Proposal for HR Employee Attrition Analysis

1.1 Project 1. Proposal and Data Selection

Milestone 1

Jennifer Barrera Conde

Bellevue University

Applied Data Science

August 29, 2024

Topic

"Understanding Key Drivers of Employee Attrition in Organizations."

I selected this topic due to its relevance to virtually every business, as effective employee retention is critical to organizational success.

The cost of hiring new talent significantly outweighs the investment required to improve and retain existing employees. Recruiting new employees involves direct expenses like advertising, interviewing, and onboarding and indirect costs such as lost productivity, training time, and the integration period for new hires to reach full performance. In contrast, focusing on employee development and retention strategies helps build a loyal, skilled workforce, reducing turnover and preserving institutional knowledge.

Business Problem

Employee attrition can be costly for businesses, leading to loss of talent, reduced productivity, and increased hiring costs. This project aims to identify the primary factors contributing to employee attrition, helping organizations implement strategies to retain valuable employees and reduce turnover.

According to the Society for Human Resource Management (SHRM), the average cost-per-hire is around $4,129, and it takes approximately 42 days to fill a position. Additionally, replacing an employee can be as high as 50% to 60% of their annual salary for entry-level employees and up to 150% to 200% for high-level positions. Retaining current talent through career development, training programs, and employee engagement initiatives reduces costs and enhances overall organizational performance and morale.

Critical strategies for retention include fostering a positive work environment, providing opportunities for career growth, recognizing employee contributions, and offering competitive compensation packages.

Datasets

The dataset being used for this project is the HR Employee Attrition dataset. It contains detailed employee data, including demographics, job roles, tenure, salary, and attrition status, which will be analyzed to identify patterns associated with employee turnover. IBM has collected the data and retrieved it using Kaggle.

The dataset contains 1,470 entries and 35 columns, covering various employee-related factors potentially linked to attrition. Here is a brief overview of key attributes in the dataset:

- **Age**: Employee age (numerical).

- **Attrition**: Whether the employee left the company (Yes/No).

- **BusinessTravel**: Frequency of business travel (Travel_Rarely, Travel_Frequently, Non-Travel).

- **Department**: The department where employees work (e.g., Research & Development, Sales).

- **DistanceFromHome**: The distance between the employee's home and the workplace (numerical).

- **Education**: Level of education (1-5 scale).

- **JobRole**: Employee's job title (e.g., Sales Executive, Research Scientist).

- **MonthlyIncome**: Monthly salary of the employee (numerical).

- **YearsAtCompany**: Total number of years the employee has worked there.

Other essential factors include job satisfaction, work-life balance, performance rating, and overtime status. These variables will help analyze potential drivers of employee attrition.

Methods

The analysis could include, but is not limited to, the following methods:

- Descriptive statistics to understand general trends and characteristics of the data. To summarize the main characteristics of the dataset, such as averages, counts, or distributions of the variables.

- Correlation analysis to identify relationships between variables. To determine relationships between variables, helping identify which factors are most associated with employee attrition. Using libraries such as "matplotlib" and "Seaborn."

- Logistic regression and decision tree classification to predict employee attrition based on various factors. The logistic regression model will be used to find the probability of employee attrition (binary outcome) based on multiple independent variables using libraries from Python such as "statsmodels" and "sklearn." The decision tree map classifies whether an employee will leave based on key features.

- Data visualization (using tools like Tableau, Power BI, Python, or R) to highlight key findings and trends. To create visual insights from the data, such as trends and patterns related to employee attrition.

Ethical Considerations

Potential ethical concerns include:

- **Privacy of employee data:** Ensuring the dataset is anonymized, and personal information is protected.

- **Bias in the analysis:** Care should be taken to avoid reinforcing biases, such as discrimination based on age, gender, or ethnicity, when concluding the data.

Challenges/Issues

- **Data quality:** Missing or incomplete data may affect the accuracy of the analysis.

- **Model accuracy:** Predictive models may face challenges in achieving high accuracy due to the complexity of factors influencing employee attrition.

- **Interpretability:** Ensuring that the insights generated are understandable and actionable for business decision-makers.

References

For references, I will primarily rely on the provided dataset. However, I will supplement my findings with additional credible sources to ensure a comprehensive and well-substantiated analysis. For example:

- Industry reports and research studies on employee retention and turnover.

- Academic papers on attrition prediction models and HR analytics.

- Any internal documentation or HR-related case studies from reliable sources to validate the findings.

Reference For This Milestone

Society for Human Resource Management. (2017). The recruiting cost ratio. SHRM.

https://www.shrm.org

https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/code