

CAPITAL BIKESHARE: USE PATTERNS

1. Problem

Understand patterns of bikeshare use in Washington, D.C.

To better understand use patterns in a bikeshare fleet, the goal is to define high and low demand times and describe frequency and duration of rentals for casual and registered users.

2. Client

The client is Capital Bikeshare, and results from this analysis will allow a clearer understanding of how users interact with the bikeshare program. This information could be helpful for growing the business, for example, targeting advertising dollars on days or seasons with a high proportion of unregistered users might be the most efficient way to increase registration.

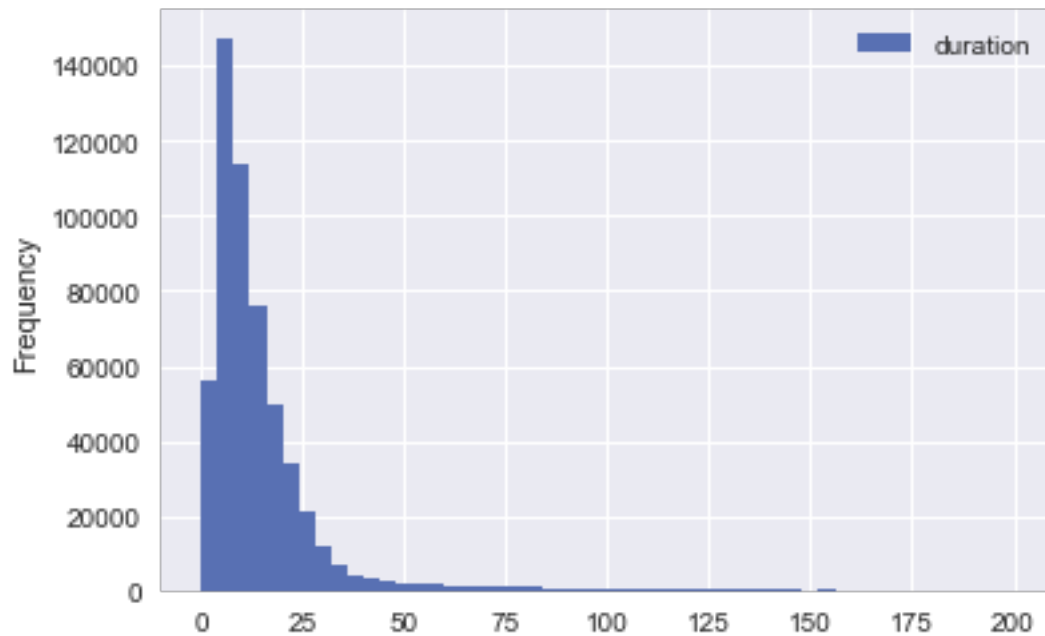
3. Data

Capital Bikeshare posts quarterly data reports of bike trip times, start and end locations, and type of user (registered or casual). Each trip is on one line of data, and includes start time, end time, duration, start station ID and address, end station ID and address, and user type (registered or casual). These data are readily and publicly available at <https://www.capitalbikeshare.com/system-data>.

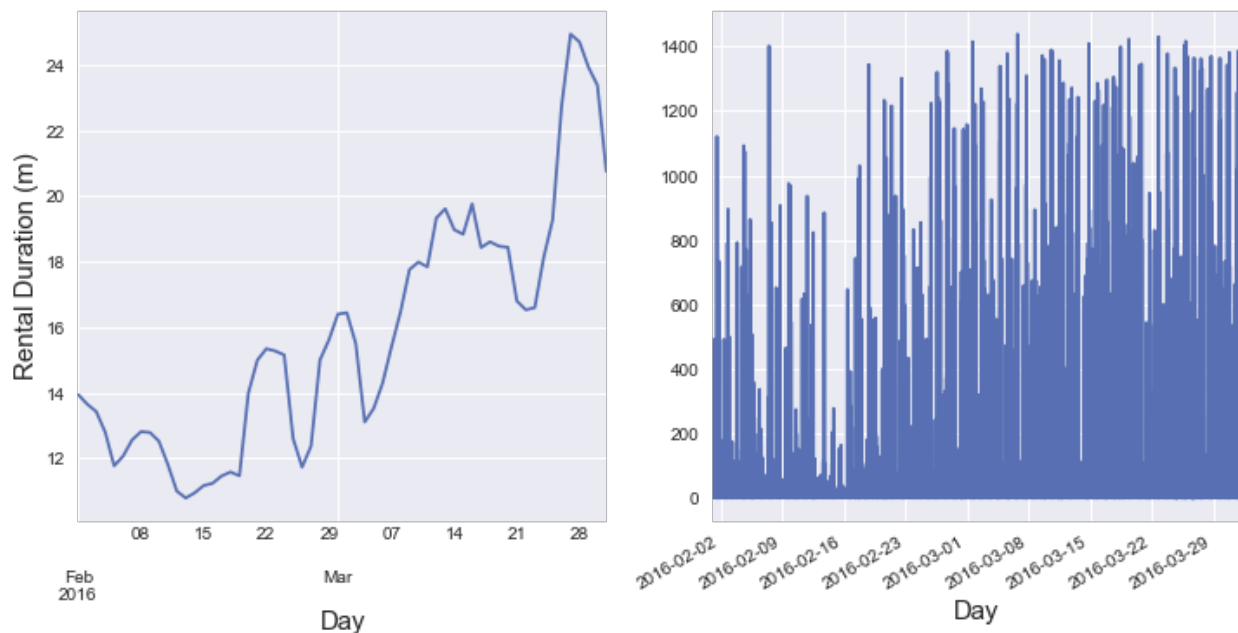
These data were provided in a very manageable format. When importing, I created a time index based on the start time of each rental, and converted the duration column from milliseconds to minutes for improved interpretability. To simplify the exploration process, I have focused on one data file which contains data from the first quarter of 2016 (Jan-March). There are a little over 550,000 rows in this subset.

4. Initial Findings

One interesting characteristic of the test data is that the duration of rentals has a non-normal distribution. Although most rentals range from approximately 5-25 minutes, there is a long tail of much longer durations, even up to 1400 minutes (~24 hours).

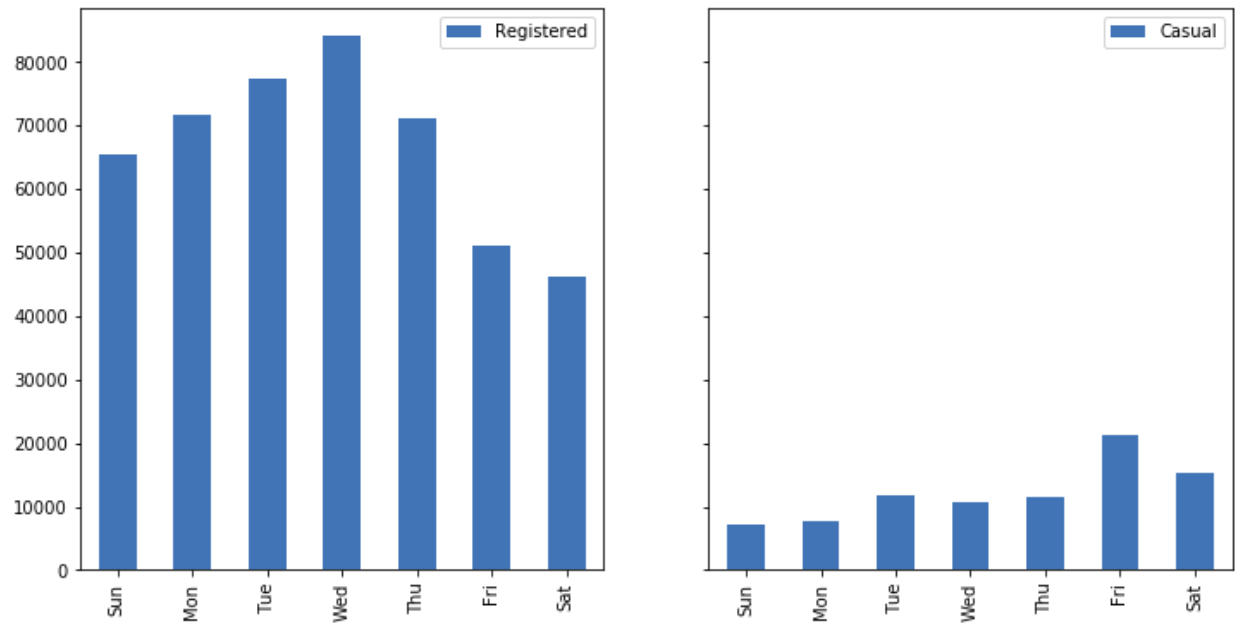


Rental duration varies quite a bit across the quarter, but a rolling mean shows a general trend for increasing rental durations.



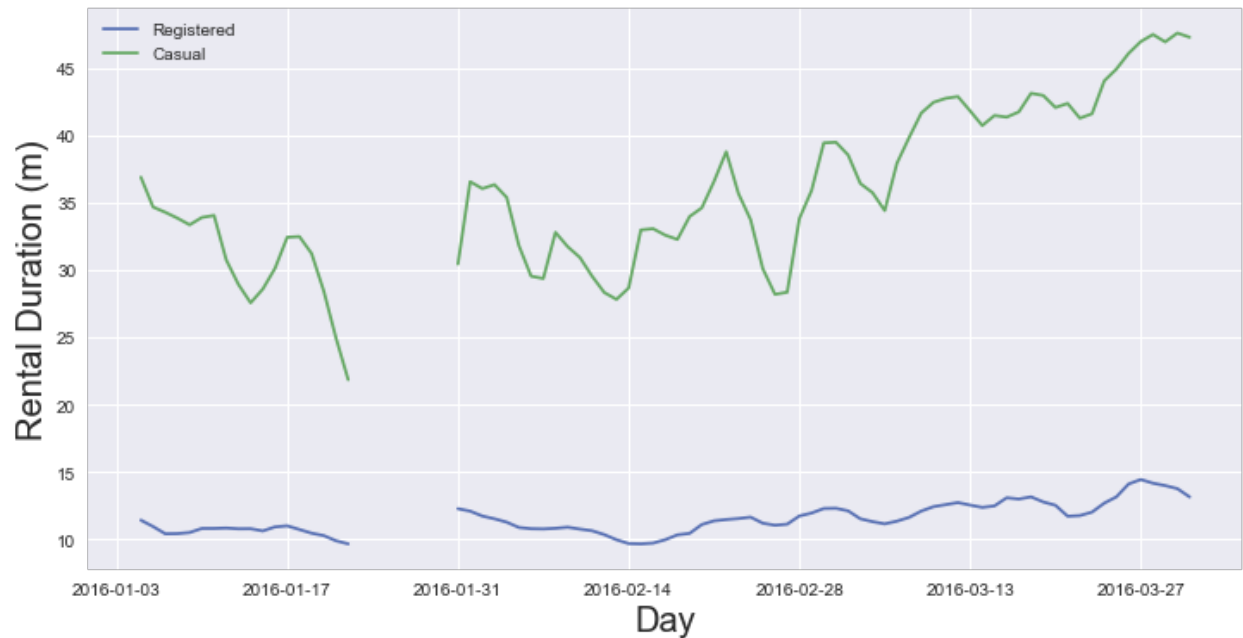
However, it seems that whether a user is registered can have a big influence on this usage behavior. For example, whether rentals occur on the weekend or on weekdays:

Frequency of rentals by user type

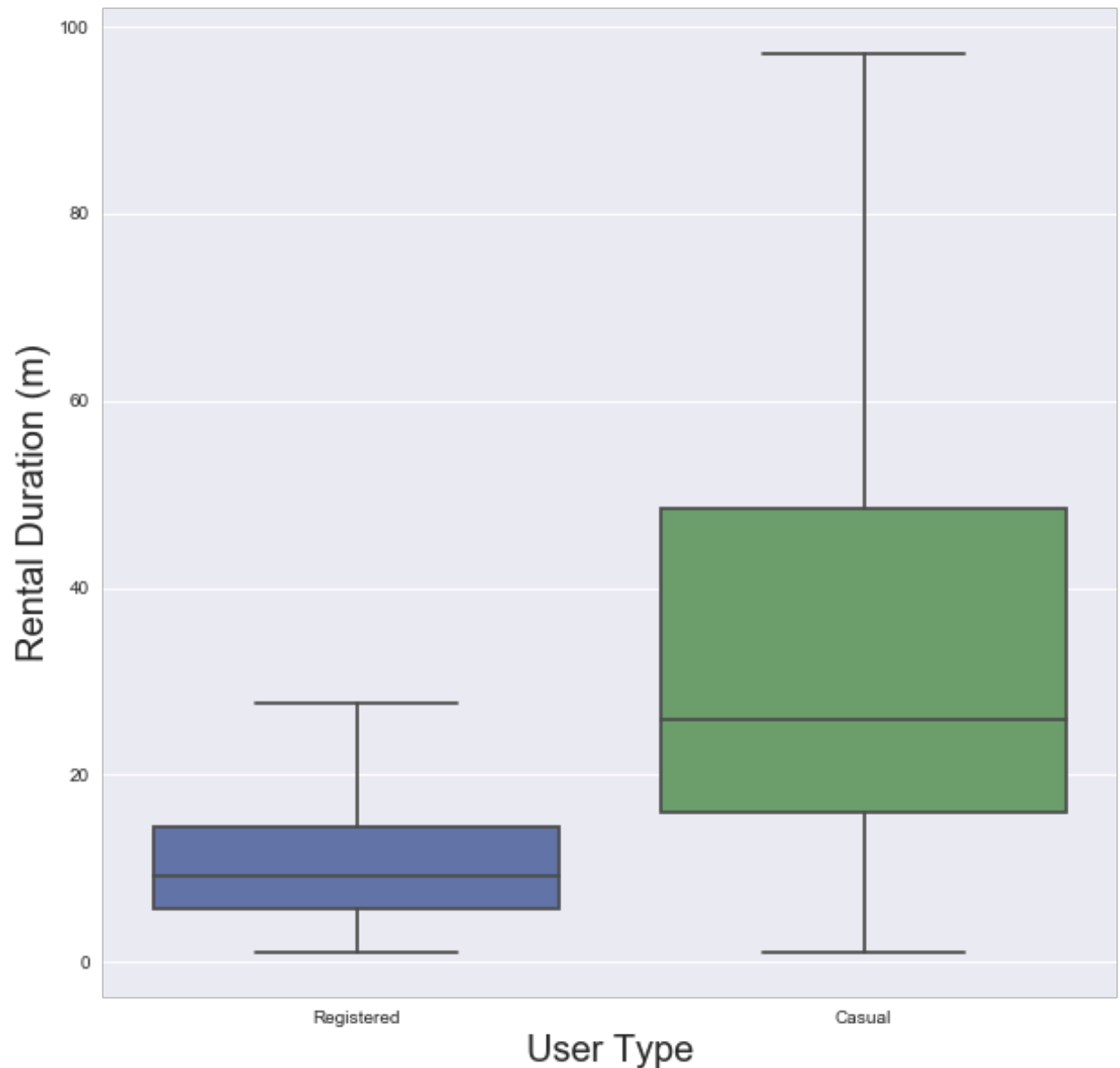


Or how long rentals last (duration):

Duration of rentals by user type



Another way to show this difference in rental duration is using a box plot, shown here:



A simple t-test tells us the means of these two populations are statistically different, but this is not surprising given the high sample size.

My initial hypotheses for these differences based on user type is that users of each category have predictable characteristics. For example, local residents may tend to register for the program, bike shorter distances, and rent for shorter times. We could speculate that trip purpose might be personal errands, lunch on a workday, or transport to work meetings. Non-residents (i.e. tourists), on the other hand, may rent a bike to facilitate longer distance sight-seeing agendas, or even circular trips (no destination). Perhaps even unregistered locals use the bikeshare program only on weekends for purposes similar to those postulated for tourists.

Next steps:

To move this story forward, we can ask some questions, such as:

1. Can we predict user type by the bike ride duration?
2. Can we predict user type by the day of the week?
3. Compare this winter/spring data to summer or fall data - do the answers to the above questions change?