

Employee Attritions using R

Why Employee leave?

Table of Contents:

I. Summary of our Data:

II. Gender Analysis:

- a) Age Distribution by Gender
- b) Job Satisfaction Distribution by Gender
- c) Monthly Income by Gender

III. Analysis by Generation and Education:

- a) Understanding Attrition by Generation
- b) Attrition by Educational Level

IV. The Impact of Income towards Attrition:

- a) Average income by Department
- b) Satisfaction by Income
- c) Income and the Levels of Attrition
- d) Average Daily Rates
- e) Attrition by Overtime

V. Working Environment:

- a) Mean Salary by Job Role
- b) Attrition by Job Role
- c) Current Managers and Average Satisfaction Score
- d) Average Environment Satisfaction
- e) Distance from Work Status
- f) Employees have Stockoption levels
- g) Attrition due to Business Travels

VI. Analysis and Models

- a) Correlation Matrix
- b) Bi-variate Analysis
- c) Decision Trees
- d) Feature Importance
- e) Confusion Matrix

VII. Conclusion:

Top Reasons why Employees Leave the Organization

Introduction:

What is Attrition and what determines it?

Attrition: It is basically the turnover rate of employees inside an organization.

This can happen for many reasons:

1. Employees looking for better opportunities.
2. A negative working environment.
3. Bad management.
4. Excessive working hours.

Structure of the Project:

This project will be structured in the following way:

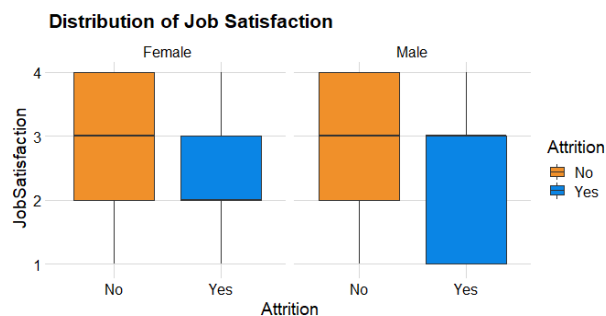
- **Questions:** Questions will be asked previous to the visualization to make sure the visualizations shown in this project are insightful.
- **Summary:** After each section I will provide a summary to understand what we got from the visualizations.
- **Recommendations:** What recommendations could be given to the organization to reduce the attrition rate .

I . Summary of our Data:

- **Dataset Structure:** 1470 observations (rows), 35 features (variables) Missing Data: Luckily for us, there is no missing data! This will make it easier to work with the dataset.
- **Data Type:** We only have two data types in this dataset: factors and integers Label" Attrition is the label in our dataset and we would like to find out why employees are leaving the organization!
- **Imbalanced dataset:** 1237 (84% of cases) employees did not leave the organization while 237 (16% of cases) did leave the organization making our dataset to be considered imbalanced since more people stay in the organization than they actually leave.

Distribution of our Labels:

This is an important aspect that we are dealing with an imbalanced dataset will help us determine what will be the best approach to implement our predictive model. 84% of employees did not quit the organization while 16% did leave the organization.

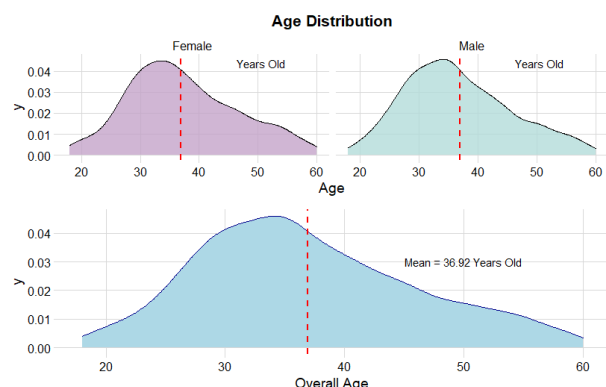


II . Gender Analysis:

In this section, we will try to see if there are any discrepancies between male and females in the organization. Also, we will look at other basic information such as the age, level of job satisfaction and average salary by gender.

a) Distribution of the Age of employees:

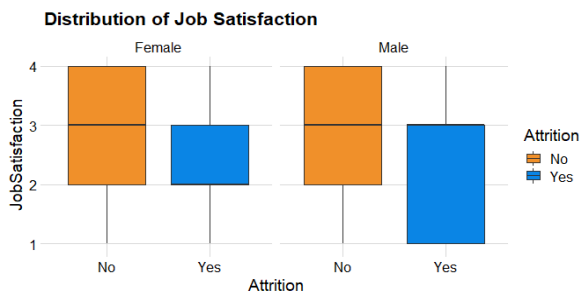
The average age of females is 37.33 and for males is 36.65 and both distributions are similar.



b) Distribution of Job Satisfaction:

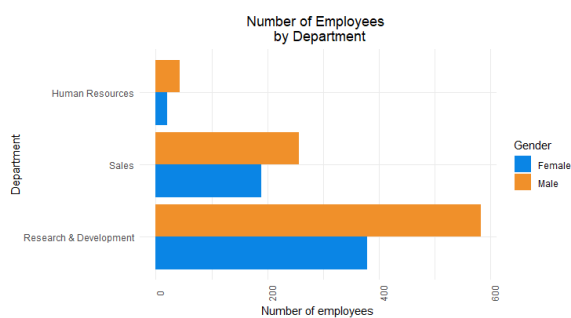
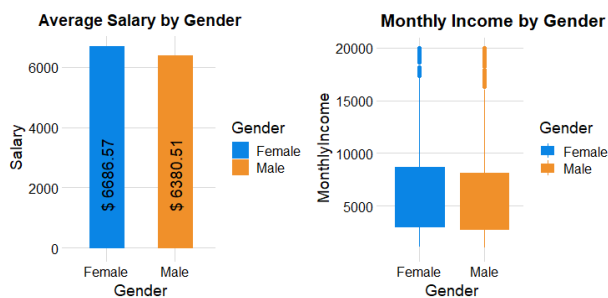
For individuals who didn't leave the organization, job satisfaction levels are practically the same.

However, for people who left the organization, females had a lower satisfaction level as opposed to males.



c) Monthly Income by Gender:

The average salaries for both genders are the same with males having an average of 6380.51 and females 6686.57.

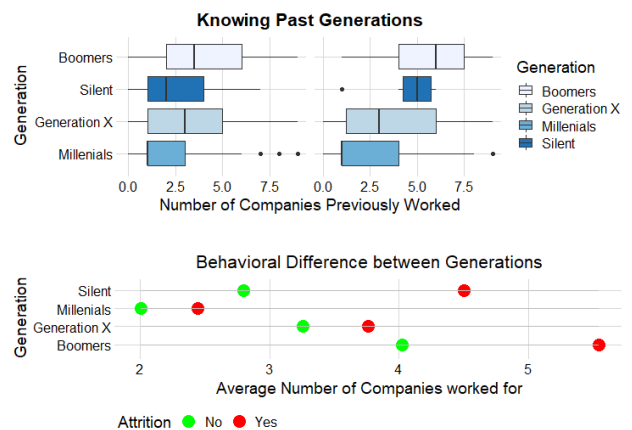


III. Generation and Education:

Each type of generation has their particular peculiarities and that is why we should explore in this dataset.

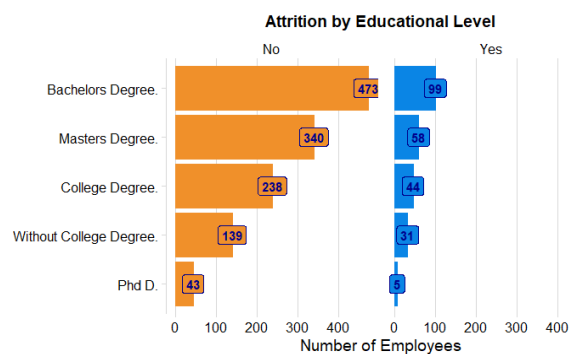
a) Generational Behaviour:

1. Most millennial are still relatively young, so that explains why the number of companies for millennial is relatively low however.
2. It seems that millennial are the ones with the highest turnover rate, followed by the boomers.
3. The newer generation which are the millennial opt to look more easy for other jobs that satisfy the needs on the other side we have the boomers which are approximating retirement and could be one of the reasons why the turnover rate of boomers is the second highest.



b) Attrition by Educational Level:

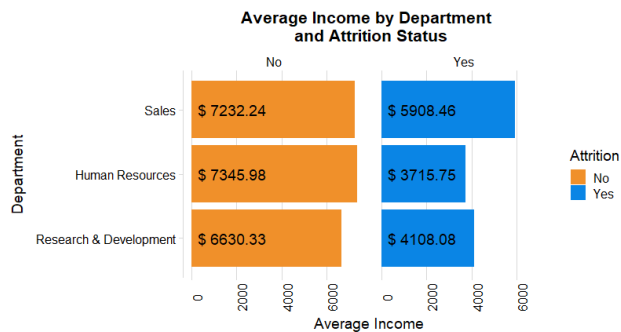
The bachelors are the ones showing the highest level of attrition which makes sense since Millennial create the highest turnover rate inside the organization.



IV. Impact of Income towards Attrition:

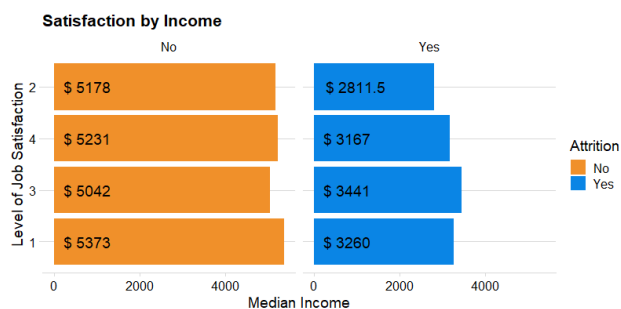
a) Average Income by Department:

There is a huge difference in each department by attrition.



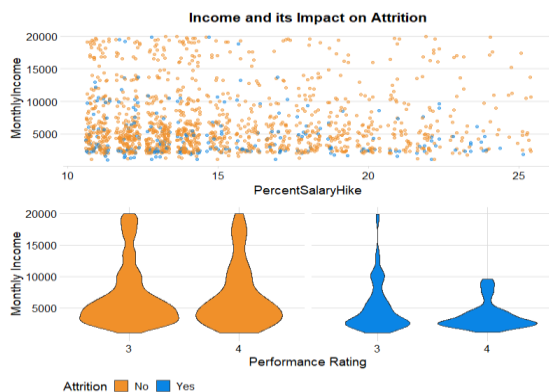
b) Satisfaction by Income:

The lower the job satisfaction the wider the gap by attrition status in the levels of income.



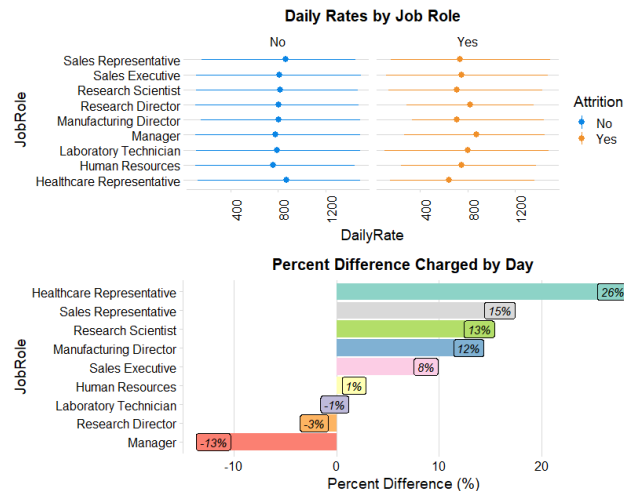
c) Income and the Level of Attrition:

This might indicate that at least for these roles, the sample population that left the organization was mainly because of income.



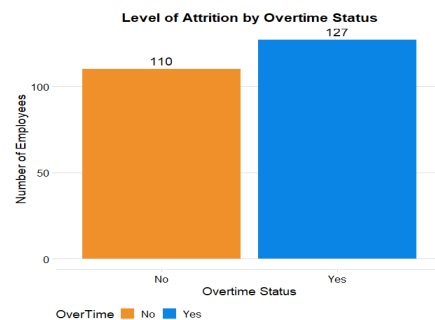
d) Average and Percent Difference of Daily Rates:

HealthCare Representatives, Sales Representatives, and Research Scientists have the highest daily rates differences in terms of employees who quit or didn't quit the organization.



e) Attrition due to Overtime:

Over 54% of workers who left the organization worked overtime!!!

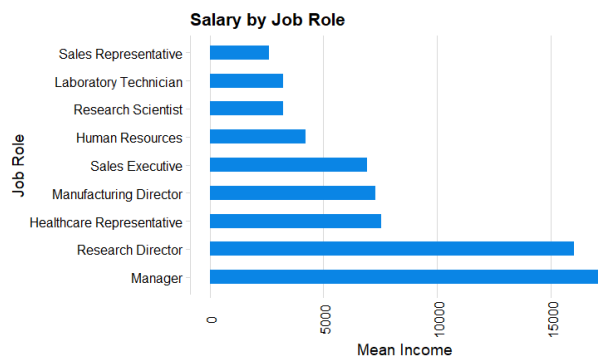


V. Working Environment:

In this section, we will explore everything that is related to the working environment and the structure of the organization.

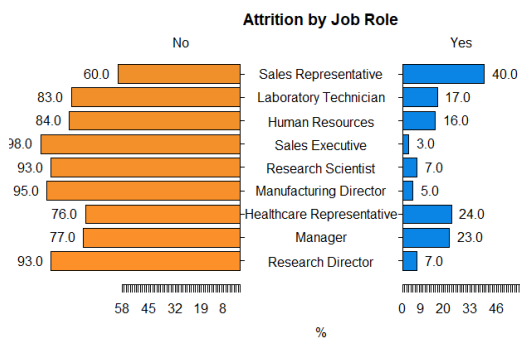
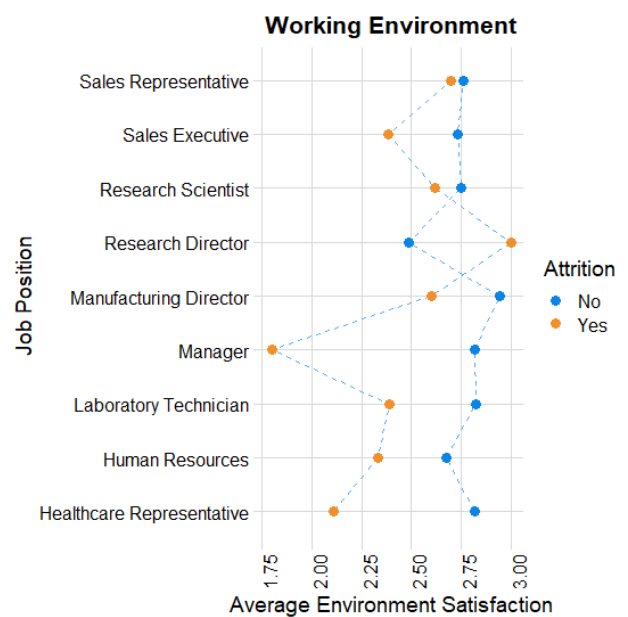
a) Mean Salary by Job Role:

Managers and Research Directors have the highest salary on average.



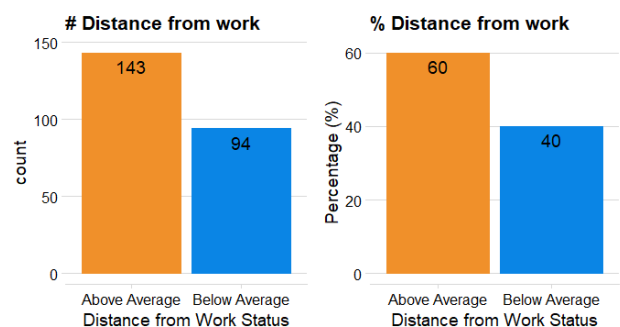
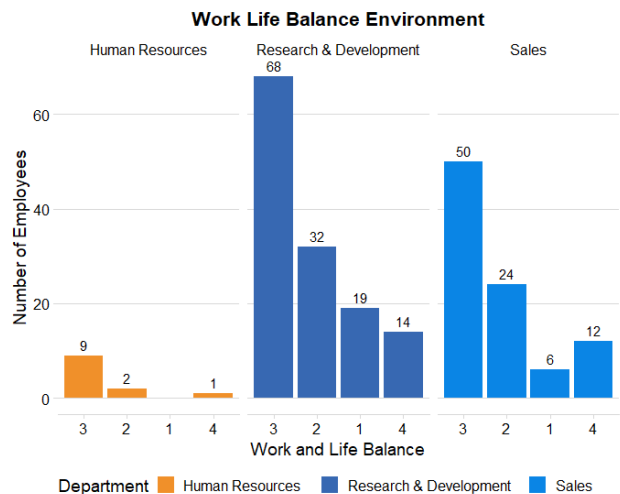
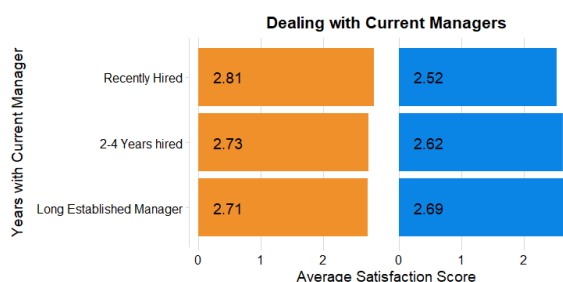
b) Attrition by Job Role:

Sales Representatives, HealthCare Representatives and Managers have the highest attrition rates. This could give us a hint that in these departments we are experiencing certain issues with employees.



c) Current Managers and Average Satisfaction Score:

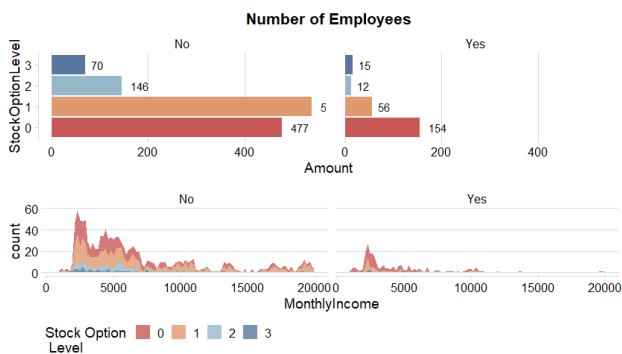
Employees that are dealing with recently hired managers have a lower satisfaction score than managers that have been there for a longer time.



d) Average Environment Satisfaction:

Managers and healthcare representatives are dealing with a lower working environment however; we don't see the same with sales representatives that could be because most sales representatives work outside the organization.

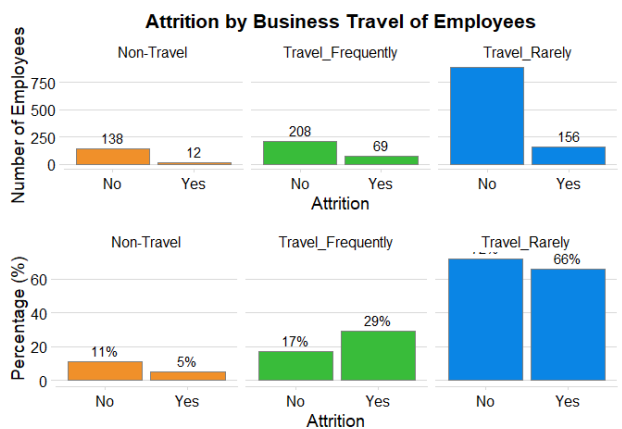
Stock option:



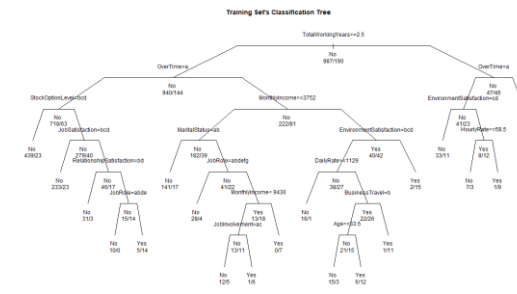
b) Bi-Variate Analysis:



Attrition due to Business Travels:



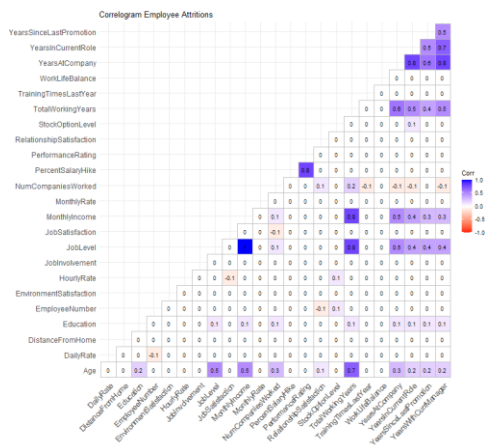
c) Decision Tree:



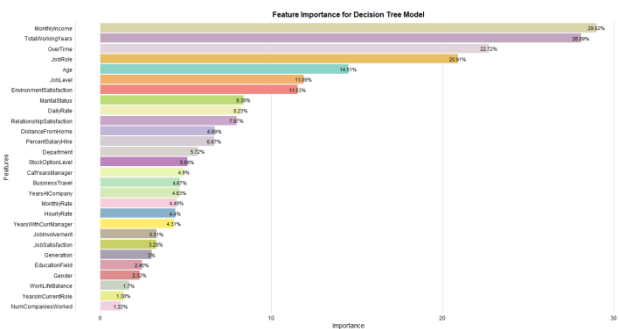
VI. Analysis and Models:

a) Correlation Matrix:

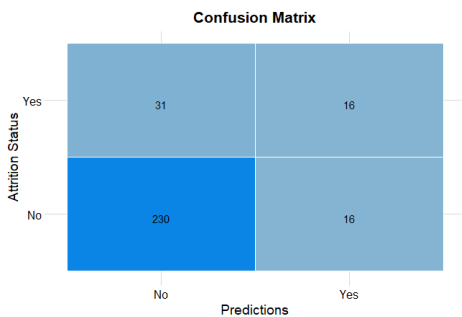
In this section we will understand what features have a positive correlation with each other. This tells us whether there is an association between two variables.



d) Feature Importance:



e) Confusion Matrix:



VII. Conclusion:

Top Reasons why Employees leave the Organization:

- **Monthly Income:** Income is a huge factor as why employees leave the organization in search for a better salary.
- **Opportunities:** It was found that employees receiving low income or no promotion or no career growth develops stress and as a result decides to look out for better opportunities.
- **Environment:** it was analysed that is no good understanding and support between superiors then employees would prefer to change to an organization with better environment.
- **Overtime:** If workload is high for the employees it is noticed that the satisfaction is lowered thus leading to employee leaving the organization.
- **No Overtime:** This was a surprise, employees who don't have overtime are most likely to leave the organization. This could be that employees would like to have a higher amount of income or employees could feel that they are underused.
- **Age:** This could also be expected, since people who are aiming to retire will leave the organization. Knowing the most likely reasons why employees leave the organization, can help the organization take action and reduce the level of Attrition inside the organization.