Caitlin Alano

Professor Mentch

STAT 1361

19 April 2022

Final Report

Introduction

The goal of this project was to find out the factors that would best help us understand when people are more likely to rent bikes than not. We are provided with collected data from every hour of every day for a couple months, with factors including date, seasons, holiday, functioning, ID, hour of the day, temperature in Celsius, humidity percentage, wind speed, visibility, dew point temperature in Celsius, solar radiation, rainfall, and snowfall. The first file titled 'train.csv' contains data with these variables in addition to the count of bikes that were rented at each hour of each day in this dataset which adds up to 6552 total observations. Some of the data in this file will be used to come up with a couple formulas that would allow us to get a sense of how many bikes will be rented at a specific hour of a day by generating a guess from this formula and comparing it to the actual data in the rented bike counts column in the file. The second file titled 'test.csv' contains only the factors that were explored in the first file for a total of 2208 observations, and we will be guessing values for rented bike count for each hour of everyday using the formula we came up with using data from the first file.

There are three main takeaways from this project. First, it appears that the best "formula" was one that took the average of many guesses for rental bike count using the data in the first file, with this average being a lot closer to the true amount in the rental bike count column than the guesses made from other formulas. Secondly, the amount of bike rentals is expected to

increase the most for every increase in unit for hour of the day and temperature in Celsius, so it is safe to conclude that these two are very important when making guesses for rental bike counts. Functioning also appears to be an important predictor for rental bike counts due to it having the largest effect on how far our guess for rental bike count is from the true count in the rental bike count column in the first file. Lastly, after seeing how high values for temperature are associated with high values for dew point temperature, as well as its lack of inclusion from two of the formulas we generated, it is fair to conclude that dew point temperature does not have a significant influence when predicting rental bike counts. For future analysis, it is important to investigate whether the presence of a holiday and functioning hours help us generate more accurate guesses for how many rental bikes will be rented.