Web-Data-Analysis-R-Programming-Project.R

labsuser

2022-09-18

```
rm(list = ls(all = TRUE))
library(readxl)
webdata <- read_excel('1555058318_internet_dataset.xlsx')

head(webdata)
```

```
## # A tibble: 6 x 8
##   Bounces Exits Continent Sourcegroup Timeinpage Uniquepageviews Visits
##     <dbl> <dbl> <chr>     <chr>            <dbl>            <dbl>  <dbl>
## 1       0     0 OC        (direct)            18                1      0
## 2       0     0 N.America (direct)             4                1      0
## 3       0     0 N.America Others              35                1      0
## 4       0     0 N.America public.tab…         70                1      0
## 5       0     0 N.America public.tab…         81                1      0
## 6       0     0 N.America public.tab…         75                1      0
## # … with 1 more variable: BouncesNew <dbl>
```

```
summary(webdata)
```

```
##     Bounces           Exits         Continent         Sourcegroup       
##  Min.   : 0.000   Min.   : 0.000   Length:32109       Length:32109      
##  1st Qu.: 0.000   1st Qu.: 1.000   Class :character   Class :character  
##  Median : 1.000   Median : 1.000   Mode  :character   Mode  :character  
##  Mean   : 0.713   Mean   : 0.906                                        
##  3rd Qu.: 1.000   3rd Qu.: 1.000                                        
##  Max.   :30.000   Max.   :36.000                                        
##    Timeinpage       Uniquepageviews      Visits         BouncesNew       
##  Min.   :    0.00   Min.   : 1.000   Min.   : 0.000   Min.   :0.00000  
##  1st Qu.:    0.00   1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:0.00000  
##  Median :    0.00   Median : 1.000   Median : 1.000   Median :0.01000  
##  Mean   :   73.18   Mean   : 1.114   Mean   : 0.906   Mean   :0.00713  
##  3rd Qu.:   10.00   3rd Qu.: 1.000   3rd Qu.: 1.000   3rd Qu.:0.01000  
##  Max.   :46745.00   Max.   :45.000   Max.   :45.000   Max.   :0.30000
```

```
cor(webdata$Uniquepageviews, webdata$Visits)
```

```
## [1] 0.8144457
```

```
anova <- aov(Uniquepageviews ~ Visits, data = webdata)
summary(anova)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## Visits          1   8052    8052   63257 <2e-16 ***
```

```
## Residuals   32107   4087       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova <- aov(Exits ~., data = webdata)
summary(Anova)
```

```
##                    Df Sum Sq Mean Sq   F value   Pr(>F)
## Bounces             1  10578   10578 1.043e+05  < 2e-16 ***
## Continent           5      3       1 5.960e+00 1.62e-05 ***
## Sourcegroup         8      7       1 8.760e+00 4.89e-12 ***
## Timeinpage          1    130     130 1.279e+03  < 2e-16 ***
## Uniquepageviews     1   1573    1573 1.552e+04  < 2e-16 ***
## Visits              1      1       1 5.014e+00   0.0251 *
## Residuals       32091   3254       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
An <- aov(Timeinpage ~., data = webdata)
summary(An)
```

```
##                    Df    Sum Sq   Mean Sq  F value   Pr(>F)
## Bounces             1 5.947e+07  59466495  422.868  < 2e-16 ***
## Exits               1 1.304e+08 130400662  927.283  < 2e-16 ***
## Continent           5 4.767e+06    953431    6.780 2.51e-06 ***
## Sourcegroup         8 1.545e+06    193153    1.374    0.202
## Uniquepageviews     1 1.791e+08 179133934 1273.826  < 2e-16 ***
## Visits              1 1.073e+08 107321113  763.163  < 2e-16 ***
## Residuals       32091 4.513e+09    140627
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
webdata$Bounces <- webdata$Bounces * 0.01
log_reg <- glm(Bounces ~ Timeinpage + Exits + Continent + Sourcegroup +
Uniquepageviews + Visits, data = webdata, family = "binomial")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
glm!
```

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

```
summary(log_reg)
```

```
##
## Call:
## glm(formula = Bounces ~ Timeinpage + Exits + Continent + Sourcegroup +
##     Uniquepageviews + Visits, family = "binomial", data = webdata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.26149  -0.02406   0.00206   0.00895   1.81288
##
```

```
## Coefficients:
##                                          Estimate Std. Error z value
Pr(>|z|)
## (Intercept)                            -4.9667681  0.6784678  -7.321 2.47e-
13
## Timeinpage                             -0.0010294  0.0005774  -1.783
0.0746
## Exits                                   1.3907608  0.3356504   4.143 3.42e-
05
## ContinentAS                             0.0022768  0.6932044   0.003
0.9974
## ContinentEU                            -0.0069240  0.6786600  -0.010
0.9919
## ContinentN.America                      0.0101334  0.6674188   0.015
0.9879
## ContinentOC                             0.0201123  0.7333671   0.027
0.9781
## ContinentSA                             0.0237507  0.7914250   0.030
0.9761
## Sourcegroupfacebook                    -0.0241949  1.1045171  -0.022
0.9825
## Sourcegroupgoogle                      -0.0783631  0.1720157  -0.456
0.6487
## SourcegroupOthers                      -0.0767919  0.2182692  -0.352
0.7250
## Sourcegrouppublic.tableausoftware.com -0.2528285  0.4923123  -0.514
0.6076
## Sourcegroupreddit.com                  -0.0092792  0.4709304  -0.020
0.9843
## Sourcegroupt.co                         0.0148690  0.2760157   0.054
0.9570
## Sourcegrouptableausoftware.com         -0.1129305  0.3190762  -0.354
0.7234
## Sourcegroupvisualisingdata.com         -0.0822525  0.4614866  -0.178
0.8585
## Uniquepageviews                        -3.2363108  0.5791664  -5.588 2.30e-
08
## Visits                                  2.1941121  0.5202216   4.218 2.47e-
05
##
## (Intercept)                            ***
## Timeinpage                             .
## Exits                                  ***
## ContinentAS
## ContinentEU
## ContinentN.America
## ContinentOC
## ContinentSA
## Sourcegroupfacebook
## Sourcegroupgoogle
```

```
## SourcegroupOthers
## Sourcegrouppublic.tableausoftware.com
## Sourcegroupreddit.com
## Sourcegroupt.co
## Sourcegrouptableausoftware.com
## Sourcegroupvisualisingdata.com
## Uniquepageviews                         ***
## Visits                                  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.937  on 32108  degrees of freedom
## Residual deviance:  96.514  on 32091  degrees of freedom
## AIC: 506.56
##
## Number of Fisher Scoring iterations: 11
```