Marketing time series Analysis using R

Student's name: Jinning Yan
Professor's name: Thomas Kurnicki
Course: Forecasting & Predicting the Future Using Data - DAT-3531 - NYC1
Date: June 18th, 2023

Essay Wordcount: 1969

Table of Content:

Introduction:

In the era of data-driven decision making, businesses are increasingly leveraging analytics to gain insights, optimize operations, and boost profitability. This essay presents an analytical study conducted on a sales dataset, employing various statistical and machine learning techniques to extract valuable business insights.

The data comes from https://www.kaggle.com/datasets/denis6715/marketing-sales.

Data Preparation and Initial Analysis:

The data includes a total of 14 columns: date (transaction time, from = January 1, 2020 to September 30, 2020), source (the source of the transaction includes Google, Yandex, etc.) medium (the medium used when the transaction occurred)), delivery_available (whether it can be delivered, mainly including the two categories of yes and no data), device_type (the type of device used, mainly including PC and mobile, promo_activated (whether it is a promotional product, including the two categories of True and False), filter_used (Whether a filter is used when shopping), pageviews (page views), visits (page visits), productClick (product clicks), addToCart (number of times added to the shopping cart), checkout (number of times to enter the checkout) , transactions (commodity transactions) and revenue (income, If the transaction has not been completed, then the value is 0).

The 'transaction' variable, indicating whether a transaction occurred, serves as our response variable. Initial data cleaning revealed no missing values, ensuring the robustness of subsequent analyses.

Referring to the dataset, there are several pieces of data every day, and I plan to analyze this data from two angles. One is to combine daily transactions, calculate the total daily income, and then build a time series model to forecast its future income. The other path is to analyze directly in the form of raw data, using logistics regression and decision trees to predict whether a transaction will occur.

Predictive Modeling and Evaluation:

In order to get the forecast for future sales, it is important to combine the individual incomes into a total daily income, where the codes is as followed.

```
total <- sales %>% group_by(date) %>% summarise(total = sum(revenue))

daily_revenue <- ts(total$total, start = c(2020, 1, 1), frequency = 3
65)

plot(daily_revenue)
```

The diagrams below can observe that there is no obvious trend in this time series, and one can argue that the shape of this data is between non-stationary and somewhat semi-stationary because of the fluctuations. However, after performing the Augmented Dickey-Fuller Test, the data is stationary. The transformation of normalization does not change the distribution form, and the data after normalization still has the same shape as the previous data.

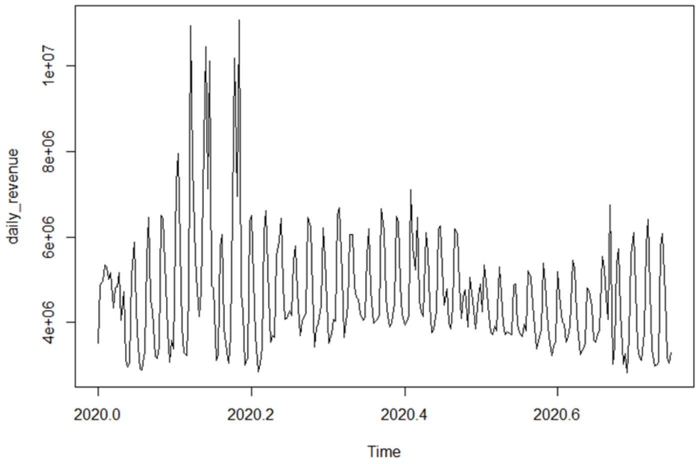

Exhibit 1: Non-normalized

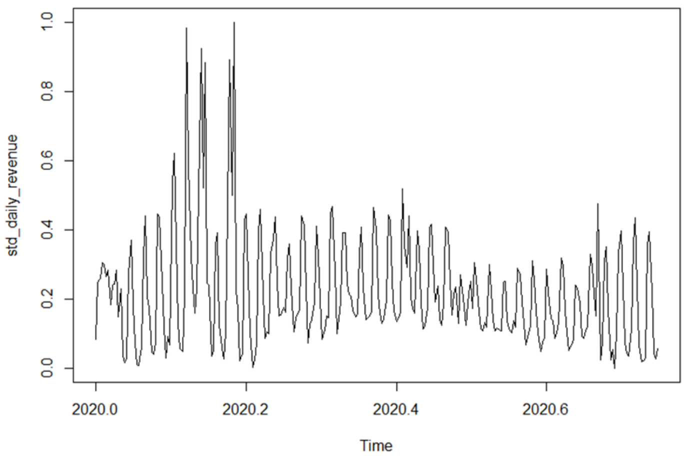

Exhibit 2: Normalized

By analyzing the autocorrelation and partial autocorrelation plots in Exhibit 3, there is a partial autocorrelation between the data and the first three lags. This suggests that the revenue values are influenced by their past values up to three days ago. This can help the business understand the temporal dynamics of revenue and anticipate short-term fluctuations. Through the previous exhibitions, it is observed that there is no obvious trend but frequent fluctuations. This suggests that the revenue does not exhibit a clear long-term increasing or decreasing trend. However, it is important to note that the ARMA model does not explicitly capture seasonality.

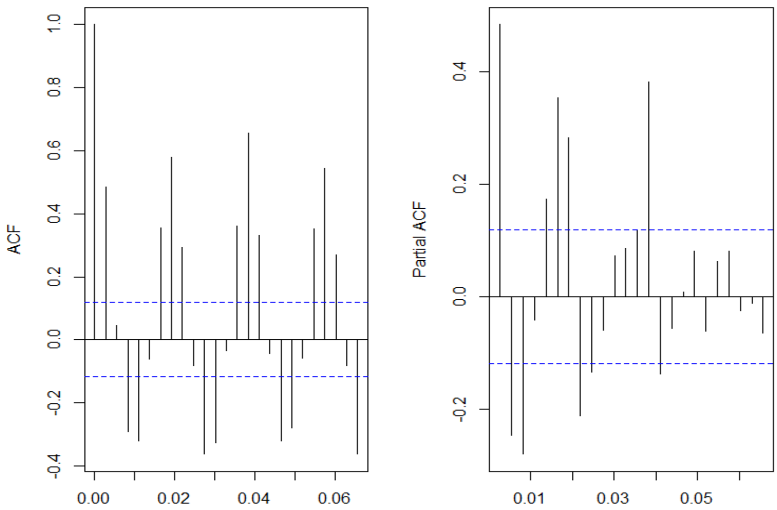

Exhibit 3: ACF & pACF

After the not ideal box test and plotting the residuals of the AR(3) model (Exhibition 4), we can observe that the residuals fluctuate around 0 in a wide range. This suggests that the model may not fully capture all patterns.

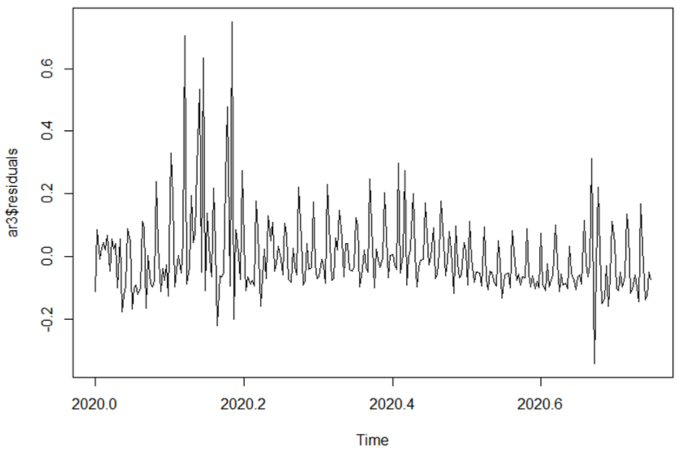

Exhibit 4

By considering the lagged revenue values, the business can make informed decisions on pricing adjustments or promotional activities to maximize revenue. Monitoring the fluctuations in revenue and analyzing the residuals can help identify unusual patterns in the data, which can guide the business in detecting potential issues or opportunities and taking appropriate actions. However, due to the need to improve the model, other models such as GARCH to address heteroscedasticity. This would provide a better understanding of the volatility in revenue and will help the business with risk management.

The analysis of the daily revenue data using ARIMA modeling techniques revealed valuable business insights and forecasts. The ARIMA(12,1,1) model was optimized to capture the temporal patterns and fluctuations in the daily revenue. The Box-Pierce test indicated that the residuals of the model no longer exhibited significant autocorrelation, suggesting that the model effectively captures the underlying dynamics of the revenue data. By considering the critical value of 1.645, coefficients that were not statistically significant were removed. The resulting ARIMA(12,1,1) model with sparse coefficients demonstrated that the remaining coefficients were significant, strengthening the model's predictive power. After removing the insignificant coefficients, the remaining coefficients are significant, and the built model is:

$$\begin{aligned} Y_t - Y_{t-1} = & -4.5541(Y_{t-3} - Y_{t-4}) - 4.9741(Y_{t-4} - Y_{t-5}) - 4.7865(Y_{t-5} - Y_{t-6}) \\ & - 3.0816(Y_{t-8} - Y_{t-9}) - 4.8989(Y_{t-9} - Y_{t-10}) \\ & - 4.2846(Y_{t-10} - Y_{t-11}) - 2.6598(Y_{t-11} - Y_{t-12}) \\ & - 2.6995(Y_{t-12} - Y_{t-13}) - 15.6181\varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

With this improved model, a 5-step forecast of the total daily revenue was generated (Exhibit 5). The forecasted values aligned well with the previous fluctuation patterns, indicating that the model captured the underlying dynamics of the revenue data. Leveraging this forecasted information, the business can make more accurate revenue projections and gain insights into future revenue trends. This enables them to make data-driven decisions regarding many things, 3 examples are provided below.

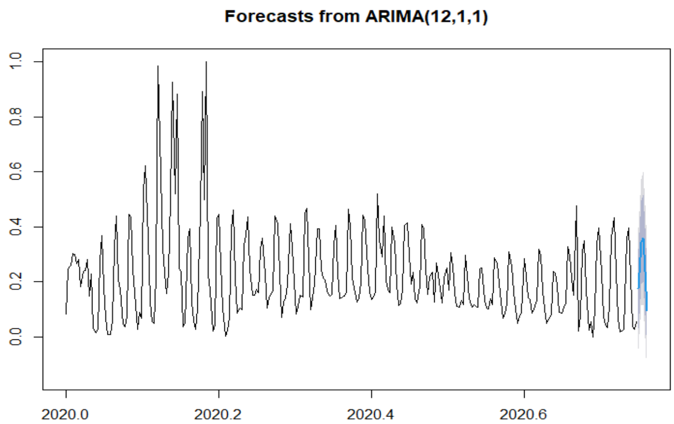

Exhibit 5

Since the forecasted values exhibit a continuation of the previous patterns, the business understands that the underlying revenue trend is expected to remain consistent in the near future. This business can more confidently use this insight to assess the overall revenue trajectory and make informed decisions regarding financial planning, budgeting and set their future revenue targets.

The 5-step forecast also highlights the inherent variability in daily revenue, which is crucial for businesses to assess and manage their financial risks. By considering the forecasted range and potential revenue fluctuations, this business can implement appropriate risk management such as maintaining cash reserves or diversifying their revenue streams to mitigate the impact of low revenue or unexpected revenue changes even though the trend is relatively steady. You can never be too careful.

The forecasted values can also be compared to actual revenue data to evaluate the accuracy of the model and identify any discrepancies. This evaluation will allow the business to assess its effectiveness of the forecasting methods and make necessary adjustments for future predictions. Additionally, comparing the forecasted revenue with the actual performance can help identify potential gaps and opportunities for improving revenue generation strategies.

Predictive Modeling and Evaluation:

Logistic Regression:

To analyze transaction predictions, the daily data was transformed by assigning a value of 1 if a transaction occurred and 0 if there was no transaction. The date variable was removed as it was not considered crucial for the analysis, except during specific shopping events like Black Friday. The transaction variable, which was used to create the target variable indicating the occurrence of a transaction, was also eliminated. Additionally, the revenue variable, which is zero when there is no transaction, was excluded.

Three predictive models were developed: a logistic regression model and two decision tree models. The logistic regression model initially utilized all available predictors and was then refined by selecting only the significant ones. The decision tree models provide visual representation of the decision rules used for predicting transactions, and in the end, the challenger model with the higher accuracy rate will be selected.

The logistic regression model for predicting transactions yielded an accuracy of 86.05% on the test set, indicating a high level of predictive performance. Referencing Exhibit 6, the area under the ROC curve (AUC) was calculated to be 0.93, further supporting the effectiveness of the model. A higher AUC signifies better discrimination between positive and negative cases. In this case, the AUC of 0.93 indicates a strong ability to differentiate between transactions and non-transactions.

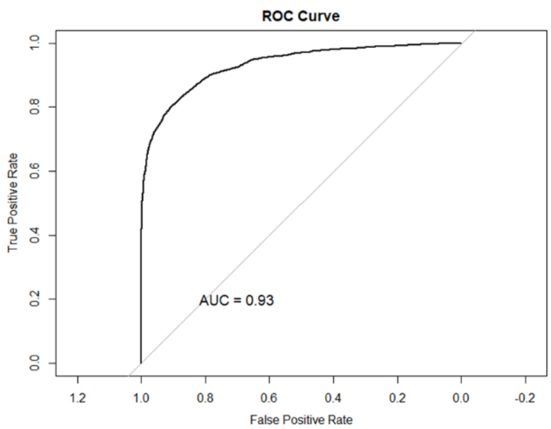

Exhibit 6

Several variables stood out as significant predictors of transactions from the coefficients of the logistic regression model. Notably, the variables related to user activity such as visits, productClick, addToCart, and checkout were prominent factors influencing transactions. Additionally, the source of user traffic, including platforms like eLama, Facebook, and Google, demonstrated significance. Investing in these platforms may lead to increased transaction rates. Moreover, the "delivery_available" variable showed that customers tend to be more inclined to make a purchase when free shipping is not available, which is an interesting finding, and could be better explored through Neuromarketing and some A/B testing. Finally, the "mediumcpc" variable indicated that increasing investment in cost-per-click advertising could potentially yield positive results.

Decision Tree 1:

The decision tree model constructed for the classification problem revealed several key variables that play a significant role in predicting transactions. The main variables identified in the decision tree model were checkout, visits, productClick, and pageviews. These variables provide valuable insights into user behavior and engagement, indicating that the likelihood of a transaction is influenced by factors such as the number of checkouts, the number of visits, the occurrence of product clicks, and the number of pageviews. By analyzing these variables, this business can gain a better understanding of customer actions and tailor their strategies accordingly. Graphically, the decision tree (Exhibit 7) plot provides a visual representation of the decision rules, allowing stakeholders to interpret the classification process and identify the most important variables affecting transactions.

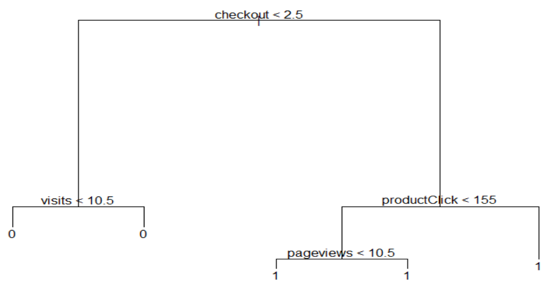

Exhibit 7

When evaluating the performance of the decision tree model, the accuracy rate was found to be 86.77% on the test set, showing a slight improvement compared to the logistic regression model. This suggests that the decision tree model effectively captures patterns and rules that differentiate transactions from non-transactions and being more visible. Furthermore, the area under the ROC curve (AUC) was calculated to be 0.8 (Exhibit 8), indicating the model's ability to distinguish between positive and negative cases. The ROC curve plot visualizes the trade-off between true positive rate and false positive rate, while the AUC value provides a measure of the model's overall performance. Although the AUC of 0.8 is lower than the logistic regression model’s, it still indicates a good discriminatory power. By considering the decision tree model's accuracy and AUC, businesses can gain confidence in its ability to predict transaction outcomes and make informed decisions based on the model's insights.

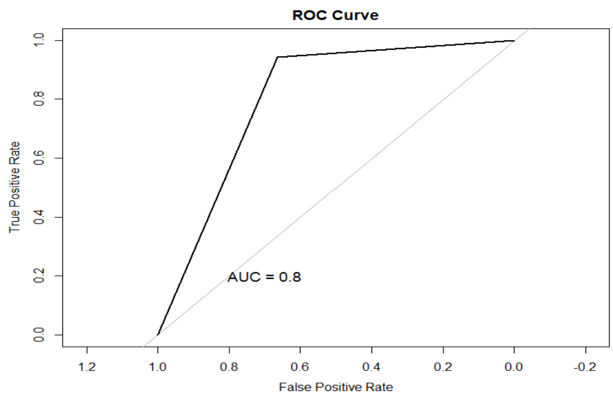

Exhibit 8

Decision Tree 2 (Exhibit 9):

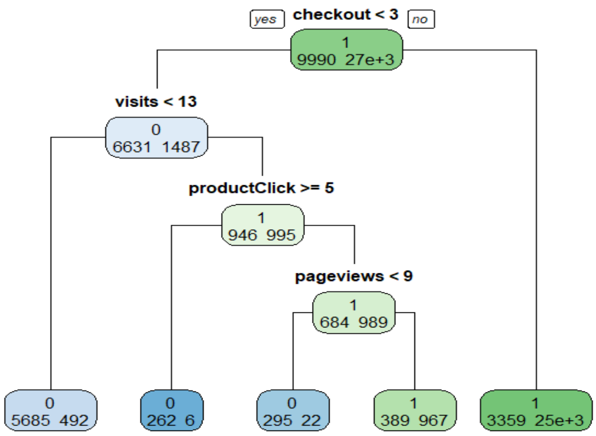

Exhibit 9

For the second decision tree, the checkout, visits, productClick, and pageviews, variables are identified as important. The importance of these variables shows the relative importance of each variable in predicting transactions. This information can guide businesses in understanding the key factors driving transactions and assist in decision-making. The performance of this model is evaluated using the confusion matrix, and the result is 83.69%, which is not bad, but comparing to the two previous models, falls a bit short.

While the second decision tree model identified the variables checkout, visits, productClick, and pageviews as important predictors of transactions, its performance of 83.69% falls slightly short when compared to the logistic regression and the first decision tree model. However, it is important to note that the first decision tree model encountered challenges during the training process due to the presence of NA values, as indicated by the warning message. Despite these challenges, the first decision tree model outperformed the other two challenger models in terms of accuracy, achieving an accuracy rate of 86.77%. Through optimization of this tree, this business can allocate their resources more efficiently, tailor their marketing efforts, and enhance customer experiences to maximize the likelihood of transactions.

Conclusion:

This essay conducted a dual-pronged analysis of a sales dataset using time series forecasting and predictive modeling. The time series analysis, despite its stationary nature and frequent fluctuations, suggested room for improvement. Predictive modeling, using logistic regression and decision trees, identified key predictors of transactions, such as checkout, visits, productClick, and pageviews. While the logistic regression model demonstrated good discriminatory power, the first decision tree model's performance was slightly better. These findings provide valuable insights for businesses to optimize operations and boost profitability, underscoring the importance of data-driven decision-making.

Appendix: R code & output

```
> library(tidyverse)
── Attaching core tidyverse packages ────────────────────────────
──── tidyverse 2.0.0 ──
✔ dplyr     1.1.2     ✔ readr     2.1.4
✔ forcats   1.0.0     ✔ stringr   1.5.0
✔ ggplot2   3.4.2     ✔ tibble    3.2.1
✔ lubridate 1.9.2     ✔ tidyr     1.3.0
✔ purrr     1.0.1
── Conflicts ────────────────────────────────────────────
tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package to force all conflicts to become errors
> library(forecast)
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
> library(caTools)
> library(pROC)
Type 'citation("pROC")' for a citation.

载入程辑包：‘pROC’

The following objects are masked from ‘package:stats’:

    cov, smooth, var

> library(texreg)
Version:  1.38.6
Date:     2022-04-06
Author:   Philip Leifeld (University of Essex)

Consider submitting praise using the praise or praise_interactive func
tions.
Please cite the JSS article in your publications -- see citation("texr
eg").

载入程辑包：‘texreg’

The following object is masked from ‘package:tidyr’:

    extract
```

```
> library(tree)
> library(pROC)
> library(rpart)
> library(rpart.plot)
> library(caret)
载入需要的程辑包：lattice

载入程辑包：‘caret’

The following object is masked from ‘package:purrr’:

    lift

> library(ggplot2)
> library(tseries)

    ‘tseries’ version: 0.10-54

    ‘tseries’ is a package for time series analysis and computational f
inance.

    See ‘library(help="tseries")’ for details.

> library(rugarch)
载入需要的程辑包：parallel

载入程辑包：‘rugarch’

The following object is masked from ‘package:purrr’:

    reduce

The following object is masked from ‘package:stats’:

    sigma

> library(dplyr)
> library(corrplot)
corrplot 0.92 loaded
>
> sales <- read.csv("C:/Users/yanji/OneDrive/桌面/sales.csv")
> head(sales)
```

```
       date source  medium delivery_available device_type promo_activat
ed filter_used
1 2020/5/11 google organic            no data          PC             no
       no
2 2020/5/11 yandex     cpc            no data      mobile            yes
        no
3 2020/5/11 google     cpc            no data      mobile             no
       no
4 2020/5/11 google     cpc            no data          PC             no
       no
5 2020/5/11 yandex organic            no data          PC             no
       no
6 2020/5/11 yandex     cpc            no data          PC             no
       no
  pageviews visits productClick addToCart checkout transactions revenu
e
1      3120   1233         5240      1048      525           90   86649
2      3302    544         9930      1984     1416          217  244478
3      2970   1450         5460      1090      599          100  105150
4      1875    854         4250       848      407           71   79003
5      2159   1000         4110       824      351           62   61861
6      2775   1441         5990      1196      549           76   90862
> 
> 
> ####ANNALYSIS####
> sales$transaction <- ifelse(sales$transactions>0, 1, 0)
> 
> sub_sales <- sales %>% select(-c(date, transactions, revenue))
> 
> sum(is.na(sub_sales))
[1] 0
> 
> ##Logistic Regression##
> set.seed(202306)
> 
> split <- sample.split(sub_sales$transaction, SplitRatio = 0.7)
> train <- subset(sub_sales, split == TRUE)
> test <- subset(sub_sales, split == FALSE)
> 
> model <- glm(transaction ~., data = train, family = "binomial")
Warning message:
glm.fit:拟合機率算出来是数值零或一
> predictions <- predict(model, newdata = test, type = "response")
Warning message:
```

```
In predict.lm(object, newdata, se.fit, scale = 1, type = if (type ==  :
  prediction from rank-deficient fit; attr(*, "non-estim") has doubtfu
l cases
> predicted_classes <- ifelse(predictions > 0.5, 1, 0)
>
> accuracy <- sum(predicted_classes == test$transaction) / length(test
$transaction)
> accuracy
[1] 0.860521
>
> roc_obj <- roc(test$transaction, predictions)
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> plot(roc_obj, main = "ROC Curve", xlab = "False Positive Rate", ylab
 = "True Positive Rate")
>
> auc <- auc(roc_obj)
> text(0.7, 0.2, paste0("AUC = ", round(auc, 2)), cex = 1.2)
>
> screenreg(model)

=========================================
                        Model 1
-----------------------------------------
(Intercept)                 -1.37 ***
                           (0.08)
sourceactionpay             -0.48
                           (0.38)
sourceadmitad                0.85 *
                           (0.36)
sourceadvertise              0.25
                           (0.39)
sourcebaidu                -13.50
                         (294.25)
sourcebing                  -0.25
                           (0.25)
sourcecityads                0.62
                           (0.36)
sourceco-promo               0.37
                           (0.37)
sourceDuckDuckGo             0.09
                           (0.25)
sourceeLama                  0.85 ***
                           (0.23)
```

sourceexponea	-0.71
	(0.75)
sourcefacebook	-2.70 ***
	(0.15)
sourcegoogle	0.83 ***
	(0.23)
sourceinstagram	-3.01 ***
	(0.40)
sourcemytarget	-0.93 ***
	(0.23)
sourcenewsletter	0.31
	(0.35)
sourceopmcpa	0.47
	(0.38)
sourceother	-0.19
	(0.24)
sourcepromo	0.59
	(0.36)
sourcesailplay	-0.37
	(0.56)
sourcevk	-1.47 ***
	(0.12)
sourceyandex	0.79 ***
	(0.23)
sourceyandex_direct	-2.74 ***
	(0.53)
sourceyoutube	-13.57
	(105.52)
mediumcpa	-0.61
	(0.35)
mediumcpc	-0.61 **
	(0.21)
mediumemail	-0.80 *
	(0.34)
mediumorganic	-0.43 *
	(0.21)
delivery_availableno data	0.73 ***
	(0.07)
delivery_availableyes	2.59 ***
	(0.07)
device_typeno data	-0.14 *
	(0.07)
device_typePC	0.37 ***
	(0.04)

```
promo_activatedyes             -0.01
                              (0.03)
filter_usedyes                 -0.60 ***
                              (0.05)
pageviews                      -0.00
                              (0.00)
visits                          0.01 ***
                              (0.00)
productClick                    0.08 ***
                              (0.02)
addToCart                      -0.30 **
                              (0.11)
checkout                        0.05 ***
                              (0.00)
----------------------------------------
AIC                         22160.24
BIC                         22492.36
Log Likelihood              -11041.12
Deviance                     22082.24
Num. obs.                    36905
========================================
*** p < 0.001; ** p < 0.01; * p < 0.05
> step(model)
Start:  AIC=22160.24
transaction ~ source + medium + delivery_available + device_type + 
    promo_activated + filter_used + pageviews + visits + productClick + 

    addToCart + checkout

                   Df Deviance   AIC
- promo_activated     1    22082 22158
- pageviews          1    22084 22160
<none>                    22082 22160
- addToCart          1    22091 22167
- medium             4    22103 22173
- productClick        1    22097 22173
- visits             1    22202 22278
- device_type         2    22206 22280
- filter_used         1    22221 22297
- checkout           1    22270 22346
- source            22    22734 22768
- delivery_available  2    24244 24318

Step:  AIC=22158.4
```

```
transaction ~ source + medium + delivery_available + device_type +
    filter_used + pageviews + visits + productClick + addToCart +
    checkout

                    Df Deviance   AIC
- pageviews          1    22084 22158
<none>                    22082 22158
- addToCart          1    22092 22166
- medium             4    22103 22171
- productClick       1    22097 22171
- visits             1    22202 22276
- device_type        2    22206 22278
- filter_used        1    22221 22295
- checkout           1    22271 22345
- source            22    22734 22766
- delivery_available  2    24244 24316

Step:  AIC=22158.11
transaction ~ source + medium + delivery_available + device_type +
    filter_used + visits + productClick + addToCart + checkout

                    Df Deviance   AIC
<none>                    22084 22158
- addToCart          1    22093 22165
- productClick       1    22099 22171
- medium             4    22105 22171
- device_type        2    22208 22278
- filter_used        1    22224 22296
- checkout           1    22272 22344
- visits             1    22488 22560
- source            22    22744 22774
- delivery_available  2    24248 24318

Call:  glm(formula = transaction ~ source + medium + delivery_availabl
e +
    device_type + filter_used + visits + productClick + addToCart +
    checkout, family = "binomial", data = train)

Coefficients:
             (Intercept)           sourceactionpay             sourceadm
itad
               -1.369604                 -0.510319                0.8608
24
```

```
        sourceadvertise              sourcebaidu              sourceb
ing
              0.258126               -13.491653               -0.2423
30
          sourcecityads             sourceco-promo         sourceDuckDu
ckGo
              0.617156                 0.370879               0.08792
0
            sourceeLama              sourceexponea          sourceface
book
              0.852938                -0.702757              -2.7041
36
           sourcegoogle            sourceinstagram           sourcemyta
rget
              0.826394                -3.009434              -0.9319
15
       sourcenewsletter              sourceopmcpa              sourceo
ther
              0.316089                 0.484170             -0.18298
4
            sourcepromo             sourcesailplay                sourc
evk
              0.591877                -0.365022              -1.4757
37
           sourceyandex     sourceyandex_direct              sourceyo
utube
              0.793667                -2.748956             -13.5708
78
              mediumcpa                  mediumcpc            mediumema
il
             -0.623038                -0.621510              -0.8105
76
          mediumorganic               mediumsocial  delivery_availablen
o data
             -0.442604                       NA             0.724884

   delivery_availableyes        device_typeno data              device_
typePC
              2.591014                -0.135169               0.37574
5
          filter_usedyes                     visits           productCl
ick
             -0.600403                 0.006108               0.07541
3
```

```
                 addToCart                checkout
                 -0.304322                0.044266

Degrees of Freedom: 36904 Total (i.e. Null);  36868 Residual
Null Deviance:      43100
Residual Deviance: 22080        AIC: 22160
There were 32 warnings (use warnings() to see them)
> model2 <- glm(formula = transaction ~ source + medium + delivery_ava
ilable +
+                device_type + filter_used + visits + productClick + add
ToCart +
+                checkout, family = "binomial", data = train)
Warning message:
glm.fit:拟合機率算出来是数值零或一
>
> predictions <- predict(model2, newdata = test, type = "response")
Warning message:
In predict.lm(object, newdata, se.fit, scale = 1, type = if (type ==  :
  prediction from rank-deficient fit; attr(*, "non-estim") has doubtfu
l cases
> predicted_classes <- ifelse(predictions > 0.5, 1, 0)
> accuracy <- sum(predicted_classes == test$transaction) / length(test
$transaction)
> accuracy
[1] 0.8607107
> roc_obj <- roc(test$transaction, predictions)
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> plot(roc_obj, main = "ROC Curve", xlab = "False Positive Rate", ylab
 = "True Positive Rate")
>
> auc <- auc(roc_obj)
> text(0.7, 0.2, paste0("AUC = ", round(auc, 2)), cex = 1.2)
>
> ##Decision Tree##
> model <- tree(factor(transaction) ~., data = train)
Warning message:
In tree(factor(transaction) ~ ., data = train) : NAs introduced by coe
rcion
> plot(model)
> text(model, pretty = 0)
> predictions <- predict(model, newdata = test, type = "class")
Warning message:
```

```
In pred1.tree(object, tree.matrix(newdata)) : NAs introduced by coerci
on
> accuracy <- sum(predictions == test$transaction) / length(test$trans
action)
> accuracy
[1] 0.8676657
>
> auc <- roc(test$transaction, as.numeric(predictions))$auc
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> roc_obj <- roc(test$transaction, as.numeric(predictions))
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> plot(roc_obj, main = "ROC Curve", xlab = "False Positive Rate", ylab
 = "True Positive Rate")
> text(0.7, 0.2, paste0("AUC = ", round(auc, 2)), cex = 1.2)
>
> ##Tree 2##
> # 'transaction' is business success/failure
> which(sub_sales$transaction==1)
   [1]    1    2    3    4    5    6    7    8    9   10   11   12   13   14
 15   16   17
  [18]   18   19   20   21   22   23   24   25   26   27   28   29   30   31
  32   33   34
  [35]   35   36   37   38   39   40   41   42   43   44   45   46   47   48
  49   50   51
  [52]   52   53   54   55   56   57   58   59   60   61   62   63   64   65
  66   67   68
  [69]   69   70   71   72   73   74   75   76   77   78   79   80   81   82
  83   84   85
  [86]   86   87   88   89   90   91   92   93   94   95   96   97   98   99
  100  101  102
 [103]  103  104  105  106  107  108  109  110  111  112  113  114  115
116  117  118  119
 [120]  120  121  122  123  124  125  126  127  128  129  130  131  132
133  134  135  136
 [137]  137  138  139  140  141  142  143  144  145  146  147  148  149
150  151  152  153
 [154]  154  155  156  157  158  159  160  161  162  163  164  165  166
167  168  169  170
 [171]  171  172  173  174  175  176  177  178  179  180  181  182  183
184  185  186  187
 [188]  188  189  190  191  192  193  194  195  196  197  198  199  200
201  202  203  204
```

[205] 205 207 208 209 210 211 212 213 214 215 216 217 218
219 220 221 222
[222] 223 224 225 226 227 228 229 230 231 232 233 234 235
236 237 238 239
[239] 240 241 242 243 244 245 246 247 248 249 250 251 252
253 254 255 256
[256] 257 258 259 260 261 262 263 264 265 266 267 268 269
270 271 272 273
[273] 274 275 276 277 278 279 280 281 282 283 284 285 286
287 288 289 290
[290] 291 292 293 294 295 296 297 298 299 300 301 302 304
305 306 307 308
[307] 309 310 311 312 313 314 315 316 317 318 319 320 321
322 323 324 325
[324] 326 327 328 329 330 331 332 333 334 335 336 337 338
339 340 341 342
[341] 343 344 345 346 347 348 349 350 351 352 353 354 355
356 357 358 359
[358] 360 361 362 363 364 365 366 367 368 369 370 371 372
373 374 375 376
[375] 377 378 379 380 381 382 383 384 385 386 387 388 389
390 391 392 393
[392] 394 395 396 397 398 399 400 401 402 403 404 405 406
407 408 409 410
[409] 411 412 413 414 415 416 417 418 419 420 421 422 423
424 425 426 427
[426] 428 429 430 431 432 433 434 435 436 437 438 439 440
441 442 443 444
[443] 445 446 447 448 449 450 451 452 453 454 455 456 457
458 459 460 461
[460] 462 463 464 465 466 467 468 469 470 471 472 473 474
475 476 477 478
[477] 479 480 481 482 483 484 485 486 487 488 489 490 491
492 493 494 495
[494] 496 497 498 499 500 501 502 503 504 505 506 507 508
509 510 511 512
[511] 513 514 515 516 517 518 519 520 521 522 523 524 525
526 527 528 529
[528] 530 531 532 533 534 535 536 537 538 539 540 541 542
543 544 545 546
[545] 547 548 549 550 551 552 553 554 555 556 557 558 559
560 561 562 563
[562] 564 565 566 567 568 569 570 571 572 573 574 575 576
577 578 579 580

[579] 581 582 583 584 585 586 587 588 589 590 591 592 593
594 595 596 597
[596] 598 599 600 601 602 603 604 605 606 607 608 609 610
611 612 613 614
[613] 615 616 617 618 619 620 621 622 623 624 625 626 627
628 629 630 631
[630] 632 633 634 635 636 637 638 639 640 641 642 643 644
645 646 647 648
[647] 649 650 651 652 653 654 655 656 657 658 659 660 661
662 663 664 665
[664] 666 667 668 669 670 671 672 673 674 675 676 677 678
679 680 681 682
[681] 683 684 685 686 687 688 689 690 691 692 693 694 695
696 697 698 699
[698] 700 701 702 703 704 705 706 707 708 709 710 711 712
713 714 715 716
[715] 717 718 719 720 721 722 723 724 725 726 727 728 729
730 731 732 733
[732] 734 735 736 737 738 739 740 741 742 743 744 745 746
747 748 749 750
[749] 751 752 753 754 755 756 757 758 759 760 761 762 763
764 765 766 767
[766] 768 769 770 771 772 773 774 775 776 777 778 779 780
781 782 783 784
[783] 785 786 787 788 789 790 791 792 793 794 795 796 797
798 799 800 801
[800] 802 803 804 805 806 807 808 809 810 811 812 813 814
815 816 817 818
[817] 819 820 821 822 823 824 825 826 827 828 829 830 831
832 833 834 835
[834] 836 837 838 839 840 841 842 843 844 845 846 847 848
849 850 851 852
[851] 853 854 855 856 857 858 859 860 861 862 863 864 865
866 867 868 869
[868] 870 871 872 873 874 875 876 877 878 879 880 881 882
883 884 885 886
[885] 887 888 889 890 891 892 893 894 895 896 897 898 899
900 901 902 903
[902] 904 905 906 907 908 909 910 911 912 913 914 915 916
917 918 919 920
[919] 921 922 923 924 925 926 927 928 929 930 931 932 933
934 935 936 937
[936] 938 939 940 941 942 943 944 945 946 947 948 949 950
951 952 953 954

```
 [953]  955  956  957  958  959  960  961  962  963  964  965  966  967
968  969  970  971
 [970]  972  973  974  975  976  977  978  979  980  981  982  983  984
985  986  987  988
 [987]  989  990  991  992  993  994  995  996  997  998  999 1000 1001 1
002
 [ reached getOption("max.print") -- omitted 37450 entries ]
> business_success <- sub_sales[which(sub_sales$transaction==1),]
> summary(business_success)
    source             medium          delivery_available device_type

 Length:38450       Length:38450       Length:38450       Length:38450

 Class :character   Class :character   Class :character   Class :charac
ter
 Mode  :character   Mode  :character   Mode  :character   Mode  :charac
ter



 promo_activated    filter_used          pageviews           visits
 productClick
 Length:38450       Length:38450       Min.   :    0.0   Min.   :   1.0
Min.   :    0
 Class :character   Class :character   1st Qu.:   29.0   1st Qu.:    3.0
  1st Qu.:   60
 Mode  :character   Mode  :character   Median :  103.0   Median :  15.0
  Median :  220
                                       Mean   :  603.5   Mean   : 170.7   Mean
   : 1213
                                       3rd Qu.:  467.0   3rd Qu.:  72.0   3rd
 Qu.:  980
                                       Max.   :26589.0   Max.   :6975.0   Ma
x.   :32460
   addToCart          checkout        transaction
 Min.   :   0.0   Min.   :   0.0   Min.   :1
 1st Qu.:   12.0   1st Qu.:    9.0   1st Qu.:1
 Median :   44.0   Median :  35.0   Median :1
 Mean   : 242.4   Mean   : 199.6   Mean   :1
 3rd Qu.: 196.0   3rd Qu.: 161.0   3rd Qu.:1
 Max.   :6486.0   Max.   :3857.0   Max.   :1
>
> which(sub_sales$transaction==0)
```

```
  [1]   206   303  2242  2269  5061  6514  6603  7537  8547  9744 10574
11469 11793 12045
 [15] 12375 12480 12616 12686 13833 13996 14495 14649 15121 15897 1603
6 16148 16180 16467
 [29] 16603 16646 16858 16993 17129 17198 17317 17348 17351 17488 1749
1 17532 17537 17564
 [43] 17586 17597 17598 17599 17600 17602 17606 17610 17612 17615 1761
9 17627 17639 17641
 [57] 17642 17645 17650 17654 17657 17664 17665 17667 17669 17670 1767
1 17673 17675 17677
 [71] 17680 17681 17688 17689 17693 17696 17697 17702 17704 17705 1770
7 17709 17712 17713
 [85] 17717 17719 17720 17723 17725 17728 17729 17730 17731 17734 1773
5 17736 17741 17744
 [99] 17746 17749 17751 17753 17754 17755 17758 17759 17761 17764 1776
5 17766 17767 17771
[113] 17773 17777 17778 17779 17781 17783 17784 17786 17788 17790 177
93 17794 17796 17797
[127] 17798 17802 17803 17810 17816 17817 17821 17828 17832 17834 178
37 17842 17843 17847
[141] 17857 17858 17865 17871 17876 17878 17880 17884 17889 17890 178
96 17906 17908 17910
[155] 17913 17916 17918 17919 17921 17924 17925 17926 17929 17930 179
32 17934 17935 17936
[169] 17937 17942 17944 17945 17948 17955 17958 17959 17964 17970 179
77 17979 17980 17997
[183] 18000 18009 18016 18017 18022 18033 18036 18039 18043 18044 180
50 18054 18061 18062
[197] 18065 18075 18082 18086 18091 18092 18095 18097 18100 18105 181
10 18111 18112 18114
[211] 18115 18118 18125 18131 18136 18137 18141 18142 18145 18147 181
48 18149 18151 18155
[225] 18158 18161 18162 18165 18166 18167 18169 18170 18172 18173 181
76 18178 18180 18183
[239] 18186 18187 18193 18196 18197 18198 18200 18201 18204 18206 182
13 18214 18215 18227
[253] 18228 18232 18236 18237 18238 18240 18241 18243 18245 18248 182
50 18254 18255 18257
[267] 18258 18262 18264 18267 18272 18275 18276 18279 18283 18293 182
95 18298 18302 18308
[281] 18310 18314 18316 18318 18321 18322 18327 18329 18330 18334 183
38 18339 18345 18347
[295] 18353 18357 18358 18359 18360 18363 18365 18366 18367 18368 183
69 18376 18379 18380
```

[309] 18383 18385 18386 18387 18388 18389 18390 18394 18395 18399 184
00 18403 18404 18405
[323] 18406 18407 18408 18409 18412 18413 18418 18421 18422 18424 184
26 18429 18431 18433
[337] 18434 18437 18438 18439 18442 18443 18444 18445 18449 18451 184
58 18459 18461 18462
[351] 18463 18465 18466 18468 18472 18473 18476 18477 18478 18480 184
81 18482 18483 18484
[365] 18487 18488 18492 18493 18494 18495 18496 18497 18500 18501 185
03 18505 18506 18507
[379] 18508 18509 18511 18512 18514 18515 18517 18518 18521 18523 185
26 18527 18529 18531
[393] 18532 18536 18537 18538 18539 18540 18545 18547 18549 18551 185
53 18554 18555 18558
[407] 18559 18561 18562 18566 18567 18568 18572 18573 18574 18575 185
80 18587 18588 18589
[421] 18591 18592 18593 18598 18599 18603 18605 18606 18608 18609 186
10 18611 18613 18615
[435] 18616 18618 18619 18620 18624 18627 18628 18629 18630 18631 186
33 18634 18637 18640
[449] 18641 18642 18643 18644 18648 18649 18650 18653 18654 18655 186
57 18658 18659 18660
[463] 18661 18662 18665 18666 18668 18671 18680 18683 18685 18686 186
92 18694 18695 18706
[477] 18708 18710 18712 18719 18722 18723 18730 18731 18732 18734 187
37 18738 18740 18747
[491] 18748 18749 18754 18755 18756 18757 18760 18761 18762 18765 187
66 18768 18774 18775
[505] 18776 18778 18779 18781 18783 18784 18785 18786 18788 18789 187
90 18792 18794 18795
[519] 18796 18797 18798 18799 18801 18802 18803 18805 18806 18811 188
13 18817 18818 18819
[533] 18820 18824 18825 18826 18827 18828 18830 18831 18835 18837 188
38 18840 18841 18842
[547] 18843 18844 18846 18847 18848 18851 18852 18853 18855 18858 188
59 18860 18863 18864
[561] 18865 18867 18872 18873 18875 18876 18877 18878 18882 18886 188
91 18892 18893 18897
[575] 18898 18899 18903 18904 18905 18906 18907 18910 18911 18912 189
13 18914 18917 18918
[589] 18919 18920 18923 18924 18926 18928 18930 18931 18934 18936 189
37 18939 18940 18941
[603] 18942 18946 18947 18949 18955 18956 18957 18960 18961 18962 189
63 18966 18967 18968

[617] 18969 18970 18971 18973 18975 18979 18980 18981 18982 18983 189
85 18986 18990 18992
[631] 18995 18998 18999 19001 19004 19006 19007 19012 19017 19018 190
19 19025 19029 19030
[645] 19035 19036 19038 19040 19041 19043 19045 19047 19048 19049 190
51 19052 19053 19054
[659] 19055 19056 19057 19058 19063 19064 19065 19066 19067 19068 190
75 19077 19079 19080
[673] 19082 19083 19084 19085 19087 19089 19090 19094 19095 19096 190
97 19098 19099 19101
[687] 19103 19105 19106 19109 19110 19111 19112 19113 19114 19115 191
16 19121 19122 19123
[701] 19124 19125 19126 19129 19130 19132 19133 19135 19138 19143 191
47 19148 19149 19150
[715] 19151 19152 19153 19156 19157 19159 19161 19162 19163 19165 191
67 19169 19170 19173
[729] 19175 19177 19178 19179 19181 19183 19184 19187 19190 19192 191
94 19195 19198 19199
[743] 19203 19204 19206 19207 19211 19212 19213 19215 19217 19220 192
23 19224 19225 19227
[757] 19228 19229 19231 19232 19234 19235 19237 19241 19242 19243 192
44 19245 19246 19247
[771] 19248 19249 19250 19251 19253 19255 19257 19258 19259 19260 192
62 19263 19266 19267
[785] 19268 19270 19272 19273 19275 19276 19277 19278 19279 19280 192
82 19283 19286 19290
[799] 19291 19293 19294 19299 19300 19302 19303 19304 19306 19308 193
09 19310 19314 19315
[813] 19316 19317 19320 19322 19325 19326 19330 19332 19335 19337 193
39 19340 19341 19344
[827] 19346 19347 19348 19351 19352 19353 19355 19356 19358 19363 193
64 19367 19369 19370
[841] 19371 19372 19373 19377 19378 19380 19381 19382 19384 19386 193
87 19390 19393 19394
[855] 19401 19404 19405 19406 19409 19410 19411 19412 19416 19419 194
20 19421 19422 19424
[869] 19425 19426 19427 19429 19431 19432 19433 19434 19436 19438 194
40 19441 19442 19444
[883] 19446 19448 19449 19450 19451 19452 19453 19454 19455 19456 194
57 19458 19459 19461
[897] 19464 19467 19468 19469 19470 19471 19473 19475 19476 19478 194
79 19480 19482 19484
[911] 19486 19488 19489 19492 19493 19495 19497 19498 19500 19506 195
09 19512 19513 19517

```
 [925] 19518 19519 19521 19522 19523 19527 19528 19529 19530 19531 195
34 19535 19536 19538
 [939] 19539 19540 19541 19542 19543 19544 19546 19547 19549 19553 195
54 19556 19558 19560
 [953] 19561 19562 19563 19567 19568 19571 19572 19573 19574 19575 195
76 19578 19579 19581
 [967] 19582 19583 19584 19585 19587 19589 19591 19592 19593 19594 195
96 19597 19599 19602
 [981] 19603 19605 19607 19608 19609 19610 19611 19613 19614 19615 196
16 19618 19619 19621
 [995] 19622 19623 19625 19626 19627 19628
 [ reached getOption("max.print") -- omitted 13271 entries ]
> business_failure <- sub_sales[which(sub_sales$transaction==0),]
> summary(business_failure)
    source             medium          delivery_available device_type

 Length:14271       Length:14271       Length:14271       Length:14271

 Class :character   Class :character   Class :character   Class :charac
ter
 Mode  :character   Mode  :character   Mode  :character   Mode  :charac
ter



 promo_activated    filter_used          pageviews           visits
   productClick
 Length:14271       Length:14271       Min.   :   0.00   Min.   :   1.000
  Min.   :  0.00
 Class :character   Class :character   1st Qu.:   2.00   1st Qu.:   1.00
0   1st Qu.:  0.00
 Mode  :character   Mode  :character   Median :   6.00   Median :   2.00
0   Median :  0.00
                                       Mean   :  20.17   Mean   :   9.686   Me
an   : 21.35
                                       3rd Qu.:  15.00   3rd Qu.:   5.000   3
rd Qu.: 30.00
                                       Max.   :1894.00   Max.   :1057.000   M
ax.   :930.00
   addToCart          checkout        transaction
 Min.   : 0.000   Min.   : 0.000   Min.   :0
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:0
 Median : 0.000   Median : 0.000   Median :0
 Mean   : 4.269   Mean   : 3.528   Mean   :0
```

```
 3rd Qu.:  6.000   3rd Qu.:  4.000   3rd Qu.:0
 Max.   :186.000   Max.   :259.000   Max.   :0
>
> logit <- glm(transaction ~ pageviews + visits + productClick + addTo
Cart + checkout, data=train, family="binomial")
Warning message:
glm.fit:拟合機率算出来是数值零或一
> summary(logit)

Call:
glm(formula = transaction ~ pageviews + visits + productClick +
    addToCart + checkout, family = "binomial", data = train)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8105695  0.0206170 -39.316  < 2e-16 ***
pageviews    -0.0013006  0.0004278  -3.040  0.00236 **
visits        0.0058532  0.0006022   9.720  < 2e-16 ***
productClick  0.0816004  0.0201922   4.041 5.32e-05 ***
addToCart    -0.3446355  0.1009423  -3.414  0.00064 ***
checkout      0.0799951  0.0032769  24.412  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43101  on 36904  degrees of freedom
Residual deviance: 25710  on 36899  degrees of freedom
AIC: 25722

Number of Fisher Scoring iterations: 11

>
> tree <- rpart(transaction ~ pageviews + visits + productClick + addT
oCart + checkout, data = train, method = "class")
> summary(tree)
Call:
rpart(formula = transaction ~ pageviews + visits + productClick +
    addToCart + checkout, data = train, method = "class")
  n= 36905

          CP nsplit rel error    xerror        xstd
1 0.51491491      0 1.0000000 1.0000000 0.008544209
2 0.01928595      1 0.4850851 0.4850851 0.006494693
```

```
3 0.01000000      4 0.4272272 0.4278278 0.006153530

Variable importance
    checkout productClick    addToCart    pageviews       visits 
          31           24           24           18            3 

Node number 1: 36905 observations,    complexity param=0.5149149
  predicted class=1  expected loss=0.270695  P(node) =1
    class counts:  9990 26915
   probabilities: 0.271 0.729 
  left son=2 (8118 obs) right son=3 (28787 obs)
  Primary splits:
      checkout     < 2.5   to the left,  improve=6208.158, (0 missing)
      pageviews    < 10.5  to the left,  improve=5428.159, (0 missing)
      productClick < 25    to the left,  improve=4969.168, (0 missing)
      addToCart    < 5     to the left,  improve=4969.029, (0 missing)
      visits       < 3.5   to the left,  improve=2100.584, (0 missing)
  Surrogate splits:
      productClick < 5     to the left,  agree=0.952, adj=0.784, (0 spli
t)
      addToCart    < 1     to the left,  agree=0.952, adj=0.783, (0 spli
t)
      pageviews    < 5.5   to the left,  agree=0.888, adj=0.492, (0 spli
t)

Node number 2: 8118 observations,    complexity param=0.01928595
  predicted class=0  expected loss=0.1831732  P(node) =0.2199702
    class counts:  6631  1487
   probabilities: 0.817 0.183 
  left son=4 (6177 obs) right son=5 (1941 obs)
  Primary splits:
      visits       < 12.5  to the left,  improve=553.73730, (0 missing)
      pageviews    < 20.5  to the left,  improve=528.81440, (0 missing)
      addToCart    < 1     to the right, improve= 88.01460, (0 missing)
      productClick < 5     to the right, improve= 84.87056, (0 missing)
      checkout     < 0.5   to the right, improve= 18.69954, (0 missing)
  Surrogate splits:
      pageviews < 18.5  to the left,  agree=0.913, adj=0.635, (0 split)

Node number 3: 28787 observations
  predicted class=1  expected loss=0.1166846  P(node) =0.7800298
    class counts:  3359 25428
   probabilities: 0.117 0.883 
```

```
Node number 4: 6177 observations
  predicted class=0  expected loss=0.07965032  P(node) =0.1673757
    class counts:  5685   492
   probabilities: 0.920 0.080

Node number 5: 1941 observations,    complexity param=0.01928595
  predicted class=1  expected loss=0.4873776  P(node) =0.0525945
    class counts:   946   995
   probabilities: 0.487 0.513
  left son=10 (268 obs) right son=11 (1673 obs)
  Primary splits:
      productClick < 5     to the right, improve=149.45200, (0 missing)
      addToCart    < 1     to the right, improve=149.45200, (0 missing)
      pageviews    < 8.5   to the left,  improve=148.85750, (0 missing)
      visits       < 466.5 to the left,  improve=101.27170, (0 missing)
      checkout     < 0.5   to the right, improve= 98.15414, (0 missing)
  Surrogate splits:
      addToCart < 1    to the right, agree=1.000, adj=1.000, (0 split)
      checkout  < 0.5  to the right, agree=0.942, adj=0.578, (0 split)

Node number 10: 268 observations
  predicted class=0  expected loss=0.02238806  P(node) =0.007261889
    class counts:   262     6
   probabilities: 0.978 0.022

Node number 11: 1673 observations,    complexity param=0.01928595
  predicted class=1  expected loss=0.4088464  P(node) =0.04533261
    class counts:   684   989
   probabilities: 0.409 0.591
  left son=22 (317 obs) right son=23 (1356 obs)
  Primary splits:
      pageviews < 8.5   to the left,  improve=212.939100, (0 missing)
      visits    < 434   to the left,  improve= 73.989920, (0 missing)
      checkout  < 0.5   to the right, improve=  6.515696, (0 missing)

Node number 22: 317 observations
  predicted class=0  expected loss=0.06940063  P(node) =0.008589622
    class counts:   295    22
   probabilities: 0.931 0.069

Node number 23: 1356 observations
  predicted class=1  expected loss=0.2868732  P(node) =0.03674299
    class counts:   389   967
   probabilities: 0.287 0.713
```

```
> rpart.plot(tree, type=1, extra=1)
>
> my_prediction <- predict(logit, test, type="response")
> confusionMatrix(data=as.factor(as.numeric(my_prediction>0.5)) , ref
erence=as.factor(as.numeric(test$transaction)))
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 3323 1622
         1  958 9913

               Accuracy : 0.8369
                 95% CI : (0.831, 0.8426)
    No Information Rate : 0.7293
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.606

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.7762
            Specificity : 0.8594
         Pos Pred Value : 0.6720
         Neg Pred Value : 0.9119
             Prevalence : 0.2707
         Detection Rate : 0.2101
   Detection Prevalence : 0.3127
      Balanced Accuracy : 0.8178

       'Positive' Class : 0

>
>
> ####Forecasting####
> ##ARMA##
> total <- sales %>% group_by(date) %>% summarise(total = sum(revenu
e))
> daily_revenue <- ts(total$total, start = c(2020, 1, 1), frequency =
365)
> plot(daily_revenue)
>
> min_rev <- min(daily_revenue)
```

```
> max_rev <- max(daily_revenue)
> std_daily_revenue <- (daily_revenue-min_rev)/(max_rev - min_rev)
> plot(std_daily_revenue)
>
> adf.test(std_daily_revenue)

        Augmented Dickey-Fuller Test

data:  std_daily_revenue
Dickey-Fuller = -5.004, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(std_daily_revenue) : p-value smaller than printed p-value
> adf.test(daily_revenue)

        Augmented Dickey-Fuller Test

data:  daily_revenue
Dickey-Fuller = -5.004, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(daily_revenue) : p-value smaller than printed p-value
> par(mar=c(3, 5, 1, 1), mfrow=c(1,2))
> acf(std_daily_revenue)
> pacf(std_daily_revenue)
>
> ar3 = arima(std_daily_revenue, c(3, 0, 0))
> Box.test(ar3$residuals, 6)

        Box-Pierce test

data:  ar3$residuals
X-squared = 11.134, df = 6, p-value = 0.08432

>
> ts.plot(ar3$residuals)
>
> ar3

Call:
arima(x = std_daily_revenue, order = c(3, 0, 0))
```

```
Coefficients:
         ar1      ar2      ar3  intercept
      0.4658  -0.0167  -0.1807     0.2209
s.e.  0.0595   0.0658   0.0595     0.0119

sigma^2 estimated as 0.02058:  log likelihood = 143.04,  aic = -276.09
> (ar3$coef/sqrt(diag(ar3$var.coef)))
       ar1        ar2        ar3  intercept
 7.8301514 -0.2545875 -3.0358993 18.6310696
>
> #daily\_revenue_t = 0.2213 + 0.534\times daily\_revenue_{t-1} -0.279
1\times  daily\_revenue_{t-3}
>
> ##ARIMA##
> arima12 = arima(std_daily_revenue, c(12, 1, 1))
> Box.test(arima12$residuals, 12)

        Box-Pierce test

data:  arima12$residuals
X-squared = 2.3994, df = 12, p-value = 0.9985

>
> ts.plot(arima12$residuals)
>
> arima12

Call:
arima(x = std_daily_revenue, order = c(12, 1, 1))

Coefficients:
          ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
     ar9
      -0.1489  -0.2888  -0.3472  -0.4767  -0.3286  -0.2375  -0.1637  -0.
1145  -0.1691
s.e.   0.3356   0.1880   0.1782   0.1886   0.2379   0.2086   0.1659   0.1
233   0.0862
         ar10     ar11     ar12      ma1
      -0.0739  -0.1878  -0.0596  -0.4184
s.e.   0.0886   0.0705   0.0928   0.3331

sigma^2 estimated as 0.01863:  log likelihood = 155.18,  aic = -282.37
> (arima12$coef/sqrt(diag(arima12$var.coef)))
```

```
       ar1        ar2        ar3        ar4        ar5        ar6        ar7
     ar8
-0.4437743 -1.5363204 -1.9481282 -2.5272641 -1.3811263 -1.1388407 -0.
9869416 -0.9287913
       ar9       ar10       ar11       ar12        ma1
-1.9604696 -0.8344784 -2.6652265 -0.6417988 -1.2561722
>
> fixed_i = c(0,0,NA,NA,NA,0,0,NA,NA,NA,NA,NA,NA)
> arima12_sparse = arima(std_daily_revenue, order=c(12,1,1), fixed=fi
xed_i)
Error in optim(init[mask], armafn, method = optim.method, hessian = TR
UE,  :
  non-finite finite-difference value [1]
In addition: Warning message:
In arima(std_daily_revenue, order = c(12, 1, 1), fixed = fixed_i) :
  一些 AR 参数是固定的：把 transform.pars 设成 FALSE
```