MAST6251 HW1

Bike share Regression

1 Introduction

For this assignment, we constructed a Regression model to predict the number of daily bike users based on the given dataset. A key advantage is that by predicting the number of daily bike users, we can create an efficient system to manage the bike inventory to ensure that demand can be met during different times and conditions throughout the year.

2 Regression Model Interpretation

- Dependent variable - Total(cnt)
- Casual & Registered are two different categories of bike users. Since our goal is to predict the number of daily customers, we decide "total" is most valuable as a dependent variable.

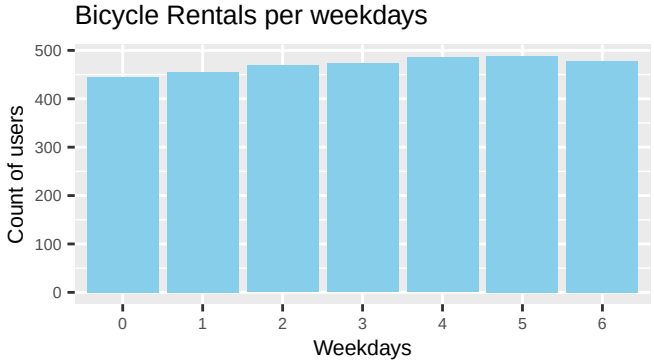
```
##
## Call:
## lm(formula = cnt ~ ., data = bikeshare_new)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3692.3  -382.1    76.9   503.8  2850.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1386.50     225.93   6.137 1.39e-09 ***
## yr1           1983.75      59.53  33.324  < 2e-16 ***
## workingday1    413.97      79.78   5.189 2.75e-07 ***
## temp          6120.24     181.68  33.687  < 2e-16 ***
## hum          -1699.87     287.53  -5.912 5.23e-09 ***
## windspeed    -3046.72     412.82  -7.380 4.38e-13 ***
## season_2       795.73      74.15  10.732  < 2e-16 ***
## season_4      1089.83      88.87  12.264  < 2e-16 ***
## mnth_9         957.70     114.92   8.334 3.97e-16 ***
## mnth_10        446.35     128.27   3.480 0.000532 ***
## weathersit_2  -427.35      78.64  -5.434 7.54e-08 ***
## weathersit_3 -1908.16     200.49  -9.518  < 2e-16 ***
## weekday_6      517.75     105.38   4.913 1.11e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 794.9 on 718 degrees of freedom
## Multiple R-squared:  0.8344, Adjusted R-squared:  0.8316
## F-statistic: 301.5 on 12 and 718 DF,  p-value: < 2.2e-16
```

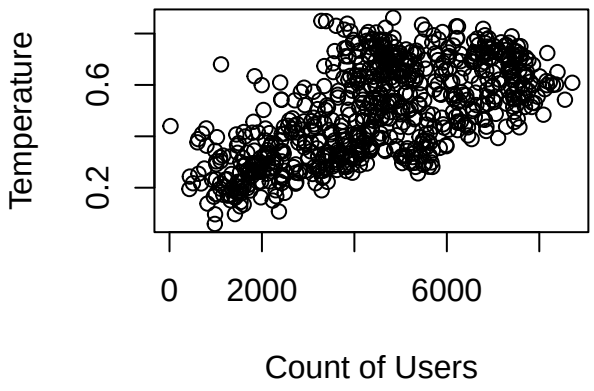
- F-statistic: 301.5, p-value: $<$ 2.2e-16 (The overall model is significant)
- R-squared:0.8344 (83.44% of the total count can be explained by predictor variables)
- There is no multicollinearity existing between the predictor variables, as all the values are within permissible range of below 5

Below are the insights and reccomendations from Model

- With our model one of the important variables was Saturday, otherwise known as weekday 6. Saturday is shown to increase the baseline by **517.75** units and is the only statistically significant day of the week based on our model. This can also be seen in the below chart that shows **Saturday** having one of the highest counts of bike rentals for any day of the week. *With this finding we believe that Bike Sharing Systems can take advantage of the higher sales on Saturday.*



- There is evidence that there is a linear relationship between temperature and Number of bikes rentals. Temperature can be used as a strong predictor to predict the number of bike hires in the region. As we can analyse from the below scatter plot, the number of bikes hired acts as a function of temperature and the number increases substantially with the variation in the **temperature** i.e. for a single unit variation in temperature can increase the number of bike rentals by **6120.24** units.



- *One way for Bike Sharing Systems to take advantage of this is by offering a discount on the price of a bike rental when a user signs up for their membership program.* This will allow Bike Sharing Systems to convert "casual" users into registered users by leveraging the most popular day of the week. Bike Sharing Systems can then take the Saturday promotion a step further by offering a smaller discount for registered users to continue to rent bikes on an ongoing basis every Saturday and drive continued use.
- Looking into our model, one of the key variables is the time of year that users rent bikes through Bike Sharing Systems. One way for us to look at the impact that **seasons** have on bike rentals. Based on the model the most value-added seasons are **Spring and Fall**, which typically includes milder temperatures that are better for bike riding. By looking further into our model, we can narrow down the specific months during each season that are the most valuable.
- By doing this we can see that the most valuable season of the year is **Fall (season 4)** and that the best months during this period are **September (month 9)** and **October (month 10)**. This makes sense as the temperatures would be cooling down from the summer highs, but not quite the cold temperatures that come with the winter months. *To take advantage of this Bike Sharing Systems can run a marketing campaign to encourage people to get outside and enjoy the last few months of nice weather during the year.*