```
R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> install.packages("tidyverse")  # For data manipulation
Installing package into 'C:/Users/Neha/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
Error in contrib.url(repos, "source") :
  trying to use CRAN without setting a mirror
> install.packages("tidyverse")  # For data manipulation
Installing package into 'C:/Users/Neha/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'https://mirror.niser.ac.in/cran/bin/windows/contrib/4.4/tidyverse_2.0.0.zip'
Content type 'application/zip' length 431663 bytes (421 KB)
downloaded 421 KB

package 'tidyverse' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Neha\AppData\Local\Temp\RtmpKwCP7S\downloaded_packages
> library(tidyverse)
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become e
rrors
>
> setwdsetwd("path_to_your_CSV_folder")  # Replace with your folder path
Error in setwdsetwd("path_to_your_CSV_folder") :
  could not find function "setwdsetwd"
> setwd("C:\Users\Neha\Desktop\Cyclistic_Data")  #Replace with your folder path
Error: '\U' used without hex digits in character string (<input>:1:11)
> setwd("G:\Cyclistic_Data")
Error: '\C' is an unrecognized escape in character string (<input>:1:11)
> file.choose(C:\Users\Neha\Desktop\Cyclistic_Data)
Error: unexpected symbol in "file.choose(C:\Users"
> file.choose("C:\Users\Neha\Desktop\Cyclistic_Data")
Error: '\U' used without hex digits in character string (<input>:1:17)
> setwd("C:/Users/YourUsername/Documents/Cyclistic_Data")
Error in setwd("C:/Users/YourUsername/Documents/Cyclistic_Data") :
  cannot change working directory
> getwd("G:\Cyclistic_Data")
Error: '\C' is an unrecognized escape in character string (<input>:1:11)
> data_path <- dirname(file.choose())
> setwd(data_path)
```

```
> # Get list of all CSV files in the folder
> file_list <- list.files(pattern = "*.csv")
>
> # Read and combine all files into one dataframe
> combined_data <- file_list %>%
+   map_df(~ read_csv(.))
[1mindexing[0m [34m202101-divvy-tripdata.csv[0m [=----------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202101-divvy-tripdata.csv[0m [======================]
 [32m99.33MB/s[0m, eta: [36m 0s[0m
                                                    Rows: 96834 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
dttm (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.
[1mindexing[0m [34m202102-divvy-tripdata.csv[0m [==----------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202102-divvy-tripdata.csv[0m [======================]
[32m102.38MB/s[0m, eta: [36m 0s[0m
                                                    Rows: 49622 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
dttm (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.
[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [======================]
 [32m99.96MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [=====
=====----------] [32m99.98MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-trip
data.csv[0m [=====================] [32m100.42MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [3
4m202103-divvy-tripdata.csv[0m [==========----------] [32m100.28MB/s[0m, eta: [36m 0s[0m
[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [==========----------] [32m100.35MB/s[0m,
eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [====================] [
32m100.39MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [=======
==========----------] [32m99.53MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripda
ta.csv[0m [==================-----] [32m99.74MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m
202103-divvy-tripdata.csv[0m [====================----] [32m100.08MB/s[0m, eta: [36m 0s[0m[1
mindexing[0m [34m202103-divvy-tripdata.csv[0m [====================---] [32m99.46MB/s[0m, et
a: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [====================---] [3
2m99.54MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata.csv[0m [========
==========--] [32m99.37MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202103-divvy-tripdata
.csv[0m [=====================-] [32m99.83MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2103-divvy-tripdata.csv[0m [======================] [32m102.32MB/s[0m, eta: [36m 0s[0m

             Rows: 228496 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
dttm (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.
[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [=======-------------]
[32m111.53MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [=====
=====-------------] [32m111.37MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-trip
data.csv[0m [=========-----------] [32m111.07MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [3
4m202104-divvy-tripdata.csv[0m [==========----------] [32m110.80MB/s[0m, eta: [36m 0s[0m
[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [==========----------] [32m111.28MB/s[0m,
eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [====================] [
32m111.21MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [=======
==========----------] [32m110.90MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripda
ta.csv[0m [====================] [32m111.39MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m
202104-divvy-tripdata.csv[0m [==================-----] [32m111.18MB/s[0m, eta: [36m 0s[0m[1
```

```
mindexing[0m [34m202104-divvy-tripdata.csv[0m [==================------] [32m111.31MB/s[0m, et
a: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [====================] [32
m111.17MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [=========
============] [32m112.92MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata
.csv[0m [====================] [32m110.57MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2104-divvy-tripdata.csv[0m [====================] [32m109.48MB/s[0m, eta: [36m 0s[0m[1mi
ndexing[0m [34m202104-divvy-tripdata.csv[0m [====================--] [32m110.86MB/s[0m, eta:
 [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [====================--] [32m1
09.33MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [==========
==========--] [32m110.79MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.c
sv[0m [====================-] [32m109.43MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m2021
04-divvy-tripdata.csv[0m [====================-] [32m110.68MB/s[0m, eta: [36m 0s[0m[1mind
exing[0m [34m202104-divvy-tripdata.csv[0m [====================] [32m109.82MB/s[0m, eta: [
[36m 0s[0m[1mindexing[0m [34m202104-divvy-tripdata.csv[0m [====================] [32m111
.94MB/s[0m, eta: [36m 0s[0m
                                        Rows: 337230 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
dttm (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.
[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [====================]
[32m111.83MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [=====
=----------------] [32m109.75MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202105-divvy-trip
data.csv[0m [=====================] [32m109.60MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [3
4m202105-divvy-tripdata.csv[0m [=====================] [32m108.88MB/s[0m, eta: [36m 1s[0m
[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [========-------------] [32m107.88MB/s[0m,
eta: [36m 1s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [=========-----------] [
32m108.54MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [=======
====-----------] [32m108.67MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripda
ta.csv[0m [===========----------] [32m108.25MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m
202105-divvy-tripdata.csv[0m [============---------] [32m107.58MB/s[0m, eta: [36m 0s[0m[1
mindexing[0m [34m202105-divvy-tripdata.csv[0m [==============-------] [32m107.44MB/s[0m, et
a: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [===============-------] [32
m106.79MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [========
=============-------] [32m106.64MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata
.csv[0m [==================-----] [32m105.95MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2105-divvy-tripdata.csv[0m [==================-----] [32m105.44MB/s[0m, eta: [36m 0s[0m[1mi
ndexing[0m [34m202105-divvy-tripdata.csv[0m [==================----] [32m105.40MB/s[0m, eta:
 [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [====================---] [32m1
05.44MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata.csv[0m [==========
==========--] [32m105.53MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202105-divvy-tripdata.c
sv[0m [====================--] [32m105.50MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m2021
05-divvy-tripdata.csv[0m [====================-] [32m105.53MB/s[0m, eta: [36m 0s[0m[1mind
exing[0m [34m202105-divvy-tripdata.csv[0m [====================] [32m108.56MB/s[0m, eta: [
[36m 0s[0m
                                        Rows: 531633 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
dttm (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.
[1mindexing[0m [34m202106-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202106-divvy-tripdata.csv[0m [====================]
[32m110.69MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202106-divvy-tripdata.csv[0m [=====
----------------] [32m108.18MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202106-divvy-trip
data.csv[0m [=====================] [32m108.81MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [3
4m202106-divvy-tripdata.csv[0m [=====================] [32m108.18MB/s[0m, eta: [36m 1s[0m
[1mindexing[0m [34m202106-divvy-tripdata.csv[0m [======================] [32m109.88MB/s[0m,
eta: [36m 1s[0m[1mindexing[0m [34m202106-divvy-tripdata.csv[0m [=======------------] [
32m109.81MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202106-divvy-tripdata.csv[0m [=======
=============] [32m109.21MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202106-divvy-tripda
ta.csv[0m [=========----------] [32m109.50MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m
```

```
202106-divvy-tripdata.csv [0m [======================] [32m109.10MB/s [0m, eta: [36m 1s [0m [1
mindexing [0m [34m202106-divvy-tripdata.csv [0m [======================] [32m108.84MB/s [0m, et
a: [36m 1s [0m [1mindexing [0m [34m202106-divvy-tripdata.csv [0m [======================] [32
m108.31MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202106-divvy-tripdata.csv [0m [========
==========] [32m108.46MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202106-divvy-tripdata
.csv [0m [======================] [32m109.57MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m20
2106-divvy-tripdata.csv [0m [======================] [32m109.90MB/s [0m, eta: [36m 0s [0m [1mi
ndexing [0m [34m202106-divvy-tripdata.csv [0m [======================] [32m110.15MB/s [0m, eta:
 [36m 0s [0m [1mindexing [0m [34m202106-divvy-tripdata.csv [0m [======================] [32m1
09.78MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202106-divvy-tripdata.csv [0m [========
==========] [32m109.52MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202106-divvy-tripdata.c
sv [0m [======================] [32m109.65MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m2021
06-divvy-tripdata.csv [0m [======================] [32m109.85MB/s [0m, eta: [36m 0s [0m [1mind
exing [0m [34m202106-divvy-tripdata.csv [0m [======================] [32m110.06MB/s [0m, eta:
[36m 0s [0m [1mindexing [0m [34m202106-divvy-tripdata.csv [0m [======================] [32m109
.87MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202106-divvy-tripdata.csv [0m [========
==========] [32m112.97MB/s [0m, eta: [36m 0s [0m
                                                                    Rows: 729595 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
dttm (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.
[1mindexing [0m [34m202107-divvy-tripdata.csv [0m [----------------------] [32m2.15GB/s [0m
, eta: [36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [======================]
[32m108.73MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [=====
----------------] [32m106.51MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-trip
data.csv [0m [======================] [32m105.80MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [3
4m202107-divvy-tripdata.csv [0m [=====----------------] [32m104.85MB/s [0m, eta: [36m 1s [0m
[1mindexing [0m [34m202107-divvy-tripdata.csv [0m [======---------------] [32m103.91MB/s [0m,
eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [=======--------------] [
32m104.02MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [=======
==-------------] [32m104.14MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripda
ta.csv [0m [=========----------] [32m104.21MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m
202107-divvy-tripdata.csv [0m [======================] [32m103.46MB/s [0m, eta: [36m 1s [0m [1
mindexing [0m [34m202107-divvy-tripdata.csv [0m [==========----------] [32m102.99MB/s [0m, et
a: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [==========---------] [32
m102.66MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [========
==========] [32m102.67MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202107-divvy-tripdata
.csv [0m [======================] [32m102.15MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m20
2107-divvy-tripdata.csv [0m [======================] [32m101.76MB/s [0m, eta: [36m 0s [0m [1mi
ndexing [0m [34m202107-divvy-tripdata.csv [0m [======================] [32m101.66MB/s [0m, eta:
 [36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [======================] [32m1
01.58MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [========
==========] [32m101.60MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.c
sv [0m [======================] [32m100.56MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m2021
07-divvy-tripdata.csv [0m [======================] [32m99.58MB/s [0m, eta: [36m 0s [0m [1mind
exing [0m [34m202107-divvy-tripdata.csv [0m [======================] [32m99.75MB/s [0m, eta:
[36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [======================] [32m99
.55MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv [0m [========
==========] [32m99.68MB/s [0m, eta: [36m 0s [0m [1mindexing [0m [34m202107-divvy-tripdata.csv
[0m [======================] [32m102.52MB/s [0m, eta: [36m 0s [0m
                                                                    Rows: 822
410 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
dttm (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.
[1mindexing [0m [34m202108-divvy-tripdata.csv [0m [----------------------] [32m2.15GB/s [0m
, eta: [36m 0s [0m [1mindexing [0m [34m202108-divvy-tripdata.csv [0m [======================]
[32m116.62MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202108-divvy-tripdata.csv [0m [=====
----------------] [32m116.43MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [34m202108-divvy-trip
data.csv [0m [======================] [32m116.93MB/s [0m, eta: [36m 1s [0m [1mindexing [0m [3
```

```
4m202108-divvy-tripdata.csv[0m [======----------------] [32m116.07MB/s[0m, eta: [36m 1s[0m
[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [=======---------------] [32m115.66MB/s[0m,
eta: [36m 1s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [========----------------] [
32m115.26MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [=======
==--------------] [32m114.09MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202108-divvy-tripda
ta.csv[0m [=========----------] [32m113.43MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m
202108-divvy-tripdata.csv[0m [==========---------] [32m114.21MB/s[0m, eta: [36m 1s[0m[1
mindexing[0m [34m202108-divvy-tripdata.csv[0m [===========----------] [32m115.82MB/s[0m, et
a: [36m 1s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [============---------] [32
m113.44MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [=========
====---------] [32m112.71MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202108-divvy-tripdata
.csv[0m [==============--------] [32m112.50MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2108-divvy-tripdata.csv[0m [===============-------] [32m112.11MB/s[0m, eta: [36m 0s[0m[1mi
ndexing[0m [34m202108-divvy-tripdata.csv[0m [================------] [32m111.15MB/s[0m, eta:
 [36m 0s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [=================-----] [32m1
11.39MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [==========
=========----] [32m111.85MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202108-divvy-tripdata.c
sv[0m [====================---] [32m111.95MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m2021
08-divvy-tripdata.csv[0m [====================--] [32m111.72MB/s[0m, eta: [36m 0s[0m[1mind
exing[0m [34m202108-divvy-tripdata.csv[0m [===================--] [32m111.49MB/s[0m, eta: [
[36m 0s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [====================-] [32m110
.75MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202108-divvy-tripdata.csv[0m [==========
=========] [32m113.90MB/s[0m, eta: [36m 0s[0m
                                                                    Rows: 804352 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
dttm (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.
[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [=====-------------------]
[32m114.06MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [=====
----------------] [32m112.75MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-trip
data.csv[0m [======------------------] [32m111.75MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [3
4m202109-divvy-tripdata.csv[0m [=======----------------] [32m110.68MB/s[0m, eta: [36m 1s[0m
[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [========---------------] [32m108.95MB/s[0m,
eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [=========--------------] [
32m107.97MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [=======
==--------------] [32m107.01MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripda
ta.csv[0m [==========-------------] [32m106.36MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m
202109-divvy-tripdata.csv[0m [==========-------------] [32m105.60MB/s[0m, eta: [36m 1s[0m[1
mindexing[0m [34m202109-divvy-tripdata.csv[0m [===========------------] [32m105.58MB/s[0m, et
a: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [============-----------] [32
m105.76MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [=========
====----------] [32m106.12MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202109-divvy-tripdata
.csv[0m [==============--------] [32m106.41MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2109-divvy-tripdata.csv[0m [===============-------] [32m106.87MB/s[0m, eta: [36m 0s[0m[1mi
ndexing[0m [34m202109-divvy-tripdata.csv[0m [================------] [32m107.70MB/s[0m, eta:
 [36m 0s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [=================-----] [32m1
08.23MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [==========
==========----] [32m108.30MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202109-divvy-tripdata.c
sv[0m [====================---] [32m108.51MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m2021
09-divvy-tripdata.csv[0m [====================--] [32m108.57MB/s[0m, eta: [36m 0s[0m[1mind
exing[0m [34m202109-divvy-tripdata.csv[0m [===================--] [32m108.59MB/s[0m, eta: [
[36m 0s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [====================-] [32m108
.69MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202109-divvy-tripdata.csv[0m [==========
=========] [32m111.68MB/s[0m, eta: [36m 0s[0m
                                                                    Rows: 756147 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
dttm (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.
[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
```

, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [======================]
[32m120.84MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [=====
=----------------] [32m122.44MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202110-divvy-trip
data.csv[0m [======================] [32m123.23MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [3
4m202110-divvy-tripdata.csv[0m [======================] [32m122.47MB/s[0m, eta: [36m 1s[0m
[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [=======--------------] [32m122.01MB/s[0m,
eta: [36m 1s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [=======================] [
32m120.29MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [=======
==--------------] [32m119.02MB/s[0m, eta: [36m 1s[0m[1mindexing[0m [34m202110-divvy-tripda
ta.csv[0m [==========-----------] [32m118.42MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m
202110-divvy-tripdata.csv[0m [==========-----------] [32m118.13MB/s[0m, eta: [36m 0s[0m[1
mindexing[0m [34m202110-divvy-tripdata.csv[0m [=========----------] [32m118.31MB/s[0m, et
a: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [============---------] [32
m118.18MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [========
=====--------] [32m118.32MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata
.csv[0m [================-------] [32m118.44MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2110-divvy-tripdata.csv[0m [================------] [32m118.19MB/s[0m, eta: [36m 0s[0m[1mi
ndexing[0m [34m202110-divvy-tripdata.csv[0m [==================-----] [32m118.27MB/s[0m, eta:
[36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [====================----] [32m1
19.74MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [==========
==========----] [32m118.03MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.c
sv[0m [====================----] [32m118.79MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m2021
10-divvy-tripdata.csv[0m [==================---] [32m117.56MB/s[0m, eta: [36m 0s[0m[1mind
exing[0m [34m202110-divvy-tripdata.csv[0m [====================---] [32m118.73MB/s[0m, eta:
[36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [==================--] [32m117
.58MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [============
========--] [32m118.96MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv
[0m [====================--] [32m117.93MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202110
-divvy-tripdata.csv[0m [====================-] [32m117.98MB/s[0m, eta: [36m 0s[0m[1mindex
ing[0m [34m202110-divvy-tripdata.csv[0m [====================-] [32m116.17MB/s[0m, eta: [3
6m 0s[0m[1mindexing[0m [34m202110-divvy-tripdata.csv[0m [====================] [32m118.8
4MB/s[0m, eta: [36m 0s[0m

Rows: 631226 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
dttm (2): started_at, ended_at

ⓘ Use `spec()` to retrieve the full column specification for this data.
ⓘ Specify the column types or set `show_col_types = FALSE` to quiet this message.

[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [----------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [=======--------------]
[32m114.96MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [=====
====-------------] [32m113.49MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-trip
data.csv[0m [=========------------] [32m112.69MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [3
4m202111-divvy-tripdata.csv[0m [==========-----------] [32m112.50MB/s[0m, eta: [36m 0s[0m
[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [==========----------] [32m112.44MB/s[0m,
eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [==========----------] [
32m112.46MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [=======
=====---------] [32m112.24MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripda
ta.csv[0m [==========----------] [32m111.80MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m
202111-divvy-tripdata.csv[0m [==========----------] [32m111.29MB/s[0m, eta: [36m 0s[0m[1
mindexing[0m [34m202111-divvy-tripdata.csv[0m [==========------] [32m111.22MB/s[0m, et
a: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [==========----------] [32
m111.20MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [========
=====---------] [32m111.39MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata
.csv[0m [==========---] [32m111.67MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m20
2111-divvy-tripdata.csv[0m [==========--] [32m111.17MB/s[0m, eta: [36m 0s[0m[1mi
ndexing[0m [34m202111-divvy-tripdata.csv[0m [==========--] [32m111.47MB/s[0m, eta:
[36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [==========-] [32m1
11.60MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202111-divvy-tripdata.csv[0m [==========
==========] [32m114.43MB/s[0m, eta: [36m 0s[0m

Rows: 359978 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
dttm (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.
[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [========================]
[32m112.30MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [=====
==================] [32m112.49MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-trip
data.csv[0m [========================] [32m111.25MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [3
4m202112-divvy-tripdata.csv[0m [========================] [32m110.55MB/s[0m, eta: [36m 0s[0m
[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [================--------] [32m110.08MB/s[0m,
eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [==================------] [
32m110.51MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [=======
==============----] [32m110.50MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripda
ta.csv[0m [====================----] [32m108.42MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m
202112-divvy-tripdata.csv[0m [====================---] [32m107.91MB/s[0m, eta: [36m 0s[0m[1
mindexing[0m [34m202112-divvy-tripdata.csv[0m [=====================--] [32m107.40MB/s[0m, et
a: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [====================--] [32
m107.35MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata.csv[0m [=========
============-] [32m107.36MB/s[0m, eta: [36m 0s[0m[1mindexing[0m [34m202112-divvy-tripdata
.csv[0m [========================] [32m110.33MB/s[0m, eta: [36m 0s[0m
                                                                                Rows:
 247540 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
dttm (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.
[1mindexing[0m [34m202201-divvy-tripdata.csv[0m [=------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202201-divvy-tripdata.csv[0m [========================]
[32m119.79MB/s[0m, eta: [36m 0s[0m
                                                    Rows: 103770 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
dttm (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.
[1mindexing[0m [34m202202-divvy-tripdata.csv[0m [=------------------------] [32m2.15GB/s[0m
, eta: [36m 0s[0m[1mindexing[0m [34m202202-divvy-tripdata.csv[0m [========================]
[32m113.12MB/s[0m, eta: [36m 0s[0m
                                                    Rows: 115609 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
dttm (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.
> # Remove rows with missing critical columns
> cleaned_data <- combined_data %>%
+   drop_na(ride_id, started_at, ended_at, member_casual)
>
> # Convert time columns to date-time format
> cleaned_data <- cleaned_data %>%
+   mutate(
+     started_at = as.POSIXct(started_at, format = "%Y-%m-%d %H:%M:%S"),
+     ended_at = as.POSIXct(ended_at, format = "%Y-%m-%d %H:%M:%S")
+   )
>
> # Calculate ride length in minutes
> cleaned_data <- cleaned_data %>%
+   mutate(ride_length = as.numeric(difftime(ended_at, started_at, units = "mins")))
>
> # Remove negative and outlier ride lengths
> cleaned_data <- cleaned_data %>%
```

```
+   filter(ride_length > 0, ride_length <= 1440)  # Max 24 hours
>
> # Add a column for the day of the week (1 = Sunday, 7 = Saturday)
> cleaned_data <- cleaned_data %>%
+   mutate(day_of_week = wday(started_at))
> write_csv(cleaned_data, "Cleaned_Cyclistic_Data.csv")
```

[1mwrote[0m [32m12.24MB[0m in [36m 0s[0m, [32m816.14MB/s[0m[1mwrote[0m [32m48.62MB[0m in [36m 0s[0m, [32m231.33MB/s[0m[1mwrote[0m [32m60.72MB[0m in [36m 0s[0m, [32m216.47MB/s[0m[1mwrote[0m [32m73.02MB[0m in [36m 0s[0m, [32m210.05MB/s[0m[1mwrote[0m [32m84.20MB[0m in [36m 0s[0m, [32m200.22MB/s[0m[1mwrote[0m [32m95.56MB[0m in [36m 0s[0m, [32m193.43MB/s[0m[1mwrote[0m [32m107.89MB[0m in [36m 1s[0m, [32m193.55MB/s[0m[1mwrote[0m [32m120.12MB[0m in [36m 1s[0m, [32m187.09MB/s[0m[1mwrote[0m [32m132.45MB[0m in [36m 1s[0m, [32m185.99MB/s[0m[1mwrote[0m [32m144.81MB[0m in [36m 1s[0m, [32m186.34MB/s[0m[1mwrote[0m [32m156.95MB[0m in [36m 1s[0m, [32m186.39MB/s[0m[1mwrote[0m [32m169.22MB[0m in [36m 1s[0m, [32m182.17MB/s[0m[1mwrote[0m [32m181.56MB[0m in [36m 1s[0m, [32m179.37MB/s[0m[1mwrote[0m [32m193.89MB[0m in [36m 1s[0m, [32m181.25MB/s[0m[1mwrote[0m [32m205.63MB[0m in [36m 1s[0m, [32m177.91MB/s[0m[1mwrote[0m [32m218.01MB[0m in [36m 1s[0m, [32m172.19MB/s[0m[1mwrote[0m [32m230.19MB[0m in [36m 1s[0m, [32m169.38MB/s[0m[1mwrote[0m [32m242.51MB[0m in [36m 1s[0m, [32m167.41MB/s[0m[1mwrote[0m [32m253.80MB[0m in [36m 2s[0m, [32m168.80MB/s[0m[1mwrote[0m [32m263.96MB[0m in [36m 2s[0m, [32m167.49MB/s[0m[1mwrote[0m [32m276.29MB[0m in [36m 2s[0m, [32m166.99MB/s[0m[1mwrote[0m [32m288.63MB[0m in [36m 2s[0m, [32m165.00MB/s[0m[1mwrote[0m [32m300.90MB[0m in [36m 2s[0m, [32m159.49MB/s[0m[1mwrote[0m [32m313.14MB[0m in [36m 2s[0m, [32m157.57MB/s[0m[1mwrote[0m [32m325.50MB[0m in [36m 2s[0m, [32m157.19MB/s[0m[1mwrote[0m [32m336.89MB[0m in [36m 2s[0m, [32m158.49MB/s[0m[1mwrote[0m [32m347.97MB[0m in [36m 2s[0m, [32m155.66MB/s[0m[1mwrote[0m [32m360.21MB[0m in [36m 2s[0m, [32m155.93MB/s[0m[1mwrote[0m [32m372.45MB[0m in [36m 2s[0m, [32m155.31MB/s[0m[1mwrote[0m [32m384.77MB[0m in [36m 2s[0m, [32m156.20MB/s[0m[1mwrote[0m [32m397.02MB[0m in [36m 3s[0m, [32m156.35MB/s[0m[1mwrote[0m [32m407.55MB[0m in [36m 3s[0m, [32m155.86MB/s[0m[1mwrote[0m [32m419.82MB[0m in [36m 3s[0m, [32m156.28MB/s[0m[1mwrote[0m [32m432.23MB[0m in [36m 3s[0m, [32m156.09MB/s[0m[1mwrote[0m [32m444.40MB[0m in [36m 3s[0m, [32m156.67MB/s[0m[1mwrote[0m [32m456.70MB[0m in [36m 3s[0m, [32m156.23MB/s[0m[1mwrote[0m [32m468.99MB[0m in [36m 3s[0m, [32m156.63MB/s[0m[1mwrote[0m [32m481.34MB[0m in [36m 3s[0m, [32m156.37MB/s[0m[1mwrote[0m [32m493.70MB[0m in [36m 3s[0m, [32m155.66MB/s[0m[1mwrote[0m [32m506.02MB[0m in [36m 3s[0m, [32m156.66MB/s[0m[1mwrote[0m [32m518.33MB[0m in [36m 3s[0m, [32m157.19MB/s[0m[1mwrote[0m [32m529.33MB[0m in [36m 3s[0m, [32m156.64MB/s[0m[1mwrote[0m [32m540.22MB[0m in [36m 3s[0m, [32m157.15MB/s[0m[1mwrote[0m [32m552.46MB[0m in [36m 4s[0m, [32m157.44MB/s[0m[1mwrote[0m [32m564.65MB[0m in [36m 4s[0m, [32m158.12MB/s[0m[1mwrote[0m [32m574.86MB[0m in [36m 4s[0m, [32m157.94MB/s[0m[1mwrote[0m [32m585.45MB[0m in [36m 4s[0m, [32m157.64MB/s[0m[1mwrote[0m [32m597.56MB[0m in [36m 4s[0m, [32m158.23MB/s[0m[1mwrote[0m [32m609.25MB[0m in [36m 4s[0m, [32m158.18MB/s[0m[1mwrote[0m [32m621.55MB[0m in [36m 4s[0m, [32m157.80MB/s[0m[1mwrote[0m [32m633.83MB[0m in [36m 5s[0m, [32m132.30MB/s[0m[1mwrote[0m [32m646.15MB[0m in [36m 5s[0m, [32m121.46MB/s[0m[1mwrote[0m [32m658.44MB[0m in [36m 5s[0m, [32m120.97MB/s[0m[1mwrote[0m [32m670.97MB[0m in [36m 6s[0m, [32m118.71MB/s[0m[1mwrote[0m [32m683.27MB[0m in [36m 6s[0m, [32m117.04MB/s[0m[1mwrote[0m [32m695.66MB[0m in [36m 6s[0m, [32m116.70MB/s[0m[1mwrote[0m [32m706.54MB[0m in [36m 6s[0m, [32m117.25MB/s[0m[1mwrote[0m [32m717.16MB[0m in [36m 6s[0m, [32m117.60MB/s[0m[1mwrote[0m [32m729.55MB[0m in [36m 6s[0m, [32m118.54MB/s[0m[1mwrote[0m [32m741.74MB[0m in [36m 6s[0m, [32m119.26MB/s[0m[1mwrote[0m [32m753.85MB[0m in [36m 6s[0m, [32m119.47MB/s[0m[1mwrote[0m [32m766.22MB[0m in [36m 6s[0m, [32m120.07MB/s[0m[1mwrote[0m [32m778.35MB[0m in [36m 6s[0m, [32m120.50MB/s[0m[1mwrote[0m [32m790.73MB[0m in [36m 7s[0m, [32m121.20MB/s[0m[1mwrote[0m [32m803.10MB[0m in [36m 7s[0m, [32m121.45MB/s[0m[1mwrote[0m [32m815.47MB[0m in [36m 7s[0m, [32m121.67MB/s[0m[1mwrote[0m [32m827.84MB[0m in [36m 7s[0m, [32m122.36MB/s[0m[1mwrote[0m [32m840.23MB[0m in [36m 7s[0m, [32m122.93MB/s[0m[1mwrote[0m [32m852.39MB[0m in [36m 7s[0m, [32m122.99MB/s[0m[1mwrote[0m [32m864.92MB[0m in [36m 7s[0m, [32m123.63MB/s[0m[1mwrote[0m [32m877.24MB[0m in [36m 7s[0m, [32m124.19MB/s[0m[1mwrote[0m [32m887.55MB[0m in [36m 7s[0m, [32m124.41MB/s[0m[1mwrote[0m [32m896.29MB[0m in [36m 7s[0m, [32m124.70MB/s[0m[1mwrote[0m [32m907.66MB[0m in [36m 7s[0m, [32m124.54MB/s[0m[1mwrote[0m [32m920.06MB[0m in [36m 8s[0m, [32m122.35MB/s[0m[1mwrote[0m [32m932.39MB[0m in [36m 8s[0m, [32m120.50MB/s[0m[1mwrote[0m [32m944.78MB[0m in [36m 8s[0m, [32m119.26MB/s[0m[1mwrote[0m [32m957.16MB[0m in [36m 8s[0m, [32m117.14MB/s[0m[1mwrote[0m [32m969.55MB[0m in [36m 8s[0m, [32m117.74MB/s[0m[1mwrote[0m [32m981.99MB[0m in [36m 8s[0m, [32m118.51MB/s[0m[1mwrote[0m [32m994.03MB[0m in [36m 8s[0m, [32m118.98MB/s[0m[1mwrote[0m [32m1.00GB[0m in [36m 8s[0m, [32m118.88MB/s[0m[1mwrote[0m [32m1.01GB[0m in [36m 9s[0m, [32m119.15MB/s[0m[1mwrote[0m [32m1.03GB[0m in [36m 9s[0m, [32m119.57MB/s[0m[1mwrote[0m [32m1.04GB[0m in [36m 9s[0m, [32m119.79MB/s[0m[1mwrote[0m [32m1.05GB[0m in [36m 9s[0m, [32m120.25MB/s[0m[1mwrote[0m [32m1.06GB[0m in [36m 9s[0m, [32m120.70MB/s[0m[1mwrote[0m [32m1.08GB[0m in [36m 9s[0m, [32m120.96MB/s[0m[1mwrote[0m [32m1.09GB[0m in [36m 9s[0m, [32m121.45MB/s[0m[1mwrote

```
[0m [32m1.10GB[0m in [36m 9s[0m, [32m121.37MB/s[0m[1mwrote[0m [32m1.11GB[0m in [36m
9s[0m, [32m121.59MB/s[0m[1mwrote[0m [32m1.12GB[0m in [36m 9s[0m, [32m121.99MB/s[0m[
1mwrote[0m [32m1.13GB[0m in [36m 9s[0m, [32m122.17MB/s[0m[1mwrote[0m [32m1.14GB[0m in
 [36m 9s[0m, [32m122.22MB/s[0m[1mwrote[0m [32m1.15GB[0m in [36m 9s[0m, [32m122.61MB/s
[0m[1mwrote[0m [32m1.17GB[0m in [36m 9s[0m, [32m123.12MB/s[0m[1mwrote[0m [32m1.18GB
[0m in [36m10s[0m, [32m123.20MB/s[0m[1mwrote[0m [32m1.19GB[0m in [36m10s[0m, [32m123
.87MB/s[0m                                                                    [1mwr
ote[0m [32m2.15GB[0m in [36m10s[0m, [32m2.15GB/s[0m
                                        > save.image("C:\\Users\\Neha\\Desktop\\Cyclistic_Data\\
.RData")
>
>
> # Remove rows with missing critical columns
> > cleaned_data <- combined_data %>%
Error: unexpected '>' in ">"
> +   drop_na(ride_id, started_at, ended_at, member_casual)
Error: object 'ride_id' not found
> >
Error: unexpected '>' in ">"
> > # Convert time columns to date-time format
Error: unexpected '>' in ">"
> > cleaned_data <- cleaned_data %>%
Error: unexpected '>' in ">"
> +   mutate(
+ +     started_at = as.POSIXct(started_at, format = "%Y-%m-%d %H:%M:%S"),
Error: unexpected '=' in:
"+   mutate(
+     started_at ="
> +     ended_at = as.POSIXct(ended_at, format = "%Y-%m-%d %H:%M:%S")
Error: object 'ended_at' not found
> +   )
Error: unexpected ')' in "+   )"
> >
Error: unexpected '>' in ">"
> > # Calculate ride length in minutes
Error: unexpected '>' in ">"
> > cleaned_data <- cleaned_data %>%
Error: unexpected '>' in ">"
> +   mutate(ride_length = as.numeric(difftime(ended_at, started_at, units = "mins")))
Error: object 'ended_at' not found
> >
Error: unexpected '>' in ">"
> > # Remove negative and outlier ride lengths
Error: unexpected '>' in ">"
> > cleaned_data <- cleaned_data %>%
Error: unexpected '>' in ">"
> +   filter(ride_length > 0, ride_length <= 1440)  # Max 24 hours
Error: object 'ride_length' not found
> >
Error: unexpected '>' in ">"
> > # Add a column for the day of the week (1 = Sunday, 7 = Saturday)
Error: unexpected '>' in ">"
> > cleaned_data <- cleaned_data %>%
Error: unexpected '>' in ">"
> +   mutate(day_of_week = wday(started_at))
Error: object 'started_at' not found
> > write_csv(cleaned_data, "Cleaned_Cyclistic_Data.csv")
Error: unexpected '>' in ">"
> # Average ride length for members and casual riders
> cleaned_data %>%
+   group_by(member_casual) %>%
+   summarise(
+     average_ride_length = mean(ride_length),
+     max_ride_length = max(ride_length),
+     total_rides = n()
+   )
# A tibble: 2 × 4
  member_casual average_ride_length max_ride_length total_rides
  <chr>                      <dbl>           <dbl>       <int>
1 casual                      26.7            1440.     2564937
```

```
2 member                      13.3          1440.     3244650
> cleaned_data %>%
+   group_by(member_casual, day_of_week) %>%
+   summarise(average_ride_length = 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 × 3
# Groups:   member_casual [2]
   member_casual day_of_week average_ride_length
   <chr>               <dbl>               <dbl>
 1 casual                  1                31.0
 2 casual                  2                27.1
 3 casual                  3                24.4
 4 casual                  4                23.2
 5 casual                  5                23.1
 6 casual                  6                24.8
 7 casual                  7                29.1
 8 member                  1                15.1
 9 member                  2                12.9
10 member                  3                12.5
11 member                  4                12.6
12 member                  5                12.5
13 member                  6                13.0
14 member                  7                14.8
> cleaned_data %>%
+   filter(member_casual == "casual") %>%
+   count(day_of_week) %>%
+   arrange(desc(n))
# A tibble: 7 × 2
  day_of_week      n
        <dbl>  <int>
1           7 563721
2           1 486948
3           6 368678
4           2 292745
5           5 290073
6           4 283574
7           3 279198
> write_csv(cleaned_data, "Cleaned_Data_Summary.csv")
[1mwrote[0m [32m12.24MB[0m in [36m 0s[0m, [32m2.15GB/s[0m[1mwrote[0m [32m60.72MB[0m
 in [36m 0s[0m, [32m278.00MB/s[0m[1mwrote[0m [32m73.02MB[0m in [36m 0s[0m, [32m268.49
MB/s[0m[1mwrote[0m [32m84.20MB[0m in [36m 0s[0m, [32m263.65MB/s[0m[1mwrote[0m [32m9
5.56MB[0m in [36m 0s[0m, [32m252.34MB/s[0m[1mwrote[0m [32m107.89MB[0m in [36m 0s[0m,
[32m244.73MB/s[0m[1mwrote[0m [32m120.12MB[0m in [36m 1s[0m, [32m240.17MB/s[0m[1mwrot
e[0m [32m132.45MB[0m in [36m 1s[0m, [32m236.20MB/s[0m[1mwrote[0m [32m144.81MB[0m in
[36m 1s[0m, [32m233.11MB/s[0m[1mwrote[0m [32m156.95MB[0m in [36m 1s[0m, [32m229.86MB/s
[0m[1mwrote[0m [32m169.22MB[0m in [36m 1s[0m, [32m228.34MB/s[0m[1mwrote[0m [32m181.
56MB[0m in [36m 1s[0m, [32m227.27MB/s[0m[1mwrote[0m [32m193.89MB[0m in [36m 1s[0m, [
32m226.20MB/s[0m[1mwrote[0m [32m205.63MB[0m in [36m 1s[0m, [32m224.24MB/s[0m[1mwrote
[0m [32m218.01MB[0m in [36m 1s[0m, [32m222.94MB/s[0m[1mwrote[0m [32m230.19MB[0m in [3
6m 1s[0m, [32m222.35MB/s[0m[1mwrote[0m [32m242.51MB[0m in [36m 1s[0m, [32m221.62MB/s[
0m[1mwrote[0m [32m253.80MB[0m in [36m 1s[0m, [32m222.13MB/s[0m[1mwrote[0m [32m263.96
MB[0m in [36m 1s[0m, [32m219.92MB/s[0m[1mwrote[0m [32m276.29MB[0m in [36m 1s[0m, [32
m219.71MB/s[0m[1mwrote[0m [32m288.63MB[0m in [36m 1s[0m, [32m219.77MB/s[0m[1mwrote[0
m [32m300.90MB[0m in [36m 1s[0m, [32m219.45MB/s[0m[1mwrote[0m [32m313.14MB[0m in [36m
 1s[0m, [32m219.21MB/s[0m[1mwrote[0m [32m325.50MB[0m in [36m 2s[0m, [32m215.61MB/s[0m
[1mwrote[0m [32m336.89MB[0m in [36m 2s[0m, [32m216.19MB/s[0m[1mwrote[0m [32m347.97MB
[0m in [36m 2s[0m, [32m215.45MB/s[0m[1mwrote[0m [32m360.21MB[0m in [36m 2s[0m, [32m2
15.37MB/s[0m[1mwrote[0m [32m372.45MB[0m in [36m 2s[0m, [32m214.83MB/s[0m[1mwrote[0m
[32m384.77MB[0m in [36m 2s[0m, [32m214.41MB/s[0m[1mwrote[0m [32m397.02MB[0m in [36m 2
s[0m, [32m214.16MB/s[0m[1mwrote[0m [32m407.55MB[0m in [36m 2s[0m, [32m213.29MB/s[0m
[1mwrote[0m [32m419.82MB[0m in [36m 2s[0m, [32m204.71MB/s[0m[1mwrote[0m [32m432.23MB[
0m in [36m 2s[0m, [32m205.16MB/s[0m[1mwrote[0m [32m444.40MB[0m in [36m 2s[0m, [32m205
.44MB/s[0m[1mwrote[0m [32m456.70MB[0m in [36m 2s[0m, [32m205.19MB/s[0m[1mwrote[0m [
32m468.99MB[0m in [36m 2s[0m, [32m205.50MB/s[0m[1mwrote[0m [32m481.34MB[0m in [36m 2s
[0m, [32m205.27MB/s[0m[1mwrote[0m [32m493.70MB[0m in [36m 2s[0m, [32m204.88MB/s[0m[1
mwrote[0m [32m506.02MB[0m in [36m 2s[0m, [32m204.80MB/s[0m[1mwrote[0m [32m518.33MB[0m
 in [36m 3s[0m, [32m205.40MB/s[0m[1mwrote[0m [32m529.33MB[0m in [36m 3s[0m, [32m205.8
1MB/s[0m[1mwrote[0m [32m540.22MB[0m in [36m 3s[0m, [32m205.28MB/s[0m[1mwrote[0m [32
```

```
m552.46MB[0m in [36m 3s[0m, [32m204.86MB/s[0m[1mwrote[0m [32m564.65MB[0m in [36m 3s[0
m, [32m204.98MB/s[0m[1mwrote[0m [32m574.86MB[0m in [36m 3s[0m, [32m204.74MB/s[0m[1mw
rote[0m [32m585.45MB[0m in [36m 3s[0m, [32m204.14MB/s[0m[1mwrote[0m [32m597.56MB[0m i
n [36m 3s[0m, [32m204.65MB/s[0m[1mwrote[0m [32m609.25MB[0m in [36m 3s[0m, [32m204.42M
B/s[0m[1mwrote[0m [32m621.55MB[0m in [36m 3s[0m, [32m203.96MB/s[0m[1mwrote[0m [32m6
33.83MB[0m in [36m 3s[0m, [32m203.77MB/s[0m[1mwrote[0m [32m646.15MB[0m in [36m 3s[0m,
 [32m203.84MB/s[0m[1mwrote[0m [32m658.44MB[0m in [36m 3s[0m, [32m203.86MB/s[0m[1mwro
te[0m [32m670.97MB[0m in [36m 3s[0m, [32m203.88MB/s[0m[1mwrote[0m [32m683.27MB[0m in
[36m 3s[0m, [32m203.84MB/s[0m[1mwrote[0m [32m695.66MB[0m in [36m 4s[0m, [32m197.92MB/
s[0m[1mwrote[0m [32m706.54MB[0m in [36m 4s[0m, [32m187.12MB/s[0m[1mwrote[0m [32m717
.16MB[0m in [36m 4s[0m, [32m176.09MB/s[0m[1mwrote[0m [32m729.55MB[0m in [36m 5s[0m, [
[32m161.87MB/s[0m[1mwrote[0m [32m741.74MB[0m in [36m 5s[0m, [32m154.38MB/s[0m[1mwrote
[0m [32m753.85MB[0m in [36m 5s[0m, [32m154.48MB/s[0m[1mwrote[0m [32m766.22MB[0m in [
36m 5s[0m, [32m155.28MB/s[0m[1mwrote[0m [32m778.35MB[0m in [36m 5s[0m, [32m155.84MB/s[
[0m[1mwrote[0m [32m790.73MB[0m in [36m 5s[0m, [32m156.47MB/s[0m[1mwrote[0m [32m803.1
0MB[0m in [36m 5s[0m, [32m154.32MB/s[0m[1mwrote[0m [32m815.47MB[0m in [36m 5s[0m, [3
2m150.04MB/s[0m[1mwrote[0m [32m827.84MB[0m in [36m 6s[0m, [32m148.32MB/s[0m[1mwrote[
0m [32m840.23MB[0m in [36m 6s[0m, [32m148.61MB/s[0m[1mwrote[0m [32m852.39MB[0m in [36
m 6s[0m, [32m148.65MB/s[0m[1mwrote[0m [32m864.92MB[0m in [36m 6s[0m, [32m148.41MB/s[0
m[1mwrote[0m [32m877.24MB[0m in [36m 6s[0m, [32m148.88MB/s[0m[1mwrote[0m [32m887.55M
B[0m in [36m 6s[0m, [32m148.40MB/s[0m[1mwrote[0m [32m896.29MB[0m in [36m 6s[0m, [32m
148.16MB/s[0m[1mwrote[0m [32m907.66MB[0m in [36m 6s[0m, [32m146.67MB/s[0m[1mwrote[0m
 [32m920.06MB[0m in [36m 6s[0m, [32m143.92MB/s[0m[1mwrote[0m [32m932.39MB[0m in [36m
7s[0m, [32m141.15MB/s[0m[1mwrote[0m [32m944.78MB[0m in [36m 7s[0m, [32m138.47MB/s[0m
[1mwrote[0m [32m957.16MB[0m in [36m 7s[0m, [32m137.79MB/s[0m[1mwrote[0m [32m969.55MB[
[0m in [36m 7s[0m, [32m136.49MB/s[0m[1mwrote[0m [32m981.99MB[0m in [36m 7s[0m, [32m13
7.54MB/s[0m[1mwrote[0m [32m994.03MB[0m in [36m 7s[0m, [32m137.82MB/s[0m
                                                                [1mwrote[0m [32m1.00GB[0m in
[36m 7s[0m, [32m138.08MB/s[0m[1mwrote[0m [32m1.01GB[0m in [36m 7s[0m, [32m138.63MB/s[
[0m[1mwrote[0m [32m1.03GB[0m in [36m 7s[0m, [32m139.10MB/s[0m[1mwrote[0m [32m1.04GB[
[0m in [36m 7s[0m, [32m139.37MB/s[0m[1mwrote[0m [32m1.05GB[0m in [36m 7s[0m, [32m140.
20MB/s[0m[1mwrote[0m [32m1.06GB[0m in [36m 8s[0m, [32m140.46MB/s[0m[1mwrote[0m [32m
1.08GB[0m in [36m 8s[0m, [32m140.82MB/s[0m[1mwrote[0m [32m1.09GB[0m in [36m 8s[0m, [
32m140.90MB/s[0m[1mwrote[0m [32m1.10GB[0m in [36m 8s[0m, [32m140.84MB/s[0m[1mwrote[0
m [32m1.11GB[0m in [36m 8s[0m, [32m141.63MB/s[0m[1mwrote[0m [32m1.12GB[0m in [36m 8s[
[0m, [32m141.56MB/s[0m[1mwrote[0m [32m1.13GB[0m in [36m 8s[0m, [32m142.00MB/s[0m[1mw
rote[0m [32m1.14GB[0m in [36m 8s[0m, [32m142.40MB/s[0m[1mwrote[0m [32m1.15GB[0m in [
36m 8s[0m, [32m142.67MB/s[0m[1mwrote[0m [32m1.17GB[0m in [36m 8s[0m, [32m143.02MB/s[0
m[1mwrote[0m [32m1.18GB[0m in [36m 8s[0m, [32m143.32MB/s[0m[1mwrote[0m [32m1.19GB[0
m in [36m 8s[0m, [32m144.15MB/s[0m
                      [1mwrote[0m [32m2.15GB[0m in [36m 8s[0m, [32m2.15GB/s[0m
                                                                  > save.image("C:\\Users\\Neha
\\Desktop\\Cyclistic_Data\\data")
> install.packages("ggplot2")
Warning: package ‘ggplot2’ is in use and will not be installed
> library(ggplot2)
> ggplot(cleaned_data, aes(x = member_casual, y = ride_length, fill = member_casual)) +
+   geom_boxplot() +
+   labs(title = "Ride Length Comparison: Members vs Casual Riders",
+        x = "User Type",
+        y = "Ride Length (minutes)") +
+   theme_minimal()
> ggplot(cleaned_data, aes(x = factor(day_of_week, labels = c("Sun", "Mon", "Tue", "Wed", "Thu",
"Fri", "Sat")),
+                          fill = member_casual)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Rides by Day of the Week",
+        x = "Day of the Week",
+        y = "Number of Rides") +
+   theme_minimal()
> save.image("C:\\Users\\Neha\\Desktop\\Cyclistic_Data\\graph cyc")
> ggsave("Ride_Comparison_Boxplot.png")
Saving 5.76 x 5.75 in image
>
```