

Rental Bike Users Prediction Report

Authors: Yin Wang, Karen Tabet, Cooper Clark, Kieran Furse, Sai Nathani

2023-01-18

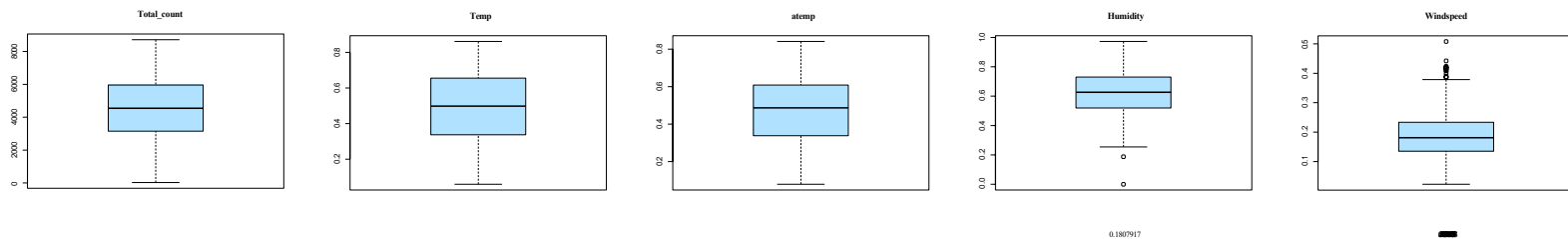
1. Data Exploration

1.1 Summary Statistics: We analyzed and summarized the data to gain a deeper understanding of it and identify the variables that have an impact on rental behaviors.

- We conducted a correlation analysis, excluding the casual and registered variables.
- The moderately correlated variables (0.3 - 0.7) are the following: temp & season, atemp & season, cnt & temp, cnt & atemp, hum & weathersit, temp & casual, atemp & casual.
- The strongly correlated variables (≥ 0.7) are mnth & season.
- The variables correlated with cnt are season, yr, temp, atemp

2. Clean the Data

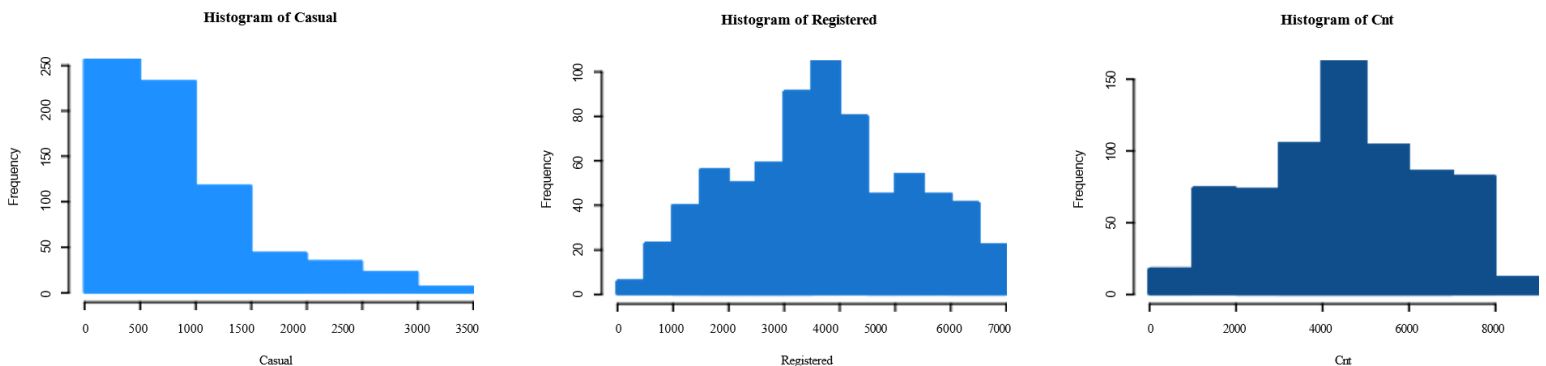
- There are no missing values in the given data set.
- There seems to be outliers in the two following variables: hum (Humidity), windspeed (Wind speed). These outliers are visually represented and can be seen below.



- Next, we removed the outliers in variables windspeed and humidity.

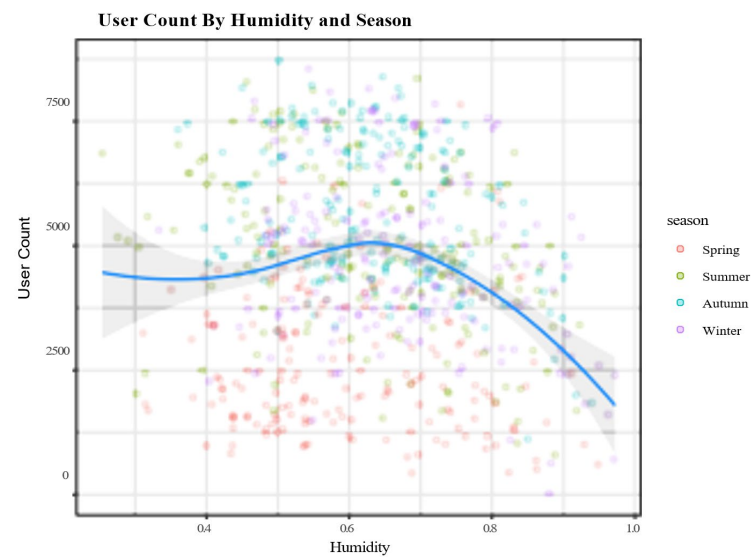
3. Graph Showing the Distribution of Dependent Variable (y)

- Results show that Casual bike users follows a lognormal distribution, while both registered and total users follow a normal distribution.
- It is also shown that registered users account for the majority (81.2%) of total users. Therefore, we selected total user (cnt) as the dependent variable to build our predictive models, as we would like to predict the number of people that rent bikes regardless of whether they are registered or not. It is still important to keep registered users as the primary target segment when advertising and developing marketing campaigns. We recommend to develop incentive programs for registered users to stay engaged and encourage friends and family to try out a bike and get a discounted day pass. We also suggest improving referral reward programs to increase loyalty and retention among users. Another method to increase brand awareness in neighborhoods would be to partner with local and small businesses.



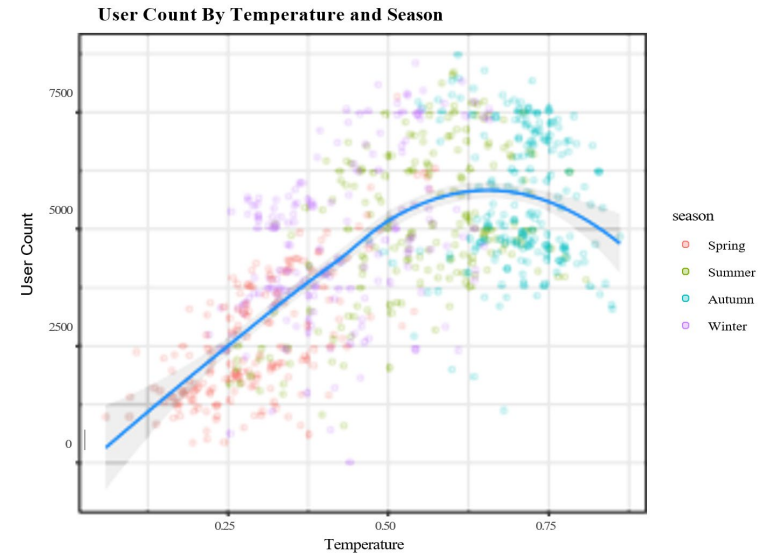
3.2 Scatterplot of Users by Humidity and Season

- We realized that Season does not necessarily tell the humidity level. We also noticed that people tend to use bikes more when humidity is moderate or low. As a result, bike rentals decrease as the humidity increases.



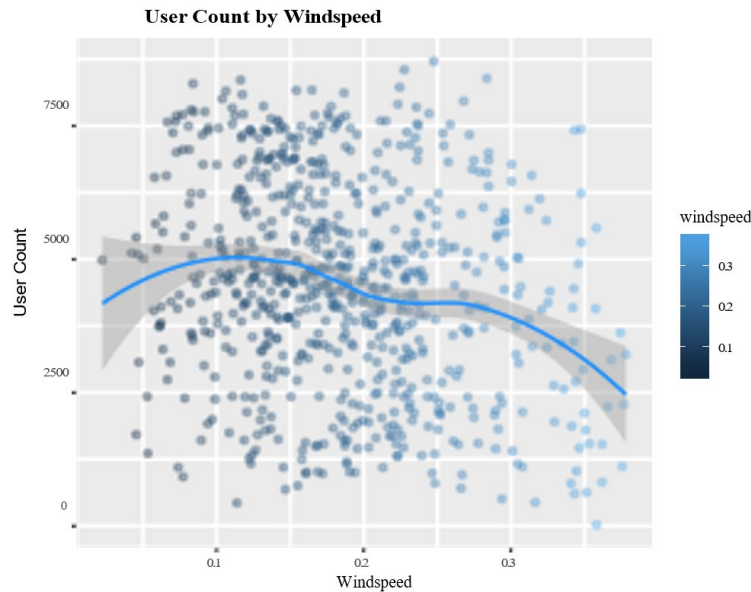
3.3 Scatterplot of Users by Temperature and Season

- We see that in the Summer and Fall seasons, when temperature is modestly warm, more people tend to rent bikes. However, bike rentals is not very common in Winter, when it is cold and the temperatures are low. This is a useful insight allowing the marketing department to cut down on budgets for Winter seasons, and increase budgets during warm weather seasons to attract the larger amount of people that are predicted to rent bikes. That being said, an alternative strategy is to offer discounts during slow periods to boost user activity.



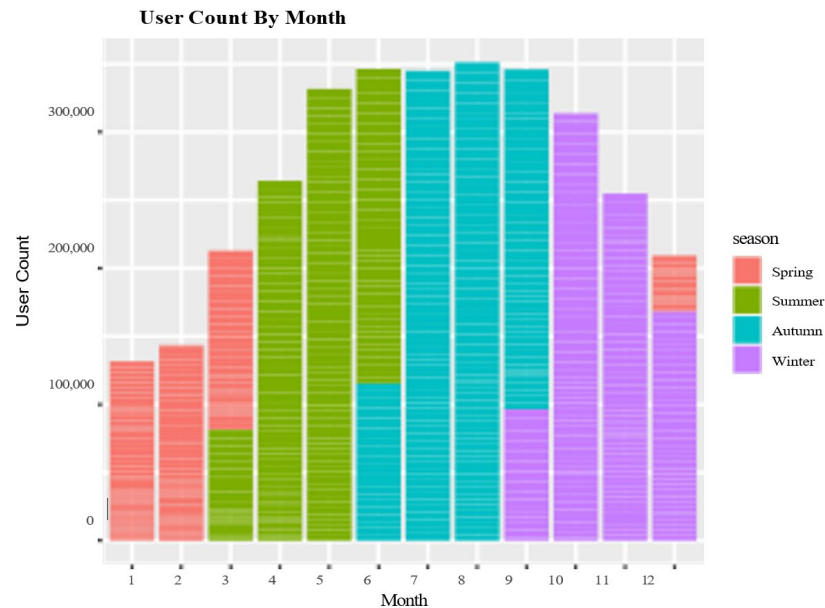
3.4 Plot of Users by Windspeed

- In addition, when it is windy, less people are willing to rent bikes as their chosen means of transportation. This is also an actionable insight when deciding the number of bikes to be made available on days with high windspeed.



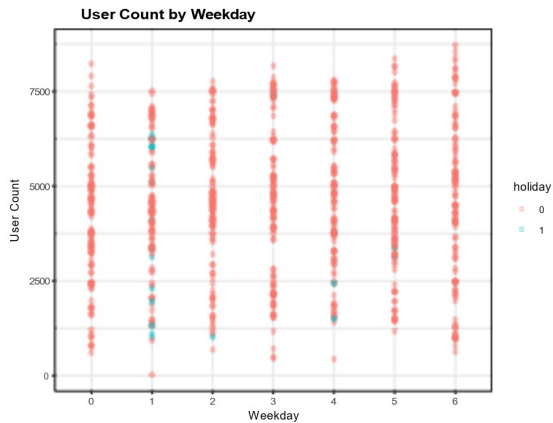
3.5 Bar Plot of Users by Month and Season

- Looking at the bar plots, we see that bike rentals increase between the months of May and September.
- Summer is actually the season with the highest bike rentals, followed by fall and winter, with winter being the least preferred season to rent bikes.
- This is also an actionable insights in terms of marketing campaigns and offers that could be launched for the Summer season and Back to School season in the Fall. We suggest hosting a series of events such as group rides, bike tours, and races during Summer to engage with the community. An example of a marketing campaign to start off the Summer could be Cycle Into The Summer, where the brand would organize cycling events on weekends, sponsored by fitness brands.

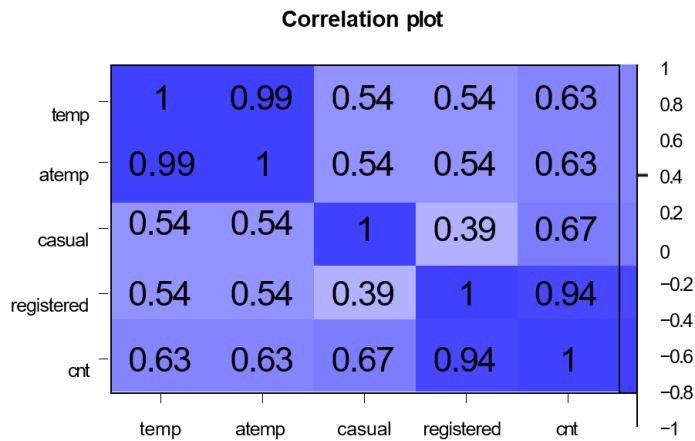


3.6 Scatter Plot of Users by Weekday and Holiday

- Looking more closely at bike rentals per weekday and holiday, we see that Holidays are likely to fall on a Monday. Generally, holidays would not lead to an increase in bike rentals and users. (Holiday: 0= non-holiday, 1= holiday).



4. Regression models



- The Adjusted R-squared of Model_1, Model_2, and Model_3 are 0.5738, 0.5743, 0.8204, respectively. The larger the Adjusted R-squared the more accurate the model; therefore, Model_3 is the best performing model in predicting total rental bikes.

5. Model Predictions

- The Residual plot of Model_3 is a horizontal line, which is the best among the three models as its variance is constant.

