**HR Analytics Project- Understanding the Attrition in HR**

**HR (Human Resources)** in a company is a department responsible for managing employee relations, hiring, training, compensation, benefits, and ensuring compliance with labor laws and company policies. HR plays a crucial role in maintaining a positive work environment, fostering employee development, and supporting the company's overall goals and objectives.

**key functions typically handled by HR**:

1. **Recruitment and Hiring:** HR manages the recruitment process, including job postings, screening resumes, interviewing candidates, and making job offers.
2. **Employee Onboarding:** HR assists new hires with the orientation process, providing information about company policies, benefits, and procedures.
3. **Employee Relations:** HR handles employee grievances, conflicts, and disciplinary actions, striving to resolve issues in a fair and respectful manner.
4. **Training and Development:** HR organizes training programs to enhance employee skills and knowledge, promoting career development and succession planning.
5. **Compensation and Benefits:** HR manages employee compensation, including salaries, bonuses, and benefits such as health insurance, retirement plans, and other perks.
6. **Performance Management:** HR oversees performance evaluation processes, setting goals, providing feedback, and identifying opportunities for improvement or recognition.
7. **Compliance:** HR ensures compliance with labor laws, regulations, and company policies, including matters related to equal employment opportunity, workplace safety, and privacy.
8. **HR Information Systems:** HR utilizes technology to manage employee data, payroll, and other HR-related functions efficiently.

**Meaning of HR ATTRITION:**

"Attrition" in the context of HR refers to the natural reduction in the size of the workforce over time due to various factors such as resignations, retirements, terminations, or deaths. It is a measure of the rate at which employees leave the organization voluntarily or involuntarily.

Understanding and managing attrition is crucial for HR departments because high attrition rates can have negative implications for the organization, including increased recruitment and training costs, loss of institutional knowledge, decreased morale among remaining employees, and potential impacts on productivity and performance.

HR professionals often analyze attrition rates to identify trends, patterns, and underlying reasons for employee departures. By understanding the causes of attrition, HR can develop strategies to mitigate turnover and improve retention, such as enhancing employee engagement, offering competitive compensation and benefits, providing opportunities for career development, addressing workplace issues, and fostering a positive organizational culture.

Additionally, HR may conduct exit interviews with departing employees to gather feedback and insights that can help identify areas for improvement and inform retention strategies. By proactively addressing factors contributing to attrition, HR can work towards maintaining a stable and engaged workforce, which is essential for the long-term success of the organization.

In this project we will discuss about the attrition for the following dataset given for a firm.

Data Analysis:

1. Firstly analysing the data set given
2. Evaluating the null values and the data type of the dataset
3. As by using the proper code its evaluated that the null values are not present in the dataset

Age 0

Attrition 0

BusinessTravel 0

DailyRate 0

Department 0

DistanceFromHome 0

Education 0

EducationField 0

EmployeeCount 0

EmployeeNumber 0

EnvironmentSatisfaction 0

Gender 0

HourlyRate 0

JobInvolvement 0

JobLevel 0

JobRole 0

JobSatisfaction 0

MaritalStatus 0

MonthlyIncome 0

MonthlyRate 0

NumCompaniesWorked 0

Over18 0

OverTime 0

PercentSalaryHike 0

PerformanceRating 0

RelationshipSatisfaction 0

StandardHours 0

StockOptionLevel 0

TotalWorkingYears 0

TrainingTimesLastYear 0

WorkLifeBalance 0

YearsAtCompany 0

YearsInCurrentRole 0

YearsSinceLastPromotion 0

YearsWithCurrManager 0

1. Statistical analysis to determine the mean median of the dataset of all the numerical values.
2. After evaluating the statistics converting the all categorical values into their respective numerical values as machine learning models analysis only the numerical values.

**Graphical repreesntaion of the data in the dataset.**

1. Evaluating the duplicate values of the values present each column
2. Plotting the colunt plot for the attrition column
3. Ploting the attrition rate in the firm for each department.
4. Ploting the attrition rate in the firm for Education field of each employee
5. Ploting the attrition rate in the firm for job role an employee is doing in the company.
6. Checking with the help of graph how the attrition is affected by the gender of the employee in the company
7. Evaluation with the help of graph if the marital status of the employee affects attrition

* After checking the correlation f attrition with the other variables in the dataset now checking the outliers present in the data set using **Z-Score** .
* Creating the new data frame using the Zscore and having dataframe free of the outliers
* Featuring the x and y variables.
* Dropping the attrition from x and adding it to y

**Feature scaling the data**

1. In standard scaling we convert the data in the values ranging from -1 to 1 so that the evaluation process on the data can be dine in the proper manner.
2. Scaling is the best method to increase the accuracy of the model

**Variance inflation factor:**

The Variance Inflation Factor (VIF) is a measure used in statistics to quantify the severity of multicollinearity in regression analysis. Multicollinearity occurs when two or more independent variables in a regression model are highly correlated with each other. It can cause issues such as inflated standard errors, misleading coefficient estimates, and reduced interpretability of the regression model.

The VIF measures how much the variance of an estimated regression coefficient is increased due to multicollinearity. Specifically, it calculates the ratio of the variance of the coefficient when multicollinearity is present to the variance of the coefficient when there is no multicollinearity.

**Modelling:**

1. Checking the accuracy of the different model.
2. Considering the model where the accuracy is highest.
3. After the model selection the cross validation score is calculate
4. Cross validation:

Cross-validation is a technique used in machine learning and statistical modeling to evaluate the performance of a predictive model. It is particularly useful for assessing how well a model generalizes to unseen data.

The basic idea behind cross-validation is to partition the available data into subsets, often called "folds." The model is trained on a subset of the data (training set) and then evaluated on the remaining data (validation set or test set). This process is repeated multiple times, with different partitions of the data used for training and validation. The performance metrics obtained from each iteration are then averaged to provide a more robust estimate of the model's performance.

1. After the cross validating the score the hyper parameter tunning is done

**Hyper parameter tunning:**   
Hyperparameter tuning is a crucial step in machine learning where you adjust the hyperparameters of a model to optimize its performance. Hyperparameters are configuration settings that are not learned from the data directly, but rather set prior to the training process. These parameters can significantly impact the performance and behavior of the model

After completion of the process the model is tested on an unknown data and downloaded.