Mid-Term Data Analysis Project: Employee Attrition

Harshil Saini

03/20/2023

## 1. Introduction.

For this company report, we have gathered data from our human resources team to determine the leading factors of employee attrition within our company. We have gathered data of employees in our company such as whether they have left the company, department in which they work in, job level of their current occupation, salary, weekly hours, business travel, years at the company, years since promotion, previous companies, job satisfaction, performance, marital status, and distance from home to see if these factors have any correlation to the employee attrition data we have acquired. After carefully analyzing the data, we had received from the Human Resources team, we have decided to focus on six main factors that may influence attrition at our company. The six factors are job satisfaction, salary, weekly hours, number of previous companies, years at our company, and distance of commute.

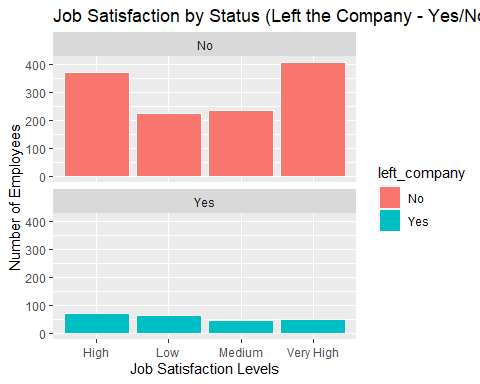
After carefully analyzing the data, we had found that the three factors that most contributed to employees leaving the company were salary distribution, number of companies worked for previously, and years served at the organization among employees. Research shows that factors such as job satisfaction, weekly hours worked, and commuting distance did not have a strong enough correlation in the data to draw conclusions about our attrition rates. Generally, employees that made less money yearly, have previously worked for many different companies, or have been with the company for less time were more likely to leave our company. For the remainder of this report, we will cover the data and model section of our chosen factors, the conclusions drawn from our data, and the appendices/data visualization that allowed us to draw the conclusions about our problem of attrition rates in our organization.

## 2. Data and Model

As mentioned in the introductory portion of the company report, we will be analyzing data regarding job satisfaction, salary, weekly hours, number of previous companies, years at our company, and distance of commute of the employees at our company to determine whether any of the listed factors have any correlation to the number of employees our company is losing. Here are our big questions as follows:

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

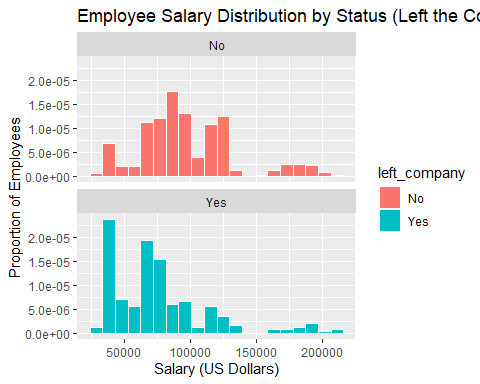
## # A tibble: 2 x 7  
## left\_company n\_employees min\_job\_satisfacti~ avg\_job\_satisfa~ max\_job\_satisfa~  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 No 1233 1 2.78 4  
## 2 Yes 237 1 2.47 4  
## # ... with 2 more variables: sd\_job\_satisfaction <dbl>,  
## # pct\_less\_medium\_satisfaction <dbl>



Answer: Yes, the data does make a slight indication that there are more employees that leave the company who have lower job satisfaction than those that have higher job satisfaction. Our data suggests that job satisfaction among employees that had stayed was averaged to be 2.78/4 (~70% satisfied with their job). According to our data, job satisfaction among employees who had left our organization was slightly lower at 2.47/4 (~62% satisfied with their job). In summary, data suggests that employees who had left the company were ~8% less satisfied with their jobs on average.

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

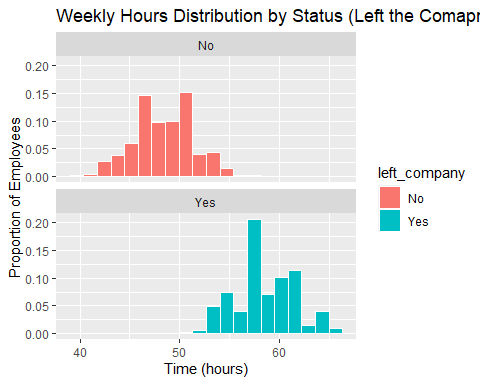
## # A tibble: 2 x 7  
## left\_company n\_employees min\_salary avg\_salary max\_salary sd\_salary  
## <chr> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 No 1233 29849. 97431. 212135. 36470.  
## 2 Yes 237 30488. 76626. 211621. 38567.  
## # ... with 1 more variable: pct\_less\_60k <dbl>



Answer: Yes, the data does make a strong indication that there are more employees that leave the company who make less annual salary than those that make more salary annually. Our data suggests that annual salary among employees that had stayed was averaged to be ~$97,430 annually while employees that chose to leave had an average salary of ~$76,625 annually. That is ~27% difference in annual salary between averages. In summary, due to the large difference in averages, it is safe to conclude that on average, employees making less salary annually are more likely to leave the company than those that make more.

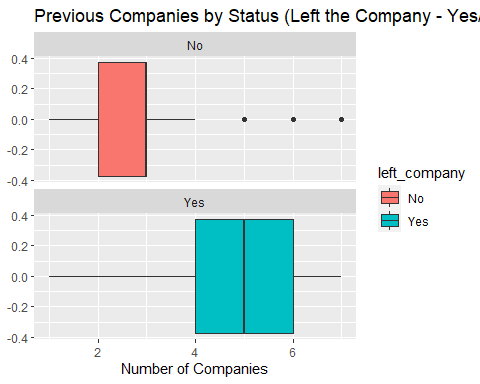
1. Is there a relationship between employees leaving the company and the number of hours worked weekly?

## # A tibble: 2 x 7  
## left\_company n\_employees min\_weekly\_hours avg\_weekly\_hours max\_weekly\_hours  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 No 1233 40 48.4 58  
## 2 Yes 237 51 58.6 66  
## # ... with 2 more variables: sd\_weekly\_hours <dbl>, pct\_less\_55\_hours <dbl>

 Answer: Yes, the data does make a slight indication that there are more employees that leave the company who overwork than those that work less hours on a weekly basis. Our data suggests that weekly hours worked among employees that had stayed was averaged to be ~48 hours a week while those that had left the company were averaging ~59 hours per week. Again, that is ~23% increase in hours from those who did not leave the company. In summary, the average hours worked weekly by those who left the company as compared to those who stayed in the company indicates that there is a higher chance of employees leaving if overworked.

1. Is there a relationship between employees leaving the company and the number of previous companies they worked for?

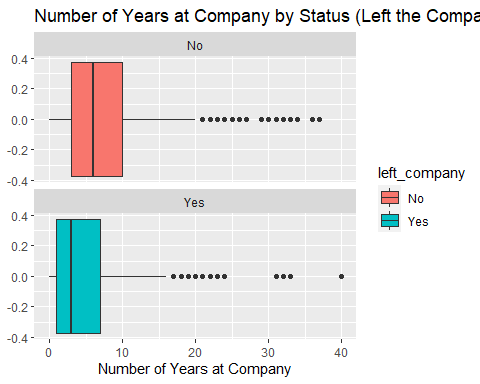
## # A tibble: 2 x 7  
## left\_company n\_employees min\_previous\_compa~ avg\_previous\_co~ max\_previous\_co~  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 No 1233 1 2.97 7  
## 2 Yes 237 1 4.65 7  
## # ... with 2 more variables: sd\_previous\_companies <dbl>,  
## # pct\_less\_4\_companies <dbl>



Answer: Yes, the data does make a strong indication that employees are more likely to leave the company if they have also worked for many other companies previously. Our data suggests that companies worked for among employees that had stayed was averaged to be ~2.97 previous companies while number of previous companies of employees that had left averaged at ~4.65. In other words, employees that left our company saw ~57% increase in number of companies worked for previously than those who had stayed. These observations allowed our team to conclude that employees that have worked for more than three companies previously were more likely to leave our organization.

1. Is there a relationship between employees leaving the company and the number of years served at our organization?

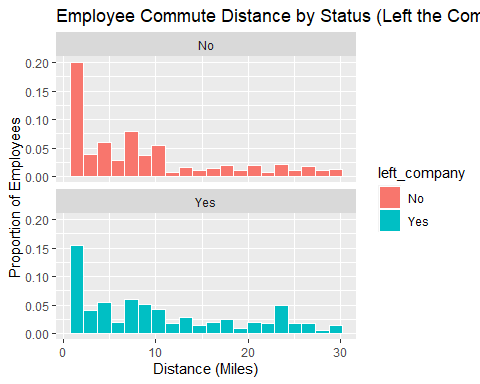
## # A tibble: 2 x 7  
## left\_company n\_employees min\_yrs\_at\_company avg\_yrs\_at\_compa~ max\_yrs\_at\_comp~  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 No 1233 0 7.37 37  
## 2 Yes 237 0 5.13 40  
## # ... with 2 more variables: sd\_yrs\_at\_companys <dbl>, pct\_less\_15\_years <dbl>



Answer: Yes, the data does make a strong indication that there are more employees that leave the company who have stayed with the company for less time than those who have stayed for longer. Our data suggests that length of service among employees that had stayed was averaged to be ~7.37 years while average years served at the company decreased by ~30% to ~5.13 years served at the company. This portion of our data indicated that the longer employees worked for our company, the more likely they were to stay with the company.

1. Does the commute distance for employees at our company have any correlation to the number of employees lost at our organization?

## # A tibble: 2 x 7  
## left\_company n\_employees min\_miles\_from\_home avg\_miles\_from\_~ max\_miles\_from\_~  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 No 1233 1 8.92 29  
## 2 Yes 237 1 10.6 29  
## # ... with 2 more variables: sd\_miles\_from\_home <dbl>, pct\_less\_15\_miles <dbl>



Answer: Yes, the data does make a slight indication that there are more employees that leave the company who must commute further rather than those who have a closer commute. Our data suggests that the average commute in miles for employees that have stayed with our company is ~8.92 while the average for people that have left the company was ~10.63 miles. That is ~19% increase in distance of commute on average. In summary, employees are more likely to stay if their commute distance is shorter.

## 3. Conclusion(s)/Discussion.

For the organizational report, we have explored the topics of employee attrition at our company and have done analysis on many different factors to determine if these factors have an influence on the company’s ability to retain employees. In this report, we have explored the factors of job satisfaction, salary, weekly hours, number of previous companies, years at our company, and distance of commute of the employees at our company. As shown from our analysis and the data we have acquired from our Human Resources team, our study found that employees earning less money annually, employees that have worked for more companies previously, and employees that have worked for our company for a shorter amount of time were all more likely to leave our organization.

These results found in our analysis were very significant to our study and have led our team to discuss possible recommendations for the organization. The first recommendation for our company based on our results is to request our accounting department to focus our salary budget resources towards people making less money annually. Secondly, we have concluded that since employees are company hopping, the recruiting department should try to avoid hiring people than have worked for more than three companies previously. Lastly, our organization must come up with an alternative strategy to retain employees that have been with the company for a short amount of time. By implementing these strategies in our company, I believe that our organization will be able to retain more valuable employees over time.

After discussing the recommendations for our organization, our team also came up with new questions about the future of our study and our organization. Are other variables such as PTO time, workspace, and building amenities another reason why employees are leaving our company? Also, would providing a hybrid schedule to those that have longer commute distances to work help retain the employees lost for that reason? Additional data to answer these questions may help support our conclusions on why employees keep leaving our organization. To answer such questions and learn more about employee behavior, I believe the future work of our human resources team should be guided towards doing extensive research on employee behavior in other companies to better grasp the factors that keep employees satisfied and efficient at our company.

## 4. Appendix/Appendices.

employee\_data <- employee\_data %>% add\_column(job\_satisfaction3 = employee\_data$job\_satisfaction)  
employee\_data[employee\_data$job\_satisfaction3 == "Low",]$job\_satisfaction3 = "1"  
employee\_data[employee\_data$job\_satisfaction3 == "Medium",]$job\_satisfaction3 = "2"  
employee\_data[employee\_data$job\_satisfaction3 == "High",]$job\_satisfaction3 = "3"  
employee\_data[employee\_data$job\_satisfaction3 == "Very High",]$job\_satisfaction3 = "4"  
  
employee\_data$job\_satisfaction3 <- as.numeric(employee\_data$job\_satisfaction3)  
  
  
employee\_data <- employee\_data %>% add\_column(performance\_rating3 = employee\_data$performance\_rating)  
employee\_data[employee\_data$performance\_rating3 == "Not Effective",]$performance\_rating3 = "1"  
employee\_data[employee\_data$performance\_rating3 == "Minimally Effective",]$performance\_rating3 = "2"  
employee\_data[employee\_data$performance\_rating3 == "Meets Expectations",]$performance\_rating3 = "3"  
employee\_data[employee\_data$performance\_rating3 == "Exceeds Expectations",]$performance\_rating3 = "4"  
employee\_data[employee\_data$performance\_rating3 == "Exceptional",]$performance\_rating3 = "5"  
  
employee\_data$performance\_rating3 <- as.numeric(employee\_data$performance\_rating3)  
  
  
  
employee\_data <- employee\_data %>% add\_column(business\_travel3 = employee\_data$business\_travel)  
employee\_data[employee\_data$business\_travel3 == "None",]$business\_travel3 = "1"  
employee\_data[employee\_data$business\_travel3 == "Rarely",]$business\_travel3 = "2"  
employee\_data[employee\_data$business\_travel3 == "Frequently",]$business\_travel3 = "3"  
  
employee\_data$business\_travel3 <- as.numeric(employee\_data$business\_travel3)  
  
  
employee\_data <- employee\_data %>% add\_column(job\_level3 = employee\_data$job\_level)  
employee\_data[employee\_data$job\_level3 == "Associate",]$job\_level3 = "1"  
employee\_data[employee\_data$job\_level3 == "Manager",]$job\_level3 = "2"  
employee\_data[employee\_data$job\_level3 == "Senior Manager",]$job\_level3 = "3"  
employee\_data[employee\_data$job\_level3 == "Vice President",]$job\_level3 = "4"  
employee\_data[employee\_data$job\_level3 == "Director",]$job\_level3 = "5"  
  
employee\_data$job\_level3 <- as.numeric(employee\_data$job\_level3 )  
  
  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_miles\_from\_home = min(miles\_from\_home),  
 avg\_miles\_from\_home = mean(miles\_from\_home),  
 max\_miles\_from\_home = max(miles\_from\_home),  
 sd\_miles\_from\_home = sd(miles\_from\_home),  
 pct\_less\_15\_miles = mean(miles\_from\_home <= 15))  
  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_weekly\_hours = min(weekly\_hours),  
 avg\_weekly\_hours = mean(weekly\_hours),  
 max\_weekly\_hours = max(weekly\_hours),  
 sd\_weekly\_hours = sd(weekly\_hours),  
 pct\_less\_55\_hours = mean(weekly\_hours <= 55))  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_yrs\_at\_company = min(yrs\_at\_company),  
 avg\_yrs\_at\_company = mean(yrs\_at\_company),  
 max\_yrs\_at\_company = max(yrs\_at\_company),  
 sd\_yrs\_at\_companys = sd(yrs\_at\_company),  
 pct\_less\_15\_years = mean(yrs\_at\_company <= 20))  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_yrs\_since\_promotion = min(yrs\_since\_promotion),  
 avg\_yrs\_since\_promotion = mean(yrs\_since\_promotion),  
 max\_yrs\_since\_promotion = max(yrs\_since\_promotion),  
 sd\_yrs\_since\_promotion = sd(yrs\_since\_promotion),  
 pct\_less\_7\_years = mean(yrs\_since\_promotion <= 7))  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_previous\_companies = min(previous\_companies),  
 avg\_previous\_companies = mean(previous\_companies),  
 max\_previous\_companies = max(previous\_companies),  
 sd\_previous\_companies = sd(previous\_companies),  
 pct\_less\_4\_companies = mean(previous\_companies <= 4))  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_job\_satisfaction = min(job\_satisfaction3),  
 avg\_job\_satisfaction = mean(job\_satisfaction3),  
 max\_job\_satisfaction = max(job\_satisfaction3),  
 sd\_job\_satisfaction = sd(job\_satisfaction3),  
 pct\_less\_medium\_satisfaction = mean(job\_satisfaction3 <= 2))  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_performance\_rating = min(performance\_rating3),  
 avg\_performance\_rating = mean(performance\_rating3),  
 max\_performance\_rating = max(performance\_rating3),  
 sd\_performance\_rating = sd(performance\_rating3),  
 pct\_less\_meets\_expectations = mean(performance\_rating3 <= 3))  
  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_business\_travel = min(business\_travel3),  
 avg\_business\_travel = mean(business\_travel3),  
 max\_business\_travel = max(business\_travel3),  
 sd\_business\_travel = sd(business\_travel3),  
 pct\_less\_rarely\_travelling = mean(business\_travel3 <= 2))  
  
  
employee\_data %>% group\_by(left\_company) %>%  
 summarise(n\_employees = n(),  
 min\_job\_level = min(job\_level3),  
 avg\_job\_level = mean(job\_level3),  
 max\_job\_level= max(job\_level3),  
 sd\_job\_level = sd(job\_level3),  
 pct\_less\_senior\_manager = mean(job\_level3 <= 3))  
  
  
  
#bar chart  
ggplot(data = employee\_data, aes(x = job\_satisfaction, fill = left\_company)) +  
geom\_bar(color = "white") +  
facet\_wrap(~ left\_company, nrow = 2) +  
labs(title = "Job Satisfaction by Status (Left the Company - Yes/No)",  
x = "Job Satisfaction Levels", y = "Number of Employees")  
  
  
  
#histogram  
ggplot(data = employee\_data, aes(x = salary, fill = left\_company)) +  
geom\_histogram(aes(y = ..density..), color = "white", bins = 20) +  
facet\_wrap(~ left\_company, nrow = 2) +  
labs(title = "Employee Salary Distribution by Status (Left the Comapny - Yes/No)",  
x = "Salary (US Dollars)", y = "Proportion of Employees")  
  
ggplot(data = employee\_data, aes(x = weekly\_hours, fill = left\_company)) +  
geom\_histogram(aes(y = ..density..), color = "white", bins = 20) +  
facet\_wrap(~ left\_company, nrow = 2) +  
labs(title = "Weekly Hours Distribution by Status (Left the Comapny - Yes/No)",  
x = "Time (hours)", y = "Proportion of Employees")  
  
ggplot(data = employee\_data, aes(x = miles\_from\_home, fill = left\_company)) +  
geom\_histogram(aes(y = ..density..), color = "white", bins = 20) +  
facet\_wrap(~ left\_company, nrow = 2) +  
labs(title = "Employee Commute Distance by Status (Left the Comapny - Yes/No)",  
x = "Distance (Miles)", y = "Proportion of Employees")  
  
  
  
#boxplot  
ggplot(employee\_data, aes(x = previous\_companies, fill = left\_company)) +  
 geom\_boxplot()+  
facet\_wrap(~ left\_company, nrow = 2) +  
labs(title = "Previous Companies by Status (Left the Company - Yes/No)",  
x = "Number of Companies")  
  
ggplot(employee\_data, aes(x = yrs\_at\_company, fill = left\_company)) +  
 geom\_boxplot()+  
facet\_wrap(~ left\_company, nrow = 2) +  
labs(title = "Number of Years at Company by Status (Left the Company - Yes/No)",  
x = "Number of Years at Company")

— End —