Employees Attrition Prediction Analysis

By: Monika Mangla

Contents :

Overview 3

Introduction.....3

Dataset Overview.....4

Exploratory Data Analysis and Discovery 6

Tableau Dashboard……10

Overview:

The goal of this project is to analyse about various factors that leads to attrition. Employee attrition is defined as employees leaving their organizations for unpredictable or uncontrollable reasons. Employee attrition analytics allows organizations to identify potential risk areas or areas for improvement, such as high turnover rates in specific departments or job roles. By addressing these issues, businesses can improve employee morale, lower turnover costs, and ultimately build a more successful and sustainable business. We will examine the most important factors that influence attrition within an organization. We will consider if these factors are within the control of the organization and what actions can used to mitigate or combat attrition. We will also analyse current trends in HR and how these apply to our analysis, where finally, based on our results, we will conclude with insights and recommendations.

Introduction:

Employee attrition is gradual yet deliberate reduction of employees in an organization. This decline in number of employees over a period is defined as Rate of Attrition. Employees leave company due to non-fulfilment of the expectation from the organization in return of the service towards their job. Employees leave the organization for various reasons. The cause could be higher salary in other organization, family mobility, technology preference, higher position etc. This dissatisfaction prevails at individual level and no organization can have control over it. However, attrition leads to losses and extra expense to the organizations. Organizations spend lot of time, resources and efforts in training and developing the employees to increase the efficiency level of their work. If employee leaves the company, he needs to be replaced with another one. This process demands same investment of time, efforts, and resources in recruiting a new one and train him. Most of IT organizations today are being hit badly with high rate of attrition resulting in various productivity and quality related issues. Hence it has become critical to hire a right candidate for a job position.

High rate of employee attrition directly indicates frequent changes in the team and resources. These changes are not welcome by the service seekers of the organization. Frequent changes in their sources leads to decline in quality of the service deliverables. Hence not only cost but quality and productivity is hampered. That is why it is important for organizations to know the attrition reasons and the techniques that could avoid or reduce the rate of attrition.

In this report we will be discussing the validity of these reasons through a general analysis of the available employee dataset. We will analyse the dataset through a generally understood set of procedures. These procedures are as follows: research, exploratory data analysis, data manipulation and cleaning, descriptive and predictive analysis, data visualization using Tableau Dashboard. This process can be approached procedurally or iteratively.

Dataset Overview:

Acme Corporation has provided historical data on employee demographics, job satisfaction, work environment, performance metrics, and turnover status. This dataset spans the last five years and includes information on employees who have left the company and those who are still currently employed.

The dataset typically includes several features that provide insights into employee characteristics, job satisfaction, and performance., here's a general list of common features in a dataset:

- **Employee ID:** A unique identifier for each employee.
- **Age:** The age of the employee generally between 18-60 years.
- **Attrition:** A binary variable indicating whether the employee has left the company (1) or is still employed (0).
- **Business Travel:** The frequency and nature of business-related travel (e.g., "Travel_Rarely","Travel_Frequently","Non-Travel").
- **Department:** The department to which the employee belongs (e.g., "Sales," "Research & Development," "Human Resources").
- **Distance From Home:** The distance of the employee's residence from the workplace categorizes as ‘Near’,’Far’,’Very far’.
- **Education:** The employee's level of education (e.g., "1: 'Below College'," "2: 'College'," "3: 'Bachelor'," "4: 'Master'," "5: 'Doctor').
- **Education Field:** The field in which the employee's education lies (e.g., "Life Sciences," "Medical," "Marketing").
- **Environment Satisfaction:** The level of satisfaction with the work environment on a scale like ‘yes’ or ‘no’.
- **Gender:** The gender of the employee either ‘Male’ or ‘Female’.
- **Job Involvement:** The degree to which the employee is involved in their job (e.g.-low, medium, high).
- **Job Level:** The level or rank of the employee's position on a scale of 1-5.

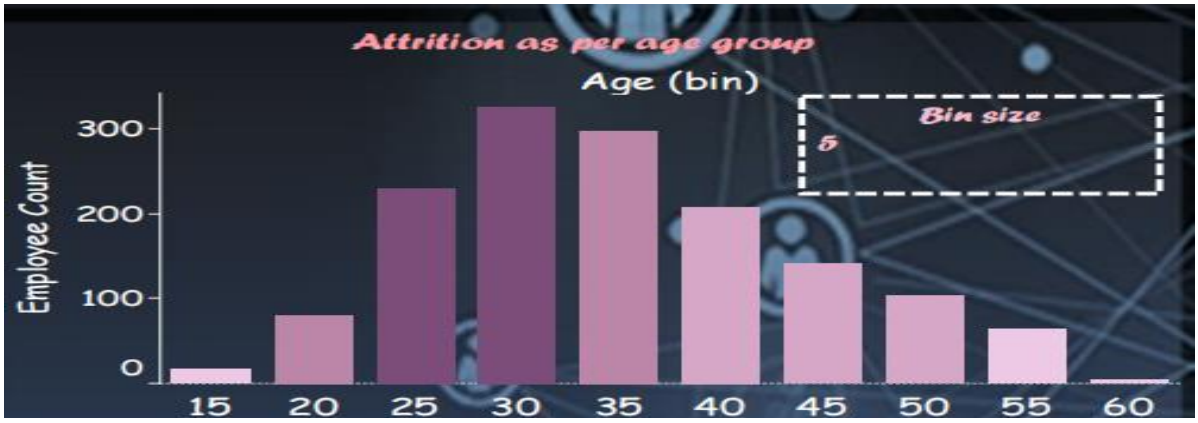
- **Job Role:** The specific role or title of the employee's job (e.g.- sales rep, lab technician, HR, sales exec, research scientist, manufacturing director, healthcare rep, manager, research director).
- **Job Satisfaction:** The level of satisfaction with the job on a scale (e.g.- low, med, high).
- **Marital Status:** The marital status of the employee. (like married, divorced, unmarried).
- **Monthly Income:** The monthly salary of the employee.
- **Num Companies Worked:** The number of companies the employee has worked for basically from 0 to 9.
- **Over Time:** Whether the employee works overtime or not.
- **Performance Rating:** The performance rating of the employee on a scale of 1-4.
- **Relationship Satisfaction:** The level of satisfaction with relationships at the workplace. (e.g.- low, med, high, very high)
- **Stock Option Level:** The level of stock options provided to the employee. (e.g.- 0-3)
- **Total Working Years:** The total number of years the employee has been working generally having 0-40 years of experience.
- **Training Times Last Year:** The number of training sessions the employee attended last year. (like 0-6 times).
- **Work-Life Balance:** The balance between work and personal life. (such as bad, ok, good, best)
- **Years At Company:** The number of years the employee has been with the current company (like 0-40 years).
- **Years In Current Role:** The number of years the employee has been in their current role probably 0-18 years.
- **Years Since Last Promotion:** The number of years since the last time the employee was promoted such as 0-15 years.
- **Years With Current Manager:** The number of years the employee has been working under the current manager ranging from 0-17 years.

Within the dataset we have a mix of numeric and categorical datatypes. The initial dataset contains twenty-six numeric variables and it contains nine categorical data types. Examples of numeric data types include employee age, their monthly income, and the years that they've been working with the company. Examples of available categorical variables include education, gender, job role, or the department that the employee works with.

Exploratory Data Analysis and Discovery:

When analysing a dataset, it is important to initially explore the constructs of the data, getting a sense for how the content within each observation varies, the ranges, the most influential variables, the least. The simplest way to approach exploratory data analysis is through visualizations and basic statistical analysis.

In the section below, we will discuss the exploratory data analysis process on each variable, the resulting connection to attrition, and our general consensus towards the importance and use of the respective variable.



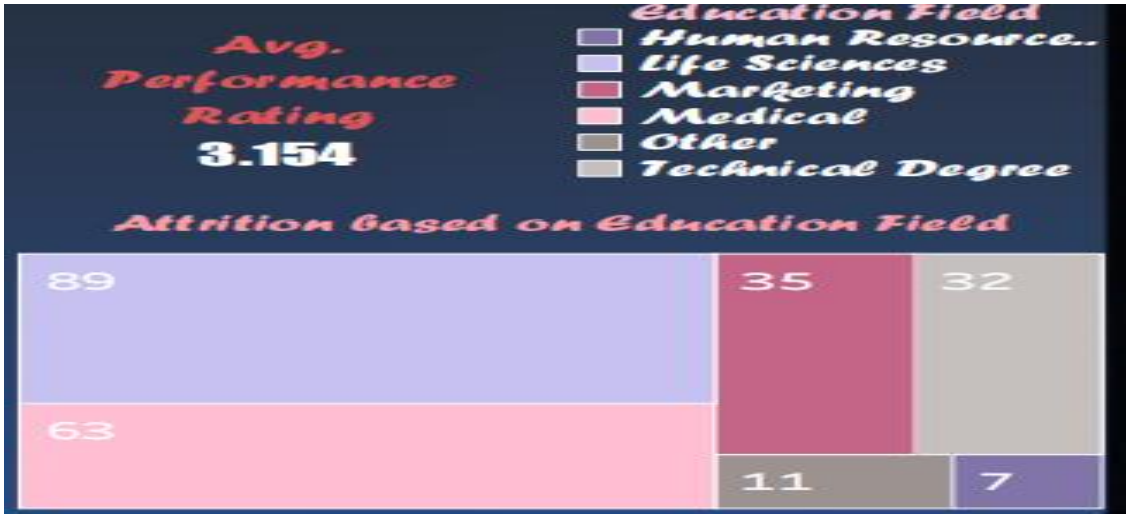
Above we see a cropped section of a graph that illustrates the attrition rate along with an employee's age. We see how between the ages of 18 and 21, there is a major increase in attrition, while as the years continue, and especially looking at age 27, the attrition rate is significantly lower. This same pattern is apparent when looking at the years worked by employees. We notice that the first three years of employment incur the highest attrition rates.

Below we will briefly discuss the variables available to us through the dataset. We can begin by examining demographic data, including age, gender, marital status, academic field of study, and education level. The age range within this dataset is between 18 to 60. This gives a full age range from completely green—employees that have no work experience—to employees who are ready to retire. The marital status includes whether an employee is single, married, or reportedly divorced.

When analysing marital status, we do see a greater proportion of single employees leaving the company versus those that are married or divorced. We see that on average, married and divorced employees have an attrition rate of 40%, where employees that indicated being single, have an attrition rate of 64%. Concerning environment satisfaction, we do see a positive correlation between a reportedly low environmental satisfaction versus a medium, high, and very high. We see that the average attrition rate for medium, high, and very high is 14%. The attrition rate jumps to 25% with a environmental rating of low.

Lastly, we examined work-life balance, since this is presumably an important factor in modern day company culture. This variable carries a rating of between 1 and 4, where 1 is equivalent to bad work-life balance and a 4 is equivalent to the best work-life balance. Looking at the data, we find that the attrition rate when work-life balance is bad, is 31.25%; furthermore, on average, when work-life balance is good, better, or best, the attrition rate is 16.2%. This is approximately a two-fold difference. When reviewing the statistical correlation between work-life balance and attrition, we find that there is again, a very low correlation of 0.064. For this reason, and due to the fact that the dataset does not state what factors lead to the a bad or a good work-life balance, we chose not to utilize this variable in later portions of our analysis.

Education includes five levels, from no college experience to having a masters or doctorate degree. We see that the variance in attrition is not large when comparing employees based on their degree. Having no college experience is corresponds to a 52% attrition rate, while having a masters or doctorate degree roughly corresponds to a 45% attrition rate. We do see a slightly higher increase in age per the degree, where an employee with no college experience has an average age of 32, while employees that have a Masters or Doctorate both have an average age of 39 years. All education levels also have similar average years at the company, where employees with no college experience have an average of 6.5 years at the company, and those with a higher degree have between 7.5 and 8.3 years. This information on education leads us to believe that education level does not have a significant impact on the attrition rate. If there was a greater variance in the years worked or the age group, then we would expect a greater impact, especially if they average down at the lower age group where attrition rates are highest.



Looking at job involvement, we see that there is also a high correlation to attrition rate. Employees who maintain a job involvement rating of 1, or low, have an attrition ratio of 33.5%. This is a stark contrast to those who have a very high job involvement rating, where the attrition rate hovers around 9%. If we average job involvement, combining medium, high, and very high, we see an attrition ratio of 13.6%. This is roughly two and a half times lower than employees with a low job involvement rating. Additionally, after oversampling for this variable, we found that a low job involvement corresponded to an attrition rate of 70%. This is a significant finding.

Job satisfaction is generally understood to be an important factor on employee attrition.

Job satisfaction rating

Job Role	Job Satisfaction 1	2	3	4
Healthcare..	26	19	43	43
Human Re..	10	16	13	13
Laboratory ..	56	48	75	80
Manager	21	21	27	33
Manufactu..	26	32	49	38
Research D..	15	16	27	22
Research Sc..	54	53	90	95
Sales Exec..	69	54	91	112
Sales Repr..	12	21	27	23

After examining for this variable, we see that the range between employees who report a low job satisfaction and those that report a very high job satisfaction, is 22.8% versus 11.3 %. This is approximately double and is notably concerning, although it is not as drastically different as we might have expected. It is also important to note that while this variable is inherently indicative of an employee's sentiment towards the job, it is also very difficult to control for. There can be an array of circumstances that affects an employee's degree of satisfaction where,

furthermore, while certain circumstances may affect one employee drastically, it may affect another employee very little.

For instance, when relationship satisfaction is low and job involvement is low, the attrition rate is 23.5%. When relationship satisfaction is low and job involvement is high, the attrition rate drops to 7%. When relationship satisfaction is very high and job involvement is low, the attrition rate is 43.5%. But when relationship satisfaction is high and job involvement is also high, the attrition rate is 12.5%. While we do notice this relationship, after reviewing the statistical correlation between the two variables, we see that it is only 0.03, which is very low. Additionally, relationship status is another variable that is difficult to control for, since there are many factors that go into determining an employee's degree of satisfaction towards their manager or colleagues.

In the figure below, we included attrition rates as they correspond to several key variables. These include years at the company, overtime, stock option level, and lastly, stock option level in conjunction with overtime. We see in the figure that when overtime is clocked in by employees, that the attrition rate is nearly 75%, while the attrition rate is 37% when they do not. We can see that attrition rates are exceedingly high from when an employee start with the company at year zero, to the tenth year. This proportion is much higher within the first three years. If we look at stock option level in conjunction to overtime, we see that when stock option level is not offered –equal to zero—and overtime is clocked in, that the attrition rates are above 80%. The pattern is successively high even while stock options increase. When overtime is not clocked in and when stock options are not given, the attrition rate is high at 54%, which is expected due to correlation to no stock options, but as stock options are increase and while overtime is equal to no, then the proportion of attrition rates begin to drop.
We see that stock option level two is the best alternative, with an attrition rate of 26%, nearly half that of an employee without stock options.

Tableau Dashboard:

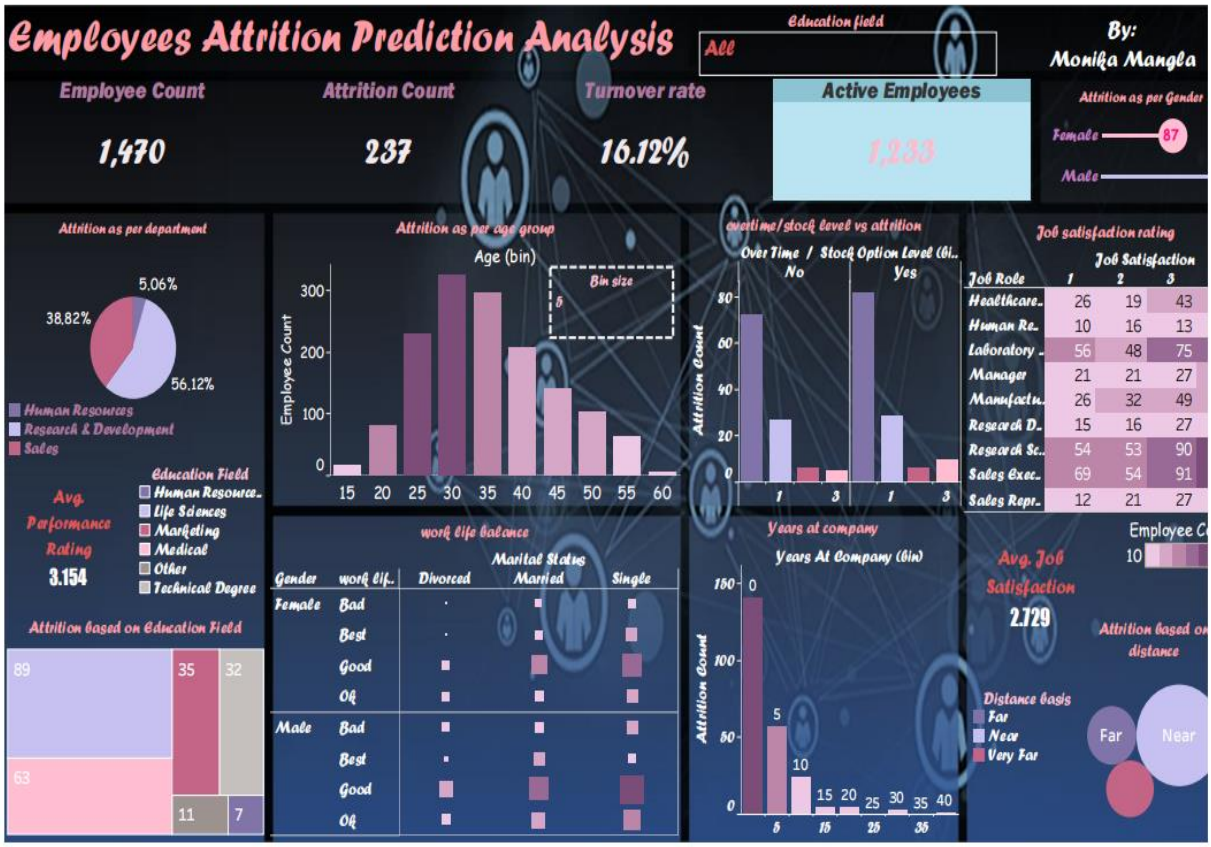


Tableau Public Server Link:

https://public.tableau.com/app/profile/monika.mangla/viz/attritiondashboard_17176794992920/EmployeeAttritionDashboard