**Prof Neeraj Hatekar Assignment Submission(PGCDS\_2015)**

**By –(i) Sadanand Manekar**

**(ii) Sonali Nimkar**

**Assignment Statement** :-

**Data Provided** :- Year 2014 & Year 2015 data listed under 49 Headings for 40 Movies

**Budget** to make successful movie at box office as per highly optimised model based given data of 2014 and 2015 🡪 **Rs 20 Crores**

**Rate List** :-

1. Lead Male actor A rated – Rs 12 Crores
2. Lead Male actor B rated – Rs 03 Crores
3. Lead Male actor C rated – Rs 0.5 Crores
4. Lead Female actor A rated – Rs 03 Crores
5. Lead Female actor B rated – Rs 1.5 Crores
6. Lead Female actor C rated – Rs 0.3 Crores
7. Cost per GRP – Rs 5000/-
8. Director A rated – Rs 02 Crores
9. Director B/C rated – Rs 1 Cr/Rs 0.5 Cr
10. News per item – Rs 2.5 Lakhs

**Assignment** :- Discover a optimised model using highly influential explanatory variables from data provided to make successful film to generated maximum Lifetime collection in the budget of Rs 20 Crores to be spent on explanatory variable activities as emerged from optimised model

**Solution** :-

Packages used (i) **AER** (ii) **glmulti**

About glmulti package :-

glmulti: An R Package for Easy Automated

Model Selection with ( Generalized ) Linear

Models

**Methodology** :-

Step 1 – Data conditioning

Step 2 - Dividing data as (a) Training Data(2014)

&

(b) Testing data(2015)

Step 3 - Grouping data for time bound analysis

(04 Groups)

Step 4 - Listing most influential explanatory variables emerged from each group data analysis

Step 5 - Testing model on test data

Step 6 - Distribution of budget based on coefficients of significance level of influential explanatory variable

**Step 1 – Data conditioning**

1. All explanatory variables which was listed as rated data was converted into binomial with added parameter for each rating. For example, larm is having three ratings A, B and C. Hence larm is splitted into three explanatory variables as larmA, larmB and larmC. Similarly, larf, dirr, genre is also done. Opening day earnings, weekly earnings columns were dropped. Space and “—“ were replaced with “0”.

2. Long column headings were replaced with H1 to H25. The list is as follows :-

|  |  |  |
| --- | --- | --- |
| **S No** | **Old Column Name** | **New Column Names** |
| 1 | movies | movies |
| 2 | weekend | weekend |
| 3 | holiday | holiday |
| 4 | Lifetime | Lifetime |
| 5 | genre | Drama,Romance,Action,sci,Comedy,Biography,Thriller |
| 6 | sequel | sequel |
| 7 | larm | larmA,larmB,larmC |
| 8 | larf | larfA,larfB,larfC |
| 9 | dirr | dirrA,dirrB,dirrC |
| 10 | fg | fg |
| 11 | IMDb Rating | H1 |
| 12 | cr | H2 |
| 13 | GRPs | H3 |
| 14 | Approx TV Spend (Crs) | H4 |
| 15 | Trailer Likes | H5 |
| 16 | Trailer Dislikes | H6 |
| 17 | Trailer Comments | H7 |
| 18 | Movie Songs Likes | H8 |
| 19 | Movie Songs Dislikes | H9 |
| 20 | Movie Page Likes | H10 |
| 21 | Movie Page PTAT | H11 |
| 22 | New Likes on Page Posts | H12 |
| 23 | Followers of movie handle | H13 |
| 24 | Total tweets on movie handle | H14 |
| 25 | Favourites on movie handle | H15 |
| 26 | Search volume of movie keywords Month of release | H16 |
| 27 | Search volume of cast keywords M, m-1, m+1 | H17 |
| 28 | Twitter | H18 |
| 29 | Comments | H19 |
| 30 | News | H20 |
| 31 | Aggregator | H21 |
| 32 | Blogs | H22 |
| 33 | Videos | H23 |
| 34 | Sentiment Anaylsis - Twitter - Positive (%) | H24 |
| 35 | Sentiment Anaylsis - Twitter - Negative (%) | H25 |

**Step 2 - Dividing data as (a) Training Data(2014) & (b) Testing data(2015)**

(a) Training Data(2014)

|  |  |  |
| --- | --- | --- |
| **S No** | **Movie Name** | **Date of Release** |
| 1 | Ugly | 26-Dec-14 |
| 2 | PK | 19-Dec-14 |
| 3 | Ungli | 28-Nov-14 |
| 4 | Haider | 20-Oct-14 |
| 5 | Action Jackson | 05-Dec-14 |
| 6 | Zid | 28-Nov-14 |
| 7 | Zed Plus | 28-Nov-14 |
| 8 | Happy Ending | 21-Nov-14 |
| 9 | Kill Dil | 14-Nov-14 |
| 10 | The Shaukeens | 07-Nov-14 |
| 11 | Rang Rasiya | 07-Nov-14 |
| 12 | Super Nani | 31-Oct-14 |
| 13 | Roar | 31-Oct-14 |
| 14 | Happy New Year | 24-Oct-14 |
| 15 | Bang Bang | 02-Oct-14 |
| 16 | Daawat-e-Ishq | 19-Sep-14 |
| 17 | Khoobsurat | 19-Sep-14 |
| 18 | Finding Fanny | 12-Sep-14 |
| 19 | Creature 3D | 12-Sep-14 |
| 20 | Mary Kom | 05-Sep-14 |
| 21 | Raja Natwarlal | 29-Aug-14 |
| 22 | Mardaani | 22-Aug-14 |
| 23 | Singham Returns | 15-Aug-14 |
| 24 | Entertainment | 08-Aug-14 |
| 25 | Kick | 25-Jul-14 |

(b) Testing data(2015)

|  |  |  |
| --- | --- | --- |
| **S No** | **Movie Name** | **Date of release** |
| 1 | Hunterrr | 20-Mar-15 |
| 2 | Dilliwaali Zaalim Girlfriend | 20-Mar-15 |
| 3 | NH10 | 13-Mar-15 |
| 4 | Dirty Politics | 06-Mar-15 |
| 5 | Dum Laga Ke Haisha | 27-Feb-15 |
| 6 | Ab Tak Chhappan 2 | 27-Feb-15 |
| 7 | Badlapur | 20-Feb-15 |
| 8 | Roy | 13-Feb-15 |
| 9 | Shamitabh | 30-Jan-15 |
| 10 | Hawaizaada | 30-Jan-15 |
| 11 | Khamoshiyan | 30-Jan-15 |
| 12 | Baby | 23-Jan-15 |
| 13 | Dolly Ki Doli | 23-Jan-15 |
| 14 | Alone | 16-Jan-15 |
| 15 | Tevar | 09-Jan-15 |

**Step 3 - Grouping data for time bound analysis (04 Groups)**

a) Group 1 🡪 Lifetime~movies+weekend+holiday+Drama+Romance+Action+sci+Comedy+Biography+Thriller

b) Group 2 🡪

Lifetime~sequel+larmA+larmB+larmC+larf+larfA++larfB+larfC+dirrA+dirrB+dirrC

c) Group 3 🡪

Lifetime~fg+H1+H2+H3+H5+H6+H7+H8+H9+H10+H11+H12+H13+H14+H15

d) Group 4 🡪

Lifetime~H16+H17+H18+H19+H20+H21+H22+H23+H24+H25

a) The common R code for each model is as follows :-

>rm(list=ls())

>library(AER)

>library(glmulti)

>data1<-read.csv(file.choose(),header=T,sep=",")

>attach(data1)

b) **Group 1 Analysis code as follows** :-

>model1<-glm(Lifetime~movies+weekend+holiday+Drama+Romance+Action+sci+Comedy

+Biography+Thriller-1,data=data1)

>global.model1<-glmulti(model1,level=1,crit="aicc")

>summary(global.model1)

>weightable (global.model1)



model aicc weights

1 Lifetime ~ 1 + movies 205.7789 0.079450183

2 Lifetime ~ 1 206.2357 0.063229725

3 Lifetime ~ 1 + movies + Thriller 206.4123 0.057883849

4 Lifetime ~ 1 + Action 207.2797 0.037515249

5 Lifetime ~ 1 + holiday 207.8440 0.028292787

6 Lifetime ~ 1 + movies + Drama 207.8985 0.027531702

7 Lifetime ~ 1 + Romance 208.0216 0.025888860

.

.

.

.

.

.

.

90 Lifetime ~ 1 + movies + Romance + Comedy 212.5903 0.002636540

91 Lifetime ~ 1 + movies + weekend + sci 212.6060 0.002615948

92 Lifetime ~ 1 + movies + weekend + Comedy 212.6197 0.002598014

93 Lifetime ~ 1 + movies + weekend + holiday 212.6341 0.002579424

94 Lifetime ~ 1 + movies + holiday + Biography 212.6626 0.002542831

95 Lifetime ~ 1 + Drama + Action + Biography 212.6763 0.002525516

96 Lifetime ~ 1 + Drama + Action + sci 212.6824 0.002517834

97 Lifetime ~ 1 + movies + Romance + Biography + Thriller 212.7765 0.002402120

98 Lifetime ~ 1 + movies + sci + Biography 212.7784 0.002399815

99 Lifetime ~ 1 + movies + holiday + Drama + Thriller 212.8315 0.002336965

100 Lifetime ~ 1 + movies + Comedy + Biography 212.8707 0.002291582

**Further R code**:-

>model1<-glm(Lifetime~movies+holiday+Action+Thriller-1,data=data1)

> summary(model1)

**Output :-**

Call:

glm(formula = Lifetime ~ movies + holiday + Action + Thriller -

1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-69.01 -11.34 25.07 64.00 339.50

Coefficients:

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

movies 33.23 51.41 0.646 0.529

holiday 91.41 72.71 1.257 0.231

Action 101.07 68.01 1.486 0.161

Thriller -12.96 89.05 -0.146 0.887

(Dispersion parameter for gaussian family taken to be 13215.16)

Null deviance: 286525 on 17 degrees of freedom

Residual deviance: 171797 on 13 degrees of freedom

AIC: 215

Number of Fisher Scoring iterations: 2

**Further R code**:-

> model1<-glm(Lifetime~movies+holiday+Action-1,data=data1)

> summary(model1)

**Output :-**

Call:

glm(formula = Lifetime ~ movies + holiday + Action - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-71.16 -9.29 25.07 64.00 339.50

Coefficients:

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

movies 33.23 49.58 0.670 0.514

holiday 87.10 64.01 1.361 0.195

Action 103.22 64.01 1.613 0.129

(Dispersion parameter for gaussian family taken to be 12291.21)

Null deviance: 286525 on 17 degrees of freedom

Residual deviance: 172077 on 14 degrees of freedom

AIC: 213.03

Number of Fisher Scoring iterations: 2

**Further R code**:-

> model1<-lm(Lifetime~movies+holiday+Action-1)

> summary(model1)

**Output :-**

Call:

lm(formula = Lifetime ~ movies + holiday + Action - 1)

Residuals:

Min 1Q Median 3Q Max

-71.16 -9.29 25.07 64.00 339.50

Coefficients:

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

movies 33.23 49.58 0.670 0.514

holiday 87.10 64.01 1.361 0.195

Action 103.22 64.01 1.613 0.129

Residual standard error: 110.9 on 14 degrees of freedom

Multiple R-squared: 0.3994, Adjusted R-squared: 0.2707

F-statistic: 3.104 on 3 and 14 DF, p-value: 0.06085

c) **Group 2 Analysis code as follows** :-

> model2<-glm(Lifetime~sequel+larmA+larmB+larmC+larf+larfA++larfB+larfC+dirrA+dirrB+dirrC-+1,data=data1)

> global.model2<-glmulti(model2,level=1,crit="aicc")

>summary(global.model2)

> weightable(global.model2)

model aicc weights

|  |
| --- |
| 0.199956 |
| 0.090378 |
| 0.064226 |
| 0.03277 |
| 0.032549 |
| 0.032083 |
| 0.032083 |
| 0.032083 |
| 0.032083 |
| 0.018991 |

1 Lifetime ~ 1 + sequel + larmA + dirrA 187.5724

2 Lifetime ~ 1 + larmA 189.1606

3 Lifetime ~ 1 + larmA + dirrA 189.8438

4 Lifetime ~ 1 + sequel + larmA 191.1896

5 Lifetime ~ 1 + larmA + dirrC 191.2031

6 Lifetime ~ 1 + sequel + larmA + dirrB + dirrC 191.2320

7 Lifetime ~ 1 + sequel + larmA + dirrA + dirrB + dirrC 191.2320

8 Lifetime ~ 1 + sequel + larmA + dirrA + dirrB 191.2320

9 Lifetime ~ 1 + sequel + larmA + dirrA + dirrC 191.2320

10 Lifetime ~ 1 + sequel + larmA + larmB + dirrA 192.2807

.

.

.

|  |
| --- |
| 0.001273 |
| 0.001273 |
| 0.001233 |
| 0.001233 |
| 0.001233 |
| 0.001233 |
| 0.001226 |
| 0.001082 |
| 0.001074 |
| 0.001074 |
| 0.001074 |

.

90 Lifetime ~ 1 + larmA + larmC + dirrA + dirrC 197.6864

91 Lifetime ~ 1 + larmA + larmC + dirrA + dirrB + dirrC 197.6864

92 Lifetime ~ 1 + larmA + larfB + dirrA + dirrB + dirrC 197.7501

93 Lifetime ~ 1 + larmA + larfB + dirrA + dirrC 197.7501

94 Lifetime ~ 1 + larmA + larfB + dirrA + dirrB 197.7501

95 Lifetime ~ 1 + larmA + larfB + dirrB + dirrC 197.7501

96 Lifetime ~ 1 + larmA + larmB + larfA + dirrA 197.7603

97 Lifetime ~ 1 + sequel + larmA + larmB + larfA + dirrA 198.0104

98 Lifetime ~ 1 + larmA + larfC + dirrA + dirrC 198.0263

99 Lifetime ~ 1 + larmA + larfC + dirrA + dirrB + dirrC 198.0263

100 Lifetime ~ 1 + larmA + larfC + dirrB + dirrC 198.0263



**Further R code**:-

> model2<-glm(Lifetime~sequel+larmA+dirrA-1,data=data1)

> summary(model2)

**Output :-**

Call:

glm(formula = Lifetime ~ sequel + larmA + dirrA - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-89.68 21.89 31.15 55.25 70.82

Coefficients:

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

sequel -130.40 65.81 -1.981 0.067534 .

larmA 162.18 31.02 5.227 0.000128 \*\*\*

dirrA 109.22 49.05 2.227 0.042900 \*

---

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

(Dispersion parameter for gaussian family taken to be 2887.427)

Null deviance: 286525 on 17 degrees of freedom

Residual deviance: 40424 on 14 degrees of freedom

AIC: 188.4

Number of Fisher Scoring iterations: 2

**Further R code**:-

> model2<-glm(Lifetime~larmA+dirrA-1,data=data1)

> summary(model2)

**Output :-**

Call:

glm(formula = Lifetime ~ larmA + dirrA - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-89.68 21.89 31.15 55.25 111.57

Coefficients:

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

larmA 162.18 33.91 4.782 0.000242 \*\*\*

dirrA 65.76 47.96 1.371 0.190544

---

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

(Dispersion parameter for gaussian family taken to be 3450.673)

Null deviance: 286525 on 17 degrees of freedom

Residual deviance: 51760 on 15 degrees of freedom

AIC: 190.6

Number of Fisher Scoring iterations: 2

**Further R code**:-

> model2<-lm(Lifetime~sequel+larmA+dirrA-1)

> summary(model2)

**Output :-**

Call:

lm(formula = Lifetime ~ sequel + larmA + dirrA - 1)

Residuals:

Min 1Q Median 3Q Max

-89.68 21.89 31.15 55.25 70.82

Coefficients:

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

sequel -130.40 65.81 -1.981 0.067534 .

larmA 162.18 31.02 5.227 0.000128 \*\*\*

dirrA 109.22 49.05 2.227 0.042900 \*

---

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

Residual standard error: 53.73 on 14 degrees of freedom

Multiple R-squared: 0.8589, Adjusted R-squared: 0.8287

F-statistic: 28.41 on 3 and 14 DF, p-value: 3.273e-06

d) **Group 3 Analysis code as follows** :-

> model3<-glm(Lifetime~fg+H1+H2+H3+H5+H6+H7+H8+H9+H10+H11+H12+H13+H14+H15-1,data=data1)

> global.model3<-glmulti(model3,level=1,crit="aicc")

After 50 models:

Best model: Lifetime~1+H7

Crit= 187.257022213821

Mean crit= 197.865707317101

After 33900 models:

Best model: Lifetime~1+H9+H13+H14

Crit= 183.498407669745

Mean crit= 190.618550295829

Completed.

> summary(global.model3)

> weightable(global.model3)

model aicc weights

1 Lifetime ~ 1 + H9 + H13 + H14 183.4984 0.163060132

2 Lifetime ~ 1 + H7 + H13 + H14 184.2466 0.112171103

3 Lifetime ~ 1 + H7 + H9 + H13 + H14 185.3755 0.063788266

4 Lifetime ~ 1 + H13 + H14 185.9434 0.048020087

5 Lifetime ~ 1 + H8 + H13 + H14 186.4865 0.036600735

6 Lifetime ~ 1 + H1 + H9 + H13 + H14 186.6101 0.034407090

7 Lifetime ~ 1 + H2 + H13 + H14 186.8899 0.029914918

8 Lifetime ~ 1 + H7 187.2570 0.024898605

9 Lifetime ~ 1 + H2 + H7 + H13 + H14 187.7209 0.019744664

10 Lifetime ~ 1 + H2 + H7 187.7532 0.019428304

.

.

.

.

90 Lifetime ~ 1 + H1 + H7 + H14 193.1738 0.001292302

91 Lifetime ~ 1 + H1 + H7 + H10 + H13 + H14 193.1949 0.001278766

92 Lifetime ~ 1 + H2 + H7 + H10 + H13 + H14 193.1995 0.001275830

93 Lifetime ~ 1 + H6 + H10 + H13 + H14 193.2486 0.001244869

94 Lifetime ~ 1 + H5 + H6 + H7 193.3104 0.001206982

95 Lifetime ~ 1 + H2 + H7 + H13 + H14 + H15 193.3282 0.001196295

96 Lifetime ~ 1 + H1 + H2 + H7 + H13 + H14 193.3995 0.001154362

97 Lifetime ~ 1 + H7 + H10 + H13 + H14 + H15 193.4522 0.001124364

98 Lifetime ~ 1 + H3 + H9 + H10 + H13 + H14 193.5521 0.001069569

99 Lifetime ~ 1 + H2 + H9 + H10 + H13 + H14 193.6379 0.001024651

100 Lifetime ~ 1 + H2 + H5 + H7 + H13 + H14 193.6603 0.001013247



**Further R code**:-

f3<-glm(Lifetime~H7+H9+H13+H14-1,data=data1)

> summary(f3)

**Output :-**

Call:

glm(formula = Lifetime ~ H7 + H9 + H13 + H14 - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-41.729 -22.183 6.352 25.984 61.331

Coefficients:

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

H7 0.011510 0.005844 1.970 0.07056 .

H9 0.003925 0.001927 2.037 0.06260 .

H13 0.004411 0.001150 3.835 0.00207 \*\*

H14 -0.004583 0.001145 -4.001 0.00151 \*\*

---

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

(Dispersion parameter for gaussian family taken to be 1318.648)

Null deviance: 286525 on 17 degrees of freedom

Residual deviance: 17142 on 13 degrees of freedom

AIC: 175.82

Number of Fisher Scoring iterations: 2

**Further R code**:-

f3<-lm(Lifetime~H7+H9+H13+H14-1)

> summary(f3)

**Output :-**

Call:

lm(formula = Lifetime ~ H7 + H9 + H13 + H14 - 1)

Residuals:

Min 1Q Median 3Q Max

-41.729 -22.183 6.352 25.984 61.331

Coefficients:

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

H7 0.011510 0.005844 1.970 0.07056 .

H9 0.003925 0.001927 2.037 0.06260 .

H13 0.004411 0.001150 3.835 0.00207 \*\*

H14 -0.004583 0.001145 -4.001 0.00151 \*\*

---

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

Residual standard error: 36.31 on 13 degrees of freedom

Multiple R-squared: 0.9402, Adjusted R-squared: 0.9218

F-statistic: 51.07 on 4 and 13 DF, p-value: 7.977e-08

e) **Group 4 Analysis code as follows** :-

>model4<-glm(Lifetime~H16+H17+H18+H19+H20+H21+H22+H23+H24+H25-1,data=data1)

> global.model4<-glmulti(model4,level=1,crit="aicc")

Initialization...

TASK: Exhaustive screening of candidate set.

Fitting...

After 50 models:

Best model: Lifetime~1+H17

Crit= 178.667250459805

Mean crit= 191.839016286958

After 1050 models:

Best model: Lifetime~1+H16+H21+H22+H24+H25

Crit= 177.571970893231

Mean crit= 183.635824466697

Completed.

> summary(global.model4)

> weightable(global.model4)

model aicc weights

1 Lifetime ~ 1 + H16 + H21 + H22 + H24 + H25 177.5720 0.116112531

2 Lifetime ~ 1 + H16 + H24 + H25 178.1828 0.085554790

3 Lifetime ~ 1 + H17 178.6673 0.067149403

4 Lifetime ~ 1 + H16 178.8581 0.061037026

5 Lifetime ~ 1 + H16 + H21 + H22 + H23 + H24 + H25 179.6021 0.042076591

6 Lifetime ~ 1 + H16 + H18 + H22 + H23 + H24 + H25 180.7769 0.023385284

7 Lifetime ~ 1 + H16 + H18 + H24 + H25 181.0070 0.020843117

8 Lifetime ~ 1 + H16 + H18 + H22 + H24 + H25 181.0128 0.020783398

9 Lifetime ~ 1 + H17 + H21 + H22 + H24 + H25 181.0846 0.020050456

10 Lifetime ~ 1 + H17 + H25 181.1385 0.019517557

.

.

.

.

.

90 Lifetime ~ 1 + H16 + H18 + H21 185.5053 0.002198812

91 Lifetime ~ 1 + H17 + H19 + H21 185.6539 0.002041366

92 Lifetime ~ 1 + H17 + H20 + H21 185.6949 0.001999884

93 Lifetime ~ 1 + H16 + H20 + H22 + H24 + H25 185.7673 0.001928793

94 Lifetime ~ 1 + H16 + H18 + H20 + H22 + H24 + H25 185.7682 0.001927900

95 Lifetime ~ 1 + H16 + H17 + H22 185.8838 0.001819693

96 Lifetime ~ 1 + H16 + H19 + H23 185.8965 0.001808119

97 Lifetime ~ 1 + H16 + H17 + H19 185.9206 0.001786446

98 Lifetime ~ 1 + H17 + H21 + H22 + H25 185.9427 0.001766801

99 Lifetime ~ 1 + H16 + H20 + H22 + H25 185.9716 0.001741465

100 Lifetime ~ 1 + H16 + H19 + H20 185.9775 0.001736393



**Further R code**:-

> model4<-glm(Lifetime~H16+H21+H22+H24+H25-1,data=data1)

> summary(model4)

**Output :-**

Call:

glm(formula = Lifetime ~ H16 + H21 + H22 + H24 + H25 - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-43.22 -19.03 -15.74 19.66 78.78

Coefficients:

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

H16 7.699e-05 1.021e-05 7.543 6.83e-06 \*\*\*

H21 -4.076e-01 3.954e-01 -1.031 0.323

H22 7.295e-01 6.174e-01 1.182 0.260

H24 1.391e-01 2.234e-01 0.623 0.545

H25 6.473e-01 4.854e-01 1.334 0.207

---

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

(Dispersion parameter for gaussian family taken to be 1576.688)

Null deviance: 286525 on 17 degrees of freedom

Residual deviance: 18920 on 12 degrees of freedom

AIC: 179.5

Number of Fisher Scoring iterations: 2

**Further R code**:-

> model4<-lm(Lifetime~H16+H21+H22+H24+H25-1)

> summary(model4)

**Output :-**

Call:

lm(formula = Lifetime ~ H16 + H21 + H22 + H24 + H25 - 1)

Residuals:

Min 1Q Median 3Q Max

-43.22 -19.03 -15.74 19.66 78.78

Coefficients:

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

H16 7.699e-05 1.021e-05 7.543 6.83e-06 \*\*\*

H21 -4.076e-01 3.954e-01 -1.031 0.323

H22 7.295e-01 6.174e-01 1.182 0.260

H24 1.391e-01 2.234e-01 0.623 0.545

H25 6.473e-01 4.854e-01 1.334 0.207

---

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

Residual standard error: 39.71 on 12 degrees of freedom

Multiple R-squared: 0.934, Adjusted R-squared: 0.9065

F-statistic: 33.95 on 5 and 12 DF, p-value: 1.114e-06

**Step 4 - Listing most influential explanatory variables emerged from each group data analysis**

The most influential explanatory variables are listed below :-

|  |  |
| --- | --- |
| **S No** | **Most Influential explanatory variable** |
| 1 | movies |
| 2 | Holiday |
| 3 | Action |
| 4 | Thriller |
| 5 | sequel |
| 6 | larmA |
| 7 | dirrA |
| 8 | Trailer Comments |
| 9 | Movie Songs Dislikes |
| 10 | Followers of movie handle |
| 11 | Total tweets on movie handle |
| 12 | Search volume of movie keywords Month of release |
| 13 | Aggregator |
| 14 | Blogs |
| 15 | Videos |
| 16 | Sentiment Anaylsis - Twitter - Positive (%) |
| 17 | Sentiment Anaylsis - Twitter - Negative (%) |

**Final analysis in Two stages**

>rm(list=ls())

>library(AER)

>library(glmulti)

>data1<-read.csv(file.choose(),header=T,sep=",")

>attach(data1)

>model51<-glm(Lifetime~movies+holiday+Action+Thriller+sequel+larmA+dirrA-1,data=data1)

>global.model51<-glmulti(model51,level=1,crit="aicc")

>summary(global.model51)

>weightable(global.model51)

> weightable(global.model51)

model aicc weights

1 Lifetime ~ 1 + sequel + larmA + dirrA 187.5724 2.617959e-01

2 Lifetime ~ 1 + Action + sequel + larmA + dirrA 189.1297 1.201732e-01

>rm(list=ls())

>library(AER)

>library(glmulti)

>data1<-read.csv(file.choose(),header=T,sep=",")

>attach(data1)

>model52<-glm(Lifetime~movies+holiday+Action+Thriller+sequel+larmA+dirrA-1,data=data1)

>global.model52<-glmulti(model52,level=1,crit="aicc")

>summary(global.model52)

>weightable(global.model51)

> weightable(global.model52)

model aicc weights

1 Lifetime ~ 1 + H7 + H9 + H14 + H16 170.5457 0.1384340368

2 Lifetime ~ 1 + H9 + H14 + H16 171.8388 0.0725178135

3 Lifetime ~ 1 + H9 + H13 + H14 + H16 172.1224 0.0629297518

4 Lifetime ~ 1 + H7 + H9 + H16 + H21 172.3627 0.0558052494

5 Lifetime ~ 1 + H7 + H9 + H13 + H16 172.5314 0.0512923870



>model53<-glm(Lifetime~H7+H9+H14+H16+Action+sequel+larmA+dirrA-1,data=data1)

>global.model53<-glmulti(model53,level=1,crit="aicc")

>summary(global.model53)

>weightable(global.model53)

> weightable(global.model53)

model aicc weights

1 Lifetime ~ 1 + H16 + Action + larmA 155.1007 5.973291e-01

2 Lifetime ~ 1 + H16 + Action + larmA + dirrA 158.8831 9.013037e-02

3 Lifetime ~ 1 + H14 + H16 + Action + larmA 159.8763 5.485209e-02

4 Lifetime ~ 1 + H16 + Action + sequel + larmA 159.8874 5.454895e-02

5 Lifetime ~ 1 + H9 + H16 + Action + larmA 160.0242 5.094199e-02



model54<-glm(Lifetime~H14+H16+Action+sequel+larmA+dirrA-1,data=data1)

> summary(model54)

Call:

glm(formula = Lifetime ~ H14 + H16 + Action + sequel + larmA +

dirrA - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-16.690 -4.049 -0.166 6.584 45.694

Coefficients:

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

H14 5.181e-04 2.100e-04 2.467 0.0313 \*

H16 1.029e-04 1.262e-05 8.153 5.45e-06 \*\*\*

Action 3.683e+01 1.407e+01 2.617 0.0239 \*

sequel 1.682e+02 6.046e+01 2.781 0.0179 \*

larmA 3.846e+01 1.617e+01 2.378 0.0366 \*

dirrA -1.590e+02 5.182e+01 -3.069 0.0107 \*

---

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

(Dispersion parameter for gaussian family taken to be 295.529)

Null deviance: 286524.7 on 17 degrees of freedom

Residual deviance: 3250.8 on 11 degrees of freedom

AIC: 151.55

Number of Fisher Scoring iterations: 2

> model54<-lm(Lifetime~H14+H16+Action+sequel+larmA+dirrA-1)

> summary(model54)

Call:

lm(formula = Lifetime ~ H14 + H16 + Action + sequel + larmA +

dirrA - 1)

Residuals:

Min 1Q Median 3Q Max

-16.690 -4.049 -0.166 6.584 45.694

Coefficients:

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

H14 5.181e-04 2.100e-04 2.467 0.0313 \*

H16 1.029e-04 1.262e-05 8.153 5.45e-06 \*\*\*

Action 3.683e+01 1.407e+01 2.617 0.0239 \*

sequel 1.682e+02 6.046e+01 2.781 0.0179 \*

larmA 3.846e+01 1.617e+01 2.378 0.0366 \*

dirrA -1.590e+02 5.182e+01 -3.069 0.0107 \*

---

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

Residual standard error: 17.19 on 11 degrees of freedom

Multiple R-squared: 0.9887, Adjusted R-squared: 0.9825

F-statistic: 159.8 on 6 and 11 DF, p-value: 4.788e-10

With training data we have Adjusted R-squared: 0.9825

With most influential explanatory variables as :-

Lifetime~Tweets+Search volume+Action+sequel+larmA+dirrA

**Now for using test data we have** :-

model54<-glm(Lifetime~H14+H16+Action+sequel+larmA+dirrA-1,data=data1)

> summary(model54)

Call:

glm(formula = Lifetime ~ H14 + H16 + Action + sequel + larmA +

dirrA - 1, data = data1)

Deviance Residuals:

Min 1Q Median 3Q Max

-18.1976 0.0000 0.4371 4.1109 17.3802

Coefficients:

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

H14 6.087e-03 2.436e-03 2.499 0.03390 \*

H16 2.759e-05 8.679e-06 3.179 0.01121 \*

Action -3.545e+00 1.144e+01 -0.310 0.76374

sequel 1.015e+01 1.484e+01 0.684 0.51122

larmA 6.946e+01 1.665e+01 4.171 0.00241 \*\*

dirrA -6.115e+01 2.125e+01 -2.878 0.01825 \*

---

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

(Dispersion parameter for gaussian family taken to be 89.10962)

Null deviance: 19040.10 on 15 degrees of freedom

Residual deviance: 801.99 on 9 degrees of freedom

AIC: 116.25

Number of Fisher Scoring iterations: 2

> model54<-lm(Lifetime~H14+H16+Action+sequel+larmA+dirrA-1)

> summary(model54)

Call:

lm(formula = Lifetime ~ H14 + H16 + Action + sequel + larmA +

dirrA - 1)

Residuals:

Min 1Q Median 3Q Max

-18.1976 0.0000 0.4371 4.1109 17.3802

Coefficients:

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

H14 6.087e-03 2.436e-03 2.499 0.03390 \*

H16 2.759e-05 8.679e-06 3.179 0.01121 \*

Action -3.545e+00 1.144e+01 -0.310 0.76374

sequel 1.015e+01 1.484e+01 0.684 0.51122

larmA 6.946e+01 1.665e+01 4.171 0.00241 \*\*

dirrA -6.115e+01 2.125e+01 -2.878 0.01825 \*

---

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

Residual standard error: 9.44 on 9 degrees of freedom

Multiple R-squared: 0.9579, Adjusted R-squared: 0.9298

F-statistic: 34.11 on 6 and 9 DF, p-value: 1.077e-05

Hence total budget allocation of Rs 20 Crores shall be as follows :-

i) Total Tweets on movie handle – **Rs 05 Crores**

ii) Search Volume of movie keyword in month of release – **Rs 01 Crores**

iii) Leading Male actor Male – **Rs 12 Crores**

iv) Director A Grade – **Rs 02 Crores**