library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.5 v dplyr 1.0.3  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(tidymodels)

## -- Attaching packages -------------------------------------- tidymodels 0.1.2 --

## v broom 0.7.3 v recipes 0.1.15  
## v dials 0.0.9 v rsample 0.0.8   
## v infer 0.5.4 v tune 0.1.2   
## v modeldata 0.1.0 v workflows 0.2.1   
## v parsnip 0.1.5 v yardstick 0.0.7

## -- Conflicts ----------------------------------------- tidymodels\_conflicts() --  
## x scales::discard() masks purrr::discard()  
## x dplyr::filter() masks stats::filter()  
## x recipes::fixed() masks stringr::fixed()  
## x dplyr::lag() masks stats::lag()  
## x yardstick::spec() masks readr::spec()  
## x recipes::step() masks stats::step()

bike = read\_csv("bike\_cleaned-2.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## instant = col\_double(),  
## dteday = col\_character(),  
## season = col\_character(),  
## mnth = col\_character(),  
## hr = col\_double(),  
## holiday = col\_character(),  
## weekday = col\_character(),  
## workingday = col\_character(),  
## weathersit = col\_character(),  
## temp = col\_double(),  
## atemp = col\_double(),  
## hum = col\_double(),  
## windspeed = col\_double(),  
## casual = col\_double(),  
## registered = col\_double(),  
## count = col\_double()  
## )

bike = bike %>% mutate(dteday = mdy(dteday))  
 bike1 <- sapply(bike, is.character)  
 bike[bike1] <- lapply(bike[bike1], factor)  
bike = bike %>% mutate(hr = as\_factor(hr))

set.seed(1234)  
bike\_split = initial\_split(bike, prob = 0.70, strata = count)  
train = training(bike\_split)  
test = testing(bike\_split)

Training has 13,036 rows and testing has 4,343 rows.

bike\_recipe = recipe(count ~ season + mnth + hr + holiday + weekday + temp + weathersit, train)  
  
lm\_model =   
 linear\_reg() %>%  
 set\_engine("lm")  
  
lm\_wkflow =   
 workflow() %>%  
 add\_model(lm\_model) %>%  
 add\_recipe(bike\_recipe)  
  
lm\_fit = fit(lm\_wkflow, train)

summary(lm\_fit$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -413.11 -61.65 -10.20 52.16 493.99   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -98.9730 79.7296 -1.241 0.214495   
## seasonSpring -25.8270 7.1838 -3.595 0.000325 \*\*\*  
## seasonSummer -31.0608 6.4153 -4.842 1.30e-06 \*\*\*  
## seasonWinter -62.3139 6.1169 -10.187 < 2e-16 \*\*\*  
## mnthAug -19.7272 8.0977 -2.436 0.014857 \*   
## mnthDec -4.5940 8.0633 -0.570 0.568865   
## mnthFeb 9.1888 8.0952 1.135 0.256358   
## mnthJan 8.0391 8.2890 0.970 0.332141   
## mnthJul -38.6963 8.2668 -4.681 2.88e-06 \*\*\*  
## mnthJun -16.0399 5.7001 -2.814 0.004901 \*\*   
## mnthMar 14.2386 6.2523 2.277 0.022783 \*   
## mnthMay 0.1209 4.9805 0.024 0.980638   
## mnthNov -9.9480 8.8162 -1.128 0.259180   
## mnthOct 5.1698 8.6346 0.599 0.549363   
## mnthSep 8.3652 7.6251 1.097 0.272635   
## hr1 -17.6394 6.7829 -2.601 0.009318 \*\*   
## hr2 -24.7408 6.7860 -3.646 0.000268 \*\*\*  
## hr3 -36.3172 6.7857 -5.352 8.85e-08 \*\*\*  
## hr4 -39.8317 6.8741 -5.794 7.01e-09 \*\*\*  
## hr5 -23.5341 6.8326 -3.444 0.000574 \*\*\*  
## hr6 34.9075 6.7378 5.181 2.24e-07 \*\*\*  
## hr7 170.4187 6.7576 25.219 < 2e-16 \*\*\*  
## hr8 310.2081 6.7874 45.703 < 2e-16 \*\*\*  
## hr9 167.5555 6.6896 25.047 < 2e-16 \*\*\*  
## hr10 112.2824 6.7742 16.575 < 2e-16 \*\*\*  
## hr11 139.9731 6.7959 20.597 < 2e-16 \*\*\*  
## hr12 180.4694 6.8816 26.225 < 2e-16 \*\*\*  
## hr13 182.6847 6.8514 26.664 < 2e-16 \*\*\*  
## hr14 163.6753 6.8350 23.947 < 2e-16 \*\*\*  
## hr15 168.7255 6.8956 24.469 < 2e-16 \*\*\*  
## hr16 228.8081 6.8944 33.187 < 2e-16 \*\*\*  
## hr17 380.6338 6.8048 55.936 < 2e-16 \*\*\*  
## hr18 355.7561 6.8635 51.833 < 2e-16 \*\*\*  
## hr19 244.4088 6.7834 36.031 < 2e-16 \*\*\*  
## hr20 160.9975 6.8198 23.607 < 2e-16 \*\*\*  
## hr21 110.3631 6.7372 16.381 < 2e-16 \*\*\*  
## hr22 73.3439 6.7251 10.906 < 2e-16 \*\*\*  
## hr23 34.8460 6.7667 5.150 2.65e-07 \*\*\*  
## holidayNotHoliday 27.9348 6.1853 4.516 6.35e-06 \*\*\*  
## weekdayMonday -8.6556 3.7208 -2.326 0.020018 \*   
## weekdaySaturday -1.2942 3.6295 -0.357 0.721414   
## weekdaySunday -20.1141 3.6426 -5.522 3.42e-08 \*\*\*  
## weekdayThursday -4.5677 3.6400 -1.255 0.209557   
## weekdayTuesday -7.7072 3.6594 -2.106 0.035212 \*   
## weekdayWednesday -4.4203 3.6433 -1.213 0.225043   
## temp 289.3663 11.7834 24.557 < 2e-16 \*\*\*  
## weathersitLightPrecip -50.8770 78.7639 -0.646 0.518328   
## weathersitMisty 20.5312 78.7197 0.261 0.794241   
## weathersitNoPrecip 41.2406 78.7048 0.524 0.600294   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 111.1 on 12987 degrees of freedom  
## Multiple R-squared: 0.6243, Adjusted R-squared: 0.6229   
## F-statistic: 449.6 on 48 and 12987 DF, p-value: < 2.2e-16

Most of the seasons are significant, along with most of the hours and the temp. The adjusted R-squared is 0.6229.

lm\_fit %>% predict(test) %>% bind\_cols(test) %>% metrics(truth = count, estimate = .pred)

## # A tibble: 3 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 112.   
## 2 rsq standard 0.623  
## 3 mae standard 81.6

Performance on the test set is similar to the training set and that suggests that our model is not overfitting.

The R-squared value is 0.6229 which alines with the training set.