🧠 Activity: Feature Engineering Showdown

⏱ **Time Required:** 15 - 20 minutes
🎯 **Purpose:** Apply key ideas from today's lecture on feature design and data preparation, build practical thinking skills, and get ready for the lab work.

🔍 Scenario: Designing Smart Features for Predictive Models

Imagine you're part of a data team at a hospital tech department. Your team's goal is to **build a predictive system** that helps flag patients who are likely to return to the hospital within 30 days after discharge.

You've been given a sample of raw data that includes:

patient_id	age	gender	blood_type	last_visit_date	monthly_cost	notes_ length	diagnosis_ code
...	...	...	...	...	...	...	...

The data includes:

- **Static features**

- **Time-based features**

- **Potentially risky features**

⚠️ Your Task:

Decide which features to use in your model, how to structure the data over time, and how to avoid future-looking values that could lead to inaccurate results (a.k.a. data leakage).

👥 Group Work Instructions (15 minutes)

Form groups of 3–4 and complete these steps:

1. 🔍 Feature Selection

Pick the features you would include in your predictive model.
 Use what we discussed in the lecture to support your choices:

- Is the feature useful and meaningful?

- Does it vary enough across records?

- Could it include future info that would cause data leakage?

2. 🔗 Designing Joins Over Time

Many features might come from separate tables (like lab results or prescriptions).
 How would you **join** this data safely over time?

3. ⚠️ Identify Risks

Discuss features that could be problematic:

- Are any features too constant or unrelated?

- Do any **reveal future info** (after discharge)?

🎯 Challenge: Find the Trap!

Can you find a feature that looks helpful—but might leak future information?

📢 Sharing and Discussion

✅ What You Need:

- Pen and paper, or a shared Google Doc

- Your group (3–4 people)

- Critical thinking and curiosity