



AMAZON SALES ANALYSIS

By
Priyadharshini

AGENDA



Introduction

Key variables

Data cleaning

EDA

Data Visualization

Conclusion

INTRODUCTION

- The dataset contains details about products, including discounted prices, actual prices, and user ratings
- The goal of the analysis is to explore the relationships between these features and identify key insights.

OBJECTIVES

DATA CLEANING:

Ensure the dataset is free from missing values and duplicates.

DATA VISUALIZATION:

Create visual representations to understand distributions and relationships.

ANALYSIS

Derive actionable insights from the cleaned data



KEY VARIABLES

DISCOUNTED PRICE

The price after applying discounts.

ACTUAL PRICE

The original price of the product before any discount

RATINGS



Customer ratings of the product.



DATA CLEANING



CONVERTING PRICE COLUMNS

Converted the `discounted_price` and `actual_price` columns to numeric format after removing currency symbols and commas.

HANDLING MISSING VALUES

Dropped rows with missing values in critical columns such as `discounted_price`, `actual_price`, and `rating` to ensure the accuracy of the analysis.

REMOVING DUPLICATES

Identified and removed duplicate entries to maintain data integrity and avoid redundancy in the analysis.

DATA VALIDATION

Checked for any inconsistencies in the data types and ensured that all columns were in the correct format for analysis.

EXPLORATORY DATA ANALYSIS (EDA)



DESCRIPTIVE STATISTICS

calculated summary statistics (mean, median, standard deviation) for key numerical columns such as `discounted_price`, `actual_price`, and `rating`.

Observed differences between the mean and median prices, indicating potential skewness in the price distribution.

DESCRIPTIVE STATISTICS

Created histograms to visualize the distribution of `discounted_price` and `actual_price`

Noticed that the price distributions are right-skewed, with most products having lower discounted prices.

CORRELATION ANALYSIS

Analyzed correlations between `discounted_price`, `actual_price`, and `rating`.

Found a moderate positive correlation between `discounted_price` and `actual_price`, as expected..

DATA VISUALIZATION



BOX PLOT

A box plot was used to compare the distributions of `discounted_price` and `actual_price`.

BAR CHART

Created bar charts to visualize the distribution of products across different categories.

SCATTER PLOT

A scatter plot was used to examine the relationship between `discounted_price` and `rating`.

CONCLUSION

The analysis provided a comprehensive overview of the dataset, focusing on pricing patterns, customer ratings, and category distributions

While price adjustments are important, they may not directly influence customer satisfaction, as evidenced by the lack of a strong relationship between prices and ratings.

RECOMMENDATION

- **Focus on Quality:** Since ratings are not heavily influenced by price, efforts should be placed on improving product quality and customer experience to boost ratings and customer satisfaction.
- **Review Outliers:** Investigate products with extremely high or low prices to understand if they are affecting overall pricing strategies.
- **Category Expansion:** Consider expanding offerings in underrepresented categories to balance product distribution and explore new market opportunities.
- **Further Analysis:** Perform deeper analysis on other factors such as brand, reviews, and marketing efforts that could be influencing ratings and sales.



THANK YOU
