

FermaLogis

By:  Chandrashila Chattopadhyay

Vinny Franchin

  Ashwini Nadupuri

Haley Reierson

  Stephanie Wallace

**Table of Contents**

* **Introduction \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 3**
  + Business Problem
  + Analytical Objective
  + Data Source
  + Tools Used (SAS Base)
* **External Research \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 2**
* **Data Description \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 3**
* **Data Preprocessing \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 4**
  + Data Cleaning
  + Data Transformation
  + Variable Creation
  + Recoding
* **Data Exploration \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 5**
* **Data Modeling \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 6**
* **Model Comparison \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 7**
* **Findings & Results \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 8**
  + What are the major factors identified by final model?
  + Who is leaving the company?  Why are they leaving?
  + When is the biggest danger for employees to leave?
  + Is there any difference in attrition between different employee groups or categories?
  + Any other findings of the company.
  + Can I combine different event types together? Or do all need to be handled separately?
  + What attributes increase/decrease the hazard rates for certain event types?
  + Does bonus affect employee turnover? If yes, how?
  + Are there any variables which affect hazards non-proportionally?
* **Recommendations \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 9**
* **Conclusion \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 10**
* **References \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 11**
* **Appendix \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Page 12**

**Introduction**

To many of the top-ranking companies, employees are the most valued asset. Without an engaged and committed workforce, the operations would be unable to deliver customer satisfaction. It is because of this, that year after year, companies make significant investments in both the development and retention of their workforce. The hope being, the more the company invests in its people, the more committed they will be to the company. One common measure used to assist in the evaluation of employee engagement and satisfaction is “employee turnover”, the reciprocal of which is “employee retention”. Turnover serves as a barometer for concern if it starts to trend up. When it becomes elevated, it is important for the company to get ahead of it by looking for the potential factors of causation. Without company action, the problem will only exacerbate itself.

FermaLogis, a successful pharmaceutical company, prides itself on its commitment to the development and retention of its workforce. With training programs available to both new and experienced personnel, FermaLogis understands the value their employees bring and strives to provide the tools and skills necessary to be successful at the company. The professional training offered to new employees requires a rather sizable investment that does not see a return for the first three years. When new employees leave, it not only provides a vacancy, but it also means that the return will not be realized. The Executive Training program is offered to employees who have a tenure with the company of five years or more. Each year, two weeks are designated for participants to be trained by industry leading consultants. The knowledge gained from the training sessions are believed to be invaluable. Employees who partake in these training programs become prime targets for poaching from competitors.

While the company is passionate about developing its workforce, FermaLogis is becoming increasingly challenged by employee turnover. With the rise in, and new-found knowledge of survival analysis, FermaLogis’ COO, Larry Hansen is seeking analytical and technical assistance to hopefully uncover the drivers of turnover so the company can get ahead of the problem before it gets worse. Overall, Larry Hansen would like to improve the workplace environment at FermaLogis and is open to adopting new ideas and techniques that will achieve this and lead to employee retainment. Hansen has publicly remarked, “We care about our employees so much. Employees are very important for the FermaLogis operations as the main resource and input.” Using workforce data provided by FermaLogis’ COO, in combination with survival analysis techniques, the hope is to gain an understanding around who is leaving the company and identify potential reasons why they are leaving. Using SAS Base, the data will be analyzed by a variety of segmentations not limited to; demographics, tenure, performance, and compensation. The hope is that the analyses performed will not only realize pockets of opportunity but identify turnover related trends that the company can both react to and risk manage against.

**External Research:**

Employee engagement and retention are consistently hot topics with countless sources providing feedback and recommendations for how companies can foster engagement and company loyalty. The constant battle is that “over the long term, just like customers, employees do have a choice, which is why it is in the company’s best interest to obsess a bit over why they would stay or go.” (Wagner, 2017) At the end of the day, it seems that employees value happiness the most. Employees want to wake up wanting to go to work, not dreading it, and when employee happiness is compromised, engagement suffers, and quite quickly everything spirals downhill from there.

To get a pulse on the current state of the workforce, companies tend to utilize anonymous surveys led by independent consulting firms. Surveying can be a useful mechanism to get employee insight, however, the results can oftentimes be biased. For example, “the importance of compensation tends to be underestimated in employee surveys because, by definition, they include only those at least satisfied enough with the company’s pay to still be at the organization. The talented people who took higher wages elsewhere are not part of the analysis because they are no longer employees.” (Wagner, 2017) As a result, it is important for companies to not only ask their workforce how they feel, but to benchmark the employee benefits being offered to those of their competitors. Offering at least the same or higher benefits will certainly reduce turnover associated with inadequate compensation.

“Retention is more of a strategy than an outcome. It takes careful attention and maintenance to cultivate an engaged and loyal workforce.” (McDougall, 2018) Different demographics may require focus on different workforce benefits than others. For example, millennials greatly appreciate flexible hours, the ability to work remote, and work life balance. Whereas, those closer to retirement, might be most concerned with what their compensation benefits look like. Either way, one key thing to remember is that employees want to feel valued and that they matter. There are a variety of ways to show employees their worth that extends beyond compensation. In addition to providing competitive pay, giving real-time feedback, as well as having conversations around career trajectory can have a huge impact on employee retention. Employees do not want to feel stuck, and they also want to be treated like adults. Allowing employees to have some level of responsibility and control over their schedule, through flexible work hours is another great way to showcase employee appreciation and trust.

Another aspect to consider when seeking to maximize employee retention and minimize employee turnover, is hiring selection. Sometimes employees leave because the job is just simply not a good fit. Being careful and selective during the hiring process will allow companies to put greater focus on how good of a fit the candidate is for the job opportunity, as well as how good of a fit is the candidate relative to the company’s overall culture. It benefits both parties to dig into the overall fit of a prospective employee. According to a recent study, the cost to replace an employee is roughly one-third of their salary. Carefully vetting candidates will not only lead to better odds of finding the right candidate, but it will have definite cost savings in the long run.

**Data Description**

For this project initiative, FermaLogis has provided details of its employee workforce in a CSV file. Each row of information is at an individual employee level and contains demographic information surrounding age, gender, marital status, and education, as well as job, tenure, performance, and compensation information. From a job perspective, information is provided around a variety of factors including; what the employee’s current role is, what their operational involvement-level is, how many hours they work, and how satisfied they are with their job. From a tenure perspective, information has been included beyond simply how long the employee has been in the company. In addition, there are details around how many years the employee has been in their current role, as well as how many years the employee has worked for their current manager. Performance and compensation go hand-in-hand, and there is information related to the employees pay, their manager’s evaluation of their performance, the last 40 years of bonus payment history, how many years since the employee’s last promotion, and the number of training programs attended by the employee.

The above provides a high-level overview of some of the core areas that can be analyzed with the provided dataset. The CSV is extremely comprehensive and will allow for tons of analytical flexibility. There are a variety of ways turnover can be segmented and analyzed. It will be interesting to see which variables prove to be the most impactful and the most telling when it comes to uncovering drivers of employee turnover.

**Data Preprocessing**

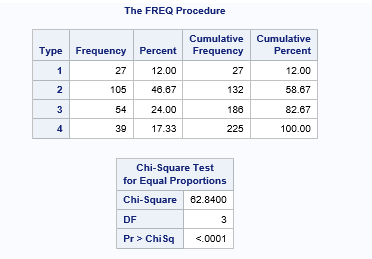
Before beginning our model construction efforts, we analyzed the dataset provided by FermaLogis, as well as the corresponding Metadata to determine if any data clean-up was necessary. Metadata is a useful tool when analyzing data, you are unfamiliar with. It provides definitions and various details about the columns of variables included in the dataset. Upon review of the Metadata, it became clear that our dataset contained approximately 1,470 rows and 79 columns of variables. With a clearer understanding of the data structure, we began data preprocessing which unveiled components of the data that needed to be cleaned due to any of the following; noise that needed smoothing, missing values or data inconsistencies that needed to be handled or areas of that required transformation.

Our data preprocessing efforts ultimately reduced complexity by dropping unwanted columns such as; unnamed columns, columns that seemed to provide the exact same information (duplicates), and columns that had the same exact value for every individual employee. By removing unwanted columns, it simplified our dataset since those columns were not going to be used as model inputs. The dataset provided by FermaLogis surprisingly contained minimal missing values. It is rather unusual for data to be so “complete”, however, this was great as it meant there wasn’t a lot of extensive data clean-up needed. We did notice for “Bonus” columns 1-40 that there were several NA’s. It makes sense that not every employee would receive a bonus each year, however, we do want to clean up the NA’s to be inline with the rest of the information in these columns. We chose to handle the NA’s by transforming NA’s within the “Bonus” columns to a 0 since a 1 indicates a bonus was received, and so therefore, a 0 would mean a bonus was not received. The transformation of NA’s allowed for the “Bonus” columns to be transformed from text to numeric columns. The last data preprocessing involved the creation of a new numeric column titled “Turnover Type”. The possible values for “Turnover Type” included four options; 1 = Retirement, 2 = Voluntary Resignation, 3 = Involuntary Resignation, and 4 = Job Termination. “Turnover Type” utilized the column “Type” to recategorize the type of turnover into a number which would be much easier to utilize for modeling purposes. One important thing to note about the newly created column “Turnover Type” is that employees who are still with the company are given an input of 0, so the total possible values for this variable range from 0 to 4. This distinction allowed us to observe that out of the 1,470 employees, 225 had a turnover type of either 1,2,3, or 4; meaning that 1,245 employees are still with the company.

With the data preprocessing steps taken we have the flexibility to compare not only those that are still with the company against those that have left the company, but we can also compare the different turnover populations. As we have previously mentioned, survival analysis is used to find the time to an event. In the case of our business problem, “Years At Company” represents tenure or the time at which the turnover will happen. In the real world, it is impossible to track or observe the data more fully, in which case, survival analysis provides a better solution than other common regression/ classification techniques. With survival analysis, if the targeted event happens in the given time period, it is a “failure”, otherwise, if the event does not occur, the targeted event is considered to be “censored/ not observed”. With regards to our project objective, when the newly created column “Turnover Type” yields a value between 1-4, we considered this to be a “failure”1,2,3,4, a value of 0 is a censored event.

**Data Exploration:**

To begin data exploration efforts, we first created a sample size of about 225 records. From there, we continued with a deep dive into exploration by running a FREQ procedure on the column “Turnover Type” to understand which turnover type has the greatest frequency at FermaLogis. Based on the results, we see that “Turnover Type” = 2 (Voluntary Resignation) accounts for almost half of the observations at 46.67% (105 out of 225 sample records). The below procedure confirms that FermaLogis is experiencing higher turnover, and thus it makes sense for them to be concerned. “Turnover Type” = 1 (Retirement) accounts for the smallest percentage of turnover, at 12% (27 out of 225 sample records). It seems that on average, fewer employees tend to stay at FermaLogis long enough to reach retirement. An interesting avenue to dig deeper into with our modeling, would be to understand exactly why voluntary turnover is so high, and why so few FermaLogis employees opt to stay at the company until they are retirement eligible.



To further our analytical exploratory efforts on turnover, we developed a series of frequency plots which uncovered very interesting findings:

* 80% of employees that voluntarily resigned were not satisfied with the company’s facilities which potentially suggests that the company’s facilities are outdated or not conducive to fostering learning and productivity.
* 70% of employees that voluntarily resigned either received or was intending on pursuing a higher education suggesting that really smart and driven individuals wanted to leave the company.

Analyzing the uncovered 70% statistic, we thought that these employees’ drive to leave the company could be since FermaLogis is not providing employees with higher education support or highlighting the pathway for internal growth opportunities which would include potential promotions or pay increases. Without this clarity, employees might be leaving to work for a company that will be more supportive and present in their pursuit for growth and career progression. w them to grow their careers and take on new roles/responsibilities. With this caliber of employees voluntarily resigned at such a high rate, FermaLogis could be in trouble if they don’t act to stop this trend from continuing. Talent retention is particularly important in the pharma industry; competitors are always on the lookout and looking for ways to poach top talent.

Sticking with the theme of education, our next step was to analyze the educational backgrounds of FermaLogis employees such as human resources, life sciences, marketing, and medical, just to name a few in order to see if there were any identifiable patterns or trends related to voluntary turnover. What we saw was that the two largest fields that made up the voluntarily turnover population included Life Sciences and Medical fields. While it takes a variety of talents to successfully operate a pharmaceutical company; quite obviously both the Life Science and Medical fields are hugely important and are arguably two of the most essential fields. Typically, Life Science & Medical backgrounds are associated with higher education. At FermaLogis, this was also the case. When looking at the industry as a whole, research and development accounts for approximately 17% of total revenues at pharmaceutical companies. FermaLogis, like its competitors, relies heavily on research and development to development new drugs that can hopefully become FDA approved and brought to market.

Another variable we analyzed was “Overtime” and its interaction with “Turnover Type” and uncovered that 60% of employees who voluntarily left FermaLogis were working more than 10 hours of overtime a week. Comparing “Overtime” to termination, we see that about two-thirds of those employees that were terminated, did not work overtime. Rationalizing both statistics makes sense because terminated employees are typically not high performers and thus are not going to go above and beyond, so you would expect not to see a lot of overtime being worked with this group. However, it is important to focus on the fact that we have already seen that educated talent is leaving, and not only that, they were high-performing from the perspective that they were working overtime hours. If employees are not feeling valued or that they are being paid in correspondence to their performance, they are going to ultimately feel overworked and seek employment elsewhere. Also, important to consider is the fact that non-exempt employees get paid for the hours of overtime worked, whereas exempt employees do not, that is until performance compensation (i.e. bonuses, stocks, etc.) are awarded, if they are awarded.

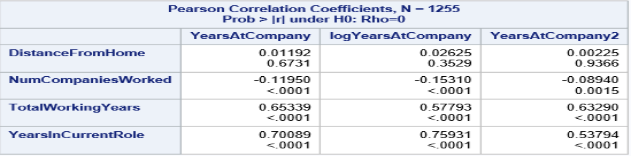
Finally, there were a few other frequency plots we ran which compared variables like “Employee Environment Satisfaction”, “Business Travel”, “Employee Job Involvement”, and “Gender” to “Turnover Type”. We found that variables such as “Employee Environment Satisfaction” and “Employee Job Involvement” were evenly split down the middle when comparing them to our turnover populations.

**Data Modeling**

After completing the data preprocessing and exploration efforts, it was time to get into the very heart of our project objectives. The COO of FermaLogis tasked us with identifying the causation and factors driving the recently experienced elevated turnover and provide recommendations for how to overcome this hurtle and thus improve the overall retention rate at the company. With our dataset ready for modeling, it was time to evaluate the input variables to determine which showed statistical significance, and from there, which would be best suited for inclusion in the various models we would be developing. With a variety of survival models at our disposal, we made sure to explore all of our options before zeroing in on a particular direction.

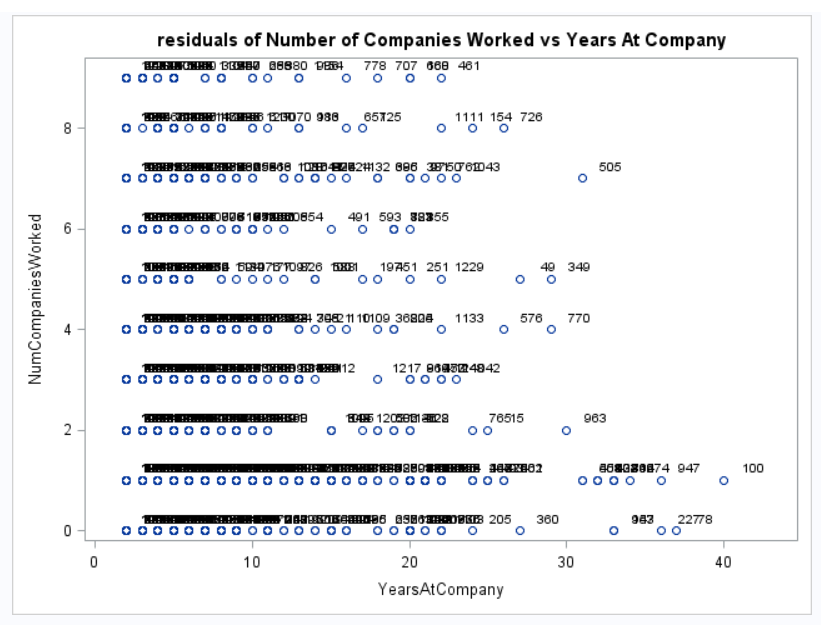
One model type we debated using was Accelerated Failure Time (AFT). AFT models are used to describe the relationship between the survivor functions of any two individuals. Associated assumptions for AFT Models are; there is a constant mean, constant variance, and there is independence across different observations (i.e. survival time of individual 1 does not depend on the survival time of individual 2). Unlike with other survival models, there is no normality assumption which means this model can follow a wide array of distributions including; Weibull, Exponential, Gamma, Log-logistic, and Log-Normal. The expansive set of distributions is made possible by the fact that each distribution has different implications for hazard functions which lead to varying substantive interpretations. As a result, selecting the most appropriate distribution for your dataset is essential to developing an accurate model because the chosen distribution has an impact on which variables prove to be significant, thus leading to different results and conclusions depending. Though AFT was a valid route to go, we ultimately decided to use Cox Proportional Hazards (PH) Model. Unlike AFT, Cox PH models do not require an analyst to select the probability distribution to be used to represent survival times. Additionally, Cox PH does not depend on the timing around when events took place, but rather, places emphasis on the order in which events occurred. In context to our objective, the Cox PH model would focus on the ranking of drivers of what led to turnover, as oppose to when the resignation took place. For our project, the timing of when an employee chose to leave the company is not as important as understanding why employees are leaving, making ranking the appropriate to take. Furthermore, we opted to use the EFRON method, which is best suited for tied and small datasets.

Before diving into the development of our various survival analysis models, there were some important assumptions we needed to consider in order to maximize the potential accuracy of our results. To aid in the determination of the assumptions we needed to make, the residual method was used. Common residual methodology for the Cox model include Schoenfeld residuals which are used to check the proportional hazards assumption including the time dependent covariates (if any) and Martingale residual to assess any potential nonlinearity. The proportional hazards (PH) assumption can be checked by using statistical tests and graphical diagnostics based on the scaled Schoenfeld residuals.

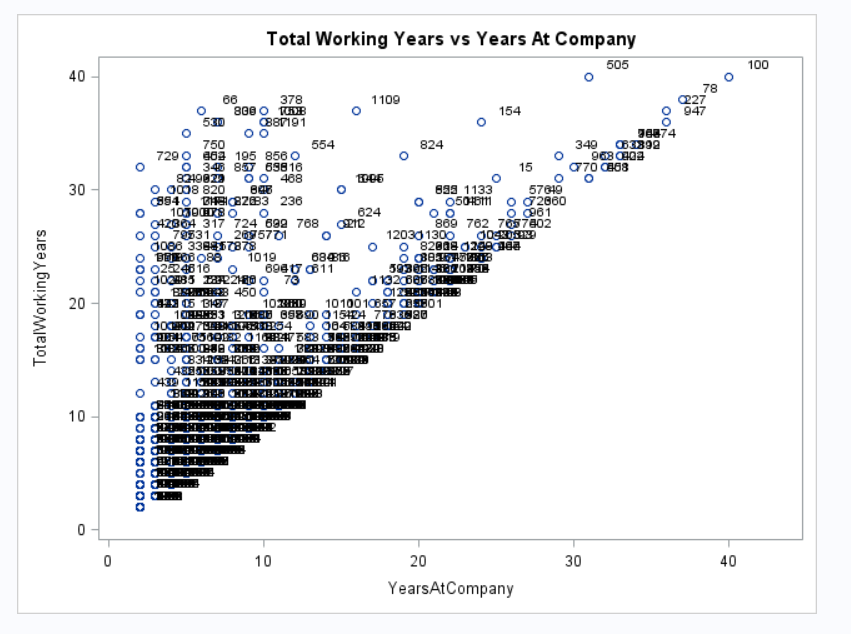
**

The correlation matrix above shows that “Distance from Home” is not correlated with any of the time dependent interaction of “Years At Company”. The remaining three variables: “Number of Companies Worked”, “Total Working Years”, and “Years In Current Role” have significant time dependent interactions with “Years At Company”. Additionally, is the fact that the variables “Total Working Years” and “Years In Current Role” have a positive correlation with the variable “Years At Company”, whereas “Number of Companies Worked” has a negative correlation with the variable “Years At Company”. Logically speaking, the presence of this negative correlation seems to make sense as employees who have worked at multiple companies are less likely to stay at one company for a long time. On average, individuals who tend to jump from one company to the next are typically driven by higher pay and title.

The residual plot below shows the **“Number of Companies Worked”** vs. “Years at Company”. From the diagram you can see that higher “Number of Companies Worked” corresponds more often to “Years at Company” of 5 years or less. Thus, there is clearly a time dependency with this variable.



The Residual Plot showing “**Total Working Years”** vs. “Years at Company” exuded an *interesting* pattern. As the total working years of an employee increases, we observe an increase in the years at the company which means the total working years are directly correlated with the metric of interest also making the variable total working years a time dependent variable. When rationalizing this result, we concluded that this can be explained by the fact that the more work experiences an employee has, typically, the older the employee is and thus the closer to retirement compared to a newly college graduate just beginning their career.



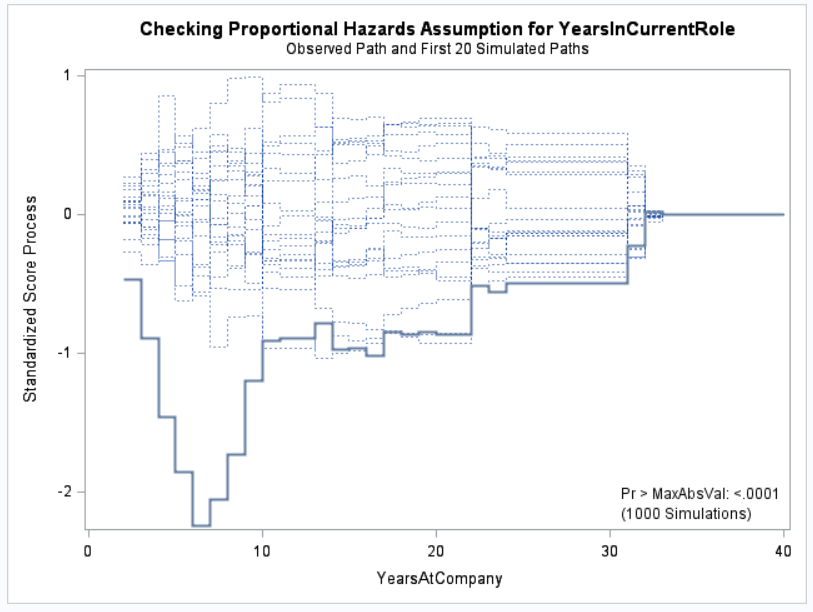
The last Residual Plot contrasted **“Years In Current Role”** vs. “Years At Company”. The resulting Residual Plot showed that as an employee’s years at the company are increasing, the years in current role are also increasing. We think this positive correlation makes sense as longevity in a particular role are associated with promotions and growth opportunities, all of which are inter-departmental retainment measures. This analysis also resulted in the signification that he variable “Years In Current Role” is a time dependent variable.



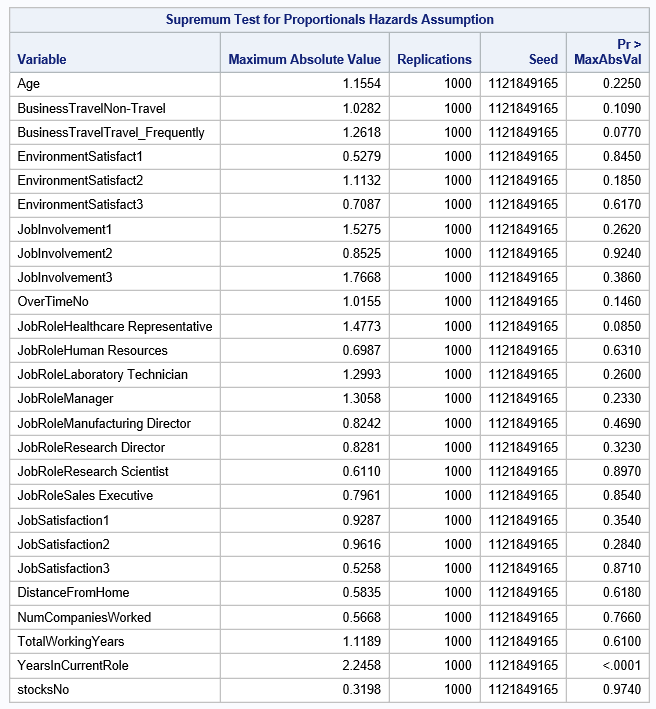
Analysis of the previously described residual plots bring to light evidence of non-proportional relationships of covariates with time. Since all of the residual plots proved to have a time dependent variable, we have added interactions so that we can use these three variables (“Number of Companies Worked”, “Total Working Years”, and “Years In Current Role”) in our model development.

Plotting the Martingale Residuals against continuous covariates (which are not time dependent) is a common approach used to detect nonlinearity or, in other words, to assess the functional form of a covariate. For a given continuous covariate, patterns in the plot may suggest that the variable is not properly fit. Unlike continuous variables, nonlinearity is not an issue for categorical variables, so we only examine plots of martingale residuals and partial residuals against continuous variables.

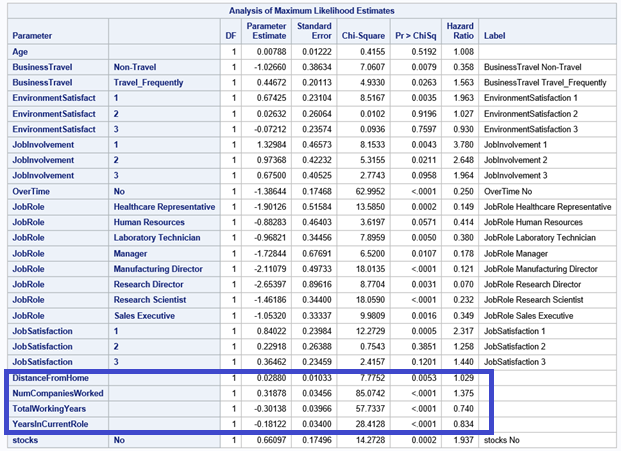
Martingale residuals may present any value in the range (-INF, +1) where a value of martingale residuals near 1 represents individuals that “died too soon”, and large negative values correspond to individuals that “lived too long”. Like we found in our Schoenfeld residuals test, “Years In Current Role” is a non-proportional variable in the below Martingale residual plot of the Standardized Score Process of “Years In Current Role” against “Years at Company”. Note, there are several deviations of actual vs observed in the diagram below, with the largest deviation occurring around year 7.



The Supremum Test for Proportional Hazards Assumption (as given in below screenshot) tells us that the maximum absolute value of covariates. This test yielded non-significant P-values for all covariates except for “Years In Current Role”. The significance of one covariate in this test tells us that the proportional hazards assumption holds for most of the covariates. In other words, we have evidence to conclude that “Years In Current Role” has significant time interaction with the target variable.



Our Schoenfeld statistics (screenshot below) shows that the parameter estimates for “Distance From Home”, “Number of Companies Worked”, “Total Working Years”, and “Years In Current Role” are all significant, which means that their hazard function changes over time. In other words, these variables are not proportional. However, it is important to note, that in our correlation matrix, we found that “Distance From Home” had no significant time interaction with the variable “Years At Company”.



When building the Cox PH models, it is important for us to be thoughtful when it comes to deciding on variable inputs because all variable inputs must satisfy the proportional assumption. The proportional assumption is not satisfied when we review the variables; “Number of Companies Worked”, “Total Working Years”, and “Years In Current Role”. To satisfy the necessary condition, or rather, to include time interaction with these three variables, we multiplied each of the three variables by “Years At Company” to create three new variables titled; “Time Interact Working Years”, “Time Interact Current Role”, and “Time Interact Number of Companies”.

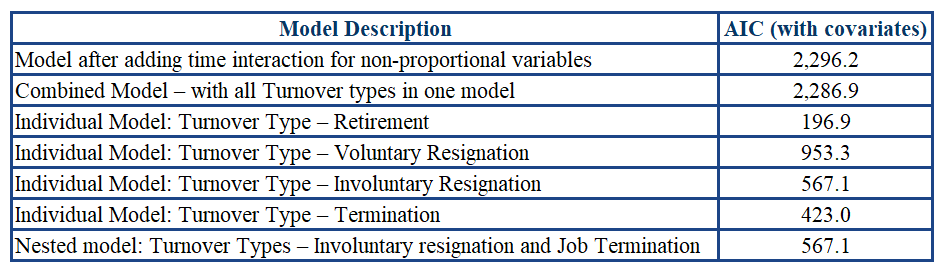
Three types of models were developed to analyze the response variable with respect to each turnover type; retirement, voluntary resignation, involuntary resignation, and termination. The three models we developed are:

1. *Combined model:* All four turnover types studied together in one model.
2. *Nested Model:* Similar turnover types were analyzed together; involuntary resignation and termination had similar survival curves in PROC LIFETEST.
3. *Individual Models:* One model developed to analyze each of the four turnover types on its own.

Except for involuntary resignation and termination, the other turnover types are considerably different from one another. In other words, the major factors driving each of these turnover types were quite different. As a result, we found the individual models performed far better than the combined and nested models; meaning, the individual models were superior in analyzing the factors responsible for employees leaving the company. In the following sections, we will provide a more in depth look into how each of the developed models compare to one another and provide additional observations and insight based on the model results.

**Model Comparison**

To benchmark our developed models and determine which was the best performing, we utilized the comparative statistic AIC. The metric AIC stands for Akaike Information Criterion and is model tuning technique that focuses on in-sample fit and uses this to estimate the likelihood of a model’s ability to predict/estimate future values. The lower the AIC, the better the model is at predicting/estimating future values.



The chart above shows each model we developed, the description, as well as the associated AIC. The “Individual Model: Turnover Type – Retirement” proved to be the best model as the AIC, 196.917 is the lowest compared to the others. From an AIC perspective, it seems that the individual models are performing better than the combined models. Also, noteworthy, is the fact that nested models have better AIC performance than the developed combined models. At the end of the day, the model we selected was an individual model and as was previously explained, had the lowest AIC overall.

**Findings & Results**

**What are the major factors identified by the final model?** As discussed previously, we believe that analyzing each of the four turnover types; retirement, voluntary turnover, involuntary turnover, and termination separately allows for more accurate results as shown by the markedly lower AIC’s compared to the combined and nested models.

**Turnover Type: Retirement**

* “Age” was the most significant factor for this model. Every year increase in age leads to a 51% risk of employees leaving the organization.
* “Business Travel” was found to increase the risk of leaving the organization by over 830 times compared to employees who do not regularly travel.
* “Low Job Involvement” and “Number Of Companies Worked” were showed to increase the risk of turnover due to retirement.
* Employees who do not work overtime have an 11.5% less chance of leaving the organization than those employees who do work overtime.

**Turnover Type: Voluntary Resignation**

* Frequent “Business Travel”, “Overtime”, “Job Satisfaction”, “Stocks”, and “Years In Current Role” showed the most significance for this turnover type.
* Low “Job Satisfaction” had the highest hazard ratio among all of the significant factors.
* Employees identified as having low “Job Satisfaction” are 584 times more likely to leave the company than medium or highly satisfied employees.

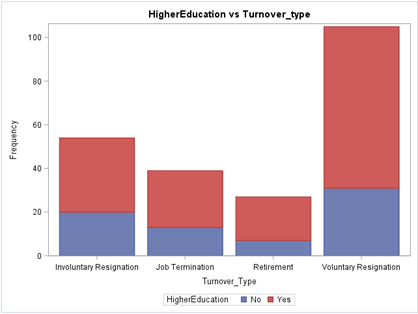
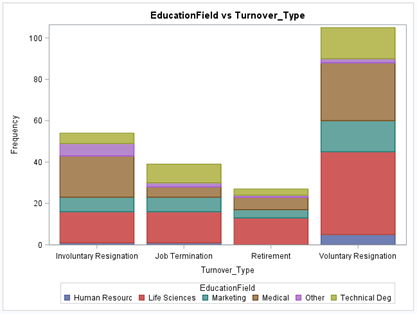
**Turnover Type: Involuntary Resignation**

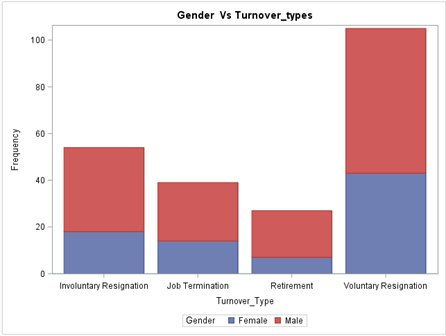
* “Overtime” was the biggest factor for this turnover type.
* Employees who work overtime were 75% more likely to involuntarily resign than employees who do not work overtime.
* “Number Of Years Worked”, “Current Role”, and “Number Of Companies Worked” were also determined to be significant factors.
* Examples of “Job Roles” that showed significance were: Laboratory Technician, Manufacturing Director and Research Scientist.

**Turnover Type: Termination**

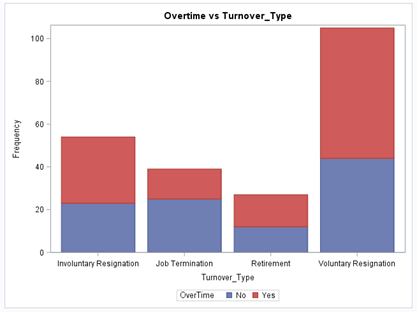
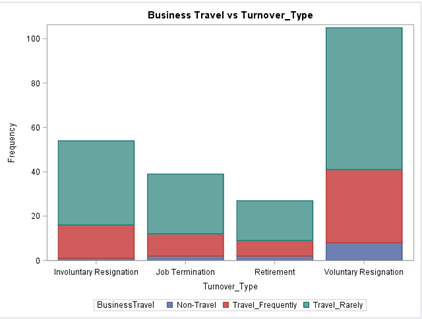
* “Years In Current Role” and “Age” were identified as the most significant factors for this turnover type.
* A one-year increase in “Years In Current Role” is leads to a 7.2% increase in the risk of leaving the organization.
* An increase in age decreases the risk of getting terminated by 87.8%.

**Who is leaving the company?** The employees who are leaving the company are by and large voluntarily resigning. This would not be a surprise to the company but should indicate that it is not the employees themselves who are the main issue. Primarily, it appears that most people who are voluntarily resigning are relatively new hires, in their first few years on the job. Additionally, those who are leaving typically have or are receiving a higher education. There are also more men leaving than women. Lastly, those leaving often have an educational background in Medicine or Life sciences. Below are graphs which illustrate who is leaving FermaLogis. According to our analysis, of those resigning voluntarily, about 70% had received or were receiving a higher education. Additionally, a large portion of those who were voluntarily leaving the company had educational backgrounds in either Life Sciences or Medical fields. Men appear to be slightly more prone to attrition at the company than women.

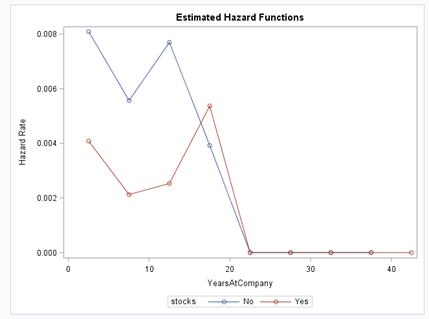


**Why are they leaving?** Employees are leaving the company for a multitude of compounding reasons, many of which are mentioned in the section on factors above. A poor Environment hurts employees' production, and when combined with the fact that they often do not feel that they are involved with their jobs, low satisfaction is a logical follow. Another key factor not yet discussed is the impact of working overtime on employee retainment. Employees who were terminated generally were not working more than 10 hours of overtime a week, while a majority of those who voluntarily resigned were working overtime. This tendency toward overtime suggests that there may be greater issues with culture at play.

Another factor in employee turnover is business travel. Although most employees who left were infrequent travelers, those who voluntarily resigned tended to be more likely to travel frequently, which may tie in with those working overtime. One final factor stems from a reason the CEO noted as why other companies were snagging their employees: training. Employees who did not attend any trainings in the past year had a hazard ratio of over 4.5, while those who attended 5 trainings had a ratio of only 0.875. Although the drop-in ratio as trainings increase is not linear, getting employees into training consistently clearly helps ensure that they stay with FermaLogis.

**When is the biggest danger for employees to leave?** Generally, as the number of years an employee works increases, they are slightly more likely to leave, and after about 30 to 35 years the survival curve drops sharply. Employees who leave voluntarily have the highest hazard rates at the first 5 years, around year 15, and again around year 20. Nonetheless, they are most likely to leave in the first 5 years so hanging on to them beyond those 5 years will help with retention. This can be seen on the following graph, which adds the factor of stock options, which as we mentioned above, helps mitigate turnover.



**Is there any difference in attrition between different employee groups or categories?** Employees with stock options have a slightly lower hazard rate than those without. It seems that those who have higher education, as well as those in the medical and life sciences fields, are more likely to leave than those who are not.

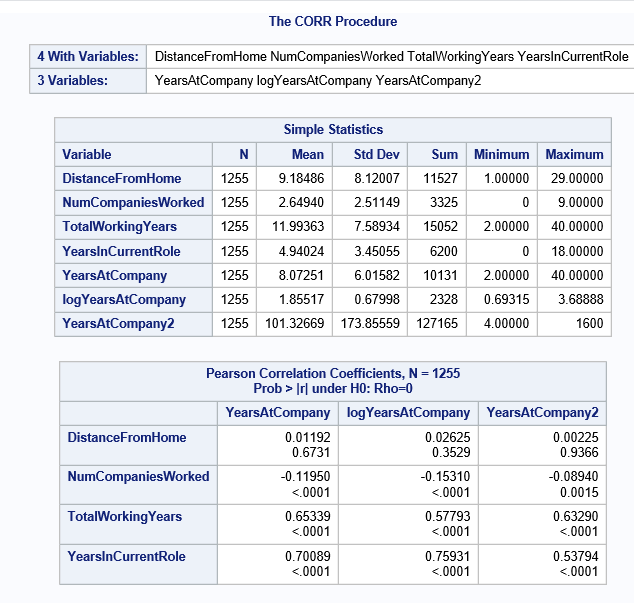
**Any other findings of the company.** There does not appear to be a great culture of growth at FermaLogis, as employees with higher education are leaving early in their time at the company. FermaLogis needs to pivot their culture to encourage learning and foster growth and development within the company.

**Can I combine different event types together? Or do all need to be handled separately?** For a general understanding of why the turnover is high, combining the different types of turnover together is reasonable. However, it is only on observing event types individually, like Voluntary Resignation or Retirement, that the greater issues come to light. This is because when looking at turnover individually, you can identify Voluntary Designation as the primary source of turnover, and then delve into its relevant issues.

**What attributes increase/decrease the hazard rates for certain event types?** Stock options are a compensation example that changes the hazard rate depending on the event type. For example, voluntary resignation shows a decrease in hazard rate for stock options.

**Does bonus affect employee turnover? If yes, how?** Bonuses showed to have some effect on employee turnover. They had a stronger influence on those who retire but proved to be insignificant for those employees who are terminated. This makes sense because if you are getting terminated, you are most likely not the best performer and therefore are not getting the best bonuses compared to your peers.

**Are there any variables which affect hazards non-proportionally?** The variables which showed to affect hazards non-proportionally were “Distance From Home”, “Number Of Companies Worked For”, “Total Working Years”, and “Years In Current Role”. This is displayed in the following table which looks at correlations and proportionality.



**Recommendations**

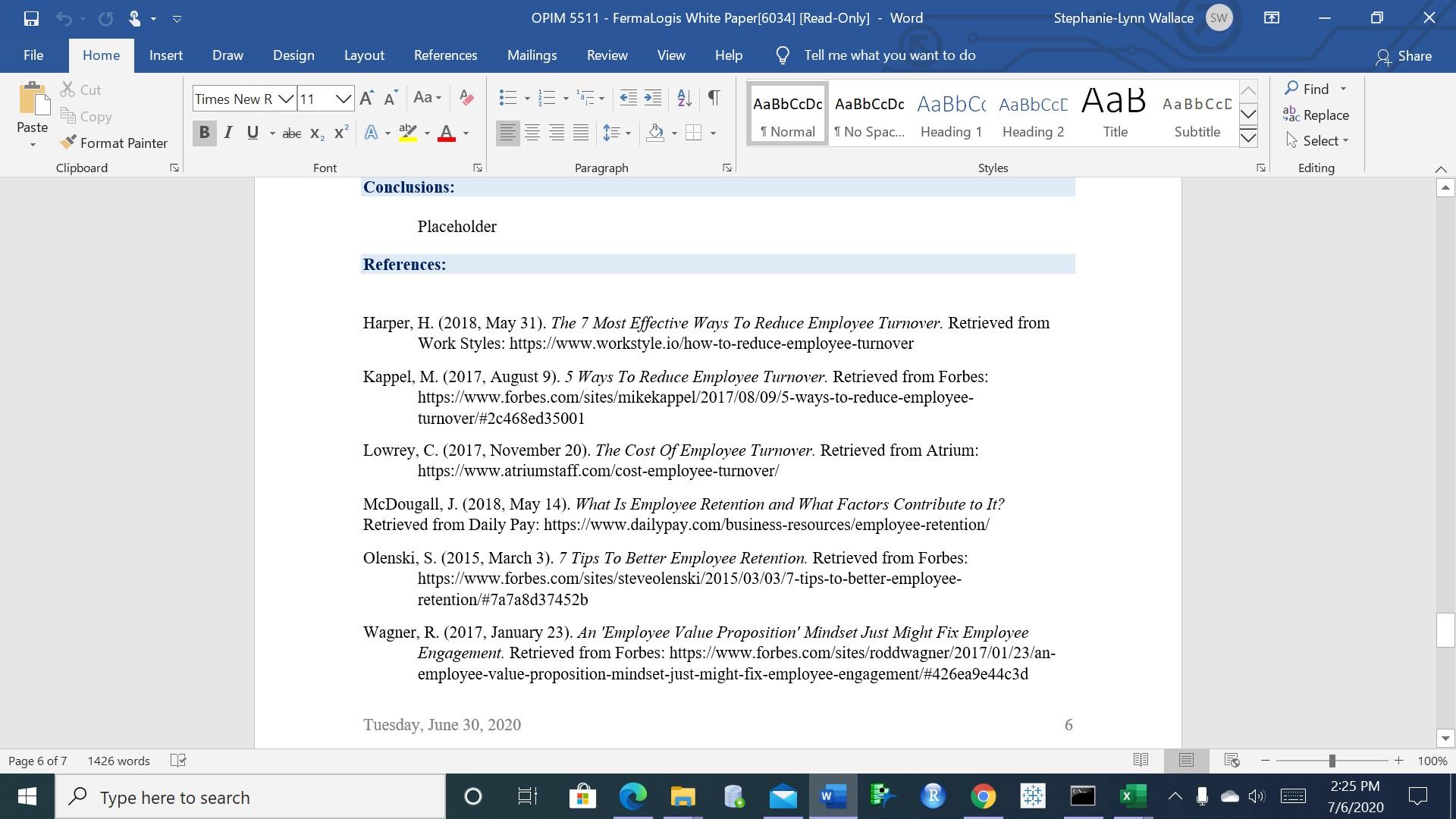
FermaLogis is going to need to rapidly change their work culture and environment if they want to stay current. However, we think that there are three areas where they will be able to address issues and help their employees stay on.

* **Motivate:** Foster an environment that promotes collaboration and creativity. Give employees the ability to have a voice and feel heard. Simply asking “How are you doing?” can give you, the manager, a clear pulse on things.
* **Retain:** Make employees feel valued. Not only show their value through compensation but give them insight into the potential growth opportunities that exist and walk them through what you need to see to take them to the next level. Employees who leave were frequently working ten or more hours of overtime a week, and employees who were terminated or involuntarily resigned generally did not. Work-life balance is highly valued, and a culture which punishes those who are unable or uninterested in overtime will lose employees fast.
* **Hire:** Take special care when hiring. Make sure the expectations of the position align with the expectations of the individual being hired. Validate that the prospect will be a good culture fit. Understand their near-future plan.
* **Train:** Take the time to onboard new employees of the company and those new to their current role. Give the employee the tools to succeed. Set them up with a mentor and make sure they feel empowered to ask questions and think outside of the box. While the training you provide to your employees is very valuable, there is also a disconnect between what is in those trainings and who can attend them. Employees who were able to attend many trainings on average were more likely to stay on, while there were employees who did not attend any trainings and were likely to leave. Job Involvement is one of the key factors, and those who do not feel involved, and are not attending trainings are seeking further growth. There are clearly many bright minds at FermaLogis, and providing them with challenging, engaging work with room to grow will help them feel fulfilled and improve your products.

**Conclusions**

In our analysis of the data, we have identified multiple key factors to that influence employee turnover. Despite having an abundance of bright minds, FermaLogis is struggling to hang on to them, and it is our hope that the provided analysis gives FermaLogis the tools and inspiration to rethink some of their current practices and implement actions that will foster a collaborative and positive work environment. Employees can only be as loyal as their company is to them. Part of establishing loyalty is providing a landscape for the employee so they can continuously see the value they add, as well as the potential growth opportunities that exist for them, should they want to work for it. Employee engagement is the nucleus that gets individuals excited to come to work, wanting to maximize their productivity, and passionate about doing what they can to see the company succeed. Without this, employees will ride the wave and leave at the first chance they find for a better professional experience. It is our hope that FermaLogis can find a way to minimize future turnover and retain the employees that most proudly represent and add value to their brand.

**References**



**Additional Resources:**

* <https://www.linkedin.com/pulse/using-survival-analysis-predict-reduce-employee-attrition-kumar/>
* <https://stats.stackexchange.com/questions/136704/can-i-use-survival-analysis-to-model-employee-attrition-when-employees-have-diff>
* <https://www.casact.org/education/spring/2013/handouts/Paper_1619_handout_953_0.pdf>
* <https://square.github.io/pysurvival/tutorials/employee_retention.html>
* <https://intellihr.zendesk.com/hc/en-us/articles/360000372915-Survival-Analysis-Understanding-Survival-Curves>
* <https://ttu-ir.tdl.org/bitstream/handle/2346/19664/31295013268429.pdf?sequence=1>
* <https://corpgov.law.harvard.edu/2018/01/22/the-option-to-quit-the-effect-of-employee-stock-options-on-turnover/>
* <https://www.strategy-business.com/article/re00229?gko=b0552>
* <https://github.com/nareshv00/Survival-Analysis-for-Predicting-Employee-Turnover>
* <https://hbr.org/2016/03/28-years-of-stock-market-data-shows-a-link-between-employee-satisfaction-and-long-term-value>
* <http://www.sthda.com/english/wiki/cox-model-assumptions>
* <https://stats.idre.ucla.edu/sas/seminars/sas-survival/>
* <https://stats.idre.ucla.edu/stata/seminars/stata-survival/>
* <https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/BS704_Survival_print.html>
* <https://towardsdatascience.com/survival-analysis-part-a-70213df21c2e>
* <http://www.sthda.com/english/wiki/cox-proportional-hazards-model>

**Appendix**



