

Project Retention

A Data-Driven Strategy to Reduce Employee Attrition

Proof of Concept (PoC) for Company I

Presented By : Sambath Seakty

Annual Revenue

\$12M



Today's Expenses

\$250K



Loan Approval Rate

>65%

Equity

55%

from last month



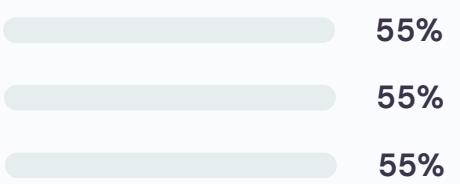
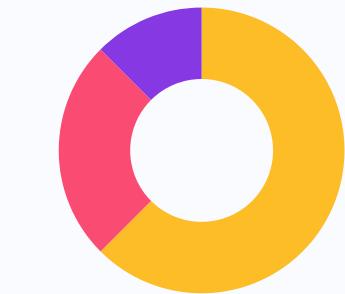
Current YTD (Q3)

72%

Invalid

Total Income

\$345M



Content

- 1. Executive Summary**
- 2. Situation Analysis**
- 3. Predictive Modeling Strategy**
- 4. Strategic Proposals & ROI**
- 5. Implementation Roadmap**

Executive Summary

The Challenge

- Company I faces a **16% attrition rate**, threatening project continuity and incurring high replacement costs.

The Discovery

- Our analysis reveals that attrition is not driven by base salary, but by the ***Incentive Ratio*** (Bonus vs. Income) and ***Hidden Burnout***.

The Solution

- We propose a ***Proportional Reward System*** and a "***Pulse Check***" Strategy, powered by a predictive XGBoost model.

Potential annual savings of \$5.88 Million by retaining key talent.

The "*Turnover Tax*" is Rising in Tech.

- ⊕ Tech industry turnover is high (13–15%), costing companies **150–200%** of an employee's salary to replace them.

2024 Trends show a shift back to "**Total Rewards**"—employees want L compensation that reflects their performance output.

- ⌚ "**Silent Burnout**" is an epidemic, with 42% of tech workers reporting stress that managers often miss.



Data Landscape & Demographics



Source: Historical HR Data (~1,500 Employees).



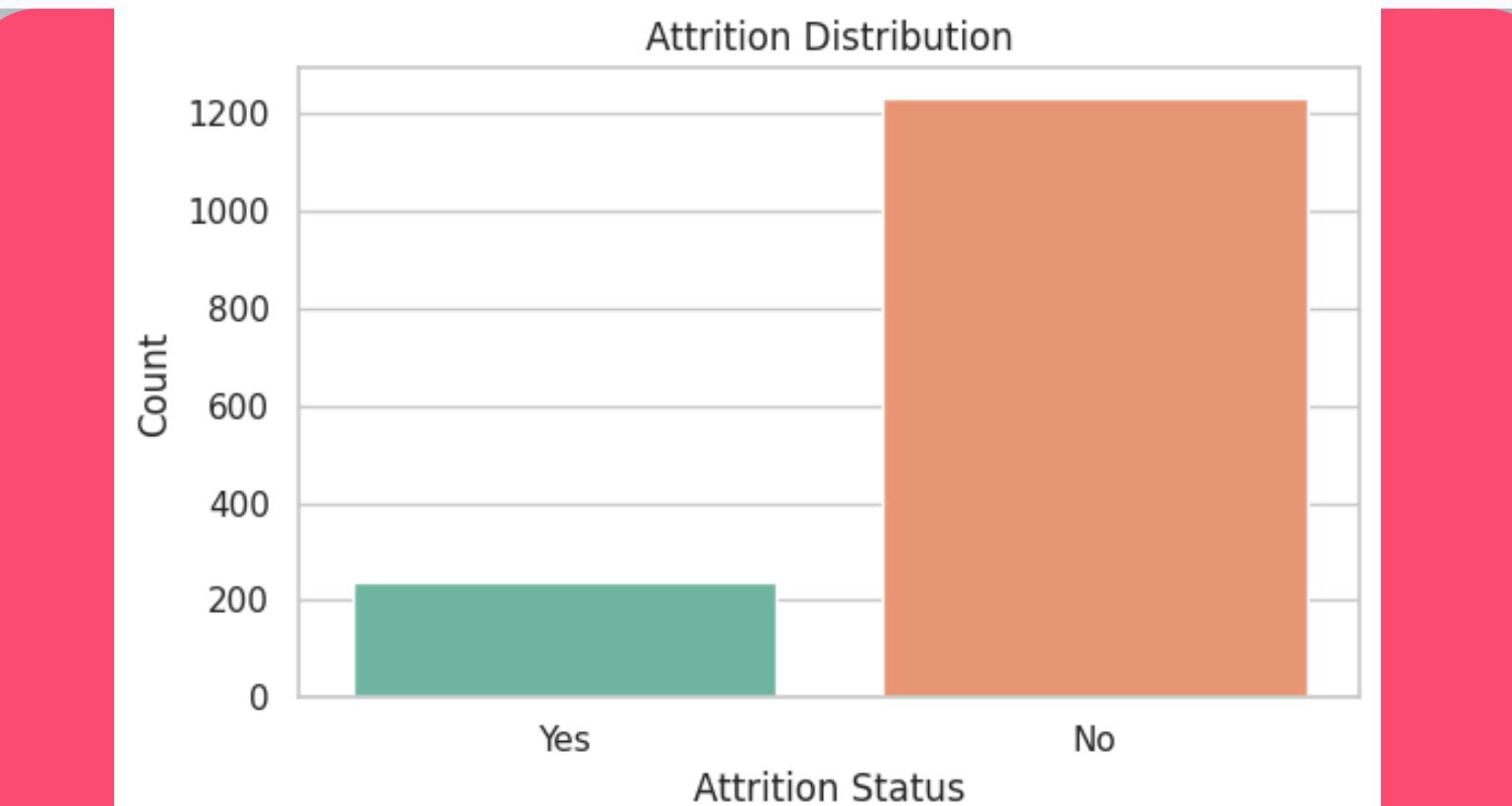
Target: *Attrition* (Yes/No).



Key Challenge: The data is **Imbalanced** (Only 16% Quit). This requires specialized modeling techniques to detect the risk.

Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...	RemoteWork	MonthlyIncome	StressRating	WelfareBenefits	InHouseFacility	ExternalFacility	ExtendedLeave	FlexibleWork	StressSelfReported	Year
0	41	Yes	Travel_Rarely	Sales	1	2	Life Sciences	1	1	2	...	1	5224	4	1	0	0	0	0	1
1	49	No	Travel_Frequently	Research & Development	8	1	Life Sciences	1	2	3	...	3	6863	2	4	1	0	0	1	1
2	37	Yes	Travel_Rarely	Research & Development	2	2	Other	1	4	4	...	2	7612	3	2	1	0	0	0	1
3	33	No	Travel_Frequently	Research & Development	3	4	Life Sciences	1	5	4	...	2	11245	1	4	1	1	1	1	1
4	27	No	Travel_Rarely	Research & Development	2	1	Medical	1	7	1	...	2	3029	3	2	0	0	0	0	3
5	32	No	Travel_Frequently	Research & Development	2	2	Life Sciences	1	8	4	...	3	9313	3	2	0	0	0	0	3
6	59	No	Travel_Rarely	Research & Development	3	3	Medical	1	10	3	...	2	5562	3	1	0	0	0	1	3
7	30	No	Travel_Rarely	Research & Development	24	1	Life Sciences	1	11	4	...	4	7362	2	2	0	0	0	1	1
8	38	No	Travel_Frequently	Research & Development	23	3	Life Sciences	1	12	4	...	4	6266	1	3	1	1	0	1	1
9	36	No	Travel_Rarely	Research & Development	27	3	Medical	1	13	3	...	5	8258	3	3	1	1	0	1	3

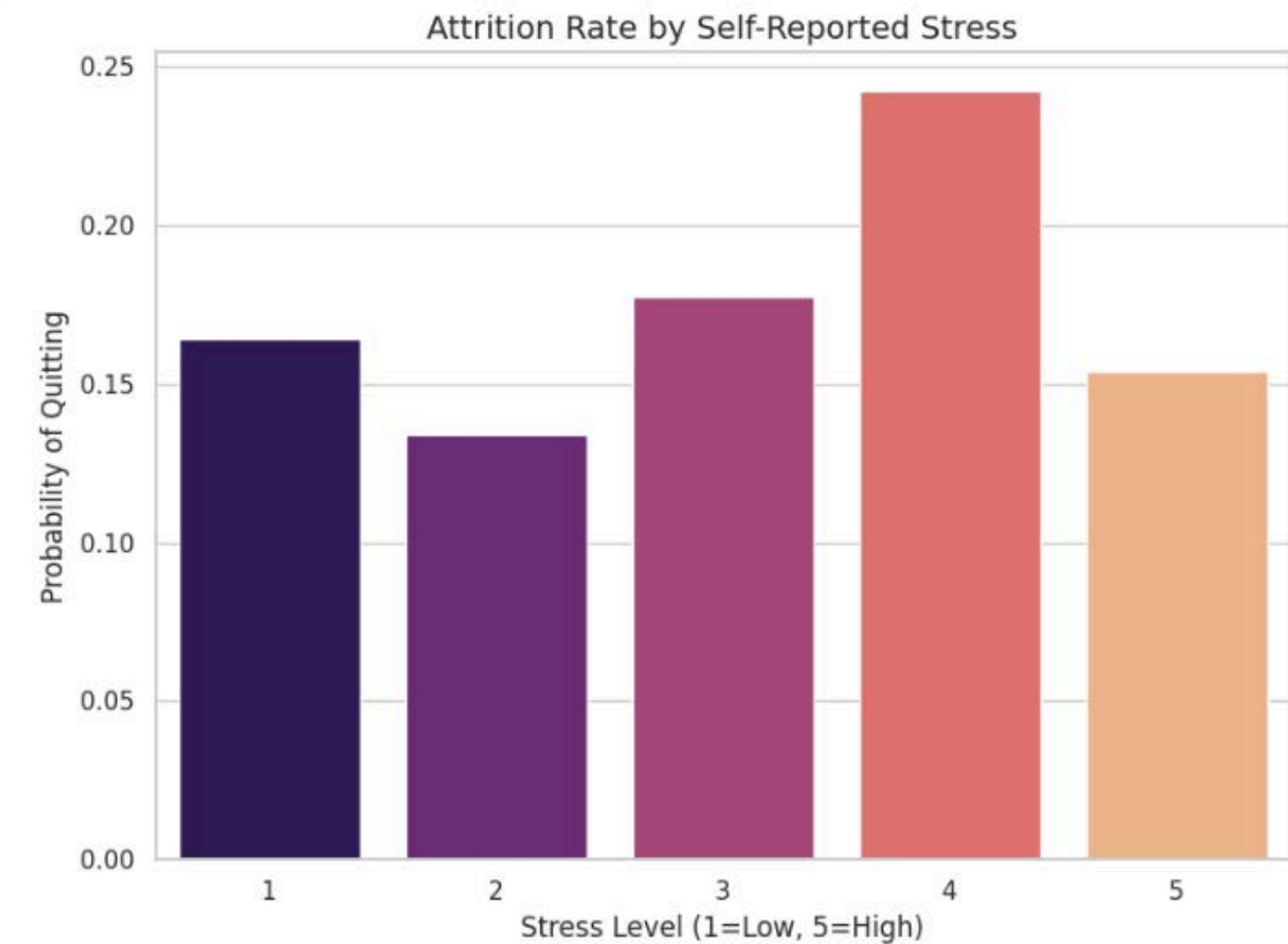
10 rows x 44 columns



The "Stress Disconnect" (EDA)

Managers are Blind to Employee Stress.

- Employees with high Self-Reported Stress quit significantly more often.
- **Critical Gap:** Data shows Managers consistently rate team stress lower than employees rate themselves.
- **Conclusion:** The current "Manager Evaluation" system is failing to detect burnout.



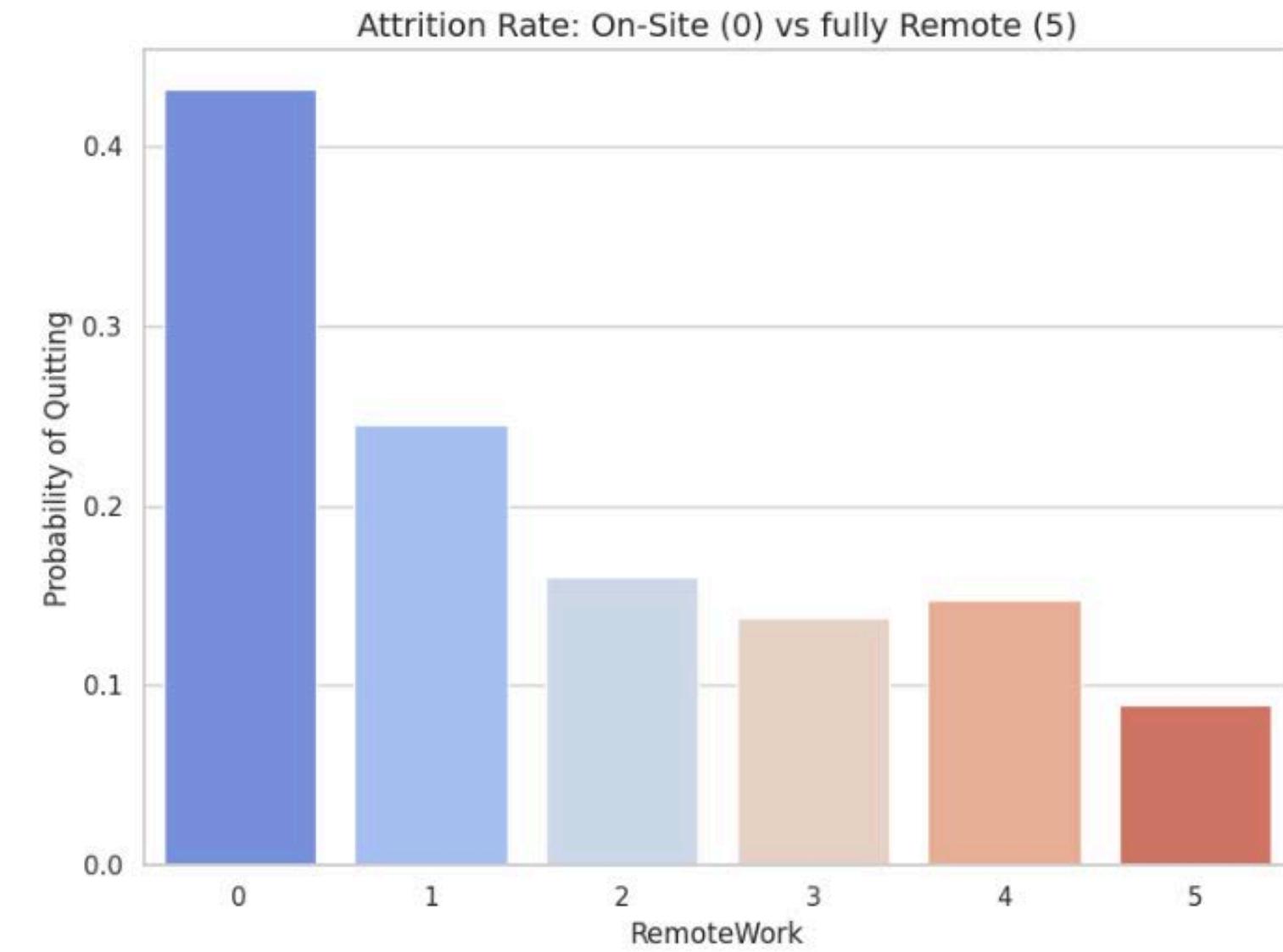
--- INSIGHT CHECK: Stress Mismatch ---

StressRating	
1	1.308642
2	1.783961
3	2.448669
4	2.930070
5	3.000000

Flexibility Matters (EDA)

Flexibility is the Strongest Retention Lever.

- **The Trend:** The data shows a clear, linear relationship: as Remote Work days increase, Attrition decreases drastically.
- **The Danger Zone:** Employees with Zero Remote Days (On-Site) have the highest probability of quitting (>40%).
- **The Safe Zone:** Employees with Full Remote (Level 5) status have the lowest attrition rate (<10%).
- **Conclusion:** The "Return to Office" mandate is the primary driver of turnover.



Model Selection Strategy

We conducted a comparative tournament of three algorithms:

--- 🌟 STARTING CROSS-VALIDATION TOURNAMENT (5 Folds) 🌟 ---

- ✓ Logistic Regression Finished. (Avg AUC: 0.796)
- ✓ Random Forest Finished. (Avg AUC: 0.806)
- ✓ XGBoost Finished. (Avg AUC: 0.826)

--- 🏆 FINAL ROBUST RANKINGS 🏆 ---

Model	Avg Accuracy	Avg AUC	Avg Recall	Avg Precision
XGBoost	0.874830	0.826145	0.457801	0.666347
Random Forest	0.857143	0.805693	0.151330	0.850000
Logistic Regression	0.718367	0.796302	0.739716	0.334533

Logistic Regression

High Recall (73%) but extremely low Precision (33%). Too many false alarms.

Random Forest

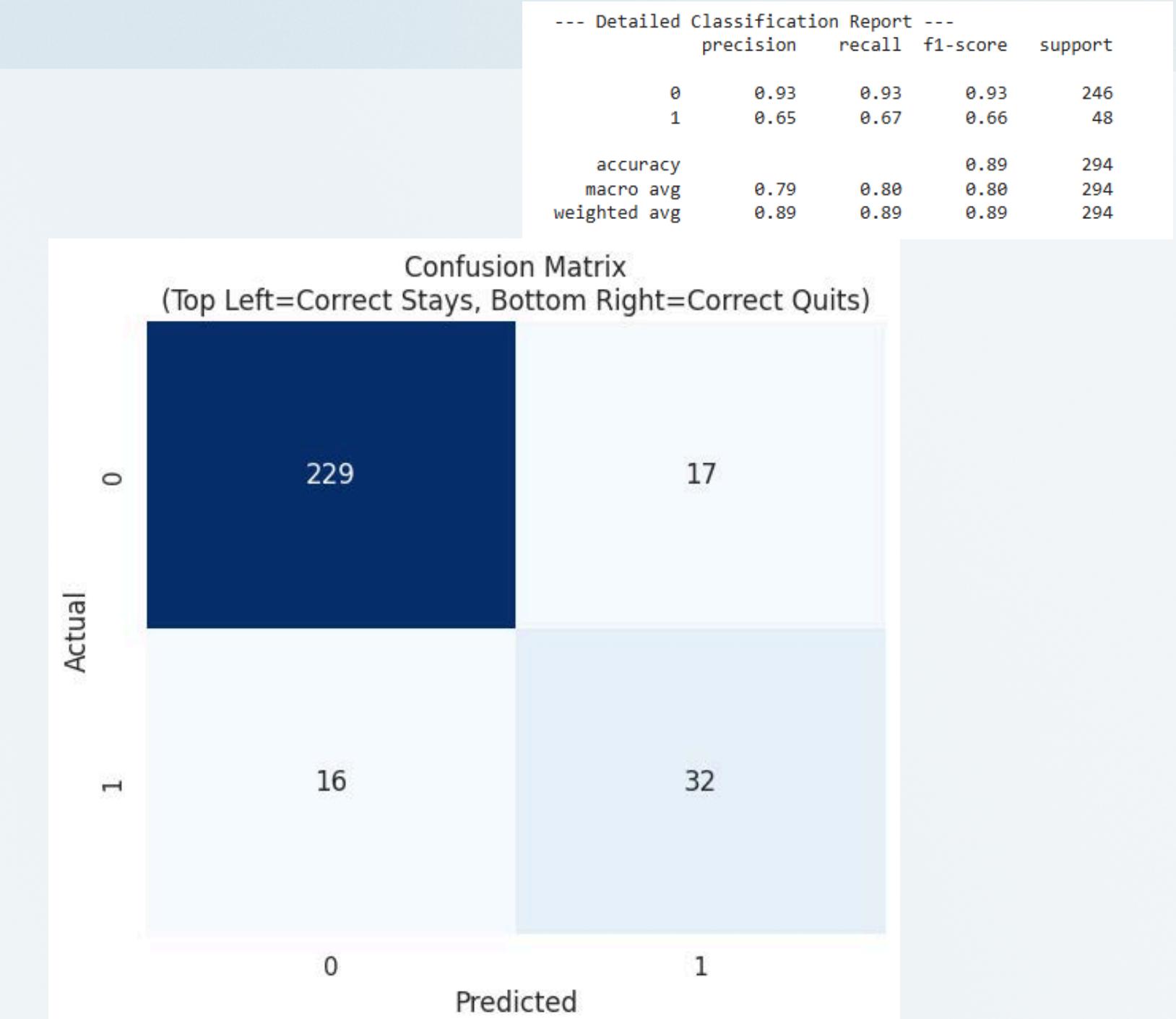
High Precision but failed to catch quitters (Recall 12%). Too conservative.

XGBoost (Winner)

The optimal balance. High predictive power (AUC 0.83) and high sensitivity to at-risk employees.

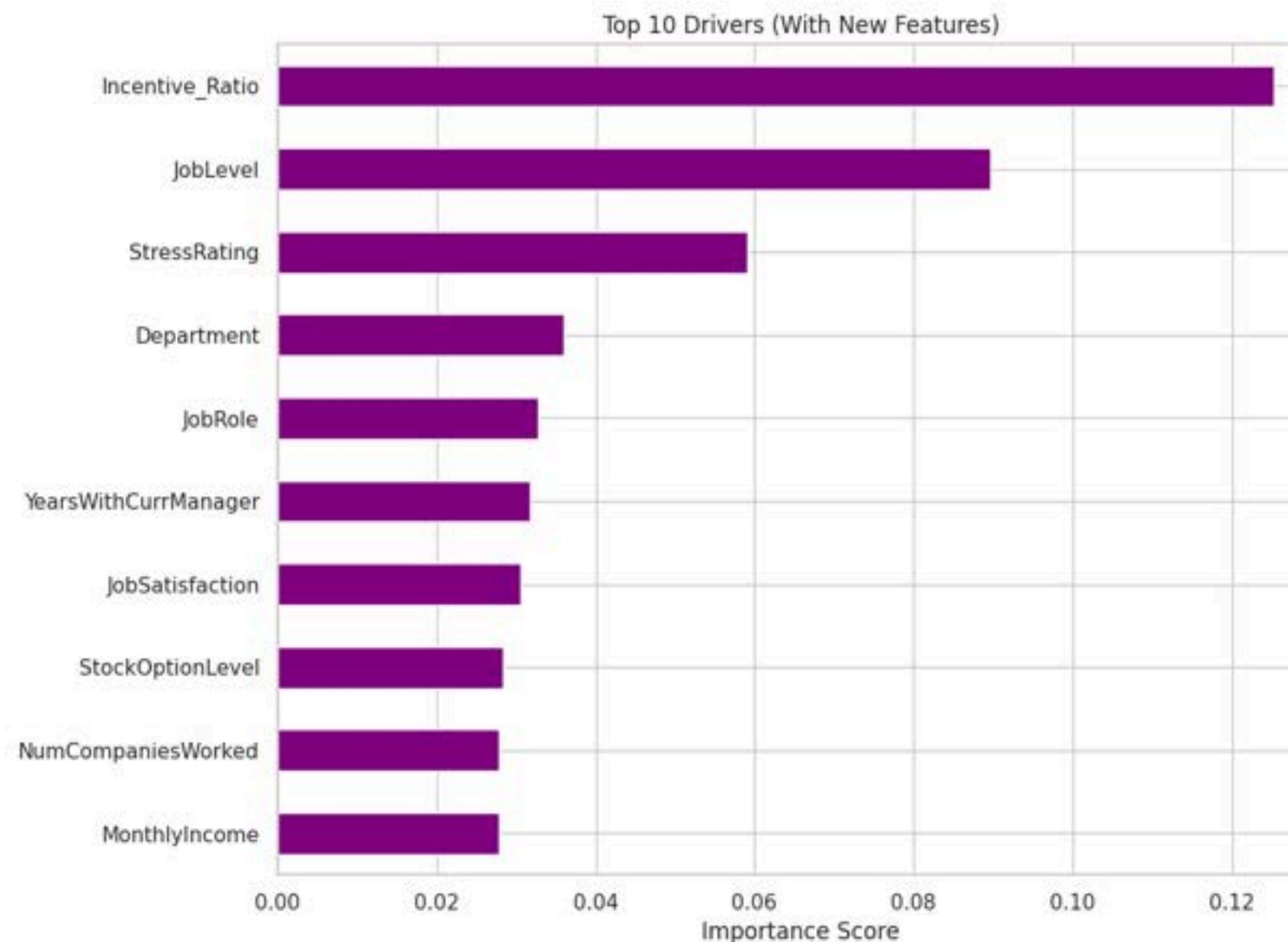
Predictive Capabilities (XGBoost)

- **AUC Score:** 0.83 (Excellent ability to rank risk).
- **Recall:** 67% (Identifying 2 out of 3 quitters).
- **Precision:** 65% (Reliable predictions).



Note: "By engineering custom features, we improved Recall from 46% to 67%."

Top Drivers of Attrition



- The model reveals that **Incentive Ratio** (Bonus relative to Income) is the **#1 Predictor** of attrition
- **Job Level** and **Stress Rating** follow closely.
- **Insight:** Employees don't just want "more money"; they want rewards that feel significant compared to their base pay.

Proposal 1 - The Proportional Reward System

Strategy: Restructure Incentives

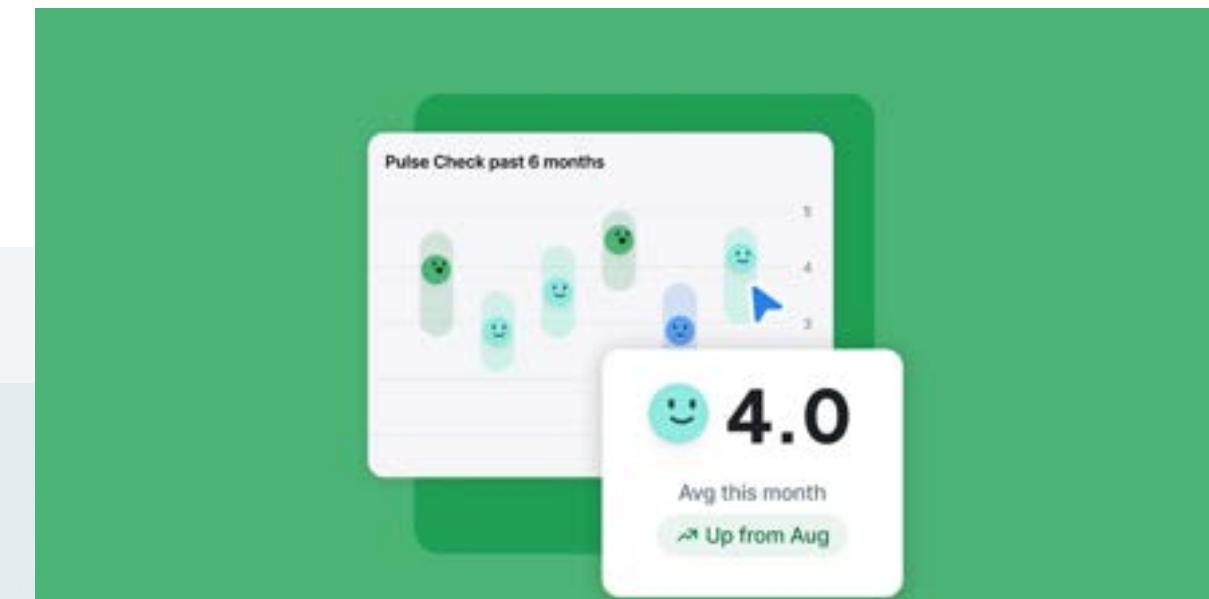
- **The Logic:** *Incentive_Ratio* is the top driver. Flat bonuses do not motivate junior staff effectively.
- **The Action:** Move from fixed-amount bonuses to ***Percentage-Based Bonuses***.
- **The Goal:** Ensure high-performers feel the financial impact of their success relative to their salary.



Proposal 2 - The "Pulse Check" System

Strategy: Close the Stress Gap

- **The Logic:** Managers cannot see burnout (The "Disconnect" found in EDA).
- **The Action:** Implement an Anonymous Monthly Pulse Check.
- **The Goal:** Use data (Self-Reported Stress) to trigger interventions, rather than relying on manager guesses.



Proposal 3 - Remote-First Policy

Strategy: Formalize Flexibility

- **The Logic:** On-site employees are flight risks.
- **The Action:** Make "Remote-First" a core benefit, not just a temporary perk.
- **The Goal:** Improve retention without increasing payroll costs.



The Financial Case

- **Prediction:** Our XGBoost model identified ~245 employees company-wide who are at high risk of quitting.
- **Cost Basis:** Based on the average salary of this risk group (\$80k/year) and SHRM standard replacement costs (150%).**Target:** Prevent 20% of Attrition (50 people).
- The "Bleeding": The total cost of doing nothing is \$29.4 Million annually.
- **Implementing this proposal pays for itself in the first month.**

Total Risk Exposure

\$29.4 Million / Year

Success Target

Retain 20% of At-Risk Talent

PROJECTED SAVINGS

\$5.88 Million

Roadmap

Phase 1 (Month 1)



Deploy XGBoost Model to HR Dashboard. Identify "High Risk" list.

Phase 2 (Month 3)



Launch "Pulse Check" Survey to gather better stress data.

Phase 3 (Month 6)



Roll out new Incentive Structure for FY2026.

References

1. Data Source

- IBM HR Analytics Employee Attrition & Performance dataset: modified version
 - [Link:https://drive.google.com/file/d/1MRgT5pYTSaaEKV37VLJ92iR49eWuBvGM/view?usp=drive_link](https://drive.google.com/file/d/1MRgT5pYTSaaEKV37VLJ92iR49eWuBvGM/view?usp=drive_link)

2. Industry & Market Research

- SHRM (Society for Human Resource Management): "Cost of Turnover" Analysis (Estimating 150–200% replacement cost).
 - [Link:https://drive.google.com/file/d/1MRgT5pYTSaaEKV37VLJ92iR49eWuBvGM/view?usp=drive_link](https://drive.google.com/file/d/1MRgT5pYTSaaEKV37VLJ92iR49eWuBvGM/view?usp=drive_link)
- Mercer: "Global Talent Trends 2024" (identifying Compensation as a top driver).
 - [Link:https://www.mercer.com/en-us/insights/talent-and-transformation/global-talent-trends/](https://www.mercer.com/en-us/insights/talent-and-transformation/global-talent-trends/)
- Yerbo: "The State of Burnout in Tech" (identifying the 42% burnout rate).
 - [Link: https://yerbo.co](https://yerbo.co)

3. Technical Methodology

- XGBoost Documentation: Chen, T., & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System."
 - [Link: https://xgboost.readthedocs.io/en/stable/](https://xgboost.readthedocs.io/en/stable/)
- Scikit-Learn: Machine Learning in Python.
 - [Link: https://scikit-learn.org/stable/](https://scikit-learn.org/stable/)