IBM HR ANALYTICS

Aravind

March 1, 2018

```
ibm_data <- read.csv("IBM_HR_Attrition.csv")</pre>
str(ibm_data)
                    1470 obs. of 35 variables:
## 'data.frame':
   $ i..Age
##
                              : int 41 49 37 33 27 32 59 30 38 36 ...
## $ Attrition
                              : Factor w/ 2 levels "No", "Yes": 2 1 2 1 1 1 1 1 1 1 ...
                              : Factor w/ 3 levels "Non-Travel", "Travel_Frequently",..: 3 2 3 2 3 2 3 3
## $ BusinessTravel
## $ DailyRate
                                     1102 279 1373 1392 591 1005 1324 1358 216 1299 ...
## $ Department
                              : Factor w/ 3 levels "Human Resources",..: 3 2 2 2 2 2 2 2 2 2 ...
## $ DistanceFromHome
                                     1 8 2 3 2 2 3 24 23 27 ...
## $ Education
                              : int
                                     2 1 2 4 1 2 3 1 3 3 ...
   $ EducationField
                              : Factor w/ 6 levels "Human Resources",..: 2 2 5 2 4 2 4 2 2 4 ...
## $ EmployeeCount
                                    1 1 1 1 1 1 1 1 1 1 ...
## $ EmployeeNumber
                              : int
                                    1 2 4 5 7 8 10 11 12 13 ...
                                    2 3 4 4 1 4 3 4 4 3 ...
## $ EnvironmentSatisfaction : int
                              : Factor w/ 2 levels "Female", "Male": 1 2 2 1 2 2 1 2 2 2 ...
## $ Gender
## $ HourlyRate
                                    94 61 92 56 40 79 81 67 44 94 ...
## $ JobInvolvement
                              : int
                                    3 2 2 3 3 3 4 3 2 3 ...
## $ JobLevel
                                    2 2 1 1 1 1 1 1 3 2 ...
## $ JobRole
                              : Factor w/ 9 levels "Healthcare Representative",...: 8 7 3 7 3 3 3 3 5 1
## $ JobSatisfaction
                                    4 2 3 3 2 4 1 3 3 3 ...
## $ MaritalStatus
                              : Factor w/ 3 levels "Divorced", "Married", ...: 3 2 3 2 2 3 2 1 3 2 ...
## $ MonthlyIncome
                                    5993 5130 2090 2909 3468 3068 2670 2693 9526 5237 ...
## $ MonthlyRate
                                    19479 24907 2396 23159 16632 11864 9964 13335 8787 16577 ...
                              : int
## $ NumCompaniesWorked
                                    8 1 6 1 9 0 4 1 0 6 ...
## $ Over18
                              : Factor w/ 1 level "Y": 1 1 1 1 1 1 1 1 1 1 ...
## $ OverTime
                              : Factor w/ 2 levels "No", "Yes": 2 1 2 2 1 1 2 1 1 1 ...
## $ PercentSalaryHike
                                    11 23 15 11 12 13 20 22 21 13 ...
                              : int
## $ PerformanceRating
                                    3 4 3 3 3 3 4 4 4 3 ...
                              : int
## $ RelationshipSatisfaction: int
                                    1 4 2 3 4 3 1 2 2 2 ...
## $ StandardHours
                                    80 80 80 80 80 80 80 80 80 80 ...
                              : int
## $ StockOptionLevel
                                    0 1 0 0 1 0 3 1 0 2 ...
                              : int
                                     8 10 7 8 6 8 12 1 10 17 ...
## $ TotalWorkingYears
                              : int
                                    0 3 3 3 3 2 3 2 2 3 ...
## $ TrainingTimesLastYear
                              : int
## $ WorkLifeBalance
                              : int
                                    1 3 3 3 3 2 2 3 3 2 ...
                                    6 10 0 8 2 7 1 1 9 7 ...
## $ YearsAtCompany
                              : int
## $ YearsInCurrentRole
                              : int
                                    4707270077...
## $ YearsSinceLastPromotion : int
                                    0 1 0 3 2 3 0 0 1 7 ...
## $ YearsWithCurrManager
                              : int
                                    5700260087...
head(ibm_data)
                         BusinessTravel DailyRate
                                                              Department
     i...Age Attrition
## 1
         41
                  Yes
                          Travel_Rarely
                                             1102
                                                                   Sales
## 2
                  No Travel_Frequently
         49
                                              279 Research & Development
```

1373 Research & Development

1392 Research & Development

591 Research & Development

Travel_Rarely

Travel Rarely

No Travel Frequently

3

4

5

37

33

27

Yes

No

```
No Travel_Frequently
                                                1005 Research & Development
## 6
     DistanceFromHome Education EducationField EmployeeCount EmployeeNumber
## 1
                                2 Life Sciences
                     1
## 2
                     8
                                   Life Sciences
                                                                                2
## 3
                     2
                                2
                                            Other
                                                                                4
## 4
                     3
                                4
                                   Life Sciences
                                                               1
                                                                                5
                     2
## 5
                                          Medical
                     2
                                2 Life Sciences
## 6
     EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel
## 1
                             2 Female
                                               94
                                                                 3
## 2
                                 Male
                                               61
                                                                 2
                                                                          2
## 3
                                 Male
                                               92
                                                                 2
                                                                          1
## 4
                             4 Female
                                               56
                                                                 3
                                                                          1
## 5
                                                                 3
                                 Male
                                               40
## 6
                                 Male
                                               79
                                                                 3
##
                    JobRole JobSatisfaction MaritalStatus MonthlyIncome
## 1
           Sales Executive
                                            4
                                                      Single
                                                                       5993
## 2
                                                                       5130
        Research Scientist
                                                     Married
                                                                       2090
## 3 Laboratory Technician
                                            3
                                                      Single
        Research Scientist
                                            3
                                                                       2909
                                                     Married
## 5 Laboratory Technician
                                            2
                                                     Married
                                                                       3468
## 6 Laboratory Technician
                                            4
                                                      Single
                                                                       3068
     MonthlyRate NumCompaniesWorked Over18 OverTime PercentSalaryHike
##
## 1
           19479
                                    8
                                            Y
                                                    Yes
## 2
           24907
                                                                        23
                                    1
                                            γ
                                                     No
## 3
            2396
                                            Y
                                                    Yes
                                                                        15
## 4
           23159
                                    1
                                            Y
                                                    Yes
                                                                        11
## 5
           16632
                                    9
                                            Y
                                                     No
                                                                        12
                                    0
                                            Y
## 6
           11864
                                                     No
                                                                        13
     PerformanceRating RelationshipSatisfaction StandardHours
## 1
                      3
                                                 1
## 2
                      4
                                                 4
                                                               80
## 3
                      3
                                                 2
                                                               80
                      3
                                                 3
## 4
                                                               80
## 5
                      3
                                                 4
                                                               80
                      3
## 6
                                                 3
                                                               80
     StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance
## 1
                     0
                                         8
                                                                 0
                                                                                  1
## 2
                                                                 3
                                        10
                                                                                  3
## 3
                     0
                                         7
                                                                 3
                                                                                  3
## 4
                     0
                                                                 3
                                                                                  3
                                         6
                                                                 3
                                                                                  3
## 5
                     1
## 6
                     0
                                         8
                                                                                  2
     YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion
## 1
                   6
                                        4
## 2
                  10
                                        7
                                                                  1
## 3
                   0
                                        0
                                                                  0
## 4
                   8
                                        7
                                                                  3
## 5
                   2
                                        2
                                                                  2
## 6
                   7
                                                                  3
     YearsWithCurrManager
## 1
                          5
## 2
                          7
## 3
                          0
```

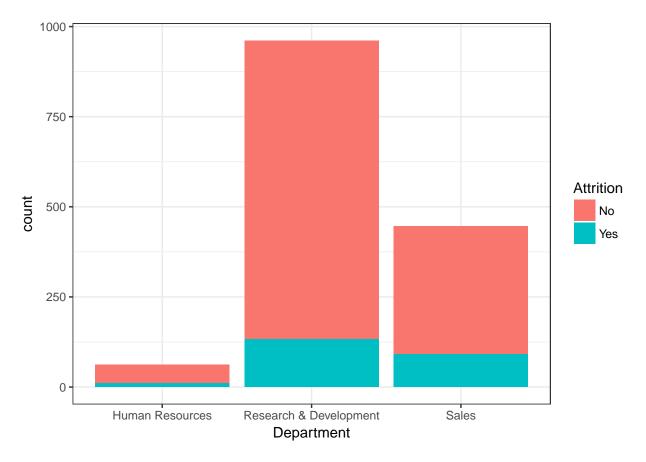
```
## 4 0
## 5 2
## 6 6
```

Check for missing values

```
colSums(is.na(ibm_data))
##
                                                                   BusinessTravel
                      ï..Age
                                             Attrition
##
##
                   DailyRate
                                            Department
                                                                 DistanceFromHome
##
##
                   Education
                                        EducationField
                                                                    EmployeeCount
##
##
             EmployeeNumber
                              EnvironmentSatisfaction
                                                                           Gender
##
                  HourlyRate
                                        JobInvolvement
                                                                         JobLevel
##
##
                                       JobSatisfaction
                                                                    MaritalStatus
                     JobRole
##
##
##
              MonthlyIncome
                                           MonthlyRate
                                                               NumCompaniesWorked
##
                                               OverTime
##
                      Over18
                                                                PercentSalaryHike
##
                                                                    StandardHours
          PerformanceRating RelationshipSatisfaction
##
##
           StockOptionLevel
                                     TotalWorkingYears
                                                           TrainingTimesLastYear
##
##
             WorkLifeBalance
                                                               YearsInCurrentRole
##
                                        YearsAtCompany
##
                                  YearsWithCurrManager
    YearsSinceLastPromotion
##
##
table(ibm_data$Attrition)
##
##
     No
         Yes
## 1233
         237
```

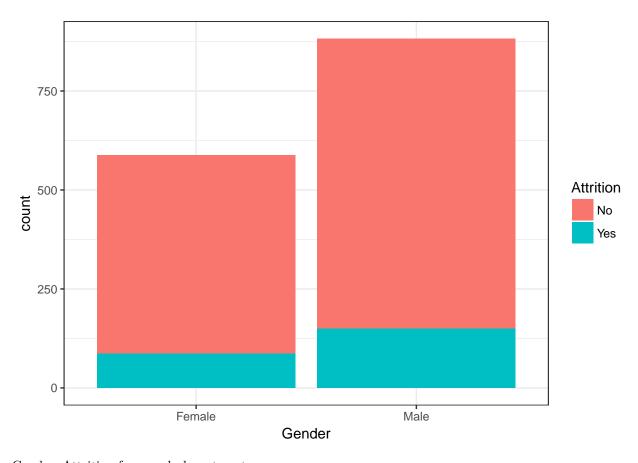
Visualization

```
require(ggplot2)
## Loading required package: ggplot2
ggplot(ibm_data, aes(x=Department, fill=Attrition))+geom_bar() + theme_bw()
```

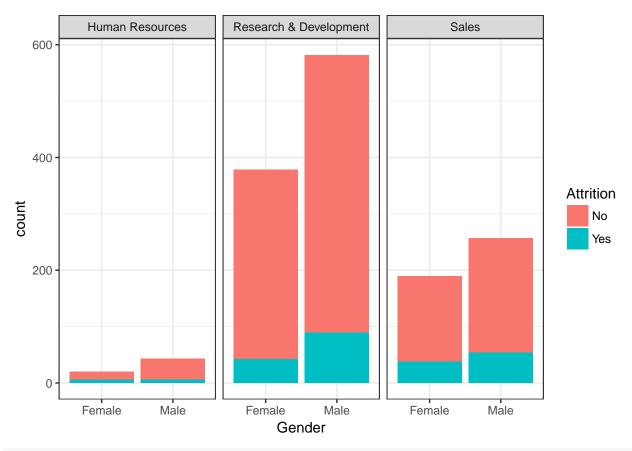


HR department has the least count. R&D department has more $Attrition = No(High\ proportion\ of\ no)$ Lets see Gender and Attrition

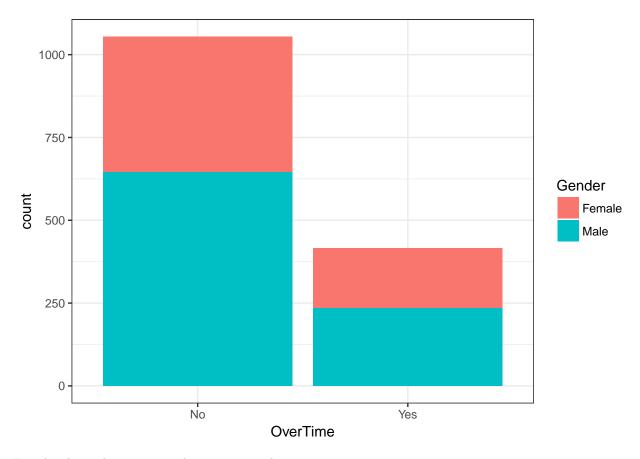
```
ggplot(ibm_data, aes(x=Gender, fill=Attrition))+geom_bar() + theme_bw()
```



Gender, Attrition from each department ggplot(ibm_data, aes(x=Gender, fill=Attrition))+geom_bar() + theme_bw() + facet_wrap(~Department)



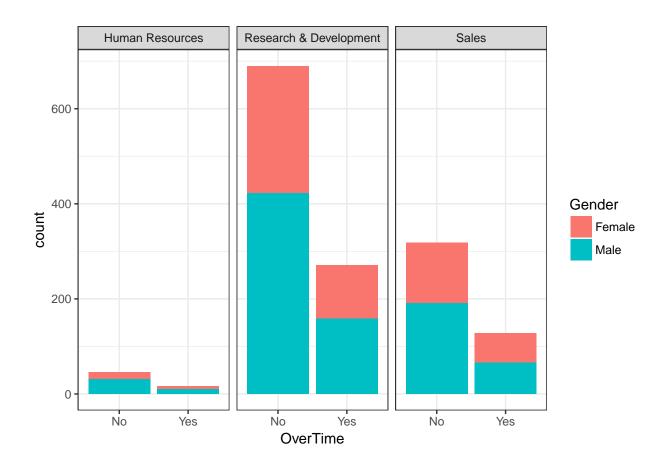
ggplot(ibm_data, aes(x=OverTime, fill=Gender))+geom_bar() + theme_bw()



Females do work overtime. There is quite a lot.

Department wise.

```
ggplot(ibm_data, aes(x=OverTime, fill=Gender))+geom_bar() + theme_bw() + facet_wrap(~Department)
```



Attrition = Yes - 1, No - 0

Use dummy_cols{fastDummies} for label encoding. I have done manually.

ibm_data\$Attrition = as.integer(ibm_data\$Attrition)
head(ibm_data)

##		ïAge At	ttrition	Busine	ssTravel	DailyRate	е	Department	
##	1	41	2	Trave	l_Rarely	1102	2	Sales	
##	2	49	1	Travel_Fr	equently	279	9 Research &	Development	
##	3	37	2	Trave	l_Rarely	1373	3 Research &	Development	
##	4	33	1	Travel_Fr	equently	1392	2 Research &	Development	
##	5	27	1	Trave	l_Rarely	59:	1 Research &	Development	
##	6	32	1	Travel_Fr	equently	100!	5 Research &	Development	
##		Distance	FromHome	Education	Educati	onField Er	mployeeCount	EmployeeNumb	er
##	1		1	2	Life S	ciences	1		1
##	2		8	1	Life S	ciences	1		2
##	3		2	2		Other	1		4
##	4		3	4	Life S	ciences	1		5
##	5		2	1		Medical	1		7
##	6		2	2	Life S	ciences	1		8
##		Environme	entSatisf	action Ge	nder Hou	rlyRate Jo	obInvolvement	t JobLevel	
##	1			2 Fe	male	94	;	3 2	
##	2			3	Male	61	2	2 2	
##	3			4	Male	92	2	2 1	

```
## 4
                             4 Female
                                                56
                                                                 3
                                                                           1
## 5
                                 Male
                                                40
                                                                 3
                                                                           1
## 6
                                                79
                                 Male
                                                                 3
##
                    JobRole JobSatisfaction MaritalStatus MonthlyIncome
## 1
           Sales Executive
                                                      Single
## 2
        Research Scientist
                                            2
                                                     Married
                                                                        5130
## 3 Laboratory Technician
                                            3
                                                      Single
                                                                        2090
        Research Scientist
                                                                        2909
                                            3
                                                     Married
## 5 Laboratory Technician
                                            2
                                                     Married
                                                                        3468
## 6 Laboratory Technician
                                                                        3068
                                            4
                                                      Single
     MonthlyRate NumCompaniesWorked Over18 OverTime PercentSalaryHike
## 1
            19479
                                     8
                                            Y
                                                    Yes
## 2
           24907
                                            Y
                                     1
                                                     No
                                                                         23
## 3
                                            Y
             2396
                                     6
                                                    Yes
                                                                         15
## 4
           23159
                                     1
                                            Y
                                                    Yes
                                                                         11
## 5
            16632
                                     9
                                            Y
                                                     No
                                                                         12
## 6
            11864
                                     0
                                            Y
                                                     No
                                                                         13
     PerformanceRating RelationshipSatisfaction StandardHours
## 1
                      3
## 2
                      4
                                                  4
                                                                80
## 3
                      3
                                                  2
                                                                80
## 4
                      3
                                                  3
                                                                80
## 5
                       3
                                                  4
                                                                80
## 6
                       3
                                                  3
     StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance
## 1
                     0
                                         8
                                                                 0
## 2
                     1
                                        10
                                                                 3
                                                                                   3
## 3
                     0
                                         7
                                                                 3
                                                                                   3
                                                                                   3
## 4
                     0
                                         8
                                                                 3
## 5
                                                                 3
                                                                                   3
                     1
                                         6
                                                                                   2
                                                                 2
## 6
                     0
                                         8
     YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion
## 1
                   6
                                        4
                                                                  0
                                        7
## 2
                  10
                                                                  1
## 3
                                        0
                                                                  0
                   0
## 4
                   8
                                        7
                                                                  3
## 5
                   2
                                        2
                                                                  2
## 6
                   7
                                                                  3
     YearsWithCurrManager
## 1
                          5
## 2
                          7
## 3
                          0
## 4
                          0
## 5
                          2
## 6
                          6
ibm_data$Attrition[ibm_data$Attrition == 1] <- 0</pre>
ibm_data$Attrition[ibm_data$Attrition == 2] <- 1</pre>
ibm_data$Attrition = as.factor(ibm_data$Attrition)
head(ibm_data)
                           BusinessTravel DailyRate
                                                                   Department
##
     i..Age Attrition
## 1
                            Travel Rarely
```

279 Research & Development

0 Travel_Frequently

2

49

```
## 3
                            Travel_Rarely
                                                1373 Research & Development
## 4
         33
                     0 Travel_Frequently
                                                1392 Research & Development
## 5
                            Travel Rarely
         27
                                                 591 Research & Development
## 6
                     0 Travel_Frequently
                                                1005 Research & Development
         32
     DistanceFromHome Education EducationField EmployeeCount EmployeeNumber
## 1
                     1
                                2 Life Sciences
## 2
                     8
                                1 Life Sciences
                                                                                2
                     2
## 3
                                2
                                            Other
                                                               1
                                                                                4
## 4
                     3
                                   Life Sciences
                                                                                5
## 5
                     2
                                1
                                          Medical
                                                                                7
## 6
                     2
                                2 Life Sciences
##
     EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel
## 1
                             2 Female
                                               94
                                                                3
## 2
                                                                 2
                                                                          2
                                 Male
                                               61
## 3
                                 Male
                                               92
                                                                 2
                                                                          1
## 4
                             4 Female
                                               56
                                                                 3
## 5
                                 Male
                                               40
                                                                 3
                                                                 3
## 6
                                 Male
                                               79
##
                    JobRole JobSatisfaction MaritalStatus MonthlyIncome
## 1
           Sales Executive
                                            4
                                                      Single
## 2
        Research Scientist
                                            2
                                                    Married
                                                                       5130
## 3 Laboratory Technician
                                            3
                                                      Single
                                                                       2090
        Research Scientist
                                                                       2909
## 4
                                            3
                                                    Married
## 5 Laboratory Technician
                                            2
                                                    Married
                                                                       3468
## 6 Laboratory Technician
                                            4
                                                                       3068
                                                      Single
     MonthlyRate NumCompaniesWorked Over18 OverTime PercentSalaryHike
## 1
           19479
                                    8
                                            Y
                                                   Yes
## 2
           24907
                                    1
                                            Y
                                                                        23
                                                    No
## 3
            2396
                                    6
                                            Y
                                                   Yes
                                                                        15
## 4
           23159
                                            Y
                                    1
                                                   Yes
                                                                        11
## 5
           16632
                                    9
                                            Y
                                                    No
                                                                        12
## 6
            11864
                                            Y
                                                     No
                                                                        13
     PerformanceRating RelationshipSatisfaction StandardHours
## 1
                      3
                                                 1
## 2
                      4
                                                 4
                                                               80
## 3
                      3
                                                 2
                                                               80
## 4
                      3
                                                 3
                                                               80
## 5
                      3
                                                 4
                                                               80
                      3
## 6
                                                 3
     StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance
                                                                0
                                         8
## 2
                     1
                                        10
                                                                 3
                                                                                  3
## 3
                     0
                                        7
                                                                 3
                                                                                  3
## 4
                     0
                                         8
                                                                 3
                                                                                  3
## 5
                     1
                                                                 3
                                                                                  3
                                                                 2
                                                                                  2
                     0
## 6
                                         8
     YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion
## 1
                   6
                                        4
                                        7
## 2
                  10
                                                                  1
## 3
                                        0
                                                                  0
                   0
## 4
                   8
                                        7
                                                                  3
                   2
                                        2
                                                                  2
## 5
                   7
## 6
                                        7
                                                                 3
     YearsWithCurrManager
```

```
## 1 5
## 2 7
## 3 0
## 4 0
## 5 2
```

Turning numeric variables into factors

ibm_data\$Education <- as.factor(ibm_data\$Education)</pre>

```
ibm data$EnvironmentSatisfaction <- as.factor(ibm data$EnvironmentSatisfaction)</pre>
ibm_data$JobInvolvement <- as.factor(ibm_data$JobInvolvement)</pre>
ibm data$JobSatisfaction <- as.factor(ibm data$JobSatisfaction)</pre>
ibm_data$PerformanceRating <- as.factor(ibm_data$PerformanceRating)</pre>
ibm_data$RelationshipSatisfaction <- as.factor(ibm_data$RelationshipSatisfaction)
ibm_data$WorkLifeBalance <- as.factor(ibm_data$WorkLifeBalance)</pre>
str(ibm data)
## 'data.frame':
                   1470 obs. of 35 variables:
                             : int 41 49 37 33 27 32 59 30 38 36 ...
## $ i..Age
## $ Attrition
                              : Factor w/ 2 levels "0", "1": 2 1 2 1 1 1 1 1 1 1 ...
## $ BusinessTravel
                             : Factor w/ 3 levels "Non-Travel", "Travel_Frequently", ..: 3 2 3 2 3 2 3 3
## $ DailyRate
                             : int 1102 279 1373 1392 591 1005 1324 1358 216 1299 ...
                             : Factor w/ 3 levels "Human Resources",..: 3 2 2 2 2 2 2 2 2 ...
## $ Department
                             : int 1 8 2 3 2 2 3 24 23 27 ...
## $ DistanceFromHome
                             : Factor w/ 5 levels "1","2","3","4",...: 2 1 2 4 1 2 3 1 3 3 ...
## $ Education
                             : Factor w/ 6 levels "Human Resources",..: 2 2 5 2 4 2 4 2 2 4 ...
## $ EducationField
## $ EmployeeCount
                             : int 111111111...
## $ EmployeeNumber
                             : int 1 2 4 5 7 8 10 11 12 13 ...
## $ EnvironmentSatisfaction : Factor w/ 4 levels "1","2","3","4": 2 3 4 4 1 4 3 4 4 3 ...
## $ Gender
                             : Factor w/ 2 levels "Female", "Male": 1 2 2 1 2 2 1 2 2 2 ...
## $ HourlyRate
                             : int 94 61 92 56 40 79 81 67 44 94 ...
## $ JobInvolvement
                             : Factor w/ 4 levels "1", "2", "3", "4": 3 2 2 3 3 3 4 3 2 3 ...
## $ JobLevel
                             : int 2 2 1 1 1 1 1 1 3 2 ...
## $ JobRole
                             : Factor w/ 9 levels "Healthcare Representative",..: 8 7 3 7 3 3 3 3 5 1
                             : Factor w/ 4 levels "1", "2", "3", "4": 4 2 3 3 2 4 1 3 3 3 ...
## $ JobSatisfaction
## $ MaritalStatus
                             : Factor w/ 3 levels "Divorced", "Married", ...: 3 2 3 2 2 3 2 1 3 2 ...
## $ MonthlyIncome
                             : int 5993 5130 2090 2909 3468 3068 2670 2693 9526 5237 ...
## $ MonthlyRate
                             : int 19479 24907 2396 23159 16632 11864 9964 13335 8787 16577 ...
## $ NumCompaniesWorked
                             : int 8 1 6 1 9 0 4 1 0 6 ...
## $ Over18
                             : Factor w/ 1 level "Y": 1 1 1 1 1 1 1 1 1 1 ...
                             : Factor w/ 2 levels "No", "Yes": 2 1 2 2 1 1 2 1 1 1 ...
## $ OverTime
## $ PercentSalaryHike
                             : int 11 23 15 11 12 13 20 22 21 13 ...
## $ PerformanceRating
                             : Factor w/ 2 levels "3", "4": 1 2 1 1 1 1 2 2 2 1 ...
## $ RelationshipSatisfaction: Factor w/4 levels "1","2","3","4": 1 4 2 3 4 3 1 2 2 2 ...
## $ StandardHours
                             : int 80 80 80 80 80 80 80 80 80 80 ...
                             : int 0 1 0 0 1 0 3 1 0 2 ...
## $ StockOptionLevel
                             : int 8 10 7 8 6 8 12 1 10 17 ...
## $ TotalWorkingYears
## $ TrainingTimesLastYear
                             : int 0 3 3 3 3 2 3 2 2 3 ...
## $ WorkLifeBalance
                             : Factor w/ 4 levels "1", "2", "3", "4": 1 3 3 3 3 2 2 3 3 2 ...
## $ YearsAtCompany
                             : int 6 10 0 8 2 7 1 1 9 7 ...
## $ YearsInCurrentRole
                             : int 4707270077...
```

```
## $ YearsSinceLastPromotion : int 0 1 0 3 2 3 0 0 1 7 ...
## $ YearsWithCurrManager : int 5 7 0 0 2 6 0 0 8 7 ...
```

head(ibm_data)

##		ïAge Attrition	BusinessTra	avel Dai	lyRate		Department	
##	1	41 1	Travel_Ran	rely	1102		Sales	
##	2	49 0	Travel_Frequer	ntly	279	Research &	Development	
##	3	37 1	Travel_Rai	•	1373	Research &	Development	
##	4	33 0	Travel_Frequer	•			Development	
##	5	27 0	Travel_Ran	-	591	Research &	Development	
##	6		Travel_Frequer	-			Development	
##		DistanceFromHome			_	ployeeCount	EmployeeNum	ber
##	1	1		fe Scien		1		1
##		8		fe Scien		1		2
##	-	2	2		her	1		4
##	_	3		e Scien		1		5
##	5	2	1	Medi		1		7
	6	2		fe Scien		1		8
##		EnvironmentSatis		Hourly				
##	1		2 Female		94		3 2	
##	_		3 Male		61		2 2	
##			4 Male		92		2 1	
	4		4 Female		56		3 1	
##	5		1 Male 4 Male		40		3 1 3 1	
##	6	To			79			
##	1	Sales Exec	bRole JobSatisi	4	Marital		5993	
##	2	Research Scie		2	1	Single Married	5130	
##	_	Laboratory Techn		3	I	Single	2090	
##	4	Research Scient		3	N	Married	2909	
##	_	Laboratory Techn		2		Married	3468	
##		Laboratory Techn		4	1	Single	3068	
##	Ü	MonthlyRate NumC			OverTin	_		
	1	19479	8	Y		es	11	
	2	24907	1	Y		Vo	23	
##	3	2396	6	Y	Υe	es	15	
	4	23159	1	Y	Υe	es	11	
##	5	16632	9	Y	ľ	No.	12	
##	6	11864	0	Y	ľ	1o	13	
##		PerformanceRatin	g RelationshipS	Satisfac	tion St	tandardHour	s	
##	1		3		1	8	0	
##	2		4		4	8	0	
##	3		3		2	8	0	
##	4		3		3	8	0	
##	5		3		4	8	0	
##	6		3		3	8	0	
##		${\tt StockOptionLevel}$	TotalWorkingYe	ears Tra	iningTi	imesLastYea	r WorkLifeBa	lance
##	1	0		8			0	1
##	2	1		10			3	3
##	3	0		7			3	3
##		0		8			3	3
##		1		6			3	3
##	6	0		8			2	2
##		YearsAtCompany Y	earsInCurrentRo	ole Year	sSincel	LastPromoti	on	

```
## 1
                  6
                                     4
                                                              0
## 2
                 10
                                     7
                                                              1
## 3
                                     0
                                                              0
                 0
## 4
                  8
                                     7
                                                              3
                  2
                                     2
                                                              2
## 5
## 6
                  7
                                     7
                                                              3
## YearsWithCurrManager
## 1
## 2
                        7
## 3
                        0
## 4
                        0
                        2
## 5
## 6
```

Gender

```
ibm_data$Gender <- as.integer(ibm_data$Gender)
ibm_data$MaritalStatus <- as.integer(ibm_data$MaritalStatus)
ibm_data$OverTime <- as.integer(ibm_data$OverTime)
head(ibm_data)</pre>
```

##		ïAge Attrition	Busine	ssTrav	el DailyR	ate		Department	
##	1	41 1	Trave	l_Rare	ly 1	102		Sales	
##	2	49 0	Travel_Fr	equent	ly :	279 Re	esearch & l	Development	
##	3	37 1	Trave	l_Rare	ly 13	373 Re	esearch & l	Development	
##	4	33 0	Travel_Fr	equent	ly 13	392 Re	esearch & l	Development	
##	5	27 0	Trave	l_Rare	ly	591 Re	esearch & l	Development	
##	6	32 0	Travel_Fr	equent	ly 1	005 Re	esearch & l	Development	
##		DistanceFromHome	Education	Educa	tionField	Emplo	oyeeCount 1	EmployeeNumb	er
##	1	1	2	Life	Sciences		1		1
##	2	8	1	Life	Sciences		1		2
##	3	2	2		Other		1		4
##	4	3	4	Life	Sciences		1		5
##	5	2	1		Medical		1		7
##	6	2	2		Sciences		1		8
##		EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel					JobLevel		
	1		2	1	94		3	2	
	2		3	2	61		2	2	
##	3		4	2	92		2	1	
##	_		4	1	56		3	1	
##	5		1	2	40		3	1	
	6		4	2	79		3	1	
##				atisfa		italSt	tatus Montl	•	
##	1	Sales Exec			4		3	5993	
	2	Research Scie			2		2	5130	
##		Laboratory Techn			3		3	2090	
	4	Research Scie			3		2	2909	
##		Laboratory Techn			2		2	3468	
	6	Laboratory Techn			4	m ·	3	3068	
##	4	MonthlyRate NumC	ompaniesWo				rercentSa.	•	
	1	19479		8	Y	2		11	
##	2	24907		1	Y	1		23	

```
## 3
             2396
                                     6
                                                                         15
## 4
            23159
                                     1
                                            Υ
                                                      2
                                                                         11
                                     9
## 5
            16632
                                            Y
                                                      1
                                                                         12
## 6
            11864
                                     0
                                            Y
                                                      1
                                                                         13
##
     PerformanceRating RelationshipSatisfaction StandardHours
## 1
                       3
                                                  1
## 2
                       4
                                                  4
                                                                80
## 3
                       3
                                                  2
                                                                80
## 4
                       3
                                                  3
                                                                80
## 5
                       3
                                                  4
                                                                80
## 6
                       3
                                                  3
                                                                80
##
     StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance
## 1
                     0
                                         8
                                                                 0
                                                                                  1
## 2
                                        10
                                                                 3
                                                                                  3
                     1
## 3
                     0
                                         7
                                                                 3
                                                                                  3
## 4
                     0
                                         8
                                                                 3
                                                                                  3
## 5
                     1
                                         6
                                                                 3
                                                                                  3
                     0
                                                                                  2
## 6
                                         8
                                                                 2
##
     YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion
## 1
                   6
                                        4
## 2
                  10
                                        7
                                                                  1
## 3
                   0
                                        0
                                                                  0
                   8
                                        7
                                                                  3
## 4
## 5
                   2
                                        2
                                                                  2
                                                                  3
## 6
                   7
                                        7
     YearsWithCurrManager
## 1
## 2
                          7
## 3
                          0
## 4
                          0
## 5
                          2
## 6
                          6
ibm_data$Gender <- as.factor(ibm_data$Gender)</pre>
ibm_data$MaritalStatus <- as.factor(ibm_data$MaritalStatus)</pre>
ibm_data$0verTime <- as.factor(ibm_data$0verTime)</pre>
head(ibm_data)
##
     ï..Age Attrition
                           BusinessTravel DailyRate
                                                                   Department
## 1
                            Travel_Rarely
                                                                         Sales
## 2
         49
                     0 Travel_Frequently
                                                  279 Research & Development
## 3
         37
                            Travel_Rarely
                                                1373 Research & Development
                     1
## 4
         33
                     0 Travel_Frequently
                                                1392 Research & Development
## 5
         27
                            Travel_Rarely
                                                  591 Research & Development
## 6
         32
                     0 Travel_Frequently
                                                 1005 Research & Development
     DistanceFromHome Education EducationField EmployeeCount EmployeeNumber
## 1
                     1
                                2 Life Sciences
                                                                                1
## 2
                     8
                                1
                                   Life Sciences
                                                                1
                                                                                2
                     2
## 3
                                2
                                            Other
                                                                                4
                                                                1
                     3
## 4
                                4
                                   Life Sciences
                                                                1
                                                                                5
## 5
                     2
                                                                                7
                                          Medical
## 6
                     2
                                2 Life Sciences
     EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel
```

94

61

3

2

2

2

2

3

1

2

1

2

```
## 3
                                                92
## 4
                                     1
                                                56
                                                                  3
                                                                            1
## 5
                                     2
                                                40
                                                                  3
## 6
                                     2
                                                79
                                                                            1
                     JobRole JobSatisfaction MaritalStatus MonthlyIncome
## 1
            Sales Executive
                                             4
                                                            3
        Research Scientist
                                             2
                                                            2
                                                                        5130
## 3 Laboratory Technician
                                                            3
                                                                        2090
                                             3
        Research Scientist
                                             3
                                                            2
                                                                        2909
## 5 Laboratory Technician
                                                            2
                                                                        3468
## 6 Laboratory Technician
                                             4
                                                                        3068
##
     MonthlyRate NumCompaniesWorked Over18 OverTime PercentSalaryHike
## 1
            19479
                                     8
                                             Y
## 2
                                             Y
            24907
                                     1
                                                       1
                                                                         23
## 3
             2396
                                     6
                                             Y
                                                       2
                                                                         15
## 4
            23159
                                             Y
                                                                         11
## 5
            16632
                                             Y
                                                       1
                                                                         12
            11864
                                             Y
## 6
                                                       1
                                                                         13
     PerformanceRating RelationshipSatisfaction StandardHours
##
## 1
                       3
## 2
                       4
                                                  4
                                                                80
## 3
                       3
                                                  2
                                                                80
## 4
                       3
                                                  3
                                                                80
## 5
                       3
                                                                80
## 6
                       3
                                                  3
                                                                80
     StockOptionLevel TotalWorkingYears TrainingTimesLastYear WorkLifeBalance
## 1
                     0
## 2
                     1
                                        10
                                                                  3
                                                                                   3
## 3
                     0
                                                                                   3
                                         7
                                                                  3
## 4
                     0
                                         8
                                                                  3
                                                                                   3
                                                                                   3
## 5
                     1
                                                                  3
## 6
                     0
                                                                  2
                                                                                   2
     YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion
## 1
                   6
                                        4
                                        7
## 2
                  10
                                                                   1
## 3
                                        0
                                                                   0
                   0
## 4
                   8
                                        7
                                                                   3
## 5
                   2
                                        2
                                                                   2
                                                                   3
## 6
##
     YearsWithCurrManager
## 2
## 3
                          0
## 4
                          0
## 5
                          2
                          6
## 6
```

Train Test Split

```
require(caret)
```

Loading required package: caret

```
## Loading required package: lattice
set.seed(1)
ind = createDataPartition(ibm_data$Attrition, p=0.80, list = F)
train = ibm_data[ind,]
test = ibm_data[-ind,]
```

Logistic Model

```
\verb|model <- glm(Attrition ~DailyRate+EnvironmentSatisfaction+JobInvolvement+RelationshipSatisfaction , dat| | data | dat
summary(model)
##
## Call:
## glm(formula = Attrition ~ DailyRate + EnvironmentSatisfaction +
                JobInvolvement + RelationshipSatisfaction, family = "binomial",
                data = train)
##
##
## Deviance Residuals:
                                                Median
                                      1Q
                                                                                                   Max
## -1.4632 -0.5991 -0.4977 -0.4105
                                                                                             2.4402
## Coefficients:
                                                                        Estimate Std. Error z value Pr(>|z|)
                                                                     0.9345089 0.3660083 2.553 0.01067 *
## (Intercept)
## DailyRate
                                                                   -0.0004937 0.0002033 -2.428 0.01518 *
## EnvironmentSatisfaction2 -0.6771158 0.2430255 -2.786 0.00533 **
## EnvironmentSatisfaction3 -0.9069009 0.2216922 -4.091 4.30e-05 ***
## EnvironmentSatisfaction4 -0.9430097 0.2228605 -4.231 2.32e-05 ***
## JobInvolvement2
                                                                 -0.9906657  0.3019406  -3.281  0.00103 **
## JobInvolvement3
                                                                  ## JobInvolvement4
                                                                   -1.8239490 0.4117666
                                                                                                                      -4.430 9.44e-06 ***
## RelationshipSatisfaction2 -0.4745642 0.2454090
                                                                                                                       -1.934 0.05314 .
## RelationshipSatisfaction3 -0.5148690 0.2244683 -2.294 0.02181 *
## RelationshipSatisfaction4 -0.5869789 0.2314739 -2.536 0.01122 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
                Null deviance: 1040.54 on 1176 degrees of freedom
## Residual deviance: 978.95 on 1166 degrees of freedom
## AIC: 1000.9
## Number of Fisher Scoring iterations: 5
```

Prediction

```
prediction <- predict(model, newdata = test, type = 'response')
head(prediction)</pre>
```

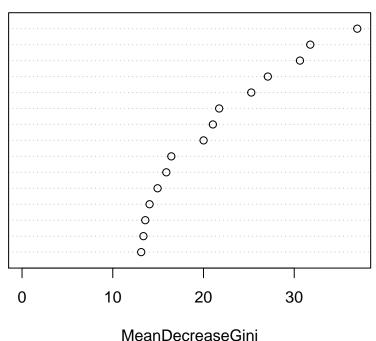
```
##
                                           15
## 0.14139655 0.22474550 0.10183574 0.18415052 0.09359310 0.09795004
prediction <- ifelse(prediction > 0.5,1,0)
head(prediction)
## 11 13 14 15 31 34
## 0 0 0 0 0 0
tab = table(predicted = prediction, original = test$Attrition)
##
            original
## predicted 0
          0 245
                 46
##
              1
print(sum(diag(tab))/sum(tab))
## [1] 0.8395904
```

Lets try random forest

```
require(randomForest)
## Loading required package: randomForest
## randomForest 4.6-12
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:ggplot2':
##
##
       margin
model_rf <- randomForest(Attrition ~DailyRate+EnvironmentSatisfaction+JobInvolvement+</pre>
                           RelationshipSatisfaction+Education+MonthlyIncome+MonthlyRate+
                           PercentSalaryHike+TotalWorkingYears+YearsAtCompany+
                           YearsInCurrentRole+YearsWithCurrManager+NumCompaniesWorked+
                           JobRole+HourlyRate,
                           data=train)
varImpPlot(model_rf)
```

model_rf

MonthlyIncome
DailyRate
MonthlyRate
HourlyRate
TotalWorkingYears
JobRole
PercentSalaryHike
YearsAtCompany
NumCompaniesWorked
EnvironmentSatisfaction
YearsWithCurrManager
RelationshipSatisfaction
JobInvolvement
Education
YearsInCurrentRole



Prediction

```
prediction <- predict(model_rf, newdata = test)
head(prediction)

## 11 13 14 15 31 34
## 0 0 0 0 0 0 0
## Levels: 0 1</pre>
```

Accuracy

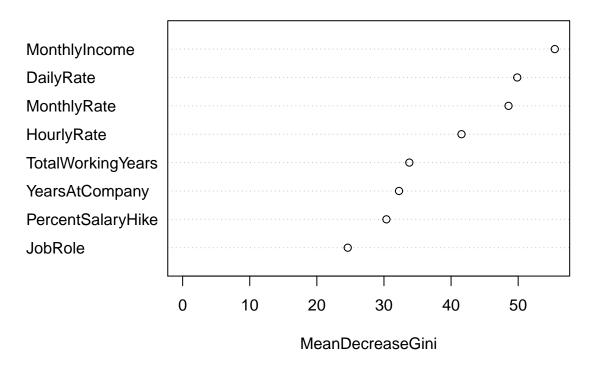
```
library(caret)
pl1 <- data.frame(original = test, predicted = prediction)
confusionMatrix(table(pl1$original.Attrition,pl1$predicted))

## Confusion Matrix and Statistics
##
##
## 0 1
## 0 246 0
## 1 42 5
##</pre>
```

```
##
                  Accuracy : 0.8567
##
                    95% CI : (0.8112, 0.8947)
       No Information Rate: 0.9829
##
##
       P-Value [Acc > NIR] : 1
##
##
                     Kappa : 0.1666
   Mcnemar's Test P-Value : 2.509e-10
##
##
##
               Sensitivity: 0.8542
##
               Specificity: 1.0000
##
            Pos Pred Value : 1.0000
##
            Neg Pred Value: 0.1064
##
                Prevalence: 0.9829
            Detection Rate: 0.8396
##
##
      Detection Prevalence : 0.8396
##
         Balanced Accuracy : 0.9271
##
##
          'Positive' Class : 0
##
```

Drop some variables and see how accuracy changes.

model_rf



Prediction

```
prediction <- predict(model_rf, newdata = test)
head(prediction)

## 11 13 14 15 31 34
## 0 0 0 0 0 0
## Levels: 0 1</pre>
```

accuracy

```
library(caret)
pl1 <- data.frame(original = test, predicted = prediction)
confusionMatrix(table(pl1$original.Attrition,pl1$predicted))

## Confusion Matrix and Statistics
##
##
## 0 1
## 0 244 2
## 1 43 4
##</pre>
```

```
##
                  Accuracy : 0.8464
##
                    95% CI : (0.7999, 0.8857)
       No Information Rate : 0.9795
##
##
       P-Value [Acc > NIR] : 1
##
##
                     Kappa: 0.1189
    Mcnemar's Test P-Value : 2.479e-09
##
##
               Sensitivity: 0.85017
##
##
               Specificity: 0.66667
##
            Pos Pred Value: 0.99187
##
            Neg Pred Value : 0.08511
##
                Prevalence: 0.97952
            Detection Rate: 0.83276
##
##
      Detection Prevalence : 0.83959
##
         Balanced Accuracy : 0.75842
##
##
          'Positive' Class : 0
##
```

Therefore, not much difference in accuracy.