

HR ANALYTICS CASE STUDY

SUBMISSION

Group Name:

1. Amit Gupta
2. Raghavendra D
3. Navneet Kishore

Business Understanding

- The company **XYZ** employs at any given point of time, around 4000 employees.
- The attrition rate at XYZ is at 15% which is very high
- The attrition is due to voluntary and involuntary separation.
- The company faces the following problems
 - The former employees' projects get delayed, which makes it difficult to meet **timelines**, resulting in a reputation loss among consumers and partners
 - A sizeable department has to be maintained, for the purposes of recruiting new talent
 - More often than not, the new employees have to be trained for the job and/or given time to acclimatise themselves to the company

AIM of the Analytics Project

- To find the possible reasons for high attrition
- Suggest changes to the current setup to make employee stay
- Variables that need immediate attention by the management

Data Understanding

Dataset	Description	Details
employee_survey_data.csv	Responses from all the employees of an internal satisfaction survey	4410 observations with 4 variables
general_data.csv	All the details of the employees of the company	4410 observations with 24 variables
manager_survey_data.csv	Responses from the managers on the performance of all the employees of the company	4410 observations with 3 variables
in_time.csv	In-Punch time of all the employees for the past year (262 working days)	4410 observations with 262 variables
out_time.csv	Out-Punch time of all the employees for the past year (262 working days)	4410 observations with 262 variables

Data Preparation

- Import the data from the data files and check if EmployeeId is the key
- Convert the in_time and out_time data frames data into date time datatype and rename the headers
- Combine the in_time and out_time data frames and find the difference between in_time and out_time for each day.

deskTime=mean(in_time - out_time) across 262 days

- Average the working hours for each employee across 262 days to get the **“deskTime” derived metric.**
- Combine the desktime with other dataframes to get the final **“employeeData”** data frame.

Data Preparation Contd.,

- Drop the following columns as the data across all the employees is the same:
 - Over18
 - StandardHours
 - EmployeeCount
- Convert the following categorical attributes mapped in numerical variables to categories to get the final dataset
 - Education
 - JobLevel
 - StockOptionLevel
 - EnvironmentSatisfaction
 - JobSatisfaction
 - WorkLifeBalance
 - JobInvolvement
 - PerformanceRating

Exploratory Data Analysis – Categorical Attributes

Upon plotting various categorical features with the attrition the following categories seems to have more attrition

- Travel Rarely
- Education Level - Bachelor
- Research and Development Department
- Lifescience and Medical Field
- Male Employees
- Job level 1 & 2
- Research Scientist, Sales & Executive Job Role
- Marital Status – Single
- Stock Option Level 0 & 1
- All environment satisfaction level except medium
- Low and high Job Satisfaction Levels
- Better work-life balance
- High Job Involvement
- Low Performance Rating

Exploratory Data Analysis – Quantitative Attributes

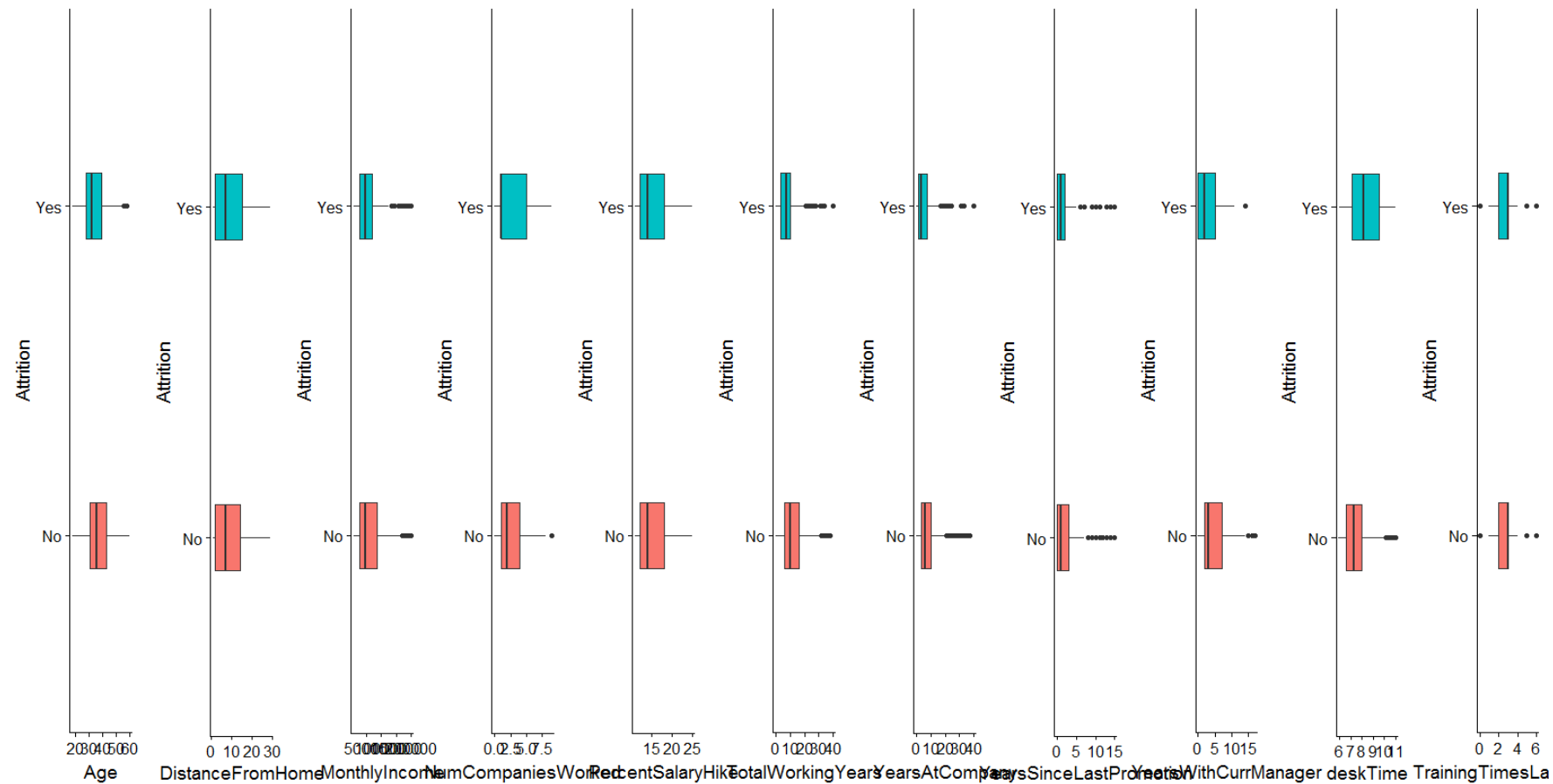
Upon plotting boxplots for various quantitative features with the attrition the following categories seems to have more attrition

- Age
- MonthlyIncome
- NumCompaniesWorked
- PercentSalaryHike
- TotalWorkingYears
- YearsAtCompany
- YearsSinceLastPromotion
- YearsWithCurrManager
- deskTime
- TrainingTimesLastYear

Data Preparation

- Outlier treatment - All the outliers $<1\%$ and $>95\%$ are pulled back to 1% and 95% respectively to avoid any skewing of the results.
- Missing Value Treatment – All the categorical missing values were removed as the % of missing values were at 0.025
- Scaling – All the quantitative variables were scaled to avoid skewing
- Converting categorical values into numeric representation to make it ready for logistic regression
- Train and test set creation was done taking 70% of the data as train data and 25% as test data

Plot – Check for outlier in Quantitative Attributes



Logistic Regression – Model Building

- Attrition rate in the given data was found to be 16.2%
- The final data set ready for Logistic regression had the following statistics.
 - Total: 4327 obs. of 57 variables
 - Train Set: 3029 obs. of 57 variables
 - Test Set: 1298 obs. of 57 variables
- The logistic regression run on the train data provided the following stats
 - AIC: 2108.3....57 coeff..nullDev 2684.5...resDev 1994.3
- Step AIC method was used to perform step wise model selection and the final model of AIC had 36 co-efficients.

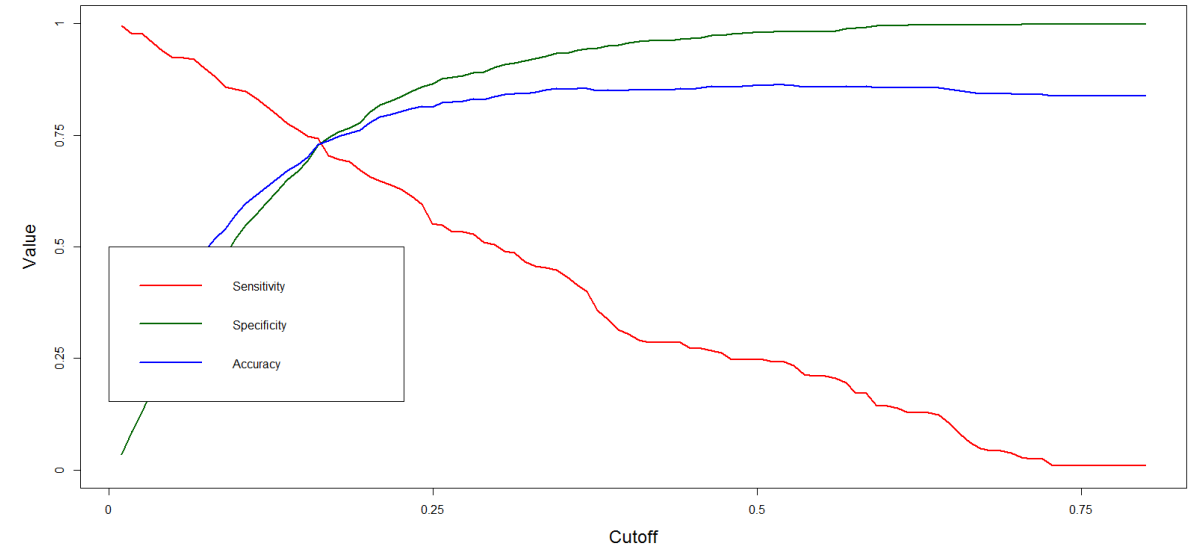
Logistic Regression – Model Building Contd...

- The logistic regression continued to provide the final model with the following 9 important attributes

Coefficients:	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.92434	0.10759	-27.181	< 2e-16	***
deskTime	0.5634	0.05245	10.741	< 2e-16	***
MaritalStatus.xSingle	1.03772	0.11074	9.371	< 2e-16	***
TotalWorkingYears	-0.732	0.09062	-8.078	6.60E-16	***
EnvironmentSatisfaction.xLow	0.91605	0.12718	7.203	5.90E-13	***
YearsSinceLastPromotion	0.52852	0.07515	7.033	2.03E-12	***
JobSatisfaction.xLow	0.79316	0.12465	6.363	1.98E-10	***
YearsWithCurrManager	-0.50664	0.08459	-5.99	2.10E-09	***
BusinessTravel.xTravel_Frequently	0.73601	0.12693	5.798	6.70E-09	***
NumCompaniesWorked	0.25797	0.05804	4.445	8.80E-06	***

Model Evaluation

- Looking at the graph the point where the three line graph meet is the optimal cut off value where the sensitivity, specificity and accuracy are well balanced in the model.
- Hence we choose cutoff value of 0.2334 for final model
- The final accuracy, specificity and sensitivity of the model are as follows:
 - accuracy 79.66%
 - sensitivity 63.33%
 - specificity 82.81%



Model Evaluation Contd.,

- The KS Statistic for the model is 0.4607931
- The following are the lift and gain charts for the model
- It shows that if we sort all the employees according to the probability we among the top 40% of the employees this model can identify 76% of the employees that are bound to leave the company.
- The model is outperforming the random model by 2.2 time by the 3rd decile.

bucket	total	totalresp	Cumresp	Gain	Cumlift
1	130	73	73	34.7619	3.47619
2	130	43	116	55.2381	2.761905
3	130	25	141	67.14286	2.238095
4	130	19	160	76.19048	1.904762
5	129	15	175	83.33333	1.666667
6	130	9	184	87.61905	1.460317
7	130	9	193	91.90476	1.312925
8	130	4	197	93.80952	1.172619
9	130	8	205	97.61905	1.084656
10	129	5	210	100	1

Conclusions / Recommendations

Top Factors that need immediate attention:

1	Desk Time	The employees that are most likely to move out are the one who seem to be working overtime in the organization. Need to do something to reduce their workload.
2	Single Employees	Unmarried employees are more likely to change jobs. Need to look deeper into the reasons that they are giving while quitting to address this issue. Better carrier prospects and pay might be one of the reasons.
3	Work Experience	Employees will less experience are more likely to change jobs as compared to the experienced ones. This is again tied to the fact that the junior employees are often unmarried.
4	Low Environment Satisfaction	There seems to be something not right with the work environment could be both physical and team. Further digging into these aspects could help retain people.
5	Years Since Last Promotion	People who haven't been promoted for a long time are likely to move out. Look at each of them on case by case basis and see what can be done to retain them.

Secondary Factors

These are some of the other important factors that needs to be looked into:

6	Low Job Satisfaction	There are employees with very low job satisfaction. This means that they are not feeling rewarded for the work that they are doing for the company. Additional surveys need to be done to understand the concerns and solve them.
7	Years with Current Manager	Employees are leaving the company if they are not able to set a good working relationship with their manager. If they are with a manager for a longer time they are not leaving the company. Look for some training to managers to make them more approachable.
8	Frequent Business Travels	Employees who travel frequently are leaving the company due to excessive travel. Try to provide a good travel-work-life balance.
9	Number of Jobs changed	Frequent job hoppers are the ones who will change the job quickly. Look at the hiring process and try hire less people who frequently switch jobs