Stat 220 Lab 4

Background

With the increasing focus on sustainability and public health, bike-sharing services have gained immense popularity. A local bike-sharing company named CycleHub aims to better understand the factors that influence the demand for their bikes. Accurate estimation and prediction can help them allocate resources more efficiently, thereby optimizing operations and maximizing customer satisfaction.

Scenario

CycleHub collects data on the number of bikes rented per hour, along with several associated variables like temperature, humidity, wind speed, and whether it is a holiday or not. This data is publicly available at `https://richardson.byu.edu/220/bike_sharing_data.csv`.

CycleHub is particularly interested in the following:

- Identifying the variables that significantly influence the demand for bikes.
- Estimating the impact of these significant variables.
- Creating predictive models that can forecast future demand accurately.

Your analysis will help to start the process of fulfilling these tasks.

Your Role

As students of statistics and data science, you are tasked to help CycleHub analyze this data using simple and multiple regression as well as regression trees. Your findings will aid in decision-making processes and strategic planning.

Instructions

1. Load the provided dataset from `https://richardson.byu.edu/220/bike_sharing_data.csv`
2. Conduct exploratory data analysis to get a sense of the data.
3. Perform simple linear regression:
   (a) Choose one variable that you believe would most influence bike demand.
   (b) Estimate the parameters using Ordinary Least Squares (OLS).
   (c) Interpret the coefficients.

4. Perform multiple linear regression:
    (a) Add more variables to your model from the simple linear regression.
    (b) Estimate the new parameters.
    (c) Interpret the significance of each variable.
5. Implement a regression tree model:
    (a) Build a regression tree with the variables you found significant.
    (b) Interpret the tree.
6. Special Case Prediction:
    (a) Based on the weather forecast, the temperature tomorrow will be 72F, the humidity will be 45%, and the wind speed will be 10 mph. Also, it is not a holiday.
    (b) Use your models to predict the number of bikes that will be rented during a peak hour tomorrow.
    (c) Compare the estimated values of the simple and multiple linear regression models against the regression tree with respect to the special case prediction.

Submission

Please submit a detailed report containing all your analyses, code, visualizations, and conclusions. Additionally, provide recommendations for CycleHub based on your findings, which they can implement to improve their service.