

**Final Assignment Project**

**Bike Share Rental Analysis**

Yi Yang

**ALY6070 Communication and Visualization for Data Analytics**

**Summer 2020**

**Professor Jack Bergersen**

**08/16/2020**

**INTRODUCTION**

*Background*

As a relatively new way of transportation, bike sharing programs are becoming more and more common in major cities of the U.S. and they are usually nestled into parking spaces or next to subway station entrances for people who need to get to their destination quicker and within a short distance.

Bike share business has grown rapidly. According to a report from the National Association of City Transportation Officials (NACTO), since 2010, 123 million trips have been taken on bike share bikes in the U.S. And there are 35 million trips taken in 2017, which is 25% more than in 2016. Therefore, we can observe that there is a giant market for this business to expand.

*Dataset Introduction*

Our two given datasets are in CSV file named “day” and “hour” which are obtained from the Capital Bikeshare system that contain bike rental information from years 2011 and 2012. They also include attributes of “instant” which is record index, “dteday” which is dates from 1/1/2011 to 12/31/2012, “season” of 4, “yr” which is “year”, “mnth” which is month, “hr” which indicates hour of day, “holiday”, “weekday” or “workingday”, “weathersit” with 4 levels from Few clouds to Heavy Rain, weather factors like “temp” in Celsius, “atemp” which is feeling temperature in Celsius, “humidity”, and “wind speed”. Besides, there are both total bike rental users as in “cnt”, and “registered” or “casual” users recorded to help the company differentiate among all users.

*Business Problem and Goals*

Our business problem is that the bike share business/company still needs to grow and they are not sure how to find better ways to attract more loyal/ registered customers.

Hence, our business goal is to help the bike share company understand the distribution of registered and casual users and explore how these are changed over time. Additionally, we will observe what weather factors have the last impact on total rentals. Finally, we will predict how much rental volume will grow in the future.

To achieve this, in this assignment, we will propose 4 research questions and using visualization tools in Tableau and Excel to seek for answers and visualize our findings toward our datasets. Moreover, we will conclude our findings and try to provide some insights to help the bike share company grow.

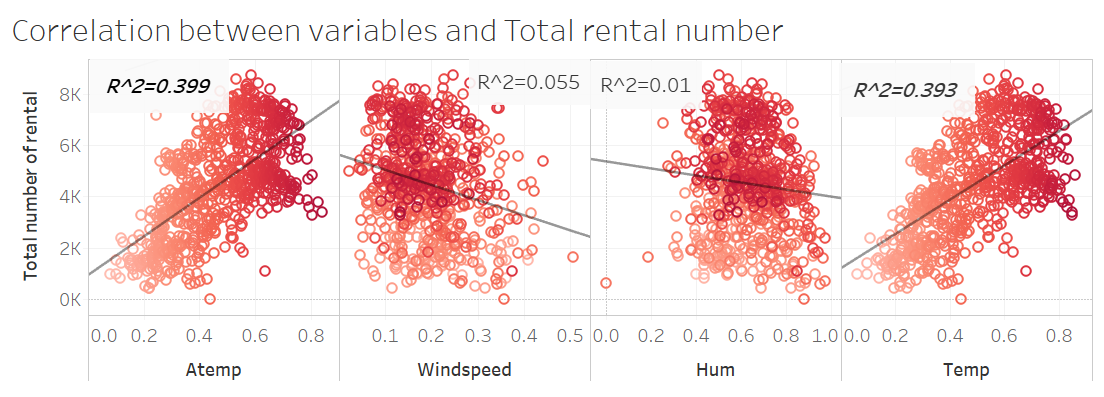
**Research question 1:** *“which weather factors have the greatest impact on rental volume?”*

**Step 1: Data processing**

Since the “day” dataset doesn’t have any null value or error, we can just start the analysis part. First of all, we run the statistical test (regression) on every weather variable and plot the dashboard on Tableau. We removed some factors that don’t show obvious linear relation and kept the factors that have the most impact based on the p-value and linear relation with the rental numbers.

**Step 2: Choose a Display**

Next, we decided to choose a scatter plot to display our results since it can easily show the linear regression results between the factors and rental numbers. On the other hand, the “weatheesit” attribute is the key factor that directly impacts the rental numbers. However, it didn’t perform well on the scatter plot. Instead, we chose a bar chart to display the average rental number on different weather.

****

*Figure 1.* Correlation plot

**Step 3: Draw Attention**

In the scatter plot, we used red layered color to drag the audience’s attention and put all the variables together, so they could easily tell the linear regression between each variable. In the bottom, the bar chart, instead of using the 1,2,3, we changed the name to clear, Mist+ cloudy, and heavy rain+thunder. Also, we used the average number to replace the total number. By doing this, we decreased the cognitive load of the audience and helped them directly connect with the variables.

**Step 4: The Story**

By running the statistical test, we can conclude that “temperature”, “wind speed”, and the weather are the most important weather factors that affect the rental numbers. When the weather is clear, the variable “temperature” is higher but “wind speed” is lower. Our story does match with common knowledge. When the weather is warm and clear, people will more like to go out and rent the bike.

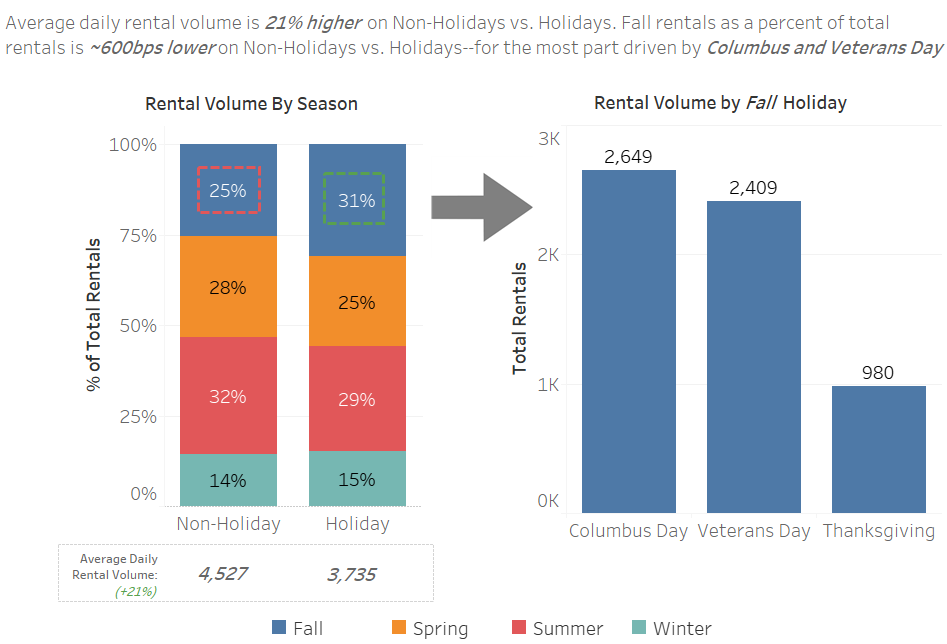
**Research Question 2:** *“How do holidays affect rental volume?”*

**Step 1: Data Processing**

While the data did not need in-depth cleansing of the fields used to answer this question, we did need to create a few calculated fields in Tableau. For example, we had to create a daily average and percent of total. We also had to use IF and ELSEIF statements to rename binary columns and give the holidays a name based on the date.

**Step 2: Choose a Display**

Using Tableau, we created a dashboard to display our findings. The left chart shows the composition of Non-Holiday and Holiday rentals by season. Additionally, the average daily rental volume is given under the chart. This is important to include because it establishes the quick answer--rental volume is lower on holidays--before getting into the depths of the analysis. The stacked bar allows the reader to quickly conclude that fall rentals make a larger portion of total rentals on holidays as compared to non-holiday. This is the only season that displays this significant trend, which is why the arrow guides the reader to the right hand chart where we see the volume by fall holiday. We’re then able to conclude that Columbus and Veterans Day drive much of the volume fall holiday volume.

****

*Figure 2.* Rental Volume by season and holiday

**Step 3: Draw Attention**

The dashboard is set up in such a way that the reader can walk through the story left to right. Additionally, the consistent use of blue for fall and the arrow walking the reader from the left to right chart helps create an understanding without needing to hear the talk track.

**Step 4: The Story**

Average daily rental volume is 21% higher on Non-Holidays vs. Holidays. However, Fall rentals, as a percent of total rentals, is ~600bps lower on Non-Holidays vs. Holidays--for the most part driven by Columbus Day and Veterans Day.

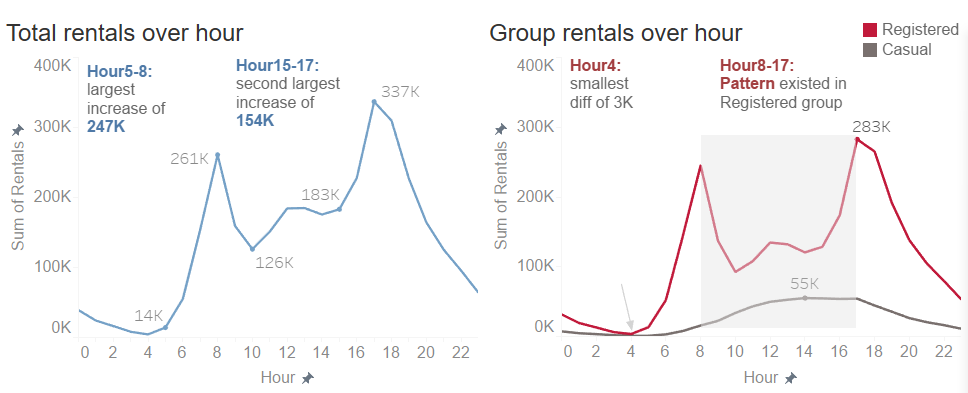
**Research question 3:** *“How does time-of-day affect rental volume?”*

**Step 1: Data Processing**

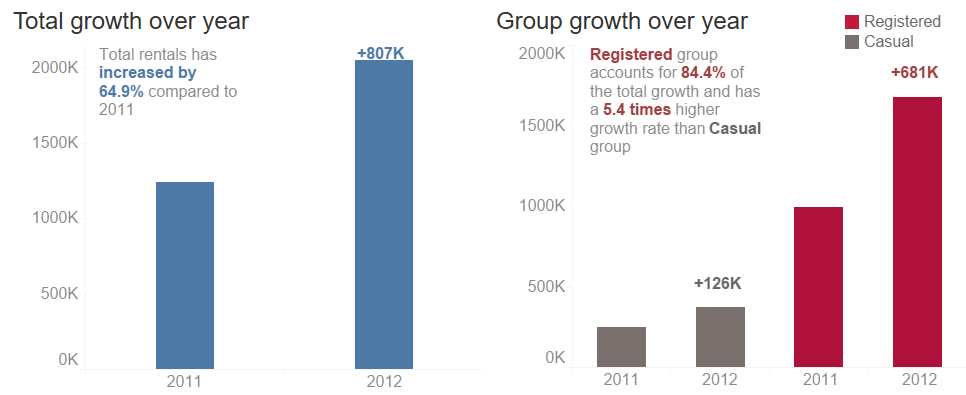
We were interested in how to increase user volume based on hourly usage, and how do the results contribute to our long-term strategy. We split the “hour” dataset into two parts, one for 2011 and the other for 2012. Therefore, we had three datasets ready for analysis: 2011, 2012 and the total.

**Step 2: Choose a Display**

To better examine the relationship between the total and user groups, we plotted line charts respectively for the trends of rentals using Tableau. We expanded the range of y-axis in group trends because that way we can horizontally compare it with the total trend.



*Figure 3.* Rental volume over hour

Next, we wanted to know more details about the majority user, leading to this question: how much of an impact the registered group can have? Vertical visualization might be more helpful in answering this. Corresponding to Figure 3, we plotted a bar chart containing total rentals in 2011 and 2012 to calculate the difference and growth rate. Then we did the same to each group, and concluded that the registered group indeed was the vast majority of our users as it accounted for 84.4% of the total growth. Hence, with the existing data, the research question had become “how to increase the registered user volume?”. More information will be needed if we want to focus more on the casual group.

*Figure 4.* Rental Growth over year

**Step 3: Draw Attention**

Overall, we used three colors on all charts, blue for total, red for registered group, and dark grey for casual group. Red and grey generated a clear visualization and strong contrast. We decluttered by using light grey on labels to push them to the background and also removing grid lines. Highlighting important numbers and texts effectively reduced the cognitive load for the audience as they do not need to extract information by themselves. With the “z” pattern in mind, we aligned the titles and textboxes to the top left and legends to the top right.

**Step 4: The Story**

From Figure 3, we saw the registered group and total rentals had a very similar hourly trend, meaning that they had a significant relationship with each other. Further, Figure 3 showed that the registered group reached the peak twice during rush hours, and had a sharp decline from 8-10 a.m. followed by a significant fluctuation around lunch time. On the other hand, the trend of the casual group had slightly changed within a whole day and peaked in the afternoon. Moreover, both groups had a sharp decrease after 5pm. From then, rental volume for each group tended to be equal until 4am, when the smallest gap between groups occurred. Therefore, we concluded that registered users had a certain routine to perform and this was why they rented bikes, but there were not enough variables to examine the casual trend.

Figure 4 told us that the registered users should definitely be our target group for long-term development. In the meantime, combining two figures helped us grow vision on the casual group and started to think about how to increase market share among casual users.

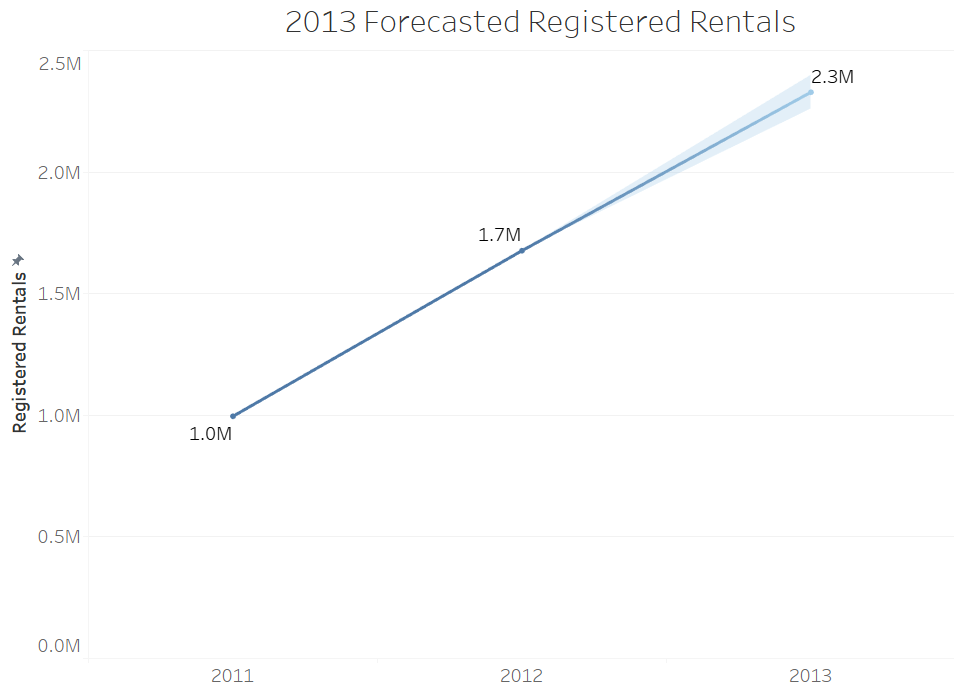
**Research Question 4:** *“How many registered accounts are there over time (growth)?”*

**Step 1: Data Processing**

We used Tableau to respond to our question. The data did not require a great deal of cleansing. Manipulation of the “DTEday” field allowed for easy calculation of registered accounts by month or year.

**Step 2: Choose a Display**

We created a dashboard in Tableau to display our findings and used a series of bar and line charts to show the trend of growth over time in bicycle registration accounts. A line chart illustrates actual month-by-month growth. An area chart shows the forecast of registered accounts by month, and another line chart shows the total for the year 2013.



*Figure 5.* Line Chart of Forecast

**Step 3: Draw Attention**

The dashboard is designed so that the reader can view the story starting with the actual number of registered accounts and ending with the forecast. The line chart of actual registrations uses the colors orange and blue to emphasize the differences in the years 2011 and 2012 as well as similar month-to-month patterns in registered accounts. The use of trend lines in the area chart brings attention to the upward trends.

**Step 4: The Story**

The number of registrations of bicycle rentals increased dramatically by 63.4%, from 996k to 1,724k.  A forecast with a confidence interval of 95% shows an expected increase in registered accounts to 2,329k in 2013 (39%). 

**SUMMARY**

Utilizing Tableau and the Capital Bikeshare system rental data, we were able to identify opportunities for the company to increase rental volume in year two of operation.

Through our analysis, we were able to conclude that temperature is the weather factor that correlates best with rental volume (R^2= 0.4).

In addition, rental volume is higher on non-holidays as compared to holidays and peaks around commute times-particularly among the registered cohort.

Lastly, registered rental volume increased 63% Y/Y and is forecasted to increase 39% next year.

Commuters appear to be the key demographic as rental trends point to: fair weather conditions, peak volumes during rush-hour & on non-holidays.

Besides, 39% Y/Y forecasted growth suggests great potential as Capital Bikeshare enters year 2 in the market, especially if marketing is focused toward this cohort of customers.

**REFERENCES**

http://capitalbikeshare.com/system-data

(2019, April 17). Bike Share in the U.S.:2017. Retrieved 08.12.2020 from <https://nacto.org/bike-share-statistics-2017/#:~:text=In%202017%2C%20the%20number%20of,came%20from%20new%20dockless%20systems.>