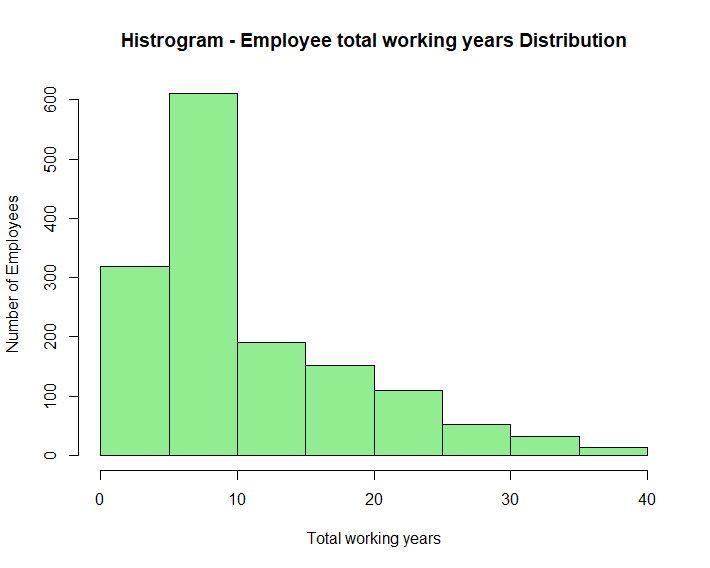
**3.Total Working Years (experience)**

This is the 3rd category that we are selected to conduct an analysis to find out how it affects to a business. We try to identify patterns, trends and finally get valuable insights using this column.

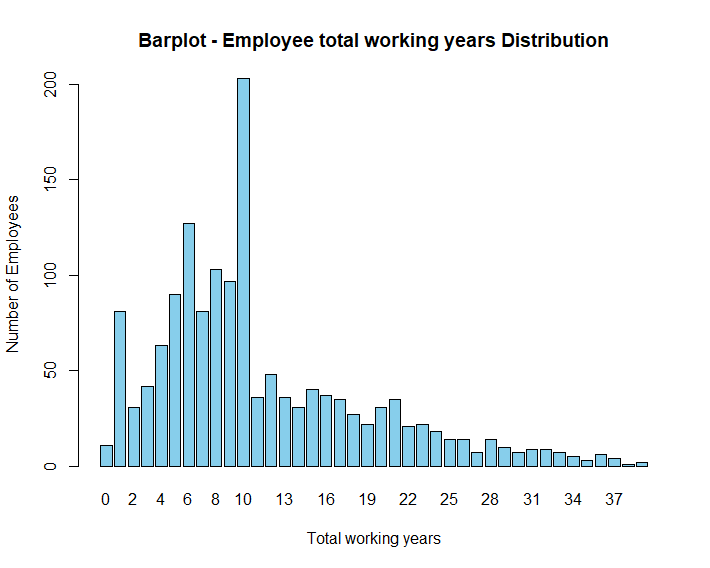
**1.Data visualizations**

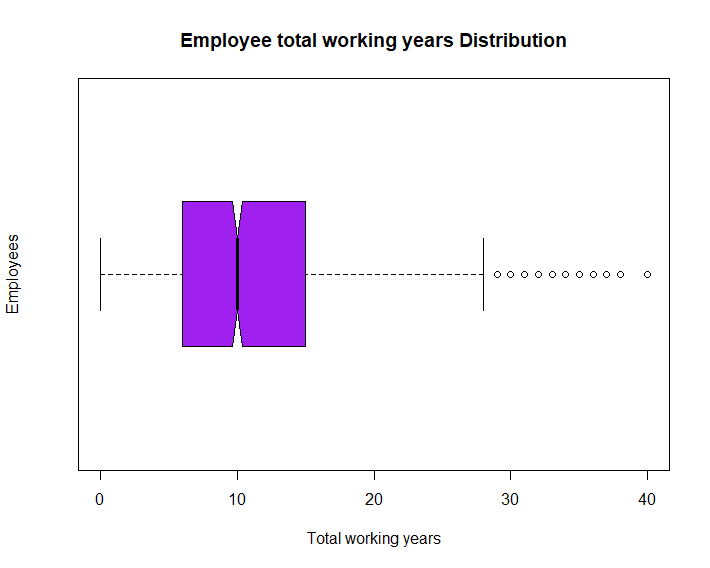
**1. Main Column Introduction**

This is an introduction about **“Total working Years”** column, **“Experience”** of an employee. This column includes numerical data. So, the most suitable visualization methods are Boxplots, Histograms, violin plots and Whisker plots.



This histogram represents distribution of “total working years” among employees. A larger proportion of employees(approximately 950) included in 0-10 years range. Upon that we can assume this company recruits more fresh graduates and young employees.





Above bar plot shows employee count that belongs to each experience year. And the box plot show min value is ‘0 years’, 1st quartile is ‘5 years’, median experience is ‘10 years’, 3rd quartile is ’15 years’, max value is ’28 years’ and other years are as outliers. The graphs and the visualizations are generated by using the ggplot2 libraries. From now **“Total working Years”** is called as, **“Experience”** of an employee.

**1. How experience vary on employee age distribution**

Above scatter plot shows the total experience of employees against age distribution. It has a positive correlation. The average number of working years increases with age. This is because people typically start working in their 20’s and continue working until, they retire in 60’s.

There is significant spread of variations in the data around the trendline. Based on the observations we can say, the average number of working years will continue to increase over time the employees ages, this can be helpful for get discissions about future workforce, and making future business policies.

**code:**

**2. More analysis about experience with boxplots**

This is a Bar-plot comparison of total working years of employees with age, the data are divided into 5 age groups. The box plot shows the spread of the data in each age group. The median age in each group increases as the age group gets older.

The spread of experience is greater in older age groups. We can assume that there is more variation in the number of years that people work in later life than in earlier life. There may be factors other than age that influence the experience of employees. Such as, education level, health, economic conditions and monthly income.

**Code:**

Let’s explore more detailed visualizations of experience for different age and age groups.

**Age Group 18-25**

**Age Group 26-35**

**Age Group 46-55**

**Age Group 36-45**

**Age Group 55+**

By observing each box-plot that plotted to age groups. We can get an idea about minimum value, 1st quartile, median value, 3rd quartile of total working years (experience) of employees in each age.

**Code:**

**3. Employee attrition rate with their experience**

This plot shows the experience by age distribution with attrition. There are two sets of bars for each age group. Red implements employees who haven’t left the company (Attrition = “No”), Green shows employees who have left (Attrition = “Yes”). When the experience increases the attrition rate gets low. Employees who are having 27 or more years of experience tends work until they become 60 and they retire in 60. There are lot of variations below 26 years of experience.

**code:**

**4. How job level varies on employee experience display with age**

The scatter plot shows the relationship between total working years and job level distribution, with age groups as a secondary factor. Here are some observations and insights we can get from this. There is positive correlation between these two variables. As employee gain more experience and become more older, they tend to move up in job level. But we can observe some variations like young employees get into higher job levels but, it may affect various factors. There are lot of variability in the data. Even within the same age group and total working years, there are wide range of job levels spread.

**Code:**

**4. Employee job satisfaction rate with their experience and job level**

Above graph shows the level of job satisfaction for total working years. Here are some observations and insights we found out. New employees tend to be more satisfied with their jobs. The highest level of job satisfaction is for employees with 10-30 years of experience. After that, job satisfaction starts to decline slightly. Employees who have been working longer and have higher job level seems to have low job satisfaction rate. This could be due to a number of factors such as people in this age group are approaching retirement and may be starting to think about slowing down or changing careers.

**Code:**

**5. Employee monthly income with their experience and job level**

Above graph shows how monthly income will differ with experience. Here are some observations and insights we found out. New employees seem to have low income. Employees who are having more work experience have higher monthly income. So, there is a positive correlation. Finally, we identified when your work experience and job level increase you paid more amount monthly. Factors such as, job role, education level and responsibility can be affected to your monthly income.

**Code:**

**5. More analysis about job level with boxplots**

This is a comparison of job level and total working years, with each box representing a different age group. Median job level increases with total working years across all age groups. The median job level tends to be higher for those with more total working years. This suggests that, when employees move up in job level as they gain experience. We can get more information using these boxplots.

**Code:**

**2. Data analytics through models**

**1. Predict attrition rate using Logistic regression model**

Logistic Regression is a supervised machine learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome ( Yes or No ). In this data set there are two columns (Attrition & Over Time) that compatible for a binary classification. So, we decided to choose attrition as our dependent variable (Y) and total working years plus Job level as independent variables (X).

At first, we load the necessary libraries. Then did the data cleaning and preprocessing. We checked for null values in the dataset and removed them.

In logistic regression the dependent variable always must be a binary outcome. So, we converted “Yes” & “No” to “1” & “0” in Attrition column.

**Code:**

After the cleaning & preprocessing we continued to the model building process. First of all, we split the dataset in to training and testing sets. Then trained the logistic regression model.

We trained the model using two independent variables so, it called as multivariate logistic regression model. Then we did the model evaluation using confusion matrix and ROC curve. Finally, we got the predicted output for attrition.

**Final outputs: ( model summery )**

As shown in the above the model accuracy is 0.853. and our test data set have 450 observations. The confusion matrix shows out of 450, 384 are stay at the company and 66 of them will leave the company. So, that is the prediction about attrition.

ROC curve image is shown below. ROC curve provides a more evaluation of a model's performance.