**#Install & Load Required Libraries**

install.packages(c("readxl", "dplyr", "tidyr", "lubridate", "ggplot2", "janitor", “modeest”))

library(readxl) # To read Excel files

library(dplyr) # For data manipulation

library(tidyr) # For data reshaping

library(lubridate) # For date/time cleaning

library(janitor) # For cleaning column names

library(modeest) # For Calculating mode

**#Import Your Excel Files**

# Read 2019 and 2020 data

data\_2019 <- read\_excel("C:\\Rproj\\data\\Divvy\_Trips\_2019\_Q1.xlsx")

data\_2020 <- read\_excel("C:\\Rproj\\data\\Divvy\_Trips\_2020\_Q1.xlsx")

**#Clean column names**

data\_2019 <- data\_2019 %>% clean\_names()

data\_2020 <- data\_2020 %>% clean\_names()

**#Standardize column names**

#Rename usertype to member\_casual in 2019.

#Convert values for consistency (Subscriber → member, Customer → casual).

data\_2019 <- data\_2019 %>%

rename(member\_casual = usertype) %>%

mutate(member\_casual = case\_when(

member\_casual == "Subscriber" ~ "member",

member\_casual == "Customer" ~ "casual",

TRUE ~ member\_casual

))

**#Align columns between both datasets**

#2019 has start\_time and end\_time

#2020 has started\_at and ended\_at

#Let's rename 2020 columns to match 2019:

data\_2020 <- data\_2020 %>%

rename(

start\_time = started\_at,

end\_time = ended\_at

)

**#Add derived columns for both datasets**

# Function to add derived fields

add\_derived\_fields <- function(df) {

df %>%

mutate(

start\_time = ymd\_hms(start\_time),

end\_time = ymd\_hms(end\_time),

ride\_length = as.numeric(difftime(end\_time, start\_time, units = "mins")),

ride\_year = year(start\_time),

ride\_month = month(start\_time, label = TRUE, abbr = TRUE),

day\_of\_week = wday(start\_time, label = TRUE),

hour\_of\_day = hour(start\_time)

) %>%

filter(ride\_length > 0, !is.na(ride\_length))

}

# Apply to both datasets

data\_2019 <- add\_derived\_fields(data\_2019)

data\_2020 <- add\_derived\_fields(data\_2020)

# **Safely merge both cleaned datasets**

all\_trips <- bind\_rows(data\_2019, data\_2020)

**# Export cleaned dataset to CSV for Tableau**

write.csv(all\_trips, "Cyclistic\_cleaned\_data.csv", row.names = FALSE)