---

title: "Project Attrition"

author: "Anthony Yeung"

date: "April 4, 2019"

output: html\_document

library(dplyr)

library(plyr)

library(ggplot2)

#read attrition csv file

atr <- read.csv("C:/Users/AY/Desktop/attrition.csv", header=TRUE)

names(atr)

str(atr)

#1 replacing "Attrition column yes and no with 1 and 0

atr$Attrition <- revalue(atr$Attrition, c("Yes"=1))

atr$Attrition<- revalue(atr$Attrition, c("No"=0))

#replacing "Attrition"" column with atr$Attrition

Attrition <- atr$Attrition

# change "Attrition" column into numeric

atr$Attrition <- as.numeric(as.character(atr$Attrition))

str(atr)

atr[1:3,1:3]

#2 replacing "Gender" column Female and Male with 1 and 0

atr$Gender <- revalue(atr$Gender, c("Female"=1))

atr$Gender <- revalue(atr$Gender, c("Male"=0))

#replacing "Gender" column with atr$Gender

Gender <- atr$Gender

# change "Gender" column into numeric

atr$Gender <- as.numeric(as.character(atr$Gender))

str(atr)

atr[1:3,11:13]

#3 replacing "MaritalStatus" column Single and Married with 1 and 0

atr$MaritalStatus <- revalue(atr$MaritalStatus, c("Single"=1))

atr$MaritalStatus <- revalue(atr$MaritalStatus, c("Married"=0))

#replacing "MaritalStatus" column with atr$MaritalStatus

MaritalStatus <- atr$MaritalStatus

# change "MaritalStatus" column into numeric

atr$MaritalStatus <- as.numeric(as.character(atr$MaritalStatus))

str(atr)

atr[1:3,17:19]

#4 replacing "Over18" column Y with 1

atr$Over18 <- revalue(atr$Over18, c("Y"=1))

#replacing "Over18" column with atr$Over18

Over18 <- atr$Over18

# change "Over18" column into numeric

atr$Over18 <- as.numeric(as.character(atr$Over18))

str(atr)

atr[1:3,21:23]

#5 replacing "OverTime" column Yes and No with 1 and 0

atr$OverTime <- revalue(atr$OverTime, c("Yes"=1))

atr$OverTime <- revalue(atr$OverTime, c("No"=0))

#replacing "OverTime" column with atr$OverTime

OverTime <- atr$OverTime

# change "Overtime" column into numeric

atr$OverTime <- as.numeric(as.character(atr$Overtime))

str(atr)

atr[1:3,22:24]

####Correlation Calculation###

#Correlation between Attrition and age

corr\_age<- cor(atr$Attrition,atr$Age )

corr\_age

corr\_DailyRate <- cor(atr$Attrition, atr$DailyRate)

corr\_DailyRate

corr\_DistanceFromHome <- cor(atr$Attrition, atr$DistanceFromHome)

corr\_DistanceFromHome

corr\_Education <- cor(atr$Attrition, atr$Education)

corr\_Education

corr\_EmployeeCount <- cor(atr$Attrition, atr$EmployeeCount)

corr\_EmployeeCount

corr\_EmployeeNumber <- cor(atr$Attrition, atr$EmployeeNumber)

corr\_EmployeeNumber

corr\_EnvironmentSatisfaction <- cor(atr$Attrition, atr$EnvironmentSatisfaction)

corr\_EnvironmentSatisfaction

corr\_Gender <- cor(atr$Attrition, atr$Gender)

corr\_Gender

corr\_HourlyRate <- cor(atr$Attrition, atr$HourlyRate)

corr\_HourlyRate

corr\_JobInvolvement <- cor(atr$Attrition, atr$JobInvolvement)

corr\_JobInvolvement

corr\_JobLevel <- cor(atr$Attrition, atr$JobLevel)

corr\_JobLevel

corr\_JobSatisfaction <- cor(atr$Attrition, atr$JobSatisfaction)

corr\_JobSatisfaction

corr\_MaitalStatus <- cor(atr$Attrition, atr$MaritalStatus)

corr\_MaitalStatus

corr\_MonthlyIncome <- cor(atr$Attrition, atr$MonthlyIncome)

corr\_MonthlyIncome

corr\_MonthlyRate <- cor(atr$Attrition, atr$MonthlyRate)

corr\_MonthlyRate

corr\_NumCompaniesWorked <- cor(atr$Attrition, atr$NumCompaniesWorked)

corr\_NumCompaniesWorked

corr\_Over18 <- cor(atr$Attrition, atr$Over18)

corr\_Over18

corr\_OverTime <- cor(atr$Attrition, atr$OverTime)

corr\_OverTime

corr\_PercentSalaryHike <- cor(atr$Attrition, atr$PercentSalaryHike)

corr\_PercentSalaryHike

corr\_PerformanceRating <- cor(atr$Attrition, atr$PerformanceRating)

corr\_PerformanceRating

corr\_RelationshipSatisfaction <- cor(atr$Attrition, atr$RelationshipSatisfaction)

corr\_RelationshipSatisfaction

corr\_StandardHours <- cor(atr$Attrition, atr$StandardHours)

corr\_StandardHours

corr\_StockOptionLevel <- cor(atr$Attrition, atr$StockOptionLevel)

corr\_StockOptionLevel

corr\_TotalWorkingYears <- cor(atr$Attrition, atr$TotalWorkingYears)

corr\_TotalWorkingYears

corr\_TrainingTimesLastYear <- cor(atr$Attrition, atr$TrainingTimesLastYear)

corr\_TrainingTimesLastYear

corr\_WorkLifeBalance <- cor(atr$Attrition, atr$WorkLifeBalance)

corr\_WorkLifeBalance

corr\_YearsAtCompany <- cor(atr$Attrition, atr$YearsAtCompany)

corr\_YearsAtCompany

corr\_YearsInCurrentRole <- cor(atr$Attrition, atr$YearsInCurrentRole)

corr\_YearsInCurrentRole

corr\_YearsSinceLastPromotion <- cor(atr$Attrition, atr$YearsSinceLastPromotion)

corr\_YearsSinceLastPromotion

corr\_YearsWithCurrManager <- cor(atr$Attrition, atr$YearsWithCurrManager)

corr\_YearsWithCurrManager

#### PCA Analysis

atr\_select <- atr %>% select(Age, Attrition, DailyRate,DistanceFromHome,Education,EmployeeNumber,EnvironmentSatisfaction,Gender, HourlyRate,JobInvolvement, JobLevel, JobSatisfaction, MonthlyIncome, MonthlyRate,NumCompaniesWorked, PercentSalaryHike, PerformanceRating,RelationshipSatisfaction, StockOptionLevel,TotalWorkingYears, TrainingTimesLastYear, WorkLifeBalance, YearsAtCompany,YearsInCurrentRole, YearsSinceLastPromotion,YearsWithCurrManager)

str(atr\_select)

names(atr\_select)

#cor

summary(atr\_select)

cor(atr\_select)

# correlation of Attrition against all variables

cor(atr\_select,atr\_select$Attrition)

p <- princomp(atr\_select, scores=TRUE, cor=TRUE)

p

summary(p)

plot(p, type="l")

screeplot(p,type="l",main="Scree Plot")

biplot(p)

p$scores[1:10,]

###############

atr\_select <- atr %>% select(Age, Attrition, DailyRate,DistanceFromHome,Education,EmployeeNumber,EnvironmentSatisfaction,Gender, HourlyRate,JobInvolvement, JobLevel, JobSatisfaction, MonthlyIncome, MonthlyRate,NumCompaniesWorked, PercentSalaryHike, PerformanceRating,RelationshipSatisfaction, StockOptionLevel,TotalWorkingYears, TrainingTimesLastYear, WorkLifeBalance, YearsAtCompany,YearsInCurrentRole, YearsSinceLastPromotion,YearsWithCurrManager)

> p <- princomp(atr\_select, scores=TRUE, cor=TRUE)

>

> p

Call:

princomp(x = atr\_select, cor = TRUE, scores = TRUE)

Standard deviations:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9

2.1688786 1.3551618 1.3252642 1.1345965 1.0994319 1.0657088 1.0399280 1.0263119 1.0237611

Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18

1.0060019 1.0001628 0.9841417 0.9758031 0.9556473 0.9468013 0.9409413 0.9290087 0.8832916

Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25 Comp.26

0.8248539 0.7275478 0.7136439 0.5289118 0.4748938 0.4507012 0.3759526 0.2199300

26 variables and 1470 observations.

>

> summary(p)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6

Standard deviation 2.1688786 1.35516176 1.32526423 1.13459652 1.0994319 1.06570881

Proportion of Variance 0.1809244 0.07063321 0.06755097 0.04951189 0.0464904 0.04368213

Cumulative Proportion 0.1809244 0.25155760 0.31910858 0.36862047 0.4151109 0.45879300

Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12

Standard deviation 1.03992804 1.02631185 1.02376109 1.00600194 1.00016282 0.98414172

Proportion of Variance 0.04159424 0.04051215 0.04031103 0.03892461 0.03847406 0.03725134

Cumulative Proportion 0.50038724 0.54089940 0.58121043 0.62013504 0.65860910 0.69586045

Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18

Standard deviation 0.97580308 0.95564733 0.94680128 0.94094131 0.92900871 0.88329162

Proportion of Variance 0.03662276 0.03512545 0.03447818 0.03405271 0.03319451 0.03000785

Cumulative Proportion 0.73248320 0.76760866 0.80208684 0.83613955 0.86933406 0.89934191

Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24

Standard deviation 0.82485391 0.72754784 0.71364390 0.52891178 0.474893844 0.450701200

Proportion of Variance 0.02616861 0.02035869 0.01958799 0.01075953 0.008674006 0.007812753

Cumulative Proportion 0.92551052 0.94586921 0.96545719 0.97621672 0.984890726 0.992703479

Comp.25 Comp.26

Standard deviation 0.375952608 0.219929971

Proportion of Variance 0.005436168 0.001860354

Cumulative Proportion 0.998139646 1.000000000

>

> plot(p, type="l")

# correlation of Attrition against all variables

cor(atr\_select,atr\_select$Attrition)

> cor(atr\_select,atr\_select$Attrition)

[,1]

Age -0.159205007

Attrition 1.000000000

DailyRate -0.056651992

DistanceFromHome 0.077923583

Education -0.031372820

EmployeeNumber -0.010577243

EnvironmentSatisfaction -0.103368978

Gender -0.029453253

HourlyRate -0.006845550

JobInvolvement -0.130015957

JobLevel -0.169104751

JobSatisfaction -0.103481126

MonthlyIncome -0.159839582

MonthlyRate 0.015170213

NumCompaniesWorked 0.043493739

PercentSalaryHike -0.013478202

PerformanceRating 0.002888752

RelationshipSatisfaction -0.045872279

StockOptionLevel -0.137144919

TotalWorkingYears -0.171063246

TrainingTimesLastYear -0.059477799

WorkLifeBalance -0.063939047

YearsAtCompany -0.134392214

YearsInCurrentRole -0.160545004

YearsSinceLastPromotion -0.033018775

YearsWithCurrManager -0.156199316