Alternate Transformation - New Public/Private Feature

Ryan Rogers

2022-10-10

Concept

The idea behind this notebook is to determine if the process of converting the initial input format (with columns for public and private data) into a new format (with columns for mean/median values and a public/private indicator) yields better results with simple regression models. No other supplementary features will be included at this time. Standard filtering of rows (remove private rows with invalid counts, etc. will be applied.)

Data Load and Transformation

Data Load

```
original_data <- read.csv("priv_mcare_f_pay.csv")
```

Data Filtering

Note: I presume that all data from forbidden MSAs is off limits. Therefore, we will simply drop all rows where priv_count is NA or 0

```
filtered_data <- original_data %>%
  filter(!is.na(priv_count) & (priv_count > 0) & !is.na(lon) & !is.na(lat))
```

Data Preprocessing

One thing we will do before much of the transformation is deal with the categorical variables.

```
filtered_data_important_fields <- filtered_data %>% select(!c(CBSA_NAME, FIPS.State.Code))
cat_encoder <- dummyVars(" ~ .", data=filtered_data_important_fields)
cat_encoded <- data.frame(predict(cat_encoder, filtered_data_important_fields))

cat_encoded$index <- 1:nrow(cat_encoded)
```

Data Splitting

For filters applied, we have:

- Public (mcare):
  - Drop all NAs
- Private:
  - Drop all NAs
  - Drop all rows with no mcare_los
  - Drop all rows with priv_count < 50

For general transformations:

- public_private column added
- SD/IQR columns dropped. They aren't exactly comparable
- Columns renamed for dataset recombination.

```
filtered_data_public <-
  cat_encoded %>%
  select(!c(priv_count, priv_pay_iqr, priv_pay_mean, priv_pay_median, mcare_pay_sd)) %>%
  mutate(public_private = 'public') %>%
  rename(pay_mean = mcare_pay_mean) %>%
  rename(pay_median = mcare_pay_median) %>%
  filter(!is.na(pay_mean))

filtered_data_private <-
  cat_encoded %>%
  filter(priv_count >= 50) %>%
  select(!c(priv_count, priv_pay_iqr, mcare_pay_mean, mcare_pay_median, mcare_pay_sd)) %>%
  mutate(public_private = 'private') %>%
  rename(pay_mean = priv_pay_mean) %>%
  rename(pay_median = priv_pay_median) %>%
  filter(!is.na(pay_mean) & !is.na(mcare_los))
```

Separate out test set

I arbitrarily grabbed 25% of the private records post-filtering (we are only interested in predicting using these).

```
test_set <-
  filtered_data_private %>%
  sample_frac(0.25)

filtered_data_private <- anti_join(filtered_data_private,
                                   test_set,
                                   by = ('index'))
```

Data Recombination

```
dev_set <- rbind(filtered_data_public, filtered_data_private)
```

Create Development and Test Sets for Original Data

```
untransformed_data <- cat_encoded %>%
  filter((priv_count >= 50) & !is.na(mcare_los)) %>%
  select(!c(priv_pay_iqr, mcare_pay_sd))

untransformed_test_set <-
  untransformed_data %>%
  sample_frac(0.25)

untransformed_dev_set <- anti_join(untransformed_data,
                                   untransformed_test_set,
                                   by = ('index'))
```

Comparing Performance

Original Dataset

```
orig_lm <- lm(formula = priv_pay_median ~ ., data = (untransformed_dev_set %>% select(!c(priv_pay_mean,
#train(
#  priv_pay_median ~ .,
#  data = (untransformed_dev_set %>% select(!c(priv_pay_mean, msa, index))),
#  method = 'lasso'
#)
summary(orig_lm)
```

Linear Regression

```
##
## Call:
## lm(formula = priv_pay_median ~ ., data = (untransformed_dev_set %>%
##     select(!c(priv_pay_mean, msa, index))))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34151  -2984   -246   2445  44147
##
## Coefficients: (24 not defined because of singularities)
##                                           Estimate Std. Error t value
## (Intercept)                             -1.437e+06  3.336e+05  -4.306
## year                                     6.834e+02  1.651e+02   4.138
## siteASC                                         NA         NA      NA
## siteInpatient                            1.425e+03  1.145e+03   1.244
## siteOutpatient                                  NA         NA      NA
## groupankle_fix                          -7.314e+03  6.447e+03  -1.134
## groupant_cerv_fusion                    -6.416e+02  6.393e+03  -0.100
## groupant_tls_fusion                      3.479e+04  7.165e+03   4.855
## groupbariatric                          -6.308e+03  6.398e+03  -0.986
## groupbreast.reconstruction              -7.212e+03  6.420e+03  -1.123
## groupbsp                                -8.475e+03  7.356e+03  -1.152
## groupbunionectomy                       -1.055e+04  6.456e+03  -1.634
## groupcardiac.ablation                    1.531e+03  6.396e+03   0.239
## groupcardiac.ablation_additional_discrete  1.388e+04  6.714e+03   2.067
## groupcardiac.ablation_linear_focal      -1.324e+03  8.979e+03  -0.148
## groupcardiac_ablaton_anesthesia          4.686e+03  6.673e+03   0.702
## groupcardiac_ablaton_ice                 2.582e+03  6.419e+03   0.402
## groupclavicle.fixation                  -9.996e+03  7.826e+03  -1.277
## groupcolorect                           -9.591e+03  6.660e+03  -1.440
## groupfemoral.shaft.fixation                     NA         NA      NA
## groupfess                               -9.838e+03  6.422e+03  -1.532
## grouphepat                              -7.857e+03  7.020e+03  -1.119
## grouphernia                             -9.368e+03  6.465e+03  -1.449
## grouphip_fracture_fixation                      NA         NA      NA
## grouphysterect                          -9.512e+03  6.386e+03  -1.490
## groupintracranial_thromb                        NA         NA      NA
## groupkidney.ablation                            NA         NA      NA
## grouplaac                                       NA         NA      NA
## grouplap.appendectomy                   -9.740e+03  6.436e+03  -1.513
## groupliver.ablation                             NA         NA      NA
```

```
## grouplung.ablation                          NA         NA      NA
## groupmastectomy                         -8.781e+03  6.427e+03  -1.366
## groupnavigation                          7.912e+03  7.776e+03   1.017
## grouporthovisc_monovisc                 -1.043e+04  7.394e+03  -1.410
## grouppartial.shoulder.arthroplasty              NA         NA      NA
## grouppka                                 1.809e+03  8.932e+03   0.202
## grouppnn                                        NA         NA      NA
## grouppost_cerv_fusion                           NA         NA      NA
## grouppost_tls_fusion                     2.378e+04  6.479e+03   3.670
## groupprostatectomy                      -4.778e+03  6.960e+03  -0.686
## groupprox_tibia_fixation                 8.902e+02  7.812e+03   0.114
## groupproximal.humerus                           NA         NA      NA
## groupradius.ulna.internal.fixation      -7.828e+03  6.456e+03  -1.212
## grouprevision_tha                               NA         NA      NA
## grouprevision_tka                               NA         NA      NA
## grouprobotic_assisted_surgery           -7.482e+03  6.449e+03  -1.160
## grouprtc_slap_bank                      -8.485e+03  6.411e+03  -1.324
## groupseptoplasty                        -1.041e+04  6.445e+03  -1.616
## grouptha                                -1.255e+03  6.364e+03  -0.197
## groupthoracic                            5.668e+03  8.981e+03   0.631
## grouptka                                -7.870e+02  6.342e+03  -0.124
## grouptpa                                -9.518e+03  7.845e+03  -1.213
## grouptsa                                        NA         NA      NA
## priv_count                               2.994e+00  9.228e-01   3.245
## mcare_los                                9.900e+02  4.064e+02   2.436
## mcare_pay_mean                           2.227e-01  1.669e-01   1.334
## mcare_pay_median                         7.554e-01  1.837e-01   4.113
## StateAlabama                            -2.520e+03  1.906e+03  -1.322
## StateAlaska                                     NA         NA      NA
## StateArizona                            -1.126e+04  3.453e+03  -3.260
## StateArkansas                           -4.480e+03  2.015e+03  -2.224
## StateCalifornia                         -1.164e+04  3.905e+03  -2.981
## StateColorado                           -8.083e+03  2.469e+03  -3.273
## StateDelaware                            4.736e+03  2.140e+03   2.214
## StateFlorida                             1.349e+04  2.249e+03   5.998
## StateGeorgia                             2.395e+03  1.800e+03   1.330
## StateHawaii                                     NA         NA      NA
## StateIllinois                           -6.460e+03  1.416e+03  -4.563
## StateIowa                               -6.362e+03  1.927e+03  -3.302
## StateKansas                             -7.603e+03  1.800e+03  -4.223
## StateMaryland                           -6.524e+03  2.032e+03  -3.210
## StateMassachusetts                       1.233e+03  2.341e+03   0.526
## StateMichigan                           -7.036e+03  1.423e+03  -4.944
## StateMinnesota                          -9.521e+03  1.751e+03  -5.438
## StateMississippi                         1.005e+03  3.104e+03   0.324
## StateMissouri                           -1.618e+03  3.536e+03  -0.458
## StateNebraska                           -4.056e+03  4.724e+03  -0.858
## StateNevada                             -1.493e+04  3.684e+03  -4.053
## StateNew.Jersey                          3.051e+03  1.969e+03   1.550
## StateNew.York                            6.581e+03  1.975e+03   3.333
## StateNorth.Carolina                      8.594e+03  1.785e+03   4.813
## StateNorth.Dakota                               NA         NA      NA
## StateOhio                               -1.157e+03  1.462e+03  -0.791
## StateOklahoma                           -4.128e+03  2.125e+03  -1.943
```

```
## StateOregon                                  -1.913e+04  3.958e+03  -4.833
## StatePennsylvania                            -1.123e+03  1.857e+03  -0.605
## StatePuerto.Rico                                    NA         NA      NA
## StateRhode.Island                             2.019e+03  2.825e+03   0.715
## StateSouth.Dakota                            -8.885e+03  3.940e+03  -2.255
## StateTennessee                                8.327e+02  1.804e+03   0.462
## StateTexas                                    2.238e+03  2.349e+03   0.952
## StateUtah                                    -1.478e+04  3.086e+03  -4.791
## StateVermont                                        NA         NA      NA
## StateVirginia                                 3.202e+03  1.785e+03   1.794
## StateWashington                              -1.893e+04  3.813e+03  -4.964
## StateWest.Virginia                                  NA         NA      NA
## StateWisconsin                                      NA         NA      NA
## StateWyoming                                        NA         NA      NA
## lon                                          -4.666e+02  1.060e+02  -4.403
## lat                                           8.253e+02  1.242e+02   6.643
##                                             Pr(>|t|)
## (Intercept)                                 1.73e-05 ***
## year                                        3.62e-05 ***
## siteASC                                           NA
## siteInpatient                               0.213444
## siteOutpatient                                    NA
## groupankle_fix                              0.256727
## groupant_cerv_fusion                        0.920068
## groupant_tls_fusion                         1.28e-06 ***
## groupbariatric                              0.324269
## groupbreast.reconstruction                  0.261343
## groupbsp                                    0.249377
## groupbunionectomy                           0.102299
## groupcardiac.ablation                       0.810892
## groupcardiac.ablation_additional_discrete   0.038818 *
## groupcardiac.ablation_linear_focal          0.882748
## groupcardiac_ablaton_anesthesia             0.482581
## groupcardiac_ablaton_ice                    0.687528
## groupclavicle.fixation                      0.201578
## groupcolorect                               0.149970
## groupfemoral.shaft.fixation                       NA
## groupfess                                   0.125655
## grouphepat                                  0.263181
## grouphernia                                 0.147460
## grouphip_fracture_fixation                        NA
## grouphysterect                              0.136446
## groupintracranial_thromb                          NA
## groupkidney.ablation                              NA
## grouplaac                                         NA
## grouplap.appendectomy                       0.130316
## groupliver.ablation                               NA
## grouplung.ablation                                NA
## groupmastectomy                             0.172010
## groupnavigation                             0.309073
## grouporthovisc_monovisc                     0.158588
## grouppartial.shoulder.arthroplasty                NA
## grouppka                                    0.839550
## grouppnn                                          NA
```

```
## grouppost_cerv_fusion                       NA
## grouppost_tls_fusion                  0.000247 ***
## groupprostatectomy                    0.492521
## groupprox_tibia_fixation              0.909288
## groupproximal.humerus                       NA
## groupradius.ulna.internal.fixation    0.225471
## grouprevision_tha                           NA
## grouprevision_tka                           NA
## grouprobotic_assisted_surgery         0.246085
## grouprtc_slap_bank                    0.185745
## groupseptoplasty                      0.106316
## grouptha                              0.843662
## groupthoracic                         0.528023
## grouptka                              0.901250
## grouptpa                              0.225156
## grouptsa                                    NA
## priv_count                            0.001192 **
## mcare_los                             0.014923 *
## mcare_pay_mean                        0.182252
## mcare_pay_median                      4.04e-05 ***
## StateAlabama                          0.186222
## StateAlaska                                 NA
## StateArizona                          0.001130 **
## StateArkansas                         0.026256 *
## StateCalifornia                       0.002898 **
## StateColorado                         0.001077 **
## StateDelaware                         0.026942 *
## StateFlorida                          2.30e-09 ***
## StateGeorgia                          0.183535
## StateHawaii                                 NA
## StateIllinois                         5.30e-06 ***
## StateIowa                             0.000974 ***
## StateKansas                           2.50e-05 ***
## StateMaryland                         0.001344 **
## StateMassachusetts                    0.598608
## StateMichigan                         8.16e-07 ***
## StateMinnesota                        5.93e-08 ***
## StateMississippi                      0.746198
## StateMissouri                         0.647248
## StateNebraska                         0.390717
## StateNevada                           5.21e-05 ***
## StateNew.Jersey                       0.121328
## StateNew.York                         0.000872 ***
## StateNorth.Carolina                   1.57e-06 ***
## StateNorth.Dakota                           NA
## StateOhio                             0.428858
## StateOklahoma                         0.052166 .
## StateOregon                           1.43e-06 ***
## StatePennsylvania                     0.545547
## StatePuerto.Rico                            NA
## StateRhode.Island                     0.474986
## StateSouth.Dakota                     0.024198 *
## StateTennessee                        0.644426
## StateTexas                            0.340967
```

```
## StateUtah                                    1.75e-06 ***
## StateVermont                                       NA
## StateVirginia                                0.073014 .
## StateWashington                              7.36e-07 ***
## StateWest.Virginia                                 NA
## StateWisconsin                                     NA
## StateWyoming                                       NA
## lon                                          1.11e-05 ***
## lat                                          3.77e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6292 on 2453 degrees of freedom
## Multiple R-squared:  0.8453, Adjusted R-squared:  0.8406
## F-statistic: 178.8 on 75 and 2453 DF,  p-value: < 2.2e-16
```

```
orig_lm_pred <- predict(orig_lm, newdata = untransformed_test_set)
```

```
## Warning in predict.lm(orig_lm, newdata = untransformed_test_set): prediction
## from a rank-deficient fit may be misleading
```

```
print("")
```

```
## [1] ""
```

```
print("MAPE is:")
```

```
## [1] "MAPE is:"
```

```
MAPE(orig_lm_pred, untransformed_test_set$priv_pay_median)
```

```
## [1] 0.2651394
```

```
orig_tree <- rpart(formula = priv_pay_median ~ ., data = (untransformed_dev_set %>% select(!c(priv_pay_m
summary(orig_tree)
```

Decision Tree Regression

```
## Call:
## rpart(formula = priv_pay_median ~ ., data = (untransformed_dev_set %>%
##     select(!c(priv_pay_mean, msa, index))))
##   n= 2529
##
##           CP nsplit rel error    xerror       xstd
## 1 0.48565102      0 1.0000000 1.0007136 0.06190039
## 2 0.23315220      1 0.5143490 0.5208618 0.03707932
## 3 0.03353983      2 0.2811968 0.2870148 0.01460363
## 4 0.01790218      3 0.2476569 0.2589322 0.01437185
## 5 0.01000000      4 0.2297548 0.2352728 0.01279367
##
## Variable importance
##     mcare_pay_median       mcare_pay_mean            mcare_los
##                   30                   25                   11
##        siteInpatient       siteOutpatient grouppost_tls_fusion
##                   11                   11                    8
##             grouptha
##                    3
```

```
##
## Node number 1: 2529 observations,    complexity param=0.485651
##   mean=18954.27, MSE=2.482977e+08
##   left son=2 (1780 obs) right son=3 (749 obs)
##   Primary splits:
##       mcare_pay_median     < 10421.8   to the left,  improve=0.4856510, (0 missing)
##       mcare_pay_mean       < 9643.043  to the left,  improve=0.4617235, (0 missing)
##       grouppost_tls_fusion < 0.5       to the left,  improve=0.4025051, (0 missing)
##       mcare_los            < 0.2629382 to the left,  improve=0.3252547, (0 missing)
##       siteInpatient        < 0.5       to the left,  improve=0.3252547, (0 missing)
##   Surrogate splits:
##       mcare_pay_mean < 10222.65  to the left,  agree=0.955, adj=0.848, (0 split)
##       mcare_los      < 0.8368019 to the left,  agree=0.867, adj=0.550, (0 split)
##       siteInpatient  < 0.5       to the left,  agree=0.866, adj=0.549, (0 split)
##       siteOutpatient < 0.5       to the right, agree=0.866, adj=0.549, (0 split)
##       grouptha       < 0.5       to the left,  agree=0.747, adj=0.147, (0 split)
##
## Node number 2: 1780 observations,    complexity param=0.03353983
##   mean=11831, MSE=3.533881e+07
##   left son=4 (1552 obs) right son=5 (228 obs)
##   Primary splits:
##       mcare_pay_median < 8001.902  to the left,  improve=0.3348193, (0 missing)
##       mcare_pay_mean   < 7739.329  to the left,  improve=0.3258993, (0 missing)
##       siteOutpatient   < 0.5       to the right, improve=0.1615661, (0 missing)
##       mcare_los        < 0.2629382 to the left,  improve=0.1615661, (0 missing)
##       siteInpatient    < 0.5       to the left,  improve=0.1615661, (0 missing)
##   Surrogate splits:
##       mcare_pay_mean < 7908.753  to the left,  agree=0.984, adj=0.877, (0 split)
##       mcare_los      < 0.8845599 to the left,  agree=0.937, adj=0.504, (0 split)
##       siteInpatient  < 0.5       to the left,  agree=0.936, adj=0.500, (0 split)
##       siteOutpatient < 0.5       to the right, agree=0.936, adj=0.500, (0 split)
##       grouptka       < 0.5       to the left,  agree=0.908, adj=0.281, (0 split)
##
## Node number 3: 749 observations,    complexity param=0.2331522
##   mean=35882.74, MSE=3.47236e+08
##   left son=6 (658 obs) right son=7 (91 obs)
##   Primary splits:
##       mcare_pay_median     < 24200.04  to the left,  improve=0.56293020, (0 missing)
##       mcare_pay_mean       < 25868.47  to the left,  improve=0.51263140, (0 missing)
##       grouppost_tls_fusion < 0.5       to the left,  improve=0.50274680, (0 missing)
##       mcare_los            < 3.1957    to the left,  improve=0.12668500, (0 missing)
##       groupbariatric       < 0.5       to the right, improve=0.04341436, (0 missing)
##   Surrogate splits:
##       grouppost_tls_fusion < 0.5       to the left,  agree=0.980, adj=0.835, (0 split)
##       mcare_pay_mean       < 26497.53  to the left,  agree=0.980, adj=0.835, (0 split)
##       groupant_tls_fusion  < 0.5       to the left,  agree=0.885, adj=0.055, (0 split)
##
## Node number 4: 1552 observations
##   mean=10512.58, MSE=2.132907e+07
##
## Node number 5: 228 observations
##   mean=20805.49, MSE=3.832979e+07
##
## Node number 6: 658 observations,    complexity param=0.01790218
```

```
##   mean=30683.41, MSE=1.179562e+08
##   left son=12 (394 obs) right son=13 (264 obs)
##   Primary splits:
##       mcare_pay_median < 15633.84  to the left,  improve=0.14483740, (0 missing)
##       mcare_pay_mean   < 14744.56  to the left,  improve=0.08703534, (0 missing)
##       StateNew.York    < 0.5       to the left,  improve=0.08000622, (0 missing)
##       groupbariatric   < 0.5       to the right, improve=0.04851706, (0 missing)
##       lat              < 33.19977  to the right, improve=0.04656688, (0 missing)
##   Surrogate splits:
##       mcare_pay_mean        < 15625.54  to the left,  agree=0.839, adj=0.598, (0 split)
##       siteInpatient         < 0.5       to the right, agree=0.796, adj=0.492, (0 split)
##       siteOutpatient        < 0.5       to the left,  agree=0.796, adj=0.492, (0 split)
##       mcare_los             < 0.5738636 to the right, agree=0.796, adj=0.492, (0 split)
##       groupcardiac.ablation < 0.5       to the left,  agree=0.734, adj=0.337, (0 split)
##
## Node number 7: 91 observations
##   mean=73477.89, MSE=3.962397e+08
##
## Node number 12: 394 observations
##   mean=27300, MSE=5.049393e+07
##
## Node number 13: 264 observations
##   mean=35732.89, MSE=1.760568e+08
```

```
orig_tree_pred <- predict(orig_tree, newdata = untransformed_test_set)

print("")
```

```
## [1] ""
```

```
print("MAPE is:")
```

```
## [1] "MAPE is:"
```

```
MAPE(orig_tree_pred, untransformed_test_set$priv_pay_median)
```

```
## [1] 0.4618645
```

Transformed Dataset

```
transformed_lm <- lm(formula = pay_median ~ ., data = (dev_set %>% select(!c(pay_mean, msa, index))))
#train(
#  priv_pay_median ~ .,
#  data = (untransformed_dev_set %>% select(!c(priv_pay_mean, msa, index))),
#  method = 'lasso'
#)
summary(transformed_lm)
```

Linear Regression

```
##
## Call:
## lm(formula = pay_median ~ ., data = (dev_set %>% select(!c(pay_mean,
##     msa, index))))
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
## -37698  -1492    174   1562  69553
##
## Coefficients: (4 not defined because of singularities)
##                                          Estimate Std. Error  t value
## (Intercept)                            -1.394e+06  5.285e+04  -26.374
## year                                    6.960e+02  2.617e+01   26.600
## siteASC                                        NA         NA       NA
## siteInpatient                           3.482e+03  8.269e+01   42.111
## siteOutpatient                                 NA         NA       NA
## groupankle_fix                         -3.145e+03  1.809e+02  -17.386
## groupant_cerv_fusion                    2.598e+03  1.792e+02   14.494
## groupant_tls_fusion                     1.585e+04  2.274e+02   69.688
## groupbariatric                         -4.268e+03  1.854e+02  -23.022
## groupbreast.reconstruction             -3.398e+03  1.846e+02  -18.407
## groupbsp                               -3.188e+03  2.789e+02  -11.430
## groupbunionectomy                      -3.421e+03  2.084e+02  -16.415
## groupcardiac.ablation                   8.866e+03  1.870e+02   47.413
## groupcardiac.ablation_additional_discrete  1.073e+04  2.311e+02   46.451
## groupcardiac.ablation_linear_focal      1.069e+04  2.442e+02   43.772
## groupcardiac_ablaton_anesthesia         1.022e+04  5.026e+02   20.338
## groupcardiac_ablaton_ice                1.081e+04  2.135e+02   50.636
## groupclavicle.fixation                 -2.612e+03  2.382e+02  -10.965
## groupcolorect                          -5.160e+03  1.943e+02  -26.557
## groupfemoral.shaft.fixation            -3.871e+03  2.508e+02  -15.432
## groupfess                              -4.629e+03  1.957e+02  -23.660
## grouphepat                             -7.952e+03  2.266e+02  -35.100
## grouphernia                            -5.676e+03  1.817e+02  -31.248
## grouphip_fracture_fixation             -4.254e+03  2.235e+02  -19.035
## grouphysterect                         -3.503e+03  1.727e+02  -20.285
## groupintracranial_thromb                1.247e+04  3.014e+02   41.367
## groupkidney.ablation                   -2.875e+03  3.687e+02   -7.798
## grouplaac                               6.053e+03  4.459e+02   13.575
## grouplap.appendectomy                  -4.528e+03  1.967e+02  -23.024
## groupliver.ablation                    -1.287e+03  3.076e+02   -4.185
## grouplung.ablation                     -4.156e+03  7.896e+02   -5.263
## groupmastectomy                        -4.561e+03  1.838e+02  -24.821
## groupnavigation                         2.172e+03  2.773e+02    7.832
## grouporthovisc_monovisc                -7.135e+03  3.979e+02  -17.934
## grouppartial.shoulder.arthroplasty     -3.899e+02  2.454e+02   -1.589
## grouppka                                3.408e+02  2.098e+02    1.624
## grouppnn                               -5.085e+03  3.156e+02  -16.111
## grouppost_cerv_fusion                   2.671e+03  2.332e+02   11.457
## grouppost_tls_fusion                    1.324e+04  1.865e+02   71.012
## groupprostatectomy                     -3.000e+03  1.934e+02  -15.511
## groupprox_tibia_fixation               -3.390e+03  2.039e+02  -16.626
## groupproximal.humerus                   3.462e+02  2.018e+02    1.716
## groupradius.ulna.internal.fixation     -3.359e+03  1.868e+02  -17.985
## grouprevision_tha                      -1.985e+03  2.642e+02   -7.513
## grouprevision_tka                       5.631e+03  5.420e+02   10.389
## grouprobotic_assisted_surgery          -3.619e+03  2.797e+02  -12.940
## grouprtc_slap_bank                     -3.890e+03  1.925e+02  -20.214
## groupseptoplasty                       -4.695e+03  2.034e+02  -23.085
## grouptha                               -1.406e+03  1.795e+02   -7.832
```

```
## groupthoracic                -2.387e+03  2.289e+02  -10.429
## grouptka                      5.224e+02  1.712e+02    3.051
## grouptpa                     -3.753e+03  2.313e+02  -16.226
## grouptsa                             NA         NA       NA
## mcare_los                     1.332e+03  1.689e+01   78.892
## StateAlabama                 -2.677e+03  5.217e+02   -5.132
## StateAlaska                  -2.411e+03  8.473e+02   -2.846
## StateArizona                 -1.261e+03  5.076e+02   -2.484
## StateArkansas                -2.549e+03  4.983e+02   -5.115
## StateCalifornia               7.630e+02  5.119e+02    1.490
## StateColorado                -1.631e+03  4.581e+02   -3.560
## StateDelaware                 6.325e+02  6.349e+02    0.996
## StateFlorida                 -5.675e+02  5.637e+02   -1.007
## StateGeorgia                 -1.173e+03  5.405e+02   -2.170
## StateHawaii                   1.658e+03  1.298e+03    1.278
## StateIllinois                -1.955e+03  5.010e+02   -3.902
## StateIowa                    -2.232e+03  4.910e+02   -4.545
## StateKansas                  -2.220e+03  4.800e+02   -4.625
## StateMaryland                 3.680e+03  6.087e+02    6.046
## StateMassachusetts            1.444e+03  6.714e+02    2.150
## StateMichigan                -1.560e+03  5.395e+02   -2.893
## StateMinnesota               -9.943e+02  5.017e+02   -1.982
## StateMississippi             -1.972e+03  5.518e+02   -3.573
## StateMissouri                -2.204e+03  5.180e+02   -4.254
## StateNebraska                -2.138e+03  5.556e+02   -3.848
## StateNevada                  -1.810e+03  5.305e+02   -3.412
## StateNew.Jersey               4.908e+02  6.227e+02    0.788
## StateNew.York                -4.851e+02  6.202e+02   -0.782
## StateNorth.Carolina          -5.794e+02  5.686e+02   -1.019
## StateNorth.Dakota            -2.333e+03  7.427e+02   -3.142
## StateOhio                    -2.001e+03  5.485e+02   -3.647
## StateOklahoma                -2.468e+03  4.924e+02   -5.012
## StateOregon                  -1.872e+03  5.180e+02   -3.615
## StatePennsylvania            -1.055e+03  5.967e+02   -1.768
## StatePuerto.Rico             -1.237e+03  1.292e+03   -0.957
## StateRhode.Island             5.151e+01  7.184e+02    0.072
## StateSouth.Dakota            -1.998e+03  5.372e+02   -3.719
## StateTennessee               -2.033e+03  5.383e+02   -3.776
## StateTexas                   -1.278e+03  4.818e+02   -2.653
## StateUtah                    -2.502e+03  4.852e+02   -5.157
## StateVermont                 -3.835e+03  7.821e+02   -4.903
## StateVirginia                -6.845e+02  5.818e+02   -1.176
## StateWashington              -2.219e+03  4.998e+02   -4.440
## StateWest.Virginia           -1.706e+03  6.080e+02   -2.806
## StateWisconsin               -1.496e+03  5.166e+02   -2.895
## StateWyoming                         NA         NA       NA
## lon                          -5.127e+01  1.422e+01   -3.605
## lat                           1.014e+02  1.861e+01    5.451
## public_privatepublic         -1.135e+04  9.071e+01 -125.138
##                              Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## year                          < 2e-16 ***
## siteASC                           NA
## siteInpatient                 < 2e-16 ***
```

```
## siteOutpatient                                    NA
## groupankle_fix                                < 2e-16 ***
## groupant_cerv_fusion                          < 2e-16 ***
## groupant_tls_fusion                           < 2e-16 ***
## groupbariatric                                < 2e-16 ***
## groupbreast.reconstruction                    < 2e-16 ***
## groupbsp                                      < 2e-16 ***
## groupbunionectomy                             < 2e-16 ***
## groupcardiac.ablation                         < 2e-16 ***
## groupcardiac.ablation_additional_discrete     < 2e-16 ***
## groupcardiac.ablation_linear_focal            < 2e-16 ***
## groupcardiac_ablaton_anesthesia               < 2e-16 ***
## groupcardiac_ablaton_ice                      < 2e-16 ***
## groupclavicle.fixation                        < 2e-16 ***
## groupcolorect                                 < 2e-16 ***
## groupfemoral.shaft.fixation                   < 2e-16 ***
## groupfess                                     < 2e-16 ***
## grouphepat                                    < 2e-16 ***
## grouphernia                                   < 2e-16 ***
## grouphip_fracture_fixation                    < 2e-16 ***
## grouphysterect                                < 2e-16 ***
## groupintracranial_thromb                      < 2e-16 ***
## groupkidney.ablation                         6.46e-15 ***
## grouplaac                                     < 2e-16 ***
## grouplap.appendectomy                         < 2e-16 ***
## groupliver.ablation                          2.86e-05 ***
## grouplung.ablation                           1.42e-07 ***
## groupmastectomy                               < 2e-16 ***
## groupnavigation                              4.94e-15 ***
## grouporthovisc_monovisc                       < 2e-16 ***
## grouppartial.shoulder.arthroplasty           0.112162
## grouppka                                     0.104312
## grouppnn                                      < 2e-16 ***
## grouppost_cerv_fusion                         < 2e-16 ***
## grouppost_tls_fusion                          < 2e-16 ***
## groupprostatectomy                            < 2e-16 ***
## groupprox_tibia_fixation                      < 2e-16 ***
## groupproximal.humerus                        0.086156 .
## groupradius.ulna.internal.fixation            < 2e-16 ***
## grouprevision_tha                            5.88e-14 ***
## grouprevision_tka                             < 2e-16 ***
## grouprobotic_assisted_surgery                 < 2e-16 ***
## grouprtc_slap_bank                            < 2e-16 ***
## groupseptoplasty                              < 2e-16 ***
## grouptha                                     4.94e-15 ***
## groupthoracic                                 < 2e-16 ***
## grouptka                                     0.002283 **
## grouptpa                                      < 2e-16 ***
## grouptsa                                          NA
## mcare_los                                     < 2e-16 ***
## StateAlabama                                 2.88e-07 ***
## StateAlaska                                  0.004431 **
## StateArizona                                 0.012997 *
## StateArkansas                                3.15e-07 ***
```

```
## StateCalifornia                          0.136133
## StateColorado                            0.000372 ***
## StateDelaware                            0.319130
## StateFlorida                             0.314124
## StateGeorgia                             0.029987 *
## StateHawaii                              0.201411
## StateIllinois                            9.57e-05 ***
## StateIowa                                5.50e-06 ***
## StateKansas                              3.76e-06 ***
## StateMaryland                            1.50e-09 ***
## StateMassachusetts                       0.031540 *
## StateMichigan                            0.003823 **
## StateMinnesota                           0.047504 *
## StateMississippi                         0.000353 ***
## StateMissouri                            2.10e-05 ***
## StateNebraska                            0.000119 ***
## StateNevada                              0.000646 ***
## StateNew.Jersey                          0.430624
## StateNew.York                            0.434078
## StateNorth.Carolina                      0.308216
## StateNorth.Dakota                        0.001681 **
## StateOhio                                0.000265 ***
## StateOklahoma                            5.41e-07 ***
## StateOregon                              0.000301 ***
## StatePennsylvania                        0.077008 .
## StatePuerto.Rico                         0.338569
## StateRhode.Island                        0.942840
## StateSouth.Dakota                        0.000200 ***
## StateTennessee                           0.000159 ***
## StateTexas                               0.007982 **
## StateUtah                                2.52e-07 ***
## StateVermont                             9.46e-07 ***
## StateVirginia                            0.239415
## StateWashington                          9.02e-06 ***
## StateWest.Virginia                       0.005011 **
## StateWisconsin                           0.003794 **
## StateWyoming                                   NA
## lon                                      0.000313 ***
## lat                                      5.05e-08 ***
## public_privatepublic                      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4246 on 39623 degrees of freedom
##   (2142 observations deleted due to missingness)
## Multiple R-squared:  0.768,  Adjusted R-squared:  0.7675
## F-statistic:  1411 on 93 and 39623 DF,  p-value: < 2.2e-16
```

```
transformed_lm_pred <- predict(transformed_lm, newdata = test_set)
```

```
## Warning in predict.lm(transformed_lm, newdata = test_set): prediction from a
## rank-deficient fit may be misleading
```

```
print("")
```

```
## [1] ""
```

```
print("MAPE is:")
```

```
## [1] "MAPE is:"
```

```
MAPE(transformed_lm_pred, test_set$pay_median)
```

```
## [1] 0.6292419
```

```
transformed_tree <- rpart(formula = pay_median ~ ., data = (dev_set %>% select(!c(pay_mean, msa, index)))
summary(transformed_tree)
```

Decision Tree Regression

```
## Call:
## rpart(formula = pay_median ~ ., data = (dev_set %>% select(!c(pay_mean,
##     msa, index))))
##   n= 41859
##
##            CP nsplit rel error    xerror        xstd
## 1  0.27040939      0 1.0000000 1.0000321 0.019049771
## 2  0.09919233      1 0.7295906 0.7296743 0.016323518
## 3  0.06078039      2 0.6303983 0.6305335 0.013027902
## 4  0.04770343      3 0.5696179 0.5698055 0.011755718
## 5  0.04436619      4 0.5219145 0.5305009 0.010390980
## 6  0.03918482      5 0.4775483 0.4777951 0.009351369
## 7  0.03307511      6 0.4383635 0.4386349 0.008891424
## 8  0.02816571      7 0.4052883 0.4055676 0.008430314
## 9  0.02728049      8 0.3771226 0.3803345 0.008262183
## 10 0.02367028      9 0.3498422 0.3501809 0.008006010
## 11 0.01946429     10 0.3261719 0.3265164 0.007945781
## 12 0.01000000     11 0.3067076 0.3070790 0.007763071
##
## Variable importance
##                              siteInpatient
##                                         23
##                             siteOutpatient
##                                         20
##                                  mcare_los
##                                         20
##                       grouppost_tls_fusion
##                                          9
##                             public_private
##                                          8
##                        groupant_tls_fusion
##                                          5
##                      groupcardiac.ablation
##                                          4
##                    groupcardiac_ablaton_ice
##                                          3
## groupcardiac.ablation_additional_discrete
##                                          2
```

```
##                  groupintracranial_thromb
##                                         2
##        groupcardiac.ablation_linear_focal
##                                         2
##                                grouphepat
##                                         1
##                                  grouptha
##                                         1
##
## Node number 1: 41859 observations,    complexity param=0.2704094
##   mean=10808, MSE=7.576857e+07
##   left son=2 (25021 obs) right son=3 (16838 obs)
##   Primary splits:
##       siteInpatient        < 0.5        to the left,  improve=0.27040940, (0 missing)
##       mcare_los            < 0.9459064  to the left,  improve=0.25741680, (2142 missing)
##       siteOutpatient       < 0.5        to the right, improve=0.19690010, (0 missing)
##       grouppost_tls_fusion < 0.5        to the left,  improve=0.12304970, (0 missing)
##       groupant_tls_fusion  < 0.5        to the left,  improve=0.07672191, (0 missing)
##   Surrogate splits:
##       siteOutpatient       < 0.5        to the right, agree=0.949, adj=0.873, (0 split)
##       mcare_los            < 0.05555556 to the left,  agree=0.948, adj=0.870, (0 split)
##       grouppost_tls_fusion < 0.5        to the left,  agree=0.614, adj=0.041, (0 split)
##       grouphepat           < 0.5        to the left,  agree=0.614, adj=0.041, (0 split)
##       grouptha             < 0.5        to the left,  agree=0.613, adj=0.038, (0 split)
##
## Node number 2: 25021 observations,    complexity param=0.04436619
##   mean=7094.795, MSE=3.366384e+07
##   left son=4 (24111 obs) right son=5 (910 obs)
##   Primary splits:
##       groupcardiac.ablation                     < 0.5        to the left,  improve=0.16705590, (0 mis
##       groupcardiac_ablaton_ice                  < 0.5        to the left,  improve=0.13661250, (0 mis
##       public_private                            splits as  RL, improve=0.09723133, (0 missing)
##       groupcardiac.ablation_additional_discrete < 0.5        to the left,  improve=0.09121639, (0 mis
##       groupcardiac.ablation_linear_focal        < 0.5        to the left,  improve=0.07256291, (0 mis
##
## Node number 3: 16838 observations,    complexity param=0.09919233
##   mean=16325.75, MSE=8.740137e+07
##   left son=6 (15849 obs) right son=7 (989 obs)
##   Primary splits:
##       grouppost_tls_fusion     < 0.5        to the left,  improve=0.21377030, (0 missing)
##       public_private           splits as  RL, improve=0.15509860, (0 missing)
##       groupant_tls_fusion      < 0.5        to the left,  improve=0.11611440, (0 missing)
##       groupintracranial_thromb < 0.5        to the left,  improve=0.04467537, (0 missing)
##       mcare_los                < 2.751442   to the left,  improve=0.04419843, (0 missing)
##
## Node number 4: 24111 observations,    complexity param=0.03918482
##   mean=6634.087, MSE=2.704411e+07
##   left son=8 (23328 obs) right son=9 (783 obs)
##   Primary splits:
##       groupcardiac_ablaton_ice                  < 0.5        to the left,  improve=0.19059350, (0 mis
##       groupcardiac.ablation_additional_discrete < 0.5        to the left,  improve=0.12771710, (0 mis
##       groupcardiac.ablation_linear_focal        < 0.5        to the left,  improve=0.10174840, (0 mis
##       public_private                            splits as  RL, improve=0.09581826, (0 missing)
##       groupbunionectomy                         < 0.5        to the right, improve=0.02332043, (0 mis
```

```
##
## Node number 5: 910 observations
##   mean=19301.53, MSE=5.442947e+07
##
## Node number 6: 15849 observations,    complexity param=0.06078039
##   mean=15245.98, MSE=5.873582e+07
##   left son=12 (15320 obs) right son=13 (529 obs)
##   Primary splits:
##       groupant_tls_fusion      < 0.5        to the left,  improve=0.20707920, (0 missing)
##       public_private           splits as  RL, improve=0.11166180, (0 missing)
##       groupintracranial_thromb < 0.5        to the left,  improve=0.08094680, (0 missing)
##       mcare_los                < 2.733688   to the left,  improve=0.04199526, (0 missing)
##       StateCalifornia          < 0.5        to the left,  improve=0.03498900, (0 missing)
##
## Node number 7: 989 observations,    complexity param=0.04770343
##   mean=33629.32, MSE=2.286777e+08
##   left son=14 (896 obs) right son=15 (93 obs)
##   Primary splits:
##       public_private  splits as  RL, improve=0.66897140, (0 missing)
##       mcare_los       < 3.754265   to the left,  improve=0.06697020, (0 missing)
##       StateCalifornia < 0.5        to the left,  improve=0.03953899, (0 missing)
##       lon             < -104.8334  to the right, improve=0.03935104, (0 missing)
##       StateAlabama    < 0.5        to the right, improve=0.01489424, (0 missing)
##
## Node number 8: 23328 observations,    complexity param=0.02816571
##   mean=6218.145, MSE=2.107729e+07
##   left son=16 (22744 obs) right son=17 (584 obs)
##   Primary splits:
##       groupcardiac.ablation_additional_discrete < 0.5        to the left,  improve=0.18167970, (0 mis
##       groupcardiac.ablation_linear_focal        < 0.5        to the left,  improve=0.14491470, (0 mis
##       public_private                            splits as  RL, improve=0.10955960, (0 missing)
##       grouptka                                  < 0.5        to the left,  improve=0.03396334, (0 mis
##       groupant_cerv_fusion                      < 0.5        to the left,  improve=0.02417109, (0 mis
##
## Node number 9: 783 observations
##   mean=19026.28, MSE=4.60933e+07
##
## Node number 12: 15320 observations,    complexity param=0.03307511
##   mean=14597.92, MSE=4.364144e+07
##   left son=24 (14736 obs) right son=25 (584 obs)
##   Primary splits:
##       public_private           splits as  RL, improve=0.15689940, (0 missing)
##       groupintracranial_thromb < 0.5        to the left,  improve=0.12180610, (0 missing)
##       mcare_los                < 5.274457   to the left,  improve=0.05869344, (0 missing)
##       StateCalifornia          < 0.5        to the left,  improve=0.03947224, (0 missing)
##       groupcardiac.ablation    < 0.5        to the left,  improve=0.03767129, (0 missing)
##
## Node number 13: 529 observations
##   mean=34014.13, MSE=1.314674e+08
##
## Node number 14: 896 observations
##   mean=29644.55, MSE=5.494393e+07
##
## Node number 15: 93 observations
```

```
##   mean=72020.2, MSE=2.756603e+08
##
## Node number 16: 22744 observations,    complexity param=0.02367028
##   mean=5904.576, MSE=1.711468e+07
##   left son=32 (22258 obs) right son=33 (486 obs)
##   Primary splits:
##       groupcardiac.ablation_linear_focal < 0.5        to the left,  improve=0.19286160, (0 missing)
##       public_private                      splits as  RL, improve=0.14476160, (0 missing)
##       grouptka                            < 0.5        to the left,  improve=0.05030443, (0 missing)
##       groupant_cerv_fusion                < 0.5        to the left,  improve=0.03714102, (0 missing)
##       groupcardiac_ablaton_anesthesia     < 0.5        to the left,  improve=0.03185870, (0 missing)
##
## Node number 17: 584 observations
##   mean=18430.17, MSE=2.243893e+07
##
## Node number 24: 14736 observations,    complexity param=0.02728049
##   mean=14076.99, MSE=3.477435e+07
##   left son=48 (14464 obs) right son=49 (272 obs)
##   Primary splits:
##       groupintracranial_thromb < 0.5        to the left,  improve=0.16884620, (0 missing)
##       mcare_los                < 4.28244    to the left,  improve=0.08987019, (0 missing)
##       lon                      < -114.4738  to the right, improve=0.05542764, (0 missing)
##       StateCalifornia          < 0.5        to the left,  improve=0.05541694, (0 missing)
##       groupcardiac.ablation    < 0.5        to the left,  improve=0.05157433, (0 missing)
##
## Node number 25: 584 observations
##   mean=27742.41, MSE=8.775876e+07
##
## Node number 32: 22258 observations,    complexity param=0.01946429
##   mean=5636.114, MSE=1.389775e+07
##   left son=64 (20572 obs) right son=65 (1686 obs)
##   Primary splits:
##       public_private                  splits as  RL, improve=0.19956570, (0 missing)
##       grouptka                        < 0.5        to the left,  improve=0.07187799, (0 missing)
##       groupant_cerv_fusion            < 0.5        to the left,  improve=0.05451575, (0 missing)
##       groupcardiac_ablaton_anesthesia < 0.5        to the left,  improve=0.04183239, (0 missing)
##       grouppka                        < 0.5        to the left,  improve=0.02938855, (0 missing)
##
## Node number 33: 486 observations
##   mean=18199.68, MSE=9974495
##
## Node number 48: 14464 observations
##   mean=13744.7, MSE=2.845857e+07
##
## Node number 49: 272 observations
##   mean=31746.93, MSE=5.252753e+07
##
## Node number 64: 20572 observations
##   mean=5159.348, MSE=8959678
##
## Node number 65: 1686 observations
##   mean=11453.46, MSE=3.753544e+07
```

```r
transformed_tree_pred <- predict(transformed_tree, newdata = test_set)

print("")
```

```
## [1] ""
```

```r
print("MAPE is:")
```

```
## [1] "MAPE is:"
```

```r
MAPE(transformed_tree_pred, test_set$pay_median)
```

```
## [1] 0.507283
```