Data Lifecycle Scenario: E-commerce Marketplace

You are given a short scenario to analyze.

Step 1: Individual work Draw the lifecycle diagram, indicating:

- the sources of data,
- potential issues with quality, trustworthiness, or usability,
- the expected type(s) of data (structured / semi-structured / unstructured)

Step 2: Group Discussion: Compare your individual diagrams and notes with group members. Reflect on each stage of the lifecycle with your team and answer the guiding questions provided, completing the **lifecycle-step boxes** together. Prepare a brief summary (keywords or 1–2 sentences) to present.

Time guideline: 5 minutes individually, 10 minutes in groups.

Scenario Description: E-commerce Marketplace

An online marketplace collects extensive data on purchases, product listings, user reviews, and browsing behaviors. Additional information may come from affiliate programs, social media campaigns, and external market trends. The platform aims to detect fraud, understand customer preferences, improve product recommendations, and optimize pricing strategies. Data can be structured (transaction logs) or semi/unstructured (reviews, images, product descriptions). Timely integration and cleaning of data from multiple sources is essential to ensure accurate analytics and reliable recommendations, while respecting user privacy and data regulations.

Scenario Analysis (individual work): Draw the lifecycle diagram, indicating:

- Source(s) of the data (human/machine, internal/external)
- Data type (structured / semi-structured / unstructured)
- Any obvious quality, reliability, or trustworthiness issues

Lifecycle diagram

Data Lifecycle Reflection (group work): For each stage, reflect on what it could involve in this scenario. Consider:

- How can inconsistent product IDs or duplicate transactions be resolved?
- How can user behavior be anonymized while still enabling insights?
- What analytics could improve sales, recommendations, or marketing campaigns?
- How to integrate and validate external data sources effectively?

Collect

Clean / Preprocess

Store

Analyze

Communicate / Report