

Bike-Sharing Service's Clients

Anna Savinova

2023-12-20

Business task

Since maximizing the number of annual memberships will ensure Cyclistic's future success, the Chicago-based bike-sharing service's new marketing strategy aims at converting casual riders into annual members. To develop a strategy, there is a need to understand the differences between these types of clients.

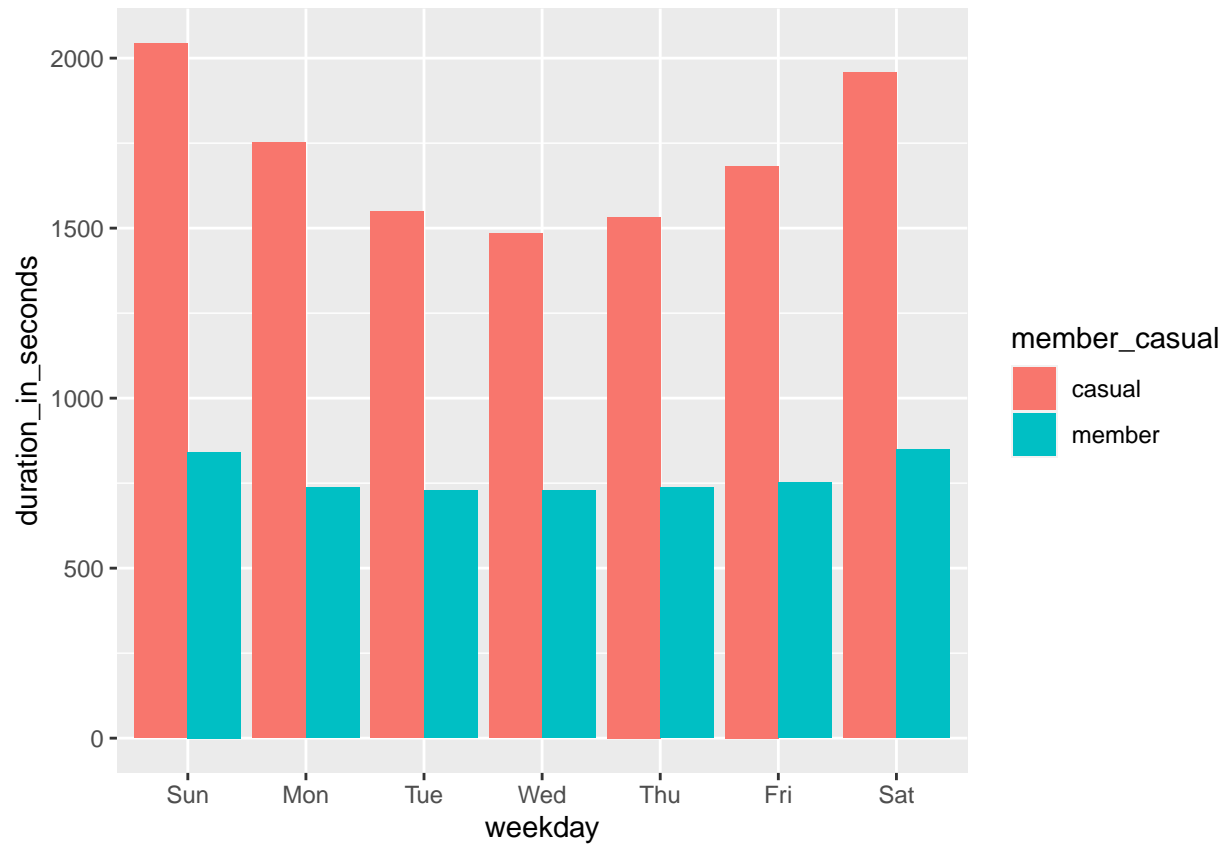
Recommendations

developed by the author based on data from the New York bicycle-sharing system Motivate International Inc., as well as information about the weather forecast and federal holidays in the United States:

- The difference between the number of trips of casual riders in January and July by almost **22 times** indicates the need to increase interest in bike-sharing services in winter - for example, through collaboration with coffee shops.
- The **16 times** drop in demand for bike-sharing services at Christmas among casual riders and the **12 times** drop among annual members (compared to December 21) requires the creation of a "New Year's offer" – for example, a holiday rates.
- Since a decline in demand is inevitable due to extreme weather, as evidenced by the **3.5 times** drop in the number of trips because of the flood on September 11, interest in the service can be maintained by synchronizing the app with a weather forecast.
- To strengthen interest of casual riders in bicycle-sharing service is possible by synchronizing the app with a city traffic map, as well as by putting forward the slogan "bicycle is a freedom of movement" that is relevant to them, as evidenced by the increase of demand of casual riders on days of citywide events (Chicago Marathon day saw **the highest number of rides for the entire month of October**).
- To use the everyday nature of members traffic as a basis for changing the user interest of casual riders - for example, to demonstrate them benefits of commuting to work by bicycle.
- The first-line target audience is among casual users who maintain healthy lifestyle and live in the office areas of Chicago.

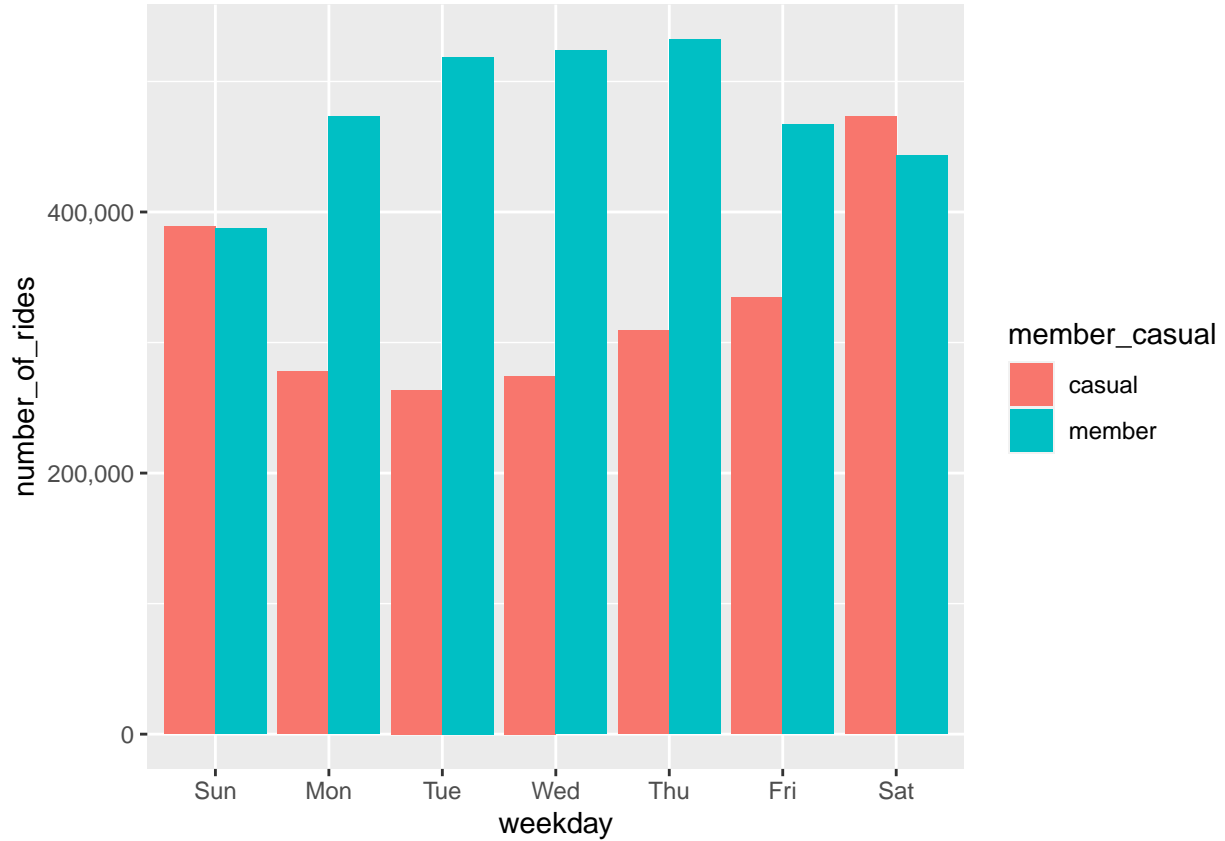
These recommendations are based on the following findings:

*1. The average trip duration of a casual rider exceeds the duration of a member's trip by **2.3 times** (for a casual rider the average trip duration is 29 minutes 15 seconds, for a member – 12 minutes 43 seconds).*



Membership	Avg duration
casual	0:29:15
member	0:12:43

2. Annual members use bike-sharing service more often than casual riders, since the number of trips they make on weekdays is **1.7 times** greater than that of casual riders, and the difference on weekends is not decisive.



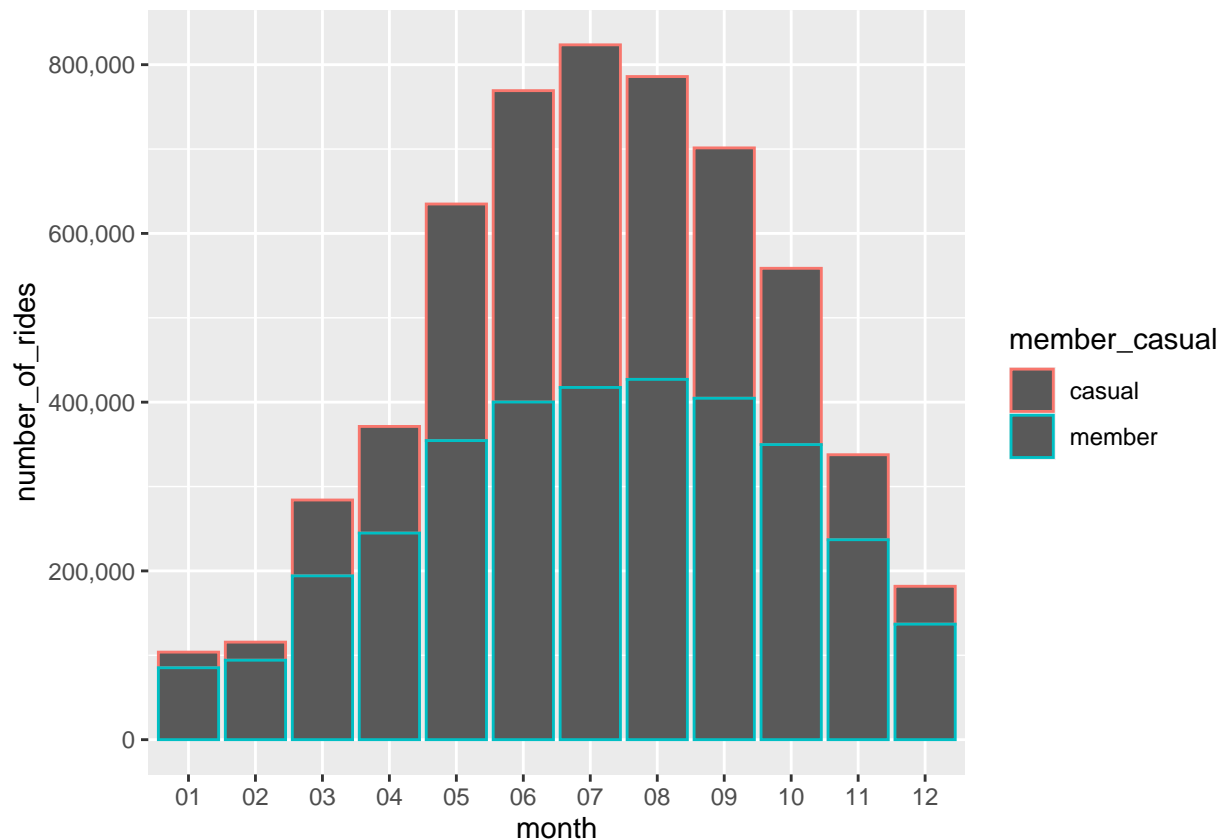
Membership	Weekday	Number_of_rides
casual	Sunday	389036
casual	Monday	277675
casual	Tuesday	263746
casual	Wednesday	274354
casual	Thursday	309330
casual	Friday	334701
casual	Saturday	473190
member	Sunday	387223
member	Monday	473339
member	Tuesday	518626
member	Wednesday	523869
member	Thursday	532261
member	Friday	467086
member	Saturday	443281

3. Based on the above graphs, the use nature of bicycle-sharing service by two categories of clients can be characterized as follow:

- Maintaining the average duration of trips throughout the week with a slight increase on Saturday and Sunday, as well as their regularity and short duration, suggests that the use of a bicycle by annual members is part of their lifestyle and covers basic needs such as commuting, shopping, going to cafes.
- The increase in the average duration of trips and their number on weekends compared to weekdays suggests the entertaining nature of the service using by some casual riders.

4. A possible “portrait” of an annual member is a person living near the office and prone to a healthy lifestyle.
5. There is a dependence between the use of bike sharing services and weather:

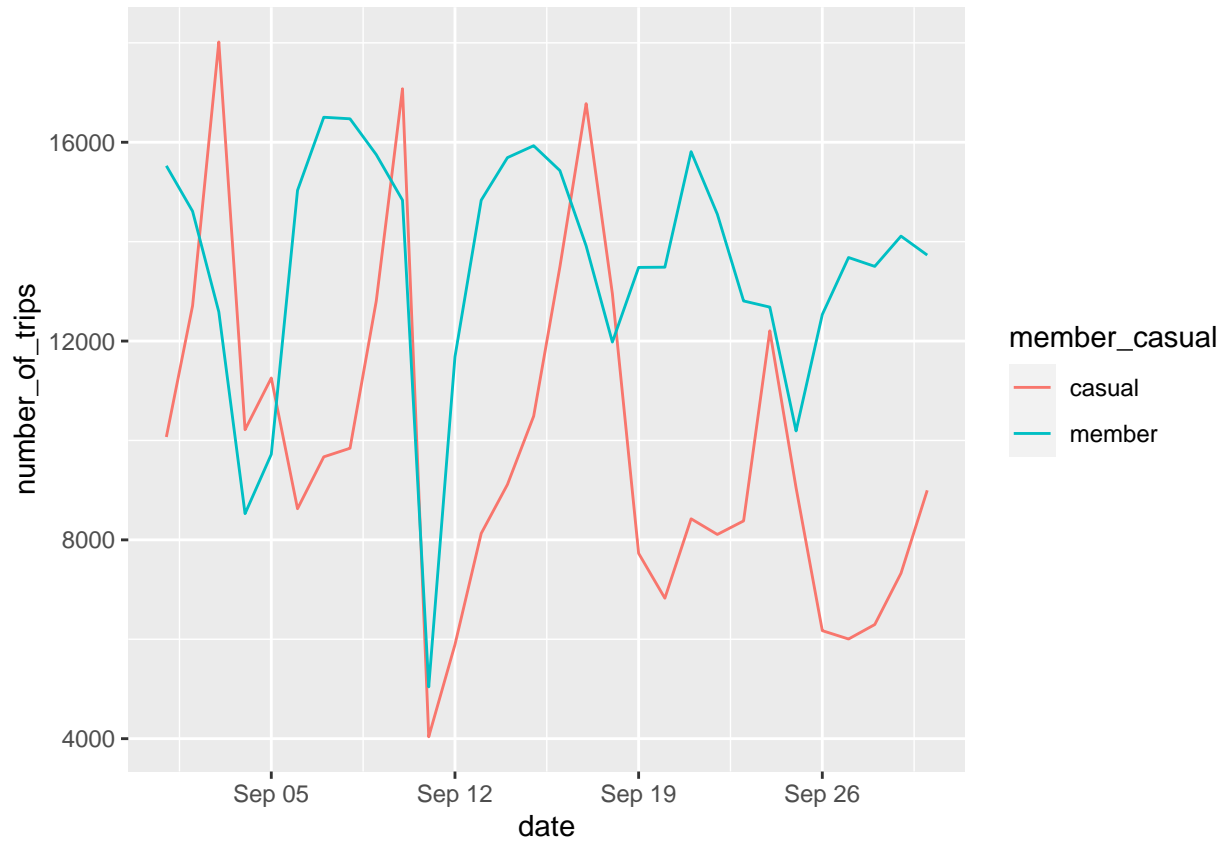
- The number of trips made by casual riders in the 2nd quarter increased almost by **6 times** compared to the 1st quarter, members – by **2.7 times**, with an average temperature in the 1st quarter of -4 degrees and in the 2nd quarter of +15.6. At the same time, there is **3 times** decrease in the number of trips of casual riders in the 4th quarter compared to the 3rd quarter, with an average temperature in the 3rd quarter of +22, and in the 4th quarter of +5. Even the average trip duration of a casual rider decreased in the 4th quarter.



Quarter	Membership	Number of trips
1 quarter	casual	129818
	member	373603
2 quarter	casual	775883
	member	999428
3 quarter	casual	1061676
	member	1249083
4 quarter	casual	354655
	member	723571

- The most popular month by the number of trips was July (823,488 trips) with an average temperature of +24, the most unpopular month by the number of trips was January (103,770) with an average temperature of -6.

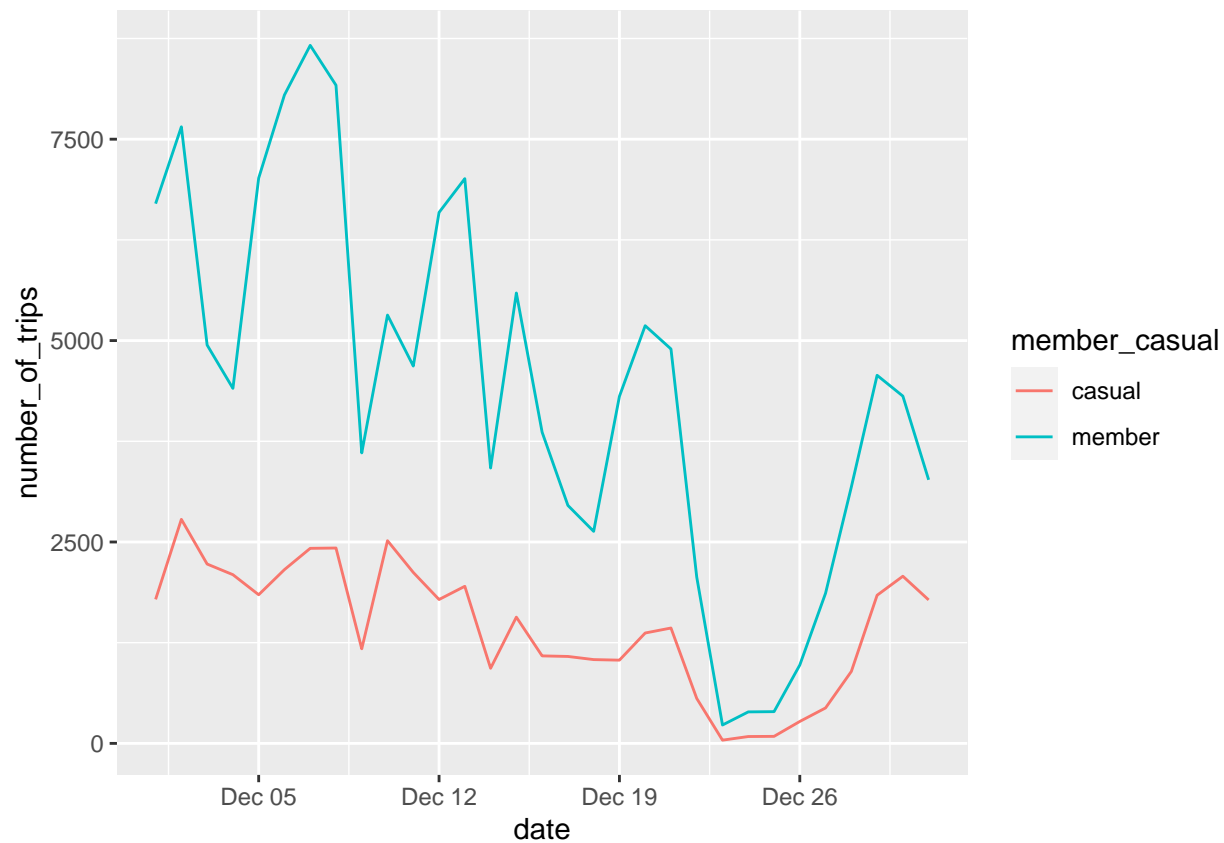
- A **3.5 times** drop in the total number of trips on September 11 compared to the day before was likely due to the flood.



Date	Number
Sept 10	31911
Sept 11	9080
Sept 12	17577

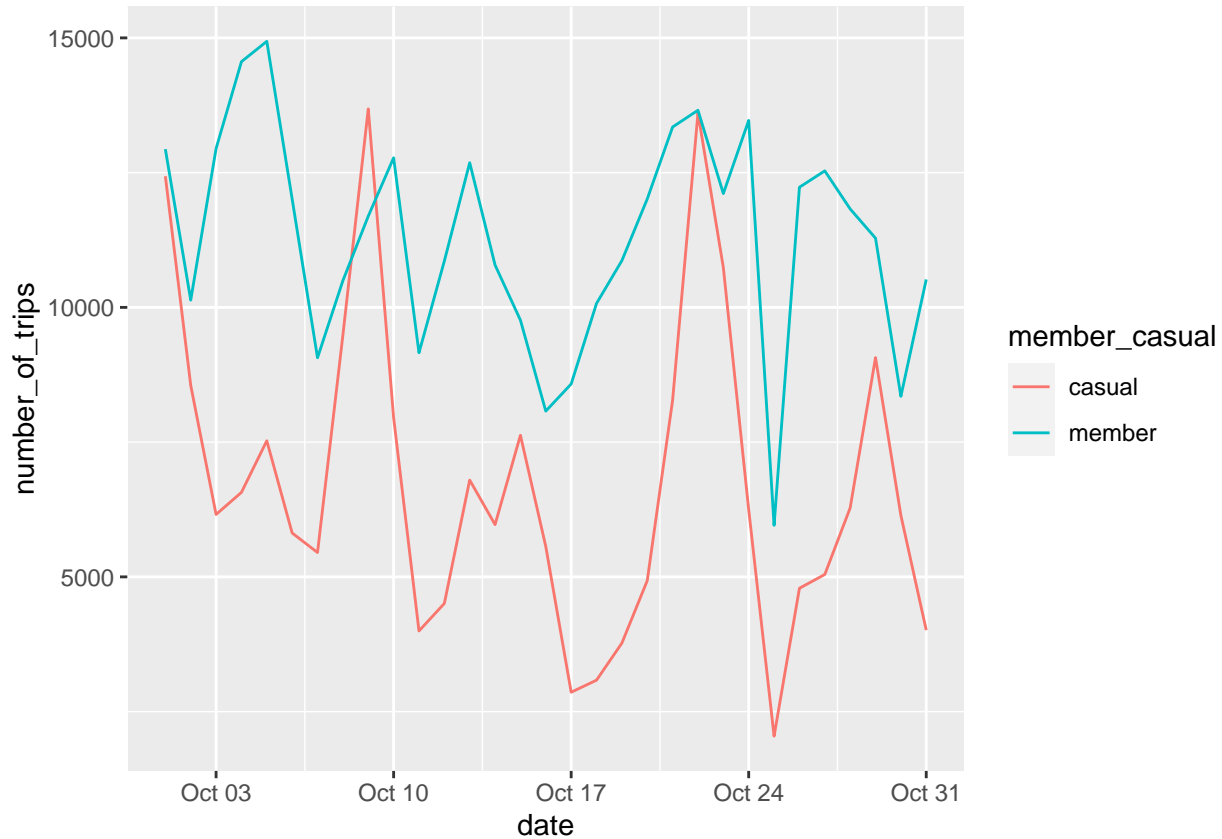
6. *There is a relationship between the use of bike sharing services and holidays/city events:*

- a significant drop in the number of trips for both types of users on the eve and on Christmas Day, with a consistent increase immediately after December 25, 2022.



Date	Membership	Number
Dec 21	casual	1432
	member	4896
Dec 22	casual	558
	member	2065
Dec 23	casual	40
	member	228
Dec 24	casual	85
	member	391
Dec 25	casual	87
	member	394
Dec 26	casual	272
	member	973

- Increased demand on the day of the Chicago Marathon on October 9, 2022 from casual riders.



Given the traditional increase in the number of trips of casual riders on weekends, demand on Sunday, October 9, the day of the Chicago Marathon, can still be called peak, since this day saw the highest number of trips for the entire October (13,683).

The second peak (13,608) occurred on October 22, which may have been due to the unusually warm weather for this time of year in Chicago: +21 degrees with an average temperature for this month of +12.

Characteristics of data sources

1. Data for this study was provided by Motivate International Inc., a New York- based bicycle sharing service. The data used is available to the public under a data license agreement. The data sets located on the website in the form of CSV files are data from a real-life Chicago bicycle sharing service. The materials used meet the requirements of this study as they are reliable and authentic. Data shortcomings: inconsistency (column names differ in files for different years), incompleteness (quarterly data for 2021 and 2022 is missing, although monthly data for these periods is available).
2. Chicago weather data for 2022 was taken from the National Weather Service website.
3. Data on federal holidays in the United States were taken from the official US government website.

Data analysis

Excel

- Prepared data: saved it in XLS format; unified column names; checked if all IDs consist of 8 digits (= LEN and conditional formatting); checked for spaces in columns with text data (=TRIM); filled empty

cells (“find and replace” or filter - select empty cells - remove format - “find and highlight” empty cells - fill using CTRL and Enter).

- Calculated:
 - ride_length - duration of each ride (subtracted start_time from end_time, formatted as time in hours, minutes and seconds).
 - day_of_week - on what day of the week each trip began (=WEEKDAY).
 - mean_ride_length - average ride duration (=AVERAGE).
 - max_ride_length - maximum ride duration (=MAX).
 - most_popular_day_of_week - day of the week on which bicycles are most often taken (=MODE).
- Created pivot tables: AVG_for_usertype (average trip duration for a member and a casual rider), AVG_by_day_of_week (average trip duration for a member and a casual rider by day of the week), Number_of_rides_by_day_of_week (number of trips by day of the week).
- Created graphs for pivot tables.

Big Query

1. *Imported CSV files into Big Query using Google Cloud Storage.*
2. *Calculated the number of trips and the average trip duration of casual riders and members for each month separately:*

```
SELECT member_casual,
COUNT(ride_id) as number_of_trips_jan,
AVG(ended_at - started_at) as avg_ride_length_jan
FROM 2022_tripdata.2022_01
GROUP BY member_casual
```

Membership	Number of trips JAN	AVG ride length JAN
casual	18520	0:30:23
member	85250	0:11:59

3. *Calculated the number of trips and the average trip duration of casual riders and members by quarter:*

```
SELECT
SUM(number_of_trips) as number_of_trips_1_quarter,
SUM(total_ride_length)/ SUM(number_of_trips) as avg_ride_length_1_quarter FROM
(
SELECT
COUNT(ride_id) as number_of_trips,
SUM(ended_at - started_at) as total_ride_length
FROM my-project-cyclistic-404211.2022_tripdata.2022_01
WHERE
member_casual = 'member'
UNION ALL
```



```

SELECT
COUNT(ride_id) as number_of_trips,
SUM(ended_at - started_at) as total_ride_length
FROM my-project-cyclistic-404211.2022_tripdata.2022_02
WHERE
member_casual = 'member'
UNION ALL
SELECT
COUNT(ride_id) as number_of_trips,
SUM(ended_at - started_at) as total_ride_length
FROM my-project-cyclistic-404211.2022_tripdata.2022_03
WHERE
member_casual = 'member'
)

```

Number of trips 1 quarter	AVG ride length 1 quarter
373603	0:11:49

4. *Calculated the number of trips made on each day of the month:*

```

SELECT
EXTRACT(day from started_at) as day_of_month,
COUNT(ride_id) as number_of_trips
FROM my-project-cyclistic-404211.2022_tripdata.2022_01
GROUP BY day_of_month
ORDER BY day_of_month

```

Day of month JAN	Number of trips JAN
1	2563
2	2062
3	3135
...	...

5. *Calculated the number of trips made by both types of users on each day of the month:*

```

SELECT
member_casual,
COUNT(ride_id) as number_of_trips,
EXTRACT(day from started_at) as day_of_month
FROM my-project-cyclistic-404211.2022_tripdata.2022_09

```

GROUP BY member_casual, day_of_month
ORDER BY day_of_month

Membership	Number of trips SEP	Day of month SEP
casual	10070	1
member	15526	1
casual	12710	2
member	14612	2
...

6. *Calculated which month was the most popular by the number of trips:*

```

SELECT
COUNT(ride_id) as number_of_trips,
EXTRACT(month from started_at) as month
FROM my-project-cyclistic-404211.2022_tripdata.2022_01
GROUP BY month
UNION ALL
SELECT
COUNT(ride_id) as number_of_trips,
EXTRACT(month from started_at) as month
FROM my-project-cyclistic-404211.2022_tripdata.2022_02
GROUP BY month
UNION ALL
SELECT
COUNT(ride_id) as number_of_trips,
EXTRACT(month from started_at) as month
FROM my-project-cyclistic-404211.2022_tripdata.2022_03
GROUP BY month
... (repeat for all months)
ORDER BY month

```

Number of trips	Month
103770	1
115609	2
284042	3
371249	4
634858	5
769204	6
823488	7
785932	8
701339	9

Number of trips	Month
558685	10
337735	11
181806	12

R Studio

```
#setting up my environment
install.packages("tidyverse")
install.packages("lubridate")
install.packages("ggplot2")
library(tidyverse)
library(lubridate)
library(ggplot2)

#loaded data files and created data frames for 12 months
m1_2022<-read_csv("/Users/annasavinova/Desktop/my_project/csv_files/202201-divvy-tripdata.csv")
m2_2022<-read_csv("/Users/annasavinova/Desktop/my_project/csv_files/202202-divvy-tripdata.csv")
m3_2022<-read_csv("/Users/annasavinova/Desktop/my_project/csv_files/202203-divvy-tripdata.csv")

and etc.

#combined all months into one document
all_trips<-bind_rows(m1_2022, m2_2022, m3_2022, m4_2022, m5_2022, m6_2022, m7_2022, m8_2022, m9_2022, m10_2022, m11_2022, m12_2022)

#removed unnecessary columns
all_trips<-all_trips %>%
select(-c(start_lat, start_lng, end_lat, end_lng))

#checked the new table
head(all_trips)
colnames(all_trips)
dim(all_trips)
str(all_trips)
summary(all_trips)

#added columns - day, month, year of trip, day of week
all_trips$date <- as.Date(all_trips$started_at)
all_trips$month <- format((all_trips$date), "%m")
all_trips$day <- format((all_trips$date), "%d")
all_trips$year <- format((all_trips$date), "%Y")
all_trips$day_of_week <- format(as.Date(all_trips$date), "%A")

#added trip duration
all_trips$ride_length <- difftime(all_trips$ended_at,all_trips$started_at)

#converted it into numerical data
all_trips$ride_length <- as.numeric(all_trips$ride_length)

#checked for NA
na_rows<-which(is.na(all_trips$ride_length))
```

```

print(all_trips[na_rows, ])

#checked for negative values
negative_rows<-which(all_trips$ride_length<0)
print(all_trips[negative_rows, ])

#switched columns 'started_at' and 'ended_at' for the rows where 'ride_length' is negative
all_trips[negative_rows, c("started_at", "ended_at")] <- all_trips[negative_rows, c("ended_at", "started_at")]

#calculated trip duration ensuring the correct conversion of 'ride_length' to numeric
all_trips$ride_length <- as.numeric(as.character(difftime(all_trips$ended_at,all_trips$started_at)))

#calculated the average trip duration in seconds
mean(all_trips$ride_length)

#calculated the median
median(all_trips$ride_length)

#calculated the longest trip
max(all_trips$ride_length)

#calculated the shortest trip
min(all_trips$ride_length)

#Compared members and casual users
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = mean)
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = median)
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = max)
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = min)

#calculated the average trip time by day for members and casual riders
aggregate(all_trips$ride_length ~ all_trips$member_casual + all_trips$day_of_week, FUN = mean)

#arranged the weekdays in order
all_trips$day_of_week<-ordered(all_trips$day_of_week, levels = c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
aggregate(all_trips$ride_length ~ all_trips$member_casual + all_trips$day_of_week, FUN = mean)

#analyzed data by type and day of week
all_trips %>%
mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)

#visualized the number of trips by user type
all_trips %>%
mutate(weekday = wday(started_at, label = TRUE)) %>% group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(),
average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday) %>%
ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  scale_y_continuous(name="number_of_rides", labels = scales::comma)

```

```

#visualized the average trip duration by user type
all_trips %>%
mutate(weekday = wday(started_at, label = TRUE)) %>% group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(),
average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday) %>%
ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge") +
labs(y = "duration_in_seconds")

#visualized the number of trips by user type in a particular month
all_trips %>%
filter(month == "12") %>%
group_by(member_casual, date) %>%
summarise(number_of_trips = n(), .groups = 'drop') %>% ggplot(aes(x = date, y = number_of_trips, color
  geom_line()

#visualized the number of trips by user type for the entire year
all_trips %>%
group_by(member_casual, month) %>%
summarise(number_of_trips = n(), .groups = 'drop') %>%
ggplot(aes(x = month, y = number_of_trips, color = member_casual)) +
  geom_col() +
  scale_y_continuous(name="number_of_rides", labels = scales::comma)

#saved a csv file in the project folder that can be visualized in other programs for further analysis
counts<-aggregate(all_trips$ride_length ~ all_trips$member_casual + all_trips$day_of_week, FUN = mean)
write.csv(counts, file = '~/Desktop/my_project/avg_ride_length.csv')

```