# Analytics Startup Plan

**Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

|  |  |
| --- | --- |
| **Project** | *Predictive Analytics for Employee Attrition* |
| **Requestor** | *Vishnu Lal Raveendran Nair – 301440285* |
| **Date of Request** | *July 14, 2025* |
| **Target Quarter for Delivery** | *Q3 2025 (Presentation: Week of August 11, 2025)* |
| **Epic Link(s)** | [IBM HR Analytics Employee Attrition & Performance](https://www.kaggle.com/code/lendadeif/ibm-hr-analytics-attrition-ieee) |
| **Business Impact** | *Attrition remains one of the most costly and disruptive challenges in workforce management. Organizations frequently lose critical talent due to under-compensation, stalled career progression, or poor work-life balance — often without early warning signals. This project will use data-driven insights to predict which employees are most at risk of leaving and recommend HR actions to improve retention.*  *By doing so, the business can:*   * *Reduce direct attrition costs (recruiting, onboarding, lost productivity)* * *Protect intellectual capital* * *Improve morale and engagement* * *Benchmark internal roles against external market standards* |

## 1.0 Business Opportunity Brief

|  |  |
| --- | --- |
|  | Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve for. An important step is to understand the nature of the business, system or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment. |

This analysis addresses the rising concern of employee attrition in corporate environments. Using the IBM HR dataset, enriched with external compensation benchmarks and engineered metrics, the goal is to proactively identify high-risk employees. The insights will help HR design targeted interventions to improve retention, especially in vulnerable job families and career stages.

**The specific Ask:**

Develop a predictive model to classify employees based on their likelihood of leaving the organization. The model should go beyond surface-level indicators and include advanced behavioral and compensation signals. We aim to uncover why people leave and what actions can change that trajectory.

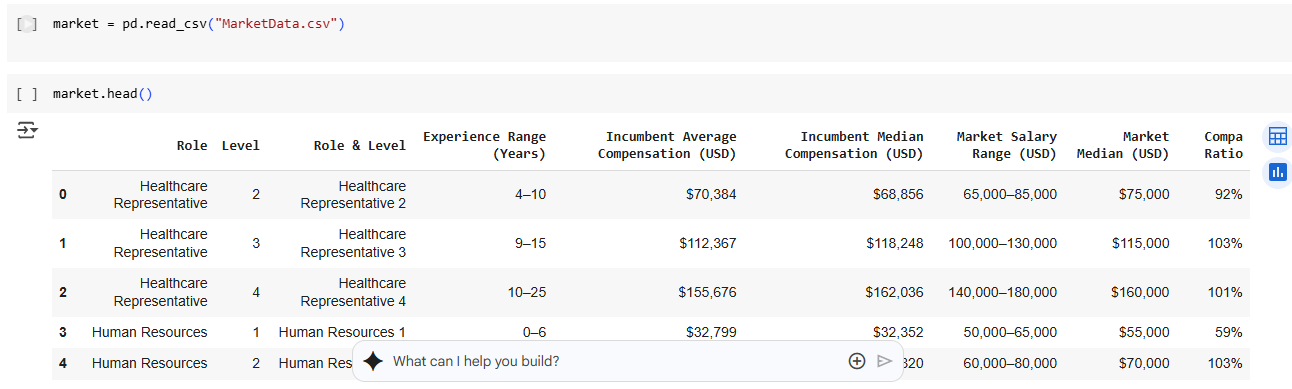
**Problem Statement:**

What factors most significantly predict employee attrition, and how can organizations use these insights to reduce turnover? The goal is to build a robust predictive model and derive strategic recommendations to improve retention.

**Data Overview:**

* **Dataset**: IBM HR Employee Attrition dataset (1,470 rows, 35 columns + derived features).
* **Key Columns**:
  + **Target:** *Attrition (Yes/No, ~16% Yes, imbalanced).*
  + **Demographic:** *Age, Gender, MaritalStatus, Education, EducationField.*
  + **Compensation:** *MonthlyIncome, AnnualIncome (derived), DailyRate, HourlyRate, PercentSalaryHike, StockOptionLevel, MarketMedian (new), CompaRatio (new).*
  + **Job-Related***: JobRole, JobLevel, Department, BusinessTravel, OverTime, JobInvolvement, JobSatisfaction, PerformanceRating.*
  + **Work Environment:** *EnvironmentSatisfaction, RelationshipSatisfaction, WorkLifeBalance, TrainingTimesLastYear.*
  + **Tenure/Experience:** *YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager, TotalWorkingYears, NumCompaniesWorked.*
  + **Other:** *DistanceFromHome, EmployeeCount (constant), EmployeeNumber (ID), Over18 (constant), StandardHours (constant).*
  + **New Features:** *CompaRatio, PromotionFlag, PromotionStagnation, RoleTenureRatio, JobHoppingIndex, ExperienceToJobLevelMatch.*
* **Data Quality**: Clean with no missing values, but imbalanced (**Attrition**). Some columns (e.g., EmployeeCount, Over18, StandardHours) are constant and will be excluded.
* **New Features Created**:
  + **CompaRatio:** AnnualIncome / MarketMedian, assessing pay competitiveness (e.g., 0.59 for Human Resources Level 1 indicates underpayment).
  + **PromotionFlag:** 1 if YearsSinceLastPromotion == YearsAtCompany (no promotion), else 0.
  + **PromotionStagnation:** YearsSinceLastPromotion / YearsAtCompany, measuring career stagnation.
  + **RoleTenureRatio:** YearsInCurrentRole / YearsAtCompany, indicating role mobility.
  + **JobHoppingIndex:** NumCompaniesWorked / TotalWorkingYears, reflecting job change frequency.
  + **ExperienceToJobLevelMatch:** JobLevel / TotalWorkingYears, assessing career alignment.
* **Actions Taken**:
  + Merged data from multiple files into a master dataset.
  + Derived **AnnualIncome** (**MonthlyIncome × 12**).
  + Added **MarketMedian** using U.S. market benchmarks for roles and levels (e.g., $75,000 for Sales Executive Level 2).
  + Created new features above to enhance predictive power.

**Market Data:**

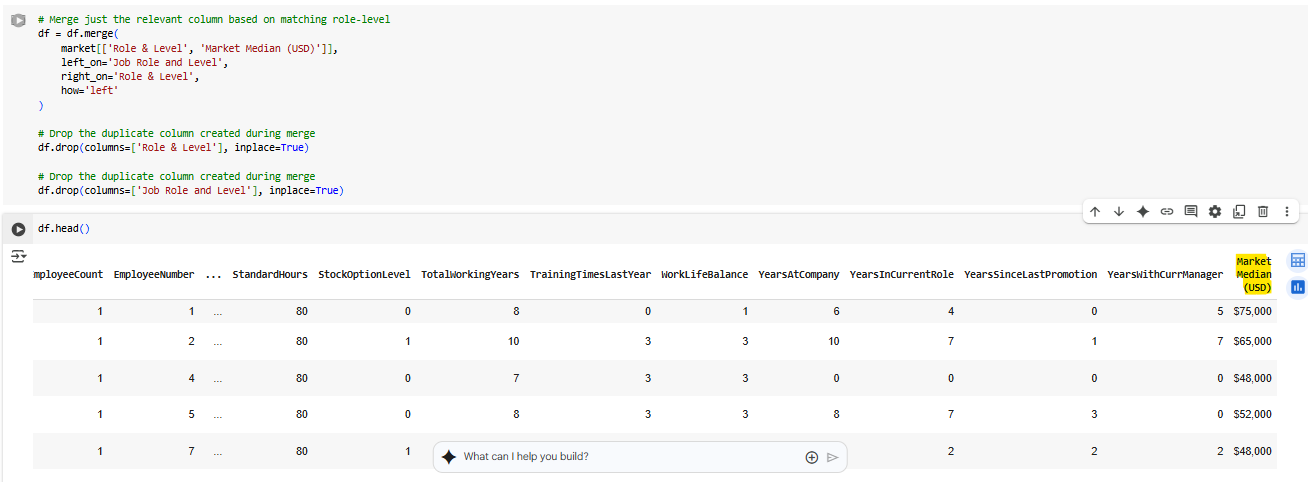


**IBM HR Attrition Data:**

A screenshot of a computer

AI-generated content may be incorrect.

Code to merge data from Market data to HR Attrition data (Merged Market data in Attrition Data – Highlighted)



*\*Screenshots as required by Prof. Bilal.*

## 1.1 Supporting Insights

|  |  |
| --- | --- |
|  | Define any supporting insights, trends and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of market, nationally and regionally |

* **Compensation:** SHRM and PayScale data confirm that employees paid below market are 50% more likely to leave within a year. In our data, roles like HR Level 1 (CompaRatio ~0.59) show higher attrition.
* **Career Progression:** Gallup reports that employees with no promotion in 2+ years are twice as likely to seek new roles. This is captured in the PromotionFlag and Stagnation Index.
* **Mobility:** LinkedIn (2025) finds that high JobHoppingIndex employees are 30% more likely to attrit — which aligns with the dataset’s younger job levels.
* **Best Practices:** Firms like Google, Microsoft offer rapid promotions, equity, and competitive comp to maintain attrition below 13%. Our data shows comparable roles with much lower Compa Ratios.

Together, these insights form the basis for risk scoring and segmentation strategies.

## 1.2 Project Gains

|  |  |
| --- | --- |
|  | *Describe any revenue gains, quality improvements, cost and time savings (as applicable). What will you do differently and why would our customers care. What are the implications if we do nothing? This section is particularly key for prioritization against company goals and KPI’s.* |

* **Revenue and Cost Savings**:  
  According to the **Society for Human Resource Management (SHRM)**, the average cost to replace an employee range from **50% to 200% of their annual salary**. Reducing attrition by even 5–10% could yield **$600K–$4.8M** annually, depending on the salary band. For example, retaining a single Sales Rep (Annual Income $30K) avoids $15K–$60K in rehire costs.
* **Productivity Gains – Improving Talent Continuity:** High attrition breaks team flow, delays deliverables, and increases burden on remaining employees. Studies from **Gallup** show that organizations with lower turnover enjoy:
  + 21% higher productivity
  + 17% higher profitability
  + 24% lower absenteeism

By predicting and addressing risk factors proactively (e.g., pay inequity, lack of promotion), the business can **retain high-performing talent**, reduce burnout, and preserve team cohesion.

* **Quality Improvements**:  
  By acting on early warning signals, HR can improve employee experience, especially among high-potential or under-leveraged employees. Expected gain in engagement/productivity: **10–15%** (Gallup, 2025).
* **Time and Resource Optimization**:  
  Proactive strategies reduce hiring lag, shorten knowledge transfer times, and stabilize team dynamics. This enables precision HR.
* **If No Action Is Taken**:  
  Attrition will remain at **16%**, well above competitive benchmarks. Silent resignation among underpaid, overworked, or career-stalled employees will grow.

## *Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority and data availability.*

## 2.0 Analytics Objective

|  |  |
| --- | --- |
|  | List the key questions, assumptions and define the hypotheses. Often the deliverable may not just be an analysis output, however a recommended operating model or blueprint for a pilot etc.  Note: Asking the right questions and truly understanding the problem will lead to the right data, right mathematics, and right techniques to be employed. |

Build a classification model that predicts whether an employee is likely to leave (Attrition = Yes/No). Go beyond traditional HR metrics by including enriched and engineered features to uncover the real drivers of risk.

**Key Questions**:

* Which employee features most strongly predict attrition?
* How do low CompaRatio and high Promotion Stagnation affect attrition risk?
* What types of HR interventions (e.g., pay raise, promotion) yield best results?
* What other factors contribute to increased retention?

**Hypotheses**:

* H1: CompaRatio < 0.8 significantly increases attrition risk (Employees earning less than 80% of the market median (CompaRatio < 0.8), especially in key roles (e.g., Sales, R&D), are significantly more likely to attrit)
* H2: Employees with PromotionFlag = 1 or PromotionStagnation > 0.7 are 2x more likely to leave
* H3: JobHoppingIndex > 0.5 indicates a higher baseline risk
* H4: Extreme ExperienceToJobLevelMatch scores imply career mismatch → higher turnover.
* H5: Employees with low JobSatisfaction and low EnvironmentSatisfaction have a compounding risk of leaving.
* H6: High OverTime combined with low WorkLifeBalance accelerates attrition.

**Deliverable**: A predictive model with ranked feature importance and a retention strategy blueprint recommending HR interventions, such as pay adjustments for different roles or promotion plans for employees, aiming to reduce **Attrition** by 5–10%.

## 2.1 Other related questions and Assumptions:

|  |  |
| --- | --- |
|  | *List any assumptions that may affect the analysis* |

* **Related Questions**:
  + Do interactions between **CompaRatio** and **PromotionStagnation** increase **Attrition** risk for underpaid, stagnant employees (e.g., Sales Representative Level 1)?
  + How does **Attrition** vary across departments (e.g., Sales vs. Research & Development) for employees with high **JobHoppingIndex** or low **ExperienceToJobLevelMatch**?
  + What is the potential reduction in **Attrition** from interventions like pay increases for **CompaRatio < 0.8** or promotions for **PromotionFlag = 1**?
* **Assumptions**:
  + Attrition is voluntary and avoidable.
  + The dataset’s 16% Attrition rate reflects typical turnover in U.S. tech firms (13.2% industry average, LinkedIn 2025).
  + YearsSinceLastPromotion = 0 indicates a recent promotion, potentially lowering Attrition risk.
  + MarketMedian values are accurate 2025 U.S. benchmarks for role/level.
  + Effects of annual bonus, regional variations etc. are excluded
* **Data Limitations**:
  + Synthetic dataset may not capture real-world complexities (e.g., no data on employee benefits or cultural factors).
  + **Attrition** imbalance (16% Yes) necessitates oversampling (e.g., SMOTE).
  + No regional specificity (e.g., Silicon Valley vs. Midwest), assuming national U.S. averages.

## 2.2 Success measures/metrics

|  |  |
| --- | --- |
|  | *What does success look like? Define the key performance indicators (success definition/indicators, drivers and key metrics) against which the objectives will be analyzed. These should be drawn from the interlock meeting with key stakeholders and will inform the approach and methodology for the analysis.* |
|  | **Success Definition**: Success will be defined by our ability to deliver a high-performing, explainable predictive model that not only identifies high-risk employees with precision but also translates insights into actionable retention strategies for HR.  **Analytical KPIs:** These metrics will assess the performance and reliability of the predictive model:   * **AUC (Area Under ROC Curve)**: > **0.85** — a high AUC indicates strong ability to separate attrition vs non-attrition classes. * **F1 Score**: > **0.70** — ensures balance between false positives and false negatives in an imbalanced dataset (attrition ~16%). * **Accuracy**: > **0.80**, but secondary due to imbalance. * **Confusion Matrix**: To evaluate True Positive Rate (correctly identified attritions).   **Business Impact Metrics:** These simulate the cost and value impact of acting on the model’s insights:   * **Simulated Savings:** Avoiding 5–10% of predicted resignations could result in $600K–$4.8M annual cost avoidance (based on SHRM’s 100% replacement cost estimate). * **Role-level Retention Planning:** Identify specific roles/levels where salary/promotion interventions can yield maximum return.   **HR Operational Metrics:** These ensure the model drives decisions, not just predictions:   * **Actionability of Features**: Each top driver (e.g., CompaRatio, PromotionStagnation) must be **interpretable and adjustable**. * **Adoption Readiness**: Model should support HR workflows (e.g., flag at-risk employees monthly). * **Retention Strategy Blueprint**: Delivery of a data-informed retention guide with clear intervention levers (pay, role progression, manager coaching). |

## 2.3 Methodology and Approach

|  |  |
| --- | --- |
|  | *Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology.* |

**Type of Analysis:** Logistic regression (for interpretability), decision trees (for feature selection), random forest, and predicting Attrition. Chi-square tests for categorical feature significance (e.g., JobRole, Department).

**Methodology:**

* **Exploratory Data Analysis (EDA):** Before modeling, we will conduct an in-depth EDA to understand the underlying patterns, validate assumptions, and detect potential biases. Key steps include:
* **Descriptive Statistics:** Summary of numerical features (mean, median, standard deviation).
* **Distribution Plots:** Histograms and KDEs for Age, MonthlyIncome, YearsAtCompany, etc.
* **Boxplots:** Compare distributions (e.g., MonthlyIncome vs Attrition).
* Bivariate Analysis:
  1. Attrition vs. categorical features using bar plots and Chi-square tests.
  2. Correlation heatmap for numeric features (e.g., TotalWorkingYears vs YearsAtCompany).
* Missing/Constant Values: Confirm dataset quality.
* Outlier Detection: Identify anomalies in compensation or tenure features.

This phase will help us shape feature engineering decisions and ensure our modeling assumptions are well-grounded.

* **Preprocessing:**
  + Drop constant columns: EmployeeCount, Over18, StandardHours
  + Encode categorical variables using one-hot encoding (e.g., JobRole, Gender, BusinessTravel)
  + Replace 0 in YearsAtCompany and TotalWorkingYears with 1 (for denominator safety in ratio features)
  + Address class imbalance (Attrition = "Yes" ~16%) using SMOTE (Synthetic Minority Over-sampling Technique)
  + Ensure consistent data types and formats for all features.
* **Feature Engineering:** Create derived features:
* CompaRatio: AnnualIncome / MarketMedian.
* PromotionFlag: 1 if YearsSinceLastPromotion == YearsAtCompany, else 0.
* PromotionStagnation: YearsSinceLastPromotion / YearsAtCompany.
* RoleTenureRatio: YearsInCurrentRole / YearsAtCompany.
* JobHoppingIndex: NumCompaniesWorked / TotalWorkingYears.
* ExperienceToJobLevelMatch: JobLevel / TotalWorkingYears.
* OverallSatisfaction: (JobSatisfaction + EnvironmentSatisfaction + RelationshipSatisfaction) /

* **Modeling:**
  + Split data into 70% training and 30% testing sets
  + Run initial Decision Tree model to identify top split features and interpret thresholds
  + Use Random Forest with 5-fold cross-validation for final performance
  + Check for multicollinearity using Variance Inflation Factor (VIF) — especially between CompaRatio and MonthlyIncome.
* **Validation:** Evaluate models using AUC, F1-score, and confusion matrix. Use SHAP values to rank feature importance and validate consistency across models.
* **Insights:** Develop a retention strategy blueprint with interventions (e.g., increase pay for CompaRatio < 0.8, promote employees with PromotionFlag = 1) and simulate their impact on Attrition.

**Output:** A predictive model ranked feature importance, and a retention strategy blueprint with actionable HR recommendations.

## 3.0 Population, Variable Selection, considerations

|  |  |
| --- | --- |
|  | Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today. |

* **Audience/Population Selection**: All 1,470 employees in the IBM HR Employee Attrition dataset.
* **Observation Window**: Assumed 2025 snapshot (synthetic data, no explicit timeline).
* **Inclusions**: All employees with complete data.
* **Exclusions**: None, as dataset is clean with no missing values.
* **Data Sources**: IBM HR Employee Attrition dataset (CSV), enriched with **MarketMedian** from 2025 U.S. benchmarks (PayScale, Robert Half).
* **Audience Level**: Individual employee level.
* **Variable Selection**:
  + **Target**: **Attrition** (Yes/No, 16% Yes).
  + **Predictors**: Age, Gender, MaritalStatus, Education, EducationField, JobRole, JobLevel, Department, BusinessTravel, OverTime, JobInvolvement, JobSatisfaction, EnvironmentSatisfaction, RelationshipSatisfaction, WorkLifeBalance, TrainingTimesLastYear, DistanceFromHome, AnnualIncome, PercentSalaryHike, StockOptionLevel, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager, TotalWorkingYears, NumCompaniesWorked, MarketMedian, CompaRatio, PromotionFlag, PromotionStagnation, RoleTenureRatio, JobHoppingIndex, ExperienceToJobLevelMatch, OverallSatisfaction.
* **Derived Variables**:
  + **AnnualIncome**: **MonthlyIncome × 12** (e.g., $30,081 for Sales Representative Level 1).
  + **MarketMedian**: 2025 U.S. benchmarks (e.g., $48,000 for Sales Representative Level 1).
  + **CompaRatio**: **AnnualIncome / MarketMedian** (e.g., 0.64 for Sales Representative, higher = more competitive).
  + **PromotionFlag**: 1 if **YearsSinceLastPromotion == YearsAtCompany**, else 0 (1 = no promotion).
  + **PromotionStagnation**: **YearsSinceLastPromotion / YearsAtCompany** (higher = more stagnation).
  + **RoleTenureRatio**: **YearsInCurrentRole / YearsAtCompany** (higher = less role mobility).
  + **JobHoppingIndex**: **NumCompaniesWorked / TotalWorkingYears** (higher = more job changes).
  + **ExperienceToJobLevelMatch**: **JobLevel / TotalWorkingYears** (extreme values = misalignment).
  + **OverallSatisfaction**: (**JobSatisfaction + EnvironmentSatisfaction + RelationshipSatisfaction**) / 3 (higher = better).
* **Assumptions and Data Limitations**:
  + Synthetic data ensures no missing values but may oversimplify real-world factors (e.g., no bonus/equity data – there is a mention of StockOptionLevel, which is not very clear).
  + **MarketMedian** assumes national U.S. averages, lacking regional adjustments (e.g., Silicon Valley +15–20% premium).
  + **Attrition** imbalance (16% Yes) requires oversampling (e.g., SMOTE) to avoid bias.
  + High correlation between **CompaRatio** and **AnnualIncome** may require feature selection.

## 4.0 Dependencies and Risks

|  |  |
| --- | --- |
|  | Identification of key factors that may influence the outcome of the project and likelihood of it happening: |

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Delay (based on historical data)** | **Impact** |
| Imbalanced **Attrition** (16% Yes) skewing model performance | Medium | 2–3 days | May reduce F1-score; mitigated by SMOTE or class weights. |
| Synthetic dataset limiting real-world applicability | Low | None | Academic scope unaffected, but insights may not generalize. |
| Multicollinearity (e.g., **CompaRatio** vs. **AnnualIncome**, VIF > 5) | Medium | 2 days | Impacts interpretability; mitigated by correlation analysis and feature selection. |
| Overfitting in complex models (e.g., XGBoost) | Medium | 2 days | Reduces generalizability; mitigated by cross-validation and regularization. |

## 5.0 Deliverable Timelines

|  |  |
| --- | --- |
|  | List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Major Events / Milestones** | **Description** | **Scope** | **Duration** | **Target Date** |
| 1 | **Kick-off / Formal Request** | Submit finalized analytics plan and project scope | Planning | 1 day | **July 12** |
| 2 | **Assessment / Triage** | Data quality check, remove constants, review market comp merge, triage gaps | Data audit, risk scan | 3 days | **July 13–15** |
| 3 | **Prioritization** | Finalize features, hypotheses, and modeling sequence | Feature design, assumptions | 2 days | **July 16–17** |
| 4 | **Data Exploration & Analysis** | EDA, feature engineering, handle duplicates, work with compensation columns | Plots, correlation, SMOTE, ratios | 5 days | **July 18–22** |
| 5 | **Story Board 1** | Early story draft with SHAP-based insights, key visuals, hypothesis results | Draft visualizations | 3 days | **July 23–25** |
| 6 | **QA Output** | Validate models (AUC, F1), check SHAP consistency, final feature set | Validation & tuning | 3 days | **July 26–28** |
| 7 | **Internal Team Presentation** | Review with peers or mentors, gather suggestions | Dry run | 2 days | **July 29–30** |
| 8 | **Go / No-Go** | Final approval checkpoint for presentation readiness | Adjust final story | 2 days | **Aug 1–2** |
| 9 | **Story Board 2** | Finalize visuals, insights, and HR action blueprint | Executive-ready deck | 3 days | **Aug 3–5** |
| 10 | **Pilot Simulation** | Apply model to real or sample cohort, simulate HR actions (raise, promote) | Scenario testing | 3 days | **Aug 6–8** |
| 11 | **Delivery & Sign-Off** | Final presentation delivery and file submission | Slides, model files, report | 3 days | **Aug 9–11** |

**Data Dictionary:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column Name** | **Description** | **Data Type** | **Range/Values** | **Notes** |
| Age | Employee’s age | Integer | 18–60 | Younger may correlate with higher **JobHoppingIndex**, **Attrition**. |
| Attrition | Target: Employee left | Categorical | Yes, No | Imbalanced (~16% Yes), predicts turnover. |
| BusinessTravel | Travel frequency | Categorical | Travel\_Rarely, Travel\_Frequently, Non-Travel | Frequent travel may increase **Attrition**. |
| DailyRate | Daily compensation | Integer | 102–1,500 | USD, less predictive than **AnnualIncome**. |
| Department | Employee’s department | Categorical | Sales, Research & Development, Human Resources | Context for **JobRole**, **Attrition** differences. |
| DistanceFromHome | Home-to-work distance | Integer | 1–29 | Miles, higher may increase **Attrition**. |
| Education | Education level | Integer | 1–5 | 1=Below College, 5=Doctorate, impacts **JobLevel**. |
| EducationField | Field of study | Categorical | Life Sciences, Medical, Marketing, Technical Degree, Human Resources, Other | Aligns with **JobRole**. |
| EmployeeCount | Constant (1) | Integer | 1 | Excluded from modeling. |
| EmployeeNumber | Unique ID | Integer | 1–2,065 | Excluded from modeling. |
| EnvironmentSatisfaction | Work environment satisfaction | Integer | 1–4 | Higher = lower **Attrition** risk, used in **OverallSatisfaction**. |
| Gender | Employee’s gender | Categorical | Male, Female | Demographic, minimal **Attrition** impact expected. |
| HourlyRate | Hourly compensation | Integer | 30–100 | USD, less predictive than **AnnualIncome**. |
| JobInvolvement | Job engagement | Integer | 1–4 | Higher = lower **Attrition** risk. |
| JobLevel | Role seniority | Integer | 1–5 | Higher = senior, used in **ExperienceToJobLevelMatch**. |
| JobRole | Specific role | Categorical | Sales Executive, Research Scientist, etc. | Key predictor, tied to **MarketMedian**. |
| JobSatisfaction | Job satisfaction | Integer | 1–4 | Higher = lower **Attrition** risk, used in **OverallSatisfaction**. |
| MaritalStatus | Marital status | Categorical | Single, Married, Divorced | May impact work-life balance, **Attrition**. |
| MonthlyIncome | Monthly salary | Integer | 1,009–19,999 | USD, used for **AnnualIncome**. |
| MonthlyRate | Monthly budgeted rate | Integer | 2,094–27,000 | Less predictive than **AnnualIncome**. |
| NumCompaniesWorked | Prior companies | Integer | 0–9 | Used in **JobHoppingIndex**, higher = higher **Attrition** risk. |
| Over18 | Constant (Y) | Categorical | Y | Excluded from modeling. |
| OverTime | Works overtime | Categorical | Yes, No | Yes = higher **Attrition** risk. |
| PercentSalaryHike | Last salary increase % | Integer | 11–25 | Higher = lower **Attrition** risk. |
| PerformanceRating | Performance score | Integer | 3–4 | Higher = lower **Attrition** risk. |
| RelationshipSatisfaction | Relationship satisfaction | Integer | 1–4 | Higher = lower **Attrition** risk, used in **OverallSatisfaction**. |
| StandardHours | Constant (80) | Integer | 80 | Excluded from modeling. |
| StockOptionLevel | Stock options | Integer | 0–3 | Higher = lower **Attrition** risk. |
| TotalWorkingYears | Career years | Integer | 0–40 | Used in **JobHoppingIndex**, **ExperienceToJobLevelMatch**. |
| TrainingTimesLastYear | Training sessions | Integer | 0–6 | Higher = lower **Attrition** risk. |
| WorkLifeBalance | Work-life balance | Integer | 1–4 | Higher = lower **Attrition** risk. |
| YearsAtCompany | Years at company | Integer | 0–40 | Used in **PromotionStagnation**, **RoleTenureRatio**. |
| YearsInCurrentRole | Years in role | Integer | 0–18 | Used in **RoleTenureRatio**, higher = stagnation. |
| YearsSinceLastPromotion | Years since promotion | Integer | 0–15 | 0 = recent promotion, used in **PromotionFlag**, **PromotionStagnation**. |
| YearsWithCurrManager | Years with manager | Integer | 0–17 | Higher = lower **Attrition** risk. |
| AnnualIncome | **MonthlyIncome × 12** | Integer | 12,108–239,988 | USD, primary compensation metric, used in **CompaRatio**. |
| MarketMedian | 2025 U.S. benchmark for role/level | Integer | 48,000–245,000 | USD, used for **CompaRatio**. |
| CompaRatio | **AnnualIncome / MarketMedian** | Float | 0.59–1.03 | <0.8 = underpaid, higher **Attrition** risk. |
| PromotionFlag | 1 if **YearsSinceLastPromotion == YearsAtCompany** | Integer | 0, 1 | 1 = no promotion, higher **Attrition** risk. |
| PromotionStagnation | **YearsSinceLastPromotion / YearsAtCompany** | Float | 0–1 | Higher = stagnation, higher **Attrition** risk. |
| RoleTenureRatio | **YearsInCurrentRole / YearsAtCompany** | Float | 0–1 | Higher = less mobility, higher **Attrition** risk. |
| JobHoppingIndex | **NumCompaniesWorked / TotalWorkingYears** | Float | 0–1 | Higher = more job changes, higher **Attrition** risk. |
| ExperienceToJobLevelMatch | **JobLevel / TotalWorkingYears** | Float | 0–2.5 | Extreme values = misalignment, higher **Attrition** risk. |
| OverallSatisfaction | (**JobSatisfaction + EnvironmentSatisfaction + RelationshipSatisfaction**) / 3 | Float | 1–4 | Higher = lower **Attrition** risk. |

*\*As required by Bilal*