NYC Traditional

Hannah Khuong

Contents

Traditional Model is industry standard model. In this project, traditional predictions for testing data are means of cancellation rate by Days Prior and Product Types / Product Type Regrouped in training data.

Model

Group 1

```
# TRAIN_NYC
# Prediction of cancellation rate by product type and days prior (on train_nyc)

# Calculate model 1's cxl rate
# Average cxl rate by days prior and by product type (on train_nyc)
train_nyc <- train_nyc %>%
  group_by(days_prior, product_type_regroup)%>%
  mutate(mod1_cxl_rate = mean(cxl_rate))%>%
  mutate(mod1_survive = OTB - mod1_cxl_rate * OTB)

# TEST_NYC
# Make cxl rate table to join predition to test_nyc
mod1_cxl_rate_table <- train_nyc %>%
  select(days_prior, product_type_regroup, cxl_rate) %>%
  group_by(days_prior, product_type_regroup) %>%
  summarise(mod1_cxl_rate = mean(cxl_rate))

# Join cxl rate table to test_nyc
test_nyc <- test_nyc %>%
  left_join(mod1_cxl_rate_table, c("days_prior"="days_prior", "product_type_regroup"="product_type_regr
  mutate(mod1_survive = OTB - mod1_cxl_rate * OTB)
```

```
# In-sample
## MAE
mod1_MAE_in <- train_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod1_survive - OTB_to_survive)))
```

Table 1: In-sample metrics - Model 1

days_prior_cat	MAE	MAPE	MASE
Day 01-07	1.827776	0.0397380	0.2675097
Day 08-14	2.288588	0.0654104	0.5680857
Day 15-20	2.160051	0.0728238	0.6184229
Day 21-27	1.837865	0.0851900	0.6083262
Day 28-60	1.405906	0.1255250	0.6609618

```
## MAPE
mod1_MAPE_in <- train_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod1_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod1_MASE_in <- train_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod1_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

kable(
mod1_MAE_in %>%
  left_join(mod1_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod1_MASE_in, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - Model 1"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
# Out-sample
## MAE
mod1_MAE_out <- test_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod1_survive - OTB_to_survive)))

## MAPE
mod1_MAPE_out <- test_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod1_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod1_MASE_out <- test_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod1_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

# Metrics table
kable(
mod1_MAE_out %>%
  left_join(mod1_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
```

Table 2: Out-sample metrics - Model 1

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.387535	0.0324971	0.2718865
Day 08-14	3.459872	0.0552561	0.5553543
Day 15-20	3.028246	0.0616108	0.5974813
Day 21-27	2.707730	0.0752866	0.6011950
Day 28-60	2.457607	0.1260883	0.6070289

```
  left_join(mod1_MASE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - Model 1"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Group 2

```
# TRAIN_NYC
# Prediction of cancellation rate by product type and days prior (on train_nyc)

# Calculate model 2's cxl rate
# Average cxl rate by days prior and by product type (on train_nyc)
train_nyc <- train_nyc %>%
  group_by(days_prior, product_type_regroup2)%>%
  mutate(mod2_cxl_rate = mean(cxl_rate))%>%
  mutate(mod2_survive = OTB - mod2_cxl_rate * OTB)

# TEST_NYC
# Make cxl rate table to join predition to test_nyc
mod2_cxl_rate_table <- train_nyc %>%
  select(days_prior, product_type_regroup2, cxl_rate) %>%
  group_by(days_prior, product_type_regroup2) %>%
  summarise(mod2_cxl_rate = mean(cxl_rate))

# Join cxl rate table to test_nyc
test_nyc <- test_nyc %>%
  left_join(mod2_cxl_rate_table,
            c("days_prior"="days_prior", "product_type_regroup2"="product_type_regroup2")) %>%
  mutate(mod2_survive = OTB - mod2_cxl_rate * OTB)
```

```
# In-sample
## MAE
mod2_MAE_in <- train_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod2_survive - OTB_to_survive)))

## MAPE
mod2_MAPE_in <- train_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod2_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))
```

Table 3: In-sample metrics - Model 2

days_prior_cat	MAE	MAPE	MASE
Day 01-07	1.685583	0.0389763	0.2466986
Day 08-14	2.190970	0.0643804	0.5438545
Day 15-20	2.156842	0.0718734	0.6175043
Day 21-27	1.898972	0.0854067	0.6285526
Day 28-60	1.436970	0.1245059	0.6755661

```
## MASE
mod2_MASE_in <- train_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod2_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

kable(
mod2_MAE_in %>%
  left_join(mod2_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod2_MASE_in, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - Model 2"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
# Out-sample
## MAE
mod2_MAE_out <- test_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod2_survive - OTB_to_survive)))

## MAPE
mod2_MAPE_out <- test_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod2_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod2_MASE_out <- test_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod2_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

# Metrics table
kable(
mod2_MAE_out %>%
  left_join(mod2_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod2_MASE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - Model 2"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Table 4: Out-sample metrics - Model 2

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.215237	0.0322057	0.2522657
Day 08-14	3.411809	0.0544318	0.5476396
Day 15-20	3.105898	0.0617076	0.6128023
Day 21-27	2.843179	0.0742573	0.6312687
Day 28-60	2.414521	0.1269974	0.5963866

Group 3

```
# TRAIN_NYC
# Prediction of cancellation rate by product type and days prior (on train_nyc)

# Calculate model 3's cxl rate
# Average cxl rate by days prior and by product type (on train_nyc)
train_nyc <- train_nyc %>%
  group_by(days_prior, product_type_regroup3)%>%
  mutate(mod3_cxl_rate = mean(cxl_rate))%>%
  mutate(mod3_survive = OTB - mod3_cxl_rate * OTB)

# TEST_NYC
# Make cxl rate table to join predition to test_nyc
mod3_cxl_rate_table <- train_nyc %>%
  select(days_prior, product_type_regroup3, cxl_rate) %>%
  group_by(days_prior, product_type_regroup3) %>%
  summarise(mod3_cxl_rate = mean(cxl_rate))

# Join cxl rate table to test_nyc
test_nyc <- test_nyc %>%
  left_join(mod3_cxl_rate_table,
            c("days_prior"="days_prior", "product_type_regroup3"="product_type_regroup3")) %>%
  mutate(mod3_survive = OTB - mod3_cxl_rate * OTB)
```

```
# In-sample
## MAE
mod3_MAE_in <- train_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod3_survive - OTB_to_survive)))

## MAPE
mod3_MAPE_in <- train_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod3_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod3_MASE_in <- train_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod3_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))
```

Table 5: In-sample metrics - Model 3

days_prior_cat	MAE	MAPE	MASE
Day 01-07	1.760044	0.0394568	0.2575967
Day 08-14	2.356356	0.0660062	0.5849075
Day 15-20	2.349338	0.0745953	0.6726159
Day 21-27	2.001923	0.0874932	0.6626289
Day 28-60	1.472426	0.1254780	0.6922352

Table 6: Out-sample metrics - Model 3

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.319891	0.0323000	0.2641834
Day 08-14	3.535673	0.0555164	0.5675213
Day 15-20	3.190198	0.0624065	0.6294349
Day 21-27	2.894829	0.0766988	0.6427364
Day 28-60	2.432627	0.1300724	0.6008586

```
kable(
mod3_MAE_in %>%
  left_join(mod3_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MASE_in, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - Model 3"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
# Out-sample
## MAE
mod3_MAE_out <- test_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod3_survive - OTB_to_survive)))

## MAPE
mod3_MAPE_out <- test_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod3_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod3_MASE_out <- test_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod3_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

# Metrics table
kable(
mod3_MAE_out %>%
  left_join(mod3_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MASE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - Model 3"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Group 4

```
# TRAIN_NYC
# Prediction of cancellation rate by product type and days prior (on train_nyc)

# Calculate model 4's cxl rate
# Average cxl rate by days prior and by product type (on train_nyc)
train_nyc <- train_nyc %>%
  group_by(days_prior, product_type_regroup4)%>%
  mutate(mod4_cxl_rate = mean(cxl_rate))%>%
  mutate(mod4_survive = OTB - mod4_cxl_rate * OTB)

# TEST_NYC
# Make cxl rate table to join predition to test_nyc
mod4_cxl_rate_table <- train_nyc %>%
  select(days_prior, product_type_regroup4, cxl_rate) %>%
  group_by(days_prior, product_type_regroup4) %>%
  summarise(mod4_cxl_rate = mean(cxl_rate))

# Join cxl rate table to test_nyc
test_nyc <- test_nyc %>%
  left_join(mod4_cxl_rate_table,
            c("days_prior"="days_prior", "product_type_regroup4"="product_type_regroup4")) %>%
  mutate(mod4_survive = OTB - mod4_cxl_rate * OTB)
```

```
# In-sample
## MAE
mod4_MAE_in <- train_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod4_survive - OTB_to_survive)))

## MAPE
mod4_MAPE_in <- train_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod4_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod4_MASE_in <- train_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod4_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

kable(
mod4_MAE_in %>%
  left_join(mod4_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MASE_in, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - Model 4"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
# Out-sample
## MAE
mod4_MAE_out <- test_nyc %>% group_by(days_prior_cat) %>%
```

Table 7: In-sample metrics - Model 4

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.169122	0.0429586	0.3174685
Day 08-14	3.195795	0.0735887	0.7932777
Day 15-20	3.313679	0.0837860	0.9487070
Day 21-27	3.010936	0.0974081	0.9966082
Day 28-60	2.454313	0.1410874	1.1538521

Table 8: Out-sample metrics - Model 4

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.640365	0.0340604	0.3006781
Day 08-14	4.584600	0.0607919	0.7358877
Day 15-20	4.723256	0.0691062	0.9319115
Day 21-27	4.677099	0.0854692	1.0384524
Day 28-60	4.761807	0.1441918	1.1761661

```
  summarise(MAE = mean(abs(mod4_survive - OTB_to_survive)))

## MAPE
mod4_MAPE_out <- test_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod4_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod4_MASE_out <- test_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod4_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

# Metrics table
kable(
mod4_MAE_out %>%
  left_join(mod4_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MASE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - Model 4"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Benchmark (original product types)

```
# TRAIN_NYC
# Prediction of cancellation rate by product type and days prior (on train_nyc)

# Calculate model bm's cxl rate
# Average cxl rate by days prior and by product type (on train_nyc)
train_nyc <- train_nyc %>%
```

```
  group_by(days_prior, product_type)%>%
  mutate(modbm_cxl_rate = mean(cxl_rate))%>%
  mutate(modbm_survive = OTB - modbm_cxl_rate * OTB)

# TEST_NYC
# Make cxl rate table to join predition to test_nyc
modbm_cxl_rate_table <- train_nyc %>%
  select(days_prior, product_type, cxl_rate) %>%
  group_by(days_prior, product_type) %>%
  summarise(modbm_cxl_rate = mean(cxl_rate))

# Join cxl rate table to test_nyc
test_nyc <- test_nyc %>%
  left_join(modbm_cxl_rate_table,
            c("days_prior"="days_prior", "product_type"="product_type")) %>%
  mutate(modbm_survive = OTB - modbm_cxl_rate * OTB)
```

```
# In-sample
## MAE
modbm_MAE_in <- train_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(modbm_survive - OTB_to_survive)))

## MAPE
modbm_MAPE_in <- train_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((modbm_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
modbm_MASE_in <- train_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - modbm_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))


modbm_MAE_in %>%
  left_join(modbm_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MASE_in, c("days_prior_cat" = "days_prior_cat"))
```

```
## # A tibble: 5 x 4
##   days_prior_cat   MAE   MAPE  MASE
##   <fct>          <dbl>  <dbl> <dbl>
## 1 Day 01-07       1.51 0.0376 0.221
## 2 Day 08-14       2.00 0.0626 0.497
## 3 Day 15-20       2.02 0.0697 0.578
## 4 Day 21-27       1.80 0.0821 0.596
## 5 Day 28-60       1.55 0.125  0.726
```

```
# Out-sample
## MAE
modbm_MAE_out <- test_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(modbm_survive - OTB_to_survive)))
```

```
## MAPE
modbm_MAPE_out <- test_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((modbm_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
modbm_MASE_out <- test_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - modbm_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

# Metrics table

modbm_MAE_out %>%
  left_join(modbm_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MASE_out, c("days_prior_cat" = "days_prior_cat"))
```

```
## # A tibble: 5 x 4
##   days_prior_cat   MAE   MAPE  MASE
##   <fct>          <dbl>  <dbl> <dbl>
## 1 Day 01-07       1.85 0.0310 0.210
## 2 Day 08-14       2.63 0.0496 0.422
## 3 Day 15-20       2.65 0.0571 0.522
## 4 Day 21-27       2.69 0.0710 0.597
## 5 Day 28-60       2.56 0.127  0.634
```

No product grouping

```
# TRAIN_NYC
# Prediction of cancellation rate by product type and days prior (on train_nyc)

# Calculate model 5's cxl rate
# Average cxl rate by days prior and by product type (on train_nyc)
train_nyc <- train_nyc %>%
  group_by(days_prior)%>%
  mutate(mod5_cxl_rate = mean(cxl_rate))%>%
  mutate(mod5_survive = OTB - mod5_cxl_rate * OTB)

# TEST_NYC
# Make cxl rate table to join predition to test_nyc
mod5_cxl_rate_table <- train_nyc %>%
  select(days_prior, cxl_rate) %>%
  group_by(days_prior) %>%
  summarise(mod5_cxl_rate = mean(cxl_rate))

# Join cxl rate table to test_nyc
test_nyc <- test_nyc %>%
  left_join(mod5_cxl_rate_table,
            c("days_prior"="days_prior")) %>%
  mutate(mod5_survive = OTB - mod5_cxl_rate * OTB)
```

Table 9: In-sample metrics - Model 5

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.160147	0.0433695	0.3161548
Day 08-14	3.193365	0.0749007	0.7926745
Day 15-20	3.261915	0.0865161	0.9338870
Day 21-27	2.970904	0.1025268	0.9833578
Day 28-60	2.282424	0.1477814	1.0730415

```
# In-sample
## MAE
mod5_MAE_in <- train_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod5_survive - OTB_to_survive)))

## MAPE
mod5_MAPE_in <- train_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod5_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod5_MASE_in <- train_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod5_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))

kable(
mod5_MAE_in %>%
  left_join(mod5_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod5_MASE_in, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - Model 5"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
# Out-sample
## MAE
mod5_MAE_out <- test_nyc %>% group_by(days_prior_cat) %>%
  summarise(MAE = mean(abs(mod5_survive - OTB_to_survive)))

## MAPE
mod5_MAPE_out <- test_nyc %>%
  filter(OTB_to_survive != 0) %>%
  group_by(days_prior_cat) %>%
  summarise(MAPE = mean(abs((mod5_survive  - (OTB_to_survive)))/
                          (OTB_to_survive)))

## MASE
mod5_MASE_out <- test_nyc %>%
  group_by(days_prior_cat) %>%
  summarise(MASE = sum(abs(OTB_to_survive - mod5_survive))/
              sum(abs(OTB_to_survive - naive_survive_pred)))
```

Table 10: Out-sample metrics - Model 5

days_prior_cat	MAE	MAPE	MASE
Day 01-07	2.630673	0.0340747	0.2995745
Day 08-14	4.457273	0.0611716	0.7154501
Day 15-20	4.471857	0.0698780	0.8823097
Day 21-27	4.338367	0.0861207	0.9632441
Day 28-60	4.548237	0.1504068	1.1234143

Table 11: In-sample metrics - MAE

days_prior	Group 1	Group 2	Group 3	Group 4	Benchmark	Baseline
Day 01-07	1.827776	1.685583	1.760044	2.169122	1.513036	2.630673
Day 08-14	2.288588	2.190970	2.356356	3.195795	2.001785	4.457273
Day 15-20	2.160051	2.156842	2.349338	3.313679	2.019937	4.471857
Day 21-27	1.837865	1.898972	2.001923	3.010936	1.800086	4.338367
Day 28-60	1.405906	1.436970	1.472426	2.454313	1.545164	4.548237

```
# Metrics table
kable(
mod5_MAE_out %>%
  left_join(mod5_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod5_MASE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - Model 5"
)%>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Metrics Summary

In-sample

```
kable(
mod1_MAE_in %>%
  left_join(mod2_MAE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MAE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MAE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MAE_in, c("days_prior_cat" = "days_prior_cat"))%>%
  left_join(mod5_MAE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - MAE",
  col.names = c("days_prior", "Group 1", "Group 2", "Group 3", "Group 4", "Benchmark", "Baseline")
) %>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
kable(
mod1_MAPE_in %>%
  left_join(mod2_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MAPE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MAPE_in, c("days_prior_cat" = "days_prior_cat"))%>%
```

Table 12: In-sample metrics - MAPE

days_prior	Group 1	Group 2	Group 3	Group 4	Benchmark	Baseline
Day 01-07	0.0397380	0.0389763	0.0394568	0.0429586	0.0375723	2.630673
Day 08-14	0.0654104	0.0643804	0.0660062	0.0735887	0.0625861	4.457273
Day 15-20	0.0728238	0.0718734	0.0745953	0.0837860	0.0697334	4.471857
Day 21-27	0.0851900	0.0854067	0.0874932	0.0974081	0.0820558	4.338367
Day 28-60	0.1255250	0.1245059	0.1254780	0.1410874	0.1246017	4.548237

Table 13: In-sample metrics - MASE

days_prior	Group 1	Group 2	Group 3	Group 4	Benchmark	Baseline
Day 01-07	0.2675097	0.2466986	0.2575967	0.3174685	0.2214451	2.630673
Day 08-14	0.5680857	0.5438545	0.5849075	0.7932777	0.4968939	4.457273
Day 15-20	0.6184229	0.6175043	0.6726159	0.9487070	0.5783083	4.471857
Day 21-27	0.6083262	0.6285526	0.6626289	0.9966082	0.5958216	4.338367
Day 28-60	0.6609618	0.6755661	0.6922352	1.1538521	0.7264314	4.548237

```
  left_join(mod5_MAE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - MAPE",
  col.names = c("days_prior", "Group 1", "Group 2", "Group 3", "Group 4", "Benchmark", "Baseline")
) %>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
kable(
mod1_MASE_in %>%
  left_join(mod2_MASE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MASE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MASE_in, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MASE_in, c("days_prior_cat" = "days_prior_cat"))%>%
  left_join(mod5_MAE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "In-sample metrics - MASE",
  col.names = c("days_prior", "Group 1", "Group 2", "Group 3", "Group 4", "Benchmark", "Baseline")
) %>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Out sample

```
kable(
mod1_MAE_out %>%
  left_join(mod2_MAE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MAE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MAE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MAE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod5_MAE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - MAE",
  col.names = c("days_prior", "Group 1", "Group 2", "Group 3", "Group 4", "Benchmark", "baseline")
) %>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Table 14: Out-sample metrics - MAE

days_prior	Group 1	Group 2	Group 3	Group 4	Benchmark	baseline
Day 01-07	2.387535	2.215237	2.319891	2.640365	1.845129	2.630673
Day 08-14	3.459872	3.411809	3.535673	4.584600	2.628379	4.457273
Day 15-20	3.028246	3.105898	3.190198	4.723256	2.647036	4.471857
Day 21-27	2.707730	2.843179	2.894829	4.677099	2.690483	4.338367
Day 28-60	2.457607	2.414521	2.432627	4.761807	2.564916	4.548237

Table 15: Out-sample metrics - MAPE

days_prior	Group 1	Group 2	Group 3	Group 4	Benchmark	Baseline
Day 01-07	0.0324971	0.0322057	0.0323000	0.0340604	0.0310414	2.630673
Day 08-14	0.0552561	0.0544318	0.0555164	0.0607919	0.0495565	4.457273
Day 15-20	0.0616108	0.0617076	0.0624065	0.0691062	0.0570629	4.471857
Day 21-27	0.0752866	0.0742573	0.0766988	0.0854692	0.0710436	4.338367
Day 28-60	0.1260883	0.1269974	0.1300724	0.1441918	0.1270305	4.548237

```
kable(
mod1_MAPE_out %>%
  left_join(mod2_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MAPE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MAPE_out, c("days_prior_cat" = "days_prior_cat"))%>%
  left_join(mod5_MAE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - MAPE",
  col.names = c("days_prior", "Group 1", "Group 2", "Group 3", "Group 4", "Benchmark", "Baseline")
) %>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

```
kable(
mod1_MASE_out %>%
  left_join(mod2_MASE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod3_MASE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(mod4_MASE_out, c("days_prior_cat" = "days_prior_cat")) %>%
  left_join(modbm_MASE_out, c("days_prior_cat" = "days_prior_cat"))%>%
  left_join(mod5_MAE_out, c("days_prior_cat" = "days_prior_cat")) ,
  caption = "Out-sample metrics - MASE",
  col.names = c("days_prior", "Group 1", "Group 2", "Group 3", "Group 4", "Benchmark", "Baseline")
) %>%
  kable_styling(bootstrap_options = 'condensed', full_width = F)
```

Table 16: Out-sample metrics - MASE

days_prior	Group 1	Group 2	Group 3	Group 4	Benchmark	Baseline
Day 01-07	0.2718865	0.2522657	0.2641834	0.3006781	0.2101187	2.630673
Day 08-14	0.5553543	0.5476396	0.5675213	0.7358877	0.4218889	4.457273
Day 15-20	0.5974813	0.6128023	0.6294349	0.9319115	0.5222675	4.471857
Day 21-27	0.6011950	0.6312687	0.6427364	1.0384524	0.5973656	4.338367
Day 28-60	0.6070289	0.5963866	0.6008586	1.1761661	0.6335340	4.548237

```r
train_trad_pred <- train_nyc['modbm_cxl_rate']
names(train_trad_pred)[1] <- "traditional_pred"

test_trad_pred <- test_nyc['modbm_cxl_rate']
names(test_trad_pred)[1] <- "traditional_pred"

write.csv(train_trad_pred, file = "../../treated data/train_trad_pred.csv")
write.csv(test_trad_pred, file = "../../treated data/test_trad_pred.csv")
```