

CYCLISTIC BIKE-SHARE ANALYSIS

(MICROSOFT POWER BI)

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.



ASK

Three questions will guide the future marketing program:

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?

BUSINESS TASK

Analyze historical data to determine how annual and casual riders use Cyclistic bikes differently.

KEY STAKEHOLDERS:

- **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.

- **Cyclistic executive team:** The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

PREPARE

Dataset source:

This is public data that you can use to explore how different customer types are using Cyclistic bikes. But note that data-privacy issues prohibit you from using riders' personally identifiable information. This means that you won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes. [Link to Data Source](#)

Lisence:The data has been made available by Motivate International Inc. under this license: [Link to Lisence](#)

Size of data: rows: 5 779 438; **columns:** 13

Organization data: year month-divvy-tripdata.zip

Metadata:

- ride_id Primary Key String Not Null
- rideable_type String enum ("electric_bike", "classic_bike", "docked_bike") Not Null
- start_station_id String Nullable start station id
- start_station_name String Nullable start station name
- started_at Timestamp Not Null start date and time
- start_lat Float Nullable trip start latitude
- start_lng Float Nullable trip start longitude
- end_station_id String Nullable end station id
- end_station_name String Nullable end station name
- ended_at Timestamp Not Null end date and time
- end_lat Float Nullable end latitude
- end_lng Float Nullable end longitude
- member_casual String enum("casual", "member") Not Null

The dataset contains information about 12 months without missing information:

No	YEAR	MONTH	ROWS
1	2022	July	823 487
2	2022	August	785 931
3	2022	September	701 338

4	2022	October	558 684
5	2022	November	337 735
6	2022	December	181 806
7	2023	January	190 301
8	2023	February	190 445
9	2023	March	258 678
10	2023	April	426 590
11	2023	May	604 826
12	2023	June	719 617

Query:

```
select year, month, rows
from
(
select extract(year from started_at) year, extract(month from started_at) month, count(1)
rows
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
group by extract(year from started_at),extract(month from started_at)
)
order by year, month;
```

Data ROCCC:

- Reliability. The data has been made available by Motivate International Inc. under this license [Link to Lisence](#).

It can be used to explore how different customer types are using Cyclistic bikes.

- Originality. The dataset is a first party source.
- Comprehensiveness. For the purposes of this case study, the dataset is appropriate and will enable you to answer the business questions.
- Current. Dataset contains information for the last 12 months.
- Cited. The Google Data Analytics certification cites the Cyclistic bikes dataset.

PROCESS

- Downloaded the previous 12 months of Cyclistic trip data [Link to Data Source](#)
 Uploaded:
 202207-divvy-tripdata.zip
 202208-divvy-tripdata.zip
 202209-divvy-tripdata.zip
 202210-divvy-tripdata.zip
 202211-divvy-tripdata.zip
 202212-divvy-tripdata.zip
 202301-divvy-tripdata.zip
 202302-divvy-tripdata.zip
 202303-divvy-tripdata.zip
 202304-divvy-tripdata.zip
 202305-divvy-tripdata.zip
 202306-divvy-tripdata.zip
- Created "csv" subfolder for the .csv file and uploaded into Microsoft Power BI
Home > New Source > Folder > Create Table "FACT_CYCLISIC"

Folder

Folder path

D:\PowerBI_projects\cyclistic\csv

Browse...

OK

Cancel

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes
Binary	202207-divvy-tripdata.csv	.csv	02.10.2023 13:23:59	05.08.2022 12:31:56	11.08.2023 16:11:25	Record
Binary	202208-divvy-tripdata.csv	.csv	02.10.2023 13:23:57	08.09.2022 09:56:24	11.08.2023 16:11:25	Record
Binary	202209-divvy-publictripdata.csv	.csv	02.10.2023 13:23:56	06.10.2022 13:18:42	11.08.2023 16:11:25	Record
Binary	202210-divvy-tripdata.csv	.csv	02.10.2023 13:23:49	08.11.2022 07:40:48	11.08.2023 16:11:25	Record
Binary	202211-divvy-tripdata.csv	.csv	02.10.2023 13:23:37	02.12.2022 10:09:56	11.08.2023 16:11:26	Record
Binary	202212-divvy-tripdata.csv	.csv	02.10.2023 13:23:28	03.01.2023 07:27:02	11.08.2023 16:11:26	Record
Binary	202301-divvy-tripdata.csv	.csv	02.10.2023 13:23:28	07.02.2023 08:36:46	11.08.2023 16:11:26	Record
Binary	202302-divvy-tripdata.csv	.csv	02.10.2023 13:23:28	07.03.2023 10:54:50	11.08.2023 16:11:26	Record
Binary	202303-divvy-tripdata.csv	.csv	02.10.2023 13:23:43	06.04.2023 06:38:25	11.08.2023 16:11:26	Record
Binary	202304-divvy-tripdata.csv	.csv	02.10.2023 13:23:52	04.05.2023 11:51:15	11.08.2023 16:11:26	Record
Binary	202305-divvy-tripdata.csv	.csv	02.10.2023 13:24:00	08.06.2023 14:38:23	11.08.2023 16:11:26	Record
Binary	202306-divvy-tripdata.csv	.csv	02.10.2023 13:24:06	12.07.2023 21:05:53	11.08.2023 16:11:26	Record

- What tools are you choosing and why?
I used Microsoft Power BI because it is a popular BI tool that provides the ability to transform and visualize a large amount of data.
- Clean Data:
 - Checked data for duplicates:

```
select ride_id
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
group by ride_id
having count(ride_id) > 1;
```
 - Excluded from analyze values where started_at = ended_at

```
select count(1)
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
where started_at = ended_at;
```
 - Swapped started_at and ended_at in case started_at > ended_at

```
select count(1)
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
where started_at > ended_at;
```
 - Cleared asterisk in the end start_station_name and end_station_name

```
select count(1)
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
```

where start_station_name like "%*"
or end_station_name like "%*";

5. Checked error coordinates (these coordinates not available in Chicago)

```
select count(1)
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
where start_lng=start_lat and end_lng = end_lat;
```

```
select count(1)
from `healthy-battery-390020.Cyclistic.cycle_trip_data`
where round(start_lat)= round(start_lng)
and round(end_lat)= round(end_lng);
```

6. Excluded from analyze values

```
[Start Station Name] <> "DIVVY CASSETTE REPAIR MOBILE STATION"
and [Start Station Name] <> "OH Charging Stx - Test"
and [Start Station Name] <> "646"
and [End Station Name] <> "DIVVY CASSETTE REPAIR MOBILE
STATION"
and [End Station Name] <> "OH Charging Stx - Test"
and not Text.EndsWith([Start Station Name], "**")
and not Text.EndsWith([End Station Name], "**")
```

- Changed Types:

COLUMN NAME	ORIGINAL TYPE	NEW TYPE
Started at	text	datetime
Ended at	text	datetime
Start Lat	text	number
Start Lng	text	number
End Lat	text	number
End Lng	text	number

- Added Calculated Columns:

NAME	TYPE	DESCRIPTION
Day	date	Extract from "Started at" date without time; Join Column for Calculated "DIMM_Date" Table
Time to the Minute	time	Extract from "Started at" time without seconds; Join Column for Calculated "DIMM_Time" Table
Duration of ride (min)	minute	Calculate subtraction between "Ended at" and "Ended at" in minutes Number.Round(Duration.TotalMinutes([ended_at] - [started_at]), 2))

- Added Calculated Tables:

DIMM_Time {0..1439}

COLUMN NAME	COLUMN TYPE
Minute	number
Time to the Minute	time
5 min bucket	number
10 min bucket	number

1 hour bucket	number
5 min time slot	time
10 min time slot	time
1 hour time slot	time
Hour	number
Time of Day	text enum {night/morning/afternoon/evening}

- *DIMM_Date [07/01/2022; 06/30/2023]*

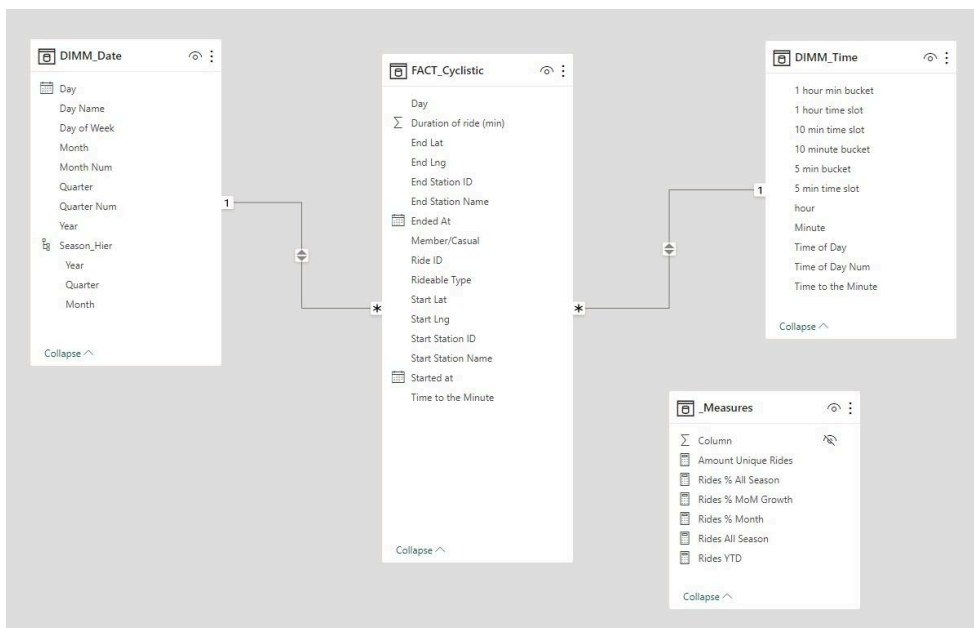
COLUMN NAME	COLUMN TYPE
Day	date
Day Name	text {Sunday, Monday, ect}
Day of Week	number {0,...,6}
Quarter Num	number {1,...,4}
Quarter	text {Qrt1, Qrt2,Qrt3, Qrt4}
Month	text {January, February,...}

Month Num number (1,.. 12)

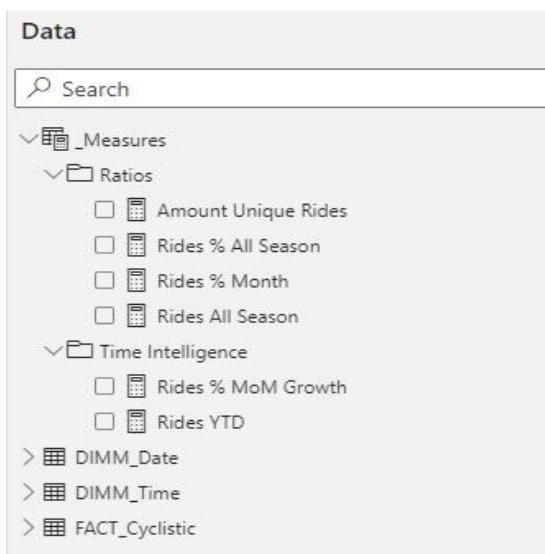
Year number

- Removed rows with empty "Start Station Name" and "End Station Name"
- Trimmed text fields in the "FACT_Cyclistic" Table

Model View:



Measures:



Amount Unique Rides = *DISTINCTCOUNT*(*FACT_Cyclistic*[*Ride ID*])

Rides % All Season = *DIVIDE*(
 DISTINCTCOUNT(*FACT_Cyclistic*[*Ride ID*]),
 CALCULATE(
 DISTINCTCOUNT(*FACT_Cyclistic*[*Ride ID*]),
 REMOVEFILTERS(*FACT_Cyclistic*)
)
)

Rides % Month = *IF*(
 ISINSCOPE(*DIMM_Date*[*Quarter*]),
 DIVIDE(
 [Amount Unique Rides],
 CALCULATE(
 [Amount Unique Rides],
 REMOVEFILTERS(*DIMM_Date*[*Month*], *DIMM_Date*[*Month*
Num])
)
)
)

Rides All Season = *CALCULATE*(
 DISTINCTCOUNT(*FACT_Cyclistic*[*Ride ID*]),
 REMOVEFILTERS(*DIMM_Date*[*Month*])
)

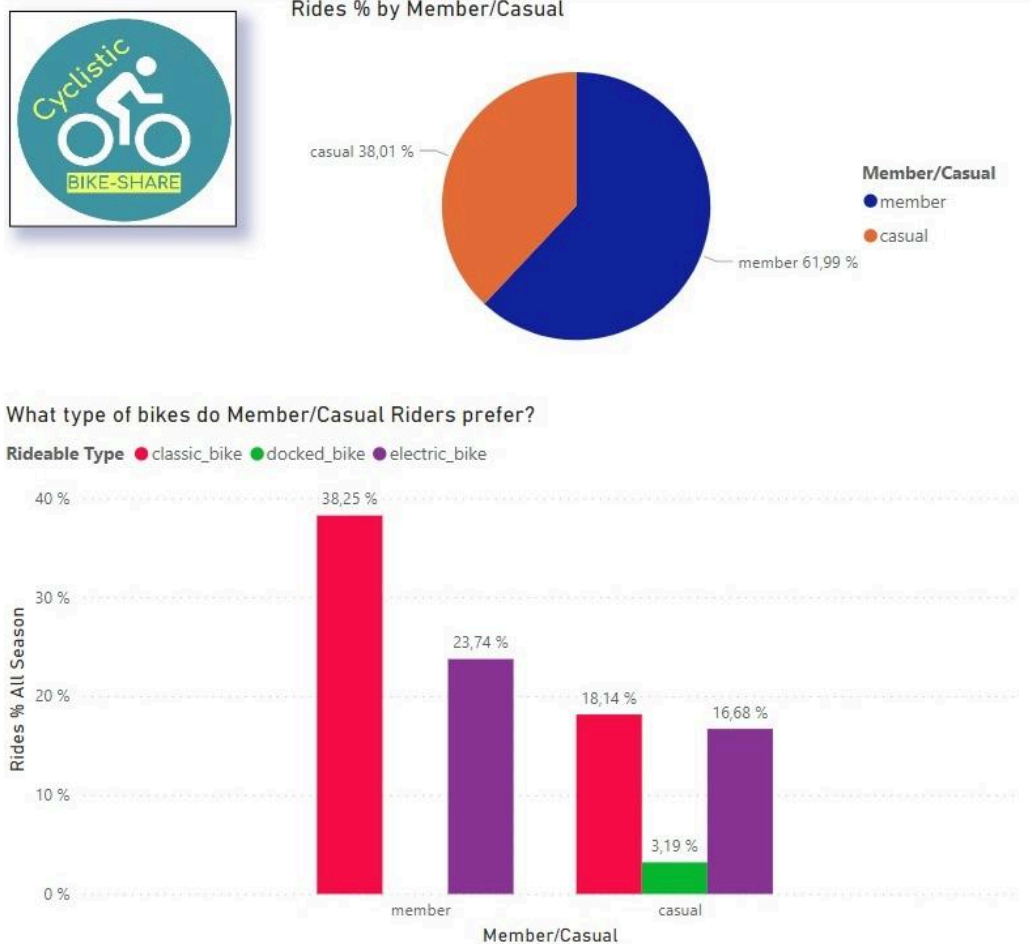
Rides % MoM Growth =
 VAR RidesPriorMonth = *CALCULATE*(*[Amount Unique Rides]*,
 PARALLELPERIOD('DIMM_Date'[*Day*], -1, MONTH)
)
 VAR RidesGrowthMonth = *IF*(
 ISINSCOPE(*DIMM_Date*[*Month*]),
 DIVIDE(
 (*[Amount Unique Rides]* - *RidesPriorMonth*),
 RidesPriorMonth
)
)
RETURN RidesGrowthMonth

```

Rides YTD = IF(
    ISINSCOPE(DIMM_Date[Quarter])||ISINSCOPE(DIMM_Date[Month]),
    TOTALYTD(
        [Amount Unique Rides],
        DIMM_Date[Day],
        "06-30"
    )
)

```

ANALYZE



Member customers made up 62% of all rides, while Casual customers accounted for the remaining 38%.

Among Members, 38.25% preferred classic bikes and 23.74% preferred electric bikes. For Casual customers, 18.14% chose classic bikes, 16.68% preferred electric bikes, and 3.19% opted for docked bikes. It's important to note that Members did not use docked bikes at all.

Member/Casual

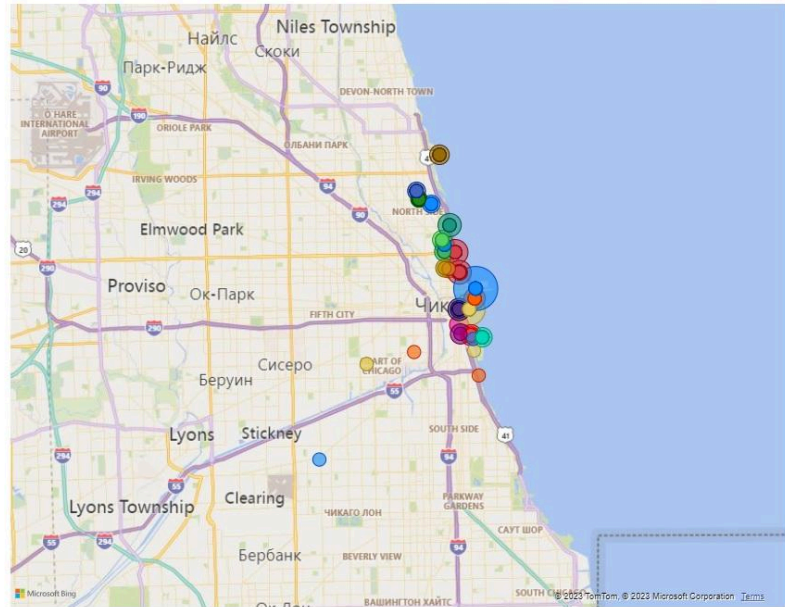
☒ casual

☐ member

Start Station Name	casual	Total
Streeter Dr & Grand Ave	1.15 %	1.15 %
DuSable Lake Shore Dr & Monroe St	0.67 %	0.67 %
Michigan Ave & Oak St	0.51 %	0.51 %
Millennium Park	0.50 %	0.50 %
DuSable Lake Shore Dr & North Blvd	0.47 %	0.47 %
Shedd Aquarium	0.43 %	0.43 %
Theater on the Lake	0.37 %	0.37 %
Wells St & Concord Ln	0.31 %	0.31 %
Dusable Harbor	0.30 %	0.30 %
Indiana Ave & Roosevelt Rd	0.28 %	0.28 %
Clark St & Armitage Ave	0.26 %	0.26 %
Montrose Harbor	0.26 %	0.26 %
Clark St & Lincoln Ave	0.25 %	0.25 %
Clark St & Elm St	0.25 %	0.25 %
Adler Planetarium	0.25 %	0.25 %
Wells St & Elm St	0.24 %	0.24 %
Michigan Ave & 8th St	0.23 %	0.23 %
Broadway & Barry Ave	0.23 %	0.23 %
Clark St & Newport St	0.23 %	0.23 %
Wilton Ave & Belmont Ave	0.23 %	0.23 %
Total	7.41 %	7.41 %

What is the Most Popular Start Stations (Top 20) Among Riders ?

Start Station Name ● Adler Pl... ● Broadw... ● Clark St ... ● Clark St ... ● Clark St ... ● Clark St ... ● Dusable... ● DuSable... ● DuSable...



The most popular start station among Casual Riders is "Streeter Dr & Grand Ave", accounting for 1.15% of Casual Rides. Casual Riders tend to prefer rides along the Chicago Lakefront Trail.

Member/Casual

☐ casual

☒ member

Start Station Name	member	Total
Kingsbury St & Kinzie St	0.54 %	0.54 %
Clark St & Elm St	0.50 %	0.50 %
Clinton St & Washington Blvd	0.48 %	0.48 %
Loomis St & Lexington St	0.45 %	0.45 %
Wells St & Concord Ln	0.45 %	0.45 %
University Ave & 57th St	0.43 %	0.43 %
Ellis Ave & 60th St	0.43 %	0.43 %
Clinton St & Madison St	0.42 %	0.42 %
Wells St & Elm St	0.41 %	0.41 %
Canal St & Adams St	0.38 %	0.38 %
Broadway & Barry Ave	0.38 %	0.38 %
Streeter Dr & Grand Ave	0.37 %	0.37 %
St. Clair St & Erie St	0.36 %	0.36 %
State St & Chicago Ave	0.35 %	0.35 %
Dearborn St & Erie St	0.35 %	0.35 %
Wells St & Huron St	0.34 %	0.34 %
Morgan St & Polk St	0.34 %	0.34 %
Clinton St & Jackson Blvd	0.34 %	0.34 %
DuSable Lake Shore Dr & North Blvd	0.34 %	0.34 %
Halsted St & Polk St	0.34 %	0.34 %
Total	8.00 %	8.00 %

What is the Most Popular Start Stations (Top 20) Among Riders ?

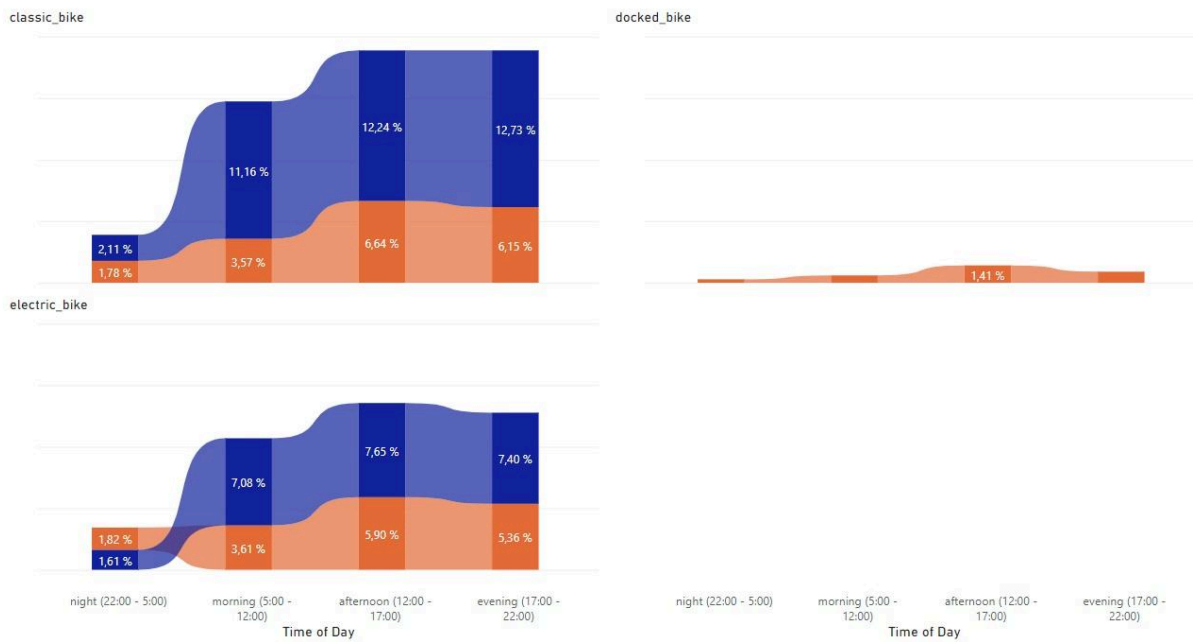
Start Station Name ● Broadw... ● Canal St... ● Clark St ... ● Clinton ... ● Clinton ... ● Clinton ... ● Dearbor... ● DuSable... ● Ellis Ave...



The most popular start station among Member Riders is 'Kingsbury St & Kinzie St,' accounting for 0.54% of Member Rides. Member Riders don't have as strong a trend as Casual Riders.

What part of day and type of bike Member/Casual customers prefer?

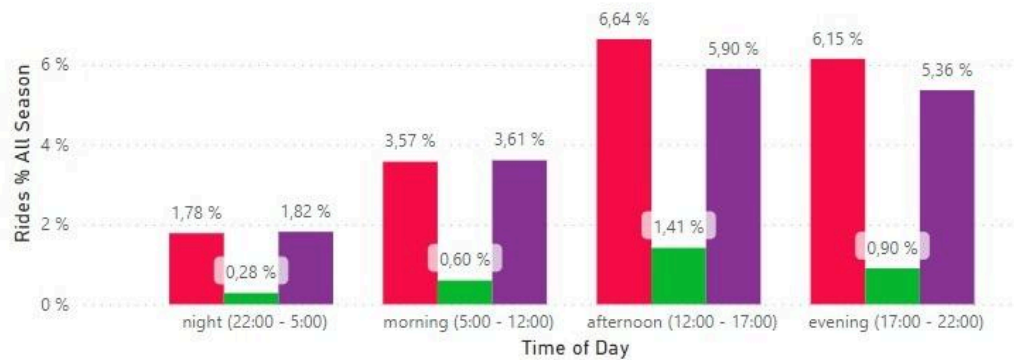
Member/Casual ● casual ● member



Casual customers ride more than member customers during the night (22:00 - 5:00) on the electric bikes; in other cases, member customers ride more.

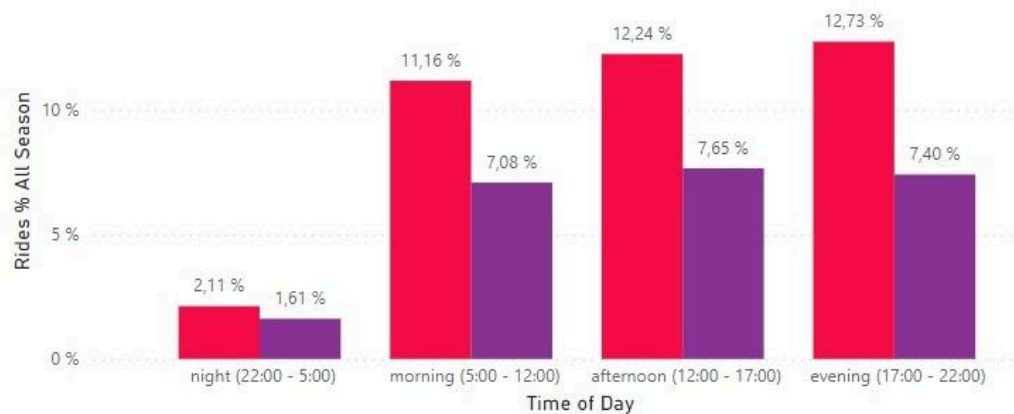
Casual Rides % by Time of Day

Rideable Type ● classic_bike ● docked_bike ● electric_bike



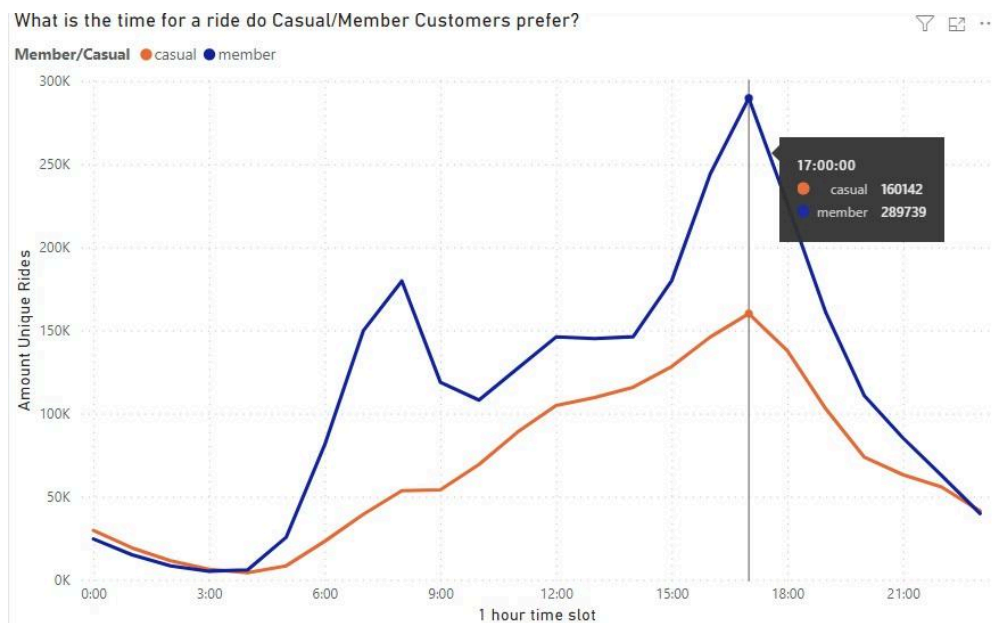
Member Rides % by Time of Day

Rideable Type ● classic_bike ● electric_bike



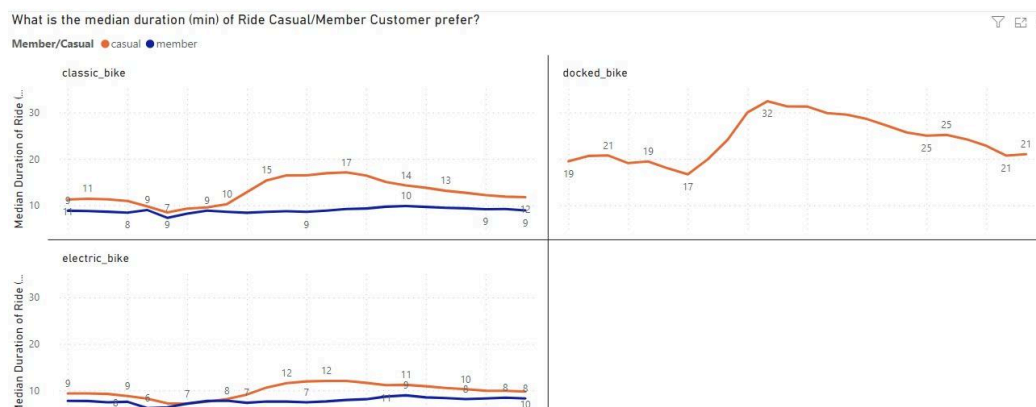
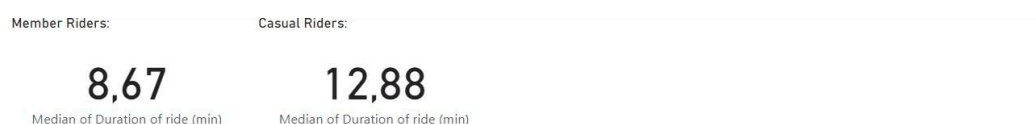
Classic and electric bikes are almost equally popular among casual riders during the night (22:00 - 5:00) and morning (5:00 - 12:00). However, in the afternoon, casual riders use classic bikes slightly more than electric bikes, with a difference of 0.74%, and in the evening, the difference is 0.79%.

Member riders prefer to use classic bikes during the morning(5:00 - 12:00), afternoon(12:00 - 17:00), and evening(17:00 - 12:00). During the night (22:00 - 5:00), member riders use classic bikes more than electric bikes, with only a 0.5% difference in favor of classic bikes.

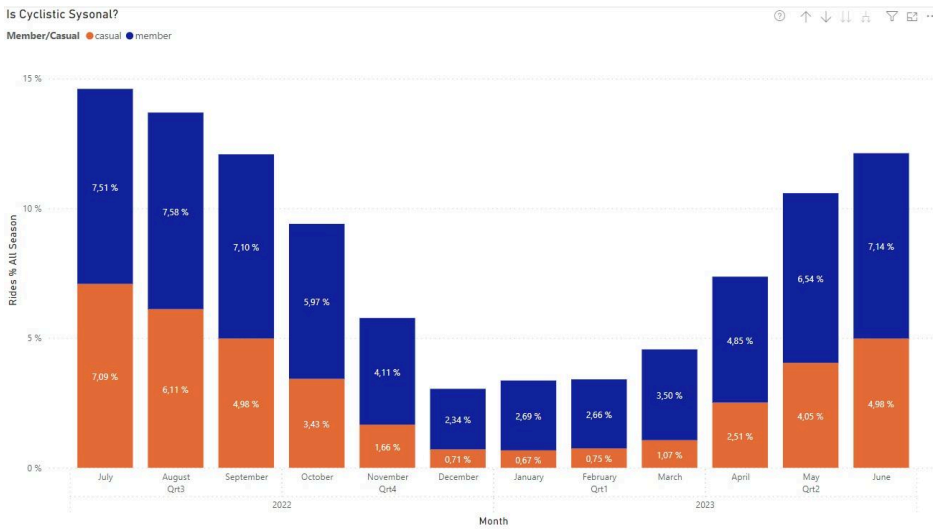


Member rides have two peaks on the graph at 8:00 and 17:00. It appears that they use classic bikes for their daily commute to and from work.

Casual Riders have only one peak value at 17:00.



The median duration is 12.88 minutes for casual customers and 8.67 minutes for member customers. In general, the median duration of rides for casual customers is higher than that for members, especially between 9:00 and 15:00.



Cyclistic is a seasonal service. Warm months, especially summer months, are popular among Member and Casual riders. July leads in ridership for Casual Riders, and August leads for Member Riders.

Member/Casual

casual

Ride Trends by Time Period

Year	Quarter	Member/Casual Month	casual Rides All Season	Rides % All Season	Rides % Month	Rides % MoM Growth	Rides YTD
2022	Qrt3	July	307041	7.09 %	38.98 %		307041
		August	264885	6.11 %	33.63 %	-13.73 %	571926
		September	215834	4.98 %	27.40 %	-18.52 %	787760
		Total	787760	18.18 %	100.00 %		787760
	Qrt4	October	148527	3.43 %	59.10 %	-31.18 %	936287
		November	72067	1.66 %	28.68 %	-51.48 %	1008354
		December	30704	0.71 %	12.22 %	-57.40 %	1039058
		Total	251298	5.80 %	100.00 %		1039058
	Total		1039058	23.98 %			
	2023	Qrt1	January	29147	0.67 %	27.06 %	-5.07 %
February			32405	0.75 %	30.09 %	11.18 %	1100610
March			46149	1.07 %	42.85 %	42.41 %	1146759
Total			107701	2.49 %	100.00 %		1146759
Qrt2		April	108909	2.51 %	21.79 %	135.99 %	1255668
		May	175274	4.05 %	35.06 %	60.94 %	1430942
		June	215740	4.98 %	43.15 %	23.09 %	1646682
		Total	499923	11.54 %	100.00 %		1646682
Total		607624	14.03 %				
Total		1646682	38.01 %				

Member/Casual

member

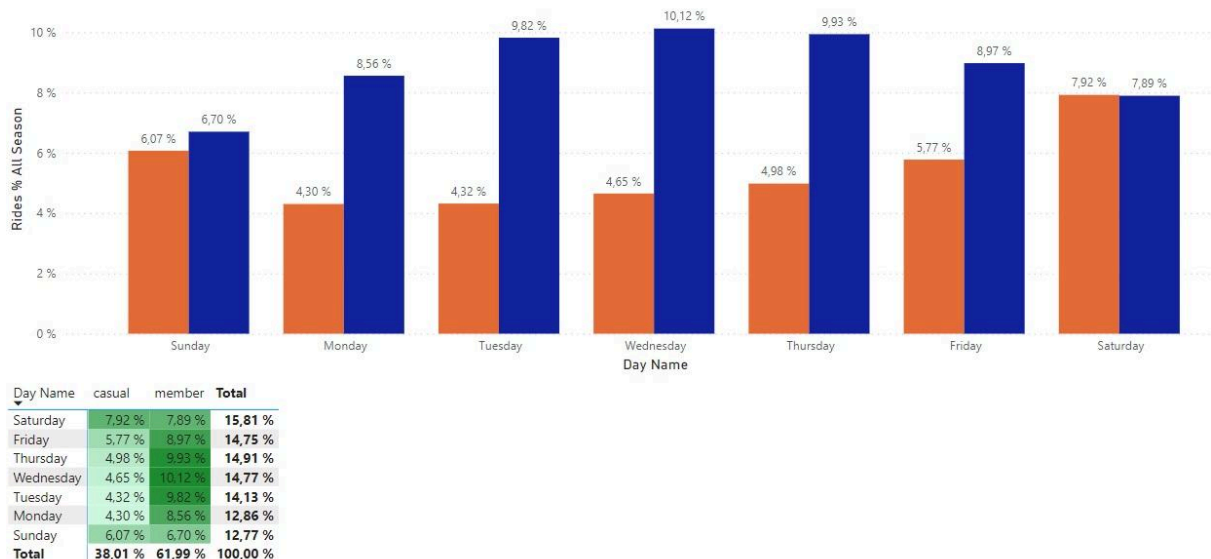
Ride Trends by Time Period

Year	Quarter	Member/Casual Month	member Rides All Season	Rides % All Season	Rides % Month	Rides % MoM Growth	Rides YTD
2022	Qrt3	July	325577	7,51 %	33,87 %		325577
		August	328179	7,58 %	34,14 %	0,80 %	653756
		September	307548	7,10 %	31,99 %	-6,29 %	961304
		Total	961304	22,19 %	100,00 %		961304
	Qrt4	October	258723	5,97 %	48,09 %	-15,88 %	1220027
		November	178076	4,11 %	33,10 %	-31,17 %	1398103
		December	101191	2,34 %	18,81 %	-43,18 %	1499294
		Total	537990	12,42 %	100,00 %		1499294
	Total		1499294	34,61 %			
2023	Qrt1	January	116590	2,69 %	30,41 %	15,22 %	1615884
		February	115312	2,66 %	30,07 %	-1,10 %	1731196
		March	151527	3,50 %	39,52 %	31,41 %	1882723
		Total	383429	8,85 %	100,00 %		1882723
	Qrt2	April	210236	4,85 %	26,18 %	38,74 %	2092959
		May	283242	6,54 %	35,27 %	34,73 %	2376201
		June	309480	7,14 %	38,54 %	9,26 %	2685681
		Total	802958	18,53 %	100,00 %		2685681
	Total		1186387	27,38 %			
Total			2685681	61,99 %			

In April, Casual Riders experienced a 136% growth, while Member Riders had a 38.74% growth in the number of rides, making it the largest growth in April.

What is the Day of Week do Casual/Member Customers prefer?

Member/Casual casual member



The most popular days for rides among Casual Riders are Saturday and Sunday, while Wednesday is the most popular day for Member Riders. Member riders are more consistent throughout the week.

Cyclistic Bike-Share Dashboard:

Cyclistic Case Study Dashboard

ACT

1. Begin a marketing campaign promoting membership in March.
2. Offer customized membership options specifically for weekends.
3. Pair custom membership options with other entertainment offerings along the Chicago Lakefront Trail.
4. Concentrate your marketing efforts along the Chicago Lakefront Trail.
5. Create special offers for Docker memberships.
6. Introduce a night-time special offer for electric bike rentals.
7. Morning special offer to attract people to ride to work with the help of bikes.