

Team Yellow technical document R

2022-08-25

This document shows how to explain the relationship between influencing factors and attributed donation amount. The granularity of the donation amount is media mention level, which means for each media mention, how much donation amount will be brought.

Data Preparation

```
library(readxl)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

load the library we need for the model building.

```
live_case_ninth_cut <- read_excel("live_case_ninth_cut.xlsx")
```

Read the data(already merged) into R.

```
live_case_ninth_cut$`SentimentF`<-factor(live_case_ninth_cut$`Sentiment`)
live_case_ninth_cut$`influencer_flagF`<-factor(live_case_ninth_cut$`influencer_flag`)
live_case_ninth_cut$`wkday_F`<-factor(live_case_ninth_cut$`wkday`)
live_case_ninth_cut$`media_typeF`<-factor(live_case_ninth_cut$`media_type`)
live_case_ninth_cut$`holiday_flagF`<-factor(live_case_ninth_cut$`holiday_flag`)
live_case_ninth_cut$`pc_flagF`<-factor(live_case_ninth_cut$`pc_flag`)
live_case_ninth_cut$`stateF`<-factor(live_case_ninth_cut$`state`)
live_case_ninth_cut$`cityF`<-factor(live_case_ninth_cut$`city`)
live_case_ninth_cut$`time_3F`<-factor(live_case_ninth_cut$`time_3`)
live_case_ninth_cut$`minnesota_flagF`<-factor(live_case_ninth_cut$`minnesota_flag`)
live_case_ninth_cut$`month_yearF`<-factor(live_case_ninth_cut$`month_year`)
```

Select influencing factors in the dataset and change the data type of some of the data points into dummy variables.

```
attach(live_case_ninth_cut)
levels(live_case_ninth_cut$media_typeF)
```

```
## [1] "Newspaper" "Radio"      "TV"         "Website"
```

Our analysis bases on the different type of media so that we divide media mentions into 4 different channels.

```
live_case_ninth_cut <- within(live_case_ninth_cut, `wkday_F` <- relevel(`wkday_F`, ref = 6))
live_case_ninth_cut <- within(live_case_ninth_cut, month_yearF <- relevel(month_yearF, ref = 4))
live_case_ninth_cut <- within(live_case_ninth_cut, `SentimentF` <- relevel(`SentimentF`, ref = 3))
live_case_ninth_cut <- within(live_case_ninth_cut, `media_typeF` <- relevel(`media_typeF`, ref = 3))
```

As for the 4 factors (weekdays, month of year, sentiment and media type), transfer their data type into factors (one basic level and how several other levels effect the donation amount compared with the basic level).

Model Building

```
attach(live_case_ninth_cut)
```

```
## The following objects are masked from live_case_ninth_cut (pos = 3):
##
##      # event_date, city, cityF, Facebook_Social_Echo, Headline,
##      holiday_flag, holiday_flagF, Influencer, influencer_flag,
##      influencer_flagF, media_type, media_typeF, minnesota_flag,
##      minnesota_flagF, month_year, month_yearF, pc_flag, pc_flagF, Reach,
##      Sentiment, SentimentF, state, stateF, sum_amount_attribution, time,
##      time_2, time_3, time_3F, Twitter_Social_Echo, wkday, wkday_F
```

```
detach(live_case_ninth_cut)
```

```
linefit <- lm(`sum_amount_attribution` ~ `Reach` + `wkday_F` + `SentimentF` + `media_typeF` + `time_3`
```

Build the linear regression model to explain the relationship between the donation amount and the influencing factors.

```
summary(linefit)
```

```
##
## Call:
## lm(formula = sum_amount_attribution ~ Reach + wkday_F + SentimentF +
##      media_typeF + time_3 + pc_flagF + minnesota_flagF + Facebook_Social_Echo +
##      month_yearF)
##
## Residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -115243 -10296  -5333    3038  681613
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.405e+03  4.682e+03   1.795 0.072722 .
## Reach            2.826e-03  6.215e-04   4.547 5.71e-06 ***
## wkday_F1        -7.899e+03  2.490e+03  -3.172 0.001532 **
## wkday_F2        -9.616e+03  2.316e+03  -4.152 3.40e-05 ***
## wkday_F3        -4.798e+03  2.627e+03  -1.826 0.067905 .
## wkday_F4         8.223e+03  2.864e+03   2.871 0.004125 **
## wkday_F5         9.961e+03  3.535e+03   2.817 0.004878 **
## wkday_F6         8.903e+03  3.710e+03   2.400 0.016479 *
## SentimentFNeutral -3.195e+02  2.923e+03  -0.109 0.912978
## SentimentFPositive  8.460e+03  3.251e+03   2.602 0.009320 **
## media_typeFRadio  -2.521e+03  3.044e+03  -0.828 0.407591
## media_typeFTV     -1.199e+04  2.157e+03  -5.560 2.99e-08 ***
## media_typeFWebsite  6.771e+03  2.209e+03   3.065 0.002196 **
## time_3morning     -8.267e+03  1.884e+03  -4.388 1.19e-05 ***
## time_3noon        -6.738e+03  2.156e+03  -3.125 0.001797 **
## pc_flagF1         8.950e+03  1.861e+03   4.808 1.61e-06 ***
## minnesota_flagF1   1.886e+04  1.674e+03  11.267 < 2e-16 ***
## Facebook_Social_Echo 1.361e+01  9.796e-01  13.895 < 2e-16 ***
## month_yearF2       5.863e+03  4.189e+03   1.399 0.161788
## month_yearF3       2.226e+04  6.095e+03   3.652 0.000265 ***
## month_yearF4       4.956e+03  4.066e+03   1.219 0.222927
## month_yearF5      -5.050e+03  3.927e+03  -1.286 0.198623
## month_yearF6      -4.928e+03  4.019e+03  -1.226 0.220272
## month_yearF7      -3.219e+03  4.500e+03  -0.715 0.474468
## month_yearF8       1.194e+03  4.296e+03   0.278 0.781101
## month_yearF9      -8.095e+03  3.952e+03  -2.048 0.040651 *
## month_yearF10      2.545e+03  4.608e+03   0.552 0.580785
## month_yearF11     -4.891e+03  3.600e+03  -1.359 0.174359
## month_yearF12      2.720e+04  3.832e+03   7.099 1.63e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32150 on 2492 degrees of freedom
## Multiple R-squared:  0.2496, Adjusted R-squared:  0.2412
## F-statistic: 29.6 on 28 and 2492 DF, p-value: < 2.2e-16

```

The result and performance of the linear regression is shown.