|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| DigData Consulting Inc.  Survival Analysis of Employee Attrition at FermaLogis |
| |  |  |  | | --- | --- | --- | | Team 8 | PROJECT 1 | OPIM 5894 | |

**Team 8**:

Avinav Neer

Lav Bist

Pavithra Rao

Ruthwick Kuchibhotla

Srinivas Godavarthi

Sujith Vallepalli

Contents

[**EXCECUTIVE SUMMARY** 2](#_Toc457594420)

[Background 2](#_Toc457594421)

[**METHODOLOGY** 3](#_Toc457594422)

[Exploratory Data Analysis 3](#_Toc457594423)

[Survival Curve analysis 13](#_Toc457594424)

[Hazard Functions Plots 19](#_Toc457594425)

[Modeling Phase 23](#_Toc457594426)

[**Conclusions** 38](#_Toc457594427)

[**Recommendations** 40](#_Toc457594428)

[**References** 41](#_Toc457594429)

[**Appendix** 41](#_Toc457594430)

**The work presented is our team’s work and our team’s work alone**

# **EXCECUTIVE SUMMARY**

Our firm DigData Consulting Inc. was contracted to look into the problem of employee attrition at FermaLogis using advanced data analytic techniques and survival analytics. After a detailed analysis we have identified that job satisfaction and job involvement, salary and stock options, years in current role and years with current manager are important factors which contribute to whether the employee would stay with the company or not.

This report further discusses how each of these factors affect the attrition rate differently. Based on our analysis, we would like to recommend FermaLogis to take steps to improve the job satisfaction levels of the employees by starting and focusing on employee engagement programs and effective distribution of tasks and responsibilities which would help in a holistic development of employee skills.

## Background

FermaLogis is a renowned pharmaceutical company which now wishes to focus on the resources which it considers to be the most important- employees. Increased attrition, or voluntary resignations from the company deplete the talent pool and prevent the company from gaining returns from all the investments it has made to train the employees. Currently, the company’s primary focus is on young employees having 3 years of experience and above or experienced employees having more than 5 years of experience. The young employees are generally fresh grads, out of the college who do not have much practical and industrial ready skill set. By investing in them on the job training the company expected to harness their talent as they gain experience. However, if those employees leave soon thereafter, it results in huge loss for the company.

On the other hand, the company also invests a large amount of money to train the experienced employees by offering trainings on latest and modern skills required for the better functioning of the organization. Similarly, for the more experienced and senior employees as well, the company invests money in the form of executive programs and retreats. The rival companies consider these well-trained employees as a huge asset for them and try to attract them to join their companies.

Larry Hansen, the COO, is worried about the increased attrition and contacted our firm to look into the trends and the possible conclusions which can be drawn. Based on survival analytic techniques, we have derived some conclusions which would help FermaLogis to answer the following four questions:

1. **Who are leaving the company?**
2. **Why are they leaving?**
3. **When is the biggest danger for employees to leave?**
4. **Is there any difference of attrition between different employee groups or categories?**

# **METHODOLOGY**

## Exploratory Data Analysis

The Employee attrition data received from Larry Hansen, contains 76 variables like Age, Business Travel, etc, and 1470 observations. The variables are detailed in the Appendix. For our analysis we considered:

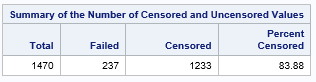
Years at the Company - Time to event

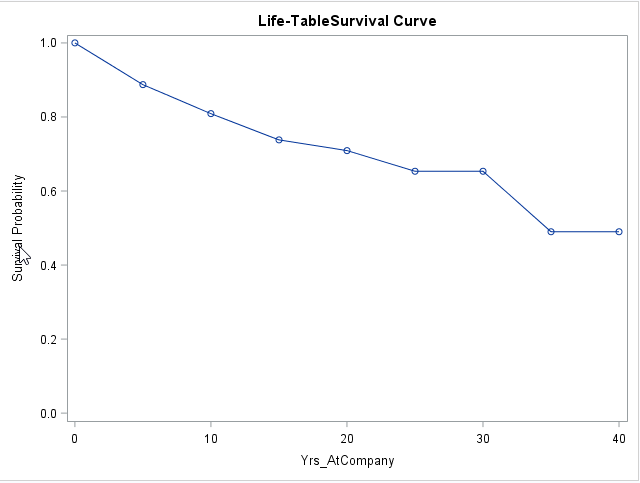
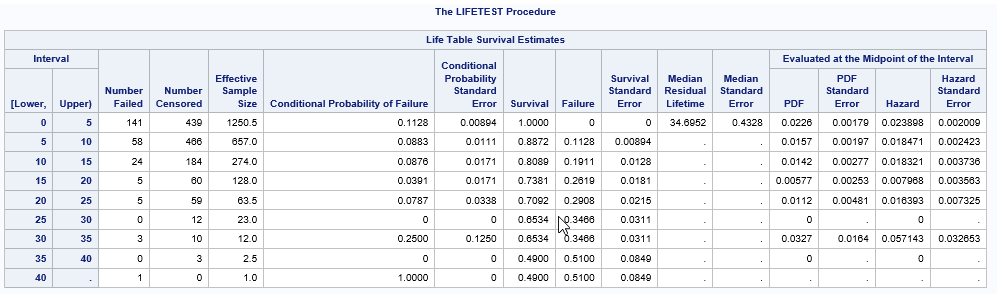
Attirition – Event (Censored=0 employee did not leave ; Censored=1 employee left)



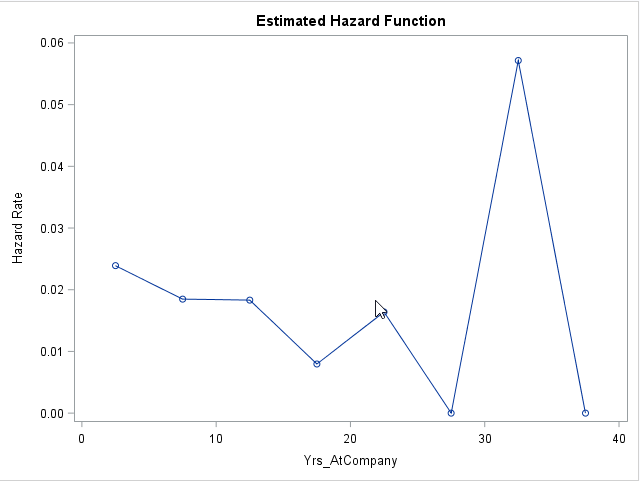
As shown in the above table, we created an additional variable “Attr1” with nominal values 0 and 1 which maps to “Attrition” variable ‘No’ and ‘Yes’ respectively, for ease of its further usage in survival analysis.

**LIFE TABLE SURVIVAL CURVES AND HAZARD FUNCTION**





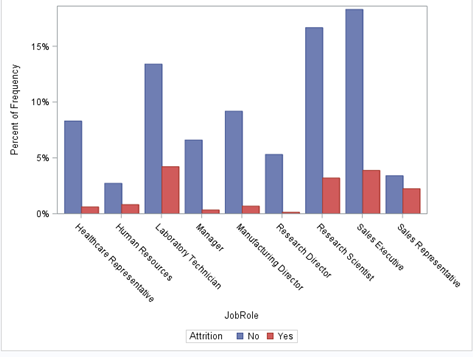
Here we used life table test to give an in-depth information about the standard error and conditional probability of failure which tells us what is the probability that the employee between intervals of 5 years at company will leave the company in the next 5 years at company. The Survival column gives the probability of survival (probability that he will not leave the company). Failure probability is (1- probability of survival). It gives us the probability of failure (probability that the employee will leave the company).



We used life table function to analyze the hazard rate of the company. Higher the survival probability lowers the hazard rate of the company. We observed that there was a *high hazard rate between the ages 30 and 40.* This is a serious implication that there have been more employees leaving in this age group.

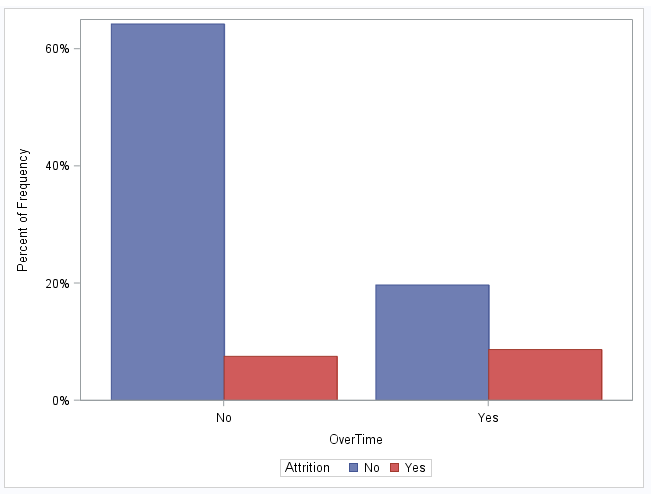
In order to undestand how each variable contributes to Attrition rate, we plotted Vertical box plots and Survival plots using Kaplan Maier Method and Life Test Method.

1. **Job Role**



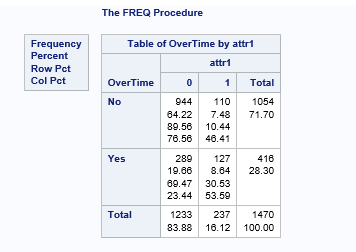
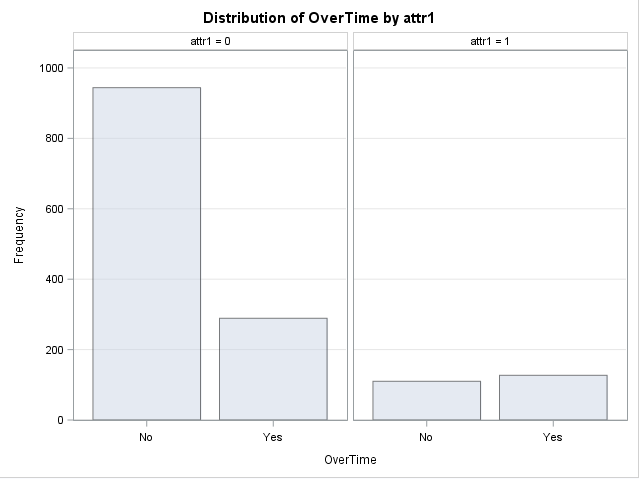
Here we observed that *laboratory technicians, Sales Executives, Research Scientist and Sales Representatives* have higher attirition rate suggesting that these job roles have something that is making these employee groups feel unwelcomed.

1. **Over Time**



Here we tried to visualize the relation between the employees who worked over time and not worked over time with the Attrition of the employees. We observed that the people who were working overtime were having higher propensity to leave the company than the ones who are not working overtime.

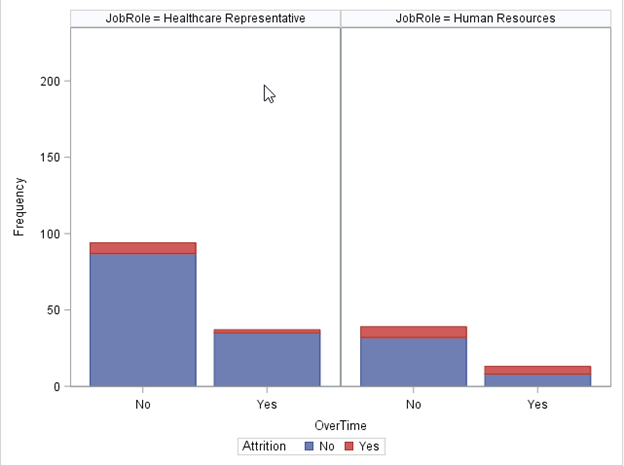
Here we tried to visualize the relation between overtime and Attrition of the company using Frequency procedure. Here we can actually see the exact count of the employees who quit the company and the ones for a particular reason like overtime, work life balance etc.



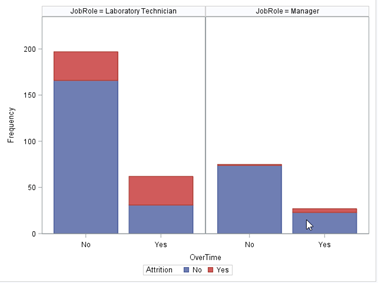
We observed that 30% of the employees left the company who did overtime and also, around 54% of the employees who left the company did overtime.

1. **Job Role and Over time**

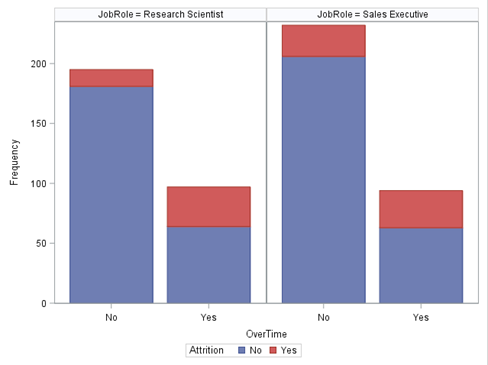
Extensive Exploratory Data Analysis on how Attrition is related to the employee job roles, who work over time or not, was done. This gave more insights about the job roles that have issues with working overtime.



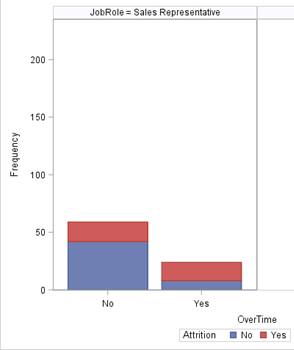
Healthcare Representatives and Human Resources employees who don’t work overtime tend to attrite more in the company.



Laboratory Technicians don’t appear to be affected by overtime as the ration of those leaving was equally divided between those who did or didn’t do overtime. We have seen a considerably high number of issues with this job role as many employees left the company from this role.

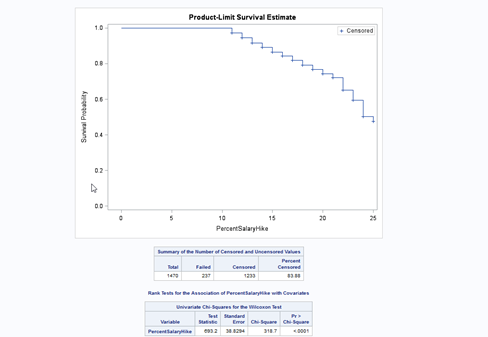


Research scientists and Sales Executives who worked over time left the company and this count seems to be high suggesting a higher concern level for the company.



Sales Representatives who either work or do not work over time equally tend to leave the company.

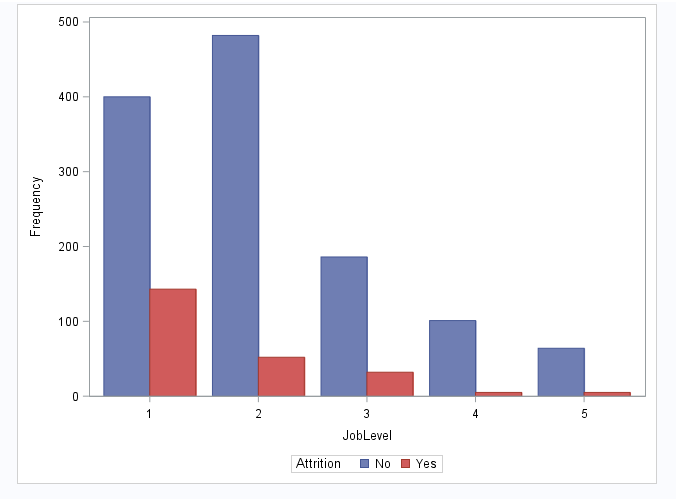
1. **Percent Salary Hike**



This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

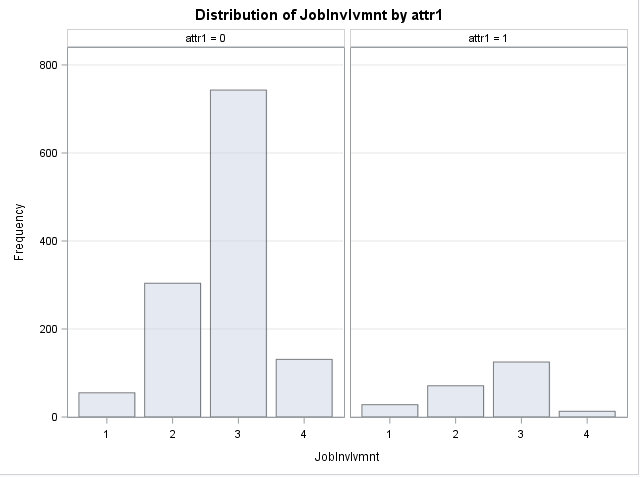
From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and there is a sudden decrease in the survival probability at 21% salary hike and at 24% salary hike suggesting that these have greatly affected the attrition rate of the company. It appears that employees with the highest salary hikes tend to start looking for more lucrative opportunities elsewhere. Sad but true! High performing employees receive more salary hikes and they are the people in demand in the market.

**5. Job level**



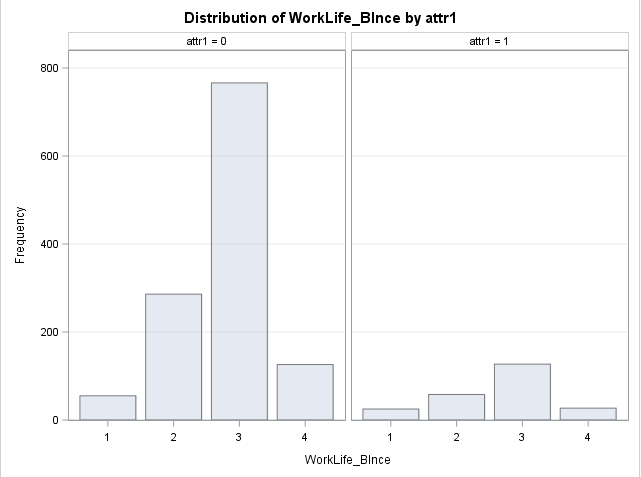
We observed that attrition is higher for Job level1 and attrition decreasing as Job Role is increasing.

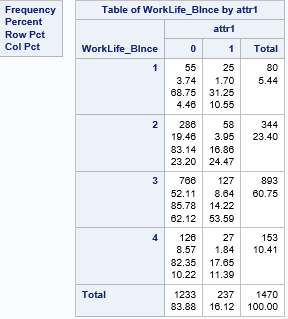
**6. Job Involvement**



From the above graph, it looks like when attrition is increasing with job involvement (1= low, 2=Medium, 3= High) which seems to be quite obvious as more & more involved an employee is with work higher the chance of attrition. This is probably because of new skills an emplyee might add which could result in higher attrition. But surprisingly, when an employee is very highly (4= very high) involved in job, his/her attrition is very less.

7. **Work life balance**

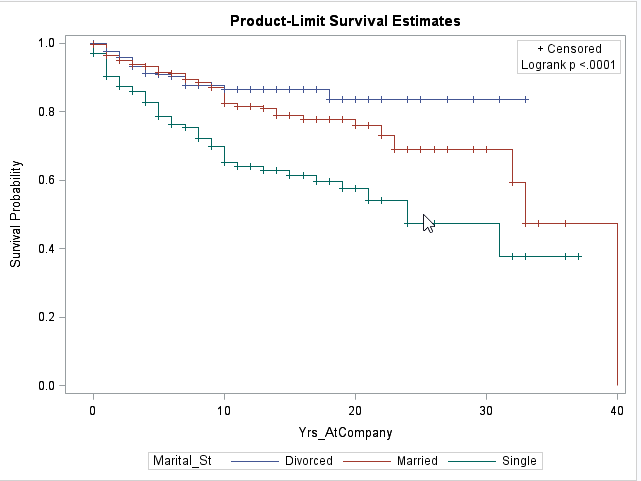




From the above graph and table, we can observe that highest attrition occurs when the work life balance of an employee falls under category 3 (3= better) . This is a point of interest, around 60% of total employee have responded work life balance as “better” and only 14.22% in that category are leaving the company. On the contrary, only around 5% of employees who have responded that work life balance as “bad” in the total of 31.25% are actually leaving the company. Company should look into this insight to better understand the impact of work life balance as an important parameter for employee attrition. Likewise for extensive analysis of variables impacting Attrition please refer Appendix section

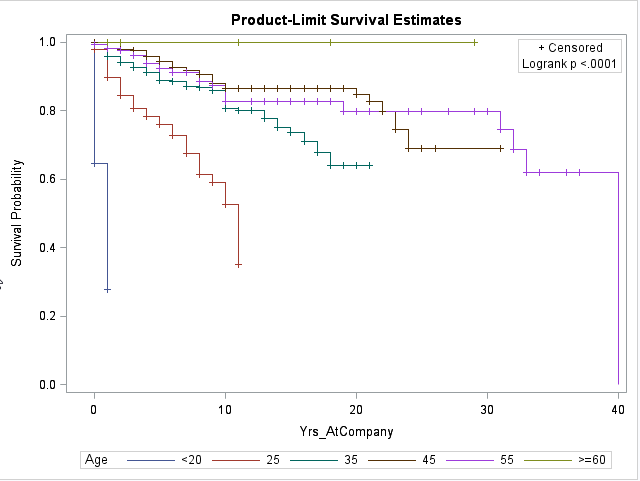
## Survival Curve analysis

1. **Years at the company with Marital status**



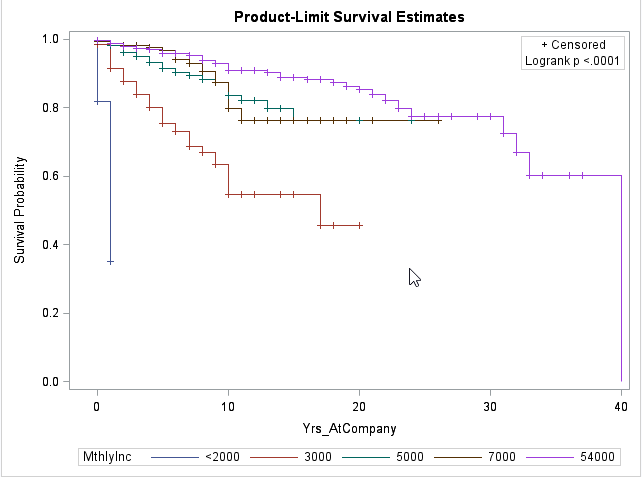
From above survival curve, we observed that the employees who are single have less survival probability than the employees who are married or divorced. It means that employees who are single have higher attrition in comparison to Married or divorced employees. The p-value suggests that the variable could be significant for our model prediction.

1. **Years at company with Age**



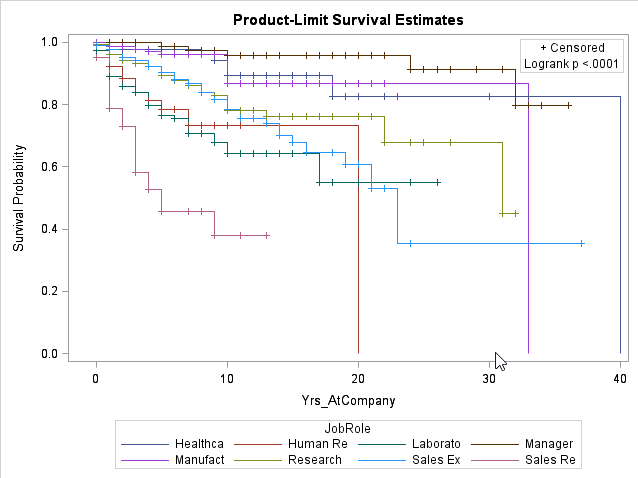
Here we observed that employees below the age of 20 and also employees between 20 and 25 age group have higher attrition in comparison to other age group members.

1. **Years at company with Monthly Income**



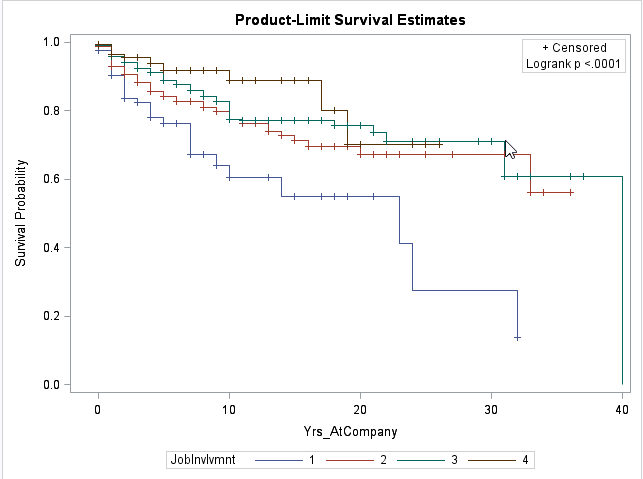
Here we observed that employees with a monthly income less than 2000 and monthly income between 2000 and 3000 are leaving the company and their survival probability is alarmingly low. The P value suggests that the variable could be significant for model prediction.

1. **Years at company with Job Role**



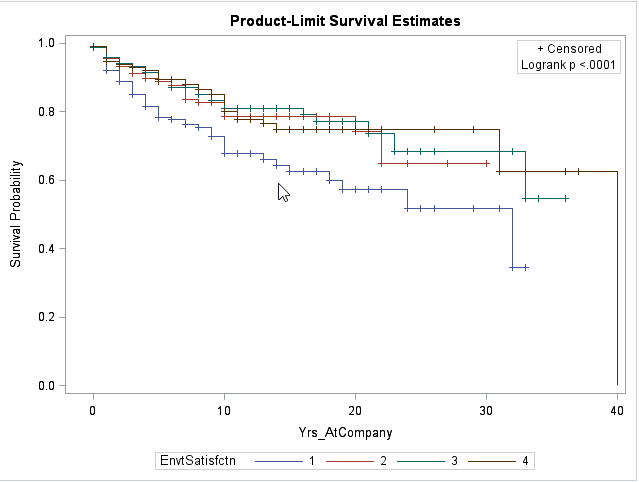
Here we observed that Sales Representative employees have very minimum survival probability than the rest of the job roles. The P value suggests that the variable could be significant for our model prediction.

1. **Years with company with Job involvement**



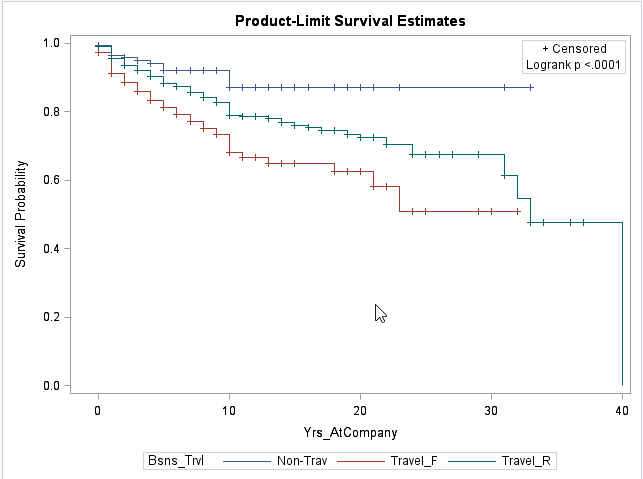
Here we tried to visualize the relation between different Job involvement levels and years at the company. Here we observed that there was a drastic decline of survival probability for the employees with very less job involvement in the company. The P value suggest that this variable could be significant for our model.

1. **Years at company with Environment satisfaction**



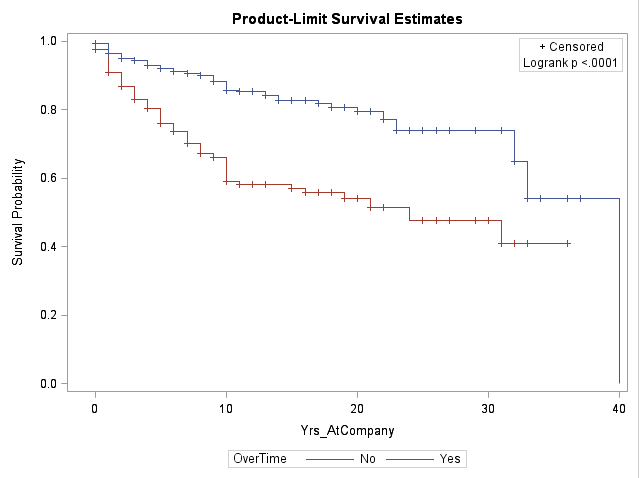
Here we tried to visualize the relation between different levels of environmental satisfaction and years at the company. We observed that people with very less environmental satisfaction left the company. Rest of the levels in the environmental satisfaction seem to go together. From the test of equality table, it is evident that the graphs are not equal. This variable could be significant due to its P value less than 0.05.

1. **Years at company with Business travel**



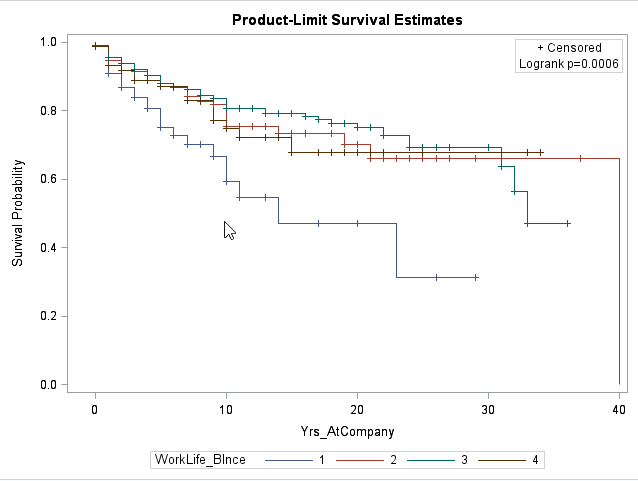
Here we tried to observe the significance of Business Travel variable on the survival probability of employees. We observed that the people or employees who were travelling frequntly had the least surivival probability in comparison to people who do not travel and travel rarely. Until 10 years in the company the surivival probability flattened out for the employees who were non travellers. The P value suggest that this variable could be significant for our model.

1. **Years at company with Overtime**



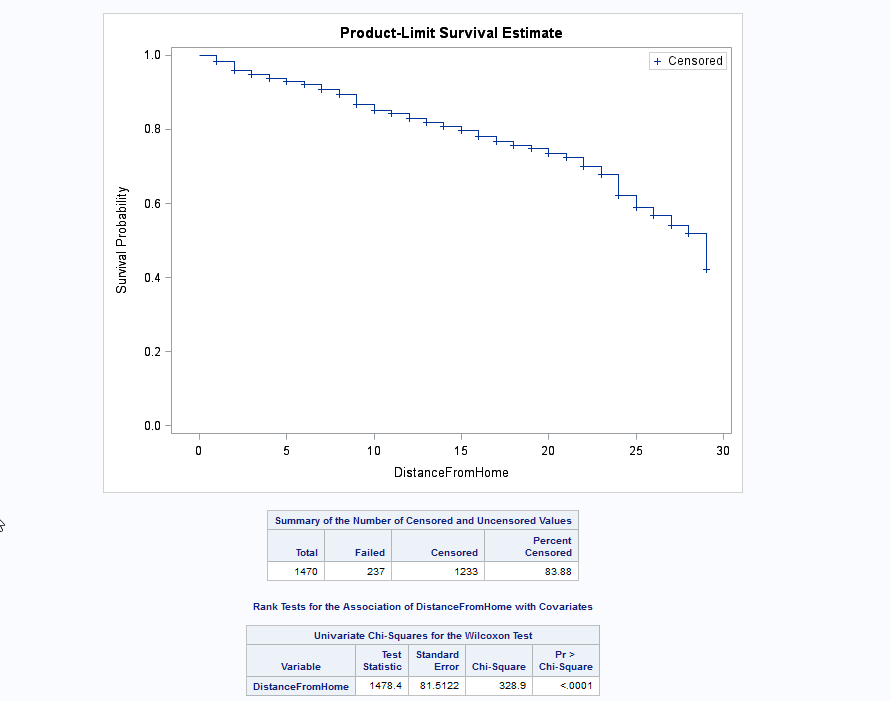
Here we tried to observe the significance of overtime variable on the survival probability of the employees. We observed that the people or employees who were working overtime, their survival probability is drastically going down than the employees who are not doing an over time. P-value suggest that this variable could be significant for our model.

1. **Years at company with Work life balance**



Here we observed that the Employees with a work life balance of 1 (1=Bad) have a very minimum survival probability than higher work life balance. The P value suggests that the variable could be significant for our model prediction.

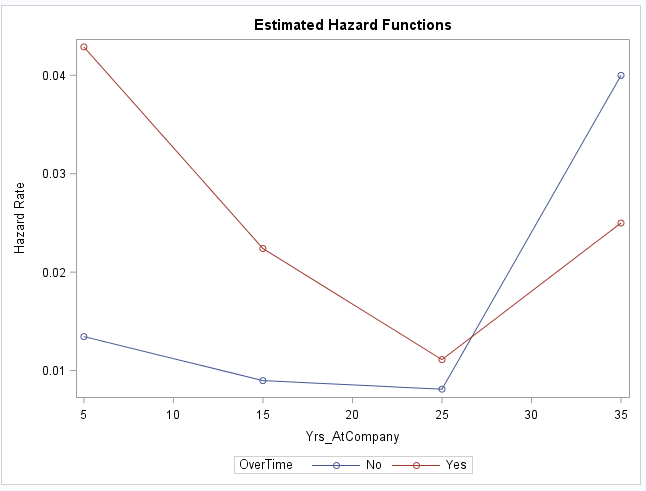
1. **Distance from home**



From above survival graph, we observe that employees living far away from company have lower survival probability in comparison to employees living near. P-value suggest that distance from home variable could be significant in our model. Survival curves, Frequency Plots and Hazard Function curves were analyzed for more co variates which is briefed in Appendix section.

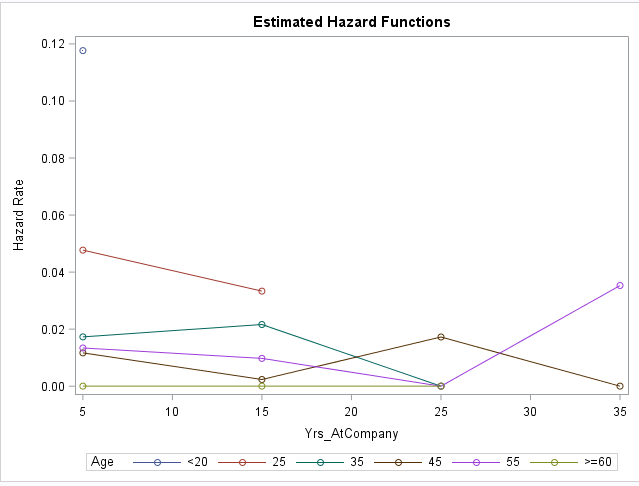
## Hazard Functions Plots

1. **Overtime:**



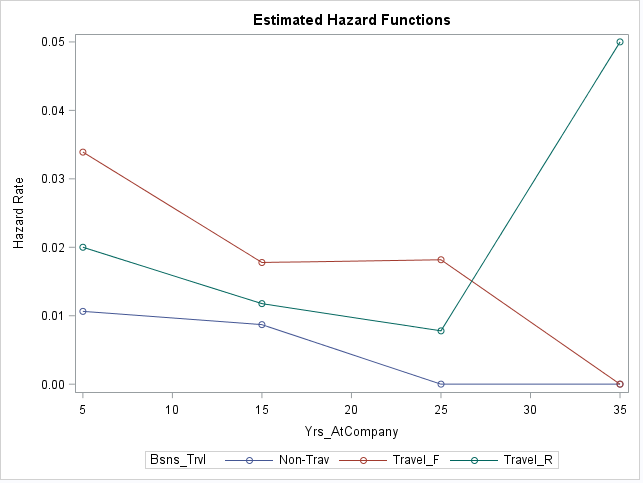
Employees who joined recently and do over time have high hazard rates compared to the employees who were with the company since long time (up to 25 years). Employees with more than 25 years of stay are hazardous irrespective of whether they do over time or not.

1. **Age:**



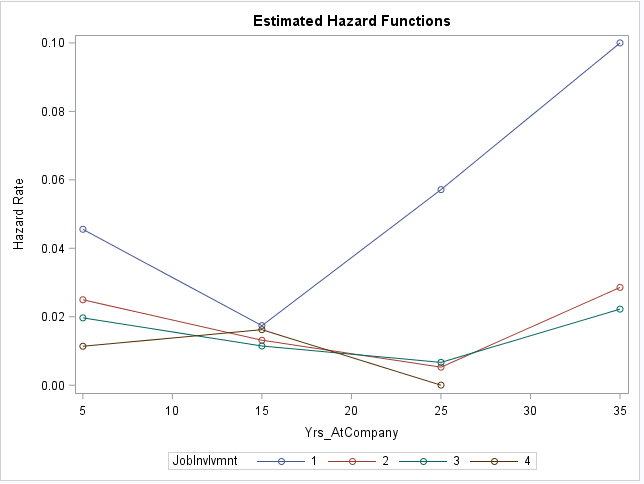
Employees under age of 25 and above 55 are having high hazard rates. Older employees might be retiring from job may be one genuine reason for their attrition.

1. **Business Travel**



Employees who travel regularly have high hazard rate when compared to other categories of travel as shown above.

1. **Job Involvement**



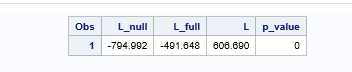
Employees with Job Involvement 1 are hazardous when compared to those employees who are highly involved in their job as shown above.

## Modeling Phase

This section details out the approach that we followed to address the employee attrition problem faced by FermaLogis. We used LIFETEST and LIFEREG PROCEDURE in SAS Base to plot survival curves and build regression models. The analysis on the complete dataset was done by following three approaches:

1. Compare full model (model built using all covariates or variables) and null model (model built using only time and censored variable).
2. Build final model with only significant covariates obtained using survival curves.
3. Dataset was divided into three groups based on tenure of employee in the company (Years at company). We referred to these groups as new employees, Medium employees and experienced employees.

We started out by comparing Full model (all covariates in the model) and Null model (no covariates in model) using Log Likelihood ratio test.

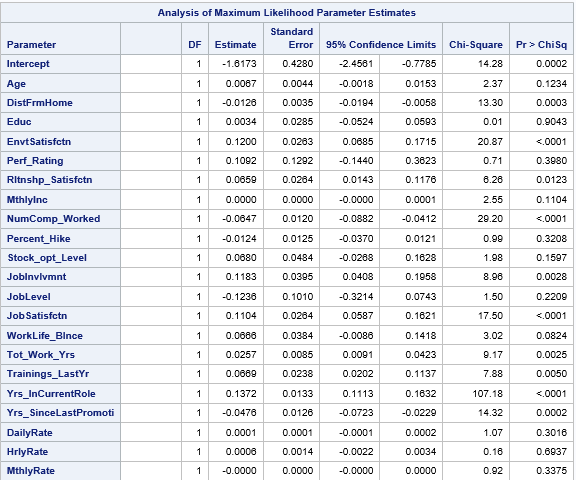


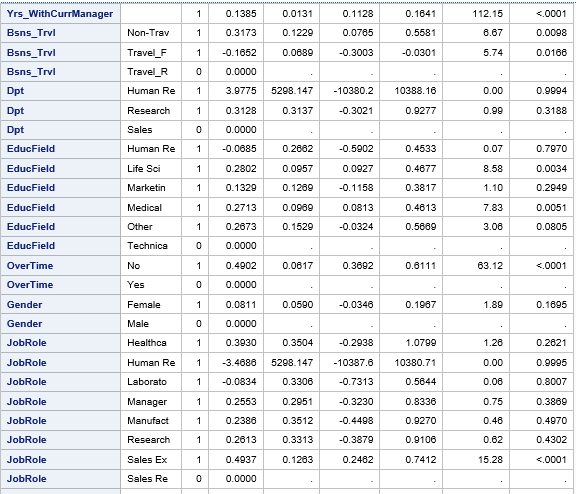
From above test we observe that p-value = 0. This implies at least one of the variables has co efficient different from 0. Hence full model will perform much better as compared to the null model.

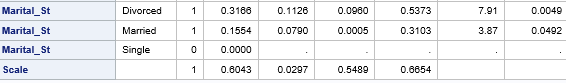
Please refer to appendix for further details of this test.

Further, Full model was run using different distribution like Weibull, Gamma, Lognormal and Exponential to conduct goodness of fit test to decide which shape is the best description for the true hazard function using Log Likelihood Ratio test.

Log Likelihood Ratio of WEIBULL distribution is almost close to that of GAMMA. But model built using WEIBULL Distribution encountered an error *Validity of the model fit is questionable*. Hence we have built our final model using *Lognormal distribution.*



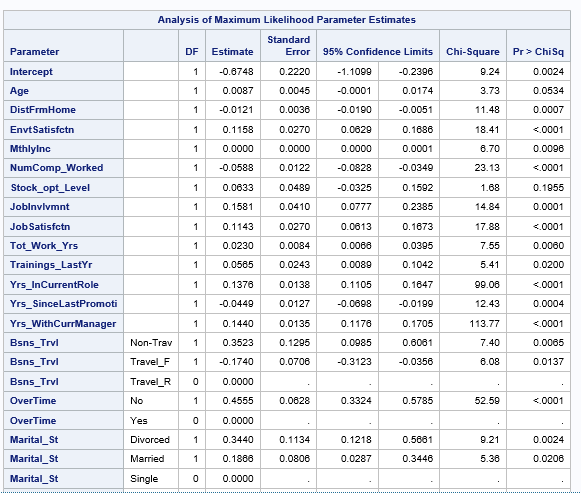




*Parameter estimates of covariates used in Lognormal model*

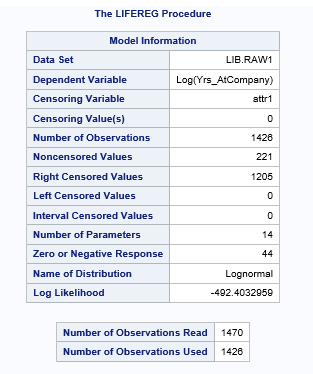
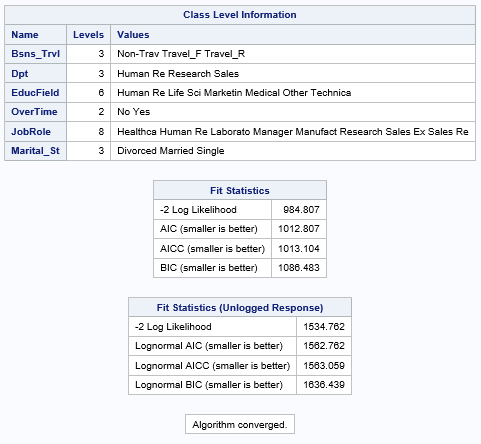
Finally, the full model using lognormal distribution was analyzed for parameters having high significance based on their p-values. Covariates having p-value > 0.05 (considering α = 5%) were ignored due to their non-contribution to attrition event.

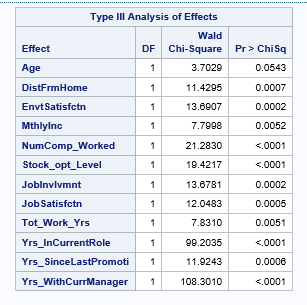
For more details about the models built please refer to Appendix.

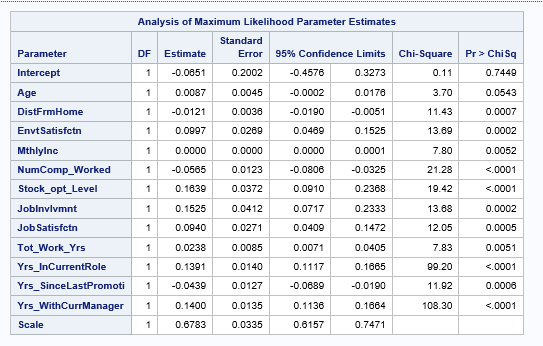


*Parameter estimates of the covariates used in final model (Lognormal Full model)*

For second approach, we built model using covariates having significance to attrition event analyzed using the survival curves as briefed in EDA section of this report. This was done as a part of our execution strategy to filter out covariates that could have significant impact on attrition event and would be useful in our model building approach. From first approach it was clear that Full model using lognormal distribution performed superior to other distribution hence we extended the same approach and followed a Lognormal distribution using only significant covariates.

Below screenshots shows the results of Log Normal model built using only significant co variates.  

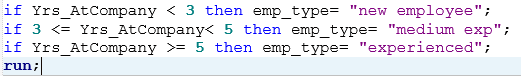




From above table we can observe that the covariates are significant (based on p-value) hence the above covariates are contributing to the attrition rate.

As third approach, we created a new variable “Employee Type” (Emp\_type) with 3 levels based on the tenure of employee in company (Years in the company) to address the Larry’s concern of young employees and Experienced employees leaving the company.

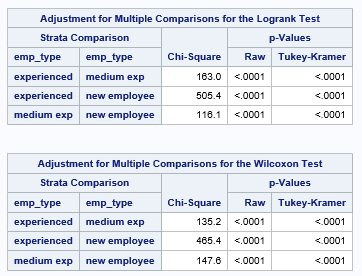
1. *New employees*: Years in company less than 3 years
2. *Medium Experienced*: Years in company greater than or equal to 3 years and less than 5 years
3. *Experienced:* Years in company greater than 5 years.

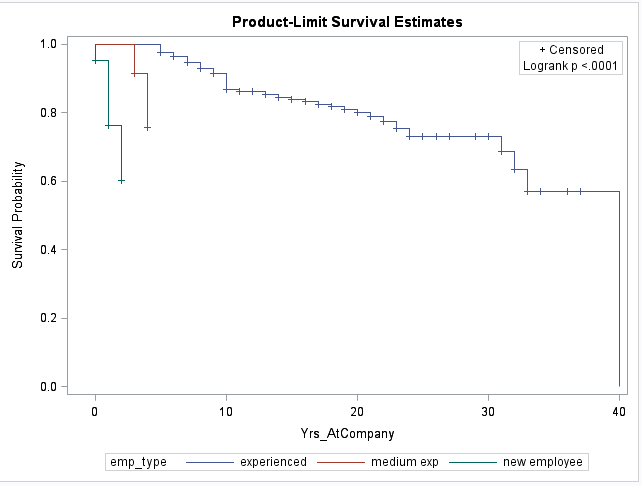


Further we applied competing risks concept to find out if we can combine the employee (event) types to build the model or we need to model them individually.

1. **Lifetest procedure to graphically test for linear relation between type hazards – Without covariates**







From above table and graphs we observe that p value is significant, indicating that each employee group are not the same, thus each group needs to be modelled individually.

1. **Life test procedure to graphically test for linear relation between type hazards – With covariates**

We divided the given data set into three subsets based on the levels in the variable “Employee type”. We conducted Log likelihood ratio test to investigate if co-efficient found for each event is same for the entire population as well. If there is a difference, then we need to build separate models for each event (employee) type.

C:\Users\pavra\Desktop\unnamed.png

Since p-value is 0, same co-efficient cannot be used to model different event types.

After conducting these tests, it was evident that we need to build separate models for each event type.

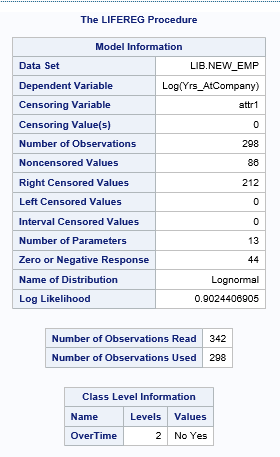
For more details about the test conducted, please refer Appendix section.

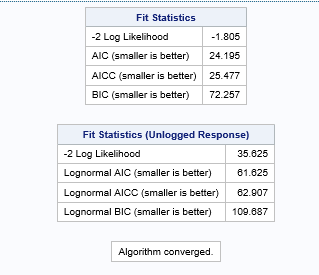
***Also team found that, this is best approach to address Larry’s question “who are leaving company and why they are leaving”.***

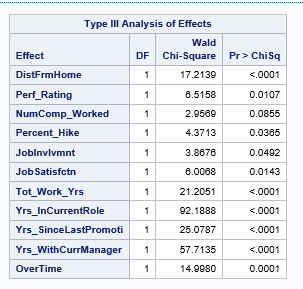
From first approach it was clear that Full model using lognormal distribution performed superior to other distribution hence we extended same approach to build models for three employee groups.

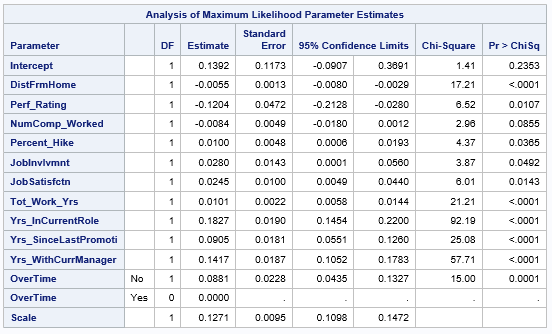
1. **Model - New Employee group**

As a part of final model selection, covariates having p-value > 0.05 were ignored due to their non-contribution to attrition event. For more details about non-significant co- variates in each of this sub group, please refer to Appendix section.



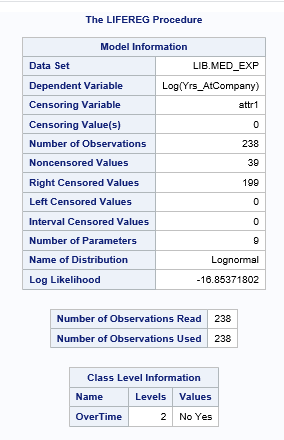


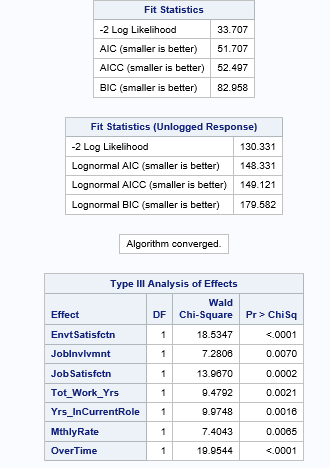


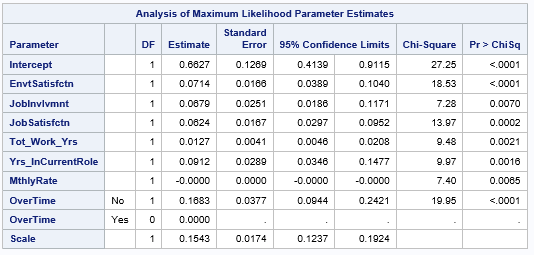


From above table we can observe that the covariates are significant (p < 0.05) hence these variable are contributing to the attrition rate in New employee group.

1. **Model - Medium Experienced Employees group**

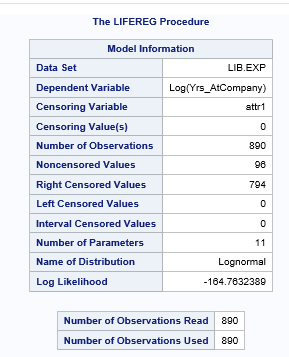


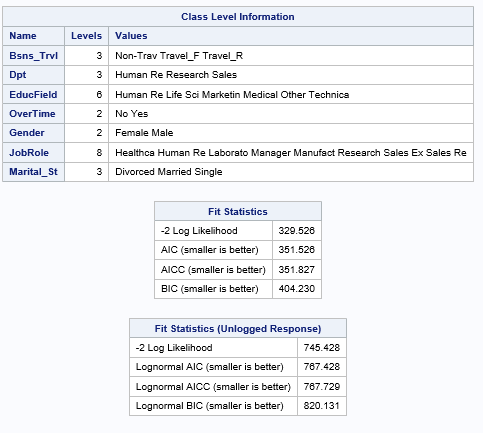


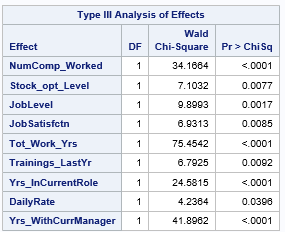


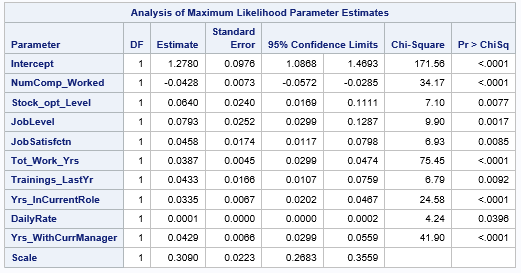
From above table we can observe that the covariates are significant hence these variable are contributing to the attrition rate.

1. **Model - Experienced Employees group**









From above table we can observe that the covariates are significant hence these variable are contributing to the attrition rate in experienced employee group.

Table below summarizes the variables which contribute significantly for attrition of employees in each employee type group.

|  |  |
| --- | --- |
| **Model – Employee type** | **Significant Covariates** |
| **All the employees**  **(Years in company 0 to 40)** | Business Travel  Over Time  Marital Status  Age  Distance From Home  Environment Satisfaction  Monthly Income  Number of Companies Worked  Stock option Level  Job Involvement  Job Satisfaction  Total Working Years  Years In Current Role  Years Since Last Promotion  Years With Current Manager |
| **New Employee**  **(Years in company < 3)** | Over Time  Distance From Home  Performance Rating  Number of Companies Worked  Percent Salary Hike  Job Involvement  Job Satisfaction  Total Working Years  Years In Current Role  Years Since Last Promotion  Years With Current Manager |
| **Medium Experience**  **(3<= Years in company < 5)** | Environment Satisfaction  Job Involvement  Job Satisfaction  Total Working Years  Years In Current Role  Monthly Rate  Over Time |
| **Experience**  **(Years in company => 5)** | Over Time  Number of Companies Worked  Stock option Level  Job Level  Job Satisfaction  Total Working Years  Training Times Last Year  Years In Current Role  Daily Rate  Years With Current Manager |

**Model Equations**

1. Model equation for full model with significant variables:

Yrs\_AtCompany = -0.6748+ Age\* 0.0087 - DistFrmHome\* 0.0121 + EnvtSatisfctn\* 0.1158 + MthlyInc\* 0.0000 - NumComp\_Worked\* 0.0588 + Stock\_opt\_Level \* 0.0633 + JobInvlvmnt \* 0.1581 + JobSatisfctn \* 0.1143 + Tot\_Work\_Yrs \* 0.0230 + Trainings\_LastYr \* 0.0565 + Yrs\_InCurrentRole \* 0.1376 - Yrs\_SinceLastPromoti \* 0.0449 + Yrs\_WithCurrManager \* 0.1440 + Bsns\_Trvl(Non-Travel) \* 0.3523 - Bsns\_Trvl(Travel\_F) \* 0.1740 + Bsns\_Trvl(Travel\_R) \* 0.0000 + OverTime(No) \* 0.4555 + OverTime(Yes) \* 0.0000 + Marital\_St(Divorced) \* 0.3440 + Marital\_St(Married) \* 0.1866 + Marital\_St(Single) \* 0.0000

1. Model equation for full model with significant strata variables:

Yrs\_AtCompany = -0.0651+ Age\* 0.0087 - DistFrmHome\* -0.0121 + EnvtSatisfctn\* 0.0997+ MthlyInc\* 0.0000 - NumComp\_Worked\* 0.0565+ Stock\_opt\_Level \* 0.1639+ JobInvlvmnt \* 0.1525+ JobSatisfctn \* 0.0940 + Tot\_Work\_Yrs \* 0.0238 + Yrs\_InCurrentRole \* 0.1391 - Yrs\_SinceLastPromoti \* 0.0439+ Yrs\_WithCurrManager \* 0.1400 + Scale \* 0.6783

1. Model equation for new employees with significant covariates:

Yrs\_AtCompany = 0.1392 - DistFrmHome\* -0.0055 - Perf\_Rating \* 0.1204 - NumComp\_Worked \* 0.0084+ Percent\_Hike \* 0.0100 + JobInvlvmnt \* 0.0280 + JobSatisfctn \* 0.0245+ Tot\_Work\_Yrs \* 0.0101+ Yrs\_InCurrentRole \* 0.1827- Yrs\_SinceLastPromoti \* 0.0905+ Yrs\_WithCurrManager \* 0.1417 + OverTime(No) \* 0.0881+ OverTime(Yes) \* 0.0000 + Scale \* 0.1271

1. Model equation for med level employees with significant covariates:

Yrs\_AtCompany = 0.6627 + EnvtSatisfctn\* 0.0714 + JobInvlvmnt \* 0.0679 + Job Satisfaction \* 0.0624 Tot\_Work\_Yrs \* 0.0127+ Yrs\_InCurrentRole \* 0.0912 - MthlyRate\* 0.0000 + OverTime(No) \* 0.1683+ OverTime(Yes) \* 0.0000 + Scale \* 0.1543

1. Model equation for Experienced level employees with significant covariates:

Yrs\_AtCompany = 1.2780 - NumComp\_Worked\* 0.0448 + Stock\_opt\_Level \* 0.0640+ JobLevel \* 0.0793 + JobSatisfctn \* 0.0458 + Tot\_Work\_Yrs \* 0.0387 + Trainings\_LastYr \* 0.0433 + Yrs\_InCurrentRole \* 0.0335 + DailyRate \* 0.0001 + Yrs\_WithCurrManager \* 0.0429 + Scale \* 0.3090

# **Conclusions**

Based on the analysis conducted, certain conclusions can be drawn about the trends being observed in the attrition of the employees at FermaLogis:

1. Laboratory Technicians, Sales Executives, Research Scientist and Sales Representatives have higher attirition rate as is evident from the exploratory data analysis.
2. Unmarried employees and those having lower salaries are also more likely to leave the company.
3. New employees (having less than 3 years of experience) are the most likely to leave the company followed by medium experienced employees and then the highly experienced ones (more than 5 years)
4. For new employees, distance from home is a very significant factor in attrition and has a negative coefficient, indicating larger distances from home propel people to stay in the company for a lower time. The other factors, apart from distance from home, which would make up the top 3, are Years in current role and Years with Current manager.
5. For medium experienced employees the top 3 factors are – Environment Satisfaction, overtime and Job satisfaction.
6. For highly experienced employees, the most affecting factors are – Number of companies worked, years in current role and years with current manager.
7. It can be seen that the differently experienced employee classes have different priorities. While younger and newly joined employees are more concerned about the distance from home, medium experienced employees care more about the job satisfaction. Highly experienced employees tend to look for more responsibilities and if they feel that their careers have stagnated, due to working the same job for many years, they would consider changing jobs.
8. Years with current manager appears as a chief concern for two sub-groups. It shows that the immediate boss has a very important role to play in ensuring a proper work environment for the subordinates and in their adequate job satisfaction and skill enhancement.
9. Number of companies worked appears as a significant attribute in the model for more experienced employees. It points out the fact that a person who has a track record of switching jobs and changing companies would be more likely to do so once again.
10. Marital status and monthly salary appear to be significant factors for the combined model, but when considered over individual classes, none of the sub-groups appear to have significant effect of these factors. This means that over large data, these factors tend to draw more significance.
11. Overtime, Job involvement and job satisfaction are reported to be significant by all the three subgroups.

# **Recommendations**

After thorough analysis, below are our recommendations to FermaLogis which could help reduce the employee attrition at their company.

1. Job Involvement, Job satisfaction and Years at a current role appear to be concerning almost all the employees. Based on the models as well as the conclusions drawn from the exploratory data analysis as well we can see that the appropriate levels of job involvement are needed. It makes the employees more associated and related to the tasks he/she is doing and makes them feel that their work is valued. Very high job involvement reflects high dependence of the company on those employees. This may make them feel overworked and look for more relaxed opportunities elsewhere. FermaLogis should, therefore, try to equalize job involvement and redistribute tasks effectively.
2. Distance from Home is a crucial factor for new hires. Employees who have recently graduated or just started working doesn’t seem to welcome the change where they have to be a large distance from their homes. At the same time the company should not focus exclusively on hiring everyone from the local areas as it will have an effect on the talent pool of the company. To counter this problem, the company needs to strike a balance between effective compensation in the form of rewards and bonuses among new employees and developing efficient talent pools in the local business ecosystem.
3. Years with current manager appears to be a significant factor. The coefficient for the variable is positive, indicating higher the years with a particular manager, higher the tendency of employees to stay. Combined with the emphasis employees place on job involvement and job satisfaction, this reflects a very important role which the immediate supervisors need to play. Better training of supervisors in these areas can help them better understand their subordinates and would help in boosting employee morale across the board.
4. Number of companies worked is a significant factor. It shows that employees who have a tendency to regularly switch companies would be very reluctant to hold on to their jobs at FermaLogis as well. The company should make certain modifications in its job involvement approach towards such employees- to give them better and challenging opportunities. Also, the hiring contracts and incentives should be appropriately designed to protect the company from eventual leaving of the employee.

# APPENDIX

**Data Dictionary**

**Age:** age of the employee when this dataset was created

**Attrition:** shows whether the employee left the company or not

**Business Travel:** shows how much travel employee makes

**DailyRate**: daily compensation of employee before any cuts/taxes

**Department**: shows the department of the employee when this dataset was created

**DistanceFromHome:** commuting distance for the employee in miles

**Education:** 1 'Below College'

2 'College'

3 'Bachelor'

4 'Master'

5 'Doctor'

**EducationField:** shows the education field of the employee

**EmployeeCount:** a field used for aggregation calculations

**EmloyeeNumber:** the ID of the employee

**EnvironmentSatisfaction:** a score showing how much the employee is satisfied with company's facilities

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

**Gender:** shows the gender of the employee

**HourlyRate:** hourly compensation of employee before any cuts/taxes

**JobInvolvement**: a score given to the employee by supervisors how much the employee is involved in company's operations

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

**JobRole:** shows the job role of the employee in the company

**JobLevel:** shows the management level of the employee

**JobSatisfaction:** shows the last survey result of the employee about his\her job satisfaction

**MaritalStatus:** shows the marital status of the company

**MonthlyIncome:** shows the monly income of the employee

**MontlyRate:** monthly compensation of employee before any cuts/taxes

**NumCompaniesWorked:** the number of companies the employee worked before starting in the company

**Over18:** shows whether the employee is over 18 years old.

**OverTime:** shows whether employee works overtime more than 10 hours a week

**PercentSalaryHike:** shows the agreed yearly salary rise percent

**PerformanceRating:** a score given to the employee by supervisors how good was the performance of the employee last year

1 'Low'

2 'Good'

3 'Excellent'

4 'Outstanding'

**RelationshipSatisfaction:** shows the last survey result of the employee about his\her satisfaction with other employees in the company

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

**StandardHours:** number of hours employee works for one payroll period (two weeks)

**StockOptionLevel:** shows the stock option for the employee. If your analyses give significant results for this variable, you can refer to that group as "employees having stock option level x"

**TotalWorkingYears**: shows the time the employee worked as a professional (at any company)

**TrainingTimesLastYear:** shows the number of training programs employee has attended last year

**WorkLifeBalance:** shows the employee satisfaction of the work load (4 is the highest satisfaction level)

1 'Bad'

2 'Good'

3 'Better'

4 'Best'

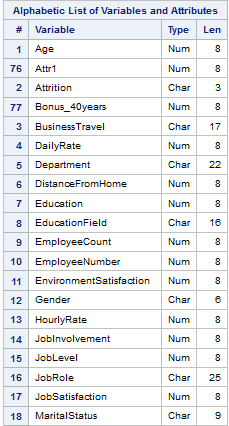
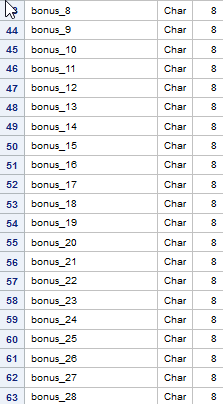
**YearsAtCompany:** Tenure at the company

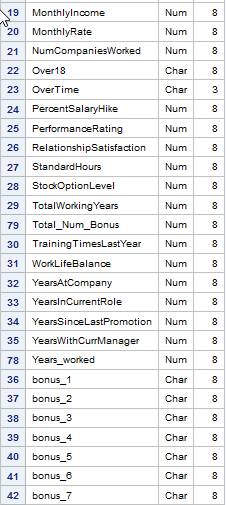
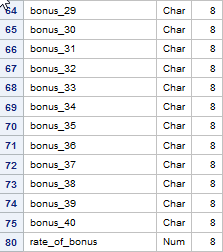
**YearsInCurrentRole:** the number of years employee works in the current position

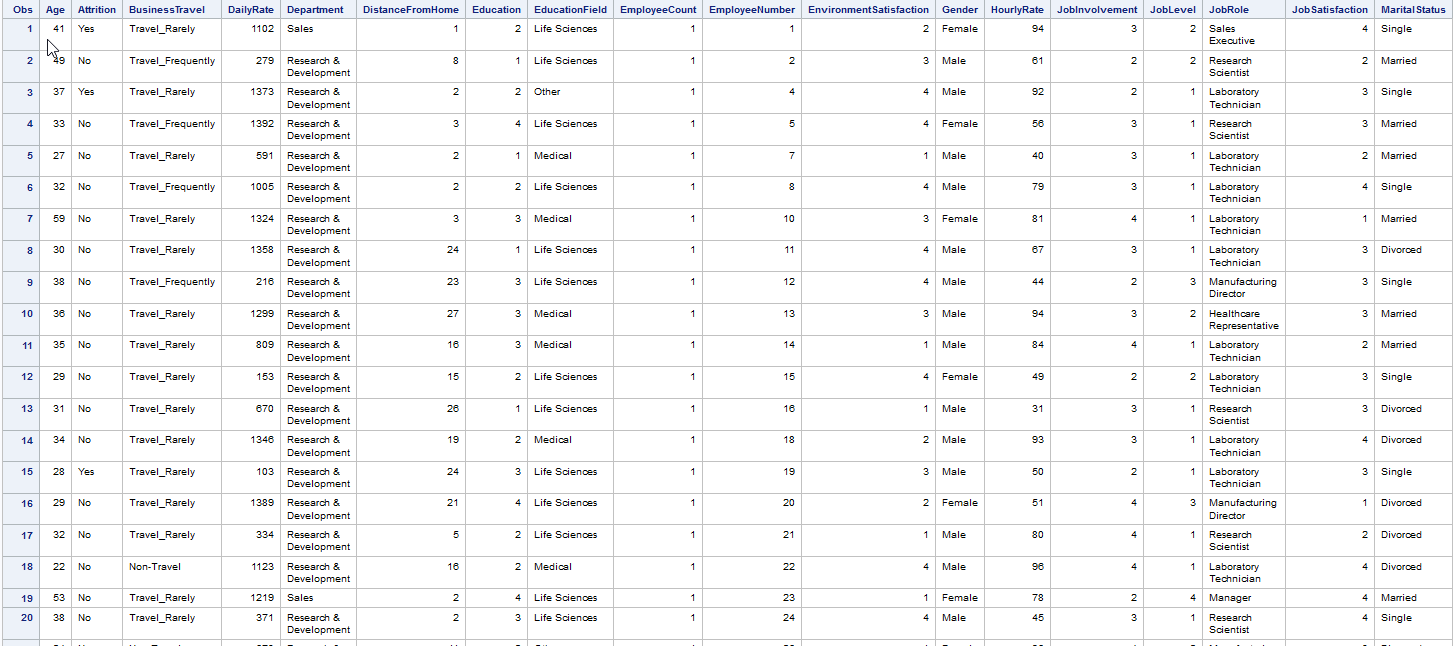
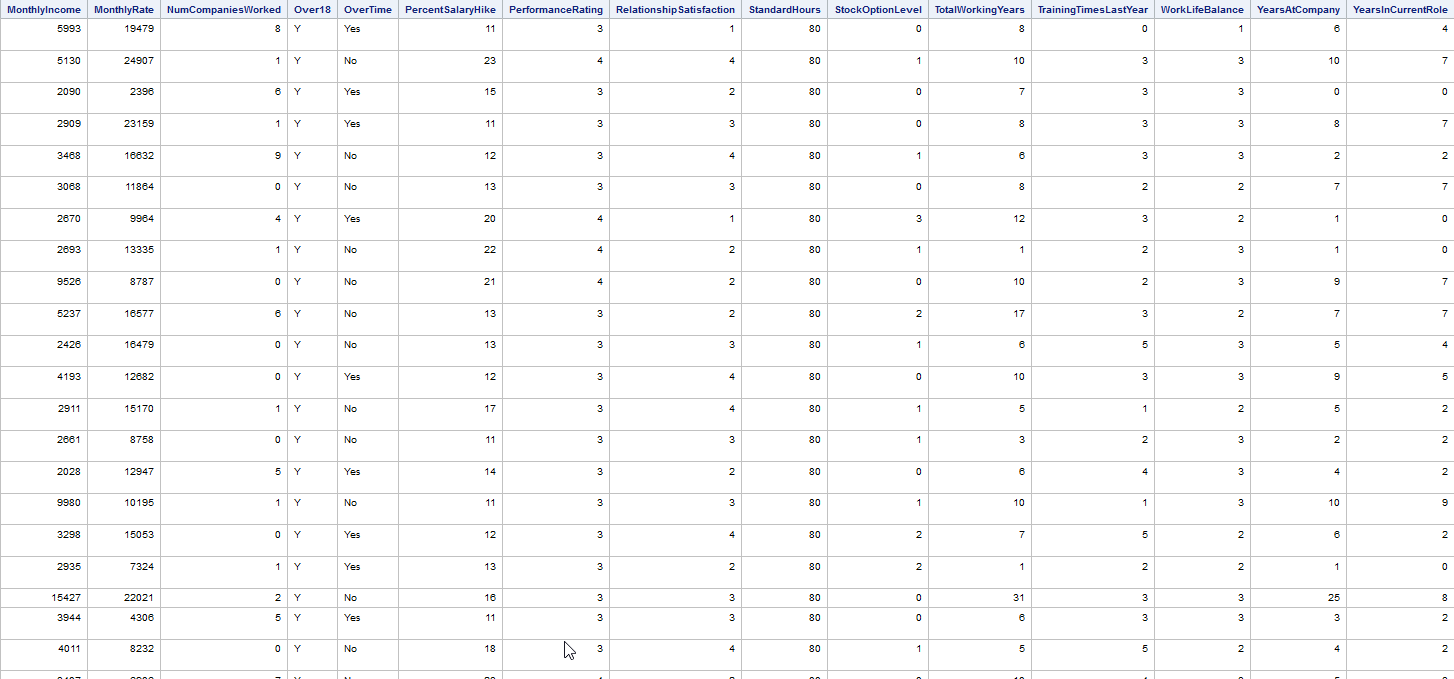
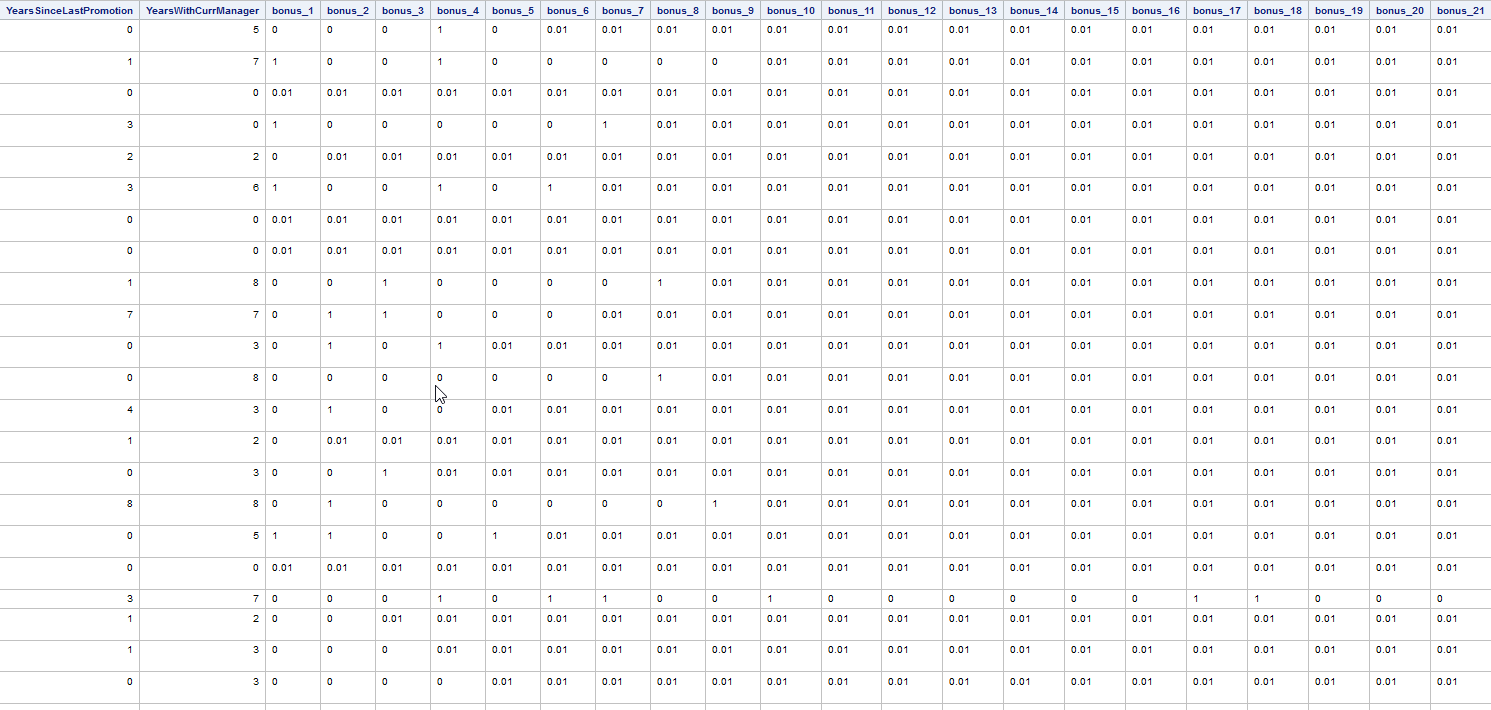
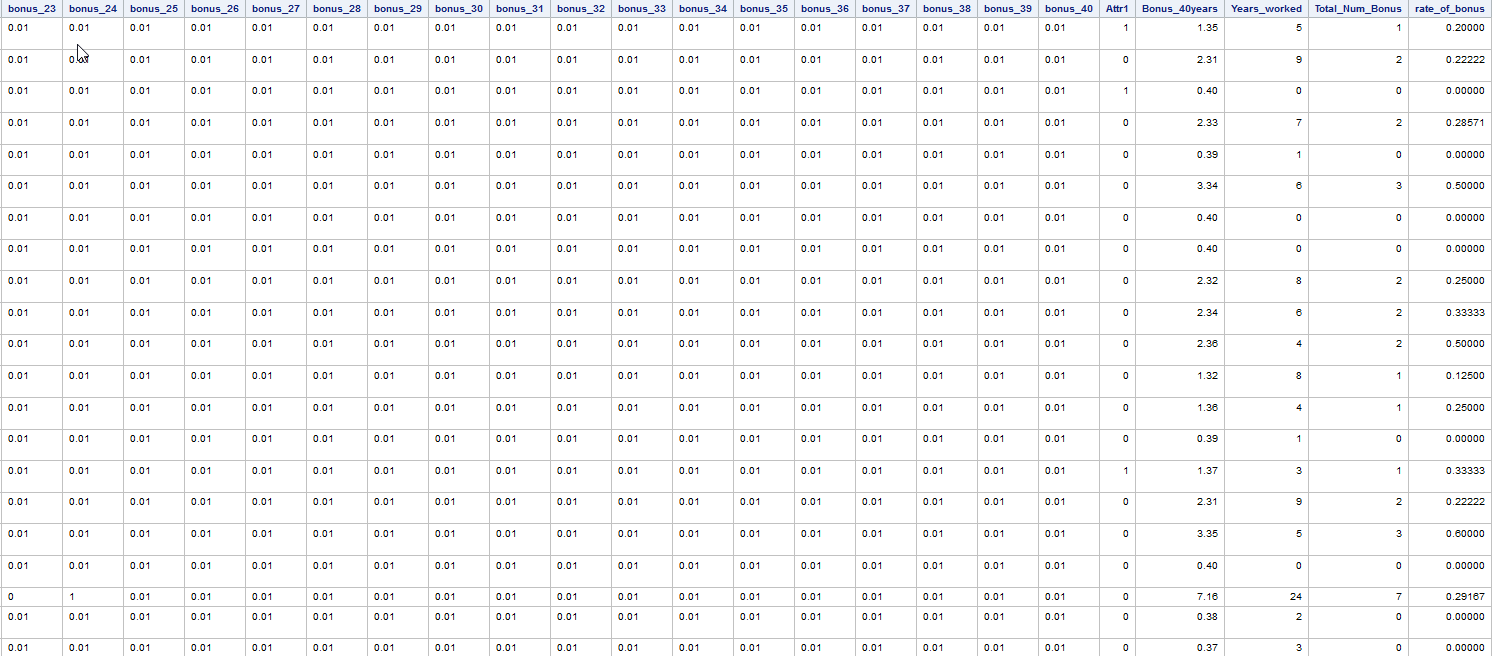
**YearsSinceLastPromotion:** shows the number of years passed since the last promotion

**YearsWithCurrentManager:** shows the number of years with the current supervisor.

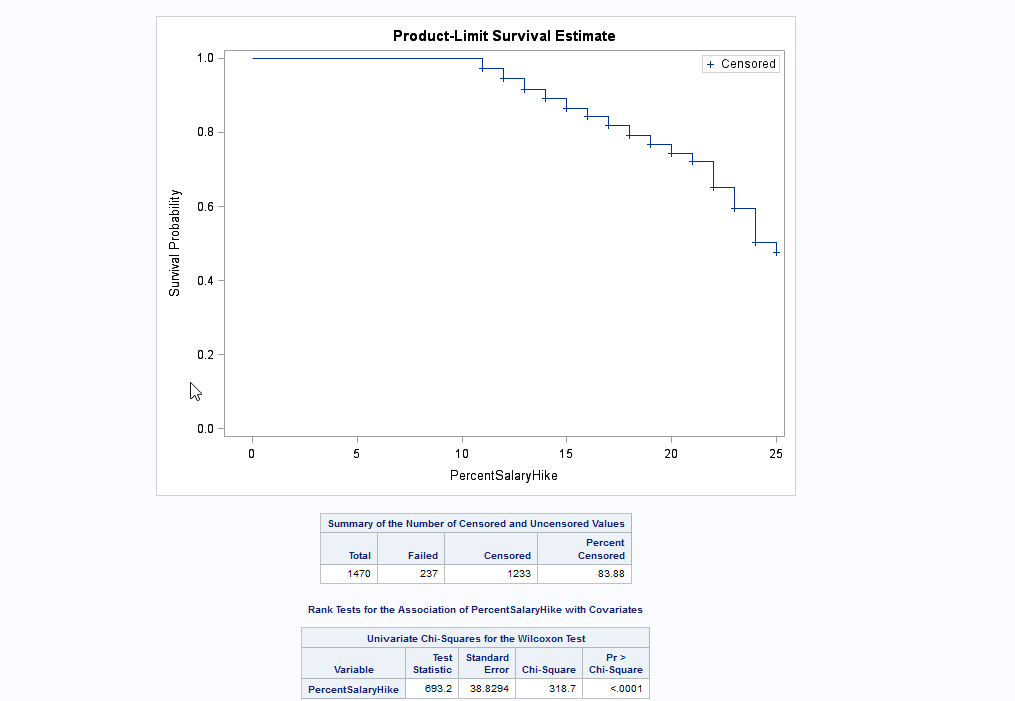
**bonus\_1-40:** shows whether the employee received bonus payments in the last 40 years. bonus\_1 is last year.

**Data Read in SAS**



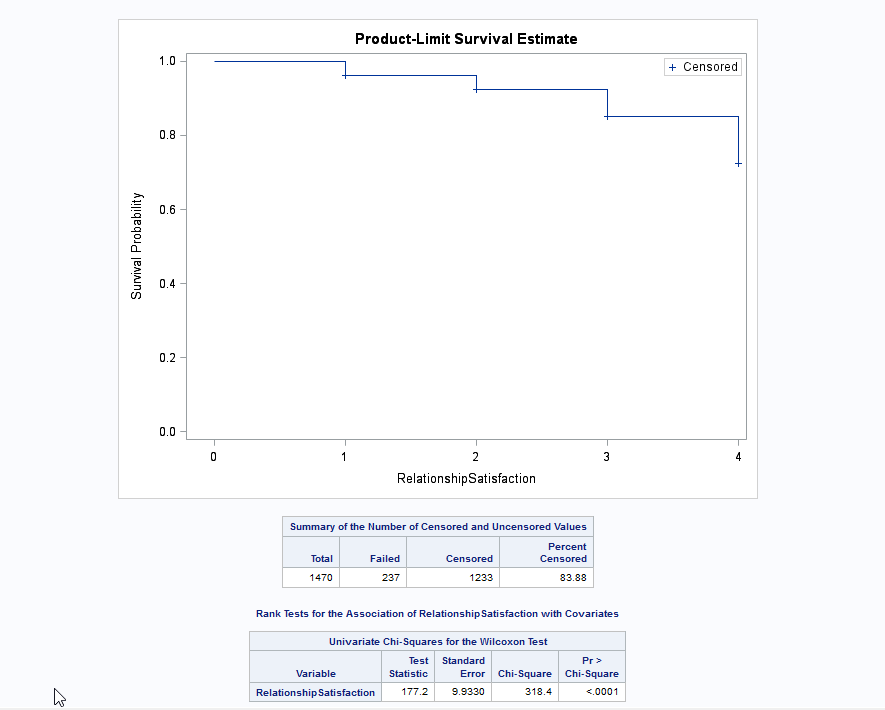
   

**Exploratory Data Analysis**



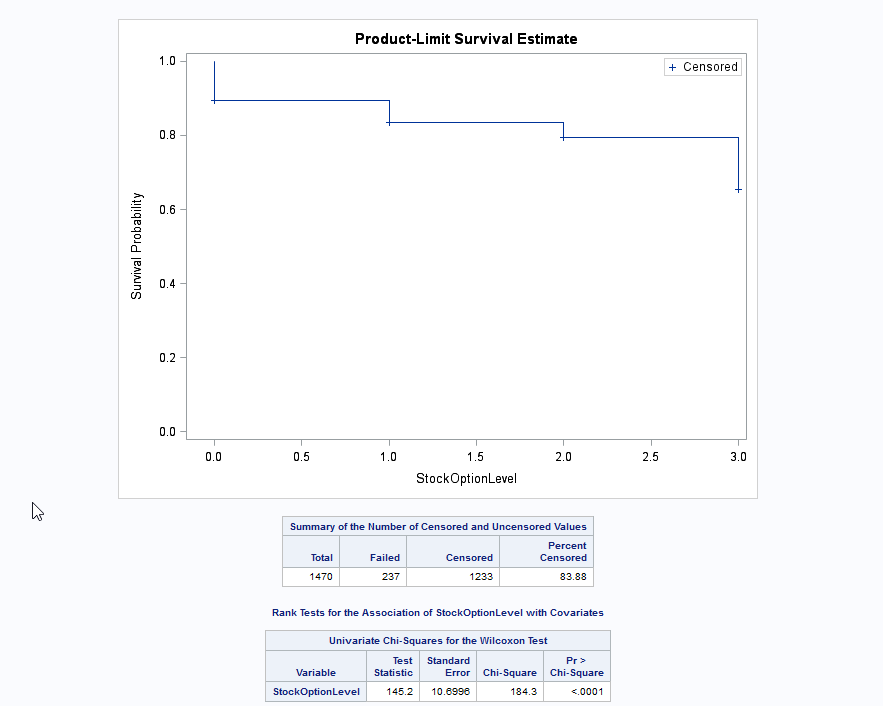
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and there is a sudden decrease in the survival probability at %21 salary hike and at %24 salary hike suggesting that these have greatly affected the attrition rate of the company.



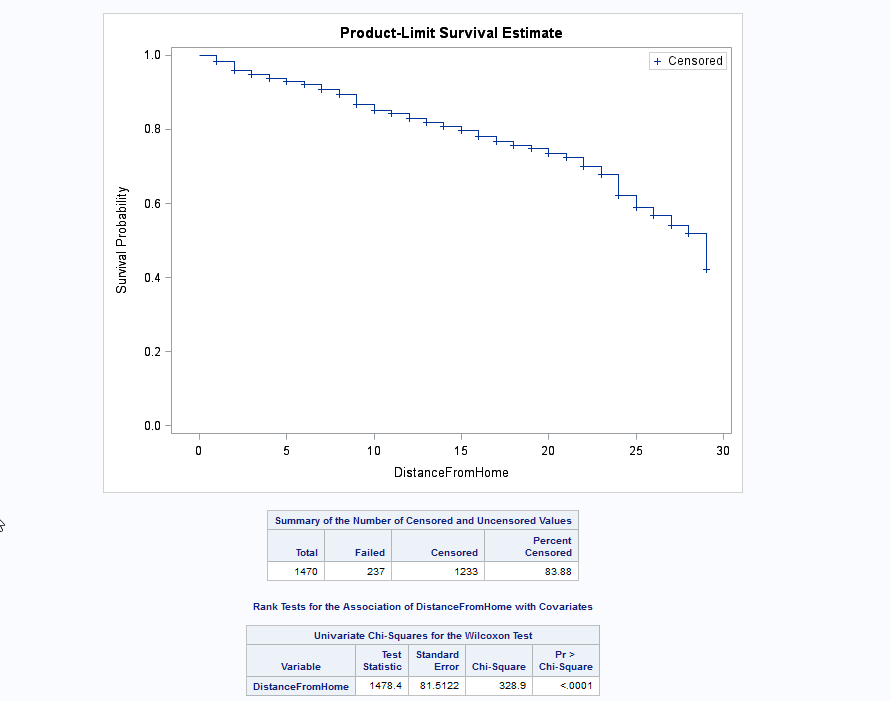
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable affected the attrition rate very little of the company.



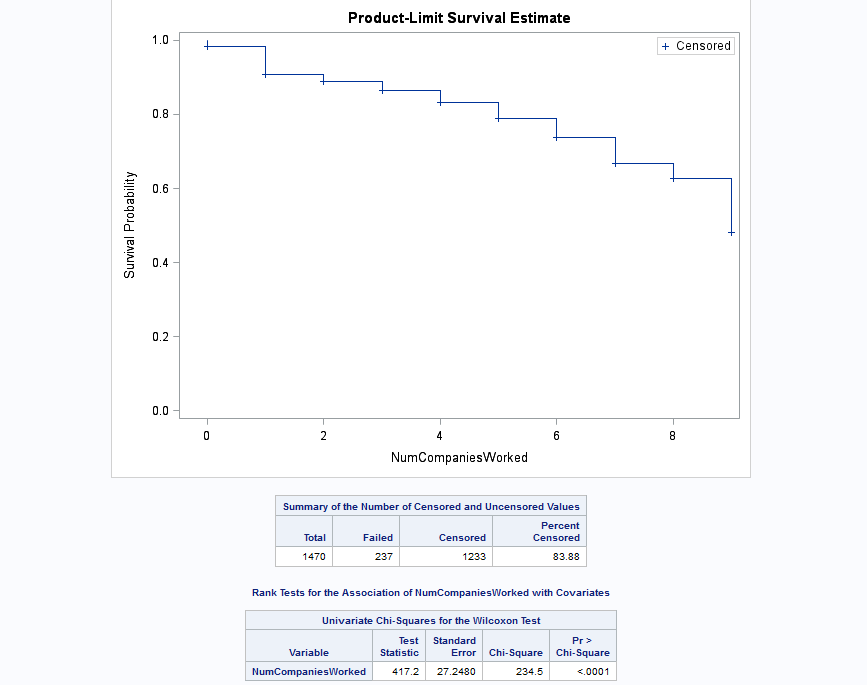
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable affected the attrition rate very little of the company.



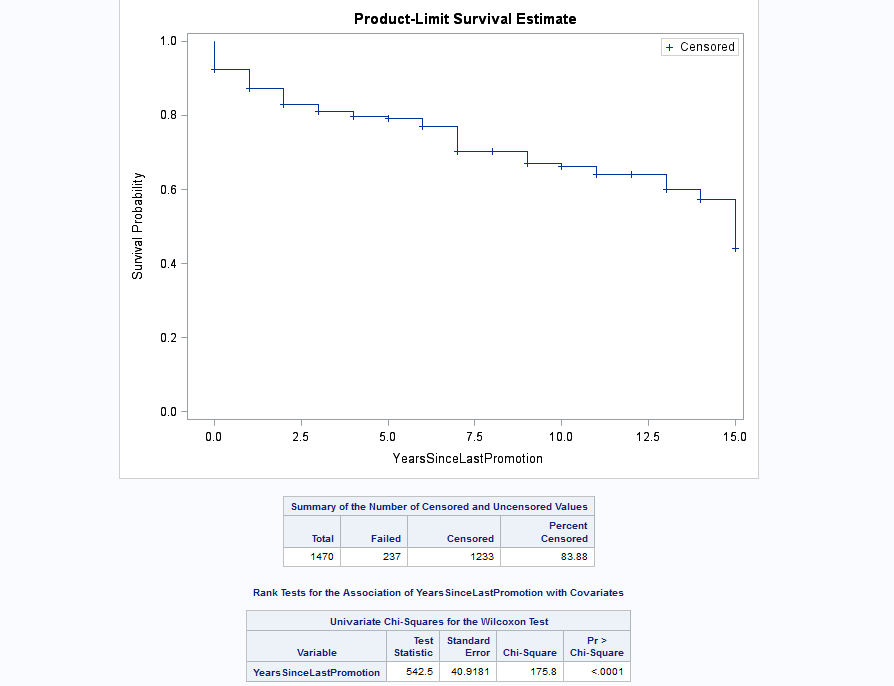
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable affected greatly the attrition rate of the company due to the survival probability reaching 0.4.



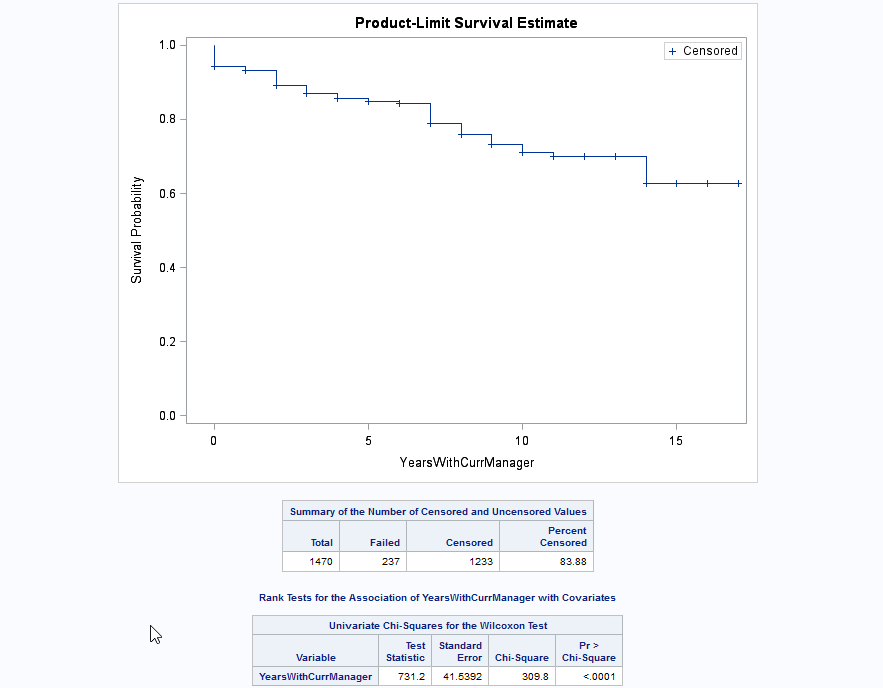
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable is affecting moderately the attrition rate of the company.



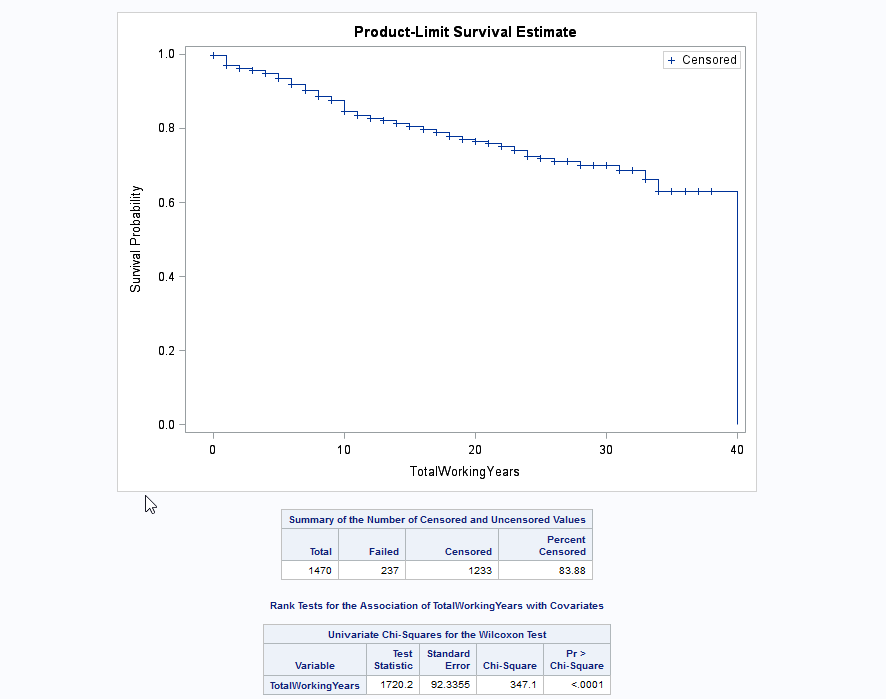
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also people are leaving the company just after they have been promoted since the survival probability went down at the 0 years since last promotion. The range of survival probability suggests that this variable is affecting moderately high the attrition rate of the company.



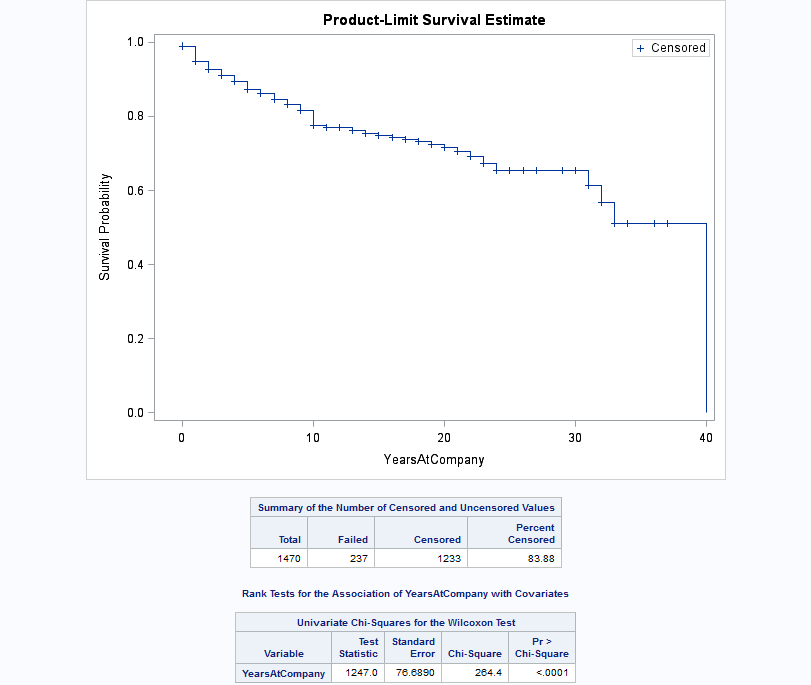
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also we have seen that more people are leaving the company with years under the same manager between 0 and 14. We could also see that there is a constant survival probability from 11 to 14 and above it too. Although it is hard to say what might have been the reason for this constant attrition rate. This suggests that this variable is affecting moderately the attrition rate of the company.



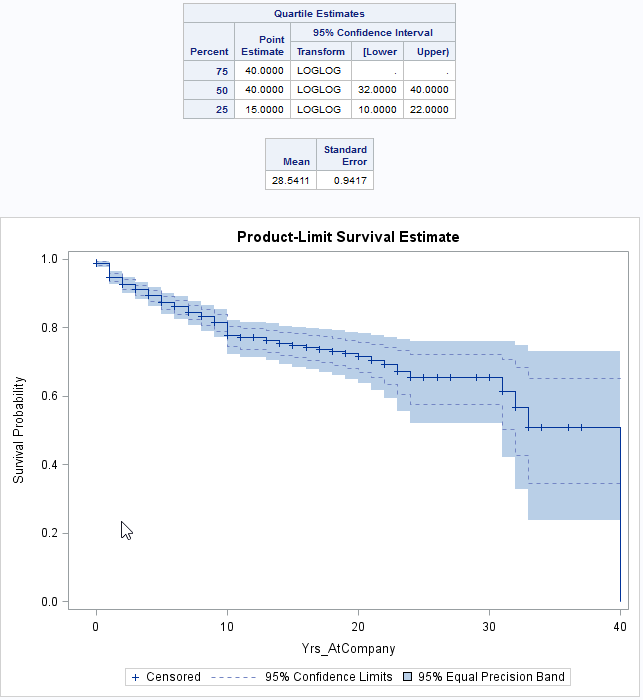
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable is affecting moderately high the attrition rate of the company. We could see that the attrition rate is constant from 34 to 40 too.

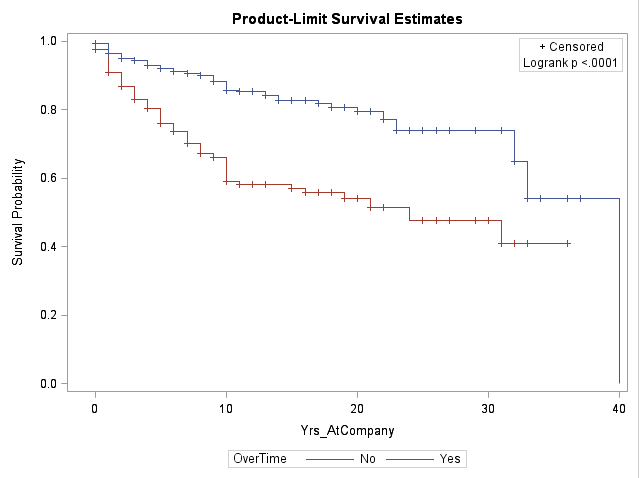
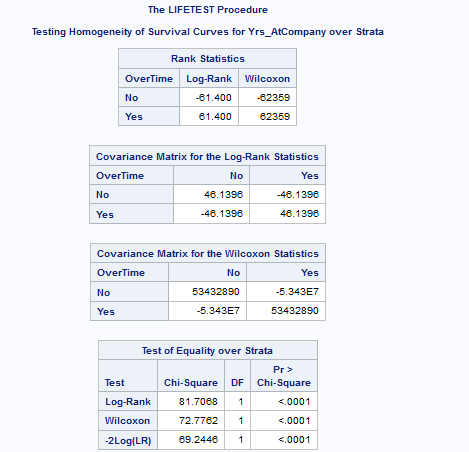


This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

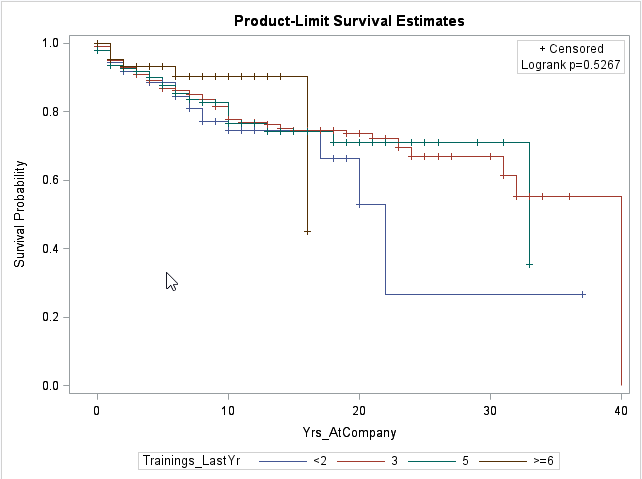
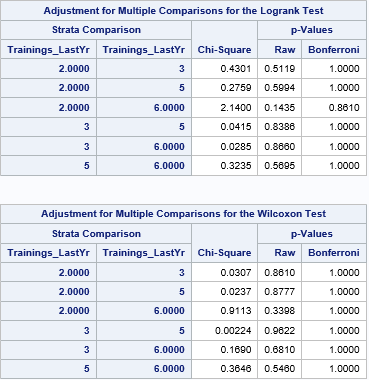
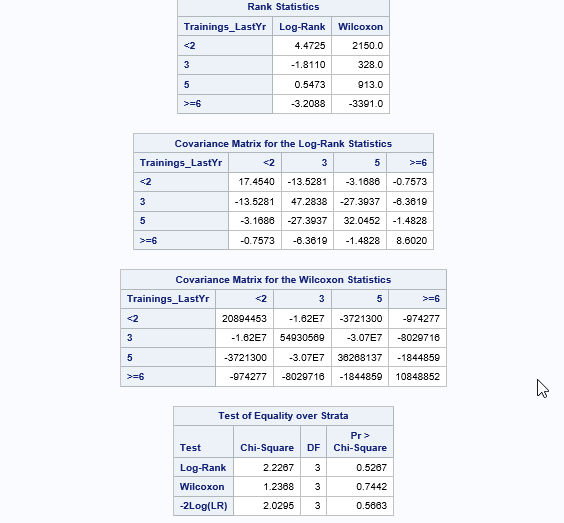
From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable is affecting moderately high the attrition rate of the company. We could see that the attrition rate is constant from 23 to 30 and 33 to 40 too.



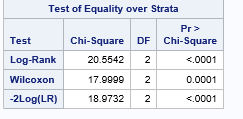
This visualization uses the procedure Life test to understand the survival probability of the attrition of the employees in the company. It also provides us with confidence intervals

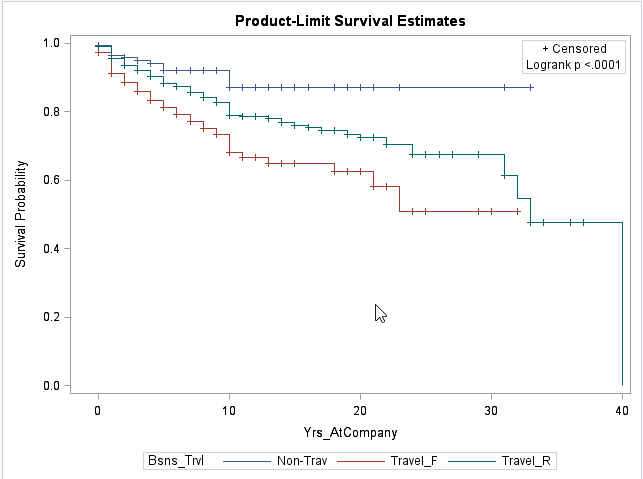


Here we tried to observe the significance of overtime variable on the survival probability of the Attrition of the employees. We observed that the people or employees who were working overtime, their survival probability is drastically goin down than the employees who are not doing an over time. From the test of equality table it is evident that thse two graphs are not equal and that these variables are significant.

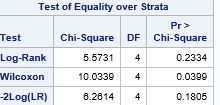


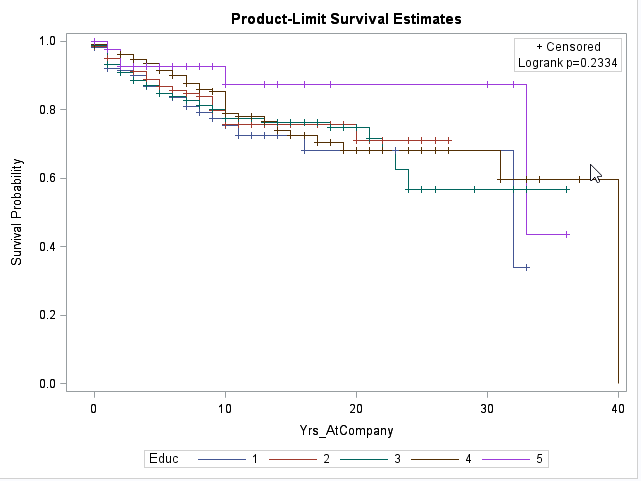
Here we tried to observe the significance of Trainings last year variable on the survival probability of the Attrition of the employees. We observed that the people or employees who had 6 trainings had a steady decline and then it flatened out. The rest of the level’s survival probability was running the same until 17 years at the company and from then people with less than 2 trainings quit the company and their survival probability went down dastically. From the test of equality table it is evident that thse two graphs are not equal and that these variables are insignificant due to the P values being more than 0.05.



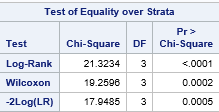


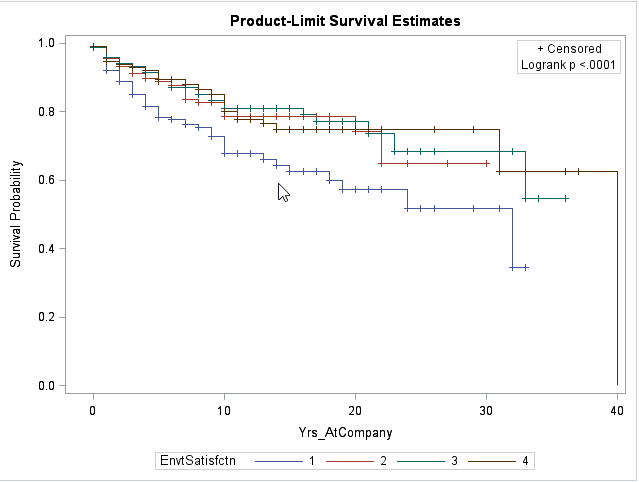
Here we tried to observe the significance of Business Travel variable on the survival probability of the Attrition of the employees. We observed that the people or employees who were travelling frequntly had the least surivival probability. Until 10 years in the company the surivival probability flattened out for the employees who were non travellers. The surivival probability of the travel rarely group stays between these two. From the test of equality table it is evident that thse two graphs are not equal and that these variables are significant due to the P values being less than 0.05.



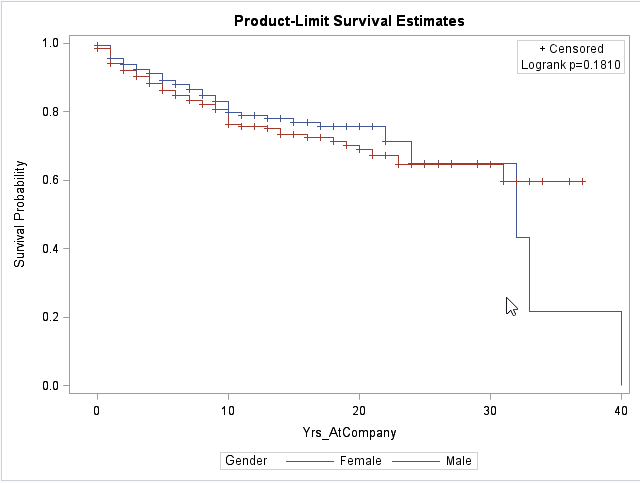
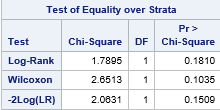


Here we tried to visualize the relation between different levels of education and years at the company by grouping Attrition. We observed that employees with very high education haven’t left the company and their survival probability flattened out after 10 years in the company. Through the graphs there seems to be no difference in the steady decline of survival probability of employees with education level less than 5. From the test of equality table, it is evident that the graphs are not equal. But this variable seems to be insignificant due to its P value greater than 0.05.

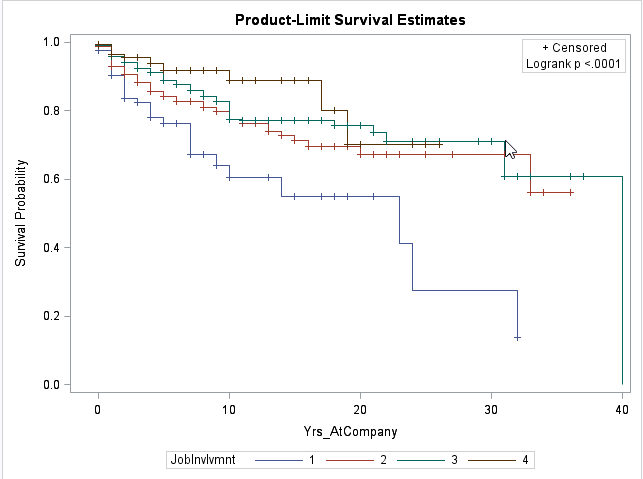
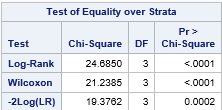




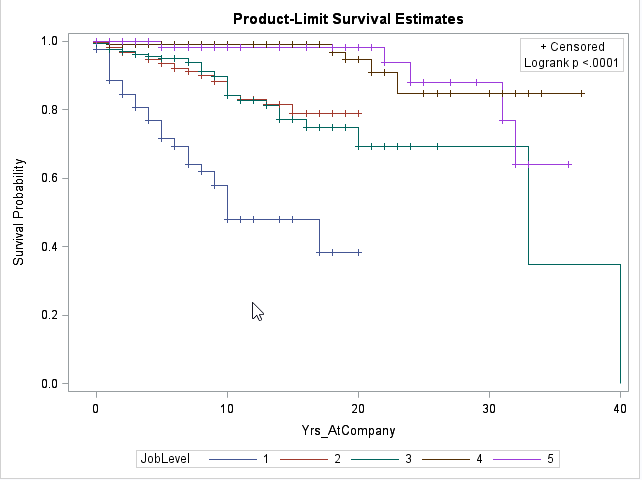
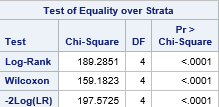
Here we tried to visualize the relation between different levels of environmental satisfaction and years at the company by grouping Attrition. We observed that people with very less environmental satisfaction left the company. Rest of the levels in the environmental satisfaction seem to go together. From the test of equality table, it is evident that the graphs are not equal. This variable seems to be significant due to its P value less than 0.05.



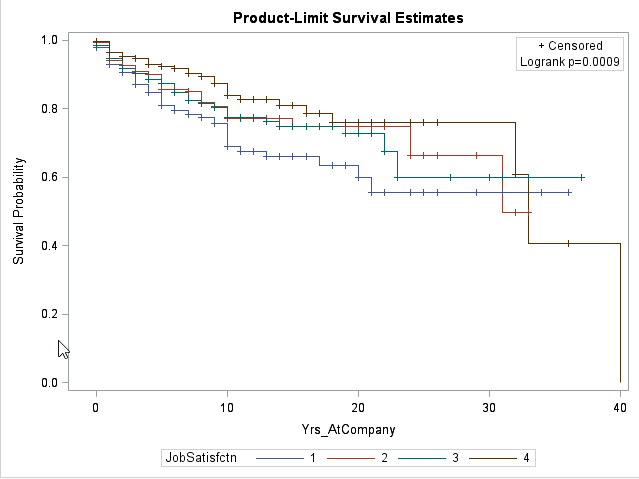
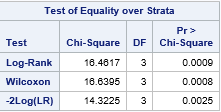
Here we tried to visualize the relation between different Genders and years at the company by grouping Attrition. Here we observed that survival probability was affected a little due to the gender and the test of equality table shows that we fail to say that they are not equal. We observed that there was a drastic decline in the survival probability of the Attrition after the female employees reached 30 years of age. The P values suggest that this variable is not significant for our model.



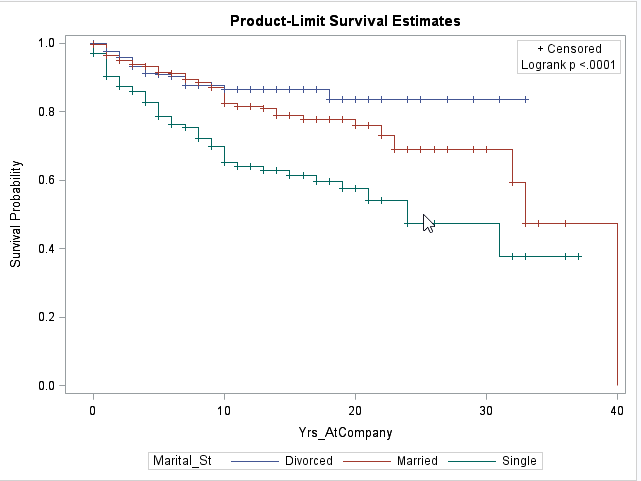
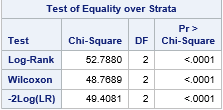
Here we tried to visualize the relation between different Job involvement levels and years at the company by grouping Attrition. Here we observed that there was a drastic decline of survival probability for the employees with very less job involvement in the company. The P value suggest that this variable is significant for our model.



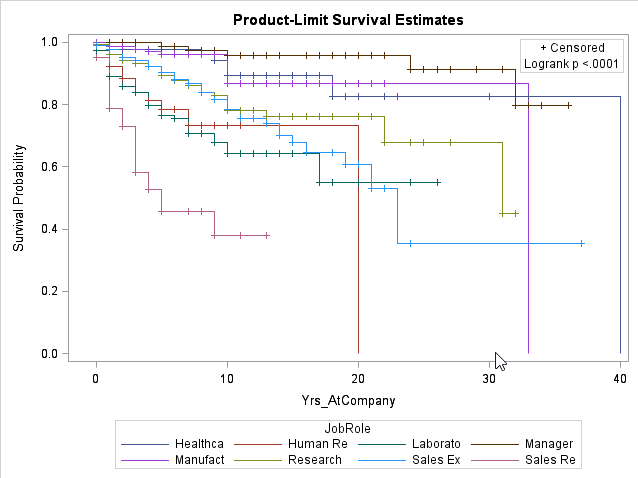
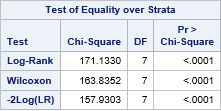
Here we tried to visualize the relation between different Job levels and years at the company by grouping Attrition. Here we observed that there was a drastic decline in the survival probability of the employees with very low job level.



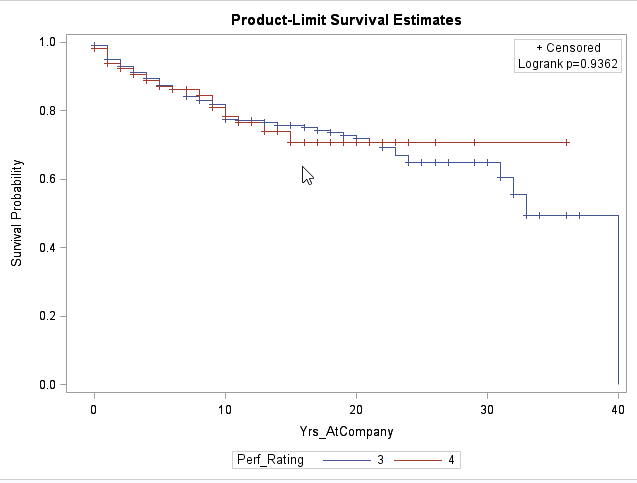
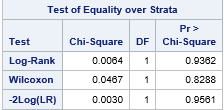
Here we observed that employees with job satisfaction equal to 1 have lower survival probability than the employees with higher job satisfaction level. The P value suggest that this variable is significant for our model.



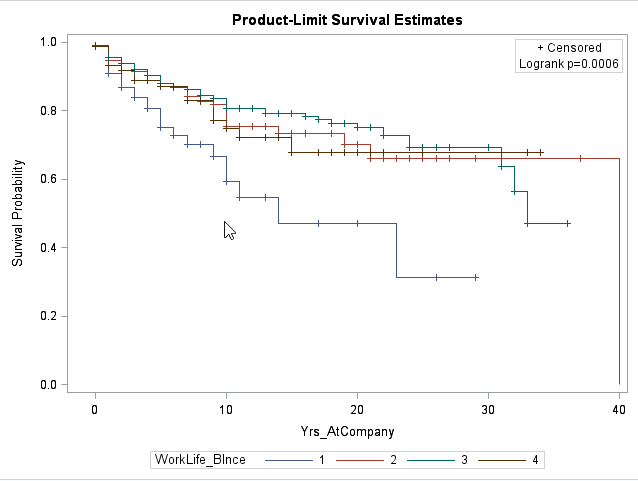
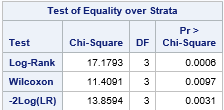
Here we observed that the employees who are single have less survival probability than the employees who are married or divorced. The P value suggests that the variable is significant for our model prediction.



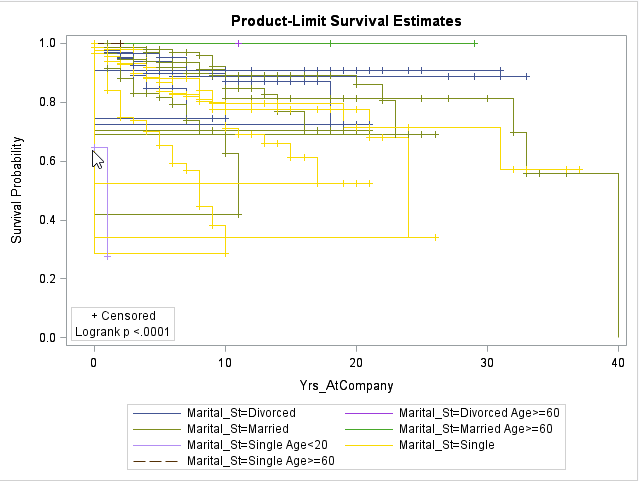
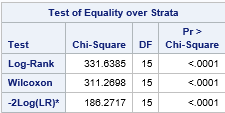
Here we observed that Sales Representative employees have very minimum survival probability than the rest of the job roles. The P value suggests that the variable is significant for our model prediction.



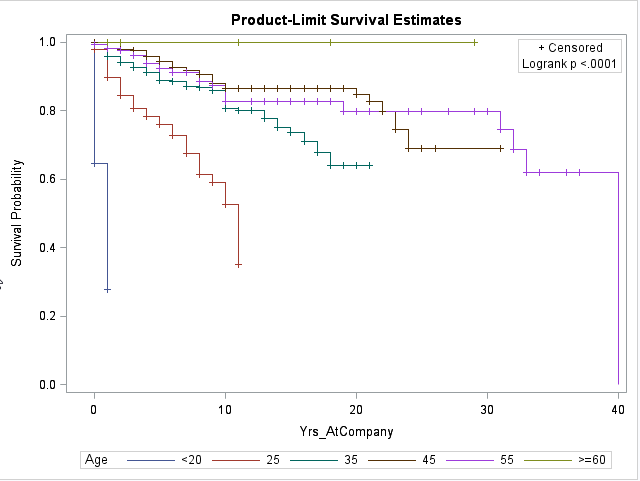
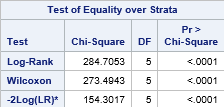
Here we observed that Performance rating of 3 had some drastic decline in the survival probability of the employees leaving the company. The P value suggests that the variable is insignificant for our prediction model.



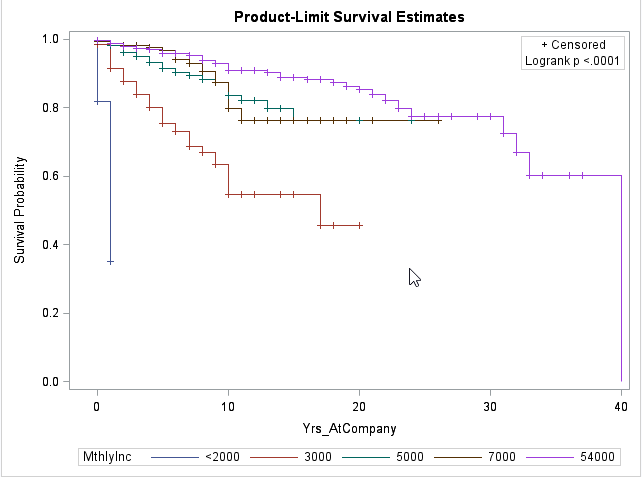
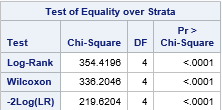
Here we observed that the Employees with a work life balance of 1 have a very minimum survival probability than higher work life balance. The P value suggests that the variable is significant for our model prediction.



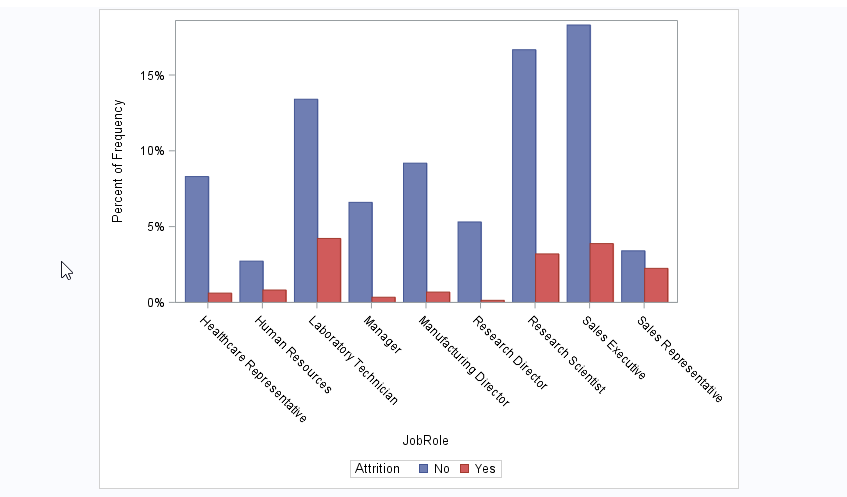
Here we observed that employees who are single and less than 20 years of age have comparatively the lowest survival probability suggesting a big concern for the management.



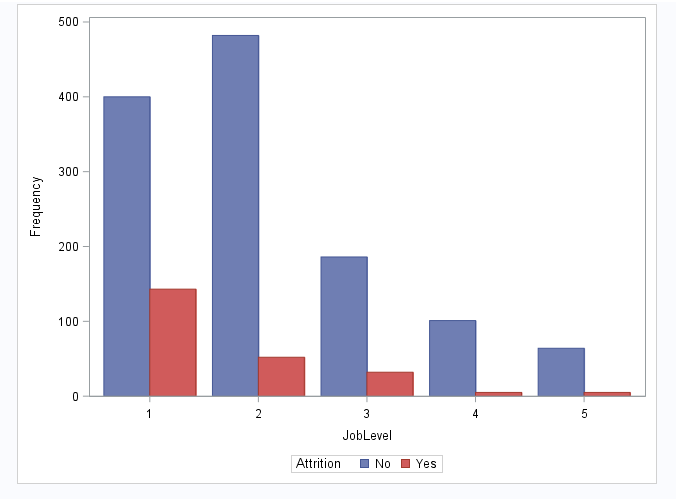
Here we observed that employees below the age of 20 have very minimum survival probability in the company.



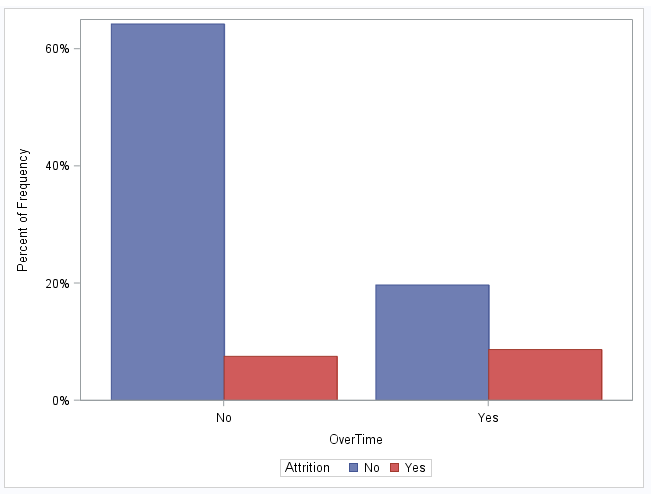
Here we observed that employees with a monthly income less than 2000 are leaving the company and their survival probability is alarmingly low. The P value suggests that the variable is significant for model prediction.



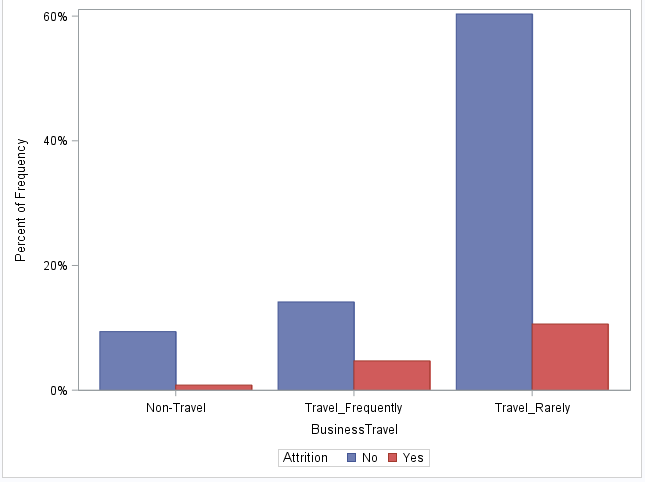
Here we observed that laboratory technicians, Sales Executives, Research Scientist and Sales Representatives left the company suggesting that these job roles have something that is making the employees unwelcoming. We suggest the Management team to take a look at these job roles and their characteristics.



Here we tried to visualize the relation between companies Attrition and the Job levels of the employees. We observed that lower job level employees are moving out of the company than the ones with higher job roles.

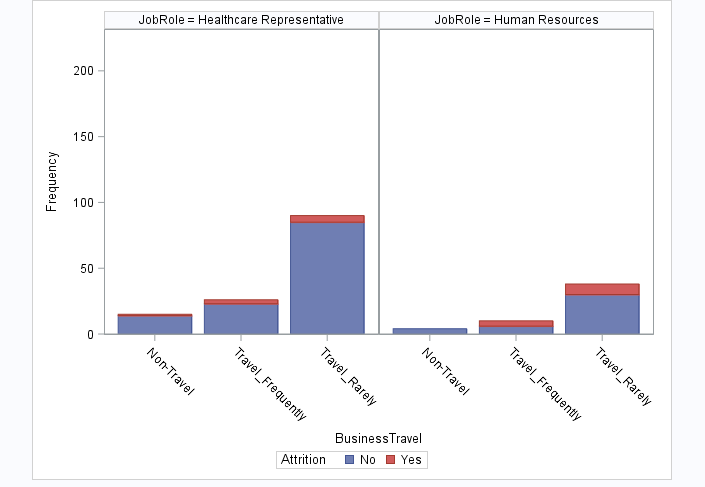


Here we tried to visualize the relation between the employees who worked over time and not worked over time with the Attrition of the employees. We observed that the people who were working overtime were having higher propensity to leave the company than the ones who are not working overtime. This is a wonderful insight to suggest the management that people are having issues to work overtime and should look into it.

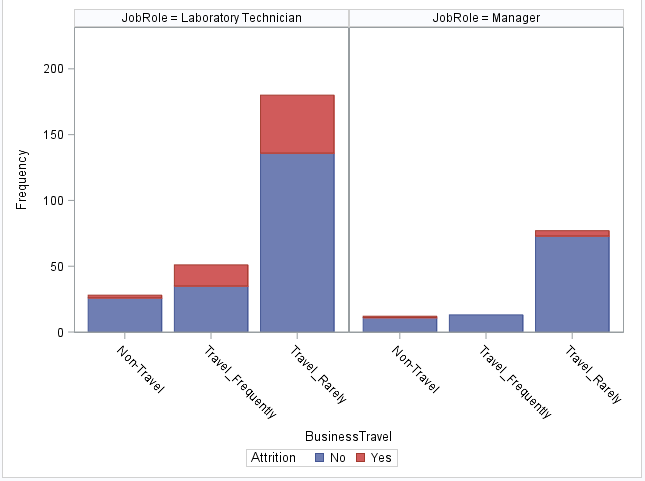


Here we tried to visualize the relation between Business travel of the employees and their Attrition. We observed that people who travel very seldom seem to be leaving the company. This may be due to one of the reasons that the employees who left may be proponents of travel than the rest. The next are the employees who travel frequently. They tend to have lesser frequency of leaving the company. Knowing better about their state of mind and opinion about the travel can give better understanding. A simple survey would be better. There is also a possibility that the people who don’t want to travel are in the job roles that requires travel and the employees who love to travel are the job roles that need travelling but have got no opportunity to travel.

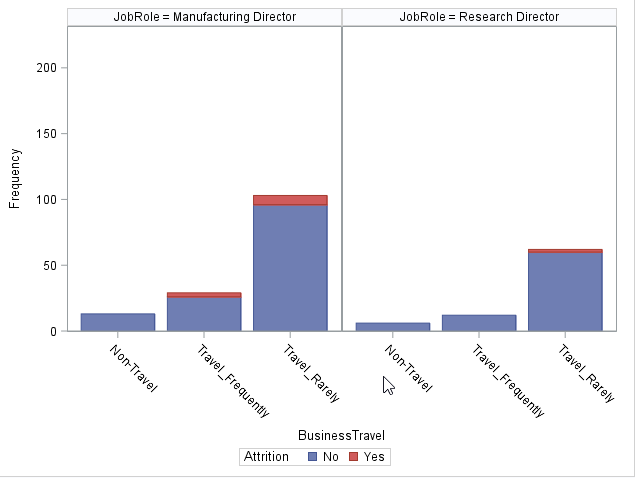
Here we did some in-depth Exploratory data analysis to see which job roles attrite more with the state of their business travel to give the management more insights about which jobs roles are of major concern in the company for the employees.



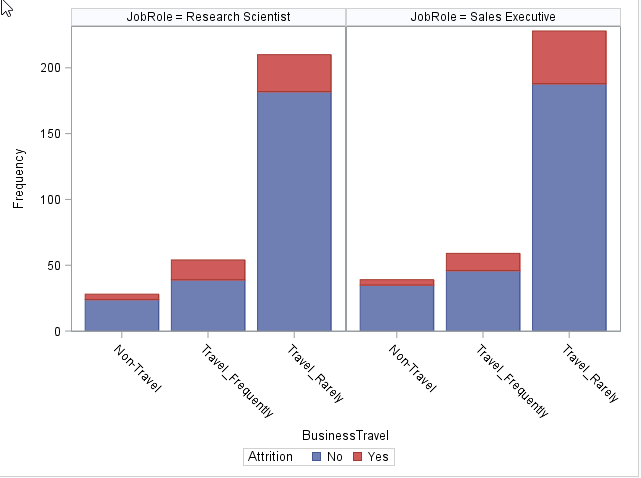
Here we tried to visualize the relation between Business travel and by grouping the Attrition and paneling the Job role to see how employees in each job role affects the Attrition of the company. We can see that Human resources who travel rarely tend to leave the company more.



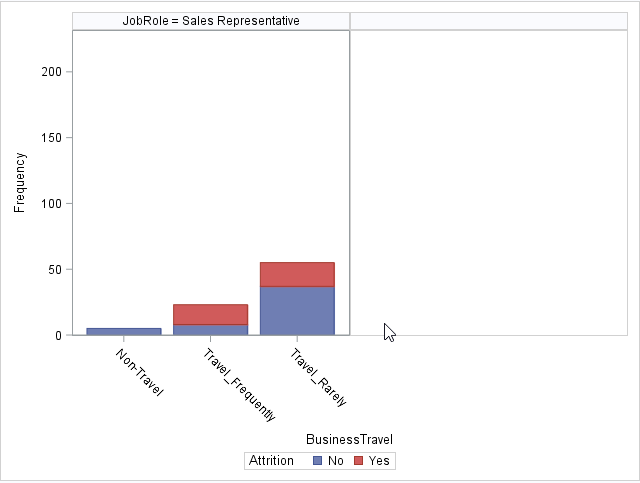
Laboratory Technicians who travel rarely seem to attrite more. The next comes the Travel frequently employees under the job role of Laboratory technician. This seems to be very minimum for Manager.



Manufacturing director and Research director job roles have very minimum attrition.

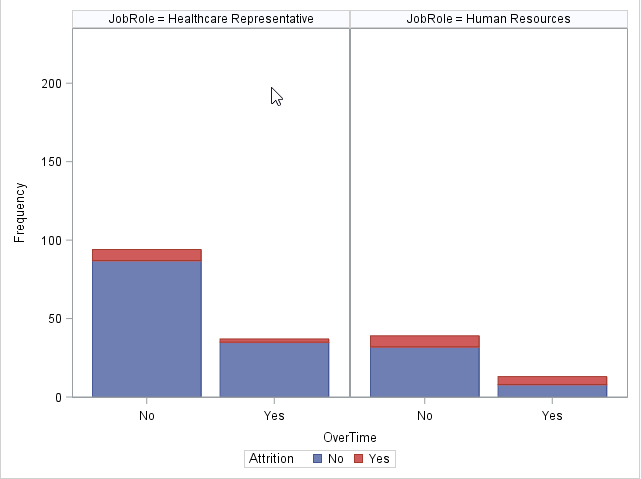


Research Scientists and Sales Executives who travel rarely attrite more in the company.

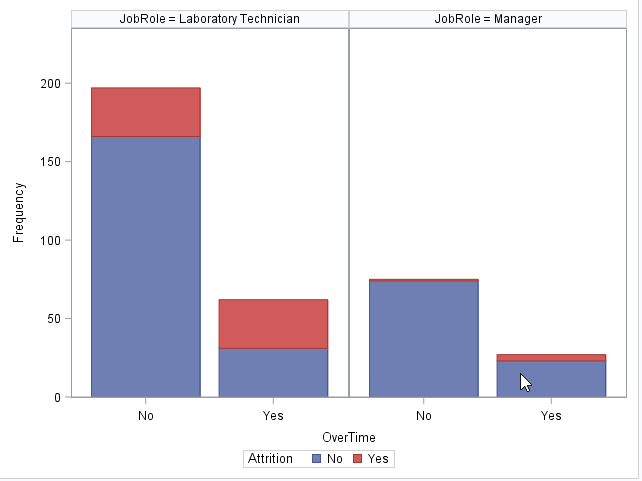


Sales Representatives who travel rarely and frequently tend to attrite more in the company.

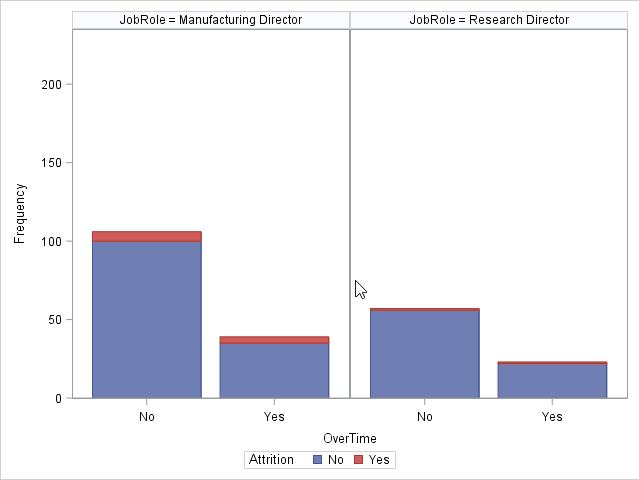
Here we did extensive Exploratory Data Analysis on how Attrition is for the employees in the job roles who work over time or not. This gives more insights about the job roles that have issues with working overtime.

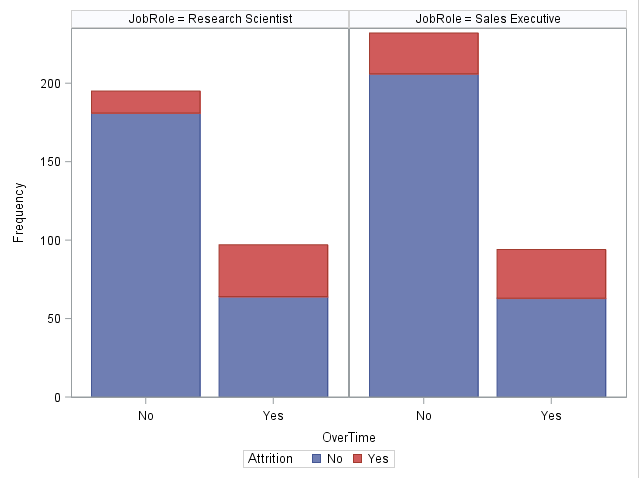


Healthcare Representatives and Human Resources employees who don’t work overtime tend to attrite more in the company.

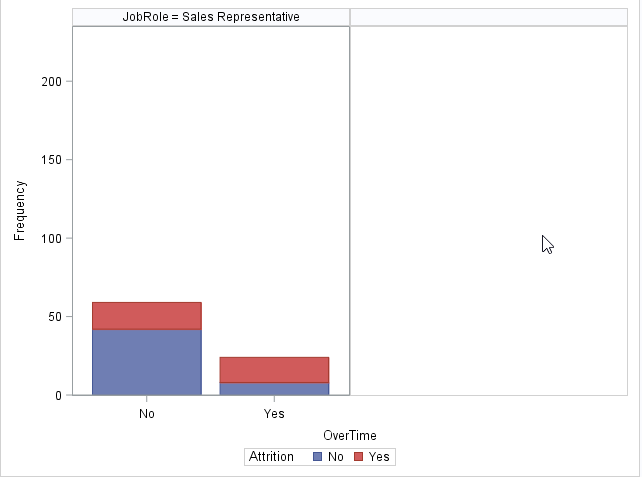


Laboratory Technicians who work over time and don’t work over time are attrition more in the company. We have seen a considerably high number of issues with this job role as many employees left the company from this role.

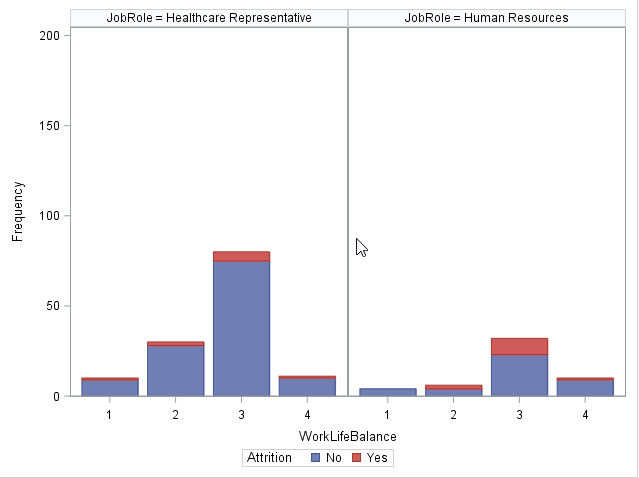


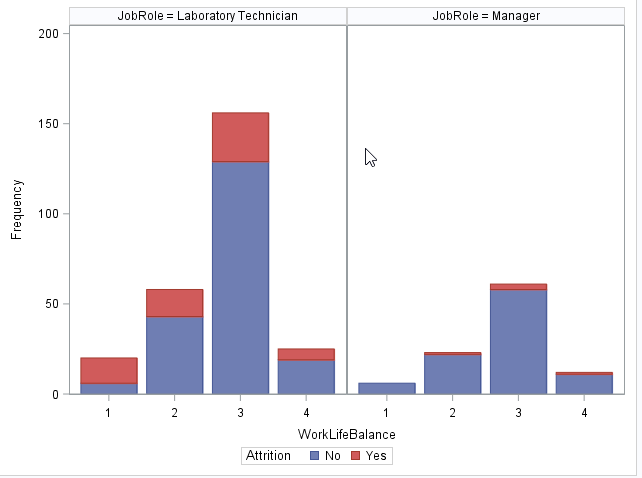


Here we can see that Research scientists and Sales Executives who worked over time left the company and this count seems to be high suggesting a higher concern level for the company.

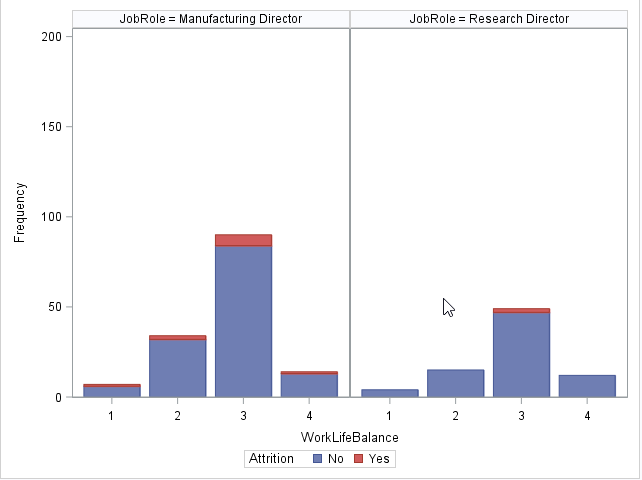


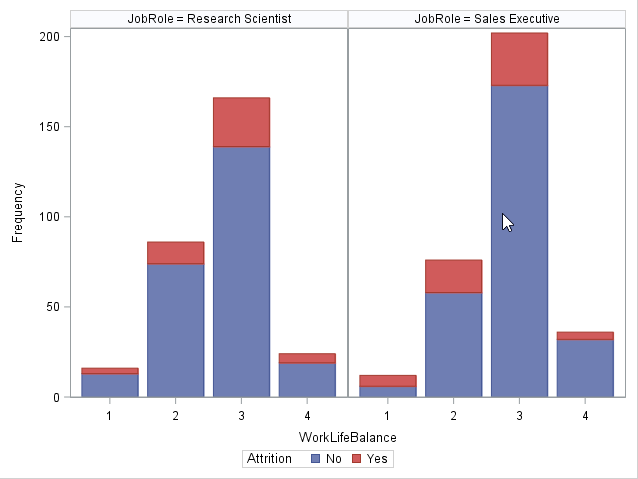
Sales Representatives who either work or do not work over time tend to leave the company. We suggest the management to look into this job role relating to overtime work.



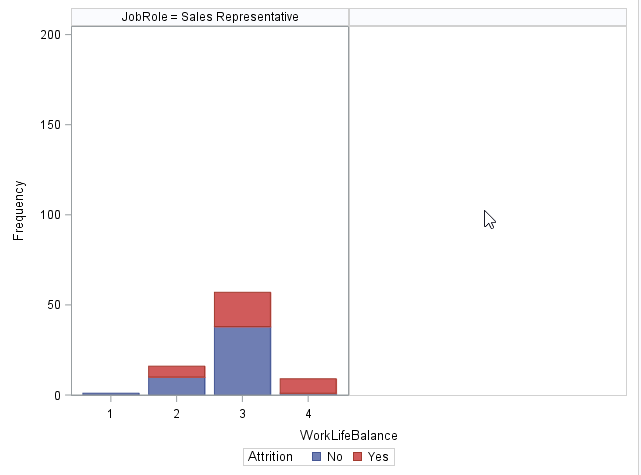


Here we observed that Laboratory Technicians with a work life balance of 3 left the company and the next goes to work life balance of 2 and 1. This variable has higher significance and we suggest the management to do further investigation on the work life balance of the employees under this job role.



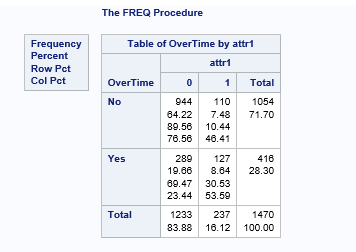
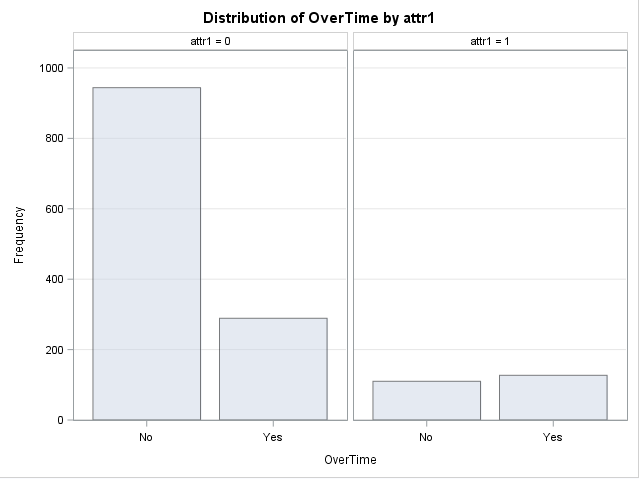


Research scientists and Sales executives with a work life balance of 3 are moving out of the company and this amount of people leaving seems to be high.

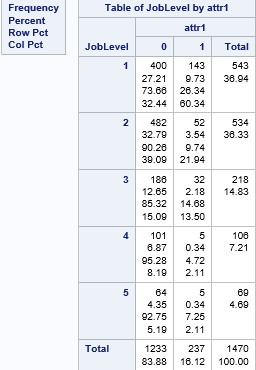
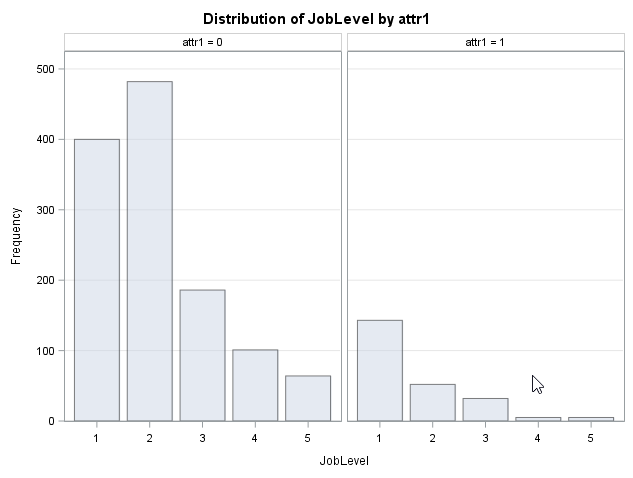


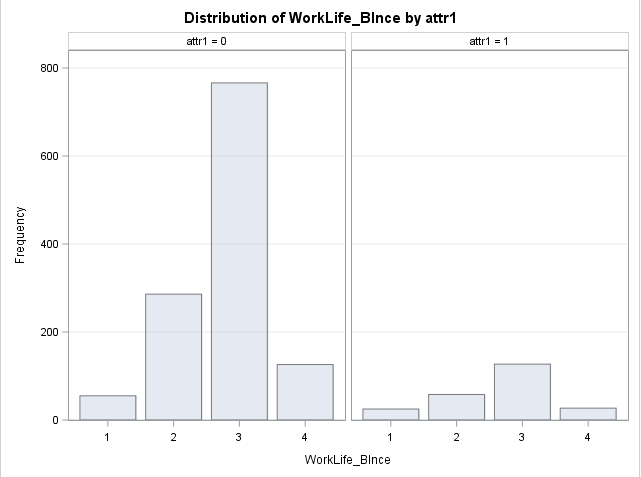
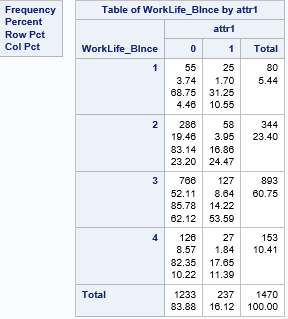
Sales Representatives with work life balance of 3 left the company and this count seems to be high. The next goes for the people with work life balance of 4. Although a further investigation is suggesting to the management from us since having a work life balance of 4 means they have the best time in the company and wouldn’t me much inclined to leave the company.

Here we tried to visualize the relation between overtime and Attrition of the company using Frequency procedure. Here we can actually see the exact count of the employees who quit the company and the ones for a particular reason like overtime, work life balance etc.

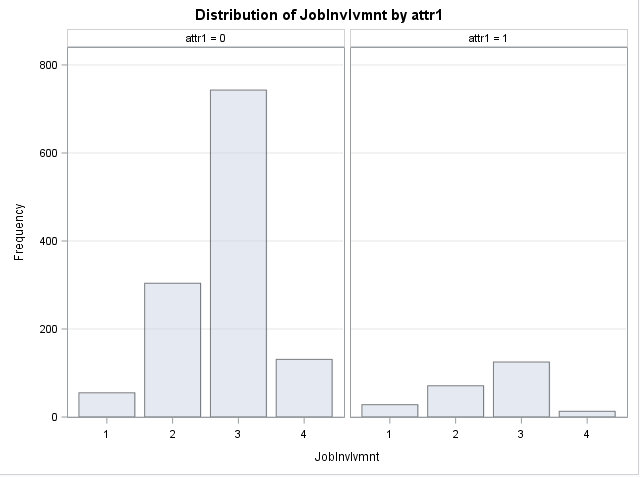
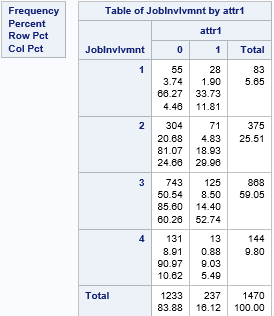


Here we observed that 30% of the employees left the company who did overtime.

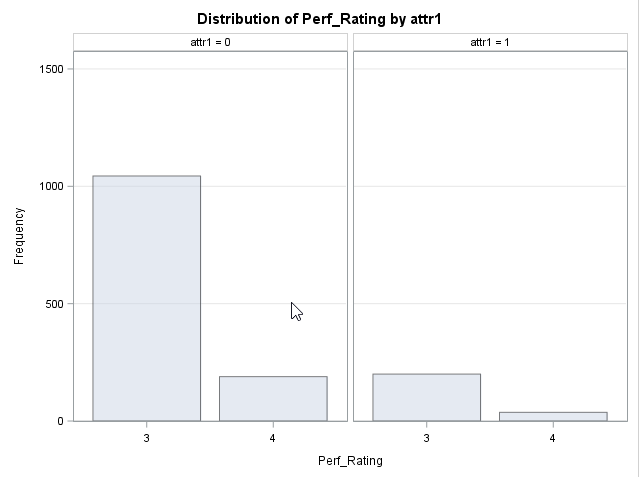
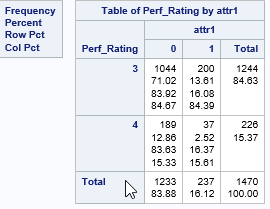
  
Here we observed that lower job level employees are leaving the company.

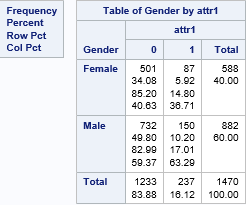
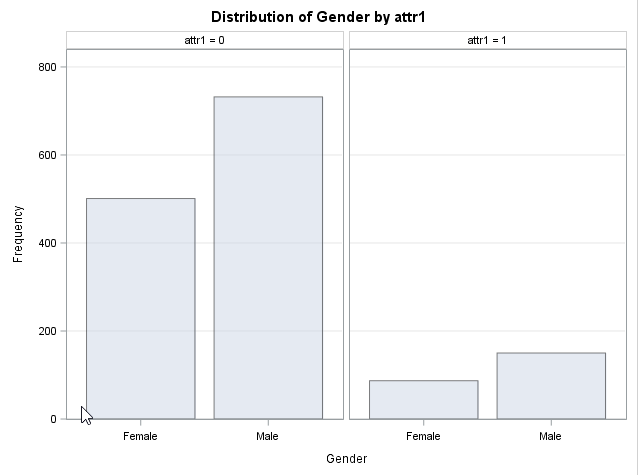
As stated before in our exploratory data analysis these graphs give us the information about the exact count of the employees leaving the company with different work life balance. This is for a quantitative understanding.

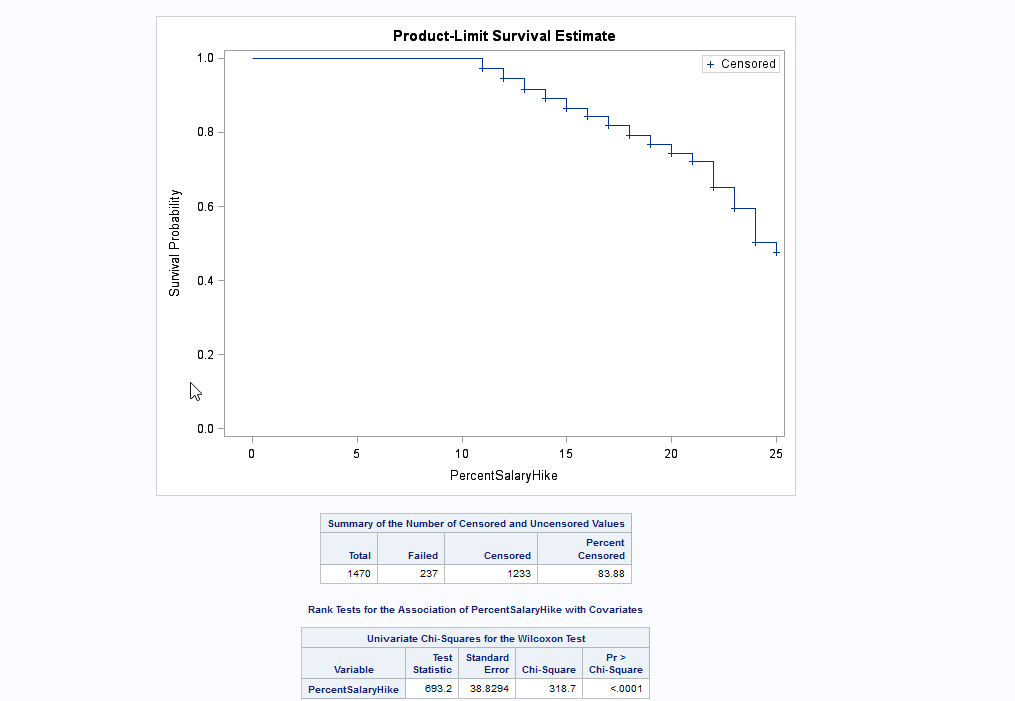
As stated before in our exploratory data analysis these graphs give us the information about the exact count of the employees leaving the company with different Job involvement levels. This is for a quantitative understanding.

As stated before in our exploratory data analysis these graphs give us the information about the exact count of the employees leaving the company with different Performance Rating. This is for a quantitative understanding.

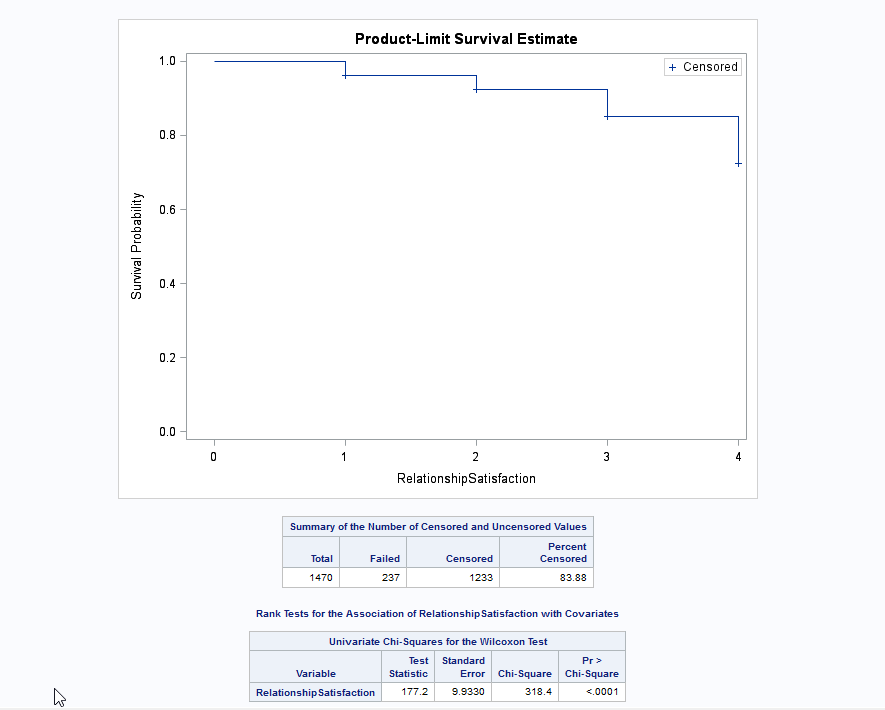


As stated before in our exploratory data analysis these graphs give us the information about the exact count of the employees leaving the company with different Genders. This is for a quantitative understanding.



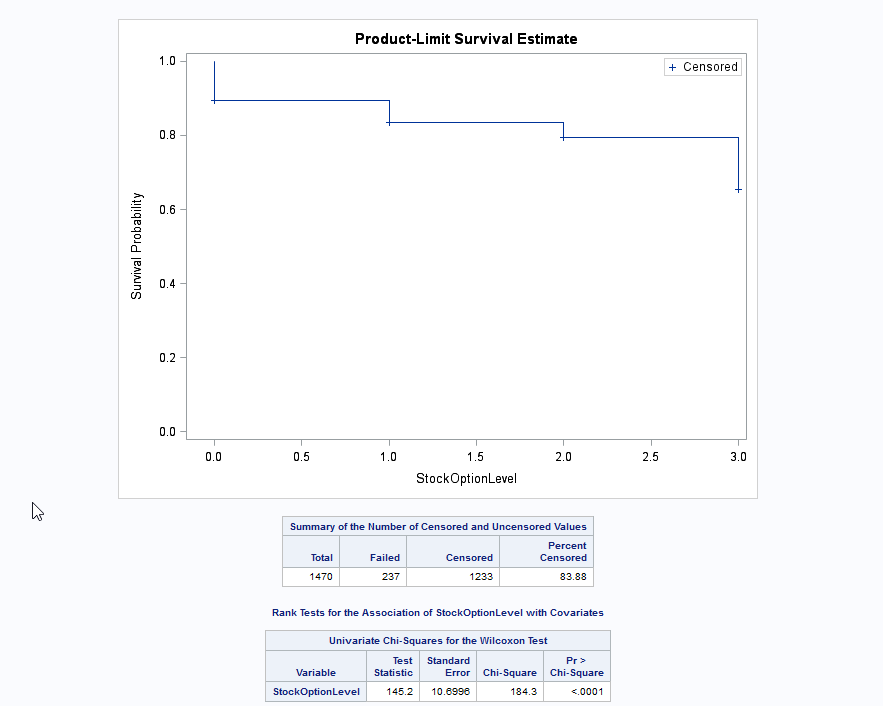
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and there is a sudden decrease in the survival probability at %21 salary hike and at %24 salary hike suggesting that these have greatly affected the attrition rate of the company.



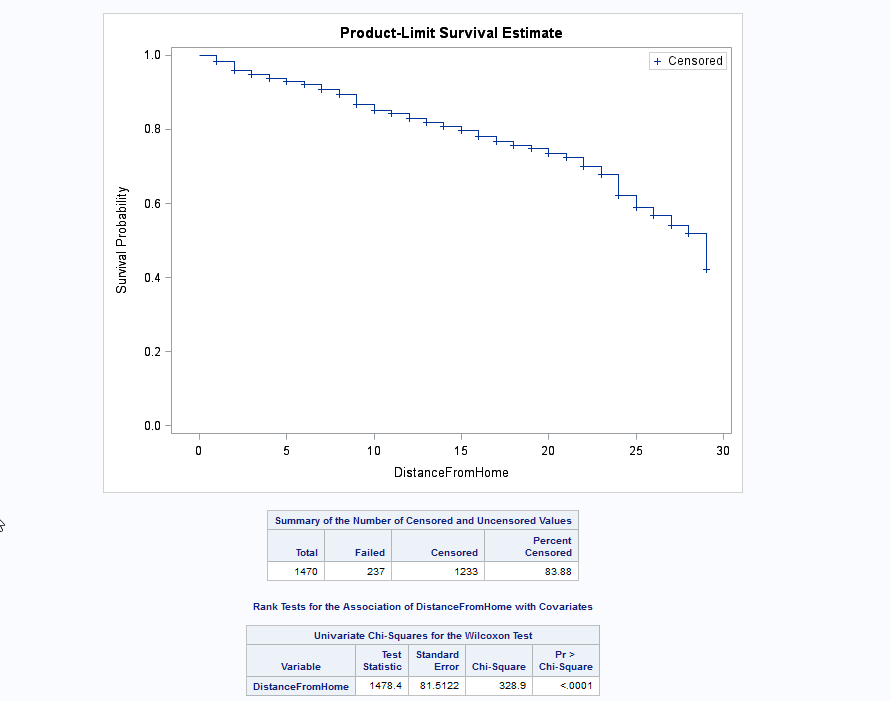
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable affected the attrition rate very little of the company.



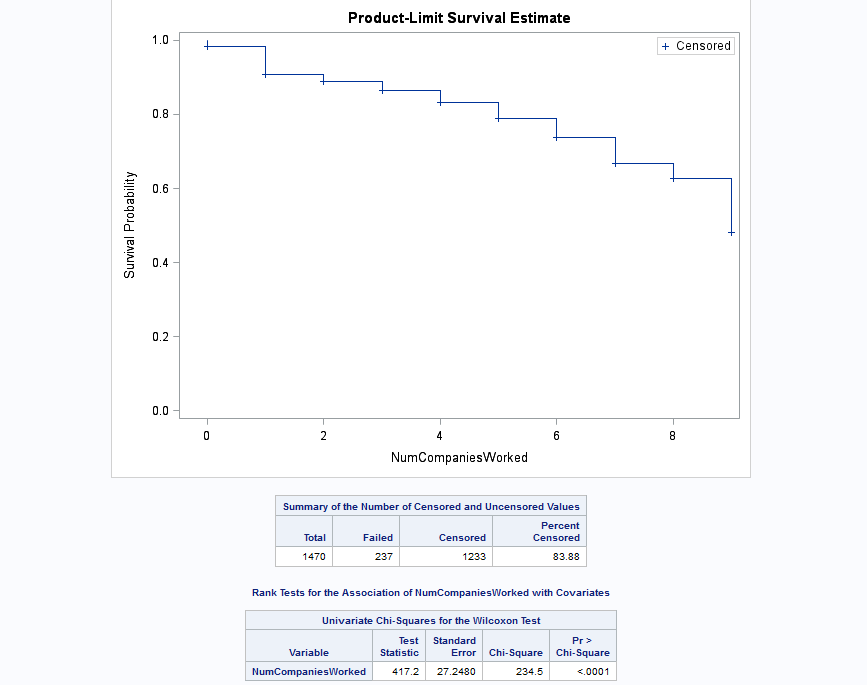
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable affected the attrition rate very little of the company.



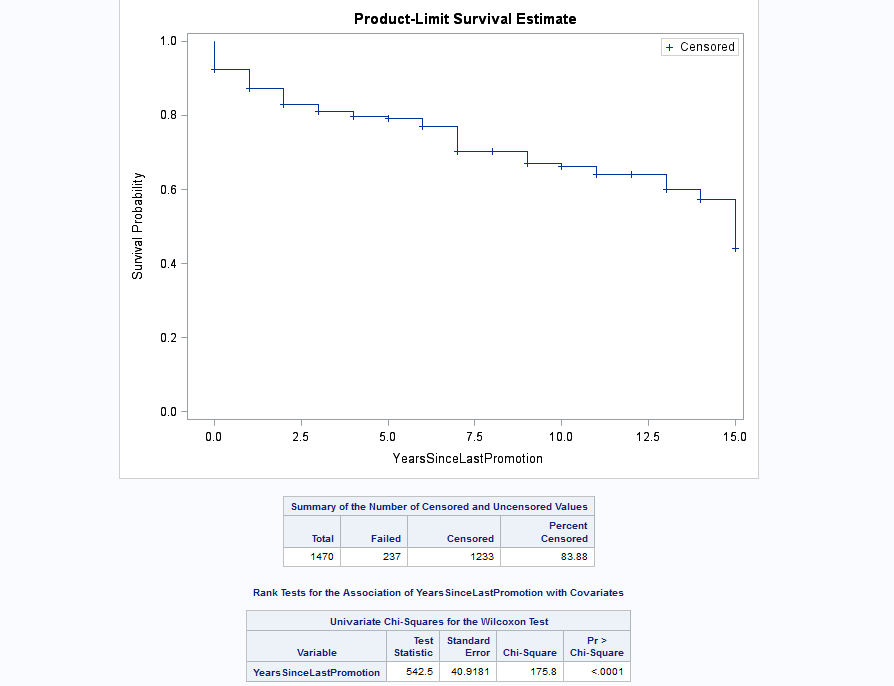
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable affected greatly the attrition rate of the company due to the survival probability reaching 0.4.



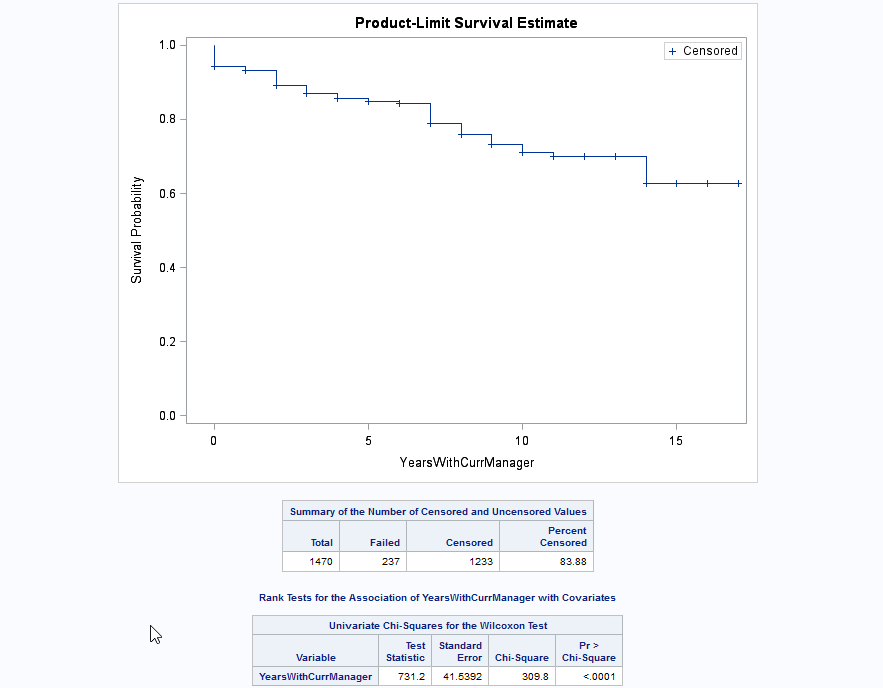
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable is affecting moderately the attrition rate of the company.



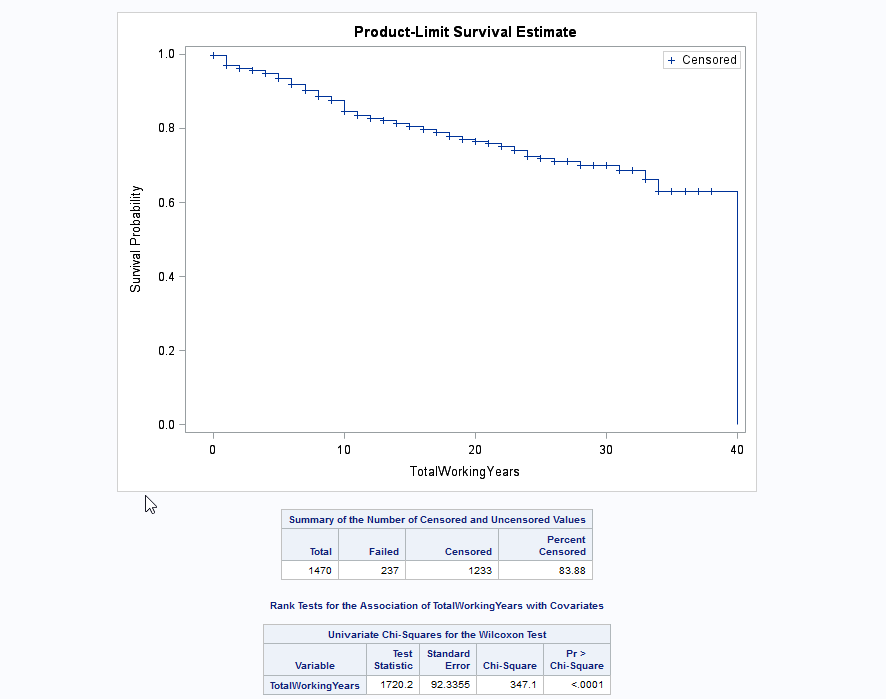
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also people are leaving the company just after they have been promoted since the survival probability went down at the 0 years since last promotion. The range of survival probability suggests that this variable is affecting moderately high the attrition rate of the company.



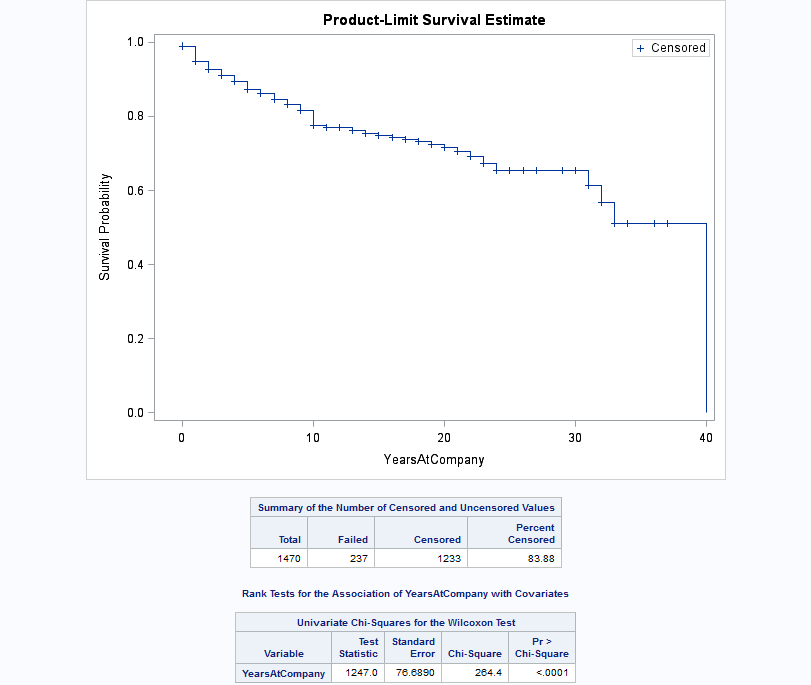
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also we have seen that more people are leaving the company with years under the same manager between 0 and 14. We could also see that there is a constant survival probability from 11 to 14 and above it too. Although it is hard to say what might have been the reason for this constant attrition rate. This suggests that this variable is affecting moderately the attrition rate of the company.



This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable is affecting moderately high the attrition rate of the company. We could see that the attrition rate is constant from 34 to 40 too.

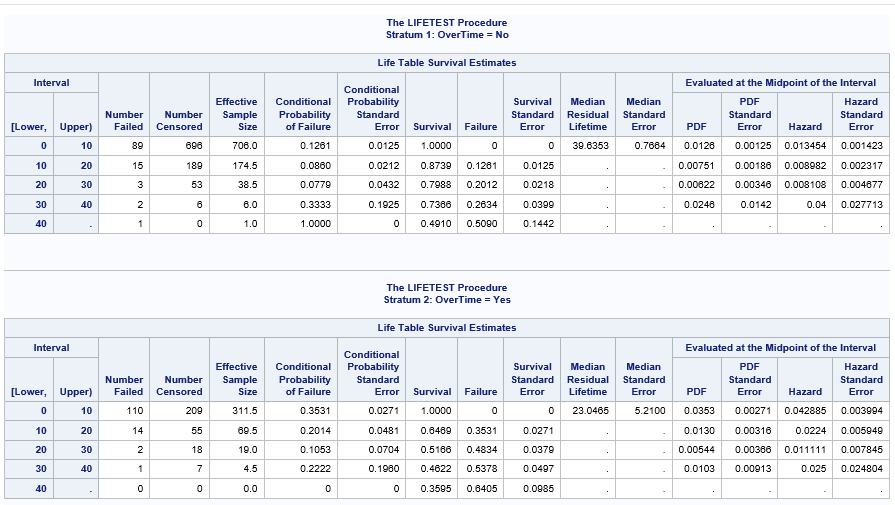


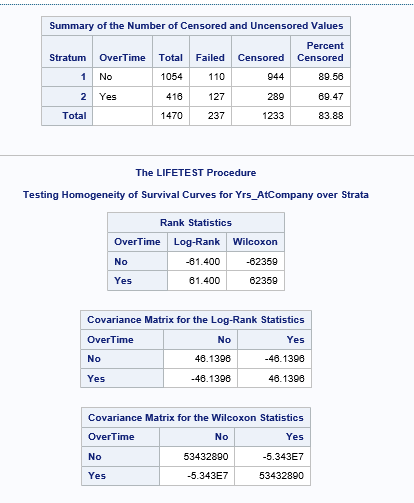
This variable is significant with a P value less than 0.05 with a value less than 0.0001 making it a significant value for our model prediction.

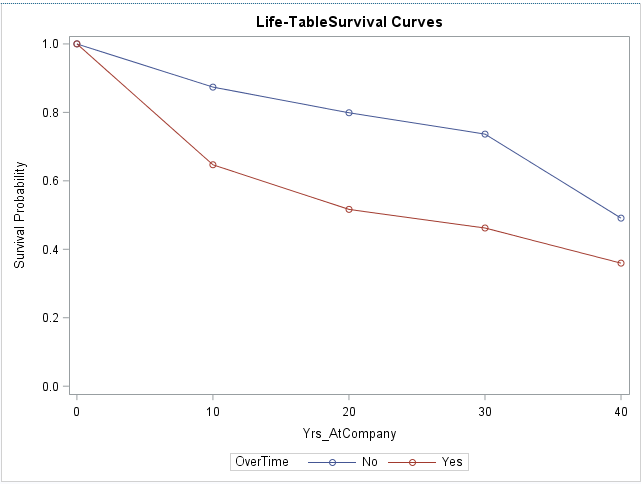
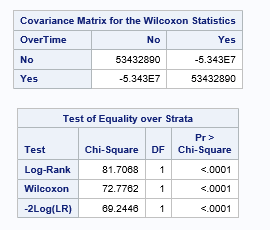
From the Life test graph or survival graph we can see that there is a steady decline in the survival probability and also suggesting that this variable is affecting moderately high the attrition rate of the company. We could see that the attrition rate is constant from 23 to 30 and 33 to 40 too.

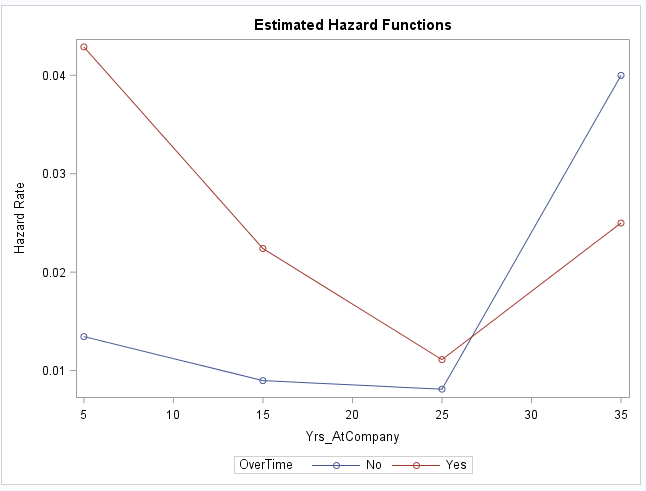
**LIFE TABEL SURVIVAL CURVES AND HAZARD FUNCTION**

1. Overtime



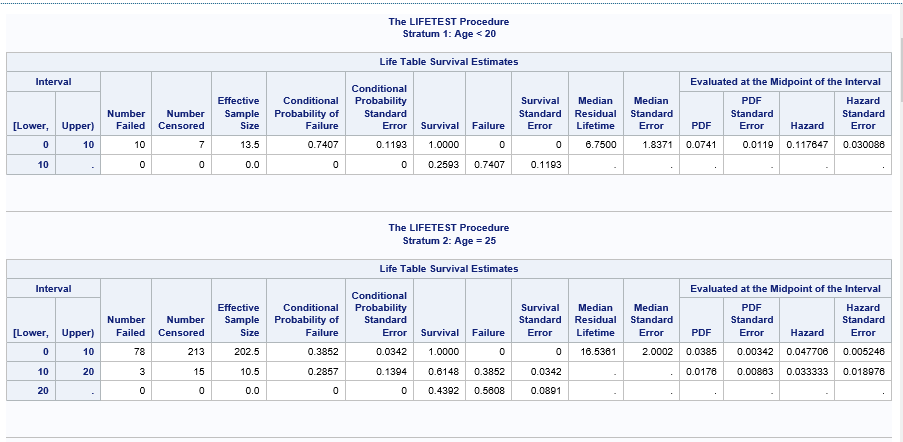


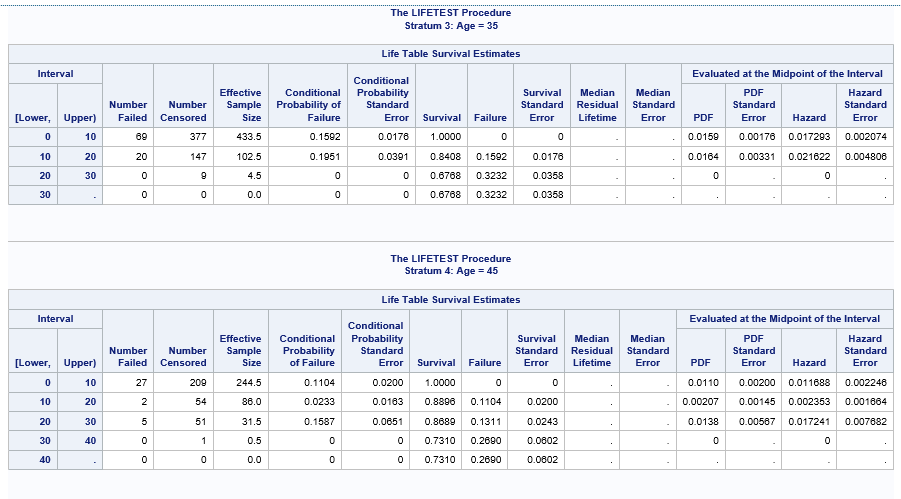


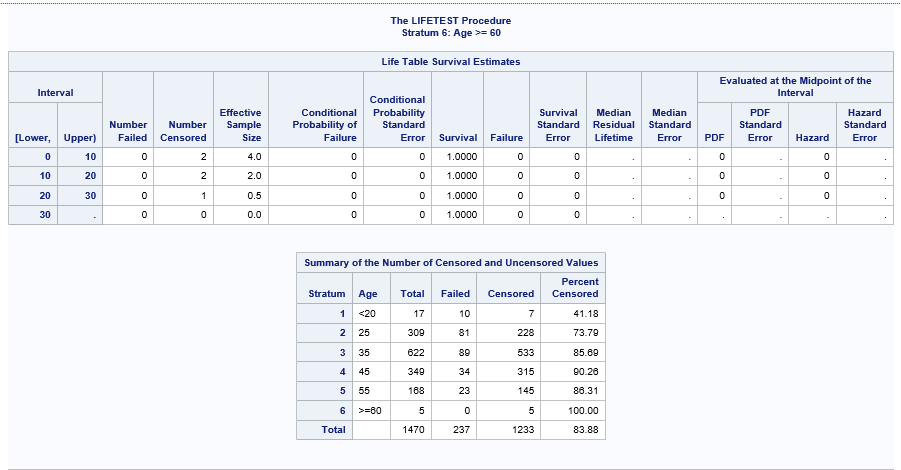


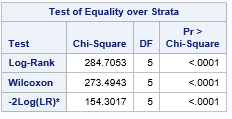
Employees who joined recently and do over time have high hazard rates compared to the employees who were with the company since long time (up to 25 years). Employees with more than 25 years of stay are hazardous irrespective of whether they do over time or not.

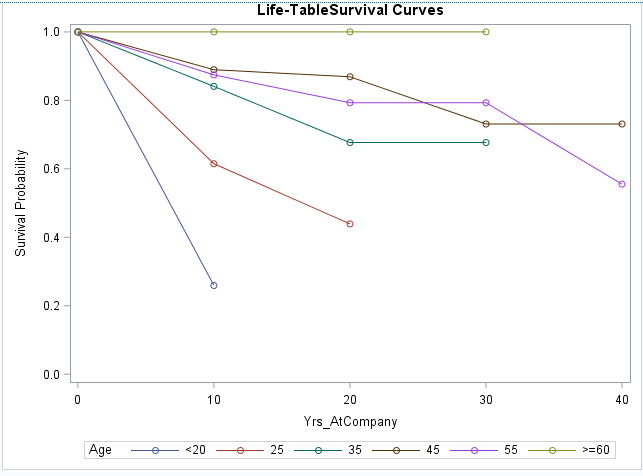
1. Age

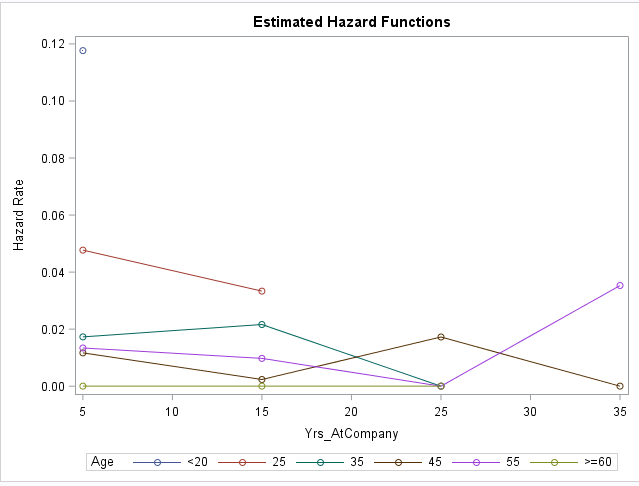






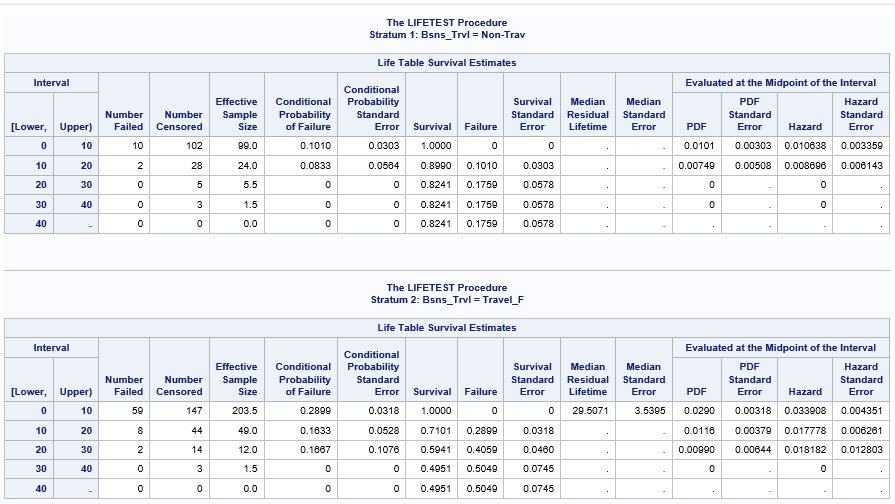


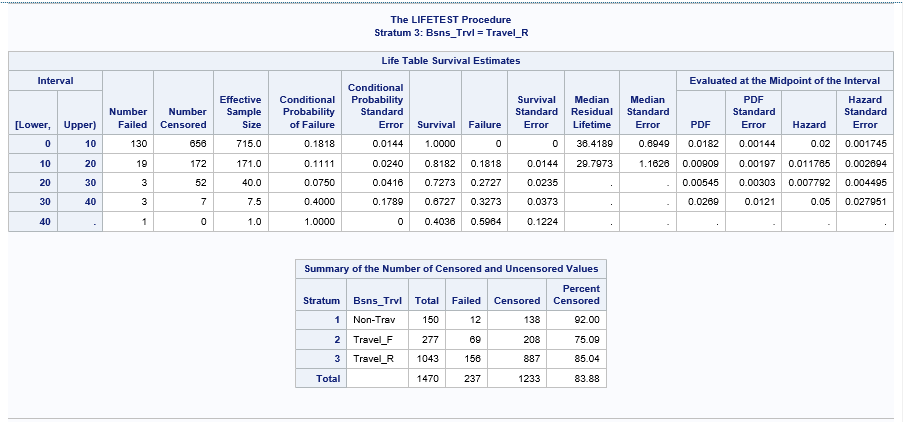


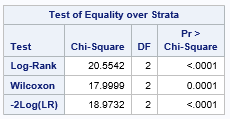


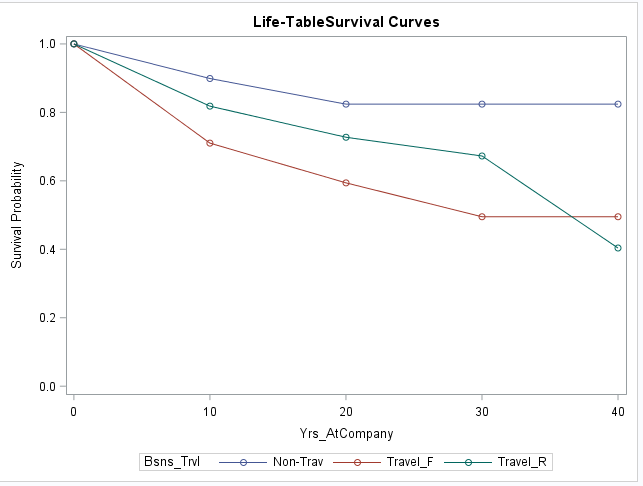
Employees under age of 25 and above 55 are having high hazard rates. Older employees might be retiring from job may be one genuine reason for their attrition.

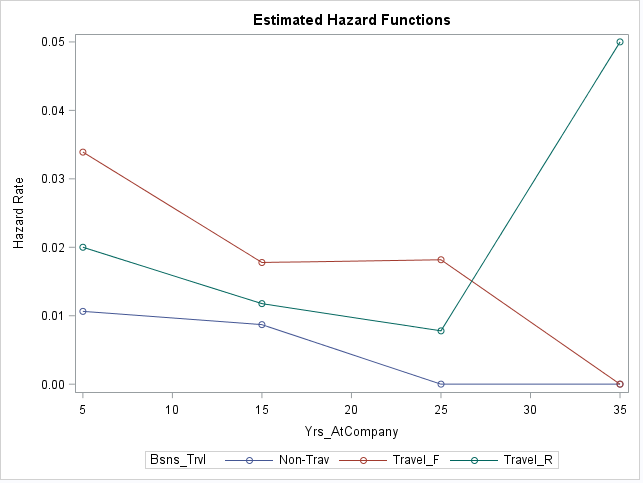
1. Business Travel





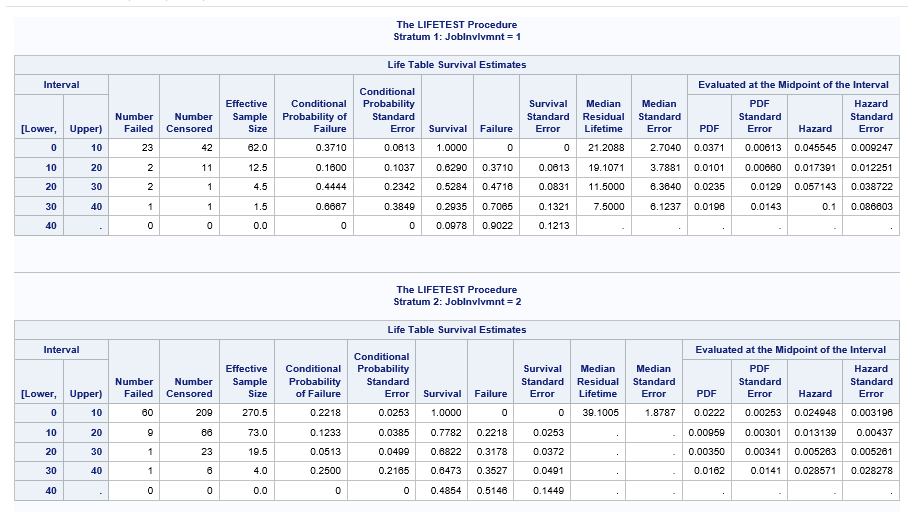


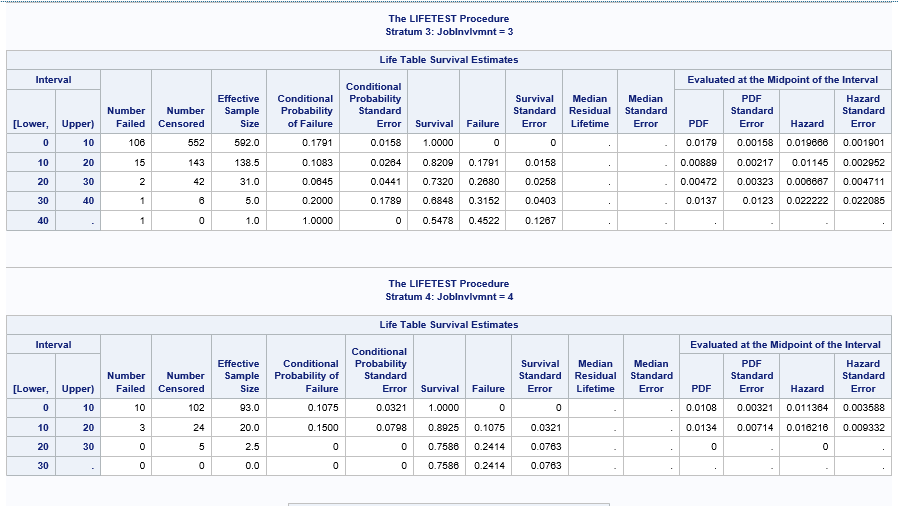


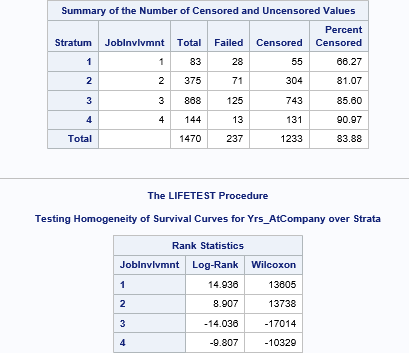


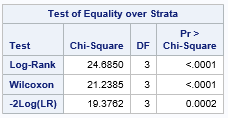
Employees who travel regularly have high hazard rate when compared to other categories of travel as shown above.

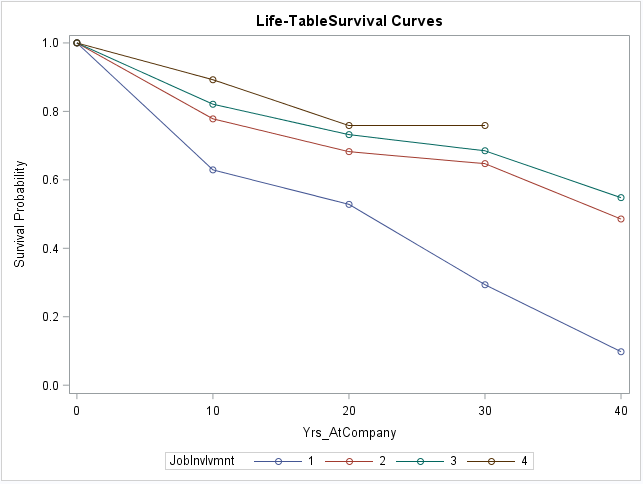
1. Job Involvement

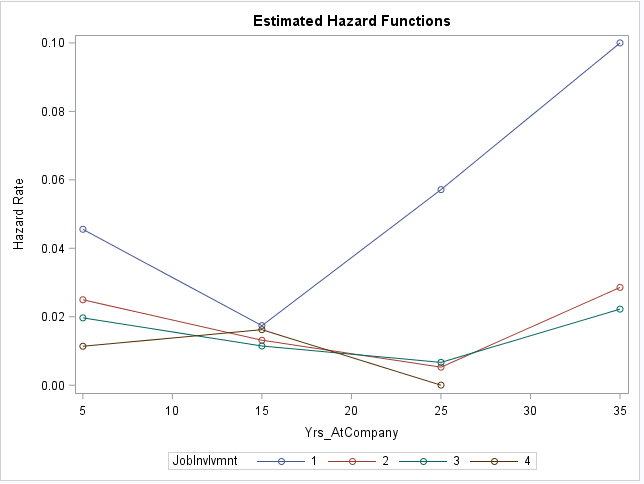












Employees with Job Involvement 1 are hazardous when compared to those employees who are highly involved in their job as shown above.

**Modeling Phase**

*Modeling using all covariates – (except Bonus variables)*

*Censored: Attrition*

*Time: Yrs at company*

(CLASS Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St

Age DistFrmHome Educ EnvtSatisfctn Perf\_Rating

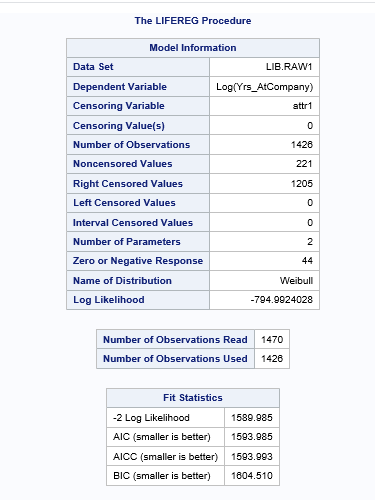
Rltnshp\_Satisfctn MthlyInc NumComp\_Worked Percent\_Hike Stock\_opt\_Level JobInvlvmnt

JobLevel JobSatisfctn WorkLife\_Blnce Tot\_Work\_Yrs Trainings\_LastYr Yrs\_InCurrentRole

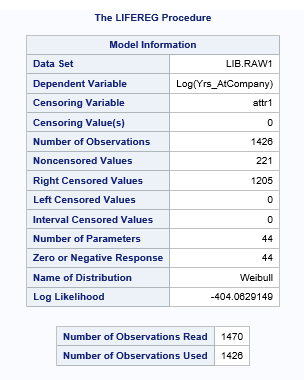
Yrs\_SinceLastPromotion DailyRate HrlyRate MthlyRate NumComp\_Worked Yrs\_WithCurrManager Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St)

1. Comparison of Null models vs Full model using PROC LIFEREG and WEIBULL distribution

Null Model



Full Model



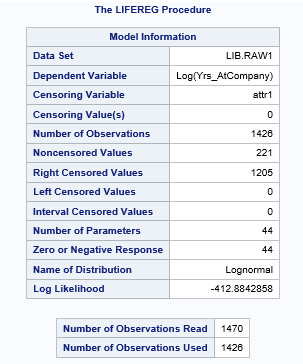
Comparison of Full model and Null Model



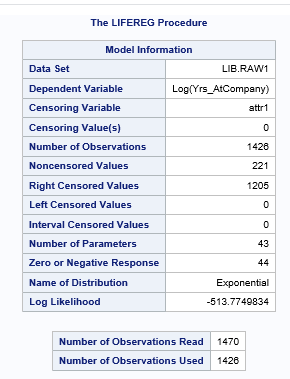
p\_value = 0 --> Full model is better than null model i.e at least one of the variables have co efficient different from 0

1. Goodness of fit curve for full model using Log likelihood ratio test

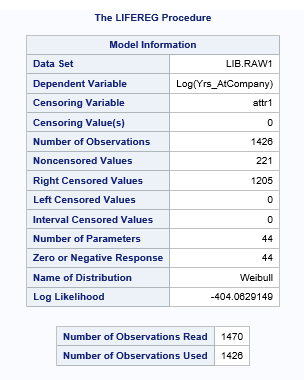
Lognormal



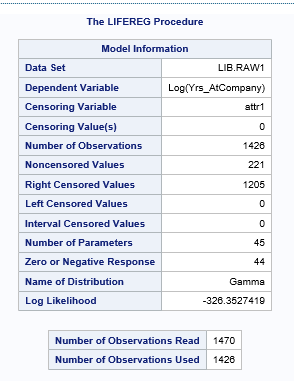
Exponential Distribution



WEIBULL Distribution



Gamma Model

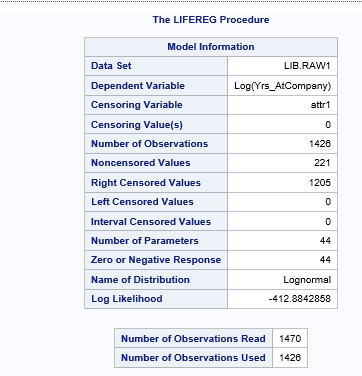


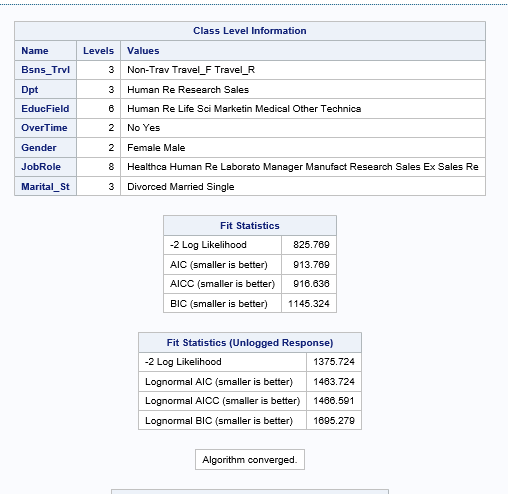
Comparison of the Models built using different Distribution



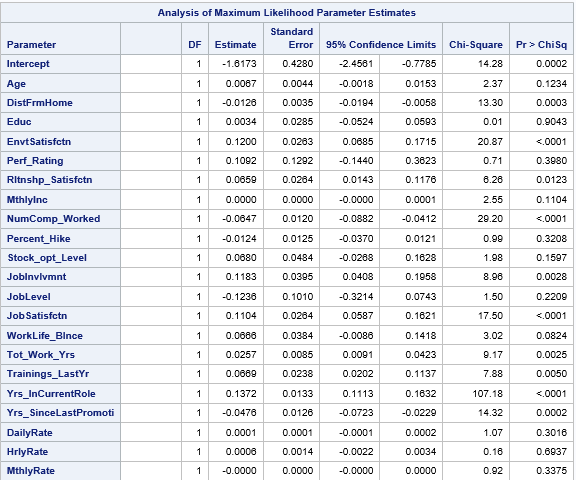
*Log Likelihood Ratio of WEIBULL distribution is almost close to that of GAMMA. But model built using WEIBULL Distribution encountered an error “*Validity of the model fit is questionable*. Hence we will have built model using Lognormal distribution*

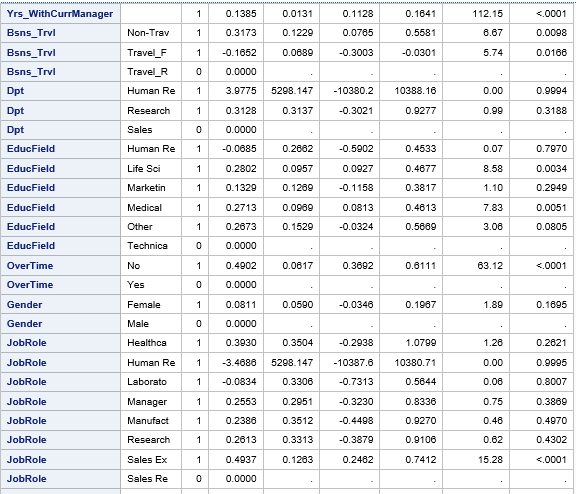
1. Results of Lognormal Model

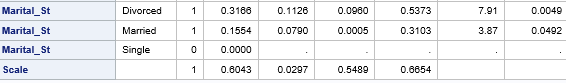






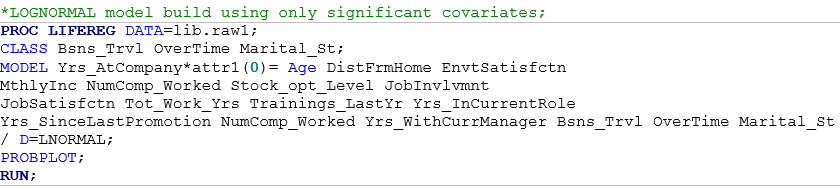




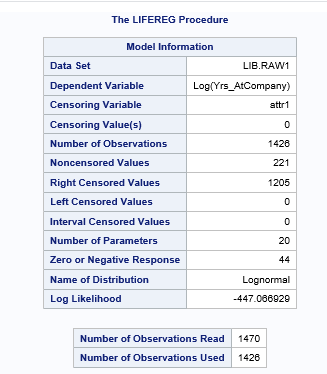


From table we observe that few covariates have p-value > 0.05(alpha = 5%). Hence we considered to remove these variables as they are not significant

1. LOGNORMAL model build using only significant covariates

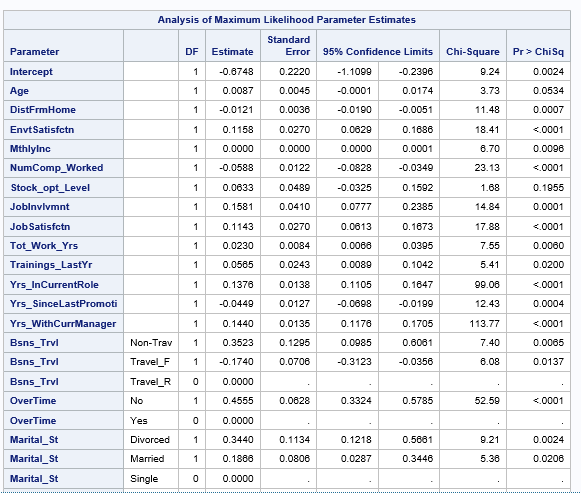


Results









From above table we can observe that the covariates are significant hence these variable are contributing to the attrition rate.

*Modeling using covariates found significant from Strata analysis– (except Bonus variables)*

*Censored: Attrition*

*Time: Yrs at company*

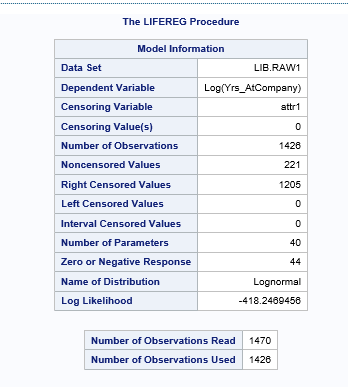
(CLASS Bsns\_Trvl Dpt EducField OverTime JobRole Marital\_St;

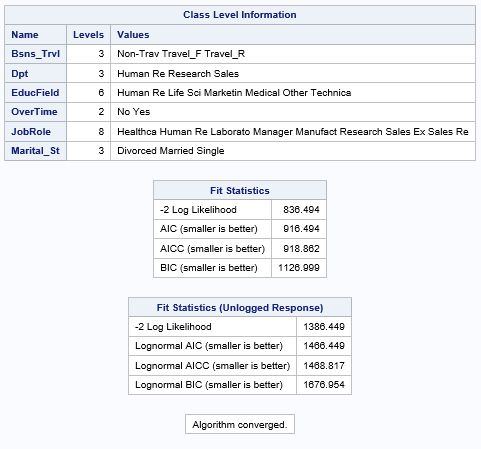
Rltnshp\_Satisfctn MthlyInc NumComp\_Worked Percent\_Hike Stock\_opt\_Level JobInvlvmnt

JobLevel JobSatisfctn WorkLife\_Blnce Tot\_Work\_Yrs Yrs\_InCurrentRole

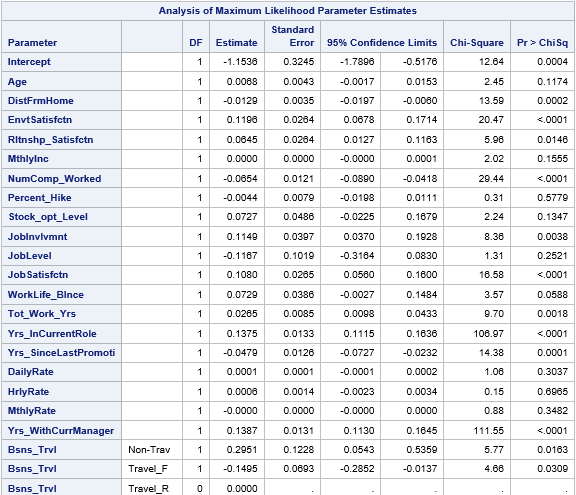
Yrs\_SinceLastPromotion DailyRate HrlyRate MthlyRate NumComp\_Worked Yrs\_WithCurrManager)

1. Results of Lognormal Model





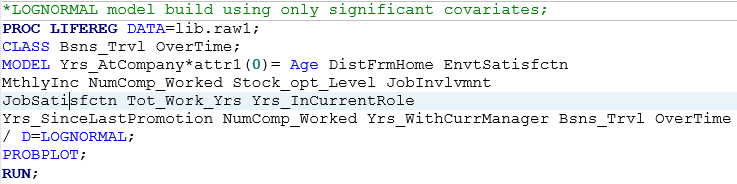




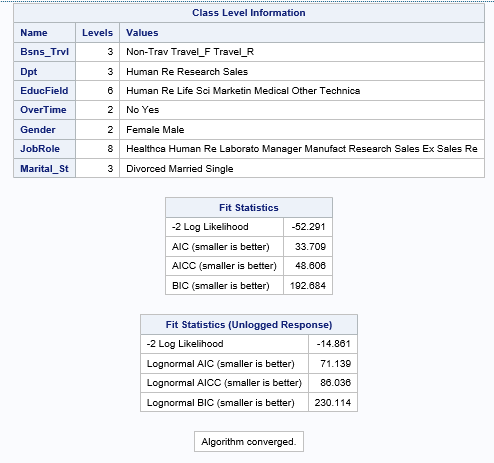


From table we observe that few covariates have p-value > 0.05(alpha = 5%). Hence we considered to remove these variables as they are not significant

1. LOGNORMAL model build using only significant covariates



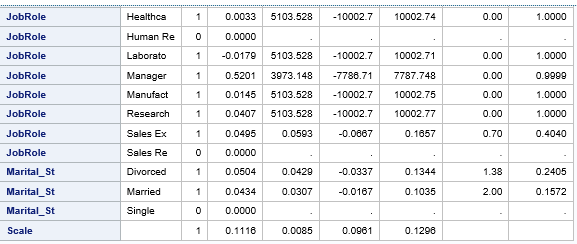




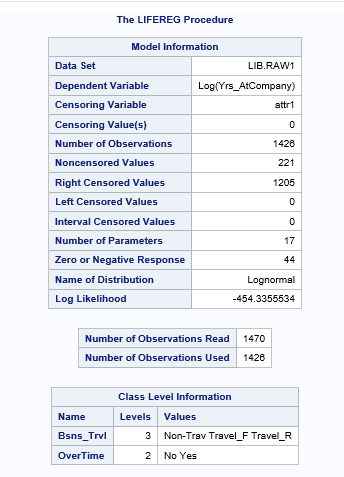


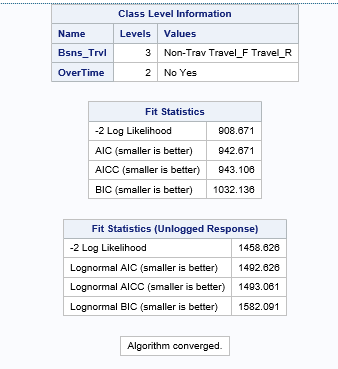


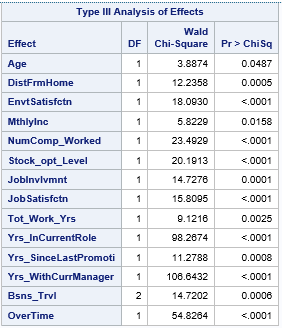


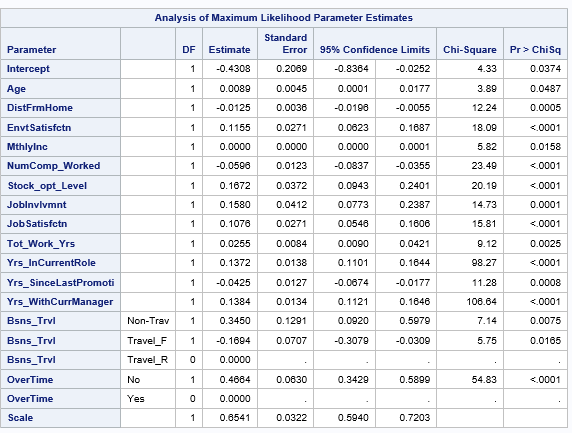


Results





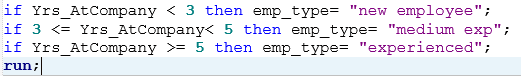




From above table we can observe that the covariates are significant hence these variable are contributing to the attrition rate.

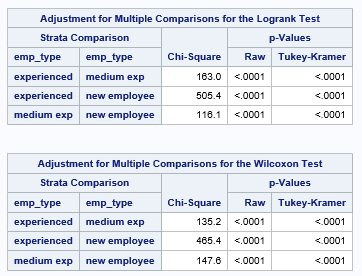
Approach 3

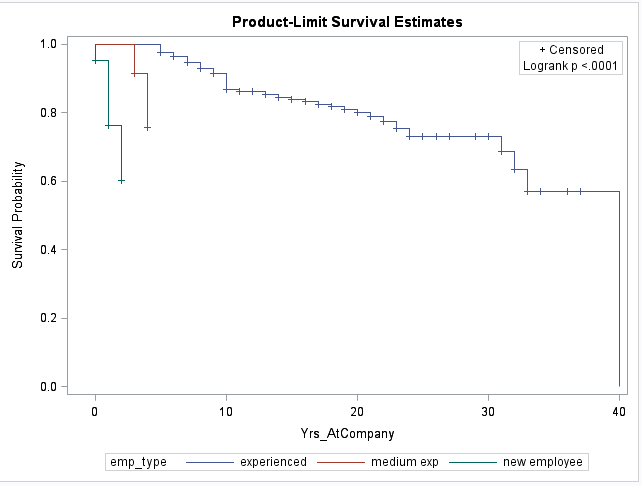
Create a variable Employee type (emp\_type) to group the employees based on their Tenure in the company



Lifetest procedure to test for linear relation between hazard types







From above table and graphs we observe that p value is significant, each employee group are not the same.

Subset the data into 3 employee groups

Log likelihood ratio test to investigate if we need to create different models for each employee type.

C:\Users\pavra\Desktop\unnamed.png

We have p-value = 0 which is not significant. Hence we can not combine employee types and need to model them separately.

*Modeling New Employees type – (except Bonus variables)*

*Censored: Attrition*

*Time: Yrs at company*

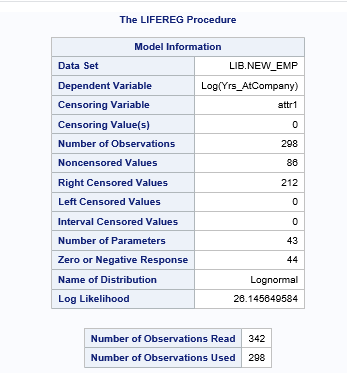
(CLASS Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St

Age DistFrmHome Educ EnvtSatisfctn Perf\_Rating

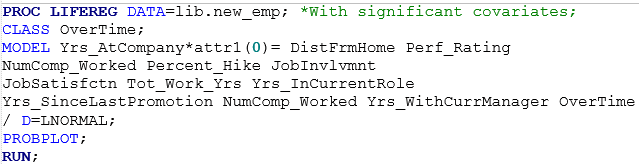
Rltnshp\_Satisfctn MthlyInc NumComp\_Worked Percent\_Hike Stock\_opt\_Level JobInvlvmnt

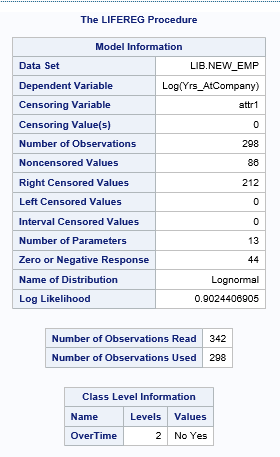
JobLevel JobSatisfctn WorkLife\_Blnce Tot\_Work\_Yrs Trainings\_LastYr Yrs\_InCurrentRole

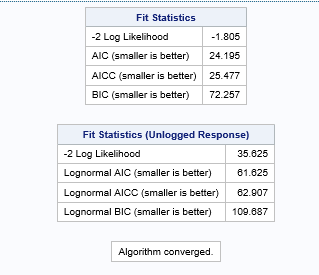
Yrs\_SinceLastPromotion DailyRate HrlyRate MthlyRate NumComp\_Worked Yrs\_WithCurrManager Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St)

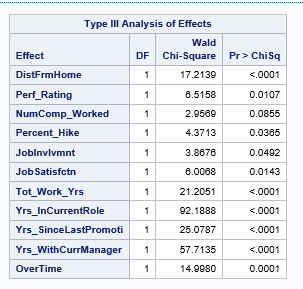


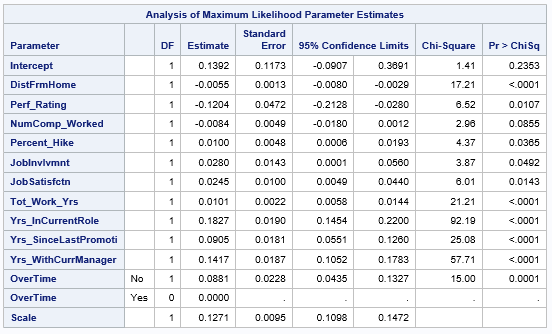
Model for New Employees with significant co variates











From above table we can observe that the covariates are significant hence these variable are contributing to the attrition rate.

**Modeling Medium Experienced Employees type – (except Bonus variables)**

**Censored: Attrition**

**Time: Yrs at company**

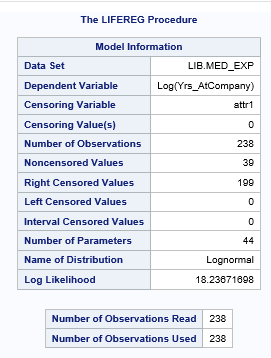
(CLASS Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St

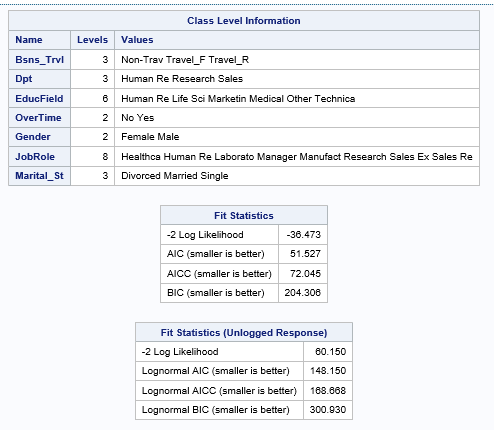
Age DistFrmHome Educ EnvtSatisfctn Perf\_Rating

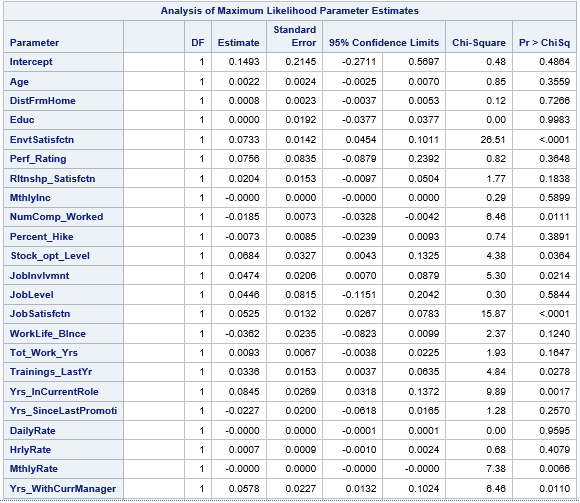
Rltnshp\_Satisfctn MthlyInc NumComp\_Worked Percent\_Hike Stock\_opt\_Level JobInvlvmnt

JobLevel JobSatisfctn WorkLife\_Blnce Tot\_Work\_Yrs Trainings\_LastYr Yrs\_InCurrentRole

Yrs\_SinceLastPromotion DailyRate HrlyRate MthlyRate NumComp\_Worked Yrs\_WithCurrManager Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St)



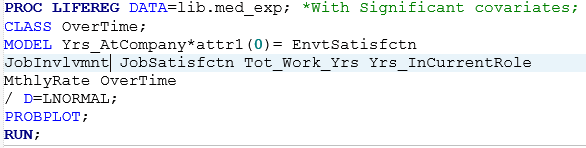


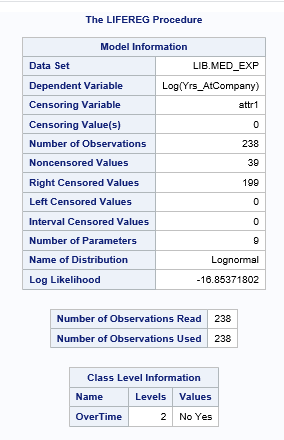


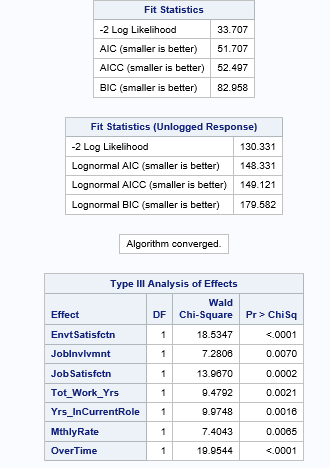


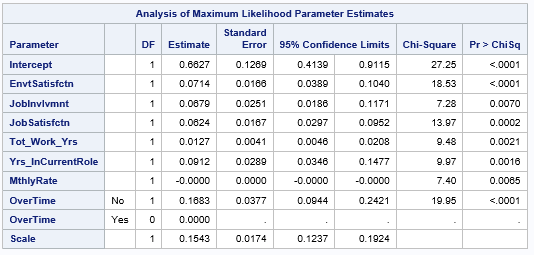
From table we observe that few covariates have p-value > 0.05(alpha = 5%). Hence we considered to remove these variables as they are not significant

Model for Medium Experienced employees with significant covariates









From above table we can observe that the covariates are significant hence these variable are contributing to the attrition rate.

**Modeling Experienced Employees type – (except Bonus variables)**

**Censored: Attrition**

**Time: Yrs at company**

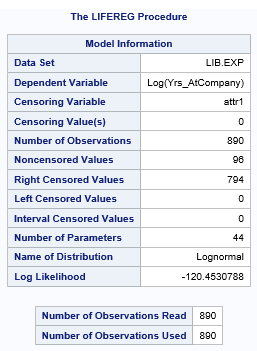
(CLASS Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St

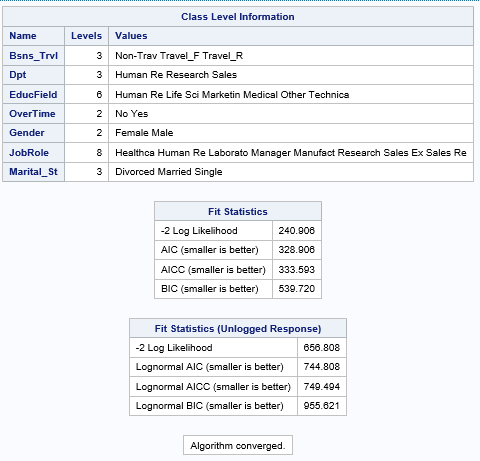
Age DistFrmHome Educ EnvtSatisfctn Perf\_Rating

Rltnshp\_Satisfctn MthlyInc NumComp\_Worked Percent\_Hike Stock\_opt\_Level JobInvlvmnt

JobLevel JobSatisfctn WorkLife\_Blnce Tot\_Work\_Yrs Trainings\_LastYr Yrs\_InCurrentRole

Yrs\_SinceLastPromotion DailyRate HrlyRate MthlyRate NumComp\_Worked Yrs\_WithCurrManager Bsns\_Trvl Dpt EducField OverTime Gender JobRole Marital\_St)

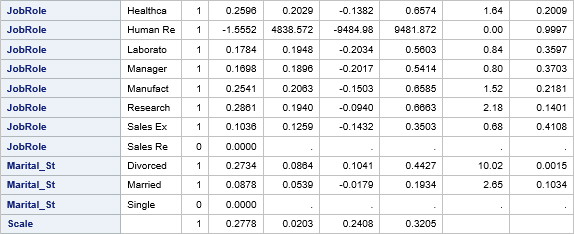












From table we observe that few covariates have p-value > 0.05(alpha = 5%). Hence we considered to remove these variables as they are not significant

Model for Experienced employees with significant covariates

