ANALYSING CAPITAL SHAREBIKE DATA

Kimball Wu, Pei-Yi Cheng, Ling Tan, Sumeet Mishra
Indiana University Bloomington
School of Informatic, Computing and Engineering

April 25th 2019

ABSTRACT

Bike sharing systems are a fast growing smart transportation trend today, with more and more cities around the world implementing them. They are emerging as an innovative agent that can help reduce the burden on the public transport system as well as reduce traffic congestions in cities. They are also a healthy and eco-friendly way to get from point A to point B for daily commuters.

The growing demand for bike sharing makes it a hot topic for investors, entrepreneurs and start-ups to invest their time and money in it. However, along with all the advantages, bike sharing systems also pose new sets of challenges to be addressed and solved. Some of the primary challenges are identifying with over or undersupply of bikes, and how to interest more user to choose the bike share system.

In this paper we will address these problems by explore users' travel patterns and predicting the number of bikes needed for the system at any given point of time under different influencing factors. We are using system usage record generated by *Captial Bikeshare* from the city of Washington D.C.

INTRODUCTION

Capital Bikeshare is a bike sharing system that operates in greater Washington, D.C. and Arlington area. The company has provided an environmental friendly alternative way to commute around D.C other than taking the Metro or driving. In order to help the company to interest more potential users to grow their profit, save on the maintenance cost and be more competitive in the bike sharing market. The company has made their usage data public for researcher and professional to analyze their data to help them thrive in the market.

The main purpose of this paper are:

- Analyze traveling behavior under different conditions and discover user's profile
- Fit a model and predict bike usage count under different conditions.
- Using the prediction model to help the campany to avoid over or undersupply of bike and save on bike maintenance cost.
- Introduce new promotions that would interest potential users.

DATA DESCRIPTION

The dataset was compiled by *Fanaee-T and Gama* and made available through the UCI machine learning repository[1]. It consists of two years of hourly aggregated bike system usage count data.

It has a total of 17379 data points collected from dates between 2011-01-01 and 2012-12-31.

The dataset consists of three sets of variables, with a total of 16 variables. First is the usage count data generated by the system in a specific hour of the day for the registered, unregistered and the total users. The second set of variables are the date-time information about the hour for which usage count was generated including the hour, the date, day-of-the-week, season-of-the-year (Winter, Spring, Summer, and Fall) and weekday

or weekend, work or non work day, and holiday. The last set of variables are the hourly weather information including temperature, humidity, wind speed and the weather conditions like sunny, cloudy, mist, rain, and snow, etc.

The following table shows the attributes and variables:

Temporal variables	Statistic	Weather variables	Statistic	Response variable
Hour	0-23	Temperature	-7.06 - 39 °C	Casual rider*3 Count
Workingday	0, 1*1	Humidity	0-100%	Registered rider*3 Count
Month	1-12	Windspeed	0-56 mph	Total Usage Count
Year	2011,2012	Weather	1,2,3,4 *2	

Table 1: Attributes

*1: 1 = working day, 0 = holiday or weekend

*2: 1 =Clear or partly Cloudy,2=Light Snow or Rain',3=Misty and Cloudy,4=Heavily Snow or Rain

*3: **Casual Rider** is a user purchased a Single Trip, 24-Hour Pass, 3-Day Pass or 5-Day Pass, and a **Registered Rider** rider are Annual Member, 30-Day Member or Day Key Member. **Casual Rider** +**Registered Rider** = **Total Usage Count**

TRAVELING BEHAVIOR ANALYSIS

How is the Bike Rental Demand Affected by Different Type of Days?

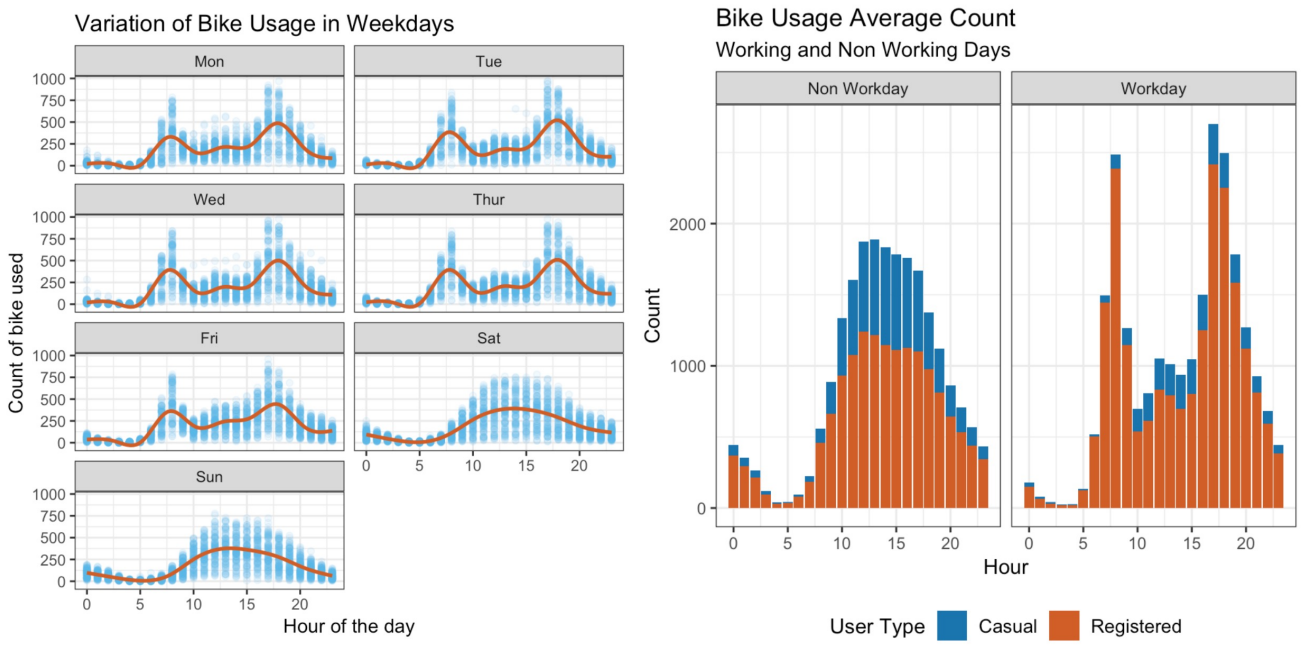


Figure 1: A figure

The line plot of hours of the day against bike rental count categorize by days of the week shows the difference in rental demand for weekday and weekend in different hours.

The peak of demand during weekdays is around 6 to 10 A.M in the morning, and 4 to 5 P.M. in the afternoon, possibly due to people are commuting from home to work, and done working and in need of transportation to go home. We can also see that the bike usage count has a dip at around 12 P.M. during weekdays, whereas around the same time during a non-work day shows a peak of demand of the day. Moreover, the usage count remains active later in the midnight during the weekend than a weekday.

As the same pattern shows more clearly in the bar plot of the hours of the day against bike rental count categorize by working and non-working days, we can also observe that the most demanding usage of the bike is at 13 P.M on a weekday and 17 P.M on a working day. In addition to the number of bike usage, there are more unregistered(casual) user using the bike on non-working days than a working day, possibly mostly used by tourist traveling to D.C.

How is the Bike Rental Demand Affected by Seasons and Weather Conditions?

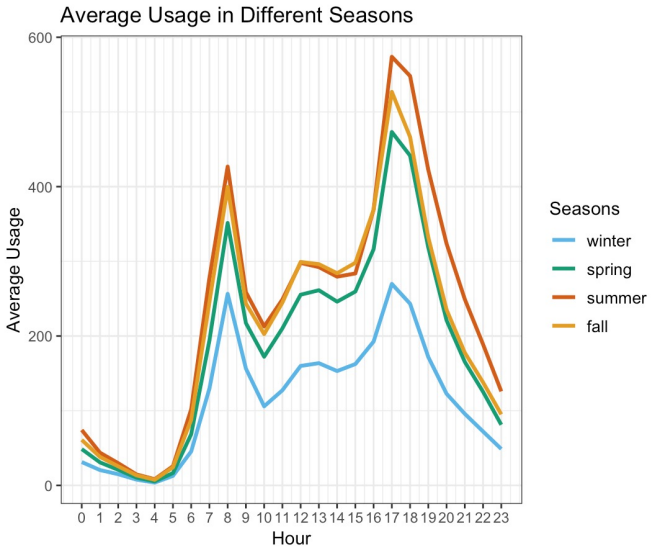


Figure 2: A figure

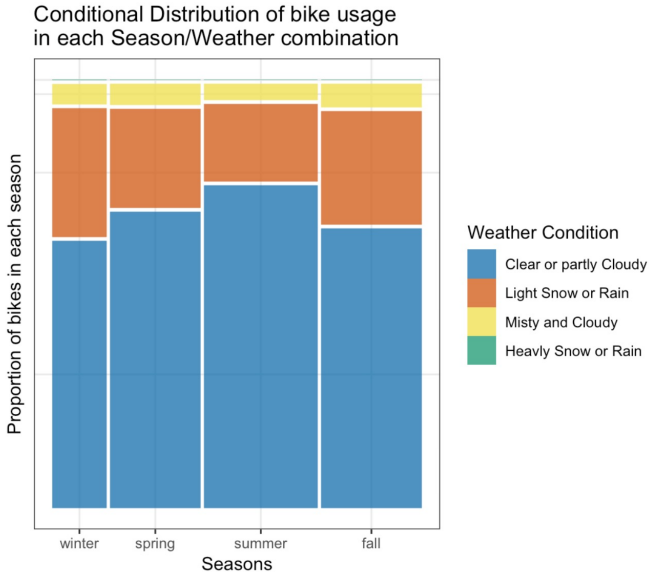


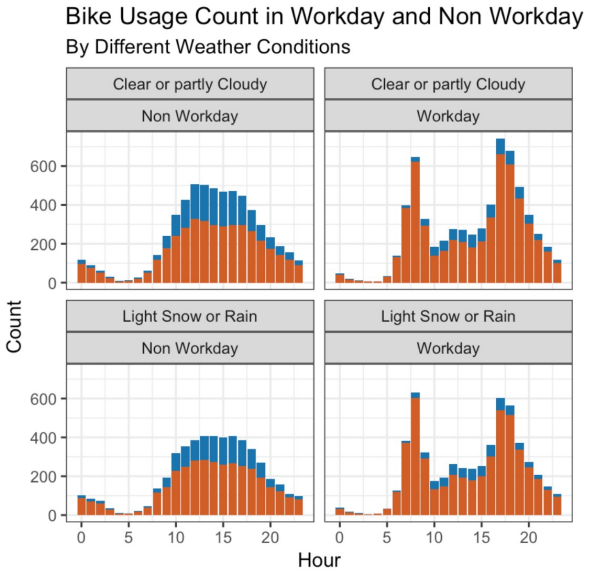
Figure 3: Another figure

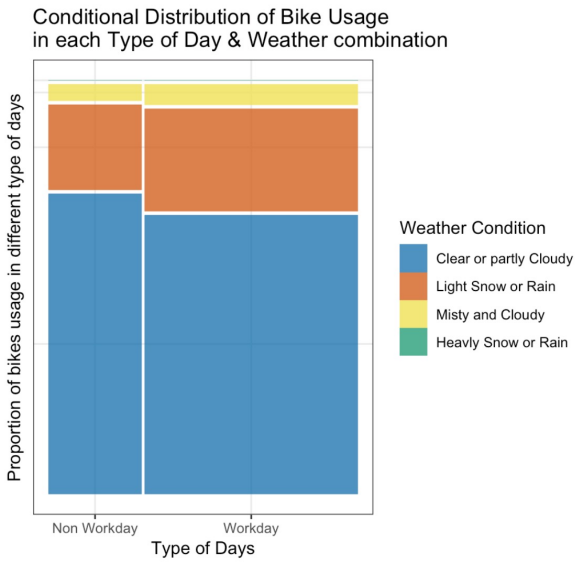
The line plot of bike usage against different hours in a day grouped by different seasons count reveals that there is a seasonal trend with bike usage. Usage demand is generally low in Winter and it peaks in Summer. Therefore, seasons can be one of the determining factors that affect bike rental count.

We can also see similar seasonality pattern to the conditional distribution of the count in each season by different weather conditions; winter shows the lowest in usage and Summer shows the highest. Also in different weather conditions, it shows different overall count. For example, rainy and snowy days show significantly lower average usage count whereas for heavy snowing or raining, there is almost no usage of the bike.

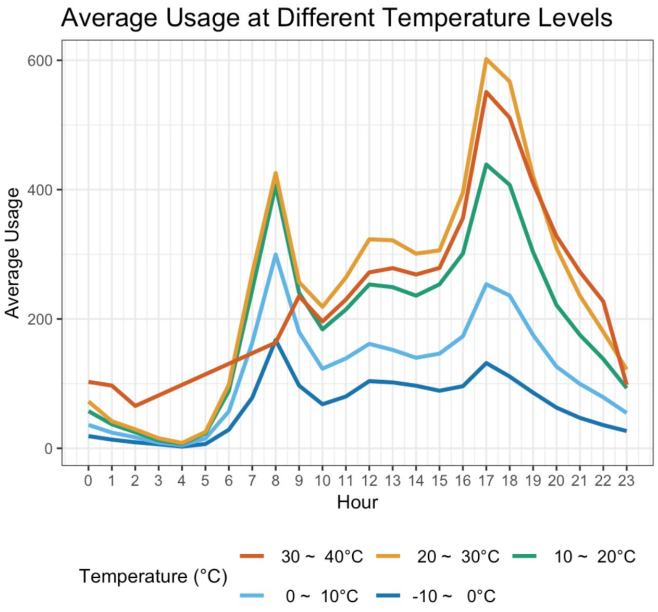
What would happen if it's raining and snowing outside and it's a workday? Would the bike usage have a similar pattern overall? To answer these questions we plotted the following graphs to compare the bike usage in different weather condition such as clear and partly cloudy, light snow and rain, etc against hours of the day grouped by working and non-working days.

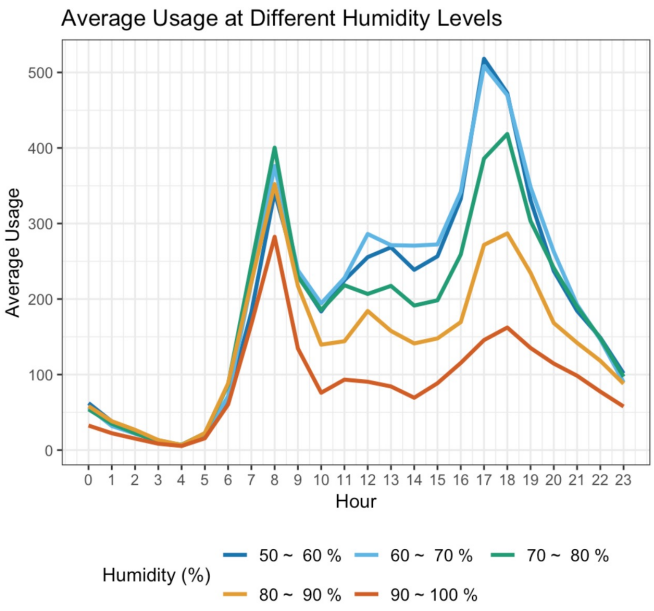
We observed that during non-working days, it show a slightly lower average usage count throughout the day. However, it's very interesting to observe that during working days even if it's snowing or raining outside the usage of the bike did not drop exponentially. In fact, the average usage during morning rush hours is almost the same as if it's clear or partly cloudy. This result confirms our early assumption that during working days people mostly use the bike to commute between home and work.





How is the Bike Rental Demand affected by Levels of Temperature and Humidity?





The humidity and temperature plots show that generally the higher the relative humidity and temperature, the lower is the bike rental demand. However, when higher the general trend of average bike usage does not change, it still shows two peaks from 8 A.M and 18 P.M.

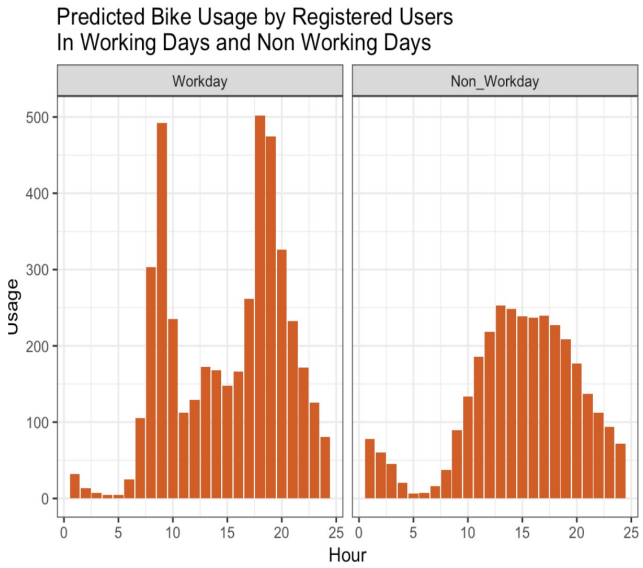
FITTED MODEL

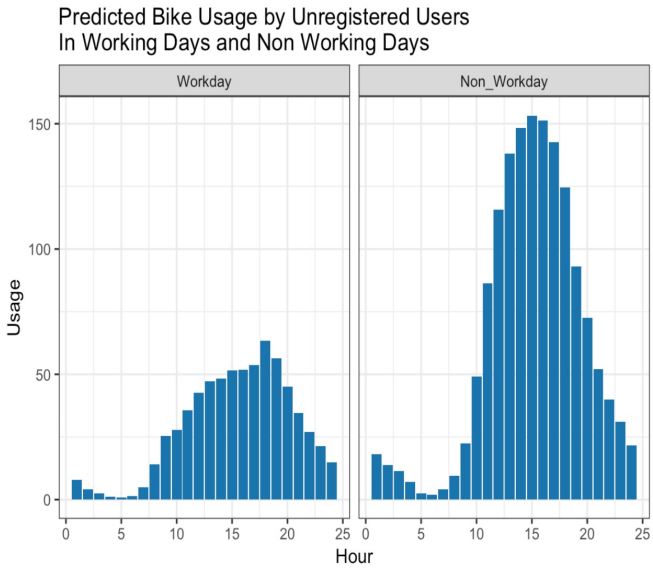
From the result we observed from above, we can say that there is an interaction in play. We have tried to fit a *linear* and *loess* model, however it does not shows the pattern we are looking for. Therefore, we have decided to use a *poisson* model, since our response variable is the count of bike usage.

First we are using **weather** and **workingdays** with interaction with **hours** to predict **casual** and **registered** bike user count. Here is the model:

glm(registered ~weather + workingday * factor(hr), family = poisson, data = bike)
glm(casual ~weather + workingday * factor(hr), family = poisson, data = bike)

The following figures are the predicted bike usage count by registered and unregistered users.

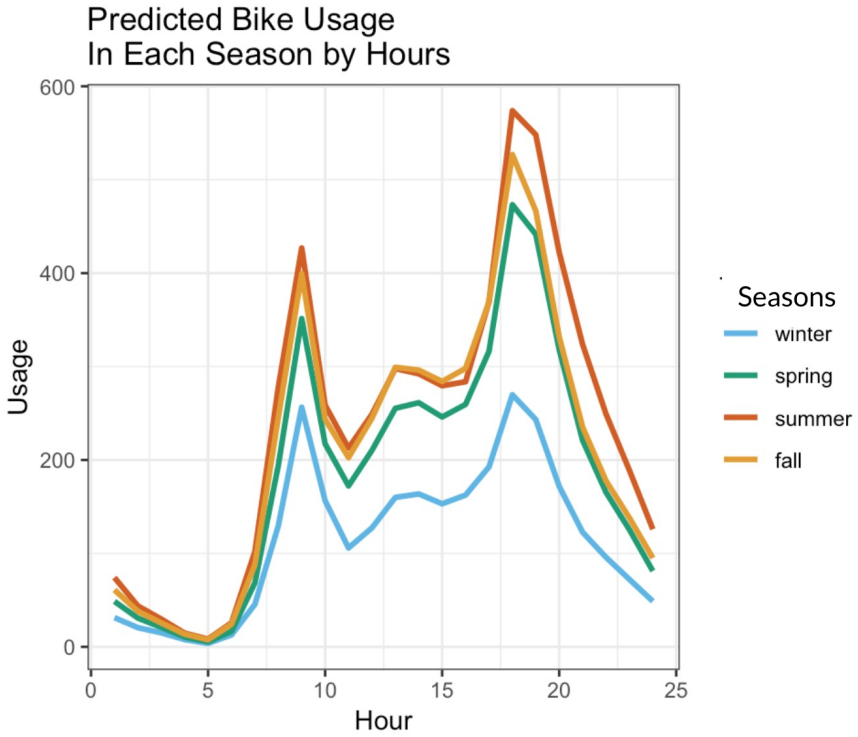




The predicted model follows our previous observation, shows that there is a higher demand for bike usage by the unregistered bike user on the non-working days other than the working days.

Another model we fitted is predicting the overall bike usage in a different season of a year.

glm(count ~seasons * factor(hr), family = poisson, data = bike)



CONCLUSION

We can conclude the following from our observations and prediction:

Bike rental demand difference in work and non-work day.

- On a working day there is a higher demand for bike rental during morning and afternoon rush hours.
- On a non-working day, the rental demand increases starting from morning and reaches its peak around 1 P.M. Then slowly decreases.

Bike user groups difference in work and non-work day.

- Registered users are using the bike mostly for commuting between home and office on working days. Also, using it for personal travel needs on the non-working days. The unregistered users using it in non-working days are mostly tourists and visitors.

Temperature and weather condition affects bike usage.

- There is a significant seasonal trend with bike rental demand, such as during the summer and fall, there is higher usage of bikes whereas during winter and spring the usage of bikes drops.
- When the weather condition gets more severe, the bike usage drops.

After analyzing the usage pattern of the bike user, we have the following business suggestions for the Capital Bikeshare company to consider :

- In order to interest tourists and visitors to choose Capital Bikeshare on their trip to D.C, we would advise offering a promotional **Weekend Pass** for $12, pass valid starting 0:01 A.M Saturday till 11:59 P.M Sunday.
- We would suggest offering a pass or a **Seasonal Membership** for $35, valid for 3 consecutive months to people willing to register themselves and use it to commute to work using Capital Bikeshare during summer and fall. This is to increase the number of users and to create a habit of using bikes.
- Also we would suggest initiating an outreach program and building cooperation with companies located in D.C and Arlington area to set up business cooperation and offer **Corporate Memberships** to their employees who will register for a healthier lifestyle.
- Finally, in order to reduce bike maintenance cost, we would adjust the number of bikes operational on the street according to the rental demand in different seasons. For example, in winter, we would suggest removing half of the number of bikes than used in summer from being operational.

We are hoping the Capital Bikeshare can interest more potential users and provides residents and visitors with a convenient, fun, affordable and environmentally friendly transportation option for roaming around greater Washington Metropolitan Area.

REFERENCES

Dua, D. and Graff, C. (2019). *UCI Machine Learning Repository Bike Sharing Dataset Data Set* . http://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset