Google capstone project

Scenario

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Characters and teams

Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

Lily Moreno: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.

Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic's mission and business goals — as well as how you, as a junior data analyst, can help Cyclistic achieve them. ● Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

About the company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets

all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends

Report

According to the scenario and situations the question that keeps coming for the report of the project are

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

These three question are the question which needed the most attention and we have to find answer to solve the business decision.

Data Resources

So for that we have been already provided the data resources which is to be used for the project.

We are using previous 12 months data for analyzation and for the business decision.

The data set was located in this site - Index of bucket "divvy-tripdata"

In that we have used the the data of January 2022 to December 2022 data to answer the question

(Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this license.)

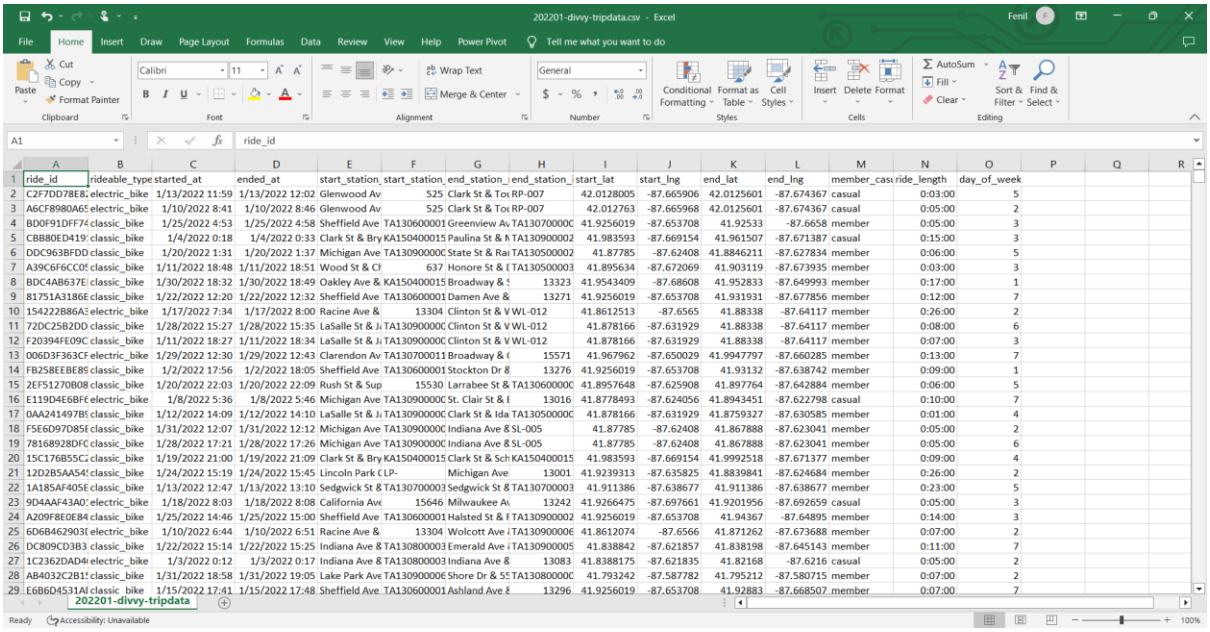
Excel

We are using excel for the data cleaning. Which is good tool for the cleaning and formatting of the data

Cleaning , Formatting and Filtering

1) We have created the two new column named ride_length and day_of week
   Ride_length : It was created for knowing how much the rider has ridded the bike
   This is formatted to hh:mm:ss for just knowing the time
   By applying formula **=(D2-C2)**

   Day_of_week : this column was created to know the day of week. On which day the rider has rided the bike most. It contains the numeric values starting from 1 to 7 in that 1 equals Sunday and 7 equals Saturday
   By applying formula **=WEEKDAY(C2,1))**



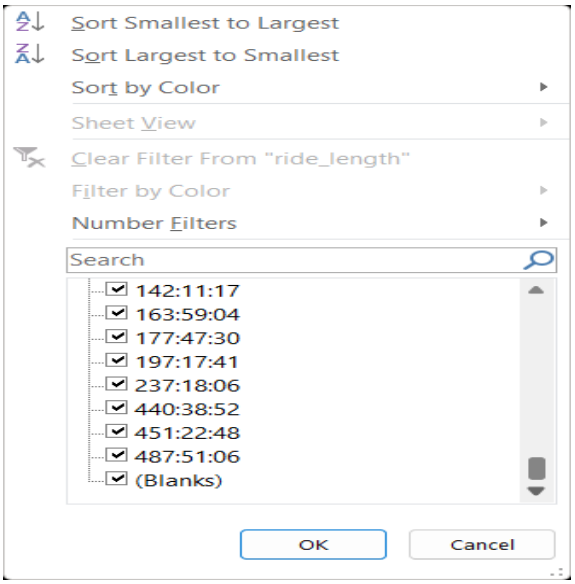
ride_id	rideable_type	started_at	ended_at	start_station	start_station	end_station_	end_station_	start_lat	start_lng	end_lat	end_lng	member_cas	ride_length	day_of_week
C2F7DD78E8	electric_bike	1/13/2022 11:59	1/13/2022 12:02	Glenwood Av	525	Clark St & To	RP-007	42.0128005	-87.665906	42.0125601	-87.674367	casual	0:03:00	5
A6CF8980A6	electric_bike	1/10/2022 8:41	1/10/2022 8:46	Glenwood Av	525	Clark St & To	RP-007	42.012763	-87.665968	42.0125601	-87.674367	casual	0:05:00	2
BD0F91DFF7	classic_bike	1/25/2022 4:53	1/25/2022 4:58	Sheffield Ave	TA130600001	Greenview Av	TA130700000	41.9256019	-87.653708	41.92533	-87.6658	member	0:05:00	3
CBB80ED419	classic_bike	1/4/2022 0:18	1/4/2022 0:33	Clark St & Bry	KA150400015	Paulina St & M	TA130900002	41.983593	-87.669154	41.961507	-87.671387	casual	0:15:00	3
DDC963BFDD	classic_bike	1/20/2022 1:31	1/20/2022 1:37	Michigan Ave	TA130900000	State St & Ra	TA130500002	41.87785	-87.62408	41.8846211	-87.627834	member	0:06:00	5
A39C6F6CC0	classic_bike	1/11/2022 18:48	1/11/2022 18:51	Wood St & Ch	637	Honore St & I	TA130500003	41.895634	-87.672069	41.903119	-87.673935	member	0:03:00	3
BDC4AB637E	classic_bike	1/30/2022 18:32	1/30/2022 18:49	Oakley Ave &	KA150400015	Broadway & S	13323	41.9543409	-87.68608	41.952833	-87.649993	member	0:17:00	1
81751A3186	classic_bike	1/22/2022 12:20	1/22/2022 12:32	Sheffield Ave	TA130600001	Damen Ave &	13271	41.9256019	-87.653708	41.931931	-87.677856	member	0:12:00	7
154222B86A	electric_bike	1/17/2022 7:34	1/17/2022 8:00	Racine Ave &	13304	Clinton St & W	L-012	41.8612513	-87.6565	41.88338	-87.64117	member	0:26:00	2
72DC25B2DD	classic_bike	1/28/2022 15:27	1/28/2022 15:35	LaSalle St & J	TA130900000	Clinton St & W	L-012	41.878166	-87.631929	41.88338	-87.64117	member	0:08:00	6
F20394FE09	classic_bike	1/11/2022 18:27	1/11/2022 18:34	LaSalle St & J	TA130900000	Clinton St & W	L-012	41.878166	-87.631929	41.88338	-87.64117	member	0:07:00	3
006D3F363C	electric_bike	1/29/2022 12:30	1/29/2022 12:43	Clarendon Av	TA130700011	Broadway & S	15571	41.967962	-87.650029	41.9947797	-87.660285	member	0:13:00	7
FB258EEBE8	classic_bike	1/2/2022 17:56	1/2/2022 18:05	Sheffield Ave	TA130600001	Stockton Dr &	13276	41.9256019	-87.653708	41.93132	-87.638742	member	0:09:00	1
2EF51270B0	classic_bike	1/20/2022 22:03	1/20/2022 22:09	Rush St & Sup	15530	Larrabee St &	TA130600000	41.8957648	-87.625908	41.897764	-87.642884	member	0:06:00	5
E119D4E6BF	electric_bike	1/8/2022 5:36	1/8/2022 5:46	Michigan Ave	TA130900000	St. Clair St & E	13016	41.8778493	-87.624056	41.8943451	-87.622798	casual	0:10:00	7
0AA241497B	classic_bike	1/12/2022 14:09	1/12/2022 14:10	LaSalle St & J	TA130900000	Clark St & Ida	TA130500000	41.878166	-87.631929	41.8759327	-87.630585	member	0:01:00	4
F5E6D97D85	classic_bike	1/31/2022 12:07	1/31/2022 12:12	Michigan Ave	TA130900000	Indiana Ave &	SL-005	41.87785	-87.62408	41.867888	-87.623041	member	0:05:00	2
78168928DF	classic_bike	1/28/2022 17:21	1/28/2022 17:26	Michigan Ave	TA130900000	Indiana Ave &	SL-005	41.87785	-87.62408	41.867888	-87.623041	member	0:05:00	6
15C176B55C	classic_bike	1/19/2022 21:00	1/19/2022 21:09	Clark St & Bry	KA150400015	Clark St & Sch	KA150400015	41.983593	-87.669154	41.9992518	-87.671377	member	0:09:00	4
12D2B5AA54	classic_bike	1/24/2022 15:19	1/24/2022 15:45	Lincoln Park (	LP-	Michigan Ave	13001	41.9239313	-87.635825	41.8839841	-87.624684	member	0:26:00	2
1A185AF405	classic_bike	1/13/2022 12:47	1/13/2022 13:10	Sedgwick St &	TA130700003	Sedgwick St &	TA130700003	41.911386	-87.638677	41.911386	-87.638677	member	0:23:00	5
9D4AAF43A0	electric_bike	1/18/2022 8:03	1/18/2022 8:08	California Ave	15646	Milwaukee Av	13242	41.9266475	-87.697661	41.9201956	-87.692659	casual	0:05:00	3
A209F8E0E8	classic_bike	1/25/2022 14:46	1/25/2022 15:00	Sheffield Ave	TA130600001	Halsted St & F	TA130900002	41.9256019	-87.653708	41.94367	-87.64895	member	0:14:00	3
6D6B462903	electric_bike	1/10/2022 6:44	1/10/2022 6:51	Racine Ave &	13304	Wolcott Ave (	TA130900006	41.8612074	-87.6566	41.871262	-87.673688	member	0:07:00	2
DC809CD3B3	classic_bike	1/22/2022 15:14	1/22/2022 15:25	Indiana Ave &	TA130800003	Emerald Ave (	TA130900005	41.838842	-87.621857	41.838198	-87.645143	member	0:11:00	7
1C2362DAD4	classic_bike	1/3/2022 0:12	1/3/2022 0:17	Indiana Ave &	TA130800003	Indiana Ave &	13083	41.8388175	-87.621835	41.82168	-87.6216	casual	0:05:00	2
AB4032C2B1	classic_bike	1/31/2022 18:58	1/31/2022 19:05	Lake Park Ave	TA130900006	Shore Dr & 55	TA130800000	41.793242	-87.587782	41.795212	-87.580715	member	0:07:00	2
E6B6D4531A	classic_bike	1/15/2022 17:41	1/15/2022 17:48	Sheffield Ave	TA130600001	Ashland Ave &	13296	41.9256019	-87.653708	41.92883	-87.668507	member	0:07:00	7

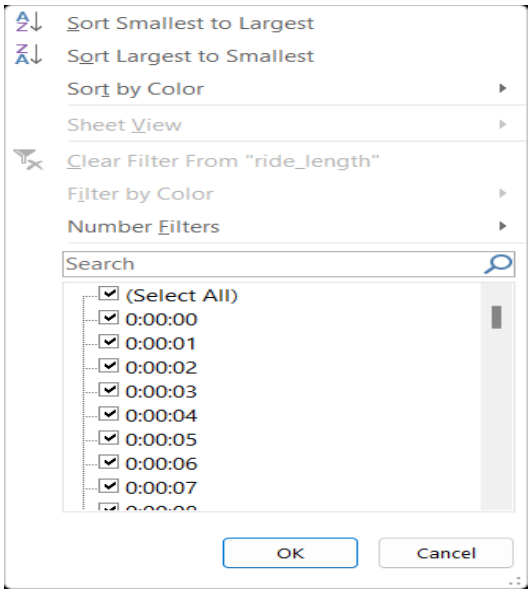
As you can see above the two columns are created and other data which is given in the dataset for analyzing it

2)Changed the format of ride_length to hh:mm:ss for easier calculation

3) I have removed the row which were containing the value 00:00:00 and more than 23:00:00 in ride length Which helped us to reduce some waste data.

Reason for removing the data : The data which may be entered the by mistake or was some error in the system which created this data. We only need the data for the ride which is done on the same day note more than that





Calculation

CALCULATED **MINIMUM MAXIMUM SUM AND AVERAGE OF THE TOTAL RIDES**

sum of ride
21857:57:00

average ride length
0:12:47
maximum ride length
23:00:00
minimum ride length
0:01:00

PIVOT TABLE

CREATED VARIOUS PIVOT TABLE AND CALCULATED MANY QUESTIONS WITH EACH FIELD

1) **THIS PIVOT TABLE SHOWS THE TOTAL NUMBER OF MEMBER AND CASUAL USERS OF THE JANUARY DATASET**

1		
2		
3	**Row Labels**	**Count of ride_length**
4	member	85250
5	casual	18520
6	**Grand Total**	**103770**
7		
8		
9		
10		

2) The pivot table with total number of rides each day of week

Row Labels	Count of ride_length
1	11509
2	15803
3	16149
4	15174
5	16554
6	13811
7	14770
Grand Total	**103770**

3) This pivot tbale consist of daysof week tyoes of ride and total number of ride length by each one

Count of ride_length	Column Labels							
Row Labels	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Grand Total**
classic_bike	5810	8281	8472	8450	8970	7585	7499	55067
docked_bike	162	95	98	125	107	114	260	961
electric_bike	5537	7427	7579	6599	7477	6112	7011	47742
Grand Total	**11509**	**15803**	**16149**	**15174**	**16554**	**13811**	**14770**	**103770**

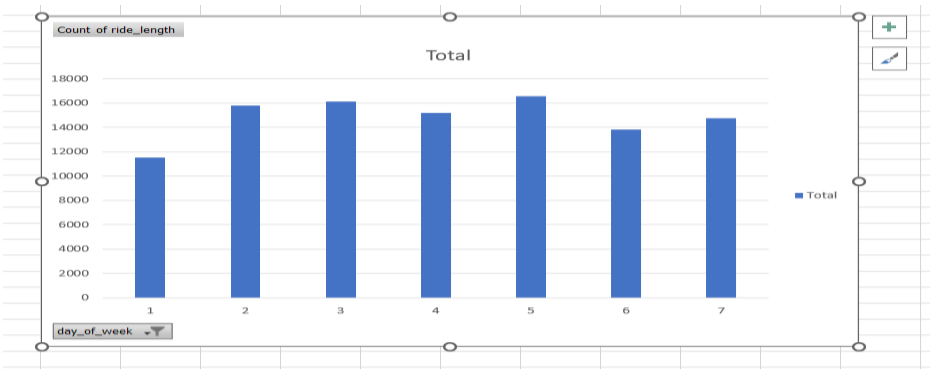
4) This pivot tbale shows the busiest time and ride length on that time period

Count of ride_length	Column Labels																								
Row Labels	**12 AM**	**1 AM**	**2 AM**	**3 AM**	**4 AM**	**5 AM**	**6 AM**	**7 AM**	**8 AM**	**9 AM**	**10 AM**	**11 AM**	**12 PM**	**1 PM**	**2 PM**	**3 PM**	**4 PM**	**5 PM**	**6 PM**	**7 PM**	**8 PM**	**9 PM**	**10 PM**	**1**
1	11320	104	40	13	7	1	2	6	2	1	2	1		1	1	1	1			2	1		3	
2	15606	114	41	14	2	6	1	2		5	3			1		1	1	1	1		1	1		
3	16005	76	26	10	5	1	1	2	2		4	3	1	1	2	4	1	3	1	1				
4	15021	90	17	11	7	4	2	1	2	1		1	2		4	1		2	4	1	1		1	
5	16369	101	31	15	9	4	3	1	4	2	1	1			5		1		1	2	2	1	1	
6	13641	96	32	12	5	2			2		2				5	2	6	2	3				1	
7	14558	139	44	6	3	1		4	1	2	2	2	1		1		1	2	2	1				
Grand Total	**102520**	**720**	**231**	**81**	**38**	**19**	**9**	**16**	**13**	**11**	**14**	**8**	**4**	**3**	**18**	**9**	**11**	**10**	**12**	**7**	**5**	**2**	**6**	

This all the pivot table are created for the future analyses and for the summary of the data which will help us to find the final conclusion on the data and the scenarios given.

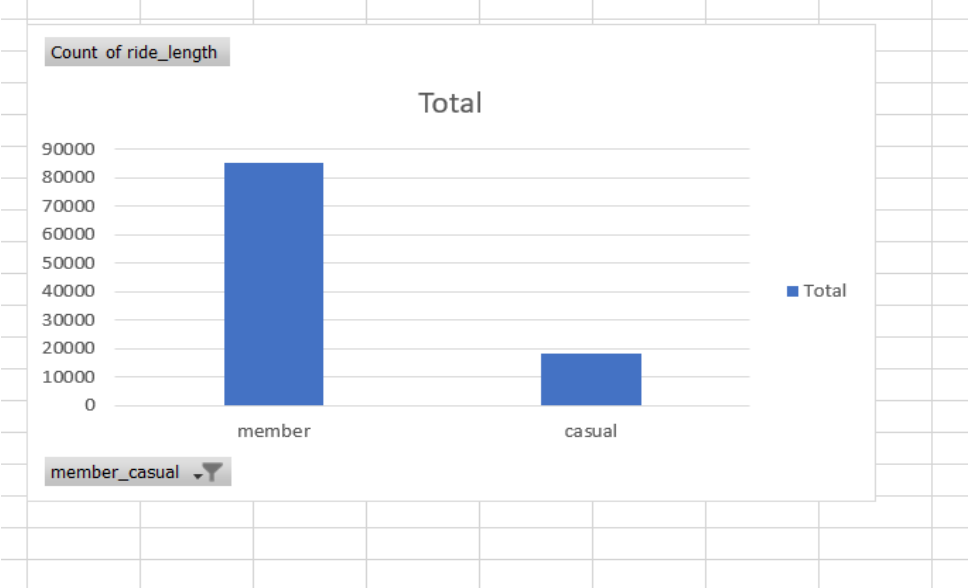
Charts

1) This chart consist of days of week types of ride and total number of ride length by each o



2) This chart consist of total number of rides each day of week

3) **THIS CHART SHOWS THE TOTAL NUMBER OF MEMBER AND CASUAL USERS OF THE JANUARY DATASET**



THIS WHOLE PROCESS OF EXCEL IS REPEATED FOR ALL THE DATASET FROM FEBRUARY TO DECEMBER FOR THE FURTHER ANALYSES

Merging the data

After all the sorting and filtering we have merged all the data set into one which consists of **row more than 6 million** roughly around **(6426040)** and 8 different column field.

Python

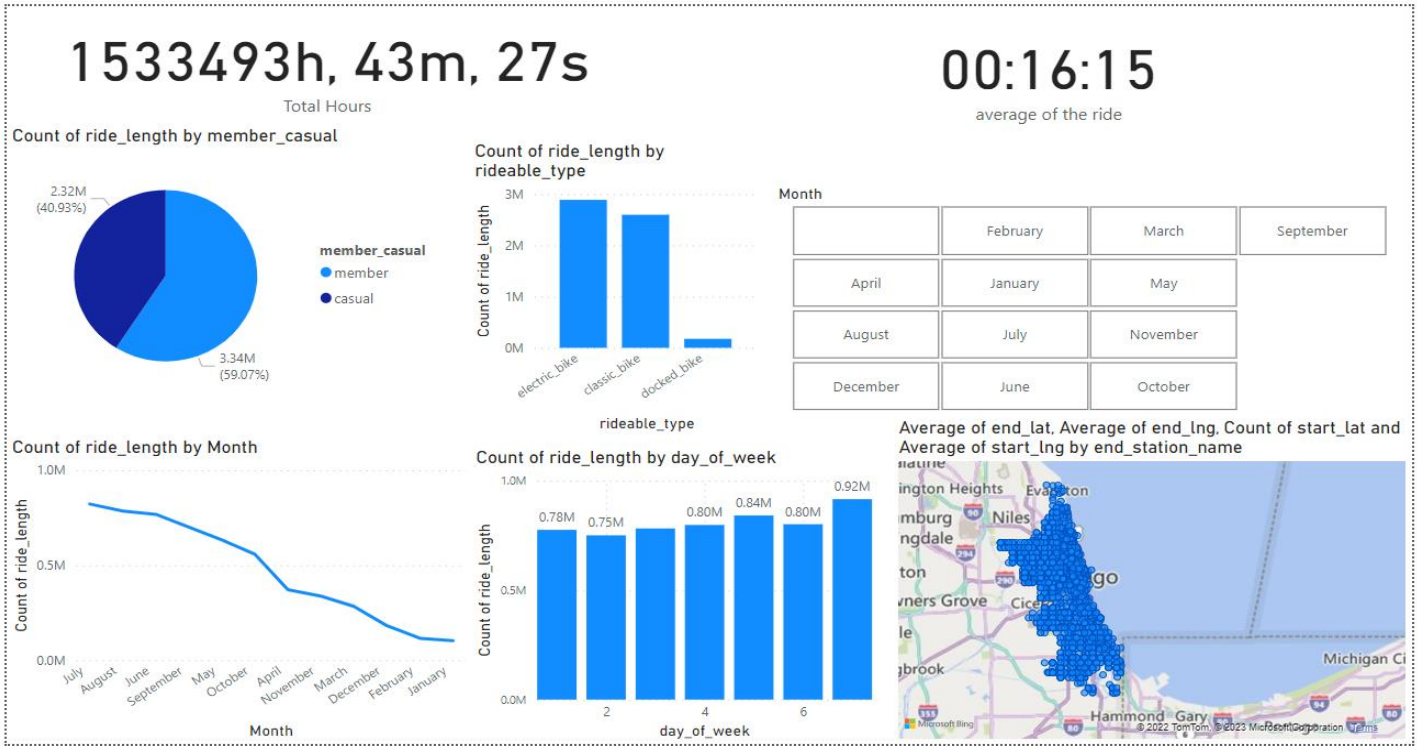
Python program for rechecking the the data set and confirming the values that everything is correct before visualizing

There's an separate python file for the codes and that checked various things for analyzation

Visualization

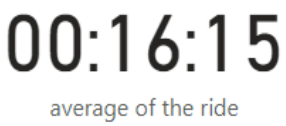
We have used the Power bi for the visualization.

As you can see above is the visual I have created for the year 2022 trip data for the cyclistic company and we can see the clear visual about the data.

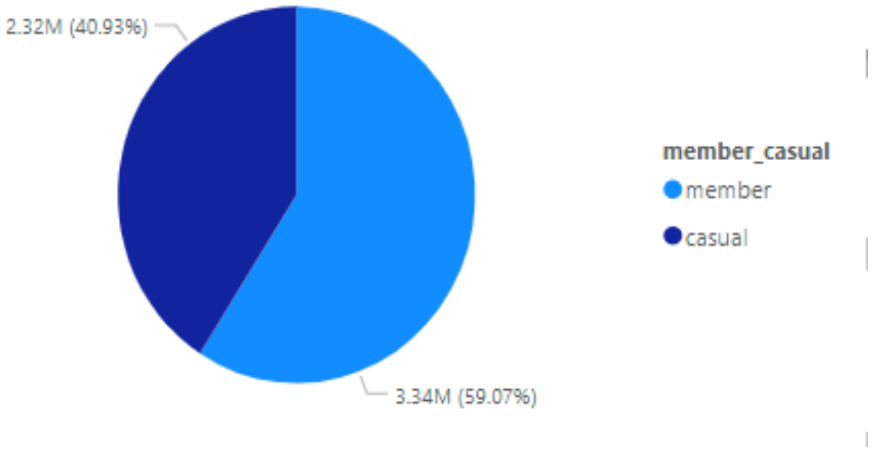


You can see that total hours of ride which was done in whole year.

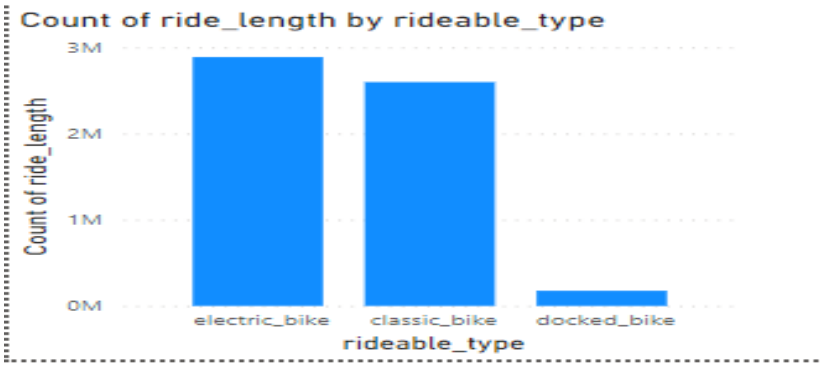
I have also added average of the total ride in the visualization



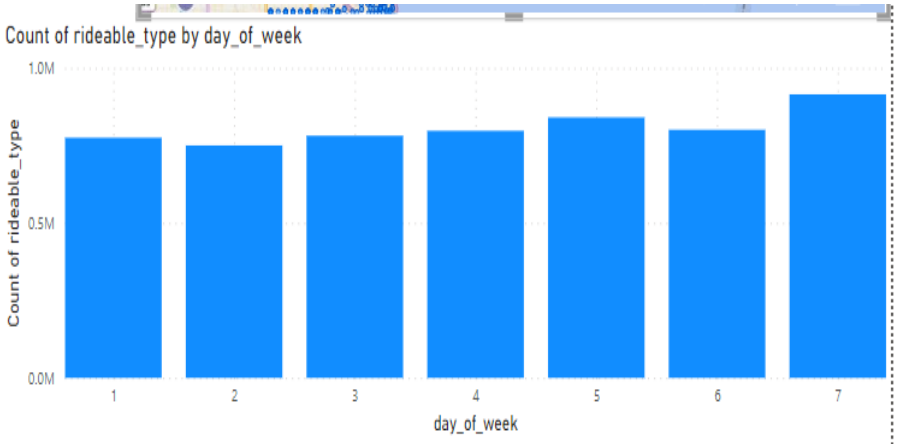
We can see that total number of member and casual people for the company. We can see that there are 40.93% of casual riders. Which is note the small number which to be ignored, it can bea huge beneficial for the company to convert them into a members and make them permanent customers



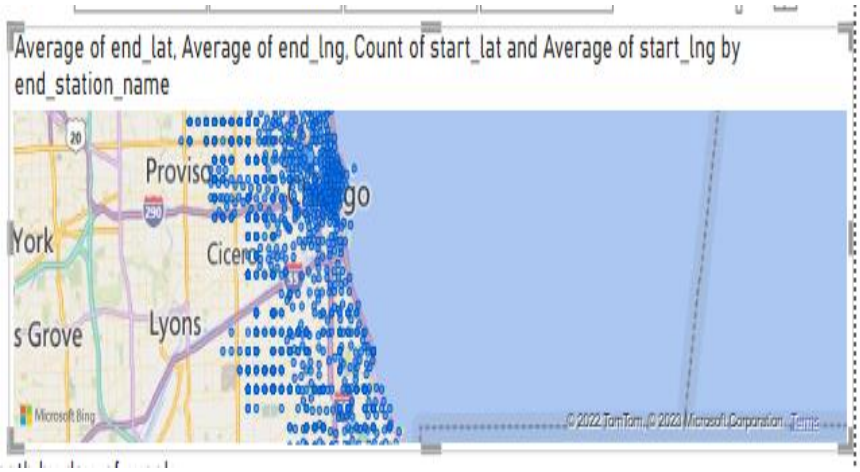
As you can see in the data there three major types of ride or bikes are available by the company and there over all use by the users. Its an over all use of the bikes by the users through out the users which we can see that electric and classic bikes are more in demand then docked bike the company can increase there supply for more profit and use the marketing campaign for the bikes to be used.

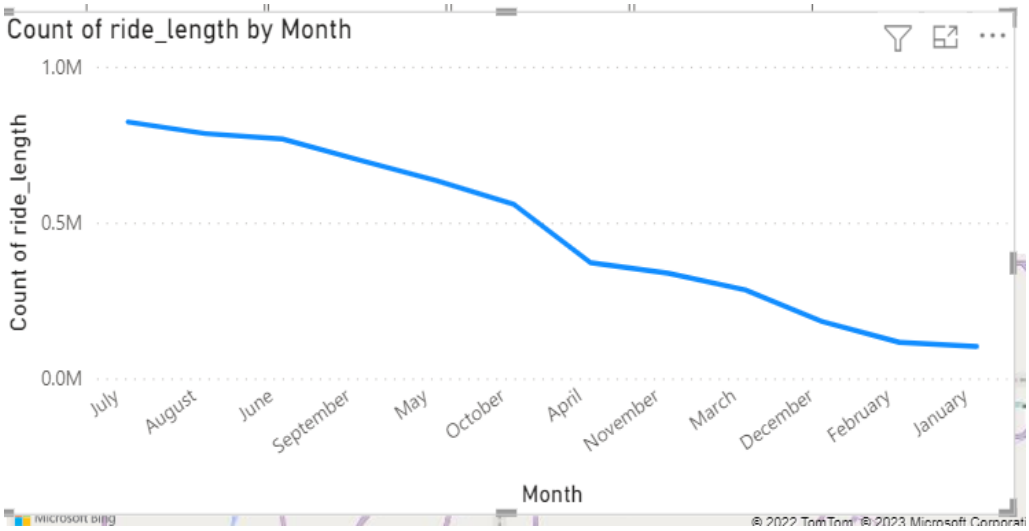


Now we look at the use of bike in all over week. Here you can assume that 1 = Sunday and 7 = Saturday so that you don't confuse by the numbers here we can see that over all performance through out the week was almost equal we can't see the busiest day for the use of bikes but the maximum use of the bike was on Saturday with busiest day in the week.



These are the most end station in the chicago here the data set is focusing on the particual tarea or sate in the nation as we can see in the map visualization the most comman endstation was near the michigan lake area which can be said that the users uses bikes more for ride to the lake helps us to know that it can be profitable for doing the business near lake area for there convinence and profit in the business.





We can see here the most number of ride or ride length is done on July month and with very least rides we can see that January is the month with very least number of ride length . we can here say that users prefer more bikes on the **summer season** then on

winter season cause due to the heat they prefer to go on the bike then on the car or any other vehicle or ride like metro or bus . here we can use this opportunity to make the business more profitable by giving discounts or more benefits to the members so that casual member by the benefits they will buy the subscription plan which is the main goal of the company.

Conclusion and Solutions :

Conclusion : Company needs to focus on the winter season on that time the casual users are very low so to increase the members users we need to plan more genuine pricing system on that particular season cause the member are also using less bikes on the winter season.

Solutions

By seeing the visualization above I can give the answers based on the data result

1) Based on the season we can plan or make different subscription according to the season that can be helpful fo the members and the users we can see that the total riders and ride length both are very low on the winter season we can plan a budget friendly plan that can casual people can buy and use that plan or offers on the winter season

2) There are many end station near the Michigan lake we can provide other external services or marketing campaign other then the lake so that more and more people know about the company and there plans and offers .we can also collab with the local shops or the market people which gives us the more customer.

3) There are more users with electric and casual bike type then the docked bike so we can focus more on electric and classic bike and make it more available to the users so that due to the busiest season or the busiest hour we can supply our product to the users without any supply chain problem to the users.

4) They can plan more membership or offers with plan by making collab with the local store on the winter season and the spring season

On the basis of the conclusion the stake holders and the team leader can see the report and apply the solutions based on the data.

Data needed

I needed the data on pricing of the membership so that more insights can be generated that can answer more question and bring more solution for that company.

Thank you for your time hope you have liked the report and the solution.