Honors Project Part 3

Amit Grewal

4/9/2021

# Introduction

This dataset consists of an observation for every day from January 1st, 2011 to December 31st, 2012. The observation includes such independent variables as the temperature, humidity, and wind speed plus some categorical variables like day of the week, season, year, month, working days, holidays, and weather. The dependent variables are casual, registered, and total which refer to the number of casual/random, registered, and total (casual plus registered) bike renters on a specific day.

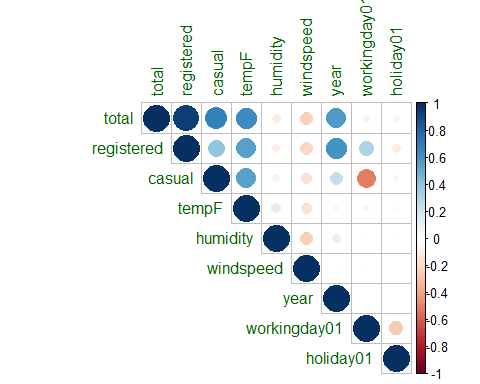
Obviously, the casual and registered variables have a strong positive relationship with the total variable as the total variable is the sum of casual and registered. I think that the quantitative variable “tempF” will be strongly positively related to “total” and “casual” and a little less strongly to “registered” since less casual renters would want to rent a bike on a cold day, but registered renters are probably commuting or have plans that they can’t alter as much to accomodate alternate forms of transport.

If the bike rental company is growing, then “year” may also be moderately positively correlated to “registered” as more renters are likely to become regular/registered renters over time. However, there are only 2 years in the dataset, so the sample may be too narrow in terms of “year” to make this difference significant.

Weather should also affect “total”, “registered”, and “casual” with a weak relationship between “misty…” and the 3 dependent variables and moderate positive and negative relationships between “clear…” and “light snow…” and the 3 dependent variables respectively.

I will be analyzing the total dependent variable, but may split it into the registered and casual variables for deeper analytics.

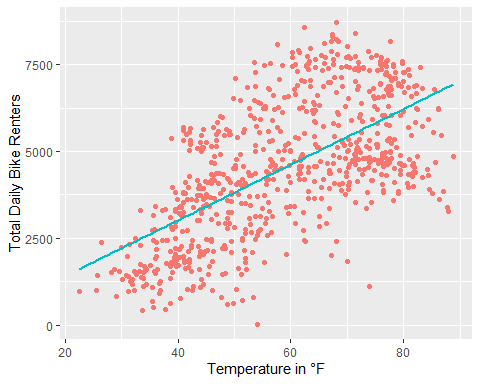
# Correlation Plot



## Insights

1. The large, dark blue dot for the correlation between registered and total users means that they are strongly and positively correlated, so as one increases, the other does very similarly if all else stays the same. However, the casual and total dot is smaller and lighter and thus the two are not as strongly correlated.
   1. This suggests that more users register in advance than just casually rent the bike, all other things equal.
2. The correlations between tempF and both casual and registered users are moderately positive, thus as the temperature increases so do the number of registered and casual users, albeit at a slower rate (given that all other variables stay constant).
   1. This also tells us that registered and casual users react to the temperature in almost the same way. Both groups prefer higher temperatures when deciding whether or not to rent a bike, all other things equal.
3. Year is moderately positively correlated with both total and registered users, but only slightly positively correlated to casual users.
   1. This seems to tell a story of moderate growth in registered users and slow growth in casual users, when all other variables are accounted for. This could mean that users are quickly becoming registered and many casual users are converting to being registered, while new casual users are also growing, albeit at a slower pace.
   2. This is good news for the company, as more total users provide more revenue, and more registered users provide more stability and predictability which can reduce costs and thus boost profits.
4. The correlation between workingday01 and casual users is moderately negative, implying that casual users rent bikes less often on work days, all else being equal. (i.e., as workday01 increases from 0 to 1, or non-work day to work day, casual users decrease)
   1. This makes complete sense as people who rent bikes on a workday are probably going to know of the need for a bike in advance and thus register to rent it, while casual users would use a bike for more leisurely activities. Such activities can’t take place during work/a work day, so they rent bikes less often on work days.
   2. This is valuable data for the company, as they can market the bikes to consumers with leisure time on work days to make up for the reduced demand from consumers who must work on a work day.

# One Independent Variable



This plot shows the relationship between the “tempF” and “total” variables which represent the outside temperature in degrees Fahrenheit and the total number of bike renters on a specific day respectively.

The plot is made using a linear model as it fits the data best and does not exhibit significant patterns in its residual plot. The equation for the linear model is:

Ŷ = -197.83 + 80.217(tempF)

In other words, the predicted total number of bike renters on a specific day is equal to 197.83 subtracted from the temperature in °F times 80.217

### Intercept Discussion

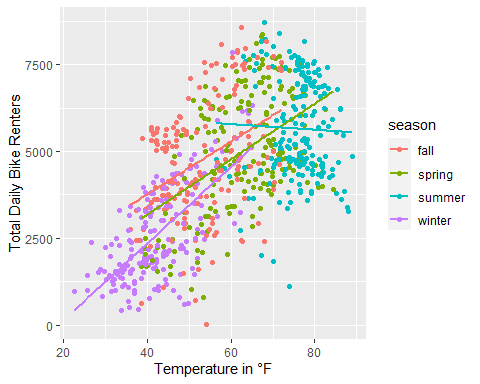
The interpretation of the intercept is not relevant here as it is not possible to have negative bike rentals, even though it is possible for the temperature to be 0°F. However, the dataset’s lowest recorded temperature is 22.5°F, so the model is still relevant in the range of temperatures displayed in the “bikes” dataset.

### Interpretation

The equation can be interpreted as: the predicted total number of bike renters increases by 80.217 (or 80 if considering that people dont come in fractions) for every 1°F increase in outside temperature all else considered equal.

# Two Independent Variables

## Continuation of “tempF” vs. “total”



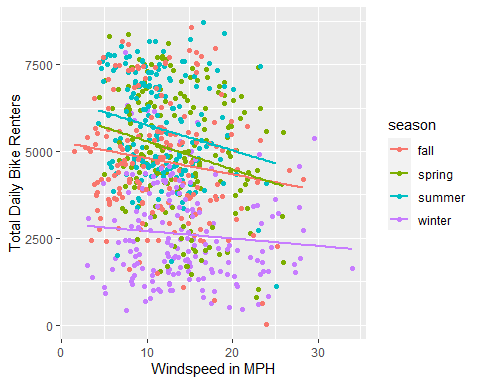
Parallel trend lines do not seem appropriate for this data as the plot shows. Only fall and spring seem to have parallel trend lines while summer has a fairly horizontal trend line and winter has a steeper upward slope than that of fall and spring. This suggests that there is an interaction taking place or that other variables are needed to create parallel trend lines.

The equation for this model is:

Predicted total daily bike renters = 760.31 + 75.411(tempF) - 494.493(spring) - 853.388(summer) - 1342.859(winter)

Fall is the reference dummy variable and all of the p-values for the season dummy variable are below 0.05 suggesting that each season’s predicted difference in temperature from Fall is statistically significant. i.e., at the same temperature, a day in Winter will have about 1343 less bike renters than in Fall, a day in Summer will have about 853 less bike renters than in Fall, and a day in Spring will have about 494 less bike renters than in Fall.

## Windspeed as Independent Variable



In this second model with two independent variables (windspeed and season), the trend lines for each season are much more parallel than before and all have decreasing slopes that are relatively close to each other in steepness. The resulting equation is:

Predicted total daily bike renters = 5308.95 - 50.40(windspeed) + 370.57(spring) + 916.48(summer) - 1980.37(winter)

### Interpretation

For each 1 MPH increase in windspeed, assuming all else is held constant (i.e. season) the predicted total daily bike renters decreases by 50.4. At the same windspeed, the predicted total daily bike renters is about 371 more in Spring than in fall, about 916 more in Summer than in fall, and about 1980 less in Winter than in Fall.

All of the p-values (intercept, windspeed, and dummy seasons) are below 0.05 thus making them statistically significant. However, the non-parallel slopes indicate some sort of interaction or other variable influencing the results.

The equation does bring up some interesting results about windspeed and temperature. It seems as though hotter months lead to higher total daily bike renters when accounting for windspeed. This may be due to the intense heat in hotter months and how the wind provides a nice breeze or wicks away sweat when bicycling but chills you down significantly more in colder months when the outside temperature is already cold enough.

# Many Independent Variables

The reduced model equation is as follows:

Predicted daily total bike renters = 589.547 + 2017.294(year2012) - 730.924(holidayYes) - 1789.6(WeatherLightSnow…) - 472.113(weatherMist…) + 72.284(tempF) - 7.077(humidity) - 51.408(windspeed)

All of the p-values for this model are below 0.05

### Predictions

On a non-holiday in 2012 with clear weather, 75 degrees Fahrenheit, 80% humidity, and 12MPH windspeed, the predicted daily total bike renters is 6845.085 or about 6845 total bike renters.

On a holiday in 2011 with light snow… weather, 40 degrees Fahrenheit temperature, 47% humidity, and 11MPH windspeed, the predicted total bike renters is 62.276 or about 62 bike renters.

### Insights

The coefficients in the equation can tell a business that bike rentals decrease on holidays and days with bad/cold weather and high humidity and windspeed. A business could capitalize on this information and make related accessories available for sale near the bike rental locations. For example, the business could sell sweat-wicking and wind-proof clothing which reduces the problems associated with humidity and windspeed (and colder weather to an extent). Further, they could offer deep discounts on holidays and on days with a poor weather forecast.

# Conclusion

1. The variables “year” and “registered” are strongly positively correlated. This means that more users are becoming registered/regulars as time progresses. Thus, the bike rental business should continue its efforts to expand and convert casual renters into registered renters.
   1. Further, “registered” is highly positively, and moreso than “casual”, related to “total”. This means that the company’s bike rentals consist mostly of rentals by registered users. (This is further backed up by the raw data and the sheer disparity between rentals by casual and registered users). Thus the company should focus its retention efforts on registered users while also attempting to convert casual users (who are more likely to leave the service) into registered users.
2. Casual rentals are moderately negatively correlated with working days. I.e., casual bike rentals decrease on work days. Thus, the company should attempt to reach out to new customer bases like retirees and seniors who are more likely to have the time and need for bike rentals on work days (like for seniors who want to explore the city or go on a leisurely ride).
3. Temperature is strongly positively correlated with total daily bike rentals. Further, this holds true for all seasons but summer, in which the temperature/total rentals trend line is somewhat flat.
   1. The company should investigate the flat summer trend line to better understand its customers and the market during summer.
   2. The company should anticipate more bike rentals on hotter days/periods like summer and plan accordingly. The company could have extra maintenance technicians on standby, deploy more reserve bikes into the field, and reduce promotions to increase profit since the demand will be high enough to cover lost customers due to the higher rental prices.
4. Bike rentals go down when the windspeed increases. The company should take this into account and check local weather forecasts to plan accordingly. Further, the company could capitalize on this to increase accessory sales and increase bike rentals.
   1. The company could sell wind-breakers or offer rentals of wind-proof clothing and other apparel for poor riding conditions that would entice prospective renters to rent a bike when conditions aren’t perfect.
5. Customers tend to rent bikes less in poor weather conditions like snow, rain, cold, high humidity, and high winds.
   1. The company could alleviate customer concerns by offering easy rental cancellations that would make customers’ uncertainty about riding conditions less of a concern as they could always cancel if conditions do turn out to be poor.