

Bike-Sharing Analysis Report

Rishabh Jaiswal

2024-05-23

Introduction

Background

As 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, my team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, we will design a new marketing strategy to convert casual riders into annual members.

Objective of the Analysis

Analyze the behavioral patterns and usage metrics of casual riders compared to annual paid members in our bike share program to identify key differences. Use these insights to develop targeted strategies and recommendations aimed at converting casual riders into annual members

Data description

This analysis is based on the Divvy case study “‘Sophisticated, Clear, and Polished’: Divvy and Data Visualization” written by Kevin Hartman (found here: <https://artscience.blog/home/divvy-dataviz-case-study>).

Data preparation

-- Creating a table named bike_share

```
CREATE TABLE bike_share(ride_id VARCHAR(100), rideable_type VARCHAR(50), started_at  
TIMESTAMP, ended_at TIMESTAMP, start_station_name VARCHAR(255),  
end_station_name VARCHAR(255), member_casual VARCHAR(50),  
ride_duration INTERVAL, day_of_week INTEGER);
```

-- Importing data : csv files using copy command

```
COPY bike_share(ride_id, rideable_type, started_at, ended_at, start_station_name,  
end_station_name, member_casual, ride_duration, day_of_week) FROM  
'C:/Users/Saurabh/Documents/bike_share_success/csv/202301-divvy  
tripdata.csv' DELIMITER ',' CSV HEADER;
```

```
COPY bike_share(ride_id, rideable_type, started_at, ended_at, start_station_name,  
end_station_name, member_casual, ride_duration, day_of_week) FROM  
'C:/Users/Saurabh/Documents/bike_share_success/csv/202302-divvy-tripdata.csv'  
DELIMITER ',' CSV HEADER;
```

Similarly the data for every month has been imported into the bike_share table created earlier.

Data Cleaning

-- Removing rows with duplicate ride_id

```
DELETE FROM bike_share  
WHERE ride_id IN (  
    SELECT ride_id  
    FROM bike_share  
    GROUP BY ride_id  
    HAVING COUNT(*) > 1);
```

-- Removing rows with multiple or relevant empty columns

```
DELETE FROM bike_share  
WHERE ride_duration IS NULL OR start_station_name IS NULL OR end_station_name IS NULL OR  
rideable_type IS NULL OR member_casual IS NULL OR ride_id IS NULL;
```

--Inspecting the dataset

```
SELECT * FROM bike_share LIMIT 10;
```

```

SELECT COUNT(*) num_of_rows,
COUNT(DISTINCT ride_id) num_of_dis_rows
FROM bike_share;

```

ride_id	rideable_type	started_at	ended_at	start_station_name	end_station_name	member_casual	ride_duration	day_of_week
F7333D8 DD77C05 73	electric_bike	10/1/20 23 18:30	10/1/20 23 18:50	Pine Grove Ave & Waveland Ave	California Ave & Byron St	member	0:20:00	1
F5B8796 DBDD553 6F	electric_bike	10/8/20 23 18:32	10/8/20 23 18:52	Broadway & Waveland Ave	California Ave & Byron St	member	0:20:00	1
8F97128A 4399DB5 8	classic_bike	10/22/20 023 16:00	10/22/20 023 16:20	Wabash Ave & 16th St	Clinton St & Jackson Blvd	casual	0:20:00	1
5B876A4 70D2943 99	electric_bike	10/21/20 023 16:36	10/21/20 023 16:56	State St & 33rd St	Racine Ave & 35th St	member	0:20:00	7
CD7FFFF F3D8DA6 50	electric_bike	10/18/20 023 9:25	10/18/20 023 9:45	Spaulding Ave & Armitage Ave	St. Louis Ave & Balmoral Ave	member	0:20:00	4
4685B49 8A88EB8 C1	classic_bike	10/2/20 23 21:30	10/2/20 23 21:50	Sangamon St & Lake St	Dearborn St & Erie St	member	0:20:00	2
10F1E8A B7369A6 C8	classic_bike	10/21/20 023 11:32	10/21/20 023 11:52	Ritchie Ct & Banks St	Dusable Harbor	member	0:20:00	7
8163CC9 D7E4FFD 25	classic_bike	10/3/20 23 16:40	10/3/20 23 17:00	Halsted St & Fulton St	Fairbanks St & Superior St	member	0:20:00	3
15439029 D90A538 8	classic_bike	10/5/20 23 12:08	10/5/20 23 12:28	Millennium Park	Cityfront Plaza Dr & Pioneer Ct	casual	0:20:00	5
88772BD FE3865D 04	electric_bike	10/23/20 023 8:03	10/23/20 023 8:23	Lincoln Ave & Diversey Pkwy	Cityfront Plaza Dr & Pioneer Ct	member	0:20:00	2

num_of_rows	num_of_dis_rows
4330300	4330300

Performing analysis

-- Average ride duration by user type

```
SELECT member_casual,  
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration  
FROM bike_share  
GROUP BY member_casual;
```

member_casual	avg_ride_duration
casual	22.93712674
member	12.13076574

-- Average ride duration per user type by days of the week

```
SELECT member_casual, TO_CHAR(started_at, 'Day') AS day,  
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration  
FROM bike_share GROUP BY TO_CHAR(started_at, 'Day'), member_casual  
ORDER BY EXTRACT(DAY FROM MIN(started_at));
```

member_casual	day	avg_ride_duration
casual	Sunday	26.55642248
member	Sunday	13.6122201
casual	Monday	22.53409136
member	Monday	11.54927854
member	Tuesday	11.64676111
casual	Tuesday	20.51392471
member	Wednesday	11.58547089
casual	Wednesday	19.59170172
casual	Thursday	20.00544563
member	Thursday	11.60063693
member	Friday	12.03986024
casual	Friday	22.31992379
casual	Saturday	25.91961374

member	Saturday	13.58330052
--------	----------	-------------

/*In this query:

EXTRACT(DAY FROM MIN(started_at)) extracts the day from the smallest started_at value in the grouped rows. This ensures that the output is ordered chronologically by the earliest occurrence of each day.

We use MIN(started_at) to ensure that we're extracting the day from the earliest date within each group, as extracting from any other date in the group would yield the same day.*/

-- Average ride duration for casual riders by days of the week, sorted in descending order

```
SELECT member_casual, TO_CHAR(started_at, 'Day') AS day,
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration
FROM bike_share WHERE member_casual = 'casual'
GROUP BY TO_CHAR(started_at, 'Day'), member_casual
ORDER BY avg_ride_duration DESC;
```

member_casual	day	avg_ride_duration
casual	Sunday	26.55642248
casual	Saturday	25.91961374
casual	Monday	22.53409136
casual	Friday	22.31992379
casual	Tuesday	20.51392471
casual	Thursday	20.00544563
casual	Wednesday	19.59170172

-- Seasonal trends in ride duration

```
SELECT member_casual, TO_CHAR(started_at, 'Month') AS month,
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration
FROM bike_share GROUP BY TO_CHAR(started_at, 'Month'), member_casual
ORDER BY EXTRACT(MONTH FROM MIN(started_at));
```

member_casual	month	avg_ride_duration
member	January	10.00449179

casual	January	14.88011883
casual	February	17.69222567
member	February	10.41696034
casual	March	16.72077964
member	March	10.17157919
casual	April	22.62286878
member	April	11.55563672
casual	May	24.52611263
member	May	12.69299423
casual	June	24.07290007
member	June	12.94513869
casual	July	25.21376757
member	July	13.35025555
member	August	13.31027358
casual	August	24.3608572
member	September	12.67266648
casual	September	23.53691486
casual	October	21.36332926
member	October	11.64808644
member	November	11.05473847
casual	November	17.88659136
member	December	10.80452287
casual	December	16.5388601

-- Seasonal trends in ride duration for casual riders, sorted in descending order

```
SELECT member_casual, TO_CHAR(started_at, 'Month') AS month,
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration
FROM bike_share WHERE member_casual = 'casual'
GROUP BY TO_CHAR(started_at, 'Month'), member_casual
ORDER BY avg_ride_duration DESC;
```

member_casual	month	avg_ride_duration
---------------	-------	-------------------

casual	July	25.21376757
casual	May	24.52611263
casual	August	24.3608572
casual	June	24.07290007
casual	September	23.53691486
casual	April	22.62286878
casual	October	21.36332926
casual	November	17.88659136
casual	February	17.69222567
casual	March	16.72077964
casual	December	16.5388601
casual	January	14.88011883

/* AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration: This calculates the average ride duration in minutes.

EXTRACT(EPOCH FROM ride_duration): This extracts the duration of each ride in seconds using the EXTRACT function with the EPOCH parameter, as discussed earlier.

/60: This divides the extracted ride durations (in seconds) by 60 to convert them from seconds to minutes.

AVG: This calculates the average of the ride durations.

FROM bike_share: This specifies the table from which the data is being retrieved, in this case, the bike_share table. */

-- Ride frequency by day of the week

```
SELECT member_casual, TO_CHAR(started_at, 'Day') AS day_week,
COUNT(*) AS ride_count
FROM bike_share GROUP BY TO_CHAR(started_at, 'Day'), member_casual
ORDER BY EXTRACT(DAY FROM MIN(started_at));
```

member_casual	day_week	ride_count
casual	Sunday	254668
member	Sunday	307739
casual	Monday	175367

member	Monday	386578
member	Tuesday	448719
casual	Tuesday	181476
member	Wednesday	452581
casual	Wednesday	183050
casual	Thursday	198875
member	Thursday	452483
member	Friday	400399
casual	Friday	227798
casual	Saturday	310053
member	Saturday	350514

-- Ride frequency for casual riders by day of the week

```
SELECT member_casual, TO_CHAR(started_at, 'Day') AS day_week,
COUNT(*) AS ride_count
FROM bike_share WHERE member_casual = 'casual'
GROUP BY TO_CHAR(started_at, 'Day'), member_casual
ORDER BY ride_count DESC;
```

member_casual	day_week	ride_count
casual	Saturday	310053
casual	Sunday	254668
casual	Friday	227798
casual	Thursday	198875
casual	Wednesday	183050
casual	Tuesday	181476
casual	Monday	175367

-- Number of rides per user type across months

```
SELECT member_casual, TO_CHAR(started_at, 'Month') AS month,
COUNT(*) AS ride_count FROM bike_share
GROUP BY TO_CHAR(started_at, 'Month'), member_casual
```


ORDER BY EXTRACT(MONTH FROM MIN(started_at));

member_casual	month	ride_count
member	January	118661
casual	January	29621
casual	February	32826
member	February	116908
casual	March	46791
member	March	153655
casual	April	110500
member	April	213546
casual	May	176945
member	May	286050
casual	June	219698
member	June	314812
casual	July	245243
member	July	328503
member	August	350900
casual	August	233785
member	September	309565
casual	September	196899
casual	October	130251
member	October	273391
member	November	202618
casual	November	72058
member	December	130404
casual	December	36670

-- Number of rides for casual riders across months arranged in descending order

SELECT member_casual, TO_CHAR(started_at, 'Month') AS month,

COUNT(*) AS ride_count FROM bike_share

WHERE member_casual = 'casual'

GROUP BY TO_CHAR(started_at, 'Month'), member_casual

ORDER BY ride_count DESC;

member_casual	month	ride_count
casual	July	245243
casual	August	233785
casual	June	219698
casual	September	196899
casual	May	176945
casual	October	130251
casual	April	110500
casual	November	72058
casual	March	46791
casual	December	36670
casual	February	32826
casual	January	29621

-- Preferred routes and docking stations by casual riders with long rides

SELECT start_station_name || ' to ' || end_station_name AS route,

COUNT (*) AS ride_count FROM bike_share

WHERE member_casual = 'casual' AND

EXTRACT(EPOCH FROM ride_duration) > 10 * 3600

GROUP BY route

ORDER BY ride_count DESC;

route	ride_count
Streeter Dr & Grand Ave to Streeter Dr & Grand Ave	19
Millennium Park to Millennium Park	15
Dusable Harbor to Dusable Harbor	14
Wabash Ave & Grand Ave to Wabash Ave & Grand Ave	12
Shore Dr & 55th St to Shore Dr & 55th St	11
Museum of Science and Industry to Museum of Science and Industry	9
McClurg Ct & Ohio St to McClurg Ct & Ohio St	9

DuSable Lake Shore Dr & Monroe St to DuSable Lake Shore Dr & Monroe St	8
Dusable Harbor to Streeter Dr & Grand Ave	8
Indiana Ave & Roosevelt Rd to Indiana Ave & Roosevelt Rd	8
Lakefront Trail & Wilson Ave to Lakefront Trail & Wilson Ave	7
Ellis Ave & 60th St to Ellis Ave & 60th St	6
Michigan Ave & 8th St to Michigan Ave & 8th St	6
Clark St & Drummond Pl to Clark St & Drummond Pl	6
Clark St & Lincoln Ave to Clark St & Lincoln Ave	6
LaSalle Dr & Huron St to LaSalle Dr & Huron St	6
Montrose Harbor to Montrose Harbor	6
DuSable Lake Shore Dr & Monroe St to Streeter Dr & Grand Ave	6
Theater on the Lake to Theater on the Lake	6
Michigan Ave & Pearson St to Michigan Ave & Pearson St	6
Chicago Ave & Sheridan Rd to Chicago Ave & Sheridan Rd	6
McCormick Place to McCormick Place	6
University Library (NU) to University Library (NU)	6
Southport Ave & Roscoe St to Southport Ave & Roscoe St	5
Wells St & Hubbard St to Streeter Dr & Grand Ave	5
Michigan Ave & Oak St to Michigan Ave & Oak St	5
Racine Ave & Fullerton Ave to Racine Ave & Fullerton Ave	5
Michigan Ave & Oak St to Streeter Dr & Grand Ave	5
California Ave & Montrose Ave to California Ave & Montrose Ave	5
Lakefront Trail & Bryn Mawr Ave to Lakefront Trail & Bryn Mawr Ave	5
Ritchie Ct & Banks St to Ritchie Ct & Banks St	5
Aberdeen St & Jackson Blvd to Canal St & Adams St	5
Michigan Ave & Jackson Blvd to Michigan Ave & Jackson Blvd	5
Mies van der Rohe Way & Chestnut St to Mies van der Rohe Way & Chestnut St	5
Franklin St & Jackson Blvd to Franklin St & Jackson Blvd	5
Adler Planetarium to Streeter Dr & Grand Ave	5

/* ||: This is the concatenation operator in SQL. It combines two strings into a single string.

' to ': This is a string literal that adds the text " to " between the start station name and the end station name in the output. */

-- Average ride duration for each user type by different bikes

```
SELECT member_casual, rideable_type,  
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration  
FROM bike_share  
GROUP BY member_casual, rideable_type;
```

member_casual	rideable_type	avg_ride_duration
casual	classic_bike	25.74943249
casual	docked_bike	53.9556477
casual	electric_bike	14.664367
member	classic_bike	12.99734026
member	electric_bike	10.52712039

-- Average ride duration for each user type by different bikes across months

```
SELECT TO_CHAR(started_at, 'Month') AS month,  
member_casual, rideable_type,  
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration  
FROM bike_share GROUP BY TO_CHAR(started_at, 'Month'),  
member_casual, rideable_type  
ORDER BY EXTRACT(MONTH FROM MIN(started_at));
```

month	member_casual	rideable_type	avg_ride_duration
January	casual	docked_bike	38.44887039
January	casual	electric_bike	9.749485049
January	member	classic_bike	10.67069593
January	member	electric_bike	8.801976219
January	casual	classic_bike	17.23160173
February	casual	docked_bike	42.93398419
February	casual	electric_bike	11.39002566
February	member	classic_bike	11.14352631

February	member	electric_bike	9.15029919
February	casual	classic_bike	20.37328767
March	casual	docked_bike	41.91236413
March	casual	electric_bike	10.8634896
March	member	classic_bike	11.17476825
March	member	electric_bike	8.841323665
March	casual	classic_bike	20.28869969
April	casual	classic_bike	26.34796174
April	casual	docked_bike	51.18724731
April	casual	electric_bike	14.55207529
April	member	classic_bike	12.83895336
April	member	electric_bike	9.886400121
May	casual	docked_bike	53.78278528
May	member	electric_bike	10.99240102
May	member	classic_bike	13.73936687
May	casual	electric_bike	16.06146122
May	casual	classic_bike	27.06318095
June	casual	docked_bike	56.22130023
June	casual	electric_bike	15.90349517
June	member	classic_bike	13.88241004
June	member	electric_bike	11.38296683
June	casual	classic_bike	26.2955156
July	casual	classic_bike	26.66133487
July	casual	electric_bike	16.39410575
July	member	classic_bike	14.20009224
July	member	electric_bike	11.65076843
July	casual	docked_bike	55.64413508
August	member	classic_bike	14.02357063
August	casual	electric_bike	15.37783558
August	casual	docked_bike	57.06385861
August	casual	classic_bike	25.20597461
August	member	electric_bike	11.58966131

September	member	electric_bike	11.32256667
September	casual	classic_bike	27.63152051
September	casual	electric_bike	15.3058248
September	member	classic_bike	13.26720516
October	casual	classic_bike	25.98800997
October	casual	electric_bike	13.33292716
October	member	classic_bike	12.33621748
October	member	electric_bike	10.20979519
November	casual	classic_bike	22.45787554
November	casual	electric_bike	11.41390062
November	member	classic_bike	11.78974086
November	member	electric_bike	9.639327692
December	member	electric_bike	9.076647842
December	casual	classic_bike	21.59476125
December	casual	electric_bike	10.28851079
December	member	classic_bike	11.7587311

-- Average ride duration for casual riders by different bikes across months, arranged in descending order

```

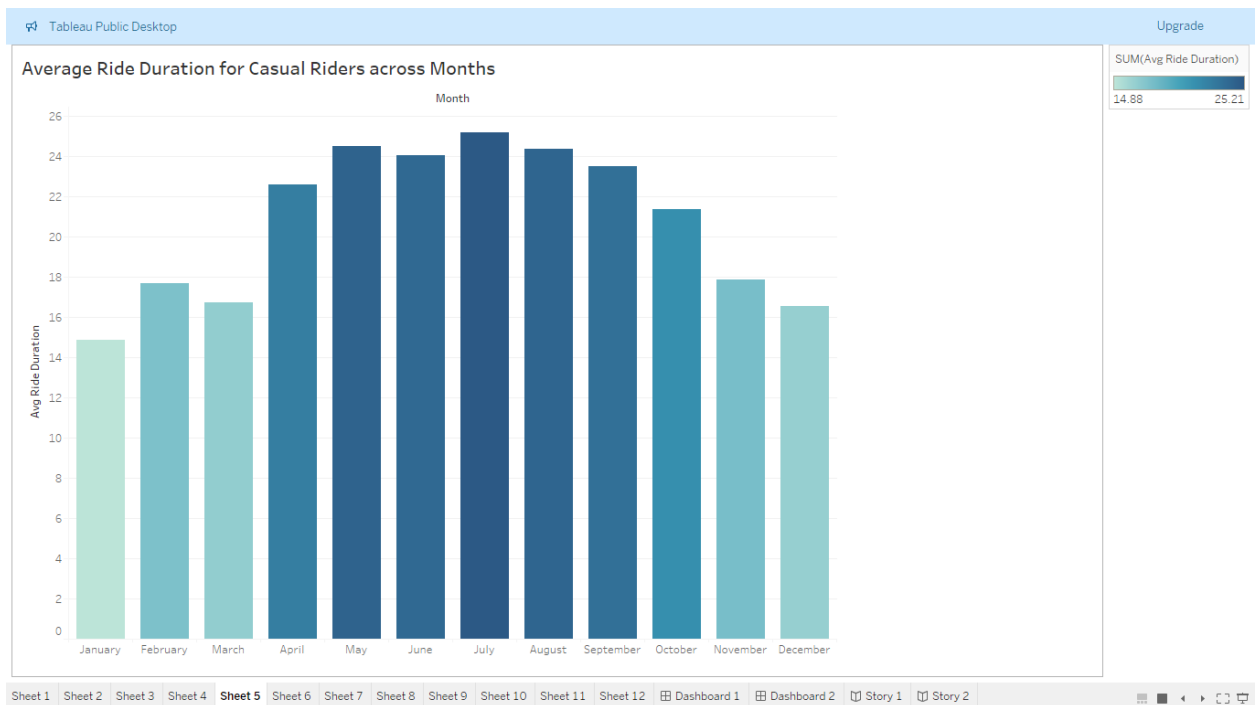
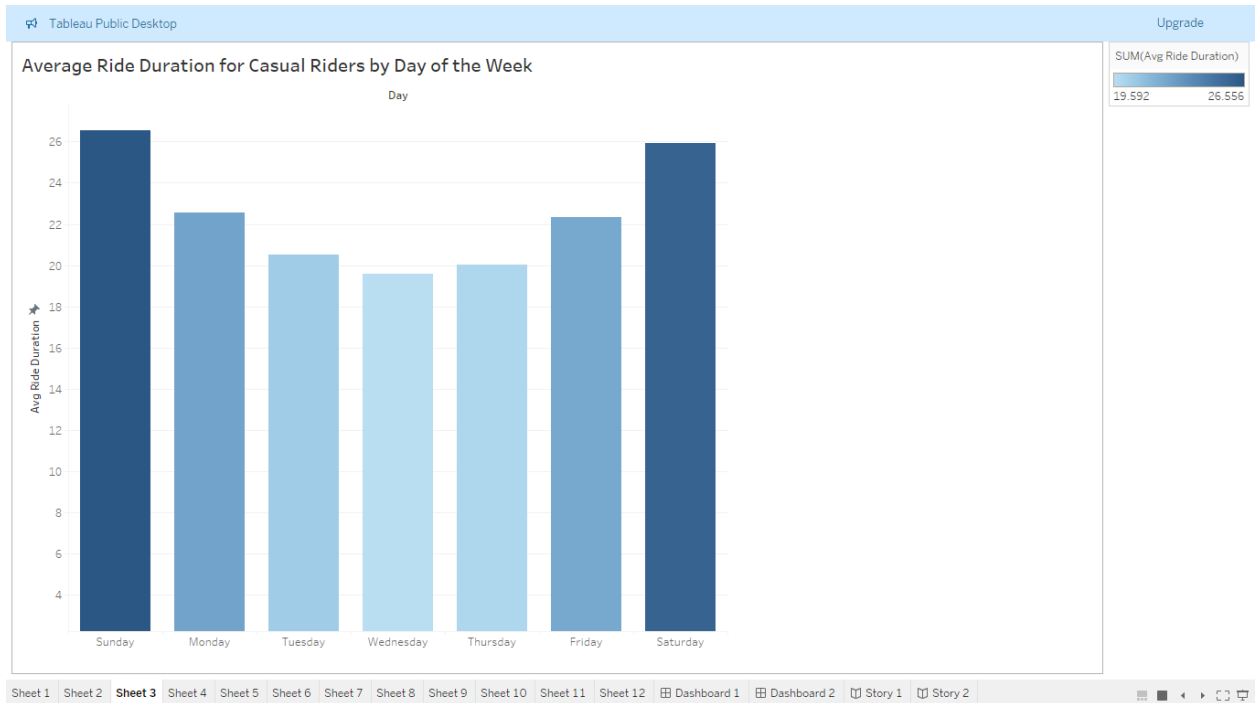
SELECT TO_CHAR(started_at, 'Month') AS month,
member_casual, rideable_type,
AVG(EXTRACT(EPOCH FROM ride_duration)/60) AS avg_ride_duration
FROM bike_share WHERE member_casual = 'casual'
GROUP BY TO_CHAR(started_at, 'Month'),
member_casual, rideable_type
ORDER BY avg_ride_duration DESC;

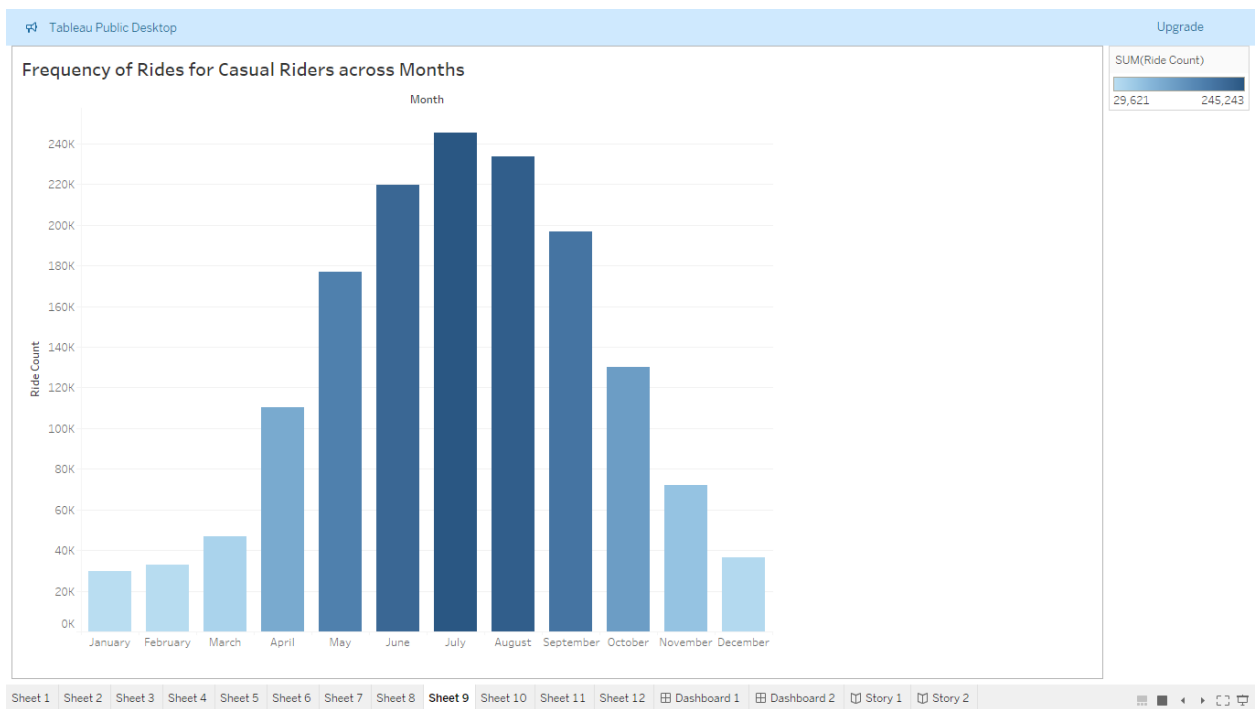
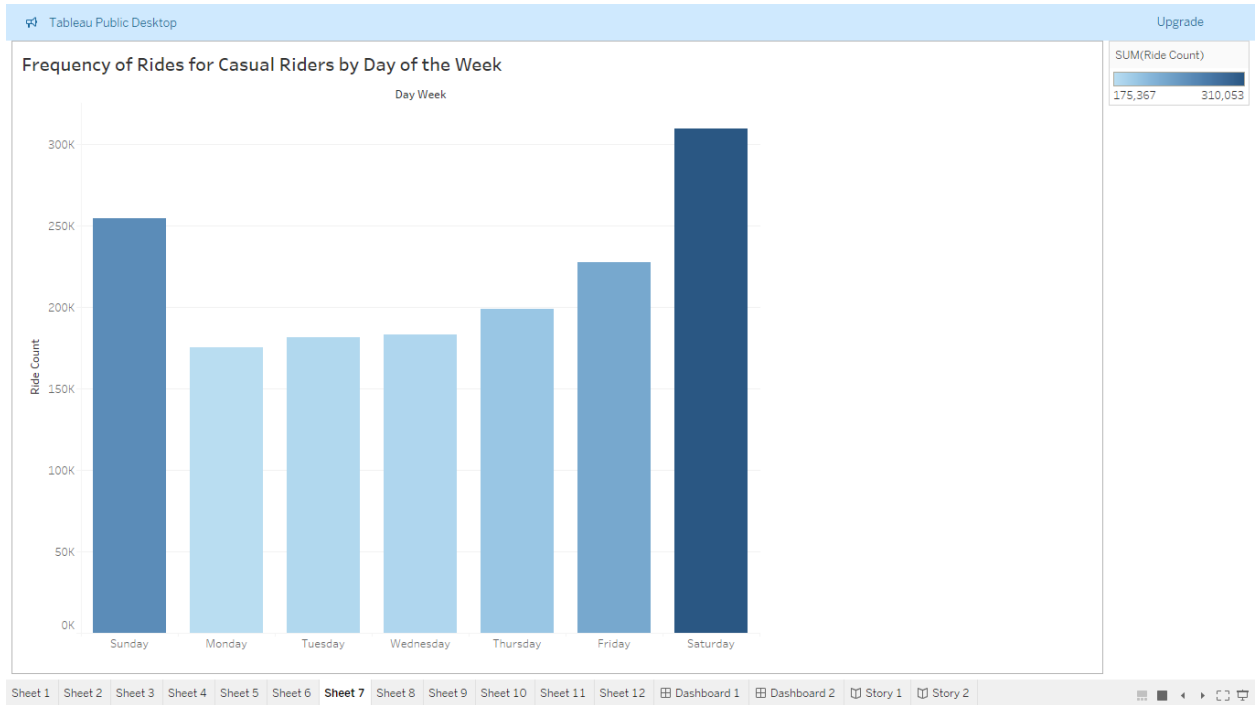
```

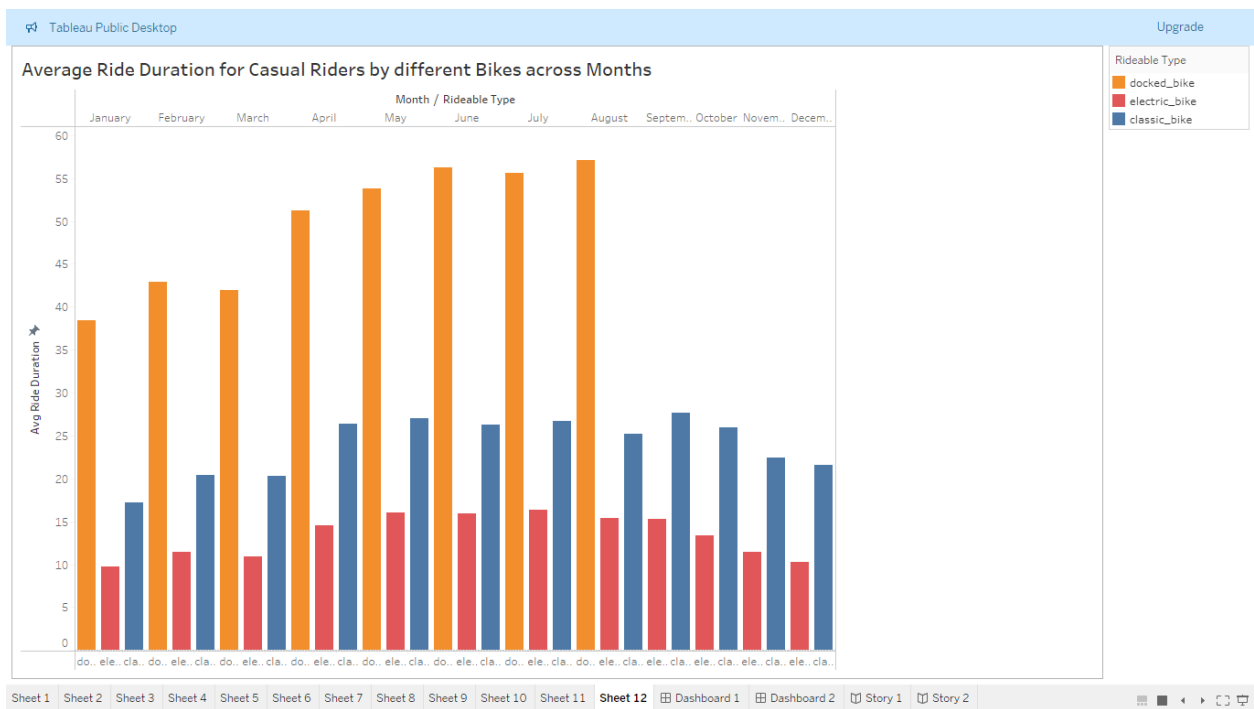
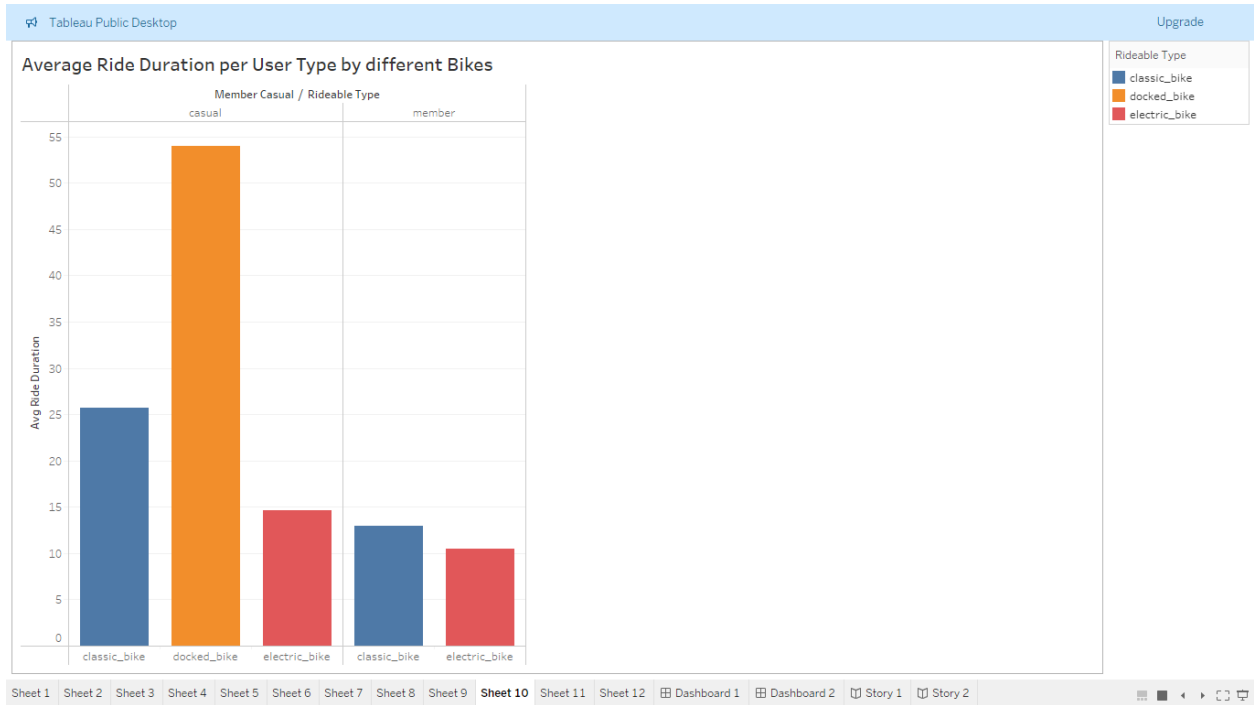
month	member_casual	rideable_type	avg_ride_duration
August	casual	docked_bike	57.06385861
June	casual	docked_bike	56.22130023
July	casual	docked_bike	55.64413508
May	casual	docked_bike	53.78278528

April	casual	docked_bike	51.18724731
February	casual	docked_bike	42.93398419
March	casual	docked_bike	41.91236413
January	casual	docked_bike	38.44887039
September	casual	classic_bike	27.63152051
May	casual	classic_bike	27.06318095
July	casual	classic_bike	26.66133487
April	casual	classic_bike	26.34796174
June	casual	classic_bike	26.2955156
October	casual	classic_bike	25.98800997
August	casual	classic_bike	25.20597461
November	casual	classic_bike	22.45787554
December	casual	classic_bike	21.59476125
February	casual	classic_bike	20.37328767
March	casual	classic_bike	20.28869969
January	casual	classic_bike	17.23160173
July	casual	electric_bike	16.39410575
May	casual	electric_bike	16.06146122
June	casual	electric_bike	15.90349517
August	casual	electric_bike	15.37783558
September	casual	electric_bike	15.3058248
April	casual	electric_bike	14.55207529
October	casual	electric_bike	13.33292716
November	casual	electric_bike	11.41390062
February	casual	electric_bike	11.39002566
March	casual	electric_bike	10.8634896
December	casual	electric_bike	10.28851079
January	casual	electric_bike	9.749485049

Visualizing different metrics calculated to find some insights (Using Tableau)







Conclusion

Possible reasons for longer ride durations of Casual riders

- Casual riders might be using bike sharing service primarily for leisure activities, sightseeing or tourism, which typically involve longer ride duration.
- Weekends generally offer more free time for leisure activities, therefore rise in ride duration on weekends especially for casual riders.

Targeted marketing strategies

- Emphasize leisure and recreational benefits in marketing campaigns.
- Highlight scenic routes, local attractions, and weekend events. Offer weekend specific promotions to encourage casual riders to use the service more frequently and potentially convert them to members.
- Develop membership plans tailored to casual riders who primarily use the service on weekends.
- Increase the distribution of bikes in parks, tourist spots, and recreational areas on weekends to meet the higher demand for casual riders.

Potential factors contributing to popularity of docked bikes among casual riders

- Due to its availability, convenience, or specific routes where these bikes are docked.

Marketing and service strategies for casual riders

- Highlight the availability and benefits of docked bikes in marketing campaigns targeted at casual riders.
- Offer packages or passes specifically for docked bikes, which could include longer ride times or discounts for casual riders.
- Ensure that docked bikes are well maintained and available in popular leisure areas, especially on weekends.

Reasons that might have caused disappearance of docked bikes from September onwards

- Further investigation is required, if there were any operational changes made by the company around september, this could include removal or reduction of docked bikes from the fleet, changes in the areas where they were deployed, or modifications in the service offerings.
- Consulting internal reports from company regarding changes in bike availability.
- Ensure that data collection methods remained consistent throughout the year, any changes in how data was recorded or processed could affect the presence of certain bike types in the data.

- Examine if casual riders' preferences or behavior changed around september, leading to drop in the use of docked bikes. Although this is less likely, it's worth considering.

Seasonal influence on usage (from April to October)

- The period from april to october generally corresponds to warmer weather and longer daylight hours, making it more conducive to outdoor activities and longer rides.
- These months might coincide with peak tourism and recreational activities, leading to mor leisurely and longer rides by casual riders.

Operational adjustments

- Ensure that there are enough bikes available, especially docked bikes, during the peak months to meet the higher demand for longer rides.
- Schedule maintenance activities in the off-peak months (November to March) to ensure the fleet is in optimal condition during the peak season.

Recommendations

- Offer promotions and discounts during the peak months (April to October) to attract more casual riders.
- Introduce seasonal passes or packages specifically for peak months, encouraging casual riders to purchase longer term access during the high demand period.
- Provide enhance services such as guided tours, themed rides, or partnerships with local attractions during the peak months to capitalize on the increased interest.
- Consider scaling services e.g., additional bike stations, more docking points during the high demand season and potentially reducing them during the off-peak months to optimize resource allocation.

Recommendations for popular docking stations and routes among casual riders with long ride durations

- Ensure a higher number of bikes are available at these popular docking stations to meet the increased demand from casual riders, especially during peak times and weekends.
- Improve the amenities at the docking stations to enhance the user experience. This can include better lighting, seating areas, informational kiosks, and bike maintenance tools.
- Utilize these high-traffic stations as venues for promoting membership benefits. Install signage and distribute brochures that highlight the advantages of becoming a member.
- Introduce special discounts or promotional offers for rides starting or ending at these popular docking stations to attract more casual riders.

