

Sarah Liang
09/04/2024

The why behind this project

During my job search, I came across a LinkedIn post describing Intuit's most recent wave of layoffs. While this wouldn't usually catch my eye, what was significant about this instance was the fact that Intuit very publicly described 1,050 out of 1,800 total layoffs as "low performers". Understandably, this led to outrage as (1) these public statements can have detrimental effects on the career of the now unemployed job searcher, (2) low employee performance can be due to many interlaced factors such as poor project management, excessive workloads, unclear direction, etc., and (3) low employee performance can be addressed several time prior to immediate layoff.

Regardless of the questionable management methods of Intuit's recent fumble, layoffs are undoubtedly becoming a more and more aggressive cycling season in the tech industry. In order to mitigate the need for aggressive and inhumane layoffs, assessing the natural employee attrition behaviors within a company can aid in mitigating the intensity of layoff cycles. Honest and transparent employee performance and HR analytics to assess employee attrition is evermore needed to have a good control of the hiring periods for a company, especially in the current corporate climates.

Motivating research questions

- 1. How can employee performance be assessed based on different factors?
- 2. How does employee performance and other factors affect employee attrition?
- 3. Can forecasting of employee attrition inform hiring practices?

Datasets

The collection of datasets describing employee attrition and performance at Atlas Labs is sourced from Kaggle:

- [HR Analytics Employee Attrition & Performance](#) from Mahmoud Emad Abdallah on Kaggle

This data is documented until the end of year 2022.

Excel Workbook files

See the [Github Readme](#) for this project for a full description (including variables) of all datasets in the Excel workbook.

File Name	Description
Date Events	Time series dataset with date and counts of events for particular variables, like Hire Count, Leave Count, etc.
EducationLevel	Lookup table for categorical highest completed education level.
Employee	Dataset with employee information.
PerformanceRating	Dataset with performance assessment information.

File Name	Description
RatingLevel	Lookup table for categorical rating level information.
SatisfiedLevel	Lookup table for categorical satisfaction level information.

Data Cleaning and Manipulation in Microsoft Excel

A combination of techniques (VLOOKUP(), Pivot Tables, etc.) were used to combine necessary data into one or two main datasets for further visualization purposes. The **Employee** had the most useful information pertaining to employee demographics, so supplemental performance data from other datasets was pulled and appended to the former, as well as computation of meaningful variables. Here is a general workflow of steps achieved:

1. **VLOOKUP()** was used to pull variable **Education Level** from the **EducationLevel** lookup table.

For example the **Education Level** variable was filled using the formula:

```
= VLOOKUP(  
    Education,  
    EducationLevel!$A$2:$B$6,  
    2,  
    FALSE)
```

where **Education** was a dummy variable and column 2 in the **EducationLevel** lookup table denoted the name of highest accomplished education level.

2. **IF()**, **YEAR()**, **DATE()**, and conditional logic were used to include **Leave Date** for all listed employees. **Hire Date**, **Attrition**, and **Years At Company** were used to calculate **Leave Date**.

An example formula like this was used to add variable **Leave Date**. Recall that the data was collected up to the end of year 2022.

```
= IF(  
    Attrition="No",  
    "",  
    IF(  
        YEAR(Hire Date)=2022,  
        DATE(  
            YEAR(Hire Date) + Years At Company,  
            MONTH(Hire Date) + ((12-MONTH(Hire Date))/2),  
            DAY(Hire Date) + ((30-DAY(Hire Date))/2)),  
        DATE(  
            YEAR(Hire Date) + Years At Company,  
            MONTH(Hire Date),  
            DAY(Hire Date))  
    )  
)
```

For those that left the company before completing one full year of work (i.e. having a hire date in 2022), they are assigned an arbitrary leave date that lands in the middle of the time they had left depending on when they were hired.

For example, someone hired on February 2, 2022 but left the company before completing a year of work would be assigned a leave date with a month exactly in the middle of February and December and a day exactly in the middle of the 2nd and 30th of that month. So, the assigned leave date would be May 14, 2022.

3. The variable **Promotion Date** was calculated similarly with **Hire Date**, **Years Since Last Promotion**, and **Years At Company**, with the following formula for example.

```
= IF(
    AND(
        Years Since Last Promotion < Years At Company,
        Years Since Last Promotion > 0),
    DATE(
        2022 - Years Since Last Promotion,
        MONTH(Hire Date),
        DAY(Hire Date)),
    ""
)
```

4. **Pivot Tables** were used for aggregating purposes over groups, like dates, in order to prepare for the above operations and also construct the **Date Events** dataset.

For example, a Pivot Table grouping along dates and taking the average of **Self Rating** and **Manager Rating** was used to insert such variables into **Date Events**. Other variables like **Hire Count**, **Leave Count**, **Head Count**, and **Promotion Count** used the **COUNTIF()** function in a manner like:

```
Hire Count = COUNTIF(
    Employee!Hire Date,
    Date
)
```

Interactive data visualization Power BI report

Due to the data size, aesthetic reasons, and resources, I want to leverage Power BI to create the final interactive data visualization.

Through data visualization, I would like to convey the (1) company hiring, head count, and performance metrics over time and (2) employee attrition and performance behavior dependent on demographic and other characteristics. These will be displayed through a combination of the following dashboard(s) and plots:

- **Overall Company Dashboard**
 - **time series line plot**

- measures company wise metrics/behavior (head count, hire count, leave count, promotion count, average self rating, average manager rating)
- timeline features: monthly, quarterly, yearly
- grouping features: diversity demographics (ethnicity, gender, age), job role, department
- **Diversity Demographics Dashboard**
 - **bar plots**
 - measuring diversity among total employees, and within departments, job roles
 - convey salary comparisons between job roles, genders, ethnicities, and depending on age
- **Attrition and Performance Dashboard**
 - **line plots**
 - measure attrition rate and/or count over time
 - measure employee performance rating metric, depending on location, department, job role, education level, etc.

Results and meaningful insights

The final Power BI report is in the [GitHub](#) repository for this project.

The final Power BI data dashboard report is in pdf and pbix file format. If you wish to experience the interactivity of the dashboard please download the pbix file and explore with the free application Microsoft Power BI Desktop. Otherwise, the pdf file has sufficient insights and aesthetics.

Some final results, especially regarding the original motivating questions and curiosities, are listed here:

1. How can employee performance be assessed based on different factors?

First of all, employee performance rating can vary depending on whether it is rated by the employee themselves or their manager. For analysis purposes, we continued with the managers' ratings for clarity and less bias. Company-wide-wise, average employee performance over time does not change. However, performance ratings can have a higher variance depending on job role and attrition. When exploring job roles with low attrition, there are more employees not meeting expectations as opposed to the job roles with high attrition. This may indicate for more analysis and exploration needed to understand why people are leaving the company even when they are highly rated in performance. This may be due to reasons unincorporated in the data, such as more promising outside opportunities, unsupportive work environment, etc.

2. How does employee performance and other factors affect employee attrition?

Like mentioned prior, job roles with more varying performance ratings can indicate lower attrition counts within that job role. It is important to note that this company is in a growth period. They are hiring at very high numbers, so much so that they've grown from 0 to about 1200 employees in ten years. So, first and foremost, the company's rapidly rising head count is increasing their attrition rate. Some specific job roles that have the most attrition are data scientists, sales executives, software engineers, and sales representatives. Another factor that may indicate attrition is whether or not an employee works over time. Out of all attrition counts, 53.59% of them worked over time and 46.41% didn't.

3. Can forecasting of employee attrition inform hiring practices?

As we can see on the Overview of Company Behavior and Trajectory report page, employee attrition is not frequent and/or seasonal enough to provide insightful forecasting results. The attrition is very close to zero and very minimally increased over the ten years the company has existed. So, unfortunately forecasting is not feasible or reliable in this situation. On the other hand, simply looking at the company-wide attrition rate over time on the Employee Attrition and Performance report page can tell us more. A good attrition rate should be under 10%. With the growth period of the company, the attrition rate is at a good place. As the head count increases in the future, attrition rate should be expected to increase as well, but preventative measures should be taken to keep attrition rate under 10%. These preventative measures will relate to the factors mentioned before that can impact attrition. Some potential actionables to manage and work with the attrition rate can include:

- exploring possible incentives for employees in job roles with the most attrition (data scientists, sales executives, software engineers, and sales representatives)
- dispersing workload so that employees are not working over time
- introduce a ceiling/cap for hiring rates so that head count can maintain and level out (also leveling out attrition rate) once attrition rate approaches 10%

Utilizing this discovered knowledge can help in efforts to retain employees and manage attrition.