

Reasons for Employee Attrition and Strategies to Reduce It

By Luis Gomez

INTRODUCTION

According to [Indeed](#), “employee attrition refers to the loss of employees due to life events such as retirement, resignation initiated by the employee, elimination of a position, or other similar event. As opposed to turnover, when a job is vacated and refilled, with attrition, the employer will not refill the position instantly”. The employee attrition rate measures the number of people who move out of a company and are not replaced immediately. On average, a company will lose 18% of its workforce annually. Some common reasons for high attrition rates include compensation, recognition, career progression, company culture, poor training, and stress. According to [go2HR](#), “40% of employees who receive poor job training leave their positions within the first year”.

PROBLEM STATEMENT

The problem: why are employees leaving the company? I will be looking at the biggest factors causing employees to leave. A high attrition rate suggests that a company is not retaining high-quality, talented employees. Managing retention and keeping the employee attrition rate below target is one of the most challenging issues facing business. Although employee attrition is a natural process in any business, it is costly and detrimental. Not only is it expensive, but it leads to less productivity and efficiency. Therefore, I want to help businesses minimize their attrition rate.

OBJECTIVES

In this project, the objectives are 1) to identify the various reasons for employee attrition and 2) to build a model to identify which employees are most likely to leave. At the end, I will offer strategies to reduce employee attrition for the company.

METHOD

Data Collection

To collect the data, I found a dataset on Kaggle about employee attrition for a company in the healthcare industry. The dataset contained 35 variables, such as attrition, monthly income, and job satisfaction. There were 1470 observations and no missing rows.

Descriptive Statistics

```
> # Preview summary
> summary(df)
```

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
Min. :18.00	Min. :0.0000	Non-Travel : 150	Min. : 102.0	Human Resources : 63	Min. : 1.000	1:170	Human Resources : 27
1st Qu.:30.00	1st Qu.:0.0000	Travel_Frequently: 277	1st Qu.: 465.0	Research & Development:961	1st Qu.: 2.000	2:282	Life Sciences :606
Median :36.00	Median :0.0000	Travel_Rarely :1043	Median : 802.0	Sales :446	Median : 7.000	3:572	Marketing :159
Mean :36.92	Mean :0.1612		Mean : 802.5		Mean : 9.193	4:398	Medical :464
3rd Qu.:43.00	3rd Qu.:0.0000		3rd Qu.:1157.0		3rd Qu.:14.000	5: 48	Other : 82
Max. :60.00	Max. :1.0000		Max. :1499.0		Max. :29.000		Technical Degree:132

EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome
1:284	Female:588	Min. : 30.00	1: 83	1:543	Sales Executive :326	1:289	Divorced:327	Min. : 1009
2:287	Male :882	1st Qu.: 48.00	2:375	2:534	Research Scientist :292	2:280	Married :673	1st Qu.: 2911
3:453		Median : 66.00	3:868	3:218	Laboratory Technician :259	3:442	Single :470	Median : 4919
4:446		Mean : 65.89	4:144	4:106	Manufacturing Director :145	4:459		Mean : 6503
		3rd Qu.: 83.75		5: 69	Healthcare Representative:131			3rd Qu.: 8379
		Max. :100.00			Manager (Other) :215			Max. :19999

MonthlyRate	NumCompaniesWorked	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StockOptionLevel	TotalWorkingYears
Min. : 2094	Min. :0.000	Min. :0.000	Min. :11.00	1: 0	1:276	0:631	Min. : 0.00
1st Qu.: 8047	1st Qu.:1.000	1st Qu.:0.000	1st Qu.:12.00	2: 0	2:303	1:596	1st Qu.: 6.00
Median :14236	Median :2.000	Median :0.000	Median :14.00	3:1244	3:459	2:158	Median :10.00
Mean :14313	Mean :2.693	Mean :0.283	Mean :15.21	4: 226	4:432	3: 85	Mean :11.28
3rd Qu.:20462	3rd Qu.:4.000	3rd Qu.:1.000	3rd Qu.:18.00				3rd Qu.:15.00
Max. :26999	Max. :9.000	Max. :1.000	Max. :25.00				Max. :40.00

TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
Min. :0.000	1: 80	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.:2.000	2:344	1st Qu.:3.000	1st Qu.:2.000	1st Qu.:0.000	1st Qu.:2.000
Median :3.000	3:893	Median :5.000	Median :3.000	Median :1.000	Median :3.000
Mean :2.739	4:153	Mean :7.008	Mean :4.229	Mean :2.188	Mean :4.123
3rd Qu.:3.000		3rd Qu.:9.000	3rd Qu.:7.000	3rd Qu.:3.000	3rd Qu.:7.000
Max. :6.000		Max. :40.000	Max. :18.000	Max. :15.000	Max. :17.000

Figure 1: Descriptive Statistics of Employee Attrition Data

Based on Figure 1, we can infer many things about the employees at the company. For example, the median age of the employees is 37 years old and 673 employees are married. At the company, 60% of the employees are men. The biggest department is research and development. The company is mainly composed of sales executives, research scientists, and laboratory technicians. Additionally, based on job satisfaction, employees seem satisfied with the majority of the responses being high (3) or very high (4). Employees here are generally educated as the majority of the employees have Bachelor's (3) or Master's degrees (4). The company's work life balance is good with the majority of employees answering better (3) or best (4). However, 28.3% of the employees work overtime, and the company had an attrition rate of 16.12%. Since there are 1470 observations, this means that 237 employees left the company ($1470 * 0.1612$).

Data Visualization

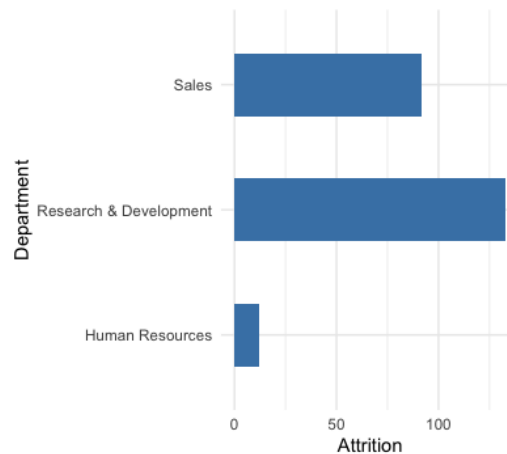


Figure 2: Attrition by Department

Based on Figure 2, the research and development department had the highest attrition at 133 employees. The sales department had the second highest attrition at 92 employees, while the human resources department had the lowest attrition at 12 employees.



Figure 3: Attrition by Job Role

From Figure 3, the following job roles had the highest number of attrition: laboratory technician, sales executive, research scientist, and sales representatives. These four job roles made up the majority (~84%) of the attrition at the company. On the other hand, research directors and managers had the lowest attrition.

Data Analysis

To conduct the analysis, I have chosen to do binomial logistic regression because we are trying to predict a binary outcome: will the employee leave (Yes = 1) or not leave (No = 0). Additionally, I used the odds ratio to determine how much each variable increases or decreases the odds of attrition. I checked for multicollinearity, used only significant variables, and kept the variables with the highest correlation with attrition. Also, I removed five outliers in the dataset and assumed a significance level of 0.05.

RESULTS

Final Model

$$\hat{y} = \frac{e^{1.22 - 0.05x_1 + 0.03x_2 + \dots + 0.18x_8 + 1.65x_9}}{1 + e^{1.22 - 0.05x_1 + 0.03x_2 + \dots + 0.18x_8 + 1.65x_9}}$$

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.220e+00	6.962e-01	1.753	0.079583 .
Age	-4.393e-02	1.264e-02	-3.476	0.000508 ***
DistanceFromHome	3.241e-02	1.090e-02	2.974	0.002944 **
EnvironmentSatisfaction	-3.971e-01	8.440e-02	-4.706	2.53e-06 ***
JobInvolvement	-5.334e-01	1.259e-01	-4.238	2.25e-05 ***
JobSatisfaction	-3.607e-01	8.299e-02	-4.347	1.38e-05 ***
MaritalStatus	6.089e-01	1.325e-01	4.596	4.31e-06 ***
MonthlyIncome	-1.324e-04	3.235e-05	-4.092	4.27e-05 ***
NumCompaniesWorked	1.754e-01	3.690e-02	4.754	1.99e-06 ***
OverTime	1.650e+00	1.920e-01	8.592	< 2e-16 ***

Figure 4: Summary of Final Model

Based on Figure 4, all of the variables in the final model are significant because all the p-values are less than the significance level, 0.05, and the slope intercept is 1.22.

Model Performance

fitting null model for pseudo-r2

llh	llhNull	G2	McFadden	r2ML	r2CU
-380.8560928	-491.8148580	221.9175305	0.2256108	0.1785921	0.3069169

Figure 5: Final Model's McFadden's Pseudo R²

Based on Figure 5, McFadden's Pseudo R² is 0.23. Since the value ranges between 0.2 to 0.4, it indicates that the final model has excellent fit.

	Reference	
Prediction	0	1
0	280	3
1	44	10

Accuracy : 0.8605
 95% CI : (0.8189, 0.8957)
 No Information Rate : 0.9614
 P-Value [Acc > NIR] : 1

Figure 6: Confusion Matrix and Accuracy of Final Model's Predictions

Based on Figure 6, the final model incorrectly predicted 47 rows and correctly predicted 290 rows. From this, the accuracy is 86.05%. Generally, values between 70% to 90% are considered good, so the final model performs well.

	predictor	oddsratio	ci_low (2.5)	ci_high (97.5)	increment
1	Age	0.957	0.933	0.981	1
2	DistanceFromHome	1.033	1.011	1.055	1
3	EnvironmentSatisfaction	0.672	0.569	0.792	1
4	JobInvolvement	0.587	0.458	0.750	1
5	JobSatisfaction	0.697	0.592	0.819	1
6	MaritalStatus	1.838	1.423	2.393	1
7	MonthlyIncome	0.718	0.609	0.836	2500
8	NumCompaniesWorked	1.192	1.109	1.281	1
9	OverTime	5.206	3.585	7.618	1

Figure 7: Odds Ratio with 95% Confidence Interval

Based on Figure 7, all of the odds ratios are significant because they do not include 1 in the 95% confidence interval.

- If an employee works overtime, the odds of attrition is 5.21 times more likely to occur.
- As age increases by one, it is 0.96 times less likely that an employee will leave.
- If the monthly income of an employee increases by \$2500, the odds of attrition is 0.72 times less likely to occur.
- If job satisfaction increases by one, the odds that the employee will leave the company is 0.70 times less likely to happen.

Model Assumption Checks

Assumption #1: The response variable is binary

The response variable (attrition) only had two possible outcomes: yes or no.

Assumption #2: The observations are independent

The observations in the data did not come from repeated measurements of the same individual and were not related to each other in any way – no duplicate values were found.

Assumption #3: There is no multicollinearity among explanatory variables

After I built the final model, I checked for multicollinearity. None of the correlation coefficients between independent variables were greater than 0.7 or less than -0.7 (refer to R file).

Assumption #4: There are no extreme outliers

I calculated Cook's distance for each observation. Five observations were identified as outliers, and I chose to remove them because it made up a small percentage of the dataset.

Assumption #5: There is a linear relationship between the explanatory variables and the logit of the response variable

To determine the relationship, I performed correlation analysis. Each of the nine independent variables had a linear relationship with attrition (refer to R file).

Assumption #6: The sample size is sufficiently large

The sample size needed was 559 ((10 minimum cases * 9 explanatory variables) / 0.1612). After dropping the outliers, there were 1465 observations. Therefore, this requirement is satisfied

CONCLUSION & RECOMMENDATIONS

I found the biggest factors contributing to employee attrition for the company. They are age, distance from home, environment satisfaction, job involvement, job satisfaction, marital status, monthly income, number of companies worked, and overtime. I have built a model that identifies which employees are most likely to leave.

Recommendation #1

Since overtime is one of the biggest factors contributing to employee attrition, the company should reduce the amount of overtime employees work. One way to do this is cross-training employees. Spreading out responsibilities and specialties among the whole team will reduce overtime hours because instead of relying on a single skilled employee, there will be more team members to step in and pick up the load. The company can also match staffing to meet demand by scheduling smarter or hiring more.

Recommendation #2

The company should focus on employee well-being to increase job satisfaction. For example, the company can encourage more work life balance, such as having more days off and maintaining a regular schedule for working hours. With better work life balance, employees should experience less stress and burnout. This would make them more productive, committed, and motivated. In addition, they can emphasize making the work environment more collaborative. It will give employees more sense of belonging, which can lead to more happiness at work. These factors should make employees not want to leave the company.

Recommendation #3

Every six months, the company should collect environmental satisfaction, job involvement, and job satisfaction data about their employees because their feelings will inevitably change over time. Then, they can input this data to the final model to monitor which employees are most likely to leave the company. Whichever employees the model predicts will leave, the company can spend resources to prevent those employees from leaving by catering towards their needs. For instance, sitting down with the employees and discussing what they are lacking or need currently. This gives the company a better chance to prevent employees from leaving.