GCI FINAL PROJECT

EXECUTIVE SUMMARY

- Business Challenge

The company is experiencing a high employee attrition rate of 16.19%, leading to significant replacement and productivity costs. Current HR strategies are reactive, identifying resignations only after employees decide to leave.

- Our Solution

We developed a Predictive Attrition Model using company workforce data to identify employees at high risk of leaving. The model analyzes key factors such as stress, benefits, job role, and work-life balance to generate actionable insights.

- Key Outcomes

1. ML model (Random Forest) achieved ~0.81 ROC-AUC, strong predictive ability
2. Identified 238 high-risk employees who are likely to leave
3. Estimated annual cost of turnover ≈ $9.9M
4. Preventing just 30% attrition among high-risk employees yields:
5. Net savings: ~$2.65M/year
6. ROI: ~479%

WHY RETENTION MATTER

High Cost of Employee Turnover

- Replacing an employee costs 50%–100% of annual salary
- Includes hiring, onboarding, lost productivity, and knowledge loss
- Current estimated cost of attrition: $9.9M per year

Productivity & Customer Impact

- Loss of experienced staff reduces project efficiency
- High churn damages customer relationships and delays delivery
- Remaining employees face increased workload → more future churn

Challenges with Current Approach

- HR teams often detect resignations too late
- No system to proactively flag talent at risk
- Retention budgets are used broadly, not targeted where impact is highest

PROBLEM DEFINITION

High Attrition is Driving Financial Loss

- Current annual attrition: 16.19%
- Cost to replace each employee: $41,534 on average
- Estimated total cost of turnover: ~$9.9M/year

Reactive Retention Strategy

- Employees only identified after resignation notice
- HR cannot prioritize who needs intervention most
- Limited ability to justify retention budgets with data

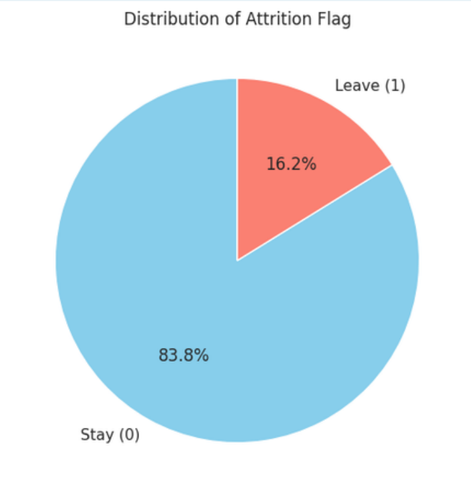
Key Attrition Pain Points

- Highest turnover in Sales, HR, and Lab Tech roles
- Strong relationship between stress, low flexibility, and quitting
- Early-career employees leave fastest → weak talent pipeline

DATA OVERVIEW

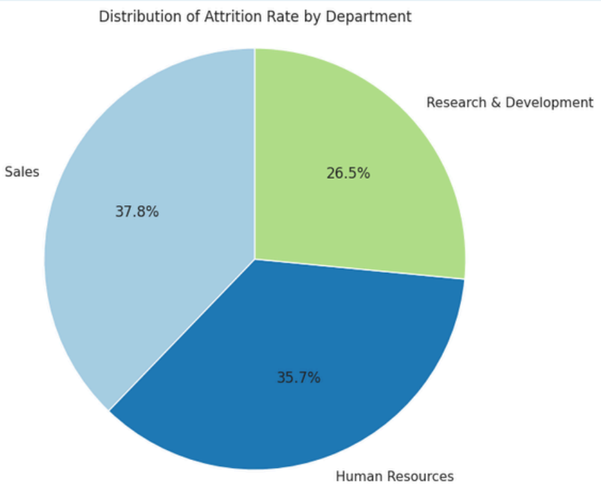
Target Variable

- Attrition
  - Leave = employee left
  - Stay = currently active employee
  - Imbalanced dataset:
    - 238 Leave (16.19%)
    - 1,232 Stay (83.81%)

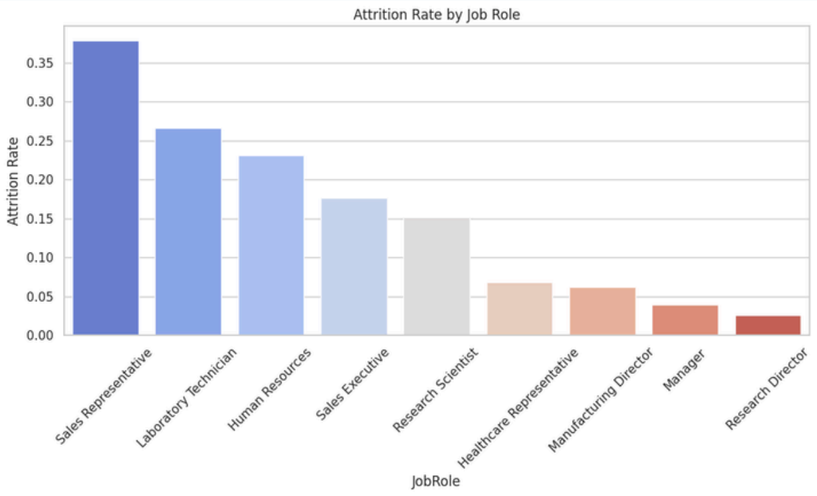


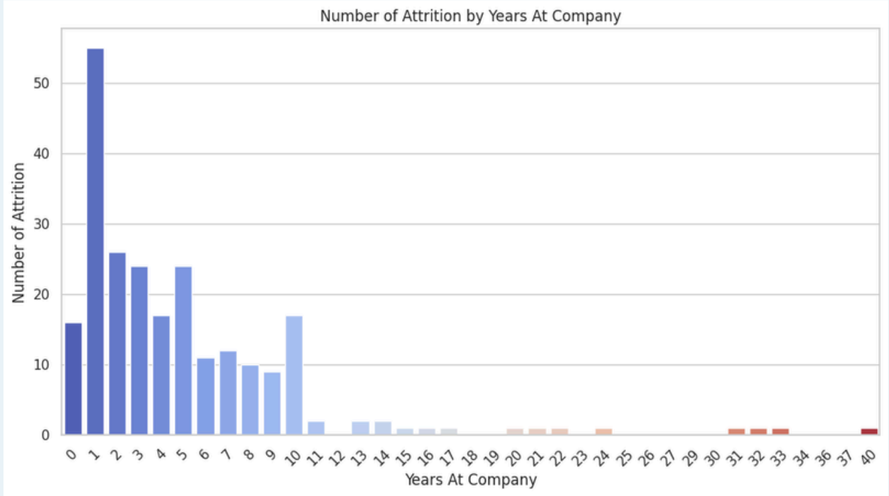
Attrition Rate each Department

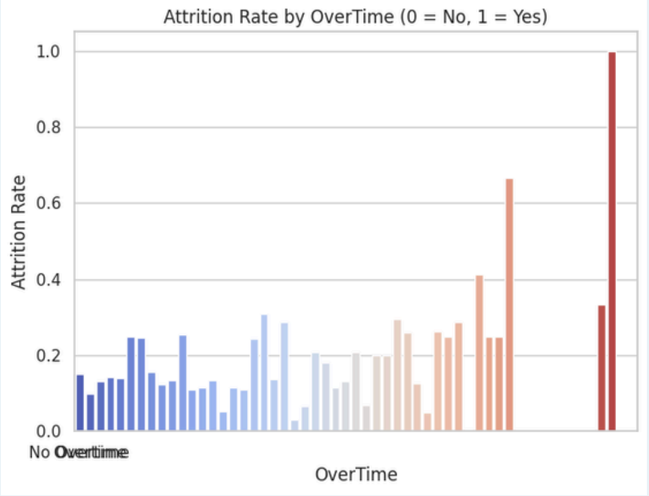
- Sales 20.17%
- HR 19.04%
- R&D 14.15%



KEY INSIGHTS FROM DATA EXPLORATION







High Attrition Concentrated in Specific Roles

- Roles with highest churn:
  - Laboratory Technician & Sales Executive
  - Account for ~45% of all resignations
- These roles show high workload + low job satisfaction scores

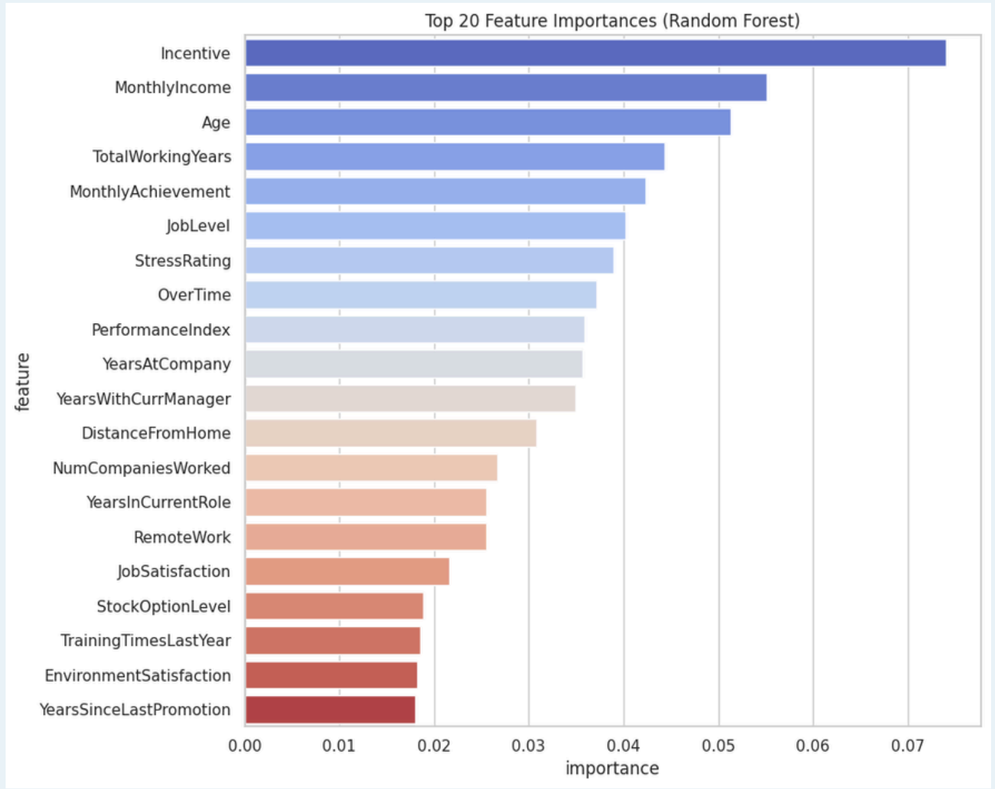
Early-career Attrition is Critical

- Employees with < 5 years at the company = most likely to leave
- Indicates challenges with onboarding, training, and career progression

Overtime Strongly Drives Turnover

- ~80% of employees who left worked frequent overtime
- Employees without overtime had significantly lower attrition

PREDICTIVE MODEL: KEY INSIGHTS



Interpretation of Feature Importance Chart

The Random Forest model highlights the top drivers of employee attrition:

Incentive (#1 strongest predictor)

- Employees with low incentive opportunities are significantly more likely to leave

Monthly Income

- Lower income → higher exit probability

Age & Total Working Years

- Younger professionals and those earlier in their career show higher turnover

Monthly Achievement & Job Level

- Employees with limited growth or lower achievement feel undervalued

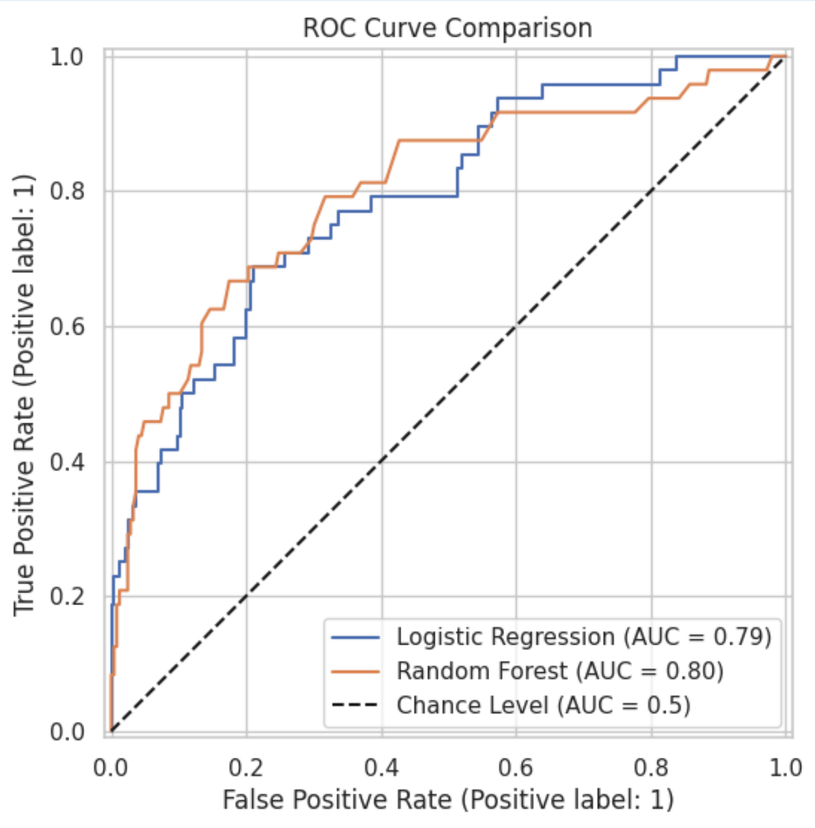
Stress Rating & OverTime

- High workload and poor work-life balance drive resignations

Other meaningfully contributing factors:

- YearsAtCompany – biggest risk early in tenure
- DistanceFromHome – long commute reduces satisfaction
- RemoteWork – flexibility reduces attrition
- YearsSinceLastPromotion – stagnant career affects morale
- Job Satisfaction & Environment Satisfaction – emotional wellbeing matters

PREDICTIVE MODEL: KEY INSIGHTS



ROC-AUC Model Performance Analysis

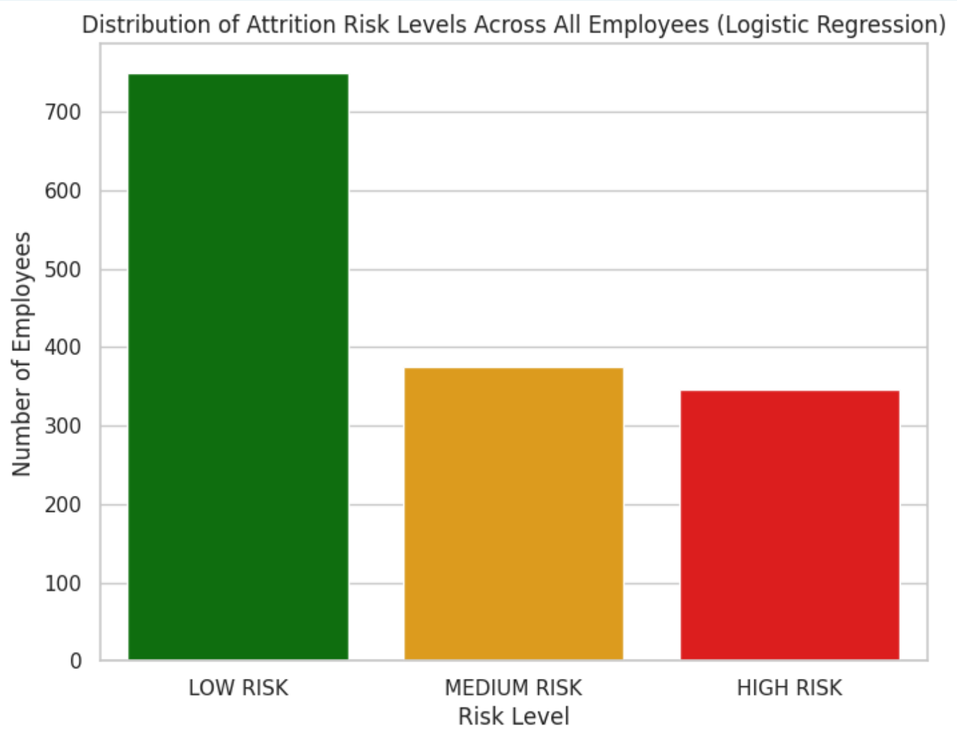
This ROC curve compares the predictive performance of two models:

- Logistic Regression has score 0.79
- Random Forest has score 0.80

Key Insights

- Both models perform significantly better than random chance
- Random Forest demonstrates stronger ability to detect employees at risk of leaving
- High recall in the Random Forest makes it suitable for proactive HR decisions
- Better to catch a risk early than miss it

EMPLOYEE RISK SEGMENTATION



Using the Logistic Regression model, each employee receives a predicted attrition risk score (0-1).

We classify employees (including leavers) into three actionable risk bands:

- if score < 0.4 = Low Risk (749 Employees)
- if score < 0.6 = Medium Risk (375 Employees)
- if score > 0.6 = High Risk (346 Employees)

If we look at the all employee compare with actual leaver:

- Low Risk → 749 / 40 which is 5.34%
- Medium Risk → 375 / 46 which is 12.26%
- High Risk → 346 / 152 which is 43.93%

Why It Matters

By focusing on only the High-Risk group, even a modest 30% retention success can save the business ~$4.3M annually.

PROPOSED DATA-DRIVEN RETENTION STRATEGY

Predictive Attrition Monitoring System

- Automated scoring for every employee (updated monthly/quarterly)
- Dashboard highlighting High-Risk individuals & teams
- HR receives alerts for sudden risk spikes

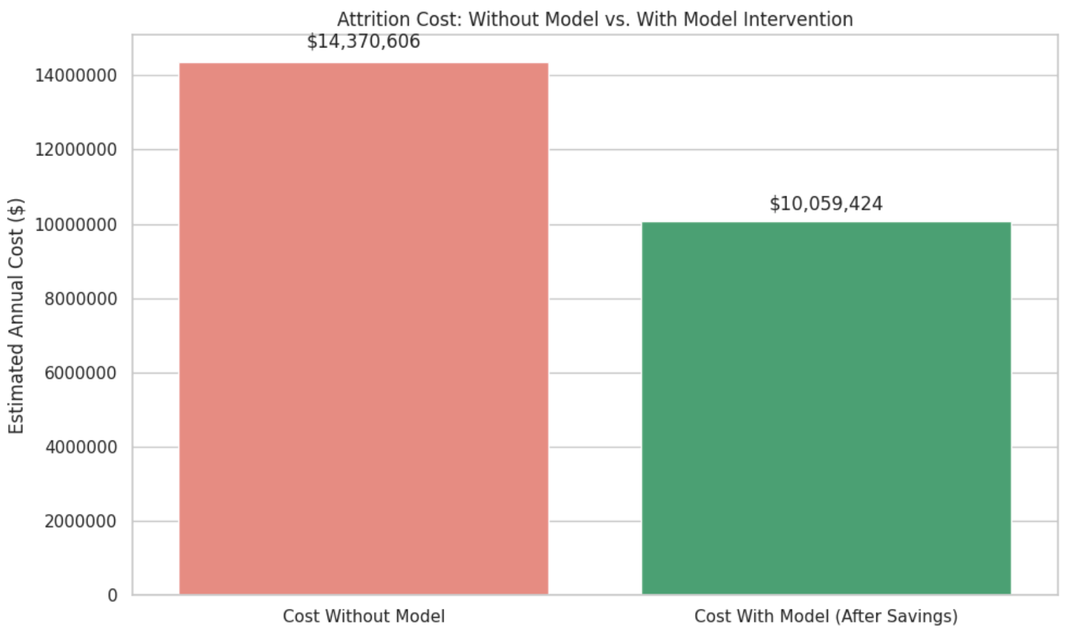
Targeted Retention Actions Focus on the top 20% high-risk employees with the highest ROI interventions:

- Flexible work (hybrid schedules, workload balancing)
- Compensation adjustments (performance bonuses, promotions)
- Career development (mentorship, skill training, clear pathways)
- Well-being programs (stress management, wellness support)

Priority Focus Areas

- Sales, HR, and Lab Technician roles
- Employees within first 3 years of tenure
- Younger employees with high overtime & low incentives

FINANCIAL IMPACT



Current Cost of Attrition (Doing Nothing)

- Estimated annual turnover cost: $14,370,606
- Driven by recruitment, onboarding, productivity loss, and knowledge drain

With Predictive Retention Model

- Targeted intervention focused on high-risk employees
- Annual turnover cost reduced to $10,059,424
- Savings: ~$4.31 Million per year

Interpretation

- Predictive analytics allows HR to invest retention budget strategically
- Preventing even a small portion of resignations creates major financial gains
- Business impact is measurable and immediate

RISKS & MITIGATION STRATEGY

Risk 1 – False Predictions (False Positives / Negatives)

- Model may sometimes misclassify employee risk

Mitigation Strategy

- Continuous model monitoring & retraining every quarter
- HR decision-making remains augmented, not automated
- Use confidence thresholds to reduce noise

Risk 2 – Employee Trust & Privacy Concerns

- Employees may fear being monitored unfairly

Mitigation Strategy

- Transparent communication: model is used to support employee wellbeing
- Strict data access controls (HR-only visibility)
- Comply with organizational data privacy policies

Risk 3 – HR Adoption Challenges

- Managers may ignore or misinterpret model insights

Mitigation Strategy

- User-friendly dashboards and training sessions
- HR guidance on what actions align to each risk factor
- Feedback loop between managers & analytics team

Risk 4 – Unequal Resource Allocation

- Focusing too much on high-risk employees may be questioned

Mitigation Strategy

- Align investments with ROI-driven retention strategy
- Monitor fairness and impact across all departments