DATA ANALYTICS REPORT

EXPLORATORY DATA ANALYSIS OF AMAZON CONSUMER BEHAVIOR DATASET

Purchasing Patterns, Consumer Intent & User Engagement

amazon

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

GITHUB

TABLE OF CONTENTS

Problem Statement & Objective	3
EDA Workflow	4
Key Questions Explored	5
Data Overview	6
Insights	11
Business/Developer Takeaways	27
Future Work	29
Conclusion	30

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

The Amazon platform hosts millions of products across diverse categories. With massive variations in purchases, ratings & reviews, it becomes challenging to understand consumer trends, customer preferences & what drives product success.

Objective:

To explore the Amazon consumer behaviour dataset and uncover patterns, insights, and trends that help businesses and sellers make data driven decisions.

EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.

01 Data Collection

- Gathered the Google Play Store dataset from Kaggle

02 Data Understanding & Anomaly Detection

- Looked at data distributions
- Found missing values, outliers, and unusual patterns

03 Data Cleaning & Treatment

- Fixed missing or incorrect values
- Standardized formats for consistency

04 Exploratory Analysis

- Univariate Analysis: e.g Purchase frequency, Search Method, Reviews
- **Bivariate Analysis:** Age vs Purchase Frequency
- Multivariate Analysis: Shopping Satisfaction vs Purchase Frequency vs Age Groups

05 Insights & Reporting

- Summarized patterns and trends
- Highlighted key findings for developers and businesses

KEY QUESTIONS EXPLORED

- How does upload timing (day, hour) influence reach and engagement and when are the most effective publishing windows?
- Which content categories drive the highest reach & where does competition versus niche opportunity exist?
- Is consistent publishing more reliable than relying on rare viral hits for long-term performance?
- How are views, likes, comments & dislikes related & which engagement signals matter most for trending?
- Do regional differences affect engagement patterns and where is country-specific optimization most effective?
- How do platform policies and external factors influence video longevity beyond engagement metrics?

DATA OVERVIEW

The dataset provides insights into Amazon consumer behavior, covering demographics, purchasing and browsing patterns, product search behavior, cart interactions, and the influence of reviews and personalized recommendations. It also captures overall shopping satisfaction and user feedback for service improvement.

Data Source: Kaggle

Dataset Size

602
Records

23
Features

Purchase Diversity

29
Product Categories

4
Search Methods

5
Satisfaction Indices

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
Timestamp	Date	Response time
Age	Numerical (Discrete)	Age of the user
Trending_date	Categorical	Date the video trended on YouTube
Purchase_Frequency	Categorical	How often the user buys on Amazon
Purchase_Categories	Categorical	Types of products the user buys
Personalized_Recommendation_Frequency	Categorical	How often purchases are influenced by recommendations
Browsing_Frequency	Categorical	How often the user browses Amazon
Product_Search_Method	Categorical	How the user searches for products
Search_Result_Exploration	Categorical	Whether the user explores multiple pages of results
Customer_Reviews_Importance	Categorical	Importance of reviews in purchase decisions
Add_to_Cart_Browsing	Categorical	Whether the user adds items to the cart while browsing
Cart_Completion_Frequency	Categorical	How often cart items are purchased
Cart_Abandonment_Factors	Categorical	Reasons for leaving items in the cart
Saveforlater_Frequency	Categorical	Frequency of using "Save for Later
Review_Left	Categorical	Whether the user has left a product review
Review_Reliability	Categorical	Trust in product reviews
Review_Helpfulness	Categorical	Usefulness of other customers' reviews

DATA OVERVIEW

Continued..

Dataset Features	Type	Feature Description
Personalized_Recommendation_Frequency	Categorical	How often user receives personalized recommendations
Recommendation_Helpfulness	Categorical	USefulness of personalized recommendations
Rating_Accuracy	Categorical	Relevance of recommendations
Shopping_Satisfaction	Categorical	Overall satisfaction with Amazon
Service_Appreciation	Categorical	Most Appreciated Amazon Services
Improvement_Areas	Categorical	Suggested areas for Amazon to improve

DATA QUALITY CHALLENGES & ANOMALIES

Few inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- **Columns requiring cleaning or type conversion**: timestamp, Gender, Purchase_Categories, Product_Search_Method, Search_Result_Exploration, Customer_Reviews_Importance, Add_to_Cart_Browsing, Cart_Abandonment_Factors, Review_Left, Service_Appreciation, Improvement_Areas, Rating_Accuracy, Shopping_Satisfaction
- **Personalized_Recommendation_Frequency** appears twice with different types (str and int64) resolve duplication (rename or remove).
- **Columns with notable anomalies or invalid entries**:
  - Improvement_Areas (has an entry like '.')
- Two missing values in **Product_Search_Method** column.

DATA CLEANING & TREATMENT

Inconsistent and missing values were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

- Features such as t**imestamp, Gender, Purchase_Categories, Product_Search_Method, Search_Result_Exploration, Customer_Reviews_Importance, Add_to_Cart_Browsing, Cart_Abandonment_Factors, Review_Left, Service_Appreciation, Improvement_Areas, Rating_Accuracy, Shopping_Satisfaction** were standardized with respect to data type conversion.
- **Improvement_Areas** were cleaned to handle invalid entries (. replaced with mode).
- Personalized_Recommendation_Frequency were resolved and renamed as:
  - '**Personalized_Recommendation_Frequency** ' to Number_of_times_Personalized_Recommendation_Received
  - '**Personalized_Recommendation_Frequency**' to Purchase_made_on_Personalized_Recommendation
- Missing values in **Product_Search_Method** were replaced with mode.

INSIGHTS