> #=======================================================================

> # Project Name: Statistical Analysis of Smartphone Usage

> #

> # 1. Environment Setup and Data Import

> # 2. Data Pre Processing and Cleansing

> # 3. Data Visualization

> # 4. Model Building - Linear Regression

> #

> #=========================================================================

> # 1. Environment Set up and Data Import

> #==========================================================================

> # Install Libraries and Packages

> #==========================================================================

> # loading the libraries

> install.packages('ggplot2')

install.packages('ggplot2')

install.packages('corrplot')

install.packages('ROCR')

install.packages('lmtest')

install.packages('pscl')

install.packages('Deducer')

install.packages('caret')

install.packages('rlang')

install.packages('pillar')

library(ggplot2)

library(corrplot)

library(ROCR)

library(lmtest)

library(pscl)

library(Deducer)

library(caret)

library(dplyr)

library(rlang)

library(pillar)

#

# Setup Working Directory

#=========================================================================

getwd()

setwd("D:\\Geeta")

#

Project\_Data<-read.csv("Prj\_data.csv", header=T)

| > dim(Project\_Data)  [1] 200 9  > summary(Project\_Data)  Addict No\_of\_Q Hrs Photo\_.\_Cam Social  Min. :0.00 Min. : 0.000 Min. :1.00 Min. : 1 Min. :1.000  1st Qu.:0.00 1st Qu.: 2.000 1st Qu.:3.00 1st Qu.: 2 1st Qu.:2.000  Median :1.00 Median : 4.000 Median :4.00 Median : 4 Median :3.000  Mean :0.54 Mean : 4.735 Mean :3.69 Mean : 5 Mean :3.145  3rd Qu.:1.00 3rd Qu.: 7.250 3rd Qu.:4.00 3rd Qu.: 7 3rd Qu.:4.000  Max. :1.00 Max. :13.000 Max. :8.00 Max. :13 Max. :5.000  Game\_.\_Ent Occupation Age Gender  Min. : 2.00 Min. :1.000 Min. :1.000 Min. :1.000  1st Qu.: 7.00 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000  Median : 9.00 Median :1.000 Median :2.000 Median :1.000  Mean : 8.93 Mean :1.605 Mean :2.475 Mean :1.445  3rd Qu.:11.00 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:2.000  Max. :13.00 Max. :4.000 Max. :5.000 Max. :2.000  > str(Project\_Data)  'data.frame': 200 obs. of 9 variables:  $ Addict : int 0 1 0 1 0 0 1 0 1 1 ...  $ No\_of\_Q : int 4 6 2 8 1 3 11 2 7 10 ...  $ Hrs : int 3 4 2 6 4 2 7 4 5 8 ...  $ Photo\_.\_Cam: int 6 9 7 11 5 8 4 3 7 9 ...  $ Social : int 4 5 5 3 5 3 5 3 1 3 ...  $ Game\_.\_Ent : int 11 9 11 11 8 4 7 12 9 7 ...  $ Occupation : int 1 1 1 1 4 3 1 1 1 1 ...  $ Age : int 1 2 2 1 5 3 2 2 2 2 ...  $ Gender : int 2 1 2 1 1 1 2 1 2 2 ...  > names(Project\_Data)  [1] "Addict" "No\_of\_Q" "Hrs" "Photo\_.\_Cam" "Social"  [6] "Game\_.\_Ent" "Occupation" "Age" "Gender"  > sum(is.na(Project\_Data)) # Null data is not present  [1] 0  > #  > #Scatter Plots and Correlation  > names(Project\_Data)  [1] "Addict" "No\_of\_Q" "Hrs" "Photo\_.\_Cam" "Social"  [6] "Game\_.\_Ent" "Occupation" "Age" "Gender" |
| --- |
|  |
| | > | | --- | |

> cor(Project\_Data)

Addict No\_of\_Q Hrs Photo\_.\_Cam Social

Addict 1.00000000 0.82101609 0.652763362 0.01659171 -0.004736730

No\_of\_Q 0.82101609 1.00000000 0.541398887 0.11100987 0.043701495

Hrs 0.65276336 0.54139889 1.000000000 0.12787762 -0.007036831

Photo\_.\_Cam 0.01659171 0.11100987 0.127877621 1.00000000 0.016564008

Social -0.00473673 0.04370150 -0.007036831 0.01656401 1.000000000

Game\_.\_Ent -0.01317132 -0.02065966 -0.071609267 -0.04544708 -0.012245584

Occupation -0.22972779 -0.16834945 -0.333267159 -0.24577586 -0.046491421

Age -0.16118638 -0.05951329 -0.302466060 -0.24114852 -0.098912302

Gender 0.14009647 0.12826818 0.251990602 0.52248047 -0.042501466

Game\_.\_Ent Occupation Age Gender

Addict -0.013171317 -0.22972779 -0.161186382 0.14009647

No\_of\_Q -0.020659661 -0.16834945 -0.059513291 0.12826818

Hrs -0.071609267 -0.33326716 -0.302466060 0.25199060

Photo\_.\_Cam -0.045447078 -0.24577586 -0.241148516 0.52248047

Social -0.012245584 -0.04649142 -0.098912302 -0.04250147

Game\_.\_Ent 1.000000000 -0.11070700 -0.007709458 -0.04135574

Occupation -0.110707002 1.00000000 0.844434887 -0.51928271

Age -0.007709458 0.84443489 1.000000000 -0.52583204

Gender -0.041355738 -0.51928271 -0.525832040 1.00000000

> #install.packages("corrplot")

> correlations<- cor(Project\_Data)

> corrplot::corrplot(correlations, method="circle")

> Project\_Data.df <- Project\_Data

> #

> # split 70 percent of the data into the training dataset and

> # 30 percent of the data into the testing dataset:

> #

> set.seed(100)

> ind = sample(2, nrow(Project\_Data.df), replace = TRUE, prob=c(0.7,0.3))

> train.df = Project\_Data.df[ind == 1,]

> test.df = Project\_Data.df[ind == 2,]

> dim(train.df)

[1] 140 9

> dim(test.df)

[1] 60 9

> prop.table(table(train.df$Addict))

0 1

0.4785714 0.5214286

> prop.table(table(test.df$Addict))

0 1

0.4166667 0.5833333

> #Model 1 is with all the variables

> Model1 <- glm(Addict ~ ., data = train.df, family = binomial(link="logit"))

> summary(Model1)

Call:

glm(formula = Addict ~ ., family = binomial(link = "logit"),

data = train.df)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.91637 -0.02413 0.00006 0.01327 1.87004

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -22.3018 9.1761 -2.430 0.01508 \*

No\_of\_Q 1.9892 0.7408 2.685 0.00725 \*\*

Hrs 4.3297 1.5907 2.722 0.00649 \*\*

Photo\_.\_Cam -0.7154 0.3922 -1.824 0.06815 .

Social -0.7601 0.6434 -1.181 0.23743

Game\_.\_Ent 0.5575 0.3124 1.785 0.07433 .

Occupation 1.9166 1.7321 1.107 0.26850

Age -1.3239 1.5844 -0.836 0.40339

Gender -0.1257 1.8172 -0.069 0.94487

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 193.824 on 139 degrees of freedom

Residual deviance: 21.351 on 131 degrees of freedom

AIC: 39.351

Number of Fisher Scoring iterations: 9

> # the significant coefficients are No of Questions, and Hours

> #Likelihood ratio test

> lrtest(Model1)

Error in lrtest(Model1) : could not find function "lrtest"

> # Odds:

> exp(coef(Model1))

(Intercept) No\_of\_Q Hrs Photo\_.\_Cam Social Game\_.\_Ent

2.062780e-10 7.309368e+00 7.592287e+01 4.890017e-01 4.676147e-01 1.746329e+00

Occupation Age Gender

6.797742e+00 2.660994e-01 8.819231e-01

> # Probability:

> exp(coef(Model1))/(1+exp(coef(Model1)))

(Intercept) No\_of\_Q Hrs Photo\_.\_Cam Social Game\_.\_Ent

2.062780e-10 8.796539e-01 9.870000e-01 3.284091e-01 3.186222e-01 6.358776e-01

Occupation Age Gender

8.717577e-01 2.101726e-01 4.686286e-01

> #train data

> pred<-predict(Model1,newdata=train.df,type="response")

> y\_pred\_num <- ifelse(pred>0.5,1,0)

> y\_pred <- factor(y\_pred\_num, levels=c(0,1))

> y\_act <- train.df$Addict

> ##test data

> pred<-predict(Model1,newdata=test.df,type="response")

> y\_pred\_num <- ifelse(pred>0.5,1,0)

> y\_pred <- factor(y\_pred\_num, levels=c(0,1))

> y\_act <- test.df$Addict

> # Step function

> Model2 <- step(glm(Addict ~ ., data = train.df, family = binomial(link="logit")))

Start: AIC=39.35

Addict ~ No\_of\_Q + Hrs + Photo\_.\_Cam + Social + Game\_.\_Ent +

Occupation + Age + Gender

Df Deviance AIC

- Gender 1 21.355 37.355

- Age 1 22.136 38.136

- Occupation 1 22.822 38.822

- Social 1 23.102 39.102

<none> 21.351 39.351

- Game\_.\_Ent 1 25.668 41.668

- Photo\_.\_Cam 1 26.389 42.389

- Hrs 1 40.799 56.799

- No\_of\_Q 1 89.100 105.100

Step: AIC=37.36

Addict ~ No\_of\_Q + Hrs + Photo\_.\_Cam + Social + Game\_.\_Ent +

Occupation + Age

Df Deviance AIC

- Age 1 22.198 36.198

- Occupation 1 22.827 36.827

- Social 1 23.212 37.212

<none> 21.355 37.355

- Game\_.\_Ent 1 25.681 39.681

- Photo\_.\_Cam 1 28.610 42.610

- Hrs 1 41.017 55.017

- No\_of\_Q 1 89.440 103.440

Step: AIC=36.2

Addict ~ No\_of\_Q + Hrs + Photo\_.\_Cam + Social + Game\_.\_Ent +

Occupation

Df Deviance AIC

- Occupation 1 22.858 34.858

- Social 1 23.444 35.444

<none> 22.198 36.198

- Game\_.\_Ent 1 25.730 37.730

- Photo\_.\_Cam 1 28.893 40.893

- Hrs 1 41.856 53.856

- No\_of\_Q 1 91.028 103.028

Step: AIC=34.86

Addict ~ No\_of\_Q + Hrs + Photo\_.\_Cam + Social + Game\_.\_Ent

Df Deviance AIC

- Social 1 23.797 33.797

<none> 22.858 34.858

- Game\_.\_Ent 1 25.827 35.827

- Photo\_.\_Cam 1 29.297 39.297

- Hrs 1 41.929 51.929

- No\_of\_Q 1 92.186 102.186

Step: AIC=33.8

Addict ~ No\_of\_Q + Hrs + Photo\_.\_Cam + Game\_.\_Ent

Df Deviance AIC

<none> 23.797 33.797

- Game\_.\_Ent 1 27.235 35.235

- Photo\_.\_Cam 1 29.760 37.760

- Hrs 1 42.064 50.064

- No\_of\_Q 1 92.421 100.421

> Model2 <- glm(Addict ~ No\_of\_Q +

+ Hrs +

+ Photo\_.\_Cam +

+ Game\_.\_Ent,

+ data = train.df, family = binomial(link="logit"))

> #Individual coefficients

> summary(Model2)

Call:

glm(formula = Addict ~ No\_of\_Q + Hrs + Photo\_.\_Cam + Game\_.\_Ent,

family = binomial(link = "logit"), data = train.df)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.03600 -0.04186 0.00055 0.03343 1.84973

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -20.0324 6.7479 -2.969 0.00299 \*\*

No\_of\_Q 1.5935 0.4859 3.279 0.00104 \*\*

Hrs 3.4399 1.1918 2.886 0.00390 \*\*

Photo\_.\_Cam -0.5188 0.2335 -2.222 0.02626 \*

Game\_.\_Ent 0.4246 0.2498 1.700 0.08921 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 193.824 on 139 degrees of freedom

Residual deviance: 23.797 on 135 degrees of freedom

AIC: 33.797

Number of Fisher Scoring iterations: 9

> # Odds:

> exp(coef(Model2))

(Intercept) No\_of\_Q Hrs Photo\_.\_Cam Game\_.\_Ent

1.995524e-09 4.920908e+00 3.118365e+01 5.952291e-01 1.528961e+00

> # Probability:

> exp(coef(Model2))/(1+exp(coef(Model2)))

(Intercept) No\_of\_Q Hrs Photo\_.\_Cam Game\_.\_Ent

1.995524e-09 8.311070e-01 9.689283e-01 3.731308e-01 6.045807e-01

> #Classification Table

> #train data

> pred<-predict(Model2,newdata=train.df,type="response")

> y\_pred\_num <- ifelse(pred>0.5,1,0)

> y\_pred <- factor(y\_pred\_num, levels=c(0,1))

> y\_act <- train.df$Addict

> ##test data

> pred<-predict(Model2,newdata=test.df,type="response")

> y\_pred\_num <- ifelse(pred>0.5,1,0)

> y\_pred <- factor(y\_pred\_num, levels=c(0,1))

> y\_act <- test.df$Addict