Effect of Seasons on Rental Bike Demand

Effect of Seasons on Rental Bike Demand

Alejandra Robles Lopez, Fabienne Ishimwe and Sameer Kulkarni

Harrisburg University of Science and Technology

Effect of Seasons on Rental Bike Demand

Abstract

Over the past few years bike sharing systems have grown rapidly in the United States. However, it can be challenging for company owners to know just how many bikes they need to have throughout year considering riders do not have to register for access to this service. We conducted data analysis using two year sample bike sharing data from Capital Bike Share Company based in Washington DC with the hypothesis that there will be a significant difference in the means of total rental between summer and winter. One- way ANOVA was used to test this hypothesis. We rejected the null hypothesis, supporting our claim and then developed the best fit multiple linear regression model.

Effect of Seasons on Rental Bike Demand

Effects of Seasons on Rental Bike Demand

The idea of bike sharing started back in 1965 by a group of activists in Amsterdam. August of 2008 that Washington DC launched a 120-bike pilot program called SmartBike DC that would last a couple of years until in 2010 it was expanded and change its name to Capital Bikeshare. Nowadays, Capital Bikeshare (abbreviated CaBi) also serves in various cities of Virginia such as Fairfax County, Arlington County and the city of Alexandria, and also serves Montgomery County in the state of Maryland. Starting with 400 bikes at 50 stations it now operates more than 3,700 bicycles at 400 stations.

Throughout the years many different researchers have analyzed the different effects weather has over bike sharing in different cities and countries. A great example, is a Swedish analysis that was able to find that bike trips decreased from the summer to the winter time by 47%; precipitation and temperature were the factors that most affected the seasonal bike sharing riders (Bergstrom and Magnusson, 2003). Most significantly in the United States and Canada, we found a study that estimates the percentage of bike trips on a work day in different cities in both countries. The elements that drastically had a big impact were once again, temperature and precipitation (Pucher and Buehler, 2006).

Another survey was given to about 800+ people from 35 different countries from Europe. The survey was to determine if people in the sample were interested in holiday cycling. Through the questionnaire they were able to determine that “the respondents in the sample are interested in holiday cycling, as 47.8% visited cycling-friendly cities once or twice and 16.5% visit such cities often.” Also, “of the respondents who visited cycling-friendly cities, 44.2% visited either Copenhagen or Amsterdam. While on holiday, 32.2% of the respondents rented a bicycle once or

twice and 7.0% rented a bicycle often." (Kaplan, S., Manca, F., Nielsen, T. S., & Prato, C. G., 2014).

We would like to investigate this research better as we believe that in addition to specific weather details, seasons overall highly affect rental demand. Our claim is that rental demand changes significantly between summer and winter. Our goal is to also to research more and determine exactly which factors are good predictors of rental count by creating a multiple linear regression model. We have used Capital Bikeshare dataset which we found on the UCI machine learning website. The insights acquired from this research can be used by similar companies to better allocate resources and make their business decisions accordingly.

Method

Our sample dataset obtained electronically from UCI website, contained daily rental information between 2011 and 2012 with 731 observations, 13 explanatory variables and 1 response variable. The response variable count is a sum of casual and registered users, named "causal" and "registered:" respectively. However for the purpose of our research, we did not focus on the type of users but rather the total rental count "cnt". The independent variables can be grouped in three categories, mainly time, type of day and weather. We used R integrated development environment, R studio to perform descriptive, exploratory and inferential analysis of the data. We started by performing Data preprocessing, which involved making sure there are no missing values and changing some columns data types. Most of the independent variables were changed to nominal categorical variables.

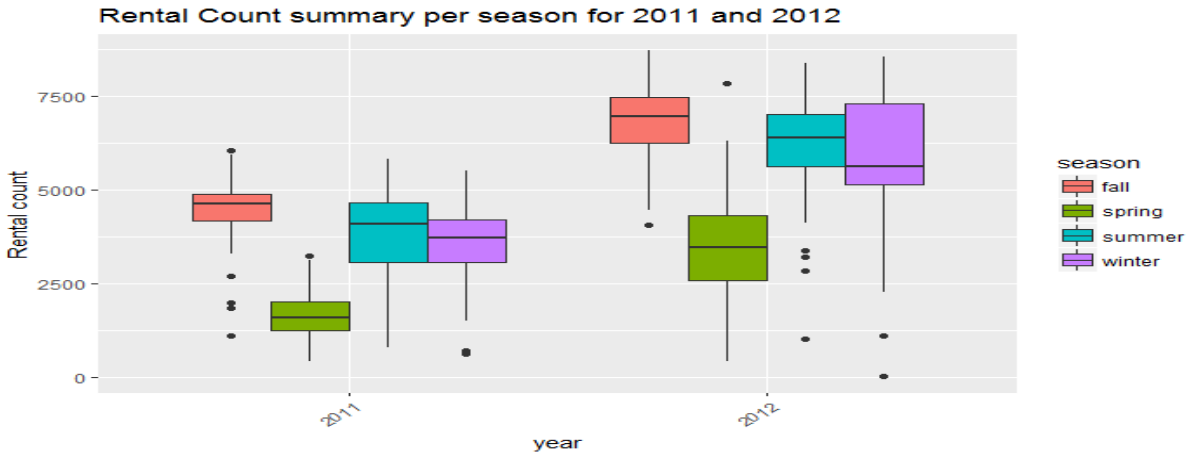
Effect of Seasons on Rental Bike Demand

Results

Descriptive and well as graphical exploratory data analysis was performed on the season independent variable and the rest of the independent variables. The table below and box plot below shows that bike demand was significantly in higher in 2012 than 2011. The average daily rental count seems to have been high for 2012 than 2011 and significantly different across seasons. The box plot also shows a pattern for rental bike demand across seasons for these 2 years; we can see that spring bike demand seemed to have differed a lot from the rest of the seasons

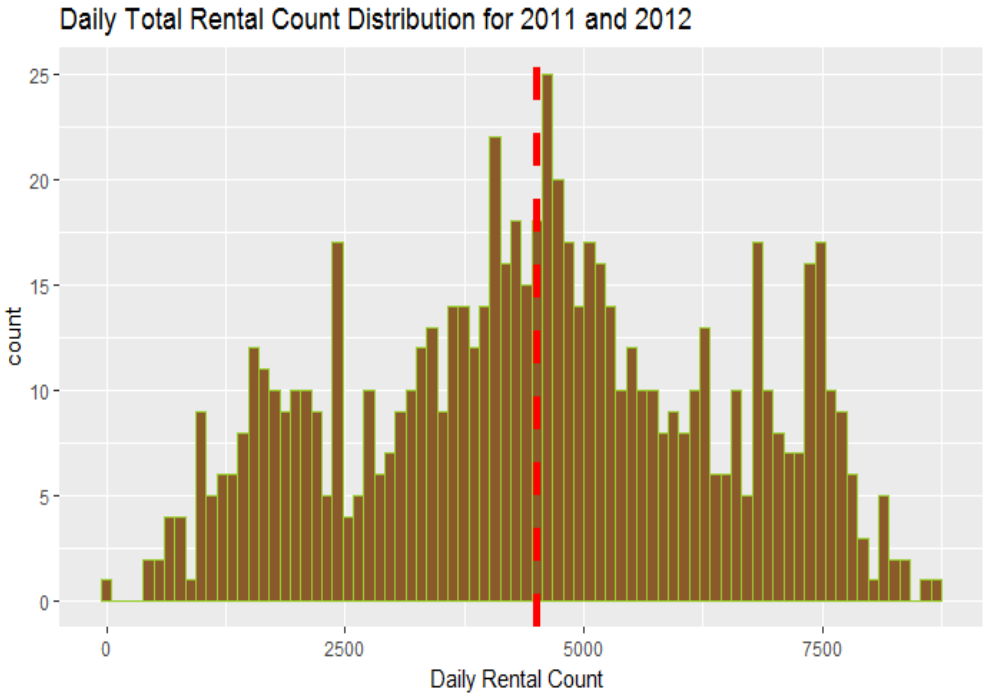
Rental count statistics for 2011 and 2012 per season

	2011			2012		
	Total count	Average daily rental count	rental count sd	Total count	Average daily rental count	rental count sd
Fall	419650.00	4464.36	798.35	641479.00	6824.25	912.76
spring	150000.00	1666.67	614.01	321348.00	3531.30	1341.20
summer	347316.00	3775.17	1138.90	571273.00	6209.49	1221.35
winter	326137.00	3664.46	983.01	515476.00	5791.87	1598.52



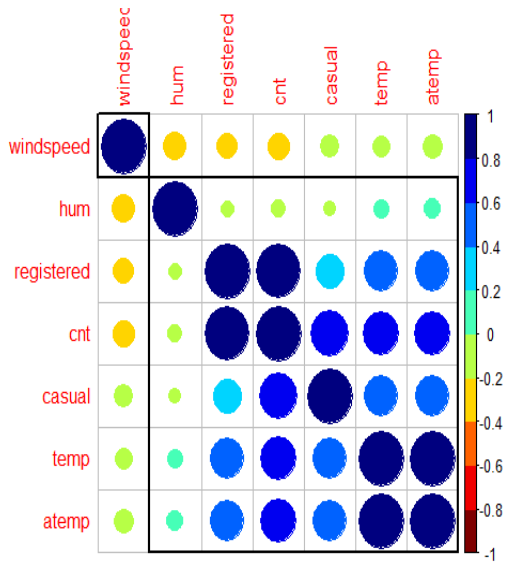
Effect of Seasons on Rental Bike Demand

We can also see that there is a difference in rental count means between summer and winter and in order to test if the difference is statistically significant and deduct inferences, we used one-way ANOVA technique. The AVOVA model of rental count vs season returned a p-value < 0.05. We then used Tukey HSD function for performing multiple pairwise-comparisons between the means of season groups. The function returned of p-value 0.34 for difference in means between winter and summer. To test the validity of the ANOVA model, we used Using Levene's test to check the homogeneity of variances and the normality plot. We also observed the distribution of our response variable count using a histogram.



Additionally, as part of our goal was to develop a multiple linear regression model, we plotted a correlation matrix for numerical variables and used Chi-Square test for collinearity of

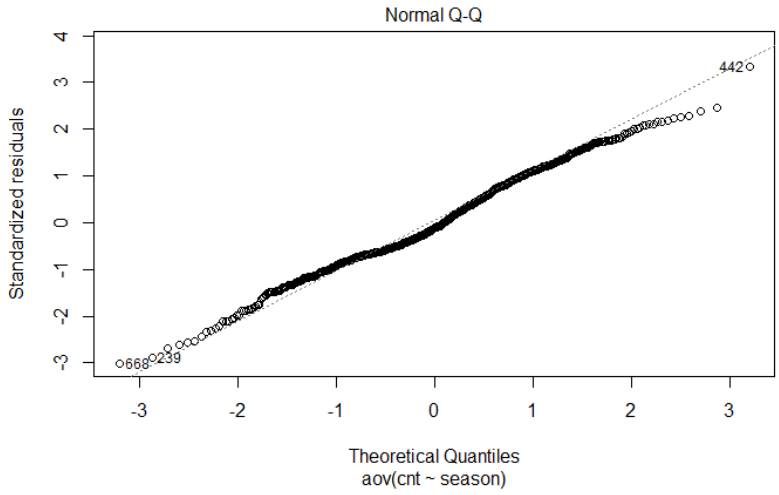
categorical variables. We then developed the best fit multiple linear regressions using backward elimination, tested its validity and made some predicts.



Discussion

After performing the descriptive and visualizing the relationship between rental count and season, we could see that our claim that there is a significant difference between rental count during summer and winter might hold. Additionally, The AVOVA model of rental count vs season proved there was a significant difference in the means of rental count between season groups since it returned p-value < 0.05. However surprisingly, the Tukey HSD function returned a p-value 0.34 for difference in means between winter and summer, implying that there was no significant difference in means of daily rental count during summer and winter, forcing us to fail rejecting the null hypothesis. This was unexpected as we presumed people would be much less likely to bike during winter. This also conforms to previous literature as they have suggested that more detailed weather conditions such as temperature and humidity affect bike use. However,

further testing was needed before we could make a final conclusion on this. Using the Levene's test to check the homogeneity of variance across season groups, it returned a p-value is > than the significance level of 0.05. Therefore the variance across season groups was not statistically significantly different and we assumed the homogeneity of variances in season levels. The normal probability plot of the residuals approximately followed straight line, suggesting normality of rental count across different seasons. Therefore we concluded our plot to be valid and failing to reject the claim that there is no significant difference between the means of total rental count during summer and winter.



Finally, as previous literature has suggested, rental count was best predicted by detailed weather conditions and part of our goal was to find the best fit model involving those variables. The final best fit multiple linear regression models created accounted for 50% of variation in the data and suggested the rental count was best predicted by a combination of four variables,

season, holiday, temp and hum. Evidently, rental count was not be best predicted by seasons themselves but, combined with type of day, temperature and hum, decent predictions can be made. However, one limitation with this research is that the data was not enough to make inferences on the whole country given the difference in weather across the United States throughout the year. As the bike rental systems are rapidly growing, future studies should be done with enough samples from different states. This would insure a better predictive model and perhaps more insights that we were able to gather.

Effect of Seasons on Rental Bike Demand

References

Bergström, A., & Magnusson, R. (2003, October). *Potential of Transferring Car Trips to Bicycle During Winter.* Retrieved from https://doi.org/10.1016/S0965-8564(03)00012-0

El-Assi, W., Mahmoud, M. S., & Habib, K. N. (2017, November 5th). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Toronto, Canada.

Fanaee-T, Hadi, and Gama, Joao, 'Event labeling combining ensemble detectors and background knowledge', Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg.

Kaplan, S., Manca, F., Nielsen, T. A., & Prato, C. G. (2014). *Intentions to use bike-sharing for holiday cycling:*. Trafikdage på Aalborg Universitet .

Lazo, L. (2017, October 2nd). *The Washington Post*. Retrieved from Capital Bikeshare gears up for another expansion — and competition: https://www.washingtonpost.com/local/trafficandcommuting/capital-bikeshare-gears-up-for-another-expansion/2017/10/02/bcf81b4a-a2fe-11e7-ade1-76d061d56efa_story.html?utm_term=.d2fb28e30e0a

Pucher, J., & Buehler, R. (2006). Why Canadians Cycle More Than Americans: A Comparative Analysis of Bicycling Trends and Policies. *Transport Policy, Volume 13, Issue 3,* 265-279.