Midterm Analysis

Patrick T Doyne

2023-06-10

```
#R Libraries
library(ggplot2)
library(tidyverse)

## ── Attaching core tidyverse packages ────────────────────────
tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ lubridate 1.9.2     ✔ tibble    3.2.1
## ✔ purrr     1.0.1     ✔ tidyr     1.3.0
## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to
force all conflicts to become errors

library(readxl)
options(dplyr.summarise.inform = FALSE)
```

Introduction:

The following data consists of 1,200 employee records for a U.S. based product company. It also contains a column for if the person left the company, their salary, department, job level, weekly hours, years at the company, years since last promotion, performance, job satisfaction, if they travel for work, are married, how far their home is from the job, and lastly the other jobs they worked at.

```
emp_data <- read_excel("Downloads/Employee Data Formatted - MIS 431
Summer 2023 (1100 Records) f.xlsx")
emp_data

## # A tibble: 1,100 × 13
##    left_company department         job_level salary weekly_hours
business_travel
```

```
##    <chr>         <chr>                <chr>      <dbl>         <dbl>
<chr>
##  1 Yes          Sales                Director  1.19e5            56
Rarely
##  2 No           Sales                Senior M… 8.56e4            42
Frequently
##  3 Yes          Product Developme… Associate 4.62e4            56
Rarely
##  4 No           IT and Analytics     Director  1.17e5            50
Frequently
##  5 No           Sales                Associate 3.66e4            46
Rarely
##  6 No           Marketing            Senior M… 8.35e4            48
Frequently
##  7 No           Marketing            Senior M… 8.86e4            44
Rarely
##  8 No           Sales                Director  1.22e5            47
Rarely
##  9 No           Finance and Opera… Senior M… 9.46e4            50
Frequently
## 10 No           Product Developme… Director  1.25e5            51
Rarely
## # ℹ 1,090 more rows
## # ℹ 7 more variables: yrs_at_company <dbl>, yrs_since_promotion
<dbl>,
## #   previous_companies <dbl>, job_satisfaction <chr>,
performance_rating <chr>,
## #   marital_status <chr>, miles_from_home <dbl>
```

The Problem:

We are concerned with using employee data to optimize objectives such as employee satisfaction, productivity, hiring and employee attrition. Ideally, companies would like to keep attrition rates as low as possible due to costs and business disruptions that come with replacing productive employees on short notice.

The Big Questions Being Asked:

-Does the department being worked in have any impact on employees leaving the company? Which are the worst? -Do those who make more and have higher job ratings tend to leave as much as those who don’t? -Is there an element of burnout? Are those leaving the company people who are overworked -Does this distance traveled by the employee affect retention? -Is there a benchmark amount of years since a promotion that seems to cause people to quit? -How effective is the company survey? Do those who post being happy actually stay with the company?

Whats to Come?

The rest of this briefing will show how we used the data in R, through both visuals such as graphs, and summary tables. This will cover the methods used, as well as an analysis of our results for each of the listed questions above.

Analysis 1:

Does the department being worked in have any impact on employees leaving the company?

What I Found:

Finance and Operations lost about 26% of workers. IT and Analytics held stronger losing only 7.8% of workers. Marketing lost around 20% of workers. Prdouct Development lost the most with about 30% of workers quitting. Research lost a fraction of a percent of workers, and Sales lost nearly 29%. The three worst performing are Finance and Ops, Sales, and Product Development. Another to keep and eye on is Marketing.

```
emp_data%>%
  group_by(department, left_company)%>%
  summarize(count = n())

## # A tibble: 12 × 3
## # Groups:   department [6]
##    department             left_company count
##    <chr>                  <chr>        <int>
##  1 Finance and Operations No              66
##  2 Finance and Operations Yes             24
##  3 IT and Analytics       No             269
##  4 IT and Analytics       Yes             23
##  5 Marketing              No             134
##  6 Marketing              Yes             35
##  7 Product Development    No              98
##  8 Product Development    Yes             41
##  9 Research               No             214
## 10 Research               Yes              8
```

```
## 11 Sales                     No              134
## 12 Sales                     Yes              54
```

##What I Recommend: It is clear the department plays a large role in workers leaving. I recommend you look into Finance and Ops, Sales, and Product to development. There is likely an issue with the managers in charge of these departments, or the training/ onboarding the new employees of this department are recieving. I would also allow the resources to do the same for the Marketing department after the first three are handled effectively and turnover numbers in the departments improve.

Analysis 2:

Do those who make more and have higher job ratings tend to leave as much as those who don't?

What I Found:

Employees in the upper three levels of performance tend to leave the company when they are extremely underpaid. You have employees deemed as not effective or minimally effective who are being paid more.

```
ggplot(emp_data, aes(x = left_company, y = mean(salary), color =
left_company))+
  geom_col()+
  facet_wrap(~performance_rating)+
  labs(title="Wages and Ratings of Employees vs Retention", y ="Mean
Salary",x="Quit the Company")
```

What I Recommend:

Cutting lose those with poor performance and using the saved capital to increase the wages of higher performing under paid employees. This will likely increase employee retention over time, as well as foster a more productive working environment. Short term sacrifice long term gain.

Analysis 3:

Is there an element of burnout? Are those leaving the company people who are overworked

What I Found:

The average weekly hours worked for those who left the company was a staggering 58.7, and for those who stayed 48.4. When looking at the scatter plot, the outliers for those who stayed was the bottom quartile for those who left, and the lower end outliers for those who did leave the company were right around the third quartile of those who stayed. This shows there is clearly an element of employee burnout, particularly starting at around the 50 hour mark.

```
emp_data%>%
  group_by(left_company)%>%
  summarize(avg_weekly_hours=mean(weekly_hours), total_emps = n())

## # A tibble: 2 × 3
##   left_company avg_weekly_hours total_emps
##   <chr>                   <dbl>      <int>
## 1 No                       48.4        915
## 2 Yes                      58.7        185

ggplot(emp_data, aes(x=left_company,y=weekly_hours))+
  geom_boxplot()+
  labs(title="Burnout for Employees", y ="Number of Hours Worked Per
Week",x="Left the Company")
```

What I Recommend: Reducing the hours of all employees down to a maximum of 50 hours. Prioritizing the highest rated and most efficient employees first. This will increase retention rates and reduce burnout.

Analysis 4:

Does the distance traveled by the employee to the job site affect retention?

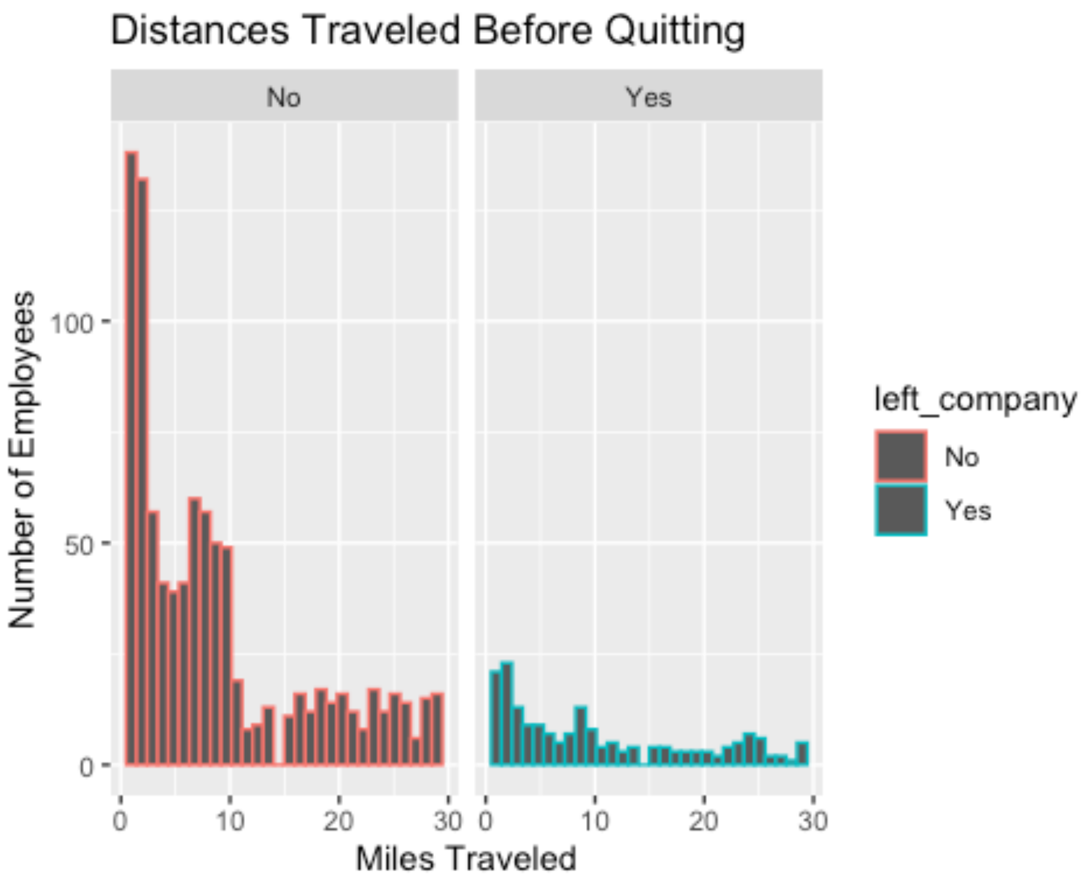
What I Found:

Those who chose to leave the company had on average 1.52 miles further to drive. Employees who stayed averaged a drive of about 8.78 miles while those that left averaged 10.3 miles. These numbers comes from pools of 915 employees and 185 employees respectively. Additionally, when looking at the data within the histograms for those who left and stayed at the company you can see that at every level of distance where someone left the company, there was a higher frequency of employees who stayed.

```
emp_data %>%
  group_by(left_company)%>%
  summarize(avg_miles_from_home = mean(miles_from_home), total_emps =
n())

## # A tibble: 2 × 3
##   left_company avg_miles_from_home total_emps
##   <chr>                      <dbl>      <int>
## 1 No                          8.78        915
## 2 Yes                        10.3         185

ggplot(emp_data, aes( x = miles_from_home, color =left_company))+
  geom_histogram(bins=30)+
  facet_wrap(~left_company)+
  labs(title="Distances Traveled Before Quitting", y ="Number of
Employees",x="Miles Traveled")
```



What I Recommend:

Although there are no distances that seem to "force" employees to leave the company, hiring people within the 0-10 mile range will likely lead to better employee retention. The highest frequencies in which employees didn't leave when compared to those who did are within this range of 0-10 miles.

Analysis 5:

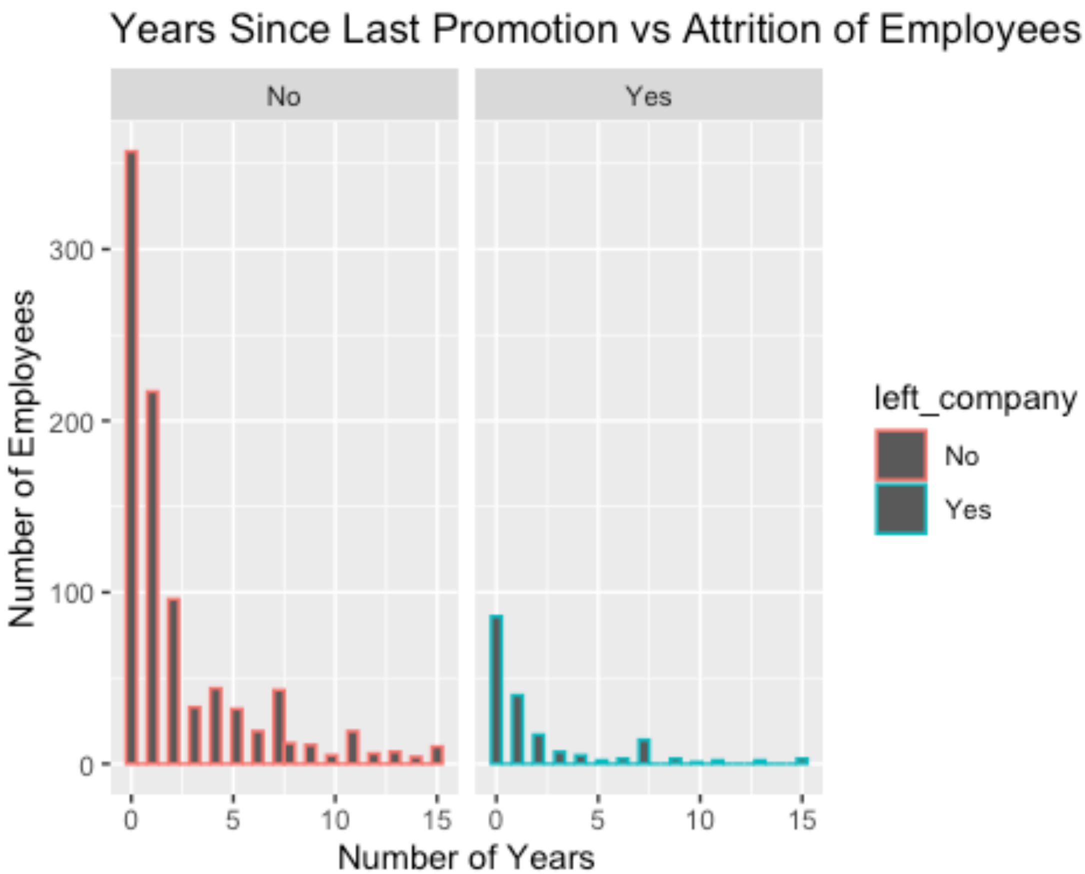
Is there a benchmark amount of years since a promotion that seems to cause people to quit?

What I Found:

The distribution is relatively the same across employees who left the company and those who stayed. It seems the time since promotion is not a major factor in employee attrition.

```
ggplot(emp_data, aes( x = yrs_since_promotion,color=left_company))+
  geom_histogram()+
  facet_wrap(~left_company)+
  labs(title="Years Since Last Promotion vs Attrition of Employees", y
="Number of Employees",x="Number of Years")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



What

I Recommend: Continue to promote employees at the rate, standard, and levels you currently are. It wouldn't hurt to rehighlight expectations for promotion to fuel employee effectiveness and improve attrition rates a little bit.

Analysis 6:

How effective is the company survey? Do those who post being happy actually stay with the company?

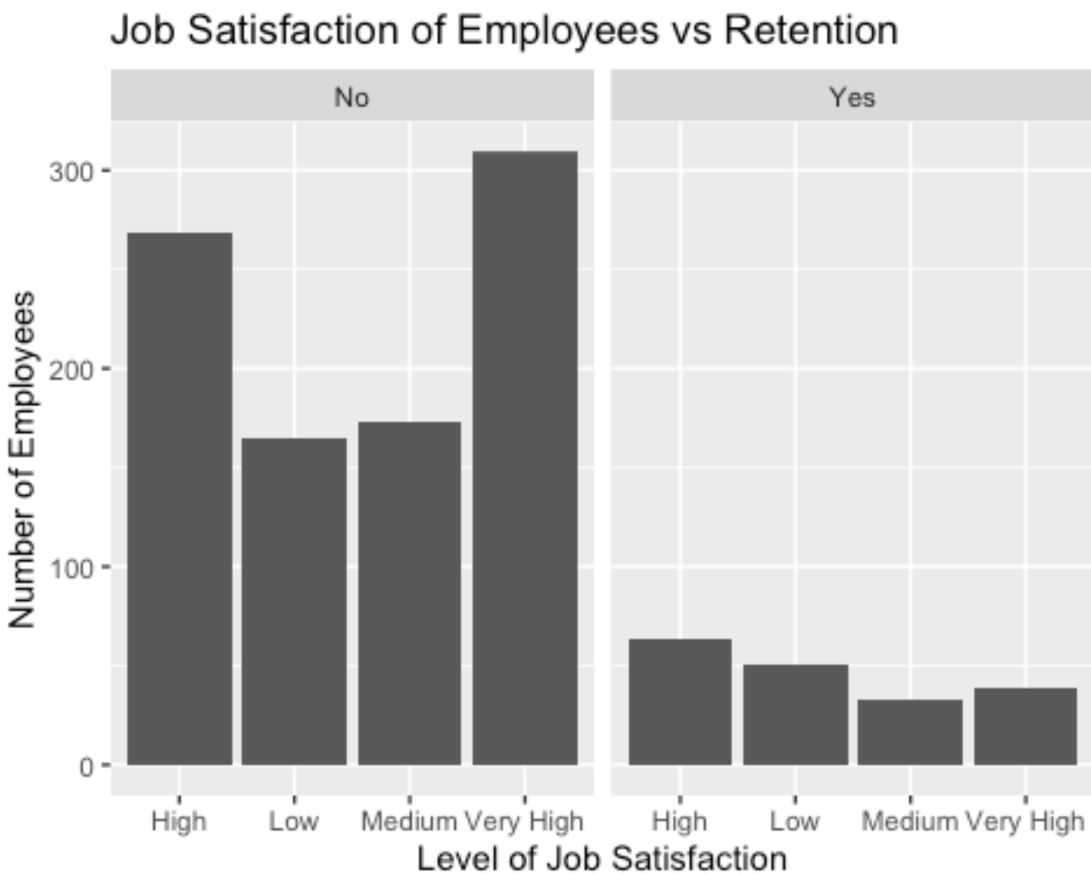
What I Found:

The company survey doesn't seem to be a functional/viable one in regard to employee attrition. There were more people who posted high and very high satisfaction levels that quit than those who posted low and medium and quit as well. This point to an ineffective data sampling technique, or employees not taking it seriously.

```
emp_data_6 <- emp_data%>%
  summarise(left_company=left_company,
job_satisfaction=job_satisfaction, count = 1)
```

```
## Warning: Returning more (or less) than 1 row per `summarise()`
group was deprecated in
## dplyr 1.1.0.
## ℹ Please use `reframe()` instead.
## ℹ When switching from `summarise()` to `reframe()`, remember that
`reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this
warning was
## generated.
```

```
ggplot(emp_data_6, aes(x = job_satisfaction, y=count))+
  geom_col()+
  facet_wrap(~left_company)+
  labs(title="Job Satisfaction of Employees vs Retention", y ="Number
of Employees",x="Level of Job Satisfaction")
```



What I Recommend: Create a new method that encompasses key factors of job satisfaction, and is able to accurately report it. One way is to create incentive around the method used to gather data, as this well help give better quality data. You can then use this to make adjustments to the organization on a need be basis.

Conclusion:

To conclude, there are many things that can be done to reduce employee attrition. We dialed in on a few that seem to be the most pressing: increasing pay for employees who are out performing their peers and remaining underpaid, improving training and getting new department managers for those with the highest rates of attrition, reducing the number of hours worked weekly to a maximum of 50 to eliminate burnout and aid in employee retention, creating a new method for recieving feedback on job satisfaction that is incentivized by extrinsic rewards to ensure quality data that is far more extractable for solutions to problems as they come, and finally only hiring people who are within 0-9 miles of the job site. On the other hand, we had one analysis that yielded a minimal connection to employee attrition; years since a promotion and its effect on employees. Going forward, my strongest recommendation is to conduct more analysis and create predictive models, this will put the company in the best position to be proactive in employee retention.