HR Analytics Case Study

(Logistics Regression)

Group Name:

1. Deepak Aneja
2. Suresh Balla
3. Merin Jose
4. Fayiz Mayam Veettil

25 Mar 2018

Problem Statement

Objective

- Understand the factors in order to curb attrition of the Company **XYZ**
- Know which of the variables is most important and needs to be addressed right away
- Company want to know what changes they should make to their workplace, in order to get most of their employees to stay in the company

Data Exploration

Dataset

- Contains information about the employees working in the Company XYZ
- Around 4,400 employees records from Jan ‘15 to Dec ‘15
- 30 different variables to understand the employees behaviour and status in the dataset
- Dataset across five different sheets including
  - General Information
  - Daily time in and time out details
  - Employees’ and Managers’ job satisfaction and work life balance survey data
- The aim is to model the **probability of attrition** using a logistic regression

Step by Step Approach for Analysis

- Cleaning and formatting data
- Derive useful metrics out of data
- Data Binning based on WOE and IV
- Check the data density and correlations
- Data Scaling
- Splitting the data between train and test
- Data Modeling
- Conclusion

Data Cleaning and Manipulation

- Check for possible data inconsistencies
  - Check for duplicate records
  - Check for NA values in the columns used for analysis
- Status of data cleaning
  - No duplicate records found
  - Uses the column ‘Attrition’ as targeted variable
  - 38 dummy variables were created for modeling

Binning of Continuous Variable using WOE and IV

Variables tested with WOE and IV Factors

Bins Worked

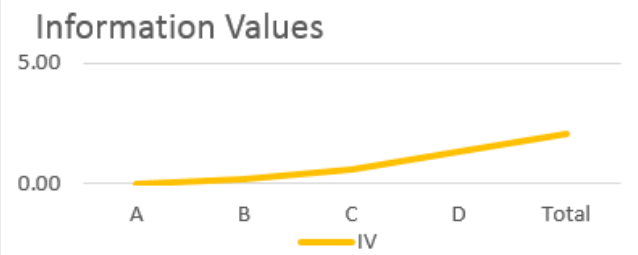
- Age
- Distance From Home
- Total Working Year

Bins Didn't Work

- Year At Company
- Years with Cur. Manager
- Monthly Income

Age Groups - WOE and IV Pattern

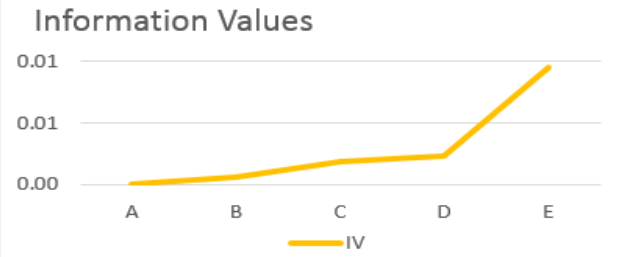
Group	No of good	No of bad	WOE	IV
A	132	237	0.0641346	0.00
B	348	1470	-0.7914224	0.18
C	177	1617	-1.5627853	0.56
D	1200	231	2.2970519	1.34
Total	**1857**	**3555**		**2.08**



Age Bins
18-25
26-30
31-35
36-50
50+

Distance From Home - WOE and IV Pattern

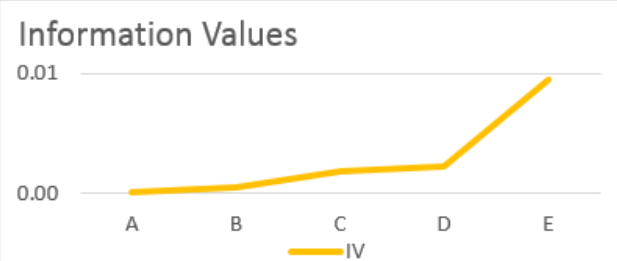
Group	No of good	No of bad	WOE	IV
A	303	1593	-0.0104961	0.00
B	198	984	0.0457865	0.00
C	63	282	0.150373	0.00
D	69	306	0.1596668	0.00
E	78	534	-0.2745417	0.01
Total	**711**	**3699**		**0.01**



Bins (KM)
1-5
6-10
11-15
16-20
20+

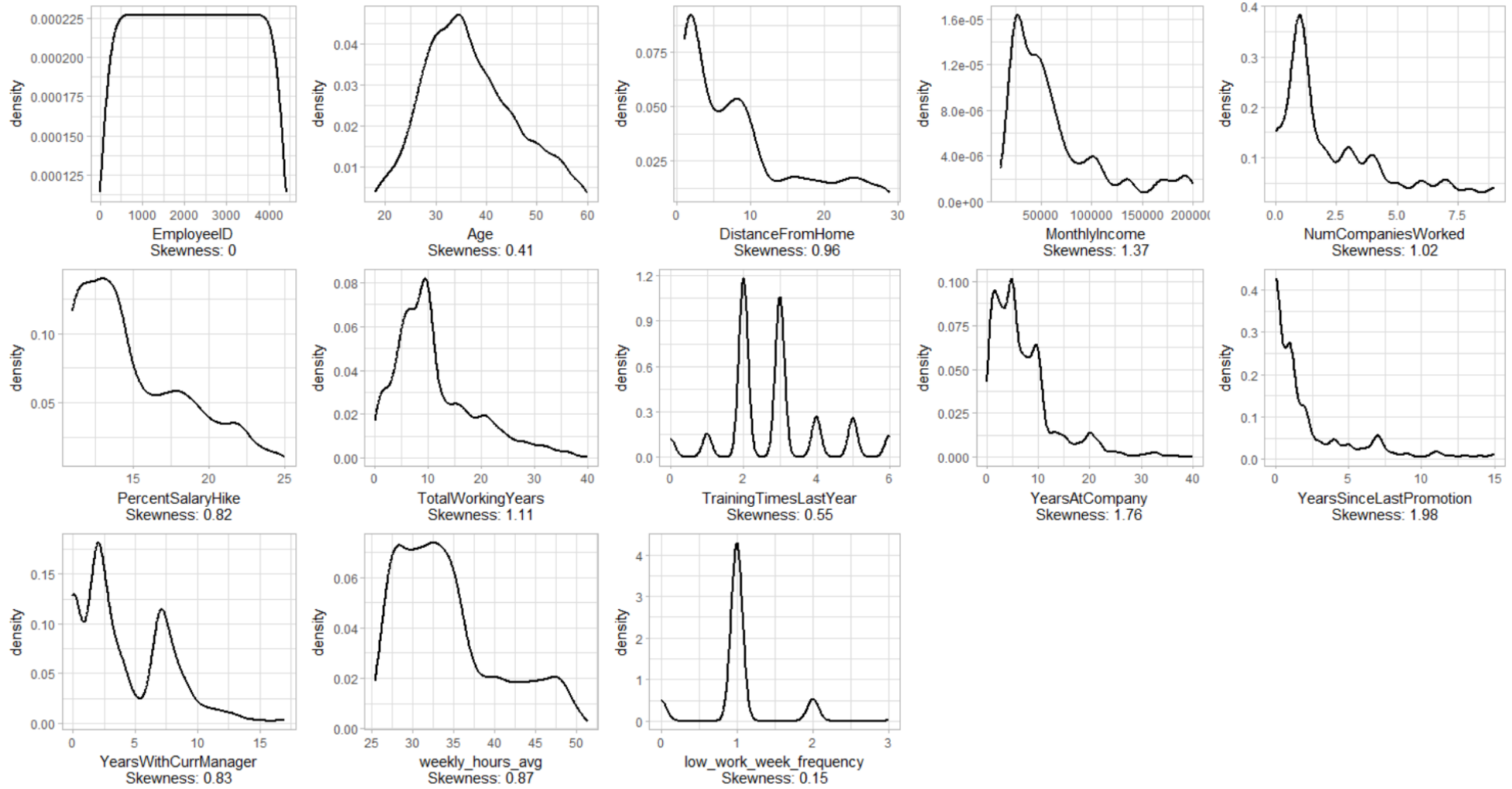
Total Working Years - WOE and IV Pattern

Group	No of good	No of bad	WOE	mation V
A	161	207	1.398754	0.24
B	111	468	0.21113	0.01
C	272	1545	-0.08691	0.00
D	165	1472	-0.53836	0.09
Total	**709**	**3692**		**0.34**



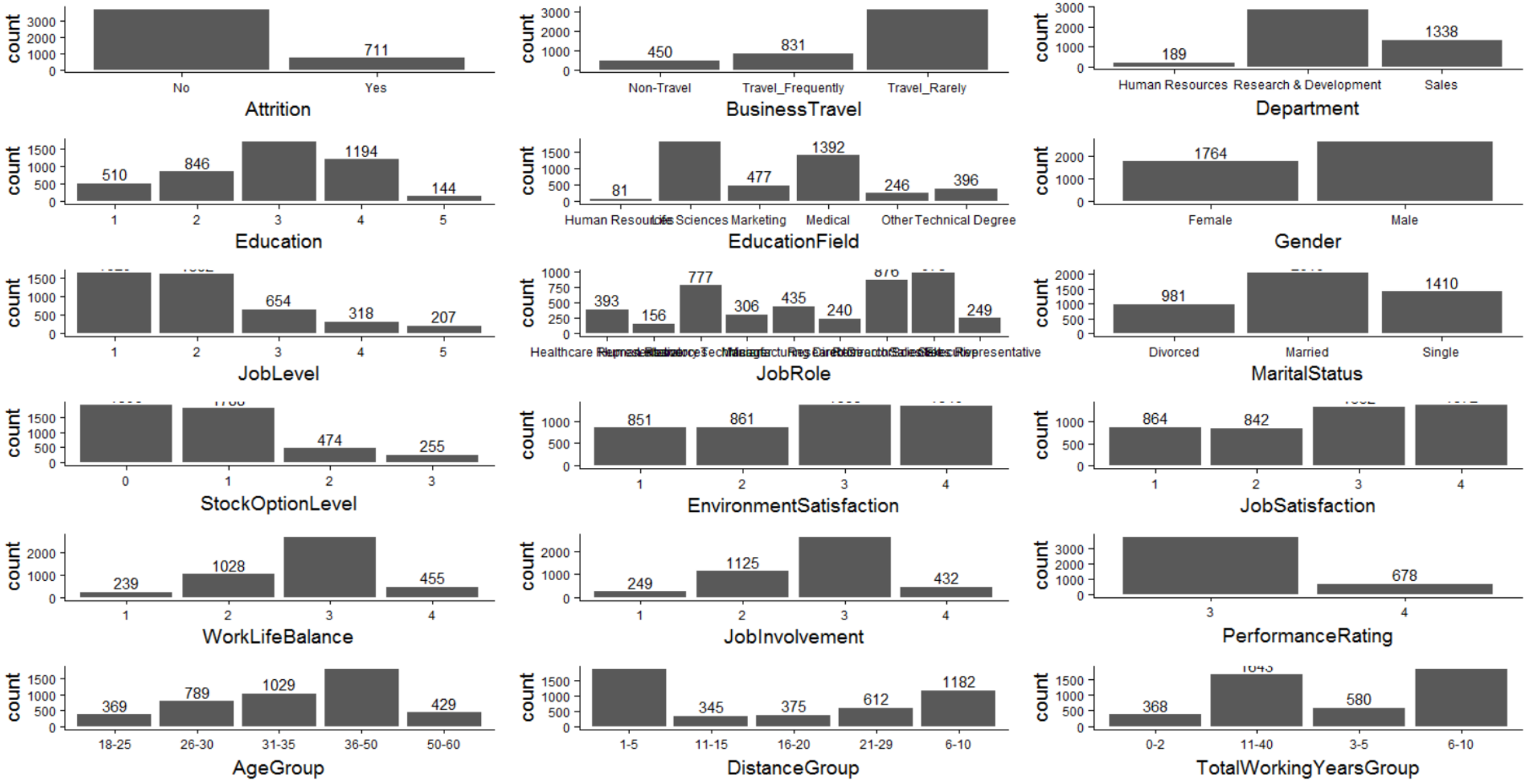
Bins (Yrs)
1-2
3-5
6-10
11-40
20+

Data Density - I



Data density check on numerical data for further modeling

Data Density - II



Data density check on categorical data for modeling

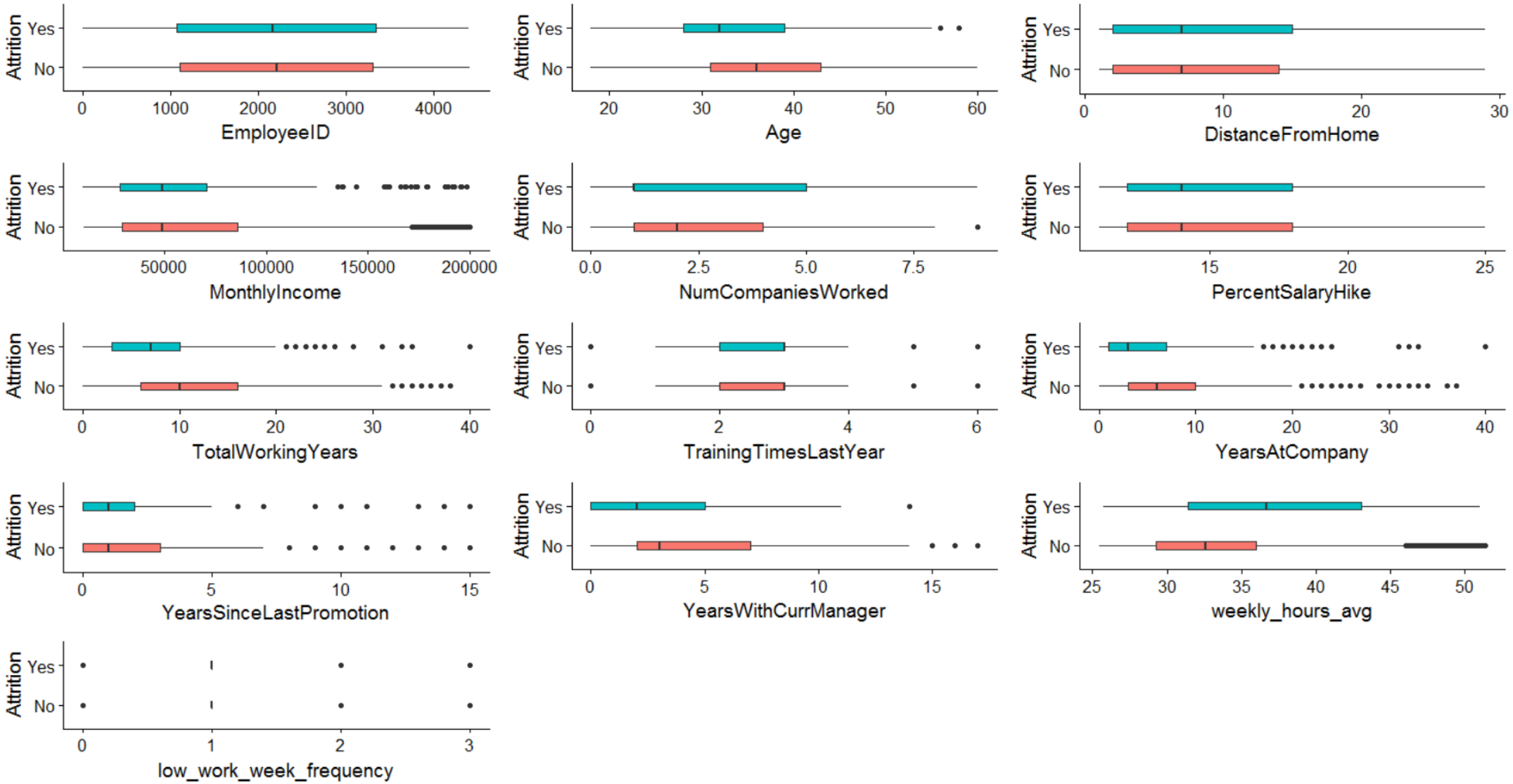
Univariate Analysis - I

Status of variables relative to attrition

Univariate Analysis - II

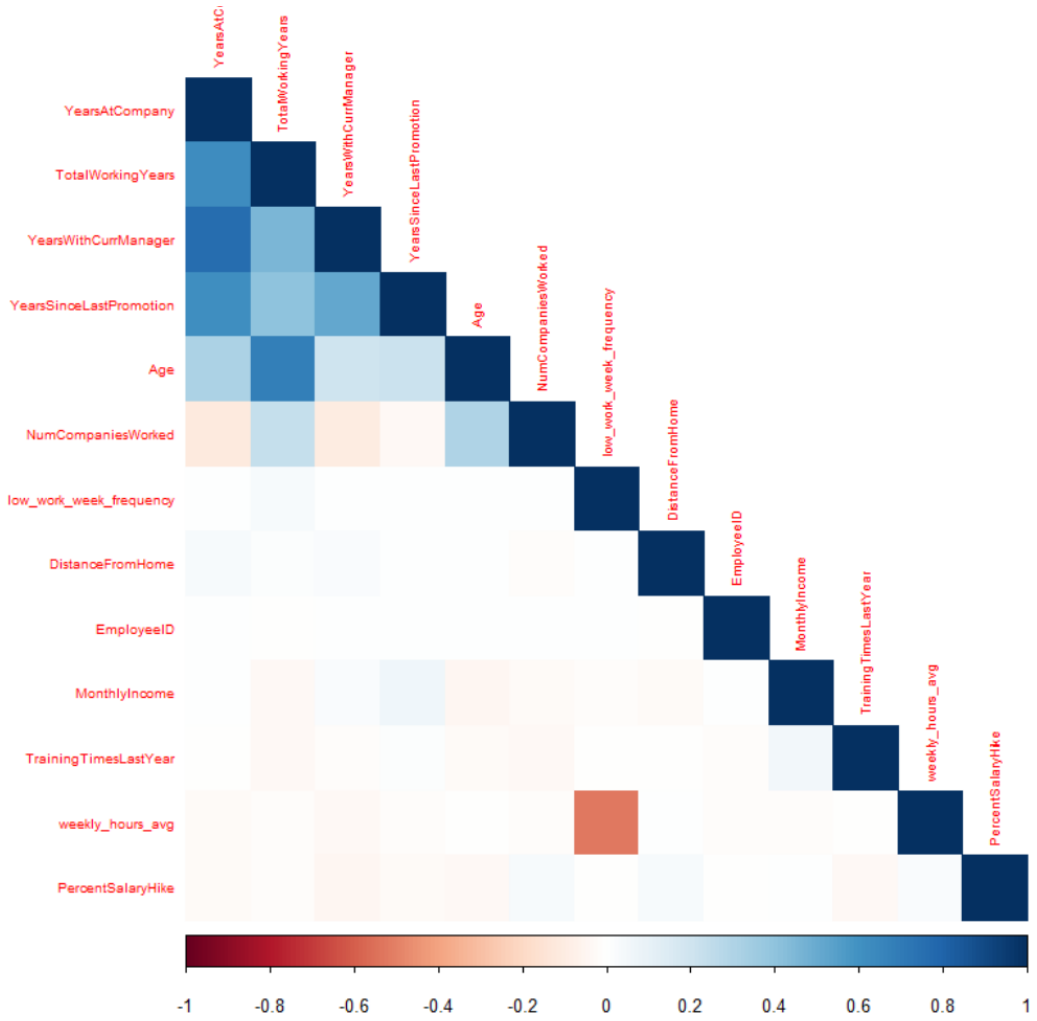
Percentage of variables relative to attrition

Attrition Status



Numeric box plots against attrition status

Data Correlation



Checks the correlation between each variable for further modeling

Steps for Model Creation (Logistics Regression)

- Monthly Income Scaled
- Data split by 70% for training and 30% for testing
- Initial Model run with 65 Variables (AIC 2094, Null Deviance 2728 and Residual Div 1960)
- Performs step wise model selection by StepAIC and reduced the variable length to 41
- VIF comparison with variable significance
- Conclude the final model
- Model Evaluation

Model Evaluation

Probability Cut-off of 50%

Actual	Predicted	
	Non-Attrition	Attrition
Non-Attrition	1086	24
Attrition	157	56

Accuracy	86.32%
Sensitivity	26.29%
Specificity	97.84%

Confusion Matrix and Statistics (40%)

Actual	Predicted	
	Non-Attrition	Attrition
Non-Attrition	1046	133
Attrition	64	80

Accuracy	85.11%
Sensitivity	37.56%
Specificity	94.23%

Confusion Matrix and Statistics (20%)

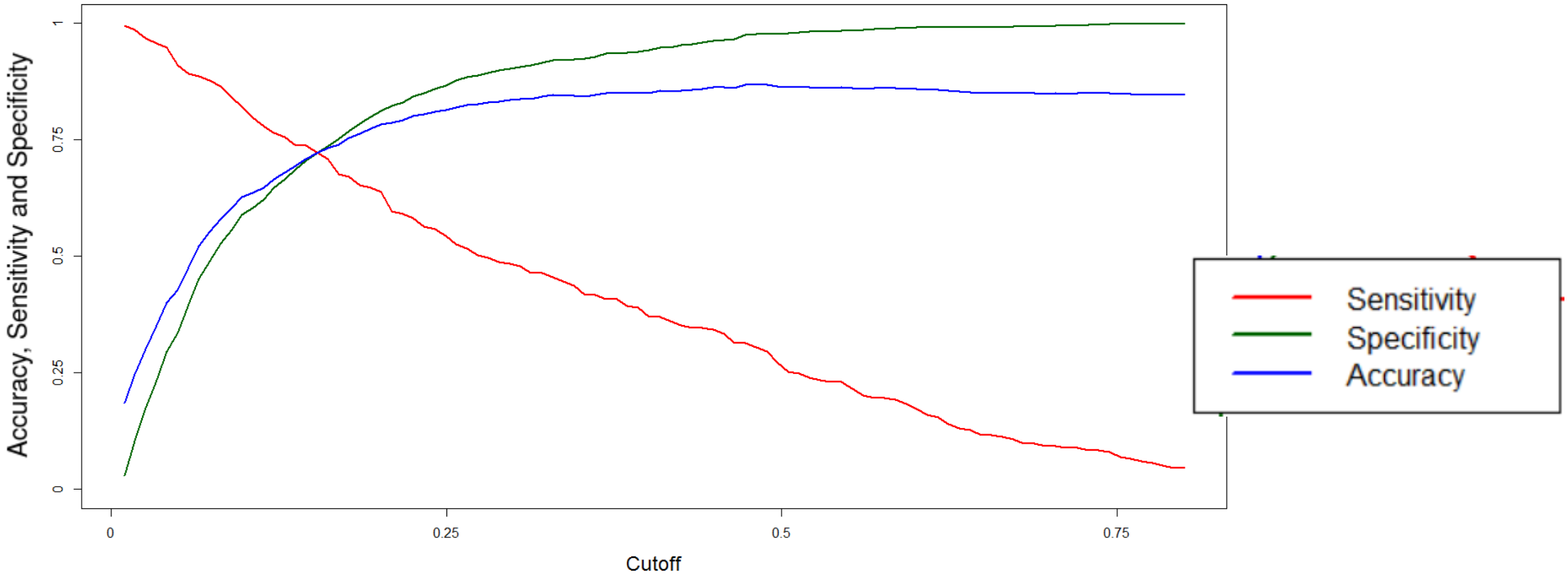
Actual	Predicted	
	Non-Attrition	Attrition
Non-Attrition	794	58
Attrition	316	155

Accuracy	71.73%
Sensitivity	72.77%
Specificity	71.53%

Calculation Gain, Lift and KS Statistics

Bucket	Total	Attrition	NonAttrition	CumAttrition	CumNonAttrition	GAIN	Cum Lift	%cum attr	%non-attr	%cum_attr - %cum Non attr
1	133	77	56	77	56	36.2	3.62	36.15%	5.05%	31.11%
2	132	39	93	116	149	54.5	2.72	54.46%	13.42%	41.04%
3	132	26	106	142	255	**66.7**	**2.22**	66.67%	22.97%	**43.69%**
4	133	19	114	161	369	75.6	1.89	75.59%	33.24%	42.34%
5	132	16	116	177	485	83.1	1.66	83.10%	43.69%	39.40%
6	132	12	120	189	605	88.7	1.48	88.73%	54.50%	34.23%
7	133	5	128	194	733	91.1	1.3	91.08%	66.04%	25.04%
8	132	10	122	204	855	95.8	1.2	95.77%	77.03%	18.75%
9	132	3	129	207	984	97.2	1.08	97.18%	88.65%	8.53%
10	132	6	126	213	1110	100	1	100.00%	100.00%	0.00%
Total	1323	**213**	**1110**							

Model Evaluation Metrics



20% was finalized as the cut-off value due to acceptable sensitivity and maximum KS Statistics value

Problem Identified and Corresponding Resolution

Variables	Problems	Recommendations
Weekly Hours	Tends to be the most impacting factor on attrition. People who working more than 45 hours a week tends to be leaving the company	Ensure better work load distribution and better work life balance
Age Group 50+ & Age Group 36-50	Young crowd, below 35 years tends to be leaving the company faster than the ones above 35 year of age	Recognize the young talents, understand their concerns, reward and promote them as deserved

Problem Identified and Corresponding Resolution

Variables	Problems	Recommendations
Business Travel Rarely & Business Travel Frequently	Employees who travel add to the attrition rate	Try to reduce travel frequency, try to give more opportunities for single people to travel than married ones as they may have family responsibilities, try to provide better or more disturbance allowance in case business travel is required. Try to hire onsite employees for the kind of role employees are travelling the most
Years With Curr Manager	It seems people who spend more time with same managers tend to stay	Try to get a 360 degree feedback on managers whose subordinates are leaving too fast and try to understand their concerns if any

Problem Identified and Corresponding Resolution

Variables	Problems	Recommendations
Years Since Last Promotion	It seems employees who have not been promoted for a while tend to leave	Have a certain promotion criteria which is measurable, Ensure that people are being promoted on fair grounds and when they well deserve it
Environment Satisfaction 2, Environment Satisfaction 3 & Environment Satisfaction 4	While employees who rated 2 and above seems to stay, rating 1 tends to leave the company	look into detailed feedback from people who have rated 1 and take actions accordingly

Problem Identified and Corresponding Resolution

Variables	Problems	Recommendations
Department Research Development & Department Sales	Employees working in Development and Sales department tend to be staying a bit longer	Look into other departments concerns, understand from sales and development what is working in favor of them and apply the same to other departments
Job Satisfaction 4	Rating below 4 tend to be addition on to attrition rate	Look into detailed feedback for below 4 rating and take actions accordingly

Problem Identified and Corresponding Resolution

Variables	Problems	Recommendations
Total Working Years Group 11-40, Total Working Years Group 3-5 & Total Working Years Group 6-10	Employees with less than 3 years experience tend to be leaving faster	Once an employee completed 2.5 years, understand their concerns , their ambitions, career path they want to achieve, expose them to new opportunities, provide them with training, promote them if they are performing well and give a remarkable hire which is not less than 15%
Number of Companies Worked	Rating below 4 tend to be addition on to attrition rate	screen for candidates who have stayed with companies fro longer and not the ones who kept changing companies one after other

Problem Identified and Corresponding Resolution

Variables	Problems	Recommendations
Marital Status Single	Single Employees ted to have less liability and hence may think of changing jobs more frequently.	Like some companies have family days, Have a singles day as well for single employees to interact and know each other. Provide a Marriage bonus which could be one months salary and also a marriage leave of 5 days to appease the single employees to stay in the company until they get married.