Proposal for HR Employee Attrition Analysis

Jennifer Barrera Conde

Bellevue University

Applied Data Science

Sept 19, 2024

Business Problem

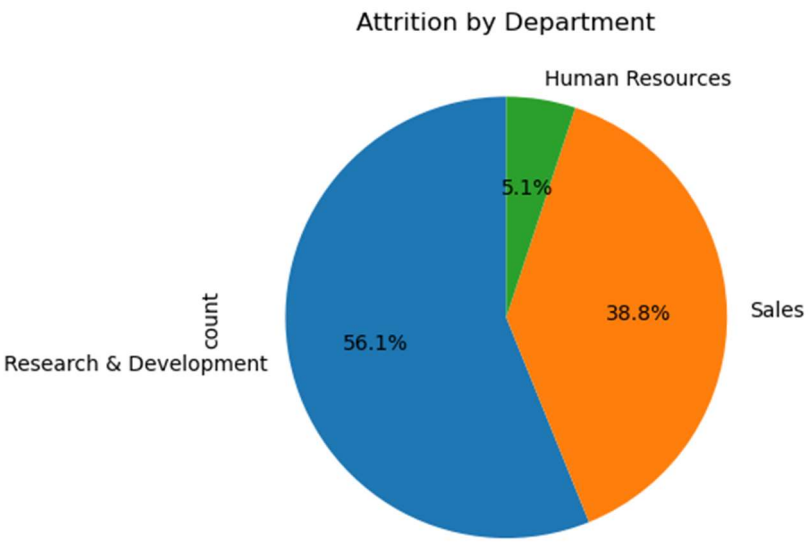
Employee attrition can have a significant financial impact on organizations, resulting in the loss of valuable talent, reduced productivity, and increased hiring costs. This project aims to identify the primary factors contributing to employee attrition, enabling organizations to implement strategies to retain employees and reduce turnover.

According to the Society for Human Resource Management (SHRM), the average cost per hire is approximately $4,129, and it takes 42 days to fill a position. Moreover, replacing an employee can cost between 50% to 60% of an entry-level employee's annual salary and up to 150% to 200% for high-level positions. Implementing effective retention strategies such as career development, training programs, and employee engagement initiatives can reduce these costs, improve organizational performance, and boost employee morale.

Key strategies for retention include:

- Fostering a positive work environment
- Providing career growth opportunities
- Recognizing employee contributions
- Offering competitive compensation packages

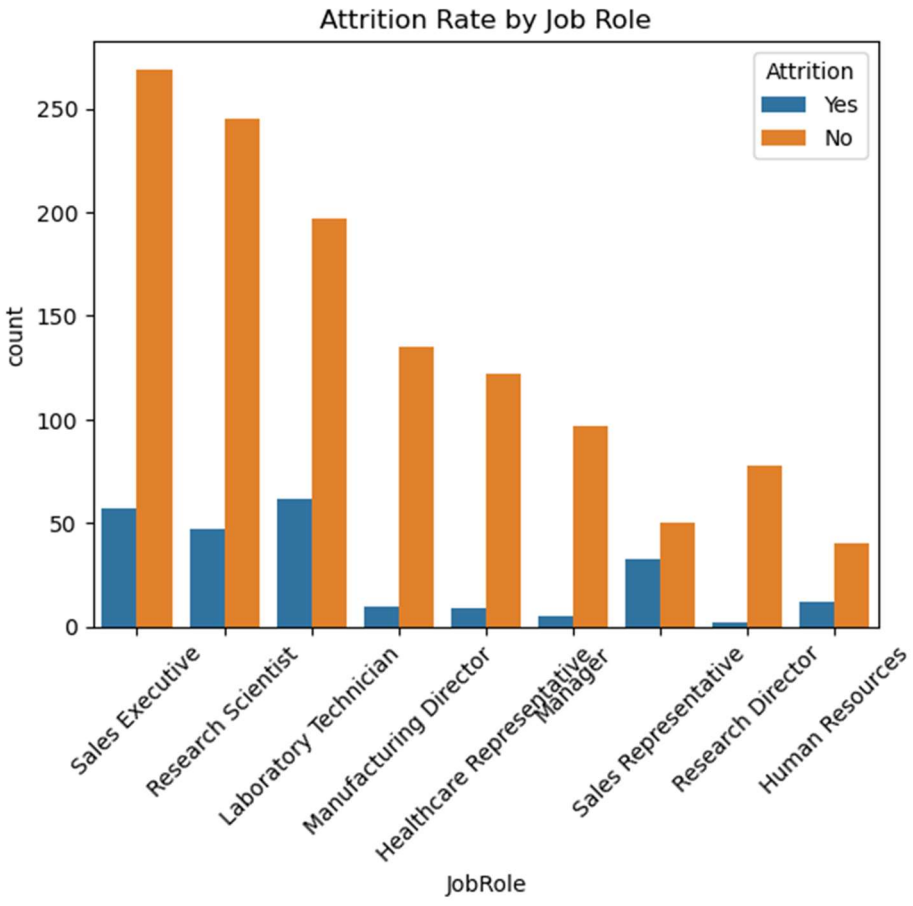
From the dataset's initial analysis, most employee attrition occurs in the research and development (R&D) department. Analyzing job roles within this department can reveal whether specific roles are more



susceptible to attrition. The job roles identified within R&D include:

- Research Scientist
- Laboratory Technician
- Manufacturing Director
- Healthcare Representative
- Research Director
- Manager

The highest attrition rates have been observed among Laboratory Technicians, Sales Executives, and Research Scientists, while roles such as Research Director, Manager, and Healthcare Representative have lower attrition rates. This analysis can help companies focus on roles with higher attrition and develop targeted retention strategies.



Background/History

Employee attrition has been a longstanding issue in human resource management, significantly affecting business performance and profitability. Research indicates that job satisfaction, work-life balance, compensation, and career development opportunities are key factors influencing whether employees remain with a company.

With advancements in data analytics, organizations can now predict employee behavior and take preemptive actions to retain key employees. Studies show that investing in retention efforts is more cost-effective than recruiting and onboarding new hires. Recruitment involves direct costs like advertising and interviewing, as well as indirect costs such as lost productivity and training. Focusing on employee development builds a skilled workforce and reduces turnover.

Datasets

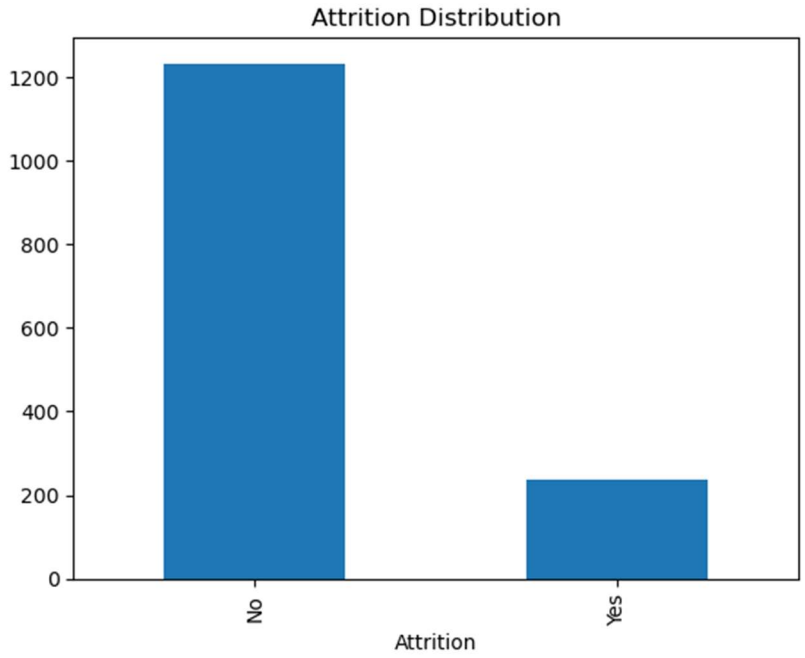
The dataset contains 1,470 entries and 35 columns, covering various employee-related factors potentially linked to attrition. Here is a brief overview of key attributes in the dataset:

- **Age**: Employee age (numerical).
- **Attrition**: Whether the employee left the company (Yes/No).
- **Business Travel**: Frequency of business travel (Travel_Rarely, Travel_Frequently, Non-Travel).
- **Department**: The department where employees work (e.g., Research & Development, Sales).
- **Distance From Home**: The distance between the employee's home and the workplace (numerical).
- **Education**: Level of education (1-5 scale).

- **Job Role**: Employee's job title (e.g., Sales Executive, Research Scientist).
- **Monthly Income**: Monthly salary of the employee (numerical).
- **Years At Company**: 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. The use of the variables will be decided as the project develops.

From a quick inspection of the data, we can inherit that about 200 of the 1470 participants of Attrition, which means that 7% of the 1470 participants retired or resigned from their jobs.



Data Preparation

The data is clean, with no missing values. It was preprocessed by encoding categorical variables (e.g., converting "Attrition" into binary values: Yes = 1, No = 0). Data was also normalized where necessary to ensure uniformity in the analysis.

Methods

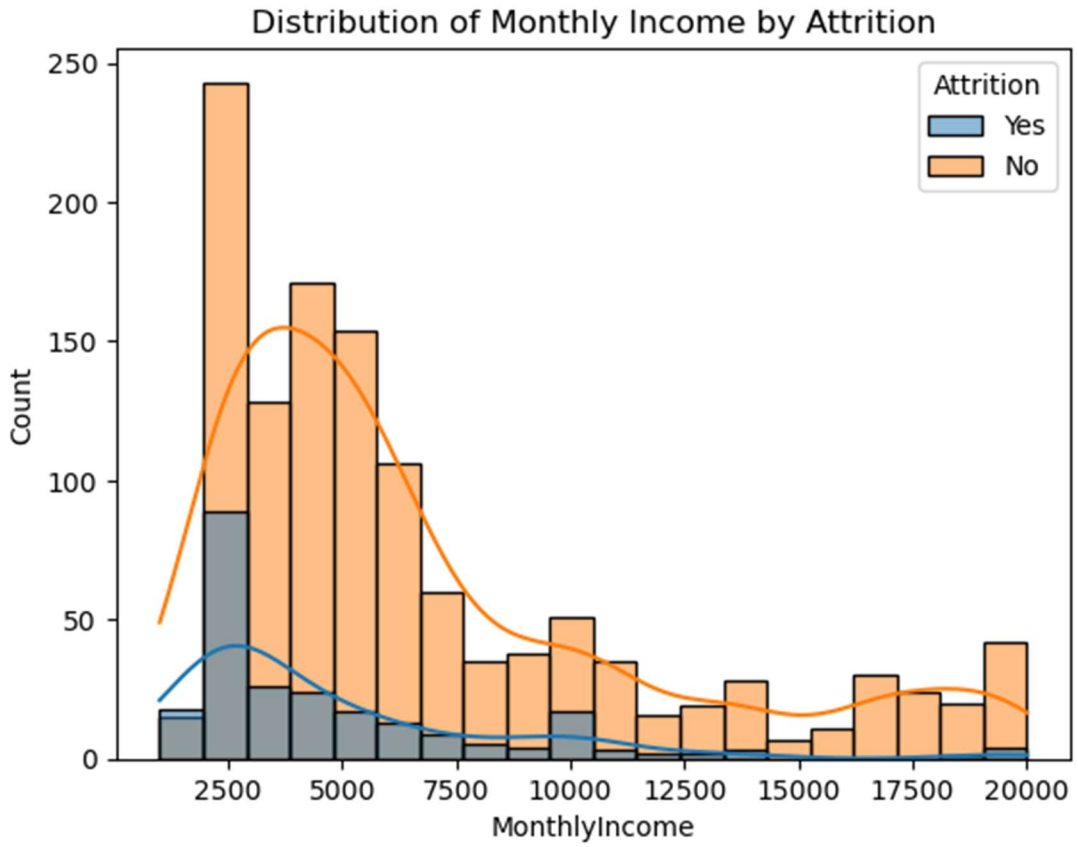
The analysis could include, but is not limited to, the following methods:

- 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.

Analysis

Initial descriptive statistics revealed that age, monthly income, and distance from home correlate with attrition. Correlation analysis showed moderate correlations between job satisfaction, work-life balance, and employee attrition. The logistic regression and decision tree models indicated that employees with lower job satisfaction and work-life balance are likelier to leave the company.

I could also view attrition from an income perspective to further my earlier analysis. Does attrition occur more in higher or lower-income job roles? I would assume it would happen most often in lower-income job roles, and I may be on the correct assumption, but not entirely.



If my previous assumption had been entirely correct, there would be a decline of attrition on higher paying roles where the graph is right-skewed, which is accurate, but there is more to it. Under a quick impression, I would be 100% correct. Still, upon closer inspection, we can see an attrition spike around the monthly income of $10,000 when it was on a steady decline and another on the $20,000 monthly income, the highest income from the collected data. One could compare the individuals with a monthly income of $10,000 and $20,000 to discover similarities. After that, they could be compared to the lowest attritions between the monthly incomes of $15,000 and $17,500.

This is just another example of how interesting but challenging it is to understand attrition since it becomes human behavior and quantifying it may become difficult. Still, the data does provide scorings that could be related to human behavior, such as Job Satisfaction, Work-Life Balance, Environment Satisfaction, Relationship Satisfaction, and more.

Conclusion

This analysis highlights that job satisfaction, work-life balance, and compensation play critical roles in employee retention. By addressing these areas, companies can reduce turnover and enhance employee retention.

Challenges

- **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. It is difficult to predict human behavior, such as attrition, due to its complexity, and assumptions can be misleading with human behavior.
- **Interpretability:** Ensuring that the insights generated are understandable and actionable for business decision-makers.
- **Data bias:** There could be unintentional bias in the data, especially regarding gender, age, and marital status.

Future Uses/Additional Applications

The models can be applied to other organizations or industries to predict attrition. This analysis could be expanded to identify the effectiveness of employee retention programs by tracking changes over time. Similar models could forecast other HR issues, such as employee performance and engagement.

Recommendations

- Implement targeted retention strategies focusing on job satisfaction and work-life balance.
- Develop personalized career development plans and competitive salary packages.
- Conduct periodic employee satisfaction surveys to monitor changes in crucial attrition factors.

Implementation Plan

- **Data-Driven Strategy:** Use the findings from the data analysis to inform HR retention strategies, focusing on areas like work-life balance and job satisfaction.
- **Employee Engagement Initiatives:** Launch programs to improve employee engagement, such as flexible working arrangements, mentorship, and professional development opportunities.
- **Continuous Monitoring:** Establish regular data monitoring and analysis processes to track attrition trends.

Ethical Assessment

- **Data Privacy:** Ensure that employee data used in the analysis is anonymized to protect privacy.
- **Bias:** Guard against any analysis or predictive model that could perpetuate or introduce bias against specific employee demographics (e.g., age, gender).
- **Transparency:** Communicate findings and actions derived from the analysis in a transparent way to employees.

10 Audience Questions for Milestone 4

Business Problem:

- How do you plan to quantify the financial impact of attrition on the company?

Background/History:

- What common retention strategies have other companies successfully implemented to reduce attrition?

Data Explanation:

- What specific factors in the dataset are most likely to predict attrition?
- Did the dataset include any biases, such as overrepresenting a particular demographic, and how did you address them?

Methods:

- Why did you choose logistic regression and decision tree models for this analysis?
- How did you determine which variables were the most important predictors of attrition?

Analysis:

- What challenges did you face in building predictive models, and how did you overcome them?

Future Uses/Applications:

- How could the attrition models be applied to other departments or industries?

Recommendations:

- What specific recommendations would you give to reduce attrition in the high-risk job roles identified in your analysis?

Ethical Assessment:

- How will you ensure that the data analysis and its outcomes do not unintentionally discriminate against specific groups of employees?

Reference

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