

AI for Business: Insights from Corporate Data

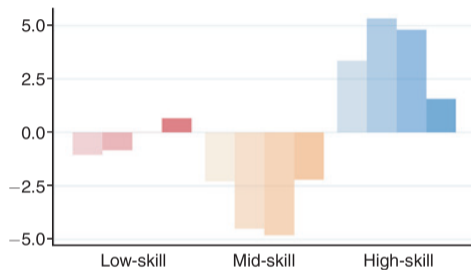
Topic 1: Overview of Business Analysis with AI

Miao Liu

Boston College

January 15, 2026

Before we start...



■ Health, personal, cleaning and security, operators, laborers
■ Production, clerical, admin, sales
■ Professional, technical, managerial

■ 1970-1980
■ 1980-1990
■ 1990-2000
■ 2000-2016

Course Roadmap (Topics 1–5)

Topic	
1	The AI Era of Business Analysis: Overview
2	Financial Statements, Ratios, and LLM-Aided Interpretation
3	Operating Assets & Quality, and LLM-Aided Hypothesis Generator
4	Earnings Quality, and Corporate Valuation using an AI Agent
5	In-class Lab

Course Roadmap (Topics 6–9)

Topic	
6	Foundations of Machine Learning
7	Trees and Neural Networks
8	Applications of Machine Learning
9	Foundations of Natural Language Processing
10	Large Language Models

Course Roadmap (Weeks 10–13)

Topic

- | | |
|----|---|
| 11 | LLM-Aided Deep Market Analysis |
| 12 | Integrative Business Analysis: Overview |
| 13 | Integrative Business Analysis: Building the Model |

Assessment and Grading

Component	Weight
Lab Assignments and Cases (8 write-ups)	40%
Quiz and in-class Lab	35%
Final Project	20%
Class Participation (discussion + in-class activities)	5%

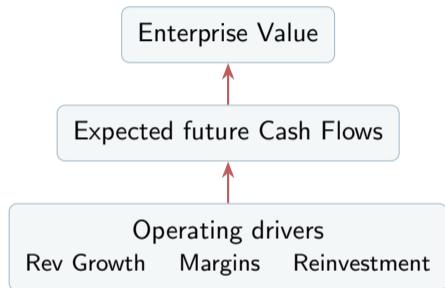
Overview

- 1 Foundations of Business Analysis
- 2 From Data to Drivers
- 3 What an LLM Is (and Isn't) for Business Analysis
- 4 Human-AI collaboration
- 5 Business & Industry Analysis

Overview

- 1 Foundations of Business Analysis
- 2 From Data to Drivers
- 3 What an LLM Is (and Isn't) for Business Analysis
- 4 Human-AI collaboration
- 5 Business & Industry Analysis

Foundations of Business Analysis: The Driver Tree

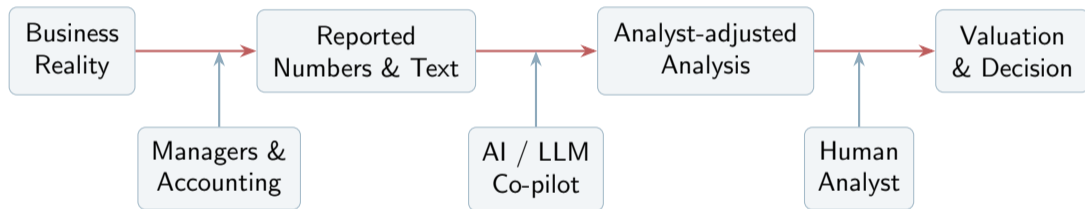


- ▶ **Revenue Growth:** Sustainable expansion of sales driven by changes in volume, price, and mix over time
- ▶ **Operating Margins:** Portion of revenue that remains after operating costs, reflecting pricing power, cost structure, and operating leverage
- ▶ **Reinvestment:** Incremental capital and operating spending required to sustain and grow the business (capex, working capital, and intangible investment such as R&D)

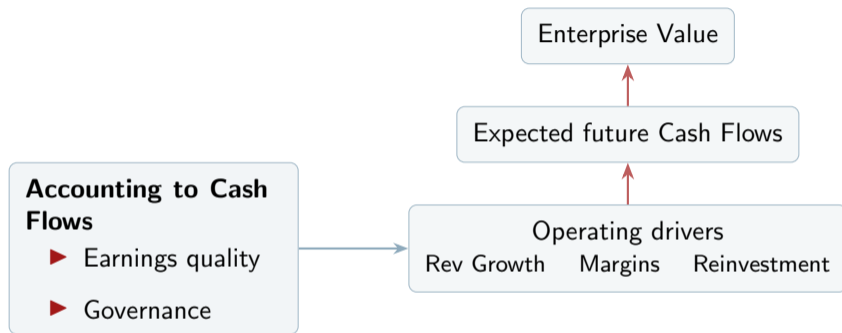
Business Analysis as a Translation Pipeline



Business Analysis as a Translation Pipeline

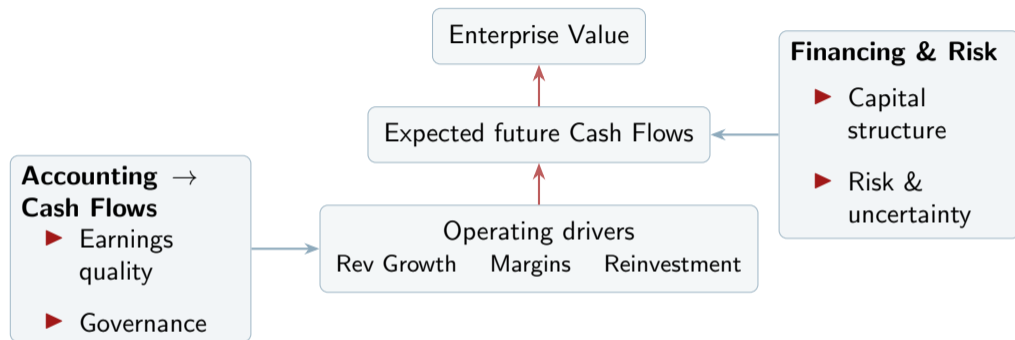


The Complete Driver Tree



- ▶ **Earnings quality:** Extent to which reported earnings reflect sustainable economic performance and predict future cash flows rather than transitory items/discretion
- ▶ **Governance:** Incentive and oversight structure that shapes disclosure credibility, managerial decisions, and value-destructive behavior

The Complete Driver Tree

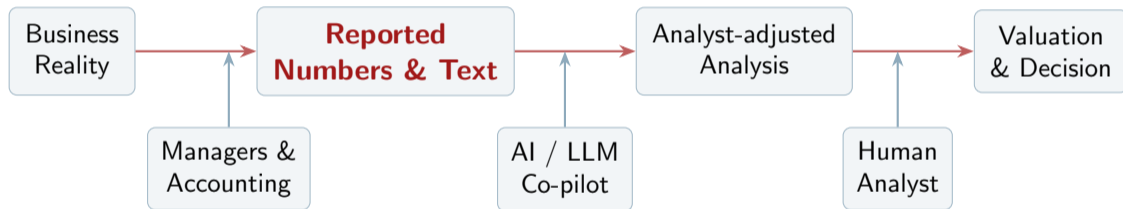


- ▶ **Capital structure:** The debt–equity mix that determines financial flexibility and downside fragility
- ▶ **Risk & uncertainty:** The uncertainties that shape scenario design and the required return (macro, competition, regulation, climate)

Overview

- 1 Foundations of Business Analysis
- 2 From Data to Drivers**
- 3 What an LLM Is (and Isn't) for Business Analysis
- 4 Human-AI collaboration
- 5 Business & Industry Analysis

Business Analysis as a Translation Pipeline



Data: Disclosure Ecosystem (Roadmap)

- ▶ **10-K:** baseline architecture
- ▶ **10-Q:** inflection points and updating priors
- ▶ **8-K:** material events and expectation resets
- ▶ **Proxy Statement:** incentives and credibility of guidance
- ▶ **Calls/Decks:** forward view and KPI discipline

Data: Disclosure Ecosystem — Form 10-K

Form 10-K (annual): the baseline disclosure architecture

Typically contains: Audited financials; MD&A; risk factors; business description; segment discussion; legal proceedings; accounting policies and estimates.

Why it matters for valuation: Sets the baseline for growth, margin, reinvestment, and risk analysis.

Data: Disclosure Ecosystem — Form 10-Q

Form 10-Q (quarterly): inflection points and updating priors

Typically contains: Unaudited interim financials; MD&A updates; risk updates; selected footnotes.

Why it matters for valuation: High-frequency updates and inflection points (margin compression, working-capital swings, guidance changes).

Data: Disclosure Ecosystem — Form 8-K

Form 8-K (current): material events and expectation resets

Typically contains: Event-driven disclosures (earnings releases, M&A announcements, leadership changes, major contracts, restructuring)

Why it matters for valuation: Often the first place material changes appear.

Data: Disclosure Ecosystem — Proxy (DEF 14A)

Proxy statement (DEF 14A): incentives and governance

Typically contains: Executive compensation; governance; board composition; shareholder proposals; related-party transactions; CD&A.

Why it matters for valuation: Incentives and governance are inputs to earnings-quality assessment and the credibility of forward guidance.

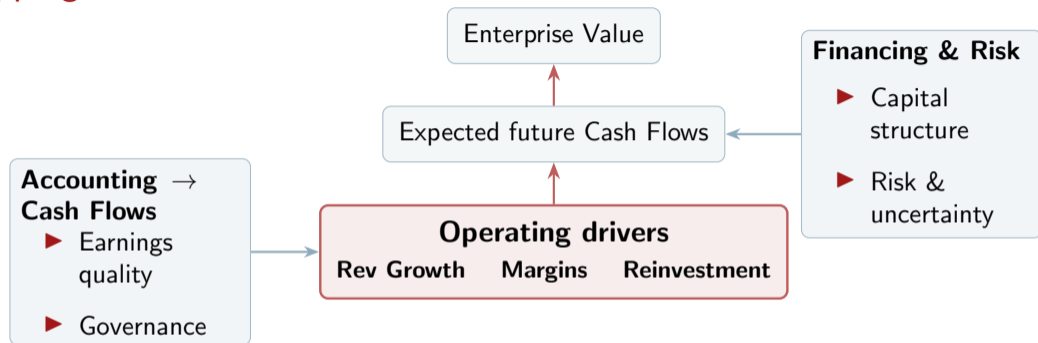
Data: Disclosure Ecosystem — Calls / Investor Days / Presentations

Earnings calls, investor days, presentations: forward view and KPI discipline

Typically contains: Prepared remarks; Q&A; KPI narratives; strategy and competitive positioning.

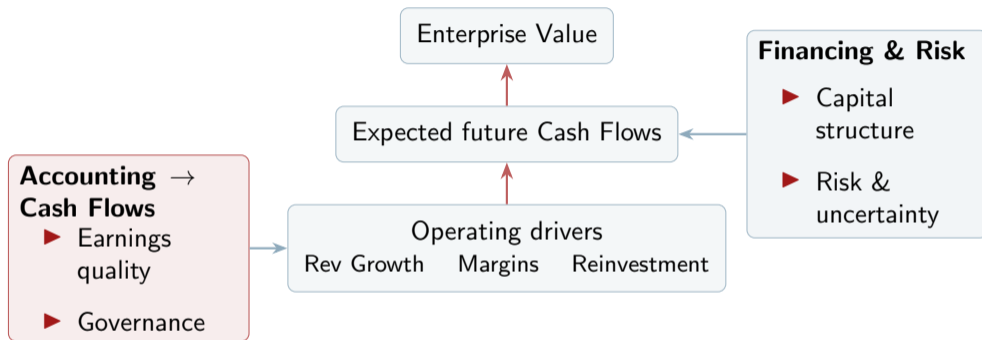
Why it matters for valuation: Forward-looking information; Q&A often reveals the “edges” of the story.

Mapping Drivers Data



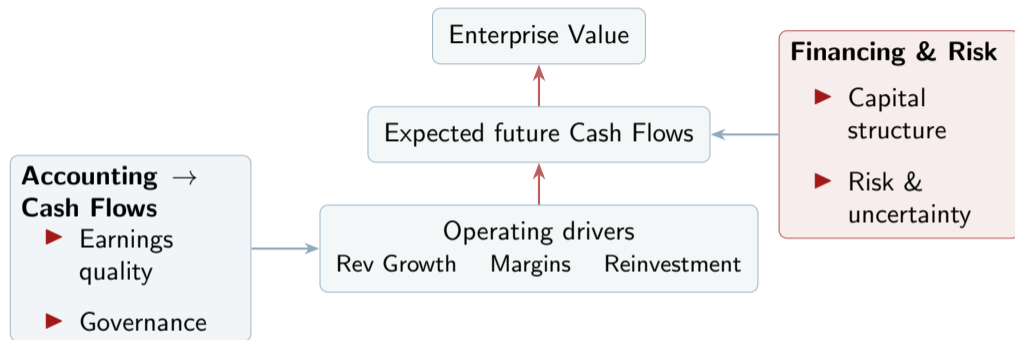
- ▶ **Rev Growth:** rev recog footnotes (10K); demand drivers in MD&A (10K/Q)
- ▶ **Margins:** segment footnotes (10K); margin walk in MD&A (10K/Q); cost shocks/restructuring (8K); Q&A nuance (*earnings calls*)
- ▶ **Reinvestment:** cash flow statement (10K/Q); capex/leases in notes (10K); working-capital drivers in MD&A (10K/Q)

Mapping Drivers Data



- ▶ **Earnings quality:** special items footnotes (*10K*); non-GAAP reconciliations (*8K/10K*); accrual components (*10K*); “one-time” discussion in MD&A (*10K/Q*)
- ▶ **Governance:** proxy statement (CD&A) (*DEF 14A*)

Mapping Drivers Data



- ▶ **Capital structure:** debt footnotes (10K); covenant in credit agreements (10K); maturity schedules (10K); liquidity discussion (10K/Q)
- ▶ **Risk & uncertainty:** risk factors (10K/10Q); legal proceedings (10K); sensitivity disclosures (10K)

Overview

- 1 Foundations of Business Analysis
- 2 From Data to Drivers
- 3 What an LLM Is (and Isn't) for Business Analysis**
- 4 Human-AI collaboration
- 5 Business & Industry Analysis

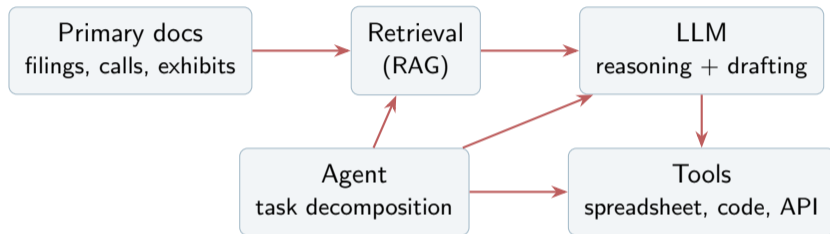
What is an LLM (intuitive)

- ▶ An LLM predicts the **next token** given context:

$$\Pr(\text{token}_t \mid \text{token}_{1:t-1})$$

- ▶ Modern LLMs are typically built on the **Transformer** architecture.

From model to system: Retrieval, Tools, and Agents



- ▶ **RAG:** grounded to retrieved text.
- ▶ **Tools:** offload computation, parsing, and API calling.
- ▶ **Agents:** task decomposition.

Two silent failure modes (and the rules they motivate)

Failure mode	Why it matters in valuation
Definition drift	KPI/non-GAAP terms quietly change meaning across sections/time (e.g., “free cash flow,” “adjusted EBITDA,” “bookings”).
False precision	Specific numbers/citations/causal claims appear plausible but are unsupported; they propagate into models.

- ▶ **Rule 1.1A:** preserve definitions as first-class objects.
- ▶ **Rule 1.1B:** treat specificity as suspect until tied to source lines/cells/assumptions.

The Fluency Trap of LLMs in Corporate Valuation

- ▶ **Persuasion becomes cheap.** The same tools that accelerate drafting and summarization also make it easy to produce fluent, rhetorically effective narratives—even when evidence is thin.
- ▶ **The fluency trap.** Coherence on demand can turn a tentative hypothesis into an unchallenged “fact” simply because it reads well.

Overview

- 1 Foundations of Business Analysis
- 2 From Data to Drivers
- 3 What an LLM Is (and Isn't) for Business Analysis
- 4 Human-AI collaboration**
- 5 Business & Industry Analysis

Task frontier I: information extraction and synthesis

Task	LLM comparative advantage
Map disclosures; build reading plan	Outline sections; identify cross-references; generate “what to read next” questions
Extract definitions (KPIs; non-GAAP)	Build a definition ledger; flag inconsistencies; track definition changes
Summarize MD&A and risks	Compress into driver language; map narrative → drivers; propose hypotheses
Footnote interpretation (leases, rev. rec., pensions)	Extract key definitions; surface implications as hypotheses; propose verification steps

Task frontier II: modeling, judgment, and communication

Task	LLM comparative advantage
Compute ratios/multiples	Not a core LLM task. Use spreadsheets/code; LLM can explain interpretation
Peer set design	Propose comparability criteria; generate candidate peer logic; draft exclusion rules
Scenario design (bull/base/bear)	Draft scenario narratives; surface missing evidence; act as a red-team critic
Draft memo/deck narrative	Improve clarity and logic; rewrite for audience; generate counterarguments
Model auditing	Generate audit checklists; highlight consistency questions; propose sensitivity grids

LLMs: Pros, Cons, and the Required Reframe

Side	Summary
Pros	Information extraction and synthesis: disclosure mapping, definition ledgers, MD&A/risk summarization, footnote hypothesis generation.
	Modeling, judgment, communication support: peer-set logic drafts, scenario narratives, red-team critique, memo/deck rewriting for clarity.
Cons	Truth is not guaranteed: hallucinations, invented citations, scope errors.
	Fluency trap: persuasive coherence can outrun evidence and accelerate overconfidence.

Reframe: This tradeoff requires a new framework—treat corporate valuation as **work-flow design**, not tool adoption.

Human-AI Collaboration in Corporate Valuation: Audit Trail

- ▶ Without a record of what was prompted, accepted, and verified, teams cannot separate **evidence** from **inference**.

Minimum audit trail	What it contains
Objective of LLM use	Evidence mapping, red-team critique, writing assistance
Inputs provided	Document names/dates, excerpts, constraints
Prompts	Key prompts used
Outputs used	Only outputs that materially shaped the work product
Verification log	What you verified and how (filing cites; spreadsheet tabs/cells)
Rejected outputs	What you rejected and why (unsupported; fabricated)

Rule: Numbers and factual claims must be verified against primary documents or trusted data providers.

Overview

- 1 Foundations of Business Analysis
- 2 From Data to Drivers
- 3 What an LLM Is (and Isn't) for Business Analysis
- 4 Human-AI collaboration
- 5 Business & Industry Analysis**

Why Business & Industry Analysis Comes First

- ▶ Business Analysis and Valuation is often taught as a technical model capturing various business value drivers.
- ▶ In practice, this is a **economic problem before a technique problem**.
- ▶ If you do not first understand **what the business is** and **what game it is playing**, you can build a spreadsheet that is internally consistent but **economically wrong**.

Failure Example 1: eBay (Marketplace \neq Retailer)

- ▶ **Retailer archetype:**
Revenue engine: product sales (gross revenue); margins driven by COGS and product gross margin.
- ▶ **eBay archetype:**
eBay primarily connects buyers and sellers; it monetizes via fees on transactions - GMV(Gross Merchandise Value).
- ▶ **What you miss if you treat eBay as Walmart:**
 - ▶ **Wrong revenue model:** buyer/seller retention and repeat behavior drive long-run GMV durability.
 - ▶ **Wrong cost model:** (COGS/inventory) instead of platform operating costs.

Failure Example 2: Duke/NextEra (Regulation \neq Commodity Competition)

- ▶ **Commodity competition archetype:**
Margins set by market price competition; key threat is entry.
- ▶ **Regulated utility archetype:**
Earnings governed by regulatory frameworks: allowed returns.
- ▶ **What you miss if you model it like a commodity producer:**
 - ▶ **Wrong Rev growth and margin model:** growth and margin tied to regulation more than entry of competitors.
 - ▶ **Wrong risk model:** entry is not the primary threat; political and regulatory shifts can change the rules of the game.

Data (10-K): Where to Extract What You Need

10-K location	High-yield extracts for business & industry analysis
Item 1 (Business)	Business model; segments; customers; distribution; seasonality; strategy; competitive landscape; principal offerings
Item 1A (Risk Factors)	Constraints and bottlenecks; dependence on partners/suppliers; regulatory exposure; competitive threats
Item 7 (MD&A)	Management's causal narrative: what drove revenue/cost changes; KPI commentary; forward framing
Financial statement notes	Segment definitions; revenue disaggregation; major customers; geographic exposure; contracts; accounting policies interacting with the business model

But we should treat it as an architecture, not a narrative.

Why we need a Business Model Blueprint

- ▶ Business & industry analysis is **messy by default**: filings are long and cross-referenced.
- ▶ The constraint is rarely access to information; it is the lack of a **compact economic representation**.

Business Model Blueprint = a systematic, organized way to map messy narrative disclosures into a structured representation of the firm's business archetype.

Business Model Blueprint

- | # | Blueprint field |
|---|---|
| 1 | Customers and use cases (who pays, who uses, who decides) |
| 2 | Value proposition (why customers choose the firm) |
| 3 | Monetization (pricing metrics and revenue recognition logic) |
| 4 | Cost structure (fixed vs. variable; major cost pools) |
| 5 | Key assets/capabilities (tangible, intangible, organizational) |
| 6 | Key constraints and risks (competition, regulation, technology, execution) |
| 7 | KPI spine (the smallest set of KPIs that track the economic model) |

eBay Business Model Blueprint (1/3): Customers & Value Proposition

(A) Field	(B) Evidence (10-K)	(C) Interpretation	(D) Driver-tree implication
Customers & use cases	Item 1: buyers & sellers; "active buyers"; global markets; listings/transactions	Two-sided marketplace: demand (buyers) and supply (sellers). Sellers are also paying customers via fees/ads; activity is transaction-driven.	Growth: active buyers, GMV; Margins: take rate, ads mix; Reinv.: platform/payments; Risk: trust/fraud shocks.
Value proposition	Item 1: "value and unique selection"; focus on "enthusiasts"; marketplace discovery	Differentiation via selection/-value + category depth. Strategy leans on discovery and trusted transactions to sustain repeat behavior.	Growth: conversion, repeat; Margins: category/ads mix; Reinv.: discovery & trust tools; Risk: competition/quality erosion.

eBay Business Model Blueprint (2/3): Monetization, Costs, Capabilities

(A) Field	(B) Evidence (10-K)	(C) Interpretation	(D) Driver-tree implication
Monetization	Item 1: marketplace + advertising; take rate; seller fees; promoted listings/ads	Revenue scales with marketplace activity (GMV) and monetization (take rate + ads). Key lever is take-rate and ads penetration, not inventory margin.	Growth: GMV, ads adoption; Margins: take rate, ads mix; Reinv.: monetization tools; Risk: take-rate compression.
Cost structure	Item 7: cost of net revenues (payments/site ops/support); S&M; product dev; transaction losses	Platform cost base + demand generation + trust costs. Unlike a retailer, costs are not COGS-heavy; payments and trust/safety are central.	Growth: S&M efficiency; Margins: cost ratios, losses; Reinv.: product dev; Risk: fraud/loss spikes.
Assets/capabilities:	Item 1: platform/apps; managed payments; seller tools; ads stack; trust features	Core capabilities: discovery + marketplace matching + integrated payments + seller tools/ads. These support conversion and monetization.	Growth: retention/cohorts; Margins: payment efficiency, ads effectiveness; Reinv.: tech/payments; Risk: execution/quality.

eBay Business Model Blueprint (3/3): Constraints/Risks & KPI Spine

(A) Field	(B) Evidence (10-K)	(C) Interpretation	(D) Driver-tree implication
Constraints & risks	Item 1A: competition; trust/safety; regulatory/privacy; platform dependence	Key constraints are competitive intensity and marketplace trust. Risk regimes include fraud/transaction losses and regulatory constraints on payments/data/marketplaces.	Risk: scenario weights; Growth: demand shocks; Margins: take-rate pressure; Reinv.: compliance & safety spend.
KPI spine	Item 7: GMV, take rate, active buyers; Item 7: transaction losses; revenue mix	Minimum spine should track: activity (GMV), demand (active buyers), monetization (take rate + ads), and trust costs (transaction losses).	Growth knobs: GMV, buyer retention; Margin knobs: take rate, ads mix; Risk knobs: loss rate; Reinv.: product/safety investment.

Why a KPI Spine Ledger Matters

- ▶ A **driver tree** tells you what must be forecast (growth, margins, reinvestment, risk).
- ▶ A **KPI spine ledger** tells you how to **observe** those drivers in the 10-K:
 - ▶ the **definition** (what the KPI means),
 - ▶ the **location** (where it lives in the filing),
 - ▶ the **control link** (which driver node/model knob it informs).
- ▶ It prevents **definition drift** (same term, different meaning) and keeps your analysis auditable.

Rule of thumb: If a KPI is not defined and sourced, it should not become a forecast input.

KPI Spine Ledger (1/2)

KPI	Definition / computation	Where in 10-K	What it controls
Active buyers	Defined count of buyers who paid for a transaction in the trailing period (per company definition).	Item 1 / Item 7 (Key metrics)	Growth: demand base; cohort health proxy.
GMV	Gross Merchandise Volume (total value of goods transacted on platform; per definition).	Item 7 (Key metrics)	Growth: marketplace activity; volume driver.
Take rate	Marketplace revenue divided by GMV (per definition / computed if components disclosed).	Item 7 (Key metrics)	Margins: monetization efficiency; pricing power vs competition.
Advertising revenue (Promoted listings / ads)	Net revenues from advertising activities (as defined/discussed).	Item 1 / Item 7	Growth & margins: monetization mix uplift; seller demand sensitivity.

KPI Spine Ledger (2/2)

KPI	Definition / computation	Where in 10-K	What it controls
Transaction losses	Provision / losses related to fraud, chargebacks, and transaction integrity (as defined).	Item 7 (Expenses / discussion)	Risk & margins: trust/safety shock; scenario trigger.
Sales & marketing expense ratio	Sales & marketing expense divided by net revenues.	Item 7 (Operating expenses)	Growth efficiency: CAC-like proxy; operating leverage.
Product development expense ratio	Product development expense divided by net revenues.	Item 7 (Operating expenses)	Reinvestment: product/tech intensity to sustain platform.
Cost of net revenues ratio	Cost of net revenues divided by net revenues (payments, ops, support).	Item 7 (Cost of net revenues)	Margins: payments/site-ops efficiency; scaling effects.

Reading Assignment and Lab 1

- ▶ Chapter 1 and Chapter 2: 2.3-2.5
- ▶ **Lab Assignment 1**