

Employee Attrition Factors - Assignment

Group Member:

1. Atul Mehra
2. Ramgopal C R
3. Shafali Sharma
4. Surabhi Gupta



Strategy and Business Objectives

Business Objective

- Identify the driving variables or the factors which are strong indicators of causing attrition in the firm
- With the identification of the factors, changes can be made by employer for employees to stay in the company

Strategy

- Find out the indicators where attrition is more, more likely it caused the attrition and also which type of people with their income, job satisfaction , level, gender, department etc., are more likely to leave the company

Data - 4410 employees data

Data

- 4410 observations of employee data
- Age, income, Education, gender, marital status, experience and last hike of the employee are all been given
- Other metrics like distance from home, business travel, job role, stock option level, years with current manager are also given
- Other metrics which is also been given are survey data from employee and manager and the in time and out time of each employees during different period

Data cleaning

The data has to be cleaned for further analysis

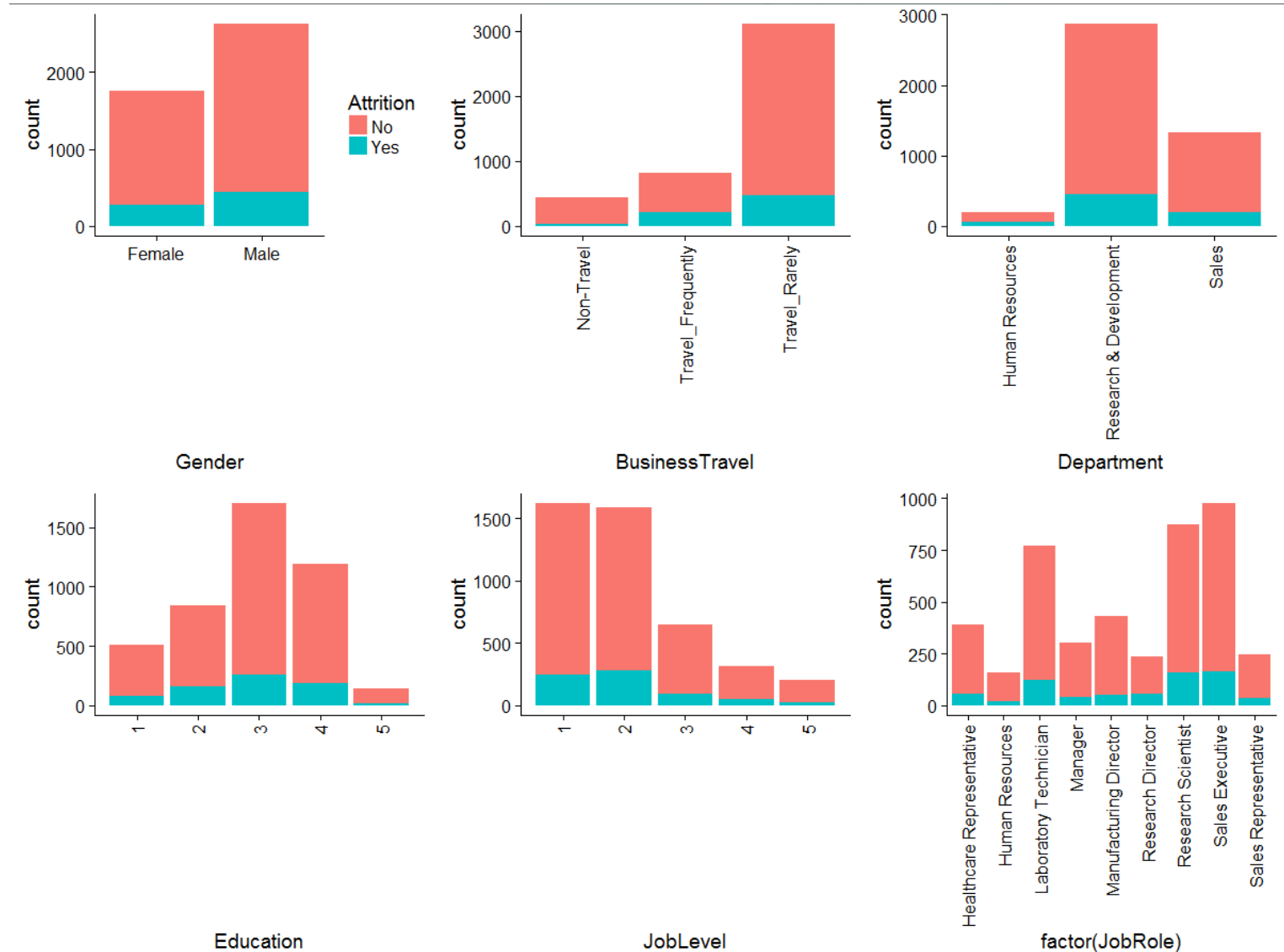
- There are no duplicates which are observed and there are mostly NA's in some metrics, however, they are not a considerable number around 2% and hence removing the same.
- Employee count and standard hours are not really useful for the analysis and hence removed the same
- Income range outliers greater than 165000 are set to the limit of 165K
- Working years greater than 27 is limited to 27
- Years at company also been limited to 17 based on the outlier analysis
- Years since last promotion is restricted to 9 and years with current manager to 13

Data preparation

- Derive the additional data variables like average working hours based on the in time and out time
- Convert all the categorical as factor variables like education, job role, etc., to dummy variables to make it all numeric
- Scale all the continuous variables like income, age, distance from home, stock option level, work life balance etc., to have uniformity in the range

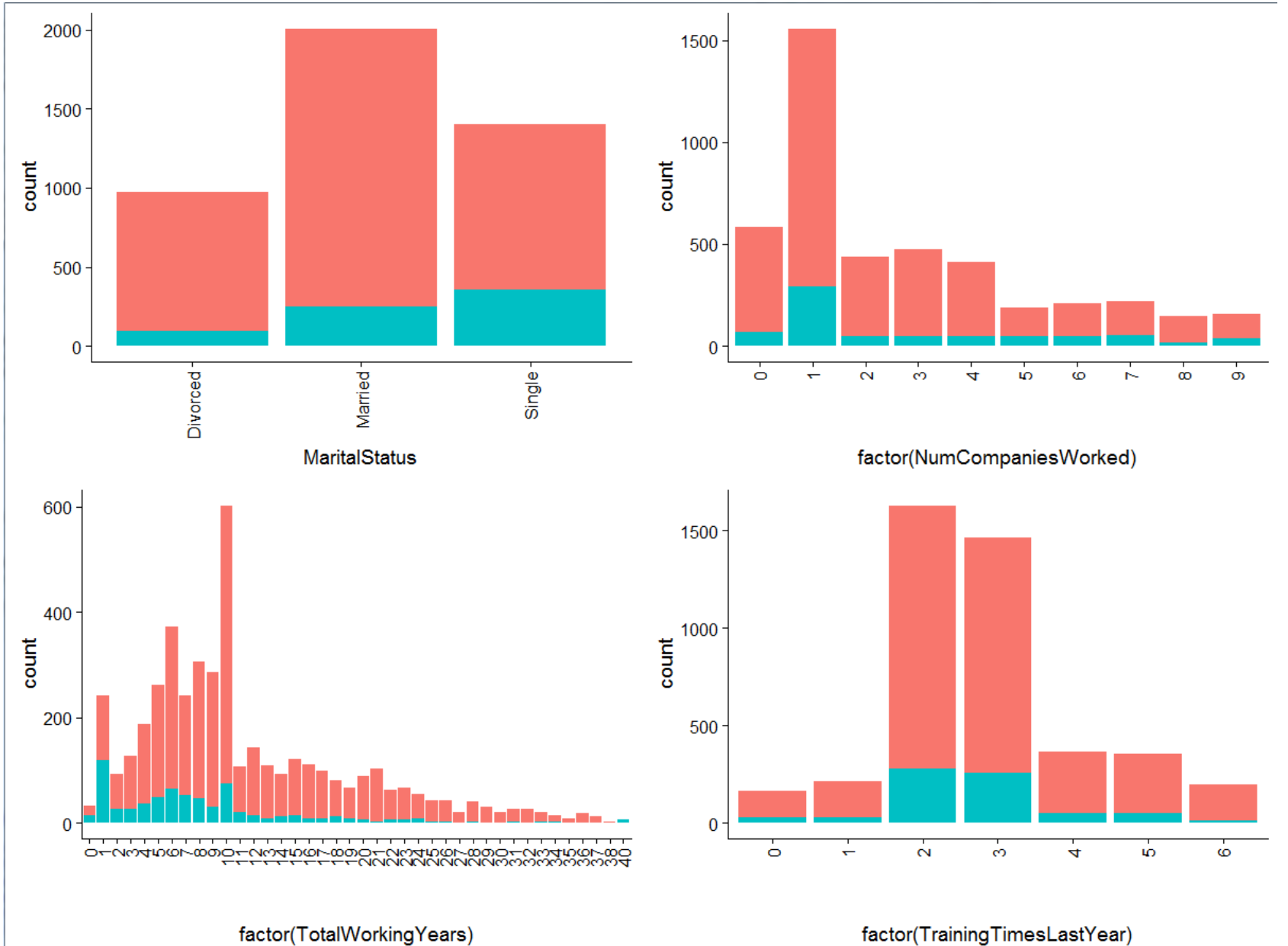
EDA

- Strong possibility if the job level is low, the attrition is more and is in R&D department
- Sales Executive & Research Scientist are more prone to attrition.
- Those who travels frequently having higher chances of attrition.



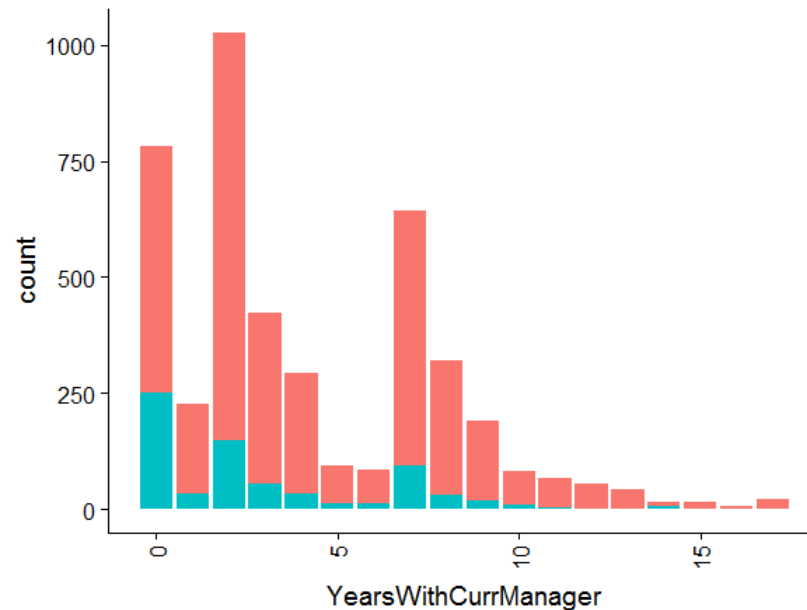
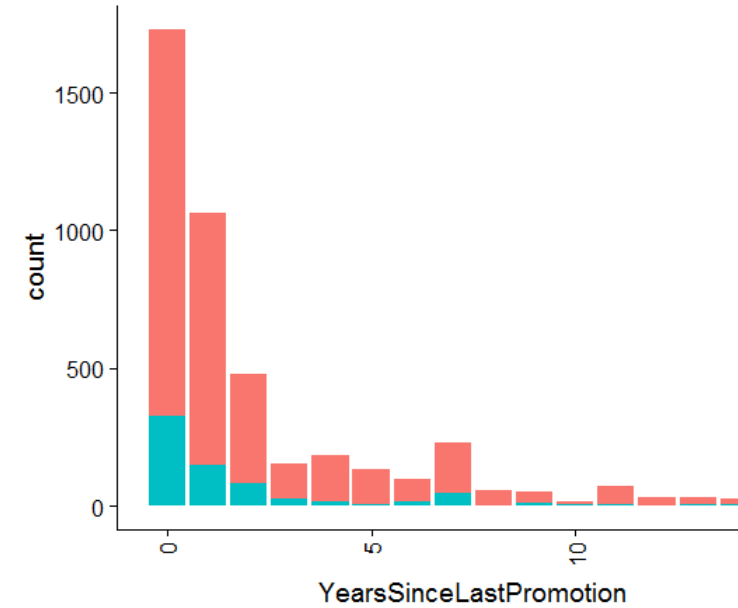
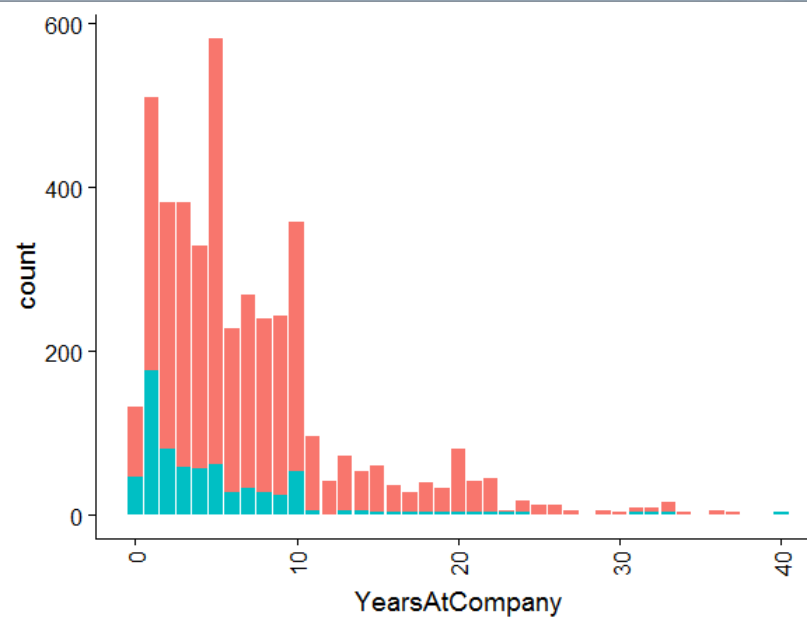
Attrition and Marital Status

- More attrition in the single status
- Less experienced people are also tend to leave the company.



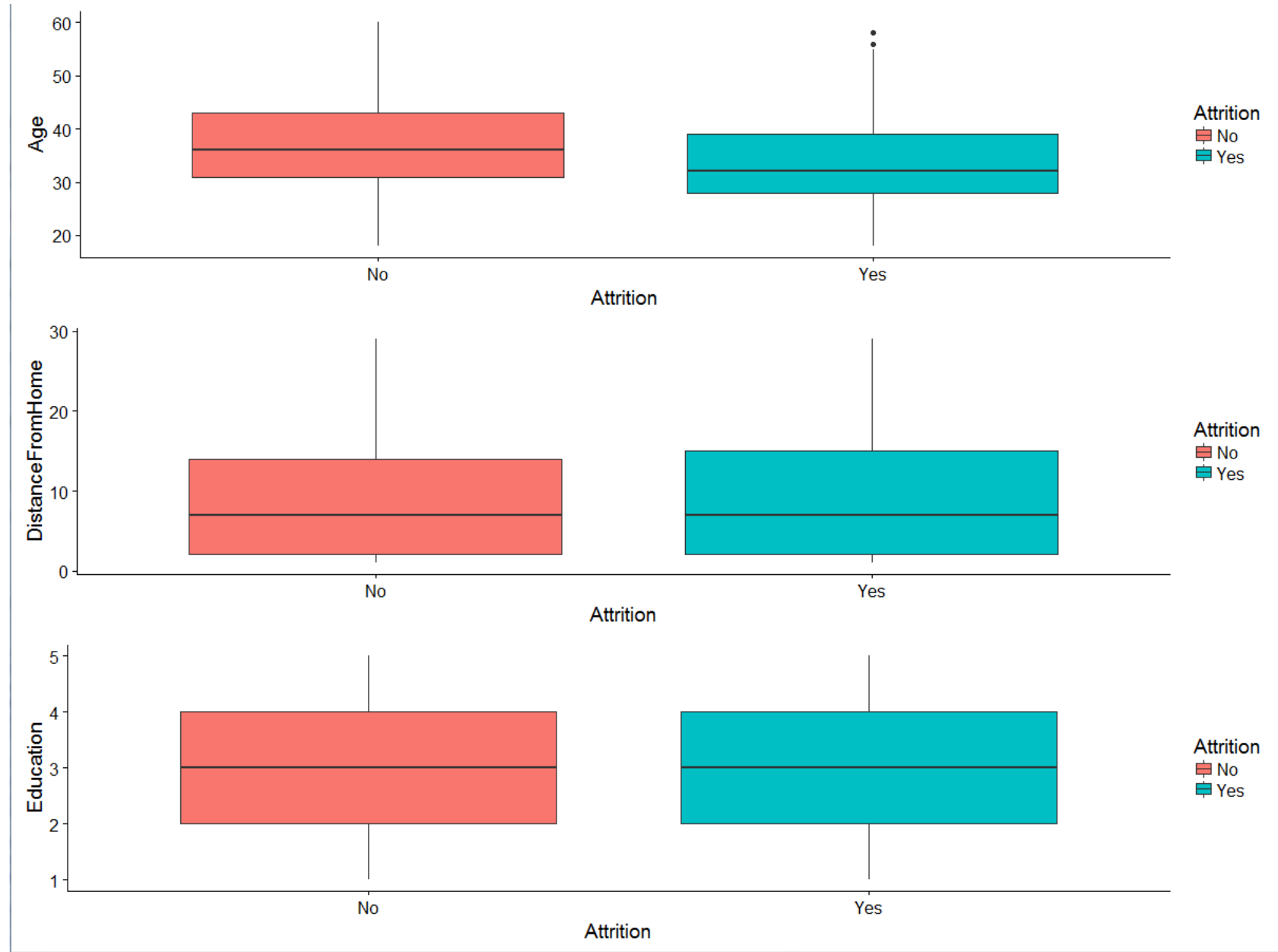
Attrition and Years at company

- Less attrition as they spend more years at the same company and with same manager



Attrition and age and Education

- Age or education or even the distance from home does not have much effect on employees leaving the company



Build the model

- From the all merged data, split the data in 70% for training and 30% for testing.
- Applying the stepAIC function, the model comes out with the result which has 26 variables
- Further reducing the model variables based on the vif value and the insignificance or probability, the final model is derived

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.93133    0.41180  -4.690 2.73e-06 ***
Environmentsatisfaction -0.35598    0.05672  -6.276 3.48e-10 ***
Jobsatisfaction -0.38136    0.05768  -6.611 3.81e-11 ***
WorkLifeBalance -0.24815    0.05657  -4.387 1.15e-05 ***
Age           -0.32220    0.08064  -3.996 6.45e-05 ***
JobLevel      -0.10159    0.05872  -1.730 0.083600 .
NumCompaniesworked  0.34669    0.06027   5.752 8.80e-09 ***
StockoptionLevel -0.08159    0.05790  -1.409 0.158806
TotalWorkingYears -0.55684    0.10341  -5.385 7.26e-08 ***
TrainingTimesLastYear -0.23093    0.05942  -3.886 0.000102 ***
YearssinceLastPromotion  0.68299    0.08259   8.270 < 2e-16 ***
YearswithCurrManager -0.57578    0.09334  -6.169 6.88e-10 ***
BusinessTravelTravel_Frequently  1.25814    0.24558   5.123 3.01e-07 ***
BusinessTravelTravel_Rarely    0.58864    0.22581   2.607 0.009140 **
EducationFieldLife.Sciences -1.34174    0.34253  -3.917 8.96e-05 ***
EducationFieldMarketing -1.38712    0.37241  -3.725 0.000196 ***
EducationFieldMedical -1.40476    0.34528  -4.068 4.73e-05 ***
EducationFieldother -1.63341    0.41616  -3.925 8.67e-05 ***
EducationFieldTechnical.Degree -1.68609    0.38926  -4.332 1.48e-05 ***
JobRoleHuman.Resources -0.56633    0.34478  -1.643 0.100468
JobRoleManager -0.57061    0.25301  -2.255 0.024114 *
JobRoleManufacturing.Director -0.80251    0.22589  -3.553 0.000381 ***
JobRoleResearch.Director  0.55466    0.22952   2.417 0.015667 *
JobRolesales.Executive  0.29416    0.14072   2.090 0.036580 *
MaritalStatusMarried  0.23843    0.16548   1.441 0.149643
MaritalStatusSingle  1.10269    0.16655   6.621 3.58e-11 ***
`Average Hours`      0.65091    0.05424  12.000 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2661.4  on 3009  degrees of freedom
Residual deviance: 2068.1  on 2983  degrees of freedom
AIC: 2122.1

Number of Fisher Scoring iterations: 6
```

Final Model

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.44885    0.09170  -26.704 < 2e-16 ***
EnvironmentsSatisfaction -0.32877    0.05470   -6.011 1.85e-09 ***
JobSatisfaction -0.35057    0.05482   -6.395 1.61e-10 ***
WorkLifeBalance -0.20959    0.05462   -3.837 0.000125 ***
NumCompaniesWorked  0.34583    0.05547    6.234 4.54e-10 ***
TotalWorkingYears -0.93078    0.08356  -11.140 < 2e-16 ***
YearsSinceLastPromotion  0.38374    0.06651    5.770 7.93e-09 ***
BusinessTravelTravel_Frequently  0.68074    0.12971    5.248 1.54e-07 ***
JobRoleManufacturing_Director -0.74998    0.21189   -3.539 0.000401 ***
MaritalStatusSingle  0.89603    0.11180    8.015 1.11e-15 ***
`Average Hours`      0.64488    0.05204   12.392 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2661.4  on 3009  degrees of freedom
Residual deviance: 2188.7  on 2999  degrees of freedom
AIC: 2210.7

Number of Fisher Scoring iterations: 5
```

- Final model has a total of 10 variables which are highly significant and less vif values.

The Factors Which Need focus to curb Attrition

These Factors can be Classify in two categories considering their correlation coefficient:

“From the coefficients obtained, we can predict the nature of the variable with respect to Attrition”

- **Positive Correlated Factors:**

The positively related variable raises the chance of Attrition of an employee, means if a Positive related variable increases, the Attrition of the employee increases.

- **Negatively Correlated Factors:**

The Negatively related variable decreases the chance of Attrition of an employee, means if a Negatively related variable increases, the Attrition of the employee increases.

The Factors Which Need focus to curb Attrition

Positive Correlated Factors:

1. MaritalStatusSingle:

From the Plot as well as from positive coefficient, we can infer that an unmarried employee have higher chance of leaving the company.

2. BusinessTravelTravel frequently:

Those who travels frequently are having more tendency to leave the company. This could also be infer from the plot. Also the sales executive need to travel a lot have high chances of attrition.

3. Average Hours :

Spending more time at office is also turn out to be the prominent factor of attrition.

4. YearsSinceLastPromotion :

As duration since last promotion raised the chances of Attrition also get increased.

5. NumCompaniesWorked :

The name of variable self descriptive in this case. As employee is changing company frequently, which means that its attrition level is high.

The Factors Which Need focus to curb Attrition

Negatively Correlated Factors

1. TotalWorkingYears :

As the Number of total Working Year increases the chances of leaving the company decreases. As with time the stability of an employee in the company increases.

2. JobRoleManufacturing.Director:

Out of the all the job roles, the jobs of higher level as that of Manufacturing Director are having lesser chance of attrition.

3. JobSatisfaction:

As Job Satisfaction rating of an employee is better, better be the chance he would remain in the company for longer time.

4. EnvironmentSatisfaction :

Similar to job satisfaction, high rated Environment satisfaction leads to lesser chance of attrition.

5. WorkLifeBalance :

Lesser be the rating in work & life balance for an employee, more be the chance he would quit the Job.

Final Model Performance Measures

Confusion Matrix and Statistics

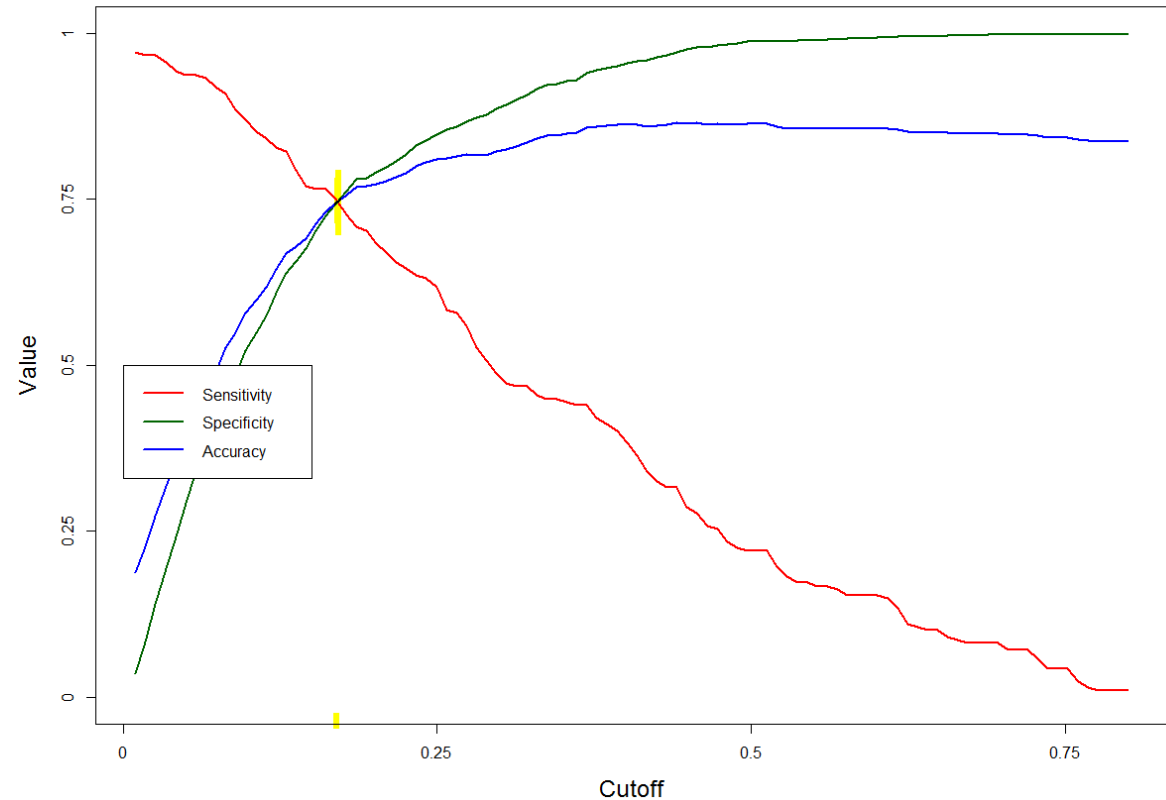
Prediction \ Reference	Reference	
	No	yes
No	1068	163
yes	13	46

Accuracy : 0.8636
95% CI : (0.8436, 0.8818)
No Information Rate : 0.838
P-Value [Acc > NIR] : 0.00617

Kappa : 0.2928
McNemar's Test P-Value : < 2e-16

Sensitivity : 0.22010
Specificity : 0.98797
Pos Pred Value : 0.77966
Neg Pred Value : 0.86759
Prevalence : 0.16202
Detection Rate : 0.03566
Detection Prevalence : 0.04574
Balanced Accuracy : 0.60403

'Positive' class : yes



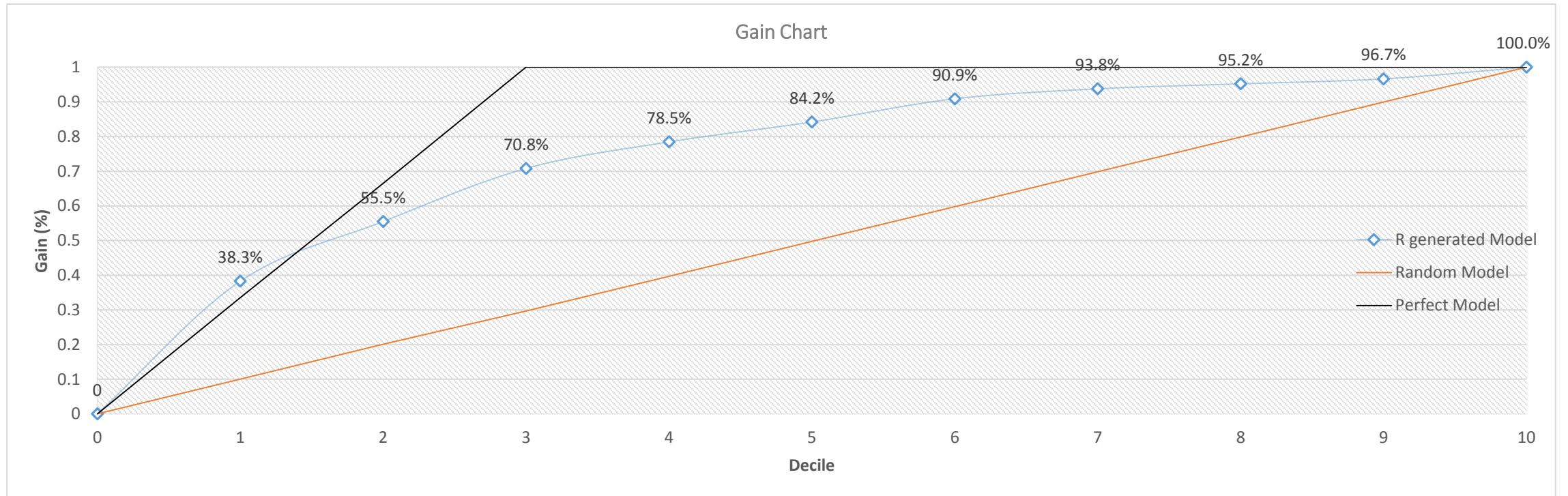
- The final model with 50% probability cut off has an accuracy of 86%, sensitivity of 22 % and specificity of 98 %
- In order to calculate the best cut off, the graph is useful and based on the same, the optimal cut off is at 16.9% and with this accuracy, sensitivity and specificity of this model is around 75%

Final Model Performance Measures

```
> churn_decile
# A tibble: 10 x 6
  bucket total totalresp cumresp      Gain      Cumlift
  <int> <int>    <dbl>    <dbl>    <dbl>    <dbl>
1     1    129      80      80  38.27751  3.827751
2     2    129     36    116  55.50239  2.775120
3     3    129     32    148  70.81340  2.360447
4     4    129     16    164  78.46890  1.961722
5     5    129     12    176  84.21053  1.684211
6     6    129     14    190  90.90909  1.515152
7     7    129      6    196  93.77990  1.339713
8     8    129      3    199  95.21531  1.190191
9     9    129      3    202  96.65072  1.073897
10    10    129      7    209 100.00000  1.000000
> |
```

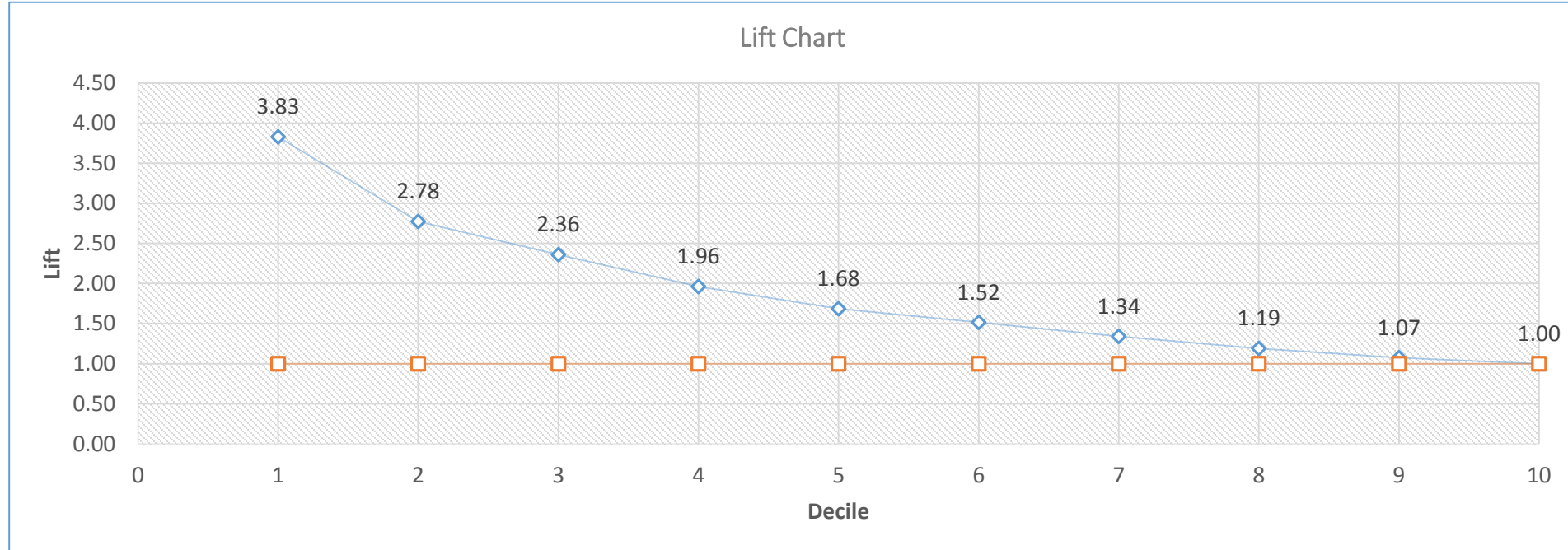
- 70% of attrition can be found at decile 3rd. By lift we get the factor by which our model outperform the random model. thus at 3rd decile, lift is 2.3

Gain Charts



- From Gain Chart shown above the Regenerated model we built is very close to the perfect model away from the random model. Thus from this Gain chart we can evaluate that the model we built is tends to be perfect.

Lift Charts



- The lift basically just tells you the factor by which your model is outperforming a random model. Thus, we can see from the Chart that at 3rd decile the performance is 2.3 times better than the random one.

KS Statistic

Decile	Observations	Churn	Cum- Churn	% Cum-Churn	Non- Churn	Cum-Non-Churn	%Cum-Non-Churn	(%Cum-Churn) - (%Cum-Non-Churn)
1	129	80	80	38.3%	49	49	4.5%	33.7%
2	129	36	116	55.5%	93	142	13.1%	42.4%
3	129	32	148	70.8%	97	239	22.1%	48.7%
4	129	16	164	78.5%	113	352	32.6%	45.9%
5	129	12	176	84.2%	117	469	43.4%	40.8%
6	129	14	190	90.9%	115	584	54.0%	36.9%
7	129	6	196	93.8%	123	707	65.4%	28.4%
8	129	3	199	95.2%	126	833	77.1%	18.2%
9	129	3	202	96.7%	126	959	88.7%	7.9%
10	129	7	209	100.0%	122	1081	100.0%	0.0%
Total	1290	209			1081			

- **From KS Statistic: The Model we built is a good model since,**
 1. *Its KS Statistic value is 49% which is greater than 40%.*
 2. *And It lies in the top 3rd deciles.*