Problems and background

The online retail market is fiercely competitive. Understanding customer preferences, product performance, and market trends can help a retailer stay ahead. The problem here is to analyse the dataset to extract insights that can help in enhancing product visibility, improving customer satisfaction, and eventually, increasing sales.

Project scope

This project focuses on digging into Amazon's product reviews to find useful insights. We want to understand how discounts affect ratings, which categories get the best ratings, and whether product prices impact ratings. We'll also look at how review length and the presence of product images influence ratings. Our goal is to provide clear findings and recommendations to help boost sales and improve customer satisfaction.

Methodology

Data Sources: Utilising the provided Amazon product reviews dataset as the primary source of information.

Data Wrangling: Organising and refining the raw dataset, including tasks like handling missing values, removing duplicates, and formatting data for analysis.

Data Analysis: Employing analytical methods to extract insights from the dataset, such as examining correlations, identifying trends, and conducting statistical analysis.

Data Visualization: Presenting the analysed findings visually through graphs, charts, and diagrams to enhance comprehension and decision-making.

Technical Processes

- Tools used : Excel, Python, PowerBi
- Formulas used in Excel: Sum, Max, Min, Pivot table, Average, Vlookup
- Python library: Numpy, Pandas. Used Matplotlib, Plotly and Seaborn for visualisation in python, NLTK (for sentiment analysis)
- Used PowerBi for visualisation and making Dashboard

Business Concepts Used

Market Understanding: Utilising the dataset to comprehend market dynamics, including product performance, customer preferences, and emerging trends.

Customer Demographic: Analysing demographic data to profile customers based on factors such as age, gender, location, and purchasing habits.

Customer Behaviour: Examining customer interactions with products, including review ratings, content, and purchase history, to understand their preferences and behaviours.

Customer Retention: Implementing strategies informed by data insights to enhance customer satisfaction, address concerns, and foster loyalty to retain existing customers.

New Customer Acquisition: Leveraging data-driven approaches to attract new customers through targeted marketing initiatives, personalised offers, and optimised product recommendations.

Recommended Analysis:

How does the discount percentage affect the rating of a product?

➜ It appears that there is no discernible correlation between the discount percentage offered and the rating of a product. The scatter plot depicted below, along with the linear trend line, illustrates that the discount percentage does not significantly influence the product's rating. Instead, it seems to depend largely on individual customer preferences. This observation is further supported by statistical analysis, which

indicates that across discounts ranging from 10% to 90%, the rating pattern remains consistent, averaging around 4.8 to 4.9 out of 5.

Which category has the highest average rating?

→ The statistics indicate that the categories Office Products, Toys & Games, and Home Improvement boast the highest average ratings, each garnering a rating of 4.3. Following closely are Computer & Accessories and Electronics, ranked as the second highest-rated categories with ratings of 4.2 and 4.1, respectively. Other categories fall within the range of 4.0 to 3.8.

Is there a correlation between the product's price and its rating?

→ Based on the data provided, there appears to be a positive correlation between the product's price and its rating. As the price range increases, the average rating of the products also tends to increase. For instance, products in the higher price ranges, such as 70039-80039 and 80039-90039, have higher average ratings of 4.3, compared to products in lower price ranges, which generally have ratings ranging from 4.1 to 4.2. However, it's essential to note that the highest-priced range, 130039-140039, stands out significantly with an average rating of 4.7, indicating that price may not be the sole determinant of a product's rating, and other factors could also influence customer satisfaction and perception.

What is the most common word in the positive and negative reviews?

(The findings presented here also include results obtained from Python analysis. For a detailed examination, please refer to the Python file provided in this folder.)

Python Results:

→ Most common words in positive reviews: [('good', 4751), ('product', 2754), ('quality', 1786), ('use', 1244), ('cable', 1157)]
→ Most common words in negative reviews: [('heater', 2), ('heating', 1), ('capacity', 1), ('zero', 1), ('initiated', 1)]

Excel Results:

Sentiments	Count of review_content

negative   984

neutral   69

positive   410

Which product has the highest number of reviews and what is its rating?

➔ The top 10 products with the highest ratings are as follows: at the forefront are REDTECH USB-C, Amazon Basic Wireless Mouse, and Syncwire LTG Charger, all boasting a perfect rating of 5.0. Following closely are products with the second-highest rating of 4.8, including Instant Pot Air Fryer Appliance and Oratech Electric Coffee Frother, among six others, each receiving a rating of 4.7.

Is there a correlation between the length of a review and the rating given?

➔ The scatter plot conclusively demonstrates a correlation between the length of a review and its corresponding rating. This relationship is clearly depicted in the chart, where a review length of 626 characters corresponds to a rating of 3.0, while a review length of 11411 characters corresponds to a higher rating of 4.6. It can be inferred that customers tend to provide more extensive reviews when they are satisfied with a product, whereas they may be less inclined to compose detailed reviews if they are dissatisfied with it.

Can the length of the product description be correlated to the product's Rating?

➔ Upon observing the scatter plot, it appears that there exists a correlation between the product description and the product's rating. This correlation suggests that individuals tend to exhibit a preference for comprehensively described products. Specifically, as

the length of the product descriptions increases, so does the corresponding rating. For instance, according to the chart, a product description with a length of 1143 characters yields a rating of 3.0, whereas a description of 2646 characters is associated with a higher rating of 4.3, and so forth.

Conclusion

In conclusion, the analysis of Amazon product reviews uncovers valuable insights. Higher-priced products generally receive higher ratings, suggesting a correlation between price and perceived value. Additionally, longer reviews tend to accompany higher ratings, indicating that customers appreciate detailed feedback. Furthermore, there's an indication that longer product descriptions might lead to higher ratings, highlighting the importance of providing comprehensive information. Understanding these correlations can empower sellers to optimise their product listings, enhance customer satisfaction, and ultimately drive sales on the platform.

Project owner

Name: Ritika Pathak

Email Id: ritupat707@gmail.com

Peerlist profile: https://peerlist.io/ritika_p

Discord username: r.k.0.7