cyclics

ahmed_el3sfory

2023-07-31

How Does a Bike-Share Navigate Speedy Success?

About the task

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

loading the packages

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.1
```

```
## Warning: package 'lubridate' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(janitor)
```

```
## Warning: package 'janitor' was built under R version 4.3.1
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(lubridate)
```

import the csv files

"we will use read_csv to import csv files into our R script"

```
jun2023 <- read_csv("202306-divvy-tripdata.csv")
```

```
## Rows: 719618 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
may2023 <- read_csv("202305-divvy-tripdata.csv")
```

```
## Rows: 604827 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
apr2023 <- read_csv("202304-divvy-tripdata.csv")
```

```
## Rows: 426590 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
mar2023 <- read_csv("202303-divvy-tripdata.csv")
```

```
## Rows: 258678 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
```

```
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
feb2023 <- read_csv("202302-divvy-tripdata.csv")
```

```
## Rows: 190445 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
jan2023 <- read_csv("202301-divvy-tripdata.csv")
```

```
## Rows: 190301 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dec2022 <- read_csv("202212-divvy-tripdata.csv")
```

```
## Rows: 181806 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
nov2022 <- read_csv("202211-divvy-tripdata.csv")
```

```
## Rows: 337735 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
oct2022 <- read_csv("202210-divvy-tripdata.csv")
```

```
## Rows: 558685 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
sep2022 <- read_csv("202209-divvy-tripdata.csv")
```

```
## Rows: 701339 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
aug2022 <- read_csv("202208-divvy-tripdata.csv")
```

```
## Rows: 785932 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
jul2022 <- read_csv("202207-divvy-tripdata.csv")
```

```
## Rows: 823488 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

"We would also need to check if there are any discrepancies with formatting as all the twelve tables should have the same number and types of columns"

```r
str(jun2023)
```

```
## spc_tbl_ [719,618 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:719618] "6F1682AC40EB6F71" "622A1686D64948EB" "3C88859D926253B4" "EAD8
##  $ rideable_type     : chr [1:719618] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:719618], format: "2023-06-05 13:34:12" "2023-06-05 01:30:22" ...
##  $ ended_at          : POSIXct[1:719618], format: "2023-06-05 14:31:56" "2023-06-05 01:33:06" ...
##  $ start_station_name: chr [1:719618] NA NA NA NA ...
##  $ start_station_id  : chr [1:719618] NA NA NA NA ...
##  $ end_station_name  : chr [1:719618] NA NA NA NA ...
##  $ end_station_id    : chr [1:719618] NA NA NA NA ...
##  $ start_lat         : num [1:719618] 41.9 41.9 42 42 42 ...
##  $ start_lng         : num [1:719618] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:719618] 41.9 41.9 41.9 42 42 ...
##  $ end_lng           : num [1:719618] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:719618] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
str(may2023)
```

```
## spc_tbl_ [604,827 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:604827] "0D9FA920C3062031" "92485E5FB5888ACD" "FB144B3FC8300187" "DDEB
##  $ rideable_type     : chr [1:604827] "electric_bike" "electric_bike" "electric_bike" "classic_bike"
##  $ started_at        : POSIXct[1:604827], format: "2023-05-07 19:53:48" "2023-05-06 18:54:08" ...
##  $ ended_at          : POSIXct[1:604827], format: "2023-05-07 19:58:32" "2023-05-06 19:03:35" ...
##  $ start_station_name: chr [1:604827] "Southport Ave & Belmont Ave" "Southport Ave & Belmont Ave" "Ha
##  $ start_station_id  : chr [1:604827] "13229" "13229" "13162" "13196" ...
##  $ end_station_name  : chr [1:604827] NA NA NA "Damen Ave & Cortland St" ...
##  $ end_station_id    : chr [1:604827] NA NA NA "13133" ...
##  $ start_lat         : num [1:604827] 41.9 41.9 41.9 41.9 42 ...
##  $ start_lng         : num [1:604827] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:604827] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:604827] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:604827] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
```

```
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(apr2023)
```

```
## spc_tbl_ [426,590 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:426590] "8FE8F7D9C10E88C7" "34E4ED3ADF1D821B" "5296BF07A2F77CB5" "4075
##  $ rideable_type     : chr [1:426590] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:426590], format: "2023-04-02 08:37:28" "2023-04-19 11:29:02" ...
##  $ ended_at          : POSIXct[1:426590], format: "2023-04-02 08:41:37" "2023-04-19 11:52:12" ...
##  $ start_station_name: chr [1:426590] NA NA NA NA ...
##  $ start_station_id  : chr [1:426590] NA NA NA NA ...
##  $ end_station_name  : chr [1:426590] NA NA NA NA ...
##  $ end_station_id    : chr [1:426590] NA NA NA NA ...
##  $ start_lat         : num [1:426590] 41.8 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:426590] -87.6 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:426590] 41.8 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:426590] -87.6 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:426590] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(mar2023)
```

```
## spc_tbl_ [258,678 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
```

```
##  $ ride_id           : chr [1:258678] "6842AA605EE9FBB3" "F984267A75B99A8C" "FF7CF57CFE026D02" "6B61
##  $ rideable_type     : chr [1:258678] "electric_bike" "electric_bike" "classic_bike" "classic_bike"
##  $ started_at        : POSIXct[1:258678], format: "2023-03-16 08:20:34" "2023-03-04 14:07:06" ...
##  $ ended_at          : POSIXct[1:258678], format: "2023-03-16 08:22:52" "2023-03-04 14:15:31" ...
##  $ start_station_name: chr [1:258678] "Clark St & Armitage Ave" "Public Rack - Kedzie Ave & Argyle S
##  $ start_station_id  : chr [1:258678] "13146" "491" "620" "TA1306000003" ...
##  $ end_station_name  : chr [1:258678] "Larrabee St & Webster Ave" NA "Clark St & Randolph St" "Sheff
##  $ end_station_id    : chr [1:258678] "13193" NA "TA1305000030" "13154" ...
##  $ start_lat         : num [1:258678] 41.9 42 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:258678] -87.6 -87.7 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:258678] 41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:258678] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:258678] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(feb2023)
```

```
## spc_tbl_ [190,445 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:190445] "CBCD0D7777F0E45F" "F3EC5FCE5FF39DE9" "E54C1F27FA9354FF" "3D56
##  $ rideable_type     : chr [1:190445] "classic_bike" "electric_bike" "classic_bike" "electric_bike"
##  $ started_at        : POSIXct[1:190445], format: "2023-02-14 11:59:42" "2023-02-15 13:53:48" ...
##  $ ended_at          : POSIXct[1:190445], format: "2023-02-14 12:13:38" "2023-02-15 13:59:08" ...
##  $ start_station_name: chr [1:190445] "Southport Ave & Clybourn Ave" "Clarendon Ave & Gordon Ter" "S
##  $ start_station_id  : chr [1:190445] "TA1309000030" "13379" "TA1309000030" "TA1309000030" ...
##  $ end_station_name  : chr [1:190445] "Clark St & Schiller St" "Sheridan Rd & Lawrence Ave" "Aberdee
##  $ end_station_id    : chr [1:190445] "TA1309000024" "TA1309000041" "13156" "TA1309000008" ...
##  $ start_lat         : num [1:190445] 41.9 42 41.9 41.9 41.8 ...
##  $ start_lng         : num [1:190445] -87.7 -87.6 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:190445] 41.9 42 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:190445] -87.6 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:190445] "casual" "casual" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
```

```
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(jan2023)
```

```
## spc_tbl_ [190,301 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:190301] "F96D5A74A3E41399" "13CB7EB698CEDB88" "BD88A2E670661CE5" "C907
##  $ rideable_type     : chr [1:190301] "electric_bike" "classic_bike" "electric_bike" "classic_bike"
##  $ started_at        : POSIXct[1:190301], format: "2023-01-21 20:05:42" "2023-01-10 15:37:36" ...
##  $ ended_at          : POSIXct[1:190301], format: "2023-01-21 20:16:33" "2023-01-10 15:46:05" ...
##  $ start_station_name: chr [1:190301] "Lincoln Ave & Fullerton Ave" "Kimbark Ave & 53rd St" "Western
##  $ start_station_id  : chr [1:190301] "TA1309000058" "TA1309000037" "RP-005" "TA1309000037" ...
##  $ end_station_name  : chr [1:190301] "Hampden Ct & Diversey Ave" "Greenwood Ave & 47th St" "Valli Pi
##  $ end_station_id    : chr [1:190301] "202480.0" "TA1308000002" "599" "TA1308000002" ...
##  $ start_lat         : num [1:190301] 41.9 41.8 42 41.8 41.8 ...
##  $ start_lng         : num [1:190301] -87.6 -87.6 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:190301] 41.9 41.8 42 41.8 41.8 ...
##  $ end_lng           : num [1:190301] -87.6 -87.6 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:190301] "member" "member" "casual" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(dec2022)
```

```
## spc_tbl_ [181,806 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:181806] "65DBD2F447EC51C2" "0C201AA7EA0EA1AD" "E0B148CCB358A49D" "54C5
##  $ rideable_type     : chr [1:181806] "electric_bike" "classic_bike" "electric_bike" "classic_bike"
##  $ started_at        : POSIXct[1:181806], format: "2022-12-05 10:47:18" "2022-12-18 06:42:33" ...
##  $ ended_at          : POSIXct[1:181806], format: "2022-12-05 10:56:34" "2022-12-18 07:08:44" ...
##  $ start_station_name: chr [1:181806] "Clifton Ave & Armitage Ave" "Broadway & Belmont Ave" "Sangamo
```

```
##  $ start_station_id  : chr [1:181806] "TA1307000163" "13277" "TA1306000015" "KA1503000038" ...
##  $ end_station_name  : chr [1:181806] "Sedgwick St & Webster Ave" "Sedgwick St & Webster Ave" "St. Cl
##  $ end_station_id    : chr [1:181806] "13191" "13191" "13016" "13134" ...
##  $ start_lat         : num [1:181806] 41.9 41.9 41.9 41.8 41.9 ...
##  $ start_lng         : num [1:181806] -87.7 -87.6 -87.7 -87.6 -87.7 ...
##  $ end_lat           : num [1:181806] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:181806] -87.6 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:181806] "member" "casual" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(nov2022)
```

```
## spc_tbl_ [337,735 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:337735] "BCC66FC6FAB27CC7" "772AB67E902C180F" "585EAD07FDEC0152" "91C4
##  $ rideable_type     : chr [1:337735] "electric_bike" "classic_bike" "classic_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:337735], format: "2022-11-10 06:21:55" "2022-11-04 07:31:55" ...
##  $ ended_at          : POSIXct[1:337735], format: "2022-11-10 06:31:27" "2022-11-04 07:46:25" ...
##  $ start_station_name: chr [1:337735] "Canal St & Adams St" "Canal St & Adams St" "Indiana Ave & Roos
##  $ start_station_id  : chr [1:337735] "13011" "13011" "SL-005" "SL-005" ...
##  $ end_station_name  : chr [1:337735] "St. Clair St & Erie St" "St. Clair St & Erie St" "St. Clair St
##  $ end_station_id    : chr [1:337735] "13016" "13016" "13016" "13016" ...
##  $ start_lat         : num [1:337735] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:337735] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:337735] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:337735] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:337735] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
```

```
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(oct2022)
```

```
## spc_tbl_ [558,685 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:558685] "A50255C1E17942AB" "DB692A70BD2DD4E3" "3C02727AAF60F873" "47E6
##  $ rideable_type     : chr [1:558685] "classic_bike" "electric_bike" "electric_bike" "electric_bike"
##  $ started_at        : POSIXct[1:558685], format: "2022-10-14 17:13:30" "2022-10-01 16:29:26" ...
##  $ ended_at          : POSIXct[1:558685], format: "2022-10-14 17:19:39" "2022-10-01 16:49:06" ...
##  $ start_station_name: chr [1:558685] "Noble St & Milwaukee Ave" "Damen Ave & Charleston St" "Hoyne A
##  $ start_station_id  : chr [1:558685] "13290" "13288" "655" "KA1504000133" ...
##  $ end_station_name  : chr [1:558685] "Larrabee St & Division St" "Damen Ave & Cullerton St" "Western
##  $ end_station_id    : chr [1:558685] "KA1504000079" "13089" "TA1307000140" "620" ...
##  $ start_lat         : num [1:558685] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:558685] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:558685] 41.9 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:558685] -87.6 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:558685] "member" "casual" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(sep2022)
```

```
## spc_tbl_ [701,339 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:701339] "5156990AC19CA285" "E12D4A16BF51C274" "A02B53CD7DB72DD7" "C82E
##  $ rideable_type     : chr [1:701339] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:701339], format: "2022-09-01 08:36:22" "2022-09-01 17:11:29" ...
##  $ ended_at          : POSIXct[1:701339], format: "2022-09-01 08:39:05" "2022-09-01 17:14:45" ...
##  $ start_station_name: chr [1:701339] NA NA NA NA ...
##  $ start_station_id  : chr [1:701339] NA NA NA NA ...
##  $ end_station_name  : chr [1:701339] "California Ave & Milwaukee Ave" NA NA NA ...
##  $ end_station_id    : chr [1:701339] "13084" NA NA NA ...
##  $ start_lat         : num [1:701339] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:701339] -87.7 -87.6 -87.6 -87.7 -87.7 ...
```

```
##  $ end_lat           : num [1:701339] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:701339] -87.7 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:701339] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(aug2022)
```

```
## spc_tbl_ [785,932 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:785932] "550CF7EFEAE0C618" "DAD198F405F9C5F5" "E6F2BC47B65CB7FD" "F5978
##  $ rideable_type     : chr [1:785932] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:785932], format: "2022-08-07 21:34:15" "2022-08-08 14:39:21" ...
##  $ ended_at          : POSIXct[1:785932], format: "2022-08-07 21:41:46" "2022-08-08 14:53:23" ...
##  $ start_station_name: chr [1:785932] NA NA NA NA ...
##  $ start_station_id  : chr [1:785932] NA NA NA NA ...
##  $ end_station_name  : chr [1:785932] NA NA NA NA ...
##  $ end_station_id    : chr [1:785932] NA NA NA NA ...
##  $ start_lat         : num [1:785932] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:785932] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:785932] 41.9 41.9 42 42 41.8 ...
##  $ end_lng           : num [1:785932] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:785932] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(jul2022)
```

```
## spc_tbl_ [823,488 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:823488] "954144C2F67B1932" "292E027607D218B6" "57765852588AD6E0" "B5B6
##  $ rideable_type     : chr [1:823488] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:823488], format: "2022-07-05 08:12:47" "2022-07-26 12:53:38" ...
##  $ ended_at          : POSIXct[1:823488], format: "2022-07-05 08:24:32" "2022-07-26 12:55:31" ...
##  $ start_station_name: chr [1:823488] "Ashland Ave & Blackhawk St" "Buckingham Fountain (Temp)" "Buc
##  $ start_station_id  : chr [1:823488] "13224" "15541" "15541" "15541" ...
##  $ end_station_name  : chr [1:823488] "Kingsbury St & Kinzie St" "Michigan Ave & 8th St" "Michigan A
##  $ end_station_id    : chr [1:823488] "KA1503000043" "623" "623" "TA1307000164" ...
##  $ start_lat         : num [1:823488] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:823488] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:823488] 41.9 41.9 41.9 41.8 41.9 ...
##  $ end_lng           : num [1:823488] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:823488] "member" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

“after check every thing is ok”

“Now merge all the tables into one dataset”

```
 cyclics <- bind_rows(jun2023,may2023 ,apr2023 ,mar2023 ,feb2023 ,jan2023 ,dec2022 ,
nov2022 ,oct2022 ,sep2022 ,aug2022 ,jul2022 )
```

” to ensure that the new datast has been created ”

```
str(cyclics)
```

```
## spc_tbl_ [5,779,444 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:5779444] "6F1682AC40EB6F71" "622A1686D64948EB" "3C88859D926253B4" "EAD8
##  $ rideable_type     : chr [1:5779444] "electric_bike" "electric_bike" "electric_bike" "electric_bik
##  $ started_at        : POSIXct[1:5779444], format: "2023-06-05 13:34:12" "2023-06-05 01:30:22" ...
##  $ ended_at          : POSIXct[1:5779444], format: "2023-06-05 14:31:56" "2023-06-05 01:33:06" ...
##  $ start_station_name: chr [1:5779444] NA NA NA NA ...
##  $ start_station_id  : chr [1:5779444] NA NA NA NA ...
```

```
##  $ end_station_name  : chr [1:5779444] NA NA NA NA ...
##  $ end_station_id    : chr [1:5779444] NA NA NA NA ...
##  $ start_lat         : num [1:5779444] 41.9 41.9 42 42 42 ...
##  $ start_lng         : num [1:5779444] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:5779444] 41.9 41.9 41.9 42 42 ...
##  $ end_lng           : num [1:5779444] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:5779444] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

cleaning the dataset "cyclics"

first ensure that the variables names is clean and consistent

```
cyclics <- clean_names(cyclics)
```

We should also remove any empty columns and rows in our data.frame, we can do so by using remove_empty().

```
cyclics<-remove_empty(cyclics,which=c("rows","cols"),quiet=F)
```

```
## No empty rows to remove.
```

```
## No empty columns to remove.
```

Add some important columns

"we need to add some columns (day_of_week ,start_hour , month ,trip_duration)"

"first add the column day_of_week by using wday() function"

```
cyclics$day_of_week<-wday(cyclics$started_at ,label=T,abb=T)
```

“second add the column starting_hour by using format(as.posixct)”

```
cyclics$starting_hour <- format(as.POSIXct(cyclics$started_at),'%H')
```

“Now let’s create month column by using format(as.POSIXct)”

```
cyclics$month <- format(as.Date(cyclics$started_at),format='%Y-%m')
```

” Next let’s get the trip_length column ”

```
cyclics$trip_length <- difftime(cyclics$ended_at, cyclics$started_at, units ='min')
```

“Now we will remove any row that has atrip_length equals or less than zero”

```
cyclics <- cyclics%>%
filter(trip_length>0)
```

build our new csv file

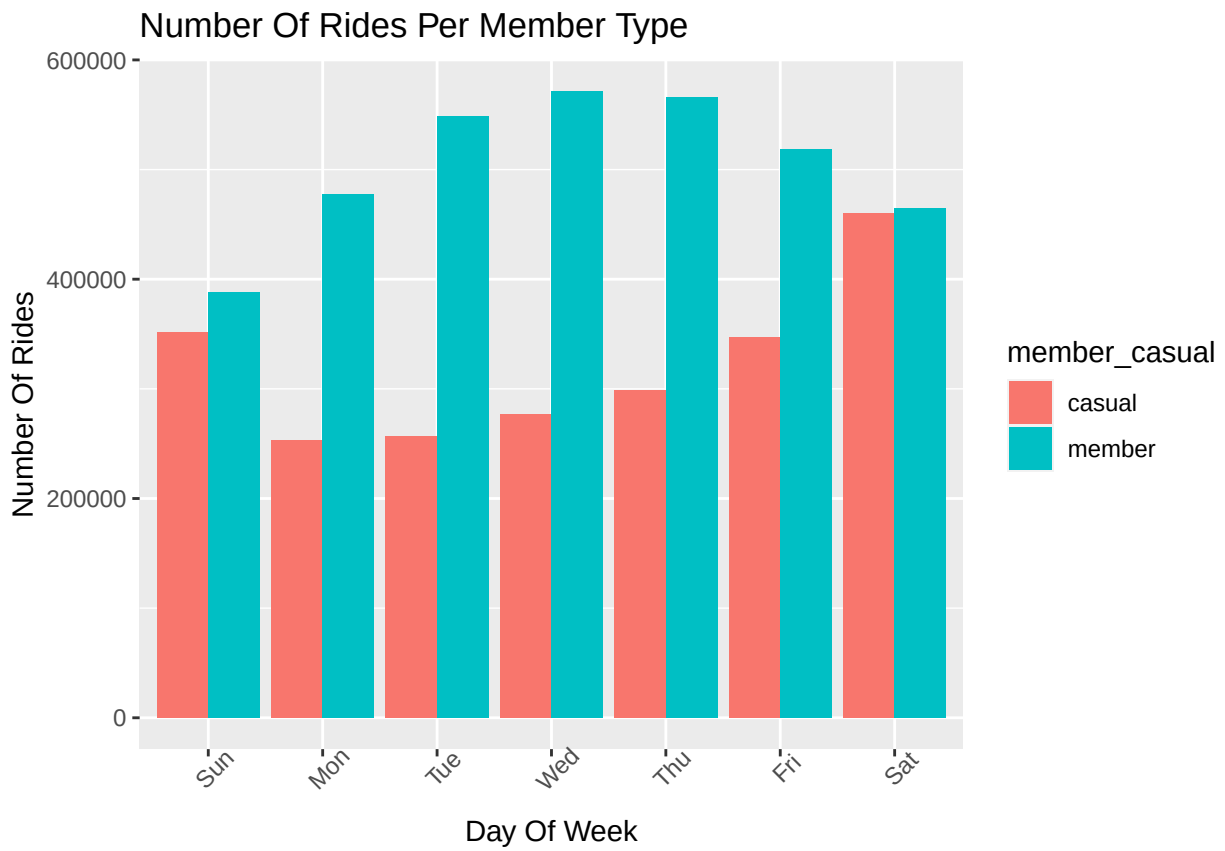
```
write.csv(cyclics, file ='cyclics.csv')
```

share&analyze

” we have some analysis questions to answer 1-How do casual and members use their bikes differently throughout the week 2-Peak hours of bike usage between casual and annual members 3-Bike usage throughout the year 4-The average trip duration between casual and annual members 5-Most popular stations among casual and annual members”

Now we will use ggplot2 library which is apart of tidyverse library to create a chart about the number of rides per day for each type od rider

```
options(scipen = 999)
ggplot(cyclics)+
aes(x=day_of_week,fill=member_casual)+
  geom_bar(position='dodge')+
labs(title='Number Of Rides Per Member Type', x='Day Of Week',y='Number Of Rides')+
theme(axis.text.x=element_text(angle=45))
```



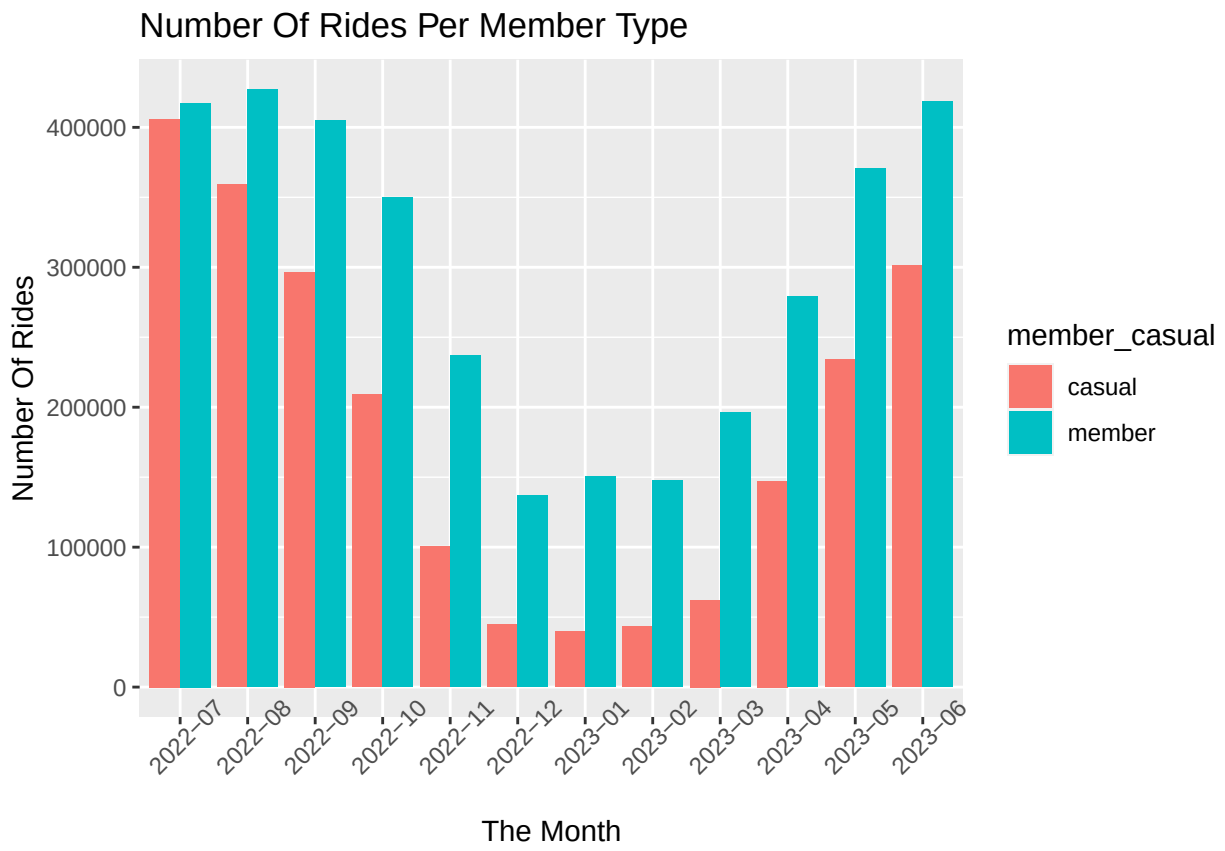
Now let's save the image

```
ggsave("Number Of Rides Per day.png")
```

```
## Saving 6.5 x 4.5 in image
```

to create a chart about the number of rides per month for each type of rider

```
ggplot(cyclics)+
  aes(x=month,fill=member_casual)+
  geom_bar(position='dodge')+
  labs(title='Number Of Rides Per Member Type', x='The Month',y='Number Of Rides')+
  theme(axis.text.x=element_text(angle=45))
```



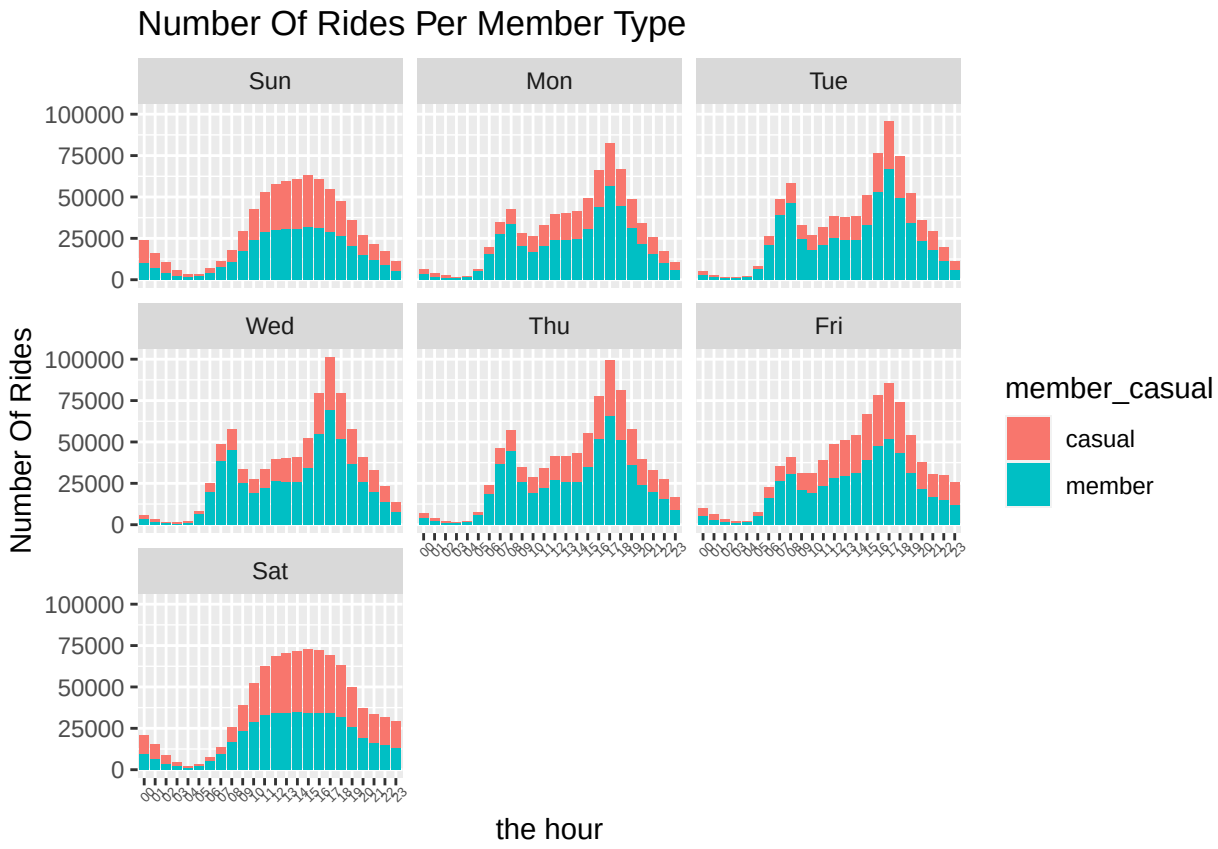
Now let's save the image

```
ggsave("Number Of Rides Per month.png")
```

```
## Saving 6.5 x 4.5 in image
```

create arelations between the hour of day and number of rides

```
ggplot(cyclics)+
  aes(x=starting_hour,fill=member_casual)+
  geom_bar()+
  facet_wrap(~day_of_week)+
  labs(title='Number Of Rides Per Member Type', x='the hour',y='Number Of Rides')+
theme(axis.text.x=element_text(angle=45,size=5))
```



Casual Vs Member

Now we have aquick look about the data but let's make it more professional , so what about getting the percent of member and casual per trips

```
cyclics %>%
  group_by(member_casual) %>%
  summarise(count = n(), Percentage = n()/nrow(cyclics)*100)
```

```
## # A tibble: 2 x 3
##   member_casual   count Percentage
##   <chr>           <int>      <dbl>
## 1 casual        2243990       38.8
## 2 member        3534880       61.2
```

It clears that the number of users as members is 61.17 % and the casuals is 38.83 % so now we know that members is more than casuals as users.

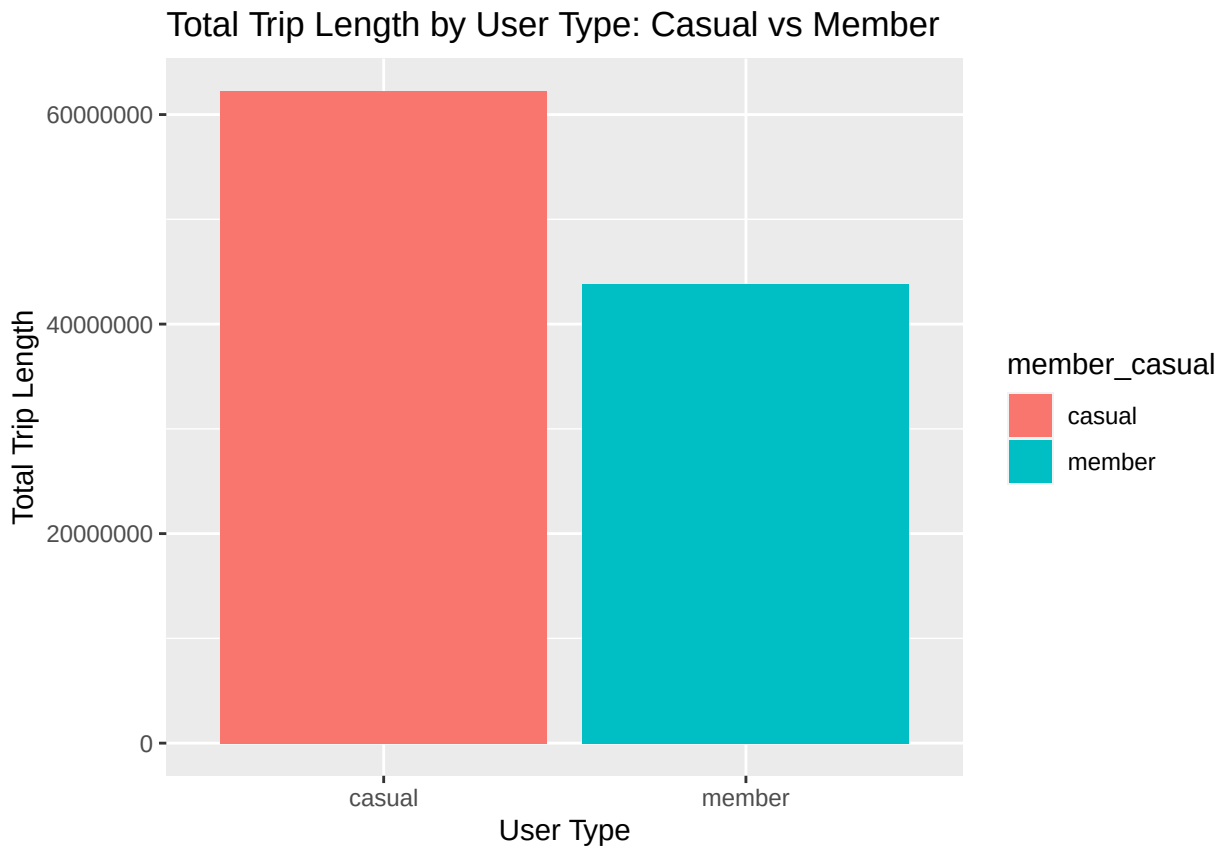
```
cyclics %>%
  group_by(member_casual) %>%
  summarise(sum(trip_length))
```

```
## # A tibble: 2 x 2
```

```
##   member_casual `sum(trip_length)`
##   <chr>         <drtn>
## 1 casual        62232084 mins
## 2 member        43776409 mins
```

```
ggplot(cyclics, aes(x=member_casual,y=trip_length, fill=member_casual)) +
  geom_col() +
  labs(x="User Type", y="Total Trip Length", title = "Total Trip Length by User Type: Casual vs Member")
```

```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```



It's so important to know whose spend more time casuals or members Now we discovered that casuals using bike for so much time than members.

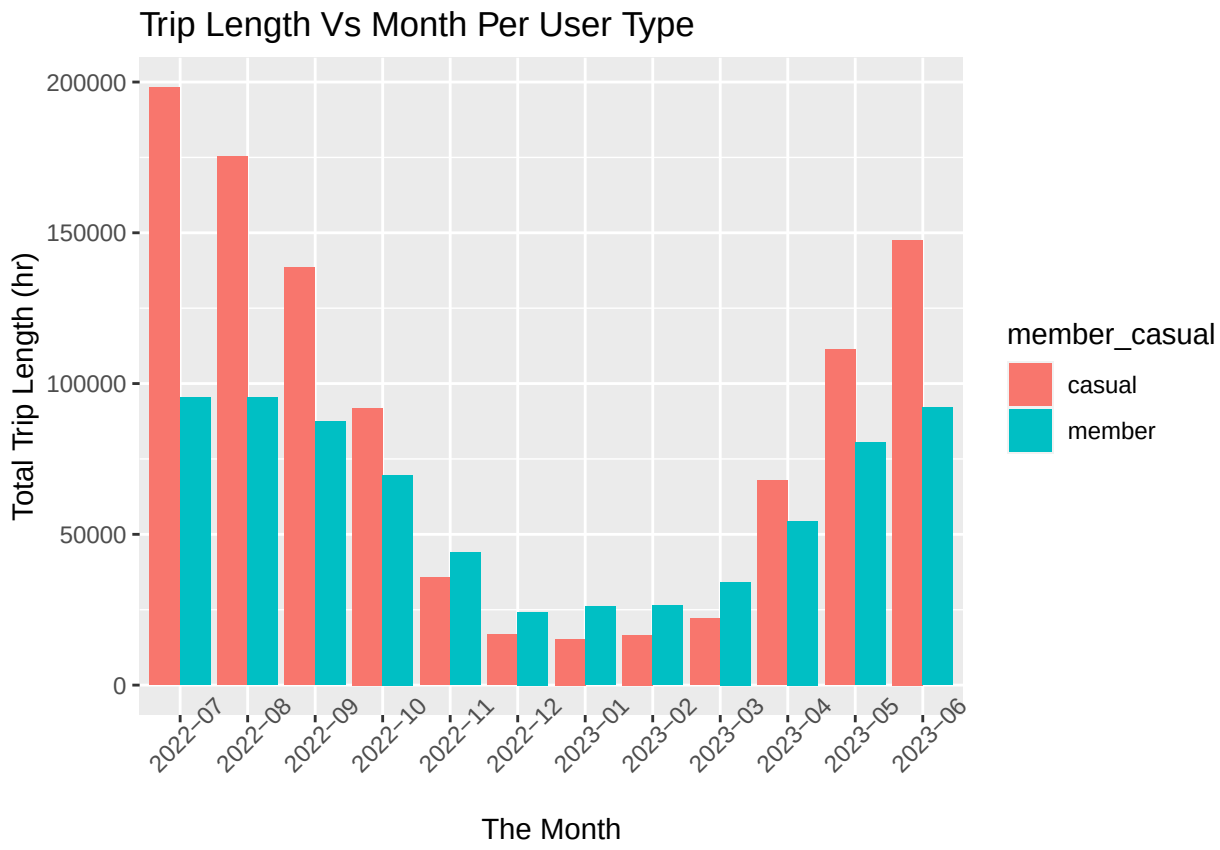
Trips_Length is our new scale

–create a chart about trips length per month for each type of user

```
cyclics %>%
  group_by(member_casual,month) %>%
  summarise(total_trip_length=sum(trip_length)/60,.groups = "drop") %>%
ggplot()+
aes(x=month,y=total_trip_length,fill=member_casual)+
```

```
  geom_col(position ='dodge')+
  labs(title ='Trip Length Vs Month Per User Type',x='The Month',y='Total Trip Length (hr)')+
  theme(axis.text.x=element_text(angle=45))
```

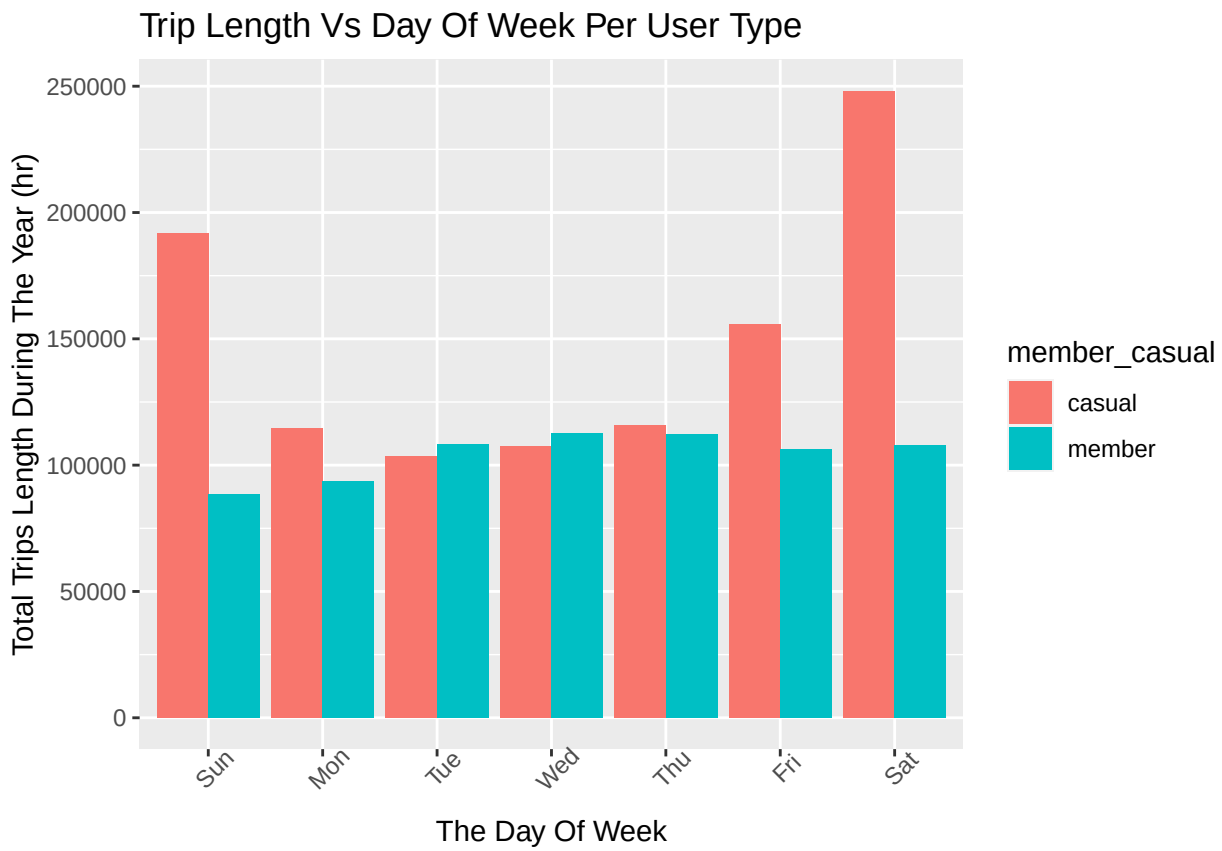
```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```



Let's take a look at the weekdays

```
cyclics %>%
  group_by(member_casual,day_of_week) %>%
  summarise(total_trip_length=sum(trip_length)/60,.groups = 'drop') %>%
ggplot()+
  aes(x=day_of_week,y=total_trip_length,fill=member_casual)+
  geom_col(position = 'dodge')+
 labs(title ='Trip Length Vs Day Of Week Per User Type',x='The Day Of Week',y='Total Trips Length During
  theme(axis.text.x=element_text(angle=45))
```

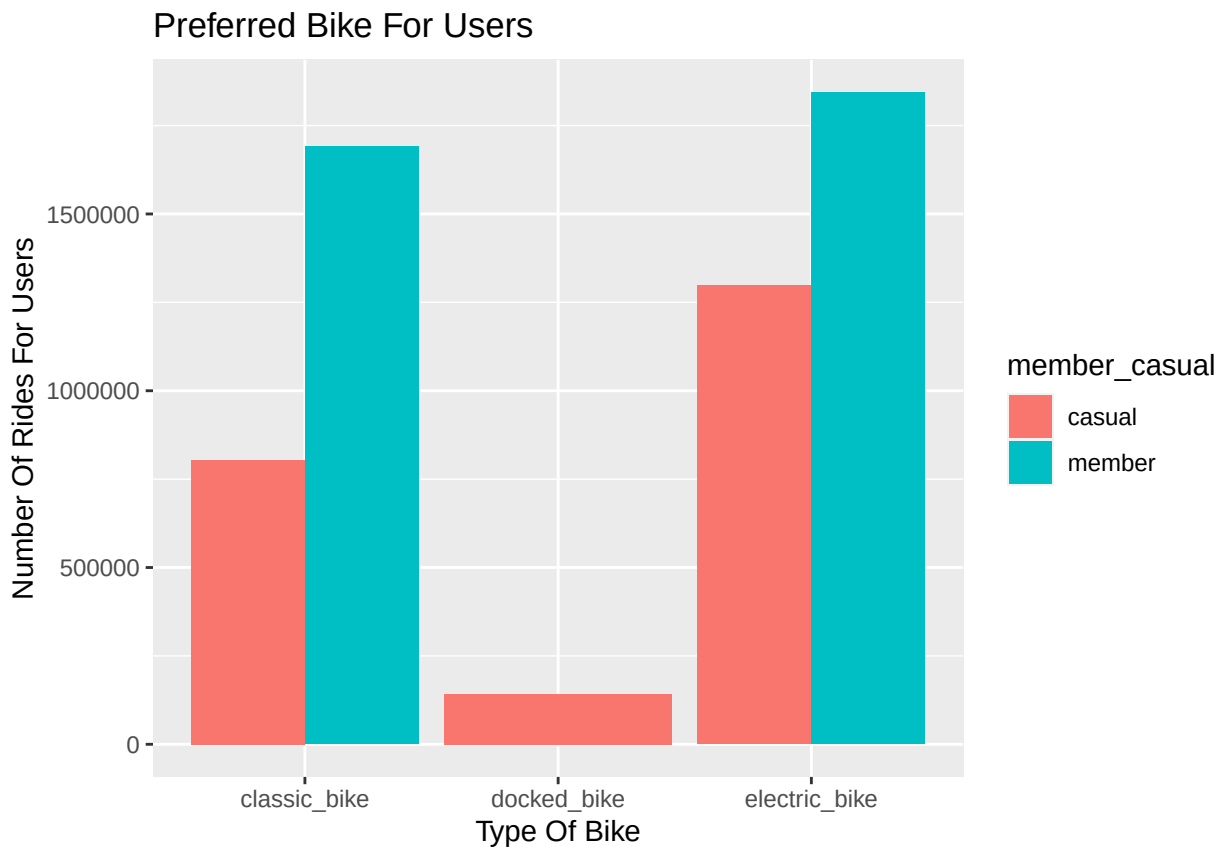
```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```



Type Of Bike

what's the most preferred type of bike ? –as number of rides–

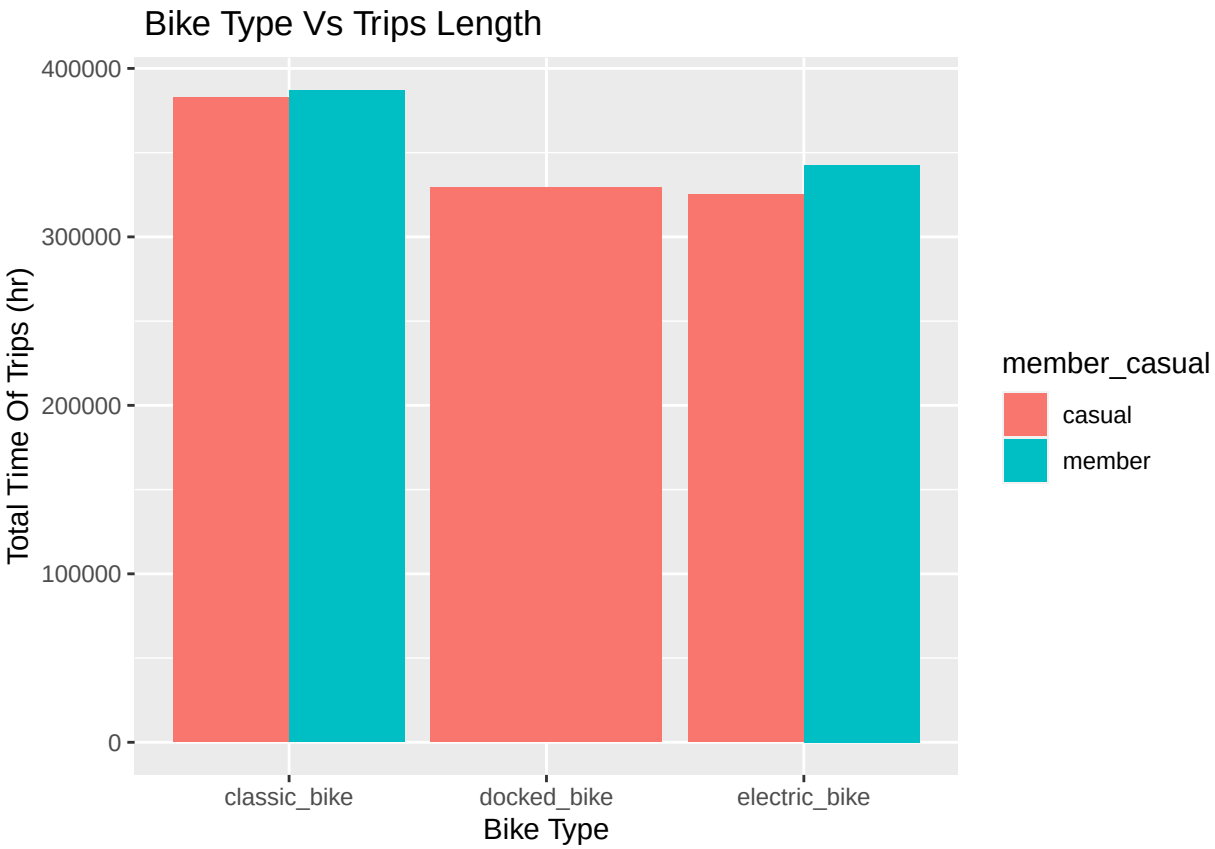
```
ggplot(cyclics)+
  aes(x=rideable_type,fill=member_casual)+
  geom_bar(position='dodge')+
labs(title = 'Preferred Bike For Users',x='Type Of Bike' ,y='Number Of Rides For Users' )
```



what's the most preferred type of bike ? –as length of trips–

```
cyclics %>%
  group_by(rideable_type,member_casual) %>%
summarise(time_of_trips=sum(trip_length)/60,.groups ='drop' ) %>%
  ggplot()+
  aes(x=rideable_type,y=time_of_trips,fill=member_casual)+
  geom_col(position = 'dodge')+
  labs(title=' Bike Type Vs Trips Length',x='Bike Type',y='Total Time Of Trips (hr)')
```

```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```



So we estimate that classic_bikes have trips length more than others

Conclusions

- Members have trips more than Casuals.
- casuals spent time more than members in their trips.
- People prefer bikes in summer months from may to September.
- Casuals prefer bikes in weekend days (Saturday and Sunday).
- Member usage almost fixed during weekdays.
- People prefer electric bikes.

Recommendations

- Focusing in making advertisements and offers of membership at Saturday and Sunday.
- Making a true offers in months from May to September.
- Give attention to get more electric bikes.