**Employee attrition in healthcare**

**Background**

Employee attrition is the gradual but intentional loss of staff members within an organization.The unprecedented covid-19 pandemic had been a burden to various sector of the economy, but most especially to the healthcare industry. The covid-19 pandemic has led to an increase in healthcare market around the world such as creation of new health facilities, treatment of new diseases, etc. This has left healthcare workers with more options as to where they want to work and consequently increased the attrition rate in the health care industry. However, losing highly skilled health care workers may incur substantial costs connected with the recruitment process and also directly or indirectly affect the patients who use the service of a healthcare provider.

**Motivation**

The increase in attrition rate experienced in healthcare sector due to the advent of the covid-19 pandemic have forced decision makers, like managers, director, human resource persons in the healthcare industry to think about better ways to increase retention among healthcare workers. The adverse effect of this turnover can be seen in the decrease in the quality of healthcare services, loss of qualified personnel, loss of customers, decrease in profit, etc. Hence, any healthcare organization that wants to thrive in this new normal should be able to understand why healthcare workers leave and find ways to increase retention among healthcare workers so as to increase profit and productivity. This analysis through the following research questions, aims to understand the dynamics of healthcare attrition and provide recommendation for decision makers in the healthcare industry:

**Research Questions**

1. **What is the proportion of health care employees in attrition?**

Understanding the proportion of health care employees in attrition will help understand the gravity of the problem.

1. **Are there any differentials in the attrition rate of healthcare employees?**

Considering demographic factors such as age, sex and characteristics of the healthcare workers such as the department, education, etc, will help in understanding factors associated with employee attrition in health care.

1. **Does employee associated variables compared to organizational associated variables contribute more to healthcare attrition?**

Many factors play an important role in the employee attrition rate of healthcare workers, and these can stem from both the employer and the employees. Few research pointed out attrition as the correlation between job dissatisfaction and decreased organizational commitment. This analysis also aims to find out if the employer or employees are contributing more to employee attrition. With this, organization can know where to focus on their evaluation process.

1. **What variables can be used to accurately predict employee attrition among health care workers?**

Aside using guess work or intuition, it’s important for decision makers to use machine learning to accurately predict employees that are more likely to leave the company. Knowing this can save the organization some cost, by ensuring they hire right and also ensure that the organization do the right thing to increase retention.

**Description of the Data**

The data is synthetic and based on the IBM Watson dataset for attrition and downloaded from Kaggle. Employee roles and departments were changed to reflect the healthcare domain.

At first glance, the data set contains 1676 observation and 35 variables. The target variable is attrition- whether an employee left or not. There are 26 numerical variables and 9 factor variables (outcome variable inclusive). Also, the data had proper labelling, was clean and did not contain any missing value. The data contains employee associated variables (such as age, sex, education, etc.) and organizational associated variables (such as department, job satisfaction, monthly income, etc.). Some variables like employee ID, employee count, and Over18 will not be used in this analysis because of their distribution. The employee ID is the unique employee identification, also respondent were above 18 years and the employee count for all healthcare employee in the data is 1. All other variables will be explored for this analysis.

**Proposed Analysis**

The analysis will comprise of descriptive and inferential analysis. An exploratory analysis of the variables will be conducted to see the distribution of the variables. Also, cross-tabulation, chi-square tests, and bar charts of some of the predictors and the target variables will be done. Histogram will be done for examine the skewness of the numerical variables. The models that would be explored for this analysis will be logistic regression (Generalized Linear Model), support vector machine, K nearest model, decision trees, random forest.

The logistic regression is a strong foundation for testing more sophisticated models. The logistic model presupposes that the underlying data have the following characteristics: 1) a binary outcome variable, 2) no multicollinearity between variables, 3) no extreme outliers in continuous predictors, 4) independent observations, and 5) a sufficient sample size.

The support vector machine is suitable for predicting complex variables and assumes a large data margin. However, the support vector machine fails to perform optimally when the data set has more noise and when there is huge data. These assumptions were explored and addressed in the Data section of this report. The decision tree is bias towards multi-level categorical predictors and overfitting in noisy data sets.

In conclusion, to determine if attrition is correlated with employee associated variables or organizational associated variables, principal component analysis (PCA) will be carried out to group all employee associated variables and all organizational associated variables and see which is strongly correlated with employee’s attrition.

**References:**

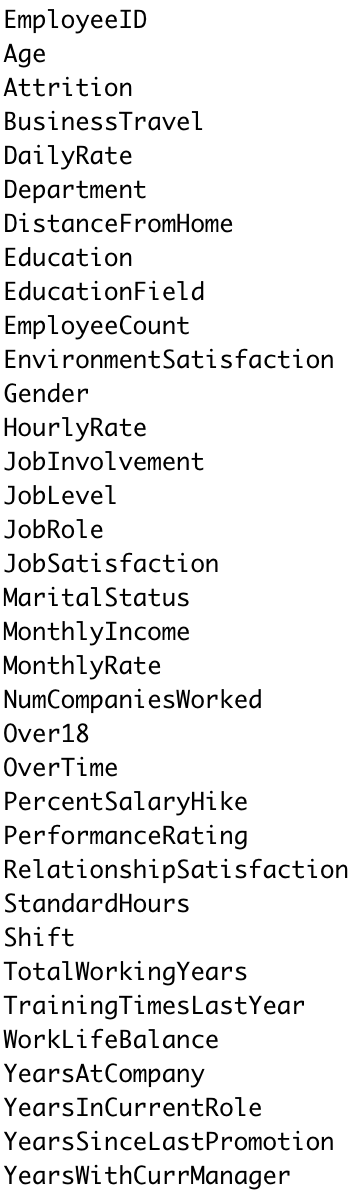
Employee attrition in healthcare <https://www.kaggle.com/datasets/jpmiller/employee-attrition-for-healthcare>

Dr. I. Anand Pawar\*1, V. Chakravarthy2 Factors influencing Employee Turnover in Fusion Healthcare Organization

**Appendix – Variable Descriptions**

The raw data: <https://www.kaggle.com/datasets/jpmiller/employee-attrition-for-healthcare>

The data came as one CSV file with 1676 observations with 35 variables. The variables are described as follows:



# 3.1 Employee Demographics data

As company owner or HR Department, it is important to know the employees demography to know their characteristics and become valuable to analyze further employees development. In this part i break the analysis into three segments: Overall Demography, Experiences, and Company Survey Result .

In this section i try to visualize the overall employees demography data like Gender,Education Background, Age Departments, Job Level, and Job Role. I also display the number of attrition rate per categories to find out whether

Overall Demography,

Experiences,

and Company Survey Result .

# Insights :

There are no significant discrepancies between male and female employee composition. 61 % of the employees are male and the rest is female.

From three departments, the most employees are in R&D (65.37%) and Sales Department (30..34%) while Human Resource Department has the least employee.

Most of our employees were graduated from university.

Most of the employees are work as Sales Executive, Research Scientist, and also Lab Technician. However those job role has the most highest level of attrition compare to any Job Role

Most of the employees are in productive age. From the employee’s age distribution, 60% of employee age are between 20-40 years.

Most of the employees are on Job Level 1 and 2. When we see the attrition rate, the highest rate among all Job Levels are from Job Level 1 (26.3 %).

# Worker Experiences variables

In this section i try to visualize several employees experience feature that are available in the dataset, such as total working experience, total company work, duration work in this company, and number of completed training.

#Insights :

The employees mostly have more than five years working experience and mostly have been work 0-9 years in this company.

Around 48 % of the employees probably fresh-graduate and start their career by working in this company, 44 % of them have been work for 2-5 different companies, and about 16 % have been work for more than 5 different companies

Around 70% (1038 persons) of the employees had completed 2-3 training. Interestingly, the highest attrition rate among the number of completed training groups are also from those group (around 59 % on group of employees that had completed 2-3 training).

# Insights :

The proportion of each score for Employee Satisfaction and Job Satisfaction are quite similar. From Employee Satisfaction survey, 30.84 % were very satisfied (score = 4), 30.82 % were satisfied (score = 3), 19.52 % were partly-satisfied (score = 2), and 19.32 % are not satisfied (score = 1). While in in Job Satisfaction survey, 31.22 % were very satisfied (score = 4), 30.07 % were satisfied (score = 3), 19.05 % were partly-satisfied (score = 2), and 19.66 % are not satisfied (score = 1).

From the Job Involvement survey, around 59 % of them had high involvement in their job (score = 3). Among that group, the attrition rate is around 8 %.

Looks like the employees can manage their work-life balance. Based on Work-Life balance survey, around 60 % of them feels they had better (score = 3) work-life balance and only 5 % of them feels that they had bad (score = 1) work-life balance.

# 3.2 Employee Attrition against several factor

In this section, i would like to analyze several factor that probably leads to employee attrition such as income, home to workplace distance and overtime working, . I also breakdown the attrition rate per several variable like departments and company survey.

# 3.2.1 Overall Attrition Rate

Before we deep dive into several factor, let’s see the overall attrition in this company

# As shown in above diagram, the attrition rate is about 16 % . Several companies has different metrics or threshold to decide whether this rate is good, moderate, or bad.

# From the Departments, the highest employees (consecutively) in this follwing department: Research & Developments, Sales, and HRD. However the highest attrition rate is from Sales department (20.63 %), followed by Research & Developments department (13.84 %) and HRD department (19.05 %).

# I try to analyze the income per each job satisfaction score. In the plot above shows us that for the same level of Job Satisfaction Score there is a big gap of (median) income between those who leave the company and those who stay.

It also seems like the gap of income between the attrition status group tend to increase when Job Score Satisfaction Score is lower. Probably the gap of income become the reason for employee to leave.

# Insights :

Most of the employees get salary hike between 11 - 14 %.

Employees that have salary hike between 11 to 14 percent have higher chance to leave.

# Insights:

As seen on the plot above, for almost every job role those who leave the company has lower Environmental Satisfaction Score compare to those who stay. There is an exception for Research Director, probably Environmental Satisfaction is not their main reason to leave the company

On average, Managers and Healthcare Representatives that leave has lowest score and big gap compare to those who stay. Probably the environment job become the main issue for those role.

On average, the gap of Environmental Satisfaction Score between who leave and those who stay as Research Scientist and Sales Representativeis small. Probably the Environment for those role is good and not the main reason to leave.

# 3.2.5 Gender

I was wondering whether specific marital status increase the chance to leave the company. I assume that the married woman or the employee that has divorce tend to leave the company.

After I agggregate the data between the attrition status and the marital status the assumption is not true.

#Insights:

As seen on the plot above, the attrition rate for divorced employee is still lower than those who married or single.

Also the married woman has lower attrition rate than those who married. If compared to married male, the attrition rate for woman is still lower than married man.

#### NUMBER 222

# Chi-Square Test for Feature Selection

To decide which categorical variables should be kept in the attrition model, Chi-Square will be used to test whether there is a relationship between the categorical variables and attrition. The null hypothesis for this test is the two variables are independent, and the alternative hypothesis is the variables are not independent. In order to reject the null hypothesis and keep variables in the model, the p-value of this test must have a p-value below .05.

The variables we will leave out of the model are education, gender, performance rating, and relationship satisfaction. These variables all have a p-value above .05 so they are independent from attrition.