Untitled

2023-11-04

#loading necessary packages

```
library(party)
```

```
## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich
```

```
library(randomForest)
```

```
## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.
```

```
library(cluster)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:randomForest':
##
##     margin
```

```
library(caret)
```

```
## Loading required package: lattice
```

#loading dataset

```
hr<-read.csv("/Users/jannyvelazquez/Desktop/HR_Comma_sep.csv")
head(hr)
```

```
##   satisfaction_level last_evaluation number_project average_montly_hours
## 1               0.38            0.53              2                  157
## 2               0.80            0.86              5                  262
## 3               0.11            0.88              7                  272
## 4               0.72            0.87              5                  223
## 5               0.37            0.52              2                  159
## 6               0.41            0.50              2                  153
##   time_spend_company Work_accident left promotion_last_5years sales salary
## 1                  3             0    1                     0 sales    low
## 2                  6             0    1                     0 sales medium
## 3                  4             0    1                     0 sales medium
## 4                  5             0    1                     0 sales    low
## 5                  3             0    1                     0 sales    low
## 6                  3             0    1                     0 sales    low
```

#dividing data into training and testing.

```
index_hr <- createDataPartition(hr$left, p = 0.7, list = FALSE)

# Split the data into training (70%) and testing (30%) sets
train_hr <- hr[index_hr, ]
test_hr <- hr[-index_hr, ]
```

#making a full multiple lnear regression model with dependent variable “left”

```
full_model <- lm(formula = left~satisfaction_level + last_evaluation+ number_project + average_montly_h
summary(full_model)
```

```
##
## Call:
## lm(formula = left ~ satisfaction_level + last_evaluation + number_project +
##     average_montly_hours + time_spend_company + Work_accident +
##     promotion_last_5years + salary + sales, data = test_hr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8358 -0.2633 -0.1025  0.2928  1.0839
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             0.3927398  0.0453191   8.666  < 2e-16 ***
## satisfaction_level     -0.6563524  0.0236768 -27.721  < 2e-16 ***
```

```
## last_evaluation        0.1074178  0.0374399   2.869 0.004136 **
## number_project        -0.0382116  0.0054253  -7.043 2.17e-12 ***
## average_montly_hours   0.0005624  0.0001291   4.356 1.36e-05 ***
## time_spend_company     0.0408258  0.0040900   9.982  < 2e-16 ***
## Work_accident         -0.1645131  0.0160735 -10.235  < 2e-16 ***
## promotion_last_5years -0.1348184  0.0400566  -3.366 0.000770 ***
## salarylow              0.1829887  0.0223053   8.204 3.01e-16 ***
## salarymedium           0.0856636  0.0223769   3.828 0.000131 ***
## saleshr                0.0163903  0.0367157   0.446 0.655321
## salesIT               -0.0503119  0.0317115  -1.587 0.112685
## salesmanagement       -0.0840531  0.0375176  -2.240 0.025116 *
## salesmarketing         0.0214338  0.0347327   0.617 0.537195
## salesproduct_mng      -0.0431979  0.0343403  -1.258 0.208480
## salesRandD            -0.1095398  0.0357749  -3.062 0.002212 **
## salessales            -0.0442731  0.0271328  -1.632 0.102809
## salessupport           0.0081366  0.0289298   0.281 0.778528
## salestechnical        -0.0064454  0.0281929  -0.229 0.819175
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3831 on 4480 degrees of freedom
## Multiple R-squared:  0.2259, Adjusted R-squared:  0.2228
## F-statistic: 72.62 on 18 and 4480 DF,  p-value: < 2.2e-16
```

#making a reduced multiple lnear regression model with dependent variable "left"

```
reduced_model <- lm(formula = left~satisfaction_level + average_montly_hours + promotion_last_5years + s
summary(reduced_model)
```

```
##
## Call:
## lm(formula = left ~ satisfaction_level + average_montly_hours +
##     promotion_last_5years + salary + sales, data = test_hr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6849 -0.2632 -0.1252  0.2734  1.0517
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.4668739  0.0421011  11.089  < 2e-16 ***
## satisfaction_level    -0.6619370  0.0235308 -28.131  < 2e-16 ***
## average_montly_hours   0.0004747  0.0001159   4.097 4.25e-05 ***
## promotion_last_5years -0.1243422  0.0409711  -3.035 0.002420 **
## salarylow              0.1787618  0.0228762   7.814 6.84e-15 ***
## salarymedium           0.0863893  0.0229547   3.763 0.000170 ***
## saleshr                0.0180657  0.0376625   0.480 0.631483
## salesIT               -0.0520616  0.0325373  -1.600 0.109656
## salesmanagement       -0.0807953  0.0383781  -2.105 0.035325 *
## salesmarketing         0.0220248  0.0356146   0.618 0.536329
## salesproduct_mng      -0.0498076  0.0352179  -1.414 0.157352
## salesRandD            -0.1217470  0.0366948  -3.318 0.000914 ***
## salessales            -0.0402857  0.0278256  -1.448 0.147745
```

```
## salessupport           0.0027846  0.0296692   0.094 0.925230
## salestechnical        -0.0126810  0.0289190  -0.439 0.661045
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.393 on 4484 degrees of freedom
## Multiple R-squared:  0.1842, Adjusted R-squared:  0.1817
## F-statistic: 72.34 on 14 and 4484 DF,  p-value: < 2.2e-16
```

#making a full model with interactions

```
full_model_interact <- lm(formula = left~.*., data = test_hr)
summary(full_model_interact)
```

```
##
## Call:
## lm(formula = left ~ . * ., data = test_hr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30881 -0.14224 -0.00827  0.13843  1.17346
##
## Coefficients: (1 not defined because of singularities)
##                                          Estimate Std. Error t value
## (Intercept)                             6.080e+00  1.783e-01  34.096
## satisfaction_level                     -1.931e+00  1.409e-01 -13.708
## last_evaluation                        -4.526e+00  2.194e-01 -20.627
## number_project                         -8.754e-01  3.128e-02 -27.984
## average_montly_hours                   -1.516e-02  7.839e-04 -19.335
## time_spend_company                     -4.004e-01  2.795e-02 -14.327
## Work_accident                          -8.080e-02  1.029e-01  -0.785
## promotion_last_5years                  -2.017e-01  2.742e-01  -0.736
## saleshr                                -4.979e-02  1.787e-01  -0.279
## salesIT                                -4.697e-02  1.530e-01  -0.307
## salesmanagement                         1.895e-01  1.947e-01   0.974
## salesmarketing                          3.120e-01  1.560e-01   2.000
## salesproduct_mng                       -2.669e-02  1.577e-01  -0.169
## salesRandD                             -3.440e-01  1.820e-01  -1.890
## salessales                              6.641e-02  1.268e-01   0.524
## salessupport                           -5.723e-02  1.373e-01  -0.417
## salestechnical                          4.414e-02  1.313e-01   0.336
## salarylow                               1.447e-01  1.230e-01   1.176
## salarymedium                            1.201e-01  1.242e-01   0.967
## satisfaction_level:last_evaluation      1.052e+00  1.203e-01   8.749
## satisfaction_level:number_project       4.450e-02  1.670e-02   2.664
## satisfaction_level:average_montly_hours 1.578e-03  4.072e-04   3.875
## satisfaction_level:time_spend_company   1.423e-01  1.383e-02  10.292
## satisfaction_level:Work_accident        1.102e-01  5.181e-02   2.127
## satisfaction_level:promotion_last_5years 2.149e-01  1.371e-01   1.568
## satisfaction_level:saleshr              8.988e-02  1.131e-01   0.795
## satisfaction_level:salesIT              3.630e-02  9.455e-02   0.384
## satisfaction_level:salesmanagement      2.613e-04  1.200e-01   0.002
## satisfaction_level:salesmarketing       1.676e-02  1.088e-01   0.154
```

```
## satisfaction_level:salesproduct_mng         1.274e-01  1.031e-01   1.235
## satisfaction_level:salesRandD                3.750e-01  1.106e-01   3.392
## satisfaction_level:salessales                3.701e-03  8.155e-02   0.045
## satisfaction_level:salessupport              4.597e-02  8.760e-02   0.525
## satisfaction_level:salestechnical           -5.492e-03  8.400e-02  -0.065
## satisfaction_level:salarylow                -1.546e-01  7.134e-02  -2.168
## satisfaction_level:salarymedium             -9.753e-02  7.205e-02  -1.354
## last_evaluation:number_project               5.015e-01  2.492e-02  20.124
## last_evaluation:average_montly_hours         7.439e-03  5.704e-04  13.043
## last_evaluation:time_spend_company           1.673e-01  1.841e-02   9.087
## last_evaluation:Work_accident               -5.161e-02  7.354e-02  -0.702
## last_evaluation:promotion_last_5years       -8.166e-02  2.051e-01  -0.398
## last_evaluation:saleshr                     -3.586e-01  1.822e-01  -1.968
## last_evaluation:salesIT                     -6.107e-02  1.544e-01  -0.395
## last_evaluation:salesmanagement             -1.118e-01  1.904e-01  -0.587
## last_evaluation:salesmarketing              -2.198e-01  1.753e-01  -1.254
## last_evaluation:salesproduct_mng            -1.706e-01  1.632e-01  -1.046
## last_evaluation:salesRandD                   1.596e-02  1.746e-01   0.091
## last_evaluation:salessales                  -1.484e-01  1.339e-01  -1.108
## last_evaluation:salessupport                -8.152e-02  1.439e-01  -0.567
## last_evaluation:salestechnical              -2.489e-01  1.397e-01  -1.782
## last_evaluation:salarylow                    1.825e-01  1.077e-01   1.695
## last_evaluation:salarymedium                 1.461e-01  1.077e-01   1.356
## number_project:average_montly_hours          1.899e-03  8.338e-05  22.778
## number_project:time_spend_company            3.498e-02  3.108e-03  11.254
## number_project:Work_accident                 1.805e-02  1.199e-02   1.506
## number_project:promotion_last_5years         1.156e-03  3.228e-02   0.036
## number_project:saleshr                       2.317e-03  2.701e-02   0.086
## number_project:salesIT                       4.259e-03  2.204e-02   0.193
## number_project:salesmanagement               2.061e-02  2.447e-02   0.842
## number_project:salesmarketing               -7.106e-03  2.404e-02  -0.296
## number_project:salesproduct_mng              2.438e-02  2.281e-02   1.069
## number_project:salesRandD                    6.648e-03  2.590e-02   0.257
## number_project:salessales                    1.418e-03  1.851e-02   0.077
## number_project:salessupport                  1.618e-02  2.002e-02   0.808
## number_project:salestechnical                4.291e-02  1.915e-02   2.241
## number_project:salarylow                    -4.058e-02  1.558e-02  -2.604
## number_project:salarymedium                 -3.120e-02  1.562e-02  -1.998
## average_montly_hours:time_spend_company      5.146e-04  6.492e-05   7.926
## average_montly_hours:Work_accident          -1.743e-05  2.540e-04  -0.069
## average_montly_hours:promotion_last_5years -4.074e-04  6.575e-04  -0.620
## average_montly_hours:saleshr                 2.948e-04  6.154e-04   0.479
## average_montly_hours:salesIT                 5.313e-04  5.158e-04   1.030
## average_montly_hours:salesmanagement        -7.666e-04  6.189e-04  -1.239
## average_montly_hours:salesmarketing          2.394e-04  5.937e-04   0.403
## average_montly_hours:salesproduct_mng       -1.285e-04  5.494e-04  -0.234
## average_montly_hours:salesRandD              4.257e-04  5.688e-04   0.748
## average_montly_hours:salessales              2.150e-04  4.449e-04   0.483
## average_montly_hours:salessupport           -7.654e-05  4.704e-04  -0.163
## average_montly_hours:salestechnical          1.885e-04  4.659e-04   0.405
## average_montly_hours:salarylow               1.548e-04  3.677e-04   0.421
## average_montly_hours:salarymedium            2.550e-04  3.666e-04   0.695
## time_spend_company:Work_accident            -1.721e-02  7.782e-03  -2.212
## time_spend_company:promotion_last_5years    -3.707e-02  1.741e-02  -2.130
```

```
## time_spend_company:saleshr                 3.900e-02  2.382e-02   1.637
## time_spend_company:salesIT                -1.738e-02  1.924e-02  -0.903
## time_spend_company:salesmanagement        -2.818e-02  1.914e-02  -1.473
## time_spend_company:salesmarketing         -1.774e-02  1.990e-02  -0.891
## time_spend_company:salesproduct_mng       -3.163e-02  1.992e-02  -1.587
## time_spend_company:salesRandD              4.284e-03  2.361e-02   0.181
## time_spend_company:salessales             -2.330e-02  1.672e-02  -1.394
## time_spend_company:salessupport            1.488e-02  1.790e-02   0.831
## time_spend_company:salestechnical         -6.928e-03  1.759e-02  -0.394
## time_spend_company:salarylow               8.984e-03  1.053e-02   0.853
## time_spend_company:salarymedium           -1.375e-02  1.024e-02  -1.343
## Work_accident:promotion_last_5years        9.352e-02  7.334e-02   1.275
## Work_accident:saleshr                     -6.007e-02  8.052e-02  -0.746
## Work_accident:salesIT                      3.932e-02  7.006e-02   0.561
## Work_accident:salesmanagement              3.756e-02  7.811e-02   0.481
## Work_accident:salesmarketing               2.829e-02  7.262e-02   0.390
## Work_accident:salesproduct_mng             1.128e-02  7.304e-02   0.154
## Work_accident:salesRandD                   2.091e-02  7.504e-02   0.279
## Work_accident:salessales                  -5.835e-04  5.998e-02  -0.010
## Work_accident:salessupport                -4.621e-02  6.261e-02  -0.738
## Work_accident:salestechnical               1.327e-02  6.189e-02   0.214
## Work_accident:salarylow                   -7.282e-02  4.563e-02  -1.596
## Work_accident:salarymedium                -1.508e-02  4.615e-02  -0.327
## promotion_last_5years:saleshr              1.718e-01  2.054e-01   0.837
## promotion_last_5years:salesIT              5.830e-01  2.565e-01   2.273
## promotion_last_5years:salesmanagement      4.261e-01  1.509e-01   2.824
## promotion_last_5years:salesmarketing       2.070e-01  1.632e-01   1.269
## promotion_last_5years:salesproduct_mng            NA         NA      NA
## promotion_last_5years:salesRandD           3.745e-02  1.649e-01   0.227
## promotion_last_5years:salessales           2.429e-01  1.497e-01   1.622
## promotion_last_5years:salessupport         9.218e-02  1.764e-01   0.522
## promotion_last_5years:salestechnical       5.397e-01  2.036e-01   2.651
## promotion_last_5years:salarylow           -1.062e-02  1.080e-01  -0.098
## promotion_last_5years:salarymedium        -2.329e-02  9.444e-02  -0.247
## saleshr:salarylow                          5.678e-02  1.055e-01   0.538
## salesIT:salarylow                         -4.230e-02  9.091e-02  -0.465
## salesmanagement:salarylow                  7.710e-02  8.774e-02   0.879
## salesmarketing:salarylow                  -1.218e-01  9.893e-02  -1.231
## salesproduct_mng:salarylow                 1.070e-01  8.982e-02   1.192
## salesRandD:salarylow                      -1.252e-01  1.057e-01  -1.184
## salessales:salarylow                       8.762e-02  7.599e-02   1.153
## salessupport:salarylow                     3.061e-02  8.381e-02   0.365
## salestechnical:salarylow                  -6.717e-02  7.796e-02  -0.862
## saleshr:salarymedium                       7.980e-02  1.049e-01   0.760
## salesIT:salarymedium                      -3.108e-02  9.065e-02  -0.343
## salesmanagement:salarymedium               1.294e-02  8.199e-02   0.158
## salesmarketing:salarymedium               -1.345e-01  9.808e-02  -1.371
## salesproduct_mng:salarymedium              7.994e-02  8.990e-02   0.889
## salesRandD:salarymedium                   -6.982e-02  1.056e-01  -0.661
## salessales:salarymedium                    1.796e-02  7.539e-02   0.238
## salessupport:salarymedium                 -2.970e-02  8.353e-02  -0.356
## salestechnical:salarymedium               -4.008e-02  7.748e-02  -0.517
##                                           Pr(>|t|)
## (Intercept)                                < 2e-16 ***
```

```
## satisfaction_level                           < 2e-16 ***
## last_evaluation                              < 2e-16 ***
## number_project                               < 2e-16 ***
## average_montly_hours                         < 2e-16 ***
## time_spend_company                           < 2e-16 ***
## Work_accident                               0.432399
## promotion_last_5years                       0.462018
## saleshr                                     0.780540
## salesIT                                     0.758956
## salesmanagement                             0.330342
## salesmarketing                              0.045539 *
## salesproduct_mng                            0.865656
## salesRandD                                  0.058833 .
## salessales                                  0.600483
## salessupport                                0.676829
## salestechnical                              0.736779
## salarylow                                   0.239467
## salarymedium                                0.333383
## satisfaction_level:last_evaluation           < 2e-16 ***
## satisfaction_level:number_project           0.007745 **
## satisfaction_level:average_montly_hours     0.000108 ***
## satisfaction_level:time_spend_company        < 2e-16 ***
## satisfaction_level:Work_accident            0.033494 *
## satisfaction_level:promotion_last_5years    0.116910
## satisfaction_level:saleshr                  0.426660
## satisfaction_level:salesIT                  0.701040
## satisfaction_level:salesmanagement          0.998263
## satisfaction_level:salesmarketing           0.877535
## satisfaction_level:salesproduct_mng         0.216838
## satisfaction_level:salesRandD               0.000701 ***
## satisfaction_level:salessales               0.963801
## satisfaction_level:salessupport             0.599818
## satisfaction_level:salestechnical           0.947875
## satisfaction_level:salarylow                0.030245 *
## satisfaction_level:salarymedium             0.175943
## last_evaluation:number_project               < 2e-16 ***
## last_evaluation:average_montly_hours         < 2e-16 ***
## last_evaluation:time_spend_company           < 2e-16 ***
## last_evaluation:Work_accident               0.482845
## last_evaluation:promotion_last_5years       0.690526
## last_evaluation:saleshr                     0.049105 *
## last_evaluation:salesIT                     0.692508
## last_evaluation:salesmanagement             0.557098
## last_evaluation:salesmarketing              0.210027
## last_evaluation:salesproduct_mng            0.295799
## last_evaluation:salesRandD                  0.927156
## last_evaluation:salessales                  0.267861
## last_evaluation:salessupport                0.570999
## last_evaluation:salestechnical              0.074761 .
## last_evaluation:salarylow                   0.090232 .
## last_evaluation:salarymedium                0.175226
## number_project:average_montly_hours          < 2e-16 ***
## number_project:time_spend_company            < 2e-16 ***
## number_project:Work_accident                0.132207
```

```
## number_project:promotion_last_5years       0.971439
## number_project:saleshr                      0.931663
## number_project:salesIT                      0.846776
## number_project:salesmanagement              0.399619
## number_project:salesmarketing               0.767568
## number_project:salesproduct_mng             0.285235
## number_project:salesRandD                   0.797425
## number_project:salessales                   0.938946
## number_project:salessupport                 0.419152
## number_project:salestechnical               0.025048 *
## number_project:salarylow                    0.009240 **
## number_project:salarymedium                 0.045772 *
## average_montly_hours:time_spend_company     2.85e-15 ***
## average_montly_hours:Work_accident          0.945287
## average_montly_hours:promotion_last_5years 0.535575
## average_montly_hours:saleshr                0.631924
## average_montly_hours:salesIT                0.303085
## average_montly_hours:salesmanagement        0.215561
## average_montly_hours:salesmarketing         0.686857
## average_montly_hours:salesproduct_mng       0.815129
## average_montly_hours:salesRandD             0.454234
## average_montly_hours:salessales             0.628931
## average_montly_hours:salessupport           0.870747
## average_montly_hours:salestechnical         0.685853
## average_montly_hours:salarylow              0.673838
## average_montly_hours:salarymedium           0.486789
## time_spend_company:Work_accident            0.027008 *
## time_spend_company:promotion_last_5years    0.033238 *
## time_spend_company:saleshr                  0.101686
## time_spend_company:salesIT                  0.366486
## time_spend_company:salesmanagement          0.140955
## time_spend_company:salesmarketing           0.372951
## time_spend_company:salesproduct_mng         0.112510
## time_spend_company:salesRandD               0.856041
## time_spend_company:salessales               0.163419
## time_spend_company:salessupport             0.405774
## time_spend_company:salestechnical           0.693642
## time_spend_company:salarylow                0.393483
## time_spend_company:salarymedium             0.179253
## Work_accident:promotion_last_5years         0.202300
## Work_accident:saleshr                       0.455698
## Work_accident:salesIT                       0.574625
## Work_accident:salesmanagement               0.630626
## Work_accident:salesmarketing                0.696849
## Work_accident:salesproduct_mng              0.877325
## Work_accident:salesRandD                    0.780582
## Work_accident:salessales                    0.992238
## Work_accident:salessupport                  0.460472
## Work_accident:salestechnical                0.830277
## Work_accident:salarylow                     0.110570
## Work_accident:salarymedium                  0.743890
## promotion_last_5years:saleshr               0.402776
## promotion_last_5years:salesIT               0.023069 *
## promotion_last_5years:salesmanagement       0.004769 **
```

```
## promotion_last_5years:salesmarketing       0.204642
## promotion_last_5years:salesproduct_mng           NA
## promotion_last_5years:salesRandD           0.820356
## promotion_last_5years:salessales           0.104781
## promotion_last_5years:salessupport         0.601398
## promotion_last_5years:salestechnical       0.008048 **
## promotion_last_5years:salarylow            0.921677
## promotion_last_5years:salarymedium         0.805191
## saleshr:salarylow                          0.590373
## salesIT:salarylow                          0.641734
## salesmanagement:salarylow                  0.379598
## salesmarketing:salarylow                   0.218473
## salesproduct_mng:salarylow                 0.233485
## salesRandD:salarylow                       0.236423
## salessales:salarylow                       0.248919
## salessupport:salarylow                     0.714967
## salestechnical:salarylow                   0.388962
## saleshr:salarymedium                       0.447014
## salesIT:salarymedium                       0.731713
## salesmanagement:salarymedium               0.874597
## salesmarketing:salarymedium                0.170350
## salesproduct_mng:salarymedium              0.373915
## salesRandD:salarymedium                    0.508503
## salessales:salarymedium                    0.811750
## salessupport:salarymedium                  0.722211
## salestechnical:salarymedium                0.604962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2718 on 4365 degrees of freedom
## Multiple R-squared:  0.6201, Adjusted R-squared:  0.6086
## F-statistic: 53.58 on 133 and 4365 DF,  p-value: < 2.2e-16
```

#making a reduced model with interactions

```
reduced_model_interactions <- lm(formula = left~time_spend_company*Work_accident + time_spend_company*pr
number_project*time_spend_company  , data = test_hr)
summary(reduced_model_interactions)
```

```
##
## Call:
## lm(formula = left ~ time_spend_company * Work_accident + time_spend_company *
##     promotion_last_5years + satisfaction_level * time_spend_company +
##     average_montly_hours * time_spend_company + number_project *
##     average_montly_hours + number_project * time_spend_company,
##     data = test_hr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46415 -0.19050 -0.05254  0.19867  1.34041
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                                4.074e+00  8.264e-02  49.302  < 2e-16
## time_spend_company                        -4.132e-01  2.109e-02 -19.590  < 2e-16
## Work_accident                             -7.743e-05  3.141e-02  -0.002 0.998033
## promotion_last_5years                      4.831e-02  7.249e-02   0.666 0.505176
## satisfaction_level                        -1.152e+00  6.100e-02 -18.886  < 2e-16
## average_montly_hours                      -1.284e-02  3.875e-04 -33.147  < 2e-16
## number_project                            -7.693e-01  1.918e-02 -40.120  < 2e-16
## time_spend_company:Work_accident          -2.829e-02  8.078e-03  -3.502 0.000466
## time_spend_company:promotion_last_5years -3.986e-02  1.540e-02  -2.588 0.009686
## time_spend_company:satisfaction_level      2.414e-01  1.486e-02  16.240  < 2e-16
## time_spend_company:average_montly_hours    6.281e-04  7.068e-05   8.887  < 2e-16
## average_montly_hours:number_project        2.870e-03  7.918e-05  36.246  < 2e-16
## time_spend_company:number_project          4.582e-02  3.396e-03  13.493  < 2e-16
##
## (Intercept)                               ***
## time_spend_company                        ***
## Work_accident
## promotion_last_5years
## satisfaction_level                        ***
## average_montly_hours                      ***
## number_project                            ***
## time_spend_company:Work_accident          ***
## time_spend_company:promotion_last_5years **
## time_spend_company:satisfaction_level     ***
## time_spend_company:average_montly_hours   ***
## average_montly_hours:number_project       ***
## time_spend_company:number_project         ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3143 on 4486 degrees of freedom
## Multiple R-squared:  0.4781, Adjusted R-squared:  0.4767
## F-statistic: 342.4 on 12 and 4486 DF,  p-value: < 2.2e-16
```

#Multiple Linear Regression Conclussion: After comparing all three models, it's evident that the full model with interactions significantly outperforms both the full model and the reduced model in explaining the factors behind employee retention. This tells us that it's crucial to consider how these factors interact with each other when trying to understand why employees leave.

##Logistic Regression

#making a full logistic regression model

```r
full_logistic_model <- glm(left ~ ., data = test_hr, family = binomial)

# Print a summary of the logistic regression model
summary(full_logistic_model)
```

```
##
## Call:
## glm(formula = left ~ ., family = binomial, data = test_hr)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)           -0.7817844  0.3198025  -2.445 0.014502 *
## satisfaction_level    -4.1343012  0.1769546 -23.364  < 2e-16 ***
## last_evaluation        0.8730544  0.2721175   3.208 0.001335 **
## number_project        -0.3283080  0.0389994  -8.418  < 2e-16 ***
## average_montly_hours   0.0038091  0.0009244   4.121 3.78e-05 ***
## time_spend_company     0.2894917  0.0284546  10.174  < 2e-16 ***
## Work_accident         -1.5326630  0.1571664  -9.752  < 2e-16 ***
## promotion_last_5years -1.5980316  0.4593108  -3.479 0.000503 ***
## saleshr                0.1189071  0.2434679   0.488 0.625274
## salesIT               -0.3448161  0.2167613  -1.591 0.111663
## salesmanagement       -0.7248557  0.2922411  -2.480 0.013126 *
## salesmarketing         0.1929950  0.2289623   0.843 0.399278
## salesproduct_mng      -0.2539511  0.2363639  -1.074 0.282640
## salesRandD            -0.8689426  0.2674126  -3.249 0.001156 **
## salessales            -0.3144149  0.1819358  -1.728 0.083959 .
## salessupport           0.0487708  0.1922903   0.254 0.799780
## salestechnical        -0.0460957  0.1883219  -0.245 0.806634
## salarylow              1.4969166  0.2013553   7.434 1.05e-13 ***
## salarymedium           0.8571542  0.2036696   4.209 2.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5084.5  on 4498  degrees of freedom
## Residual deviance: 3942.9  on 4480  degrees of freedom
## AIC: 3980.9
##
## Number of Fisher Scoring iterations: 5
```

```
# Make predictions on your test data
full_logistic_predictions <- predict(full_logistic_model, newdata = test_hr, type = "response")

# Convert predicted probabilities to binary classes (0 or 1) using a threshold (e.g., 0.5)
full_predicted_classes <- ifelse(full_logistic_predictions > 0.5, 1, 0)

# Calculate accuracy
full_accuracy <- mean(full_predicted_classes == test_hr$left)

# Print the accuracy
cat("Accuracy:", full_accuracy, "\n")
```

```
## Accuracy: 0.7899533
```

#making a reduced logistic regression model

```
red_logistic_model <- glm(left~satisfaction_level + average_montly_hours + promotion_last_5years + sala

# Print a summary of the logistic regression model
summary(red_logistic_model)
```

```
##
## Call:
```

```
## glm(formula = left ~ satisfaction_level + average_montly_hours +
##     promotion_last_5years + salary + sales, family = binomial,
##     data = test_hr)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.2669706  0.2896637  -0.922 0.356708
## satisfaction_level    -3.7939204  0.1603455 -23.661  < 2e-16 ***
## average_montly_hours   0.0020522  0.0007305   2.809 0.004962 **
## promotion_last_5years -1.4488323  0.4470021  -3.241 0.001190 **
## salarylow              1.3910851  0.1935227   7.188 6.56e-13 ***
## salarymedium           0.8217199  0.1964052   4.184 2.87e-05 ***
## saleshr                0.1375145  0.2360135   0.583 0.560125
## salesIT               -0.3179196  0.2102429  -1.512 0.130495
## salesmanagement       -0.6748605  0.2849005  -2.369 0.017848 *
## salesmarketing         0.1899668  0.2207747   0.860 0.389538
## salesproduct_mng      -0.3000291  0.2301221  -1.304 0.192308
## salesRandD            -0.8614842  0.2594302  -3.321 0.000898 ***
## salessales            -0.2473517  0.1765330  -1.401 0.161165
## salessupport           0.0261374  0.1864212   0.140 0.888497
## salestechnical        -0.0808851  0.1828131  -0.442 0.658166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5084.5  on 4498  degrees of freedom
## Residual deviance: 4220.8  on 4484  degrees of freedom
## AIC: 4250.8
##
## Number of Fisher Scoring iterations: 5
```

```
# Make predictions on your test data
red_logistic_predictions <- predict(red_logistic_model, newdata = test_hr, type = "response")

# Convert predicted probabilities to binary classes (0 or 1) using a threshold (e.g., 0.5)
red_predicted_classes <- ifelse(red_logistic_predictions > 0.5, 1, 0)

# Calculate accuracy
red_accuracy <- mean(red_predicted_classes == test_hr$left)

# Print the accuracy
cat("Accuracy:", red_accuracy, "\n")
```

```
## Accuracy: 0.7717271
```

#making a full logistic regression model with interactions

```
int_full_logistic_model <- glm(left ~ .*., data = test_hr, family = binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# Print a summary of the logistic regression model
summary(int_full_logistic_model)
```

```
##
## Call:
## glm(formula = left ~ . * ., family = binomial, data = test_hr)
##
## Coefficients: (1 not defined because of singularities)
##                                          Estimate Std. Error z value
## (Intercept)                             5.086e+01  3.178e+00  16.001
## satisfaction_level                     -3.001e+01  3.057e+00  -9.818
## last_evaluation                        -3.470e+01  4.398e+00  -7.891
## number_project                         -8.805e+00  6.577e-01 -13.387
## average_montly_hours                   -1.182e-01  1.499e-02  -7.886
## time_spend_company                     -3.555e+00  5.066e-01  -7.017
## Work_accident                          -1.700e+01  9.168e+02  -0.019
## promotion_last_5years                   5.516e+01  3.677e+03   0.015
## saleshr                                -1.605e+00  3.929e+00  -0.408
## salesIT                                 3.106e-01  2.652e+00   0.117
## salesmanagement                        -3.111e+00  3.636e+00  -0.855
## salesmarketing                          3.717e+00  2.596e+00   1.432
## salesproduct_mng                        1.404e+00  2.645e+00   0.531
## salesRandD                             -1.176e+00  2.931e+00  -0.401
## salessales                              6.082e-01  2.407e+00   0.253
## salessupport                            4.546e-01  2.477e+00   0.183
## salestechnical                          2.950e+00  2.332e+00   1.265
## salarylow                              -1.401e+00  2.457e+00  -0.570
## salarymedium                           -1.545e+00  2.465e+00  -0.627
## satisfaction_level:last_evaluation      1.717e+01  2.352e+00   7.300
## satisfaction_level:number_project      -1.032e+00  3.174e-01  -3.253
## satisfaction_level:average_montly_hours 3.960e-02  8.880e-03   4.460
## satisfaction_level:time_spend_company   2.489e+00  2.475e-01  10.057
## satisfaction_level:Work_accident       -3.104e+00  1.302e+00  -2.384
## satisfaction_level:promotion_last_5years -7.137e-01  3.265e+03   0.000
## satisfaction_level:saleshr              2.947e+00  1.976e+00   1.491
## satisfaction_level:salesIT             -8.990e-02  1.697e+00  -0.053
## satisfaction_level:salesmanagement      1.326e+00  2.369e+00   0.560
## satisfaction_level:salesmarketing       3.368e+00  1.904e+00   1.769
## satisfaction_level:salesproduct_mng     1.056e+00  1.959e+00   0.539
## satisfaction_level:salesRandD           1.083e+00  2.007e+00   0.540
## satisfaction_level:salessales          -5.028e-01  1.478e+00  -0.340
## satisfaction_level:salessupport         9.293e-01  1.532e+00   0.607
## satisfaction_level:salestechnical       1.057e+00  1.489e+00   0.710
## satisfaction_level:salarylow            2.955e-01  1.481e+00   0.200
## satisfaction_level:salarymedium        -5.621e-01  1.509e+00  -0.372
## last_evaluation:number_project          5.044e+00  4.988e-01  10.114
## last_evaluation:average_montly_hours    1.059e-02  1.222e-02   0.867
## last_evaluation:time_spend_company      7.247e-01  3.408e-01   2.126
## last_evaluation:Work_accident           2.203e+00  2.020e+00   1.091
## last_evaluation:promotion_last_5years  -9.925e+01  8.445e+03  -0.012
## last_evaluation:saleshr                -8.084e+00  3.417e+00  -2.366
## last_evaluation:salesIT                -2.475e+00  2.982e+00  -0.830
## last_evaluation:salesmanagement         8.440e-01  4.197e+00   0.201
```

```
## last_evaluation:salesmarketing             -6.905e+00  3.372e+00  -2.048
## last_evaluation:salesproduct_mng           -4.045e+00  3.437e+00  -1.177
## last_evaluation:salesRandD                 -4.936e-02  3.515e+00  -0.014
## last_evaluation:salessales                 -2.645e+00  2.566e+00  -1.031
## last_evaluation:salessupport               -4.168e+00  2.733e+00  -1.525
## last_evaluation:salestechnical             -7.169e+00  2.659e+00  -2.696
## last_evaluation:salarylow                   8.971e+00  2.773e+00   3.235
## last_evaluation:salarymedium                9.433e+00  2.825e+00   3.339
## number_project:average_montly_hours         2.182e-02  1.694e-03  12.884
## number_project:time_spend_company           4.327e-01  5.922e-02   7.306
## number_project:Work_accident               -4.126e-01  3.457e-01  -1.194
## number_project:promotion_last_5years       -3.557e+00  1.401e+03  -0.003
## number_project:saleshr                      9.691e-01  5.265e-01   1.841
## number_project:salesIT                      2.478e-01  4.712e-01   0.526
## number_project:salesmanagement              7.297e-01  6.995e-01   1.043
## number_project:salesmarketing               9.420e-01  4.821e-01   1.954
## number_project:salesproduct_mng             7.099e-01  5.142e-01   1.380
## number_project:salesRandD                  -5.589e-01  5.850e-01  -0.955
## number_project:salessales                   1.275e-01  3.960e-01   0.322
## number_project:salessupport                 5.455e-01  4.182e-01   1.304
## number_project:salestechnical               1.107e+00  4.053e-01   2.731
## number_project:salarylow                   -1.226e+00  3.414e-01  -3.592
## number_project:salarymedium                -1.407e+00  3.513e-01  -4.006
## average_montly_hours:time_spend_company     3.099e-03  1.183e-03   2.620
## average_montly_hours:Work_accident          2.602e-03  7.212e-03   0.361
## average_montly_hours:promotion_last_5years  1.607e-01  2.700e+01   0.006
## average_montly_hours:saleshr                1.808e-03  1.138e-02   0.159
## average_montly_hours:salesIT                1.462e-02  1.065e-02   1.372
## average_montly_hours:salesmanagement       -8.915e-03  1.634e-02  -0.546
## average_montly_hours:salesmarketing        -2.394e-03  1.117e-02  -0.214
## average_montly_hours:salesproduct_mng      -1.087e-02  1.213e-02  -0.897
## average_montly_hours:salesRandD             1.645e-02  1.235e-02   1.332
## average_montly_hours:salessales             8.824e-03  8.951e-03   0.986
## average_montly_hours:salessupport           2.673e-03  9.407e-03   0.284
## average_montly_hours:salestechnical         1.671e-03  9.268e-03   0.180
## average_montly_hours:salarylow              2.016e-03  9.175e-03   0.220
## average_montly_hours:salarymedium           9.619e-03  9.415e-03   1.022
## time_spend_company:Work_accident            1.309e-01  1.874e-01   0.699
## time_spend_company:promotion_last_5years   -5.622e+00  1.026e+03  -0.005
## time_spend_company:saleshr                  4.087e-01  3.871e-01   1.056
## time_spend_company:salesIT                 -5.166e-01  3.019e-01  -1.711
## time_spend_company:salesmanagement         -7.731e-01  4.069e-01  -1.900
## time_spend_company:salesmarketing          -5.805e-01  3.202e-01  -1.813
## time_spend_company:salesproduct_mng        -2.933e-01  3.271e-01  -0.897
## time_spend_company:salesRandD               2.315e-01  4.185e-01   0.553
## time_spend_company:salessales              -5.506e-01  2.685e-01  -2.051
## time_spend_company:salessupport             2.234e-02  2.759e-01   0.081
## time_spend_company:salestechnical          -3.858e-01  2.694e-01  -1.432
## time_spend_company:salarylow                1.002e-01  2.533e-01   0.396
## time_spend_company:salarymedium            -2.895e-01  2.567e-01  -1.128
## Work_accident:promotion_last_5years        -2.575e+01  5.461e+03  -0.005
## Work_accident:saleshr                      -6.321e-01  1.397e+00  -0.453
## Work_accident:salesIT                      -3.336e-01  1.235e+00  -0.270
## Work_accident:salesmanagement              -2.272e-01  1.500e+00  -0.152
```

```
## Work_accident:salesmarketing               -2.451e-01  1.226e+00  -0.200
## Work_accident:salesproduct_mng             -1.634e+00  1.402e+00  -1.166
## Work_accident:salesRandD                   -2.391e+00  1.979e+00  -1.208
## Work_accident:salessales                   -1.269e+00  1.025e+00  -1.238
## Work_accident:salessupport                 -2.891e+00  1.260e+00  -2.294
## Work_accident:salestechnical               -3.750e-01  1.038e+00  -0.361
## Work_accident:salarylow                     1.692e+01  9.168e+02   0.018
## Work_accident:salarymedium                  1.787e+01  9.168e+02   0.019
## promotion_last_5years:saleshr              -3.813e+01  4.761e+03  -0.008
## promotion_last_5years:salesIT               7.353e+00  7.115e+03   0.001
## promotion_last_5years:salesmanagement      -1.619e+01  4.589e+03  -0.004
## promotion_last_5years:salesmarketing       -5.666e+01  6.705e+03  -0.008
## promotion_last_5years:salesproduct_mng             NA         NA      NA
## promotion_last_5years:salesRandD           -5.773e+01  3.234e+03  -0.018
## promotion_last_5years:salessales           -5.031e+01  3.810e+03  -0.013
## promotion_last_5years:salessupport         -7.852e+01  4.739e+03  -0.017
## promotion_last_5years:salestechnical       -1.815e+01  5.770e+03  -0.003
## promotion_last_5years:salarylow             5.052e+01  1.956e+03   0.026
## promotion_last_5years:salarymedium          1.343e+01  2.161e+03   0.006
## saleshr:salarylow                           8.011e-01  3.459e+00   0.232
## salesIT:salarylow                          -7.663e-01  2.279e+00  -0.336
## salesmanagement:salarylow                   3.587e+00  2.569e+00   1.396
## salesmarketing:salarylow                   -1.248e+00  2.228e+00  -0.560
## salesproduct_mng:salarylow                  1.058e+00  2.350e+00   0.450
## salesRandD:salarylow                       -2.809e+00  2.420e+00  -1.161
## salessales:salarylow                        1.620e+00  2.145e+00   0.755
## salessupport:salarylow                     -2.171e-03  2.162e+00  -0.001
## salestechnical:salarylow                   -1.429e+00  2.034e+00  -0.703
## saleshr:salarymedium                        1.350e+00  3.454e+00   0.391
## salesIT:salarymedium                       -4.570e-01  2.290e+00  -0.200
## salesmanagement:salarymedium                3.114e+00  2.558e+00   1.217
## salesmarketing:salarymedium                -8.646e-01  2.243e+00  -0.386
## salesproduct_mng:salarymedium               1.395e+00  2.363e+00   0.590
## salesRandD:salarymedium                    -2.050e+00  2.405e+00  -0.853
## salessales:salarymedium                     8.940e-01  2.149e+00   0.416
## salessupport:salarymedium                  -1.556e-01  2.172e+00  -0.072
## salestechnical:salarymedium                -8.392e-01  2.039e+00  -0.412
##                                           Pr(>|z|)
## (Intercept)                                < 2e-16 ***
## satisfaction_level                         < 2e-16 ***
## last_evaluation                           3.00e-15 ***
## number_project                             < 2e-16 ***
## average_montly_hours                      3.11e-15 ***
## time_spend_company                        2.27e-12 ***
## Work_accident                             0.985208
## promotion_last_5years                     0.988030
## saleshr                                   0.682938
## salesIT                                   0.906761
## salesmanagement                           0.392327
## salesmarketing                            0.152164
## salesproduct_mng                          0.595698
## salesRandD                                0.688375
## salessales                                0.800502
## salessupport                              0.854407
```

```
## salestechnical                                0.205931
## salarylow                                     0.568660
## salarymedium                                  0.530744
## satisfaction_level:last_evaluation            2.88e-13 ***
## satisfaction_level:number_project             0.001143 **
## satisfaction_level:average_montly_hours       8.21e-06 ***
## satisfaction_level:time_spend_company          < 2e-16 ***
## satisfaction_level:Work_accident              0.017113 *
## satisfaction_level:promotion_last_5years      0.999826
## satisfaction_level:saleshr                    0.135978
## satisfaction_level:salesIT                    0.957759
## satisfaction_level:salesmanagement            0.575767
## satisfaction_level:salesmarketing             0.076961 .
## satisfaction_level:salesproduct_mng           0.589985
## satisfaction_level:salesRandD                 0.589372
## satisfaction_level:salessales                 0.733662
## satisfaction_level:salessupport               0.544106
## satisfaction_level:salestechnical             0.477753
## satisfaction_level:salarylow                  0.841854
## satisfaction_level:salarymedium               0.709531
## last_evaluation:number_project                 < 2e-16 ***
## last_evaluation:average_montly_hours          0.385951
## last_evaluation:time_spend_company            0.033477 *
## last_evaluation:Work_accident                 0.275374
## last_evaluation:promotion_last_5years         0.990623
## last_evaluation:saleshr                       0.017976 *
## last_evaluation:salesIT                       0.406557
## last_evaluation:salesmanagement               0.840607
## last_evaluation:salesmarketing                0.040569 *
## last_evaluation:salesproduct_mng              0.239278
## last_evaluation:salesRandD                    0.988797
## last_evaluation:salessales                    0.302646
## last_evaluation:salessupport                  0.127194
## last_evaluation:salestechnical                0.007015 **
## last_evaluation:salarylow                     0.001217 **
## last_evaluation:salarymedium                  0.000840 ***
## number_project:average_montly_hours            < 2e-16 ***
## number_project:time_spend_company             2.74e-13 ***
## number_project:Work_accident                  0.232573
## number_project:promotion_last_5years          0.997975
## number_project:saleshr                        0.065689 .
## number_project:salesIT                        0.598924
## number_project:salesmanagement                0.296897
## number_project:salesmarketing                 0.050727 .
## number_project:salesproduct_mng               0.167435
## number_project:salesRandD                     0.339419
## number_project:salessales                     0.747508
## number_project:salessupport                   0.192104
## number_project:salestechnical                 0.006310 **
## number_project:salarylow                      0.000329 ***
## number_project:salarymedium                   6.17e-05 ***
## average_montly_hours:time_spend_company       0.008790 **
## average_montly_hours:Work_accident            0.718241
## average_montly_hours:promotion_last_5years    0.995251
```

```
## average_montly_hours:saleshr               0.873800
## average_montly_hours:salesIT               0.170078
## average_montly_hours:salesmanagement       0.585266
## average_montly_hours:salesmarketing        0.830279
## average_montly_hours:salesproduct_mng      0.369884
## average_montly_hours:salesRandD            0.182711
## average_montly_hours:salessales            0.324241
## average_montly_hours:salessupport          0.776309
## average_montly_hours:salestechnical        0.856951
## average_montly_hours:salarylow             0.826096
## average_montly_hours:salarymedium          0.306959
## time_spend_company:Work_accident           0.484860
## time_spend_company:promotion_last_5years   0.995628
## time_spend_company:saleshr                 0.291088
## time_spend_company:salesIT                 0.087095 .
## time_spend_company:salesmanagement         0.057399 .
## time_spend_company:salesmarketing          0.069790 .
## time_spend_company:salesproduct_mng        0.369834
## time_spend_company:salesRandD              0.580081
## time_spend_company:salessales              0.040299 *
## time_spend_company:salessupport            0.935479
## time_spend_company:salestechnical          0.152071
## time_spend_company:salarylow               0.692424
## time_spend_company:salarymedium            0.259323
## Work_accident:promotion_last_5years        0.996238
## Work_accident:saleshr                      0.650840
## Work_accident:salesIT                      0.787005
## Work_accident:salesmanagement              0.879576
## Work_accident:salesmarketing               0.841552
## Work_accident:salesproduct_mng             0.243639
## Work_accident:salesRandD                   0.227016
## Work_accident:salessales                   0.215686
## Work_accident:salessupport                 0.021799 *
## Work_accident:salestechnical               0.717993
## Work_accident:salarylow                    0.985278
## Work_accident:salarymedium                 0.984448
## promotion_last_5years:saleshr              0.993610
## promotion_last_5years:salesIT              0.999175
## promotion_last_5years:salesmanagement      0.997185
## promotion_last_5years:salesmarketing       0.993258
## promotion_last_5years:salesproduct_mng           NA
## promotion_last_5years:salesRandD           0.985758
## promotion_last_5years:salessales           0.989464
## promotion_last_5years:salessupport         0.986780
## promotion_last_5years:salestechnical       0.997490
## promotion_last_5years:salarylow            0.979398
## promotion_last_5years:salarymedium         0.995041
## saleshr:salarylow                          0.816833
## salesIT:salarylow                          0.736705
## salesmanagement:salarylow                  0.162703
## salesmarketing:salarylow                   0.575320
## salesproduct_mng:salarylow                 0.652472
## salesRandD:salarylow                       0.245779
## salessales:salarylow                       0.450175
```

```
## salessupport:salarylow                    0.999199
## salestechnical:salarylow                  0.482288
## saleshr:salarymedium                      0.695967
## salesIT:salarymedium                      0.841828
## salesmanagement:salarymedium              0.223496
## salesmarketing:salarymedium               0.699859
## salesproduct_mng:salarymedium             0.554973
## salesRandD:salarymedium                   0.393898
## salessales:salarymedium                   0.677468
## salessupport:salarymedium                 0.942866
## salestechnical:salarymedium               0.680662
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5084.5  on 4498  degrees of freedom
## Residual deviance: 1660.1  on 4365  degrees of freedom
## AIC: 1928.1
##
## Number of Fisher Scoring iterations: 18
```

```
# Make predictions on your test data
int_full_logistic_predictions <- predict(int_full_logistic_model, newdata = test_hr, type = "response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
# Convert predicted probabilities to binary classes (0 or 1) using a threshold (e.g., 0.5)
int_full_predicted_classes <- ifelse(int_full_logistic_predictions > 0.5, 1, 0)

# Calculate accuracy
int_full_accuracy <- mean(int_full_predicted_classes == test_hr$left)

# Print the accuracy
cat("Accuracy:", int_full_accuracy, "\n")
```

```
## Accuracy: 0.9439876
```

#Logistic Regression Conclussion: After performing logistic regression analysis on our dataset, it is evident that the full model with interactions consistently outperforms the reduced model, demonstrating a remarkable accuracy of 0.945 in predicting the likelihood of employee retention. This outstanding accuracy underscores the model's ability to make precise predictions based on the complex interplay of various factors. These findings highlight the significance of considering variable interactions in understanding employee retention and indicate the potential practical application of the model in developing effective retention strategies.

##Decision Tree

#making categorical variables into factors

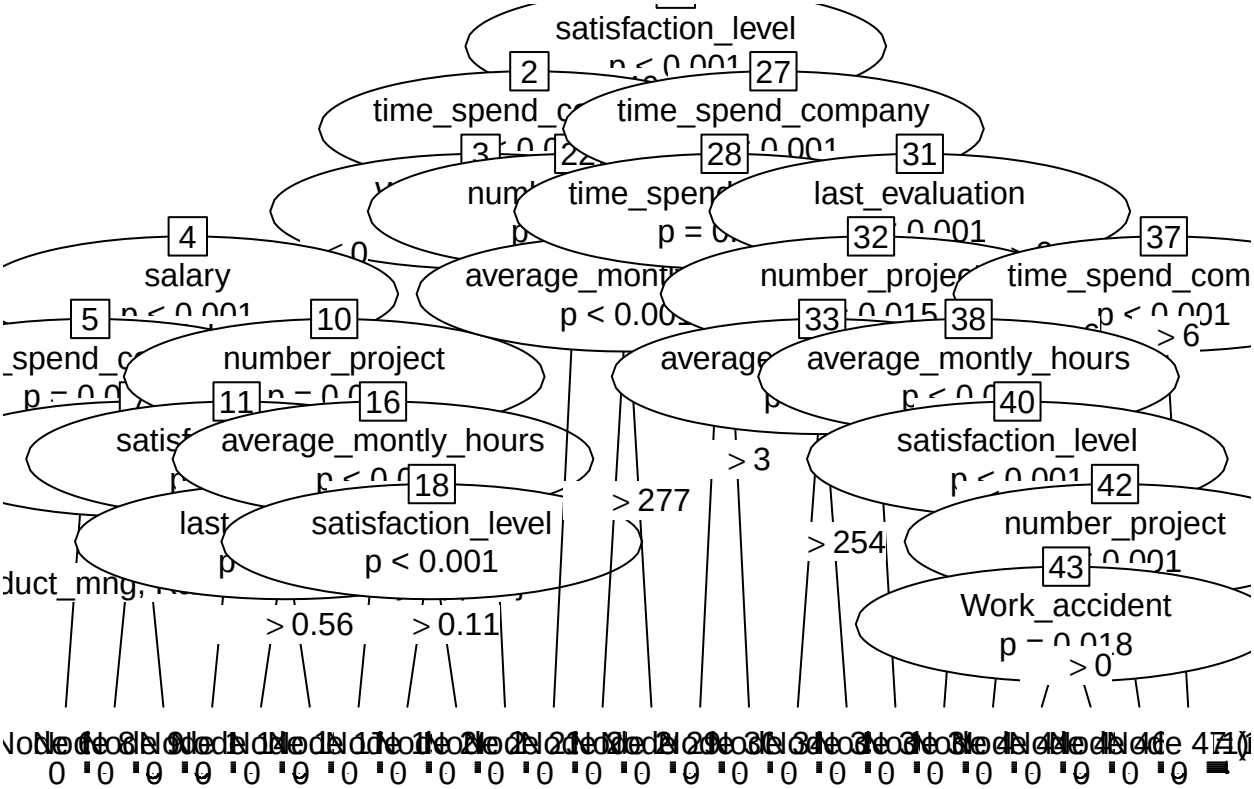
```
test_hr$sales <- as.factor(test_hr$sales)
test_hr$salary <- as.factor(test_hr$salary)
```

```r
full_decision_tree_model <- ctree(left ~ ., data = test_hr)
decision_tree_model2 <- ctree(left ~ ., data = test_hr, control = ctree_control(maxdepth = 3))
print(full_decision_tree_model)
```

```
##
##   Conditional inference tree with 24 terminal nodes
##
## Response:  left
## Inputs:  satisfaction_level, last_evaluation, number_project, average_montly_hours, time_spend_compar
## Number of observations:  4499
##
## 1) satisfaction_level <= 0.46; criterion = 1, statistic = 688.135
##   2) time_spend_company <= 4; criterion = 1, statistic = 139.881
##     3) Work_accident <= 0; criterion = 1, statistic = 41.964
##       4) salary == {low}; criterion = 1, statistic = 31.756
##         5) time_spend_company <= 2; criterion = 0.999, statistic = 27.105
##           6)*  weights = 29
##         5) time_spend_company > 2
##           7) sales == {accounting, IT, management, marketing, sales, support}; criterion = 0.998, sta
##             8)*  weights = 371
##           7) sales == {hr, product_mng, RandD, technical}
##             9)*  weights = 171
##       4) salary == {high, medium}
##         10) number_project <= 2; criterion = 0.996, statistic = 12.536
##           11) satisfaction_level <= 0.33; criterion = 1, statistic = 34.28
##             12)*  weights = 8
##           11) satisfaction_level > 0.33
##             13) last_evaluation <= 0.56; criterion = 1, statistic = 18.062
##               14)*  weights = 169
##             13) last_evaluation > 0.56
##               15)*  weights = 10
##         10) number_project > 2
##           16) average_montly_hours <= 242; criterion = 1, statistic = 99.749
##             17)*  weights = 102
##           16) average_montly_hours > 242
##             18) satisfaction_level <= 0.11; criterion = 1, statistic = 61.363
##               19)*  weights = 89
##             18) satisfaction_level > 0.11
##               20)*  weights = 18
##     3) Work_accident > 0
##       21)*  weights = 84
##   2) time_spend_company > 4
##     22) number_project <= 5; criterion = 1, statistic = 29.717
##       23)*  weights = 154
##     22) number_project > 5
##       24) average_montly_hours <= 277; criterion = 0.999, statistic = 16.221
##         25)*  weights = 47
##       24) average_montly_hours > 277
##         26)*  weights = 19
## 1) satisfaction_level > 0.46
##   27) time_spend_company <= 4; criterion = 1, statistic = 389.102
##     28) time_spend_company <= 3; criterion = 0.992, statistic = 11.157
##       29)*  weights = 2220
```

```
##     28) time_spend_company > 3
##       30)*  weights = 410
##   27) time_spend_company > 4
##     31) last_evaluation <= 0.81; criterion = 1, statistic = 238.349
##       32) number_project <= 4; criterion = 0.985, statistic = 18.643
##         33) average_montly_hours <= 254; criterion = 0.97, statistic = 23.349
##           34)*  weights = 138
##         33) average_montly_hours > 254
##           35)*  weights = 26
##       32) number_project > 4
##         36)*  weights = 53
##     31) last_evaluation > 0.81
##       37) time_spend_company <= 6; criterion = 1, statistic = 156.994
##         38) average_montly_hours <= 212; criterion = 1, statistic = 60.401
##           39)*  weights = 31
##         38) average_montly_hours > 212
##           40) satisfaction_level <= 0.68; criterion = 1, statistic = 53.132
##             41)*  weights = 15
##           40) satisfaction_level > 0.68
##             42) number_project <= 4; criterion = 1, statistic = 24.975
##               43) Work_accident <= 0; criterion = 0.982, statistic = 71.512
##                 44)*  weights = 105
##               43) Work_accident > 0
##                 45)*  weights = 11
##             42) number_project > 4
##               46)*  weights = 172
##       37) time_spend_company > 6
##         47)*  weights = 47
```

```
plot(full_decision_tree_model)
```

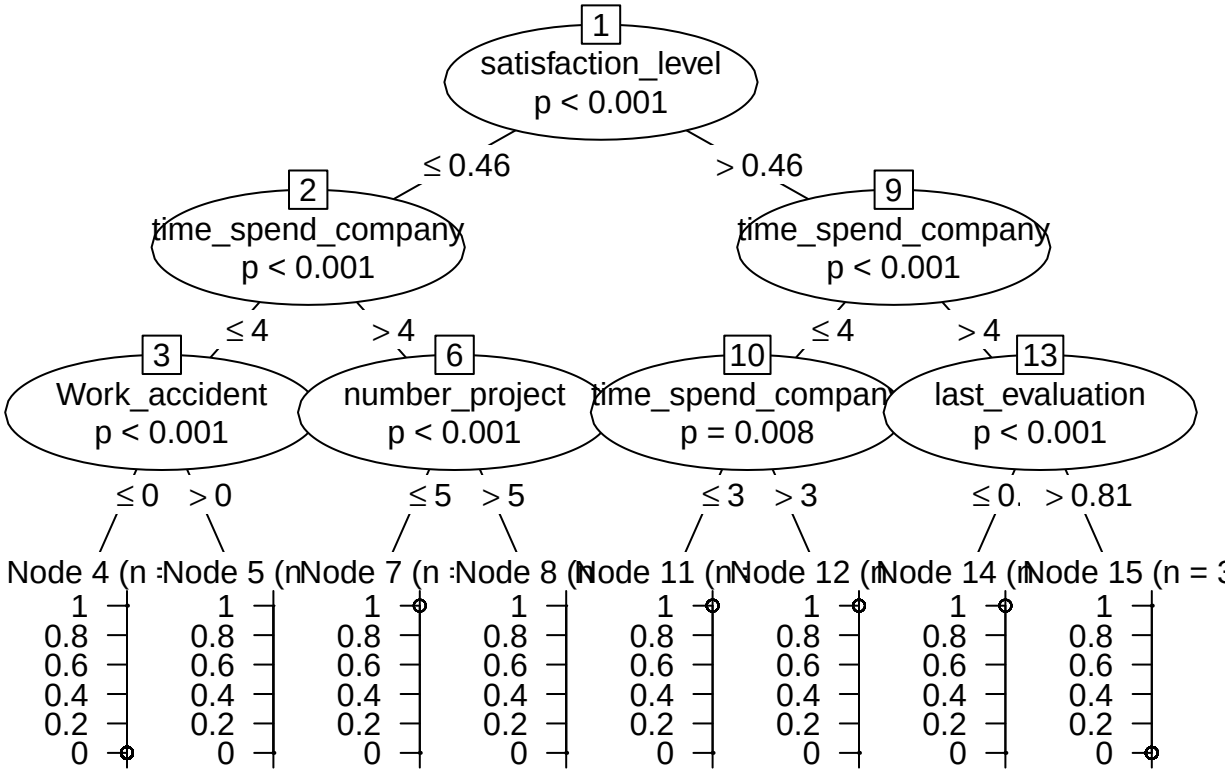



```
print(decision_tree_model2)
```

```
##
##   Conditional inference tree with 8 terminal nodes
##
## Response:  left
## Inputs:  satisfaction_level, last_evaluation, number_project, average_montly_hours, time_spend_compar
## Number of observations:  4499
##
## 1) satisfaction_level <= 0.46; criterion = 1, statistic = 688.135
##   2) time_spend_company <= 4; criterion = 1, statistic = 139.881
##     3) Work_accident <= 0; criterion = 1, statistic = 41.964
##       4)*  weights = 967
##     3) Work_accident > 0
##       5)*  weights = 84
##   2) time_spend_company > 4
##     6) number_project <= 5; criterion = 1, statistic = 29.717
##       7)*  weights = 154
##     6) number_project > 5
##       8)*  weights = 66
## 1) satisfaction_level > 0.46
##   9) time_spend_company <= 4; criterion = 1, statistic = 389.102
##     10) time_spend_company <= 3; criterion = 0.992, statistic = 11.157
##       11)*  weights = 2220
##     10) time_spend_company > 3
```

```
##       12)*  weights = 410
##   9) time_spend_company > 4
##     13) last_evaluation <= 0.81; criterion = 1, statistic = 238.349
##       14)*  weights = 217
##     13) last_evaluation > 0.81
##       15)*  weights = 381
```

```
plot(decision_tree_model2)
```




```
library(partykit)
```

```
## Loading required package: libcoin
```

```
##
## Attaching package: 'partykit'
```

```
## The following objects are masked from 'package:party':
##
##     cforest, ctree, ctree_control, edge_simple, mob, mob_control,
##     node_barplot, node_bivplot, node_boxplot, node_inner, node_surv,
##     node_terminal, varimp
```

#Decision Tree Conclusion TBD

##Clustering

Select the relevant variables for clustering

```r
clustering_data_train <- train_hr[, c("satisfaction_level", "last_evaluation", "number_project", "averag
```

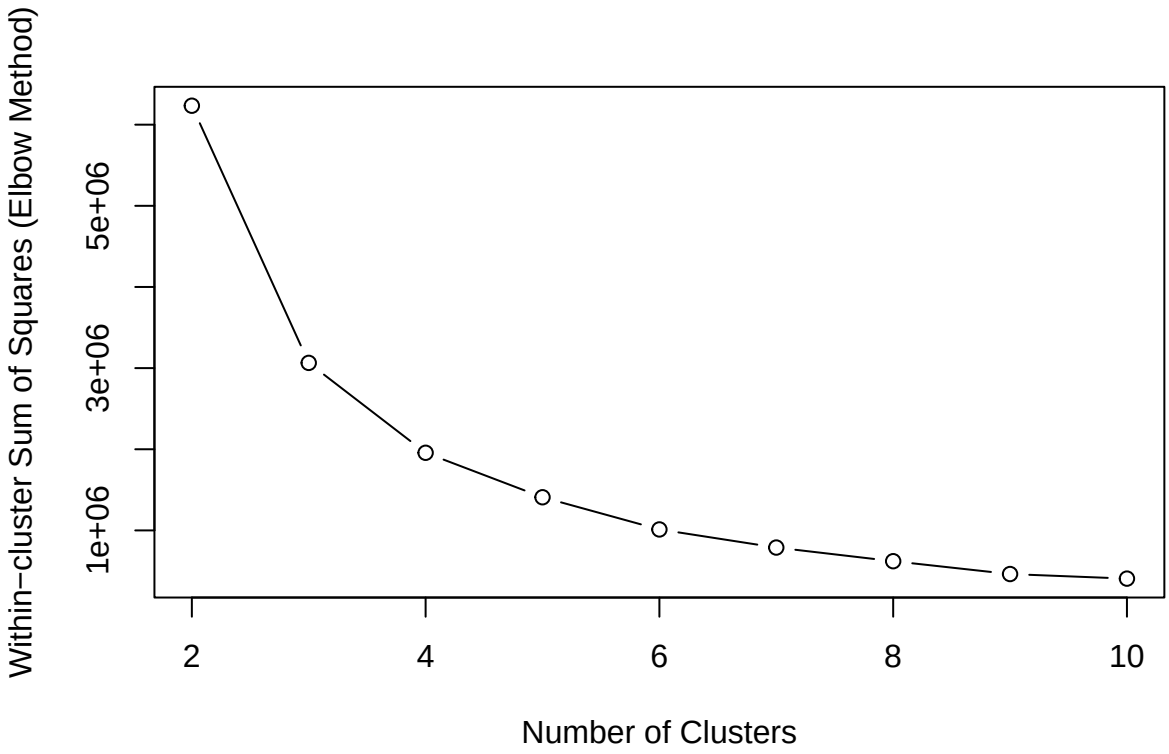
Initialize empty vectors to store results

```r
wss <- numeric(10)  # For the Elbow Method
silhouette_scores <- numeric(10)  # For Silhouette Score


# Determine the range of cluster numbers to test (e.g., from 2 to 10 clusters)

for (i in 2:10) {
  # Perform K-Means clustering on the training data
  kmeans_result <- kmeans(clustering_data_train, centers = i)

  # Calculate the sum of squared distances (SSD) for the Elbow Method
  wss[i] <- kmeans_result$tot.withinss

  # Calculate the Silhouette Score
  if (i > 1) {
    silhouette_scores[i] <- silhouette(kmeans_result$cluster, dist(clustering_data_train))
  }
}
```

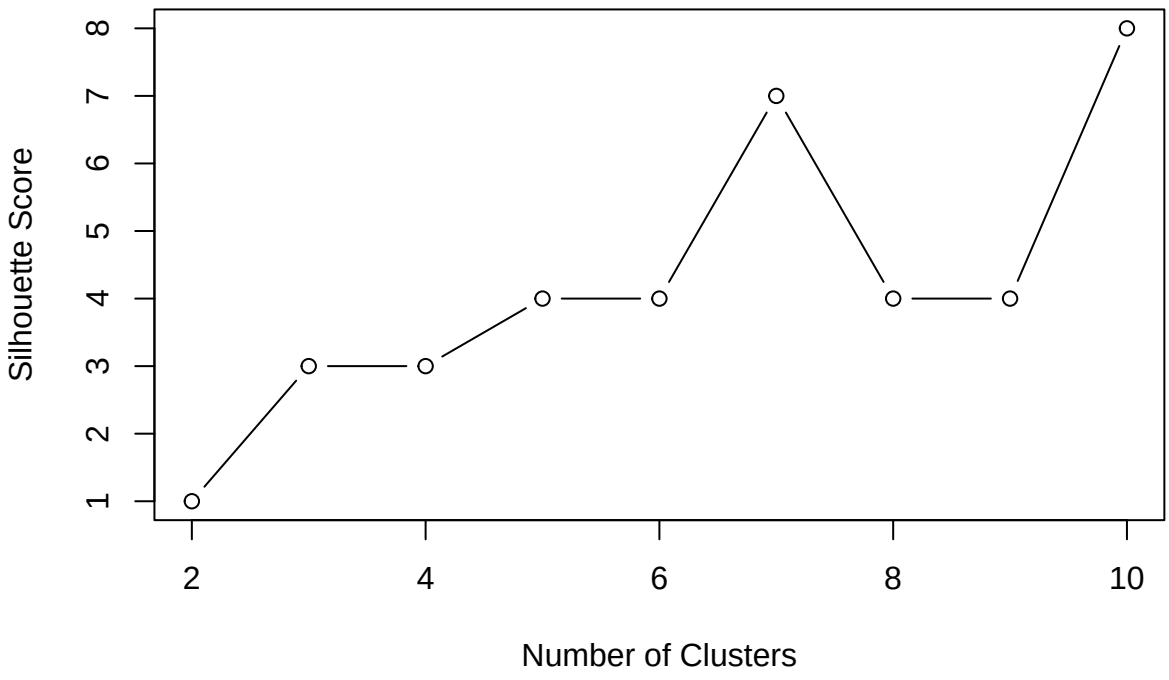
```
## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length
```

```
## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length

## Warning in silhouette_scores[i] <- silhouette(kmeans_result$cluster,
## dist(clustering_data_train)): number of items to replace is not a multiple of
## replacement length
```

```
# Plot the Elbow Method results
plot(2:10, wss[-1], type = "b", xlab = "Number of Clusters", ylab = "Within-cluster Sum of Squares (Elb
```




```
# Plot the Silhouette Score results
plot(2:10, silhouette_scores[-1], type = "b", xlab = "Number of Clusters", ylab = "Silhouette Score")
```




```
clustering_data <- test_hr[, c("satisfaction_level", "last_evaluation", "number_project", "average_montl
```

Standardize the data

```
scaled_data <- scale(clustering_data)
```

Determine the number of clusters (you can choose an appropriate number)

```
k <- 6
```

Perform K-Means clustering

```
kmeans_result <- kmeans(scaled_data, centers = k)
```

Assign cluster labels to the original dataset

```
test_hr$cluster <- kmeans_result$cluster
```

View the cluster centers

```
print(kmeans_result$centers)
```

```
##   satisfaction_level last_evaluation number_project average_montly_hours
## 1         0.05999707     -0.67345847    -0.12633562         -0.862535694
## 2        -0.79290623     -1.12455630    -1.41446849         -1.098449210
## 3         0.40541711     -0.48778543     0.01855098          0.816332592
## 4         0.48082659      0.97185959    -0.03490626          0.009291057
## 5        -0.63229226      0.96615958     1.28170240          1.108702230
## 6        -0.34032509     -0.04554361     0.15012149         -0.056773238
##   time_spend_company       left
## 1         -0.3767188 -0.5811356
## 2         -0.3197152  1.7203864
## 3         -0.4064120 -0.5706503
## 4         -0.4259675 -0.5706836
## 5          0.7400169  1.6125025
## 6          2.3292382 -0.5811356
```

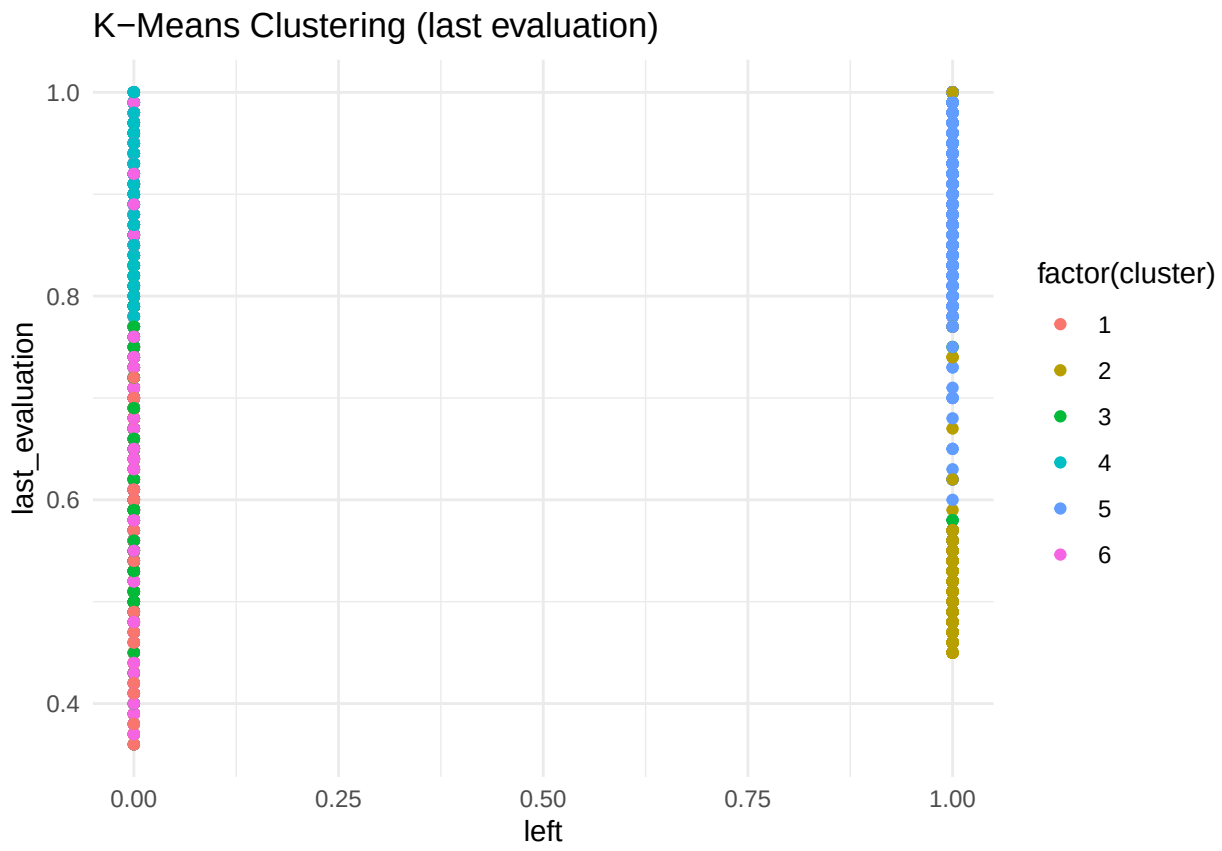
View the cluster sizes

```
table(test_hr$cluster)
```

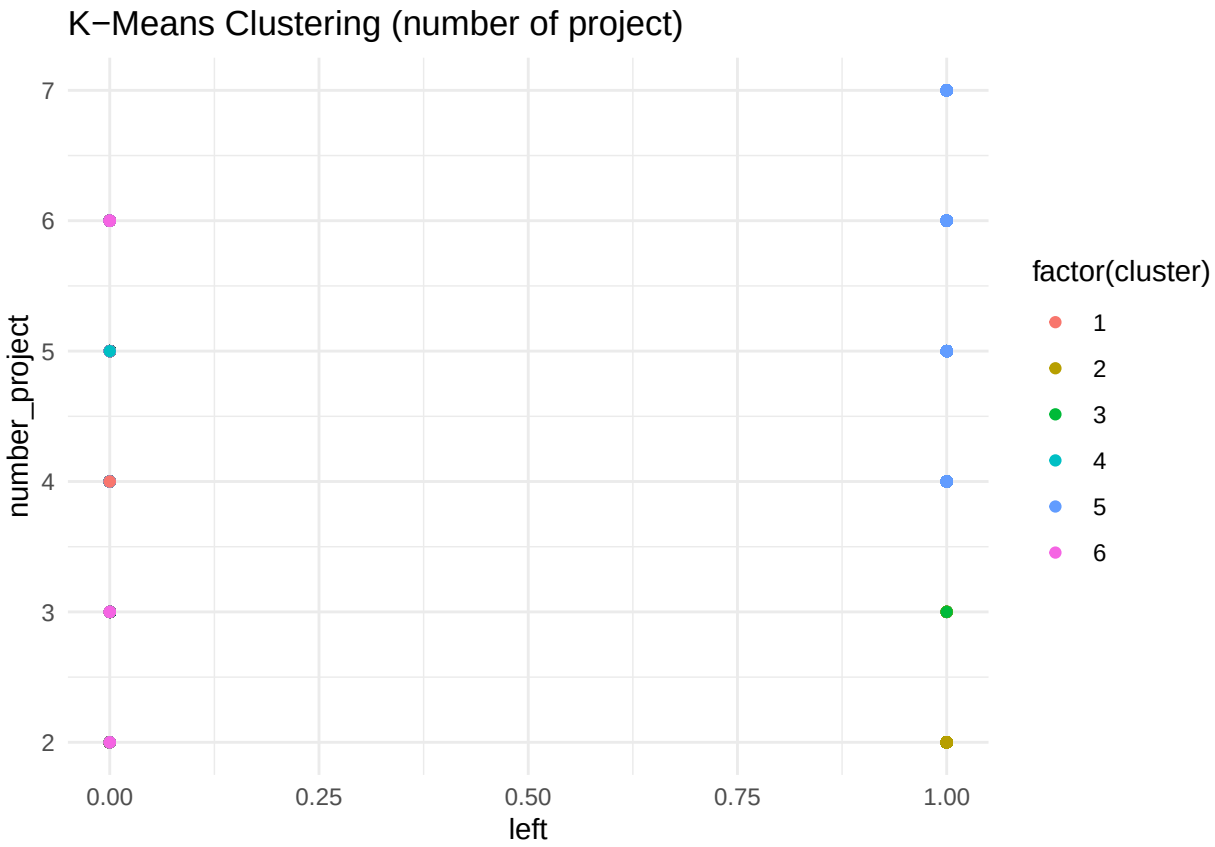
```
##
##    1    2    3    4    5    6
##  982  517  878 1101  640  381
```

#plotting k-mean clustering to predict dependent variable “left”

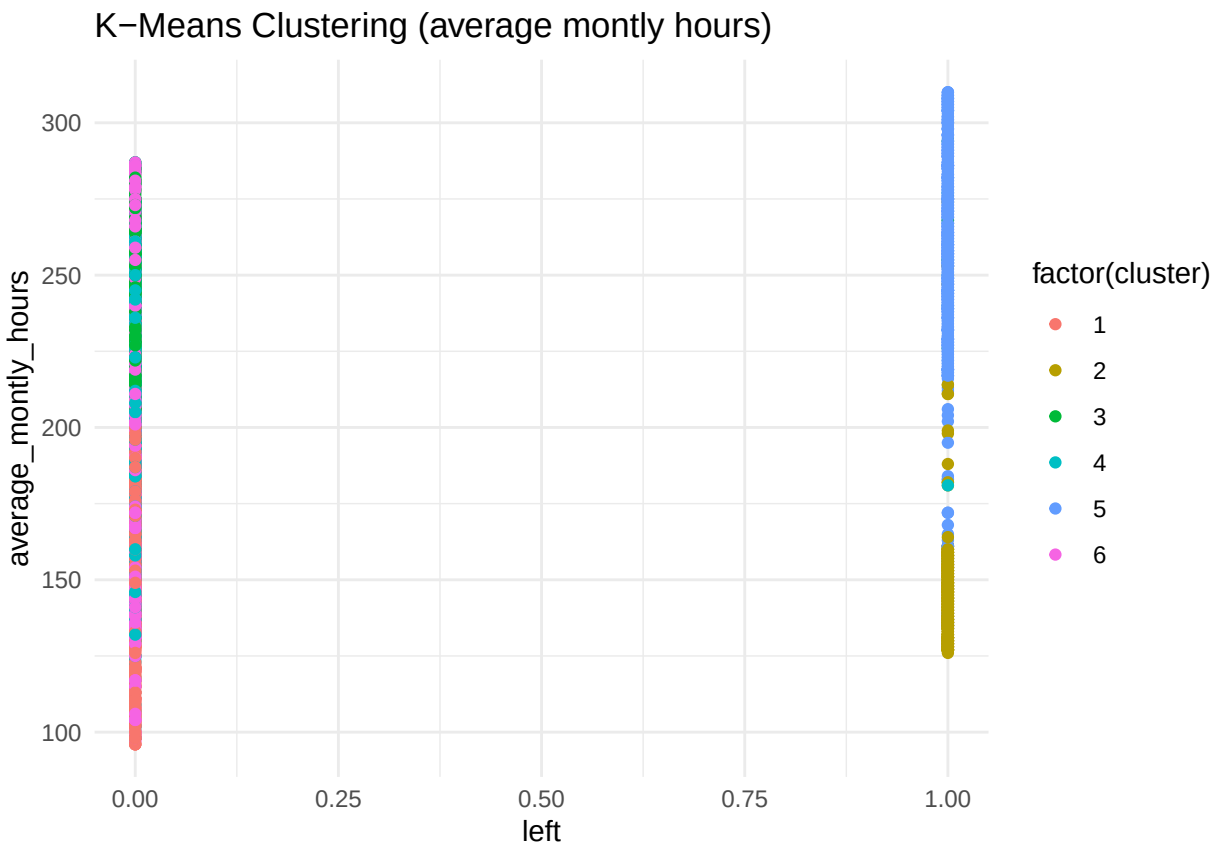
```
ggplot(test_hr, aes(x = left, y = last_evaluation, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (last evaluation)") +
  theme_minimal()
```



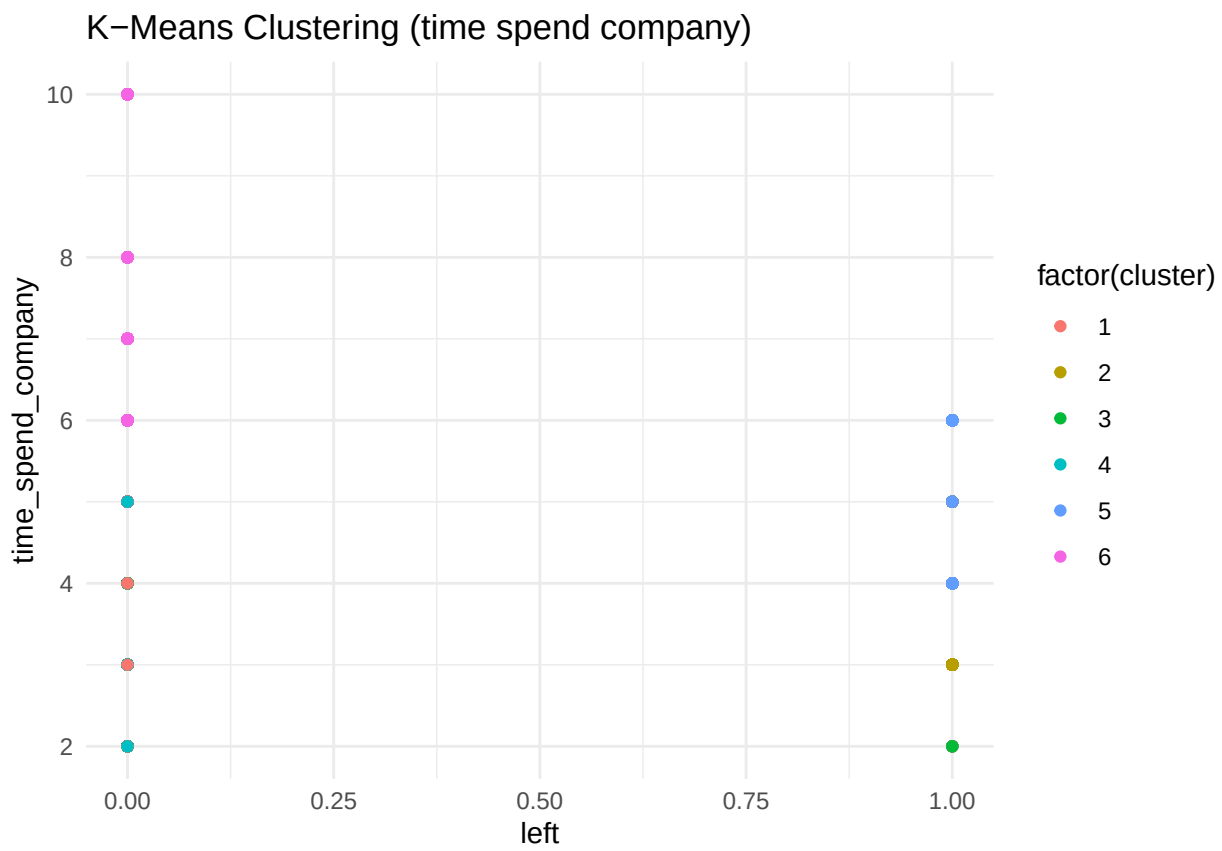

```
ggplot(test_hr, aes(x = left, y = number_project, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (number of project)") +
  theme_minimal()
```



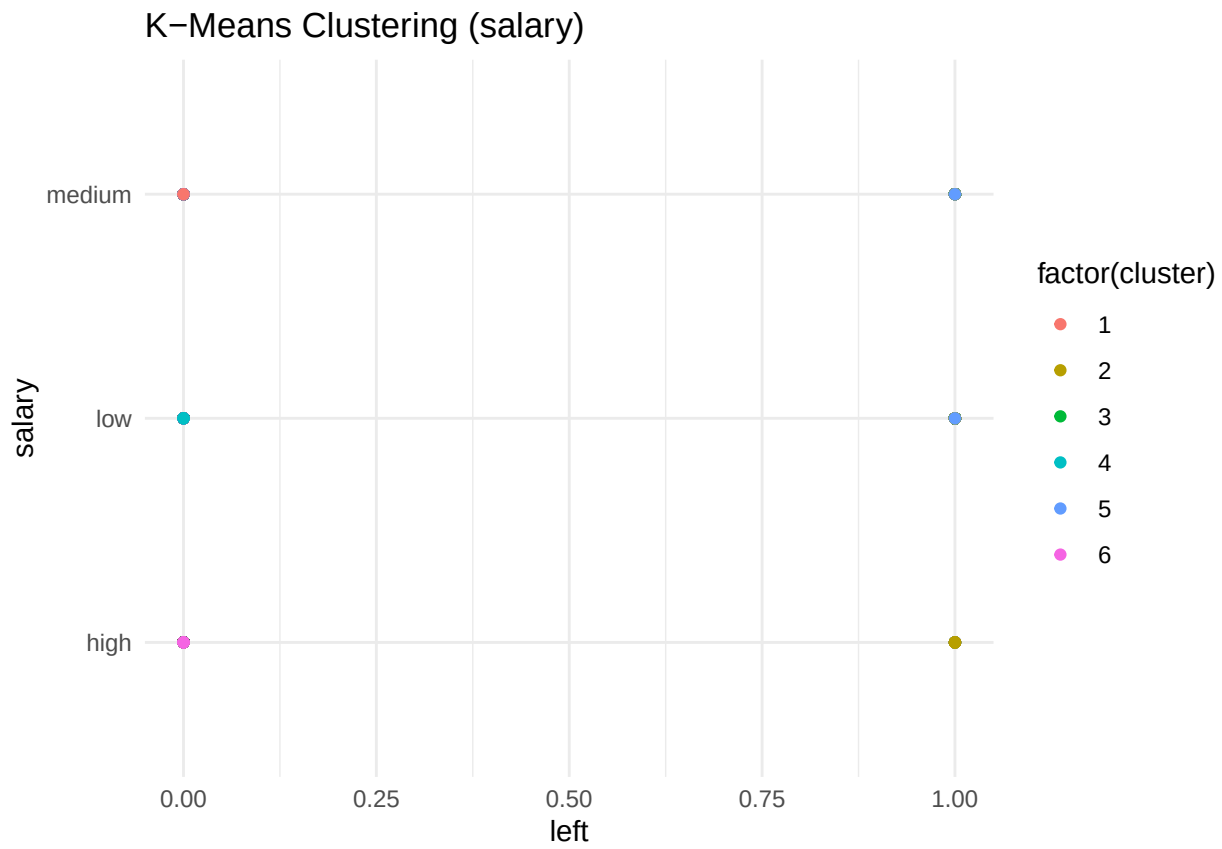

```
ggplot(test_hr, aes(x = left, y = average_montly_hours, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (average montly hours)") +
  theme_minimal()
```



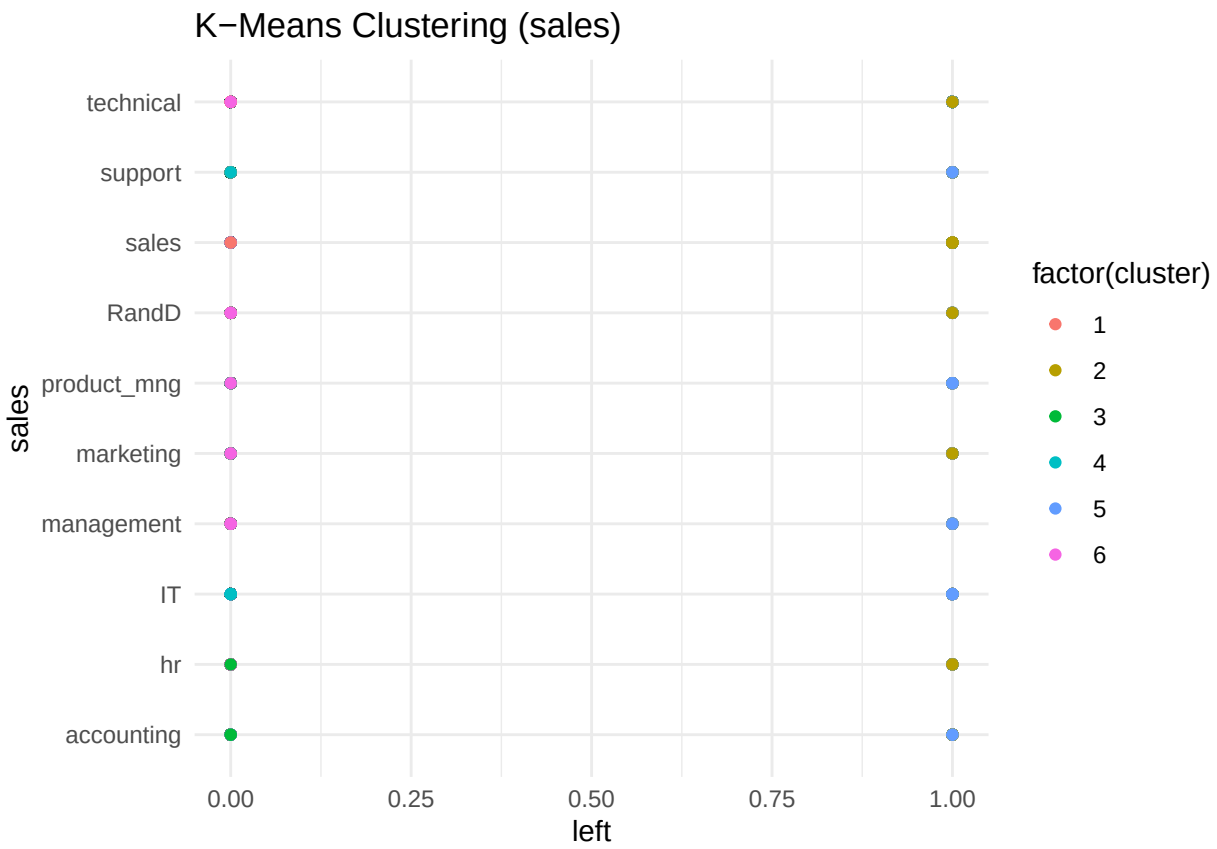

```
ggplot(test_hr, aes(x = left, y = time_spend_company, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (time spend company)") +
  theme_minimal()
```




```r
ggplot(test_hr, aes(x = left, y = salary, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (salary)") +
  theme_minimal()
```

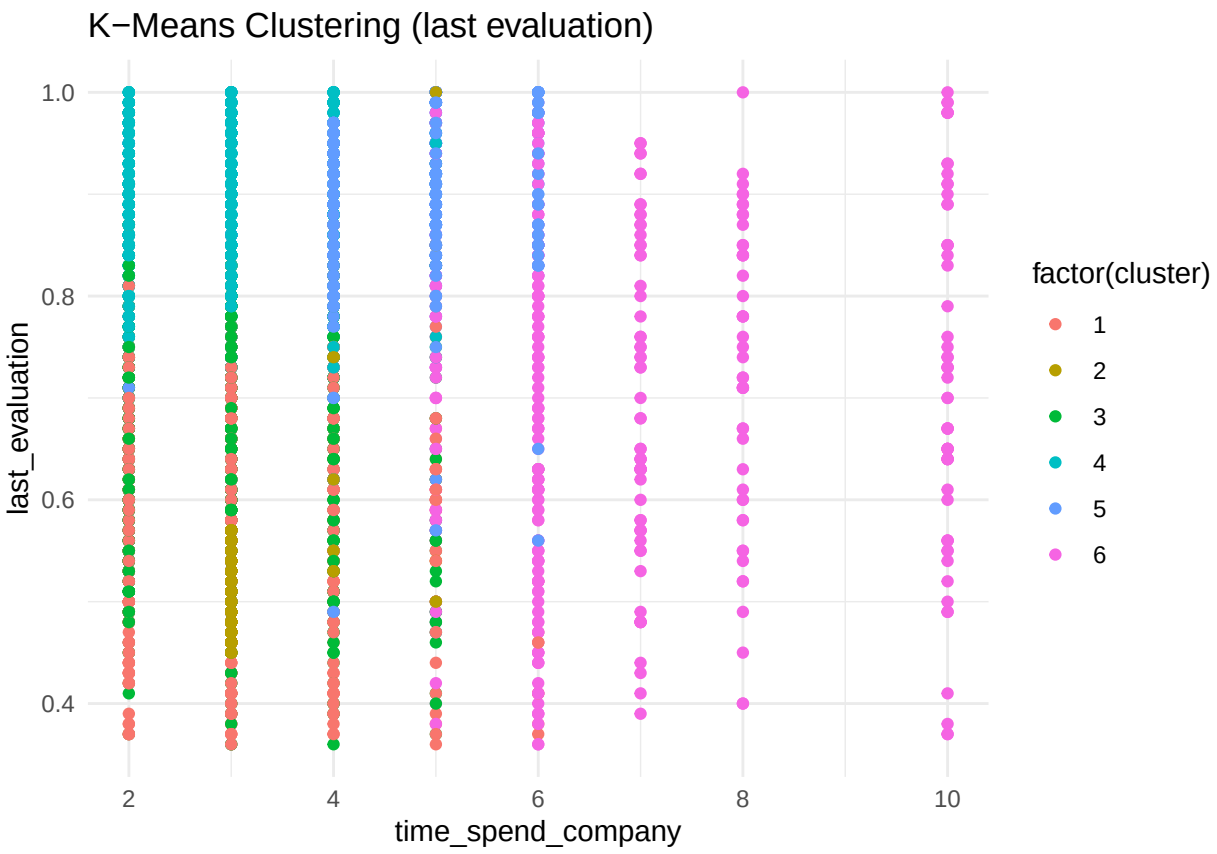



```
ggplot(test_hr, aes(x = left, y = sales, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (sales)") +
  theme_minimal()
```

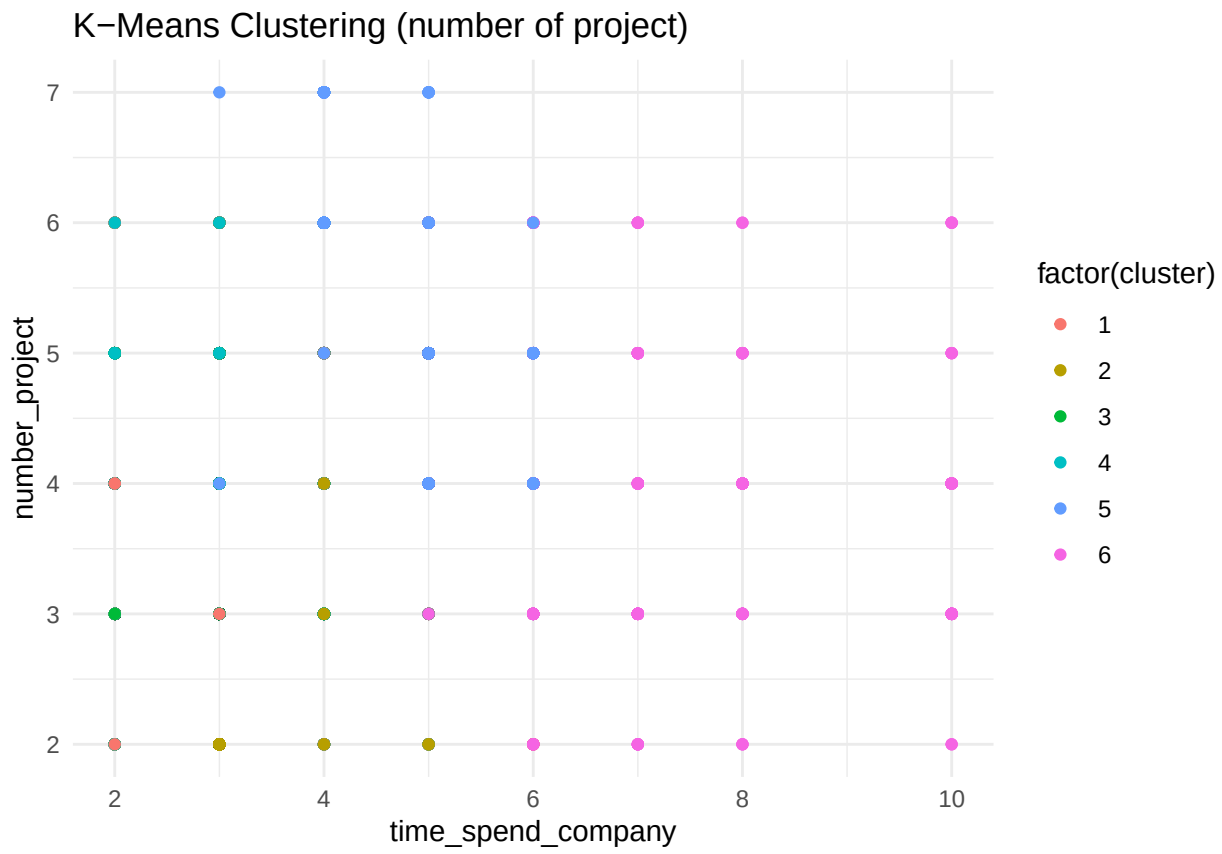


#plotting k mean clustering, considering "time spent at company" as a dependent variable

```
ggplot(test_hr, aes(x = time_spend_company, y = last_evaluation, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (last evaluation)") +
  theme_minimal()
```

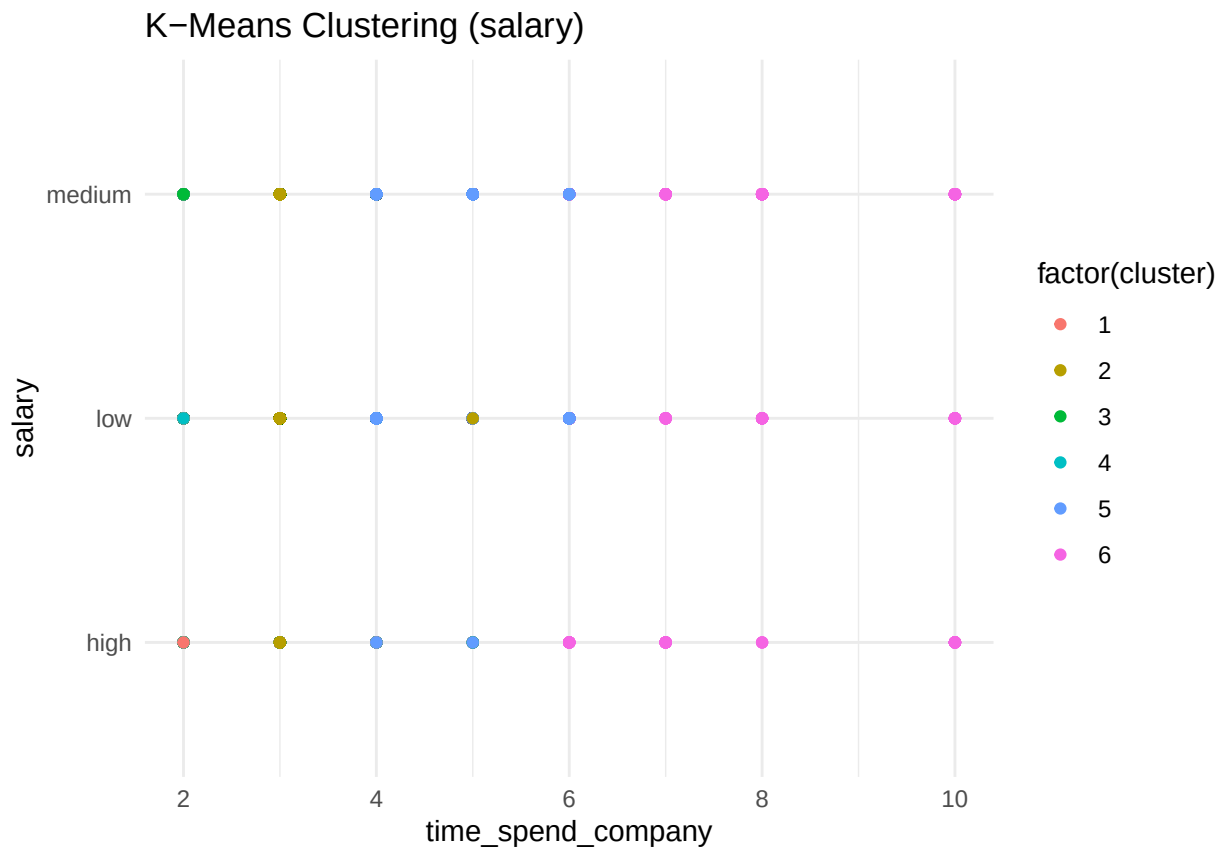



```
ggplot(test_hr, aes(x = time_spend_company, y = number_project, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (number of project)") +
  theme_minimal()
```

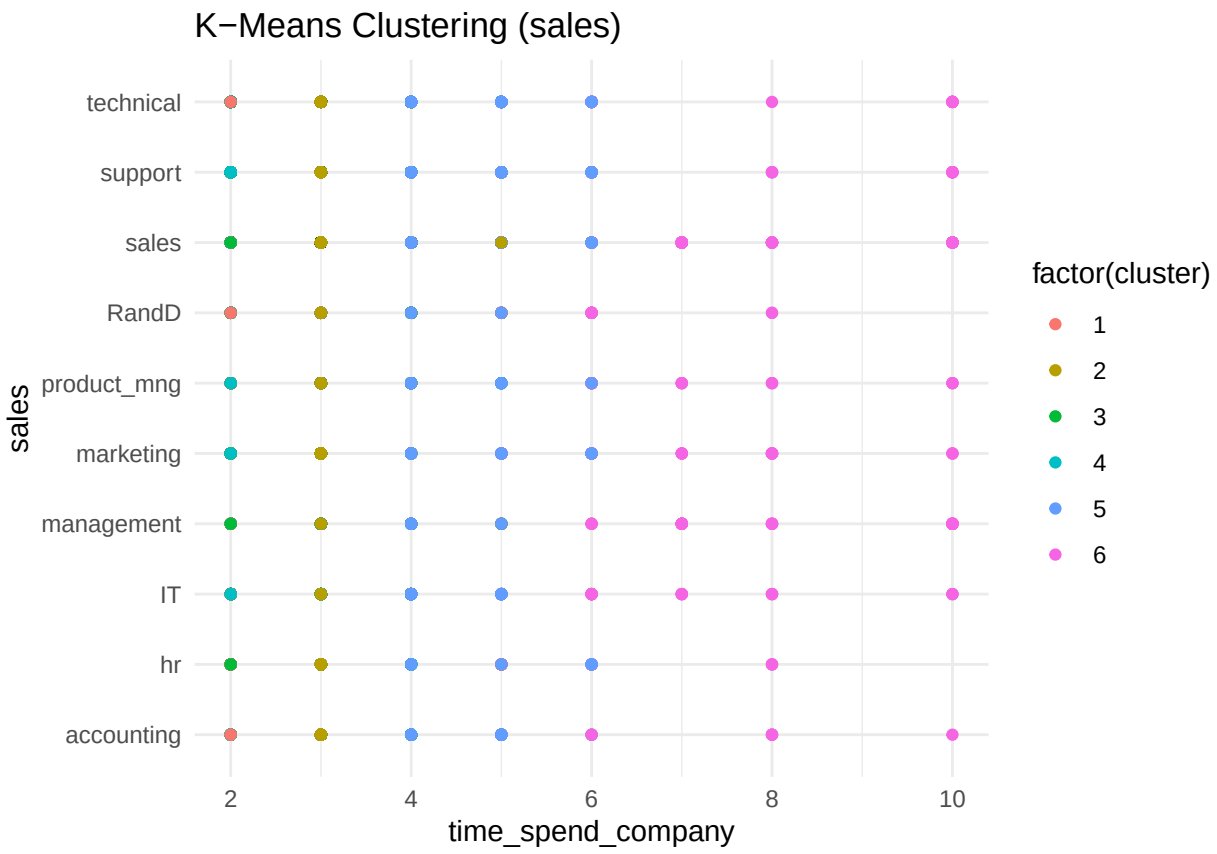



```
ggplot(test_hr, aes(x = time_spend_company, y = average_montly_hours, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (average montly hours)") +
  theme_minimal()
```


```
ggplot(test_hr, aes(x = time_spend_company, y = salary, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (salary)") +
  theme_minimal()
```




```
ggplot(test_hr, aes(x = time_spend_company, y = sales, color = factor(cluster))) +
  geom_point() +
  labs(title = "K-Means Clustering (sales)") +
  theme_minimal()
```


K−Means Clustering (sales)
technical
support
sales
RandD
product_mng
marketing
management
IT
hr
accounting
sales
2
4
6
8
10
time_spend_company
factor(cluster)
1
2
3
4
5
6