Cyclistic\_data\_2022\_10

Kevin Ajith

2022-12-06

## Cyclistic Data

The data is taken from [link](https://divvy-tripdata.s3.amazonaws.com/index.html) for the file updated on Nov 9th 2022, 04:17:10 am. And is copyrighted under Motivate International Inc [[lisence](https://ride.divvybikes.com/data-license-agreement)].

### Data Modification before Data Anlysis Process

* 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.
* Create a column called “day\_of\_week,” and calculate the day of the week that each ride started.
* 1 = Sunday and 7 = Saturday
* Converted data for “started\_at” and “ended\_at” column to Date and Time Format
* Removed “latitude” and “longitude”, Since the data is not required for the analysis.

### Library Required For Analysis

library(tidyverse)

### Reading and storing csv file

The data\_trip data is the consolidated data taken from December 2021 to November 2022, And shows the latest trend among the people using **Cyclistic**.

This part of the code accumulated the data into one variable.

data\_trip <- list.files(path="C:/Users/kevin/Desktop/DA/Case Study DA/1/Data Transform/2022-cleaned", full.names = TRUE) %>%   
 lapply(read\_csv) %>%   
 bind\_rows

## Rows: 247540 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 103770 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 115609 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 284042 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 371249 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 634858 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 769204 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 823488 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 785932 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 701339 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 558685 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.  
## Rows: 337735 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, ...  
## dbl (1): day\_of\_week  
## time (1): ride\_length  
##   
## 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.

head(data\_trip)

## # A tibble: 6 × 11  
## ride\_id ridea…¹ start…² ended…³ start…⁴ start…⁵ end\_s…⁶ end\_s…⁷ membe…⁸  
## <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 46F8167220E44… electr… 07-12-… 07-12-… Laflin… 13307 Morgan… TA1307… member   
## 2 73A77762838B3… electr… 11-12-… 11-12-… LaSall… KP1705… Claren… TA1307… casual   
## 3 4CF42452054F5… electr… 15-12-… 15-12-… Halste… KA1504… Broadw… 13137 member   
## 4 3278BA87BF698… classi… 26-12-… 26-12-… Halste… KA1504… LaSall… KP1705… member   
## 5 6FF54232576A3… electr… 30-12-… 30-12-… Leavit… 18058 Clark … TA1307… member   
## 6 93E8D79490E3A… classi… 01-12-… 01-12-… Wabash… SL-012 Wells … SL-011 member   
## # … with 2 more variables: ride\_length <time>, day\_of\_week <dbl>, and  
## # abbreviated variable names ¹​rideable\_type, ²​started\_at, ³​ended\_at,  
## # ⁴​start\_station\_name, ⁵​start\_station\_id, ⁶​end\_station\_name, ⁷​end\_station\_id,  
## # ⁸​member\_casual

### Calculations on ride\_length and day\_of\_week

Mean and max can be found using the function using R as well as the Excel code

* =MEAN()
* =MAX()

This Code gets some basic information about the data like mean and max ride lengths.

data\_trip %>% summarise(mean\_ride\_length = mean(ride\_length),  
 max\_ride\_length = max(ride\_length))

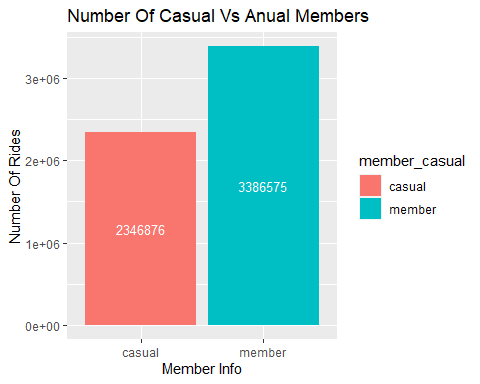
## # A tibble: 1 × 2  
## mean\_ride\_length max\_ride\_length  
## <drtn> <drtn>   
## 1 989.8742 secs 86396 secs

### Plot For Member Vs Casual riders

Before going for further analysis we need to find out how many rides are taken by casual and annual Cyclistic users.

This part of the code take the information from data\_trip and plots the bar graph for the same, also adding title and label layer

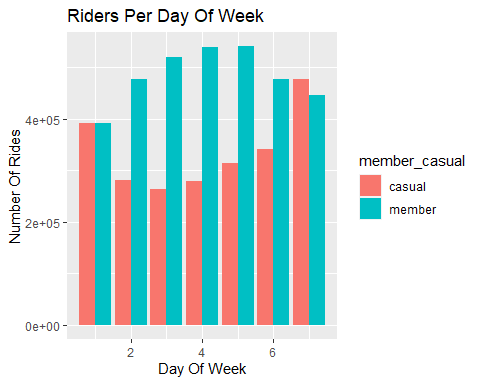
ggplot(data\_trip,aes(x=member\_casual, fill=member\_casual)) + geom\_bar() +   
 stat\_count(geom = "text",   
 colour = "white", size = 3.5,   
 aes(label = ..count..),   
 position = position\_stack(vjust = 0.5)) +  
 ggtitle("Number Of Casual Vs Anual Members") + xlab("Member Info") + ylab("Number Of Rides")

 Cyclistic has 2,346,876 casual users, Which is a great number when we compare to annual users of 3,386,575. This makes the percentage of casual users to be **40.93%**.

### Plot: Bar Chart for day\_of\_week with respect to member\_casual

This Code finds the number of rides taken by Cyclistic ride user as per the days.

ggplot(data\_trip,aes(x=day\_of\_week, fill=member\_casual))+geom\_bar(position = "dodge") +  
 ggtitle("Riders Per Day Of Week") + xlab("Day Of Week") + ylab("Number Of Rides")

 The Graph Clearly Shows that the casual riders use Cyclistic the most during Saturdays and Sundays which is significantly more than other days.

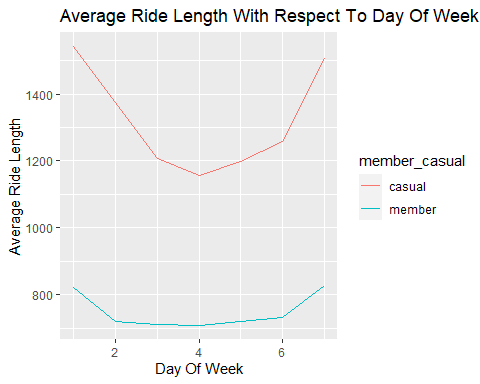
### Plot: Bar Chart for Ride\_length Vs day\_of\_week

This code creates another data frame where it finds the mean ride length for each day with respect to the type of membership they hold.

data\_trip\_mean <- data\_trip %>%   
 group\_by(day\_of\_week, member\_casual) %>%   
 summarise(mean\_ride\_length = mean(ride\_length), .groups = "rowwise" )

This code prints the bar plot for the above calculated data frame.

ggplot(data\_trip\_mean,aes(x = day\_of\_week, y = mean\_ride\_length, color = member\_casual)) +   
 geom\_line() +  
 ggtitle("Average Ride Length With Respect To Day Of Week") +   
 xlab("Day Of Week") + ylab("Average Ride Length")



The plot clearly shows that the casual riders on average take longer timed trips as compared to annual members.

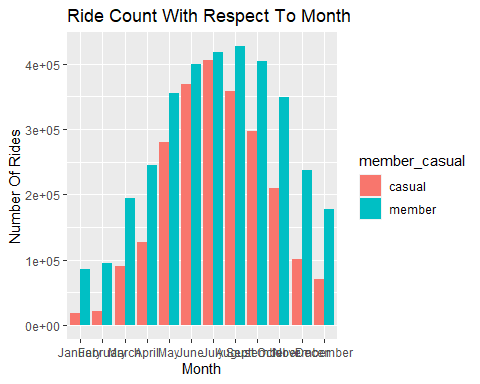
### Plot: Bar Chart for Ride\_length Vs month

This code snippet compiles the membership status and month from “start\_at” column.

data\_month <- data\_trip %>%   
 arrange(started\_at) %>%   
 summarize(month = months(as.Date(started\_at)), member\_casual)

This Code is to plot the bar chart to find the number of rides per month with respect to the membership.

month\_vals = c('January', 'February','March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December')  
  
ggplot(data\_month,aes(x=factor(month, level=month\_vals), fill= member\_casual)) +   
 geom\_bar(position = "dodge") + ggtitle("Ride Count With Respect To Month") +  
 xlab("Month") + ylab("Number Of Rides")



This chart Specifies that casual as well as annual members peak during June, July and August. It is the least when it during January and February.