

# *WEB ANALYTICS ASSIGNMENT 2*

## **Group B**

Mariamawit Anteneh

Drithi Iyer

Yifan Ruan

1.

- a. Summary table showing the top 10 *domain names* that generated the most volume of transactions.

*Domain name* is the website name where the transaction for booking hotels online was made. The table shows that marriott.com generated the most volume with 524 transactions, while orbitz.com ranks at 10 with 109 transactions.

Rank	Domain Names	# of Transactions
1	marriott.com	524
2	hilton.com	438
3	hotels.com	400
4	expedia.com	380
5	priceline.com	309
6	choicehotels.com	297
7	jetblue.com	229
8	hotwire.com	217
9	bestwestern.com	120
10	orbitz.com	109

- b. Summary table showing the top 10 *reference domain names* that generated the most volume of transactions.

*Reference domain names* are the referring website name through which the final purchase website was reached. The table shows that google.com generated the most volume with 620 transactions, which mywebsearch.com ranks at 10 with 17 transactions.

Rank	Domain Names	# of Transactions
1	google.com	620
2	yahoo.com	222
3	bing.com	129
4	aol.com	53
5	comfortinn.com	48
6	jetblue.com	43
7	qualityinn.com	29
8	comfordsuites.com	22
9	kayak.com	20
10	mywebsearch.com	17

### c. Summary Statistics

Variables Summary Statistics	DIRECTP_D	REF_D	DURATION	PAGES_VIEWED	LOG_PRICE	TRANS_FREQ
Mean	0.481461723	0.449986663	25.31750846	18.33128834	1.826136318	2.981328354
Standard Deviation	0.499722866	0.497558741	22.80649727	16.63652277	0.979831222	4.120927309
Minimum	0	0	0.029998779	1	-1	1
Maximum	1	1	227.34375	141	3.454393568	30
Count	3749	3749	3749	3749	3749	3749

- The average of all variables is shown as the mean in the table above.
- The standard deviation shows how the numbers are spread out from the average (mean) or expected value.
- The minimum shows the least value from all variables in the set of observations.
- Maximum shows the highest value from all variables in the set of observations.
- Count indicates the number of observations in the dataset provided. In this case, there was a total of 3749 observations.

### 2. Binary Outcome (Logit) Regression

We performed the Binary outcome regression analysis on our data set. The coefficient for each variable is understood as the change in log odds of the dependent variable (DIRECTP\_D). Upon one unit increase in the independent variables. For instance, based on our results, when there is an increase of 1 minute (DURATION), we expect the log odds of DIRECTP\_D which is a dummy variable indicating whether the transaction is incurred directly from a hotel website (1) or other third party travel websites (0), to decrease by 0.019. We can also take the exponent of the coefficient to get a odds-ratio interpretation:  $\exp(-0.019) = 0.98$ . This means a shopper who has 1 more minute to shop than another user will have a 100-98= 2% lesser odds.

Interpreting p value: If  $p > 0.05$ , then the correlation is not statistically significant. If  $p < 0.05$ , it is statistically significant.

```
attach(HotelClickStream)
glm.model1 <-
glm(DIRECTP_D~REF_D+LOG_PRICE+TRANS_FREQ+DURATION+HOUSEHOLD_SIZE+CHILDREN_D+CONNECTIONSPEED_D,family = binomial(logit),data = HotelClickStream)
summary(glm.model1)

##
## Call:
## glm(formula = DIRECTP_D ~ REF_D + LOG_PRICE + TRANS_FREQ + DURATION +
##      HOUSEHOLD_SIZE + CHILDREN_D + CONNECTIONSPEED_D, family =
```

```

binomial(logit),
##      data = HotelClickStream)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.2173  -1.0791  -0.6308   1.1150   2.2067
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.486987   0.400982  -1.214 0.224562
## REF_D          0.735643   0.069666  10.560 < 2e-16 ***
## LOG_PRICE      0.024439   0.035518   0.688 0.491403
## TRANS_FREQ     0.115516   0.011628   9.934 < 2e-16 ***
## DURATION      -0.018726   0.001706 -10.976 < 2e-16 ***
## HOUSEHOLD_SIZE -0.011058   0.008454  -1.308 0.190875
## CHILDREN_D     0.255806   0.072464   3.530 0.000415 ***
## CONNECTIONSPEED_D 0.050238   0.391118   0.128 0.897795
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5192.1  on 3748  degrees of freedom
## Residual deviance: 4825.5  on 3741  degrees of freedom
## AIC: 4841.5
##
## Number of Fisher Scoring iterations: 4

#odds ratio
exp(glm.model1$coefficients)

##      (Intercept)          REF_D          LOG_PRICE          TRANS_FREQ
##      0.6144750          2.0868240          1.0247405          1.1224524
##      DURATION    HOUSEHOLD_SIZE    CHILDREN_D CONNECTIONSPEED_D
##      0.9814484          0.9890033          1.2915018          1.0515216

```

### 3. Count Data Models

- a. *Poisson Regression Analysis:* we performed Count data(Poisson) analysis for our Hotel click stream data applying the TRANS\_FREQ as our dependent variable and a series of independent variable. This analysis show the factors that are involved in determining frequencies of booking the hotel. From this we can see that most of the variables are statistically based on each p-value. This means that these variables count towards peoples' influence in making booking transactions. The coefficients show that a unit increase in the variables will either increase or decrease frequencies depending on whether the coefficient is positive or negative. For instance REF\_D is very significant based on its value and in regard to the coefficient, if it increased by one unit, we will see that log(frequency) will decrease by 0.23 because of the negative value. The actual numbers of the variable

coefficients are also shown using exponential function. For REF\_D its 1.26, which means if REF\_D is 1, it will decrease transaction frequency by 1.26.

```
glm.model2 <-
glm(TRANS_FREQ~REF_D+LOG_PRICE+PAGES_VIEWED+HOUSEHOLD_SIZE+CHILDREN_D+CONNECTIONSPEED_D,family = poisson(log),data = HotelClickStream)
summary(glm.model2)

##
## Call:
## glm(formula = TRANS_FREQ ~ REF_D + LOG_PRICE + PAGES_VIEWED +
##      HOUSEHOLD_SIZE + CHILDREN_D + CONNECTIONSPEED_D, family =
##      poisson(log),
##      data = HotelClickStream)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2678  -1.3049  -0.9396   0.1156   9.9369
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.3685323   0.1666643    2.211  0.02702 *
## REF_D          -0.2280717   0.0193607  -11.780 < 2e-16 ***
## LOG_PRICE       0.0263280   0.0099196    2.654  0.00795 **
## PAGES_VIEWED    0.0024343   0.0005517    4.412 1.02e-05 ***
## HOUSEHOLD_SIZE -0.0117769   0.0036903   -3.191  0.00142 **
## CHILDREN_D      -0.2330816   0.0198819  -11.723 < 2e-16 ***
## CONNECTIONSPEED_D 0.9023729   0.1648290    5.475 4.38e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 11646  on 3748  degrees of freedom
## Residual deviance: 11268  on 3742  degrees of freedom
## AIC: 21065
##
## Number of Fisher Scoring iterations: 5

#actual number
exp(abs(glm.model2$coefficients))

##      (Intercept)          REF_D          LOG_PRICE          PAGES_VIEWED
##      1.445611         1.256175         1.026678         1.002437
##      HOUSEHOLD_SIZE      CHILDREN_D CONNECTIONSPEED_D
##      1.011847         1.262485         2.465447
```

- b. *Negative Binomial Regression*: We also performed a Quasipoisson regression using similar dependent and independent variables as we did for the Poisson regression. The coefficient values are the same where if a predictor's value increases by a unit, there will be a change in the log of the TRANS\_FREQ. When we look at the p-value for this regression only two of the independent variables seem to be statistically significant, i.e. REF\_D and CHILDREN\_D. Relatively both the coefficients for these two variables are negative, which means a unit increase could lead to a decrease in the value of the respective variables. The values can also be better interpreted by looking at the exponential function results for the actual numbers of the coefficients. For example, if REF\_D is 1, it will lead to a decrease in transaction frequency of 1.26.

```
glm.model3 <-  
glm(TRANS_FREQ~REF_D+LOG_PRICE+PAGES_VIEWED+HOUSEHOLD_SIZE+CHILDREN_D+CONNECTIONSPEED_D,family = quasipoisson(log),data = HotelClickStream)  
summary(glm.model3)  
  
##  
## Call:  
## glm(formula = TRANS_FREQ ~ REF_D + LOG_PRICE + PAGES_VIEWED +  
##     HOUSEHOLD_SIZE + CHILDREN_D + CONNECTIONSPEED_D, family =  
quasipoisson(log),  
##     data = HotelClickStream)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.2678  -1.3049  -0.9396   0.1156   9.9369   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    0.368532   0.391937   0.940    0.3471      
## REF_D          -0.228072   0.045530  -5.009 5.72e-07 ***  
## LOG_PRICE       0.026328   0.023328   1.129   0.2591      
## PAGES_VIEWED    0.002434   0.001297   1.876   0.0607 .     
## HOUSEHOLD_SIZE -0.011777   0.008678  -1.357   0.1748      
## CHILDREN_D      -0.233082   0.046755  -4.985 6.47e-07 ***  
## CONNECTIONSPEED_D 0.902373   0.387621   2.328   0.0200 *     
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for quasipoisson family taken to be 5.530276)  
##  
##      Null deviance: 11646  on 3748  degrees of freedom  
## Residual deviance: 11268  on 3742  degrees of freedom  
## AIC: NA  
##  
## Number of Fisher Scoring iterations: 5
```

*#actual number*

```
exp(abs(glm.model3$coefficients))
```

##	(Intercept)	REF_D	LOG_PRICE	PAGES_VIEWED
##	1.445611	1.256175	1.026678	1.002437
##	HOUSEHOLD_SIZE	CHILDREN_D	CONNECTIONSPEED_D	
##	1.011847	1.262485	2.465447	

c. Summary of the Count Data Models

The main comparison between the two methods lays in the variance. We know that Poisson assumes the variance to be the same value of the mean and that is what is shown in the summary model for Poisson regression analysis. Whereas, in the case of the Negative Binomial regression model, the actual dispersion value is taken to be greater than the mean. Although the data type and other factors are also involved for better analysis, the Quasipoisson method may be more suitable for our Hotel Clicks data. However the standard error for the Quasipoisson regression seems to be greater in comparison. The Another comparison can be the p-values for each regression. For negative binary we can see that only two of the variables are indicated to be significant, however, for Poisson regression almost all the variables are shown as significant.

4.

a. Linear Regression (Duration as DV)

We first run the linear regression model using the entire dataset. Then we interpreted the p-values for each variable in the model and picked those significant variables for our final model as shown above. From the table above, we can see that some DOMAIN\_NAMES, PROD\_QTY, PAGES\_VIEWED, and REF\_D are all statistically significant. Specifically, if domain name where the transaction was made is aol.com, the duration will increase 36 mins; choicehotels.com will increase duration 13.4 mins; expedia.com will increase duration 16.5mins; hotels.com will increase duration 16.2 mins; hotwire.com will increase duration 15.2 mins; jetblue.com will increase duration 28.8 mins; Marriott.com will increase duration 12.3 mins; orbitz.com will increase duration 14 mins; priceline.com will increase duration 14 mins; southwest.com will increase duration 23.9 mins; starwoodhotels.com will increase duration 12.5 mins; and travelnow.com will increase duration 12.6 mins. In terms of product quantity, each 1 unit increase in PROD\_QTY, duration will increase 0.77 min. If customer visit 1 more page at the site, it will increase the duration for about 1 min. In addition, if the transaction has a referring website, it will also increase the duration for 1.4 min.

```
##
## Call:
## lm(formula = DURATION ~ DOMAIN_NAME + PROD_QTY + PAGES_VIEWED +
##     REF_D, data = newfinal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -125.003   -8.515   -3.117    5.640   135.999
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -7.96141     5.63858   -1.412  0.15805
## DOMAIN_NAMEaol.com    35.99176    16.75534    2.148  0.03177 *
## DOMAIN_NAMEbestwestern.com    6.49363     5.77387    1.125  0.26081
## DOMAIN_NAMEcheaptickets.com   12.13246     6.48798    1.870  0.06156 .
## DOMAIN_NAMEchoicehotels.com   13.43813     5.67448    2.368  0.01793 *
## DOMAIN_NAMEcountryinns.com    8.63239     6.19590    1.393  0.16363
## DOMAIN_NAMEdaysinn.com    6.98951     5.90290    1.184  0.23646
## DOMAIN_NAMEexpedia.com   16.50343     5.64562    2.923  0.00349 **
## DOMAIN_NAMEhilton.com    7.34004     5.64038    1.301  0.19322
## DOMAIN_NAMEhojo.com   -2.88530     7.33943   -0.393  0.69425
## DOMAIN_NAMEhotels.com   16.21156     5.65273    2.868  0.00416 **
## DOMAIN_NAMEhotwire.com   15.16515     5.69807    2.661  0.00781 **
## DOMAIN_NAMEhyatt.com    8.93740     6.02761    1.483  0.13823
## DOMAIN_NAMEichotelsgroup.com  9.14275     5.83143    1.568  0.11700
## DOMAIN_NAMEjetblue.com   28.82586     5.70408    5.054 4.55e-07
## ***
## DOMAIN_NAMEmarriott.com   12.27729     5.63790    2.178  0.02950 *
## DOMAIN_NAMEnetbooker.com    4.86280    10.72334    0.453  0.65023
## DOMAIN_NAMEorbitz.com   14.02910     5.79022    2.423  0.01544 *
## DOMAIN_NAMEpriceline.com  13.95315     5.66031    2.465  0.01374 *
## DOMAIN_NAMeradisson.com   10.87526     7.10326    1.531  0.12585
## DOMAIN_NAMeramada.com    7.79522     6.42609    1.213  0.22518
## DOMAIN_NAMeres99.com   10.64943     5.89695    1.806  0.07101 .
## DOMAIN_NAMeritzcarlton.com 15.25935     9.68009    1.576  0.11503
## DOMAIN_NAMESouthwest.com  23.86544     9.00720    2.650  0.00809 **
## DOMAIN_NAMEstarwoodhotels.com 12.45673     5.98223    2.082  0.03738 *
## DOMAIN_NAMEsuper8.com    7.92147     6.02569    1.315  0.18872
## DOMAIN_NAMetravelnow.com  12.59695     6.14990    2.048  0.04060 *
## DOMAIN_NAMetravelocity.com  8.23478     5.78969    1.422  0.15502
## DOMAIN_NAMetravelodge.com  3.49897     8.18410    0.428  0.66902
## DOMAIN_NAMEwingatehotels.com 1.39160     7.90565    0.176  0.86028
```



```

## DOMAIN_NAMEwwte1.com          7.55232    7.67503    0.984 0.32517
## DOMAIN_NAMEwyndham.com        3.60800    7.89868    0.457 0.64785
## DOMAIN_NAMEyahoo.net          1.47341    6.24574    0.236 0.81352
## PROD_QTY                      0.77399    0.19529    3.963 7.53e-05
***
## PAGES_VIEWED                  0.99290    0.01716   57.846 < 2e-16
***
## REF_D1                        1.39433    0.53963    2.584 0.00981 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.79 on 3713 degrees of freedom
## Multiple R-squared:  0.5249, Adjusted R-squared:  0.5204
## F-statistic: 117.2 on 35 and 3713 DF,  p-value: < 2.2e-16

```

b. Linear Regression (Pages Viewed as DV)

We did the same practice as the previous question to pick out the variables that are significant in the linear model and include those variables in the final model. Variables TRANS\_FREQ, some DOMAIN\_NAMES, DURATION, and LOG\_PRICE are all statistically significant. More specifically, if transaction frequency increase by one, the customer will view 0.16 more pages. Certain domain name has negative effect on pages viewed. On Bestwestern.com, customer will view 8.6 pages less; choicehotels.com will decrease 16.4 pages viewed; countryinns.com will decrease 10.8 pages; daysinn.com will decrease 10.3 pages viewed; expedia.com will decrease 9.3 pages viewed; Hilton.com will decrease 7.8 pages viewed; hotels.com will decrease 15 pages viewed; hotwire.com will decrease 14.4 pages viewed; ichotelsgroup.com will decrease 11.1 pages viewed; jetblue.com will decrease 26.0 pages viewed; Marriott.com will decrease 11.3 pages viewed; netbooker.com will decrease 15.3 pages viewed; orbitz.com will decrease 9.1 pages viewed; priceline.com will decrease 9 pages viewed; ramada.com will decrease 12.4 pages viewed; res99.com will decrease 13.7 pages viewed; starwoodshotels.com will decrease 11.9 pages viewed; super8.com will decrease 11.8 pages viewed; travelnow.com will decrease 18 pages viewed.

Duration and log\_price also have impact on pages viewed. 1 more minute on site will increase 0.5 pages viewed, while log\_price increase by 1 will decrease 0.6 pages viewed.

```
##
## Call:
## lm(formula = PAGES_VIEWED ~ TRANS_FREQ + DOMAIN_NAME + DURATION +
##     LOG_PRICE, data = newfinal)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -50.865  -5.562   -1.283    3.878   123.813
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    18.556096     3.878652   4.784 1.78e-06
***
## TRANS_FREQ      0.159992     0.045241   3.536 0.000411
***
## DOMAIN_NAMEaol.com    -11.402233     11.603973  -0.983 0.325861
## DOMAIN_NAMEbestwestern.com  -8.608127     4.030903  -2.136 0.032783
*
## DOMAIN_NAMEcheaptickets.com  -5.011491     4.509952  -1.111 0.266551
## DOMAIN_NAMEchoic-hotels.com -16.399421     3.954204  -4.147 3.44e-05
***
## DOMAIN_NAMEcountryinns.com  -10.786190     4.284075  -2.518 0.011853
*
## DOMAIN_NAMEdaysinn.com    -10.288729     4.119524  -2.498 0.012548
*
## DOMAIN_NAMEexpedia.com    -9.265070     3.942744  -2.350 0.018830
*
```

## DOMAIN_NAMEhilton.com	-7.840388	3.940454	-1.990	0.046695
*				
## DOMAIN_NAMEhojo.com	5.266142	5.115202	1.030	0.303308
## DOMAIN_NAMEhotels.com	-15.022735	3.949458	-3.804	0.000145
***				
## DOMAIN_NAMEhotwire.com	-14.364143	3.980133	-3.609	0.000311
***				
## DOMAIN_NAMEhyatt.com	-7.082213	4.208302	-1.683	0.092476
.				
## DOMAIN_NAMEichotelsgroup.com	-11.070925	4.060451	-2.727	0.006431
**				
## DOMAIN_NAMEjetblue.com	-25.951213	3.977533	-6.524	7.75e-11
***				
## DOMAIN_NAMEmarriott.com	-11.289094	3.936172	-2.868	0.004154
**				
## DOMAIN_NAMEnetbooker.com	-15.283170	7.421347	-2.059	0.039530
*				
## DOMAIN_NAMEorbitz.com	-9.083360	4.039505	-2.249	0.024594
*				
## DOMAIN_NAMEpriceline.com	-8.973373	3.948798	-2.272	0.023117
*				
## DOMAIN_NAMEradisson.com	-10.663645	4.950525	-2.154	0.031301
*				
## DOMAIN_NAMEramada.com	-12.392340	4.480378	-2.766	0.005704
**				
## DOMAIN_NAMEres99.com	-13.262458	4.091485	-3.241	0.001200
**				
## DOMAIN_NAMEritzcarlton.com	-13.664052	6.724732	-2.032	0.042234
*				
## DOMAIN_NAMEsouthwest.com	-3.664911	6.237266	-0.588	0.556848

```

## DOMAIN_NAMEstarwoodhotels.com -11.832073 4.172470 -2.836 0.004597
**

## DOMAIN_NAMESuper8.com -11.751001 4.202187 -2.796 0.005194
**

## DOMAIN_NAMEtravelnow.com -17.989257 4.288479 -4.195 2.80e-05
***

## DOMAIN_NAMEtravelocity.com 0.522534 4.032745 0.130 0.896911

## DOMAIN_NAMEtravelodge.com -9.766795 5.656395 -1.727 0.084308
.

## DOMAIN_NAMEwingatehotels.com -9.402376 5.464180 -1.721 0.085383
.

## DOMAIN_NAMEwwte1.com -2.420732 5.313436 -0.456 0.648714

## DOMAIN_NAMEwyndham.com -3.175064 5.499121 -0.577 0.563720

## DOMAIN_NAMEyahoo.net 0.677657 4.344586 0.156 0.876059

## DURATION 0.475348 0.008205 57.935 < 2e-16
***

## LOG_PRICE -0.591199 0.199941 -2.957 0.003127
**

## ---

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 10.92 on 3713 degrees of freedom

## Multiple R-squared: 0.5732, Adjusted R-squared: 0.5691

## F-statistic: 142.5 on 35 and 3713 DF, p-value: < 2.2e-16

```

c. Count Data (Poisson) Regression

The variables' significance is the same between these two models, as well as the positive and negative signs. However, coefficients are very different between two models. The linear model has bigger larger coefficients than the negative binomial model, especially for domain name variable. Most domain name variables' coefficients have double digits in linear model, while those in negative binomial have single digit after conversion and often very small under 2. The difference is huge, since linear regression assumes outcome variable is distributed from negative infinity to positive infinity and Count data only hold values from zero and above.

```
##
## Call:
## glm(formula = PAGES_VIEWED ~ TRANS_FREQ + DOMAIN_NAME + DURATION +
##     LOG_PRICE, family = quasipoisson(log), data = newfinal)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -20.9478  -1.6537   -0.6041    0.8698   18.6064
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.8648536   0.1481673  19.335  < 2e-16
***
## TRANS_FREQ      0.0093502   0.0022276   4.197 2.76e-05
***
## DOMAIN_NAMEaol.com    -0.2688607   0.4107817  -0.655 0.512824
## DOMAIN_NAMEbestwestern.com -0.2788207   0.1578296  -1.767 0.077378
.
## DOMAIN_NAMEcheaptickets.com -0.0759119   0.1736121  -0.437 0.661956
## DOMAIN_NAMEchoicehotels.com -0.8895801   0.1556348  -5.716 1.18e-08
***
## DOMAIN_NAMEcountryinns.com -0.3747581   0.1750259  -2.141 0.032326
*
## DOMAIN_NAMEdaysinn.com    -0.4063401   0.1663462  -2.443 0.014623
*
## DOMAIN_NAMEexpedia.com    -0.2945803   0.1507726  -1.954 0.050799
.
## DOMAIN_NAMEhilton.com    -0.2461691   0.1513254  -1.627 0.103874
```

## DOMAIN_NAMEhojo.com	0.1719951	0.1921423	0.895	0.370768
## DOMAIN_NAMEhotels.com	-0.6454361	0.1530290	-4.218	2.53e-05
***				
## DOMAIN_NAMEhotwire.com	-0.5898563	0.1554172	-3.795	0.000150
***				
## DOMAIN_NAMEhyatt.com	-0.2032694	0.1654643	-1.228	0.219345
## DOMAIN_NAMEichotelsgroup.com	-0.4014847	0.1608792	-2.496	0.012619
*				
## DOMAIN_NAMEjetblue.com	-1.7744554	0.1654381	-10.726	< 2e-16
***				
## DOMAIN_NAMEmarriott.com	-0.4051513	0.1511952	-2.680	0.007402
**				
## DOMAIN_NAMEnetbooker.com	-1.0913616	0.5548327	-1.967	0.049256
*				
## DOMAIN_NAMEorbitz.com	-0.2660047	0.1561704	-1.703	0.088596
.				
## DOMAIN_NAMEpriceline.com	-0.3590638	0.1516770	-2.367	0.017970
*				
## DOMAIN_NAMEradisson.com	-0.3478131	0.2177117	-1.598	0.110220
## DOMAIN_NAMEramada.com	-0.5902498	0.2039388	-2.894	0.003823
**				
## DOMAIN_NAMEres99.com	-0.6043610	0.1664228	-3.631	0.000286
***				
## DOMAIN_NAMEritzcarlton.com	-0.4903791	0.3322147	-1.476	0.140004
## DOMAIN_NAMEsouthwest.com	-0.0981261	0.2168156	-0.453	0.650879
## DOMAIN_NAMEstarwoodhotels.com	-0.4433109	0.1678349	-2.641	0.008292
**				
## DOMAIN_NAMEsuper8.com	-0.5305875	0.1765741	-3.005	0.002674
**				



```
## DOMAIN_NAMEtravelnow.com      -1.3475455  0.2195274 -6.138 9.21e-10
***

## DOMAIN_NAMEtravelocity.com    -0.0127420  0.1529882 -0.083 0.933627
## DOMAIN_NAMEtravelodge.com     -0.3690059  0.2740388 -1.347 0.178209
## DOMAIN_NAMEwingatehotels.com  -0.3779797  0.2655319 -1.423 0.154681
## DOMAIN_NAMEwwte1.com          -0.0307213  0.2011122 -0.153 0.878598
## DOMAIN_NAMEwyndham.com        -0.1755796  0.2244577 -0.782 0.434124
## DOMAIN_NAMEyahoo.net          0.0911712  0.1661822  0.549 0.583298
## DURATION                      0.0159393  0.0002751 57.948 < 2e-16
***

## LOG_PRICE                     -0.0221749  0.0090022 -2.463 0.013812
*

## ---

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for quasipoisson family taken to be 5.70724)

##

##    Null deviance: 45009  on 3748  degrees of freedom
## Residual deviance: 19043  on 3713 degrees of freedom
## AIC: NA

##

## Number of Fisher Scoring iterations: 5
```

*#actual number*

`exp(abs(glm.model4$coefficients))`

##	(Intercept)	TRANS_FREQ
##	17.546485	1.009394
##	DOMAIN_NAMEaol.com	DOMAIN_NAMEbestwestern.com
##	1.308473	1.321570
##	DOMAIN_NAMEcheaptickets.com	DOMAIN_NAMEchoicehotels.com
##	1.078868	2.434107
##	DOMAIN_NAMEcountryinns.com	DOMAIN_NAMEdaysinn.com
##	1.454640	1.501313
##	DOMAIN_NAMEexpedia.com	DOMAIN_NAMEhilton.com
##	1.342563	1.279116
##	DOMAIN_NAMEhojo.com	DOMAIN_NAMEhotels.com
##	1.187672	1.906818
##	DOMAIN_NAMEhotwire.com	DOMAIN_NAMEhyatt.com
##	1.803729	1.225403
##	DOMAIN_NAMEichotelsgroup.com	DOMAIN_NAMEjetblue.com
##	1.494041	5.897069
##	DOMAIN_NAMEmarriott.com	DOMAIN_NAMEnetbooker.com
##	1.499529	2.978326
##	DOMAIN_NAMEorbitz.com	DOMAIN_NAMEpriceline.com
##	1.304741	1.431988
##	DOMAIN_NAMeradisson.com	DOMAIN_NAMeramada.com
##	1.415968	1.804439

##	DOMAIN_NAMEeres99.com	DOMAIN_NAMEritzcarlton.com
##	1.830082	1.632935
##	DOMAIN_NAMESouthwest.com	DOMAIN_NAMEstarwoodhotels.com
##	1.103102	1.557857
##	DOMAIN_NAMESuper8.com	DOMAIN_NAMEtravelnow.com
##	1.699931	3.847969
##	DOMAIN_NAMEtravelocity.com	DOMAIN_NAMEtravelodge.com
##	1.012824	1.446296
##	DOMAIN_NAMEwingatehotels.com	DOMAIN_NAMEwwte1.com
##	1.459333	1.031198
##	DOMAIN_NAMEwyndham.com	DOMAIN_NAMEyahoo.net
##	1.191937	1.095457
##	DURATION	LOG_PRICE
##	1.016067	1.022423

#### d. Summary

From these models, we can see that Pages viewed and duration are strongly correlated. It means that more pages viewed can increase duration and vice versa. The website could focus on either increasing the pages viewed or duration.

Domain name plays a strong role in both of the variables. Domain name generally has positive effect in duration, but negative effect on pages viewed. It is an interesting observation, but we can interpret that certain website performs better, can retain users longer

and has less decrease in page views. So, the website should improve its landing page and user experience to help the performance.

To increase pages viewed, main variables we should focus on improving are TRANS\_FREQ, DOMAIN\_NAME, DURATION, and LOG\_PRICE. Because LOG\_PRICE has the largest absolute coefficient and is negative, it has the largest impact on pages viewed. Thus, the website should work on cost efficiency and lower the price of the products. Transaction frequency also played a role in pages viewed, as a more frequent visited customer trends to view more pages.

To increase duration, main variables we should focus on improving are DOMAIN\_NAME, PROD\_QTY, PAGES\_VIEWED, and REF\_D. Whether the transaction has a referred website has a significant impact on duration, so improve on the relationship and linkage with other websites is essential. Also, increasing product quantity will also increase the duration, so increasing the product variety will help.

In sum, improving TRANS\_FREQ, DOMAIN\_NAME, LOG\_PRICE, REF\_D, PROD\_QTY will help increase the pages viewed and duration.