**Exploratory Data Analysis of Employee Attrition Dataset**

## **Introduction:**

Employee turnover poses a significant challenge for organizations, impacting productivity, team dynamics, and overall performance. Understanding the underlying factors contributing to employee attrition is crucial for developing effective retention strategies. This project aims to delve into the Employee Attrition dataset, employing data wrangling and exploratory data analysis techniques to uncover insights into employee turnover and inform retention strategies.

## **Dataset Description:**

The Employee Attrition dataset contains information about employee demographics, job characteristics, and attrition status. The variables included in the dataset are as follows:

|  |  |  |
| --- | --- | --- |
| Attribute | Type | Description |
| Age | Numerical | Age of the employee |
| Gender | Categorical | Gender of the employee |
| MaritalStatus | Categorical | Marital status of the employee |
| BusinessTravel | Categorical | Frequency of business travel |
| DailyRate | Numerical | Daily rate of pay |
| MonthlyIncome | Numerical | Monthly income |
| MonthlyRate | Numerical | Monthly rate of pay |
| HourlyRate | Numerical | Hourly rate of pay |
| Department | Categorical | Department in which the employee works |
| JobRole | Categorical | Role of the employee within the organization |
| JobLevel | Numerical | Level of the employee's job within the organization |
| DistanceFromHome | Numerical | Distance from home to workplace |
| Education | Numerical | Education level of the employee |
| EducationField | Categorical | Field of education of the employee |
| EnvironmentSatisfaction | Numerical | Satisfaction with the work environment |
| JobSatisfaction | Numerical | Satisfaction with the job |
| RelationshipSatisfaction | Numerical | Satisfaction with work relationships |
| WorkLifeBalance | Numerical | Satisfaction with work-life balance |
| NumCompaniesWorked | Numerical | Number of companies the employee has worked for |
| TotalWorkingYears | Numerical | Total number of years the employee has worked |
| OverTime | Categorical | Whether the employee works overtime |
| PerformanceRating | Numerical | Performance rating of the employee |
| PercentSalaryHike | Numerical | Percentage increase in salary |
| StockOptionLevel | Numerical | Level of stock options granted to the employee |
| TrainingTimesLastYear | Numerical | Number of training sessions attended last year |
| YearsAtCompany | Numerical | Number of years the employee has worked at the company |
| YearsInCurrentRole | Numerical | Number of years in the current role |
| YearsSinceLastPromotion | Numerical | Number of years since the last promotion |
| YearsWithCurrManager | Numerical | Number of years with the current manager |

**Initial Data Summaries:**

Initial data summaries have been conducted to comprehend the distribution and characteristics of key attributes. Summary statistics, such as mean, median and standard deviation have been calculated for numerical variables like Age, DailyRate, and MonthlyIncome. Additionally, categorical variables like Gender, MaritalStatus, and Department have been summarized to identify prevalent categories and their distributions.

**Goal:**

The revised goal of this project is to explore the Employee Attrition dataset using Tableau to uncover insights into employee turnover trends and factors influencing attrition. By conducting exploratory data analysis, the aim is to identify key drivers of employee attrition which can provide valuable insights for understanding employee behavior. The dataset is relevant and useful for achieving this goal as it contains comprehensive information about employee demographics, job characteristics, and attrition status, allowing for a thorough analysis of factors contributing to attrition.

**Literature review:**

1. Title: [An Effectiveness of Human Resource Management Practices on Employee Retention in Institute of Higher learning: - A Regression Analysis](https://www.researchgate.net/profile/Charles-Ramendran/publication/267725979_An_Effectiveness_of_Human_Resource_Management_Practices_on_Employee_Retention_in_Institute_of_Higher_learning_-A_Regression_Analysis/links/54d0dacf0cf298d6566937a0/An-Effectiveness-of-Human-Resource-Management-Practices-on-Employee-Retention-in-Institute-of-Higher-learning-A-Regression-Analysis.pdf)

Authors: Eric Ng Chee Hong, Lam Zheng Hao, Ramesh Kumar, Charles Ramendran, Vimala Kadiresan

Summary:

This paper investigates the impact of human resource management practices on employee retention, focusing on factors such as empowerment, training, appraisal systems, and compensation. Through quantitative analysis involving 278 individuals, it finds that while training, compensation, and appraisal systems significantly influence retention, employee empowerment has less significance, especially in Asian cultures. The study emphasizes the importance of addressing employees' needs for training, fair compensation, and recognition to enhance retention strategies. It also highlights cultural differences in retention preferences between Eastern and Western countries. Ultimately, prioritizing these factors can potentially improve employee retention rates and enhance organizational productivity.

1. Title: [Boosting employee retention through CSR: A configurational analysis](https://onlinelibrary.wiley.com/doi/full/10.1002/csr.1511?saml_referrer)

Authors: Leemen Lee, Li-Fei Chen

Summary:

This paper explores the interplay between corporate social responsibility (CSR) initiatives, employee job needs (existence, relatedness, and growth), and their impact on job satisfaction and retention intention within a leading chain department store in Asia. By utilizing fuzzy-set qualitative comparative analysis (fsQCA), the study uncovers complex relationships between CSR, employee needs, and organizational outcomes. Findings suggest that positive perceptions of CSR practices coupled with high fulfillment of employee ERG needs correlate with increased job satisfaction and retention intention. Additionally, the study emphasizes the importance of considering environmental and consumer-focused CSR factors for optimal employee outcomes. The research underscores the need for a balanced approach to CSR, integrating both top-down and bottom-up perspectives. Overall, the study contributes to understanding the psychological dimensions of CSR and its implications for employee retention strategies.

1. Title: [IBM Employee Attrition Analysis](https://arxiv.org/pdf/2012.01286.pdf)

Authors: Shenghuan Yang, Md Tariqul Islam

Summary:

This paper delves into the analysis of the IBM Employee Attrition dataset to uncover the primary factors influencing employee resignations. Employing various techniques including correlation matrix, Random Forest, K-means Clustering, and binary logistic regression, the study identifies key determinants of attrition such as monthly income, age, job level, job satisfaction, and frequency of travel. The findings reveal that older individuals with higher job levels, job satisfaction, and monthly income, as well as those who have worked for multiple companies, are less likely to leave. However, certain demographics such as females, married or divorced individuals, and employees in the Human Resources department show a higher tendency to resign. The study emphasizes the need for further qualitative analysis to understand diverse employee intentions. Model evaluation indicates that logistic regression outperforms Random Forest in predicting attrition. The paper concludes with recommendations for the company to improve job satisfaction, particularly among Human Resources employees, and to prioritize employee well-being to enhance retention rates.

1. Title: [Predicting Employee Attrition Using Machine Learning Techniques](https://www.mdpi.com/2073-431X/9/4/86)

Authors: Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano, Ernesto William De Luca

Summary:

This paper explores the application of machine learning techniques in human resources management to predict employee attrition. By analyzing a dataset provided by IBM analytics containing 35 features and around 1500 samples, the study aims to identify key indicators and predict the likelihood of employees leaving the company. The methodology follows the Team Data Science Process framework, involving data collection, cleaning, descriptive analysis, model training, and performance evaluation. Various classification algorithms are tested, and the Gaussian Naïve Bayes classifier is identified as the most suitable, achieving a recall rate of 0.54 and a false negative rate of 4.5%. Factors such as monthly income, age, overtime, and distance from home are found to be significant in predicting attrition. The study concludes by highlighting the potential for further research to improve attrition classifiers through larger datasets, feature engineering, and consideration of external labor market factors. The findings offer insights for HR practitioners to implement proactive measures to mitigate employee turnover risks.

**Questions:**

No questions or concerns have been identified at this point in the project. However, further exploration may be required during the data analysis phase to ensure comprehensive understanding and interpretation of the findings.