Cyclistics_divvy_trip

Olivia Akabogu

5/24/2021

Holla!

Here is my first case study from the Google data analytics certificate.

Problem

The company wants to improve her earnings by making their casual riders convert to an annual member

Solution

Design marketing strategies targeted at the casual riders. But how and what strategies should be used?

Business Task

first we need to know how they differ, then we will decide on the right strategy to use.

ASK PHASE

"Why and what would convert casual riders to subscribe?"

PREPARE PHASE

Where was the data gotten? It is a public data from a bike sharing company.

Data Collection

For this analysis, i used the past year (apr2020-apr2021) cause that would be recent and relevant to the objective.

```
library(tidyverse)
```

Load the packages

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(here)
```

```
## here() starts at C:/Users/ENGR OBINNA/Documents/Divvy trip
```

```
library(skimr)
library(janitor)
```

```
##
## Attaching package: 'janitor'

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

```
library(readr)
library(ggplot2)
```

```


apr2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

Load the data

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
may2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_rides
```

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
jun2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
jul2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_rides
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
aug2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_rides
```

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
sept2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
oct2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
nov2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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_double(),
##   end_station_name = col_character(),
##   end_station_id = col_double(),
##   start_lat = col_double(),
##   start_lng = col_double(),
##   end_lat = col_double(),
##   end_lng = col_double(),
##   member_casual = col_character()
## )
```

```
dec2020 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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()
## )
```

```
jan2021 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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()
## )
```

```
feb2021 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_ride
```

```
##
## -- Column specification --------------------------------------------------------
## 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()
## )
```

```
mar2021 <- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_rides
```

```
##
## -- Column specification --------------------------------------------------------
## 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()
## )
```

```
apr2021<- read_csv("C:/Users/ENGR OBINNA/Desktop/Data Analysis/Google Capstone project/Cyclistics_rides
```

```
##
## -- Column specification --------------------------------------------------------
## 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()
## )
```

PROCESS STAGE

Cleaning the data for analysis

I noticed that some columns have different data types when i tried to combine them. so i did this:

```
data_with_double <- bind_rows(apr2020,may2020,jun2020,jul2020,aug2020,sept2020,oct2020,nov2020)
data_with_char <- bind_rows(dec2020,jan2021,feb2021,mar2021,apr2021)
```

Change 'data_with_double' to character

```
data_with_double <-  mutate(data_with_double, start_station_id = as.character(start_station_id)
                                ,end_station_id = as.character(end_station_id))
```

Finally, To bind them all together:

```
all_trips <- bind_rows(data_with_double, data_with_char)
```

```
all_trips_v2 <- all_trips %>%
  # new column for the length of each ride and day,month and year
  mutate(ride_length = (ended_at - started_at),day_of_week =format(as.Date(started_at), "%A"), month = 
  # removing all empty values in the 'start-station_name'
  drop_na(start_station_name)%>%
  # remove unwanted columns
  select(-c(start_station_id, end_station_id))
```

Change month column to numeric in order to abbreviate

```
all_trips_v2$month <- as.numeric(all_trips_v2$month)
all_trips_v2$month <- month.abb[all_trips_v2$month]
```

Unite month and year column

```
  new_date <- unite(all_trips_v2, 'new_date', month, year, sep = ' ')
all_trips_v2 <- mutate(all_trips_v2, new_date)
View(all_trips_v2)
```

I want only the ride lenths that are not negative.

```
all_trips_v2 <- all_trips_v2[!(all_trips_v2$ride_length < 0),]
```

ANALYZE PHASE

Identify the behaviour of the user types

```
all_trips_v2 %>%
  group_by(member_casual)%>%
  summarise(mean = mean(ride_length), median = median(ride_length),
            max_ride = max(ride_length), min_ride = min(ride_length))
```

```
## # A tibble: 2 x 5
##   member_casual mean           median    max_ride     min_ride
##   <chr>         <drtn>         <drtn>    <drtn>       <drtn>
## 1 casual        2717.6094 secs 1278 secs 3341033 secs 0 secs
## 2 member         966.1052 secs  689 secs 3523202 secs 0 secs
```

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_week, FUN = mean)
```

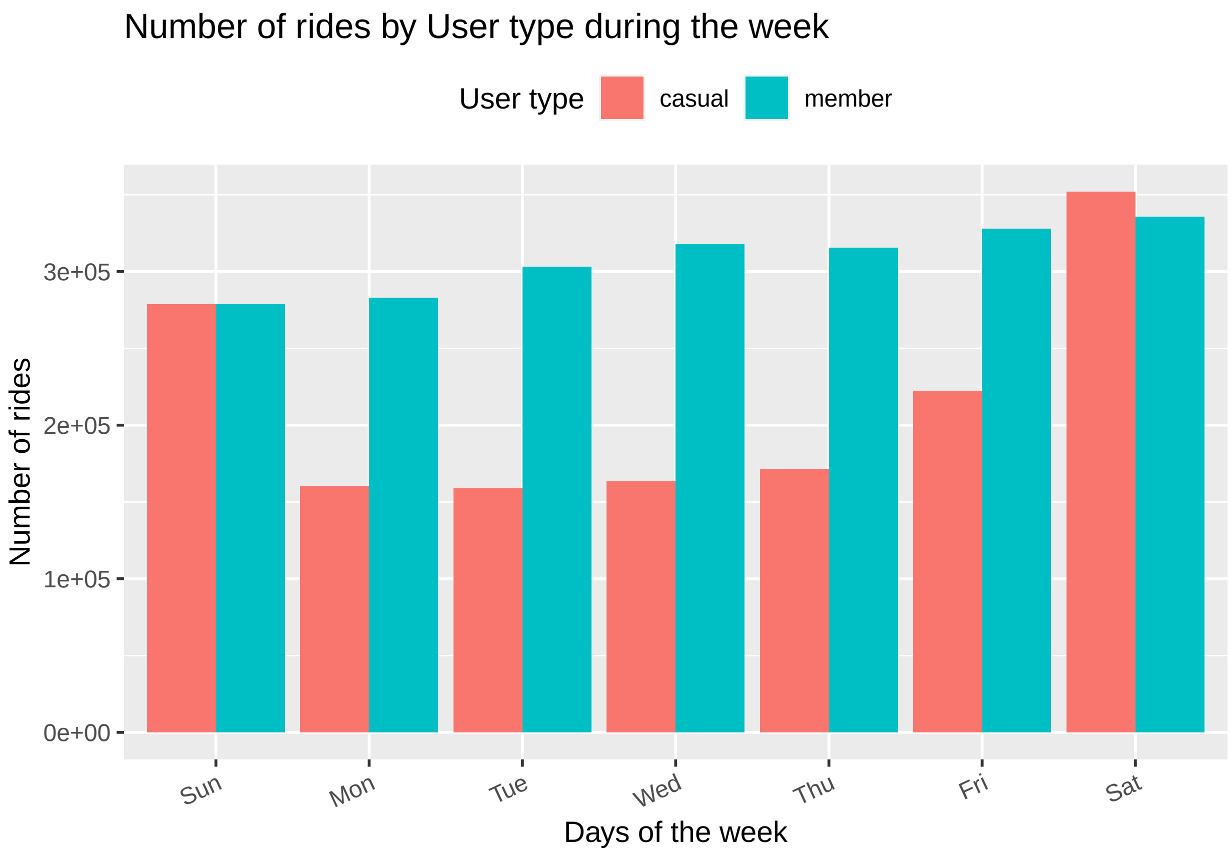
```
##    all_trips_v2$member_casual all_trips_v2$day_of_week all_trips_v2$ride_length
## 1                      casual                   Friday           2615.8820 secs
## 2                      member                   Friday            945.1852 secs
## 3                      casual                   Monday           2715.2761 secs
## 4                      member                   Monday            921.1990 secs
## 5                      casual                 Saturday           2817.0078 secs
## 6                      member                 Saturday           1070.2137 secs
## 7                      casual                   Sunday           3053.4045 secs
## 8                      member                   Sunday           1095.5406 secs
## 9                      casual                 Thursday           2562.8716 secs
## 10                     member                 Thursday            908.3416 secs
## 11                     casual                  Tuesday           2477.4782 secs
## 12                     member                  Tuesday            909.7689 secs
## 13                     casual                Wednesday           2466.9220 secs
## 14                     member                Wednesday            915.2750 secs
```

The we check the number of rides differences by weekday:

```
rides_by_type <- all_trips_v2 %>%
   mutate(weekday = wday(started_at, label = TRUE))%>%
  group_by(member_casual, weekday) %>%  #groups by usertype and weekday
  summarise(number_of_rides = n()                           #calculates the number of rides and average
  ,average_duration = mean(ride_length),.groups = 'drop') %>%       # calculates the average duration
  arrange(member_casual, weekday)                               # sorts
```

Lets see it

```
ggplot(data = rides_by_type) + geom_col(mapping = aes(x = weekday, y = number_of_rides, fill = member_ca
  labs(title = "Number of rides by User type during the week",x="Days of the week",y="Number of rides",
  theme(legend.position="top", axis.text.x = element_text(angle=25, hjust = 1))
```



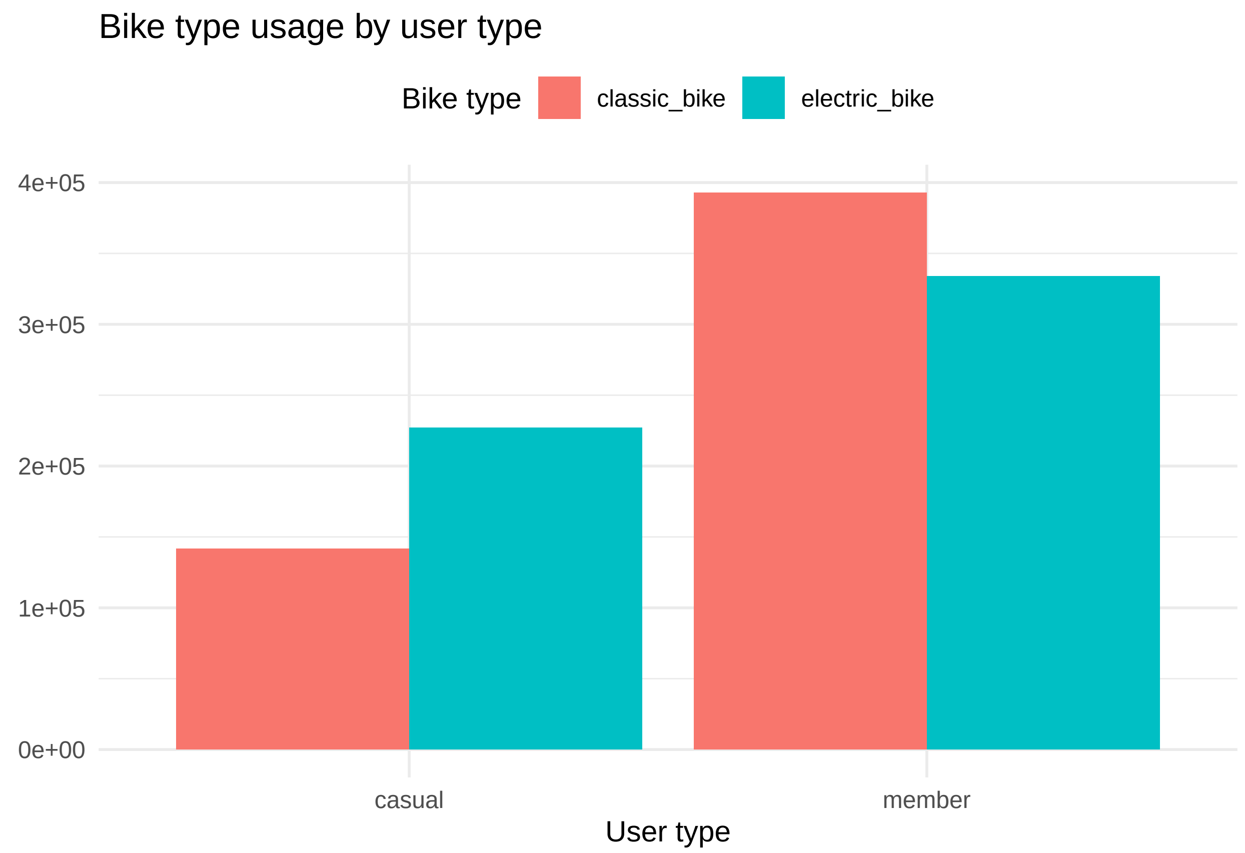
Insights

- It is seen that casual riders rides the most during the weekends especially saturday. IT is assumed that this is more of a leisure activity while the member riders uses it as public transport during the week.

```
#Create a new data frame with only the rows with info in the "bike type" column:

with_bike_type <- all_trips_v2%>% filter(rideable_type=="classic_bike" | rideable_type=="electric_bike"
```

Lets see

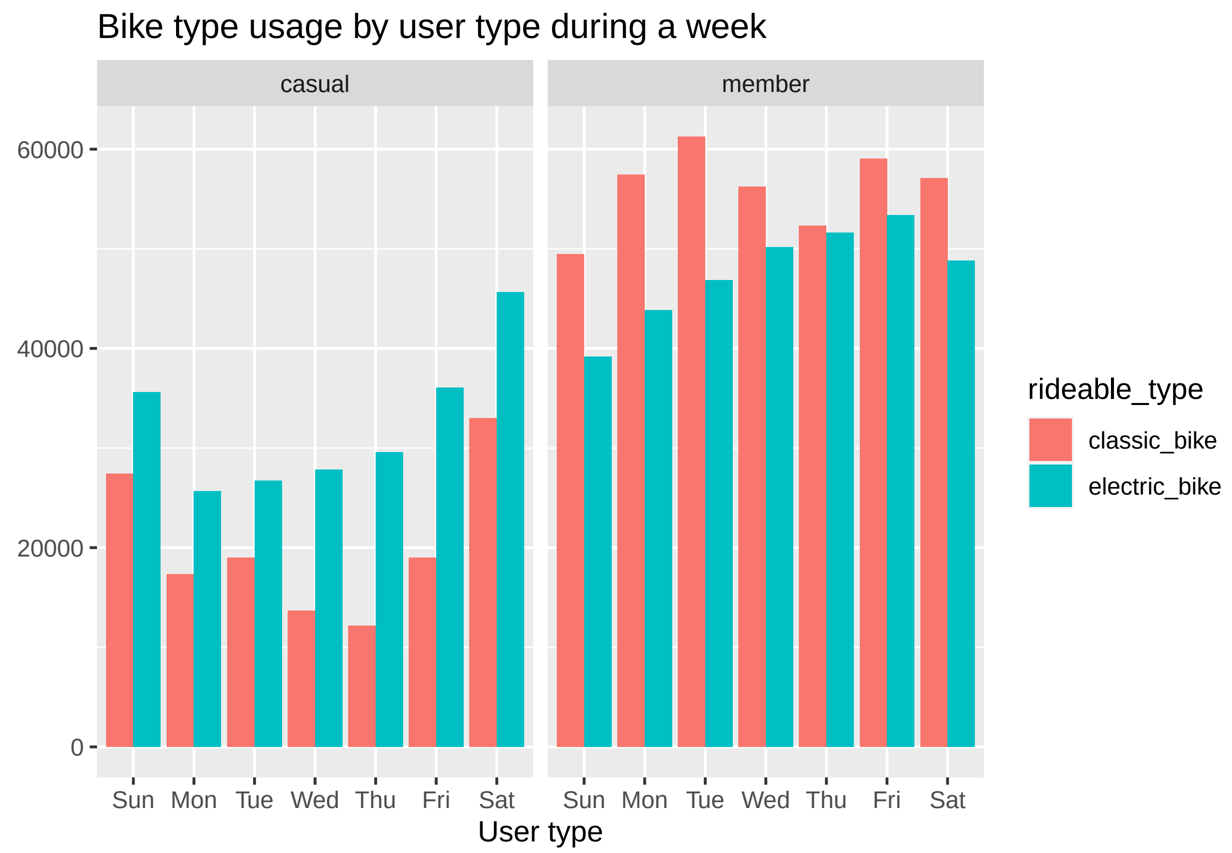
```
#Then lets check the bike type usage by user type:

with_bike_type %>%
    group_by(member_casual,rideable_type) %>%
    summarise(no_of_rides = n(), .groups = "drop")%>%
  ggplot()+
    geom_col(aes(x=member_casual,y=no_of_rides,fill=rideable_type), position = "dodge") +
    labs(title = "Bike type usage by user type",x="User type",y=NULL, fill="Bike type") +
    theme_minimal() +
    theme(legend.position="top")
```



```
with_bike_type <- with_bike_type %>%
    mutate(weekday = wday(started_at, label = TRUE)) %>%
    group_by(member_casual,rideable_type,weekday) %>%
    summarise(totals=n(), .groups="drop")
```

Lets see

```
ggplot(data = with_bike_type) + geom_col(aes(x=weekday,y=totals, fill=rideable_type), position = "dodge
  labs(title = "Bike type usage by user type during a week",x="User type",y=NULL)
```



### Insights * It is shown that annual members prefer classic bikes to electric bikes which casual is the reverse * Annual members use of electric bikes towards the end of working days increases * Just as seen earlier, casual riders are more of weekend riders

SHARE PHASE

I would share my findings with these conclusions:

- Casual riders are rides more during the weekends using electric bikes
- Annual riders use this service as a commute or public transport during the week prefering classic bikes

ACT PHASE

Recommendations

- Now that it has been known how these users differ, it would be recommended to target their marketing strategies to the weekends or electric bikes.

Thank You