Project 1

Survival Analysis for 'Fermalogis'

OPIM 5894 - Survival Analysis with SAS

Group 9

AUTHORS

Rajarshi Das
Xueling Chen
Yu Li
Vinit Gupta
Rahul Manchanda
Pawan Shivhare

UNIVERSITY OF CONNECTICUT, SCHOOL OF BUSINESS
SUMMER, 2017

Table of Contents

Executive Summary

The investigation on employee attrition is focused on the ability to predict the employee survival time and describe the reasons behind the attrition measuring variables such as job satisfaction level, business travel, working overtime and employee working department. In this Survival analysis time is a variable of interest and is designed to measure the attrition rate in Fermalogis.

From the Fermalogis dataset in the given time frame, observed that around 16% of employees are leaving the company and the key focus is on two segments of employees. First, is the group of young employees associated with the company for almost 3 years and second is the group of highly experienced employees who are with the company for more than 5 years. Even though the employees are treated equally, different other factors are extremely important for being a contributor towards the attrition. The analysis is capable of indicating the differences between the young and experienced employees based on predictor variables collected from the employee records of this company.

Based on all these analysis, from the different sets of the analysis results, it has been identified, that, working overtime and frequent travels are few major factors which are driving the employee to leave the company. Further, the results also show that inside the company, there is culture of bias, between the young and experienced employees, when it comes to decision on the special perks like, granting stock options, which contributed as well for the attrition of young employees. On the other hand, the attrition of experienced employee is influenced by work environment dissatisfaction, specially, when it comes to interaction with coordination with departments like Sales, HR etc. And have also identified, that young employee group leaves the company comparatively early than the experienced ones.

This report explains the sequential process, and followed, to arrive at the conclusion on how various factors contributed towards being the reasons for employee attrition and at what instance of time. The report also emphasized the results with significant and industry standard statistical methods.

1. Project Objective

The objective of this survival analysis project is to solve essential business questions which will help chief operating officer of "FermaLogis" company to understand who are leaving the company and for what reason, and come out with better strategies to improve situation of staff turnover.

2. Description

In order to come up with the answers, few important insights are taken out of the project dataset which will apparently answer the question - who are leaving the company and why are they leaving.

Following are the subset of these additional questions which are answered from the analysis done by the project:

a. Who are leaving the company?
b. Why are they leaving?
c. When is the biggest danger for employees to leave?
d. Is there any difference of attrition between different employee groups or categories?
e. Any other findings about the company.

3. Data Source

Employee dataset is provided by the chief operating officer of the company named "FermaLogis". The dataset consists of information about the past and present employees of the company and different attributes of the employee which have been explained in the following section. (Section 4 of this report).

The holistic details of the different attributes of the employee data set can be find in Table 1 in Appendix.

4. Project Approach

- ✓ For this project conventional SEMMA aspect of data analysis is followed in order to get the business insights from the dataset.
- ✓ Actual dataset is provided in .CSV file format. The project team converted this to a compatible SAS .bat file format, which was necessary to continue the analysis of the data in SAS base environment.
- ✓ Brief details of SEMMA aspect of data analysis are provided in this section of the report.
- ✓ Detailed steps and inferences drawn from those steps are also explained in subsequent sections of this report as applicable.

- **Sample:**

  Sampling is the first step of the dataset analysis and modeling process where whole dataset is divided into 3 main part i.e. 'Training dataset', 'Validation dataset', and 'testing dataset'. But for this given project, haven't done any sampling as it was not necessary for the survival analysis approach so planned to execute with the data.

- **Explore:**

  Under the explore phase, past and present employee data is analyzed. After understanding the data, deeper understanding is gained for individual variables and also relationships between the variables. COO explained that business is getting affected by attrition in two different ways based on the employee's tenure at the company. Employees with tenure at the company less or equal to 3 years are considered young and greater than 3 years are considered Experienced. These two sectors are explored separately to understand the behavior and factors that contributed towards their attrition.

- **Modify:**

  After the data exploration which included paying a close look over each of the variable, few new variables are introduced and few variables are omitted from the project execution steps. Few important variables are identified for building the survival models. The dataset is divided into sections as have observed that the reason behind employee's attrition according to the exploration are different for Young and Experienced employees.

- **Model:**

  Under the modelling phase, the project has executed the following 3 industry standard modeling algorithms for analyzing the survival analysis aspect of the data:

  - ✓ Exponential Model
  - ✓ Weibull Model
  - ✓ Log-Normal Model

- **Assess:**

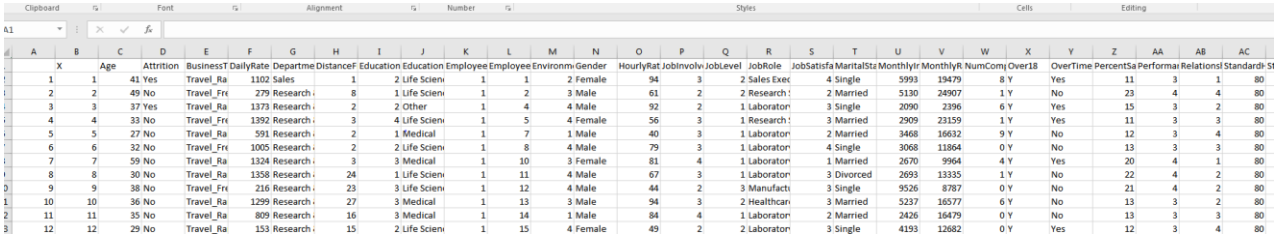
  Under the assess phase, statistical factors are compared among different models. This helped in evaluating the **effectiveness, reliability and usefulness of the survival models**. Better survival model leads to better survival analysis of the existing employee attrition behavior.

5. Assumptions

- ✓ Employees having tenure of less than or equal to 3 years are considered as 'Young Employees'

6. Data Preparation (Explore-Modify Phase)-dataset format change

- ✓ The original data is provided as Comma Separated value file.
- ✓ In order to access dataset easily the CSV file is converted into SAS7bdat file.
- ✓ Here is the sample view of CSV file: FermaLogis.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
		X	Age	Attrition	BusinessT	DailyRate	Departme	DistanceF	Education	Education	Employee	Employee	Environm	Gender	HourlyRat	JobInvolv	JobLevel	JobRole	JobSatisfa	MaritalSta	MonthlyIr	MonthlyR	NumComp	Over18	OverTime	PercentSa	Performar	RelationsI	StandardH
	1	1	41	Yes	Travel_Ra	1102	Sales	1	2	Life Scien	1	1	2	Female	94	3	2	Sales Exec	4	Single	5993	19479	8	Y	Yes	11	3	1	80
	2	2	49	No	Travel_Fre	279	Research	8	1	Life Scien	1	2	3	Male	61	2	2	Research	2	Married	5130	24907	1	Y	No	23	4	4	80
	3	3	37	Yes	Travel_Ra	1373	Research	2	2	Other	1	4	4	Male	92	2	1	Laboratory	3	Single	2090	2396	6	Y	Yes	15	3	2	80
	4	4	33	No	Travel_Fre	1392	Research	3	4	Life Scien	1	5	4	Female	56	3	1	Research	3	Married	2909	23159	1	Y	Yes	11	3	3	80
	5	5	27	No	Travel_Ra	591	Research	2	1	Medical	1	7	1	Male	40	3	1	Laboratory	2	Married	3468	16632	9	Y	No	12	3	4	80
	6	6	32	No	Travel_Fre	1005	Research	2	2	Life Scien	1	8	4	Male	79	3	1	Laboratory	4	Single	3068	11864	0	Y	No	13	3	3	80
	7	7	59	No	Travel_Ra	1324	Research	3	3	Medical	1	10	3	Female	81	4	1	Laboratory	1	Married	2670	9964	4	Y	Yes	20	4	1	80
	8	8	30	No	Travel_Ra	1358	Research	24	1	Life Scien	1	11	4	Male	67	3	1	Laboratory	3	Divorced	2693	13335	1	Y	No	22	4	2	80
0	9	9	38	No	Travel_Fre	216	Research	23	3	Life Scien	1	12	4	Male	44	2	3	Manufact	3	Single	9526	8787	0	Y	No	21	4	2	80
1	10	10	36	No	Travel_Ra	1299	Research	27	3	Medical	1	13	3	Male	94	3	2	Healthcar	3	Married	5237	16577	6	Y	No	13	3	2	80
2	11	11	35	No	Travel_Ra	809	Research	16	3	Medical	1	14	1	Male	84	4	1	Laboratory	2	Married	2426	16479	0	Y	No	13	3	3	80
3	12	12	29	No	Travel_Ra	153	Research	15	2	Life Scien	1	15	4	Female	49	2	2	Laboratory	3	Single	4193	12682	0	Y	Yes	12	3	4	80

- ✓ Here is the sample view of SAS7bdat file: team9_csv.sas7bdat

Total rows: 1470 Total columns: 85 Rows 1-100

	,	X	Age	Attrition	BusinessTravel	DailyRate	Department	Dis
1	1	1	41	Yes	Travel_Rarely	1102	Sales	1
2	2	2	49	No	Travel_Frequently	279	Research & Development	8
3	3	3	37	Yes	Travel_Rarely	1373	Research & Development	2
4	4	4	33	No	Travel_Frequently	1392	Research & Development	3
5	5	5	27	No	Travel_Rarely	591	Research & Development	2
6	6	6	32	No	Travel_Frequently	1005	Research & Development	2
7	7	7	59	No	Travel_Rarely	1324	Research & Development	3
8	8	8	30	No	Travel_Rarely	1358	Research & Development	24
9	9	9	38	No	Travel_Frequently	216	Research & Development	23
10	10	10	36	No	Travel_Rarely	1299	Research & Development	27
11	11	11	35	No	Travel_Rarely	809	Research & Development	16
12	12	12	29	No	Travel_Rarely	153	Research & Development	15
13	13	13	31	No	Travel_Rarely	670	Research & Development	26
14	14	14	34	No	Travel_Rarely	1346	Research & Development	19

7. Data Preparation (Explore-Modify Phase)-*addition of manufactured variables*

- ✓ For better usage of the dataset, 4 additional manufactured variables are added. The new variable details are provided below:

Table 2:

Column name	Derived based on column	Data type	Condition	Sample value
NumCompaniesWorked1	NumCompaniesWorked	Character	If the value of variable 'NumCompaniesWorked' is greater than equal to 4 years then the value of computed column would be 'Less than or equal to 4 Companies'. For all other cases new column will have value -'Greater than 4 Companies'.	Greater than 4 Companies
NumBonusAwarded	bonus_1 to bonus_40	Numeric	Sum of value (bonus_1 to bonus_40)	5
BonusReceivedRatio	NumBonusAwarded & YearsAtCompany	Numeric	(NumBonusAwarded / YearsAtCompany)	0.2777777778

Datasplit	YearsAtCompany	Character	If the value of 'YearsAtCompany' field is less than equal to 3 Then the value of new 'Datasplit' field will be 'YOUNG_EMP'. And if the value of 'YearsAtCompany' field is greater than 3 Then new 'Datasplit' will be 'Exp_Emp'.	YOUNG_EMP

8. Data Preparation (Explore-Modify Phase) – *Identified insignificant variables.*

- ✓ There are 75 variables in the dataset and 4 variables are added additionally. However, for analysis and modeling purposes only important variables are considered. There are few variables only consists of static values, these variables are neither helpful to analysis data, nor helpful to build the survival model. Following are the list of variables which are not considered significant in any of the subsequent project phases.

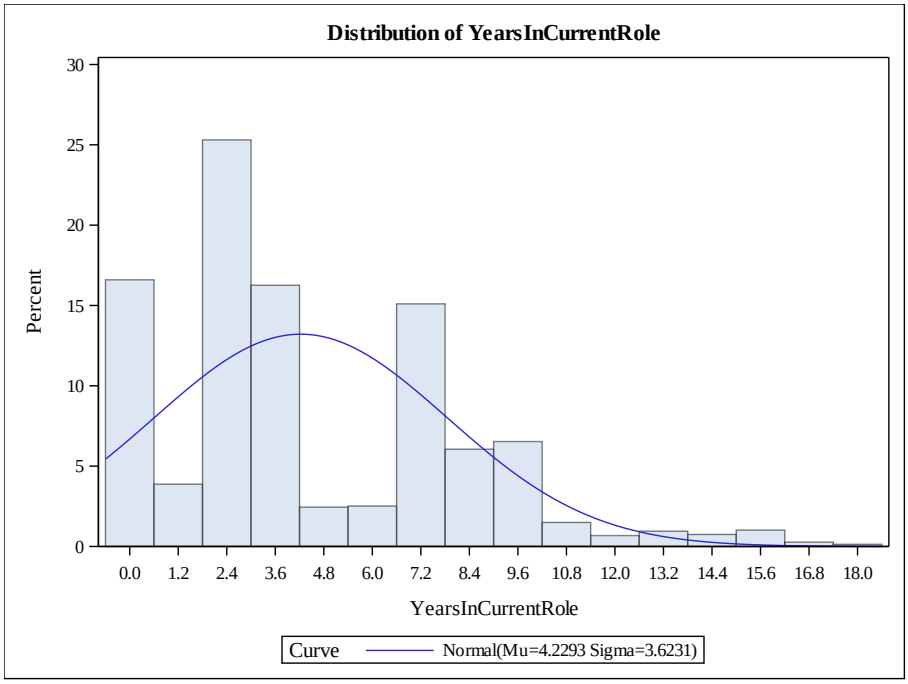
Table 3:

Column name	Data type	Reason
EmployeeCount	Numeric	All rows have value as 1. The variance is 0.
EmployeeNumber	Numeric	This is just incremental sequence starting from 1. This doesn't help in analysis. Everything is unique.
Over18	Character	This value is 'Y' for every employee.
StandardHours	Numeric	This holds value 80 for every employee. The variance is 0.

9. Explore Phase: Distribution and key observations

Variable: YearsInCurrentRole

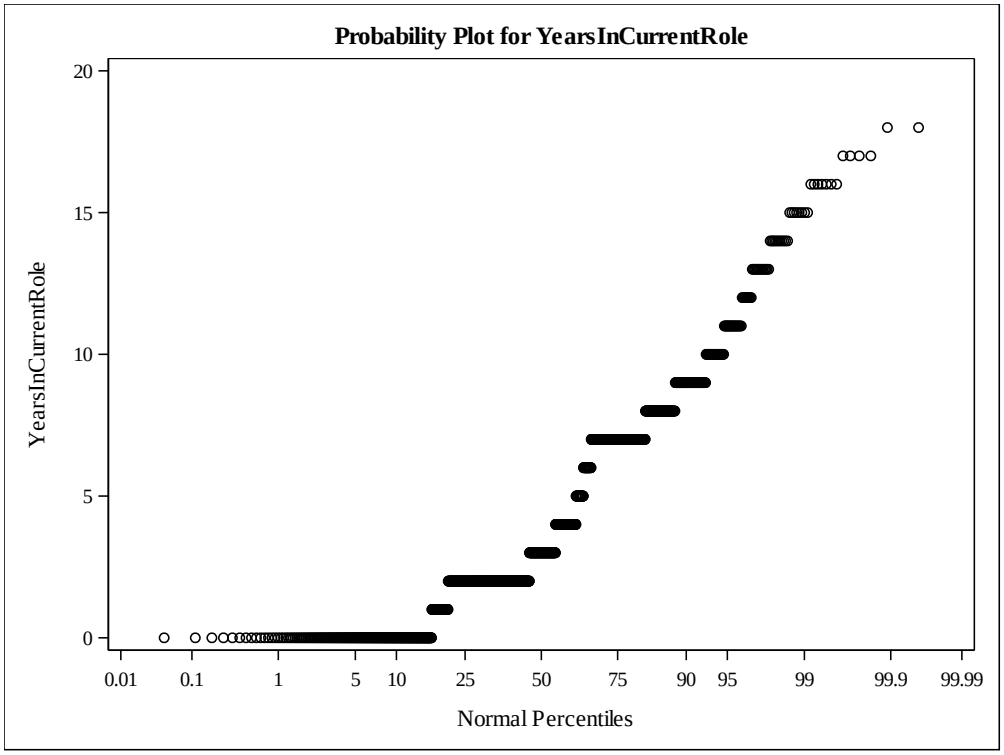
Here is the distribution for variable YearsInCurrentRole. As per the plot over 25% of the employees are around 2.4 years with the company. Around 43% of the employees are in the current role for less than or equal to 3 years.



As given in statistical measurement table, mean number of years in current role is 4 years and employees are in range of 18 years or the current role.

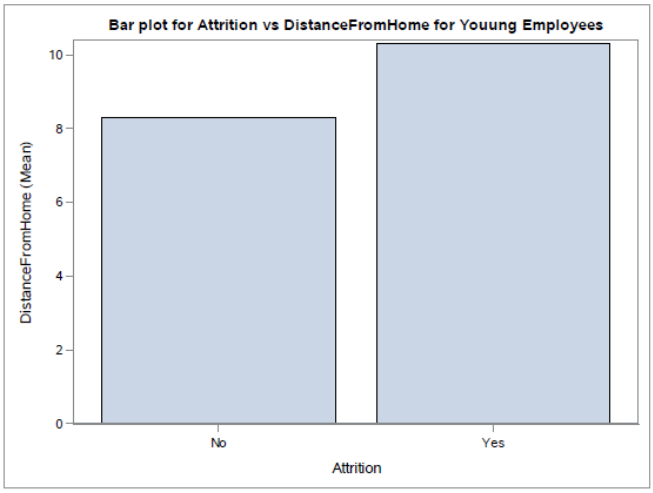
Basic Statistical Measures			
Location		**Variability**	
Mean	4.229252	**Std Deviation**	3.62314
Median	3.000000	**Variance**	13.12712
Mode	2.000000	**Range**	18.00000
		Interquartile Range	5.00000

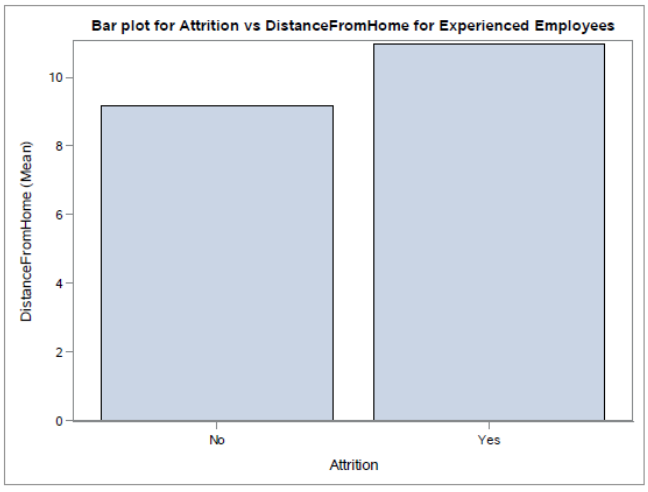
Here is the probability plot for the current variable:



Variable: DistanceFromHome

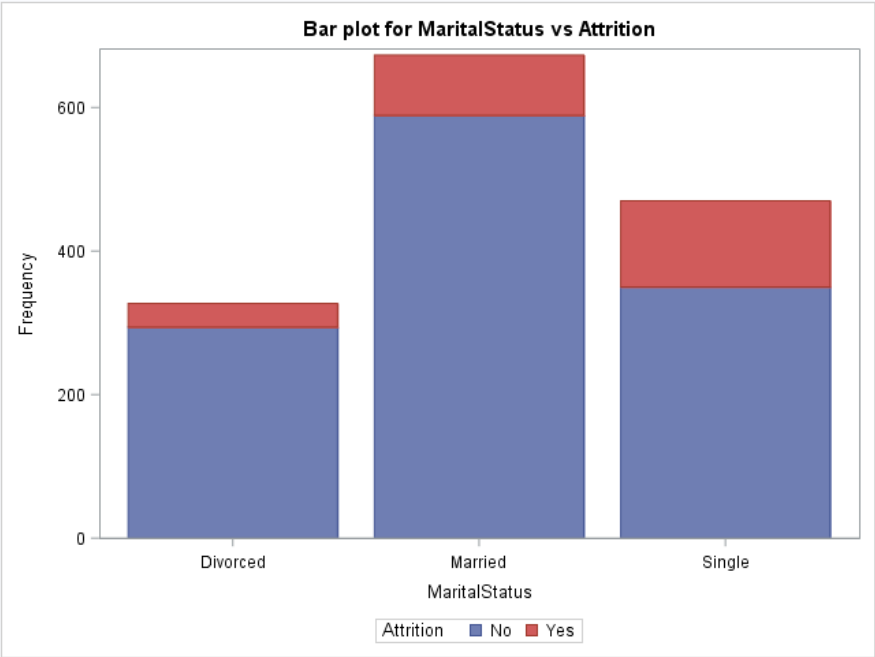
As commute distance is a very important factors among all employees, below plots are generated to visualize the distance from home effect on both Young and experienced employees. As the mean distance from home increases, employees tend to leave the company. In later part of the project, the impact of distance from home will be checked while considering employee attrition rate.



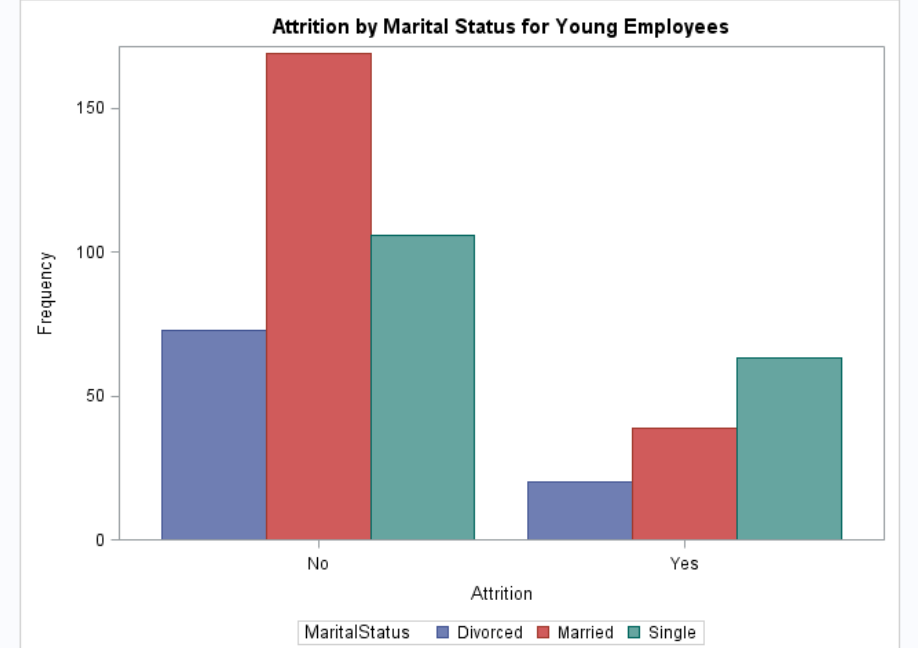


Variable: MaritalStatus Vs Attrition

In below plot, team tried to find the connection between attrition amounts with Marital Status. From below plot it is visible that employees having status single are leaving company.

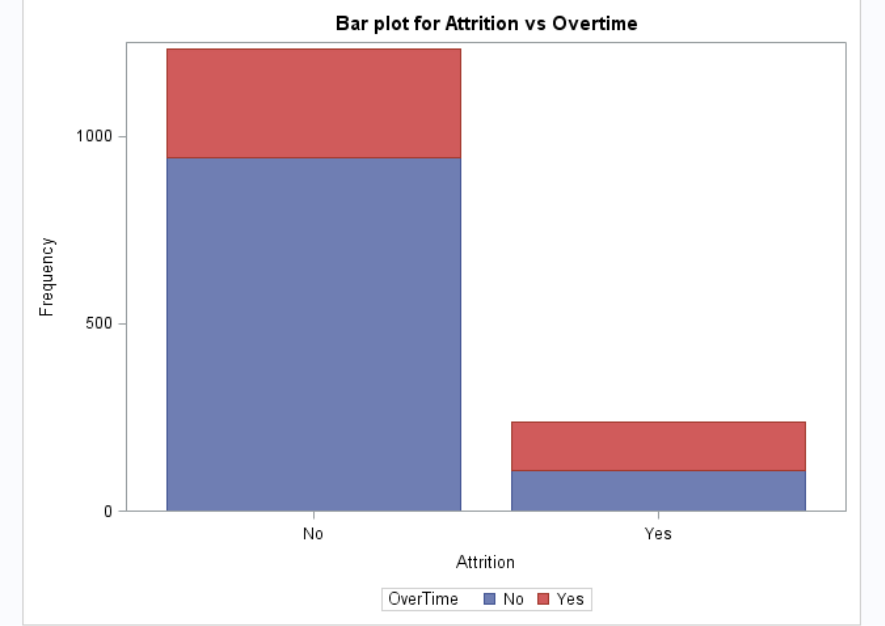


Further, by checking the impact of Marital Status in both Young and Experienced employees cluster. Found that in Young employee cluster, the impact of Marital Status is significant visually. Around half of the single young employees are leaving the company.

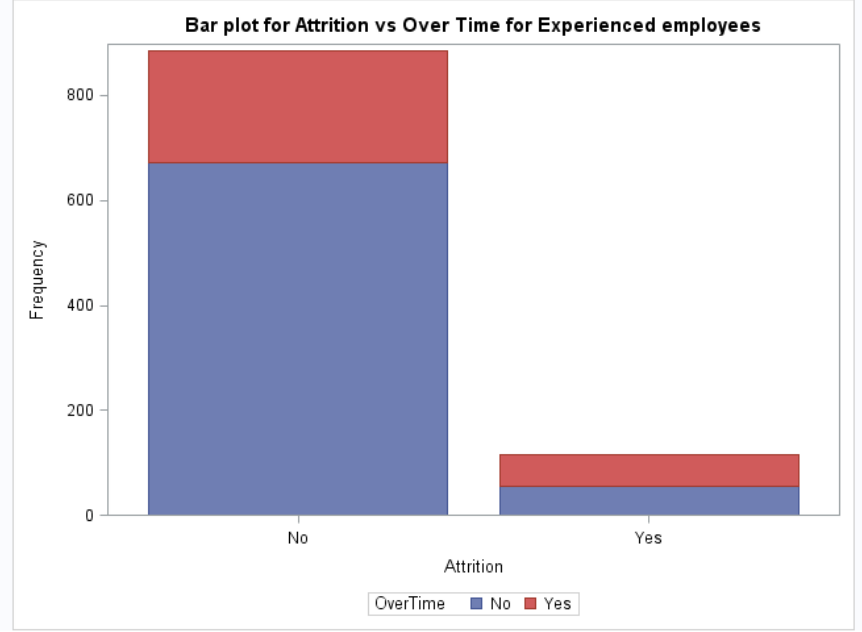


Variable: Attrition Vs OverTime

It is also valuable if some connection is found between Attrition rate and Over Time hours. Below plot shows that almost 50% of people who left the company does over time. Rate of over time employees ate less among the people having attrition 'NO'. It is assumed that Overtime is having a great impact on attrition. It can only be confirmed after building survival model.

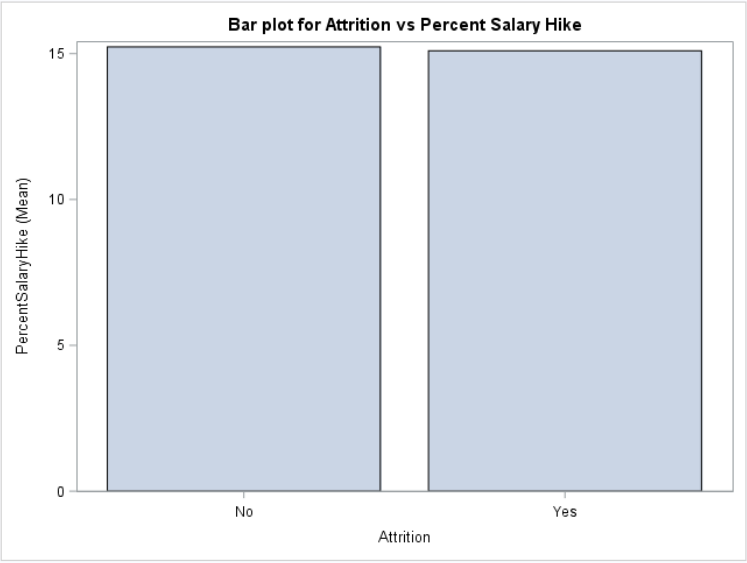


In Young employees, this ratio is more.



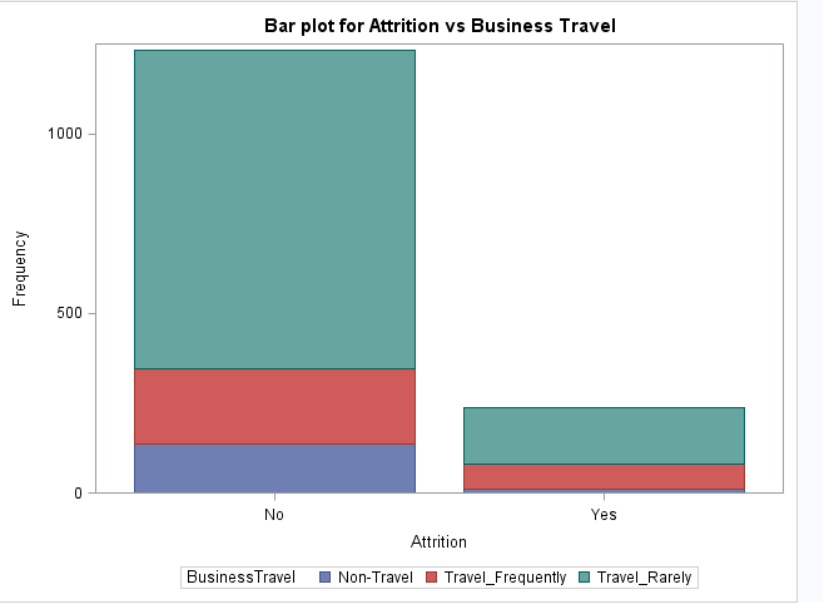
Variable: PercentSalaryHike Vs Attrition

When employees are leaving the company, salary hike might be an important factor. But as it is noticed in below bar plot, the employees which are with the company or leaving the company, both the set of people are getting average 15% of salary hike.

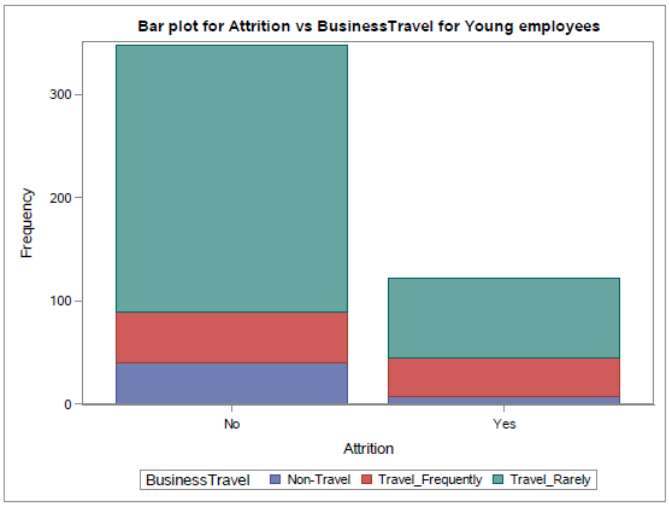


Variable: Attrition Vs Business travel

In the below plot it is shown that the employees who are leaving the company are getting frequent chances for business travel. This can be an important factor for employee attrition rate. In later part of the document, Survival model are build which will confirm this doubt.

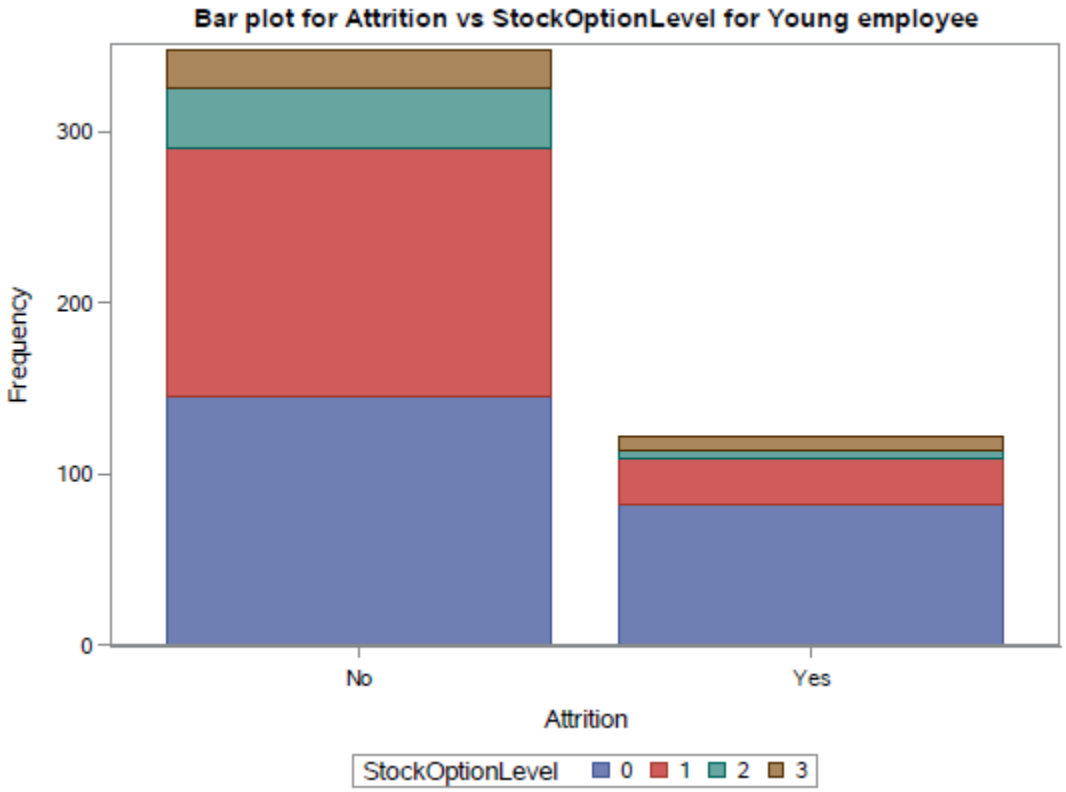


The experienced employee cluster follows the same pattern as the overall as shown above. In the young employee cluster, the employees who are travelling frequently are leaving more.



Variable: Attrition Vs StockOptionLevel

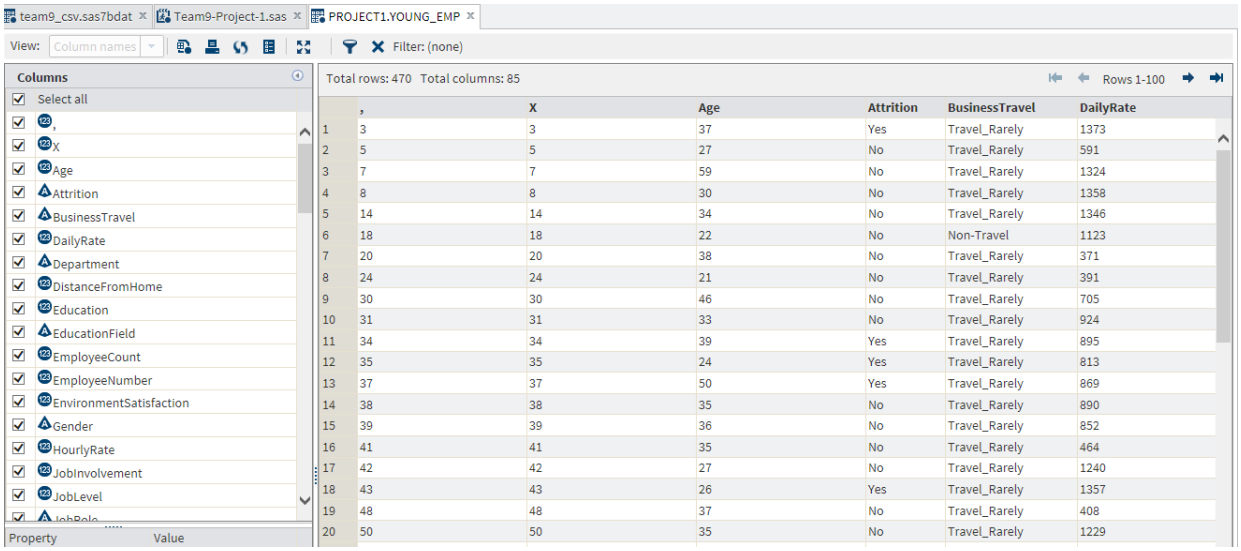
Young employees with stock option level 0 i.e. who are not having stocks offered from company are leaving more. In later part of the document, Survival model are build which will confirm this doubt.



10. Data Preparation (Modify Phase)

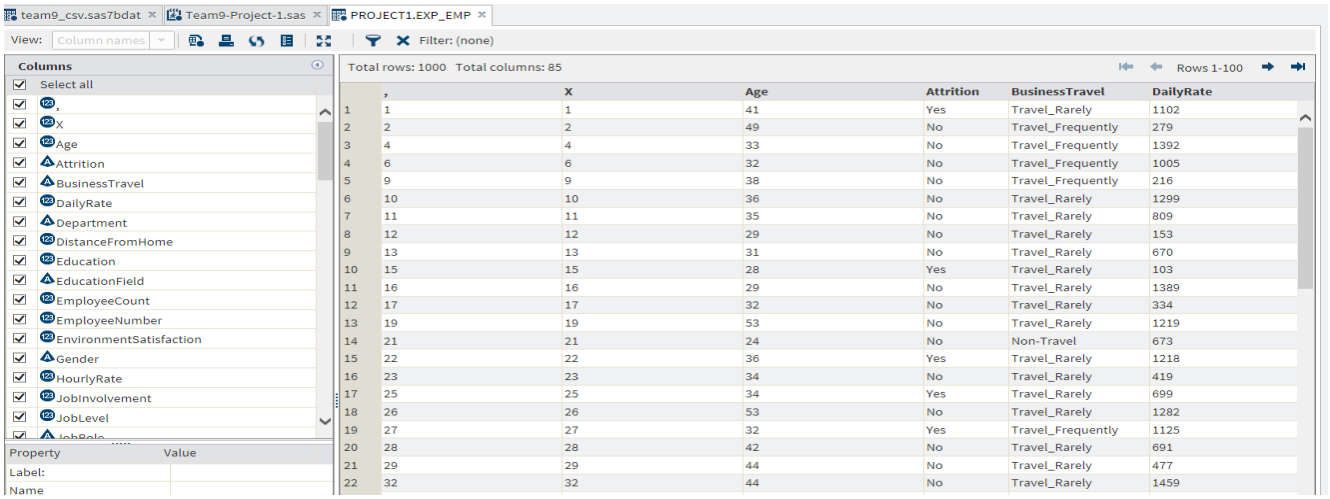
- ✓ So far few visualization is shown above in order to understand the dataset and the associated factors among employees.

- ✓ In problem statement of the project, the issue is mainly defined for 2 groups of employees. One is group of young employees and another is group of experience employees.
- ✓ **Young employees:** Company is investing a lot on the new employees and it takes three years to get returns from the employees. But, after 3 years employees become professionally competent enough to get the new career opportunities with higher salary. This is one of the biggest problem for the organization now.
- ✓ **Experienced employees**: These are the employees who are with the company for 5 years or more. These set of employees get executive training by professors and consultant from top universities and firms. They gain a lot and they become even more competent each year. These group of employees are another point of concern as they are being poached by rival companies.
- ✓ So to make project findings more insightful, full dataset is divided into 2 parts. One part is holding employee information having less than equal to 3 years of experience with the company. Another part is employees who are more than 3 years with the company. Here are the sample dataset.
- ✓ **Young employees dataset :** With company less than equal to 3 years:



	,	X	Age	Attrition	BusinessTravel	DailyRate
1	3	3	37	Yes	Travel_Rarely	1373
2	5	5	27	No	Travel_Rarely	591
3	7	7	59	No	Travel_Rarely	1324
4	8	8	30	No	Travel_Rarely	1358
5	14	14	34	No	Travel_Rarely	1346
6	18	18	22	No	Non-Travel	1123
7	20	20	38	No	Travel_Rarely	371
8	24	24	21	No	Travel_Rarely	391
9	30	30	46	No	Travel_Rarely	705
10	31	31	33	No	Travel_Rarely	924
11	34	34	39	Yes	Travel_Rarely	895
12	35	35	24	Yes	Travel_Rarely	813
13	37	37	50	Yes	Travel_Rarely	869
14	38	38	35	No	Travel_Rarely	890
15	39	39	36	No	Travel_Rarely	852
16	41	41	35	No	Travel_Rarely	464
17	42	42	27	No	Travel_Rarely	1240
18	43	43	26	Yes	Travel_Rarely	1357
19	48	48	37	No	Travel_Rarely	408
20	50	50	35	No	Travel_Rarely	1229

- ✓ **Experienced employees dataset:** With company more than 3 years:



	,	X	Age	Attrition	BusinessTravel	DailyRate
1	1	1	41	Yes	Travel_Rarely	1102
2	2	2	49	No	Travel_Frequently	279
3	4	4	33	No	Travel_Frequently	1392
4	6	6	32	No	Travel_Frequently	1005
5	9	9	38	No	Travel_Frequently	216
6	10	10	36	No	Travel_Rarely	1299
7	11	11	35	No	Travel_Rarely	809
8	12	12	29	No	Travel_Rarely	153
9	13	13	31	No	Travel_Rarely	670
10	15	15	28	Yes	Travel_Rarely	103
11	16	16	29	No	Travel_Rarely	1389
12	17	17	32	No	Travel_Rarely	334
13	19	19	53	No	Travel_Rarely	1219
14	21	21	24	No	Non-Travel	673
15	22	22	36	Yes	Travel_Rarely	1218
16	23	23	34	No	Travel_Rarely	419
17	25	25	34	Yes	Travel_Rarely	699
18	26	26	53	No	Travel_Rarely	1282
19	27	27	32	Yes	Travel_Frequently	1125
20	28	28	42	No	Travel_Rarely	691
21	29	29	44	No	Travel_Rarely	477
22	32	32	44	No	Travel_Rarely	1459

11. Exploration phase: Survival Methods

To explain when can be the biggest danger for the employees to leave the company, the hazard plot and the function estimates have been chosen to better explain the probability of losing the employees in the coming period. These estimates are calculated separately for the two intervals, Young (tenure less than 3 years) and experienced employees (tenure greater than 3years) with 37 years being the longest number of years survived.

APPROACH:

For young employees, ran the Procedure Life test – Life table model with the interval gap of 1. To analyse the survival time of censored employees with 4 years at company and added the interval 4 where attrition is 0.

For Experienced employees ran the Procedure Life test – Life table model with the interval with interval 4 10 20 30 and 40. At 40 years at company it hasonly one employee whose attrition is yes that is he left the company, showing that there is no censored data at point 40 years at the company

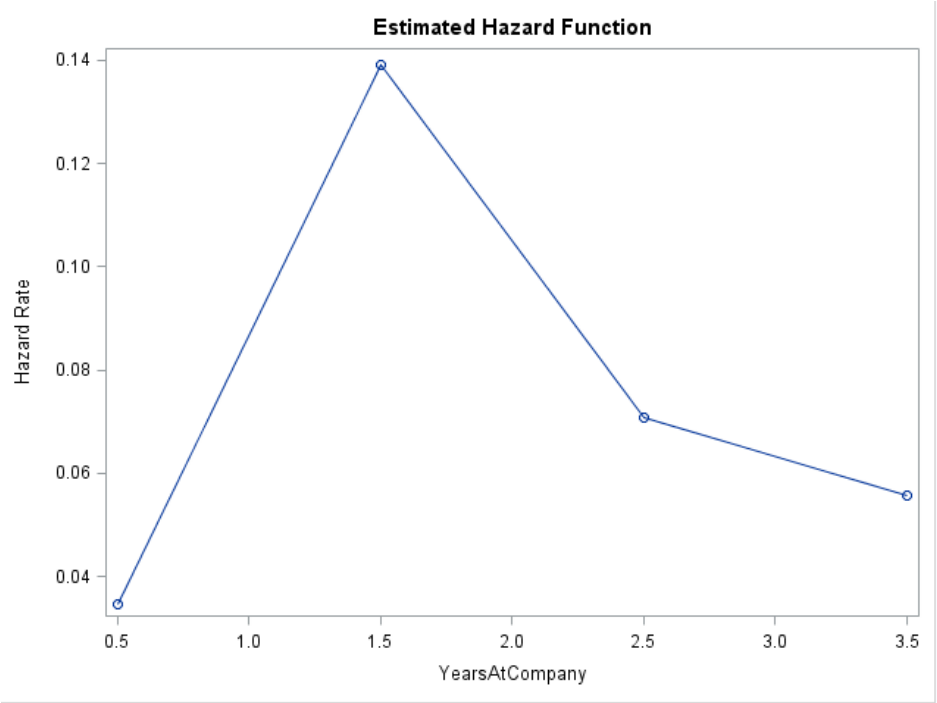
Young Employees:

The hazard rate or probability of failure, shows an increase during the first interval of **0-1(0.034)** and **1-2(0.1300)**. The first drop in the hazard rate occurred between **2-3(0.0684)** and it continued decrease overall till the last interval.

The LIFETEST Procedure

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval: PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	16	0	470.0	0.0340	0.00836	1.0000	0	0	.	.	0.0340	0.00836	0.034632	0.008657
1	2	59	0	454.0	0.1300	0.0158	0.9660	0.0340	0.00836	.	.	0.1255	0.0153	0.138987	0.018051
2	3	27	0	395.0	0.0684	0.0127	0.8404	0.1596	0.0169	.	.	0.0574	0.0107	0.070773	0.013612
3	4	20	0	368.0	0.0543	0.0118	0.7830	0.2170	0.0190	.	.	0.0426	0.00931	0.055866	0.012487
4	.	0	348	174.0	0	0	0.7404	0.2596	0.0202	.	.	.	.	.	.



Through the examination of the above estimates and curves, it appears that more number of employees are leaving during an interval of 1.5-2 years after they joined the company. Calculating the mean of the employees who are still in the survival period (0-3yrs) and it came out to be **1.82** (348 observations).

So, the existing employees in this period have the highest probability of leaving the company in the coming months and for the new employees who would be joining can leave the company in the period of 2years.

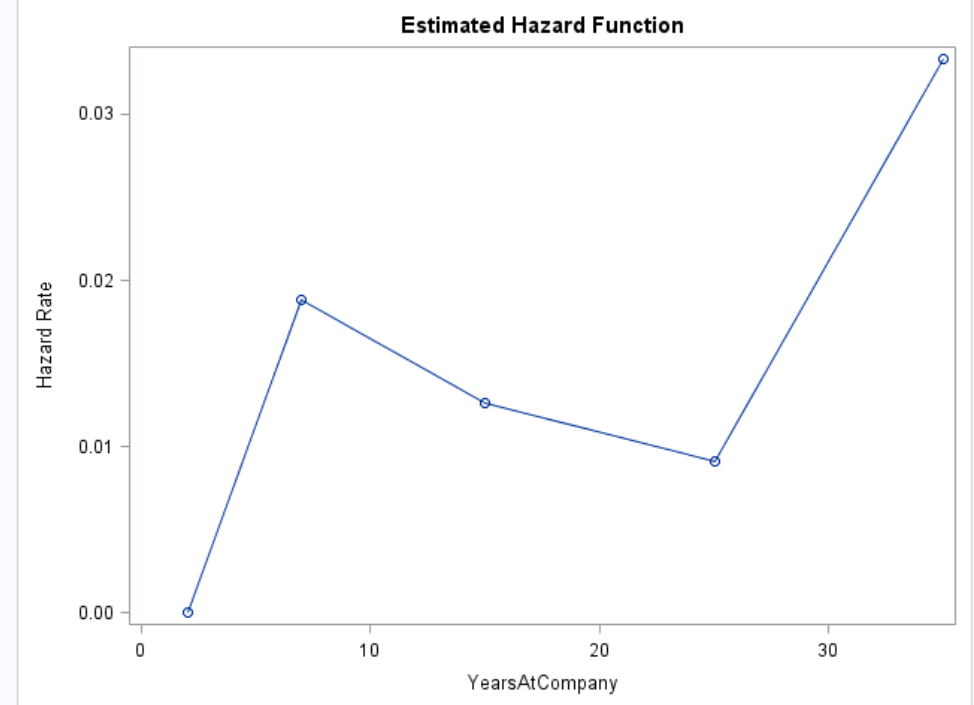
Experienced Employees:

The hazard rate or probability of failure, shows an increase during the first interval of **4-10(0.1067)** and **10-20(0.1189)**. The first drop in the hazard rate occurred between **20-30(0.087)**, where it stayed until it rose to **0.2857**.

The LIFETEST Procedure

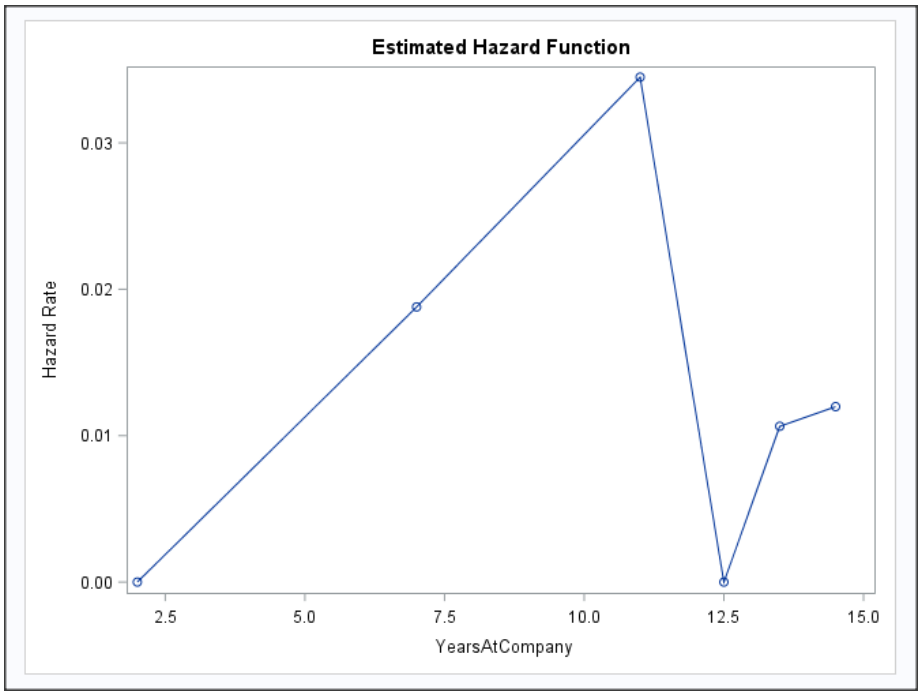
Life Table Survival Estimates

Interval												Evaluated at the Midpoint of the Interval			
[Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	1000.0	0	0	1.0000	0	0	.	.	0	.	0	.
4	10	77	557	721.5	0.1067	0.0115	1.0000	0	0	.	.	0.0178	0.00192	0.01879	0.002138
10	20	29	244	244.0	0.1189	0.0207	0.8933	0.1067	0.0115	.	.	0.0106	0.00186	0.012636	0.002342
20	30	5	71	57.5	0.0870	0.0372	0.7871	0.2129	0.0211	.	.	0.00684	0.00293	0.009091	0.004061
30	40	3	13	10.5	0.2857	0.1394	0.7187	0.2813	0.0350	.	.	0.0205	0.0101	0.033333	0.018976
40	.	1	0	1.0	1.0000	0	0.5133	0.4867	0.1033	.	.	.	.	.	.



Through the examination of the above estimates and curves, it appears that more number of employees are leaving during an interval of 7-15 years and if sustained in this period, they would probably stay in the company for long. Calculating the mean of the employees who are still in the survival period (0-3yrs) and it came out to be **9.32** (885 observations).

So, to be more precise on the point where employees are leaving the company, should consider the duration between 4 and 15 with a 2 interval difference. Below is the plot depicting the same



It can be observed that the highest probability of failure is showing at around 11.5years of tenure. So, conclude saying that the employees who have survived the interval of 0-4, have the highest probability of leaving at 11.5 years of experience.

Analyzing Contribution of different employee groups or categories to Attrition:

YOUNG EMPLOYEES:

Category: GENDER

The LIFETEST Procedure
Stratum 1: Gender = Female

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	4	0	179.0	0.0223	0.0110	1.0000	0	0	.	.	0.0223	0.0110	0.022599	0.011299
1	2	21	0	175.0	0.1200	0.0246	0.9777	0.0223	0.0110	.	.	0.1173	0.0241	0.12766	0.027801
2	3	10	0	154.0	0.0649	0.0199	0.8603	0.1397	0.0259	.	.	0.0559	0.0172	0.067114	0.021211
3	4	7	0	144.0	0.0486	0.0179	0.8045	0.1955	0.0296	.	.	0.0391	0.0145	0.049822	0.018825
4	.	0	137	68.5	0	0	0.7654	0.2346	0.0317	.	.	.	.	.	.

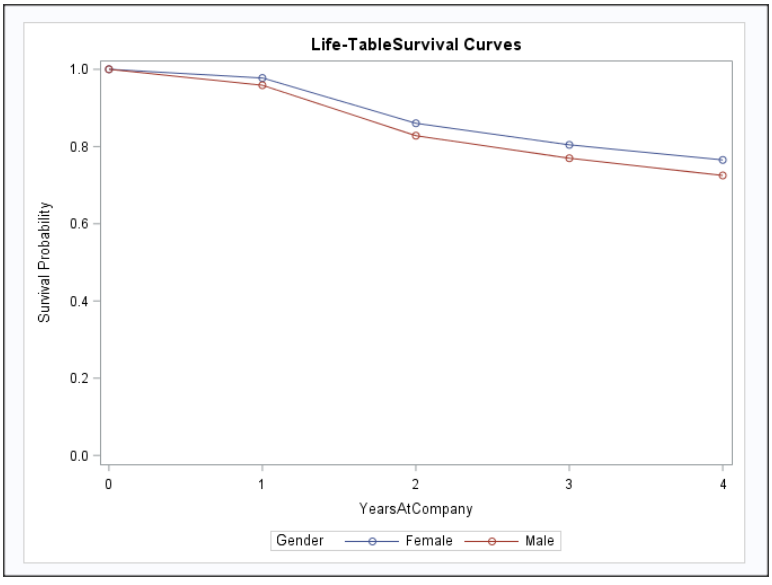
The LIFETEST Procedure
Stratum 2: Gender = Male

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	12	0	291.0	0.0412	0.0117	1.0000	0	0	.	.	0.0412	0.0117	0.042105	0.012152
1	2	38	0	279.0	0.1362	0.0205	0.9588	0.0412	0.0117	.	.	0.1306	0.0198	0.146154	0.023646
2	3	17	0	241.0	0.0705	0.0165	0.8282	0.1718	0.0221	.	.	0.0584	0.0137	0.073118	0.017722
3	4	13	0	224.0	0.0580	0.0156	0.7698	0.2302	0.0247	.	.	0.0447	0.0121	0.05977	0.01657
4	.	0	211	105.5	0	0	0.7251	0.2749	0.0262	.	.	.	.	.	.

Interpretation:

Among young **FEMALE** employees who are in the interval of 1-2 years at the company, the probability of leaving the company is 12%. Among young **MALE** employees, the probability of leaving in that interval is 13.6%.



Interpretation:

The survival probability of male employees within 3 years of working at the company falls from 1 to less than 0.8 and the survival probability of female employees within 3 years of working at the company falls to 0.8

Conclusion:

From the above interpretations, it can be concluded that young male employees contribute more to attrition than young female employees.

Category : OVERTIME:

The LIFETEST Procedure

Stratum 1: OverTime = No

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	6	0	326.0	0.0184	0.00744	1.0000	0	0	.	.	0.0184	0.00744	0.018576	0.007583
1	2	31	0	320.0	0.0969	0.0165	0.9816	0.0184	0.00744	.	.	0.0951	0.0162	0.101806	0.018261
2	3	12	0	289.0	0.0415	0.0117	0.8865	0.1135	0.0176	.	.	0.0368	0.0104	0.042403	0.012238
3	4	6	0	277.0	0.0217	0.00875	0.8497	0.1503	0.0198	.	.	0.0184	0.00744	0.021898	0.008939
4	.	0	271	135.5	0	0	0.8313	0.1687	0.0207	.	.	.	.	.	.

The LIFETEST Procedure

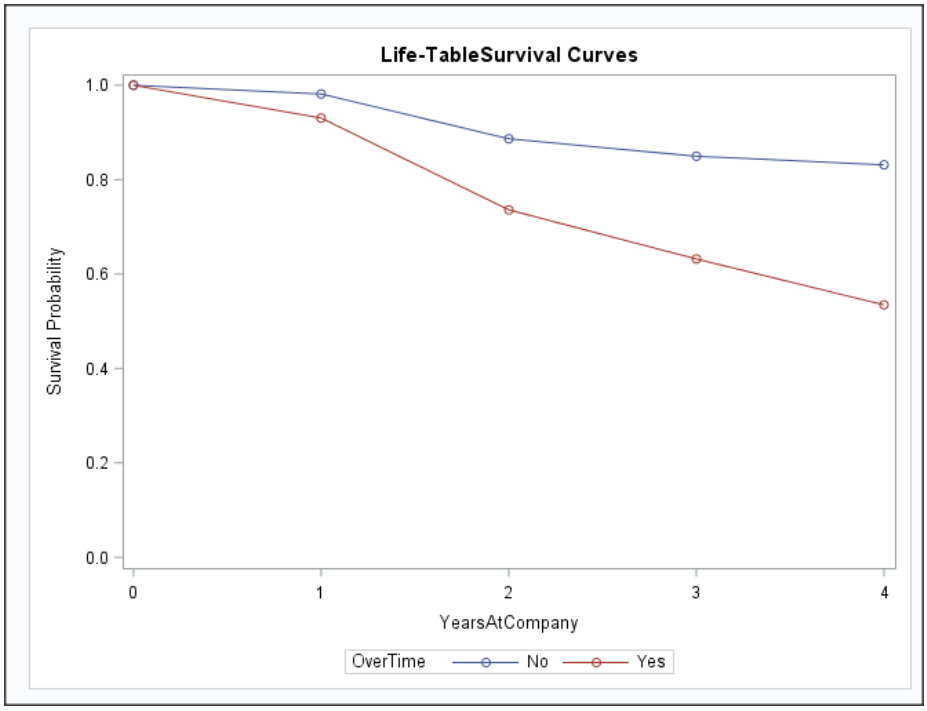
Stratum 2: OverTime = Yes

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	10	0	144.0	0.0694	0.0212	1.0000	0	0	.	.	0.0694	0.0212	0.071942	0.022735
1	2	28	0	134.0	0.2090	0.0351	0.9306	0.0694	0.0212	.	.	0.1944	0.0330	0.233333	0.043795
2	3	15	0	106.0	0.1415	0.0339	0.7361	0.2639	0.0367	.	.	0.1042	0.0255	0.152284	0.039205
3	4	14	0	91.0	0.1538	0.0378	0.6319	0.3681	0.0402	.	.	0.0972	0.0247	0.166667	0.044389
4	.	0	77	38.5	0	0	0.5347	0.4653	0.0416	.	.	.	.	.	.

Interpretation:

Among young employees who do overtime and who don't, the conditional probablity of an employee doing overtime to leave the company in the interval of 1-2 years at the company is 20.9% whereas for the young employees who don't perform overtime in that interval is only 9% . This shows that young employees who do overtime is having attrition probablity more than the young employees who don't do overtime.



Interpretation:

Young employees doing overtime have survival probability less than 60% whereas the employees not performing overtime tend to survive with more than 80% probability.

Conclusion:

From the above interpretations, it can be concluded that employees doing overtime have less survival probability tending more to leave the company within four years and employees not performing overtime have very less attrition rate tending to stay back in company.

CATEGORY : MARITAL STATUS

The LIFETEST Procedure

Stratum 1: MaritalStatus = Divorced

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	0	0	93.0	0	0	1.0000	0	0	.	.	0	.	0	.
1	2	8	0	93.0	0.0860	0.0291	1.0000	0	0	.	.	0.0860	0.0291	0.089888	0.031748
2	3	5	0	85.0	0.0588	0.0255	0.9140	0.0860	0.0291	.	.	0.0538	0.0234	0.060606	0.027091
3	4	7	0	80.0	0.0875	0.0316	0.8602	0.1398	0.0360	.	.	0.0753	0.0274	0.091503	0.034549
4	.	0	73	36.5	0	0	0.7849	0.2151	0.0426	.	.	.	.	.	.

The LIFETEST Procedure

Stratum 2: MaritalStatus = Married

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	2	0	208.0	0.00962	0.00677	1.0000	0	0	.	.	0.00962	0.00677	0.009662	0.006832
1	2	21	0	206.0	0.1019	0.0211	0.9904	0.00962	0.00677	.	.	0.1010	0.0209	0.107417	0.023406
2	3	9	0	185.0	0.0486	0.0158	0.8894	0.1106	0.0217	.	.	0.0433	0.0141	0.049861	0.016615
3	4	7	0	176.0	0.0398	0.0147	0.8462	0.1538	0.0250	.	.	0.0337	0.0125	0.04058	0.015335
4	.	0	169	84.5	0	0	0.8125	0.1875	0.0271	.	.	.	.	.	.

The LIFETEST Procedure

Stratum 3: MaritalStatus = Single

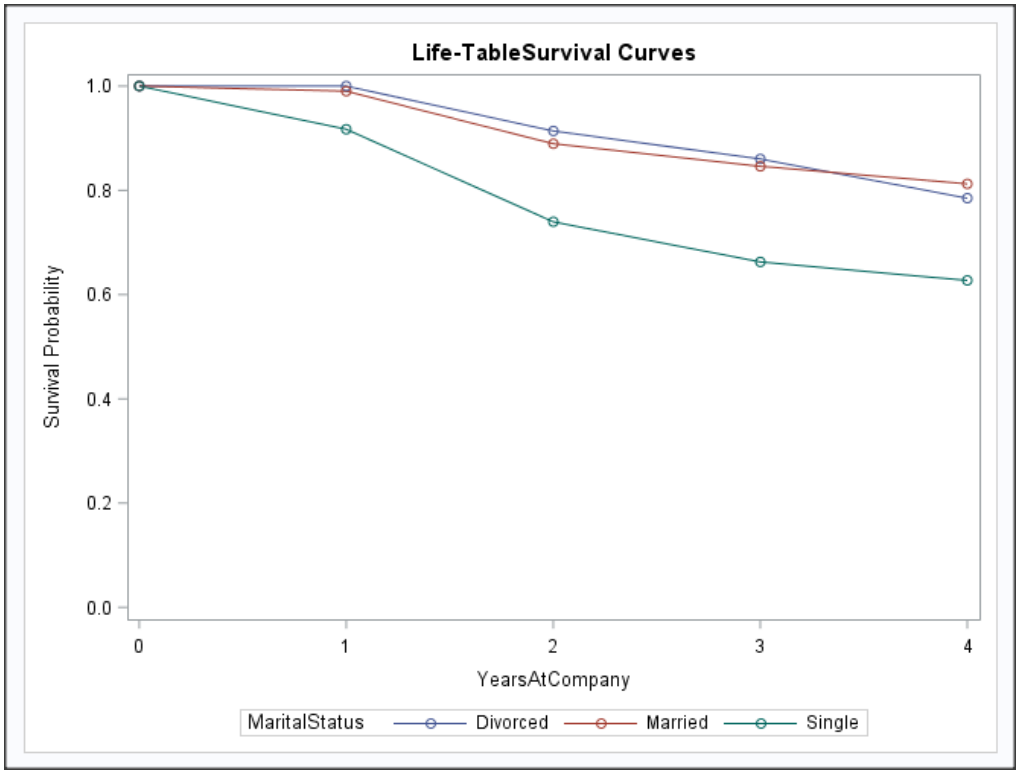
Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	14	0	169.0	0.0828	0.0212	1.0000	0	0	.	.	0.0828	0.0212	0.08642	0.023075
1	2	30	0	155.0	0.1935	0.0317	0.9172	0.0828	0.0212	.	.	0.1775	0.0294	0.214286	0.038898
2	3	13	0	125.0	0.1040	0.0273	0.7396	0.2604	0.0338	.	.	0.0769	0.0205	0.109705	0.030381
3	4	6	0	112.0	0.0536	0.0213	0.6627	0.3373	0.0364	.	.	0.0355	0.0142	0.055046	0.022464
4	.	0	106	53.0	0	0	0.6272	0.3728	0.0372	.	.	.	.	.	.

Interpretation:

The different marital status of the employees contributes differently to attrition rate.

For example, **divorced employees** with one-year of experience have the probability of **8.6%** to leave the company before entering in their second year at the company, whereas for **married employees** the probability to leave at that year is **10.1%** and probability of **single employees** to leave the company in that interval is **19.35%**



Interpretation:

The probability of 'single' employees to survive after 3 years falls to 70% whereas 'married' and 'divorced' employee tend to survive with more than 80% of probability.

Conclusion:

Young employees who are 'single' tend to leave the company more than the young 'married' and 'divorced employees

Experienced Employees:

CATEGORY: GENDER

The LIFETEST Procedure

Stratum 1: Gender = Female

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	409.0	0	0	1.0000	0	0	34.7505	0.5068	0	.	0	.
4	10	30	230	294.0	0.1020	0.0177	1.0000	0	0	30.7505	0.5978	0.0170	0.00294	0.017921	0.003267
10	20	10	101	98.5	0.1015	0.0304	0.8980	0.1020	0.0177	25.7964	0.9273	0.00912	0.00274	0.010695	0.003377
20	30	2	33	21.5	0.0930	0.0626	0.8068	0.1932	0.0316	16.7308	1.7834	0.00751	0.00506	0.009756	0.00689
30	40	2	0	3.0	0.6667	0.2722	0.7317	0.2683	0.0581	7.5000	4.3301	0.0488	0.0203	0.1	0.061237
40	.	1	0	1.0	1.0000	0	0.2439	0.7561	0.2001	.	.	.	.	.	.

The LIFETEST Procedure

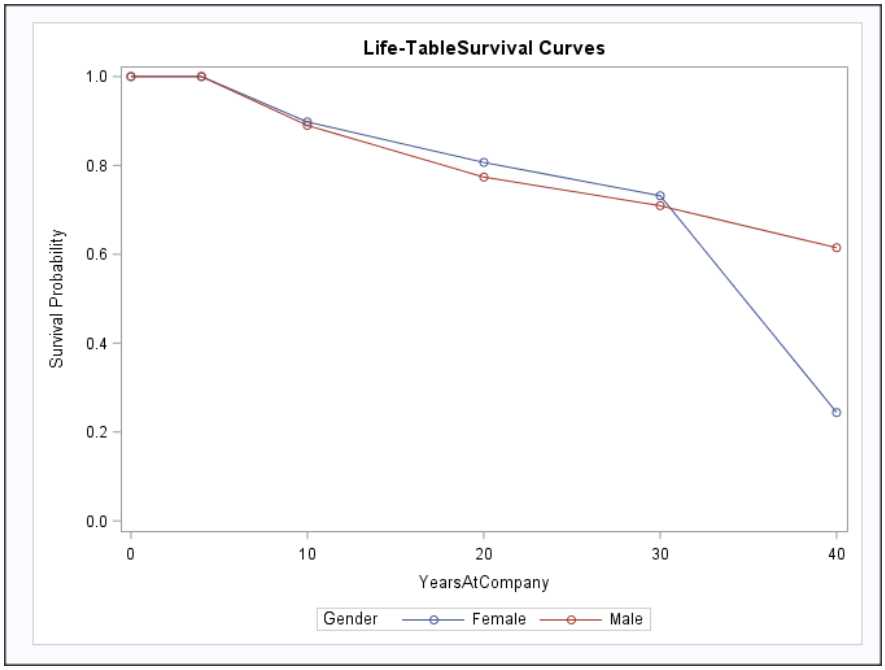
Stratum 2: Gender = Male

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	591.0	0	0	1.0000	0	0	.	.	0	.	0	.
4	10	47	327	427.5	0.1099	0.0151	1.0000	0	0	.	.	0.0183	0.00252	0.019389	0.002823
10	20	19	143	145.5	0.1306	0.0279	0.8901	0.1099	0.0151	.	.	0.0116	0.00249	0.013971	0.003197
20	30	3	38	36.0	0.0833	0.0461	0.7738	0.2262	0.0281	.	.	0.00645	0.00357	0.008696	0.005016
30	40	1	13	7.5	0.1333	0.1241	0.7093	0.2907	0.0440	.	.	0.00946	0.00882	0.014286	0.014249
40	.	0	0	0.0	0	0	0.6148	0.3852	0.0959	.	.	.	.	.	.

Interpretation:

Among experienced employees , the probablity of female employees to leave the company between the interval of 30-40 years at company is 66.6% whereas the the probablity of male employees to leave in that interval is only 13.3%.



Interpretation:

Between 10 – 30 years at the company, experienced male employees tend to survive marginally less than female employees but after 30 years the probability of survival of the female employees falls very low to less than 20%

Conclusion:

Female employees after working 30 years at the company have more probability of leaving the company compared to male employees.

CATEGORY : OVERTIME

Stratum 1: OverTime = No

Life Table Survival Estimates

Interval [Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	728.0	0	0	1.0000	0	0	.	.	0	.	0	.
4	10	34	425	515.5	0.0660	0.0109	1.0000	0	0	.	.	0.0110	0.00182	0.011367	0.001948
10	20	15	189	174.5	0.0860	0.0212	0.9340	0.0660	0.0109	.	.	0.00803	0.00198	0.008982	0.002317
20	30	3	53	38.5	0.0779	0.0432	0.8538	0.1462	0.0222	.	.	0.00665	0.00369	0.008108	0.004677
30	40	2	6	6.0	0.3333	0.1925	0.7872	0.2128	0.0422	.	.	0.0262	0.0152	0.04	0.027713
40	.	1	0	1.0	1.0000	0	0.5248	0.4752	0.1541	.	.	.	.	.	.

The LIFETEST Procedure

Stratum 2: OverTime = Yes

Life Table Survival Estimates

Interval [Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	272.0	0	0	1.0000	0	0	35.2022	2.4131	0	.	0	.
4	10	43	132	206.0	0.2087	0.0283	1.0000	0	0	31.2022	2.7728	0.0348	0.00472	0.038844	0.005883
10	20	14	55	69.5	0.2014	0.0481	0.7913	0.2087	0.0283	.	.	0.0159	0.00385	0.0224	0.005949
20	30	2	18	19.0	0.1053	0.0704	0.6319	0.3681	0.0443	.	.	0.00665	0.00447	0.011111	0.007845
30	40	1	7	4.5	0.2222	0.1960	0.5654	0.4346	0.0596	.	.	0.0126	0.0112	0.025	0.024804
40	.	0	0	0.0	0	0	0.4397	0.5603	0.1201	.	.	.	.	.	.

Interpretation:

Between the employee's years at company 4-10, 10-20 and 20-30 , with performing overtime the probability of employees leaving the company is 20.8%, 20.14% and 10.53% respectively , whereas the probability in those intervals for the employees not performing overtime is low , tending more to stay back at the company. But after working for 30 years at company in the time with ten years the tendency of the employees without overtime to leave the company is 33.3%

Conclusion:

Experienced employees with experience of 10 – 30 years at the company tend to leave the company because of overtime, but with 30 + years of experience at the company employees tend to look for overtime and behaving to leave the company without overtime.

CATEGORY : MARITAL STATUS

The LIFETEST Procedure

Stratum 1: MaritalStatus = Divorced

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	234.0	0	0	1.0000	0	0	.	.	0	.	0	.
4	10	11	145	161.5	0.0681	0.0198	1.0000	0	0	.	.	0.0114	0.00330	0.011752	0.003541
10	20	2	54	51.0	0.0392	0.0272	0.9319	0.0681	0.0198	.	.	0.00365	0.00253	0.004	0.002828
20	30	0	18	13.0	0	0	0.8953	0.1047	0.0317	.	.	0	.	0	.
30	40	0	4	2.0	0	0	0.8953	0.1047	0.0317	.	.	0	.	0	.
40	.	0	0	0.0	0	0	0.8953	0.1047	0.0317	.	.	.	.	.	.

The LIFETEST Procedure

Stratum 2: MaritalStatus = Married

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	465.0	0	0	1.0000	0	0	38.6938	0.8721	0	.	0	.
4	10	24	253	338.5	0.0709	0.0140	1.0000	0	0	34.6938	1.0222	0.0118	0.00233	0.012251	0.002499
10	20	15	127	124.5	0.1205	0.0292	0.9291	0.0709	0.0140	.	.	0.0112	0.00272	0.012821	0.003303
20	30	3	35	28.5	0.1053	0.0575	0.8172	0.1828	0.0298	.	.	0.00860	0.00471	0.011111	0.006405
30	40	2	5	5.5	0.3636	0.2051	0.7311	0.2689	0.0540	.	.	0.0266	0.0151	0.044444	0.030641
40	.	1	0	1.0	1.0000	0	0.4653	0.5347	0.1539	.	.	.	.	.	.

The LIFETEST Procedure

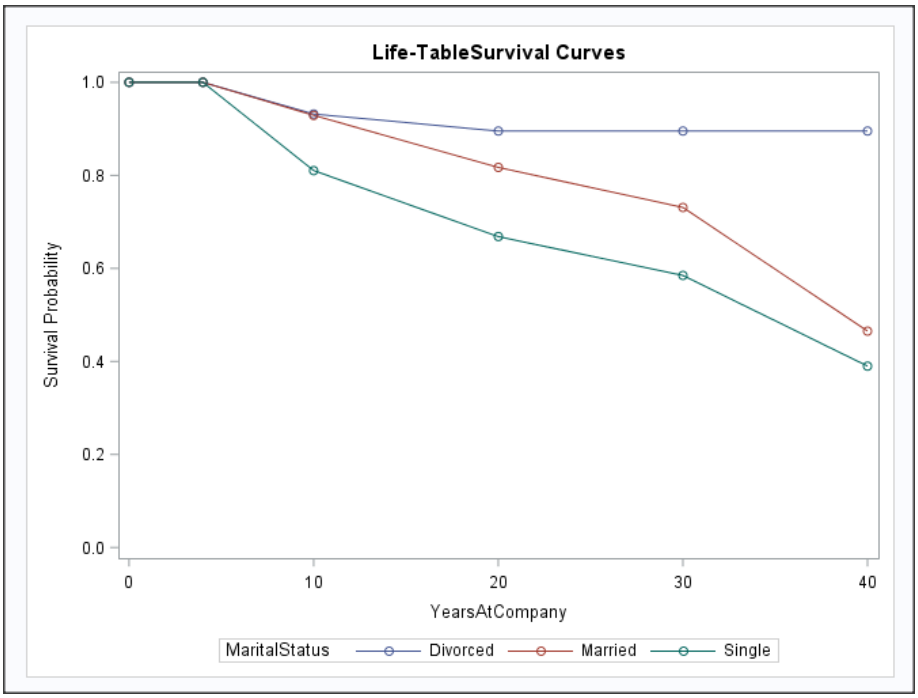
Stratum 3: MaritalStatus = Single

Life Table Survival Estimates

Interval [Lower,	Interval Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	4	0	0	301.0	0	0	1.0000	0	0	34.3531	1.4783	0	.	0	.
4	10	42	159	221.5	0.1896	0.0263	1.0000	0	0	30.3531	1.7232	0.0316	0.00439	0.034913	0.005358
10	20	12	63	68.5	0.1752	0.0459	0.8104	0.1896	0.0263	29.2162	2.5112	0.0142	0.00375	0.0192	0.005517
20	30	2	18	16.0	0.1250	0.0827	0.6684	0.3316	0.0431	.	.	0.00836	0.00555	0.013333	0.009407
30	40	1	4	3.0	0.3333	0.2722	0.5849	0.4151	0.0669	.	.	0.0195	0.0161	0.04	0.039192
40	.	0	0	0.0	0	0	0.3899	0.6101	0.1653	.	.	.	.	.	.

Interpretation:

Both ‘married’ and ‘single’ employees in the experience interval of 30-40 years at the company leave the company at the probability more than 30%



Interpretation:

Divorced employees tend to survive more, whereas single and divorced employee's survival probability fall to 40%.

Conclusion:

Single and married experienced employees tend to leave the company more after working for 30 – 40 years in the company.

SUMMARY OF ATTRITION BEHAVIOUR – YOUNG AND EXPERIENCED EMPLOYEES

- Young experienced **male** employees have more tendency to leave and experienced **female** employees (30- 40 years at the company) have more tendency to leave the company.
- Young employees who does overtime leave the company more whereas with experience of 30 years at the company employees who does not do overtime tend to leave the company as well.
- Single employees in both young and experienced groups tend to leave the company very quickly.

12. Modeling phase: Exponential Model

Young Employees:

In the process of building the model, important parameters have been removed by setting up the cut-off value for the probability of the chi2 value as 0.05 in several steps.

The exponential model with stepwise regression shows that the following variables are statistically significant. So, the below covariates are the ones which should retain in the model.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Age	1	10.8510	0.0010
DistanceFromHome	1	6.2748	0.0122
OverTime	1	29.7478	<.0001
StockOptionLevel	3	11.2130	0.0106
BusinessTravel	2	10.1048	0.0064
EnvironmentSatisfact	3	16.6409	0.0008
Department	2	6.9355	0.0312
JobSatisfaction	3	14.0907	0.0028
MonthlyIncome	1	10.7378	0.0010
NumCompaniesWorked	1	3.8578	0.0495
WorkLifeBalance	3	10.6031	0.0141
YearsInCurrentRole	1	5.4486	0.0196
YearsWithCurrManager	1	6.0200	0.0141

Also, the below variables found to be not statistically significant and hence decided to remove it in the following order through step wise regression.

1. JobRole
2. Education
3. HourlyRate
4. PerformanceRating
5. PercentSalaryHike
6. YearsSinceLastPromot
7. EducationField
8. JobLevel
9. RelationshipSatisfaction
10. Marital Status
11. Daily Rate
12. Job Involvement
13. TrainingTimesLastYea
14. TotalWorkingYears
15. Monthly Rate
16. BonusReceivedRatio

For the exponential distribution, found the log likelihood for the null model as below.

The LIFEREG Procedure

Model Information	
Data Set	PROJECT1.YOUNG_EMP
Dependent Variable	Log(YearsAtCompany)
Censoring Variable	lAttrition
Censoring Value(s)	0
Number of Observations	426
Noncensored Values	106
Right Censored Values	320
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	1
Zero or Negative Response	44
Name of Distribution	Exponential
Log Likelihood	-280.7429216

Also, the log ratio for the exponential is showed as follows

Obs	L_null	L_full	L	p_value
1	-280.742	-188.267	184.95	0

Interpretation of log ratio for Exponential:

The p-value zero indicates that at least one of the coefficient of covariate is not zero. This specifies that the model can be explained by at least one of the covariate. This also suggests that the full model is significantly different from the null model.

Experienced Employees:

In the process of building the model, the important parameters have been removed by setting up the cut-off value for the probability of the chi -square as 0.05 in several steps.

The exponential model with stepwise regression shows that the following variables are statistically significant. So, the below covariates are the ones which have been retained in the model.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DistanceFromHome	1	11.7094	0.0006
OverTime	1	34.8028	<.0001
StockOptionLevel	3	17.4527	0.0006
BusinessTravel	2	11.2157	0.0037
EnvironmentSatisfact	3	5.5540	0.1354
Department	2	9.2206	0.0099
EducationField	5	9.8404	0.0799
JobInvolvement	3	9.8239	0.0201
JobLevel	4	34.4641	<.0001
JobSatisfaction	3	4.5505	0.2078
NumCompaniesWorked	1	14.1533	0.0002
RelationshipSatisfac	1	5.1458	0.0233
TrainingTimesLastYea	1	6.0049	0.0143
WorkLifeBalance	3	5.0922	0.1652
YearsInCurrentRole	1	8.1310	0.0044
YearsSinceLastPromot	1	7.0330	0.0080
YearsWithCurrManager	1	6.3302	0.0119

Also, the below variables found to be not statistically significant and hence decided to remove it in the following order through step wise regression

1. JobRole
2. Education
3. BonusReceivedRatio
4. PerformanceRating
5. PercentSalaryHike
6. MaritalStatus
7. hourlyrate
8. monthlyrate
9. dailyrate
10. TotalWorkingYears
11. MonthlyIncome

For the exponential distribution, found the log likelihood for the null model as below.

The LIFEREG Procedure

Model Information	
Data Set	PROJECT1.EXP_EMP
Dependent Variable	Log(YearsAtCompany)
Censoring Variable	lAttrition
Censoring Value(s)	0
Number of Observations	1000
Noncensored Values	115
Right Censored Values	885
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	1
Name of Distribution	Exponential
Log Likelihood	-388.2480908

Also, found the log ratio for the exponential as follows

Obs	L_null	L_full	L	p_value
1	-388.25	-280.83	214.84	0

Interpretation of log ratio for Exponential:

The p-value zero indicates that at least one of the coefficient of covariate is not zero. This specifies that the model can be explained by at least one of the covariate. This also suggests that the full model is significantly different from the null model.

13. Modeling phase: Weibull Model

Young Employees:

The Weibull model with stepwise regression shows that the following variables are statistically significant. So, the below covariates are the ones which have been retained in the model.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DistanceFromHome	1	27.9863	<.0001
MonthlyRate	1	13.9903	0.0002
OverTime	1	37.6010	<.0001
StockOptionLevel	3	30.3902	<.0001
BusinessTravel	2	14.4385	0.0007
EnvironmentSatisfact	3	34.7065	<.0001
Department	2	9.6124	0.0082
JobSatisfaction	3	18.7509	0.0003
MonthlyIncome	1	6.8493	0.0089
NumCompaniesWorked	1	21.8276	<.0001
TotalWorkingYears	1	16.2708	<.0001
WorkLifeBalance	3	10.8378	0.0126
YearsInCurrentRole	1	118.6539	<.0001
YearsWithCurrManager	1	97.1968	<.0001
BonusReceivedRatio	1	14.5118	0.0001

Also, the below variables found to be not statistically significant and hence decided to remove it in the following order through step wise regression

1. Job Involvement
2. Job Role
3. Job Level
4. Marital Status
5. Daily rate
6. Relationship Statistic
7. Education Filed
8. Education
9. YearsSinceLastPromot
10. Hourly rate
11. TrainingTimesLastYear
12. PercentSalaryHike
13. Performance Rating
14. Age

For the weibull distribution found the log likelihood for the null model as below.

The LIFEREG Procedure

Model Information	
Data Set	PROJECT1.YOUNG_EMP
Dependent Variable	Log(YearsAtCompany)
Censoring Variable	lAttrition
Censoring Value(s)	0
Number of Observations	426
Noncensored Values	106
Right Censored Values	320
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	2
Zero or Negative Response	44
Name of Distribution	Weibull
Log Likelihood	-253.1648599

Also, found the log ratio for the weibull as follows

Obs	L_null	L_full	L	p_value
1	-253.164	-44.656	417.016	0

Interpretation of log ratio for Weibull:

The p-value zero indicates that at least one of the coefficient of covariate is not zero. This specifies that the model can be explained by at least one of the covariate. This also suggests that the full model is significantly different from the null model.

Experienced Employees:

The Weibull model with stepwise regression shows that the following variables are statistically significant. So, the below covariates are the ones which have been retained in the model.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DistanceFromHome	1	10.5463	0.0012
OverTime	1	26.2367	<.0001
BusinessTravel	2	12.9256	0.0016
Department	2	9.5623	0.0084
EducationField	5	14.3029	0.0138
JobInvolvement	3	12.1852	0.0068
JobLevel	4	37.2569	<.0001
MaritalStatus	2	18.8270	<.0001
NumCompaniesWorked	1	79.5345	<.0001
TotalWorkingYears	1	66.7370	<.0001
TrainingTimesLastYea	1	6.5843	0.0103
YearsInCurrentRole	1	31.3794	<.0001
YearsWithCurrManager	1	35.2585	<.0001
BonusReceivedRatio	1	4.1960	0.0405
EnvironmentSatisfact	3	7.7004	0.0526

Also, the below variables found to be not statistically significant and hence decided to remove it in the following order through step wise regression

1. JobRole
2. PerformanceRating
3. YearsSinceLastPromotion
4. MonthlyIncome
5. Education
6. StockOptionLevel
7. JobSatisfaction
8. WorkLifeBalance
9. PercentSalaryHike
10. Age
11. Hourly Rate
12. dailyrate
13. RelationshipSatisfaction

For the Weibull distribution found the log likelihood for the null model as below.

The LIFEREG Procedure

Model Information	
Data Set	PROJECT1.EXP_EMP
Dependent Variable	Log(YearsAtCompany)
Censoring Variable	lAttrition
Censoring Value(s)	0
Number of Observations	1000
Noncensored Values	115
Right Censored Values	885
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	2
Name of Distribution	Weibull
Log Likelihood	-366.9739729

Also, found the log ratio for the Weibull as follows

Obs	L_null	L_full	L	p_value
1	-366.97	-145.941	442.058	0

Interpretation of log ratio for Weibull:

The p-value zero indicates that at least one of the coefficient of covariate is not zero. This specifies that the model can be explained by at least one of the covariate. This also suggests that the full model is significantly different from the null model.

14. Modeling phase: Log-Normal Model

Young Employees:

In the process of building the model, important parameters have been removed by setting up the cut-off value for the probability of the chi2 value as 0.05 in several steps.

Firstly, ran Log Normal model with stepwise regression and found that the following variables are statistically significant. So, the below covariates are the ones which have been retained in the model.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Age	1	6.4215	0.0113
DistanceFromHome	1	19.3521	<.0001
OverTime	1	33.3338	<.0001
StockOptionLevel	3	20.7769	0.0001
BusinessTravel	2	12.4910	0.0019
EnvironmentSatisfact	3	24.6655	<.0001
Department	2	8.2379	0.0163
JobSatisfaction	3	15.9387	0.0012
MonthlyIncome	1	9.4984	0.0021
NumCompaniesWorked	1	14.4367	0.0001
TotalWorkingYears	1	6.4858	0.0109
WorkLifeBalance	3	12.2380	0.0066
YearsInCurrentRole	1	80.8269	<.0001
YearsWithCurrManager	1	62.8393	<.0001
BonusReceivedRatio	1	19.6577	<.0001

Also, the below variables found to be not statistically significant and hence decided to remove itin the following order through step wise regression

1. JobRole
2. Marital Status
3. HourlyRate
4. Education
5. YearsSinceLastPromot
6. DailyRate
7. Training Times Last Year
8. Relationship satisfaction
9. JobLevel
10. EducationField
11. PerformanceRating
12. PercentSalaryHike
13. JobInvolvement
14. Monthly Rate

For the Lognormal distribution found the log likelihood for the null model as below.

The LIFEREG Procedure

Model Information	
Data Set	PROJECT1.YOUNG_EMP
Dependent Variable	Log(YearsAtCompany)
Censoring Variable	IAttrition
Censoring Value(s)	0
Number of Observations	426
Noncensored Values	106
Right Censored Values	320
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	2
Zero or Negative Response	44
Name of Distribution	Lognormal
Log Likelihood	-246.6895335

Also, found the log ratio for the Lognormal as follows

Obs	L_null	L_full	L	p_value
1	-246.68	-42.28	408.8	0

Interpretation of log ratio for Lognormal:

The p-value zero indicates that at least one of the coefficient of covariate is not zero. This specifies that the model can be explained by at least one of the covariate. This also suggests that the full model is significantly different from the null model.

Experienced Employees:

Ran the Log Normal model with stepwise regression and found that the following variables are statistically significant. So, the below covariates are the ones which have been retained in the model.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DistanceFromHome	1	10.3820	0.0013
OverTime	1	27.6175	<.0001
StockOptionLevel	3	20.1802	0.0002
BusinessTravel	2	13.1815	0.0014
EnvironmentSatisfact	3	2.9574	0.3982
Department	2	14.4961	0.0007
EducationField	5	12.4304	0.0293
JobInvolvement	3	7.3436	0.0617
JobLevel	4	105.5547	<.0001
JobSatisfaction	3	10.4453	0.0151
NumCompaniesWorked	1	23.4318	<.0001
RelationshipSatisfac	3	11.1951	0.0107
TrainingTimesLastYea	1	11.3000	0.0008
YearsInCurrentRole	1	74.3484	<.0001
YearsWithCurrManager	1	73.8726	<.0001

Also, the below variables found to be not statistically significant and hence decided to remove itin the following order through step wise regression

1. JobRole
2. MonthlyIncome
3. PercentSalaryHike
4. hourlyrate
5. MaritalStatus
6. Education
7. PerformanceRating
8. BonusReceivedRatio
9. WorkLifeBalance
10. monthlyrate
11. TotalWorkingYears
12. dailyrate
13. YearsSinceLastPromotion

For the Log Normal distribution, found the log likelihood for the null model as below.

The LIFEREG Procedure

Model Information	
Data Set	PROJECT1.EXP_EMP
Dependent Variable	Log(YearsAtCompany)
Censoring Variable	IAttrition
Censoring Value(s)	0
Number of Observations	1000
Noncensored Values	115
Right Censored Values	885
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	2
Name of Distribution	Lognormal
Log Likelihood	-355.5976268

Also, found the log ratio for the Log Normal as follows

Obs	L_null	L_full	L	p_value
1	-355.59	-156.37	398.44	0

Interpretation of log ratio for Log Normal:

The p-value zero indicates that at least one of the coefficient of covariate is not zero. This specifies that the model can be explained by at least one of the covariate. This also suggests that the full model is significantly different from the null model.

15. Model comparison (Model Phase)

Young Employees:

Model Comparison

Comparing the exponential vs Weibull, lognormal vs Weibull and lognormal vs exponential models and the result obtained is as shown below.

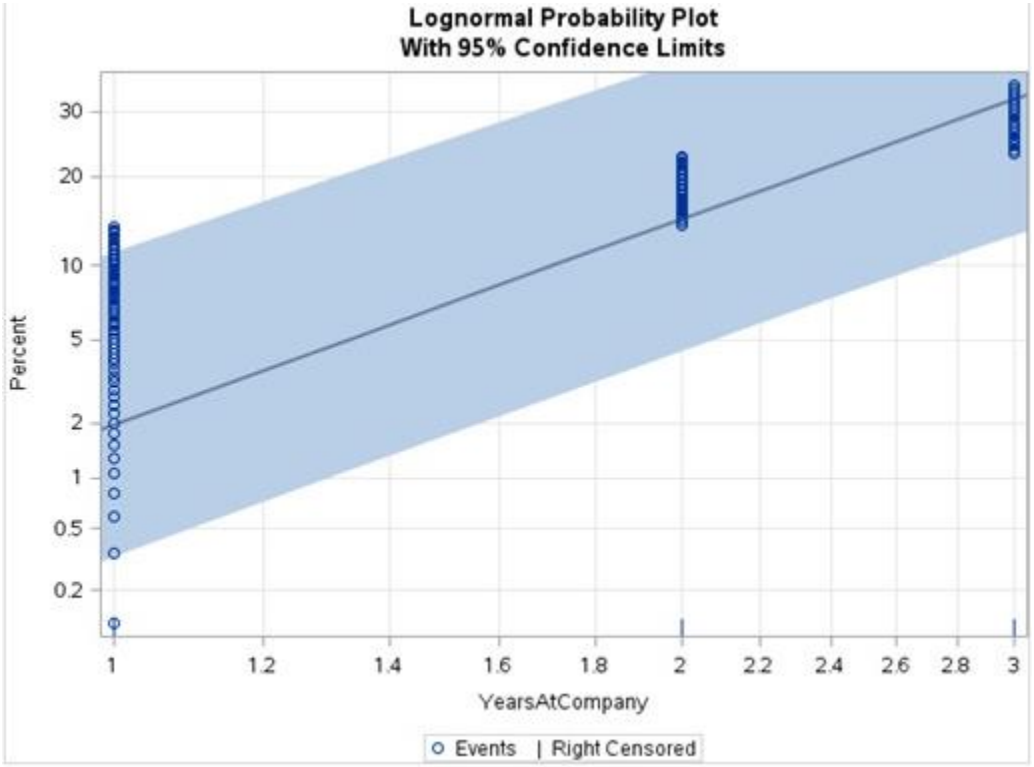
Obs	L_exponential	L_weibull	L_lognormal	LRTEW	LRTLW	LRTLE	p_valueEW	p_valueLW	p_valueLE
1	-188.26	-44.65	-42.28	287.22	-4.74	-291.96	0	.	.

Interpretation

The pvalue zero indicates that the coefficient of covariates for the model essentially differs and explanatory variables differ in explaining the model.

Since the log ratio for the LogNormal distribution is better than other distributions, so select **LogNormal distribution as the best model.**

Graphical Evaluation of the Models for Young Employees:



The above plot illustrates that the Lognormal distribution is the true distribution of the data and hence corresponding plot is returning a straight line.

Experienced Employees:

Model Comparison

Comparing the exponential vs Weibull, lognormal vs Weibull and lognormal vs exponential models and the result obtained is as shown below.

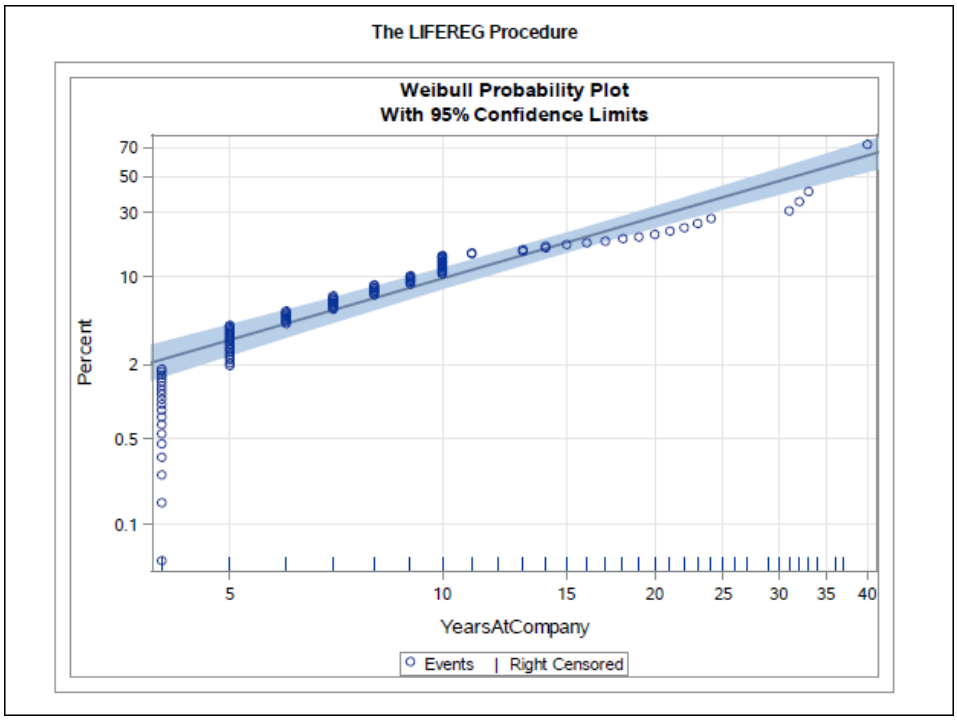
Obs	L_exponential	L_weibull	L_lognormal	LRTEW	LRTLW	LRTLE	p_valueEW	p_valueLW	p_valueLE
1	-280.83	-145.941	-156.37	269.778	20.8581	-248.92	0	.000004946	.

Interpretation

The pvalue zero indicates that the coefficient of covariates for the model essentially differs and explanatory variables differ in explaining the model.

Since the log ratio for the Weibull distribution is better than other distributions, **so select Weibull distribution as the best model.**

Graphical Evaluation of the Models for Experienced Employees:



The above plot illustrates that the Weibull distribution is the true distribution of the data and hence corresponding plot is returning a straight line.

16. Analysis (Assess Phase)

Young Employees:

Following are the estimates of the covariates of the **Log Normal model – best model selected**. It could be seen that most of the variables are highly significant.

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-0.0925	0.1246	-0.3366	0.1516	0.55	0.4575
Age		1	0.0066	0.0026	0.0015	0.0117	6.42	0.0113
DistanceFromHome		1	-0.0096	0.0022	-0.0139	-0.0053	19.35	<.0001
OverTime	No	1	0.2296	0.0398	0.1517	0.3075	33.33	<.0001
OverTime	Yes	0	0.0000	.	.	.	.	.
StockOptionLevel	0	1	-0.0434	0.0739	-0.1883	0.1015	0.35	0.5568
StockOptionLevel	1	1	0.1466	0.0796	-0.0094	0.3026	3.39	0.0655
StockOptionLevel	2	1	0.1041	0.1035	-0.0987	0.3070	1.01	0.3143
StockOptionLevel	3	0	0.0000	.	.	.	.	.
BusinessTravel	Non-Travel	1	0.0806	0.0720	-0.0605	0.2218	1.25	0.2626
BusinessTravel	Travel_Frequently	1	-0.1374	0.0432	-0.2221	-0.0528	10.12	0.0015
BusinessTravel	Travel_Rarely	0	0.0000	.	.	.	.	.
EnvironmentSatisfact	1	1	-0.2133	0.0493	-0.3100	-0.1166	18.70	<.0001
EnvironmentSatisfact	2	1	-0.1149	0.0545	-0.2218	-0.0081	4.45	0.0350
EnvironmentSatisfact	3	1	-0.0018	0.0517	-0.1031	0.0994	0.00	0.9716
EnvironmentSatisfact	4	0	0.0000	.	.	.	.	.
Department	Human Resources	1	0.1077	0.0772	-0.0435	0.2589	1.95	0.1628
Department	Research & Development	1	0.1140	0.0404	0.0348	0.1932	7.96	0.0048
Department	Sales	0	0.0000	.	.	.	.	.
JobSatisfaction	1	1	-0.1935	0.0576	-0.3064	-0.0806	11.28	0.0008
JobSatisfaction	2	1	-0.1329	0.0574	-0.2454	-0.0204	5.36	0.0206
JobSatisfaction	3	1	-0.1951	0.0519	-0.2968	-0.0933	14.13	0.0002
JobSatisfaction	4	0	0.0000	.	.	.	.	.
MonthlyIncome		1	0.0000	0.0000	0.0000	0.0001	9.50	0.0021
NumCompaniesWorked		1	-0.0310	0.0082	-0.0471	-0.0150	14.44	0.0001
TotalWorkingYears		1	0.0142	0.0056	0.0033	0.0251	6.49	0.0109
WorkLifeBalance	1	1	-0.0947	0.0797	-0.2509	0.0616	1.41	0.2350
WorkLifeBalance	2	1	-0.0483	0.0633	-0.1724	0.0757	0.58	0.4451
WorkLifeBalance	3	1	0.0832	0.0572	-0.0288	0.1953	2.12	0.1454
WorkLifeBalance	4	0	0.0000	.	.	.	.	.
YearsInCurrentRole		1	0.2686	0.0299	0.2101	0.3272	80.83	<.0001
YearsWithCurrManager		1	0.2317	0.0292	0.1744	0.2890	62.84	<.0001
BonusReceivedRatio		1	0.5699	0.1285	0.3180	0.8218	19.66	<.0001
Scale		1	0.2201	0.0151	0.1924	0.2519		

Complied the following list of potential reasons which can answer the following questions for the young employees –

1. **Who are leaving the company?**
2. **Why are they leaving?**

The explanations are written as per the contributing attribute

1. **Age**

As already considered the young employees as the employees with less than 3 years of work experience, there is a chance that employees could be older than expected. Employees in this cluster who are having lower age tend to leave the company.

It could be noticed that the coefficient of the 'Age' is '0.0066'. It means that the employees of greater age tend to work longer for this company.

2. **Distance from home**

This is another significant parameter that observed in the model results. With the increase in the distance between office and home, employees are tending to leave, however, the impact is less.

It shows that the coefficient of the parameter 'Distance from home' is '-0.0096' which is very less.

This inference is also matching with the thought process. And expected that if the distance between the office and home is less, employee would have no difficulties in commuting.

3. **Overtime**

Young Employees have a tendency to leave organization when they work for over time.

The coefficient of the 'Overtime' for level 'No' is 0.2296 which implies that employees are more likely to stay in the company for long time if overtime working hours are not imposed on them.

The inference is in line with the thought process. And also thought that if employees working hours are more, it could disturb their work-life balance.

4. **Stock Option Level**

Employees who are awarded with stock of option level '1' tend to stay in the company when compared to the employees who are awarded with stocks of other types.

The estimate of the 'StockOptionLevel' for level '1' is '0.1466' which implies that such employees are enjoying this stock level option and continuing with the company for long time.

5. **Business Travel**

Young employees tend to leave the company if they are being offered with opportunity to travel at frequent intervals.

The coefficient of the 'Business Travel' for level 'Travel_Frequently' is '-0.1374' which implies that they could leave the company.

Expected that when they travel frequently, they get enough exposure to switch the companies and the results are in line with the thoughts.

6. **Environment Satisfaction**

Employees who are not satisfied with the company facilities tend to leave the company.

The coefficient of the 'Environment Satisfaction' for level 'Low' and 'Medium' is '-0.2133' and '-0.1149' respectively. So, conclude that employees who are not enjoying the facilities of the company are leaving.

It is general that every employee would like to enjoy the company's facilities they are working with and if they are unhappy, it could lead to their attrition.

7. **Department**

Young employees in Research and Development department are not leaving the company.

Observed that the coefficient of the 'Business Travel' for level 'Research & Development' is '0.1140'. So, can conclude that such employees are not leaving the organization.

Usually, employees from R&D department would not have as much exposure as the sales department employees would have and this could be a reason for them to stay back in the company for more time.

8. **Number of companies worked**

Employees who has worked with more companies in their past experience tend to leave from the FermaLogis Company.

Noticed that the coefficient of the 'Number of companies worked' is '-0.0310' and it is significant based on the chi2 probability test. It means that employees who have their past history of working for different companies tend to leave more than employees who start their career with this organization.

9. **Total Working Years**

Young employees who does not have much experience are leaving the organization.

Saw that the coefficient of 'Total Working Years' is '0.0142' which implies that more experience people would tend stay in the organization for longer time.

10. **Job Satisfaction**

It is interesting to find out that young employees are leaving irrespective of their satisfaction with their current job.

Saw that the coefficient of 'Job Satisfaction' for level '1', '2' and '3' is '-0.1935', '-0.1329' and '-0.1951' respectively. Expected to see that if employees are more satisfied they would stay in the company for longer time. However, it is not observed with the given data.

11. **Work Life Balance**

Employees with good work life balance staying in the company for longer time.

Saw that that the coefficient of 'Work Life Balance' for level 'Bad', 'Good' and 'Very Good' is '-0.0947', '-0.0483' and '0.0832' respectively. It shows that employees who are rating the work life balance as 'Very Good' are retained more in the company. It also shows that even if the employees rate the company as 'Good', they are leaving the organization.

12. **Years in Current Role**

Young Employees who are in the same role for the longer period are staying longer in the company.

Observed that the coefficient of 'Years in Current Role' is '0.2686' which means that the employees are staying in the company if they are in the same role for the long time. It also implies that employees are working for longer periods even if they are not getting promotions. Such employees could be enjoying the work and highly satisfied with the company's facilities etc.

13. **Years with Current Manager**

Young employees who are working with the same manager for longer time tend to stay in the organization.

Noticed that the coefficient for 'Years with Current Manager' is '0.2317'. It means that employees with same manager tend to stay with the company longer which could be because of the employee rapport with his/her manager.

14. **Bonus Received Ratio**

Employees who receive bonus are staying with the company for the longer time.

Identified that the estimate for 'Bonus Received Ratio' is '0.5699'. It means that the employees who receive bonus would continue to work with the organization for longer time when compared to the employees who did not receive any bonus.

Experienced Employees:

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	1.4401	0.1830	1.0815	1.7988	61.93	<.0001
DistanceFromHome		1	-0.0069	0.0021	-0.0111	-0.0027	10.55	0.0012
OverTime	No	1	0.1971	0.0385	0.1217	0.2726	26.24	<.0001
OverTime	Yes	0	0.0000	.	.	.	.	.
BusinessTravel	Non-Travel	1	0.2399	0.0866	0.0702	0.4097	7.67	0.0056
BusinessTravel	Travel_Frequently	1	-0.0827	0.0412	-0.1634	-0.0019	4.03	0.0448
BusinessTravel	Travel_Rarely	0	0.0000	.	.	.	.	.
Department	Human Resources	1	0.1826	0.1145	-0.0418	0.4071	2.54	0.1107
Department	Research & Development	1	0.1487	0.0495	0.0517	0.2457	9.02	0.0027
Department	Sales	0	0.0000	.	.	.	.	.
EducationField	Human Resourc	1	0.1686	0.2158	-0.2544	0.5916	0.61	0.4347
EducationField	Life Sciences	1	0.1852	0.0554	0.0767	0.2937	11.19	0.0008
EducationField	Marketing	1	0.0778	0.0708	-0.0610	0.2166	1.21	0.2717
EducationField	Medical	1	0.1819	0.0596	0.0651	0.2987	9.32	0.0023
EducationField	Other	1	0.0781	0.0880	-0.0945	0.2506	0.79	0.3751
EducationField	Technical Deg	0	0.0000	.	.	.	.	.
JobInvolvement	1	1	-0.2406	0.0896	-0.4162	-0.0650	7.21	0.0073
JobInvolvement	2	1	-0.0855	0.0791	-0.2405	0.0696	1.17	0.2799
JobInvolvement	3	1	-0.0433	0.0746	-0.1895	0.1028	0.34	0.5613
JobInvolvement	4	0	0.0000	.	.	.	.	.

JobLevel	1	1	-0.1909	0.1267	-0.4393	0.0574	2.27	0.1319
JobLevel	2	1	0.1039	0.1176	-0.1266	0.3344	0.78	0.3769
JobLevel	3	1	0.0199	0.1120	-0.1996	0.2395	0.03	0.8589
JobLevel	4	1	0.3256	0.1270	0.0768	0.5744	6.58	0.0103
JobLevel	5	0	0.0000	.	.	.	.	.
MaritalStatus	Divorced	1	0.2392	0.0599	0.1218	0.3567	15.94	<.0001
MaritalStatus	Married	1	0.1137	0.0387	0.0378	0.1896	8.62	0.0033
MaritalStatus	Single	0	0.0000	.	.	.	.	.
NumCompaniesWorked		1	-0.0598	0.0067	-0.0729	-0.0466	79.53	<.0001
TotalWorkingYears		1	0.0470	0.0058	0.0357	0.0582	66.74	<.0001
TrainingTimesLastYea		1	0.0397	0.0155	0.0094	0.0700	6.58	0.0103
YearsInCurrentRole		1	0.0345	0.0062	0.0224	0.0466	31.38	<.0001
YearsWithCurrManager		1	0.0383	0.0065	0.0257	0.0510	35.26	<.0001
BonusReceivedRatio		1	-0.2373	0.1158	-0.4644	-0.0102	4.20	0.0405
EnvironmentSatisfact	1	1	-0.1000	0.0489	-0.1958	-0.0043	4.19	0.0407
EnvironmentSatisfact	2	1	0.0394	0.0543	-0.0671	0.1459	0.53	0.4685
EnvironmentSatisfact	3	1	-0.0139	0.0471	-0.1061	0.0784	0.09	0.7684
EnvironmentSatisfact	4	0	0.0000	.	.	.	.	.
Scale		1	0.1780	0.0126	0.1549	0.2046		
Weibull Shape		1	5.6185	0.3991	4.8882	6.4578		

Compiled the below potential factors which could answer the questions for Experienced employees.

a) **Who are leaving the company?**
b) **Why are they leaving?**

The explanations are written as per the contributing attribute

1. **<u>Overtime:</u>**

Experienced Employees have a tendency to leave organization when they work for over time.

Saw the coefficient of the 'Overtime' for level 'No' as 0.1971 which implies that employees are more likely to stay in the company for long time if overtime working hours are not imposed on them.

The inference is in line with the thought process. And also thought that if employees working hours are more, it could disturb their work-life balance.

2. **<u>Distance from Home:</u>**

Experienced Employees have a tendency to leave organization when their distance between home and office is more.

Saw the coefficient of the 'Distance From Home' as -0.0069 which implies that employees are more likely to stay in the company for long time if the commuting distance is less.

The inference is in line with the thought process.

3. **<u>Business Travel:</u>**

Experienced Employees have a tendency to leave organization when they do not get an opportunity to travel from the organization.

Saw the coefficient of the 'Business Travel' as -0.0827 which implies that employees are more likely to stay in the company for long time if the business travel frequency is more.

The inference is in line with the thought process. As long as the employees are allowed to travel from the company, experienced employees feel good to be as a part of the company.

4. **Department:**

Experienced Employees will stay in the organization if they are working for 'Research & Development' department.

Saw the coefficient of the 'Research & Development' as 0.1487 which implies experienced employees tend to continue in the same company.

The inference is in line with the thought process. Assumed that as long as the employees are working in Research department, they will have job satisfaction.

5. **Environment Satisfaction:**

Experienced Employees will tend to leave the organization if the work environment is not great.

Saw the coefficient of the 'Environment Satisfaction' for 1 as -0.1 which implies experienced employees are willing to work in the better workplace.

The inference is in line with the thought process. Assumed that if the working environment is better, then their efficiency will be good.

6. **Education Field:**

Experienced Employees will not tend to leave the organization if their background is from Life Science and Medical.

Saw the coefficient of the 'Education Field' for Life Science and Medical as 0.1852 and 0.1819 respectively which implies employees from the medical field tend to stay in the same organization.

Assumed that if the experienced employees are having the above qualifications then there is high chance they will not leave the organization. Also, this is a good sign in the company's perspective as the domain is related to pharmaceuticals.

7. **Job Involvement:**

Experienced Employees will tend to leave the organization if the rating given by their supervisors is less.

Saw the coefficient of the 'Job Involvement' for 1 as -0.24 which means that the lesser the rating for the employee more is the chance of leaving the organization.

Expected that if the job involvement factor is good for the employee then they will stay in the company.

8. **Job Level:**

Employees having more experience will stay in the organization if they are in the higher positions.

Saw the coefficient of the 'Job Level 'for 4 as 0.3256 which means that the higher job role employees are sticking to the same company. Also the coefficient of the 'Job Level 'for 1 as -0.1959 which means that the lower level employees are inclined to change the company.

9. **Bonus Received Ratio:**

Employees will tend to leave the organization if the ratio of average Bonus Received over the years is more.

Saw the coefficient of the 'Bonus Received Ratio' is -0.2373 which means that the bonus received may be during the initial period of the employee.

Also, the reason for leaving the organization may differ due to the other prominent factors as well.

10. **Marital Status:**

Employees will stay in the organization if the marital status of employees is either 'Divorced' or 'Married'.

Saw the coefficient of the 'Divorced' and 'Married' as 0.24 and 0.11 which implies that the experienced employees are more stable.

11. **Number of Companies Worked:**

Employees will tend to leave the organization if the employee has worked for more number of companies.

Saw the coefficient of the 'Num of Companies Worked 'as -0.0598 which means that the employees often shifting to other companies cannot be retained.

This is in line with the understanding. Also, experienced employees will not change the company very often.

12. **Total Working Years:**

Employees will be in the organization if the employee has spent most in the same company.

Saw the coefficient of the 'Total Working Years 'as 0.0470 which implies that the employee associated with the company for more period do not prefer to change the organization.

This is in line with the assumption. Also, it's a good sign for the company to retain the experienced employees.

13. **Years in Current Role:**

Employees will not tend to change the organization if the employee in the same role is continuing.

Saw the coefficient of the 'Years in Current Role 'is positive which means that the employees will not tend to change over the company.

This is quite contradictory. But more domain knowledge may aid in looking in depth of the behavior.

14. **Years with Current Manager:**

Employees will stay in the organization if the employee is working under same manager.

Saw the coefficient of the 'Years in Current Manager 'is positive which means that the employees tend to work with the same manager instead of different managers.

This is quite obvious. It takes time for the employee and manager to build a rapport for the better efficiency when working as a team.

15. **Training Times Last Year:**

Employees' willingness to be associated with the organization will be better if they receive the professional trainings every year.

Saw the coefficient of the 'Training Times Last Year 'is 0.0397 which implies that the experienced employees wish to attend the training sessions.

This is a positive aspect for the company. If the company provides the training for the experienced employees it will help in retaining them.

17. Conclusion and Recommendations

Generic Recommendations for all group of employees.

Based on all the analysis above, the following recommendation towards the company management are proposed, to retain the employees for a longer period of time.

- Travelling benefits, like ride sharing options, shuttle services, travel allowances etc., could be effective for employees who travel longer distances to come to work.
- Also business travel opportunities, which entails more experience in the job function should be offered to all level of employees and not only be constricted to young employees only. This will boost confidence level of the comparatively younger generation of employees and they will tend to stay in the company for more valuable on the job experience through these business travels.

Recommendation for retaining young employees

Following are the recommendations to retain the young employees in FermaLogis Company.

- As more number of employees are leaving from Sales team, HR team should identify the reasons for the attrition and identify the competitors in the market who are offering more opportunities to these people, so that HR have an idea to effectively strategize and incentivized process for the sales people to retain them.

- HR & Management team should concentrate on the facilities provided to employees which will positively increase employee satisfaction, like for example, young people like more amenities, better work place perks, excellent benefits etc.
- It shows, that long hours are a key factor of leaving for young employees. So management should focus on correct work allocation and distribution and staffing priorities, to optimize the level of work allocated to a specific employee, which would ensure that the work will be finished during normal hours in majority of the cases.

Recommendation for retaining experienced employees

Following are the recommendations to retain the experienced employees in FermaLogis Company.

- Experience employees who stayed in the company longer, did not just stay because they love the job, but also they love the reporting manager and team they work with. So, frequent reshuffling of the staffing and manager change is not recommended for the experienced employees as it gives them a sense of discomfort towards adjusting to new manager and team, which can lead to frequent attrition.
- Also management should think of allocating budget for continuous training for experience group of employees as this section of the employee population values new learning to advance their competencies, which would eventually be used for the company itself and will keep these employees motivated and hence would play a significant role in retaining them.

Appendix

Table 1:

Attribute-Name	Attribute Definition	Sample value(s)
Age	age of the employee when this dataset was created	41
Attrition	shows whether the employee left the company or not	'Yes'
BusinessTravel	shows how much travel employee makes	Travel_Rarely
DailyRate	daily compensation of employee before any cuts/taxes	1102
Department	shows the department of the employee when this dataset was created	Sales
DistanceFromHome	commuting distance for the employee in miles	1
Education	1 - 'Below College' , 2 -'College',3 - 'Bachelor',4 - 'Master',5 - 'Doctor'	2
EducationField	shows the education field of the employee	Life Sciences
EmployeeCount	a field used for aggregation calculations	1
EmloyeeNumber	the ID of the employee	1
EnvironmentSatisfaction	a score showing how much the employee is satisfied with company's facilities (1 - 'Low',2 - 'Medium',3 - 'High', 4 - 'Very High')	2
Gender	shows the gender of the employee	Female
HourlyRate	hourly compensation of employee before any cuts/taxes	94
JobInvolvement	a score given to the employee by supervisors how much the employee is involved in company's operations (1 -'Low',2 -'Medium',3 -'High',4 -'Very High')	3
JobRole	shows the job role of the employee in the company	Sales Executive
JobLevel	shows the management level of the employee	2
JobSatisfaction	shows the last survey result of the employee about his\her job satisfaction	4
MaritalStatus	shows the marital status of the company	Single
MonthlyIncome	shows the monthly income of the employee	5993
MontlyRate	monthly compensation of employee before any cuts/taxes	19479
NumCompaniesWorked	the number of companies the employee worked before starting in the company	8
Over18	shows whether the employee is over 18 years old	Y
OverTime	shows whether employee works overtime more than 10 hours a week	Yes
PercentSalaryHike	shows the agreed yearly salary rise percent	11
PerformanceRating	a score given to the employee by supervisors how good was the performance of the employee last year (1 -'Low',2 - 'Good',3 -'Excellent',4 - 'Outstanding')	3

RelationshipSatisfaction	shows the last survey result of the employee about his\her satisfaction with other employees in the company(1- 'Low',2- 'Medium',3- 'High',4- 'Very High')	1
StandardHours	number of hours employee works for one payroll period (two weeks)	80
StockOptionLevel	shows the stock option for the employee. If your analyses give significant results for this variable, you can refer to that group as "employees having stock option level x"	0
TotalWorkingYears	shows the time the employee worked as a professional (at any company)	8
TrainingTimesLastYear	shows the number of training programs employee has attended last year	0
WorkLifeBalance	shows the employee satisfaction of the work load (4 is the highest satisfaction level) (1- 'Bad',2 - 'Good',3- 'Better',4 -'Best')	1
YearsAtCompany	Tenure at the company	6
YearsInCurrentRole	the number of years employee works in the current position	4
YearsSinceLastPromotion	shows the number of years passed since the last promotion	0
YearsWithCurrentManager	Shows the number of years with the current supervisor.	5
bonus_1-40	Shows whether the employee received bonus payments in the last 40 years. bonus_1 is last year	0……

SAS Code

```
/*=============================================================================================*/
/* This is the SAS code file for Project 1 under the survival analysis class for Summer-2017 Semester *
/* Author of this code : Team 9 * /
/* Team members are as follows */
/* 1. Rajarshi Das, 2. Xueling Chen, 3. Yu Li */
/* 4. Vinit Gupta, 5. Rahul Manchanda, 6. Pawan Shivhare */
/* We have used the analysis done by this code and put all the details under the project reaport
document named : Team9-Final-Report-Project1.docx */
/* This code is written by the members of team 9 only, we have used class material and
internet as resourecs for reference only */
/*********************************************************************************************/
/* With the following statement, we have created our project library */

Libname Project1 'P:\FINAL-Project1';

/* Data preparation step, creating the sas file from the .CSV file*/
PROC IMPORT OUT=project1.Team9_csv FILE='P:\FINAL-Project1\FermaLogis.csv' DBMS=CSV
        REPLACE;
RUN;
PROC PRINT data= project1.Team9_csv;
Title' SAS data file to be used by team 9 for Project 1 Analysis';
RUN;
/*Adding few more variables based on certain given conditions*/
/*These variables will help, our further analysis of the data */
DATA project1.Team9_csv;
    SET project1.Team9_csv;

    IF NumCompaniesWorked=0 or NumCompaniesWorked=1 or NumCompaniesWorked=2 or
        NumCompaniesWorked=3 or NumCompaniesWorked=4 THEN
            NumCompaniesWorked1='Less than or equal to 4 Companies';

    IF NumCompaniesWorked=5 or NumCompaniesWorked=6 or NumCompaniesWorked=7 or
        NumCompaniesWorked=8 or NumCompaniesWorked=9 THEN
            NumCompaniesWorked1='Greater than 4 Companies';
    NumBonusAwarded=SUM(bonus_1, bonus_2, bonus_3, bonus_4, bonus_5, bonus_6,
        bonus_7, bonus_8, bonus_9, bonus_10, bonus_11, bonus_12, bonus_13, bonus_14,
        bonus_15, bonus_16, bonus_17, bonus_18, bonus_19, bonus_20, bonus_21,
        bonus_22, bonus_23, bonus_24, bonus_25, bonus_26, bonus_27, bonus_28,
        bonus_29, bonus_30, bonus_31, bonus_32, bonus_33, bonus_34, bonus_35,
        bonus_36, bonus_37, bonus_38, bonus_39, bonus_40);
    BonusReceivedRatio=NumBonusAwarded/YearsAtCompany;

    IF YearsAtCompany <=3 Then
        Datasplit='YOUNG_EMP';

    IF YearsAtCompany > 3 Then
        Datasplit='Exp_Emp';

    IF Attrition="No" THEN
        IAttrition=0;
    ELSE
        IAttrition=1;

    IF BonusReceivedRatio=. THEN
        BonusReceivedRatio=0;

    IF overtime="No" THEN
        Iovertime=0;
    ELSE
        Iovertime=1;

    IF BusinessTravel="Travel_Frequently" then
        IBusinessTravel=2;
    ELSE IF BusinessTravel="Travel_Rarely" then
        IBusinessTravel=1;
    ELSE IF BusinessTravel="Non-Travel" then
        IBusinessTravel=0;

    IF MaritalStatus='Married' then
        IMaritalStatus=1;
    ELSE IF MaritalStatus='Divorced' then
        IMaritalStatus=0;
    ELSE IF MaritalStatus='Single' then
        IMaritalStatus=2;
RUN;
/* Now, we are creating two separate data files, one for Young Employee and one for
experienced employee */
```

```
78  DATA project1.Young_emp project1.Exp_emp;
79      SET project1.Team9_csv;
80
81      If UPCASE(Datasplit)='YOUNG_EMP' THEN
82          OUTPUT project1.Young_emp;
83      Else if UPCASE(Datasplit)='EXP_EMP' THEN
84          output project1.Exp_emp;
85  RUN;
86  /* Now we will try to explore the data, to analyze the contribution of different parameters
87  towards employee attrition for the Fermalogis company */
88  /* Analyzing contribution of Marital status towards attrition */
89  PROC SGPLOT data=project1.Young_emp;
90      VBAR attrition / GROUP=MaritalStatus groupdisplay=cluster;
91      TITLE 'Attrition by Marital Status for Young Employees';
92  RUN;
93  /* Analyzing contribution of Department towards attrition */
94  PROC SGPLOT data=project1.Young_emp;
95      VBAR attrition / GROUP=department groupdisplay=cluster;
96      TITLE 'Attrition by department for Young Employees';
97  RUN;
98  /*Analyzing Contribution of Jobrole towards Attrition*/
99  PROC SGPLOT data=project1.Young_emp;
100     VBAR attrition / GROUP=jobrole groupdisplay=cluster;
101     TITLE 'Attrition by jobrole for Young Employees';
102 RUN;
103
104 /* Analyzing Contribution of Age towards Attrition*/
105 proc sql ;
106     create table rec_tablevisual as select count(employeeNumber) as empCount, Age,
107         Attrition from project1.Young_emp group by Age, Attrition;
108
109 PROC SGPLOT DATA=rec_tablevisual;
110     SERIES X=Age Y=empCount / group=attrition;
111     XAXIS TYPE=DISCRETE;
112     XAXIS LABEL='Age of Employee';
113     YAXIS LABEL='Count of Employees' VALUES=(0 TO 30 BY 2);
114     TITLE 'Attrition rate by Age for Young Employees';
115 RUN;
116
117 PROC SGPLOT DATA=project1.team9_csv;
118     VBAR Attrition/ RESPONSE=MonthlyIncome STAT=MEAN;
119     Title 'Bar plot for Attrition vs Monthly Income of the employee';
120 run;
121
122 PROC SGPLOT DATA=project1.Young_emp;
123     VBAR Attrition/ RESPONSE=MonthlyIncome STAT=MEAN;
124     Title 'Bar plot for Attrition vs Monthly Income for Young employees';
125 run;
126
127 PROC SGPLOT DATA=project1.Exp_emp;
128     VBAR Attrition/ RESPONSE=MonthlyIncome STAT=MEAN;
129     Title 'Bar plot for Attrition vs Monthly Income for Experienced employees';
130 run;
131
132 PROC SGPLOT DATA=project1.team9_csv;
133     VBAR Attrition/ GROUP=OverTime;
134     Title 'Bar plot for Attrition vs Overtime ';
135 run;
136
137 PROC SGPLOT DATA=project1.Young_emp;
138     VBAR Attrition/ GROUP=OverTime;
139     Title 'Bar plot for Attrition vs Over Time for Young employees';
140 run;
141
142 PROC SGPLOT DATA=project1.Exp_emp;
143     VBAR Attrition/ GROUP=OverTime;
144     Title 'Bar plot for Attrition vs Over Time for Experienced employees';
145 run;
146
147 PROC FREQ DATA=project1.team9_csv;
148     TABLES Attrition * OverTime / CHISQ;
149     TITLE 'Association of Attrition and OverTime ';
150 RUN;
151
152 PROC FREQ DATA=project1.Young_emp;
153     TABLES Attrition * OverTime / CHISQ;
154     TITLE 'Association of Attrition and OverTime for Young employees ';
155 RUN;
156
157 PROC FREQ DATA=project1.Exp_emp;
158     TABLES Attrition * OverTime / CHISQ;
159     TITLE 'Association of Attrition and OverTime for Experienced employee';
160 RUN;
```

```
161
162 PROC SGPLOT DATA=project1.team9_csv;
163     VBAR Attrition/ GROUP=StockOptionLevel;
164     Title 'Bar plot for Attrition vs StockOptionLevel ';
165 run;
166
167 PROC SGPLOT DATA=project1.Young_emp;
168     VBAR Attrition/ GROUP=StockOptionLevel;
169     Title 'Bar plot for Attrition vs StockOptionLevel for Young employee ';
170 run;
171
172 PROC SGPLOT DATA=project1.Exp_emp;
173     VBAR Attrition/ GROUP=StockOptionLevel;
174     Title 'Bar plot for Attrition vs StockOptionLevel for Experienced employee ';
175 run;
176
177 PROC FREQ DATA=project1.team9_csv;
178     TABLES Attrition * StockOptionLevel / CHISQ;
179     TITLE 'Association of Attrition and Stock Option Level ';
180 RUN;
181
182 PROC FREQ DATA=project1.Young_emp;
183     TABLES Attrition * StockOptionLevel / CHISQ;
184     TITLE 'Association of Attrition and StockOptionLevel for Young employees ';
185 RUN;
186
187 PROC FREQ DATA=project1.Exp_emp;
188     TABLES Attrition * StockOptionLevel / CHISQ;
189     TITLE
190         'Association of Attrition and StockOptionLevel for Experienced employees';
191 RUN;
192
193 PROC SGPLOT DATA=project1.team9_csv;
194     VBAR Attrition/ GROUP=BusinessTravel;
195     Title 'Bar plot for Attrition vs Business Travel ';
196 run;
197
198 PROC SGPLOT DATA=project1.Young_emp;
199     VBAR Attrition/ GROUP=BusinessTravel;
200     Title 'Bar plot for Attrition vs BusinessTravel for Young employees ';
201 run;
202
203 PROC SGPLOT DATA=project1.Exp_emp;
204     VBAR Attrition/ GROUP=BusinessTravel;
205     Title 'Bar plot for Attrition vs BusinessTravel for Experienced employees ';
206 run;
207
208 PROC FREQ DATA=project1.team9_csv;
209     TABLES Attrition * BusinessTravel / CHISQ;
210     TITLE 'Association of Attrition and BusinessTravel ';
211 RUN;
212
213 PROC FREQ DATA=project1.Young_emp;
214     TABLES Attrition * BusinessTravel / CHISQ;
215     TITLE 'Association of Attrition and BusinessTravel for Young employees ';
216 RUN;
217
218 PROC FREQ DATA=project1.Exp_emp;
219     TABLES Attrition * BusinessTravel / CHISQ;
220     TITLE 'Association of Attrition and BusinessTravel for Experienced employees';
221 RUN;
222
223 PROC FREQ DATA=project1.team9_csv;
224     TABLES Attrition * EnvironmentSatisfaction / CHISQ;
225     TITLE 'Association of Attrition and EnvironmentSatisfaction ';
226 RUN;
227
228 PROC FREQ DATA=project1.Young_emp;
229     TABLES Attrition * EnvironmentSatisfaction / CHISQ;
230     TITLE
231         'Association of Attrition and EnvironmentSatisfaction for Young employees ';
232 RUN;
233
234 PROC FREQ DATA=project1.Exp_emp;
235     TABLES Attrition * EnvironmentSatisfaction / CHISQ;
236     TITLE 'Association of Attrition and EnvironmentSatisfaction for Experienced emmployees ';
237 RUN;
238
239 PROC FREQ DATA=project1.team9_csv;
240     TABLES Attrition * JobSatisfaction / CHISQ;
241     TITLE 'Association of Attrition and JobSatisfaction ';
242 RUN;
243
244 PROC FREQ DATA=project1.Young_emp;
245     TABLES Attrition * JobSatisfaction / CHISQ;
246     TITLE 'Association of Attrition and JobSatisfaction for Young emplloyees';
247 RUN;
```

```
248
249 PROC FREQ DATA=project1.Exp_emp;
250     TABLES Attrition * JobSatisfaction / CHISQ;
251     TITLE 'Association of Attrition and JobSatisfaction for Experienced employees';
252 RUN;
253
254 PROC CORR DATA=project1.team9_csv PLOTS=scatter;
255     VAR YearsAtCompany;
256     WITH MonthlyIncome;
257     TITLE 'Correlations for MonthlyIncome with YearsAtCompany';
258 RUN;
259
260 PROC CORR DATA=project1.Young_emp PLOTS=scatter;
261     VAR YearsAtCompany;
262     WITH MonthlyIncome;
263     TITLE 'Correlations for MonthlyIncome with YearsAtCompany for Young employees';
264 RUN;
265
266 PROC CORR DATA=project1.Exp_emp PLOTS=scatter;
267     VAR YearsAtCompany;
268     WITH MonthlyIncome;
269     TITLE
270         'Correlations for MonthlyIncome with YearsAtCompany for Experienced employees';
271 RUN;
272
273 PROC anova data=project1.team9_csv;
274     CLASS Attrition;
275     MODEL MonthlyIncome=Attrition;
276     MEANS Attrition/SCHEFFE;
277     TITLE 'MonthlyIncome based on Attrition';
278
279 PROC anova data=project1.Young_emp;
280     CLASS Attrition;
281     MODEL MonthlyIncome=Attrition;
282     MEANS Attrition/SCHEFFE;
283     TITLE 'MonthlyIncome based on Attrition for Young employees';
284
285 PROC anova data=project1.Exp_emp;
286     CLASS Attrition;
287     MODEL MonthlyIncome=Attrition;
288     MEANS Attrition/SCHEFFE;
289     TITLE 'MonthlyIncome based on Attrition for Experienced employees';
290
291 PROC SGPLOT DATA=project1.team9_csv;
292     VBAR NumCompaniesWorked1/ GROUP=Attrition GROUPDISPLAY=cluster STAT=percent;
293     Title 'Bar plot for Attrition vs Num of CompaniesWorked ';
294 run;
295
296 PROC SGPLOT DATA=project1.Young_emp;
297     VBAR NumCompaniesWorked1/ GROUP=Attrition GROUPDISPLAY=cluster STAT=percent;
298     Title 'Bar plot for Attrition vs Num of Companies Worked for Young Employees';
299 run;
300
301 PROC SGPLOT DATA=project1.Exp_emp;
302     VBAR NumCompaniesWorked1/ GROUP=Attrition GROUPDISPLAY=cluster STAT=percent;
303     Title
304         'Bar plot for Attrition vs Num of Companies Worked for Experienced employees ';
305 run;
306
307 PROC FREQ DATA=project1.team9_csv;
308     TABLES Attrition * NumCompaniesWorked1 / CHISQ;
309     TITLE 'Association of Attrition and Num of CompaniesWorked ';
310 RUN;
311
312 PROC FREQ DATA=project1.Young_emp;
313     TABLES Attrition * NumCompaniesWorked1 / CHISQ;
314     TITLE 'Association of Attrition and Num of CompaniesWorked for Young Employees ';
315 RUN;
316
317 PROC FREQ DATA=project1.Exp_emp;
318     TABLES Attrition * NumCompaniesWorked1 / CHISQ;
319     TITLE
320         'Association of Attrition and Num of CompaniesWorked for Experienced employees';
321 RUN;
322
323 PROC SGPLOT DATA=project1.team9_csv;
324     VBAR Attrition/ Response=NumBonusAwarded STAT=Mean;
325     Title 'Bar plot for Attrition vs Num of BonusAwarded ';
326 run;
327
328 PROC SGPLOT DATA=project1.Young_emp;
329     VBAR Attrition/ Response=NumBonusAwarded STAT=Mean;
330     Title 'Bar plot for Attrition vs Num of BonusAwarded for Youung Employees ';
331 run;
```

```
PROC SGPLOT DATA=project1.Exp_emp;
    VBAR Attrition/ Response=NumBonusAwarded STAT=Mean;
    Title 'Bar plot for Attrition vs Num of BonusAwarded for Experienced Employees ';
run;

PROC SGPLOT DATA=project1.team9_csv;
    VBAR Attrition/ Response=DistanceFromHome STAT=Mean;
    Title 'Bar plot for Attrition vs DistanceFromHome ';
run;

PROC SGPLOT DATA=project1.Young_emp;
    VBAR Attrition/ Response=DistanceFromHome STAT=Mean;
    Title 'Bar plot for Attrition vs DistanceFromHome for Young Employees ';
run;

PROC SGPLOT DATA=project1.Exp_emp;
    VBAR Attrition/ Response=DistanceFromHome STAT=Mean;
    Title 'Bar plot for Attrition vs DistanceFromHome for Experienced Employees ';
run;
/* In the following sections, we are working different Survival Plots for
further analysis of the data from survival perspective */
/* In the survival plots as well, we have categorized the data at different level
to closely analyze our data */
/*Surival plots for employees till 3years of Tenure */
DATA project1.project1.rec_upto3 ;
    set project1.Young_emp;
    IF Iattrition=0 then
        YearsAtCompany=4;
RUN;

proc lifetest data=project1.rec_upto3  method=life intervals=0 1 2 3 4 plots=(s, h);
    time YearsAtCompany*IAttrition(0);
    strata overtime;
    test StockOptionLevel monthlyrate TotalWorkingYears EnvironmentSatisfaction
        age DistanceFromHome IBusinessTravel IMaritalStatus NumCompaniesWorked
        JobSatisfaction HourlyRate;
    title 'Survival plot for employees with 3 years of Tenure';
run;
/*************************************************************************************************/
/*Surival plots for employees with greater than 3years of Tenure */
proc lifetest data=project1.Exp_emp method=life intervals=4 10 20 30 40
        plots=(s, h);
    time YearsAtCompany*IAttrition(0);
    strata overtime;
    test StockOptionLevel monthlyrate TotalWorkingYears EnvironmentSatisfaction
        age DistanceFromHome IBusinessTravel IMaritalStatus NumCompaniesWorked
        JobSatisfaction HourlyRate;
run;

/*Hazard plots for both the categories (YearsAtcompany <= 3 and > 3) */
proc lifetest data=project1.rec_upto3  method=life intervals=0 1 2 3 4 plots=(h);
    time YearsAtCompany*IAttrition(0);
run;

proc lifetest data=project1.Exp_emp method=life intervals=4 10 20 30 40
        plots=(h);
    time YearsAtCompany*IAttrition(0);
run;

/* Strata plot for gender based analysis */
proc lifetest data=project1.rec_upto3  method=life intervals=0 1 2 3 4 plots=(s);
    time YearsAtCompany*IAttrition(0);
    strata gender;
    test StockOptionLevel TotalWorkingYears EnvironmentSatisfaction age
        DistanceFromHome IBusinessTravel IMaritalStatus JobSatisfaction;
run;

proc lifetest data=project1.Exp_emp method=life intervals=4 10 20 30 40
        plots=(s);
    time YearsAtCompany*IAttrition(0);
    strata gender;
    test StockOptionLevel TotalWorkingYears EnvironmentSatisfaction age
        DistanceFromHome IBusinessTravel IMaritalStatus JobSatisfaction;
run;

/* Strata plot for Overtime*/
proc lifetest data=project1.rec_upto3  method=life intervals=0 1 2 3 4 plots=(s);
    time YearsAtCompany*IAttrition(0);
    strata Overtime;
    test StockOptionLevel TotalWorkingYears EnvironmentSatisfaction age
        DistanceFromHome IBusinessTravel IMaritalStatus JobSatisfaction;
run;
```

```
/* Strata plot for Business Travel*/
proc lifetest data=project1.rec_upto3  method=life intervals=0 1 2 3 4 plots=(s);
    time YearsAtCompany*IAttrition(0);
    strata BusinessTravel;
    test StockOptionLevel TotalWorkingYears EnvironmentSatisfaction age
        DistanceFromHome IBusinessTravel IMaritalStatus JobSatisfaction;
run;

proc lifetest data=project1.Exp_emp method=life intervals=4 10 20 30 40
        plots=(s);
    time YearsAtCompany*IAttrition(0);
    strata BusinessTravel;
    test StockOptionLevel TotalWorkingYears EnvironmentSatisfaction age
        DistanceFromHome IBusinessTravel IMaritalStatus JobSatisfaction;
run;

/* Now we are working on analyzing the data running various statistical models
available for each section and categories of employees (like for example, Young vs. experienced
and also if young, we are running models for employees having less than 3 years of tenure */

***Analysis of Young Employees;
**Start of LogNormal Model;
*LogNormal for less than 3 years for null;

PROC LIFEREG DATA=project1.Young_emp;
    MODEL YearsAtCompany*IAttrition(0)= /DISTRIBUTION=lnormal;
RUN;

*LogNormal for less than 3 years;

PROC LIFEREG DATA=project1.Young_emp;
    CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
        Department JobSatisfaction WorkLifeBalance;
    MODEL YearsAtCompany*IAttrition(0)=Age distancefromhome overtime
        StockOptionLevel BusinessTravel EnvironmentSatisfaction Department
        JobSatisfaction MonthlyIncome NumCompaniesWorked TotalWorkingYears
        WorkLifeBalance YearsInCurrentRole YearsWithCurrManager BonusReceivedRatio
        /DISTRIBUTION=lnormal;
RUN;

*Find Loglikelihood statistic for LogNormal;

DATA project1.calculateLogRatioForLogNormal;
    L_null=-246.68;
    L_full=-42.28;
    L=2 * ABS(L_full - L_null);
    p_value=1 - probchi(L, 14);
RUN;

PROC PRINT DATA=project1.calculateLogRatioForLogNormal;
RUN;

**End of LogNormal Model;
**Start of Exponential Model;
*Exponential Model for less than 3 years;

PROC LIFEREG DATA=project1.Young_emp;
    CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
        Department JobSatisfaction WorkLifeBalance;
    MODEL YearsAtCompany*IAttrition(0)=Age distancefromhome overtime
        StockOptionLevel BusinessTravel EnvironmentSatisfaction Department
        JobSatisfaction MonthlyIncome NumCompaniesWorked WorkLifeBalance
        YearsInCurrentRole YearsWithCurrManager /DISTRIBUTION=exponential;
RUN;

*Exponential for less than 3 years for null;

PROC LIFEREG DATA=project1.Young_emp;
    MODEL YearsAtCompany*IAttrition(0)= /DISTRIBUTION=exponential;
RUN;

*Find Loglikelihood statistic for Exponential;

DATA project1.calculateLogRatioForExponential;
    L_null=-280.742;
    L_full=-188.267;
    L=2 * ABS(L_full - L_null);
    p_value=1 - probchi(L, 12);
RUN;

PROC PRINT DATA=project1.calculateLogRatioForExponential;
RUN;

**End of Exponential model;
```

```
**Start of Exponential Model;
*Exponential Model for less than 3 years;

PROC LIFEREG DATA=project1.Young_emp;
    CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
        Department JobSatisfaction WorkLifeBalance;
    MODEL YearsAtCompany*IAttrition(0)=Age distancefromhome overtime
        StockOptionLevel BusinessTravel EnvironmentSatisfaction Department
        JobSatisfaction MonthlyIncome NumCompaniesWorked WorkLifeBalance
        YearsInCurrentRole YearsWithCurrManager /DISTRIBUTION=exponential;
RUN;

*Exponential for less than 3 years for null;

PROC LIFEREG DATA=project1.Young_emp;
    MODEL YearsAtCompany*IAttrition(0)= /DISTRIBUTION=exponential;
RUN;

*Find Loglikelihood statistic for Exponential;

DATA project1.calculateLogRatioForExponential;
    L_null=-280.742;
    L_full=-188.267;
    L=2 * ABS(L_full - L_null);
    p_value=1 - probchi(L, 12);
RUN;

PROC PRINT DATA=project1.calculateLogRatioForExponential;
RUN;

**End of Exponential model;
**Start of Weibull Model;
*Weibull for less than 3 years;

PROC LIFEREG DATA=project1.Young_emp;
    CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
        Department JobSatisfaction WorkLifeBalance;
    MODEL YearsAtCompany*IAttrition(0)=distancefromhome monthlyrate overtime
        StockOptionLevel BusinessTravel EnvironmentSatisfaction Department
        JobSatisfaction MonthlyIncome NumCompaniesWorked TotalWorkingYears
        WorkLifeBalance YearsInCurrentRole YearsWithCurrManager BonusReceivedRatio
        /DISTRIBUTION=weibull;
RUN;

*Weibull for less than 3 years for null;

PROC LIFEREG DATA=project1.Young_emp;
    MODEL YearsAtCompany*IAttrition(0)=/DISTRIBUTION=weibull;
RUN;

*Find Loglikelihood statistic for Weibull;

DATA project1.calculateLogRatioForWeibull;
    L_null=-253.164;
    L_full=-44.656;
    L=2 * ABS(L_full - L_null);
    p_value=1 - probchi(L, 14);
RUN;

PROC PRINT DATA=project1.calculateLogRatioForWeibull;
RUN;

**End of Weibull Model
```

```
**Start of Exponential Model;
*Exponential Model for less than 3 years;

PROC LIFEREG DATA=project1.Young_emp;
    CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
        Department JobSatisfaction WorkLifeBalance;
    MODEL YearsAtCompany*IAttrition(0)=Age distancefromhome overtime
        StockOptionLevel BusinessTravel EnvironmentSatisfaction Department
        JobSatisfaction MonthlyIncome NumCompaniesWorked WorkLifeBalance
        YearsInCurrentRole YearsWithCurrManager /DISTRIBUTION=exponential;
RUN;

*Exponential for less than 3 years for null;

PROC LIFEREG DATA=project1.Young_emp;
    MODEL YearsAtCompany*IAttrition(0)= /DISTRIBUTION=exponential;
RUN;

*Find Loglikelihood statistic for Exponential;

DATA project1.calculateLogRatioForExponential;
    L_null=-280.742;
    L_full=-188.267;
    L=2 * ABS(L_full - L_null);
    p_value=1 - probchi(L, 12);
RUN;

PROC PRINT DATA=project1.calculateLogRatioForExponential;
RUN;

**End of Exponential model;
**Start of Weibull Model;
*Weibull for less than 3 years;

PROC LIFEREG DATA=project1.Young_emp;
    CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
        Department JobSatisfaction WorkLifeBalance;
    MODEL YearsAtCompany*IAttrition(0)=distancefromhome monthlyrate overtime
        StockOptionLevel BusinessTravel EnvironmentSatisfaction Department
        JobSatisfaction MonthlyIncome NumCompaniesWorked TotalWorkingYears
        WorkLifeBalance YearsInCurrentRole YearsWithCurrManager BonusReceivedRatio
        /DISTRIBUTION=weibull;
RUN;

*Weibull for less than 3 years for null;

PROC LIFEREG DATA=project1.Young_emp;
    MODEL YearsAtCompany*IAttrition(0)=/DISTRIBUTION=weibull;
RUN;

*Find Loglikelihood statistic for Weibull;

DATA project1.calculateLogRatioForWeibull;
    L_null=-253.164;
    L_full=-44.656;
    L=2 * ABS(L_full - L_null);
    p_value=1 - probchi(L, 14);
RUN;

PROC PRINT DATA=project1.calculateLogRatioForWeibull;
RUN;

**End of Weibull Model
```

```
643
644 PROC PRINT data=project1.calculateLogRatio_weibull_exp;
645 RUN;
646
647 **End of Weibull Model;
648 **Start of Exponential Model;
649 * Null Model for exponential for Experienced Employees;
650
651 PROC LIFEREG DATA=project1.Exp_emp;
652     *null model to test null hypothesis;
653     MODEL YearsAtCompany*IAttrition(0)= / D=exponential;
654
655     /*Loglikelihood of null hypothesis*/
656 RUN;
657
658 *exponential for Experienced Employees;
659
660 PROC LIFEREG DATA=project1.Exp_emp;
661     CLASS overtime StockOptionLevel BusinessTravel EnvironmentSatisfaction
662         Department Education EducationField JobInvolvement JobLevel JobSatisfaction
663         WorkLifeBalance;
664     MODEL YearsAtCompany*IAttrition(0)=distancefromhome overtime StockOptionLevel
665         BusinessTravel EnvironmentSatisfaction Department EducationField
666         JobInvolvement JobLevel JobSatisfaction NumCompaniesWorked
667         RelationshipSatisfaction TrainingTimesLastYear WorkLifeBalance
668         YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
669         /DISTRIBUTION=exponential;
670 RUN;
671
672 *Log ratio for exponential for Experienced Employees;
673
674 DATA project1.calculateLogRatio_exponen_exp;
675     L_null=-388.25;
676     L_full=-280.83;
677     L=2 * ABS(L_full - L_null);
678     p_value=1 - probchi(L, 16);
679 RUN;
680
681 PROC PRINT data=project1.calculateLogRatio_exponen_exp;
682 RUN;
683
684 **End of Exponential Model;
685 **Comparing models that are created;
686
687 DATA Project1.CompareModels_exp;
688     L_exponential=-280.83;
689     L_weibull=-145.9409387;
690     L_lognormal=-156.37;
691     LRTEW=-2*(L_exponential - L_weibull);
692     LRTLW=-2*(L_lognormal - L_weibull);
693     LRTLE=-2*(L_lognormal - L_exponential);
694     p_valueEW=1 - probchi(LRTEW, 1);
695     p_valueLW=1 - probchi(LRTLW, 1);
696     p_valueLE=1 - probchi(LRTLE, 2);
697 RUN;
698
699 PROC PRINT DATA=Project1.CompareModels_exp;
700 RUN;
```

References

Arda Zuber, (2017). *OPIM-5894 - Survival Analysis with SAS* Course Materials. UConn School of Business.

Delwiche, L. D., & Slaughter, S. J. (2012). *The Little SAS Book: A Primer: A Primer*. SAS Institute.

Allison, P. D. (2010). *Survival analysis using SAS: a practical guide.* Sas Institute.