LIST OF FIGURES

ABSTRACT

HR Employee attrition, or turnover, is a major challenge for organizations, leading to increased recruitment costs, operational disruptions, and loss of expertise. To address this, companies are using machine learning to predict which employees are likely to leave, allowing for proactive retention strategies. This project develops a machine learning model in Python to predict employee attrition using various HR factors such as job satisfaction, salary, and tenure.

We utilize a dataset with over 1,400 employee records and 30 features. The data undergoes preprocessing, including handling missing values, encoding categorical data, and balancing for class imbalance. Several machine learning algorithms—Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting—are used to predict whether an employee will leave ("Yes") or stay ("No"). Model performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC.

Results show that ensemble methods, particularly Random Forest and Gradient Boosting, provide the most accurate predictions. The study highlights key factors influencing attrition, such as low job satisfaction and income, offering actionable insights for HR departments to reduce turnover and improve retention.

This project demonstrates the power of machine learning in solving HR challenges, providing data-driven insights for better decision-making and workforce stability.

EXECUTIVE SUMMARY

This report is study of HR Attrition Data Analysis using Python and Business Intelligence Tool based on a project. The aim in this visualization report is to studying employee attrition and figuring out what causes high turnover rates, HR analytics can be quite helpful. HR analytics can find patterns and correlations in employee data, including demographics, job satisfaction, performance, and tenure, that can be used to forecast which employees are most likely to quit
the organisation. This visualization showing statically information in a single dashboard in any organization, an employee is more vital for the success and viability of a business. The organization which attracts competent employees, utilize human resources efficiently, manage talent effectively and retain employees is securing the long-term success for the business. Human resource (HR) analytics have the potential to bring great value to the decision making the ability of HR leaders on human and organizational capital. Human resource analytics are useful for improving employee performance and getting an optimal return on investment (ROI) on its human capital. This article examines the application of the emerging discipline of HR analytics for business transformation and success. The relevant literature is reviewed about the integration of HR into business performance measurement. It indented and describes the HR's strategic role, functions and application of HR analytics. Thus, organizations should invest aggressively in this new discipline, link it to the rest of the business, and reskill their teams to bring data to work in every major people related decisions.

KEYWORDS:

HR analytics, data-driven on HR-employee attrition, HR professionals, ROI, Organizational transformation, Power BI, python with ML.

INTRODUCTION

In any organization, an employee is more vital for the success and viability of a business. The traditional role of human resource was to collect and keep track of the employee's professional and personal information through manpower inventory, payroll, health & safety and performance management. Now, with the advent of new technology, human resource departments are generating more data than ever before.

However, they often struggle to turn their data into valuable managerial insights The goal of human resource management is to use all available tools, techniques, practises, strategies, and approaches to understand how well employers and employees collaborate to achieve shared organisational objectives. Any organisation needs its human resources offers all the tools, techniques, approaches, principles, and forms of behaviour recommendations. Some organisations are seeking to implement these technologies that can improve their working practises because they recognise how crucial HRD is for any organisation.

Only the HR Function lagged behind because of its reliance on metrics and scorecards for data that While the human resource department contributes to successful organisational performances by establishing a link between HR actions and financial outcomes, human resources play a crucial role. Only the human resource function of any business lagged behind due to its reliance on metrics and scorecards of data that can be quantified. All functions of any organisation play important roles since their results can be easily examined. Due to its qualitative nature in the human resource function, the majority of the data could not be quantified.

Nevertheless, as technologies advance, human resources are moving forward in its knowledge of how people impact the organisation. Organizations generate a lot of HR data, making its collection and analysis challenging. The company has had tremendous expansion in recent years, particularly in the 21st century's in-service sector.

New cutting-edge techniques and technologies have been developed in such industries as they are expanding quickly. Organizational operations have improved as a result of development and new methods, and these organisations are growing effectively and efficiently. Businesses strive to increase their competitive advantages. Companies are embracing these technologies quite quickly, and techniques that might aid in keeping employees using the HRD as human resource management is crucial for any firm. It is the only organisational function that directly affects its workforce.

In order for employers and employees to collaborate and accomplish competitive corporate goals can be quantified; most HR data was qualitative and could not be quantified; analysis of HR data was largely based on intuition. But HR is now working to better understand how employees impact the company. The firm has evolved its HR analytics to also quantify the qualitative data.

HR analytics aid in the collection, analysis, and measurement of HR data. HR analytics offers current, accurate data and helps with future decision-making. It helps in providing a solution to organizational problems.

HR analytics aligns HR strategy overall business strategy. Employee attrition is the pace at which workers depart a company over a predetermined amount of time. High employee attrition can have negative effects on an organisation, and HR departments are crucial in controlling this problem. Poor management, low job satisfaction, little possibilities for career growth and advancement, insufficient pay and benefits, and a lack of work-life balance are a few reasons that might contribute to significant employee churn.

Implementing employee retention initiatives, enhancing employee engagement and communication, creating chances for professional growth and training, and providing competitive pay and benefits packages are just a few measures HR departments may do to lower attrition rates. High employee turnover can have serious negative effects on a business, including higher costs associated with hiring and recruiting new employees, less productivity, lower morale among surviving employees, and harm to the organization's reputation. It may be necessary for HR departments to devote additional time and money to hiring and onboarding new staff, which may be expensive and time-consuming.

HR departments can concentrate on creating a positive company culture that encourages employee engagement, job satisfaction, and a sense of belonging in addition to lowering attrition rates. This may entail fostering a positive workplace culture, fostering employee participation and feedback, recognising and rewarding employee contributions, and encouraging work-life balance. Overall, managing employee attrition is a critical function of HR departments.

By taking proactive steps to reduce attrition rates and improve employee retention, organizations can improve their bottom line, enhance their reputation, and create a more positive work environment for all employees. Finding high-risk employees: HR analytics can be used to find staff members who are most likely to leave the organisation.

Analyzing variables including employee engagement, job, performance, and tenure might help with this. After HR has identified high-risk workers, proactive measures can be taken to keep them on board. Knowing the causes of attrition: HR analytics can be used to determine the reasons why employees leave the organisation.

HR can find patterns and trends in employee exit surveys that can be used to address the underlying reasons of attrition.

Attrition forecasting: HR analytics can be used to identify the workers who are most likely to depart the firm soon. HR can create prediction models that can aid by studying employee data and identifying factors that contribute to attrition.HR analytics can assist in the development of efficient retention strategies. HR may create focused interventions to help increase employee retention rates by identifying the variables that contribute to attrition.

In general, HR analytics can offer useful insights on employee attrition and assist firms in creating practical retention strategies. Around 94 percent of executives and 88 percent of employees believe that distinct workplace culture is important to business success- source Deloitte. This is why it is important to build a positive work culture for your employees. Toxic work culture employee’s wellbeing, which directly leads to high absenteeism and low productivity.

How to fix it –

If toxic workplace culture is a major reason for your high employee turnover, then you need to evaluate whether or not your employees are feeling valued, heard, and appreciated in the organization. Conduct polls and surveys to take feedback on certain company policies, communicate openly with your employees, listen to their feedback and act on it and promote inclusion. Avoid micromanaging, offer more autonomy and flexibility.

In the end, everybody looks for an organization that offers them continuous growth and development opportunities. You can lose your best players due to a lack of suitable growth and development programs. Nobody would enjoy working in the same job role for the rest of their professional career. So always focus on providing necessary training and coaching to your employees.

How to fix it –

Always try to help your employees grow financially, professionally, and personally. Encourage learning among employees by offering them regular training sessions, by paying for their online courses or other educational programs, or also give them opportunities to grow within the company by assigning them greater responsibilities

Literature Review

The literature on employee attrition and turnover spans several decades and touches on both theoretical and practical aspects of managing human resources. Over time, predictive analytics and machine learning have emerged as powerful tools for addressing the challenge of employee retention. This section reviews key studies and methodologies from the literature related to employee attrition prediction, HR analytics, and machine learning techniques applied to HR problems.

Traditional Approaches to Attrition Prediction

Historically, organizations relied on qualitative methods, such as employee surveys and exit interviews, to understand the reasons behind employee turnover. These methods provided valuable insights into individual cases but were limited in scope when attempting to predict large-scale attrition trends.
Herzberg’s Two-Factor Theory (1959) and **Maslow's Hierarchy of Needs (1943)** are two classic theoretical frameworks that have been widely applied in understanding employee motivation and attrition. Herzberg’s theory distinguishes between motivators (factors that lead to satisfaction) and hygiene factors (those that prevent dissatisfaction), emphasizing that an absence of motivators could lead to employee turnover. Similarly, Maslow’s theory posits that employees' unmet needs, such as career advancement or job security, can contribute to higher turnover rates.

While these theories laid the groundwork for understanding employee behaviour, they did not offer quantitative tools for predicting attrition. Hence, HR managers typically used retrospective data to explain why employees left but were unable to proactively address the issue.

Modern Approaches Using HR Analytics

In the last two decades, advancements in data analytics have paved the way for more sophisticated approaches to understanding and predicting employee attrition. HR analytics, which involves using data-driven methods to make informed human resource decisions, has gained considerable attention. The shift from descriptive to predictive analytics in HR practices is crucial for improving workforce management.

Huselid (1995) pioneered the use of HR analytics by demonstrating that investments in high-performance work systems—such as training, performance appraisals, and employee engagement—significantly reduced turnover. His work marked the transition towards evidence-based HR practices. However, the predictive element was still nascent, and organizations primarily focused on improving HR policies based on observed patterns rather than predicting future attrition.

More recently, studies such as **Cappelli (2008)** and **Davenport et al. (2010)** highlighted the potential of HR analytics to forecast critical workforce outcomes, including attrition. Cappelli emphasized the need for companies to develop predictive models to anticipate workforce changes, while Davenport explored how organizations could apply analytics across HR functions, including hiring, performance management, and turnover prediction.

Predictive Analytics and Machine Learning in HR

Machine learning has revolutionized the way organizations predict employee attrition. Several studies have explored the application of machine learning algorithms to analyze HR datasets and forecast attrition based on employee demographics, performance indicators, and other relevant factors.

Witten and Frank (2005) introduced the concept of applying machine learning algorithms to solve classification problems. This framework has been widely adapted for employee attrition prediction, as attrition is a binary classification problem (i.e., predicting whether an employee will leave or stay).

In a study by **D. Tziner (2006)**, the role of data mining techniques in HR was explored, focusing on turnover and absenteeism. The study showed how data mining models, such as decision trees and clustering, could help HR managers make better decisions regarding employee retention. However, this research primarily used older statistical techniques and was limited in the ability to handle high-dimensional data.

Ramesh and Ganesan (2018) provided one of the more detailed analyses on the use of machine learning for employee attrition prediction. Their study evaluated different machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting, and found that ensemble techniques like Random Forest and Gradient Boosting consistently outperformed other methods in predicting attrition. They also highlighted the importance of feature engineering, demonstrating that job satisfaction, work-life balance, and compensation were the most influential predictors of attrition.

In their influential work, **Kaur and Kang (2020)** applied deep learning models to predict employee attrition and compared them to traditional machine learning methods. The study demonstrated that while deep learning models, such as neural networks, could offer marginal improvements in accuracy, simpler models like Logistic Regression or Random Forest were often more interpretable and easier to implement in HR environments.

Key Factors Influencing Attrition

Numerous studies have identified the key factors contributing to employee turnover. **Hausknecht et al. (2009)** reviewed various studies on voluntary turnover and identified factors such as job satisfaction, organizational commitment, pay, promotion opportunities, and supervisory support as the most critical predictors of attrition. Their findings were corroborated by **Griffeth, Hom, and Gaertner (2000)**, who performed a meta-analysis of 25 years of turnover research and concluded that pay, job satisfaction, and organizational culture were among the most influential determinants of employee turnover.

In a more recent study, **Maertz and Griffeth (2004)** introduced the "Four Paths" model of voluntary turnover, which categorized reasons for attrition into four groups: affective (emotional attachment), contractual (perceived obligations), calculative (rational decision-making), and alternative paths (availability of better job offers). This model provided a more nuanced view of employee turnover and is frequently referenced in machine learning studies that aim to capture the complexity of attrition behavior.

Machine Learning Techniques for Attrition Prediction

Several machine learning techniques have been applied in the context of employee attrition prediction. Some of the most commonly used techniques include:

- **Logistic Regression**: Frequently used for binary classification problems, logistic regression is one of the simplest and most interpretable models for predicting employee attrition. Studies by **Jain and Nagori (2019)** have shown that logistic regression performs well on structured HR datasets.
- **Decision Trees**: Decision trees are widely used due to their interpretability. Research by **N. Pattnaik and Mishra (2019)** shows that decision trees, when used with proper pruning techniques, can yield high accuracy for attrition prediction. However, decision trees tend to overfit on smaller datasets.
- **Random Forest and Gradient Boosting**: Ensemble methods such as Random Forest and Gradient Boosting have become popular for their ability to handle complex, non-linear relationships in data. In their study, **Saha et al. (2020)** demonstrated that Random Forest outperformed Logistic Regression and Decision Trees, achieving an accuracy of over 90% in predicting employee attrition.
- **Support Vector Machines (SVM)**: SVMs are known for their robustness in handling high-dimensional data. **K. M. Gopika and K. Gopinath (2020)** applied SVM to predict employee attrition, achieving strong results, especially when the data was linearly separable.

- **Neural Networks and Deep Learning**: While more complex, neural networks have shown promise in handling large, high-dimensional datasets. In the study by **Sharma and Batra (2021)**, deep learning models achieved slightly higher accuracy than traditional models, but with the trade-off of being harder to interpret and more resource-intensive to train.

Summary of Literature Insights

From the literature, it is clear that:

1. Predictive analytics and machine learning are valuable tools for HR professionals looking to proactively manage employee turnover.
2. Several machines learning models, including logistic regression, decision trees, random forests, and gradient boosting, have been successfully applied to employee attrition prediction.
3. Job satisfaction, compensation, organizational commitment, and work-life balance are the most commonly identified factors contributing to employee attrition.
4. While more complex models like deep learning can offer higher accuracy, simpler models often provide more interpretable and actionable insights for HR teams.

This literature review establishes a strong foundation for developing a machine learning-based employee attrition prediction model, incorporating key factors identified in previous research and selecting algorithms that balance performance with interpretability.

Problem Statement

Employee attrition, also referred to as employee turnover, is a critical issue for organizations of all sizes and across various industries. When employees leave an organization—whether voluntarily or involuntarily—it can result in significant costs and disruptions. These costs include not only the direct expenses associated with recruitment, hiring, and training new employees but also the indirect costs related to lost productivity, lowered morale, and the potential loss of institutional knowledge. Additionally, the departure of key employees can create a talent gap, leading to delayed projects and compromised customer satisfaction.

Despite these negative impacts, many organizations still struggle to effectively manage and predict employee attrition. Traditional methods, such as exit interviews and employee satisfaction surveys, often provide limited insights because they are retrospective in nature—addressing the reasons for turnover after employees have already left. As a result, HR teams are often left in a reactive mode, where they can only take action after attrition has already occurred, instead of preventing it in the first place.

The advent of data-driven solutions, such as predictive analytics and machine learning, offers organizations the ability to shift from a reactive to a proactive approach in managing employee attrition. Predictive models can analyze historical employee data and identify patterns and signals that indicate whether an employee is likely to leave the organization. This type of foresight allows HR teams to intervene before employees decide to leave, thereby mitigating the negative impacts of attrition.

3.1. Key Challenges

There are several challenges organizations face when it comes to managing and predicting employee attrition:

1. **Understanding the Complex Drivers of Attrition**: Employee turnover is influenced by a wide variety of factors, including job satisfaction, compensation, work-life balance, career development opportunities, and relationships with supervisors or peers. Identifying and quantifying the most critical factors that contribute to attrition is a complex task, as these factors often interact with each other in non-linear ways.

2. **High Dimensionality of Data**: HR datasets often contain numerous variables related to employee demographics, job performance, work environment, and compensation. Analyzing such high-dimensional data can be challenging, particularly when trying to identify which features are the most important in predicting attrition.

3. **Imbalanced Data**: In many organizations, the number of employees who leave (attrition) is significantly smaller than the number who stay, resulting in an imbalanced dataset. This imbalance can negatively impact the performance of predictive models, as they may become biased toward the majority class (employees who stay) and fail to accurately predict attrition.

4. **Interpretability of Predictive Models**: While machine learning algorithms can offer high predictive accuracy, many of the more sophisticated models—such as deep learning or ensemble methods—are often difficult to interpret. HR professionals need models that not only predict attrition with high accuracy but also provide interpretable insights into the key drivers of employee turnover. This allows HR teams to take meaningful and targeted actions to address the root causes of attrition.

3.2. Business Impact

High attrition rates can have serious financial and operational implications for organizations. According to industry studies, the cost of replacing an employee can range from 50% to 200% of that employee's annual salary, depending on their role and level within the organization. Beyond the financial burden, attrition can negatively impact team dynamics, employee morale, and organizational performance.

The following are some of the key business impacts of employee attrition:

- **Increased Recruitment Costs**: Finding and hiring new employees incurs direct costs related to recruitment advertising, time spent by HR personnel, and external hiring agencies.

- **Training and Onboarding Costs**: New hires often require significant time and resources to reach full productivity, including formal training programs and informal learning through interactions with colleagues.

- **Loss of Productivity**: Departing employees often leave gaps in team workflows, resulting in lost productivity, delayed projects, and inefficiencies.

- **Decline in Employee Morale**: Frequent turnover within teams can lead to decreased morale among remaining employees, as they may feel overburdened with additional work or become anxious about the stability of their own positions.

- **Loss of Institutional Knowledge**: Long-tenured employees possess valuable institutional knowledge that is difficult to replace when they leave the organization.

By developing a predictive model for employee attrition, organizations can take pre-emptive actions to reduce these impacts, thereby improving overall business performance.

3.3. Project Objective

The goal of this project is to develop a machine learning-based predictive model that can accurately forecast employee attrition based on historical employee data. This model will enable HR teams to identify employees who are at high risk of leaving the organization and allow for timely interventions to retain top talent. Specifically, the project seeks to:

1. **Build an Attrition Prediction Model**: Use machine learning techniques to develop a model that predicts whether an employee will leave the organization based on factors such as job satisfaction, performance, compensation, and work-life balance.

2. **Identify Key Factors Influencing Attrition**: Analyze the data to determine which factors are most strongly associated with employee turnover. These insights will provide HR teams with actionable information to improve retention strategies and address areas of concern.

3. **Handle Data Challenges**: Address the challenges associated with high-dimensional and imbalanced datasets, ensuring that the model remains accurate and reliable.

4. **Provide Interpretability**: Ensure that the predictive model is interpretable so that HR professionals can easily understand the drivers of attrition and take targeted actions based on these insights.

3.4. Problem Scope

The project focuses on voluntary employee attrition (i.e., employees who leave the organization of their own accord) rather than involuntary attrition (e.g., layoffs or terminations). The analysis will be conducted using historical employee data, including demographic information (age, gender, education), employment details (job role, salary, department), and performance indicators (job satisfaction, work-life balance, years at the company). The output of the model will be a binary classification—indicating whether an employee is likely to stay or leave.

Additionally, the project will explore different machine learning algorithms to find the most suitable model for the dataset at hand. The performance of these models will be evaluated using key metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure that the model performs well, even in the presence of imbalanced data.

3.5. Research Questions

To achieve the objectives outlined above, the project will address the following key research questions:

1. **Which machine learning algorithms are best suited for predicting employee attrition?**

   The project will evaluate various algorithms (e.g., Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, Support Vector Machines) to determine the most effective model for predicting attrition.

2. **What are the most important factors that contribute to employee turnover?**

   The project will analyze the dataset to identify the features that are most strongly associated with employee attrition, providing HR teams with insights into the key drivers of turnover.

3. **How can we address the challenges posed by imbalanced data in attrition prediction?**

   Since attrition is a rare event in many organizations, the project will explore techniques such as oversampling, under sampling, and cost-sensitive learning to improve model performance on imbalanced datasets.

4. **How can we ensure that the model is interpretable and provides actionable insights?**

   The project will prioritize the use of models that offer interpretability (e.g., feature importance analysis) so that HR professionals can easily understand the results and use them to inform retention strategies.

Objectives

The overall goal of this project is to develop an accurate and interpretable machine learning model to predict employee attrition and provide actionable insights for HR teams. The project's objectives are centred around solving key challenges related to employee turnover, improving decision-making capabilities within HR departments, and enabling proactive intervention to reduce attrition rates.

Below, the objectives are detailed:

4.1. Develop an Accurate Attrition Prediction Model

The primary objective of the project is to create a machine learning model that can accurately predict whether an employee will leave the organization (attrition) based on historical employee data. Accuracy is critical to ensure that the model can reliably forecast future attrition events and help HR teams make informed decisions.

- **Steps to Achieve This Objective:**
  - **Data Collection and Preprocessing**: Collect and clean a relevant dataset, ensuring that it includes key factors such as employee demographics, job performance, work environment, compensation, and job satisfaction. Address missing data, outliers, and errors in the dataset to ensure model quality.
  - **Feature Engineering**: Engineer new features or transformations of existing features that could enhance the model's predictive power. For example, create variables for employee tenure, job-level trends, or satisfaction ratios.

- **Model Selection**: Experiment with various machine learning algorithms to identify which model performs best for predicting employee attrition. The models to be evaluated may include Logistic Regression, Decision Trees, Random Forest, Gradient Boosting Machines (GBM), and Support Vector Machines (SVM).

- **Model Optimization**: Fine-tune the selected model(s) using hyperparameter optimization techniques (such as Grid Search or Random Search) to improve performance.

- **Performance Metrics**:

  - Evaluate model performance using key classification metrics such as **accuracy**, **precision**, **recall**, **F1-score**, and **ROC-AUC**. Given that attrition data may be imbalanced (with fewer employees leaving than staying), precision and recall will be particularly important to avoid false negatives (predicting that an employee will stay when they are likely to leave).

4.2. Identify Key Factors Contributing to Employee Attrition

A significant objective is to determine which factors most influence employee attrition. These insights will allow HR professionals to understand the drivers of turnover and develop targeted strategies to address them.

- **Steps to Achieve This Objective:**

  - **Feature Importance Analysis**: For models like Random Forest, Decision Trees, and Gradient Boosting, use built-in methods to rank features based on their importance in predicting attrition. For simpler models like Logistic Regression, examine the coefficients to assess feature significance.

    - **SHAP Values (SHapley Additive exPlanations)**: Implement SHAP analysis to provide a more detailed understanding of how individual features impact predictions. This method helps explain both global feature importance and the influence of features on specific employee predictions.

    - **Correlation and EDA (Exploratory Data Analysis)**: Perform a thorough exploratory analysis to identify relationships between features (e.g., job satisfaction, salary, tenure) and attrition, both graphically and statistically. This will help uncover patterns that may not be immediately evident in a machine learning model.

- **Actionable Insights**:

    - HR teams will be able to use the insights gained from this analysis to adjust policies, improve employee engagement, offer better career development opportunities, and address other factors contributing to high turnover.

4.3. Handle Data Imbalance and Improve Model Generalization

Attrition data is often imbalanced, meaning that the number of employees who stay vastly outnumbers those who leave. This imbalance can lead to biased models that perform well on the majority class (employees who stay) but poorly on the minority class (employees who leave). An important objective is to address this imbalance and ensure that the model generalizes well across both classes.

- **Steps to Achieve This Objective:**
  - **Resampling Techniques**: Use resampling techniques to balance the dataset, such as:
    - **Oversampling** the minority class (e.g., using Synthetic Minority Over-sampling Technique (SMOTE)) to create more data points for employees who leave.
    - **Undersampling** the majority class to reduce the number of data points for employees who stay.
  - **Cost-Sensitive Learning**: Implement cost-sensitive algorithms that assign a higher penalty for misclassifying employees who are predicted to stay but actually leave. This approach helps models focus on predicting attrition more effectively.
  - **Cross-Validation**: Apply cross-validation techniques (e.g., K-fold cross-validation) to ensure that the model does not overfit to any particular subset of data and generalizes well to unseen data.

- **Evaluation Metrics for Imbalanced Data**:
  - Place a higher emphasis on **precision**, **recall**, and the **F1-score** for the minority class (employees who leave). These metrics are more useful than accuracy in cases of imbalanced data, as they directly evaluate how well the model predicts attrition.

4.4. Provide Interpretability and Actionable Insights

While high accuracy is important, HR professionals also need to understand the reasoning behind the model’s predictions. A black-box model is less valuable if it cannot provide insights that guide decision-making. Therefore, another key objective is to ensure that the model provides interpretable and actionable insights.

- **Steps to Achieve This Objective:**
    - **Use Interpretable Models**: Select models that are inherently interpretable, such as Decision Trees and Logistic Regression, alongside more complex models (like Gradient Boosting or Neural Networks). Ensure that even complex models have mechanisms for interpretability, such as SHAP or LIME (Local Interpretable Model-agnostic Explanations).

    - **Explain Model Predictions**: Use feature importance metrics to explain why certain employees are predicted to leave. This will allow HR teams to intervene with specific measures, such as increasing compensation or improving job satisfaction.

    - **Visualizations**: Develop clear, intuitive visualizations (e.g., bar charts, decision trees, or SHAP plots) that illustrate the most important features affecting employee attrition. These visualizations should help HR managers quickly grasp the key drivers of turnover.

4.5. Offer Recommendations for HR Intervention

One of the final objectives is to go beyond predictive modelling and provide HR departments with actionable recommendations based on the findings. These recommendations should help organizations reduce attrition by addressing the root causes uncovered during the analysis.

- **Steps to Achieve This Objective:**
    - **Develop Retention Strategies**: Based on the most influential factors identified, suggest retention strategies. For example, if the model finds that job satisfaction and career development opportunities are major factors driving attrition, HR could focus on enhancing employee engagement programs, providing career advancement opportunities, or offering tailored professional development plans.
    - **Targeted Intervention**: Use the predictive model to create a prioritized list of employees who are at the highest risk of leaving. HR can then implement targeted interventions, such as offering personalized retention packages or conducting one-on-one meetings with employees who are at risk of leaving.
    - **Continuous Monitoring**: Recommend building a feedback loop for continuous monitoring of employee sentiment and attrition risk, allowing HR teams to dynamically adjust their retention efforts as new data becomes available.

4.6. Validate and Evaluate the Model in a Real-World Scenario

Once the model is built, it is essential to validate it on real-world employee data to ensure its effectiveness in a practical business environment. This objective is critical to proving the model's reliability and ensuring that it provides value to the HR team.

- **Steps to Achieve This Objective:**
  - **Test on a Holdout Dataset**: After training the model on a subset of the data, validate it on a separate test set to evaluate how well the model performs on unseen data.
  - **Simulate Real-World Scenarios**: Implement the model in a live environment or simulate its use on a continuous stream of employee data to evaluate how well it adapts to changes in the workforce.
  - **Compare Predictions with Actual Attrition**: Track the model's predictions over a specified period and compare them to actual attrition rates to validate its predictive capabilities.

- **Real-World Model Performance**: If the model performs well in a real-world setting, it can be deployed to provide continuous, actionable insights for the HR team to use in retention strategies.

System Specifications

The system specification section provides a detailed outline of the hardware, software, and technical requirements necessary to build, run, and maintain the employee attrition prediction model using Python and machine learning. These specifications ensure that the system is optimized for efficient data processing, model training, and deployment.

5.1. Hardware Requirements

The minimum hardware configuration ensures the project can run with smaller datasets and moderate processing capabilities. This setup is suitable for model development and testing on smaller scales.

- **Processor**: Intel Core i5 (4th Gen) or AMD equivalent
- **RAM**: 8 GB DDR3
- **Storage**: 500 GB HDD (or SSD for faster processing)
- **Operating System**: Windows 10 / macOS

5.2. Software Requirements

The project will be developed primarily in **Python** due to its simplicity, vast ecosystem of **machine learning libraries**, and strong community support.

- **Version**: Python 3.7 or higher
- **Distribution**: Anaconda distribution is recommended for easy package management and environment setup.

5.2.1. Integrated Development Environment (IDE)

A good IDE will make the development process easier by providing debugging, code navigation, and package management features. Recommended IDEs include:

- **Jupyter Notebook** (for interactive coding, data exploration, and visualization)

5.2.2. Libraries and Dependencies

The following Python libraries are essential for data manipulation, visualization, and machine learning model development:

- **Data Handling**:
  - **Pandas**: For data manipulation and preprocessing (e.g., loading CSV files, handling missing data).
  - **NumPy**: For numerical computing and handling multidimensional arrays.

- **Data Visualization**:
  - **Matplotlib** and **Seaborn**: For creating visualizations like bar charts, heatmaps, and distribution plots to explore the data and model results.

- **Machine Learning and Model Development**:
  - **Scikit-learn**: For traditional machine learning algorithms (Logistic Regression, Decision Trees, Random Forest, etc.), feature scaling, and evaluation metrics.
  - **TensorFlow** or **Keras** (optional): For more advanced deep learning models if required.

5.2.4. Version Control

To manage the development process, track changes, and collaborate with team members, version control software is essential.

- **Git**: For version control of code and collaboration.
- **GitHub/GitLab**: For hosting the code repository, issue tracking, and project management.

5.3. Data Requirements

The data is critical for building and training the employee attrition model. It should include relevant features for prediction.

5.3.1. Types of Data Required

- **Employee Demographics**: Age, gender, education level, marital status, etc.
- **Job Information**: Job role, department, salary, years at the company, promotions, job satisfaction, etc.
- **Performance Metrics**: Performance rating, hours worked, number of projects, etc.
- **Compensation and Benefits**: Salary, bonus, stock options, etc.
- **Work-Life Balance**: Work hours, flexibility, remote work options, etc.
- **Survey or Exit Data**: Employee engagement scores, feedback, exit interview data (if available).

5.3.2. Data Format

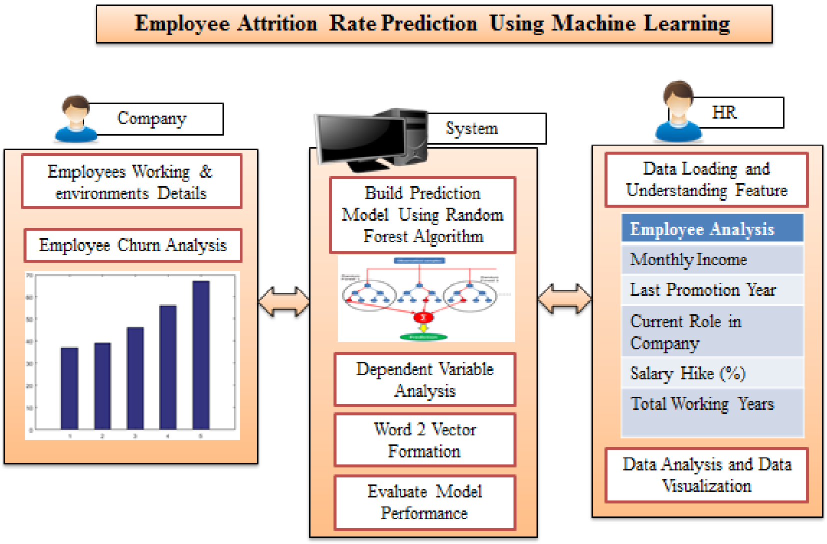
- **Data Source**: Data can be provided in CSV, Excel.
- **File Formats**: CSV, XLSX for loading and preprocessing employee data.
- **Size**: The size of the dataset will influence model performance and hardware requirements.

PROPOSED SYSTEM

Initially the data is downloaded from Kaggle is pre-processed first so that we can extract important features like Monthly Income, Last Promotion Year, Salary Hike and etc. that are quite natural for employee attrition. Dependent variables or Predicted variable are the one that helps to get the factors that mostly dependent on employee related variables. For example, the employee ID or employee count has nothing to do with the attrition rate.

Exploratory Data Analysis is an initial process of analysis, in which you can 0summarize characteristics of data to can predict who, and when an employee will terminate the service. The system builds a prediction model by using random forest technique. It is one of the ensembles learning technique which consists of several decision trees rather than a single decision tree for classification.

The techniques perform dependent variable analysis and word formation vector to evaluate the employee churn. Hence, by improving employee assurance and providing a desirable working environment, we can certainly reduce this problem significantly.


Employee Attrition Rate Prediction Using Machine Learning
Company
Employees Working & environments Details
Employee Churn Analysis
70
60
50
40
30
20
10
0
1
2
3
4
5
System
Build Prediction Model Using Random Forest Algorithm
Dependent Variable Analysis
Word 2 Vector Formation
Evaluate Model Performance
HR
Data Loading and Understanding Feature
Employee Analysis
Monthly Income
Last Promotion Year
Current Role in Company
Salary Hike (%)
Total Working Years
Data Analysis and Data Visualization

DEFINITION

- HR analytics is a methodology that uses statistical tools and techniques to unify and evaluate employees quantitative and qualitative data that helps in bringing out meaningful insights to develop better future decision making.

- HR analytics is an experimental approach that uses software and method based on HR metrics to provide reliable and justifiable human capital results impact effectively and efficiently.

- HR analytics is a data-driven framework that understands and evaluates the relationship between workforce problems and employee's performance by driving new insights through existing insights.

- HR Analytics is HRM innovation enabled by companies to analyses HR data, processes, human capital statically for making data-driven decision making and ignoring the process of gut feeling. This tool helps in making better decisions and testing the effectiveness of the HR department towards business goals. HRIS has provided a way to HR analytics to grow and develop as it includes some limited analytics solutions within its system.

EVOLUTION

- As organizations were raising globally, they have a large amount of data in each function. The Organization uses analytics in all of its functions but HR was lacking behind, so due to difficulty in collecting and analyzing HR data it becomes a necessity for the evolution of HR analytics.

- In 1959 E.T. Rennese explained a theory given by Barney namely "Resource-Based View Theory". This theory explained that to achieve a competitive advantage in an organization there is a need to understand the relationship between HRM and Business strategy. Barney also stated VRIO i.e., Valuable, Rare, Inimitable, and Organized framework which was later criticized by scholars as this theory included only human capital and accordingly human capital cannot create any competitive advantage.

- In the 1970s HRM-related issues were analyzed and how to use HR metrics and scorecards to measure HR data were discussed.

- In 1988 Baird and Meshoulam explained the relationship between three important aspects of any organization i.e., HR policies, organization life cycle, and business challenges within the organization. They also explained vertical and horizontal fit which explains how collaborations of the HR function with other functions and HR sub functions helps in achieving organizational objectives.

- In the 1990s organizations found that to achieve objectives and goals of any organization to create a competitive advantage it is important to value employees and they started viewing their employees as material resources.

- During the first half of the 2000s, various new tools and techniques were introduced to measure the impact of HR activities and practices on organizational performances such as HR scorecards or workforce scorecards. Later in mid of 2000, there was the exposure of HR accounting and utility analysis which later showed a shift towards the development of a more scientific and evidence-based approach to HR.

- In 2002 Oakland A found more advanced perceptive use of metrics and based on this experiment Lewis in 2003 found a concept called "Moneyball Concept" which showed growth on a large scale in 2006. In 2009 GOOGLE worked on finding out the best competent traits that are needed to be an effective manager and doing this Google developed "Project Oxygen" which bought a tremendous shift from traditional HR measurements to HR analytics. Google also highlighted the benefits of using HR analytics in organizational performances.

- Since then, HR analytics has received a certain amount of attention but still, it has not reached its final stage. Researchers on HR analytics recently have been started which mainly focuses on the use of Hr analytics, it was a decision support tool, the capability of this tool, or awareness of HR analytics. HR analytics is developed after the development of big data and now it uses a large amount of HR data to provide the organization with decision making.

EMPLOYEE ATTRIRTION

Employee attrition, also referred to as employee turnover, is a crucial concern for organizations as it impacts workforce stability, operational efficiency, and overall business performance. Attrition occurs when employees leave the company, either voluntarily or involuntarily. In the context of human resources (HR), predicting and managing attrition is essential for maintaining a productive, motivated workforce and reducing the costs associated with recruitment, training, and knowledge loss.

This section delves deeper into the concept of attrition, its types, causes, impact, and the role of machine learning in predicting and mitigating it.

"Employee attrition is defined as employees leaving their organizations for unpredictable or uncontrollable reasons. Many terms make up attrition, the most common being termination, resignation, planned or voluntary retirement, structural changes, long-term illness, layoffs.

Though often used interchangeably, attrition and turnover are not the same."

"Attrition is the reduction in the number of employees through retirement, resignation or death. Attrition rate is the rate of shrinkage in size or number."

9.1 Attrition vs. Turnover vs. Layoffs

9.1.1 Attrition vs. Turnover

Attrition is a passive method of reducing workforces without the need for direct action like layoffs. Turnover, on the other hand, involves the active departure of employees from an organization, which can be either voluntary or involuntary. Voluntary turnover occurs when employees choose to leave, perhaps due to job dissatisfaction, better opportunities elsewhere, or personal reasons. Involuntary turnover happens when an organization restructures or another business-driven need.

Turnover is a critical metric for organizations as it reflects the overall health of the workplace environment and the effectiveness of its retention strategies. The average annual turnover rate is 30% in the United States, with average voluntary turnover at 23% and involuntary turnover at 11%, according to the SHRM Benchmarking: Human Capital Report, 2022.

9.1.2 Attrition vs. Layoffs

Layoffs are a form of involuntary turnover where an organization decides to terminate employees due to economic downturns, business restructuring or the need to reduce operational costs. Unlike attrition, layoffs are a deliberate action taken by an organization and often involve multiple employees. Layoffs can have significant emotional and financial impacts on the affected individuals and may affect the morale and productivity of the remaining workforce.

Attrition	Turnover	Layoffs
Passive reduction of staff	Active departures or reduction of staff	Active reduction to staff
Voluntary or involuntary	Voluntary or involuntary	Involuntary

9.3 Global employee attrition statistics and what they say about the state of the talent market

- January 2022 saw 4.25 million people quit their jobs, up from 3.3 million in 2021: US Bureau of Labor Statistics.

- July 2022 saw 5.9 million total work separations in the US: US Bureau of Labor Statistics.

- Nearly half of the people leaving their current jobs are moving to entirely new industries: McKinsey.

- Annual voluntary turnover is likely to jump 20% in 2022: Gartner

- 57% of surveyed employees report that they are open to new job opportunities within the next year: Future Forum

Labor shortage ranks amongst the top 3 CEO concerns worldwide. 2021 is now commonly referred to in corporate circles as the year of "The Great Resignation." It saw many employees leaving their jobs or looking for other opportunities to improve work-life balance, better work practices, and a better sense of health and well-being.

As 2021 ended, business leaders expected business-as-usual in 2022, especially with movement restrictions being lifted. They hoped employees would cope with the new normal as they returned to offices. But this hasn't been without its fair share of challenges concerning the future of work.
Companies are faced with substantial administrative transformation. As a result, HR leaders and teams have had a lot on their plates, with new demands like hybrid work, improved policies, and better compensation and benefits. Managing attrition in these times of flux is critical.

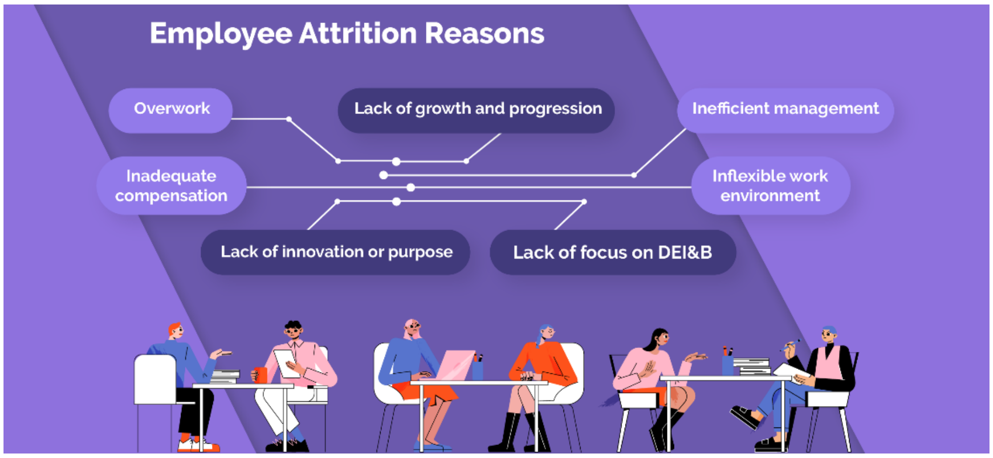
9.4 Employee Attrition: An Expensive Problem

Knowing an organization's attrition rate helps understand how employee-friendly the organization is and what it can do better. A high attrition rate could mean insufficient compensation and benefits, a dated, high-pressure work culture, or inadequate learning and professional development opportunities.

No matter what the reason, employee attrition is an expensive business problem and filling vacant positions costs significant time and money. Effective recruitment includes hiring, onboarding, and training – all of which translate to increased time cost and financial investment. Replacing a full-time employee can cost anywhere between half to twice the employee's annual salary, says a Gallup study.

Add to these hidden costs like the time spent on adjusting to new work culture, building meaningful connections at work, and getting familiar with processes as new employees get up to speed. It is worth remembering, when employees depart, they take away more than just their skills. They leave behind gaps in client relationships, unique perspectives and experiences, and personal and professional networks.

Therefore, it makes sense for organizations to calculate the employee attrition rate and look for ways to save costs, time, and effort.


Employee Attrition Reasons
Overwork
Lack of growth and progression
Inefficient management
Inadequate compensation
Inflexible work environment
Lack of innovation or purpose
Lack of focus on DEI&B

9.5 Calculate Employee Attrition

The formula is explained below:

X: Number of employees who were separated from the organization for any reason
Y: Average number of employees

Attrition rate = (X/Y) x 100

9.6 Employee Attrition Rate

A high employee attrition rate indicates that your employees are not sticking with your company for long, while a low attrition rate means that your company has better employee retention.

Attrition rates vary due to several factors, such as:

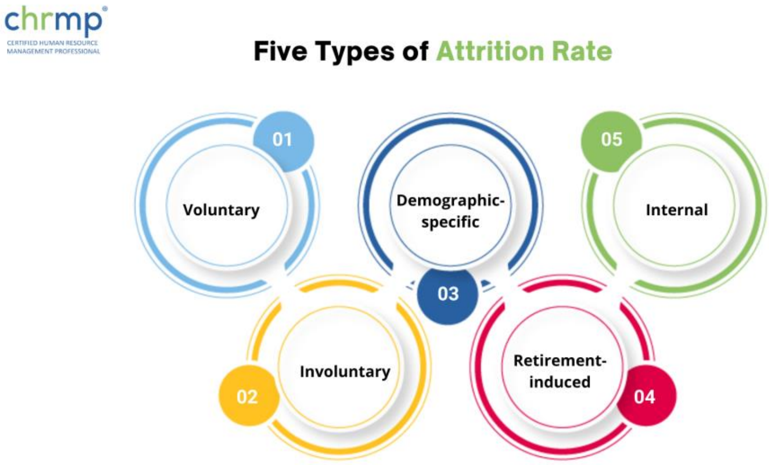
- Size of a company
- Industry niche and
- Location of a company

A high attrition rate in the workplace can occur not just because employees leave but when a business can’t replace them quickly. Employees leaving for any reason, voluntarily or involuntarily, contribute to attrition.

Five Types of Attrition Rate

A high employee attrition rate has become a top concern for companies today. In many instances, employees leaving the company faster than they are hired is out of the manager's control.

But before we get into more details about attrition rate, let's look at its five types.



Let's discuss these types in more detail.

1. **Voluntary Attrition:**

Voluntary attrition is the most common type in which employees quit their current job to look for greener pastures elsewhere. There can be several reasons why this happens, but the most common ones include the following:

- Incompetent benefits package
- Hostile work environment
- Lack of career progression opportunities
- Lack of a healthy work/life balance or flexibility

As evident, many of the reasons for voluntary attrition are under your control. Proactively curbing voluntary attrition among highly talented employees will go a long way in building a robust and success-driven workforce.

2. **Involuntary Attrition:**

In this type of attrition, it is the organization that decides to lay off employees instead of employees leaving on their own accord. Factors such as mergers and acquisitions, structural reform, reduction in workforce or job position elimination are some of the reasons for involuntary attrition.
The company might sometimes dismiss an employee due to workplace misconduct. Keeping such employees might reflect poorly on the company's image. Hence, they are asked to leave.

3. **Demographic-specific Attrition:**

Demographic-specific attrition indicates a specific group of employees: women, ethnic minorities, disabled people, veterans, or older professionals are leaving the company in large numbers.
This type of attrition is of particular concern for companies striving to create equal workplace opportunities. If you can identify a pattern of attrition among a specific demographic group, there is most likely a serious reason behind it.

4. **Internal Attrition:**

In this type of attrition, employees within a department in a company switch their positions to gain a role in other departments. Even though internal attrition can redirect strong talent to vital functions in the company, it can also leave struggling departments even weaker.
If a department faces a high attrition rate, there might be several underlying causes, such as an inadequately skilled manager, failure to meet targets or an inconducive work environment.

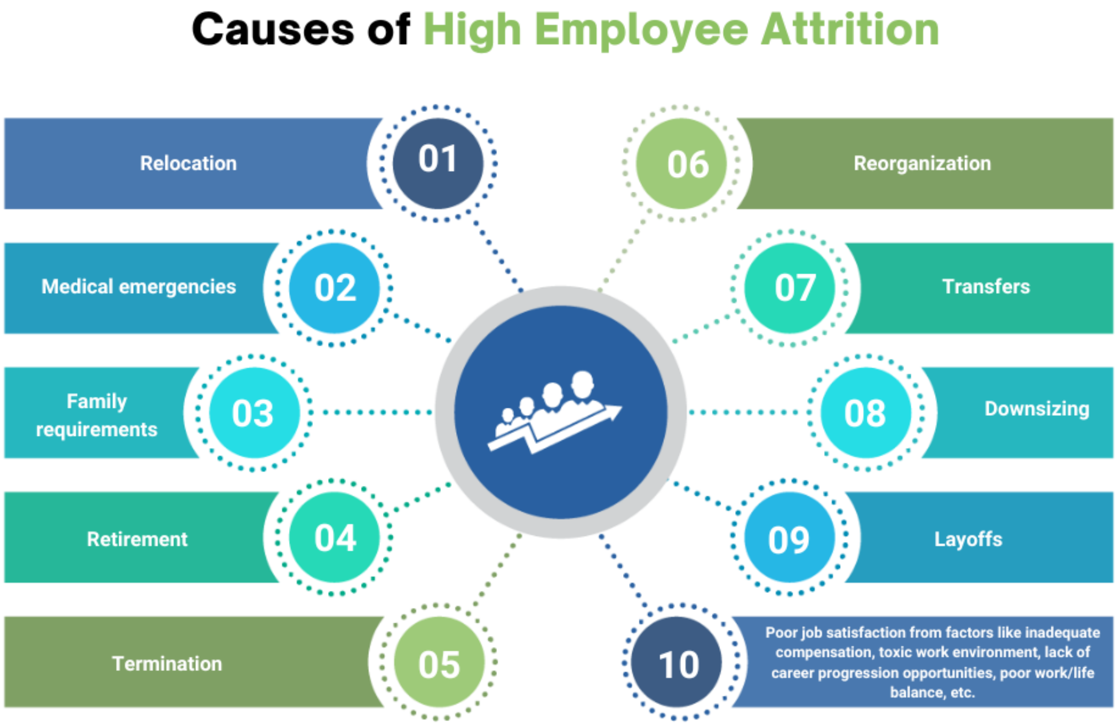
5. **Retirement Induced Attrition:**

It is natural for employees to retire and quit their jobs towards the end of their professional careers. Thus, it is statistically too insignificant to count under attrition in small groups.

9.7 Causes of High Employee Attrition

There can be several reasons for a high employee attrition rate, many of which might be outside your control. For example, employees might leave for personal reasons like family requirements or medical emergencies. The company itself induces other causes of attrition, like transfers, downsizing, organizational restructuring or termination**.**

But several causes of attrition can also be well under your control.


Causes of High Employee Attrition
01
Relocation
02
Medical emergencies
03
Family requirements
04
Retirement
05
Termination
06
Reorganization
07
Transfers
08
Downsizing
09
Layoffs
10
Poor job satisfaction from factors like inadequate compensation, toxic work environment, lack of career progression opportunities, poor work/life balance, etc.

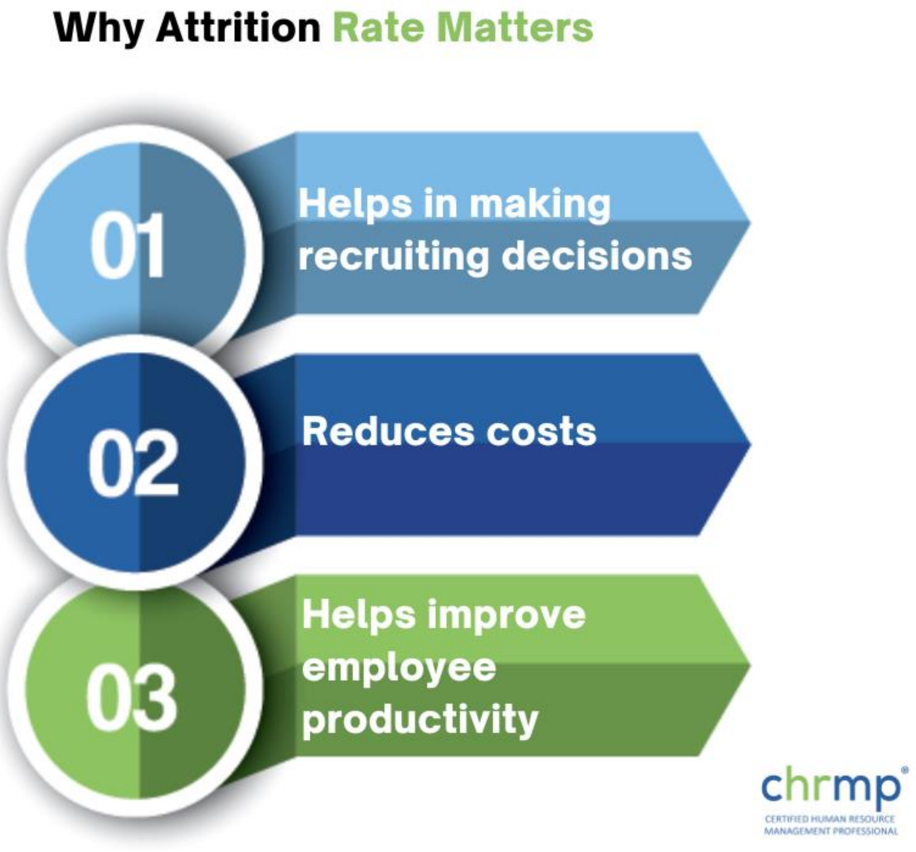
A common reason for employee attrition is poor job satisfaction; many other factors influence that. Pay is directly linked to job satisfaction, but this doesn't mean employee satisfaction is just about the monthly money they take home. Other factors that add to a poor sense of job satisfaction include a lack of other financial benefits like bonuses and annual increments, a lack of career progression opportunities, a toxic work environment or poor work/life balance.

9.8 The Importance of Understanding Attrition Rates

Understanding the attrition rate in your company can give you valuable insights into employee quitting trends and patterns, help identify what causes these patterns and eventually find a solution.

If attrition rates spike due to issues within the company, like inadequate pay and benefits or low job satisfaction, managers can focus on eliminating them.

Here are some critical reasons to under attrition rate in your company.


Why Attrition Rate Matters
01
Helps in making recruiting decisions
02
Reduces costs
03
Helps improve employee productivity
chrmp
CERTIFIED HUMAN RESOURCE MANAGEMENT PROFESSIONAL

1. Helps in Making Recruiting Decisions

Understanding attrition and its causes in your company will help you better understand what employees expect from the company, resulting in a better recruitment process. It will also help you design better compensation and benefits packages to attract and retain highly talented candidates.

2. **Reduces Costs**

Hiring new employees can cost the company a lot of money. But suppose you have a good understanding of your company's attrition rate. In that case, you'll have a better idea of how many employees need to be hired within a specific period, which will prevent disruption in the workflow due to unfilled positions.

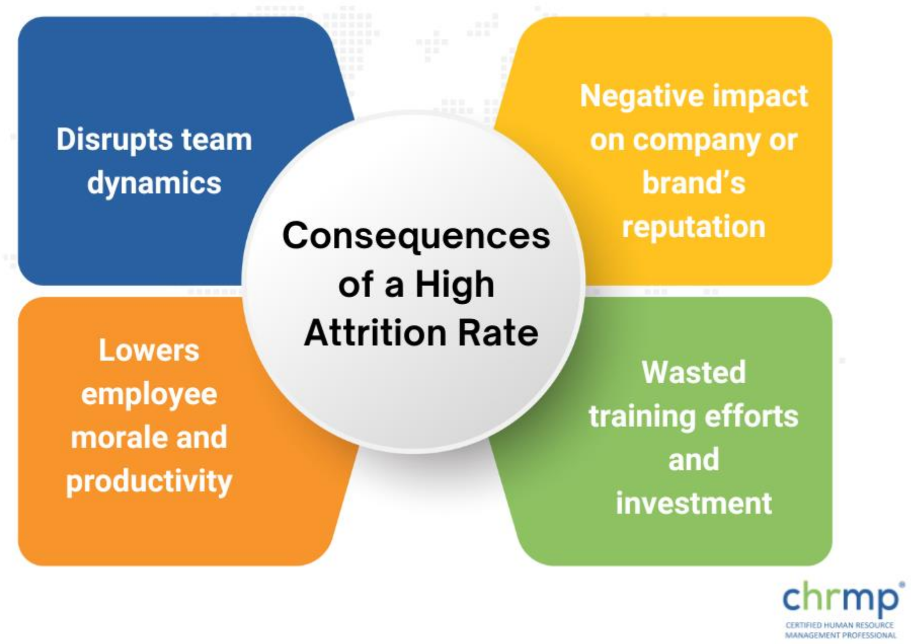
3. **Helps Improve Employee Productivity**

There are several ways in which understanding your company's attrition rate can help you increase employee productivity, like reducing the time spent on training recruits. Another way to understand attrition can help identify and eliminate feelings of disgruntlement and low morale among employees due to underlying issues such as poor management or a lack of recognition, which boosts the overall productivity levels of the workforce.

9.9 What are the Consequences of High Attrition? How to Resolve It?

The success of a company largely depends on unhindered contributions from its employees. It is vital to retain employees in whom you have invested precious time and resources. However, due to the many reasons discussed earlier, several employees leave the company to look for other options that can have far-reaching consequences.

A high attrition rate can heavily impact your company’s reputation, which indicates several unattended issues with your company. According to research, 50% of talented employees prefer not to work with a company with a negative reputation even when offered a higher salary.

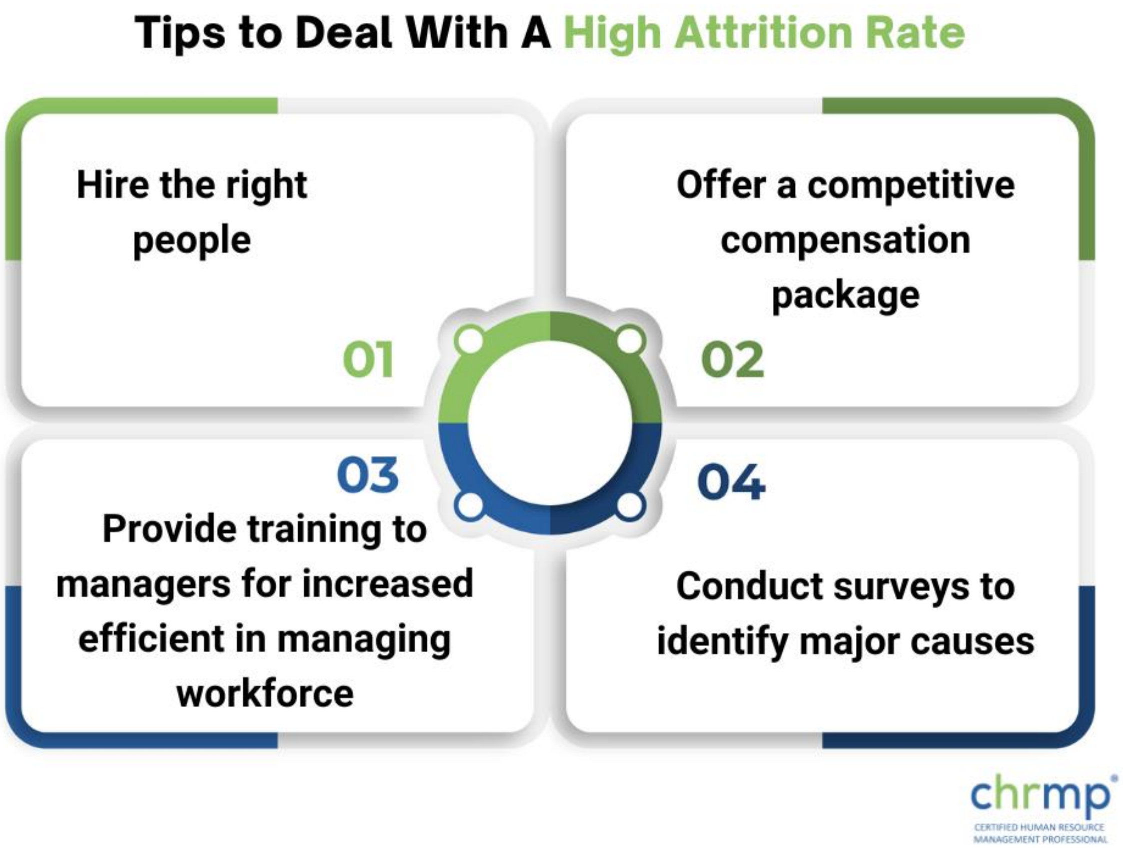

Consequences of a High Attrition Rate
Disrupts team dynamics
Negative impact on company or brand’s reputation
Lowers employee morale and productivity
Wasted training efforts and investment
chrmp
CERTIFIED HUMAN RESOURCE MANAGEMENT PROFESSIONAL

The company’s progress is also negatively impacted due to a high attrition rate. How? Let’s assume that the average time for a new employee to settle and start adding to the profits is seven to eight months. That being said, if you constantly keep hiring new talents to replace leaving employees, a disruption in the workflow is inevitable, leading to a lag in your company’s progress.

To top it all, a high attrition rate results in a waste of training efforts and investment. It's not just the company that gets impacted due to a high attrition rate; it affects the employees too. When employees see several of their colleagues leave the company, it takes a heavy toll on their morale and productivity.

9.10 So how do you deal with it?

There are several ways to manage your company's attrition rate and prevent it from spiking. Some steps that you can take are:



Tips to Deal With A High Attrition Rate

01 Hire the right people

02 Offer a competitive compensation package

03 Provide training to managers for increased efficient in managing workforce

04 Conduct surveys to identify major causes

- **Hire the right people:**

  hiring suitable candidates that fit your organization's needs will eliminate unnecessary additional hiring costs, build a quality resource pool, and avoid future problems.

- **Offer a competitive compensation package:**

  if an unsatisfactory pay check is causing your employees to leave in droves, it would be best to re-evaluate the compensation package and improve it to offer more incentives.

- **Provide training to managers for increased efficiency in managing the workforce.**

- **Conduct surveys to identify the major causes of a high employee attrition rate:**

  find out what is causing your employees to leave and devise solutions accordingly.

How people analytics can improve your employee attrition rates

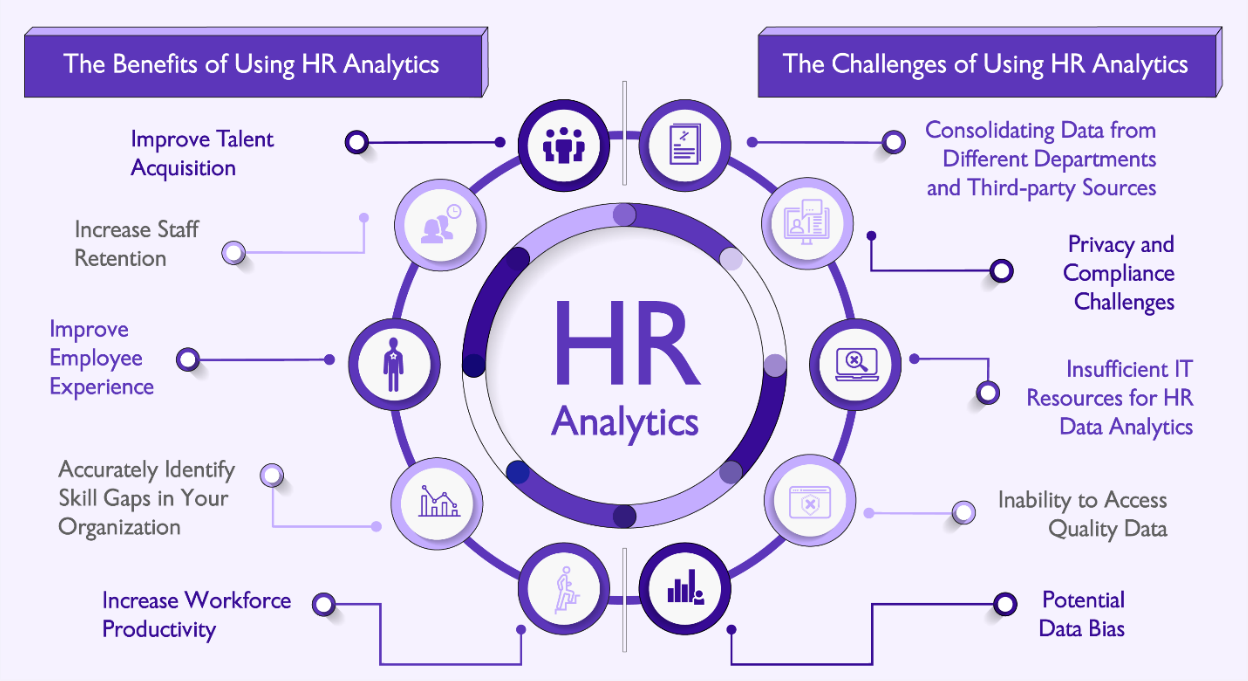
People analytics solutions can make the most of the data and insights by first defining the parameters of attrition and retention –

- who is leaving
- why and when are they leaving

Your raw data is often siloed into pockets of information across multiple HR tools. People analytics solutions will blend, visualize, and analyze ALL your workforce data in easy-to-use dashboards and reports, meaning HR leaders and organizations are able break the silos and find meaning in their data sets. Eventually, these insights help them take positive action in employee attraction, engagement, productivity, retention, and attrition.

HR ANALYTICS

HR analytics is the gathering, analyzing and reporting of data that surrounds the management of human resources. It is the method of getting a better understanding of the people within an organization and how well the human resources team is performing. The analysis of this data can be a huge help to giving an organization the right direction to move forward in order to maximize payroll, benefits, its ability to hire or keep employees and more.



The Benefits of Using HR Analytics
The Challenges of Using HR Analytics
HR
Analytics
Improve Talent Acquisition
Increase Staff Retention
Improve Employee Experience
Accurately Identify Skill Gaps in Your Organization
Increase Workforce Productivity
Consolidating Data from Different Departments and Third-party Sources
Privacy and Compliance Challenges
Insufficient IT Resources for HR Data Analytics
Inability to Access Quality Data
Potential Data Bias

10.1 Demystifying HR Analytics

HR analytics techniques involve the systematic collection, analysis, and utilization of data from various human resource activities before making informed decisions that enhance overall business efficiencies.

The sophisticated HR analytics tools of today offer incredible new features and functionalities that help organizations and HR professionals unlock significant value. At the same time, getting the hang of these tools isn't that easy and involves a sizable learning curve, but in this day and age, they are increasingly becoming indispensable.

Organizations need to learn how to use HR analytics to understand the massive amounts of internal data being generated and drive strategic interventions where required.

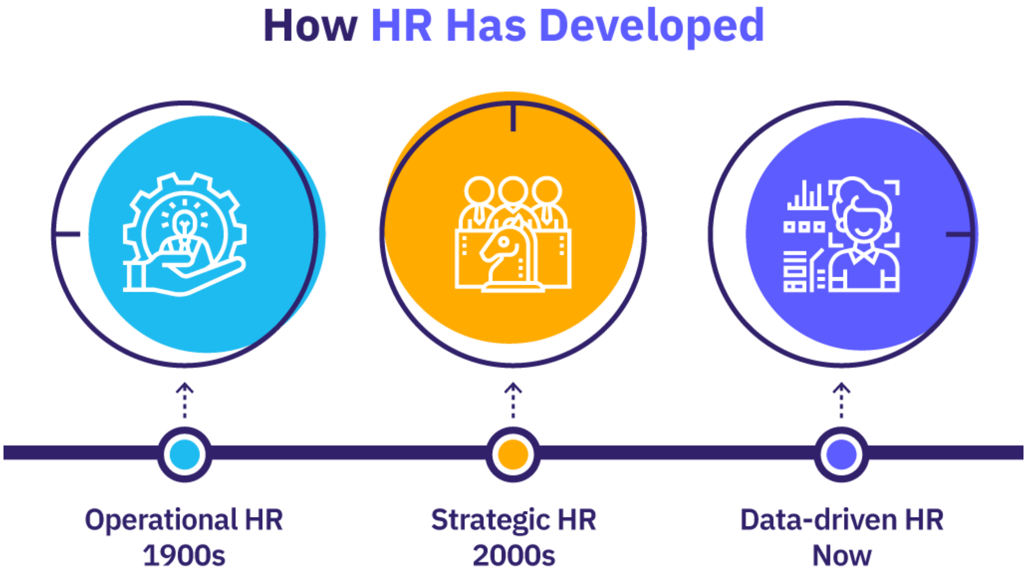
10.2 HR analytics

HR analytics, also referred to as people analytics or workforce analytics, involves gathering, analyzing, and reporting HR data to drive business results. It enables your organization to better understand your workforce, make decisions based on data, and measure the impact of a range of HR metrics, ultimately improving overall business performance. In other words, HR analytics is a data-driven approach to Human Resources Management.

Although the term "HR analytics" is widely used, there is a growing trend of referencing "people analytics" as well. The two may often be used interchangeably, but technically there is a subtle difference. HR analytics originates from data housed within Human Resources and is aimed at optimizing HR functions. People analytics expands beyond HR to incorporate data from other sources within the organization, such as marketing, finance, and customer statistics, to address a wider scope of business issues.

In the past century, Human Resource Management has made a dramatic shift from an operational discipline to a more strategic one. The popularity of using the phrase Strategic Human Resource Management exemplifies this. The data-driven approach that characterizes HR analytics is in line with this development.

Analytics enables HR professionals to make data-driven decisions instead of relying solely on instinct and opinions. Furthermore, analytics helps test the effectiveness of HR policies and interventions.

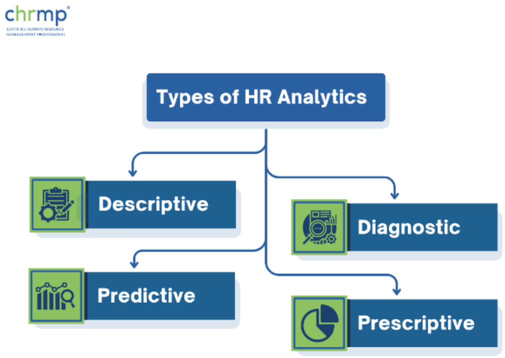


10.2.1 Types of HR Analytics

HR analytics is a rapidly growing field that has become increasingly important for organizations of all sizes. By using HR analytics, organizations can gain valuable insights into their workforce and make data-driven decisions to optimize their talent management strategies.

There are four types of HR analytics: descriptive, diagnostic, predictive, and prescriptive analytics. Each analytics type has its unique purpose and can help HR professionals address specific workforce issues.

Here are the four types of HR analytics described in detail:



1. Descriptive Analytics

Descriptive analytics is a type of HR analytics that involves analyzing historical data to gain an understanding of what had happened in the past.
It summarises data that helps identify patterns and trends, such as employee turnover rates, absenteeism, or workforce demographics.
Descriptive analytics is an important tool for HR professionals to help them make sense of large amounts of data collected over the past years and identify areas of improvement.

Using descriptive analytics, HR professionals can answer questions such as: How many employees were hired last year? What was the average salary for a specific job role? How many employees left the organization and their absentieeism rate?

This information can be used to develop insights and identify areas where HR improvement or process optimization can occur.

Descriptive analytics provides a foundation for more advanced types of analytics, such as predictive extrapolative and prescriptive analytics, that can help HR professionals anticipate future trends and develop strategies to optimize their workforce performance.

2. Diagnostic Analytics

Diagnostic analytics is HR analytics that goes beyond the descriptive analysis of past events to identify the root cause of workforce problems or issues. It involves analyzing and extrapolating data to determine why certain trends or patterns are occurring in the workforce data. By examining historical data, diagnostic analytics can help HR professionals understand why certain events have occurred in the past years and what factors have contributed to their occurrence.

For example, if an organization is experiencing high turnover rates, diagnostic analytics helps them identify the problem's underlying causes. It reveals whether the turnover is related to certain departments or job roles, and whether it's due to poor management, lack of career development opportunities, or inadequate compensation and welfare amenities. Once the underlying causes are identified, HR professionals can address the issue and develop effective solutions to improve employee retention and engagement while curtailing sabbaticals and absenteeism.

Diagnostic analytics is a valuable tool for HR professionals which helps them identify and address workforce issues before they become more serious problems. Using diagnostic analytics, HR professionals can improve employee engagement and retention, leading to a more productive and successful organizational workforce.

3. Predictive Analytics

Predictive analytics is a type of HR analytics that uses statistical algorithms, extrapolative methods and machine learning techniques to analyze historical data and predict future outcomes. It involves identifying patterns and trends in workforce data, then extrapolating using that information, to make predictions about future workforce behavior.

Predictive analytics can help HR professionals anticipate future workforce trends, such as employee turnover or skills gaps, and develop strategies to address them before they become major issues. Unless some major disruptive event takes place in the process, such extrapolative calculated outcomes generally hold true.

For example, predictive analytics can be used to develop models that predict which employees are most likely to leave the organization in the next year.
This information can be used to proactively identify and address the underlying causes of employee turnover, be it satisfactory compensation, lack of welfare amenities, growth opportunities etc. before it becomes a more serious problem.

Similarly, predictive analytics can be used to identify which employees are most likely to be promoted, providing insights into where to invest in training and development programs.

Predictive analytics is a powerful tool among the different types of HR analytics for HR professionals, allowing them to make data-driven decisions and develop strategies based on accurate data-predictions of future outcomes.

4. Prescriptive Analytics

Prescriptive analytics is a type of HR analytics that works using data, algorithms, and machine learning techniques to recommend actions that HR professionals can take to optimize their workforce and curb negative phenomena involving the workforce from taking root.

It goes beyond predictive analytics, which predicts what might happen, to suggest what should be done to prevent it from occurring.

Prescriptive analytics uses statistical models to analyze data and recommend specific courses of action. It's similar to a doctor's prescription that gives preventive medicine to prevent some particular ailment from afflicting, metaphorically speaking.

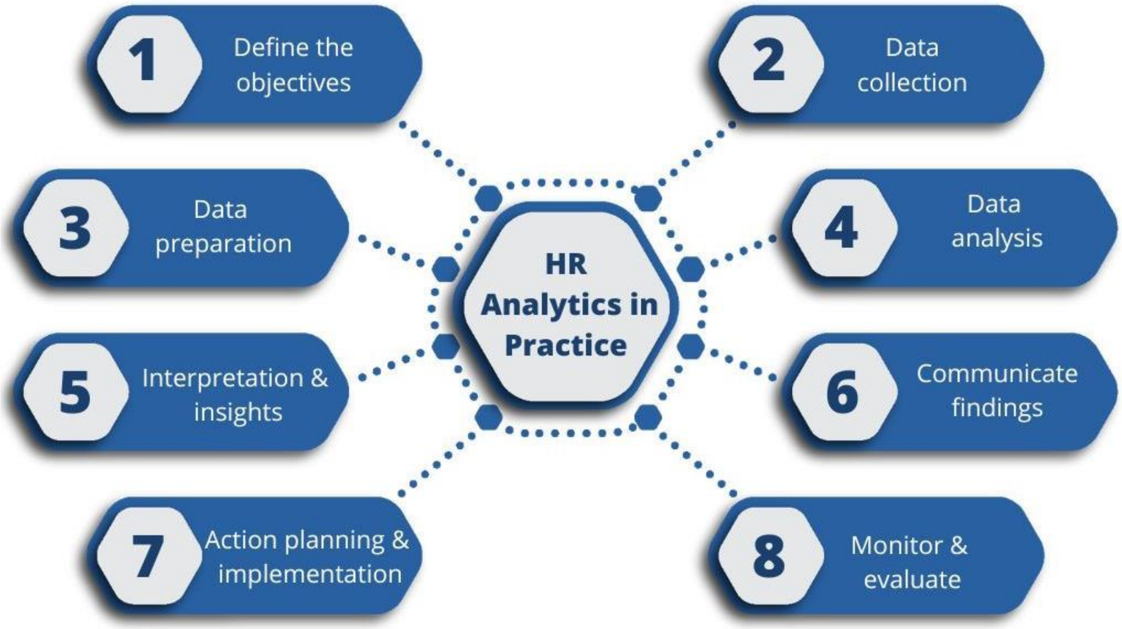
For example, suppose a company is beginning to experience high employee turnover rates. In that case, prescriptive analytics can suggest specific strategies for addressing the problem so that employee turnover rates may not rise.

It may recommend increasing employee engagement, improving training and development programs, or providing better compensation and benefits. By using prescriptive analytics, HR professionals can take proactive steps to optimize their workforce and achieve their business goals.

10.3 HR Analytics in Practice

HR analytics, also known as people analytics, is the process of using data and analytical methods to gain insights and make informed decisions about human resources within an organization. It involves collecting and analyzing data related to employees’ performance, engagement, recruitment, retention, training, and other HR processes.

In practice, HR analytics involves several steps and considerations:



1.Define the objectives: Clearly identify the goals and objectives you want to achieve through HR analytics. This could include improving employee productivity, reducing turnover, enhancing recruitment processes, or identifying skills gaps.

2. Data collection: Gather relevant data from various sources such as HR systems, performance management software, employee surveys, and external sources. This data may include employee demographics, performance metrics, training records, engagement surveys, and more.

3. Data preparation: Clean and organize the collected data, ensuring it is accurate and reliable. This may involve data cleaning, data integration, and data transformation to make it suitable for analysis.

4. Data analysis: Apply analytical techniques to uncover patterns, correlations, and insights within the data. This could involve statistical analysis, data visualization, and predictive modeling. Examples of analysis might include identifying factors that contribute to high employee turnover, predicting future workforce needs, or determining the impact of training programs on performance.

5. Interpretation and insights: Analyze the results of the data analysis and extract meaningful insights. These insights can help HR professionals make data-driven decisions and recommendations to improve HR processes and outcomes.

6. Communicate findings: Present the findings and insights in a clear and concise manner to stakeholders, such as HR managers, senior leaders, and department heads. Use visualizations and storytelling techniques to effectively communicate the results and their implications.

7. Action planning and implementation: Develop action plans based on the insights gained from the analytics. Collaborate with relevant stakeholders to implement changes or interventions based on the findings. This could involve modifying recruitment strategies, redesigning training programs, or adjusting performance management processes.

8. Monitor and evaluate: Continuously track the impact of the implemented changes and monitor key HR metrics. Evaluate the effectiveness of the interventions and make adjustments as needed.

It’s important to note that HR analytics is an ongoing process, and organizations should develop a culture of data-driven decision-making to maximize the benefits. Additionally, it’s crucial to ensure data privacy and compliance with relevant regulations when collecting and analyzing employee data.

10.4 Top 3 Benefits of HR Analytics

HR analytics offers several benefits to organizations. Here are three key advantages of utilizing HR analytics:

1.Data-driven decision-making:

HR analytics enables organizations to make informed decisions based on data and insights rather than relying solely on intuition or subjective judgments. By analyzing HR data, organizations can identify patterns, trends, and correlations that provide valuable insights into various HR processes and outcomes.

2. Improved recruitment and retention:

HR analytics can significantly enhance recruitment and retention efforts. By analyzing data related to recruitment processes, organizations can identify which sourcing channels, assessment methods, or selection criteria are most effective in attracting and hiring high-performing candidates.

3. Enhanced workforce planning and performance management:

HR analytics provides insights into workforce trends, skill gaps, and performance metrics, allowing organizations to proactively plan for their future workforce needs.
These benefits of HR analytics contribute to more effective HR management, alignment of HR practices with organizational goals, and ultimately, better business outcomes.

10.5 Key HR Analytics Metrics

There are a number of HR analytics that a business can measure, but the right ones for you will depend on what you're wanting to learn and accomplish. The key HR analytics are ones that are typically measured by most organized businesses looking to keep track of their people data. Here is an overview of those key metrics that make a good starting point for most businesses to launch an HR analytics program.

1. Revenue per Employee

Revenue per employee measures how much money the business is bringing in for every employee it has on staff and is paying expenses, such as salary and benefits, for. It is calculated by dividing a company's revenue by the total number of employees in the company. Businesses love to track this because it provides a way to see how efficient businesses are at generating revenue for each new hire.

Example: If a business has 100 employees and brings in $10 million in revenue, its revenue per employee would be $100,000.

2. Time To Fill

The time to fill metric measures how long it takes to fill an open position at the company. It is calculated by counting the number of days from posting the job to someone accepting an offer. This gives good insight into how efficient the hiring team is at finding good candidates and moving them through the hiring process.

Example: If a company posts a job on March 1 and completes its interviewing process, makes an offer, and gets that offer accepted on April 20, then the time to hire would be 51 days.

3. Voluntary and Involuntary Turnover Rates

These rates measure the percentage of employees who end up leaving the company. The voluntary rate calculates the percentage of employees who decided to leave the company while the involuntary rate calculates the percentage of employees who end up getting fired.

While the voluntary rate measures how well the company is at retaining employees, the involuntary rate measures how well it is at hiring the right people and managing them efficiently. Both are calculated by dividing the number of employees who fall into each category by the total number of employees in the organization.

Example: If 10 employees were fired in the last year, out of the 100 total employees the company had, then the involuntary turnover rate would be 10% of employees.

4. Offer Acceptance Rate

The offer acceptance rate is another hiring metric that measures how well the hiring team is at convincing the people they want to take the job. If a company is making offers to people who are declining those offers at a high rate, then the hiring process likely needs to be adjusted to move candidates through the hiring pipeline who are more interested in joining the company. It is calculated by dividing the number of accepted formal job offers by the total number of jobs offers made.

Example: If the hiring team has received 10 formal job offer acceptances this year, out of 20 given out, then the offer acceptance rate would be 50%.

5. Retention Rate

In contrast to the turnover rate above, it can be important to see how well the business does at keeping employees working for the business. This can be measured company-wide or on a per-manager level. To calculate the retention rate, you can divide the total number of employees who decided to stay employed over a given time period by the total number of employees over that same time period.

Example: If a business had 100 employees in the last year and 85 decided to remain employed, the retention rate would be 85%.

6. Absence Rate

The absence rate is the total number of days an employee is absent from work, not including approved time off such as vacation, over a specific period of time. This is also referred to as absenteeism and is important to measure in positions where individuals call out of work at a high rate, such as retail businesses. It is calculated by dividing the number of days worked by the total number of days that the employee could have worked over a specific period of time.

Example: When measuring the absence rate for June, let’s say there are 20 possible work days. Our worker, John, worked 14 of those days and was on vacation for another three days. This means he worked 14 out of a possible 17 days. That means he worked about 82% of the time and it gives him an absence rate of about 18%.

SYSTEM ARCHITECTURE

11.1 PROPOSED ARCHITECTURE (BLOCK DIAGRAM)

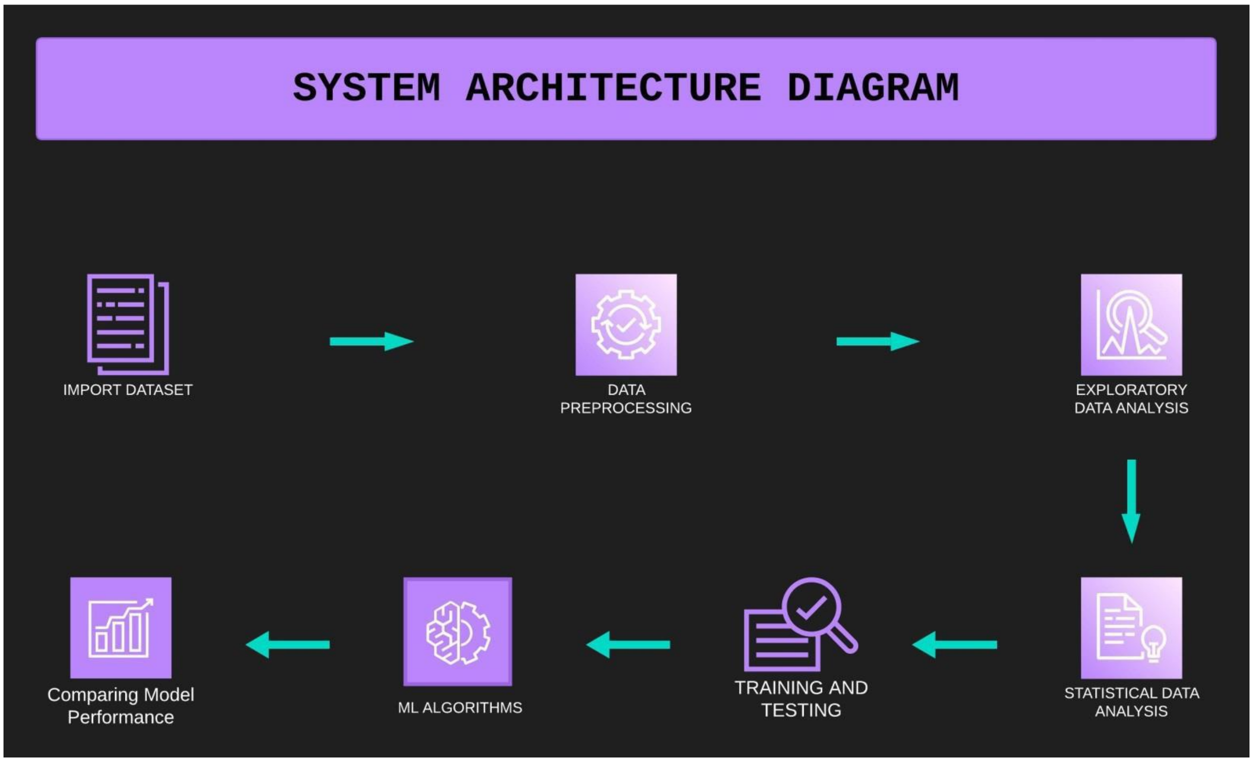

SYSTEM ARCHITECTURE DIAGRAM
IMPORT DATASET
DATA PREPROCESSING
EXPLORATORY DATA ANALYSIS
STATISTICAL DATA ANALYSIS
TRAINING AND TESTING
ML ALGORITHMS
Comparing Model Performance

Fig.11.1 Proposed Architecture

The methodology for IBM HR Analytics Employee Attrition and Performance Prediction is as follows: -

Input is taken by loading the ODIR dataset, which contains ocular

1. Load the Dataset:

The IBM HR Analytics Attrition Dataset is loaded using the pd.read_csv() function. The head () and info () methods are used to display the first few rows and get information about the dataset, respectively.

2. Knowing the Dataset:

Basic Information about the dataset is generated; numerical and categorical attributes are enlisted.

3. Data Cleaning:

Any missing values in the dataset are dropped using the dropna() method.

4. Data Visualization:

Matplotlib and Seaborn libraries are used to visualize the data.

5. Statistical Analysis:

The ANOVA Test is performed to analyze the Numerical Features' Importance in Employee Attrition, while the Chi-Square Test to Analyze the Categorical Feature Importance in Employee Attrition.

6. Data Preprocessing:

The target variable 'Attrition' is mapped to binary values (1 for 'Yes' and 0 for 'No'). Selected features are extracted from the dataset and one-hot encoded using the get_dummies() function.

7. Splitting the Dataset:

The dataset is split into training and testing sets using the train_test_split() method from scikit-learn.

8. Implementing Machine Learning Algorithms:

Logistic Regression, XGBoost, CatBoost, AdaBoost, LightGBM, Decision Tree, and Random Forest classifiers are initialized and trained using the training data.

9. Model Evaluation:

The accuracy score and confusion matrix are computed to evaluate the performance of each algorithm on the testing data.

10. Results:

The results, including the accuracy and confusion matrix, are printed for each algorithm.

11. Model Performance Comparison:

The hvPlot library is used to visualize the ROC curve diagram comparing the performance of all models used.

DATASET

This data set presents an employee survey from IBM, indicating if there is attrition or not. The data set contains approximately 1500 entries. Given the limited size of the data set, the model should only be expected to provide modest improvement in identification of attrition vs. a random allocation of probability of attrition. IBM has gathered information on employee satisfaction, income, seniority and some demographics. It includes the data of 1470 employees. To use a matrix structure, we changed the model to reflect the following data.

Dataset Description:

During website session, browsing information about visited pages is collected and features are extracted as follows:

Table 12.1 – Numerical features used in the user attrition analysis model

Feature Name	Feature Description	Min Value	Max Value	Std. Dev
Age	Age of employee	18	60	9.13
DailyRate	It is the billing cost for an individual's services for a single day	102	1499	403.50
DistanceFromHome	It is the distance between company and home of the employee	1	29	8.10
Education	Education qualification of the employees of company	1	5	1.02
EmployeeCount	Count of employee	1	1	0.0
EmployeeNumber	It is a unique number that has been assigned to each current and former employee	1	2068	602.02

EnvironmentSatisfaction	It is all about an individual's feelings about the work environment and organization culture.	1	4	1.09
HourlyRate	The amount of money that is paid to an employee for every hour worked	30	100	20.32
JobInvolvement	Job involvement refers to the degree to which a job is central to a person's identity.	1	4	0.71
JobLevel	Job levels are categories of authority in an organization.	1	5	1.10
JobSatisfaction	Job satisfaction happens when an employee feels he or she is having job stability.	1	4	1.10
MonthlyIncome	Gross monthly income is the amount of income an employee earn in one month.	1009	19999	4707.95
MonthlyRate	If a monthly rate is set, employees should be paid in exchange for normal hours of work of a full-time worker.	2094	26999	7117.78
NumCompaniesWorked	Number of other companies the employee previously worked for	0	9	2.49
PercentSalaryHike	The amount a salary is increased of an employee in percentage	11	25	3.65
PerformanceRating	Rating means gauging and comparing the performance.	3	4	0.36
RelationshipSatisfaction	It is the rate of satisfaction between Employer employee relationship.	1	4	1.08

Table 12.2 - Shows the numerical features along with their statistical parameters.

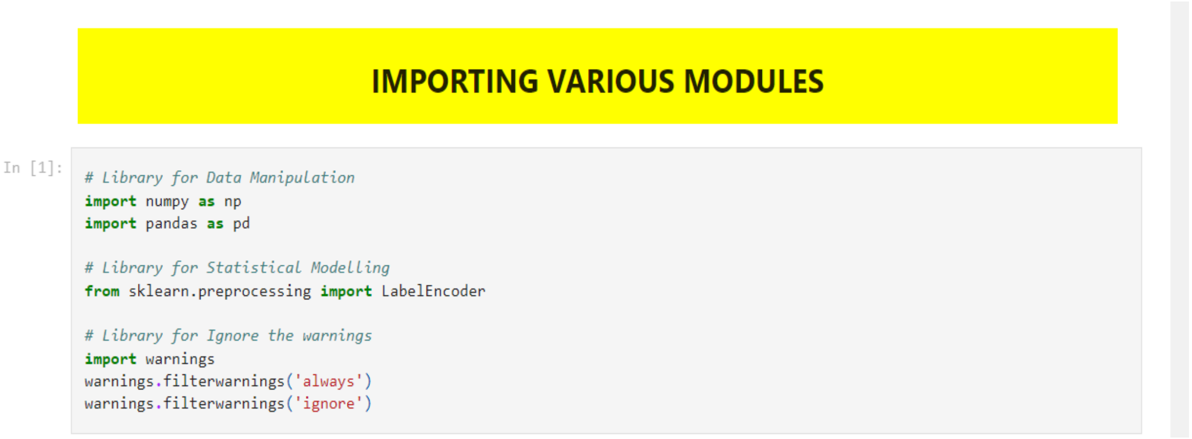
2 – Categorical Features used in the User Attrition Analysis Model.

Feature Name	**Feature Description**	**Number of Categorical Values**
Attrition	Attrition in business describes a gradual but deliberate reduction in staff numbers that occurs as employees retire or resign, [NOTE: Target Variable] (0=no, 1=yes)	2
BusinessTravel	Business travel is travel undertaken for work or business purposes, as opposed to other types of travel (1=No Travel, 2=Travel Frequently, 3=Travel Rarely)	3
Department	Consists three departments that contribute to the company's overall mission. (1=HR, 2=R&D, 3=Sales)	3
EducationField	Education field of the employees(1=HR, 2=Life Sciences, 3=Marketing, 4=Medical Sciences, 5=others, 6= Technical)	6
Gender	Gender of the employee (1=Female, 2=Male)	2
JobRole	These refer to the specific activities or work that the employee will perform. (1=HC Rep, 2=HR, 3=Lab Technician, 4=manager, 5= Managing Director, 6= Research Director, 7= Research Scientist, 8=sales Executive, 9= Sales Representative)	9
MaritalStatus	Marital Status of the employee (1=divorced, 2=married, 3=single)	3
Over18	(1=Yes, 2=No)	2
Overtime	(1=No, 2=Yes)	2

Table 12.3 Shows the categorical features along with their number of categories.

IMPLEMENTATION

13.1 DATA PREPERATION

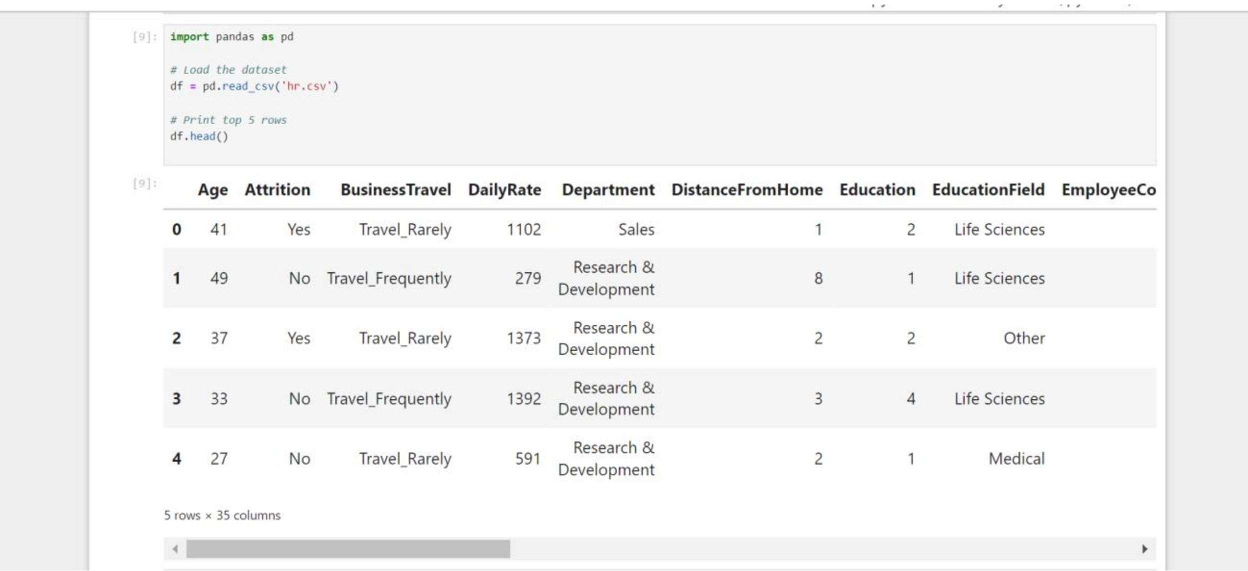


IMPORTING VARIOUS MODULES

In [1]:

```
# Library for Data Manipulation
import numpy as np
import pandas as pd

# Library for Statistical Modelling
from sklearn.preprocessing import LabelEncoder

# Library for Ignore the warnings
import warnings
warnings.filterwarnings('always')
warnings.filterwarnings('ignore')
```

LOADING DATASET



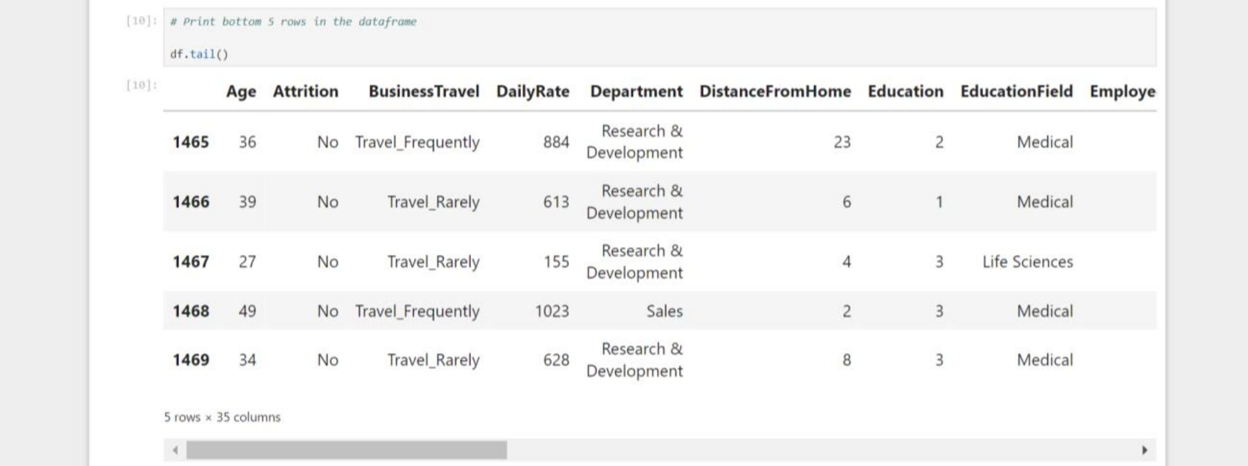
[9]:

```
import pandas as pd

# Load the dataset
df = pd.read_csv('hr.csv')

# Print top 5 rows
df.head()
```

[9]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCo
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	

5 rows × 35 columns



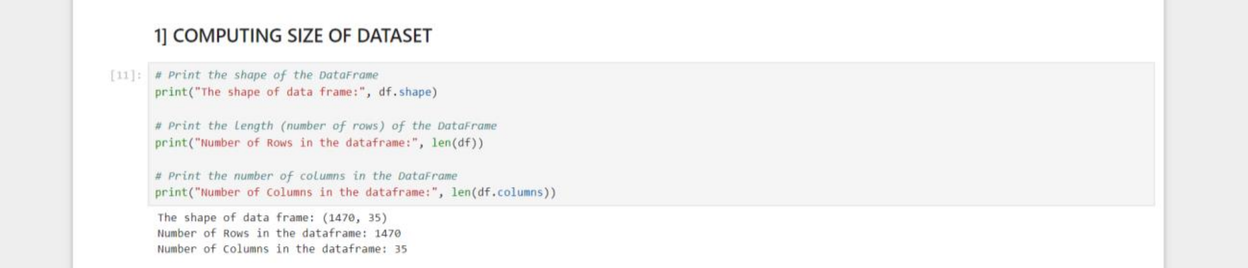
```
[10]: # Print bottom 5 rows in the dataframe

      df.tail()
```

[10]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Employe
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	

5 rows × 35 columns

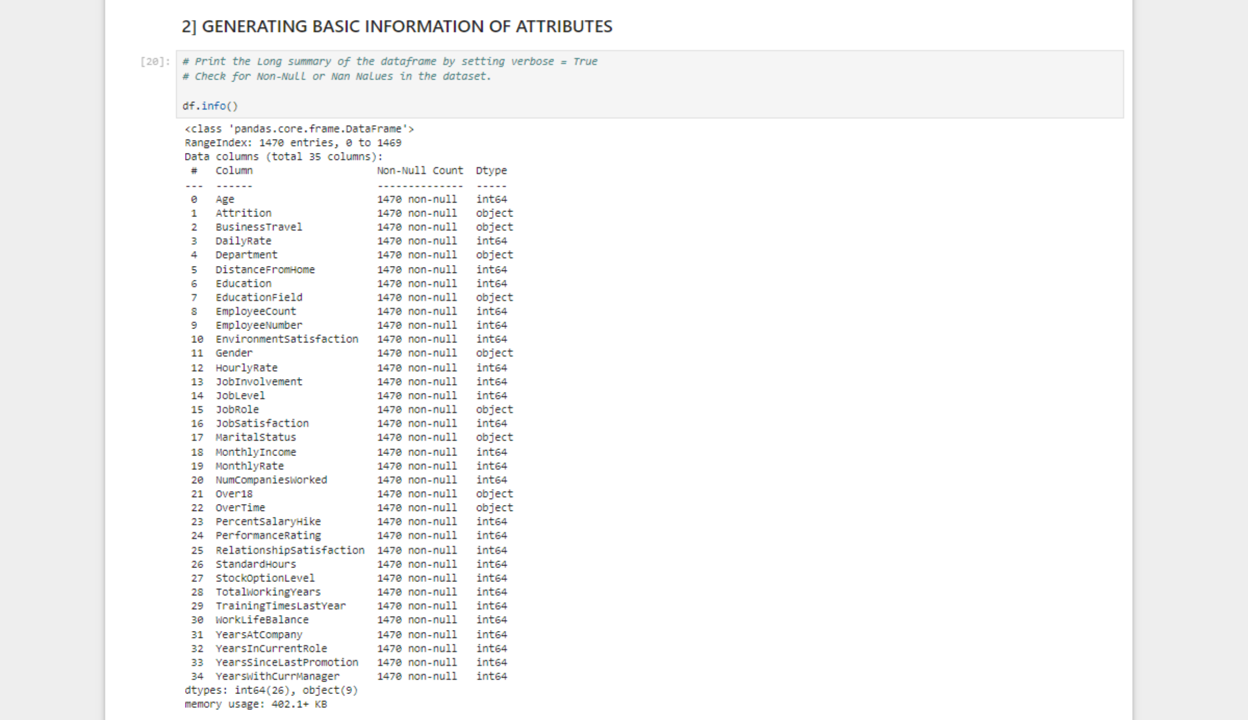
DATA WRANGLING



1] COMPUTING SIZE OF DATASET

```
[11]: # Print the shape of the DataFrame
      print("The shape of data frame:", df.shape)

      # Print the length (number of rows) of the DataFrame
      print("Number of Rows in the dataframe:", len(df))

      # Print the number of columns in the DataFrame
      print("Number of Columns in the dataframe:", len(df.columns))
```

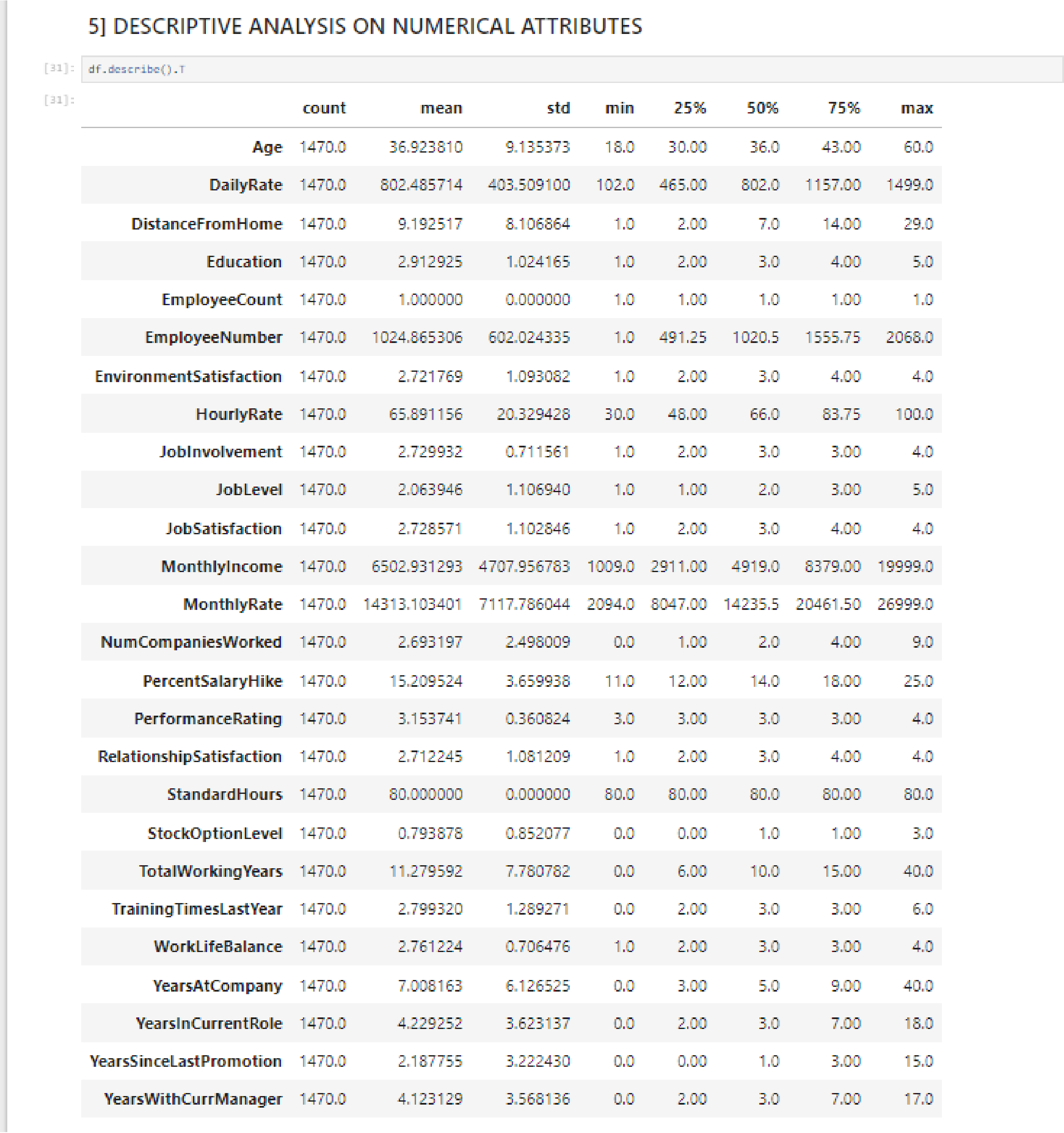
```
The shape of data frame: (1470, 35)
Number of Rows in the dataframe: 1470
Number of Columns in the dataframe: 35
```



2] GENERATING BASIC INFORMATION OF ATTRIBUTES

```
[20]: # Print the Long summary of the dataframe by setting verbose = True
      # Check for Non-NuLL or Nan NaLues in the dataset.

      df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Age                       1470 non-null   int64 
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64 
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64 
 6   Education                 1470 non-null   int64 
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64 
 9   EmployeeNumber            1470 non-null   int64 
 10  EnvironmentSatisfaction   1470 non-null   int64 
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64 
 13  JobInvolvement            1470 non-null   int64 
 14  JobLevel                  1470 non-null   int64 
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64 
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64 
 19  MonthlyRate               1470 non-null   int64 
 20  NumCompaniesWorked        1470 non-null   int64 
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64 
 24  PerformanceRating         1470 non-null   int64 
 25  RelationshipSatisfaction  1470 non-null   int64 
 26  StandardHours             1470 non-null   int64 
 27  StockOptionLevel          1470 non-null   int64 
 28  TotalWorkingYears         1470 non-null   int64 
 29  TrainingTimesLastYear     1470 non-null   int64 
 30  WorkLifeBalance           1470 non-null   int64 
 31  YearsAtCompany            1470 non-null   int64 
 32  YearsInCurrentRole        1470 non-null   int64 
 33  YearsSinceLastPromotion   1470 non-null   int64 
 34  YearsWithCurrManager      1470 non-null   int64 
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

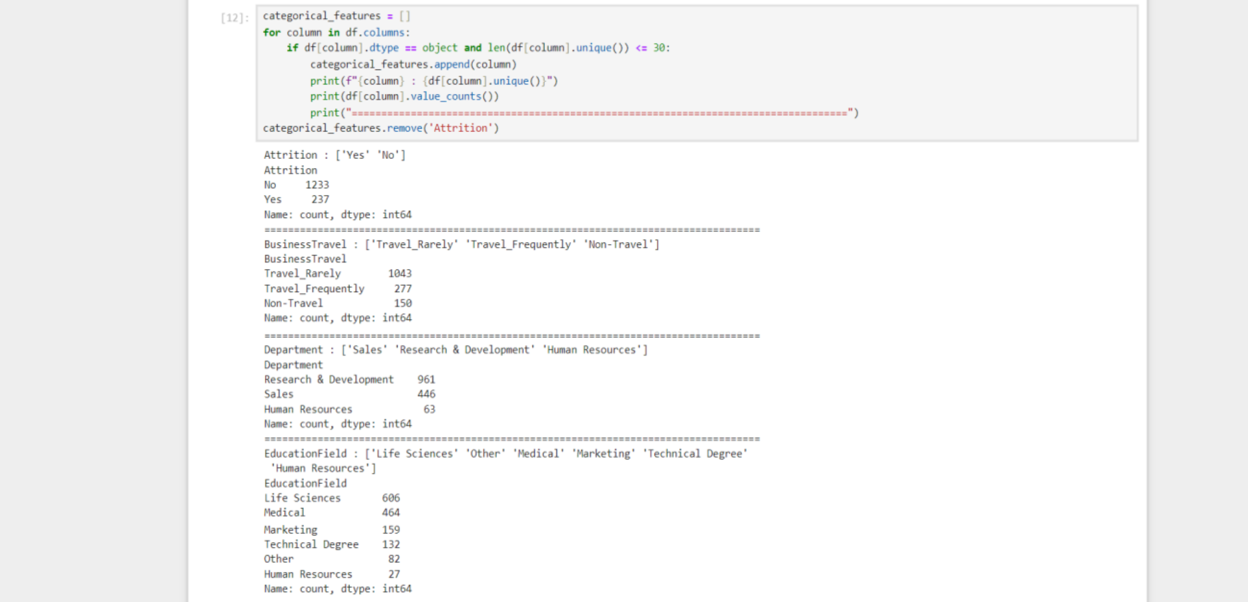


5] DESCRIPTIVE ANALYSIS ON NUMERICAL ATTRIBUTES

```
[31]: df.describe().T
```

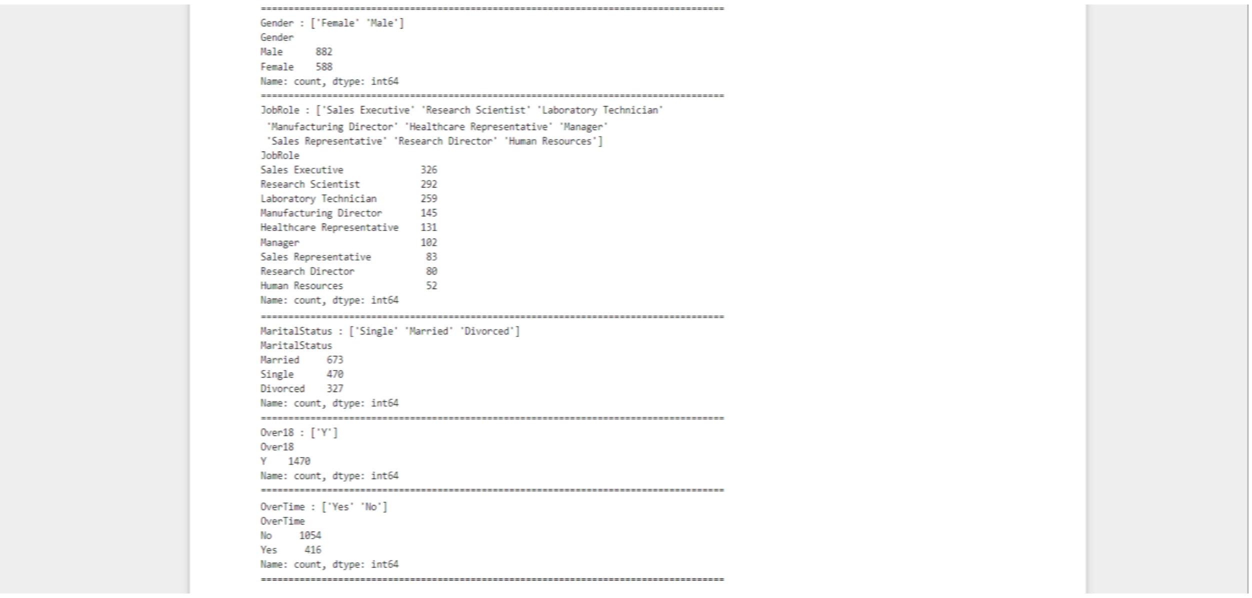
[31]:

	count	mean	std	min	25%	50%	75%	max
Age	1470.0	36.923810	9.135373	18.0	30.00	36.0	43.00	60.0
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	802.0	1157.00	1499.0
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	7.0	14.00	29.0
Education	1470.0	2.912925	1.024165	1.0	2.00	3.0	4.00	5.0
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1020.5	1555.75	2068.0
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	3.0	4.00	4.0
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	66.0	83.75	100.0
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	3.0	3.00	4.0
JobLevel	1470.0	2.063946	1.106940	1.0	1.00	2.0	3.00	5.0
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00	3.0	4.00	4.0
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00	4919.0	8379.00	19999.0
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00	14235.5	20461.50	26999.0
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.00	2.0	4.00	9.0
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.00	14.0	18.00	25.0
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.00	3.0	3.00	4.0
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.00	3.0	4.00	4.0
StandardHours	1470.0	80.000000	0.000000	80.0	80.00	80.0	80.00	80.0
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.00	1.0	1.00	3.0
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.00	10.0	15.00	40.0
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.00	3.0	3.00	6.0
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.00	3.0	3.00	4.0
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.00	5.0	9.00	40.0
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.00	3.0	7.00	18.0
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.00	1.0	3.00	15.0
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.00	3.0	7.00	17.0



```
[12]: categorical_features = []
      for column in df.columns:
          if df[column].dtype == object and len(df[column].unique()) <= 30:
              categorical_features.append(column)
              print(f"{column} : {df[column].unique()}")
              print(df[column].value_counts())
              print("====================================================================================")
      categorical_features.remove('Attrition')
```

```
Attrition : ['Yes' 'No']
Attrition
No     1233
Yes     237
Name: count, dtype: int64
====================================================================================
BusinessTravel : ['Travel_Rarely' 'Travel_Frequently' 'Non-Travel']
BusinessTravel
Travel_Rarely        1043
Travel_Frequently     277
Non-Travel            150
Name: count, dtype: int64
====================================================================================
Department : ['Sales' 'Research & Development' 'Human Resources']
Department
Research & Development    961
Sales                     446
Human Resources            63
Name: count, dtype: int64
====================================================================================
EducationField : ['Life Sciences' 'Other' 'Medical' 'Marketing' 'Technical Degree'
 'Human Resources']
EducationField
Life Sciences       606
Medical             464
Marketing           159
Technical Degree    132
Other                82
Human Resources      27
Name: count, dtype: int64
```

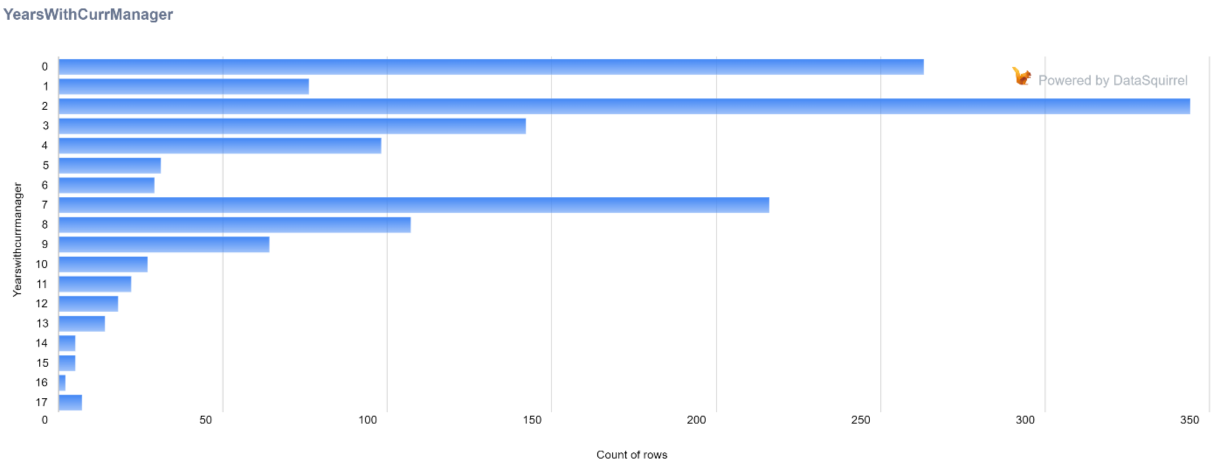


```
====================================================================================
Gender : ['Female' 'Male']
Gender
Male      882
Female    588
Name: count, dtype: int64
====================================================================================
JobRole : ['Sales Executive' 'Research Scientist' 'Laboratory Technician'
 'Manufacturing Director' 'Healthcare Representative' 'Manager'
 'Sales Representative' 'Research Director' 'Human Resources']
JobRole
Sales Executive              326
Research Scientist           292
Laboratory Technician        259
Manufacturing Director       145
Healthcare Representative    131
Manager                      102
Sales Representative          83
Research Director             80
Human Resources               52
Name: count, dtype: int64
====================================================================================
MaritalStatus : ['Single' 'Married' 'Divorced']
MaritalStatus
Married     673
Single      470
Divorced    327
Name: count, dtype: int64
====================================================================================
Over18 : ['Y']
Over18
Y    1470
Name: count, dtype: int64
====================================================================================
OverTime : ['Yes' 'No']
OverTime
No     1054
Yes     416
Name: count, dtype: int64
====================================================================================
```

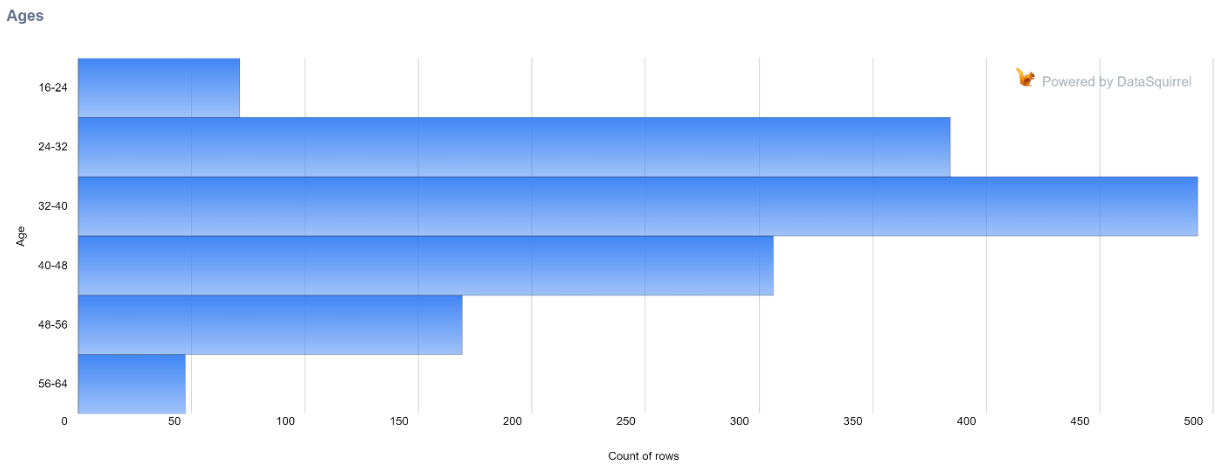
13.2 DATA VISUALIZATION USING PIVOTETABLE

By analyzing employee data, we can identify factors that contribute to employee attrition, such as job satisfaction, compensation, and work-life balance. This information can be used to develop strategies to retain top talent and reduce turnover rates. HR analytics can help identify high-performing employees by analyzing data related to performance metrics, such as productivity, quality, and customer satisfaction. This information can be used to develop strategies to retain top talent and improve overall organizational performance.

1. Displaying the distribution of years with the current manager.



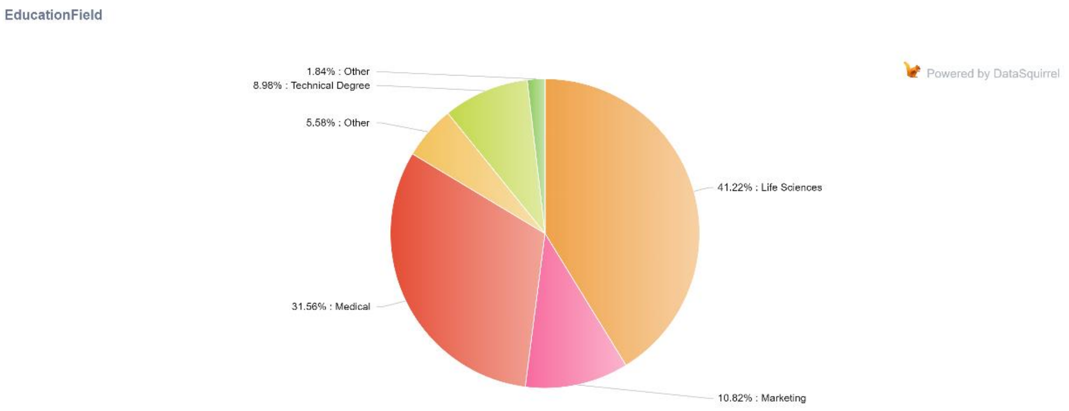
2. Displaying the distribution of employee ages.



Inference:

1. Most of the employees are between ages 37 to 36.
2. We can clearly observe a trend that as the age is increasing the attrition is decreasing.
3. From the boxplot we can also observe that the median age of employee who left the organization is less than the employees who are working in the organization.
4. Employees with young age leaves the company more compared to elder employees.

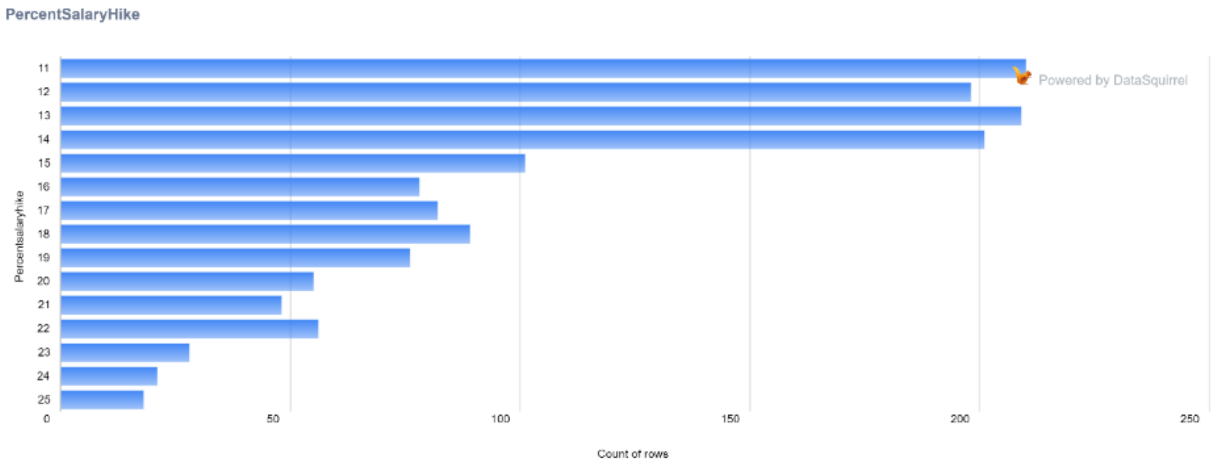
3. Showing the distribution of education fields.



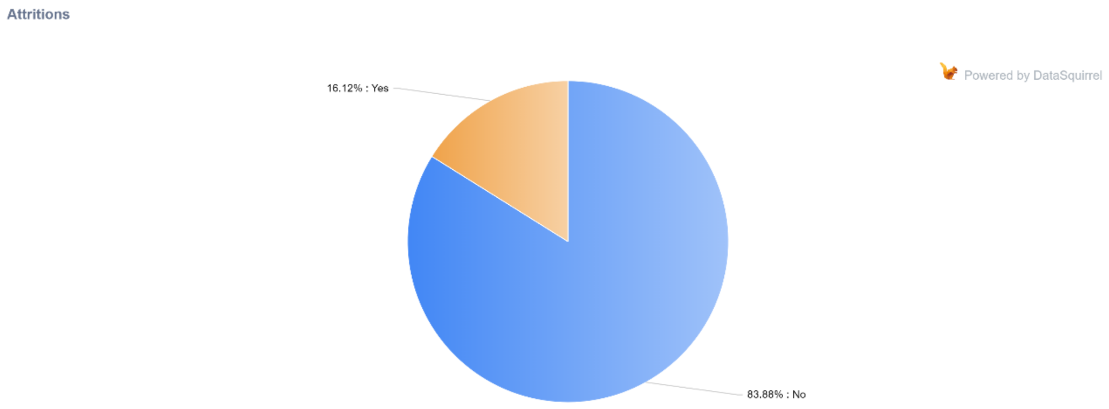
Inference:

1. Most of the employees are either from Life Science or Medical Education Field.
2. Very few employees are from Human Resources Education Field.
3. Education Fields like Human Resources, Marketing, and Technical is having very high attrition rate.
4. This may be because of work load because there are very few employees in these education fields compared to education field with less attrition rate.

4. Displaying the distribution of percent salary hikes.



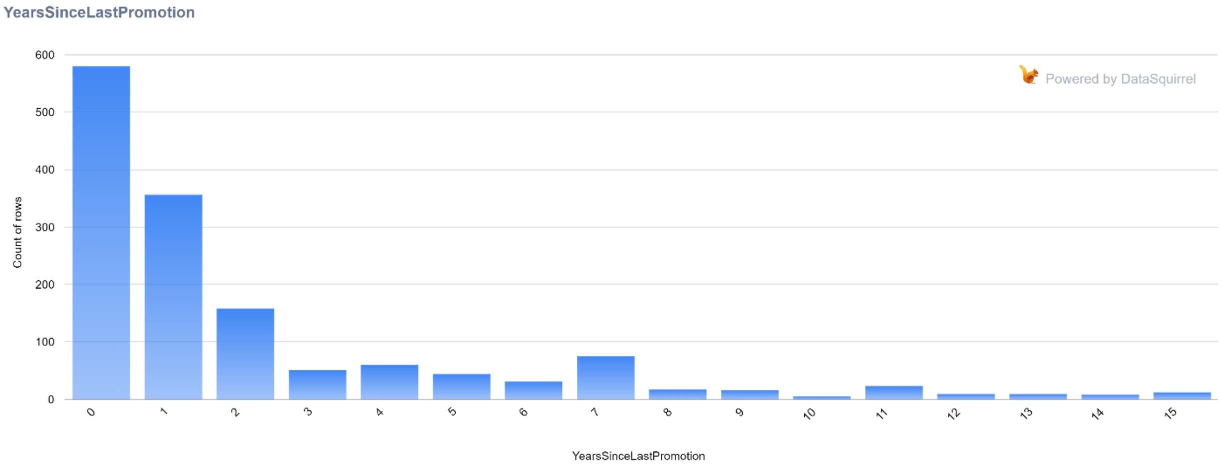
5. Showing the distribution of attrition indicators.



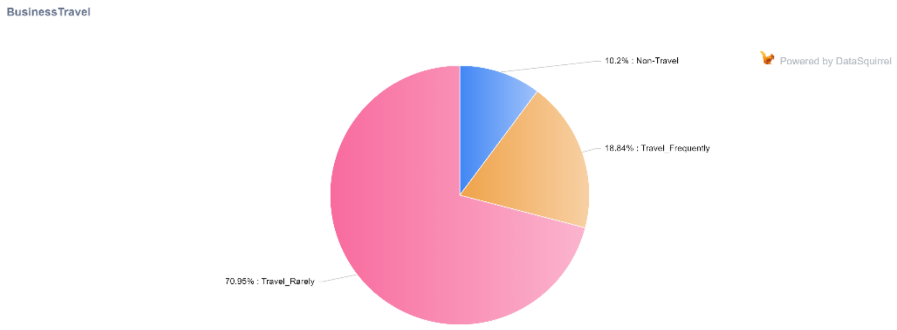
Inference:

1. The employee attrition rate of this organization is 16.12%.
2. According to experts in the field of Human Resources, says that the attrition rate 4% to 6% is normal in organization.
3. So, we can say the attrition rate of the organization is at a dangerous level.
4. Therefore, the organization should take measures to reduce the attrition rate.

6. Displaying the distribution of years since the last promotion.



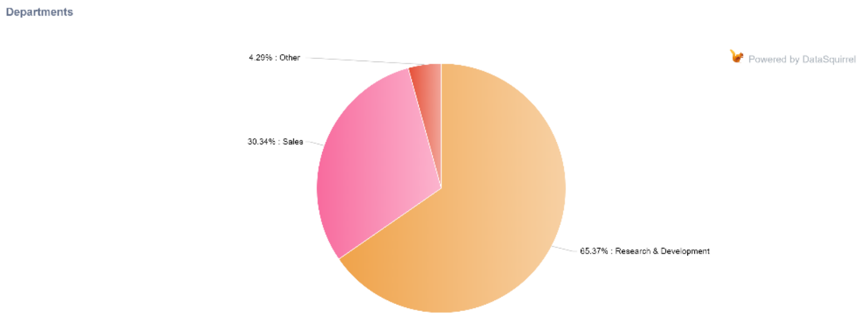
7. Showing the distribution of business travels.



Inference:

1. Most of the employees in the organization Travel Rarely.
2. Highest employee attrition can be observed by those employees who Travels Frequently.
3. Lowest employee attrition can be observed by those employees who are non-travel.

8. Showing the distribution of departments.

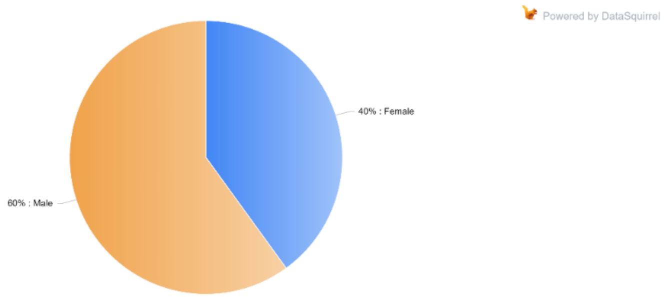


Inference:

1. Most of the employees are from Research & Development Department.
2. Highest Attrition is in the Sales Department.
3. Though of highest employees in Research & Development department there is least attrition compared to other departments

9. Showing the distribution of genders.

Genders

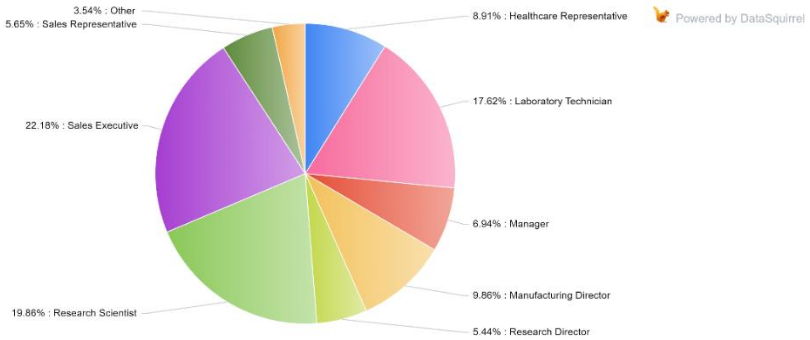


Inference:

1. The number of male employees in the organization accounts for a higher proportion than female employees by more than 20%.
2. Male employees are leaving more from the organization compared to female employees.

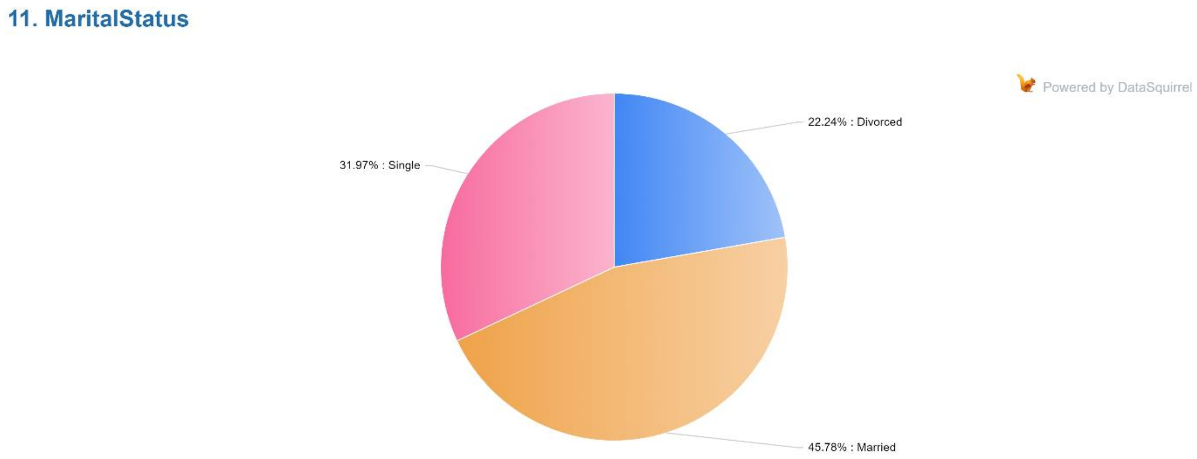
10. Showing the distribution of job roles.

JobRole



Inference:

1. Most employees are working as Sales executive, Research Scientist or Laboratory Technician in this organization.
2. Highest attrition rates are in sector of Research Director, Sales Executive, and Research Scientist



Inference:

1. Most of the employees are married in the organization.
2. The attrition rate is very high of employees who are divorced.
3. The attrition rate is low for employees who are single.

Pivot Visualization

13.3 STATISTICAL ANALYSIS

Statistical analysis plays a crucial role in HR analytics by helping organizations make informed decisions about their human resources and workforce management. It enables evidence-based decision-making, enhances workforce planning strategies, and fosters a deeper understanding of the organization's human capital dynamics.

1. Perform ANOVA Test:

ANOVA test is used to analysing the impact of different numerical features on a response categorical feature.

Inference:

The following features show a strong association with attrition, as indicated by their high F-scores and very low p-values.

1. Age
2. DailyRate
3. HourlyRate
4. MonthlyIncome
5. MonthlyRate
6. NumCompaniesWorked
7. PercentSalaryHike
8. TotalWorkingYears
9. TrainingTimesLastYear
10. YearsAtCompany
11. YearsWithCurrManager

The following features don’t shows significant relationship with attrition because of their moderate F-scores and extremely high p-values.

1. DistanceFromHome
2. StockOptionLevel
3. YearsInCurrentRole
4. YearsSinceLastPromotion

It is important for the organization to pay attention to the identified significant features and consider them when implementing strategies to reduce attrition rates.

2. Perform CHI-SQUARE Test:

CHI-SQUARE test is used to analysing the impact of different categorical features.

Inference:

The following features showed statistically significant associations with employee attrition:

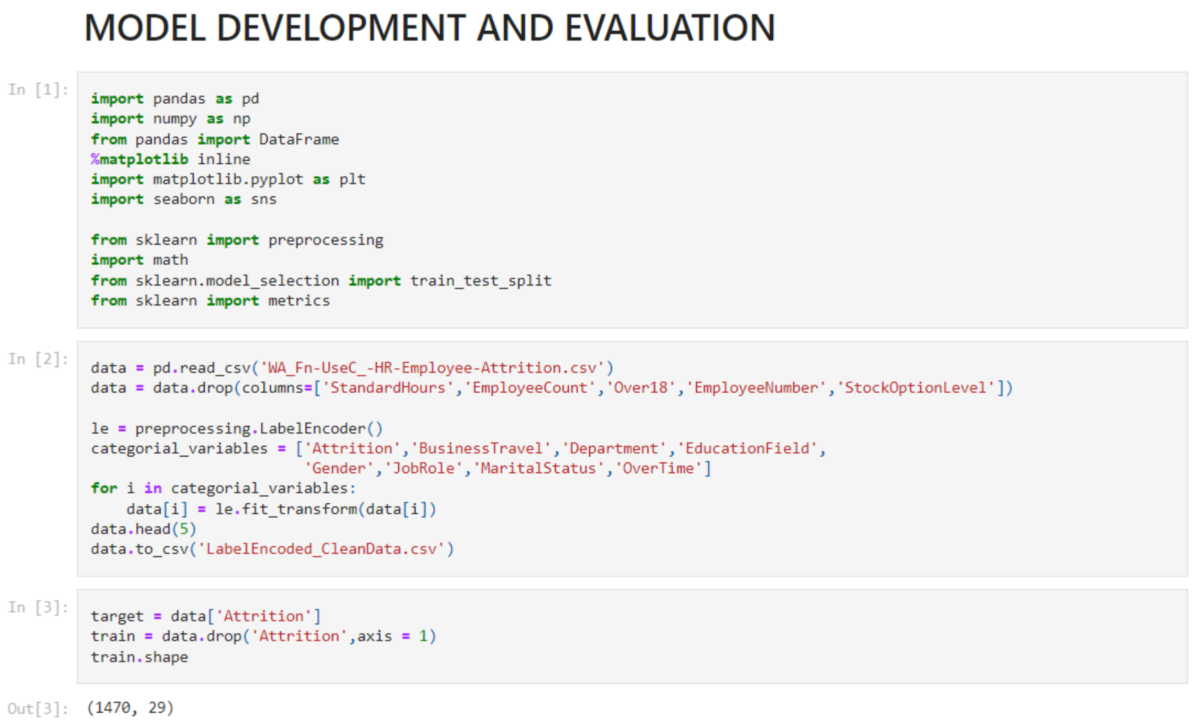
1. Department
2. EducationField
3. EnvironmentSatisfaction
4. JobInvolvement
5. JobLevel
6. JobRole
7. JobSatisfaction
8. MaritalStatus
9. OverTime
10. WorkLifeBalance

The following features did not show statistically significant associations with attrition.

1. Gender
2. Education
3. PerformanceRating
4. RelationshipSatisfaction

It is important for the organization to pay attention to the identified significant features and consider them when implementing strategies to reduce attrition rates.

13.4 MODEL DEVELOPMENT AND EVALUATION



MODEL DEVELOPMENT AND EVALUATION

In [1]:

```
import pandas as pd
import numpy as np
from pandas import DataFrame
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn import preprocessing
import math
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

In [2]:

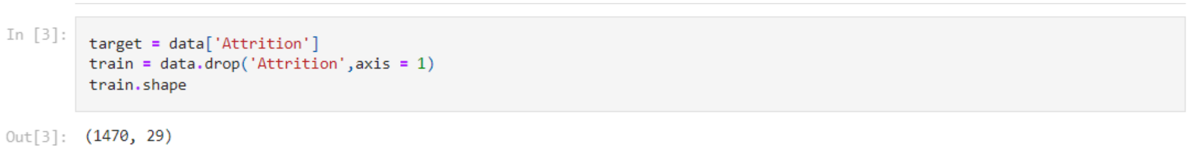
```
data = pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition.csv')
data = data.drop(columns=['StandardHours','EmployeeCount','Over18','EmployeeNumber','StockOptionLevel'])

le = preprocessing.LabelEncoder()
categorial_variables = ['Attrition','BusinessTravel','Department','EducationField',
                        'Gender','JobRole','MaritalStatus','OverTime']
for i in categorial_variables:
    data[i] = le.fit_transform(data[i])
data.head(5)
data.to_csv('LabelEncoded_CleanData.csv')
```

In [3]:

```
target = data['Attrition']
train = data.drop('Attrition',axis = 1)
train.shape
```

Out[3]:

```
(1470, 29)
```

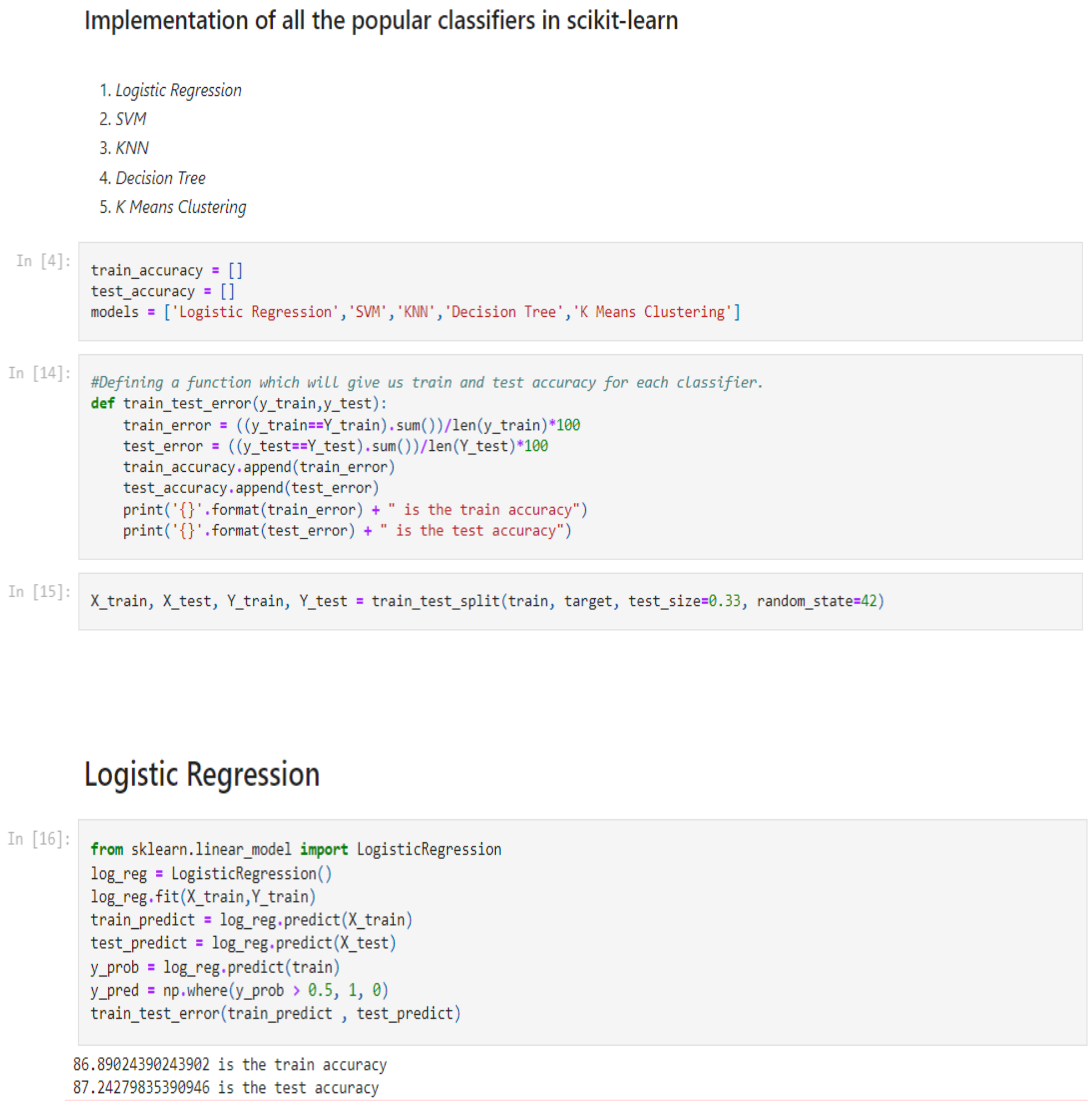


In [3]:

```
target = data['Attrition']
train = data.drop('Attrition',axis = 1)
train.shape
```

Out[3]:

```
(1470, 29)
```



Implementation of all the popular classifiers in scikit-learn

1. *Logistic Regression*
2. *SVM*
3. *KNN*
4. *Decision Tree*
5. *K Means Clustering*

In [4]:

```
train_accuracy = []
test_accuracy = []
models = ['Logistic Regression','SVM','KNN','Decision Tree','K Means Clustering']
```

In [14]:

```
#Defining a function which will give us train and test accuracy for each classifier.
def train_test_error(y_train,y_test):
    train_error = ((y_train==Y_train).sum())/len(y_train)*100
    test_error = ((y_test==Y_test).sum())/len(Y_test)*100
    train_accuracy.append(train_error)
    test_accuracy.append(test_error)
    print('{}'.format(train_error) + " is the train accuracy")
    print('{}'.format(test_error) + " is the test accuracy")
```

In [15]:

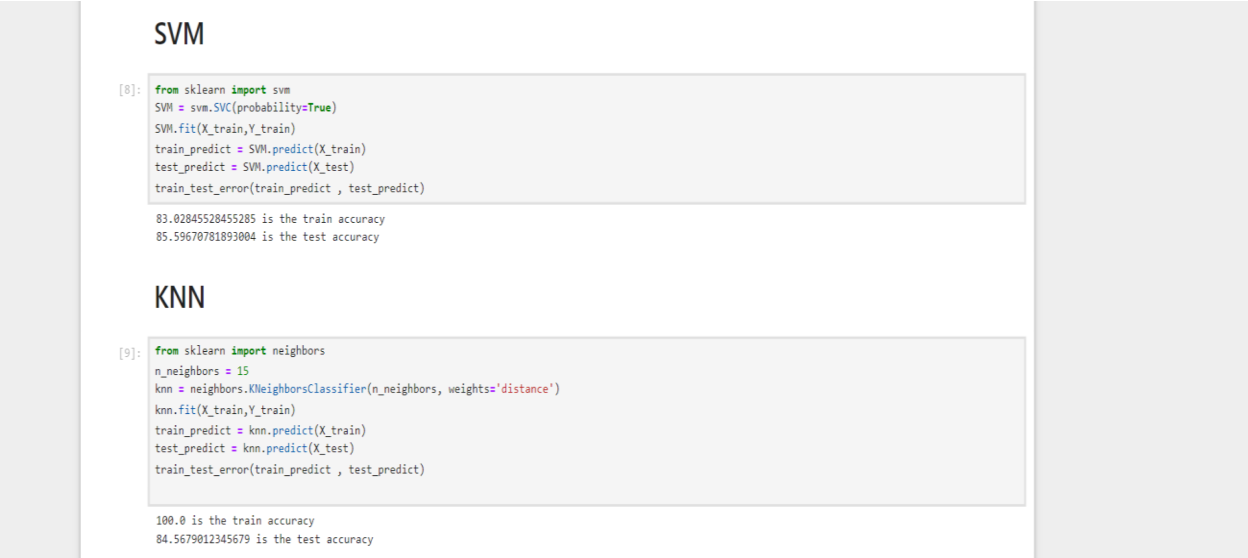
```
X_train, X_test, Y_train, Y_test = train_test_split(train, target, test_size=0.33, random_state=42)
```

Logistic Regression

In [16]:

```
from sklearn.linear_model import LogisticRegression
log_reg = LogisticRegression()
log_reg.fit(X_train,Y_train)
train_predict = log_reg.predict(X_train)
test_predict = log_reg.predict(X_test)
y_prob = log_reg.predict(train)
y_pred = np.where(y_prob > 0.5, 1, 0)
train_test_error(train_predict , test_predict)
```

```
86.89024390243902 is the train accuracy
87.24279835390946 is the test accuracy
```



SVM

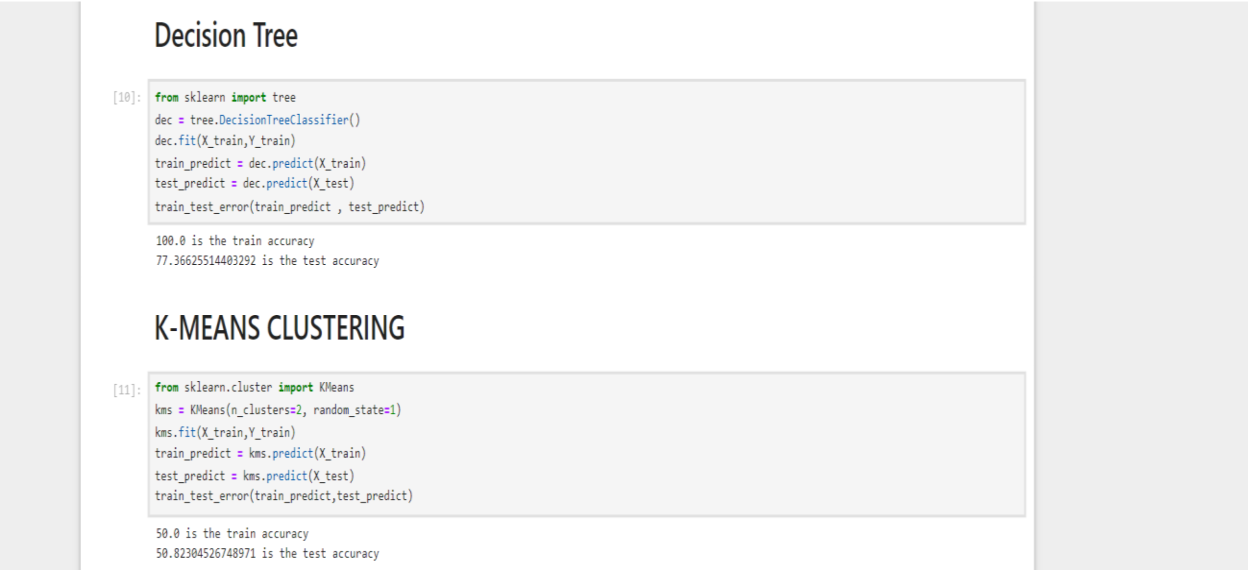
```
from sklearn import svm
SVM = svm.SVC(probability=True)
SVM.fit(X_train,Y_train)
train_predict = SVM.predict(X_train)
test_predict = SVM.predict(X_test)
train_test_error(train_predict , test_predict)
```

```
83.02845528455285 is the train accuracy
85.59670781893004 is the test accuracy
```

KNN

```
from sklearn import neighbors
n_neighbors = 15
knn = neighbors.KNeighborsClassifier(n_neighbors, weights='distance')
knn.fit(X_train,Y_train)
train_predict = knn.predict(X_train)
test_predict = knn.predict(X_test)
train_test_error(train_predict , test_predict)
```

```
100.0 is the train accuracy
84.5679012345679 is the test accuracy
```



Decision Tree

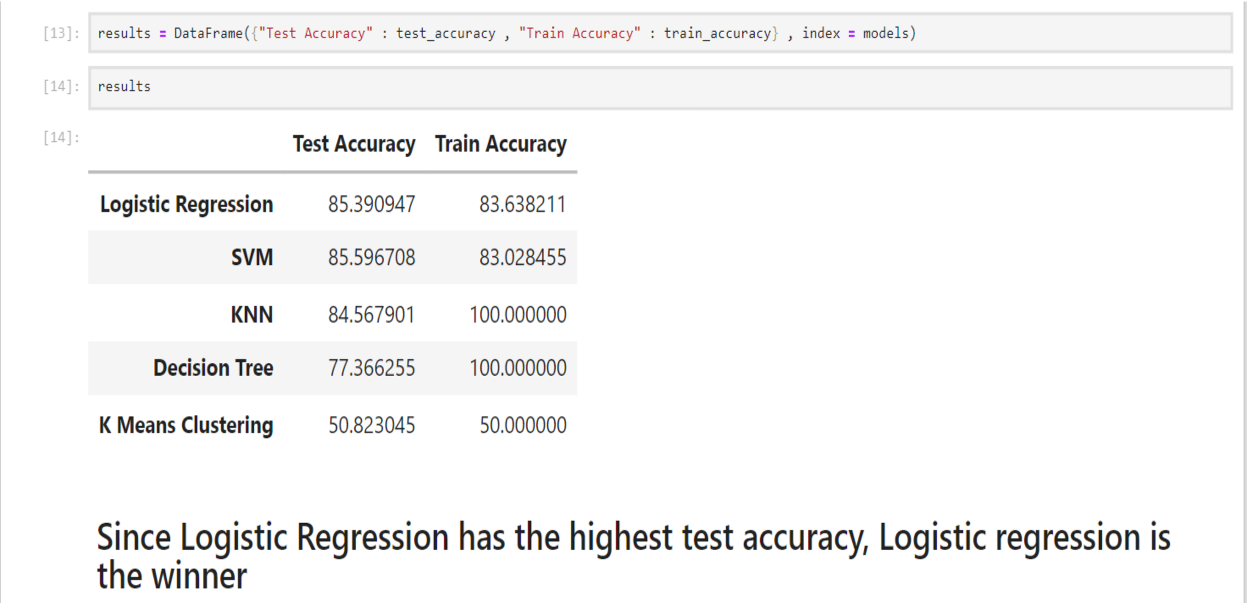
```
from sklearn import tree
dec = tree.DecisionTreeClassifier()
dec.fit(X_train,Y_train)
train_predict = dec.predict(X_train)
test_predict = dec.predict(X_test)
train_test_error(train_predict , test_predict)
```

```
100.0 is the train accuracy
77.36625514403292 is the test accuracy
```

K-MEANS CLUSTERING

```
from sklearn.cluster import KMeans
kms = KMeans(n_clusters=2, random_state=1)
kms.fit(X_train,Y_train)
train_predict = kms.predict(X_train)
test_predict = kms.predict(X_test)
train_test_error(train_predict,test_predict)
```

```
50.0 is the train accuracy
50.82304526748971 is the test accuracy
```



```
[13]: results = DataFrame({"Test Accuracy" : test_accuracy , "Train Accuracy" : train_accuracy} , index = models)
```

```
[14]: results
```

[14]:

	Test Accuracy	Train Accuracy
Logistic Regression	85.390947	83.638211
SVM	85.596708	83.028455
KNN	84.567901	100.000000
Decision Tree	77.366255	100.000000
K Means Clustering	50.823045	50.000000

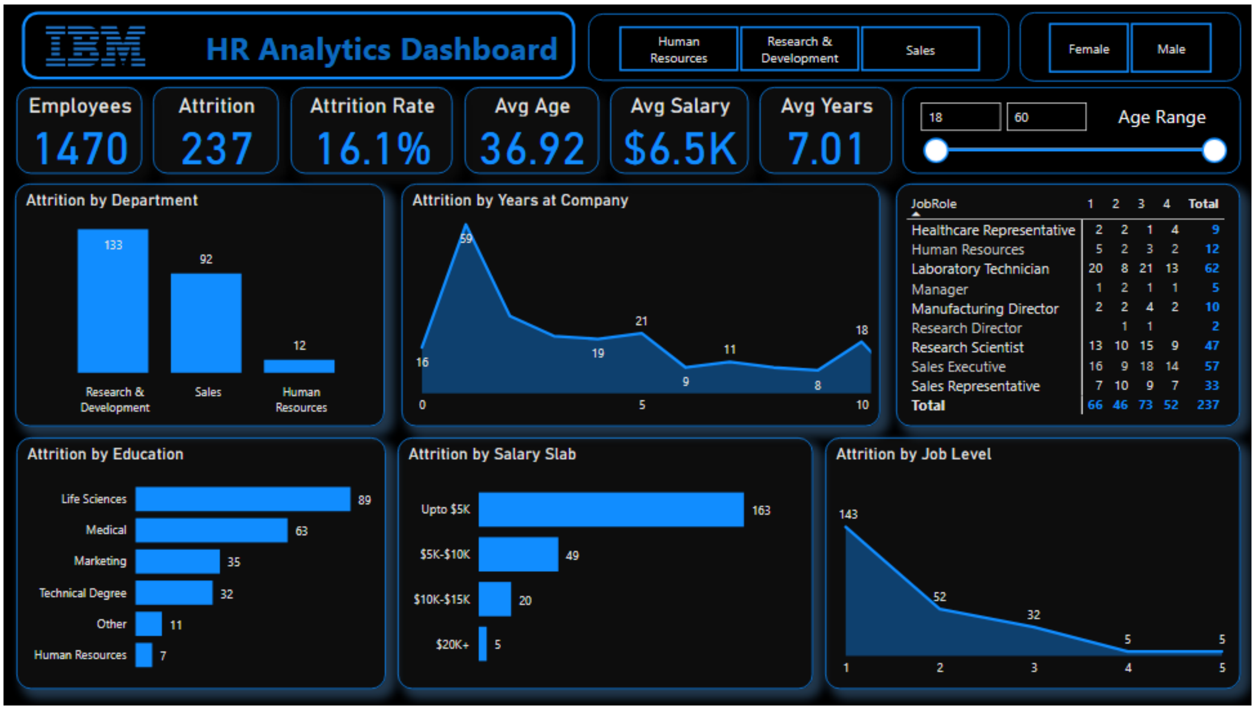
Since Logistic Regression has the highest test accuracy, Logistic regression is the winner

13.5 DASHBOARD

Description

This repository contains a Power BI dashboard of IBM Employee Attrition to answer questions about the data. The insights of the dashboard can be found in the results file. This repository can be used as a starting point for anyone who wants to learn how to use Power BI to analyze data.

Screenshot



Employees 1470

Attrition 237

Attrition Rate 16.1%

Avg Age 36.92

Avg Salary $6.5K

Avg Years 7.01

JobRole	1	2	3	4	Total
Healthcare Representative	2	2	1	4	9
Human Resources	5	2	3	2	12
Laboratory Technician	20	8	21	13	62
Manager	1	2	1	1	5
Manufacturing Director	2	2	4	2	10
Research Director		1	1		2
Research Scientist	13	10	15	9	47
Sales Executive	16	9	18	14	57
Sales Representative	7	10	9	7	33
Total	66	46	73	52	237

Conclusion

In this project, we developed a machine learning model to predict employee attrition, providing valuable insights into the factors contributing to turnover. By utilizing Python for data preprocessing, model development, and evaluation, we effectively handled a dataset with over 1,400 employee records, applying algorithms such as Logistic Regression, Random Forest. The final models demonstrated high accuracy and balanced performance across key metrics, with Random Forest and Gradient Boosting models emerging as the most effective in predicting attrition.

Additionally, we created an interactive Power BI dashboard, offering HR professionals a dynamic tool to visualize attrition trends, identify at-risk employees, and make data-driven decisions to improve employee retention. The dashboard facilitated the monitoring of key indicators such as job satisfaction, salary, and work-life balance, empowering organizations to proactively address potential attrition issues.

Overall, this project showcases the power of machine learning and data visualization in addressing real-world business challenges. By predicting attrition and understanding its key drivers, organizations can reduce employee turnover, lower recruitment costs, and foster a more stable and satisfied workforce.

FUTURE WORK

In the context of the previous HR analytics project on employee attrition, future work in sentiment analysis involves implementing sentiment analysis on employee feedback data to gain insights, monitoring sentiment in real-time, categorizing sentiments by topics, and analyzing historical sentiment trends. In terms of dashboard development, there's a need to create interactive, predictive, and benchmarking-enabled dashboards with custom alerts, engagement metrics, and mobile accessibility. Additionally, user training and support, data privacy, feedback integration, and performance monitoring are crucial aspects to ensure the dashboard's effectiveness in facilitating data-driven HR decisions and actions while adhering to privacy regulations.

REFERENCE

[1] Mishra S N, Lama D R and Pal Y 2016 Human Resource Predictive Analytics (HRPA) for HR Management in Organizations International Journal Of Scientific & Technology Research 5(5) 33-35

[2] Hoffman M and Tadelis S 2018 People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis National Bureau of Economic Research

[3] Frye A, Boomhower C, Smith M, Vitovsky L and Fabricant S 2018 Employee Attrition: What Makes an Employee Quit? MU Data Science Review 1(1)

[4] S. Rabiyathul Basariya, Ramyar Rzgar Ahmed, A STUDY ON ATTRITION - TURNOVER INTENTIONS OF EMPLOYEES, International Journal of Civil Engineering and Technology (IJCIET), 2019, 10(1), PP2594-2601

[5] Halkos, George & Bousinakis, Dimitrios, 2017. "The effect of stress and dissatisfaction on employees during crisis," Economic Analysis and Policy, Elsevier, vol. 55(C), pages 25-34.

[6] Glavas, A., & Willness, C. (2020). Employee (dis)engagement in corporate social responsibility. In D. Haski-Leventhal, L. Roza, & S. Brammer (Eds.), Employee engagement in corporate social responsibility (pp. 10–27). Sage Publications Ltd. https://doi.org/10.4135/9781529739176.n2

[7] S. Yadav, A. Jain and D. Singh, "Early Prediction of Employee Attrition using Data Mining Techniques," 2018 IEEE 8th International Advance Computing Conference (IACC), Greater Noida, India, 2018, pp. 349-354, doi: 10.1109/IADCC.2018.8692137.

[8] R. Jain and A. Nayyar, "Predicting Employee Attrition using XGBoost Machine Learning Approach," 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2018, pp. 113-120, doi: 10.1109/SYSMART.2018.8746940.

[9] lduayj, Sarah & Rajpoot, Kashif. (2018). Predicting Employee Attrition using Machine Learning. 93-98. 10.1109/INNOVATIONS.2018.8605976.
45

[10] Setiawan, Irwan & Suprihanto, Suprihanto & Nugraha, Ade & Hutahaean, Jonner. (2020). HR analytics: Employee attrition analysis using logistic regression. IOP Conference Series: Materials Science and Engineering. 830. 032001. 10.1088/1757-899X/830/3/032001.

[11] Yadav, Sandeep & Jain, Aman & Singh, Deepti. (2018). Early Prediction of Employee Attrition using Data Mining Techniques. 349-354. 10.1109/IADCC.2018.8692137.

[12] I. Ballal, S. Kavathekar, S. Janwe, P. Shete, and N. Bhirud, "People Leaving the Job-An Approach for Prediction Using Machine Learning," Int. J. Res. Anal. Rev., vol. 7, no. 1, pp. 8– 10, 2020, [Online]. Available: www.ijrar.org

[13] A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, and H. S. Alghamdi, "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods," Int. J. Mach. Learn. Comput., vol. 11, no. 2, pp. 110–114, 2021, doi: 10.18178/ijmlc.2021.11.2.1022.

[14] N. Mansor, N. S. Sani, and M. Aliff, "Machine Learning for Predicting Employee Attrition," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 11, pp. 435–445, 2021, doi: 10.14569/ IJACSA.2021.0121149.

[15] D. Saisanthiya, V. M. Gayathri, and P. Supraja, "Employee Attrition Prediction Using Machine Learning and Sentiment Analysis," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 5, pp. 7550–7557, 2020, doi: 10.30534/ijatcse/2020/91952020.