## Introduction

- Welcome to the Cyclistic bike-share analysis case study!

- In this case study, you work for a fictional company, Cyclistic, along with some key team members.

- In order to answer the business questions, follow the steps of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act.

- Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path.

---------------------------------------------------------------------------------------------------------

## Scenario

- You are a junior data analyst working on 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 **diferently**.

- 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 the bikes 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 solid opportunity 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 team needs to beer 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.

-------------------------------------------------------------------------------------------------------

Ask

Three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes dierently?

2. Why would casual riders buy Cyclistic annual memberships?

3. How can Cyclistic use digital media to influence casual riders to become members?

Moreno has assigned you the first question to answer:

**How do annual members and casual riders use Cyclistic bikes dierently?**

You will produce a report with the following deliverables:

1. A clear statement of the business task

2. A description of all data sources used

3. Documentation of any cleaning or manipulation of data

4. A summary of your analysis

5. Supporting visualizations and key findings

6. Your top three recommendations based on your analysis

# Case Study: How does a bike-share navigate speedy success?"

# Background

This analysis is for case study 1 from the Google Data Analytics Certificate (Cyclistic). It's originally based on the case study "'Sophisticated, Clear, and Polished': Divvy and Data Visualization" written by Kevin Hartman click here to visit the original blog post. We will be using the Divvy dataset for the case study.

## Purpose

The purpose of this script is to consolidate downloaded Divvy data into a single dataframe and then conduct simple exploratory data analysis to help answer the key question: *"In what ways do members and casual riders use Divvy bikes differently?"*

Use the following Case Study Roadmap as a guide. Note: Completing this case study within a week is a reasonable goal.

| Case Study Roadmap - Ask |
| --- |
| **Guiding questions**<br>• What is the problem you are trying to solve?<br>• How can your insights drive business decisions? |
| **Key tasks**<br>• Identify the business task<br>• Consider key stakeholders |
| **Deliverable**<br>• A clear statement of the business task |

# Business Task

To analyze user behaviours on how annual members and casual riders use Cyclistic bikes differently to make recommendations on how to convert casual riders into annual members

----------------------------------------------------------------------------------------------------------

Prepare

- Use Cyclistic's historical trip data to analyze and identify trends.

- Download the previous 12 months of Cyclistic trip data here.

- (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.)

- 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 identiable 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.

- Now, prepare your data for analysis using the following Case Study Roadmap as a guide:

| **Case Study Roadmap - Prepare** |
|---|
| **Guiding questions**<br>• Where is your data located?<br>• How is the data organized?<br>• Are there issues with bias or credibility in this data? Does your data ROCCC?<br>• How are you addressing licensing, privacy, security, and accessibility?<br>• How did you verify the data's integrity?<br>• How does it help you answer your question?<br>• Are there any problems with the data? |
| **Key tasks**<br>• Download data and store it appropriately.<br>• Identify how it's organized.<br>• Sort and filter the data.<br>• Determine the credibility of the data. |
| **Deliverable**<br>• A description of all data sources used |

## Information about the dataset

Regarding the dataset, we have the following information:

- It is external data stored in the cloud
- It used the long data format
- First-party data (reliable and original)
- There is a file for each past 12 months (current)
- It is distributed under a license
- There is not Personally Identifiable Information (PII)

## Data acquisition process

The following process is utilized:

- Each dataset is downloaded
- Appropriately stored in a folder for original datasets

## Identifying

### In order to identify issues with the data, we:

- Evaluate the ride length and spot unusual observations
- Filtered the data and identified missing values
- Sorted the data and found inconsistent attribute format

-----------------------------------------------------------------------------------------------------------

Process

Then, process your data for analysis using the following Case Study Roadmap as a guide:

| **Case Study Roadmap - Process** |
| --- |
| **Guiding questions**<br>• What tools are you choosing and why?<br>• Have you ensured your data's integrity?<br>• What steps have you taken to ensure that your data is clean?<br>• How can you verify that your data is clean and ready to analyze?<br>• Have you documented your cleaning process so you can review and share those results? |
| **Key tasks**<br>• Check the data for errors.<br>• Choose your tools.<br>• Transform the data so you can work with it effectively.<br>• Document the cleaning process. |

| **Deliverable** |
| --- |
| • Documentation of any cleaning or manipulation of data |

Follow these steps:

1. Download the previous 12 months of trip data.

- Note: If you are planning on using Posit's RStudio, use the Divvy 2019 Q1 and Divvy 2020 Q1 datasets.

- Choosing other data might lead to errors because the data exceeds the memory available in the free plan.

2. Unzip the files.

3. Create a folder on your desktop or Drive to house the files.

- Use appropriate file-naming conventions.

4. Create subfolders for the .csv file and the .xls or Sheets file so that you have a copy of the original data.

- Move the downloaded files to the appropriate subfolder.

5. Follow these instructions for either Excel (a) or Google Sheets (b):

a. Launch Excel, open each file, and choose to Save As an Excel Workbook file.

- Put it in the subfolder you created for .xls files.

b. Open each .csv file in Google Sheets and save it to the appropriate subfolder.

6. Open your spreadsheet and create a column called ride_length.

- Calculate the length of each ride by subtracting the column started_at from the column ended_at (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 373055.

7. Create a column called day_of_week, and calculate the day of the week that each ride started using the WEEKDAY command (for example, =WEEKDAY(C2,1)) in each file.

- Format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday.

8. Proceed to the analyze step.

- If you like, continue working with the data to beer familiarize yourself, and perhaps even identify new approaches to answering the business questions.

-----------------------------------------------------------------------------------------------------------

Analyze

Now that your data is stored appropriately and has been prepared for analysis, start pushing it to work. Use the following Case Study Roadmap as a guide:

| Case Study Roadmap - Analyze |
| --- |
| **Guiding questions**<br>• How should you organize your data to perform analysis on it?<br>• Has your data been properly formatted?<br>• What surprises did you discover in the data?<br>• What trends or relationships did you find in the data?<br>• How will these insights help answer your business questions? |
| **Key tasks**<br>• Aggregate your data so it's useful and accessible.<br>• Organize and format your data.<br>• Perform calculations.<br>• Identify trends and relationships. |
| **Deliverable**<br>• A summary of your analysis |

Follow these steps for using spreadsheets

Open your spreadsheet application, then complete the following steps:

1. Where relevant, make columns consistent and combine them into a single worksheet.

2. Clean and transform your data to prepare for analysis.

3. Conduct descriptive analysis.

4. Run a few calculations in one file to get a better sense of the data layout. Options:

- Calculate the mean of ride_length

- Calculate the max ride_length

- Calculate the mode of day_of_week

5. Create a pivot table to quickly calculate and visualize the data. Options:

- Calculate the average ride_length for members and casual riders. Try rows = member_casual; Values = Average of ride_length.

- Calculate the average ride_length for users by day_of_week. Try columns = day_of_week; Rows = member_casual; Values = Average of ride_length.

- Calculate the number of rides for users by day_of_week by adding Count of trip_id to Values.

6. Open another file and perform the same descriptive analysis steps. Explore different seasons to make some initial observations.

7. Once you have spent some time working with the individual spreadsheets, merge them into a full-year view.

Do this with the tool you have chosen to use to perform your final analysis, either a spreadsheet, a database and SQL, or R Studio.

8. Export a summary file for further analysis.

Follow these steps for using SQL

Open your SQL tool of choice, then complete the following steps:

1. Import your data.

2. Explore your data, perhaps looking at the total number of rows, distinct values, maximum, minimum, or mean values.

3. Where relevant, use JOIN statements to combine your relevant data into one table.

4. Create summary statistics.

5. Investigate interesting trends and save that information to a table.

Follow these steps for using R

Open your preferred version of R, click this link, and select “Use template.” Then, copy and paste the text from the template into an R script.

1. Import your data from Divvy 2019 Q1 and Divvy 2020 Q1.

2. Make columns consistent and merge them into a single dataframe.

3. Clean up and add data to prepare for analysis.

4. Conduct descriptive analysis.

5. Export a summary file for further analysis.

-------------------------------------------------------------------------------------------------------------

Share

- Now that you have performed your analysis and gained some insights into your data, create visualizations to share your findings.

- Moreno has reminded you that they should be sophisticated and polished in order to efectively communicate to the executive team.

- Use the following Case Study Roadmap as a guide:

| Case Study Roadmap - Share |
|---|
| **Guiding questions**<br>• Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?<br>• What story does your data tell?<br>• How do your findings relate to your original question?<br>• Who is your audience? What is the best way to communicate with them?<br>• Can data visualization help you share your findings?<br>• Is your presentation accessible to your audience? |
| **Key tasks**<br>• Determine the best way to share your findings.<br>• Create effective data visualizations.<br>• Present your findings.<br>• Ensure your work is accessible. |
| **Deliverable**<br>• Supporting visualizations and key findings |

Follow these steps:

1. Take out a piece of paper and a pen and sketch some ideas for how you will visualize the data.

2. Once you choose a visual form, open your tool of choice to create your visualization.

Use a presentation soware, such as PowerPoint or Google Slides, your spreadsheet program, Tableau, or R.

3. Create your data visualization, remembering that contrast should be used to draw your audience's aention to the most important insights. Use artistic principles including size, color, and shape.

4. Ensure clear meaning through the proper use of common elements, such as headlines, subtitles, and labels.

5. Rene your data visualization by applying deep aention to detail.

--------------------------------------------------------------------------------------------------------------

Act

Now that you have finished creating your visualizations, act on your findings.

Prepare the deliverables Morena asked you to create, including the three top recommendations based on your analysis.

Use the following Case Study Roadmap as a guide:

| **Case Study Roadmap - Act** |
| --- |
| **Guiding questions**<br>• What is your final conclusion based on your analysis?<br>• How could your team and business apply your insights?<br>• What next steps would you or your stakeholders take based on your findings?<br>• Is there additional data you could use to expand on your findings? |
| **Key tasks**<br>• Create your portfolio.<br>• Add your case study.<br>• Practice presenting your case study to a friend or family member. |
| **Deliverable**<br>• Your top three recommendations based on your analysis |

Follow these steps:

1. If you do not have one already, create an online porolio.

2. Consider how you want to feature your case study in your porolio.

3. Upload or link your case study findings to your portfolio.

4. Write a brief paragraph describing the case study, your process, and your discoveries.

5. Add the paragraph to introduce your case study in your portfolio.

-----------------------------------------------------------------------------------------------------------

Wrap-up

- Congratulations on nishing the Cyclistic bike-share case study!
- If you like, complete another case study to continue growing your porolio.
- Or, use the steps from the Ask, Prepare, Process, Analyze, Share, and Act Case Study Roadmap to create a new project that's all your own.
- Best of luck on your job search!

-----------------------------------------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------

# Process

Here, we perform the data cleaning, ensure integrity, and that it is complete, correct, and relevant.

## Tools

We selected tools to perform specific tools, as described next.

- Spreadsheets: perform initial data inspection
- R: develop scripts to clean, transform, organize, and summarize the datasets
- Tableau: create data visualizations
- R Notebook: provide a complete report of the data analysis process

# Install Required Packages

- tidyverse for data import and wrangling
- libridate for date functions
- ggplot for visualization

```
library(tidyverse)  #tidyverse includes core packages like ggplot2 and readr
                    #which are helpful to wrangle data

library(lubridate)  #helps wrangle date attributes

library(ggplot2)  #helps visualize data
```

# Exploratory Data Analysis

## Step 1: Collect Data - Download Data and store it appropriately

### Upload Divvy datasets

```
df_q2_2019 <- read_csv("../input/last-12-months-data-of-cyclistic-trip-data/Divvy_Trips_20

df_q3_2019 <- read_csv("../input/last-12-months-data-of-cyclistic-trip-data/Divvy_Trips_20

df_q4_2019 <- read_csv("../input/last-12-months-data-of-cyclistic-trip-data/Divvy_Trips_20

df_q1_2020 <- read_csv("../input/last-12-months-data-of-cyclistic-trip-data/Divvy_Trips_20
```

## Step 2: Data Wrangling

### Compare column names each of the files

While the names don't have to be in the same order, they DO need to match perfectly before we can use a command to join them into one file.

```
colnames(df_q2_2019)

colnames(df_q3_2019)

colnames(df_q4_2019)

colnames(df_q1_2020)
```

## Metrics

*After comparing column names, we can infer that there are 13 variables/columns available, below is the metadata:*

**ride_id** : Unique id of each ride trip

**rideable_type** : type of bicycle ridden, split between 3 categories - classic, docked and electric

**started_at** : date and time of the start of the trip

**ended_at** : date and time of the end of the trip

**start_station_name** : Start station name

**start_station_id** : Start station id

**end_station_name** : End station name

**end_station_id** : End station id

**start_lat** : latitude of the start location

**start_lng** : longitude of the start location

**end_lat** : latitude of the end location

**end_lng** : longitude of the end location

**member_casual** : type of membership, either casual or member

## Rename columns

```
df_q4_2019 <- rename(df_q4_2019
                     ,ride_id = trip_id
                     ,rideable_type = bikeid
                     ,started_at = start_time
                     ,ended_at = end_time
                     ,start_station_name = from_station_name
                     ,start_station_id = from_station_id
                     ,end_station_name = to_station_name
                     ,end_station_id = to_station_id
                     ,member_casual = usertype)
```

```
df_q3_2019 <- rename(df_q3_2019
                    ,ride_id = trip_id
                    ,rideable_type = bikeid
                    ,started_at = start_time
                    ,ended_at = end_time
                    ,start_station_name = from_station_name
                    ,start_station_id = from_station_id
                    ,end_station_name = to_station_name
                    ,end_station_id = to_station_id
                    ,member_casual = usertype)
```

```
df_q2_2019 <- rename(df_q2_2019
                    ,ride_id = "01 - Rental Details Rental ID"
                    ,rideable_type = "01 - Rental Details Bike ID"
                    ,started_at = "01 - Rental Details Local Start Time"
                    ,ended_at = "01 - Rental Details Local End Time"
                    ,start_station_name = "03 - Rental Start Station Name"
                    ,start_station_id = "03 - Rental Start Station ID"
                    ,end_station_name = "02 - Rental End Station Name"
                    ,end_station_id = "02 - Rental End Station ID"
                    ,member_casual = "User Type")
```

## Inspect the dataframes & look for discrepancy

```
str(df_q1_2020)

str(df_q4_2019)

str(df_q3_2019)

str(df_q2_2019)
```

### Convert ride_id and rideable_type to character so that they can stack correctly

```
df_q4_2019 <-  mutate(df_q4_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
df_q3_2019 <-  mutate(df_q3_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
df_q2_2019 <-  mutate(df_q2_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
```

## Combining dataframes into one big dataframe

### # Stack individual quarter's data frames into one big data frame

```
all_trips <- bind_rows(df_q2_2019, df_q3_2019, df_q4_2019, df_q1_2020)
```

### Remove lat, long, birthyear, and gender fields as this data was dropped beginning in 2020

```
all_trips <- all_trips %>%
  select(-c(start_lat, start_lng, end_lat, end_lng, birthyear, gender, "01 - Rental Detai
```

```
"01 - Rental Details Duration In Seconds Uncapped", "05 - Member Details Member Birthday
```

```
Uncapped", "05 - Member Details Member Birthday Year", "Member Gender", "tripduration"))
```

## Step 3: Clean up and prepare data for analysis

### Inspect the new table that has been created

```
# Inspect the new table that has been created

colnames(all_trips)  #List of column names

nrow(all_trips)  #How many rows are in data frame?

dim(all_trips)  #Dimensions of the data frame?

head(all_trips)  #See the first 6 rows of data frame.  Also tail(qs_raw)

str(all_trips)  #See list of columns and data types (numeric, character, etc)

summary(all_trips)  #Statistical summary of data. Mainly for numerics
```

## Assumption and limitation

### There are a few problems we will need to fix:

- In the "member_casual" column, there are two names for members ("member" and "Subscriber") and two names for casual riders ("Customer" and "casual"). We will need to consolidate that from four to two labels.
- The data can only be aggregated at the ride-level, which is too granular. We will want to add some additional columns of data – such as day, month, year – that provide additional opportunities to aggregate the data.
- We will want to add a calculated field for length of ride since the 2020Q1 data did not have the "tripduration" column. We will add "ride_length" to the entire dataframe for consistency.
- There are some rides where tripduration shows up as negative, including several hundred rides where Divvy took bikes out of circulation for Quality Control reasons. We will want to delete these rides.

## Consolidating in just two labels

```
# Consolidating in just two labels
all_trips <-  all_trips %>%
  mutate(member_casual = recode(member_casual
                                ,"Subscriber" = "member"
                                ,"Customer" = "casual"))

# Check to make sure the proper number of observations were reassigned

table(all_trips$member_casual)
```

```
# Adding columns that list the date, month, day, and year of each ride
# This will allow us to aggregate ride data for each month, day, or year
#... before completing these operations we could only aggregate at the ride level
# https://www.statmethods.net/input/dates.html more on date formats in R found at that link
all_trips$date <- as.Date(all_trips$started_at) #The default format is yyyy-mm-dd
all_trips$month <- format(as.Date(all_trips$date), "%m")
all_trips$day <- format(as.Date(all_trips$date), "%d")
all_trips$year <- format(as.Date(all_trips$date), "%Y")
all_trips$day_of_week <- format(as.Date(all_trips$date), "%A")


# Adding a "ride_length" calculation to all_trips (in seconds)
# https://stat.ethz.ch/R-manual/R-devel/library/base/html/difftime.html
all_trips$ride_length <- difftime(all_trips$ended_at,all_trips$started_at)


# Inspect the structure of the columns
str(all_trips)
```

```
# Convert "ride_length" from Factor to numeric so we can run calculations on the data
is.factor(all_trips$ride_length)

all_trips$ride_length <- as.numeric(as.character(all_trips$ride_length))
is.numeric(all_trips$ride_length)


# Remove "bad" data
# The dataframe includes a few hundred entries when bikes were taken out of docks and
# checked for quality by Divvy or ride_length was negative
# We will create a new version of the dataframe (v2) since data is being removed
# https://www.datasciencemadesimple.com/delete-or-drop-rows-in-r-with-conditions-2/
all_trips_v2 <- all_trips[!(all_trips$start_station_name == "HQ QR" | all_trips$ride_lengt
```

```
"HQ QR" | all_trips$ride_length<0),]
```

## Step 4: It is time to play with numbers - Descriptive analysis

### Statistics in action

```
# Descriptive analysis on ride_length (all figures in seconds)
mean(all_trips_v2$ride_length) #straight average (total ride length / rides)

median(all_trips_v2$ride_length) #midpoint number in the ascending array of ride lengths

max(all_trips_v2$ride_length) #longest ride

min(all_trips_v2$ride_length) #shortest ride


# You can condense the four lines above to one line using summary() on the specific
# attribute
summary(all_trips_v2$ride_length)
```

```
# Compare members and casual users
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = mean)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = median)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = max)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = min)

# See the average ride time by each day for members vs casual users
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual + all_trips_v2$day_of_weel
```

```
+ all_trips_v2$day_of_week, FUN = mean)
```

```
# Notice that the days of the week are out of order. Let's fix that.
all_trips_v2$day_of_week <- ordered(all_trips_v2$day_of_week, levels=c("Sunday", "Monday",
```

```
levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
```

```
# Now, let's run the average ride time by each day for members vs casual users
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual + all_trips_v2$day_of_weel
```

```
casual + all_trips_v2$day_of_week, FUN = mean)
```

```
# analyze ridership data by type and weekday
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%  #creates weekday field using wday(
  group_by(member_casual, weekday) %>%  #groups by usertype and weekday
  summarise(number_of_rides = n()                          #calculates the number of ride
            ,average_duration = mean(ride_length)) %>%      # calculates the average durat
  arrange(member_casual, weekday)
```

```
abel = TRUE)) %>%  #creates weekday field using wday()
%  #groups by usertype and weekday
                    #calculates the number of rides and average duration
ride_length)) %>%      # calculates the average duration
```

## Step 5: Data Visualization

### Let's visualize the number of rides by rider type

```
# Let's visualize the number of rides by rider type
all_trips_v2 %>%
  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") +
  labs(
      title = "Number of rides by rider type",
      subtitle = "Sorted by weekday"
      )
```

#### Let's create a visualization for average duration

```
# Let's create a visualization for average duration
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, month)  %>%
  ggplot(aes(x = month, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title = "Average duration by rider type",
       subtitle = "Sorted by month")
```

# KEY TAKEAWAYS

- Annual members rode on average 1.4 times more than casual riders throughout the year and almost twice as many times on some weekdays.
- Saturday has highest count of rented bikes.
- The maximum ride duration is 2607.5 Hours and the rider who has the maximum ride duration is casual.
- Throughout the week: Casual riders' usage peaks on the weekends while dropping off the rest of the week, Member Riders' usage remains consistent over the week.
- Over the Year “Monthly”: In January and February the average trip duration is highest in Casual riders while the average trip duration remains constant for member riders

# Conclusion

- Casual riders spent more time in bikes and member riders spent less time biking than casual riders.
- On weekends casual bile riders increases. The most popular day for bike renting for all subscribers was Saturday. Sunday was the second most popular day for casual users.
- It seems like casual rider use bikes for picnic and short trips on weekends and casual users use bike for general purpose like for going to school or school.
- The data clearly showed seasonal fluctuations in ridership of all users, both in the number of rides and the average trip duration. It peaked during the summer months and was at its lowest point from the late fall to early spring, especially among casual users.

# Recommendations

- Company may give discount on weekends if casual user take annual membership.Company may Offer weekend passes. This option may lead local users to purchase an annual membership.
- Company may gave more discount on long duration of rides. Company should also offer seasonal passes as well. If the winter season pass is offered at a discount, some casual users might be willing to try it at a fraction of a membership cost before deciding to become year-long subscribers.
- Provide a greener and echo friendly option for rider by Offering discounts to businesses, It might also motivate more casual users to cycle to work more often or all the time.