# Employee Attrition: Analysis and Prediction

# Abstract

## Key Words: Attrition, Turnover, Employees, Prediction, Machine Learning

# Introduction

|  |
| --- |
| A common popular perception is that all businesses have employees, however according to the U.S. Government around 21.5% of businesses in the United States employ workers (U.S. Small Business Administration, 2020). While some businesses may not require employees, those that do hope to find high performing employees and retain them. The relationship between employer and employee relies on a balance between multiple factors including work-life balance, employee satisfaction, salary, and other more nuanced factors. When employees exit the relationship with the business en masse, an employer is faced with the potential problem of unwanted “attrition.” Analysis of variables that describe employee factors is performed using a dataset published by IBM specifically to analyze and track employee attrition. Prediction of employee attrition is performed using the dataset. |
|  |

|  |
| --- |
|  |
|  |

# Literature review

# Methodology

We obtained the data set from Kaggle.com using this link:

[https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset]. The fictional data set was originally created by IBM data scientists to uncover the facts that lead to employee attrition and explore important question like what the important factors are that influence attrition among employees. Also, the original dataset can be accessed from the Link- https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/. Then it was saved in our group github repository as a .csv file for the convenience of the analysis purposes. The Attrition dataset had 1470 observations with 35 variables. Out of those 35 there exists the target variable Attrition with possible outcomes "Yes" and "No". With our experiment results We will do the analysis based on Gender, Education, Income, Working Environment, and lastly, build a predictive model to determine whether an employee is going to quit or not.

# Experimentation and Results

In preparation of data, first we checked if there are any missing values in the dataset. As we realized that there were no missing data or duplicate data, then we moved forward by changing the categorical data that was represented as integers back to categorical. Next step was to check each variable with Attrition.

Chart, bar chart, waterfall chart, treemap chart

Description automatically generated

1. Age: We see that majority of employees leaving the org are around 30 Years (year 28-36). Average age is between 30 to 40.

2. Business Travel: Among people who leave, most travel frequently or rarely.

3. Department: Among people attritted employees from HR dept. are less. It is because of low proportion of HR in the organization(Fig 1).

4. Distance From Home: Contrary to normal assumptions, a majority of employees who have left the organization are near to the Office.

5. Daily Rate: We are not able to see any distinguishable feature here(Fig 1).

Chart, bar chart

Description automatically generated

6. Education: From the data we know that, 1-‘Below College’, 2-‘College’, 3-‘Bachelor’, 4-‘Master’ 5 ‘Doctor’. Looking at the plot we see that very few Doctors attrite. May be because of less number. Based on the data most of the employees have Bachelor’s degree level education.

7. Education Field: On lines of the trend in Departments, a minority of HR educated employees leave and it is majorly because of low proportion of the HR in the organization.

8. Employee Count: It is an insignificant variable for us.

9. Employee Number: It is also an insignificant variable for us.

10. Environment Satisfaction: Ratings stand for: 1-‘Low’, 2-‘Medium’, 3-‘High’, 4-‘Very High’. We don’t see any distinguishable feature (Fig 2).

Chart, bar chart

Description automatically generated

11. Gender: Majority of separated employees are Male and the reason might be because around 61% of employees in the dataset are Male.

12. HourlyRate : There seems to be no straightforward relation with the Daily Rate of the employees.

13. Job Involvement: Ratings stand for 1-‘Low’, 2-‘Medium’, 3-‘High’, 4-‘Very High’. Majority of employees who leave are either very highly involved or least involved in their Jobs.

14. JobLevel: Job Level increases the number of people quitting decreases.

15. Job Satisfaction: As per data 1-‘Low’, 2-‘Medium’, 3-‘High’, 4-‘Very High’. We see higher attrition levels among lower Job Satisfaction levels.

Chart, waterfall chart

Description automatically generated

16. Marital Status: Attrition is on higher side for Single and lowest for Divorced employees. Most employees are married.

17. Monthly Income: We see higher levels of attrition among the lower segment of monthly income. If looked at in isolation, might be due to dissatisfaction of income. Higher number of employees earn less.

18. Monthly Rate: We don’t see any inferable trend from this. Also no straightforward relation with Monthly Income.

19. Number of Companies Worked: We see a clear indication that many people who have worked only in One company before quitting a lot.

Chart, bar chart

Description automatically generated

20. Over Time: Larger Proportion of Overtime Employees are quitting.

21. Percent Salary Hike: We see that people with less than 15% hike have more chances to leave.

22. Performance Rating: 1-‘Low’, 2-‘Good’, 3-‘Excellent’, 4-‘Outstanding’. We see that we have employees of only 3 and 4 ratings. Lesser proportion of 4 raters quit.

23. Relationship Satisfaction: 1-‘Low’, 2-‘Medium’, 3-‘High’, 4-‘Very High’. Higher number of people with 3 or more rating are quitiing. There are considerable amount of low and medium relationship satisfaction in this organization.

Chart, bar chart, waterfall chart

Description automatically generated

24. Stock Option Level: Larger proportions of levels 1 & 2 tend to quit more.

25. Total Working Years: We see larger proportions of people with 1 year of experiences quitting the organization also in bracket of 1-10 Years. Higher the number of experiences you have, you tend to stay in the job.

26. Training Times Last Year: This indicates the no of training interventions the employee has attended. People who have been trained 2-4 times is an area of concern.

27. Work Life Balance: Ratings as per Metadata is 1 ‘Bad’ 2 ‘Good’ 3 ‘Better’ 4 ‘Best’. As expected, larger proportion of 1 rating quit, but absolute number wise 3 is on higher side.

Chart, bar chart

Description automatically generated

28. Years at Company: Larger proportion of newcomers are quitting the organization, which sidelines the recruitment efforts of the organization.

29. Years In Current Role: Plot shows a larger proportion with just 0 years quitting. May be a role change is a trigger for Quitting.

30. Years Since Last Promotion: Larger proportion of people who have been promoted recently have quit the organization.

31. Years With Current Manager: As expected a new Manager is a big cause for quitting.

# Discussion and Conclusions

# References

Fallucchi, F. C. (2020). Predicting Employee Attrition Using Machine Learning Techniques. *Computers*.

Raman, R. S. (2019). Predict employee attrition by using predictive analytics. *Benchmarking: An International Journal*.

Saradhi, V. V. (2011). Employee churn prediction. *Expert Systems with Applications 38.3* , 1999-2006.

Sisodia, D. S. (2017). Evaluation of machine learning models for employee churn prediction. *International Conference on Inventive Computing and Informatics (ICICI)*.

U.S. Small Business Administration, O. o. (2020). *Frequently Asked Questions About Small Business, 2020*. Retrieved from https://advocacy.sba.gov/2020/10/22/frequently-asked-questions-about-small-business-2020/

Yadav, S. A. (2018). Early Prediction of Employee Attrition using Data Mining Techniques. *IEEE 8th International Advance Computing Conference (IACC)*.

Yedida, R. R. (2018). Employee Attrition Prediction. *arXiv preprint* , 1806.10480.

# Appendices