

Cyclistic Data Cleaning and Exporting

Maxwell Dill

2022-04-04

Install appropriate packages for cleaning and manipulating

```
install.packages("tidyverse")
install.packages("here")
install.packages("skimr")
install.packages("janitor")

library(tidyverse)
library(here)
library(skimr)
library(janitor)
```

Combining the data from all 12 months into one data frame

```
twelve_months <- rbind(Apr_21, Aug_21, Dec_21, Feb_22, Jan_22, Jul_21, Jun_21, Mar_21, May_21, Nov_21,
                        Oct_21, Sep_21)
```

Cleaning

Ensured all rows were unique

```
unique(cleaned_twelve_months)
glimpse(cleaned_twelve_months)
distinct(cleaned_twelve_months)
```

Created a new data frame which removed rows where the ride length was less than 1 minute or greater than 24 hours

```
cleaned_twelve_months <- filter(twelve_months, ride_length > 1, ride_length < 1440)
cleaned_twelve_months %>%
  summarize(min(ride_length), max(ride_length))
```

Created a new column “ride_length” that takes the difference between the end_time and start_time and rounded

```
cleaned_twelve_months <- cleaned_twelve_months %>%  
  mutate(ride_length = difftime(ended_at, started_at, units = "mins"))  
  
cleaned_twelve_months <- cleaned_twelve_months %>%  
  mutate(round(ride_length, 2))
```

Added a new column “weekday” that converts the started_at date to the corresponding weekday

```
cleaned_twelve_months <- cleaned_twelve_months %>%  
  mutate(weekday = weekdays(started_at))
```

Created a new data frame “main data” that contained only the necessary columns for data visualization

```
main_data <- data.frame(cleaned_twelve_months$rideable_type, cleaned_twelve_months$member_casual,  
                        cleaned_twelve_months$ride_length, cleaned_twelve_months$weekday)
```

Exporting

Exported “main data” data frame as a csv for PowerBI visualization

```
write.csv(main_data, "C:\\Users\\Max - Work\\Documents\\Case Studies\\Cyclistic_Mar_2022\\main_data.csv",  
          , row.names = FALSE)
```