Google Data Analytics Capstone: Cyclistic Bike-Share Analysis



Author: Tyler Johnson

Date: 7/15/2023

For this Google Data Analytics Capstone, I was tasked to imagine I was a marketing analyst for the fictional company “Cyclistic,” a bike-share company located in Chicago.

I was taught to break the analysis process down to 6 phases:

**Ask, Prepare, Process, Analyze, Share, and Act**

I will explain my journey through this Capstone using those 6 stages to demonstrate my process.

**Phase 1: Ask**



Cyclistic has over 5,800 bikes and 600 docking stations across the city, and has a variety of different types of bikes available to be more inclusive to their riders. The company offers three pricing plans: single-ride passes, full-day passes, and an annual membership.

Lily Moreno, the director of marketing, believes that the company’s future success relies on converting casual riders to annual memberships, and it is my goal to compare the behaviors between causal riders and members to help design a new marketing strategy to increase membership.

****

Moreno has asked the Cyclistic marketing team to answer these 3 questions:

* **How do annual members and casual riders use Cyclistic bikes differently?**
* **Why would casual riders buy Cyclistic annual memberships?**
* **How can Cyclistic use digital media to influence casual riders to become members?**

**Phase 2: Prepare**

****

Does the data **“ROCC?”**

**Reliable:** Yes, it also is a large sample size

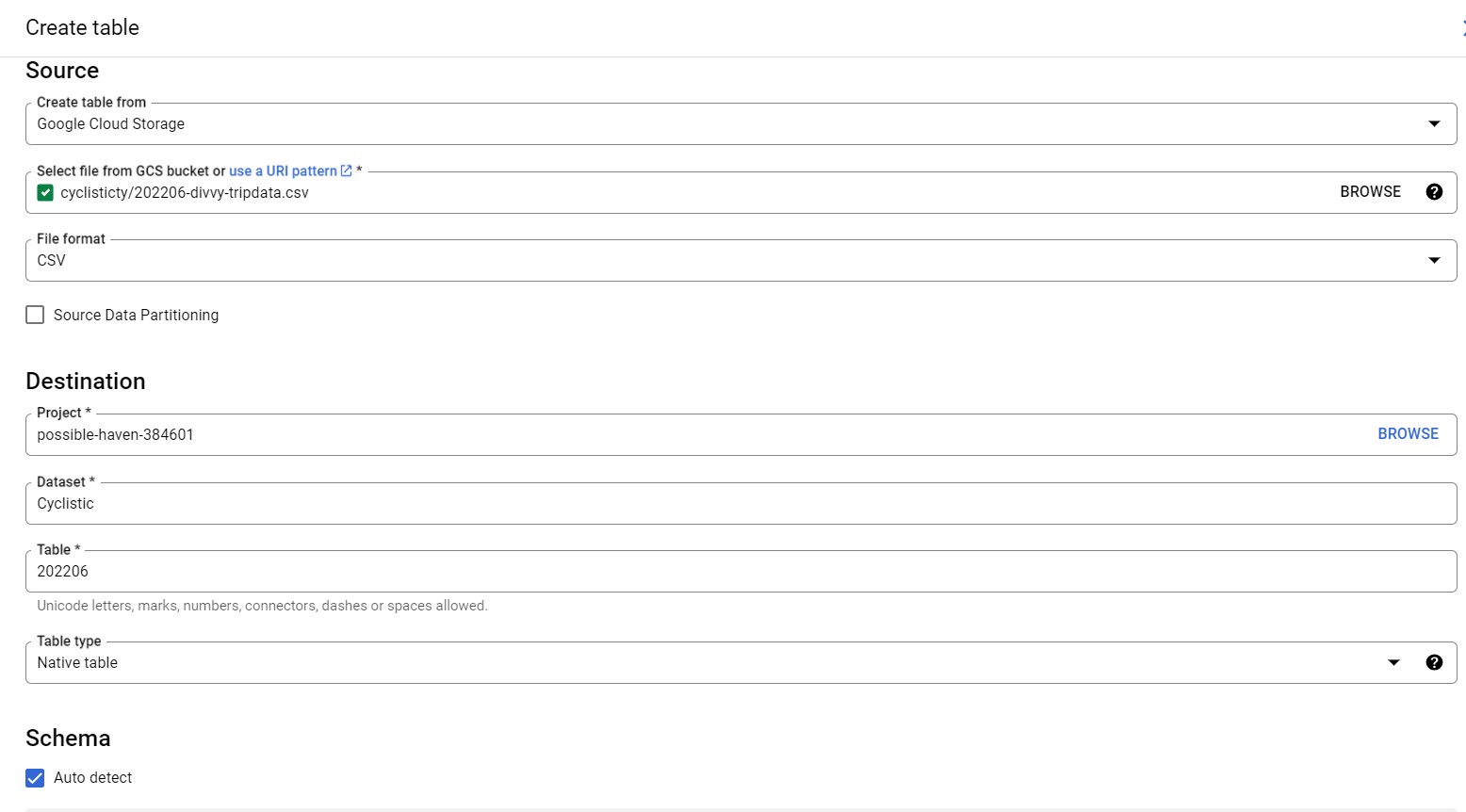
**Original:** Yes, the data is easily to locate, verify and [download](https://divvy-tripdata.s3.amazonaws.com/index.html) online

**Comprehensive:** Yes, it appears unbiased but missing some non-vital information

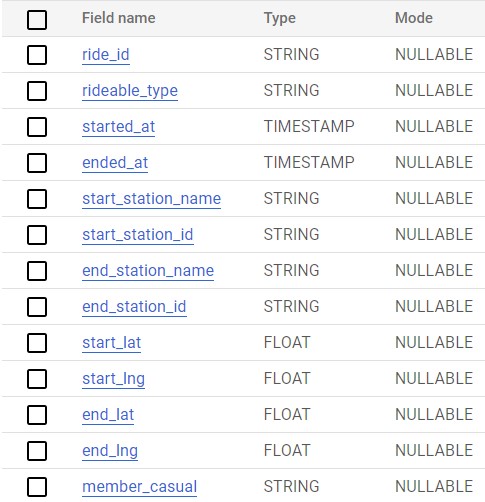
**Current:** Yes, the data is updated each month

**Cited:** Yes, it was made available by Motivate International Inc. under [this license](https://ride.divvybikes.com/data-license-agreement)

I used the most recent 12 files available (June 2022 - May 2023) and downloaded them as .csv files. The file sizes were too large to import locally to BigQuery (100MB limit), so I used Google Cloud Storage Buckets to import the data.

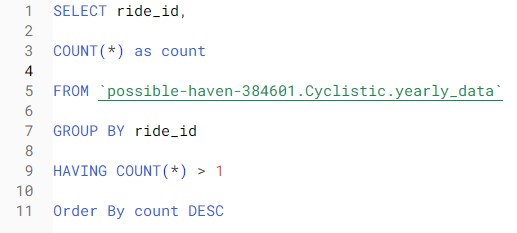
****

In the tables were 13 columns:



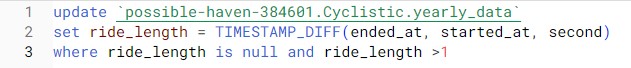
**Phase 3: Process**

After merging the 12 tables into one year long table, the first step I took was to be sure there weren’t any duplicates.

****

There were none, and I was able to ensure that each “ride\_id” was unique.

The first column I created was “ride\_length” to easily see how much time each ride took. I took the difference from the “ended\_at” column and the “started\_at” column, then converted it to seconds.



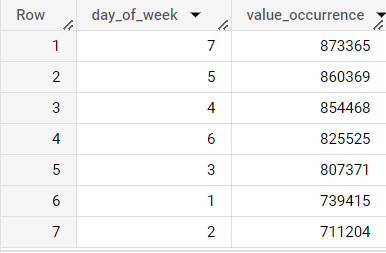
I discovered early that some of the “ended\_at” times were actually before the “started\_at” times, creating a negative ride length, so I removed any row that had a ride length lower than 0 second.

****

I then decided to create another column, “day\_of\_week,” to help analyze how members and casuals rode differently each day of the week.



From there, I decided to get a glimpse at what the most common days were for riders to use Cyclistic.

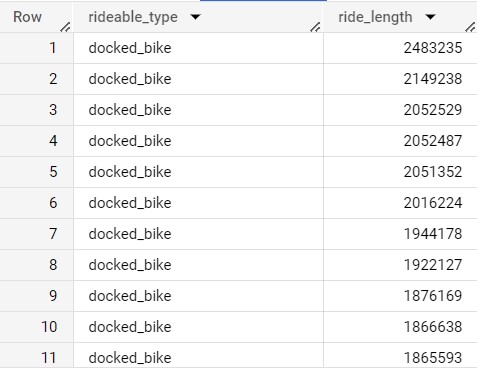
****

Saturday, Thursday and Wednesday were the most popular days in total, with Sunday and Monday quite significantly on the lower end of the spectrum.

I decided to look at ride length per bike type and examine the longest rides of the year, but I noticed something strange in the data regarding docked bikes.

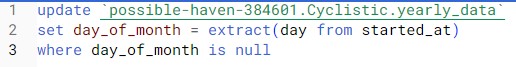


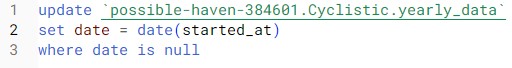
The max ride length came out to 28 days, and it took me 2000 rows before I found my first ride length that wasn’t a docked bike. I also found that they were consistently starting and ending at the same location and only used by casuals.



Since docked bikes gave a lot of weird data that was greatly skewing results, I decided that removing docked bikes from this study, which only accounted for .3% of yearly rides, would help give a more accurate analysis of how riders differ between members and casuals.

I finished up by creating columns for day of the month, the date, and hour for each ride and uploaded the updated table to GCS to prepare analyzing in Tableau.







**Phase 4: Analyze**

****

With the data imported to Tableau, I decided to use visuals to find the answer to these questions:

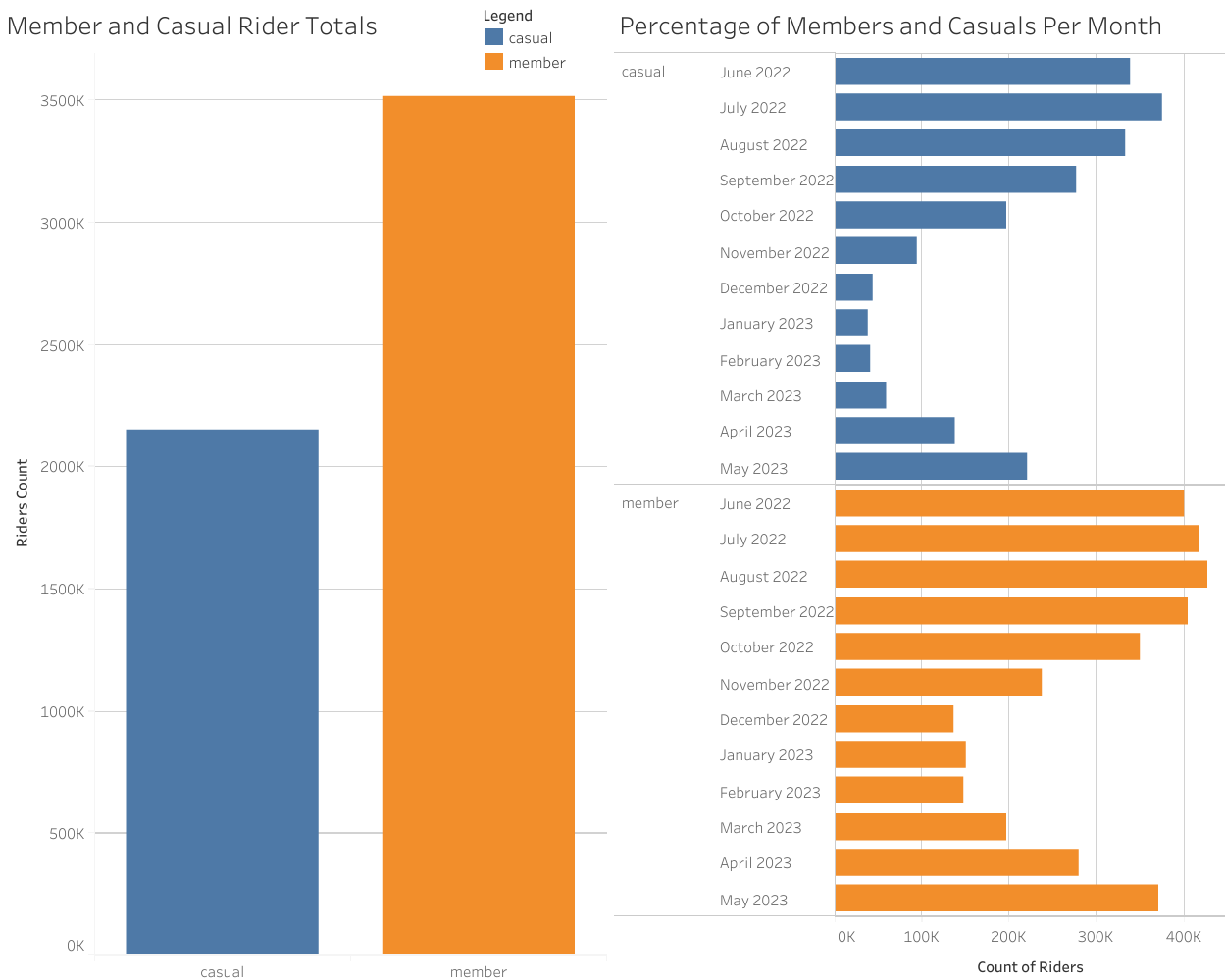
* How many members and casuals rode this last year? Per Month?
* What relations does ride length have with members and casuals?
* How do rides per day, week day and month differ between both?
* How does the season affect both groups?
* Does one group favor one bike type over the other? Does season affect the type of bike used?

**Phase 5: Share**

****

(Interactive Dashboard can be accessed [here](https://public.tableau.com/app/profile/tyler.johnson1873/viz/TylerJohnsonCyclisticCaseStudy/RideLengthandCountPerType))

**Membership and Casual Totals:**

****

In my first page, we can see the totals of members and casuals for the last year, then see it broken down by month.

* There are about 1.3 million more members than casual riders
* Each month has more members riding than casuals
* Both groups peak during the warmer months (May-September)

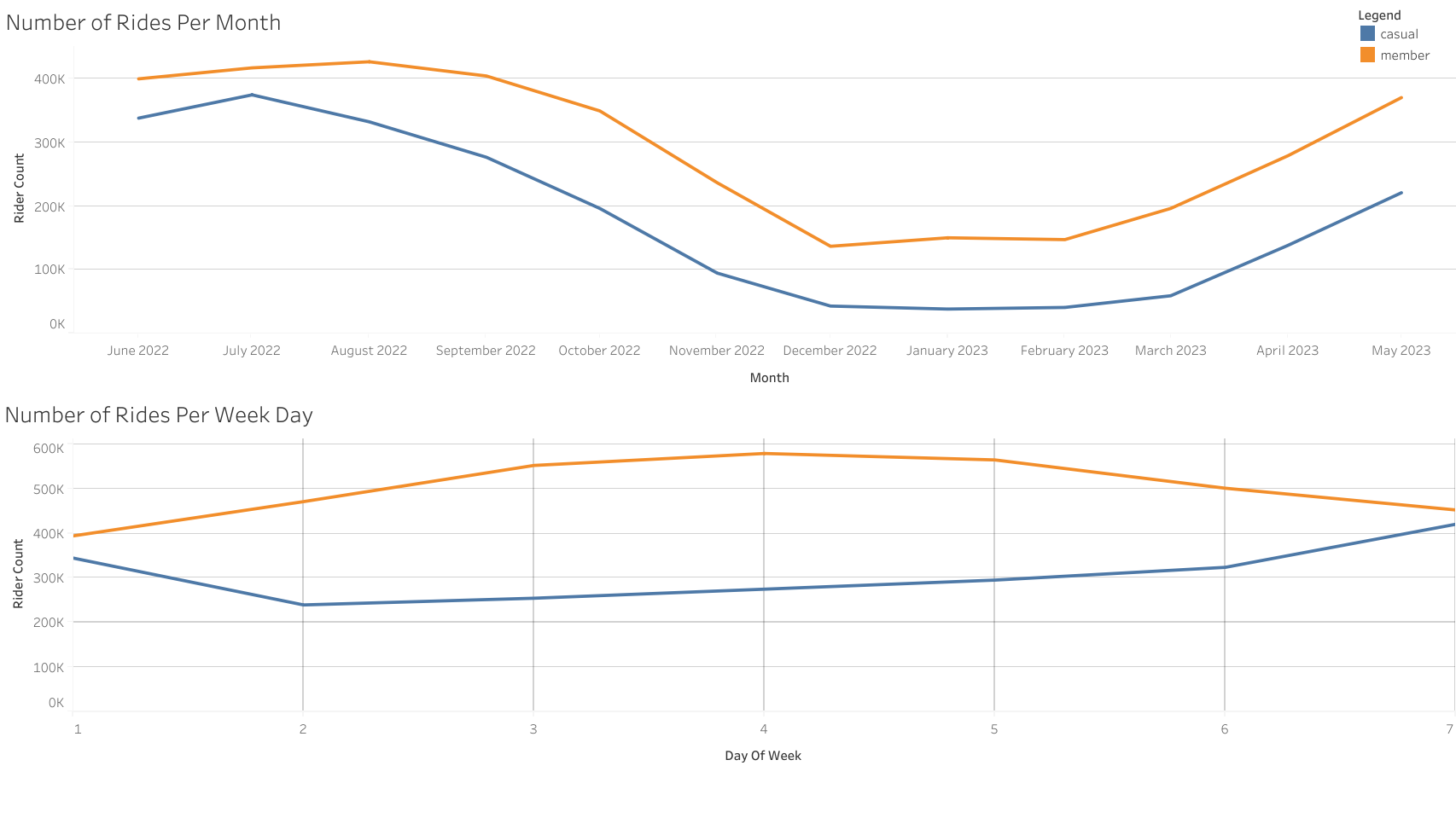
**Ride Length:**

****

My second page is focused on ride length and how it relates to membership and total rider count.

* Ride length for both groups increases on the weekend
* Casual riders regularly have greater ride lengths
* Ride length per hour differs much less for members, implying they ride more for work than leisure compared to casuals
* Ride length and rider count both dip as it gets colder, but the count drops more drastically

**Rides Per Day and Month:**

****

Here we see the rider count by month and by week day broken down by membership status.

* Members and casual riders follow a similar monthly trend on rider count
* July is the peak for casual riders
* Casual riders use Cyclistic the most on weekends
* Members use Cyclistic the most on weekdays, implying many use it mostly for work transportation

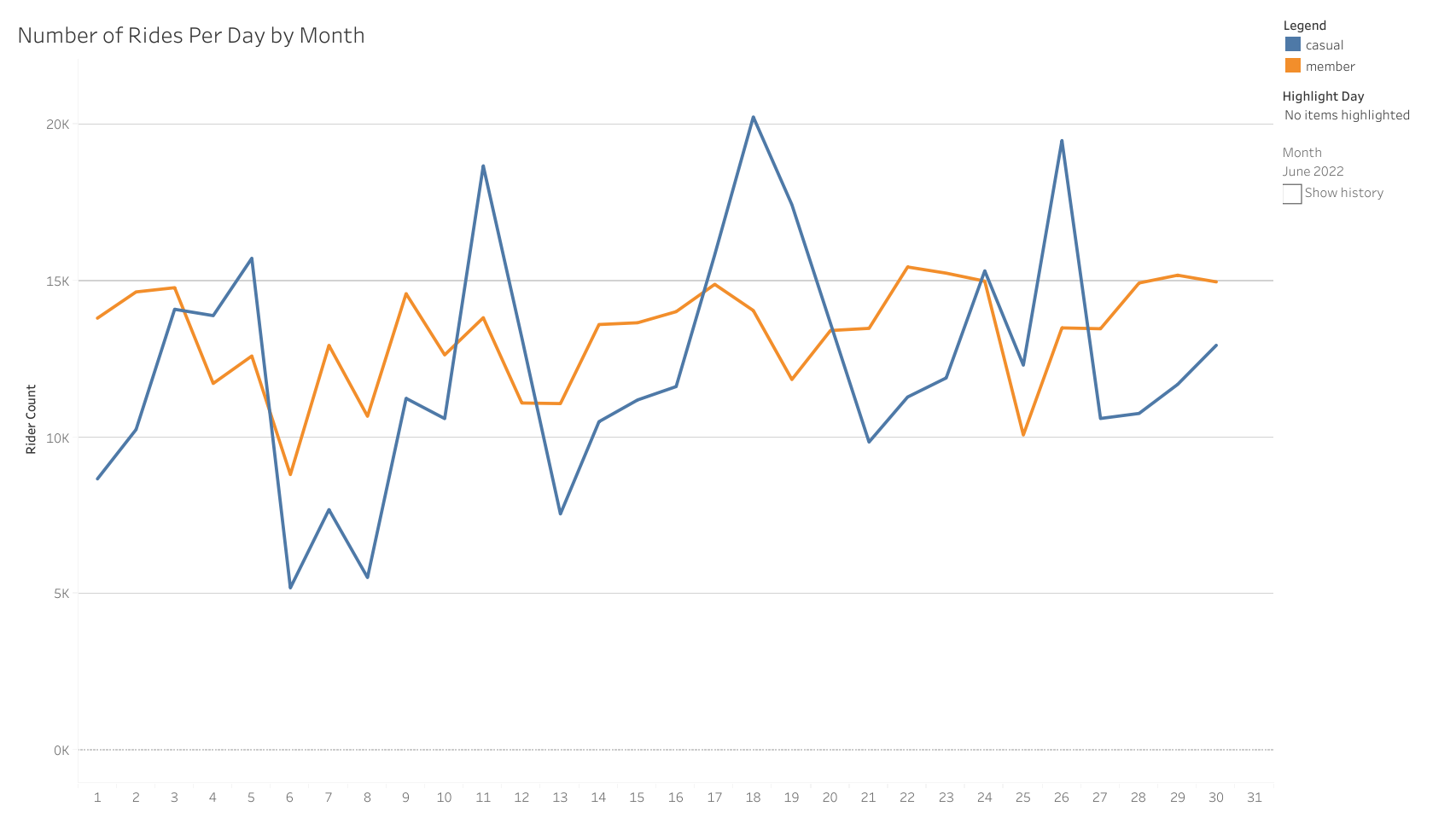
**Rides Per Month and Week:**

****

Here we see a more detailed overview of monthly rider count (broken down by week) and the number of riders per week day.

* Casuals and members share peaks in the weeks of April 9th (Easter weekend) and May 21st.
* Weekends share the closest ratio of casuals and members
* Mondays are the least popular day for casuals
* Saturdays are the most popular for casuals

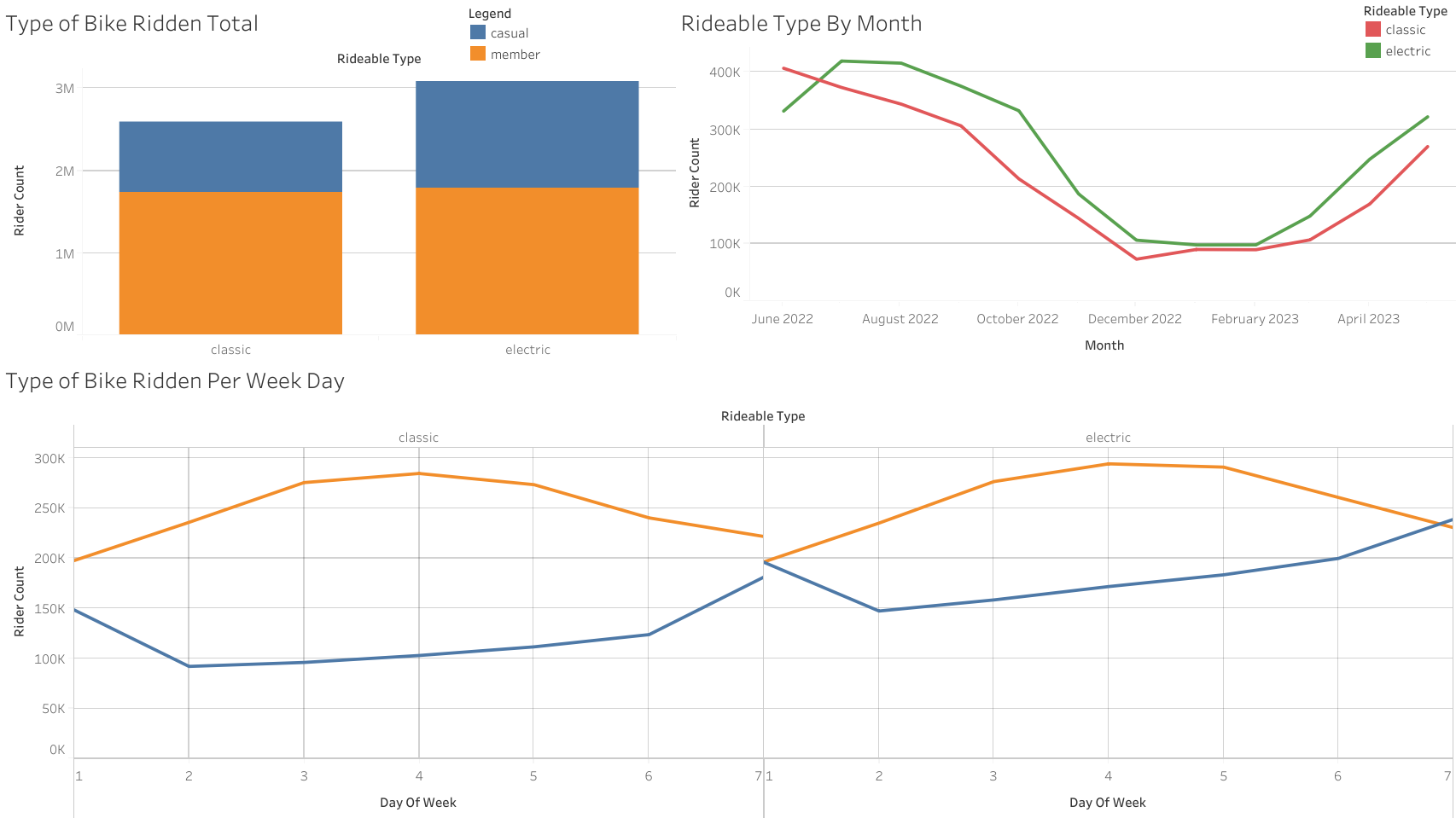
**Rides Per Day by Month:**

****

Here I have broken down the rider count by members and month, which can be scrolled in the dashboard [here](https://public.tableau.com/app/profile/tyler.johnson1873/viz/TylerJohnsonCyclisticCaseStudy/RideLengthandCountPerType)

* Casual rider count is more sporadic, with peaks on weekends
* Member rider counts are more consistent, but share similar dips and climbs with casuals
* Casuals level out when the weather gets colder, dip below 5,000 riders after November
* The week of April 9th was the first time casuals rose above 5,000 daily riders in 2023, peaking on Saturday April 15th with almost 15,000

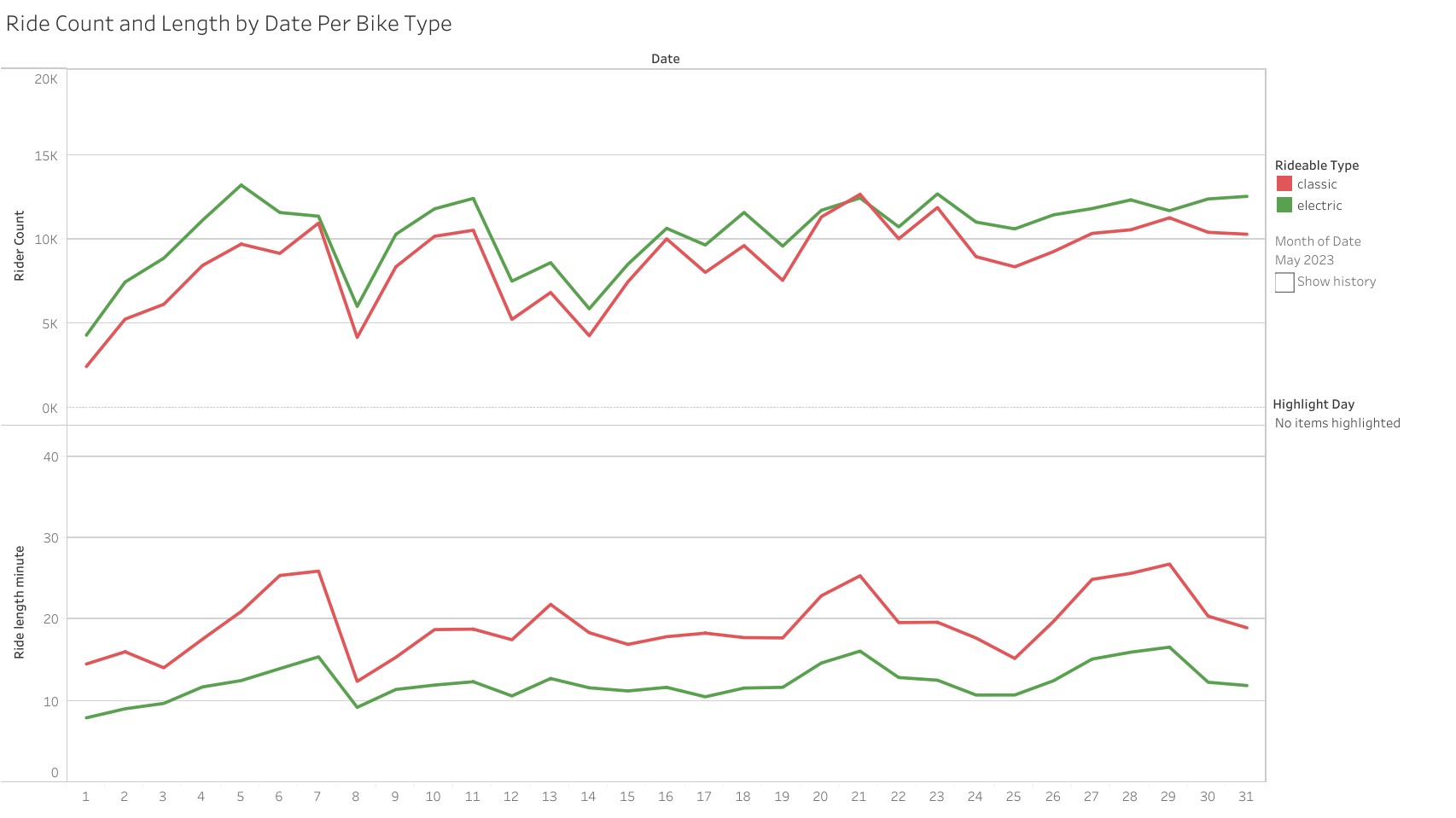
**Bike Types:**

****

Here I created graphs to compare bike type usage in total, by membership per day of the week, and by overall bike type used each month.

* Members use both bikes equally, while casuals tend to favor electric bikes
* Aside from a flip in popularity in July, both bike types seem to follow the same trend as weather changes
* Each bike type follows the same trend of casuals peaking on weekends, and members peaking in the middle of the week day.

**Ride Count and Length by Date Per Bike Type**



Here I dove deeper to see if there were any more trends I could identify with bike types used. The dashboard can be selected [here](https://public.tableau.com/app/profile/tyler.johnson1873/viz/TylerJohnsonCyclisticCaseStudy/RideLengthandCountPerType)

* Rider count followed almost the exact trends every month, with electric being slightly more popular in warmer months.
* Classic bikes were consistently used for longer bike rides, but still followed the same trends as electric bikes aside from a few spikes due to holidays.

**Phase 6: Act**

****

**How do annual members and casual riders use Cyclistic bikes differently?**

* Members tend to skew towards using Cyclistic to get to work, and have shorter rides
* Casual members use it more for leisure, especially on weekends, and have longer rides
* Casuals slightly favor electric bikes over classic

**Why would casual riders buy Cyclistic annual memberships?**

* A 4th pricing plan to get casual riders to become members: **The Weekend Pass**

**Weekend Pass Includes:**

* Unlimited rides Friday - Sunday that week for the price of two day passes
* A discount offer for the first year of membership

Casual riders already gravitate towards riding on the weekends, and by giving a discounted membership for anyone with a **Weekend Pass**, it may incentivize them to give membership a shot. The cost of one weekend pass plus the membership should also make up for the cost of the discount.

* Offer a discounted first year membership after a fixed amount of time ridden

As it has been shown that casual riders have the longer ride averages, perhaps offering a membership discount to those who regularly have long rides would convince them to give membership a try.

**How can Cyclistic use digital media to influence casual riders to become members?**

* Advertise heavily as weather begins to warm, with a strong focus in April
* Focus advertisements on weekends and promote the **Weekend Pass**
* Create videos highlighting members and how/why they use their membership