FORGE LAUNCH | SPRING 2026

Predicting Employee Attrition: A Diagnostic Approach

From Descriptive Stats to Machine Learning Drivers

Generated by: Aaron Guo
Date: Jan 30, 2026

Baseline Analysis: Attrition aligns with industry norms (16.1%), providing a stable control group.

HISTORICAL ATTRITION
16.1%
Industry Avg: ~15%

POPULATION SCOPE
1,470
Full-Time Employees

DATA DIMENSIONS	FEATURES INCLUDED
Demographics	Age, Gender, Education, Marital Status
Engagement	JobSatisfaction, WorkLifeBalance, Involvement
Comp & Tenure	MonthlyIncome, YearsAtCompany, StockOptionLevel

Business Objective

Can we identify **preventable turnover** before it impacts the bottom line?

TARGET VARIABLE

```
Attrition == "Yes"
```

SUCCESS CRITERIA

- Identify >70% of at-risk employees
- Isolate actionable drivers (e.g., Overtime vs. Pay)

Data Source: IBM HR Analytics (Synthetic) via Kaggle
Note: Synthetic dataset used for blind benchmarking. No PII involved.

Data Snapshot: Fiscal Year 2025
Scope: All Regions

Methodology: Moving beyond correlation to isolate causal risk drivers.

1

Data Cleaning & Prep

Handled categorical variables (One-Hot Encoding) for 'Department' and 'JobRole'. Checked for class imbalance (Attrition 'Yes' is minority class).

2

Modeling Strategy

Moved from basic correlation matrices to **Logistic Regression**. This allowed me to isolate the *coefficient weights* of each feature to understand directionality.

3

Diagnostic Analytics

Distinguished between **Risk Accelerators** (drivers of exit) and **Protective Factors** (anchors of retention).

Exposure: High-risk employees represent a potential $7.7M liability to the organization.

Our model predicts significant exposure. High-risk employees are concentrated in R&D roles with specific tenure profiles.

TOTAL EMPLOYEES

1,470

84% Accuracy on Test Set

HIGH RISK SEGMENT

237

~16.1% of Workforce

EST. ATTRITION COST

$3.3M

Based on 50% Repl. Cost

Insight: While only 16% are "High Risk", this group accounts for a disproportionate amount of potential turnover cost.

Drivers: Manager Stability is the primary retention anchor; Overtime is the lead risk accelerator.

The model reveals distinct **Accelerators** and **Protectors**.

*Coefficients extracted from Logistic Regression model (C=1.0, L2 Penalty).

Roadmap: Operationalizing these insights into daily HR workflows

Prescriptive Analytics

Phase 2: The "What-If" Calculator

Move from *predicting* risk to *preventing* it. This tool will allow managers to toggle variables (e.g., "Reduce Travel") to instantly see the reduction in Risk Probability.

Deployment Strategy

Integration: The "Monday Warning"

Automate the model to run weekly. Push alerts to managers only when a high-performer crosses the **70% Risk Threshold**, ensuring proactive retention conversations.