University of Warsaw

Faculty of Economic Sciences

Jarosław Leski 411174

Robert Kowalczyk 397108

Determinants of job attrition.

Analysis using binary dependent variable models

Research project

Advanced Econometrics

A paper written under the direction of

dr Rafał Woźniak

Department of Statistics and Econometrics

Faculty of Economic Sciences

Warsaw, June 2022

Abstract

The purpose of this study was to examine the determinants of quitting. In particular, attention was focused on non-wage factors. A logit model was used for this purpose, which proved to be better than a probit model in this case. Our results reveal that non-wage factors are as important as wage factors and should be taken into account when describing this phenomenon. It was also found that factors beyond the control of the employer such as matrimonial status, age, education level are important and affect the propensity to change jobs. However, employers can reduce turnover rates by increasing job satisfaction.

Table of contents

1. Introduction

A number of prior studies have shown a relationship between inadequate job satisfaction and increased odds of an employee leaving (Shields and Ward 2001). Most studies focus their attention mainly on pay factors, confirming their impact on job satisfaction. However, employee departure is a more complex topic and may also result from non-wage characteristics. Many studies have analyzed the influence of other factors such as personal characteristics: age, gender, and characteristics of the work environment and employment culture (Lopes, Lagoa, and Calapez, 2014).

Undoubtedly, the phenomenon of employees leaving is extremely costly for companies. The loss of an employee is associated not only with additional costs resulting from hiring and training a new person, but also in the loss of know-how possessed by the departing employee. In extreme cases, this can lead to the disruption of critical processes for the company, such as when a highly skilled employee is lost and difficult to replace. In view of the above, there is no illusion that it pays for companies to keep their employee turnover rate as low as possible.

The research problem of this paper is to analyze the determinants of employee quits. Our research hypothesis is that both wage and non-wage factors must be taken into account when analyzing departures. In particular, we will verify the significance and direction of employment-related variables. We believe that valuing employees by increasing the frequency of promotions has a positive effect on decreasing the probability of quitting. We will also verify whether the level of education has an impact on exits, as Weiss (1984) shows. We will use personal characteristics such as age, matrimonial status as secondary research hypotheses.

This study provides a complete literature review of previous research studies related to the determinants of job satisfaction and quitting. We then formulate once again the research hypotheses. The next section provides an extensive analysis and description of the dataset used in the study. In the next step, we turn to the empirical study conducted with three models: linear probability model (LPM), probit and logit regression. We present the initial forms of the models, carry out the procedure from general-to-specific, and look at the diagnostics of the models. We then verify the research hypotheses posed by referring to previous scientific achievements. The final part of the paper contains the conclusions of the study and possibilities for further work on the issue of job attritions.

2. Related literature

In the literature, a topic on which much research is being generated is the determinants of employee job satisfaction. This issue may be one of the reasons for an employee's decision to quit their job (Kristensen and Westergård-Nielsen, 2004). To begin our discussion, it is worth citing a study by Lopes, Lagoa, and Calapez (2014), which noted over the period 1995-2010 that average job satisfaction decreased while job pressure increased. In the ordered logit regression model conducted, the authors also prove the negative effect of age and age squared variables on job satisfaction, and that women are less satisfied with their jobs than men.

Diaz-Serrano and Cabral Vieira (2005) examined job satisfaction among low- and well-paid workers in European Union countries. The results of the ordered probit model show an unequivocal positive relationship between hourly pay and job satisfaction in all examined EU countries except the UK. In most of the countries studied, men on average rated their job satisfaction lower.

An important issue regarding job satisfaction is the so-called work-life balance (WLB). One study that has emerged in this area is that of Haar, Russo, Sune, and Ollier (2014). Using SEM analysis on a sample of 1416 employees, the authors showed the presence of a positive relationship between WLB and job satisfaction and a negative relationship with anxiety and depression in each of the 7 cultures studied.

The distance from home to the place of employment is also a factor taken into account when considering job satisfaction. Spies (2006) study was conducted on a group of oil industry workers in northwestern Russia and attempted to explain job satisfaction using 26 indicators including distance from home to the place of employment. The research group was divided into three subgroups, of which the one with the longest distance to work rated the best satisfaction. In addition, the researcher was able to prove a statistically significant weak positive Spearman correlation between distance and job satisfaction.

In research on job satisfaction, the issue of employees' education often comes up. Vila and Garcia-Mora (2007) investigated on a group of Spanish employees whether the level of education has an impact on job satisfaction. They confirmed that the impact of education level on job satisfaction varies, both in magnitude and direction, depending on the aspect of work considered.

Using panel data, the determinants of job satisfaction were also explained by D'Addio and Frijters (2011), confirming previous research findings. They proved on Danish employees data that the determinants of job satisfaction differ significantly by gender, especially after accounting for individual fixed effects. Additionally, being married was found to be a statistically significant factor that increases job satisfaction.

Of course, studies on the determinants of work attrition have also emerged. One such study is that of Corneließen (2006), in which the author examined the determinants of quits using data from the German Socio-Economic Panel (GSOEP) in a fixed effects framework. The explanatory variables used included gender, age, number of previous job changes, material status, earnings, job satisfaction, development opportunities, promotion opportunities, working time including overtime. Two models were used in the study: Logit and Linear probability model. For both models, the influence of none of the above mentioned factors on leaving the job was confirmed. The key aspects were found to be job autonomy, interpersonal relationships and job security.

However, using a probit model, some of the relationships mentioned above were demonstrated by researcher Weiss (1984). The conclusions of his work are as follows: employees who resigned from their previous jobs are less likely to leave their jobs than employees who were unemployed when they applied for the studied positions. This is consistent with another finding from the paper which is that more educated people are more likely to quit. What is more, younger workers are more likely to quit than older workers. Moreover, the negative influence of salary on the probability of leaving was proved. However, it should be emphasized that the study was conducted on a group of employees up to 6 months from the moment of employment.

Kristensen and Westergård-Nielsen (2004) also addressed the issue of determinants of leaving a job. The analysis was based on data from the European Community Household Panel (ECHP) on Danish employees. The hypothesis was that low job satisfaction is the main cause of employees' leaving, which has been confirmed. It also indirectly proved that women are more prone to change jobs. Seniority in the company also proved to be an important factor. Employees with more than 10 years of work experience in the company were more prone to change compared to the base level - employees with 4-10 years of experience in the company.

3. Research hypotheses

The following study will attempt to verify the main hypothesis and secondary hypotheses presented below. The list of proposed hypotheses is presented as follows:

Main hypothesis:

1. A worker's willingness to leave a job is not determined solely by the wages factors, but also by non-wages factors.

Second hypotheses:

1. Married people are less likely to change jobs because they care more about the financial stability of their family.
2. People who are more satisfied with their job are less likely to decide to change jobs.
3. The age of the employee has a significant impact on the employee attrition.
4. An employee's level of education has a significant impact on employee attrition.

4. Data analysis

The dataset was obtained through shared data via the Kaggle website[1]. The dataset consists of 1470 observations and contains 34 independent variables describing the characteristics of each observation and one dependent variable indicating whether the individual decided to change jobs.

In the following analysis and econometric modeling, only 14 variables out of 34 independent variables were used. This decision was dictated by the fact that many variables did not have a precise description of what they meant and we used our economic intuition and the literature to eliminate unnecessary variables.

A detailed description of each variable can be found in Table 1.

[1] https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset (access 02.05.2022)

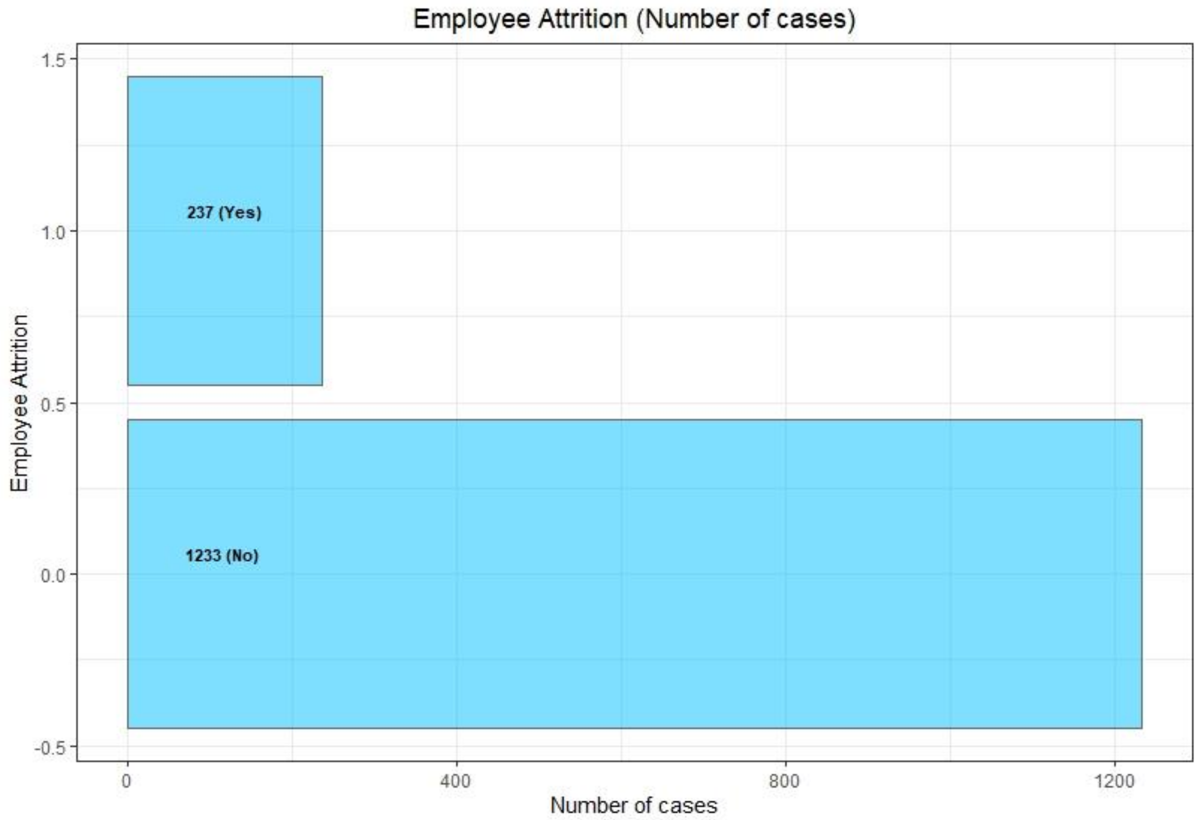
Table 1. Detailed information on the variable under consideration and the characteristics of the employee

Variable name	Dependent or independent	Variable type	Details
Attrition	Dependent	Binary	Information whether the person has decided to change job: 1 = ‘Yes’ 0 = ‘No’
Age	Independent	Continuous/Discrete	Age of person
Age2	Independent	Continuous/Discrete	Age of person squared
AgeOverTime	Independent	Continuous/Discrete	Interaction between age of a person and whether person is taking overtime
AgexMarried	Independent	Continuous/Discrete	Interaction between age of a person and whether person is married
Divorced	Independent	Binary	Information whether the person is divorced: 1 = ‘Yes’ 0 = ‘No’
Education	Independent	Ordinal	Educational level of the person: 1 = 'Below College' 2 = 'College' 3 = 'Bachelor' 4 = 'Master' 5 = 'Doctor'
Female	Independent	Binary	Information whether person is female: 1 = ‘Yes’ 0 = ‘No’
JobInvolvement	Independent	Ordinal	Information on how involved the person is in the work: 1 = 'Low' 2 = 'Medium' 3 = 'High' 4 = 'Very High'
JobSatisfaction	Independent	Ordinal	Job satisfaction level: 1 = 'Low' 2 = 'Medium' 3 = 'High' 4 = 'Very High'
lnMonthlyIncome	Independent	Continuous	Natural logarithm of person monthly income

Married	Independent	Binary	Information whether the person is married: 1 = ‘Yes’ 0 = ‘No’
NumCompaniesWorked	Independent	Discrete	Information on how many companies the person has worked with so far
Over7KM	Independent	Binary	Information whether the person has further to go to work than the population median (7 KM): 1 = ‘Yes’ 0 = ‘No’
OverTime	Independent	Binary	Information on whether someone is taking overtime: 1 = ‘Yes’ 0 = ‘No’
TotalWorkingYears	Independent	Discrete	Information on how much work experience the person has
WorkLifeBalance	Independent	Ordinal	Work Life Balance level: 1 = 'Bad' 2 = 'Good' 3 = 'Better' 4 = 'Best'
YearsAtCompany	Independent	Discrete	Information on the number of years in the current company
YearsSinceLastPromotion	Independent	Discrete	Information on the number of years since last promotion

Source: Own elaboration

Figure 1. Distribution of the attrition level binary variable under study.



Source: Own elaboration using external library ggplot2 in RStudio software

A first look at the dependent variable reveals that our dataset regarding the characteristics of employees and their willingness to change jobs is highly unbalanced. This may be due to the fact that relatively fewer employees choose to change jobs compared to those who want to continue working in their current job. The abundance of these two sets is shown in Figure 1.

4.1. Missing values and outliers detection

Preliminary analysis of the dataset of workers and their characteristics did not reveal that there were any missing observations in the dataset. Additionally, almost all variables showed no outlier observations except for monthly income. Therefore, we decided to transform this variable accordingly.

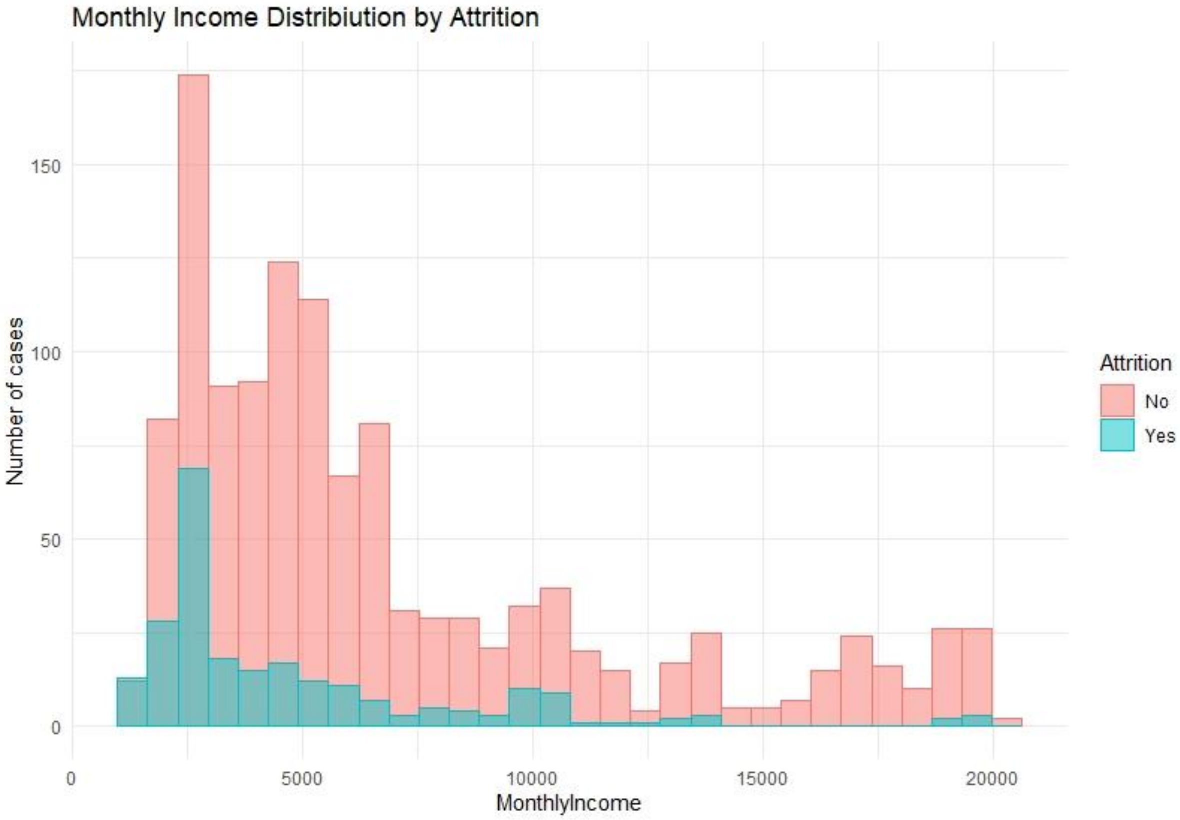
4.2. Data transformation

In this section we would like to reflect on and present appropriate transformations of variables, the addition of new non-linear variables and interactions.

4.2.1. Natural logarithm of Monthly Income

The variable describing the monthly income of a worker is characterized by a right skewness of the distribution (long right tail). In most of the datasets on workers' earnings we are confronted with the above problem. The distribution of continuous variables should be as close as possible to a normal distribution. The distribution of workers' monthly earnings by job attritions is shown in Figure 2.

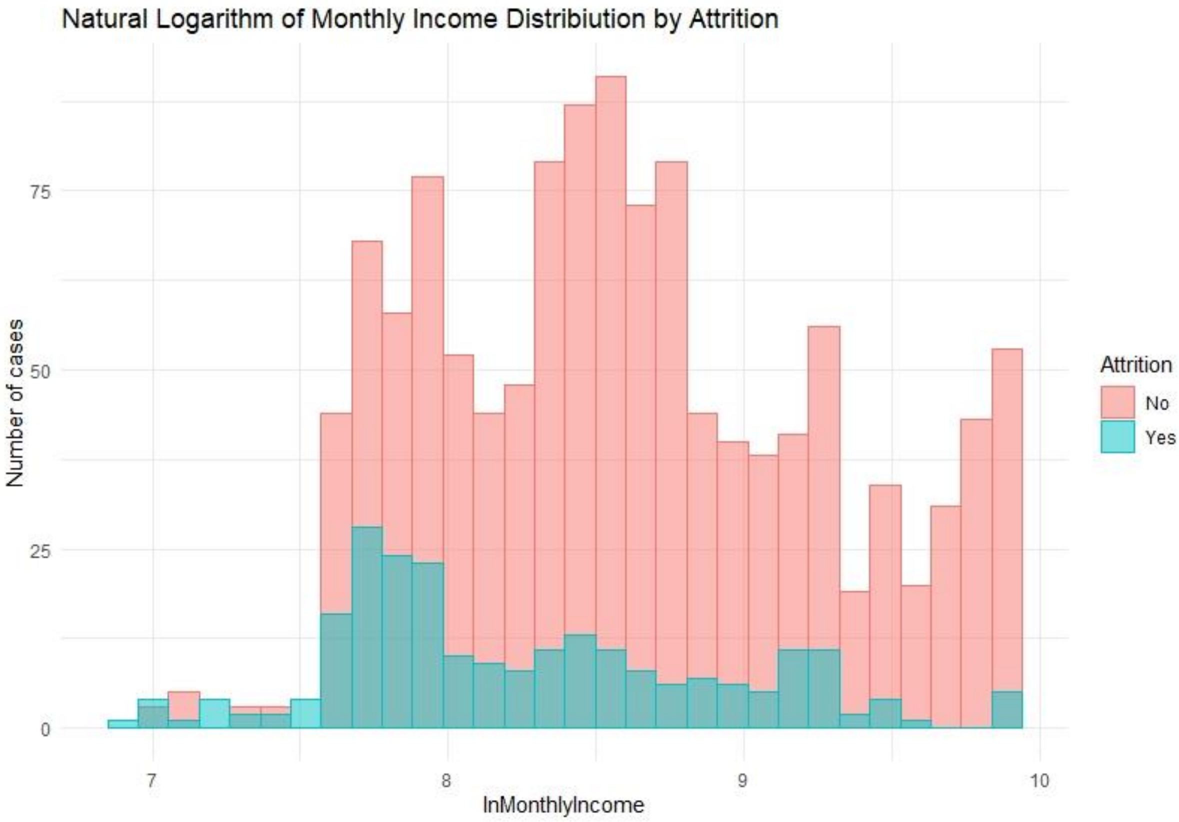
Figure 2. Distribution of workers' monthly income by attrition.



Source: Own elaboration using external library ggplot2 in RStudio software

Using the natural logarithm on the variable of employees' monthly income did not lead to a complete approximation to the normal distribution. On the other hand, we got rid of the right-hand skewness. The transformation using the natural logarithm can be seen in Figure 3.

Figure 3. Distribution of natural logarithm of workers' monthly income by Attrition.



Source: Own elaboration using external library ggplot2 in RStudio software

4.2.2. Age squared

We also used a non-linear variable in our model. Many econometric models use the age of the worker squared. In this case, the age variable has a double effect on the dependent variable and therefore the partial derivative must be calculated to get an accurate effect.

4.2.3. Interaction between variable Age and Married

The first interaction of interest was the interaction of the worker's age with whether they are married. This complementary effect could allow us to verify whether married people differ significantly from unmarried people according to their age.

4.2.4. Interaction between variable Age and OverTime

In addition, we were interested in the interaction effect of employee age on whether they work after hours. This complementary effect could allow us to verify whether there is a significant difference in people working after hours according to their age.

4.3. Analysis of the relationship of the variables included in the research hypotheses

4.3.1. Job satisfaction in comparison to Attrition

Figure 4. Relationship between Job Satisfaction to Attrition.

Source: Own elaboration using external library ggplot2 in RStudio software

Graphs of the relationship of job satisfaction by Attrition variable are shown in Figure 4. The numerical values for the ordinal variable do not clearly indicate a one-sided positive or negative relationship. Much more can be deduced from the graph of the percentage for the category. It can be seen that as an employee's level of job satisfaction increases, the lower the percentage of employees thinking about changing jobs.

4.3.2 Age in comparison to Attrition

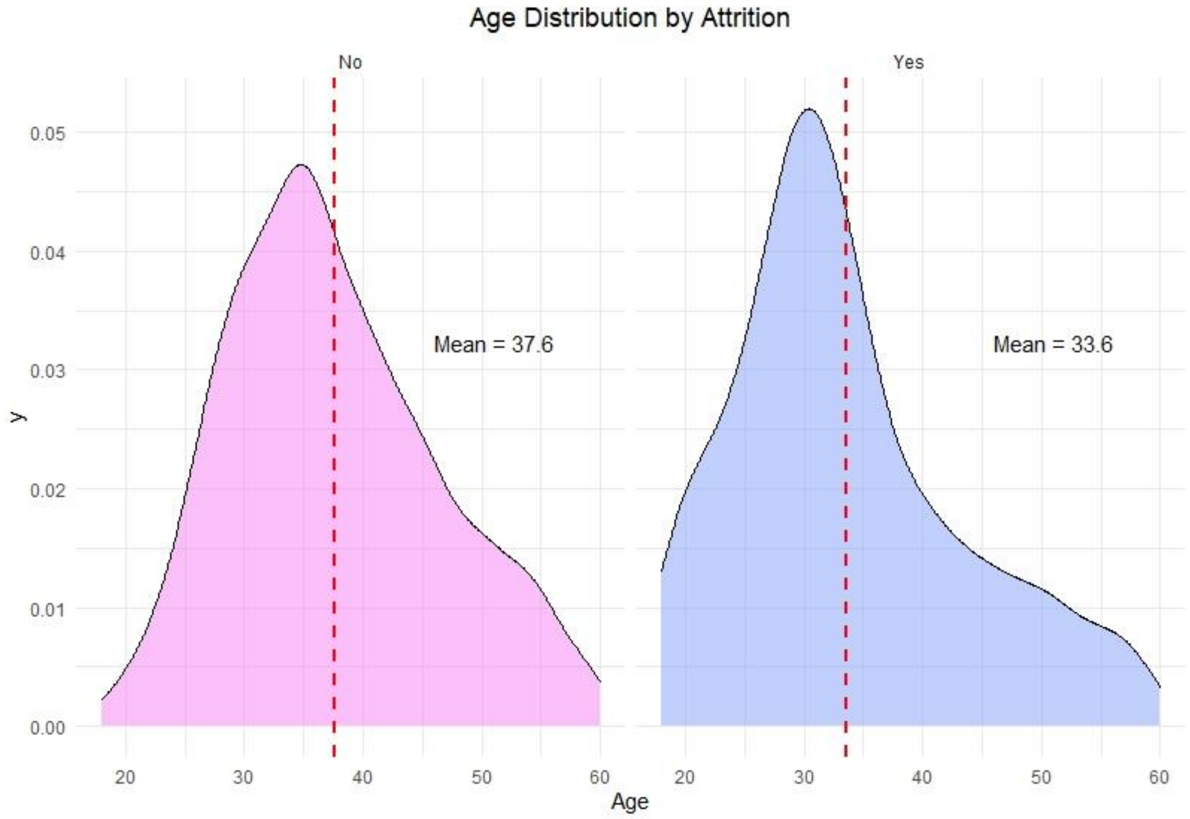


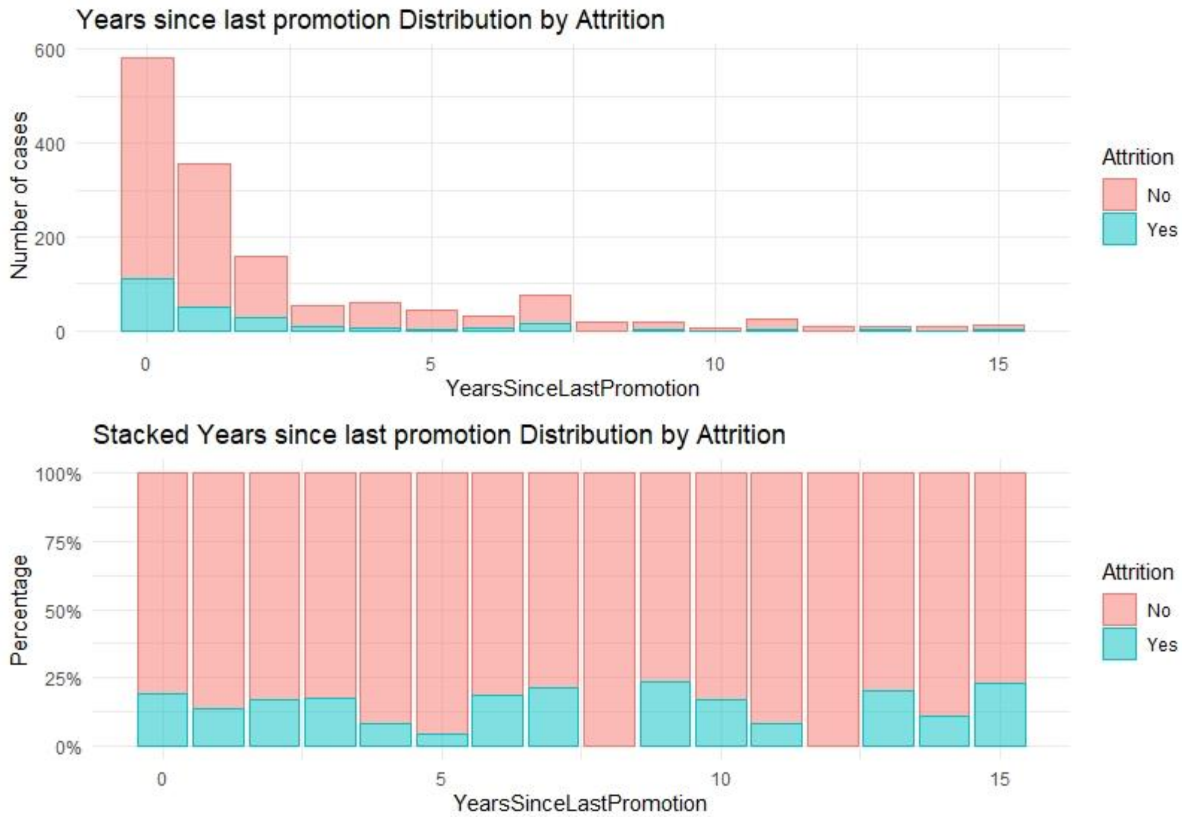
Figure 5. Distribution of the employee age variable by Attrition.

Source: Own elaboration using external library ggplot2 in RStudio software

In Figure 5 showing the distribution of the employee age variable according to the value of the Attrition variable, it can be seen that the average age for the group of people who do not choose to change jobs is higher compared to those who do, however, choose to change. The distribution is more shifted to the right, indicating that older people are less likely to decide to change jobs. The younger people are less afraid of challenges and are more likely to reach for changes in their lives because they have not yet found themselves in their target environment.

4.3.3. Frequency of promotions in comparison to Attrition

Figure 6. The number of years since the last promotion compared to the Attrition variable.

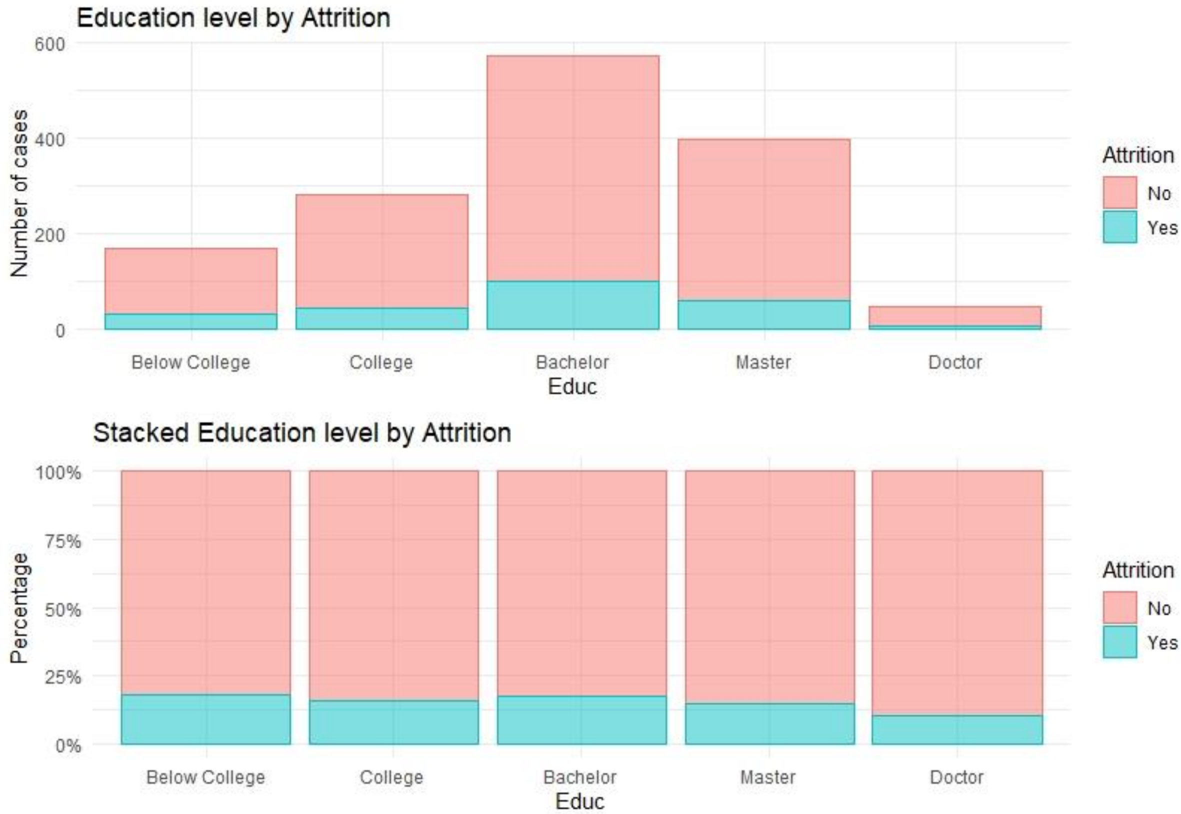


Source: Own elaboration using external library ggplot2 in RStudio software

Figure 6 shows the relationship of the number of years since the last promotion by both job changers and nonchangers. From the graph, it is not clear that the number of years since the last promotion significantly stands out in terms of substantive significance. On the other hand, it can be seen that a very large number of observations in the current or previous year have been promoted and do not want to leave the workplace.

4.3.4. Education level in comparison to Attrition

Figure 7. Employee education level by Attrition variable.



Source: Own elaboration using external library ggplot2 in RStudio software

Figure 7 shows the relationship of education level by jobs attrition. It can be seen that most observations have a bachelor's degree. Additionally, it can be seen that as the educational degree increases, the proportion of those willing to change jobs decreases.

5. Econometric analysis

The Ordinary Least Squares model for binary dependent variable (Linear Probability Model) using White's robust matrix was determined by the following equation:

$$\begin{aligned} Attrition_i = {} & \beta_0 + \beta_1 Age_i + \beta_2 Age2_i + \beta_3 AgeOverTime_i + \beta_4 AgexMarried_i \\ & + \beta_5 Education_i + \beta_6 JobInvolvement_i + \beta_7 JobSatisfaction_i \\ & + \beta_8 lnMonthlyIncome_i + \beta_9 NumCompaniesWorked_i \\ & + \beta_{10} TotalWorkingYears_i + \beta_{11} WorkLifeBalance_i \\ & + \beta_{12} YearsAtCompany_i + \beta_{13} YearsSinceLastPromotion_i \\ & + \beta_{14} Over7KM_i + \beta_{15} Married_i + \beta_{16} Female_i + \varepsilon_i \end{aligned}$$

The logit and probit model, on the other hand, was estimated using the following formula. [2]

$$
\begin{aligned}
ln\left(\frac{P(Attrition = 1|X_i)}{1 - P(Attrition = 1|X_i)}\right)
&= \beta_0 + \beta_1 Age_i + \beta_2 Age2_i + \beta_3 AgeOverTime_i + \beta_4 AgexMarried_i \\
&+ \beta_5 Education_i + \beta_6 JobInvolvement_i + \beta_7 JobSatisfaction_i \\
&+ \beta_8 lnMonthlyIncome_i + \beta_9 NumCompaniesWorked_i \\
&+ \beta_{10} TotalWorkingYears_i + \beta_{11} WorkLifeBalance_i \\
&+ \beta_{12} YearsAtCompany_i + \beta_{13} YearsSinceLastPromotion_i \\
&+ \beta_{14} Over7KM_i + \beta_{15} Married_i + \beta_{16} Female_i
\end{aligned}
$$

In order to choose between a logit and a probit model, we used a comparison of two measures that similarly interpret which model is better or more preferred. These are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) respectively. In both cases, the logit model is the better model:

Table 2. Values of AIC and BIC statistics for probit and logit model.

	Probit	**Logit**
AIC	1079.723	1068.938
BIC	1169.704	1158.919

Source: Own elaboration

Thus, in the following section, we will look at the estimation of the logit model taking into account the General-to-Specific procedure. It allows the iterative elimination of successive least significant variables in order to obtain a restricted model with all significant variables.

[2] Details of the derivation of the formula are available in Appendix A.

Table 3. Estimation of Linear Probability Model, probit model and logit model for all variables.

	Dependent variable:		
	Attrition		
	OLS (LPM)	probit	logistic
	(1)	(2)	(3)
Age	-0.044***	-0.165***	-0.269***
	(0.008)	(0.034)	(0.060)
Age2	0.0005***	0.002***	0.003***
	(0.0001)	(0.0004)	(0.001)
AgeOverTime	0.005***	0.021***	0.041***
	(0.001)	(0.002)	(0.005)
AgexMarried	0.004	0.013	0.018
	(0.002)	(0.010)	(0.018)
Education	0.008	0.035	0.064
	(0.009)	(0.044)	(0.081)
JobInvolvement	-0.070***	-0.324***	-0.579***
	(0.013)	(0.060)	(0.109)
JobSatisfaction	-0.035***	-0.159***	-0.305***
	(0.008)	(0.039)	(0.071)
lnMonthlyIncome	-0.072***	-0.362***	-0.654***
	(0.020)	(0.101)	(0.188)
NumCompaniesWorked	0.017***	0.083***	0.154***
	(0.004)	(0.019)	(0.034)
TotalWorkingYears	-0.002	-0.020	-0.039
	(0.002)	(0.013)	(0.024)
WorkLifeBalance	-0.027	-0.112	-0.248*
	(0.014)	(0.060)	(0.109)
YearsAtCompany	-0.001	-0.004	-0.025
	(0.002)	(0.014)	(0.027)
YearsSinceLastPromotion	0.010**	0.056**	0.120**
	(0.003)	(0.019)	(0.036)
Over7KM	0.049**	0.277**	0.476**
	(0.018)	(0.088)	(0.161)
Married	-0.194*	-0.661	-1.056
	(0.082)	(0.360)	(0.657)
Female	-0.029	-0.138	-0.293
	(0.018)	(0.090)	(0.165)
Constant	1.979***	6.641***	11.824***
	(0.223)	(0.970)	(1.781)
Observations		1,470	1,470
Log Likelihood		-522.861	-517.469
Akaike Inf. Crit.		1,079.723	1,068.938
Bayesian Inf. Crit.		1,169.704	1,158.919
Note:	*p<0.05; **p<0.01; ***p<0.001		

5.1. “General - to - Specific” procedure

We use the Likelihood Ratio Test to verify the combined insignificance of the eliminated variables for the logit model. It is calculated using the formula below, and the value of the test statistic is then compared with a chi-square distribution with g degrees of freedom.

$$LR = 2\left[lnL\left(\hat{\theta}\right) - lnL(\widehat{\theta_R})\right] \rightarrow \chi_g^2$$

where $\hat{\theta}$ means unrestricted maximum likelihood estimator and $\widehat{\theta_R}$ means restricted maximum likelihood estimator.

In the first iteration, we see that the least significant variable is the level of education (the p-value for this variable is 0.43). Let's check the pooled insignificance of the rejection of this variable using the Likelihood Ratio Test. The p-value for this test is 0.4292, so we decide not to reject the null hypothesis of the insignificance of the parameter standing by the Education variable.

$$\begin{cases} H_0: \beta_{Education} = 0 \\ H_1: all\ variables\ are\ jointly\ significant \end{cases}$$

In the second iteration, we see that the least significant variable is the number of years in the current company (p-value for this variable is 0.348). Let's check the cumulative rejection insignificance of this variable and all the previous ones using the Likelihood Ratio Test. The p-value for this test is 0.4703, so we choose not to reject the null hypothesis of the insignificance of the parameters standing by the Education and YearsAtCompany variables.

$$\begin{cases} H_0: \beta_{Education} = 0\ \wedge\ \beta_{YearsAtCompany} = 0 \\ H_1: all\ variables\ are\ jointly\ significant \end{cases}$$

In the third iteration, we see that the least significant variable is the interaction between the worker's age and whether they are married (the p-value for this variable is 0.2777). Let's check the cumulative rejection insignificance of this variable and all the previous ones using the Likelihood Ratio Test. The p-value for this test is 0.4413, so we choose not to reject the null hypothesis of the insignificance of the parameters standing by the AgexMarried, Education and YearsAtCompany variables.

$$\begin{cases} H_0: \beta_{Education} = 0\ \wedge\ \beta_{YearsAtCompany} = 0 \wedge\ \beta_{AgexMarried} = 0 \\ H_1: all\ variables\ are\ jointly\ significant \end{cases}$$

In the fourth iteration, we see that the least significant variable is whether the person is female (the p-value for this variable is 0.083). Let's check the cumulative rejection insignificance of this variable and all the previous ones using the Likelihood Ratio Test. The p-value for this test is 0.2197, so we choose not to reject the null hypothesis of insignificance of the parameters standing by the variables Female, AgexMarried, Education and YearsAtCompany.

$$\begin{cases} H_0: \beta_{Education} = 0 \ \wedge \ \beta_{YearsAtCompany} = 0 \wedge \ \beta_{AgexMarried} = 0 \wedge \ \beta_{Female} = 0 \\ H_1: all\ variables\ are\ jointly\ significant \end{cases}$$

Let us now check that the resulting model has the correct functional form. For this purpose we should use the LINKTEST test, the Hosmer-Lemeshow test and the Osius-Rojek test.

LINKTEST test:

Table 4. LINKTEST result for the model before removing the interaction

	Estimate	Std. Error	Z Value	Pr(>|Z|)
(Intercept)	0.13796	0.13719	1.006	0.314599
yhat	1.41693	0.15202	9.321	< 2e-16 ***
yhat2	0.13782	0.04025	3.425	0.000617 ***

Source: Own elaboration

Hosmer-Lemeshow and Osius-Rojek tests:

Table 5. Hosmer-Lemeshow and Osius-Rojek test results before removal of the interaction.

Test	Stat	Val	df	P-value
HL	chiSq	24.8308	8	0.00166
OsRo	Z	4.90706	NA	9.2e-7

Source: Own elaboration

Unfortunately, the form of our proposed model is inadequate so that we cannot consistently interpret the estimated parameters. In spite of all the relevant variables, we decided to remove one more variable which has the worst effect on the functional form of the logit model. Unfortunately, the proposed interaction between an employee's age and whether they work overtime gave the worst results. So let's check by Likelihood Ratio Test the combined insignificance of the rejected variables. The p-value for this test is close to zero, so the previously rejected variables are jointly significant.

$$\begin{cases} H_0: \beta_{Education} = 0 \ \land \ \beta_{YearsAtCompany} = 0 \land \ \beta_{AgexMarried} = 0 \land \ \beta_{Female} = 0 \land \ \beta_{AgeOverTime} = 0 \\ H_1: all\ variables\ are\ jointly\ significant \end{cases}$$

Let us now check that the resulting model has the correct functional form. For this purpose we should use the LINKTEST test, the Hosmer-Lemeshow test and the Osius-Rojek test.

LINKTEST test:

Table 6. LINKTEST result for the model before removing the interaction

	Estimate	Std. Error	Z Value	Pr(>|Z|)
(Intercept)	0.11444	0.16254	0.704	0.481
yhat	1.26662	0.20071	6.311	2.78e-10 ***
yhat2	0.09179	0.06099	1.505	0.132

Source: Own elaboration

Hosmer-Lemeshow and Osius-Rojek tests:

Table 7. Hosmer-Lemeshow and Osius-Rojek test results after removal of the interaction.

Test	Stat	Val	df	P-value
HL	chiSq	6.2444	8	0.61987
OsRo	Z	1.6473	NA	0.09948

Source: Own elaboration

By removing the interactions, our functional form of the model is correct and we can confidently interpret the estimated parameters. Unfortunately, this is a certain trade-off we had to make in order to get the correct functional form, which we think is much more important than introducing interactions into the model.

Finally, we completed the General-to-Specific procedure with the following final model:

$$ln\left(\frac{P(Attrition = 1|X_i)}{1 - P(Attrition = 1|X_i)}\right)$$
$$= \beta_0 + \beta_1 Age_i + \beta_2 Age2_i + \beta_6 JobInvolvement_i + \beta_7 JobSatisfaction_i$$
$$+ \beta_8 lnMonthlyIncome_i + \beta_9 NumCompaniesWorked_i$$
$$+ \beta_{10} TotalWorkingYears_i + \beta_{11} WorkLifeBalance_i$$
$$+ \beta_{13} YearsSinceLastPromotion_i + \beta_{14} Over7KM_i + \beta_{15} Married_i$$

5.2. Interpretation of Pseudo R^2 statistics

Let's check the values for the Pseudo R^2 statistics and try to interpret them for the final logit model.

Table 8. Table of calculated R^2 statistics.

McKelvey & Zavoina	Count R^2	Adjusted Count R^2
0.24561632	0.84829932	0.05907173

Source: Own elaboration

For models that explain variation in the latent variable, the McKelvey and Zavoina R^2 is a measure of fit. It shows what, at least, part of the variation in the dependent variable could be explained by the variables included in the model if the values of the dependent variable were observed. From our model, the value of this measure was 24.5% which translates to that what, at least, part of the variation in the dependent variable could be explained by the variables included in the model if the values of the latent variable were observed.

For a model with discrete dependent variables, a Count R^2 (Count-R^2) is also defined. It shows what percentage of the value of the dependent variable can be correctly predicted using the estimated values of the model parameters. Our model can correctly predicts 85% values of the dependent

Without knowing anything about the predictors, one could always predict the more common outcome and be right the majority of the time. An effective model should improve on this null model, and so this null model is the baseline for which the Count R^2 is adjusted. Our model, therefore, after adjusting by removing the most frequent observations, has an Adjusted Count R^2 statistic of 5.9%

Table 9. Table of the estimated models using the General-to-specific procedure.

	Dependent variable: Attrition					
	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.269***	-0.259***	-0.265***	-0.270***	-0.268***	-0.260***
	(0.060)	(0.059)	(0.059)	(0.058)	(0.058)	(0.056)
Age2	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AgeOverTime	0.041***	0.041***	0.041***	0.041***	0.041***	
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	
AgexMarried	0.018	0.019	0.020			
	(0.018)	(0.018)	(0.018)			
Education	0.064					
	(0.081)					
JobInvolvement	-0.579***	-0.576***	-0.577***	-0.573***	-0.568***	-0.538***
	(0.109)	(0.109)	(0.109)	(0.109)	(0.109)	(0.104)
JobSatisfaction	-0.305***	-0.305***	-0.307***	-0.305***	-0.301***	-0.280***
	(0.071)	(0.071)	(0.071)	(0.071)	(0.071)	(0.069)
lnMonthlyIncome	-0.654***	-0.653***	-0.659***	-0.662***	-0.665***	-0.625***
	(0.188)	(0.188)	(0.188)	(0.189)	(0.188)	(0.180)
NumCompaniesWorked	0.154***	0.156***	0.164***	0.164***	0.162***	0.143***
	(0.034)	(0.034)	(0.032)	(0.032)	(0.032)	(0.031)
TotalWorkingYears	-0.039	-0.039	-0.051*	-0.051*	-0.051*	-0.044*
	(0.024)	(0.024)	(0.022)	(0.022)	(0.022)	(0.020)
WorkLifeBalance	-0.248*	-0.249*	-0.249*	-0.254*	-0.258*	-0.258*
	(0.109)	(0.109)	(0.109)	(0.109)	(0.109)	(0.105)
YearsAtCompany	-0.025	-0.026				
	(0.027)	(0.027)				
YearsSinceLastPromotion	0.120**	0.121***	0.103***	0.103***	0.102***	0.091**
	(0.036)	(0.036)	(0.031)	(0.031)	(0.031)	(0.029)
Over7KM	0.476**	0.481**	0.478**	0.473**	0.471**	0.440**
	(0.161)	(0.160)	(0.160)	(0.160)	(0.160)	(0.153)
Married	-1.056	-1.077	-1.100	-0.412*	-0.403*	-0.415**
	(0.657)	(0.656)	(0.657)	(0.164)	(0.164)	(0.158)
Female	-0.293	-0.290	-0.291	-0.285		
	(0.165)	(0.165)	(0.165)	(0.165)		
Constant	11.824***	11.793***	11.911***	11.881***	11.750***	11.330***
	(1.781)	(1.779)	(1.774)	(1.772)	(1.768)	(1.696)
Observations	1,470	1,470	1,470	1,470	1,470	1,470
Log Likelihood	-517.469	-517.781	-518.223	-518.816	-520.337	-560.908
Akaike Inf. Crit.	1,068.938	1,067.563	1,066.447	1,065.632	1,066.674	1,145.815

Note: *p<0.05; **p<0.01; ***p<0.001

5.3. Marginal effects of variables in the final model

Table 10. Table of calculated marginal effects for averaged observation.

Variable name	Marginal effects (dF/dx)
Age	-0.032977736
Age2	0.000393992
JobInvolvement	-0.062928774
JobSatisfaction	-0.031899565
lnMonthlyIncome	-0.071664322
NumCompaniesWorked	0.016284306
TotalWorkingYears	-0.004347906
WorkLifeBalance	-0.027117779
YearsSinceLastPromotion	0.009974947
Over7KM	0.055141588
Married	-0.044178458

Source: Own elaboration

The marginal effects from the final model were counted for the averaged observation. Let's analyze the effect of each variable on the change in probability.

To determine the marginal effect of an additional year of a worker's life, calculate the first derivative on the Age variable. For our averaged observation, an additional year of life will reduce the probability of leaving a job by 0.23 percentage points. A one-point increase in job involvement will decrease the likelihood of leaving a job by 6.29 percentage points. A one-point increase in job satisfaction will decrease the likelihood of leaving a job by 3.18 percentage points. A 1% increase in monthly income will reduce the likelihood of leaving a job by 7.16 percentage points. An additional unit in the number of firms for which the worker worked will increase the probability of leaving by 1.62 percentage points. An additional year of service for an employee will translate into a 0.43 percentage point decrease in the likelihood of leaving a job. Increasing the quality of work-life balance by a unit will translate into a 2.71 percentage point decrease in the likelihood of leaving a job. An extra year waiting for promotion will translate into a 0.99 percentage point increase in the probability of leaving your job. If an employee has to commute to a workplace at least 7 kilometers away, this will increase the probability of leaving the workplace by 5.51 percentage points. Married people will be 4.41 percentage points less willing to change jobs.

5.4. Validation of hypotheses

At this stage, we can proceed to verify the research hypotheses set. First we analyze the main research hypothesis about the influence of non-wage factors on employees' quitting. For this purpose we conducted Likelihood Ratio Test for unrestricted model (final version of the logit model) and restricted model (logit model with one explanatory variable - lnMonthlyIncome). P-value for the conducted test is close to 0, so we reject the null hypothesis of total non-suitability of non-wage factors.

Turning to secondary hypotheses, we can divide them into one-sided and two-sided hypotheses. The one-sided hypotheses will be: 'Married people are less likely to change jobs because they care more about the financial stability of their family' and 'People who are more satisfied with their job are less likely to decide to change jobs'. In this case, we assume that the alternative hypothesis is that the parameter β is statistically significant and greater or less than zero, respectively. To verify the hypothesis that being married has a negative effect on willingness to change jobs, we need to look at the parameter estimate standing by the variable Married ($\beta = -0.415$), as well as the test statistic with. The one-tailed p-value in this case is 0.004294, which means that the parameter estimate is statistically significant and supports the hypothesis. Unlike in the study of Corneließen (2006) or D'Addio and Frijters (2011), we were able to confirm the existence of a negative effect of matrimonial status on job quits.

Analogous to the study by Kristensen and Westergård-Nielsen (2004), it was also possible to confirm the hypothesis of a negative effect of increased job satisfaction on the decision to leave. In this case, the standing parameter estimate for the JobSatisfaction variable was $\beta = -0.2798605$, while the test z-statistic = -4.079, for which the corresponding one-tailed p-value is close to zero.

In the case of the remaining hypotheses for which we tested whether the parameter estimate is statistically different from zero (alternative hypothesis, where β != 0) we were able to confirm the statistical significance of the age effect on the decision to quit work. Similar conclusions were reached by the authors of Weiss (1984) and Kristensen and Westergård-Nielsen (2004), while the study by Corneließen (2006) failed to confirm this hypothesis.

In our research, it was not possible to prove the statistical significance of the parameter estimate next to the Education variable. For both the logit and probit model, the mentioned variable had a p-value that did not allow us to reject the null hypothesis of insignificance of this parameter. The conclusion therefore differs from the previous studies (Weiss 1984).

Summary

The topic of attrition or employees leaving the workplace is constantly analysed in the wider literature and is also a major issue for employees and HR managers to reflect on. When analysing attrition, we should not only consider wage factors, as in many situations these do not play the main role in the decision to leave a job. We should also consider non-salary factors such as relationship with the superior, frequency of promotions, age of the employee, job and task satisfaction. These are just a few of the many potentially important factors that HR can control.

From the econometric analysis presented, it was possible to extract interesting relationships between a number of relevant employee characteristics and employee attrition. Factors that reduce the likelihood of job change include employee age, job and task engagement, job satisfaction, monthly income, work-life balance and being married. On the other hand, characteristics such as the number of companies an employee has worked for, the number of years since the last promotion and a distance of more than 7KM to the workplace all contribute to the likelihood that an employee will be more likely to change jobs.

More accurate machine learning models such as K Nearest Neighbours, Support Vector Machines, Decision Trees, Random Forests or even neural networks could be used to explore this issue in even greater depth. With such algorithms, we would be able to get to know our employees even more accurately and provide them with the right working conditions and motivate them to learn more about reality.

Bibliography

Cornelißen, T. (2006). *Job characteristics as determinants of job satisfaction and labour mobility* (No. 334). Diskussionsbeitrag.

D'Addio, A. C., Eriksson, T., & Frijters, P. (2007). An analysis of the determinants of job satisfaction when individuals' baseline satisfaction levels may differ. *Applied economics*, *39*(19), 2413-2423.

Diaz-Serrano, L., & Cabral Vieira, J. A. (2005). Low pay, higher pay and job satisfaction within the European Union: Empirical evidence from fourteen countries. *Higher Pay and Job Satisfaction within the European Union: Empirical Evidence from Fourteen Countries (April 2005)*.

Haar, J. M., Russo, M., Suñe, A., & Ollier-Malaterre, A. (2014). Outcomes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of Vocational Behavior*, *85*(3), 361-373.

Krishna, S., & Sidharth, S. (2022). HR Analytics: Employee Attrition Analysis using Random Forest. *International Journal of Performability Engineering*, *18*(4).

Kristensen, N., & Westergård-Nielsen, N. (2004). Does low job satisfaction lead to job mobility?. *Available at SSRN 511722*.

Lopes, H., Lagoa, S., & Calapez, T. (2014). Work autonomy, work pressure, and job satisfaction: An analysis of European Union countries. *The Economic and Labour Relations Review*, *25*(2), 306-326.

Spies, M. (2006). Distance between home and workplace as a factor for job satisfaction in the North-West Russian oil industry. *Fennia-International Journal of Geography*, *184*(2), 133-149.

Weiss, A. (1984). Determinants of quit behavior. *Journal of Labor Economics*, *2*(3), 371-387.

Veall, Michael R. & Zimmermann, Klaus F. (1996): Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Collaborative Research Center 386*, Discussion Paper 18

Vila, L. E., & García-Mora, B. (2005). Education and the determinants of job satisfaction. *Education Economics*, *13*(4), 409-425.

Appendix A

A mathematical transformation of the probability formula to a log of the odds ratio is shown below:

$$p_i = P(y = 1|X_i) = f\big(g(x, \beta)\big) = \frac{1}{1 + e^{-g(x,\beta)}} = \frac{e^{g(x,\beta)}}{1 + e^{g(x,\beta)}}$$

where $g(x, \beta) = \beta_0 + \sum_{j=1}^{m} \beta_j x_j + \varepsilon_i$ is a linear combination of the estimated parameters and the independent variables. By performing some transformations of the above formula, a logit model was finally obtained.

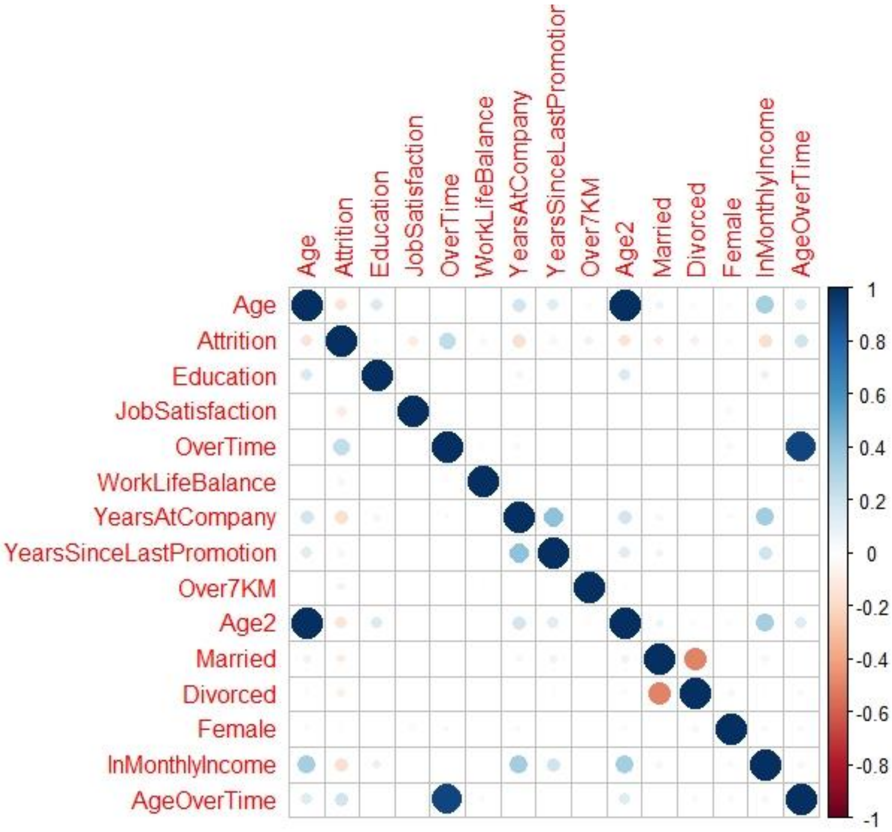
$$p_i * \big(1 + e^{g(x,\beta)}\big) = e^{g(x,\beta)}$$

$$p_i = e^{g(x,\beta)} * (1 - p_i)$$

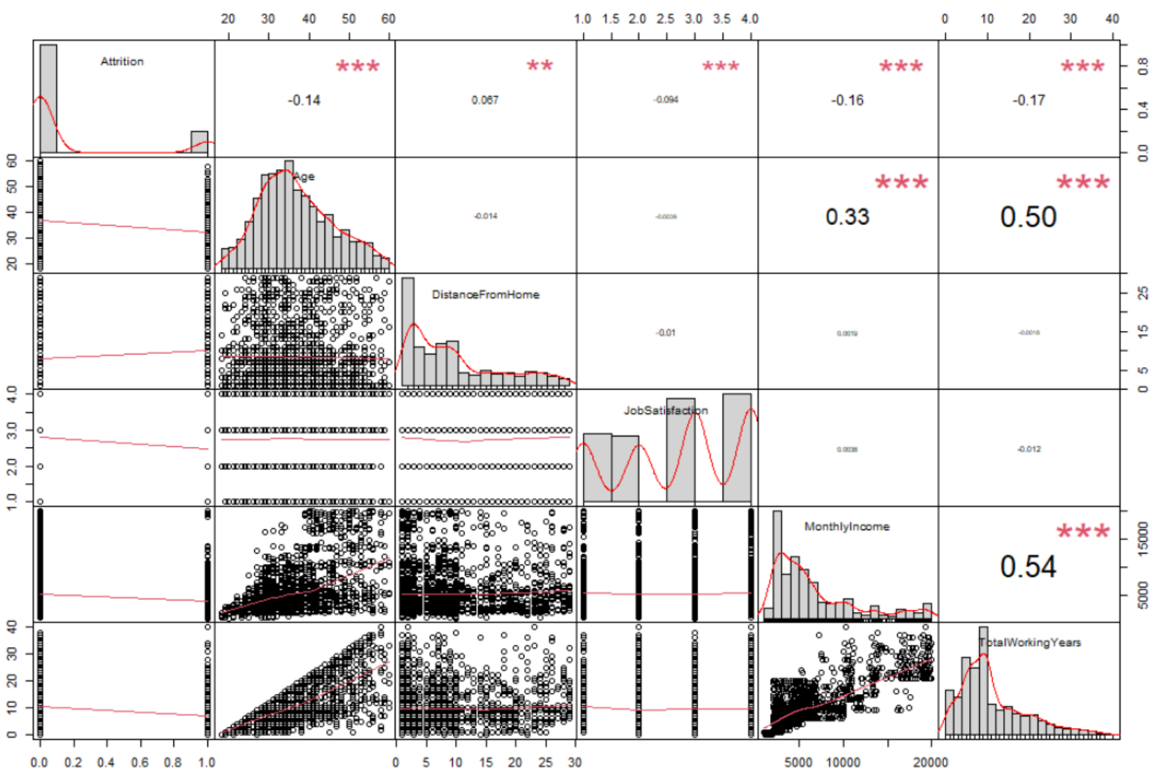
$$\frac{p_i}{(1 - p_i)} = e^{g(x,\beta)}$$

$$\ln\left(\frac{p_i}{(1 - p_i)}\right) = g(x, \beta) = \beta_0 + \sum_{j=1}^{m} \beta_j x_j$$

Appendix B



Age
Attrition
Education
JobSatisfaction
OverTime
WorkLifeBalance
YearsAtCompany
YearsSinceLastPromotion
Over7KM
Age2
Married
Divorced
Female
lnMonthlyIncome
AgeOverTime
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1



Attrition
Age
DistanceFromHome
JobSatisfaction
MonthlyIncome
TotalWorkingYears
-0.14
0.067
-0.094
-0.16
-0.17
-0.014
0.33
0.50
-0.01
-0.012
0.54