Table of Contents

[What predicts employee attrition? 1](#_Toc192489320)

[Business Objective 1](#_Toc192489321)

[Notebook path 1](#_Toc192489322)

[Observations 1](#_Toc192489323)

[Initial Observations: 1](#_Toc192489324)

[Cleaned Dataset Summary 1](#_Toc192489325)

[Key Findings from Visualizations 1](#_Toc192489326)

[Target variable analysis 1](#_Toc192489327)

[Correlation Matrix 2](#_Toc192489328)

[Numerical features vs Target (Attrition) 5](#_Toc192489329)

[Categorical features vs Target (Attrition) 23](#_Toc192489330)

[Key Findings & Observations 31](#_Toc192489331)

# What predicts employee attrition?

## Business Objective

Employee attrition poses a significant challenge for organizations, leading to high recruitment costs, loss of expertise, and reduced team productivity. This study aims to identify key factors that predict employee attrition and develop a machine learning model to accurately flag high-risk employees.

Using an internal HR dataset containing demographic, job role, and performance data, we will apply machine learning techniques such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and AdaBoost. Feature importance analysis and SHAP (SHapley Additive exPlanations) values will help interpret the most influential predictors, such as job satisfaction, salary, and tenure.

By uncovering these insights, HR teams can implement targeted retention strategies, reducing turnover costs and fostering a more engaged workforce. This data-driven approach will enable proactive decision-making, ensuring long-term organizational stability and employee satisfaction.

## **Notebook path**

## **Observations**

### **Initial Observations:**

**Dataset Overview**

* **Total Rows:** 23,436
* **Total Columns:** 37, 19 numeric columns and 18 categorical columns
* **Data Types:** A mix of numerical (float64) and categorical (object) columns.
* **Warning:** Some columns have mixed data types (DistanceFromHome, EmployeeCount, HourlyRate, etc.), which may require data cleaning.

**Missing Values**

* Several columns have missing values, with counts ranging from 3 to 15 missing entries.
* Columns such as Attrition, DailyRate, MonthlyIncome, and YearsInCurrentRole have missing data, which should be handled during preprocessing.

**Potential Data Quality Issues**

* **Unusual Values:** EnvironmentSatisfaction and PerformanceRating have extreme maximum values (e.g., 129,588 and 13, respectively), which may indicate data entry errors.
* **High Variance:** NumCompaniesWorked has an extreme max value (23,258), which seems unrealistic.
* **Duplicate or Redundant Columns:** EmployeeCount seems to have a constant value of 1, making it redundant.

**Statistical Summary**

* **Age Distribution:** Ranges from 18 to 60, with a mean age of ~37.
* **Total Working Years:** Averages ~11 years, with a max of 40.
* **Years at Company:** Mean of ~7 years, with some employees staying up to 40 years.
* **Work-Life Balance:** Averages around 2.76 (on a 1–4 scale).

### **Cleaned Dataset Summary**

**Dataset Size:**

* **Total Records:** 23,188 (after removing duplicates and missing values)
* **Total Columns:** 33, 25 numeric and 8 categorical columns

**Data Quality Improvements:**

* Removed all missing values.
* Eliminated duplicate records.
* Dropped two records with invalid ApplicationID values.
* Converted categorical columns (DistanceFromHome, HourlyRate, MonthlyIncome, PercentSalaryHike, JobSatisfaction, ApplicationID) to int64.
* Removed StandardHours, EmployeeNumber, EmployeeCount, and Over18 as they contained only unique values.
* Encoded the target variable:
  + 'Current employee' → 0
  + 'Voluntary Resignation' → 1
* Ensured all columns now contain complete data.

**Statistical Summary:**

* **Age:** Ranges from 18 to 60, with a mean of ~37.
* **Total Working Years:** Averages ~11 years, with a max of 40.
* **Years at Company:** Mean of ~7 years, ranging from 0 to 40.
* **Job Level:** Median is 2, with a max of 5.
* **Work-Life Balance:** Median of 3 (on a scale of 1-4).
* **Attrition:** Cleaned dataset maintains all attrition-related data.

**Key Observations:**

* PerformanceRating max value is now 4 (previously had potential outliers).
* EnvironmentSatisfaction has values between 1-4, with a mean of 2.72.
* StockOptionLevel now ranges from 0-3 (previous max was 80, indicating prior data issues).

## **Key Findings from Visualizations**

### **Target variable analysis**

A graph of a graph showing a number of employees

AI-generated content may be incorrect.

Key observations from the chart:

1. **Class Imbalance**:
   * The majority of employees are **current employees** (significantly higher count).
   * A smaller proportion has **voluntarily resigned**.
2. **Potential Impact**:
   * If this dataset is used for predictive modeling, class imbalance could affect model performance.
   * Oversampling (SMOTE) or undersampling techniques may be required to balance the dataset.

### Correlation Matrix

Correlation Heatmap of Numerical Features

A graph with numbers and a red line

AI-generated content may be incorrect.

The correlation heatmap provides insights into the relationships between numerical features in the dataset. Here are some key takeaways:

**1. Strong Positive Correlations**

* **Monthly Income & Job Level (~0.94)**: Higher job levels are strongly associated with higher monthly income.
* **Monthly Income & Monthly Rate (~0.94)**: Employees earning higher monthly income also have higher monthly rates.
* **Total Working Years & Job Level (~0.77)**: Employees with more experience tend to have higher job levels.
* **Years at Company & Years in Current Role (~0.76)**: Employees who have been in the company longer tend to stay in the same role for an extended period.
* **Years in Current Role & Years Since Last Promotion (~0.72)**: Employees in the same role for longer periods are likely to have gone longer without a promotion.

**2. Moderate Positive Correlations**

* **Years at Company & Total Working Years (~0.63)**: Employees with longer total work experience tend to have been at the company for a longer time.
* **Years with Current Manager & Years in Current Role (~0.77)**: Employees who have been in a role for a long time likely have had the same manager for a long time.

**3. Weak or Negligible Correlations**

* **DailyRate & Most Other Variables (~close to 0)**: The daily rate of pay does not show strong relationships with other features.
* **Distance from Home & Attrition (~-0.07)**: The correlation is weak, meaning distance may not be a strong predictor of attrition.
* **Performance Rating & Most Variables (~0.01 - 0.04)**: Performance ratings seem to have little correlation with other attributes.

**4. Interesting Insights Related to Attrition**

* **Attrition & Age (~-0.15)**: Younger employees are slightly more likely to leave, but the correlation is weak.
* **Attrition & Total Working Years (~-0.17)**: Employees with more experience are slightly less likely to leave.

**Final Thoughts**

* The **strongest correlations** are primarily between salary-related features, job level, and experience.
* **Attrition-related correlations are weak**, suggesting multiple factors influence an employee’s decision to leave.
* Further analysis with models like logistic regression or decision trees could reveal non-linear patterns.

### **Numerical features vs Target (Attrition)**

1. **Age vs Attrition**

A graph showing an attrition and an attrition

AI-generated content may be incorrect.

* Younger employees are more likely to resign, possibly for better opportunities.
* Older employees tend to stay, likely due to career stability or company benefits.
* The company might need retention strategies focused on younger employees.

1. **DailyRate vs Attrition**

A graph showing a comparison between an attrition and an attrition

AI-generated content may be incorrect.

* The **median DailyRate is slightly lower for employees who resigned** compared to current employees.
* **DailyRate alone may not be a strong factor** influencing attrition.

1. **DistanceFromHome vs Attrition**

A graph showing a couple of rectangular boxes

AI-generated content may be incorrect.

* **Longer commutes might contribute to higher attrition** due to work-life balance issues.
* Employees closer to work may be more **satisfied and less likely to leave**.
* The **company could consider remote work or relocation benefits** for employees with long commutes.

1. **Education vs Attrition**

A diagram of an attrition and attrition

AI-generated content may be incorrect.

* **Education does not appear to be a significant factor** in predicting attrition.
* Employees with higher education levels **do not show a clear trend of leaving or staying**.

1. **EnvironmentSatisfaction vs Attrition**

A graph showing different colored squares

AI-generated content may be incorrect.

* **Dissatisfaction with the environment does not seem to be a major factor in attrition.**

1. **HourlyRate vs Attrition**

A graph showing an attrition and attrition

AI-generated content may be incorrect.

* Hourly pay alone does not appear to drive attrition.

1. **JobInvolvement vs Attrition**

A graph showing a graph of an attrition

AI-generated content may be incorrect.

* **Employees who resigned were not necessarily less involved in their work.**

1. **JobLevel vs Attrition**

A graph showing a job level

AI-generated content may be incorrect.

* **Lower-level employees may leave more frequently**, possibly due to:
  + **Better external job opportunities**.
  + **Limited growth or promotion opportunities**.
  + **Lower salaries compared to higher job levels**.
* **Higher-level employees tend to stay**, possibly due to better pay and stability.

1. **JobSatisfaction vs Attrition**

A graph showing different colored squares

AI-generated content may be incorrect.

* **Job satisfaction alone does not strongly correlate with attrition.**
* **Employees might be leaving for other reasons, such as:** 
  + **Compensation (e.g., MonthlyIncome, JobLevel)**
  + **Career Growth (e.g., YearsAtCompany, Promotions)**
  + **Work-Life Balance**

1. **MonthlyIncome vs Attrition**

A graph showing an individual income

AI-generated content may be incorrect.

* **Lower-income employees may be more likely to leave for higher-paying jobs**.
* Employees in **higher income brackets tend to stay longer**.
* **Compensation might be a strong factor influencing attrition**.

1. **MonthlyRate vs Attrition**

A diagram of an employee attrition

AI-generated content may be incorrect.

* **MonthlyRate does not seem to be a strong predictor of attrition.**

1. NumCompaniesWorked vs Attrition

A graph showing a diagram

AI-generated content may be incorrect.

* Employees with a history of switching jobs frequently are more likely to resign.
* This suggests that past job changes can indicate a higher likelihood of future attrition.
* Employees with fewer job changes may value stability and are less likely to leave.

1. PercentSalaryHike vs Attrition

A graph showing a graph of a person's salary

AI-generated content may be incorrect.

* Employees may **resign despite salary hikes**, likely due to:
  + **Better external job offers**
  + **Dissatisfaction with role or growth opportunities**
  + **Work-life balance issues**
* Salary hikes may help **reduce attrition**, but they are **not the sole factor** in retaining employees.

1. PerformanceRating vs Attrition

A graph showing a performance rating

AI-generated content may be incorrect.

* Employees may be **leaving for reasons unrelated to performance**, such as:
  + **Compensation differences** (e.g., MonthlyIncome, SalaryHike)
  + **Lack of career growth** (e.g., JobLevel, Promotions)
  + **Work-life balance issues**
* Performance evaluations **may not be a key factor in attrition** decisions.

1. RelationshipSatisfaction vs Attrition

A graph showing a couple of colored squares

AI-generated content may be incorrect.

* Employees **are not necessarily leaving due to relationship satisfaction issues**.
* Other factors like **salary, career growth, work-life balance, or job satisfaction** may have a stronger influence on attrition.
* This feature might have **low predictive power for attrition**.

1. StockOptionLevel vs Attrition

A chart with a graph and text

AI-generated content may be incorrect.

* **Stock options do not appear to be a strong predictor of attrition**.
* Employees may value **salary, promotions, work-life balance, or job satisfaction more** when deciding to leave.
* **Other incentives beyond stock options might be more effective** for retention.

1. TotalWorkingYears vs Attrition

A graph showing an attrition and an attrition

AI-generated content may be incorrect.

* **Newer employees (fewer working years) may leave more frequently**, likely due to:
  + Seeking **better opportunities elsewhere**.
  + **Lack of career progression** or dissatisfaction with current growth.
* **Longer-tenured employees tend to stay**, suggesting **stability increases with experience**.

1. TrainingTimesLastYear vs Attrition

A graph with a bar chart and a row of rectangular objects

AI-generated content may be incorrect.

* **Lack of training does not appear to be a major factor in attrition**.
* Employees **who received more training (5-6 sessions) still resigned**, suggesting **other factors like career growth or promotions may play a bigger role**.
* Training alone may not **retain employees if they don’t see clear career advancement opportunities**.

1. WorkLifeBalance vs Attrition

A graph showing a diagram of work life balance

AI-generated content may be incorrect.

* **Work-life balance is not a strong driver of attrition**.
* Employees may be **leaving for reasons other than work-life balance**, such as:
  + **Compensation (Salary, JobLevel)**
  + **Career Growth (Promotions, JobSatisfaction)**
  + **Company culture or management issues**
* Work-life balance **may be a factor but is not the primary reason for voluntary resignations**.

1. YearsAtCompany vs Attrition

A graph showing an attrition and attrition

AI-generated content may be incorrect.

* **Newer employees (fewer years at the company) are more likely to leave**, possibly due to:
  + **Limited career growth opportunities**.
  + **Better job offers elsewhere**.
  + **Dissatisfaction with role or compensation**.
* **Employees who have been with the company longer tend to stay**, indicating a sense of stability or loyalty.

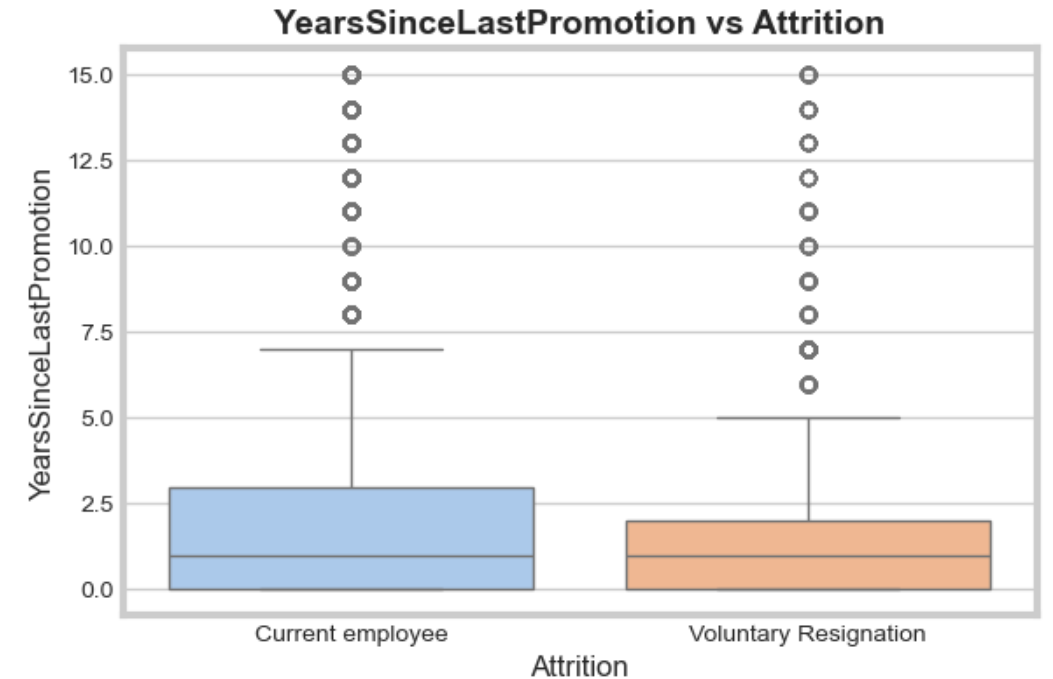
1. YearsInCurrentRole vs Attrition

A graph showing an attrition and an attrition

AI-generated content may be incorrect.

* Employees **who have been in their current role for a short period are more likely to leave**, possibly due to:
  + **Lack of growth opportunities or career advancement.**
  + **Dissatisfaction with their current position.**
  + **Seeking new challenges elsewhere.**
* Employees **who have been in the same role for many years tend to stay**, likely due to job stability and familiarity.

1. YearsSinceLastPromotion vs Attrition



* **Lack of promotion does not seem to be the primary reason for resignation.**
* Employees with **longer waits for promotion don’t necessarily leave**; they may **value stability or other factors like salary and job satisfaction**.
* Those who resign may be leaving for **better opportunities elsewhere** rather than waiting for a promotion

1. YearsWithCurrManager vs Attrition

A graph with a blue and orange rectangular object

AI-generated content may be incorrect.

* **Short tenure under a manager is correlated with higher attrition**, possibly due to:
  + **Poor managerial relationships** leading to early exits.
  + **Lack of mentorship or guidance** from new managers.
  + **Employees switching teams frequently before leaving.**
* Employees who **stay with the same manager for many years tend to remain with the company**, indicating a **stronger sense of stability and job satisfaction**.

### **Categorical features vs Target (Attrition)**

A graph of different colored squares

AI-generated content may be incorrect.

* **Frequent travelers have higher attrition** – Employees who travel frequently are more likely to resign, possibly due to work stress and poor work-life balance.
* **Rare travelers have the largest population** – While they experience some attrition, the proportion is lower compared to frequent travelers.
* **Non-travelers have the lowest attrition** – Employees who don’t travel tend to stay, suggesting job stability reduces turnover.

A graph of a graph showing different types of employment

AI-generated content may be incorrect.

* **Sales and R&D have the highest attrition** – Both departments show a significant number of voluntary resignations, with Sales having a relatively higher proportion.
* **Human Resources has the lowest attrition** – Fewer employees in HR are leaving, indicating better job stability.
* **Research & Development is the largest department** – The majority of employees belong to R&D, yet attrition is still present, suggesting that even technical roles face retention challenges.

A graph of different colored bars

AI-generated content may be incorrect.

* **Life Sciences and Medical fields have the highest attrition** – These fields also have the largest number of employees, making them the biggest contributors to overall attrition.
* **Marketing and Technical Degree fields show moderate attrition** – Although the total employee count is lower, attrition is still noticeable in these areas.
* **Human Resources and "Other" fields have the least attrition** – These fields have the smallest number of employees, and fewer resignations.

💡 **Implication:** The high attrition in Life Sciences and Medical fields could indicate **job dissatisfaction, better opportunities elsewhere, or demanding work conditions**. Employers may need to focus on **career growth, compensation, and work-life balance** to retain talent in these fields.

A graph showing gender and attrition

AI-generated content may be incorrect.

* **Attrition is present in both genders, with a slightly higher count for males** – This could be due to a higher overall male workforce in the dataset.
* **The proportion of voluntary resignations appears similar across genders** – There is no significant indication that one gender is resigning at a much higher rate than the other.

💡 **Implication:** Gender does not seem to be a strong predictor of attrition. Retention strategies should focus on **workplace culture, career growth, and job satisfaction** rather than gender-specific interventions.

A graph of a job role

AI-generated content may be incorrect.

* **Sales Executives and Laboratory Technicians have the highest attrition** – These roles show a noticeably higher number of voluntary resignations compared to others.
* **Managers and Directors have lower attrition rates** – Leadership roles (Manager, Research Director, Manufacturing Director) see fewer resignations, indicating **better job stability** or **higher job satisfaction** at senior levels.
* **Job roles like Research Scientist and Healthcare Representative show moderate attrition** – Suggesting some variability in job satisfaction or external opportunities in these fields.

💡 **Implication:** **Retention efforts should focus on Sales Executives and Laboratory Technicians, potentially improving work conditions, compensation, or growth opportunities in these roles.**

A graph of a person with a divorce

AI-generated content may be incorrect.

* **Single employees have the highest attrition rate** – A larger proportion of single employees are voluntarily resigning compared to married or divorced employees.
* **Married employees have the lowest attrition rate** – This suggests that marital stability may be linked to **higher job retention**, possibly due to financial responsibilities or long-term career planning.
* **Divorced employees fall in between** – Their attrition rate is **lower than single employees but higher than married employees**.

💡 **Implication:** **Retention strategies could focus on single employees, offering better work-life balance, career growth opportunities, or benefits that improve job satisfaction and reduce turnover.**

A graph of an employee

AI-generated content may be incorrect.

* **Employees working overtime show higher attrition rates** compared to those who do not work overtime.
* **A significant portion of employees who do not work overtime remain with the company**, indicating that excessive work hours may contribute to voluntary resignations.
* **Although more employees do not work overtime**, those who do seem to be at a greater risk of leaving, likely due to **burnout or work-life imbalance**.

💡 **Implication:** Organizations could explore **workload management strategies**, such as fair overtime compensation, flexible work schedules, or enforcing work-hour limits, to **reduce burnout and improve retention**.

A graph of different colored bars

AI-generated content may be incorrect.

* **Referral hires have the lowest attrition rate** compared to other sources, suggesting that employees brought in through referrals tend to stay longer.
* **Company website, Seek, and LinkedIn are major hiring sources**, but they also see notable attrition.
* **Job boards (e.g., Indeed, Glassdoor, Adzuna, and Jora) show moderate attrition**, indicating that employees recruited from these sources may not have as strong a long-term commitment.

💡 **Implication:** Investing in **employee referral programs** could improve **retention rates**. Additionally, companies might consider **adjusting their recruitment strategies** to attract candidates who align better with long-term career growth.

## **Key Findings & Observations**

**1.** **General Observations**

* The dataset contains **23,436 records** with **37 columns** (19 numeric, 18 categorical).
* After data cleaning, **23,188 records** remained with **33 columns**.
* Major issues like **missing values, duplicate records, and extreme outliers** were handled before analysis.

**2. Attrition Insights**

* The dataset exhibits a **class imbalance**, with significantly **more current employees than those who resigned**.
* Employees who resigned were generally **younger, lower in job level, and had lower income**.
* Compensation-related variables, such as **MonthlyIncome, JobLevel, and TotalWorkingYears**, showed a weak but noticeable correlation with attrition.
* **Attrition was higher among employees who frequently changed jobs in the past**.

**3. Salary & Job Level Impact**

* **Higher-income employees were more likely to stay**.
* **Job Level and Monthly Income were highly correlated (~0.94)**, meaning that **higher job levels translate into higher income** and possibly **better job stability**.
* Employees at **lower job levels left more frequently**, likely for better opportunities.

**4. Work-Life Balance & Job Satisfaction**

* **Work-life balance had little impact on attrition**, suggesting **employees left for better compensation or career growth** rather than dissatisfaction with work-life balance.
* **Job satisfaction was not a strong predictor** of attrition, indicating **other factors like salary, promotions, and career growth were more influential**.

**5. Impact of Promotions & Career Growth**

* Employees **stuck in the same role for too long** were more likely to leave.
* **Years Since Last Promotion had no strong correlation with attrition**, meaning **employees left for better opportunities rather than waiting for a promotion**.
* Employees with **more experience (Total Working Years)** were less likely to leave.

**6. Managerial Influence**

* **Short tenure under a manager was correlated with higher attrition**, suggesting that **poor managerial relationships or lack of mentorship could drive employees away**.
* Employees who **stayed with the same manager for a long time** were more likely to remain with the company.

**7. Other Notable Findings**

* **Longer commutes (DistanceFromHome) showed weak correlation (-0.07) with attrition**, indicating it might not be a major factor.
* **Overtime work was linked to higher attrition**, possibly due to burnout.
* Employees **hired via referrals had the lowest attrition rates**, meaning **referral programs could be a strong retention strategy**.
* **Frequent travelers had higher attrition**, possibly due to work-related stress and dissatisfaction.
* **Sales Executives and Laboratory Technicians had the highest attrition**, indicating a need for targeted retention efforts in these roles.

**Recommendations**

1. **Compensation Adjustments**
   * Improve salaries for lower-level employees to **reduce voluntary resignations**.
   * Consider **performance-based raises or bonuses** to retain high performers.
2. **Career Growth & Development**
   * Provide **clear career progression paths** to encourage long-term employee retention.
   * Invest in **training programs that align with career growth**.
3. **Retention-Focused Hiring**
   * Strengthen **employee referral programs**, as referrals have lower attrition rates.
   * Evaluate **hiring sources** and prioritize **platforms that bring long-term employees**.
4. **Workload & Managerial Improvements**
   * **Optimize workload and reduce excessive overtime** to minimize burnout.
   * Provide **leadership training for managers** to improve employee-manager relationships.
5. **Flexible Work Arrangements**
   * Explore **hybrid or remote work options** for employees with **long commutes**.
   * **Review travel policies** for frequent travelers to reduce stress-related turnover.