HR ANALYTICS CASE STUDY

(Attrition Propensity Model)

SUBMISSION

Group Name : Quantitative-Soothsayers

1. Asma Singh
2. Hemanth Ponnada
3. Chayan Chatterjee
4. Himanshu Srivastava

Problem Solving Methodology

Business Understanding

To understand what factors company should focus on, in order to curb attrition.

Data Understanding

5 data files(in_time,out_time,employee_survey_data,

general_data,manager_survey_data)

Each file has 4410 observations

Contains NA values too

Data Preparation And EDA

Removing duplicates

Removing colums with complete NA values

Assigning Names to levels of factor variables eg.Low','Medium’e etc

Removing Outliers

Deriving Metrics

Treating NA values

Plotting Graphs

Problem Solving Methodology

Model Building considering both technical and business aspects and by correct variables selection method

Model Evaluation by Various method:

1.Indentifying Accuracy of the Model

2.Senstivity and Specificty

3.Gain And Lift charts

4.KS statistic

Quick Summary

Key Take-Away --

1.Understanding the structures of the data file

2.Removing duplicates: Each data file has 4410 observation.No duplicates observed

3.Checking columns for NA values –in_time and out_time dataset has 12 columns with complete NA values. After seeing the pattern it is observed that those columns indicates holidays

4. Assigning names to levels of categorical variables as per data dictionary provided e.g 'Low‘, 'Medium‘, 'High', 'Very High‘

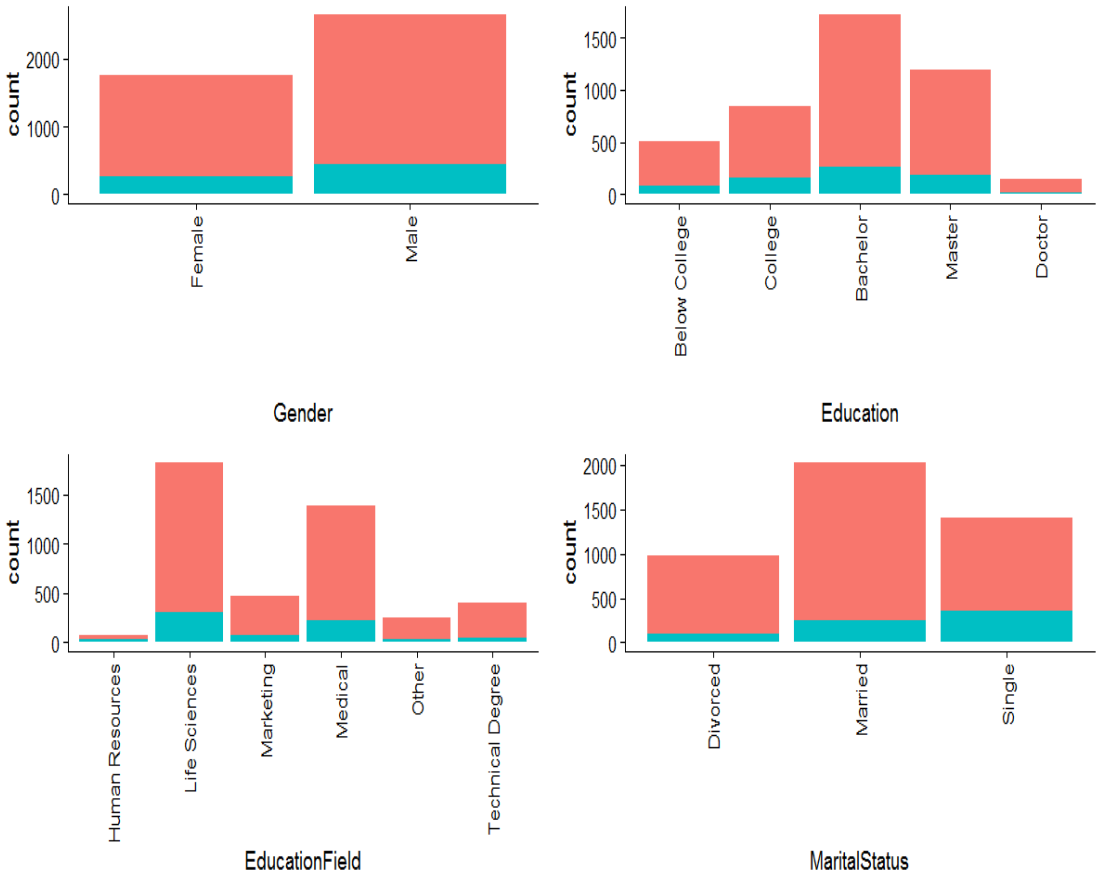
5.Deriving metrics –Calculated Average working hours for each employee from in_time and out_time dataset.

6. Replacing NA values for categorical values by "Mode" and numeric values by "Median“

7. Plotted graphs to show factors affecting Attrition, grapically

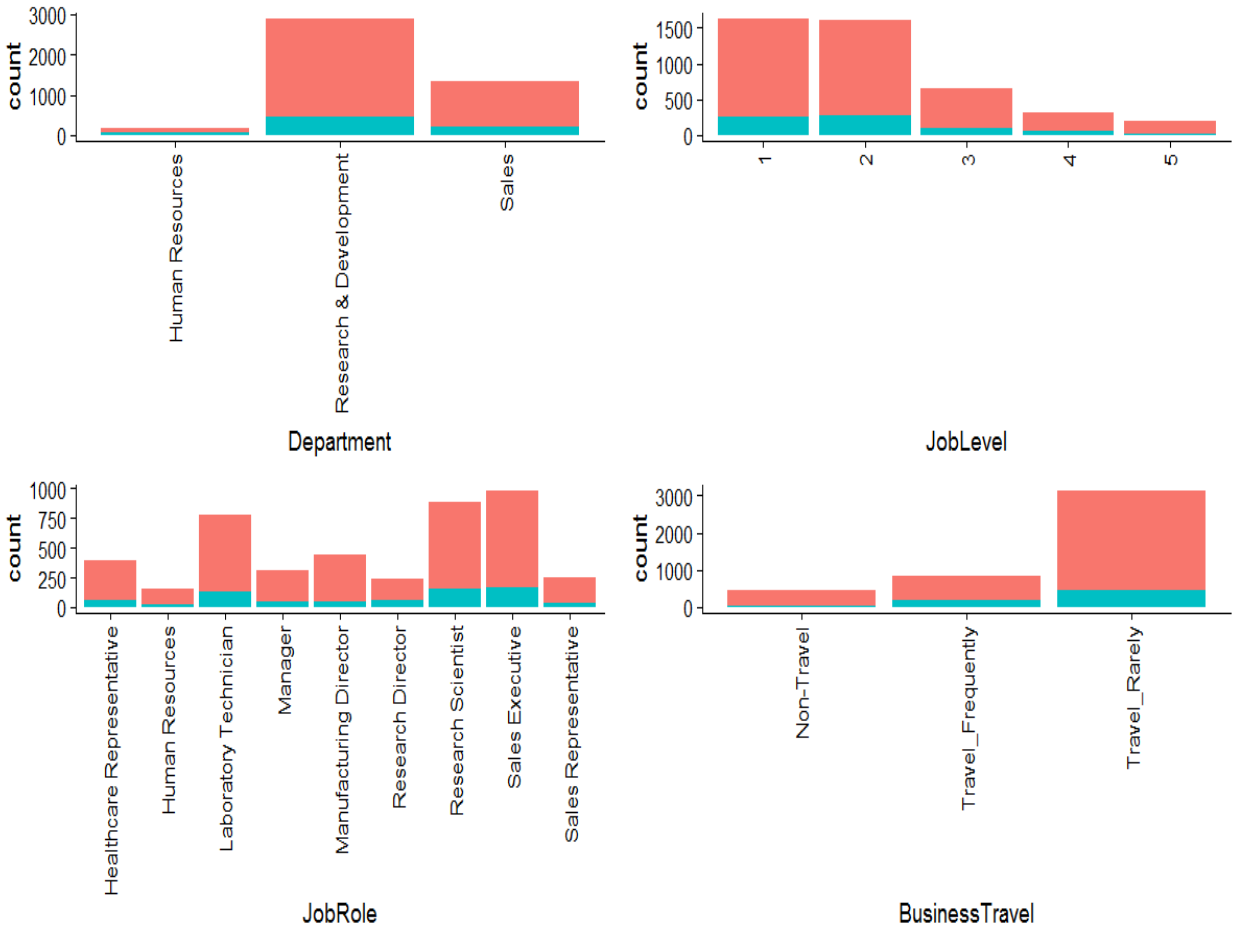
✓ Programming done in : **R**

Visualizing attrition rate for each categorical variables(1/4)



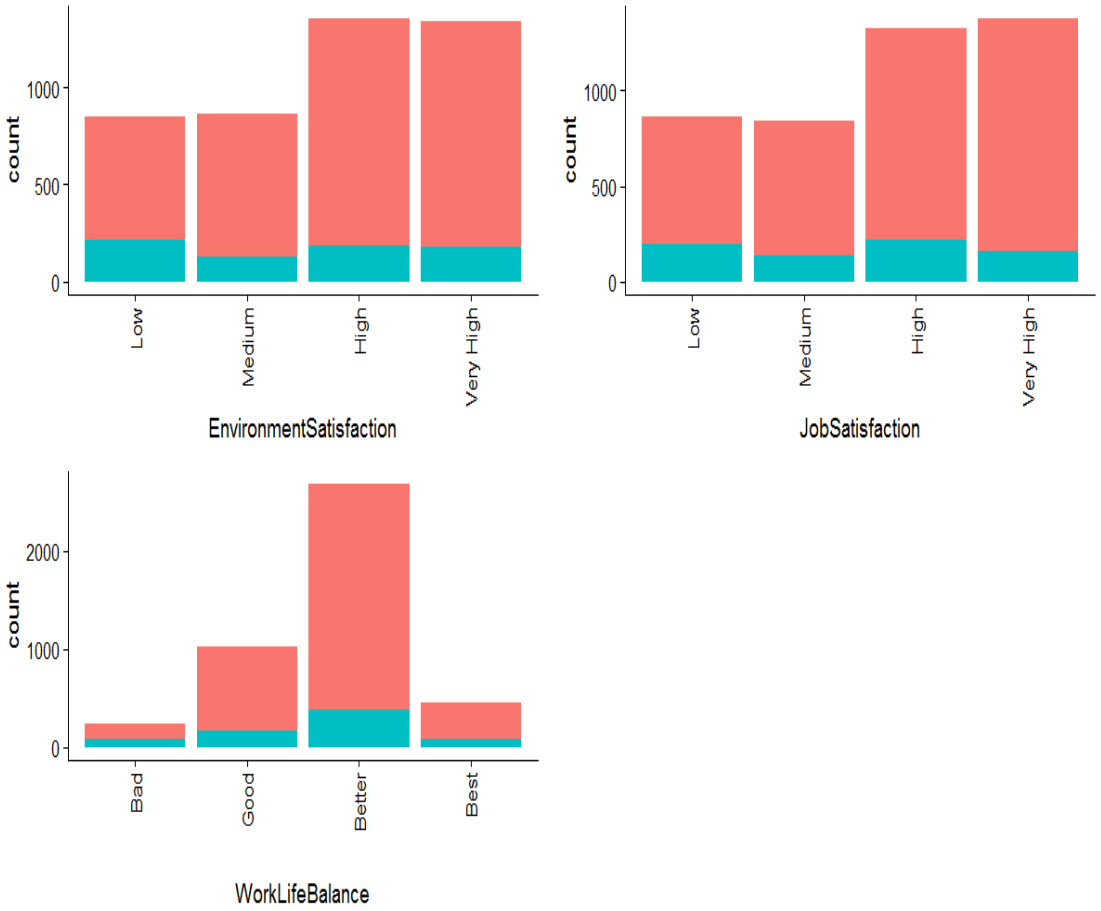
- In the **first plot** attrition rate is more in males compared to females.
- In the **second plot** attrition rate is more in people with bachelor's degree followed by people with master's degree then followed by people who completed the college
- In the **third plot** attrition rate is maximum in people from field of life sciences and minimum in Human resources.
- In the **fourth plot** attrition rate is maximum among married people and least among divorced ones.

Visualizing attrition rate for each categorical variables(2/4)



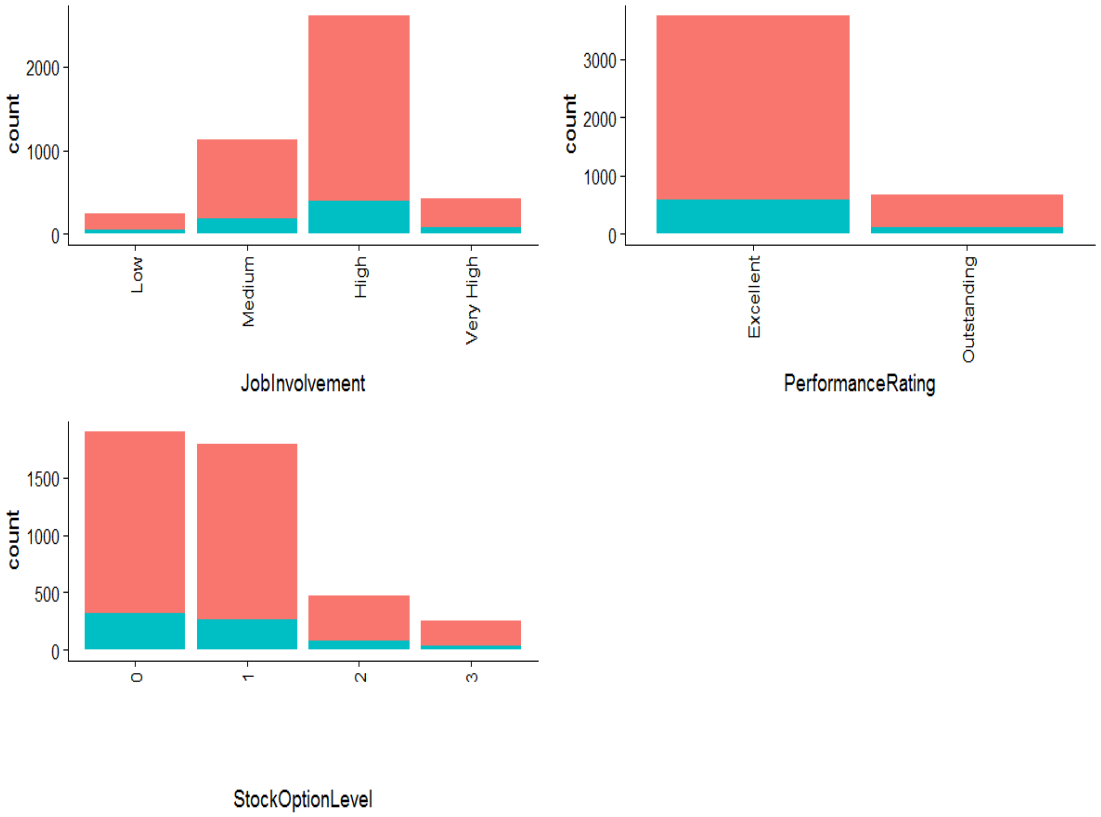
- In the **first plot** attrition rate is maximum in people from Research and development department and least in Human resources department
- In **second plot** attrition rate is maximum in JobLevel 2 and minimum in JobLevel 5
- In **third plot** attrition rate is maximum in Research Scientist and minimum in Human Resources JobRole
- In **fourth plot** attrition rate is maximum in Travel_Rarely and least in Non-Travel

Visualizing attrition rate for each categorical variables(3/4)



- In **first plot** attrition rate is maximum in people with high Environment satisfaction and least in low Environment satisfaction
- In **second plot** attrition rate is maximum in people with Very High JobSatisfaction and least in people with medium JobSatisfaction
- In **third plot** attrition rate is maximum in people with better WorkLifeBalance and least in bad WorkLifeBalance

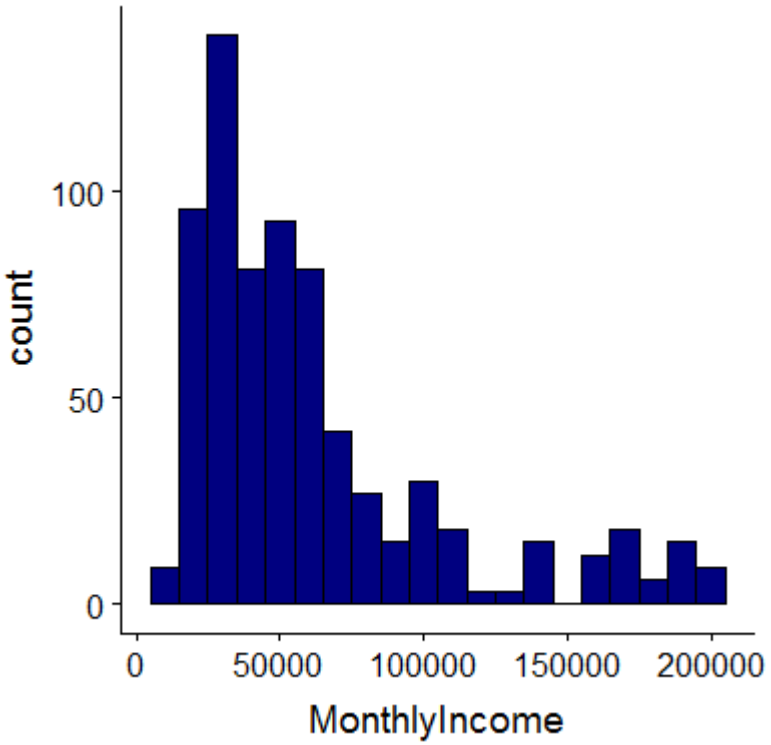
Visualizing attrition rate for each categorical variables(4/4)



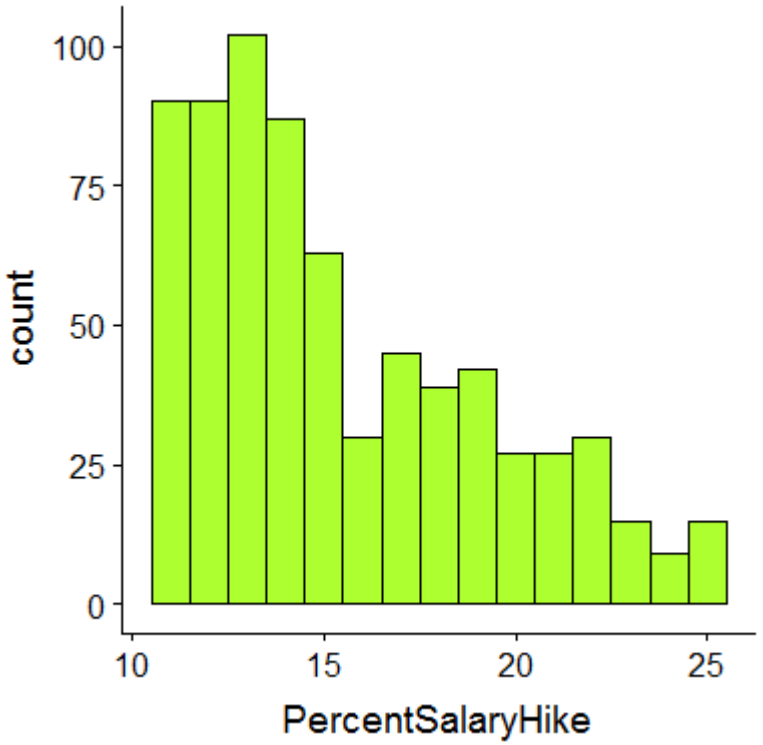
- In the **first plot** attrition rate in people with **high** JobInvolvement and least in low JobInvolvement
- In the **second plot** attrition rate is maximum among Excellent PerformanceRating people and least among outstanding PerformanceRating people
- In **third plot** attrition rate is maximum in StockOptionLevel 0 and least in StockOptionLevel 3

Visualizing attrition rate for each numeric variables(1/2)

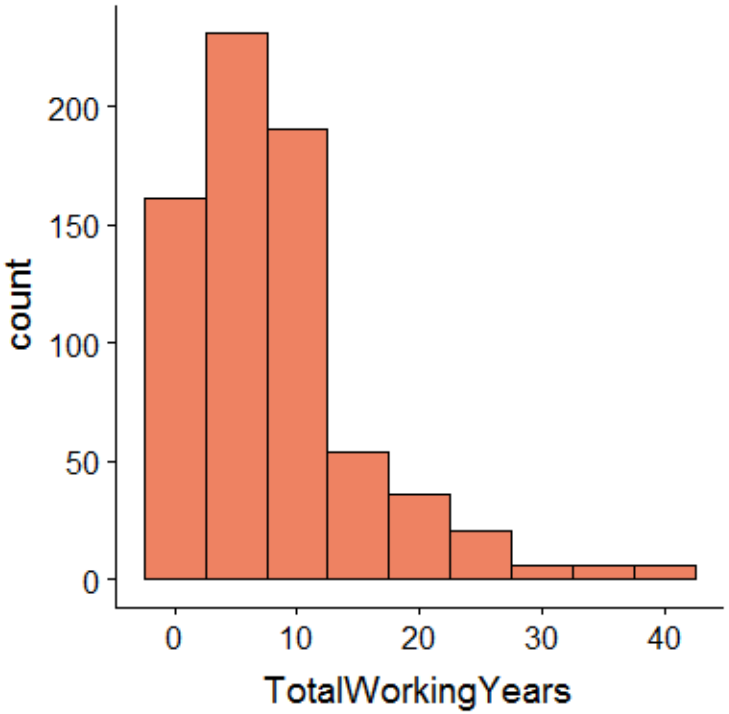
MONTHLY INCOME: lower monthly income more no. of attrition can be seen



PERCENT SALARY HIKE: lower percent salary hike attrition rate is high

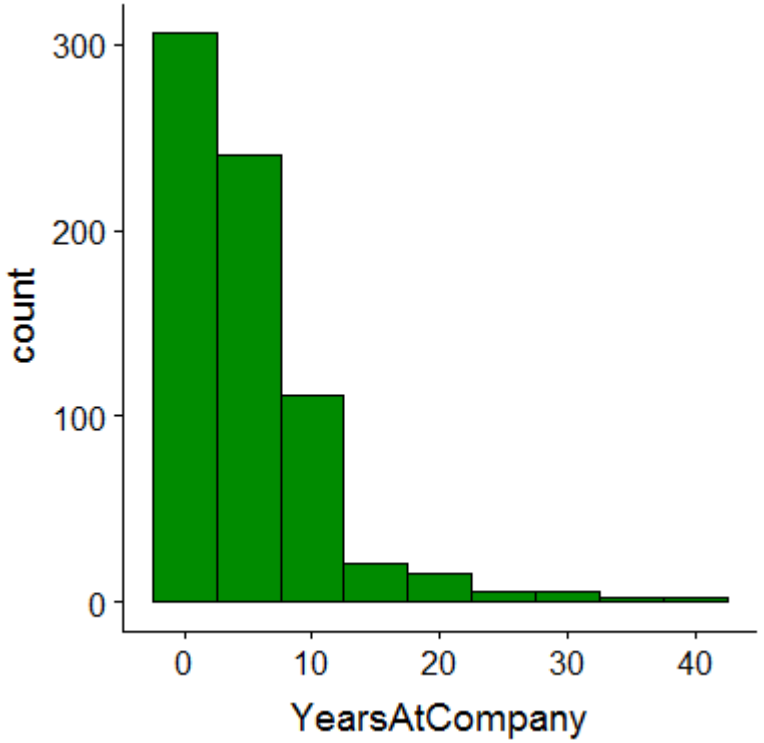


TotalWorkingYears: Higher working years people have low attrition rate compared to people with less working years

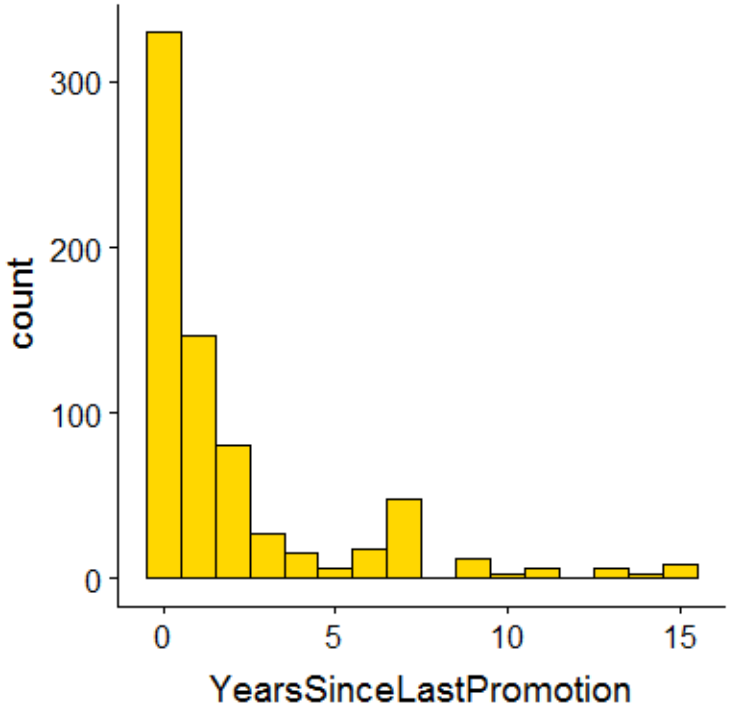


Visualizing attrition rate for each numeric variables(2/2)

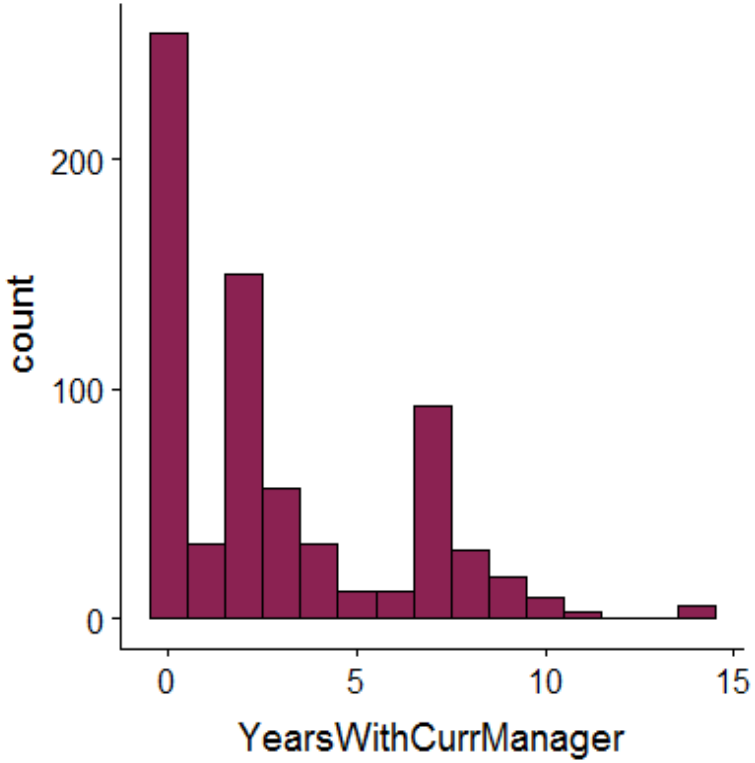
YearsAtCompany:Freshers are more likely to leave company



YearsSincsLastPromotiom:Employees are more likely to leave company after getting promoted



YearsWithCurrManager:There are more count of attrition for lesser the no. of years spent with current manager.This may be due to change in manager or team.



Model Building

17 iterations took to come up with the final model

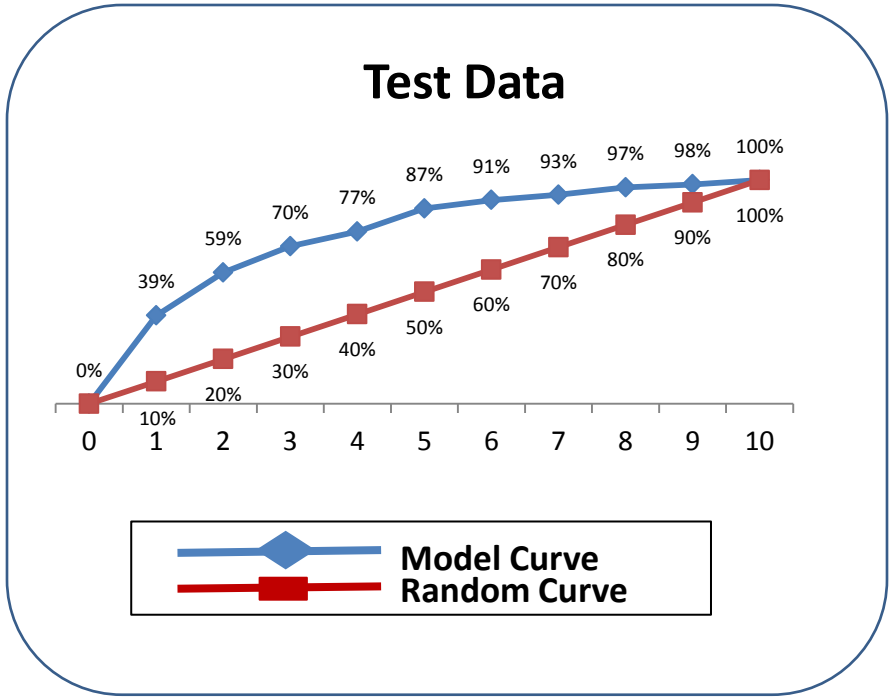
Ratio of train and test dataset is 70:30.

14 variables comes out to be significant , affecting the attrition:

- ✓ Age
- ✓ NumCompaniesWorked
- ✓ TotalWorkingYears
- ✓ TrainingTimesLastYear
- ✓ YearsSinceLastPromotion
- ✓ Overtime
- ✓ Engagement
- ✓ BusinessTravel.xTravel_Frequently
- ✓ maritalStatus.xSingle
- ✓ EnvironmentSatisfaction.xMedium
- ✓ EnvironmentSatisfaction.xHigh
- ✓ EnvironmentSatisfaction.xVery.High
- ✓ JobSatisfaction.xVery.High
- ✓ WorkLifeBalance.xBetter

Model Evaluation Results

Gain Chart



- Top 4 deciles capture 77% of the attritions showing the effectiveness of the model

Some Important Model Evaluation Metrics :

Model Evaluation Metrics	Test Data Results
AUC	0.8161
Optimal Cutoff	0.1680
Sensitivity	0.7277
Specificity	0.7252
Accuracy	0.7256
KS Statistics	0.4529

- KS statistics being 45% which is the maximum difference between cumulative %Attrition & cumulative %Non-Attritions
- At Optimal cutoff - Sensitivity, Specificity and Accuracy coincide with each other (~72% each)
- Area under the curve coming from ROC curve is 81%

Recommendations

- Odds of Attritions are lower with increase in Age
- Odds of Attritions are higher if the employee has worked for more number of companies
- Odds of Attritions are lower with increase in Total Working Years
- Odds of Attritions are lower if the employee is Single etc.

```
Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -1.67319    0.14397 -11.622  < 2e-16 ***
Age                              -0.27017    0.07801  -3.463 0.000533 ***
NumCompaniesWorked                0.37151    0.05661   6.562 5.29e-11 ***
TotalWorkingYears                -0.74537    0.10130  -7.358 1.86e-13 ***
TrainingTimesLastYear            -0.20318    0.05816  -3.493 0.000477 ***
YearsSinceLastPromotion           0.38629    0.06997   5.520 3.38e-08 ***
overtime                          1.56668    0.11720  13.368  < 2e-16 ***
Engagement                       -0.39263    0.05426  -7.237 4.59e-13 ***
BusinessTravel.xTravel_Frequently 0.85972    0.13023   6.602 4.07e-11 ***
MaritalStatus.xSingle             0.94026    0.11520   8.162 3.30e-16 ***
EnvironmentSatisfaction.xMedium  -0.92449    0.16616  -5.564 2.64e-08 ***
EnvironmentSatisfaction.xHigh    -1.12482    0.15427  -7.291 3.07e-13 ***
EnvironmentSatisfaction.xVery.High -1.34992  0.15896  -8.492  < 2e-16 ***
JobSatisfaction.xVery.High       -0.81313    0.13230  -6.146 7.94e-10 ***
WorkLifeBalance.xBetter          -0.45271    0.11300  -4.006 6.17e-05 ***
```