My Code:{Solution}

#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.

> Internet <- read\_excel("Downloads/1555058318\_internet\_dataset.xlsx")

> View(Internet)

> summary(Internet)

Bounces Exits Continent Sourcegroup Timeinpage

Min. : 0.000 Min. : 0.000 Length:32109 Length:32109 Min. : 0.00

1st Qu.: 0.000 1st Qu.: 1.000 Class :character Class :character 1st Qu.: 0.00

Median : 1.000 Median : 1.000 Mode :character Mode :character Median : 0.00

Mean : 0.713 Mean : 0.906 Mean : 73.18

3rd Qu.: 1.000 3rd Qu.: 1.000 3rd Qu.: 10.00

Max. :30.000 Max. :36.000 Max. :46745.00

Uniquepageviews Visits BouncesNew

Min. : 1.000 Min. : 0.000 Min. :0.00000

1st Qu.: 1.000 1st Qu.: 1.000 1st Qu.:0.00000

Median : 1.000 Median : 1.000 Median :0.01000

Mean : 1.114 Mean : 0.906 Mean :0.00713

3rd Qu.: 1.000 3rd Qu.: 1.000 3rd Qu.:0.01000

Max. :45.000 Max. :45.000 Max. :0.30000

> cor(Internet$Uniquepageviews, Internet$Visits)

[1] 0.8144457

> UPV\_Value <- aov(Uniquepageviews~Visits, data = Internet)

> summary(UPV\_Value)

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

> Exit\_M <- aov(Exits~.,data = Internet)

> summary(Exit\_M)

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

> #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.

> Time\_on\_Page\_M <- aov(Timeinpage~., data = Internet)

> summary(Time\_on\_Page\_M)

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

> Bounce\_M <- lm(Bounces~., data = Internet)

> summary(Bounce\_M)

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 \*\*\*

---

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