61070507218
เฟพรรณ มาสถิตทรัพย์

- Given the dataset of daily demand forecast
  - The dataset was collected during 60 days, this is a real database of a Brazilian logistics company.
  - The dataset has twelve predictors and a target that is the total of orders for daily treatment.

- Experiment and create the best regression model for predicting daily

```r
library(tidyverse)
library(dplyr)
library(ggplot2)


demand <- read.csv("L8-demand.csv", sep=";")
demand

demand2 <- tibble::rowid_to_column(demand, "day")

model <- lm(Target..Total.orders.~ Urgent.order, data = demand2)
summary(model)

model2 <- lm(Target..Total.orders. ~ Urgent.order + Non.urgent.order , data = demand2)
summary(model2)

model3 = lm(Target..Total.orders.~Urgent.order*Non.urgent.order,data=demand2)
summary(model3)

predict(model,demand2)
predict(model2,demand2)
predict(model3,demand2)
```

```
> model <- lm(Target..Total.orders.~ Urgent.order, data = demand2)
> summary(model)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order, data = demand2)

Residuals:
     Min       1Q   Median       3Q      Max
-117.177  -39.350   -8.785   26.678  179.242

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    14.676     36.098   0.407    0.686
Urgent.order    2.407      0.296   8.129 3.72e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 61.78 on 58 degrees of freedom
Multiple R-squared:  0.5326,	Adjusted R-squared:  0.5245
F-statistic: 66.09 on 1 and 58 DF,  p-value: 3.719e-11
```

```
> predict(model,demand2)
       1        2        3        4        5        6        7        8        9       10       11
552.0025 245.8125 217.7345 321.9220 287.8899 246.5778 300.8689 288.7178 314.0138 257.8529 262.0596
      12       13       14       15       16       17       18       19       20       21       22
234.1115 290.0222 354.9866 356.0648 313.7130 395.9040 276.2467 408.6880 415.7370 252.7556 263.8839
      23       24       25       26       27       28       29       30       31       32       33
225.5656 355.0612 269.9774 219.9846 256.3415 243.6730 335.6590 331.8397 236.4700 331.0576 450.6331
      34       35       36       37       38       39       40       41       42       43       44
281.5076 315.7899 200.8785 351.7016 282.7542 283.8781 307.5545 377.3730 261.4267 274.7233 271.3203
      45       46       47       48       49       50       51       52       53       54       55
241.6563 417.6768 350.1180 290.9873 285.9093 268.0281 304.2502 296.1519 224.7594 254.7507 205.0010
      56       57       58       59       60
395.2422 334.9226 278.5354 275.5416 306.1322
```

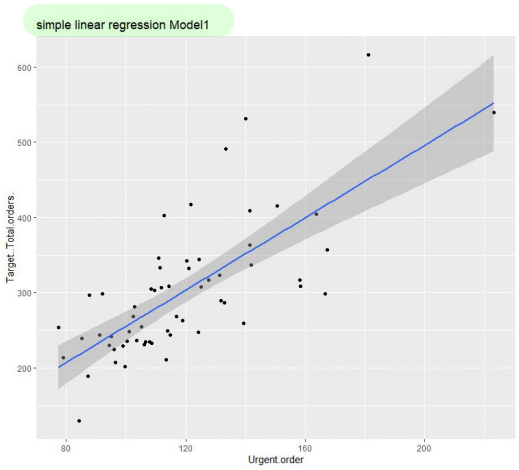
```
>
> model2 <- lm(Target..Total.orders. ~ Urgent.order + Non.urgent.order , data = demand2)
> summary(model2)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order + Non.urgent.order,
    data = demand2)

Residuals:
    Min      1Q  Median      3Q     Max
-10.700  -9.123  -6.669   3.006 159.158

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      13.94606   13.76304   1.013    0.315
Urgent.order      0.97804    0.13678   7.151 1.78e-09 ***
Non.urgent.order  0.98878    0.05347  18.493  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.56 on 57 degrees of freedom
Multiple R-squared:  0.9332,	Adjusted R-squared:  0.9309
F-statistic: 398.3 on 2 and 57 DF,  p-value: < 2.2e-16
```

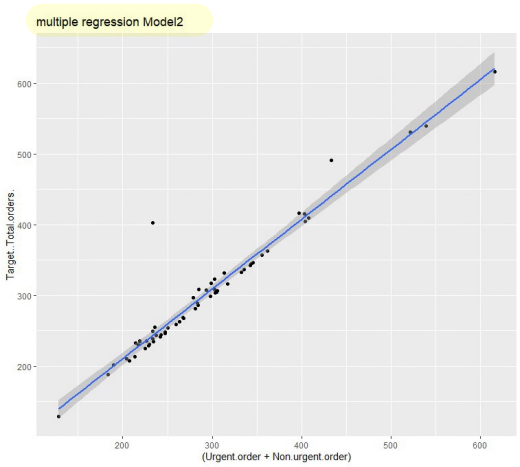
```
> predict(model2,demand2)
       1        2        3        4        5        6        7        8        9       10       11
545.0693 235.0680 139.6290 308.1836 214.4946 217.8697 272.7597 243.3543 351.9950 257.6454 290.9118
      12       13       14       15       16       17       18       19       20       21       22
247.3514 294.4969 370.1163 344.2778 256.8315 311.3092 224.4788 412.0292 306.5309 239.3688 237.1251
      23       24       25       26       27       28       29       30       31       32       33
288.3249 416.0239 239.4601 243.8588 229.3790 252.3188 440.7973 292.4820 307.9736 311.4720 621.5349
      34       35       36       37       38       39       40       41       42       43       44
354.8522 302.6085 260.4616 528.3946 341.8443 315.1284 404.9589 410.8104 277.8390 244.2433 229.2311
      45       46       47       48       49       50       51       52       53       54       55
228.1077 363.4253 268.7855 252.6080 243.4489 245.9077 351.4146 276.9069 194.7338 200.0323 224.2094
      56       57       58       59       60
325.5406 293.1841 312.5389 314.3094 322.3520
```

```
> model3 = lm(Target..Total.orders.~Urgent.order*Non.urgent.order,data=demand2)
> summary(model3)

Call:
lm(formula = Target..Total.orders. ~ Urgent.order * Non.urgent.order,
    data = demand2)

Residuals:
    Min      1Q  Median      3Q     Max
-10.115  -9.504  -7.113   3.152 159.283

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    5.4292067 36.3823335   0.149 0.881911
Urgent.order                   1.0439799  0.2946657   3.543 0.000806 ***
Non.urgent.order               1.0338506  0.1859750   5.559 7.83e-07 ***
Urgent.order:Non.urgent.order -0.0003293  0.0013004  -0.253 0.801011
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.75 on 56 degrees of freedom
Multiple R-squared:  0.9333,	Adjusted R-squared:  0.9297
F-statistic: 261.2 on 3 and 56 DF,  p-value: < 2.2e-16
```

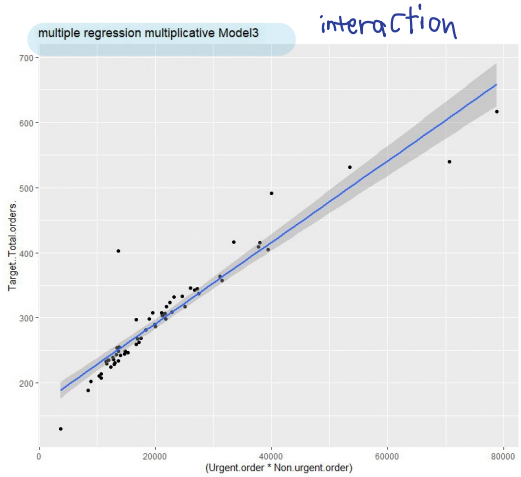
```
> predict(model3,demand2)
       1        2        3        4        5        6        7        8        9       10       11
542.2766 234.6141 137.4308 308.6049 214.1601 217.1871 272.9369 243.2508 352.5807 257.5205 291.1761
      12       13       14       15       16       17       18       19       20       21       22
247.0421 294.7864 370.5955 344.7975 257.0185 312.2168 224.1077 412.1811 307.7183 239.0017 236.7896
      23       24       25       26       27       28       29       30       31       32       33
288.6810 416.4332 239.1873 243.4830 228.9073 252.0958 441.4212 292.9092 308.5705 311.9288 618.6170
      34       35       36       37       38       39       40       41       42       43       44
355.6598 302.9935 260.4325 528.7145 342.5312 315.5857 405.8421 411.0824 277.9548 244.0481 228.8641
      45       46       47       48       49       50       51       52       53       54       55
227.5384 364.0597 269.3610 252.5908 243.3242 245.6906 352.0438 277.0881 193.5489 199.1880 223.4659
      56       57       58       59       60
326.3405 293.6313 312.9879 314.7838 322.8189
```

```r
g_model1 <- ggplot(data = demand2, aes(x = Urgent.order, y = Target..Total.orders.)) +
    geom_point() +
    stat_smooth(method = "lm") +
    ggtitle("simple linear regression Model1")

g_model2 <- ggplot(data = demand2, aes(x = (Urgent.order + Non.urgent.order), y = Target..Total.orders.)) +
   geom_point() +
   stat_smooth(method = "lm") +
   ggtitle("multiple regression Model2")


g_model3 <- ggplot(data = demand2, aes(x = (Urgent.order* Non.urgent.order), y = Target..Total.orders.)) +
   geom_point() +
   stat_smooth(method = "lm") +
   ggtitle("multiple regression multiplicative Model3")
```





interaction



ตอบ the best regression model คือ model 2

เพราะ มี error Residual standard น้อยที่สุด และมี Adjusted R-square มากที่สุด แสดงถึงอธิบายความแปรปรวนของข้อมูลได้เยอะ จึงคิดว่า การทำนาย ที่ model 2 ทำนายได้ดีที่สุด จากทั้งหมด 3 model ที่ได้ทดลอง

- Why is sometime adding predictors do not help prediction?

เพราะ แต่ละ adding predictor ก็จะทำให้ model มีคุณภาพที่สามารถ predictors ได้ต่างกัน ขึ้นอยู่กับลักษณะความสัมพันธ์ และตัวแปร และบาง adding predictor ก็ไม่สามารถ predictors model ได้ หรือมีคุณภาพการ predictors ต่ำ