HR Analytics Case Study

Payel Das
Shalini
Sridar Ganesh
Ravishankar

HR Analytics : Attrition

Business problem :

XYZ Company

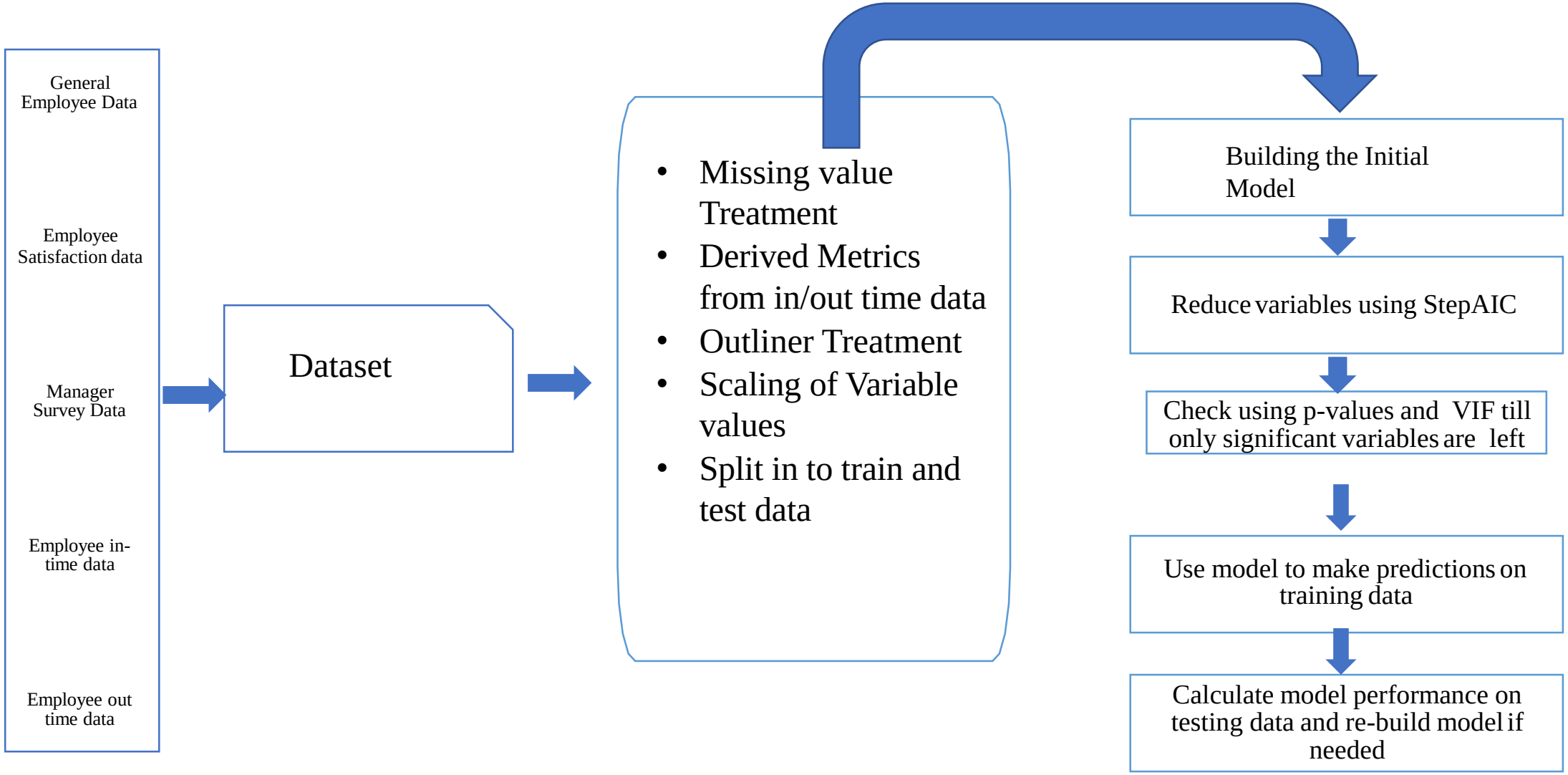
- ❑ Current Workforce strength ~ 4000 employees
- ❑ Each year nearly 15% of employees leave the company
- ❑ This attrition rate is much higher than the ideal turnover rate of ~10%.

- ❖ Delay in Projects making it difficult to meet timelines resulting in a reputation loss among customers and partners
- ❖ Maintenance cost behind large department for recruiting new talent
- ❖ Training new employees incurs additional costs
- ❖ Time involved in acclimatizing for new employees and getting used to company culture leads to further delays in meeting timelines.

Business objectives :

- ❖ Model the **probability of attrition** using a logistic regression. The results thus obtained will be used by the management to understand what changes they should make to their workplace, in order to get most of their employees to stay.
- ❖ Help **HR analytics** firm to understand what factors they should focus on, in order to curb attrition.

Data Analysis Methodology



Data Sourcing and Cleaning

1.The Manager Survey Data – Collected from a company wide survey.
2.The Employee Survey Data – Collected from a company wide survey.
3.In-Time Data – Collected from company's attendance Log sheet/ Machine
4.Out – Time Data – Collected from company's attendance Log sheet/ Machine
5.General Data – General data includes employee wide their personal data along with education and their satisfaction level for association with XYZ org, etc. This data has following fields

- **Working on in_time and out_time data.**
  - Total columns : 250 (249 days + 1 Employee ID)
  - Total rows : 4410
  - Number of columns with all NAs (holidays) : 12 days
- **Data processing :**
  - For every employee build a monthly (%) working hours.
  - For every employee build a total number of yearly offs
  - For every employee build a yearly (%) working hours

Understanding The Data

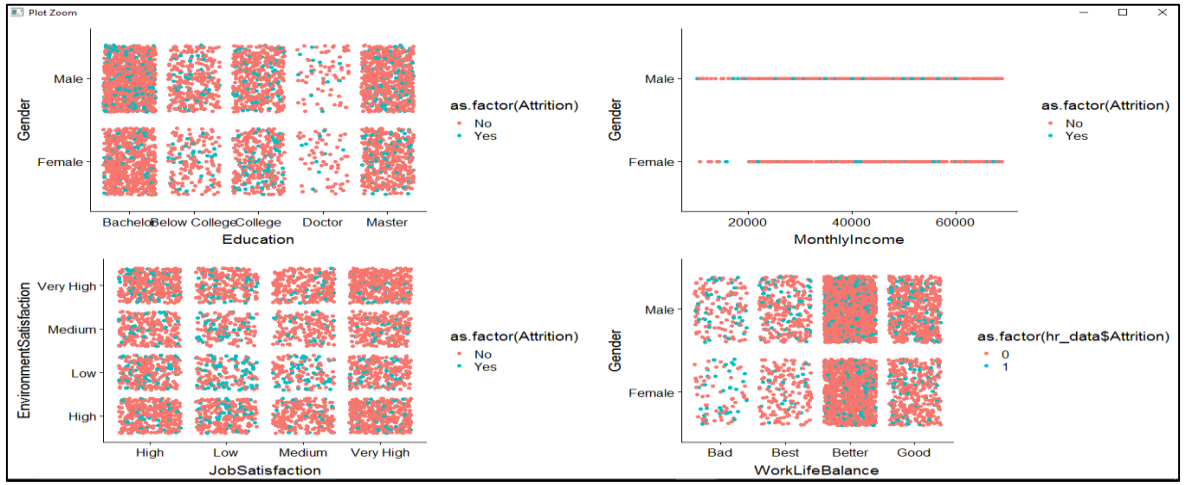
- Employee survey (3 features) : Survey of employee satisfaction on different parameters
- Manager survey (2 features) : Employee opinion of managers
- In time and out time data (Daily) : Employee's in-time and out-time data (Period : 12 months). The columns with only NA values we can consider as holidays and are removed
- Some dates (which are holidays) have in_time, out_time as NA.
- "2015.01.01", "2015.01.14", "2015.01.26", "2015.03.05" ,"2015.05.01" "2015.07.17","2015.09.17", "2015.10.02", "2015.11.09", "2015.11.10" "2015.11.11", "2015.12.25"
- General Data (24 features) : Various information about employees, department, Job Role etc.
- Primary key linking all data : Employee ID (Total employee list – 4410 employees)
- Employee count over 18, standard hours has unique values which were removed

Variables In The Final Dataset Used For Modelling

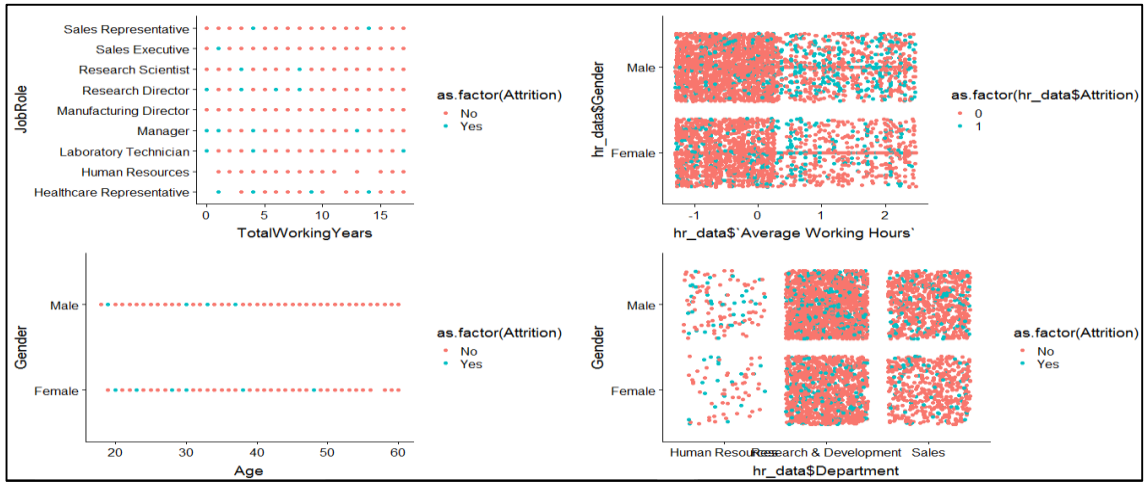
S.No	Feature Name	Remarks
1	Attrition	Possible values = Yes/No
2	BusinessTravel	Possible values = Non-Travel, Travel_Frequently, Travel_Rarely
3	Department	Possible values = Human Resources, Research & Development, Sales
4	EducationField	Possible values = Human Resources, Life Sciences, Marketing, Medical, Other, Technical Degree
5	Gender	Possible values = Female, Male
6	JobRole	Possible values = Healthcare Representative, Human Resources, Laboratory Technician, Manager, Manufacturing Director, Research Director, Research Scientist, Sales Executive, Sales Representative
7	MaritalStatus	Possible values = Divorced, Married, Single
8	StockOptionLevel	Possible values = 0, 1, 2, 3

S.No	Feature Name	Remarks
1	Age, DistanceFromHome	Numeric/continuous values
2	Education	Ordered Categorical – higher number means better education (1 – belowCollege, 5 – Doctor/PhD)
3	JobLevel	Ordered Categorical – higher number means better higher employee band
4	MonthlyIncome, NumCompaniesWorked PercentSalaryHike TotalWorkingYears TrainingTimesLastYear YearsAtCompany YearsSinceLastPromot YearsWithCurrManager	Numerical/continuous values
5	EnvironmentalSatisfaction JobSatisfaction WorkLifeBalance JobInvolvement PerformanceRating	Ordered Categorical

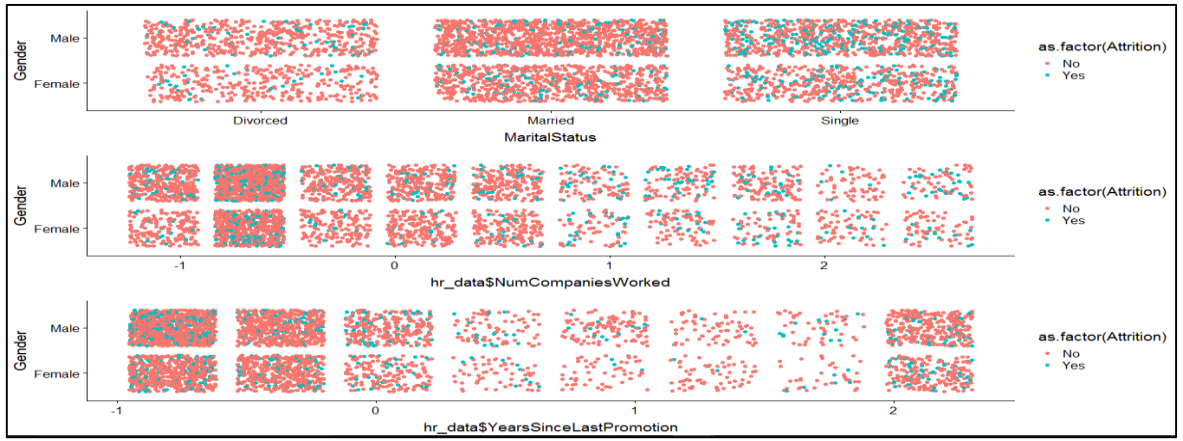
Univariate and Segmented Univariate Analysis



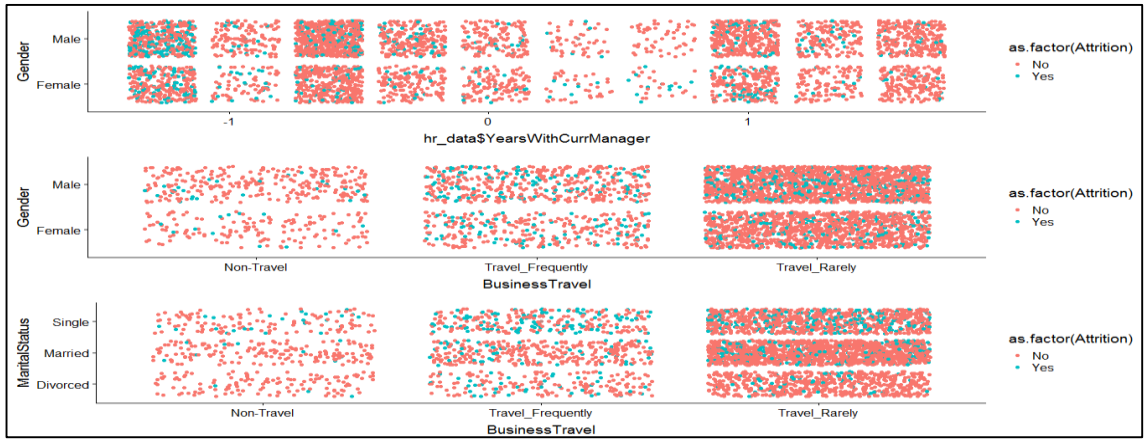
- Higher attrition in those with education less than level 4(Masters)
- Lower the job satisfaction and environment satisfaction, the higher is the attrition



- Less attrition in employees who have worked for more than 10 years
- Higher the average working hours, higher the attrition
- Higher the age, lower the attrition. Low after the age of 40



- Attrition is higher among those single compared to married and divorced employees
- Higher the gap in promotion duration, the higher the attrition



- Higher attrition in those who spend less than 2.5 years with a given manager
- Those who are single and travel frequently are highly likely to leave

Model Building

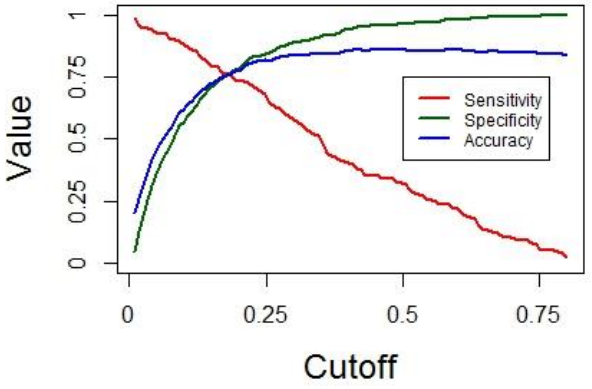
Information related to data used for model building	
Number of independent variables in the final dataset including dummy variables.	56
Dependent variable	“Attrition”
Final Dimensions of data frame used for modeling = 4300 X 57 (train = 70% of 4300 & test = 30% of 4300)	

Operation	AIC value
Model 1 (56 independent + 1 dependent)	2107.4
StepAIC execution Number of columns suggested : 37	2082.4
Model 22 (Final Model) Number of significant columns : 16 (excluding intercept)	2129.6

Model Evaluation – Choosing the best cutoff

- Statistical Criteria- Ideally model should have high Accuracy, Sensitivity and Specificity.
- Business Goals- Our primary goal is to predict correctly the people who are mostly likely to attrite. Even if such a model has lower accuracy it can be acceptable as it will identify employees likely to resign and enable HR of XYZ company to take preventive action.

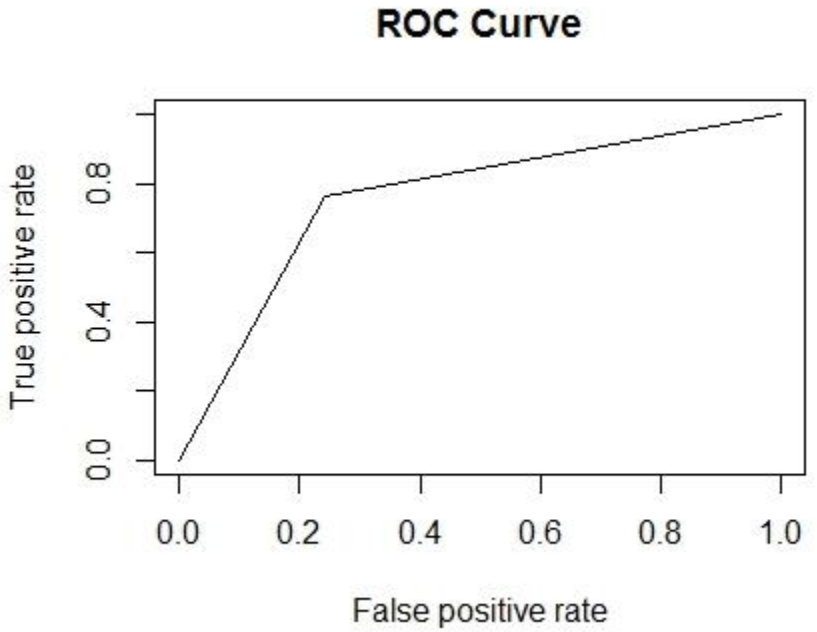
Criterion	Cutoff Value	Accuracy	Sensitivity	Specificity	Confusion Matrix	Comments
Maximum KS value : 0.52	0.1775758	0.7596899	0.7607656	0.759482	Reference Prediction No Yes No 821 50 Yes 260 159	1) High Accuracy but balanced Specificity and Sensitivity 2) Best model for Business objective



Cut-off Value: The point where 3 curves intersect.

Model Evaluation Metrics

```
> Attrition_decile
   bucket total totalresp Cumresp      Gain  Cumlift
1       1   129        79      79  37.79904 3.779904
2       2   129        48     127  60.76555 3.038278
3       3   129        27     154  73.68421 2.456140
4       4   129        17     171  81.81818 2.045455
5       5   129        14     185  88.51675 1.770335
6       6   129         4     189  90.43062 1.507177
7       7   129         6     195  93.30144 1.332878
8       8   129         3     198  94.73684 1.184211
9       9   129         3     201  96.17225 1.068581
10     10   129         8     209 100.00000 1.000000
```



- Our final model is able to capture 81.8% of the bad within top 4 deciles.
- If Model K-S statistic is >40% the model can discriminate between good and bad features quite well. Our model has K-S statistic = 52% which is higher as compared to the acceptable range. So our model is good.
- Area under the curve for a good model should be >=0.8. For our model AUC=0.8 which is expected from a good model and our curve is also rising steeply as expected.

Final Model Coefficients & Their Interpretation:

Variables	Coeff. Estimate	Std. Error	z value	Pr(>|z|)	Significance	Interpretation
(Intercept)	-1.51798	0.21237	-7.148	8.83E-13	***	Negative coefficient indicates, in absence of any controlling factor, an employee is less likely to attrite.
Age	-0.27731	0.07691	-3.605	0.000312	***	Higher the age = lesser attrition probability
NumCompaniesWorked	0.34905	0.05988	5.829	5.57E-09	***	Employee who switches job frequently = higher attrition probability
TotalWorkingYears	-0.58321	0.09144	-6.378	1.79E-10	***	Higher the total experience of employee = lesser the attrition probability
TrainingTimesLastYear	-0.21391	0.05823	-3.673	0.000239	***	More trainings and development = lesser the attrition probability
YearsSinceLastPromotion	0.60775	0.07484	8.121	4.64E-16	***	Delay in promotion = higher the attrition probability
YearsWithCurrManager	-0.43363	0.08664	-5.005	5.58E-07	***	Led by the same manager for a long period of time/ no change in team=lesser the attrition probability
`Average Working Hours`	0.66461	0.05384	12.344	< 2e-16	***	Long working hours=lesser attrition probability
EnvironmentSatisfaction.xLow	0.83081	0.13259	6.266	3.70E-10	***	Satisfactory work environment = higher attrition probability
JobSatisfaction.xLow	0.62196	0.13923	4.467	7.92E-06	***	Lower job satisfaction level = higher attrition probability
JobSatisfaction.xVery.High	-0.52733	0.13857	-3.805	0.000142	***	Very high job satisfaction = lesser attrition probability
WorkLifeBalance.xBest	-1.09214	0.27015	-4.043	5.28E-05	***	Work life balance best = lesser attrition probability
WorkLifeBalance.xBetter	-1.27484	0.21193	-6.015	1.80E-09	***	Work life balance better = lesser attrition probability
WorkLifeBalance.xGood	-1.04080	0.22705	-4.584	4.56E-06	***	Work life balance good = lesser attrition probability
BusinessTravel.xTravel_Frequently	0.69088	0.13259	5.211	1.88E-07	***	Frequent business travel = higher attrition probability
JobRole.xManufacturing.Director	-0.82486	0.21719	-3.798	0.000146	***	Higher the position(such as Director) = lesser attrition probability
MaritalStatus.xSingle	0.91858	0.11575	7.936	2.09E-15	***	Single employee = higher attrition probability

Our Recommendations

❖ Key Recommendations

The Following are the key factors which are influencing the attrition. XYZ Corporation needs to focus in the following to curb the attrition.

1. The Major influential factor is the number of employees putting high number of hours. XYZ corporation should plan better and control late working hours.
2. Those who change managers is contributing to high attrition. Employees are concern and reservations against working with new managers frequently. XYZ needs to plan for a robust transition plan in case of any change in managers.
3. Promotions is an important factor for attrition. XYZ management needs to focus better management of promotions and appreciations.
4. Work life balance and environmental satisfactions plays an important role in attrition. XYZ HR should focus on team building, outings, events, and provide a flexible working options for their employees
5. Job Satisfaction and rotating of job after a period of time will help to reduce the attrition
6. Business Travel has a impact on attrition. XYZ needs to focus on avoiding frequent long travels for employees. They may adopt modern collaborations tools as far as possible.
7. Employees who are unmarried are prone to leaving the company. This may in the case of fresher's exploring new job opportunities frequently. XYZ HR can come our with a strategy to inform such fresher's on their growth path in the company.