SkyGeni – Sales Intelligence Challenge Submission

PART 1 - Problem Framing

1.1 What is the real business problem here?

The core issue is not simply that the win rate has declined — the real business problem is that revenue outcomes are becoming less predictable despite a healthy-looking pipeline.

A drop in win rate can be caused by multiple underlying factors such as:

- Changes in deal quality entering the pipeline
- Competitive pressure in certain segments
- Poor execution in later stages of the sales cycle
- Shifts in product-market fit across industries or regions
- Operational inefficiencies among reps or teams

The CRO's challenge is that pipeline volume alone is a misleading indicator. The company needs a system that explains:

What is driving losses, where risk is concentrated, and what actions will improve conversion.

So the real business problem is:

> Diagnosing the drivers of declining sales effectiveness and translating them into actionable decisions for revenue leadership.

1.2 What key questions should an AI insight system answer for the CRO?

A decision intelligence system should help the CRO answer questions such as:

- **Where exactly is the win rate dropping?**
  - Specific industries? Regions? Product types? Lead sources?

- **Are we losing more deals, or are we winning fewer high-value deals?**
  - Is the decline uniform or concentrated in certain deal sizes?

- **Which stages of the sales funnel are breaking down?**

    - Are deals stalling in negotiation? Are more deals being lost late-stage?

- **Are certain sales reps or teams underperforming relative to historical baselines?**

- **What deal characteristics are most associated with loss risk today?**
    - Industry + product + lead source combinations

- **What are the biggest actionable levers?**
    - Improve qualification? Focus on certain segments? Coaching for specific reps?

Ultimately, the AI system should move from reporting to recommendations:
“These deals look risky, and these actions are most likely to improve outcomes.”

1.3 What metrics matter most for diagnosing win rate issues?

To understand the win-rate decline, the CRO needs metrics beyond just pipeline volume:

Core Outcome Metrics

- Win Rate (% deals won vs closed)
- Loss Rate by segment (industry, region, product)

Pipeline Health Metrics

- Stage Conversion Rates
    - How deals progress from stage to stage
- **Late-Stage Loss Rate**
    - Losing deals in final stages is more damaging than early disqualification

Deal Velocity Metrics

- **Sales Cycle Length**
    - Longer cycles often indicate friction or weak buyer intent

- **Time-in-Stage**
    - Deals stuck too long are high-risk

Revenue Impact Metrics

- **ACV-weighted Win Rate**
    - Losing fewer big deals can hurt more than many small losses

- **Forecast Risk Concentration**
    - Are losses clustered in key quarters or segments?

Rep-Level Performance Metrics

- Win rate per rep vs historical baseline
- Pipeline quality per rep

1.4. What assumptions are you making about the data or business?

Some assumptions required for this analysis include:

- All deals in the dataset are labeled as Won/Lost, so the analysis focuses on historical closed outcomes rather than live open pipeline scoring.
- 2024Q3 is excluded from core trend interpretation because only partial data is available through July.
- The goal is interpretability and decision usefulness rather than complex modeling accuracy.
- The dataset represents the majority of the company's deal flow over the last few quarters
- Deal stages follow a consistent process across reps and regions
- External factors (pricing changes, competitor moves) are not explicitly captured
- Win rate decline could reflect either:
    - Worse deal quality entering the pipeline
    - Execution issues later in the funnel
    - Market shifts affecting specific customer segments

Part 2 — Data Exploration & Business Insights

2.1 Objective

The CRO reports:

> "Our win rate has dropped over the last two quarters, but pipeline volume looks healthy."

The goal of Part 2 is to use exploratory data analysis (EDA) to diagnose why win rate is declining, identify the most important business drivers behind the change, and generate actionable insights that a sales leader can use immediately.

This analysis focuses on closed historical deals, segmented by key attributes such as:

- Lead source
- Region
- Industry
- Product type

- Deal stage
- Deal amount (ACV)

The complete EDA workflow is provided in the code file part2_eda_insights.py

---

2.2 Baseline vs Recent Quarter Comparison

Since the CRO specifically mentions a decline over the last two quarters, the analysis focuses on comparing:

- **Baseline Quarter:** 2024Q1
- **Recent Quarter:** 2024Q2

This quarter-over-quarter framing mirrors how CROs typically evaluate performance trends.

Note on 2024Q3
The dataset contains a small number of deals closed in July 2024, which technically fall into **2024Q3**. However, since Q3 is incomplete (only partial data is available), it is excluded from the core win rate diagnosis to avoid misleading conclusions.

2.3 Key Business Metric

Win Rate

The primary metric of interest is win rate:

$$Win\ Rate\ =\ \frac{Deals\ Won}{Total\ Deals\ Closed}$$

Win rate is the CRO's most important indicator of sales effectiveness and pipeline conversion quality.

A declining win rate despite stable pipeline volume suggests that deal quality, execution, or segmentation performance may be shifting.

2.4 Custom Metrics Designed for CRO Decision-Making

Beyond standard win rate reporting, I designed two custom metrics to make the analysis more decision-oriented.

Custom Metric 1: QAPI — Quality-Adjusted Pipeline Inflow

Pipeline volume alone (deal count or ACV) can be misleading.

A company may generate the same amount of pipeline, but if the pipeline mix shifts toward weaker-converting segments, win rate will decline.

QAPI estimates the expected revenue value of pipeline inflow, adjusted by historical win-rate quality.

$$QAPI(q) \; = \; \Sigma \, ACV_{created\ in\ q} \; \times \; BaselineWinRate\ (segment)$$

meaning for every deal created in quarter q, multiply its ACV by the historical baseline win rate of its segment, then sum across all deals.

This metric helps answer:

> "Is the pipeline we are generating still high-quality, or just high-volume?"

Action it enables:

- Detect early deterioration in lead quality
- Improve qualification and scoring upstream
- Prevent future win rate drops before deals close

Custom Metric 2: WRSI — Win Rate Shock Index

Sales leaders need early warning signals when performance drops unusually sharply in specific segments.

WRSI is a statistical "shock score" that highlights segments where win rate decline exceeds normal variation.

- Large negative values = unusually sharp drop
- Useful for automated alerts and prioritization

Rule of thumb:

- **WRSI < −1.5 → high-risk segment requiring immediate attention**

Action it enables:

- Segment-level anomaly detection
- Faster win/loss root-cause analysis
- Targeted intervention instead of broad guessing

2.5 Key Business Insights

The following insights were generated using driver analysis across multiple pipeline dimensions.

Insight 1 — Inbound Lead Performance Collapsed

Lead source analysis shows that Inbound deals experienced the sharpest win rate decline:

- Baseline win rate (2024Q1): **48.8%**
- Recent win rate (2024Q2): **40.5%**
- Change: **−8.34 percentage points**
- Impact on overall decline: **−2.14 pts**
- WRSI shock score: **−1.68** (statistically significant)

Why this matters

Inbound is typically one of the largest pipeline contributors.
If inbound conversion weakens, pipeline volume may remain high while revenue outcomes deteriorate.

CRO Action

- Audit inbound lead quality (ICP fit, intent strength)
- Tighten qualification and routing rules
- Align marketing + sales on scoring and handoff

Insight 2 — Partner and Outbound Channels Also Weakened

Other acquisition channels also declined quarter-over-quarter:

- Partner win rate dropped **−4.10 pts**
- Outbound win rate dropped **−3.43 pts**

Why this matters

This suggests the decline is not isolated to one channel.
It may reflect broader execution issues such as:

- weaker discovery conversations
- competitive pressure
- pricing objections
- reduced buyer urgency

CRO Action

- Review outbound targeting and messaging
- Improve partner qualification and deal support
- Add coaching around early-stage conversion

Insight 3 — Referral Deals Are a Bright Spot

Referral performance improved:

- Referral win rate increased by **+4.13 pts**

Why this matters

Referrals represent a high-trust pipeline with stronger buyer intent and faster close likelihood.

CRO Action

- Invest in referral loops and advocacy programs
- Encourage reps to source more customer-driven introductions

---

Insight 4 — Regional Performance Decline is Not Uniform

Region-level driver analysis shows that win rate decline is concentrated in specific geographies, particularly:

- India region showing one of the strongest negative shifts

Why this matters

Pipeline may look healthy globally, but regional execution issues can drive overall decline.

Possible drivers include:

- increased competitive intensity
- pricing sensitivity
- inconsistent enablement across regions

CRO Action

- Run focused win/loss reviews by region
- Introduce region-specific playbooks
- Evaluate rep performance and competitor patterns

Insight 5 — Core and Pro Product Motions are Dragging Win Rate

Product-type segmentation indicates that win rate deterioration is strongest in:

- **Core**
- **Pro**

while Enterprise performance is more stable.

Why this matters

This suggests the issue may not be company-wide, but concentrated in mid-market motions due to:

- packaging friction
- weaker differentiation
- competitive displacement

CRO Action

- Reassess qualification criteria for Core/Pro opportunities
- Improve objection handling and positioning
- Explore pricing or bundling experiments

Part 3 — Decision Engine (Option B: Win Rate Driver Analysis)

3.1 Why Option B

The CRO's concern is:

> "Win rate has dropped over the last two quarters, but pipeline volume looks healthy.
> I don't know what exactly is going wrong."

This is not primarily a prediction problem.

The CRO is not asking:

- "Will this specific deal close?"

They are asking:

- "What has changed in our business performance?"
- "Which areas are driving the decline?"
- "Where should my team focus to recover the win rate?"

Therefore, I chose:

Option B — Win Rate Driver Analysis

This option directly supports executive decision-making by identifying the factors that are hurting or improving win rate.

3.2 Problem Definition

The goal of this decision engine is:

> To explain why win rate declined quarter-over-quarter, and generate prioritized, actionable insights for sales leadership.

Specifically, the system should answer:

- Which segments contributed most to the win rate drop?
- Are these changes statistically meaningful or just noise?
- What actions should the CRO take next?

3.3 Approach: Interpretable Rule-Based Decision Engine

Instead of building a complex black-box ML model, I implemented a lightweight, interpretable **rule-based driver engine**.

This is intentional:

- CROs need explanations, not opaque probabilities
- The output must map directly to business actions
- SkyGeni's value is decision intelligence, not model complexity

The engine works by comparing segment performance across two quarters:

- **Baseline Quarter:** 2024Q1
- **Recent Quarter:** 2024Q2

3.4 Core Driver Logic

For each business dimension (lead source, region, product type), the engine computes three key decision metrics:

---

1. Segment Win Rate Change

For each segment s:

$$\Delta WR(s) \ = \ WR(s)_{recent} \ - \ WR(s)_{baseline}$$

This identifies which segments improved or deteriorated.

Example:

- Inbound win rate dropped sharply from baseline to recent quarter.

---

2. Impact Attribution (Business Contribution)

Not every segment matters equally.

A segment with high volume and declining win rate will drive the overall drop more than a small segment.

Impact is computed as:

$$Impact(s) \ = \ Share(s)_{baseline} \ \times \ \Delta WR(s)$$

This produces a ranked list of true win rate drivers.

---

3. Win Rate Shock Index (WRSI)

To detect unusually sharp declines, I introduced WRSI.

It is a statistical shock score that flags segments where win rate changes exceed expected variance.

Rule:

- **WRSI < −1.5 → high-risk performance shock**

This makes the engine usable as an alerting system, not just reporting.

3.5 Actionable Outputs Produced

The engine generates CRO-readable driver tables across:

- Lead Source
- Region
- Product Type

Each segment is scored with:

- Baseline win rate
- Recent win rate
- Win rate change
- Impact contribution
- Shock score

This transforms raw pipeline data into prioritized decision signals.

3.6 Key Drivers Identified in This Dataset

Driver 1 — Inbound is the Primary Source of Win Rate Decline

Inbound deals experienced the sharpest drop:

- Win rate fell from **48.8%** → **40.5%**
- Change: **−8.34 pts**
- Impact contribution: **−2.14 pts**
- WRSI shock score: **−1.68** (significant decline)

Interpretation:

> The win rate decline is primarily driven by inbound conversion deterioration.

Recommended CRO Action:

- Audit inbound lead quality and ICP alignment
- Improve qualification and routing
- Align marketing + sales on scoring

Driver 2 — Regional Weakness is Concentrated (India)

Region-level analysis shows the decline is not uniform.

The India region shows one of the strongest negative shifts.

Interpretation:

> Pipeline may look healthy globally, but performance issues are localized.

Recommended CRO Action:

- Conduct India-specific win/loss reviews
- Introduce region-specific enablement

- Investigate competitor pressure and pricing sensitivity

Driver 3 — Core and Pro Motions are Dragging Performance

Product-type analysis indicates deterioration in:

- Core
- Pro

while Enterprise performance is more stable.

Interpretation:

> Mid-market product motions may face packaging friction or competitive displacement.

Recommended CRO Action:

- Improve objection handling for Core/Pro
- Reassess qualification criteria
- Consider pricing/bundling experiments

Positive Signal — Referral Deals are Improving

Referral win rate increased quarter-over-quarter.

Interpretation:

> Referral is the strongest-performing channel and should be scaled.

Recommended CRO Action:

- Invest in referral loops and advocacy programs
- Encourage reps to source more customer-driven pipeline

3.7 How a CRO Would Use This Engine

This driver engine can be operationalized in weekly or monthly revenue reviews:

1. Run driver analysis each period

2. Identify top negative impact segments

3. Trigger targeted interventions (not generic coaching)

4. Monitor whether win rate recovers in the next quarter

Part 4 — Mini System Design: Sales Insight & Alert System

4.1 Goal

Design a lightweight system that operationalizes the Win Rate Driver Analysis (Option B) so a CRO and RevOps team can:

- Understand *why* win rate is changing
- Get alerted when a meaningful shift happens
- Take action quickly (channel, region, product playbooks)

The system should produce **interpretable, segment-level insights** (not black-box predictions) and run continuously with minimal maintenance.

4.2 High-Level Architecture

Sources

- CRM / Sales system (Salesforce, HubSpot)
- Optional enrichment (product usage, intent data, marketing attribution)

Pipeline

1. **Ingestion**
   - Daily extract of deals + relevant fields (deal_id, created_date, closed_date, ACV, rep, region, product, lead_source, outcome, etc.)
2. **Validation & Standardization**
   - Schema checks (required columns present)
   - Value normalization (consistent labels for region, lead_source, stages)
   - Date sanity checks (created_date ≤ closed_date)
3. **Feature/Metric Layer**
   - Compute closed-quarter win rates by segment
   - Compute driver metrics:
     - Win Rate Change (baseline → recent)
     - Impact Attribution (share × delta)
     - WRSI shock score (signal vs noise)
   - Compute QAPI (quality-adjusted pipeline inflow)
4. **Insights Store**

    - Persist daily/weekly results in a metrics DB (Postgres / warehouse table)
    - Store both “raw tables” and “top insights”
5. **Delivery**
    - Slack/Email alerts for high-impact changes
    - Executive dashboard (Looker/PowerBI/internal web UI)

4.3 Data Flow (Step-by-Step)

Step 1 — Data ingestion (daily)

- Pull last N months of deals (or incremental changes since last run)
- Partition into:
    - Closed deals (used for win rate driver analysis)
    - Open deals (optional future work; not required for Option B)

Step 2 — Clean & standardize

- Standardize categorical fields:
    - lead_source, product_type, region, industry, stage
- Handle missing values:
    - If a segment label is missing, map to “Unknown”
- Dedupe by deal_id (keep latest state)

Step 3 — Compute metrics and drivers

For each dimension (lead_source, region, product_type, etc.):

- Compare **baseline quarter** vs **recent quarter**
- Produce a ranked driver table including:
    - Baseline Win Rate
    - Recent Win Rate
    - Win Rate Change (pts)
    - Impact on Overall (pts)
    - WRSI Shock Score

Compute custom pipeline quality metric:

- **QAPI** per created quarter (skip first quarter due to missing historical baseline)

Step 4 — Generate Insights + recommended actions

Convert metrics into plain-language insights:

- “Inbound win rate dropped 8.3 pts QoQ (impact -2.14 pts)”
- “India region shows abnormal decline; investigate competitors/pricing”

- “Core/Pro motions are weakening; review qualification and enablement”
- “Referral is improving; scale referral loops”

Step 5 — Alert & dashboard delivery

- Alerts delivered weekly for execs + daily for RevOps (configurable)
- Dashboard updated after each run (or on schedule)

4.4 Example Alerts

Alert Type A — High-impact driver

Trigger condition:
- Impact on overall win rate ≤ -1.0 pts (configurable)

Example:
- “Inbound is driving win rate decline: -8.3 pts QoQ (Impact: -2.14 pts). Action: audit inbound lead quality + scoring + routing.”

Alert Type B — Shock detection

Trigger condition:
- WRSI Shock Score < -1.5

Example:
- “India region shows abnormal win rate drop (Shock: -1.7). Action: run win/loss review + competitor/pricing assessment.”

Alert Type C — Positive opportunity

Trigger condition:
- Impact ≥ +0.75 pts or win rate improvement ≥ +4 pts

Example:
- “Referral performance improved QoQ (+4.1 pts). Action: scale referral program and customer advocacy motion.”

Alert Type D — Pipeline quality mismatch

Trigger condition:
- Pipeline ACV stable/up, but QAPI down materially

Example:

- “Pipeline volume was stable, but quality-adjusted expected revenue declined. Action: tighten qualification and improve ICP match.”

4.5 How Often It Runs

Recommended schedule:

- **Daily**: compute metrics and refresh dashboards (for RevOps)
- **Weekly**: send driver alerts and action summaries (for CRO/Sales leadership)
- **Quarterly**: produce a formal “Win Rate Driver Review” report for QBRs

(Your assignment implementation can run locally via a Python script; in production it becomes a scheduled job.)

4.6 Failure Cases & Limitations

1) Partial-quarter bias

If the “recent quarter” is incomplete, win rates can be misleading.

- Mitigation: lock comparisons to complete quarters and explicitly label completeness.

2) Data quality issues

CRMs often have:

- missing outcomes
- inconsistent stage definitions
- mislabeled lead sources
- duplicate deals

Mitigation:

- validation layer + data contracts
- “Unknown” bucket
- alert when missingness exceeds thresholds

3) Correlation ≠ causation

Drivers show *where* performance changed, not the true root cause.
Mitigation:

- recommend follow-up investigation steps
- enrich with “loss reason”, competitor, pricing discount, activity logs (future)

4) Segment sparsity

Small segments can show large win rate swings due to low volume.
Mitigation:

- minimum sample thresholds (e.g., $n \geq 30$)
- rely on WRSI / confidence scoring

Part 5 — Reflection (Most Important)

This assignment focused on building an interpretable decision intelligence system to help a CRO understand why win rate is declining despite healthy pipeline volume. While the driver-based approach produces actionable insights, several assumptions and limitations would need to be addressed in a real production deployment.

5.1 Weakest Assumptions in My Solution

The weakest assumption is that historical closed deals alone are sufficient to explain win rate decline.

In practice, win rate shifts are often driven by factors not present in this dataset, such as:

- competitor involvement
- pricing discounts
- buyer intent strength
- sales activity levels (calls, meetings, follow-ups)
- loss reasons and objection categories

My current solution identifies *where* performance changed, but not always the deeper causal reason behind it.

5.2 What Would Break in Real-World Production

Several real-world challenges could break or distort the system:

1. CRM Data Quality Issues

Sales CRM data is often messy:

- inconsistent stage definitions
- missing or delayed outcomes

- misclassified lead sources
- duplicated or reopened deals

Without strong validation and standardization, driver metrics can become unreliable.

---

2. Partial Quarter and Seasonality Effects

Quarter-over-quarter comparisons can be misleading when:

- the “recent quarter” is incomplete
- seasonal buying patterns shift
- deal cycles span multiple quarters

Production systems would require time-window controls and completeness checks.

---

3. Segment Sparsity and Noise

Some segments may have low deal counts, which can produce large swings in win rate due to randomness rather than true performance change.

While WRSI helps reduce this risk, production systems should enforce:

- minimum sample thresholds
- confidence intervals
- smoothing across time windows

---

4. Correlation vs Causation

Driver analysis highlights correlations:

- “Inbound win rate dropped”
- “India region weakened”

But it does not prove causality.

In production, sales leaders will ask:

> “Is inbound worse because of lead quality, competition, rep execution, or pricing?”

Additional enrichment is required to move from diagnosis → root cause.

5.3 What I Would Build Next With 1 Month More Time

If given one month, the next iteration would focus on turning this into a more complete SkyGeni-grade decision product:

1. Add Activity and Execution Signals

Integrate sales engagement data:

- number of touches
- meeting velocity
- response time
- stage duration

This would help explain whether declines are due to execution breakdowns.

---

2. Add Deal-Level Root Cause Explainability

Enrich with:

- loss reason fields
- competitor mentions
- pricing/discounting
- product usage intent signals

This would allow insights like:

> "Inbound win rate dropped primarily due to pricing objections vs competitor displacement."

---

3. Build an Alerting + Playbook Layer

Productize the driver engine into automated workflows:

- weekly CRO digest
- anomaly alerts when WRSI crosses threshold
- recommended actions tied to driver type

Example:

- "Inbound shock detected → tighten qualification + run marketing audit"

---

4. Extend Into Forward-Looking Decision Support

The next logical extension is combining driver insights with:

- deal risk scoring for open pipeline
- revenue forecasting adjusted by driver health

This would connect diagnosis directly to future outcomes.

5.4 Least Confident Part of My Solution

The area I am least confident about is the extent to which segment-level win rate drivers alone capture the full complexity of sales performance.

In real B2B SaaS environments, win rate is influenced heavily by:

- competitive dynamics
- rep behavior
- multi-threading and champion strength
- pricing and procurement friction

Without those signals, driver analysis may surface the correct “where” but not always the full “why.”

However, I believe the framework is still highly valuable as a first decision layer that can be expanded with richer production data.