HR ANALYTICS CASE STUDY

SUBMISSION

Group Name:
1. Abhay Channe
2. Naresh Mepperla
3. Phani Mandalika
4. Sunil Bangaru

Abstract – HR analytics in XYZ

Problem Statement:

A large company named **XYZ** required a model from HR Analytics firm for the probability of attrition using a logistic regression. The results thus obtained will be used by the XYZ management to understand what changes they should make to their workplace, in order to get most of their employees to stay.

Data Collection andAnalysis:

The team has collected the below data to start working on the problem statement. The main variables considered for the analysis but not

- Unique Id of the employee
- Attrition
- Department, Job role & Level
- Education level & field
- Monthly income, stock options and performance evaluation ratings
- Experience (total, with XYZ & with current Manager)
- Business travel details
- Various survey results for Job Involvement, Environment satisfaction etc.

Steps followed -

Data Cleaning:

• Delete columns with no information

• Check and remove all duplicates

• Check for missing and NA values. Work to eliminate, where possible

Data Preparation:

Clean the data as per above

Converting values to numeric for easy calculation, where needed

Creating dummy variables for factor variable with >2 levels

Addition of columns for univariate and bivariate analysis

Univariate Analysis - 1

Environment Satisfaction	Attrition		
Levels	**No**	**Yes**	**Grand Total**
High	86.28%	13.72%	100.00%
Low	74.58%	25.42%	100.00%
Medium	84.98%	15.02%	100.00%
Very High	86.51%	13.49%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**
Job Satisfaction	**Attrition**		
Levels	**No**	**Yes**	**Grand Total**
High	83.49%	16.51%	100.00%
Low	77.10%	22.90%	100.00%
Medium	83.60%	16.40%	100.00%
Very High	88.61%	11.39%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**
Work-Life Balance	**Attrition**		
Levels	**No**	**Yes**	**Grand Total**
Bad	69.07%	30.93%	100.00%
Best	82.22%	17.78%	100.00%
Better	85.63%	14.37%	100.00%
Good	83.38%	16.62%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Environment Satisfaction survey gives us some indications which are quite obvious –
- With Low satisfaction, the Attrition is quite high (~25.5%) than average (16.16%)

Job Satisfaction survey gives us similar indications –
- With Low satisfaction ratings, the Attrition is quite high (~23%) than average (16.16%)
- With Medium and High ratings as well, the Attrition is above the average

Work-Life Balance survey gives indications on the same line –
- With Bad satisfaction, the Attrition is quite high than average (31%)
- Only the better ratings gives below Average attrition percentage

Univariate Analysis - 2

Business Travel	Attrition		
Levels	**No**	**Yes**	**Grand Total**
Non-Travel	91.82%	8.18%	100.00%
Travel_Frequently	75.40%	24.60%	100.00%
Travel_Rarely	84.92%	15.08%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Department	Attrition		
Levels	**No**	**Yes**	**Grand Total**
Human Resources	70.97%	29.03%	100.00%
Research & Development	84.22%	15.78%	100.00%
Sales	84.85%	15.15%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Educations	Attrition		
Levels	**No**	**Yes**	**Grand Total**
Bachelor	84.49%	15.51%	100.00%
Below Collage	84.57%	15.43%	100.00%
Collage	81.17%	18.83%	100.00%
Doctor	85.71%	14.29%	100.00%
Master	84.25%	15.75%	100.00%

Business Travels frequency gives us some indications –
- With Frequent Travel employee, the Attrition is quite high (~25%) above than average (16.16%)

Department wise trend –
- Department attrition trend shows HR department have higher attrition than others

Education survey shows that the education level does not play any significant role in the Attrition numbers

Univariate Analysis - 3

Education Field	Attrition		
Levels	**No**	**Yes**	**Grand Total**
Human Resources	60.00%	40.00%	100.00%
Life Sciences	83.30%	16.70%	100.00%
Marketing	84.22%	15.78%	100.00%
Medical	83.94%	16.06%	100.00%
Other	87.34%	12.66%	100.00%
Technical Degree	88.28%	11.72%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Job Role	Attrition		
Levels	**No**	**Yes**	**Grand Total**
Healthcare Representative	85.41%	14.59%	100.00%
Human Resources	86.36%	13.64%	100.00%
Laboratory Technician	83.88%	16.12%	100.00%
Manager	86.96%	13.04%	100.00%
Manufacturing Director	88.63%	11.37%	100.00%
Research Director	77.02%	22.98%	100.00%
Research Scientist	81.61%	18.39%	100.00%
Sales Executive	83.05%	16.95%	100.00%
Sales Representative	85.06%	14.94%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Education Field confirms indications on the Department trend–
- HR area of education contributes more in Attrition numbers (40%)

Jobe role don't tilt towards any particular Job role for attrition.
If at all, we look deeply, Research Director & Research Scientist are prospective employees who show Attrition trend

Univariate Analysis - 4

Marital Status	Attrition		
Levels	No	Yes	Grand Total
Divorced	90.09%	9.91%	100.00%
Married	87.25%	12.75%	100.00%
Single	74.67%	25.33%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Job Involvement	Attrition		
Levels	No	Yes	Grand Total
High	84.69%	15.31%	100.00%
Low	78.42%	21.58%	100.00%
Medium	83.88%	16.12%	100.00%
Very High	81.67%	18.33%	100.00%
Grand Total	**83.84%**	**16.16%**	**100.00%**

Marital status indicates, Single employee tends to switch the job more often than other categories

Low Job Involvement also leads employee to switch the job.

Overall, the attrition for XYZ company is 16.16%

Logistic Regression Model Building

- Dummy Variables created for Categorical variables with >2 levels
- Dropped the columns with not-useful data e.g. Employee Count, Over 18 etc.
- Divided the Total dataset in 70% training & 30% test data set
- Build the first module with glm() function
  - First module with AIC = 2237.2 & 30 elements
  - StepAIC model building results in next model with AIC = 2201.2
- First model built with 0.5 probability for Attrition
  - Accuracy –   84.42%
  - Sensitivity – 15 %
  - Specificity – 98.62%
- Optimization of model is needed to increase the Sensitivity
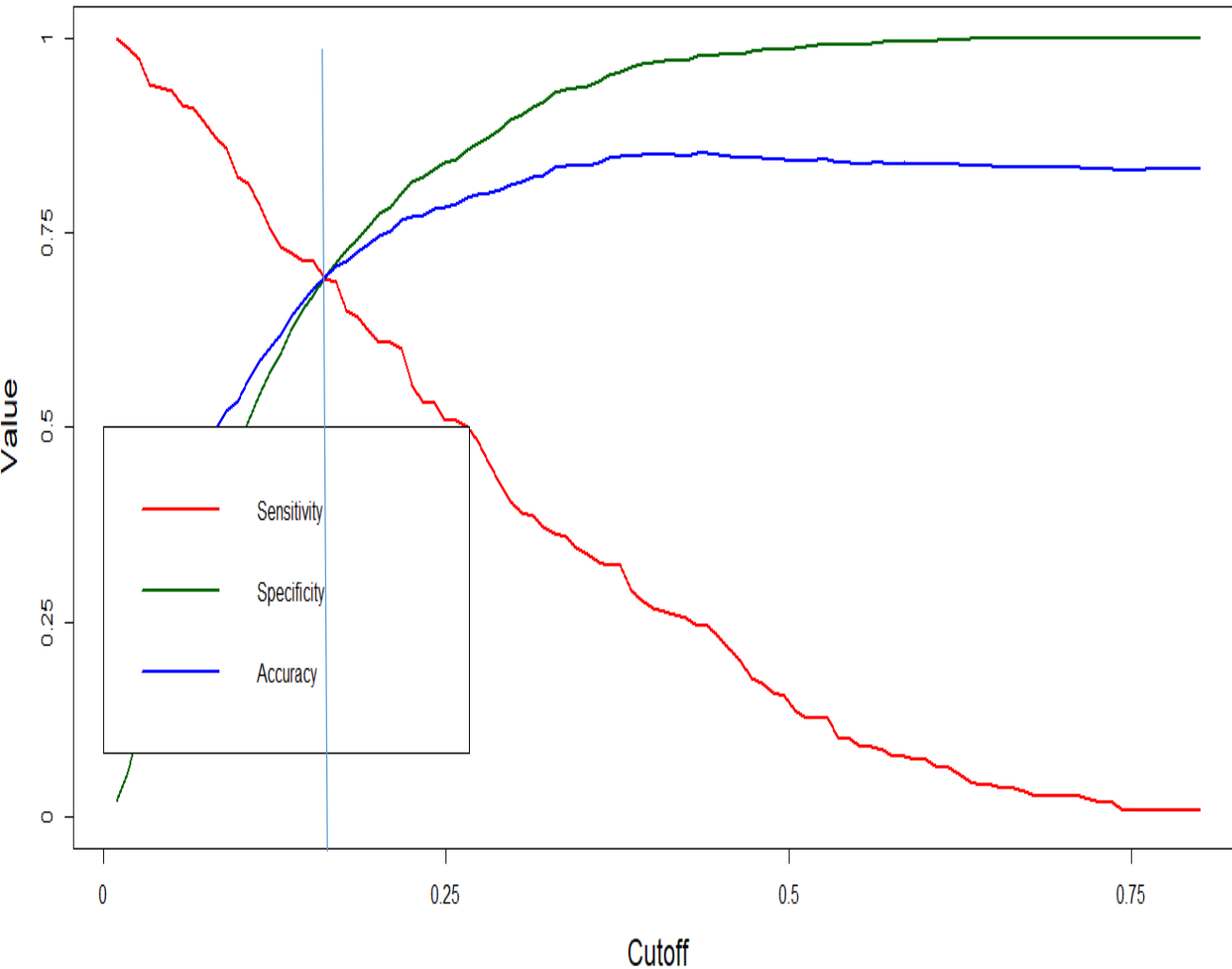
- Plotting the Sensitivity, Specificity & Accuracy to find the optimal cutoff value
- The Cutoff value of Attrition probability is ~0.1616
- With the optimized cutoff probability –
  - Accuracy –   69.14%
  - Sensitivity –   69.09%
  - Specificity –   69.15%
  - KS-Stats –   38.24%
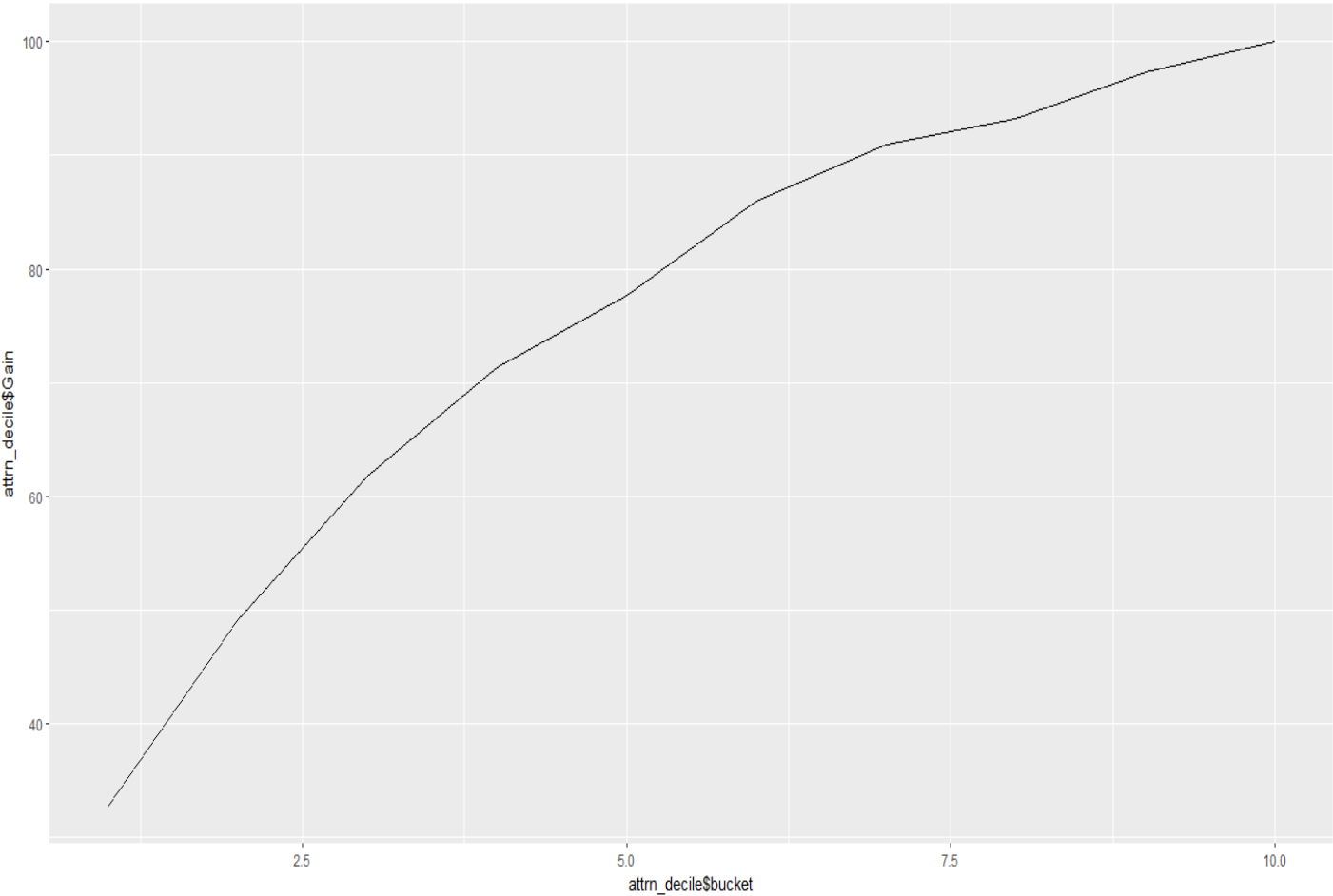
	bucket	total	totalresp	Cumresp	Gain	Cumlift
1	1	129	72	72	32.72727	3.272727
2	2	129	36	108	49.09091	2.454545
3	3	129	28	136	61.81818	2.060606
4	4	129	21	157	71.36364	1.784091
5	5	129	14	171	77.72727	1.554545
6	6	129	18	189	85.90909	1.431818
7	7	129	11	200	90.90909	1.298701
8	8	129	5	205	93.18182	1.164773
9	9	129	9	214	97.27273	1.080808
10	10	129	6	220	100.00000	1.000000

Recommendations

- Work on the survey results, specially when the survey results are indicating Low/Medium satisfactions

- HR department are the department to be focused on, since it shows ~40% attrition rate

- Single status employee tend to switch more, so other parameters need to improve like Job Satisfaction, Work-Life Balance etc

- The lower experience employee are the segment who looks for switch in the companies
  - Lower total experience
  - Less number of companies (<2) worked