**Case Study –** **Employee Attrition Prediction**

Singidi Sreeja Reddy

03-05-2023

**Executive Summary:**

This report aims to answer several questions that may be summarized from the Employee Attrition Prediction provided as a data set for analysis. Employee attrition prediction can help organizations improve workforce planning and management, foster a more engaged and productive workforce, and ultimately achieve long-term success.

**Definition:**

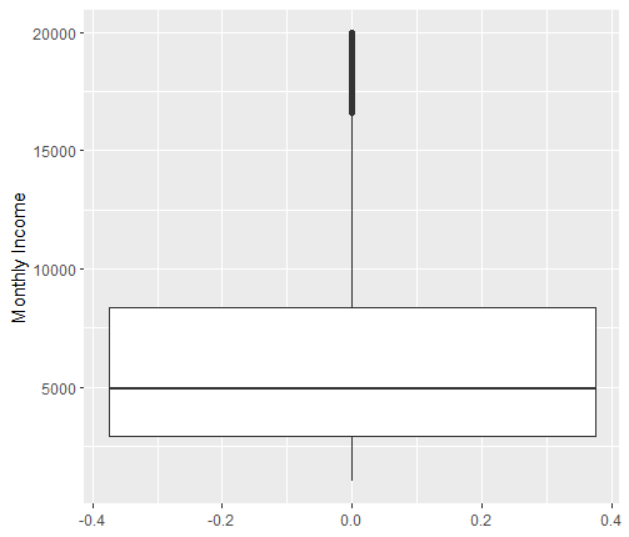
Employee attrition prediction uses data analysis and machine learning techniques to forecast which employees will likely leave a company within a given time frame. It involves collecting and analyzing various employee data, such as job performance, satisfaction, and demographics, to identify patterns and factors influencing their decision to stay or leave. Employee attrition prediction enables organizations to take proactive measures to retain their top talent, reduce the cost and disruption associated with employee turnover, and improve overall workforce planning and management.

It also enables companies to proactively retain their top talent and reduce the cost and disruption associated with employee turnover. This may include implementing targeted retention strategies, offering career development opportunities, improving workplace culture and communication, or adjusting compensation and benefits packages.

**Preparation:**

Employee Attrition determines the specific reasons for wanting to predict employee attrition, such as reducing turnover, improving retention, or enhancing workforce planning. The dataset gathers data on job satisfaction, salary, business travel, performance, hourly rate, and other factors influencing their decision to stay or leave.

I cleaned and preprocessed the data to remove missing values, outliers, and irrelevant variables. This may involve data transformation, normalization, or scaling, identifying the most relevant features affecting employee attrition using feature selection techniques such as correlation analysis, univariate or multivariate feature selection, or feature importance analysis.



**Figure 1**

I used the MonthlyIncome variable from this dataset to create a whisker plot shown in Figure 1.

**Analysis:**

Analyzing employee attrition prediction typically involves evaluating the prediction model’s performance to determine its effectiveness in identifying employees likely to leave the company.

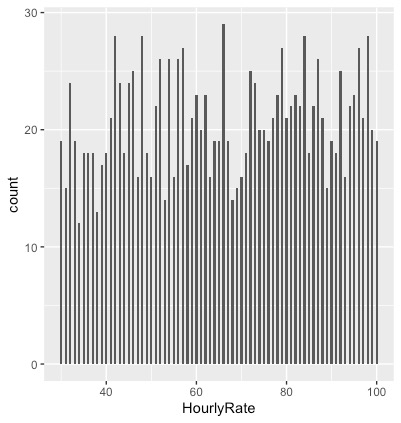
It also includes predicting the model’s performance using appropriate evaluation metrics such as accuracy and precision. These metrics can help us understand how well the model performs overall and how well it identifies employees likely to leave.

I am analyzing the importance of different features in predicting employee attrition. This can help identify which factors are most important in determining whether an employee is likely to leave and may inform retention strategies.

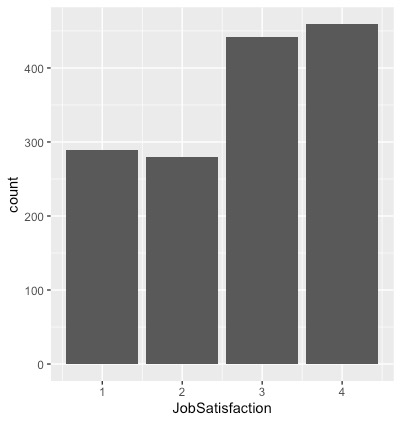
I can also use the model to identify high-risk groups of employees who are most likely to leave. This can help prioritize retention efforts and resources for those employees who are most at risk of leaving.

Here are some graphs that I generated with the codes provided below:

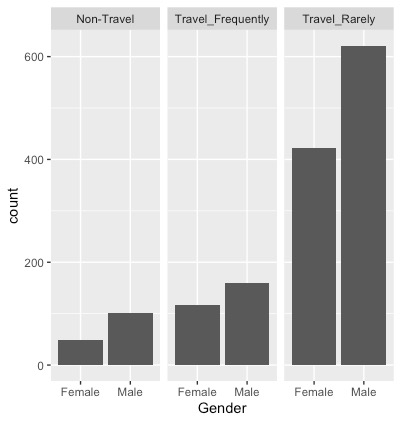
Hourly Rate:



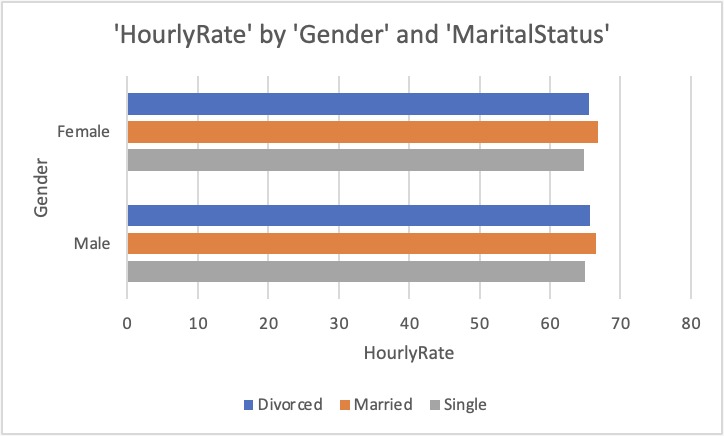
Job satisfaction rate:



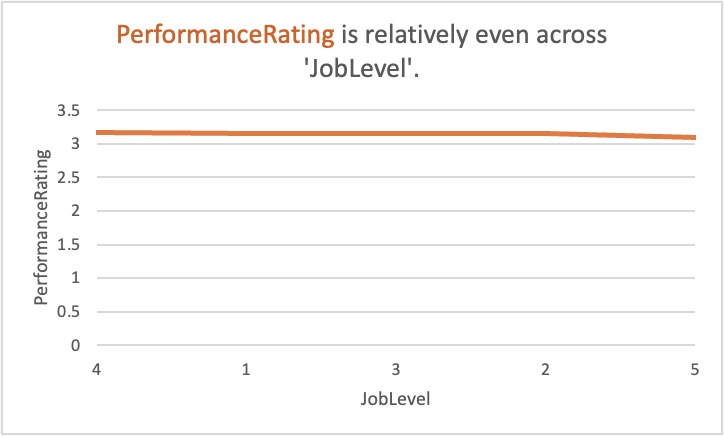
Gender:

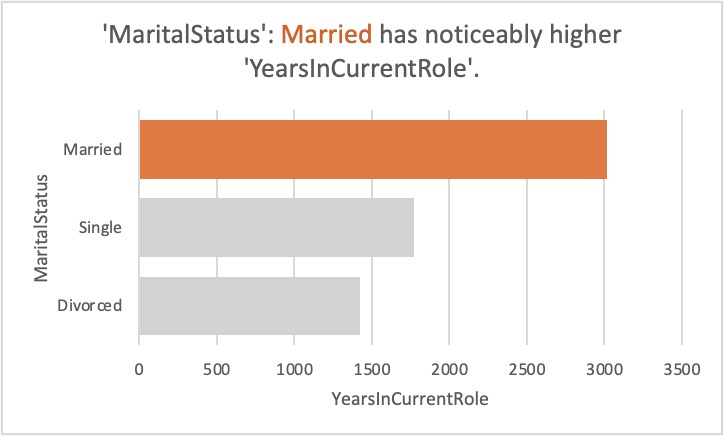


Graph for Hourly Rate by gender and Marital Status

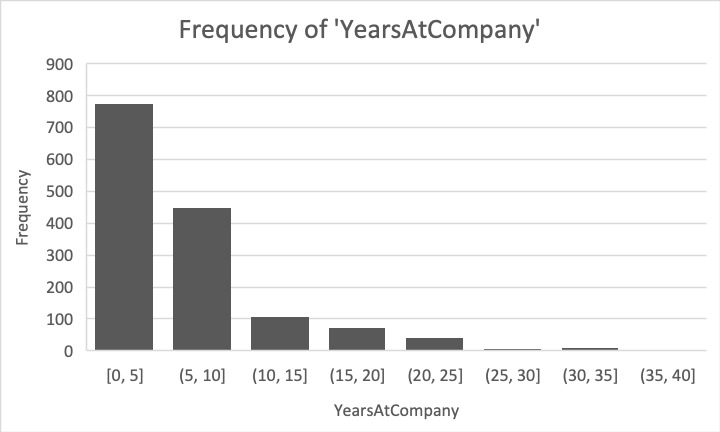


I have noticed that the Performance Rating is relatively even across Job Level. Here is the graph for the same.



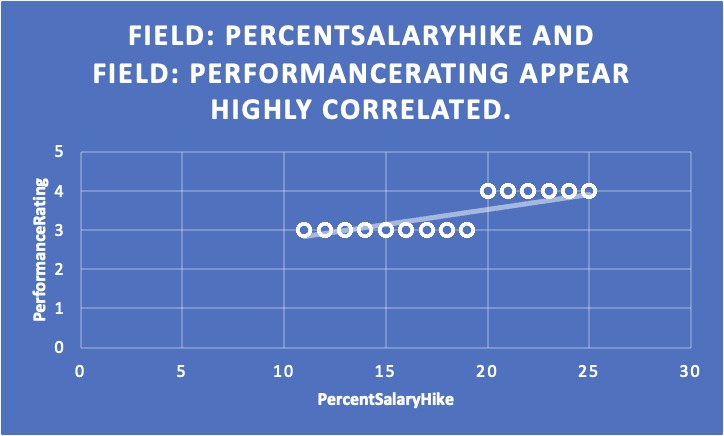


The graph shows the number of employees who have worked for the company for a specific number of years. This information can help understand employee turnover rates and patterns and identify trends or factors contributing to high or low retention rates for employees at different stages of their tenure with the company.



In this dataset, one could analyze the relationship between performance ratings and salary increases by examining the data to see how the two variables are related. For example, one could create a scatterplot of the data, with performance ratings on the x-axis and salary increases on the y-axis.

Analyzing this relationship can help organizations understand how they are compensating their employees and whether their compensation strategy is aligned with their performance management strategy. It can also help them identify any disparities in compensation between high-performing and low-performing employees, which may impact employee morale and retention.



**Model 1: Linear regression** assumes a linear relationship between the response and predictor variables and that the residuals are normally distributed with constant variance.

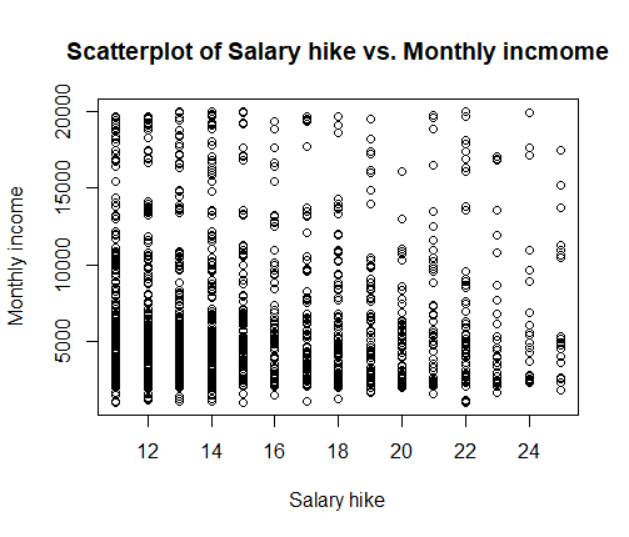
Using lm, we can fit a linear regression model to predict Monthly Income based on Age and Performance Rating based on Percent Salary Hike. Finally, I used the summary to print the model summary, which includes information about the estimated coefficients, standard errors, t-values, and p-values.

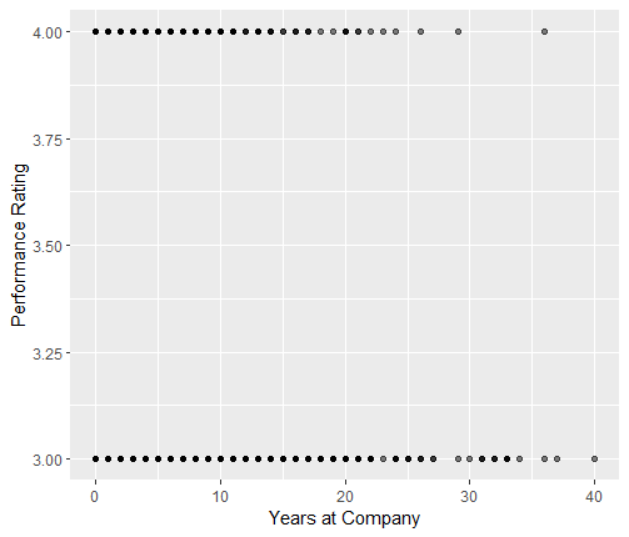
I split the dataset into two groups based on the Attrition variable: employees who have left (left) and those who have stayed (stayed). Finally, we can use a t-test to perform a t-test to compare the means of Monthly Income between the two groups.

The output of the t-test includes the estimated difference in means between the two groups, the standard error of the difference, the t-value, the degrees of freedom, and the p-value. A p-value less than 0.05 is typically considered statistically significant, suggesting strong evidence of a difference between the two groups.

Then, I generated the scatter plots from the dataset using the Salary hike and monthly income variables using the ggplot2 package.

Also, I have used geom\_point() to add the points to the plot and “labs” to add a title and axis labels to the plot.

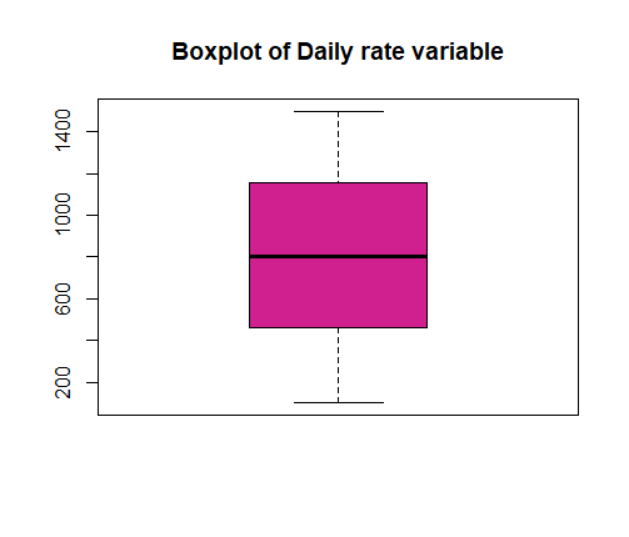


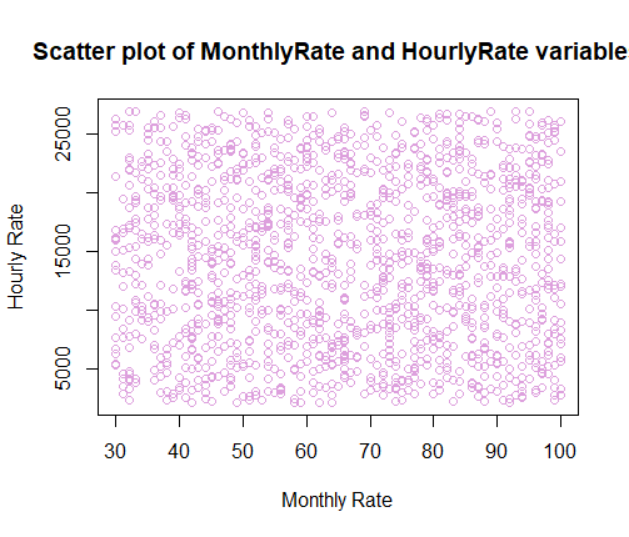


**Outliers**:

To identify outliers in my dataset, I used a boxplot and also a scatter plot to visualize the distribution of the variables. Outliers are data points that fall outside the range of "normal" values and are often indicated by points far away from the other data points on a plot.

Here's an example of how you can use a boxplot to identify outliers in the " DailyRate" variable



****

**Model 2:** An ANOVA (Analysis of Variance) model can be used to analyze the relationship between a continuous response variable and one or more categorical predictor variables. In the HR Attrition dataset, an ANOVA model can be used to examine the impact of various factors on an employee's monthly income.

The ANOVA table provides information about the model's overall significance and each predictor variable's significance. The table includes the following columns:

Df (degrees of freedom): The degrees of freedom associated with each source of variation in the model.

Sum Sq (sum of squares): The sum of squared deviations of the response variable from its mean.

Mean Sq (mean square): The sum of squares divided by the corresponding degrees of freedom represents the average variation for that source of variation.

F value: The F-statistic is the ratio of the predictor variable's mean square to the residual variation's mean square.

Pr(>F) (p-value): The p-value is associated with the F-statistic, which indicates the significance of the predictor variable in explaining the variation in the response variable.

A predictor variable with a small p-value (less than 0.05) suggests that the variable is significantly associated with the monthly income and has a non-zero effect. On the other hand, if a predictor variable has a significant p-value (greater than 0.05), it suggests that the variable is not significantly associated with the monthly income and can be dropped from the model.

**Conclusion:**

Employee attrition prediction is a critical process that can help organizations identify employees who are at risk of leaving and take proactive measures to retain their top talent. By using data analysis and machine learning techniques to identify patterns and factors that may influence employee turnover; organizations can develop targeted retention strategies, improve workforce planning, and reduce the cost and disruption associated with employee turnover.

Overall, employee attrition prediction is a valuable tool for organizations looking to improve their workforce management and retain their top talent. By understanding the factors that contribute to employee turnover and taking proactive steps to address these issues; organizations can foster a more engaged and productive workforce and achieve tremendous long-term success.