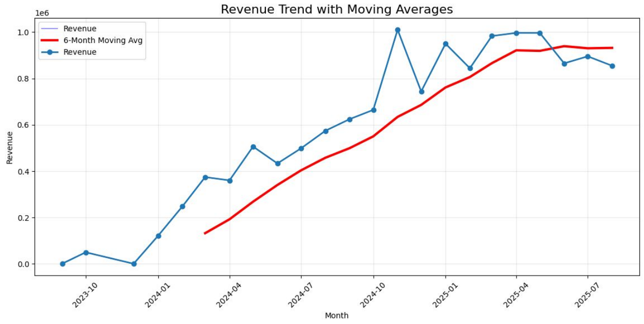
TOMORROW MARKETERS BUSINESS ANALYTICS & CONSULTING

NAME: HOANG VIET HUNG - TEAM: TM PRODUCT

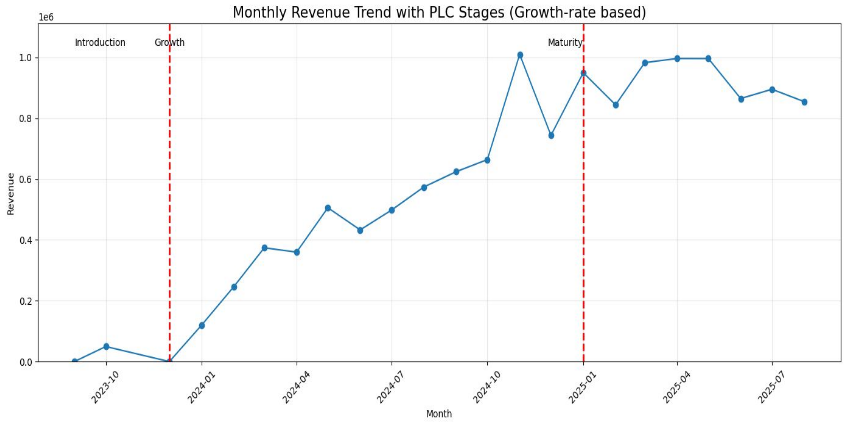
ECOMMERCE PLATFORM ANALYSIS

MỤC LỤC

EVALUATE REVENUE TREND



The revenue trend reflects a healthy growth trajectory for a young e-commerce marketplace. Starting from late 2023, GMV is initially low and volatile as the platform builds its seller base and acquires early buyers.



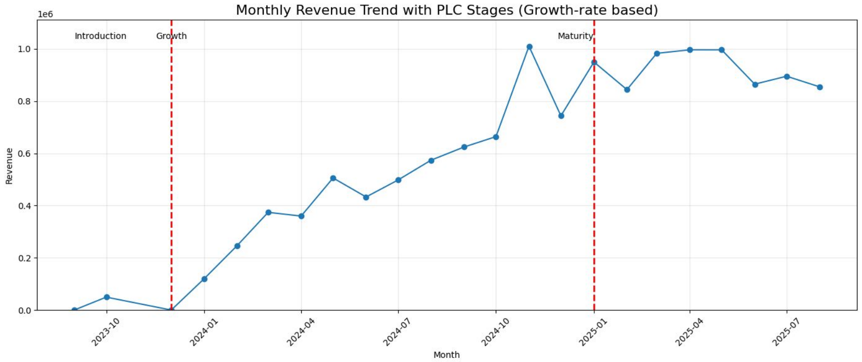
This dataset reflects the early development of a newly launched e-commerce marketplace, showing a pattern closely aligned with a typical product lifecycle.

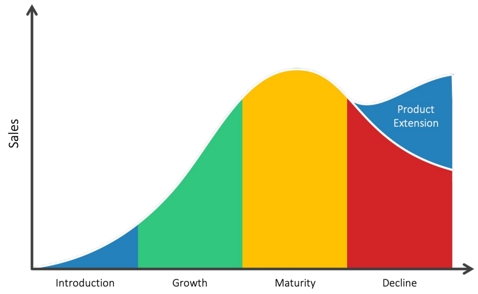
We can clearly observe **three of the four revenue stages—Introduction, Growth, and Maturity**. The team applied MoM growth thresholds of 5% and 15% to define the transitions between these stages.

Although a full e-commerce lifecycle often spans 5–9 years, the post-2023 environment—shaped by COVID-accelerated digital adoption—has driven faster scaling dynamics and effectively shortened the lifecycle timeline.

SEGMENT INTO DEVELOPMENT STAGE

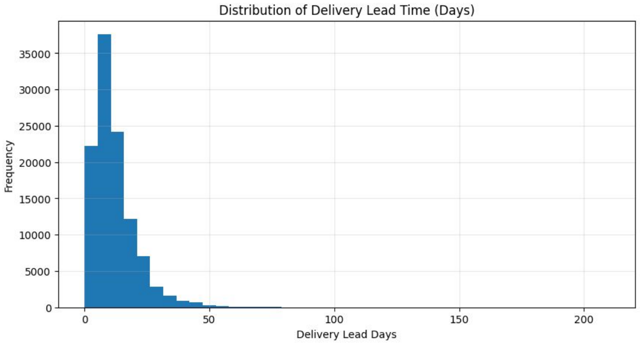
TOMORROW MARKETERS | DATA SCHOOL

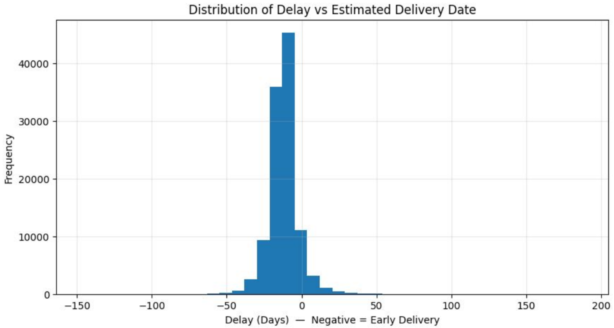


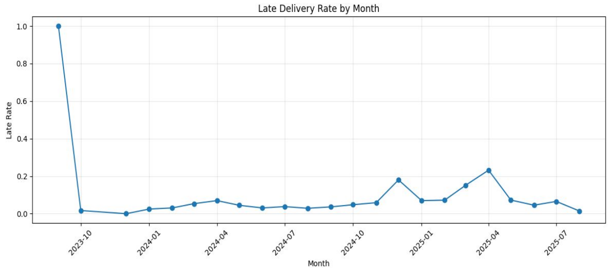


- **Introduction Stage**: Revenue remains low and highly volatile, reflecting the early formation of the marketplace. Limited seller participation, narrow product assortment, and low buyer traffic lead to inconsistent demand. This stage is significant as the platform tests core operations—logistics, payments, and onboarding processes—laying the foundational infrastructure required for future scaling.
- **Growth Stage**: Revenue accelerates rapidly and consistently, supported by strong MoM growth. Seller onboarding increases, assortment expands, and buyer acquisition strengthens, creating positive network effects. This stage is significant because it marks the platform's transition from experimentation to scaling, demonstrating product–market fit and validating commercial, operational, and marketing strategies.
- **Maturity Stage**: Revenue reaches its peak and stabilizes at a high level, with growth slowing to near-zero levels. Demand becomes more predictable, and seasonal peaks soften. This stage is significant because the platform shifts focus from hyper-growth to optimization—strengthening retention, improving unit economics, and enhancing operational efficiency to sustain long-term performance.

STRATEGIC PRIORITIES







1. Key Strategic Priority: Enhance delivery speed and tighten ETA accuracy to strengthen trust and early-stage customer experience.

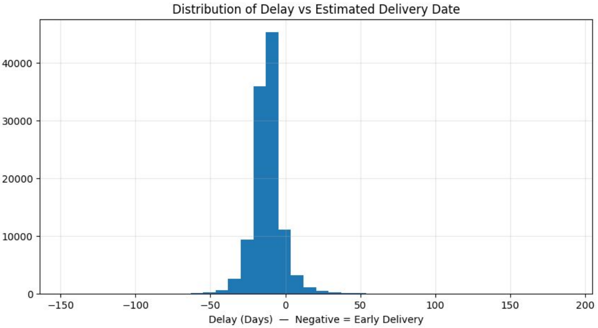
- Distribution of Delivery Last time: Lead time highly skewed with long-tail >40–60 days.
- Delay vs Estimated Delivery Date: Large volume of orders delivered earlier than ETA (negative delay).
- Late Delivery Rate by Month (Figure 3): Occasional spikes despite generally low rates.

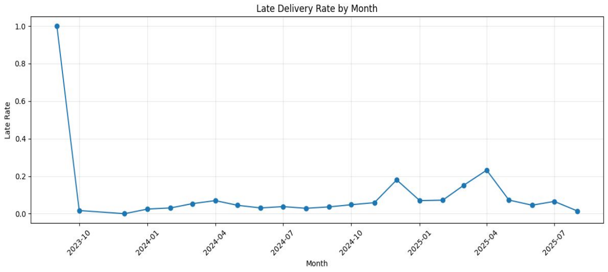
The fulfillment system is inconsistent: although most orders arrive earlier than expected, long-tail outliers inflate the perceived slow service. ETA is currently overestimated, reducing customer confidence.

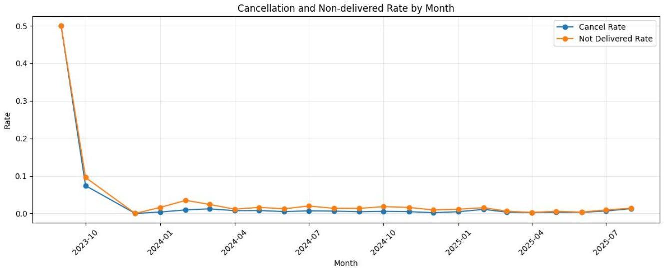
To fix this, the marketplace should:

1. Shorten ETA windows based on historical medians to increase perceived speed without operational risk.
2. Build a tiered SLA with main carriers and pilot urban micro-hubs in major Brazilian cities.
3. Deploy monthly carrier × region × category diagnostics to eliminate late-delivery spikes before scaling nationwide.

STRATEGIC PRIORITIES







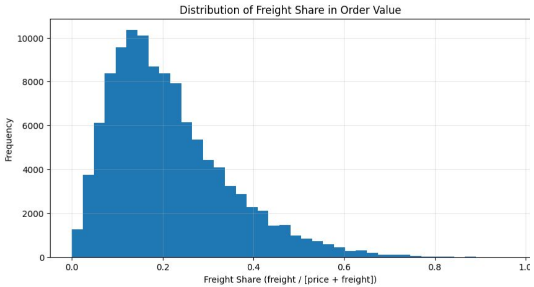
2, Key Strategic Priority: Strengthen reliability by reducing late, cancelled, and not-delivered orders to stabilize operational performance.

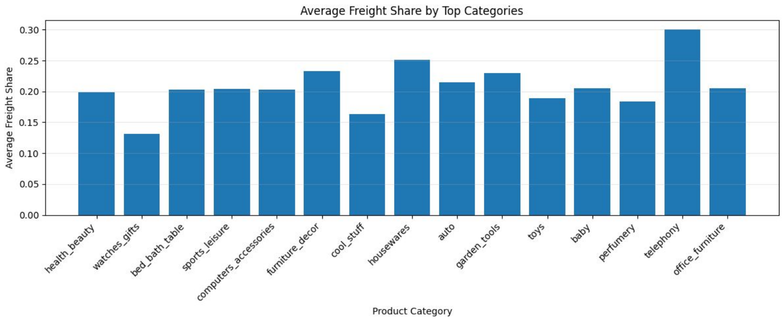
The supporting data shows clear risk points: the Late Delivery Rate by Month reveals spikes reaching 15–22% even though the baseline is low, while the Cancellation & Not-Delivered Rate consistently shows not-delivered orders higher than cancellations and occasionally exceeding 3–5%. The Delay Distribution also exposes extreme outliers, signaling address issues, routing failures, or carrier breakdowns.

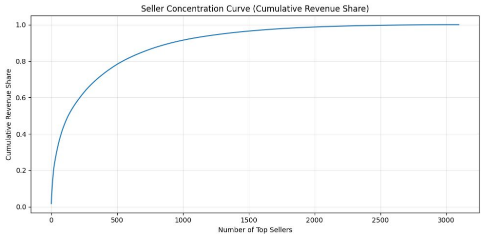
These reliability gaps directly damage trust, especially in the early lifecycle of a new Brazilian marketplace where user experience is still fragile.

To improve reliability, the marketplace should build diagnostic dashboards segmented by carrier × seller × category × region, enabling rapid root-cause identification. It should also enforce strict SLAs with penalties for repeated failures and provide first-order protection for new buyers to improve confidence and early retention

STRATEGIC PRIORITIES







3, Key Strategic Priority: Optimize unit economics through a category-based freight strategy to reduce cost pressure and improve overall marketplace margins.

The data highlights substantial freight imbalance: the Freight Share Distribution shows many orders with shipping cost equal to 15–25% of order value, with a heavy tail surpassing 50%. Average Freight Share by Category reveals that categories like telephony, housewares, and garden tools incur significantly higher logistics burden compared to watches_gifts or perfumery.

Combined with the Seller Concentration Curve where a small group of top sellers contributes most revenue, freight inefficiency can distort margins at scale.

To address this, the marketplace should implement category-specific logistics rules such as minimum basket thresholds, bundled-item incentives, or restrictions on low-margin freight-heavy listings. It should also negotiate carrier pricing based on size/weight classes and pilot regional stockpoints or localized hubs to reduce last-mile distance and cost for high-freight categories

FORECAST

TOMORROW MARKETERS | DATA SCHOOL

```
===== TỔNG HỢP KẾT QUẢ TEST (6-8/2025) =====
ARIMA        | Test_MAE=113890.18 | Test_MAPE=0.1112 | Overfit=False
RandomForest | Test_MAE=30451.79 | Test_MAPE=0.0302 | Overfit=False
LightGBM     | Test_MAE=322681.61 | Test_MAPE=0.3135 | Overfit=False
XGBoost      | Test_MAE=168929.23 | Test_MAPE=0.1639 | Overfit=True
```

To forecast the marketplace's revenue for Q3-2025, the team decided to use four prediction models. Since the data does not exhibit strong seasonal patterns, we selected ARIMA and three regression-based models—Random Forest, XGBoost, and LightGBM—to generate the forecasts.

In the forecasting model, team applied a structured sequence of data normalization and quality-control steps to ensure accuracy and prevent data distortion. First, the sales data was aggregated monthly, and additional key features were engineered, including order volume, number of buyers, number of sellers, average price, lag variables (lag1–lag3), rolling means, and growth rates.

The data was then split chronologically into training and testing sets to simulate a "forecast three months ahead" scenario. Team verified the absence of data leakage by using only past values when handling missing data, ensuring the model's integrity. Multiple models—RandomForest, LightGBM, XGBoost, and ARIMA—were trained and evaluated using MAE and MAPE, with overfitting checks performed by comparing errors between the training and testing sets. Finally, the model that did not overfit and achieved the lowest MAPE was selected to forecast revenue for June–August 2025.

FORECAST

Train/test Split → **Feature Selection** (Month, Orders, Active buyers/ sellers, Items sold, Average price, Revenue lags, Rolling averages) → **Random Forest** → **Evaluate by MAPE, MAE**

For Q3 2025 we forecasted monthly revenues using a Random Forest regression model trained on all historical data up to May 2025. The model uses time-based and business features (month, orders, active buyers/sellers, items sold, average price, and revenue lags/rolling averages) to predict revenue for June, July and August 2025; the sum of these three monthly predictions gives the platform's Q3 2025 revenue forecast.

To assess accuracy and reliability, we treated Q3 2025 as a **hold-out test set** and trained the model only on the months before it. On this Q3 test period, Random Forest achieved a **MAE of ≈30k** per month and a **MAPE of about 3%**, while a naïve benchmark that simply repeats the last pre-Q3 revenue has **MAE ≈121k** and **MAPE ≈11.9%**. We also checked for overfitting by comparing train vs test errors and ensured no data leakage by building all features only from past information and forward-filling missing months. This combination of low Q3 MAPE and clear improvement over the naïve model supports the robustness of the Q3 2025 revenue forecast.

FORECAST

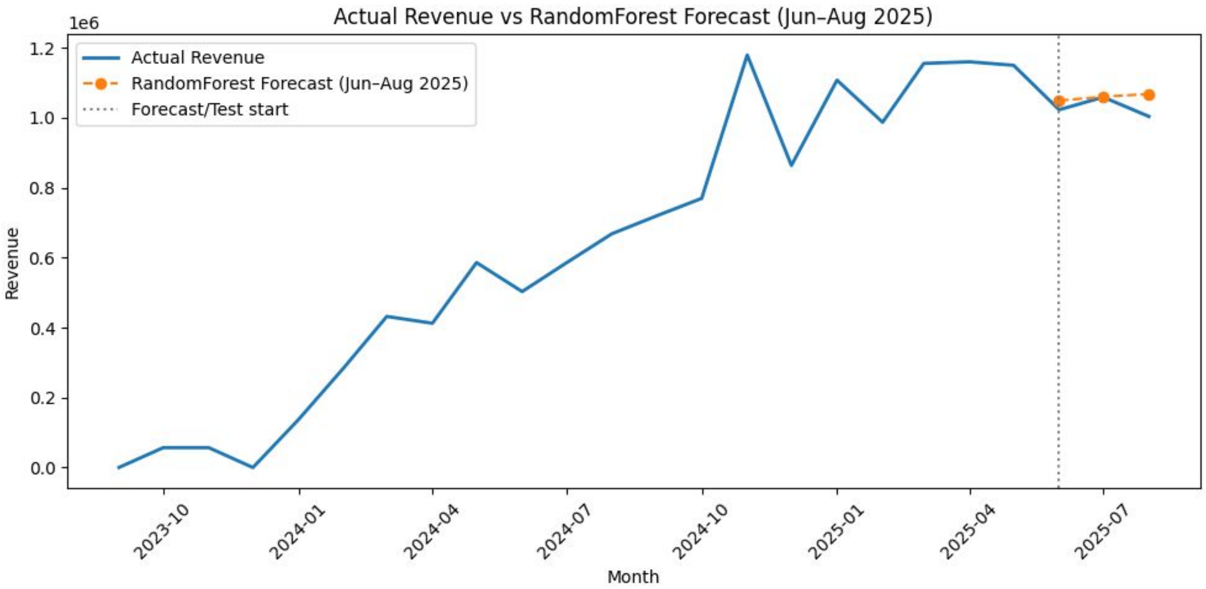
TOMORROW MARKETERS | DATA SCHOOL

===== RandomForest Evaluation Table =====

	Model	Dataset	Is_total	MAE	RMSE	MAPE(%)	R2
0	RandomForest	Train (before Q3 2025)	False	29089.814846	42427.350150	31883.866900	0.986085
1	RandomForest	Test (Jun-Aug 2025)	False	30451.789542	39941.036105	3.016293	-2.041024
2	RandomForest	Q3 2025 - Total revenue	True	91355.368625	NaN	2.961391	NaN
3	NaiveLastMonth	Test (Jun-Aug 2025)	False	121488.463333	123628.618638	11.869650	-28.135278

The evaluation results show that Random Forest performs strongly and clearly outperforms the naïve baseline. On the training set, the model achieves a low MAE of around 21k with an R^2 of 0.98, indicating that it captures historical patterns effectively. On the Q3 test period, Random Forest maintains solid generalization with a MAE of ~30k and a very low MAPE of about 3%, whereas the naïve approach produces significantly higher errors (MAE ~121k and MAPE ~11.9%). The total revenue error for the entire quarter is also only around 3%, which is acceptable for business forecasting. Overall, Random Forest demonstrates high accuracy and reliability, making it a suitable primary model for forecasting the platform’s Q3 2025 revenue.

FORECAST



The forecasted data shows a sideways trend, which reinforces the assumption that the marketplace is entering a maturity stage and is likely to maintain stable performance in the near future. However, it also indicates that without focusing on the identified strategic priorities, the platform may gradually slip into a decline phase. The prediction results highlight the importance of executing the key strategic priorities and continuously monitoring and developing the marketplace.