Citi Bike: Mini Exploratory Visual Analysis

One of the main justifications for bringing Bike Share services to New York City was that such services provide an alternative transportation option for the local workforce and supplementary connections for commuters using the various forms of mass transit in and out of the city, and Manhattan in particular.

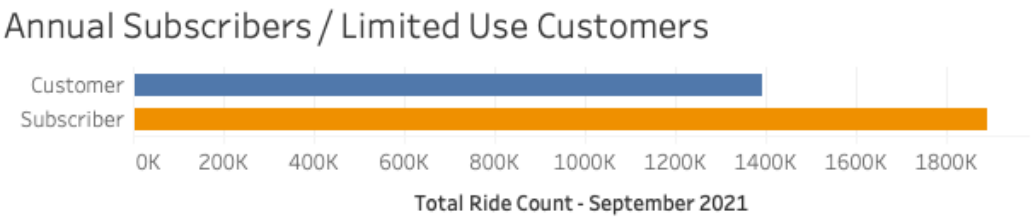
The Citi Bike program has now been in operation for over eight years and by all measures can be considered established and mature. Its operators have amassed a vast store of ride data, collected on a monthly basis, detailing each and every use of the system during that period. Drawing on this resource, one obvious question to be asked of the data is whether or not the system is being used as its advocates imagined it would; that is, was the local workforce making use of Citi Bike in daily commuting? What is the proportion of commuters to casual riders and tourists among all users?

Since there is nothing in the data regarding the purpose of each individual ride, it is necessary to use what data is present to make some educated guesses and use these assumptions to differentiate between the two types of user. It is here where some exploratory visual analysis is useful.

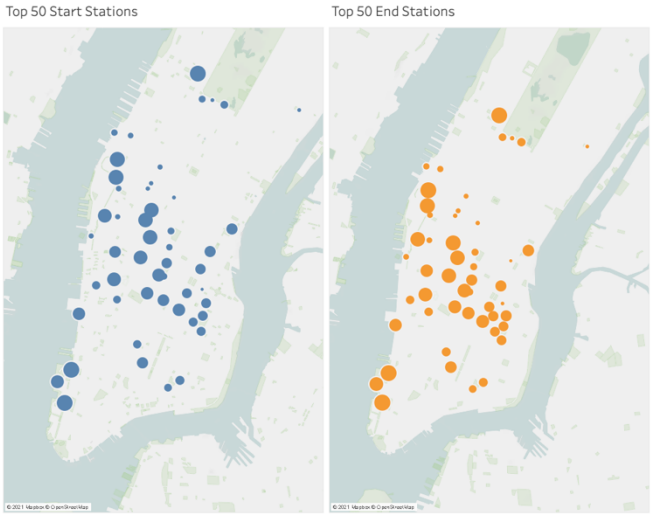
Given the enormous amount of data collected (monthly figures can measure upwards of three million rides), I chose to use one month's data for my initial visual analysis, September 2021, the most recent month for which both ride data and a monthly Operating Report had been published. This dataset is composed of 3,280,221 records including: Ride ID, Rideable Type (classic or e-bike), Start date/time, End date/time, Start station name/ID/latitude and longitude, End station name/ID/latitude and longitude, and User type. Citi Bikes provides its data in CSV format which was imported directly into Tableau (the data load proved too heavy for Excel conversion).

Using the September 2021 dataset, I considered three options that might reveal usage patterns and numbers for both types of user. The first, and most obvious was the number of rides by annual

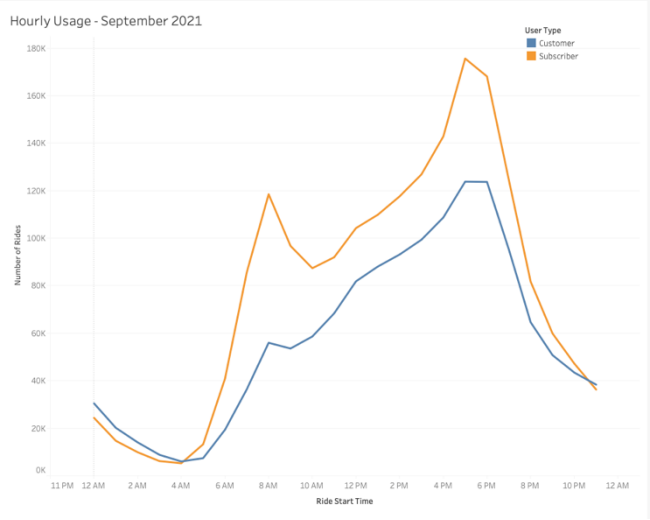
subscribers versus 24 hour or 3-day pass users, the assumption here is that daily commuters would be more likely to take advantage of the cost savings that come with a yearly subscription, while tourists and other casual users would opt for the short-term day pass.



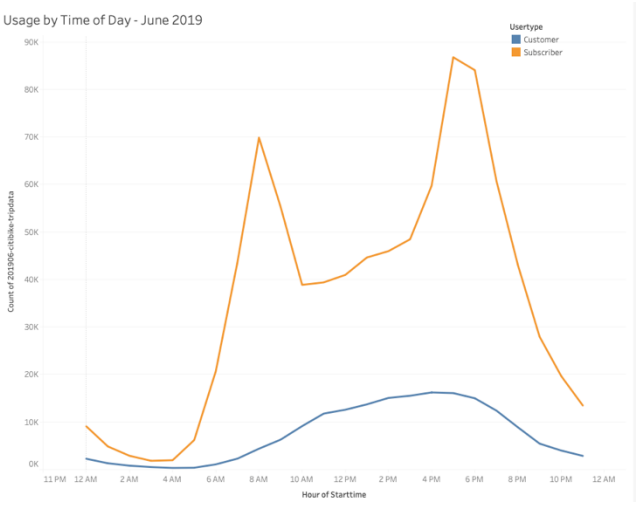
While this chart did show more use by annual subscribers, it was a bit too general for my purposes. The second option was to map the 50 most used start and end stations, with the idea of revealing travel patterns toward either commercial or recreational areas of the city. This proved even less successful than Option 1 as the visualization produced no usable patterns for differentiation.



The third option was to chart usage based on time of day and revealed a clear pattern indicating that the Citi Bike program is indeed being used by commuters as usage tended to spike during both morning and evening commute times (8am and 6pm) and dropped off significantly at other times.



An unexpected revelation was the fact that this pattern appeared for both subscribers and day pass customers, though not as pronounced in the latter. This was in marked contrast with a similar visualization using data from two years previous, in which the rush hour spike was exclusive to the subscribers, while day pass use slowly grew then declined throughout the day.



This difference opens another avenue of inquiry for visual analysis, that is, how usage patterns have evolved over the years, which may be worth exploring as this project progresses.