HUMAN RESOURCES DATASET OF COMPANY X
EXPLORATORY DATA ANALYSIS REPORT

Serra Uzun
Northwestern University

Date: August 27th, 2020

1. INTRODUCTION

Employee turnover has long been one of the hidden yet impactful cost for businesses of any size. The motivational, financial, and cultural impact of the negative changes that come with high employee turnover can be minimized or even eliminated by understanding the source of the problem in the first place. The Human Resources data collected over a specific time on all employees in all departments and at all levels gives companies the ability to analyze the implications of various and numerous parameters on employee satisfaction and departure. Through what is called People Analytics, businesses nowadays have a better understanding of conditions, events, and circumstances that keep their employees happy and satisfied working at their firm and what causes them to look for other opportunities.

The main reasons for selecting the HR dataset for Exploratory Data Analysis are:

- A high employee turnover rate is a globally experienced issue that all companies may face and are not industry-focused. The analytics of this type of HR dataset can prove to be versatilely beneficial for any company anywhere.
- It brings attention to the importance of collecting and storing HR related data for long term success.
- It helps create more sustainable and long-term teams where happy and full-filled employees can thrive and contribute more and eagerly towards companies' long-lasting success.
- It puts the spotlight on a hidden and often not prioritized burden for any company that has a high potential to provide a decrease in financial loss and increase motivation, efficiency, and productivity.

The Human Resources dataset used in this EDA was obtained through Kaggle, an online platform for data scientists with one of the largest open dataset libraries. The company name that the HR dataset comes from was kept anonymous, and the format of the data file was comma-separated values (.csv).

The hypothesis set out to test was that the number of work accidents employee experiences, their salary range, and in-company recognition they received (promotion) have a significant and collective impact on their likelihood of departure. The goal of this analysis did not only give the company leadership and HR department a valuable insight on the potential causes for employee dissatisfaction and departure, where they can invest time and financial resources to minimize turnover rate, but also to help set more thorough and lasting standards and paths of advancement for current junior and prospective employees.

2. DATA PREPARATION AND ANALYSIS

The HR dataset used for the analysis originally came in the size of 14,999 rows (observations) by ten columns (variables). The information in the dataset was of the variables listed below, together with their explanation and type:

	VARIABLE	DESCRIPTION	TYPE
1	satisfaction_level	Satisfaction Level	decimal
2	last_evaluation	Last Evaluation Score	decimal
3	number_project	Number of projects the employee has worked on	integer
4	average_monthly_hours	Average number of hours that employee works (monthly)	integer
5	time_spend_company	the time employee has spent in the company (in years)	integer
6	work_accident	If the employee had a work accident	binary
7	left	If the employee had departed the company	binary
8	promotion_last_5years	If the employee received a promotion in the past 5 years	binary
9	department	The department employee is a part of	string
10	salary	The salary range group the employee is a part of.	string

Table 1

The dataset's cleaning process involved identifying and removing duplicated entries, missing values, and removing outliers. Fortunately, there were no variables with missing values in them and no outliers under any of the variables listed above. 3,008 duplicated rows were removed, which made the dataset for analysis to be 11,991 rows in total.

Out of 11,991 employee records, 10,000 are current employees, whereas 1,991 were of past employees who have departed the company. The dataset consists of information in Sales, Technical, Support, Information Technology (IT), Research & Development (RandD), Product Management (product_mng), Marketing, Accounting, Human Resources (HR), and Management departments. The department with most employees is identified to be Sales, with 2,689 current employees, which account for 22.5% of the company followed by Technical with 1,854 and Support with 1,509 employees. On average, employees work on 3.8 projects during their time at the company, work 201 hours monthly, have a satisfaction level of 0.61 (61%), and have been with the company for 3.5 years. If we look at the minimum and maximum values for some of the variables, we see that the minimum time spends with the company is two years, the monthly hours worked is 96, and the number of projects worked on is two. On the other hand, the maximum time spent by an employee at the company is ten years, the maximum monthly hours worked is 310, and the maximum number of projects completed by an employee is seven.

When we looked into the number of current employees in each salary group (low/medium/high), we saw that only 9% of the current employees are in the high salary group and 91% of the employees are either in low or medium salary group. This finding was the initial indicator of any potential dissatisfaction, yet it did not give us a definite and concrete outcome as the dataset does not provide the information on the lower and upper limits of these salary groups are.

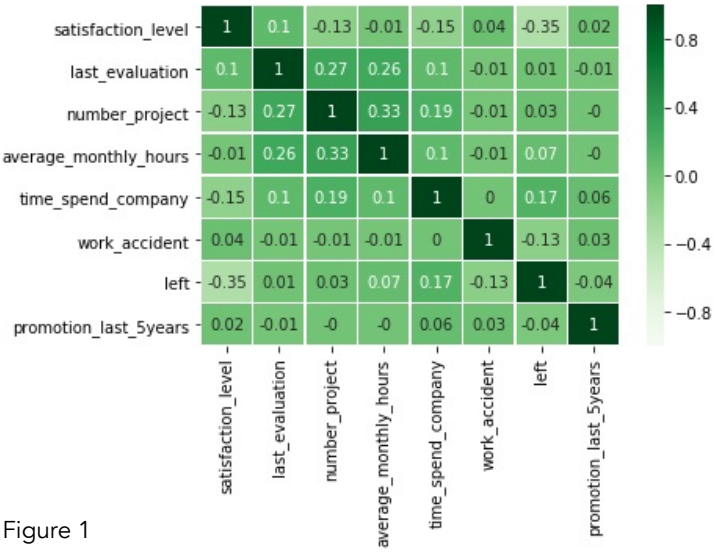

Figure 1

The correlation analysis we performed on the dataset with all departments and employees included, we have seen that the highest correlation is between departure rate and satisfaction level, which are negatively correlated (-0.35), meaning that the higher employees' satisfaction level, the less likely they are to leave the company. (see Figure1 below) Some other correlations we observed yet are not significant indicators are the positive correlation between the number of projects an employee has on record and the average monthly hours they worked (0.33), and the positive correlation between the number of projects the employee worked on and their last evaluation score (0.27). These were not only expected correlations, but they were also not the optimal variables to indicate an employee departure.

A As we realized the negative correlation between satisfaction level and departure rate, we then looked at these variables in the departmental context. The departmental approach to looking at satisfaction levels and departure rates showed us that Technical, Human Resources (HR), and Accounting have the lowest satisfaction levels and highest departure rates. For the remainder of the study, we focused our analysis on determining if the work accidents, salary range, and promotions impact the satisfaction and departure rate of the employees in these given departments listed above.

department	satisfaction_level	left
support	0.634822	0.171334
marketing	0.63477	0.166419
IT	0.634016	0.161885
management	0.631995	0.119266
sales	0.631349	0.169805
product_mng	0.629825	0.16035
technical	0.627937	0.173797
RandD	0.627176	0.122478
hr	0.621947	0.18802
accounting	0.607939	0.175523

Table 2

-The total number of employees in the Technical, HR, and Accounting departments is 3,466, which accounts for 29% of the total number of past and present employees within all departments. Within the 3,466 employees, 65% (2,244) is part of the Technical department, 18% is a part of Accounting, and 17% are in the HR department.

department	satisfaction_level	work_accident	promotion_last_5years	left
accounting	0.607939	0.140097	0.016103	0.175523
hr	0.621947	0.133111	0.018303	0.188020
technical	0.627937	0.151070	0.009804	0.173797

Table 3

From the above table, we can determine that Accounting department employees have the lowest satisfaction level (0.61 – 61%). While the average work accident, promotions, and departure rate of the Accounting employees are not higher or lower than the rest, the interesting insight we obtain from the above table is regarding the Technical department. The Technical department has the highest satisfaction level out of these three departments, yet has the highest number of average work accidents, the lowest rate of promotions in the last five years yet, interestingly, the lowest departure rate. While mean values for the above variables do not present a positive picture for the Technical department, we see that they are the least likely to leave or have left than Accounting and HR. (see Table 3)

work_accident	left	
0	0	80.3041
	1	19.6959
1	0	94.2688
	1	5.73123

Table 4

promotion_last_5years	left	
0	0	82.1502
	1	17.8498
1	0	97.6744
	1	2.32558

Table 5

salary	left	
high	0	94.0075
	1	5.99251
low	0	79.2486
	1	20.7514
medium	0	83.8666
	1	16.1334

Table 6

As we further investigated the departure of employees in Accounting, HR, and Technical departments in the context of a work accident, promotion, and salary, we obtained insightful results, most significantly related to work accidents' impact on employee departure. As seen in Table 4, from all employees in the departments listed, the occurrence of a work accident 94% of the employees stayed with the company despite the incident, while this percentage is only 80% for employees who have never had a

work accident. This result is interesting and parallel with the negative correlation we observed between work accidents and departure rates in our overall correlation matrix (Figure 1). Our analysis also shows that employees who received a promotion in the last five years have stayed with the company in the rate that is 15% more than the employees who haven't received a promotion. (see Table 5) Finally, as expected, our analysis indicates that employees in the 'high' salary group have stayed and are thus more likely to remain with the company than employees in the 'medium' or 'low' salary group. (see Table 6)

Lastly, the correlation matrix that we plotted only with the data of departments Technical, Accounting, and HR shows similar correlations with the overall correlation matrix presented in Figure 1. While the most significant correlation (negative) remains between satisfaction level and departure, the matrix re-iterates our findings above.

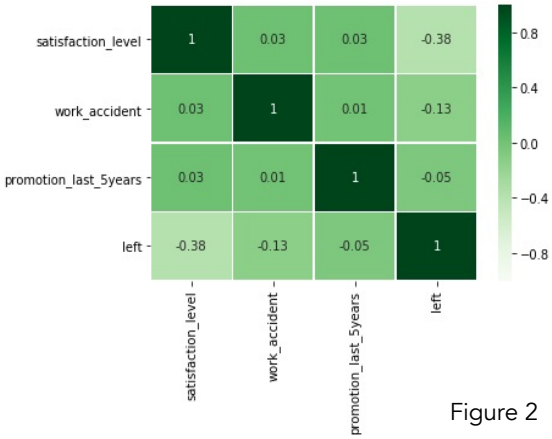

Figure 2

3. CONCLUSION

Even though there are no significant negative or positive correlations between other variables, we have observed from our exploratory analysis above that employees who have been promoted in the past 5 years of their experience in the company, and employees who are in the high/medium salary group are less likely to leave than the ones who haven't been promoted and are in the low salary group. As departure is negatively correlated with satisfaction level, lowering the departure rate is highly likely to positively impact the satisfaction level amongst Accounting, HR, and Technical department employees. Our recommended next steps are:

1. Re-evaluate the in-place salary and promotion thresholds and procedures in order to better respond to employee recognition and salary expectations
2. Put in place a more transparent and straightforward path for promotion, set by the time recruits join the company to avoid any new turnover.
3. Regularly communicate and discuss employee performance expectations and provide guidance for advancement within the firm.
4. Look further into the 'Accounting' department as their employees are neither the least promoted or least paid group yet have the lowest satisfaction level in the whole organization and have one of the highest departure rates.
5. Collect more detailed data on employees for more accurate analyses in the future.

Source of dataset: https://www.kaggle.com/giripujar/hr-analytics