MKTG 5983 - Database Marketing - Final Project

Group-14

Dataset- Human Resources Analytics

Why are our best and most experienced employees leaving prematurely?

Overview

Scope of the project:

Human resources data which contains the employee's features in his stint at the organization. The behavioural aspects of an employee are analysed against organizational features to make informed decisions. There might be several aspects that influence employee's productivity at the company like high job satisfaction, good pay etc. Inversely there might be factors that drive him to leave the company. A good analysis is necessary to understand what factors are impacting the most. In our dataset we deal with around 15000 employees in company and 10 features.

Primary focus is on the following questions:

1. Why do employees leave?
2. What are the factors that keep employees satisfied in the job?
3. Do employees who have good last evaluation end up with low salary or no promotion?
4. Whom do we need to retain?

Business Question-1 Why Do Employees Leave?
Method- Logistic regression analysis

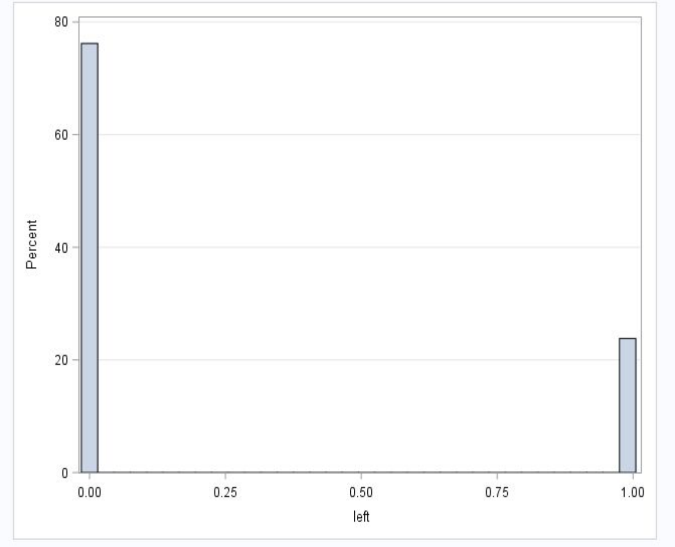
Variables used in dataset-

- Independent Continuous Variables- Satisfaction_level, Last_evaluation, Number_project, Average_montly_hours and Time_spend_company.
- Independent Categorical Variables- Work_accident, promotion_last_5_years, salary, sales.
- Dependent Variables- Left.

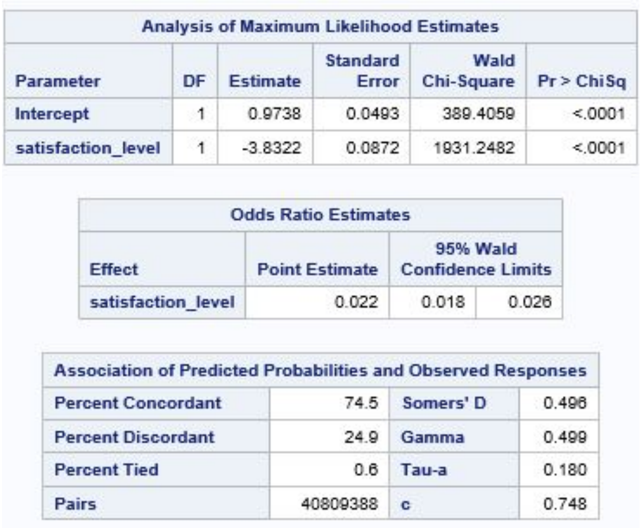
The company had a turnover(left) rate of about 24%

Mean satisfaction of employees is 0.61

Let us analyze each variable with the dependent variable left to determine the factors for an employee to leave the company.



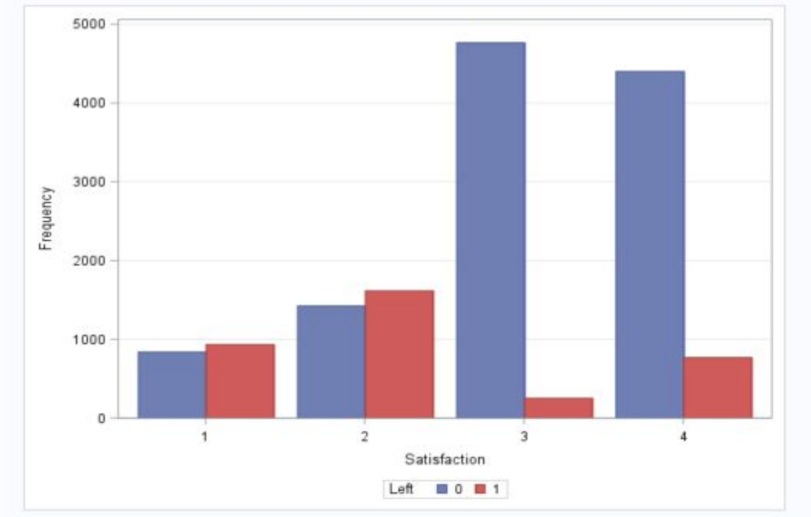
Logistic regression for Satisfaction_level Vs Left gives the following output:



Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.9738	0.0493	389.4059	<.0001
satisfaction_level	1	-3.8322	0.0872	1931.2482	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
satisfaction_level	0.022	0.018	0.026

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	74.5	Somers' D	0.496
Percent Discordant	24.9	Gamma	0.499
Percent Tied	0.6	Tau-a	0.180
Pairs	40809388	c	0.748



We observe that the satisfaction level is a significant predictor of left.

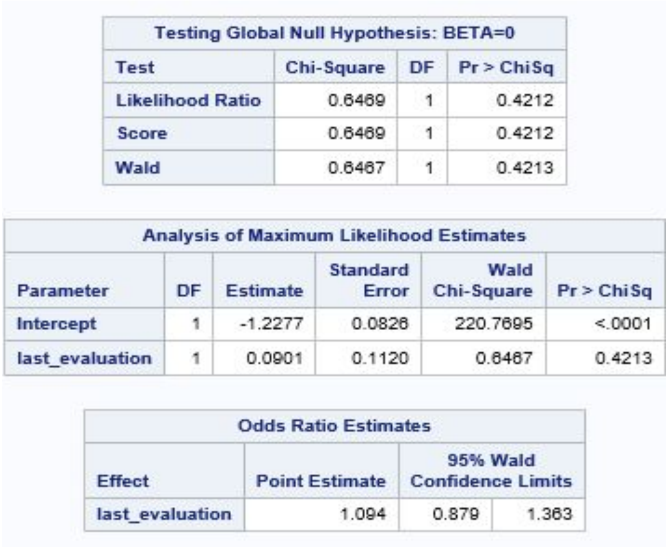
Also, we observe that the correlation between satisfaction level and left is strong as the value –3.8322 signifies that decrease in the satisfaction level increases the probability of employee leaving the company and vice versa.

We transformed the variable as:

- Satisfaction_level from (0 to 0.25) as category1
- Satisfaction_level from (0.26 to 0.50) as category2
- Satisfaction_level from (0.51 to 0.75) as category3
- Satisfaction_level from (0.76 to 1) as category4

From the above graph, we see that employees in category 1 & 2 likely left the company more and the employees in category 3 & 4 did not leave the company in comparison.

Logistic regression for Last_evalutation Vs Left gives the following output:



Testing Global Null Hypothesis: BETA=0

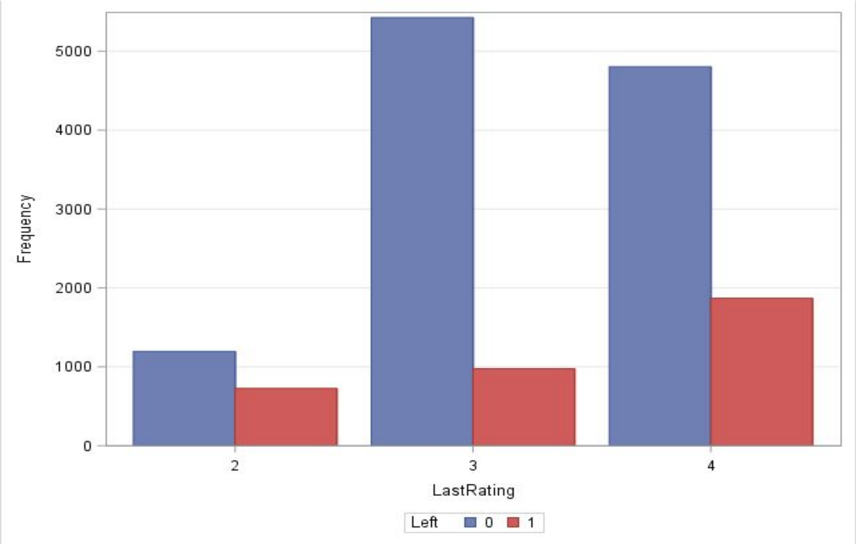
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.6469	1	0.4212
Score	0.6469	1	0.4212
Wald	0.6467	1	0.4213

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.2277	0.0826	220.7695	<.0001
last_evaluation	1	0.0901	0.1120	0.6467	0.4213

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
last_evaluation	1.094	0.879	1.363



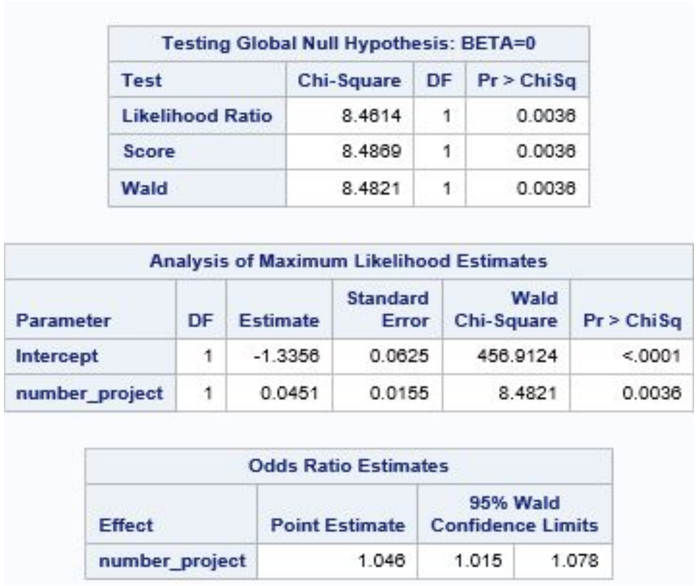
We transformed the variable as:

- Last_Evaluation from (0 to 0.25) as category1
- Last_Evaluation from (0.26 to 0.50) as category2
- Last_Evaluation from (0.51 to 0.75) as category3
- Last_Evaluation from (0.76 to 1) as category4

From the above graph, we observe that the distribution is bimodal.

Employees with Category 2(low) & 4(high) more likely left the company and employees with Category 3(medium) rarely left the company.

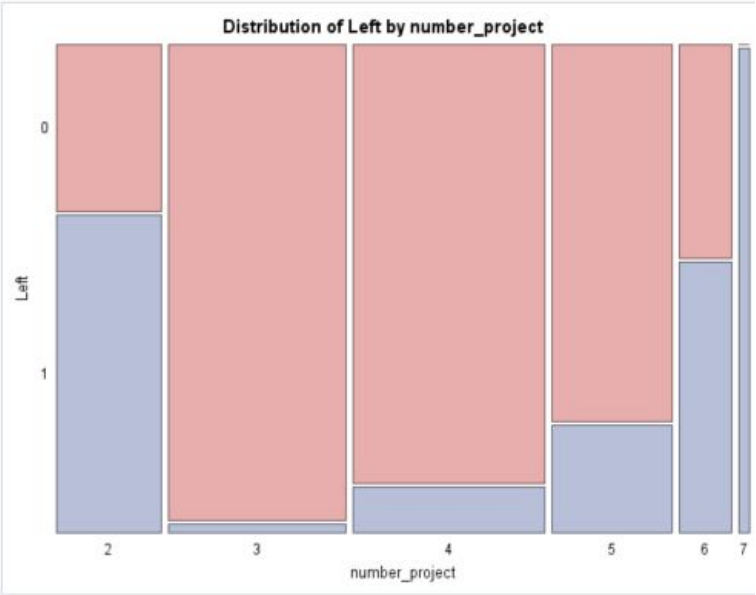
Logistic regression for Number_project Left gives the following output:



Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.4614	1	0.0036
Score	8.4869	1	0.0036
Wald	8.4821	1	0.0036

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.3356	0.0625	456.9124	<.0001
number_project	1	0.0451	0.0155	8.4821	0.0036

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
number_project	1.046	1.015	1.078

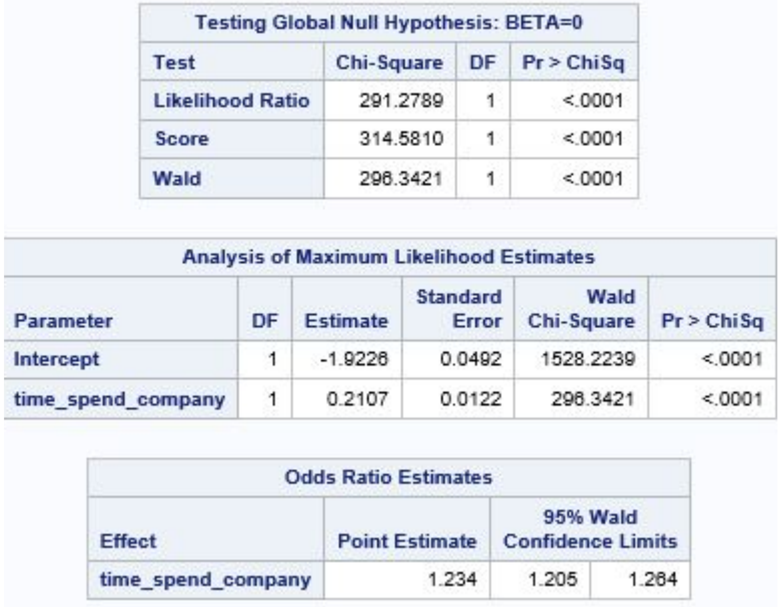


We observe from the above table that Numer_project is a significant predictor of left.

We also observe that the correlation between project number and left is 0.0451. Hence, there is not much correlation between the two variables.

Further, looking into the plot above we interpret that the employees having Number_project as 2, 6 and 7 most likely left the company.

Logistic regression for Time_Spend_Company Vs Left gives the following output:



Testing Global Null Hypothesis: BETA=0

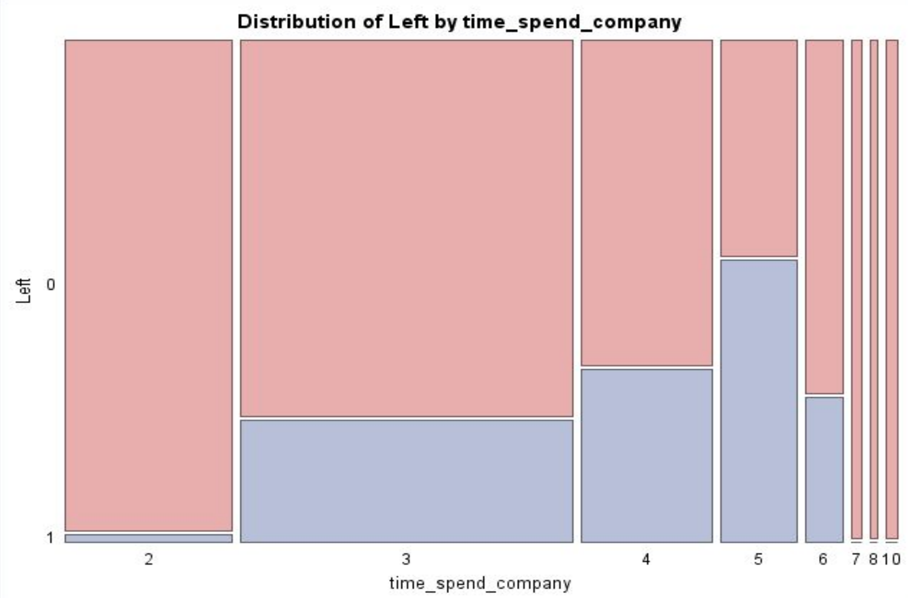
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	291.2789	1	<.0001
Score	314.5810	1	<.0001
Wald	296.3421	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.9226	0.0492	1528.2239	<.0001
time_spend_company	1	0.2107	0.0122	296.3421	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
time_spend_company	1.234	1.205	1.264



We observe Time_Spend_Company is a significant predictor of left since the p-value is less than 0.05.

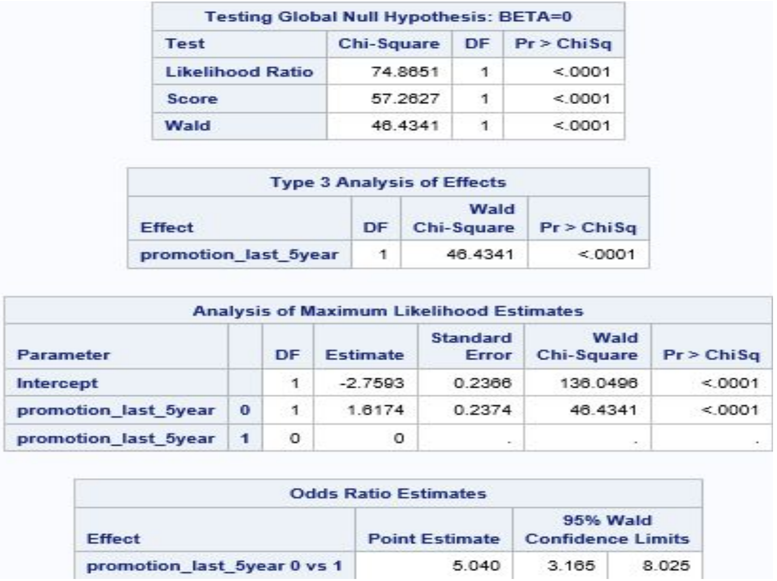
We also see the correlation between Time_Spend_Company and left is 0.2107 which means for every 1 year increase in the time spend at the company, the probability of employee leaving the company increased by 0.2107.

From the above graph, we see that the employees who worked for 5 years left the company more likely.

In contrast, the employees who worked for 6 years or more did not leave the company.

The employees who spent 2 years at the company rarely left.

Logistic regression for Promotion_last_5 years Vs Left gives the following output:

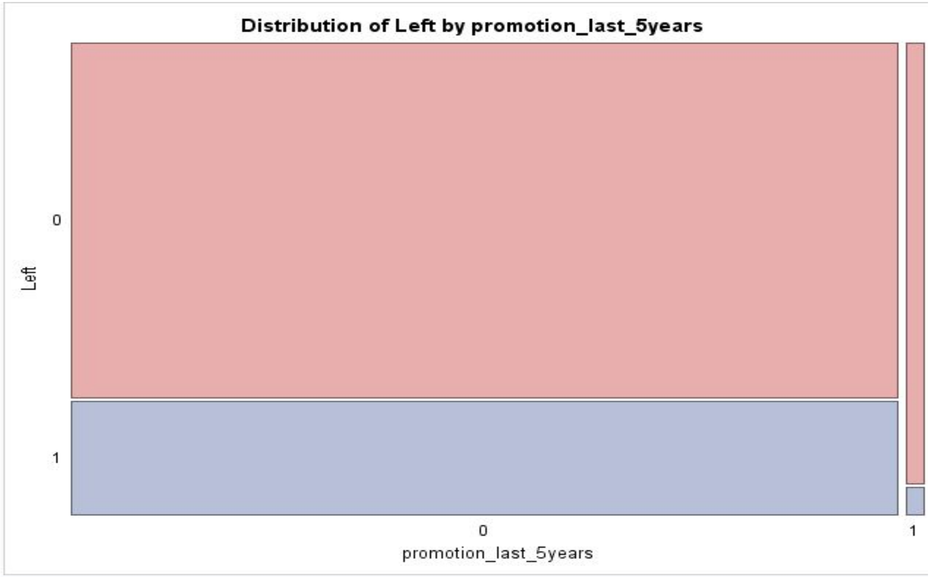


Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	74.8651	1	<.0001
Score	57.2627	1	<.0001
Wald	46.4341	1	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
promotion_last_5year	1	46.4341	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.7593	0.2366	136.0496	<.0001
promotion_last_5year	0	1	1.6174	0.2374	46.4341	<.0001
promotion_last_5year	1	0	0	.	.	.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
promotion_last_5year 0 vs 1	5.040	3.165	8.025

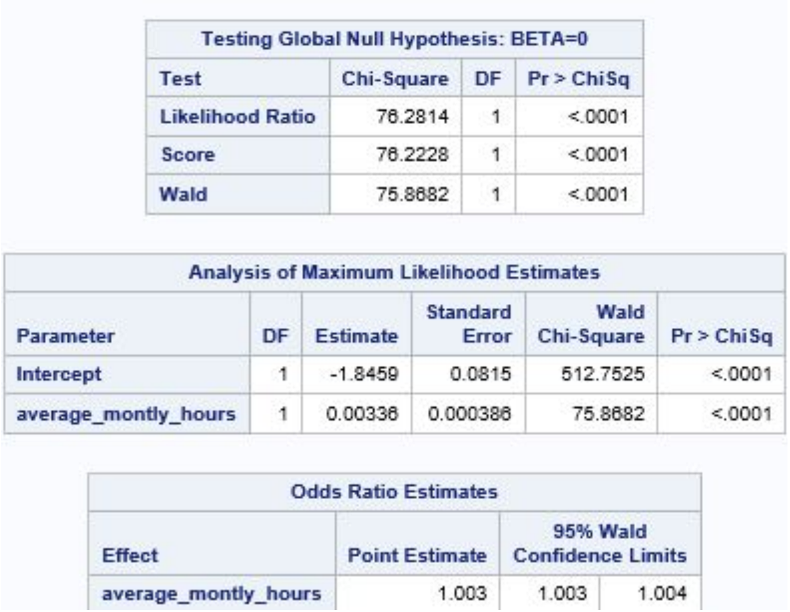


We can see from the above table that 'promotion_last_5years' is a significant predictor of left .

We also observe that the correlation between 'promotion_last_5years' and left is strong with value -1.6174 which signifies that for each promotion an employee gets, the probability of him leaving the company decreases by -1.6174.

From the above graph we observe that the employees who are promoted in last 5 years are more likely to stay in the company than the employees who are not promoted.

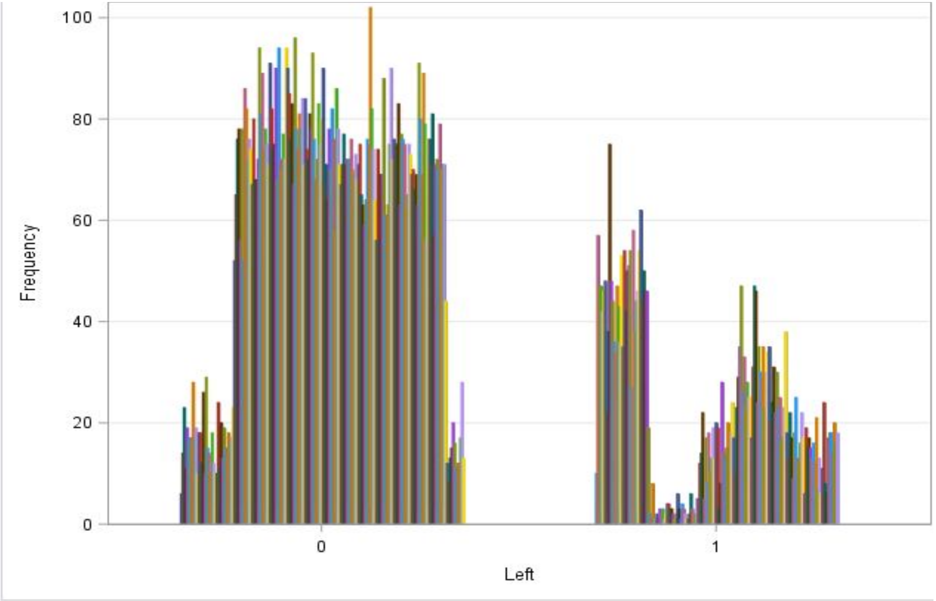
Logistic regression for Average_montly_hours Vs Left gives the following output:



Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	76.2814	1	<.0001
Score	76.2228	1	<.0001
Wald	75.8682	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.8459	0.0815	512.7525	<.0001
average_montly_hours	1	0.00336	0.000386	75.8682	<.0001

Odds Ratio Estimates			
		95% Wald	
Effect	Point Estimate	Confidence Limits	
average_montly_hours	1.003	1.003	1.004



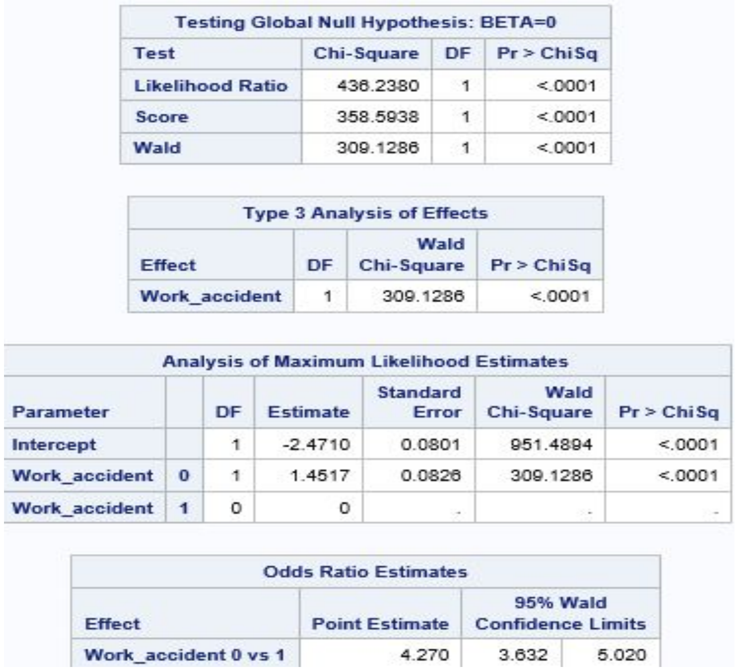
We observe from the above table that the 'Average_monthly_hours' is a significant predictor of left.

We also observe that the correlation between 'Average_monthly_hours' and left is 0.00336 i.e., for each unit increase in 'Average_monthly_hours', increase 0.00336 units of the probability of employee leaving the company.

From the above graph, it is clearly seen that the employees who spent more numbers of monthly hours more likely left the company.

In addition, we can also see that a portion of employees with low 'average_monthly_hours' also likely left the company.

Logistic regression for Work_accident Vs Left gives the following output:

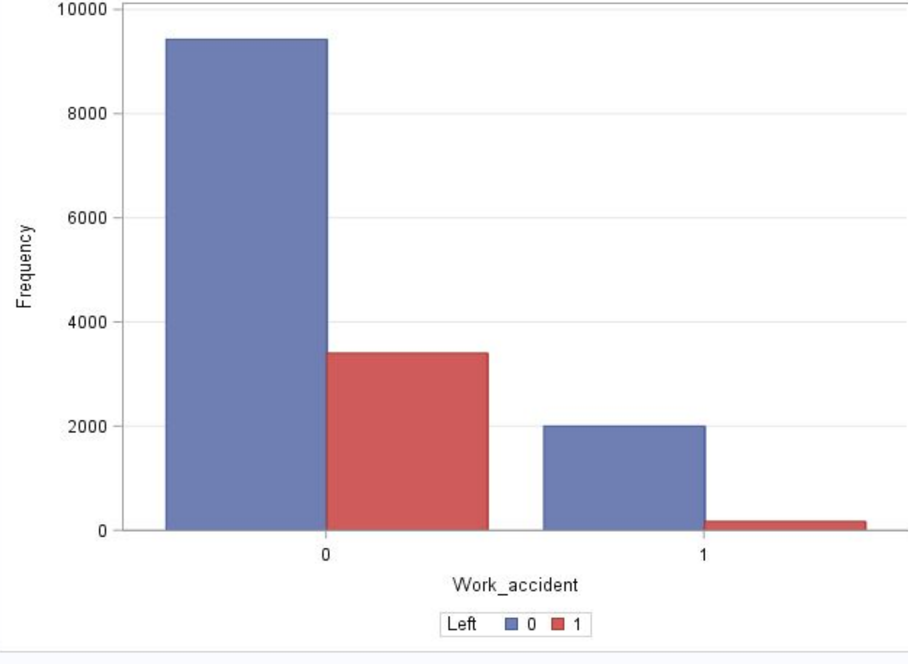


Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	436.2380	1	<.0001
Score	358.5938	1	<.0001
Wald	309.1286	1	<.0001

Type 3 Analysis of Effects			
		Wald	
Effect	DF	Chi-Square	Pr > ChiSq
Work_accident	1	309.1286	<.0001

Analysis of Maximum Likelihood Estimates						
				Standard	Wald	
Parameter		DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept		1	-2.4710	0.0801	951.4894	<.0001
Work_accident	0	1	1.4517	0.0826	309.1286	<.0001
Work_accident	1	0	0	.	.	.

Odds Ratio Estimates			
		95% Wald	
Effect	Point Estimate	Confidence Limits	
Work_accident 0 vs 1	4.270	3.632	5.020



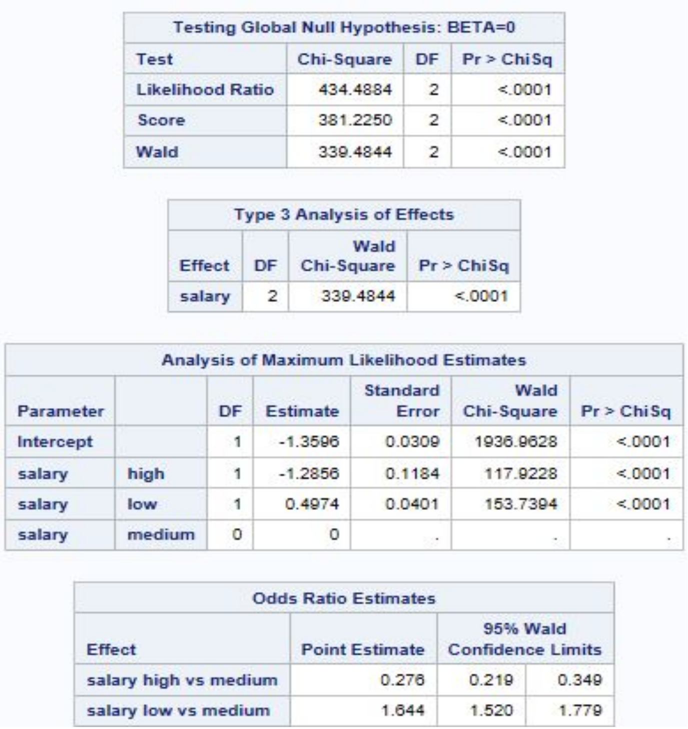
We can see from the above table that the 'Work_accident' is a significant predictor of left.

We also see that the correlation between Work accident and left is 1.4517.

The above graph says that the employees who did not leave the company are the employees who encountered the work_accident.

However, this is in contrast to the real scenario.

Logistic regression for Salary Vs Left gives the following output:

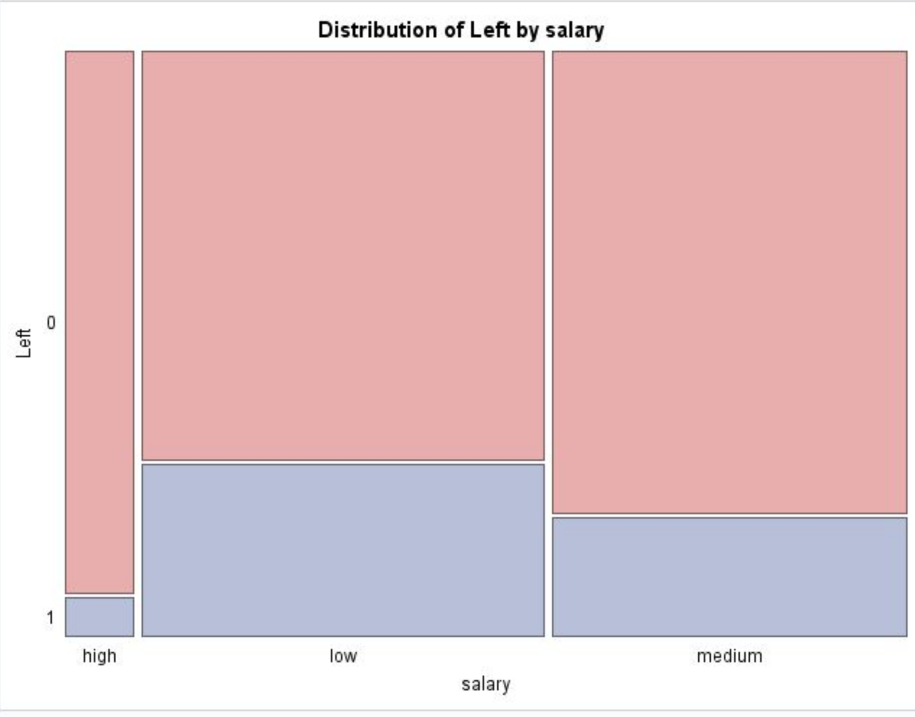


Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	434.4884	2	<.0001
Score	381.2250	2	<.0001
Wald	339.4844	2	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
salary	2	339.4844	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.3596	0.0309	1936.9628	<.0001
salary	high	1	-1.2856	0.1184	117.9228	<.0001
salary	low	1	0.4974	0.0401	153.7394	<.0001
salary	medium	0	0	.	.	.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
salary high vs medium	0.276	0.219	0.349
salary low vs medium	1.644	1.520	1.779



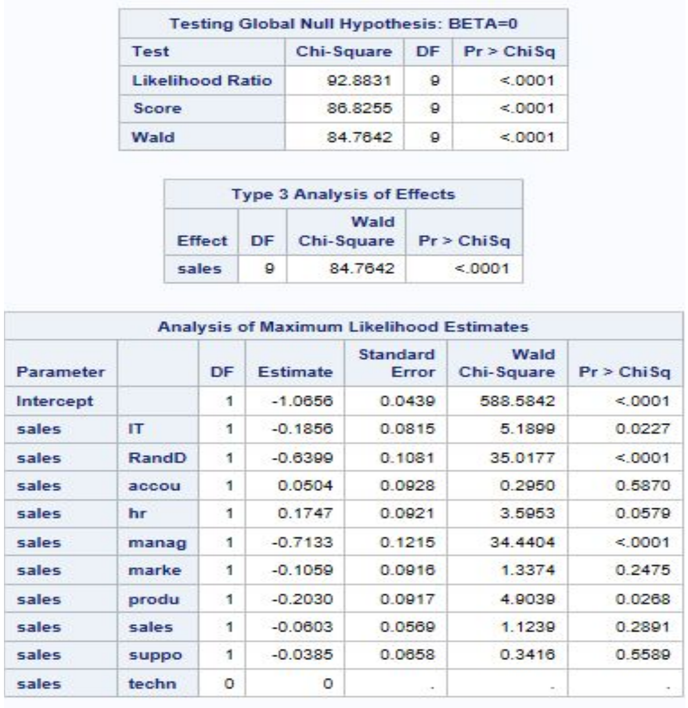
From the above table, we see that the Salary is a significant predictor of left.

Also we can see the correlation between Salary(high) and left is -1.2856 when you are controlling for Salary(low).

From the above graph, we can interpret that employees with low and medium salaries likely left the company.

And the employees with high salaries rarely left the company.

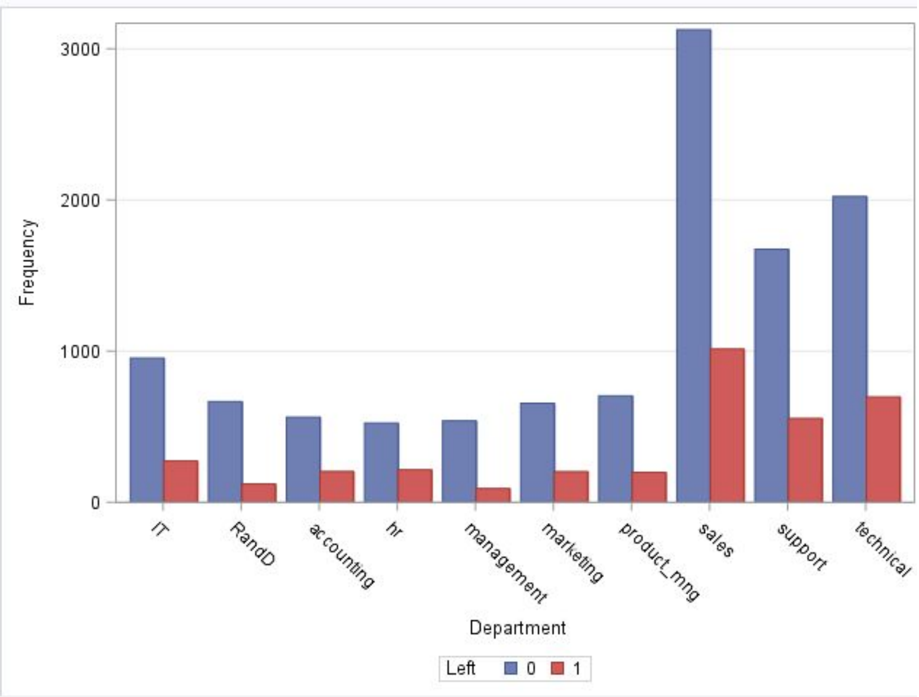
Logistic regression for Department Vs Left gives the following output:



Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	92.8831	9	<.0001
Score	86.8255	9	<.0001
Wald	84.7642	9	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sales	9	84.7642	<.0001

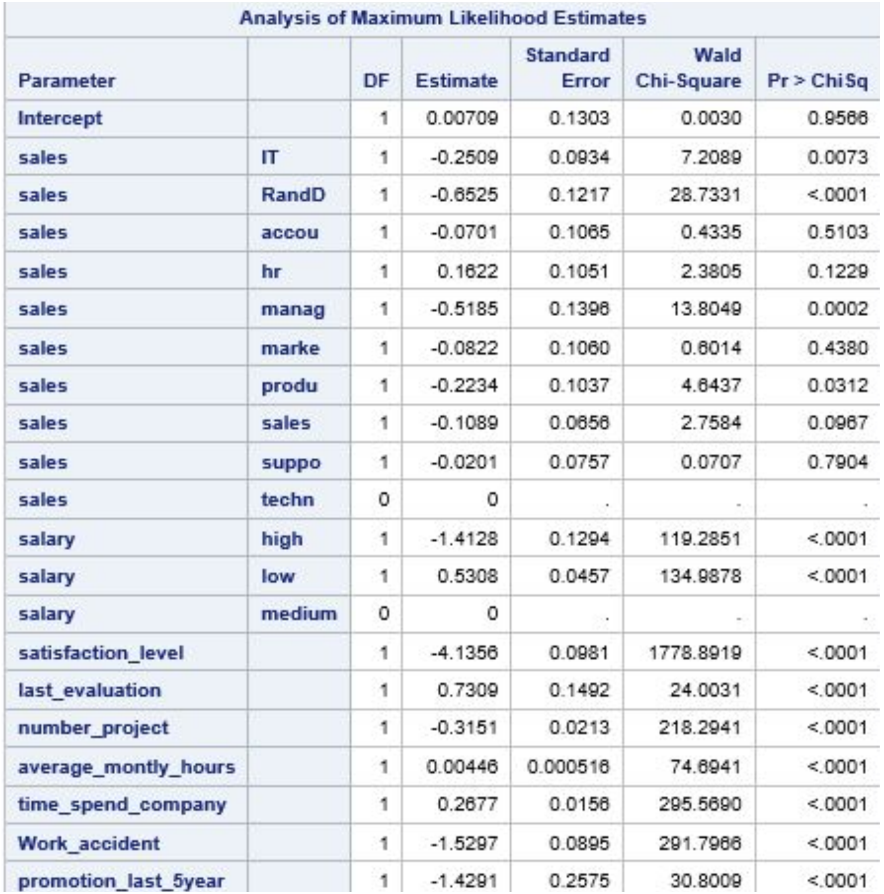
Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.0656	0.0439	588.5842	<.0001
sales	IT	1	-0.1856	0.0815	5.1899	0.0227
sales	RandD	1	-0.6399	0.1081	35.0177	<.0001
sales	accou	1	0.0504	0.0928	0.2950	0.5870
sales	hr	1	0.1747	0.0921	3.5953	0.0579
sales	manag	1	-0.7133	0.1215	34.4404	<.0001
sales	marke	1	-0.1059	0.0916	1.3374	0.2475
sales	produ	1	-0.2030	0.0917	4.9039	0.0268
sales	sales	1	-0.0603	0.0569	1.1239	0.2891
sales	suppo	1	-0.0385	0.0658	0.3416	0.5589
sales	techn	0	0	.	.	.



From the above table, we observe that overall model is significant for department.

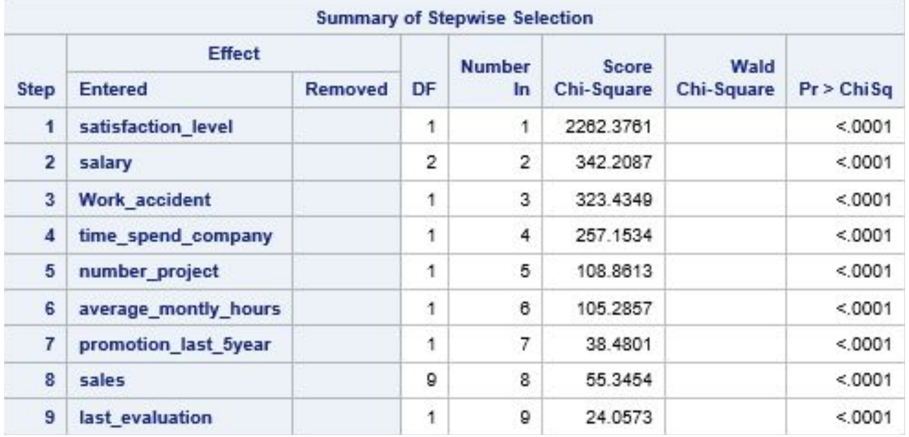
We can clearly see from the graph that the sales, support and technical departments have the highest attrition rate. The management department has the lowest attrition rate.

Stepwise logistic regression is used to check the significant predictors of the model



Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.00709	0.1303	0.0030	0.9566
sales	IT	1	-0.2509	0.0934	7.2089	0.0073
sales	RandD	1	-0.6525	0.1217	28.7331	<.0001
sales	accou	1	-0.0701	0.1065	0.4335	0.5103
sales	hr	1	0.1622	0.1051	2.3805	0.1229
sales	manag	1	-0.5185	0.1396	13.8049	0.0002
sales	marke	1	-0.0822	0.1060	0.6014	0.4380
sales	produ	1	-0.2234	0.1037	4.6437	0.0312
sales	sales	1	-0.1089	0.0656	2.7584	0.0967
sales	suppo	1	-0.0201	0.0757	0.0707	0.7904
sales	techn	0	0	.	.	.
salary	high	1	-1.4128	0.1294	119.2851	<.0001
salary	low	1	0.5308	0.0457	134.9878	<.0001
salary	medium	0	0	.	.	.
satisfaction_level		1	-4.1356	0.0981	1778.8919	<.0001
last_evaluation		1	0.7309	0.1492	24.0031	<.0001
number_project		1	-0.3151	0.0213	218.2941	<.0001
average_montly_hours		1	0.00446	0.000516	74.6941	<.0001
time_spend_company		1	0.2677	0.0156	295.5690	<.0001
Work_accident		1	-1.5297	0.0895	291.7966	<.0001
promotion_last_5year		1	-1.4291	0.2575	30.8009	<.0001



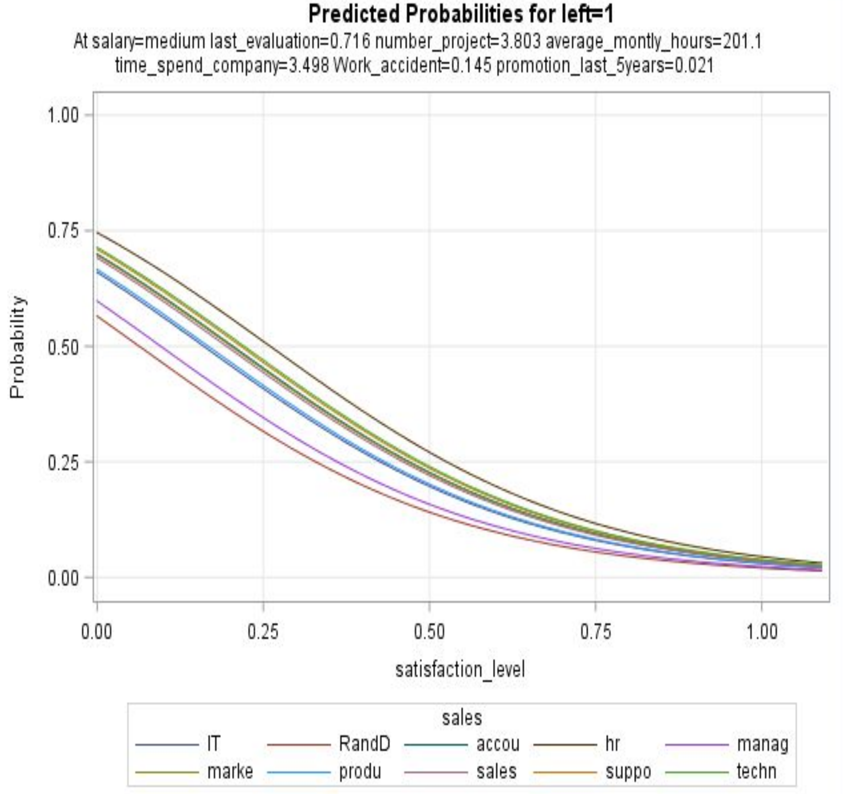
Summary of Stepwise Selection

Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	satisfaction_level		1	1	2262.3761		<.0001
2	salary		2	2	342.2087		<.0001
3	Work_accident		1	3	323.4349		<.0001
4	time_spend_company		1	4	257.1534		<.0001
5	number_project		1	5	108.8613		<.0001
6	average_montly_hours		1	6	105.2857		<.0001
7	promotion_last_5year		1	7	38.4801		<.0001
8	sales		9	8	55.3454		<.0001
9	last_evaluation		1	9	24.0573		<.0001

The above Summary table and likelihood estimates mentiones all the significant predictors which fit the model.

Summary

- Barely any employees left with **high** salary
- Employees with low to average salaries tend to leave the company.
- More than half of the employees with **2,6, and 7** projects left the company
- All of the employees with **7** projects left the company
- There is an increase in employee turnover rate as project count increases
- Employees with **low** performance tend to leave the company more
- Employees with **high** performance tend to leave the company more
- The employees that stayed is within **0.6-0.8** evaluation
- Employees who had less hours of work **(~150 hours or less)** left the company more
- Employees who had too many hours of work **(~250 or more)** left the company
- Employees who had really low satisfaction levels **(0.2 or less)** left the company more
- Employees who had low satisfaction levels **(0.3~0.5)** left the company more
- Employees who had really high satisfaction levels **(0.7 or more)** left the company more.



Business Question-2 What are the factors that keep employees satisfied in the job?
Method - Multiple regression analysis

The Multiple regression for Satisfaction_level as dependent variable and all the other independent variables is performed using the stepwise method.

Stepwise method gives all significant factors and determines the satisfaction level of the employees.

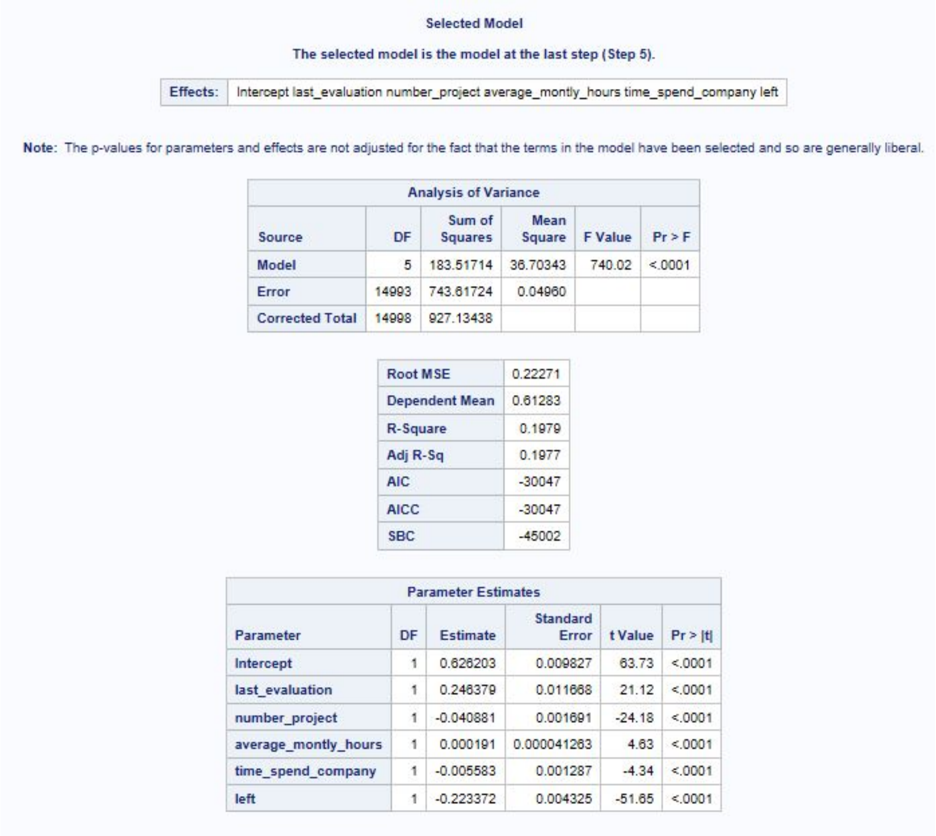
We clearly see the factors that determine the satisfaction level of the employee are the variables-

- Last_Evaluation
- Number_Project
- Average_monthly_hours
- Time_spend_company
- Left.

The below model explains 19.79% of the variability.

From the ANOVA model, we can conclude the overall model is significant as p<0.0001

We analyse each of the variables individually on how they determine the satisfaction level of the employee without taking into account the controlling of other variables.



Selected Model

The selected model is the model at the last step (Step 5).

Effects:	Intercept last_evaluation number_project average_montly_hours time_spend_company left

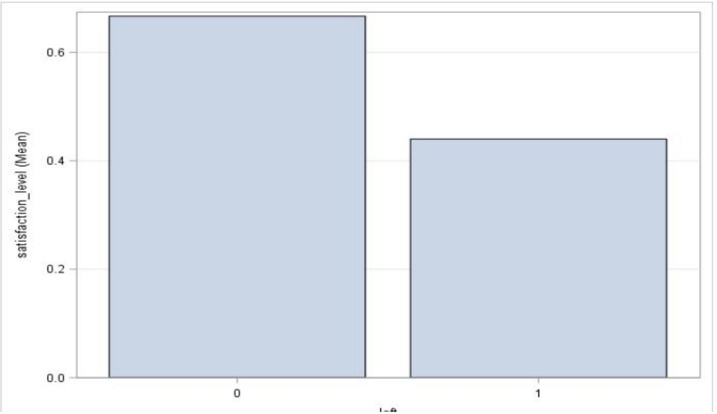
Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	183.51714	36.70343	740.02	<.0001
Error	14993	743.61724	0.04960		
Corrected Total	14998	927.13438			

Root MSE	0.22271
Dependent Mean	0.61283
R-Square	0.1979
Adj R-Sq	0.1977
AIC	-30047
AICC	-30047
SBC	-45002

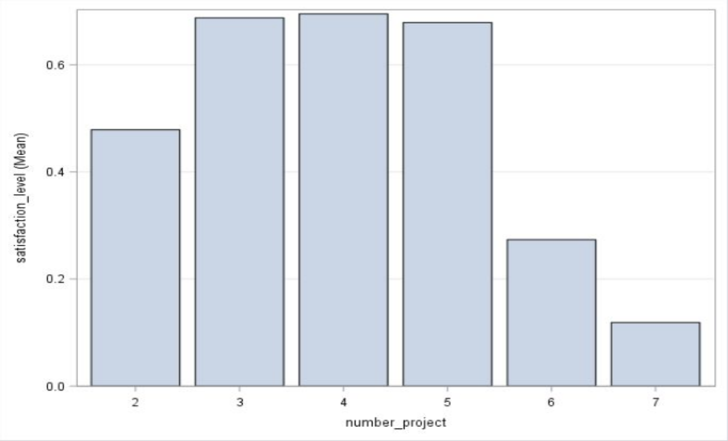
Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.626203	0.009827	63.73	<.0001
last_evaluation	1	0.246379	0.011668	21.12	<.0001
number_project	1	-0.040881	0.001691	-24.18	<.0001
average_montly_hours	1	0.000191	0.000041263	4.63	<.0001
time_spend_company	1	-0.005583	0.001287	-4.34	<.0001
left	1	-0.223372	0.004325	-51.65	<.0001

Bar chart of Satisfaction_level(Mean) vs Left & Satisfaction_level(Mean) vs Number_project:



From the graph, we observe that all the employees who haven't left the company have a mean satisfaction level of 0.67.

All employees who left the company have a mean satisfaction level of 0.44 which makes logical sense as only the people who are least satisfied tend to leave the company.

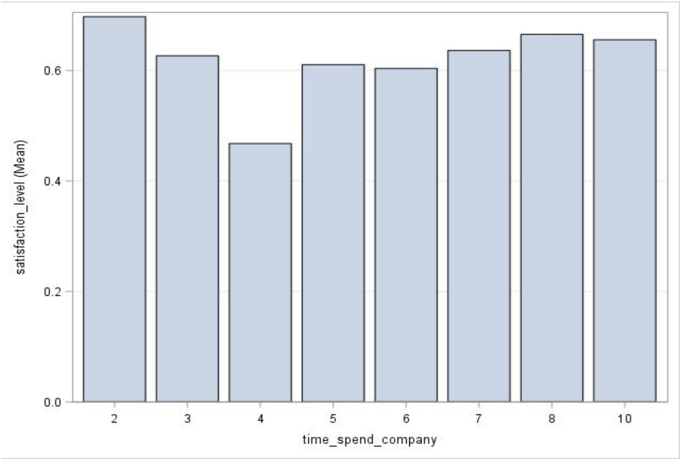


From the above graph, employees having 4 projects have the highest mean satisfaction level.

The most satisfying employees in the organization have projects between 3 to 5.

Those Employees who have 7 projects have the least mean satisfaction levels which makes sense.

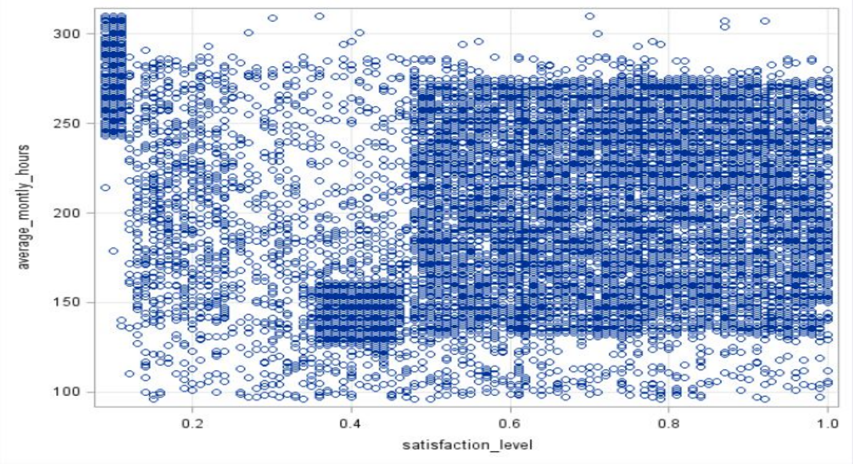
Bar chart of Satisfaction_level(Mean) Vs Time_spent_company & Scatterplot of Satisfaction_level vs Average_monthly_hours:



From the above graph, we see that Satisfaction_level was high for the employees who are just about 2 to 3 years old in the company.

Mean satisfaction_level was the lowest for the employees who have worked for 4 years.

The mean satisfaction level gradually increases as the employee spends longer than 4 years with the company.



The above graph points out that the least satisfied employees are the ones who work more than 240 monthly hours on an average.

The employees who work on an average of 140 to 275 monthly hours have good satisfaction levels.

A fair bunch of employees have an average satisfaction level of 0.4 who work on an average of 130 to 160 hours indicating that they would be more satisfied if they are assigned more work than the usual less amount of work.

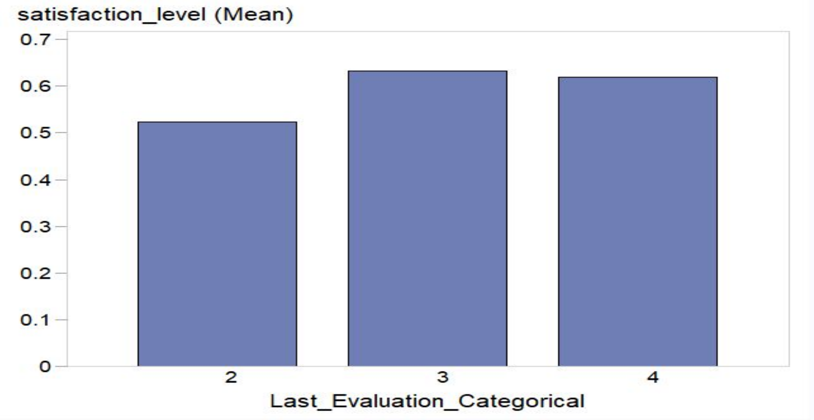
Bar chart of Satisfaction level(Mean) vs Last_Evaluation:

We interpret the graph as:

- The employees whose satisfaction level was low were the ones who couldn't perform well as their Last_Evaluation falls in category 2.

- The employees whose satisfaction level was high were the ones who were appreciated well for their performance as their Last_Evaluation falls in category 3.

- The employees whose satisfaction level was good were the ones whose last_evaluation falls in the category 4. Probably the employees are not getting enough appreciation for their work and hence their satisfaction level was not as high as the employees whose Last_Evaluation falls in category 3.



Firstly, The Last_Evaluation continuous variable has been converted to a categorical variable for easy analysis.

We transformed the variable as:

- Last_Evaluation from (0 to 0.25) as category1
- Last_Evaluation from (0.26 to 0.50) as category2
- Last_Evaluation from (0.51 to 0.75) as category3
- Last_Evaluation from (0.76 to 1) as category4

Summary

- Factors that determine the satisfaction level of the employee are the variables - Last_Evaluation, Number_Project, Average_monthly_hours, time_spend_company, Left.

- Employees who haven’t left the firm have a better satisfaction levels compared to the employees who have left.

- Employees who have the average number of projects between 3 and 5 have the most satisfaction levels.

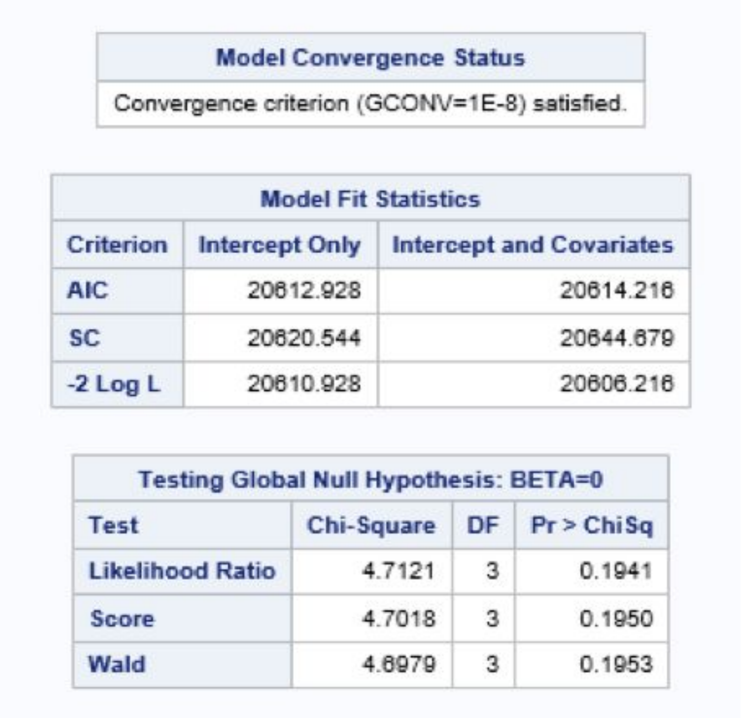
- Employees who work on an average of 140 to 275 monthly hours have the best satisfaction levels.

- Employees who have Last_Evaluation as 3 (good) or 4 (very good) are happy and have good satisfaction levels.

- Employees who spent time in the company between 2 to 3 years and over 5 years are well satisfied.

Question 3 - Do employees who have good last evaluation end up with low salary or no promotion?
Method - Logistic Regression

- A narrowed analysis
- Dependent variable - Employees with good last evaluation.
- Dependent variable - Transformed to categorical variable with good last evaluation.
- New dependent variable - 'Good Evaluation'
- Predictor 'salary' is transformed to 'LowSalary'
- Logistic regression performed-
  - 'Good Evaluation' - dependent variable
  - 'LowSalary' - Predictor 1
  - 'promotion_last_5years' - Predictor 2
- Interaction term 'LowSalary*promotion_last_5years' is also considered.

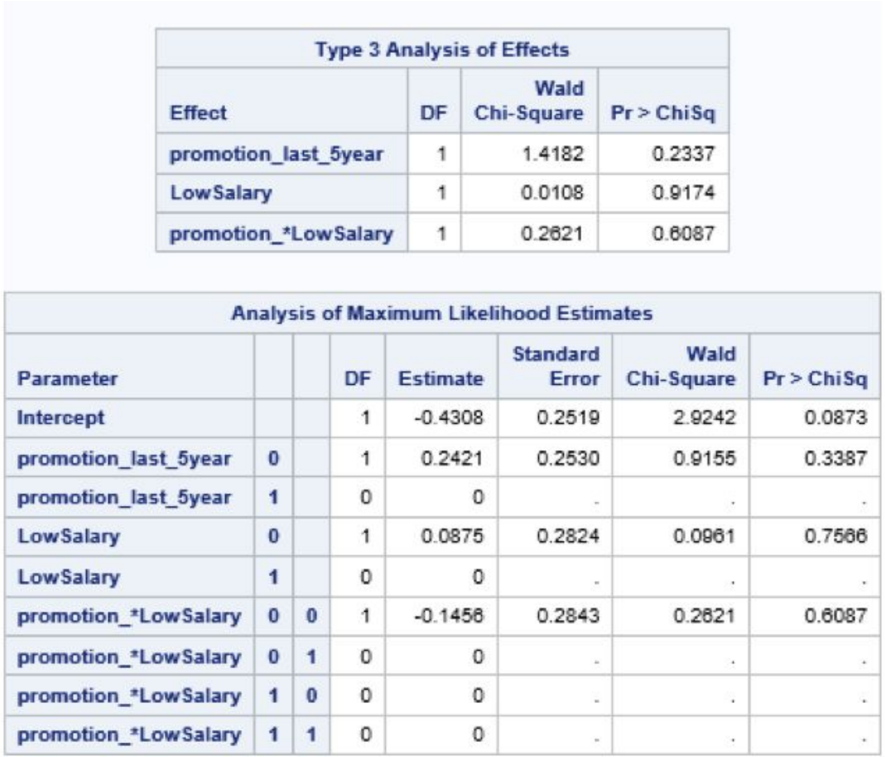


Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	20612.928	20614.216
SC	20620.544	20644.679
-2 Log L	20610.928	20606.216

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.7121	3	0.1941
Score	4.7018	3	0.1950
Wald	4.6979	3	0.1953

Effect of predictors and their interaction on the dependent variable:



Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
promotion_last_5year	1	1.4182	0.2337
LowSalary	1	0.0108	0.9174
promotion_*LowSalary	1	0.2621	0.6087

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-0.4308	0.2519	2.9242	0.0873
promotion_last_5year	0		1	0.2421	0.2530	0.9155	0.3387
promotion_last_5year	1		0	0	.	.	.
LowSalary	0		1	0.0875	0.2824	0.0961	0.7566
LowSalary	1		0	0	.	.	.
promotion_*LowSalary	0	0	1	-0.1456	0.2843	0.2621	0.6087
promotion_*LowSalary	0	1	0	0	.	.	.
promotion_*LowSalary	1	0	0	0	.	.	.
promotion_*LowSalary	1	1	0	0	.	.	.

From the tables, we observe:

- Predictor ‘LowSalary’ is not significant.
- Predictor ‘promotion_last_5years’ is not significant either.
- Interaction between the 2 predictors ‘LowSalary*promotion_last_5years’ is not significant since the p-value is greater than 0.05.

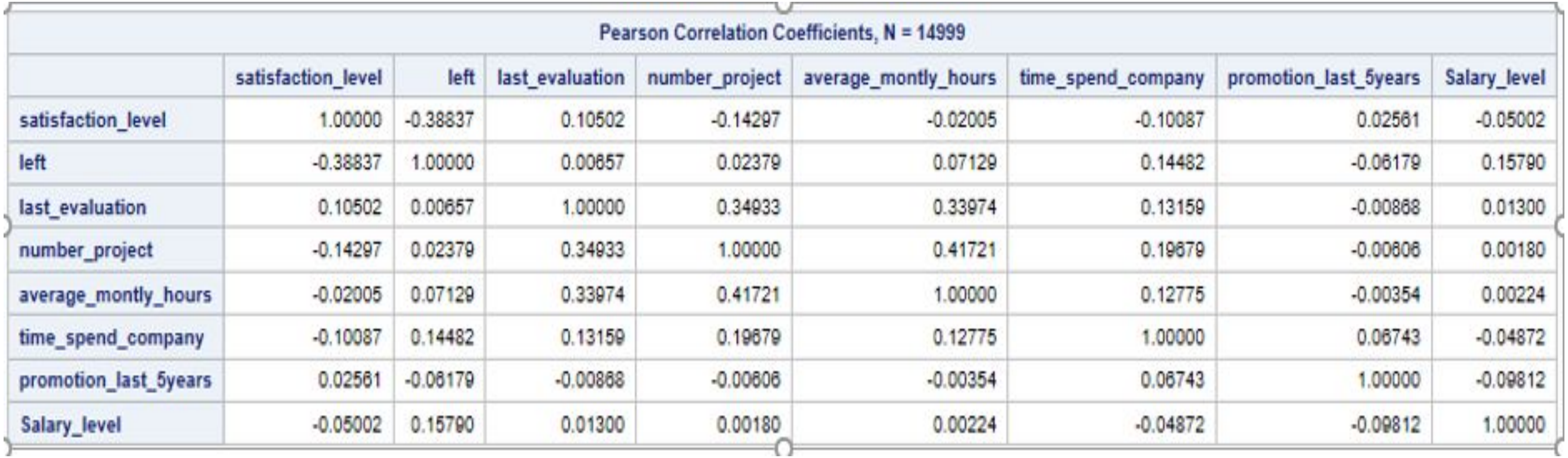
Hence, we cannot conclude that employees having good last evaluation end up having both low salary and no promotion in the last 5 years since both the factors are not significant.

For example, Employees having good last evaluation might have high salary as in the case of managers who might not have good last evaluation but might have high salary.

Business Question-4 Whom do we need to retain?
Method- Correlation Matrix

Variables used for this question:

- Satisfaction_level
- Last_evaluation
- Number_project
- Average_montly_hours
- Time_spend_company
- Salary
- Left
- Promotion_last_5years

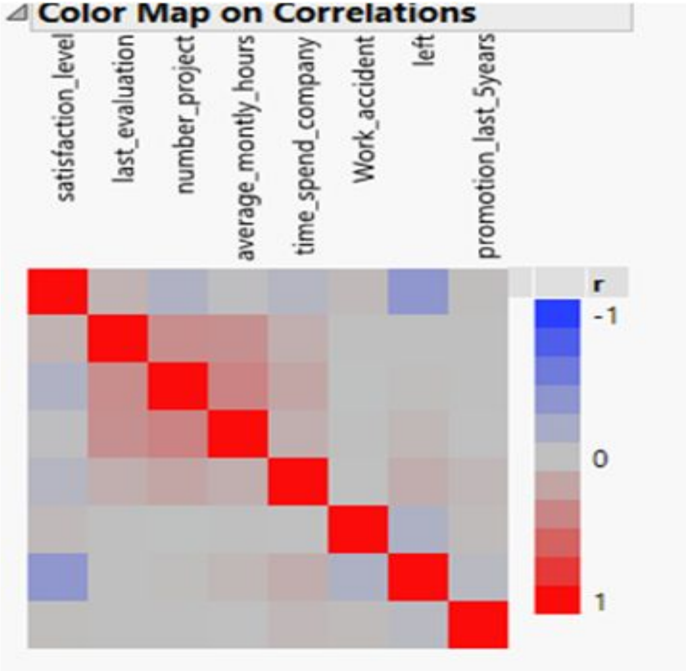


Pearson Correlation Coefficients, N = 14999

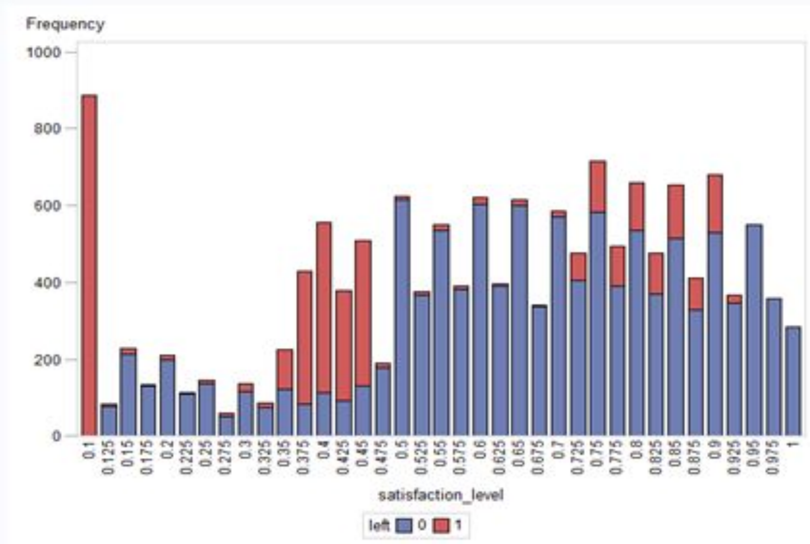
	satisfaction_level	left	last_evaluation	number_project	average_montly_hours	time_spend_company	promotion_last_5years	Salary_level
satisfaction_level	1.00000	-0.38837	0.10502	-0.14297	-0.02005	-0.10087	0.02561	-0.05002
left	-0.38837	1.00000	0.00657	0.02379	0.07129	0.14482	-0.06179	0.15790
last_evaluation	0.10502	0.00657	1.00000	0.34933	0.33974	0.13159	-0.00868	0.01300
number_project	-0.14297	0.02379	0.34933	1.00000	0.41721	0.19679	-0.00606	0.00180
average_montly_hours	-0.02005	0.07129	0.33974	0.41721	1.00000	0.12775	-0.00354	0.00224
time_spend_company	-0.10087	0.14482	0.13159	0.19679	0.12775	1.00000	0.06743	-0.04872
promotion_last_5years	0.02561	-0.06179	-0.00868	-0.00606	-0.00354	0.06743	1.00000	-0.09812
Salary_level	-0.05002	0.15790	0.01300	0.00180	0.00224	-0.04872	-0.09812	1.00000

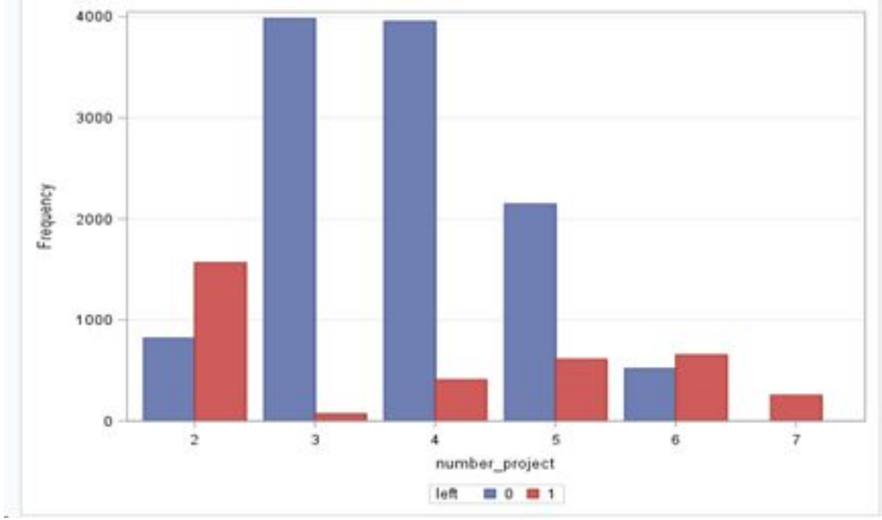
Correlation matrix for all the variables gives the following output:

- From the correlation matrix, left is highly correlated with Satisfaction_level.
- Satisfaction_Level is correlated with number_project and time_spent_company.
- Low satisfaction level is correlated with employees who have worked for greater number of hours and more number of projects.
- Employees with no promotion have left job.
- Employees having low satisfaction left the company as these are negatively correlated.
- To find whom do we need to retain, we must know the reasons why an employee has left.



Bar chart for Satisfaction_level Vs Left & Number_project Vs Left:





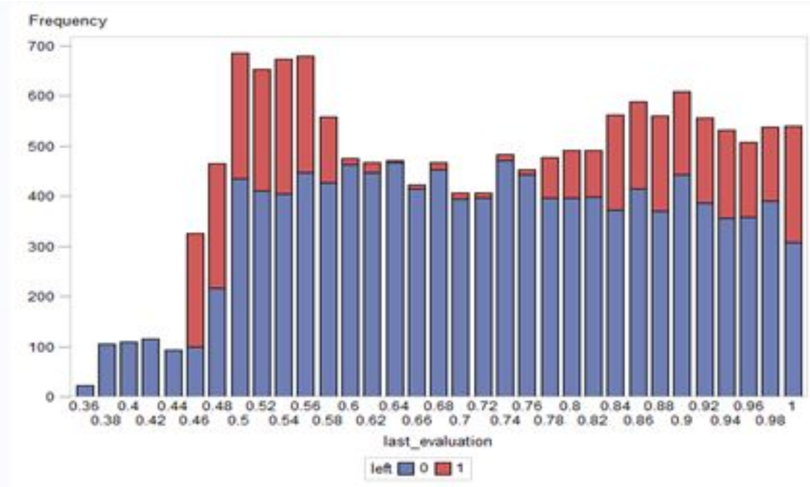
From the above graph, we see that employees with low satisfaction level leave the company. But we can also see that employee with higher satisfaction level too leave the company.

Employees with higher satisfaction level will be easy to retain.

The above graph says that employees who have worked on minimum and maximum number of projects leave the organization.

We should retain employees who are working for greater number of projects as they are assets to the company.

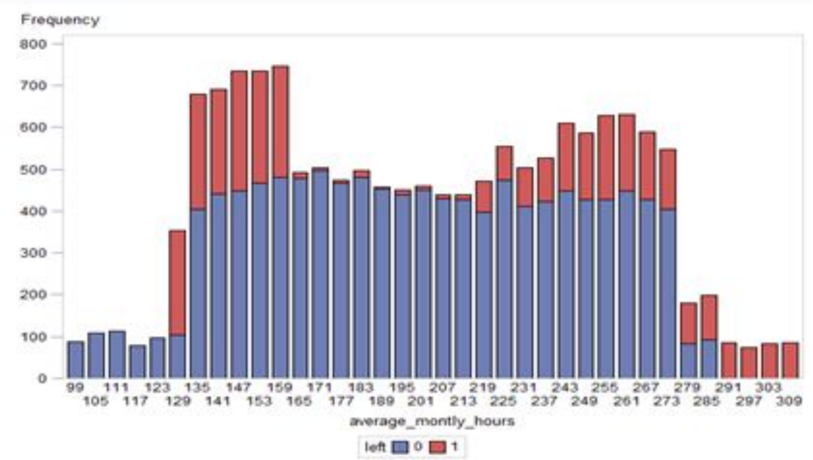
Bar chart for Last_Evaluation Vs Left & Average_monthly_hours Vs Left:



From the above graph, we observe that employees with least and higher last_evaluation leave the company.

Employees with least last_evaluation between 0.36 and 0.44 might be those who have joined company recently.

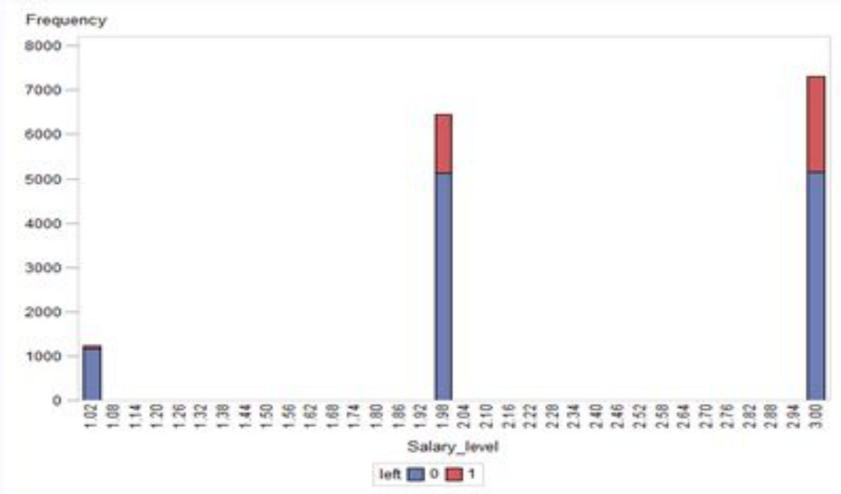
Employees with higher last_evaluation should be retained as they are an asset to the company.



The above graph suggests that employees with highest average_monthly_hours and lowest average_monthly_hours have left organization.

Hence, number of average_monthly_hours for employee on higher end can be reduced to retain the employee

Bar chart for Salary Vs Left:



From the above graph, we observe that employees with low salary level(3) and medium salary level(2) have left organization

Very few employees with high salary do not leave organization.

Salary of employees should be increased if we need to retain them.

Summary

Following are the employees organization needs to retain.

- Employees having higher last_evaluation rate are assets to an organization and should be retained.
- Employees who are working for more number of projects.
- Employees who have average_working_hours value between 159 to 213 hours are most likely to not leave organization.
- Employees who are working for higher than average_working_hours can be retained.
- Employees having higher satisfaction_level are easy to retain as they are satisfied with the organization.

THANK YOU!