

Amazon Products EDA – Team Project Instructions

Project Type: Team-Based Project

- You will work in **teams (2 students per team)**.
- Each team must create **one shared GitHub repository**.
- All team members must contribute (commits should reflect contribution).
- Work must be divided clearly between team members.



Project Title: **Amazon Products (EDA)**

You will perform a complete **Exploratory Data Analysis (EDA)** on an Amazon Products dataset.

This project simulates a **real-world business data analysis case**.



About The Dataset

Dataset Overview

This dataset contains data for **1000+ Amazon products**, including product details, pricing, ratings, and customer reviews.



Dataset Columns Description

Product Information

- **product_id** → Unique ID of the product
- **product_name** → Name of the product
- **category** → Product category
- **about_product** → Product description
- **img_link** → Product image link
- **product_link** → Official Amazon product link



Pricing Information

- **actual_price** → Original price
- **discounted_price** → Discounted price
- **discount_percentage** → Discount percentage



Rating Information

- **rating** → Product rating
- **rating_count** → Number of users who rated the product



Review Information

- **review_id** → Unique review ID
- **review_title** → Short review title
- **review_content** → Full review text
- **user_id** → Reviewer ID
- **user_name** → Reviewer name

Project Objectives

The main objective of this analysis is to:

- Analyze product categories and pricing patterns
- Understand rating behavior and customer satisfaction
- Explore discount strategies
- Identify trends that influence product popularity
- Translate insights into business recommendations

By the end of this project, you should be able to:

- ✓ Extract meaningful insights from raw data
 - ✓ Communicate findings clearly
 - ✓ Think like a data analyst
 - ✓ Provide actionable business recommendations
-

Required EDA Steps

Your analysis must include the following:

1 Data Understanding

- Load the dataset
- Inspect shape, columns, data types
- Identify missing values



2 Data Cleaning

- Handle missing values
- Fix data types if needed
- Remove duplicates (if any)
- Handle incorrect or inconsistent data



3 Descriptive Statistics

- Summary statistics
- Category distribution
- Price statistics
- Rating statistics



4 Data Visualization

Use meaningful visualizations such as:

- Category distribution charts
- Price vs Discount analysis
- Rating distribution
- Correlation heatmap
- Top-rated products
- Most reviewed products

⚠ Do NOT create random plots. Every visualization must answer a business question.



5 Outlier Detection

- Detect unusual prices
- Detect abnormal ratings
- Explain whether they should be removed or kept



6 Business Insights & Recommendations

This is the **most important part**.

You must answer questions like:

- Which category performs best?
- Do higher discounts increase ratings?
- Are expensive products rated better?
- Which features make products more popular?

Then provide:

Clear business recommendations

Marketing suggestions

Strategic improvements



GitHub Repository Structure (Mandatory)

Create a GitHub Repository

```
Amazon-EDA-Project/  
├── data/  
│   └── amazon_dataset.csv  
├── notebooks/  
│   └── amazon_eda.ipynb  
├── presentation/  
│   └── final_presentation.pdf  
└── README.md
```

README.md (Mandatory)

Each team must create a professional **README.md** file inside the repository.

You must include a section in the README explaining:

- Team members names
- Role of each member
- Contribution summary



Final Deliverables

Each team must submit:

1. GitHub Repository Link
2. Completed Jupyter Notebook
3. Clean and Organized Code
4. Clear Markdown Explanations
5. Final Presentation (5–10 slides summarizing findings)



Important Rules

- Do NOT copy from Kaggle notebooks.
- Do NOT blindly use ChatGPT without understanding.
- Every line of code must be understood.
- Plagiarism = zero.



Bonus (Optional)

Extra points if you include:

- Advanced visualizations
- Feature engineering
- Simple hypothesis testing
- Sentiment analysis on reviews



Final Advice

// This project is not just about finishing tasks or submitting a repository.
It's about learning how to think like a real data analyst.

Don't rush the process.

Understand your data.

Ask questions.

Be curious.

Great analysts are not the ones who write the most code
they are the ones who extract the most meaningful insights.

Work as a team, respect each other's effort, and take responsibility for your part.

I care more about your understanding than your results.

Make me proud ❤️