About Cyclistic Bike-Share

Since its launch in 2016, Cyclistic has grown their bike-share program to feature more than 5,800 bicycles and nearly 700 docking stations across Chicago, IL. Cyclistic sets itself apart from competing bike-share programs by offering reclining bikes, hand tricycles, and cargo bikes, making bikeshare more inclusive to people with disabilities. Although the company has gained the appeal of new customers through their flexible pricing plans: single ride passes, full day passes, and annual memberships, Cyclistic's finance analyst have determined that annual members are much more profitable than casual riders. The director, Lily Moreno and the executive team aim to convert casual riders into annual members but must first understand how the riders differ.

Questions for Analysis

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

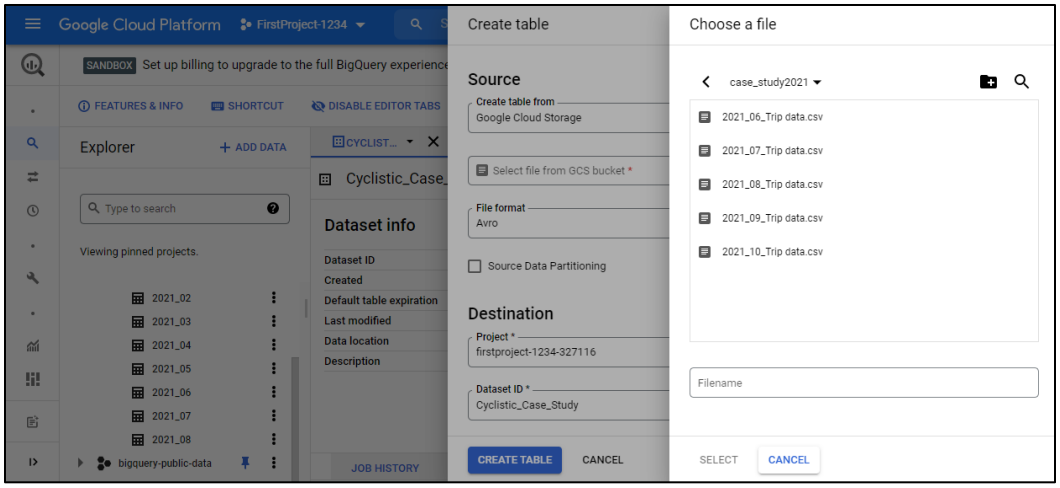
Ask- Business Task Statement

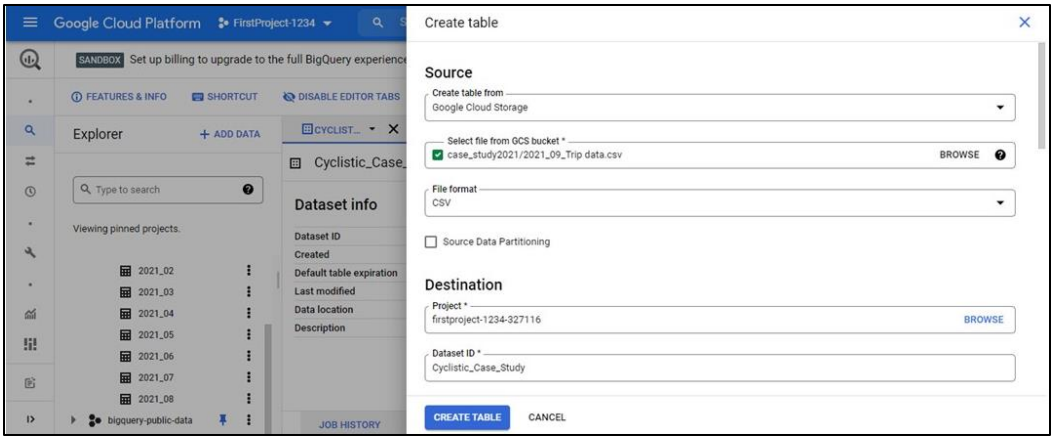
How do annual members and casual riders use Cyclistic bikes differently?

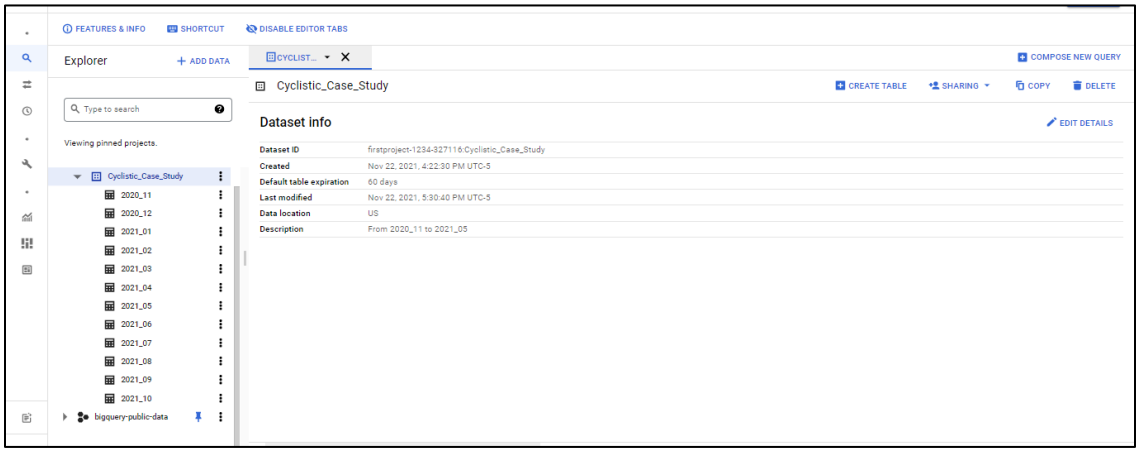
Prepare- Data Exploration Preparation

For this case study, I downloaded the previous 12 months of Cyclistic Trip Data. The quantitative data is structured in rows and columns with each table containing 13 fields. I extracted 12 files dating from November 2020 – October 2021, as this matched the required timeframe.

Although the data has been made available by Motivate International Inc. under this license, there is no information found on how the data was collected nor is there any information given on when it was last updated and its frequency. The dataset provides bikeshare information of both annual members and casual riders allowing analysis that will provide insight as to how annual members and casual riders use the service. I uploaded the files to BigQuery to begin data cleaning.







Process- Data Cleaning

I have elected to use the SQL workspace on BigQuery for cleaning and analysis. Since its introduction in week three of Course 2: Prepare Data for Exploration, I have enjoyed the user-friendly interface and the

ability to explore public datasets. I believe this platform is great place for those that are new to data analytics.

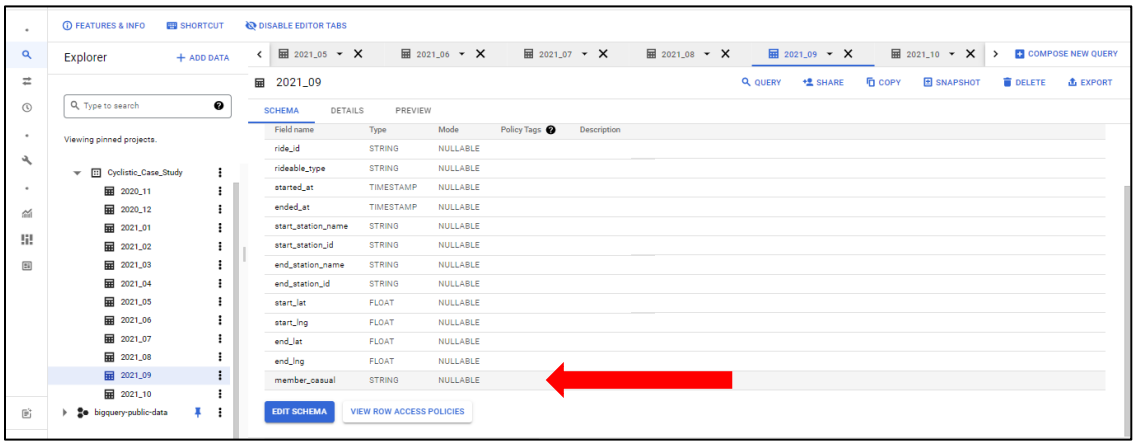
As directed on page 7 of the Case Study 1 details, I also used Excel/Google Sheets to manipulate the data to discover additional insights of the dataset.

I began the data cleaning process by following the Data Cleaning Verification Checklist, introduced in week four of Course 4: Process Data from Dirty to Clean.

Unique Strings: *member_casual*

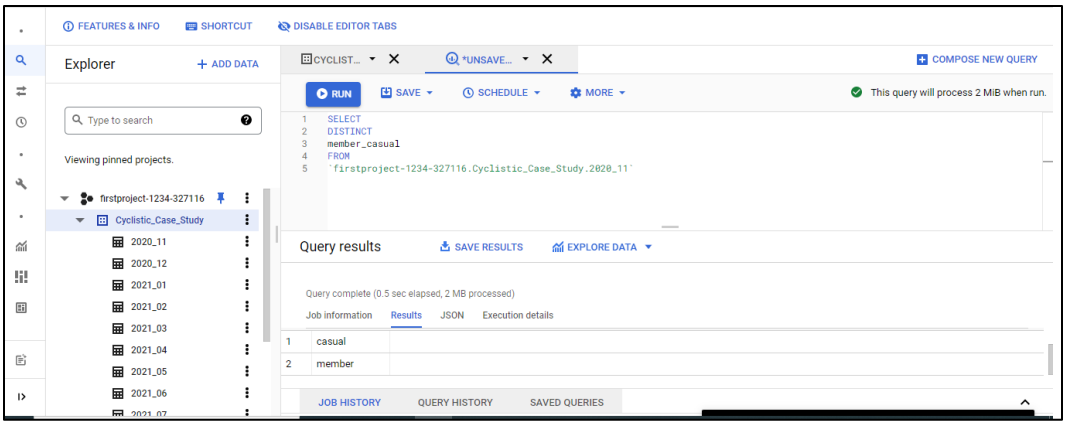
While this entire project henges on the bikeshare information of both annual members and casual riders, I ran a query to ensure that each *member_casual* column had only two unique strings: member or casual.

Upon an initial review of the schema for each table, I found that all the tables contained a *member_casual* column.



A quick query confirmed that each table contained only two unique strings for this column.

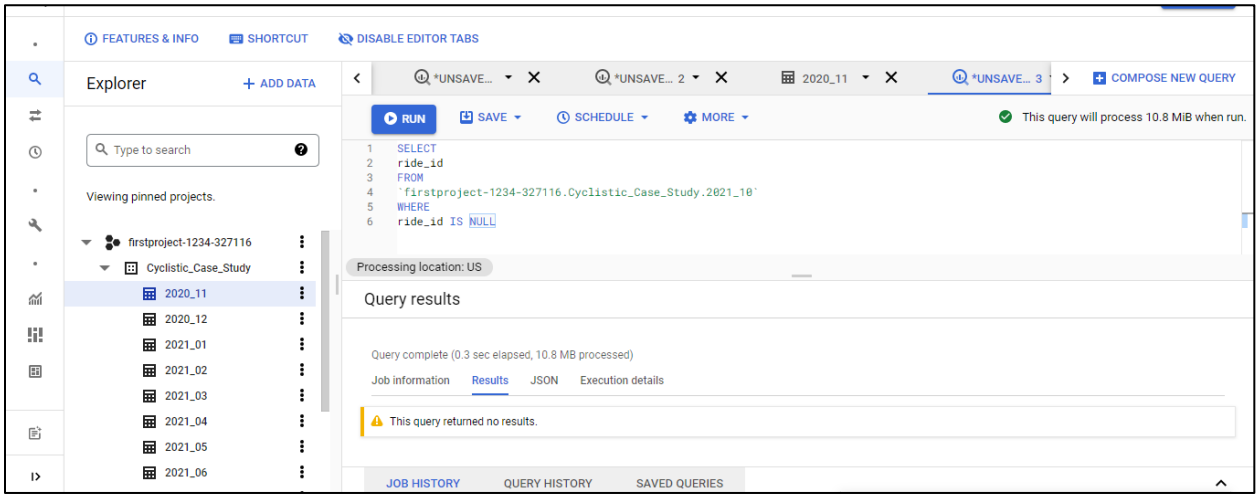
```
SELECT
DISTINCT
member_casual
FROM
`firstproject-1234-327116.Cyclistic_Case_Study.2020_11`
```



Null Values: *ride_id & rideable_type*

To gain a clear picture of how annual members differ from casual riders, I wanted to ensure that specific data points are included, and null values are either updated or determined not needed for analysis.

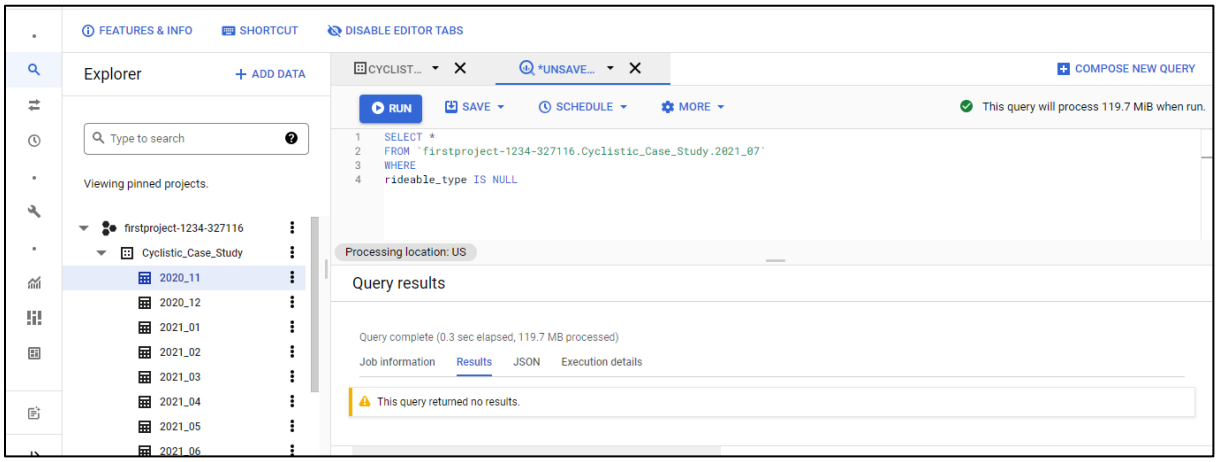
```
SELECT
ride_id
FROM
`firstproject-1234-327116.Cyclistic_Case_Study.2021_10`
WHERE
ride_id IS NULL
```



A query for null values of the *ride_id* column for each table did not return any results.

Information such as the types of bikes that members and casual riders are utilizing will be important for comparison during the analysis phase. The *rideable_type* column should include the type of bike rented by each rider. I also want to ensure that this column does not return null values for any of the tables in this dataset.

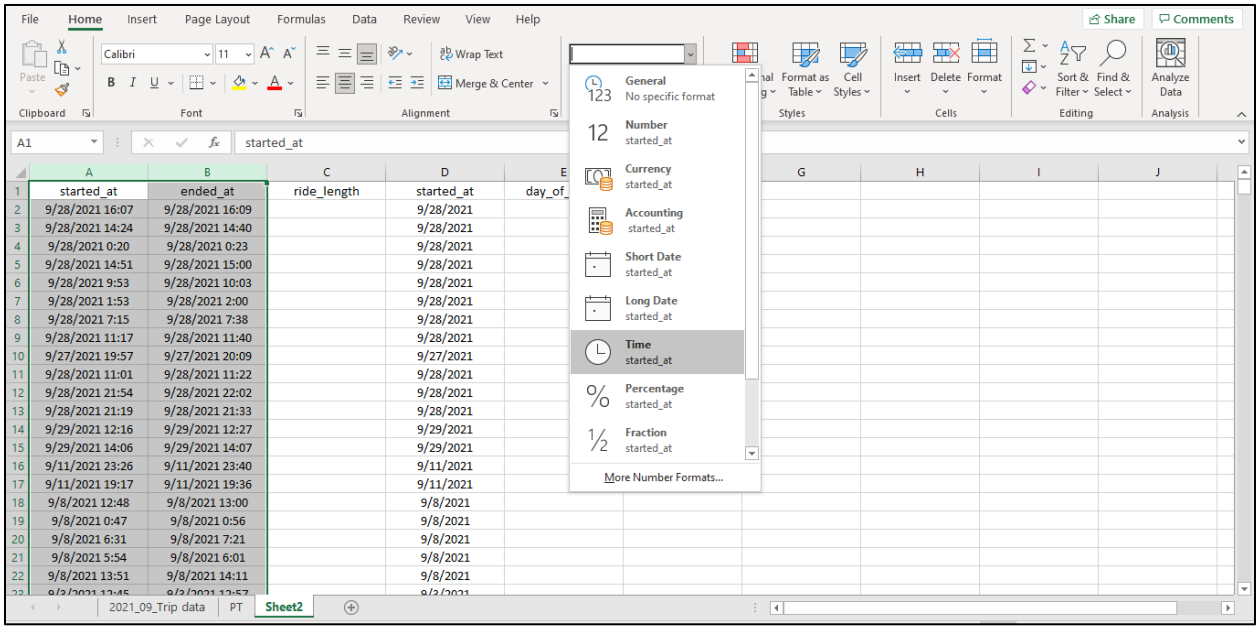
```
SELECT *
FROM `firstproject-1234-327116.Cyclistic_Case_Study.2021_10`
WHERE
rideable_type IS NULL
```



This query did not return any null values for *rideable_type*.

Excel/Google Sheets

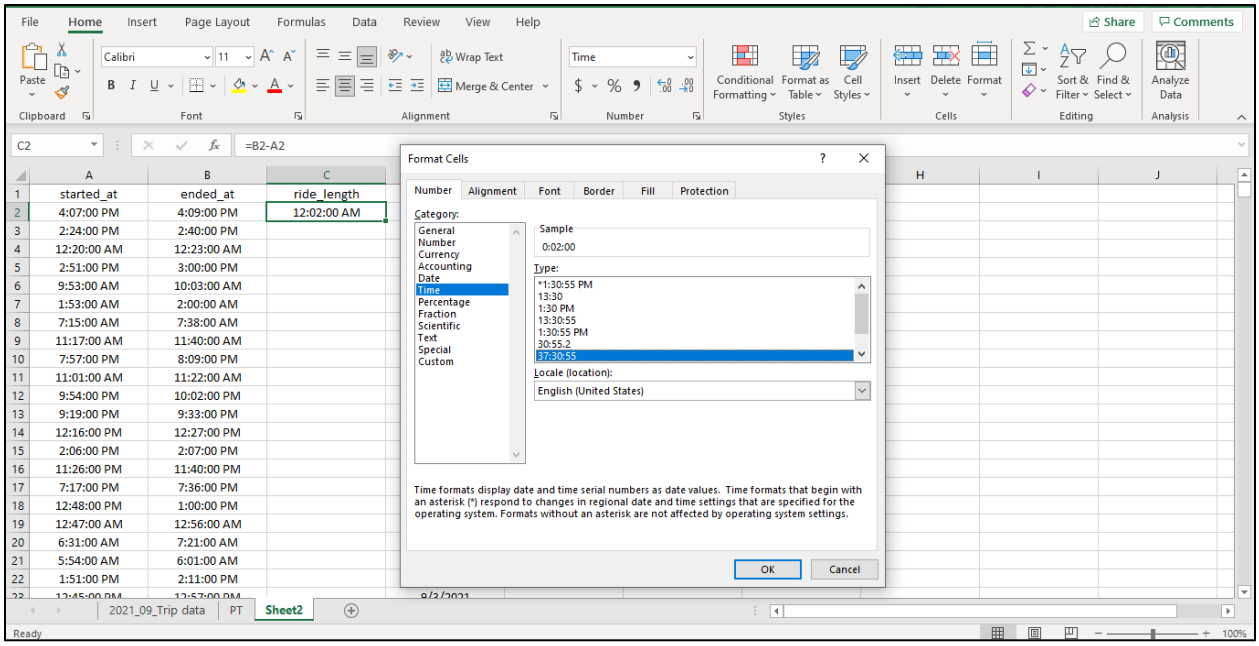
To determine the length of each ride, I copied and pasted the “started_at” and “end_at” columns into a separate spreadsheet and then converted them to Time. I then created a column called “ride_length”.



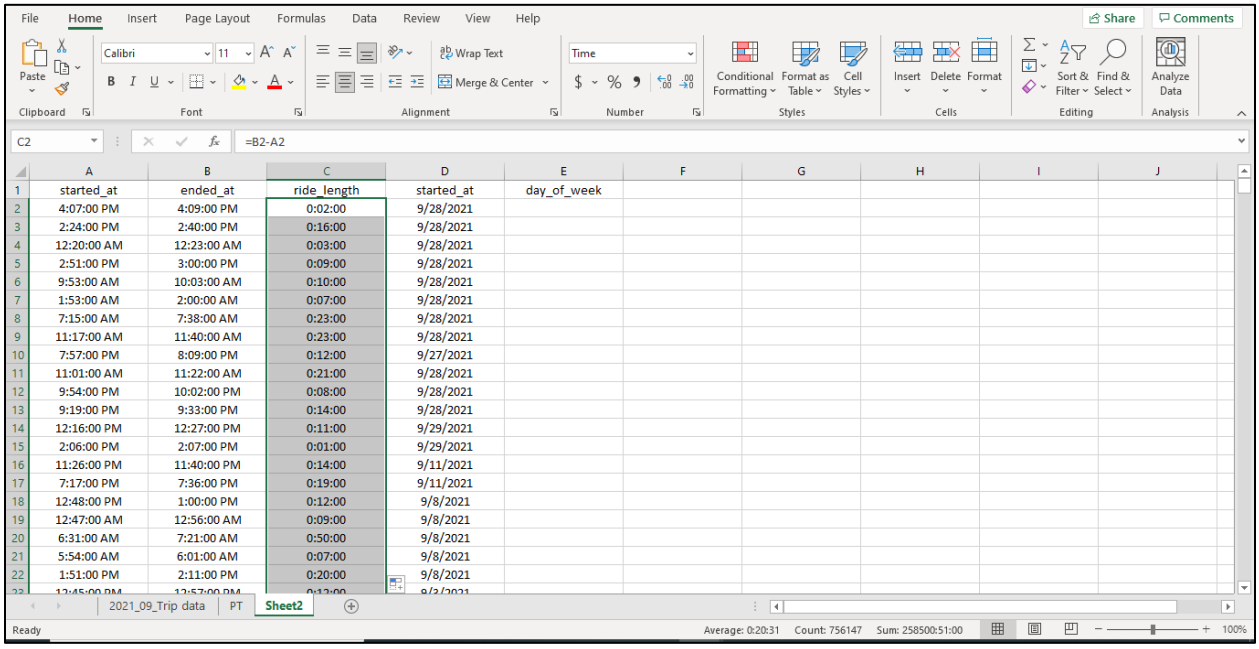
Next, I entered the following formula to determine the ride_length:

=B2-A2

I converted the output to HH:MM:SS to determine the duration of each ride.

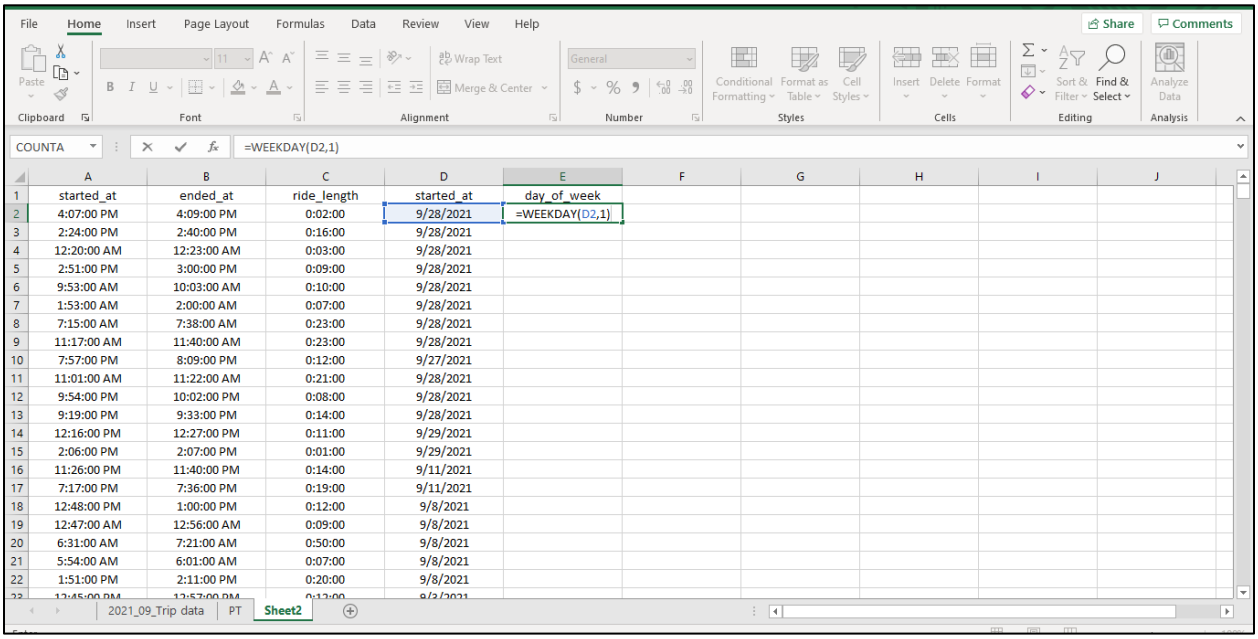


This conversion was then copied down the entire column.



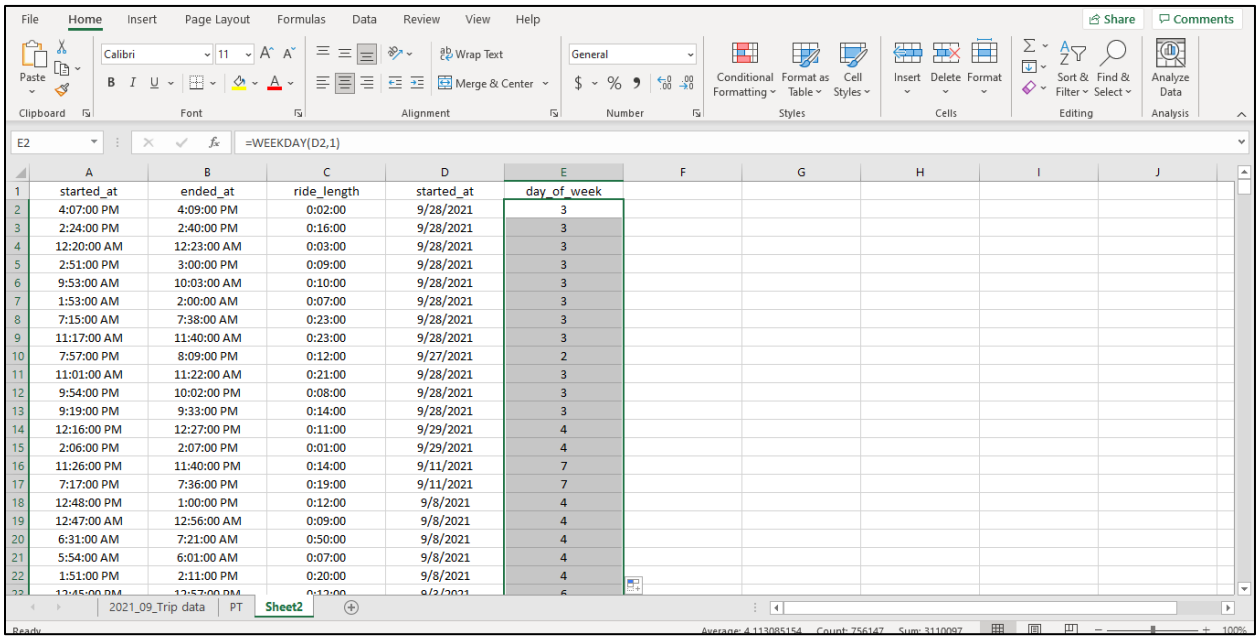
To determine the day of the week for each ride, I copied and pasted the started_at column and converted the column to Short Date. I then entered the following formula:

=WEEKDAY(D2,1)



started_at	ended_at	ride_length	started_at	day_of_week
4:07:00 PM	4:09:00 PM	0:02:00	9/28/2021	=WEEKDAY(D2,1)
2:24:00 PM	2:40:00 PM	0:16:00	9/28/2021	
12:20:00 AM	12:23:00 AM	0:03:00	9/28/2021	
2:51:00 PM	3:00:00 PM	0:09:00	9/28/2021	
9:53:00 AM	10:03:00 AM	0:10:00	9/28/2021	
1:53:00 AM	2:00:00 AM	0:07:00	9/28/2021	
7:15:00 AM	7:38:00 AM	0:23:00	9/28/2021	
11:17:00 AM	11:40:00 AM	0:23:00	9/28/2021	
7:57:00 PM	8:09:00 PM	0:12:00	9/27/2021	
11:01:00 AM	11:22:00 AM	0:21:00	9/28/2021	
9:54:00 PM	10:02:00 PM	0:08:00	9/28/2021	
9:19:00 PM	9:33:00 PM	0:14:00	9/28/2021	
12:16:00 PM	12:27:00 PM	0:11:00	9/29/2021	
2:06:00 PM	2:07:00 PM	0:01:00	9/29/2021	
11:26:00 PM	11:40:00 PM	0:14:00	9/11/2021	
7:17:00 PM	7:36:00 PM	0:19:00	9/11/2021	
12:48:00 PM	1:00:00 PM	0:12:00	9/8/2021	
12:47:00 AM	12:56:00 AM	0:09:00	9/8/2021	
6:31:00 AM	7:21:00 AM	0:50:00	9/8/2021	
5:54:00 AM	6:01:00 AM	0:07:00	9/8/2021	
1:51:00 PM	2:11:00 PM	0:20:00	9/8/2021	

I then copied the formula down the entire column. The day of the week is now shown for each ride with 1=Sunday and 7=Saturday.



started_at	ended_at	ride_length	started_at	day_of_week
4:07:00 PM	4:09:00 PM	0:02:00	9/28/2021	3
2:24:00 PM	2:40:00 PM	0:16:00	9/28/2021	3
12:20:00 AM	12:23:00 AM	0:03:00	9/28/2021	3
2:51:00 PM	3:00:00 PM	0:09:00	9/28/2021	3
9:53:00 AM	10:03:00 AM	0:10:00	9/28/2021	3
1:53:00 AM	2:00:00 AM	0:07:00	9/28/2021	3
7:15:00 AM	7:38:00 AM	0:23:00	9/28/2021	3
11:17:00 AM	11:40:00 AM	0:23:00	9/28/2021	3
7:57:00 PM	8:09:00 PM	0:12:00	9/27/2021	2
11:01:00 AM	11:22:00 AM	0:21:00	9/28/2021	3
9:54:00 PM	10:02:00 PM	0:08:00	9/28/2021	3
9:19:00 PM	9:33:00 PM	0:14:00	9/28/2021	3
12:16:00 PM	12:27:00 PM	0:11:00	9/29/2021	4
2:06:00 PM	2:07:00 PM	0:01:00	9/29/2021	4
11:26:00 PM	11:40:00 PM	0:14:00	9/11/2021	7
7:17:00 PM	7:36:00 PM	0:19:00	9/11/2021	7
12:48:00 PM	1:00:00 PM	0:12:00	9/8/2021	4
12:47:00 AM	12:56:00 AM	0:09:00	9/8/2021	4
6:31:00 AM	7:21:00 AM	0:50:00	9/8/2021	4
5:54:00 AM	6:01:00 AM	0:07:00	9/8/2021	4
1:51:00 PM	2:11:00 PM	0:20:00	9/8/2021	4

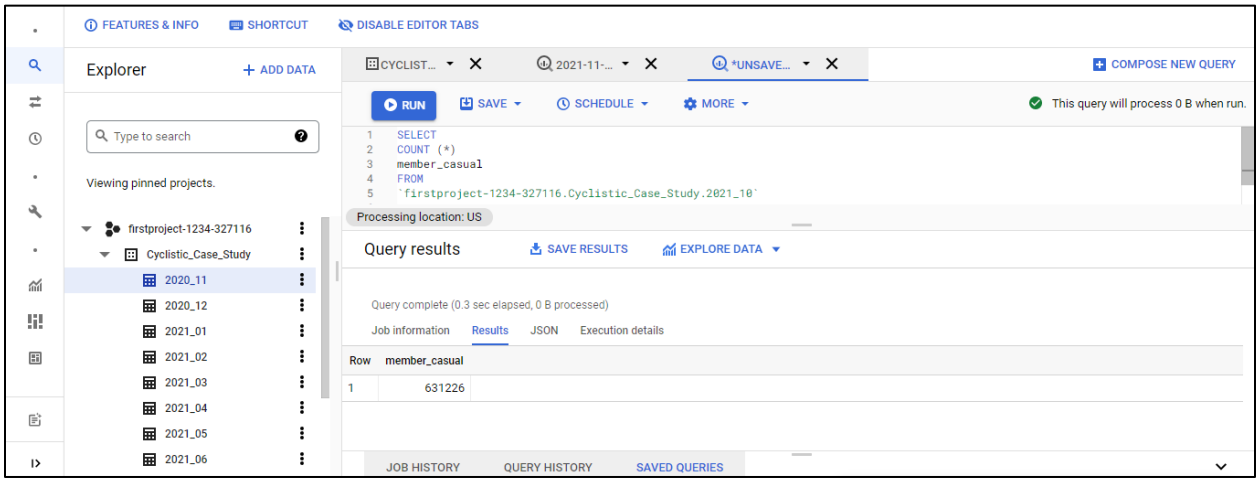
I repeated the steps under this section for each file.

Analyze

Here are a few insights that I gleaned from the dataset.

Total numbers of riders by month

```
SELECT
COUNT (*)
member_casual
FROM
`firstproject-1234-327116.Cyclistic_Case_Study.2021_10`
```

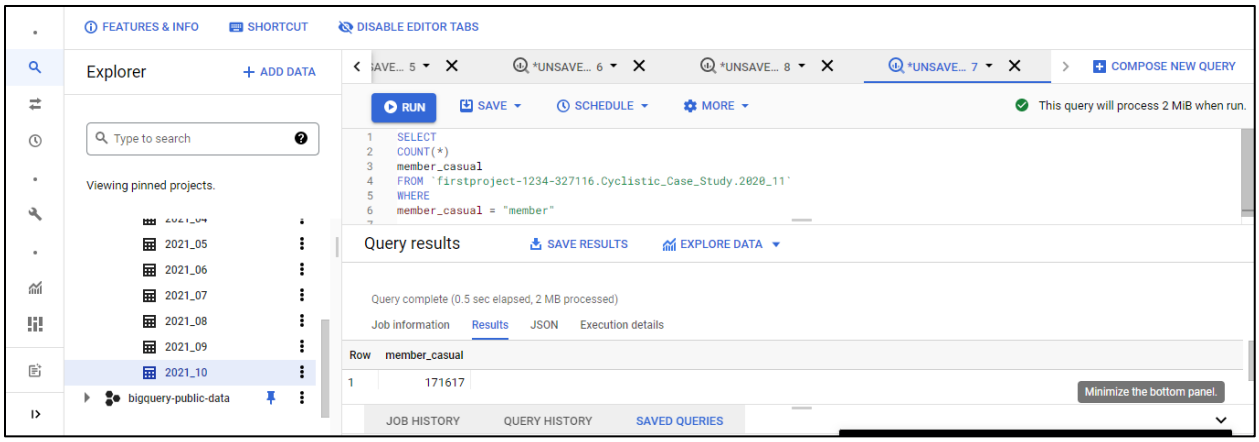


Month	Total Riders
2020_11	259,716
2020_12	131,573
2021_01	96,834
2021_02	49,622
2021_03	228,496
2021_04	337,230
2021_05	531,633
2021_06	729,595
2021_07	822,410
2021_08	804,352
2021_09	756,147
2021_10	631,226

As expected, the number of riders declines heading into the fall and winter months and rises in the spring and summer months.

Total number of annual members versus casual riders by month

```
SELECT
COUNT(*)
member_casual
FROM `firstproject-1234-327116.Cyclistic_Case_Study.2020_11`
WHERE
member_casual = "member"
```



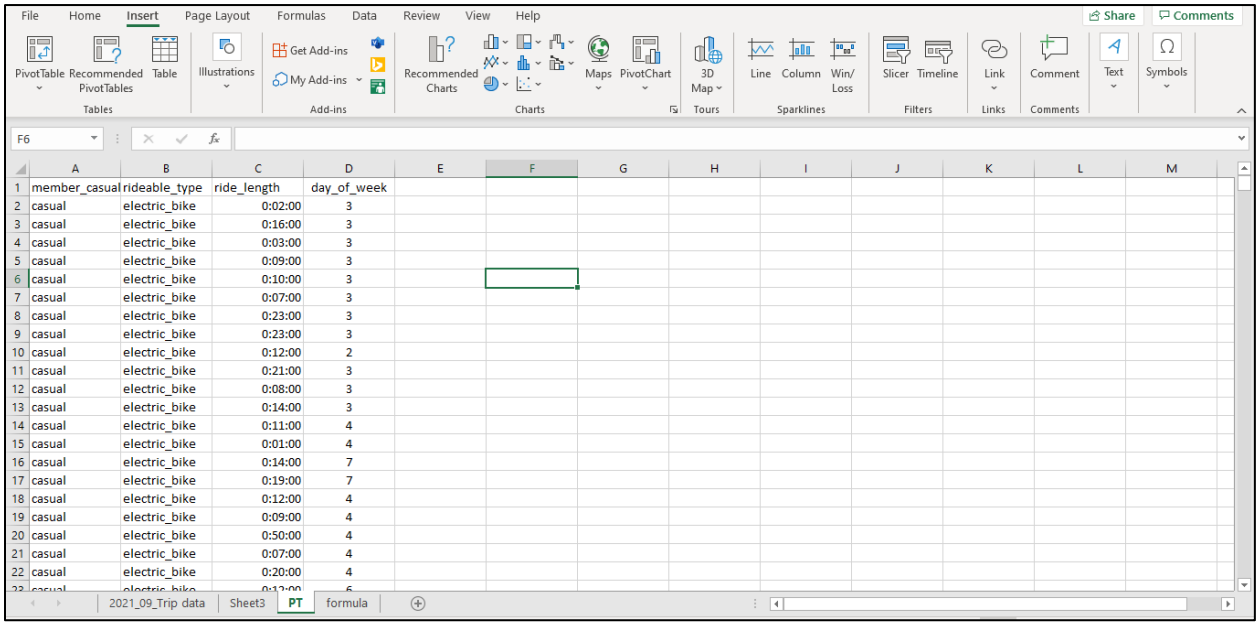
Month	*Casual*	*Annual*
2021_10	41%	59%
2021_09	48%	52%
2021_08	51%	49%
2021_07	54%	46%
2021_06	51%	49%
2021_05	48%	52%
2021_04	41%	59%
2021_03	37%	63%
2021_02	20%	80%
2021_01	19%	81%
2020_12	23%	77%
2020_11	34%	66%

At the onset of this dataset, the percentage point difference varies between casual riders and annual members significantly with 32% in November 2020, 54% in December 2020, 62% in January 2021, and 60% in February 2021. It is not until April 2021, that the percentage point difference between casual riders and annual riders decreased to a one-digit. This trend remained steady until last month, October 2021, where annual riders held a difference of 18% over casual riders.

Casual riders held a small margin over annual riders from June 2021 to August 2021.

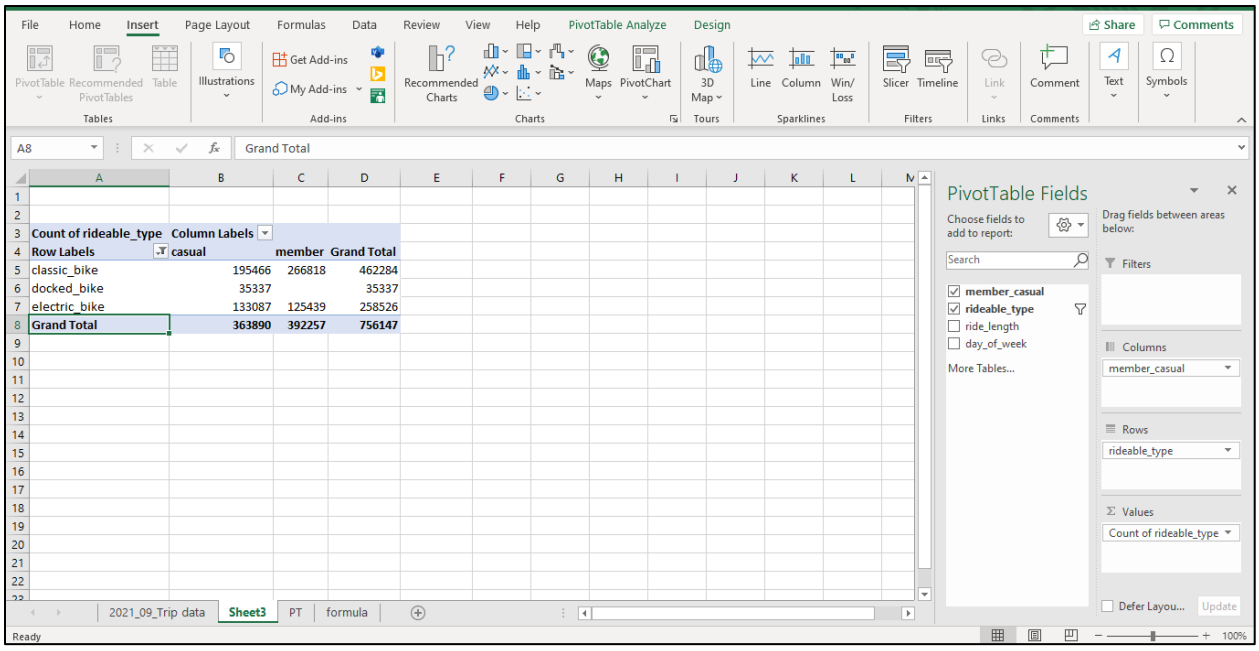
Rideable Types and Day of the Week Frequency by Riders

To determine the total rideable types and rides per day of the week, I opened each file and copied the needed columns from the file onto a separate spreadsheet.



member_casual	rideable_type	ride_length	day_of_week
casual	electric_bike	0:02:00	3
casual	electric_bike	0:16:00	3
casual	electric_bike	0:03:00	3
casual	electric_bike	0:09:00	3
casual	electric_bike	0:10:00	3
casual	electric_bike	0:07:00	3
casual	electric_bike	0:23:00	3
casual	electric_bike	0:23:00	3
casual	electric_bike	0:12:00	2
casual	electric_bike	0:21:00	3
casual	electric_bike	0:08:00	3
casual	electric_bike	0:14:00	3
casual	electric_bike	0:11:00	4
casual	electric_bike	0:01:00	4
casual	electric_bike	0:14:00	7
casual	electric_bike	0:19:00	7
casual	electric_bike	0:12:00	4
casual	electric_bike	0:09:00	4
casual	electric_bike	0:50:00	4
casual	electric_bike	0:07:00	4
casual	electric_bike	0:20:00	4

I then created a pivot table for rideable types and days of the week. I repeated this process for each file.



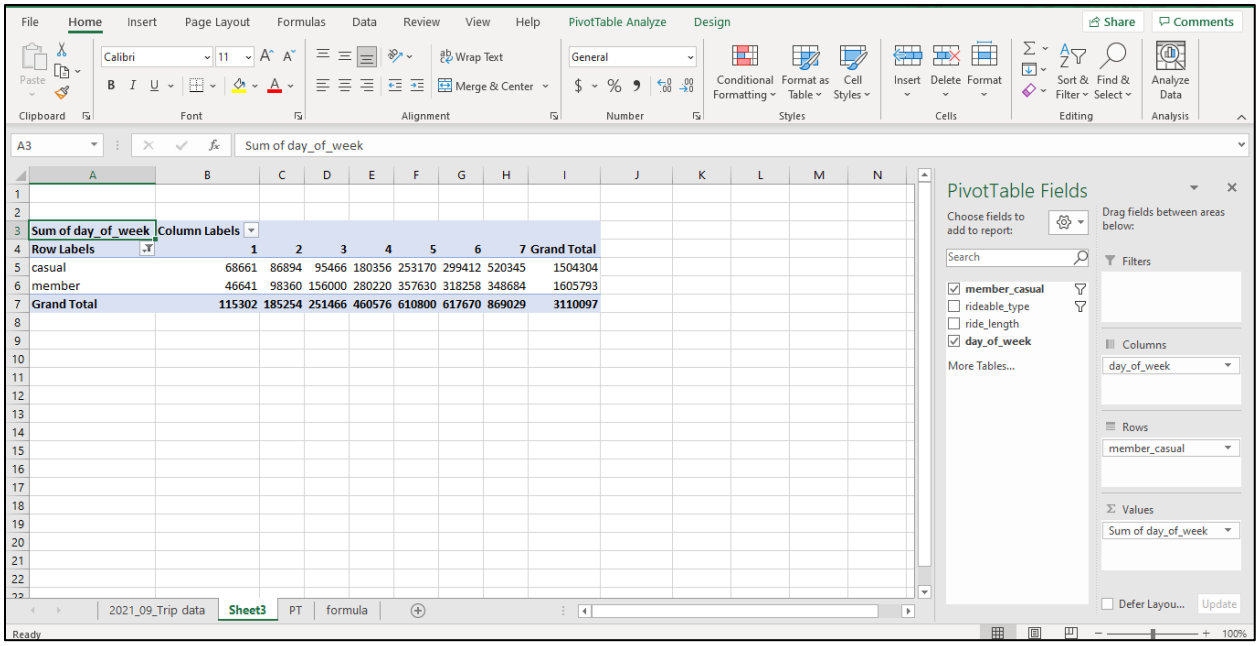
Count of rideable_type	Column Labels		
Row Labels	casual	member	Grand Total
classic_bike	195466	266818	462284
docked_bike	35337		35337
electric_bike	133087	125439	258526
Grand Total	363890	392257	756147

		2020_11	2020_12	2021_01	2021_02	2021_03	2021_04	2021_05	2021_06	2021_07	2021_08	2021_09	2021_10
docked	Casual	46,013	4,968	2,105	1,271	15,657	24,714	43,353	51,716	57,698	45,065	35,337	22,884
	Member	105,569	8,036	1	0	0	0	0	0	0	0	0	0
		2020_11	2020_12	2021_01	2021_02	2021_03	2021_04	2021_05	2021_06	2021_07	2021_08	2021_09	2021_10
electric	Casual	42,086	13,793	7,753	3,165	22,848	41,111	89,613	130,720	142,869	137,618	133,087	128,769
	Member	66,048	34,160	25,275	10,174	37,446	56,786	89,574	112,139	114,934	118,636	125,439	163,434
classic		2020_11	2020_12	2021_01	2021_02	2021_03	2021_04	2021_05	2021_06	2021_07	2021_08	2021_09	2021_10
	Casual		11,319	8,259	5,695	45,528	70,776	123,950	188,245	241,489	229,988	195,466	105,589
	Member		59,297	53,441	29,317	107,017	143,843	185,143	246,775	265,420	273,045	266,818	210,550

The data on docked bikes created the most curiosity among the rideable types. In November 2011, annual members docked a bike 105,569 times. This number decreased by 92% the following month to 8,036 docked bikes. With only one bike docked in January 2021, the remainder of the year fell flat at 0% of docked bikes among annual members.

Data on docked bikes for casual riders began with a baseline of 46,013 in November 2020. Similar to annual members, the number of docked bikes decreased by 89% the following month. Decreases continued for January 2021 (57%) and February 2021 (39%) respectively, however data shows a gain just over 1100% from February to March. Steady increases were seen over the proceeding four months for docked bikes among casual riders. Decreases were recorded at a starting mark at 21% from July to August and has since fell by 35% from September to October.

While there is not a baseline datapoint for classic bike riders from November to December 2020, there was a consistent and moderate decrease in electric and classic bikes among all riders from December to February. From February to March, electric and classic bikes yielded the largest increases of over 600% among casual riders and over 200% among annual riders. These bike types continued to increase steadily from March until July. From July to October 2021, there has been a slow decrease of casual riders utilizing electric and classic bike types. Although annual members began showing a decrease in classic bikes starting in August, and data on electric bikes among annual riders has remained positive and has consistently shown growth since March 2021.



Month	*Casual*	*Annual*
2021_10	7	7
2021_09	7	5
2021_08	1	3
2021_07	7	5
2021_06	7	4
2021_05	7	7
2021_04	7	6
2021_03	7	3
2021_02	7	4
2021_01	7	6
2020_12	4	4
2020_11	2	7

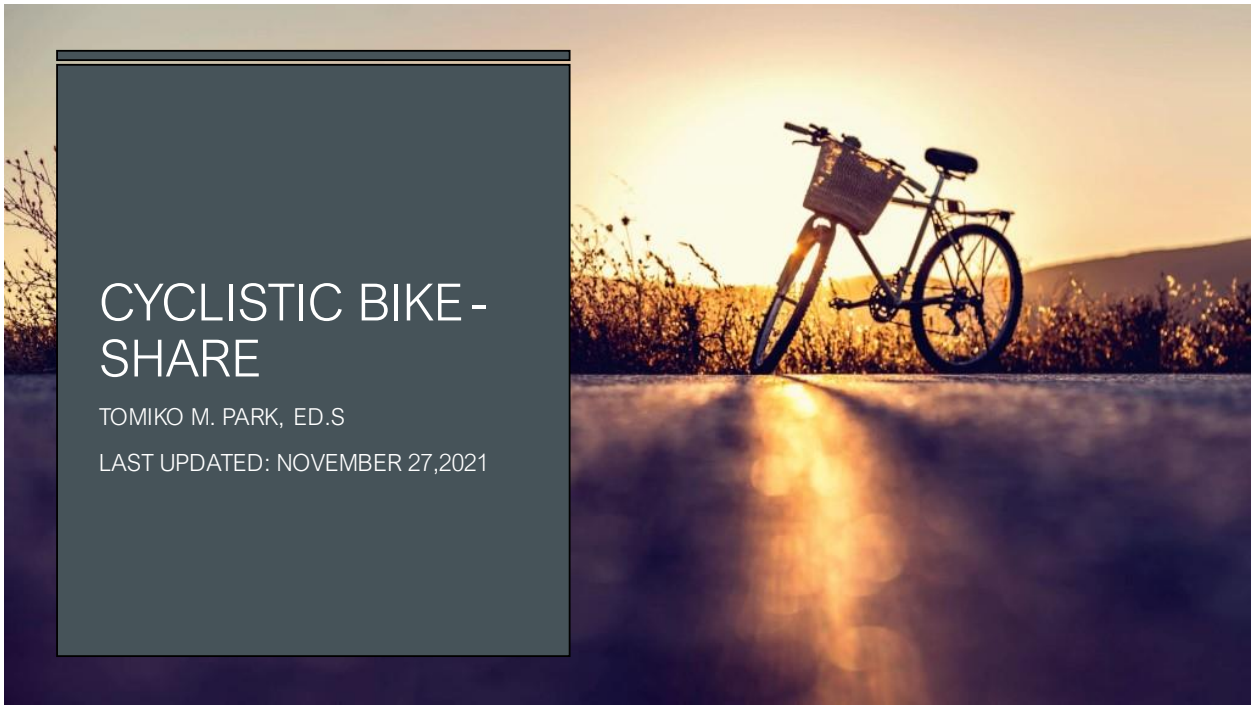
While the data is overwhelming that Saturday is the most popular day for casual riders, there is an even distribution between Wednesday and Saturday for annual members.

MAX Ride_length by Rider

MONTH	CASUAL	ANNUAL
2021_10	678:25:00	25:00:00
2021_09	547:39:00	25:00:00
2021_08	693:49:00	25:00:00
2021_07	818:27:00	25:00:00
2021_06	932:24:00	25:00:00
2021_05	898:41:00	25:00:00
2021_04	796:16:00	25:00:00
2021_03	528:02:00	26:00:00
2021_02	502:10:00	25:00:00
2021_01	330:26:00	25:00:00
2020_12	162:21:00	25:00:00
2020_11	598:54:00	25:00:00

When calculating the ride_length for each rider, I found that the MAX ride_length consistently fell at 25 minutes for annual riders. From the chart above, the MAX times for casual riders clearly exceeded beyond average duration times for casual riders which is approximately 28 minutes (Retrieved from https://nacto.org/bike-share-statistics-2017/) The average MAX ride_length time for casual riders was 624 hours (25 days).

Share



CYCLISTIC BIKE-SHARE

TOMIKO M. PARK, ED.S

LAST UPDATED: NOVEMBER 27,2021

GOALS FOR OUR DISCUSSION TODAY

1. SHARE THE OBJECTIVE AND PURPOSE
2. EXAMINE TRENDS AND RELATIONSHIPS GATHERED FROM THE DATASET
3. DISCUSS RECOMMENDATIONS AND DETERMINE NEXT STEPS

1

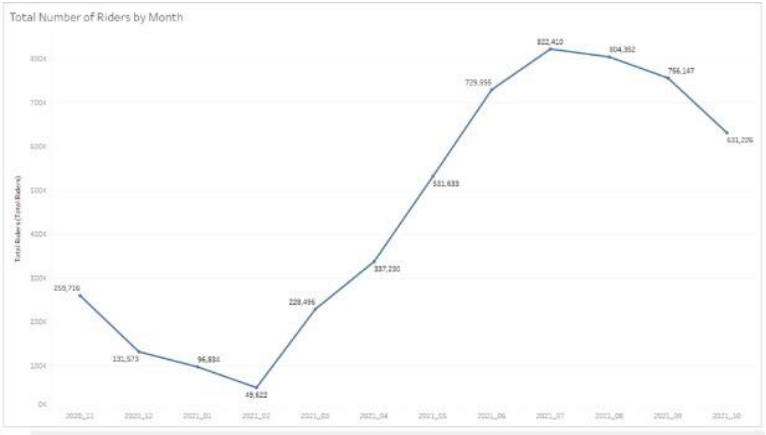
OBJECTIVE

- Cyclistic Bikeshare aims to convert casual riders to subscribing, annual members. Examine how casual riders and annual riders use Cyclistic bikes differently.

2

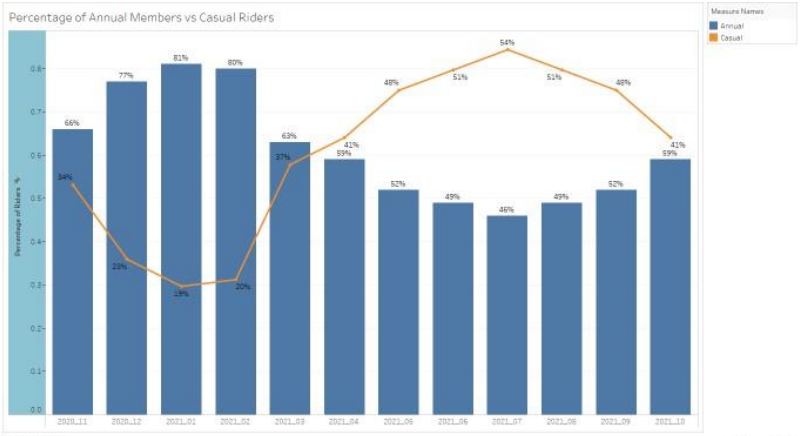
TOTAL NUMBER OF RIDERS BY MONTH

- Data shows the total number of riders decreases heading into the cooler months. As expected, the data also reflects a steady increase approaching warmer temperatures.



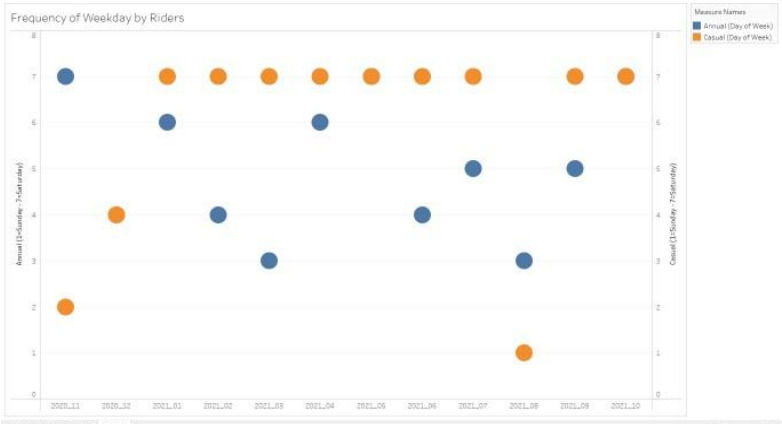
MONTHLY PERCENTAGE OF ANNUAL MEMBERS VS CASUAL RIDERS

- At the onset of this data, the percentage point difference between types of riders varied significantly with 32% in November and grew to a 62% difference in January 2021.
- Although the percentages of both types of riders nearly evens out starting in April 2021, data shows that casual riders are utilizing the service mostly in the summer months.



FREQUENCY OF WEEKDAY BY RIDERS

- While the data is clear that casual riders patronize the bikeshare service primarily on Saturdays, annual members are distributed evenly between Wednesday and Saturday.



KEY TAKEAWAYS

- Cyclistic is likely to see a decrease of usage among casual riders during the fall and winter months and an increase in annual members at the same time.
- Data shows that casual riders are “weekend” riders, whereas annual riders are likely to use the service throughout the week.
- Ride _length times varied greatly between annual members and casual riders. Where annual members are riding at a max time of twenty -five minutes , casual riders’ max times for each month are on average 26 days.

3

Problem

- There is a decreased in usage among casual users in the fall and winter months.

Recommendation

- Consider using a reward point system in the fall and winter months that can be redeemed during the spring and summer months.
- Additional research of competing programs, as well as, holding focus groups with current casual riders may uncover determining factors that are connected to why casual riders are not committing to Cyclistic bikeshare program.

Problem

Casual riders typically use the service on the weekends, while annual members utilize the service throughout the week.

Recommendation

- A comparison of the company's current plans to competing plans may reveal factors of why single -ride and full-day passes appear more feasible, while the option for annual memberships does not garner the attractiveness to gain new members.
- Consider forming partnerships with local eateries and shops who are also looking for a boost in revenue during the week. This could allow Cyclistic to see increased usage during the week and with a makeover of the current pricing plans, casual riders may take advantage of a bundled offer.

Problem

Max ride length times far exceed average ride times for casual riders. Data shows that on average max ride time for Cyclistic's casual riders is approximately 26 days.

Recommendation

- A further examination should be made to connect the rider_id, starting_at_station and ending_at_station to determine if the max ride times are attributed to dockless bikes.
- If the company's terms and polices allow for text messages, casual riders should be set up for text and email reminders to return bikes to a nearby docking station as their rental period ends.
- Include longer ride passes as an incentive to joining the bikeshare program.

Thank you for viewing my capstone project! I welcome all feedback (and job offers)!

Tomiko M. Park, Ed.S

Contact Information:
tomikompark@gmail.com
404-358-1967

View my resume here!