

HR ANALYTICS CASE STUDY

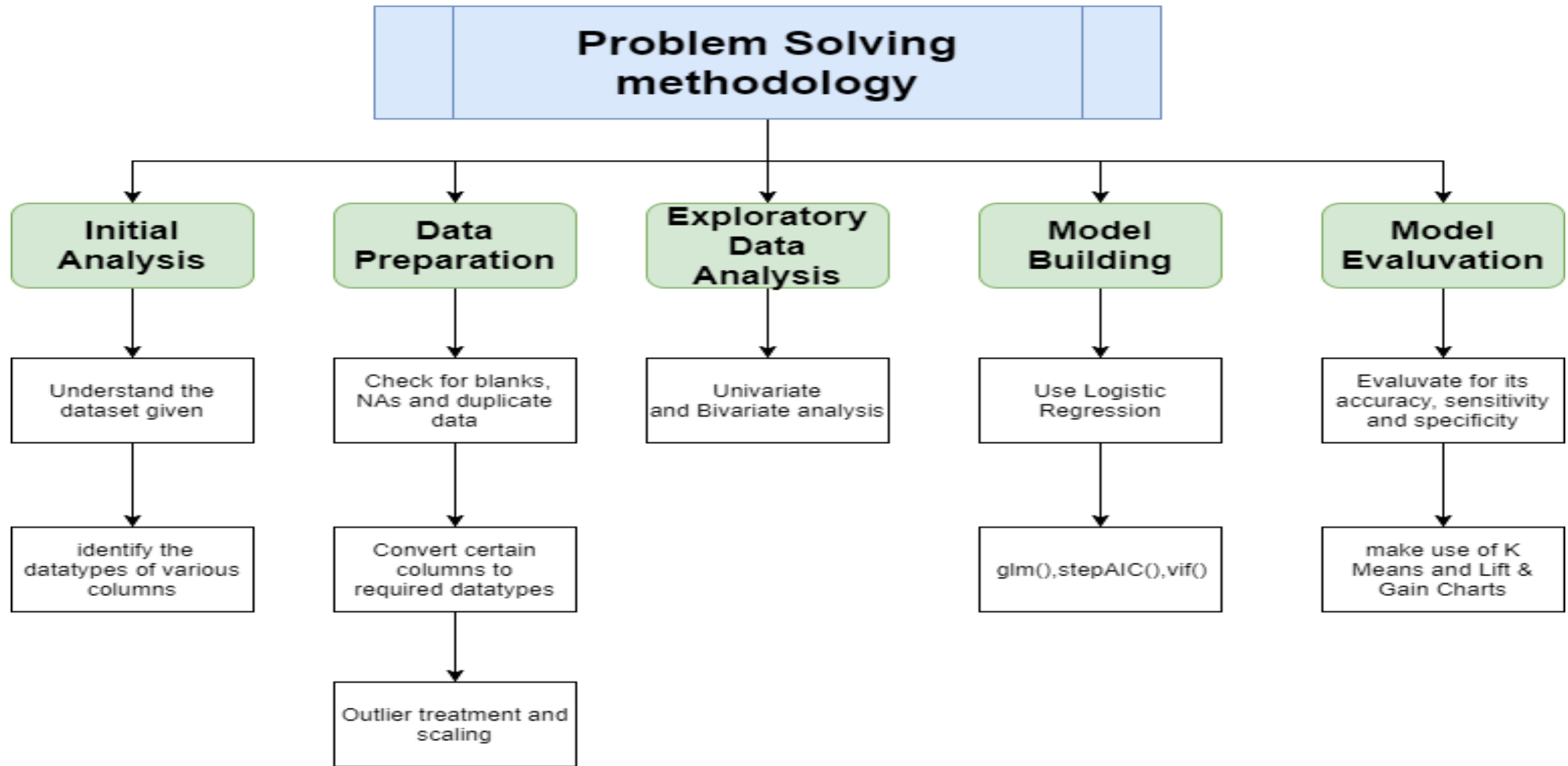
SUBMISSION

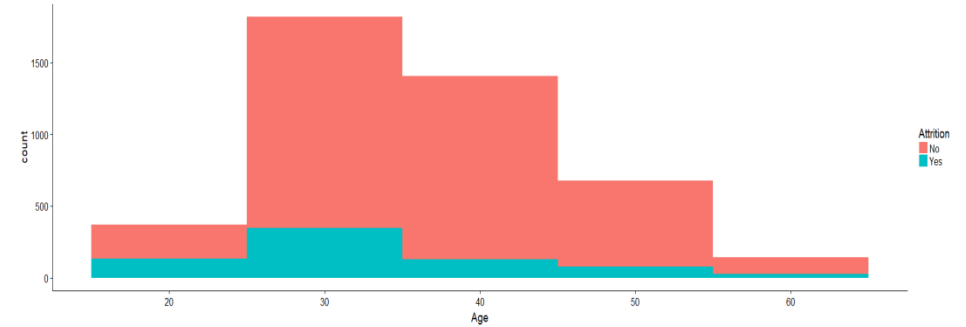
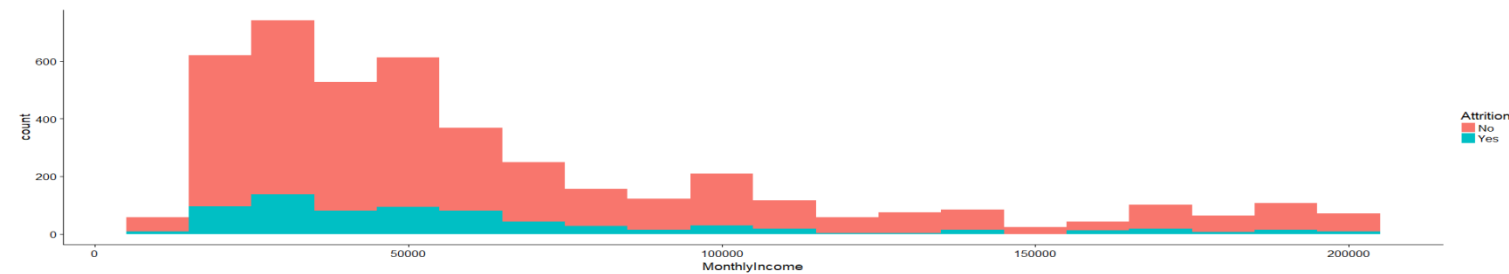
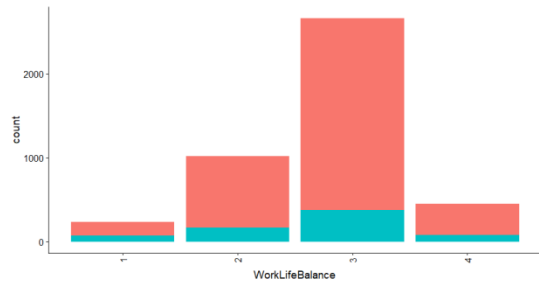
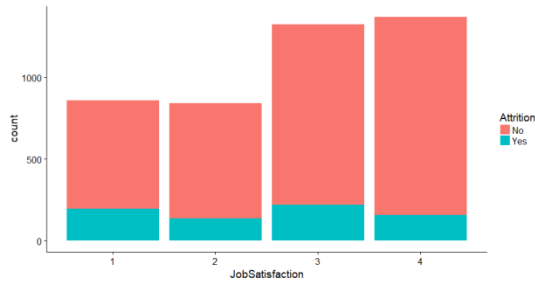
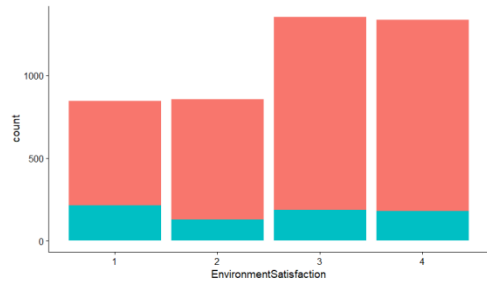
Group:

1. Anusha Muninarayanappa
2. Thejo Kishan
3. S Rakesh Kumar
4. Nikhil Chakravarthi

Abstract

- **Aim:**
 - To model the probability of attrition of a Company using logistic regression
- **Problem Statement:**
 - A large company named XYZ employs around 4000 employees, however, every year, around 15% of its employees leave the company and need to be replaced.
 - The management believes that this level of attrition is bad for the company and wants to understand what factors they should focus on, in order to curb attrition.
- **Approach**
 - The primary approach is to use build a Logistic Regression model in order to understand the important factors affecting the company.
 - We make use of EDA (Exploratory Data Analysis) operations like univariate and bivariate analysis to identify problems and tackle it building a model and evaluating it.
 - Using the data sets available, we make use of R, a statistical computing program to perform various operations and hence come up with the results.
- **Conclusion**
 - Come up with factors causing attrition and thus help the Company curb it.





- EDA was carried out on all available variables of the dataset.

- Continuous and categorical variables were segregated.

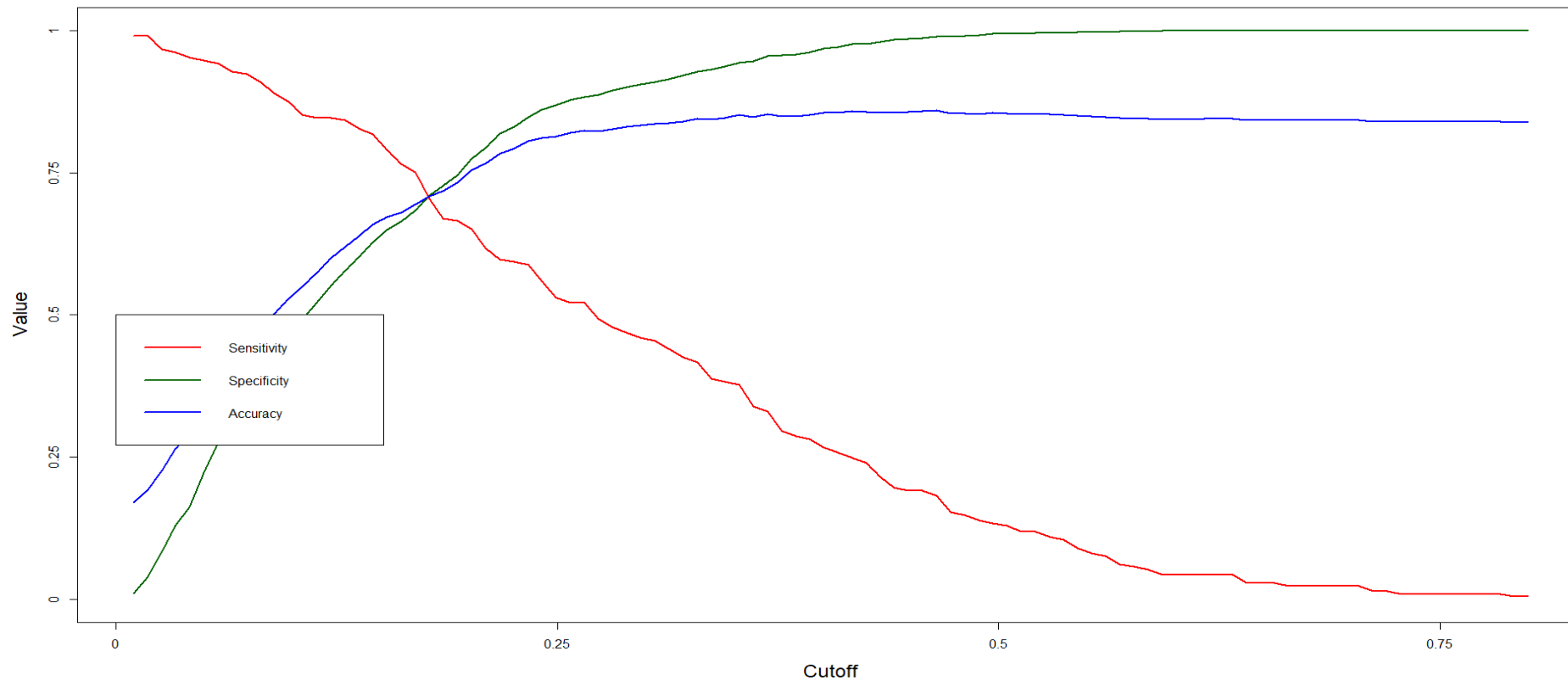
- Accordingly, outlier treatment and scaling were done

1. We make use of `glm()` function to build a logistic regression model. Attrition is dependent variable.
2. We use `stepAIC()` to remove insignificant categorical variables.
3. We run `vif()` function and look for values with higher vif and higher P-value.
4. Accordingly, in every iteration, we remove insignificant variables to build the final model.
5. The final model had 16 significant variables: `TotalWorkingYears`, `YearsSinceLastPromotion`, `YearsWithCurrManager`, `Avg.hrs`, `JobSatisfaction4`, `EnvironmentSatisfaction2`, `EnvironmentSatisfaction3`, `EnvironmentSatisfaction4`, `BusinessTravelTravel_Frequently`, `DepartmentResearch & Development`, `DepartmentSales`, `JobRoleManufacturing Director`, `MaritalStatusSingle`, `NumCompaniesWorked5`, `NumCompaniesWorked7`.

1. We start model evaluation assuming the probability cut-off of 0.5.

And Found: Accuracy = 85.81%, Sensitivity = 22.00% and Specificity = 98.14%.

2. Then, to find optimal probability cut-off, we used confusionMatrix and found the cut-off to be 0.17.



1. For the optimal probability cut-off of 0.17,

Accuracy = 70.85%

Sensitivity = 70.33% and

Specificity = 70.95%.

2. KS – Statistic

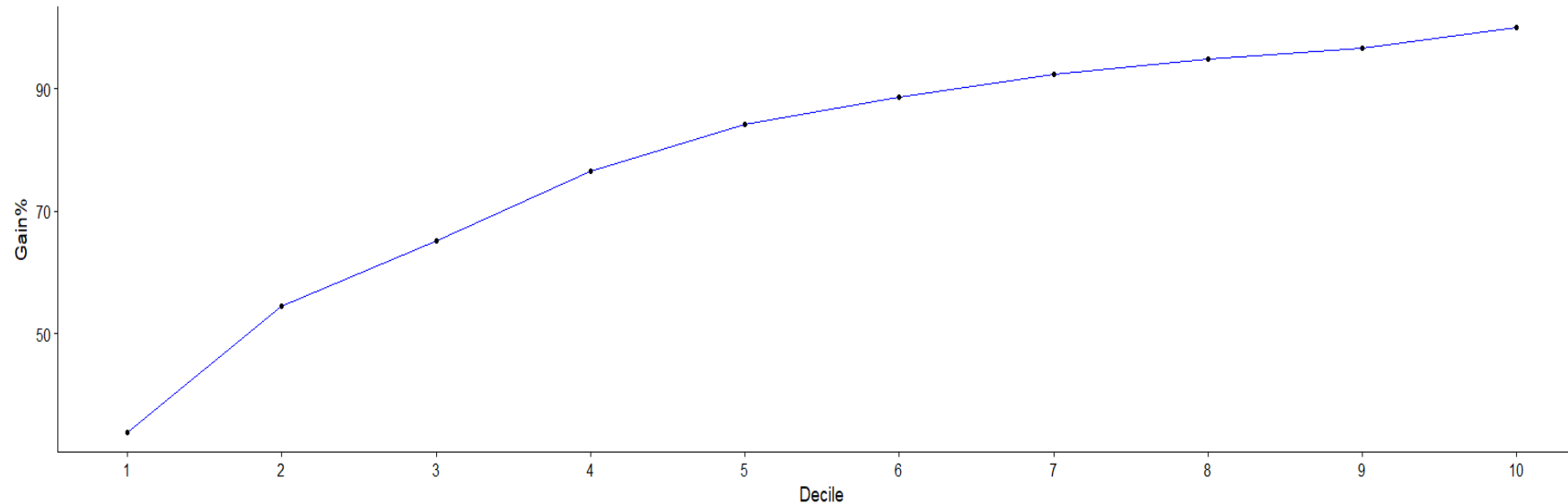
The KS Statistic was found to be 0.4128775 i.e 41.3%

- A good model is one for which the KS statistic:
- is equal to 40% or more
- lies in the top deciles, i.e. 1st, 2nd, 3rd or 4th

3. We further plot the Gain and Lift Charts.

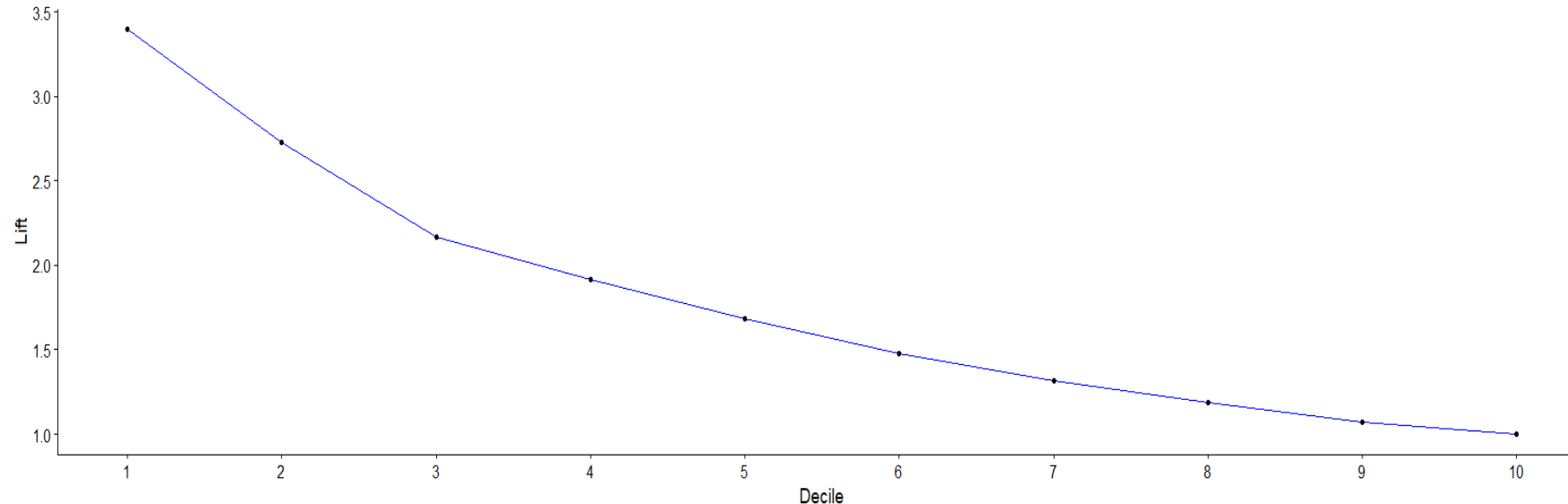
	bucket <int>	total <int>	totalresp <dbl>	Cumresp <dbl>	Gain <dbl>	Cumlift <dbl>
1	1	129	71	71	33.97129	3.397129
2	2	129	43	114	54.54545	2.727273
3	3	129	22	136	65.07177	2.169059
4	4	129	24	160	76.55502	1.913876
5	5	129	16	176	84.21053	1.684211
6	6	129	9	185	88.51675	1.475279
7	7	129	8	193	92.34450	1.319207
8	8	129	5	198	94.73684	1.184211
9	9	129	4	202	96.65072	1.073897
10	10	129	7	209	100.00000	1.000000

Gain Chart



The gain for the given model is 76.55% by the 4th decile. This means, that if you sort all employees according to probability, then among the top 40% of the employees, 76.55% are more likely to switch the company.

Lift Chart



The model's lift is equal to 1.9 by the 4th decile. It means, the model catches 1.9 times more attrition than a random model would have caught at the 4th decile.

Conclusion

1. The Logistic Regression model built for the given data set.
2. 16 Important Variables affecting attrition were found.

For the optimal probability cut-off of 0.17

- Accuracy = 70.85%
 - Sensitivity = 70.33% and
 - Specificity = 70.95%.
3. The Model was found to be a good model based on the model evaluation that was done using KS Statistic along with Gain and Lift charts.
 - The KS Statistic was found to be 41.3%
 - Gain% by the 4th decile was 76.55%
 - Lift by the 4th decile was 1.9.
 4. This Model with significant variables should help the Company to understand what changes they should make to their workplace, in order to get most of their employees to stay and cut attrition.