

Amazon Product Reviews By Customers

Summary --

In this analysis of the Amazon Product Reviews dataset, we began by importing the necessary libraries and loading the dataset, providing an initial overview of its structure. Subsequently, we meticulously cleaned the data, eliminating redundant columns, handling duplicates, and enhancing column names for better readability.

The exploratory data analysis delved into various aspects. We visually depicted the distribution of product brands and the proportion of categories, gaining insights into the diversity of the dataset. Moving to the reviews, we analysed the distribution of helpful reviews through histograms and box plots, shedding light on the level of engagement and interaction with user reviews.

The temporal evolution of reviews was explored through time series analysis, where we converted the 'reviews.date' to datetime format, set it as the index, and visualized monthly review counts. Additionally, we examined the relationship between ratings and helpful votes using scatter plots and explored correlations among numerical features through a heatmap.

Sentiment analysis added a qualitative dimension to our study. Leveraging TextBlob, we assessed the sentiment of customer reviews, creating a new column for sentiment analysis results. The subsequent visualization of sentiments using a pie chart provided a concise overview of the distribution of sentiments in the dataset.

This comprehensive analysis, encompassing data cleaning, exploratory data analysis, time series analysis, and sentiment analysis, furnishes valuable insights into the characteristics of the Amazon Product Reviews dataset and the sentiments expressed by users. Each step contributes to a thorough understanding of the dataset, providing a solid foundation for informed decision-making and further exploration.