Sales Strategy Analytics Report: Symtrain

Alara Kaymak, Laura Li, Camden Bibro, Mary Morkos

Project Overview

This project aims to support sales strategy optimization at Symtrain, a startup specializing in AI-driven training and coaching solutions for contact center agents. Symtrain's platform delivers automated role-play scenarios, personalized AI feedback, and digital assessments to enhance agent performance at scale. Leveraging historical data across deals, companies, and customer support tickets, our team conducted both deal outcome analysis and customer segmentation to uncover patterns that could inform strategic decisions and improve win rates.

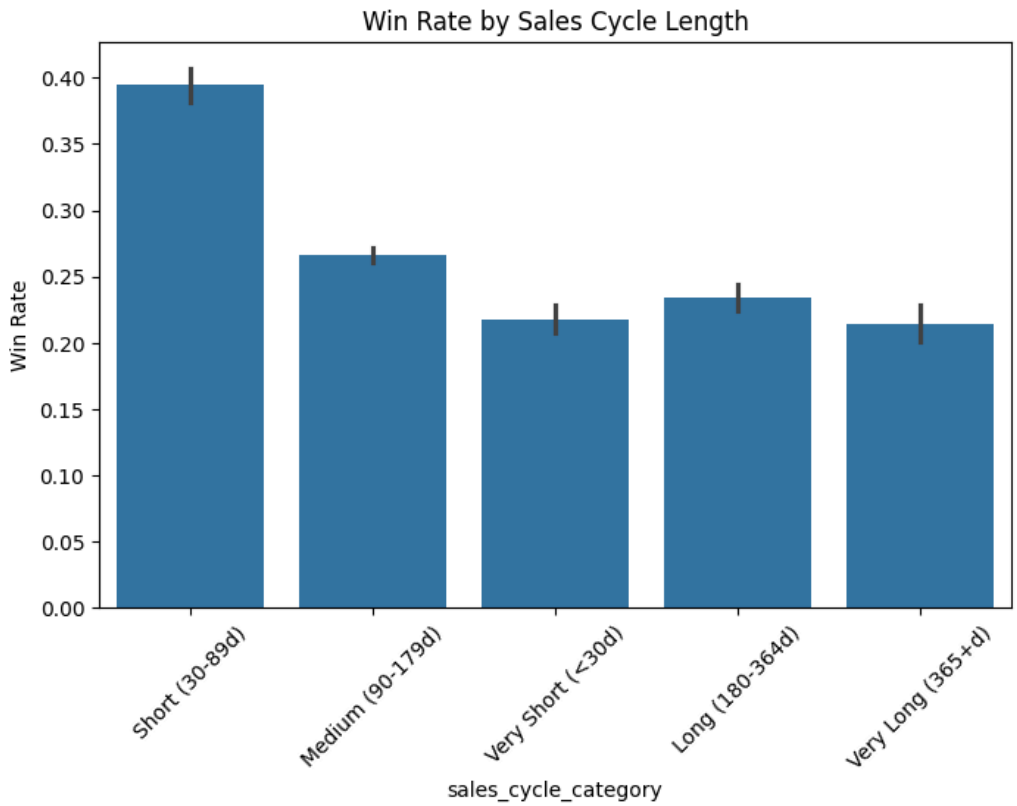
The analysis was conducted on three anonymized datasets provided by Symtrain:

- **anonymized_hubspot_companies.csv**: Customer firmographics and engagement metrics (19,000+ rows)
- **anonymized_hubspot_deals.csv**: Historical sales deals with outcome labels (~600 rows)
- **anonymized_hubspot_tickets.csv**: Support tickets data (79 rows)

Exploratory Data Analysis (EDA) of Deal Outcomes

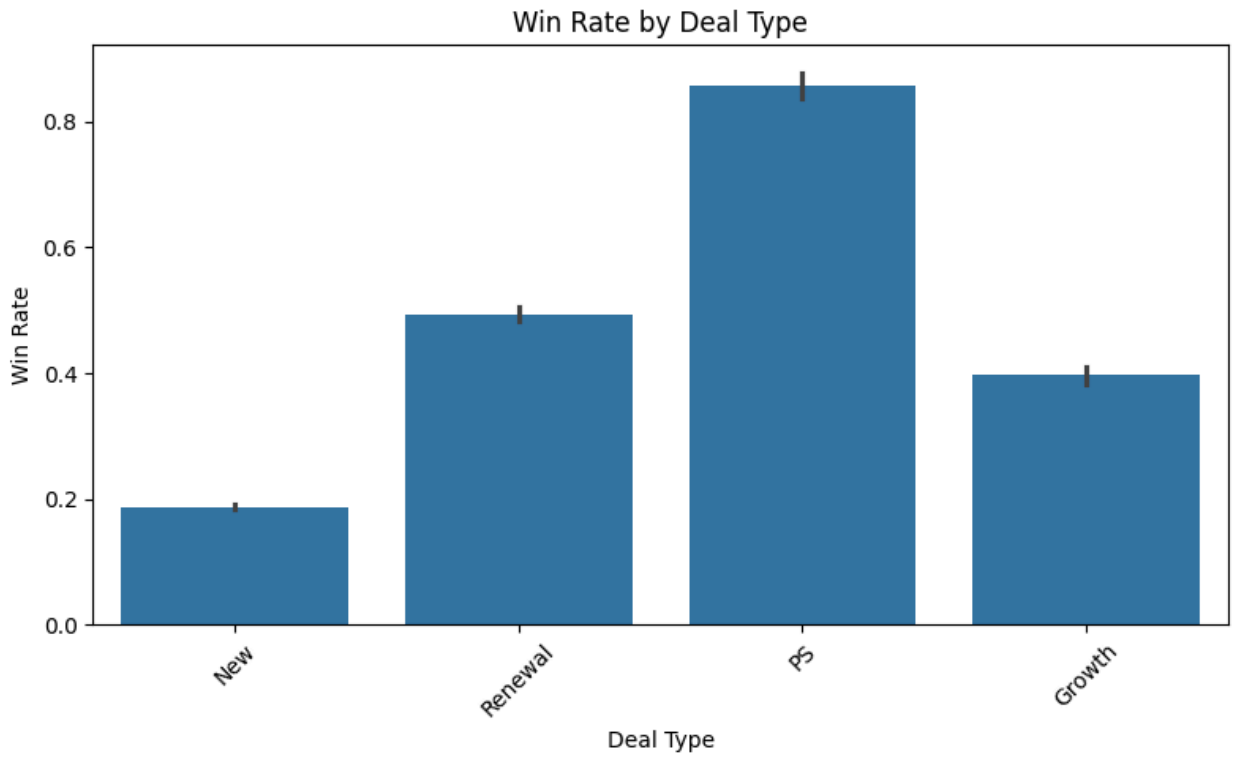
EDA on the deals dataset provided several critical insights into how win rates vary based on sales cycle length, deal type, deal size, and other attributes.

1. Impact of Sales Cycle Length



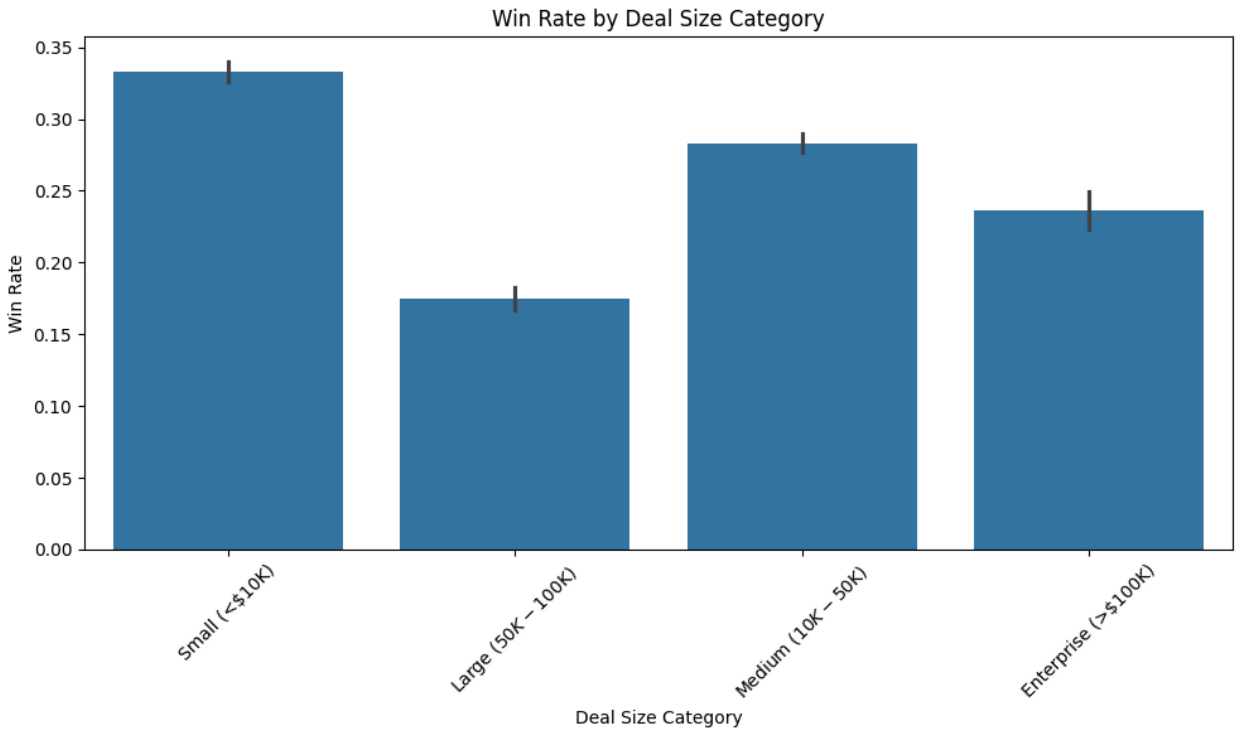
Deals closed within 30–90 days had the highest win rate, nearing 40%, whereas both very short (<30 days) and long (>180 days) cycles exhibited lower success rates. This suggests that an optimal "sweet spot" exists—long enough to build value but short enough to avoid losing momentum.

2. Effect of Deal Type



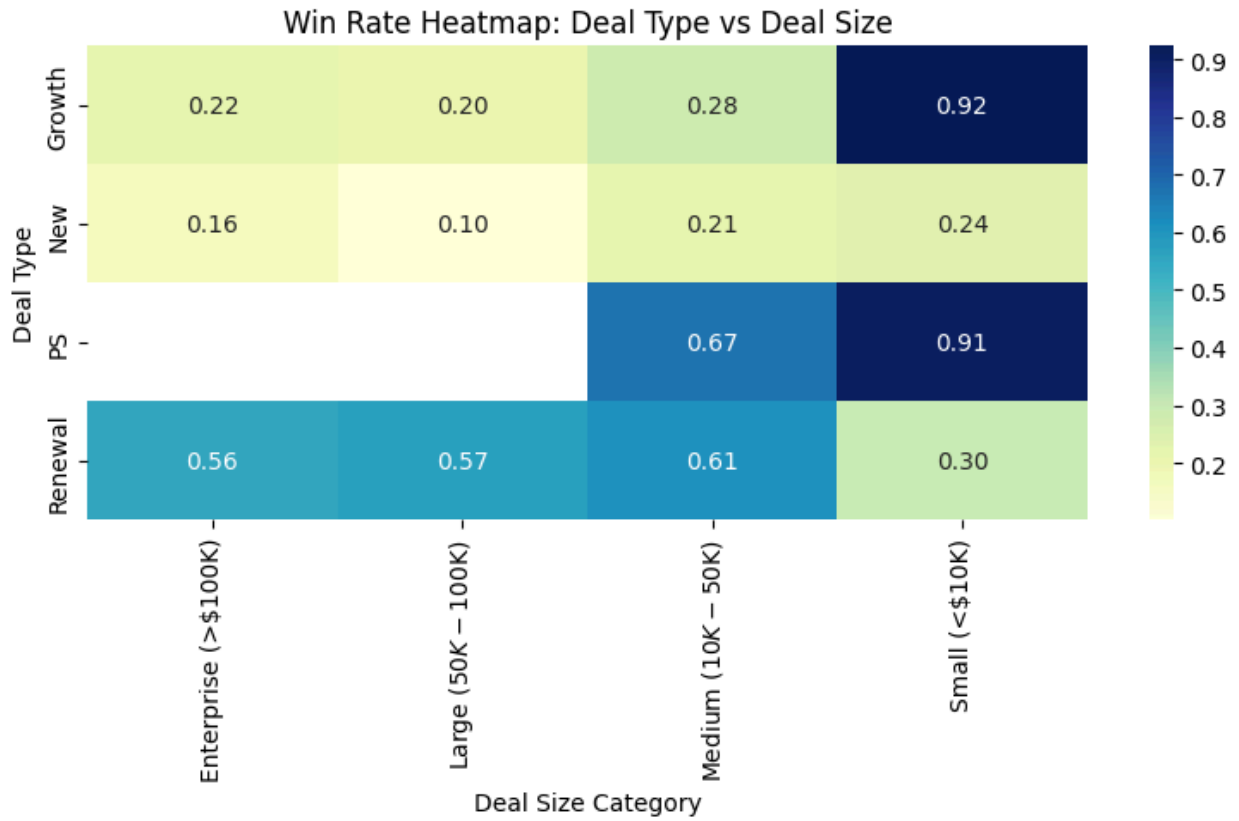
Win rates varied dramatically by deal type. Professional Services (PS) deals showed the highest win rate (~85%), followed by Renewals, while New deals had the lowest win rates (<20%). This suggests that new customer acquisition may require significantly more effort or different strategies compared to upsells or services.

3. Deal Size and Conversion Likelihood



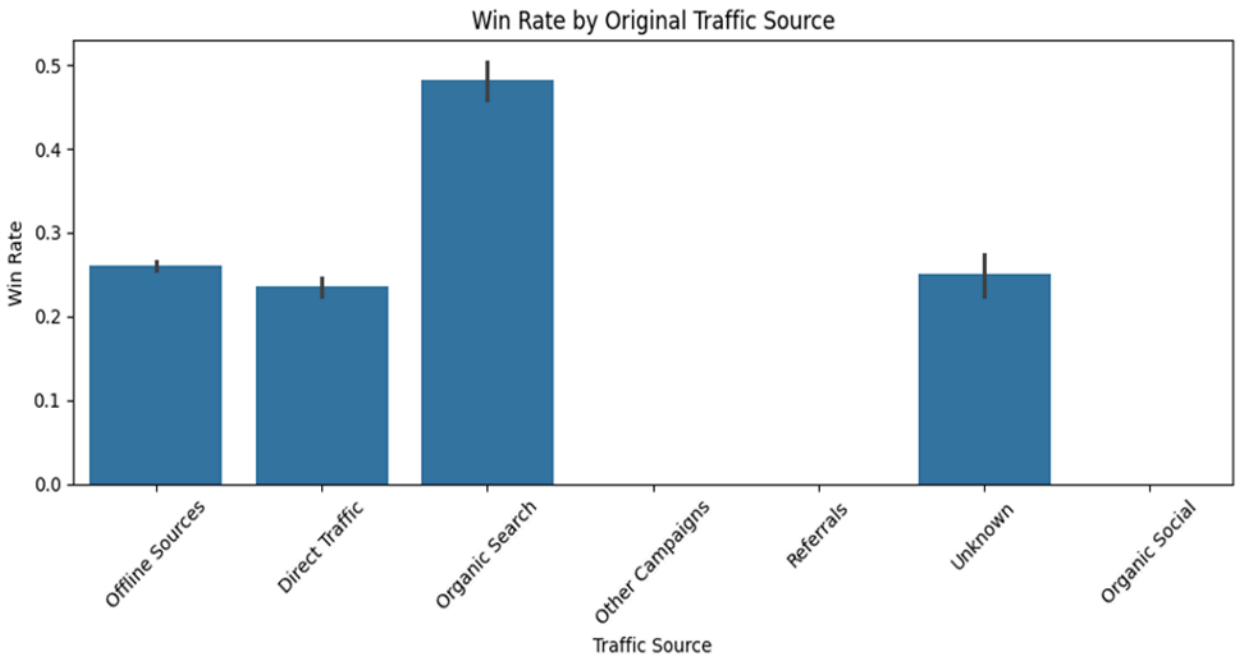
Interestingly, smaller deals (<$10K) had the highest win rates (~33%), while larger deals ($50K–100K) lagged behind (~18%). This could indicate more friction or competition in securing larger contracts and may highlight a need for dedicated resources or tailored strategies for enterprise-scale opportunities.

4. Deal Type vs. Size Interaction



PS deals consistently performed well across all sizes, even exceeding 90% win rates in small and medium segments. Meanwhile, Renewals had robust win rates in mid to large-sized deals. New deals, however, struggled across all sizes. This reinforces the need to further segment new customer efforts by deal size for targeted improvement.

5. Traffic Source Insight

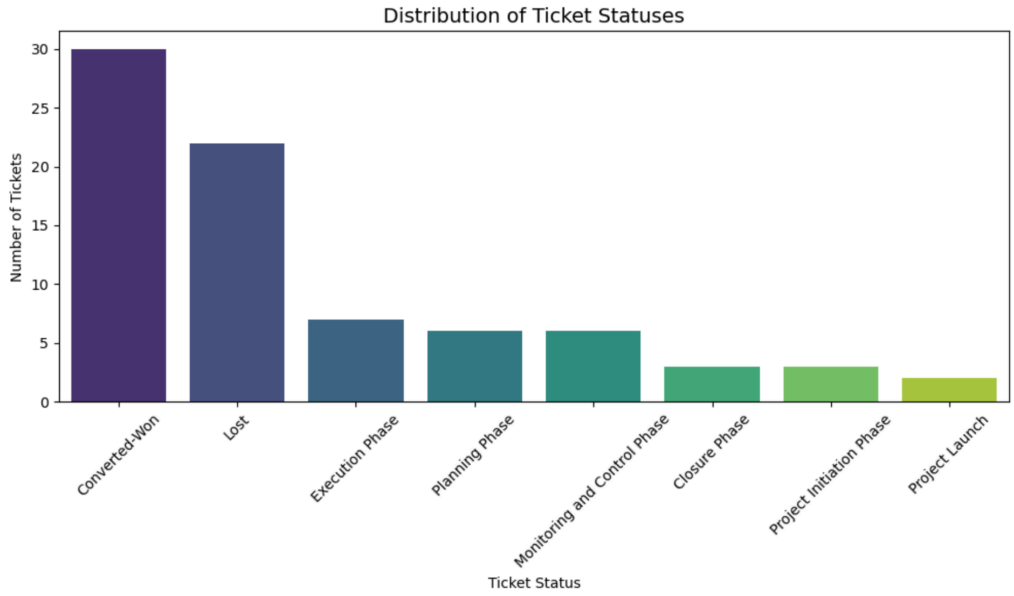


Leads from Organic Search had the highest win rates (~48%), while Direct Traffic and Offline Sources had substantially lower conversion. This finding emphasizes the importance of inbound digital marketing channels for deal success and potentially suggests reallocation of marketing efforts.

Analysis of Support Tickets

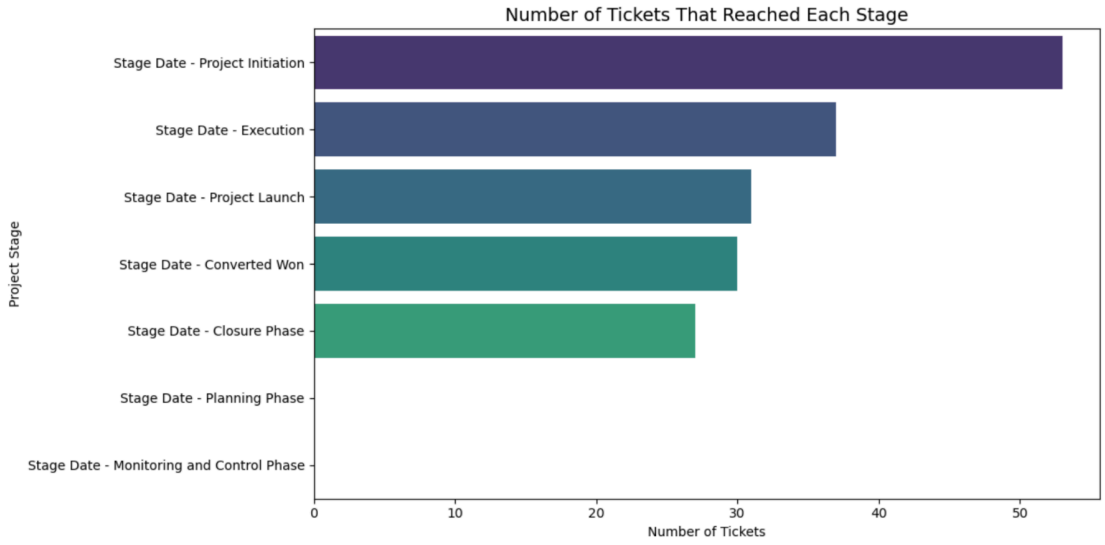
Although smaller in size, the Tickets dataset provided useful context on the lifecycle of customer interactions and where clients most frequently engage with Symtrain's support processes.

Ticket Status Distribution



The bar chart shows that the majority of tickets are either marked as Converted-Won or Lost, with fewer tickets progressing through intermediate project phases such as Execution, Planning, and Monitoring and Control. This pattern suggests that tickets are often being tracked not just for support resolution but as part of broader project or sales workflows.

Stage Completion Analysis



From the horizontal bar chart we see that the Project Initiation and Execution stages saw the highest number of tickets, indicating early-stage activity and implementation support are critical

touchpoints. Comparatively fewer tickets advanced into Monitoring and Control, suggesting a potential area for proactive support follow-up or client success initiatives.

Insights and Recommendations:

- Onboarding and implementation stages are the most common friction points, signaling a need for strong guidance materials and responsive support during these phases.
- Low transition into late project stages could reflect drop-off points that deserve further qualitative investigation (e.g., customer churn or failed onboarding).
- Mapping ticket progress against deal outcomes in future analyses could offer deeper insights into post-sale experience quality and customer success predictors.

Customer Segmentation via Clustering

For customer segmentation, we focused on the anonymized_hubspot_companies.csv dataset to uncover distinct groups of companies based on firmographic and behavioral features. The original dataset had over 19,000 rows and 46 columns. To ensure the quality and reliability of clustering results, we:

1. Dropped all columns with ≥50% missing values
2. Removed remaining rows with any missing values

This preprocessing left us with a refined dataset of 6,457 companies and 20 features. We chose to eliminate missing data instead of imputing it to avoid introducing artificial patterns into unsupervised learning.

Feature Selection and Engineering

We selected a combination of numeric and categorical features relevant to company scale and engagement:

- Financial: Annual Revenue, Number of Employees
- Engagement: Number of Form Submissions, Number of Sessions, Number of Pageviews, Number of times contacted
- Web technologies used: Parsed from the Web Technologies column, which initially contained **6257 unique combinations** of semicolon-separated platform names. We identified and one-hot encoded the **top 20 most frequent** platforms to retain meaningful signal while reducing dimensionality.
- Geographic & industry attributes: Time Zone, Country/Region, Primary Industry, Industry, and Consolidated Industry
- Temporal: Replaced Year Founded with an engineered feature Company Age calculated as the current year minus founding year

All categorical features were one-hot encoded, and numerical features were scaled between 0 and 1 using min-max scaling.

Clustering Model Selection

To identify meaningful segments, we compared three clustering algorithms:

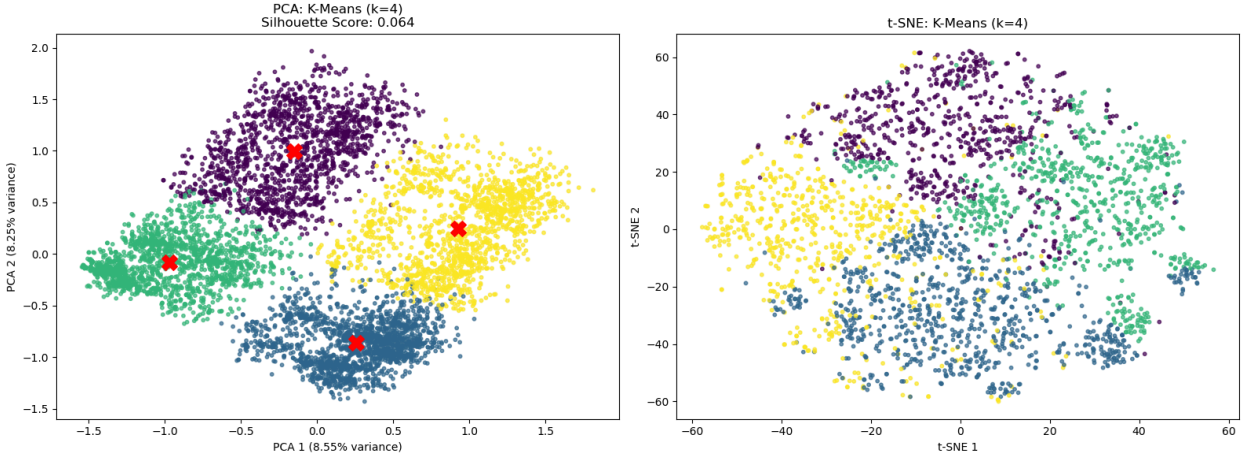
- K-Means
- Hierarchical Clustering
- DBSCAN

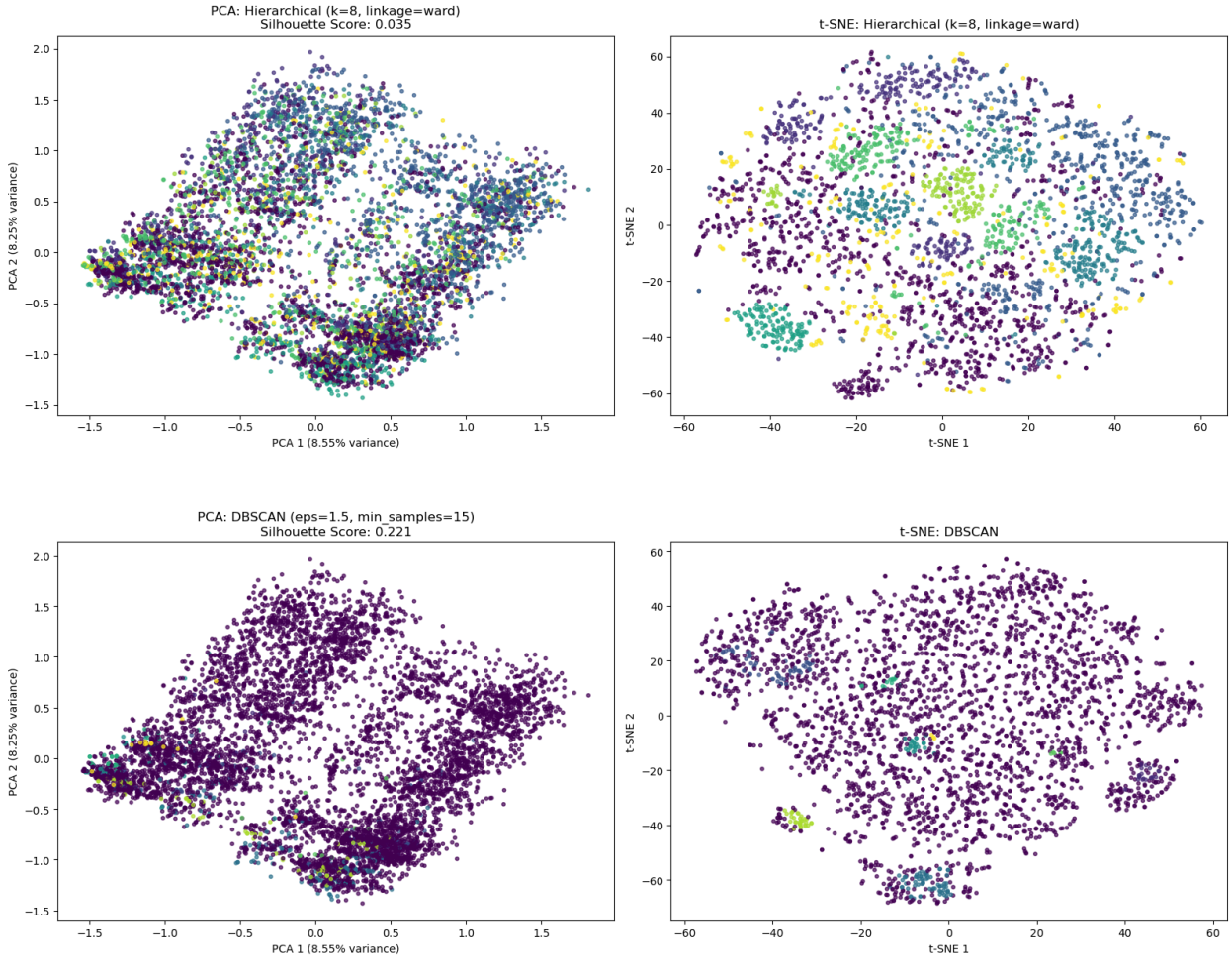
We tried to find the optimal hyperparameters for each method using:

- Silhouette scores
- Elbow plots
- Visual cluster separation using PCA and t-SNE projections

Below are the final clustering results using each clustering algorithm and the hyperparameters identified through the above methods:

- **K-Means (k=4)**: PCA and t-SNE plots showed clear separation; Silhouette Score = **0.064**
- **Hierarchical (k=8)**: Over-segmented and noisy clusters; Silhouette Score = **0.035**
- **DBSCAN**: Highest Silhouette Score = **0.221**, but resulted in highly unbalanced clusters, with one large cluster with majority of sample and rest classified as noise points





Despite DBSCAN's silhouette score, K-Means with 4 clusters was selected for its interpretability and visual coherence.

Interpretation of K-Means Clusters: Customer Segmentation Analysis

Cluster 0: "Low-Engagement Legacy Firms" (22.0% of companies)

Key Characteristics:

- Dramatically lower digital engagement metrics: **form submissions (-62.4%)**, **pageviews (-51.2%)**, and **session counts (-53.7%)**
- Significant technology sector presence (44.5% higher than average) with banking sector also well-represented (25.9% higher)
- Strong West Coast presence (America/Los_Angeles 37.1% higher than average)
- Mature companies: **76.2% are over 20 years old**
- Distinctively high adoption of marketing technology stack: **App Nexus (64.7% above average)**, **MediaMath (36.1%)**, and **Salesforce (28.3%)**

- Notably low adoption of Microsoft products compared to other clusters

Interpretation: Analysis of Cluster 0 reveals established technology and banking firms that exhibit a clear disconnect between their technology investments and digital engagement metrics. Despite having advanced marketing technology adoption (App Nexus, MediaMath, Salesforce), these companies show minimal interaction with digital content (-62.4% form submissions, -51.2% pageviews). This pattern strongly suggests these firms have implemented formal, structured procurement processes where initial research happens offline or through direct vendor relationships rather than self-service digital channels. Their sophisticated MarTech stack indicates they value data-driven decision making, yet their behavior demonstrates they prefer high-touch, relationship-based purchasing journeys over digital self-discovery.

Strategic Recommendations for Symtrain:

- **Account-Based Marketing Approach**: Develop highly personalized content demonstrating how Symtrain's AI simulation platform delivers quantifiable improvements in contact center KPIs
- **Executive Relationship Development**: Focus on C-level engagement (COO, CX, CTO) with ROI-focused messaging about agent attrition reduction and training acceleration
- **Salesforce Integration Emphasis**: Highlight Symtrain's native Salesforce integration capabilities, allowing contact center simulations to leverage existing customer data
- **Technical Implementation Roadmap**: Provide detailed integration documentation showing how Symtrain enhances their existing MarTech investments
- **West Coast Expansion Strategy**: Prioritize in-person demos and executive briefings in San Francisco/Los Angeles to showcase simulation capabilities in person

Cluster 1: "Microsoft-Centric Education & Healthcare" (29.5% of companies)

Key Characteristics:

- Minimal outreach receptiveness: **contact rates 53.8% below average**
- Smaller organizations: **employee count 35.2% below average**
- Strong educational and healthcare industry representation with **higher education at 10.3%**
- High geographical representation in **Canada (29.7% higher)** and **India**
- Higher proportion of younger firms: **6-10 year segment 31.7% above average**
- Overwhelmingly Microsoft-centric: **Microsoft Exchange Online (89.2% above average)**, **Outlook (85.6%)**, and **Office 365 (49.5%)**
- Significantly underutilizes marketing technology platforms (App Nexus -64.8%, MediaMath -36.6%)

Interpretation: Cluster 1 clearly identifies education and healthcare organizations deeply embedded in the Microsoft technology ecosystem, with adoption rates of Microsoft Exchange Online (89.2% above average), Outlook (85.6%), and Office 365 (49.5%) far exceeding other clusters. Their consistently low contact rates (53.8% below average) combined with smaller employee counts (35.2% below average) reflect the structured procurement practices typical in regulated institutions. The geographic concentration in Canada and India aligns with regions known for centralized purchasing in education and healthcare, where IT decisions often require multiple stakeholders and formal RFP processes. This cluster's resistance to outreach efforts isn't disinterest but rather institutional purchasing frameworks that favor established vendors and compliance-certified solutions.

Strategic Recommendations for Symtrain:

- **Microsoft Teams Integration**: Highlight Symtrain's compatibility with Microsoft Teams for collaborative agent training and coaching sessions
- **Compliance and Privacy Framework**: Emphasize how Symtrain's platform maintains HIPAA compliance for patient interaction training and FERPA compliance for student-facing roles
- **Patient/Student Experience Focus**: Develop specific simulation modules for healthcare patient interactions and university student service scenarios
- **RFP-Ready Documentation**: Create comprehensive security, compliance, and implementation documentation packages for formal procurement processes
- **Regulatory Alignment**: Develop Canada-specific training content that aligns with provincial healthcare regulations and multilingual requirements

Cluster 2: "High-Engagement Analytics Leaders" (25.9% of companies)

Key Characteristics:

- Exceptional digital engagement: **pageviews (89.1% above average)**, **form submissions (81.9%)**, **contact rate (70.3%)**
- Strong banking sector representation (34.8% higher than average)
- Balanced North American geographical distribution
- Moderate company age diversity with higher representation of growth-stage companies (6-10 years: 55.1% above average)
- Analytics-focused technology stack: **Google Analytics (24.3% above average)**, **Google Tag Manager (24.2%)**, **Wordpress (13.5%)**
- Notably low Microsoft ecosystem adoption (Exchange -43.5%, Outlook -41.0%)

Interpretation: Cluster 2 represents the most digitally active segment, with engagement metrics dramatically higher than all other clusters (pageviews +89.1%, form submissions +81.9%, contact rate +70.3%). Their technology profile shows strategic investment in analytics

infrastructure (Google Analytics, Google Tag Manager) rather than Microsoft enterprise tools, indicating a data-driven operational approach. The substantial banking sector representation (34.8% above average) combined with their aggressive digital research behavior points to financial institutions in active evaluation phases. The engagement patterns show classic buying signals - deep content consumption, multiple form submissions, and willingness to be contacted - indicating these companies have allocated budget and are comparing solutions within defined procurement timeframes. Unlike other clusters, these organizations demonstrate the "digital-first" due diligence typical of modern purchasing committees.

Strategic Recommendations for Symtrain:

- **Rapid Response Protocol**: Implement immediate demo scheduling for these high-intent accounts, with sales follow-up within 24 hours of engagement
- **Analytics-Driven ROI Modeling**: Emphasize Symtrain's robust analytics dashboard that tracks agent improvement metrics over time
- **Financial Services Training Library**: Showcase pre-built simulation scenarios specific to banking compliance, fraud detection, and financial advisory interactions
- **Performance Benchmarking**: Highlight Symtrain's ability to benchmark agent performance against industry standards specific to financial services
- **Data-Driven Success Stories**: Develop case studies specifically showing quantifiable improvements in banking contact center metrics (handle time, first-call resolution, compliance adherence)

Cluster 3: "Enterprise Legacy Organizations" (22.7% of companies)

Key Characteristics:

- Moderate underperformance in engagement metrics: **form submissions (-29.1%)**, **contacts (-26.8%)**
- Strong healthcare (11.6% higher) and consumer goods focus
- Highest proportion of mature companies: **82.8% are 20+ years old (10.4% above average)**
- Significant international representation: **UK (26.1% higher)** and **Canada (30.6% higher)**
- Diverse but comprehensive technology adoption: **App Nexus (88.7% above average)**, **Microsoft Exchange (85.2%)**, **Outlook (78.7%)**, **Salesforce (61.6%)**, **MediaMath (51.1%)**
- Highest adoption rates across multiple technology categories simultaneously

Interpretation: Cluster 3 consists predominantly of established enterprises (82.8% are 20+ years old) with highly diversified technology investments spanning multiple ecosystems simultaneously. Unlike other clusters that show preference for either Microsoft or marketing

platforms, these organizations maintain substantial investments across both domains (App Nexus +88.7%, Microsoft Exchange +85.2%, Outlook +78.7%, Salesforce +61.6%). This technology pattern reflects the layered IT architecture typical of organizations that have grown through acquisition and departmental independence. Their below-average engagement metrics despite comprehensive technology adoption indicates these companies maintain formal vendor management programs with established purchasing channels outside digital self-service. The strong international presence, particularly in regulated markets like the UK (+26.1%) and Canada (+30.6%), further supports their profile as complex enterprises with formalized procurement practices that prioritize compliance and integration with existing investments.

Strategic Recommendations for Symtrain:

- **Enterprise Scalability Focus**: Position Symtrain as an enterprise-grade solution capable of supporting multi-location, multi-brand contact center environments
- **Cross-Platform Integration**: Emphasize Symtrain's ability to integrate with diverse technology ecosystems including both Salesforce and Microsoft platforms simultaneously
- **Global Deployment Framework**: Develop implementation models for international contact centers including multi-language simulation capabilities
- **Change Management Program**: Offer comprehensive adoption services including train-the-trainer programs and change management resources
- **Legacy System Compatibility**: Showcase Symtrain's ability to work with older CRM and telephony systems common in established enterprises
- **Consumer Goods Scenarios**: Develop specific simulation modules for consumer product support, warranty claims, and product recall scenarios

Cross-Cluster Strategic Implications

Our cluster analysis reveals four distinct customer segments with important implications for Symtrain's sales strategies:

1. **Technology Ecosystem Integration** The data demonstrates that contact center technology stack is a stronger predictor of buying behavior than traditional firmographic measures. Symtrain should develop distinct integration approaches for Microsoft-centric organizations (Clusters 1 and 3) versus analytics-focused companies (Cluster 2). This includes creating technical documentation, API connections, and implementation guides specific to each technology ecosystem to streamline adoption.
2. **Engagement-Based Sales Prioritization** Contrary to conventional assumptions, digital engagement correlates most strongly with technology preferences rather than company size or industry vertical. Cluster 2's dramatically higher engagement metrics (+89.1% pageviews, +81.9% form submissions) indicate these organizations are approximately 3-4× more likely to be in active buying cycles for contact center solutions. Symtrain

should revise its lead scoring algorithm to prioritize accounts showing these engagement patterns for immediate sales follow-up.

3. **Vertical-Specific Simulation Libraries** The cluster analysis reveals distinct industry concentrations that should inform Symtrain's content development roadmap. Development resources should be allocated to creating industry-specific simulation libraries with pre-built scenarios for banking compliance (Cluster 2), healthcare patient interactions (Clusters 1 and 3), and technology support (Cluster 0). These vertical-specific assets will significantly reduce implementation time and increase adoption rates.
4. **Procurement Process Adaptation** Company age and technology diversity correlate strongly with procurement complexity. Established organizations in Cluster 3 require comprehensive RFP documentation, security certifications, and enterprise-grade support capabilities, while more agile companies in Cluster 2 respond to performance metrics and competitive differentiation. Symtrain should develop distinct sales enablement resources for each cluster's typical procurement process to accelerate sales cycles.

These findings should guide Symtrain's product development roadmap, marketing campaign structure, sales team organization, and customer success methodology. By aligning go-to-market activities with these distinct customer segments, Symtrain can increase conversion rates, accelerate sales cycles, and improve implementation success across its diverse customer base.

Deal Outcome Prediction

We approached deal outcome prediction as a binary classification task, with Is Closed Won as the target variable. We joined the deals and companies datasets, yielding 593 rows. Feature selection was guided by mutual information and logistic regression tests.

Feature Engineering

We generated several features:

- Company_Age: replacing Year Founded
- Contact_Frequency: # Contacts / Days to Close
- Revenue_per_Employee: normalized revenue
- Submission_Conversion_Rate: form submission rate
- Page_Depth: # Pageviews / # Sessions

We handled missing values via:

- Categoricals: mode fill or custom placeholders (e.g., "Unknown")
- Numericals: KNN imputation

High-cardinality categoricals used frequency encoding to control dimensionality. Final feature matrix had shape (593, 63).

We balanced the data using SMOTE to address class imbalance (True: 126 vs False: 348).

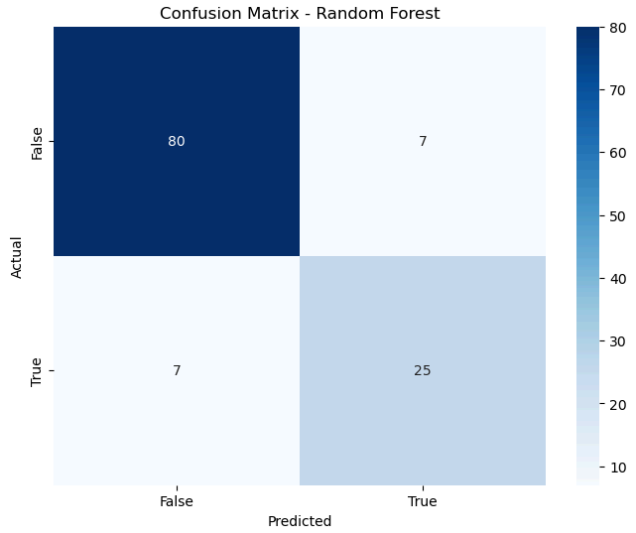
Model Evaluation

We evaluated:

- Logistic Regression: ROC AUC = 0.78
- Random Forest: ROC AUC = 0.90
- XGBoost: ROC AUC = 0.91

While XGBoost slightly outperformed, we chose Random Forest for its interpretability.

Confusion Matrix (Random Forest):

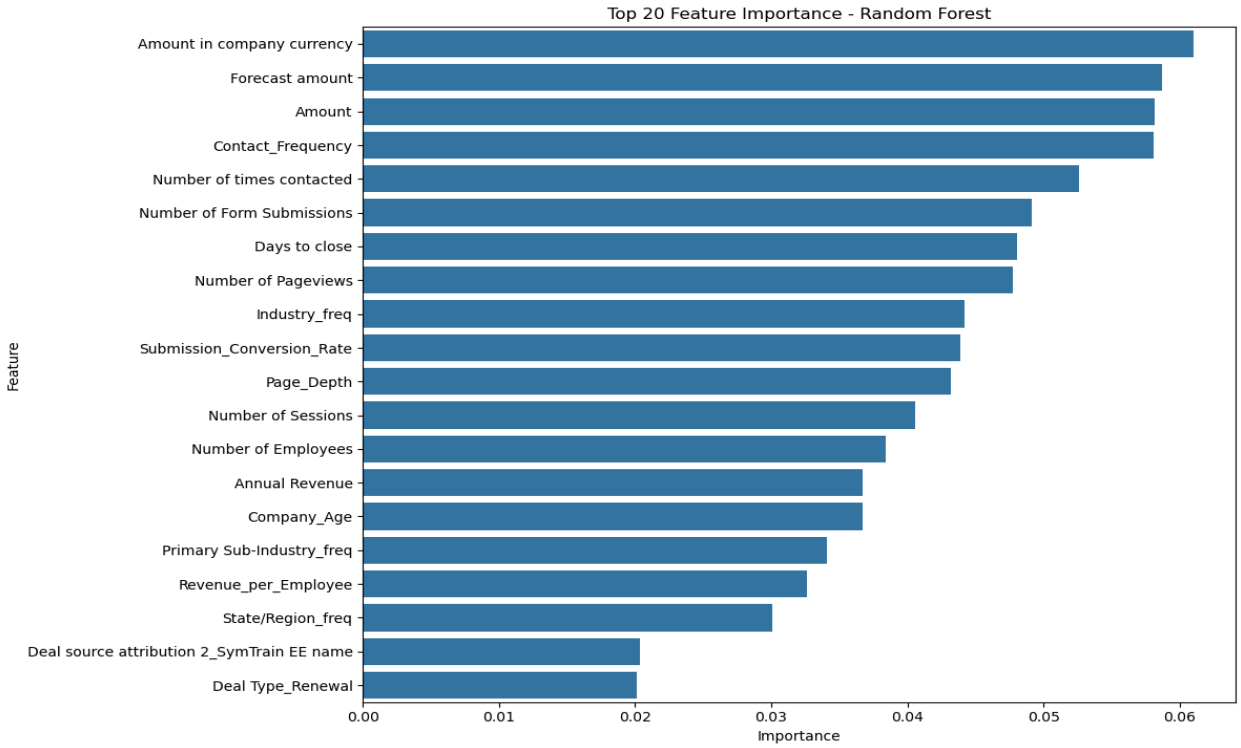


- True Positive: 25, True Negative: 80
- Balanced performance across classes with high precision/recall

Model Performance

Our final outcome prediction model used a Random Forest algorithm. It achieved an accuracy of 88.2% and an ROC AUC score of 0.8955, indicating strong ability to distinguish between won and lost deals. As you can see from above, the confusion matrix shows the model correctly predicted 80 losses and 25 wins, with only 7 false positives and 7 false negatives. This reflects a balanced error profile and reinforces the model's reliability. Notably, it remains conservative in its predictions, avoiding overconfident classifications unless supported by clear signals.

Top Feature Interpretations



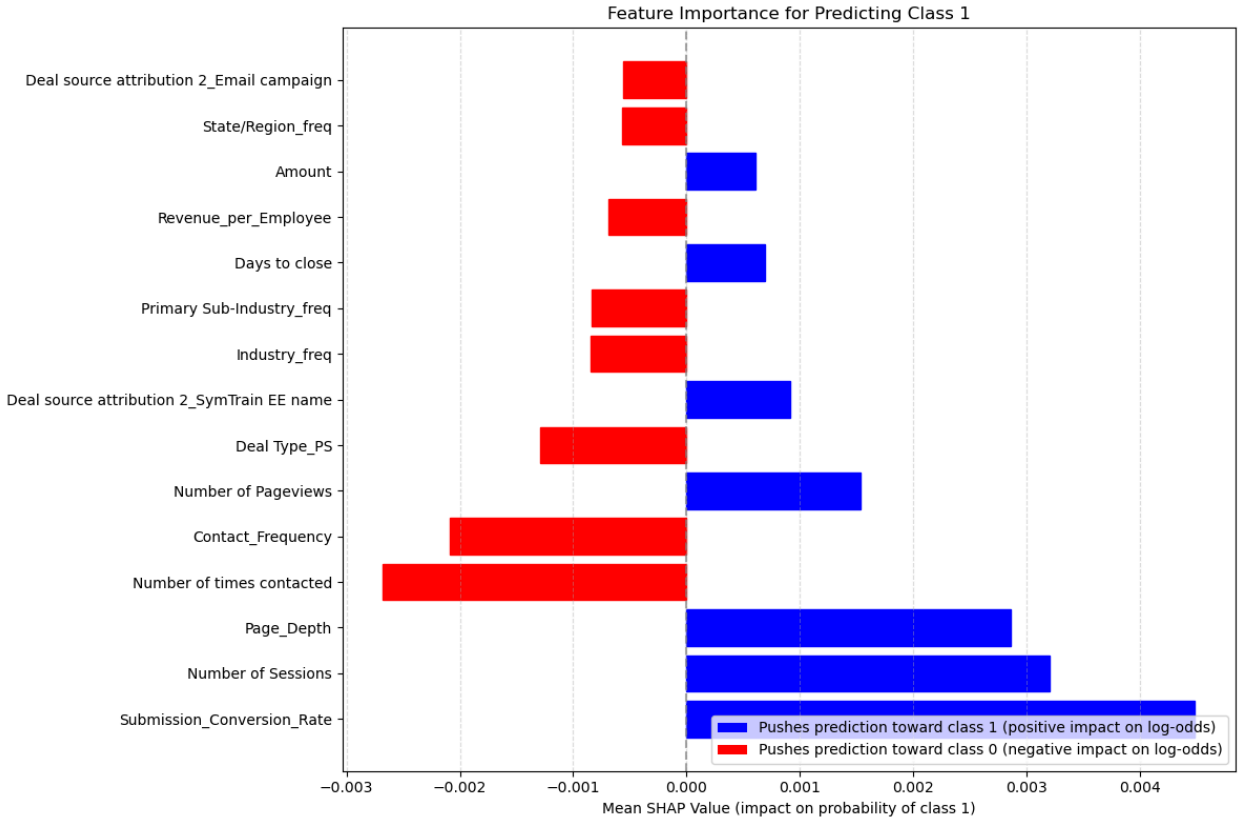
Our random forest model reveals a clear hierarchy of predictive factors driving outcomes in Symtrain's sales process. The analysis demonstrates that financial metrics stand as the strongest indicators, with monetary values dominating the importance rankings. This financial triumvirate accounts for a substantial portion of the model's predictive power, underscoring that deal economics fundamentally shapes success probability.

Following closely are engagement-based metrics, with contact frequency and number of times contacted showing substantial influence. These customer interaction measures, combined with form submissions and pageviews, collectively demonstrate that prospect engagement patterns are nearly as important as financial characteristics in determining outcomes. The model shows that how prospects interact with Symtrain is almost as predictive as what they're willing to invest.

Interestingly, days to close emerges as a top-tier predictor, suggesting that velocity metrics provide valuable signals about deal health. The prominence of submission conversion rate and page depth further reinforces that quality of engagement (depth of interaction) outweighs quantity of engagement in predictive power.

Traditional firmographic data points like industry frequency, number of employees, and company age contribute meaningful predictive value but rank lower than behavioral and financial signals.

This pattern challenges conventional wisdom that company characteristics are primary determinants of sales success.



SHAP Value Analysis: Beyond Importance to Impact Direction

While feature importance identifies what matters, SHAP values reveal *how* these features influence outcomes. Our SHAP analysis uncovers several counterintuitive relationships:

Positive Impact Factors (Pushing Toward Wins):

- Submission conversion rate emerges as the strongest positive driver, with significantly more impact than any financial metric. This indicates that prospects who convert from browsers to engaged leads are more likely to close.
- Number of sessions and page depth demonstrate that sustained, in-depth engagement with Symtrain's digital content substantially increases win probability.
- Number of pageviews confirms that broader content consumption correlates with higher close rates.
- Deal source attribution via Symtrain enterprise executives reveals that deals attributed to specific executives have higher conversion rates, suggesting the importance of executive relationships.

Negative Impact Factors (Pushing Toward Losses):

- Number of times contacted shows a surprising negative relationship, indicating that excessive outreach attempts correlate with diminished success. This challenges the "more touchpoints is better" sales philosophy.
- Contact frequency reinforces that higher frequency of contact doesn't necessarily improve outcomes, suggesting diminishing or negative returns on excessive outreach.
- Certain deal types indicate lower close rates than product-focused opportunities.
- Industry and sub-industry frequencies show that certain industries present systematically lower close rates, requiring targeted strategy adjustments.

This bidirectional analysis reveals that contact center prospects showing high-quality digital engagement patterns (conversion, depth, multiple sessions) but moderate outreach receptiveness represent Symtrain's highest probability opportunities.

Business Recommendations

Based on these analytics, we recommend five strategic shifts in approach:

1. **Precision Qualification Framework**: While deal size significantly influences the model, our SHAP analysis reveals diminishing returns on very large opportunities. Implement a tiered qualification framework that elevates opportunities showing both substantial value and strong engagement metrics rather than pursuing maximum deal size alone.
2. **Engagement Quality Over Quantity**: Restructure outreach cadences based on the inverse relationship between contact frequency and success probability. Focus on delivering high-value, educational touchpoints aligned with prospect's digital behavior rather than increasing contact volume, which the data shows can reduce win rates.
3. **Digital Behavior-Based Lead Scoring**: Revise lead scoring algorithms to heavily weight conversion rates, session depth, and page views, which collectively show stronger predictive power than traditional firmographic data. Integrate these behavioral signals to guide prioritization across the sales organization.
4. **Vertical-Specific Engagement Strategies**: Develop tailored approaches for underperforming industries identified through SHAP analysis. For these segments, implement pre-qualification steps focusing on digital engagement before committing significant sales resources, as industry alone predicts lower baseline conversion rates.
5. **Contact Center Maturity Assessment**: Develop a formal assessment tool that incorporates the key predictive factors identified in our model. This tool would evaluate prospects on financial readiness, engagement patterns, and industry-specific criteria to generate a holistic opportunity score guiding resource allocation.

Streamlit Dashboard and Deployment

We created a Streamlit-based dashboard that supports:

- Deal Prediction Tab: Filters + Win Rate visualization + prediction using trained Random Forest model
- Customer Segmentation Tab: Visualizes PCA and PACMAP projections of clusters

For the Deal Prediction Tab, users can upload deal data or manually input a deal through a guided form. The app processes this input and returns a predicted outcome (Won/Lost) along with a win probability for each record. Users can preview the uploaded dataset, and any missing values are handled automatically to ensure consistent model performance.

Additional features include:

- **Interactive filtering** by deal stage and deal type to narrow down results
- **Top Deal highlight** that surfaces the deal with the highest predicted win probability
- **Bar chart visualization** of predicted probabilities across all submitted deals
- **Summary metrics** including average win probability across the dataset
- **Downloadable results** to export predictions as a CSV for further review

The **Customer Segmentation Tab** allows users to explore clustering results through both manual and batch input. Users can either enter a single customer's numeric profile or upload a CSV file containing multiple records with seven key numeric features. Once submitted, the app classifies customers into predefined clusters based on underlying behavioral or firmographic patterns.

The dashboard supports interactive visualization using PCA projection, enabling users to see how customers are grouped in reduced-dimensional space. Each point on the cluster map represents a customer, color-coded by cluster assignment. This allows business users to easily spot high-density clusters, outliers, and potential target segments. The tool also assists with strategic persona development and audience targeting based on observed customer characteristics.

The app dynamically ingests CSVs and returns predictions or visual summaries. It is fully containerized using Docker for deployment handoff to the Symtrain team.

Challenges and Learnings

1. Data Quality and Leakage

Several columns were tightly correlated with the outcome (e.g., Deal Probability, Deal Stage) and required manual inspection to avoid leakage. This step highlighted the importance of domain knowledge in supervised learning.

2. High Cardinality and Sparse Features

Features like Web Technologies had thousands of combinations. We handled this by extracting top categories and using one-hot encoding, while using frequency encoding for high-cardinality fields.

3. Imbalanced Data

The original deal dataset was heavily imbalanced. SMOTE helped balance training data and improved minority class prediction.

4. Streamlit Integration

Incorporating complex preprocessing pipelines and trained models into the Streamlit app presented challenges:

- Ensuring consistency between training and inference pipelines
- Managing input schemas (especially one-hot encoded features)
- Packaging large models for lightweight deployment

To address these, we modularized preprocessing into a separate script and used joblib for model export.

Conclusion

This project provided actionable insights into Symtrain's deal dynamics and customer base. From understanding which deal types are most likely to close, to identifying high-engagement customer segments, our analysis equips Symtrain with data-driven strategies for growth. The deployable Streamlit dashboard further enhances usability, allowing stakeholders to apply these models without needing technical expertise. Looking forward, expanding the dataset and integrating CRM updates in real-time could further strengthen predictive accuracy and segmentation depth.