ASSIGNMENT 2

Question 2) a) This model is very well correlated with the DV.

The R square value and the Adjusted R square value is absolute one. This means that the prediction is not leaving any residuals and is perfectly predicting the correlation among the variables with the DV. The Adjusted R square value is absolute one which means there is no divergence from the original model. The correlation estimate is under strong decimals due to the presence of negative powers of 10. This indicates that some of the variable has negligible correlation or is affected with the dependent variable COUNT.

Question 2) b) The coefficient of Humidity is +ve and has high -ve power of ten. The coefficient of Humidity is 1.4 * 10 ^ (-13). This is a very small and positive value so it will surely increase the count of bike rented as it positively correlates with the COUNT. Therefore, bike share count will increase by 1.4 * 10 ^ (-13) factor if we observe a unit increase in humidity. As I mentioned above that some of the positively correlating variable have negligible value due to high negative power, so they have a very little impact or weak correlation on the Dependent Variable.

Question 3) b) We can see from the COUNT curve that when the "feels like" temperature is very low the COUNT of bike rentals decreases and reaches minimum value of COUNT at approximately 10 degrees. If we see the "NO – BADWEATHER" line, then it perfectly correlates with the COUNT curve. When the temperature is pleasant the COUNT positively correlates with the optimum temperatures. The bike rental count decreases, after the optimum
25 temperature point. The curve gradually obtains a downward slope when the temperature increases and become high enough to be unpleasant.

Question 3) c) i) Between the temperature 10-20 we can observe the maximum difference between both the ggplots. The customers who are registered have been using a greater number of bikes as compared to the ones who are not registered with the application. Both the plots touches the minima which is 0 when the

badweather has a temperature of almost 10. Both touch the main regression line when the temperature is pleasant irrespective of the bad weather. The registered users have been using a much greater number of bikes at 0 temperature with badWeather.

ii) When conditions are bad, the registered users have been using a large number of bikes. At approx. 10 both the type of users reaches the minima zero. Registered users have a maxima of 3000 during badweather whereas the casual users have only 1000 has maximum value.

iii) Indeed, during badweather or at any given temperature the registered users have been using greater amount of bikes and it logically make sense as they are obliged to give a monthly or annually rent for the bikes and they obviously have daily needs that is why in the first place they have registered with the application for bike renting on regular basis. So, at any given temperature or badweather one has to go to work.

iv) Yes, we had used the geom smooth when we were comparing the CPI and weekly_Sales. Then we compared between various stores. As each store had a unique regession line , at each temperature there is a unique point which depicts the current COUNT. When CPI was low the sale was high, when the temperature is good enough the count is the maximum.

Question 4) a) Adjusted R-squared: 0.521. It can be explained as approximately 52.1% correlation of the dependent variable COUNT with the given variables above. The value of adjusted R – square is variable and is totally dependent on the independent variable like MONTH or WEEKDAY. So, it can increase or decrease as per the correlation between the independent and dependent variables.

b) The co-efficient for BADWEATHER is -1954.835 viz. large in magnitude with a negative sign. Hence, it is negatively associated with the predicted COUNT.

c) After calculating all the values of the given variables, keeping all other variables constant, we get the value as follows:

COUNT = 3967.981 + (-858.3)MONTH + 69.745 WEEKDAY +(–1954.835)* BADWEATHER + 184.596*TEMP – 48.640 * ATEMP + (– 25.341)*HUMIDITY

= 2520.497

d) The adjusted R square and the R square value has not been reaching to a improved point. 0.521 is not a good R square value despite the presence of various independent variable which should have actually improved the R square value but instead it is not at all improving the model.

Question5) TEMP and ATEMP are co-linear. This means that one of them can be removed. The VIF values dictate which variable to remove. ATEMP has a lower VIF value and so we keep it and discard TEMP, which has a higher VIF value.

Question6) a) Q4 had the coefficient of the value -> (-1954.835)

Now in the above model the badweather has the coefficient even worse value

(--2780.95) as there has been only one IV in this case whereas in the previous Q4 case there were larger number of Independent variables present. As there is only one Independent variable available and as we already know that BADWEATHER is the most negatively correlated variable with the predicted COUNT. Therefore, we obtained such a high negative coefficient of badweather

b) If we add the weekday as an independent variable in the model and then if we observe the coefficient of weekday (YES) we will see that it is positively correlating with the predicted count, so the thought process of the consultant is going in the right direction. So if the weekday (Yes) has 180 positive coefficient and

c) Our model is pretty much the same as the old one, the r-squared value has increased by 0.01. But we can say it has improved by a little bit.

d) i) COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS

- 201.2BADWEATHERYES:WEEKDAYYES

Case1 : BADWEATHER AND WEEKDAY

COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS - 201.2 BADWEATHERYES:WEEKDAYYES

= 4452.5 -2637.1 + 185.3 – 201.2

= 1799.5 ~ 1800

Case 2: BADWEATHER AND WEEKEND

COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS - 201.2 BADWEATHERYES:WEEKDAYYES

= 4452.5 – 2637.1

= 1815.4 ~ 1815

ii) Case1 : BADWEATHER AND WEEKDAY

COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS - 201.2 BADWEATHERYES:WEEKDAYYES

= 4452.5 -2637.1 + 185.3 – 201.2

= 1799.5 ~ 1800

Case 2: BADWEATHER AND WEEKEND

COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS - 201.2 BADWEATHERYES:WEEKDAYYES

= 4452.5 – 2637.1

= 1815.4 ~ 1815

Case 3: No BADWEATHER AND WEEKDAY

COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS - 201.2 BADWEATHERYES:WEEKDAYYES

= 4452.5 + 185.3

= 4637.8 ~ 4638

Case 4: No BADWEATHER AND WEEKEND

COUNT= 4452.5 -2637.1BADWEATHERYES + 185.3 WEEKDAYS - 201.2 BADWEATHERYES:WEEKDAYYES

= 4452.5

Question7) After adding WINDSPEED, our model is better for predictive analytics. The RMSE and MAE value for this model(fitNew) is lower than the fitOrg. In fitNew we can also see that the R-squared value has improved a bit than that of fitOrg.

Question 8) No, our model is not any better as both, RMSE and MAE have increased. Even though the adjusted R-squared value is better the error is higher which does not make it a very good model for predictive analysis.

Question 9) The Capital Bikeshare company could employ a policy that would ensure bikes station would have bikes available based on the temperature and weather. Adverse temperature and weather could call for more bikes stationed at the centers which have the most rush hour traffic ensued. Also change the pricing when there's adverse weather with charging a lower price at such occasions to ensure ease of customers, ensuring more effective customer retention and also acquiring new customers in the process.

Question 10) As the trucks used for transferring the bikes are limited to only big roads the other option which could be employed for the transferring the bikes could be bike angels who are regular users who can transfer bikes from one location to another. It also provides users with incentives for their next rides. There is an app in place that tracks the places and locations which require users to transfer bikes.