Web_Data_Analysis.R

HP

2021-09-18

```
install.packages("readxl")
```

```
## Installing package into 'C:/Users/HP/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'readxl' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\HP\AppData\Local\Temp\RtmpaegxPF\downloaded_packages
```

```
install.packages('plyr', repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/HP/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'plyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\HP\AppData\Local\Temp\RtmpaegxPF\downloaded_packages
```

```
install.packages("caTools")
```

```
## Installing package into 'C:/Users/HP/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'caTools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\HP\AppData\Local\Temp\RtmpaegxPF\downloaded_packages
```

```
install.packages("e1071")
```

```
## Installing package into 'C:/Users/HP/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## 
##   There is a binary version available but the source version
##   is later:
##       binary source needs_compilation
## e1071  1.7-8  1.7-9              TRUE
## 
##   Binaries will be installed
## package 'e1071' successfully unpacked and MD5 sums checked
## 
## The downloaded binary packages are in
##   C:\Users\HP\AppData\Local\Temp\RtmpaegxPF\downloaded_packages
```

```
install.packages("rmarkdown")
```

```
## Error in install.packages : Updating loaded packages
```

```
library(rmarkdown)
library(readxl)
library(plyr)
library(caTools)
library(e1071)

data<-read_excel("C:\\Users\\HP\\Desktop\\R Projects\\3-Webdata project\\1555058318_internet_
dataset.xlsx")
summary(data)
```

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

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

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

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

```
##                    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
```

```
anov2<-aov(Timeinpage~.,data=data)
summary(anov2)
```

```
##                    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
```

```
data$Bounces=data$Bounces*0.01
rmm<-glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits,data=data,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(rmm)
```

```
## 
## Call:
## glm(formula = Bounces ~ Timeinpage + Continent + Exits + Sourcegroup +
##     Uniquepageviews + Visits, family = "binomial", data = data)
## 
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.26149  -0.02406  0.00206  0.00895  1.81288
## 
## Coefficients:
##                                     Estimate Std. Error z value
## (Intercept)                        -4.9667681  0.6784678  -7.321
## Timeinpage                         -0.0010294  0.0005774  -1.783
## ContinentAS                         0.0022768  0.6932044   0.003
## ContinentEU                        -0.0069240  0.6786600  -0.010
## ContinentN.America                  0.0101334  0.6674188   0.015
## ContinentOC                         0.0201123  0.7333671   0.027
## ContinentSA                         0.0237507  0.7914250   0.030
## Exits                               1.3907608  0.3356504   4.143
## Sourcegroupfacebook                -0.0241949  1.1045171  -0.022
## Sourcegroupgoogle                  -0.0783631  0.1720157  -0.456
## SourcegroupOthers                  -0.0767919  0.2182692  -0.352
## Sourcegrouppublic.tableausoftware.com -0.2528285  0.4923123  -0.514
## Sourcegroupreddit.com              -0.0092792  0.4709304  -0.020
## Sourcegroupt.co                     0.0148690  0.2760157   0.054
## Sourcegrouptableausoftware.com     -0.1129305  0.3190762  -0.354
## Sourcegroupvisualisingdata.com     -0.0822525  0.4614866  -0.178
## Uniquepageviews                    -3.2363108  0.5791664  -5.588
## Visits                              2.1941121  0.5202216   4.218
##                                     Pr(>|z|)
## (Intercept)                         2.47e-13 ***
## Timeinpage                          0.0746 .
## ContinentAS                         0.9974
## ContinentEU                         0.9919
## ContinentN.America                  0.9879
## ContinentOC                         0.9781
## ContinentSA                         0.9761
## Exits                               3.42e-05 ***
## Sourcegroupfacebook                 0.9825
## Sourcegroupgoogle                   0.6487
## SourcegroupOthers                   0.7250
## Sourcegrouppublic.tableausoftware.com  0.6076
## Sourcegroupreddit.com               0.9843
## Sourcegroupt.co                     0.9570
## Sourcegrouptableausoftware.com      0.7234
## Sourcegroupvisualisingdata.com      0.8585
## Uniquepageviews                     2.30e-08 ***
## Visits                              2.47e-05 ***
## ---
## 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
```