Predicting the Wave: Measuring Social Media Virality to Aid Demand Planning

by

Lucas Achman

and

Alma Zubaid

Submitted on May 8th, 2026 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management at the
Massachusetts Institute of Technology

Abstract

Forecasting demand has become increasingly difficult as TikTok and other social media platforms can create rapid demand spikes that are not captured by traditional forecasting methods. This project addresses that limitation by developing a predictive model for CosmeticCo's North American Consumer Products Division that integrates TikTok engagement signals into demand planning. The model uses a two-stage approach: logistic regression estimates the probability of entering a high-demand regime within a seven-day horizon, while quantile regression estimates the expected uplift if that event occurs.

In the proposed model, high demand was defined as actual seven-day demand reaching at least 1.6 times the baseline forecast, with alerts triggered at a 30% probability threshold and uplift estimated at the 75th percentile. During a 115-day test period, the model achieved 95% overall accuracy and correctly identified 4 of 10 actual high-demand events, resulting in 40% recall and 100% precision for the high-demand class. Although the model did not capture every spike, every high-demand alert it generated was correct, making it useful as a focused early warning signal.

Without the model, the baseline forecast failed to anticipate significant incremental demand during high-demand periods, representing a substantial volume of potential lost sales. By incorporating TikTok engagement signals, our model recovered visibility into 72.3% of that uplift, directly reducing potential lost sales exposure by approximately three-quarters of the undetected volume, which planners could now act on. These results demonstrate that incorporating TikTok engagement signals can help anticipate social-media-driven demand surges and improve visibility into their magnitude, enabling planners to identify high-risk periods earlier than the baseline forecast alone. In practice, this framework can support more informed decisions on inventory positioning, supplier readiness, production planning, and escalation of response actions when social media activity suggests a potential surge in demand.

Capstone Advisor: Dr. Jaime E Macias Aguayo

Title: Postdoctoral Associate

Capstone Co-Advisor: Dr. Maria Jesus Saenz

Title: Executive Director, MIT SCM Master's Program

Title: Director, MIT Digital Supply Chain Transformation Lab

Table of Contents

1. Introduction

CosmeticCo is a global leader in the consumer goods industry, known for its strong influence on the beauty market through a diverse portfolio of innovative products. Its offerings span key categories such as hair care, color cosmetics, fragrance, and sun protection, catering to a broad and dynamic customer base. With operations across numerous international markets, the company distributes its products through a wide range of channels, including brick-and-mortar retail, e-commerce, mobile commerce, click-and-collect, and direct-to-consumer platforms.

This extensive omnichannel presence, combined with a large and frequently evolving product assortment, adds significant complexity to demand planning and supply chain management. In North America alone, CosmeticCo manages tens of thousands of stock keeping units (SKUs), making accurate demand forecasting a critical enabler of supply chain efficiency. By maintaining high service levels while minimizing stockouts and excess inventory, the company aims to reduce operational costs, improve inventory positioning, and support overall profitability.

1.1 Motivation

Forecasting is a challenging task, and it becomes even more complex in the beauty industry, an industry historically shaped by traditional media. For decades, apart from promotional campaigns, advertisements have been used to generate and drive demand. The media landscape, however, has shifted significantly toward social media, transforming how demand generation works.

Demand forecasting at CosmeticCo is complicated by the unpredictable dynamics of social media virality, which can trigger rapid surges in awareness and demand, even for existing products. In 2020, a single TikTok post generated significant reach within two weeks, ultimately leading to a product sellout within three months, as shared during a company interview. As a result of such viral moments, CosmeticCo has experienced monthly sales surges of over 300% above forecasted levels for SKUs that gained traction on TikTok.

To boost awareness and product appeal, CosmeticCo collaborates with social media influencers, who are granted creative freedom to leverage greater trends within social media. A trend in social media refers to a content format that is widely adopted and which rapidly spreads across the platform. Users participate by creating their own versions of the content, each adding a unique personal twist. Effective influencer posts on TikTok generate curiosity, drive audience engagement (likes, comments, shares), and expand posts' reach. As engagement grows, TikTok's algorithm amplifies exposure, fueling social media virality. An inability to quantify key parameters that amplify and drive viral events introduces substantial uncertainty in demand forecasting. If demand spikes are not anticipated, stockouts and lost sales follow.

1.2 Problem Statement and Key Questions

Demand is inherently uncertain, and the inability to identify its key drivers or predict their relationship with social media virality directly impacts sales performance. Cosmetic Co. currently lacks a tool that captures the effect of social media driven demand spikes in its forecasting process. This capstone project addresses this gap by analyzing viral events over a two-year horizon to identify key metrics that drive virality and demand variability, enabling the development of early signals for potential demand surges.

The key questions addressed in this project are:

- Can TikTok engagement metrics improve the identification of high-demand periods compared with relying on the baseline forecast alone?
- Once a high-demand event is identified, how accurately can the model estimate the magnitude of demand uplift above the baseline forecast?

- How should model outputs, including high-demand probability, expected uplift, and aggregate capture rate, be presented to clearly communicate results to internal stakeholders?

The project focuses on existing products within CosmeticCo's North American Consumer Products Division (CPD), examining TikTok driven virality from paid influencer partnerships, specifically targeting products that have historically demonstrated significant demand surges and structural shifts in baseline demand following viral events. The methodology follows three main steps. First, historical viral events are analyzed using sales, forecasts, and TikTok engagement data to identify key drivers of demand variability. Second, statistical and machine learning (ML) techniques are applied to uncover patterns and build a predictive model. Finally, the model is tested on new scenarios to evaluate its accuracy and usefulness for demand planning.

1.3 Project Goals and Expected Outcomes

The project aims to leverage social media indicators with predictive power and use them to build a model that estimates the likelihood and magnitude of demand surges during TikTok driven viral events.

The sponsor company will be provided with the following outcomes:

- A predictive model to estimate the probability and magnitude of demand spikes driven by TikTok virality
- A scenario-based framework that enables planners to simulate different scenarios by adjusting model inputs and probability thresholds, supporting more proactive and data-driven decision making
- A set of recommendations to operationalize the model for improved predictive performance

Once implemented, CosmeticCo can integrate this framework into demand forecasts across parent products, divisions, and regions. By incorporating social media variables that are not currently captured in the baseline forecast, the model improves visibility into demand variability, particularly during high-demand events. This enables more informed planning decisions and supports better inventory management.

2. State of the Practice

The central problem we aim to address is how to incorporate the impact of social media virality into the existing forecasting methodology of the sponsor company. To better understand the current context and address this problem, we reviewed relevant literature on this topic. We sought to define what is meant by virality, specifically social media (TikTok) virality. We also explored the metrics used and tracked on TikTok to determine the correlation between TikTok virality and its impact on actual sales. Finally, we examined the types of predictive models that can be developed based on the social media performance data that can be gathered.

2.1 TikTok Virality and Key Metrics

TikTok virality generally refers to any clip that achieves fast and widespread popularity within a short time horizon, often measured by the number of likes, shares, views, comments, etc. "A clip is typically deemed viral if it accumulates over 1 million views within 72 hours" (fresh content society, 2025). At the core of TikTok is its algorithm, which pushes and favors engaging and relatable content to gain traction (fresh content society, 2025). Virality can be organic, driven by online trends that gain popularity and lead to user sharing, or inorganic, driven by planned promotions from individuals and organizations, which involves using a budget for paid promotion to achieve viral reach. Whether organic or amplified, TikTok's algorithm uses advanced techniques to analyze user behavior and interactions and to push the right content to the right audience.

TikTok records several metrics, including audience and reach metrics, traffic and source metrics, live stream metrics, and core content and engagement metrics. Some of the key metrics to be used in our analysis include:

- Product Impressions: Indicates product visibility and exposure. Number of times a product is shown to users.
- Product Clicks: Number of times users click on a product listing. Shows user interest in exploring the product further.

A critical concept is the distinction between engagement and virality. The former offers a perspective on popularity, while the latter captures the rate at which popularity disseminates. Both are important and can provide predictive insights; however, virality reflects the momentum of change, a factor often missing in traditional forecasting. Influencer activity is another demand driver that demonstrates how social endorsements can influence purchasing behavior, especially when influencers have large or highly engaged audiences. The impact of this influence on purchases and, thus, demand is further amplified by platforms such as TikTok, where a single piece of content can reach millions of users.

The convergence of these factors — engagement metrics (e.g., product clicks, likes, shares, impressions, comments) and viral velocity (e.g., view and engagement growth rates over time) has elevated social media from a marketing channel to an essential forecasting resource.

2.2 Social Media Metrics as a Catalyst for Improving Demand Forecasts

Traditional forecasting methods often rely on historical sales patterns, seasonality, and internal business inputs. However, these methods may not fully capture external demand signals created by social media activity. Social media data provides companies with an additional source of market information, enabling predictions that extend beyond internal variables and support improved supply chain decision making (Iftikhar et al., 2020). Prior research has shown that social media signals can be useful for predictive analytics across different fields. Oliveira et al. (2017), for example, studied how Twitter data can help predict stock market returns, volatility, trading volume, and sentiment indices. Nguyen et al. (2021) similarly highlighted the relationship between social media engagement and enrollment intentions in higher education. Together, these studies suggest that social media engagement data can provide meaningful external signals for forecasting behavior and demand-related outcomes.

According to Badulescu et al. (2024), social media activities can lead to four distinct impacts on demand:

- Transient Effects – A demand spike caused by viral trends, influencer marketing, or short-term promotions, the impact of which quickly diminishes. Companies should adjust immediate production plans and not long-term forecasts to manage its impact.
- Quantum Jumps – Permanent shifts in demand. Social media can serve as an early warning signal if it shows sustained impact across all parameters. These trends persist over time, allowing companies to proactively scale up operations.
- Transferred Impact – Demand shifts between time periods because of early discounts or promotions. The long-term forecast should be adjusted through redistribution to factor in this internal seasonal effect.

- Trend Changes – Long-term shifts in consumer preferences, either upward or downward. Recognizing these changes allows companies to plan for strategic changes that are necessary (Badulescu et al.,2024).

To understand the resultant change in demand, it is essential to track the right social media metrics that will allow the predictive model to pick up the correct signals.

2.3 Predictive Modeling Approaches for Demand Forecasting

A variety of analytical approaches can be applied to integrate and analyze social media variables for demand forecasting. These include both **traditional time-series models** and **machine learning (ML) techniques**, each offering distinct advantages. Time-series models such as ARIMA and SARIMA are effective at capturing temporal structures like trends and seasonality, while machine learning methods are better suited to modeling nonlinear relationships and incorporating external drivers such as social media signals (Alassafi et al., 2023).

We evaluated four categories of models:

Time-Series Models:

Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Prophet (developed by Facebook) are widely used for forecasting time-dependent data. These models rely on historical observations to predict future values and are effective in capturing trends, seasonality, and cyclical patterns. ARIMA models combine autoregression (dependence on past values), differencing (to achieve stationarity), and moving averages (to smooth noise), while SARIMA extends this framework to explicitly model seasonal effects. Prophet, in contrast, decomposes time series into trend, seasonality, and holiday effects, making it particularly robust to missing data and structural changes (Gubkin, 2023).

Regression Models:

Regression models estimate the relationship between a dependent variable and one or more independent variables. Examples include linear regression, logistic regression, and quantile regression. Logistic regression models the probability of a binary outcome using a logistic function, making it suitable for classifying whether demand will enter a high-demand state. This probabilistic output allows decision makers to apply thresholds based on risk tolerance. Quantile regression extends traditional regression by estimating conditional quantiles of the target variable, enabling the modeling of different points of the demand distribution (e.g., median vs. upper quantiles), which is particularly useful for capturing demand uplift during extreme events (Waldmann, 2018).

Tree-Based Models:

Tree-based models, such as decision trees, random forests, and gradient boosting machines, construct multiple decision rules to generate predictions. These models are effective at handling nonlinear relationships and capturing complex interactions between variables without requiring strong parametric assumptions. Random forest, for example, builds multiple decision trees using bootstrap samples (i.e., random samples of the dataset drawn with replacement) and introduces feature randomness at each split, improving generalization and reducing overfitting. Gradient boosting models further enhance predictive performance by sequentially correcting errors from previous trees (Oliveira et al., 2017).

Auto Regressive and Deep Learning Models:

Autoregressive (AR) models and Vector Autoregressive (VAR) models predict future values as a function of past observations, capturing temporal dependencies across one or multiple time series. Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and

Gated Recurrent Unit (GRU) architectures, extend this approach by modeling sequential data through internal memory states that retain information from prior time steps. LSTMs, in particular, use gating mechanisms (input, forget, and output gates) to control information flow, enabling them to capture long-term dependencies and delayed effects, such as lagged demand responses following viral events (Chug, 2025). However, these models typically require large datasets, extensive tuning, and may lack interpretability in business applications.

The literature indicates that forecasting demand influenced by social media virality requires combining temporal modeling with exogenous variables, such as engagement and virality metrics. While traditional time series models effectively capture baseline demand patterns, machine learning approaches are better suited to modeling nonlinear effects, threshold behaviors, and sudden demand shifts driven by viral events.

Based on insights derived from the literature, logistic regression and quantile regression were identified as the most appropriate modeling approaches for this context. Logistic regression enables probabilistic classification of demand spikes by estimating the likelihood of a high-demand state, while quantile regression captures the distribution of potential demand outcomes, providing insight into the magnitude of uplift beyond average forecasts.

Together, these approaches balance interpretability, robustness, and predictive capability, making them well suited to the specific requirements of this study, where both the likelihood and magnitude of demand surges driven by social media virality must be quantified.

3. Methodology

In this section, we explain our methodological approach, the type of data used, and how we analyzed the data to address the project’s objective: leveraging social media indicators with predictive power to develop a model that estimates both the probability and magnitude of demand surges associated with TikTok activity.

3.1 Methodology Selection

After defining the project's scope, reviewing relevant literature, conducting interviews with key stakeholders, and considering state-of-the-practice methodologies, we identified three methods to detect early warning signals of TikTok virality and analyze its impact on demand:

- Utilize logistic regression as a supervised machine learning classification technique at the parent product level to analyze historical demand, baseline forecast performance, and TikTok engagement signals, identifying the probability of being in a high-demand state within the next seven days.
- Overlay TikTok data with actual demand and the baseline forecast to allow analysis to be conducted on the correlation between virality metrics, demand spikes, and forecast accuracy.
- Use quantile regression to estimate the expected uplift in demand based on the probability estimation found from logistic regression.

Currently, CosmeticCo aims to optimize the lead time between social signal detection and planning adjustments. The marketing and demand planning teams seek to synchronize communication windows to ensure promotional activity aligns with inventory availability. The marketing team has access to crucial engagement metrics that drive TikTok virality and convert to increased demand within the platform. By utilizing regression analysis as a bridge between the two functions, the company can achieve greater agility in adjusting demand forecasts when demand spikes occur.

3.2 Methodology Overview

To provide CosmeticCo with an early warning signal to indicate potential virality and assess the impact of TikTok virality on demand, we followed the five main steps shown in Figure 1.

1. Scope Definition	2. Data Integration and Alignment	3. High-Demand Detection	4. Demand Uplift Estimation	5. Results Communication
• Conduct stakeholder interviews • Define high-demand planning problem • Identify model outputs needed by internal stakeholders	• Collect forecast, sales, and TikTok data • Align data by product and date • Create baseline vs. actual demand view	• Define high-demand threshold vs. baseline forecast • Use TikTok impressions and clicks as predictors • Estimate probability of high demand within seven days	• Calculate uplift above baseline forecast • Estimate expected uplift using quantile regression • Compare predicted uplift to actual uplift	• Present high-demand probability and alert status • Summarize expected uplift and capture rate • Translate outputs into clear stakeholder-facing results

Figure 1: Capstone Methodology Steps

Our process began with interviews with key stakeholders across CosmeticCo's US based teams to define the planning problem, identify available data sources, and determine which model outputs would be most useful for internal decision making. After the interview process, we received data from multiple functions and conducted follow-up discussions to clarify assumptions and ensure proper understanding of the data. We then integrated forecast, sales, and TikTok datasets into a common time series structure, allowing baseline forecasts, actual demand, and social media engagement signals to be compared by product and date. Once the data was cleaned and aligned, we used logistic regression to estimate the probability of a high-demand state within the next seven days. We then used quantile regression to estimate the expected incremental demand uplift above the baseline forecast when a high-demand event occurs. Finally, we translated the model outputs into stakeholder-facing results, including high-demand probability, expected uplift, aggregate capture rate, and model performance summaries, to clearly communicate how TikTok engagement signals can support demand planning.

3.3 Data Collection

The project focuses on a selected parent product within CosmeticCo's North American Consumer Products Division (CPD) that has experienced prior organic or promotional virality, resulting in demand surges above forecasted levels. For this product, the following types of data were obtained, organized as monthly or daily time series:

<u>Demand Planning</u>

- Forecast Fact: Provides the underlying monthly forecast at a granular level, including dimensions such as geography, forecast type, shipment date, and other operational attributes.
- Sales Fact: Captures actual daily sales at a similarly detailed level, incorporating many of the same dimensions available in the forecast dataset.

<u>Marketing and Promotion</u>

TikTok data: Product Impressions and Product Clicks

3.4 Data Preparation and Preprocessing

Data preprocessing was conducted to clean, structure, and align forecast and sales datasets for daily level demand analysis, addressing differences in data granularity and multiple forecast versions. This involved the following three steps:

1) **Sales and Forecast Fact Cleaning and Aggregation —** The Sales and Forecast Fact datasets were cleaned and structured to ensure consistency and alignment by retaining relevant columns, removing unnecessary fields, and correcting missing or inconsistent values. The forecast data was then aggregated across key dimensions, such as material code, forecast month, version, channel, plant, and location, while the sales data was aggregated at a daily level by material code and date to capture actual demand patterns.
2) **Daily Weight Calculation and Normalization —** To distribute monthly forecasts, daily sales patterns were derived by computing monthly sales totals for each material code and dividing daily sales by these totals to calculate daily weights. These weights were calculated across the full dataset, irrespective of month or year, to capture underlying demand patterns. They were then normalized to ensure they sum to 1 for each month, enabling accurate distribution of monthly forecasts without over or under allocation.
3) **Merging, Daily Forecast Generation, and Version Tagging —** Normalized daily weights were applied to monthly baseline forecasts to disaggregate them into daily level estimates. Sales data, originally at a more granular transactional level, was aggregated to daily totals. The actual dataset was then reshaped to match the forecast structure by adding required fields. The forecast and actual datasets were subsequently combined through vertical concatenation into a unified dataset, with a tag distinguishing baseline forecasts from actuals. Forecast versions were sorted chronologically, and only the earliest version per month was retained as the baseline for comparison against realized demand.

Overall, these steps ensured that the forecast and sales datasets were standardized and brought to the same daily granularity, making it possible to conduct a more detailed forecast accuracy and demand pattern analysis.

3.5 Logistic Regression Framework for High-Demand Regime Prediction

Our model uses binary logistic regression to estimate the probability of a high-demand regime within the next seven days based on the social media metrics observed at the time of the analysis. Logistic regression is conducted daily, with a seven-day time horizon. We define a binary demand state as follows:

$$\boldsymbol{D_t} \, \epsilon \{\boldsymbol{HD}, \boldsymbol{LD}\}$$

- HD (High Demand Regime): Actual Demand in the next seven days ≥ 1.60x baseline forecast
- LD (Low Demand Regime): Actual Demand in the next seven days < 1.60x baseline forecast

The probability of entering a high-demand regime is modeled using logistic regression as follows (Russell & Norvig, 2020):

$$\boldsymbol{P_t} = \boldsymbol{P}(\boldsymbol{D_t} = \boldsymbol{HD} | \boldsymbol{X_t}) = \frac{\mathbf{1}}{\mathbf{1} + \mathbf{exp}\,(-(\boldsymbol{\beta_0} + \boldsymbol{\beta^T X_t}))}$$

Where X_t represents the observed TikTok signals at time t, the model estimates the probability that demand D_t is in a high-demand regime (HD), with β_0 establishing the baseline and $\beta^T X_t$ adjusting that baseline based on the influence of social media signals.

This formulation maps observed TikTok engagement metrics to a probability score that indicates the likelihood of a demand spike within the next seven days. The probability threshold can then be used to trigger different levels of planning response. In the final model, the selected alert threshold was 30%, meaning the model flags a high-demand event when the predicted probability is at least 0.30. This lower threshold is consistent with the early warning objective of the model because it increases sensitivity to potential demand spikes, while leaving the final planning response to business judgment.

In practice, CosmeticCo could use the probability output to support a tiered response process. Lower probability scores would indicate continued monitoring of TikTok metrics and demand signals. Moderate scores that exceed the selected alert threshold could trigger low-risk and reversible actions, such as reserving capacity, checking supplier readiness, preparing overtime options, or increasing monitoring frequency. Higher scores, especially if they persist across multiple checks, could justify escalation to more involved decisions, such as adjusting production plans, repositioning inventory, or modifying allocation priorities. The exact thresholds for each action should be calibrated by product and by business risk tolerance, since different SKUs may have different demand volatility, lead times, and costs of stockout or overstock.

3.6 Logistic Regression Parameter Settings

To make the logistic regression model adaptable across products, the model includes both a manual parameter option and a parameter sweep option. The manual option allows the user to directly set the high-demand definition, probability threshold, train/test split, and logistic regression settings. This is useful when the business wants to test a specific planning scenario or apply a previously approved configuration. The parameter sweep option, on the other hand, tests multiple combinations of these values and selects the best-performing configuration based on a weighted score. This weighted score prioritizes high-demand classification performance by combining F1 score, recall, precision, and ROC AUC. This is important because the objective is not only to maximize overall accuracy, but to identify demand spikes in a way that is useful for planning.

The key logistic regression parameters are the baseline multiplier, probability threshold, train/test split ratio, regularization strength, class weight, and solver. The baseline multiplier defines what qualifies as high demand by comparing actual seven-day demand against the baseline forecast. The probability threshold converts the model's probability output into a high-demand alert. The train/test split ratio determines how much historical data is used for model training versus testing. The regularization parameter, class weight, and solver control how the logistic regression model is fit. For the final model, the selected settings were a baseline multiplier of 1.60, probability threshold of 0.30, train/test split of 0.65, C value of 10, no class weighting, and the liblinear solver. These values reflect a stricter definition of high demand, while still allowing the model to flag potential risk early enough to support planning decisions.

3.7 Quantile Regression Framework for Demand Uplift

Overlaying TikTok metrics indicating potential virality with actual demand and the baseline forecast enables estimation of the expected conditional demand uplift via quantile regression. We define incremental demand as:

$$\boldsymbol{\Delta D \equiv Incremental\ demand = Actual\ demand - Baseline\ Forecast}$$

Expected conditional demand uplift is defined as the incremental demand given the TikTok metrics at the time of the analysis. Formally, this can be expressed as:

$$\boldsymbol{Expected\ Conditional\ Uplift = E[\Delta D|X_t] = P(D_t = HD|X_t) \times E[\Delta D|D_t = HD, X_t]}$$

Which follows from the law of total expectation, subject to $\boldsymbol{E[\Delta D|D_t = LD, X_t] = 0}$ (i.e., there is no expected uplift in the low-demand state)

Quantile regression is used as an approximation to estimate the conditional uplift term ($\boldsymbol{E}[\Delta \boldsymbol{D} | \boldsymbol{D_t} = \boldsymbol{k}, \boldsymbol{X_t}]$). Specifically, it estimates $\boldsymbol{Q}[\Delta \boldsymbol{D} | \boldsymbol{D_t} = \boldsymbol{HD}, \boldsymbol{X_t}]$, allowing the model to capture asymmetric and extreme demand responses associated with viral events. Incorporating an estimate of the expected conditional uplift enables CosmeticCo to quantify the magnitude of potential demand increases given the probability of a high-demand state. Together, the logistic regression and quantile regression components form an integrated framework that links virality signals to both the likelihood and magnitude of demand spikes.

3.8 Quantile Regression Parameter Settings

The quantile regression model also includes a user-selected parameter, TAU, which determines the uplift scenario being estimated. Unlike the logistic regression parameters, TAU is not selected by the parameter sweep. It is treated as a business planning input because different teams may have different risk tolerances. A lower value, such as 0.50, estimates a more typical or median uplift. A higher value, such as 0.75 or 0.90, estimates a more conservative upside scenario.

For the final model, TAU was set to 0.75, meaning the uplift model estimates the 75th percentile of demand uplift during high-demand events. This value was selected because it provides an elevated planning estimate without relying only on extreme outcomes. In this context, the 75th percentile is useful because it gives planners a buffer against under forecasting while remaining practical for inventory and production decisions. Since products may differ in demand volatility, TikTok responsiveness, and operational risk, TAU can be adjusted for future products depending on the business case. Changing TAU will affect the estimated uplift and should be reviewed alongside coverage, mean pinball loss, and aggregate uplift capture rate.

4. Results and Discussion

To evaluate the effectiveness of the proposed framework, we implemented a two-stage modeling approach. First, logistic regression was used to estimate the probability of entering a high-demand regime within the next seven days. Second, quantile regression was used to estimate the magnitude of demand uplift conditional on that regime. The results provide insight into both the model's predictive capability and its practical relevance for demand planning.

4.1 Logistic Regression Results: High-Demand Detection

The logistic regression model was trained using historical daily observations for the selected parent product. Each observation represents one date and includes TikTok engagement metrics, specifically product impressions and product clicks, transformed into lagged and rolling features to capture recent social media activity before the demand outcome occurs. The dependent variable is a binary high-demand label, where a date is classified as high demand if actual seven-day demand exceeds the baseline forecast by the selected multiplier. The model was trained on the earlier portion of the time series and evaluated on a later test period using a time-based split. This structure avoids randomly mixing past and future observations and better reflects how the model would be used in practice, where historical data is used to predict future demand risk. After training, the model outputs a probability score for each test date, representing the likelihood of entering a high-demand state within the next seven days.

To measure performance, we use several classification metrics that assess how well the model distinguishes between high and low demand periods. One key metric is the Receiver Operating Characteristic Area Under the Curve (ROC AUC). The ROC curve measures how well a model can distinguish between two classes across all possible probability thresholds, and the AUC summarizes this performance as a single value between 0 and 1. A value closer to 1 indicates stronger discriminative ability (Fawcett, 2006). Figure 2 shows that the model attained an ROC AUC of 0.83, indicating that it has a good ability to rank high-demand periods above normal-demand periods.

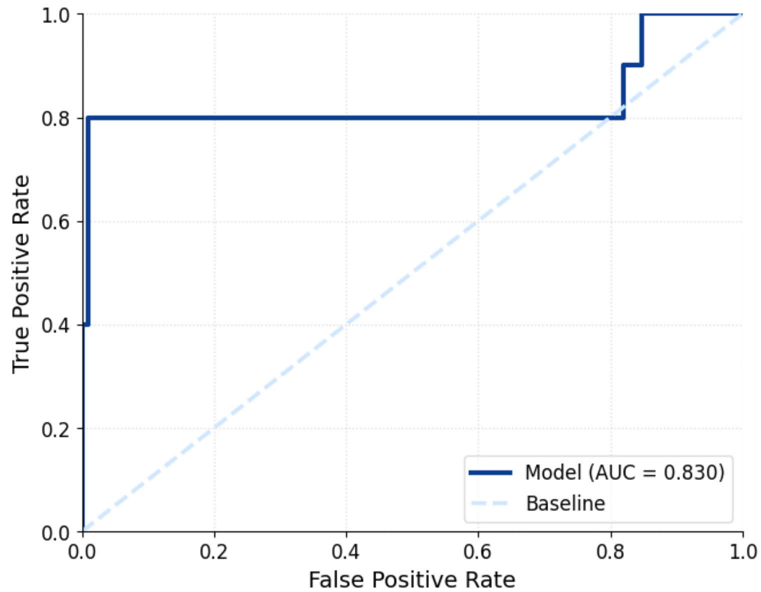

Figure 2: ROC Curve – Demand Classification Performance

While ROC AUC evaluates the model's ability to rank high-demand periods above normal-demand periods across all possible thresholds, it does not show how the model performs at the selected decision threshold. For operational interpretation, we also evaluate recall and precision for the high-demand class, as shown in Figure 3.

Recall, also known as sensitivity, measures the proportion of actual high-demand events that are correctly identified by the model:

$$\boldsymbol{Recall} = \frac{\boldsymbol{True\ Positives}}{\boldsymbol{True\ Positives} + \boldsymbol{False\ Negatives}}$$

In the final model, high-demand recall was 0.40, meaning the model correctly identified 4 of the 10 actual high-demand days in the test period. This indicates that the model captured some demand spikes but also missed 6 high-demand events. Because the model is intended to serve as an early warning signal, these missed events are important to monitor and should be a focus for future model improvement.

Precision, measures the proportion of predicted high-demand events that were actually correct:

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ Positives}}{\boldsymbol{True\ Positives} + \boldsymbol{False\ Positives}}$$

The model achieved a high-demand precision of 1.00, meaning that every high-demand alert generated by the model was correct. In other words, the model produced 4 high-demand alerts and all 4 corresponded to actual high-demand days. This suggests that the model is conservative in flagging high-demand periods, but the alerts it does generate are reliable (Chase, 2017).

The model's classification accuracy was 95%, meaning that 95% of all test-period observations were correctly classified as either high demand or normal demand. Formally, classification accuracy is calculated as:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{True\ Positives}\ + \boldsymbol{True\ Negatives}}{\boldsymbol{Total\ Observations}}$$

However, accuracy alone is less informative in this context because the test set is imbalanced. High-demand days represented only 10 of 115 test observations, or 8.7% of the test set, while normal-demand days represented the remaining 91.3%. As a result, a model can achieve high overall classification accuracy by correctly predicting normal-demand days, even if it misses some demand spikes. For this reason, high-demand recall and precision provide a more meaningful evaluation of whether the model is useful as an early-warning signal.

Figure 3 presents the confusion matrix at the selected probability threshold. A true positive represents a high-demand day that the model correctly flagged, while a false negative represents a high-demand day that the model missed. A true negative represents a normal-demand day that was correctly not flagged, and a false positive represents a normal-demand day that was incorrectly flagged as high demand. In this model run, the confusion matrix shows 4 true positives, 6 false negatives, 105 true negatives, and 0 false positives, reinforcing that the model generated highly reliable alerts but did not capture every high-demand event.

Figure 3: Confusion Matrix: High-Demand Classification

4.2 Interpretation of Probability Outputs

The logistic regression model produces a probability estimate representing the likelihood of entering a high-demand regime within the next seven days. These probabilities can be interpreted as a ranking of risk over time, with higher values indicating a greater likelihood of a spike in demand.

A probability threshold of 30% was selected through parameter tuning to convert the model's probability output into a high-demand alert. Lower thresholds generally increase the number of alerts and can improve sensitivity, while higher thresholds reduce alerts but may miss more true high-demand events. In this model run, the selected threshold produced a focused alerting pattern: the model flagged fewer high-demand days, but every alert it generated was correct, resulting in 100% precision and 40% recall for the high-demand class. This suggests that the model's probability output is useful as a reliable risk signal, while also indicating an opportunity for future refinement to capture a larger share of demand spikes.

4.3 Quantile Regression Results: Uplift Estimation

To estimate the magnitude of demand increases, we implemented quantile regression, which models specific points (quantiles) of the outcome distribution rather than the mean. This is particularly useful in contexts with asymmetric or extreme outcomes, such as demand spikes driven by social media virality (Koenker & Hallock, 2001).

The model was evaluated across multiple quantiles, with particular focus on τ= 0.75, representing an upper-bound estimate of demand. This quantile captures elevated-demand scenarios without focusing solely on extreme outliers.

One key evaluation metric for quantile regression is the mean pinball loss, which measures the deviation between predicted and actual values while penalizing underestimation more heavily at higher quantiles (Koenker & Hallock, 2001). In the final model, mean pinball loss was approximately 20,402 units of demand uplift per high-demand test observation. This value is measured against an average actual uplift of 33,353 units per high-demand event, indicating that the error is meaningful but expected given the size and volatility of TikTok driven demand spikes. Because pinball loss is technical and less intuitive from a business perspective, it is supplemented with coverage analysis.

Coverage measures the proportion of actual outcomes that fall below the predicted quantile. For a well calibrated model at τ = 0.75, approximately 75% of observations should fall below the predicted value. In our case, the model achieved coverage of 0.70, meaning that 70% of actual high-demand uplift observations were below the predicted q75 uplift estimate. This is slightly below the target coverage, indicating that the model was close to the intended quantile but slightly under-covered the actual uplift outcomes.

This result indicates that the model is close to the intended q75 estimate but slightly undercovers actual uplift outcomes. Rather than consistently overestimating demand, the model was somewhat less conservative than expected in this test period. This suggests that future refinement should focus on improving uplift calibration, so the model better captures larger demand spikes while maintaining practical planning value.

Figure 4 illustrates this relationship by comparing target quantiles with observed coverage. Points on the diagonal line represent perfect calibration, points above the line indicate over-coverage, and points below the line indicate under-coverage. In this case, the q75 point falls slightly below the diagonal, showing that the model was close to the intended quantile but somewhat less conservative than expected.

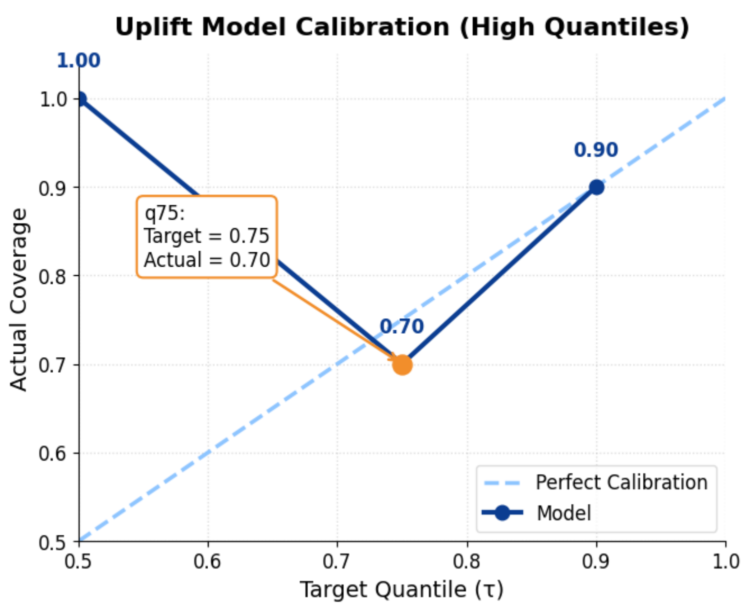

Figure 4: Quantile calibration plot with selected tau = 0.75 (q75).

4.4 Predictive vs. Actual Uplift

Predicted uplift was also compared with actual uplift during high-demand periods. Figure 5 shows this relationship relative to a 45-degree reference line, where points on the line indicate perfect prediction. At the aggregate level, the model predicted 241,174 units of uplift compared with 333,526 units of actual uplift, resulting in a net uplift gap of 92,352 units, or 27.7% of actual uplift. This corresponds to a 72.3% capture rate. The figure shows that prediction error varies across individual high-demand events, with some events underpredicted and others closer to the reference line. Overall, the model provides a useful directional estimate of uplift, but additional calibration is needed to improve event-level accuracy.

Several factors contribute to the observed variation in uplift prediction accuracy. First, the limited historical data available for the selected parent product reduces the model's ability to fully learn demand patterns associated with viral events. Second, the conversion of monthly forecasts into daily values using historical weights introduces approximation error, particularly during periods with atypical demand patterns. Finally, demand behavior may still vary across the underlying SKUs within the parent product, and this variation is not fully captured by the current feature set. These limitations contribute to the dispersion observed in the prediction results.

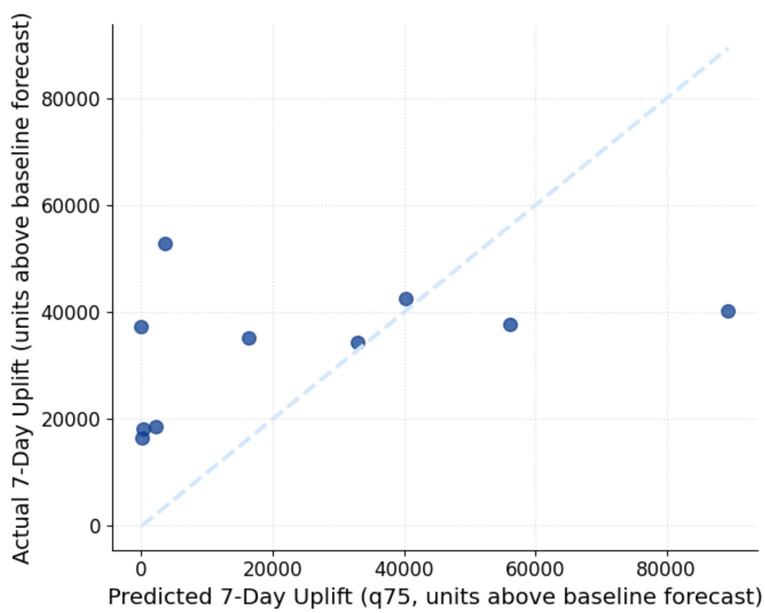

Figure 5: Predicted vs Actual Uplift

4.5 Aggregated Uplift Performance

To evaluate the model's overall effectiveness, we aggregate uplift across the actual high-demand days in the holdout test period. This analysis is specific to the selected parent product and reflects model performance across a 115-day test period, which included 10 actual high-demand days. Aggregating across these events allows us to compare how much incremental demand the model estimated relative to what was actually observed, and to benchmark that performance against the baseline forecast.

Across the high-demand days, actual demand reached an index of 258 compared to a baseline forecast index of 100. This indicates that the baseline forecast did not account for approximately 158% of incremental volume observed during viral windows. The proposed model estimated an uplift index of 114, effectively capturing 72.3% of the actual incremental demand. Unlike earlier model iterations, the final model does not overestimate aggregate uplift; instead, it underestimates total uplift but still captures a meaningful share of demand that was not reflected in the baseline forecast.

This result shows the improvement from the baseline only approach to the model assisted approach. Without the model, the entirety of the incremental volume would remain outside the baseline forecast. With the model, the majority of that uplift is estimated, reducing the unanticipated uplift gap by 72.3 percentage points relative to relying on the baseline forecast alone. This provides planners with a more structured view of demand risk during TikTok driven high-demand periods.

The capture rate, defined as predicted uplift divided by actual uplift, provides a useful aggregate measure of model performance. In this test period, the model's 72.3% capture rate indicates that it identified most, but not all, of the incremental demand above baseline. This metric is useful because it translates model performance into planning terms: the higher the capture rate, the more of the demand gap the model makes visible before or during the planning window.

When viewed alongside the baseline forecast, these results show the gap that exists when social media signals are not incorporated into demand planning. The baseline forecast alone does not account for a large share of demand variability during high-demand events, while the model provides a structured way to estimate both the likelihood and magnitude of that uplift. This supports more informed planning discussions, particularly when demand is influenced by rapid changes in consumer attention and social media engagement.

4.6 Current Limitations

While we have built a preliminary model, we have encountered several challenges along the way:

	Sales Fact	Forecast Fact	TikTok Data
Time Level Granularity	Daily	Monthly	Daily
Product Level Granularity	Material / SKU Level	Material / SKU Level	Parent Product Level

Figure 6: Data Granularity Mismatch Across Source Systems

1) **Data Granularity Mismatch:** We received data from three different sources, Sales_Fact, Forecast_Fact, and TikTok metrics—each with varying levels of granularity across time and product hierarchies (as shown in Figure 6). This required extensive data manipulation and the use of multiple assumptions (e.g., weight normalization, imputation, and feature lagging) to transform these disconnected inputs into a single usable dataset.
2) **Forecast Disaggregation:** Monthly forecasts are distributed into daily values using historical sales patterns. This approach assumes a fixed intra-month distribution that does not vary by month or year, limiting the model's ability to capture seasonal shifts such as holidays or promotional peaks.
3) **Data Limitation:** For most product families, we have limited data (<14 months) and missing values for certain SKUs, which had to be excluded from the model. This reduces the number of usable observations and weakens overall model reliability. Additionally, we were provided with very limited TikTok signals, which are currently incorporated into the model. Expanding to broader social engagement metrics (e.g., likes, shares, and comments) could significantly improve model robustness.

5. Recommendations

This project developed a predictive framework to quantify the impact of TikTok driven virality on demand, combining logistic regression to estimate the probability of high-demand events and quantile regression to estimate the magnitude of demand uplift. The results show that TikTok engagement metrics significantly improve the identification of high-demand periods compared to relying on baseline forecasts alone, enabling earlier and more reliable detection of demand spikes. Once a high-demand event is identified, the model provides a directional estimate of uplift, successfully capturing the majority of incremental demand while maintaining a conservative bias that helps mitigate the risk of under forecasting. Finally, model outputs can be structured into clear, decision-oriented signals, allowing internal stakeholders to translate predictions into actionable supply chain decisions.

While the framework demonstrates strong predictive capability, the analysis also highlights limitations related to data granularity, feature availability, and operational integration. Addressing these gaps presents a clear opportunity for CosmeticCo to further enhance forecast accuracy, reduce stockouts, and improve supply chain responsiveness.

5.1 Management Recommendations

1) Integrate Real-Time Data Pipelines: To fully realize the value of the model, CosmeticCo should transition from static datasets to automated data pipelines that continuously ingest daily sales, forecast, and TikTok metrics. This would enable:

- Daily model refresh and real time monitoring
- Faster detection of demand spikes
- Reduced lag between signal detection and operational response

This would improve responsiveness to demand spikes, reduce delays in decision-making, and enable more proactive supply chain adjustments, ultimately supporting stronger revenue performance and improved service levels.

2) Improve Forecast and Signal Granularity: Current limitations in both forecast and TikTok data reduce model precision.

- Forecast disaggregation should incorporate day-of-week effects and promotional calendars to better capture intra-month variability
- TikTok signals should be integrated at the SKU level rather than the parent level

This will:

- Improve baseline forecast accuracy
- Enable more precise uplift estimation
- Allow direct translation into SKU-level production and inventory decisions

This improves alignment between demand sensing and execution, enabling more agile and targeted supply chain decisions.

3) Establish Continuous Feedback Loops (S&OP Integration): Model performance should not be treated as static but continuously improved through structured feedback loops. In addition to tracking metrics such as recall, precision, and uplift calibration, CosmeticCo should implement:

- Regular cross-functional reviews between demand planning and TikTok marketing teams
- Integration of model outputs into Sales & Operations Planning (S&OP) processes
- Ongoing refinement of model inputs, including new parameters identified through business insights

- Continuous model retraining using updated data, including new SKUs, campaign signals, and evolving demand patterns

These feedback loops ensure that:

- Model predictions are validated against actual outcomes
- Business context (e.g., campaigns, influencer activity) is incorporated
- The model continuously learns from new data, adapts to SKU changes, and improves over time

This enables sustained improvement in forecast accuracy, leading to better inventory positioning and reduced mismatch between supply and demand.

4) Expand Data and Feature Set: The current model relies on a limited set of TikTok metrics, which constrains predictive power. CosmeticCo should expand inputs to include:

- Additional engagement signals (likes, shares, comments)
- Influencer activity and campaign spend
- Cross-channel demand signals

These enhancements will:

- Improve early detection of virality
- Capture nonlinear drivers of demand
- Increase robustness across different product categories

This strengthens the company's ability to anticipate demand shifts driven by external signals, supporting a competitive advantage in a rapidly evolving digital landscape.

5) Operationalize Model Outputs into Decision-Making: To create business impact, model outputs must be directly linked to supply chain actions. This can be achieved by:

- Defining probability thresholds tied to decisions
  - Tier 1 Threshold→ early warning actions (capacity readiness, supplier alignment)
  - Tier 2 Threshold→ escalation actions (production increases, inventory positioning)
- Embedding outputs into planning systems and workflows

This ensures:

- Clear decision rules
- Faster execution
- Reduced reliance on manual interpretation

This enables a shift from reactive to proactive planning, reducing stockouts and excess inventory while improving service levels.

5.2 Future Work

While this project demonstrates the value of integrating social media signals into demand forecasting, several opportunities exist to further enhance the framework.

Future work should focus on:

- Expanding the dataset across more parent products and longer time horizons to improve model generalizability.
- Testing more advanced models (e.g., tree-based or deep learning approaches) to capture nonlinear relationships.
- Refining forecast disaggregation methods to better capture seasonality and event driven variability.

- Exploring cross-platform signals beyond TikTok to create a more comprehensive demand sensing system.

More broadly, this study highlights a fundamental shift in demand planning: external signals such as social media are becoming critical drivers of demand variability.

The proposed framework demonstrates how these signals can be systematically integrated into forecasting, offering value not only to CosmeticCo but also to other organizations operating in industries where demand is increasingly influenced by digital and social dynamics.

References

Alassafi, M. O., Alghamdi, W., S. Sathiya Naveena, Alkhayyat, A., Absalomov Tolib, & Ibrokhimov Sarvar Muydinjon Ugli. (2023). Machine Learning for Predictive Analytics in Social Media Data. *E3S Web of Conferences*, *399*, 04046–04046. https://doi.org/10.1051/e3sconf/202339904046

Backhaus, K., Erichson, B., Gensler, S., Weiber, R., & Weiber, T. (2025). Logistic Regression. In *Multivariate Analysis: An Application Oriented Introduction* (pp. 269–358). Springer Gabler Wiesbaden.

Badulescu, Y., Cañas, F., & Cheikhrouhou, N. (2024). *Judgmental adjustment of demand forecasting models using social media data and sentiment analysis within industry 5.0 ecosystems*. International Journal of Information Management Data Insights, 4(2), 100272. https://doi.org/10.1016/j.jjimei.2024.100272

Chase, C. (2017). *Machine Learning Is Changing Demand Forecasting*. ProQuest; Journal of Business Forecasting. https://www.proquest.com/docview/1881110602/fulltext/4EA2C9C3CEB94415PQ/1?accountid=12492&sourcetype=Scholarly%20Journals

Chug, A. (2019, January 16). *What is LSTM Long Short Term Memory?* GeeksforGeeks. https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Gubkin, A. (2025). *Machine Learning for Forecasting: Use Cases & Building Your Own*. Coralogix. https://coralogix.com/ai-blog/machine-learning-for-forecasting-use-cases-and-building-your-own/

Iftikhar, R., & Khan, M. S. (2020). Social Media Big Data Analytics for Demand Forecasting. *Journal of Global Information Management*, *28*(1), 103–120. https://doi.org/10.4018/jgim.2020010106

Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, *15*(4), 143–156. https://doi.org/10.1257/jep.15.4.143

Nguyen, L., Lu, V. N., Conduit, J., Tran, T. T. N., & Scholz, B. (2020). Driving enrolment intention through social media engagement: a study of Vietnamese prospective students. *Higher Education Research & Development*, 1–16. https://doi.org/10.1080/07294360.2020.1798886

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, *73*, 125–144. https://doi.org/10.1016/j.eswa.2016.12.036

Russell, S. J., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.

Waldmann, E. (2018). Quantile regression: A short story on how and why. *Statistical Modelling*, *18*(3-4), 203–218. https://doi.org/10.1177/1471082x18759142

What Makes a TikTok Viral: Key Factors and Strategies Explained. (2025, August 15). Fresh Content Society. https://freshcontentsociety.com/what-makes-a-tik-tok-viral-key-factors-and-strategies-explained/