NYC Regression Model

Hannah Khuong

Contents

Model summary

We power transformed the regression with a fractional polynomial term as we observe curvature in cancellation rate trend (Royston & Altman, 1994). Multiple fractional power in was tried and the power 0.2 fitted our curve the best.

Because different product types have different patterns of cancellation rate along days prior axis, an interaction term of Product Type and Days Prior was included. Cumulative Gross Bookings had a negative correlation of 0.33 with Cancellation Rate. The formula included interaction term of the Regrouped Day of Week and Cumulative Gross Bookings to capture the effect of Weekdays versus Weekend.

The best regression model for New York hotel belongs to the Grouping 3 of product type. Even though this Grouping did not have the best R-squared, its performance in MASE Metrics was the best. The most important features in this regression models are Cumulative Gross Bookings, fractional polynomial transformed Days Prior, Product Type and the interaction of Product Type and Days Prior.

Metrics Functions

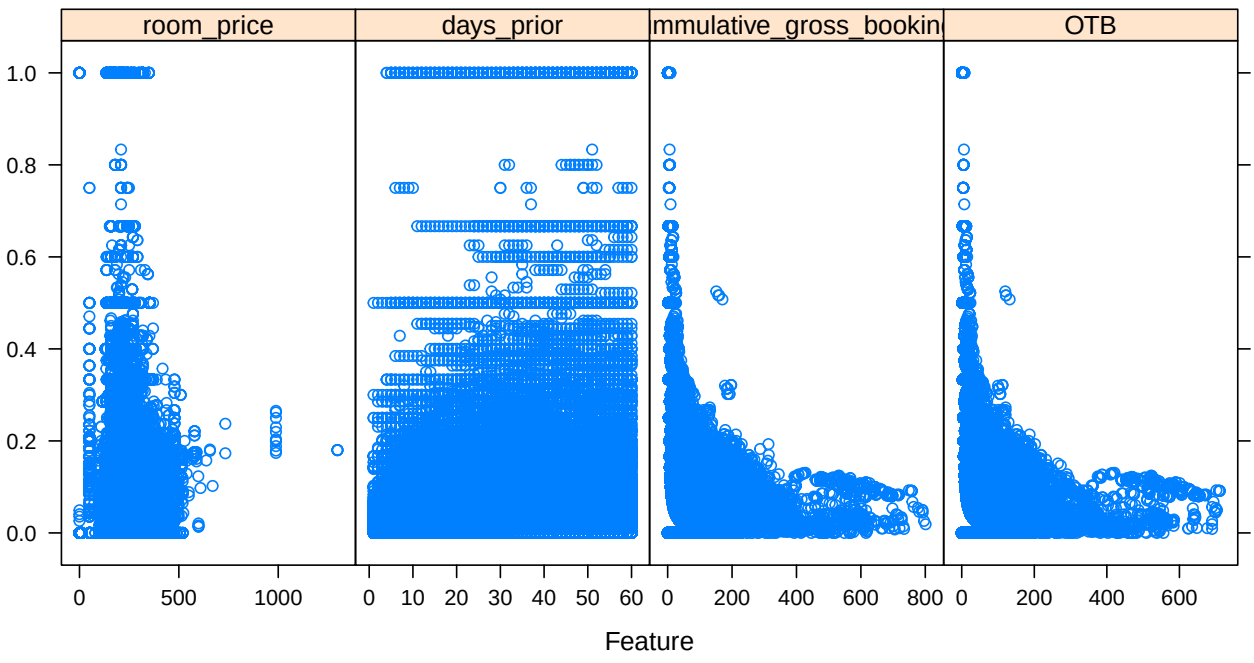
```
# Make functions

#MAE
mae_by_days = function(data, model, name){
  name <- paste(name)
  data %>%
    mutate(predict_OTB_to_survive = OTB -  predict(model, data) * OTB) %>%
    group_by(days_prior_cat) %>%
    summarise(!!name := round(mae(OTB_to_survive, predict_OTB_to_survive),4))
}

#MAPE
mape_by_days = function(data, model, name){
```

```
  name <-paste(name)
  data %>%
  mutate(predict_OTB_to_survive = OTB -  predict(model, data) * OTB) %>%
    filter(OTB_to_survive != 0)%>%
  group_by(days_prior_cat) %>%
  summarise(!!name := round(mape(OTB_to_survive, predict_OTB_to_survive) ,4))
}

#MASE
mase_by_days = function(data, model, name){
  name <-paste(name)
  data %>%
    mutate(predict_OTB_to_survive = OTB - predict(model, data) * OTB) %>%
    group_by(days_prior_cat) %>%
    summarise(!!name :=  round(sum(abs(OTB_to_survive - predict_OTB_to_survive))/
                                    sum(abs(OTB_to_survive - naive_survive_pred)),digits = 4))

}
```

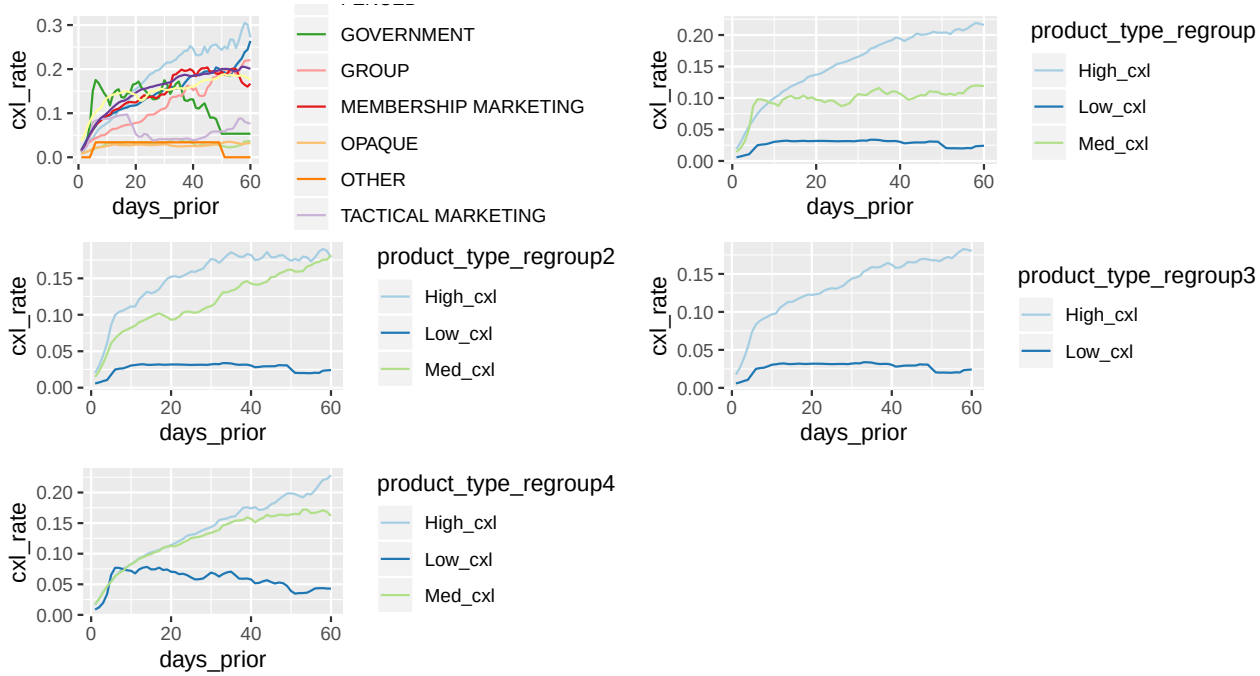
Regression model

- variables:
  - cummulative bookings
  - room price
  - product type
  - days prior
- interaction of days_prior x product type
- interaction of dow x product type

```
library(caret)
featurePlot(x = train_nyc[, c("room_price", "days_prior", "cummulative_gross_bookings", "OTB")], y = tra
```

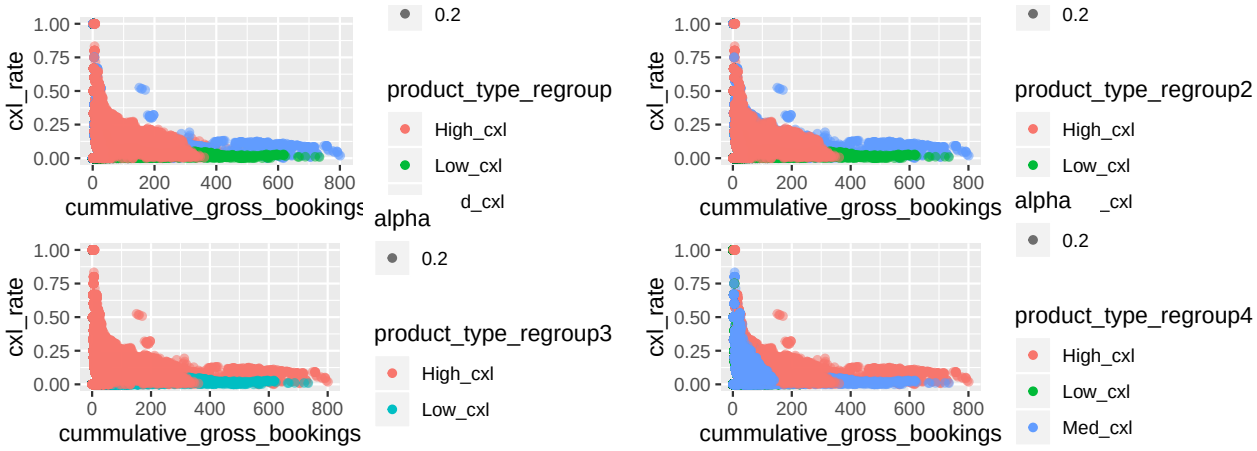


```
# Check cxl_rate trends by different groups
grid.arrange(
train_nyc %>%
  ggplot(aes(x = days_prior, color = product_type)) +
  stat_summary(aes(y = cxl_rate), fun.y = 'mean', geom = 'line')+
  scale_color_brewer(palette = "Paired"),
train_nyc %>%
  ggplot(aes(x = days_prior, color = product_type_regroup)) +
  stat_summary(aes(y = cxl_rate), fun.y = 'mean', geom = 'line')+
  scale_color_brewer(palette = "Paired"),
train_nyc %>%
  ggplot(aes(x = days_prior, color = product_type_regroup2)) +
  stat_summary(aes(y = cxl_rate), fun.y = 'mean', geom = 'line')+
  scale_color_brewer(palette = "Paired"),
train_nyc %>%
  ggplot(aes(x = days_prior, color = product_type_regroup3)) +
  stat_summary(aes(y = cxl_rate), fun.y = 'mean', geom = 'line')+
  scale_color_brewer(palette = "Paired"),
train_nyc %>%
  ggplot(aes(x = days_prior, color = product_type_regroup4)) +
  stat_summary(aes(y = cxl_rate), fun.y = 'mean', geom = 'line')+
  scale_color_brewer(palette = "Paired"),
 ncol=2)
```

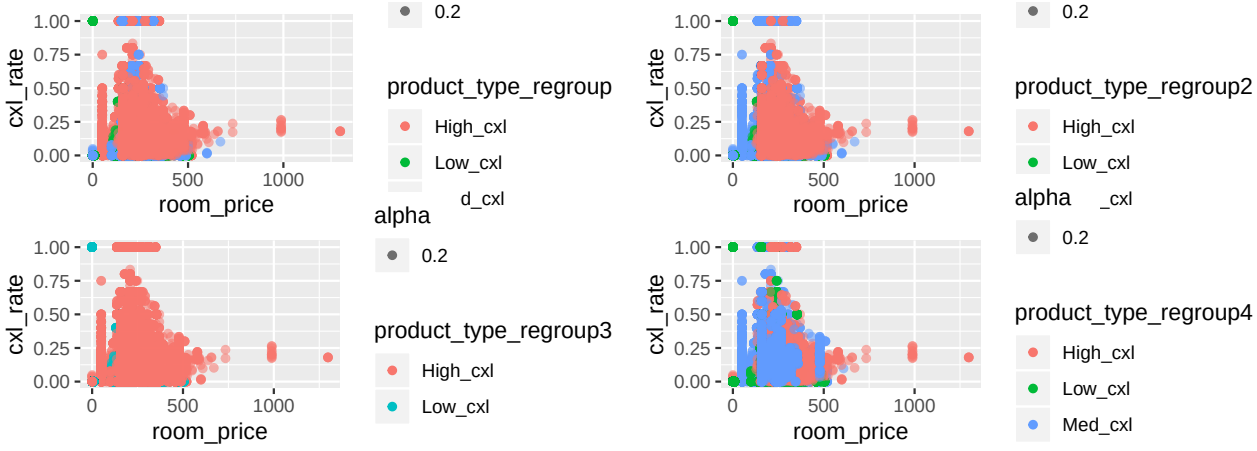



```
# Cummulative gross bookings
grid.arrange(
train_nyc %>% ggplot(aes(cummulative_gross_bookings, cxl_rate,
                         color = product_type_regroup, alpha = 0.2))+geom_point(),
train_nyc %>% ggplot(aes(cummulative_gross_bookings, cxl_rate,
                         color = product_type_regroup2, alpha = 0.2))+geom_point(),
train_nyc %>% ggplot(aes(cummulative_gross_bookings, cxl_rate,
                         color = product_type_regroup3, alpha = 0.2))+geom_point(),
```

```
train_nyc %>% ggplot(aes(cummulative_gross_bookings, cxl_rate,
                         color = product_type_regroup4, alpha = 0.2))+geom_point(),
 nrow = 3)
```




```
# Room Price
grid.arrange(
train_nyc %>% ggplot(aes(room_price, cxl_rate,
                         color = product_type_regroup, alpha = 0.2))+geom_point(),
train_nyc %>% ggplot(aes(room_price, cxl_rate,
                         color = product_type_regroup2, alpha = 0.2))+geom_point(),
train_nyc %>% ggplot(aes(room_price, cxl_rate,
                         color = product_type_regroup3, alpha = 0.2))+geom_point(),
train_nyc %>% ggplot(aes(room_price, cxl_rate,
                         color = product_type_regroup4, alpha = 0.2))+geom_point(),
 nrow = 3)
```




```
# Make functions

#MAE
mae_by_days = function(data, model, name){
  name <- paste(name)
  data %>%
    mutate(predict_OTB_to_survive = OTB -  predict(model, data) * OTB) %>%
    group_by(days_prior_cat) %>%
    summarise(!!name := round(mae(OTB_to_survive, predict_OTB_to_survive),4))
```

```
}

#MAPE
mape_by_days = function(data, model, name){
  name <-paste(name)
  data %>%
  mutate(predict_OTB_to_survive = OTB -  predict(model, data) * OTB) %>%
    filter(OTB_to_survive != 0)%>%
  group_by(days_prior_cat) %>%
  summarise(!!name := round(mape(OTB_to_survive, predict_OTB_to_survive) ,4))
}

#MASE
mase_by_days = function(data, model, name){
  name <-paste(name)
  data %>%
    mutate(predict_OTB_to_survive = OTB - predict(model, data) * OTB) %>%
    group_by(days_prior_cat) %>%
    summarise(!!name :=  round(sum(abs(OTB_to_survive - predict_OTB_to_survive))/
                                  sum(abs(OTB_to_survive - naive_survive_pred)),digits = 4))

}
```

Group 1

```
# best fit
library(car)
boxTidwell(cxl_rate ~ days_prior, data = train_nyc)
```

```
##  MLE of lambda Score Statistic (z)  Pr(>|z|)
##       0.050017             -14.163 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## iterations =  4
```

Model 1

- based on the relationship between days_prior and cxl rate, we power days prior by 0.2
- days prior has interaction with product type
- dow has interaction with product type
- High_cxl group has inverse relationship with cummulative_gross_bookings (-0.25) and room_price (-0.13) while has positive relationship with days prior (0.32). The other 2 groups have almost no relationship with other variables.

```
# model 1
g1_mod1 <- lm(cxl_rate ~ cummulative_gross_bookings  +I(days_prior^0.2) + product_type_regroup * days_pr
              , data = train_nyc)
summary(g1_mod1)
```

```
##
```

```
## Call:
## lm(formula = cxl_rate ~ cummulative_gross_bookings + I(days_prior^0.2) +
##     product_type_regroup * days_prior + last_week * cummulative_gross_bookings,
##     data = train_nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22712 -0.07634 -0.02457  0.01989  0.98292
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       -8.427e-02  1.588e-02
## cummulative_gross_bookings                        -4.735e-05  1.648e-05
## I(days_prior^0.2)                                  1.133e-01  1.255e-02
## product_type_regroupLow_cxl                       -4.332e-02  3.608e-03
## product_type_regroupMed_cxl                        3.834e-03  3.602e-03
## days_prior                                         8.077e-04  1.821e-04
## last_weeknot_last_week                             5.966e-03  4.999e-03
## product_type_regroupLow_cxl:days_prior            -2.777e-03  1.029e-04
## product_type_regroupMed_cxl:days_prior            -2.071e-03  1.026e-04
## cummulative_gross_bookings:last_weeknot_last_week -6.103e-05  1.938e-05
##                                                   t value Pr(>|t|)
## (Intercept)                                        -5.306 1.12e-07 ***
## cummulative_gross_bookings                         -2.872  0.00408 **
## I(days_prior^0.2)                                   9.025  < 2e-16 ***
## product_type_regroupLow_cxl                       -12.008  < 2e-16 ***
## product_type_regroupMed_cxl                         1.065  0.28705
## days_prior                                          4.436 9.17e-06 ***
## last_weeknot_last_week                              1.194  0.23264
## product_type_regroupLow_cxl:days_prior            -26.992  < 2e-16 ***
## product_type_regroupMed_cxl:days_prior            -20.186  < 2e-16 ***
## cummulative_gross_bookings:last_weeknot_last_week  -3.150  0.00164 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1535 on 43910 degrees of freedom
## Multiple R-squared:  0.1516, Adjusted R-squared:  0.1514
## F-statistic: 871.6 on 9 and 43910 DF,  p-value: < 2.2e-16
```

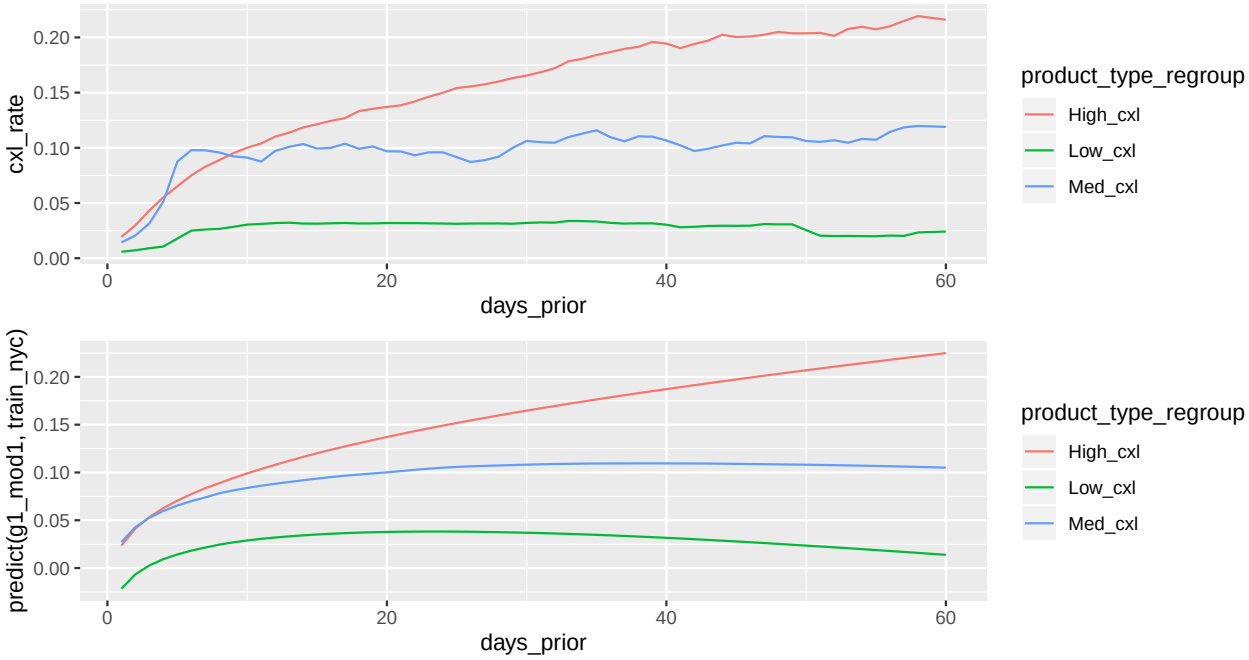
```
# TRAIN (in-group)
# MAE
g1_mod1_mae_in = mae_by_days(train_nyc,g1_mod1, 'mae_g1_mod1_in')
#MAPE
g1_mod1_mape_in = mape_by_days(train_nyc,g1_mod1, 'mape_g1_mod1_in')
#MASE
g1_mod1_mase_in = mase_by_days(train_nyc,g1_mod1, 'mase_g1_mod1_in')

# TEST (out-group)
# MAE
g1_mod1_mae_out = mae_by_days(test_nyc,g1_mod1, 'mae_g1_mod1')
#MAPE
g1_mod1_mape_out = mape_by_days(test_nyc,g1_mod1, 'mape_g1_mod1')
#MASE
g1_mod1_mase_out = mase_by_days(test_nyc,g1_mod1, 'mase_g1_mod1')
```

Table 1: Group 1 - Model 1

days_prior_cat	mae_g1_mod1_in	mape_g1_mod1_in	mase_g1_mod1_in	mae_g1_mod1	mape_g1_mod1
Day 01-07	1.9224	0.0415	0.2814	2.5058	0.0348
Day 08-14	2.2493	0.0649	0.5583	2.8396	0.0526
Day 15-20	2.1943	0.0731	0.6282	2.6786	0.0599
Day 21-27	2.0131	0.0865	0.6663	2.6268	0.0754
Day 28-60	1.5759	0.1258	0.7409	2.6090	0.1259

```
kable(
g1_mod1_mae_in %>% left_join(g1_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Group 1 - Model 1 "
)%>%
  kable_styling(bootstrap_options = 'condensed')
```

```
#visualize
grid.arrange(
ggplot(data = train_nyc, aes(color = product_type_regroup,x = days_prior,y = cxl_rate))+
  stat_summary(fun.y = 'mean', geom = 'line'),
ggplot(data = train_nyc, aes(color = product_type_regroup,x = days_prior,y = predict(g1_mod1, train_nyc)
  stat_summary(fun.y = 'mean', geom = 'line'), nrow = 2)
```



Model 2

same as model 1, except for adding dow interaction

** Result ** Higher R-squared (15.44 vs 15.27)

```r
g1_mod2 <- lm(cxl_rate ~ cummulative_gross_bookings + dow_regroup*product_type_regroup + I(days_prior^0
              , data = train_nyc)
summary(g1_mod2)
```

```
##
## Call:
## lm(formula = cxl_rate ~ cummulative_gross_bookings + dow_regroup *
##     product_type_regroup + I(days_prior^0.2) + product_type_regroup *
##     days_prior + last_week * cummulative_gross_bookings, data = train_nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22934 -0.07656 -0.02460  0.01987  0.98644
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                    -8.166e-02  1.591e-02
## cummulative_gross_bookings                     -4.932e-05  1.651e-05
## dow_regroupweekend                             -4.985e-03  2.165e-03
## product_type_regroupLow_cxl                    -4.911e-02  3.954e-03
## product_type_regroupMed_cxl                    -1.114e-03  3.926e-03
## I(days_prior^0.2)                               1.132e-01  1.255e-02
## days_prior                                      8.070e-04  1.820e-04
## last_weeknot_last_week                          5.935e-03  4.998e-03
## dow_regroupweekend:product_type_regroupLow_cxl  1.288e-02  3.597e-03
## dow_regroupweekend:product_type_regroupMed_cxl  1.119e-02  3.579e-03
## product_type_regroupLow_cxl:days_prior         -2.776e-03  1.029e-04
## product_type_regroupMed_cxl:days_prior         -2.070e-03  1.026e-04
## cummulative_gross_bookings:last_weeknot_last_week -6.067e-05  1.938e-05
##                                                t value Pr(>|t|)
## (Intercept)                                     -5.131 2.89e-07 ***
## cummulative_gross_bookings                      -2.987 0.002815 **
## dow_regroupweekend                              -2.302 0.021320 *
## product_type_regroupLow_cxl                    -12.419  < 2e-16 ***
## product_type_regroupMed_cxl                     -0.284 0.776583
## I(days_prior^0.2)                                9.015  < 2e-16 ***
## days_prior                                       4.434 9.29e-06 ***
## last_weeknot_last_week                           1.187 0.235100
## dow_regroupweekend:product_type_regroupLow_cxl   3.582 0.000341 ***
## dow_regroupweekend:product_type_regroupMed_cxl   3.126 0.001773 **
## product_type_regroupLow_cxl:days_prior         -26.987  < 2e-16 ***
## product_type_regroupMed_cxl:days_prior         -20.183  < 2e-16 ***
## cummulative_gross_bookings:last_weeknot_last_week  -3.130 0.001749 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1535 on 43907 degrees of freedom
## Multiple R-squared:  0.1519, Adjusted R-squared:  0.1517
## F-statistic: 655.4 on 12 and 43907 DF,  p-value: < 2.2e-16
```

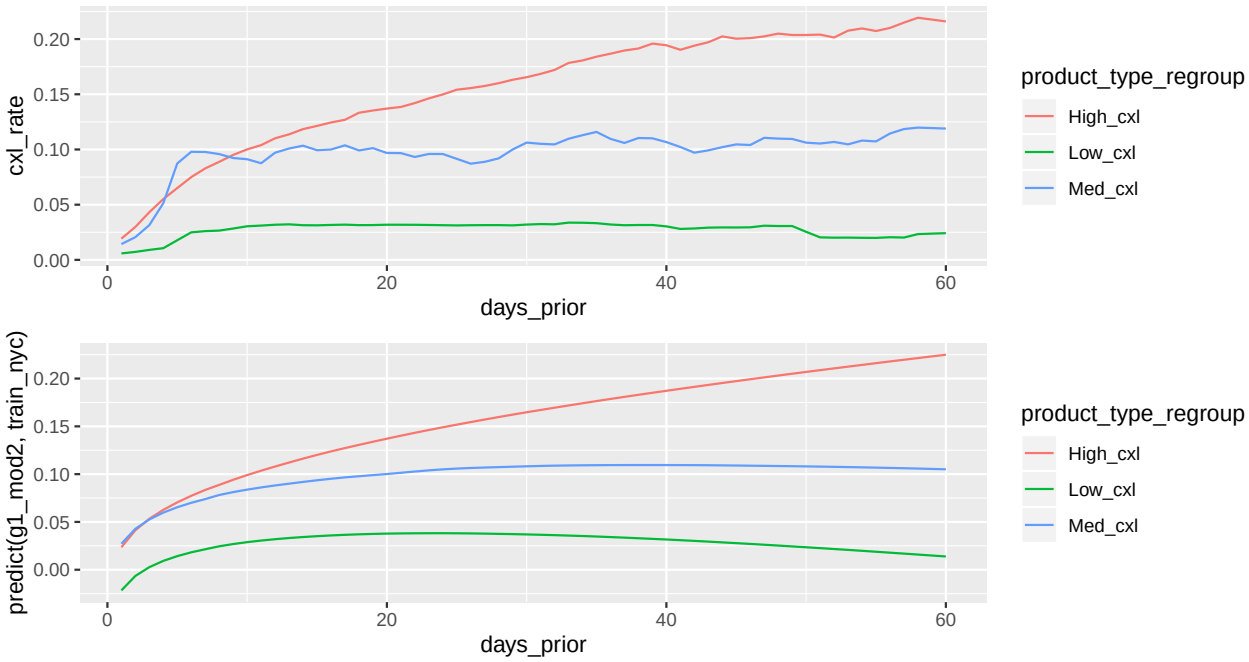
```r
# TRAIN (in-group)
# MAE
g1_mod2_mae_in = mae_by_days(train_nyc,g1_mod2, 'mae_g1_mod2_in')
#MAPE
```

Table 2: Group 1 - Model 2

days_prior_cat	mae_g1_mod1_in	mape_g1_mod2_in	mase_g1_mod2_in	mae_g1_mod2	mape_g1_mod2
Day 01-07	1.9224	0.0416	0.2810	2.5653	0.0352
Day 08-14	2.2493	0.0648	0.5565	2.8751	0.0527
Day 15-20	2.1943	0.0732	0.6308	2.6975	0.0600
Day 21-27	2.0131	0.0867	0.6713	2.6119	0.0753
Day 28-60	1.5759	0.1258	0.7471	2.6052	0.1258

```
g1_mod2_mape_in = mape_by_days(train_nyc,g1_mod2, 'mape_g1_mod2_in')
#MASE
g1_mod2_mase_in = mase_by_days(train_nyc,g1_mod2, 'mase_g1_mod2_in')

# TEST (out-group)
# MAE
g1_mod2_mae_out = mae_by_days(test_nyc,g1_mod2, 'mae_g1_mod2')
#MAPE
g1_mod2_mape_out = mape_by_days(test_nyc,g1_mod2, 'mape_g1_mod2')
#MASE
g1_mod2_mase_out = mase_by_days(test_nyc,g1_mod2, 'mase_g1_mod2')

# Summarise metrics
kable(
g1_mod1_mae_in %>% left_join(g1_mod2_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod2_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod2_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod2_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g1_mod2_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Group 1 - Model 2 "
)%>%
  kable_styling(bootstrap_options = 'condensed')
```

```
#visualize
grid.arrange(
ggplot(data = train_nyc, aes(color = product_type_regroup,x = days_prior,y = cxl_rate))+
  stat_summary(fun.y = 'mean', geom = 'line'),
ggplot(data = train_nyc, aes(color = product_type_regroup,x = days_prior,y = predict(g1_mod2, train_nyc
  stat_summary(fun.y = 'mean', geom = 'line'), nrow = 2)
```



Group 2

Model 1

- based on the relationship between days_prior and cxl rate, we power days prior by 0.2
- relationship with days prior has interaction with product type
- coefficient between cxl rate and other variables across product type groups do not differ much, and they are equally weak (<=0.1)

```
# model 1
g2_mod1 <- lm(cxl_rate ~  cummulative_gross_bookings+ product_type_regroup2 *dow_regroup +I(days_prior^
               product_type_regroup2 * days_prior+ last_week*cummulative_gross_bookings
             , data = train_nyc)
summary(g2_mod1)
```

```
##
## Call:
## lm(formula = cxl_rate ~ cummulative_gross_bookings + product_type_regroup2 *
##     dow_regroup + I(days_prior^0.2) + product_type_regroup2 *
##     days_prior + last_week * cummulative_gross_bookings, data = train_nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19749 -0.06617 -0.02170  0.02431  0.98697
##
## Coefficients:
##                                                 Estimate Std. Error
## (Intercept)                                   -6.931e-02  1.613e-02
## cummulative_gross_bookings                    -2.839e-05  1.678e-05
## product_type_regroup2Low_cxl                  -6.619e-02  4.195e-03
## product_type_regroup2Med_cxl                  -3.619e-02  3.864e-03
```

```
## dow_regroupweekend                                    2.461e-03  2.486e-03
## I(days_prior^0.2)                                     1.154e-01  1.272e-02
## days_prior                                           -3.724e-05  1.876e-04
## last_weeknot_last_week                                4.760e-03  5.064e-03
## product_type_regroup2Low_cxl:dow_regroupweekend       4.427e-03  3.826e-03
## product_type_regroup2Med_cxl:dow_regroupweekend      -5.173e-03  3.494e-03
## product_type_regroup2Low_cxl:days_prior              -1.929e-03  1.094e-04
## product_type_regroup2Med_cxl:days_prior               1.966e-04  1.006e-04
## cummulative_gross_bookings:last_weeknot_last_week -4.960e-05  1.963e-05
##                                                      t value Pr(>|t|)
## (Intercept)                                           -4.297 1.73e-05 ***
## cummulative_gross_bookings                            -1.692   0.0907 .
## product_type_regroup2Low_cxl                         -15.777  < 2e-16 ***
## product_type_regroup2Med_cxl                          -9.367  < 2e-16 ***
## dow_regroupweekend                                     0.990   0.3223
## I(days_prior^0.2)                                      9.076  < 2e-16 ***
## days_prior                                            -0.199   0.8426
## last_weeknot_last_week                                 0.940   0.3472
## product_type_regroup2Low_cxl:dow_regroupweekend        1.157   0.2473
## product_type_regroup2Med_cxl:dow_regroupweekend       -1.480   0.1388
## product_type_regroup2Low_cxl:days_prior              -17.631  < 2e-16 ***
## product_type_regroup2Med_cxl:days_prior                1.955   0.0505 .
## cummulative_gross_bookings:last_weeknot_last_week     -2.526   0.0115 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1555 on 43907 degrees of freedom
## Multiple R-squared:  0.1293, Adjusted R-squared:  0.1291
## F-statistic: 543.4 on 12 and 43907 DF,  p-value: < 2.2e-16
```

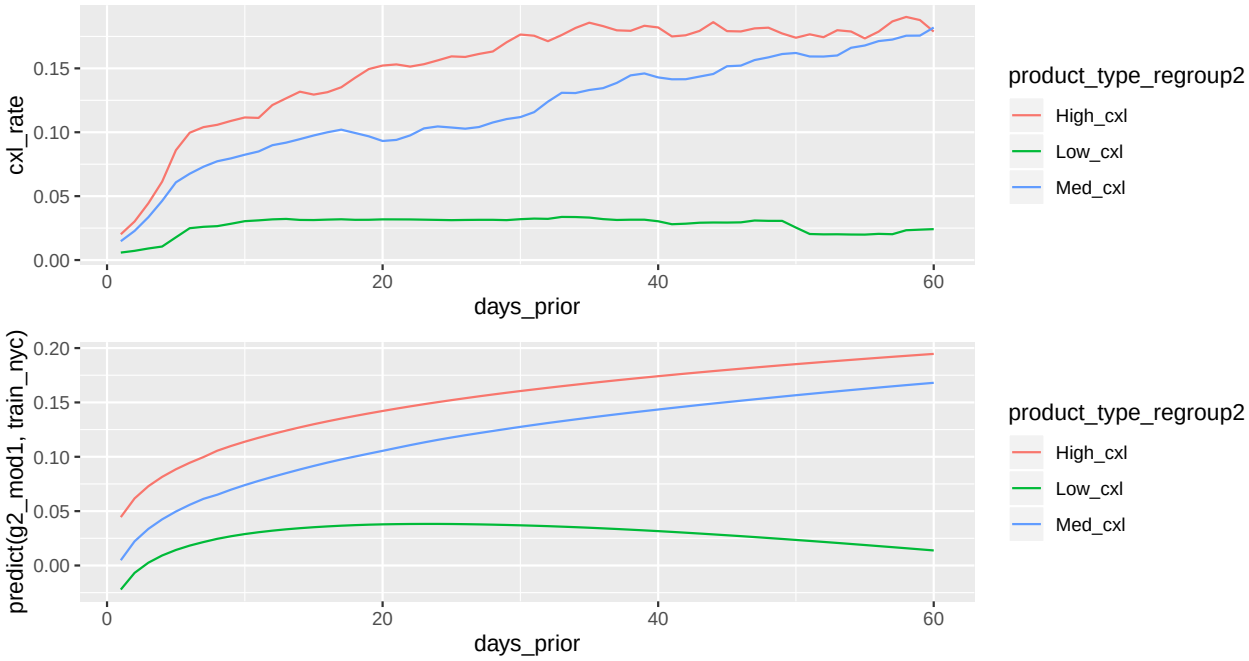
```
# TRAIN (in-group)
# MAE
g2_mod1_mae_in = mae_by_days(train_nyc,g2_mod1, 'mae_g2_mod1_in')
#MAPE
g2_mod1_mape_in = mape_by_days(train_nyc,g2_mod1, 'mape_g2_mod1_in')
#MASE
g2_mod1_mase_in = mase_by_days(train_nyc,g2_mod1, 'mase_g2_mod1_in')

# TEST (out-group)
# MAE
g2_mod1_mae_out = mae_by_days(test_nyc,g2_mod1, 'mae_g2_mod1_out')
#MAPE
g2_mod1_mape_out = mape_by_days(test_nyc,g2_mod1, 'mape_g2_mod1_out')
#MASE
g2_mod1_mase_out = mase_by_days(test_nyc,g2_mod1, 'mase_g2_mod1_out')

kable(
g1_mod1_mae_in %>% left_join(g2_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g2_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g2_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g2_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g2_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Group 2 - Model 1 "
)%>%
```

Table 3: Group 2 - Model 1

days_prior_cat	mae_g1_mod1_in	mape_g2_mod1_in	mase_g2_mod1_in	mae_g2_mod1_out	mape_g2_m
Day 01-07	1.9224	0.0424	0.3056	2.6335	
Day 08-14	2.2493	0.0647	0.5546	2.8816	
Day 15-20	2.1943	0.0723	0.5991	2.7748	
Day 21-27	2.0131	0.0862	0.6223	2.7546	
Day 28-60	1.5759	0.1242	0.6466	2.4932	

```
  kable_styling(bootstrap_options = 'condensed')
```

```
#visualize
grid.arrange(
ggplot(data = train_nyc, aes(color = product_type_regroup2,x = days_prior,y = cxl_rate))+
  stat_summary(fun.y = 'mean', geom = 'line'),
ggplot(data = train_nyc, aes(color = product_type_regroup2,x = days_prior,y = predict(g2_mod1, train_ny
  stat_summary(fun.y = 'mean', geom = 'line'), nrow = 2)
```



Group 3

Model 1

- based on the relationship between days_prior and cxl rate, we power days prior by 0.2
- relationship with days prior has interaction with product type
- coefficient between cxl rate and other variables across product type groups do not differ much, and they are equally weak (<=0.1)

```
# model 1
g3_mod1 <- lm(cxl_rate ~ cummulative_gross_bookings +dow_regroup*cummulative_gross_bookings +I(days_pri
```

```
              , data = train_nyc)
summary(g3_mod1)
```

```
##
## Call:
## lm(formula = cxl_rate ~ cummulative_gross_bookings + dow_regroup *
##     cummulative_gross_bookings + I(days_prior^0.2) + product_type_regroup3 *
##     days_prior, data = train_nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18446 -0.06761 -0.02214  0.02453  0.98504
##
## Coefficients:
##                                                Estimate Std. Error
## (Intercept)                                  -8.215e-02  1.097e-02
## cummulative_gross_bookings                   -6.453e-05  1.114e-05
## dow_regroupweekend                            3.179e-03  1.764e-03
## I(days_prior^0.2)                             1.139e-01  7.626e-03
## product_type_regroup3Low_cxl                 -4.422e-02  3.432e-03
## days_prior                                    8.484e-05  1.350e-04
## cummulative_gross_bookings:dow_regroupweekend -3.017e-05  1.733e-05
## product_type_regroup3Low_cxl:days_prior       -2.034e-03  9.750e-05
##                                               t value Pr(>|t|)
## (Intercept)                                    -7.490 7.01e-14 ***
## cummulative_gross_bookings                     -5.795 6.90e-09 ***
## dow_regroupweekend                              1.802   0.0716 .
## I(days_prior^0.2)                              14.937  < 2e-16 ***
## product_type_regroup3Low_cxl                  -12.885  < 2e-16 ***
## days_prior                                      0.628   0.5298
## cummulative_gross_bookings:dow_regroupweekend  -1.741   0.0817 .
## product_type_regroup3Low_cxl:days_prior       -20.861  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1561 on 43912 degrees of freedom
## Multiple R-squared:  0.1221, Adjusted R-squared:  0.122
## F-statistic: 872.8 on 7 and 43912 DF,  p-value: < 2.2e-16
```

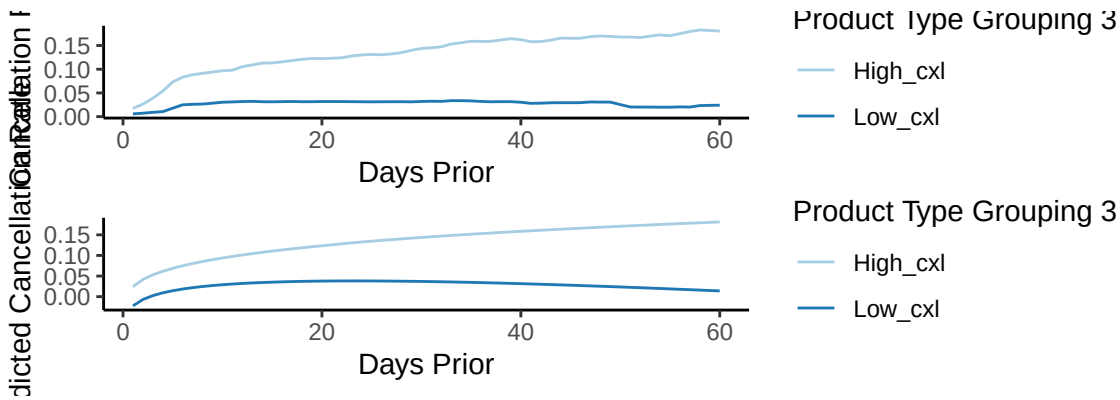
```
# TRAIN (in-group)
# MAE
g3_mod1_mae_in = mae_by_days(train_nyc,g3_mod1, 'mae_g3_mod1_in')
#MAPE
g3_mod1_mape_in = mape_by_days(train_nyc,g3_mod1, 'mape_g3_mod1_in')
#MASE
g3_mod1_mase_in = mase_by_days(train_nyc,g3_mod1, 'mase_g3_mod1_in')

# TEST (out-group)
g3_mod1_mae_out = mae_by_days(test_nyc,g3_mod1, 'mae_g3_mod1_out')
#MAPE
g3_mod1_mape_out = mape_by_days(test_nyc,g3_mod1, 'mape_g3_mod1_out')
#MASE
g3_mod1_mase_out = mase_by_days(test_nyc,g3_mod1, 'mase_g3_mod1_out')
```

Table 4: Group 3 - Model 1

days_prior_cat	mae_g3_mod1_in	mape_g3_mod1_in	mase_g3_mod1_in	mae_g3_mod1_out	mape_g3_m
Day 01-07	2.1189	0.0426	0.3101	2.7497	
Day 08-14	2.1661	0.0656	0.5377	2.7940	
Day 15-20	2.1326	0.0745	0.6106	2.8703	
Day 21-27	1.9101	0.0880	0.6322	2.8481	
Day 28-60	1.4212	0.1253	0.6681	2.5508	

```
kable(
g3_mod1_mae_in %>% left_join(g3_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Group 3 - Model 1 "
)%>%
  kable_styling(bootstrap_options = 'condensed')
```

```
#visualize
grid.arrange(
ggplot(data = train_nyc, aes(color = product_type_regroup3,x = days_prior,y = cxl_rate))+
  stat_summary(fun.y = 'mean', geom = 'line') +
  scale_color_brewer(palette = "Paired") +theme_classic() + labs(x = 'Days Prior', y = 'Cancellation Ra
ggplot(data = train_nyc, aes(color = product_type_regroup3,x = days_prior,y = predict(g3_mod1, train_ny
  stat_summary(fun.y = 'mean', geom = 'line')+
  scale_color_brewer(palette = "Paired")+theme_classic() + labs(x = 'Days Prior', y = 'Predicted Cancel
nrow = 2)
```



Group 4

Model 1

- based on the relationship between days_prior and cxl rate, we power days prior by 0.2
- relationship with days prior has interaction with product type
- coefficient between cxl rate and other variables across product type groups do not differ much, and they are equally weak (<=0.1)

```r
# model 1
g4_mod1 <- lm(cxl_rate ~ cummulative_gross_bookings +product_type_regroup4 *dow +I(days_prior^0.2) +
               product_type_regroup4 * days_prior+ last_week*cummulative_gross_bookings
             , data = train_nyc)
summary(g4_mod1)
```

```
##
## Call:
## lm(formula = cxl_rate ~ cummulative_gross_bookings + product_type_regroup4 *
##     dow + I(days_prior^0.2) + product_type_regroup4 * days_prior +
##     last_week * cummulative_gross_bookings, data = train_nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25389 -0.07126 -0.03443  0.02731  0.97539
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       -6.310e-02  1.684e-02
## cummulative_gross_bookings                        -1.105e-04  1.833e-05
## product_type_regroup4Low_cxl                      -1.293e-02  6.179e-03
## product_type_regroup4Med_cxl                       6.980e-03  5.953e-03
## dowMon                                            -5.076e-03  5.358e-03
## dowSat                                            -1.744e-02  5.207e-03
## dowSun                                             1.107e-02  5.207e-03
## dowThu                                             3.887e-02  5.221e-03
## dowTue                                             1.794e-03  5.373e-03
## dowWed                                             3.978e-03  5.226e-03
## I(days_prior^0.2)                                  1.005e-01  1.279e-02
## days_prior                                         4.877e-04  1.935e-04
## last_weeknot_last_week                             2.084e-02  5.106e-03
## product_type_regroup4Low_cxl:dowMon                1.780e-02  7.205e-03
## product_type_regroup4Med_cxl:dowMon               -2.311e-02  7.098e-03
## product_type_regroup4Low_cxl:dowSat                3.184e-02  6.946e-03
## product_type_regroup4Med_cxl:dowSat                1.532e-02  6.889e-03
## product_type_regroup4Low_cxl:dowSun               -2.479e-02  6.967e-03
## product_type_regroup4Med_cxl:dowSun               -4.909e-02  6.890e-03
## product_type_regroup4Low_cxl:dowThu               -5.639e-02  6.959e-03
## product_type_regroup4Med_cxl:dowThu               -3.687e-02  6.910e-03
## product_type_regroup4Low_cxl:dowTue               -8.872e-03  7.168e-03
## product_type_regroup4Med_cxl:dowTue               -3.765e-02  7.116e-03
## product_type_regroup4Low_cxl:dowWed               -1.485e-02  6.967e-03
## product_type_regroup4Med_cxl:dowWed               -2.337e-02  6.922e-03
## product_type_regroup4Low_cxl:days_prior           -2.724e-03  1.169e-04
## product_type_regroup4Med_cxl:days_prior           -3.426e-04  1.112e-04
## cummulative_gross_bookings:last_weeknot_last_week -2.058e-04  2.021e-05
##                                                   t value Pr(>|t|)
## (Intercept)                                        -3.747 0.000179 ***
## cummulative_gross_bookings                         -6.028 1.68e-09 ***
## product_type_regroup4Low_cxl                       -2.093 0.036361 *
## product_type_regroup4Med_cxl                        1.172 0.241042
## dowMon                                             -0.947 0.343484
## dowSat                                             -3.350 0.000809 ***
## dowSun                                              2.125 0.033599 *
```

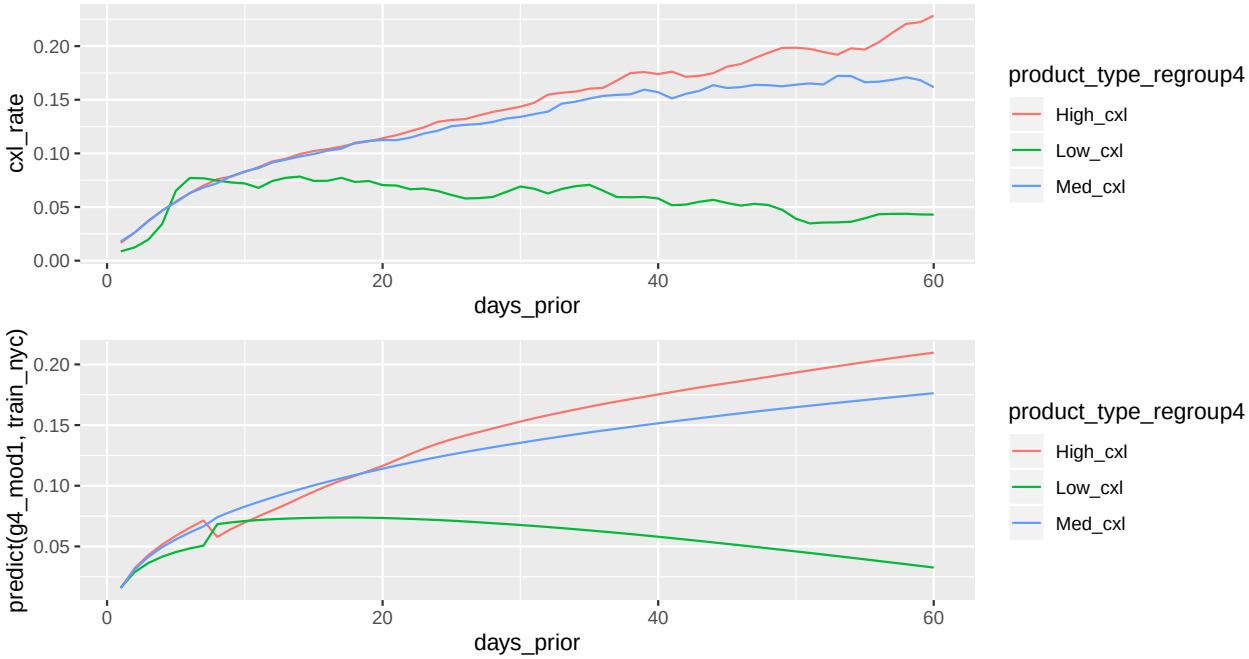
```
## dowThu                                                 7.445 9.85e-14 ***
## dowTue                                                 0.334 0.738423
## dowWed                                                 0.761 0.446533
## I(days_prior^0.2)                                      7.864 3.82e-15 ***
## days_prior                                             2.520 0.011739 *
## last_weeknot_last_week                                 4.082 4.48e-05 ***
## product_type_regroup4Low_cxl:dowMon                    2.470 0.013515 *
## product_type_regroup4Med_cxl:dowMon                   -3.256 0.001130 **
## product_type_regroup4Low_cxl:dowSat                    4.584 4.58e-06 ***
## product_type_regroup4Med_cxl:dowSat                    2.224 0.026128 *
## product_type_regroup4Low_cxl:dowSun                   -3.558 0.000374 ***
## product_type_regroup4Med_cxl:dowSun                   -7.125 1.05e-12 ***
## product_type_regroup4Low_cxl:dowThu                   -8.103 5.52e-16 ***
## product_type_regroup4Med_cxl:dowThu                   -5.336 9.57e-08 ***
## product_type_regroup4Low_cxl:dowTue                   -1.238 0.215822
## product_type_regroup4Med_cxl:dowTue                   -5.291 1.22e-07 ***
## product_type_regroup4Low_cxl:dowWed                   -2.131 0.033074 *
## product_type_regroup4Med_cxl:dowWed                   -3.377 0.000734 ***
## product_type_regroup4Low_cxl:days_prior              -23.300  < 2e-16 ***
## product_type_regroup4Med_cxl:days_prior               -3.081 0.002067 **
## cummulative_gross_bookings:last_weeknot_last_week -10.184  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1562 on 43892 degrees of freedom
## Multiple R-squared:  0.1214, Adjusted R-squared:  0.1209
## F-statistic: 224.6 on 27 and 43892 DF,  p-value: < 2.2e-16
```

```
# TRAIN (in-group)
# MAE
g4_mod1_mae_in = mae_by_days(train_nyc,g4_mod1, 'mae_g4_mod1_in')
#MAPE
g4_mod1_mape_in = mape_by_days(train_nyc,g4_mod1, 'mape_g4_mod1_in')
#MASE
g4_mod1_mase_in = mase_by_days(train_nyc,g4_mod1, 'mase_g4_mod1_in')

# TEST (out-group)
# MAE
g4_mod1_mae_out = mae_by_days(test_nyc,g4_mod1, 'mae_g4_mod1_out')
#MAPE
g4_mod1_mape_out = mape_by_days(test_nyc,g4_mod1, 'mape_g4_mod1_out')
#MASE
g4_mod1_mase_out = mase_by_days(test_nyc,g4_mod1, 'mase_g4_mod1_out')

kable(
g1_mod1_mae_in %>% left_join(g4_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Group 4 - Model 1 "
)%>%
  kable_styling(bootstrap_options = 'condensed')
```

Table 5: Group 4 - Model 1

days_prior_cat	mae_g1_mod1_in	mape_g4_mod1_in	mase_g4_mod1_in	mae_g4_mod1_out	mape_g4_m
Day 01-07	1.9224	0.0449	0.3485	2.7472	
Day 08-14	2.2493	0.0752	0.9409	3.9995	
Day 15-20	2.1943	0.0830	0.9365	3.1982	
Day 21-27	2.0131	0.0965	0.9269	3.1047	
Day 28-60	1.5759	0.1371	0.9639	3.1245	

```
#visualize
grid.arrange(
ggplot(data = train_nyc, aes(color = product_type_regroup4,x = days_prior,y = cxl_rate))+
  stat_summary(fun.y = 'mean', geom = 'line'),
ggplot(data = train_nyc, aes(color = product_type_regroup4,x = days_prior,y = predict(g4_mod1, train_ny
  stat_summary(fun.y = 'mean', geom = 'line'), nrow = 2)
```



Group 6

Model 1

- based on the relationship between days_prior and cxl rate, we power days prior by 0.2
- relationship with days prior has interaction with product type
- coefficient between cxl rate and other variables across product type groups do not differ much, and they are equally weak (<=0.1)

```
# model 1
g6_mod1 <- lm(cxl_rate ~  cummulative_gross_bookings + product_type *dow_regroup +I(days_prior^0.2) + pr
              , data = train_nyc)
summary(g6_mod1)
```

```
##
## Call:
## lm(formula = cxl_rate ~ cummulative_gross_bookings + product_type *
##     dow_regroup + I(days_prior^0.2) + product_type * days_prior,
##     data = train_nyc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30816 -0.06618 -0.02136  0.02289  0.97008
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       -9.476e-02  1.166e-02
## cummulative_gross_bookings                        -1.042e-04  1.093e-05
## product_typeCORPORATE                              3.069e-02  7.553e-03
## product_typeFENCED                                -1.922e-02  7.383e-03
## product_typeGOVERNMENT                             9.222e-02  7.594e-03
## product_typeGROUP                                 -2.763e-02  7.717e-03
## product_typeMEMBERSHIP MARKETING                   1.150e-02  7.383e-03
## product_typeOPAQUE                                -1.767e-02  7.481e-03
## product_typeOTHER                                 -5.356e-02  7.732e-03
## product_typeTACTICAL MARKETING                     5.238e-04  7.401e-03
## product_typeUNFENCED                               3.402e-02  7.533e-03
## product_typeWHOLESALE                              2.985e-02  7.384e-03
## dow_regroupweekend                                -1.895e-02  4.759e-03
## I(days_prior^0.2)                                  1.117e-01  7.383e-03
## days_prior                                         2.494e-03  1.823e-04
## product_typeCORPORATE:dow_regroupweekend           -2.285e-02  6.731e-03
## product_typeFENCED:dow_regroupweekend              1.350e-03  6.747e-03
## product_typeGOVERNMENT:dow_regroupweekend          1.461e-02  6.931e-03
## product_typeGROUP:dow_regroupweekend               3.900e-02  6.742e-03
## product_typeMEMBERSHIP MARKETING:dow_regroupweekend  2.424e-02  6.750e-03
## product_typeOPAQUE:dow_regroupweekend              1.390e-02  6.802e-03
## product_typeOTHER:dow_regroupweekend               7.244e-02  6.957e-03
## product_typeTACTICAL MARKETING:dow_regroupweekend  2.100e-02  6.727e-03
## product_typeUNFENCED:dow_regroupweekend            1.956e-02  6.728e-03
## product_typeWHOLESALE:dow_regroupweekend           4.774e-02  6.727e-03
## product_typeCORPORATE:days_prior                  -1.658e-03  1.949e-04
## product_typeFENCED:days_prior                     -4.315e-03  1.928e-04
## product_typeGOVERNMENT:days_prior                 -5.482e-03  1.984e-04
## product_typeGROUP:days_prior                      -1.430e-03  1.957e-04
## product_typeMEMBERSHIP MARKETING:days_prior       -1.879e-03  1.928e-04
## product_typeOPAQUE:days_prior                     -4.415e-03  1.938e-04
## product_typeOTHER:days_prior                      -4.594e-03  2.002e-04
## product_typeTACTICAL MARKETING:days_prior         -4.428e-03  1.929e-04
## product_typeUNFENCED:days_prior                   -2.063e-03  1.940e-04
## product_typeWHOLESALE:days_prior                  -2.738e-03  1.928e-04
##                                                   t value Pr(>|t|)
## (Intercept)                                        -8.125 4.60e-16 ***
## cummulative_gross_bookings                         -9.534  < 2e-16 ***
## product_typeCORPORATE                               4.063 4.85e-05 ***
## product_typeFENCED                                 -2.603 0.009237 **
## product_typeGOVERNMENT                             12.143  < 2e-16 ***
## product_typeGROUP                                  -3.580 0.000344 ***
```

```
## product_typeMEMBERSHIP MARKETING                        1.558 0.119294
## product_typeOPAQUE                                     -2.362 0.018185 *
## product_typeOTHER                                      -6.927 4.35e-12 ***
## product_typeTACTICAL MARKETING                          0.071 0.943580
## product_typeUNFENCED                                    4.516 6.31e-06 ***
## product_typeWHOLESALE                                   4.043 5.29e-05 ***
## dow_regroupweekend                                     -3.982 6.86e-05 ***
## I(days_prior^0.2)                                      15.127  < 2e-16 ***
## days_prior                                             13.679  < 2e-16 ***
## product_typeCORPORATE:dow_regroupweekend                -3.394 0.000689 ***
## product_typeFENCED:dow_regroupweekend                   0.200 0.841407
## product_typeGOVERNMENT:dow_regroupweekend                2.108 0.035050 *
## product_typeGROUP:dow_regroupweekend                     5.785 7.29e-09 ***
## product_typeMEMBERSHIP MARKETING:dow_regroupweekend      3.591 0.000330 ***
## product_typeOPAQUE:dow_regroupweekend                    2.044 0.040959 *
## product_typeOTHER:dow_regroupweekend                    10.412  < 2e-16 ***
## product_typeTACTICAL MARKETING:dow_regroupweekend        3.122 0.001798 **
## product_typeUNFENCED:dow_regroupweekend                  2.907 0.003650 **
## product_typeWHOLESALE:dow_regroupweekend                 7.096 1.30e-12 ***
## product_typeCORPORATE:days_prior                        -8.510  < 2e-16 ***
## product_typeFENCED:days_prior                          -22.387  < 2e-16 ***
## product_typeGOVERNMENT:days_prior                      -27.630  < 2e-16 ***
## product_typeGROUP:days_prior                            -7.307 2.78e-13 ***
## product_typeMEMBERSHIP MARKETING:days_prior             -9.748  < 2e-16 ***
## product_typeOPAQUE:days_prior                          -22.784  < 2e-16 ***
## product_typeOTHER:days_prior                           -22.946  < 2e-16 ***
## product_typeTACTICAL MARKETING:days_prior              -22.951  < 2e-16 ***
## product_typeUNFENCED:days_prior                        -10.631  < 2e-16 ***
## product_typeWHOLESALE:days_prior                       -14.203  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1508 on 43885 degrees of freedom
## Multiple R-squared:  0.1817, Adjusted R-squared:  0.1811
## F-statistic: 286.6 on 34 and 43885 DF,  p-value: < 2.2e-16
```

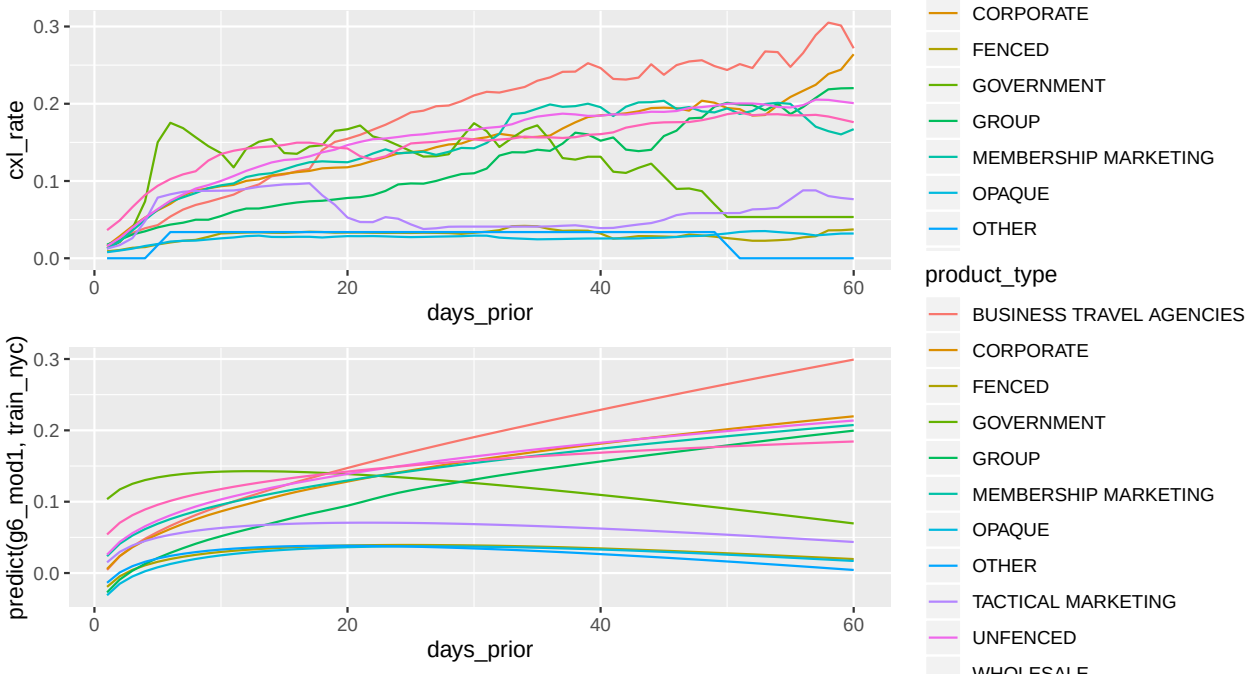
```
# TRAIN (in-group)
# MAE
g6_mod1_mae_in = mae_by_days(train_nyc,g6_mod1, 'mae_g6_mod1_in')
#MAPE
g6_mod1_mape_in = mape_by_days(train_nyc,g6_mod1, 'mape_g6_mod1_in')
#MASE
g6_mod1_mase_in = mase_by_days(train_nyc,g6_mod1, 'mase_g6_mod1_in')

# TEST (out-group)
# MAE
g6_mod1_mae_out = mae_by_days(test_nyc,g6_mod1, 'mae_g6_mod1_out')
#MAPE
g6_mod1_mape_out = mape_by_days(test_nyc,g6_mod1, 'mape_g6_mod1_out')
#MASE
g6_mod1_mase_out = mase_by_days(test_nyc,g6_mod1, 'mase_g6_mod1_out')

kable(
g1_mod1_mae_in %>% left_join(g6_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
```

Table 6: Group 6 - Model 1

days_prior_cat	mae_g1_mod1_in	mape_g6_mod1_in	mase_g6_mod1_in	mae_g6_mod1_out	mape_g6_m
Day 01-07	1.9224	0.0455	0.4205	3.7379	
Day 08-14	2.2493	0.0627	0.5827	2.9722	
Day 15-20	2.1943	0.0702	0.6243	2.7383	
Day 21-27	2.0131	0.0840	0.6527	2.7696	
Day 28-60	1.5759	0.1242	0.7010	2.6775	

```
  left_join(g6_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Group 6 - Model 1 "
)%>%
  kable_styling(bootstrap_options = 'condensed')
```

```
#visualize
grid.arrange(
ggplot(data = train_nyc, aes(color = product_type,x = days_prior,y = cxl_rate))+
  stat_summary(fun.y = 'mean', geom = 'line'),
ggplot(data = train_nyc, aes(color = product_type,x = days_prior,y = predict(g6_mod1, train_nyc)))+
  stat_summary(fun.y = 'mean', geom = 'line'), nrow = 2)
```



Metrics Results

```
# IN SAMPLE
# mae
```

Table 7: In-sample metrics - MAE

days_prior_cat	mae_g1_mod2_in	mae_g2_mod1_in	mae_g3_mod1_in	mae_g4_mod1_in	mae_g6_mod1_
Day 01-07	1.9198	2.0879	2.1189	2.3808	2.87
Day 08-14	2.2420	2.2341	2.1661	3.7906	2.34
Day 15-20	2.2033	2.0927	2.1326	3.2709	2.18
Day 21-27	2.0280	1.8802	1.9101	2.8003	1.97
Day 28-60	1.5892	1.3754	1.4212	2.0504	1.49

Table 8: In-sample metrics - MAPE

days_prior_cat	mape_g1_mod2_in	mape_g2_mod1_in	mape_g3_mod1_in	mape_g4_mod1_in	mape_g6_
Day 01-07	0.0416	0.0424	0.0426	0.0449	
Day 08-14	0.0648	0.0647	0.0656	0.0752	
Day 15-20	0.0732	0.0723	0.0745	0.0830	
Day 21-27	0.0867	0.0862	0.0880	0.0965	
Day 28-60	0.1258	0.1242	0.1253	0.1371	

```
kable(
g1_mod2_mae_in %>%
  left_join(g2_mod1_mae_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mae_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mae_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mae_in, c("days_prior_cat" = "days_prior_cat")),
  caption = "In-sample metrics - MAE"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
#mape
kable(
g1_mod2_mape_in %>%
  left_join(g2_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mape_in, c("days_prior_cat" = "days_prior_cat")),
  caption = "In-sample metrics - MAPE"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
#mase
kable(
g1_mod2_mase_in %>%
  left_join(g2_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mase_in, c("days_prior_cat" = "days_prior_cat")),
  caption = "In-sample metrics - MASE"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
# OUT SAMPLE
kable(
```

Table 9: In-sample metrics - MASE

days_prior_cat	mase_g1_mod2_in	mase_g2_mod1_in	mase_g3_mod1_in	mase_g4_mod1_in	mase_g6_mo
Day 01-07	0.2810	0.3056	0.3101	0.3485	
Day 08-14	0.5565	0.5546	0.5377	0.9409	
Day 15-20	0.6308	0.5991	0.6106	0.9365	
Day 21-27	0.6713	0.6223	0.6322	0.9269	
Day 28-60	0.7471	0.6466	0.6681	0.9639	

Table 10: Out-sample metrics - MAE

days_prior_cat	mae_g1_mod2	mae_g2_mod1_out	mae_g3_mod1_out	mae_g4_mod1_out	mae_g6_mod1_
Day 01-07	2.5653	2.6335	2.7497	2.7472	3.7
Day 08-14	2.8751	2.8816	2.7940	3.9995	2.9
Day 15-20	2.6975	2.7748	2.8703	3.1982	2.7
Day 21-27	2.6119	2.7546	2.8481	3.1047	2.7
Day 28-60	2.6052	2.4932	2.5508	3.1245	2.6

```
g1_mod2_mae_out %>%
  left_join(g2_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mae_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Out-sample metrics - MAE"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
kable(
g1_mod2_mape_out %>%
  left_join(g2_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g6_mod1_mape_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Out-sample metrics - MAPE"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
kable(
g1_mod2_mase_out %>%
  left_join(g2_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g3_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(g4_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")) %>%
```

Table 11: Out-sample metrics - MAPE

days_prior_cat	mape_g1_mod2	mape_g2_mod1_out	mape_g3_mod1_out	mape_g4_mod1_out	mape_g6_
Day 01-07	0.0352	0.0363	0.0360	0.0345	
Day 08-14	0.0527	0.0525	0.0529	0.0574	
Day 15-20	0.0600	0.0603	0.0612	0.0629	
Day 21-27	0.0753	0.0751	0.0773	0.0795	
Day 28-60	0.1258	0.1272	0.1303	0.1360	

Table 12: Out-sample metrics - MASE

days_prior_cat	mase_g1_mod2	mase_g2_mod1_out	mase_g3_mod1_out	mase_g4_mod1_out	mase_g6_m
Day 01-07	0.2921	0.2999	0.3131	0.3128	
Day 08-14	0.4615	0.4625	0.4485	0.6420	
Day 15-20	0.5322	0.5475	0.5663	0.6310	
Day 21-27	0.5799	0.6116	0.6324	0.6893	
Day 28-60	0.6435	0.6158	0.6301	0.7717	

```
  left_join(g6_mod1_mase_out, c("days_prior_cat" = "days_prior_cat")),
  caption = "Out-sample metrics - MASE"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
train_regre_pred <- data.frame(predict(g3_mod1, train_nyc))
names(train_regre_pred)[1] <- "regression_pred"

test_regre_pred <- data.frame(predict(g3_mod1, test_nyc))
names(test_regre_pred)[1] <- "regression_pred"

write.csv(train_regre_pred, file = "../../treated data/train_regre_pred.csv")
write.csv(test_regre_pred, file = "../../treated data/test_regre_pred.csv")
```