In-Class Exercise: Data-Driven Approaches in Finance

Exercise 1: "Design a Data Pipeline"

Format: Groups of three to four students. Each group receives one use case. Design a complete data pipeline using the value chain template below.

Use cases:

A: Detecting fraudulent insurance claims before payout

B: Personalizing loan offers based on customer behavior

C: Monitoring trader compliance in real time

D: Predicting which customers will leave the bank next quarter

Fill in the template for your assigned case:

Value Chain Stage	Your Design
Raw Data: What is collected?	
Cleaned Data: What is removed or fixed?	
Features: What predictive signals are extracted?	
Predictions: What does the model output?	
Decisions: What action follows a prediction?	
Outcomes: How do you observe whether the decision was correct?	
Feedback: How does the next prediction improve?	

Q1: Where in your pipeline is a human-in-the-loop, and why is that point critical?

Q2: What data are you **deliberately not** using? What is your reason for excluding it?

Q3: How would you know if your model is degrading over time?

Present your pipeline to the class (three minutes per group).

Exercise 2: Structured Debate — "Should Alternative Data Be Used for Credit Decisions?"

Team A — "It Should"

Opening position: Traditional credit data excludes billions of people worldwide. Alternative data—mobile usage, utility payments, purchasing patterns—can assess thin-file populations and bring them into the formal financial system.

Stage	Your Argument
Argument I	
Argument II	
Argument III	
Concession	
Closing Statement	

Team B — "It Should Not"

Opening position: Alternative data is collected without meaningful consent, encodes socioeconomic proxies that reproduce existing inequality, and lacks regulatory guardrails. It is surveillance disguised as financial inclusion.

Stage	Your Argument
Argument I	
Argument II	
Argument III	
Concession	
Closing Statement	

Debrief (full class):

Q1: Did any argument from the opposing side genuinely change your mind? Which one?

Q2: Is it possible to use alternative data for financial inclusion **without** enabling surveillance? What conditions would be required?

Q3: Who should decide what data is "fair game" for credit decisions—regulators, consumers, lenders, or technologists?

Exercise 3: "The Override Decision" *(if time permits)*

Scenario: You are a loan officer at a regional bank. The algorithm has denied a small-business loan application. You review the file: steady revenue, no missed payments, strong references. The model flagged an unexplainable cash flow pattern it associates with elevated default risk. The applicant is sitting across the desk from you.

Q1: Do you override the model and approve the loan? Explain your reasoning.

Q2: What additional information would you want before making this decision?

Q3: If you override the model and the loan defaults, who bears responsibility—you, the model developers, or the bank?

Q4: If you do **not** override and the business fails, is that a better outcome for the bank? For society?

