

How does a bike-share navigate speedy success?

Eric L Sammons

2023-07-14

Setup

Install required packages

Load required libraries

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v purrr      1.0.1
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.2      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::between()      masks data.table::between()
## x dplyr::filter()       masks stats::filter()
## x dplyr::first()        masks data.table::first()
## x lubridate::hour()     masks data.table::hour()
## x lubridate::isoweek()  masks data.table::isoweek()
## x dplyr::lag()          masks stats::lag()
## x dplyr::last()         masks data.table::last()
## x lubridate::mday()     masks data.table::mday()
## x lubridate::minute()   masks data.table::minute()
## x lubridate::month()    masks data.table::month()
## x lubridate::quarter()  masks data.table::quarter()
## x lubridate::second()   masks data.table::second()
## x purrr::transpose()    masks data.table::transpose()
## x lubridate::wday()     masks data.table::wday()
## x lubridate::week()     masks data.table::week()
## x lubridate::yday()     masks data.table::yday()
## x lubridate::year()     masks data.table::year()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Set variables

Functions required to download and validate data files

Acquire the data files

Validate the datafiles and create a single dataframe

```
## Rows: 634858 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dtm  (2): started_at, ended_at
```

```

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 769204 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 823488 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 785932 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 558685 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 337735 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 181806 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at

```

```

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 190301 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 190445 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 258678 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 426590 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 604827 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```

Housekeeping

Remove columns not required

```

## [1] "ride_id"           "rideable_type"      "started_at"
## [4] "ended_at"          "start_station_name" "start_station_id"

```

```
## [7] "end_station_name" "end_station_id" "member_casual"
```

Rider type summary

```
##
## casual member
## 2295791 3466758
```

Data transformations

Break down the started_at column

Create ride_length column

```
## tibble [5,762,549 x 15] (S3: tbl_df/tbl/data.frame)
## $ ride_id      : chr [1:5762549] "EC2DE40644C6B0F4" "1C31AD03897EE385" "1542FBEC830415CF" "6FF..."
## $ rideable_type : chr [1:5762549] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...
## $ started_at    : POSIXct[1:5762549], format: "2022-05-23 23:06:58" "2022-05-11 08:53:28" ...
## $ ended_at      : POSIXct[1:5762549], format: "2022-05-23 23:40:19" "2022-05-11 09:31:22" ...
## $ start_station_name: chr [1:5762549] "Wabash Ave & Grand Ave" "DuSable Lake Shore Dr & Monroe St" ...
## $ start_station_id : chr [1:5762549] "TA1307000117" "13300" "TA1305000032" "TA1305000032" ...
## $ end_station_name : chr [1:5762549] "Halsted St & Roscoe St" "Field Blvd & South Water St" "Wood S..."
## $ end_station_id   : chr [1:5762549] "TA1309000025" "15534" "13221" "TA1305000030" ...
## $ member_casual    : chr [1:5762549] "member" "member" "member" "member" ...
## $ date             : Date[1:5762549], format: "2022-05-23" "2022-05-11" ...
## $ month            : chr [1:5762549] "05" "05" "05" "05" ...
## $ day              : chr [1:5762549] "23" "11" "26" "10" ...
## $ year             : chr [1:5762549] "2022" "2022" "2022" "2022" ...
## $ day_of_week       : chr [1:5762549] "Monday" "Wednesday" "Thursday" "Tuesday" ...
## $ ride_length      : num [1:5762549] 2001 2274 1310 522 301 ...
```

Remove rows with ride_length < 0

Analysis

Ride length summary

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         0      337      599    1134    1078 2483235
```

Ride length summary member vs casual

```
comparison <- aggregate(df$ride_length, by = list(df$member_casual), FUN = function(x)
  c(mean = mean(x),
    median = median(x),
    max = max(x),
    min = min(x)))
```

comparison

```
##      Group.1      x.mean      x.median      x.max      x.min
## 1 casual    1713.7843    756.0000 2483235.0000    0.0000
## 2 member     750.1513    519.0000  93580.0000    0.0000

##      Group.1      x.min      x.mean      x.median      x.max
## 1 casual     0.0000    1713.7843    756.0000 2483235.0000
## 2 member     0.0000     750.1513    519.0000  93580.0000
```