GoogleCapstoneTT_Part3_FromCleanDataToAnalysis

TT

31/01/2022

R Markdown

This is **third** part of my capstone project. Ref to previous parts.

In the first part data have been collected and browsed. Therafter data was manipulated and processed. As as result data from various csv files has been made compatible. Thereafter data was merged into one large dataframe and exported to a big CSV file for further analysis. In the second part data was analysed in terms of compatibility, some column types were changed, data was cleaned and exported to new clean CSV file.

Loading required packages

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

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

```
library(lubridate)
```

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

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

```
library(ggplot2)
library(dplyr)
```

Loading previously created database:

```
bike_data_2019_cleaned <- read.csv("E:/Tomasz/CapstoneGoogle/capstone_bike_data_cleaned.csv")
```

Duplicate dataframe in order to receive posssibility to return to original data frame.

```
BikeData1 = bike_data_2019_cleaned

head(BikeData1)
```

```
##   X.1 X  ride_id          started_at            ended_at rideable_type
## 1   1 1 21742443 2019-01-01 00:04:37 2019-01-01 00:11:07          2167
## 2   2 2 21742444 2019-01-01 00:08:13 2019-01-01 00:15:34          4386
## 3   3 3 21742445 2019-01-01 00:13:23 2019-01-01 00:27:12          1524
## 4   4 4 21742446 2019-01-01 00:13:45 2019-01-01 00:43:28           252
## 5   5 5 21742447 2019-01-01 00:14:52 2019-01-01 00:20:56          1170
## 6   6 6 21742448 2019-01-01 00:15:33 2019-01-01 00:19:09          2437
##   tripduration start_station_id                  start_station_name
## 1          390              199              Wabash Ave & Grand Ave
## 2          441               44              State St & Randolph St
## 3          829               15                Racine Ave & 18th St
## 4         1783              123      California Ave & Milwaukee Ave
## 5          364              173 Mies van der Rohe Way & Chicago Ave
## 6          216               98          LaSalle St & Washington St
##   end_station_id               end_station_name member_casual       date month
## 1             84      Milwaukee Ave & Grand Ave    Subscriber 2019-01-01     1
## 2            624 Dearborn St & Van Buren St (*)    Subscriber 2019-01-01     1
## 3            644  Western Ave & Fillmore St (*)    Subscriber 2019-01-01     1
## 4            176              Clark St & Elm St    Subscriber 2019-01-01     1
## 5             35        Streeter Dr & Grand Ave    Subscriber 2019-01-01     1
## 6             49        Dearborn St & Monroe St    Subscriber 2019-01-01     1
##   day year day_of_week ride_length
## 1   1 2019     Tuesday    6.500000
## 2   1 2019     Tuesday    7.350000
## 3   1 2019     Tuesday   13.816667
## 4   1 2019     Tuesday   29.716667
## 5   1 2019     Tuesday    6.066667
## 6   1 2019     Tuesday    3.600000
```

```
str(BikeData1)
```

```
## 'data.frame':    3818004 obs. of  18 variables:
##  $ X.1               : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ X                 : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ ride_id           : num  21742443 21742444 21742445 21742446 21742447 ...
##  $ started_at        : chr  "2019-01-01 00:04:37" "2019-01-01 00:08:13" "2019-01-01 00:13:23" "2019-01-01 00:13:45" ...
##  $ ended_at          : chr  "2019-01-01 00:11:07" "2019-01-01 00:15:34" "2019-01-01 00:27:12" "2019-01-01 00:43:28" ...
##  $ rideable_type     : int  2167 4386 1524 252 1170 2437 2708 2796 6205 3939 ...
##  $ tripduration      : int  390 441 829 1783 364 216 177 100 1727 336 ...
##  $ start_station_id  : int  199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr  "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave & 18th St" "California Ave &
Milwaukee Ave" ...
##  $ end_station_id    : int  84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr  "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "Western Ave & Fillmore St
(*)" "Clark St & Elm St" ...
##  $ member_casual     : chr  "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
##  $ date              : chr  "2019-01-01" "2019-01-01" "2019-01-01" "2019-01-01" ...
##  $ month             : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ day               : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ year              : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ day_of_week       : chr  "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : num  6.5 7.35 13.82 29.72 6.07 ...
```

Insight No 1:

After export CSV file some unwanted colums were added and type of columns “started_at” and “ended at” where changed from date to character. This needs to addresed with dropping columns and type change.

Change type for colums:

```
BikeData1$started_at <- as_datetime(BikeData1$started_at)
BikeData1$ended_at <- as_datetime(BikeData1$ended_at)

str(BikeData1)
```

```
## 'data.frame':    3818004 obs. of  18 variables:
##  $ X.1               : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ X                 : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ ride_id           : num  21742443 21742444 21742445 21742446 21742447 ...
##  $ started_at        : POSIXct, format: "2019-01-01 00:04:37" "2019-01-01 00:08:13" ...
##  $ ended_at          : POSIXct, format: "2019-01-01 00:11:07" "2019-01-01 00:15:34" ...
##  $ rideable_type     : int  2167 4386 1524 252 1170 2437 2708 2796 6205 3939 ...
##  $ tripduration      : int  390 441 829 1783 364 216 177 100 1727 336 ...
##  $ start_station_id  : int  199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr  "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave & 18th St" "California Ave &
Milwaukee Ave" ...
##  $ end_station_id    : int  84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr  "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "Western Ave & Fillmore St
(*)" "Clark St & Elm St" ...
##  $ member_casual     : chr  "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
##  $ date              : chr  "2019-01-01" "2019-01-01" "2019-01-01" "2019-01-01" ...
##  $ month             : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ day               : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ year              : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ day_of_week       : chr  "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : num  6.5 7.35 13.82 29.72 6.07 ...
```

#Removing unwanted columns

```
BikeData1 <- BikeData1 %>%
  select(-c("X.1","X"))

str(BikeData1)
```

```
## 'data.frame':    3818004 obs. of  16 variables:
##  $ ride_id           : num  21742443 21742444 21742445 21742446 21742447 ...
##  $ started_at        : POSIXct, format: "2019-01-01 00:04:37" "2019-01-01 00:08:13" ...
##  $ ended_at          : POSIXct, format: "2019-01-01 00:11:07" "2019-01-01 00:15:34" ...
##  $ rideable_type     : int  2167 4386 1524 252 1170 2437 2708 2796 6205 3939 ...
##  $ tripduration      : int  390 441 829 1783 364 216 177 100 1727 336 ...
##  $ start_station_id  : int  199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr  "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave & 18th St" "California Ave &
Milwaukee Ave" ...
##  $ end_station_id    : int  84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr  "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "Western Ave & Fillmore St
(*)" "Clark St & Elm St" ...
##  $ member_casual     : chr  "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
##  $ date              : chr  "2019-01-01" "2019-01-01" "2019-01-01" "2019-01-01" ...
##  $ month             : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ day               : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ year              : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ day_of_week       : chr  "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : num  6.5 7.35 13.82 29.72 6.07 ...
```

Now we can see we have 16 not 18 variables.

Descriptive analysis

```
summary(BikeData1)
```

```
##     ride_id           started_at                     ended_at                  
##  Min.   :21742443   Min.   :2019-01-01 00:04:37   Min.   :2019-01-01 00:11:07  
##  1st Qu.:22873787   1st Qu.:2019-05-29 15:49:26   1st Qu.:2019-05-29 16:09:28  
##  Median :23962320   Median :2019-07-25 17:50:54   Median :2019-07-25 18:12:23  
##  Mean   :23915629   Mean   :2019-07-19 21:47:37   Mean   :2019-07-19 22:11:47  
##  3rd Qu.:24963703   3rd Qu.:2019-09-15 06:48:05   3rd Qu.:2019-09-15 08:30:13  
##  Max.   :25962904   Max.   :2019-12-31 23:57:17   Max.   :2020-01-21 13:54:35  
##                                                                                
##  rideable_type   tripduration      start_station_id start_station_name
##  Min.   :   1   Min.   :      61   Min.   :  1.0    Length:3818004    
##  1st Qu.:1727   1st Qu.:     405   1st Qu.: 77.0    Class :character  
##  Median :3451   Median :     696   Median :174.0    Mode  :character  
##  Mean   :3380   Mean   :    1500   Mean   :201.7                      
##  3rd Qu.:5046   3rd Qu.:    1257   3rd Qu.:289.0                      
##  Max.   :6946   Max.   :10628400   Max.   :673.0                      
##                 NA's   :1108163                                       
##  end_station_id  end_station_name   member_casual          date          
##  Min.   :  1.0   Length:3818004     Length:3818004     Length:3818004    
##  1st Qu.: 77.0   Class :character   Class :character   Class :character  
##  Median :174.0   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :202.6                                                           
##  3rd Qu.:291.0                                                           
##  Max.   :673.0                                                           
##                                                                          
##      month             day             year      day_of_week       
##  Min.   : 1.000   Min.   : 1.00   Min.   :2019   Length:3818004    
##  1st Qu.: 5.000   1st Qu.: 8.00   1st Qu.:2019   Class :character  
##  Median : 7.000   Median :15.00   Median :2019   Mode  :character  
##  Mean   : 7.107   Mean   :15.53   Mean   :2019                     
##  3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:2019                     
##  Max.   :12.000   Max.   :31.00   Max.   :2019                     
##                                                                    
##   ride_length      
##  Min.   :   -56.37  
##  1st Qu.:     6.85  
##  Median :    11.82  
##  Mean   :    24.17  
##  3rd Qu.:    21.40  
##  Max.   :177200.37  
##
```

Insight No 3:

A. Ride length max 177200.37 minutes seems to be unrealistic

B. Ride length cannot be obviously negative value (minimum -56).

C. Given A. and B. above other measures like mean, median etc will probably alse be affected.This needs to be adressed before further analysis of the ride_legth column.

Exploring most popular stations:

```
top_start_stations <- head(sort(table(BikeData1$start_station_name), decreasing = T), 10)
top_start_stations <- as.data.frame(top_start_stations)
colnames(top_start_stations) <- c('Station','Number of bike rides')
top_start_stations
```

```
##                          Station Number of bike rides
## 1       Streeter Dr & Grand Ave                67983
## 2           Canal St & Adams St                54389
## 3       Clinton St & Madison St                49908
## 4     Lake Shore Dr & Monroe St                49804
## 5  Clinton St & Washington Blvd                48153
## 6     Columbus Dr & Randolph St                39192
## 7  Michigan Ave & Washington St                37696
## 8         Michigan Ave & Oak St                35449
## 9    Lake Shore Dr & North Blvd                34472
## 10      Franklin St & Monroe St                34297
```

```
top_end_stations <- head(sort(table(BikeData1$end_station_name), decreasing = T), 10)
top_end_stations <- as.data.frame(top_end_stations)
colnames(top_end_stations) <- c('Station','Number of bike rides')
top_end_stations
```

```
##                          Station Number of bike rides
## 1       Streeter Dr & Grand Ave                81723
## 2  Clinton St & Washington Blvd                50686
## 3           Canal St & Adams St                50127
## 4       Clinton St & Madison St                47311
## 5    Lake Shore Dr & North Blvd                42459
## 6     Lake Shore Dr & Monroe St                40633
## 7  Michigan Ave & Washington St                40331
## 8         Michigan Ave & Oak St                37849
## 9               Millennium Park                37424
## 10          Theater on the Lake                35939
```

Exploring least popular stations:

```
bottom_start_stations <- head(sort(table(BikeData1$start_station_name)), 10)
bottom_start_stations <- as.data.frame(bottom_start_stations)
colnames(bottom_start_stations) <- c('Station','Number of bike rides')
bottom_start_stations
```

```
##                                    Station Number of bike rides
## 1                       LBS - BBB La Magie                    1
## 2                           Special Events                    1
## 3    DIVVY CASSETTE REPAIR MOBILE STATION                    3
## 4                   Elizabeth St & 59th St                    8
## 5                     Racine Ave & 61st St                    9
## 6                     Racine Ave & 65th St                   10
## 7                   Michigan Ave & 71st St                   11
## 8                   Carpenter St & 63rd St                   13
## 9           South Chicago Ave & Elliot Ave                   13
## 10 HUBBARD ST BIKE CHECKING (LBS-WH-TEST)                   15
```

```
bottom_end_stations <- head(sort(table(BikeData1$end_station_name)), 10)
bottom_end_stations <- as.data.frame(bottom_end_stations)
colnames(bottom_end_stations) <- c('Station','Number of bike rides')
bottom_end_stations
```

```
##                                  Station Number of bike rides
## 1                     LBS - BBB La Magie                    1
## 2                         Special Events                    1
## 3               TS ~ DIVVY PARTS TESTING                    2
## 4  DIVVY CASSETTE REPAIR MOBILE STATION                    5
## 5                   Racine Ave & 61st St                    8
## 6         South Chicago Ave & Elliot Ave                   10
## 7                 Elizabeth St & 59th St                   12
## 8                 Michigan Ave & 71st St                   16
## 9                   Racine Ave & 65th St                   16
## 10                 Kostner Ave & Lake St                   17
```

Analysing column ride_legth could be intersting and lead to some insights.

Longest and shortest ride

```
max(BikeData1$ride_length)
```

```
## [1] 177200.4
```

```
min(BikeData1$ride_length)
```

```
## [1] -56.36667
```

Average ride length

```
mean(BikeData1$ride_length)
```

```
## [1] 24.17419
```

Median ride length

```
median(BikeData1$ride_length)
```

```
## [1] 11.81667
```

Inspecting top ride lengths

```
top_ride_lengths <- sort(BikeData1$ride_length, decreasing = T)
top_ride_lengths <- as.data.frame(top_ride_lengths)
head(top_ride_lengths, 30)
```

```
##    top_ride_lengths
## 1         177200.37
## 2         150943.90
## 3         143038.37
## 4         136727.30
## 5         135279.75
## 6         132324.13
## 7         127791.17
## 8         126781.20
## 9         125367.70
## 10        123677.20
## 11        120795.83
## 12        115859.78
## 13        115675.75
## 14        112425.82
## 15        110497.58
## 16        108469.07
## 17        106987.07
## 18        102696.23
## 19        102198.70
## 20        101607.13
## 21        100605.72
## 22        100476.70
## 23        100354.30
## 24         98966.10
## 25         95101.70
## 26         94162.97
## 27         93812.98
## 28         93461.85
## 29         93356.23
## 30         91181.12
```

Insight no 3:

A. Longests rides duration is very unrealistic. This should be adressed and probably replaced with better quality data.

B. For purpose of this capstone all rides above 10000 min will be arbitrarily removed in next steps.

C. Keeping long rides in dataframe could significantly skew or bias results.

D: Rides with negative length will also be removed.

New dataframe “Bike_Data3” with removal of negative and long ride lengths will be created.

```
BikeData3 <- BikeData1[!(BikeData1$ride_length<0|BikeData1$ride_length>10000),]
```

Descriptive analysis of new dataframe:

```
mean(BikeData3$ride_length)
```

```
## [1] 20.22873
```

```
median(BikeData3$ride_length)
```

```
## [1] 11.81667
```

```
max(BikeData3$ride_length)
```

```
## [1] 9990.783
```

```
min(BikeData3$ride_length)
```

```
## [1] 1.016667
```

```
summary(BikeData3)
```

```
##     ride_id           started_at                     ended_at                  
##  Min.   :21742443   Min.   :2019-01-01 00:04:37   Min.   :2019-01-01 00:11:07  
##  1st Qu.:22873736   1st Qu.:2019-05-29 15:46:09   1st Qu.:2019-05-29 16:06:10  
##  Median :23962242   Median :2019-07-25 17:49:16   Median :2019-07-25 18:08:36  
##  Mean   :23915593   Mean   :2019-07-19 21:44:44   Mean   :2019-07-19 22:04:58  
##  3rd Qu.:24963706   3rd Qu.:2019-09-15 06:57:51   3rd Qu.:2019-09-15 07:53:16  
##  Max.   :25962904   Max.   :2019-12-31 23:57:17   Max.   :2020-01-02 18:33:31  
##                                                                                
##  rideable_type   tripduration     start_station_id start_station_name
##  Min.   :   1   Min.   :    61    Min.   :  1.0    Length:3817572    
##  1st Qu.:1727   1st Qu.:   405    1st Qu.: 77.0    Class :character  
##  Median :3451   Median :   696    Median :174.0    Mode  :character  
##  Mean   :3380   Mean   :  1197    Mean   :201.7                      
##  3rd Qu.:5046   3rd Qu.:  1256    3rd Qu.:289.0                      
##  Max.   :6946   Max.   :599446    Max.   :673.0                      
##                 NA's   :1108109                                      
##  end_station_id  end_station_name   member_casual          date          
##  Min.   :  1.0   Length:3817572     Length:3817572     Length:3817572    
##  1st Qu.: 77.0   Class :character   Class :character   Class :character  
##  Median :174.0   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :202.6                                                           
##  3rd Qu.:291.0                                                           
##  Max.   :673.0                                                           
##                                                                          
##      month             day             year      day_of_week       
##  Min.   : 1.000   Min.   : 1.00   Min.   :2019   Length:3817572    
##  1st Qu.: 5.000   1st Qu.: 8.00   1st Qu.:2019   Class :character  
##  Median : 7.000   Median :15.00   Median :2019   Mode  :character  
##  Mean   : 7.107   Mean   :15.53   Mean   :2019                     
##  3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:2019                     
##  Max.   :12.000   Max.   :31.00   Max.   :2019                     
##                                                                    
##   ride_length      
##  Min.   :   1.017  
##  1st Qu.:   6.850  
##  Median :  11.817  
##  Mean   :  20.229  
##  3rd Qu.:  21.383  
##  Max.   :9990.783  
## 
```

```
str(BikeData3)
```

```
## 'data.frame':    3817572 obs. of  16 variables:
##  $ ride_id           : num  21742443 21742444 21742445 21742446 21742447 ...
##  $ started_at        : POSIXct, format: "2019-01-01 00:04:37" "2019-01-01 00:08:13" ...
##  $ ended_at          : POSIXct, format: "2019-01-01 00:11:07" "2019-01-01 00:15:34" ...
##  $ rideable_type     : int  2167 4386 1524 252 1170 2437 2708 2796 6205 3939 ...
##  $ tripduration      : int  390 441 829 1783 364 216 177 100 1727 336 ...
##  $ start_station_id  : int  199 44 15 123 173 98 98 211 150 268 ...
##  $ start_station_name: chr  "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave & 18th St" "California Ave &
Milwaukee Ave" ...
##  $ end_station_id    : int  84 624 644 176 35 49 49 142 148 141 ...
##  $ end_station_name  : chr  "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "Western Ave & Fillmore St
(*)" "Clark St & Elm St" ...
##  $ member_casual     : chr  "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
##  $ date              : chr  "2019-01-01" "2019-01-01" "2019-01-01" "2019-01-01" ...
##  $ month             : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ day               : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ year              : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ day_of_week       : chr  "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : num  6.5 7.35 13.82 29.72 6.07 ...
```

Below there will be analysing of data in relation to different features.

Ride lenght by membership type

Mean

```
aggregate(BikeData3$ride_length~ BikeData3$member_casual, FUN = mean)
```

```
##   BikeData3$member_casual BikeData3$ride_length
## 1                Customer              43.11127
## 2              Subscriber              13.37094
```

Median

```
aggregate(BikeData3$ride_length~ BikeData3$member_casual, FUN = median)
```

```
##   BikeData3$member_casual BikeData3$ride_length
## 1                Customer              25.81667
## 2              Subscriber               9.80000
```

Min

```
aggregate(BikeData3$ride_length~ BikeData3$member_casual, FUN = min)
```

```
##   BikeData3$member_casual BikeData3$ride_length
## 1                Customer              1.016667
## 2              Subscriber              1.016667
```

Max

```
aggregate(BikeData3$ride_length~ BikeData3$member_casual, FUN = max)
```

```
##   BikeData3$member_casual BikeData3$ride_length
## 1                Customer              9990.783
## 2              Subscriber              9919.817
```

Ride length by date:

Mean

```
aggregate(BikeData3$ride_length~ BikeData3$date, FUN = mean)
```

```
##     BikeData3$date BikeData3$ride_length
## 1       2019-01-01              22.14639
## 2       2019-01-02              11.45478
## 3       2019-01-03              12.78807
## 4       2019-01-04              15.49217
## 5       2019-01-05              22.98920
## 6       2019-01-06              14.43493
## 7       2019-01-07              14.54389
## 8       2019-01-08              13.15567
## 9       2019-01-09              13.74168
## 10      2019-01-10              11.36214
## 11      2019-01-11              11.48342
## 12      2019-01-12              20.67730
## 13      2019-01-13              16.11526
## 14      2019-01-14              11.89209
## 15      2019-01-15              10.86616
## 16      2019-01-16              11.46395
## 17      2019-01-17              12.98396
## 18      2019-01-18              11.38110
## 19      2019-01-19              16.16004
## 20      2019-01-20              23.22623
## 21      2019-01-21              14.18466
## 22      2019-01-22              14.35121
## 23      2019-01-23              12.28973
## 24      2019-01-24              13.78439
## 25      2019-01-25              14.08958
## 26      2019-01-26              26.06556
## 27      2019-01-27              21.02830
## 28      2019-01-28              15.43748
## 29      2019-01-29              38.94805
## 30      2019-01-30              49.94071
## 31      2019-01-31              41.84143
## 32      2019-02-01              21.43275
## 33      2019-02-02              21.50503
## 34      2019-02-03              15.91251
## 35      2019-02-04              12.88729
## 36      2019-02-05              12.67378
## 37      2019-02-06              12.45671
## 38      2019-02-07              11.68014
## 39      2019-02-08              10.68665
## 40      2019-02-09              14.52549
## 41      2019-02-10              18.27355
## 42      2019-02-11              10.40417
## 43      2019-02-12              11.66502
## 44      2019-02-13              10.86830
## 45      2019-02-14              12.19524
## 46      2019-02-15              10.60238
## 47      2019-02-16              18.27401
## 48      2019-02-17              12.09108
## 49      2019-02-18              12.34826
## 50      2019-02-19              10.84570
## 51      2019-02-20              12.00478
## 52      2019-02-21              12.89657
## 53      2019-02-22              12.21521
## 54      2019-02-23              13.76271
## 55      2019-02-24              16.22505
## 56      2019-02-25              11.92194
## 57      2019-02-26              11.22830
## 58      2019-02-27              10.61898
## 59      2019-02-28              11.99725
## 60      2019-03-01              12.17621
## 61      2019-03-02              16.49392
## 62      2019-03-03              16.78116
## 63      2019-03-04              11.61487
## 64      2019-03-05              10.47886
## 65      2019-03-06              11.62675
## 66      2019-03-07              10.33856
## 67      2019-03-08              12.60523
## 68      2019-03-09              15.14612
## 69      2019-03-10              13.52190
## 70      2019-03-11              12.63015
## 71      2019-03-12              14.02792
## 72      2019-03-13              14.52423
## 73      2019-03-14              14.69882
## 74      2019-03-15              13.13825
## 75      2019-03-16              21.33011
## 76      2019-03-17              22.18365
## 77      2019-03-18              13.28452
## 78      2019-03-19              14.29464
## 79      2019-03-20              11.77107
## 80      2019-03-21              12.31119
## 81      2019-03-22              13.27002
## 82      2019-03-23              22.64908
## 83      2019-03-24              17.21464
## 84      2019-03-25              12.58608
## 85      2019-03-26              13.14971
## 86      2019-03-27              17.31102
## 87      2019-03-28              15.77009
## 88      2019-03-29              15.35277
## 89      2019-03-30              18.32835
## 90      2019-03-31              15.41291
## 91      2019-04-01              12.57095
## 92      2019-04-02              12.38352
## 93      2019-04-03              14.72519
## 94      2019-04-04              10.93617
## 95      2019-04-05              15.43004
## 96      2019-04-06              28.45972
## 97      2019-04-07              26.05714
## 98      2019-04-08              18.96555
## 99      2019-04-09              18.25690
## 100     2019-04-10              13.86956
## 101     2019-04-11              12.57035
## 102     2019-04-12              15.10075
## 103     2019-04-13              25.31116
## 104     2019-04-14              15.28112
## 105     2019-04-15              14.84661
## 106     2019-04-16              18.00130
## 107     2019-04-17              18.16615
## 108     2019-04-18              13.39782
## 109     2019-04-19              13.54395
## 110     2019-04-20              25.83927
## 111     2019-04-21              33.91959
## 112     2019-04-22              22.27518
## 113     2019-04-23              15.65292
## 114     2019-04-24              16.20465
## 115     2019-04-25              15.29988
## 116     2019-04-26              20.95795
## 117     2019-04-27              12.46115
## 118     2019-04-28              21.28276
## 119     2019-04-29              16.55089
## 120     2019-04-30              11.06976
## 121     2019-05-01              17.38113
## 122     2019-05-02              12.12755
## 123     2019-05-03              15.81201
## 124     2019-05-04              26.51677
## 125     2019-05-05              31.74186
## 126     2019-05-06              16.28479
## 127     2019-05-07              12.82556
## 128     2019-05-08              16.06818
## 129     2019-05-09              16.46801
## 130     2019-05-10              16.24780
## 131     2019-05-11              20.62274
## 132     2019-05-12              20.40660
## 133     2019-05-13              16.87098
## 134     2019-05-14              17.88362
## 135     2019-05-15              19.39333
## 136     2019-05-16              17.95475
## 137     2019-05-17              13.97065
## 138     2019-05-18              28.73339
## 139     2019-05-19              28.47655
## 140     2019-05-20              16.51659
## 141     2019-05-21              13.15388
## 142     2019-05-22              20.94925
## 143     2019-05-23              21.51891
## 144     2019-05-24              19.56950
## 145     2019-05-25                31.69990
## 146     2019-05-26                34.84245
## 147     2019-05-27                28.80619
## 148     2019-05-28                14.85851
## 149     2019-05-29                17.53571
## 150     2019-05-30                18.53965
## 151     2019-05-31                23.51705
## 152     2019-06-01                24.52509
## 153     2019-06-02                29.45516
## 154     2019-06-03                18.40445
## 155     2019-06-04                18.17116
## 156     2019-06-05                19.85806
## 157     2019-06-06                17.91561
## 158     2019-06-07                20.98239
## 159     2019-06-08                29.47857
## 160     2019-06-09                23.58256
## 161     2019-06-10                20.01584
## 162     2019-06-11                20.53976
## 163     2019-06-12                14.55753
## 164     2019-06-13                17.77995
## 165     2019-06-14                22.02162
## 166     2019-06-15                25.89234
## 167     2019-06-16                23.04340
## 168     2019-06-17                18.26853
## 169     2019-06-18                21.46774
## 170     2019-06-19                16.07398
## 171     2019-06-20                17.45708
## 172     2019-06-21                22.32245
## 173     2019-06-22                30.60184
## 174     2019-06-23                27.39860
## 175     2019-06-24                19.63936
## 176     2019-06-25                21.74416
## 177     2019-06-26                20.67307
## 178     2019-06-27                21.31644
## 179     2019-06-28                21.18814
## 180     2019-06-29                31.70122
## 181     2019-06-30                30.94025
## 182     2019-07-01                21.53971
## 183     2019-07-02                19.70120
## 184     2019-07-03                20.87521
## 185     2019-07-04                34.51781
## 186     2019-07-05                29.73861
## 187     2019-07-06                32.58750
## 188     2019-07-07                20.60204
## 189     2019-07-08                23.72166
## 190     2019-07-09                21.04382
## 191     2019-07-10                19.69019
## 192     2019-07-11                20.19965
## 193     2019-07-12                22.91619
## 194     2019-07-13                30.38426
## 195     2019-07-14                30.46272
## 196     2019-07-15                20.57266
## 197     2019-07-16                18.31829
## 198     2019-07-17                18.93402
## 199     2019-07-18                19.98032
## 200     2019-07-19                21.79175
## 201     2019-07-20                28.02739
## 202     2019-07-21                27.69852
## 203     2019-07-22                22.97409
## 204     2019-07-23                20.47742
## 205     2019-07-24                19.80688
## 206     2019-07-25                23.04322
## 207     2019-07-26                21.70466
## 208     2019-07-27                31.54879
## 209     2019-07-28                30.91370
## 210     2019-07-29                18.21427
## 211     2019-07-30                21.63162
## 212     2019-07-31                20.60153
## 213     2019-08-01                19.74344
## 214     2019-08-02                23.47565
## 215     2019-08-03                30.92067
## 216     2019-08-04                29.78196
## 217     2019-08-05                21.94389
## 218     2019-08-06              20.60997
## 219     2019-08-07              20.88833
## 220     2019-08-08              21.03278
## 221     2019-08-09              23.84575
## 222     2019-08-10              32.41693
## 223     2019-08-11              28.98195
## 224     2019-08-12              19.15986
## 225     2019-08-13              21.26964
## 226     2019-08-14              19.03280
## 227     2019-08-15              20.85822
## 228     2019-08-16              22.08797
## 229     2019-08-17              30.18977
## 230     2019-08-18              27.80548
## 231     2019-08-19              20.11626
## 232     2019-08-20              17.03339
## 233     2019-08-21              18.99055
## 234     2019-08-22              20.13216
## 235     2019-08-23              21.15147
## 236     2019-08-24              30.04899
## 237     2019-08-25              30.90064
## 238     2019-08-26              16.28474
## 239     2019-08-27              18.59196
## 240     2019-08-28              18.31122
## 241     2019-08-29              18.42035
## 242     2019-08-30              22.81537
## 243     2019-08-31              33.30987
## 244     2019-09-01              33.60513
## 245     2019-09-02              32.41449
## 246     2019-09-03              16.95762
## 247     2019-09-04              16.81272
## 248     2019-09-05              19.10135
## 249     2019-09-06              18.78087
## 250     2019-09-07              28.99832
## 251     2019-09-08              22.57644
## 252     2019-09-09              18.28225
## 253     2019-09-10              18.26855
## 254     2019-09-11              16.89872
## 255     2019-09-12              16.77426
## 256     2019-09-13              18.85131
## 257     2019-09-14              29.07808
## 258     2019-09-15              25.55102
## 259     2019-09-16              18.37644
## 260     2019-09-17              18.80660
## 261     2019-09-18              18.15554
## 262     2019-09-19              18.23374
## 263     2019-09-20              19.35152
## 264     2019-09-21              27.34413
## 265     2019-09-22              21.71583
## 266     2019-09-23              18.27245
## 267     2019-09-24              17.53694
## 268     2019-09-25              16.46297
## 269     2019-09-26              17.48570
## 270     2019-09-27              15.75319
## 271     2019-09-28              22.04403
## 272     2019-09-29              22.81629
## 273     2019-09-30              18.47392
## 274     2019-10-01              17.49536
## 275     2019-10-02              14.10790
## 276     2019-10-03              17.03366
## 277     2019-10-04              15.51953
## 278     2019-10-05              23.52748
## 279     2019-10-06              26.72655
## 280     2019-10-07              16.80171
## 281     2019-10-08              17.08289
## 282     2019-10-09              16.48035
## 283     2019-10-10              16.14467
## 284     2019-10-11              15.24534
## 285     2019-10-12              23.43263
## 286     2019-10-13              27.71245
## 287     2019-10-14              18.48901
## 288     2019-10-15              15.52133
## 289     2019-10-16              14.20063
## 290     2019-10-17              14.21635
## 291     2019-10-18              17.29351
## 292     2019-10-19              25.95441
## 293     2019-10-20              24.27565
## 294     2019-10-21              14.24772
## 295     2019-10-22              12.27351
## 296     2019-10-23              14.93622
## 297     2019-10-24              14.66240
## 298     2019-10-25              15.10848
## 299     2019-10-26              17.40958
## 300     2019-10-27              26.65632
## 301     2019-10-28              14.93062
## 302     2019-10-29              13.37802
## 303     2019-10-30              13.28634
## 304     2019-10-31              14.73225
## 305     2019-11-01              14.27658
## 306     2019-11-02              19.10275
## 307     2019-11-03              20.76497
## 308     2019-11-04              14.60448
## 309     2019-11-05              13.95757
## 310     2019-11-06              13.40011
## 311     2019-11-07              12.21767
## 312     2019-11-08              13.53587
## 313     2019-11-09              20.95106
## 314     2019-11-10              17.85776
## 315     2019-11-11              12.75153
## 316     2019-11-12              13.03971
## 317     2019-11-13              13.21006
## 318     2019-11-14              14.29909
## 319     2019-11-15              12.88100
## 320     2019-11-16              19.54694
## 321     2019-11-17              15.28140
## 322     2019-11-18              13.89459
## 323     2019-11-19              13.87371
## 324     2019-11-20              13.60412
## 325     2019-11-21              13.14009
## 326     2019-11-22              13.87987
## 327     2019-11-23              17.93453
## 328     2019-11-24              21.05071
## 329     2019-11-25              15.03363
## 330     2019-11-26              20.30509
## 331     2019-11-27              15.60520
## 332     2019-11-28              30.66354
## 333     2019-11-29              22.46235
## 334     2019-11-30              23.38391
## 335     2019-12-01              13.41982
## 336     2019-12-02              12.14463
## 337     2019-12-03              13.19833
## 338     2019-12-04              12.18444
## 339     2019-12-05              13.31259
## 340     2019-12-06              13.91131
## 341     2019-12-07              16.55732
## 342     2019-12-08              17.56003
## 343     2019-12-09              13.01831
## 344     2019-12-10              11.66182
## 345     2019-12-11              11.10670
## 346     2019-12-12              12.62874
## 347     2019-12-13              16.94559
## 348     2019-12-14              15.28016
## 349     2019-12-15              14.97722
## 350     2019-12-16              11.20587
## 351     2019-12-17              10.75739
## 352     2019-12-18              10.40417
## 353     2019-12-19              12.81144
## 354     2019-12-20              14.94906
## 355     2019-12-21              19.05580
## 356     2019-12-22              25.05126
## 357     2019-12-23              14.48553
## 358     2019-12-24              23.58070
## 359     2019-12-25              33.81302
## 360     2019-12-26              24.26599
## 361     2019-12-27              18.43508
## 362     2019-12-28              20.02945
## 363     2019-12-29              25.16288
## 364     2019-12-30               13.67412
## 365     2019-12-31               16.83269
```

Median

```
aggregate(BikeData3$ride_length~ BikeData3$date, FUN = median)
```

```
##     BikeData3$date BikeData3$ride_length
## 1       2019-01-01              9.016667
## 2       2019-01-02              8.166667
## 3       2019-01-03              8.500000
## 4       2019-01-04              9.333333
## 5       2019-01-05             12.983333
## 6       2019-01-06              8.933333
## 7       2019-01-07              8.716667
## 8       2019-01-08              8.583333
## 9       2019-01-09              7.966667
## 10      2019-01-10              7.783333
## 11      2019-01-11              8.200000
## 12      2019-01-12              8.083333
## 13      2019-01-13              8.900000
## 14      2019-01-14              8.391667
## 15      2019-01-15              8.200000
## 16      2019-01-16              8.233333
## 17      2019-01-17              8.200000
## 18      2019-01-18              8.141667
## 19      2019-01-19              9.291667
## 20      2019-01-20              8.741667
## 21      2019-01-21              8.383333
## 22      2019-01-22              8.583333
## 23      2019-01-23              8.416667
## 24      2019-01-24              8.266667
## 25      2019-01-25              7.916667
## 26      2019-01-26              8.033333
## 27      2019-01-27              7.933333
## 28      2019-01-28              9.975000
## 29      2019-01-29              9.683333
## 30      2019-01-30              8.866667
## 31      2019-01-31              9.475000
## 32      2019-02-01              9.266667
## 33      2019-02-02              9.716667
## 34      2019-02-03              9.366667
## 35      2019-02-04              8.716667
## 36      2019-02-05              8.200000
## 37      2019-02-06              8.525000
## 38      2019-02-07              8.333333
## 39      2019-02-08              7.216667
## 40      2019-02-09              7.700000
## 41      2019-02-10              7.816667
## 42      2019-02-11              8.216667
## 43      2019-02-12              8.600000
## 44      2019-02-13              8.250000
## 45      2019-02-14              8.700000
## 46      2019-02-15              7.750000
## 47      2019-02-16              8.358333
## 48      2019-02-17              8.400000
## 49      2019-02-18              8.100000
## 50      2019-02-19              8.116667
## 51      2019-02-20              8.466667
## 52      2019-02-21              8.233333
## 53      2019-02-22              8.466667
## 54      2019-02-23              7.966667
## 55      2019-02-24              7.925000
## 56      2019-02-25              7.683333
## 57      2019-02-26              7.783333
## 58      2019-02-27              8.200000
## 59      2019-02-28              7.866667
## 60      2019-03-01              8.166667
## 61      2019-03-02              8.608333
## 62      2019-03-03              7.650000
## 63      2019-03-04              7.550000
## 64      2019-03-05              7.616667
## 65      2019-03-06              7.900000
## 66      2019-03-07              7.816667
## 67      2019-03-08              8.275000
## 68      2019-03-09              8.700000
## 69      2019-03-10              8.783333
## 70      2019-03-11              8.633333
## 71      2019-03-12              9.400000
## 72      2019-03-13              9.300000
## 73      2019-03-14              9.166667
## 74      2019-03-15              8.466667
## 75      2019-03-16             11.133333
## 76      2019-03-17             12.600000
## 77      2019-03-18              9.133333
## 78      2019-03-19              9.450000
## 79      2019-03-20              8.408333
## 80      2019-03-21              8.750000
## 81      2019-03-22              8.950000
## 82      2019-03-23             13.266667
## 83      2019-03-24              9.983333
## 84      2019-03-25              8.675000
## 85      2019-03-26              9.083333
## 86      2019-03-27             10.366667
## 87      2019-03-28             10.083333
## 88      2019-03-29              9.416667
## 89      2019-03-30              9.341667
## 90      2019-03-31              9.483333
## 91      2019-04-01              9.100000
## 92      2019-04-02              9.175000
## 93      2019-04-03              9.650000
## 94      2019-04-04              8.150000
## 95      2019-04-05              9.533333
## 96      2019-04-06             16.166667
## 97      2019-04-07             14.516667
## 98      2019-04-08             12.300000
## 99      2019-04-09             11.400000
## 100     2019-04-10              8.366667
## 101     2019-04-11              8.933333
## 102     2019-04-12              9.433333
## 103     2019-04-13             14.383333
## 104     2019-04-14              7.066667
## 105     2019-04-15              9.833333
## 106     2019-04-16             11.116667
## 107     2019-04-17             11.266667
## 108     2019-04-18              8.850000
## 109     2019-04-19              8.950000
## 110     2019-04-20             15.916667
## 111     2019-04-21             22.425000
## 112     2019-04-22             13.216667
## 113     2019-04-23             10.083333
## 114     2019-04-24             10.283333
## 115     2019-04-25             10.650000
## 116     2019-04-26             12.058333
## 117     2019-04-27              7.933333
## 118     2019-04-28             13.066667
## 119     2019-04-29              9.200000
## 120     2019-04-30              8.183333
## 121     2019-05-01              9.366667
## 122     2019-05-02              8.700000
## 123     2019-05-03             10.283333
## 124     2019-05-04             16.583333
## 125     2019-05-05             19.966667
## 126     2019-05-06              9.966667
## 127     2019-05-07              8.716667
## 128     2019-05-08             10.125000
## 129     2019-05-09             10.333333
## 130     2019-05-10             10.500000
## 131     2019-05-11             11.066667
## 132     2019-05-12             11.433333
## 133     2019-05-13             10.600000
## 134     2019-05-14             11.416667
## 135     2019-05-15             12.300000
## 136     2019-05-16             11.383333
## 137     2019-05-17              9.266667
## 138     2019-05-18             16.383333
## 139     2019-05-19             15.550000
## 140     2019-05-20             10.350000
## 141     2019-05-21              9.158333
## 142     2019-05-22             13.116667
## 143     2019-05-23             12.758333
## 144     2019-05-24             11.383333
## 145     2019-05-25             20.833333
## 146     2019-05-26             20.616667
## 147     2019-05-27             16.816667
## 148     2019-05-28              9.933333
## 149     2019-05-29             11.350000
## 150     2019-05-30             12.116667
## 151     2019-05-31             13.850000
## 152     2019-06-01             14.500000
## 153     2019-06-02             17.983333
## 154     2019-06-03             11.716667
## 155     2019-06-04             11.950000
## 156     2019-06-05             12.516667
## 157     2019-06-06             11.616667
## 158     2019-06-07             12.816667
## 159     2019-06-08             18.233333
## 160     2019-06-09             14.183333
## 161     2019-06-10             12.783333
## 162     2019-06-11             12.850000
## 163     2019-06-12             10.516667
## 164     2019-06-13             11.216667
## 165     2019-06-14             13.700000
## 166     2019-06-15             13.766667
## 167     2019-06-16             14.050000
## 168     2019-06-17             11.166667
## 169     2019-06-18             12.700000
## 170     2019-06-19             10.516667
## 171     2019-06-20             11.266667
## 172     2019-06-21             12.666667
## 173     2019-06-22             19.500000
## 174     2019-06-23             16.433333
## 175     2019-06-24             12.366667
## 176     2019-06-25             13.716667
## 177     2019-06-26             12.783333
## 178     2019-06-27             12.441667
## 179     2019-06-28             12.650000
## 180     2019-06-29             19.666667
## 181     2019-06-30             18.050000
## 182     2019-07-01             13.283333
## 183     2019-07-02             11.683333
## 184     2019-07-03             12.966667
## 185     2019-07-04             21.083333
## 186     2019-07-05             18.050000
## 187     2019-07-06             21.150000
## 188     2019-07-07             15.666667
## 189     2019-07-08             14.150000
## 190     2019-07-09             13.216667
## 191     2019-07-10             12.333333
## 192     2019-07-11             12.833333
## 193     2019-07-12             14.183333
## 194     2019-07-13             18.733333
## 195     2019-07-14             18.741667
## 196     2019-07-15             12.633333
## 197     2019-07-16             11.650000
## 198     2019-07-17             12.466667
## 199     2019-07-18             12.300000
## 200     2019-07-19             12.416667
## 201     2019-07-20             16.750000
## 202     2019-07-21             16.450000
## 203     2019-07-22             13.708333
## 204     2019-07-23             12.883333
## 205     2019-07-24             13.083333
## 206     2019-07-25             13.400000
## 207     2019-07-26             13.466667
## 208     2019-07-27             19.333333
## 209     2019-07-28             18.000000
## 210     2019-07-29             11.675000
## 211     2019-07-30             13.116667
## 212     2019-07-31             13.083333
## 213     2019-08-01             12.933333
## 214     2019-08-02             14.316667
## 215     2019-08-03             19.583333
## 216     2019-08-04             18.416667
## 217     2019-08-05             13.250000
## 218     2019-08-06            12.783333
## 219     2019-08-07            13.050000
## 220     2019-08-08            13.350000
## 221     2019-08-09            14.100000
## 222     2019-08-10            19.300000
## 223     2019-08-11            17.583333
## 224     2019-08-12            11.166667
## 225     2019-08-13            12.716667
## 226     2019-08-14            12.166667
## 227     2019-08-15            12.516667
## 228     2019-08-16            13.050000
## 229     2019-08-17            16.616667
## 230     2019-08-18            17.050000
## 231     2019-08-19            12.708333
## 232     2019-08-20            11.366667
## 233     2019-08-21            11.950000
## 234     2019-08-22            12.533333
## 235     2019-08-23            12.983333
## 236     2019-08-24            18.866667
## 237     2019-08-25            17.900000
## 238     2019-08-26            10.133333
## 239     2019-08-27            11.866667
## 240     2019-08-28            11.800000
## 241     2019-08-29            12.016667
## 242     2019-08-30            12.883333
## 243     2019-08-31            19.516667
## 244     2019-09-01            22.058333
## 245     2019-09-02            20.550000
## 246     2019-09-03            11.166667
## 247     2019-09-04            11.300000
## 248     2019-09-05            11.733333
## 249     2019-09-06            11.333333
## 250     2019-09-07            17.358333
## 251     2019-09-08            13.283333
## 252     2019-09-09            11.133333
## 253     2019-09-10            11.266667
## 254     2019-09-11            10.933333
## 255     2019-09-12            11.050000
## 256     2019-09-13            11.716667
## 257     2019-09-14            18.250000
## 258     2019-09-15            15.066667
## 259     2019-09-16            11.400000
## 260     2019-09-17            11.316667
## 261     2019-09-18            11.550000
## 262     2019-09-19            11.283333
## 263     2019-09-20            11.933333
## 264     2019-09-21            15.566667
## 265     2019-09-22            12.900000
## 266     2019-09-23            11.541667
## 267     2019-09-24            11.383333
## 268     2019-09-25            11.316667
## 269     2019-09-26            11.316667
## 270     2019-09-27             9.800000
## 271     2019-09-28            12.033333
## 272     2019-09-29            12.500000
## 273     2019-09-30            11.666667
## 274     2019-10-01            11.066667
## 275     2019-10-02             9.183333
## 276     2019-10-03            10.133333
## 277     2019-10-04            10.266667
## 278     2019-10-05            13.633333
## 279     2019-10-06            15.700000
## 280     2019-10-07            11.066667
## 281     2019-10-08            10.866667
## 282     2019-10-09            10.650000
## 283     2019-10-10            10.233333
## 284     2019-10-11             8.716667
## 285     2019-10-12            12.516667
## 286     2019-10-13            13.733333
## 287     2019-10-14            10.666667
## 288     2019-10-15             9.683333
## 289     2019-10-16             9.266667
## 290     2019-10-17             9.566667
## 291     2019-10-18             10.333333
## 292     2019-10-19             15.500000
## 293     2019-10-20             13.116667
## 294     2019-10-21              9.633333
## 295     2019-10-22              8.933333
## 296     2019-10-23              9.666667
## 297     2019-10-24              9.116667
## 298     2019-10-25              9.500000
## 299     2019-10-26              9.466667
## 300     2019-10-27             14.433333
## 301     2019-10-28              9.583333
## 302     2019-10-29              8.833333
## 303     2019-10-30              8.550000
## 304     2019-10-31              8.266667
## 305     2019-11-01              8.966667
## 306     2019-11-02             10.050000
## 307     2019-11-03             10.800000
## 308     2019-11-04              9.533333
## 309     2019-11-05              8.816667
## 310     2019-11-06              8.750000
## 311     2019-11-07              8.283333
## 312     2019-11-08              8.316667
## 313     2019-11-09             10.900000
## 314     2019-11-10             10.216667
## 315     2019-11-11              9.416667
## 316     2019-11-12              8.700000
## 317     2019-11-13              8.383333
## 318     2019-11-14              8.500000
## 319     2019-11-15              8.700000
## 320     2019-11-16             10.158333
## 321     2019-11-17              8.966667
## 322     2019-11-18              8.816667
## 323     2019-11-19              8.616667
## 324     2019-11-20              8.933333
## 325     2019-11-21              8.766667
## 326     2019-11-22              8.550000
## 327     2019-11-23              9.066667
## 328     2019-11-24             10.750000
## 329     2019-11-25              9.150000
## 330     2019-11-26              8.700000
## 331     2019-11-27              7.866667
## 332     2019-11-28             10.200000
## 333     2019-11-29             10.283333
## 334     2019-11-30              8.500000
## 335     2019-12-01              8.450000
## 336     2019-12-02              8.533333
## 337     2019-12-03              8.600000
## 338     2019-12-04              8.766667
## 339     2019-12-05              8.916667
## 340     2019-12-06              8.400000
## 341     2019-12-07              8.933333
## 342     2019-12-08              9.216667
## 343     2019-12-09              8.616667
## 344     2019-12-10              7.883333
## 345     2019-12-11              8.216667
## 346     2019-12-12              8.450000
## 347     2019-12-13              8.950000
## 348     2019-12-14              8.641667
## 349     2019-12-15              8.316667
## 350     2019-12-16              8.316667
## 351     2019-12-17              8.183333
## 352     2019-12-18              7.933333
## 353     2019-12-19              8.366667
## 354     2019-12-20              8.783333
## 355     2019-12-21             10.433333
## 356     2019-12-22             12.150000
## 357     2019-12-23              9.283333
## 358     2019-12-24              9.933333
## 359     2019-12-25             21.833333
## 360     2019-12-26             13.216667
## 361     2019-12-27              9.483333
## 362     2019-12-28             10.400000
## 363     2019-12-29             13.791667
## 364     2019-12-30             8.133333
## 365     2019-12-31             8.200000
```

Mean and median ride length by day of the week:

```
aggregate(BikeData3$ride_length~ BikeData3$day_of_week, FUN = mean)
```

```
##   BikeData3$day_of_week BikeData3$ride_length
## 1                Friday              19.13664
## 2                Monday              18.34697
## 3              Saturday              27.64039
## 4                Sunday              26.68374
## 5              Thursday              17.94416
## 6               Tuesday              17.43314
## 7             Wednesday              17.22274
```

```
aggregate(BikeData3$ride_length~ BikeData3$day_of_week, FUN = median)
```

```
##   BikeData3$day_of_week BikeData3$ride_length
## 1                Friday              11.23333
## 2                Monday              11.10000
## 3              Saturday              16.06667
## 4                Sunday              15.40000
## 5              Thursday              10.95000
## 6               Tuesday              10.76667
## 7             Wednesday              10.81667
```

Insight No 4 - longest rides happens on weekends.

Mean ride length by memberhip type and weekday:

```
aggregate(BikeData3$ride_length~BikeData3$member_casual + BikeData3$day_of_week, FUN = mean)
```

```
##    BikeData3$member_casual BikeData3$day_of_week BikeData3$ride_length
## 1                 Customer                Friday              42.36915
## 2               Subscriber                Friday              12.98043
## 3                 Customer                Monday              42.70953
## 4               Subscriber                Monday              12.95919
## 5                 Customer              Saturday              44.78083
## 6               Subscriber              Saturday              15.22520
## 7                 Customer                Sunday              44.73633
## 8               Subscriber                Sunday              14.69793
## 9                 Customer              Thursday              41.39646
## 10              Subscriber              Thursday              13.06391
## 11                Customer               Tuesday              42.05689
## 12              Subscriber               Tuesday              13.04285
## 13                Customer             Wednesday              40.59286
## 14              Subscriber             Wednesday              12.98135
```

Opposite to above. Mean ride length by first by weekday and then by membership type:

```
aggregate(BikeData3$ride_length~BikeData3$day_of_week + BikeData3$member_casual, FUN = mean)
```

```
##    BikeData3$day_of_week BikeData3$member_casual BikeData3$ride_length
## 1                 Friday                Customer              42.36915
## 2                 Monday                Customer              42.70953
## 3               Saturday                Customer              44.78083
## 4                 Sunday                Customer              44.73633
## 5               Thursday                Customer              41.39646
## 6                Tuesday                Customer              42.05689
## 7              Wednesday                Customer              40.59286
## 8                 Friday              Subscriber              12.98043
## 9                 Monday              Subscriber              12.95919
## 10              Saturday              Subscriber              15.22520
## 11                Sunday              Subscriber              14.69793
## 12              Thursday              Subscriber              13.06391
## 13               Tuesday              Subscriber              13.04285
## 14             Wednesday              Subscriber              12.98135
```

As day of week is sorted alphabetically it could be wise to make order in days and months

```
BikeData4 = BikeData3

BikeData4$day_of_week <- ordered(BikeData4$day_of_week, levels=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday", "Sunday"))
```

Lets check how it looks after weekday is ordered:

```
aggregate(BikeData4$ride_length~BikeData4$member_casual + BikeData4$day_of_week, FUN = mean)
```

```
##    BikeData4$member_casual BikeData4$day_of_week BikeData4$ride_length
## 1                 Customer                Monday              42.70953
## 2               Subscriber                Monday              12.95919
## 3                 Customer               Tuesday              42.05689
## 4               Subscriber               Tuesday              13.04285
## 5                 Customer             Wednesday              40.59286
## 6               Subscriber             Wednesday              12.98135
## 7                 Customer              Thursday              41.39646
## 8               Subscriber              Thursday              13.06391
## 9                 Customer                Friday              42.36915
## 10              Subscriber                Friday              12.98043
## 11                Customer              Saturday              44.78083
## 12              Subscriber              Saturday              15.22520
## 13                Customer                Sunday              44.73633
## 14              Subscriber                Sunday              14.69793
```

```
aggregate(BikeData4$ride_length~BikeData4$day_of_week + BikeData4$member_casual, FUN = mean)
```

```
##    BikeData4$day_of_week BikeData4$member_casual BikeData4$ride_length
## 1                 Monday                Customer              42.70953
## 2                Tuesday                Customer              42.05689
## 3              Wednesday                Customer              40.59286
## 4               Thursday                Customer              41.39646
## 5                 Friday                Customer              42.36915
## 6               Saturday                Customer              44.78083
## 7                 Sunday                Customer              44.73633
## 8                 Monday              Subscriber              12.95919
## 9                Tuesday              Subscriber              13.04285
## 10             Wednesday              Subscriber              12.98135
## 11              Thursday              Subscriber              13.06391
## 12                Friday              Subscriber              12.98043
## 13              Saturday              Subscriber              15.22520
## 14                Sunday              Subscriber              14.69793
```

Analysing ride_length by membership and month of the year.

```
aggregate(BikeData4$ride_length~BikeData4$month + BikeData4$member_casual, FUN = mean)
```

```
##    BikeData4$month BikeData4$member_casual BikeData4$ride_length
## 1                1                Customer              40.94205
## 2                2                Customer              47.74435
## 3                3                Customer              42.91236
## 4                4                Customer              44.54457
## 5                5                Customer              45.85299
## 6                6                Customer              43.54886
## 7                7                Customer              43.09299
## 8                8                Customer              42.96728
## 9                9                Customer              41.04822
## 10              10                Customer              40.63957
## 11              11                Customer              45.35062
## 12              12                Customer              48.09819
## 13               1              Subscriber              13.17021
## 14               2              Subscriber              11.66227
## 15               3              Subscriber              11.67819
## 16               4              Subscriber              12.84217
## 17               5              Subscriber              13.58394
## 18               6              Subscriber              14.32693
## 19               7              Subscriber              14.81058
## 20               8              Subscriber              14.24093
## 21               9              Subscriber              13.50513
## 22              10              Subscriber              12.30207
## 23              11              Subscriber              12.11288
## 24              12              Subscriber              11.41516
```

Insight no 5.

for subscribers longest ridership occurs in summer months, which seems to be very realistic. For customers ride length looks more scattered, no pattern can be dedected. Check and ivestigate signifficant difference in ride_ duration_length between customers and subscribers.

Using pipe (%<%) in R. Analyze ridership data by type and weekday

```
BikeData4 %>%
  group_by(member_casual, day_of_week) %>%  #groups by columns
  summarise(number_of_rides = n()                          #calc no of rides and avg ride
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, day_of_week)
```

```
## `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 day_of_week number_of_rides average_duration
##    <chr>         <ord>                 <int>            <dbl>
##  1 Customer      Monday               101456             42.7
##  2 Customer      Tuesday               88615             42.1
##  3 Customer      Wednesday             89703             40.6
##  4 Customer      Thursday             101321             41.4
##  5 Customer      Friday               121085             42.4
##  6 Customer      Saturday             207991             44.8
##  7 Customer      Sunday               170119             44.7
##  8 Subscriber    Monday               458766             13.0
##  9 Subscriber    Tuesday              497013             13.0
## 10 Subscriber    Wednesday            494265             13.0
## 11 Subscriber    Thursday             486904             13.1
## 12 Subscriber    Friday               456954             13.0
## 13 Subscriber    Saturday             287153             15.2
## 14 Subscriber    Sunday               256227             14.7
```

Using pipe (%<%) in R. Analyze ridership data by type and month

```
BikeData4 %>%
  group_by(member_casual, month) %>%  #groups by columns
  summarise(number_of_rides = n()                          #calc no of rides and avg ride
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, month)
```

```
## `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 average_duration
##    <chr>         <int>           <int>            <dbl>
##  1 Customer          1            4601             40.9
##  2 Customer          2            2636             47.7
##  3 Customer          3           15917             42.9
##  4 Customer          4           47735             44.5
##  5 Customer          5           81607             45.9
##  6 Customer          6          130201             43.5
##  7 Customer          7          175567             43.1
##  8 Customer          8          186777             43.0
##  9 Customer          9          129121             41.0
## 10 Customer         10           71000             40.6
## # ... with 14 more rows
```

Using pipe (%<%) in R. Reverse of previous. Analyze ridership data by month and then by membership type

```
BikeData4 %>%
  group_by(month, member_casual) %>%  #groups by columns
  summarise(number_of_rides = n()                          #calc no of rides and avg ride
            ,average_duration = mean(ride_length)) %>%
  arrange(month, member_casual)
```

```
## `summarise()` has grouped output by 'month'. You can override using the `.groups` argument.
## # A tibble: 24 x 4
## # Groups:   month [12]
##    month member_casual number_of_rides average_duration
##    <int> <chr>                   <int>            <dbl>
##  1     1 Customer                 4601             40.9
##  2     1 Subscriber              98661             13.2
##  3     2 Customer                 2636             47.7
##  4     2 Subscriber              93543             11.7
##  5     3 Customer                15917             42.9
##  6     3 Subscriber             149681             11.7
##  7     4 Customer                47735             44.5
##  8     4 Subscriber             217561             12.8
##  9     5 Customer                81607             45.9
## 10     5 Subscriber             285832             13.6
## # ... with 14 more rows
```

Export for further analysis and visualisations in SQL/Tableau/DataViz etc

write.csv(BikeData4, file = ‘E:/Tomasz/CapstoneGoogle/capstone_bikedata_for_analysis.csv’)

Comment made to avoid consecutive exporting of this large file. Uncomment if required and copy cody to a chunk below