

HW_Assignment1_Group3

2023-01-19

MAST 6251 Homework 1 - Bike Sharing

Harsh Tandel : 49045795 Jakeline Sanchez : 47395469 Yashagra Sharma : 49000606 Parag Garg : 49057313

Overview

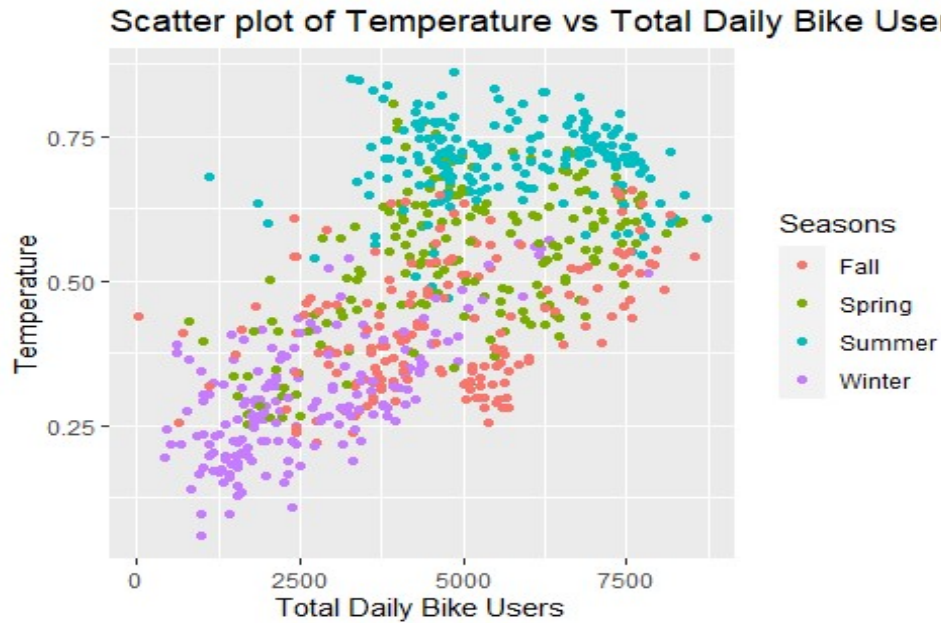
In this project, we have performed some exploratory analysis on the Bike Sharing data set, which contains historical data of bike sharing system from the beginning of 2011 to the end of 2012. Our objective is to complete a regression analysis on the data and the goal is to predict the number of daily bike users. After analyzing the relationship between each individual variable and the number of bikes rented as well as analyzing the interactions between these variables, **we recommend running a promotion in fall days with high temperature.**

Data Summary

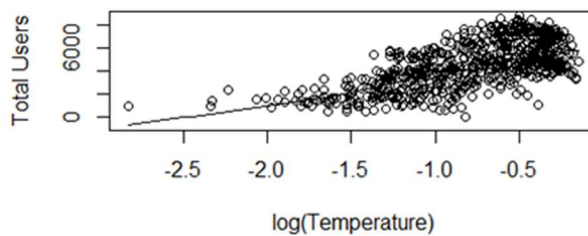
The data shows the number of registered, casual, and total users per day depending on the weather, temperature, day of the week, humidity, wind speed, month, season, if it is a weekend or holiday, and the year (2011 or 2012). Assuming that both registered and casual users pay per trip, we decided to take the total count in all our model, because to maximize revenue one must maximize the total count.

Trends

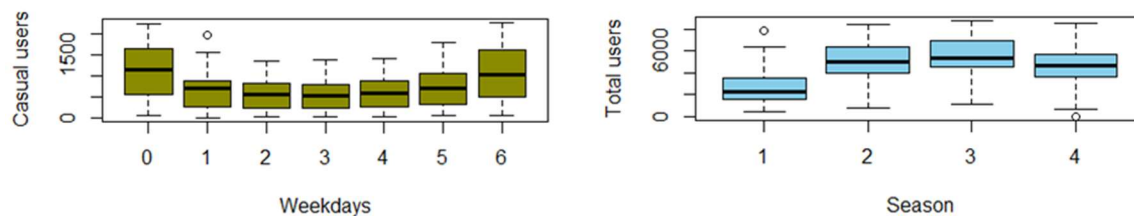
The scatter plot below shows trend for historical data collected for 2011 and 2012. This scatter plot demonstrates that we have a positive upward rise in counts of daily bike users as the temperature gets warmer in spring and summer. The daily bike user count appears to be low in the fall and wintertime.



Temperature has the strongest relationship when it comes to user count.

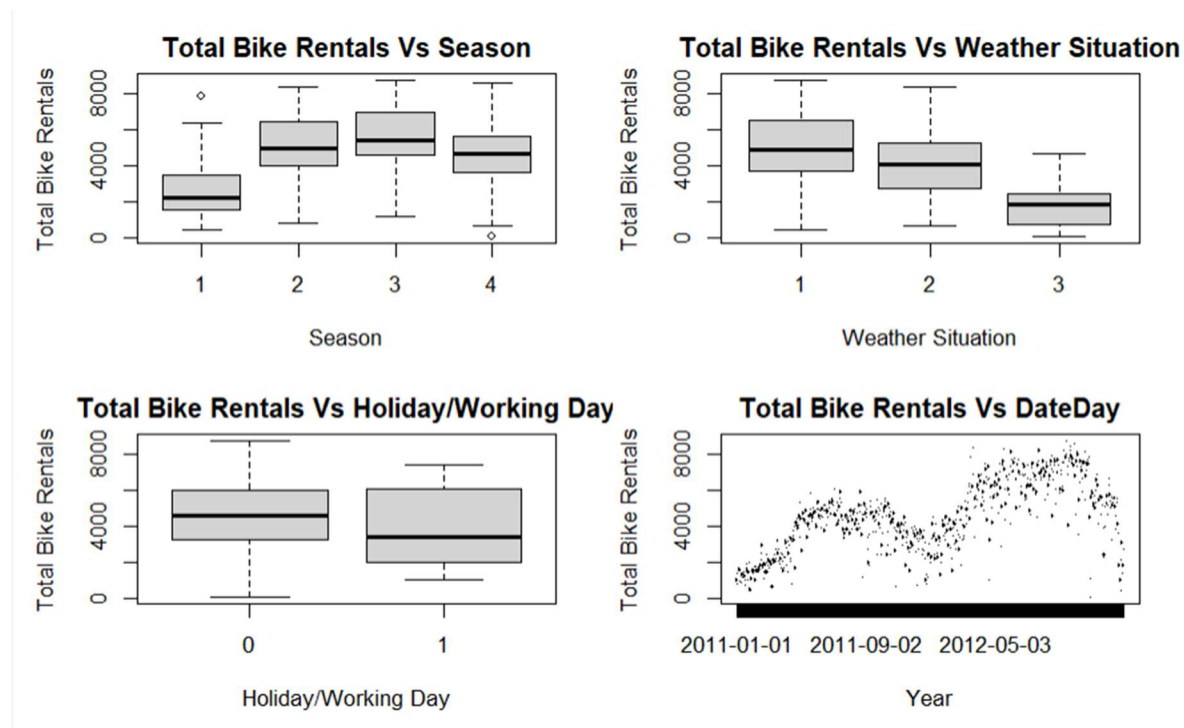


In the following figures, one can observe that the number of bike users is highest in the summer, followed by spring and fall. As expected, demand is higher when the weather is clear as compared to when it is raining/expected to rain. The count of registered bike users is higher during the weekdays (1:5), whereas the count of casual bike users is higher during the weekends (0,6). We could infer that casual users mostly use bikes for local sightseeing and recreation activities, and thus we could consider using targeted marketing strategies for these different segments of users.



Interaction

When we compare bike rentals for differ



Final Model

We created different 3 iterations of the predictive model one predicting total users, the second predicting casual users, and the third predicting registered users. For each of these models we removed statistically insignificant variables one at a time until adjusted r-squared was maximized. In other words, we picked the model that had the best balance between minimizing errors and minimizing unnecessary complexity in the model. The result was the first model iteration with the variables seen below.

In the final model, there is a statistically significant interaction between weather, temperature, humidity, and season, meaning that these 4 variables together have a greater effect in predicting the number of users than they do individually. This is expected because season influences the other three variables. In fall, the total user count goes up at a greater rate than in winter when temperature goes up.

Our final takeaway is that to get the greatest return per dollar, we should run a fall promotional campaign during warm days because it will have the greatest impact on the number of users.