Group 1
Akram Yadullahi
Manas Jani
Mihir Sanghvi
Satinder Sondhi
Srikiran Rai
Trina Maitra

EMPLOYEE TURNOVER ANALYTICS

Project Report

Contents

Executive Summary

Business Objective

FermaLogis Inc. is a leading pharmaceutics company. As a leader in the industry, FermaLogis is an organization committed to its employees. With an 'Employee First' strategy, it believes in the collective power of its employees in its progressive growth. An organization committed to investing in the growth and development of its employees, FermaLogis is currently facing a high employee attrition problem, with its young and experienced employees. Despite having a strong training program and career investment in its employees, FermaLogis has been facing huge employee turnaround. The aim of our study is to perform a detailed statistical analysis, using the concepts of survival analytics, in order to understand the dynamics of the employees who are leaving the company and to understand the tenures of employees, where there is a highest risk of attrition.

Background

FermaLogis is also one such pharmaceutical company that relies on its people and treats them like an asset. They are not only the enablers but also a central input resource for their operations. However, FermaLogis is now struggling with the problem of high employee turnover in all experience levels. Young professionals, after getting trained by the company, are leaving before FermaLogis has started benefitting from the expenditure on training. Experienced professionals are being poached by rivals who are offering them higher salaries. Employee turnover has been proven to be associated with decreased company productivity (Glebbeek & Bax, 2004) (Guthrie, 2001) and thus needs to be mitigated. FermaLogis understands the adverse effects of employee turnover and wants to take steps to minimize it. They want to understand who is leaving the company and why they are leaving. They have provided longitudinal data of past and current employees spanning 40 years. This dataset has values for employees who have left the company as well as current employees.

Approach to Analysis

In order to develop short term and long term employee retention strategies, it is imperative to first understand the dynamics of employee turnover. It is essential to answer some of the fundamental questions like – 'Who are leaving the company?', 'why they are leaving?', 'When is the biggest risk of an employee to leave the company?'. Our team of experts has attempted to answer some of these fundamental questions through the lens of 'Survival Analytics', a field of analytics that specializes in analyzing the expected duration/tenure of time until one or more events happen.

Choice of Analysis Methods

Several different perspectives on exploring employee turnover have been employed by researchers in the past. Survival Analysis is the one method that incorporates time as a variable of interest as opposed to other methods. (Morita, Lee, & Mowday, 1989). It has been used to model employee turnover in many different industries such as healthcare (Somers, 1996) (Iverson, 1999), social services (Iglehart, 1990), personal selling (William C. Moncrief III, 1989).

Major Findings

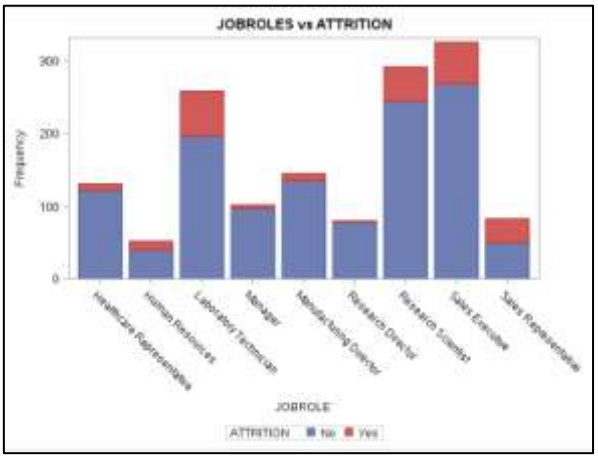
- It was observed that the overall tenure of employees who do not work overtime increases by 161% compared to the employees who work overtime
- Employees having stock option level 2 stay for a tenure, which is 131% higher than the employees having the stock option level 3
- Employees who travel frequently as a part of work, have a tenure which is 21% lower than their other counterparts.
- The tenure of female Employees is 32% higher than their male counterparts
- Employees who has a 'Low' Job Involvement tend to have a tenure which is 58% lower than the professionals with 'Very High' Job Involvement levels.

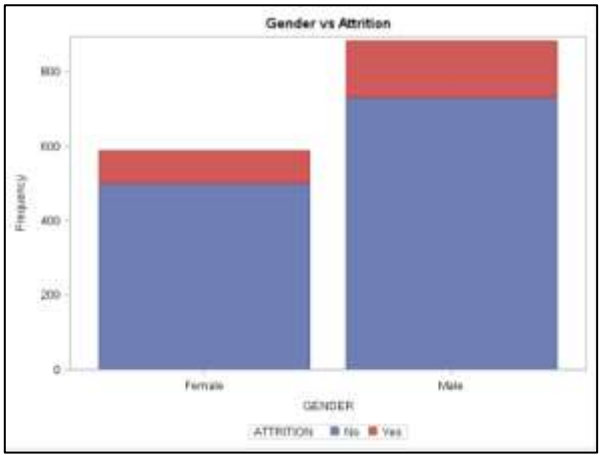
Descriptive Analytics

The graph shows the employees based on their individual job roles, also grouped together by whether they are leaving the company or not.

The highest attrition was seen in the Sales Representatives, Laboratory Technicians and Research Scientists.

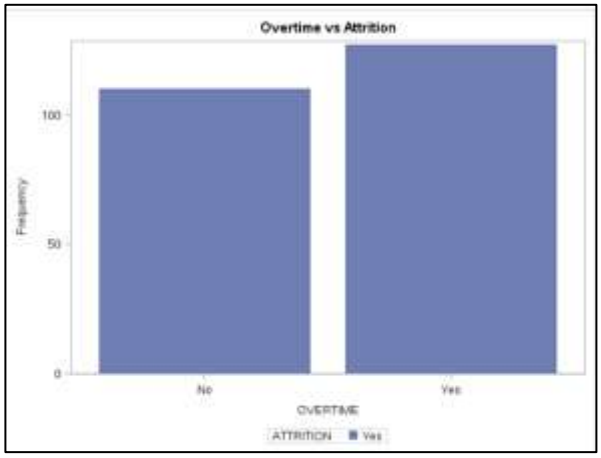
Research Directors and Managers have the lowest attrition rates.

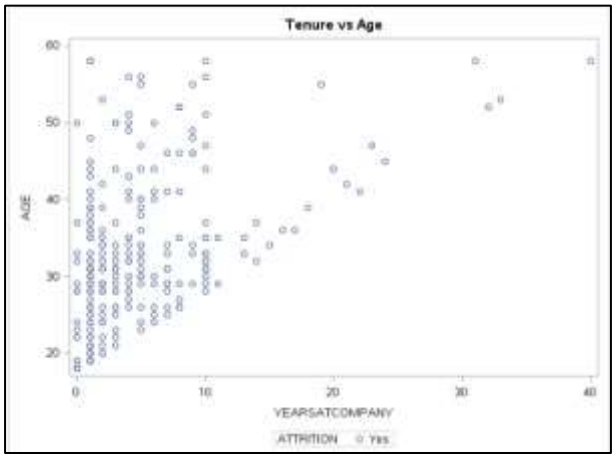




Attrition with the Male population is higher compared to the Female population

Employees who do overtime have higher attrition when compared to other employees



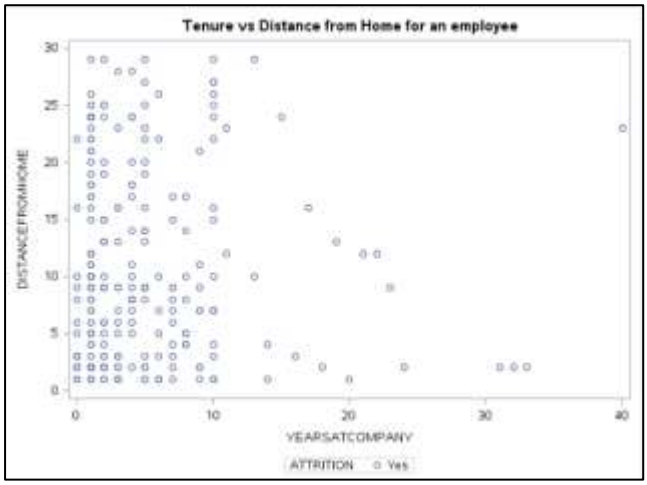


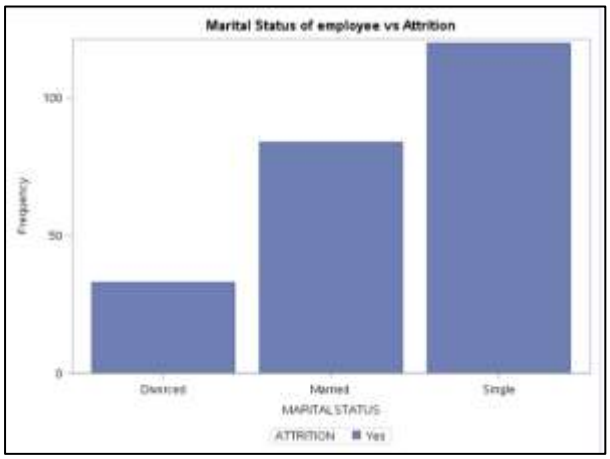
There are many employees who leave the company very early in their tenure (<1 years), across all the age groups.

There are a lot of employees in the age group of 20-30 years who leave the company

The attrition for shorter tenure isn't necessarily influenced by the distance of office from home.

As the tenure increases, majority of the employees who leave have relatively higher distance of office from home





Single employees are the ones who have the highest attrition rate.

Divorced employees have the lowest attrition rate.

After understanding the work force of FermaLogis, this study demonstrates the application of survival analytics to both understand the underlying factors influencing the turnover of an employee as well as to develop analytical models in order to predict the tenure of an employee, given his/her attributes. The study is divided into multiple modules, each of which caters to understanding a unique business objective.

Understanding the employees who are leaving the company

We have followed two survival analytics methodologies in order to understand the employees who are leaving the company

Univariate analysis to understand the impact of each of the variables on the survival probabilities of the employees

Category	Attribute	Who are leaving?	Why are they leaving? (Possible Reasons)
Demographics	Age	Young employees (age <=25 years) have the highest attrition rate amongst all the employees	Young and Single employees have the lesser responsibilities with minimal family commitments and have high appetite for risk. As such, this group is the one showing highest attrition
	Marital Status	Single Employees have higher propensity to leave Divorced Employees are the most stable employees Married Employees have the highest propensity to leave after a tenure of > 15 years	
	Distance From Home	Employees who live more than 10 miles away from the office have greater risk of leaving the company	Longer travel times can leave employees with a lot of fatigue and more time away from the family
Employee Background	Number Companies Worked	Employees who have worked for 5 or more companies have the highest risk of leaving the company	Employees who frequently change their employers, probably look for more opportunities more frequently

	Total Working Years	Employees with an overall work experience of <=5 years have the highest propensity to leave the company	In the first 5 years of their careers, employees explore their opportunities and as a result, move companies frequently.
Job Profile	Business Travel	Employees who travel frequently have the highest propensity to leave the company	Frequent travel can leave employees with a lot of fatigue and more time away from the family
	Department	Employees in the 'Sales' Department are attriting faster than in any other department	Sales department employees, probably face stiff goals as a part of their job responsibilities, as a result, are more prone to leaving
	Job Level	Junior employees are at the highest risk of leaving the company	More Junior employees can leave either due to a. Dissatisfaction with their current supervisors b. Have greater appetite for risk, thereby exploring newer opportunities
	Job Role	Sales Representatives, Laboratory Technicians and HRs are the employees who have the highest rate of attrition	
Salary & Incentives	Monthly Income	Low-Salary employees (with a monthly salary of <= $ 2,900) have significantly high propensity to leave	Salaries can be a big motivation for employees and a low monthly salary is a big reason for higher attrition
	Over Time	Employees who work 'OverTime' have a very high risk of leaving the company	Lesser Job Satisfaction due to 'burn-out' as a result of

		when compared to the other employees	consistent overtime can result in employees leaving the company
Employee Current Profile	Years In Current Role	Employees who are new to their current roles - between 0-2 years, have the highest propensity to leave the company	Employees who are unable to adjust to their new roles during the initial years of their tenure are the most susceptible to leave the company
	Years Since Last Promotion	There is a higher probability of employees quitting their jobs if they have not been promoted for a long time	One of the major motivations, apart from salary, is professional growth within the company
	Years With Current Manager	Employees having an overall tenure with their current managers between 0-2 years are at the highest risk of leaving the company	Difficulty to adjust with their supervisor/manager during the initial few years, could be the potential reason why this group is leaving the company
	Stock Option Level	Employees who are not given any stock option have the highest propensity to leave the company	Stock Options are a good motivation for employees to stay.
	Training Times Last Year	Employees who did not participate in any trainings last year had the highest attrition	A possible reason for this could be the lack of training opportunities provided to the employees. Continuous training opportunities impact the growth potential of employees and increase their overall worth

Employee Satisfaction	Environment Satisfaction	Employees with the least environment satisfaction have the highest attrition rate	These results testify that apart from monetary satisfaction, employees who do not have a good work environment and work-life balance, tend to have lesser job satisfaction and thereby quit the company more often than the other employees
	Job Involvement	There is a marked variance in the rate of attrition of employees with a very low Job Involvement compared to other employees. Employees with a low Job involvement attrite the highest	
	Job Satisfaction	Employees with a low Job satisfaction attrite the highest	
	Work Life Balance	Employees with a bad work life balance have the highest propensity to leave the company	

Survival Analytics Models to predict the tenure of Employees

We developed a survival analytics predictive model for the entire employee data. In order to better understand the dynamics of employee retention, we segregated the data into smaller subsets, for each of the departments – Sales, R&D and HR and developed individual models for the employees in the Sales and R&D department.

Survival Analytics Model on the overall data:

As per the model estimates, we can derive the following insights on the overall 'Tenure' of the employees

1. For every one year increase in the age of an employee, the overall tenure of the employee within the company increases by 9%
2. For every one mile increase in the office distance from home of an employee, the overall tenure of the employee within the company reduces by 3%
3. For every one additional company worked for, by an employee, the tenure of the employee reduces by 14%
4. Compared to Males, the overall tenure of the employee for Females increases by 32%
5. Compared to employees who travel rarely:

   a. The tenure for employees who do not travel increases by 94%
   b. The tenure for employees who travel frequently reduces by 21%

6. Compared to employees who do overtime, the tenure for other employees increases by 161%
7. Compared to employees who have the stock option-3:
   a. Tenure for employees who have no stock option reduces by 21%
   b. Tenure for employees who have the stock option-1 increases by 48%
   c. Tenure for employees who have the stock option-2 increases by 131%
8. Compared to employees who have the Very High environment satisfaction
   a. The tenure of Employees with a Low environment satisfaction reduces by 43%
   b. The tenure of Employees with Medium environment satisfaction, reduces by 7%
   c. The tenure of Employees with High environment satisfaction, increases by 5%
9. Compared to employees who have the Very High Job Involvement (Job Involvement=4)
   a. The tenure of Employees with the lowest Job Involvement (1) reduces by 58%
   b. The tenure of Employees with the medium Job Involvement (2) reduces by 47%
   c. The tenure of Employees with the High Job Involvement (3) reduces by 32%

Survival Analytics Model for the 'R&D' Department:

As per the model estimates, we can derive the following insights on the overall 'Tenure' of the employees

1. For every one year increase in the age of an employee, the overall tenure of the employee within the company increases by 2%
2. For every one mile increase in the office distance from home of an employee, the overall tenure of the employee within the company reduces by 3%
3. For every one additional company worked for, by an employee, the tenure of the employee reduces by 10%
4. Compared to Males, the overall tenure of the employee for Females increases by 36%

5. Compared to employees who do overtime, the tenure for other employees increases by 172%
6. Compared to employees who have the stock option-3:
    a. Tenure for employees who have no stock option reduces by 44%
    b. Tenure for employees who have the stock option-1 reduces by 7%
    c. Tenure for employees who have the stock option-2 increases by 31%"
7. Compared to employees who have the very high environment satisfaction
    a. The tenure of Employees with a low environment satisfaction reduces by 42%
    b. The tenure of Employees with a medium environment satisfaction, increases by 14%
    c. The tenure of Employees with a high environment satisfaction, reduces by 10%

Survival Analytics Model for the 'Sales' Department:

As per the model estimates, we can derive the following insights on the overall 'Tenure' of the employees

1. For every one mile increase in the office distance from home of an employee, the overall tenure of the employee within the company reduces by 3%
2. For every one additional company worked for, by an employee, the tenure of the employee reduces by 13%
3. For every one year increase since the last promotion, the tenure of the employee increases by 9%
4. Compared to Single Employees:
    a. The tenure of employees who are Divorced, increases by 113%
    b. The tenure of employees who are Married, increases by 74%
5. Compared to employees who do overtime, the tenure for other employees increases by 87%
6. Compared to employees who have very high Job Involvement
    a. The tenure of Employees with low Job Involvement (1) reduces by 58%
    b. The tenure of Employees with medium Job Involvement (2) reduces by 27%
    c. The tenure of Employees with high Job Involvement (3) increases by 7%
7. Compared to employees who have very high income
    a. The tenure of Employees with low income reduces by 88%
    b. The tenure of Employees with medium income reduces by 44%
    c. The tenure of Employees with high income reduces by 44%

Most Likely Employee Turnover Times

Our Approach – We have obtained year interval of highest Attrition Risk of the employees in the first 5 years & next 10 years of their employment from the hazard plots of:

a. Age Groups
b. Job Roles
c. Work Experience groups (based on Total Working Years)
d. Job Level
e. Business Travel

Since we want to capture the hazard trend of employees for 15 years of their employment, we are considering time/tenure till 15 years.

DANGER of ATTRITION: Across Age Groups

We have created 5 age groups as <25, 25-30, 30-40, 40-50, 50-60, >60 as below:

```
length age_group $20.;
if age=. then age_group="null";else
if age<=30 then age_group="<=30";else
if 30<age<=40 then age_group="30-40";else
if 40<age<=50 then age_group="40-50";else
if 50<age<=60 then age_group="50-60";else
   age_group=">60";
```

Age Group	Year Interval of highest Attrition Risk	
	0-5 years	5-15 years
<=30	0-2	9-11
30-40	0-2	9-11
40-50	3-5	9-11
50-60	3-5	9-11

The table on the Left Hand Side gives the year interval of highest risk of attrition in the first 5 years and the next 5-15 years of employment. Across all the Age Groups, the year interval which has the ***highest conditional probability of failure*** is included here. Young Employees who are <=40 years have highest risk of attrition in the years 0-2 for the first 5 years of their employment and in the years 9-11 for the next 10 years of their

employment. Attrition Risk for Employees above 40 years of age is highest in 3-5 years of their employment in the first 5 years of their job and years 9-11 for the next 10 years!

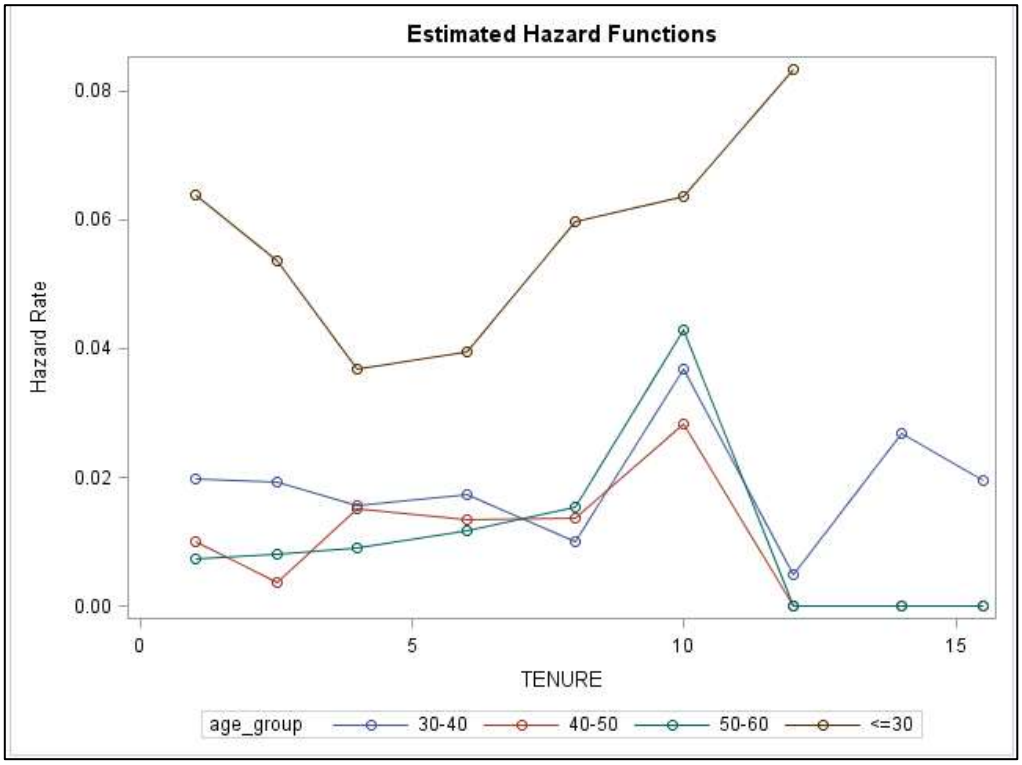
Summary of the Number of Censored and Uncensored Values					
Stratum	age_group	Total	Failed	Censored	Percent Censored
1	30-40	619	85	534	86.27
2	40-50	322	34	288	89.44
3	50-60	143	18	125	87.41
4	<=30	386	100	286	74.09
Total		1470	237	1233	83.88

The summary table on the LHS shows that **Young employees** attrite the most as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

- **26% of the Employees aged <=30 years attrite and the year which poses the biggest threat for them to leave is 0-2 years in the first 5 years of their employment!**
- **In a time-span of 15 years, High attrition risk is present in years 9-11 after 5 years of employment!**

Code used:

```
proc lifetest data=project1.emp_prep_final method=life plots=(s,h)
intervals= 2 3 5 7 9 11 13 15 16;
time tenure*censored(0);
strata age_group;
run;
```



Based on the hazard plot on the Left Hand Side, we find that young employees (<=30 years) have significantly higher risk of attrition than other age groups.

DANGER OF ATTRITION: Across Job Roles

Job Role	Year Interval of highest Attrition Risk	
	0-5 years	5-15 years
Laboratory Technician	0-2	9-11
Research Scientist	0-2	9-11
Healthcare Representative	0-2	9-11
Sales Representative	3-5	9-11
Sales Executive	3-5	9-11

Human Resources	3-5	No Risk
Manufacturing Director	3-5	9-11
Manager	No Risk	9-11
Research Director	No Risk	No Risk

The table on the Left Hand Side gives the year interval of highest risk of attrition in the first 5 years and the next 5-15 years of employment. Across all the Job Roles, the year interval which has the ***highest conditional probability of failure*** is included here. We find that for Laboratory Technicians, Research Scientists & Healthcare Representatives the biggest threat of attrition is years 0-2 in the first 5 years and years 3-5 for other job roles. If the employees have chosen to stay beyond 5 years, then years 9-11 of their employment poses the biggest risk of attrition! For senior level positions of Manager & Research Director, there is no risk of attrition during the initial 5 years. For Human Resources, there is No Risk in 5-15 years of employment. However, we ignore it since Human Resources constitutes just 4% of the employee data!

Summary of the Number of Censored and Uncensored Values

Stratum	JOBROLE	Total	Failed	Censored	Percent Censored
1	Healthcare Representative	131	9	122	93.13
2	Human Resources	52	12	40	76.92
3	Laboratory Technician	259	62	197	76.06
4	Manager	102	5	97	95.10
5	Manufacturing Director	145	10	135	93.10
6	Research Director	80	2	78	97.50
7	Research Scientist	292	47	245	83.90
8	Sales Executive	326	57	269	82.52
9	Sales Representative	83	33	50	60.24
Total		1470	237	1233	83.88

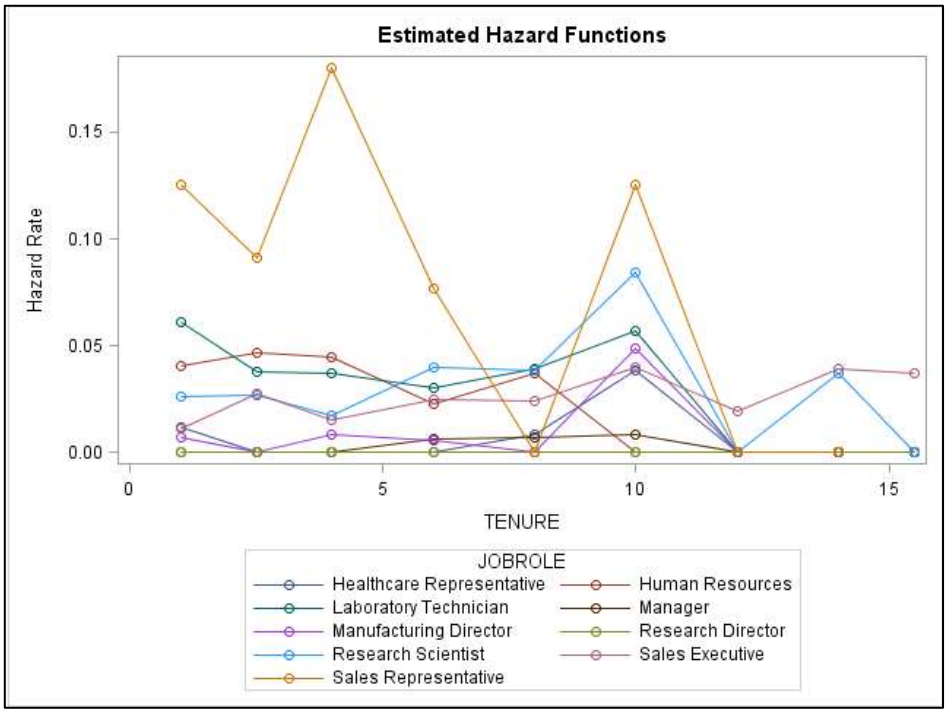
The summary table on the LHS shows that **Sales Representatives, Laboratory Technicians and Human Resources attrite the most** as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

- **40% of the Sales Representatives attrite and the year which poses the biggest threat for them to leave is 3-5 years in the first 5 years of their employment!**
- **24% of the Laboratory Technicians attrite and the year which poses the biggest threat for them to leave is 0-2 years in the first 5 years of their employment!**
- **In a time-span of 15 years, High attrition risk is present in years 9-11 after 5 years of employment!**

Code used:

```
proc lifetest data=project1.emp_prep_final method=life plots=(s,h)
intervals= 2 3 5 7 9 11 13 15 16 ;
time tenure*censored(0);
strata jobrole;
run;
```

Based on the hazard plot below, we find Sales Representatives have significantly higher risks of attrition than other job roles. ***Research Director seems to have no risk of attrition!***



DANGER OF ATTRITION: Across "Work Experience" Groups

We have created Total_Work groups as follows:

```
IF TOTALWORKINGYEARS <= 5 THEN TOTAL_WORK = "<=5 YEARS";
ELSE IF TOTALWORKINGYEARS <= 10 THEN TOTAL_WORK = "6-10 YEARS";
ELSE IF TOTALWORKINGYEARS <= 15 THEN TOTAL_WORK = "11-15 YEARS";
ELSE IF TOTALWORKINGYEARS <= 20 THEN TOTAL_WORK = "16-20 YEARS";
ELSE IF TOTALWORKINGYEARS <= 25 THEN TOTAL_WORK = "21-25 YEARS";
ELSE IF TOTALWORKINGYEARS <= 30 THEN TOTAL_WORK = "26-30 YEARS";
ELSE IF TOTALWORKINGYEARS <= 35 THEN TOTAL_WORK = "31-35 YEARS";
ELSE IF TOTALWORKINGYEARS <= 40 THEN TOTAL_WORK = "36-40 YEARS";
```

```
ELSE TOTAL_WORK = ">41 YEARS";
```

Work Experience Group	Year Interval of highest Attrition Risk	
	0-5 years	5-15 years
<=5 years	0-2	No Data
6-10 years	2-3	9-11
11-15 years	0-2	9-11
16-20 years	3-5	9-11
21-25 years	0-2	No Risk
26-30 years	No Risk	7-9
31-35 years	No Risk	9-11
36-40 years	No Risk	No Risk

The table on the Left Hand Side gives the year interval of highest risk of attrition in the first 5 years and the next 5-15 years of employment.

Across all the Work Experience Groups, the year interval which has the ***highest conditional probability of failure*** is included here.

Summary of the Number of Censored and Uncensored Values					
Stratum	TOTAL_WORK	Total	Failed	Censored	Percent Censored
1	11-15 YEA	191	24	167	87.43
2	16-20 YEA	149	15	134	89.93
3	21-25 YEA	109	9	100	91.74
4	26-30 YEA	52	2	50	96.15
5	31-35 YEA	33	3	30	90.91
6	36-40 YEA	13	2	11	84.62
7	6-10 YEAR	607	91	516	85.01
8	<=5 YEARS	316	91	225	71.20
Total		1470	237	1233	83.88

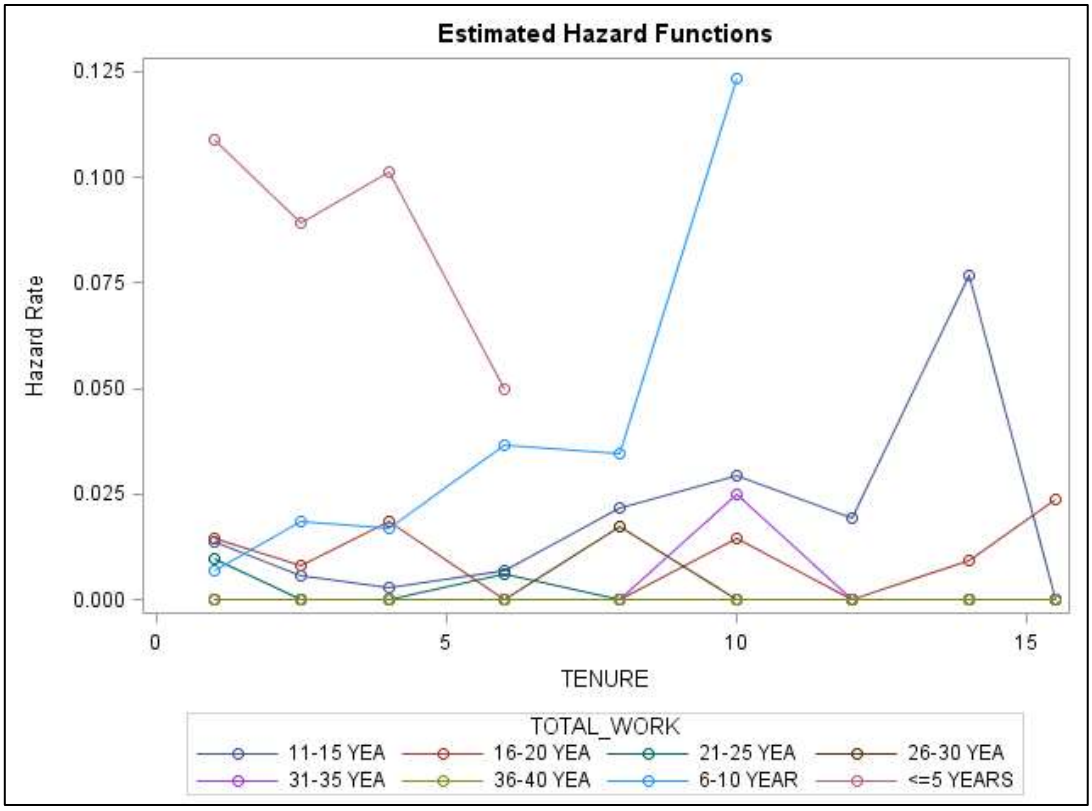
The summary table on the LHS shows that employees with work experience <=5 years attrite the most attrite the least as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

- **29% of the Employees with work experience <=5 years attrite and the year which poses the biggest threat for them to leave is 0-2 years in the first 5 years of their employment!**
- **In a time-span of 15 years, High attrition risk is present in years 9-11 after 5 years of employment!**

Team 8

Code used:

```
proc lifetest data=project1.emp_prep_final method=life plots=(s,h)
intervals= 2 3 5 7 9 11 13 15 16 ;
time tenure*censored(0);
strata total_work;
run;
```



The hazard plot on the RHS shows that employees having <=5 years of work experience have the highest risk of attrition which exponentially drops. However, attrition risk of employees having work experience of 6-10 years increases exponentially.

DANGER of ATTRITION: Across "Job Level"

Job Level	Year Interval of highest Attrition Risk	
	0-5 years	5-15 years
1	0-2	9-11
2	3-5	9-11
3	0-2	9-11
4	0-2	No Risk
5	No Risk	5-7

The table on the Left Hand Side gives the year interval of highest risk of attrition in the first 5 years and the next 5-15 years of employment. Across all the Job Level Groups, the year interval which has the ***highest***

conditional probability of failure is included here. Job Level groups are based on the seniority in management with 1 being the lowest and 5 being the highest.

Summary of the Number of Censored and Uncensored Values					
Stratum	JOBLEVEL	Total	Failed	Censored	Percent Censored
1	1	543	143	400	73.66
2	2	534	52	482	90.26
3	3	218	32	186	85.32
4	4	106	5	101	95.28
5	5	69	5	64	92.75
Total		1470	237	1233	83.88

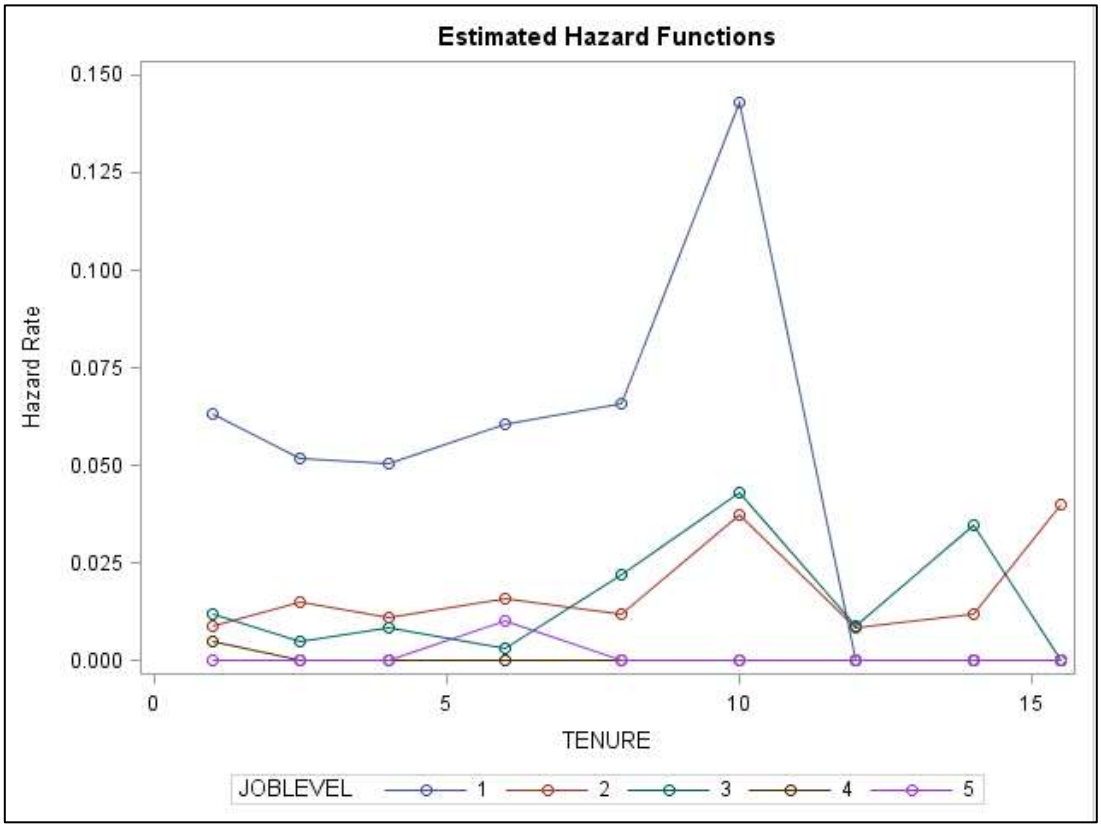
The summary table on the LHS shows that employees with Job Level 1 attrite the most while employees with job Level 4,5 attrite the least as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.** In other words, **senior management attrite the least while junior employees attrite the most**.

- **26% of the Employees with Job Level=1, i.e. holding junior positions attrite and the year which poses the biggest threat for them to leave is 0-2 years in the first 5 years of their employment!**
- **In a time-span of 15 years, High attrition risk is present in years 9-11 after 5 years of employment!**

Code used:

```
proc lifetest data=project1.emp_prep_final method=life plots=(s,h)
intervals= 2 3 5 7 9 11 13 15 16 ;
time tenure*censored(0);
strata joblevel;
run;
```

The risk of attrition is significantly higher in employees holding junior positions.



DANGER of ATTRITION: Across "Business Travel" frequency

Business Travel	Year Interval of highest Attrition Risk	
	0-5 years	5-15 years
Travel Frequently	0-2	9-11
Travel Rarely	0-2	9-11
Non Travel	0-2	9-11

The table on the Left Hand Side gives the year interval of highest risk of attrition in the first 5 years and the next 5-15 years of employment. Across all the Business Travel Groups, the year interval which has the ***highest conditional probability of failure*** is included here. We see here across all groups, years 0-2 are with highest risk of attrition in the first 5 years of the employment and for the next 10 years, it is years 9-11!

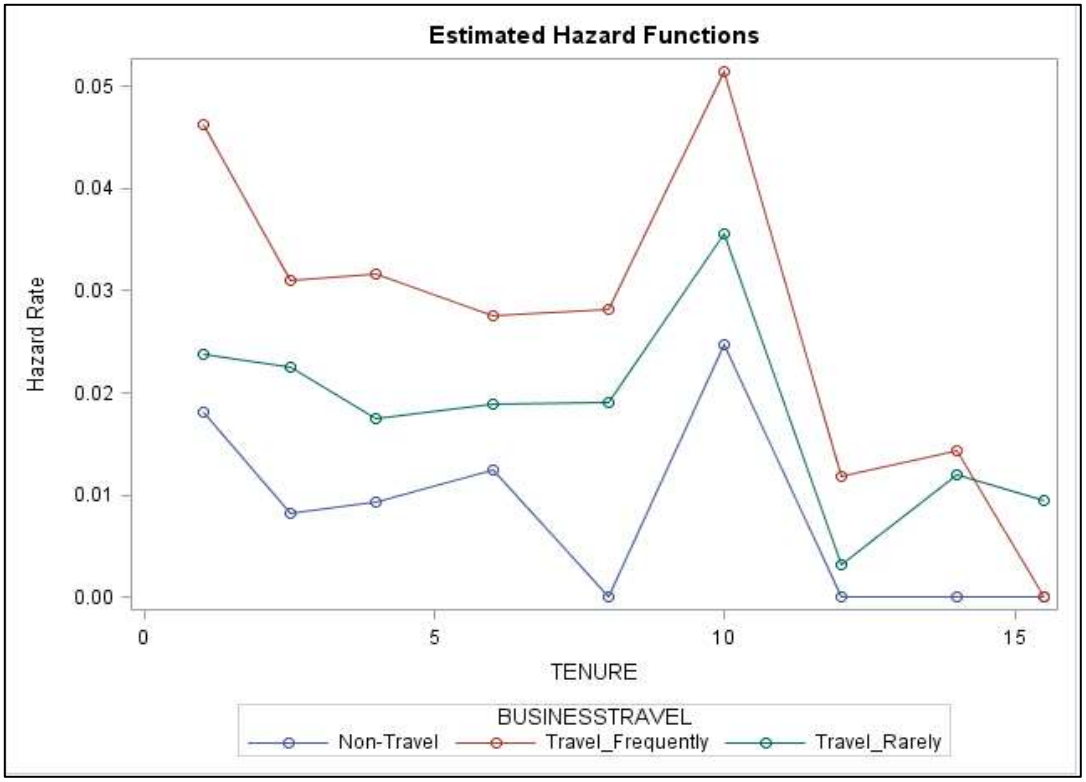
Summary of the Number of Censored and Uncensored Values					
Stratum	BUSINESSTRAVEL	Total	Failed	Censored	Percent Censored
1	Non-Travel	150	12	138	92.00
2	Travel_Frequently	277	69	208	75.09
3	Travel_Rarely	1043	156	887	85.04
Total		1470	237	1233	83.88

The summary table on the LHS shows that employees who travel frequently attrite the most as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

- **25% of the Employees Who Travel frequently attrite and the year which poses the biggest threat for them to leave is 0-2 years in the first 5 years of their employment!**
- **In a time-span of 15 years, High attrition risk is present in years 9-11 after 5 years of employment!**

Code used:

```
proc lifetest data=project1.emp_prep_final method=life plots=(s,h)
intervals= 2 3 5 7 9 11 13 15 16;
time tenure*censored(0);
strata businesstravel;
run;
```



Employees who travel frequently have the highest risk of attrition!

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

The following variables are the ones that we identified as categorical (class) or continuous variables in the dataset. The below table gives an insight about which of these variables are to be used

Category	**Attribute**	**Type**	**PROC LIFETEST**
Demographics	Age	**Continuous**	**Test**
	Gender	**Dichtomous**	**Strata**
	Marital Status	**Class/Nominal**	**Strata**
	Distance from Home	**Continuous**	**Test**
Employee background	Number of Companies worked	**Continuous**	**Test**
	Total working years	**Continuous**	**Test**
	Education	**Class/Ordinal**	**Strata**
	Education Field	**Class/Nominal**	**Strata**
Job profile	Business Travel	**Class/Nominal**	**Strata**
	Department	**Class/Nominal**	**Strata**
	Job Level	**Class/Ordinal**	**Strata**

	Job Role	**Class/Nominal**	**Strata**
Salary and incentives	Hourly Rate	**Continuous**	**Test**
	Monthly Rate	**Continuous**	**Test**
	Daily Rate	**Continuous**	**Test**
	Monthly Income	**Continuous**	**Test**
	Overtime	**Dichtomous**	**Strata**
	Percent Salary Hike	**Continuous**	
	Performance Rating	**Class/Ordinal**	**Strata**
Employee current profile	Years at Company	**Continuous**	**Test**
	Years in current role	**Continuous**	**Test**
	Years since last promotion	**Continuous**	**Test**
	Years with Current manager	**Continuous**	**Test**
	Stock option level	**Class/Ordinal**	**Strata**
Employee satisafaction	Environment Satisfaction	**Class/Ordinal**	**Strata**
	Job Involvement	**Class/Ordinal**	**Strata**
	Job Satisfaction	**Class/Ordinal**	**Strata**
	Relationship satisfaction	**Class/Ordinal**	**Strata**
	Work life Balance	**Class/Ordinal**	**Strata**

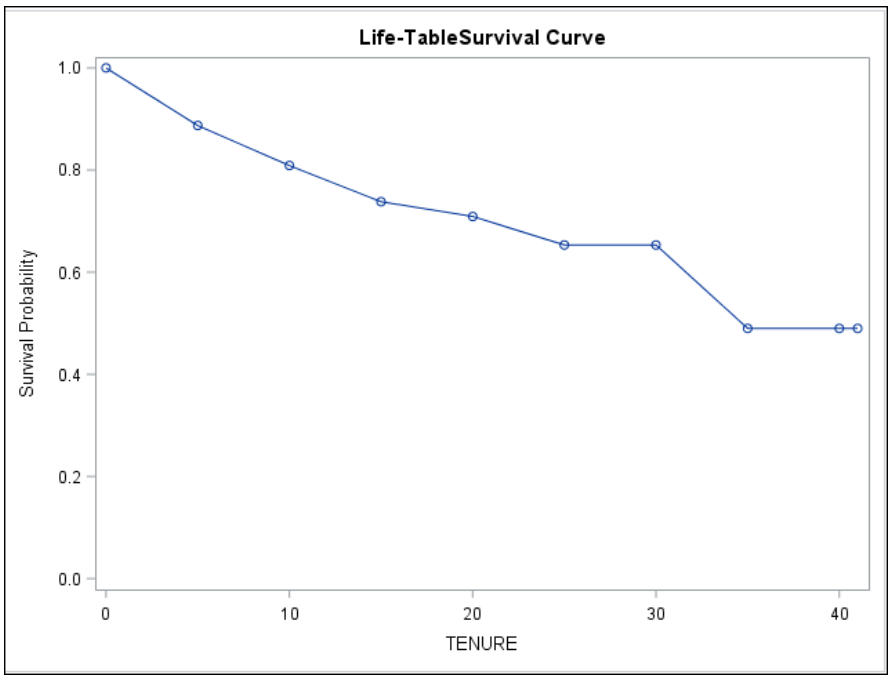
From the above table we are considering the categorical variables that have different groups and classes for the analysis of attrition of employees.

We checked the employee attrition by using the lifetest method on each of the class variables. The survival and hazard plots indicate whether that category affects the overall survival time of an employee or not. Also, the survival rates have been checked based on various groups within each of the variables.

Our analysis is divided basically into checking all individual variables effects as well as when they interact with each other amongst themselves. The analysis is only based on all those variables that have a significant effect on the model.

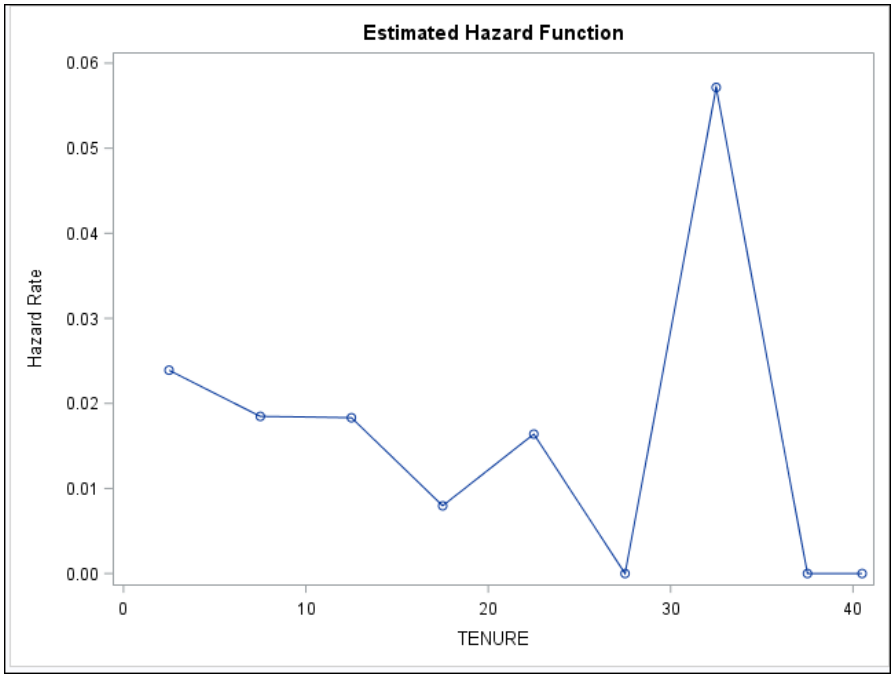
DEPARTMENT

Survival Plot:



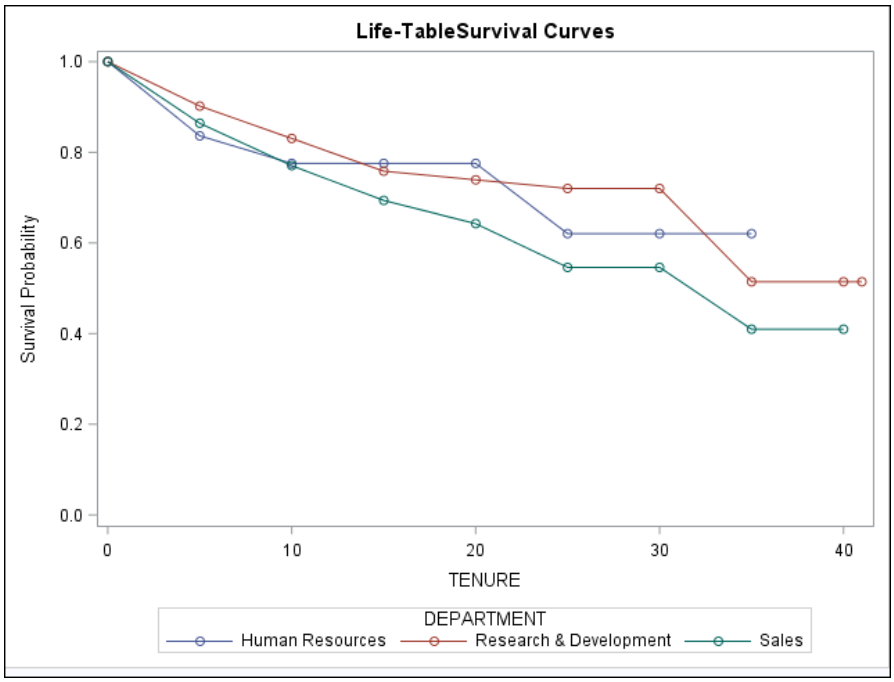
The survival plot now shows a rather clear picture regarding the survival probabilities of the employees. This indicates a much more appropriate analysis of the survival probabilities of the employees.

Hazard Plot:



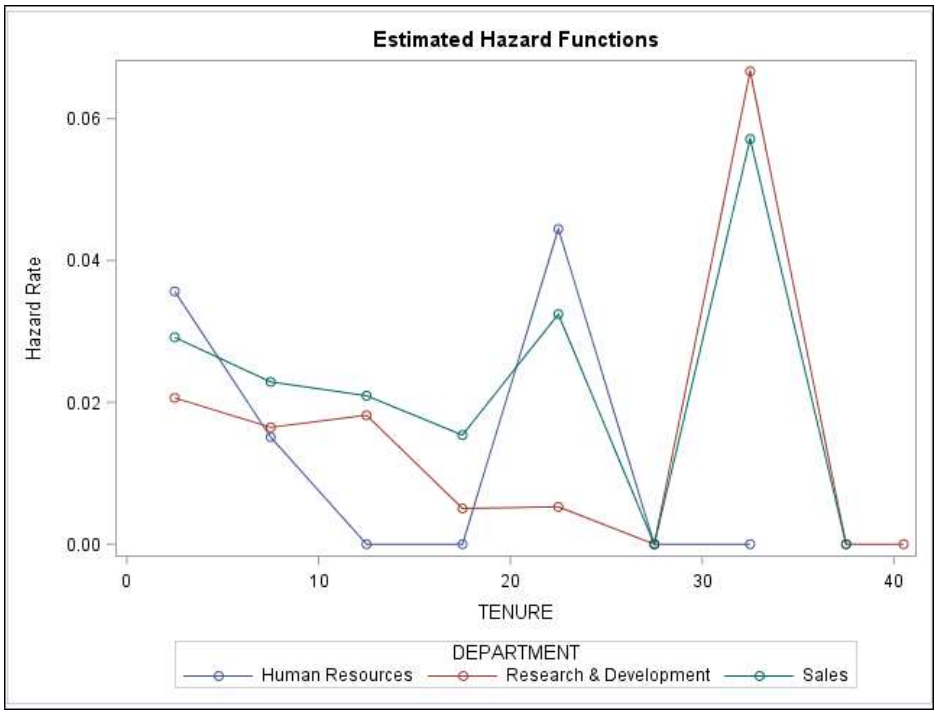
This gives a more stable hazard plot. Also, the hazard seems to significantly jump during 30-40 years of tenure, as per the data.

Understanding the impact of 'Department' on the survival probabilities



The survival plots show almost the same survival probabilities for all the 3 department types.

Hazard plots:

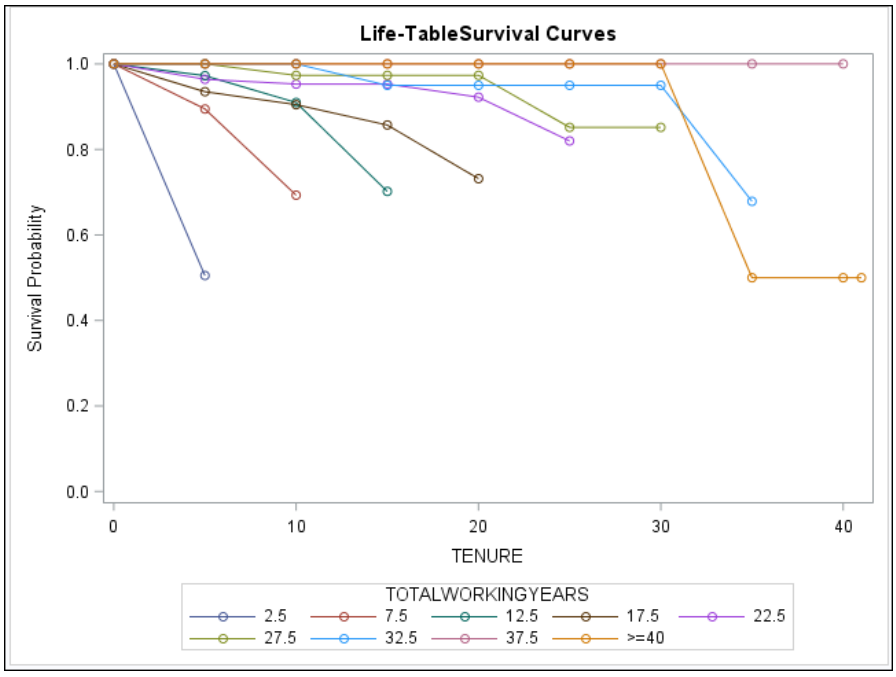


As per the above revised plots we can check that the hazard rates and the survival probabilities are all different. It does indicate that the department variable does not have an impact as a whole on the survival time of an employee.

Team 8

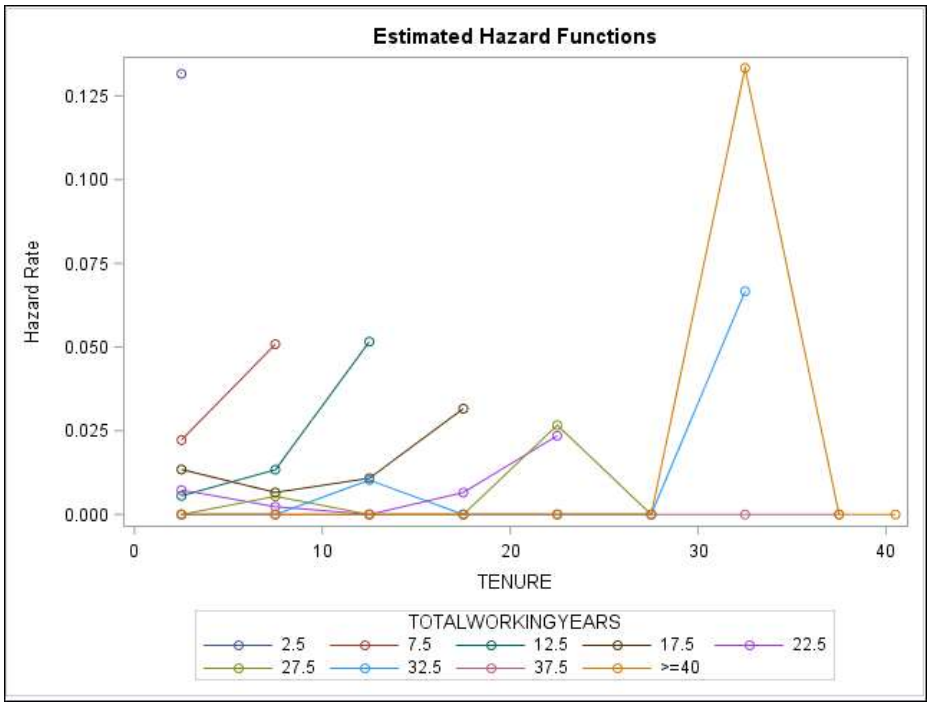
TOTAL WORKING YEARS

Survival Plot:



The survival plot indicates just a steep curve for the employees having a total work experience of up to 2.5 years.

Hazard Plot:



Both the hazard plots and the survival plots do not really show such an impact over any of their values. So we tried to recode the values.

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
TOTALWORKINGYEARS	TOTALWORKINGYEARS	Chi-Square	Raw	Bonferroni
2.5	7.5	24.7189	<.0001	<.0001
2.5	12.5	106.7	<.0001	<.0001
2.5	17.5	124.8	<.0001	<.0001
2.5	22.5	162.8	<.0001	<.0001
2.5	27.5	188.9	<.0001	<.0001
2.5	32.5	203.8	<.0001	<.0001
2.5	37.5	239.1	<.0001	<.0001
2.5	40	257.2	<.0001	<.0001
7.5	12.5	15.9623	<.0001	0.0023
7.5	17.5	12.8871	0.0003	0.0119
7.5	22.5	21.7454	<.0001	0.0001
7.5	27.5	16.8381	<.0001	0.0015
7.5	32.5	16.0200	<.0001	0.0023
7.5	37.5	13.3457	0.0003	0.0093
7.5	40	10.0080	0.0016	0.0561

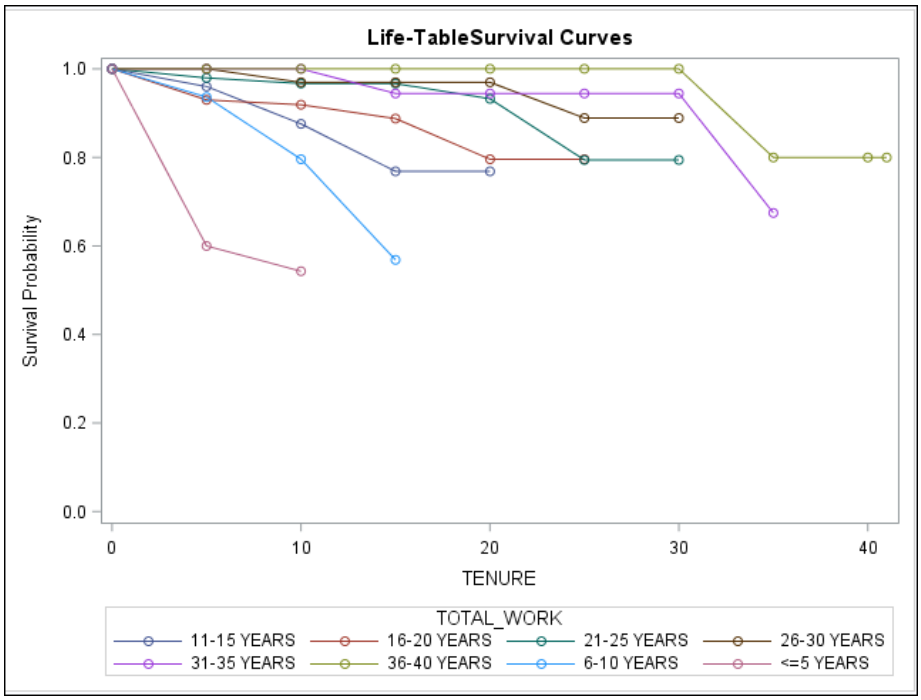
As per the 'Bonferroni' adjustment test, we note that the survival probabilities of all the employees with an overall work experience of 1-10 years is significantly different from the other employees.

In Order to understand the impact further, we classified the total work experience into multiple categories as below

```
*INTERVALS FOR TOTALWORKINGYEARS;
DATA SASDATA.EMP_LIFE_TABLE_DATA;
LENGTH TOTAL_WORK $30.;
SET SASDATA.EMP_LIFE_TABLE_DATA;
IF TOTALWORKINGYEARS <= 5 THEN TOTAL_WORK = "<=5 YEARS";
ELSE IF TOTALWORKINGYEARS <= 10 THEN TOTAL_WORK = "6-10 YEARS";
ELSE IF TOTALWORKINGYEARS <= 15 THEN TOTAL_WORK = "11-15 YEARS";
ELSE IF TOTALWORKINGYEARS <= 20 THEN TOTAL_WORK = "16-20 YEARS";
ELSE IF TOTALWORKINGYEARS <= 25 THEN TOTAL_WORK = "21-25 YEARS";
ELSE IF TOTALWORKINGYEARS <= 30 THEN TOTAL_WORK = "26-30 YEARS";
ELSE IF TOTALWORKINGYEARS <= 35 THEN TOTAL_WORK = "31-35 YEARS";
```

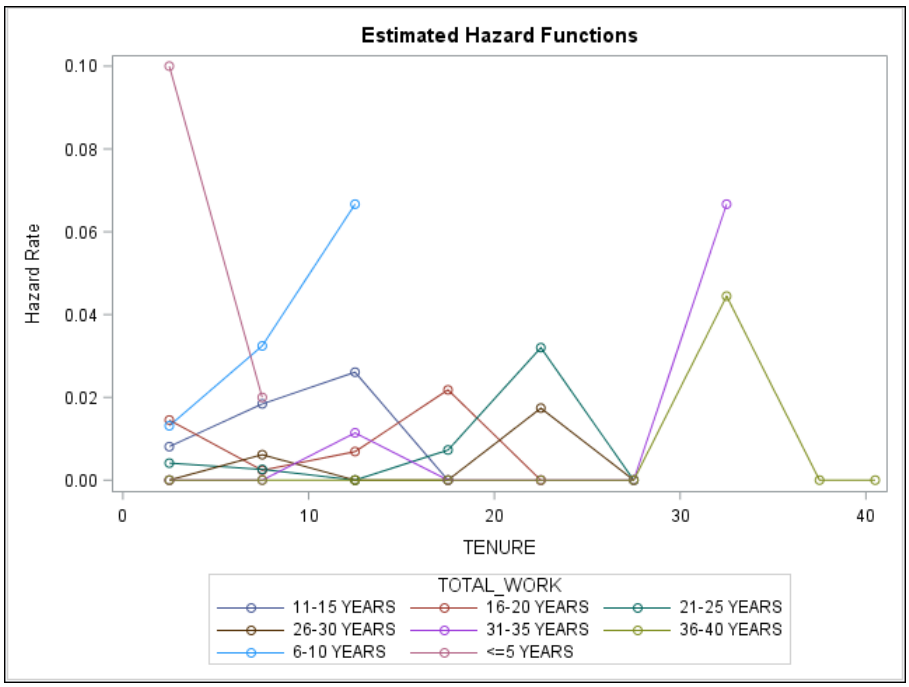
```
ELSE IF TOTALWORKINGYEARS <= 40 THEN TOTAL_WORK = "36-40 YEARS";
ELSE TOTAL_WORK = ">41 YEARS";
RUN;
```

Survival plot:



The survival curve shows a clear sign of lesser sign survival probability for employees having a total work experience of less than 5 years. These are the employees who are young and would want to switch jobs until they settle own to a better and a more secure and high paying role.

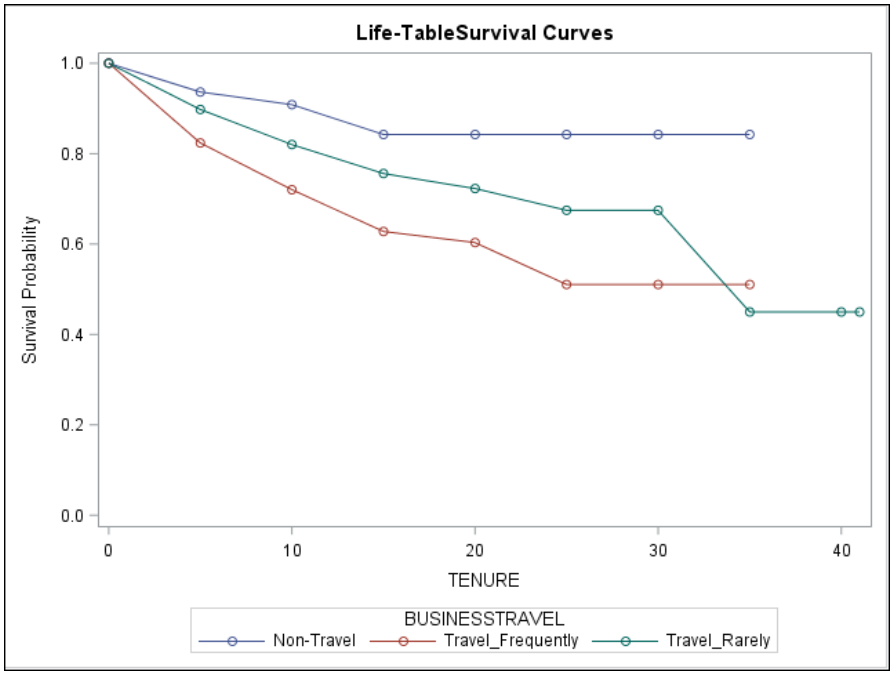
Hazard Plot:



This plot shows that there is stark contrast in the survival times of an employee with less than 5 years of experience. From the above graphs, we can see that the survival probabilities for employees with an overall <=5 years of work experience are significantly different than all the others. Hence, employees are prone to leave during their early phases of their careers.

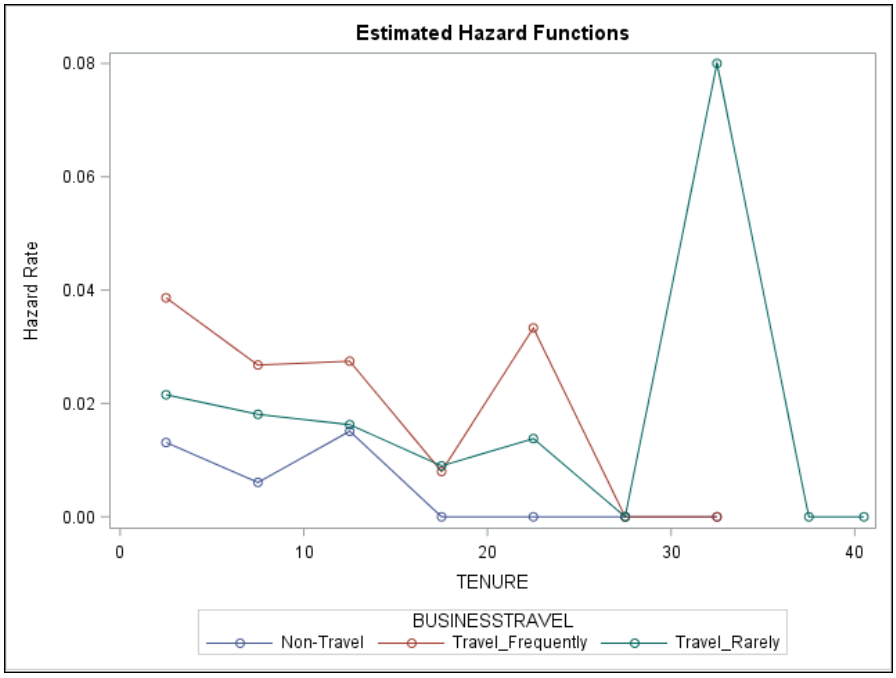
BUSINESS TRAVEL

Survival plots:



The survival plots of each of the education group (using the life-table method) is shown above. The employees with a frequent travel schedule have clearly a lower survival probability.

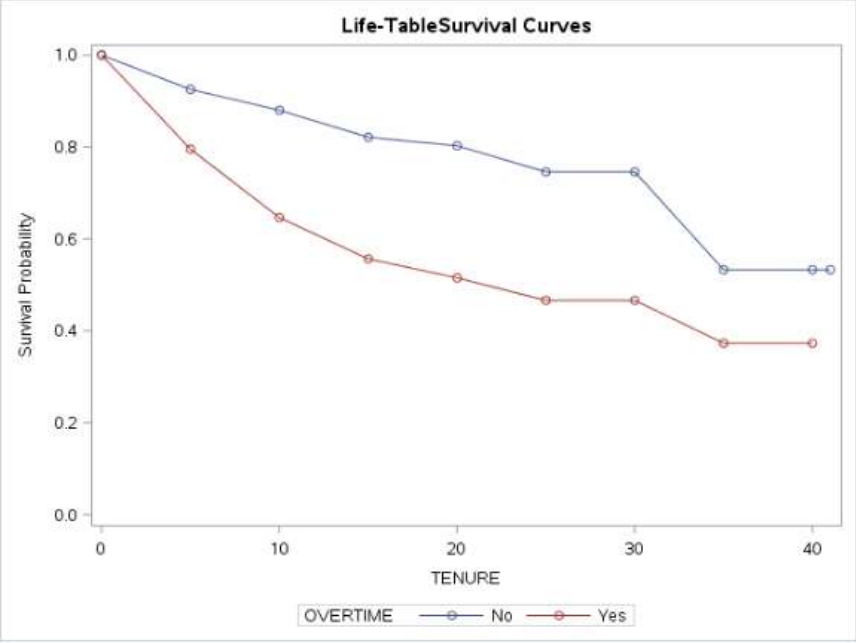
Hazard plots:



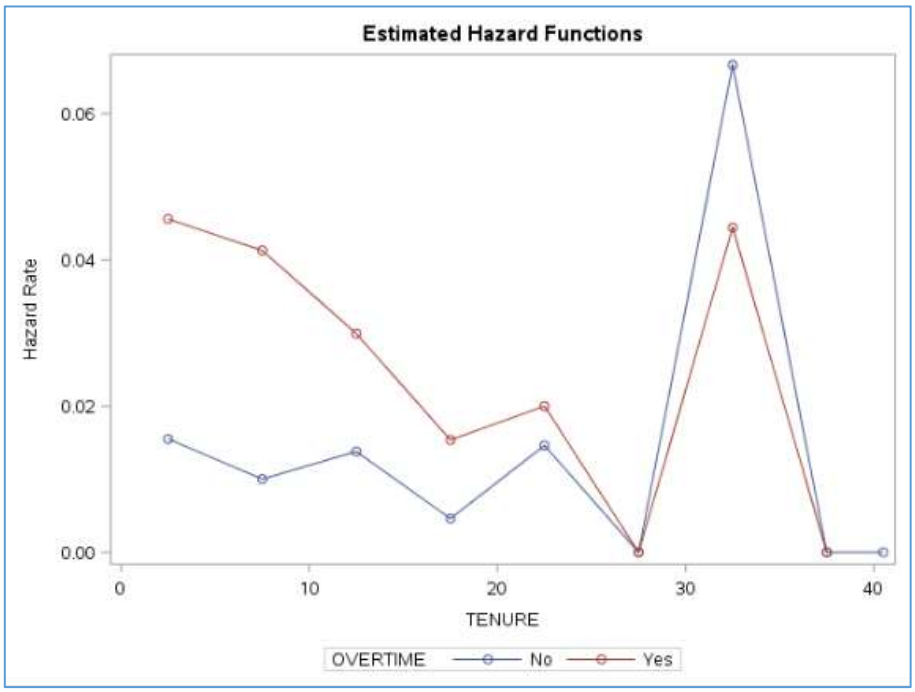
On checking the curves of both the survival and the hazard graphs, we can say that there is a statistical difference in the survival of employees who 'Travel Frequently' and those who 'Travel-Rarely'/'Non-Travel'. **This shows that the employees who travel frequently have the highest risk of leaving the company.**

OVERTIME

As-IS Survival Plot



The survival plot shows a steep curve for the employees who do have an Overtime, this means that all those employees working overtime are more prone to attrition than those who do not have an overtime.



The hazard plot shows a very high hazard rate for the employees at the start of their tenure period and also having an overtime to be done. In regards to employees with no overtime, this has a marked difference which is affecting the attrition. There is a significant increase in the hazard rate of employees with no overtime and have a tenure period of more than 30 years.

MONTHLY INCOME

Since Monthly Income is a continuous variable, we created Income groups based on the quartiles. This ensures that each of the 4 group has 25% of the employees. The 4 groups are as: low income, medium income, high income, very high income. Using proc univariate, we obtained the result on the RHS and based on 25th, 50th, 75th percentiles we created the income groups as follows:

Quantiles (Definition 5)	
Level	Quantile
100% Max	19999
99%	19827
95%	17856
90%	13798
75% Q3	8380
50% Median	4919
25% Q1	2911
10%	2316
5%	2097
1%	1359
0% Min	1009

```
length income $20.;
if monthlyincome=. then income="null";else
if monthlyincome<=2911 then income="low";else
if 2911<monthlyincome<=4919 then income="medium";else
if 4919<monthlyincome<=8380 then income="high";else
income="very high";
```

income	Frequency	Percent	Cumulative Frequency	Cumulative Percent
high	368	25.03	368	25.03
low	369	25.10	737	50.14
medium	366	24.90	1103	75.03
very high	367	24.97	1470	100.00

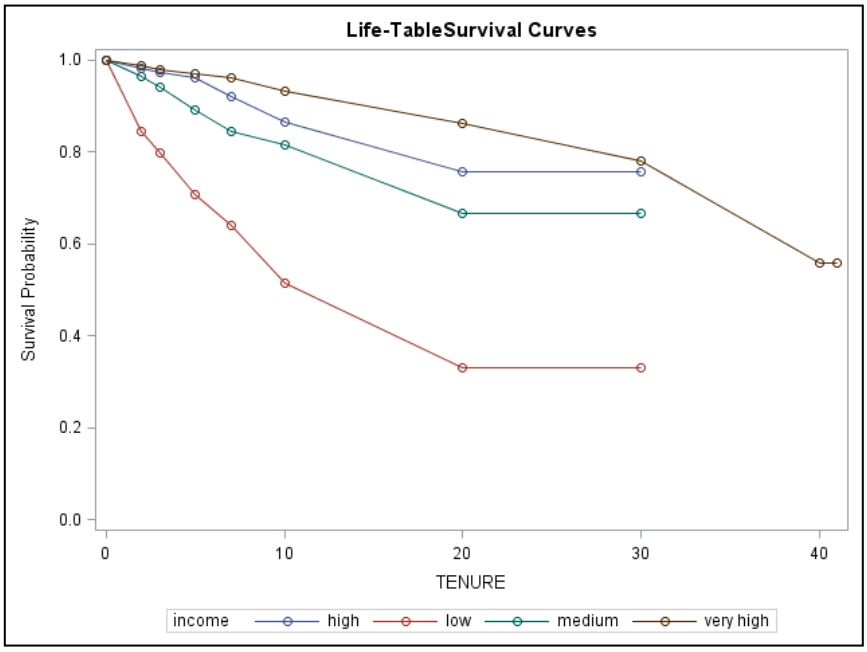
Using Frequency procedure (proc freq) we obtain the frequency table as on the LHS. Here we see that each of the 4 groups has 25% of the employees.

- Low Income group: <=$2911
- Medium Income group: $2911 - $4919

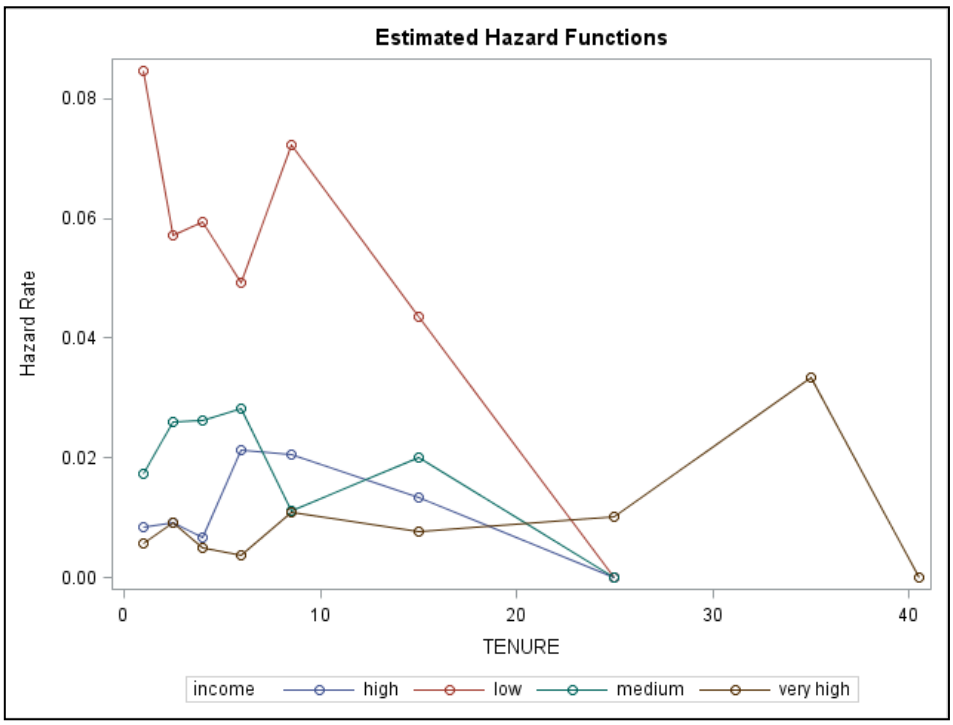
Team 8

- High Income group: $4919 - $8380
- Very High Income group: >$8380

Survival Plot:



Hazard Plot



The hazard of attrition for the low income group is the highest from the very first year. The hazard increases sharply for high income employees after year 6, thereafter declines gradually. For medium income employees the hazard increases sharply after year 2 and then declines sharply after year 8. Again picks up after year 10 and then starts to decline at year 15. For Very high income groups, the hazard remains more or less the same till year

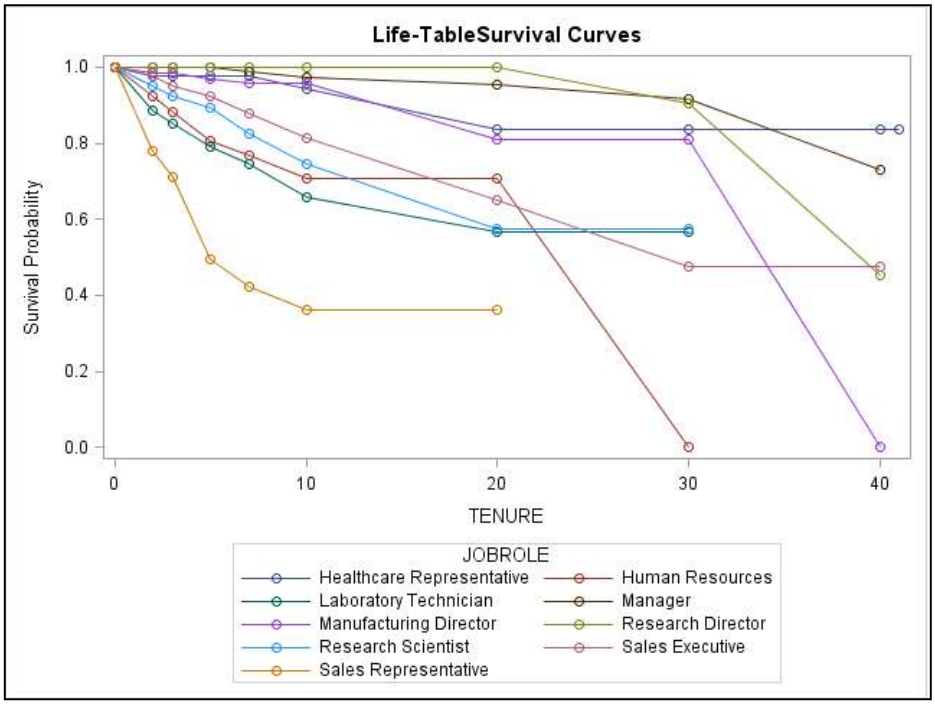
15. **For the initial 15 years, low income & high income employees have the highest & lowers risks of attrition respectively!**

JOB ROLE

The frequency table on the RHS shows that there are 9 job roles present in the company. **Sales Executive, Research Scientist, Laboratory Technician** are the 3 largest job roles accounting for around **60%** of the employees. With only 3.54% of the employees, Human Resources is the smallest group in the company.

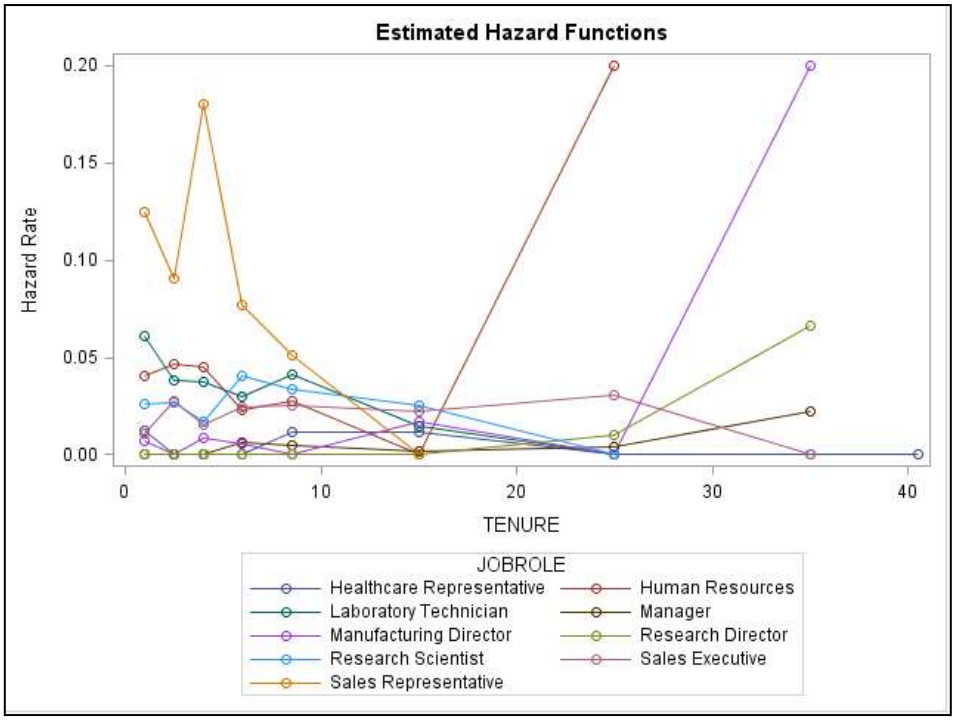
JOBROLE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Healthcare Representative	131	8.91	131	8.91
Human Resources	52	3.54	183	12.45
Laboratory Technician	259	17.62	442	30.07
Manager	102	6.94	544	37.01
Manufacturing Director	145	9.86	689	46.87
Research Director	80	5.44	769	52.31
Research Scientist	292	19.86	1061	72.18
Sales Executive	326	22.18	1387	94.35
Sales Representative	83	5.65	1470	100.00

Survival Plot:



The sales representative job role has a very steep survival probability curve as compared to the other types of jobs.

Hazard Plot



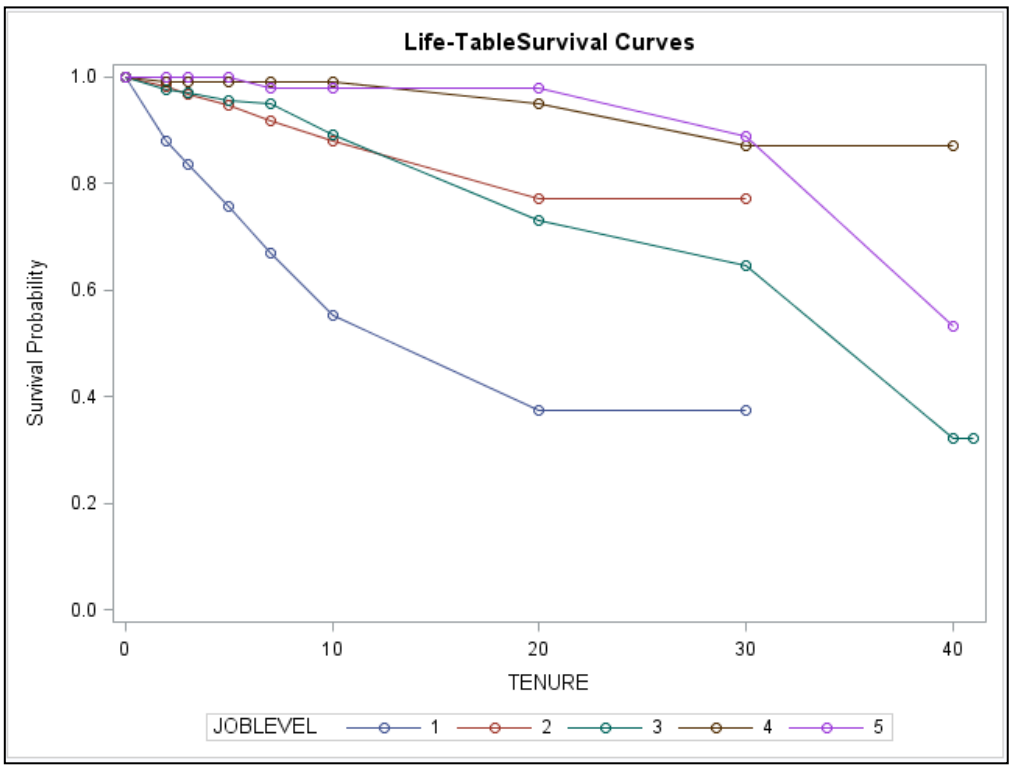
The hazard for Sales Representative is the quite high from the very start. Attrition hazard for Human Resources & Manufacturing Director increases sharply after year 15 and 25 respectively. The hazard for Sales executive increases after year 2 and then declines after year 4. For Laboratory Technician, the hazard decreases sharply after year 2 and then gradually decreases till year 8 after which it again picks up. **For the initial 10 years, Sales representative has the highest risk of attrition while Research Director has a constant 0 risk of attrition!**

JOB LEVEL

JOBLEVEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	543	36.94	543	36.94
2	534	36.33	1077	73.27
3	218	14.83	1295	88.10
4	106	7.21	1401	95.31
5	69	4.69	1470	100.00

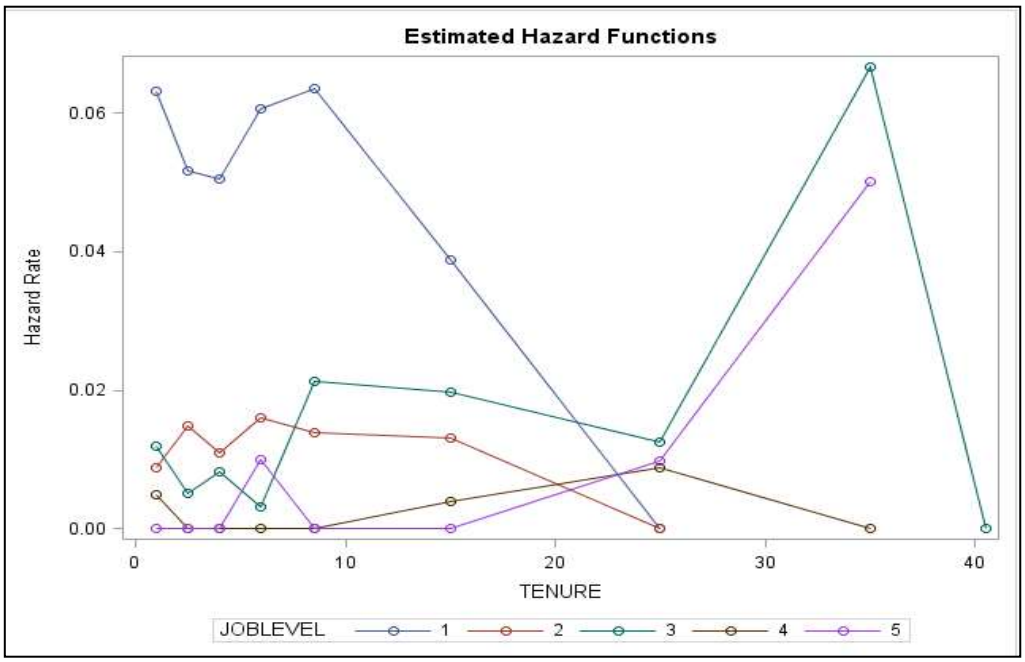
Job Level shows the management level of the employee. From the frequency table on the LHS, we can say that 73.27% of the employees belong to low management levels with Job Level 1,2. In other words, they have junior positions in the company.

Survival Plot:



The survival plot shows a steeper curve for the employees with Joblevel=1. The employees at the lowest rung of the company are more prone to attrition than the higher management.

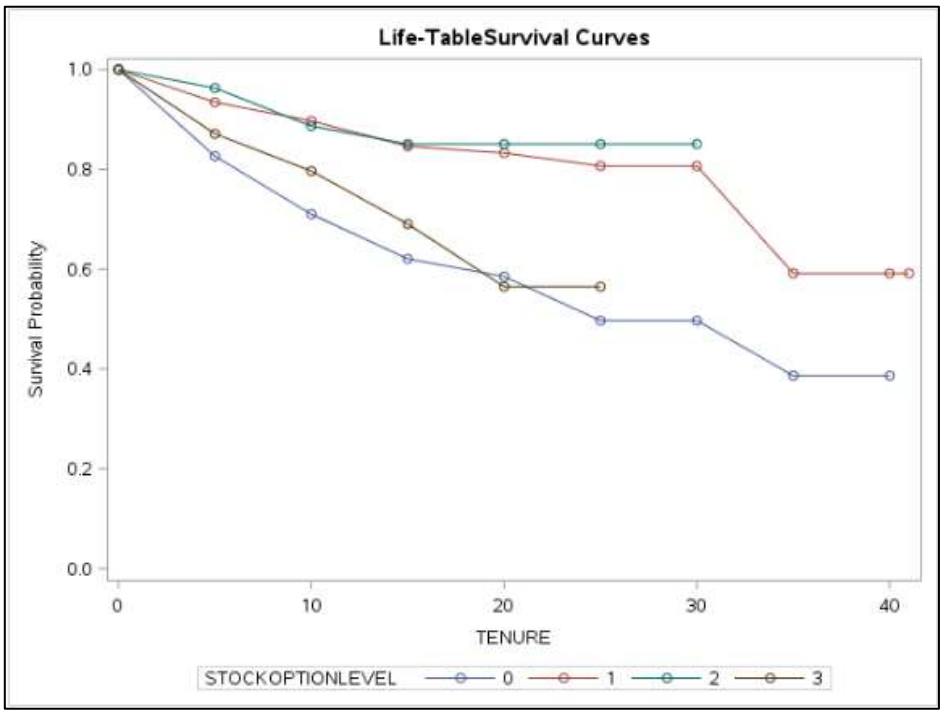
Hazard Plot



The hazard for employees with JobLevel=1 is the highest from the very start till year 20. **Senior most management level employees have 0 risk of attrition for the initial 6 years!**

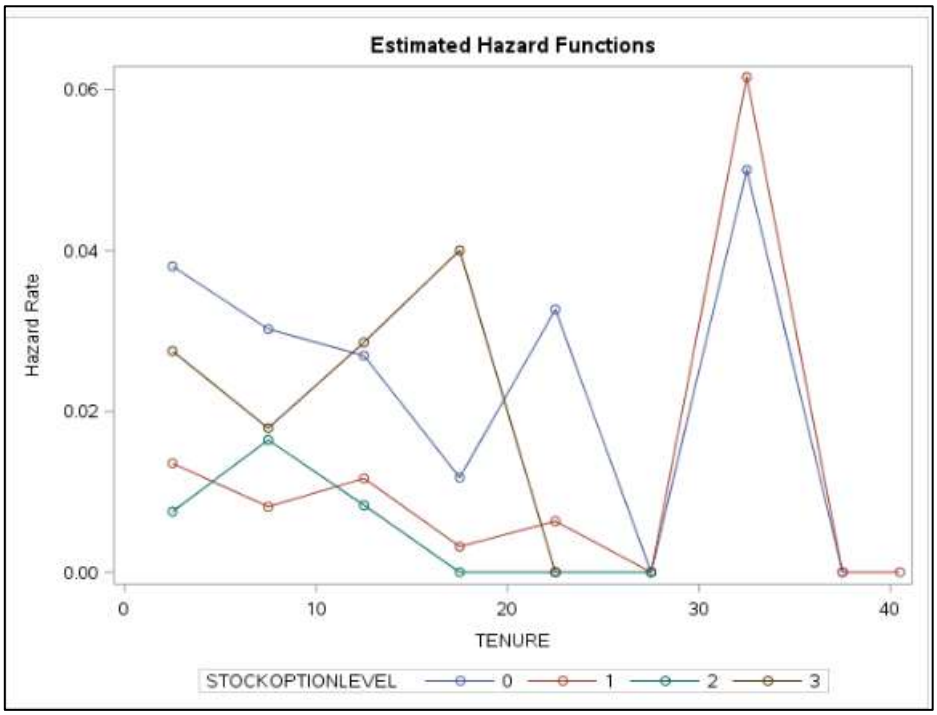
STOCKOPTIONLEVEL

Survival Plot



The survival plot for the stock option level given to each employee of the company shows that there is a sharp decrease in the survival probability of the employees having 0 stock options.

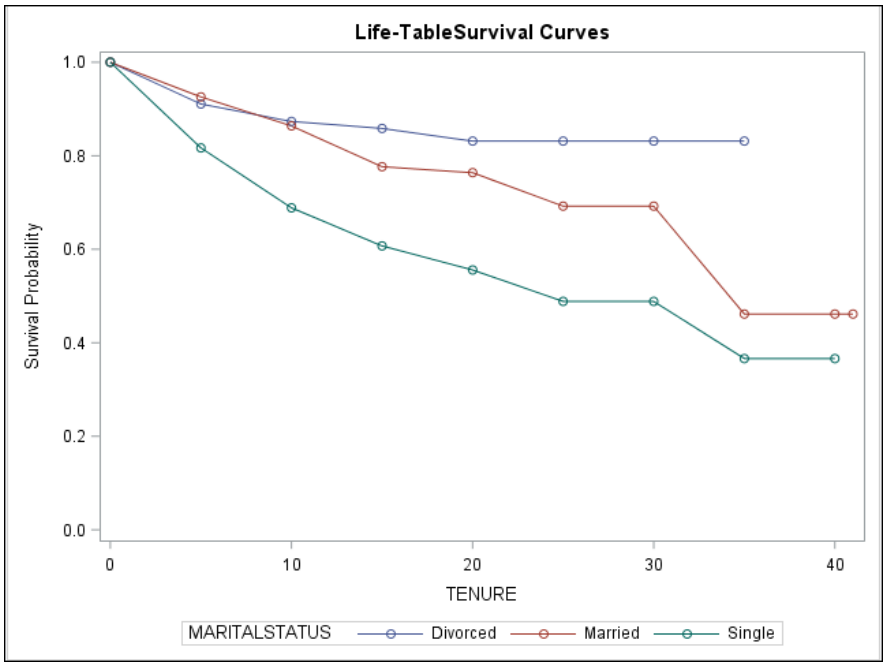
Hazard Plot:



The hazard plot shows that there is high hazard rate for those employees that are new to the company and do not have any stock options provided by the company. These are the kind of employees that are a risk of leaving the company more often.

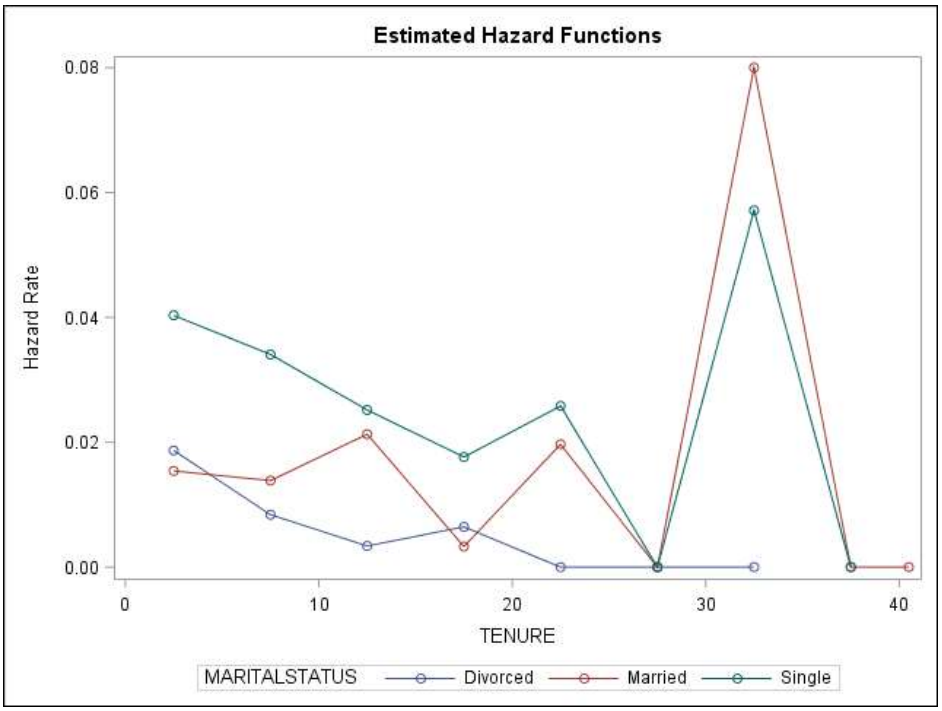
MARITALSTATUS

Survival Plot:



Survival plot shows that single employees have shorter mean survival time and divorced employees have longer mean survival time.

Hazard Plot:

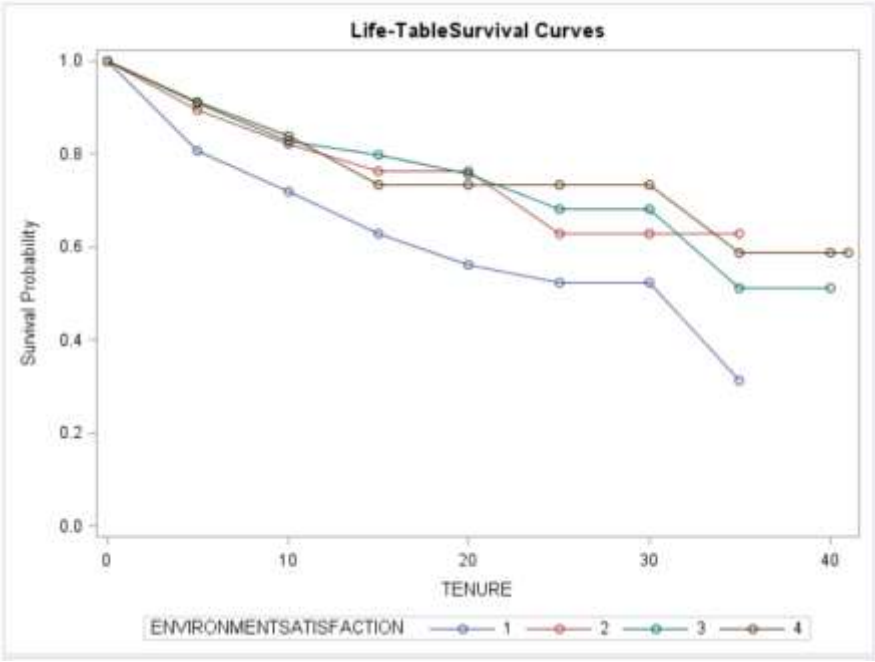


Hazard rate clearly decreases with time divorced employees. There is a spike in hazard rate for married and single employees between 30-35 year's interval.

ENVIRONMENT SATISAFACTION

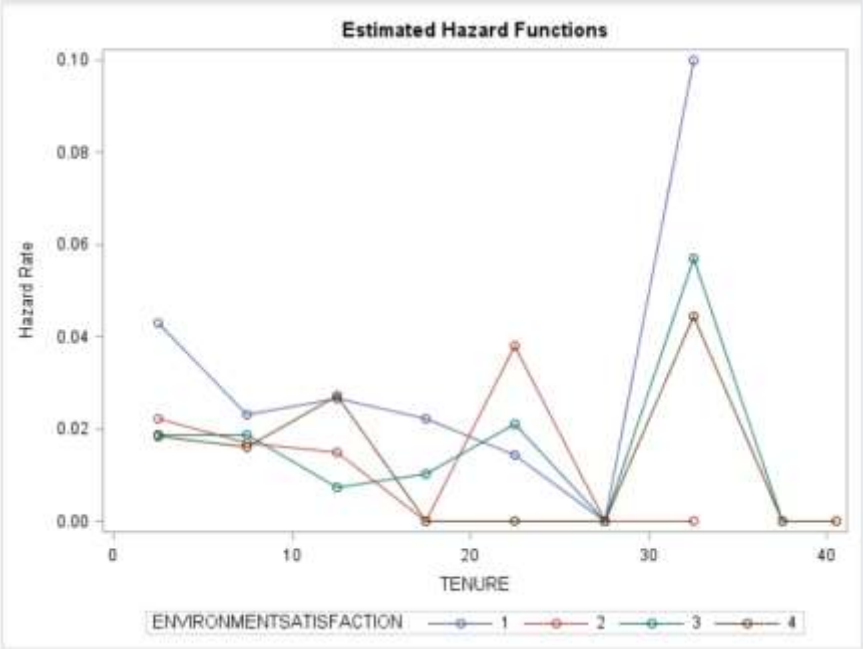
Environment Satisfaction is a categorical variable showing a score showing how much the employee is satisfied with company's facilities

1	'Low'
2	'Medium'
3	'High'
4	'Very High



The survival plots of each of the different levels of environment satisfaction (using the life-table method) is shown. It is observed that there is no significant difference between the medium, high and very high satisfaction curves but the low satisfaction curve is much farther than the rest.

Hazard plots



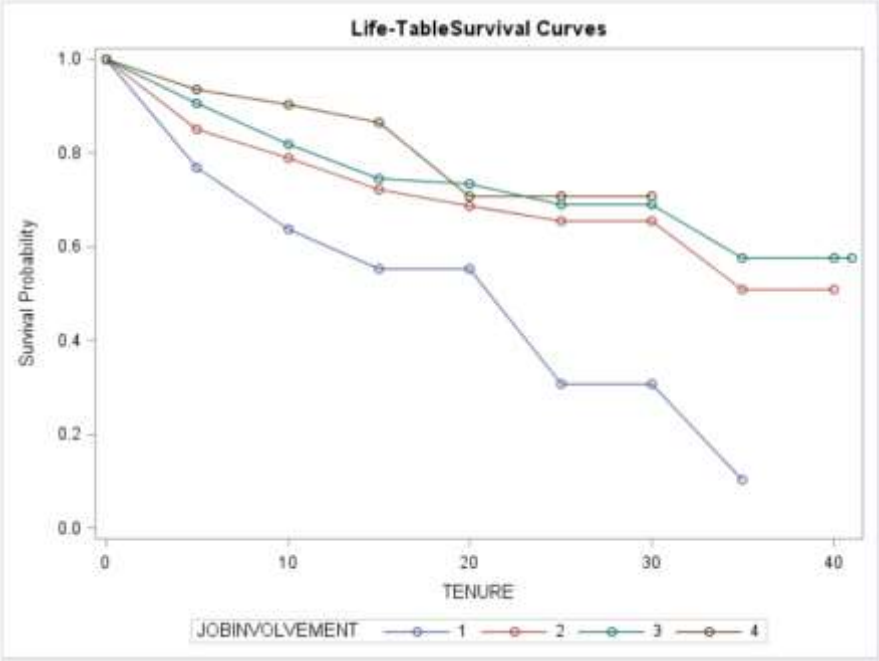
The hazard plot shows a marked increase in the hazard rate at the beginning of the employee's tenure and also those employees who are not really satisfied with their environment. **Environment satisfaction does statistically impact the survival probabilities of employees.**

JOBINVOLVEMENT

Job Involvement is a categorical variable which is a score given to the employee by supervisors how much the employee is involved in company's operations

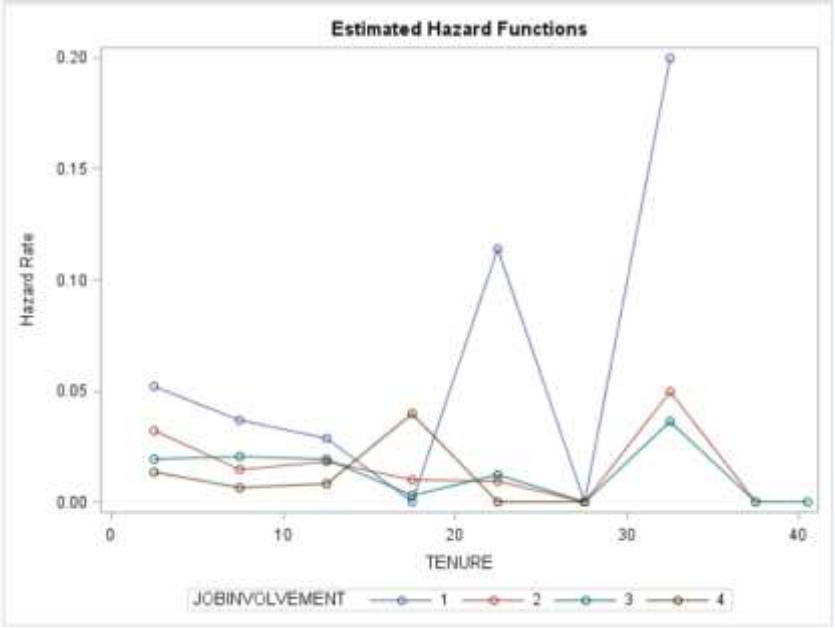
1	'Low'
2	'Medium'
3	'High'
4	'Very High

Survival Plot:



Looking at the survival plots for the different levels of job Involvement, it can be seen that there is a marked difference in the survival probability for lowest level of job involvement when compared to the other levels of job involvement. The other levels of job involvement, i.e. 2,3 and 4 have almost similar survival rates.

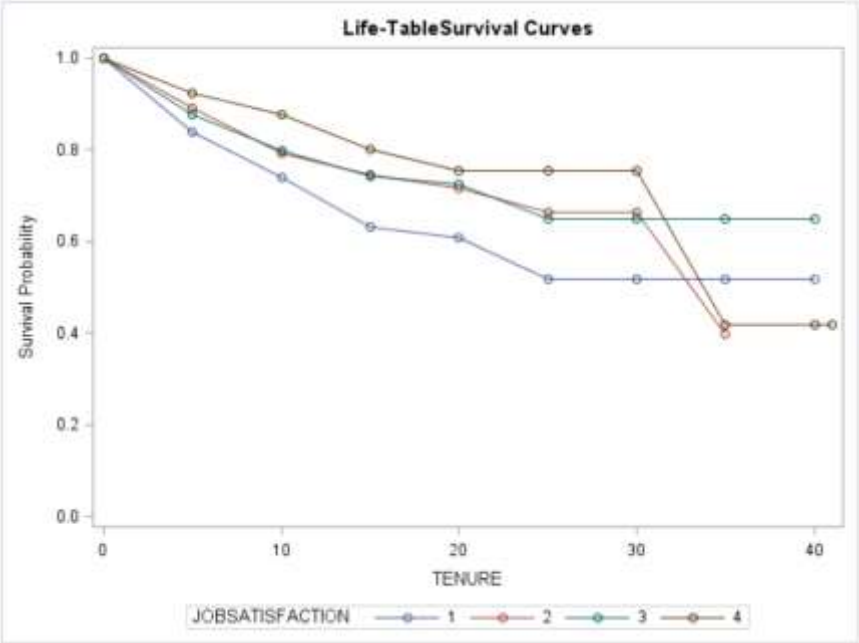
Hazard plot:



The hazard functions also reflect the same story as the survival probability curves. It is observed once again that Job Involvement level 1 has the highest hazard among all the categories.

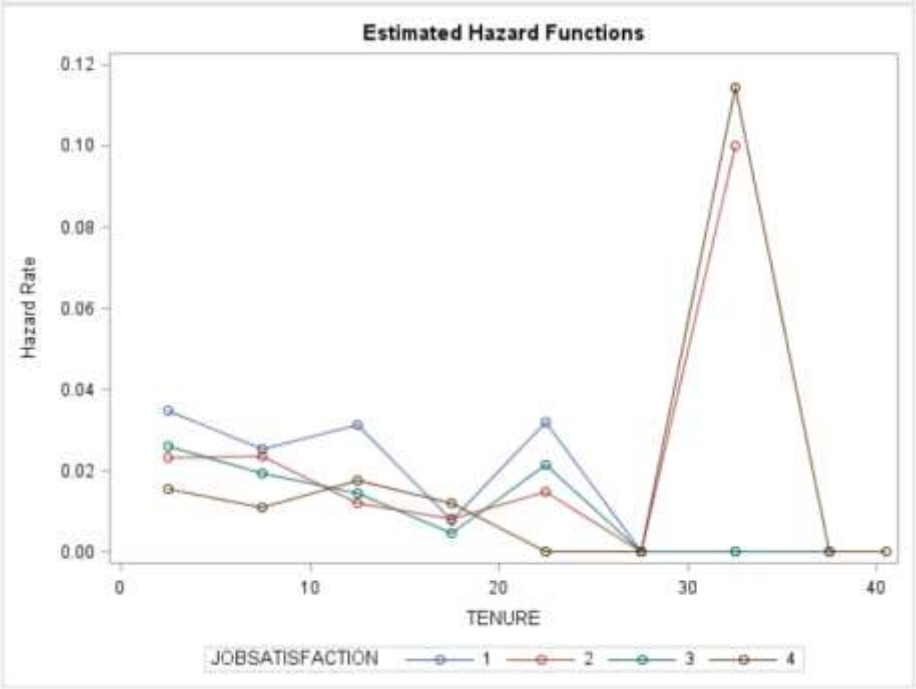
JOB SATISFACTION

Survival Plot



The survival plots for different job satisfaction levels does not show any significant differences in the Life Table method. Surprisingly, lowest satisfaction level demonstrates greater stability as tenure increases.

Hazard plot



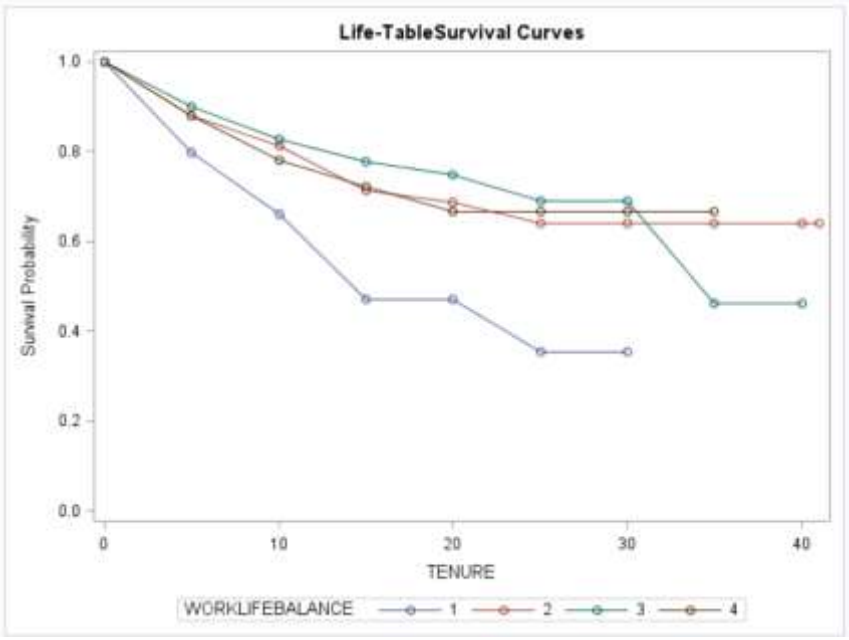
The hazard plots indicate an unstable hazard value for each of the job satisfaction levels. Not much insight can be drawn from these graphs.

WORK-LIFE BALANCE

Work-Life Balance is a categorical variable which shows the employee satisfaction of the work load (4 is the highest satisfaction level)

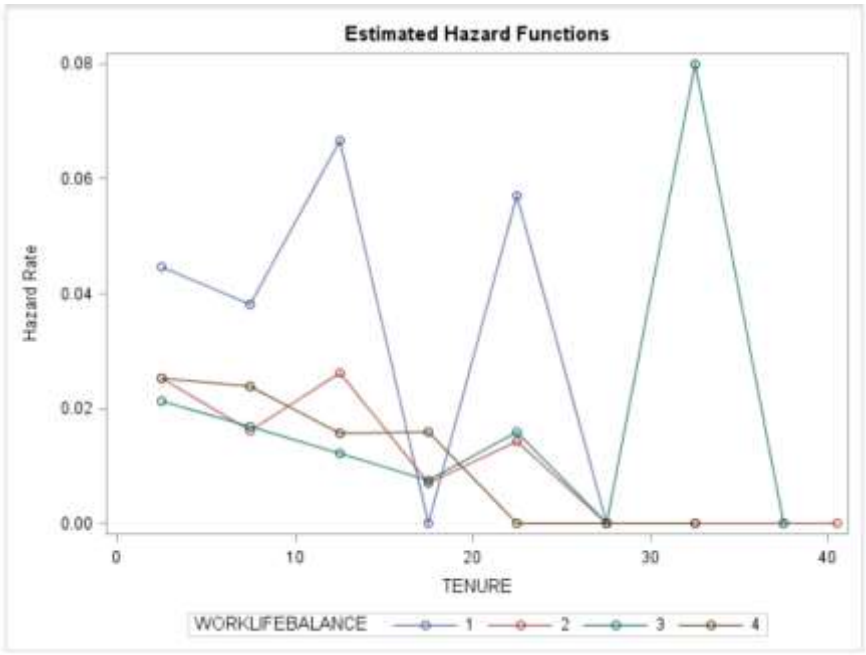
1	'Bad'
2	'Good'
3	'Better'
4	'Best'

Survival Plot:



Survival plots for Life table is as shown – a markedly lower survival rate for low work life balance employees and not much different for other levels.

Hazard plots



All the hazard functions are unstable and do not convey much information. The only insight that can be drawn from these curves is that hazard for low work life balance employees is generally high no matter what the tenure.

VARIABLE GROUPS AND THEIR EFFECTS:

We also divided the data into specific groups based on different kind of variables and checked their individual effects on the employee survival times.

JOB ROLE- "LABORATORY TECHNICIAN" & INCOME GROUP

Since Laboratory technician is one of the largest job roles (17.62% of employees) in the company & their attrition is pretty high, we decided to look at the impact of income on their survival!

Summary of the Number of Censored and Uncensored Values

Stratum	income	Total	Failed	Censored	Percent Censored
1	high	25	3	22	88.00
2	low	130	38	92	70.77
3	medium	104	21	83	79.81
Total		259	62	197	76.06

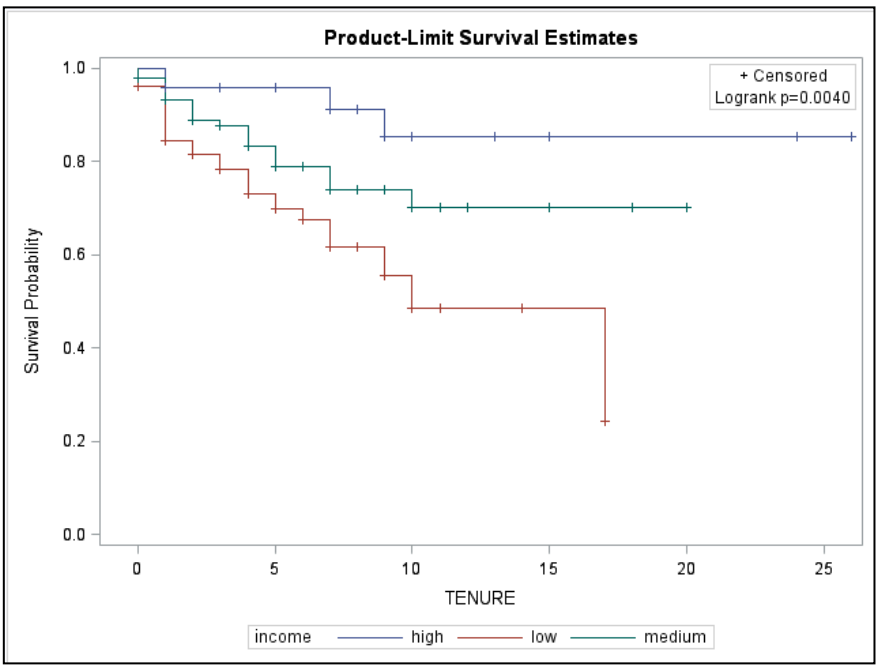
The summary table on the LHS shows that low income group among the Laboratory Technicians have the highest attrition as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.** Also around 50% of them belong to the low income group.

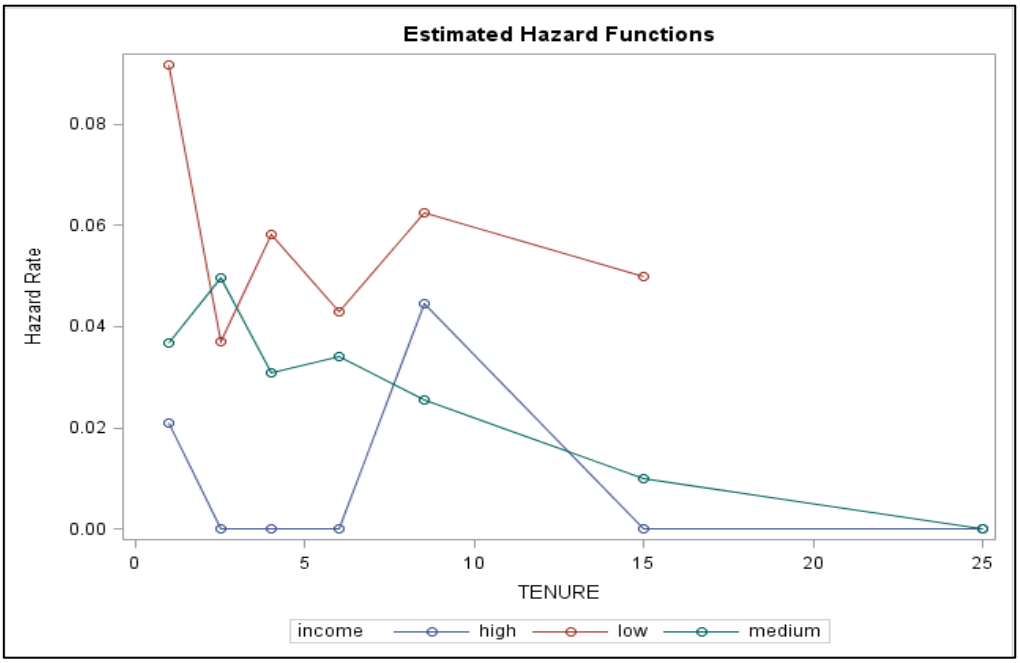
The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant <=0.05 at 5% level of significance. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the income groups!**

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	11.0283	2	0.0040
Wilcoxon	9.0354	2	0.0109
-2Log(LR)	17.3662	2	0.0002

The survival plot on the RHS shows that Low Income Laboratory Technicians attrite at a faster rate over time than the medium and high income ones. **In the first year, 10% of the low income Laboratoty Technicians leave the company.**





Tshe hazard for Low income LTs decreases sharply after 0-2 years. For medium income LTs, the hazard of attrition decreases gradually over time. For high income LTs the hazard is present only in the initial 2 years of their employment.

Thereafter, risk becomes zero except for years 7-11!

JOB ROLE- "RESEARCH SCIENTIST" & INCOME GROUP

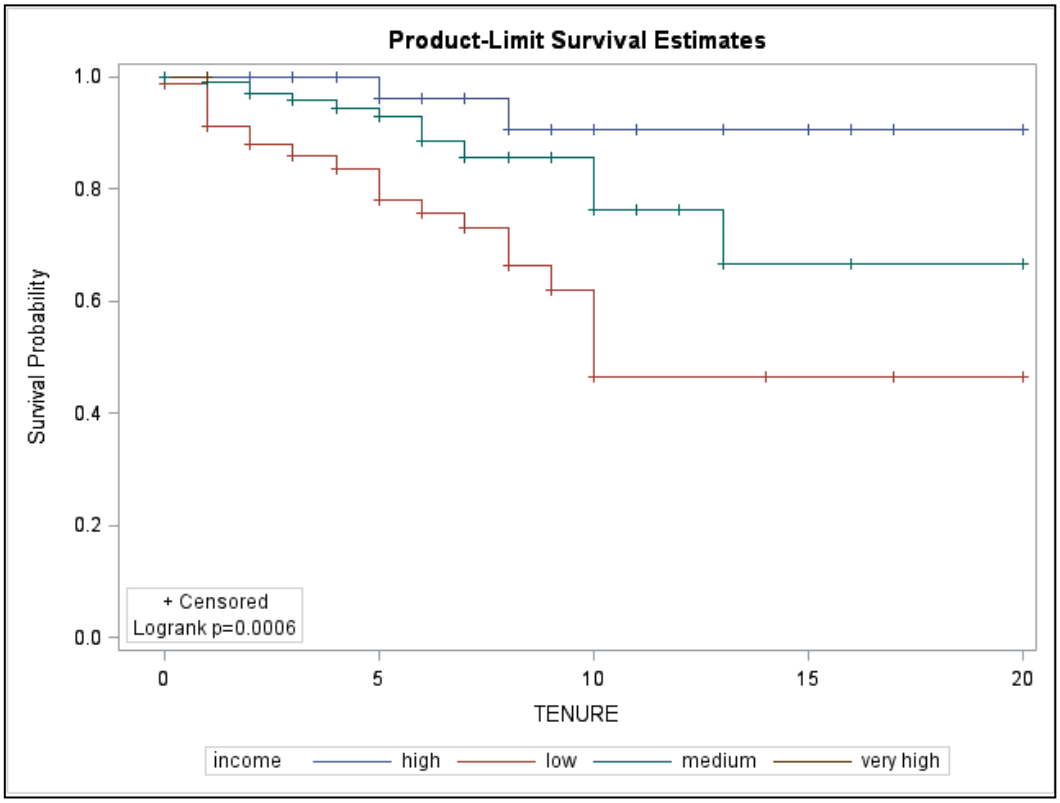
Since Research Scientist is one of the largest job roles (19.86% of employees) in the company, we decided to look at the impact of income on their survival!

Summary of the Number of Censored and Uncensored Values					
Stratum	income	Total	Failed	Censored	Percent Censored
1	high	30	2	28	93.33
2	low	152	33	119	78.29
3	medium	109	12	97	88.99
4	very high	1	0	1	100.00
Total		292	47	245	83.90

The summary table on the LHS shows that low income group among the Research Scientists have the highest attrition as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.** Also **around 86% of them belong to the low& medium income groups**.

The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant <=0.05 at 5% level of significance. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the income groups!**

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	17.4400	3	0.0006
Wilcoxon	15.6159	3	0.0014
-2Log(LR)*	17.9105	3	0.0005



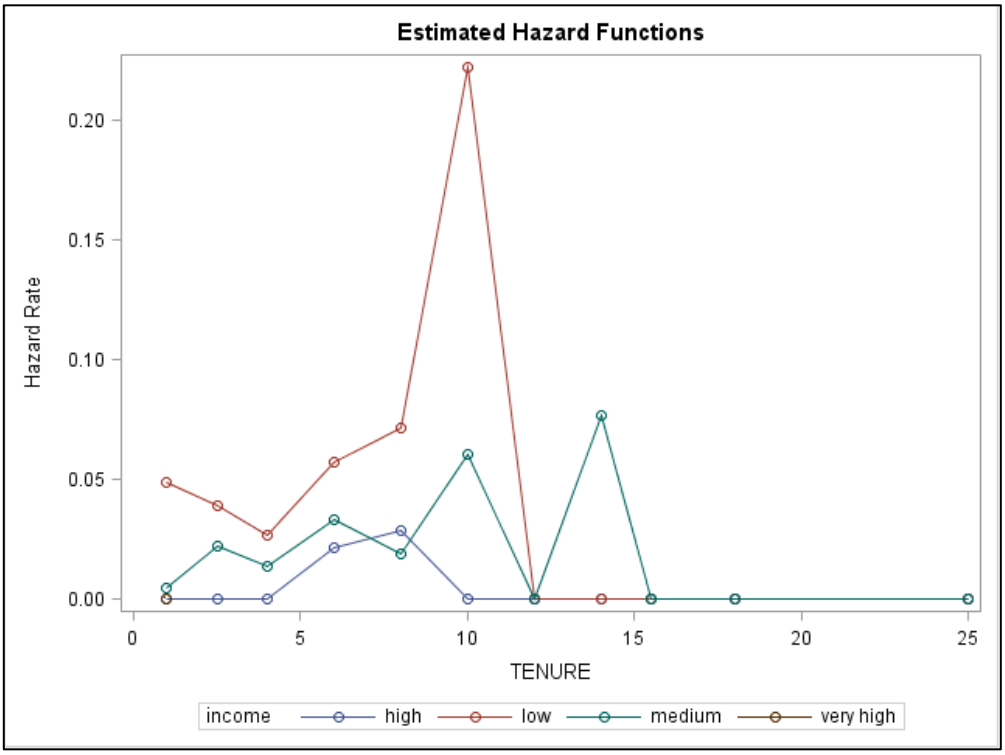
The survival plot on the LHS shows that Low Income Research Scientists attrite at a faster rate over time than the medium and high income ones.

In first year, 10% of the low income Research Scientists leave the company.

Team 8

Hazard Plot:

The hazard for low income Research Scientists decrease till year 5 and then picks up till year 11. The attrition risk for high income RSs is 0 constant for initial 5 years after which it increases in years 5-9 and thereafter becomes zero again. **Low income RS's have the highest hazard and high income RS's have 0 hazard for the initial 5 years!**



JOB ROLE- "SALES EXECUTIVE" & INCOME GROUP

Since Sales Executive is the largest job role (22.18% of employees) in the company, we decided to look at the impact of income on their survival!

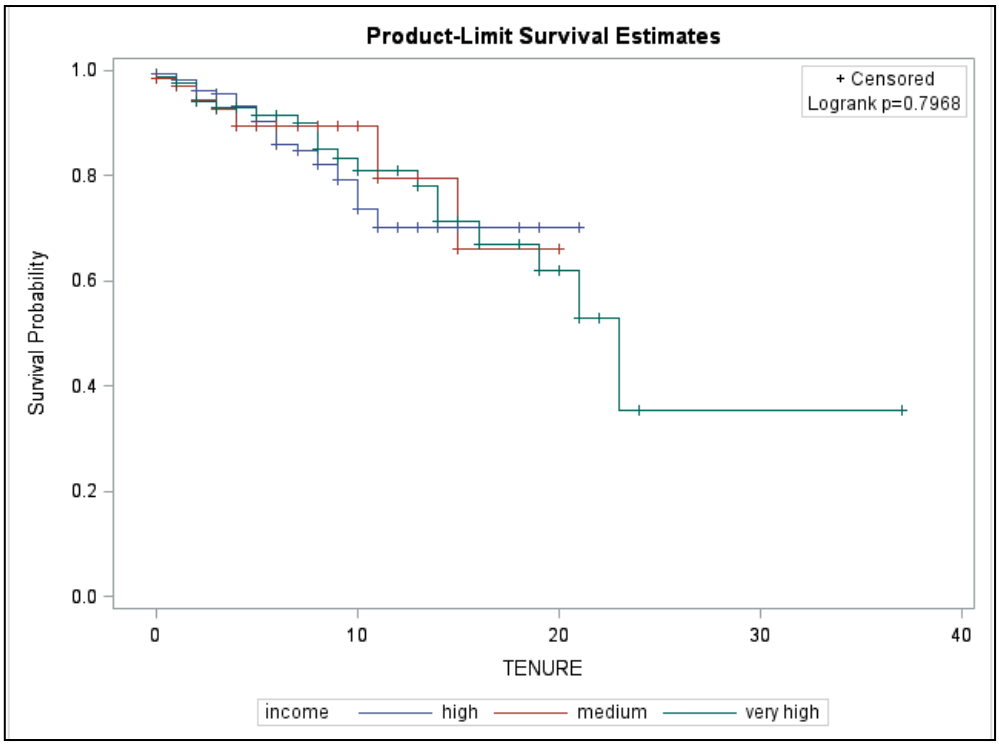
Summary of the Number of Censored and Uncensored Values

Stratum	income	Total	Failed	Censored	Percent Censored
1	high	165	28	137	83.03
2	medium	73	9	64	87.67
3	very high	88	20	68	77.27
Total		326	57	269	82.52

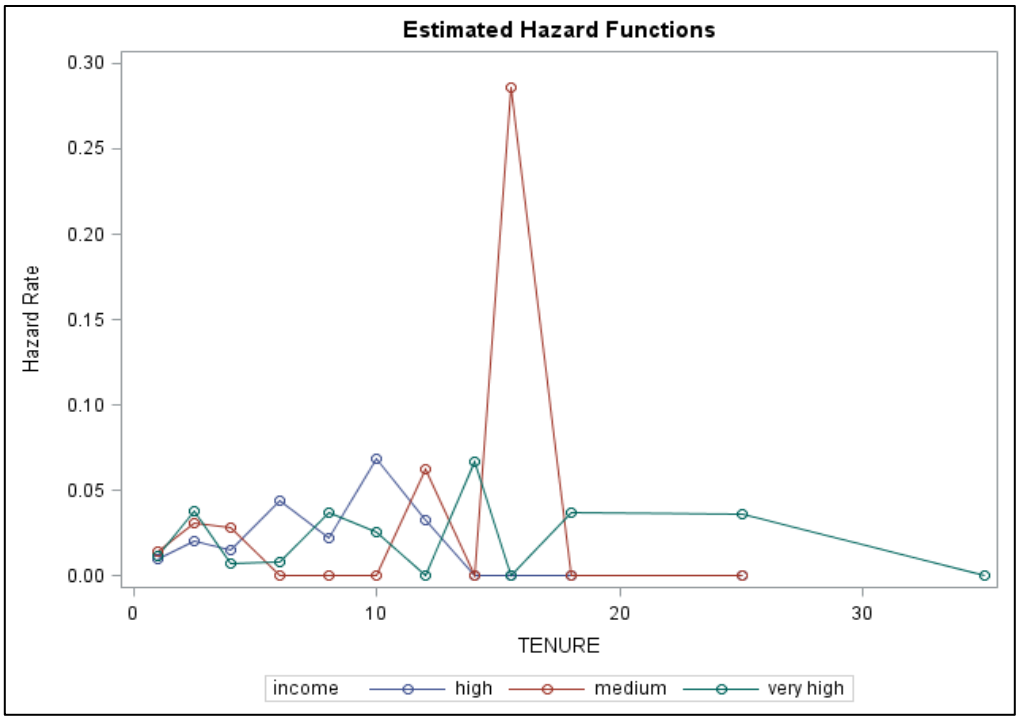
The summary table on the LHS shows that very high income group among the Sales executives have the highest attrition as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

The probability values of the Chi square statistics for all the three tests as shown on the RHS is insignificant >=0.05 at 5% level of significance. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is NOT different across the income groups!**

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.4544	2	0.7968
Wilcoxon	0.2375	2	0.8880
-2Log(LR)	0.2815	2	0.8687



The survival plot also shows that attrition pattern is not significantly different among the income groups!



Hazard rate is low for all the income groups of Sales Executives for the initial 2 years of employment after which the risk of attrition increases!

Department = "Human Resources" & Marital Status

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
MARITALSTATUS	2	1.7082	0.4257

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	5.1547	1.4435	2.3255	7.9839	12.75	0.0004
MARITALSTATUS	Divorced	1	-1.7667	1.4426	-4.5941	1.0606	1.50	0.2207
MARITALSTATUS	Married	1	-1.1023	1.3792	-3.8055	1.6008	0.64	0.4241
MARITALSTATUS	Single	0	0.0000	.	.	.	.	.
Scale		1	1.2506	0.2970	0.7852	1.9920		
Weibull Shape		1	0.7996	0.1899	0.5020	1.2736		

As per the Wald chi-square test statistics, we can conclude that Marital Status does not impact the tenure of an employee in Human Resources department.

Department = Research and Development & Marital Status

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
MARITALSTATUS	2	19.9992	<.0001

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	3.4978	0.1657	3.1731	3.8225	445.80	<.0001
MARITALSTATUS	Divorced	1	1.0788	0.2860	0.5182	1.6394	14.23	0.0002
MARITALSTATUS	Married	1	0.7448	0.2070	0.3391	1.1505	12.95	0.0003
MARITALSTATUS	Single	0	0.0000	.	.	.	.	.
Scale		1	1.0304	0.0754	0.8928	1.1893		
Weibull Shape		1	0.9705	0.0710	0.8408	1.1201		

As per the wald chi-square test statistics, we can conclude that Marital Status does impact the tenure of an employee in Human Resources department.

Estimate Coefficient Interpretation:

- Expected survival time increases by 194.12% for divorced employee compared to single employee
- Expected survival time increases by 110.60% for married employee compared to single employee

Department = Sales & Marital Status

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
MARITALSTATUS	2	16.4337	0.0003

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	3.0798	0.1575	2.7711	3.3885	382.30	<.0001
MARITALSTATUS	Divorced	1	1.0076	0.3501	0.3214	1.6938	8.28	0.0040
MARITALSTATUS	Married	1	0.8732	0.2408	0.4013	1.3451	13.15	0.0003
MARITALSTATUS	Single	0	0.0000	.	.	.	.	.
Scale		1	0.9686	0.0843	0.8167	1.1489		
Weibull Shape		1	1.0324	0.0899	0.8704	1.2245		

As per the wald chi-square test statistics, we can conclude that Marital Status does impact the tenure of an employee in Sales department.

Estimate Coefficients Interpretation:

- Expected survival time increases by 173.90% for divorced employee compared to single employee
- Expected survival time increases by 139.46% for married employee compared to single employee

Additional Findings

Finding 1

Attribute	% increase in Survival Time	
	Sales	Research & Development
OVERTIME = No	87%	172%

The Survival models predict an interesting phenomenon. It is observed that the survival is drastically improved almost about 200% if working overtime is discouraged among employees in the Research and Development department as compared to employees involved in Sales. This is legit considering the fact that the value that the R&D employees bring to the organization cannot be measured by the number of hours put in by them. Sales, on the other hand rely heavily on negotiations and the more hours are spent in that, the better it is for the organization.