



Final Project

Bike Sharing Analysis

Group 7:

Honghui Hu

Nagarjuna Lokaraju

Fatima Nurmakhamadova

Yu Qiu

Xiaolu Shen

Instructor. Jack Bergersen

ALY6070 - 21800 Communication and Visualization for Data Analytics

College of Professional Study, Northeastern University

March 30, 2022

Introduction

This project aims to analyze the demand changes for bike-sharing by different periods and weather conditions as well as the different behavior patterns of the 2 user types - casual users and registered users. We used Tableau to visualize the data needed. Based on the findings, we also provided suggestions to the bike-sharing company from a product and marketing manager perspective.

Questions

From exploratory, we found that bike-sharing users mainly have two types: casual and registered users. So, to offer more detailed suggestions for the product manager, here we raised questions from 2 aspects that would help us understand the user behaviors of bike sharing:

- 1. What are the time patterns of bike-sharing? Any difference between the 2 user types?**
- 2. How do the weather conditions influence the number of bike rentals regarding specific user types?**

At the end of this report, we drew conclusions and recommendations based on the analysis and findings.

Dataset

The data set is about the bike rental information collected mainly from Washington D.C in 2011 and 2012 (System Data | Capital Bikeshare, 2016).

The variables we used are <Date>, <Year>, <Month>, <Hour>, <Holiday>, <Weekday>, <Weathersit>, <FeelingTemperature>, <Casual>, <Registered>, <Count>, <Total>.

Method

In this report, we mainly used bar charts to compare and show the difference of values among different categories and used line charts to show trend and perform time-series analysis. Moreover, we have used scatter plot to analyze the correlations, and heat map to know the days and hours that get the most demand.

Visualizations & Analysis

Part 1: The time patterns of the 2 bike rental user types

1. The analysis of user frequency distribution by monthly rentals

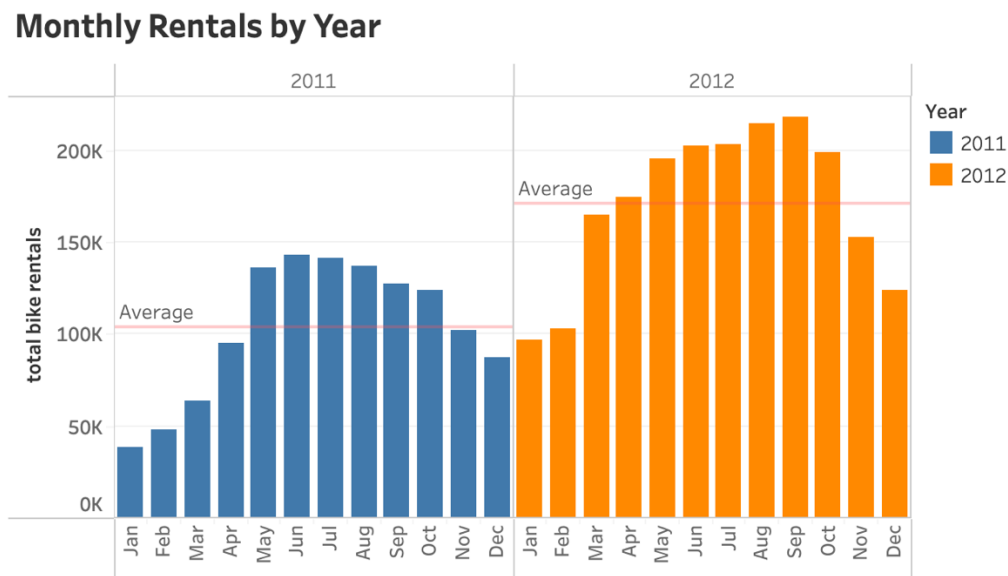


Figure 1: Bar Chart of the Total Monthly Rental by Year

Findings:

The overall patterns of the bike rentals in each month in 2011 and 2012 are similar (Figure 1). The smallest number of bikes rented were in January and February. May to October are the months with total bike rentals higher than the annual average in both 2011 and 2012. In both these 2 years, the number of rentals started to have a relatively big decrease from October.

However, there were also several changes in 2012. The number of rentals in April of 2012 was also above the annual average number, while the number of rentals in April of 2011 was below the annual average. The month with the highest total number of rentals in 2011 is June, while in 2012, the peak month is September.

Moreover, there is an increase in bike rentals in 2012 compared with the 2011 average. Also, the total rentals in each month of 2012 have an increase compared with the corresponding same month in 2011.

Reason to use bar chart with reference line:

Bar charts could quickly show the difference in values for related data sets side by side. In this case, we compared the total rentals in each month and by using a bar chart, we could effectively

observe the trends in a year. And by adding an average reference line, it is easy to show in which months the total rentals are very high, and we could also compare the 2 reference lines in 2 years to compare the general trend.

2. The analysis of bike-sharing user types by days of the week

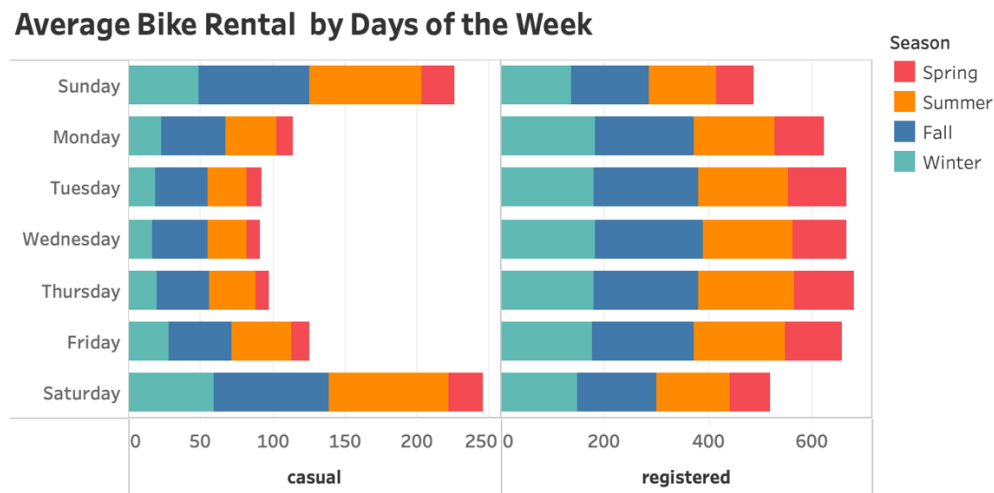


Figure 2: Bar Chart of the Average Bike Rental by Days of the Week

Findings:

The casual users' distribution looks like a C symbol, showing that casual users tend to rent mostly on weekends (Figure 2). Saturday has the most casual users which are in total around 240 users. Whereas the registered users' rent distribution seems like a 3 symbol. They tend to rent more on weekdays than on weekends with not much difference. Wednesday has the most registered users, which are around 690 users in total. In both types, Monday and Friday have almost an equal number of users, followed by Tuesday and Thursday, and lastly Wednesday.

As for the season, we can notice that there are more bikes rented in Summer, Fall, and Winter despite the day of the week. Spring seems to be the least popular season for bike-sharing. However, there are more registered users in spring than casual users.

Reason to use bar chart:

The graph below shows the users' bike-sharing by 7 Days of the Week in both years. The graph also consists of a dual combination chart for two types of users, the chart on the left is for casual users, while the chart on the right is for registered users. We chose the bar graph because it helps to compare each user type of bike-sharing on a different day of the week which is categorical data. The direction of the bar is vertical because this way it is easier to compare two types of users. Moreover, coloring the bars by season helps to compare the difference in users' size in different seasons. As shown in the legend on the right, Red is for Spring, Orange is for

Summer, Blue is for Fall, and Green is for Winter. The colors differentiate from each other thus will not be merged. The x-axis represents the scale of the average users' number. Whereas the y-axis represents the full names of 7 days in a week, starting from Sunday.

3. The analysis of bike-sharing user types by hours

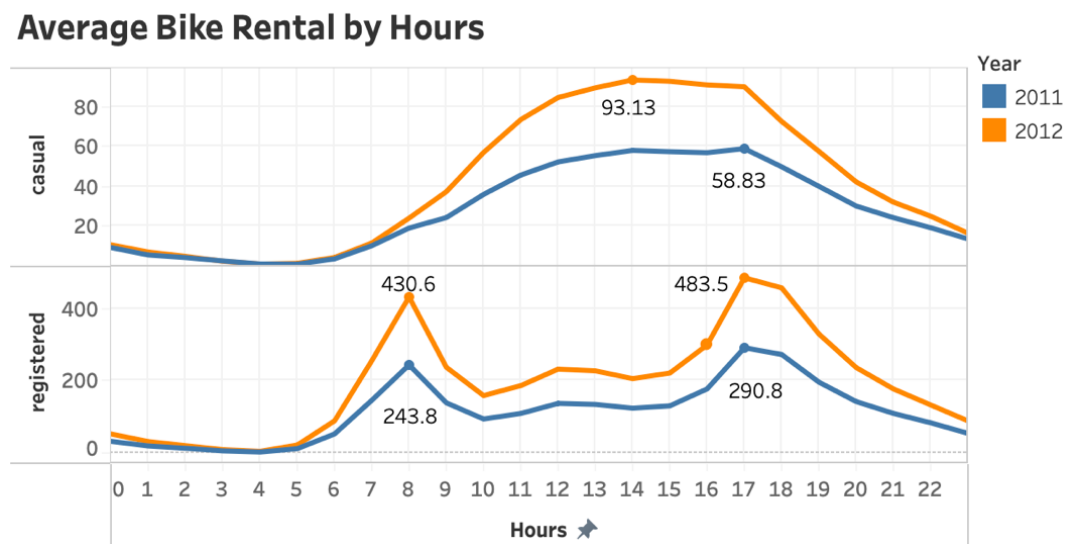


Figure 3: Line Chart of the Average Bike Rental by Hours

Findings:

As we know, there are more users in 2012 than in 2011. But both types of users have almost the same behavioral pattern of bike-sharing by hours in different years (Figure 3). We can notice that the period from midnight to early morning is still unpopular for both users to rent bikes. The same is with the late afternoon, the demand seems to fall after around 6 pm. However, casual users tend to rent bikes mostly during the afternoon only, approximately from 11 am to 6 pm. It has a unimodal distribution with one peak which was a maximum average of **58.83** users in 2011, and **93.13** users in 2012 which is almost twice more. As we know from the previous analysis, they used to rent more during the weekends. Thus, we can assume that overall, casual users tend to rent bikes during weekends afternoon hours.

Whereas, the registered users' distribution is bimodal, having two peaks: in the morning from 6 am to 9 am, and in the evening, from 4 pm to 7 pm. The morning peak shows that in 2011 there were **243.8** users, while in 2012 the number increased to **430.6** users. The evening peak shows that in 2011 there were **209.8** users, and in 2012 it became **483.5** users. The number of users increased almost twice; the evening has slightly more users than the morning. As we know from the previous analysis, they used to rent more during the weekdays. Thus, we can assume that overall, registered users tend to rent slightly more bikes on weekdays morning and evening hours.

Probably, the demand of both user types differs because of their need. We can assume that most of the registered users rent bikes to ride to work/school in the morning and come back in the evening. Whereas casual users rent bikes during the afternoon to rest.

Reason to use line chart:

The graph below demonstrates the users' bike-sharing by hours in 2011 and 2012. The graph consists of a dual combination chart for two types of users, the chart above is for casual users, while the chart below is for registered users. We chose the line graph because it can demonstrate possible changes and trends over the period. Moreover, it allows comparing the users' distribution within both years. Therefore, we colored them using pastel colors, making perception easier. As shown in the legend on the right, the blue line represents the year 2011, and the orange line represents the year 2012. We also labeled the peaks of the distribution that will help to compare changes in users' amounts each year. The x-axis represents 24 hours, starting from 0 which stands for midnight (00:00), lasting until 11 pm. Whereas the y-axis represents the scale of average users' number because comparing the average values is easier, as it results in a relatively small number. While comparing the sum of users might increase the cognitive load of the viewers as it is hard to quickly perceive and understand the difference.

4. The analysis of bike-sharing user types by days of the week vs hours

Bike-Sharing by Days of the Week vs Hours

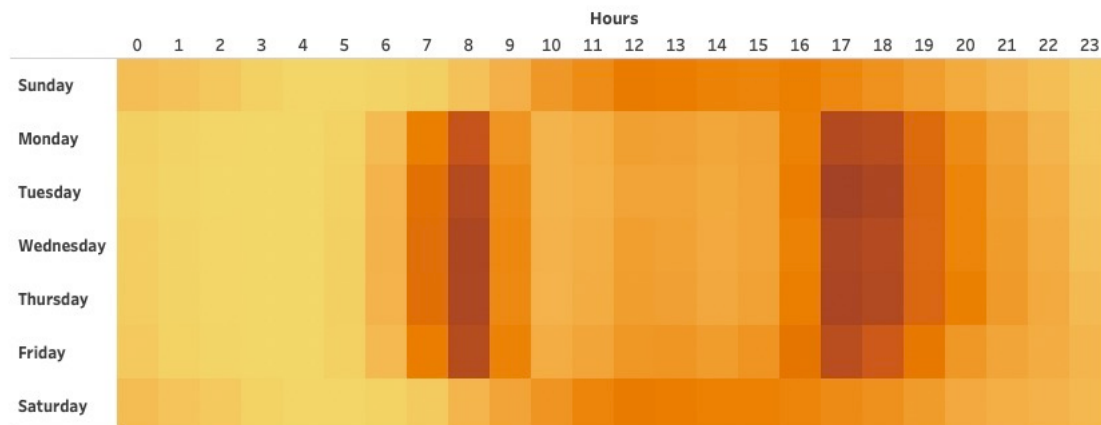


Figure 4: Heat map of Bike-sharing by Days of the Week vs Hours

Findings:

While looking at the heatmap (Figure 4), the first thing we notice is the dark-colored areas on weekdays morning (6 am - 9 pm) and evening (4 pm - 8 pm) hours. Then we see a bit less dark area on weekends afternoon (10 am - 6 pm) hours. The first probably represents the most demanded period of the registered users, while the second represents the most demanded period of the casual users. Thus, we can notice that there are more registered users than casuals. Moreover, the heatmap helps to understand that the period from midnight to 5 am has a very low number of users. Overall, the heatmap helps to view the big picture from the analysis we

did above, from a wider perspective. Thus, it gives a general idea of the bike-sharing demand by days of the week and hours.

Reason to use heat map:

The graph below shows both users' bike-sharing by 7 Days of the Week for 24 hours in both years. We chose the heat map because it represents the information by color, the more intense the color is, the more density is. Thus, it facilitates the viewers' perception of complex information. The x-axis represents 24 hours, starting from 0 which stands for midnight (00:00), lasting until 11 pm. The x-axis is located on the top of the graph because it is easier to understand the correlation between the x-axis and y-axis as the heat map might seem complicated at first. The y-axis represents the 7 days in a week, starting from Sunday. As for the color, it shows the total users' distribution, if the number of users is low - the color is light orange, while if the number increases, the color is getting darker.

5. The analysis of user types during holidays and regular days

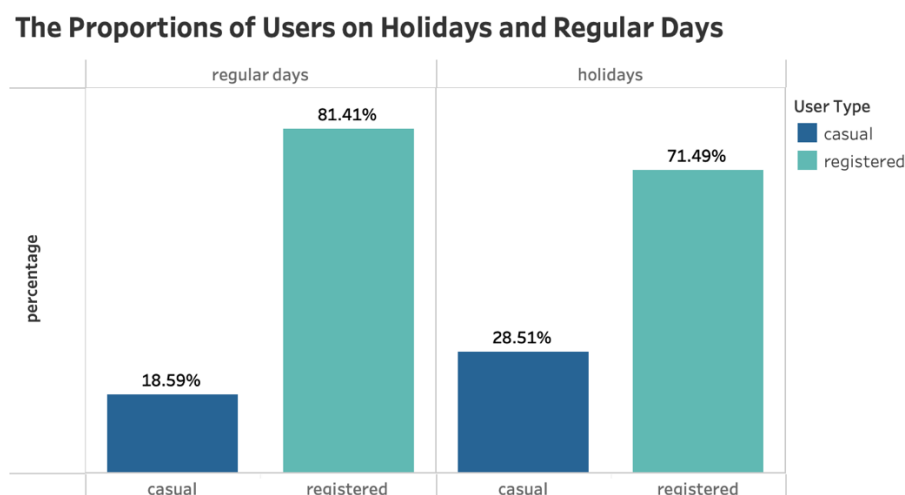


Figure 5: Bar Chart of the Proportions of Users on Holidays and Regular Days

Findings:

As we observe in Figure 5, the proportion of casual users on holidays (28.51%) is higher than the proportion of casual users on regular days (18.59%), approximately 10% more. On the other hand, the proportion of registered users shows a reducing trend from regular days to holidays (from 81.41% to 71.49%). From the regular days to holidays, the increasement in casual user proportion can be explained that more tourists are in D.C. during holidays, which has many famous scenic spots in different areas of the city. As a result, using sharing bikes is a great option for tourists to travel across the city, causing more numbers of casual users than on regular days. On the other hand, most registered sharing-bike users are local, living in Washington D.C., and they are most likely to use sharing bikes as a commuting method. So, when it comes to holidays, these registered users are less likely to use sharing bikes: they may

drive out with their families, go hiking, or just stay at homes home. This could explain why the proportion of registered users is less during the holidays than on regular days.

Reason to use bar chart:

To check whether the holiday factor would influence the user of bike-sharing, here we choose the bar chart to show the proportions of casual users and registered users during regular days and holidays. We chose this visualization form because the bar chart can clearly show the possible differences existing in the different groups and dimensions.

Part 2: How do the weather conditions influence the number of bike rentals regarding specific user types

1. The analysis of user frequency regarding the feeling temperature

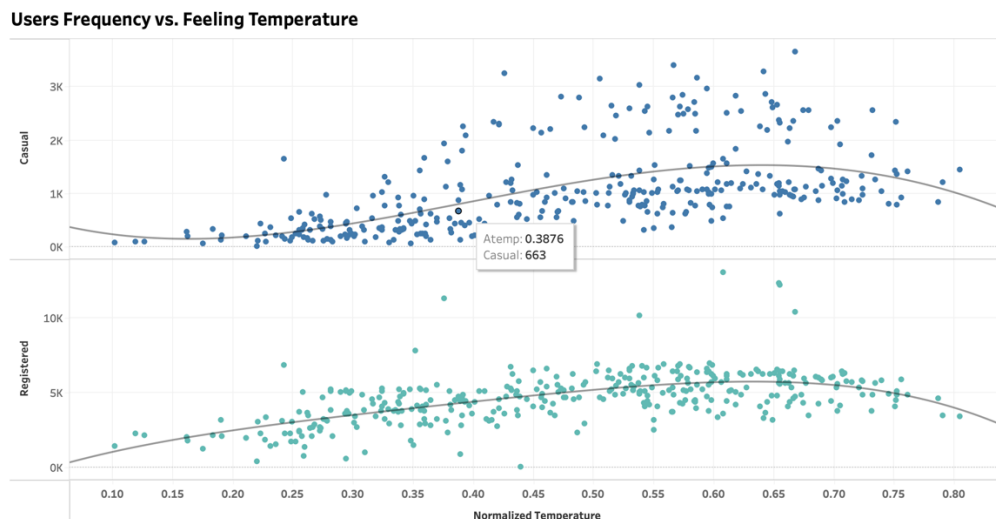


Figure 6: Line Chart with Polynomial Model of Users Frequency vs. Feeling Temperature

Findings:

In the line chart of user frequency with feeling temperatures (Figure 6), a polynomial model is performed to fit the data, almost 60% of data can be explained by this model with the degree of four. For the registered user, there is a distinct correlation between the rental frequency and feeling temperature, while for the casual users, there is no obvious pattern. Besides, in the interval of normalized feeling temperature between 0.30 and 0.70, many data points are widespread with high rental frequency.

Reason to use line chart with polynomial model:

The line chart allows the audience to catch the user frequency trend over normalized feeling temperature. Polynomial provides a better approximation of the relationship between the feeling temperature and the number of users than simple linear regression.

2. The analysis of user frequency regarding weather conditions

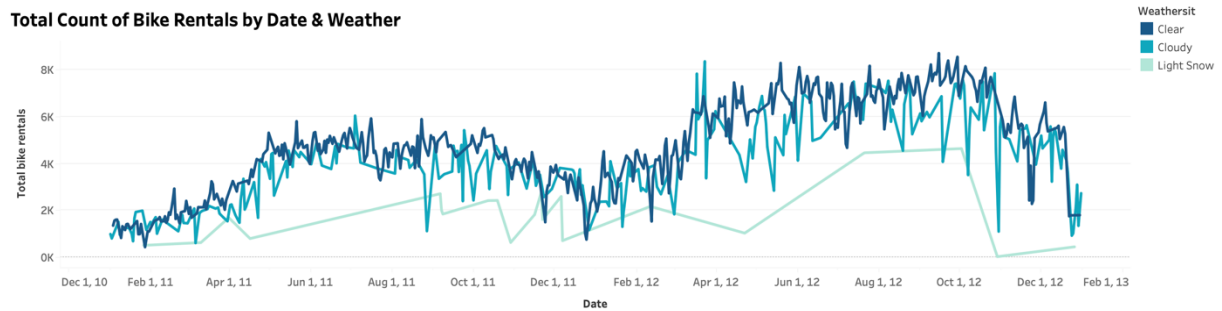


Figure 7: Line Chart of Total Count of Bike Rentals by Date & Weather

Findings:

There is a roughly increased trend of total bike rentals from 2011 to 2022 (Figure 7). The seasonality of clear days and cloudy days are very similar with a peak from June to September, and a valley in December, January, and February. However, the total rentals on light snowy days do not have an obvious seasonality or trend, so it is quite hard to predict the number of rentals on light snowy days. Besides, on light snowy days, there were fewer bike rentals compared with clear days and cloudy days.

Reason to use line chart:

Line charts could effectively show the trend over time. The above analysis is to see the changes in total bike rentals over each day from 2011 to 2012. A line chart could easily show the bike rentals trend. By further differentiating the lines by different weather, we could compare the 3 lines to tell the difference in trend and further tell how the weather will influence the total bike rentals.

Recommendations

Peak season

An increase in the availability of bikes during the good weather conditions between 0.25 and 0.65, around the peak hours of 7 am and 5 pm, and in the peak season will help in maximizing profit and this is expected to rise year on year based on the trend.

Non-peak season

In the non-peak season i.e., during the spring and winters, there are not many usage statistics which might be due to harsh cold weather. Here to attract and improve bike-sharing there can be marketing and promotion aspects, decrease in prices, etc. With harsh weather and a decrease in demand, provision for bike-sharing can be reduced which will be reducing the cost as well as maintenance.

General usage

Regularly, for all the users' schemes like a 3-day trial, etc. can be launched which will not only bring new customers but also maintain the existing ones. Based on the trends precise predictions can help the business grow and be accurate in the profit and cost-effectiveness.

References

Knafllic, C. N. (2015). *Storytelling with Data*. John Wiley & Sons.

System Data | Capital Bikeshare. (n.d.). Capital Bikeshare. Retrieved March 30, 2022, from <https://ride.capitalbikeshare.com/system-data>