**Student: Seif Kungulio**

**Date: 03/09/2025**

**Subject: Final Project**

**Class: DSCI 502**

**Section: 01W**

**Instructor: Sean Yang**

**File Name: Final\_Project\_Kungulio\_Seif.docx**

1. Load the dataset day.csv into memory.
2. Perform the following data preparations using control structures:
3. Convert numerical season (1,2,3, 4) to characters (springer, summer, fall and winter)
4. Convert numerical weathersit (1,2,3,4) to characters (Good, Mist, Bad, Severe)
5. Consider the following predictors, season, holiday, workingday, weathersit, atemp, hum, windspeed, casual and list all categorical variables from this list and convert them to factors.
6. Calculate the minimum, maximum, mean, median, standard deviation and three quartiles (25th, 50th and 75th percentiles) of cnt.
7. Calculate the minimum, maximum, mean, median, standard deviation and three quartiles (25th, 50th and 75th percentiles) of registered.
8. Calculate the correlation coefficient of the two variables: registered and cnt. Do they have a strong relationship?
9. Calculate the frequency table of season? What’s the mode of season variable?
10. Calculate the cross table of season and weathersit, then produce proportions by rows and columns respectively.
11. Please plot the histogram and density of the cnt and add the vertical line denoting the mean using ggplot2.
12. Please scatter plot of cnt (y-axis) against registered (x-axis) and add the trend line using ggplot2.
13. Please plot the barplot of season and weathersit on the same barplot using ggplot2
14. Please boxplot cnt (y-axis) against weathersit (x-axis) and save the graph in a file, cntweather.jpg, using ggplot2. Are there any differences in cnt with respect to weathersit?
15. Build the following multiple linear regression models:
16. Perform multiple linear regression with cnt as the response and the predictors are: season, weathersit, atemp, and registered. Write down the math formula with numerical coefficients for predictors atemp and registered and skip the coefficients for season and weathersit.
17. Preform multiple linear regression with cnt as the response and the predictors are: season, workingday, weathersit, atemp, and registered. Write down the math formula with numerical coefficients for predictors atemp and registered and skip the coefficients for season, workingday and weathersit.
18. Preform multiple linear regression with cnt as the response and the predictors are: season, holiday, workingday, weathersit, atemp, hum, windspeed, and registered. Write down the math formula with numerical coefficients for predictors atemp, hum, windspeed, and registered and skip the coefficients for season, holiday, workingday and weathersit.
19. Which model do you recommend to the management based on adjusted R squared? Justify your answer.
20. Summarize Question 13-C using R markdown to generate a reproducible report. Include the following scripts in your R markdown file:
21. Load the data as specified in Question 1.
22. Convert the two variables as specified in Question 2.
23. Convert the categorical variables to factors as specified in Question 3
24. Build a linear model as specified in Question 13-C. Use R markdown to report the math formula with numerical coefficients for predictors atemp, hum, windspeed, and registered. Skip the coefficients for season, holiday, workingday and weathersit.
25. Build the following logistic models:
26. forecast holiday using cnt, season, and registered.
27. forecast the holiday using cnt, season, weathersit, and registered
28. forecast the holiday using cnt, season, weathersit, workingday, and registered
29. Which model do you recommend to the management based on McFadden/pseudo-R squared to? Justify your answer