Non-Technical Report

The goal of this experiment was to try and better understand bike demand in the city of Ourra, in order to produce a model to predict the demand for rental bikes at a given time. In my initial exploration of the data there seemed to be some specific times and days where the demand for bikes far exceeded the norm and it is important to note that I did not consider these in my analysis for the sake of outfitting a more accurate model. The data provided consisted of mainly environmental data which is intuitively important when trying to consider outdoor activities such as bike rental demand however some data was more useful than others. The date and time were extremely important in building the models, which is to be expected considering the warmer months of the year will tend to have more favorable weather and hours during the afternoon on average saw more bike rentals than during the night, with the exception of the 8:00am hour which saw a high number of rentals most likely due to people using the rentals before work or in order to get to work. Other variables involving the weather such as rainfall and temperature were also very informative since high instances of rainfall correlated to low bike rental counts and high temperature hours often had high corresponding bike rental counts. One piece of information that I found to be super important in predicting the bike rental on an hourly basis was whether or not the shop was functioning or not since if it was not functioning that meant the rental place had zero rentals during that time and without that key piece of information the model suffered in performance.

Decision makers should keep in mind a few important points when handling the rental bike supply in the city based off my analysis and that is the months of January, February, and December saw a significant decrease in bike rentals due to environmental factors such as snowfall and temperature. I would also say that days where rain is expected to expect very little rentals based off the data, there were only a handful of cases where heavy rainfall days saw even a small amount of rentals. Another interesting observation regarding the outliers I observed is that mostly all of them occurred between 5-8pm and had ideal temperatures and humidities with an average temperature of $25°C$ and average humidity of 49%, so this is important to remember as these are the days where the city saw extreme amounts of bike rentals. A big takeaway from my research is that in most cases the simple linear model performed just as good as the other models I tested and produced an error rate I would most likely deem as safe considering the severity of an error in this specific case may mean that a few extra bikes don't get rented out for that day or that specific time. In regards to that scenario, I would suggest perhaps overestimating in the context of this problem would be more ideal than possibly underestimating since having extra bikes isn't as harmful to the public as not having enough bikes to rent out.

Overall, this dataset seemed to be sufficient in providing insight into the problem I was tasked with and I was able to conclude multiple findings and I was able to produce accurate models that varied in complexity. In comparison to the baseline model I calculated each of the models I tested were able to significantly approve on it and showed how variable selection was done and provided me with the insight to determine which variables were important enough to include and which ones seemed to be redundant or use not useful. A final takeaway I have from this project is that the more complex models can have either very good predictive accuracy or be worse than some other

simple approaches and that it just depends on the problem, therefore in this case I think using a simple model that is more interpretable could be very useful in order for others to understand the trends and relationships. If, for some reason, the city was very concerned with not overspending on bikes or having not enough supply of bikes the more complex model I was able to build could prove to be very useful and could accurately predict the needed bike supply for a day given all of the environmental and external data.