Veena Iyer **ST625**

Capital Bike Share in Washington D.C

Abstract:

The concept of bike sharing has been popular in the US since quite some time. This service is available for use 24*7 throughout the year and riders can have access to a bike at any station across the system. Given the environmental and health benefits many major cities have adopted this idea of bike sharing. One such system is Washington D.C's Capital Bike Share which was started back in 2010 and has now grown to about 500 docking stations with more than 4500 bikes. This study comprises of the demand for bikes and the various factors that play a major role in its demand within this system.

Introduction:

Capital Bikeshare is operated by Motivate, a global leader in bike share. This system provides residents & visitors an affordable and convenient transportation option in D.C by helping them to rent bikes to get them from point A to point B. People use bikeshare to commute to work, school, run errands, social engagements etc. There are two types of bike sharing options available for riders: 'casual' – wherein the user rents a bike for a 30 minute ride or a 24 hour period, 'registered' – wherein the user would subscribe to a 30-day plan or an annual membership.

This report analyzes the various factors that influence the demand of bikes by these users.

The data set comprises of the following variables:

- casual - Number of unregistered/non-subscription rentals
- registered - Number of subscription rentals
- season – Season (Spring, Summer, Fall, Winter)
- hr - Hour (6 to 23, where 6 means 6 a.m. - 7a.m. and 23 means 11p.m. - 12a.m.
- holiday - Federal holiday or DC Emancipation Day (1 is yes and 0 is no)
- day - Day of the week (1 to 7, where 1 is Sunday and 7 is Saturday)
- weather - Weather situation
  - 1 = Clear or Partly cloudy
  - 2 = Cloudy and/or Mist
  - 3 = Light Rain or Light Snow
  - 4 = Heavy Rain or Heavy Snow or Ice Pellets
- temp - Temperature (Fahrenheit)
- feelslike - Feels like temperature (Fahrenheit)
- hum - Humidity (percent)
- windspeed - Windspeed (miles per hour)

By exploring the data set, the demand seems to be influenced by factors such as hr, day, temp, weather etc. Various graphs were plotted to study more about the demand for each of these users.

The adjoining graphs gives a rough idea about some of these influential factors to begin with the exploratory analysis.

For casual users, there seems to be a surge in demand during weekends*(fig 2.1)* which looks obvious as bikers tend to bike for leisure during weekends; whereas for registered users, this demand seems pretty constant throughout the week, one of the reason being subscription to an annual membership(*fig 2.2*).

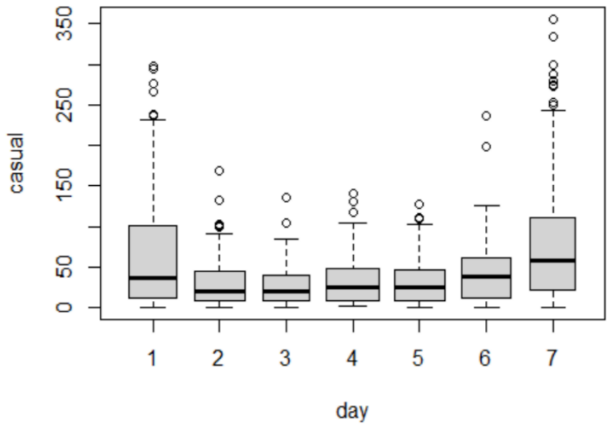


Fig 2.1: Casual users vs day

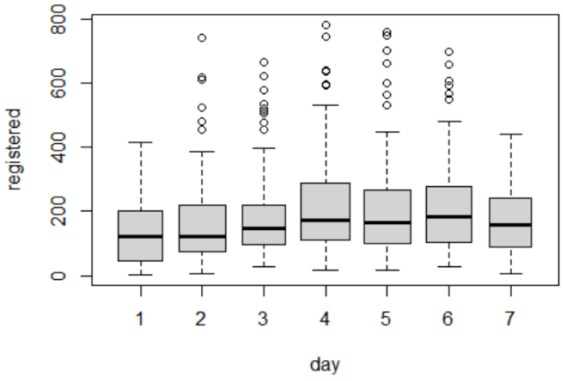


Fig 2.2: Registered users vs day

Bikes are much in demand during work hours (morning 7a.m. to 8 a.m. and evening 5 p.m. to 6 p.m.) for registered users (*fig 2.3*) but looks spread throughout the day for casual users (*fig 2.4*). This explains that hour of the day plays an important role in the demand for bikes.

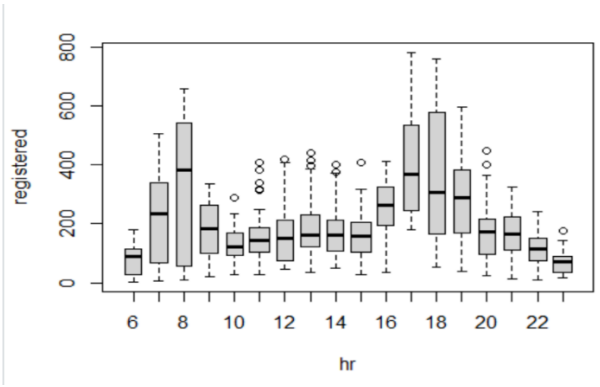


Fig 2.3: Registered users vs hour

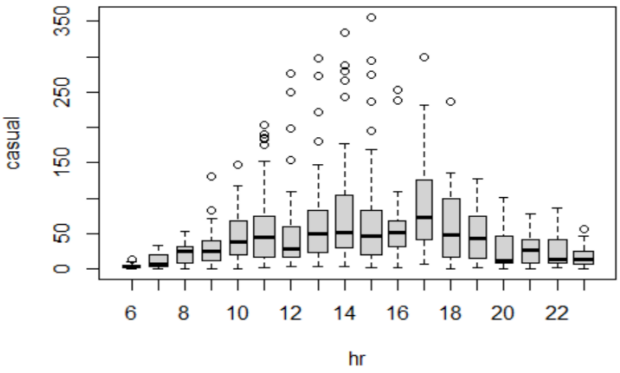


Fig 2.4: Casual users vs hour

Seasons also looks like one of the major factors that influences the demand. For casual riders, the demand is more during summer and spring seasons, while there's a dip in the demand during winters (*fig 2.5*) whereas for registered bikers, the demand persistent throughout the year which seems obvious as registered users may use bikes to commute to their workplace irrespective of the season.(*fig 2.6*)

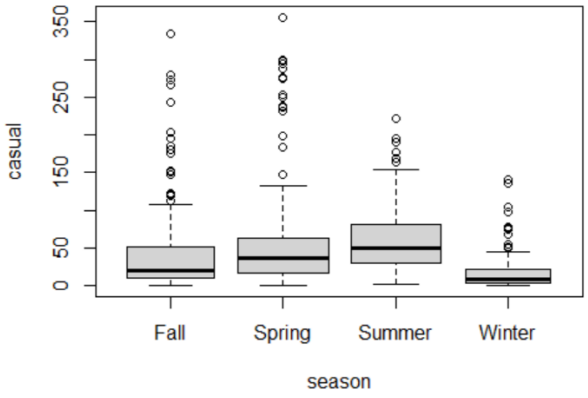


Fig 2.5: Casual bikers vs season

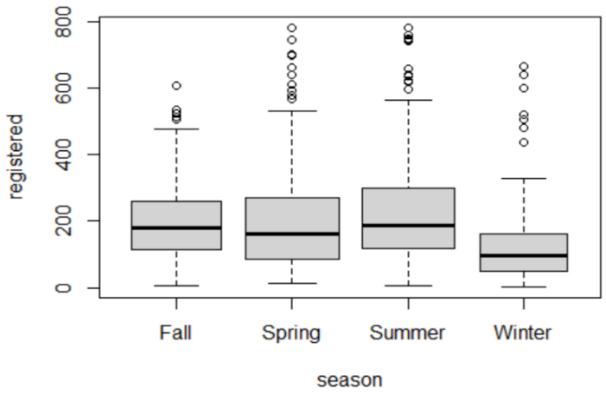


Fig 2.6: Registered bikers vs seasons

The above graphs indicate that hour, day, season could be strong predictors to anticipate the demand of bikes. With some more exploratory analysis, temperature and humidity also looks promising in predicting the demand.

Methodology: Casual users

Model 1: First Order Model

This model showcases the demand of bikes for casual users and included all variables except registered. This model only explained 31.1% of the observed variation in bike demand.

Model1 = lm(casual ~ . – registered, data = BIKESHARE)

Since the first order model did not explain much variation, the correlation between variables were analyzed to eliminate highly correlated variables. Based on the correlation matrix, 'temperature' and 'feelslike' seemed highly correlated (0.987) and hence 'feelslike' variable was dropped from further model building. Also, weather and humidity were moderately correlated with a value of 0.52. Since, these were only moderately correlated, these terms were taken into consideration.

Model 2: Higher order model with degree 2:

As the first order model did not explain much in terms of variation for day, hr and weather, these terms were remodeled using higher order and the variable 'feelslike' was dropped as 'temperature' and 'feelslike' were highly correlated.

Model2 = lm(casual ~ . - registered - feelslike, poly (temp, degree = 2, raw = TRUE) + poly(hr, degree = 2, raw = TRUE) + poly(weather, degree = 2, raw = TRUE), data = BIKESHARE)

This model only explains 50.69% of the observed variation.

Model 3:

There was a rise in demand for bikes during weekends for casual users (*fig 2.1*). Taking this into consideration, a new dummy variable was created for weekends – 'isSaturdayOrSunday'. Also, the

variables like season, holiday and weather were treated as qualitative variables per se. This helped in fitting a better model and explained about 73.01% of the demand of bikes.

model5 = lm(sqrt(casual) ~. - registered – feelslike + IsSaturdayOrSunday*poly(hr, degree = 2, raw = TRUE) + poly(temp, degree = 2, raw = TRUE)+ as.factor(season) + as.factor(holiday) + as.factor(weather), data = BIKESHARE)

The factors like day, hour, temperature, season, holiday and weather highly contribute to the demand of bikes.

Methodology: Registered Users

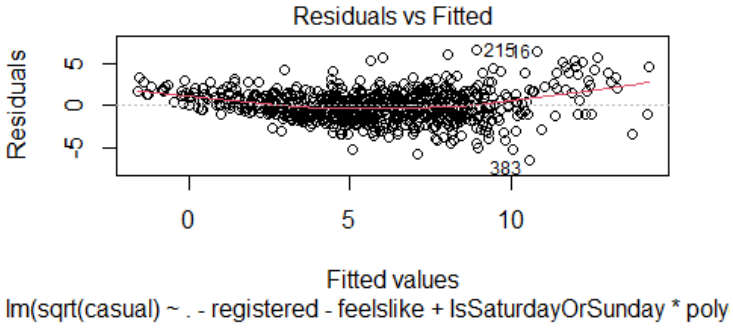
Further extending the analysis to registered users from casual users and eliminating highly correlated factors like feelslike, similar model was constructed. For registered users, hour of the day played an important role (*fig 2.3*) Hence, a new dummy variable 'IsPeakHour' was created. Taking all these pointers into consideration and after extensive modeling, the following model was constructed.

model6 = lm(registered ~ . - casual - feelslike - windspeed + temp*isPeakHour*poly(day, degree = 2, raw = TRUE) + as.factor(season) + as.factor(weather), data = BIKESHARE)

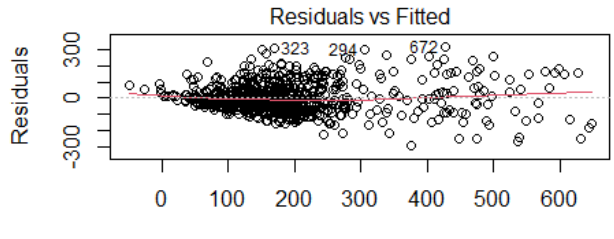
The above model explains 60.36% of the observed variation for the demand of bikes for registered users. Factors like 'holiday' and 'humidity' does not play any major part in the demand for registered users which seems transparent as bad weather would not stop registered users from using bikes to commute to their workplaces.

Results:

The demand of bikes for casual users depends on factors like day, temperature, holiday, hour, season, and weather with weekend being the most important factor that influences this demand. These results seem intuitive because for riders renting bikes for leisure, the above factors would play a huge role. For instance, season would be a hugely contributing factor as visitors/leisure riders would not rent bikes during extreme cold winter. Surge in demand during weekends also follows a similar explanation.



Similarly, for registered users, factors like day, hour, temperature, season, and weather seems valid with hour being the most important factor. Hour can be well-explained for registered users, as people using bikes for commute would tend to rent them during the peak work hours. Temperature, season and weather does not seem to impact the demand of bikes for registered users which looks satisfactory too. Again, feelslike, humidity and windspeed does not seem to be contributing factors to the demand.



lm(registered ~ . - casual - feelslike - windspeed + temp * isPeakHour *

Conclusion:

Thus, we can conclude that humidity, windspeed, feelslike are not highly contributing factors for both registered and casual users. Factors like day, hour, season and weather play an important role in the demand of bikes. Using a higher order model can provide more insight to the demand and its contributing factors, however, overfitting the data might be a problem in this case. Other complex models can therefore be used for check for stronger relationship between the various factors and the demand.

References:

https://www.capitalbikeshare.com/about

https://en.wikipedia.org/wiki/Bicycle-sharing_system