

Understanding Workplace Turnover with HR Analytics

Erez S. Sarousi

Bellevue University

Topic

In HR, it is incredibly important to find and retain good employees. This project entitled, Understanding Workplace Turnover with HR Analytics, exists to find the causes of turnover.

Business Problem

It is said that, “the best workers do the best and the most work. But many companies do an awful job of finding and keeping them” (Keller, 2021). Retaining employees can make or break a company. In fact, nearly a quarter of all US workers quit their jobs in 2006 (Allen, 2008). More than resulting in just low staffing levels; there are several causal factors to explain this.

Turnover is costly. The exact numbers vary between occupations and industries, but it can be 50-60% of a salary to directly hire with the total costs being as high as 200%. This can be as high as \$133,000 for an HR professional or \$150,000 for an accountant.

Turnover also affects performance. High turnover inhibits productivity and fosters low morale. However, reducing turnover shows immediate benefits. It has been shown that reducing turnover increases profitability and market value, in addition to improving workplace morale and sales growth. Thus, reducing turnover doesn't just save headaches; it saves companies.

Data Explanation

The dataset from Kaggle, <https://www.kaggle.com/giripujar/hr-analytics>, focused on human resources analytics. The columns are:

- Satisfaction Level – Representation of happiness within the employee's job.

- Last Evaluation - Representation of performance within their last assessment.
- Number of Projects – Number of projects managed by the employee concurrently.
- Average Monthly Hours – Amount of worked hours within a month.
- Time Spent with the Company – Amount of years employed by the company.
- Work Accident – Whether the employee experienced an accident at work.
- Left – The target variable describing if the employee left the job.
- Promotion with the Last Five Years - Represents if the employee was promoted within the last five years.
- Sales – A list of different departments including: sales, accounting, HR, technical, support, management, IT, product manager, marketing and research and development.
- Salary – How much the employee is paid. The unique values are: low, medium, high.

Work Accident, Left, and Promotion with the Last Five Years are categorical. Satisfaction Level, and Last Performance range between 0 and 1. Sales categorizes several different departments. Salary is an ordinal value.

Methods

Steps were taken in order to find out what causes turnover. Firstly, the Salary column was converted to categorical integers. Secondly, Get Dummies was utilized to create columns for each department. A list was then created for all predictors, which was cross-referenced with the Left column with statistical processes such as Pearson's Correlation, hyperparameter tuning, tests for multicollinearity, logistic regression, ridge regression, and interaction term analysis.

Analysis

Correlation analysis was conducted between the target variable and all predictors, finding moderate relationships with Satisfaction Level (~ -0.388) and small relationships with Time Spent with the Company (~ 0.145), Work Accident (~ -0.155) and Salary (~ 0.158) (Pawar, 2018).

Due to the unexpected negative correlation between the target variable and Work Accident, interaction term analysis was conducted with Time Spent with the Company as a moderator, resulting in a p value of ~ 0.006 .

Because the predictors seemed similar in nature, multicollinearity testing was conducted. The following features had high VIF levels: Satisfaction Level (~ 6.503), Last Evaluation (~ 20.260), Number of Projects (~ 13.861), Average Monthly Hours (~ 19.324) and Time Spent with the Company (~ 6.714). The remaining two features had moderate VIF scores: Work Accident (~ 1.173), Promotion with Five Years (~ 1.037).

Despite the multicollinearity, logistic regression was conducted due to the binary nature of the target variable; this was done along with Ridge Regression. Both regression analyses were done after hyperparameter tuning was implemented. Logistic regression resulted in the most important features being Satisfaction Level with a coefficient of ~ 1.027 , Work Accident ~ 0.516 , Salary ~ 0.440 , if they work in R and D (~ 0.456), and management (~ 0.414). They all had p values of < 0.001 . Ridge regression found important features being Satisfaction Level with a coefficient of ~ 0.160 , working in HR with ~ 0.099 , Support with ~ 0.071 and Technical of ~ 0.071 .

Independent t-tests have been conducted on all predicting variables deemed to be important for either ridge or logistic regression with an alpha level of .05. The following features

being considered are: Satisfaction Level, Work Accident, Salary, Research and Development, Management, HR, Support, and Technical. Based on the results of a one-tailed independent t test, Satisfaction Level had a p value of < 0.001 , Work Accident < 0.001 , Salary < 0.001 , Research and Development < 0.001 , Management < 0.001 , HR ~ 0.999 , Support ~ 0.902 , and Technical ~ 0.992 . These tests have concluded that the variables of Satisfaction Level, Work Accident, Salary and those working in Research and Development and Management reject the null hypothesis.

Conclusion

Hyperparameter tuning concluded that the features most important to the analysis are Satisfaction Level, Number of Projects, Time Spent with the Company, Work Accident, and Salary. It seems clear that low satisfaction level, a larger number of projects, longer tenure with the company, lower pay are causal factors that lead to turnover. While the data seems to suggest that being in a work accident has a strong relationship with staying with the company, it is only the case because of the time spent with the company. The longer one stays with the company, the more likely they are to get into a work accident (Bella et al, 2013).

Assumptions

There is no supplemental information that was included with the dataset that helps to explain the features or how this data was gathered. As a result, a lot of the analyses and conclusions that have been derived from this dataset came about as a result of assumptions. These analyses could not be possible without inference from potentially ambiguous data.

Limitations

There were some that prevented further study from being done. For example, the salary column had three classifications and not the spectrum of continuous data. As a result, we lost the ability to look into how important salary was. Furthermore, more analyses could have been done with more features such as reason for the departure including if they were terminated or left willingly as well as a review of the employee's impression of the job and management style. Furthermore, recommendations are limited because it is unknown if higher salaries would reduce turnover or if other benefits, such as better health insurance for instance, would be preferred.

Challenges

Some of the challenges posed in reference to this study included trying to parse the most reasonable conclusion given the lack of the data, adjusting labels on the x axis on graphics to ensure it looks professional as well as modeling the results of the data to ensure that the correct conclusions were reached.

Future Uses/Additional Applications

This project provided data on the causes of turnover; it also revealed indications on which departments were most and least satisfied. This could be also used to adjust business modeling, intended to optimize satisfaction level. Furthermore, quality staff in generally less satisfied departments could be acquired by incentives – potentially either by increased salary or more lucrative benefits.

Recommendations

Based on the conclusions of the project, it is recommended that greater incentives be placed to those who work in Research and Development, Management, as well as those with lower salaries and those who do not appear satisfied in their roles.

Implementation Plan

Employees with lower salaries should likely experience a pay raise whereas other causal factors should receive a generalized increase in benefits. Several non-pay benefits that are received well include paid time off, better benefits, options to improve work-life balance, greater contributions for 401k plans and family benefits such as subsidized child care (Miller, 2019).

Ethical Assessment

The ethical concerns in this report appear to be negligible. No names, identifying features nor company information were included in this report. Were this an actual company with data provided, this would be subject to regulatory scrutiny from the EEOC, local unions (if applicable), and by law.

Ten Questions From The Audience

1. How can time spent with the company be seen as a moderator between workplace accidents and if an employee left if employees are more likely to leave the longer they're there?
2. Would the increased salary be more costly to an employer than the cost of turnover itself?
3. What kind of benefits would be most helpful to employees?
4. Do employees prefer greater pay or other benefits such as PTO or better health insurance?
5. What kind of laws could exist in ethical considerations of this type of report?
6. Why wasn't moderation analysis done with all tests and with all predictors?
7. Why was logistic regression done despite the multicollinearity?
8. Could this dataset be somehow combined with another dataset that looks at other potential causes of turnover? Why was this not done?
9. It seems that the amount of hours worked is a causal factor for turnover. Why was this variable not identified as such through either logistic or ridge regression?
10. Could this dataset be used to make potential inferences to describe the work culture?

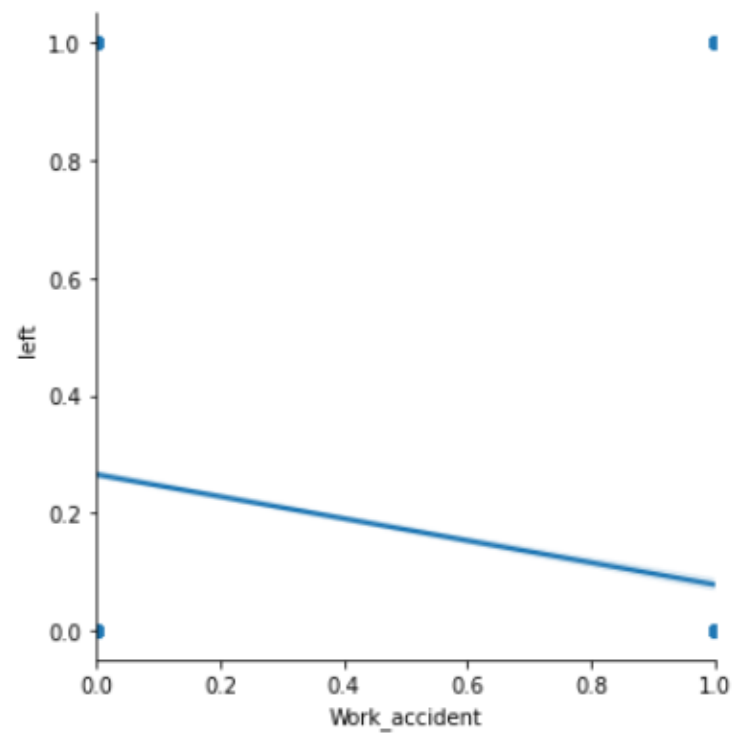
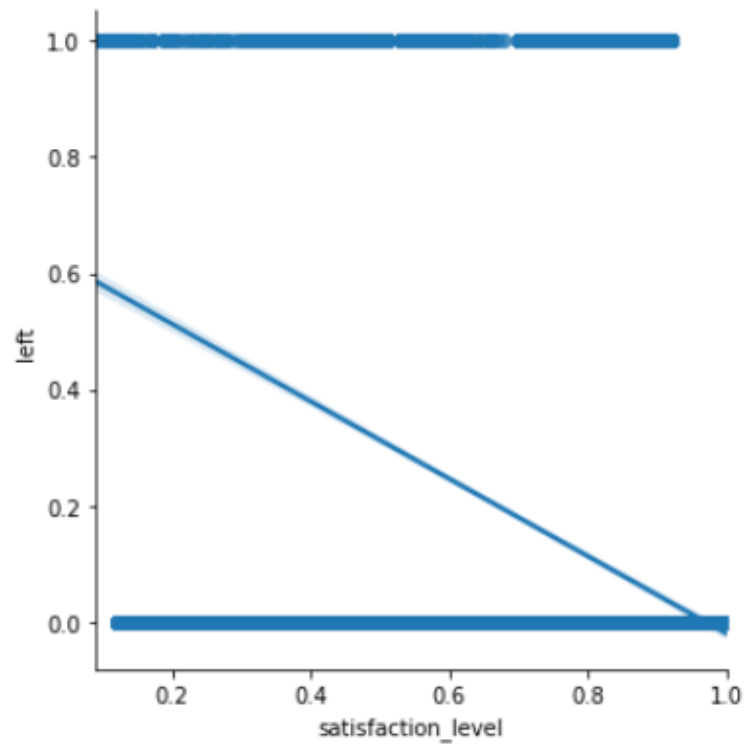
References

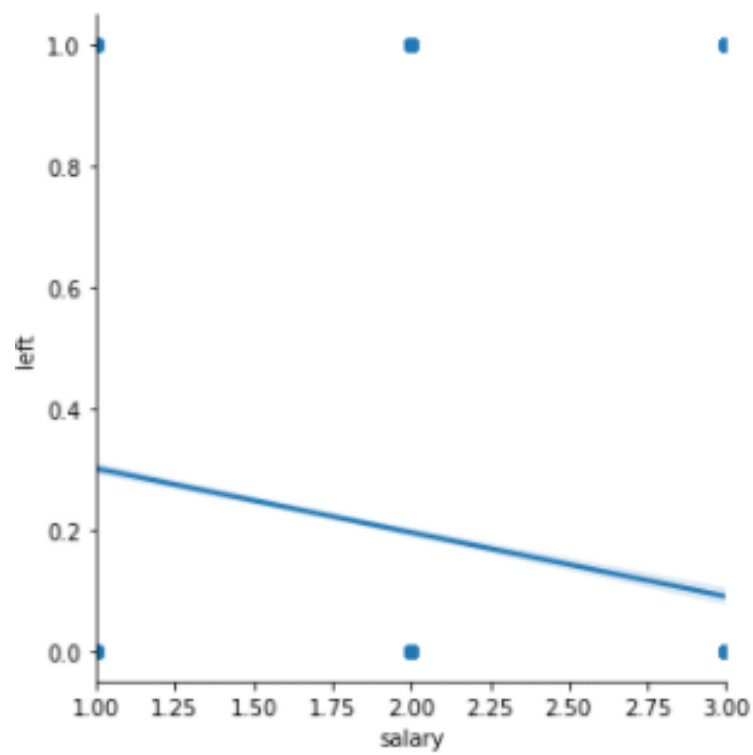
- Allen, D. G. (2008). *Retaining Talent - Society for Human Resource Management*. Retaining Talent A Guide to Analyzing and Managing Employee Turnover. Retrieved from <https://www.shrm.org/hr-today/trends-and-forecasting/special-reports-and-expert-views/Documents/Retaining-Talent.pdf>
- Bella, A., Giraudo, M., Leombruni, R., & Costa, G. (2013, September 22). *Job tenure and work injuries: a multivariate analysis of the relation with previous experience and differences by age*. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3849089/>
- Keller, S. (2021, July 7). *Attracting and retaining the right talent*. McKinsey & Company. Retrieved from <https://www.mckinsey.com/business-functions/people-and-organizational-performance/our-insights/attracting-and-retaining-the-right-talent>
- Miller, S. C. (2019, August 16). *Better Pay and Benefits Loom Large in Job Satisfaction*. SHRM. <https://www.shrm.org/resourcesandtools/hr-topics/compensation/pages/pay-benefits-satisfaction.aspx>
- Pawar, D. (2018, August 3). *Correlation — Statistical Analysis! - Dipti Pawar*. Medium. <https://medium.com/@dipti.rohan.pawar/correlation-statistical-analysis-9471411f0431>
- Sharma, G. (2021, December 16). *Effect of Multi-collinearity on Linear Regression - Analytics Vidhya*. Medium. <https://medium.com/analytics-vidhya/effect-of-multicollinearity-on-linear-regression-1cf7cfc5e8eb>

Appendix

promotion_last_5years	salary	IT	RandD	accounting	hr	management	marketing	product_mng	sales	support	technical
0.025605	0.050022	0.006373	0.006615	-0.028649	-0.012841	0.007172	0.005715	0.006919	0.004007	0.009185	-0.009345
-0.008684	-0.013002	0.001269	-0.005471	0.002193	-0.009645	0.009662	-0.000311	-0.001989	-0.023031	0.017104	0.013742
-0.006064	-0.001803	0.003287	0.009703	0.004189	-0.027356	0.009728	-0.023064	0.000829	-0.013388	0.000303	0.028596
-0.003544	-0.002242	0.006967	-0.001177	0.000524	-0.010783	0.000834	-0.008210	-0.005494	-0.001718	-0.002444	0.013638
0.067433	0.048715	-0.006053	-0.021116	0.003909	-0.022194	0.115436	0.012096	-0.003919	0.015150	-0.030111	-0.027991
0.039245	0.009247	-0.009293	0.017167	-0.012836	-0.015649	0.011242	0.011367	0.001246	-0.004955	0.012079	-0.006070
-0.061788	-0.157898	-0.010925	-0.046596	0.015201	0.028249	-0.046035	-0.000859	-0.011029	0.009923	0.010700	0.020076
1.000000	0.098119	-0.038942	0.021268	-0.004852	-0.001531	0.128087	0.049253	-0.037288	0.012353	-0.035605	-0.035799
0.098119	1.000000	-0.010959	0.002800	0.012759	0.004599	0.156665	0.011599	-0.007669	-0.035599	-0.029888	-0.018630
-0.038942	-0.010959	1.000000	-0.070240	-0.069293	-0.067949	-0.062500	-0.073524	-0.075503	-0.184302	-0.124705	-0.140484
0.021268	0.002800	-0.070240	1.000000	-0.054629	-0.053570	-0.049274	-0.057965	-0.059525	-0.145300	-0.098315	-0.110755
-0.004852	0.012759	-0.069293	-0.054629	1.000000	-0.052848	-0.048610	-0.057183	-0.058723	-0.143341	-0.096989	-0.109262
-0.001531	0.004599	-0.067949	-0.053570	-0.052848	1.000000	-0.047667	-0.056075	-0.057584	-0.140562	-0.095109	-0.107143
0.128087	0.156665	-0.062500	-0.049274	-0.048610	-0.047667	1.000000	-0.051578	-0.052966	-0.129289	-0.087482	-0.098551
0.049253	0.011599	-0.073524	-0.057965	-0.057183	-0.056075	-0.051578	1.000000	-0.062308	-0.152093	-0.102911	-0.115933
-0.037288	-0.007669	-0.075503	-0.059525	-0.058723	-0.057584	-0.052966	-0.062308	1.000000	-0.156187	-0.105682	-0.119054
0.012353	-0.035599	-0.184302	-0.145300	-0.143341	-0.140562	-0.129289	-0.152093	-0.156187	1.000000	-0.257967	-0.290608
-0.035605	-0.029888	-0.124705	-0.098315	-0.096989	-0.095109	-0.087482	-0.102911	-0.105682	-0.257967	1.000000	-0.196636
-0.035799	-0.018630	-0.140484	-0.110755	-0.109262	-0.107143	-0.098551	-0.115933	-0.119054	-0.290608	-0.196636	1.000000

	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spend_company	Work_accident	left
satisfaction_level	1.000000	0.105021	-0.142970	-0.020048	-0.100866	0.058697	-0.388375
last_evaluation	0.105021	1.000000	0.349333	0.339742	0.131591	-0.007104	0.006567
number_project	-0.142970	0.349333	1.000000	0.417211	0.196786	-0.004741	0.023787
average_monthly_hours	-0.020048	0.339742	0.417211	1.000000	0.127755	-0.010143	0.071287
time_spend_company	-0.100866	0.131591	0.196786	0.127755	1.000000	0.002120	0.144822
Work_accident	0.058697	-0.007104	-0.004741	-0.010143	0.002120	1.000000	-0.154622
left	-0.388375	0.006567	0.023787	0.071287	0.144822	-0.154622	1.000000
promotion_last_5years	0.025605	-0.008684	-0.006064	-0.003544	0.067433	0.039245	-0.061788
salary	0.050022	-0.013002	-0.001803	-0.002242	0.048715	0.009247	-0.157898
IT	0.006373	0.001269	0.003287	0.006967	-0.006053	-0.009293	-0.010925
RandID	0.006615	-0.005471	0.009703	-0.001177	-0.021116	0.017167	-0.046596
accounting	-0.028649	0.002193	0.004189	0.000524	0.003909	-0.012836	0.015201
hr	-0.012841	-0.009645	-0.027356	-0.010783	-0.022194	-0.015649	0.028249
management	0.007172	0.009662	0.009728	0.000834	0.115436	0.011242	-0.046035
marketing	0.005715	-0.000311	-0.023064	-0.008210	0.012096	0.011367	-0.000859
product_mng	0.006919	-0.001989	0.000829	-0.005494	-0.003919	0.001246	-0.011029
sales	0.004007	-0.023031	-0.013388	-0.001718	0.015150	-0.004955	0.009923
support	0.009185	0.017104	0.000303	-0.002444	-0.030111	0.012079	0.010700
technical	-0.009345	0.013742	0.028596	0.013638	-0.027991	-0.006070	0.020076





RandD:

