R Notebook

Installing Libraries

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.1.0
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(lubridate)
```

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

```r
library(ggplot2)
library(readxl)
```

Loading in Data for Each Month

```r
jan_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

feb_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

mar_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

apr_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

may_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

jun_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa
jul_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

aug_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

sep_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

oct_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

nov_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa

dec_data <- read_excel("/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclisticDa
```

Check Structure of Each Month's Data to Ensure Consistencies

```r
print("JAN")
```

```
## [1] "JAN"
```

```r
str(jan_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80
##  $ rideable_type     : chr [1:16383] "electric_bike" "electric_bike" "classic_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:16383], format: "2022-01-13 11:59:47" "2022-01-10 08:41:56" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-01-13 12:02:44" "2022-01-10 08:46:17" ...
##  $ start_station_name: chr [1:16383] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffiel
##  $ start_station_id  : chr [1:16383] "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr [1:16383] "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave &
##  $ end_station_id    : chr [1:16383] "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...
##  $ start_lat         : num [1:16383] 42 42 41.9 42 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:16383] 42 42 41.9 42 41.9 ...
##  $ end_lng           : num [1:16383] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:16383] "casual" "casual" "member" "casual" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:02:57" "1899-12-31 00:04:21" ...
##  $ day_of_week       : num [1:16383] 5 2 3 3 5 3 1 7 2 6 ...
```

```r
print("FEB")
```

```
## [1] "FEB"
```

```r
str(feb_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "E1E065E7ED285C02" "1602DCDC5B30FFE3" "BE7DD2AF4B55C4AF" "A1789
##  $ rideable_type     : chr [1:16383] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : POSIXct[1:16383], format: "2022-02-19 18:08:41" "2022-02-20 17:41:30" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-02-19 18:23:56" "2022-02-20 17:45:56" ...
##  $ start_station_name: chr [1:16383] "State St & Randolph St" "Halsted St & Wrightwood Ave" "State S
##  $ start_station_id  : chr [1:16383] "TA1305000029" "TA1309000061" "TA1305000029" "13235" ...
##  $ end_station_name  : chr [1:16383] "Clark St & Lincoln Ave" "Southport Ave & Wrightwood Ave" "Canal
##  $ end_station_id    : chr [1:16383] "13179" "TA1307000113" "13011" "13323" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:16383] "member" "member" "member" "member" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:15:15" "1899-12-31 00:04:26" ...
##  $ day_of_week       : num [1:16383] 7 1 6 2 4 2 2 3 6 1 ...
```

```
print("MAR")
```

```
## [1] "MAR"
```

```
str(mar_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "47EC0A7F82E65D52" "8494861979B0F477" "EFE527AF80B66109" "9F446
##  $ rideable_type     : chr [1:16383] "classic_bike" "electric_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:16383], format: "2022-03-21 13:45:01" "2022-03-16 09:37:16" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-03-21 13:51:18" "2022-03-16 09:43:34" ...
##  $ start_station_name: chr [1:16383] "Wabash Ave & Wacker Pl" "Michigan Ave & Oak St" "Broadway & Be
##  $ start_station_id  : chr [1:16383] "TA1307000131" "13042" "13109" "TA1307000131" ...
##  $ end_station_name  : chr [1:16383] "Kingsbury St & Kinzie St" "Orleans St & Chestnut St (NEXT Apts
##  $ end_station_id    : chr [1:16383] "KA1503000043" "620" "15578" "TA1305000025" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.6 -87.6 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.6 -87.7 -87.6 -87.7 ...
##  $ member_casual     : chr [1:16383] "member" "member" "member" "member" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:06:17" "1899-12-31 00:06:18" ...
##  $ day_of_week       : num [1:16383] 2 4 4 3 2 2 5 7 5 6 ...
```

```
print("APR")
```

```
## [1] "APR"
```

```
str(apr_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "3564070EEFD12711" "0B820C7FCF22F489" "89EEEE32293F07FF" "84D47
##  $ rideable_type     : chr [1:16383] "electric_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:16383], format: "2022-04-06 17:42:48" "2022-04-24 19:23:07" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-04-06 17:54:36" "2022-04-24 19:43:17" ...
##  $ start_station_name: chr [1:16383] "Paulina St & Howard St" "Wentworth Ave & Cermak Rd" "Halsted S
##  $ start_station_id  : chr [1:16383] "515" "13075" "TA1307000121" "13075" ...
##  $ end_station_name  : chr [1:16383] "University Library (NU)" "Green St & Madison St" "Green St & Ma
##  $ end_station_id    : chr [1:16383] "605" "TA1307000120" "TA1307000120" "KA1706005007" ...
##  $ start_lat         : num [1:16383] 42 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:16383] 42.1 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:16383] "member" "member" "member" "casual" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:11:48" "1899-12-31 00:20:10" ...
##  $ day_of_week       : num [1:16383] 4 1 4 6 7 5 2 3 6 6 ...
```

```
print("MAY")
```

```
## [1] "MAY"
```

```
str(may_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "EC2DE40644C6B0F4" "1C31AD03897EE385" "1542FBEC830415CF" "6FF59
##  $ rideable_type     : chr [1:16383] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : POSIXct[1:16383], format: "2022-05-23 23:06:58" "2022-05-11 08:53:28" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-05-23 23:40:19" "2022-05-11 09:31:22" ...
##  $ start_station_name: chr [1:16383] "Wabash Ave & Grand Ave" "DuSable Lake Shore Dr & Monroe St" "Cl
##  $ start_station_id  : chr [1:16383] "TA1307000117" "13300" "TA1305000032" "TA1305000032" ...
##  $ end_station_name  : chr [1:16383] "Halsted St & Roscoe St" "Field Blvd & South Water St" "Wood St
##  $ end_station_id    : chr [1:16383] "TA1309000025" "15534" "13221" "TA1305000030" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.6 -87.7 -87.6 -87.7 ...
##  $ member_casual     : chr [1:16383] "member" "member" "member" "member" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:33:21" "1899-12-31 00:37:54" ...
##  $ day_of_week       : num [1:16383] 2 4 5 3 3 4 6 1 2 4 ...
```

```
print("JUN")
```

```
## [1] "JUN"
```

```
str(jun_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "600CFD130D0FD2A4" "F5E6B5C1682C6464" "B6EB6D27BAD771D2" "C9C32
##  $ rideable_type     : chr [1:16383] "electric_bike" "electric_bike" "electric_bike" "electric_bike"
##  $ started_at        : POSIXct[1:16383], format: "2022-06-30 17:27:53" "2022-06-30 18:39:52" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-06-30 17:35:15" "2022-06-30 18:47:28" ...
##  $ start_station_name: chr [1:16383] NA NA NA NA ...
##  $ start_station_id  : chr [1:16383] NA NA NA NA ...
##  $ end_station_name  : chr [1:16383] NA NA NA NA ...
##  $ end_station_id    : chr [1:16383] NA NA NA NA ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.8 41.9 ...
##  $ start_lng         : num [1:16383] -87.6 -87.6 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 41.8 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:16383] "casual" "casual" "casual" "casual" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:07:22" "1899-12-31 00:07:36" ...
##  $ day_of_week       : num [1:16383] 5 5 5 5 4 5 5 5 5 5 ...
```

```r
print("JUL")
```

```
## [1] "JUL"
```

```r
str(jul_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "954144C2F67B1932" "292E027607D218B6" "57765852588AD6E0" "B5B6B
##  $ rideable_type     : chr [1:16383] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : POSIXct[1:16383], format: "2022-07-05 08:12:47" "2022-07-26 12:53:38" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-07-05 08:24:32" "2022-07-26 12:55:31" ...
##  $ start_station_name: chr [1:16383] "Ashland Ave & Blackhawk St" "Buckingham Fountain (Temp)" "Buck
##  $ start_station_id  : chr [1:16383] "13224" "15541" "15541" "15541" ...
##  $ end_station_name  : chr [1:16383] "Kingsbury St & Kinzie St" "Michigan Ave & 8th St" "Michigan Ave
##  $ end_station_id    : chr [1:16383] "KA1503000043" "623" "623" "TA1307000164" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 41.8 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:16383] "member" "casual" "casual" "casual" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:11:45" "1899-12-31 00:01:53" ...
##  $ day_of_week       : num [1:16383] 3 3 1 1 4 6 2 5 1 1 ...
```

```r
print("AUG")
```

```
## [1] "AUG"
```

```r
str(aug_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "550CF7EFEAE0C618" "DAD198F405F9C5F5" "E6F2BC47B65CB7FD" "F5978
##  $ rideable_type     : chr [1:16383] "electric_bike" "electric_bike" "electric_bike" "electric_bike"
##  $ started_at        : POSIXct[1:16383], format: "2022-08-07 21:34:15" "2022-08-08 14:39:21" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-08-07 21:41:46" "2022-08-08 14:53:23" ...
##  $ start_station_name: chr [1:16383] NA NA NA NA ...
##  $ start_station_id  : chr [1:16383] NA NA NA NA ...
##  $ end_station_name  : chr [1:16383] NA NA NA NA ...
##  $ end_station_id    : chr [1:16383] NA NA NA NA ...
##  $ start_lat         : num [1:16383] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 42 42 41.8 ...
##  $ end_lng           : num [1:16383] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:16383] "casual" "casual" "casual" "casual" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:07:31" "1899-12-31 00:14:02" ...
##  $ day_of_week       : num [1:16383] 1 2 2 2 1 2 2 1 1 1 ...
```

```r
print("SEP")
```

```
## [1] "SEP"
```

```r
str(sep_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "5156990AC19CA285" "E12D4A16BF51C274" "A02B53CD7DB72DD7" "C82E05
##  $ rideable_type     : chr [1:16383] "electric_bike" "electric_bike" "electric_bike" "electric_bike"
##  $ started_at        : POSIXct[1:16383], format: "2022-09-01 08:36:22" "2022-09-01 17:11:29" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-09-01 08:39:05" "2022-09-01 17:14:45" ...
##  $ start_station_name: chr [1:16383] NA NA NA NA ...
##  $ start_station_id  : chr [1:16383] NA NA NA NA ...
##  $ end_station_name  : chr [1:16383] "California Ave & Milwaukee Ave" NA NA NA ...
##  $ end_station_id    : num [1:16383] 13084 NA NA NA NA ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.6 -87.6 -87.7 -87.7 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.7 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:16383] "casual" "casual" "casual" "casual" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:02:43" "1899-12-31 00:03:16" ...
##  $ day_of_week       : num [1:16383] 5 5 5 5 5 5 5 5 5 5 ...
```

```r
print("OCT")
```

```
## [1] "OCT"
```

```r
str(oct_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "A50255C1E17942AB" "DB692A70BD2DD4E3" "3C02727AAF60F873" "47E65
##  $ rideable_type     : chr [1:16383] "classic_bike" "electric_bike" "electric_bike" "electric_bike"
##  $ started_at        : POSIXct[1:16383], format: "2022-10-14 17:13:30" "2022-10-01 16:29:26" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-10-14 17:19:39" "2022-10-01 16:49:06" ...
##  $ start_station_name: chr [1:16383] "Noble St & Milwaukee Ave" "Damen Ave & Charleston St" "Hoyne A
##  $ start_station_id  : chr [1:16383] "13290" "13288" "655" "KA1504000133" ...
##  $ end_station_name  : chr [1:16383] "Larrabee St & Division St" "Damen Ave & Cullerton St" "Western
##  $ end_station_id    : chr [1:16383] "KA1504000079" "13089" "TA1307000140" "620" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:16383] "member" "casual" "member" "member" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:06:09" "1899-12-31 00:19:40" ...
##  $ day_of_week       : num [1:16383] 6 7 4 2 5 5 5 4 7 2 ...
```

```r
print("NOV")
```

```
## [1] "NOV"
```

```r
str(nov_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "BCC66FC6FAB27CC7" "772AB67E902C180F" "585EAD07FDEC0152" "91C4E
##  $ rideable_type     : chr [1:16383] "electric_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:16383], format: "2022-11-10 06:21:55" "2022-11-04 07:31:55" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-11-10 06:31:27" "2022-11-04 07:46:25" ...
##  $ start_station_name: chr [1:16383] "Canal St & Adams St" "Canal St & Adams St" "Indiana Ave & Roos
##  $ start_station_id  : chr [1:16383] "13011" "13011" "SL-005" "SL-005" ...
##  $ end_station_name  : chr [1:16383] "St. Clair St & Erie St" "St. Clair St & Erie St" "St. Clair St
##  $ end_station_id    : chr [1:16383] "13016" "13016" "13016" "13016" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:16383] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:16383] "member" "member" "member" "member" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:09:32" "1899-12-31 00:14:30" ...
##  $ day_of_week       : num [1:16383] 5 6 2 6 3 6 1 3 1 3 ...
```

```
print("DEC")
```

```
## [1] "DEC"
```

```
str(dec_data)
```

```
## tibble [16,383 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:16383] "65DBD2F447EC51C2" "0C201AA7EA0EA1AD" "E0B148CCB358A49D" "54C57
##  $ rideable_type     : chr [1:16383] "electric_bike" "classic_bike" "electric_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:16383], format: "2022-12-05 10:47:18" "2022-12-18 06:42:33" ...
##  $ ended_at          : POSIXct[1:16383], format: "2022-12-05 10:56:34" "2022-12-18 07:08:44" ...
##  $ start_station_name: chr [1:16383] "Clifton Ave & Armitage Ave" "Broadway & Belmont Ave" "Sangamon
##  $ start_station_id  : chr [1:16383] "TA1307000163" "13277" "TA1306000015" "KA1503000038" ...
##  $ end_station_name  : chr [1:16383] "Sedgwick St & Webster Ave" "Sedgwick St & Webster Ave" "St. Cla
##  $ end_station_id    : chr [1:16383] "13191" "13191" "13016" "13134" ...
##  $ start_lat         : num [1:16383] 41.9 41.9 41.9 41.8 41.9 ...
##  $ start_lng         : num [1:16383] -87.7 -87.6 -87.7 -87.6 -87.7 ...
##  $ end_lat           : num [1:16383] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:16383] -87.6 -87.6 -87.6 -87.7 -87.7 ...
##  $ member_casual     : chr [1:16383] "member" "casual" "member" "member" ...
##  $ ride_length       : POSIXct[1:16383], format: "1899-12-31 00:09:16" "1899-12-31 00:26:11" ...
##  $ day_of_week       : num [1:16383] 2 1 3 3 4 6 3 3 3 4 ...
```

Upon closer inspection, we can see that much of the station data for the months of August and September is missing. Since the data is not missing at random, I will not remove rows with missing observations.

Convert all non-numerical columns to character

```
jan_data <-  mutate(jan_data, ride_id = as.character(ride_id)
                    ,rideable_type = as.character(rideable_type)
                    ,start_station_id = as.character(start_station_id)
                    ,end_station_id = as.character(end_station_id))
```

```
feb_data <-  mutate(feb_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

mar_data <-  mutate(mar_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

apr_data <-  mutate(apr_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

may_data <-  mutate(may_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

jun_data <-  mutate(jun_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

jul_data <-  mutate(jul_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

aug_data <-  mutate(aug_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

sep_data <-  mutate(sep_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

oct_data <-  mutate(oct_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

nov_data <-  mutate(nov_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
                   ,end_station_id = as.character(end_station_id))

dec_data <-  mutate(dec_data, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type)
                   ,start_station_id = as.character(start_station_id)
```

```
                    ,end_station_id = as.character(end_station_id))
```

Merge Datasets

```
year2022_data <- bind_rows(jan_data,
                       feb_data,
                       mar_data,
                       apr_data,
                       may_data,
                       jun_data,
                       jul_data,
                       aug_data,
                       sep_data,
                       oct_data,
                       nov_data,
                       dec_data
                       )
```

Inspecting Full Year Data

```
str(year2022_data)
```

```
## tibble [196,596 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:196596] "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB8
##  $ rideable_type     : chr [1:196596] "electric_bike" "electric_bike" "classic_bike" "classic_bike"
##  $ started_at        : POSIXct[1:196596], format: "2022-01-13 11:59:47" "2022-01-10 08:41:56" ...
##  $ ended_at          : POSIXct[1:196596], format: "2022-01-13 12:02:44" "2022-01-10 08:46:17" ...
##  $ start_station_name: chr [1:196596] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffie
##  $ start_station_id  : chr [1:196596] "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr [1:196596] "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave &
##  $ end_station_id    : chr [1:196596] "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...
##  $ start_lat         : num [1:196596] 42 42 41.9 42 41.9 ...
##  $ start_lng         : num [1:196596] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:196596] 42 42 41.9 42 41.9 ...
##  $ end_lng           : num [1:196596] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:196596] "casual" "casual" "member" "casual" ...
##  $ ride_length       : POSIXct[1:196596], format: "1899-12-31 00:02:57" "1899-12-31 00:04:21" ...
##  $ day_of_week       : num [1:196596] 5 2 3 3 5 3 1 7 2 6 ...
```

```
summary(year2022_data)
```

```
##    ride_id          rideable_type        started_at
##  Length:196596      Length:196596      Min.   :2022-01-01 00:03:36
##  Class :character   Class :character   1st Qu.:2022-04-01 00:36:37
##  Mode  :character   Mode  :character   Median :2022-06-30 23:56:44
##                                        Mean   :2022-07-01 14:21:10
##                                        3rd Qu.:2022-10-01 00:00:02
##                                         Max.   :2022-12-31 23:57:18
##
##     ended_at                   start_station_name start_station_id
##  Min.   :2022-01-01 00:04:02   Length:196596      Length:196596
##  1st Qu.:2022-04-01 00:51:42   Class :character   Class :character
##  Median :2022-07-01 00:52:44   Mode  :character   Mode  :character
##  Mean   :2022-07-01 14:36:24
##  3rd Qu.:2022-10-01 00:24:26
##  Max.   :2023-01-01 00:43:58
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:196596      Length:196596      Min.   :41.65   Min.   :-87.84
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.67
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.65
##                                         Mean   :41.90   Mean   :-87.65
##                                         3rd Qu.:41.93   3rd Qu.:-87.63
##                                         Max.   :45.64   Max.   :-73.80
##
##     end_lat         end_lng       member_casual
##  Min.   : 0.00   Min.   :-87.89   Length:196596
##  1st Qu.:41.88   1st Qu.:-87.67   Class :character
##  Median :41.90   Median :-87.65   Mode  :character
##  Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.12   Max.   :  0.00
##  NA's   :47      NA's   :47
##   ride_length                   day_of_week
##  Min.   :1899-12-30 23:35:23   Min.   :1.000
##  1st Qu.:1899-12-31 00:05:22   1st Qu.:2.000
##  Median :1899-12-31 00:09:19   Median :4.000
##  Mean   :1899-12-31 00:15:14   Mean   :4.069
##  3rd Qu.:1899-12-31 00:16:29   3rd Qu.:6.000
##  Max.   :1900-01-13 23:05:14   Max.   :7.000
##
```

Verifying that there are two unique values for member_casual

```r
unique(year2022_data$member_casual)
```

```
## [1] "casual" "member"
```

Creating columns for day, month, and year for aggregation purposes

```r
year2022_data$date <- as.Date(year2022_data$started_at)
year2022_data$month <- format(as.Date(year2022_data$date), "%m")
year2022_data$day <- format(as.Date(year2022_data$date), "%d")
year2022_data$year <- format(as.Date(year2022_data$date), "%Y")
```

Creating new ride_length column that shows seconds

```
year2022_data$ride_length <- difftime(year2022_data$ended_at,year2022_data$started_at)
```

Inspecting structure of updated data

```
str(year2022_data)
```

```
## tibble [196,596 x 19] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:196596] "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB8C
##  $ rideable_type     : chr [1:196596] "electric_bike" "electric_bike" "classic_bike" "classic_bike"
##  $ started_at        : POSIXct[1:196596], format: "2022-01-13 11:59:47" "2022-01-10 08:41:56" ...
##  $ ended_at          : POSIXct[1:196596], format: "2022-01-13 12:02:44" "2022-01-10 08:46:17" ...
##  $ start_station_name: chr [1:196596] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffiel
##  $ start_station_id  : chr [1:196596] "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr [1:196596] "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave &
##  $ end_station_id    : chr [1:196596] "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...
##  $ start_lat         : num [1:196596] 42 42 41.9 42 41.9 ...
##  $ start_lng         : num [1:196596] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:196596] 42 42 41.9 42 41.9 ...
##  $ end_lng           : num [1:196596] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:196596] "casual" "casual" "member" "casual" ...
##  $ ride_length       : 'difftime' num [1:196596] 177 261 261 896 ...
##   ..- attr(*, "units")= chr "secs"
##  $ day_of_week       : num [1:196596] 5 2 3 3 5 3 1 7 2 6 ...
##  $ date              : Date[1:196596], format: "2022-01-13" "2022-01-10" ...
##  $ month             : chr [1:196596] "01" "01" "01" "01" ...
##  $ day               : chr [1:196596] "13" "10" "25" "04" ...
##  $ year              : chr [1:196596] "2022" "2022" "2022" "2022" ...
```

Converting ride_length to numerical for easier calculations

```
year2022_data$ride_length <- as.numeric(as.character(year2022_data$ride_length))
```

Descriptive Analysis

Calculate Summary Statistics of Ride Length

```
summary(year2022_data$ride_length)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##   -1477.0     322.0     559.0     913.7     989.0 1206314.0
```

Subtract Observations with Negative Trip Length

```
year2022_data <- year2022_data[!(year2022_data$ride_length < 0),]
```

New Summary Statistics

```
summary(year2022_data$ride_length)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##       0.0     322.0     559.0     913.8     989.0 1206314.0
```

Comparing Members vs. Casual Riders

```
aggregate(year2022_data$ride_length ~ year2022_data$member_casual, FUN = mean)
```

```
##   year2022_data$member_casual year2022_data$ride_length
## 1                      casual                 1252.1912
## 2                      member                  708.6305
```

```
aggregate(year2022_data$ride_length ~ year2022_data$member_casual, FUN = median)
```

```
##   year2022_data$member_casual year2022_data$ride_length
## 1                      casual                       675
## 2                      member                       505
```

```
aggregate(year2022_data$ride_length ~ year2022_data$member_casual, FUN = max)
```

```
##   year2022_data$member_casual year2022_data$ride_length
## 1                      casual                   1206314
## 2                      member                     89996
```

```
aggregate(year2022_data$ride_length ~ year2022_data$member_casual, FUN = min)
```

```
##   year2022_data$member_casual year2022_data$ride_length
## 1                      casual                         0
## 2                      member                         0
```

Average Ride Time by Day of Week for Members vs. Casual Riders

```
aggregate(year2022_data$ride_length ~ year2022_data$member_casual + year2022_data$day_of_week, FUN = mea
##    year2022_data$member_casual year2022_data$day_of_week
## 1                       casual                         1
## 2                       member                         1
## 3                       casual                         2
## 4                       member                         2
## 5                       casual                         3
## 6                       member                         3
## 7                       casual                         4
## 8                       member                         4
## 9                       casual                         5
## 10                      member                         5
## 11                      casual                         6
## 12                      member                         6
## 13                      casual                         7
## 14                      member                         7
##    year2022_data$ride_length
## 1                  1494.5195
## 2                   773.6045
## 3                  1347.7381
## 4                   696.9751
## 5                  1043.0929
## 6                   685.1443
## 7                  1020.0056
## 8                   695.9388
## 9                  1088.8486
## 10                  682.8335
## 11                 1203.0795
## 12                  678.7954
## 13                 1409.3764
## 14                  771.5435
```

Average Ride Time by Month for Members vs. Casual Riders

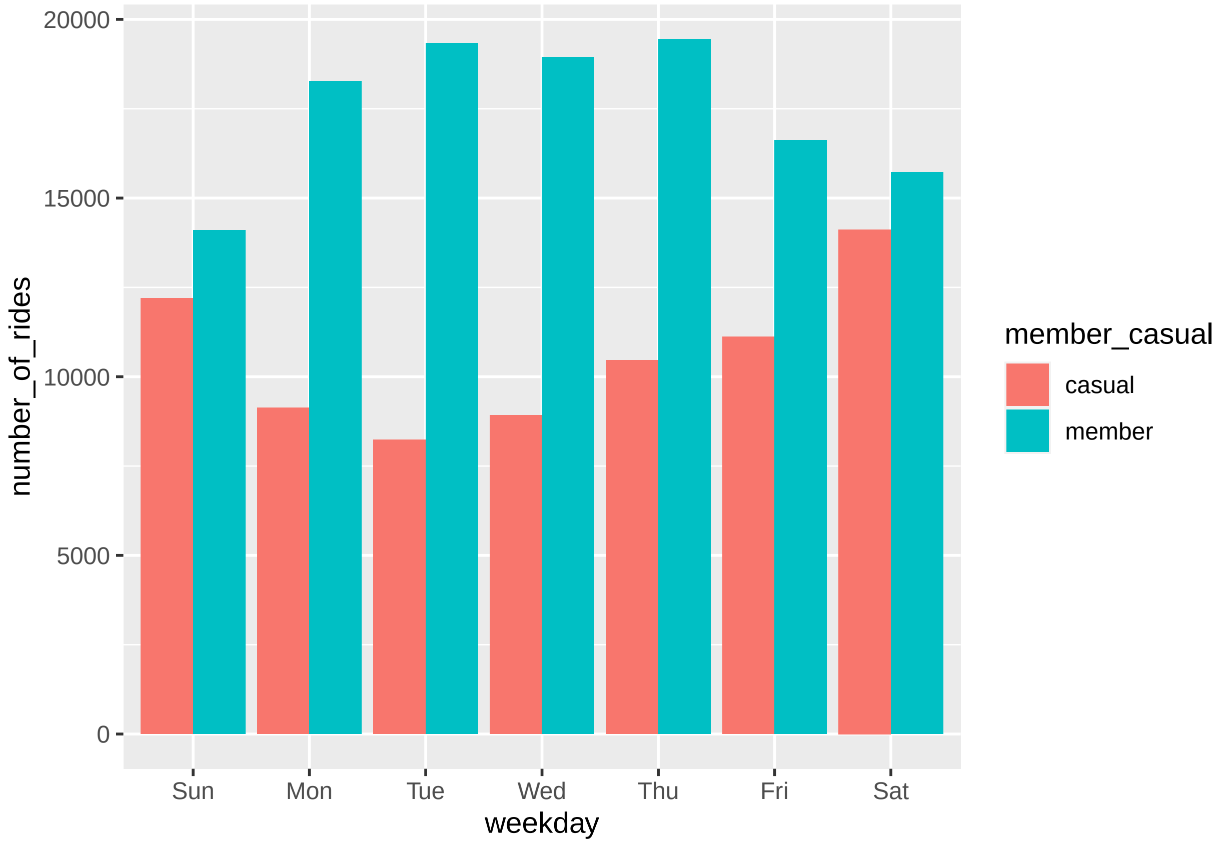
```
aggregate(year2022_data$ride_length ~ year2022_data$member_casual + year2022_data$month, FUN = mean)
```

```
##    year2022_data$member_casual year2022_data$month year2022_data$ride_length
## 1                       casual                  01                 1108.4825
## 2                       member                  01                  653.1012
## 3                       casual                  02                 1480.9223
## 4                       member                  02                  680.0305
## 5                       casual                  03                 1645.2303
## 6                       member                  03                  728.4413
## 7                       casual                  04                 1669.5367
## 8                       member                  04                  702.1838
## 9                       casual                  05                 2288.2321
## 10                      member                  05                  831.0115
## 11                      casual                  06                 1234.3803
## 12                      member                  06                  798.1082
## 13                      casual                  07                 1629.9879
## 14                      member                  07                  785.8429
## 15                      casual                  08                  848.8664
## 16                      member                  08                  717.6562
## 17                      casual                  09                  912.2009
## 18                      member                  09                  793.3800
## 19                      casual                  10                 1025.3422
## 20                      member                  10                  671.8275
## 21                      casual                  11                 1093.2376
## 22                      member                  11                  653.5631
## 23                      casual                  12                  878.4070
## 24                      member                  12                  595.5764
```

Visualization of Number of Rides by Rider Type

```
year2022_data %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge")
```

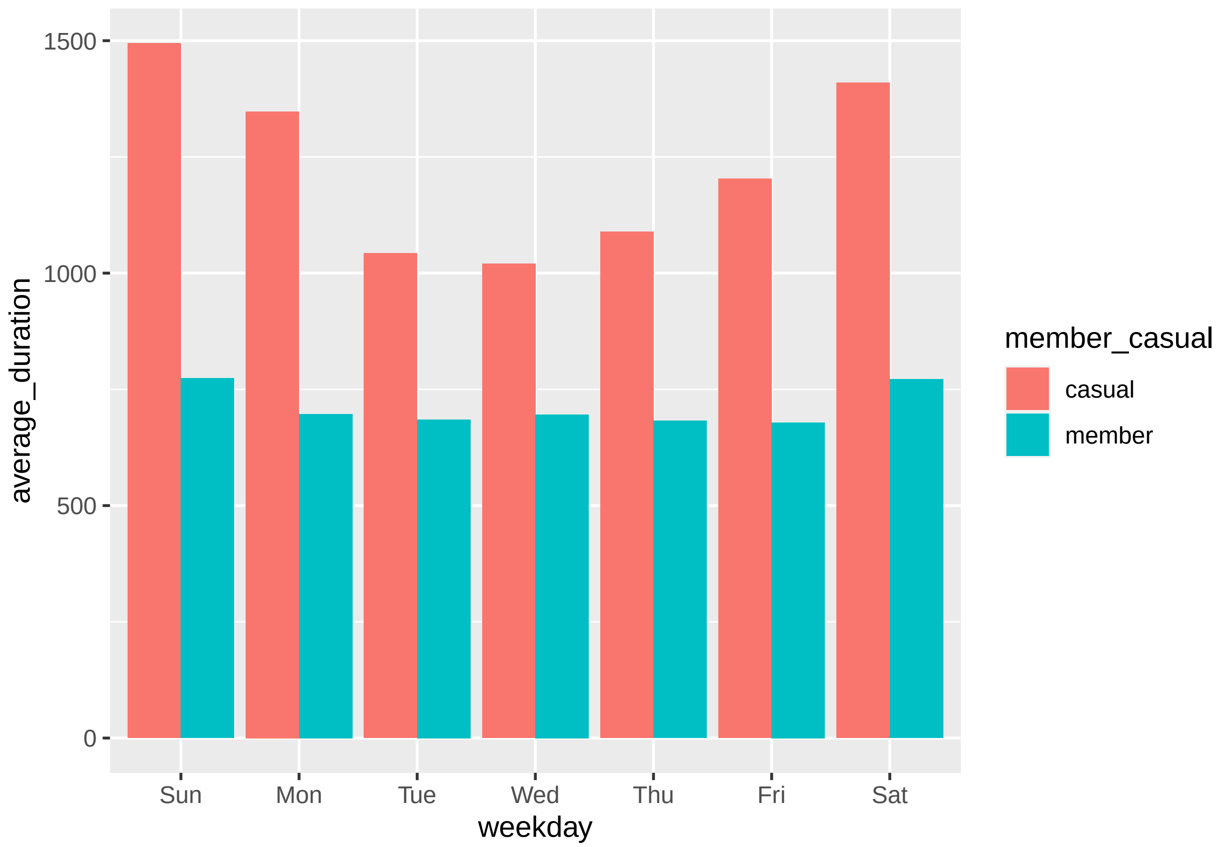
```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```



Visualization for Average Duration

```
year2022_data %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")
```

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



Exporting CSV File for Full 2022 Year Dataset

```
write.csv(year2022_data, file = ’C:/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study
```

Exporting CSV File to Further Analyze Average Ride Length

```
counts <- aggregate(year2022_data$ride_length ~ year2022_data$member_casual + year2022_data$day_of_week

write.csv(counts, file = ’C:/Users/kheng/OneDrive/Documents/Google Career Certificate Case Study/cyclis
```