Pre-Class Discovery Handout: Data-Driven Approaches in Finance

Activity 1: "What Does Your Phone Know About You?"

Scenario: You open a banking app on your phone. From the moment you unlock it to the moment you close it, data is generated at every step. Your task is to classify each data point in the table below.

Data Point	Type *(Trans / Behav / Alt / Market / Text)*	Why This Type?
Transfer money to a friend		
Spend extra seconds reading a loan offer		
App checks your GPS location		
News headline about interest rates appears		
Social media profile linked to app		

Q1: Which of these data points did you **choose** to share, and which were collected without your active decision?

Q2: If a credit model used all five data points, which one concerns you most as a borrower? Why?

Q3: Which data type most improves credit access for someone with no traditional financial history?

Activity 2: "The Automation Decision Matrix"

Scenario: You are the chief data officer at a mid-sized bank. You have the budget to automate **one** decision process this year. Rank the following from most to least suitable for automation.

a) Approving or denying small personal loans

b) Flagging potentially fraudulent transactions

c) Recommending investment portfolios to retail clients

d) Generating regulatory compliance reports

Criteria to consider: Volume of decisions, Consistency required, Explainability demanded, Risk of harm from errors.

Q1: Which process did you rank first? What makes it the strongest candidate for automation?

Q2: Which did you rank last? What makes it harder to hand over to an algorithm?

Q3: For the process you ranked first, what data would the model need, and what feedback loop would tell you the model is still working correctly?

Activity 3: "Read Before Class"

Find **one** recent news article about a financial institution using algorithms or machine learning to make decisions. Bring the article (printed or on your device) to class. Be prepared to explain:

1. What decision is being automated?
2. What data does the system use?
3. Is there a human override mechanism described?
4. Would you trust this system with **your** financial data? Why or why not?

Solutions

Activity 1: "What Does Your Phone Know About You?"

Data Point	Type	Why This Type?
Transfer money to a friend	Transactional	Records a financial event: amount, recipient, timestamp.
Spend extra seconds reading a loan offer	Behavioral	Captures how you interact with the app, not a financial event itself.
App checks your GPS location	Alternative	Location is non-financial data repurposed for financial profiling.
News headline about interest rates appears	Market or Textual	Market data if structured (rate quote); textual if the headline itself is analyzed for sentiment.
Social media profile linked to app	Alternative	Personal digital footprint used as a proxy for creditworthiness.

Q1: Only the transfer was an active choice. Reading time, GPS, news display, and social-media linkage are typically collected passively or as a condition of service—most users are unaware of the extent of behavioral and alternative data capture.

Q2: The GPS location and social media profile tend to raise the greatest concern. Both reveal personal context far beyond financial behavior, and borrowers rarely consent to their use in credit decisions explicitly. Reasonable answers may also cite behavioral data (reading time), since the borrower has no way to "perform well" on a metric they cannot see.

Q3: Alternative data (GPS patterns, utility payments, mobile-phone usage) most improves access for thin-file populations. Traditional transactional and market data require an existing banking relationship; behavioral data requires app usage history. Alternative data draws on activity that nearly everyone generates, regardless of prior financial engagement.

Activity 2: "The Automation Decision Matrix"

Most common ranking:

First — Flagging potentially fraudulent transactions. Fraud detection handles enormous volume, demands sub-second consistency, and benefits from a human-in-the-loop review stage (flagged transactions are investigated, not auto-blocked). The cost of a false negative (missed fraud) is concrete and measurable, creating a natural feedback loop.

Last — Recommending investment portfolios. Portfolio advice is deeply personal, requires understanding of a client's goals, risk tolerance, and life context. Regulators demand explainability (suitability obligations), and harm from a poor recommendation may surface only years later, making the feedback loop slow and noisy.

Q3 model answer: For fraud detection, the model needs transactional data (amount, merchant, location, time), historical patterns per account, and network-level signals (e.g., whether the merchant has seen fraud before). The feedback loop consists of investigation outcomes: confirmed fraud, false alarm, or missed fraud discovered later. Each label flows back into the training set, and drift monitoring compares current flag rates against historical baselines.

Activity 3: "Read Before Class"

No model answer—this is student research. During class discussion, listen for whether students identified the data inputs, the automated decision boundary, and the presence or absence of human oversight.

