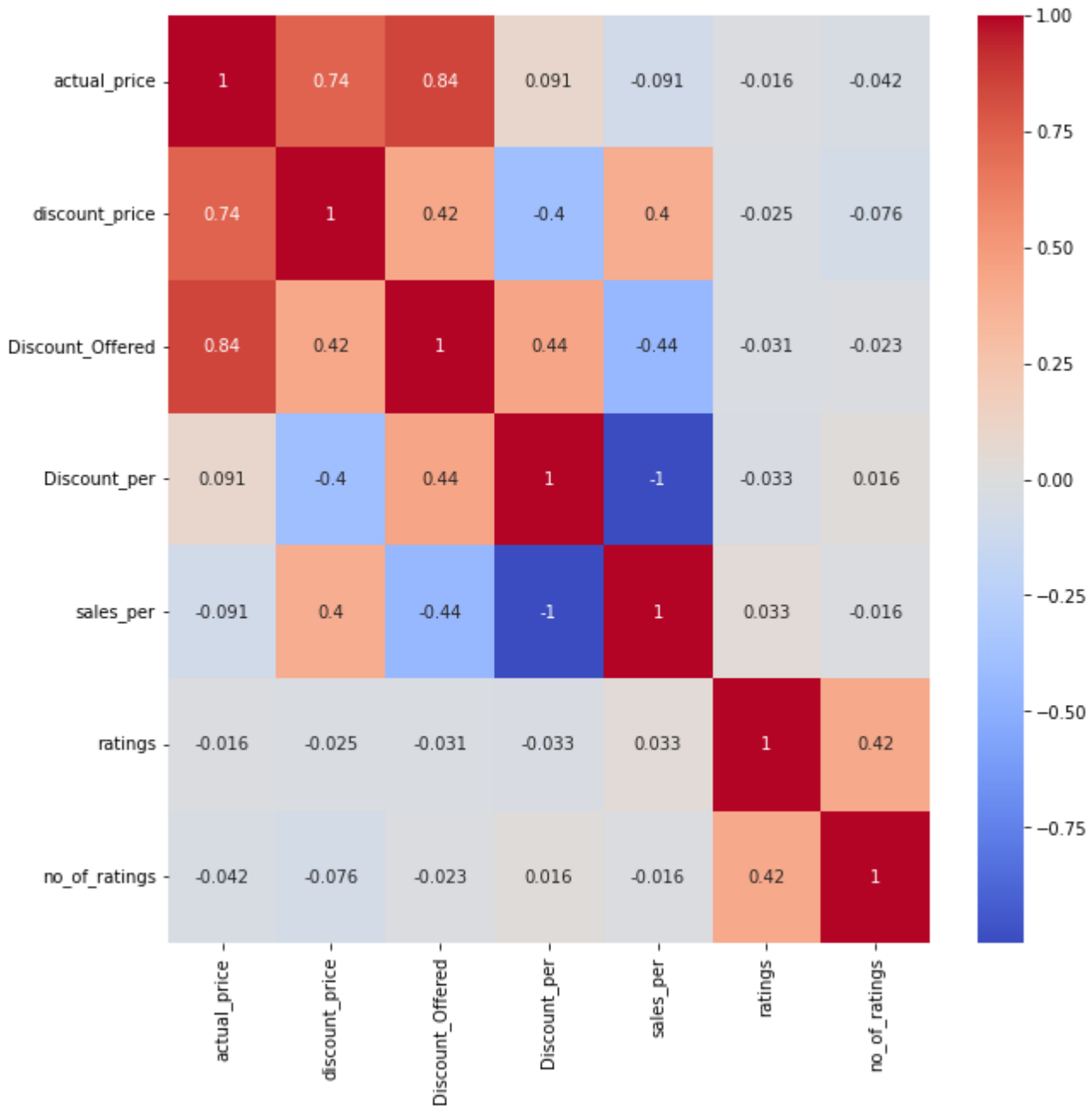




Multi – Variate Analysis



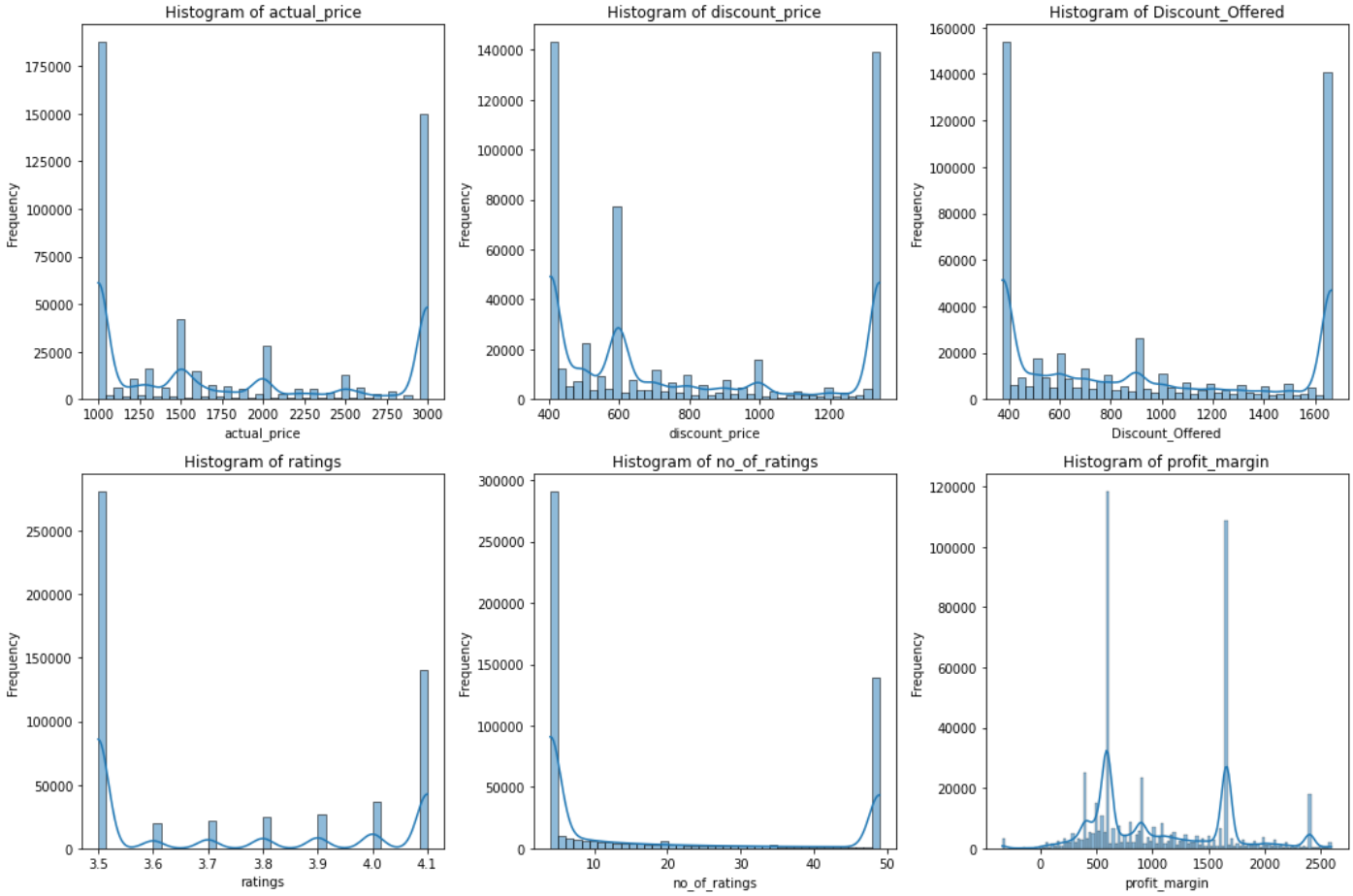
Correlation :



The correlation matrix provides a measure of the linear relationships between pairs of variables.

For example, actual_price and discount_price have a strong positive correlation, indicating that as the actual price increases, the discount price also tends to increase. On the other hand, ratings and actual_price have a negative correlation, suggesting that higher-priced items tend to have lower ratings.

Distribution:



1. **Actual Price**: The distribution is right-skewed, indicating that most products have a lower actual price, with fewer products having a higher actual price.
2. **Discount Price**: This distribution is also right-skewed, suggesting that most products have lower discount prices.
3. **Discount Offered**: The distribution appears to be right-skewed, indicating that a higher discount is offered on fewer products.
4. **Ratings**: The distribution seems to be left-skewed, suggesting that most products have high ratings.
5. **Number of Ratings**: This distribution is heavily right-skewed, indicating that most products have fewer ratings, while a few products have a large number of ratings.

6. **Profit Margin**: The distribution is right-skewed, suggesting that most products have a lower profit margin, with fewer products having a higher profit margin.

   - **None of the variables appear to follow a normal distribution.**

Hypothesis Testing:

1. I perform ANOVA test on the sales_per variable across different main categories.
   Results: (2005.7992183106674, 0.0)
   Insight: The low p-value (0.0) suggests that there is strong evidence to reject the null hypothesis.

2. Two sample T-test to compare the average discount offered between two specific brands
   Results:
   1. t-statistic: 2.477807961109393
      p-value: 0.013439243448152031
      There is a significant difference in the average discount offered on products between Brand A (LG) and Brand B (Lloyd).

   2. t-statistic: -0.34868844825295814
      p-value: 0.7275022030049723
      There is no significant difference in the average discount offered on products between Brand A (Voltas) and Brand B (Daikin).

   3. t-statistic: -5.181798059056013
      p-value: 2.8370123633526074e-07
      There is a significant difference in the average discount offered on products between Brand A (Panasonic) and Brand B (Whirlpool).

3. PEARSON CORRELATION

The test conducted on average sales percentage and the number of ratings

Results:

Correlation coefficient: -0.016155496373994296

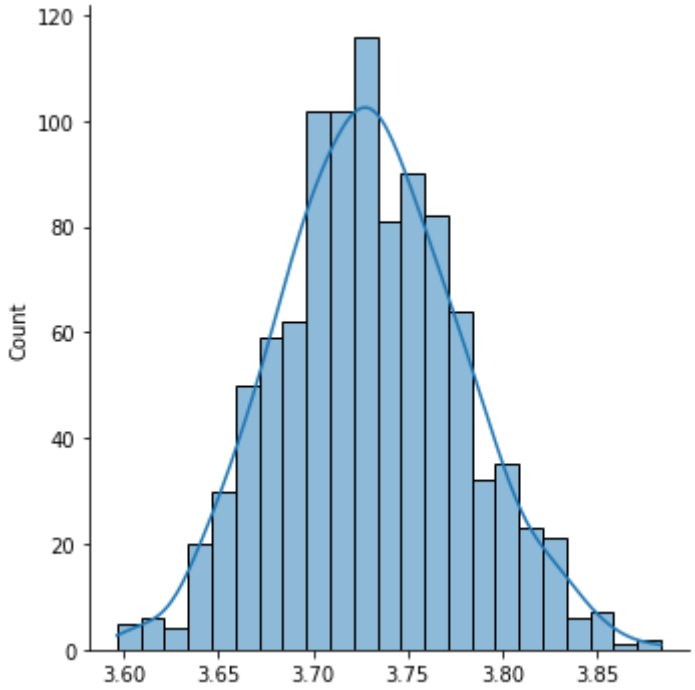
p-value: 3.585563054203087e-33

There is no significant relationship between the average sales percentage product and the number of ratings a product has.

Insight:

It appears that the correlation coefficient between the average sales percentage and the number of ratings is quite low (-0.016). The p-value is extremely small (3.585e-33), concludes that there's no significant relationship between the average sales percentage and the number of ratings a product receives. Even though the correlation coefficient indicates a weak negative relationship, it's so close to zero that it doesn't hold statistical significance.

Central Limit Theorem Influence on Average Ratings



It is a distribution of sample means, which is a histogram with a Kernel Density Estimate (KDE) overlay. This type of distribution is known as a Sampling Distribution.

From the plot, it appears that the distribution of sample means follows a Normal Distribution (also known as a Gaussian Distribution), which is characterized by its bell shape and symmetry around the mean. This is in line with the Central Limit Theorem, which states that the distribution of sample means will approximate a normal distribution as the sample size becomes large, regardless of the shape of the population distribution.

The mean of this distribution is approximately 3.75, which is the peak of the bell curve. This suggests that the average rating across all samples is around 3.75. The spread of the distribution gives us an idea about the variability of the ratings. The narrower the bell curve, the less variability there is in the ratings.

Population mean: 3.7304678336067876

Mean of sample means: 3.729793333333333

Standard deviation of your sample means: 0.047939660569302 48

standard error of the population: 0.047890091664861544

The values were very close, which supports the CLT. This implies that the sample means provide a good estimate of the population mean, and the spread of the sample means around this estimate is captured by the standard deviation of the sample means (or equivalently, the standard error of the population).

Findings and Insights:

1. Most and Least Selling Brands:
    1. Most sales are attributed to the PC brand.
    2. 'ARTELLY', 'ITDC', 'SOULMOTO-[PCS-1]', 'Orca', 'Dreamkraft', 'Kapur', 'Revant', 'S-Biv', 'Dequera®', 'Obsessions' are among the least selling brands.
2. Top Selling Main Category:
    1. Accessories stand out as the most selling main category.
3. Most Popular Product:
    1. 'Zeya Yellow Gold Ring' emerges as the most popular product on Amazon.
4. Top 10 Popular Products by Ratings:
    1. Lists the top 10 highly rated products including 'Hitachi Split Ac', 'Panasonic 2 Ton 5 Star Wi-Fi Inverter', and others.
5. Rating Distribution Insights:
    1. The concentration of ratings falls predominantly within the range of 3.5 to 4.1.
6. Pricing Strategy:
    1. The pricing strategy among top sellers appears consistent, aligning with perceived product value.
7. Analysis of Average Ratings:
    1. Indicates that the average rating of products is notably high.

8. Category-based Ratings:

    1. "Grocery & Gourmet Foods" displays notably high average ratings, while "Home, Kitchen & Pets" tends to have a comparatively lower average.

9. Sales Performance and Ratings:

    1. Products within the 3.0 to 4.6 ratings range demonstrate high sales performance.

10. Main Category Insights - Frequency:

    1. "Accessories" appears as a popular category with higher frequency, while "Home, Kitchen & Pets" has a lower frequency.

11. Competition Level within Categories:

    1. "Accessories" show a higher number of unique products, indicating higher competition. Conversely, "Home, Kitchen & Pets" exhibits a lower count of unique products, suggesting lower competition.

12. Average Ratings across Categories:

    1. "Grocery & Gourmet Foods" tends to have higher average ratings, whereas "Home, Kitchen & Pets" typically receives lower average ratings.

13. Top and Least Sales Performing Categories:

    1. "Men's Clothing" tops in sales performance, while "TV, Audio & Cameras" falls behind. Conversely, "Women's Shoes" holds relatively higher sales performance among the least performers, alongside "Home, Kitchen & Pets."

14. Discount Strategy and Implications:

    1. The average discount percentage at 44.96% signifies a reliance on discounts for sales increase. However, this strategy could impact profit margins and shape customer expectations.

15. Pricing Strategies Across Categories:

    1. Varying discount levels observed across categories, notably high in "Home, Kitchen & Pets" and lower in "Grocery & Gourmet Foods."

Trends:

1. Customers seem to consistently rate grocery and gourmet food products more positively compared to home, kitchen, and pet items.
2. Higher-rated products tend to perform better in sales, indicating a positive correlation between ratings and sales figures.
3. The high frequency and unique product count imply a competitive market landscape in the accessories category, while "Home, Kitchen & Pets" seems less competitive.
4. Retailers heavily rely on discounts to boost sales, especially evident in home, kitchen, and pet categories, potentially impacting profit margins.

Limitations:

1. **Assumption of Normality:**
    - The assumption of a normal distribution, especially in the context of sales, ratings, or discounts, might not hold true in all cases.

2. **Subjectivity in Ratings:**

- Consumer ratings are subjective and influenced by various factors. Understanding these subjective components might enhance the analysis.

3. Lack of Segmentation:

- Deeper segmentation within categories or products could reveal more nuanced insights. Aggregate analysis might overlook specific trends.

CONCLUSION:

Key Takeaways:

1. **Category Dynamics:**

- "Grocery & Gourmet Foods" consistently garners higher average ratings, while "Home, Kitchen & Pets" tends to have comparatively lower ratings.

2. **Consumer Preferences:**

- Products with ratings between 3.0 and 4.6 show stronger sales performance, indicating the influence of customer ratings on sales.

3. **Market Popularity and Competition:**

- "Accessories" stands out as a highly competitive category with both high frequency and numerous unique products, while "Home, Kitchen & Pets" appears less competitive.

4. **Sales Performance and Category Trends:**

- "Men's Clothing" excels in sales performance, while "Women's Shoes" is relatively stronger among the least performing categories.

5. **Pricing and Discounts:**

- Heavy reliance on discounts, notably in "Home, Kitchen & Pets," suggests a pricing strategy aimed at boosting sales.

6. **Rating and Sales Correlation:**

- Higher-rated products consistently drive better sales, emphasizing the impact of customer feedback on purchase decisions.

References:

1. Kaggle
   https://www.kaggle.com/code/nishubh90/amazon-products-sales-dataset-2023/notebook
2. AI ASSISTANTS(BARD, CHATGPT)
3. DAILY DOSE OF DATA SCIENCE BLOG
4. AUTOMATED EDA LIBRARIES LIKE YDATA-PROFILING,AUTOVIZ
5. MATPLOTLIB FOR ERROR CORRECTIONS
   https://matplotlib.org/stable/users/getting_started/
6. ANALYTICS VIDHYA PLATFORM

PROJECT CODE:

- Jupyter notebook link:
  https://drive.google.com/file/d/1NWd8QCqgHodR10wbS7z9wJFBGM_KOcX0/view?usp=sharing
- Dataset link:
  https://drive.google.com/file/d/1EDTL9uqiooon3OeJY2EuO9lCNuws9QZO/view?usp=sharing
- Presentation
  link:https://docs.google.com/presentation/d/1BZynVIKJ9stwxmsUURBlJDzd2laff6vF/edit?usp=sharing&ouid=105676892097579482165&rtpof=true&sd=true