**✅ 1. Total Number of Unique Products**

**Purpose:**

* Count how many distinct products (asin) are present in the dataset.

**Map:**

* Emit (asin → 1) for each product reviewed.

**Reduce:**

* Aggregate counts, giving each unique product exactly once.
* Count the total unique keys afterward.

**Insights:**

* Understand product diversity in the dataset.

**Sample Result:**

Total unique products: 128,453

**Interpretation:** You quickly see product diversity—how many different appliances are reviewed.

**✅ 2. Rating Distribution Analysis**

**Goal:**

* Identify the overall distribution of ratings (1 to 5 stars).

**Map:**

* Emit (rating → 1) for every review.

**Reduce:**

* Count how many reviews fall into each rating category.

**Insights:**

* Quickly understand customer satisfaction levels.
* **Sample Result:**
* CopyEdit
* 5 stars → 85,500 reviews
* 4 stars → 45,200 reviews
* 3 stars → 12,600 reviews
* 2 stars → 4,200 reviews
* 1 star → 1,700 reviews
* **Interpretation:** Clearly understand the overall customer satisfaction at a glance.

**✅ 3. Top Products by Helpful Reviews**

**Goal:**

* Identify products customers found most helpful, based on helpful votes.

**Map:**

* Emit (asin → helpful\_vote) for each review.

**Reduce:**

* Sum up total helpful votes per product.
* Identify products with highest helpfulness.

**Insights:**

* Find popular and influential products (positively or negatively impactful).

**Sample Result:**

mathematica

CopyEdit

Top 3 products by helpful reviews:

1. Product ID B01N0TQ0OH → 1,250 helpful votes

2. Product ID B07DD2DMXB → 900 helpful votes

3. Product ID B082W3Z9YK → 750 helpful votes

**Interpretation:** Easily identify which appliances customers discuss and find most helpful.

**✅ 4. Verified vs Unverified Purchase Analysis**

**Goal:**

* Analyze credibility by comparing verified vs unverified reviews.

**Map:**

* Emit (verified\_purchase\_status → 1) for each review.

**Reduce:**

* Count verified and unverified reviews separately.

**Insights:**

* Evaluate the reliability and credibility of customer feedback.

**Sample Result:**

Verified Purchase Reviews: 110,000 (90%)

Unverified Reviews: 19,500 reviews (10%)

**Interpretation:** Clearly evaluate the authenticity and trustworthiness of customer reviews.

**✅ 5. Product Popularity and Engagement**

**Goal:**

* Identify products with high customer engagement (number of reviews) and overall ratings.

**Map:**

* Emit (asin → [rating, helpful\_vote, 1]) for each review.

**Reduce:**

* Aggregate to get average rating, total helpful votes, and total review count.

**Insights:**

* Highlight products with strong customer interactions (popular, well-rated, highly discussed).
* **Sample Result:**
* Top 3 most popular and engaging products:
* 1. Product ID B01N0TQ0OH → 4,000 reviews, avg. rating: 4.6, total helpful votes: 3,500
* 2. Product ID B07DD2DMXB → 3,750 reviews, avg rating: 4.5, total helpful votes: 3,000
* 3. Product ID B082W3Z9YK → 3,500 reviews, avg rating: 4.3, total helpful votes: 2,800
* **Interpretation:** Identify the most reviewed, discussed, and highly-rated appliances that engage customers the most.

**Location of all files**

A screen shot of a computer program

AI-generated content may be incorrect.

**Updated HDFS Structure**

**A screenshot of a computer program

AI-generated content may be incorrect.**