Data-Driven Forecasting, Market Segmentation, and Sales Prioritization

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

Introduction and Project ScopeIn this project, I analyzed sales data and focused on three critical areas:

1. Understanding engagement patterns across the sales funnel by role to strengthen outreach and conversion.

2. Developing a near-term sales forecasting model to support planning and resource allocation.

3. Identifying the company's most profitable market segments over the past three years to refine Ideal Customer Profiles (ICPs), inform lead scoring strategies, and identify look-alike markets.

Executive Summary

Role Engagement Analysis: CEOs are crucial in moving deals forward, while mid-level contacts show engagement drop-off that could be improved.

Near-Term Sales Forecast: A weighted predictive model estimates Q2 sales revenue, adjusting for lead quality and conversion probability.

Market Segmentation & ICPs: High-performing customer profiles were identified based on recent 3-year sales data, highlighting specific industries, company sizes, regions, and channels.

Predictive Lead Scoring: A new lead scoring rubric and machine learning model were developed to prioritize high-potential leads earlier in the pipeline.

Strategic recommendations include strengthening CEO engagement, reinforcing mid-level nurturing, targeting high-value segments, and operationalizing new lead scoring logic in HubSpot. By combining descriptive analytics, predictive modeling, and segmentation analysis, this report delivers actionable insights and strategic recommendations to drive smarter growth and prioritize high-value opportunities.

Role Analysis: Contact Engagement by Role

This section examines how different contact roles engage during the sales process. Understanding who moves deals forward — and where engagement falls off — can help the company prioritize the right stakeholders and design more effective outreach strategies.To better understand how engagement varies by role and to identify potential opportunities to strengthen nurturing and sales strategies, I conducted an in-depth analysis comparing full-funnel engagement to contacts tied to deals.
Specifically, this analysis sought to answer:

- Which roles are most engaged across the full funnel?
- Are certain roles overrepresented at the top of the funnel but underrepresented in actual deals?
- Where might there be opportunities to improve conversion through more targeted outreach?

To explore the questions outlined above, I calculated an **engagement score** for each contact based on key behaviors, including:

- Website sessions and pageviews
- Form submissions
- Marketing email opens, clicks, and replies
- Sales email replies
- Active sequence enrollment

I then grouped contacts by **role** and calculated mean engagement scores for each group, both across all engaged contacts (full funnel) and for those tied to open or closed deals. This allowed me to measure **relative drop-off** in engagement by role from initial contact to active sales opportunities.

To assess statistical significance, I used Fisher's exact test to compare role distribution between the full engaged population and the subset tied to deals. The results showed a statistically significant difference overall. A post-hoc pairwise analysis revealed that the only role with a significant change was **CEO**, which became more prevalent in the deals group — suggesting they are particularly influential in moving deals forward.

Key Findings

- **CEOs become more prominent deeper in the funnel**: While CEOs make up **32.7%** of all engaged contacts, they account for **41.2%** of contacts associated with deals — an **8.5-point increase**, which is **statistically significant**. This suggests they play a crucial role in advancing or approving deals.
- **Managers and Managing Directors show drop-off**: Although **Managers (12.7%)** and **Managing Directors (5.5%)** are actively engaged across the full contact base, their representation drops to **8.7%** and **2.4%** respectively in deal-related contacts — a combined **7-point decrease**. Although the decrease in representation among Managers and Managing Directors was **not statistically significant after correction**, the directional drop (from 18% to 11%) could signal a **strategically important gap** in mid-level engagement follow-through.
- **Other roles remain stable or are underrepresented**: Roles such as VPs and Directors do not show statistically significant differences, though some qualitative shifts are present.

Engagement Analysis by Role

Role	Engaged_n	percent_all_engaged	deals_n	percent_deals	percent_diff	percent_diff_label	p_value	p_adj
CEO	626	0.327406	214	0.41233141	0.084926	8.50%	0.000172	0.00275668
Manager	245	0.128138	46	0.08863198	-0.039506	-4.00%	0.007809	0.11714067
Managing Director	127	0.066423	19	0.03660886	-0.029814	-3.00%	0.008304	0.11714067
VP	32	0.016736	19	0.03660886	0.019872	2.00%	0.009162	0.11911145
Other	108	0.056485	20	0.03853565	-0.017950	-1.80%	0.094171	1
CFO	34	0.017782	5	0.00963391	-0.008149	-0.80%	0.237040	1
Lead	25	0.013075	11	0.02119461	0.008119	0.80%	0.217894	1
Executive Director	3	0.001569	5	0.00963391	0.008065	0.80%	0.014535	0.17441839
Director	93	0.048640	22	0.04238921	-0.006251	-0.60%	0.558682	1
Engineer	8	0.004184	0	0	-0.004184	-0.40%	0.213551	1
COO	26	0.013598	5	0.00963391	-0.003964	-0.40%	0.659171	1
CRO	4	0.002092	3	0.00578035	0.003688	0.40%	0.177313	1
Consultant	4	0.002092	0	0	-0.002092	-0.20%	0.582926	1
CMO	4	0.002092	2	0.00385356	0.001762	0.20%	0.616018	1
CTO	2	0.001046	0	0	-0.001046	-0.10%	1.000000	1
Intern	1	0.000523	0	0	-0.000523	-0.10%	1.000000	1

Implementation Strategies

- **Double down on CEO engagement:** CEOs are clearly influential in getting deals to the table. Strengthen executive-level nurturing with tailored outreach, high-value content, and invitations to strategic discussions or briefings.
- **Reinforce mid-level engagement for continuity**
  The falloff of Managers and Managing Directors suggests a potential disconnect between early engagement and deal progression. Consider:
    - Segmenting mid-level contacts for dedicated nurture tracks
    - Reframing messaging to highlight how your offering supports their strategic goals (not just executive priorities)
    - Equipping them with internal advocacy resources to escalate deals
- **Audit messaging and friction points by role:** Use this insight to explore why certain roles disengage — is it messaging relevance, decision authority, or timing? Qualitative review or interview follow-ups could help illuminate causes.
- **Prioritize relationship mapping in CRM:** Ensure that multiple roles are linked to deals in your CRM. Tag decision-makers, influencers, and blockers to better track how each role moves through the funnel and target accordingly.

Near-Term Sales Forecasting

Using historical sales data and lead scoring probabilities, I developed a weighted predictive model to forecast sales revenue for the next quarter. This section outlines the forecasting approach, model performance, and key takeaways for short-term planning. The model aims to anticipate revenue from currently open deals by incorporating historical patterns and predictive insights derived from engagement, sales activity, and lead scoring data.

Approach

I trained a weighted XGBoost regression model on three years of historical sales data. This time window was selected based on leadership input and a visual analysis of sales patterns, which indicated that more recent activity better reflects the company's current business model and market position.

The model was designed to predict the expected deal amount for each open opportunity. It leverages a wide range of features including:

- Sales velocity indicators (e.g., days to close)
- Engagement metrics (e.g., number of touches, email responses)
- Historical and behavioral lead scores
- Industry and regional context

To account for lead quality, we incorporated lead score tiers into the model's weighting scheme, prioritizing data from high-quality leads during training.

Model Performance

On a held-out test set, the final model achieved the following performance metrics:

- **Mean Absolute Error (MAE):** $24,277
- **Root Mean Squared Error (RMSE):** $36,533
- **R^2 (Explained Variance):** 0.342

While the R^2 indicates moderate explanatory power, the model's MAE and RMSE are consistent with expected variability in B2B deal amounts. The model effectively differentiates between high- and low-value opportunities, even if precise deal amounts remain challenging to predict.

Forecast Adjustments

Initial raw predictions tended to overestimate total revenue by assigning non-trivial amounts to many low-score leads. To address this, we adjusted forecasts using the model's own lead conversion probability estimates. This produced more conservative and realistic revenue projections, aligning more closely with known sales conversion dynamics.

Probability-weighted forecasts by month

Month	Predicted Revenue	High-Tier Leads	Deals (n) In queue	High-Tier Lead Share (%)
April 2025	$34,665	1	5	20
May 2025	$48,492	4	66	6.1
June 2025	$227	0	2	0

Close Month	Lead Score Tier	Industry	Deals (n)	Total Expected Sales
2025-04-01	High	Software	16	$34,665
2025-04-01	Low	Manufacturing	5	$7,300
2025-05-01	High	Retail	10	$18,000
2025-05-01	Low	Manufacturing	67	$48,492
2025-05-01	Medium	Software	12	$10,230

Market Segmentation and Ideal Customer Profiles

To complement forecasting efforts and support longer-term strategic planning, I analyzed historical and recent performance data across industry, geography, company size, channel, and behavioral factors to identify **high-performing market segments**, define **updated Ideal Customer Profiles (ICPs)**, and build a **predictive scoring system** to prioritize leads at the earliest stage of the pipeline.

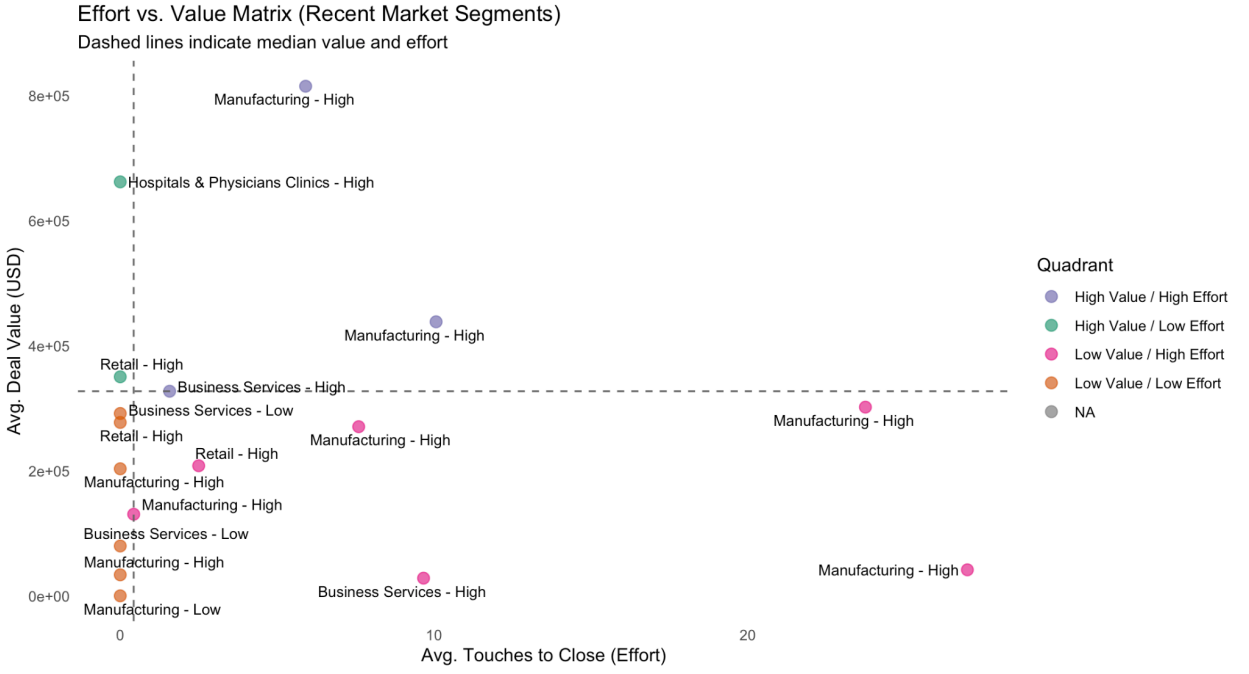
Approach

I analyzed all deals from the past 3 years, segmenting by industry, geography, company size, channel, lead score tier, sales engagement, and deal behavior. Each segment was evaluated on the following performance metrics:

- **Conversion Rate** (efficiency)
- **Average Deal Size** (value)
- **Sales Velocity** (speed to close)
- **ROI Index**: A composite metric of value per touch effort

I visualized segment-level tradeoffs using a 2x2 **Effort vs. Value Matrix** and identified top-performing combinations to guide profile development and scoring logic.

Click the image below for an interactive version:



Recent Top Performing Segment Insights

Here are the **top-performing segments** from the recent market summary (last 3 years), filtered by high conversion rate (≥30%) and fast-to-moderate sales velocity (≤45 days). All segments had high conversion rates, were all highly profitable for the company, and all came from **Offline sources**. This summary table covers deal dates from 1/1/2022 through 4/5/2025.

Industry	Region	Co Size	Co Revenue Tier	(n)	Total Sales	Avg Value	CVR	Avg Velocity	Avg Touches	Avg Touches Closed	Avg Value / Touch	Weighted Avg Value / Touch	ROI Index
Manufact.	North America	10-19	High	11	$1,435,200	$130,473	64%	24	0.27	0.43	$28,000	$28,000	65333
Manufact.	North America	250-499	High	18	$4,868,375	$270,465	83%	33	6.44	7.60	$406,018	$4,466,195	52502
Business Services	North America	10-19	Low	77	$25,192,885	$327,180	79%	3	1.52	1.57	$34,416	$344,164	17944
Manufact.	North America	20-49	Medium	7	$1,422,000	$203,143	71%	41	1.71	0.00	$8,500	$8,500	3542
Manufact.	North America	100-249	High	15	$12,224,400	$814,960	73%	7	4.33	5.91	$17,611	$52,833	2980
Manufact.	North America	20-49	High	23	$10,079,300	$438,230	65%	12	6.57	10.07	$4,245	$21,223	422
Business Services	North America	50-99	High	11	$310,804	$28,255	82%	19	9.00	9.67	$4,514	$27,083	410
Manufact.	Europe	100-249	High	7	$291,200	$41,600	57%	18	20.71	27.00	$2,018	$8,074	56

Here is the **full summary of segment performance over the past three years**.

Insights Based on Top Performing Segments

Metric	Insight	Recommendation
Industry	Manufacturing and Business Services dominate high-value, high-conversion segments.	Focus outreach and resource allocation in these industries.
Geography	All top segments are in North America.	Prioritize similar leads from this region; validate performance in other geos.
Company Size	Mid-sized companies (10–499 employees) consistently appear in top segments.	Target companies in this size range for growth and sales acceleration.
Channel	Offline Sources are the top-performing origin across all top segments.	Continue investing in offline efforts (e.g., events, partnerships, referrals).
Velocity	Sales cycles in top segments are short (avg. 2.9–41 days), allowing for quick conversions.	Use sales velocity as a qualifier for sales readiness.
Value per Touch	Some segments see values per touch exceeding $40,000 (e.g., Manufacturing, 250–499 employees).	Focus sales energy where touch-to-value returns are highest.
ROI Index	ROI Index ranges from 3,000–65,000+, indicating steeply increasing returns in targeted segments.	Rank and prioritize segments based on ROI for sales playbook development.
Emerging Opportunity	Hospitals & Clinics in N Am.: small N, but strong performance	Explore outreach tests in this vertical

Although **Hospitals and Clinics** don't appear in the table above, they convert at extraordinarily high rates (78%) with few touches (avg = 0) at high velocity (avg = 0 days) and one of the highest average values ($662k), making them an emerging high performing industry and may be worth further outreach.

Updated Ideal Customer Profiles (ICPs)

I created updated ICPs by aggregating recent top-performing segments across five key dimensions:

- **Industry**: Manufacturing, Business Services
- **Continent**: North America
- **Company Size**: 51–500 employees
- **Channel**: Offline sources, Organic search
- **Behavior**: High engagement, fast velocity (≤ 45 days), fewer touches

Each ICP was summarized in the linked Google Doc to support strategy discussions and campaign planning.

Lead Scoring Model Development

Building on the segmentation insights, I developed both a simplified pre-engagement scoring rubric and a predictive machine learning model. These tools will enable the company to prioritize new leads more efficiently and allocate sales resources based on conversion potential.

Simplified Lead Scoring Rubric

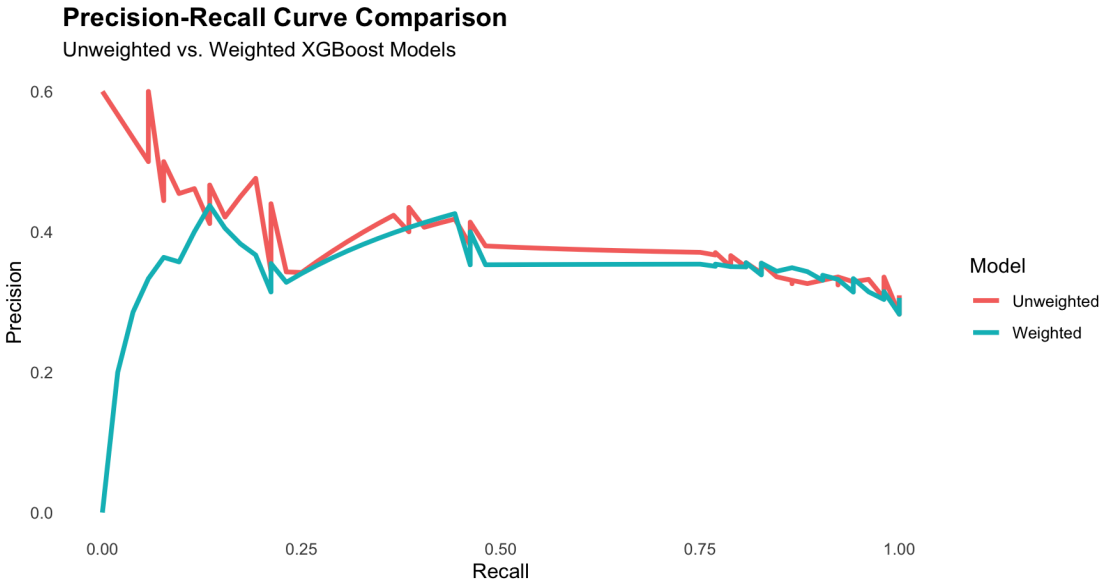
A rules-based scoring rubric was developed using pre-engagement features only (i.e., those available before a sales interaction occurs). Points are assigned for attributes like company size, traffic source, and industry alignment with top-performing segments. This can be implemented directly in HubSpot using custom properties and workflows.

Predictive Model (XGBoost Classifier)

To operationalize segmentation logic, I trained a machine learning model (XGBoost) to predict whether a new lead will convert based on pre-engagement characteristics. Model performance:

- **Accuracy**: 76.1%
- **Recall (Sensitivity)**: 26.9% (unweighted) → **improved to 92.3%** (weighted model)
- **Precision-Recall Curve**: Weighted model is more inclusive and effective at identifying likely converters.

Model Type	Strength	Use Case
Unweighted	High precision; conservative	Best when resources are limited and false positives are costly
Weighted	High recall; inclusive	Best when you want to capture all viable opportunities



How to interpret the precision-recall graph:

- **Precision** (y-axis) indicates how many of the leads predicted as “likely to convert” actually did convert.
- **Recall** (x-axis), also known as **sensitivity**, shows how many of the actual converters were correctly identified.

Recommendation: Use the **weighted model** to ensure no promising leads are missed, particularly in high-volume environments.

Implementation Options

Option	Description	Who It's For
CSV Scoring (Batch)	Export leads, score with R model, re-import to HubSpot	Fast to implement, no dev work
HubSpot Custom Scoring	Add lead score fields + rules in HubSpot UI	Great for small teams or operational staff
API Integration	Automate model scoring using HubSpot API + R or Python	Best for scaling or real-time scoring

How to operationalize the model

Step 1:Integrate ZoomInfo in your HubSpot CRM.

- **Immediate Action:** Implement the **ZoomInfo Inbound Enrich** integration to automatically enhance the quality of incoming lead data.
- **If you prefer Advanced Customization:** Utilize **Zapier** to create tailored workflows that align with specific business needs and processes.
- **Data Hygiene:** Regularly audit and clean existing HubSpot records to maintain data integrity, leveraging ZoomInfo's enrichment capabilities.

By integrating ZoomInfo with HubSpot, you ensure that you have accurate and comprehensive data, which is essential for effective segmentation and lead scoring. The model and rubric I've created both depend on information from ZoomInfo to function.

Step 2: Batch scoring via CSV export/import

This approach is manual but lightweight and doesn't require dev support.

Workflow:

1. **Add a custom field** in HubSpot like `ml_lead_score` or `ml_conversion_prob`.
2. **Export new leads from HubSpot** on a regular cadence (daily, weekly).
3. **Score the leads using your R model** (just load the exported CSV and run predictions).
4. **Append predicted scores** to the data as a new column (`predicted_conversion_prob` or `lead_score_ml`).

5. **Import the scored file back into HubSpot** using the Contact ID or Email as the matching key.
6. **Use workflows or views** to act on those scores.

Whether you use the machine learning model or the automated simple lead scoring model you can build right into HubSpot, you can:

- ➔ **Trigger workflows** in HubSpot based on score thresholds
- ➔ **Add score tiers** (e.g., High/Medium/Low) as custom properties
- ➔ **Create segmented lists** for Sales team prioritization
- ➔ **Combine model score with behavioral indicators** (e.g., email opens) for advanced prioritization

Conclusion and Strategic Next Steps

This analysis reveals clear patterns the company can leverage to accelerate revenue growth:

- Prioritizing executive (CEO-level) engagement is critical to winning deals.
- High-performing market segments cluster around mid-sized Manufacturing and Business Services companies in North America, with Hospitals and Physicians Clinics forming an emerging market in North America - all primarily sourced through offline channels.
- Updated ICPs and simplified scoring rules offer immediate improvements to lead qualification workflows.
- A predictive scoring model provides a scalable way to prioritize likely high-conversion leads early in the pipeline.

Next Steps

- **Implement ZoomInfo enrichment** in HubSpot to improve incoming lead data quality.
- **Deploy the simple lead scoring rubric** in HubSpot for faster prioritization.
- **Pilot the machine learning model** using CSV scoring while evaluating API or deeper automation integration.
- **Expand mid-level nurturing tracks** to reduce engagement falloff and support deal progression.

With these steps, the company can further sharpen its marketing and sales strategy, ensuring that resources are directed toward the highest-opportunity leads.