Amazon Product Sales

Data Analysis Report

Donatus Enebuse, August 2023

Introduction & Objectives

- Purpose of analytics:
  - What are the trends in product sales?
  - What are the factors that influence product sales?

- Research questions
  - What are the most popular product categories?
  - What are the most popular product sub-categories?
  - What are the most popular products by ratings?
  - What are the most popular products by pricing strategy?
  - What are the top selling products?
  - What are the least selling products?

Data Collection & Cleaning

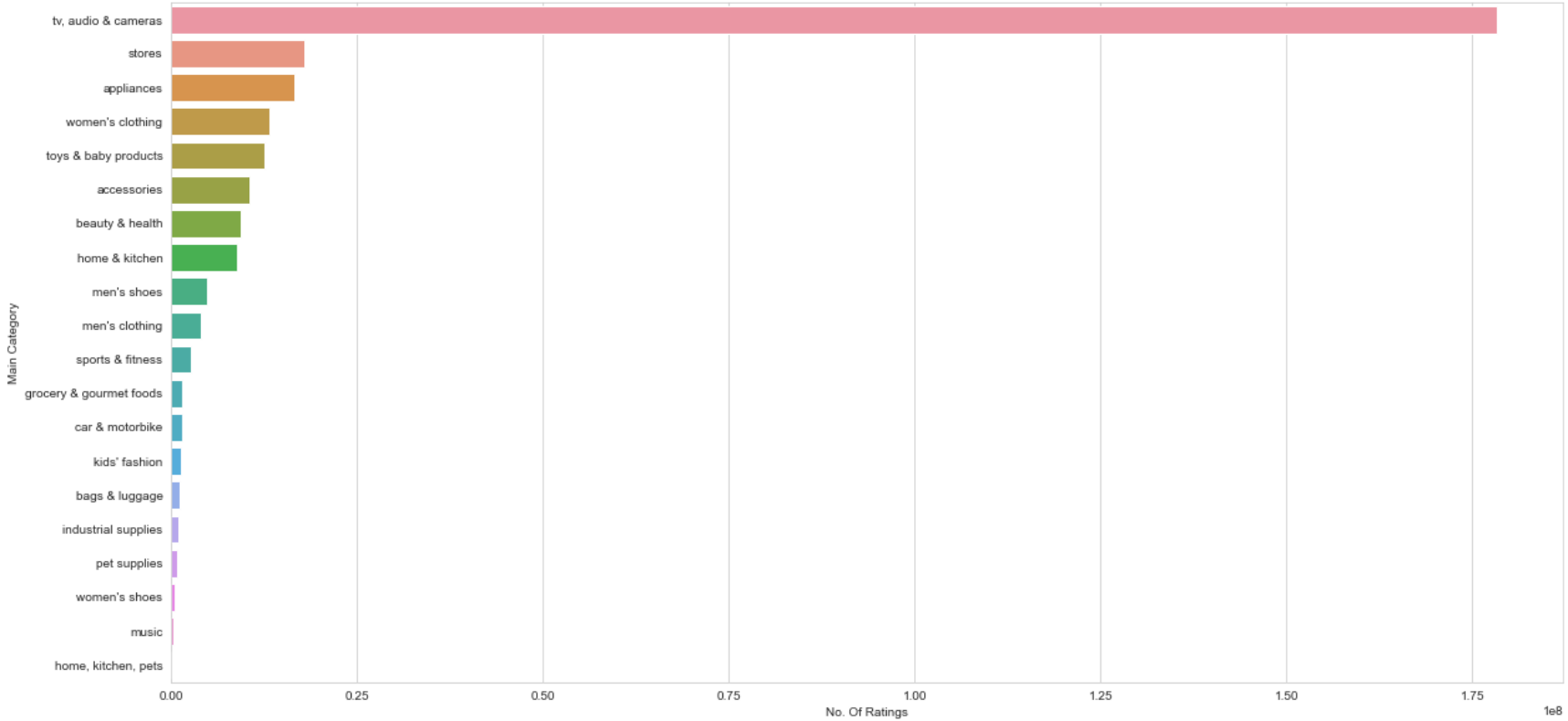
- Data Source:
  - https://www.kaggle.com/datasets/rajeevmishragwl/amazon-product-sales
  - Dataset contains 551585 records, and 10 fields
- Steps for cleaning and transformation:
  - Imported the required libraries into Jupyter Notebook
  - Read the dataset into the Python Environment
  - Removed 3 unwanted columns, several missing records, and duplicates
  - Got rid of the rupee sign in two fields, and lots of invalid entries
  - After cleaning, there was 334963 records, and 8 fields.
  - Two new fields (Total Sales, and Percentage Discount) were added to enable better analytics
- Challenges faced:
  - The dataset contained lots of duplicated records
  - The dataset has lots of records that contained invalid entries
  - Many invalid entries had to be removed using advanced techniques like regex

Data Exploration

- Description of the fields:
  - ‘Name’: Product names. 237352 unique values out of the 334963 records.
  - ‘Main Category’: 20 main categories
  - ‘Sub-category’: 112 sub-categories
  - ‘Ratings’: Ranging from 1.0 ~ 5.0
  - ‘No. Of Ratings’: Ranges from 1 ~ 589547. Taken as the ***volume of sales.***
  - ‘Discount Price’: Taken as the ***selling price.*** Ranging from 8 ~ 1249990
  - ‘Actual Price’: The original price of the item. Ranging from 10 ~ 9900000000
  - ‘Total Sales’: Field added as ‘No. Of Ratings’ × ‘Discount Price’
  - ‘Percentage Discount’: Field added as (‘Actual Price’ − ‘Discount Price’) / ‘Actual Price’ × 100%

- Special note:
  - The only item whose Actual Price is 9900000000 was found to be a Mosquito Killer. This information is apparently wrong but was ignored as it had no effect on the analytics.

Insights and Visualization 1



Insights and Visualization 2

The popular product **Categories** that seemed to generate most sales and ratings were observed to be:

1) TV, audio & cameras

2) Stores

3) Appliances

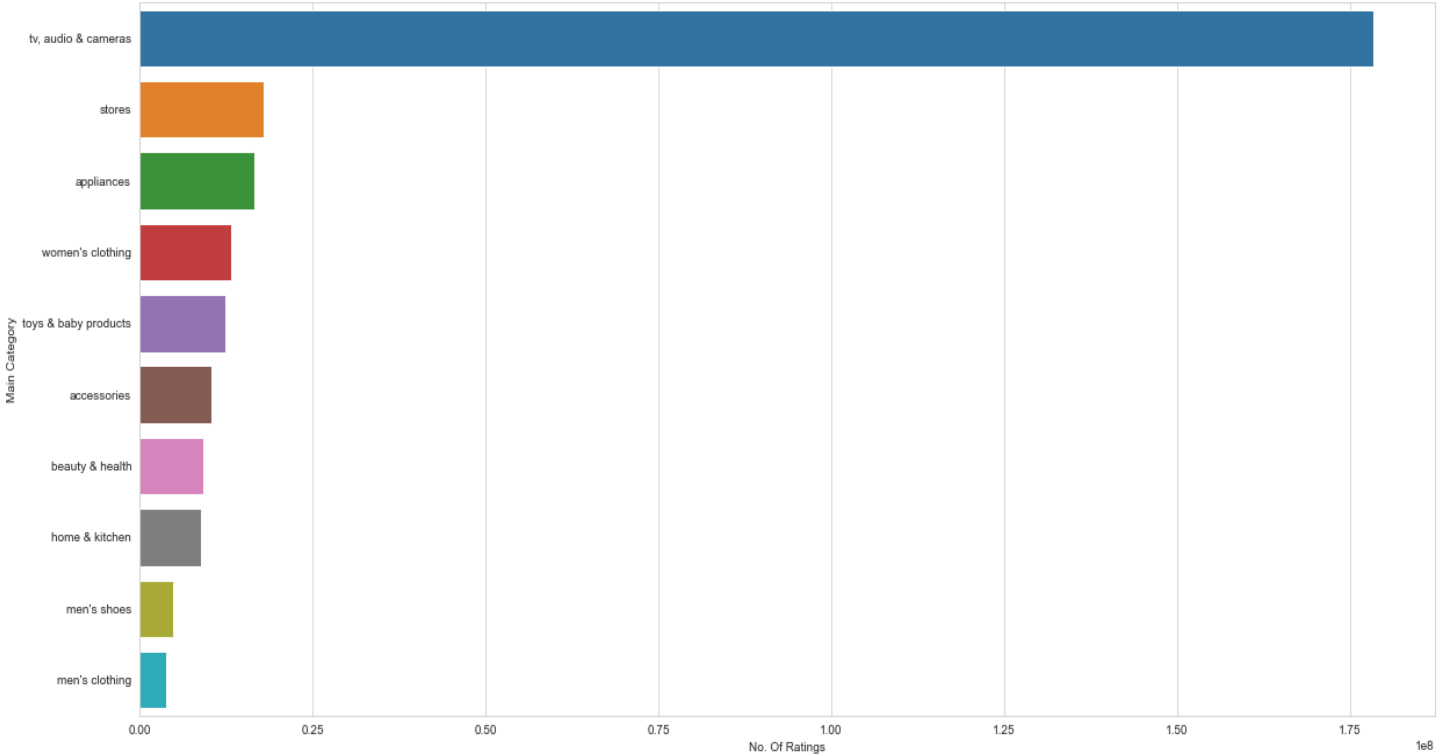
4) Women's clothing

5) Toys & baby products

6) Accessories

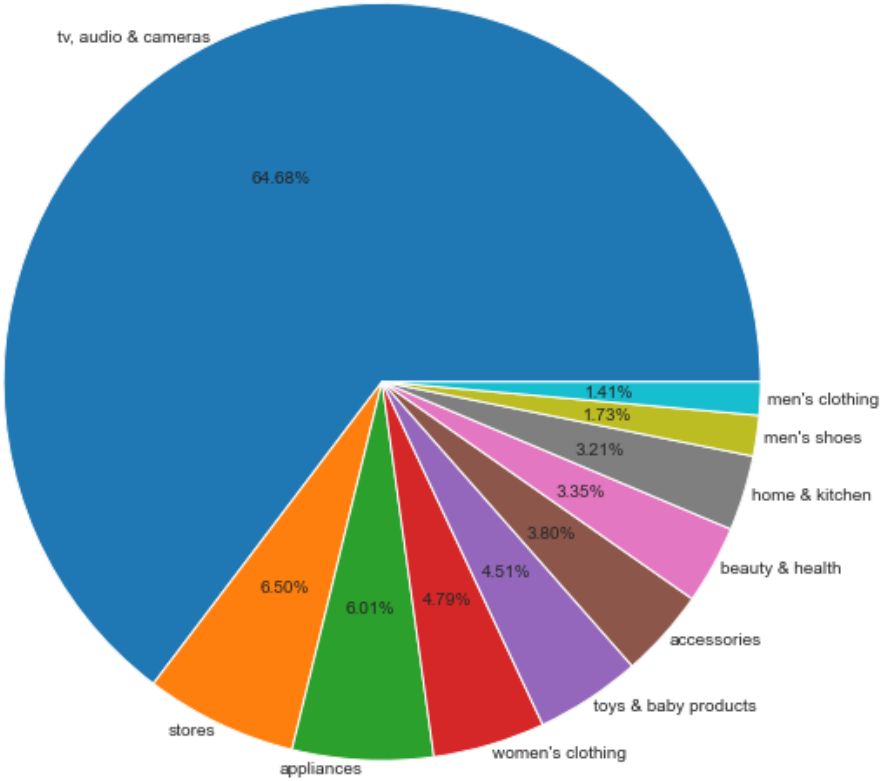
7) Beauty & health

8) Home & kitchen



Insights and Visualization 3

Electronics/Equipment/Appliances take up approx. 75% of the volume of sales!



Insights and Visualization 4

The popular product **Sub-categories** that seemed to generate most sales and ratings were observed to be:

1) All Electronics

2) Headphones

3) Cameras

4) Camera Accessories

5) Home Entertainment Systems

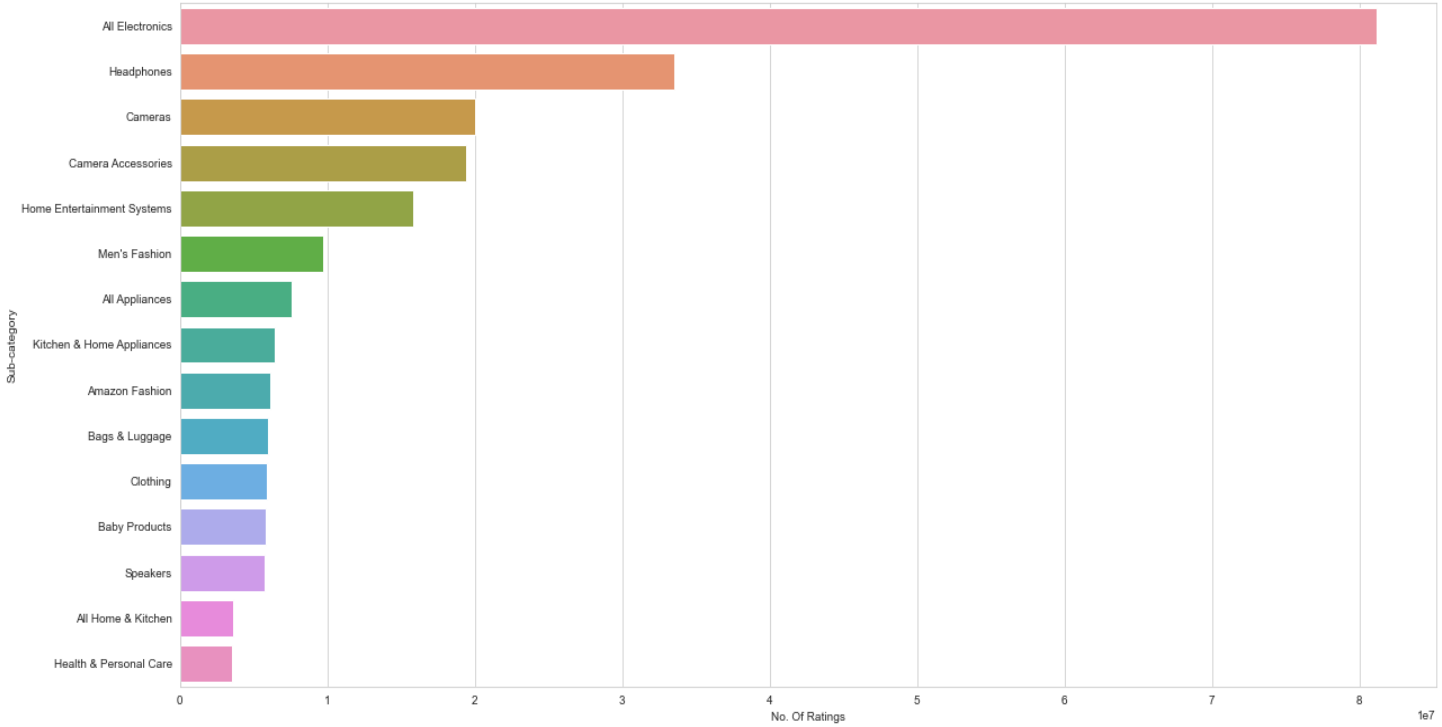
6) Men's Fashion

7) All Appliances

8) Kitchen & Home Appliances

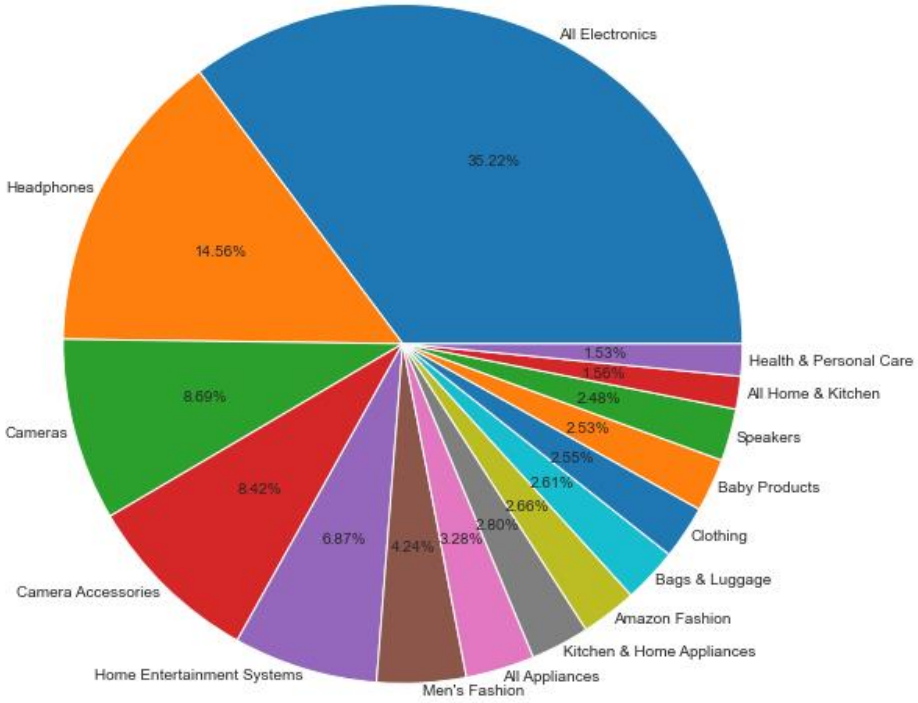
9) Amazon Fashion

10) Bags & Luggage



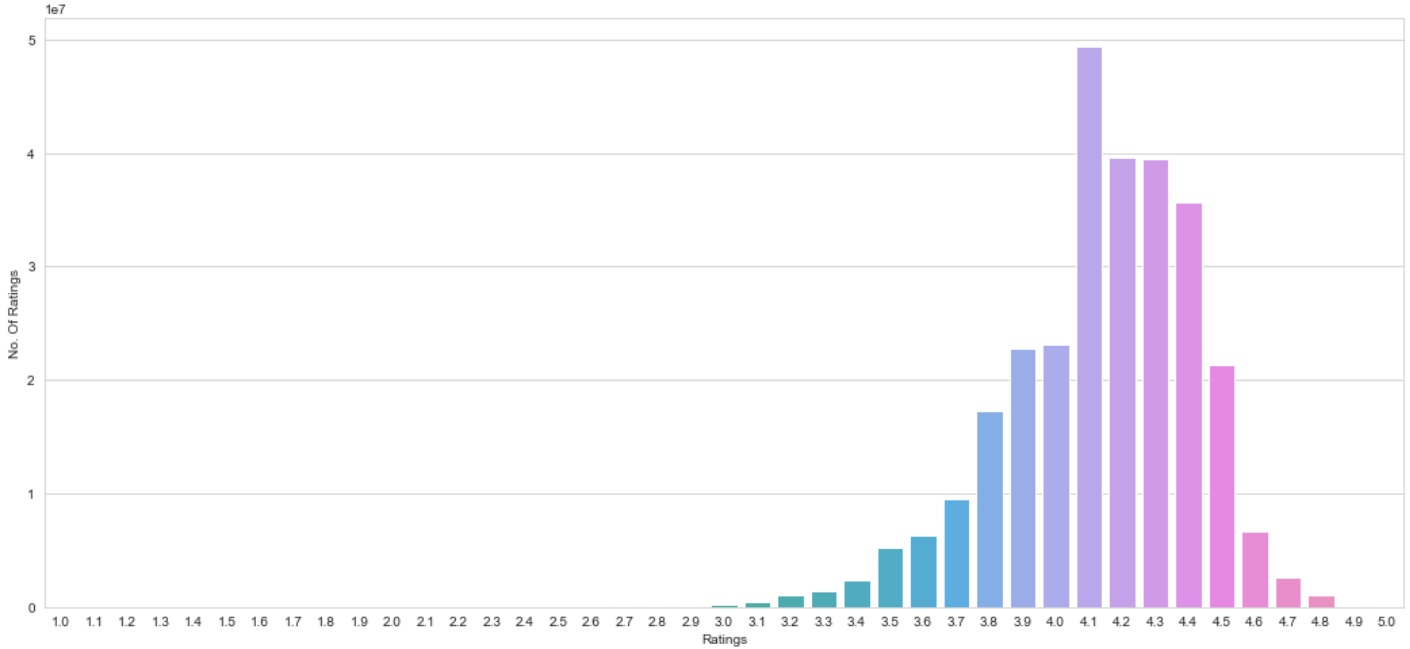
Insights and Visualization 5

As seen again, Electronics/Equipment/Appliances take up approx. 75% of the volume of sales!



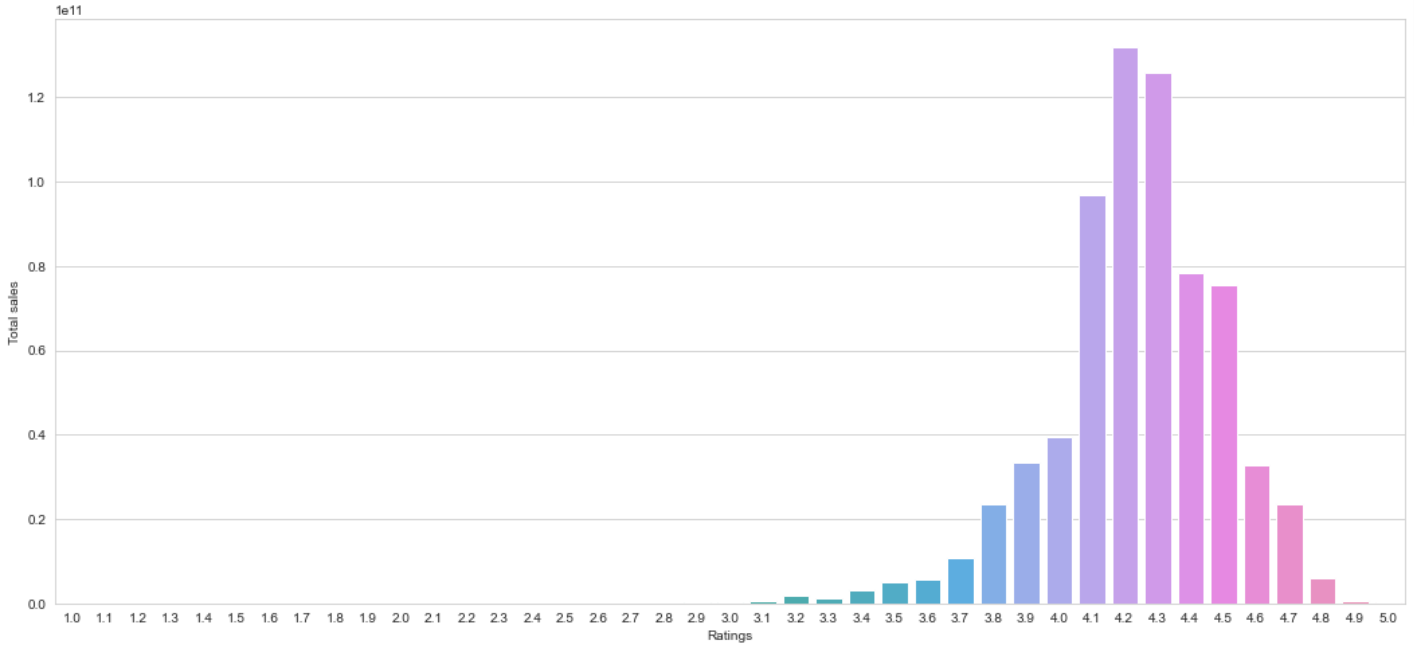
Insights and Visualization 6

The main volume of sales came from products rated from 3.2 to 4.8



Insights and Visualization 7

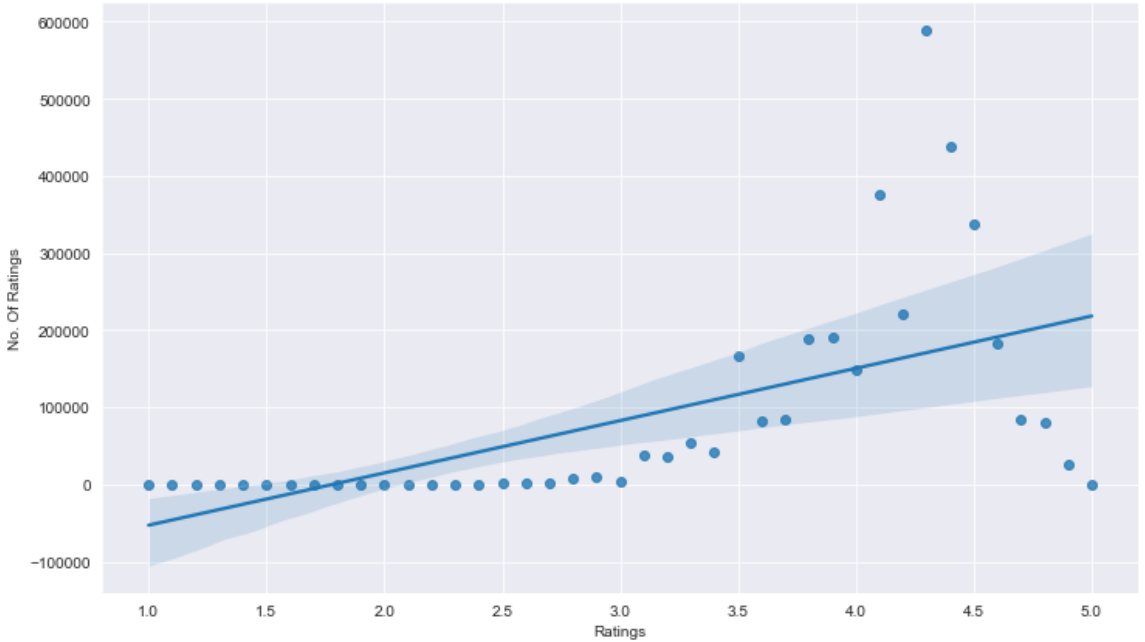
The main Total Sales came from products rated from 3.2 to 4.8



Insights and Visualization 8

There is a moderate correlation between Ratings and Sales Volume.

Correlation coefficient = 0.476932 (Pearson)



Insights and Visualization 9

The 10 most popular products by ratings:

- SanDisk 400GB Class 10 MicroSD Card (SDSQUAR-400G-GN6MA) with Adapter
- Redmi 164 cm (65 inches) 4K Ultra HD Android Smart LED TV X65 | L65M6-RA (Black)
- Redmi 9A Sport (Carbon Black, 2GB RAM, 32GB Storage) | 2GHz Octa-core Helio G25 Processor
- Redmi 139 cm (55 inches) 4K Ultra HD Android Smart LED TV X55 | L55M6-RA (Black)
- OnePlus Nord CE 2 Lite 5G (Blue Tide, 8GB RAM, 128GB Storage)
- Apple iPhone 12 (128GB) - Green
- Fitbit FB507RGPK Versa 2 Health & Fitness Smartwatch
- Samsung Galaxy M12 (Black,6GB RAM, 128GB Storage) 6 Months Free Screen Replacement for Prime
- TOZO T6 True Wireless Earbuds Bluetooth Headphones Touch Control
- MI 138.8 cm (55 inches) 5X Series 4K Ultra HD LED Smart Android TV L55M6-ES (Grey)

Insights and Visualization 10

The 10 most popular product by discount pricing strategy are:

- SanDisk 256GB Class 10 MicroSD Card with Adapter (SDSQUAR-256G-GN6MA)
- SanDisk 512GB Ultra MicroSDXC UHS-I Memory Card
- Amazon Basics High-Speed HDMI Cable, 6 Feet (2-Pack),Black
- Amazon Basics High-Speed HDMI Cable - 10 Feet (2-Pack)
- boAt BassHeads 100 Delhi Capitals Edition Wired in Ear Earphones with Super Extra Bass
- boAt BassHeads 100 in-Ear Wired Headphones with Mic (Black)
- AmazonBasics AAA Performance Alkaline Non-rechargeable Batteries (36 Count)
- Pigeon Polypropylene Mini Handy and Compact Chopper with 3 Blades
- Huggies Complete Comfort Wonder Pants Large (L) Size Baby Diaper Pants
- Huggies Wonder Pants Small (S) Size Baby Diaper Pants

Insights and Visualization 9

Top ~20 selling products are:

- The "SanDisk MicroSDXC Memory Card" products
- The "Amazon Basics High-Speed HDMI Cable" products
- The "boAt BassHeads 100 in-Ear Wired Headphones" products
- The "AmazonBasics Performance Alkaline Non-rechargeable Batteries" products
- The "Pigeon Polypropylene Mini Handy and Compact Chopper with 3 Blades" product
- The "Redmi Note 10S" products
- The "Huggies Wonder Pants, Monthly Box Pack Diapers" products
- The "SanDisk Extreme SD UHS I 32GB Card" products
- The "TOZO T6 True Wireless Earbuds Bluetooth Headphones" products
- The "JBL C100SI Wired in Ear Headphones with Mic" products
- The "SanDisk Ultra USB 3.0 256 GB Pen Drive" products
- The "AmazonBasics Microfiber Bed Sheet Set" products
- The "boAt Airdopes 121 PRO True Wireless in Ear Earbuds" products
- The "TP-LINK WiFi Dongle 300 Mbps Mini Wireless Network USB Wi-Fi Adapter" products
- The "Amazon Pay eGift Card" products
- The "Amazon Basics USB A to Lightning MFi Certified Charging Cable" products
- The "boAt Airdopes 141 Pro True Wireless in Ear Earbuds" products
- The "AmazonBasics 9 Volt Everyday Alkaline Batteries (4-Pack)" product

Insights and Visualization 11

The 10 least selling products are:

- SOOTRA BRANDED carbon film resistor 1/4 watt 4 Band Tolerance 5% - Pack of 20pcs (10K Ohms)
- 10K ohm 1/4 Watt Resistor 1% Tolerance - 10 Pieces Pack
- Okos Gold Plated Blue and Green Meenakari Radha Krishna Rakshabandhan Alloy Rakhi Bracelet
- SBWC Stylish Slim Lice Narrow Comb Fine Tooth Dust Lice Clean Remove Plastic Lice Hair Combs
- Moti Chain Gold Plated Beautiful American Diamond Heart Shape Necklace Golden Chain Pendant for Women
- Kavi The Poetry-Art Project Kavi Cork Copper Wire led String Lights for Any Bottle
- 3386P- 2K Ohm 0.5W-Through Hole Trimmer Potentiometer - Pack of 2
- MENTER Arm Sleeves Men & Women Sunlight Protection Protection from Dust, Pollution, …
- Chetran’s Herbal Tofu 200G | Masala Flavor | Fresh Paneer Alternative | High Protein
- Shelfkey.com ALUMINIUM ELECTROLYTIC CAPACITOR 1uF 63V SK-1uF/63V (B-356G)

Recommendations

- **It was seen that one important trend in product sales is that the buyers tend to buy products from the SAME brand as a top-selling product.**
- **Buyers also tend to buy from well-rated products in categories, as well as products that have higher number of ratings (as seen in the correlation data)**
- **Important factors affecting product sales are:**

a) Well-established product categories such as Electronics, Appliances, and Clothing seem to attract most sales,

b) Discount doesn't seem to encourage sales. In fact it was observed that well discounted products have lower sales (as seen in the correlation coefficient data),

c) Price of item (original price) has no effect on volume of sales.

Limitations & Future work

- **The dataset contained a lot of missing values, invalid entries, and inconsistent data that might have an effect on analysis**

- **More fields like Monthly data, etc might be needed to help obtain trends in product sales with respect to time**

References, citations, & Literature

- No. of Ratings is taken to be volume of sales because it is assumed that only those who buy the products can give ratings.
- A *regplot* is a chart used to show visual correlation between two numerical variables.
- The *most common* correlation coefficient is that of Pearson.

Contact Information

For more information on this Data Analytics Case Study, or Data Analysis Projects, you can reach me on:

Phone: +234-701-292-5971, +234-916-564-0262

My email: donatus.enebuse@gmail.com

My portfolio: https://github.com/vaxdata22

Thank you!