



HR Analytics Case Study

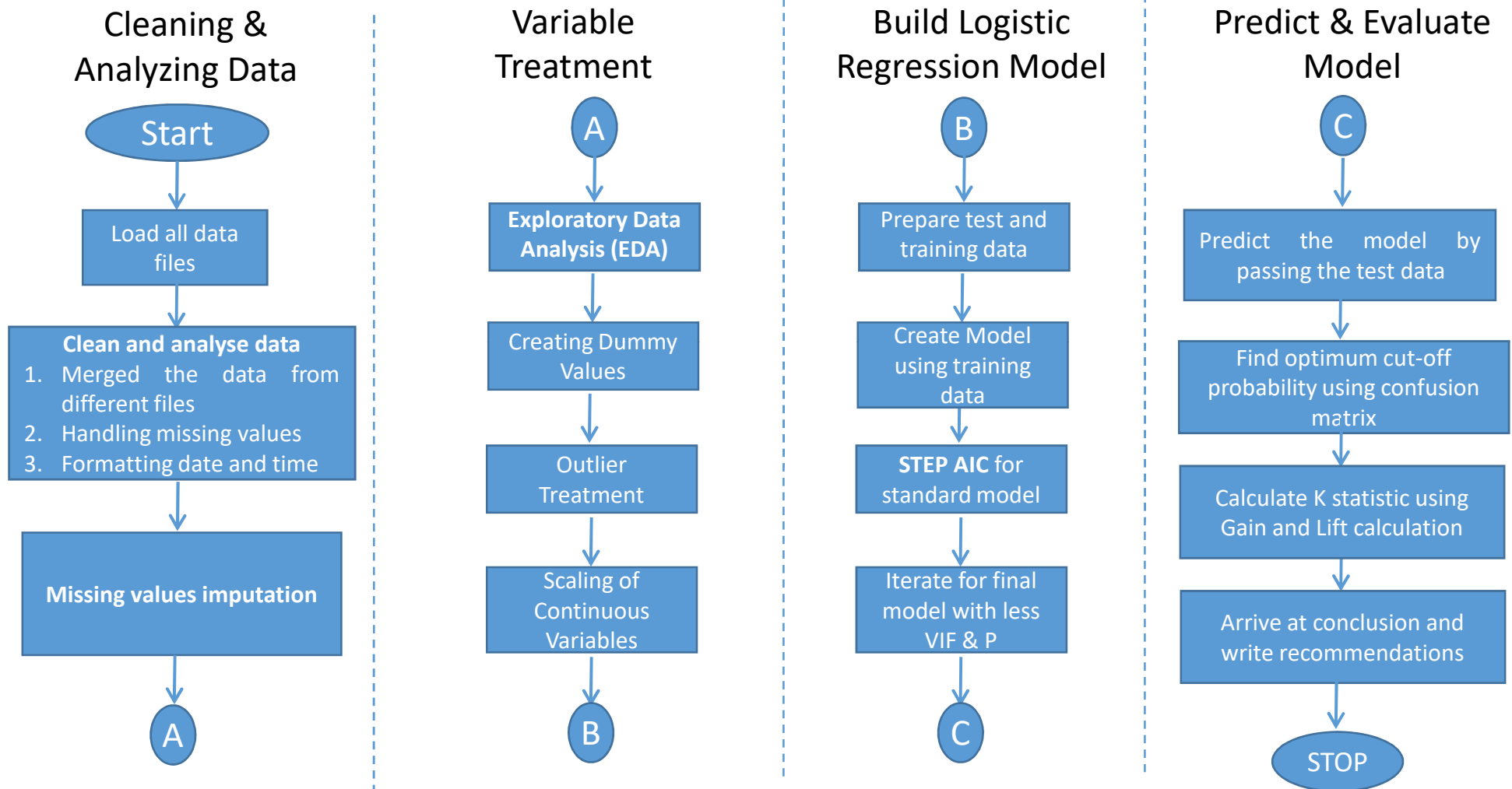
Problem Statement:

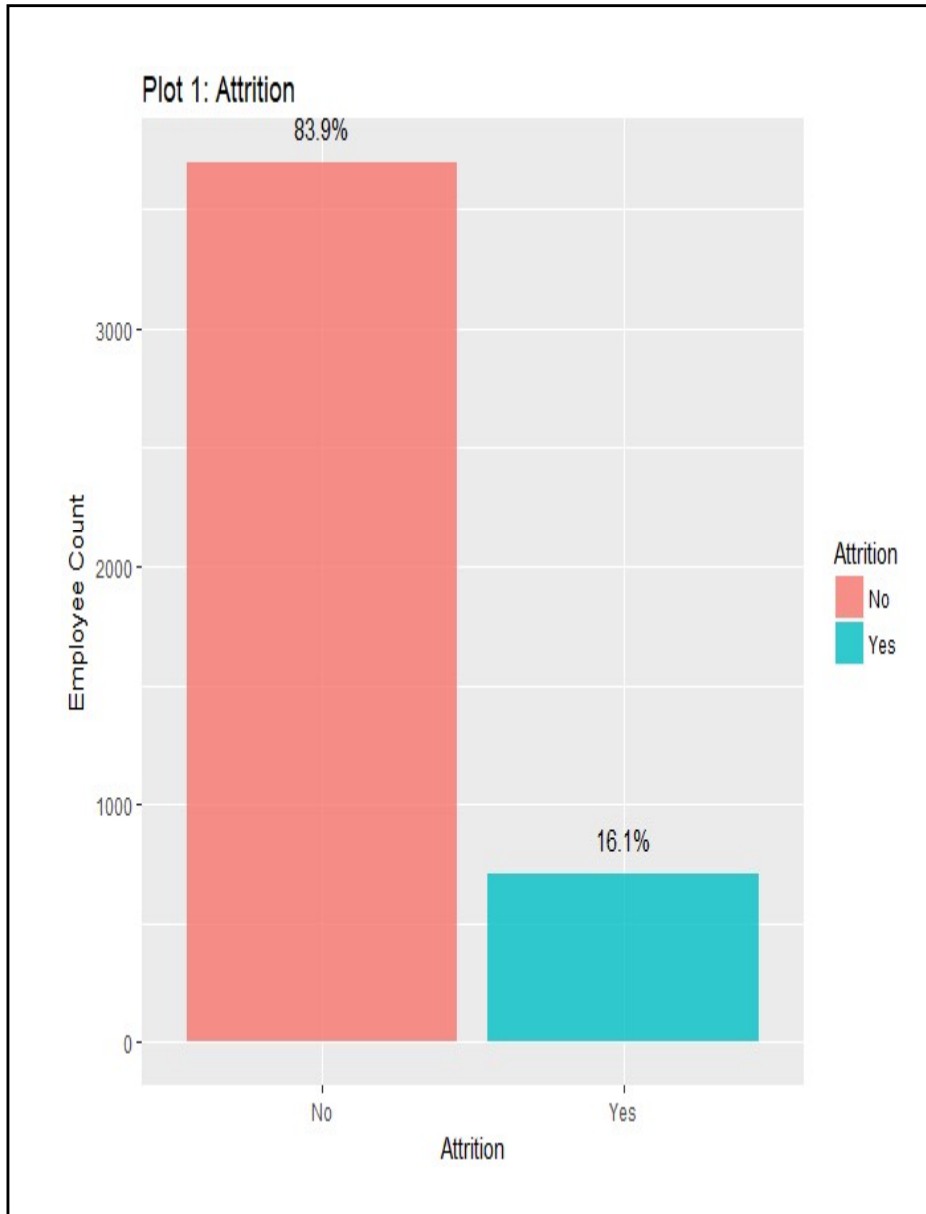
Every year, around 15% of employees leave the **XYZ** company and need to be replaced with the talent pool available in the job market. The management believes that this level of is bad for the company, because of the following reasons:

- Project delays
- Missing Deadlines and loss of business
- Loss of reputation among accounts
- Increased new recruitment cost and high HR department overheads
- Further additional Training and Development expenses
- Negative motivation levels among current employees

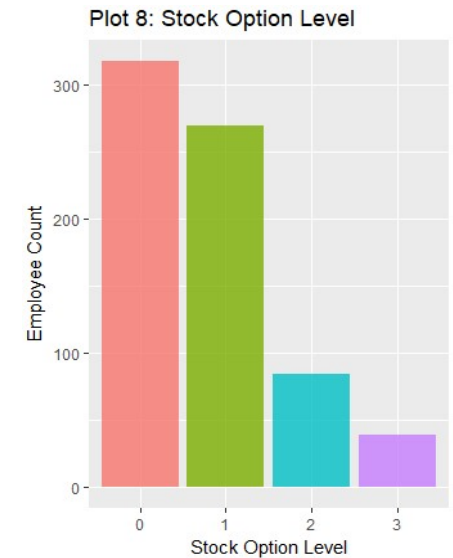
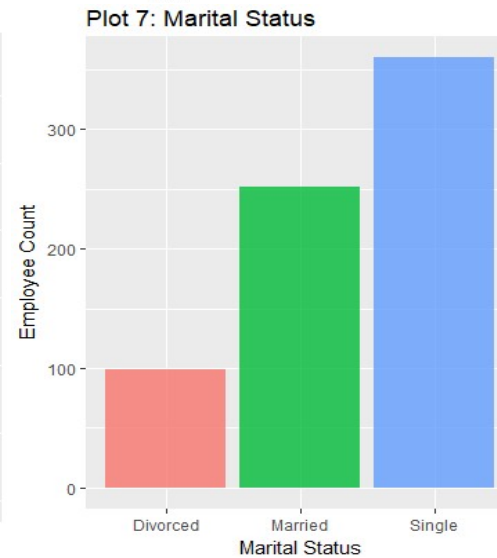
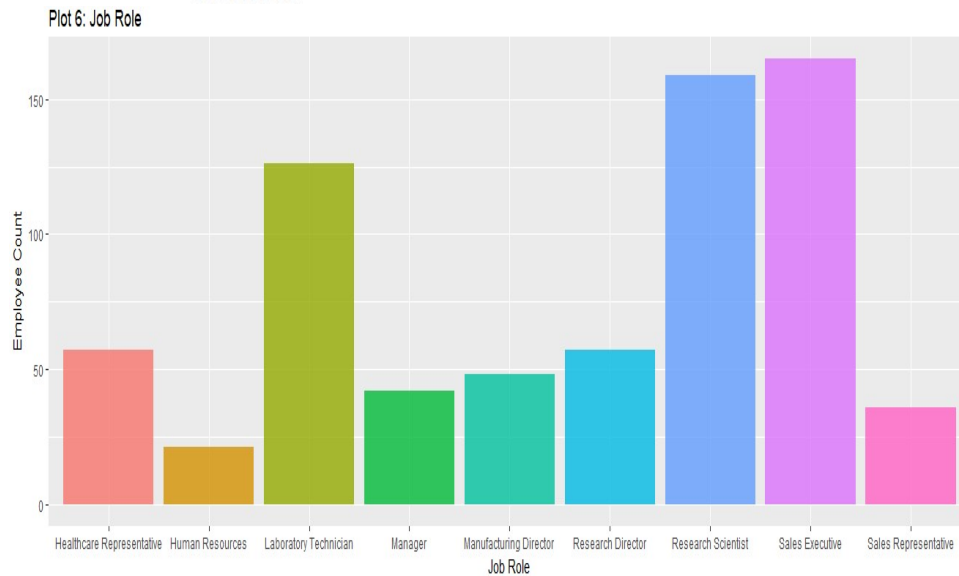
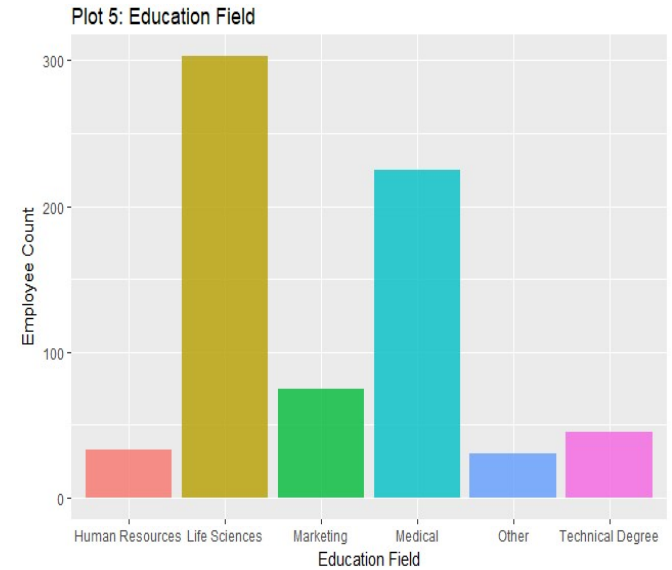
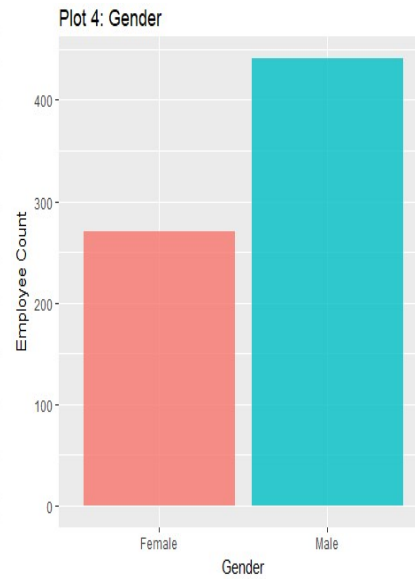
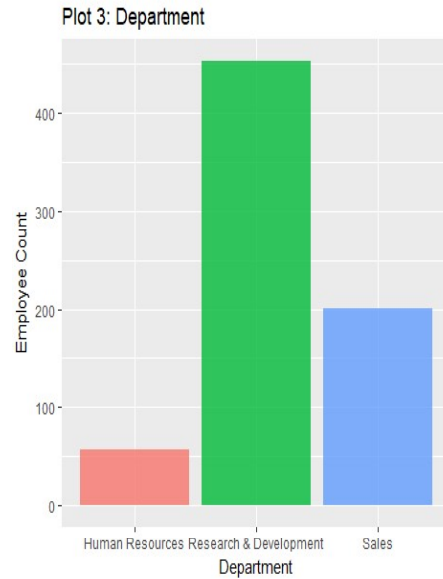
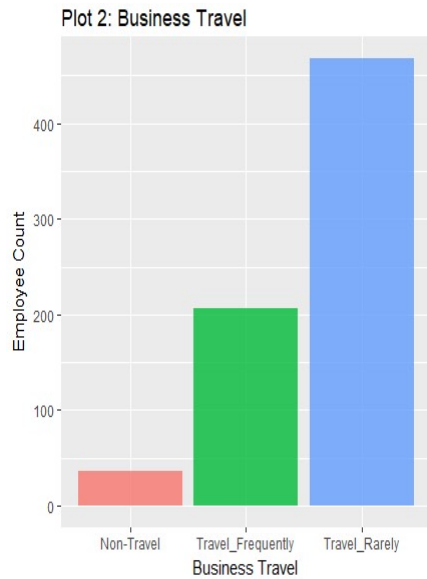
Objective:

HR analytics firm need to understand what factors **XYZ** should focus on, in order to curb attrition. In other words, they want to know what changes they should make to their workplace, in order to get most of their employees to stay. Also, they want to know which of these variables is most important and needs to be addressed right away.

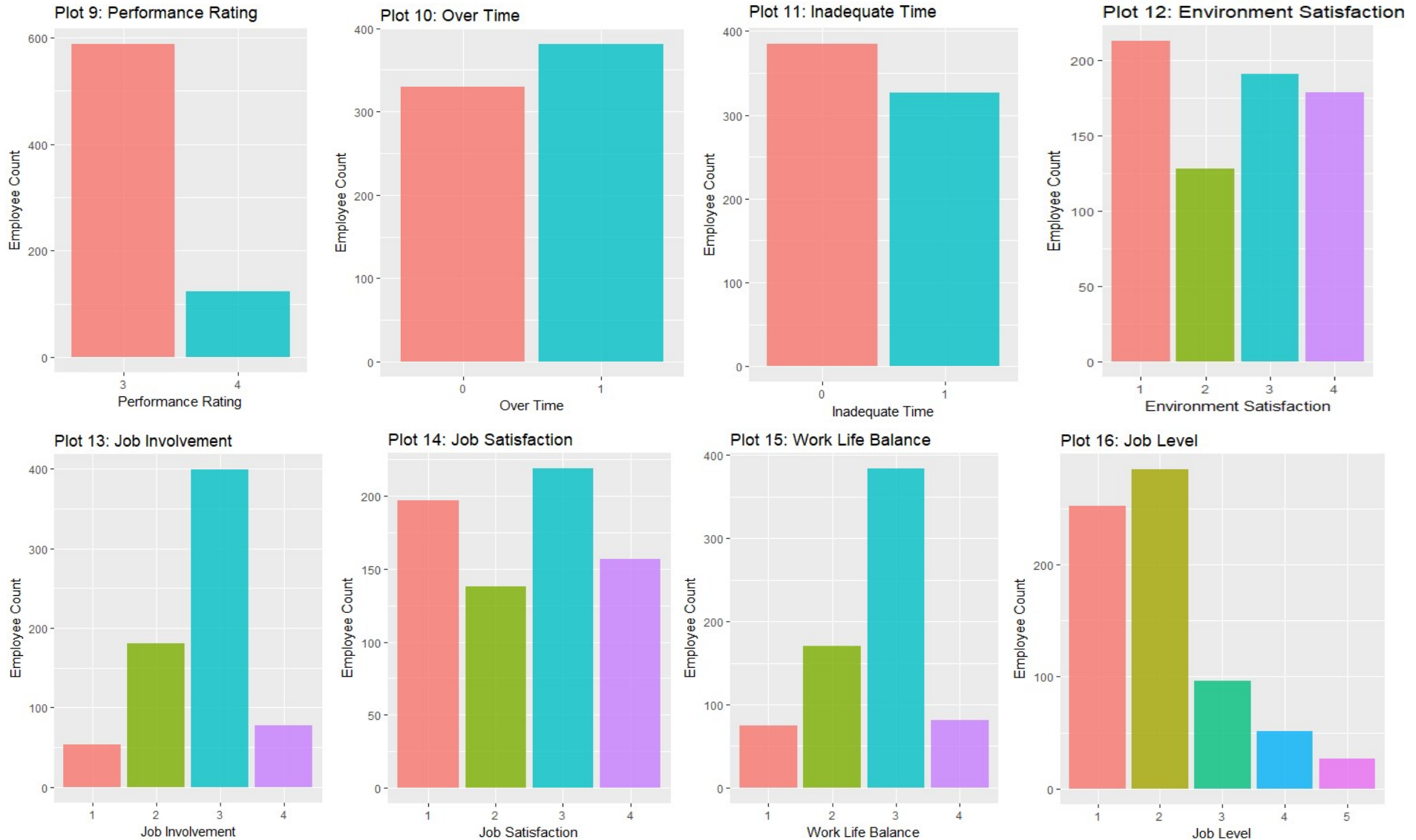




- Total Number of Employees: 4400+
- Attrition found: ~16.1%



Graphs are plotted for all the variables against number of employees who left (Attrition = Yes)



Graphs are plotted for all the variables against number of employees who left (Attrition = Yes)



Exploratory Data Analysis Cont...

UpGrad

Attrition is higher for employees having below characteristics individually:

- **Business Travel:** Rare
- **Department:** Research and Development
- **Gender:** Male
- **Education Field:** Life Sciences
- **Job Role:** Sales Executive, Research Scientist
- **Marital Status:** Single
- **Stock Option level:** 0
- **Performance Rating:** 3
- **Over Time:** Not Applicable (Same for all)
- **Inadequate time:** Not Applicable
- **Environment Satisfaction:** 1
- **Job Involvement:** 3
- **Job Satisfaction:** 3
- **Work Life Balance:** 3
- **Job Level:** 2 & 1

EDA resulted in multiple variables, however Logistic regression need to be carried out to find exact factors responsible for attrition

- Dummy variables are created for factor variables and Continuous variables are normalized and scaled
- Logistic regression model is developed for dependent variable “Attrition” with GLM function, then Step AIC has been applied to arrived at standard model.
- Based on VIF (variance inflation factor) and P value (with significance) predictors have been filtered and after another 21 iterations we could achieve our final model.

Final Logistic regression model is as below:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.43190	0.17933	-7.985	1.41e-15	***
Age	-0.59469	0.06684	-8.897	< 2e-16	***
NumCompaniesworked	0.35389	0.05785	6.117	9.53e-10	***
TrainingTimesLastYear	-0.19006	0.05815	-3.268	0.001082	**
YearsSinceLastPromotion	0.50020	0.06974	7.173	7.35e-13	***
YearswithCurrManager	-0.67075	0.07969	-8.417	< 2e-16	***
Over_time	1.69253	0.11698	14.468	< 2e-16	***
Environmentsatisfaction.x2	-0.66228	0.16841	-3.932	8.41e-05	***
Environmentsatisfaction.x3	-0.83000	0.15541	-5.341	9.26e-08	***
Environmentsatisfaction.x4	-1.21222	0.16132	-7.514	5.72e-14	***
Jobsatisfaction.x2	-0.43633	0.16891	-2.583	0.009788	**
Jobsatisfaction.x3	-0.45742	0.15064	-3.037	0.002393	**
Jobsatisfaction.x4	-1.22232	0.16407	-7.450	9.34e-14	***
workLifeBalance.x3	-0.39937	0.11276	-3.542	0.000397	***
BusinessTravel.xTravel_Frequently	0.80588	0.13074	6.164	7.10e-10	***
EducationField.xTechnical.Degree	-0.46412	0.21514	-2.157	0.030978	*
JobRole.xManufacturing.Director	-0.85512	0.22497	-3.801	0.000144	***
MaritalStatus.xSingle	0.93864	0.11597	8.094	5.78e-16	***

- Test data set is passed to final model.
- Confusion matrix was created for finding Accuracy, Sensitivity and Specificity
- Optimum cut-off probability was found to be ~**0.186**

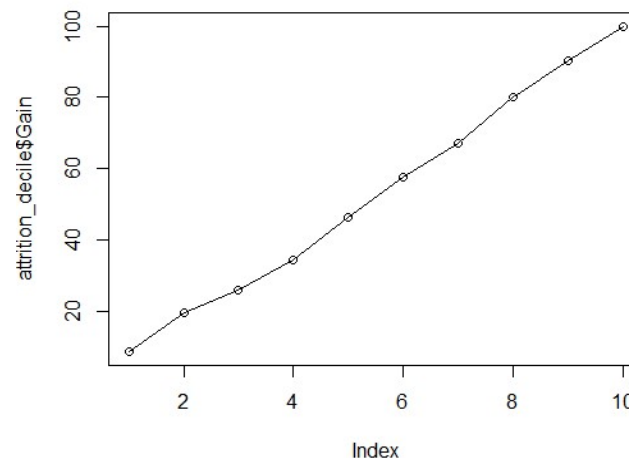
Final Cut off Matrix

Confusion Matrix and Statistics

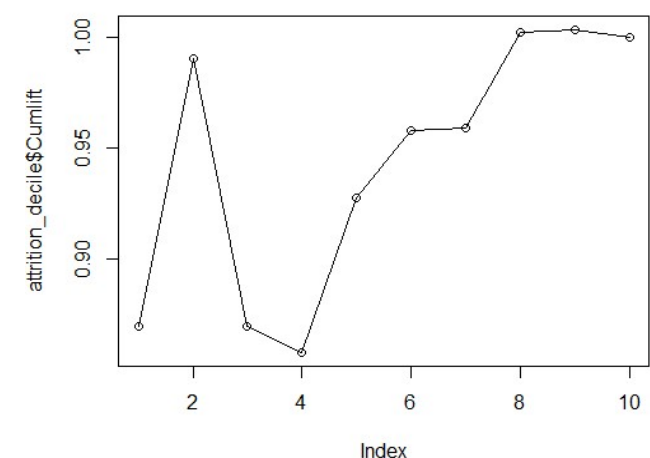
	Reference	No	Yes
Prediction	No	835	51
Yes	281	156	

Accuracy : 0.7491
 95% CI : (0.7248, 0.7722)
 No Information Rate : 0.8435
 P-Value [Acc > NIR] : 1
 Kappa : 0.3455
 Mcnemar's Test P-Value : <2e-16
 Sensitivity : 0.7536
 Specificity : 0.7482
 Pos Pred Value : 0.3570
 Neg Pred Value : 0.9424
 Prevalence : 0.1565
 Detection Rate : 0.1179
 Detection Prevalence : 0.3303
 Balanced Accuracy : 0.7509
 'Positive' Class : Yes

Gain Plot



Cumulative Lift Plot



- The model has an increasing Gain and a decreasing Lift.
- The **KS statistic (0.5018)** shows that the model is very good in distinguishing between employees who will leave the company and employees who won't.

Driving Forces for attrition:

- Employees exhibiting below characteristics tend to leave the organization:
 - More over working times
 - Unmarried
 - Promotion due for longer time
 - Worked in more number of companies
 - Frequent Business travel

- Employees with below characteristics are less likely to leave the company
 - High AGE
 - High experience
 - Having same manage for a longer period of time
 - High Environment Satisfaction, Job Satisfaction and Work life balance

Recommendations:

We recommend XYZ to implement strategies and implement key practices to:

- Improve Environment Satisfaction, Job Satisfaction and Work life balance
- Reduce working overtime
- Reduce number of business travels