

Internet Project

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com> (<http://rmarkdown.rstudio.com>).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library("openxlsx")
```

```
## Warning: package 'openxlsx' was built under R version 3.4.4
```

```
library('ggplot2')
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
library('rmarkdown')
```

```
## Warning: package 'rmarkdown' was built under R version 3.4.4
```

```
mydata=read.xlsx(file.choose(), sheet = 1, startRow = 1, colNames = TRUE)
head(mydata)
```

```
##   Bounces Exits Continent Sourcegroup Timeinpage
## 1      0     0      OC      (direct)         18
## 2      0     0 N.America      (direct)          4
## 3      0     0 N.America      Others         35
## 4      0     0 N.America public.tableausoftware.com 70
## 5      0     0 N.America public.tableausoftware.com 81
## 6      0     0 N.America public.tableausoftware.com 75
##   Uniquepageviews Visits BouncesNew
## 1                1      0          0
## 2                1      0          0
## 3                1      0          0
## 4                1      0          0
## 5                1      0          0
## 6                1      0          0
```

```
summary(mydata)
```

```
##      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
```

```
names(mydata)
```

```
## [1] "Bounces"      "Exits"        "Continent"    "Sourcegroup"
## [5] "Timeinpage"   "Uniquepageviews" "Visits"       "BouncesNew"
```

```
ano1<-aov(Uniquepageviews~Visits, data=mydata)
summary(ano1)
```

```
##              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
```

```
ano2<-aov(Exits~Timeinpage+Continent+
          Sourcegroup+Bounces+Uniquepageviews+Visits, data=mydata)
summary(ano2)
```

```
##              Df Sum Sq Mean Sq  F value Pr(>F)
## Timeinpage     1      0      0 2.690e-01 0.6039
## Continent       5      1      0 1.464e+00 0.1978
## Sourcegroup     8    325    41 4.003e+02 <2e-16 ***
## Bounces         1 10393 10393 1.025e+05 <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
```

```
ano3<-aov(Timeinpage~Exits+Continent+Sourcegroup+Bounces+
          Uniquepageviews+Visits, data=mydata)
summary(ano3)
```

```
##              Df      Sum Sq   Mean Sq   F value    Pr(>F)
## Exits          1 8.767e+03     8767      0.062    0.803
## Continent      5 5.379e+06   1075878     7.651 3.39e-07 ***
## Sourcegroup    8 6.414e+06    801769     5.701 2.85e-07 ***
## Bounces        1 1.844e+08 184377226 1311.112 < 2e-16 ***
## 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
```

```
mydata$Bounces=as.factor(mydata$Bounces)
```

```
reg1<-glm(Bounces~Timeinpage+Continent+Exits+
          Sourcegroup+Uniquepageviews+
          Visits, data=mydata,family="binomial")
```

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

```
summary(reg1)
```

```
##
## Call:
## glm(formula = Bounces ~ Timeinpage + Continent + Exits + Sourcegroup +
##       Uniquepageviews + Visits, family = "binomial", data = mydata)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7501  -0.0004   0.3530   0.4070   8.4904
##
## Coefficients:
##              Estimate Std. Error z value
## (Intercept)      -1.291e+01  5.655e-01 -22.835
## Timeinpage       -1.372e-02  3.741e-04 -36.688
## ContinentAS      -3.874e-01  2.764e-01  -1.402
## ContinentEU      -4.278e-01  2.723e-01  -1.571
## ContinentN.America -4.143e-01  2.685e-01  -1.543
## ContinentOC      -5.999e-02  2.966e-01  -0.202
## ContinentSA      -4.062e-01  3.149e-01  -1.290
## Exits             9.270e+00  3.002e-01  30.877
## Sourcegroupfacebook -1.546e-01  4.271e-01  -0.362
## Sourcegroupgoogle  -4.831e-01  6.847e-02  -7.055
## SourcegroupOthers  -1.496e-01  8.747e-02  -1.711
## Sourcegrouppublic.tableausoftware.com -1.370e+00  1.272e-01 -10.773
## Sourcegroupreddit.com  3.862e-01  2.302e-01   1.678
## Sourcegroupt.co      4.093e-01  1.391e-01   2.942
## Sourcegrouptableausoftware.com -2.870e-01  1.183e-01  -2.426
## Sourcegroupvisualisingdata.com -1.918e-01  1.752e-01  -1.095
## Uniquepageviews     -3.045e+00  1.997e-01 -15.243
## Visits              9.838e+00  3.032e-01  32.446
##
## Pr(>|z|)
## (Intercept)      < 2e-16 ***
## Timeinpage       < 2e-16 ***
## ContinentAS      0.16101
## ContinentEU      0.11622
## ContinentN.America 0.12287
## ContinentOC      0.83972
## ContinentSA      0.19704
## Exits            < 2e-16 ***
## Sourcegroupfacebook 0.71733
## Sourcegroupgoogle  1.72e-12 ***
## SourcegroupOthers   0.08712 .
## Sourcegrouppublic.tableausoftware.com < 2e-16 ***
## Sourcegroupreddit.com 0.09342 .
## Sourcegroupt.co     0.00326 **
## Sourcegrouptableausoftware.com 0.01525 *
## Sourcegroupvisualisingdata.com 0.27364
## Uniquepageviews     < 2e-16 ***
## Visits             < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##      Null deviance: 41894  on 32108  degrees of freedom
## Residual deviance: 11965  on 32091  degrees of freedom
## AIC: 12001
##
## Number of Fisher Scoring iterations: 11
```