Proposal for HR Employee Attrition Analysis

3.1 Project 1. Draft

Milestone 2

Jennifer Barrera Conde

Bellevue University

Applied Data Science

Sept 12, 2024

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.

From the pie chart, we can see that most of the attrition takes place in the research and development department. From this step, we can also view the data from the job role to discover if it is related to a specific role rather than a department. This is important as it allows the reader to determine if a particular job role should be reviewed.

Attrition by Department
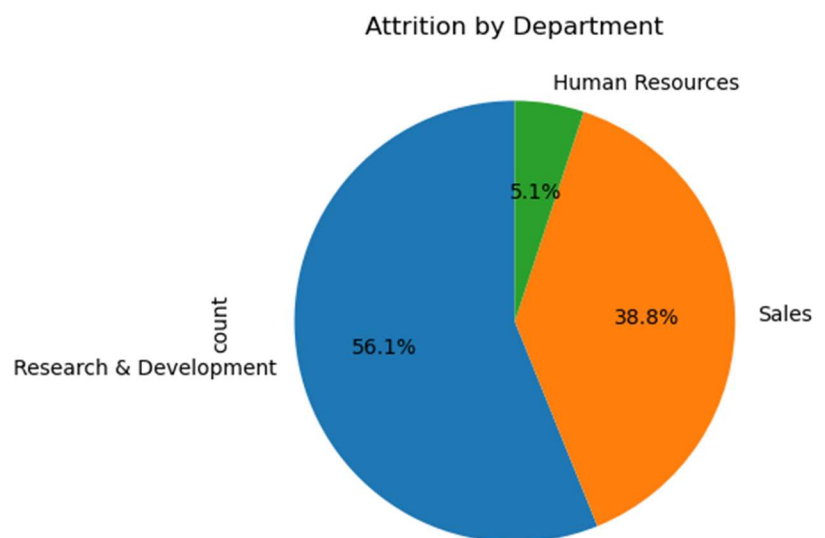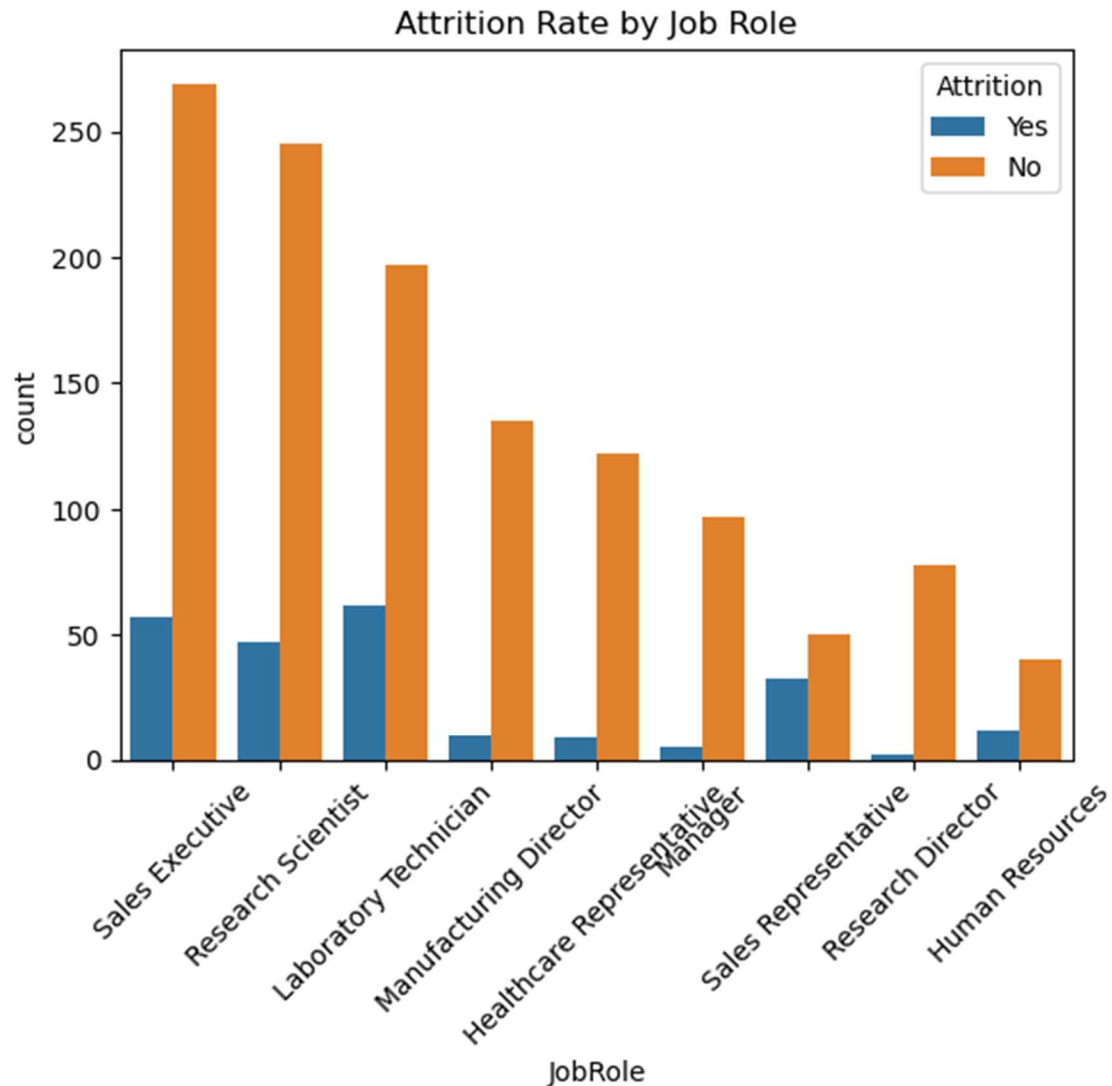
The Job Roles from the Research & Development Department are 'Research Scientist,' 'Laboratory Technician,' 'Manufacturing Director,' 'Healthcare Representative,' 'Research Director,' and 'Manager'.

From these visualizations, one can inherit that positions such as Laboratory Technician, Sales, Executive, and Research Scientist have had the highest attrition, while Research Director, Manager, and Healthcare Representative are the highest. From simple examples such as this, a Company may use this project as an example of what to look for.

From my perspective, I would compare the highest and lowest Job Roles on attrition, all the low to no attrition would be compared, and all the highest attrition would be compared to try and find if there is a meaningful reason that could explain the human behavior of attrition. Another scientist may even compare departments to other departments and many other ideas and assumptions that one would like to prove or disregard.

Background/History

Like my previous explanation, I seek to find and understand Key Drivers of Employee Attrition in Organizations.

Employee attrition has been a longstanding issue in human resource management. Businesses continuously strive to retain top talent, as high employee turnover can significantly impact a company's overall performance and profitability. Over the years, research has shown that job satisfaction, work-life balance, compensation, and career development opportunities are critical in whether employees choose to stay with a company or leave. With the rise of data analytics, companies are now using more advanced tools to predict employee behavior and take preemptive steps to retain key employees.

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.

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. 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.
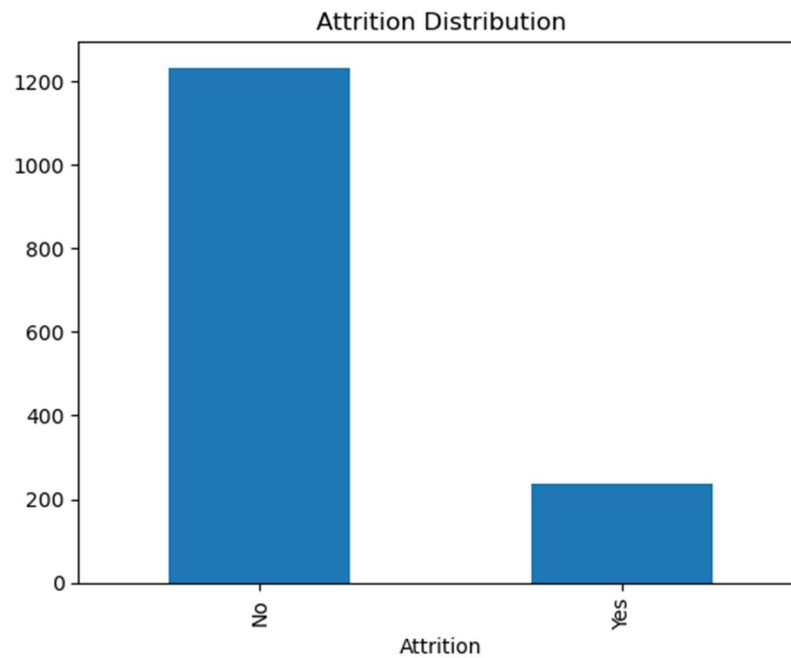
Data Preparation

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.

Distribution of Monthly Income by Attrition
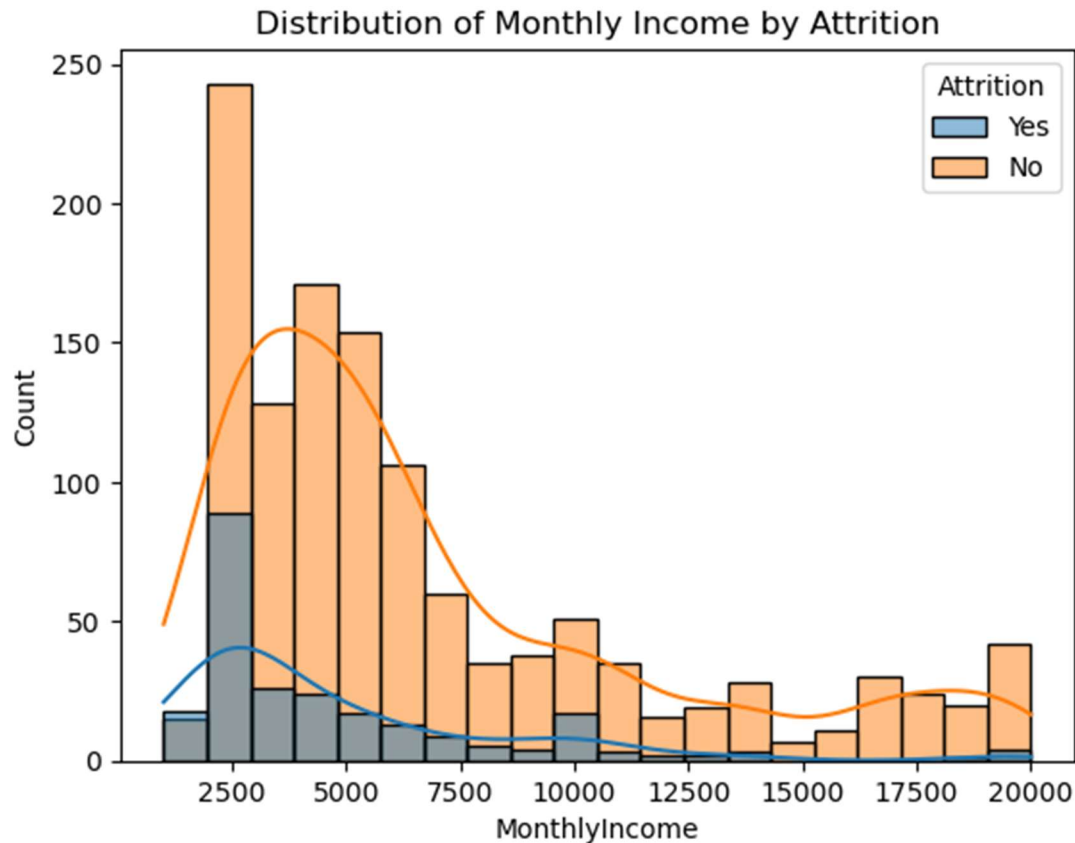
If my previous assumption had been entirely correct, there would be a decline of attrition on higher paying roles where the graph is right-skewed, which is accurate, but there is more to it. Under a quick impression, I would be 100% correct. Still, upon closer inspection, we can see an attrition spike 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

The analysis demonstrates that job satisfaction, work-life balance, and compensation are critical drivers of employee attrition. Companies can reduce turnover by focusing on these areas, implementing policies to improve job satisfaction, and offering competitive compensation packages.

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?

Other questions

1. What were the key variables most strongly associated with attrition?

2. Were there any unexpected findings in the analysis?

3. How did you address potential biases in the data?

4. What retention strategies would you recommend based on the findings?

5. How generalizable are your findings to other industries?

6. What was the biggest challenge you faced during the analysis?

7. How could this model be improved for future use?

8. What is the long-term impact of implementing the recommended retention strategies?

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