

# Cyclstic case study

Amol khaire

2023-01-21

## Scenario

You are 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, 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. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

## Problems

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

## Importing and grouping data

```
install.packages("readr")  
library(readr)  
ak1 <- read.csv("202201-divvy-tripdata.csv")  
ak2 <- read.csv("202202-divvy-tripdata.csv")  
ak3 <- read.csv("202203-divvy-tripdata.csv")  
ak4 <- read.csv("202204-divvy-tripdata.csv")  
ak5 <- read.csv("202205-divvy-tripdata.csv")  
ak6 <- read.csv("202206-divvy-tripdata.csv")  
ak7 <- read.csv("202207-divvy-tripdata.csv")  
ak8 <- read.csv("202208-divvy-tripdata.csv")  
ak9 <- read.csv("202209-divvy-publictripdata.csv")
```

```
ak10 <- read.csv("202210-divvy-tripdata.csv")
ak11 <- read.csv("202211-divvy-tripdata.csv")
ak12 <- read.csv("202212-divvy-tripdata.csv")
bike_rides <- rbind(ak1,ak2,ak3,ak4,ak5,ak6,ak7,ak8,ak9,ak10,ak11,ak12)
```

## **Cleaning and transforming data**

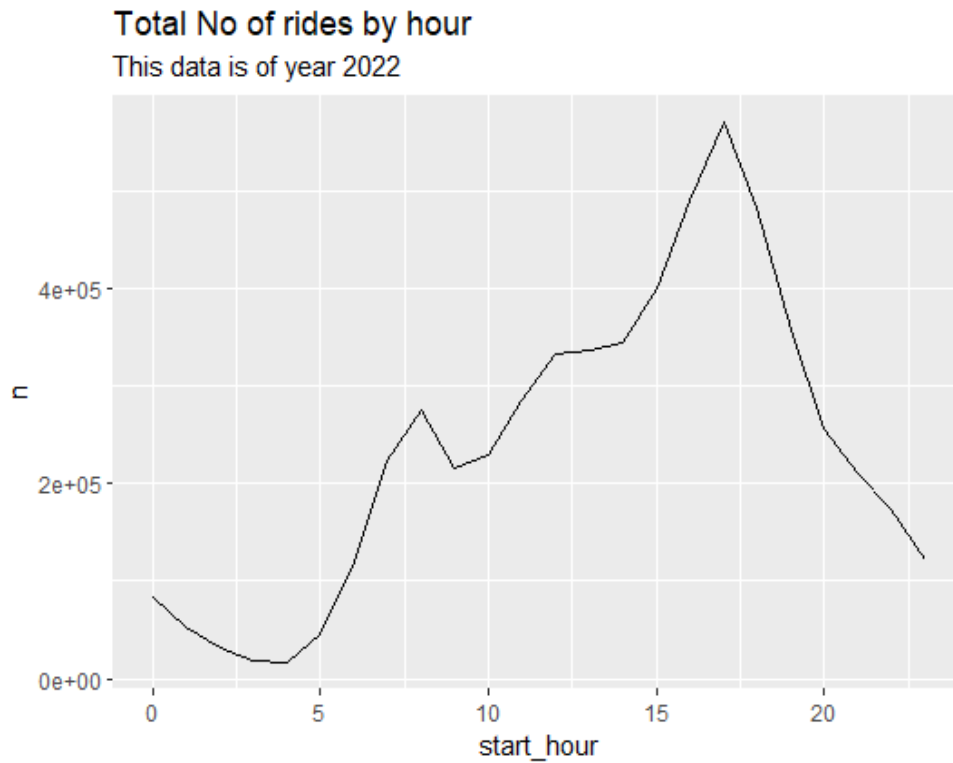
- In spreadsheet create a column called “ride\_length.” Calculate the length of each ride by subtracting the column “started\_at” from the column “ended\_at” (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37:30:55.
- Create a column called “day\_of\_week,” and calculate the 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 7 = Sunday and 1 = Monday.
- Removing empty Columns and Rows `bike_rides <- janitor::remove_empty(bike_rides,which = c("cols"))` `bike_rides <- janitor::remove_empty(bike_rides,which = c("rows"))`
- Change Date and Time format `bike_rides$started_at <- lubridate::dmy_hm(bike_rides$started_at)` `bike_rides$ended_at <- lubridate::dmy_hm(bike_rides$ended_at)`
- Creating column containing hour `bike_rides$start_hour <- lubridate::hour(bike_rides$started_at)` `bike_rides$end_hour <- lubridate::hour(bike_rides$ended_at)`
- Creating column containing month `bike_rides$start_month <- lubridate::month(bike_rides$started_at)`

## **Creating separate table name member and casual Riders**

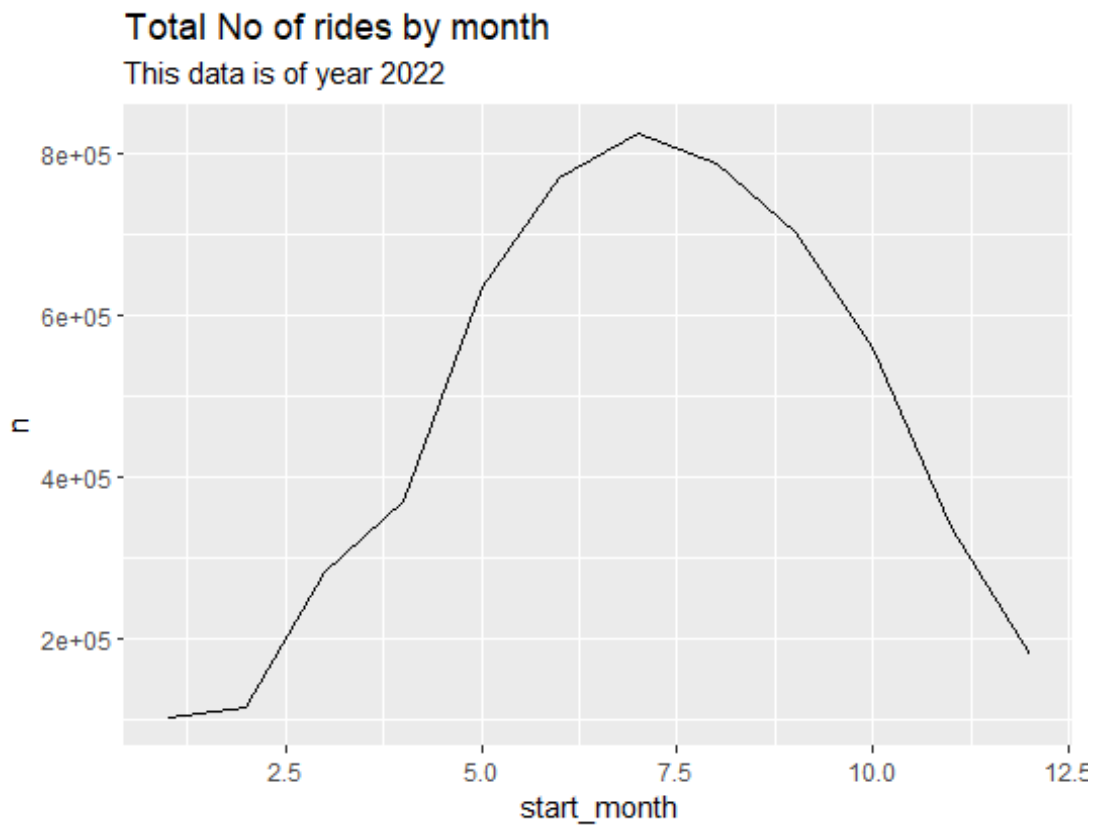
```
member_riders <- filter(bike_rides, member_casual == "member")
casual_riders <- filter(bike_rides, member_casual == "casual")
```

## Plotting data

```
bike_rides %>% count(start_hour, sort = T) %>% ggplot() +  
  geom_line(aes(x=start_hour,y=n)) + labs(title = "Total No of rides by hour",  
  subtitle = "This data is of year 2022")
```



```
bike_rides %>% count(start_month, sort = T) %>% ggplot() +  
  geom_line(aes(x=start_month,y=n)) + labs(title = "Total No of rides by  
month", subtitle = "This data is of year 2022")
```



## Finding Which days has highest no. of member and casual Riders

```
member_riders %>% count(day_of_week, sort = T) %>% ggplot() +  
  geom_line(aes(x=day_of_week, y=n)) + labs(title = "Total No of rides by day  
by members", subtitle = "(This data is of year 2022)", caption  
="(mon=1,tue=2,...,sun=7)")
```



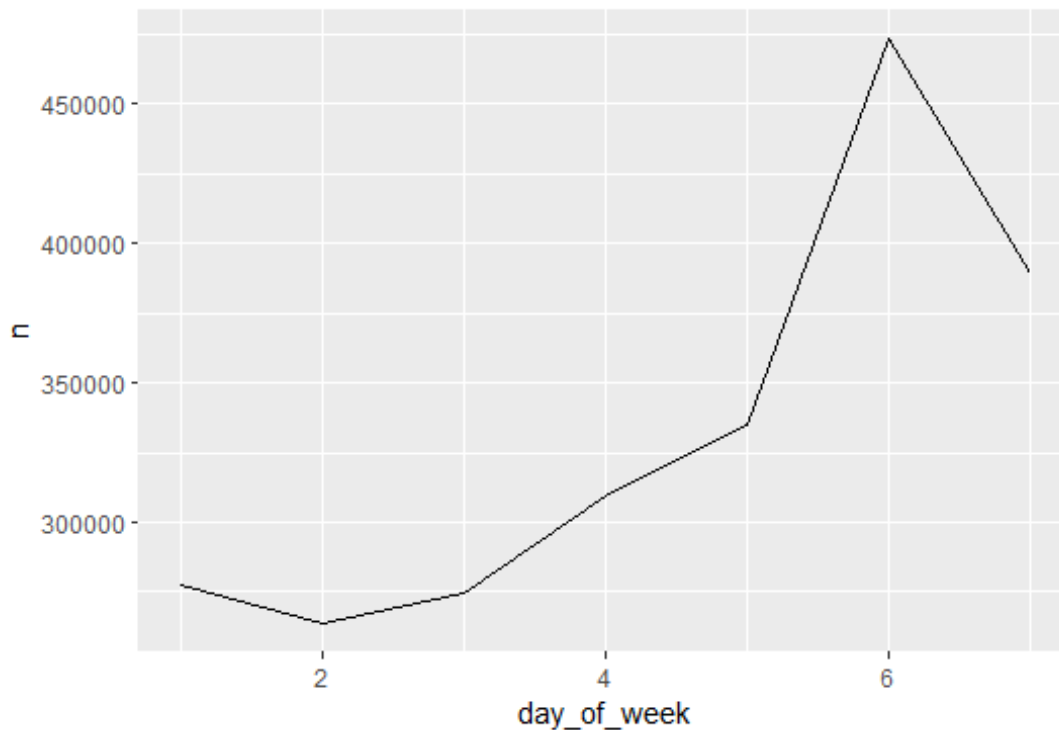
From this it is clear that mostly riders who use cycle at workdays are members.

```
casual_riders %>% count(day_of_week, sort = T) %>% ggplot() +  
  geom_line(aes(x=day_of_week, y=n)) + labs(title = "Total No of rides by day  
by casual riders", subtitle = "(This data is of year
```

```
2022)",caption="(mon=1,tue=2,...,sun=7)")
```

### Total No of rides by day by casual

This data is of year 2022



From this it is clear that mostly casual riders ride cycle at weekend.

### Count riders according to their choice of cycle type

```
member_riders %>% count(rideable_type, sort = T)
```

n

#### rideable\_type

classic\_bike 1709755

electric\_bike 1635930

```
casual_riders %>% count(rideable_type, sort = T)
```

rideable\_type n

electric\_bike 253099

classic\_bike 891459

docked\_bike 177474

## **Conclusion**

1. Most people ride during May to October, This is may be due to huge no. of tourist in Chicago during this period.
2. Most people usually rides during 1pm to 7pm and it peaks at 5pm.
3. Riders who are members more preferred classic bike and casual riders more preferred electric bike, also there is sufficient no. of casual riders who preferred docked bike.