

# Google analytics customer revenue prediction

Hao Shen

2020/12/09

## Abstract

Nowadays, more and more large company is implementing a data-driven marketing strategy and Google, as an Internet giant, is undoubtedly in a leading position. In this project provided by Google, we mainly cleaned and transformed the raw data sets and built a linear mixed-effects model combined with feature engineering to predict revenue from each customer based on their characteristic including time on sites, newcomer or old customer, etc. The model fits well at the middle of Q-Q plot, however, deviation occurs in two tails which requires further improvements.

## Introduction

Google Merchandise Store is an online shopping site that sells Google brands, including Youtube, Android and Google, products. The goods it sells include clothing, daily necessities, and stationery, covering all age groups from children to adults. As an online shop, it will naturally generate large historical customer behavior data that can be used for analysis and even further to predict future sales and support marketing decision making.

The data set of this project is available on Kaggle, which contains detailed transactions information such as date, geography, device, time, etc. from GStore and is consisted by two sub-sets:

- One is the train set from August 1st 2016 to April 30th 2018, with 13 columns, 1.71 million observations and a size of 23.67GB;
- The other is test set from May 1st 2018 to October 15th 2018, with also 13 columns but only 0.402 million observations and a size of 7.04GB.

Among all these 13 columns, 4 columns with Json format are extremely large and contain the most customer features we need for model fitting.

## Method

### Data conversion and cleaning

Since the total size of \*.csv files is larger than 30GB, it is hard to directly import them into computer memory. In order to solve this problem, we choose to firstly load them into SQLite database. Then we extract the 4 large Json columns (device, geoNetwork, totals, trafficSource) one by one through ‘getSQLquery’ function and transform them into flat data.frame format. Next, We start data cleaning also one by one column to reduce memory usage. we drop those sub-columns only contains single value like ‘not available in demo dataset’ and finally combine them four into a single data set. Finally, we get a data frame with 57 columns and 2,109,926 observations.

Then, we use geom\_col function to explore the components of 58 columns and transform their content into correct format. These transformations can be devided into tree parts:

- Date and POSIX date transformation and split them into minute, hour, day, etc.
- Replace NAs in parts of columns like transaction revenue, time on site, page views, etc. by 0.
- Replace NAs in parts of columns like is new visits, is video ad, etc. by FALSE.

- Delete other useless columns.

Thanks to all above efforts, we finally reduce the data set into 3.36GB with 24 columns and it can now be used for EDA.

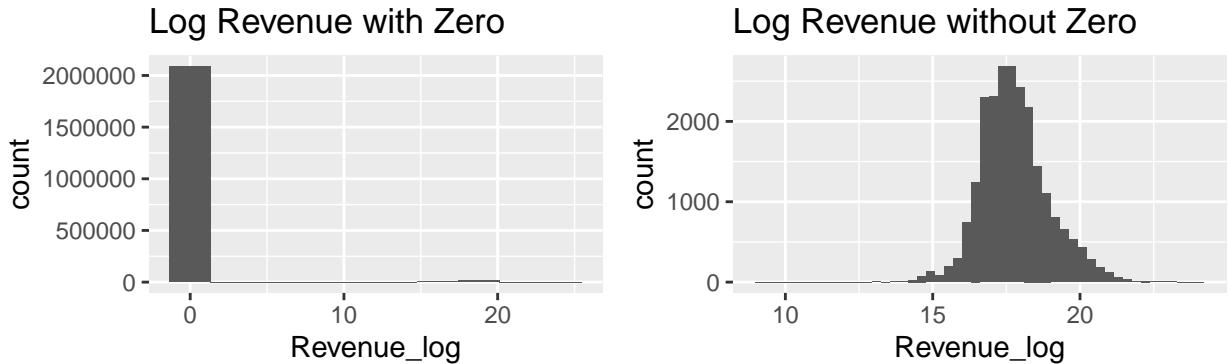
### Exploratory data analysis

At first, let us take a look of the outcome variable Revenue variables. Since it is about money we choose to apply log transformation:

$$\text{Revenue} = \ln(\text{TransactionRevenue} + 1)$$

As shown in the left plot, there are 98.9% of target variable is zero. However if we ignore those data, we can obtain right plot which is a bit similar to a normal distribution. Besides, considering my computer cannot run even a simple linear regression model on these 3.36GB data sets with more than 2 million observations, we choose to directly delete those zero observations and fit model on those with non-zero revenue.

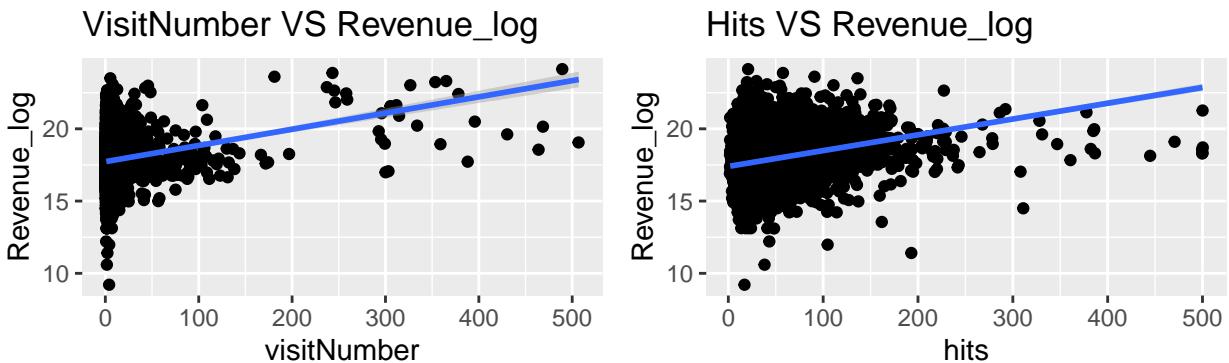
Note: I initially thought that maybe I can firstly build a model to identify those zero transactions and then go to the second model to fit the results. But finally I failed.



Then, for each of other 23 variables, we draw bar plots, line plots, scatter plots according to different types of variables (All these plot codes is in Supplement of Code) and create several new logical features:

- **visitNumLess6** which is FALSE when visit number of a customer exceeds 6.
- **hitsIn4T18** which is TRUE when a customer has total hits between 4 and 18.
- **timeOnSiteIn100T400** which is TRUE when a customer has total on site time between 100 and 400 seconds.
- ...

We choose to create these new features is because tons of data points located in these periods.



## Model fitting

Here, let's first fit a basic linear model **lm1** with all other variables.

$$\begin{aligned} Revenue_{log} = & channelGrouping + date + year + month + day + hour + minute + visitNumber + \\ & operatingSystem + deviceCategory + continent + country + hits + pageviews + bounces + \\ & newVisits + sessionDim + timeOnSite + medium + trueDirect + isVideoAd \\ & + visitNumLess6 + visitNumIn6T50 + hitsIn4T18 + hitsIn18T70 + \\ & sessionDim0 + timeOnSiteIn100T400 + timeOnSiteIn400T1800 \end{aligned}$$

Then, we need to delete some variables where collinearity exists and add some interactions to fit an updated model **lm2** as:

$$\begin{aligned} Revenue_{log} = & pageviews + medium + trueDirect + operatingSystem + \\ & visitNumber * visitNumLess6 + hits * hitsIn4_18 \end{aligned}$$

As a result, the second linear model got a just bit higher RSS and AIC but far more less variables.

Then, since each customer has a unique **fullVisitorId**, and between each customers random effects occurs, we choose to fit a linear mixed-effect model **lmer1**.

$$\begin{aligned} Revenue_{log} = & pageviews + medium + trueDirect + operatingSystem + visitNumber + \\ & hits + (1|fullVisitorId) \end{aligned}$$

This model is better than simple linear regression as the AIC decreases from 55748.05 to 54849.30.

Next, we should use the new features we created at beginning, to futher improve the model performance as **lmer2**:

$$\begin{aligned} Revenue_{log} = & pageviews + medium + trueDirect + operatingSystem + \\ & visitNumber * visitNumLess6 + hits * hitsIn4T18 + (1|fullVisitorId) \end{aligned}$$

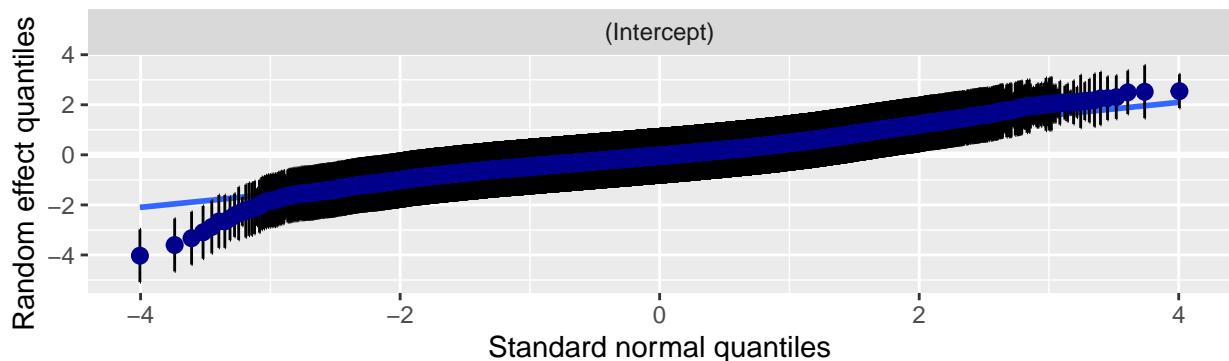
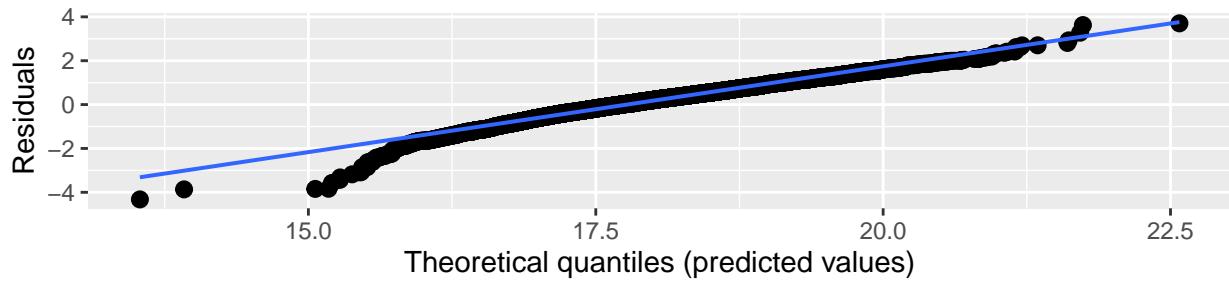
And this is the best model with lowest AIC.

	df	AIC
lm1	117	55501.12
lm2	22	55748.05
lmer1	19	54849.30
lmer2	23	54663.04

## Result

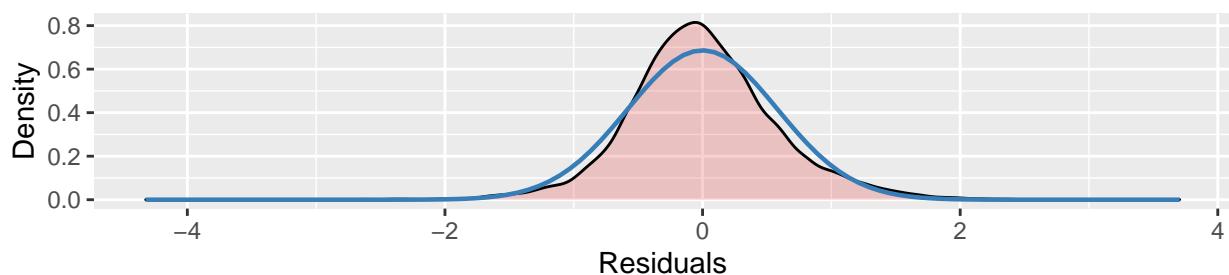
**Non-normality of residuals and outliers**

Dots should be plotted along the line



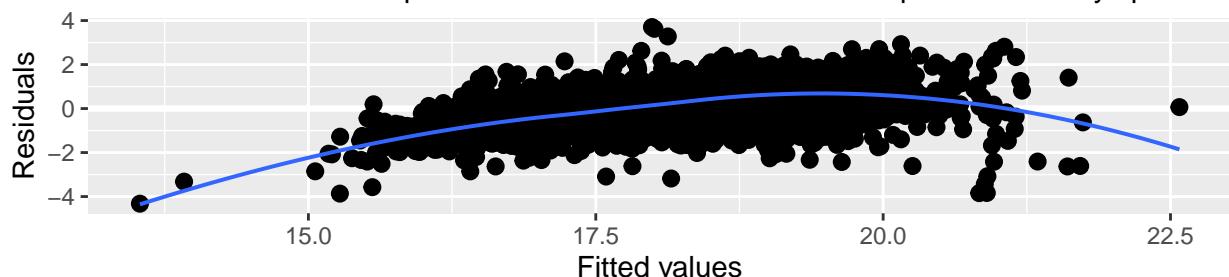
**Non-normality of residuals**

Distribution should look like normal curve



**Homoscedasticity (constant variance of residuals)**

Amount and distance of points scattered above/below line is equal or randomly spread



As all results shows, the model fits well at the middle of the data set, but deviation occurs in two tails. The residual plots also shows a curve among points which also indicates the deviation that need further improvements.

## **Discussion**

The data set used for regression is those with transaction revenue not equal to 0 which is only about 1% of the total data set. This indicates that there are lots of information about those zero transactions haven't been figured out. And if we combine this with the bad fittings of two tails, the problems maybe more interesting.

The second thing is that even there is only about 21000 observations, there are more than 16000 unique fullVisitorIDs. I am worried about whether this will have impacts on the linear mixed-effect model results.

Besides, about variables' concentration at lower values, I believe this need to dig further.

## Appendix

1. Kaggle-Google Analytics Customer Revenue Prediction
2. sjPlot package | R Documentation
3. Create Awesome HTML Table with knitr::kable and kableExtra
4. R Database Interface · DBI
5. R - Flatten JSON columns to make single data frame

### lm1 summary

```
##  
## Call:  
## lm(formula = Revenue_log ~ channelGrouping + date + year + month +  
##      day + hour + minute + visitNumber + operatingSystem + deviceCategory +  
##      continent + country + hits + pageviews + bounces + newVisits +  
##      sessionDim + timeOnSite + medium + trueDirect + isVideoAd +  
##      visitNumLess6 + visitNumIn6_50 + hitsIn4_18 + hitsIn18_70 +  
##      sessionDim0 + timeOnSiteIn100_400 + timeOnSiteIn400_1800,  
##      data = data1[1:train1, ])  
##  
## Residuals:  
##       Min     1Q   Median     3Q    Max  
## -8.5100 -0.6802 -0.0673  0.6140  5.0732  
##  
## Coefficients: (21 not defined because of singularities)  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 4.863e+03 9.113e+03  0.534 0.593588  
## channelGroupingAffiliates -1.772e-01 1.112e+00 -0.159 0.873413  
## channelGroupingDirect 8.286e-01 1.081e+00  0.766 0.443469  
## channelGroupingDisplay -3.789e-01 1.163e+00 -0.326 0.744592  
## channelGroupingOrganic Search 4.268e-01 1.081e+00  0.395 0.693026  
## channelGroupingPaid Search 1.523e-01 1.093e+00  0.139 0.889194  
## channelGroupingReferral 6.060e-01 1.081e+00  0.561 0.575138  
## channelGroupingSocial 6.402e-02 1.086e+00  0.059 0.953002  
## date 6.660e-03 1.270e-02  0.524 0.599951  
## year -2.459e+00 4.626e+00 -0.532 0.595045  
## month -2.202e-01 3.853e-01 -0.572 0.567635  
## day -6.722e-03 1.267e-02 -0.531 0.595725  
## hour -7.051e-04 1.131e-03 -0.623 0.533135  
## minute -9.390e-05 4.618e-04 -0.203 0.838895  
## visitNumber 7.135e-03 6.845e-04 10.424 < 2e-16 ***  
## operatingSystemChrome OS 3.577e-01 2.150e-01  1.664 0.096167 .  
## operatingSystemiOS -5.067e-02 5.570e-02 -0.910 0.363063  
## operatingSystemLinux -1.532e-01 2.154e-01 -0.711 0.476785  
## operatingSystemMacintosh 2.280e-01 2.137e-01  1.067 0.286116  
## operatingSystemWindows 3.904e-02 2.127e-01  0.184 0.854393  
## operatingSystemWindows Phone -4.522e-01 1.081e+00 -0.418 0.675769  
## deviceCategorymobile -3.369e-01 2.108e-01 -1.598 0.109974  
## deviceCategorytablet -1.554e-01 2.152e-01 -0.722 0.470311  
## continentAustralasia -2.282e-01 7.458e-01 -0.306 0.759636  
## continentCaribbean -8.969e-01 8.665e-01 -1.035 0.300647  
## continentCentral America -6.196e-01 8.671e-01 -0.715 0.474831  
## continentCentral Asia -1.359e+00 8.668e-01 -1.568 0.116903  
## continentEastern Africa 2.785e+00 7.459e-01  3.734 0.000189 ***  
## continentEastern Asia -4.060e-01 4.515e-01 -0.899 0.368597
```

## continentEastern Europe	-1.596e+00	5.597e-01	-2.851	0.004358	**
## continentNorthern Africa	-7.321e-01	7.475e-01	-0.979	0.327385	
## continentNorthern America	-4.378e-01	4.090e-01	-1.071	0.284371	
## continentNorthern Europe	-5.180e-01	4.627e-01	-1.120	0.262880	
## continentSouth America	9.241e-02	4.306e-01	0.215	0.830094	
## continentSoutheast Asia	-7.937e-01	5.328e-01	-1.490	0.136302	
## continentSouthern Africa	-7.768e-02	7.460e-01	-0.104	0.917065	
## continentSouthern Asia	-1.306e+00	7.462e-01	-1.751	0.080012	.
## continentSouthern Europe	2.328e-02	7.463e-01	0.031	0.975117	
## continentWestern Africa	2.997e+00	8.665e-01	3.459	0.000544	***
## continentWestern Asia	-4.895e-01	6.776e-01	-0.722	0.470002	
## continentWestern Europe	-5.434e-01	5.144e-01	-1.056	0.290827	
## countryAlgeria	3.262e-01	1.250e+00	0.261	0.794107	
## countryAnguilla	-1.548e-01	1.324e+00	-0.117	0.906985	
## countryArgentina	-5.021e-01	4.314e-01	-1.164	0.244463	
## countryArmenia	-1.920e+00	1.209e+00	-1.588	0.112356	
## countryAustralia	5.458e-01	6.587e-01	0.829	0.407364	
## countryAustria	1.070e+00	1.125e+00	0.951	0.341457	
## countryBelgium	8.562e-01	5.139e-01	1.666	0.095669	.
## countryBelize	5.383e-01	1.326e+00	0.406	0.684695	
## countryBrazil	-5.604e-01	2.736e-01	-2.048	0.040570	*
## countryCambodia	-1.256e+00	1.133e+00	-1.108	0.267881	
## countryCanada	-5.462e-02	6.181e-02	-0.884	0.376890	
## countryChile	-1.262e+00	4.315e-01	-2.924	0.003462	**
## countryChina	-3.927e-01	4.809e-01	-0.817	0.414212	
## countryColombia	-4.432e-01	3.682e-01	-1.204	0.228773	
## countryCosta Rica	-1.284e+00	1.081e+00	-1.188	0.234948	
## countryCura<U+00E7>ao	8.852e-01	1.323e+00	0.669	0.503426	
## countryCyprus	4.601e-01	1.208e+00	0.381	0.703374	
## countryCzechia	1.290e+00	1.146e+00	1.126	0.260255	
## countryDenmark	-3.202e-01	1.102e+00	-0.290	0.771441	
## countryDominican Republic	-6.930e-01	1.326e+00	-0.523	0.601181	
## countryEcuador	-3.764e-02	5.023e-01	-0.075	0.940275	
## countryEgypt	NA	NA	NA	NA	
## countryEl Salvador	7.950e-01	1.324e+00	0.601	0.548119	
## countryFinland	-8.410e-01	7.943e-01	-1.059	0.289697	
## countryFrance	1.771e-01	4.510e-01	0.393	0.694534	
## countryGeorgia	-1.519e-01	9.361e-01	-0.162	0.871099	
## countryGermany	8.147e-02	4.411e-01	0.185	0.853469	
## countryGreece	-3.366e-01	8.820e-01	-0.382	0.702757	
## countryGuadeloupe	-7.445e-01	1.323e+00	-0.563	0.573753	
## countryGuatemala	-9.124e-01	1.081e+00	-0.844	0.398649	
## countryHong Kong	4.492e-01	3.781e-01	1.188	0.234890	
## countryHungary	6.199e-01	8.543e-01	0.726	0.468073	
## countryIndia	5.491e-01	6.743e-01	0.814	0.415468	
## countryIndonesia	9.943e-01	4.357e-01	2.282	0.022489	*
## countryIreland	-2.909e-01	3.913e-01	-0.743	0.457227	
## countryIsrael	1.125e-01	6.496e-01	0.173	0.862540	
## countryItaly	-6.109e-01	8.257e-01	-0.740	0.459359	
## countryJapan	4.075e-01	2.996e-01	1.360	0.173796	
## countryKazakhstan	NA	NA	NA	NA	
## countryKenya	NA	NA	NA	NA	
## countryKuwait	5.099e-01	9.359e-01	0.545	0.585877	
## countryLebanon	-2.561e-01	1.208e+00	-0.212	0.832132	

## countryMalaysia	8.480e-01	6.395e-01	1.326	0.184845
## countryMexico	2.989e-01	7.922e-01	0.377	0.705969
## countryNetherlands	-2.053e-01	5.754e-01	-0.357	0.721249
## countryNew Zealand		NA	NA	NA
## countryNicaragua	6.040e-01	1.323e+00	0.456	0.648085
## countryNigeria		NA	NA	NA
## countryNorway	1.998e+00	7.947e-01	2.515	0.011926 *
## countryPakistan		NA	NA	NA
## countryPanama		NA	NA	NA
## countryPeru	-3.819e-01	7.761e-01	-0.492	0.622675
## countryPhilippines	8.910e-01	5.325e-01	1.673	0.094293 .
## countryPoland	9.035e-01	7.317e-01	1.235	0.216926
## countryPortugal	6.389e-01	8.826e-01	0.724	0.469137
## countryPuerto Rico	7.128e-01	8.209e-01	0.868	0.385267
## countryRomania	1.728e+00	1.151e+00	1.501	0.133284
## countryRussia	3.405e-01	5.594e-01	0.609	0.542766
## countrySaudi Arabia	6.717e-01	8.254e-01	0.814	0.415755
## countrySingapore	2.852e-01	4.152e-01	0.687	0.492103
## countrySouth Africa		NA	NA	NA
## countrySouth Korea	-1.105e-01	3.659e-01	-0.302	0.762658
## countrySpain		NA	NA	NA
## countrySt. Lucia		NA	NA	NA
## countrySweden	-7.490e-01	7.941e-01	-0.943	0.345597
## countrySwitzerland		NA	NA	NA
## countryTaiwan		NA	NA	NA
## countryThailand		NA	NA	NA
## countryTurkey	-1.361e-01	9.355e-01	-0.145	0.884351
## countryUkraine		NA	NA	NA
## countryUnited Arab Emirates		NA	NA	NA
## countryUnited Kingdom		NA	NA	NA
## countryUnited States		NA	NA	NA
## countryUruguay	-1.666e+00	7.765e-01	-2.146	0.031918 *
## countryVenezuela		NA	NA	NA
## hits	2.517e-02	1.464e-03	17.194	< 2e-16 ***
## pageviews	-2.649e-02	2.043e-03	-12.962	< 2e-16 ***
## bouncesTRUE		NA	NA	NA
## newVisitsTRUE	-2.521e-01	2.499e-02	-10.089	< 2e-16 ***
## sessionDim	3.217e-04	6.301e-04	0.511	0.609631
## timeOnSite	7.457e-05	1.746e-05	4.271	1.96e-05 ***
## medium(not set)		NA	NA	NA
## mediumaffiliate		NA	NA	NA
## mediumcpc	1.123e-01	8.809e-02	1.275	0.202261
## mediumcpm	1.273e+00	4.360e-01	2.921	0.003499 **
## mediumorganic	6.914e-02	3.678e-02	1.880	0.060136 .
## mediumreferral	1.622e-01	6.326e-02	2.564	0.010349 *
## trueDirectTRUE	1.188e-02	2.606e-02	0.456	0.648601
## isVideoAdTRUE	3.354e-01	1.507e-01	2.226	0.026039 *
## visitNumLess6TRUE	-2.410e-01	2.539e-02	-9.495	< 2e-16 ***
## visitNumIn6_50TRUE	7.501e-02	6.281e-01	0.119	0.904945
## hitsIn4_18TRUE	-2.667e-01	5.310e-02	-5.022	5.15e-07 ***
## hitsIn18_70TRUE	-7.748e-02	4.355e-02	-1.779	0.075208 .
## sessionDim0TRUE	1.566e-01	5.507e-02	2.844	0.004462 **
## timeOnSiteIn100_400TRUE	-4.302e-02	4.568e-02	-0.942	0.346354
## timeOnSiteIn400_1800TRUE	2.532e-02	3.578e-02	0.708	0.479102

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.08 on 18398 degrees of freedom
## Multiple R-squared: 0.1762, Adjusted R-squared: 0.1711
## F-statistic: 34.23 on 115 and 18398 DF, p-value: < 2.2e-16

lm2 summary

##
## Call:
## lm(formula = Revenue_log ~ pageviews + medium + trueDirect +
##     operatingSystem + visitNumber + visitNumLess6 + hits + hitsIn4_18 +
##     visitNumber:visitNumLess6 + hits:hitsIn4_18, data = data1[1:train1,
##     ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3778 -0.6881 -0.0701  0.6209  5.2549
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                17.3032018  0.0539677 320.621 < 2e-16 ***
## pageviews                 -0.0185935  0.0018983 -9.795 < 2e-16 ***
## medium(not set)           -0.6302342  1.0908880 -0.578 0.563456
## mediumaffiliate            -0.8092056  0.2645766 -3.058 0.002228 **
## mediumcpc                  -0.0711800  0.0513219 -1.387 0.165479
## mediumcpm                  0.2835493  0.0804681  3.524 0.000427 ***
## mediumorganic              -0.1412296  0.0204941 -6.891 5.71e-12 ***
## mediumreferral             -0.0368817  0.0518098 -0.712 0.476556
## trueDirectTRUE             0.1350661  0.0203477  6.638 3.27e-11 ***
## operatingSystemChrome OS    0.6761761  0.0482606 14.011 < 2e-16 ***
## operatingSystemiOS          -0.0383078  0.0548956 -0.698 0.485292
## operatingSystemLinux         0.1448761  0.0507918  2.852 0.004344 **
## operatingSystemMacintosh    0.5320580  0.0418769 12.705 < 2e-16 ***
## operatingSystemWindows        0.3710508  0.0439357  8.445 < 2e-16 ***
## operatingSystemWindows Phone -0.4863148  1.0907429 -0.446 0.655706
## visitNumber                 0.0071202  0.0006905 10.312 < 2e-16 ***
## visitNumLess6TRUE           -0.5017812  0.0330016 -15.205 < 2e-16 ***
## hits                        0.0220868  0.0013377 16.511 < 2e-16 ***
## hitsIn4_18TRUE              -0.2556153  0.0853413 -2.995 0.002746 **
## visitNumber:visitNumLess6TRUE 0.0855136  0.0084201 10.156 < 2e-16 ***
## hits:hitsIn4_18TRUE          0.0040716  0.0061047  0.667 0.504805
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.09 on 18493 degrees of freedom
## Multiple R-squared: 0.1566, Adjusted R-squared: 0.1557
## F-statistic: 171.6 on 20 and 18493 DF, p-value: < 2.2e-16

```

### lmer1 summary

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: Revenue_log ~ pageviews + medium + trueDirect + operatingSystem +
##     visitNumber + hits + (1 | fullVisitorId)

```

```

##      Data: data1[1:train1, ]
##
## REML criterion at convergence: 54811.3
##
## Scaled residuals:
##      Min      1Q  Median      3Q     Max
## -5.5143 -0.4568 -0.0382  0.4210  4.8451
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   fullVisitorId (Intercept) 0.5726    0.7567
##   Residual          0.6025    0.7762
## Number of obs: 18514, groups: fullVisitorId, 16141
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                16.837685  0.045228 372.281
## pageviews                  -0.019526  0.001865 -10.472
## medium(not set)            -0.785513  1.084785 -0.724
## mediumaffiliate             -0.760033  0.268387 -2.832
## mediumcpc                   -0.054796  0.050365 -1.088
## mediumcpm                   0.223389  0.078928  2.830
## mediumorganic               -0.124055  0.020643 -6.009
## mediumreferral              -0.034641  0.051820 -0.668
## trueDirectTRUE              0.218065  0.016909 12.897
## operatingSystemChrome OS    0.644865  0.050033 12.889
## operatingSystemiOS           -0.021816  0.055653 -0.392
## operatingSystemLinux          0.215081  0.051745  4.157
## operatingSystemMacintosh     0.588168  0.042676 13.782
## operatingSystemWindows        0.373246  0.045002  8.294
## operatingSystemWindows Phone -0.295567  1.084878 -0.272
## visitNumber                  0.007761  0.001156  6.713
## hits                         0.023023  0.001306 17.629

```

### lmer2 summary

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: Revenue_log ~ pageviews + medium + trueDirect + operatingSystem +
##           visitNumber + visitNumLess6 + hits + hitsIn4_18 + (1 | fullVisitorId) +
##           visitNumber:visitNumLess6 + hits:hitsIn4_18
## Data: data1[1:train1, ]
##
## REML criterion at convergence: 54617
##
## Scaled residuals:
##      Min      1Q  Median      3Q     Max
## -5.5789 -0.4560 -0.0391  0.4209  4.7795
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##   fullVisitorId (Intercept) 0.5596    0.7481
##   Residual          0.6006    0.7750
## Number of obs: 18514, groups: fullVisitorId, 16141
##
```

```

## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                17.233157   0.055122 312.636
## pageviews                 -0.018987   0.001858 -10.220
## medium(not set)           -0.668427   1.078208 -0.620
## mediumaffiliate            -0.783798   0.266675 -2.939
## mediumcpc                  -0.098486   0.050227 -1.961
## mediumcpm                  0.178365   0.078565  2.270
## mediumorganic               -0.144370   0.020579 -7.015
## mediumreferral              -0.059098   0.051767 -1.142
## trueDirectTRUE              0.119818   0.019537  6.133
## operatingSystemChrome OS    0.615942   0.049782 12.373
## operatingSystemiOS          -0.017359   0.055301 -0.314
## operatingSystemLinux         0.189250   0.051483  3.676
## operatingSystemMacintosh    0.555778   0.042510 13.074
## operatingSystemWindows       0.375166   0.044715  8.390
## operatingSystemWindows Phone -0.420760   1.078103 -0.390
## visitNumber                 0.002387   0.001281  1.863
## visitNumLess6TRUE            -0.391858   0.034695 -11.294
## hits                        0.021453   0.001304 16.457
## hitsIn4_18TRUE               -0.247975   0.081566 -3.040
## visitNumber:visitNumLess6TRUE 0.079533   0.008130  9.783
## hits:hitsIn4_18TRUE          0.003925   0.005829  0.673

```