**Web Data Analysis**

**Question 1.** 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.

**Solution 1.**

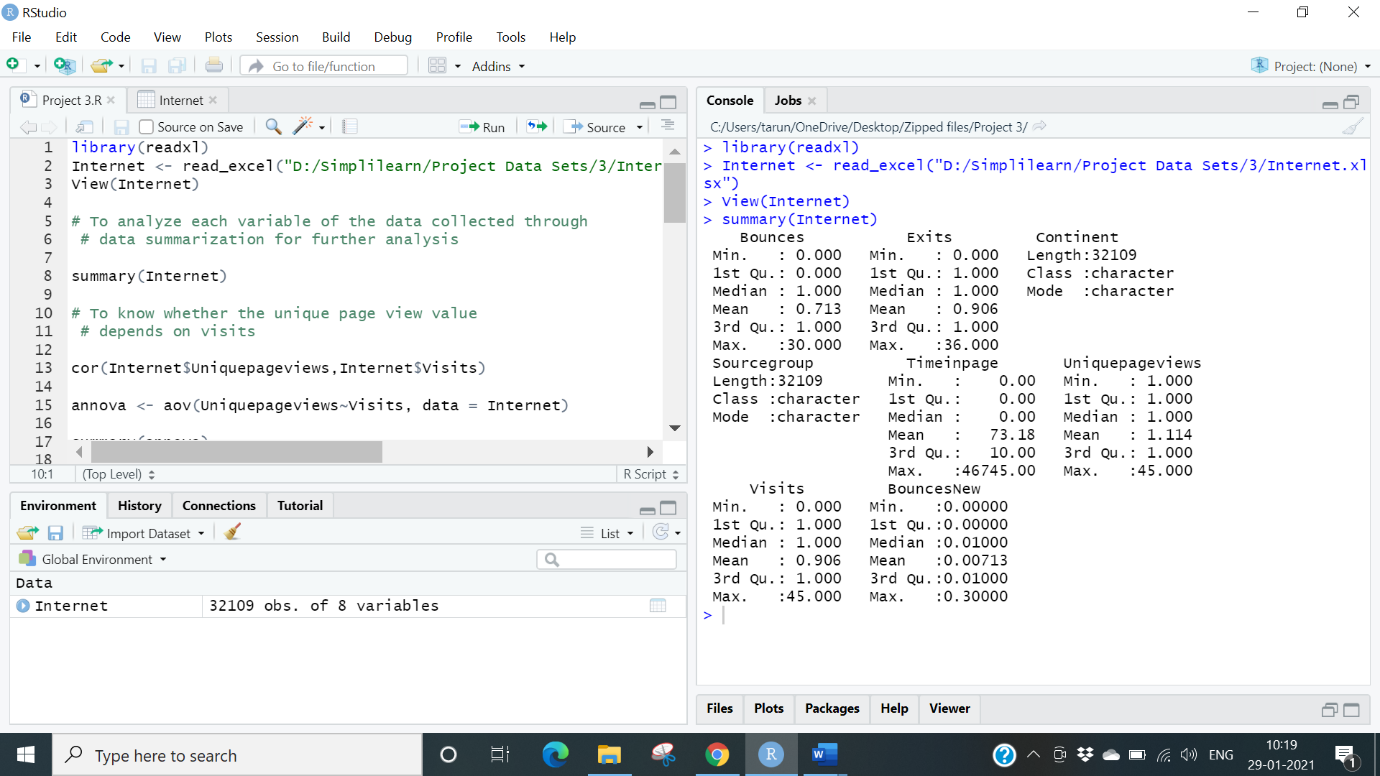
**Code.**

library(readxl)

Internet <- read\_excel ("D:/Simplilearn/Project Data Sets/3/Internet.xlsx")

View (Internet)

summary (Internet)



**Conclusion:**  From the above data we can get the Median, Mean and the several terms which were in the dataset. There are max of 30 bounces that means number of visitors were high.

**Question 2.** 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.

**Solution 2.**

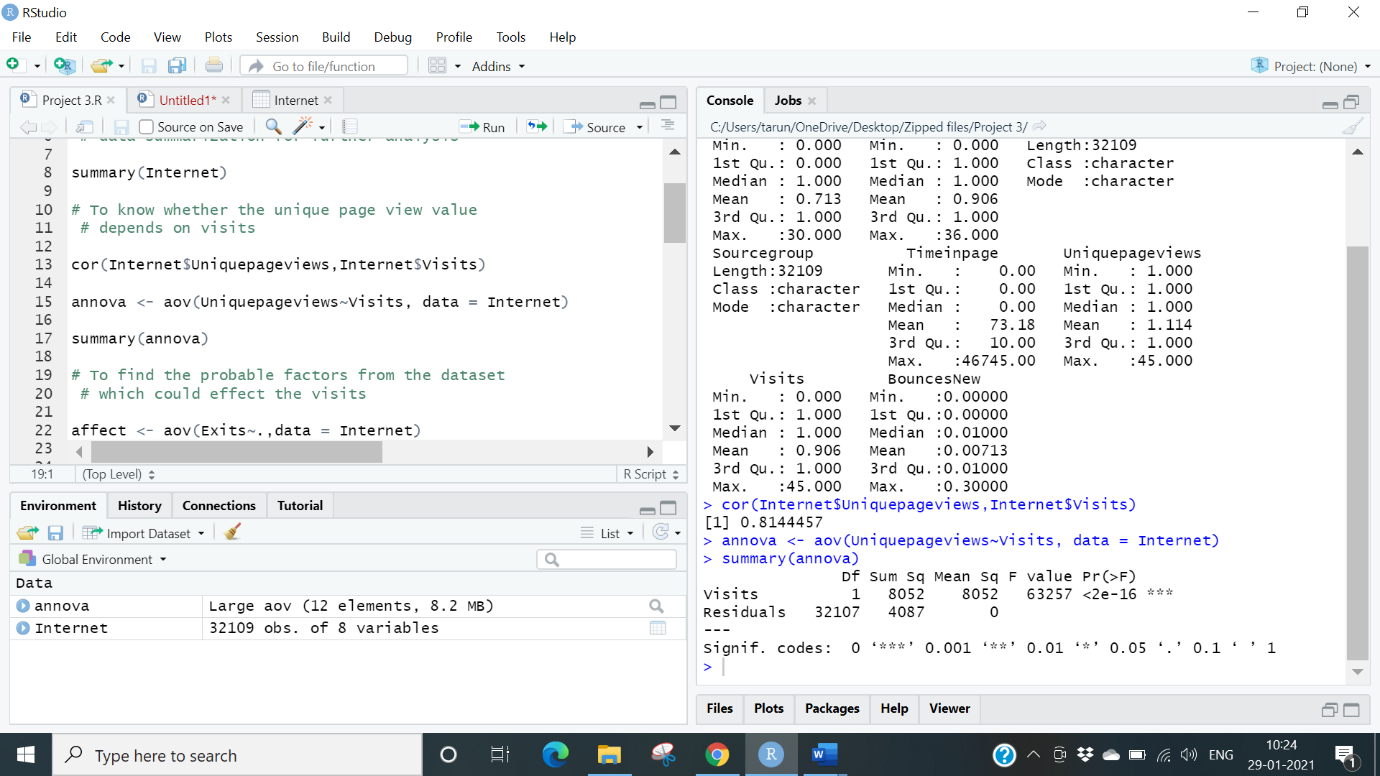
**Code:**

Cor – cor computes the variance of ‘x’ and the covariance or correlation of ‘x’ and ‘y’

cor (Internet$Uniquepageviews, Internet$Visits)

annova <- aov (Uniquepageviews~Visits, data = Internet)

summary(annova)



**Conclusion:** Visits variable has an impact of Unique pageviews. So, we can say the unique page values depends on visits.

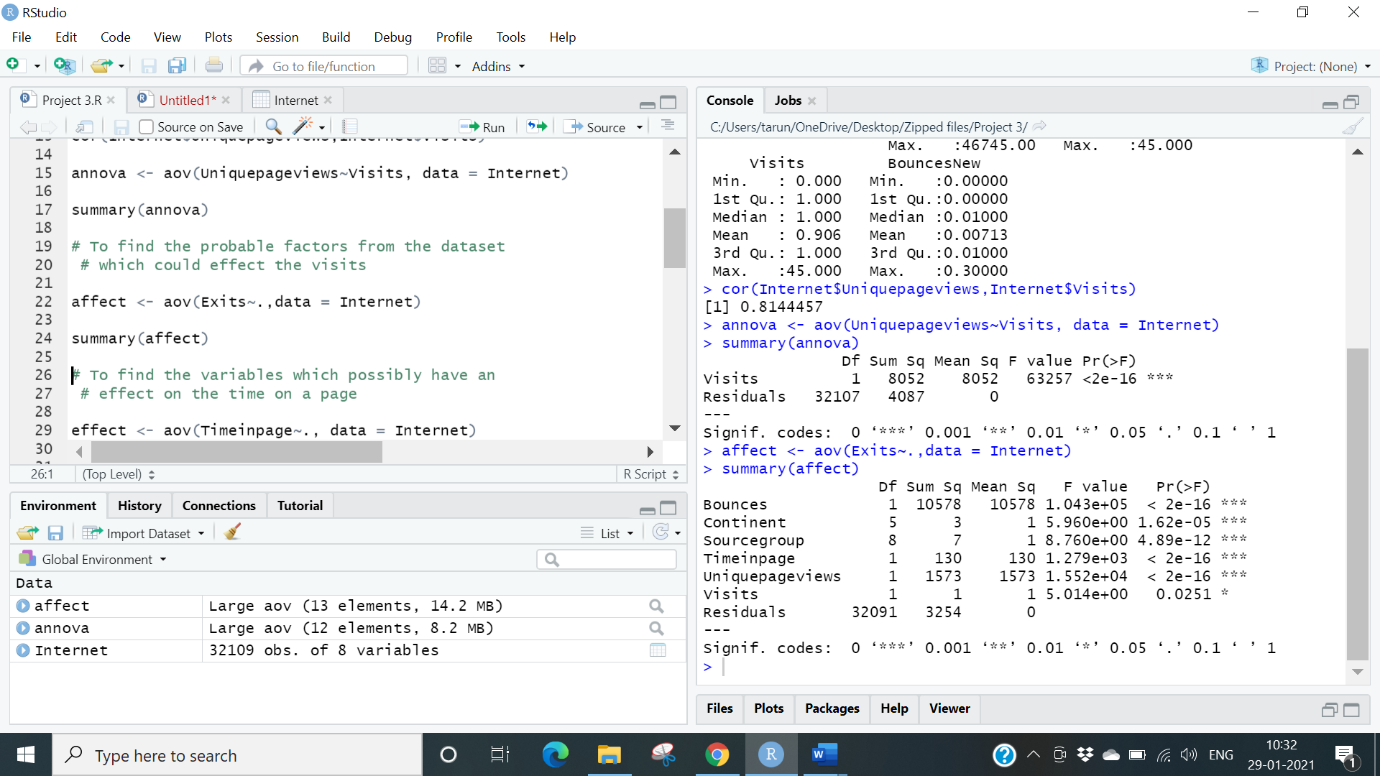
**Question 3.** Find out the probable factors from the dataset, which could affect the exits. Exit Page Analysis is usually required to get an idea about why a user leaves the website for a session and moves on to another one. Please keep in mind that exits should not be confused with bounces.

**Solution 3.**

**Code:**

affect <- aov (Exits~., data = Internet)

summary(affect)



**Conclusion:** By using Annova method we can see that visits are less in terms than Source. Group, Bounces & unique. Pageviews. So, we can say that exit from the site is affected by the factors of Source. Group, Bounces & Unique. Pageviews.

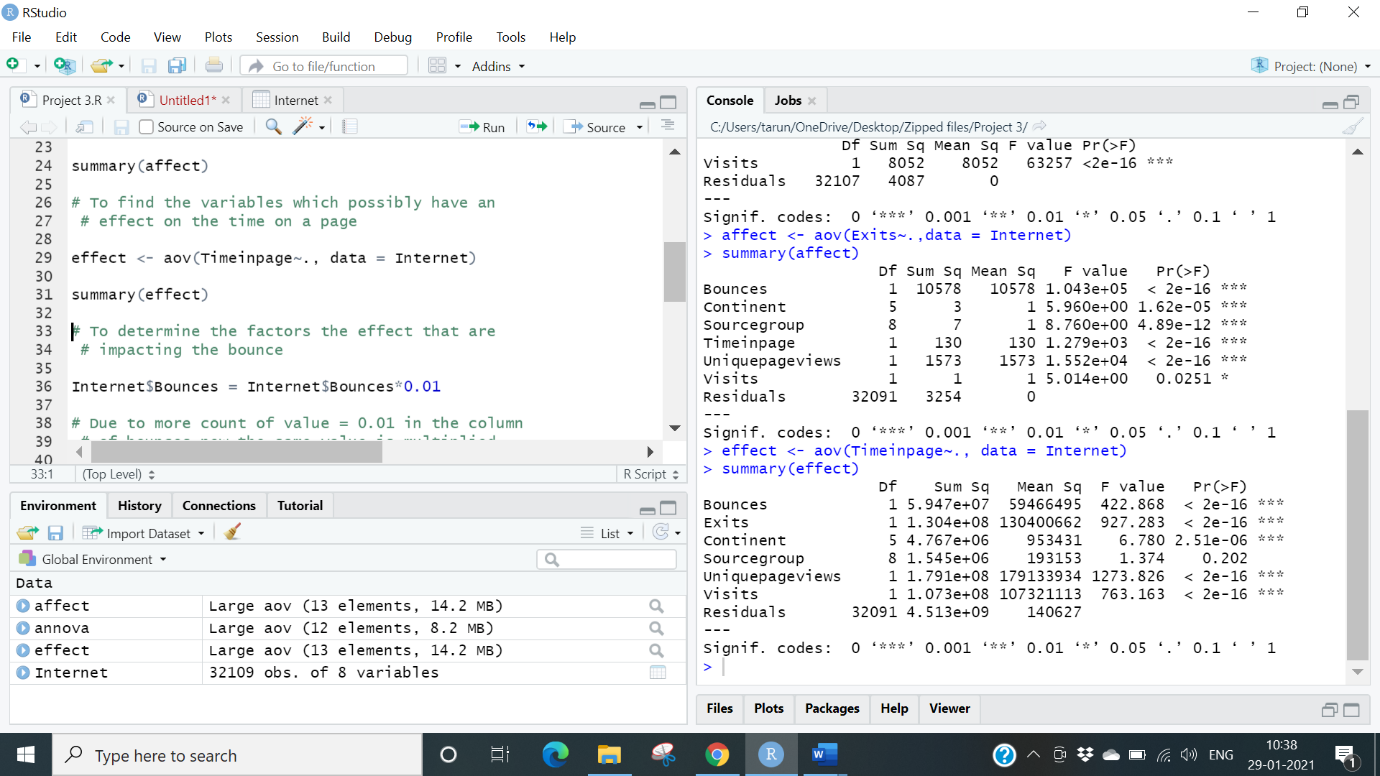
**Question 4.** Every site wants to increase the time on page for a visitor. This increases the chances of the visitor understanding the site content better and hence there are more chances of a transaction taking place. Find the variables which possibly have an effect on the time on page.

**Solution 4.**

**Code:**

effect <- aov (Timeinpage~., data = Internet)

summary(effect)



**Conclusion:** All the variables have an effect on the time on page.

**Question 5.** 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.

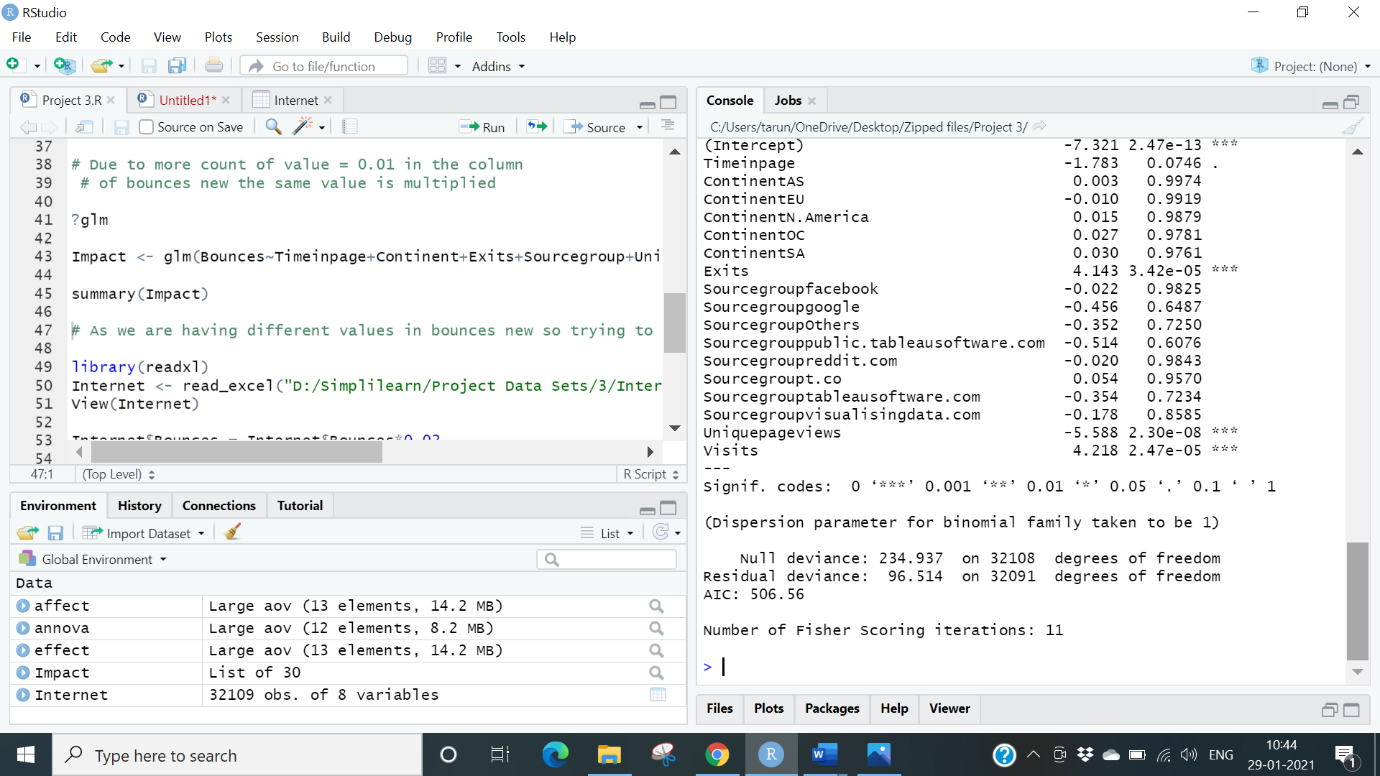
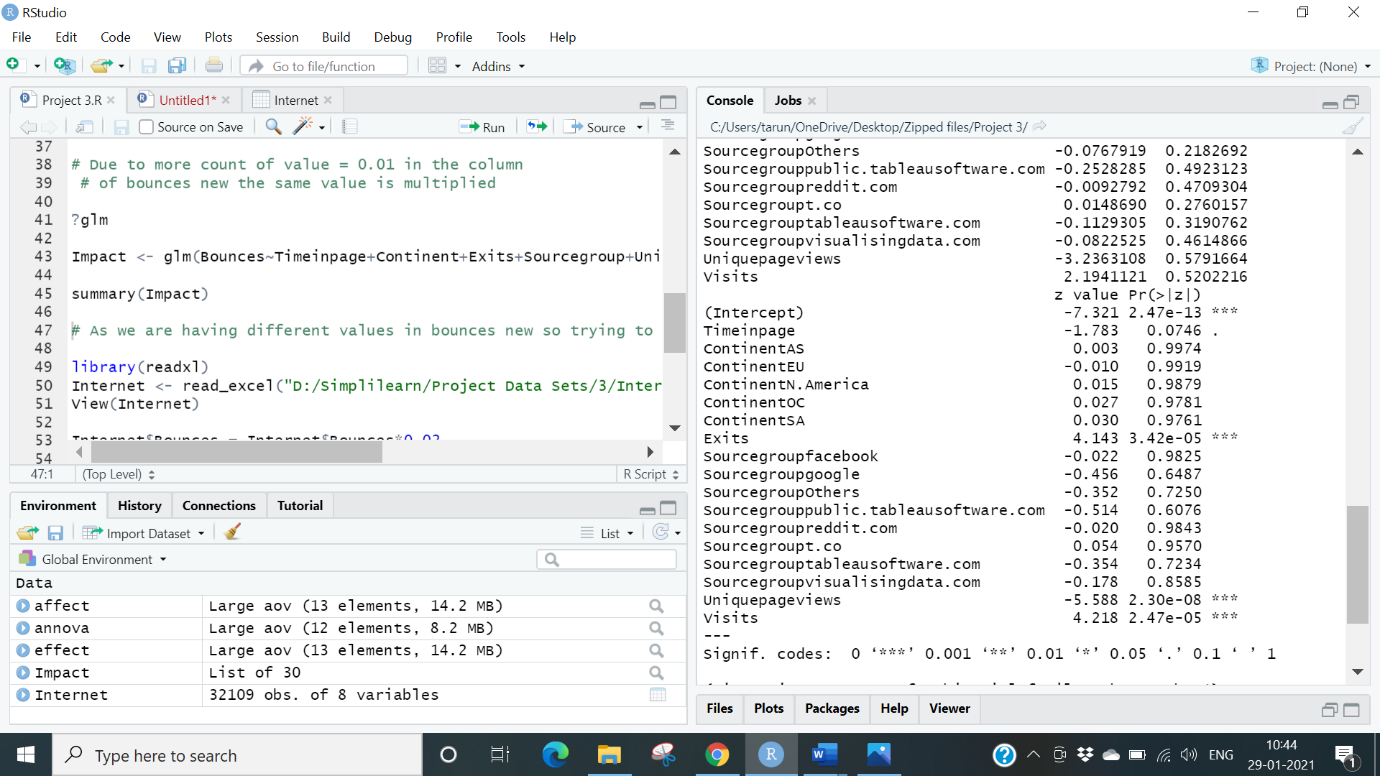
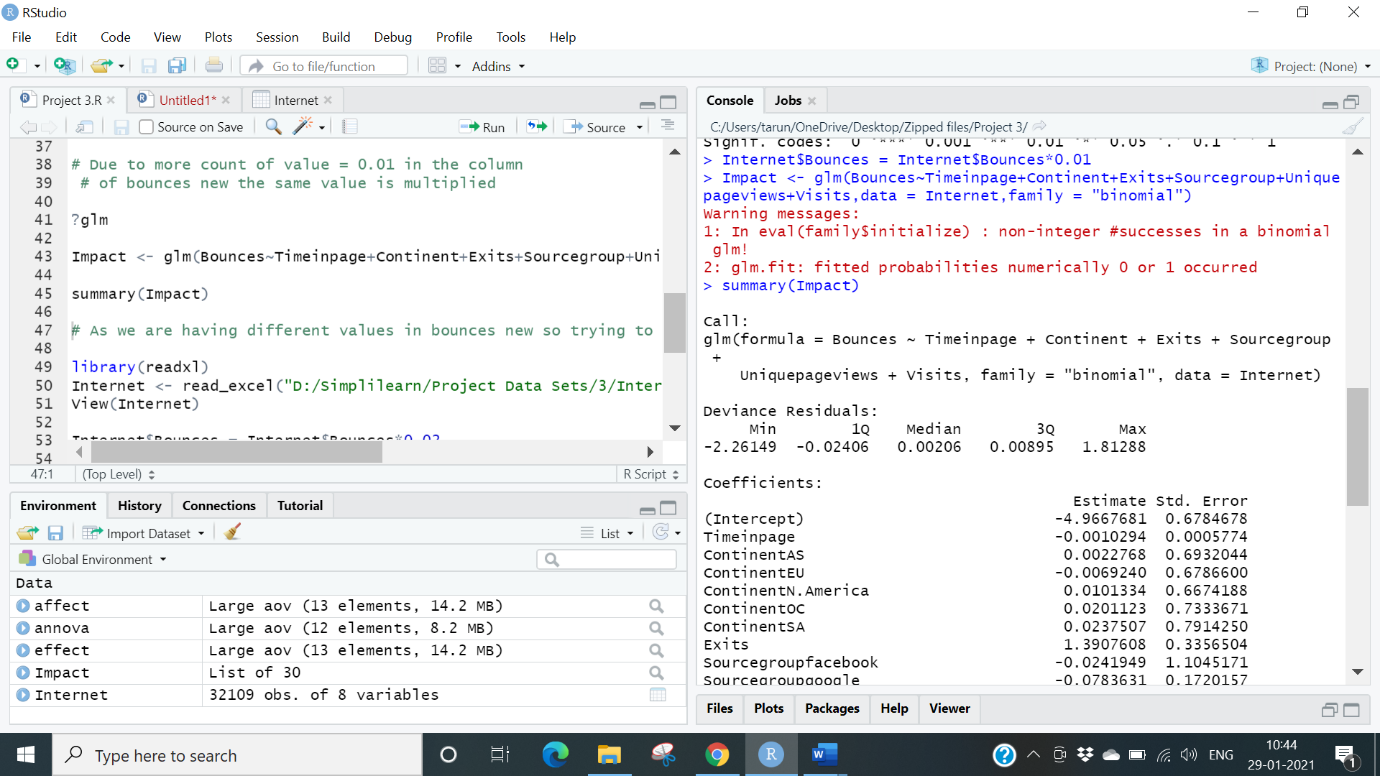
**Solution 5.**

**Code:**

Internet$Bounces = Internet$Bounces\*0.01I

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

summary (Impact)



# As we are having different values in bounces new so trying to understand the values

library(readxl)

Internet <- read\_excel ("D:/Simplilearn/Project Data Sets/3/Internet.xlsx")

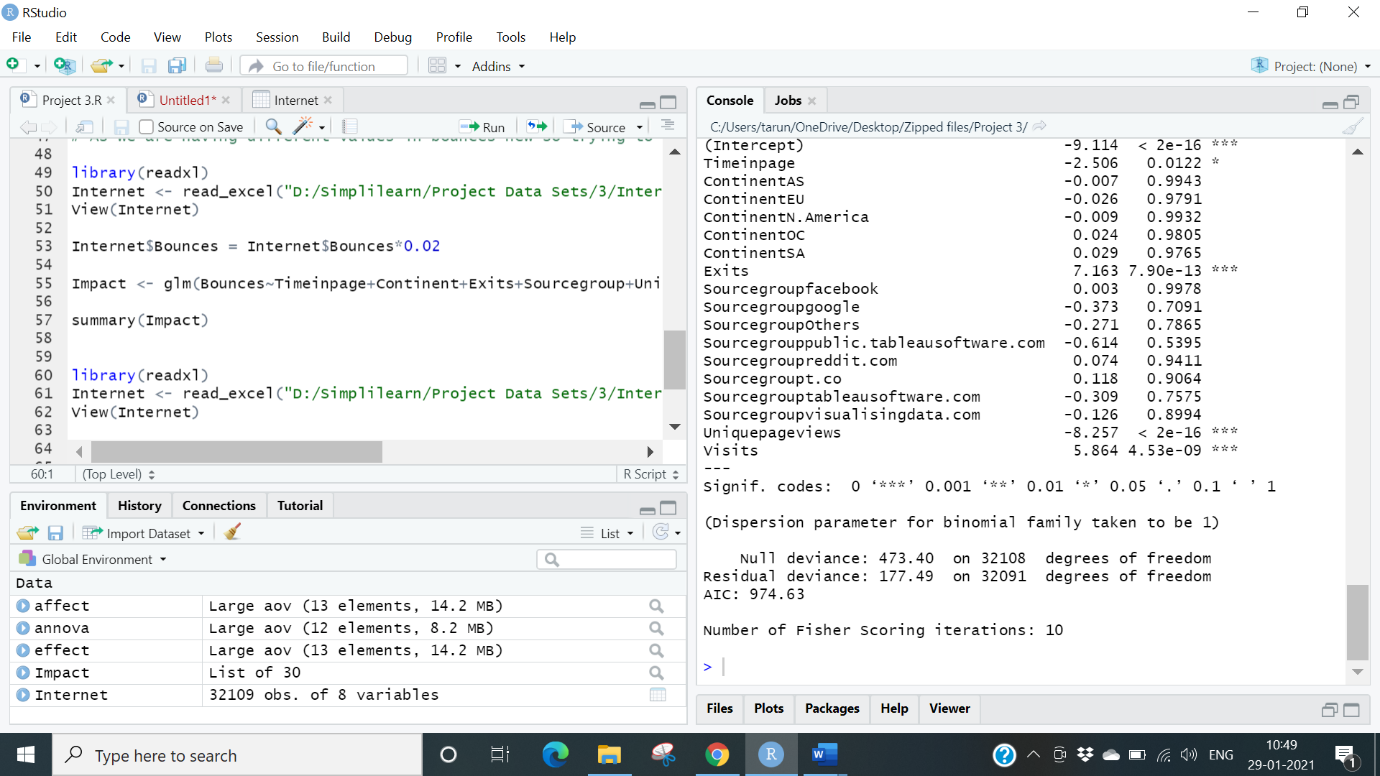
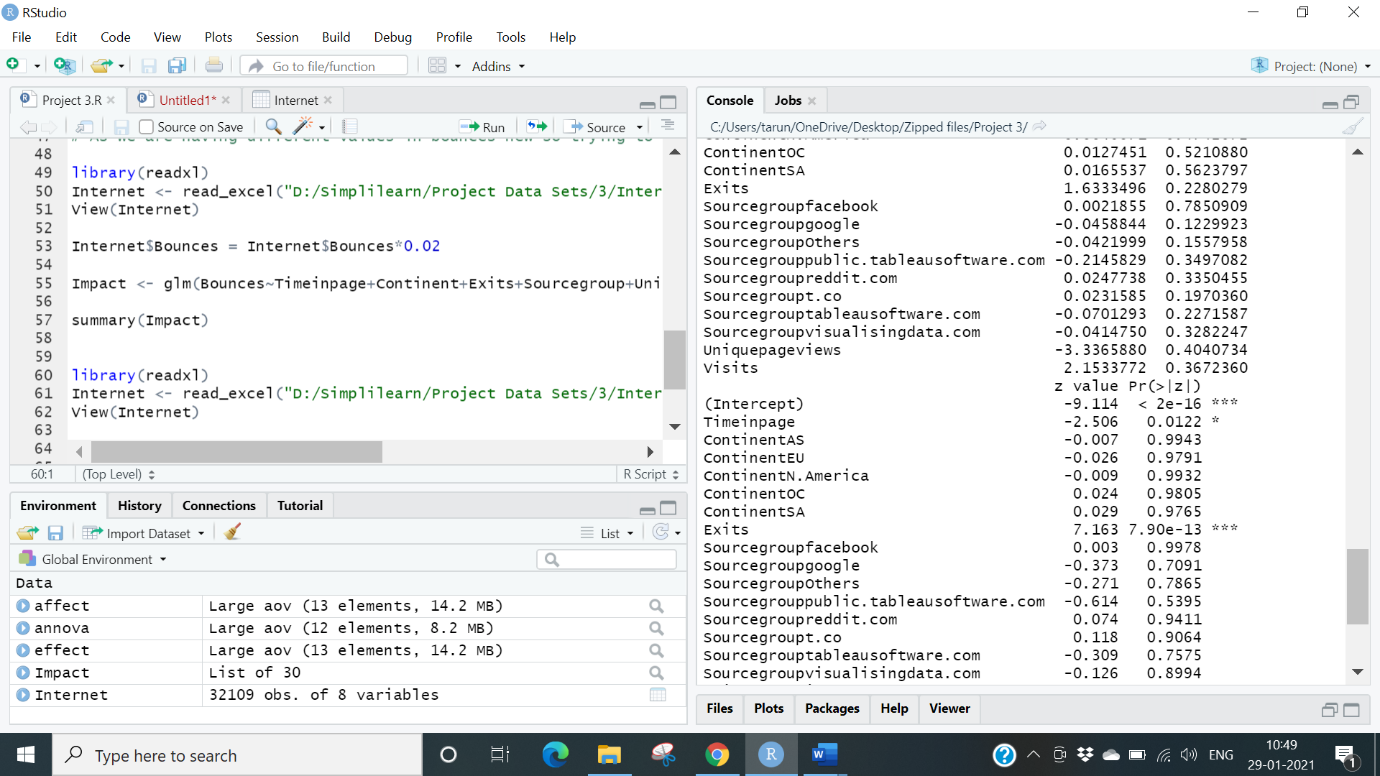
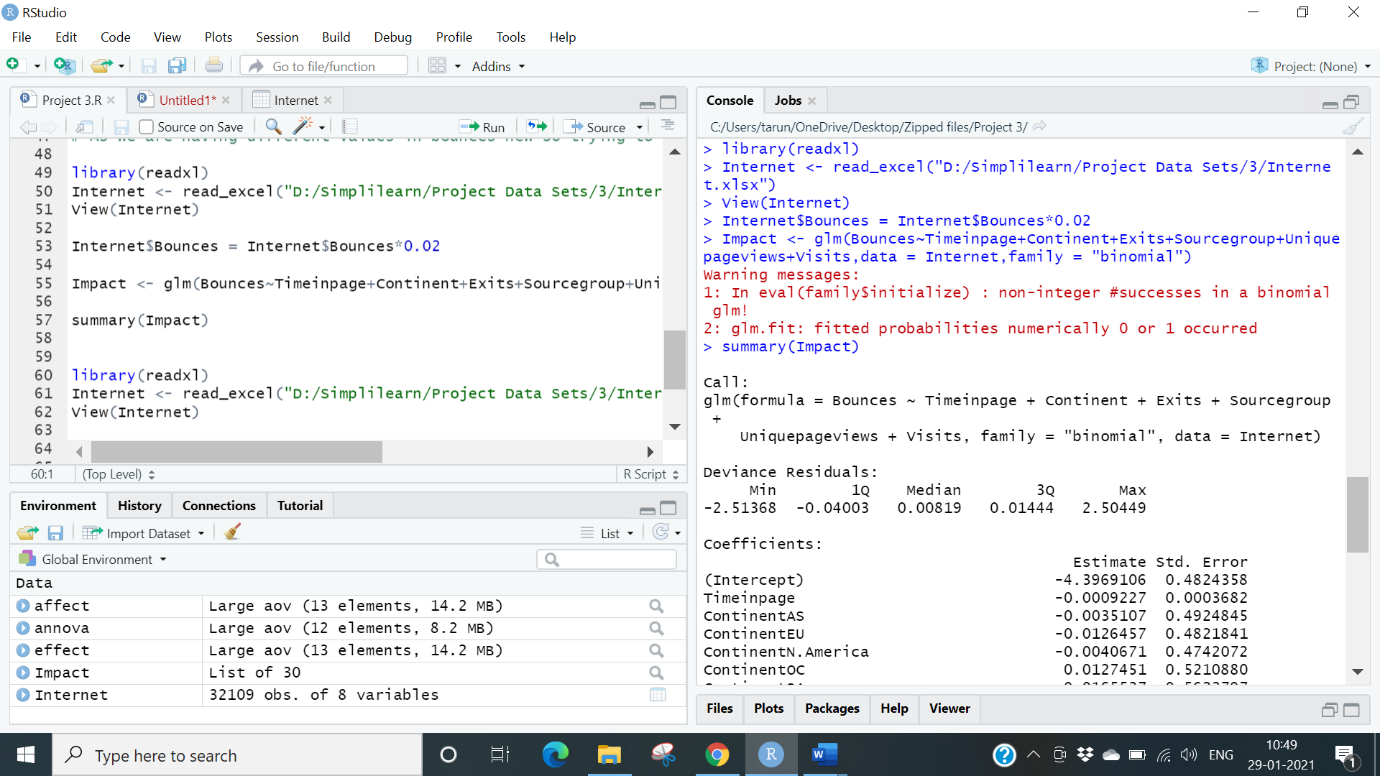
View (Internet)

Internet$Bounces = Internet$Bounces\*0.02

Impact <-

glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits, data = Internet, family = "binomial")

summary (Impact)



library(readxl)

Internet <- read\_excel ("D:/Simplilearn/Project Data Sets/3/Internet.xlsx")

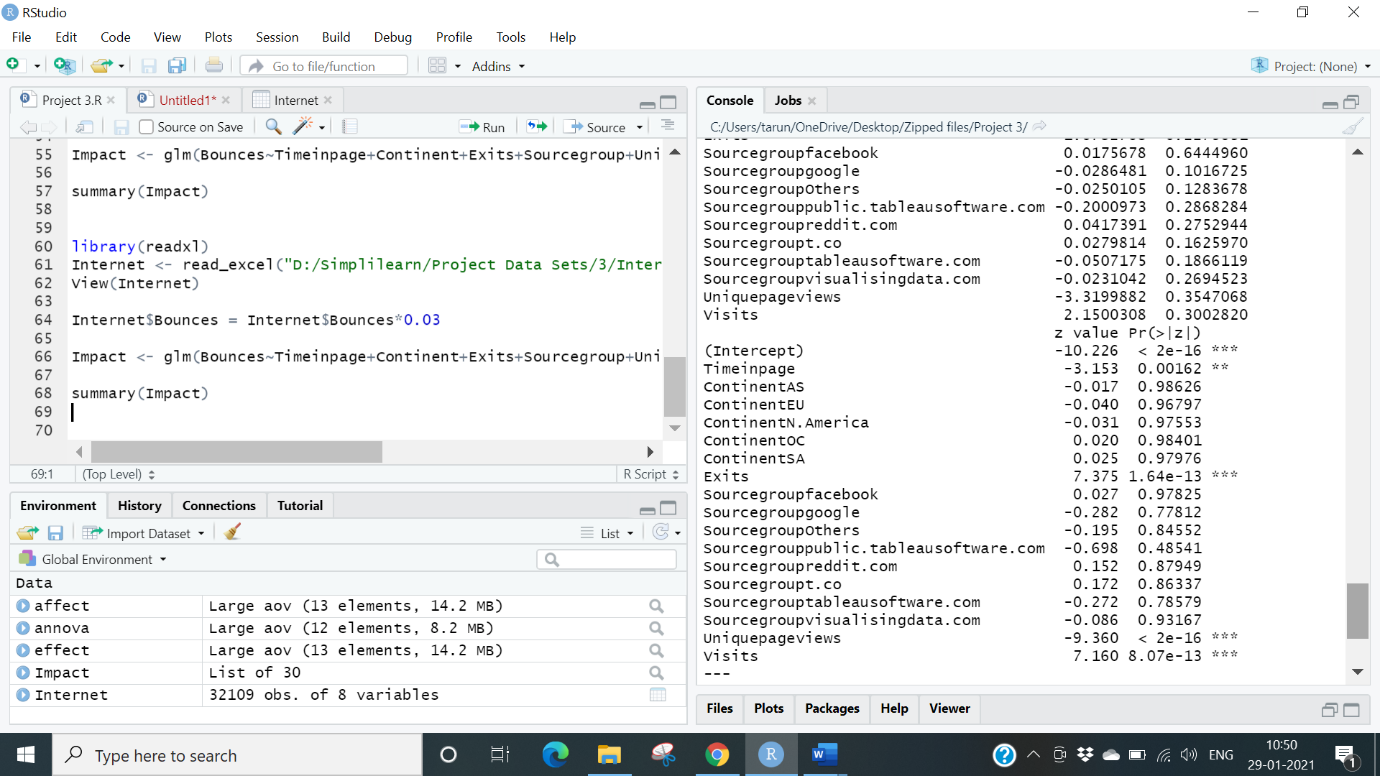
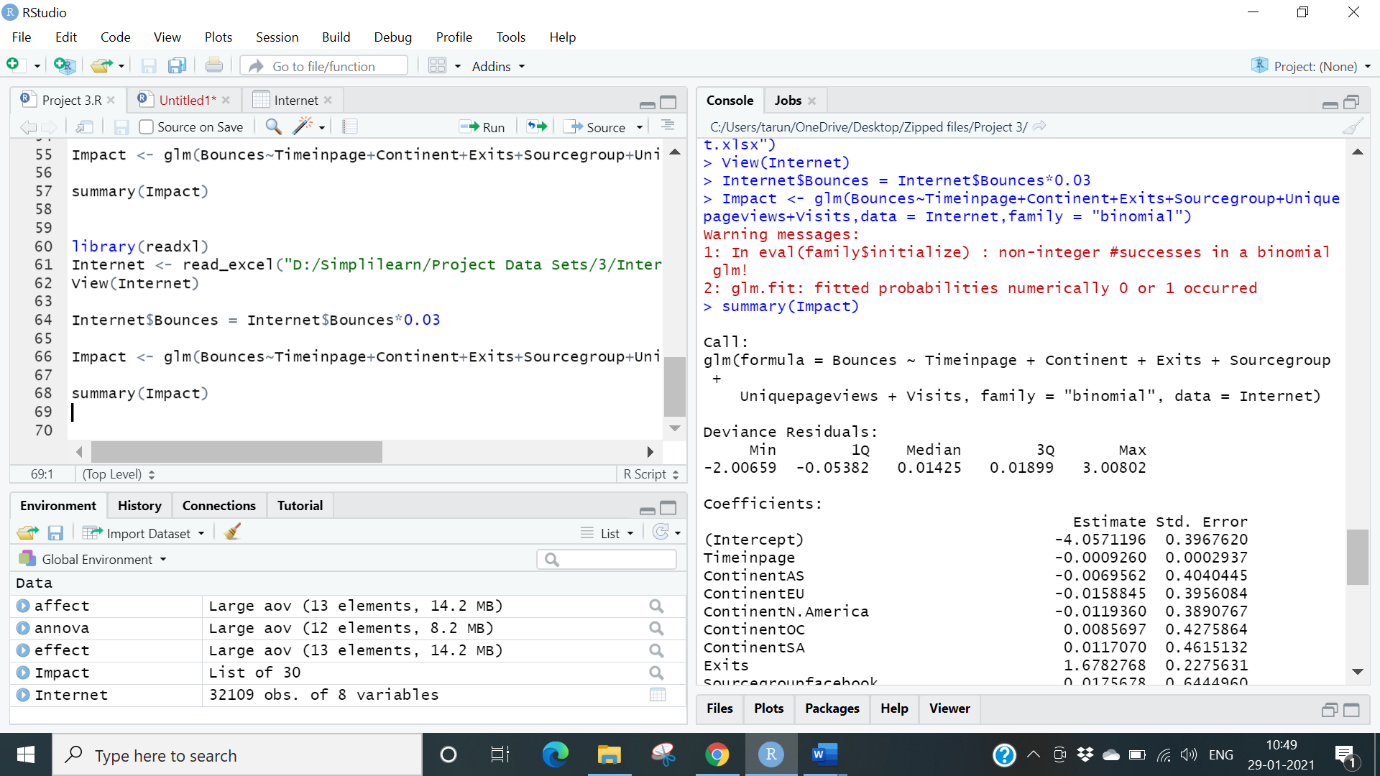
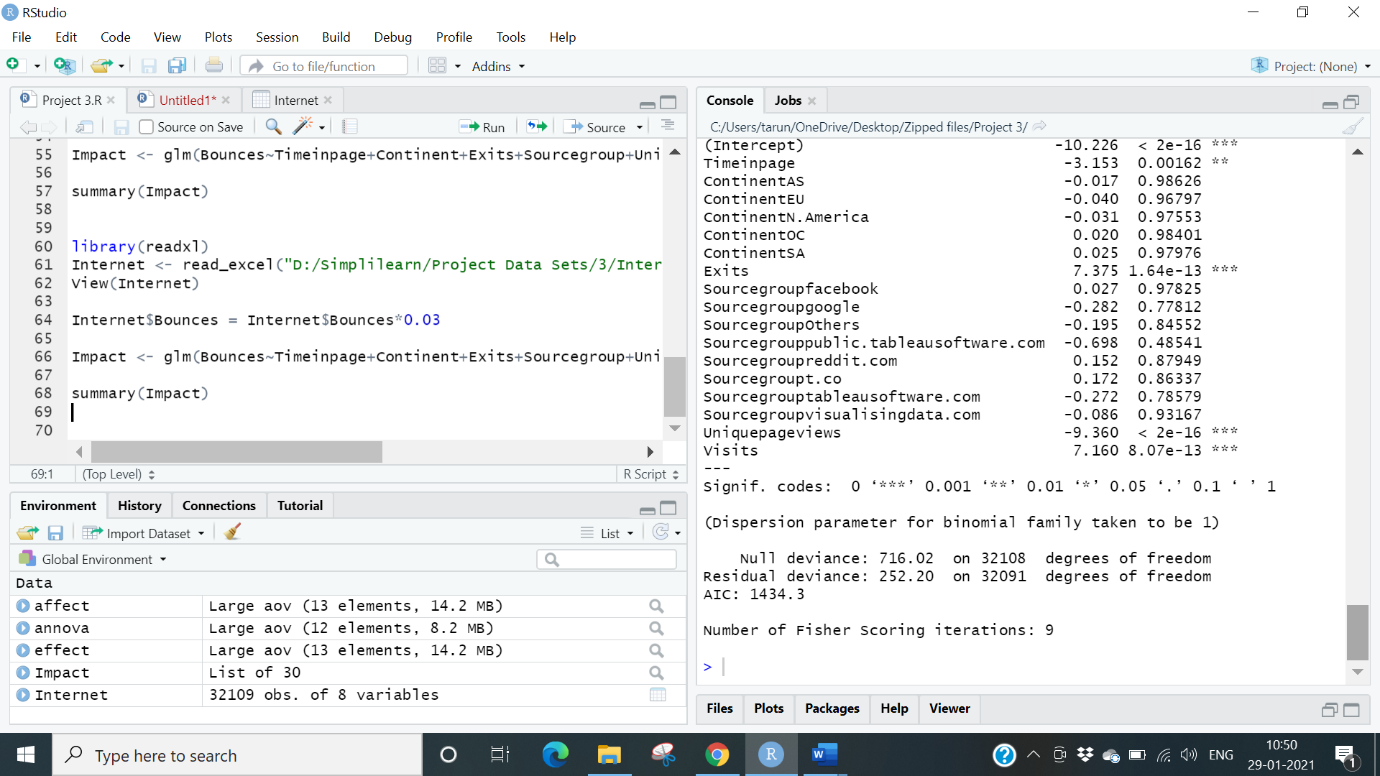
View (Internet)

Internet$Bounces = Internet$Bounces\*0.03

Impact<-

glm(Bounces~Timeinpage+Continent+Exits+Sourcegroup+Uniquepageviews+Visits, data = Internet, family = "binomial")

summary (Impact)



**Conclusion:** The Bounces New, Unique. Pageviews & Visits are the variables that impact the target variable bounces.