Subway Research

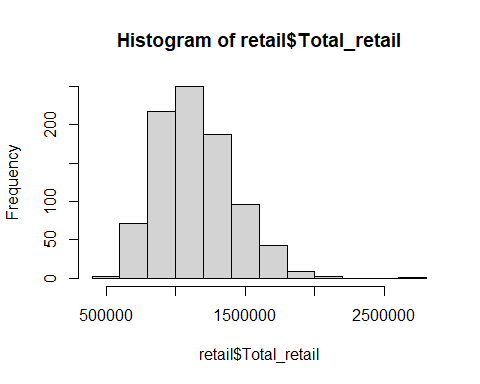
Sean Park

2022 2 13

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

hist(retail$Total\_retail)



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

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Rain + Wind + Humd + Snow +   
## G\_temp + A\_temp + Day, data = retail)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -542958 -143229 -23992 120057 1623616   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1052772.2 53346.8 19.735 < 2e-16 \*\*\*  
## Temp 30249.4 4483.7 6.747 2.77e-11 \*\*\*  
## Rain -21270.0 12453.4 -1.708 0.08800 .   
## Wind 22910.4 13215.2 1.734 0.08334 .   
## Humd -1937.0 705.4 -2.746 0.00616 \*\*   
## Snow 2383128.0 3927442.9 0.607 0.54415   
## G\_temp -7561.6 5243.9 -1.442 0.14967   
## A\_temp -11275.6 4039.6 -2.791 0.00537 \*\*   
## Day목 -39943.1 28574.0 -1.398 0.16251   
## Day수 -58757.7 28564.9 -2.057 0.03999 \*   
## Day월 4079.5 28451.0 0.143 0.88602   
## Day일 263349.4 28392.2 9.275 < 2e-16 \*\*\*  
## Day토 224748.9 28500.4 7.886 9.39e-15 \*\*\*  
## Day화 -75496.4 28474.6 -2.651 0.00816 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 225200 on 866 degrees of freedom  
## Multiple R-squared: 0.333, Adjusted R-squared: 0.323   
## F-statistic: 33.26 on 13 and 866 DF, p-value: < 2.2e-16

drop1(retail.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> 4.3919e+13 21705   
## Temp 1 2.3083e+12 4.6228e+13 21749 45.5155 2.768e-11 \*\*\*  
## Rain 1 1.4794e+11 4.4067e+13 21706 2.9172 0.088001 .   
## Wind 1 1.5242e+11 4.4072e+13 21707 3.0055 0.083339 .   
## Humd 1 3.8241e+11 4.4302e+13 21711 7.5404 0.006158 \*\*   
## Snow 1 1.8673e+10 4.3938e+13 21704 0.3682 0.544150   
## G\_temp 1 1.0545e+11 4.4025e+13 21706 2.0793 0.149670   
## A\_temp 1 3.9512e+11 4.4314e+13 21711 7.7909 0.005366 \*\*   
## Day 6 1.4450e+13 5.8369e+13 21944 47.4883 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Rain + Wind + Humd + G\_temp +   
## A\_temp + Day, data = retail)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -543824 -143185 -24531 121455 1622282   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1050617.7 53209.1 19.745 < 2e-16 \*\*\*  
## Temp 30005.6 4464.0 6.722 3.26e-11 \*\*\*  
## Rain -21502.1 12443.0 -1.728 0.08433 .   
## Wind 23854.5 13118.5 1.818 0.06935 .   
## Humd -1911.8 703.9 -2.716 0.00674 \*\*   
## G\_temp -7384.5 5233.9 -1.411 0.15863   
## A\_temp -11278.1 4038.2 -2.793 0.00534 \*\*   
## Day목 -38794.7 28500.9 -1.361 0.17381   
## Day수 -58682.7 28554.2 -2.055 0.04017 \*   
## Day월 4098.6 28440.6 0.144 0.88544   
## Day일 263323.6 28381.8 9.278 < 2e-16 \*\*\*  
## Day토 224789.6 28490.0 7.890 9.08e-15 \*\*\*  
## Day화 -75442.0 28464.1 -2.650 0.00819 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 225100 on 867 degrees of freedom  
## Multiple R-squared: 0.3328, Adjusted R-squared: 0.3235   
## F-statistic: 36.03 on 12 and 867 DF, p-value: < 2.2e-16

drop1(retail.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> 4.3938e+13 21704   
## Temp 1 2.2896e+12 4.6228e+13 21747 45.1802 3.255e-11 \*\*\*  
## Rain 1 1.5133e+11 4.4089e+13 21705 2.9862 0.084335 .   
## Wind 1 1.6757e+11 4.4105e+13 21705 3.3065 0.069351 .   
## Humd 1 3.7381e+11 4.4312e+13 21709 7.3761 0.006741 \*\*   
## G\_temp 1 1.0088e+11 4.4039e+13 21704 1.9907 0.158631   
## A\_temp 1 3.9530e+11 4.4333e+13 21710 7.8002 0.005339 \*\*   
## Day 6 1.4433e+13 5.8371e+13 21942 47.4658 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Rain + Wind + Humd + A\_temp +   
## Day, data = retail)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -541617 -145840 -27274 117615 1619616   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1052953.6 53213.6 19.787 < 2e-16 \*\*\*  
## Temp 25266.6 2942.1 8.588 < 2e-16 \*\*\*  
## Rain -19761.5 12388.7 -1.595 0.11105   
## Wind 24863.3 13106.5 1.897 0.05816 .   
## Humd -1693.2 687.1 -2.464 0.01392 \*   
## A\_temp -15209.0 2924.8 -5.200 2.49e-07 \*\*\*  
## Day목 -38142.4 28513.4 -1.338 0.18134   
## Day수 -59101.7 28569.0 -2.069 0.03887 \*   
## Day월 4206.8 28456.7 0.148 0.88251   
## Day일 262811.2 28395.7 9.255 < 2e-16 \*\*\*  
## Day토 225817.0 28496.9 7.924 7.02e-15 \*\*\*  
## Day화 -74816.6 28476.9 -2.627 0.00876 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 225200 on 868 degrees of freedom  
## Multiple R-squared: 0.3312, Adjusted R-squared: 0.3228   
## F-statistic: 39.08 on 11 and 868 DF, p-value: < 2.2e-16

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

## Single term deletions  
##   
## Model:  
## Total\_retail ~ Temp + Rain + Wind + Humd + A\_temp + Day  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 4.4039e+13 21704   
## Temp 1 3.7420e+12 4.7781e+13 21774 73.7538 < 2.2e-16 \*\*\*  
## Rain 1 1.2909e+11 4.4168e+13 21704 2.5444 0.11105   
## Wind 1 1.8258e+11 4.4221e+13 21706 3.5987 0.05816 .   
## Humd 1 3.0813e+11 4.4347e+13 21708 6.0733 0.01392 \*   
## A\_temp 1 1.3719e+12 4.5411e+13 21729 27.0397 2.487e-07 \*\*\*  
## Day 6 1.4428e+13 5.8466e+13 21941 47.3948 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Wind + Humd + A\_temp + Day,   
## data = retail)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -540688 -145322 -26421 120245 1587177   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1088653.1 48321.3 22.529 < 2e-16 \*\*\*  
## Temp 25483.3 2941.6 8.663 < 2e-16 \*\*\*  
## Wind 17753.9 12336.4 1.439 0.150467   
## Humd -2144.8 626.6 -3.423 0.000648 \*\*\*  
## A\_temp -15267.6 2927.2 -5.216 2.29e-07 \*\*\*  
## Day목 -39130.0 28532.0 -1.371 0.170591   
## Day수 -62225.5 28527.1 -2.181 0.029430 \*   
## Day월 2483.5 28461.4 0.087 0.930487   
## Day일 262658.3 28420.7 9.242 < 2e-16 \*\*\*  
## Day토 225599.6 28521.9 7.910 7.82e-15 \*\*\*  
## Day화 -76522.8 28482.0 -2.687 0.007354 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 225400 on 869 degrees of freedom  
## Multiple R-squared: 0.3293, Adjusted R-squared: 0.3216   
## F-statistic: 42.66 on 10 and 869 DF, p-value: < 2.2e-16

drop1(retail.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> 4.4168e+13 21704   
## Temp 1 3.8145e+12 4.7982e+13 21775 75.0510 < 2.2e-16 \*\*\*  
## Wind 1 1.0527e+11 4.4273e+13 21705 2.0712 0.1504673   
## Humd 1 5.9558e+11 4.4763e+13 21714 11.7181 0.0006479 \*\*\*  
## A\_temp 1 1.3827e+12 4.5551e+13 21730 27.2044 2.289e-07 \*\*\*  
## Day 6 1.4627e+13 5.8795e+13 21944 47.9641 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

retail.lm.A = update(retail.lm.A,.~.-Wind)  
summary(retail.lm.A)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Humd + A\_temp + Day, data = retail)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -525603 -147695 -26925 124056 1618317   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1130992.5 38356.8 29.486 < 2e-16 \*\*\*  
## Temp 25062.8 2928.8 8.557 < 2e-16 \*\*\*  
## Humd -2413.6 598.4 -4.033 5.99e-05 \*\*\*  
## A\_temp -14788.6 2910.0 -5.082 4.57e-07 \*\*\*  
## Day목 -39709.0 28546.7 -1.391 0.16458   
## Day수 -63747.0 28525.0 -2.235 0.02569 \*   
## Day월 2531.7 28478.9 0.089 0.92918   
## Day일 263294.2 28434.8 9.260 < 2e-16 \*\*\*  
## Day토 225383.2 28539.0 7.897 8.57e-15 \*\*\*  
## Day화 -77324.4 28494.1 -2.714 0.00679 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 225600 on 870 degrees of freedom  
## Multiple R-squared: 0.3277, Adjusted R-squared: 0.3207   
## F-statistic: 47.11 on 9 and 870 DF, p-value: < 2.2e-16

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

## Single term deletions  
##   
## Model:  
## Total\_retail ~ Temp + Humd + A\_temp + Day  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 4.4273e+13 21705   
## Temp 1 3.7265e+12 4.8000e+13 21774 73.228 < 2.2e-16 \*\*\*  
## Humd 1 8.2773e+11 4.5101e+13 21719 16.265 5.989e-05 \*\*\*  
## A\_temp 1 1.3143e+12 4.5587e+13 21728 25.827 4.572e-07 \*\*\*  
## Day 6 1.4753e+13 5.9026e+13 21946 48.318 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(retail.lm.A)

##   
## Call:  
## lm(formula = Total\_retail ~ Temp + Humd + A\_temp + Day, data = retail)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -525603 -147695 -26925 124056 1618317   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1130992.5 38356.8 29.486 < 2e-16 \*\*\*  
## Temp 25062.8 2928.8 8.557 < 2e-16 \*\*\*  
## Humd -2413.6 598.4 -4.033 5.99e-05 \*\*\*  
## A\_temp -14788.6 2910.0 -5.082 4.57e-07 \*\*\*  
## Day목 -39709.0 28546.7 -1.391 0.16458   
## Day수 -63747.0 28525.0 -2.235 0.02569 \*   
## Day월 2531.7 28478.9 0.089 0.92918   
## Day일 263294.2 28434.8 9.260 < 2e-16 \*\*\*  
## Day토 225383.2 28539.0 7.897 8.57e-15 \*\*\*  
## Day화 -77324.4 28494.1 -2.714 0.00679 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 225600 on 870 degrees of freedom  
## Multiple R-squared: 0.3277, Adjusted R-squared: 0.3207   
## F-statistic: 47.11 on 9 and 870 DF, p-value: < 2.2e-16

plot(retail.lm.A)

