Authors

Divya Murli

**PROJECT GROUP 5**

SURVIVAL ANALYSIS ON FERMA LOGI’S EMPLOYEE ATTRITION

**Authors:**

Divya Murli

Prachi Gupta

Priya Mudambi

Sneha Prabhu

Srilekha Lava

Subhashree Srinivisan

Vasanthapriya Raja

**TABLE OF CONTENTS**

# Abstract……………………………………………………………………………………………………2

# Exploratory Data Analysis……………………………………………………….......................................2

# Assumptions……………………………………………………………………………………………...22

# Identification of Insignificant Variables…………………………………………....................................22

# Identification of Significant Variables……………………………………………...................................22

# Feature Engineering……………………………………………………………………………………...23

# Data Modelling……………………………………………………………………..................................24

# Recommendations………………………………………………………………………………………..55

# Conclusions………………………………………………………………………………………………55Appendix…………………………………………………………………………………………………55

* + - 1. ***Abstract***

The COO of FermaLogis Pharmaceutical company, Larry Hansen approached us with a business problem on Employee Attrition. He believed this had a major impact on his company. He expected that the business solution answered two major questions. One being ‘Who’ are the employees i.e. the characteristics/pattern of employees who leaves the company. The second major question is ‘Why’ these employees leaves the company.

We were expected to provide other insights on other questions like; when is the biggest danger of employees to leave, difference in attrition between different employee groups or categories, etc. It was noted that, both the young employees (who worked for the company less than three years) and experienced employees (who was with the company 5 or more years) is undergoing a professional and executive training which is a big investment from the company. To avoid this investment loss through training, they need to avoid employee attrition which is the base of the project.

Data (FermaLogis\_Event\_Type.csv) and its dictionary file were provided to us to perform the analysis.

1. ***Exploratory Data Analysis***

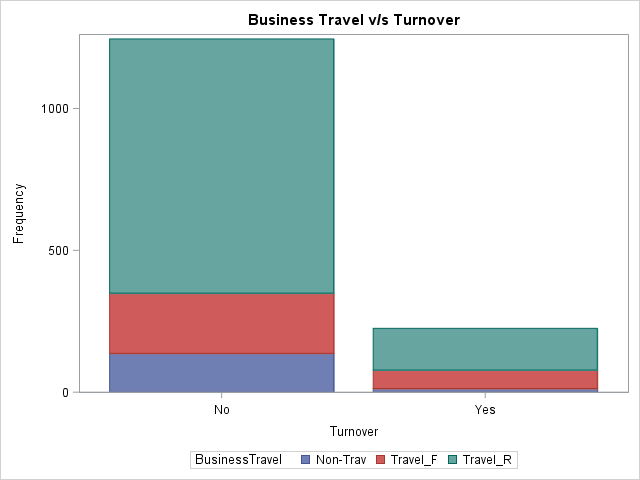
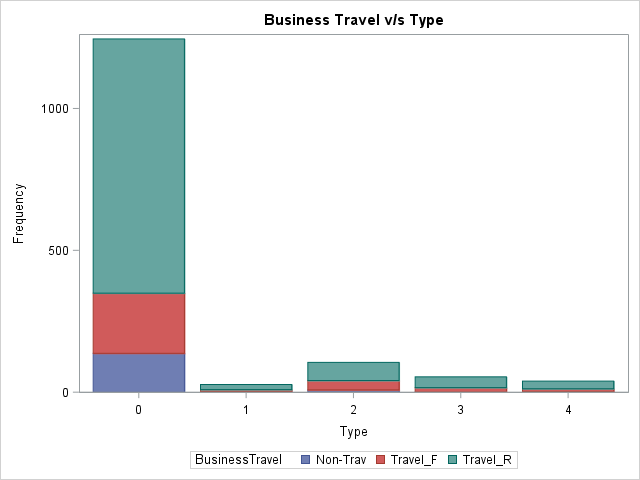
**Data Summary**

Dataset has total of 76 variables where 19 are categorical and 57 are continuous. Type variable is a target variable indicating whether an employee is currently active in company or not. There are a total of 1470 observations in the sample.

|  |  |  |
| --- | --- | --- |
| **Type** | **Observations** | **Contribution in %** |
| 0 - No turnover | 1245 | 85% |
| 1 – Retirement | 27 | 2% |
| 2 - Voluntary Resignation | 105 | 7% |
| 3 - Involuntary Resignation (Health problems, family matters) | 54 | 4% |
| 4 - Job Termination, Employee is Fired | 39 | 3% |

**Bivariate Analysis**

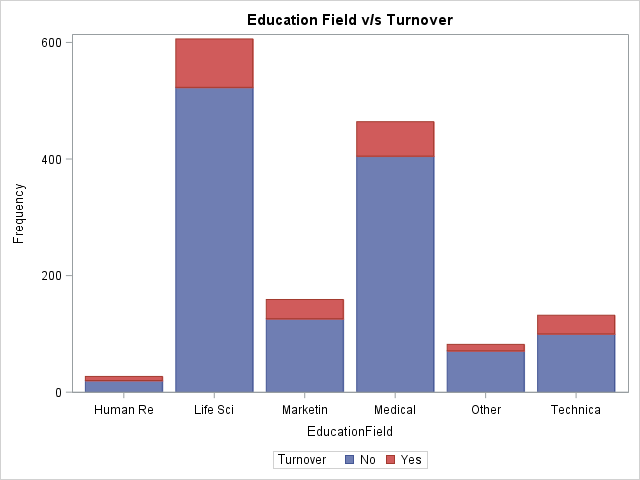
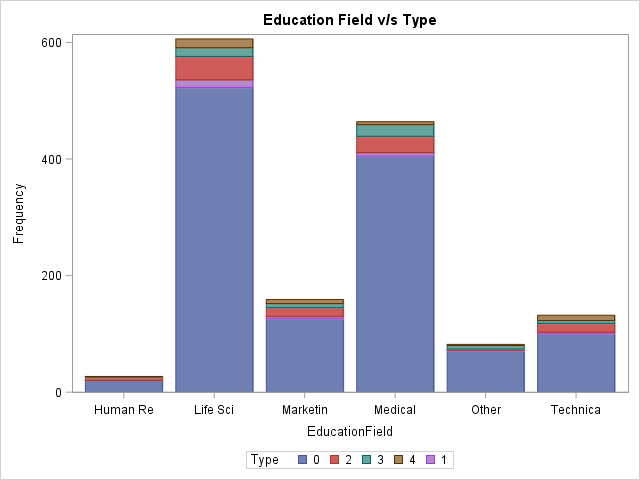
*Business Travel*

Observations:

Employees who travel rarely contribute more to voluntary resignation and other categories as well.

*Education Field*

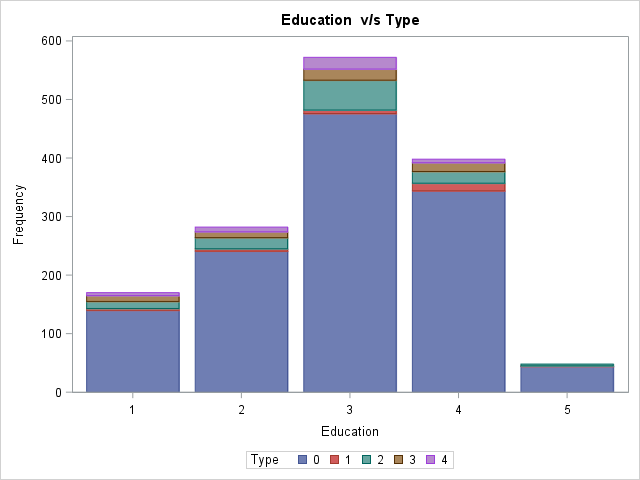
 

Observations:

Among all the fields, people who belong to “Life Science” seem to leave the company more than others though this plot shows that there is not a significant impact of education field on turnover.

The red and green stripes in most of the education fields suggest that people leave the company mainly due to type 2 and type 3.

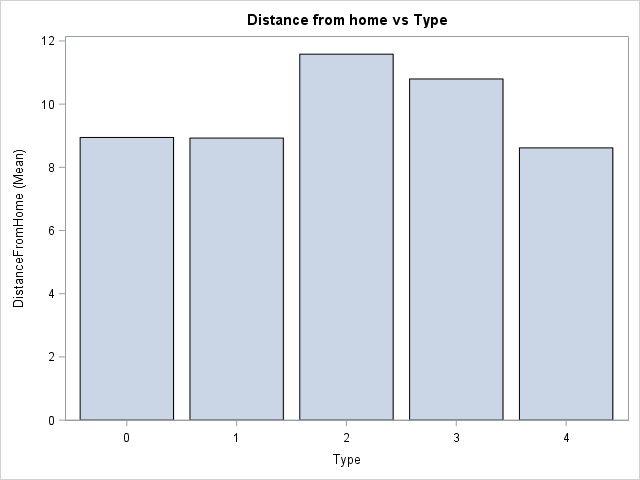
*Education*

****

Observations:

Though the major cause of leaving the company is due to voluntary resignation, there is not a significant impact of the type of education of employees on turnover.

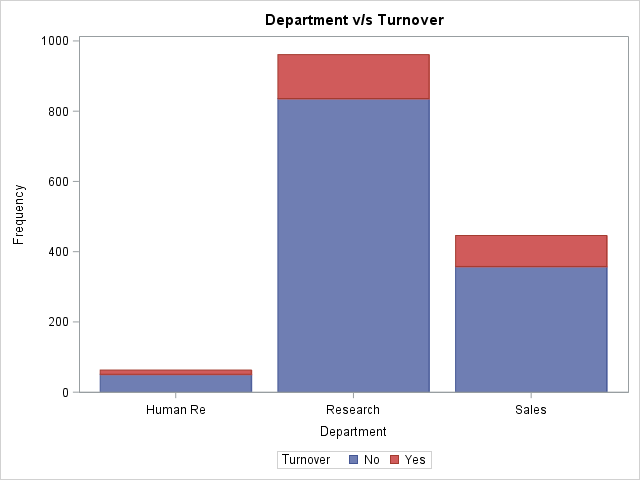
*DistanceFromHome*



Observations:

The below plot shows that as the distance from the house to office increases, people tend to leave the company more. People resign voluntarily from the company when their office is too far from their house and this might due to various reasons such as traffic, hectic travel, etc.

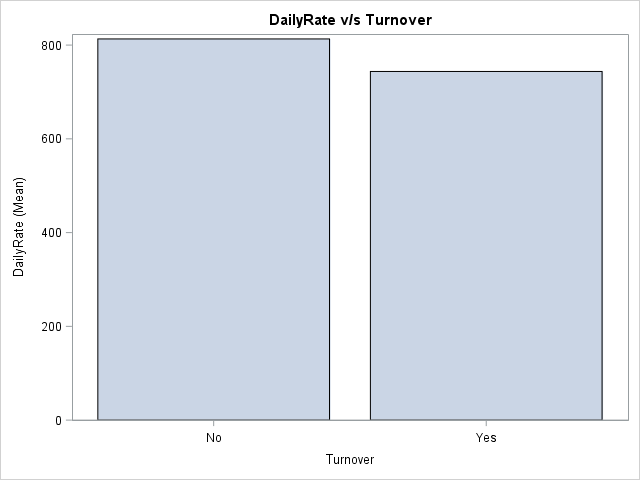
*Department*



Observations:

The people belonging to the research department leave the company more as compared to other departments but overall, there doesn’t seem to be a significant impact of department on turnover.

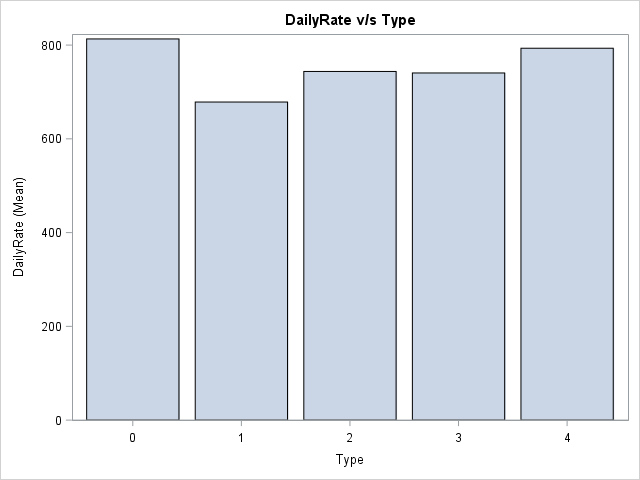
*DailyRate (vs Turnover)*



Observations:

As the daily wages of employees increases, people are more satisfied with their job, hence they do not leave the company as compared to the people who are paid less on a daily basis.

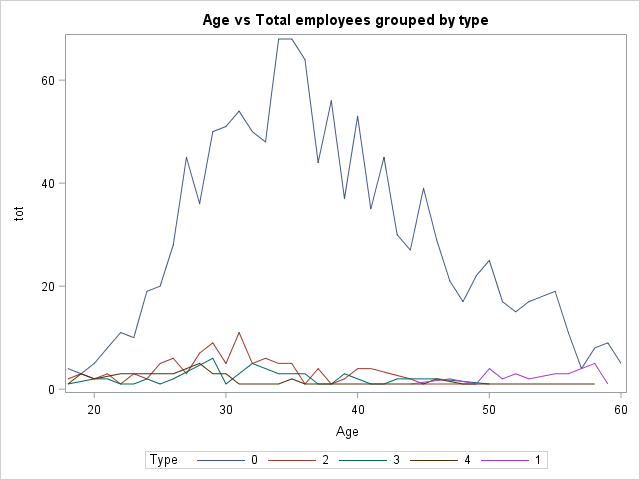
*DailyRate (vs Type)*



Observations:

Below plot shows that people whose daily rate is 800 or above do not churn at all. People who voluntarily leave the company have their daily rates between 600 and 800.

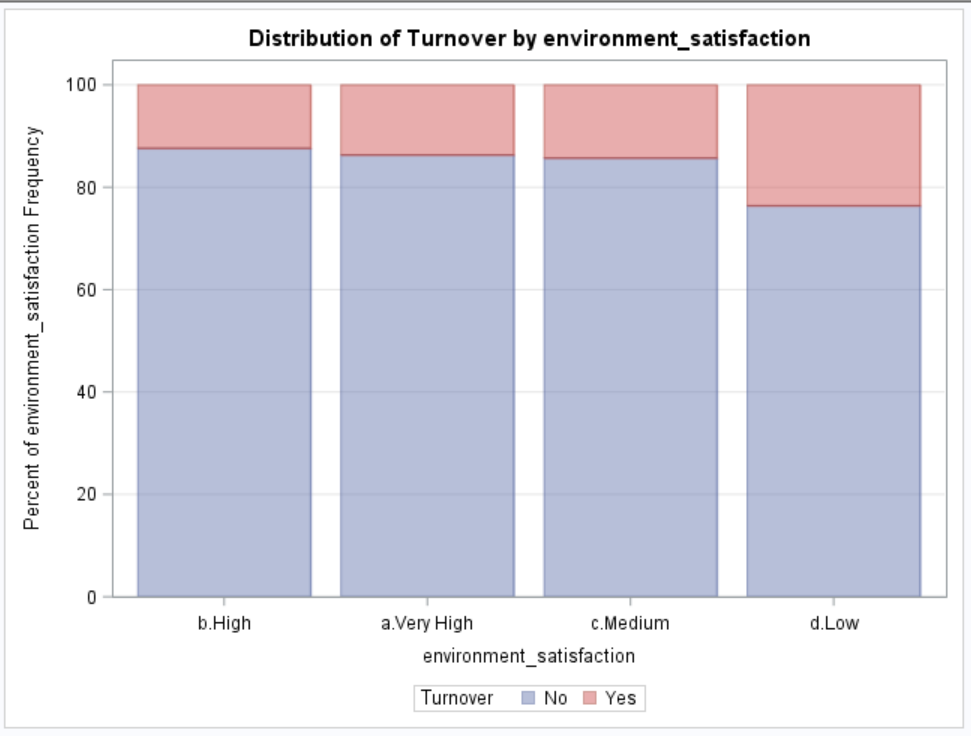
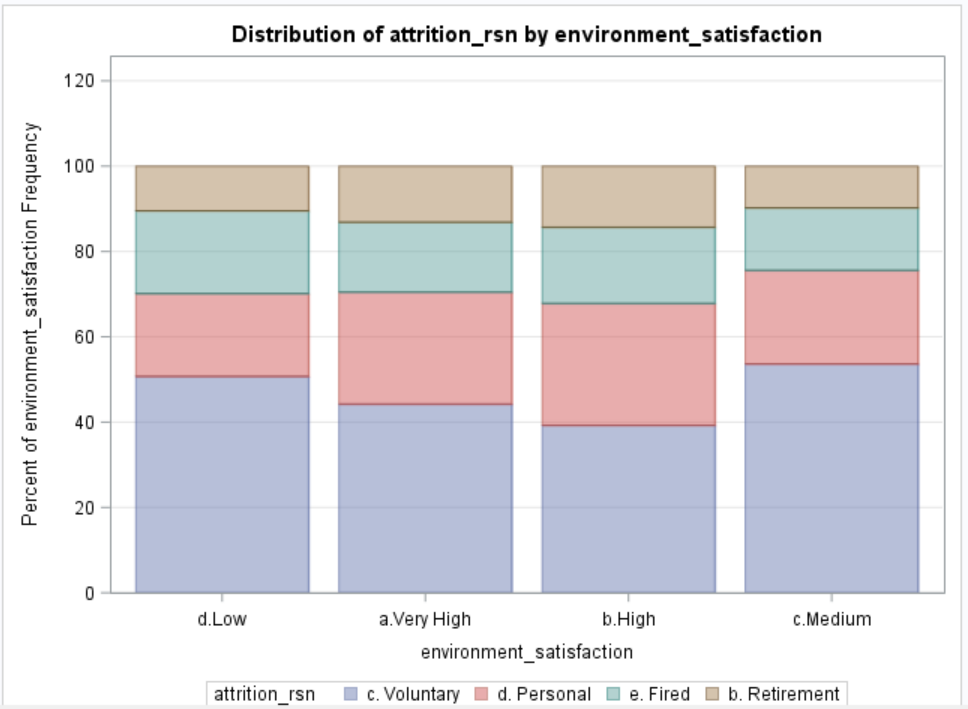
*Age v/s Total employees grouped by Type*



Observations:

Majority of employees are between 30-40 years of age and these people do not seem to churn a lot. However, a sudden spike at 30 shows that there are few people between ages 30 and 40 who leave the company voluntarily.

*Environment Satisfaction*

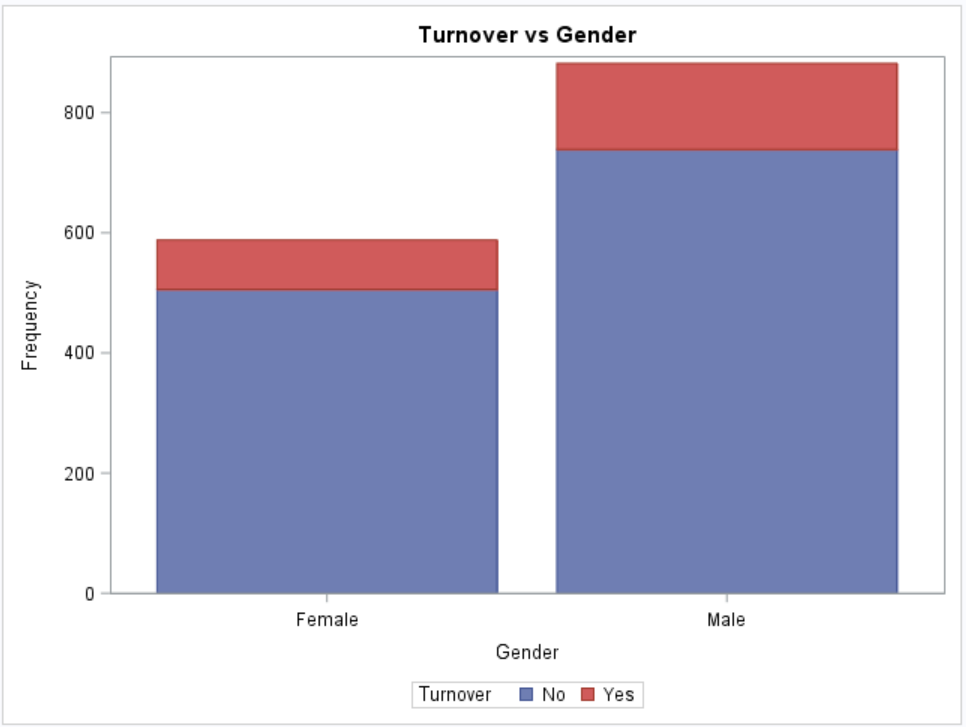
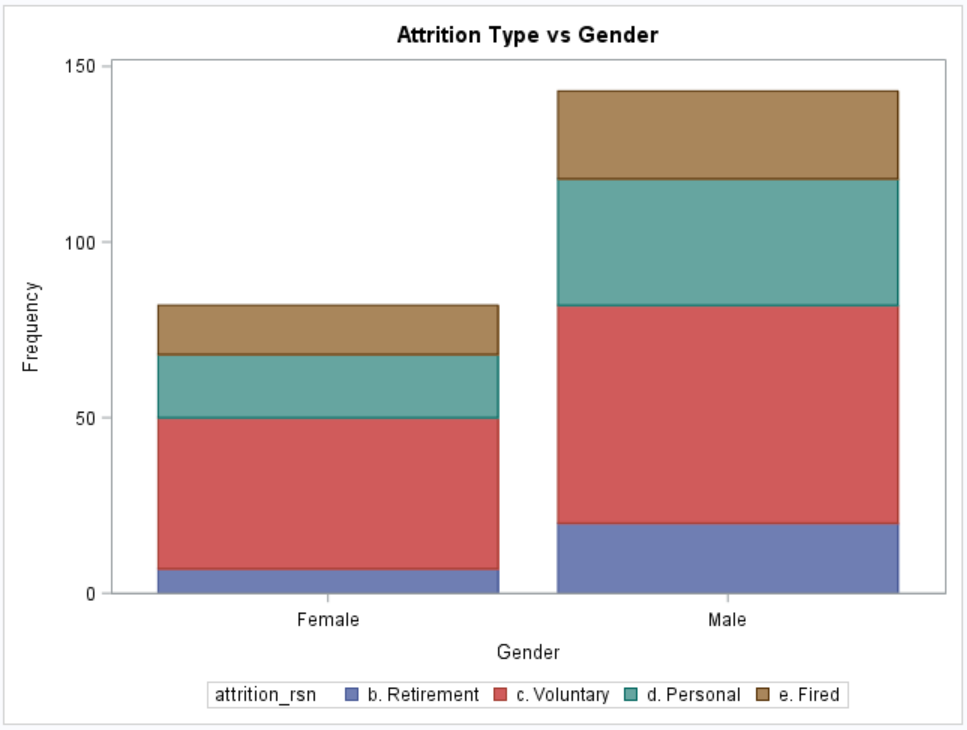
 

Observations:

Turnover is high at ‘Low’ Environment Satisfaction.

Voluntary attrition is highest at ‘Low’ Environment Satisfaction.

*Gender*

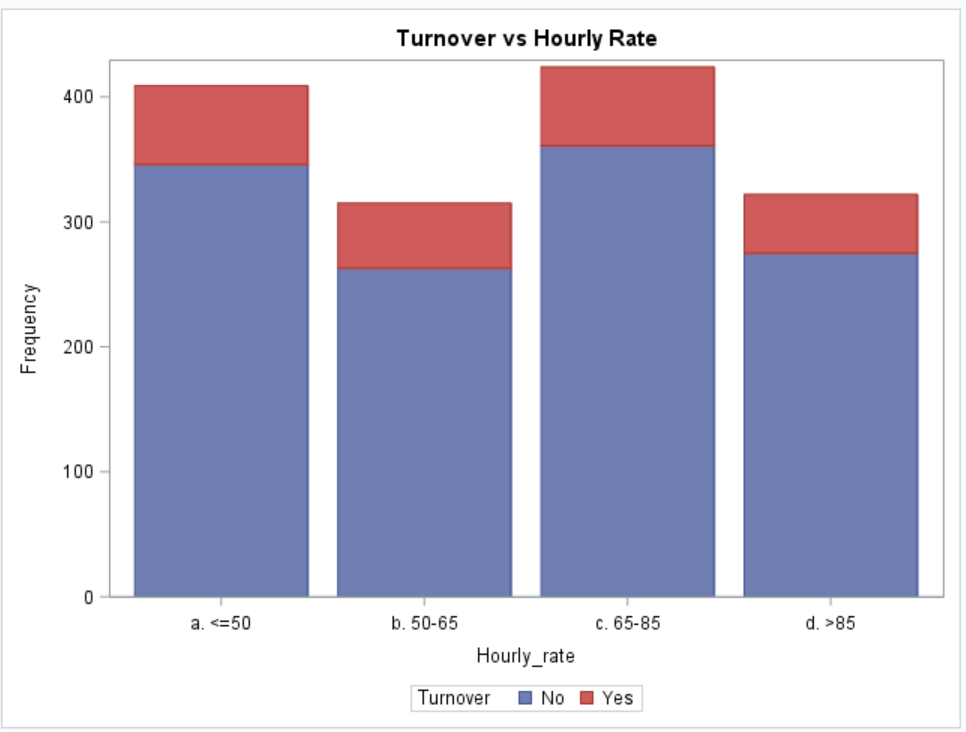
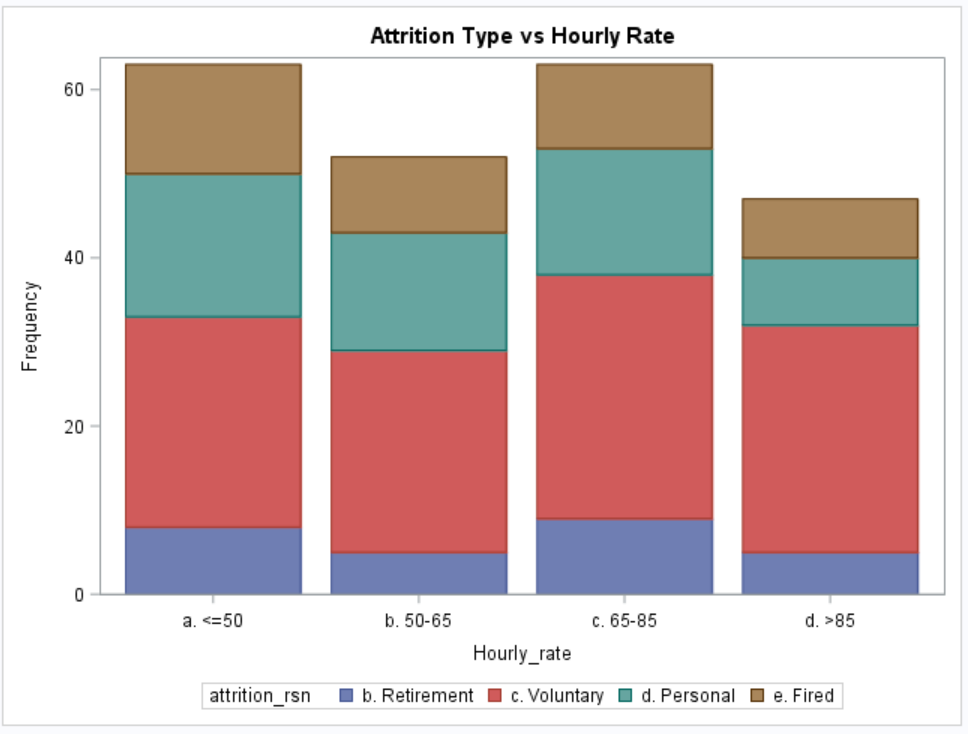
 

Observations:

Males have slightly higher turnover rate compared to females.

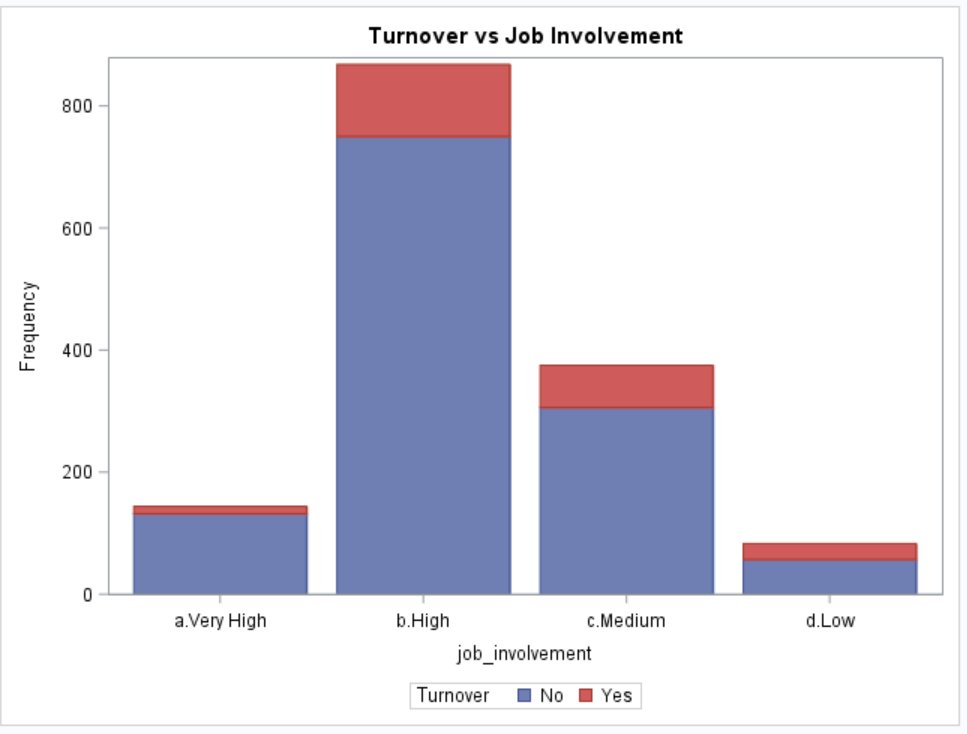
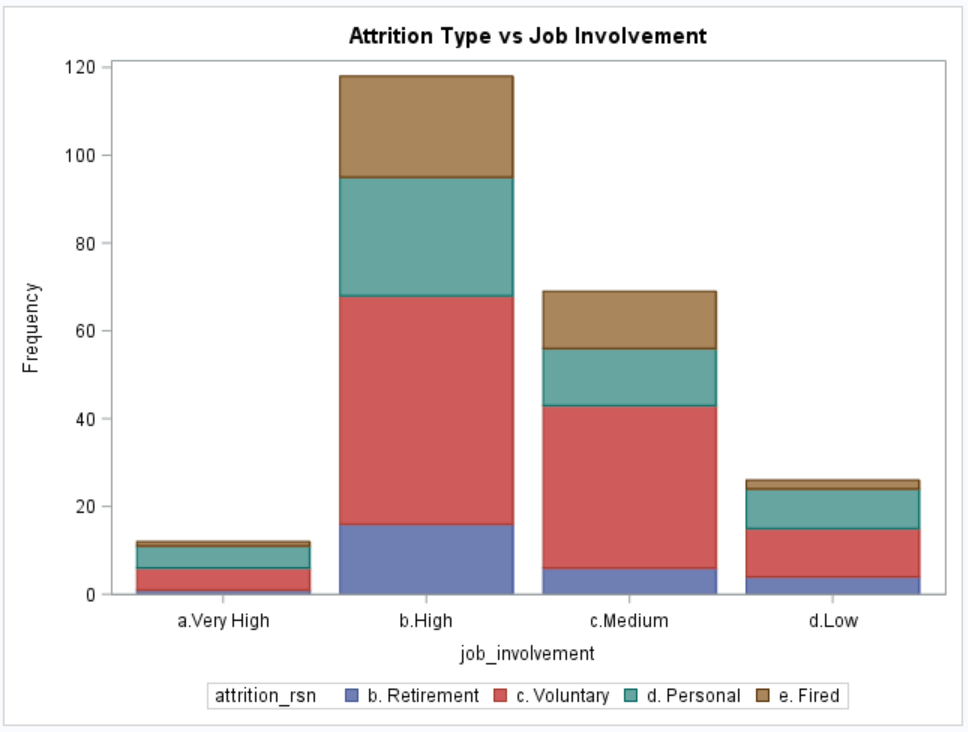
Within employees who attrite, females have higher proportion of voluntary attrition compared to males.

*Hourly Rate*

No clear pattern in turnover or its types observed in hourly rates.

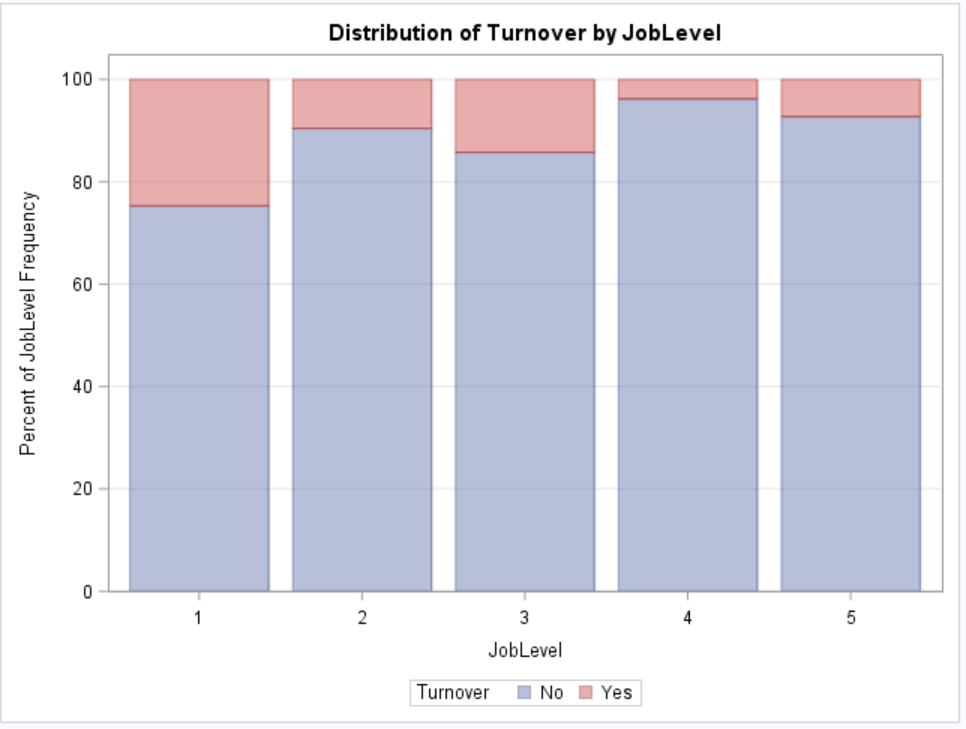
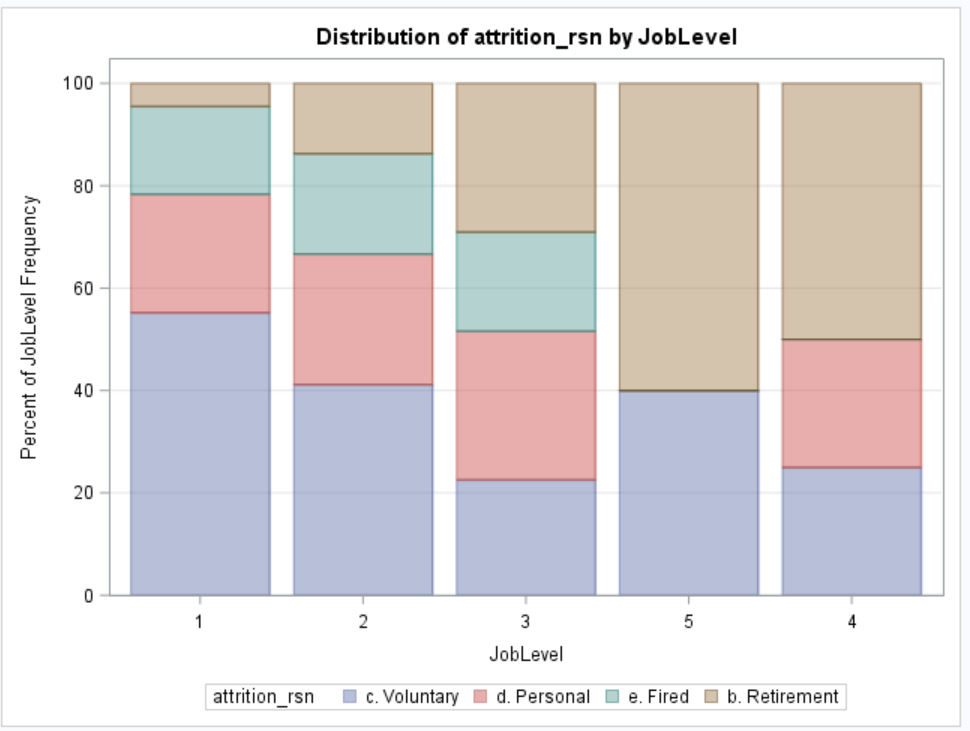
*Job Involvement*

Observations:

Employees with Lowest job involvement have highest turnover rate.

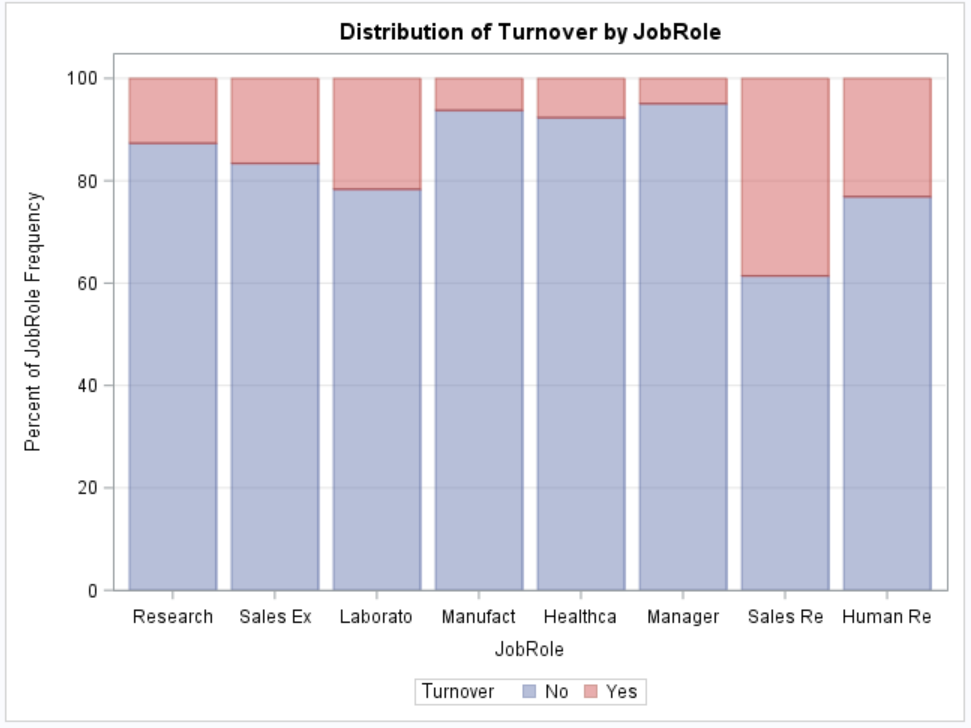
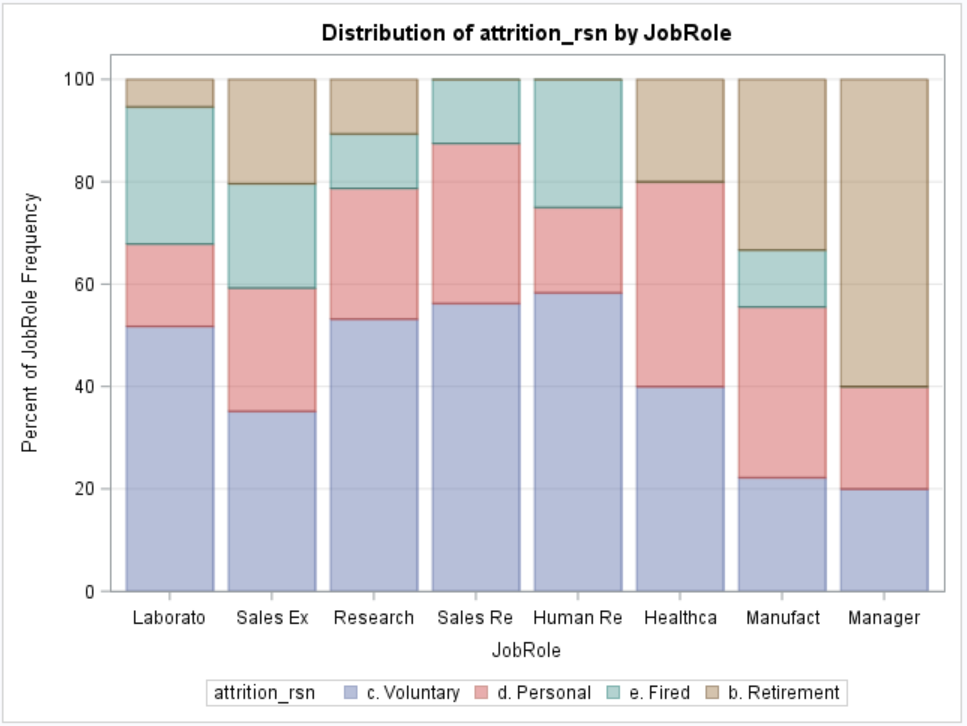
*Job Level*

Observations:

* Job Levels 1 to 3 have higher turnover rate.
* Also, all firing happens at job levels 1 to 3.
* Retirement rate is very high at job levels 4 and 5

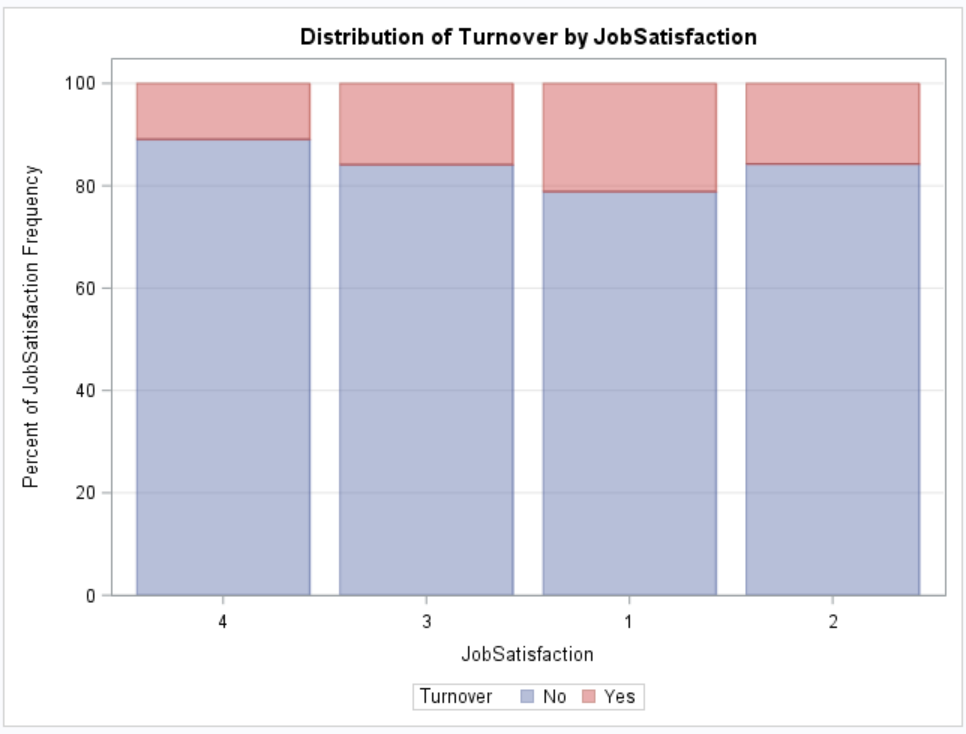
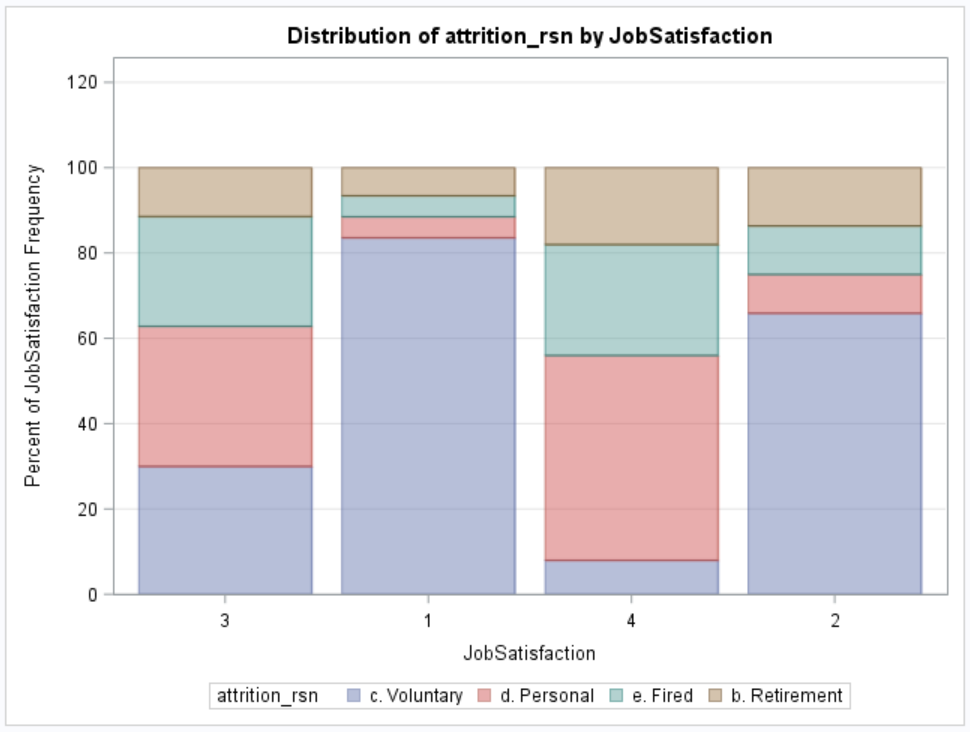
*Job Role*

Observations:

* Turnover is high for Job roles – Sales Rep, Human Resources, and Laboratories.
* Turnover is very low in Manufacturing, Healthcare and Manager roles.
* Retirement rates are very high in Manager and Manufacturing roles.
* Voluntary attrition is the highest reason for turnover in ‘Sales Rep’ and ‘Human Resources’ job roles.

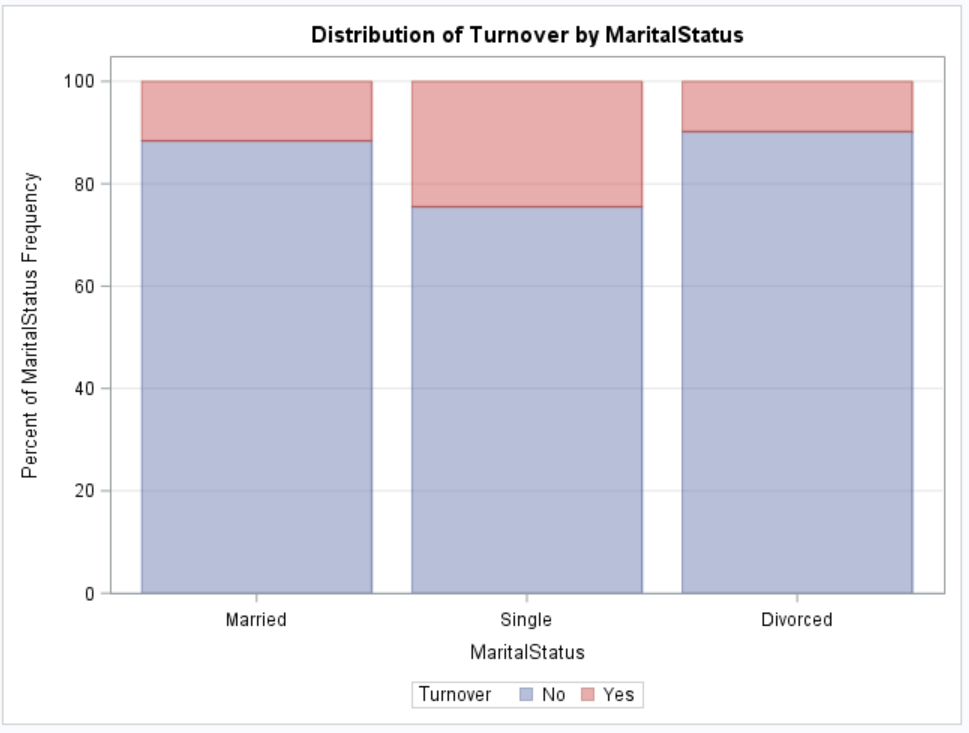
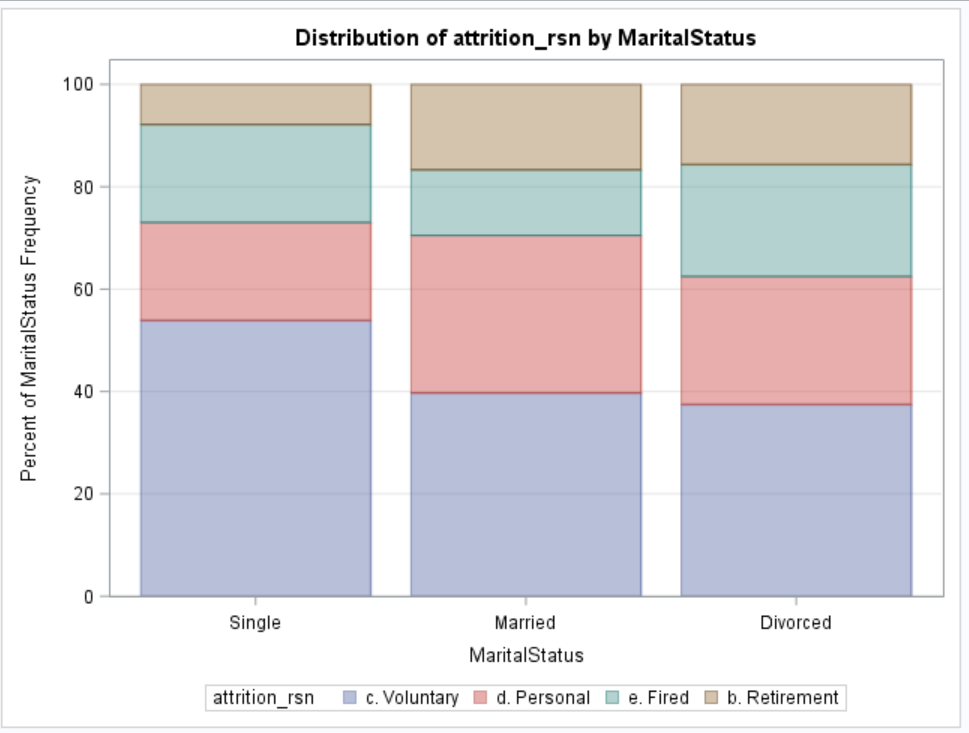
*Job Satisfaction*

Observations:

* Low job satisfaction levels correlate with high turnover rates.
* Also, low job satisfaction (level 1 and 2) have very high voluntary attrition.
* Job levels 3 and 4 are high on Retirement and Firing reasons of Attrition.

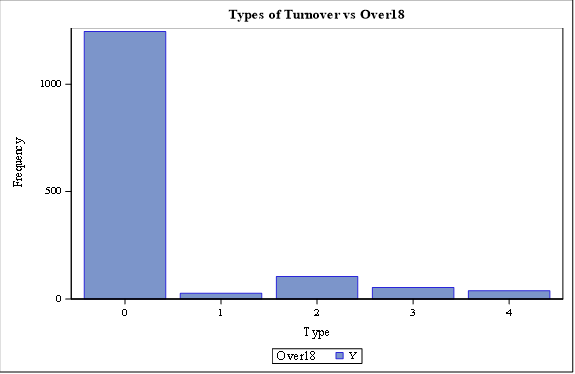
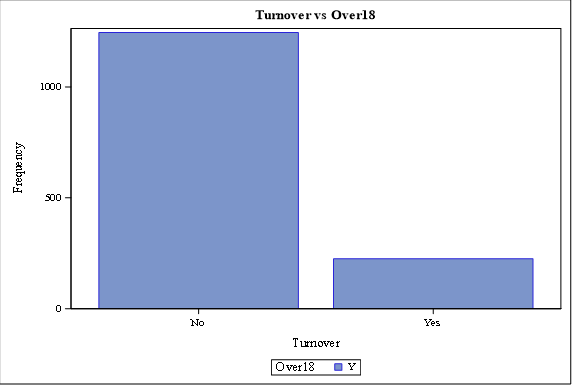
*Marital Status*

Observations:

Single employees have higher turnover rate and they also have high voluntary attrition rate

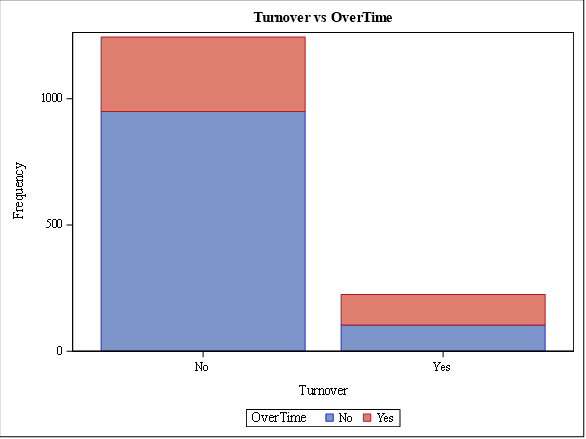
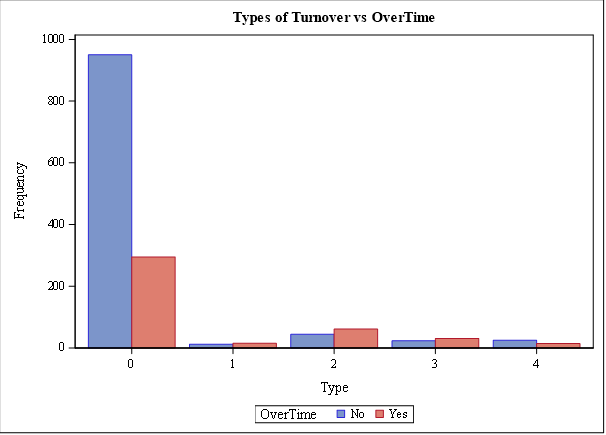
*Over18*



Observations:

All the employees from the records are over 18.

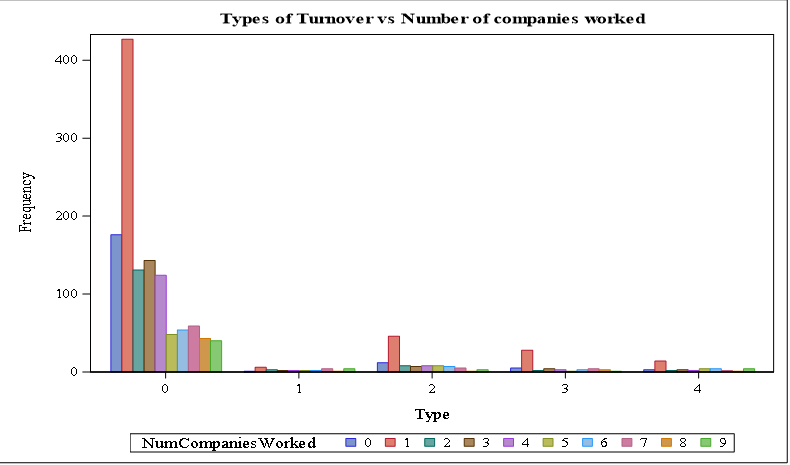
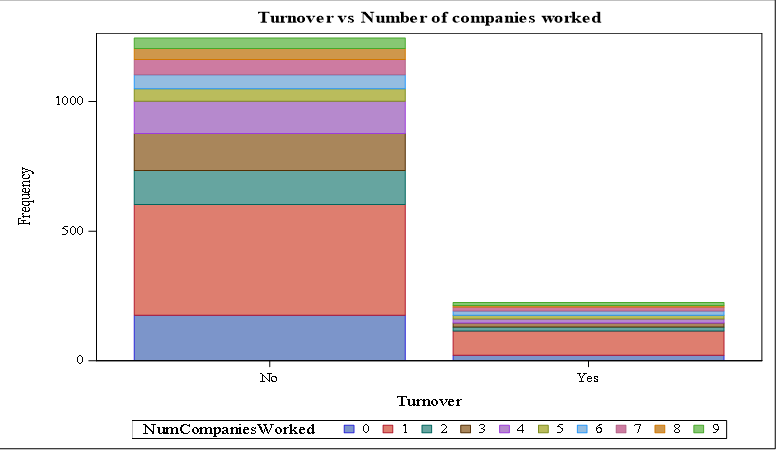
*Overtime*

Observations:

* Almost same percent of employee were present in both overtime and no- overtime category who Turnover.
* In the second graph, what type of turn over affected the most is represented, the first highest in the turn over category is voluntary retirement.
* The second highest seems to be, Involuntary retirement due to health reasons or family reasons, which is understandable as overtime causes considerable damage to work-life process.

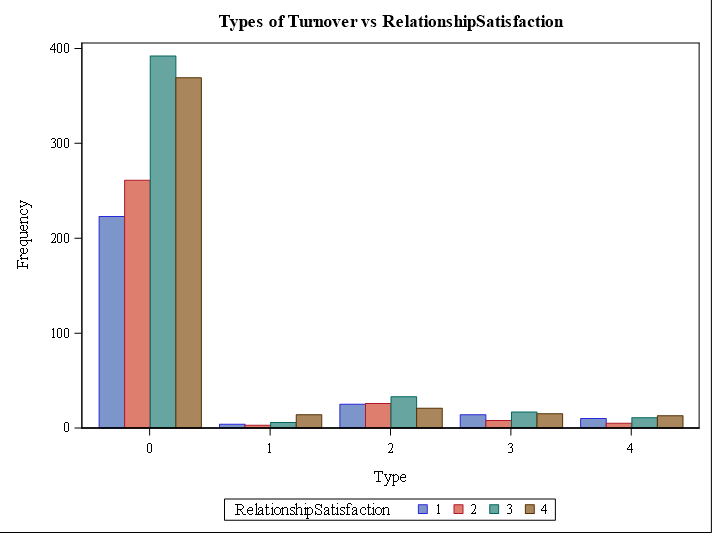
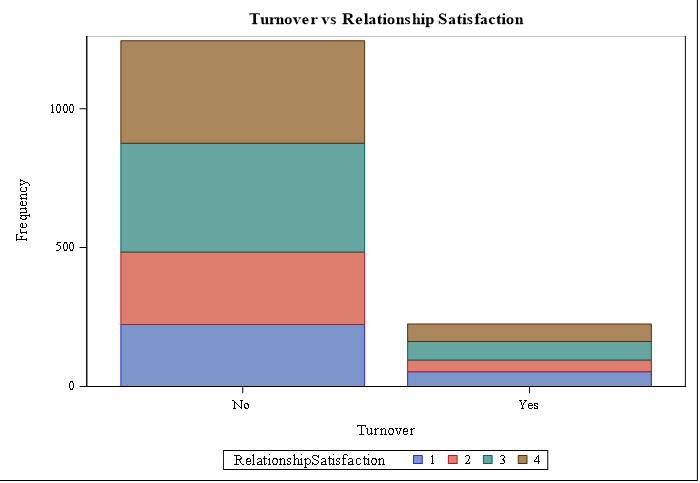
*NumCompaniesWorked*

**

Observations:

* Employees worked for lesser number of companies have turnover rate, the number of companies an employee worked is less than 3 he is the person who turn-over more and similar pattern is observed in people who do not turn-over too.
* Other than voluntary retirement, people from involuntary retirement stood out here.

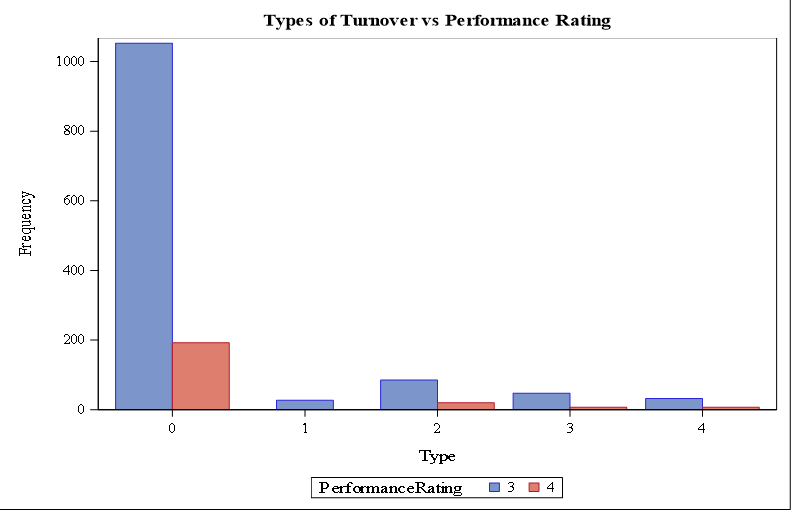
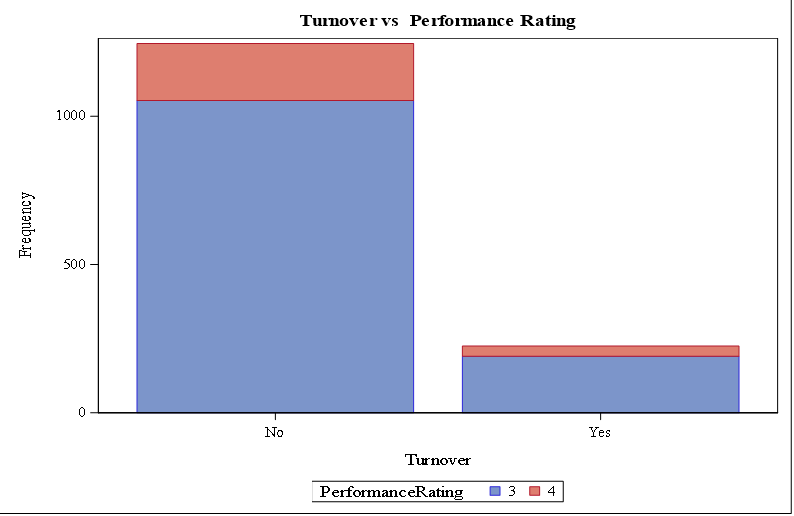
*Relationship Satisfaction*

****

Observations:

* Relationship Satisfaction seems to have no significance trend, in both the yes/no turnover categories they are segregated similarly.
* In type of turnover, it is voluntary resignation seems to be high in number.

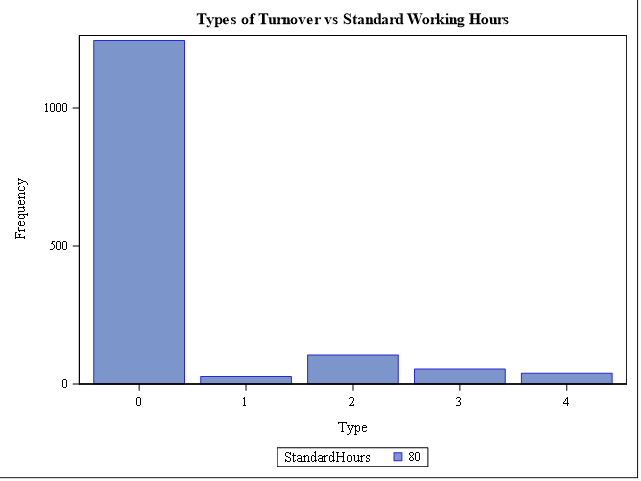
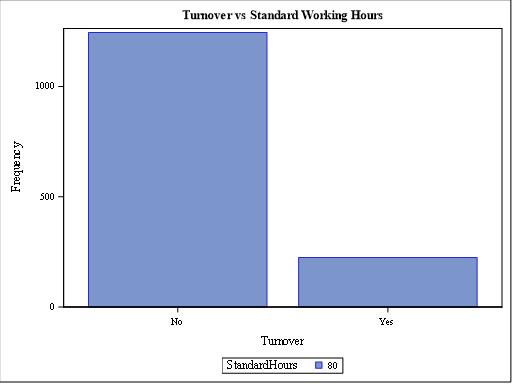
*Performance Rating*

**

Observations:

* Majority of employees churned belongs to lesser performance rating, that is 3. It seems to have similar trend in No turnover category too.
* Even though voluntary and involuntary resignation is higher, the 4th category which is termination is also considerably notable.

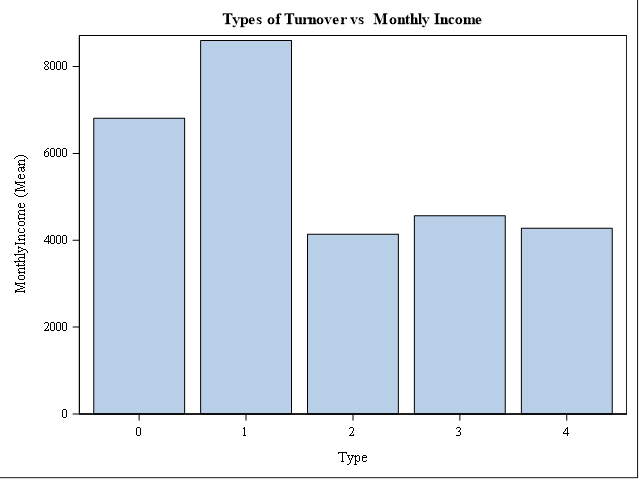
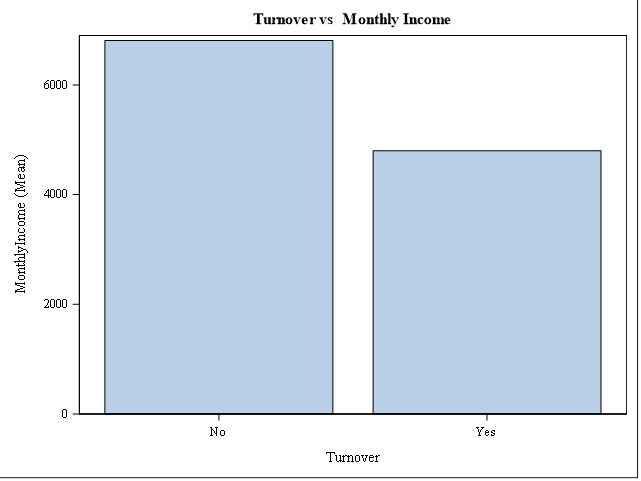
*Standard Hours*

**

Observations:

Every Employee in the data has standard working hours of 80. This variable has no impact on the target variable.

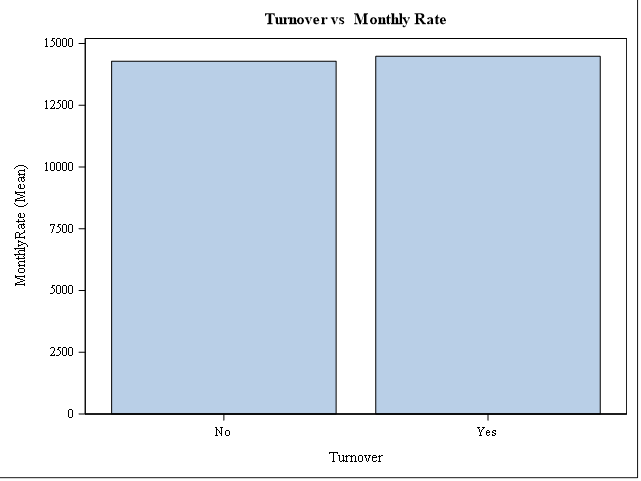
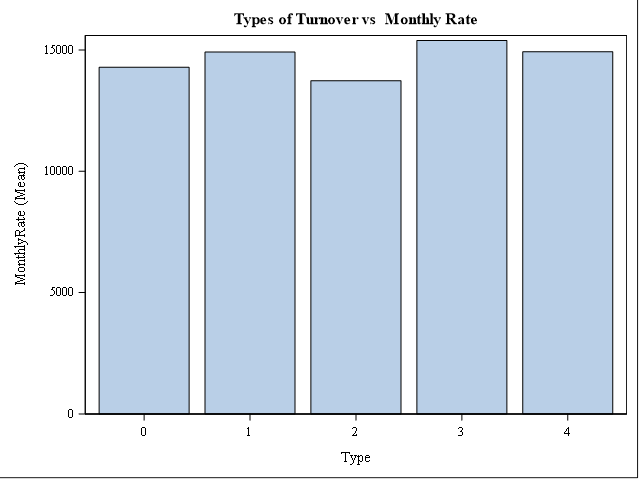
*Monthly Income*

****

Observations:

* The mean of the monthly income is observed over the variables, turnover and types of turnover. The mean income of people who is the reason for employee attrition is lower.
* The men income of people belong to retirement is high, and the lowest mean income belong to voluntary retirement.
* This Variable can be an important variable for model.

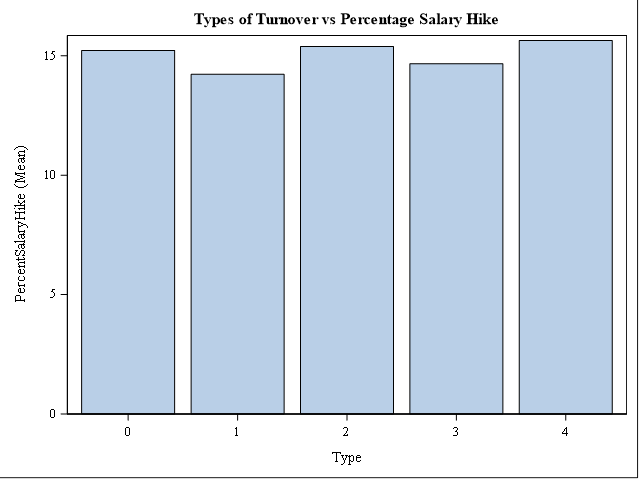
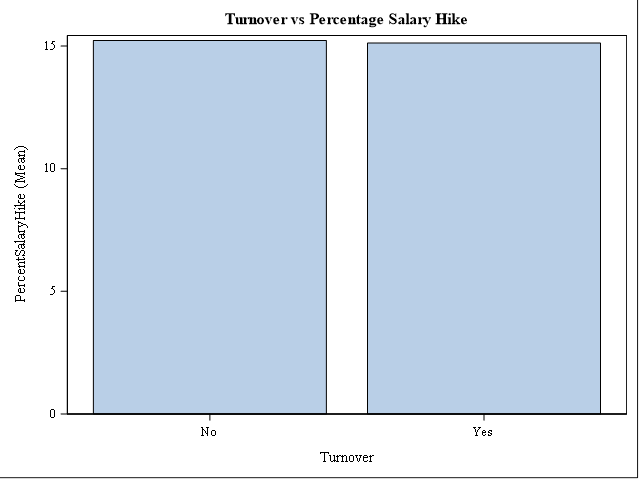
*Monthly Rate*

******

Observations:

Mean Monthly rate seems to have similar values in different types turnover and it seems to show no impact on Employee attrition.

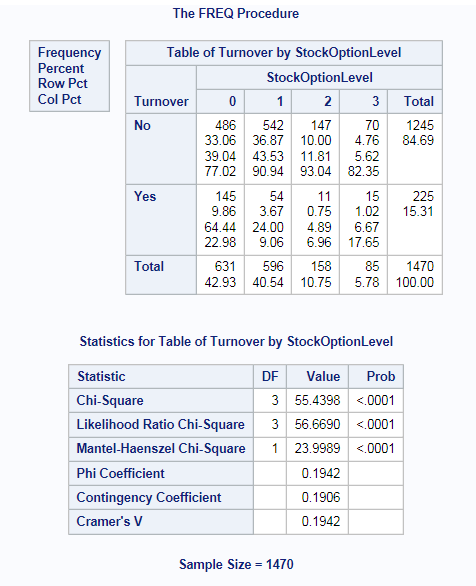
*Percent Salary Hike*

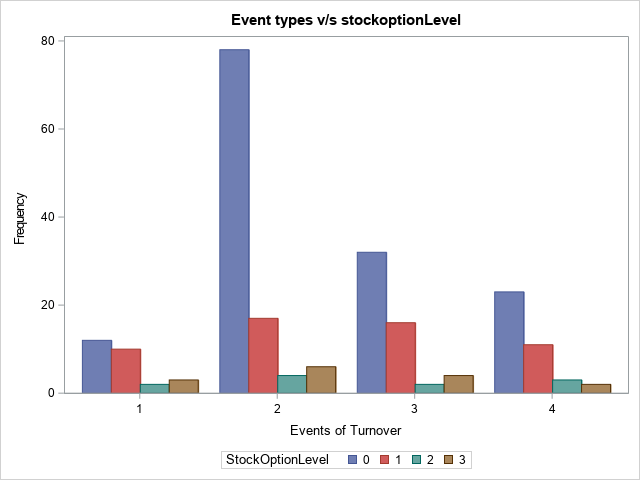
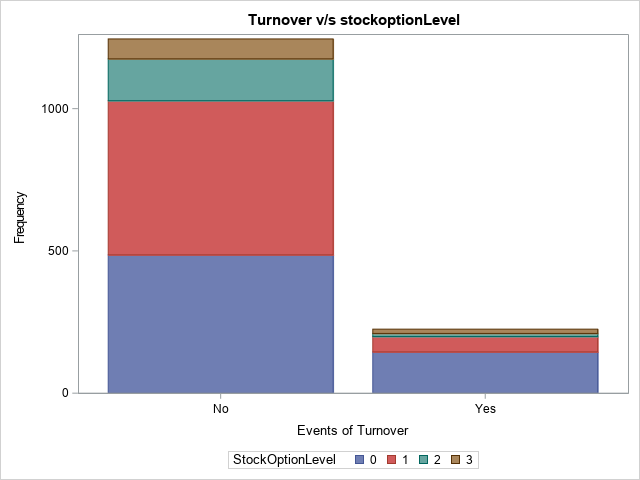
**

Observations:

The mean of percentage salary hike shows no impact on churning and it seems to have similar values in all the types of turnover.

*StockOptionLevel*

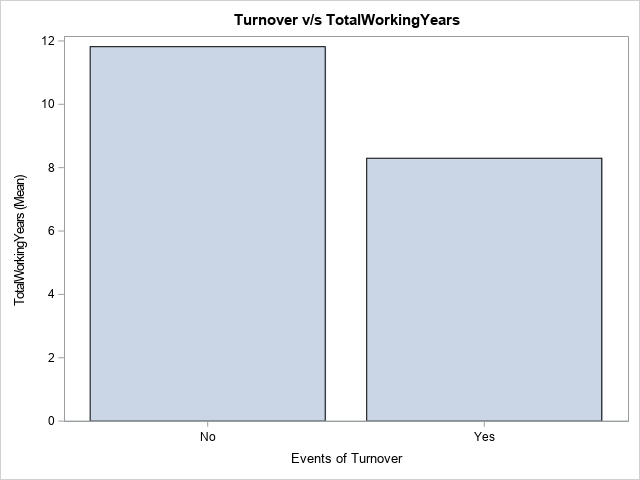
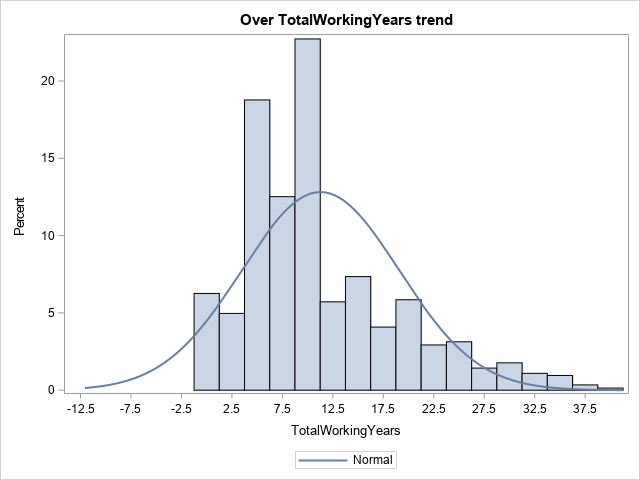


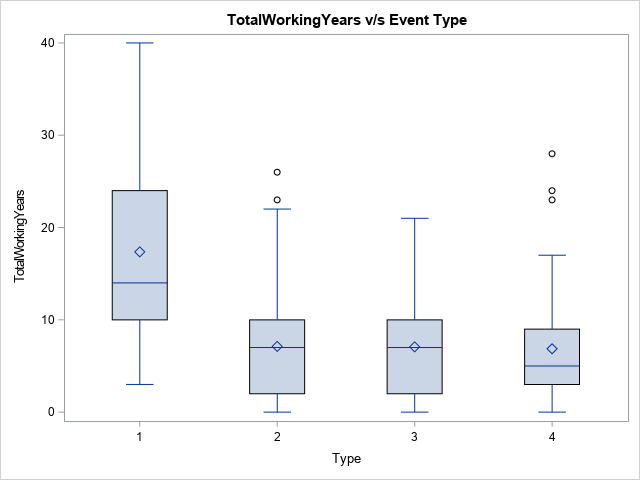
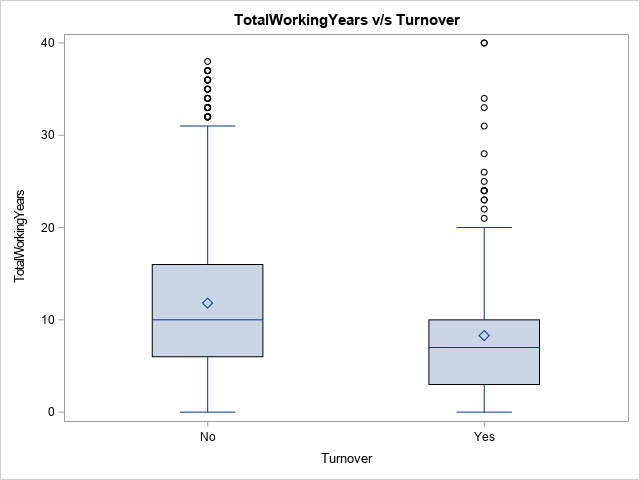


Observations:

* ChiSQ test is significant and suggests that Turnover is dependent on Stock Option level.
* StockOptionLevel 0 has maximum turnover (YES)
* Type 2 (Voluntary Resignation) has maximum frequency for StockOptionLevel 0
* We can refer the group as "employees having stock option level 0" have tendency to leave company

*TotalWorkingYears*

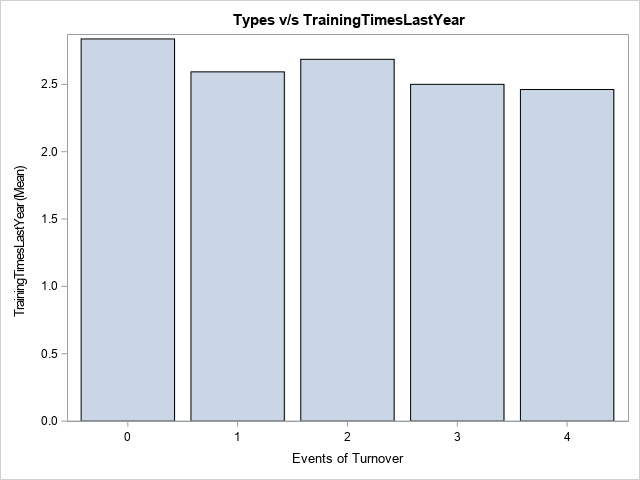
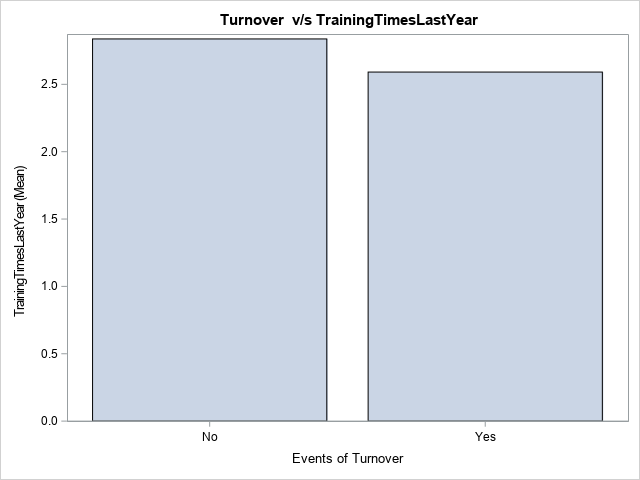
****



Observations:

* Overall Trend of TotalWorkingYrs shows that median is 10yrs
* Turnover No - Average TotalWorkingYrs approx 12 yrs
* Turnover YES- Average TotalWorkingYrs approx 8 yrs (includes all event types)
* Event Types 2 and 3 shows that 75% of of employees have TotalWorkingYrs <10 yrs are more likely to resign.

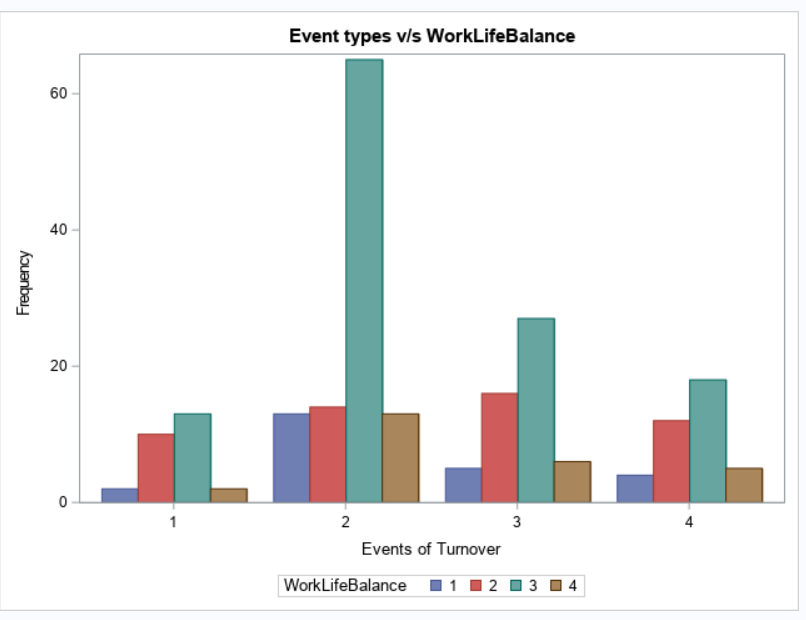
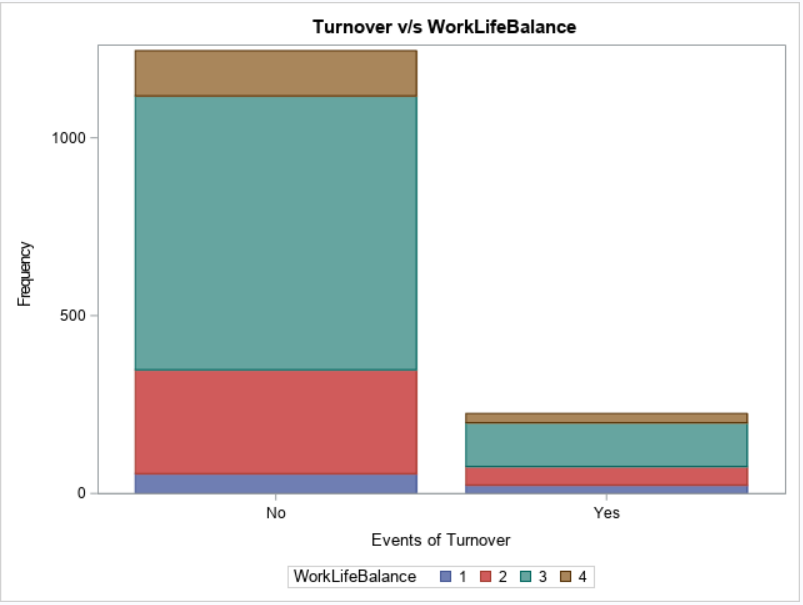
*TrainingTimesLastYear*

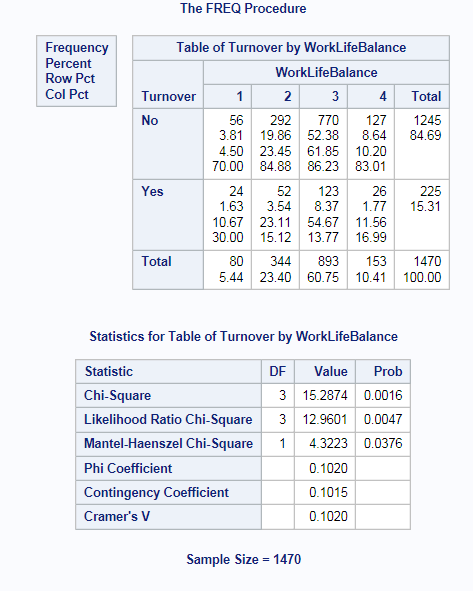


Observations:

Both Figs show that average trainings given is 2.5 to all employees, hence this variable does not seem to contribute in attrition.

*WorkLifeBalance*

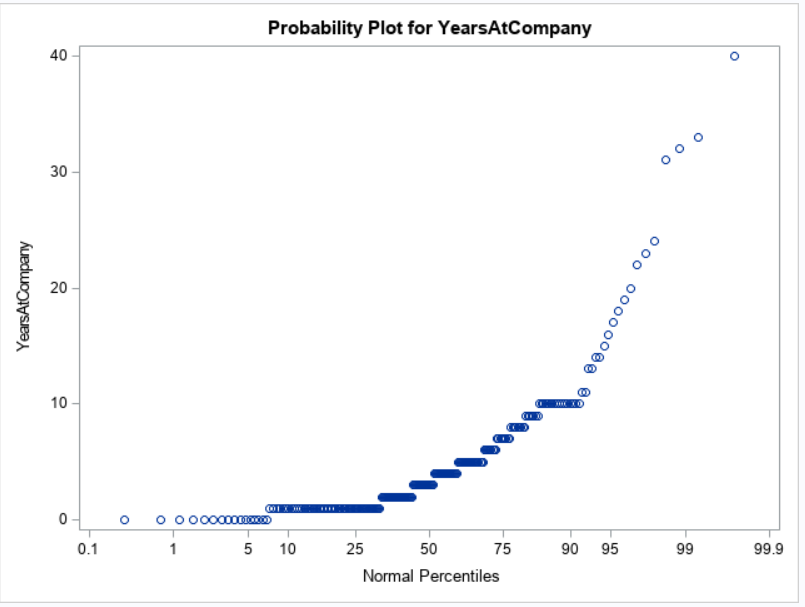
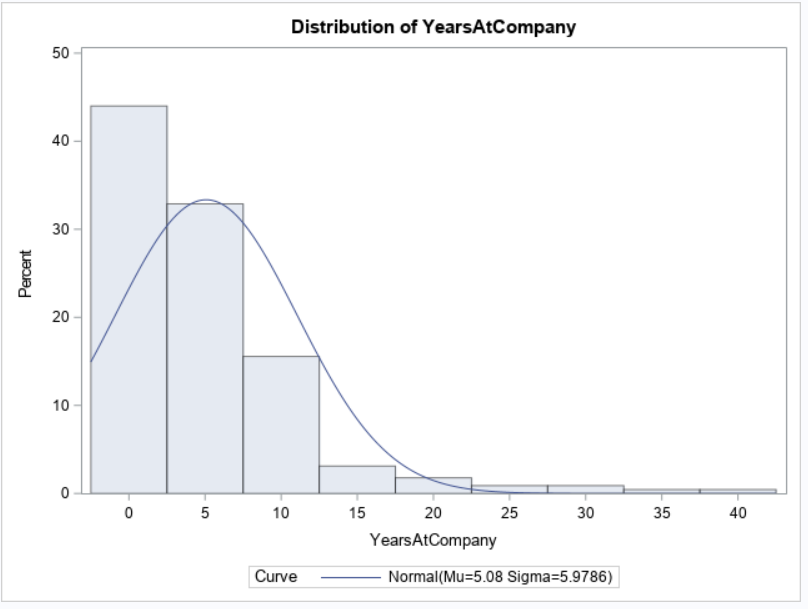


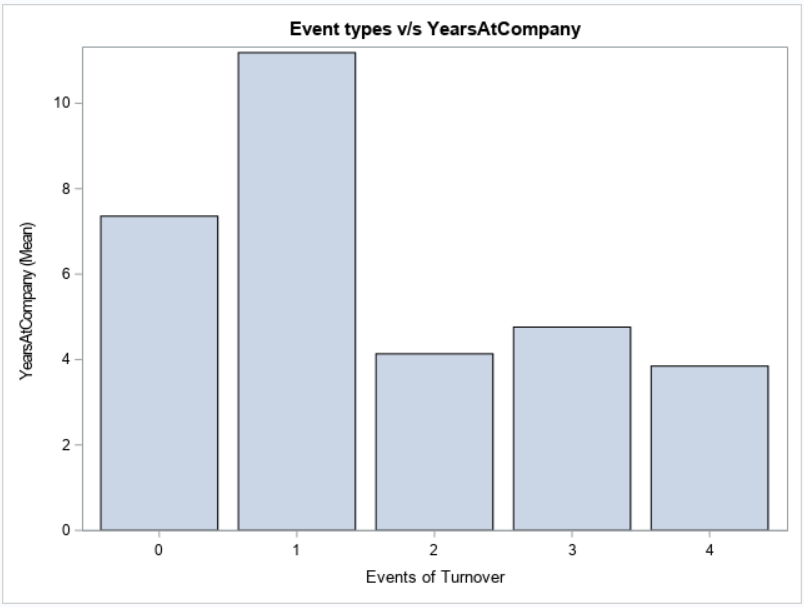


Observations:

* ChiSQ test is significant and suggests that Turnover is dependent on WorklifeBalance.
* Employees saying WorkLifeBalance as better (Category 3) are seen to be resigning.

*YearsAtCompany*

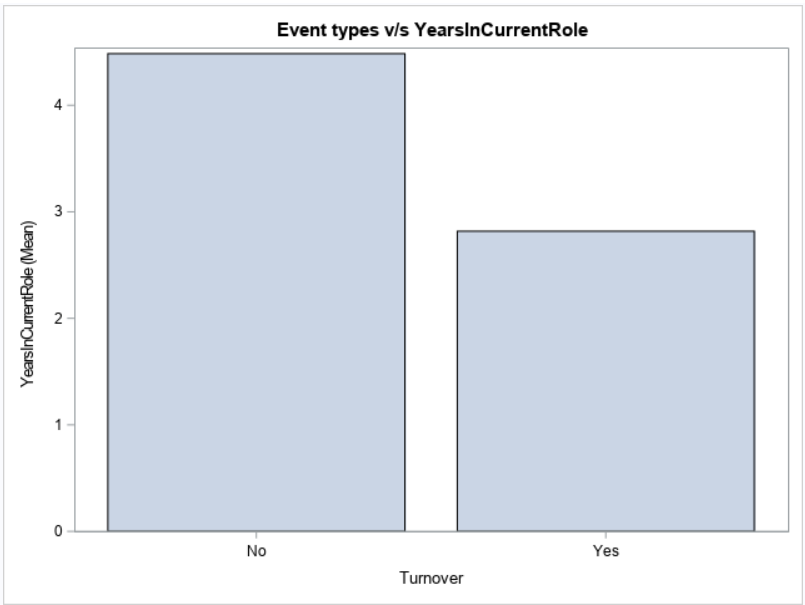
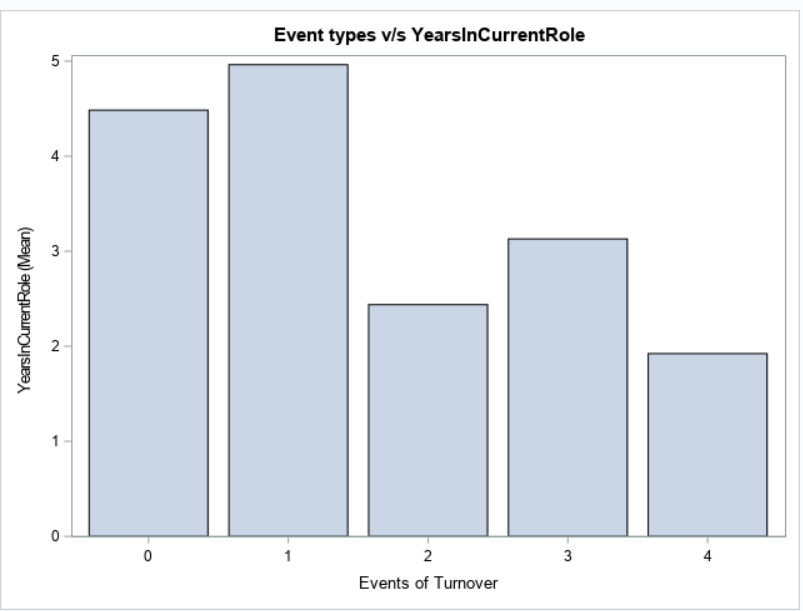




Observations:

* Average Employees working have tenure more than 6 yrs while Type 2,3,4 have approx. 4 yrs.
* Turnover = YES, skewed distribution with mean 5 yrs (All young employees to the company)
* Prob plot for employees for turnover = yes

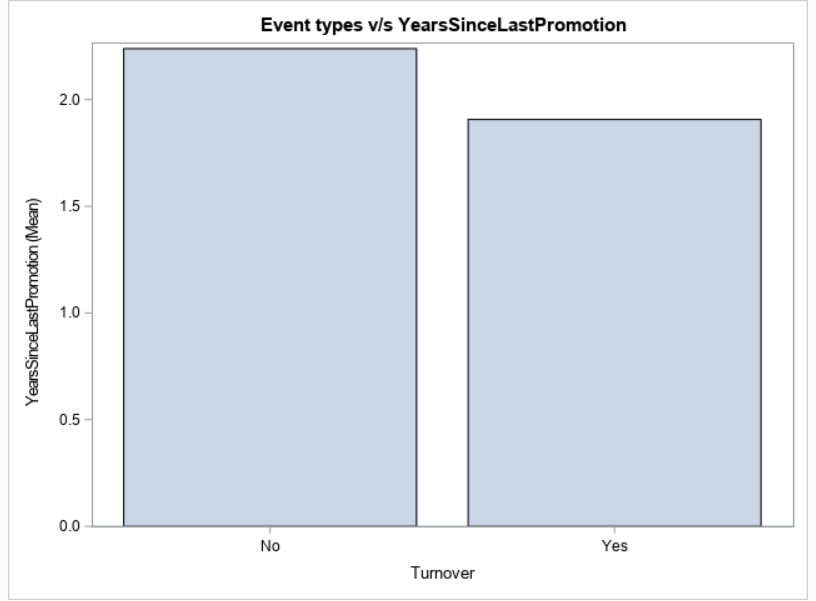
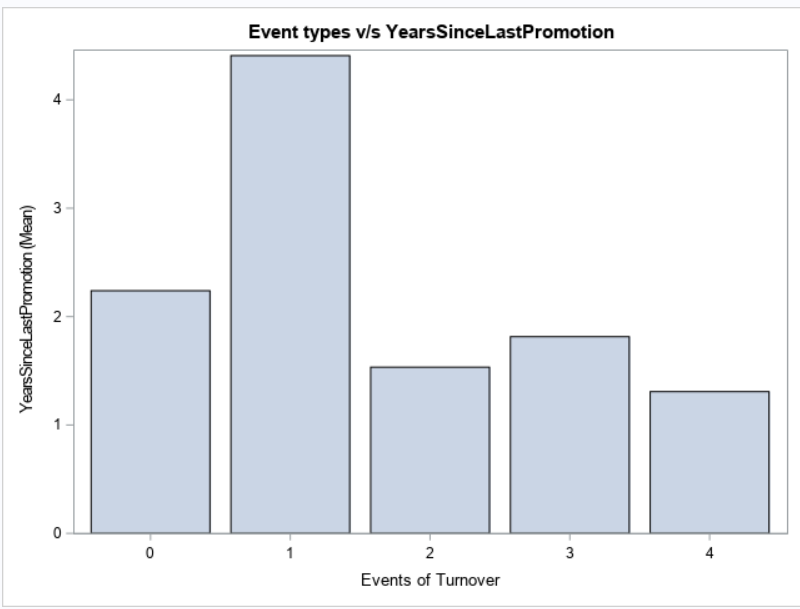
*YearsInCurrentRole*

Observations:

YearsInCurrentRole may be highly correlated to YearsAtCompany as seen in both plots.

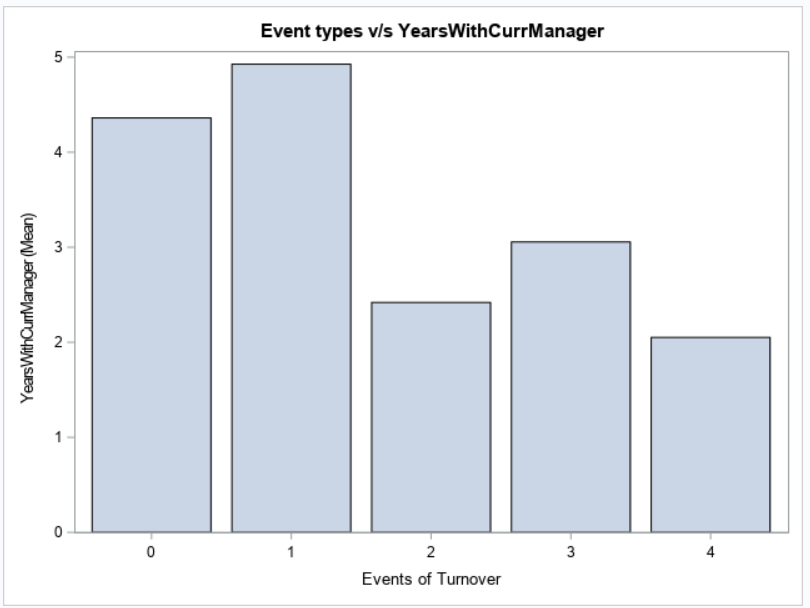
*YearsSinceLastPromotion*

Observations:

Employees with less than 2 yrs of being promoted likely to resign (Fig 2)

*YearsWithCurrentManager*



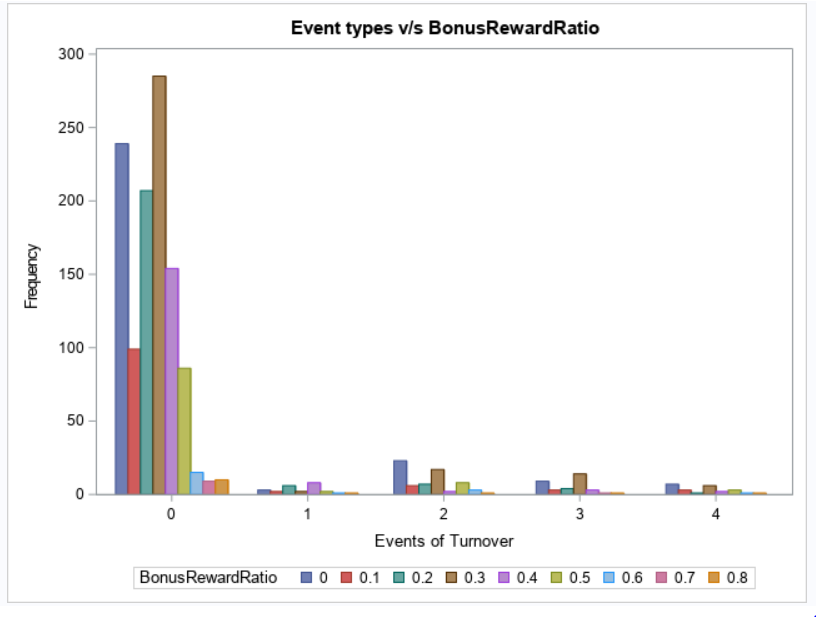
Observations:

YearsWithCurrentManager likely to be highly correlated with YearsInCurrentRole.

*Bonus Component*

All Bonus is added to find number of times an employee is awarded bonus.

This number is divided by **YearsAtCompany** to get a ratio consistent for all employees.



Observations:

If we consider Type 2,3,4 (Turnover =Yes), we can see that when bonus ratio is 30% or less, churning is more. We can focus more on employees who are still with company having bonus ratio less than 30%.

1. ***Assumptions***

There were some rows in the dataset for which the turnover has the value ‘NO’ even if they had a turnover type value other than 0. Hence, we have assumed this to be a mistake and corrected the records. Any records that had a turnover type value, we assigned the corresponding turnover column to ‘YES’ and records that had a value 0 for the turnover type column, we assigned the turnover column to ‘NO’. We have assumed turnover column to be the attrition rate. Hence, if turnover is 1, it means that the person has left the company and turnover of 0 means the person did not leave the company. We have chosen our target variable to be “Type”.

1. ***Identification of Insignificant Variables***

After we performed the exploratory data analysis (EDA), we noticed that there are few variables which do not contribute anything significant to the model. Hence, we decided to remove such insignificant variables. A list of the variables which were termed as unimportant by us are listed below.

|  |  |
| --- | --- |
| **Variable** | **Reason for removal** |
| EmployeeCount | It has a numeric value of 1 for all the rows |
| Over18 | It has a character value of ‘Y’ for all the rows |

1. ***Identification of Significant Variables***

After performing EDA, we noticed that there are few variables which are very significant and directly contribute towards the target variable, **Type**. A list of such variables are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| DistanceFromHome | Job Involvement | Marital Status | YearsatCompany |
| Daily Rate | Job Level | Stock Option Level | YearsSinceLastPromotion |
| Environment Satisfaction | Job Role | Total Working Years |  |
| Gender | Job Satisfaction | WorkLifeBalance |  |

1. ***Feature Engineering***

We derived 5 new variables based on the existing variables to enable us to understand the effects of these variables on the target variable in a better way.

|  |  |  |
| --- | --- | --- |
| Parent Variable | Derived Variable | Range |
| Age | AgeBucket | 1. Young **IF** Age <= 25 2. Mid-age **IF** 25 < Age <= 40 3. Old **IF** Age >40 |
| Monthly Income | IncomeBucket | 1. Low Pay **IF** MonthlyIncome <= 5000 2. Medium Pay **IF** 5000 < MonthlyIncome <= 10000 3. High Pay **IF** MonthlyIncome >10000 |
| DistanceFromHome | DistanceFromHomeBucket | 1. Near to Office **IF** DistanceFromHome <= 11 2. Very Far, all others |
| 40 bonus columns | YearsBonusRewarded | Summation of all the 40 bonus columns |
| YearsatCompany | Employee Type | 1. Junior **IF** YearsAtCompany < 5 2. Experienced, all others |
| Based on bonus column | BonusRewardRatio | YearsBonusRewarded/YearsAtCompany |

1. ***Data Modelling***

**Is there any difference in attrition between different employee groups or categories?**

To understand the difference in attrition, we used the Kaplan Meier analysis to compute the probability of employee turnover as a function of time and compare differences in survival probabilities for different employee groups/categories.

Once we gained useful insights on the survival probability of certain group, we performed the test of equality over strata.

Hypotheses:

Ho=> the risk of the groups are equal

Ha=> the risk of the groups are not equal

Default methods in SAS are:

* + Likelihood Ratio Test (-2LOG(LR))**\***
  + Log-Rank Test
  + Wilcoxon Test

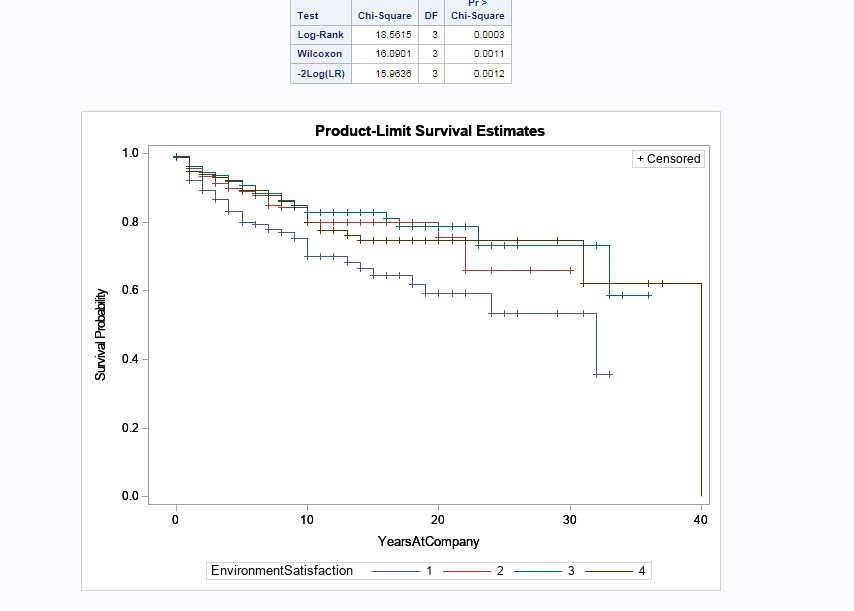
If pvalue was < 0.05, we rejected the null hypothesis. Based on this, our conclusion would suggest that that the survival experiences for all types of turnover are not equal.

**Insights**

**Note:** In most cases, the risk of employee groups were not equal

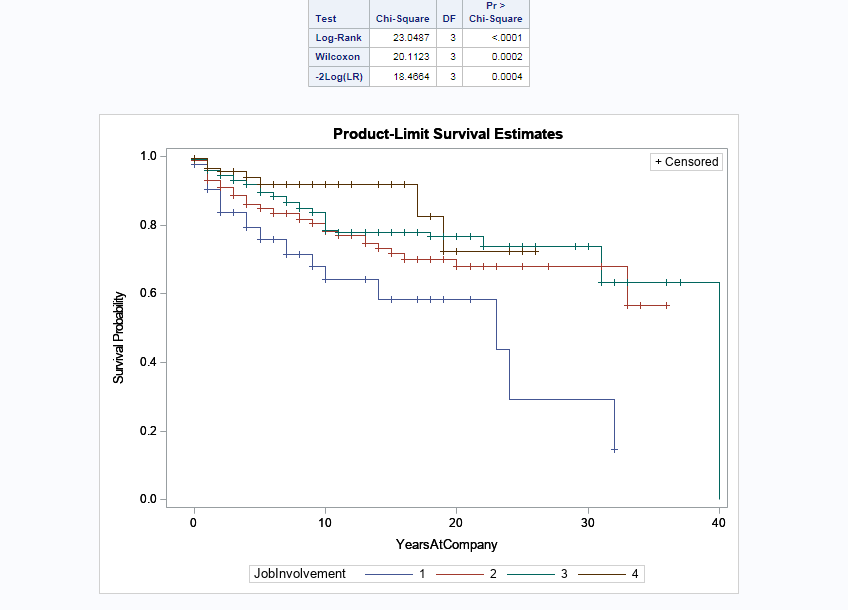
1. **Environment Satisfaction:** a score showing how much the employee is satisfied with company's facilities where 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

Employees satisfaction decreased as they continued to stay in the firm and one who had rated higher levels (3 or 4) had a better chance at survival.



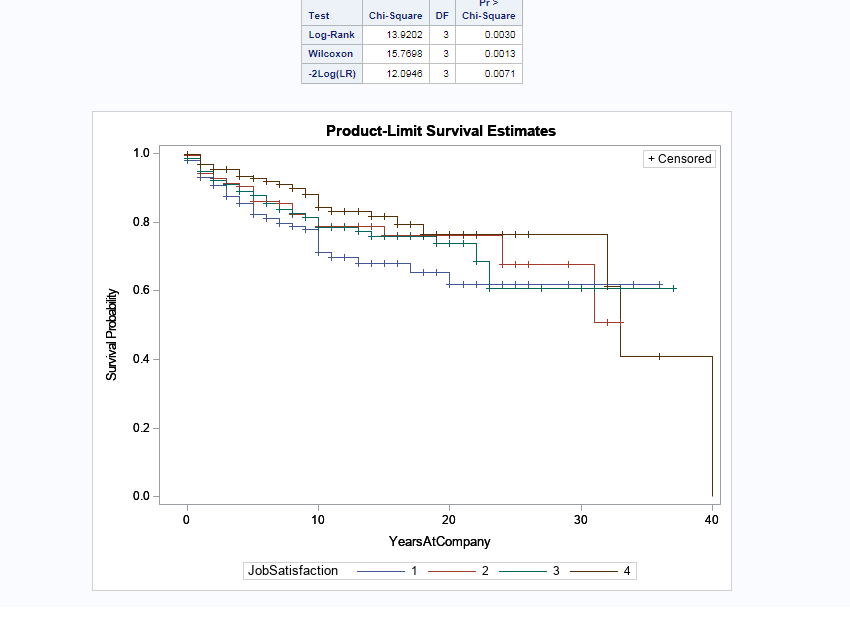
1. **Job Involvement:** a score given to the employee by supervisors how much the employee is involved in company's operations where 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

Employee attrition is directly proportional to how involved one is in the firm. Ones who rated 4 had a steady job involvement from years 5-18.



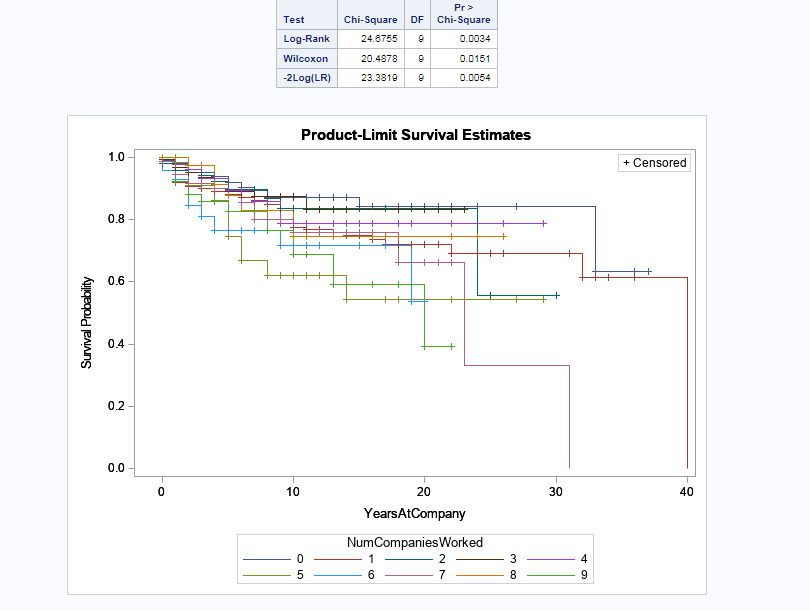
1. **Job Satisfaction:** shows the last survey result of the employee about his\her job satisfaction

An average of 60% population leave the company within 20 years. Since the job satisfaction levels vary, other reasons in conjunction with job satisfaction would give us better insights.



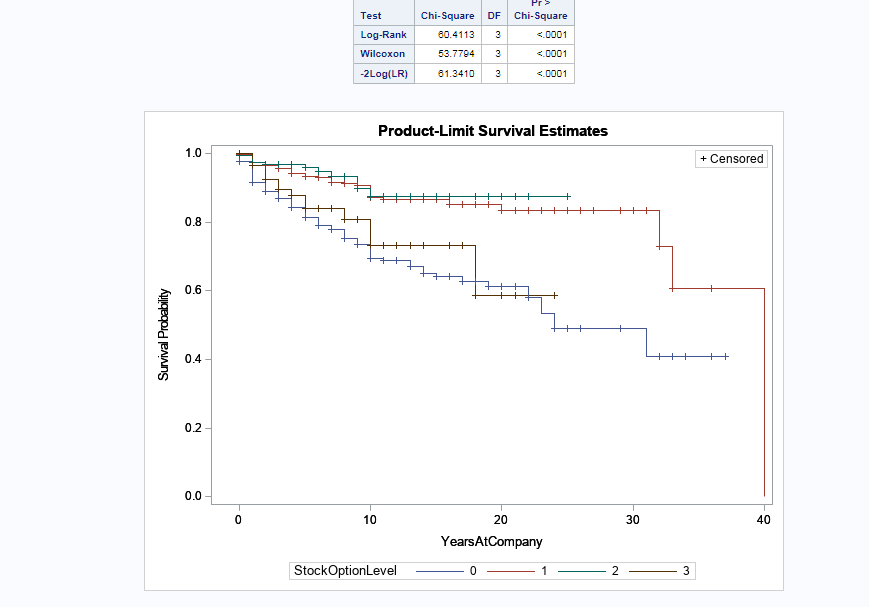
1. **Number of companies worked:** the number of companies the employee worked before starting in the company

Employees worked anywhere between 0 to 9 companies prior to joining this company. Young employees who worked between 0-2 years have a higher probability of surviving in comparison to other.



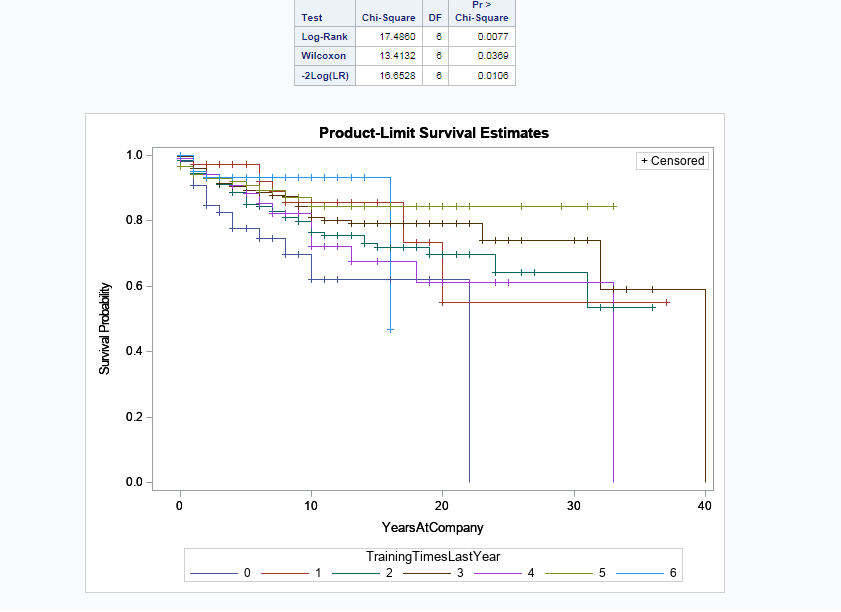
1. **Stock Option Level:** shows the stock option for the employee.

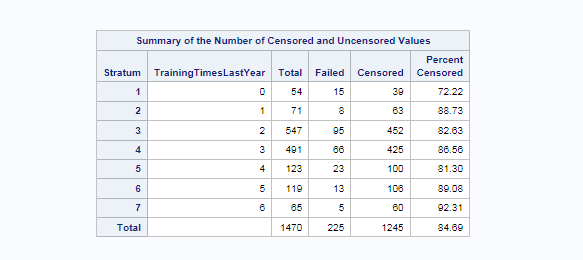
Employees who had a decent amount of stock options (1 or 2) had a higher chance of survival versus those didn’t have any stock options at all or those who had a huge amount of stocks. The huge amount of stock options may not have benefited the employee since they didn’t gain much with.



1. **Training Times Last Year:** shows the number of training programs employee has attended last year

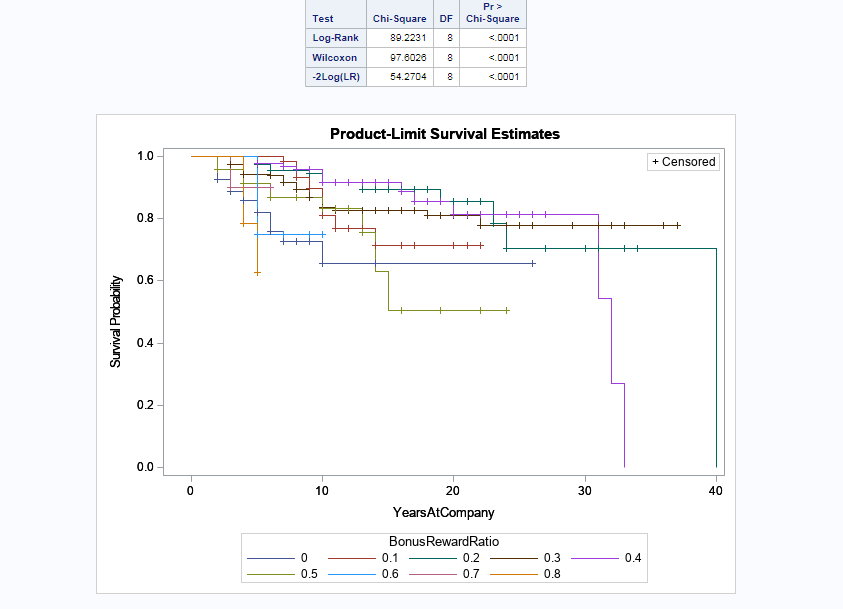
The employees who did not have any training the last year were the largest population leaving the firm.

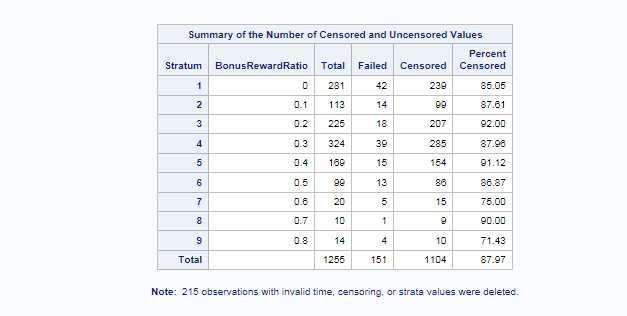




1. **Bonus Reward Ratio:** It is the ratio of Number of Years Bonus Rewarded to the Number of Years At Company

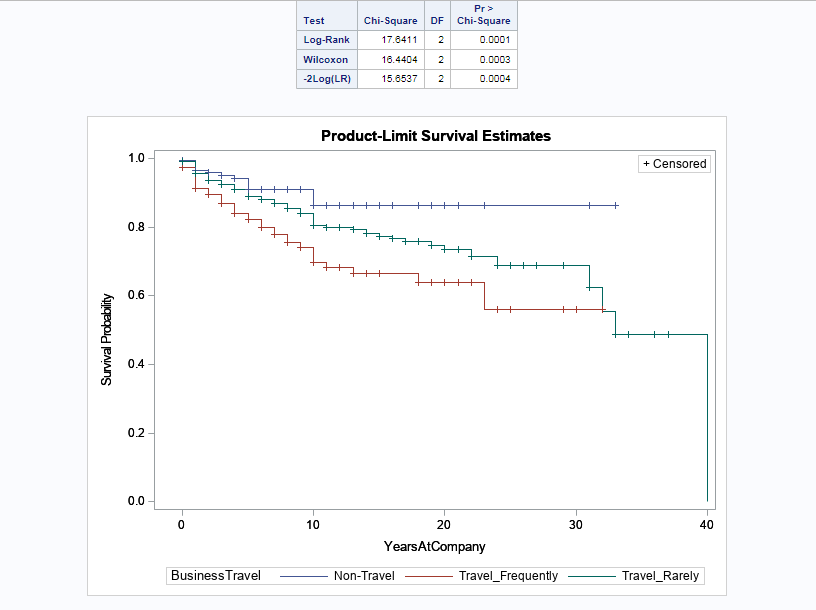
To use an example, those employees who stayed 10 years and were rewarded 8 times were the majority who left the company. This tell us that the bonus had to be re worked in order for the firm to lose their employees and money.





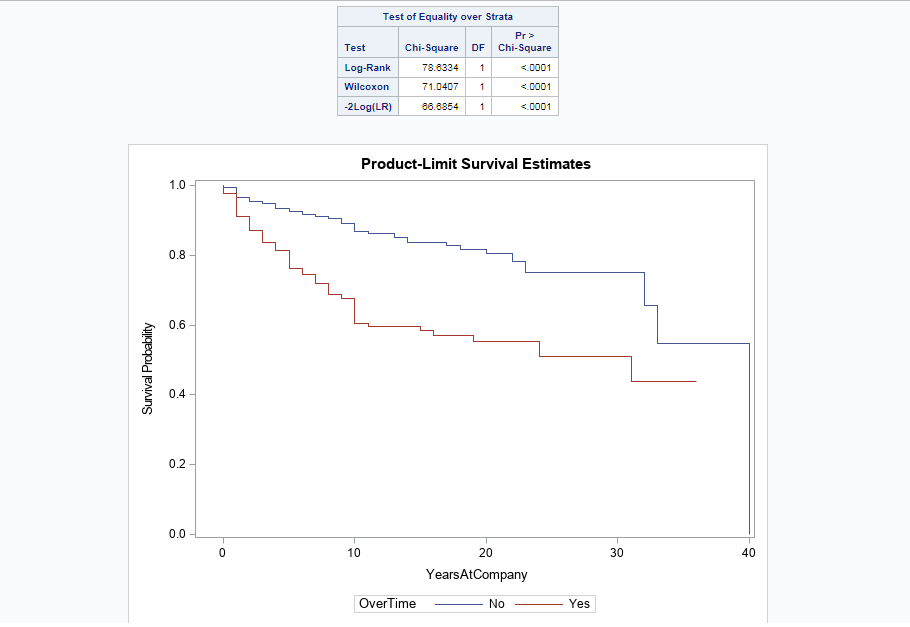
1. **Business Travel:** shows how much employees travel on work

An employee who doesn’t travel have a higher chance of survival, followed by travelling rarely and last travelling frequently. In our graph, the ones who travelled frequently were the majority who left the firm.



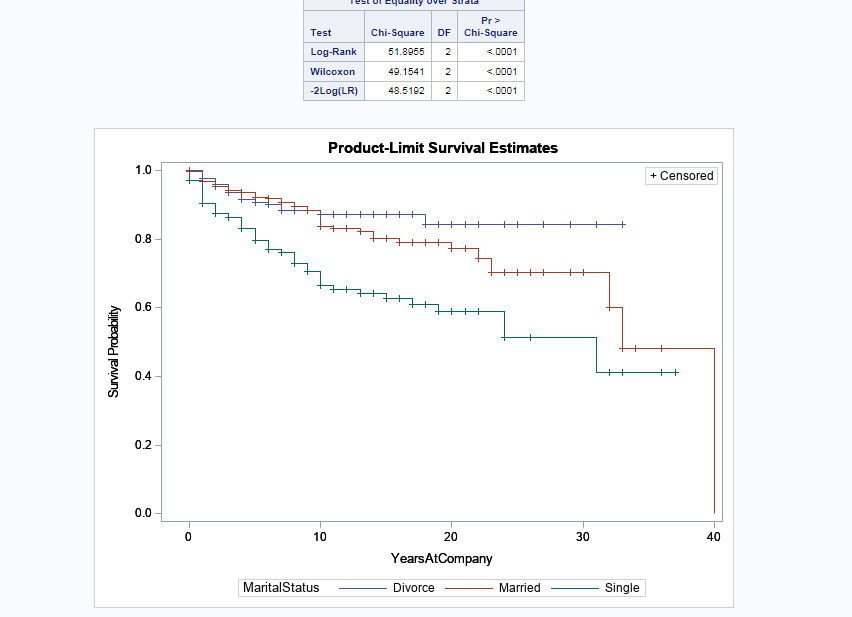
1. **Over Time:** shows whether employee works overtime more than 10 hours a week

The employees who do not work overtime have a better chance of survival over the ones who do.



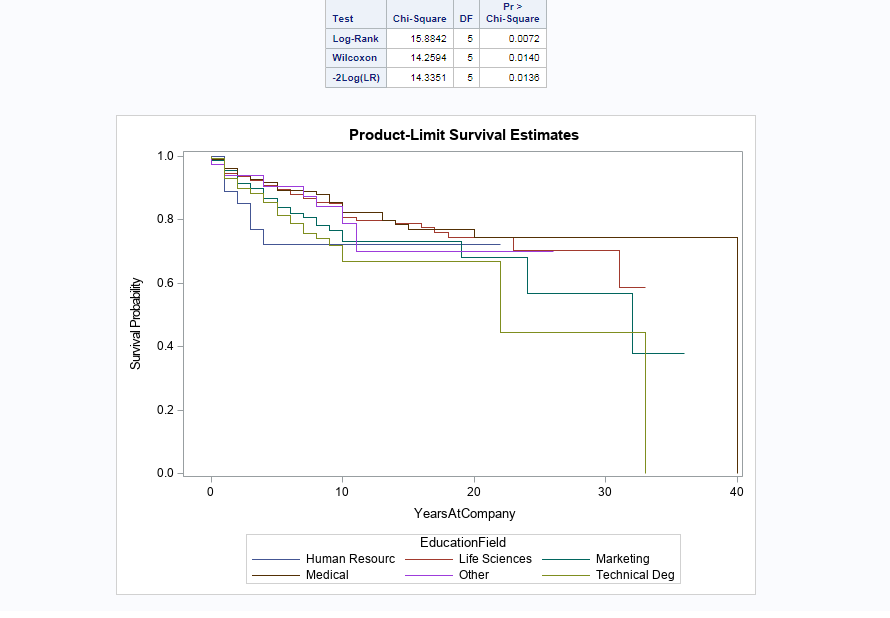
**13) Marital Status:** shows the marital status of the employee

About 58% of all employees irrespective of marital status decide to leave the firm within 23-24 years. Although, employees who are single tend to leave the firm sooner than other groups.



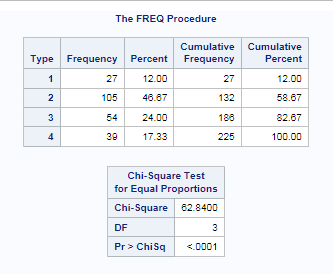
1. **Education Field:** shows the education field of the employee

Employees who have a technical degree tend to leave the firm sooner than others while those with a medical background have a higher chance of surviving.



**Can I combine different event types together? Or do all need to be handled separately?**

1. **Is the type-specific hazard function same for all event types?**



0 - No turnover

1 – Retirement

2 - Voluntary Resignation

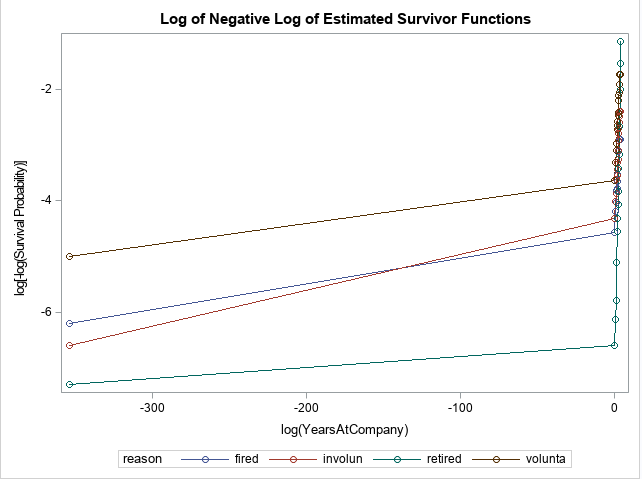
3 - Involuntary Resignation (Health problems, family matters etc.)

4 - Job Termination, Employee is Fired

From the table, the hazard rate is not same for all event types.

1. **Are the type-hazards linearly related with h(t)?**

If the hazard for one event type is linearly related to another, the graph shown below would have event types in parallel. This is not the case for ours.

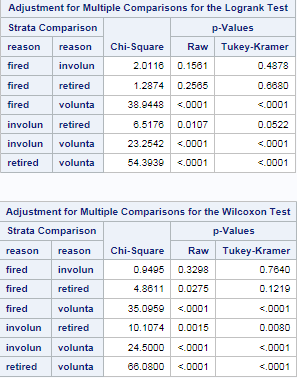


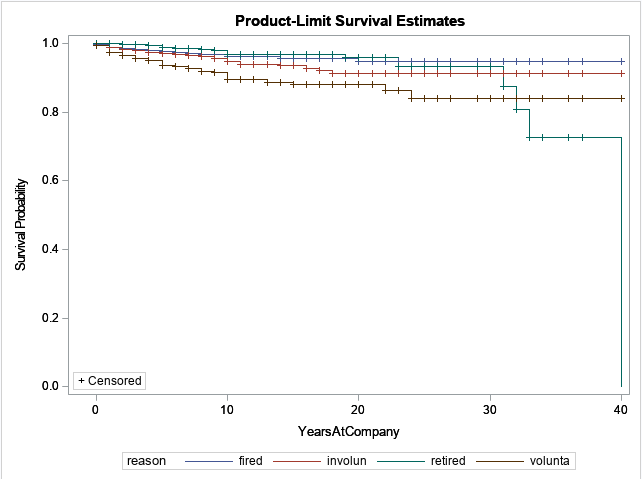
The next two screenshots test and show whether the survival of different event types are same.

From the first table, it can be observed that employees who were fired and those who left involuntarily have the same survival rate.

In addition, the employees who were fired and those who retired also have the same survival rate.

This can be observed graphically in the second screenshot.





1. **Use of log-ratio tests to determine whether we need to use separate models for each event type or combine the event types**

We need to understand how much of the variability is explained with the individual models over the nested (combined) model.

* Since the p-values is 0 which is less than 0.05, we conclude that all the event types are different.
* The difference between the combined and individual model is a positive value which means the individual models explain more than the other. (Remember the value is -2 Log L).

In conclusion, we need to handle all event types separately.



**Who is leaving the company?**

**Why are they leaving?**

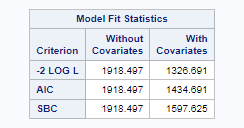
Initially to identify the effect of the covariates on the hazard rate, PROC PHREG step was executed. Years At Company was taken as the duration variable and Type as Censored variable with 0 as the censored value.

The following covariates were created from the time-dependent covariate bonus using programming statement-

1. PreviousYearBonus = bonus\_[YearsAtCompany – 1]
2. YearsBonusRewarded\_cal = sum(of bonus\_1 - bonus\_40);
3. BonusRewardRatio\_cal = YearsBonusRewarded\_cal/YearsAtCompany;

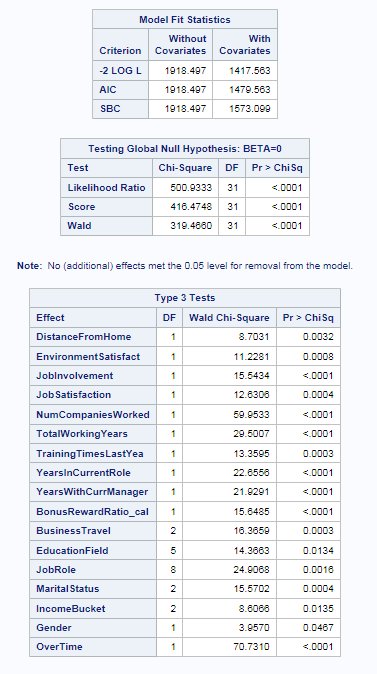
“EFRON” Method was used for Ties Handling.

The modle fit statistic is as follows-

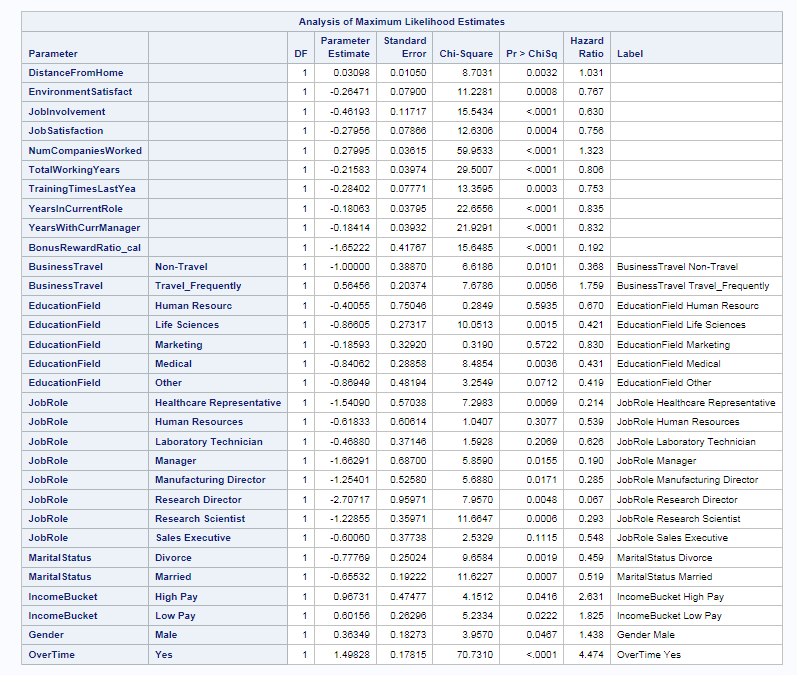


Then, Backward selection was done at 0.05 significance level to retain only the significant variables.

The Wald Tests show that the estimate of at leats one of the covariate is not equal to zero.



We see that the SBC value has reduced. 17 covariates were identified to be the most significant variable. Their effect on the hazard ratio is as seen below:



**What attributes increase/decrease the hazard rates for certain event types?**

**Does bonus affect employee turnover? If yes, how?**

We see that the following covariates **increases** the hazard of turnover:

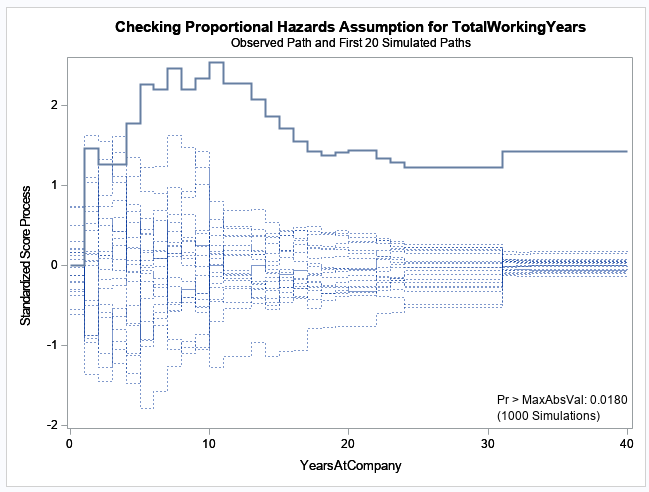
1. As **Distance from Home** increases, hazard increases by 3.1%
2. As **Num Companies worked** increases, hazard increases by ~32%
3. If an employee asked to **travel frequently**, the hazard of turnover increases by ~76% while if they are not asked to travel, the hazard reduces by 63%
4. Hazard for **male employees** increases by ~44%
5. Hazard increasesby ~47% when employees **work overtime**
6. Hazard increases for both **high and low paid** employees. The hazard of losing a highly paid employee is 163% more than an average paid employee

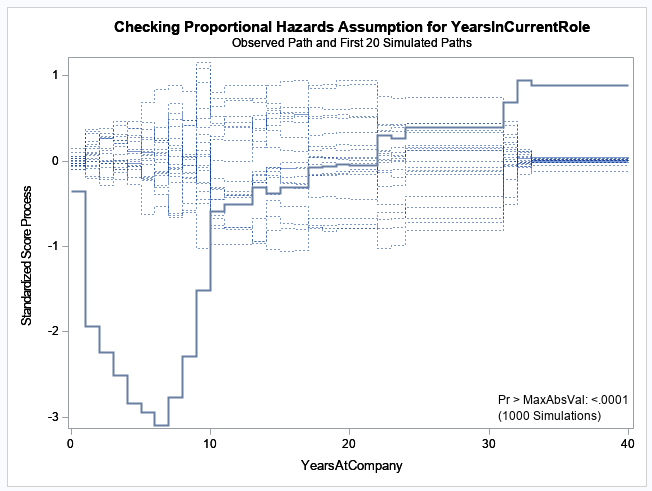
The below covariates **reduces** the hazard:

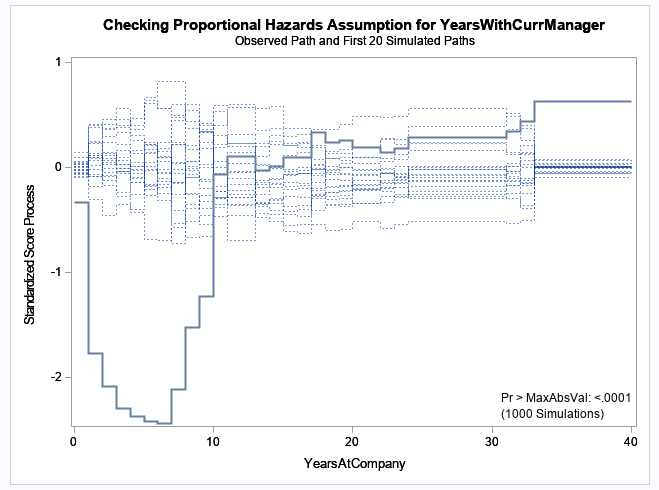
1. **As more bonus is given, hazard reduces by ~80%. Hence we could infer that to prevent employees from leaving the employee, bonus needs to be given at increased intervals, possibly every year if the performance of the employee is good**
2. As **Environment Satisfaction** increases, hazard reduces by ~23%
3. As **Job Involvement** increases, hazard reduces by 37%
4. As **Job Satisfaction** increases, hazard reduces by ~24%
5. As **Total Working Years** increases, hazard reduces by ~20%
6. As **Training Times** in the last year increases, hazard reduces by ~25%
7. As **Years in Current Role** increases, hazard reduces by 26.5%
8. As **Years with Current Manager** increases, hazard reduces by ~27%
9. We see that the hazard of turnover for employees in **Marketing and Human Resources** Education Field is reduced by ~17% and ~23% respectively
10. Hazard reduces with employees who are in **executive positions** like for a Research Director the hazard reduces by ~93%
11. The hazard is reduced for Married or divorced employees

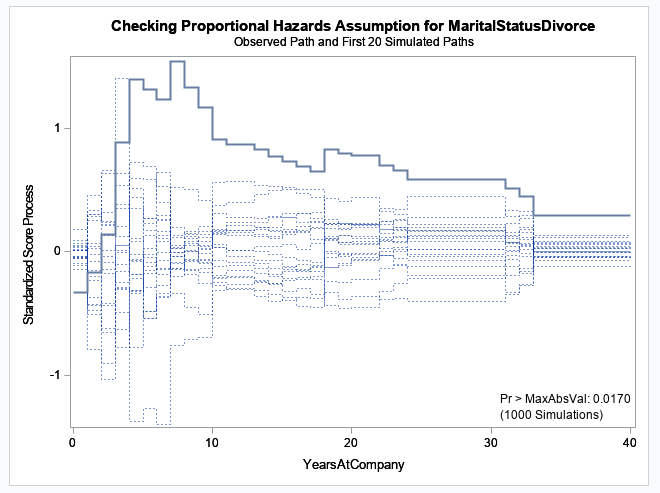
**Are there any variables which affect hazards non-proportionally?**

In order to determine variables which violate the proportionality assumption, the time-dependent covariates derived from bonus variables were neglected and ASSESS statement was used to analyse the other covariates-





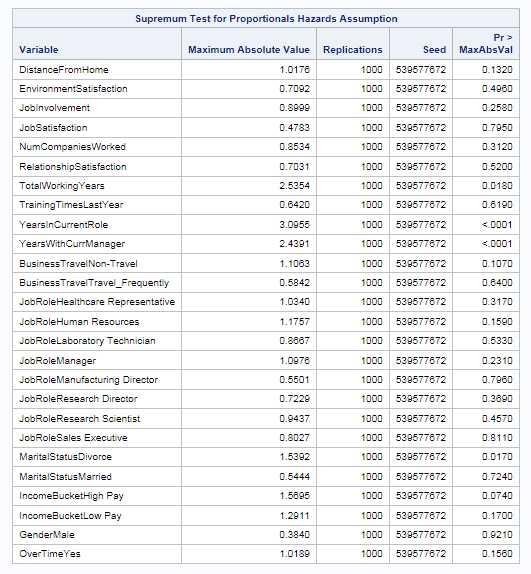




Thus, we see that the above covariates-

* Years in current role
* Years with current manager
* Marital Status (Divorce)
* Total Working Years

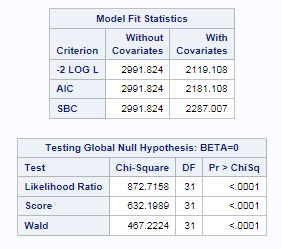
are correlated with the duration variable which caues the non-proportionality which is also supported by the below table-

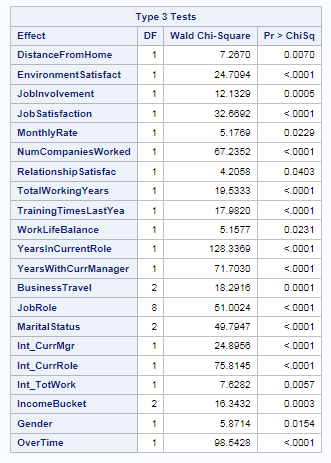


Initially, interaction variables were introduced as follows-

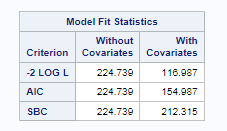
* Int\_CurrMgr = YearsWithCurrManager \* YearsAtCompany;
* Int\_CurrRole = YearsInCurrentRole \* YearsAtCompany;
* Int\_TotWork = TotalWorkingYears \* YearsAtCompany;
* Int\_MarStat = MaritalStatus\_LE \* YearsAtCompany;

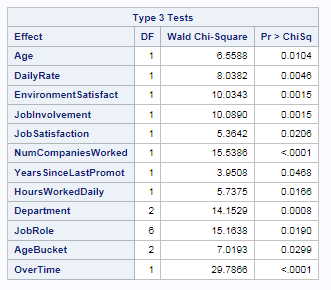
Where MaritalStatusLE is the label encoded variable of Marital Status covariate. The interaction variables Int\_CurrMgr, Int\_CurrRole and Int\_TotWork were still determined to be significant and the fit statistics after including the interaction variable were as follows-



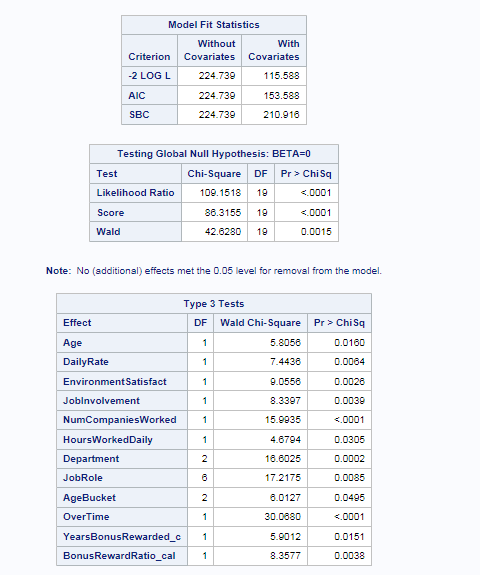


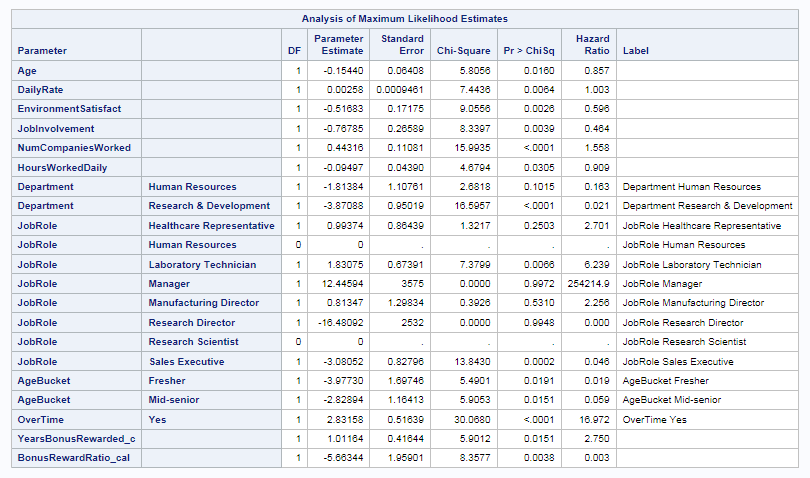
When the time dependent covariates were added as STRATA variables- the below fit statistic was obtained.





When the time-dependent covariates derived from bonus variable were also added to the above model, the fit statistic were as follows-

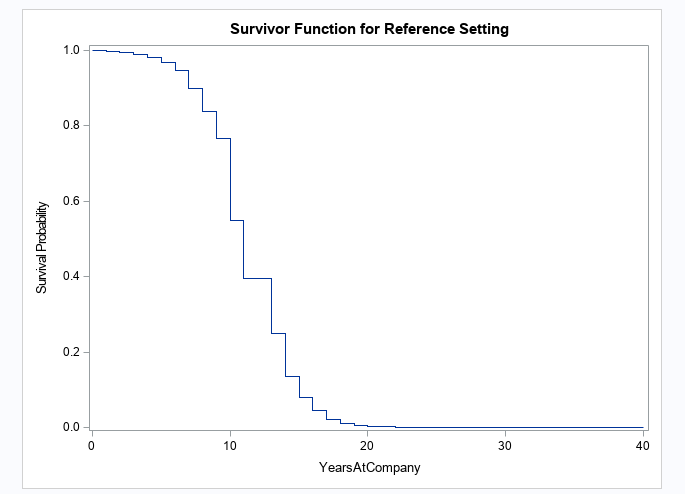




**We see that the BIC and the SBC values have reduces significantly showing that the model has been fitted well.**

**Survival Plot-**

The Survival plot shows that the survival probability reduces in a steep manner from when the tenure is around 8 years:

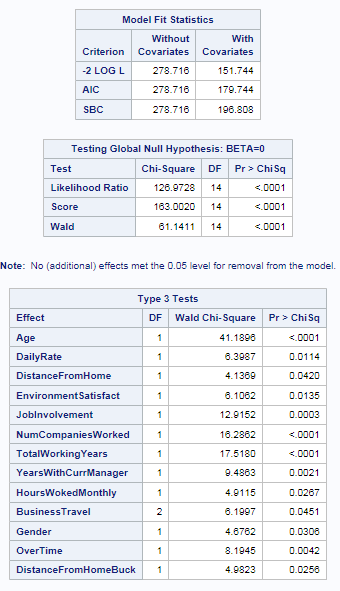


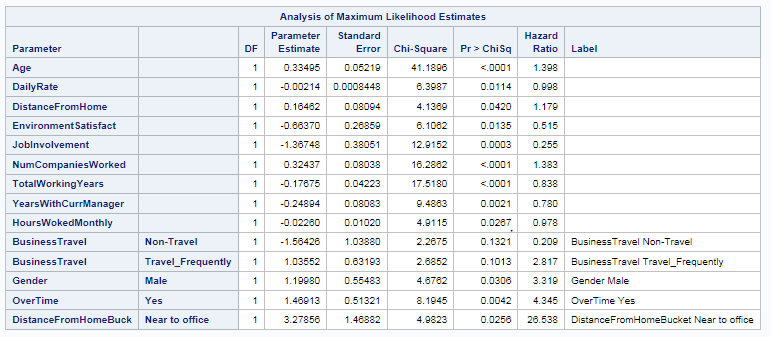
**Separate models for each Turnover Type -**

We have determined from the competing risk tests that separate model has to be built for each turnover type.

*Turnover Type 1: Retirement Model-*

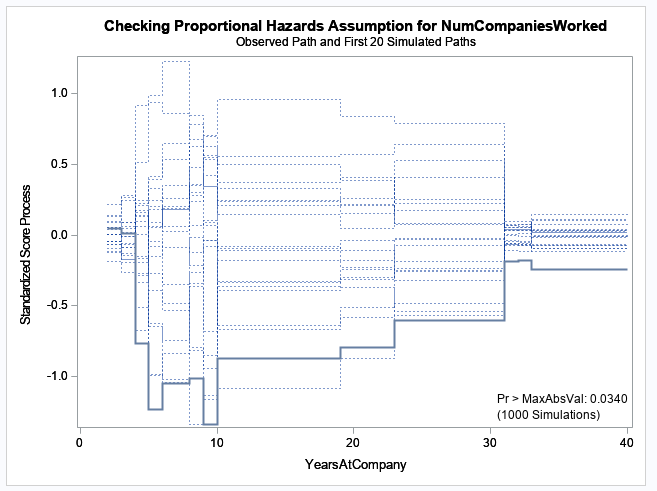
The model fit statistics and the significant variables are as follows-

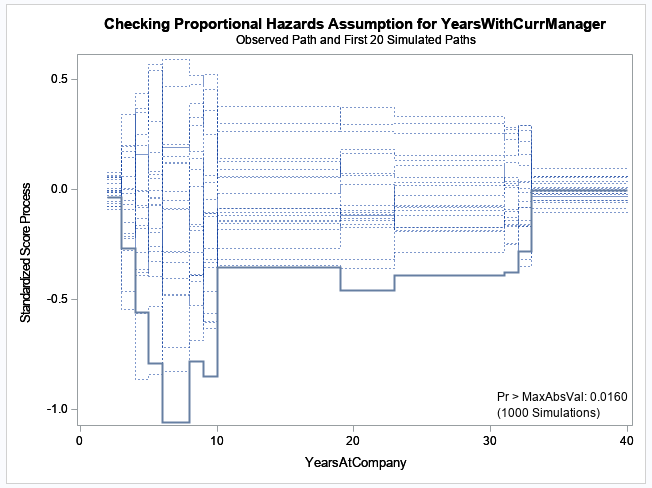


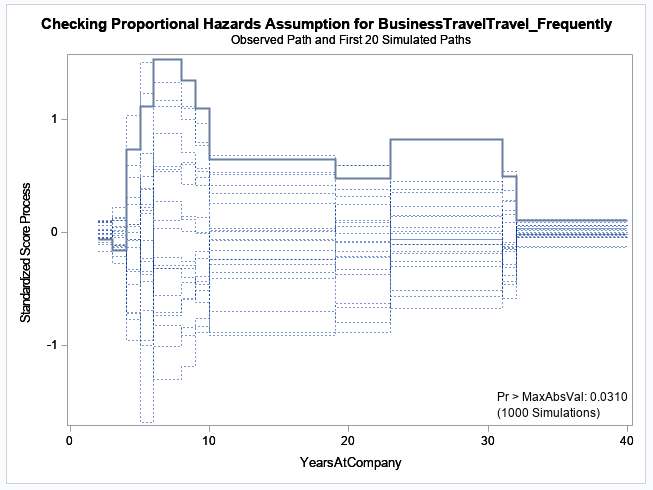


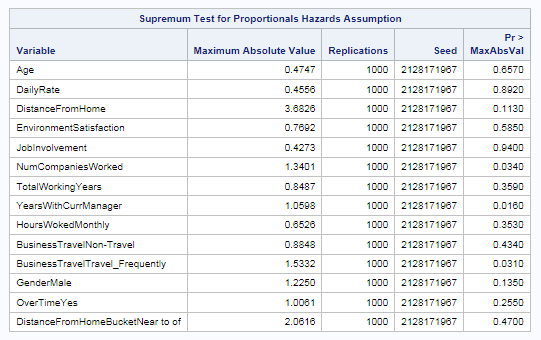
We see that the **hazard of retirement increases with age, distance from home, frequent business travels, overtime work** and the number of companies worked while the risk of retirement reduces with increased environment satisfaction and job involvement and with work involving no travel

The variables which cause non-proportionality are Number of Companies Worked, Years with Current Manager and Business Travel as shown below-

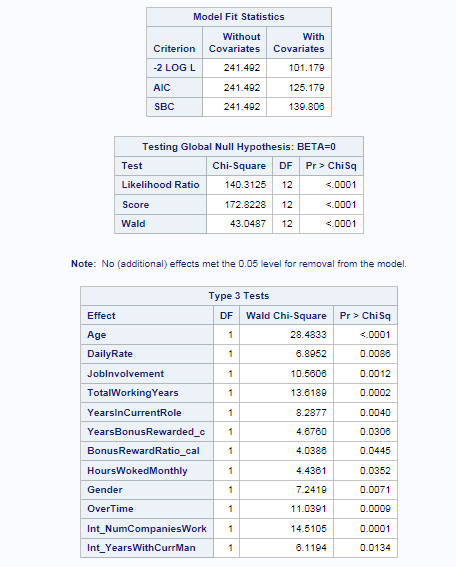


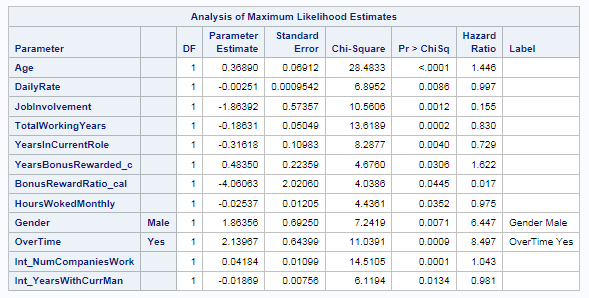




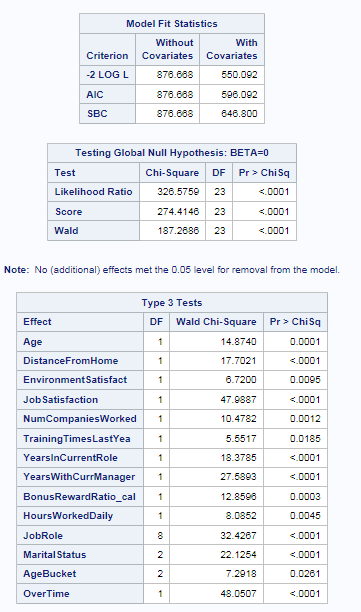


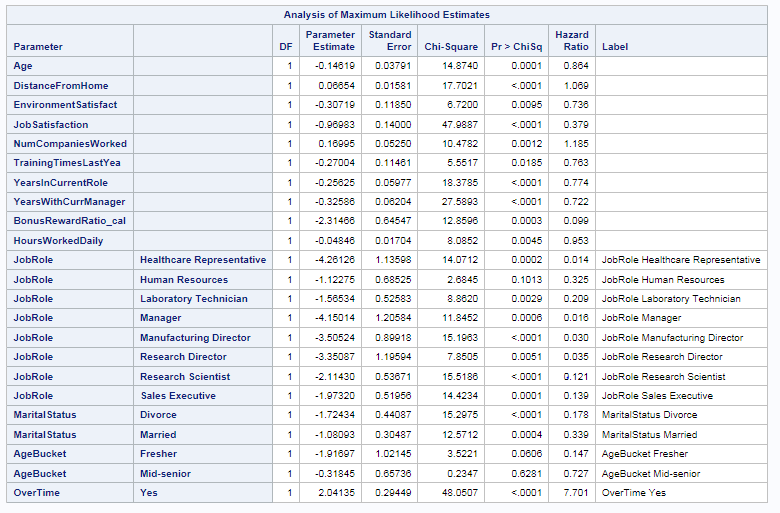
The continuous variables causing non-proportionality are added as interaction variables and the categorical Business Travel column is added as STRATA. We see that the interaction variables are still significant:





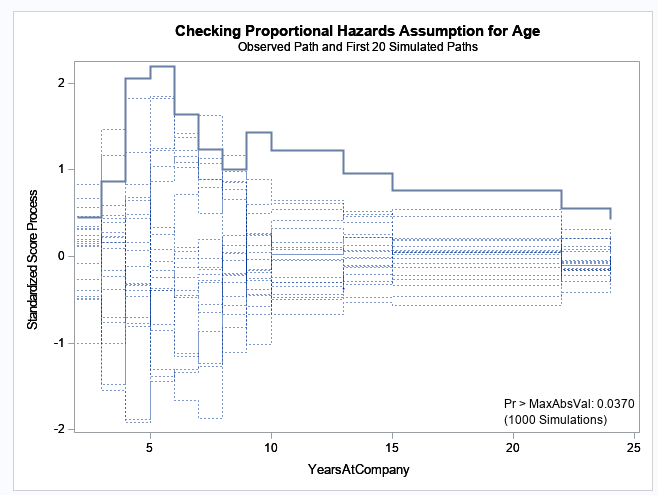
*Turnover type 2 – Voluntary Termination Model-*

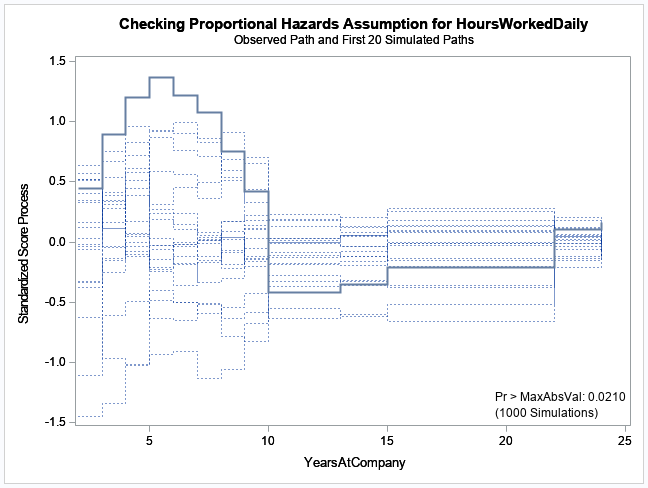


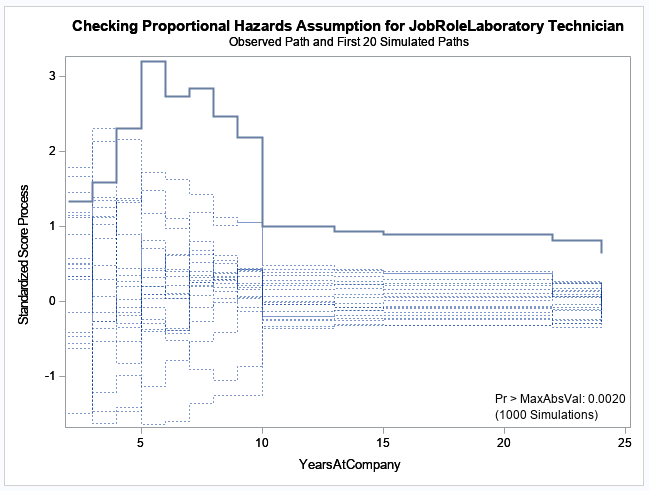


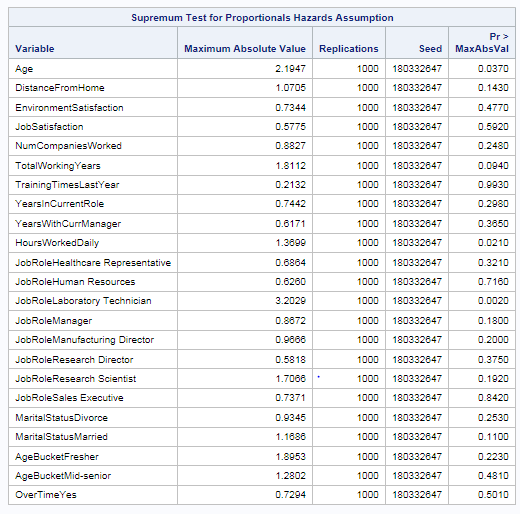
Similarly, we see that the hazard of voluntary termination increases with overtime work, greater number of companies worked and increased distance from home. The company could prevent employees from leaving by providing high bonus, great environment and job satisfaction. The company could employ a lot of freshers and mid-senior level employees who are married or divorced.

The following variables cause non-proportionality-

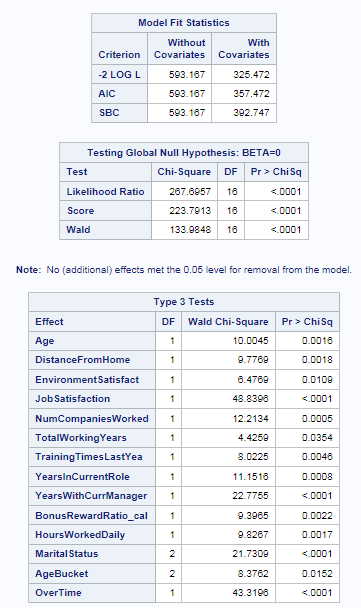


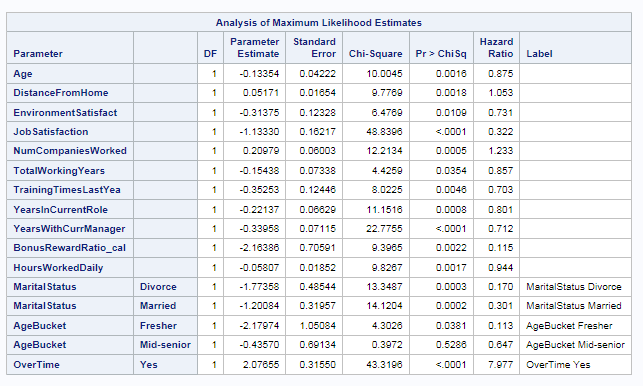




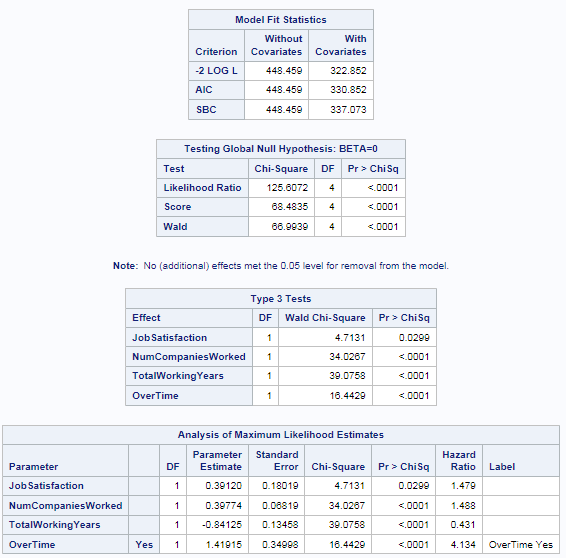


As seen before, the continuous variables age and Hours Worked Daily are added as interaction variables and the Job Role variable is added as STRATA statement. The interaction variables are not significant:





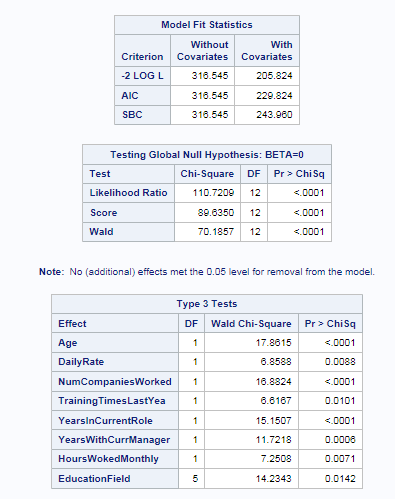
*Turnover type 3 – Involuntary Termination Model-*

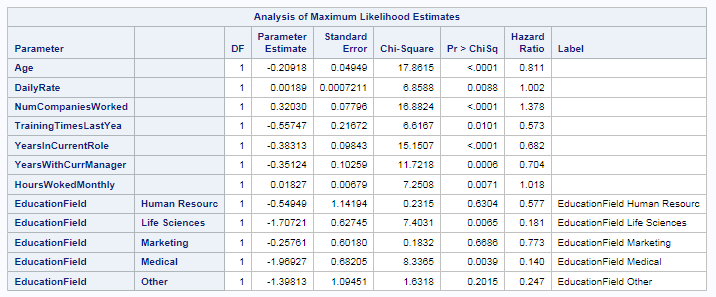


Hazard of involuntary termination increases greatly with overtime work though they are highly satisfied with the job. These employees have a history of having worked in many companies. Hazard of involuntary termination is less among experienced employees.

No variables caused non-proportionality in this model.

*Turnover type 4: Firing Model-*





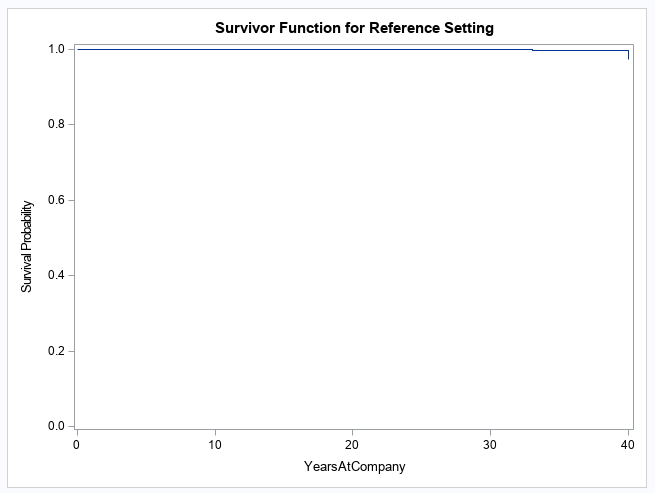
Medical, Life Sciences professionals and senior employees are less likely to get fired. More trainings and greater number of years in the current role or with current manager reduces the hazard of getting fired. Employees with increased hazard of getting fired have worked in greater number of companies.

No variables caused non-proportionality

**When is the biggest danger for employees to leave?**

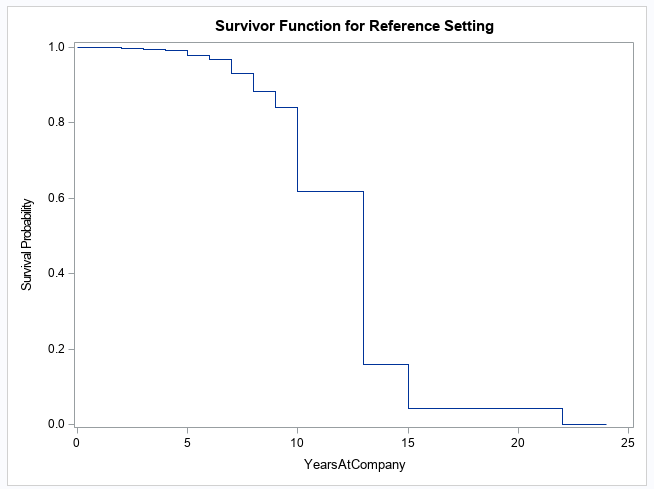
*Type 1 –*

The survival probability is almost 1 until 40 years at company for the retirement turnover type.



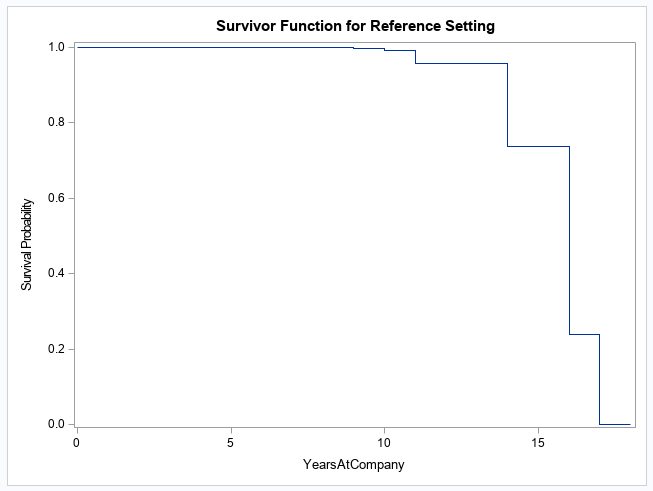
*Type 2-*

The survival probability for voluntary termination reduces increasingly from around 5 years at company to almost 0 at 24 years at the company.



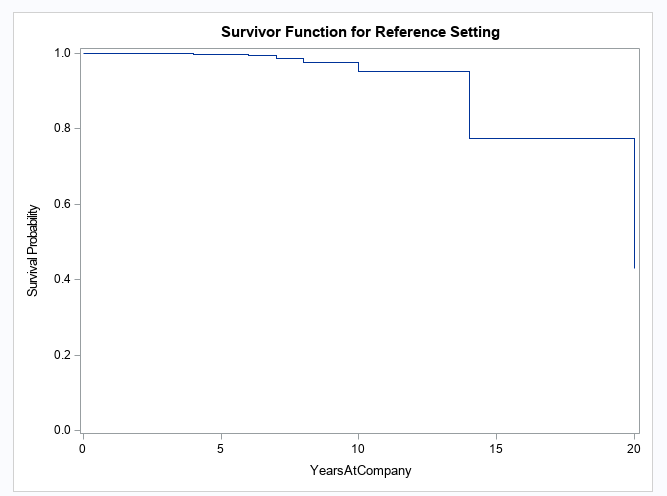
*Type 3-*

The survival probability for involuntary termination falls rapidly from 10 to 16 years at the company



*Type 4-*

The survival probability drops to 0.4 at 20 years at the company



1. ***Recommendations***

After analyzing, we found that there could be certain benefits or perks which can be offered to the employees.

* We observed that people who travel rarely churn more. Hence, more business travel opportunities must be given to people who belong to all types.
* Also, we observed that people who live far also churn more. Hence, if they are given allowances like gasoline allowances or other travel vouchers to come to office daily, it would be really helpful to them.
* We also observed that many people are leaving from the Research and Sales team. So, the company needs to figure out better strategies to retain the employees who belong to these departments.
* People tend to leave more when work for a long time in office. Hence care should be taken that experienced employees are not burdened a lot with work and it should be mutually shared among others.
* Also, employees who are jumpers i.e. people who switch more companies tend to churn more. Hence, such employees should be counselled to know their reasons for leaving the company and switching all the time.

1. ***Conclusion***

We performed a detailed analysis using univariate analysis method. Since the data was clean, we did not have to do a lot of cleaning steps other than removing certain redundant variables, correcting the turnover column. We also changed the values of the turnover column to logical values 0 and 1 so that it can be used by Survival models. Then we ran few survival models and presented our detailed results.

1. ***Appendix***

******