Untitled

Sean Park

2022 2 14

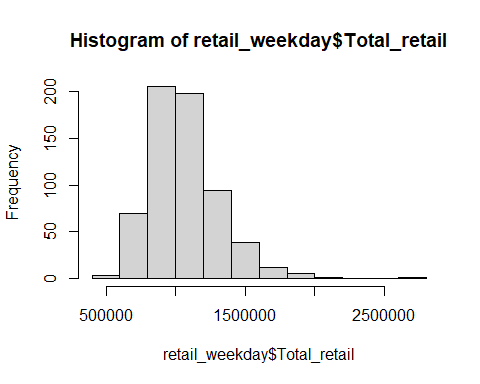
retail = read.csv("total\_data.csv", header = TRUE)  
names(retail) = c("Date","Temp","Rain","Wind","Humd","Snow","G\_temp", "A\_temp","Date1","Day","Total\_retail","Store\_retail","Deliv\_retail")  
retail = subset(retail,select = -Date1)  
summary(retail)

## Date Temp Rain Wind   
## Length:880 Min. :-7.958 Min. :0.000000 Min. :0.08333   
## Class :character 1st Qu.: 7.447 1st Qu.:0.000000 1st Qu.:0.95312   
## Mode :character Median :14.387 Median :0.000000 Median :1.28333   
## Mean :14.548 Mean :0.185928 Mean :1.43930   
## 3rd Qu.:22.240 3rd Qu.:0.004167 3rd Qu.:1.78333   
## Max. :30.629 Max. :7.058333 Max. :4.63750   
## Humd Snow G\_temp A\_temp   
## Min. : 23.08 Min. :0.000e+00 Min. :-3.792 Min. : 1.179   
## 1st Qu.: 53.73 1st Qu.:0.000e+00 1st Qu.: 6.616 1st Qu.: 7.829   
## Median : 67.98 Median :0.000e+00 Median :15.602 Median :16.163   
## Mean : 66.92 Mean :6.629e-05 Mean :15.287 Mean :16.001   
## 3rd Qu.: 80.04 3rd Qu.:0.000e+00 3rd Qu.:24.014 3rd Qu.:24.129   
## Max. :100.00 Max. :5.833e-02 Max. :34.779 Max. :29.829   
## Day Total\_retail Store\_retail Deliv\_retail   
## Length:880 Min. : 563400 Min. : 345580 Min. : 15100   
## Class :character 1st Qu.: 946500 1st Qu.: 566115 1st Qu.: 302750   
## Mode :character Median :1113750 Median : 667150 Median : 418150   
## Mean :1142001 Mean : 698353 Mean : 443648   
## 3rd Qu.:1316675 3rd Qu.: 804950 3rd Qu.: 562315   
## Max. :2771700 Max. :2533100 Max. :1068300

retail\_weekday\_mask = retail$Day != "일" & retail$Day != "토"  
retail\_weekday = retail[retail\_weekday\_mask,]  
summary(retail\_weekday)

## Date Temp Rain Wind   
## Length:628 Min. :-7.958 Min. :0.000000 Min. :0.3458   
## Class :character 1st Qu.: 7.308 1st Qu.:0.000000 1st Qu.:0.9698   
## Mode :character Median :14.481 Median :0.000000 Median :1.3042   
## Mean :14.471 Mean :0.189723 Mean :1.4380   
## 3rd Qu.:22.185 3rd Qu.:0.004167 3rd Qu.:1.7458   
## Max. :30.629 Max. :7.058333 Max. :4.6250   
## Humd Snow G\_temp A\_temp   
## Min. : 23.08 Min. :0.000e+00 Min. :-3.792 Min. : 1.179   
## 1st Qu.: 52.03 1st Qu.:0.000e+00 1st Qu.: 6.498 1st Qu.: 7.801   
## Median : 67.08 Median :0.000e+00 Median :15.404 Median :16.163   
## Mean : 66.24 Mean :9.289e-05 Mean :15.242 Mean :15.968   
## 3rd Qu.: 79.89 3rd Qu.:0.000e+00 3rd Qu.:23.999 3rd Qu.:24.136   
## Max. :100.00 Max. :5.833e-02 Max. :34.779 Max. :29.829   
## Day Total\_retail Store\_retail Deliv\_retail   
## Length:628 Min. : 563400 Min. : 345580 Min. : 15100   
## Class :character 1st Qu.: 894425 1st Qu.: 544710 1st Qu.: 282775   
## Mode :character Median :1031450 Median : 621350 Median : 382250   
## Mean :1061973 Mean : 653849 Mean : 408124   
## 3rd Qu.:1192800 3rd Qu.: 735950 3rd Qu.: 522350   
## Max. :2771700 Max. :2533100 Max. :1039500

hist(retail\_weekday$Total\_retail)



retail\_weekday.lm.full = lm(Total\_retail~Temp+Rain+Wind+Humd+Snow+G\_temp+A\_temp+Day, data=retail\_weekday)  
summary(retail\_weekday.lm.full)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Rain + Wind + Humd + Snow +   
## G\_temp + A\_temp + Day, data = retail\_weekday)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -501692 -145601 -28048 114739 1576198   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1008461.3 64041.4 15.747 < 2e-16 \*\*\*  
## Temp 35580.9 5542.8 6.419 2.73e-10 \*\*\*  
## Rain -24189.3 15302.7 -1.581 0.11445   
## Wind 41153.3 16829.8 2.445 0.01475 \*   
## Humd -1552.7 843.1 -1.842 0.06601 .   
## Snow 2138117.3 4079152.9 0.524 0.60036   
## G\_temp -10566.6 6371.0 -1.659 0.09771 .   
## A\_temp -13744.3 4910.2 -2.799 0.00528 \*\*   
## Day목 -38721.9 29529.2 -1.311 0.19024   
## Day수 -55435.4 29564.9 -1.875 0.06126 .   
## Day월 4360.2 29398.8 0.148 0.88215   
## Day화 -73986.9 29432.9 -2.514 0.01220 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 232600 on 616 degrees of freedom  
## Multiple R-squared: 0.1589, Adjusted R-squared: 0.1439   
## F-statistic: 10.58 on 11 and 616 DF, p-value: < 2.2e-16

drop1(retail\_weekday.lm.full, test="F")

## Single term deletions  
##   
## Model:  
## Total\_retail ~ Temp + Rain + Wind + Humd + Snow + G\_temp + A\_temp +   
## Day  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3.3335e+13 15532   
## Temp 1 2.2300e+12 3.5565e+13 15571 41.2081 2.734e-10 \*\*\*  
## Rain 1 1.3522e+11 3.3470e+13 15533 2.4987 0.114454   
## Wind 1 3.2357e+11 3.3658e+13 15537 5.9793 0.014754 \*   
## Humd 1 1.8354e+11 3.3518e+13 15534 3.3916 0.066011 .   
## Snow 1 1.4868e+10 3.3350e+13 15531 0.2747 0.600358   
## G\_temp 1 1.4886e+11 3.3484e+13 15533 2.7508 0.097715 .   
## A\_temp 1 4.2400e+11 3.3759e+13 15538 7.8352 0.005284 \*\*   
## Day 4 5.8567e+11 3.3920e+13 15536 2.7057 0.029571 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

retail\_weekday.lm.A = update(retail\_weekday.lm.full,.~.-Snow)  
summary(retail\_weekday.lm.A)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Rain + Wind + Humd + G\_temp +   
## A\_temp + Day, data = retail\_weekday)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -500157 -146022 -28190 116708 1574732   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1005631.3 63775.9 15.768 < 2e-16 \*\*\*  
## Temp 35287.2 5511.1 6.403 3.02e-10 \*\*\*  
## Rain -24453.5 15285.4 -1.600 0.11016   
## Wind 42386.9 16654.6 2.545 0.01117 \*   
## Humd -1520.6 840.4 -1.809 0.07088 .   
## G\_temp -10352.6 6354.2 -1.629 0.10377   
## A\_temp -13749.2 4907.3 -2.802 0.00524 \*\*   
## Day목 -37677.5 29444.6 -1.280 0.20116   
## Day수 -55338.0 29546.9 -1.873 0.06156 .   
## Day월 4381.5 29381.4 0.149 0.88150   
## Day화 -73917.7 29415.3 -2.513 0.01223 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 232500 on 617 degrees of freedom  
## Multiple R-squared: 0.1585, Adjusted R-squared: 0.1449   
## F-statistic: 11.62 on 10 and 617 DF, p-value: < 2.2e-16

drop1(retail\_weekday.lm.A, test="F")

## Single term deletions  
##   
## Model:  
## Total\_retail ~ Temp + Rain + Wind + Humd + G\_temp + A\_temp +   
## Day  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3.3350e+13 15531   
## Temp 1 2.2160e+12 3.5566e+13 15569 40.9973 3.021e-10 \*\*\*  
## Rain 1 1.3834e+11 3.3488e+13 15531 2.5593 0.110156   
## Wind 1 3.5011e+11 3.3700e+13 15535 6.4773 0.011168 \*   
## Humd 1 1.7696e+11 3.3527e+13 15532 3.2739 0.070875 .   
## G\_temp 1 1.4348e+11 3.3493e+13 15532 2.6545 0.103768   
## A\_temp 1 4.2430e+11 3.3774e+13 15537 7.8500 0.005241 \*\*   
## Day 4 5.8327e+11 3.3933e+13 15534 2.6978 0.029959 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

retail\_weekday.lm.A = update(retail\_weekday.lm.A,.~.-Rain)  
summary(retail\_weekday.lm.A)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Wind + Humd + G\_temp + A\_temp +   
## Day, data = retail\_weekday)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -491249 -141750 -25948 116778 1536001   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1049637.1 57611.4 18.219 < 2e-16 \*\*\*  
## Temp 34689.3 5505.4 6.301 5.62e-10 \*\*\*  
## Wind 33685.1 15761.2 2.137 0.03297 \*   
## Humd -2038.9 776.4 -2.626 0.00886 \*\*   
## G\_temp -9370.2 6332.4 -1.480 0.13946   
## A\_temp -14142.4 4907.3 -2.882 0.00409 \*\*   
## Day목 -38828.5 29472.9 -1.317 0.18818   
## Day수 -59266.1 29481.8 -2.010 0.04484 \*   
## Day월 2271.6 29388.8 0.077 0.93841   
## Day화 -75956.0 29424.7 -2.581 0.01007 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 232800 on 618 degrees of freedom  
## Multiple R-squared: 0.155, Adjusted R-squared: 0.1427   
## F-statistic: 12.6 on 9 and 618 DF, p-value: < 2.2e-16

drop1(retail\_weekday.lm.A, test="F")

## Single term deletions  
##   
## Model:  
## Total\_retail ~ Temp + Wind + Humd + G\_temp + A\_temp + Day  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3.3488e+13 15531   
## Temp 1 2.1514e+12 3.5639e+13 15568 39.7027 5.622e-10 \*\*\*  
## Wind 1 2.4751e+11 3.3736e+13 15534 4.5677 0.032972 \*   
## Humd 1 3.7365e+11 3.3862e+13 15536 6.8955 0.008855 \*\*   
## G\_temp 1 1.1865e+11 3.3607e+13 15532 2.1896 0.139457   
## A\_temp 1 4.5005e+11 3.3938e+13 15538 8.3054 0.004090 \*\*   
## Day 4 6.1081e+11 3.4099e+13 15535 2.8180 0.024525 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

retail\_weekday.lm.A = update(retail\_weekday.lm.A,.~.-G\_temp)  
summary(retail\_weekday.lm.A)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Wind + Humd + A\_temp + Day,   
## data = retail\_weekday)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -490758 -148220 -26108 114398 1537267   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1048758.8 57663.7 18.188 < 2e-16 \*\*\*  
## Temp 28548.0 3620.6 7.885 1.43e-14 \*\*\*  
## Wind 35189.0 15743.5 2.235 0.0258 \*   
## Humd -1696.2 741.8 -2.287 0.0226 \*   
## A\_temp -19048.4 3621.4 -5.260 1.99e-07 \*\*\*  
## Day목 -37890.0 29494.4 -1.285 0.1994   
## Day수 -59455.3 29509.9 -2.015 0.0444 \*   
## Day월 2593.0 29416.3 0.088 0.9298   
## Day화 -74974.5 29445.5 -2.546 0.0111 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 233000 on 619 degrees of freedom  
## Multiple R-squared: 0.152, Adjusted R-squared: 0.1411   
## F-statistic: 13.87 on 8 and 619 DF, p-value: < 2.2e-16

drop1(retail\_weekday.lm.A, test="F")

## Single term deletions  
##   
## Model:  
## Total\_retail ~ Temp + Wind + Humd + A\_temp + Day  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3.3607e+13 15532   
## Temp 1 3.3753e+12 3.6982e+13 15590 62.1699 1.428e-14 \*\*\*  
## Wind 1 2.7124e+11 3.3878e+13 15535 4.9959 0.02576 \*   
## Humd 1 2.8387e+11 3.3891e+13 15535 5.2286 0.02256 \*   
## A\_temp 1 1.5021e+12 3.5109e+13 15557 27.6668 1.988e-07 \*\*\*  
## Day 4 6.0390e+11 3.4211e+13 15535 2.7808 0.02609 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(retail\_weekday.lm.A)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Wind + Humd + A\_temp + Day,   
## data = retail\_weekday)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -490758 -148220 -26108 114398 1537267   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1048758.8 57663.7 18.188 < 2e-16 \*\*\*  
## Temp 28548.0 3620.6 7.885 1.43e-14 \*\*\*  
## Wind 35189.0 15743.5 2.235 0.0258 \*   
## Humd -1696.2 741.8 -2.287 0.0226 \*   
## A\_temp -19048.4 3621.4 -5.260 1.99e-07 \*\*\*  
## Day목 -37890.0 29494.4 -1.285 0.1994   
## Day수 -59455.3 29509.9 -2.015 0.0444 \*   
## Day월 2593.0 29416.3 0.088 0.9298   
## Day화 -74974.5 29445.5 -2.546 0.0111 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 233000 on 619 degrees of freedom  
## Multiple R-squared: 0.152, Adjusted R-squared: 0.1411   
## F-statistic: 13.87 on 8 and 619 DF, p-value: < 2.2e-16