Lead Insights Report:

Predicting Conversion & Accelerating Sales Cycles

Analysis by Ashley Sanders
Sanders Analytics | Business Data Strategy & Insights

This report presents the second phase of analytical work, focused on **lead engagement** and **deal velocity**.

In the first phase, I analyzed sales performance geographically and identified which customer profiles and market segments drive the most conversions. In this second phase, I've taken a **predictive approach** to understand which behaviors and attributes are most strongly associated with closed-won deals — and how quickly those deals close once initiated.

To accomplish this, I developed a machine learning–based **lead scoring model** using XGBoost and conducted two **time-to-event analyses**: a classical survival model and a machine learning–based Random Survival Forest model. Together, these methods help answer two key questions:

1. **Which contact characteristics are most predictive of conversion?**
2. **What factors influence how long it takes to close a deal?**

The report outlines my methods, findings, and practical recommendations for how the company can use these insights to prioritize leads, accelerate the sales cycle, and make more informed decisions moving forward.

I also highlight opportunities to operationalize these insights within your existing CRM workflows, including a roadmap for implementing and maintaining the lead scoring model inside HubSpot.

Sales Engagement Signals Matter

To identify which contact behaviors and attributes are most strongly associated with successful deal outcomes, I conducted a supervised machine learning analysis using a random forest classification model. The model was trained on thousands of historical contacts to predict conversion (i.e., closed-won deals) based on behavioral and demographic indicators.

Each predictor was assigned an importance score based on its contribution to the model's accuracy. The table below lists the top 20 conversion signals ranked by their predictive value.

Top 20 Variables	Importance_Score
lifecycle_stage_**Opportunities Shared**	355.0543
lifecycle_stage_unknown	30.2431
lifecycle_stage_**Quality Connection**	6.0129
engagement_score_raw	3.5005
lifecycle_stage_**Qualified Relationship**	2.902
role_unknown	1.5018
role_**Manager**	1.3127
number_of_form_submissions	1.1638
role_**VP**	1.0686
role_**Director**	1.0609
number_of_sessions	1.0478
marketing_emails_replied	0.936
marketing_emails_clicked	0.9322
marketing_emails_opened	0.9204
number_of_pageviews	0.904
role_**Lead**	0.6656
recent_sales_email_replied_TRUE	0.6405
role_Other	0.5373
lifecycle_stage_**Evangelist**	0.339
role_**Executive Director**	0.2663

Together, these findings provide a data-driven foundation for prioritizing leads and tailoring engagement strategies.

Key Findings

- ➔ Significance of Lifecycle Stage as a conversion signal
    - Stages such as "Opportunities Shared," "Quality Connection," and "Qualified Relationship" emerged as top predictors of closed-won deals.
    - These stages reflect strong engagement and meaningful progress through the funnel.

- **Interpretation:** Deals that reach these lifecycle milestones have a significantly higher probability of closing, making them powerful early indicators for prioritization

- Engagement actions matter, but not equally
  - Individual actions, including *form submissions*, *sessions*, and *marketing email opens/clicks* were moderately predictive.
  - This supports the idea that sales-readiness isn't signaled by one behavior alone but by **patterns of activity.**

- Role-level predictors (e.g., Director, VP, Manager)
  - These were relatively less predictive than lifecycle stages and engagement actions.
  - However, they may still offer valuable contextual segmentation for other analyses, such as time-to-close and prioritization strategies

---

Lead Scoring Model

To support faster, more consistent lead qualification, I developed a predictive lead scoring model that quantifies each contact's likelihood of converting based on engagement behavior and contact attributes.

Using your historical sales data, I trained and evaluated multiple machine learning models. The best-performing model — a tuned XGBoost classifier — achieved an AUC (Area Under the ROC Curve — a measure of model accuracy) of 0.905, indicating strong predictive power.

Each contact is assigned a lead score from 0 to 100, scaled from their predicted conversion probability. These scores are grouped into three tiers to help your sales team prioritize outreach and tailor engagement strategies:

Tier	Description	Quantiles	Meaning
High	Top 28% of leads	$\geq$ 72	very likely to convert
Medium	Middle 23%	$\geq$ 49	moderate opportunity
Low	Bottom 49%	< 49	minimal near-term potential

What It Means for Your Sales Team

This model allows your team to:

- **Prioritize High-Scoring Leads**
  Focus your efforts on contacts with the highest likelihood of converting.
- **Refine Nurture Strategies**
  Use lead scores to segment and target Medium- and Low-tier contacts with personalized follow-ups or drip campaigns.
- **Scale Consistently**
  Apply consistent qualification logic across all leads, reducing guesswork and manual triage.

Lead Score Performance in Historical Data

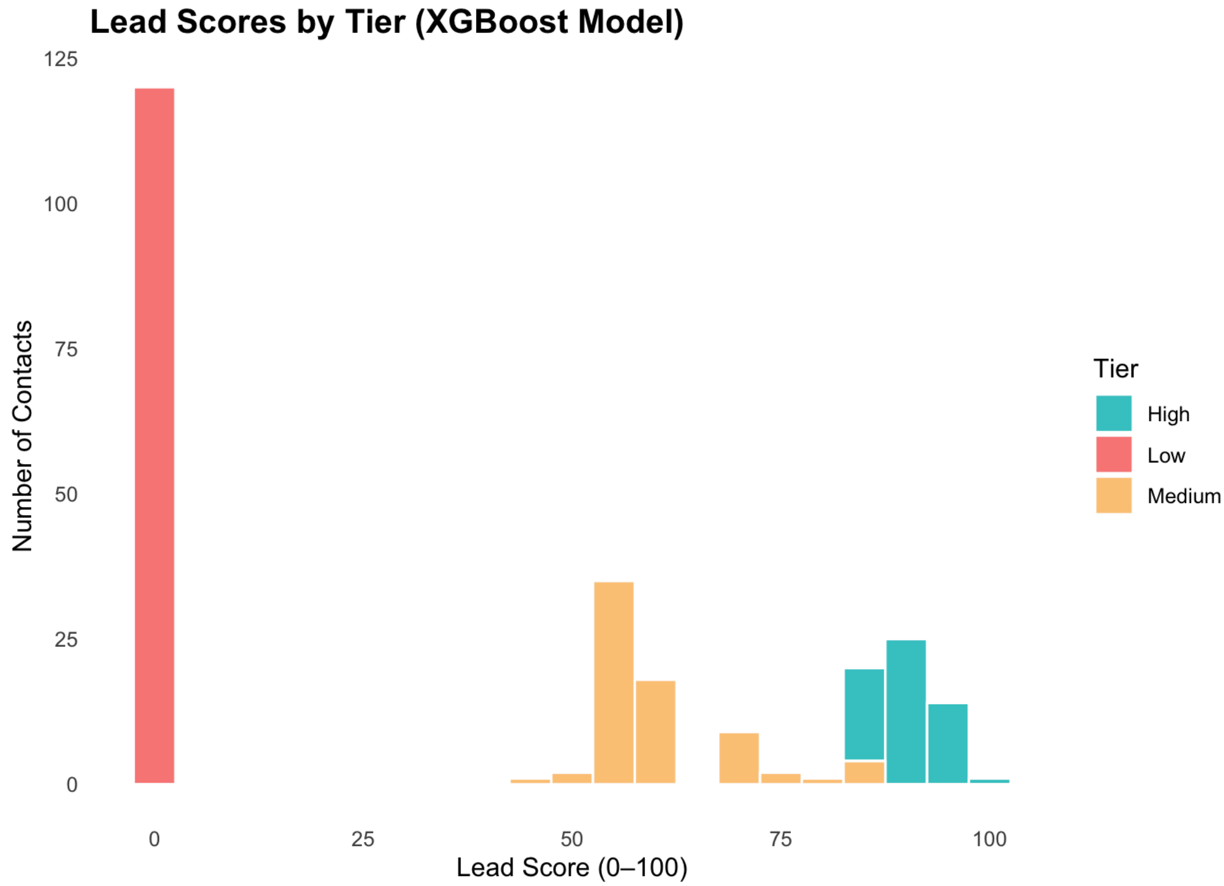
The chart and table below show how lead score tiers correlate with actual conversion outcomes in your historical data. The model cleanly separates contacts by their likelihood to close, meaning that it is highly effective at distinguishing strong leads from weak ones.

Conversion Outcomes by Score Tier

Lead_score_tier_xgb	total_contacts	closed_deals	conversion_rate
High	56	43	0.768
Medium	72	36	0.5
Low	120	0	0

Note: Tiers were based on historical conversion percentiles to align scores with actual observed performance.

The histogram below shows how the contacts are scored based on the XGBoost-predicted probabilities:



Using These Insights

- **Prioritize High Scores:** These are your best opportunities. Ensure timely and personalized outreach.
- **Nurture Medium Scores:** These leads have real potential. Consider automated follow-up sequences or product education campaigns.
- **Deprioritize Low Scores:** These contacts show low intent and may be better suited for longer-term drip marketing.
- **Monitor Trends Over Time:** We can revisit scoring thresholds quarterly to adapt to shifts in market behavior.

Questions to Consider:

- How many High-tier contacts can your team realistically follow up with each week or month?
- Would you prefer to tighten the High tier to only include the most likely deals, or broaden it to capture more potential?

Next Steps for Implementation

To operationalize this scoring system, I recommend:

- Integrating it into your **HubSpot workflows** using a field mapping + tier automation logic
- Establishing a **monthly review dashboard** to track changes and refine scoring thresholds over time
- Including additional data sources (e.g., campaign data, call notes) for ongoing model improvement.

If you would like, I can send a file with each contact's predicted probability, lead score (0–100), tier assignment, and supporting engagement fields for review and import into HubSpot.

We can discuss the possibility of providing on-demand lead scores. Or, if you would like to see the lead scores update as new clients enter the pipeline, we can explore building a custom pipeline for you. This would allow your team to continuously prioritize the most promising leads in real time.

Deal Velocity Analysis

Understanding what accelerates or delays conversion is key to shortening your sales cycle. To explore this, I used time-to-event modeling, also known as survival analysis, to identify which variables most influence the speed at which deals close.

Phase 1: Traditional Cox & AFT Models

In the first round of analysis using classical statistical models (Cox Proportional Hazards and Accelerated Failure Time), the only consistently significant predictor of faster deal closure was **sustained contact** — reinforcing a widely accepted sales principle.

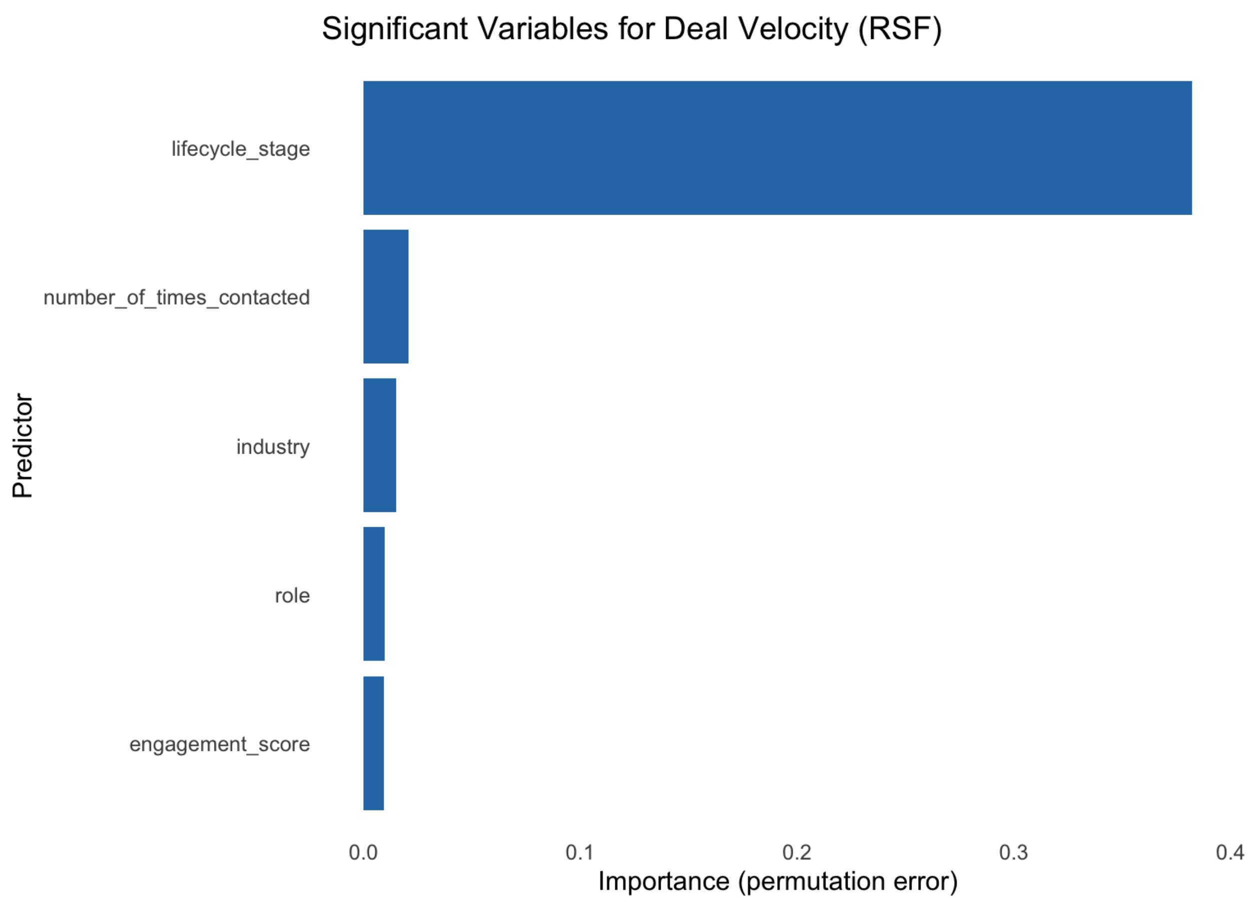
Other variables, including **role** and **industry**, showed inconclusive results, likely due to variability in data quality or insufficient sample sizes in certain categories. While the

engagement score was highly predictive of conversion, it did not significantly reduce time to close when modeled directly.

Phase 2: Random Survival Forest (RSF)

To better account for nonlinear effects and interactions between variables, I applied a **Random Survival Forest** (RSF) model — a more flexible machine learning method for time-to-event data and well-suited for complex sales processes.

Based on this analysis, the key drivers of deal velocity were Lifecycle Stage and Number of Times Contacted. The full results follow in the bar graph below.



Note: Additional variables, such as `job_domain` and `number_of_sales_activities`, were tested but excluded due to multicollinearity or data sparsity.

The engagement score included the following behavioral indicators:

- Website pageviews
- Session count
- Form submissions
- Marketing email engagement (opened, clicked, replied)
- Sales email replied
- Active sequence enrollment

To help translate the model results into practical guidance, the following table summarizes each variable's influence on deal velocity, along with what these findings mean for your sales strategy.

Variable	Model finding	Interpretation
lifecycle_stage	Strongest predictor of time-to-close	The stage of the relationship (e.g., "Quality Connection", "Qualified Relationship", etc.) is the most critical driver of deal velocity. **Focus efforts on accelerating movement between lifecycle stages.**
number_of_times_contacted	Moderate importance	Consistent outreach does reduce time-to-close. **Consider setting minimum contact targets for key segments.**
industry	Minor importance	Some industries may convert more quickly than others. This could help in prioritizing warm leads or adjusting messaging by industry.
role	Lower importance than expected	**Seniority or job function has a modest impact on deal velocity — might matter more for conversion likelihood than speed**.
engagement_score	Surprisingly low importance	Despite being predictive of conversion, engagement may not influence how fast a deal closes. **Don't over-prioritize engagement alone as a timing signal**

Recommendations

1. **Build playbooks by lifecycle stage:** Invest in tactics and automation that help move leads from "Quality Connection" to "Opportunities Shared" faster. This will have the

biggest payoff in reducing sales cycle time.

2. **Track contact frequency as a leading indicator:** Use the number and frequency of touchpoints as sales KPIs. Set activity targets by lead tier or stage to ensure high-priority contacts receive enough attention.

3. **Use industry as a segmentation layer:** Use industry as a secondary prioritization layer. Some industries may move faster and respond better to time-sensitive offers. *We could explore this question in a follow-up investigation.*

4. **Supplement Engagement Scoring:** While helpful for lead qualification, engagement score does not predict how quickly someone will convert. Always pair it with lifecycle stage insights for forecasting and follow-up planning.

The recommendations above highlight key insights and strategic priorities based on the data. To support implementation, the checklist below outlines **specific actions** your team can take to operationalize these findings and drive measurable improvements in deal velocity.

Implementation Checklist: Deal Velocity Insights

1. **Standardize and Audit Lifecycle Stage Usage**
    - Review lifecycle stage fields across HubSpot to ensure consistent labeling (e.g., “Qualified Relationship,” “Quality Connection”).
    - Train your team on how and when to update stages to improve data quality and interpretability.

2. **Add Lifecycle Stage Tracking to Sales Dashboards**
    - Include lifecycle stage as a breakdown dimension in deal velocity reports.
    - Track the average time spent in each stage to identify delays or drop-offs.

3. **Define Contact Frequency Benchmarks**
    - Set internal guidelines for minimum number of contact attempts by stage (e.g., 5 contacts within 10 days of entering “Quality Connection”).
    - Use task automation or sequences in HubSpot to support this.

4. **Monitor Time-to-Close by Industry**
    - Segment deals by industry and calculate average days to close.
    - Use this insight to prioritize fast-moving industries for quicker wins or experiment with faster CTA cycles.

5. **Supplement, Don’t Substitute, Engagement Scoring**
    - Continue using engagement scores for qualification and prioritization.
    - Pair engagement with lifecycle stage for a more accurate picture of deal timing.

6. **Tag & Monitor High-Priority Segments**

    - Create dynamic lists or filtered views in HubSpot for high-converting lifecycle stages.
    - Regularly review these lists to ensure timely follow-up and progression.

Optional Extensions

While this project has focused on priority analyses, there are a few additional areas we could explore if helpful — including the impact of **sales reps, pipeline stages**, or **campaign/first-touch sources** on deal outcomes and velocity. I’d be happy to support this as a paid extension at my standard rate.

Additionally, if you'd like to **operationalize the lead scoring system in HubSpot**, I can provide a scoped project to set up the technical and reporting components. This would include:

- A **data dictionary** for model input fields,

- **Automation logic** to assign and update lead score tiers in HubSpot,

- A **monthly dashboard** to monitor tier-based performance and recalibrate scores over time.
  If you decide to move forward with a **data pipeline**, I can automate these updates for you as well.