Project No 3: Web Data Analysis

DESCRIPTION

**Background and Objective:**  
The web analytics team of www.datadb.com is interested to understand the web activities of the site, which are the sources used to access the website. They have a database that states the keywords of time in the page, source group, bounces, exits, unique page views, and visits.

**Domain:** Web

**Dataset Description:**  
The variables in the dataset are defined here for better understanding:

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Bounces | It represents the percentage of visitors who enter the site and "bounce" (leave the site) rather than continuing to view other pages within the same site. |
| Continent | It shows the continent from which the site has been accessed. |
| Source group | It shows how the visitor has accessed the site. |
| Time on page | It shows how long the user has spent on that particular page of the website. |
| Unique pageview | It represents the number of sessions during which that page was viewed one or more times. |
| Visits | A visit counts all visitors, no matter how many times the same visitor may have been to your site. |

**Analysis Tasks:**  
The team is targeting the following issues:

* The team wants to analyze each variable of the data collected through data summarization to get a basic understanding of the dataset and to prepare for further analysis.
* As mentioned earlier, a unique page view represents the number of sessions during which that page was viewed one or more times. A visit counts all instances, no matter how many times the same visitor may have been to your site. So the team needs to know whether the unique page view value depends on visits.
* Find out the probable factors from the dataset, which could affect the exits. Exit Page Analysis is usually required to get an idea about why a user leaves the website for a session and moves on to another one. Please keep in mind that exits should not be confused with bounces.
* Every site wants to increase the time on page for a visitor. This increases the chances of the visitor understanding the site content better and hence there are more chances of a transaction taking place. Find the variables which possibly have an effect on the time on page.
* A high bounce rate is a cause of alarm for websites which depend on visitor engagement. Help the team in determining the factors that are impacting the bounce.

Project No 3

Karthik

2022-07-03

library(readxl)  
internet <- read\_excel("C:/Users/HP/Downloads/internet.xlsx")  
 View(internet)  
#We will import the dataset And we will find the summary & str of the dataset  
head(internet)

## # A tibble: 6 × 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.tableausoftw… 70 1 0  
## 5 0 0 N.America public.tableausoftw… 81 1 0  
## 6 0 0 N.America public.tableausoftw… 75 1 0  
## # … with 1 more variable: BouncesNew <dbl>

tail(internet)

## # A tibble: 6 × 8  
## Bounces Exits Continent Sourcegroup Timeinpage Uniquepageviews Visits  
## <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 1 1 N.America t.co 0 1 1  
## 2 1 1 N.America public.tableausoftw… 12 2 2  
## 3 2 2 N.America (direct) 0 2 2  
## 4 2 2 N.America (direct) 0 2 2  
## 5 2 2 N.America (direct) 0 2 2  
## 6 2 2 N.America google 0 2 2  
## # … with 1 more variable: BouncesNew <dbl>

summary(internet)

## 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

str(internet)

## tibble [32,109 × 8] (S3: tbl\_df/tbl/data.frame)  
## $ Bounces : num [1:32109] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Exits : num [1:32109] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Continent : chr [1:32109] "OC" "N.America" "N.America" "N.America" ...  
## $ Sourcegroup : chr [1:32109] "(direct)" "(direct)" "Others" "public.tableausoftware.com" ...  
## $ Timeinpage : num [1:32109] 18 4 35 70 81 75 186 710 712 344 ...  
## $ Uniquepageviews: num [1:32109] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Visits : num [1:32109] 0 0 0 0 0 0 0 0 1 1 ...  
## $ BouncesNew : num [1:32109] 0 0 0 0 0 0 0 0 0 0 ...

library(psych)  
describe(internet)

## vars n mean sd median trimmed mad min max range  
## Bounces 1 32109 0.71 0.71 1.00 0.68 0.00 0 30.0 30.0  
## Exits 2 32109 0.91 0.70 1.00 0.89 0.00 0 36.0 36.0  
## Continent\* 3 32109 3.66 0.84 4.00 3.73 0.00 1 6.0 5.0  
## Sourcegroup\* 4 32109 3.67 2.25 3.00 3.43 1.48 1 9.0 8.0  
## Timeinpage 5 32109 73.18 394.44 0.00 10.32 0.00 0 46745.0 46745.0  
## Uniquepageviews 6 32109 1.11 0.61 1.00 1.00 0.00 1 45.0 44.0  
## Visits 7 32109 0.91 0.73 1.00 0.88 0.00 0 45.0 45.0  
## BouncesNew 8 32109 0.01 0.01 0.01 0.01 0.00 0 0.3 0.3  
## skew kurtosis se  
## Bounces 7.06 219.35 0.00  
## Exits 11.11 427.57 0.00  
## Continent\* -0.46 1.27 0.00  
## Sourcegroup\* 0.78 -0.20 0.01  
## Timeinpage 57.33 6234.83 2.20  
## Uniquepageviews 24.41 1273.98 0.00  
## Visits 13.75 633.85 0.00  
## BouncesNew 7.06 219.35 0.00

# To find if there is missing values in the dataset  
sum(is.na(internet))

## [1] 0

#There is no missing values present in the dataset  
  
#Q2 As mentioned earlier, a unique page view represents the number of sessions during which that page was viewed one or more times.   
#A visit counts all instances, no matter how many times the same visitor may have been to your site.   
#So the team needs to know whether the unique page view value depends on visits.  
  
cor(internet$Bounces,internet$Uniquepageviews)

## [1] 0.6591015

cor(internet$Visits,internet$Uniquepageviews)

## [1] 0.8144457

#We find out the regression model between Visits & Uniquepageviews  
model=lm(Visits~Uniquepageviews,data=internet)  
summary(model)

##   
## Call:  
## lm(formula = Visits ~ Uniquepageviews, data = internet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.1688 0.2045 0.2045 0.2045 1.6557   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.171539 0.004893 -35.05 <2e-16 \*\*\*  
## Uniquepageviews 0.967019 0.003845 251.51 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4236 on 32107 degrees of freedom  
## Multiple R-squared: 0.6633, Adjusted R-squared: 0.6633   
## F-statistic: 6.326e+04 on 1 and 32107 DF, p-value: < 2.2e-16

ano<-aov(Uniquepageviews~Visits, data=internet)  
ano

## Call:  
## aov(formula = Uniquepageviews ~ Visits, data = internet)  
##   
## Terms:  
## Visits Residuals  
## Sum of Squares 8052.261 4087.037  
## Deg. of Freedom 1 32107  
##   
## Residual standard error: 0.3567832  
## Estimated effects may be unbalanced

summary(ano)

## 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

#Q3 Find out the probable factors from the dataset, which could affect the exits.   
#Exit Page Analysis is usually required to get an idea about why a user leaves the website for a session and moves on to another one.  
#Please keep in mind that exits should not be confused with bounces  
  
model1<-lm(Exits~.,data=internet)  
model1

##   
## Call:  
## lm(formula = Exits ~ ., data = internet)  
##   
## Coefficients:  
## (Intercept) Bounces   
## -3.223e-02 5.132e-01   
## ContinentAS ContinentEU   
## 9.433e-03 1.668e-02   
## ContinentN.America ContinentOC   
## -2.858e-03 2.562e-03   
## ContinentSA Sourcegroupfacebook   
## -1.196e-03 7.563e-02   
## Sourcegroupgoogle SourcegroupOthers   
## 3.923e-02 -4.697e-03   
## Sourcegrouppublic.tableausoftware.com Sourcegroupreddit.com   
## -5.546e-02 1.170e-02   
## Sourcegroupt.co Sourcegrouptableausoftware.com   
## -1.429e-02 -3.189e-02   
## Sourcegroupvisualisingdata.com Timeinpage   
## -5.212e-02 1.518e-05   
## Uniquepageviews Visits   
## 4.943e-01 1.254e-02   
## BouncesNew   
## NA

summary(model1)

##   
## Call:  
## lm(formula = Exits ~ ., data = internet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5364 -0.0437 -0.0045 0.0293 1.9883   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -3.223e-02 1.845e-02 -1.747 0.08058  
## Bounces 5.132e-01 4.599e-03 111.597 < 2e-16  
## ContinentAS 9.433e-03 1.865e-02 0.506 0.61305  
## ContinentEU 1.668e-02 1.823e-02 0.915 0.36031  
## ContinentN.America -2.858e-03 1.794e-02 -0.159 0.87341  
## ContinentOC 2.562e-03 1.977e-02 0.130 0.89690  
## ContinentSA -1.196e-03 2.127e-02 -0.056 0.95516  
## Sourcegroupfacebook 7.563e-02 3.341e-02 2.264 0.02360  
## Sourcegroupgoogle 3.923e-02 4.769e-03 8.225 < 2e-16  
## SourcegroupOthers -4.697e-03 5.732e-03 -0.819 0.41253  
## Sourcegrouppublic.tableausoftware.com -5.546e-02 9.542e-03 -5.812 6.22e-09  
## Sourcegroupreddit.com 1.170e-02 1.336e-02 0.876 0.38116  
## Sourcegroupt.co -1.429e-02 7.672e-03 -1.863 0.06248  
## Sourcegrouptableausoftware.com -3.189e-02 7.532e-03 -4.234 2.30e-05  
## Sourcegroupvisualisingdata.com -5.212e-02 1.090e-02 -4.782 1.74e-06  
## Timeinpage 1.518e-05 4.739e-06 3.203 0.00136  
## Uniquepageviews 4.943e-01 5.101e-03 96.909 < 2e-16  
## Visits 1.254e-02 5.599e-03 2.239 0.02515  
## BouncesNew NA NA NA NA  
##   
## (Intercept) .   
## Bounces \*\*\*  
## ContinentAS   
## ContinentEU   
## ContinentN.America   
## ContinentOC   
## ContinentSA   
## Sourcegroupfacebook \*   
## Sourcegroupgoogle \*\*\*  
## SourcegroupOthers   
## Sourcegrouppublic.tableausoftware.com \*\*\*  
## Sourcegroupreddit.com   
## Sourcegroupt.co .   
## Sourcegrouptableausoftware.com \*\*\*  
## Sourcegroupvisualisingdata.com \*\*\*  
## Timeinpage \*\*   
## Uniquepageviews \*\*\*  
## Visits \*   
## BouncesNew   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3184 on 32091 degrees of freedom  
## Multiple R-squared: 0.7907, Adjusted R-squared: 0.7906   
## F-statistic: 7132 on 17 and 32091 DF, p-value: < 2.2e-16

anoo<-aov(Exits~.,data = internet)  
summary(anoo)

## 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

#From the results of anova & linear model we find out that Bounces,Uniquepageviews,Timeinpage  
# have large amount of significant views & Visits have least amount of significance  
#Hence we can say that exit from the site is affected by the factors of source group,  
#bounces, and unique.pageviews.  
  
#Q4 Every site wants to increase the time on page for a visitor.   
#This increases the chances of the visitor understanding the site content better and h  
#hence there are more chances of a transaction taking place.   
#Find the variables which possibly have an effect on the time on page  
anova<-aov(Timeinpage~.,data=internet)  
anova

## Call:  
## aov(formula = Timeinpage ~ ., data = internet)  
##   
## Terms:  
## Bounces Exits Continent Sourcegroup Uniquepageviews  
## Sum of Squares 59466495 130400662 4767155 1545221 179133934  
## Deg. of Freedom 1 1 5 8 1  
## Visits Residuals  
## Sum of Squares 107321113 4512849757  
## Deg. of Freedom 1 32091  
##   
## Residual standard error: 375.0022  
## 1 out of 19 effects not estimable  
## Estimated effects may be unbalanced

summary(anova)

## 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

model2<-lm(Timeinpage~.,data=internet)  
model2

##   
## Call:  
## lm(formula = Timeinpage ~ ., data = internet)  
##   
## Coefficients:  
## (Intercept) Bounces   
## -22.5316 -286.2185   
## Exits ContinentAS   
## 21.0564 2.4986   
## ContinentEU ContinentN.America   
## -0.1800 6.1639   
## ContinentOC ContinentSA   
## 11.6686 75.4780   
## Sourcegroupfacebook Sourcegroupgoogle   
## 9.8873 6.7369   
## SourcegroupOthers Sourcegrouppublic.tableausoftware.com   
## -3.1952 5.7291   
## Sourcegroupreddit.com Sourcegroupt.co   
## 0.2133 6.8905   
## Sourcegrouptableausoftware.com Sourcegroupvisualisingdata.com   
## -1.0554 -10.0959   
## Uniquepageviews Visits   
## 97.8214 180.0594   
## BouncesNew   
## NA

summary(model2)

##   
## Call:  
## lm(formula = Timeinpage ~ ., data = internet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3892 -70 -3 5 46386   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -22.5316 21.7256 -1.037 0.29970   
## Bounces -286.2185 6.1776 -46.332 < 2e-16 \*\*\*  
## Exits 21.0564 6.5733 3.203 0.00136 \*\*   
## ContinentAS 2.4986 21.9661 0.114 0.90944   
## ContinentEU -0.1800 21.4717 -0.008 0.99331   
## ContinentN.America 6.1639 21.1230 0.292 0.77043   
## ContinentOC 11.6686 23.2823 0.501 0.61625   
## ContinentSA 75.4780 25.0449 3.014 0.00258 \*\*   
## Sourcegroupfacebook 9.8873 39.3518 0.251 0.80162   
## Sourcegroupgoogle 6.7369 5.6222 1.198 0.23082   
## SourcegroupOthers -3.1952 6.7505 -0.473 0.63598   
## Sourcegrouppublic.tableausoftware.com 5.7291 11.2436 0.510 0.61037   
## Sourcegroupreddit.com 0.2133 15.7333 0.014 0.98918   
## Sourcegroupt.co 6.8905 9.0360 0.763 0.44573   
## Sourcegrouptableausoftware.com -1.0554 8.8728 -0.119 0.90531   
## Sourcegroupvisualisingdata.com -10.0959 12.8408 -0.786 0.43174   
## Uniquepageviews 97.8214 6.8083 14.368 < 2e-16 \*\*\*  
## Visits 180.0594 6.5179 27.625 < 2e-16 \*\*\*  
## BouncesNew NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 375 on 32091 degrees of freedom  
## Multiple R-squared: 0.09661, Adjusted R-squared: 0.09614   
## F-statistic: 201.9 on 17 and 32091 DF, p-value: < 2.2e-16

#We found out from the summary of anova that only source group is not affecting the time in page views   
#And rest all are significantly afecting the timein page views  
  
#Q5 A high bounce rate is a cause of alarm for websites which depend on visitor engagement.   
#Help the team in determining the factors that are impacting the bounce.  
  
  
a<-lm(Bounces~.,data=internet)  
a

##   
## Call:  
## lm(formula = Bounces ~ ., data = internet)  
##   
## Coefficients:  
## (Intercept) Exits   
## 9.224e-14 2.115e-13   
## ContinentAS ContinentEU   
## -8.382e-16 -1.431e-15   
## ContinentN.America ContinentOC   
## -4.043e-15 3.825e-14   
## ContinentSA Sourcegroupfacebook   
## 7.492e-16 1.125e-14   
## Sourcegroupgoogle SourcegroupOthers   
## -1.191e-14 -7.952e-15   
## Sourcegrouppublic.tableausoftware.com Sourcegroupreddit.com   
## -1.883e-14 -2.849e-14   
## Sourcegroupt.co Sourcegrouptableausoftware.com   
## -7.856e-15 -2.122e-14   
## Sourcegroupvisualisingdata.com Timeinpage   
## -6.915e-15 2.388e-19   
## Uniquepageviews Visits   
## 3.135e-15 -1.842e-14   
## BouncesNew   
## 1.000e+02

summary(a)

##   
## Call:  
## lm(formula = Bounces ~ ., data = internet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.635e-11 -1.000e-15 1.000e-15 3.000e-15 4.813e-11   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 9.224e-14 1.776e-14 5.194e+00 2.07e-07  
## Exits 2.115e-13 5.374e-15 3.936e+01 < 2e-16  
## ContinentAS -8.382e-16 1.796e-14 -4.700e-02 0.962769  
## ContinentEU -1.431e-15 1.755e-14 -8.200e-02 0.935017  
## ContinentN.America -4.043e-15 1.727e-14 -2.340e-01 0.814848  
## ContinentOC 3.825e-14 1.903e-14 2.010e+00 0.044453  
## ContinentSA 7.492e-16 2.048e-14 3.700e-02 0.970813  
## Sourcegroupfacebook 1.125e-14 3.217e-14 3.500e-01 0.726560  
## Sourcegroupgoogle -1.191e-14 4.596e-15 -2.591e+00 0.009580  
## SourcegroupOthers -7.952e-15 5.518e-15 -1.441e+00 0.149552  
## Sourcegrouppublic.tableausoftware.com -1.883e-14 9.191e-15 -2.048e+00 0.040534  
## Sourcegroupreddit.com -2.849e-14 1.286e-14 -2.215e+00 0.026749  
## Sourcegroupt.co -7.856e-15 7.386e-15 -1.064e+00 0.287509  
## Sourcegrouptableausoftware.com -2.122e-14 7.253e-15 -2.926e+00 0.003437  
## Sourcegroupvisualisingdata.com -6.915e-15 1.050e-14 -6.590e-01 0.510057  
## Timeinpage 2.388e-19 4.563e-18 5.200e-02 0.958261  
## Uniquepageviews 3.135e-15 5.583e-15 5.620e-01 0.574394  
## Visits -1.842e-14 5.391e-15 -3.417e+00 0.000633  
## BouncesNew 1.000e+02 5.216e-13 1.917e+14 < 2e-16  
##   
## (Intercept) \*\*\*  
## Exits \*\*\*  
## ContinentAS   
## ContinentEU   
## ContinentN.America   
## ContinentOC \*   
## ContinentSA   
## Sourcegroupfacebook   
## Sourcegroupgoogle \*\*   
## SourcegroupOthers   
## Sourcegrouppublic.tableausoftware.com \*   
## Sourcegroupreddit.com \*   
## Sourcegroupt.co   
## Sourcegrouptableausoftware.com \*\*   
## Sourcegroupvisualisingdata.com   
## Timeinpage   
## Uniquepageviews   
## Visits \*\*\*  
## BouncesNew \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.065e-13 on 32090 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 9.521e+27 on 18 and 32090 DF, p-value: < 2.2e-16

internet$Bounces=internet$Bounces\*0.01  
rmm<-glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits,data = internet,family = "binomial")

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

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

rmm

##   
## Call: glm(formula = Bounces ~ Timeinpage + Continent + Exits + Sourcegroup +   
## Uniquepageviews + Visits, family = "binomial", data = internet)  
##   
## Coefficients:  
## (Intercept) Timeinpage   
## -4.966768 -0.001029   
## ContinentAS ContinentEU   
## 0.002277 -0.006924   
## ContinentN.America ContinentOC   
## 0.010133 0.020112   
## ContinentSA Exits   
## 0.023751 1.390761   
## Sourcegroupfacebook Sourcegroupgoogle   
## -0.024195 -0.078363   
## SourcegroupOthers Sourcegrouppublic.tableausoftware.com   
## -0.076792 -0.252828   
## Sourcegroupreddit.com Sourcegroupt.co   
## -0.009279 0.014869   
## Sourcegrouptableausoftware.com Sourcegroupvisualisingdata.com   
## -0.112930 -0.082252   
## Uniquepageviews Visits   
## -3.236311 2.194112   
##   
## Degrees of Freedom: 32108 Total (i.e. Null); 32091 Residual  
## Null Deviance: 234.9   
## Residual Deviance: 96.51 AIC: 506.6

summary(rmm)

##   
## Call:  
## glm(formula = Bounces ~ Timeinpage + Continent + Exits + Sourcegroup +   
## Uniquepageviews + Visits, family = "binomial", data = internet)  
##   
## 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  
## 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  
## Exits 1.3907608 0.3356504 4.143 3.42e-05  
## 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 .   
## ContinentAS   
## ContinentEU   
## ContinentN.America   
## ContinentOC   
## ContinentSA   
## Exits \*\*\*  
## 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

library(readxl) internet <- read\_excel(“C:/Users/HP/Downloads/internet.xlsx”) View(internet) #We will import the dataset And we will find the summary & str of the dataset head(internet) tail(internet) summary(internet) str(internet) library(psych) describe(internet)

# To find if there is missing values in the dataset

sum(is.na(internet)) #There is no missing values present in the dataset

#Q2 As mentioned earlier, a unique page view represents the number of sessions during which that page was viewed one or more times. #A visit counts all instances, no matter how many times the same visitor may have been to your site. #So the team needs to know whether the unique page view value depends on visits.

cor(internetUniquepageviews) cor(internetUniquepageviews) #We find out the regression model between Visits & Uniquepageviews model=lm(Visits~Uniquepageviews,data=internet) summary(model)

ano<-aov(Uniquepageviews~Visits, data=internet) ano summary(ano)

#Q3 Find out the probable factors from the dataset, which could affect the exits. #Exit Page Analysis is usually required to get an idea about why a user leaves the website for a session and moves on to another one. #Please keep in mind that exits should not be confused with bounces

model1<-lm(Exits~.,data=internet) model1 summary(model1)

anoo<-aov(Exits~.,data = internet) summary(anoo)

#From the results of anova we find out that Bounces,Uniquepageviews,Timeinpage # have large amount of significant views & Visits have least amount of significance #Hence we can say that exit from the site is affected by the factors of source group, #bounces, and unique.pageviews.

#Q4 Every site wants to increase the time on page for a visitor. #This increases the chances of the visitor understanding the site content better and h #hence there are more chances of a transaction taking place. #Find the variables which possibly have an effect on the time on page anova<-aov(Timeinpage~.,data=internet) anova summary(anova)

model2<-lm(Timeinpage~.,data=internet) model2 summary(model2)

#We found out from the summary of anova that only source group is not affecting the time in page views #And rest all are significantly afecting the timein page views

#Q5 A high bounce rate is a cause of alarm for websites which depend on visitor engagement. #Help the team in determining the factors that are impacting the bounce.

a<-lm(Bounces~.,data=internet) a summary(a)

internetBounces\*0.01 rmm<-glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits,data = internet,family = “binomial”) rmm summary(rmm)

## 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>.

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:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.