library(ggplot2)

library(tidyverse)

library(glmnet)

# Read in the bike data

bike <- read.csv('https://raw.githubusercontent.com/IAA-Faculty/statistical\_foundations/master/bike.csv')

# Split into training and test datasets

set.seed(123)

bike <- bike %>% mutate(id = row\_number())

train <- bike %>% sample\_frac(0.7)

test <- anti\_join(bike, train, by = 'id')

# Build the predictor and target elements for modeling

train\_x <- model.matrix(cnt ~ atemp +

temp +

windspeed +

hum,

data = train)[, -1]

train\_y <- train$cnt

# Run a multiple linear regression model with temperature, actual temperature, humidity, and windspeed

bike\_ridge <- glmnet(x = train\_x, y = train\_y, alpha = 0)

plot(bike\_ridge, xvar = "lambda")

# Perform a CV to select the optimal ridge regression penalty

bike\_ridge\_cv <- cv.glmnet(x = train\_x, y = train\_y, alpha = 0)

plot(bike\_ridge\_cv)

bike\_ridge\_cv$lambda.min

# Prepare the test dataset

test\_x <- model.matrix(cnt ~ atemp +

temp +

windspeed +

hum,

data = test)[, -1]

test\_y <- test$cnt

# Get test dataset predictions at the minimum MSE lambda value

test$pred\_ridge <- predict(bike\_ridge, s = bike\_ridge\_cv$lambda.min, newx = test\_x)

test %>%

mutate(ridge\_APE = 100\*abs((cnt - pred\_ridge)/cnt)) %>%

dplyr::summarise(MAPE\_ridge = mean(ridge\_APE))

# MAPE calculation for MLR of bike dataset

bike\_lm2 <- lm(cnt ~ atemp + hum + windspeed, data = train)

test$pred\_lm <- predict(bike\_lm2, newdata = test)

test %>%

mutate(lm\_APE = 100\*abs((cnt - pred\_lm)/cnt)) %>%

dplyr::summarise(MAPE\_lm = mean(lm\_APE))