About the Brand

Amazon is a global e-commerce giant offering a vast range of products and services, from books and electronics to cloud computing and AI. Known for its customer-centric approach, Amazon provides a seamless shopping experience and fast delivery. It also supports a vast marketplace for third-party sellers, enhancing its product diversity and availability.

Abstract

This article delves into the fascinating world of customer behavior on Amazon. By leveraging the power of Python and SQL, we embark on a data-driven exploration, uncovering hidden trends and insights that influence buying decisions. We analyze a vast dataset of Amazon products, dissecting the intricate relationships between product ratings, discount enticements, and actual prices across diverse categories.

Introduction

The ever-evolving landscape of online shopping necessitates a profound understanding of customer behavior. This project sheds light on customer preferences and product characteristics that shape purchasing decisions on Amazon, a leading e-commerce platform. Through this analysis, we illuminate:

Customer Champions: Which product categories consistently garner the highest ratings, revealing products that resonate most with users?

Discount Decoded: Do discounts significantly sway customer perception across different product types?

Price Perception: Is there a correlation between a product's rating and its price tag? Does a higher price tag automatically translate to higher customer satisfaction?

Objectives

1)Hypothesis testing in python using well known libraries like numpy,pandas,seaborn,matplotlib,scipy etc..

2)Creating interactive dashboards using MICROSOFT POWER BI

3)Using Postgres SQL for gaining some necessary insights about the data

Exploratory Data Analysis (Python)

Python facilitated a deep dive into a specific low-rated category, analyzing the distribution of ratings and their relation to discount percentages. Statistical tests were likely conducted to assess the significance of these relationships.

We investigated the overall correlation between ratings and discount percentages (across all categories) and the relationship between ratings and actual prices using Python. Line plots and statistical tests might have been used for visualization and evaluation.

PostGres SQL

We identified top and bottom-rated categories by grouping products by category and calculating average ratings using SQL queries (potential use case).

The distribution of discount percentages across categories was explored, potentially using SQL queries, to understand how discounts are applied.

Analyzing the relationship between ratings and other product attributes like sub-category, brand, or number of reviews.

Exploring price trends for different product categories across time periods.

Analyzing customer demographics associated with ratings or purchase behavior (if available in the data).

We aimed at finding:

1)To Find highest/lowest rated Category actual price wise/discount percentage wise

2)To Find highest/lowest rated Sub Category actual price wise/discount percentage wise

3)To Find highest/lowest rated Company actual price wise/discount percentage wise

4)To Find highest/lowest rated Products of a particular company actual price wise/discount percentage wise

5)Top discounted products category wise/subcategory wise/company wise

6)Top Trending products category wise/subcategory wise/company wise

Data Visualization

Our analysis was further enhanced by the creation of interactive Power BI dashboards. These dashboards offered a multifaceted view of the data, revealing that while Accessories dominated in terms of discount prices, Gold & Diamond Jewellery held the highest total discount within subcategories. Interestingly, unrated products (NOT_RATE) had the highest total discount, suggesting potential clearance sales. Shifting focus to actual prices, the dashboards unveiled a significant difference in price distribution. While Accessories remained a leading category, it showed a much higher actual price sum compared to discounts. Additionally, the HIGH_RATE category held the highest total of actual prices, indicating a correlation between higher ratings and higher product value. The dashboards also provided product-level insights, showcasing jewelry dominance in the top products by actual price and a positive correlation between high ratings and product prevalence.

Conclusion

This project provides valuable insights into customer behavior on Amazon. By analyzing product ratings, discounts, prices, and potentially other variables, we gain a deeper understanding of customer preferences and guide informed business decisions. This knowledge can be leveraged to enhance product development, optimize pricing strategies, and ultimately, drive customer satisfaction and sales growth.

Anyone having a certain dataset with the same analytical requirements can follow this approach of using SQL,Python and Power Bi simultaneously to fulfill their requirements.

Timeline

Initial Submission	Revision	Final Submission
Jun 28, 2024	Jul 10, 2024	Jul 12, 2024

Contributors

Parshwa Panchiwala

Het Patel

Krisha Patel

Viral Modi

Pratik Bajaj

Harsh Lohana

Guided By : Alakh Pandya

Thank you!