Global Journal of Human Resource Management
Vol.5, No.1, pp.47-59, January 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

A PREDICTIVE WORKFORCE-ANALYTICS MODEL FOR VOLUNTARY EMPLOYEE TURNOVER IN THE BANKING/ FINANCIAL-SERVICE INDUSTRY

Shipra Sharma
Sharma Consulting Co.
Boston, MA
USA

Paul D. Berger*
Bentley University, Morison Hall 351
175 Forest Street
Waltham, MA 02452
U.S.A

ABSTRACT: *This paper proposes a model for identifying key variables to identify in advance which employees may leave a company in the near future. This identification may allow potential action to take place to retain the employee, or, to potentially get advance notice of the need for a replacement. We list the key variables (both dependent and independent) in our model and detail the metric we hope to use to measure each variable. The specific statistical techniques we use to analyze the model may depend on the exact form of the data provided. We are in the process of seeking real-world data to illustrate the model's use.*

KEYWORDS: Workforce model, Employee status, Employee turnover, Banking and financial services

INTRODUCTION

The aim of the study is to propose a model to identify the key variables that help determine which employees/employee segments are at a high risk of attrition, based on the data from the separated employee's database. The specifics of this study include holistic research, covering areas such as individual attributes, organizational factors and external influences, to understand the key factors affecting attrition. Depending on the form and scope of the data available, the analysis will involve whatever statistical tools are appropriate to understand the hidden drivers of attrition. The study is envisioned to focus only on the high performing employees as defined by the Human Resources Department.

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

BACKGROUND

With the aging population, lower birth rates in most developed countries in the past decades, and the relative lack of skilled workforce, the talent bench in the Banking and Financial Services sector has never been thinner than it is today. This, along with low employee satisfaction and high disengagement level in the workforce, leads to high attrition in the organization, seriously impacting the performance of the organization. High attrition rate erodes employee morale and customer loyalty.

A study conducted by Mathis & Jackson (2011) indicated that the average cost of a single turnover case in the United States is estimated to be between $3,500 and $25,000. It is evident that employee turnover impacts negatively on the financial performance of an organization.

To reduce the cost of attrition, organizations need to know the drivers of attrition before an intervention is made. This research aims to help organizations decipher the drivers of attrition and create a strategy to combat attrition using predictive workforce analytics. Analysis of data using the proposed model in this paper will enable the organizations to employ a more targeted approach towards their employee-retention strategy. Proactive prediction of attrition using predictive workforce analytics will lead to improved decision making and, in turn, save an organization's revenue.

RESEARCH AIMS

The research aims to answer the following questions:

1. Identify factors related to voluntary employee turnover by department and/or other levels

2. Understand similarities/correlation between and within the various subgroups of the workforce

3. Create a model to predict future employee turnover (department-wise/level-wise)

The specifics of the proposed study are:

1. Focus is on Banking- and Financial-services organizations .

2. Business-wise and level-wise analysis of the data to be done to create targeted retention strategy for high-performing employees.

3. The study will consider only avoidable voluntary turnover (VTO).

4. The focus will be only on the high performing employees as defined by the Human Resources department of the organization (High performing employees are those who have an

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

average rating of less than 3 over the last 5 years or for their employment tenure [whichever is lower,] with a performance rating scale of: 1= highest, 5=lowest.)

METHODOLOGY

Data will be consolidated from an organization's human-resource information-system (HRIS), exit interviews, payroll data, performance management system, and annual employee satisfaction surveys (ESS's). The scores from the ESS can be extracted by department, by business head/manager and/or by other human-resources-designated department units. The overall conceptualization of the study is to perform a regression analysis to predict employee attrition risk in an organization, based on the data of the separated employees. The "regression line" (hyper plane) will be viewed as a "risk equation," predicting a score which will be applied to assess attrition risk with current, and future, sets of employees.

Ideally, factor analysis will be performed for grouping of similar independent variables; as noted in the earlier footnote, this may depend on the sample size and actual variables available. Assuming the factor analysis, the regression analysis would be performed on the factors identified. A regression analysis may also be performed on selected individual independent variables, in order to gain core insights. If data allows, all analyses would be conducted at an overall organization level, as well as at the department level.

As noted earlier in the specifics of the proposed study, the focus is to understand the voluntary turnover patterns for high performing employees only. While the study need not limit itself to these "high performers," it is clear that it is with these employees that an intervention can potentially have the largest payoff for the organization.

Of course, at some future point, in an "ideal world," one would need to investigate different intervention strategies to find out which might be optimal. This is a complex issue that is often "left out" of a full cogent analysis of "intervention analysis." For example, there have been studies that consider whether intervention would be useful by a health-care provider to persons who have been in the hospital during in the past, say, year (e.g., Sanky, Berger & Weinberg, 2012). However, these studies have examined only "half of the problem." They analyze the frequency with which these customers go back to the hospital in the next period, as a function of various medical and demographic variables. Results often indicate that the health-care provider is able to predict well those who will end up in the hospital at least once in the subsequent year, and very well those who will end up needing more than, say, 15 days in the hospital the next year. This high level of prediction ability is necessary to potentially make intervention with these patients (e.g., weekly phone calls to remind them to take their medication, or visits each week by medical personnel) profitable!!

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

However, these studies do not approach the other side of the decision – justifiably, since likely, there is no data available to analyze. What is the cost/benefit analysis of a specific intervention? Which intervention is "optimal" from a cost/benefit perspective? There is a dearth of literature, if any, in the public domain, that considers different intervention strategies over a sufficient period of time, in conjunction with the predictive ability of the regression analysis, to fully address the issue of who should receive which, if any, intervention.

We discussed the medical setting above at some length, to illustrate that even if we end up with an excellent predictive model to predict VTO, and we anticipate that we shall, there is that "other side of the equation," which would ideally be needed to complement any decision making involved in an intervention. Of course, the intervention strategies in the VTO case would be limited, and relatively speaking, not particularly expensive. However, while in the medical case, the vast majority of the independent variables are beyond control (e.g., age, gender, current medical condition), in the VTO study many of the independent variable would be, at least to an extent, controllable, so that VTO may be able to be reduced in the future at known, and not excessively high, costs by changing the typical/traditional values of the independent variables that appear to be prominent.

The dependent variable in this research is envisioned to be a categorical variable, the employee status: 1) Active, 2) Voluntary turnover (VTO), 3) Involuntary turnover (ITO). There are 31 independent variables that we foresee as potential useful predictors. Others may be added to the group. We list them below in Table 1:

Table 1: A listing of potential independent variables, their description and likely coding

#	Independent Variable	Description	Potential Coding
		Individual Attributes	
1	EmpNo	Unique employee identification number	Specific ID number
2	Gender	Gender of the employee	1= Male; 2= Female
3	EmpAge	Age when the employee joined the organization	0= <25; 1= 25-33; 2= 34-45; 3= >45 years
4	Ethnicity	Ethnic race to which the employee belongs	1= Asian; 2= Non-Hispanic white; 3= Black or African American; 4= Hispanic
5	Education	Highest completed education level of the employee	1= High school diploma; 2= Bachelor degree;

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

			3= Masters degree; 4= PhD
6	YoG	Year of graduation	Coded into 3-year categories
7	QS	Quality of last school attended, determined by the college rank	1= Ranking: top 20; 2= Ranking 21-50; 3= Ranking: beyond 50
8	Tenure	No. of years spent with the current organization	0= 1-5 yrs of service; 1= 5.1-10 yrs of service; 2= 10.1-15 yrs of service; 3= 15.1-20 yrs of service; 4= >20.1+ years of service
9	Commute	Time spent daily to commute to the workplace : Home Pin Code-Work Pin Code	0= <10 miles; 1= 10.1-25 miles; 2= >25 miles
10	A.Promo	Average number of promotions received during the tenure of employment	Actual value, or in categories to avoid a linearity assumption. (Depends on observed frequencies)
11	CareerPath Ratio	Career Path Ratio=#promotions/(#promotions+ #transfers). This ratio can be used at the department, level, manager levels	1= Low; 2= High
12	YPJ	Average years spent in previous jobs=Total experience in all previous jobs/no. of job hops	Actual value, or in categories to avoid a linearity assumption. (depends on observed frequencies)
13	YoRE	Total years of relevant experience at the time of hire	Actual value, or in categories to avoid a linearity assumption. (depends on observed frequencies)
14	Train	No. of hours spent on training annually (apart from the mandatory trainings)	Actual value, or in categories to avoid a linearity assumption. (depends on observed frequencies)
15	Relocation	Relocation from home town for work	1= Yes; 2= No
16	Travel%	%Travel for work	0= <20%; 1= 20-50%;

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

			2= >50%
17	VolAct	Participation in voluntary activities	1= Yes; 2= No
18	RExit	Reason for exit as captured during the exit interview	1= Commute time; 2= Better opportunity (Growth/title/role); 3= Better Salary; 4= Better work/life balance 5= Other
19	Ext.Parity	A score generated by comparing the employee's salary with the salary paid by competition. Average salary for the given role can be obtained from salary.com, compensation surveys, market pay analysis & internal recruitment database	0= salary is lower than competition; 1= salary is similar to competition; 2= salary is higher than competition
20	IncYrs	# years the employee received raise above X%/year (X=cost of living index as provided by US bureau of Labor Statistics)	Actual value, or in categories to avoid a linearity assumption. (depends on observed frequencies)
Organizational Factors			
21	Manager.R	Latest rating received by employee's manager in ESS's	1=Low; 5=High
22	WorkLifeB	Work Life balance rating received by the department in ESS's	1=Low; 5=High
23	BatchParity	A score generated by comparing employee's salary with his/her peers (performing same role, having same job code and who have graduated in the same year).	0= Employee salary is lower than comparators; 1= Employee salary is similar to comparators; 2= Employee salary is higher than comparators
24	OrgGrowth	Organization's growth projection for the next 4 quarters	Actual value, or in categories to avoid a linearity assumption. (depends on observed frequencies)
25	OrgRet	Organization's retention rate by job function & quarter	Actual value, or in categories to avoid a linearity assumption.

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

			(depends on observed frequencies)
26	JobSat	Job Satisfaction rating of the employee's department (from ESS's)	1= Low; 5= High
27	SocMedia	Time spent by employees on social media/LinkedIn/Facebook/Amazon etc.	Actual value, or in categories to avoid a linearity assumption. (depends on observed frequencies)
28	Leaves	Sudden increase in Unplanned leaves, absenteeism etc.	1= Yes; 2= No
29	Growth	Organization's growth projection above GDP= OrgGrowth - ProjGDP	Coding depends upon the observed data
External Influences			
30	ProjGDP	Projected GDP growth of the country for the next 4 quarters	Coding depends upon the observed data
31	Unemp	Unemployment rate for the given period	Coding depends upon the observed data
	Dependent Variable	**Description**	**Coding**
	EmpStatus	Current employment status of the employee with the organization	0=Active; 1=VTO; 2=ITO

Of course, whether all of these independent variables are utilized depends on certain data-related conditions. The obvious condition is their availability. However, beyond that – the Organizational-Factors variables are, of course, not usable should all of the data come from one organization, or, if there is insufficient variability among the organizations included in the data set. Likewise, the External-Influences variables would not be usable if all of the data came from the same country and/or the same time period, or, again, if there is insufficient variability among the data generated from these variables.

EXPLANATION OF VARIABLE LISTING

We now provide an explanation/discussion of the variables in Table 1:

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

1. EmpNo is a unique employee identification number. This identifier will help in within subject analysis.

2. Gender of the employee is coded as 1 for male and 2 for female. Research has indicated that career growth, and pay inequity has been a key reason for VTO among women employees (Sicherman,1996).

3. EmpAge is the start age when the employee joined the organization. This can be calculated from the employee's DoB and joining date. Age is a significant variable used in determining turnover in an organization. Research has indicated that old/vintage employees are less inclined for VTO compared to the younger/ less tenured co-workers (Cotton &Tuttle,1986).

4. Ethnicity is the ethnic race to which the employee belongs. It is coded as 1=Asian; 2=Non-Hispanic white; 3=Black or African American; 4=Hispanics. Diversity and inclusion is the top priority for the Commission on Human Rights and Opportunities (CHRO) in 2016. The reason why diversity and inclusion is given so much importance in today's workplace is because VTO among employees belonging to the non-predominant ethnic groups is often higher than for the majority groups (Sorensen, 2004). It is important to understand the relationship between Ethnicity and VTO.

5. Education refers to the highest education level attained by the employee. It is coded as 1=High school/diploma; 2=Bachelor degree; 3=Master's degree; 4=PhD. Research has indicated that more highly-educated employees have a higher propensity of leaving the organization if their career goals/aspirations are not met (Hom & Griffeth,1995). We have included education as a variable to understand how the education level is related with career progression/salary progression and VTO

6. YoG refers to year the employee attained his/her highest degree. This variable helps in determining the batch parity - i.e., to determine if employees from the same graduation year are being compensated fairly.

7. QS refers to the quality of school attended by the employee, which is determined by the college rank. It is coded as 1 if the college ranks as one of the top 20 colleges in that specific field, 2= Ranking 21-50; 3= Ranking: beyond 50. This variable has been included to understand the impact of the quality of the school attended on VTO.

8. Tenure refers to the number of years the employee has spent with the current organization. It is coded as 0= 1-5 years of service; 1= 5.1-10 years of service; 2= 10.1-15 years of service; 3= 15.1-20 years of service; 4= 20+ years of service. Tenure is likely to be a significant variable in determining VTO in the organization. Research has indicated that tenured employees are less inclined to leave compared to the younger workforce. It is also important to understand what makes these tenured employees stick with the organization over the years.

9. Commute refers to the time the employee spends daily to commute from home to the workplace. It is calculated as Home Pin Code-Work Pin Code. It is coded as 0= <10 miles; 1=

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

10.1-25 miles; 2= >25 miles. It is important to study this factor as long time spent on a daily local commute might have strong relationship with VTO.

10. A.Promo refers to the average number of promotions received by the employee during the tenure of employment. A high ratio is likely an indicator that the employee is on an accelerated career path and delays in promotion maybe a reason for VTO.

11. CareerPathRatio is determined as (no. of promotions/ [no. of promotions + no. of transfers]). This ratio can be culled out by - department, job title or manager. It is coded as 1= High, 2= Low. Career Path Ratio is best used as a leading or predictive indicator in combination with analysis of employee mobility, career path, tenure, reduced turnover, employee engagement and reduced total cost of workforce. When it is analyzed in conjunction with such metrics, Career Path Ratio can be a predictor of reduced total cost of workforce, which is based on comparing a promoted employee in a new role with the expected total market compensation cost of filling the same position externally. An employee who believes that their role offers persistent growth opportunity is less likely to leave the organization than others.

12. YPJ is the average number of years spent in previous jobs = (Total experience in all previous jobs/no. of job hops.) It is likely that YPJ is best used as a leading or predictive indicator in combination with analysis of number of A.Promo and CareerPathRatio. When it is analyzed in conjunction with such metrics, YPJ can be an indicator whether the employee is high performer.

13. YoRE refers to the total years of relevant experience at the time of hire. This metric gives an indication of the inexperienced vs. experience workforce.

14. Train refers to the number of hours spent annually on training by an employee (apart from the mandatory trainings.) A higher number indicates the organization's willingness to invest in employee development and helps in employee retention.

15. Relocation indicates if the employee has relocated from the home town for work. It is coded as 1= Yes; 2= No. In case when VTO is high due to relocation of employees, the HR department needs to revisit their recruitment strategy and focus on local hires.

16. Travel% calculates the percent of work related travel the employee undertakes per month. It is coded as 0= <20% travel; 1= 20-50% travel; 2= 50+% travel. High Travel% impacts an employee's work-life balance and can be a reason for VTO.

17. VolAct determined an employee's level of participation in voluntary activities. It is coded as 1= Yes, 2= No. Active involvement in voluntary activities indicates organization citizenship behavior, an employee's involvement in the organizational activities beyond work.

18. RExit captures the reason for exit as captured during the exit interview. It is coded as 1=Commute time; 2=Better opportunity (Growth/title/role); 3=Better Salary; 4=Better worklife balance; 5= Other. This variable can give us key information regarding the reason for employee exit. Department & level-wise analysis of the reason for exit can help us form relevant retention strategies.

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

19. ExternalParity is a comparison between the salary paid by an organization for a given role/level with the industry-standard benchmark-salary. Industry benchmarked-salaries for a given role can be obtained from salary.com, compensation surveys, market pay analysis & internal recruitment databases, etc. It can be coded as 0= salary is lower than competition; 1= salary is similar to competition; 2= salary is higher than competition.

20. IncYrs is the number of years that the employee received raise above X%/year, where X=cost of living index as provided by US bureau of Labor Statistics

21. Manager.R indicates the latest rating received by employee's manager in his/her Employee Satisfaction Survey (ESS.) A rating of 1= Low and 5= High. High VTO under a manager with low ESS rating (1) can be a root cause of high potential employee VTO for that particular department (under this particular manager.)

22. WorkLifeBal – a low rating of Work Life Balance received by any particular department in ESS's may be a driver for attrition. It is coded as 1= Low, 5= high.

23. BatchParity is a score generated by comparing an employee's salary with his/her peers (performing same role, having same job code and who have graduated in the same year.) It is coded as 0= Employee salary is lower than comparators; 1= Employee salary is similar to comparators; 2= Employee salary is higher than comparators.

24. OrgGrowth refers to the Organization's growth projection for the next 4 quarters.

25. OrgRet refers to the organization's retention rate by job function & quarter.

26. JobSat refers to the Job Satisfaction rating of the employee's department (from ESS's). It is coded as 1= Low, 5= high. A low rating of job satisfaction score received by any particular department in ESS may be a driver for attrition.

27. SocMedia refers to the metrics that calculates the time spent by employees on social media/LinkedIn/Facebook/Amazon, etc. Active involvement of employee in social media/job search sites/Amazon etc., during work hours, is a strong indicator of the employee being actively disengaged and can be an indicator of attrition.

28. Leaves: Sudden increase in unplanned leaves, absenteeism, etc. is an indicator of disengagement. Analysis of employee leave patterns over a period of time can help in timely intervention by HR.

29. Growth is the organization's growth projection above GDP = OrgGrowth – ProjGDP.

30. ProjGDP: Projected GDP growth of the country for the next 4 quarters.

31. Unemp is the unemployment rate for the given period.

The dependent variable:

Employee status, the DEPENDENT VARIABLE, is categorized as below in Figure 1:

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

Figure 1: Employee Status

The elements of Figure 1 are described below:

Active: The employee is active / employed with the organization.

Voluntary Turnover (VTO): The employee voluntarily chooses to resign from the organization.

a. Avoidable VTO: Avoidable voluntary turnover refers to turnover that is regretted by the organization and is initiated by the employee. E.g., attrition due to job dissatisfaction, better job offer, etc.

b. Unavoidable VTO: Unavoidable voluntary attrition refers to turnover that is non-regretted by the organization but is initiated by the employee. E.g., for personal reasons, family reasons, health reasons, spousal relocation, higher education. etc.

Involuntary Turnover (ITO): The employer makes the decision to discharge an employee and the employee unwillingly leaves his or her position. E.g. Resignation in force (RIF), poor performance, staff conflict, Leave of Absence without returning to work, organizational restructuring, death, etc.

DATA PREPARATION & CLEANING

Numerous techniques can be applied to deal with issues having to do with missing data. For example, if the job title of an employee is missing at the time of hire, then the salary of the employee is compared to the mean salary of each job title, to find the job title that corresponds most closely to the salary of the employee. Similarly, a missing salary can be estimated in several statistical ways by comparing it with the similar employees with the same job title and other known variables. The level of sophistication applied to estimation/replacement of missing values will depend on several factors, including the frequency of missing values and the anticipated importance of the variables containing the majority of the missing data. Data cleaning will also include the consideration of outliers, invalid data points, and the removal of individuals for whom there is such extensive missing data that his/her full exclusion from the database is warranted.

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

Assuming a sufficiently large database, two different samples of data will be collected. One sample will be utilized for running the model and creating the attrition risk equation, while the other sample will be a "holdout sample," and utilized for validating the model. The holdout sample, as the name implies, will be excluded from the initial analysis and generation of the data-analyzed equation results. And, as a final note on data-cleaning, a fixed, specific cut-off date for the data analysis will be set for each of the years under consideration.

RESEARCH QUESTIONS & HYPOTHESIS TESTING

The following research questions and hypotheses are to be tested using the data to be collected for analysis of this predictive workforce-analytics model:
The core research question is:
Which variables affect VTO of the high performing employees in the organization, and to what extent?
Specific hypotheses apply to each variable in the model. We postulate:
H1 – There is significant relationship between VTO and several employee-attribute variables;
H2 – There is significant relationship between VTO and several organizational-variables;
H3 – There is significant relationship between VTO and the two external-influences variables.

DATA ANALYSIS

We are not yet certain of the exact details of the analysis. However, we believe that our initial steps will be as follows:
1. A descriptive analysis will be performed to provide insights into any missing data, and also to provide the cell size information of the subgroups in the data set.
2. A cross-tabulation will be performed between the gender and age of the employees to get a better understanding of the employee base.
3. A cross-tabulation will be performed between the age and job tenure in order to get an overview of the proportion of experienced versus inexperienced counterparts.
4. Similar cross-tabulations will be performed with other pairs of notable independent variables to get a better understanding of the overall data set
5. A series of bivariate-correlation analyses will be conducted between VTO and all of the independent variables.

More advanced steps that are likely to occur are
6. In order to reduce the data set to a smaller set of summary variables and to explore the underlining theoretical structure of the phenomena, we will use exploratory factor- analysis.
7. Multiple regression analysis will be performed to identify the coefficients of the Master Risk Equation. Stepwise regression will also be performed to more clearly identify key independent variables and ensure that no significant variables are missed due to multi-collinearity.
8. The Master Risk Equation obtained from the above steps will be tested using the holdout sample, to check its efficacy.
9. The ultimate outcome from the analysis will be to identify high-risk clusters, initiate focus-group discussions with appropriate human-resources department members, in order to perform a "deep dive diagnostic" of VTO, and hence, create an action plan to minimize the VTO.

ISSN 2053-5686(Print), ISSN 2053-5694(Online)

REFERENCES

1. https://barbradozier.wordpress.com/2015/09/22/impact-of-staff-turnover-on-organisational-perfomance-a-case-study-of-beitbridge-town-council/, downloaded in July, 2016.
2. http://www.hcminst.com/files/CareerPathRatioExample.pdf, downloaded in July, 2016.
3. http://www.statisticssolutions.com/factor-analysis-sem-exploratory-factor-analysis/, downloaded in July, 2016.
4. Sanky, Michael, Paul D. Berger, & Bruce D. Weinberg (2012) "A Segmentation Approach To Patient Health Intervention," Journal of Medical Marketing, 12(4), p. 221-228.
5. Academy of Management Journal, 35, 1036-1056. Sicherman, N. (1996). Gender differences in departures from a large firm
6. Cotton, J. L., & Tuttle, J. M. (1986). Employee turnover: A meta-analysis and review with implications for research. The Academy of Management Review, 11(1), 55-70.
7. Sorensen, J. B. (2004). The organizational demographics of racial employee segregation. The American Journal of Sociology, 110(3), 626-671.
8. Hom, P. W., & Griffeth, R. W. (1991). Structural equations modeling test of a turnover theory: Cross-sectional and longitudinal analyses. Journal of Applied Psychology, 76(3), 350-366.

ISSN 2053-5686(Print), ISSN 2053-5694(Online)