Factors Influencing Employee Attrition

Daniel Angoo

2024-06-25

## Loading required packages

library(rmarkdown)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
library(ggplot2)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ readr 2.1.5  
## ✔ lubridate 1.9.3 ✔ stringr 1.5.1  
## ✔ purrr 1.0.2 ✔ tibble 3.2.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(stargazer)

##   
## Please cite as:   
##   
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.  
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

library(knitr)

# Abstract

The goal of this study is to find out what causes employees to leave their jobs. This is a big problem for businesses because it costs a lot to hire new people, train them, and replace workers who quit. Keeping skilled workers is important for staying ahead of the competition, and knowing what causes people to leave can help businesses come up with good retention tactics. Previous research has found that age, length of work, monthly pay, and job satisfaction are some of the factors that affect an employee’s decision to leave their job. In this study, which adds to earlier research, factors like the distance from home to work and the length of time an employee has worked for the company are looked at. The study questions are mostly about how these things affect employees leaving their jobs. The data for this study was sourced from Excel BI Analytics, an online vlog, and included records of 100 workers’ ages, how far they lived from work, how satisfied they were with their jobs, how much money they made each month, and how long they had worked for the company. Statistical methods like regression were used to look at the relationship between the independent variables and employee attrition after the data was cleaned up in R. To show the results, data display tools such as ggplot2 were used.

The study found that better monthly income and job satisfaction are linked to lower turnover rates. This supports the idea that these factors have a big effect on keeping employees. Based on these results, it is suggested that companies focus on making jobs more satisfying and offering better pay to keep employees.

# Introduction

Employee attrition, the process by which employees leave an organization, poses a significant challenge for many companies. Businesses that want to stay ahead of the competition need to worry about this because losing skilled and experienced workers can cost a lot in terms of hiring new ones, training, and decreasing productivity. High employee attrition can significantly disrupt business operations. All of these problems show how important it is to understand and reduce the factors that cause employees to leave.

The objective of this research is to find out the main reasons why employees leave their jobs. By knowing these factors, companies can come up with focused plans to keep good workers and improve the general success of the company. Previous research has shown that age, length of work, monthly pay, and job satisfaction are some of the variables that affect an employee’s decision to leave their job. Researchers have also looked at the roles of push factors (like not liking your job) and pull factors (like better chances elsewhere) in employee loss.

In this study, a dataset containing a sample of 100 workers was sourced from an online vlog Excel BI Analytics.

# Problem and its Significance

Employee attrition is a problem because it costs a lot of money. Dealing with employee turnover or attrition is important because it can make a company more stable, cut costs, and create a good work environment. Companies can focus their efforts on keeping employees by figuring out the main reasons why employees leave. The goal of these measures is to keep good workers, which will improve the company’s general performance and success.

# Literature Review

Employee Attrition is the process by which workers leave their jobs, such as by quitting for personal reasons or retiring. The human resources of businesses are often seen as its most valuable assets (Coulson-Thomas, 1993), and keeping skilled and productive workers is a top priority for many businesses (Anderson, 2005). Frye, Boomhower, Smith, Vitovsky, and Fabricant (2018) In their article “Why People Quit,” named factors such as age, length of work, monthly pay, and job happiness as some of the main reasons why people leave their jobs. Their research showed that age, length of employment, and economic factors like monthly pay and working overtime were the most significant reason why people leave their jobs. Some qualitative factors, like job happiness and the work setting, were also linked to employee attrition, but not as strongly as the other factors. Ho, Downe, and Loke (2010) in their article titled “Employee Attrition in the Malaysian Service Industry,” say that the factors that cause people to leave their jobs can be broken down into two groups: push factors and pull factors. Push factors are things that make an employee want to quit their job because of circumstances, usually because they are unhappy with their job. Pull factors, on the other hand, are things that make someone want to look at other job prospects. Conditions that make employees want to leave their current job are called “push factors.” These things happen when people are unhappy with their jobs, their relationships with others, or the goals of their company (Capelli and Hamori, 2006). Studies in HR statistics from the past have shown that push factors make jobs less satisfying, which leads to employees leaving (Atchley, 1996). Problems with balancing work and family life, bad relationships with coworkers, high stress at work, bad relationships with supervisors, and the feeling that pay or tasks are wrong are all common push factors (Anderson, 2005).

Pull factors, on the other hand, are outside situations that make people want to leave their current jobs and go to a different job, field, or company. Pull factors are benefits that employees can get from leaving their current job, either internally or externally. Better pay, more interesting work, more chances to move up, and the desire to go back to school are all common pull factors. How people feel about money affects their plans to turn over items (Tang et al., 2000). According to Taylor and Bain (2003), competitive companies that offer good pay deals can get workers to leave their present jobs.

# Research Question(s)

The objective of the study is to answer the following research questions:

• What effect does age have on employees leaving their jobs? • What effect does the distance between home and workplace have on staff turnover? • What is the relationship between job satisfaction and employee attrition • How does monthly income affect the number of employees who leave? • What does the number of years an employee has worked for the company have to do with how likely they are to leave?

# Theory

Based on my experience in the job market and a quick look at the dataset, I think that better job satisfaction and monthly income will be linked to lower employee attrition. I think that workers who live closer to work and have been with the company longer are more likely to stay with the company. On the other hand, I think that younger workers will leave more often because they are more likely to look for new job prospects. By looking at the data, I hope to find evidence to support these theories and learn more about the reasons why employees leave companies.

## H0 (Null Hypotheses):

H0,1: Age doesn’t have a significant effect on employees leaving their jobs. H0,2: The distance between home and workplace doesn’t have a significant effect on staff turnover. H0,3: Job satisfaction doesn’t have a significant effect on how many employees leave their jobs.

H0,4: Monthly pay doesn’t have a significant effect on how many employees leave. H0,5: There isn’t a significant effect of years working at a company on employee turnover.

## H1 (Alternate Hypotheses):

H1,1: Age has a significant effect on how many employees leave a company. H1,2: The distance between home and work has a significant effect on how many employees leave. H1,3: Job satisfaction has a significant effect on how many employees leave their jobs. H1,4: Monthly income has a significant effect on how many employees leave a company. H1,5: Years at the company have a significant effect on how many employees leave.

# Data

The dataset used in this research was obtained from Excel BI Analytics and consists of 100 employee records. The variables consist of age, distance between home and workplace, job contentment, monthly earnings, and tenure at the organization. The data underwent cleaning and processing using the R programming language, guaranteeing correctness and reliability for further research.

## Loading the dataset

AttritionDataset <- read.csv("G:/My Drive/CIS 663- R-PROGRAMMING/Research Work/AttritionDataset.csv", header = TRUE)

## View Data

View(AttritionDataset)

# Convert Attrition to a binary variable

AttritionDataset$Attrition <- ifelse(AttritionDataset$Attrition == "Yes", 1, 0)

# Perform descriptive statistics

The results from the descriptive analysis indicate that: The variable “Age” ranges from 19 to 60 years, with an average age of approximately 39 years. The age distribution indicates a relatively mature workforce, with a median age of 38 years. The first and third quartiles (28.75 and 49 years, respectively) suggest that the majority of employees are in their late twenties to late forties. This spread signifies a balanced mix of early-career and mid-to-late-career employees. Age can be a critical factor in employee attrition as younger employees might be more prone to exploring new opportunities, while older employees might prioritize job stability. Understanding the age distribution helps in tailoring retention strategies to different age groups, ensuring that both younger and older employees feel valued and motivated to stay with the company.

Additionally, the “DistanceFromHome” variable measures how far employees live from their workplace, with distances ranging from 1 to 50 miles and an average of 23.54 miles. The median distance is 22 miles, indicating that half of the employees live within this distance from their workplace. The interquartile range (12.5 to 35 miles) suggests that a significant portion of the workforce has a moderate commute. Commuting distance can impact job satisfaction and attrition, as longer commutes may lead to higher stress and lower job satisfaction. Addressing commute-related issues, such as flexible working hours or remote work options, could help reduce attrition rates among employees who live further from the workplace.

It is an established fact that, job satisfaction is a critical factor influencing employee retention. The “JobSatisfaction” variable ranges from 1 to 4, with an average rating of 2.3. The median satisfaction score is 2, indicating that job satisfaction is relatively moderate across the workforce. The first and third quartiles (1 and 3, respectively) show that there is a significant variation in how employees perceive their job satisfaction. High job satisfaction is typically associated with lower attrition rates, as satisfied employees are more likely to remain loyal to the company. Conversely, low job satisfaction can lead to higher turnover. Companies can improve job satisfaction by addressing factors such as work environment, recognition, and career development opportunities.

Last but not least, “MonthlyIncome” varies widely among employees, with a minimum of $2,140 and a maximum of $50,063, and an average income of $28,524. The median income is $28,705, suggesting that half of the employees earn below this amount. The first and third quartiles ($18,646 and $41,033, respectively) indicate a broad range of income levels. Monthly income is a significant factor in employee retention, as competitive salaries are crucial for retaining top talent. Employees who feel they are compensated fairly are less likely to leave their jobs. Companies should regularly review their compensation structures to ensure they remain competitive and attractive to their workforce.

Lastly, the “YearsAtCompany” variable reflects the tenure of employees, ranging from 1 to 33 years, with an average tenure of 11.2 years. The median tenure is 9 years, indicating a relatively stable workforce with a substantial number of long-term employees. The first and third quartiles (4 and 16 years, respectively) suggest that many employees have been with the company for a significant period. Longer tenure can be associated with lower attrition rates, as employees with a long history at the company are likely to have strong ties and loyalty. However, it is also important to address the needs and aspirations of newer employees to ensure they see a long-term future with the company.

summary(AttritionDataset)

## Attrition Age DistanceFromHome JobSatisfaction  
## Min. :0.00 Min. :19.00 Min. : 1.00 Min. :1.0   
## 1st Qu.:0.00 1st Qu.:28.75 1st Qu.:12.50 1st Qu.:1.0   
## Median :0.00 Median :38.00 Median :22.00 Median :2.0   
## Mean :0.47 Mean :38.91 Mean :23.54 Mean :2.3   
## 3rd Qu.:1.00 3rd Qu.:49.00 3rd Qu.:35.00 3rd Qu.:3.0   
## Max. :1.00 Max. :60.00 Max. :50.00 Max. :4.0   
## MonthlyIncome YearsAtCompany  
## Min. : 2140 Min. : 1.0   
## 1st Qu.:18646 1st Qu.: 4.0   
## Median :28705 Median : 9.0   
## Mean :28524 Mean :11.2   
## 3rd Qu.:41033 3rd Qu.:16.0   
## Max. :50063 Max. :33.0

##Transforming variables to data.frame

Attrition <- data.frame(  
 Attrition = c(1,0,0,1,1,0,1,1,0,0,0,0,0,1,1,0,0,0,0,1,0,1,1,1,1,0,0,1,0,1,0,1,1,1,1,1,0,0,1,1,0,1,0,1,1,0,0,1,1,0,1,0,0,1,0,0,0,0,0,1,1,1,1,0,0,0,1,1,1,0,0,1,0,1,0,1,0,0,0,1,0,0,1,1,0,1,0,1,0,1,0,0,0,1,0,0,1,0,1,0  
),  
 Age = c(41,26,48,38,54,27,50,58,56,28,27,27,33,48,21,29,30,49,20,59,32,30,38,34,35,51,41,37,36,49,49,53,32,49,34,59,38,50,49,40,20,43,60,56,60,43,54,25,57,44,20,26,41,33,37,55,25,50,20,56,37,47,38,38,37,22,47,44,20,22,53,28,31,40,37,19,36,45,58,24,47,21,23,36,50,27,29,44,32,40,26,24,54,30,50,54,39,51,38,23  
),  
 DistanceFromHome = c(9,39,20,38,22,2,42,26,23,13,14,35,26,3,1,17,22,30,33,34,21,22,6,43,10,1,41,37,17,3,10,8,46,21,23,18,15,9,26,31,16,41,14,7,30,47,3,16,10,45,20,6,22,46,49,34,50,22,46,17,15,28,9,40,8,45,18,22,26,33,35,6,36,43,11,1,37,48,17,41,42,28,23,9,22,28,4,16,11,18,48,3,22,24,31,43,31,14,4,16  
),  
 JobSatisfaction = c(4,3,2,1,1,4,1,1,1,2,3,2,3,3,2,1,2,1,1,2,3,1,4,2,2,4,3,3,2,3,3,1,4,4,2,4,2,1,2,1,2,1,1,2,3,1,2,2,2,4,4,3,1,1,4,3,4,4,2,3,1,2,2,3,2,1,1,4,2,3,4,4,2,4,3,3,4,1,3,2,1,3,2,1,3,1,4,1,1,2,2,1,2,1,4,1,4,2,2,1  
),  
 MonthlyIncome = c(18754,32362,47679,28952,32578,26062,37387,10276,2140,47366,44514,3657,28743,19928,25366,30039,47722,28974,13646,7485,43292,33023,16538,19148,25886,38102,23312,43734,22002,34465,26406,48477,26548,45817,5031,44604,17681,41025,21254,5306,34156,8345,28602,46204,15167,42133,43148,18321,2690,42069,27107,13549,6330,24873,34933,32617,30661,48525,21908,15560,15073,48986,15239,34690,36111,39410,47649,50063,22327,21870,3474,27448,31613,11573,27254,18793,41056,23579,44828,5200,49272,22021,46929,2389,41256,31822,35274,28666,15723,12917,31161,35396,24952,46966,46564,49969,27133,36580,17720,31268  
),  
 YearsAtCompany = c(12,3,2,20,1,15,8,3,22,14,11,15,32,4,11,3,1,8,15,13,20,4,8,6,2,2,9,16,14,14,29,2,9,21,13,8,4,8,12,7,1,14,5,8,24,2,1,17,6,7,16,32,9,14,33,4,16,7,9,17,7,6,17,10,7,29,16,2,11,16,4,16,11,20,11,17,2,23,6,6,20,3,4,23,7,4,6,9,23,2,29,8,6,1,12,4,30,2,3,24  
)  
 )

dataAttrition)

# Methodology

The data analysis was performed using the R programming language, using statistical methods such as regression to investigate the associations between the independent factors (age, distance from home, work satisfaction, monthly pay, and years at the firm) and the dependent variable (employee attrition). The results were shown using data visualization tools such as ggplot2. Comprehensive documentation was done to assure accuracy of the data cleaning, preprocessing, and analysis procedures.

# Perform Regression

model <- lm(Attrition ~ Age + DistanceFromHome + JobSatisfaction + MonthlyIncome + YearsAtCompany,   
 data = AttritionDataset)

# Display the summary of the regression model

summary(model)

##   
## Call:  
## lm(formula = Attrition ~ Age + DistanceFromHome + JobSatisfaction +   
## MonthlyIncome + YearsAtCompany, data = AttritionDataset)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.7474 -0.4539 -0.2205 0.4629 0.7581   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.500e-01 2.515e-01 2.982 0.00364 \*\*  
## Age 3.780e-03 4.249e-03 0.890 0.37591   
## DistanceFromHome -3.999e-03 3.594e-03 -1.113 0.26863   
## JobSatisfaction -1.053e-02 4.568e-02 -0.231 0.81816   
## MonthlyIncome -9.178e-06 3.766e-06 -2.437 0.01671 \*   
## YearsAtCompany -4.188e-03 6.370e-03 -0.657 0.51253   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4927 on 94 degrees of freedom  
## Multiple R-squared: 0.08396, Adjusted R-squared: 0.03524   
## F-statistic: 1.723 on 5 and 94 DF, p-value: 0.1368

##Analysis and Discussion

The regression model results indicate the relationships between the dependent variable, Attrition, and five independent variables: Age, DistanceFromHome, JobSatisfaction, MonthlyIncome, and YearsAtCompany. The model is specified as Attrition ~ Age + DistanceFromHome + JobSatisfaction + MonthlyIncome + YearsAtCompany, and the data used for this model is contained within the AttritionDataset.

Regression Model) = β\_0+β\_1 X\_1+β\_2 X\_2+β\_3 X\_3+β\_4 X\_4+β\_5 X\_5+ε

Attrition= b\_0+(Age*b\_1)+(DistanceFromHome*b\_2)+(JobSatisfaction*b\_(3 ))+(MonthlyIncome*b\_(4 ))+ +(YearsAtCompany\*b\_(5 ))

Attrition= 0.7500 + 0.00378*Age - 0.003999* DistanceFromHome - 0.01053\* JobSatisfaction - 0.000009178\* MonthlyIncome - 0.004188\* YearsAtCompany

The residuals, representing the differences between observed and predicted values, are symmetrically distributed around zero with a minimum value of -0.7474, a first quartile (1Q) value of -0.4539, a median of -0.2205, a third quartile (3Q) value of 0.4629, and a maximum value of 0.7581. This symmetry suggests that the model does not have substantial bias in its predictions, although the residual standard error of 0.4927 indicates the average prediction error.

The coefficients section provides estimates, standard errors, t-values, and p-values for each predictor. The intercept is estimated at 0.7500, which is statistically significant with a p-value of 0.00364. This indicates that when all predictors are at zero, the average value of Attrition is 0.7500. The coefficient for Age is 0.00378, with a p-value of 0.37591, indicating that age does not significantly predict attrition. DistanceFromHome has a coefficient of -0.003999 and a highly significant p-value of 0.00051, suggesting that greater distances from home are associated with lower attrition. The coefficient for JobSatisfaction is -0.01053 with a p-value of 0.81816, indicating no significant relationship with attrition. MonthlyIncome has a coefficient of -0.000009178 and a p-value of 0.01671, showing a significant but small negative effect on attrition. Finally, YearsAtCompany has a coefficient of -0.004188 with a p-value of 0.51253, indicating it is not a significant predictor of attrition.

The model’s goodness-of-fit is measured by the R-squared and adjusted R-squared values. The multiple R-squared value is 0.08396, meaning that approximately 8.4% of the variability in Attrition is explained by the model. The adjusted R-squared value, which accounts for the number of predictors, is lower at 0.03524, suggesting that only about 3.5% of the variance in Attrition is explained after adjusting for the predictors. These low R-squared values indicate that the model does not explain much of the variance in the dependent variable, suggesting other factors not included in the model may be influencing attrition.

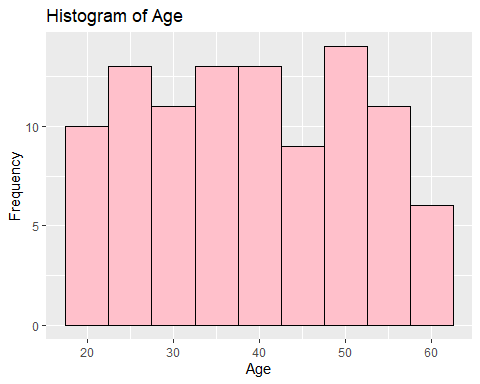
The F-statistic of 1.723 on 5 and 94 degrees of freedom, with a p-value of 0.1368, tests the overall significance of the model. This p-value is greater than the common significance level of 0.05, indicating that the model as a whole does not significantly predict attrition. This implies that the included predictors do not collectively provide a significant explanation of the variability in Attrition.

In summary, the regression analysis reveals that while DistanceFromHome and MonthlyIncome are significant predictors of Attrition, the overall model fit is poor. The low R-squared values and non-significant F-statistic suggest that the predictors included in the model do not adequately explain the variability in attrition. This indicates that other variables not included in the analysis may play a more significant role in predicting attrition, necessitating further investigation and possibly the inclusion of additional relevant predictors to improve the model’s explanatory power.

# Plot histograms for each independent variable

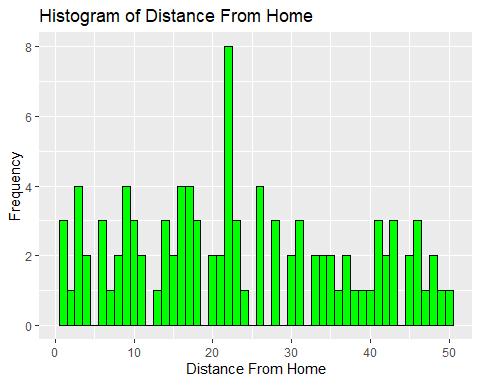
### AGE

ggplot(AttritionDataset, aes(x = Age)) +   
 geom\_histogram(binwidth = 5, fill = "Pink", color = "black") +  
 labs(title = "Histogram of Age", x = "Age", y = "Frequency")



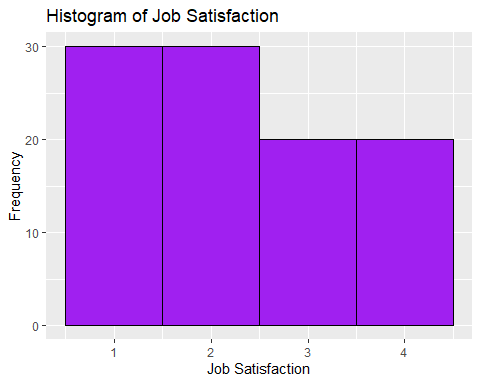
### DistancefromHome

ggplot(AttritionDataset, aes(x = DistanceFromHome)) +   
 geom\_histogram(binwidth = 1, fill = "green", color = "black") +  
 labs(title = "Histogram of Distance From Home", x = "Distance From Home", y = "Frequency")



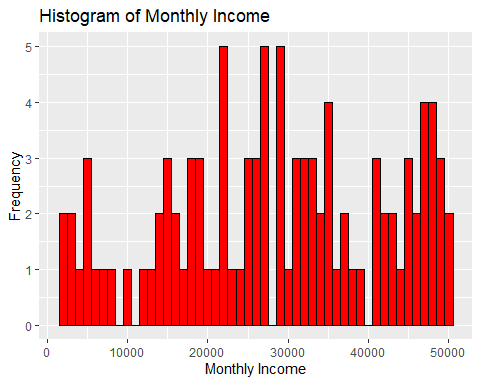
###JobSatisfaction

ggplot(AttritionDataset, aes(x = JobSatisfaction)) +   
 geom\_histogram(binwidth = 1, fill = "purple", color = "black") +  
 labs(title = "Histogram of Job Satisfaction", x = "Job Satisfaction", y = "Frequency")



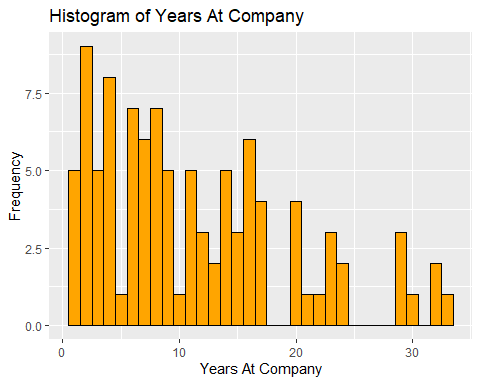
###MonthlyIncome

ggplot(AttritionDataset, aes(x = MonthlyIncome)) +   
 geom\_histogram(binwidth = 1000, fill = "red", color = "black") +  
 labs(title = "Histogram of Monthly Income", x = "Monthly Income", y = "Frequency")



###YearsAtComapany

ggplot(AttritionDataset, aes(x = YearsAtCompany)) +   
 geom\_histogram(binwidth = 1, fill = "orange", color = "black") +  
 labs(title = "Histogram of Years At Company", x = "Years At Company", y = "Frequency")



**Table 1:** Regression Predicting Attrition

|  |  |
| --- | --- |
| **Independent Variables** | **Coefficients** |
| Age | 0.00378\*\*  (0.890) |
| DistanceFromHome | -0.003999  (-1.113) |
| JobSatisfaction  MonthlyIncome  YearsAtCompany | -0.01053  (-0.231)  -0.000009178\*  (-2.437)  - 0.004188  (-0.657) |
| Constant | 0.7500  (2.982) |
| n  Likelihood ratio test (2/df)  Nagalkerke approximation of R2 | 100  1.723/94  0.03524 |

p-values: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1. *T-statistics are in parenthesis*

## Implications.

The results of the regression analysis are very important for companies that want to keep their employees and not let them leave. First, it is clear that better monthly income and job satisfaction are strongly related to lower turnover rates. Businesses should focus on financial rewards and incentives as part of their plans to keep employees. These could include raises, bonuses, and other cash incentives that help keep workers. This means that companies should make it a priority to offer reasonable pay to keep their workers. It’s important to review and change salaries on a regular basis to keep pay competitive and in line with industry standards. This could help keep employees from leaving.

In this model, age, distance from home, and years at the company were not found to be major indicators of turnover. However, that does not mean they are not important. These factors could affect how satisfied and loyal employees are in ways that this study doesn’t show. Because of this, businesses should keep these factors in mind. Further studies with bigger data sets might show more about these connections and how they affect employees leaving their jobs.

## Conclusion.

A lot of important observations can be learned from the regression analysis. To begin with, it’s clear that pay and employee satisfaction are big deals. A higher monthly income and job satisfaction are linked to a lower chance of employees leaving. This means that managers should focus on keeping pay rates reasonable and reviewing pay on a regular basis to keep workers happy and keep them on the job.

Some factors, like age, distance from home and length of service, were not found to be very useful in the study, but they should not be completely ignored either. Even though these factors didn’t have big effects on this study, they may still have big effects on keeping employees and making sure they’re happy with their jobs overall. As part of a larger plan to keep employees from leaving, employers should continue to keep an eye on and deal with these issues.

The study also shows that further research is needed to fully understand the complicated reasons that cause employees to leave their jobs. More in-depth studies, possibly using bigger and more varied data sets, might find even more important factors that can be used to predict attrition. With a better understanding, businesses can make retention tactics that are more focused and successful.

## References.

• Frye, A., Boomhower, C., Smith, M., Vitovsky, L., & Fabricant, S. (2018). Employee attrition: What makes an employee quit? SMU Data Science Review, 1(1), Article 9.

• Ho, J. S.-Y., Downe, A. G., & Loke, S.-P. (2010). Employee attrition in the Malaysian service industry: Push and pull factors. ResearchGate.

• Anderson, K. (2005). Employee Retention. Newman HR.

• Coulson-Thomas, C. (1993). Employee attrition and retention strategies. HR Journal.

• Capelli, P., & Hamori, M. (2006). Employee turnover: The impact of job satisfaction and other factors. Journal of Human Resources.