Project Report: Analysis of Amazon Product Reviews

Background

In the world of e-commerce, product reviews play a crucial role in influencing customer decisions and improving the overall quality of products. Amazon, being one of the largest online retailers, collects a vast amount of data in the form of product reviews. These reviews provide insights into product quality, customer satisfaction, and potential improvements. Understanding the relationship between product ratings, review sentiments, pricing, and other factors is key to enhancing customer experience and optimizing business strategies.

Problem Statement

The problem addressed in this project is to analyze Amazon product reviews to understand factors influencing customer ratings. Specifically, the analysis aims to explore relationships between product ratings and variables such as discount percentage, price, review length, product description length, and review sentiment. The objective is to uncover insights that can help optimize product offerings, improve customer satisfaction, and guide business strategies.

Data Overview

The dataset contains product review data collected from Amazon, including the following columns:

- **Product Name:** The name of the product.
- **Category**: The product category (e.g., Electronics, Home & Kitchen, Office Products).
- **Rating:** The product's rating (on a scale of 1-5).
- **Review Text:** The content of the customer review.
- **Review Length:** The number of characters in the review.
- **Product Description:** Description of the product.
- **Discount Percentage:** The discount applied to the product.
- **Number of Reviews:** The total number of reviews for the product.

Methodology

- **Data sources**

1. I am obtaining data from the Kaggle website, which is an open-source platform for datasets and data science projects. **Dataset Link:** https://www.kaggle.com/datasets/ahmedsayed564/amazon-sales-dataset

- **Data wrangling**

1. Data understanding
2. Data cleaning
3. Data manipulation

- **Data analysis**

1. Finding the trends and patterns

- **Data visualisation**

Technical Processes

- Use pivot tables for summarizing data.
- Calculate averages, variances, and growth rates.
- Create charts and graphs for visual representation.
- Apply filters and sorting for specific analyses.
- Use functions like Count and Sum IF for data aggregation.

Key Findings

1. **Discount and Rating**:
    - There is no apparent relationship between the discount percentage and the product's rating, as seen in the scatter plot.
2. **Top Categories by Rating**:
    - The highest average ratings are observed in the following categories:
        - **Office Products**: 4.3
        - **Toys & Games**: 4.3
        - **Home Improvement**: 4.3
    - The category with the lowest average rating is **Car & Motorbike** (3.8).
3. **Price vs. Rating**:
    - There is no significant correlation between the product's price and its rating, as indicated by the scatter plot.
4. **Review Sentiment**:
    - Positive reviews account for the majority (1115 reviews), with negative reviews (267) and neutral reviews (83) being significantly fewer.
5. **Rating Distribution**:
    - Most products fall within the **4-4.5 rating range**, with 1007 products in this category, indicating that a majority of customers give high ratings.
6. **Top Products by Reviews**:
    - The product with the highest number of reviews is **AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps, 3-Foot)** with 853,945 reviews.
7. **Review Length and Rating**:
    - There is no correlation between the length of reviews and the rating given.
8. **Product Description Length and Rating**:
    - Similarly, there is no correlation between the length of the product description and its rating.

Recommended Analysis:

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

- Which category has the highest average rating?
- Is there a correlation between the product's price and its rating?
- What is the most common word in the positive and negative reviews?
- What is the distribution of ratings across all products?
- Which product has the highest number of reviews and what is its rating?
- Identify the top 5 users who have given the most reviews.
- Is there a correlation between the length of a review and the rating given?
- Can the length of the product description be correlated to the product's rating?

Concussion:

The analysis of Amazon product reviews reveals several key insights that can inform business strategies. While there is no significant correlation between pricing, discount percentage, review length, or product description length with ratings, certain product categories stand out as having higher average ratings. Additionally, the majority of reviews are positive, suggesting a strong customer satisfaction trend. The findings provide valuable direction for product categorization, marketing strategies, and customer feedback management.

By leveraging these insights, businesses can enhance customer satisfaction, improve product offerings, and ultimately drive higher sales and better customer loyalty. Further detailed studies and experiments could refine these insights and help optimize various aspects of product marketing and customer engagement.

Project owner

Name: vikant Kumar

Date: 28-1-2024