

Scenario 1 A school district has five years of attendance records for 10,000 students. They want to understand what a "typical" chronically absent student looks like — what grade level they tend to be, whether they're more often absent on Mondays and Fridays, and what the most common reasons are. No comparison group; they just want to describe the pattern within this one group.

Answer: Class description — summarizing the general characteristics of a single group.

Scenario 2 A streaming service notices that customers who cancel their subscriptions almost always watched three or fewer shows per week for at least six consecutive weeks before canceling. They want to use this pattern to flag customers who might be about to leave.

Answer: Sequential pattern analysis — a specific ordering of events over time that predicts a future event.

Scenario 3 A grocery store chain analyzes millions of transactions and finds that customers who buy baby formula also tend to buy coffee and bananas in the same shopping trip. The store wants to use this to decide where to place products on shelves.

Answer: Association analysis — finding items that frequently co-occur, with no implied order or causation.

Scenario 4 A hospital has records for 50,000 patients. Rather than starting with any predefined categories, they run an algorithm that sorts patients into five natural groupings based on age, diagnosis history, medication use, and number of hospitalizations. The hospital didn't define these groups in advance — the algorithm found them.

Answer: Cluster analysis — discovering natural groupings in data without predefined categories.

Scenario 5 A credit card company reviews transactions from the past month. One account shows a \$4 purchase at a coffee shop in Iowa, then a \$3,800 electronics purchase in London two hours later, then another small purchase in Iowa the next morning. The company's system flags this account automatically.

Answer: Outlier analysis — identifying data points that deviate significantly from expected patterns.

Scenario 6 An insurance company wants to understand whether rural customers and urban customers differ in their claim patterns. They already know who belongs to each group. They want to find out what characteristics most distinguish the two groups from each other — not just describe one group, but contrast them.

Answer: Class discrimination — comparing two known, predefined groups to find what sets them apart.

Some Explanation and Clarification

Scenarios 1 and 6 are the pair students most often struggle with. What's the key question that distinguishes them?

Class description: "What does this group look like?" — one group, no comparison.

Class discrimination: "How are these two groups different?" — two predefined groups being contrasted.

Scenarios 2 and 3 are the other pair that students struggle with since both involve patterns across multiple items/events. The distinguishing question there is whether **order and time matter** — if yes, it's sequential pattern analysis; if it's just co-occurrence regardless of order, it's association analysis