Introduction:

The data set I considered has 14999 samples of employee data which contains different sets evaluation columns of the employees. The left column contains a value of ‘0’ and ‘1’. ‘0’ denotes the employees who are staying in the organization and ‘1’ denotes the employees who left the organization. The satisfaction levels denote the employee’s satisfaction towards the company. The average monthly hours denote the average monthly working hours in the company by the employee.

|  |  |  |
| --- | --- | --- |
| **Correlations** | | |
|  | | satisfaction\_level | | last\_evaluation | average\_monthly\_hours |
| satisfaction\_level | Pearson Correlation | 1 | | .105\*\* | -.020\* |
| Sig. (2-tailed) |  | | <.001 | .014 |
| N | 14999 | | 14999 | 14999 |
| last\_evaluation | Pearson Correlation | .105\*\* | | 1 | .340\*\* |
| Sig. (2-tailed) | <.001 | |  | .000 |
| N | 14999 | | 14999 | 14999 |
| average\_monthly\_hours | Pearson Correlation | -.020\* | | .340\*\* | 1 |
| Sig. (2-tailed) | .014 | | .000 |  |
| N | 14999 | | 14999 | 14999 |
| To prove the multicollinearity assumption, I created a correlations table between the left and the predictor variables. To prove the multicollinearity assumption the correlation values should be less than 0.8 between the different predictor values. Since in the above table the correlation values are less than 0.8 thus multicollinearity is satisfied. | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | |
|  | | B | S.E. | Wald | df | Sig. |
| Step 1a | satisfaction\_level | -4.022 | .128 | 992.916 | 1 | <.001 |
| last\_evaluation | -15.899 | .729 | 476.100 | 1 | <.001 |
| average\_monthly\_hours | .057 | .003 | 267.214 | 1 | <.001 |
| SATISFACTION\_N by satisfaction\_level | 2.260 | .354 | 40.778 | 1 | <.001 |
| LASTEVELUTAION\_N by last\_evaluation | 25.032 | 1.118 | 501.750 | 1 | <.001 |
| AVERAGE\_MONTHLY\_N | -10.640 | .656 | 262.656 | 1 | <.001 |
| Constant | 63.146 | 2.842 | 493.739 | 1 | <.001 |

To prove linearity assumptions testing I created a binary logistic regression table. In the above table, the significance column for all the predictor variables is less than 0.001.Since the relationship is significant there could be some concerns with the linearity assumption .However because of the large sample size less is the concern.

|  |  |  |  |
| --- | --- | --- | --- |
| **Case Processing Summary** | | | |
| Unweighted Casesa | | N | Percent |
| Selected Cases | Included in Analysis | 14999 | 100.0 |
| Missing Cases | 0 | .0 |
| Total | 14999 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 14999 | 100.0 |
| a. If weight is in effect, see classification table for the total number of cases. | | | |

The above table gives the summary of the data set. The dataset I considered contains 14999 entries wherein which 14999 samples are included in the analysis. The dependent variable encoding contains 0 and 1 values. In which ‘0’ denotes employees who are not promoted and ‘1’ denotes employees who got promoted.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Table**a,b | | | | | |
|  | Observed | | Predicted | | |
| left | | Percentage Correct |
| 0 | 1 |
| Step 0 | left | 0 | 11428 | 0 | 100.0 |
| 1 | 3571 | 0 | .0 |
| Overall Percentage | |  |  | 76.2 |
| a. Constant is included in the model. | | | | | |
| b. The cut value is .500 | | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 0 | Constant | -1.163 | .019 | 3681.477 | 1 | .000 | .312 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables not in the Equation** | | | | | |
|  | | | Score | df | Sig. |
| Step 0 | Variables | satisfaction\_level | 2262.376 | 1 | .000 |
| last\_evaluation | .647 | 1 | .421 |
| average\_monthly\_hours | 76.223 | 1 | <.001 |
| Overall Statistics | | 2334.188 | 3 | .000 |

Block 0 i.e., the ‘classification table’ gives the summary of the dataset I considered, and gives the percentage of the correctness of the variables that are included in the analysis. In the above example, I considered leaving as the predictor variable. In which ‘0’ denotes the employees left from the organization and ‘1’ denotes the employees who are currently working in the organization. The percentage of the correctness of the employees who stayed in the organization is cent percent while the overall percentage of the model is 76.2%.

‘Variables not in the equation’ denotes the ‘null model’ In this model there are no predictor variables included whereas only intercepts of the variable are included. The df in the above table denotes degrees of freedom whereas the sig. denotes the significance of the predictor intercepts with the model.

**Block 1: Method = Enter**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Omnibus Tests of Model Coefficients** | | | | |
|  | | Chi-square | df | Sig. |
| Step 1 | Step | 2302.049 | 3 | .000 |
| Block | 2302.049 | 3 | .000 |
| Model | 2302.049 | 3 | .000 |

The above table represents the logical regression analysis of the model. In the above model the predictor variables I considered turned out to be better fit to the model since the significance is less than 0.05, the chi squared value is 2302.049.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summary** | | | |
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 14162.641a | .142 | .214 |
| a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001. | | | |

-2 log denotes the likelihood of the final model in the above results obtained the value is 13277.045, the cox and Snell R square is the replacement of the r squared value.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Table**a | | | | | |
|  | Observed | | Predicted | | |
| left | | Percentage Correct |
| 0 | 1 |
| Step 1 | left | 0 | 10621 | 807 | 92.9 |
| 1 | 2634 | 937 | 26.2 |
| Overall Percentage | |  |  | 77.1 |
| a. The cut value is .500 | | | | | |

The observed column indicates the dependent variable left. The percentage of the correctness denotes the model accuracy of the full model. In the above results the model accuracy has increased from 76.2% in ‘Block -0’ to 77.1% in ‘Block-1’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) |
|
| Step 1a | satisfaction\_level | -3.814 | .088 | 1879.237 | 1 | .000 | .022 |
| last\_evaluation | .284 | .132 | 4.617 | 1 | .032 | 1.328 |
| average\_monthly\_hours | .002 | .000 | 17.244 | 1 | <.001 | 1.002 |
| Constant | .399 | .110 | 13.054 | 1 | <.001 | 1.491 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables in the Equation** | | | |
|  | | 95% C.I.for EXP(B) | |
| Lower | Upper |
| Step 1a | satisfaction\_level | .019 | .026 |
| last\_evaluation | 1.025 | 1.721 |
| average\_monthly\_hours | 1.001 | 1.003 |
| Constant |  |  |
|  |  |  |  |  |  |  |  |

|  |
| --- |
|  |

I had performed binary logistic regression to find the different set of values like odds ratio, statistical significance, confidence interval and degrees of freedom.

The first column B tells us the values of the logistic regression for predicting the dependent variable from the independent variable. In the above table the satisfaction level has the negative B value with the left while others have positive B value with the left column.

The B value can be also written as the for every unit increase in the average monthly hours in the company by the employees there, we can expect a decrease of satisfaction level of the employees holding other predictor variables constant. And, for every unit increase in the last evolution level we can expect a decrease in the satisfaction level.

The 95% of the confidence interval tells us the odds ratio of the predictor is significant. If the points plotted on the number line and if 1 is not included in the number line

The Exp(B) column interprets the odds ratio for the predictors. The odds ratio is examined vis these assumptions. If the odds ratio is equal to 1.0 then there is the no effect, if it is greater than 1.0 then there is a positive effect and if less than 1.0 then there is negative effect.

In the above table the odds ratio of the last evaluation level and the average monthly hours is greater than 1 indicating that the people who had worked for more hours in the monthly basis and the who had shown greater level of last evaluation are likely to leave the organization .And, when coming to satisfaction level the people who has shown lower level of satisfaction level are more likely to leave the organization.

The 95% of the confidence interval tells us the odds ratio of the predictor is significant. If the points plotted on the number line and if 1 is not included in the number line this means the adjusted odds ratio is statically significant from 1.0. In the above results the last evaluation and the satisfaction level is significant.

The odds ratio of the different predictor variables with the people who stayed in the company. The odds ratio of the employees to the satisfaction level is(1/0.22=4.54).This can be written as the employees shown lower satisfaction level are more likely to leave the organization since the odds ratio is more than 1, when coming to last evaluation level to the left variable (1/1.328=0.75).This can be written as the employees who shown greater level of last evaluation are more likely to leave the organization since the odds ratio is less than 1 .And the last predictor variable I considered is average monthly hours and odds ratio is (1/1.002=0.998).

Conclusion:

I would recommend the director of human resources to include more predictor variables for analyzing the reasons for the resignation. But the predictor variables the I included has given some valuable insights to predict the left variable. The results found that satisfaction level is one of the important reasons for the resignation while the last evaluation level and average monthly hours are the key reasons for the resignation from the organization.