

Power BI

Unlocking E-Commerce Success: Analyzing Key Sales Drivers on Amazon"



Shruti Jain

2023JULB01306

Power BI

1. Problem Statement

In the competitive e-commerce landscape, understanding the factors that influence product sales is essential for improving performance. This report analyses Amazon's sales data to identify key drivers such as product ratings, pricing, category performance, and seasonality, enabling data-driven decision-making to boost sales and optimize business strategies.

Data Requirement

The aim of this project is to identify the key factors influencing product sales in an online marketplace like Amazon. For this, we require a dataset that provides detailed information about product attributes, customer reviews, and sales-related metrics. Specifically, the data should include:

1. Product Information:

- Product names
- Brand details
- Categories and subcategories of products
- Manufacturer details (optional)

2. Sales Metrics:

- Pricing information
- Trends over time (e.g., sales dates or periods)

3. Customer Feedback:

- Ratings provided by customers
- Reviews text (to analyze sentiment)
- Helpfulness votes (e.g., number of users who found reviews helpful)
- Recommendations made by customers

4. Temporal Data:

- Review dates (to understand trends and seasonality)

By collecting these data points, the analysis will focus on identifying the factors (e.g., pricing, reviews, brand reputation) that most significantly impact sales trends.

Method of Data Collection

1. Dataset Source:

- The dataset will be sourced from Kaggle, a reliable platform for public datasets. The "**Amazon US Customer Reviews Dataset**" is chosen as it provides verified reviews, ratings, and other necessary metrics.

2. Steps to Collect Data:

- **Access Kaggle:** Visit the Kaggle website and locate the selected dataset using the search bar. The chosen dataset is [Amazon US Customer Reviews Dataset](#).
- **Download the Dataset:** After selecting the dataset, download it in CSV format.
- **Examine the Dataset:** Open the dataset using tools like Excel, Python, or Power BI to understand its structure and identify relevant columns for analysis.

3. Data Preprocessing:

- Once downloaded, inspect the dataset for missing or irrelevant data.
- Remove duplicates, fill missing values where possible, and clean the dataset to ensure accurate analysis.

This method ensures that the data collected is structured, reliable, and ready for further analysis using tools like Power BI. Let me know if you need more details for the report!

FileHomeInsertPage LayoutFormulasDataReviewViewAutomateDeveloperHelpPower Pivot

Paste

Clipboard

Font

Calibri11A⁺⁺₋

B

I

U

Alignment

Merge & Center

Number

General

%

Styles

Conditional Formatting

Format as Table

Cell Styles

Cells

Insert

Delete

Format

Editing

AutoSum

Fill

Clear

Sort & Filter

Find & Select

Add-ins

Add-ins

Analyze Data

H7

Data Validation

Data validation ensures the data collected for the analysis is accurate, consistent, and useful for deriving insights. In this project, the following steps are performed to validate the data:

1. Completeness Check:

- Verify that all necessary columns, such as product name, categories, reviews.rating, and prices, are present and contain sufficient data.
- Identify missing values in critical columns and decide how to handle them (e.g., fill with default values or drop rows).

2. Accuracy Check:

- Confirm that the data in key fields, such as reviews.rating and prices, fall within logical and expected ranges (e.g., ratings between 1 and 5).
- Check for outliers or unusual entries that could indicate data errors.

3. Consistency Check:

- Ensure uniform formats for dates, text fields, and numeric data. For example, all dates should follow a consistent format (e.g., YYYY-MM-DD).
- Validate that the same product appears under consistent names and categories.

4. Duplicates Removal:

- Identify and remove duplicate rows to avoid skewed analysis results.

5. Integrity Check:

- Cross-check related columns, such as reviews.rating and reviews.numHelpful, to ensure logical consistency (e.g., higher helpful votes for higher ratings).

Data Cleaning

After validation, data cleaning ensures the dataset is prepared for analysis by addressing issues such as missing values, duplicates, and inconsistent formats. The cleaning process includes:

1. Handling Missing Values:

- Use techniques such as:
 - Filling missing numerical values (e.g., prices) with the median or mean of the column.
 - Filling missing categorical data (e.g., categories) with the most frequent category.

- Dropping rows with excessive missing data in critical columns.

2. **Standardizing Data:**

- Convert text to lowercase (e.g., product names or categories) for uniformity.
- Ensure all numerical fields, like prices, are formatted as numbers with no special characters or trailing spaces.

3. **Removing Outliers:**

- Analyze columns like prices and reviews.rating for extreme outliers. Use visualization techniques (e.g., box plots in Power BI or Python) to identify and handle them appropriately.

4. **Deduplication:**

- Remove identical rows to avoid repetitive data. For example, duplicate product reviews or entries.

5. **Date Cleaning:**

- Ensure all dates (e.g., reviews.date) are converted to a standard format for trend analysis (e.g., YYYY-MM-DD).

6. **Text Cleaning:**

- Preprocess textual data, such as reviews.text, by:
 - Removing punctuation, special characters, and unnecessary whitespace.
 - Converting to lowercase for consistency in text analysis

1. **Top-Selling Products by Rating:**

- **Insight:** Higher-rated products often tend to have better sales performance, as customers generally prefer items with positive reviews. This analysis helps identify top-rated products that are performing well in terms of sales, indicating a strong customer preference for high-quality products.

2. **Performance by Product Category:**

- **Insight:** Certain product categories outperform others, showing which sectors are popular among customers. Categories like electronics or clothing may consistently perform better, revealing market trends and areas where more resources might be allocated for growth.

3. **Rating vs. Sales Performance:**

- **Insight:** A positive correlation between product ratings and sales can be seen, with higher ratings typically leading to higher sales. Products with better ratings attract more buyers, indicating that customer satisfaction and trust play key roles in driving sales.

4. **Pricing Strategy and Sales Correlation:**

- **Insight:** Pricing has a direct impact on sales, where products with competitive pricing often show higher sales volumes. This analysis helps understand if more expensive items are justified by their quality and ratings or if lower-priced products perform better due to affordability.

5. **Sales by Product Identification:**

- **Insight:** Comparing sales data across individual products reveals the top performers within each category. This helps identify the best-selling products, which can guide inventory management and marketing strategies to boost overall sales.

6. **Product Performance Based on Sentiment vs. Ratings:**

- **Insight:** The sentiment of reviews, whether positive, neutral, or negative, directly affects the sales of products. Positive reviews with high ratings often correlate with increased sales, while products with negative sentiments tend to have lower sales despite good ratings, showing the importance of customer satisfaction in driving purchases.

Story telling

In the ever-evolving world of e-commerce, understanding the factors that drive sales is crucial. By analyzing key metrics like product ratings, pricing strategies, and customer sentiment, we uncover a story of how customer trust and satisfaction directly influence purchasing decisions. Top-rated products consistently outperform others, proving that quality and positive feedback resonate with buyers. Categories such as electronics and fashion dominate, highlighting popular consumer preferences. Additionally, competitive pricing plays a pivotal role in driving sales, with affordability often leading to higher volumes. This analysis not only uncovers trends but also helps businesses optimize strategies for sustainable growth.