Assignment Overview:

In this assignment you will be required to create two regression models using bike sharing data, interpret the diagnostics associated with those models, and create some predictions. You will first need to download the data, familiarize yourself with the data, and perform some data preparation tasks.

Perform the following tasks using R/RStudio, create an .rdm file with all codes, output and text responses. Finally, knit the .rmd file into an **.html file** and submit the **html file** for peer review.

Step-by-step Instructions:

1. **Accessing data** - Download the folder **Bike-Sharing-Dataset.zip** from this link - <https://archive.ics.uci.edu/ml/machine-learning-databases/00275/>. The folder will contain two datafiles **hour.csv** and ***day.csv***. You will use ***day.csv*** for this assignment. The readme file in the folder has a description of the data, which you are encouraged to read so that you can successfully interpret the analytic results. There are 16 columns in the dataset. You will need to use columns: dteday, temp, and cnt. “cnt” is the outcome variable, or dependent variable.
2. **Presenting data** – Create an .rmd file in RStudio. Use a code chunk to report a summary of the data. **5 points**
3. **Preparing data** - Extract the month names from the dteday column using lubridate package and save them in a new column month\_name, which has a chr data type. **15 points** Hints: -*Convert the column dteday into a date type using the lubridate package. -Use an appropriate lubridate function to extract the month from the dteday column and save it as month\_name. -Remember the default function will extract the month numbers and not the month name. You will find a function argument to extract month names/ labels. -Also, make sure to convert the column* ***month\_name*** *to a character data type. If you want, you may convert it back to a factor. Do not keep it as an ordered factor data type.* You will use these three columns for further analysis: month\_name, temp, and cnt.
4. **Running regression models** – You will run one simple linear regression model, Model1, and one multiple regression model, Model2, as described below. Model1: **40 points** a) Use a code chunk to run a simple linear regression model where the dependent variable is **cnt** and the independent variable is **month\_name** and save the model as **Model1**. **10 points** b) Use a code chunk to report the summary for **Model1**. Below the code chunk, use regular text to comment on the R-squared. **10 points** c) From the summary of **Model1**, identify which month is set as a reference. Use regular text (outside of a code chunk) to report the reference month’s predicted **cnt**. **10 points (2 points for identifying the reference month and 8 points for reporting the correct prediction)** d) With either a code chunk or regular text, use the coefficient estimates from **Model1** to report the predicted **cnt** for the months of January and June. **10 points (5 points for each correct prediction)** Model2: **40 points** a) Use a code chunk to run a multiple linear regression model where the dependent variable is **cnt** and the independent variables are **temp** and **month\_name**. Save the model as **Model2**. **10 points** b) Use a code chunk to report the summary for **Model2**. Below the code chunk use regular text to comment on the R-squared. Please explain why the R-squared is different from the two simple regression models. **10 points (2 points for the summary, 8 points for the explanation)** c) Compare the coefficient estimates for the **month\_nameJan** variable in **Model1** and **Model2**. With regular text explain why the coefficient estimates are different. **10 points (3 points for the comparison, 7 points for the explanation)** d) With either a code chunk or regular text, use the coefficient estimates from **Model2** to report the predicted **cnt** for the month of January when the temperature is .25. **10 points**

Knit the .rmd notebook file to an **.html** file and submit the **.html** file