HR ANALYTICS CASE STUDY

Group Name:

1. Abinash Panda
2. Lipsa Satapathy
3. Prabhudatta Praharaj
4. Ankit Gupta

Business Objective

Company Background

- XYZ is a large company with an average employee strength of 4000 employees .
- The average attrition rate for XYZ is around 15% which is high

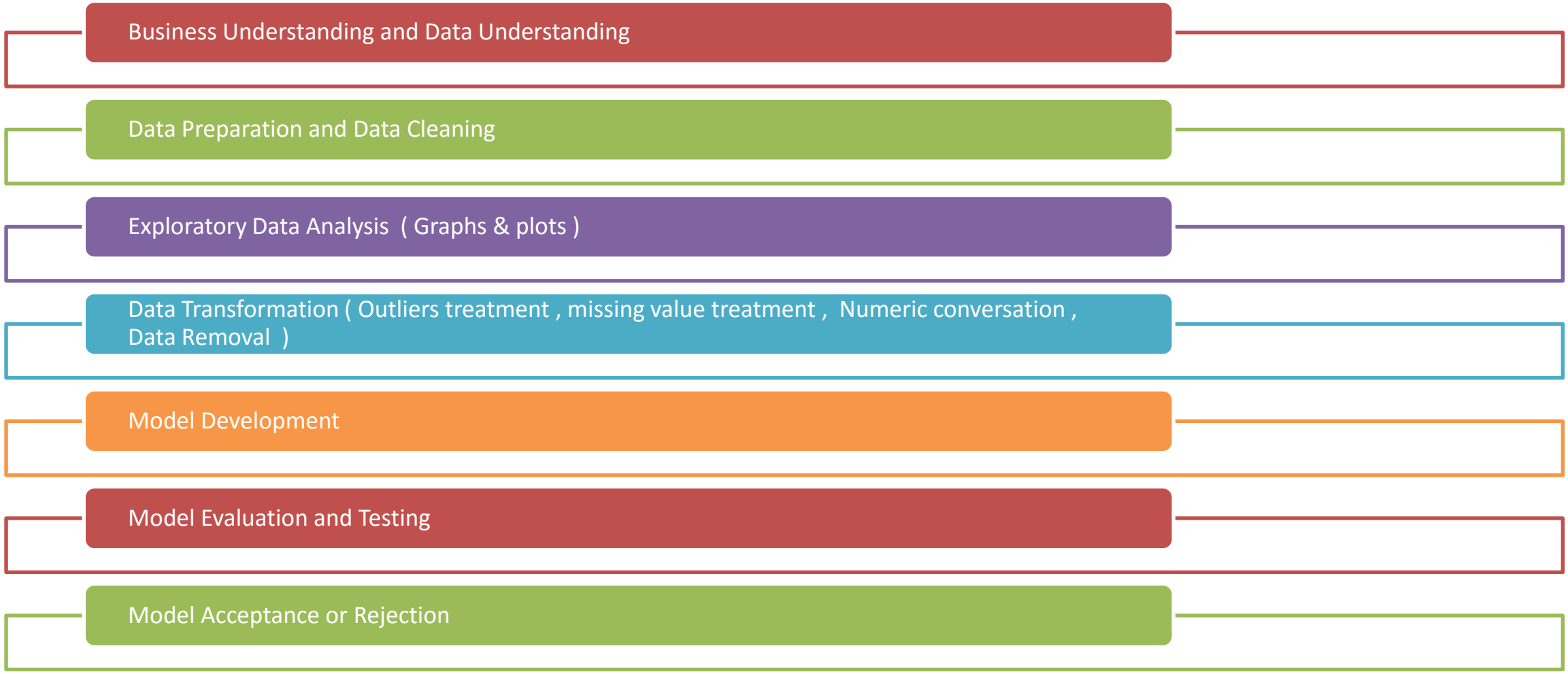
Problem Statement

- High attrition of talent is causing the following problem.
- Company reputation loss with consumer and partner due to delay in project delivery and timeline slippage.
- A sizeable department has to be maintained, for the purposes of **recruiting** new talent.
- Stiff learning curve for new joiners to be trained and acclimatize to the company.
- Difficult to source replacement from available market talent pool.

Business Objective

- Understand what factors XYZ company should focus on To curb attrition.
- What changes XYZ can make to their workplace, in order to get most of their employees to stay.
- Which of these variables is most important and needs to be addressed right away.

Model Building Methodology



Data Preparation & Cleaning

Post study of the 5 given datasets we have done the following Validation before EDA

DATA Validation

- We have verified common column is EmployeeID in all the data set.
- For working our calculation we have merged in time and out time and have removed Public Holidays.
- For Average working hour calculation we have considered employee leave cases and Public holidays.
- After data validation we have calculated actual and average working hours , overtime and Leaves for each employee for the year.

DATA Preparation

- For EDA and model building activites we have merged all the relevant data into a single Dataframe.
- We have created few calculated columns like No of overtime instances for each employee.
- We have prepared box plot for Outliers treatment , however did not find any significant ones which can be excluded.

Data Manipulation & Transformation

Post Data cleaning we have done the following data transformation

Missing Value Treatment

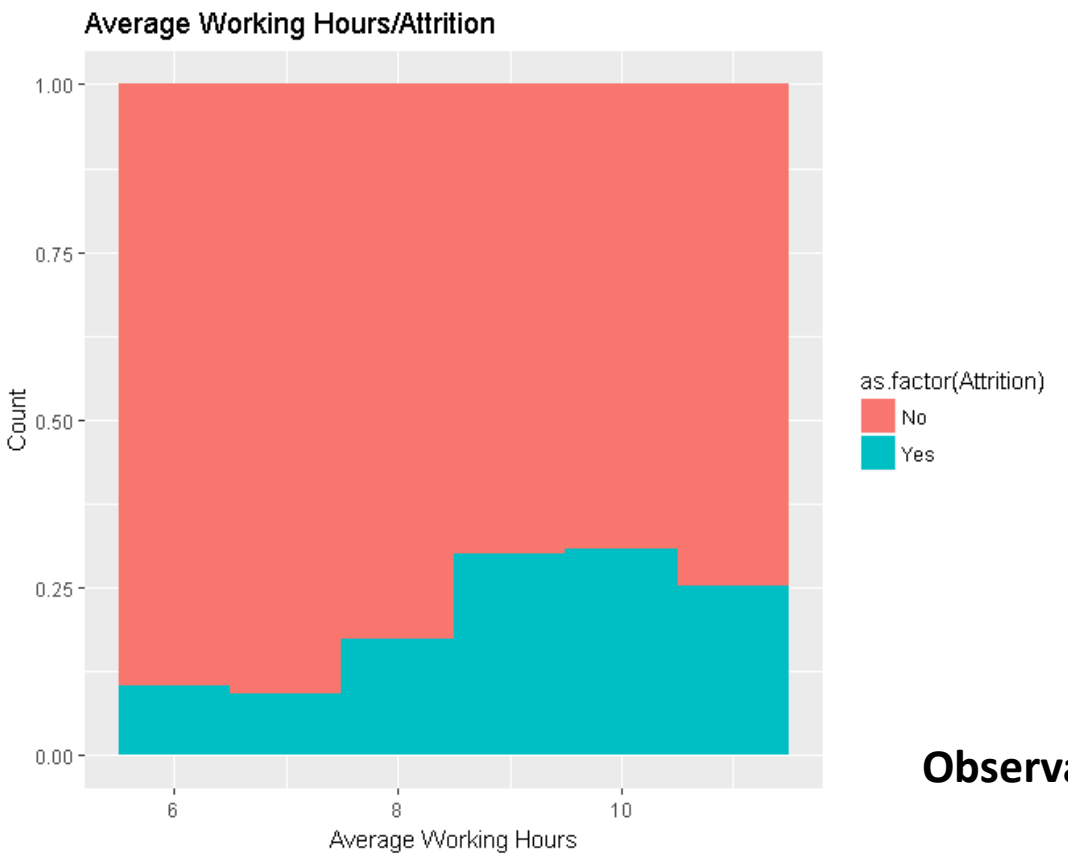
- We have removed rows with NA in NumCompaniesWorked
- We have removed rows with NA in TotalWorkingYears
- We have imputed NAs in EnvironmentSatisfaction column with its mode value, since it is a categorical variable
- We have imputed NAs in JobSatisfaction column with its mode value, as it is also a categorical value.
- We have imputed NAs in WorkLifeBalance column with its mode value, as it is also a categorical value.

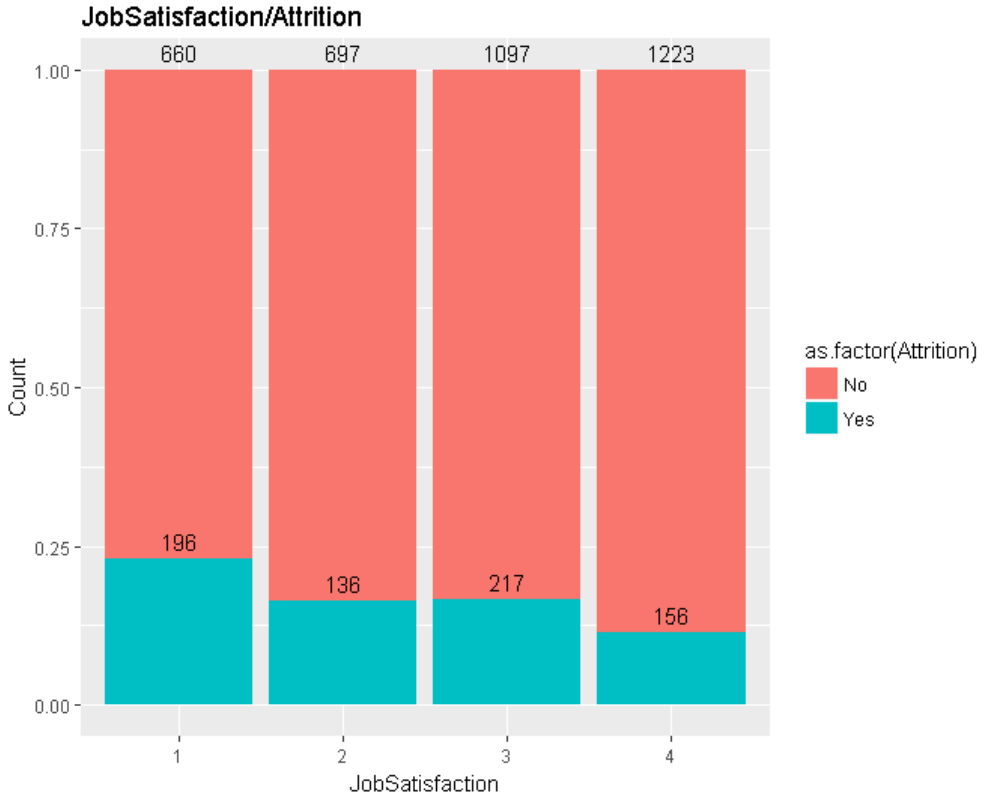
Remove Columns with Unique Value

- We have removed EmployeeCount as it has only a unique value (1).
- We have removed Over18 as it has only a unique value (Y).
- We have removed StandardHours as it has only a unique value (8).

Exploratory Data Analysis

We have plotted the following graphs to identify possible significant variables through EDA





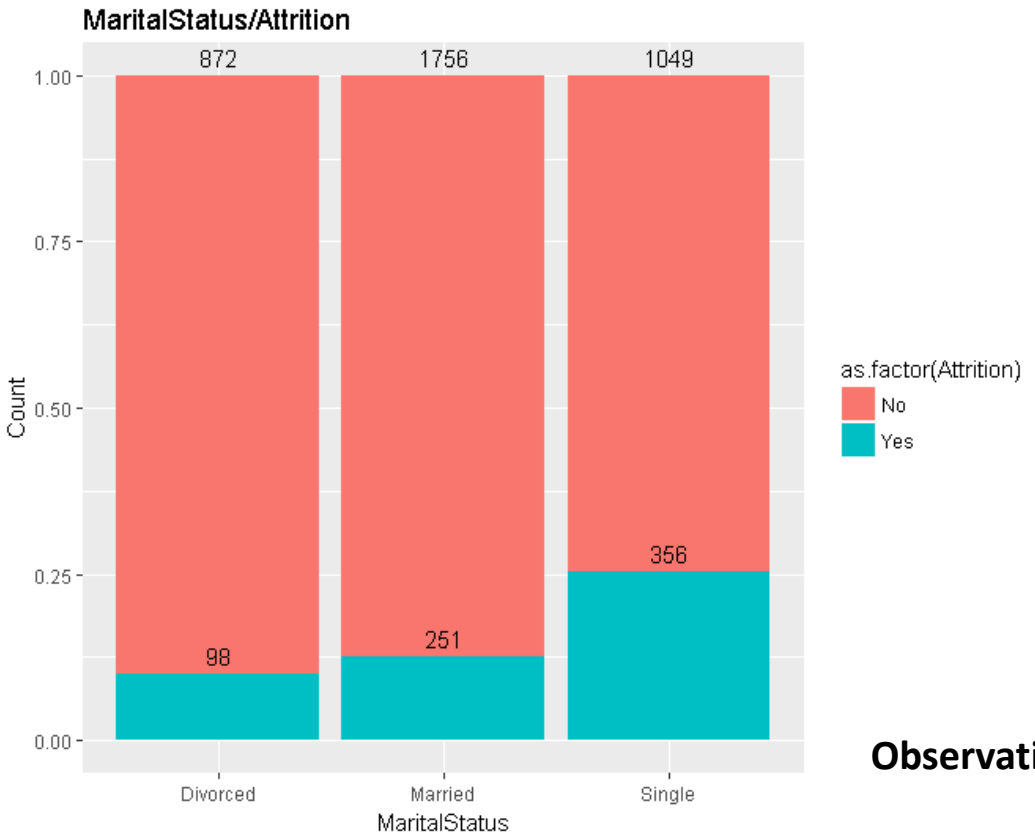
Observation : *Associates with lower job satisfaction has higher risk of leaving the company*

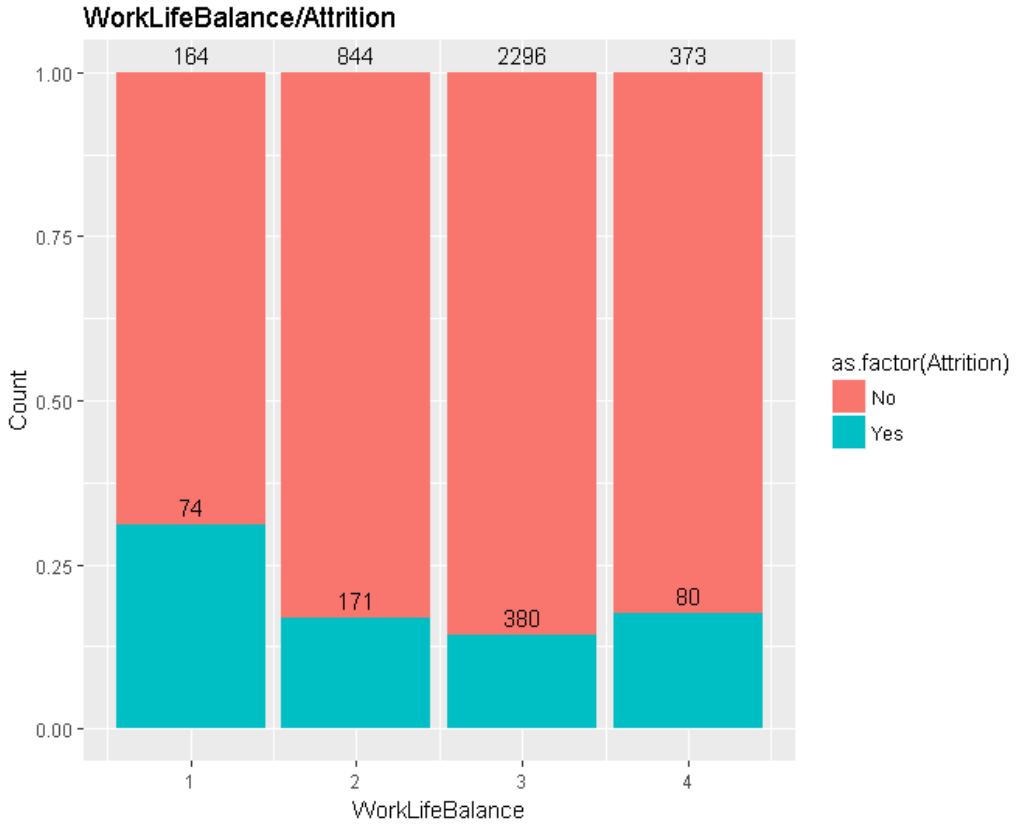
Observation : *Employees who are working more than 8 hours are more likely to leave*

EDA Continue …..

Exploratory Data Analysis ….

We have plotted the following graphs to identify possible significant variables through EDA



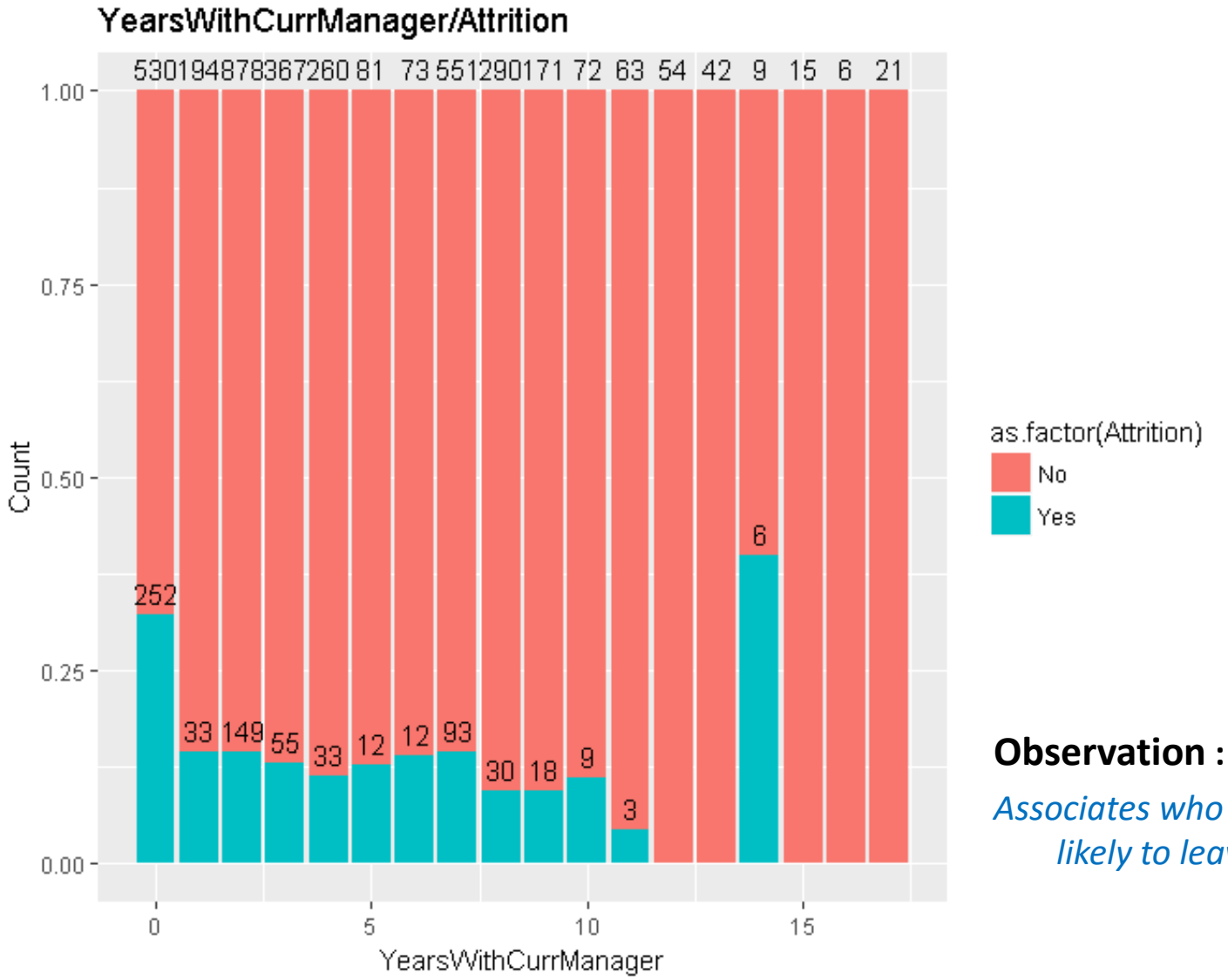


Observation : *Associates with lower work life balance has higher risk of leaving the company*

Observation : *Employee having marital status single has high risk of leaving the company*

EDA Continue …..

Exploratory Data Analysis ….



Observation :

Associates who are working with same manger for long time are less likely to leave the company.

Logistic Regression Model Building

Model Development Steps

1. Feature Standardisation
2. Dummy variable creation and checking duplicates
3. Dummy for variables having 2 level
   - Assigning 1 to Yes and 0 to No for **Attrition** variable
   - Assigning 1 to female and 0 to male for **Gender** variable
4. Split the data into train and test data in the proportion of 70 and 30 percent
5. Removing EmployeeID for training data as it is not required
6. Stepwise selection
7. Removing multicollinearity through **VIF** check
8. Iteratively remove insignificant variables based on high **p-value** & high **VIF**.

Model Summary

- Final Significant Variables

  NumCompaniesWorked, TotalWorkingYears , overtime , BusinessTravelTravel_Frequently , MaritalStatusSingle , EnvironmentSatisfaction2 , EnvironmentSatisfaction3 , EnvironmentSatisfaction4 , JobSatisfaction4 , workLifeBalance3

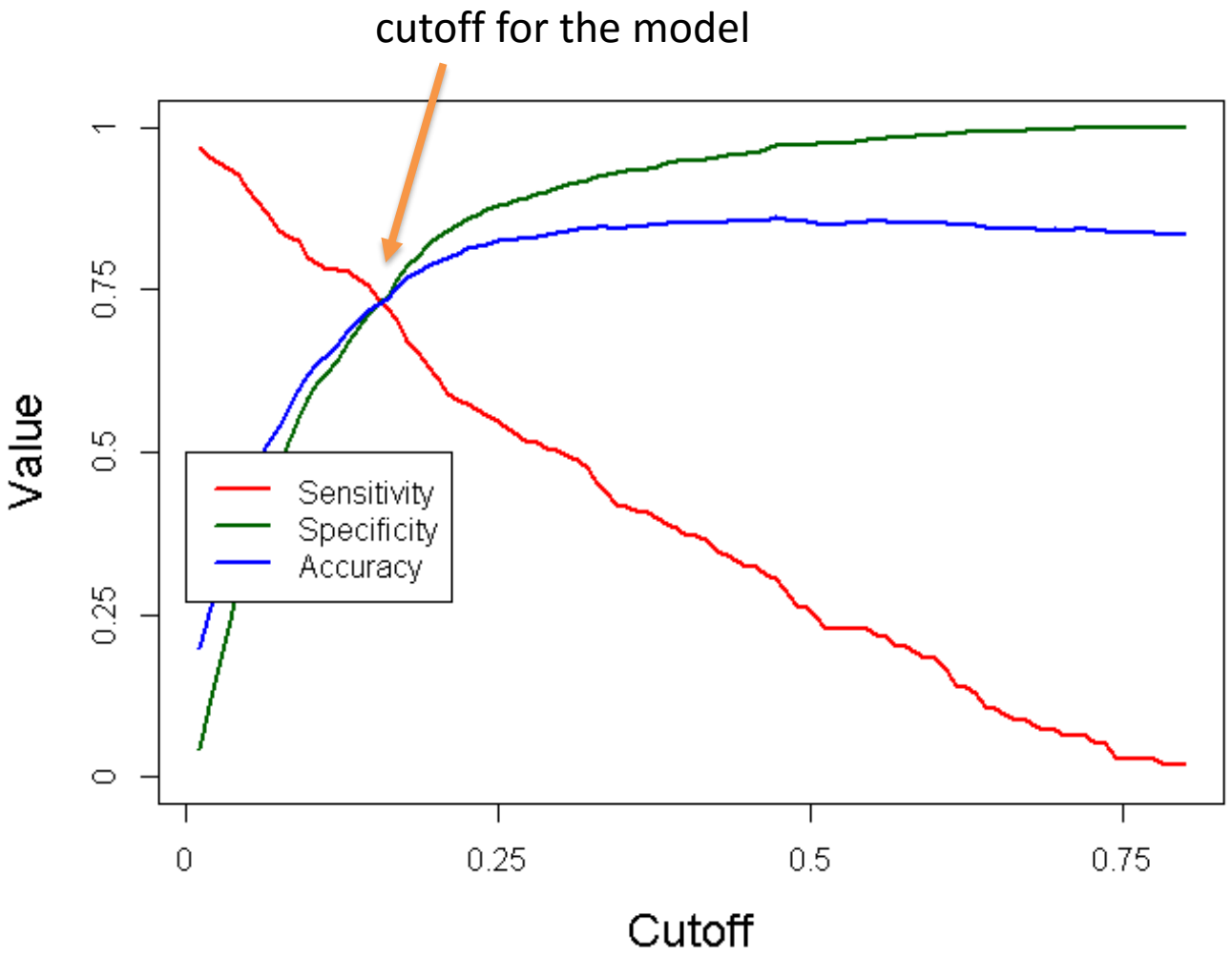
Final Model Statistics	Values
P Value of each variable	All < 0.05 ***
AIC value	2152.3
Null deviance	2681.2
Residual Deviance	2130.3

Model Testing

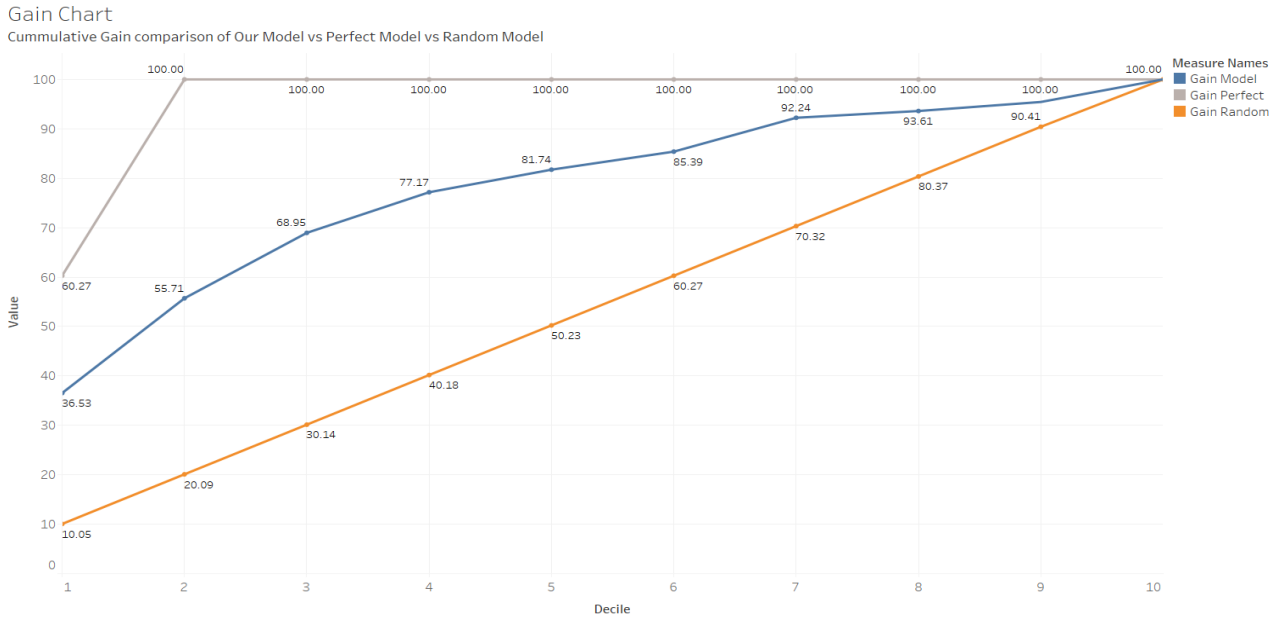
Model Thresholds Test

- We have taken the cutoff values of interception of all three as cutoff for this model and bellow are the final evaluation values for our model.
- The cutoff of the model is decided as **0.1536**.
- The Graph for the accuracy, specificity and sensitivity at various cutoff values is shown in the right.

Threshold value (0.15)	Values
Overall Accuracy	72.78 %
Sensitivity	73.52 %
Specificity	72.63 %



Model Evaluation & Acceptance



Decile	Total	TotalResp	Cumresp	Gain%	Cumlift
1	132	80	80	36.52968	3.652968
2	131	42	122	55.70776	2.785388
3	132	29	151	68.94977	2.298326
4	131	18	169	77.16895	1.929224
5	132	10	179	81.73516	1.634703
6	131	8	187	85.38813	1.423135
7	132	15	202	92.23744	1.317678
8	131	3	205	93.60731	1.170091
9	132	4	209	95.43379	1.060375
10	131	10	219	100	1
Total	1315	219			

Model Acceptance

- As per the gain chart our model is well placed in between the perfect model and Random model
- The KS-statistic value we for our model is > 40%
- Hence we are accepting it as a good model

Method	Test Dataset	
KS-statistic	KS-statistic	46.14 %
Model Evaluation	**Accepted**	

Recommendation for XYZ

Based on EDA and Our Model following are our recommending for XYZ HR Executive Management:

- *HR and management team should look at the following 8 parameter for **Attrition Risk mitigation***:
  - **NumCompaniesWorked , TotalWorkingYears, Overtime, BusinessTravel**
  - **MaritalStatus, EnvironmentSatisfaction rating, JobSatisfaction, workLifeBalance rating**
- *Employee with frequent **Business travel** and low **Work Life Balance** are in high risk of leaving the company*.
- *HR should take a closer look of associates with low **Environment Satisfaction** and **Job Satisfaction** ratings and do a deep drive to understand the employee concerns and address the same.*
- *Associates frequently working over time (**> 8 Hr**) are at high risk of leaving the company. HR team should consider incentivizing those associates and address the problem at ground level to balance the work life*.
- *Associates who are with the organization for long time and having **Marital Status** non-single are probably going to continue with the company for a longer duration. Hence they can be groomed as **Subject Mater Experts** in different areas to reduce project delivery impact in case of attrition in team*.

Thank You