

XYZ Company

Employee Attrition Prediction

HR Analytics Case Study

Sailendra Kona

Swati Bansal

Pradosh Kumar Jena

Rituraj Achuthan

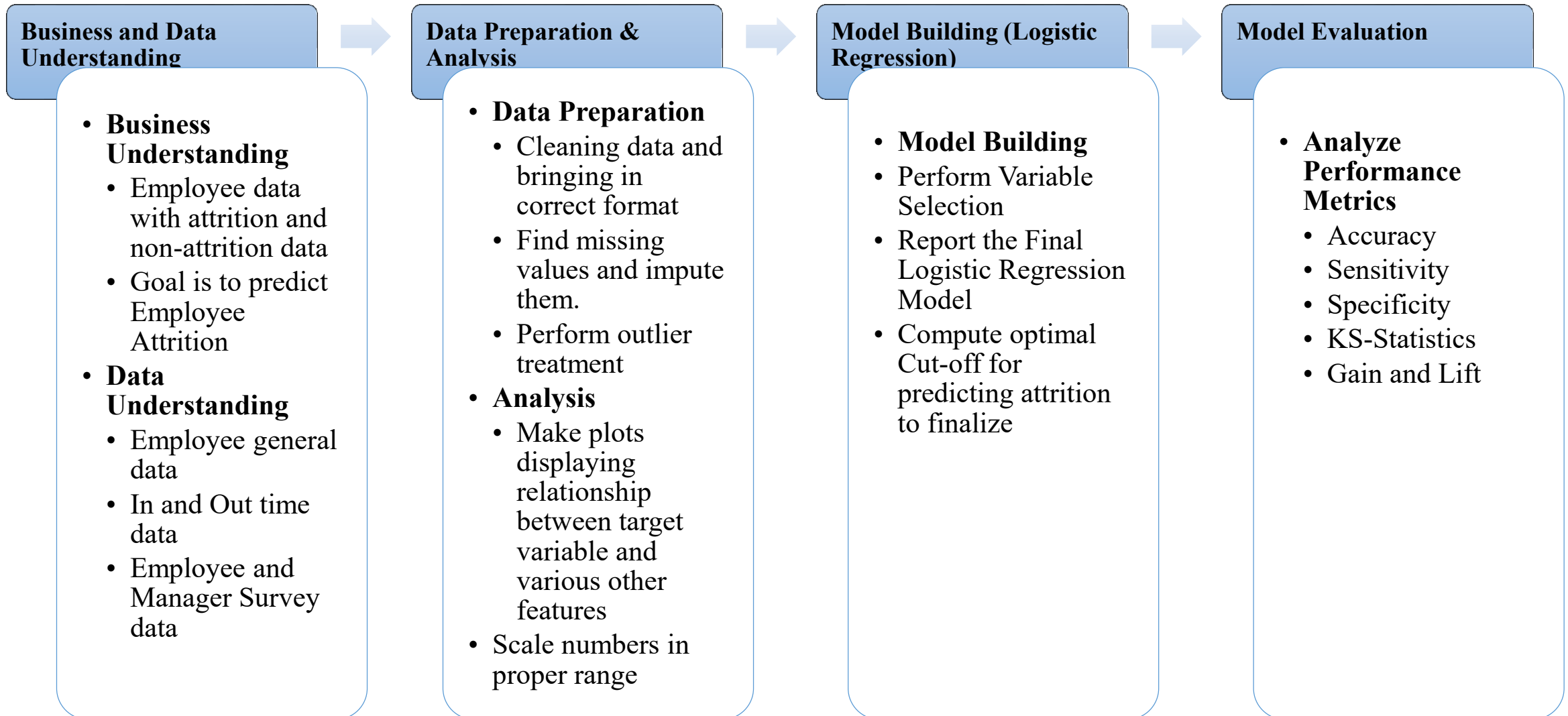
Problem Statement and Goal Objective

Problem Statement

- XYZ Company has 4000+ employees and has **15% of Employee attrition** each year.
- Attrition causing following negative effect
 - Delay in Project Deadlines.
 - Maintenance cost of sizeable Recruiting Team
 - Training of New employees results in further delay of projects

Goal of Case Study

- Predict Employee Attrition with Probability using Logistic Regression model.
- Help Company XYZ in understanding changes needed at workplace for Employee Retention



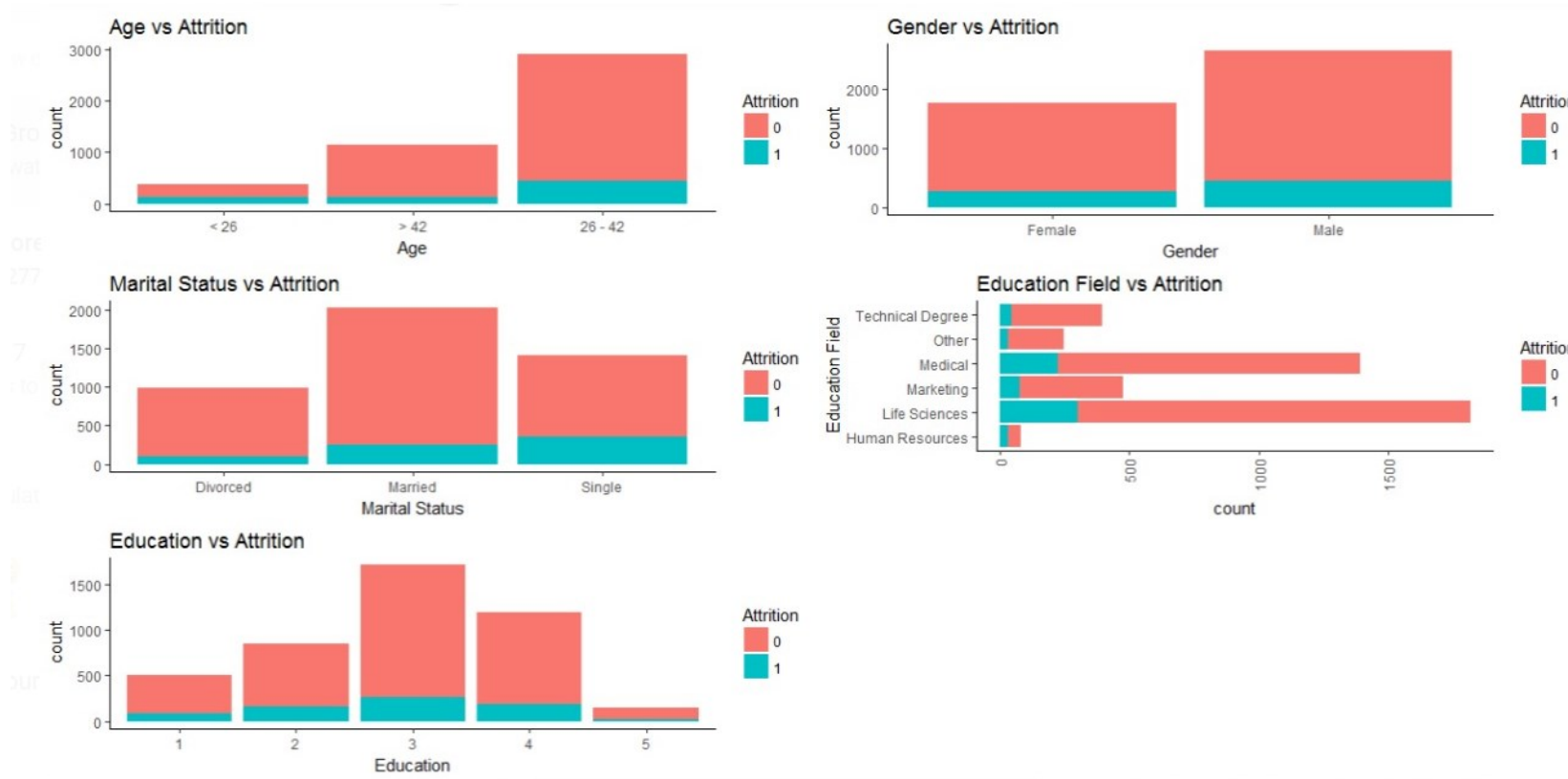
Understanding the Given Data

- **Manager Survey Data** - Manager ratings for Employee involvement in the job and Employee Performance Rating.
- **Employee Survey Data** - Contains data about whether the employee is satisfied with the Job, Office Environment and Work Life Balance.
- **In-Time and Out-Time Data** — These two datasets ranging from 1/1/2015 to 12/31/2015.
- **General Data** — Contains employee wise personal information along with education and basic in-company data.

Assumptions

- Dates having all NA values in both In-Time and Out-Time to be considered as a Holidays
- NA values on specific Dates are considered as Leaves taken

Exploratory Data Analysis



- **Important Factors that have Significance**
 - Age group 26-42
 - Married Employees has highest attrition
 - Education Levels 3 and 4
 - Departments Life Sciences and Medical have attrition
- **Factors with no Impact**
 - Gender has no effect

Exploratory Data Analysis (*Contd ...*)

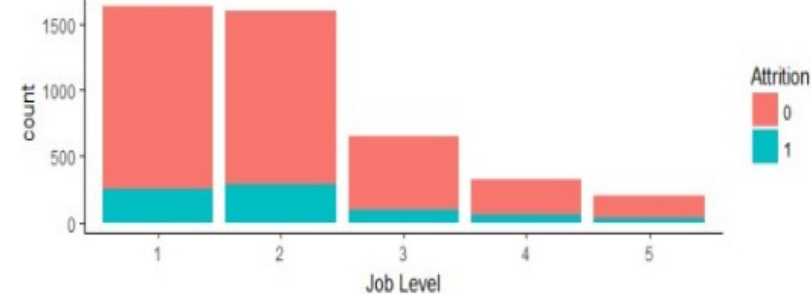
Business Travel vs Attrition



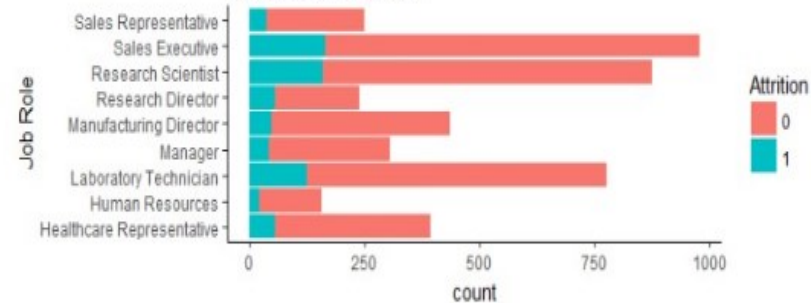
Department vs Attrition



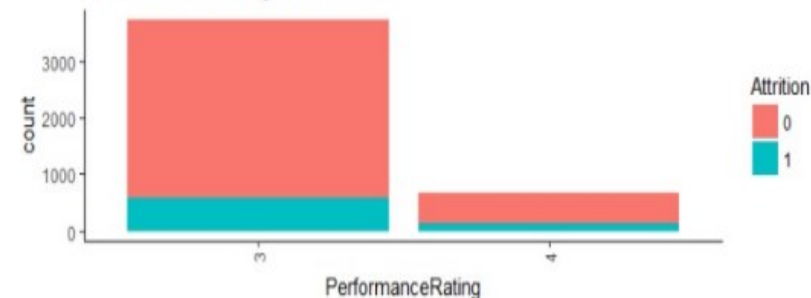
Job Level vs Attrition



Job Role vs Attrition



PerformanceRating vs Attrition



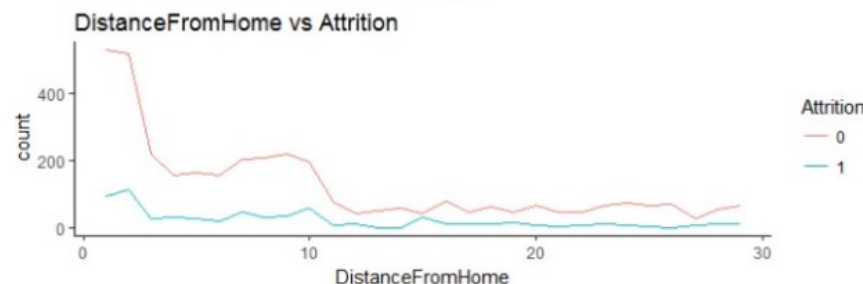
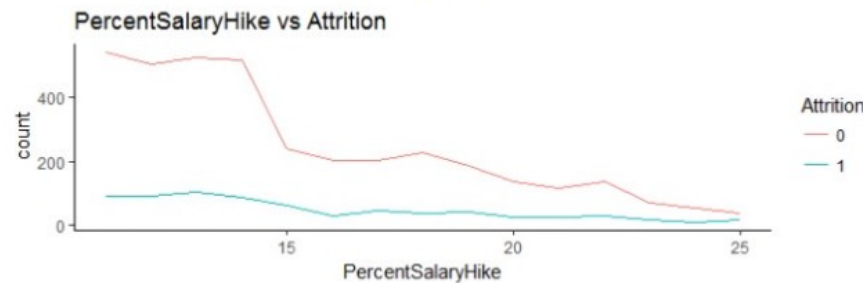
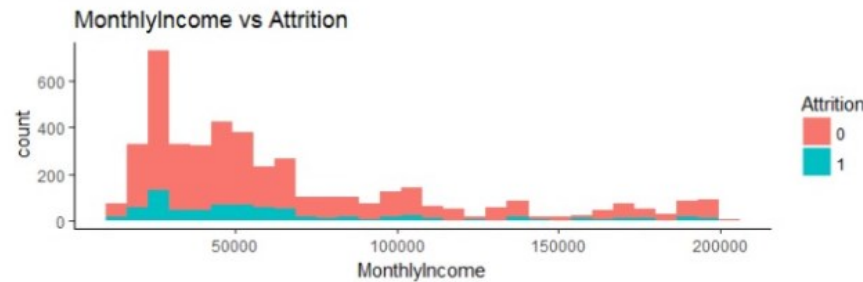
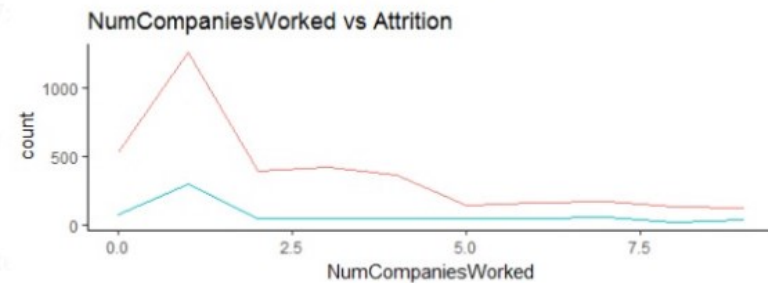
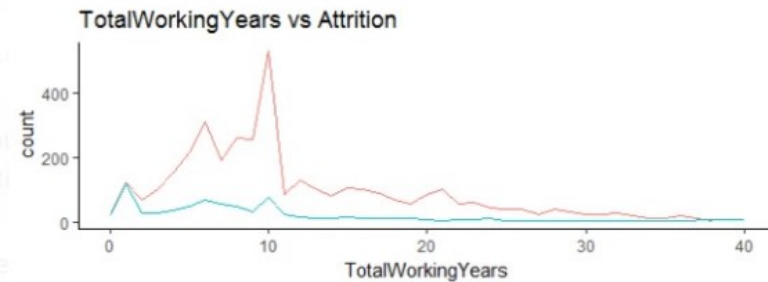
- **Important Factors that have Significance**
 - Employee Travel Rarely leave most
 - Job Levels 1 and 2 highest
- **Factors with no Impact**
 - Department has no effect, as attrition is there proportionately in all.
 - Job Role has no clear indication on attrition
 - Performance Rating has no clear indication on attrition

Exploratory Data Analysis (*Contd ...*)



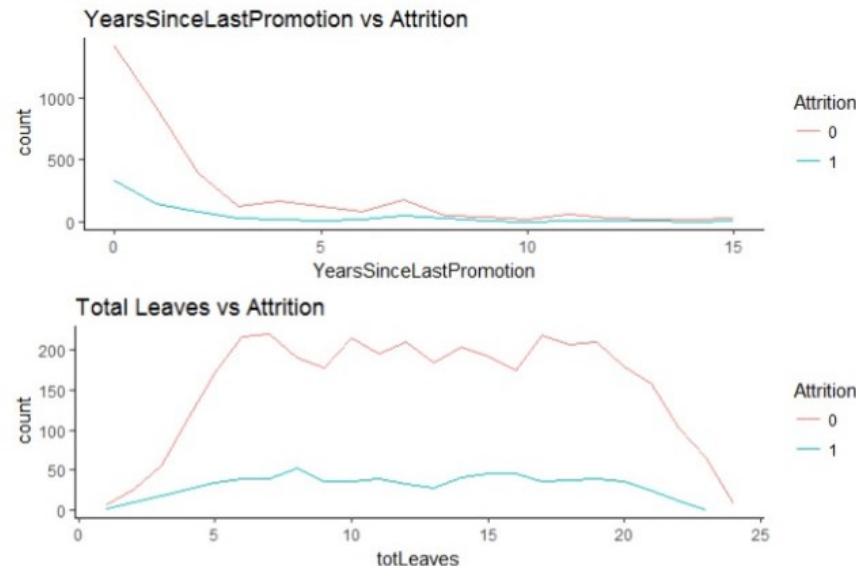
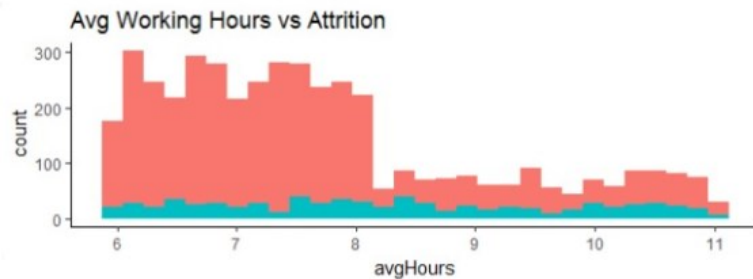
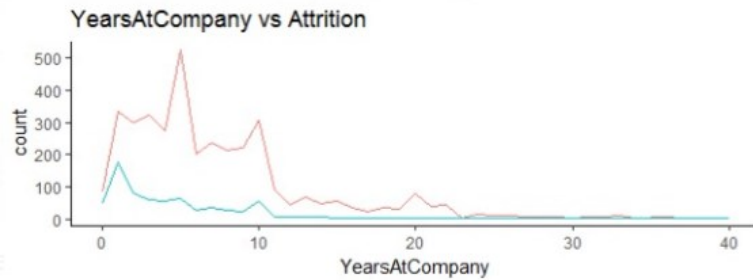
- Except **Stock Option level**, none of these factors have no logical explanation for attrition.
- The higher values for other factors here should have less attrition, but they are not.

Exploratory Data Analysis (*Contd ...*)



- **Important Factors that have Significance**
 - Total Working Years, Number of Companies worked before and Training taken last Year
- **Factors with no Impact**
 - Salary Hike
 - Distance from Home
 - Monthly Income

Exploratory Data Analysis (*Contd ...*)



- **Important Factors that have Significance**
 - Years at Company
 - Years Since last Promotion
 - Average Working Hours
 - Years with Current Manager
- **Factors with no Impact**
 - Number of Leaves

Recommended Logistic Regression Model

- Logistic model built using *glm()* and *stepAIC()* function and remove insignificant variables using *p-value*, *VIF* values, and *Correlation*.
- Final model has 10 significant variables. ***Intercept = -1.95090***

Significant Variables in Final Model	Coefficients
NumCompaniesWorked	0.31731
TotalWorkingYears	-0.67727
YearsSinceLastPromotion	0.43492
YearsWithCurrManager	-0.43509
avgHours	0.51084
BusinessTravel.xTravel_Frequently	0.74735
MaritalStatus.xSingle	1.03816
EnvironmentSatisfaction.x3	-0.48875
EnvironmentSatisfaction.x4	-0.73123
JobSatisfaction.x4	-0.77292

- The Model explains critical driving factors for Employee Attrition that match with insights as per Exploratory Data Analysis
- Higher the values for Number of years with current Manager and also Total Working Years reduce Employee attrition
- A model without Environment Satisfaction and Job Satisfaction is also performing well, but considered in model from a Business Perspective

Logistic Regression Model Performance

Confusion Matrix

Prediction	Reference	
	No	Yes
No	1063	146
Yes	47	67

Cut-off Probability value for deciding the Attrition = 0.17 or 17%

Prediction Statistics	
AIC	2284.8
Accuracy	0.739229
Sensitivity	0.741784
Specificity	0.7387387
KS Statistics	0.4805228

- Cut-Off **17%** indicates as multiple factors affecting Employee attrition, lower value is need for best prediction
- *KSStatistic* **48%** is pretty good with much higher than standard expected **40%**

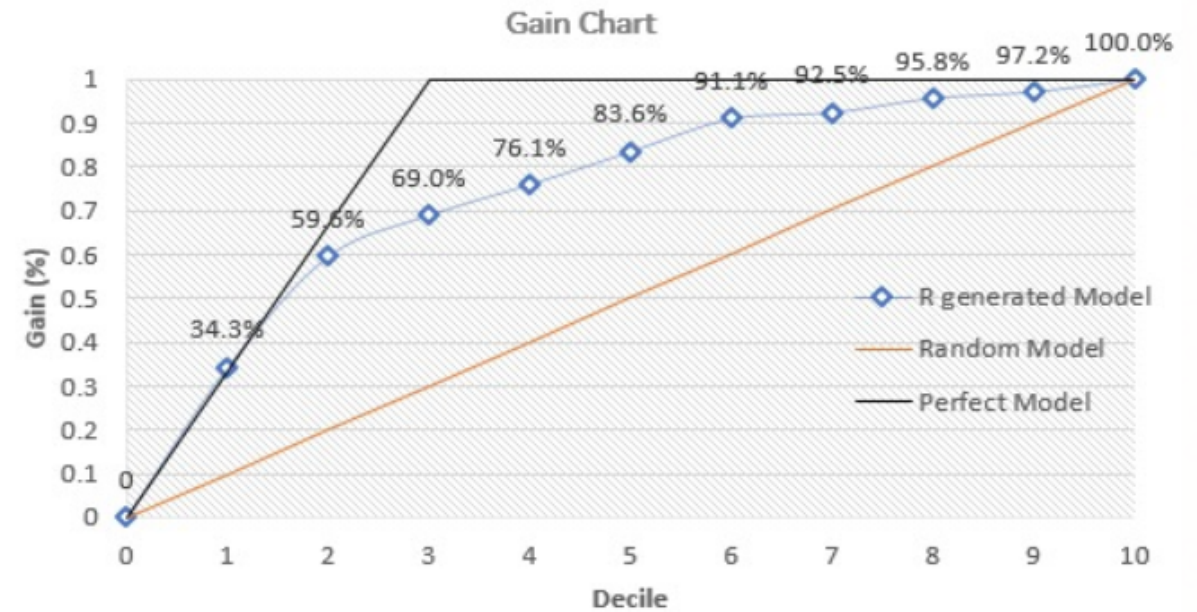
Model Assessment – KS Statistics

Decile	Observations	Attrition	Cum-Attrition	% Cum-Attrition	Non-Attrition	Cum-Non-Attrition	%Cum-Non-Attrition	(%Cum-Attrition) - (%Cum-Non-Attrition)
1	133	73	73	34.3%	60	60	5.4%	28.9%
2	132	54	127	59.6%	78	138	12.4%	47.2%
3	132	20	147	69.0%	112	250	22.5%	46.5%
4	133	15	162	76.1%	118	368	33.2%	42.9%
5	132	16	178	83.6%	116	484	43.6%	40.0%
6	132	16	194	91.1%	116	600	54.1%	37.0%
7	133	3	197	92.5%	130	730	65.8%	26.7%
8	132	7	204	95.8%	125	855	77.0%	18.7%
9	132	3	207	97.2%	129	984	88.6%	8.5%
10	132	6	213	100.0%	126	1110	100.0%	0.0%
Total	1323	213			1110			

- KS Statistics is falling at the **2nd decile** and the value is **47.2%**

Model Assessment – Gain

Gain Chart				
Decile	Observations	Attrition	Cum- Attrition	Gain(%Cum-Attrition)
1	133	73	73	34.3%
2	132	54	127	59.6%
3	132	20	147	69.0%
4	133	15	162	76.1%
5	132	16	178	83.6%
6	132	16	194	91.1%
7	133	3	197	92.5%
8	132	7	204	95.8%
9	132	3	207	97.2%
10	132	6	213	100.0%
Total	1323	213		

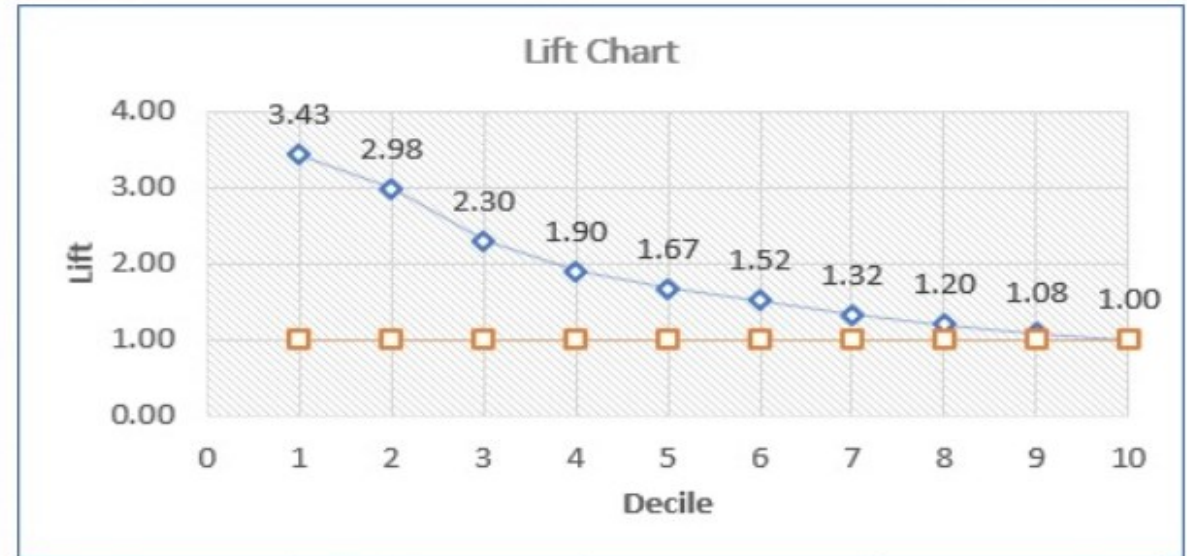


- **76%** Gain achieved with 4th decile
- **83%** Gain achieved with 5th decile

- **Model** is close to perfect model & far away from Random model

Model Assessment –Lift

Lift Chart						
Decile	Observations	Attrition	Cum- Attrition	Gain(%Cum-Attrition)	Gain (Random Model)	Lift
1	133	73	73	34.3%	10%	3.43
2	132	54	127	59.6%	20%	2.98
3	132	20	147	69.0%	30%	2.30
4	133	15	162	76.1%	40%	1.90
5	132	16	178	83.6%	50%	1.67
6	132	16	194	91.1%	60%	1.52
7	133	3	197	92.5%	70%	1.32
8	132	7	204	95.8%	80%	1.20
9	132	3	207	97.2%	90%	1.08
10	132	6	213	100.0%	100%	1.00
Total	1323	213				



- **1.9** Lift achieved with 4th decile
- **1.67** lift achieved with 5th decile

- **Model** achieves **1.9** times more attrition than the Random model at 4th decile.

Factors for Employee Attrition and Suggestions for Retention

- **Years since last promotion**
 - **0 to 2** years is a cause of concern. **40%** People leave within 1 Year after the promotion.
- **Years with Current Manager**
 - People who stay with same manager for longer than **3 Years** have high retention
- **Average working hours**
 - People who are working **8 hours and more** are leaving mostly
- **Business Travel**
 - Employees who **travel rarely** leaves 50% of the time.
- **Environment Satisfaction and Job Satisfaction**
 - As per data available, these two surveys not have clear indication on the attrition
 - However, as these critical factors and retained in model
- **Number of Companies Work Before**
 - People who join this company as their second company leave the most.
- **Marital Status**
 - Employees with Status Single leave most
- **Total Working Years**
 - Retaining employees more than first 3 years reduces the attrition