

Midterm Data Analysis Project

Manikya Deepika Eadara

G01394383

10/08/2023

1. Introduction.

Many organizations have severe concerns about employee attrition, which is the loss of staff through natural processes like retirement or resignation. It is not just about sustaining the status quo for organizations, especially in the fast-paced digital sector; it is also about assuring the continuity of innovation, productivity, and overall corporate health. A high attrition rate can result in significant expenses and interruptions for an organization due to the costly and time-consuming recruitment and training processes.

The tech corporation with a U.S. headquarters at the center of this study is coping with a rise in employee departures. Retaining personnel has proven difficult, given the high value placed on specialized talents in the tech industry. The company's capacity to maintain its high standards for product quality and operational efficiency has been compromised by this abrupt increase in attrition.

This analysis looks deeply into the nuanced human resource indicators of the organization to identify trends and possible reasons for the recent increase in attrition. Our study aims to shed

light on the elements that could affect an employee's decision to depart using a dataset of 1,200 employee records. By examining factors such as pay, job satisfaction, and frequency of business travel, we intend to provide a comprehensive knowledge of the Situation.

The primary inquiries this research seeks to address are:

1. Is there a relationship between employees leaving the company and their department?
2. Is there a relationship between employees leaving the company and their department and job level?
3. Is there a relationship between employees leaving the company and their years at the company?
4. Does business travel frequency affect employee attrition?
5. Does the distance from home affect employee attrition?
6. Does the distribution of salaries differ between employees who left and those who stayed?
7. Is there a relationship between an employee's performance rating and the decision to leave the company?

The paper then dives into an exploratory data analysis to better understand the fundamental structure and properties of the data. The investigation's findings will be presented in later sections, followed by a thorough analysis and suggestions for the company. This paper aims to deliver practical insights by the end that can direct the business' HR strategies to reduce attrition and promote a more sustainable working environment.

1. Data and Model

- **Data:**

A U.S.-based IT company provided 1,100 records for the dataset used in this study. Each record is an employee, and it has 13 qualities altogether. Personal details like marital status and distance from home are included, as well as professional indicators like work satisfaction, performance evaluations, and length of employment with the organization. The `left_company` variable, which denotes if an employee has left the company, is the main object of our attention.

- **Methods:**

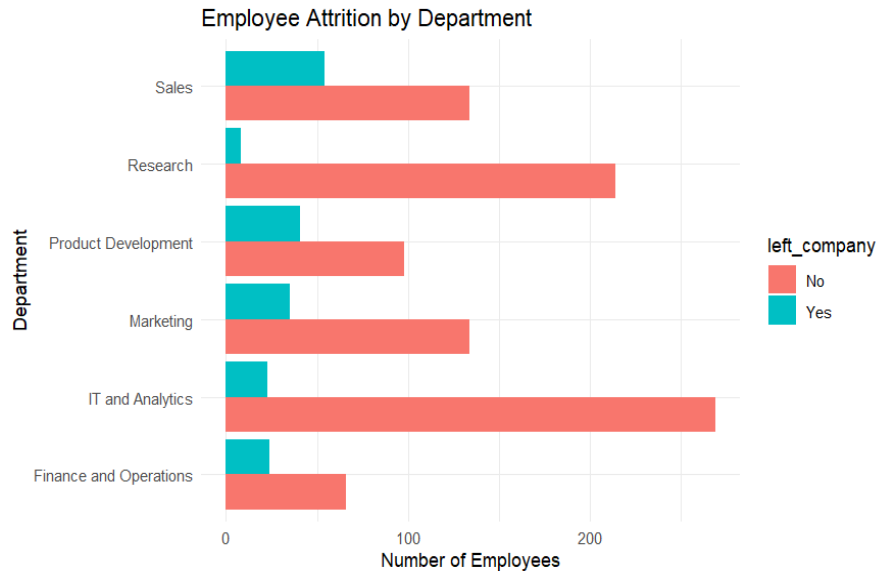
I have used a combination of summary statistics, data visualization techniques, and group-based data analysis to comprehend the complex correlations between various factors and employee attrition. Data visualization using `ggplot2`, data manipulation with `dplyr` and `tidyr`, and summary statistics generation with `skimr` are the primary toolkits for this investigation.

P

- **Analysis:**

1. Is there a relationship between employees leaving the company and their department?

Method: We combined the data based on the 'department' and 'left_company' variables.



Results: The horizontal bar graph labeled "Employee Attrition by Department" shows the variations in attrition rates between various departments. Specific department's attrition rates are higher than others. This can suggest underlying problems or difficulties unique to some departments that could be causing the rise in turnover.

After closer examination of the graph:

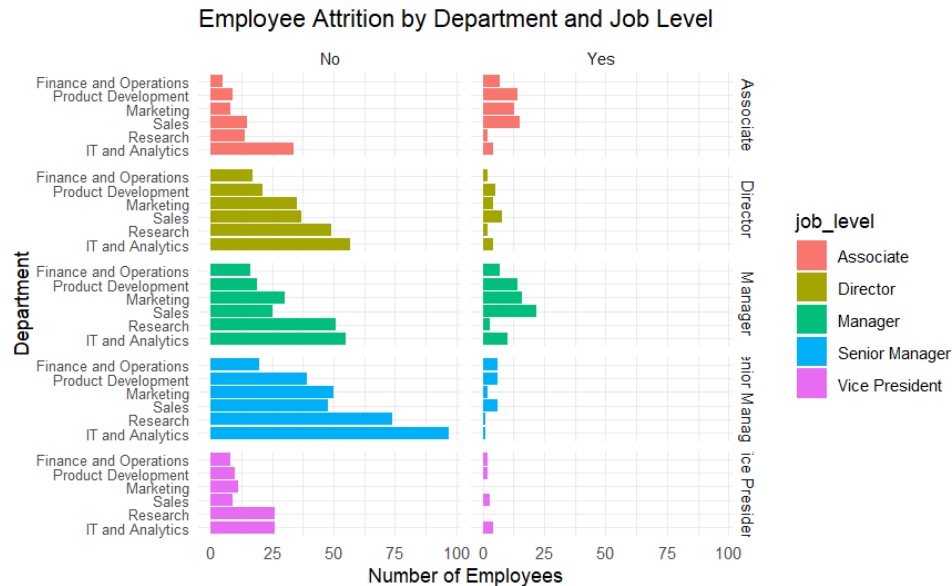
The most significant attrition rates are found in the sales department, which may indicate issues with the department's specific work requirements, sales goals, or other aspects of job satisfaction.

The marketing and product development divisions also see much turnover. The continuously shifting nature of work, project-based employment contracts, or the stress and responsibilities of these professions could all be contributing factors.

In contrast, the Department of Research experiences the most minor attrition. This can indicate a better work environment, more stable employment, or greater job satisfaction in this department.

2. Is there a relationship between employees leaving the company and their department and job level?

Data were organized by "department," "job_level," and "left_company."



Results: The "Employee Attrition by Department and Job Level" graphic sheds light on the positions within departments more prone to attrition. This depth of comprehension might be crucial for developing targeted tactics for keeping talent in particular positions and departments.

Specific findings from the graph:

Sales and Product Development departments experienced the most significant number of departures for the Associate position. On the other hand, the Research department experienced minor associate attrition.

The Sales division once more experienced the highest attrition among individuals having the Director position.

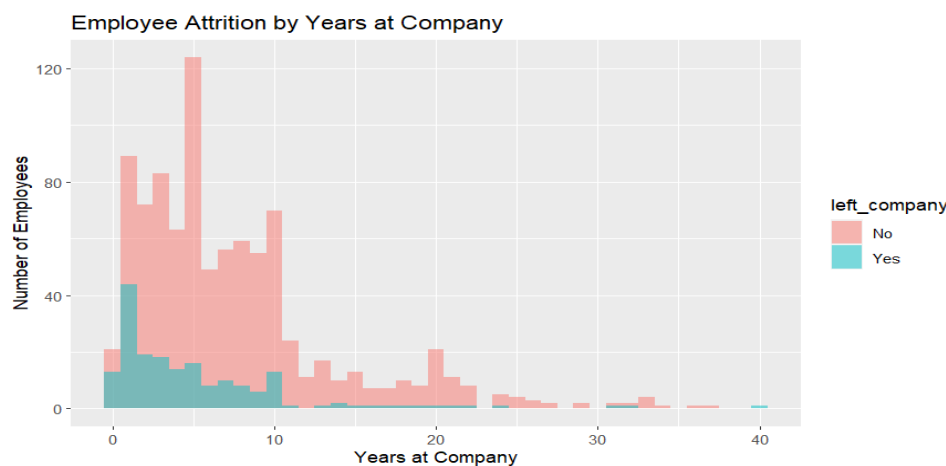
The departments with the most significant employee departures for managers were Sales, closely followed by Marketing and Product Development.

Senior Sales, Product Development, Finance, and Operations managers displayed nearly identical attrition rates, indicating a pattern to consider.

Sales were closely followed by the IT and Analytics department in terms of attrition for the Vice President position.

3. Is there a relationship between employees leaving the company and their years at the company?

Method: We used exploratory data visualization with histograms to evaluate the connection between an employee's tenure (years spent at the organization) and their departure choice.



Results: The "Employee Attrition by Years at Company" histogram shows how the attrition rates change depending on how long an employee has worked.

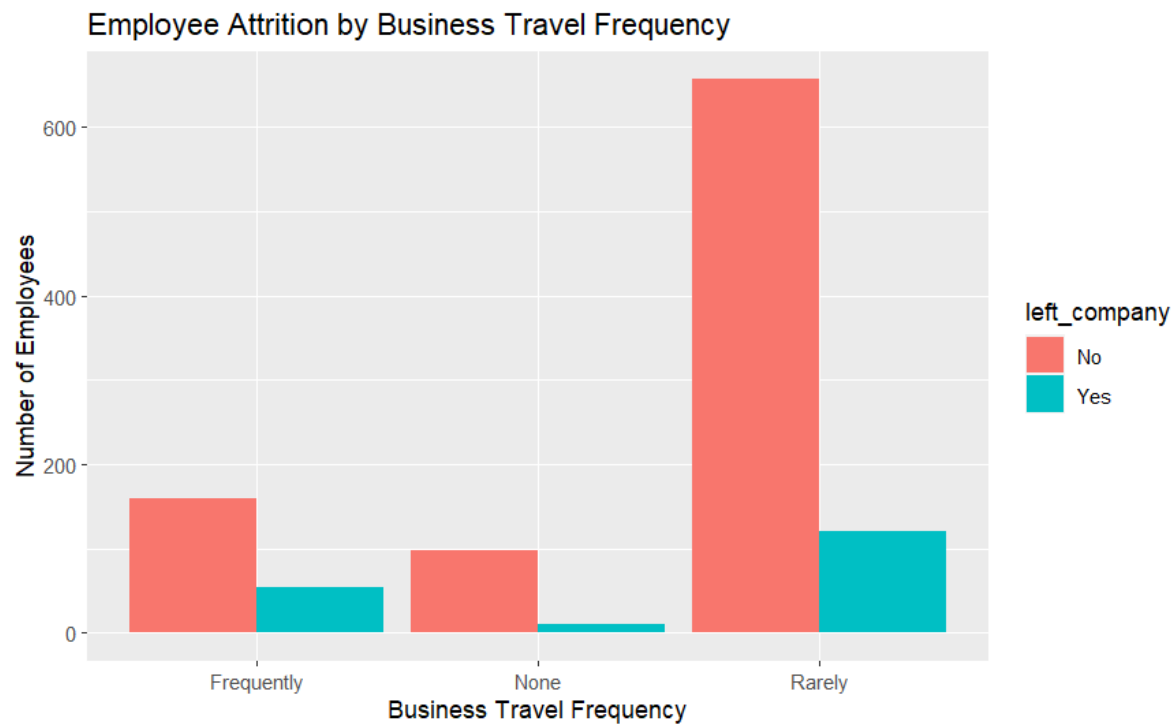
Specific findings from the histogram:

Many workers have worked for the company for 0 to 10 years and exhibit indicators of attrition.

Notably, workers with the organization for roughly 1.5 to 2 years see the highest departure rates.

4. Does business travel frequency affect employee attrition?

Method: The dataset was divided based on the 'business_travel' and 'left_company' variables to examine the association between business travel frequency and attrition.



Results: A perspective on the probable relationship between the frequency of business travel and an employee's decision to leave the organization is provided by the graphic "Employee Attrition by Business Travel Frequency."

Specific observations:

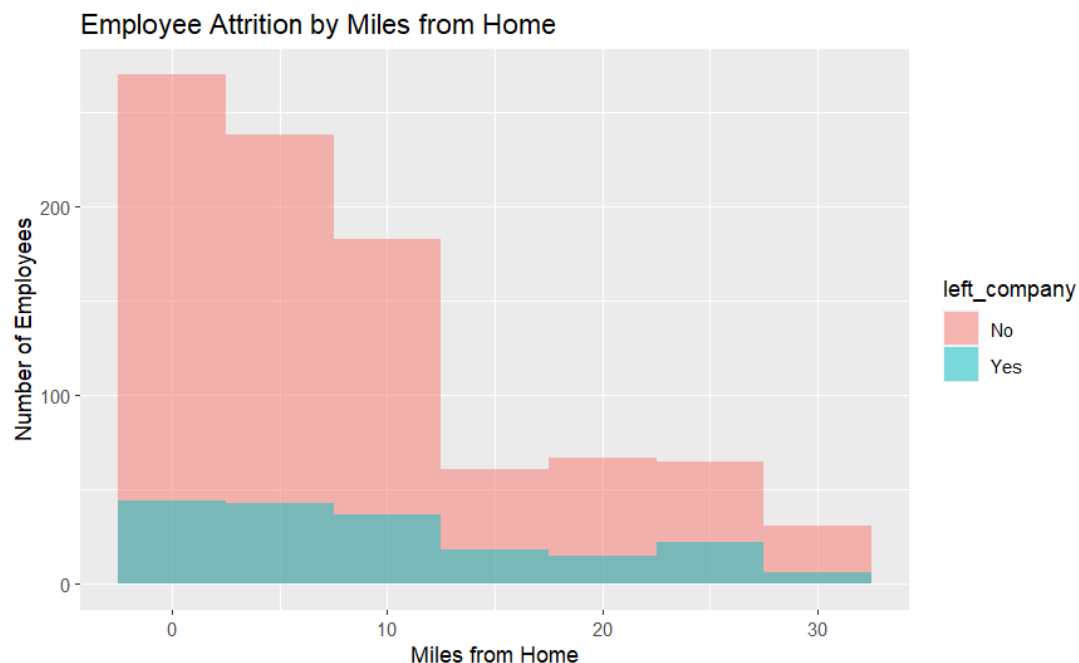
More than 650 employees stated that business travel "rarely" had no bearing on their choice to leave, while more than 100 said it "rarely" did.

While approximately 50 employees indicated business travel "is frequent" and influences their decision to stay or quit, about 175 employees said it is "not frequent" and has no bearing on their decision.

Few employees answered "none" when asked whether or not business travel impacted their decision to quit or stay, demonstrating that for most people, the existence or absence of business travel is not a deciding factor.

5. Does the distance from home affect employee attrition?

Method: Histograms were used to depict data distribution related to an employee's living situation with the company and their choice to quit or remain.



Results: The graphic "Employee Attrition by Miles from Home" sheds light on the relationship between an employee's proximity to their place of employment and their propensity to quit.

specific observations

Most of the departing employees lived less than 15 miles from their place of employment.

6. Does the distribution of salaries differ between employees who left and those who stayed?

Method: A box plot was employed to visualize the distribution of salaries among the employees, explicitly contrasting those who left the company against those who stayed.



Results: The "Distribution of Salaries by Attrition Status" graphic offers insights into the organization's salary structure and possible influence on employee retention.

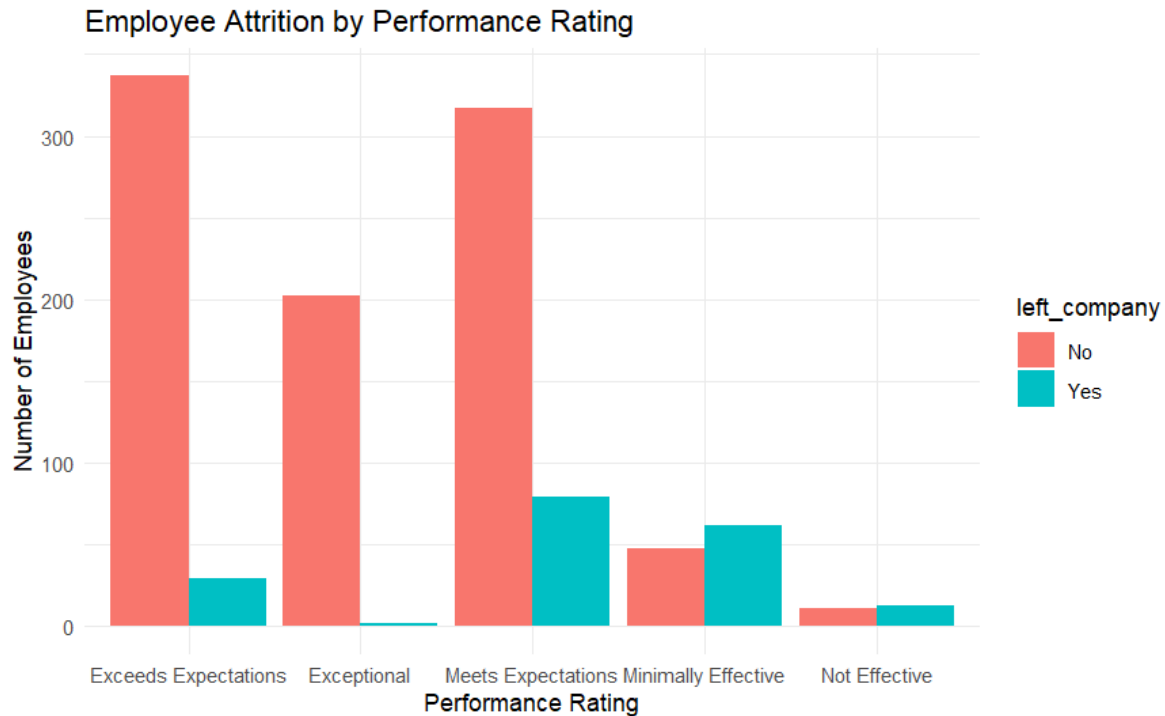
specific observations

Salary ranges for departing employees frequently ranged from under \$ 50,000 to under \$100,000.

In contrast, the earnings of individuals who remained with the company typically ranged from roughly \$75,000 to little less than \$150,000.

7. Is there a relationship between an employee's performance rating and the decision to leave the company?

Method: Data was grouped based on the 'performance_rating' and 'left_company' variables and then visualized to see the trends in attrition based on performance ratings.



Results: The visualization titled "Employee Attrition by Performance Rating" offers an insightful overview of the correlation between an employee's performance and their decision to stay or leave the company.

Detailed observations:

1. Employees rated as **meeting expectations** have the highest attrition rate, suggesting that many average performers are leaving the company.
2. Following them, the **minimally effective** employees show a notable attrition rate. This could be due to dissatisfaction or feeling undervalued.

3. Those rated as **exceeding expectations** are the next in line, indicating that even some high performers feel the need to leave, possibly for better opportunities.
4. The **not** influential group has a lower attrition rate than the above categories, which might indicate the company's ability to provide them with improvement opportunities or support.
5. Lastly, the **exceptional** performers have the lowest attrition rate, suggesting that top talent feels valued and sees growth opportunities within the company.

3. Conclusion(s)/Discussion.

Our study sought to uncover the nuanced factors influencing an employee's decision to leave a tech company. Through detailed data analysis, several key insights have emerged:

1. **Departmental Differences:** Certain departments, like Sales and Product Development, exhibit higher attrition rates. This may hint at potential department-specific challenges or work-related stressors that could be addressed through targeted HR interventions.
2. **Role & Attrition:** Delving deeper into the roles within these departments, it became evident that specific job roles, especially at the Associate and Director levels, are more susceptible to turnover.
3. **Tenure & Attrition:** Employees in the early stages of their tenure, particularly those between 1.5 to 2 years, are likelier to leave, suggesting that the company might need to bolster its employee onboarding and engagement practices during these formative years.
4. **Business Travel:** Surprisingly, business travel frequency only significantly impacts most employees' decisions to stay or leave. However, for a small segment, it does matter, which indicates a need for a flexible travel policy.

5. **Proximity to Work:** The data suggested that employees living closer to the company (within 15 miles) are likelier to leave. This counter-intuitive finding could stem from several reasons, including potential burnout from not having a clear work-life boundary or other regional factors.
6. **Salary Disparities:** A stark contrast in the salary ranges of those who stayed versus those who left was observed, emphasizing the importance of competitive remuneration in employee retention.
7. **Performance and Retention:** The company is retaining its top talent effectively. However, a significant exodus of average and above-average performers points to potential gaps in performance management or growth opportunities.

Future Recommendations:

1. **Customized Retention Strategies:** Tailor HR interventions by department, especially for those with high attrition. For instance, the Sales and Product Development departments could benefit from stress-reduction programs or clearer career growth pathways.
2. **Onboarding Programs:** Strengthen the company's onboarding process, ensuring that employees in their early years feel integrated valued, and see a clear trajectory for their growth within the organization.
3. **Revisit Remuneration:** The company should consider regular salary benchmarking exercises to ensure its compensation packages remain competitive.
4. **Performance Management:** A potential revamp of the performance management system can help recognize and reward the top talent and those who meet and exceed expectations.

Future Directions:

While the current dataset has provided significant insights, a more longitudinal study tracking employee experiences over time could offer an even deeper understanding. Integrating qualitative data, such as exit interviews, would also be beneficial to comprehend the nuanced reasons behind employee departures.

Furthermore, considering external factors, like overall job market trends or regional economic indicators, might give a holistic view of the attrition drivers.

4. Appendix/Appendices.

Detailed tables from the R output:

The following are detailed outputs from some of the R commands executed:

Dimensions of the Data: Output of `dim(employee_data)` will give the dataset's number of rows and columns.

Structure of the Data: Output of `str(employee_data)` provides an overview of the type of each variable and the first few entries in each variable.

First Few Rows of the Data: Output of `head(employee_data)` displays the dataset's first six rows to preview the data.

Detailed Summary of the Data: The command `summary(employee_data)` provides a statistical summary for each variable, which includes the mean, median, minimum, maximum, and quartiles.

Skim Summary: The `skim(employee_data)` command from the `skimr` library provides a more detailed summary than the `essential summary` function, including the number of missing values for each variable.

Plots from R output:

Several plots were generated to analyze different aspects of the data visually:

Employee Attrition by Department: A bar plot showing the number of employees that left the company from each department.

Employee Attrition by Department and Job Level: A more detailed visualization comparing attrition across departments, segmented further by job level.

Employee Attrition by Years at Company: A histogram that displays the frequency distribution of employee attrition based on the years spent at the company.

Employee Attrition by Business Travel Frequency: A bar plot to understand the influence of business travel frequency on the decision to leave the company.

Employee Attrition by Miles from Home: A histogram representing the attrition of employees based on the distance they reside from the workplace.

Distribution of Salaries by Attrition Status: A box plot illustrating the distribution of salaries between employees who left and those who stayed.

Employee Attrition by Performance Rating: A bar plot to understand the relationship between employee performance ratings and their attrition status.

R commands:

The following are the essential R commands and their descriptions used in the analysis:

`Library ()`: Used to load necessary R packages.

`read_excel()`: Reads the Excel dataset into R.

`dim()`, `str()`, `head()`, `glimpse()`: Commands for basic data exploration.

`group_by()` and `summarise()`: Used for grouping data by specific variables and summarizing it.

`ggplot()`, `geom_bar()`, `geom_histogram()`, `geom_boxplot()`: Commands from the `ggplot2` package used for creating various plots.