

Amazon Sales Dataset

By Mohammad Al Hussein

Correlation One DA9 Cohort Program



BUSINESS CONTEXT

The primary goal of this project is to analyze the Amazon Sales dataset and identify insights based on the data. The Amazon Sales dataset is a valuable resource for businesses. It provides a great deal of information about customer behavior, product trends, and market conditions. valuable insights can be gained that can help researchers and business owners make better decisions about their products and services.

ANALYSIS

- There are 1413 rows in the data set comprising 72 different product specific categories, prices and customer interactions.
- Customers interactions with products are mostly about Quality and usability.
- Ratings and review counts have no bearing on Product popularity.
- There is almost no Correlation between Discounted Price and rating.

DATA

- Data Source: Kaggle
 - <https://www.kaggle.com/datasets/karkavelrajai/amazon-sales-dataset>
- Size: 2MB
- Variables: product_id, product_name, category, Specific_category, Discounted_price, actual_price, discount_percentage, rating, rating_count, about_product, user_id, user_name, review_id, review_title, review_content, sentiment,Setiment_name, Sentiment_score.

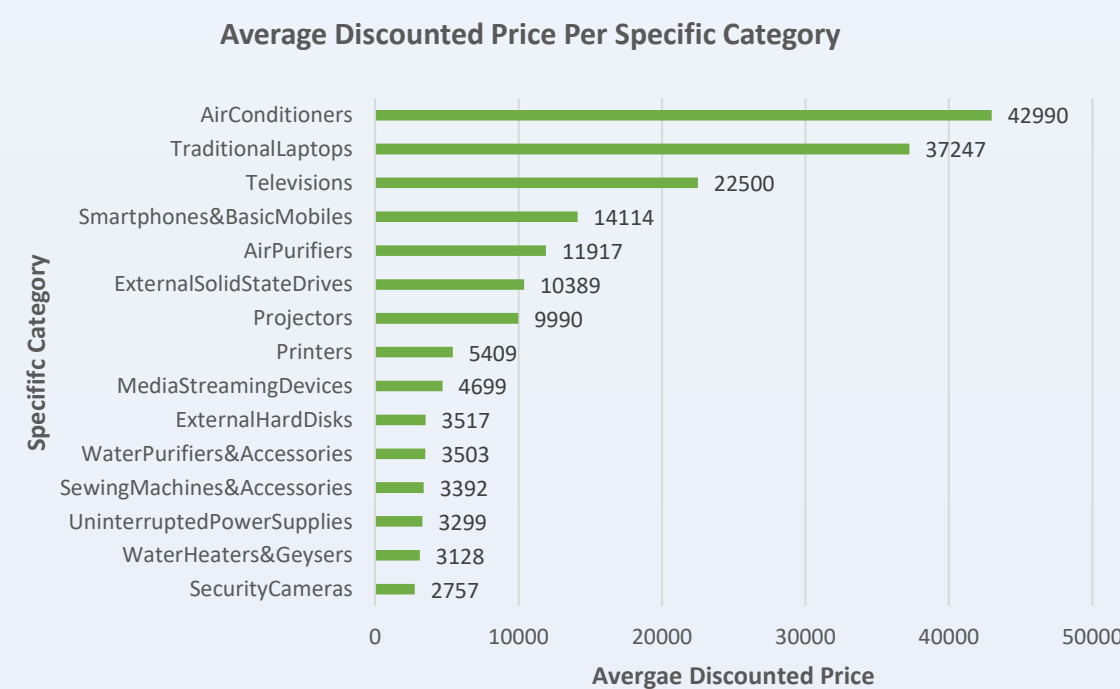
TOP 15 CATEGORIES

Judging from the count of Product ID the chart shows that Cables & Accessories are at the top of the list followed by small kitchen appliance while at the bottom of the list comes paper and water purifiers.



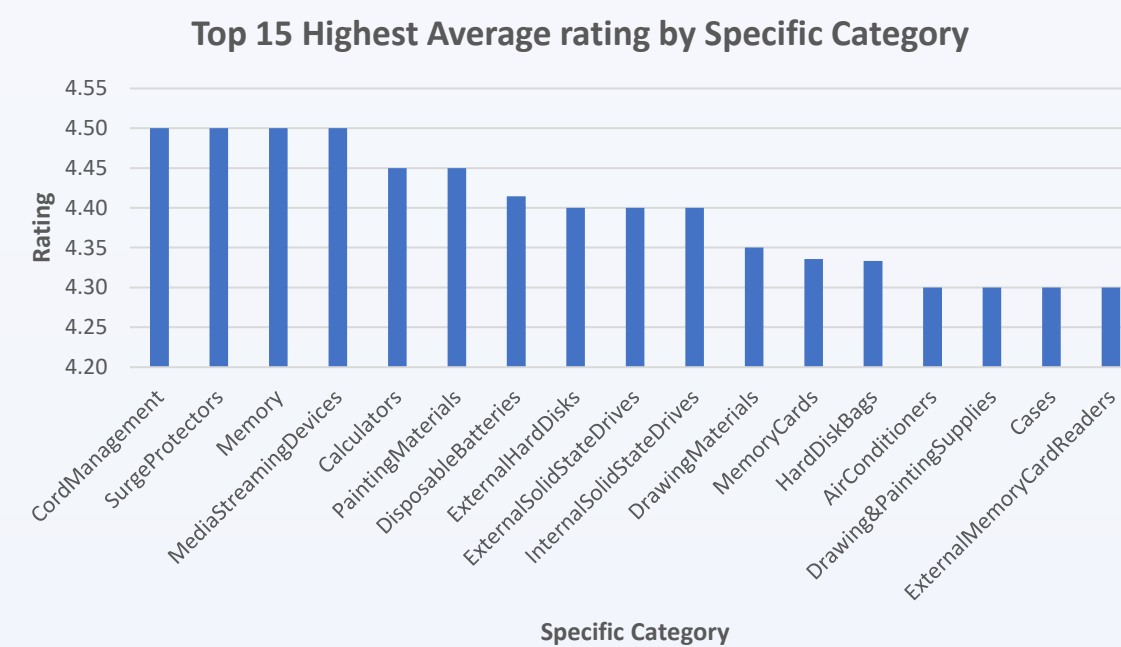
AVERAGE DISCOUNTED PRICE

Airconditioners, Laptops and TVs have the highest discounted price while sewing machine , water purifiers and hard disks have the lowest.



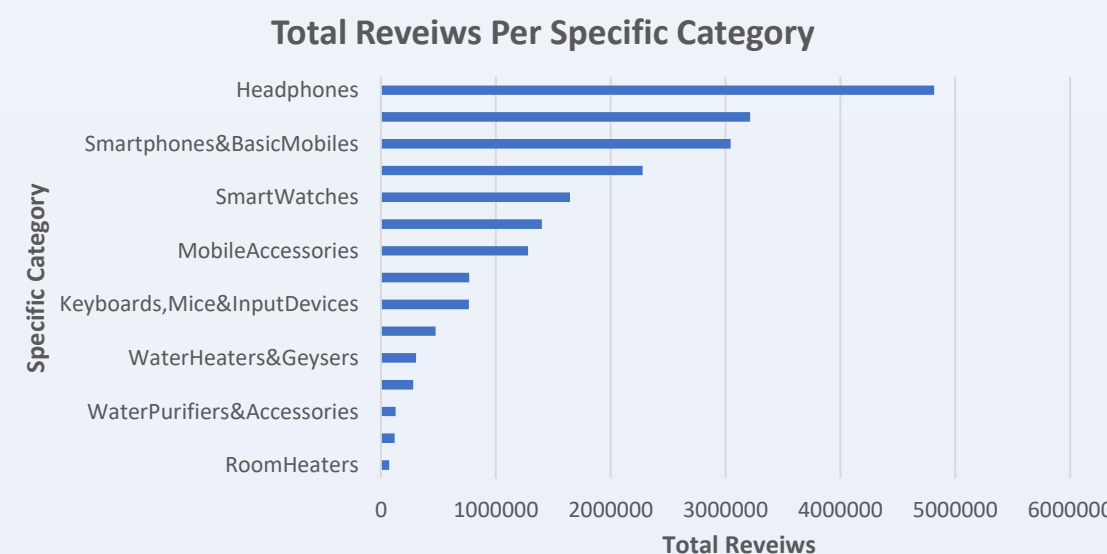
AVERAGE RATINGS

The Highest rated Category in the data is for Cord Management shared equally with surge protectors, memory and streaming devices.



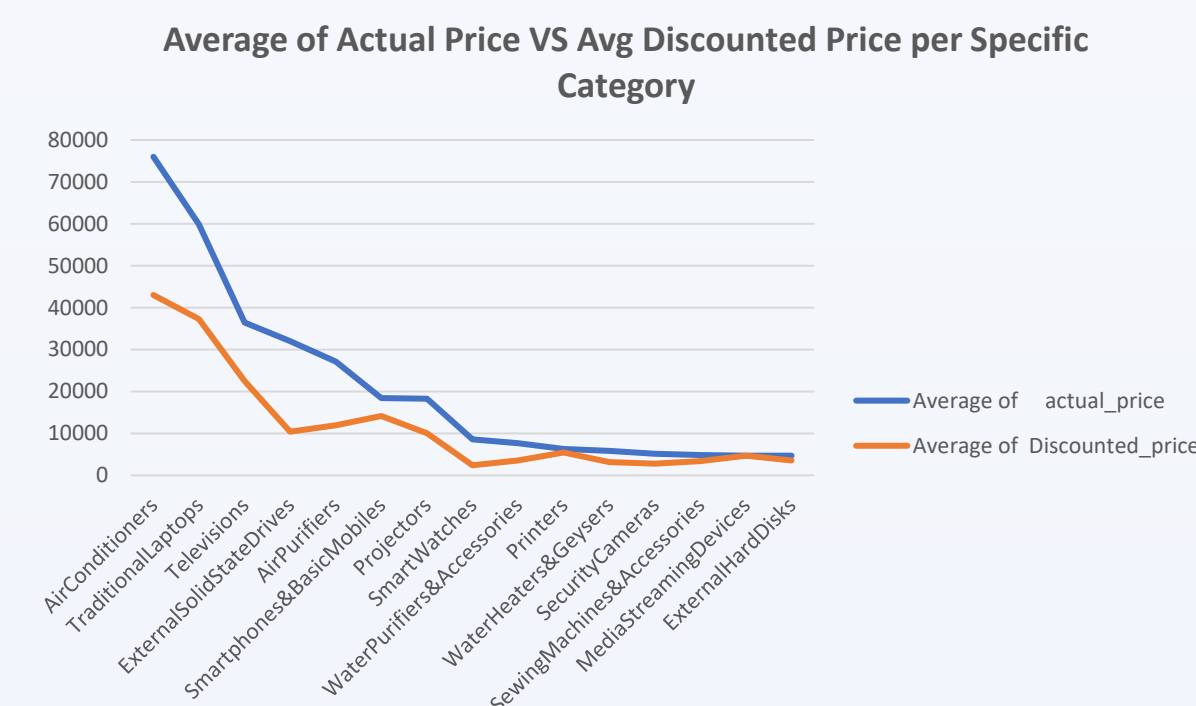
TOTAL REVIEWS

It appears that Headphones have the highest customer interactions with almost 0.5 M Total Reviews, water purifiers and Room Heaters came at the bottom with the least interactions .



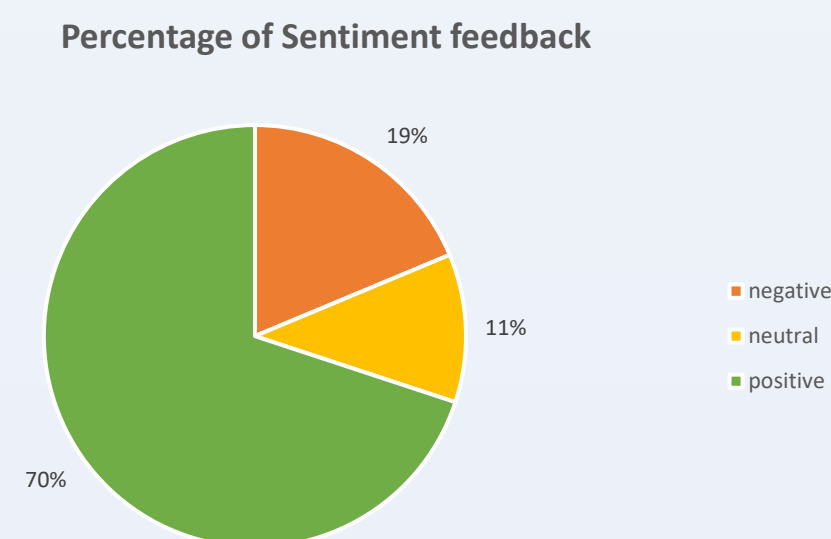
ACTUAL & DISCOUNTED PRICE

It appears that the Airconditiones and Laptops have the highest discounts percentage while categories like External Hard Disks & Media streaming devises have the lowest.



SENTIMENT FEEDBACK

70 % of the customers reacted positively about the products, 19 % were negative and only 11 % were neutral.



RESULTS

- Only 0.5% of the data listing have rating less then 3 and over 95% have ratings between 3.3 and 4.5
- 70 % of the available Sentiment reflected Positive Feedback indicating high level of customer satisfaction.

HIGHLIGHTS

- Some Categories sales are driven by high discount percentages like Air conditions , laptops and TVs while other categories like printers , security cameras and Hard Disks are sold at almost the same as actual prices
- . Products popularity is based on the quality. need and meeting expectations.

LIMITATIONS

- 85% of the dataset does not contain values for Sentiment , sentiment type and score
- The dataset does not provide transaction to derive the Revenue potential which could be lower than is shown.