Artificial Intelligence for Business Intelligence

Week 5: BI in Customer Management

Chi-Heng Yang

Ph.D. Candidate

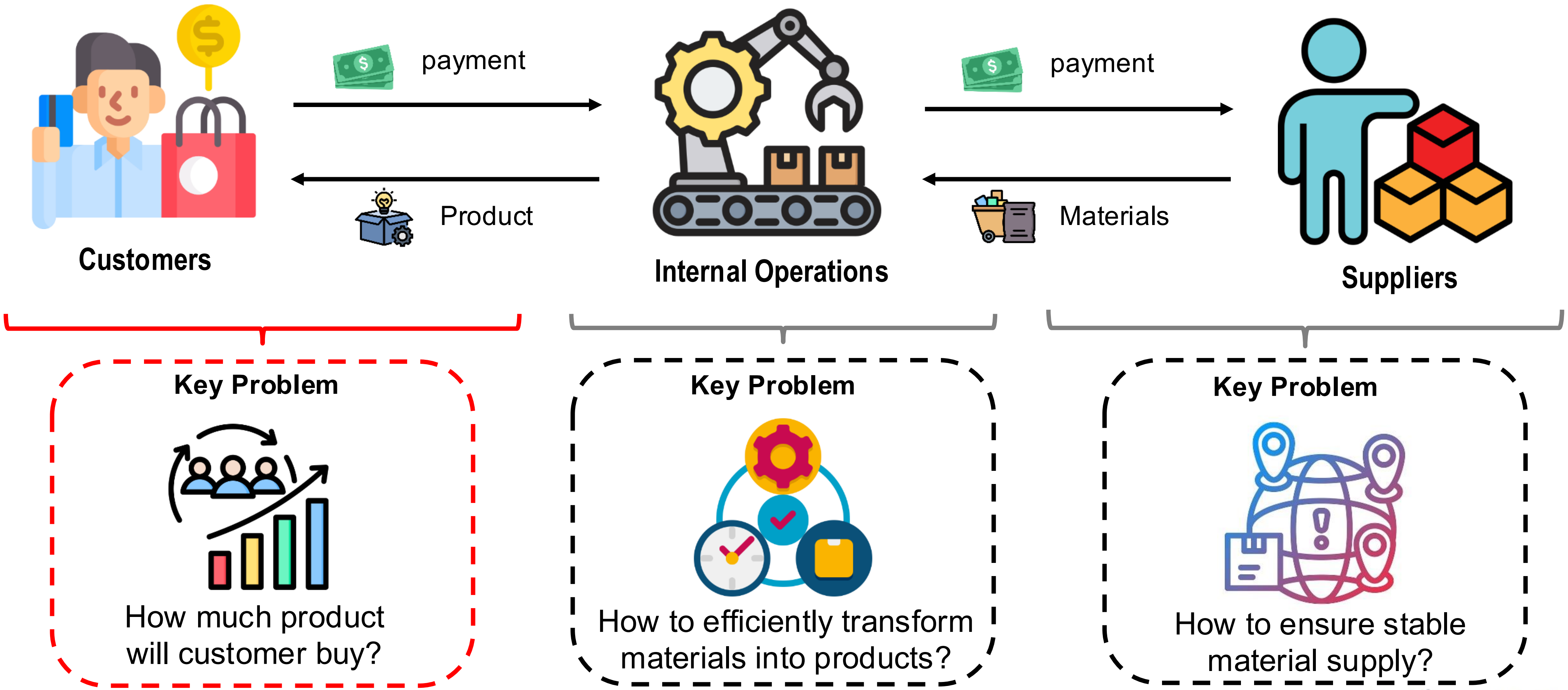
Acknowledgements: Sheng-Ya Chiu, Dr. Hsinchun Chen

Online M.S. in AI for Business

Today's Learning Objectives

- Understand the background and critical challenges in demand forecasting.
- Explain the difference between Business-to-Consumer (B2C) and Business-to-Business (B2B) demand forecasting and their specific challenges.
- Identify common demand signals and the methods used to collect them.
- Explain the pros and cons of statistical, machine learning, and deep learning methods, and the specific data characteristics they capture.
- Understand how demand forecasting delivers business value through visualization and replenishment decisions.

Business Operation



Customer-Company Interaction

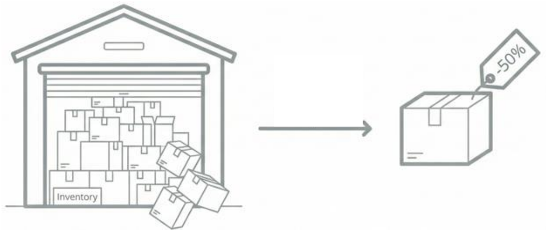
- Customer management is critical for a company as customer purchasing patterns directly drive future profit.

- Companies hold **inventory** to make products available when customers are ready to buy.
  - However, failing to balance the inventory level leads to stockouts or overstock costs.

Stockout

- Stockouts cause immediate lost sales and long-term damage to customer relationships
- Customers leave disappointed and may switch to competitors

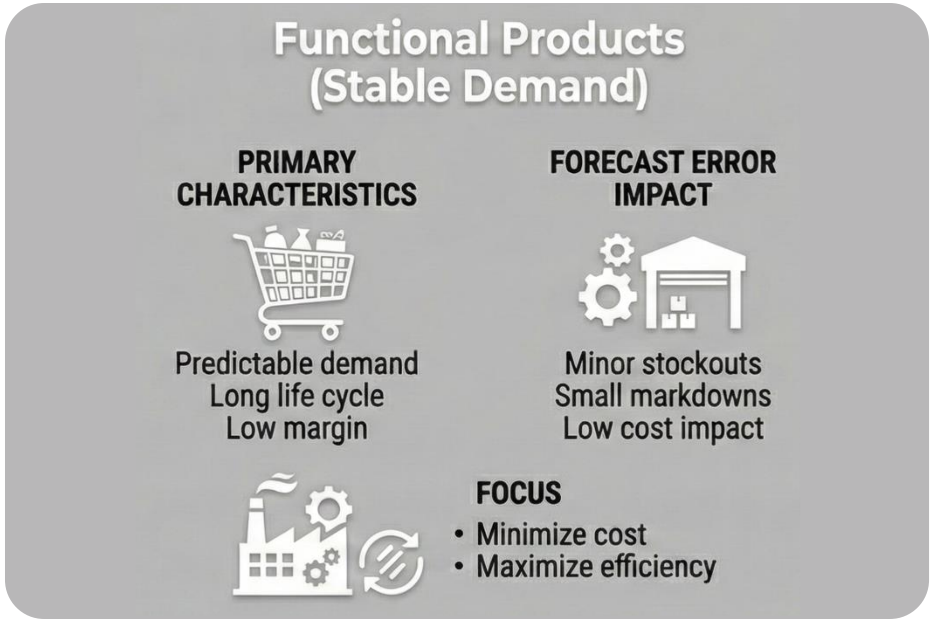
Overstock

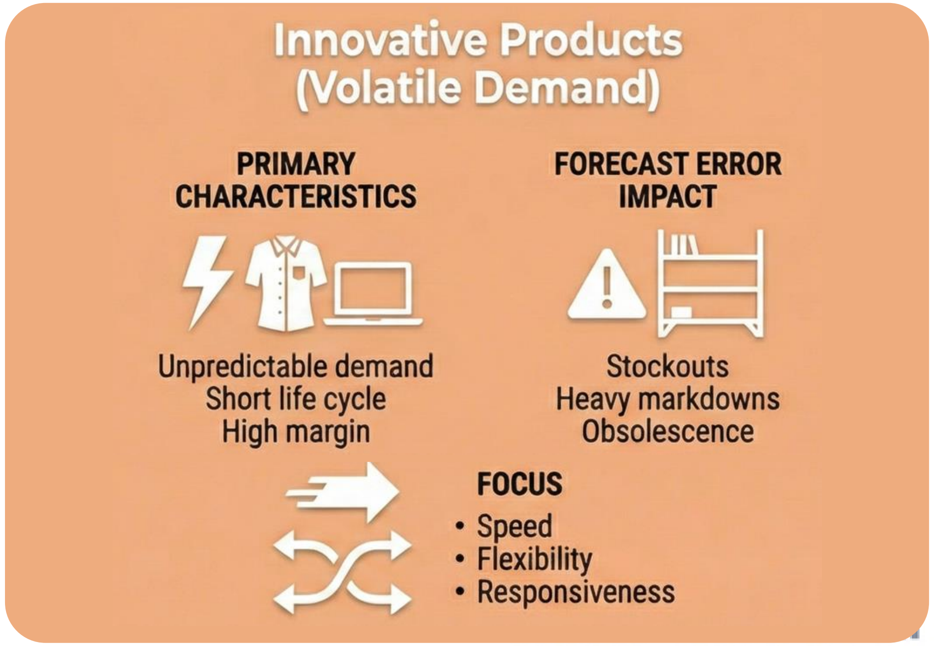


- Overstock increases holding costs and risk of obsolescence or markdowns
- Cash is tied up in unsold goods sitting in warehouses

Cost of Forecast Error

- The consequences depend heavily on product value and the criterion.
- Forecast errors are more costly for innovative than for functional products (Fisher 1997).
- Innovative items have short life cycles and high obsolescence.
    - **Innovative** (e.g., fashion apparel): a missed seasonal jacket forecast either causes lost high-margin sales or heavy discounts at season's end.
    - **Functional** (e.g., household staples): if toothpaste is over-forecasted, the extra can usually be sold later without markdown.





Reference: Fisher, M. L. (1997). What is the right supply chain for your product? Harvard Business Review, 75(2), 105–117.

Demand Forecasting

- To prepare a proper amount of inventory for customers, companies perform **demand forecasting** to estimate future customer purchases.

- Demand forecasting supports multiple decisions:
  - **Inventory replenishment**: when and how many units of each product to make or order?
  - **Pricing and promotions**: when to discount, bundle, or advertise to smooth demand peaks and valleys?
  - **Resource allocation**: how to schedule staff, warehouses, and transport to match expected demand?

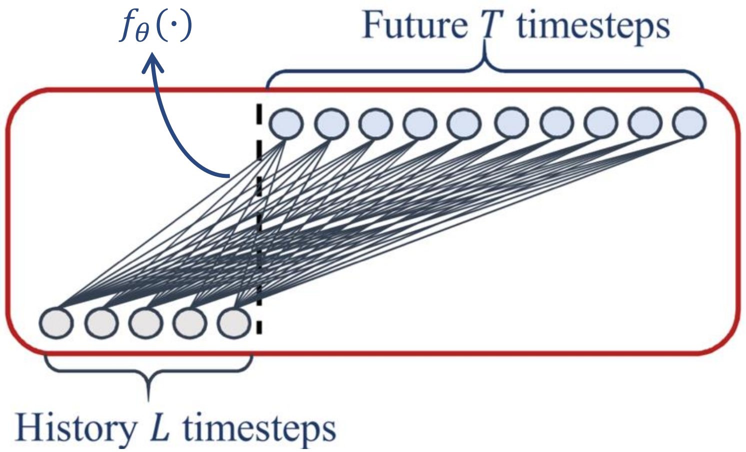
Demand Forecasting

- Demand forecasting aims to predict future demand by analyzing historical time series data.
  - Multiple historical time series are often employed (e.g., weather, economic indicators, events).

- Demand forecasting can be formally defined as

$$\hat{x}_{L+h} = f_\theta(X_{1:L}), h = 1, \dots, H$$

  - $f_\theta(\cdot)$ is a forecasting model to map history to the future
  - $x_t = (x_t^1, x_t^2, \dots, x_t^J) \in R^J$ is the history
    - $t = 1, \dots, L$ the time index
    - J is the number of channels (variables, e.g., weather, sales)
    - x_t^j is the value of channel j at time t (e.g., demand amount)



Reference: Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023, June). Are transformers effective for time series forecasting?. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 9, pp. 11121-11128).

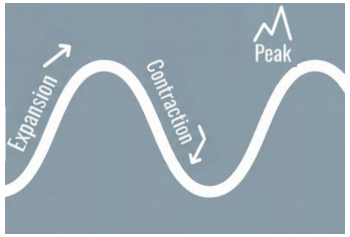
Demand Forecasting

- Multiple types of patterns exist in historical time series data.

Seasonal

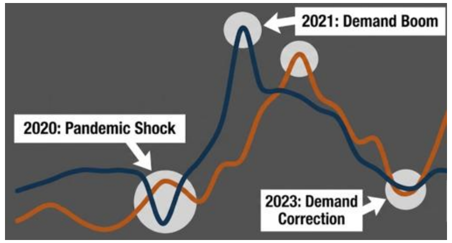
Time Scale: Within-year
Predictability: High
Scope: Many products
Example: Christmas spike

Cyclical



Time Scale: Multi-year
Predictability: Medium
Scope: Economy-wide
Example: Recession shift

Shock



Time Scale: Irregular
Predictability: Low
Scope: Product/local
Example: Pandemic

Stable

Time Scale: Persistent
Predictability: High
Scope: Necessities
Example: Milk

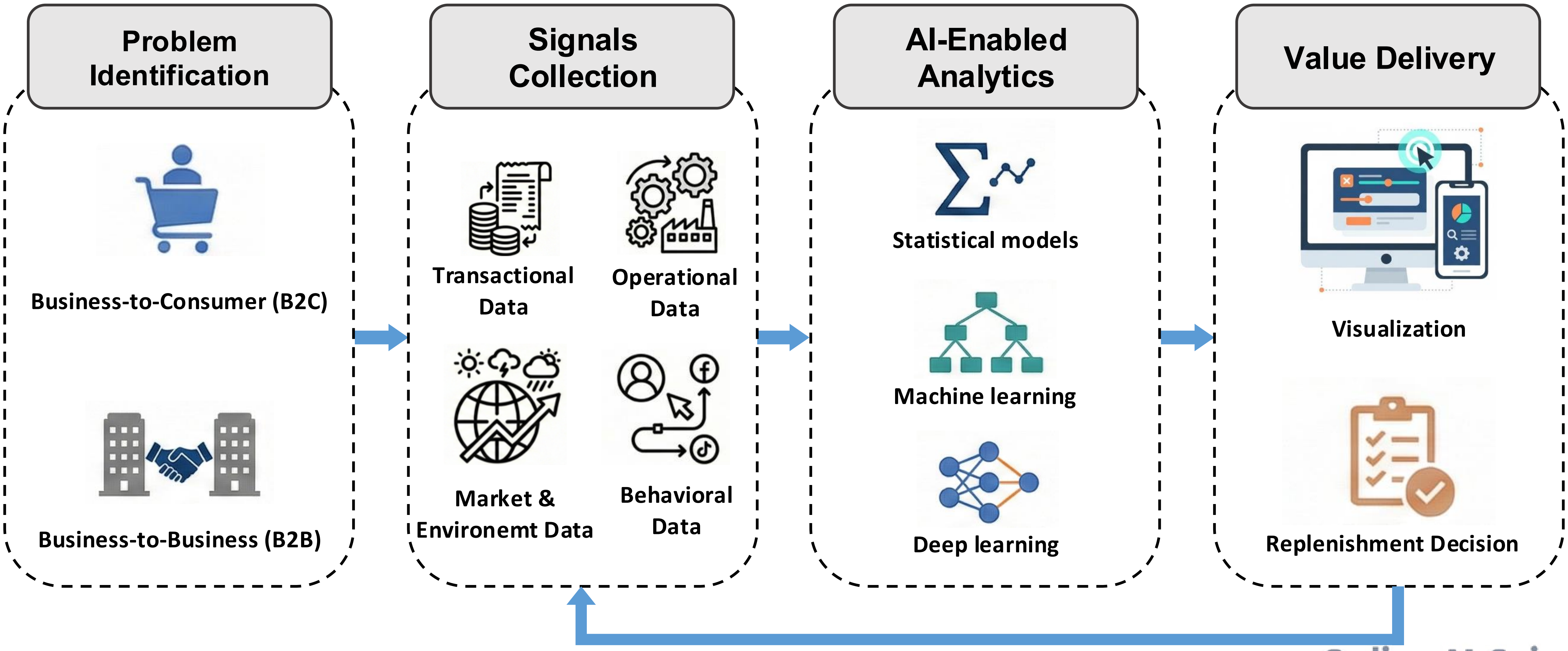
AI for Demand Forecasting

- Capturing historical time series data is challenging as:
  - There are often interactions between them (e.g., hot weather only trigger surge in air conditioner sales when the economy is good).
  - Their patterns are often nonlinear (e.g., seasonality) and complex.

- Artificial intelligence has emerged as a potential method because it can integrate multiple data sources and model the nonlinearity.

- The AI4BI development process can be applied to guide the AI development in demand forecasting problems.

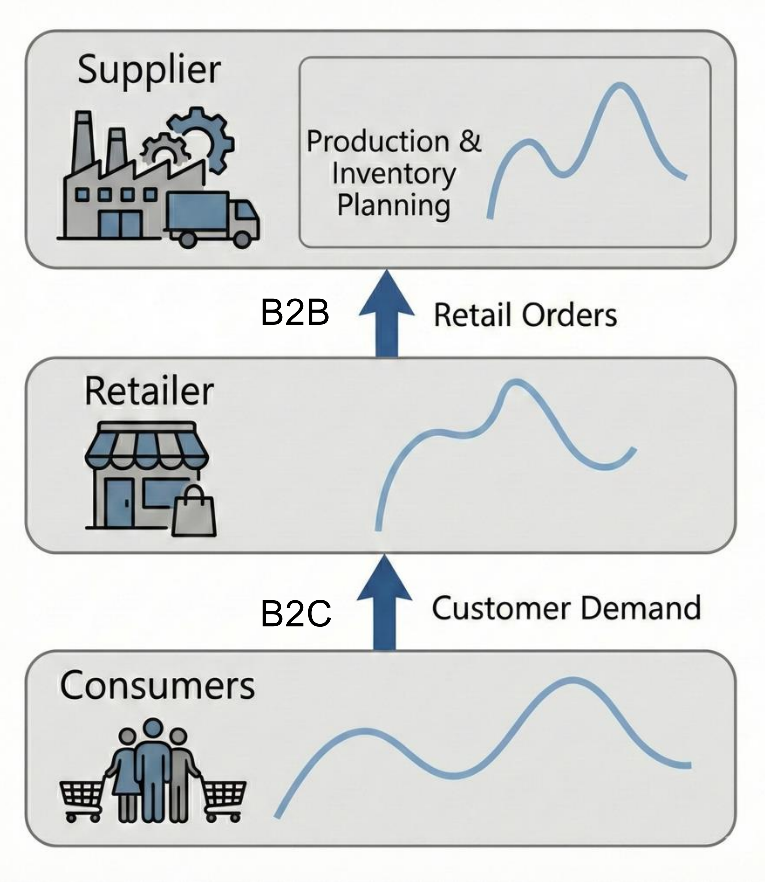
AI4BI Development Process for Demand Forecasting



Decisions change future data (feedback loop)

Step 1: Problem Identification

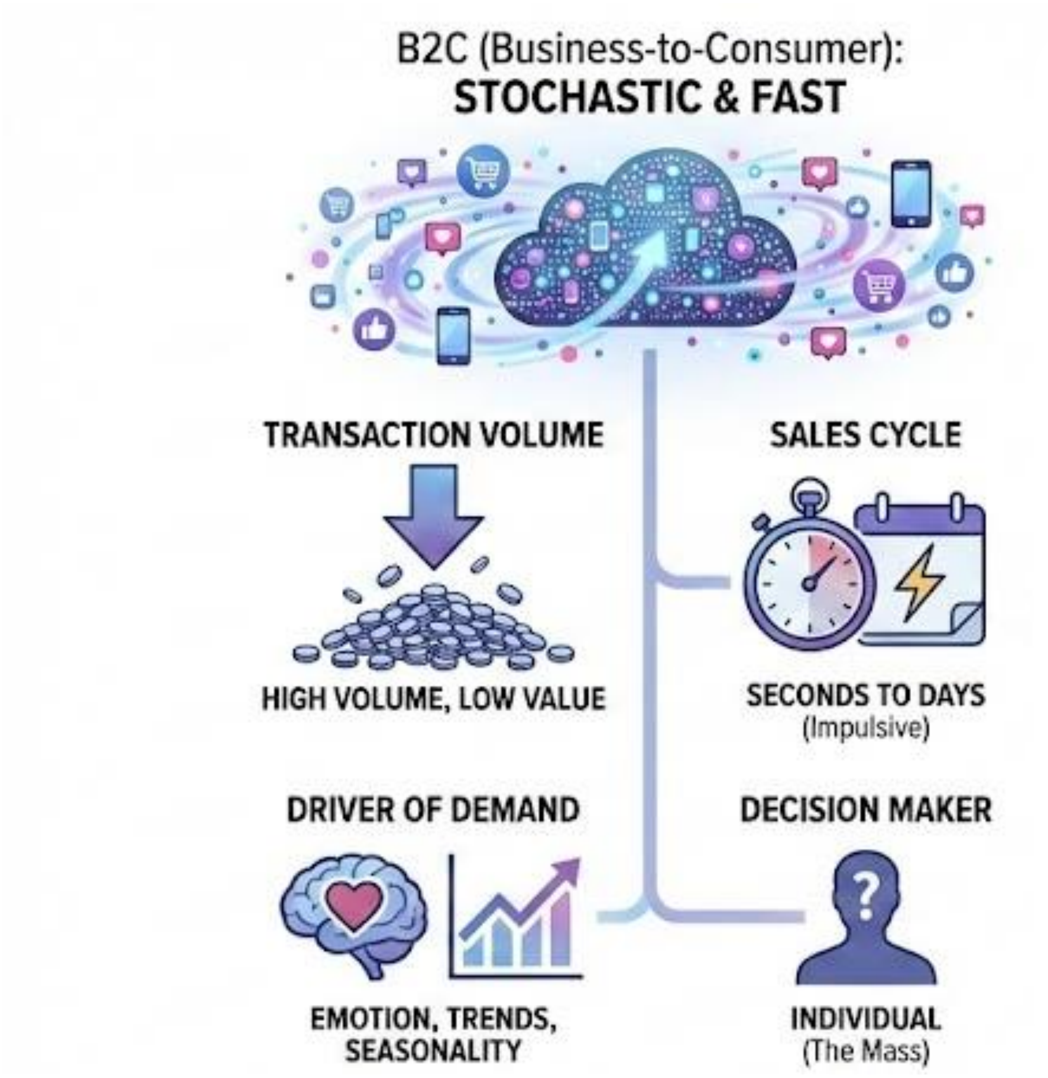
- The objective of demand forecasting depends on the target entity.
- **Business-to-Consumer (B2C)** targets direct demand from consumers.
  - Predicts individual shopping habits across large, diverse populations to manage inventory.
- **Business-to-Business (B2B)** target the order from retailers.
  - Focuses on high-value contracts and specific relationship cycles with other companies.

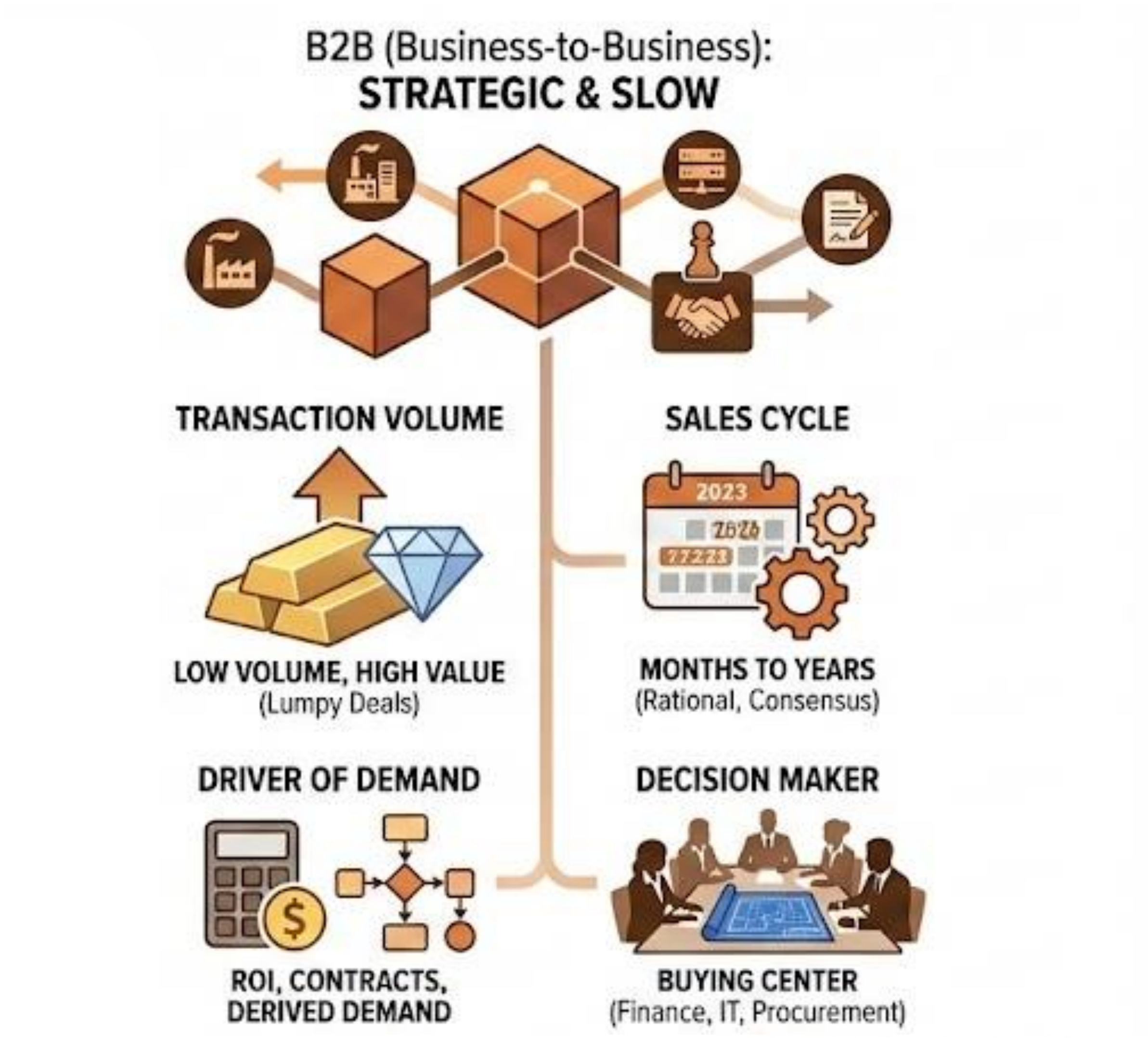


References:

- B2C: Hu, Y. J., Rombouts, J., & Wilms, I. (2025). Fast forecasting of unstable data streams for on-demand service platforms. Information Systems Research, 36(1), 552-571.
- B2B: Ali, M. M., & Boylan, J. E. (2011). Feasibility principles for downstream demand inference in supply chains. Journal of the Operational Research Society, 62(3), 474-482.

Problem Identification – B2C & B2B





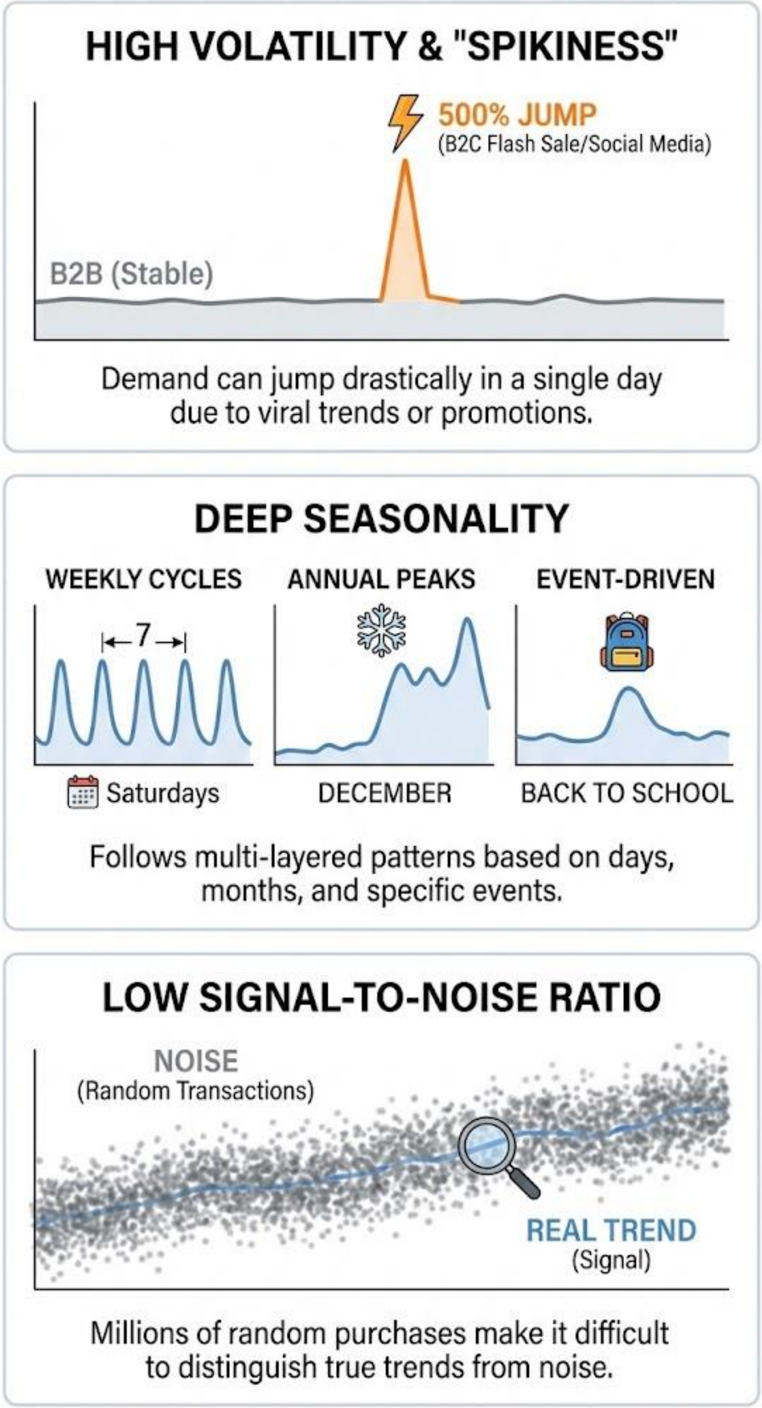
Problem Identification – B2C

- In B2C setting, demand forecasting aims to predict individual consumer behavior at scale.
  - Look for a “yes/no” question from millions of individuals aggregated together.

Problem Identification – B2C

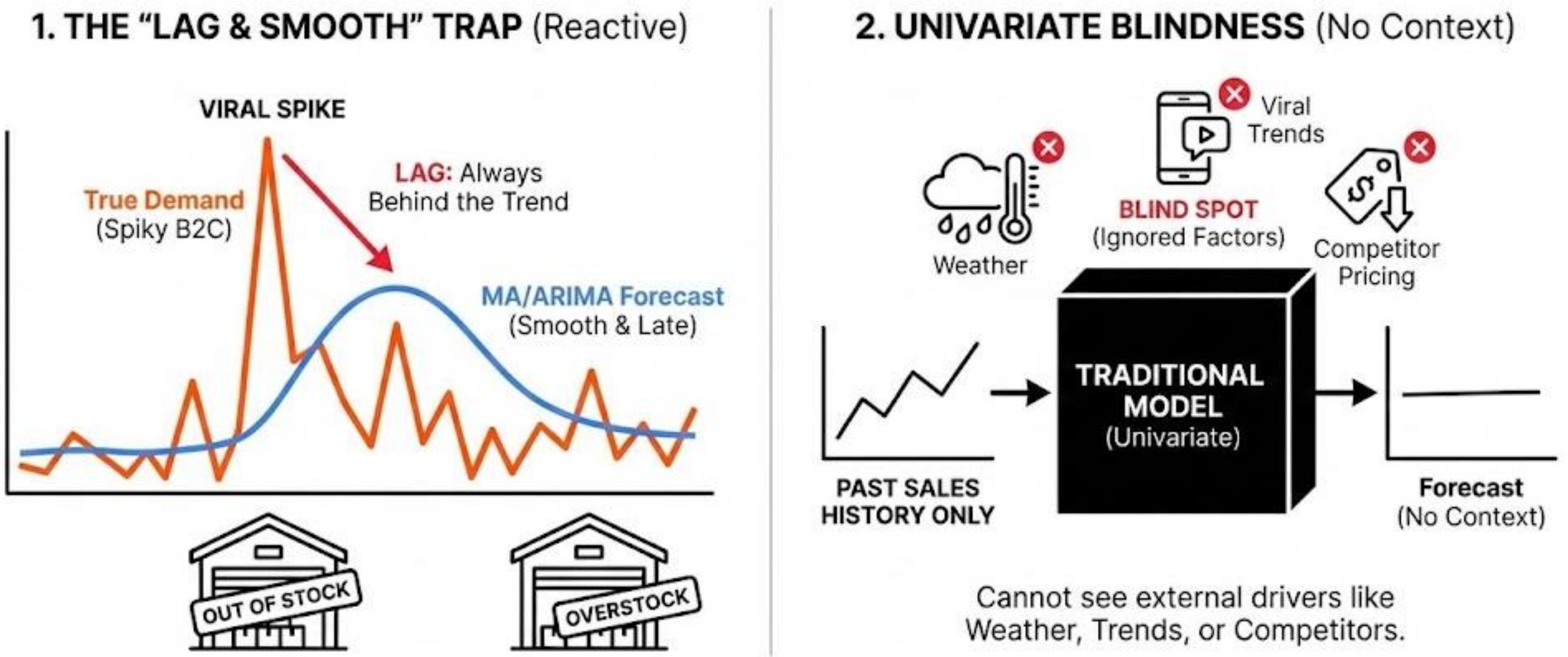
- **Characteristics of demand**: high volatility, deep seasonality, and low signal-to-noise ratio.

- **Traditional methods**:
  - **Moving Average**: Simple averages of the last n periods.
  - **Exponential smoothing**: Give more weight to recent data points while accounting for error, trend, and seasonality.
  - **ARIMA**: look at the correlation between a variable and its own past values to find patterns.



Problem Identification – B2C

- However, these methods have limitations:
  - They use historical averages to smooth out fluctuations ➔ always behind the trend or spike
  - They only look at past demand data ➔ ignore external factors (e.g., competitors, weather)



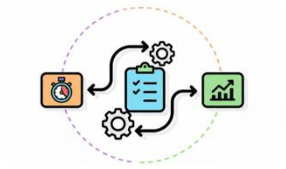
B2C – Key Components

Stakeholder

- **Primary user**: Merchandising planners or marketing teams in retailers
- **Goal**: Understand and anticipate end-consumer purchasing behavior

Scenario

- **Task**: Predict sales at product/stock keeping unit (SKU) / store level per region
- **Process**: Use past data to predict future demand



Current approach

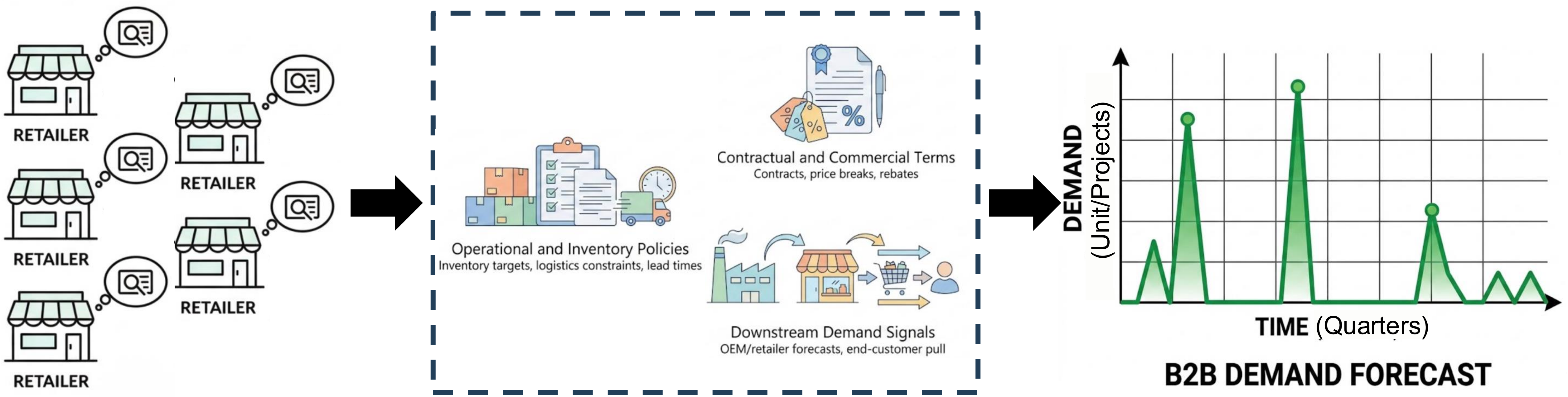
- **Data source**: Historical sales data, market trends, economic indicators, events
- **Traditional method**: statistical methods with domain expertise

Limitation of this approach

- **Data-wise**: Unable to incorporate multiple time series at a large scale
- **Process-wise**: Could be highly subjective and does not capture external factors well

Problem Identification – B2B

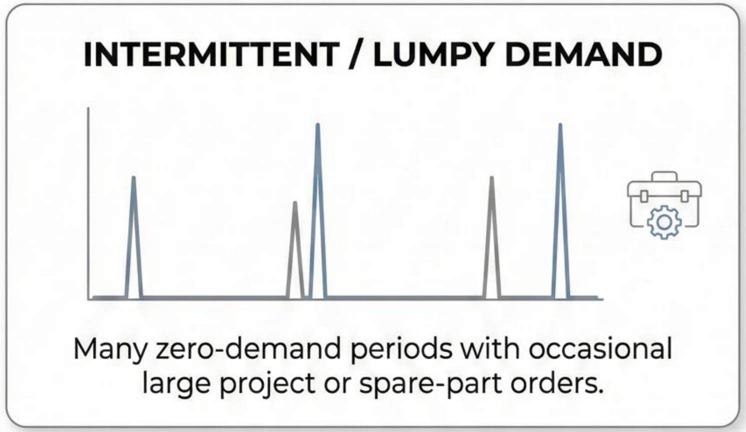
- In B2B setting, demand forecasting aims to predict how organizations place orders under contracts and supply agreements, across products and time.
  - It looks like a series of *commit/change* decisions from hundreds of accounts, each updating their order quantities.

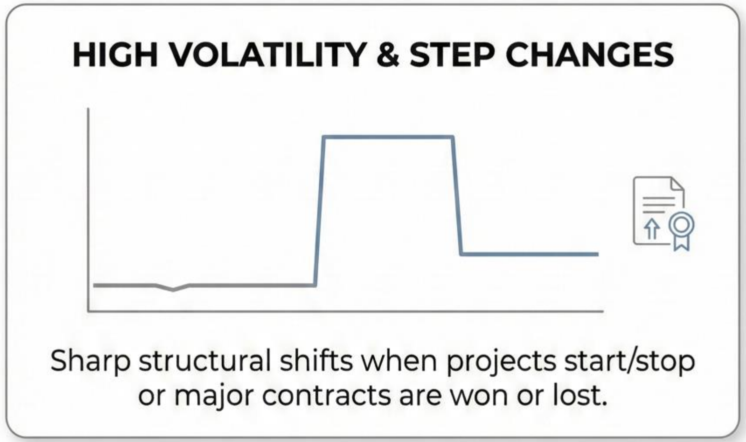


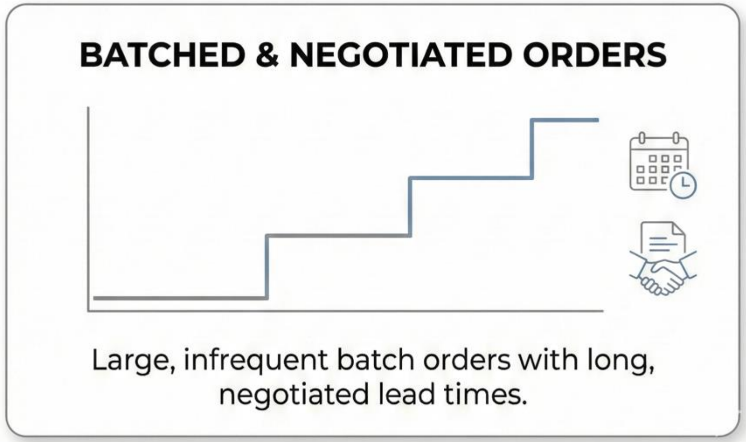
Problem Identification – B2B

- **Characteristics of demand**:
  - Intermittent/lumpy demand
  - high volatility & step changes
  - Batch & negotiated orders quantities

- Traditional methods:
  - **Intermittent-demand models (e.g., Croston's method)**: designed for slow-moving, lumpy items with many zero-demand periods
  - **Regression models**: Link demand to external drivers such as downstream forecasts, macro indicators, and pricing
  - **Manual adjustments**: planners apply manual overrides and adjustments during demand review meetings



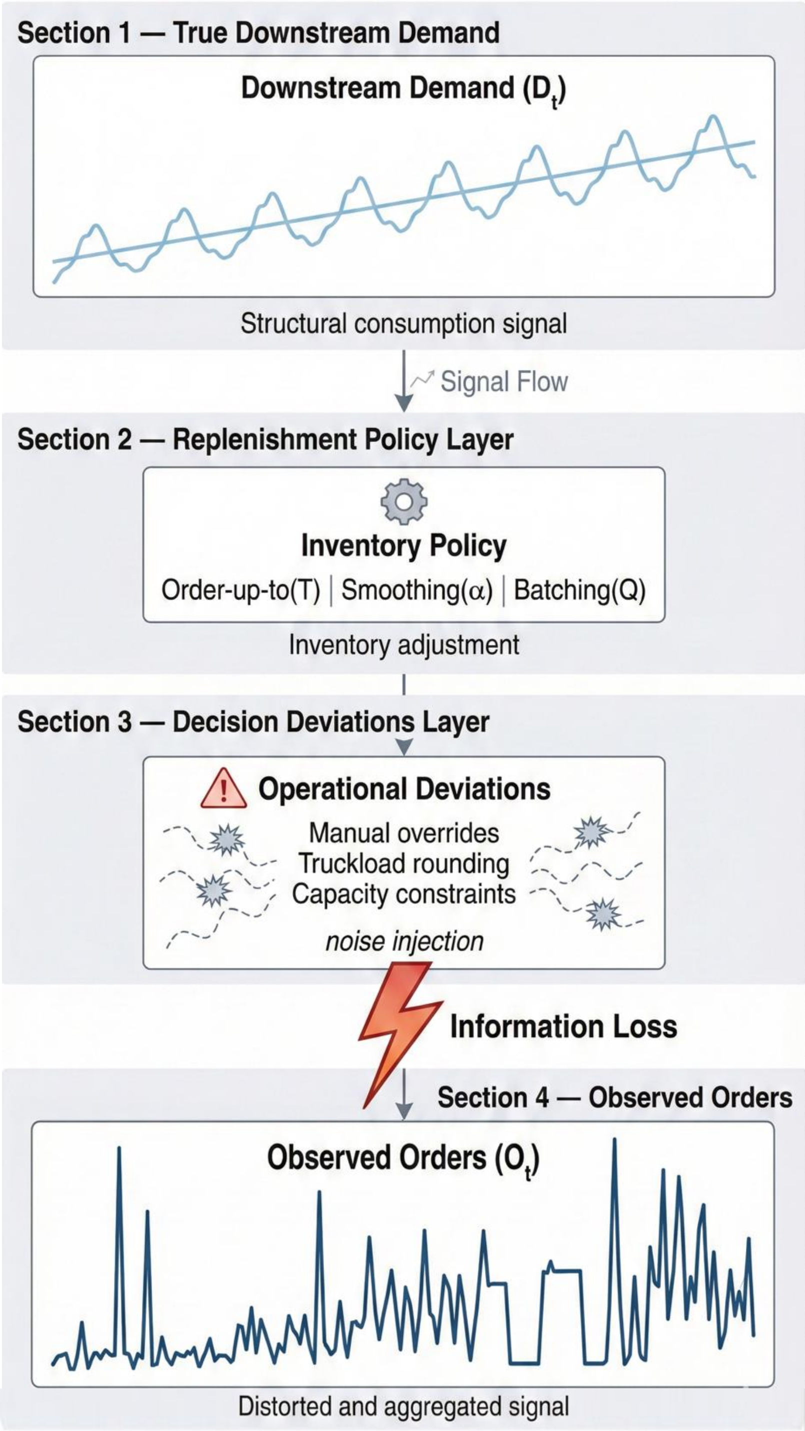




Problem Identification – B2B

- While the forecasting target is downstream orders, orders are generated from:
    - Underlying downstream demand
    - Inventory and replenishment policy
    - Operational and strategic deviations (manual overrides, batching, constraints)

- Traditional approaches assume orders are sufficiently informative, making the forecasting inaccurate.



B2B – Key Components

Stakeholder

- **Primary user**: Inventory and supply chain manager at upper stream suppliers
- **Goal**: Forecast downstream order partners for production and capacity planning

Scenario

- **Task**: Forecast orders from downstream partners
- **Process**: Use historical order data and predict future orders

Current approach

- **Data source**: Historical order data, inventory policies, lead times
- **Current method**: Intermitent demand models, regression models, manual adjustment

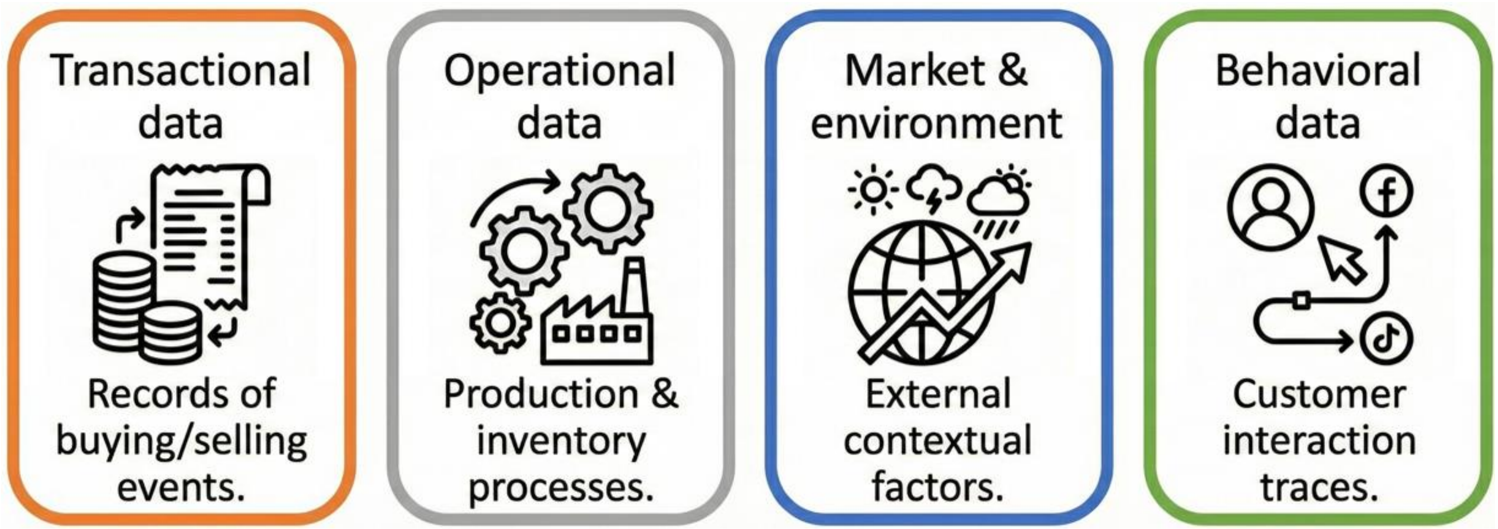
Limitation of this approach

- **Data-wise**: Orders are distorted signals with bullwhip effect and strategic ordering behavior
- **Process-wise**: Manual adjustment can be subjective

Step 2: Signal Collection

- **Signals** are the drivers for consumer & business demand that help predict the target series.

Signals (four major sources)



Target Series

Forecast target (e.g., sales, orders, demand volume)

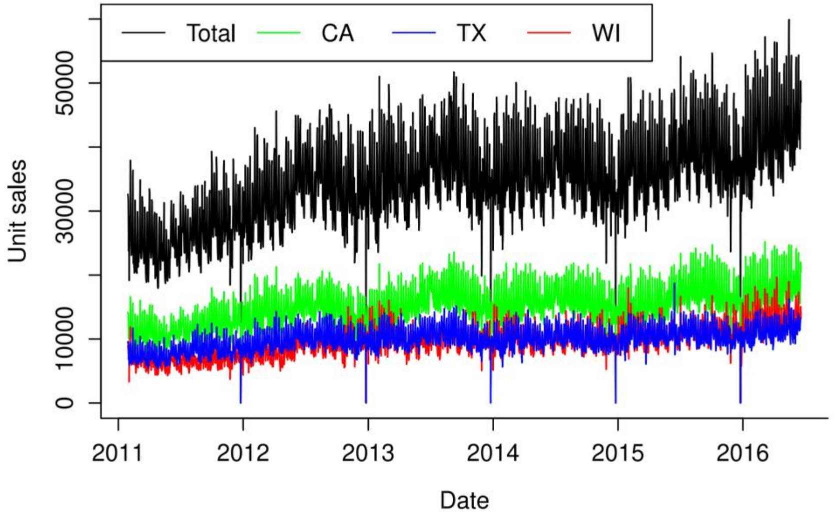
Signal Collection: Transactional data

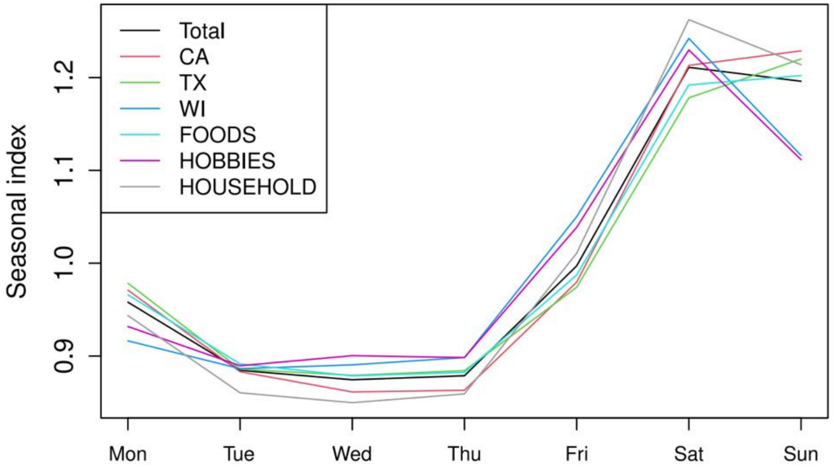
- Transactional data records actual buying and selling events, directly reflecting realized demand over time.

- **Signal provided**: Sales records, Orders, Returns

- **Common sources**:
  - Public retail datasets (e.g., M5 dataset, UCI Online Retail dataset),
  - Internal ERP / POS systems (e.g., SAP, Shopify, Square)

- **Collection methods**: Direct database query (SQL from ERP / POS), CSV export

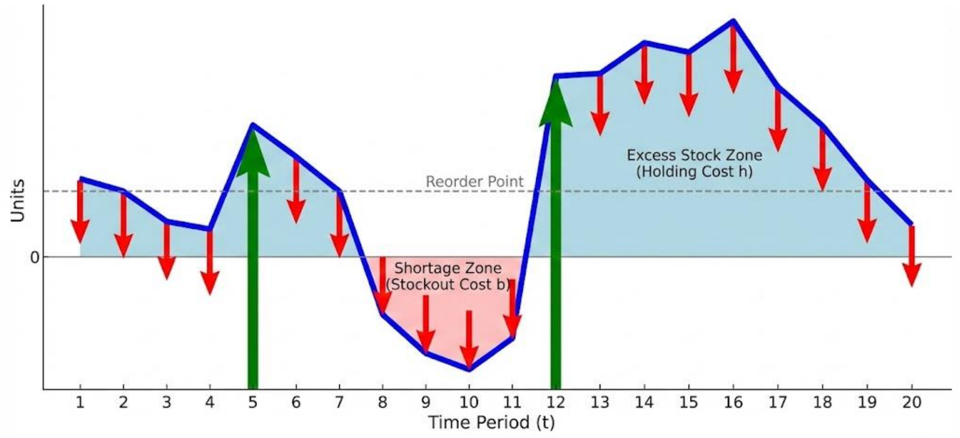


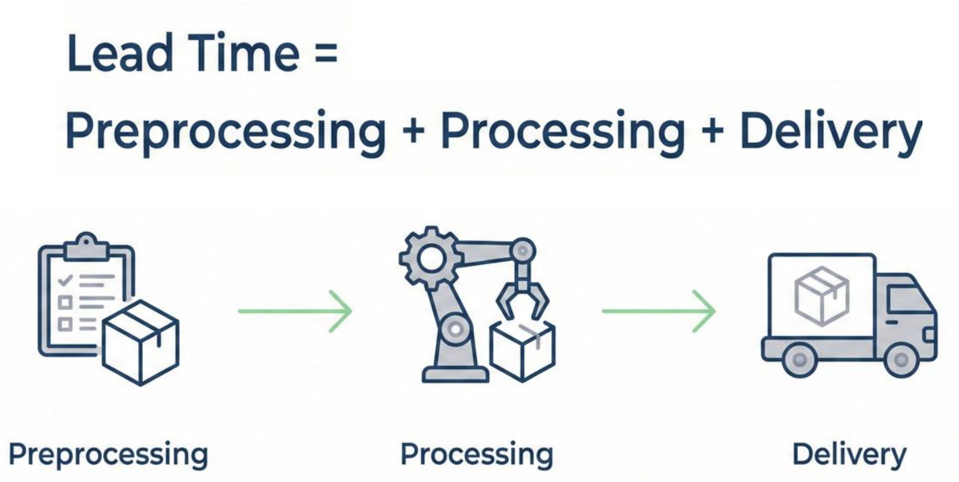


Reference: Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation. International Journal of Forecasting, 38(4), 1325-1336.

Signal Collection : Operational data

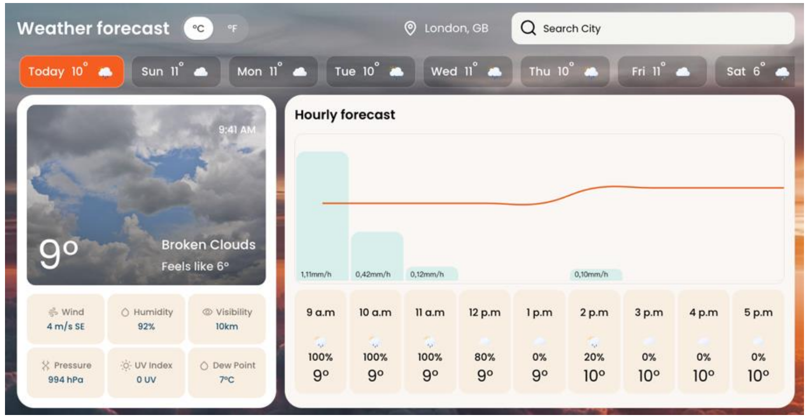
- Operational data records the real-time state of internal operations, such as inventory availability.
- **Signal provided**: Inventory levels, lead times
- **Common sources** (internal):
  - ERP systems (SAP, Oracle NetSuite),
  - Warehouse management systems (WMS)
- **Collection methods**: Internal SQL queries, ERP API endpoints

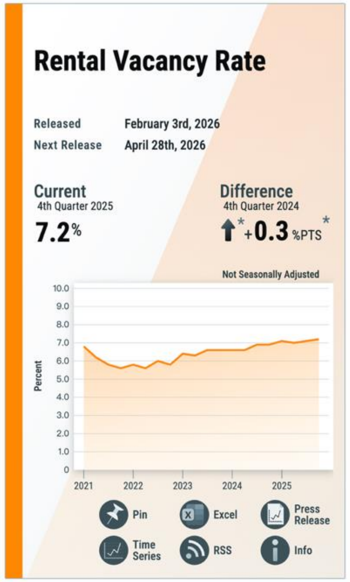




Signal Collection: Market & Environment Data

- Market and environment data represent external contextual factors that influence demand (usually out of a company’s control).
- **Signal provided**: Prices, promotions, weather, economic indicators
- **Common source**: NOAA, OpenWeather API, Census Bureau Economic, Yahoo Finance
- **Collection methods**: REST API, daily batch download





References:
- Openweather API: https://openweathermap.org/
- Census Bureau Economic Briefing Room - Index of Economic Activity: https://www.census.gov/economic-indicators/

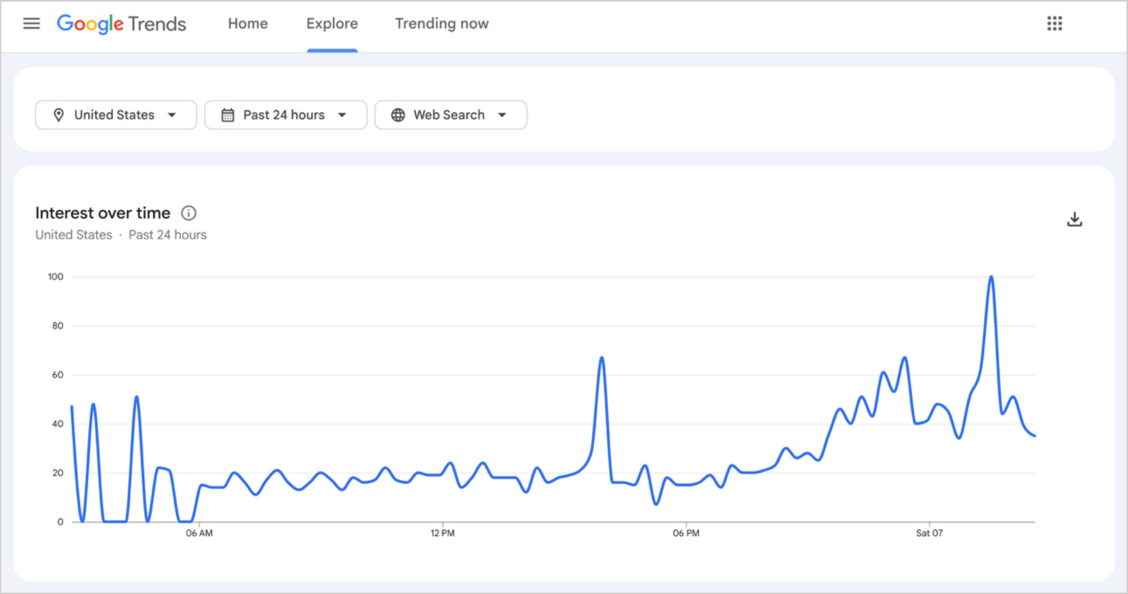
Signal Collection: Behavioral Data

- Behavioral data captures customer interaction signals that reveal attention or comments for products

- **Signal provided**: customer behavior logs, search trends, social media signals

- **Common source**: Google Trends, Amazon search rank, Reddit API

- **Collection methods**: API calls, Keyword-based scraping

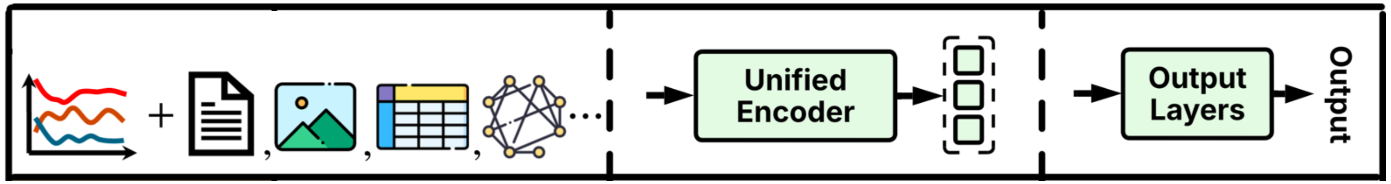


Reference: Google Trends: https://trends.google.com/home

Signal Collection: Beyond Numeric Data

- Modern signals often span multiple modalities:
  - Text (news, reviews)
  - Images (shelf photos, product images)
  - Graphs (product relationships)
  - Video/audio data (usage recordings, operational videos)

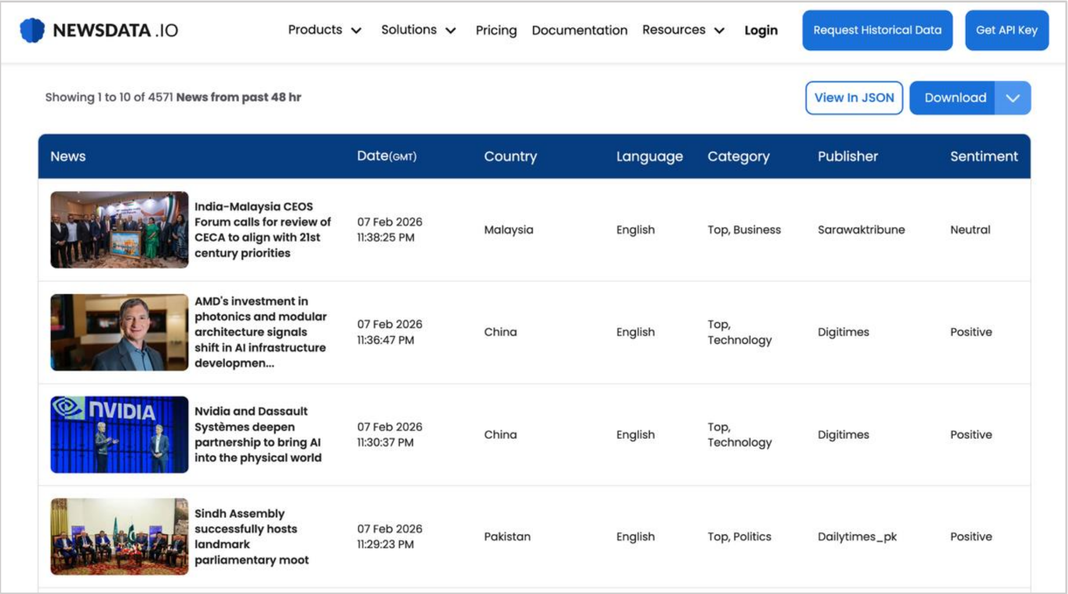
- Multimodal signals help capture cross-channel and cross-entity effects
  - E.g., social media discussions create trends that increase future demand.



Reference: Jiang, Y., Ning, K., Pan, Z., Shen, X., Ni, J., Yu, W., ... & Song, D. (2025, August). Multi-modal time series analysis: A tutorial and survey. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2 (pp. 6043-6053).

Signal Collection: Beyond Numeric Data

- **Signal provided**: Daily news, social media discussion, product images, product relations
- **Common sources**: NewsData.io, X, Shopify
- **Collection methods**: API calls, web scraping

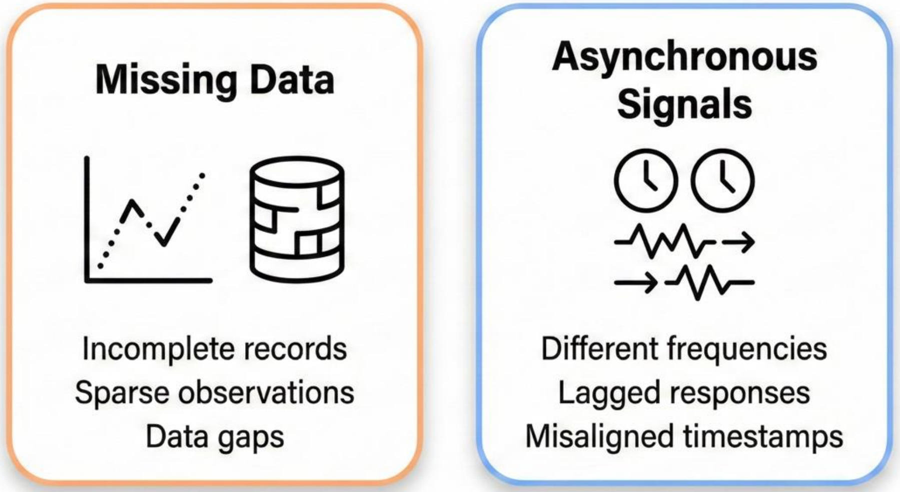


Reference: NewsData.io: https://newsdata.io/

Challenges

- Demand forecasting often requires analyzing multiple signals together.
  - However, real-world signals are imperfect and misaligned

- Leveraging multiple sources of signals often introduces issues such as missing data and asynchronous signals.

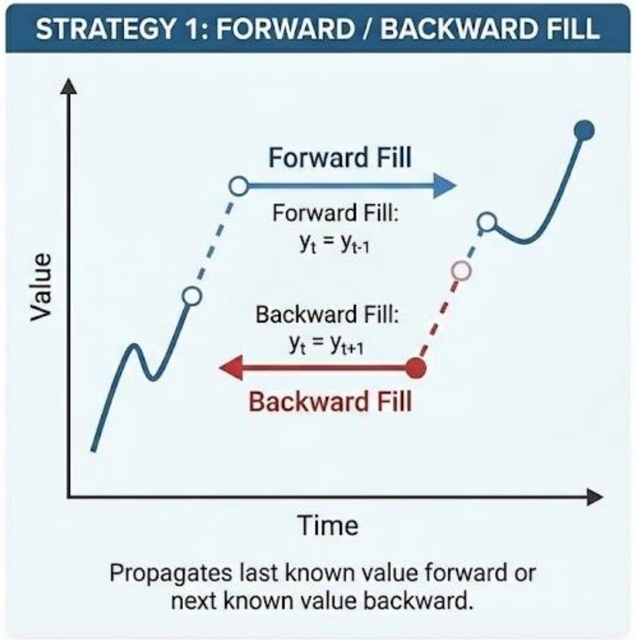


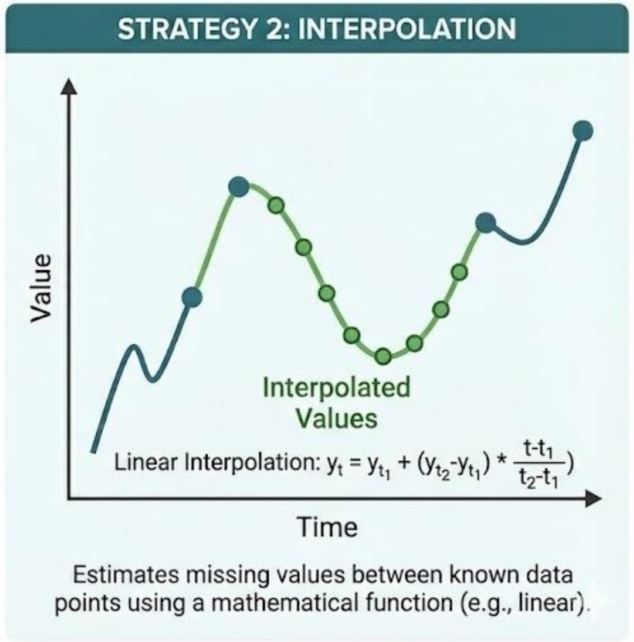
Challenges: Missing Data

- Missing data can happen when there are issues like:
  - Stockouts vs true zero demand
  - sensor failures
  - API downtime

- Unlike standard ML tasks, we cannot simply drop missing observations in time series
  - Removing timestamps breaks temporal continuity

- Common handling strategies
  - Forward / backward fill (short gaps only)
  - Interpolation (linear, spline)

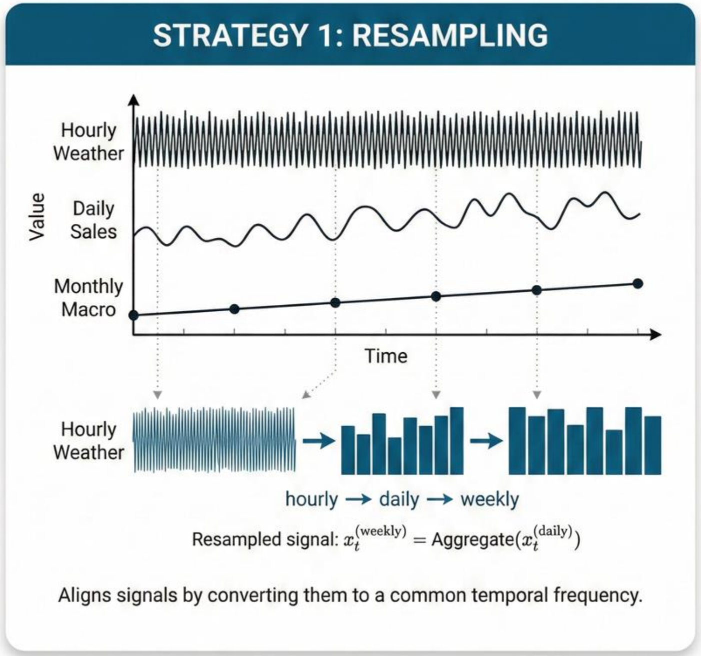




Challenges: Asynchronous Signals

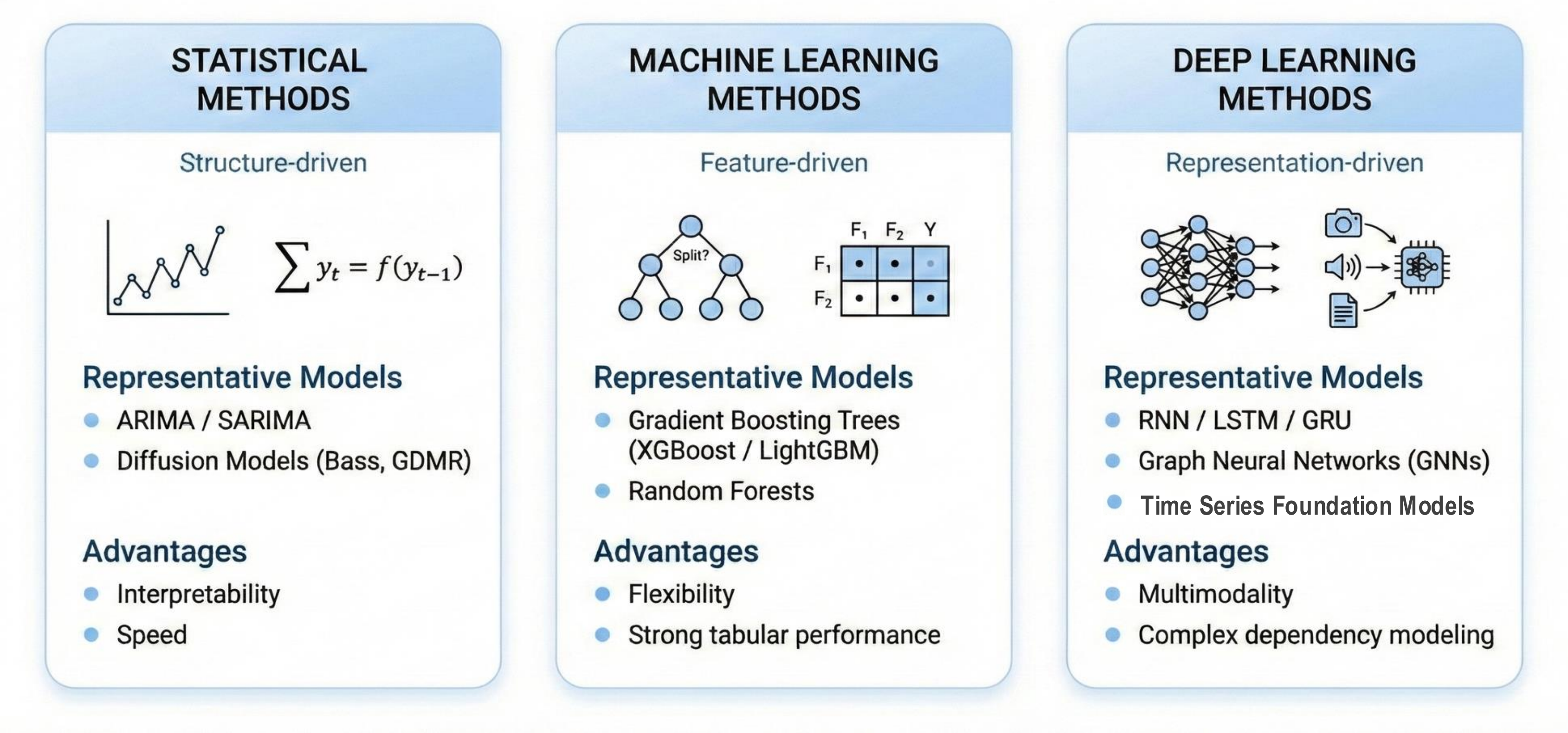
- Asynchronous signals happen when demand signals operate at different frequencies, e.g.,
  - Daily sales
  - Hourly weather
  - Monthly macroeconomic indicators

- Common synchronization strategies
  - Resampling (hourly → daily → weekly)



Step 3: AI-Enabled Analytics

Recent demand forecasting primarily advances along three methodological directions.



Statistical Methods

- **Idea**: Use explicit mathematical formulas to capture regular patterns in historical data

- However, out-of-the-box traditional models (e.g., ARIMA, Exponential smoothing) are often univariate.

- Therefore, we **inject domain knowledge** into the regression model by explicitly defining mathematical structures that mirror real-world behavior.

Statistical Methods

- Common structures to inject:
  - **Lag structure**: captures how the immediate past influences the present
    $$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}$$
  - **Seasonality**: captures recurring patterns using dummy variables
    $$Y_t = \Sigma_s \gamma_s D_{st}$$
  - **Interaction**: captures how seasonality changes depending on other conditions
    $$Y_t = \delta(Y_{weather} \times Y_{economic})$$
- These methods provides great explanability on the forecasting result and are suitable when the forecasting horizon is short.
- However, model specification is not generalizable (has to be design per case).

Statistical Methods (Example)

- **Goal**: Forecast high-frequency demand for on-demand service platforms (e.g., DoorDash, Uber) under unstable, changing environments
- **Signals**: Hourly demand time series from a large on-demand delivery platform
- **Key data characteristics**: High-frequency streaming updates with strong seasonality (hour-of-day, day-of-week)
- **Design**: A parametric regression model (trend + seasonality + autoregressive lags)
  - Provides a fast closed-form estimation for real-time updates with low computing cost compared to LSTM

Figure 1. (Color online) Performance of the New Method

Reference: Hu, Y. J., Rombouts, J., & Wilms, I. (2025). Fast forecasting of unstable data streams for on-demand service platforms. Information Systems Research, 36(1), 552-571.

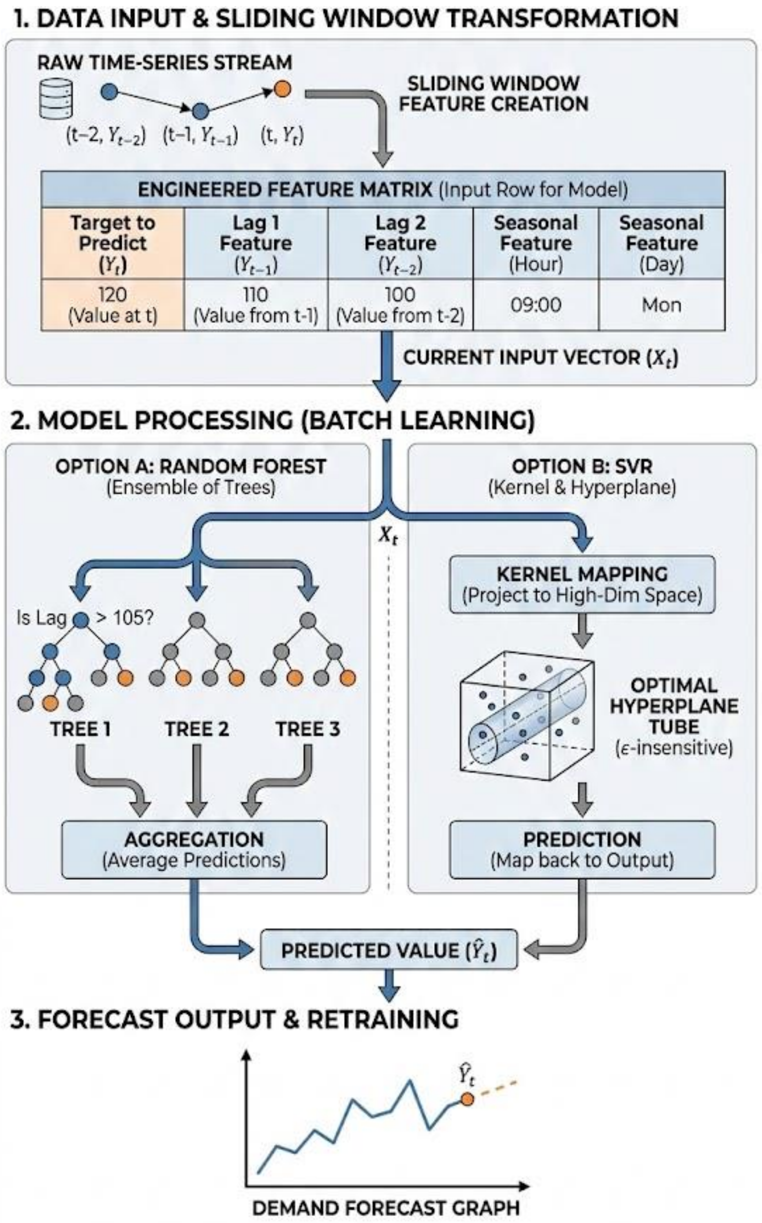
Machine Learning

- ML methods can often generalize to many cases by capturing the non-linear relationship automatically.

- Representative models:
  - Tree-based models (Random Forest, XGBoost)
  - Support Vector Regression (SVR)

- These models are good at capturing:
  - Multivariate, multiple explanatory variables
  - Nonlinear effects and interactions

- **Limitation**: they do not handle sequential context directly, which cannot capture long temporal dependencies (e.g., sales drop due to bullship effect can span months).



Online M.S. in AI

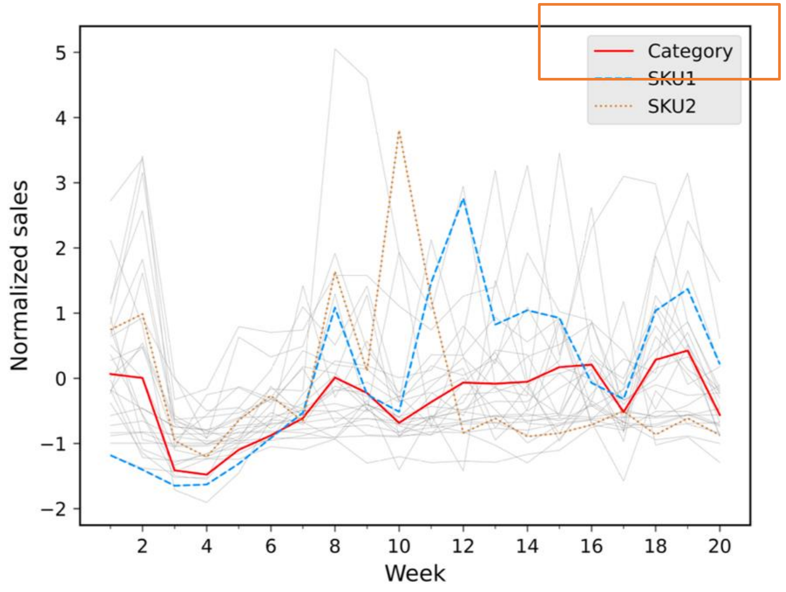
Machine Learning (Example)

- **Goal:** Predict individual product demand more accurately by leveraging aggregate category sales information

- **Signals:** Weekly retail sales data with hierarchical product categories from Walmart.

- **Key data characteristics:**
  - Individual product sales are volatile (high variance, noisy)
  - category sales are smoother (variance reduced via pooling),
  - But they share similar temporal patterns

- **Design:** A transfer learning framework implemented with gradient boosting trees (Pooling and Boosting, PAB)
  - Trains XGBoost on pooled category sales and then refines it with individual product-level data



Reference: Lei, D., Qi, Y., Liu, S., Geng, D., Zhang, J., Hu, H., & Shen, Z. J. M. (2025). Pooling and boosting for demand prediction in retail: A transfer learning approach. Manufacturing & Service Operations Management, 27(6), 1779-1794.

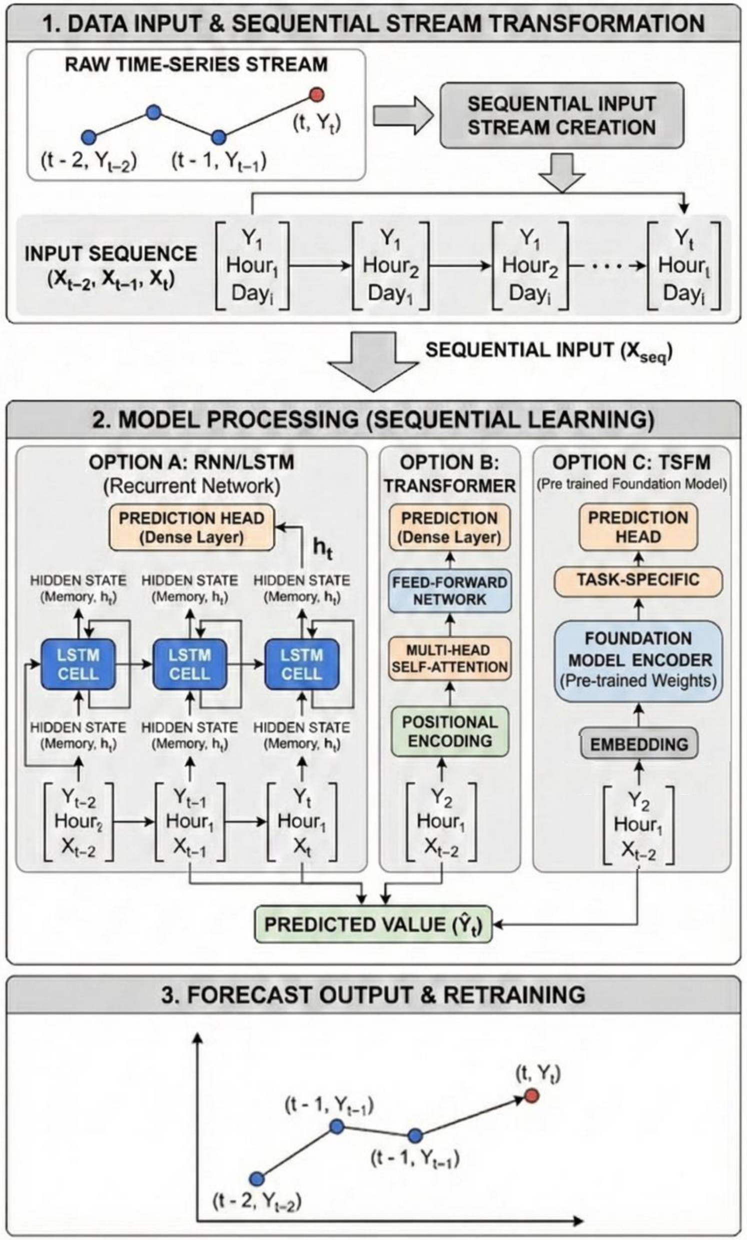
Deep Learning

- Deep learning methods can capture long temporal dependencies by modeling the signals sequentially.

- Representative models
  - Recurrent Neural Networks (RNN, LSTM, GRU)
  - Transformer-based Models (BERT, GPT, Time-Series Transformers)
  - Time Series Foundation Models (TSFMs)

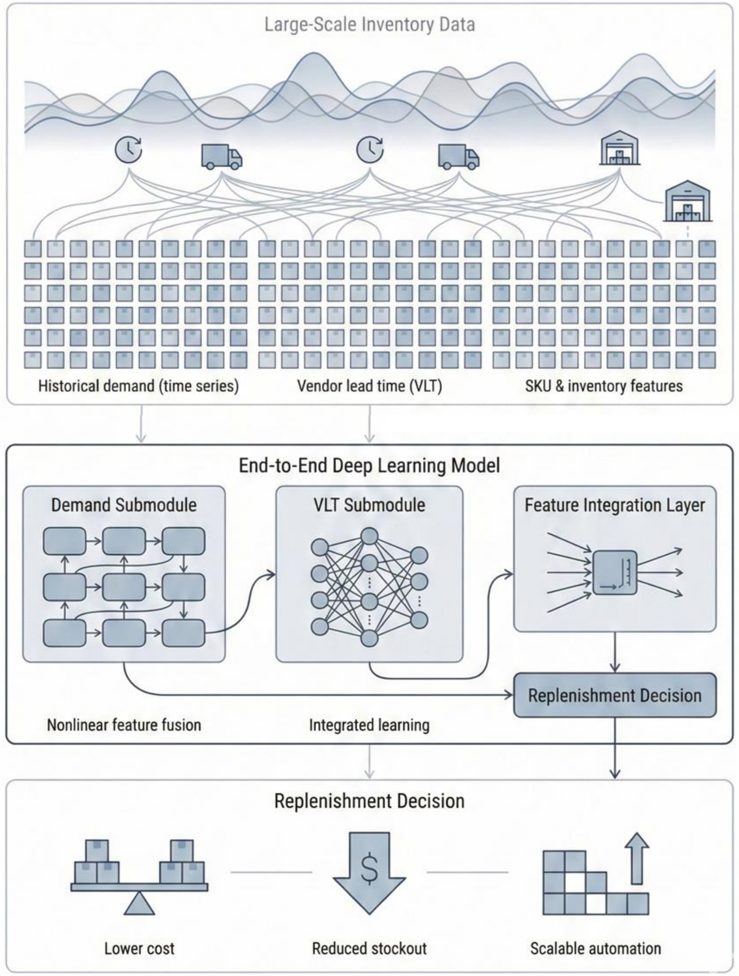
- They are well-suited to capture:
  - **Complex dependencies**: Long-term temporal patterns or non-linear interactions
  - **Large-scale datasets**: Situations where data volume exceeds the capacity of traditional ML
  - **Multimodal signals**: Integrating diverse data types

- **Limitation**: Requires massive amounts of data to generalize well, computationally expensive to train



Deep Learning – RNN (Example)

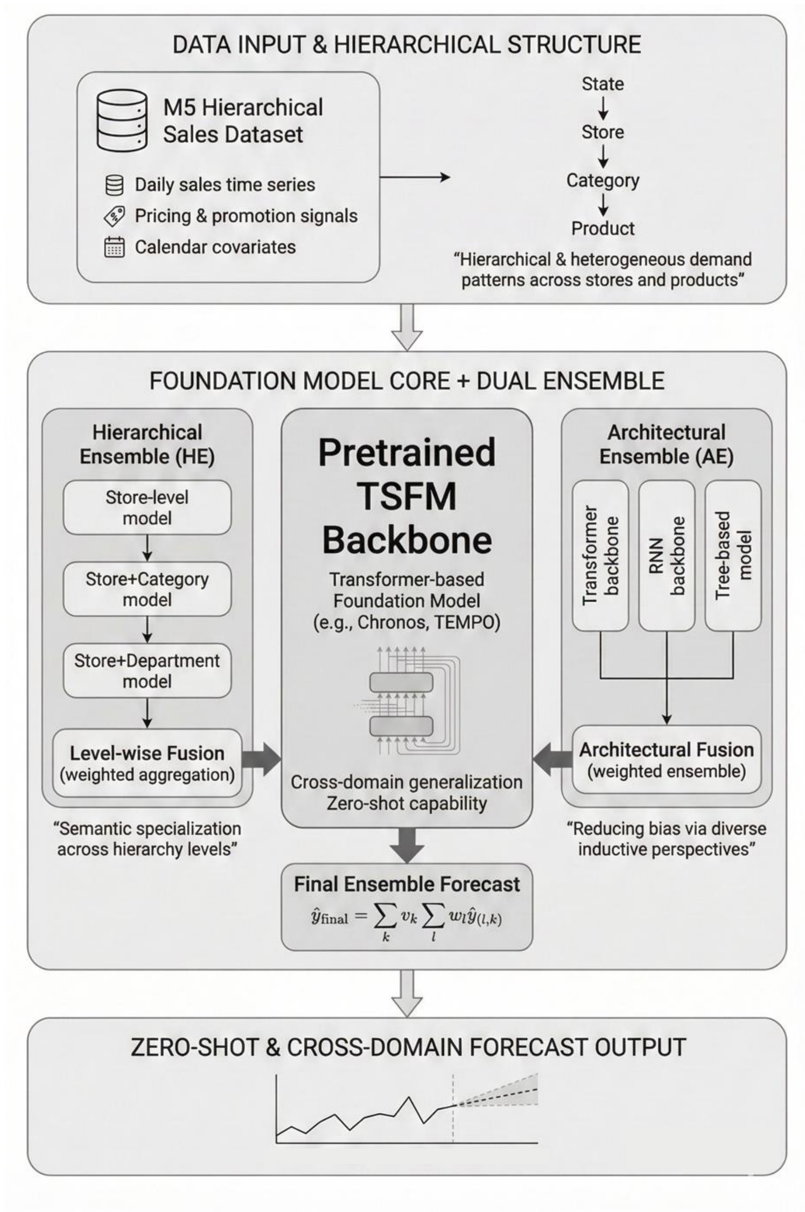
- **Goal**: Predict order quantity to minimize total inventory cost (holding + stockout)
- **Signals**: Daily sales, vendor lead times, inventory levels, and product features (category, warehouse, brand) from JD.com
- **Key data characteristics**: Complex nonlinear interaction between signals in large-scale data (thousands of products)
- **Design**: End-to-End deep learning framework
  - Combining RNN module for demand sequence modeling.



Reference: Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., & Shen, Z. J. (2023). A practical end-to-end inventory management model with deep learning. Management Science, 69(2), 759-773.

Deep Learning – TSFM (Example)

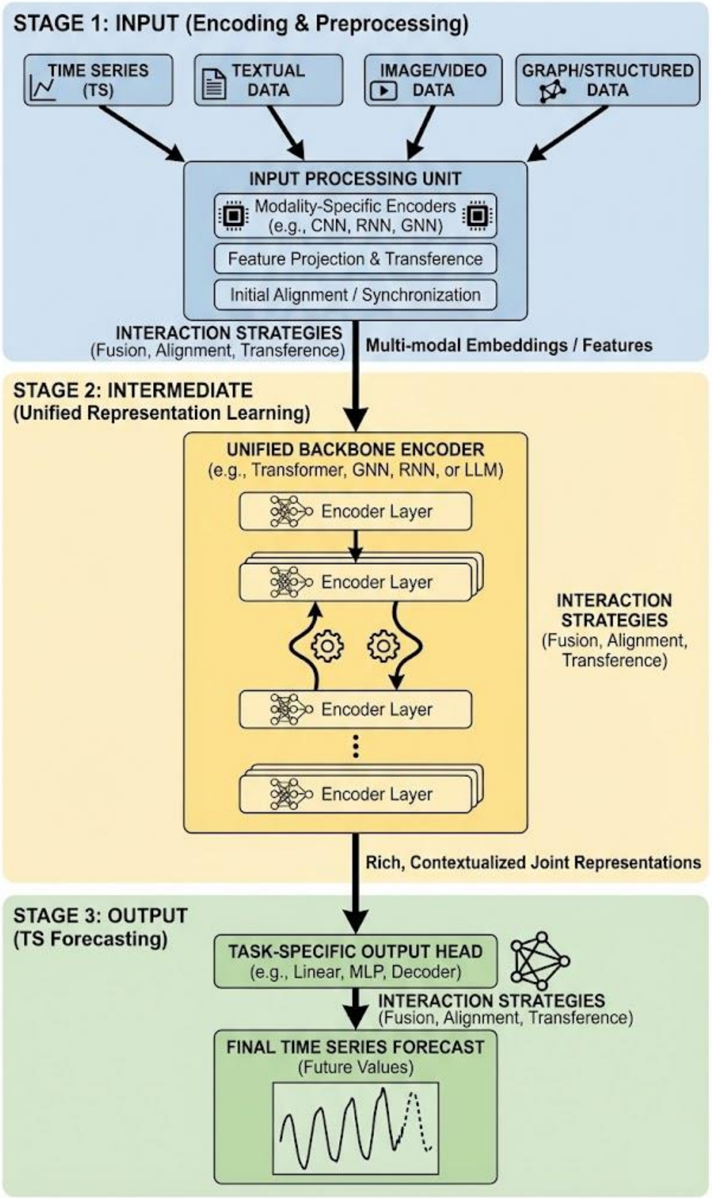
- **Goal**: Leverage a pretrained TSFM to achieve cross-domain and zero-shot demand forecasting in supply chains
- **Signals**: M5 dataset (hierarchical sales time series including store, category, and product, with pricing, promotion, and calendar covariates)
- **Key data characteristics**: Hierarchical and heterogeneous demand patterns across products and stores
- **Design**: A pretrained TSFM applied to retail forecasting, with forecasts fused across hierarchy levels and multiple model backbones



Reference: J Yang, W., Cao, D., & Liu, Y. (2025). Foundation Models for Demand Forecasting via Dual-Strategy Ensembling. KDD 2025 Workshop on AI for Supply Chain: Today and Future. Retrieved from https://openreview.net/forum?id=hcwHRlwcFA

Deep Learning – Multi-modal

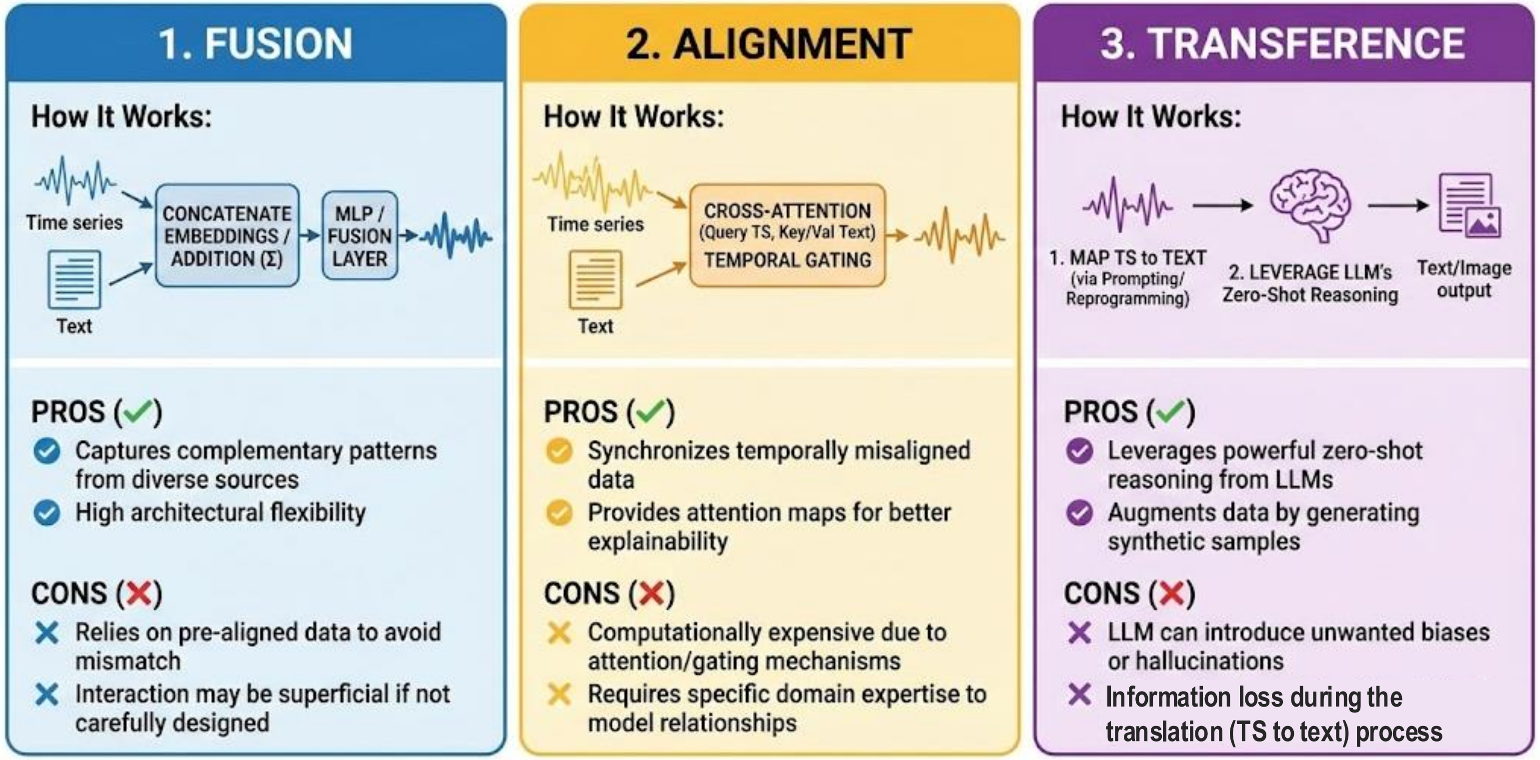
- Deep learning methods are well-suited to handle multiple modalities.
- It is achieved by adding a multi-modality interaction to a backbone architecture, such as:
  - Transformers, large language models (for textual data)
  - graph neural networks (for graph-based data)
- Interaction strategies are added to the input, intermediate, and output stages to capture multi-modal signals.



Reference: Jiang, Y., Ning, K., Pan, Z., Shen, X., Ni, J., Yu, W., ... & Song, D. (2025, August). Multi-modal time series analysis: A tutorial and survey. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2 (pp. 6043-6053).

Deep Learning – Multi-modal

- Three types of interaction strategies exist.



Reference: Jiang, Y., Ning, K., Pan, Z., Shen, X., Ni, J., Yu, W., ... & Song, D. (2025, August). Multi-modal time series analysis: A tutorial and survey. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2 (pp. 6043-6053).

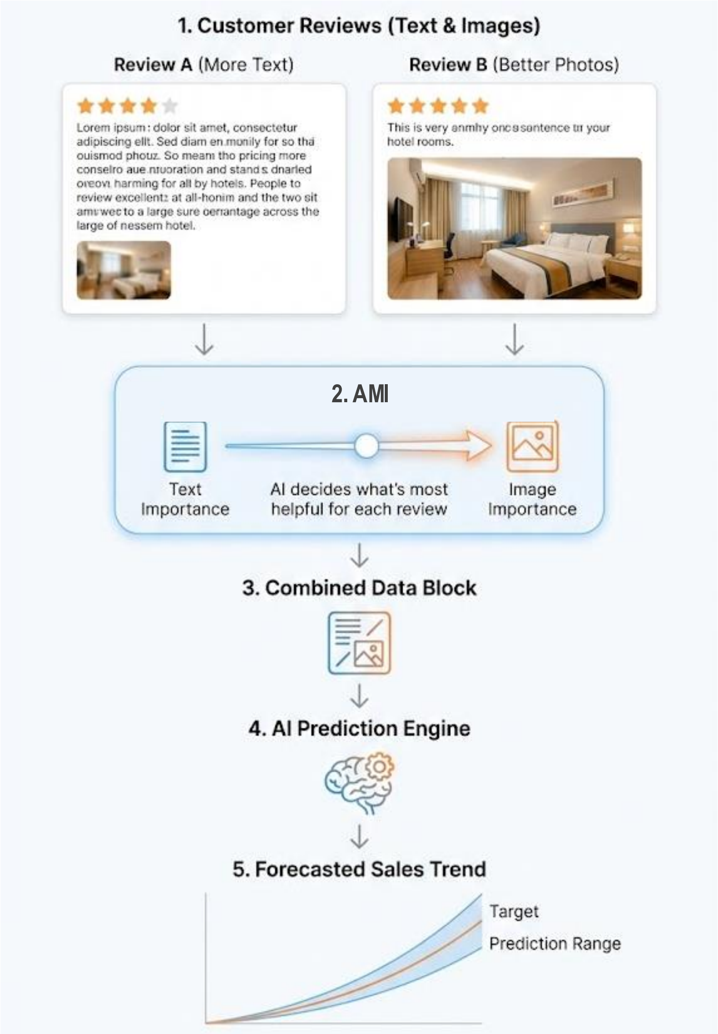
Deep Learning – Multi-modal (Example)

- **Goal**: predict future sales using multimodal online reviews

- **Signals**: Monthly hotel sales (numerical), customer review (image, textual)

- **Key data characteristics**: Sometimes a picture is more helpful than text; other times, the text is more descriptive.

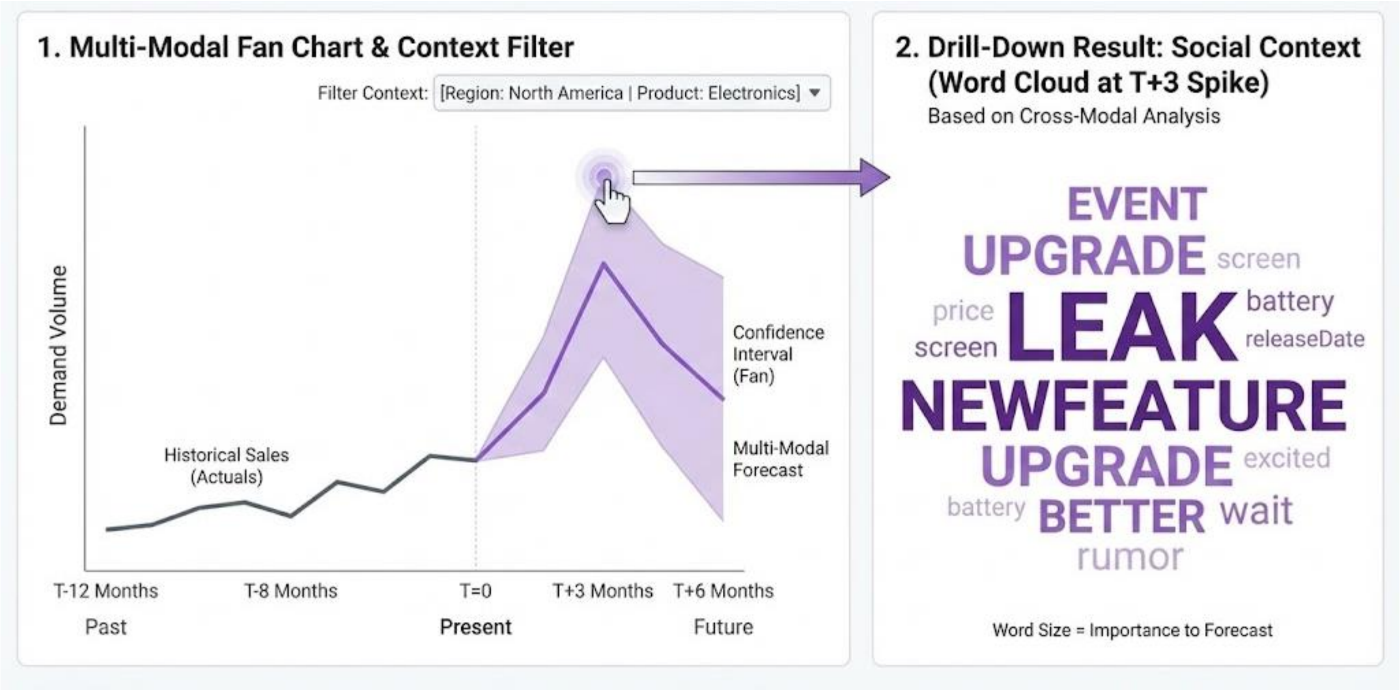
- **Design**: adaptive attention mechanism for multimodal interaction (AMI)
  - Use attention to automatically identify which modality is more important



Reference: Chen, G., Huang, L., Xiao, S., Zhang, C., & Zhao, H. (2024). Attending to customer attention: A novel deep learning method for leveraging multimodal online reviews to enhance sales prediction. *Information Systems Research*, *35*(2), 829-849.

Step 4: Value Delivery – Visualization

- Demand forecasting is often done per product, per region.
- A **fan chart** with regional and product filters can help visualize the forecasted result.
  - If textual data are incorporated into the forecast, we can provide explainability with the context words in a drill-down display (Jiang et al. 2025).



Online M.S. in AI

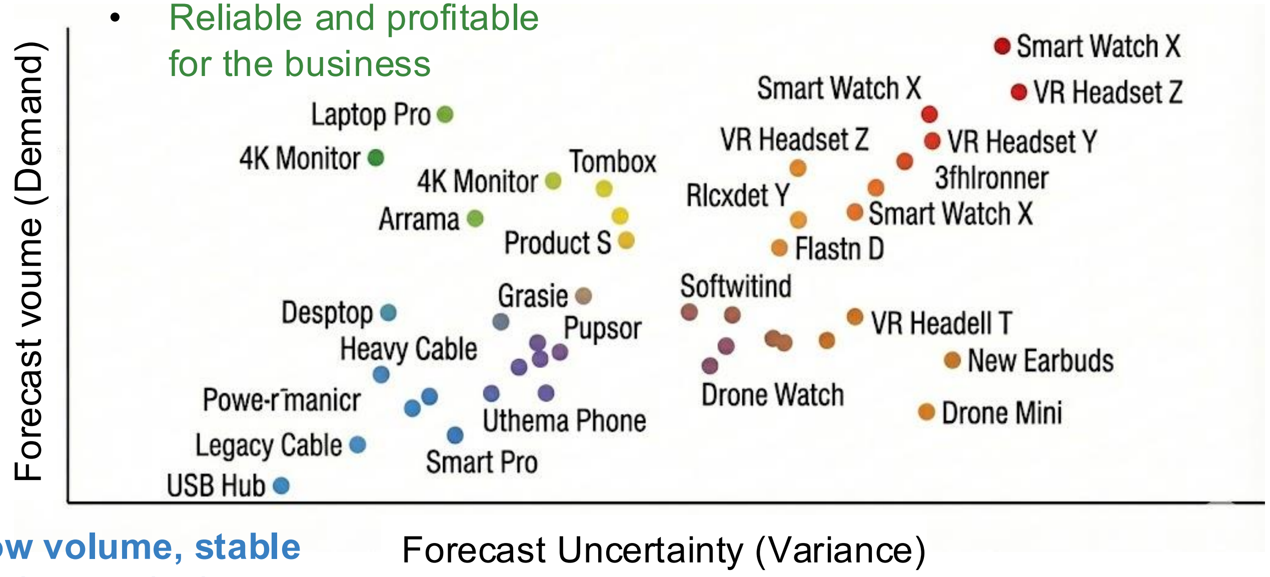
Reference: Jiang, Y., Yu, W., Lee, G., Song, D., Shin, K., Cheng, W. & Chen, H. (2025). Explainable multi-modal time series prediction with llm-in-the-loop. *arXiv preprint arXiv:2503.01013*.

Value Delivery – Visualization

- A scatter plot of each product by its forecasted demand and variance can be created based on the forecasting result at a specific time.

High volume, stable
- Reliable and profitable for the business

High volume, high variance
- Critical risks
- require the most attention because a forecast error here results in massive lost



Low volume, high variance
- Often the candidates for discontinuation
- The cost of managing their unpredictability can outweighs their profit.

Low volume, stable
- Low-stake item
- Require minimal monitoring

Value Delivery – Replenishment Decision

- After understanding the forecasted demand for each product and region, this can inform inventory replenshiment decision.

- **Traditional Method**: Predict-then-Optimize:
  1. **Prediction**: use demand forecasting model to forecast future demand
  2. **Optimization**: These forecasts are plugged into a pre-defined rule
     - E.g., You only place an order when the inventory level falls below a reorder point

- **End-to-end model**: build another deep learning model to jointly perform demand forecasting and make replenishment decisions (Qi et al. 2023).

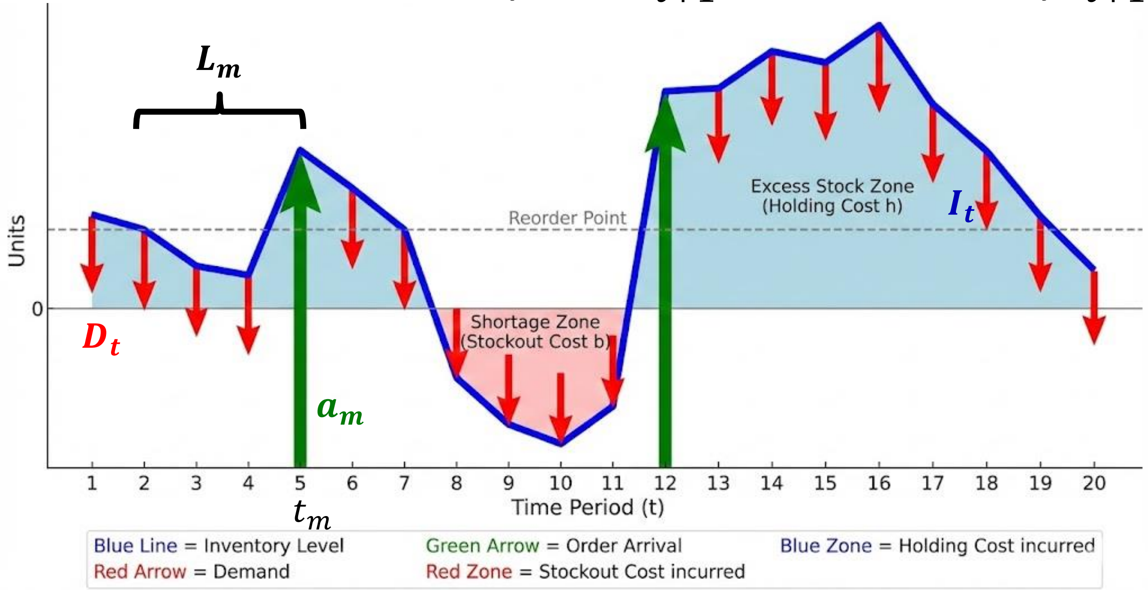
Reference: Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., & Shen, Z. J. (2023). A practical end-to-end inventory management model with deep learning. Management Science, 69(2), 759-773.

Value Delivery – Replenishment Decision

- The quality of demand forecasting and replenishment strategies can be evaluated by summing up the cost at each period t (Qi et al. 2023):

$$S_t = h[I_{t+1}]^+ + b[-I_{t+1}]^+$$

  - Where the inventory level I_{t+1} is determined by $I_{t+1} = I_t - D_t + \sum_{m=1}^{M} a_m \, \mathbb{I}\{t = t_m + L_m\}$



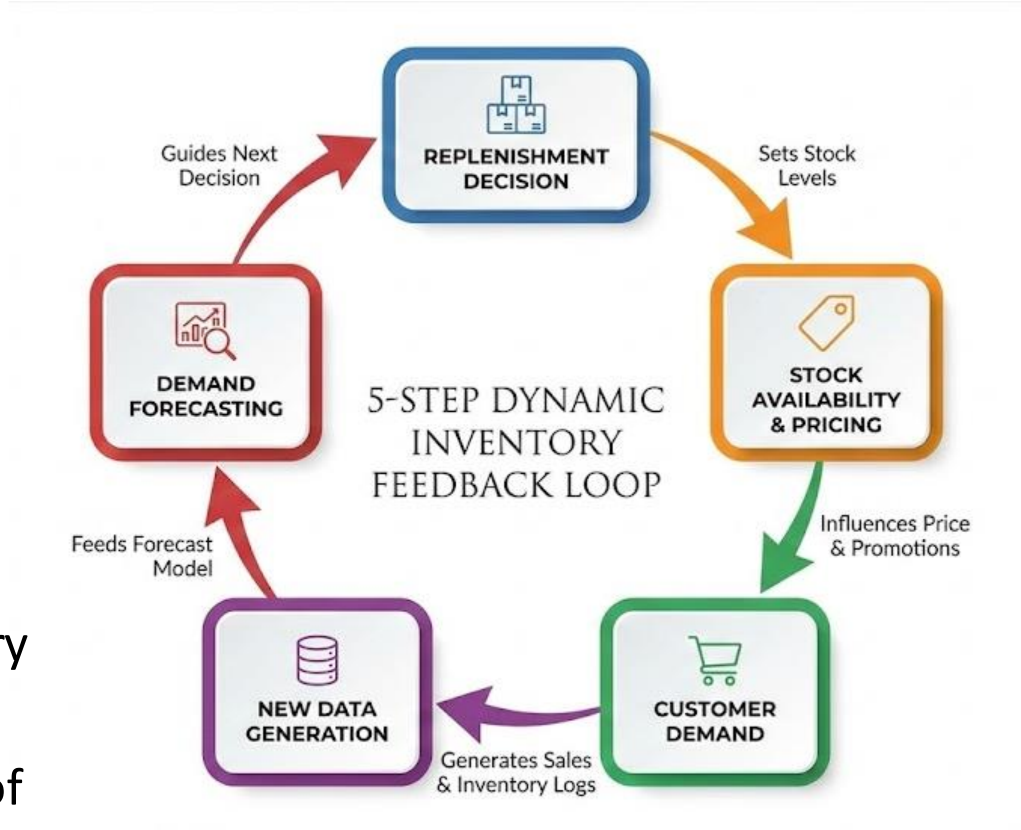
I_t : inventory level at time t
D_t : Demand at time t (by demand forecasting)
a_m: The m^{th} replenishment order quantity
L_m: The lead time for the m^{th} replenishment order

$[x]^+$: Notation for $\max(x, 0)$.
$\mathbb{I}\{\cdot\}$:1 if true, 0 otherwise

Reference: Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., & Shen, Z. J. (2023). A practical end-to-end inventory management model with deep learning. Management Science, 69(2), 759-773.

The Feedback Loop

- The decision after demand forecasting naturally forms a loop:
  - **Replenishment Decision**: Changes your current stock availability.
  - **Stock Availability**: Informs pricing strategies (e.g., clearance for overstock).
  - **Customer Demand**: Driven by those prices and external market factors.
  - **Data Generation**: The resulting sales and inventory levels create new logs.
  - **Forecasting**: This new data feeds the next round of demand forecasting to start the cycle over.



Review Questions

- What are the common time-series patterns typically observed in demand forecasting?
- How does B2C differ from B2B regarding the underlying mechanisms that generate demand?
- What types of multi-modal data are commonly incorporated into demand forecasting, and how do deep learning methods capture them?
- How to measure the stockout and holding cost after demand forecasting, given a replenishment decision?