Boat Listing Conversion Propensity Model
Internal Summary
Prepared by: Eduardo Soto
Date: 03/01/2026

OBJECTIVE

The objective of this project was to simulate an internal machine learning workflow to predict the probability that a boat listing generates a qualified inquiry within a 7-day window.

This model is intended to support:

- Listing ranking optimization
- Sales prioritization
- Pricing review triggers
- Inventory performance monitoring

The project demonstrates how a propensity model could be integrated into a marine marketplace environment in order to improve operational efficiency and marketplace performance.

DATASET CONSTRUCTION

The modeling dataset was constructed using a warehouse-style SQL layer that combines:

- listings (core listing attributes)
- engagement_events (views, saves, inquiries)
- listing_photos (photo counts)

The engagement activity was aggregated into 7-day metrics at the listing level. The final dataset contains one row per listing with listing attributes, engagement signals, freshness indicators, and a binary conversion model.

Conversion was defined as at least one inquiry within a 7-day period.

FEATURE ENGINEERING

Several derived features were created to better the capture the dynamics of the marketplace:

- Boat Age (transformed from model year)
- Price per foot (relative pricing metric)
- Engagement rate ((saves + inquiries)/views)
- Log-transformed days on site (freshness signal)
- Dealer indicator
- One-hot encoded categorical values

These features were designed to show:

- Pricing competitiveness
- Buyer engagement intensity
- Listing quality
- Seller trust signals
- Inventory freshness

MODELING APPROACH

A Logistic Regression classifier was selected as an interpretable baseline model.

Reasoning:

- Produces calibrated probability outputs
- Easily interpretable coefficients
- Appropriate for binary classification
- Lightweight and prodcution-friendly

A preprocessing timeline was implemented in order to:

- Scale numeric values
- One-hot encode categorical values
- Maintain clean feature transformation flow

This dataset was split by using a stratified 80/20 train-test approach to preserve class balance.

The model's performance was evaluated by using:

- ROC-AUC
- Precision
- Recall
- F1 Score
- Confusion Matrix

KEY FINDINGS

The strongest drivers of conversion probability were:

- Engagement intensity (views, saves, inquiries)
- Seller rating
- Dealer affiliation
- Listing freshness (shorter time on website)
- Certain boat categories with stronger baseline demand

Listings that remain posted on the site for prolonged amounts of time show a declined likelihood of generating an inquiry.

These findings align exactly as expected with marketplace behavior, where buyer engagement and listing quality strongly influence conversion outcomes.

OPERATIONAL APPLICATIONS

The model output (the predicted probability) can be implemented in several different ways:

- Ranking listings in a dynamic way based on the predicted conversion likelihood
- Prioritizing the top decile of lisitings for sales outreach
- Triggering alerts for underperforming or stale inventory
- Informing pricing optimization strategies
- Scoring inbound leads to improve sales efficiency

Threshold selection should align with business objectives, such as:

- If sale resources are limited, prioritize precision.
- If the goal is ranking optimization, prioritize AUC and relative ordering.

LIMITATIONS

- Synthetic data was used for demonstration purposes
- Conversion definition may vary depending on production environments
- Logistic regression assumes linear log-odds relationships
- Engagement-based features may require careful handling to avoid temporal leakage
- Production deployment would require monitoring for data drift and recalibration

NEXT STEPS

- Evaluate tree-based models (e.g. Random Forest, Gradient Boosting)
- Incorporate time-decay engagement features
- Implement precision@k evaluation for ranking use cases
- Test model-driven ranking via A/B experimentation
- Establish monitoring for performance drift