Cyclist case study
Omkar Mankame
Date – 25 August 2024

Steps for data analysis -

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

The marketing director of a bike rental company believes that converting more memberships to annual memberships is beneficial for the growth of the organization. To make this concrete, I as a junior data analyst am responsible to tell the company how the bike usage differs based on type of memberships.

Data viewing process –

1. The data to be used was selected for 2023 as it had all the monthly data and was recent data.
2. Data was in zip files.
3. All the data was unzipped and the spreadsheets were viewed thoroughly.
4. Data was not consistent in all the spreadsheets. Example, not all spreadsheets had start station, end station, etc.

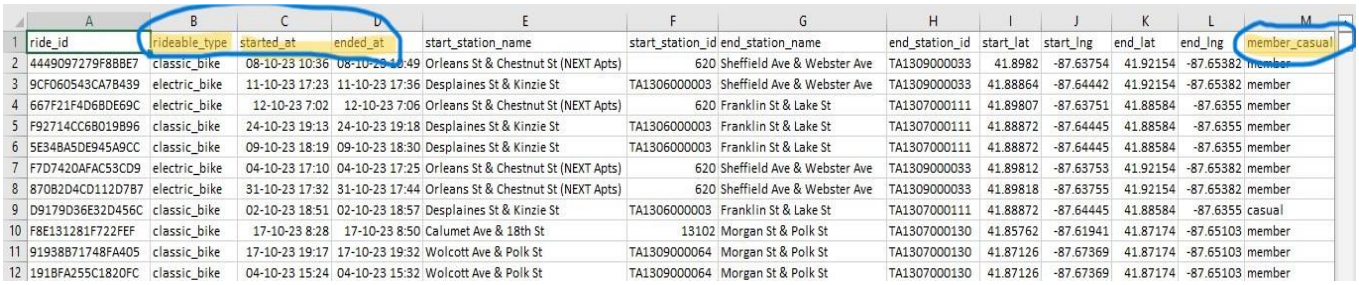
Incomplete data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ride_id	rideable_t	started_at	ended_at	start_stati	start_stati	end_statio	end_statio	start_lat	start_lng	end_lat	end_lng	member_casual			
2	6F1682AC4	electric_b	########	########					41.91	-87.69	41.91	-87.7	member			
3	622A1686I	electric_b	########	########					41.94	-87.65	41.94	-87.65	member			
4	3C88859D'	electric_b	########	########					41.95	-87.68	41.92	-87.63	member			
5	EAD8A5E0	electric_b	########	########					41.99	-87.65	41.98	-87.66	member			
6	5A36F219:	electric_b	########	########					41.98	-87.66	41.99	-87.65	member			
7	CF682EA7I	electric_b	########	########					41.99	-87.68	41.94	-87.65	member			
8	4910FBB7:	electric_b	########	########					41.88	-87.62	41.88	-87.62	member			
9	EA19D850	electric_b	########	########					41.88	-87.62	41.88	-87.62	member			

Complete data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
2	4449097279F8BBE7	classic_bike	08-10-23 10:36	08-10-23 10:49	Orleans St & Chestnut St (NEXT Apts)	620	Sheffield Ave & Webster Ave	TA1309000033	41.8982	-87.63754	41.92154	-87.65382	member
3	9CF060543CA7B439	electric_bike	11-10-23 17:23	11-10-23 17:36	Desplaines St & Kinzie St	TA1306000003	Sheffield Ave & Webster Ave	TA1309000033	41.88864	-87.64442	41.92154	-87.65382	member
4	667F21F4D6BDE69C	electric_bike	12-10-23 7:02	12-10-23 7:06	Orleans St & Chestnut St (NEXT Apts)	620	Franklin St & Lake St	TA1307000111	41.89807	-87.63751	41.88584	-87.6355	member
5	F92714CC6B019B96	classic_bike	24-10-23 19:13	24-10-23 19:18	Desplaines St & Kinzie St	TA1306000003	Franklin St & Lake St	TA1307000111	41.88872	-87.64445	41.88584	-87.6355	member
6	5E34BA5DE945A9CC	classic_bike	09-10-23 18:19	09-10-23 18:30	Desplaines St & Kinzie St	TA1306000003	Franklin St & Lake St	TA1307000111	41.88872	-87.64445	41.88584	-87.6355	member
7	F7D7420AFAC53CD9	electric_bike	04-10-23 17:10	04-10-23 17:25	Orleans St & Chestnut St (NEXT Apts)	620	Sheffield Ave & Webster Ave	TA1309000033	41.89812	-87.63753	41.92154	-87.65382	member
8	870B2D4CD112D7B7	electric_bike	31-10-23 17:32	31-10-23 17:44	Orleans St & Chestnut St (NEXT Apts)	620	Sheffield Ave & Webster Ave	TA1309000033	41.89818	-87.63755	41.92154	-87.65382	member
9	D9179D36E32D456C	classic_bike	02-10-23 18:51	02-10-23 18:57	Desplaines St & Kinzie St	TA1306000003	Franklin St & Lake St	TA1307000111	41.88872	-87.64445	41.88584	-87.6355	casual
10	F8E131281F722FEF	classic_bike	17-10-23 8:28	17-10-23 8:50	Calumet Ave & 18th St	13102	Morgan St & Polk St	TA1307000130	41.85762	-87.61941	41.87174	-87.65103	member
11	91938B71748FA405	classic_bike	17-10-23 19:17	17-10-23 19:32	Wolcott Ave & Polk St	TA1309000064	Morgan St & Polk St	TA1307000130	41.87126	-87.67369	41.87174	-87.65103	member
12	191BFA255C1820FC	classic_bike	04-10-23 15:24	04-10-23 15:32	Wolcott Ave & Polk St	TA1309000064	Morgan St & Polk St	TA1307000130	41.87126	-87.67369	41.87174	-87.65103	member

5. There were some common columns in all the spreadsheets which seemed to be useful for further analysis.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
2	4449097279F8BBE7	classic_bike	08-10-23 10:36	08-10-23 10:49	Orleans St & Chestnut St (NEXT Apts)	620	Sheffield Ave & Webster Ave	TA1309000033	41.8982	-87.63754	41.92154	-87.65382	member
3	9CF060543CA7B439	electric_bike	11-10-23 17:23	11-10-23 17:36	Desplaines St & Kinzie St	TA1306000003	Sheffield Ave & Webster Ave	TA1309000033	41.88864	-87.64442	41.92154	-87.65382	member
4	667F21F4D6BDE69C	electric_bike	12-10-23 7:02	12-10-23 7:06	Orleans St & Chestnut St (NEXT Apts)	620	Franklin St & Lake St	TA1307000111	41.89807	-87.63751	41.88584	-87.6355	member
5	F92714CC6B019B96	classic_bike	24-10-23 19:13	24-10-23 19:18	Desplaines St & Kinzie St	TA1306000003	Franklin St & Lake St	TA1307000111	41.88872	-87.64445	41.88584	-87.6355	member
6	5E34BA5DE945A9CC	classic_bike	09-10-23 18:19	09-10-23 18:30	Desplaines St & Kinzie St	TA1306000003	Franklin St & Lake St	TA1307000111	41.88872	-87.64445	41.88584	-87.6355	member
7	F7D7420AFAC53CD9	electric_bike	04-10-23 17:10	04-10-23 17:25	Orleans St & Chestnut St (NEXT Apts)	620	Sheffield Ave & Webster Ave	TA1309000033	41.89812	-87.63753	41.92154	-87.65382	member
8	870B2D4CD112D7B7	electric_bike	31-10-23 17:32	31-10-23 17:44	Orleans St & Chestnut St (NEXT Apts)	620	Sheffield Ave & Webster Ave	TA1309000033	41.89818	-87.63755	41.92154	-87.65382	member
9	D9179D36E32D456C	classic_bike	02-10-23 18:51	02-10-23 18:57	Desplaines St & Kinzie St	TA1306000003	Franklin St & Lake St	TA1307000111	41.88872	-87.64445	41.88584	-87.6355	casual
10	F8E131281F722FEF	classic_bike	17-10-23 8:28	17-10-23 8:50	Calumet Ave & 18th St	13102	Morgan St & Polk St	TA1307000130	41.85762	-87.61941	41.87174	-87.65103	member
11	91938B71748FA405	classic_bike	17-10-23 19:17	17-10-23 19:32	Wolcott Ave & Polk St	TA1309000064	Morgan St & Polk St	TA1307000130	41.87126	-87.67369	41.87174	-87.65103	member
12	191BFA255C1820FC	classic_bike	04-10-23 15:24	04-10-23 15:32	Wolcott Ave & Polk St	TA1309000064	Morgan St & Polk St	TA1307000130	41.87126	-87.67369	41.87174	-87.65103	member

6. First 4 months 2023 data was used for the analysis – Jan to April 2023
7. Next step was data cleaning.

Data cleaning process –

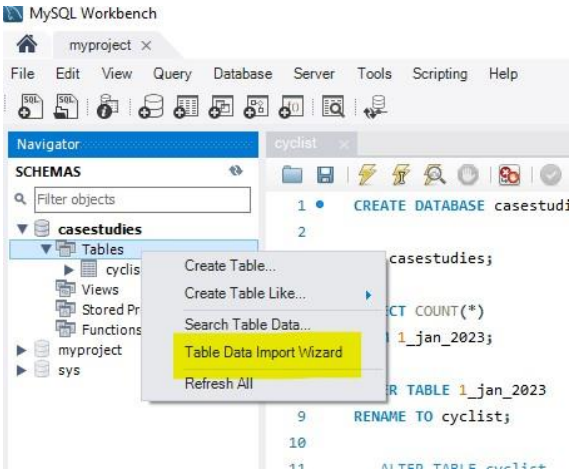
1. The columns that were used to keep the values consistent in all the spreadsheets were - rideable_type, start_at, end_at, member_casual. Total 4 columns.
2. Rideable_type had type of bicycle rides. Start_at and end_at both had time date stamp. Member_casual had the type of memberships
3. These 4 columns were used as the company was interested in understanding how the bike usage differ as per different memberships.
4. The start_at and end_at time stamps were separated in three columns each – day, time and month

	A	B	C	D	E	F	G	H	I	J	K
1	Membership Type	Bike Type	Start Time Stamp	Start Date	Start Day	Start Month	End Time Stamp	End Date	End Day	End Month	Usage Minutes
2	member	electric_bike	21-01-23 20:05	2023-01-21	Saturday	January	21-01-23 20:16	2023-01-21	Saturday	January	11
3	member	classic_bike	10-01-23 15:37	2023-01-10	Tuesday	January	10-01-23 15:46	2023-01-10	Tuesday	January	9
4	casual	electric_bike	02-01-23 7:51	2023-01-02	Monday	January	02-01-23 8:05	2023-01-02	Monday	January	-46
5	member	classic_bike	22-01-23 10:52	2023-01-22	Sunday	January	22-01-23 11:01	2023-01-22	Sunday	January	-51
6	member	classic_bike	12-01-23 13:58	2023-01-12	Thursday	January	12-01-23 14:13	2023-01-12	Thursday	January	-45
7	member	electric_bike	31-01-23 7:18	2023-01-31	Tuesday	January	31-01-23 7:21	2023-01-31	Tuesday	January	3

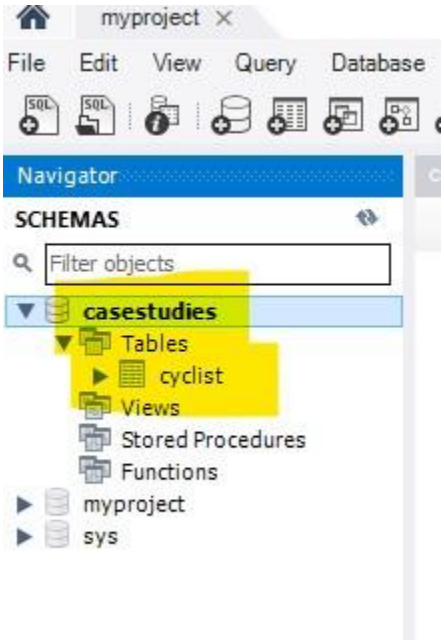
5. To separate the start date formula used was =DATE(YEAR(C2), MONTH(C2), DAY(C2))
6. To separate start day formula used was =TEXT(C2,"dddd"). Dddd gives the long format for the day where as ddd gives only first three letters of the day.
7. To separate start month formula used was =TEXT(C2,"mmmm").
8. Same was done with end time stamp.
9. Now another column was added – Usage minutes. It gave total duration of the rides.
10. A few values in usage minutes were negative. Upon understanding the data, it was seen that some start time stamps were swapped with end time stamp. They were left as it is for now.

<u>For MySQL –</u>

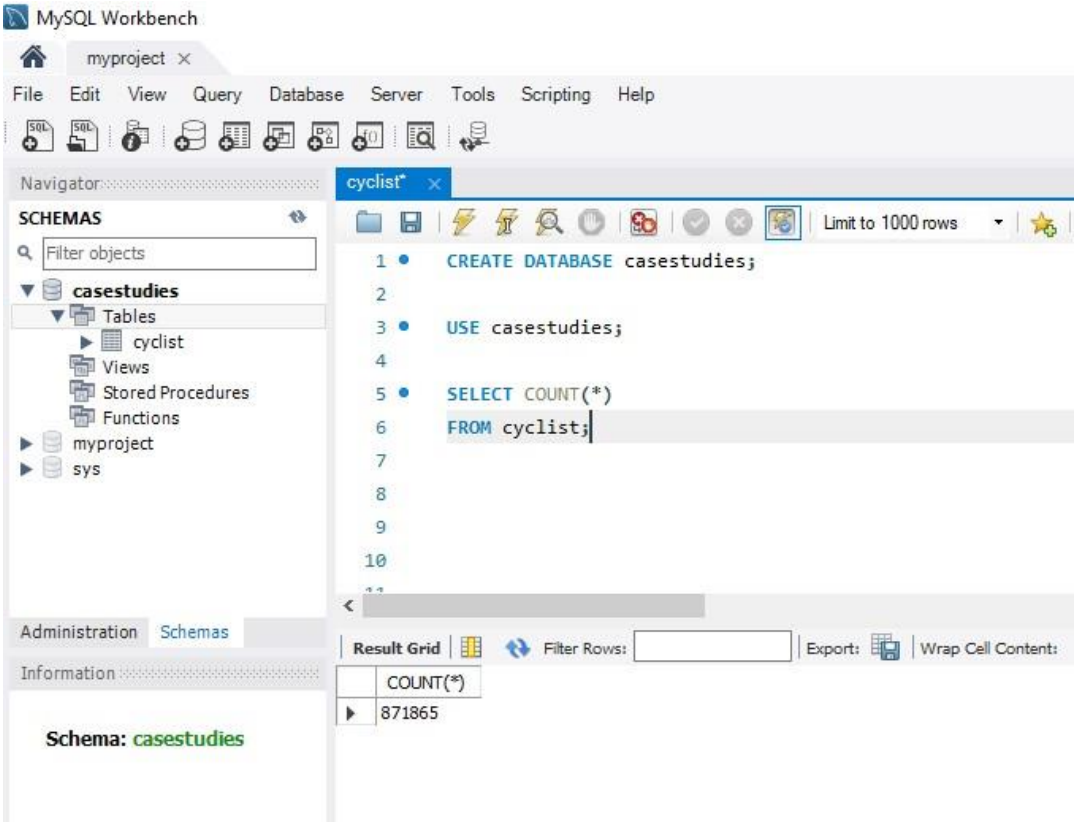
1. The 4 sheets with Jan to April 2023 data with selective columns were in xls format.
2. This format was not supported while importing data to MySQL.
3. The sheets were individually saved to csv files and then imported in MySQL using Table Data Import Wizard.



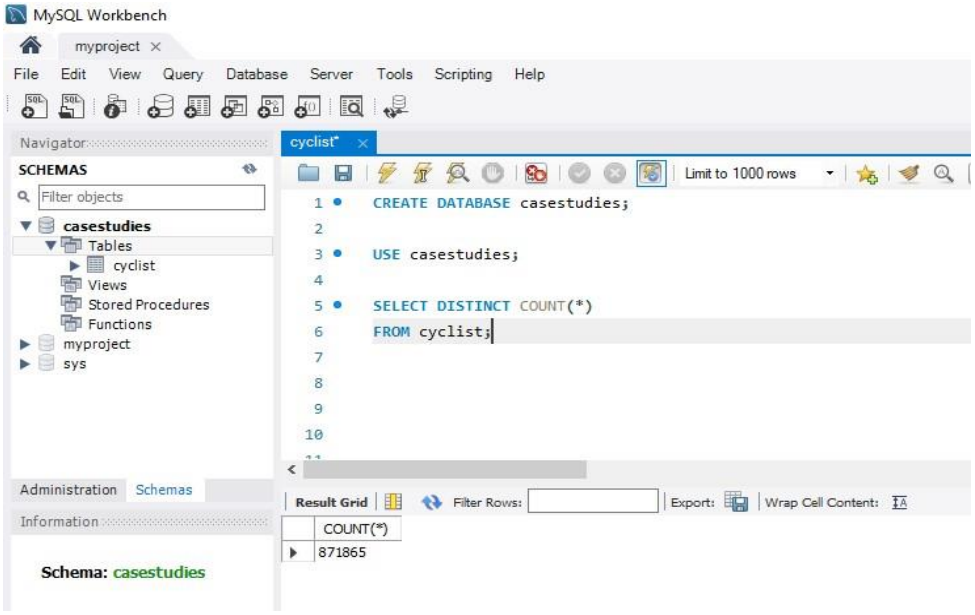
4. I created a new schema - casestudies with a new table - cyclist where in the data was imported using the Table Data Import Wizard in MySQL.



5. It took enormous time to get the data imported to MySQL likely due to my computer's system capabilities. I was successful in importing 871,865 data points in one new table that I created. Not all the data points were imported which were over 1 million.

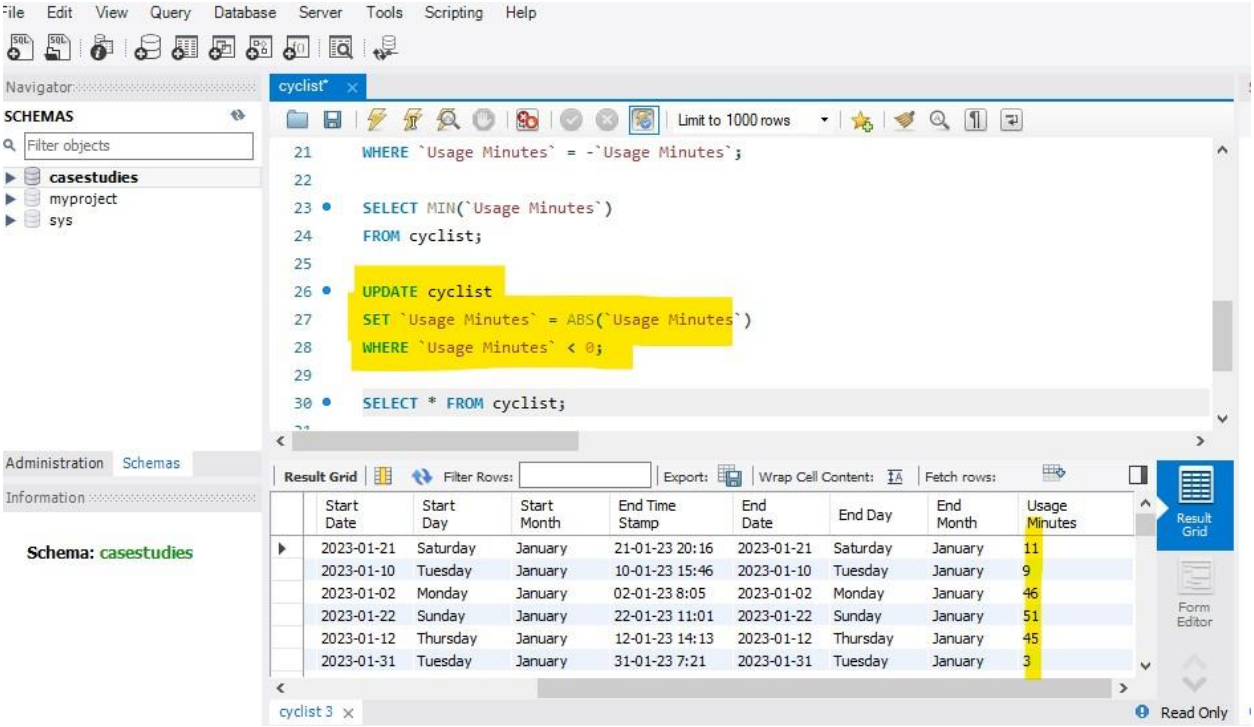


6. Distinct count was also checked to ensure that there is no duplicate data.



7. Negative entries were removed in SQL using –

```
UPDATE cyclist
        SET `Start Time Stamp` = `End Time Stamp`,
                `End Time Stamp` = `Start Time Stamp`
WHERE `Usage Minutes` < 0;
```



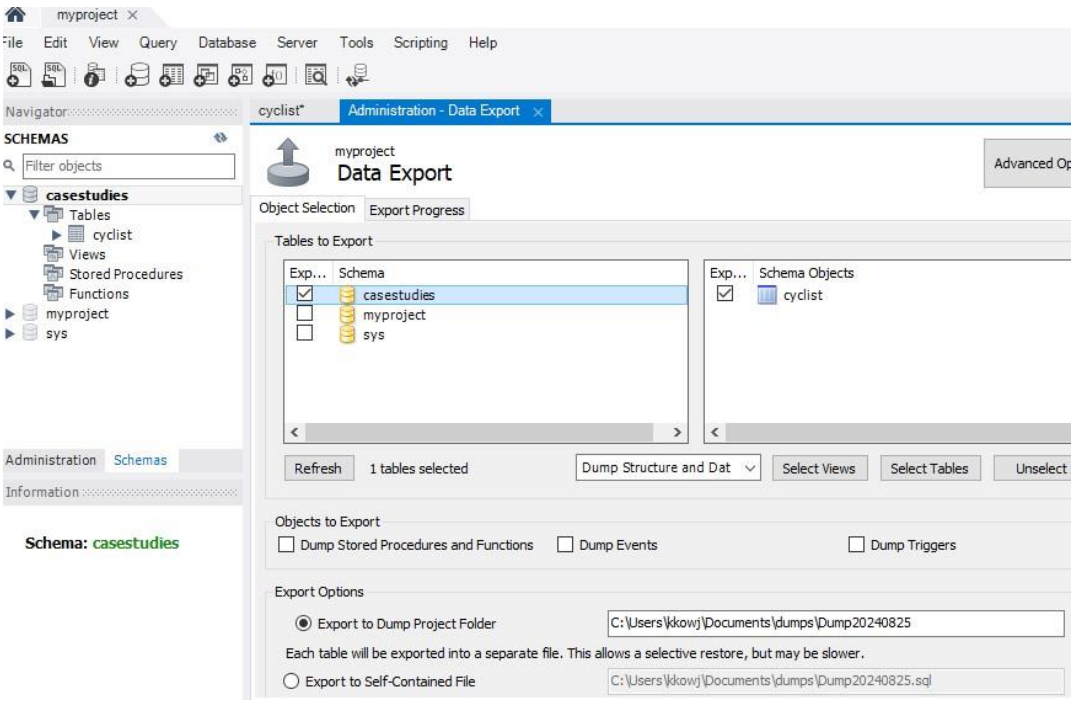
8. Clean data was now ready for further analysis.

Analyze Data –

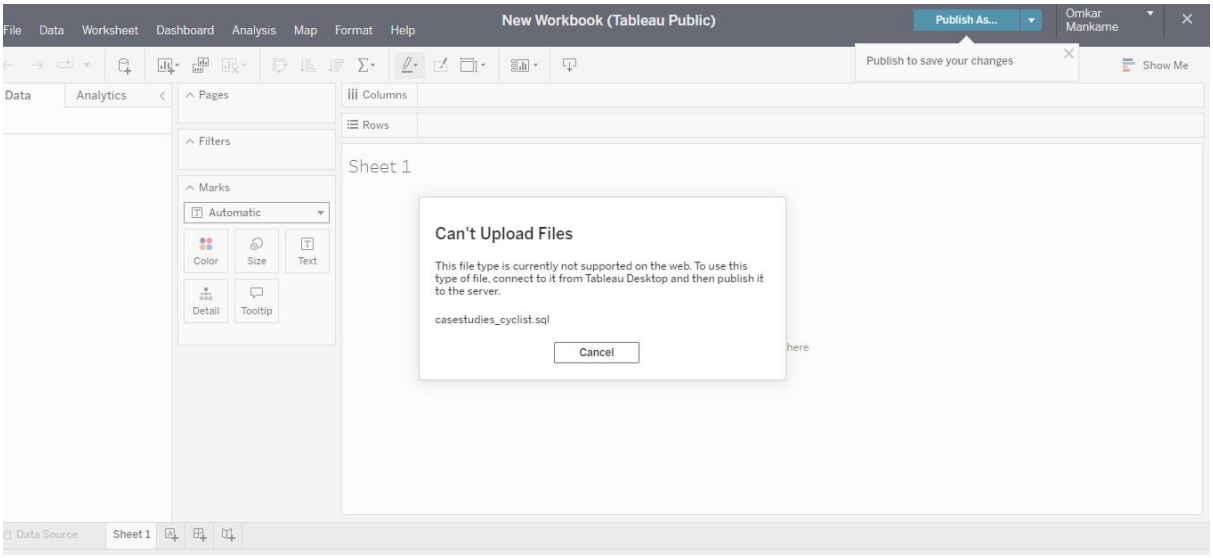
Now that reliable, original, clean, comprehensive, cited data – ROCCC data was ready, the next step was to analyze the data.

<u>For Tableau –</u>

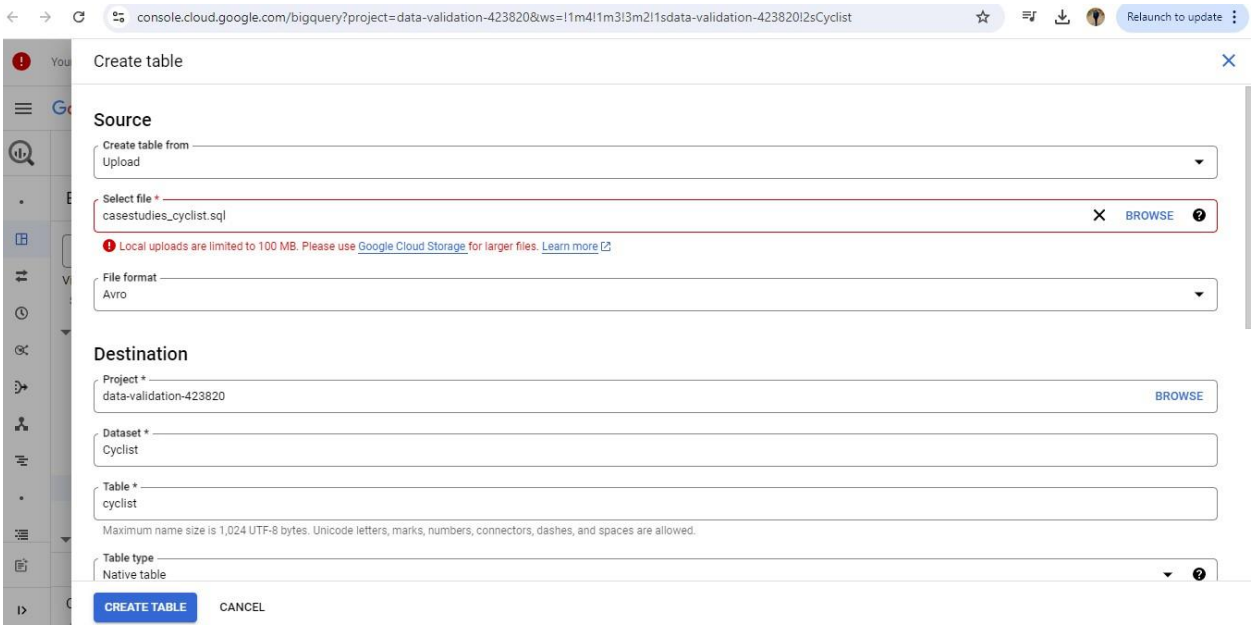
1. The clean data was first exported from MySQL.



2. The database file was 105MB which could not be uploaded to Tableau Public. Requirement is files smaller than 100MB.

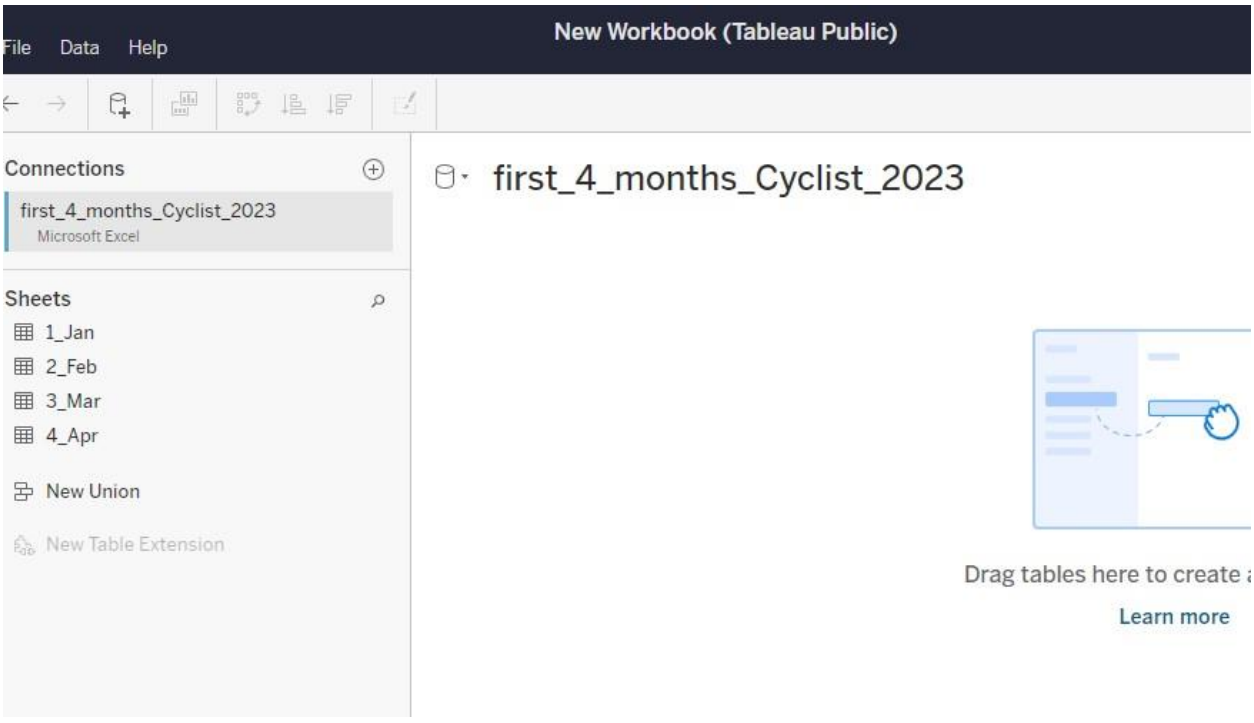


3. The next step was to find SQL platform and upload the file and then compress it to less than 100MB. Google BigQuery supports SQL file. However, again files over 100MB were not supported.

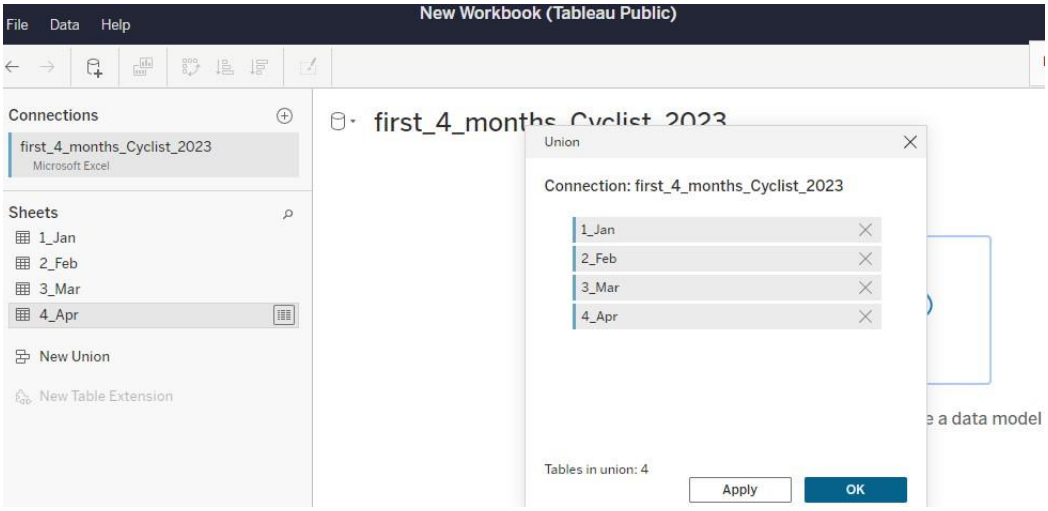


4. The then used to visualize data was done using clean spreadsheet data.

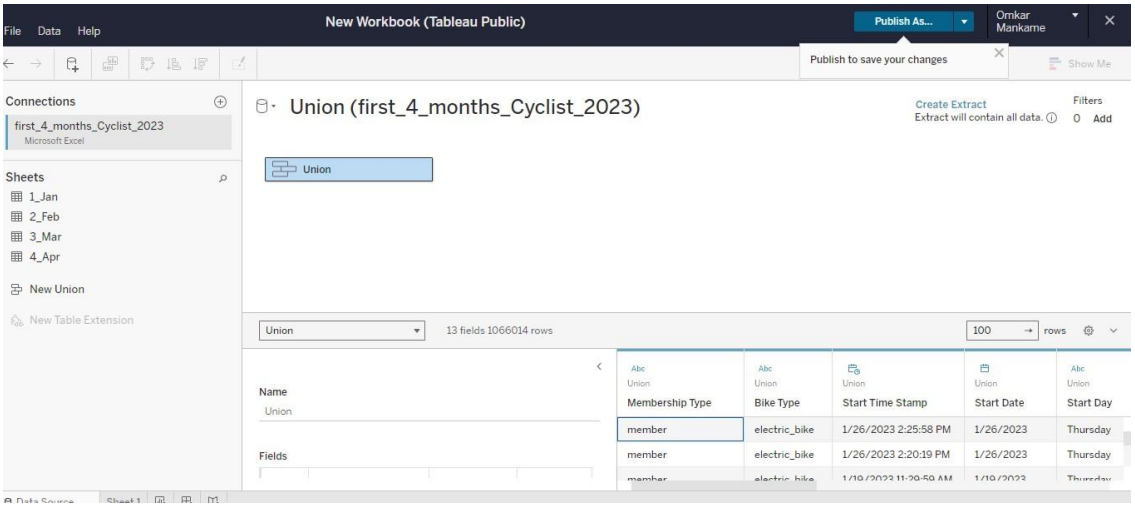
5. The workbook with all clean data for 4 months was imported to tableau.



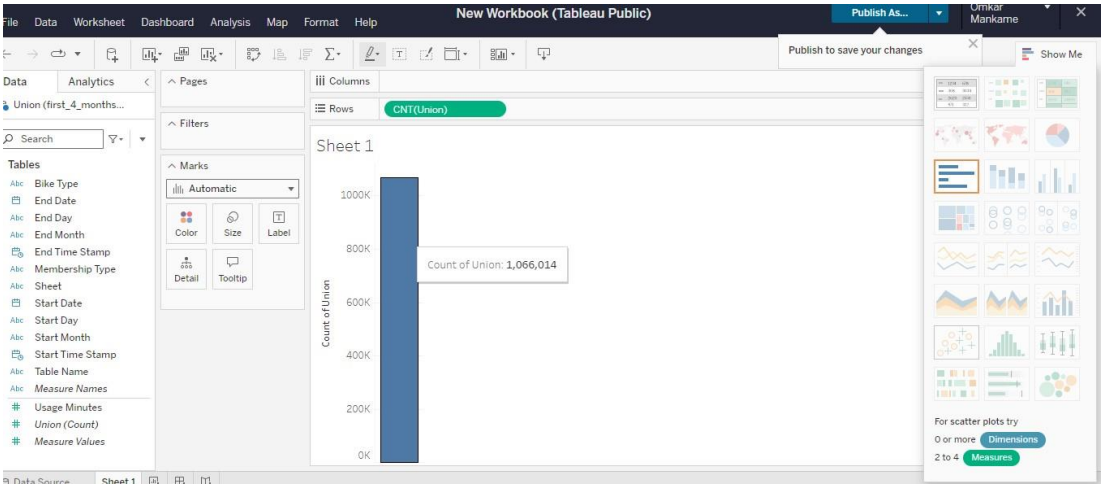
6. A new union was created with all the 4 sheets.

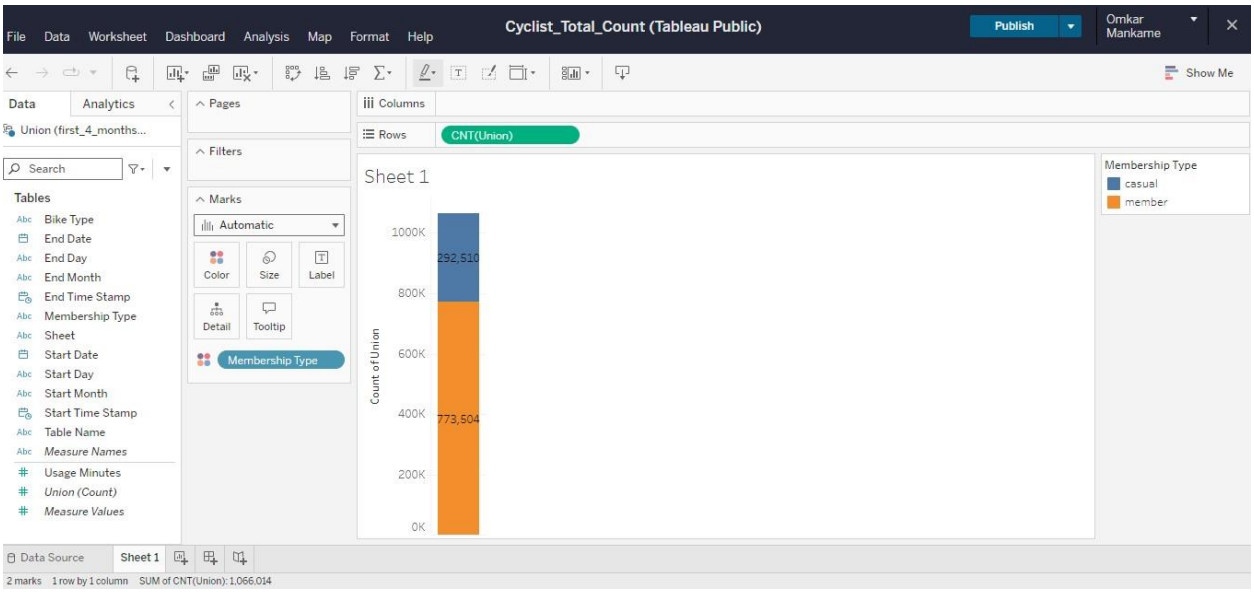


7. The union of all the data from individual worksheets is done for further analysis.

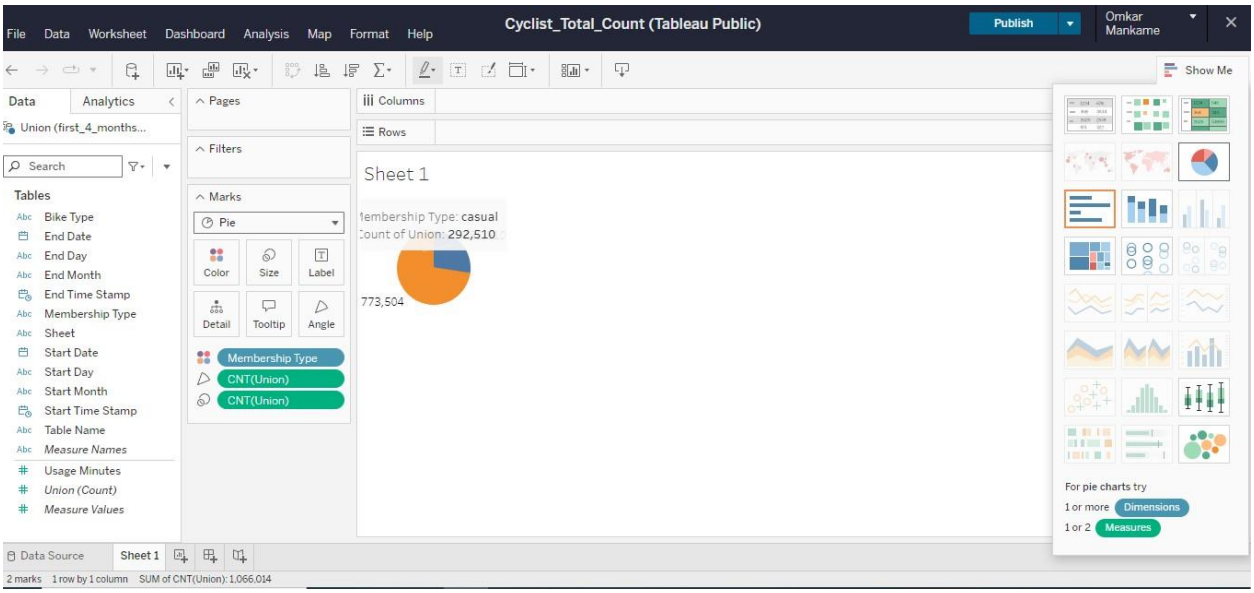


8. A new sheet is used in Tableau for further analysis.
9. The first chart that we have is the total count and the next one is the total count bifurcated by membership type.

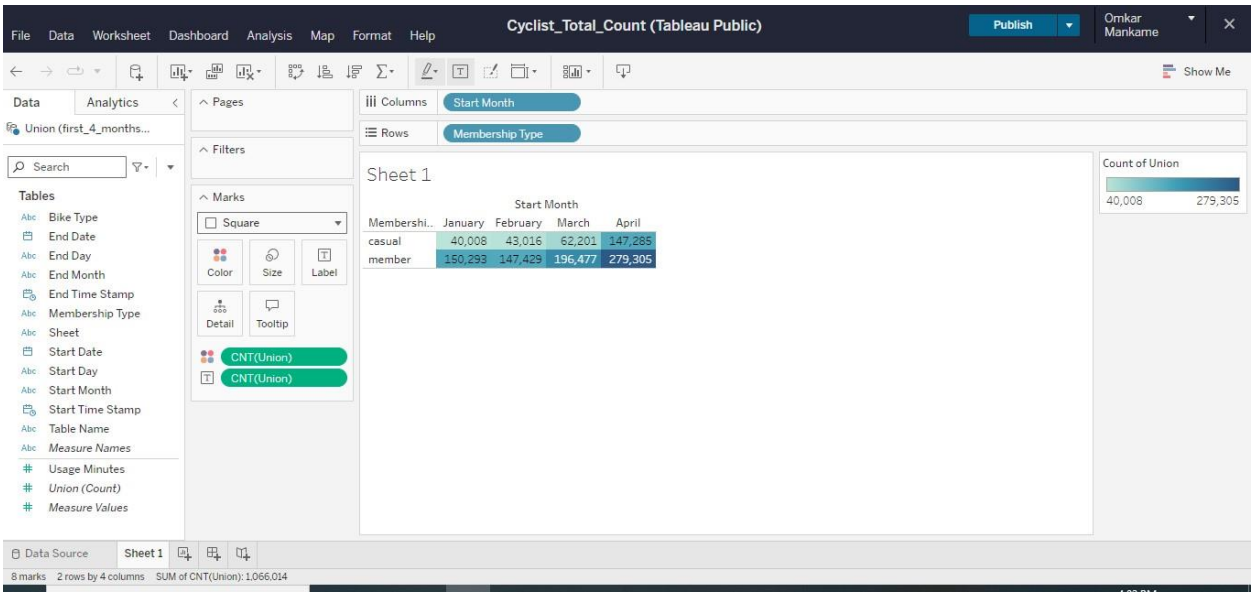




10. Now we get this information in a pie chart by using the Show Me option on the right hand top concern. This gives us a better understanding of how the membership count compares.

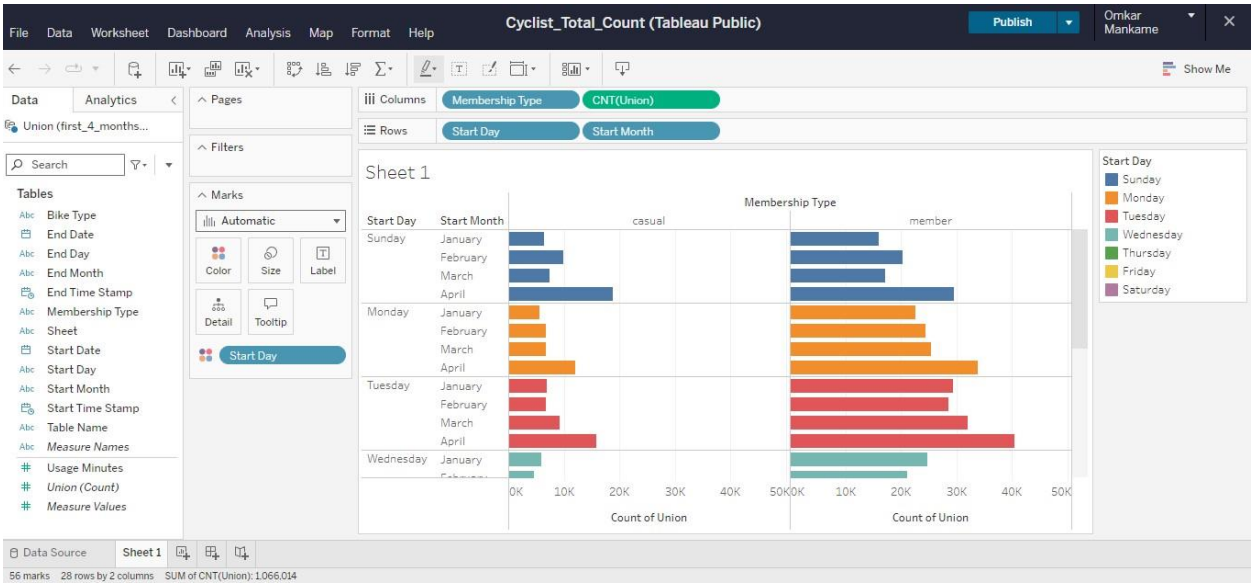


11. I have then compared the usage count in different months. It states that the annual members ride more than casual members. Also, it is evident that the casual riders increase in summer while are less as compared to winter months in Chicago.
    While the annual members are quite consistant.



12. The next few visualizations show comparisons between minute usage between the two types of members.

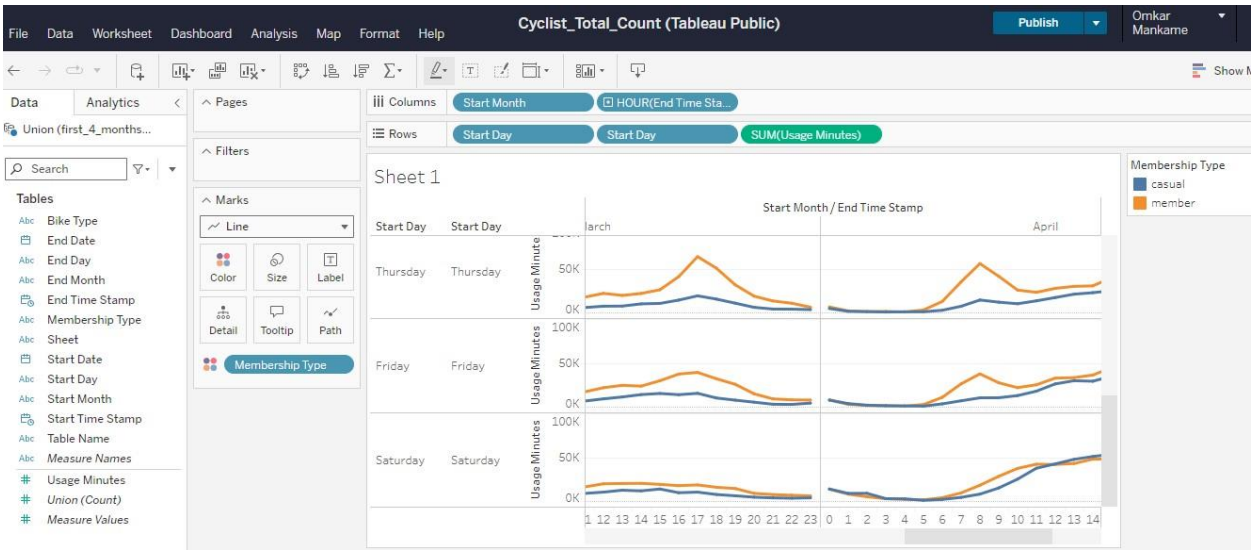
Week days comparison



Minute usage on different days

Day-Month count usage comparison

Time Usage comparison



Share results -

1. Now that we have analyzed the data. It is time to share the conclusion.
2. The data visualization with a bried description can be seen on the link below.

https://public.tableau.com/views/Cyclist_Total_Count/Sheet1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

The Analysis is that –

- There are more annual users as compared to casual users.
- The casual riders are more active in summer months in Chicago as compared to winter months.
- The usage time for casual and annual members is almost the same during the weekend
- Annual ride more on weekday while the casual ride more on the weekend
- Annual have short ride duration, casual have a bit longer duration

Result –

- It seems that the company is doing fine with more annual members. Encouraging casual riders to move to annual membership will generate more consistent revenue for the company throughout the year.
- Casual riders could be offered more discounted rides during summers to encourage more usage.
- Discounted off peak rates might encourage the riders to ride during the off peak times.