Google Data Analytics Capstone

Completed On : 25 June 2021

By Arun Solviya

Introduction

This case study is the part of the Google Data Analytics Professional Certificate offered as a capstone project. This projects demonstrates the learning and skills gained through the course and their application into a real life project. The aim of this project is to build a portfolio to be shared with the potential employer.

Phase 1 : ASK

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.

Business Task

Find the answers to the following questions :
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?

Stakeholders/Audience

Cyclistic, Lily Moreno, Cyclistic marketing analytics team and Cyclistic executive team.

Phase 2 : PREPARE

In this process we will import all the dataset that are arranged monthly wise from April 2020 to March 2021. The data has been provided by Motivate International Inc. under this License (https://www.divvybikes.com/data-license-agreement). Data is a collection of 12 months bike trips data from June 2020 to May 2021. Size of the data is approximately 600mb which is difficult to work on in spreadsheet softwares as well as SQL. The most suitable option for this analysis is the use of R which can both handle this big data and also provide powerful visualizations too.

Import libraries

```
#install.pacakages(tidyverse)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.4     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## 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()
```

```
#install.pacakages(readr)
library(readr)
#install.pacakages(ggplot2)
library(ggplot2)
#install.pacakages(tidyr)
library(tidyr)
#install.pacakages(janitor)
#install.packages(lubridate)
#install.packages(zoo)
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(lubridate)
```

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

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

```
library(dplyr)
library(zoo)
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

Import Data

```
apr20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202004-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
may20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202005-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
jun20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202006-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
jul20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202007-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
aug20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202008-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
sept20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202009-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
oct20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202010-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
nov20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202011-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
dec20 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202012-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
jan21 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202101-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
feb21 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202102-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

```
mar21 <- read_csv("C:/Users/ADMIN/Desktop/R Projects/202103-divvy-tripdata.csv")
```

```
##
## -- 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()
## )
```

Phase 3 : PROCESS

In process phase we clean our data to resolve inconsistencies or mistakes. Here we have to make sure that the columns from all tables have consistent and identical names. If this condition is true, then it will return “Columns are consistent”

```
if(colnames(apr20)==colnames(may20)
   &&colnames(may20)==colnames(jun20)
   &&colnames(jun20)==colnames(jul20)
   &&colnames(jul20)==colnames(aug20)
   &&colnames(aug20)==colnames(sept20)
   &&colnames(sept20)==colnames(oct20)
   &&colnames(oct20)==colnames(nov20)
   &&colnames(nov20)==colnames(dec20)
   &&colnames(dec20)==colnames(jan21)
   &&colnames(jan21)==colnames(feb21)
   &&colnames(feb21)==colnames(mar21)){
print("Columns are cosistent")
}else {
print("Inconsistent columns")
}
```

```
## [1] "Columns are cosistent"
```

Inspect Data

let us check the column names from one of the tables.

```
colnames(apr20)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

let us check the summary of our tables.

```
summary(apr20)
```

```
##    ride_id          rideable_type        started_at
##  Length:84776       Length:84776       Min.   :2020-04-01 00:00:30
##  Class :character   Class :character   1st Qu.:2020-04-07 22:40:26
##  Mode  :character   Mode  :character   Median :2020-04-16 15:23:46
##                                        Mean   :2020-04-15 22:41:37
##                                        3rd Qu.:2020-04-22 19:47:21
##                                        Max.   :2020-04-30 23:57:20
##
##    ended_at                  start_station_name start_station_id
##  Min.   :2020-04-01 00:10:45   Length:84776       Min.   :  2.0
##  1st Qu.:2020-04-08 00:13:07   Class :character   1st Qu.:113.0
##  Median :2020-04-16 15:48:15   Mode  :character   Median :211.0
##  Mean   :2020-04-15 23:17:28                      Mean   :236.7
##  3rd Qu.:2020-04-22 20:24:14                      3rd Qu.:324.0
##  Max.   :2020-05-30 15:30:55                      Max.   :673.0
##
##  end_station_name   end_station_id   start_lat       start_lng
##  Length:84776       Min.   :  2    Min.   :41.74   Min.   :-87.77
##  Class :character   1st Qu.:113    1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :212    Median :41.90   Median :-87.65
##                     Mean   :237    Mean   :41.91   Mean   :-87.65
##                     3rd Qu.:323    3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :673    Max.   :42.06   Max.   :-87.55
##                     NA's   :99
##    end_lat         end_lng       member_casual
##  Min.   :41.74   Min.   :-87.77   Length:84776
##  1st Qu.:41.88   1st Qu.:-87.67   Class :character
##  Median :41.90   Median :-87.65   Mode  :character
##  Mean   :41.91   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.06   Max.   :-87.55
##  NA's   :99      NA's   :99
```

```
summary(may20)
```

```
##    ride_id          rideable_type        started_at
##  Length:200274      Length:200274      Min.   :2020-05-01 00:02:07
##  Class :character   Class :character   1st Qu.:2020-05-09 18:37:16
##  Mode  :character   Mode  :character   Median :2020-05-21 12:27:12
##                                        Mean   :2020-05-18 14:44:36
##                                        3rd Qu.:2020-05-26 07:00:03
##                                        Max.   :2020-05-31 02:58:45
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-05-01 00:13:03   Length:200274      Min.   :  2.0
##  1st Qu.:2020-05-09 19:09:20   Class :character   1st Qu.:112.0
##  Median :2020-05-21 12:53:36   Mode  :character   Median :211.0
##  Mean   :2020-05-18 15:17:59                      Mean   :235.5
##  3rd Qu.:2020-05-26 07:35:25                      3rd Qu.:322.0
##  Max.   :2020-05-31 03:03:04                      Max.   :673.0
##
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:200274      Min.   :  2.0   Min.   :41.74   Min.   :-87.77
##  Class :character   1st Qu.:113.0   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :212.0   Median :41.90   Median :-87.64
##                     Mean   :237.8   Mean   :41.91   Mean   :-87.65
##                     3rd Qu.:324.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :673.0   Max.   :42.06   Max.   :-87.55
##                     NA's   :321
##     end_lat         end_lng       member_casual
##  Min.   :41.74   Min.   :-87.77   Length:200274
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.65   Mode  :character
##  Mean   :41.91   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.06   Max.   :-87.55
##  NA's   :321     NA's   :321
```

```
summary(jun20)
```

```
##    ride_id          rideable_type        started_at
##  Length:343005      Length:343005      Min.   :2020-06-03 05:59:59
##  Class :character   Class :character   1st Qu.:2020-06-11 20:17:32
##  Mode  :character   Mode  :character   Median :2020-06-18 17:54:45
##                                        Mean   :2020-06-18 06:21:04
##                                        3rd Qu.:2020-06-25 08:09:45
##                                        Max.   :2020-06-30 23:59:54
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-06-03 06:03:37   Length:343005      Min.   :  2.0
##  1st Qu.:2020-06-11 20:51:30   Class :character   1st Qu.:106.0
##  Median :2020-06-18 18:21:52   Mode  :character   Median :199.0
##  Mean   :2020-06-18 06:54:36                      Mean   :225.2
##  3rd Qu.:2020-06-25 08:32:03                      3rd Qu.:312.0
##  Max.   :2020-07-03 20:26:15                      Max.   :673.0
##
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:343005      Min.   :  2.0   Min.   :41.74   Min.   :-87.77
##  Class :character   1st Qu.:106.0   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :199.0   Median :41.90   Median :-87.64
##                     Mean   :226.6   Mean   :41.91   Mean   :-87.65
##                     3rd Qu.:313.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :674.0   Max.   :42.06   Max.   :-87.55
##                     NA's   :468
##     end_lat         end_lng       member_casual
##  Min.   :41.74   Min.   :-87.77   Length:343005
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.91   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.06   Max.   :-87.55
##  NA's   :468     NA's   :468
```

```
summary(jul20)
```

```
##    ride_id          rideable_type        started_at
##  Length:551480      Length:551480      Min.   :2020-07-01 00:00:14
##  Class :character   Class :character   1st Qu.:2020-07-09 07:57:27
##  Mode  :character   Mode  :character   Median :2020-07-17 07:38:14
##                                        Mean   :2020-07-16 21:30:52
##                                        3rd Qu.:2020-07-24 20:26:55
##                                        Max.   :2020-07-31 23:59:50
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-07-01 00:03:01   Length:551480      Min.   :  2.0
##  1st Qu.:2020-07-09 08:22:07   Class :character   1st Qu.: 94.0
##  Median :2020-07-17 08:04:42   Mode  :character   Median :195.0
##  Mean   :2020-07-16 22:09:08                      Mean   :221.6
##  3rd Qu.:2020-07-24 21:05:12                      3rd Qu.:308.0
##  Max.   :2020-08-11 01:01:26                      Max.   :683.0
##                                                   NA's   :152
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:551480      Min.   :  2.0   Min.   :41.74   Min.   :-87.77
##  Class :character   1st Qu.: 94.0   1st Qu.:41.88   1st Qu.:-87.65
##  Mode  :character   Median :195.0   Median :41.90   Median :-87.64
##                     Mean   :222.4   Mean   :41.90   Mean   :-87.64
##                     3rd Qu.:309.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :683.0   Max.   :42.06   Max.   :-87.55
##                     NA's   :969
##     end_lat         end_lng       member_casual
##  Min.   :41.73   Min.   :-87.77   Length:551480
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.91   Mean   :-87.64
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.06   Max.   :-87.55
##  NA's   :770     NA's   :770
```

```
summary(aug20)
```

```
##    ride_id          rideable_type        started_at
##  Length:622361      Length:622361      Min.   :2020-08-01 00:00:01
##  Class :character   Class :character   1st Qu.:2020-08-08 21:00:43
##  Mode  :character   Mode  :character   Median :2020-08-16 18:58:31
##                                        Mean   :2020-08-16 23:27:26
##                                        3rd Qu.:2020-08-24 11:43:27
##                                        Max.   :2020-08-31 23:58:55
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-08-01 00:04:41   Length:622361      Min.   :  2.0
##  1st Qu.:2020-08-08 21:38:05   Class :character   1st Qu.: 96.0
##  Median :2020-08-16 19:32:57   Mode  :character   Median :196.0
##  Mean   :2020-08-16 23:57:14                      Mean   :225.4
##  3rd Qu.:2020-08-24 12:07:29                      3rd Qu.:312.0
##  Max.   :2020-09-02 21:21:53                      Max.   :700.0
##                                                   NA's   :7691
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:622361      Min.   :  2.0   Min.   :41.66   Min.   :-87.87
##  Class :character   1st Qu.: 97.0   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :196.0   Median :41.90   Median :-87.64
##                     Mean   :225.8   Mean   :41.91   Mean   :-87.64
##                     3rd Qu.:312.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :700.0   Max.   :42.07   Max.   :-87.53
##                     NA's   :10110
##     end_lat         end_lng       member_casual
##  Min.   :41.66   Min.   :-87.89   Length:622361
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.91   Mean   :-87.64
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.16   Max.   :-87.53
##  NA's   :938     NA's   :938
```

```
summary(sept20)
```

```
##    ride_id          rideable_type        started_at
##  Length:532958      Length:532958      Min.   :2020-09-01 00:00:07
##  Class :character   Class :character   1st Qu.:2020-09-07 16:12:36
##  Mode  :character   Mode  :character   Median :2020-09-16 17:37:38
##                                        Mean   :2020-09-16 05:33:47
##                                        3rd Qu.:2020-09-23 18:23:47
##                                        Max.   :2020-09-30 23:58:39
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-09-01 00:04:43   Length:532958      Min.   :  2.0
##  1st Qu.:2020-09-07 16:47:07   Class :character   1st Qu.: 94.0
##  Median :2020-09-16 17:58:02   Mode  :character   Median :194.0
##  Mean   :2020-09-16 05:59:03                      Mean   :223.6
##  3rd Qu.:2020-09-23 18:45:02                      3rd Qu.:308.0
##  Max.   :2020-10-12 11:46:25                      Max.   :721.0
##                                                   NA's   :19901
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:532958      Min.   :  2.0   Min.   :41.65   Min.   :-87.84
##  Class :character   1st Qu.: 94.0   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :194.0   Median :41.90   Median :-87.64
##                     Mean   :224.2   Mean   :41.90   Mean   :-87.64
##                     3rd Qu.:309.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :721.0   Max.   :42.08   Max.   :-87.52
##                     NA's   :23524
##     end_lat         end_lng       member_casual
##  Min.   :41.64   Min.   :-87.88   Length:532958
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.64
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.12   Max.   :-87.52
##  NA's   :789     NA's   :789
```

```
summary(oct20)
```

```
##    ride_id          rideable_type        started_at
##  Length:388653      Length:388653      Min.   :2020-10-01 00:00:06
##  Class :character   Class :character   1st Qu.:2020-10-07 19:15:11
##  Mode  :character   Mode  :character   Median :2020-10-13 09:24:41
##                                        Mean   :2020-10-14 18:47:52
##                                        3rd Qu.:2020-10-21 17:21:49
##                                        Max.   :2020-10-31 23:59:50
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-10-01 00:05:09   Length:388653      Min.   :  2.0
##  1st Qu.:2020-10-07 19:34:58   Class :character   1st Qu.: 96.0
##  Median :2020-10-13 09:42:11   Mode  :character   Median :195.0
##  Mean   :2020-10-14 19:07:52                      Mean   :225.1
##  3rd Qu.:2020-10-21 17:38:20                      3rd Qu.:312.0
##  Max.   :2020-11-02 20:10:55                      Max.   :731.0
##                                                   NA's   :31405
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:388653      Min.   :  2.0   Min.   :41.64   Min.   :-87.80
##  Class :character   1st Qu.: 96.0   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :195.0   Median :41.90   Median :-87.64
##                     Mean   :224.5   Mean   :41.90   Mean   :-87.65
##                     3rd Qu.:311.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :731.0   Max.   :42.08   Max.   :-87.52
##                     NA's   :35787
##     end_lat         end_lng       member_casual
##  Min.   :41.63   Min.   :-87.84   Length:388653
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.15   Max.   :-87.52
##  NA's   :474     NA's   :474
```

```
summary(nov20)
```

```
##    ride_id          rideable_type        started_at
##  Length:259716      Length:259716      Min.   :2020-11-01 00:00:08
##  Class :character   Class :character   1st Qu.:2020-11-06 20:13:53
##  Mode  :character   Mode  :character   Median :2020-11-10 16:17:15
##                                        Mean   :2020-11-13 05:20:02
##                                        3rd Qu.:2020-11-19 15:59:47
##                                        Max.   :2020-11-30 23:56:22
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-11-01 00:02:20   Length:259716      Min.   :  2.0
##  1st Qu.:2020-11-06 20:36:08   Class :character   1st Qu.: 99.0
##  Median :2020-11-10 16:32:15   Mode  :character   Median :198.0
##  Mean   :2020-11-13 05:39:45                      Mean   :228.7
##  3rd Qu.:2020-11-19 16:18:54                      3rd Qu.:318.0
##  Max.   :2020-12-01 17:28:22                      Max.   :732.0
##                                                   NA's   :24434
##  end_station_name   end_station_id    start_lat       start_lng
##  Length:259716      Min.   :  2.0   Min.   :41.65   Min.   :-87.79
##  Class :character   1st Qu.: 99.0   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Median :198.0   Median :41.90   Median :-87.64
##                     Mean   :227.9   Mean   :41.90   Mean   :-87.65
##                     3rd Qu.:315.0   3rd Qu.:41.93   3rd Qu.:-87.63
##                     Max.   :732.0   Max.   :42.08   Max.   :-87.53
##                     NA's   :26826
##     end_lat         end_lng       member_casual
##  Min.   :41.54   Min.   :-87.87   Length:259716
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.15   Max.   :-87.44
##  NA's   :284     NA's   :284
```

```
summary(dec20)
```

```
##    ride_id          rideable_type        started_at
##  Length:131573      Length:131573      Min.   :2020-12-01 00:01:15
##  Class :character   Class :character   1st Qu.:2020-12-07 09:56:34
##  Mode  :character   Mode  :character   Median :2020-12-13 17:15:24
##                                        Mean   :2020-12-14 13:23:37
##                                        3rd Qu.:2020-12-21 08:12:40
##                                        Max.   :2020-12-31 23:59:59
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2020-11-25 07:40:56   Length:131573      Length:131573
##  1st Qu.:2020-12-07 08:32:00   Class :character   Class :character
##  Median :2020-12-13 16:33:17   Mode  :character   Mode  :character
##  Mean   :2020-12-14 12:17:25
##  3rd Qu.:2020-12-21 08:23:49
##  Max.   :2021-01-03 08:54:11
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:131573      Length:131573      Min.   :41.65   Min.   :-87.78
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.65
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.07   Max.   :-87.53
##
##     end_lat         end_lng       member_casual
##  Min.   :41.65   Min.   :-87.79   Length:131573
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.07   Max.   :-87.53
##  NA's   :111     NA's   :111
```

```
summary(jan21)
```

```
##    ride_id          rideable_type        started_at
##  Length:96834       Length:96834       Min.   :2021-01-01 00:02:05
##  Class :character   Class :character   1st Qu.:2021-01-08 20:55:02
##  Mode  :character   Mode  :character   Median :2021-01-15 06:05:04
##                                        Mean   :2021-01-15 17:57:29
##                                        3rd Qu.:2021-01-22 09:28:48
##                                        Max.   :2021-01-31 23:57:00
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2021-01-01 00:08:39   Length:96834       Length:96834
##  1st Qu.:2021-01-08 21:14:23   Class :character   Class :character
##  Median :2021-01-15 06:19:58   Mode  :character   Mode  :character
##  Mean   :2021-01-15 18:12:46
##  3rd Qu.:2021-01-22 09:41:18
##  Max.   :2021-02-01 15:33:15
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:96834       Length:96834       Min.   :41.64   Min.   :-87.78
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.65
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.06   Max.   :-87.53
##
##     end_lat         end_lng       member_casual
##  Min.   :41.64   Min.   :-87.81   Length:96834
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.07   Max.   :-87.51
##  NA's   :103     NA's   :103
```

```
summary(feb21)
```

```
##    ride_id          rideable_type        started_at
##  Length:49622       Length:49622       Min.   :2021-02-01 00:55:44
##  Class :character   Class :character   1st Qu.:2021-02-09 08:20:56
##  Mode  :character   Mode  :character   Median :2021-02-22 13:17:53
##                                        Mean   :2021-02-18 01:16:52
##                                        3rd Qu.:2021-02-26 16:02:13
##                                        Max.   :2021-02-28 23:59:41
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2021-02-01 01:22:48   Length:49622       Length:49622
##  1st Qu.:2021-02-09 08:36:02   Class :character   Class :character
##  Median :2021-02-22 13:39:20   Mode  :character   Mode  :character
##  Mean   :2021-02-18 01:41:18
##  3rd Qu.:2021-02-26 16:19:32
##  Max.   :2021-03-05 15:11:45
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:49622       Length:49622       Min.   :41.65   Min.   :-87.77
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.64
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.06   Max.   :-87.53
##
##     end_lat         end_lng       member_casual
##  Min.   :41.54   Min.   :-87.77   Length:49622
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.64
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.07   Max.   :-87.53
##  NA's   :214     NA's   :214
```

```
summary(mar21)
```

```
##    ride_id          rideable_type        started_at
##  Length:228496      Length:228496      Min.   :2021-03-01 00:01:09
##  Class :character   Class :character   1st Qu.:2021-03-10 10:45:36
##  Mode  :character   Mode  :character   Median :2021-03-19 17:37:20
##                                        Mean   :2021-03-17 23:22:08
##                                        3rd Qu.:2021-03-25 08:39:23
##                                        Max.   :2021-03-31 23:59:08
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2021-03-01 00:06:28   Length:228496      Length:228496
##  1st Qu.:2021-03-10 11:04:40   Class :character   Class :character
##  Median :2021-03-19 17:55:05   Mode  :character   Mode  :character
##  Mean   :2021-03-17 23:45:00
##  3rd Qu.:2021-03-25 08:54:12
##  Max.   :2021-04-06 11:00:11
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:228496      Length:228496      Min.   :41.65   Min.   :-87.78
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.64
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.07   Max.   :-87.53
##
##     end_lat         end_lng       member_casual
##  Min.   :41.64   Min.   :-88.07   Length:228496
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character
##  Median :41.90   Median :-87.64   Mode  :character
##  Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.08   Max.   :-87.53
##  NA's   :167     NA's   :167
```

let us check the data structure of our tables.

```
str(apr20)
```

```
## spec_tbl_df [84,776 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:84776] "A847FADBBC638E45" "5405B80E996FF60D" "5DD24A79A4E006
F4" "2A59BBDF5CDBA725" ...
##  $ rideable_type     : chr [1:84776] "docked_bike" "docked_bike" "docked_bike" "docked_bik
e" ...
##  $ started_at        : POSIXct[1:84776], format: "2020-04-26 17:45:14" "2020-04-17 17:08:5
4" ...
##  $ ended_at          : POSIXct[1:84776], format: "2020-04-26 18:12:03" "2020-04-17 17:17:0
3" ...
##  $ start_station_name: chr [1:84776] "Eckhart Park" "Drake Ave & Fullerton Ave" "McClurg C
t & Erie St" "California Ave & Division St" ...
##  $ start_station_id  : num [1:84776] 86 503 142 216 125 173 35 434 627 377 ...
##  $ end_station_name  : chr [1:84776] "Lincoln Ave & Diversey Pkwy" "Kosciuszko Park" "Indi
ana Ave & Roosevelt Rd" "Wood St & Augusta Blvd" ...
##  $ end_station_id    : num [1:84776] 152 499 255 657 323 35 635 382 359 508 ...
##  $ start_lat         : num [1:84776] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:84776] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:84776] 41.9 41.9 41.9 41.9 42 ...
##  $ end_lng           : num [1:84776] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:84776] "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_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()
##   .. )
```

```
str(may20)
```

```
## spec_tbl_df [200,274 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:200274] "02668AD35674B983" "7A50CCAF1EDDB28F" "2FFCDFDB91FE9
A52" "58991CF1DB75BA84" ...
##  $ rideable_type     : chr [1:200274] "docked_bike" "docked_bike" "docked_bike" "docked_bi
ke" ...
##  $ started_at        : POSIXct[1:200274], format: "2020-05-27 10:03:52" "2020-05-25 10:47:
11" ...
##  $ ended_at          : POSIXct[1:200274], format: "2020-05-27 10:16:49" "2020-05-25 11:05:
40" ...
##  $ start_station_name: chr [1:200274] "Franklin St & Jackson Blvd" "Clark St & Wrightwood
Ave" "Kedzie Ave & Milwaukee Ave" "Clarendon Ave & Leland Ave" ...
##  $ start_station_id  : num [1:200274] 36 340 260 251 261 206 261 180 331 219 ...
##  $ end_station_name  : chr [1:200274] "Wabash Ave & Grand Ave" "Clark St & Leland Ave" "Ke
dzie Ave & Milwaukee Ave" "Lake Shore Dr & Wellington Ave" ...
##  $ end_station_id    : num [1:200274] 199 326 260 157 206 22 261 180 300 305 ...
##  $ start_lat         : num [1:200274] 41.9 41.9 41.9 42 41.9 ...
##  $ start_lng         : num [1:200274] -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:200274] 41.9 42 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:200274] -87.6 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:200274] "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_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()
##   .. )
```

```
str(jun20)
```

```
## spec_tbl_df [343,005 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:343005] "8CD5DE2C2B6C4CFC" "9A191EB2C751D85D" "F37D14B0B5659
BCF" "C41237B506E85FA1" ...
##  $ rideable_type     : chr [1:343005] "docked_bike" "docked_bike" "docked_bike" "docked_bi
ke" ...
##  $ started_at        : POSIXct[1:343005], format: "2020-06-13 23:24:48" "2020-06-26 07:26:
10" ...
##  $ ended_at          : POSIXct[1:343005], format: "2020-06-13 23:36:55" "2020-06-26 07:31:
58" ...
##  $ start_station_name: chr [1:343005] "Wilton Ave & Belmont Ave" "Federal St & Polk St" "D
aley Center Plaza" "Broadway & Cornelia Ave" ...
##  $ start_station_id  : num [1:343005] 117 41 81 303 327 327 41 115 338 84 ...
##  $ end_station_name  : chr [1:343005] "Damen Ave & Clybourn Ave" "Daley Center Plaza" "Sta
te St & Harrison St" "Broadway & Berwyn Ave" ...
##  $ end_station_id    : num [1:343005] 163 81 5 294 117 117 81 303 164 53 ...
##  $ start_lat         : num [1:343005] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:343005] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:343005] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:343005] -87.7 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:343005] "casual" "member" "member" "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_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()
##   .. )
```

```
str(jul20)
```

```
## spec_tbl_df [551,480 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:551480] "762198876D69004D" "BEC9C9FBA0D4CF1B" "D2FD8EA432C77
EC1" "54AE594E20B35881" ...
##  $ rideable_type     : chr [1:551480] "docked_bike" "docked_bike" "docked_bike" "docked_bi
ke" ...
##  $ started_at        : POSIXct[1:551480], format: "2020-07-09 15:22:02" "2020-07-24 23:56:
30" ...
##  $ ended_at          : POSIXct[1:551480], format: "2020-07-09 15:25:52" "2020-07-25 00:20:
17" ...
##  $ start_station_name: chr [1:551480] "Ritchie Ct & Banks St" "Halsted St & Roscoe St" "La
ke Shore Dr & Diversey Pkwy" "LaSalle St & Illinois St" ...
##  $ start_station_id  : num [1:551480] 180 299 329 181 268 635 113 211 176 31 ...
##  $ end_station_name  : chr [1:551480] "Wells St & Evergreen Ave" "Broadway & Ridge Ave" "C
lark St & Wellington Ave" "Clark St & Armitage Ave" ...
##  $ end_station_id    : num [1:551480] 291 461 156 94 301 289 140 31 191 142 ...
##  $ start_lat         : num [1:551480] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:551480] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:551480] 41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:551480] -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:551480] "member" "member" "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_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()
##   .. )
```

```
str(aug20)
```

```
## spec_tbl_df [622,361 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:622361] "322BD23D287743ED" "2A3AEF1AB9054D8B" "67DC1D133E8B5
816" "C79FBBD412E578A7" ...
##  $ rideable_type     : chr [1:622361] "docked_bike" "electric_bike" "electric_bike" "elect
ric_bike" ...
##  $ started_at        : POSIXct[1:622361], format: "2020-08-20 18:08:14" "2020-08-27 18:46:
04" ...
##  $ ended_at          : POSIXct[1:622361], format: "2020-08-20 18:17:51" "2020-08-27 19:54:
51" ...
##  $ start_station_name: chr [1:622361] "Lake Shore Dr & Diversey Pkwy" "Michigan Ave & 14th
St" "Columbus Dr & Randolph St" "Daley Center Plaza" ...
##  $ start_station_id  : num [1:622361] 329 168 195 81 658 658 196 67 153 177 ...
##  $ end_station_name  : chr [1:622361] "Clark St & Lincoln Ave" "Michigan Ave & 14th St" "S
tate St & Randolph St" "State St & Kinzie St" ...
##  $ end_station_id    : num [1:622361] 141 168 44 47 658 658 49 229 225 305 ...
##  $ start_lat         : num [1:622361] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:622361] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:622361] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:622361] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:622361] "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_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()
##   .. )
```

```
str(sept20)
```

```
## spec_tbl_df [532,958 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:532958] "2B22BD5F95FB2629" "A7FB70B4AFC6CAF2" "86057FA01BAC7
78E" "57F6DC9A153DB98C" ...
##  $ rideable_type     : chr [1:532958] "electric_bike" "electric_bike" "electric_bike" "ele
ctric_bike" ...
##  $ started_at        : POSIXct[1:532958], format: "2020-09-17 14:27:11" "2020-09-17 15:07:
31" ...
##  $ ended_at          : POSIXct[1:532958], format: "2020-09-17 14:44:24" "2020-09-17 15:07:
45" ...
##  $ start_station_name: chr [1:532958] "Michigan Ave & Lake St" "W Oakdale Ave & N Broadwa
y" "W Oakdale Ave & N Broadway" "Ashland Ave & Belle Plaine Ave" ...
##  $ start_station_id  : num [1:532958] 52 NA NA 246 24 94 291 NA NA NA ...
##  $ end_station_name  : chr [1:532958] "Green St & Randolph St" "W Oakdale Ave & N Broadwa
y" "W Oakdale Ave & N Broadway" "Montrose Harbor" ...
##  $ end_station_id    : num [1:532958] 112 NA NA 249 24 NA 256 NA NA NA ...
##  $ start_lat         : num [1:532958] 41.9 41.9 41.9 42 41.9 ...
##  $ start_lng         : num [1:532958] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:532958] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:532958] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:532958] "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_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()
##   .. )
```

```
str(oct20)
```

```
## spec_tbl_df [388,653 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:388653] "ACB6B40CF5B9044C" "DF450C72FD109C01" "B6396B54A15AC
0DF" "44A4AEE261B9E854" ...
##  $ rideable_type     : chr [1:388653] "electric_bike" "electric_bike" "electric_bike" "ele
ctric_bike" ...
##  $ started_at        : POSIXct[1:388653], format: "2020-10-31 19:39:43" "2020-10-31 23:50:
08" ...
##  $ ended_at          : POSIXct[1:388653], format: "2020-10-31 19:57:12" "2020-11-01 00:04:
16" ...
##  $ start_station_name: chr [1:388653] "Lakeview Ave & Fullerton Pkwy" "Southport Ave & Wav
eland Ave" "Stony Island Ave & 67th St" "Clark St & Grace St" ...
##  $ start_station_id  : num [1:388653] 313 227 102 165 190 359 313 125 NA 174 ...
##  $ end_station_name  : chr [1:388653] "Rush St & Hubbard St" "Kedzie Ave & Milwaukee Ave"
"University Ave & 57th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : num [1:388653] 125 260 423 256 185 53 125 313 199 635 ...
##  $ start_lat         : num [1:388653] 41.9 41.9 41.8 42 41.9 ...
##  $ start_lng         : num [1:388653] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:388653] 41.9 41.9 41.8 42 41.9 ...
##  $ end_lng           : num [1:388653] -87.6 -87.7 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:388653] "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_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()
##   .. )
```

```
str(nov20)
```

```
## spec_tbl_df [259,716 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:259716] "BD0A6FF6FFF9B921" "96A7A7A4BDE4F82D" "C61526D06582B
DC5" "E533E89C32080B9E" ...
##  $ rideable_type     : chr [1:259716] "electric_bike" "electric_bike" "electric_bike" "ele
ctric_bike" ...
##  $ started_at        : POSIXct[1:259716], format: "2020-11-01 13:36:00" "2020-11-01 10:03:
26" ...
##  $ ended_at          : POSIXct[1:259716], format: "2020-11-01 13:45:40" "2020-11-01 10:14:
45" ...
##  $ start_station_name: chr [1:259716] "Dearborn St & Erie St" "Franklin St & Illinois St"
"Lake Shore Dr & Monroe St" "Leavitt St & Chicago Ave" ...
##  $ start_station_id  : num [1:259716] 110 672 76 659 2 72 76 NA 58 394 ...
##  $ end_station_name  : chr [1:259716] "St. Clair St & Erie St" "Noble St & Milwaukee Ave"
"Federal St & Polk St" "Stave St & Armitage Ave" ...
##  $ end_station_id    : num [1:259716] 211 29 41 185 2 76 72 NA 288 273 ...
##  $ start_lat         : num [1:259716] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:259716] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:259716] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:259716] -87.6 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:259716] "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_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()
##   .. )
```

```
str(dec20)
```

```
## spec_tbl_df [131,573 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:131573] "70B6A9A437D4C30D" "158A465D4E74C54A" "5262016E0F1F2
F9A" "BE119628E44F871E" ...
##  $ rideable_type     : chr [1:131573] "classic_bike" "electric_bike" "electric_bike" "elec
tric_bike" ...
##  $ started_at        : POSIXct[1:131573], format: "2020-12-27 12:44:29" "2020-12-18 17:37:
15" ...
##  $ ended_at          : POSIXct[1:131573], format: "2020-12-27 12:55:06" "2020-12-18 17:44:
19" ...
##  $ start_station_name: chr [1:131573] "Aberdeen St & Jackson Blvd" NA NA NA ...
##  $ start_station_id  : chr [1:131573] "13157" NA NA NA ...
##  $ end_station_name  : chr [1:131573] "Desplaines St & Kinzie St" NA NA NA ...
##  $ end_station_id    : chr [1:131573] "TA1306000003" NA NA NA ...
##  $ start_lat         : num [1:131573] 41.9 41.9 41.9 41.9 41.8 ...
##  $ start_lng         : num [1:131573] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:131573] 41.9 41.9 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:131573] -87.6 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:131573] "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()
##   .. )
```

```
str(jan21)
```

```
## spec_tbl_df [96,834 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:96834] "E19E6F1B8D4C42ED" "DC88F20C2C55F27F" "EC45C94683FE3F
27" "4FA453A75AE377DB" ...
##  $ rideable_type     : chr [1:96834] "electric_bike" "electric_bike" "electric_bike" "elec
tric_bike" ...
##  $ started_at        : POSIXct[1:96834], format: "2021-01-23 16:14:19" "2021-01-27 18:43:0
8" ...
##  $ ended_at          : POSIXct[1:96834], format: "2021-01-23 16:24:44" "2021-01-27 18:47:1
2" ...
##  $ start_station_name: chr [1:96834] "California Ave & Cortez St" "California Ave & Cortez
St" "California Ave & Cortez St" "California Ave & Cortez St" ...
##  $ start_station_id  : chr [1:96834] "17660" "17660" "17660" "17660" ...
##  $ end_station_name  : chr [1:96834] NA NA NA NA ...
##  $ end_station_id    : chr [1:96834] NA NA NA NA ...
##  $ start_lat         : num [1:96834] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:96834] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:96834] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:96834] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:96834] "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()
##   .. )
```

```
str(feb21)
```

```
## spec_tbl_df [49,622 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:49622] "89E7AA6C29227EFF" "0FEFDE2603568365" "E6159D746B2DBB
91" "B32D3199F1C2E75B" ...
##  $ rideable_type     : chr [1:49622] "classic_bike" "classic_bike" "electric_bike" "classi
c_bike" ...
##  $ started_at        : POSIXct[1:49622], format: "2021-02-12 16:14:56" "2021-02-14 17:52:3
8" ...
##  $ ended_at          : POSIXct[1:49622], format: "2021-02-12 16:21:43" "2021-02-14 18:12:0
9" ...
##  $ start_station_name: chr [1:49622] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave"
"Clark St & Lake St" "Wood St & Chicago Ave" ...
##  $ start_station_id  : chr [1:49622] "525" "525" "KA1503000012" "637" ...
##  $ end_station_name  : chr [1:49622] "Sheridan Rd & Columbia Ave" "Bosworth Ave & Howard S
t" "State St & Randolph St" "Honore St & Division St" ...
##  $ end_station_id    : chr [1:49622] "660" "16806" "TA1305000029" "TA1305000034" ...
##  $ start_lat         : num [1:49622] 42 42 41.9 41.9 41.8 ...
##  $ start_lng         : num [1:49622] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:49622] 42 42 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:49622] -87.7 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:49622] "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()
##   .. )
```

```
str(mar21)
```

```
## spec_tbl_df [228,496 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:228496] "CFA86D4455AA1030" "30D9DC61227D1AF3" "846D87A15682A
284" "994D05AA75A168F2" ...
##  $ rideable_type     : chr [1:228496] "classic_bike" "classic_bike" "classic_bike" "classi
c_bike" ...
##  $ started_at        : POSIXct[1:228496], format: "2021-03-16 08:32:30" "2021-03-28 01:26:
28" ...
##  $ ended_at          : POSIXct[1:228496], format: "2021-03-16 08:36:34" "2021-03-28 01:36:
55" ...
##  $ start_station_name: chr [1:228496] "Humboldt Blvd & Armitage Ave" "Humboldt Blvd & Armi
tage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...
##  $ start_station_id  : chr [1:228496] "15651" "15651" "15443" "TA1308000021" ...
##  $ end_station_name  : chr [1:228496] "Stave St & Armitage Ave" "Central Park Ave & Bloomi
ngdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...
##  $ end_station_id    : chr [1:228496] "13266" "18017" "TA1308000043" "13323" ...
##  $ start_lat         : num [1:228496] 41.9 41.9 41.8 42 42 ...
##  $ start_lng         : num [1:228496] -87.7 -87.7 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:228496] 41.9 41.9 41.8 42 42.1 ...
##  $ end_lng           : num [1:228496] -87.7 -87.7 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:228496] "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()
##   .. )
```

let us check the data in initial rows of our tables.

```
head(apr20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 A847FA~ docked_bike   2020-04-26 17:45:14 2020-04-26 18:12:03 Eckhart Park
## 2 5405B8~ docked_bike   2020-04-17 17:08:54 2020-04-17 17:17:03 Drake Ave & Ful~
## 3 5DD24A~ docked_bike   2020-04-01 17:54:13 2020-04-01 18:08:36 McClurg Ct & Er~
## 4 2A59BB~ docked_bike   2020-04-07 12:50:19 2020-04-07 13:02:31 California Ave ~
## 5 27AD30~ docked_bike   2020-04-18 10:22:59 2020-04-18 11:15:54 Rush St & Hubba~
## 6 356216~ docked_bike   2020-04-30 17:55:47 2020-04-30 18:01:11 Mies van der Ro~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(may20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 02668A~ docked_bike   2020-05-27 10:03:52 2020-05-27 10:16:49 Franklin St & J~
## 2 7A50CC~ docked_bike   2020-05-25 10:47:11 2020-05-25 11:05:40 Clark St & Wrig~
## 3 2FFCDF~ docked_bike   2020-05-02 14:11:03 2020-05-02 15:48:21 Kedzie Ave & Mi~
## 4 58991C~ docked_bike   2020-05-02 16:25:36 2020-05-02 16:39:28 Clarendon Ave &~
## 5 A79651~ docked_bike   2020-05-29 12:49:54 2020-05-29 13:27:11 Hermitage Ave &~
## 6 1466C5~ docked_bike   2020-05-29 13:27:24 2020-05-29 14:14:45 Halsted St & Ar~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(jun20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 8CD5DE~ docked_bike   2020-06-13 23:24:48 2020-06-13 23:36:55 Wilton Ave & Be~
## 2 9A191E~ docked_bike   2020-06-26 07:26:10 2020-06-26 07:31:58 Federal St & Po~
## 3 F37D14~ docked_bike   2020-06-23 17:12:41 2020-06-23 17:21:14 Daley Center Pl~
## 4 C41237~ docked_bike   2020-06-20 01:09:35 2020-06-20 01:28:24 Broadway & Corn~
## 5 4B51B3~ docked_bike   2020-06-25 16:59:25 2020-06-25 17:08:48 Sheffield Ave &~
## 6 D50DF2~ docked_bike   2020-06-17 18:07:18 2020-06-17 18:18:14 Sheffield Ave &~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(jul20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 762198~ docked_bike   2020-07-09 15:22:02 2020-07-09 15:25:52 Ritchie Ct & Ba~
## 2 BEC9C9~ docked_bike   2020-07-24 23:56:30 2020-07-25 00:20:17 Halsted St & Ro~
## 3 D2FD8E~ docked_bike   2020-07-08 19:49:07 2020-07-08 19:56:22 Lake Shore Dr &~
## 4 54AE59~ docked_bike   2020-07-17 19:06:42 2020-07-17 19:27:38 LaSalle St & Il~
## 5 54025F~ docked_bike   2020-07-04 10:39:57 2020-07-04 10:45:05 Lake Shore Dr &~
## 6 65636B~ docked_bike   2020-07-28 16:33:03 2020-07-28 16:49:10 Fairbanks St & ~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(aug20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 322BD2~ docked_bike   2020-08-20 18:08:14 2020-08-20 18:17:51 Lake Shore Dr &~
## 2 2A3AEF~ electric_bike 2020-08-27 18:46:04 2020-08-27 19:54:51 Michigan Ave & ~
## 3 67DC1D~ electric_bike 2020-08-26 19:44:14 2020-08-26 21:53:07 Columbus Dr & R~
## 4 C79FBB~ electric_bike 2020-08-27 12:05:41 2020-08-27 12:53:45 Daley Center Pl~
## 5 13814D~ electric_bike 2020-08-27 16:49:02 2020-08-27 16:59:49 Leavitt St & Di~
## 6 56349A~ electric_bike 2020-08-27 17:26:23 2020-08-27 18:07:50 Leavitt St & Di~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(sept20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 2B22BD~ electric_bike 2020-09-17 14:27:11 2020-09-17 14:44:24 Michigan Ave & ~
## 2 A7FB70~ electric_bike 2020-09-17 15:07:31 2020-09-17 15:07:45 W Oakdale Ave &~
## 3 86057F~ electric_bike 2020-09-17 15:09:04 2020-09-17 15:09:35 W Oakdale Ave &~
## 4 57F6DC~ electric_bike 2020-09-17 18:10:46 2020-09-17 18:35:49 Ashland Ave & B~
## 5 B9C471~ electric_bike 2020-09-17 15:16:13 2020-09-17 15:52:55 Fairbanks Ct & ~
## 6 378BBC~ electric_bike 2020-09-17 18:37:04 2020-09-17 19:23:28 Clark St & Armi~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(oct20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 ACB6B4~ electric_bike 2020-10-31 19:39:43 2020-10-31 19:57:12 Lakeview Ave & ~
## 2 DF450C~ electric_bike 2020-10-31 23:50:08 2020-11-01 00:04:16 Southport Ave &~
## 3 B6396B~ electric_bike 2020-10-31 23:00:01 2020-10-31 23:08:22 Stony Island Av~
## 4 44A4AE~ electric_bike 2020-10-31 22:16:43 2020-10-31 22:19:35 Clark St & Grac~
## 5 10B7DD~ electric_bike 2020-10-31 19:38:19 2020-10-31 19:54:32 Southport Ave &~
## 6 DA6C37~ electric_bike 2020-10-29 17:38:04 2020-10-29 17:45:43 Larrabee St & D~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(nov20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 BD0A6F~ electric_bike 2020-11-01 13:36:00 2020-11-01 13:45:40 Dearborn St & E~
## 2 96A7A7~ electric_bike 2020-11-01 10:03:26 2020-11-01 10:14:45 Franklin St & I~
## 3 C61526~ electric_bike 2020-11-01 00:34:05 2020-11-01 01:03:06 Lake Shore Dr &~
## 4 E533E8~ electric_bike 2020-11-01 00:45:16 2020-11-01 00:54:31 Leavitt St & Ch~
## 5 1C9F4E~ electric_bike 2020-11-01 15:43:25 2020-11-01 16:16:52 Buckingham Foun~
## 6 725958~ electric_bike 2020-11-14 15:55:17 2020-11-14 16:44:38 Wabash Ave & 16~
## # ... with 8 more variables: start_station_id <dbl>, end_station_name <chr>,
## #   end_station_id <dbl>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(dec20)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 70B6A9~ classic_bike  2020-12-27 12:44:29 2020-12-27 12:55:06 Aberdeen St & J~
## 2 158A46~ electric_bike 2020-12-18 17:37:15 2020-12-18 17:44:19 <NA>
## 3 526201~ electric_bike 2020-12-15 15:04:33 2020-12-15 15:11:28 <NA>
## 4 BE1196~ electric_bike 2020-12-15 15:54:18 2020-12-15 16:00:11 <NA>
## 5 69AF78~ electric_bike 2020-12-22 12:08:17 2020-12-22 12:10:59 <NA>
## 6 C1DECC~ electric_bike 2020-12-22 13:26:37 2020-12-22 13:34:50 <NA>
## # ... with 8 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(jan21)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 E19E6F~ electric_bike 2021-01-23 16:14:19 2021-01-23 16:24:44 California Ave ~
## 2 DC88F2~ electric_bike 2021-01-27 18:43:08 2021-01-27 18:47:12 California Ave ~
## 3 EC45C9~ electric_bike 2021-01-21 22:35:54 2021-01-21 22:37:14 California Ave ~
## 4 4FA453~ electric_bike 2021-01-07 13:31:13 2021-01-07 13:42:55 California Ave ~
## 5 BE5E8E~ electric_bike 2021-01-23 02:24:02 2021-01-23 02:24:45 California Ave ~
## 6 5D8969~ electric_bike 2021-01-09 14:24:07 2021-01-09 15:17:54 California Ave ~
## # ... with 8 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(feb21)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 89E7AA~ classic_bike  2021-02-12 16:14:56 2021-02-12 16:21:43 Glenwood Ave & ~
## 2 0FEFDE~ classic_bike  2021-02-14 17:52:38 2021-02-14 18:12:09 Glenwood Ave & ~
## 3 E6159D~ electric_bike 2021-02-09 19:10:18 2021-02-09 19:19:10 Clark St & Lake~
## 4 B32D31~ classic_bike  2021-02-02 17:49:41 2021-02-02 17:54:06 Wood St & Chica~
## 5 83E463~ electric_bike 2021-02-23 15:07:23 2021-02-23 15:22:37 State St & 33rd~
## 6 BDAA7E~ electric_bike 2021-02-24 15:43:33 2021-02-24 15:49:05 Fairbanks St & ~
## # ... with 8 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

```
head(mar21)
```

```
## # A tibble: 6 x 13
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 CFA86D~ classic_bike  2021-03-16 08:32:30 2021-03-16 08:36:34 Humboldt Blvd &~
## 2 30D9DC~ classic_bike  2021-03-28 01:26:28 2021-03-28 01:36:55 Humboldt Blvd &~
## 3 846D87~ classic_bike  2021-03-11 21:17:29 2021-03-11 21:33:53 Shields Ave & 2~
## 4 994D05~ classic_bike  2021-03-11 13:26:42 2021-03-11 13:55:41 Winthrop Ave & ~
## 5 DF7464~ classic_bike  2021-03-21 09:09:37 2021-03-21 09:27:33 Glenwood Ave & ~
## 6 CEBA85~ classic_bike  2021-03-20 11:08:47 2021-03-20 11:29:39 Glenwood Ave & ~
## # ... with 8 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
```

After the inspection of our data we have found that the column ‘start_station_id’ and ‘end_station_id’ in some tables have the wrong data types. The first 8 tables has this data type as ‘doubles’ while the later 4 has this as ‘character’. The reason behind this inconsistency might be 2 different people creating this data set. Here, we have to make the data type as ‘character’ for the tables of the following months :

April 2020
May 2020
June 2020
July 2020
August 2020
September 2020
October 2020
November 2020

Here, we will use the ‘mutate’ function to convert the “dbl” datatype to “chr” datatype for start_station_id and end_station_id.

```
apr20 <-  mutate(apr20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

may20 <-  mutate(may20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

jun20 <-  mutate(jun20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

jul20 <-  mutate(jul20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

aug20 <-  mutate(aug20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

sept20 <-  mutate(sept20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

oct20 <-  mutate(oct20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))

nov20 <-  mutate(nov20, start_station_id = as.character(start_station_id)
                        ,end_station_id = as.character(end_station_id))
```

After updating the datatypes in our tables, now we will combine the tables/data sets. This new data set will be called as yearly_trips.

```
yearly_trips <- bind_rows(apr20, may20, jun20, jul20, aug20, sept20,
                          oct20, nov20, dec20, jan21, feb21, mar21)
```

Now we have to remove the columns from ‘yearly_trips’ we do not require.

```
yearly_trips <- select(yearly_trips, -c(start_lat, start_lng, end_lat, end_lng))
```

Now we will add a new column ‘ride_len’ which will store the bike riding time for each person.

```
yearly_trips <- yearly_trips %>%
    mutate(ride_len = difftime(ended_at, started_at, units = "mins"))
```

Now we will print the head to check the columns and values in ‘yearly_trips’ data set.

```
head(yearly_trips)
```

```
## # A tibble: 6 x 10
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 A847FA~ docked_bike   2020-04-26 17:45:14 2020-04-26 18:12:03 Eckhart Park
## 2 5405B8~ docked_bike   2020-04-17 17:08:54 2020-04-17 17:17:03 Drake Ave & Ful~
## 3 5DD24A~ docked_bike   2020-04-01 17:54:13 2020-04-01 18:08:36 McClurg Ct & Er~
## 4 2A59BB~ docked_bike   2020-04-07 12:50:19 2020-04-07 13:02:31 California Ave ~
## 5 27AD30~ docked_bike   2020-04-18 10:22:59 2020-04-18 11:15:54 Rush St & Hubba~
## 6 356216~ docked_bike   2020-04-30 17:55:47 2020-04-30 18:01:11 Mies van der Ro~
## # ... with 5 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, member_casual <chr>, ride_len <drtn>
```

Changing the data type of ride_len to numeric

```
yearly_trips$ride_len <- as.numeric(as.character(yearly_trips$ride_len))
is.numeric(yearly_trips$ride_len)
```

```
## [1] TRUE
```

Adding a new columns as weekday, date, month, day and year which tells which day was it of the week,

1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6= Saturday, 7=Sunday

```
yearly_trips <- yearly_trips %>%
    mutate(weekday = as.numeric(format(yearly_trips$started_at, format = "%u")))
```

```
yearly_trips$date <- as.Date(yearly_trips$started_at)
yearly_trips$month <- as.yearmon(yearly_trips$date, "%Y-%m")
yearly_trips$day <- format(yearly_trips$date, "%d")
yearly_trips$year <- format(yearly_trips$date, "%Y")
head(yearly_trips)
```

```
## # A tibble: 6 x 15
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 A847FA~ docked_bike   2020-04-26 17:45:14 2020-04-26 18:12:03 Eckhart Park
## 2 5405B8~ docked_bike   2020-04-17 17:08:54 2020-04-17 17:17:03 Drake Ave & Ful~
## 3 5DD24A~ docked_bike   2020-04-01 17:54:13 2020-04-01 18:08:36 McClurg Ct & Er~
## 4 2A59BB~ docked_bike   2020-04-07 12:50:19 2020-04-07 13:02:31 California Ave ~
## 5 27AD30~ docked_bike   2020-04-18 10:22:59 2020-04-18 11:15:54 Rush St & Hubba~
## 6 356216~ docked_bike   2020-04-30 17:55:47 2020-04-30 18:01:11 Mies van der Ro~
## # ... with 10 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, member_casual <chr>, ride_len <dbl>, weekday <dbl>,
## #   date <date>, month <yearmon>, day <chr>, year <chr>
```

Checking if there are any rows that have ride length less than 0 and are used for test rides by Cyclistic

```
sum(yearly_trips$ride_len<0)
```

```
## [1] 10552
```

```
nrow(subset(yearly_trips, start_station_name %like% "TEST"))
```

```
## [1] 3364
```

```
nrow(subset(yearly_trips, start_station_name %like% "Test"))
```

```
## [1] 0
```

```
nrow(subset(yearly_trips, start_station_name %like% "test"))
```

```
## [1] 3
```

Created a new Dataframe which deletes all the rows that has negetive ride length or is used for test drive by Cyclistic

```
trips_updated <- yearly_trips[!(yearly_trips$start_station_name == "Test"
                    | yearly_trips$ride_len<0
                    | yearly_trips$start_station_name == "TEST"
                    | yearly_trips$start_station_name == "test" ),]
```

Dropping/ Deleting the rows that has no values or empty cells

```
trips_updated <- drop_na(trips_updated)
```

```
glimpse(trips_updated)
```

```
## Rows: 3,284,237
## Columns: 15
## $ ride_id            <chr> "A847FADBBC638E45", "5405B80E996FF60D", "5DD24A79A4~
## $ rideable_type      <chr> "docked_bike", "docked_bike", "docked_bike", "docke~
## $ started_at         <dttm> 2020-04-26 17:45:14, 2020-04-17 17:08:54, 2020-04-~
## $ ended_at           <dttm> 2020-04-26 18:12:03, 2020-04-17 17:17:03, 2020-04-~
## $ start_station_name <chr> "Eckhart Park", "Drake Ave & Fullerton Ave", "McClu~
## $ start_station_id   <chr> "86", "503", "142", "216", "125", "173", "35", "434~
## $ end_station_name   <chr> "Lincoln Ave & Diversey Pkwy", "Kosciuszko Park", "~
## $ end_station_id     <chr> "152", "499", "255", "657", "323", "35", "635", "38~
## $ member_casual      <chr> "member", "member", "member", "member", "casual", "~
## $ ride_len           <dbl> 26.816667, 8.150000, 14.383333, 12.200000, 52.91666~
## $ weekday            <dbl> 7, 5, 3, 2, 6, 4, 4, 2, 3, 6, 6, 6, 5, 6, 1, 6, 7, ~
## $ date               <date> 2020-04-26, 2020-04-17, 2020-04-01, 2020-04-07, 20~
## $ month              <yearmon> Apr 2020, Apr 2020, Apr 2020, Apr 2020, Apr 202~
## $ day                <chr> "26", "17", "01", "07", "18", "30", "02", "07", "15~
## $ year               <chr> "2020", "2020", "2020", "2020", "2020", "2020", "20~
```

```
head(trips_updated)
```

```
## # A tibble: 6 x 15
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 A847FA~ docked_bike   2020-04-26 17:45:14 2020-04-26 18:12:03 Eckhart Park
## 2 5405B8~ docked_bike   2020-04-17 17:08:54 2020-04-17 17:17:03 Drake Ave & Ful~
## 3 5DD24A~ docked_bike   2020-04-01 17:54:13 2020-04-01 18:08:36 McClurg Ct & Er~
## 4 2A59BB~ docked_bike   2020-04-07 12:50:19 2020-04-07 13:02:31 California Ave ~
## 5 27AD30~ docked_bike   2020-04-18 10:22:59 2020-04-18 11:15:54 Rush St & Hubba~
## 6 356216~ docked_bike   2020-04-30 17:55:47 2020-04-30 18:01:11 Mies van der Ro~
## # ... with 10 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, member_casual <chr>, ride_len <dbl>, weekday <dbl>,
## #   date <date>, month <yearmon>, day <chr>, year <chr>
```

Total number of members and casual riders, along with their distance covered

```
table(trips_updated$member_casual)
```

```
##
##  casual  member
## 1348018 1936219
```

```
aggregate(ride_len ~ member_casual, trips_updated, sum)
```

```
##   member_casual ride_len
## 1        casual 61270741
## 2        member 30850069
```

Phase 4 : ANALYZE

Finding the Mean, Median, Max, Min of the Ride Length. We will begin by finding the Mode of the Weekdays, meaning the days where most Cyclistic bikes were used.

```
getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}

v <- trips_updated$weekday

result <- getmode(v)
print("Mode")
```

```
## [1] "Mode"
```

```
result
```

```
## [1] 6
```

```
print("Saturday")
```

```
## [1] "Saturday"
```

```
trips_updated%>%
 group_by(member_casual)%>%
 summarize(Min= min(ride_len),
           Mean=mean(ride_len),
           Max=max(ride_len),
           Mode= result)
```

```
## # A tibble: 2 x 5
##   member_casual   Min  Mean    Max  Mode
##   <chr>         <dbl> <dbl>  <dbl> <dbl>
## 1 casual            0  45.5 55684.     6
## 2 member            0  15.9 58720.     6
```

Assigning Names of the weekdays to their respective days

```
trips_updated$weekday <- recode(trips_updated$weekday,
       "1"="Monday",
       "2"="Tuesday",
       "3"="Wednesday",
       "4"="Thursday",
       "5"="Friday",
       "6"="Saturday",
       "7"="Sunday")
```

```
str(trips_updated$weekday)
```

```
##  chr [1:3284237] "Sunday" "Friday" "Wednesday" "Tuesday" "Saturday" ...
```

```
head(trips_updated)
```

```
## # A tibble: 6 x 15
##   ride_id rideable_type started_at          ended_at            start_station_n~
##   <chr>   <chr>         <dttm>              <dttm>              <chr>
## 1 A847FA~ docked_bike   2020-04-26 17:45:14 2020-04-26 18:12:03 Eckhart Park
## 2 5405B8~ docked_bike   2020-04-17 17:08:54 2020-04-17 17:17:03 Drake Ave & Ful~
## 3 5DD24A~ docked_bike   2020-04-01 17:54:13 2020-04-01 18:08:36 McClurg Ct & Er~
## 4 2A59BB~ docked_bike   2020-04-07 12:50:19 2020-04-07 13:02:31 California Ave ~
## 5 27AD30~ docked_bike   2020-04-18 10:22:59 2020-04-18 11:15:54 Rush St & Hubba~
## 6 356216~ docked_bike   2020-04-30 17:55:47 2020-04-30 18:01:11 Mies van der Ro~
## # ... with 10 more variables: start_station_id <chr>, end_station_name <chr>,
## #   end_station_id <chr>, member_casual <chr>, ride_len <dbl>, weekday <chr>,
## #   date <date>, month <yearmon>, day <chr>, year <chr>
```

Checking how many rides were taken by different members on the basis of Days

```
trips_updated %>%
 group_by(member_casual, weekday) %>%
 summarise(number_of_rides = n(),avg_ride_len = mean(ride_len)) %>%
 arrange(member_casual, desc(number_of_rides))
```

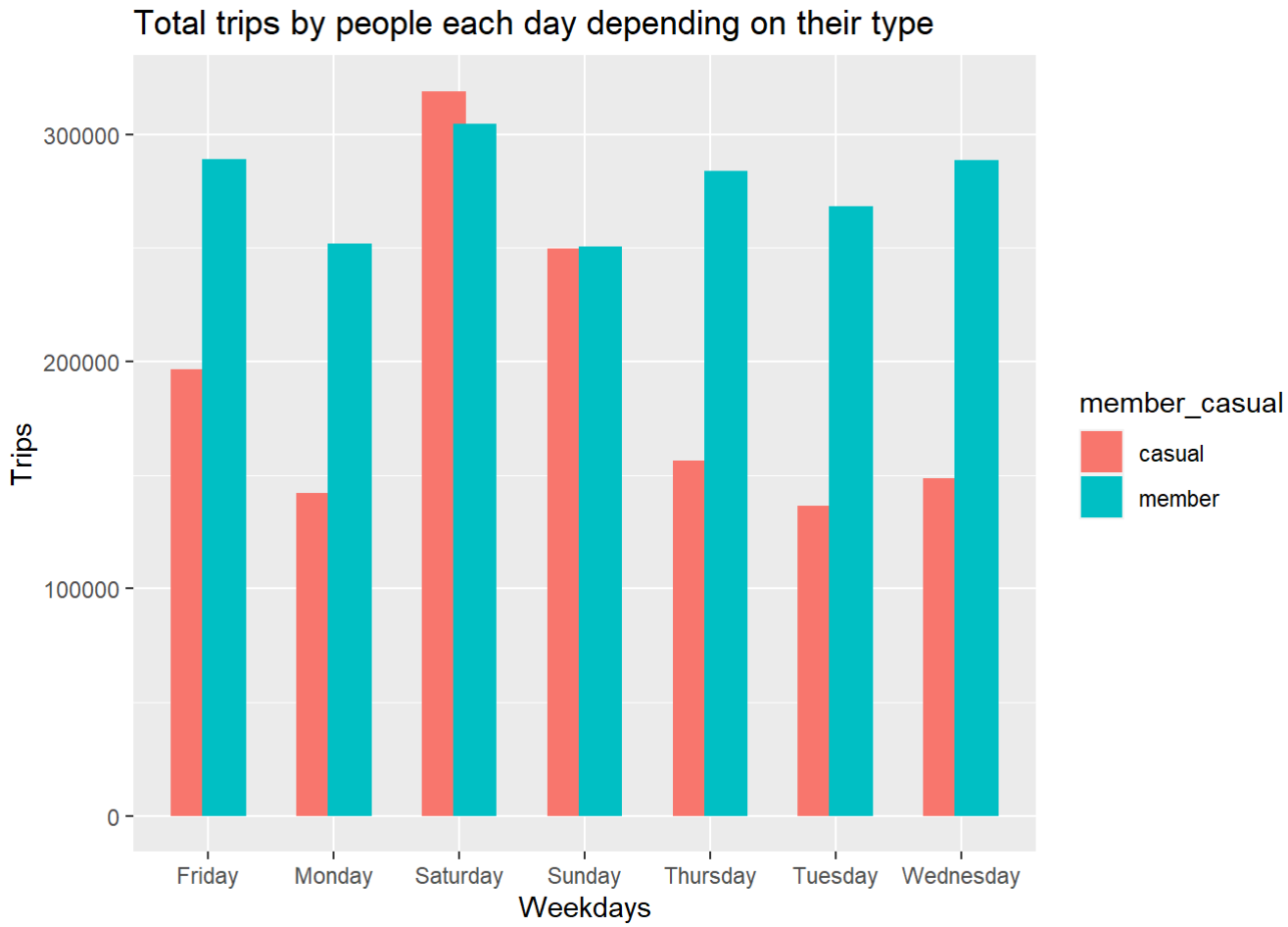
```
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups`
argument.
```

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday   number_of_rides avg_ride_len
##    <chr>         <chr>               <int>        <dbl>
##  1 casual        Saturday           319017         47.3
##  2 casual        Sunday             249761         51.1
##  3 casual        Friday             196451         43.3
##  4 casual        Thursday           156177         43.6
##  5 casual        Wednesday          148352         40.9
##  6 casual        Monday             142042         45.5
##  7 casual        Tuesday            136218         40.9
##  8 member        Saturday           304608         17.7
##  9 member        Friday             288897         15.6
## 10 member        Wednesday          288390         15.0
## 11 member        Thursday           283731         15.0
## 12 member        Tuesday            268223         15.0
## 13 member        Monday             251905         15.1
## 14 member        Sunday             250465         18.1
```

Visualizing the different riders on each day and the ride length

```
library(ggplot2)
trips_updated %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  labs(title ="Total trips by people each day depending on their type",
       x="Weekdays", y="Trips") +
  geom_col(width=0.7, position = position_dodge(width=0.5)) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups`
argument.
```



Visualizing the different Riders each day with the time they spent

```
trips_updated %>%
  #mutate(weekday = weekday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,avg_ride_len = mean(ride_len)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(y= avg_ride_len, x = weekday, fill = member_casual)) +
  geom_col(position = "dodge", alpha=0.9, size=0.5) +
 #scale_fill_viridis(discrete = T) +
#theme_ipsum() +
  ggtitle("Average ride length by the different people each day") +
    labs(y="Ride length (mins)", x="Weekdays ")
```

```
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups`
argument.
```

Checking how many rides were taken by different members on the basis of Months

```
trips_updated %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n(),`avg_ride` = mean(ride_len)) %>%
  arrange(member_casual,desc(number_of_rides))
```

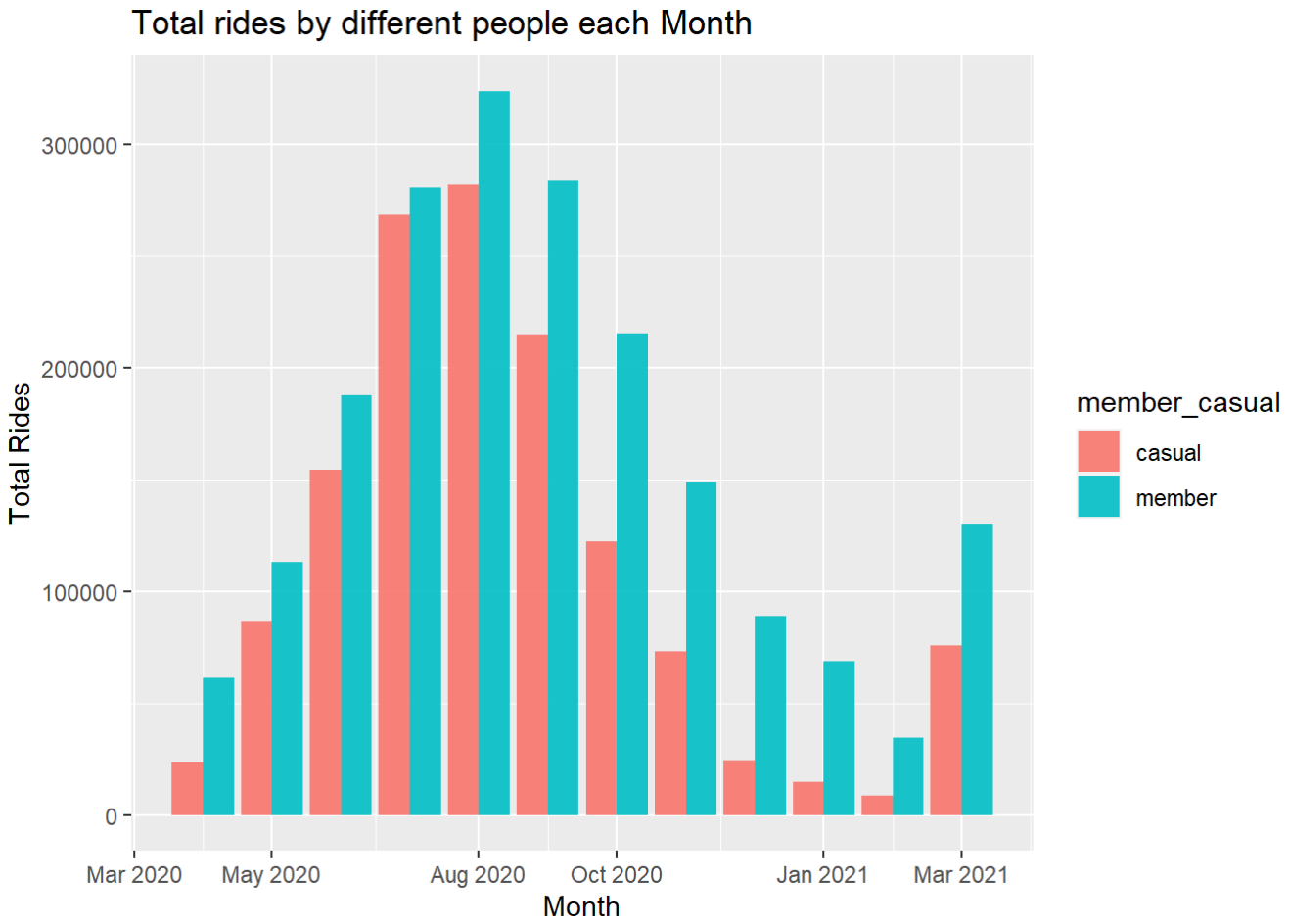
```
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups`
argument.
```

```
## # A tibble: 24 x 4
## # Groups:   member_casual [2]
##    member_casual month     number_of_rides avg_ride
##    <chr>         <yearmon>           <int>    <dbl>
##  1 casual        Aug 2020           281987     44.2
##  2 casual        Jul 2020           268125     59.3
##  3 casual        Sep 2020           214681     38.4
##  4 casual        Jun 2020           154342     51.2
##  5 casual        Oct 2020           122328     31.3
##  6 casual        May 2020            86699     50.6
##  7 casual        Mar 2021            75641     38.5
##  8 casual        Nov 2020            72850     33.5
##  9 casual        Dec 2020            24492     27.6
## 10 casual        Apr 2020            23570     72.5
## # ... with 14 more rows
```

Visualizing the Riders each month and the numbers of rides taken

```
trips_updated %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n()) %>%
  arrange(member_casual, month)  %>%
  ggplot(mapping=aes(x = month, y = number_of_rides, fill = member_casual)) +
    #facet_wrap(~member_casual)+
  labs(title ="Total rides by different people each Month",
                  x= "Month", y="Total Rides") +
  geom_col(position = "dodge", alpha=0.9, size=0.5 )+
  #theme(axis.text=element_text(angle=90))
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```

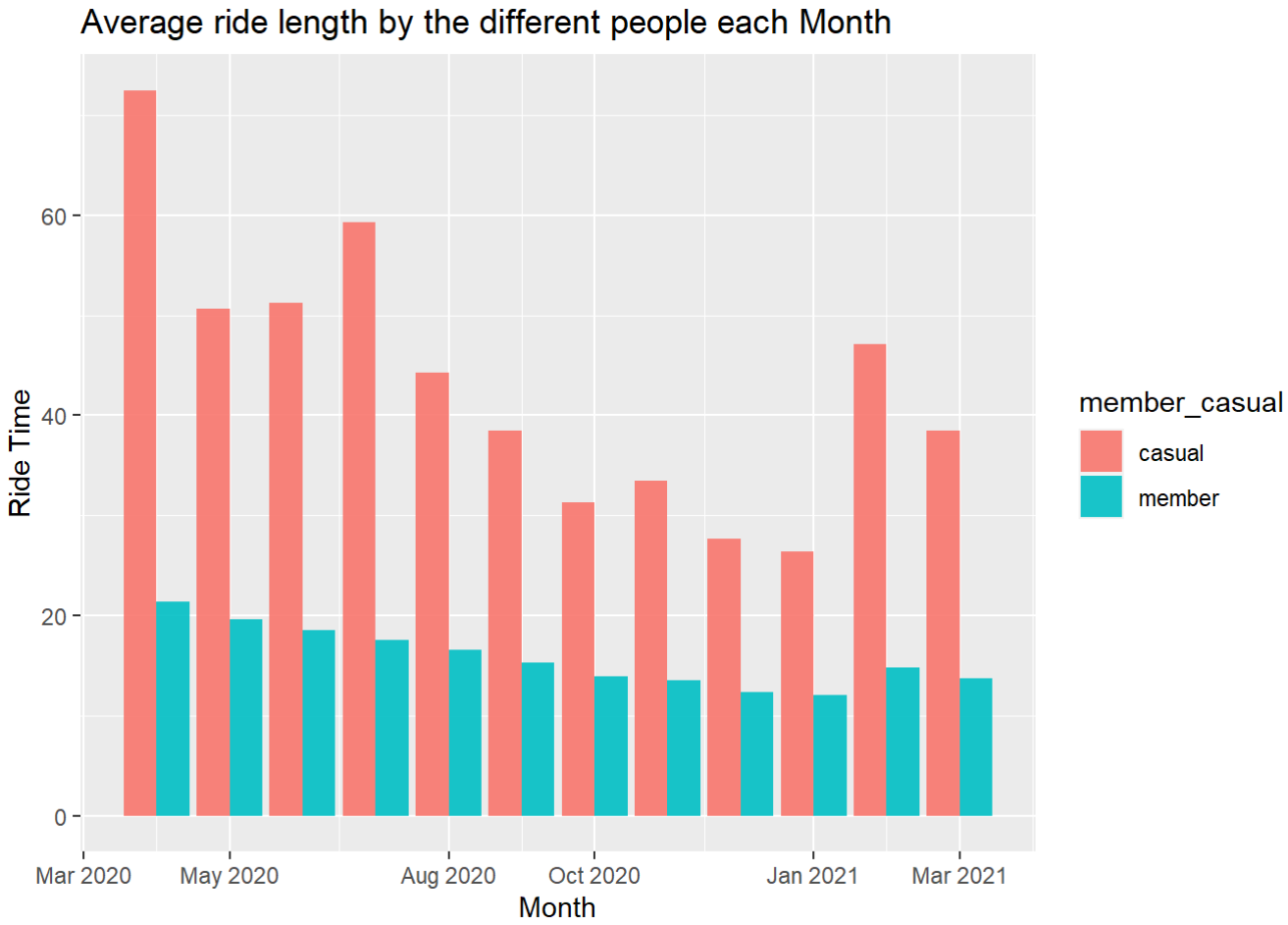
```
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups`
argument.
```



Visualizing the Riders and the Ride Length

```
trips_updated %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n(),avg_ride_len = mean(ride_len)) %>%
  arrange(member_casual, month)  %>%
  ggplot(aes(x = month, y = avg_ride_len, fill = member_casual)) +
  geom_col(position = "dodge", alpha=0.9, size=0.5) +
 #scale_fill_viridis(discrete = T) +
#theme_ipsum() +
  ggtitle("Average ride length by the different people each Month") +
    labs(x="Month", y="Ride Time")
```

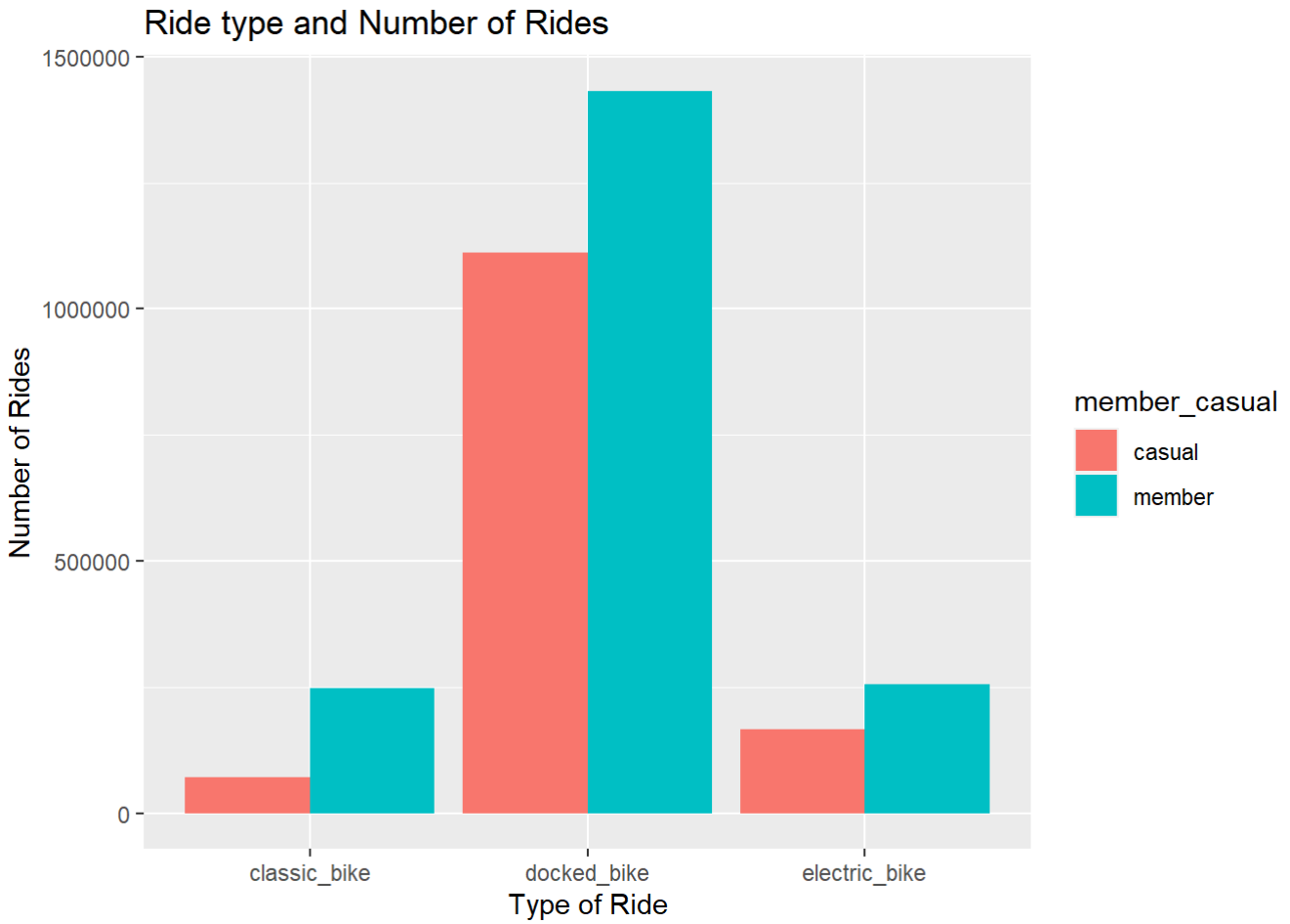
```
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups`
argument.
```



Visualizing the Different types of bikes that were used by Different Riders

```
trips_updated %>%
  group_by(rideable_type, member_casual) %>%
  summarise(number_of_rides = n()) %>%
  ggplot(aes(x= rideable_type, y=number_of_rides, fill= member_casual))+
  geom_bar(stat='identity', position = "dodge") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  labs(title ="Ride type and Number of Rides",
      x="Type of Ride", y="Number of Rides" )
```

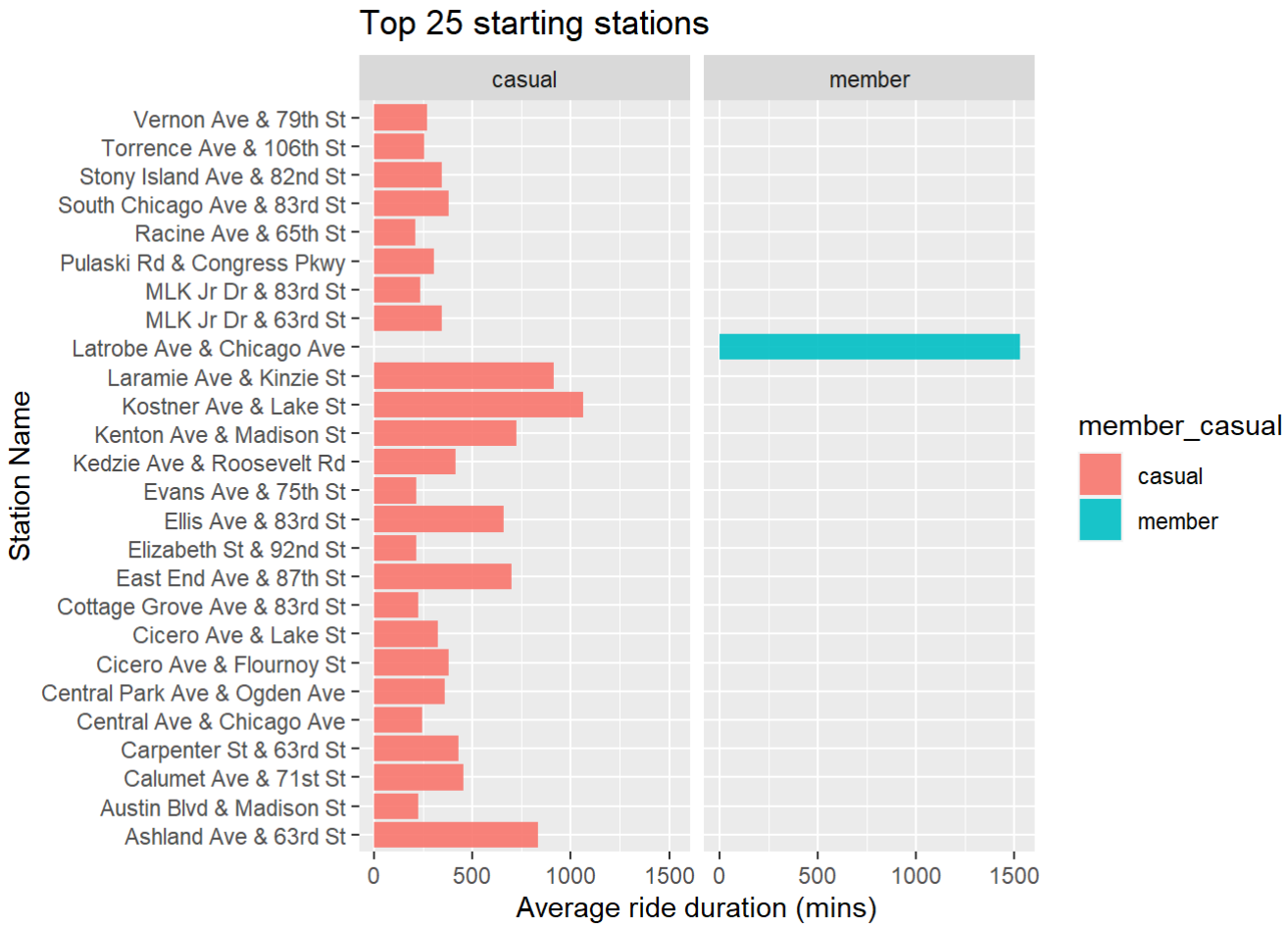
```
## `summarise()` has grouped output by 'rideable_type'. You can override using the `.groups`
argument.
```



Visualizing the Most used stations to pick up Bikes by different Rider

```
trips_updated %>%
    group_by(start_station_name, member_casual) %>%
    summarise(avg_ride = mean(ride_len)) %>%
    arrange(desc(avg_ride)) %>%
    head(26) %>%
    ggplot(aes(x = avg_ride, y = start_station_name, fill=member_casual)) +
    geom_col(position = "dodge", alpha=0.9 , size=0.5) +
    #scale_fill_viridis(discrete = T) +
    # theme_ipsum() +
    ggtitle("Top 25 starting stations") +
    labs(x="Average ride duration (mins)", y="Station Name")+
    facet_wrap(~member_casual)
```

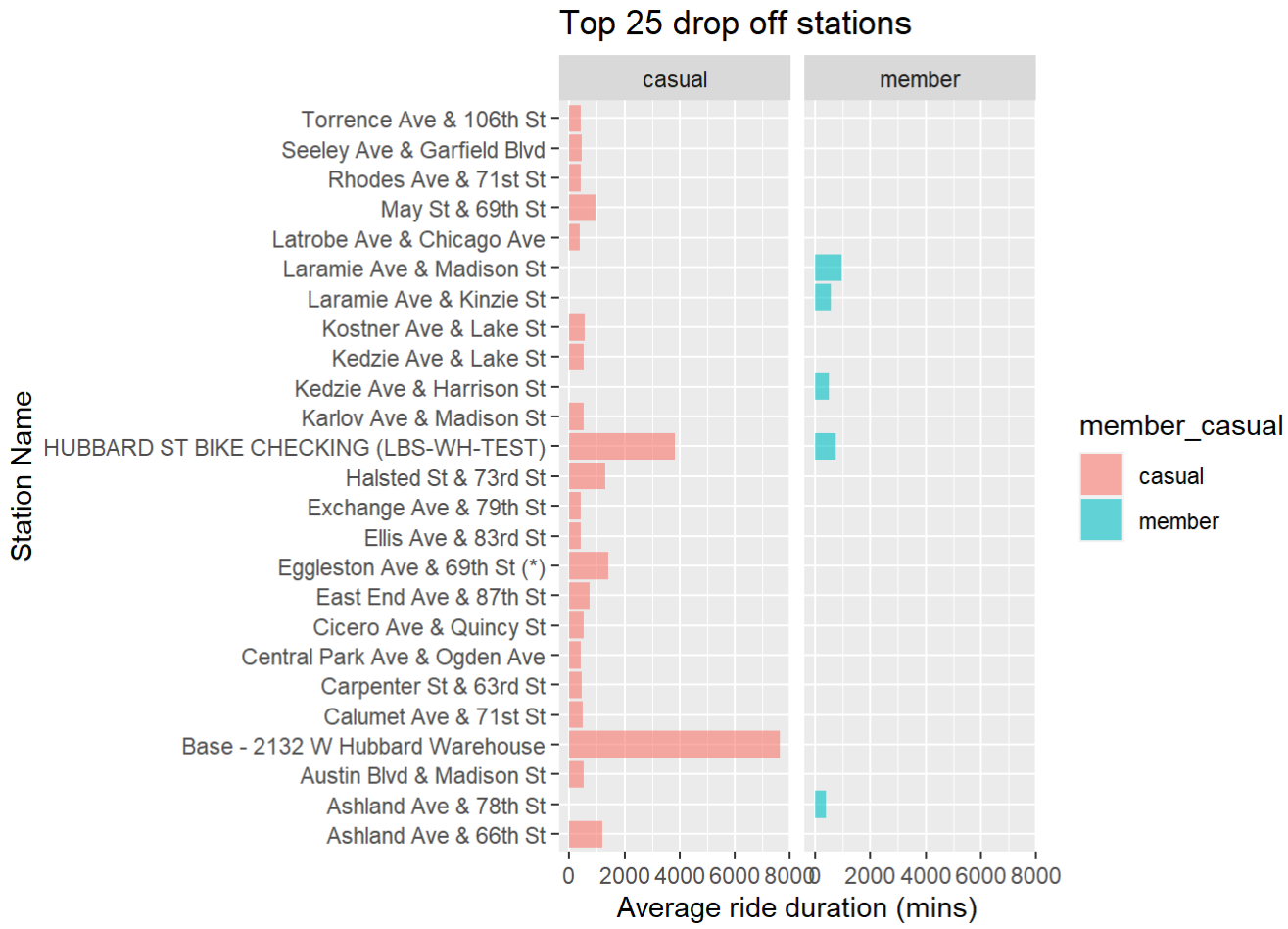
```
## `summarise()` has grouped output by 'start_station_name'. You can override using the `.gro
ups` argument.
```



Visualizing the most drop off stations by Different Riders

```
trips_updated %>%
    group_by(end_station_name, member_casual) %>%
    summarise(avg_ride = mean(ride_len)) %>%
    arrange(desc(avg_ride)) %>%
    head(26) %>%
    ggplot(aes(x = avg_ride, y = end_station_name, fill=member_casual)) +
    geom_col(position = "dodge", alpha=0.6 , size=0.5) +
    #scale_fill_viridis(discrete = T) +
    # theme_ipsum() +
    ggtitle("Top 25 drop off stations") +
    labs(x="Average ride duration (mins)", y="Station Name")+
    facet_grid(~member_casual)
```

```
## `summarise()` has grouped output by 'end_station_name'. You can override using the `.group
s` argument.
```



Phase 5 : SHARE

Conclusions

For Casual Riders

1. We can see by the visuals that we have more Casual riders than the Members.
2. Casual Riders prefer Docked Bikes.
3. Casual Riders use Cyclistic bikes more during the weekends.
4. Casual Riders also use more during the months of July, August and September.
5. Average time spent riding is 45 mins.

For Members

1. We have same amount of Rides on the weekdays, but on weekends, the riders have increased.
2. They too mostly prefer Docked Bikes.
3. The time they spent Riding is average of 15 mins.
4. Saturdays had the most number of Riders.

Phase 6 : ACT

Suggestions to the Cyclistic

1. Give offers or discounts for weekends to encourage the Casual riders to buy the Subcription.
2. Increase the number of Docked Bikes, since it is widely used.**
3. Built stations near Public places,like, Parks, Malls, etc to attract Casual Riders.
4. Gift services to the members so that they can encourage their friends to buy Subcriptions.

Thank You