



ABSTRACT

Who is Next to leave the company? Staying proactive to prevent unnecessary interruptions in service due to unexpected departures.

Brian Dunn

Table of Contents

Step 1: Business/Domain Understanding	2
Step 2: Data Understanding.....	2
Step 3: Data Preparation	4
Step 4: Modeling	4
Step 5: Evaluation	5
Sources Cited.....	9

Step 1: Business/Domain Understanding

Who is the next person to leave? And no, this is not survivor!

Turnover is a part of business. When employees leave, they take a lot of unwritten knowledge with them. If companies can find a way to predict when their employees are going to leave, they can better reduce the interruption. The Human Resources department at this company is trying to predict when employees will leave the company, so they can better utilize succession planning.

According to Jon M Werner and Randy L DeSimone in their Human Resource Management textbook, succession planning is “a way of conducting evaluations of employee potential. This process, which is done primarily for senior management positions, requires senior managers to identify employees who should be developed to replace them.”

With the provided information, I am planning to first clean up the data and then use descriptive statistics to better understand if there are ways we can increase the longevity of this company's workforce and reduce turnover. The limited nature of the data set doesn't allow for a great attempt at anomaly detection.

Based on the data attributes provided, the aspects of diversity and inclusion that I am going to focus on are:

- What is the ratio of males to female?
 - Overall
 - At headquarters
 - At store locations collectively
- What is the breakdown of age of the employees?

To help inform when and why people leave the company, I am looking for answers to the following questions as well:

- What is the length of service?
 - Each location
 - Department
- What is the breakdown of why employees left the company?

The section of questions will help set up succession planning in the future.

Step 2: Data Understanding

There are seventeen attributes in this data set. I have listed each of the attributes as well as the necessary information needed to understand each for the purpose of this assessment.

1. Employee ID

- a. The Employee ID is the number internally at the company
2. Record Date
 - a. The Record Day that the information for the employee was entered into the system.
3. Birthdate
 - a. The Birthdate is the employees date of birth.
4. Original Hire Date
 - a. The Original Hire Date that the employee was originally hired at the company.
5. Termination Date
 - a. For employees that are no longer with the company, the Termination Date is the day the employee was terminated.
 - b. For the active employees, this column is listed as '1/1/00'.
6. Age
 - a. This is the employee's age.
7. Length of Service
 - a. This tells the number of years the employee has worked at the company as of the record date.
8. City
 - a. The name of the city in which the employee works.
9. Department
 - a. The department that the employee works in.
10. Job Title
 - a. This is the employee's job title.
11. Store Name/Number
 - a. The number of the store location.
12. Gender
 - a. The gender is the gender of the employee. There are two columns, gender_short ('M' or 'F') and gender_long ('Male' or 'Female'). I filtered the gender_long column out of the data set as it is redundant.
13. Termination Reason
 - a. The termination reason is the more detailed reason for why they left the company.
 - i. For employees that are no longer with the company, this column is Layoff, Resignation, Retirement
 - ii. For active employees, this column is Not Applicable
14. Termination Type
 - a. The termination type tells if the employee was terminated voluntary or involuntary. If they are an active employee, they are listed as "Not Applicable".
15. Status year
 - a. The Status Year is the year that their 10-year anniversary occurred.
16. Status
 - a. The Status states if the employee is a current employee or a past employee.
17. Business Unit
 - a. The Business Unit differentiates the employees who work at corporate versus those that work at the individual stores.

Additionally, there were a few assumptions that made:

- The dataset sets the retirement age is 65.
- This was a somewhat randomly generated data set through a third-party program.
- Since I am not actually a part of this company, I do not know the storylines of different stores or other aspects of the company. If I had that information, it would make it easier to understand if potential outliers had reasonable explanations.
- Since I do not know the qualitative side of this information, I have to make decisions based solely on the numbers. In reality, there are times where the numbers may be a little off, but there is a perfectly good explanation for it.

Step 3: Data Preparation

The data source for this analysis is Kaggle.com. I am going to use Python and the following Python libraries to accomplish the tasks of this project:

- NumPy
- Pandas
- Matplotlib
- Seaborn

The necessary file(s) are saved on my desktop. I uploaded them into JupyterLab, filtered out columns that were not going to be used, and renamed remaining columns to make them reader-friendly. I checked for missing data, and found all of the information was in the files.

Additionally, I checked for duplicates, and found there was an additional record for each year an employee was at this company. To clean this up, I sorted and then filtered out old records, keeping only the most current record for each employee. I also corrected a spelling error on a city name.

Finally, I exported a copy of the final file to verify the updates worked and to create a backup of the file.

Step 4: Modeling

For the descriptive statistics, I ran a describe function with a few added percentiles. I included the median function, a correlation function, and a heatmap of the correlation values.

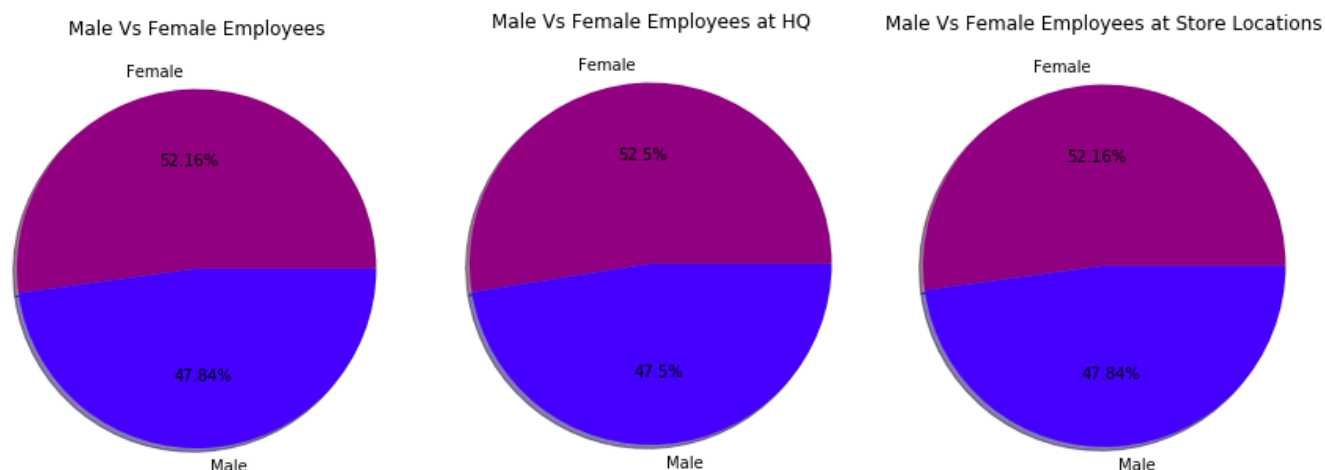
Additionally, I ran some descriptive statistics for the following groupings within this data:

- Gender ratio (male to female)
- Gender ratio based on the business unit they worked in (at individual store locations, or at headquarters)
- Termination Reason
- City Name (mainly for each individual store)
- Business Unit, then the Department Name
- Term Type then Term Reason

In addition, I created many graphics to make it easier to digest the information.

Step 5: Evaluation

The overall male to female ratio is 3006/3278. There are more females working for the company than males, and there are more females working at both the stores (in total) and headquarters than males.

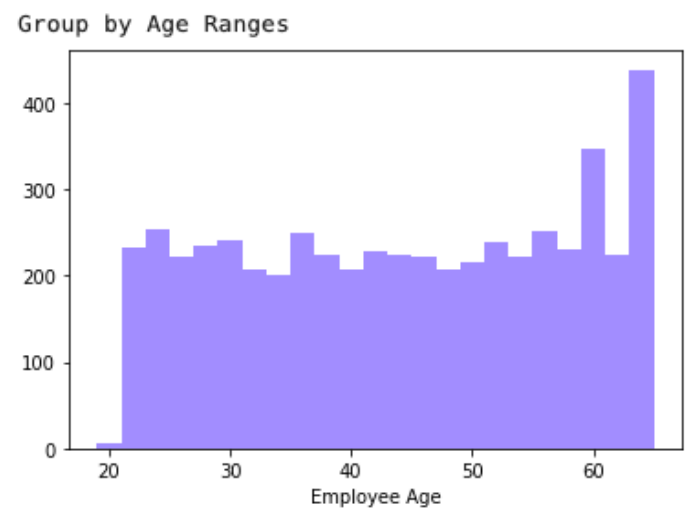
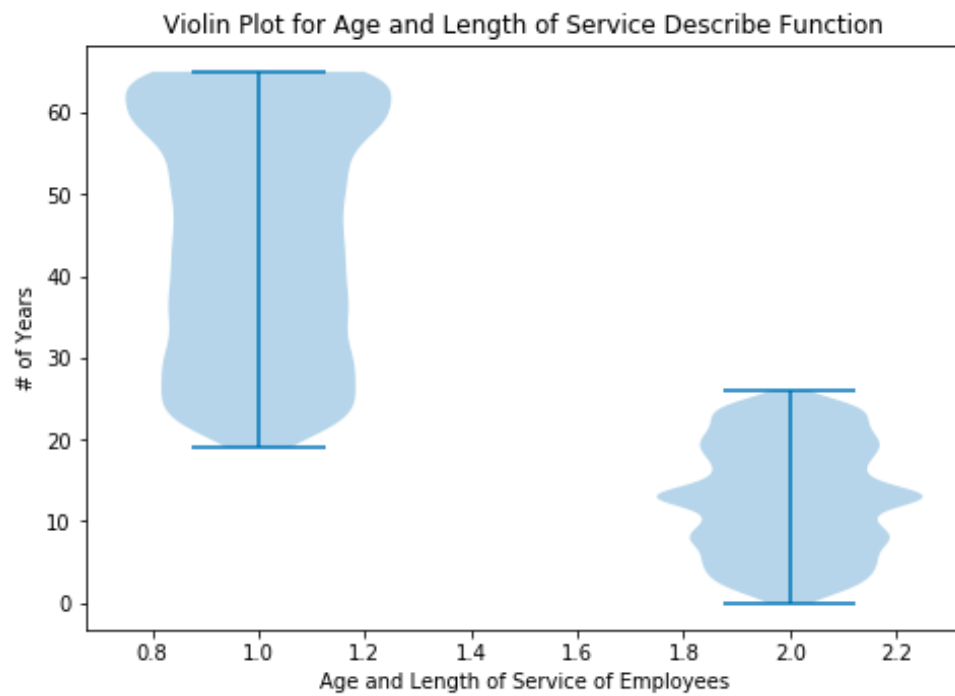


Additionally, females stay at the company for eight months longer than males on average. It is good for the continuity of business and the retention of knowledge within the company that there are more females than males. We should consider having a higher female to male ratio of our workforce.

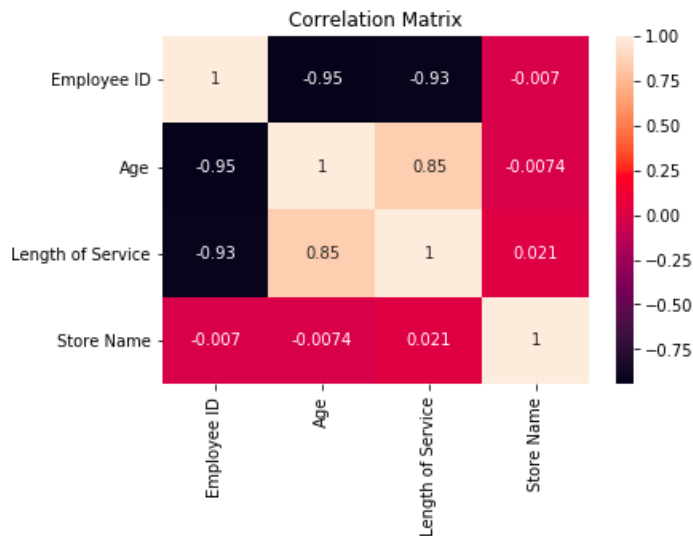
Next, the age range of the employees is very consistently distributed from about 21 years old to 65 years old. There are a lot of people with a lot of experience. The only downside of this (and this is a downside to have) is that at some time, those more experienced employees are going to retire in the near future. Now is time to start implementing succession planning. We do not want to watch experience and knowledge walk out the front door before they can pass at least some of that knowledge to their eventual successors. No one is getting pushed out, succession planning is purely to make sure that when employees retire or leave the company, that we are properly prepared to continue our excellence.

After that, the length of service for employees at the individual store locations averages to over 12.6 years overall. That is excellent, especially considering that the overall average length of service for the company is 12.83 years. Almost all of the stores have very similar averages. Most of the stores had averages in the ten- to fourteen-year range.

I noticed that the average length of service at the Valemount location is about two standard deviations lower than the average. There may well be a perfectly fine reason for that, but we should have a few conversations with both the management and employees at that location to make sure. In the graphics below, you will see that the distribution of both age and length of service for employees at the company is consistent.

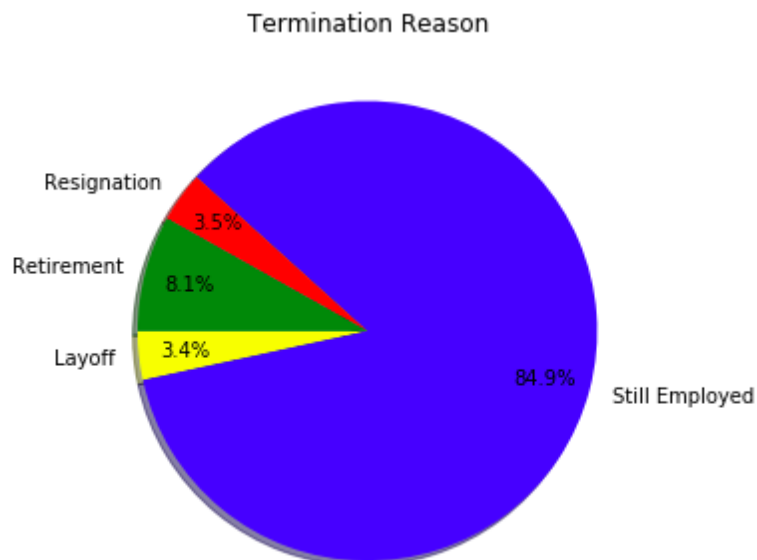


Nothing with the correlations stick out as good or bad.



The turnover rate is low. Most employees who leave do so because they choose to retire. Given the length of time that this information represents and the industry (retail), the turnover is in good shape. We can always do better!

Finally, the employee retention based on department is excellent. The two departments that I would like to see more information about are Customer Service (at the store location level) and Processed foods. While the retention is excellent, they are the two lowest values and relative outliers. There could be perfectly acceptable reasons for the shorter length of service in those two departments, but it is worth checking them out further.



Conclusion

Here is a quick summarization of the results:

- Females have a longer length of service with the company, which equates to about 8 months longer than males on average.
- The most common reason for employees departing the company is resignation (8.1%), followed by Resignation (3.5%) and Layoff (3.4%).
- The individual locations have excellence employee retention overall.

Next Steps

First, we should dedicate some time to understand what is happening within the Processed Foods and Customer Service departments as well as at the Valemount store location. While these areas are not necessarily struggling, they are the outliers. At the same time, we hire many great employees, we should approach this from the standpoint that everyone is doing the best that they can, so what can we do to make their jobs easier.

For succession planning, the first step is to complete an audit and update the education for each employee. Once we update the education information in the HRIS, we can start promoting the opportunities for upward mobility in the company and use that as additional incentive for employees. Additionally, we can start with succession planning for employees in all facets of the business.

Based on the results, we should look for employees who have worked at the company for at least five years as that is just above the average tenure of employees who leave the company by reason of resignation. We should not only continue the trend of having more females than males in our workforce, but also increase the female to male ratio.

Sources Cited

Employee Attrition: Can you forecast employee attrition?

<<https://www.kaggle.com/HRAnalyticRepository/employee-attrition-data>>

Werner, J. & DeSimone, R. (2006) Human Resource Development (4th edition). Thomson South-Western (ISBN: 0-324-31978-9)