Cyclistic Membership

Neal

2025-10-15

1. Clean and manipulate the data in excel

Data source:

Divvy Trip data: https://divvy-tripdata.s3.amazonaws.com/index.html

September 2024 to August 2025

Removed:

- start_station_id
- end_station_id
- start_lat
- start_lng
- end_lat
- end_lng

Added:

- day_of_week
  - formula: `=WEEKDAY([started_at],1)`
  - then converted the numerical values to the weekday text values using `=TEXT([day_of_week], "dddd")`. Afterwards, copied and pasted the text values to the day_of_week column to replace numerical values with the text values.
- Proper cased the member_casual values

1. Load in packages

```
library(tidyverse)
library(readxl)
library(scales)
library(knitr)
```

2. Import 12 months of data

3. Combine all sheets in order

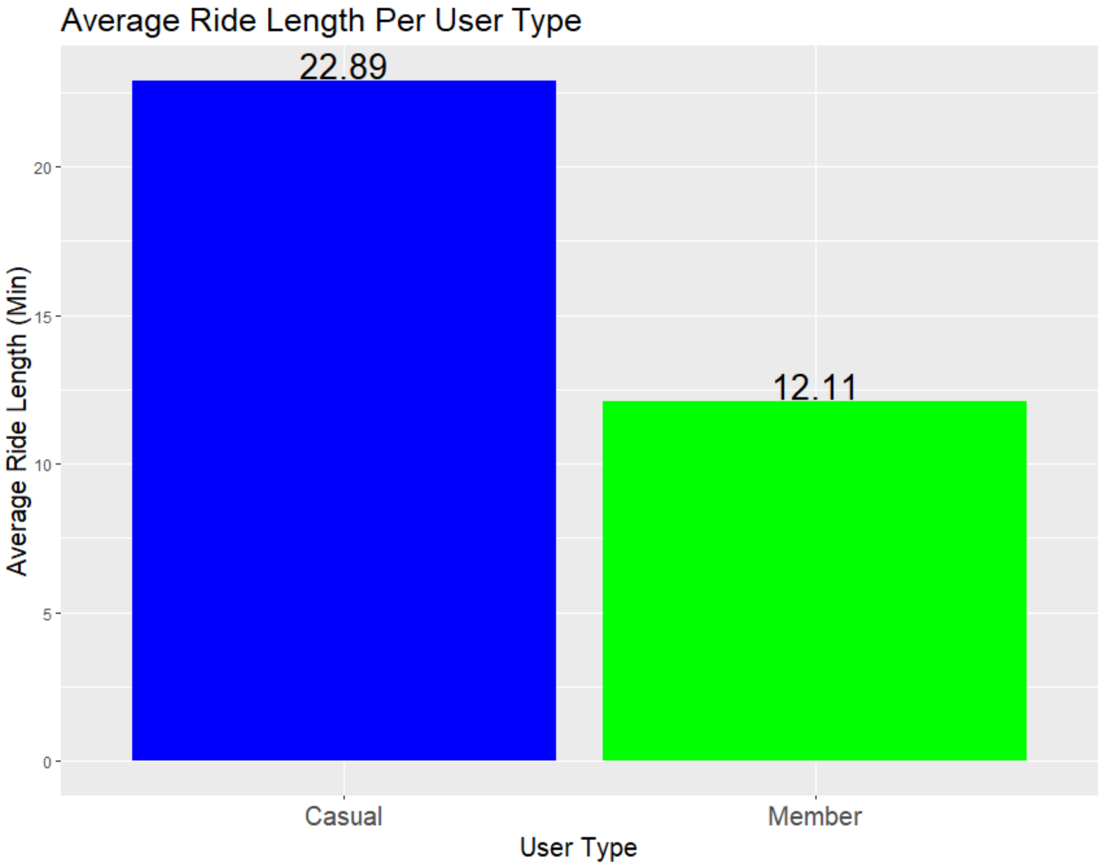
```
combine_12_months <-  bind_rows(aug_25, july_25, june_25,may_25, apr_25,
mar_25, feb_25, jan_25, dec_24, nov_24, oct_24, sep_24)
```

4. Separate the time and date from the `ride_length` column and convert the data type into doubles

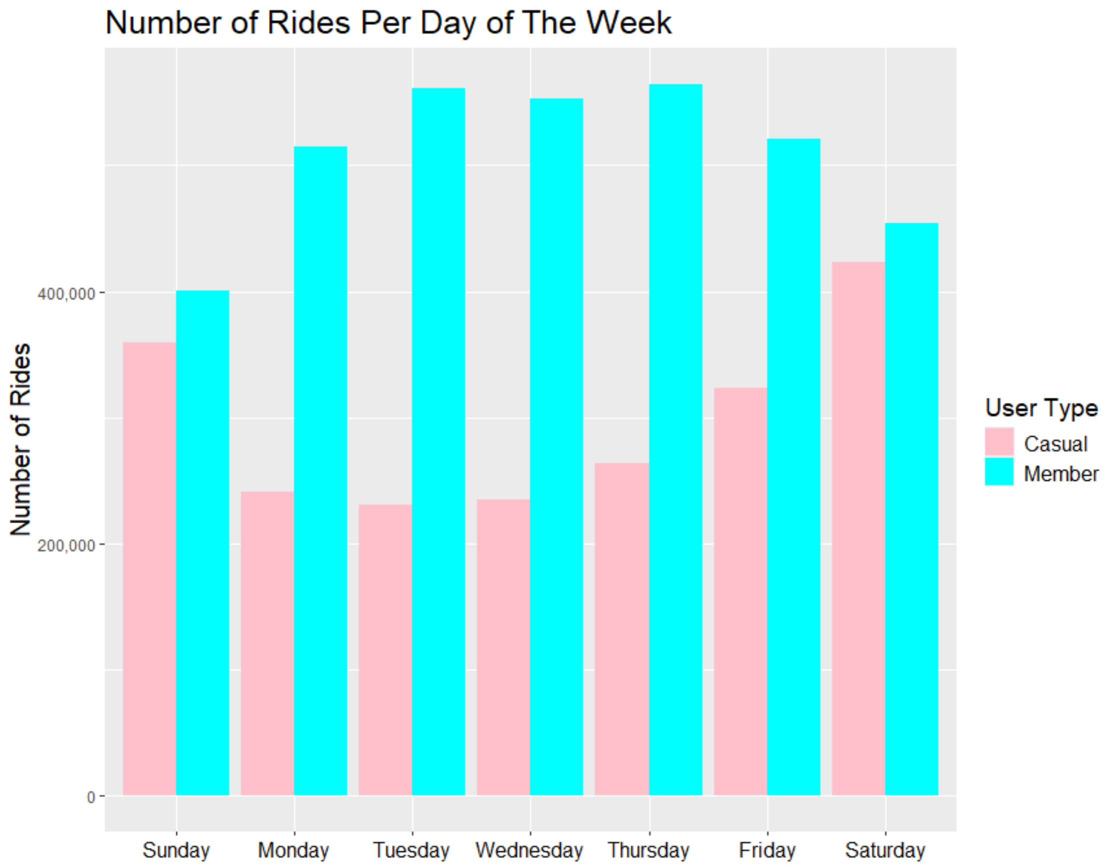
```
combine_12_months <- combine_12_months %>%
  separate(ride_length, c("ride_date", "ride_length"), sep = " ") %>%
  select(-ride_date)

combine_12_months$ride_length <- as.numeric(combine_12_months$ended_at -
combine_12_months$started_at)/60

combine_12_months$day_of_week <- factor(combine_12_months$day_of_week, levels
= c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday"))
```

5. Further analyze and plot data

I. Find the average ride length per user type and plot it

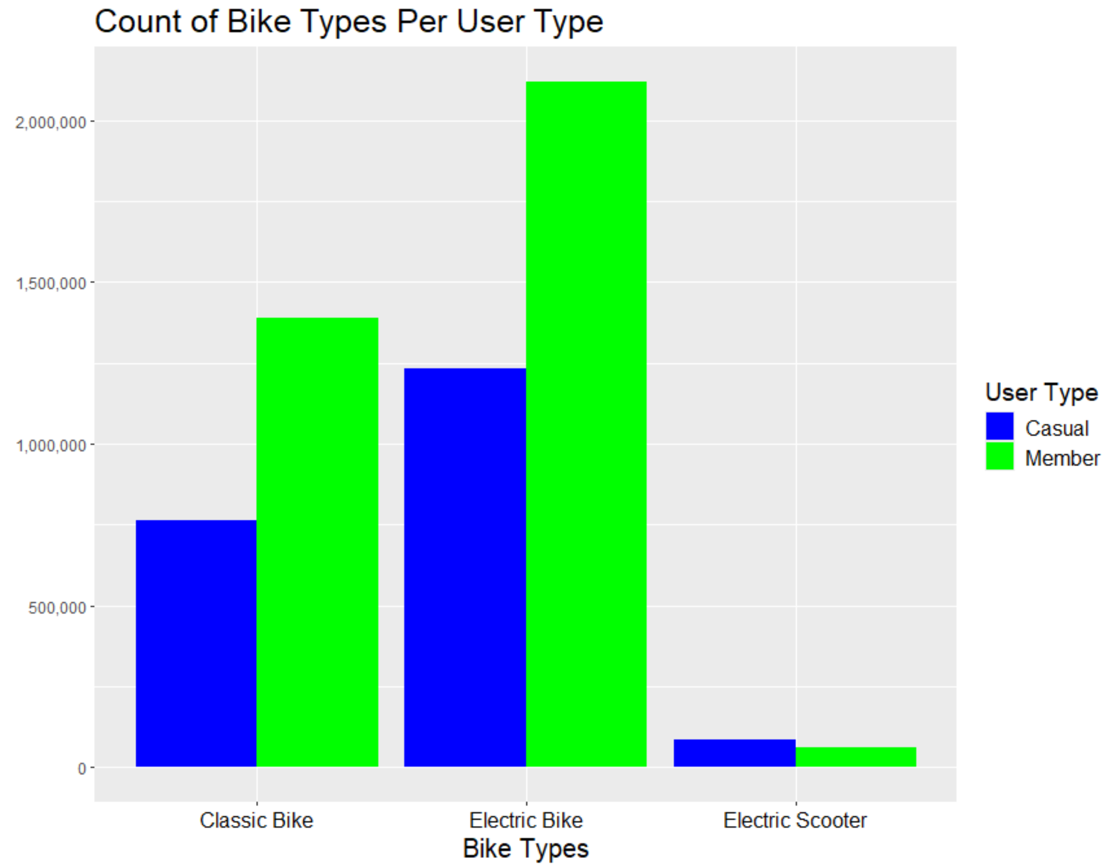


II. Find the relationship between user type and day of week, then plot



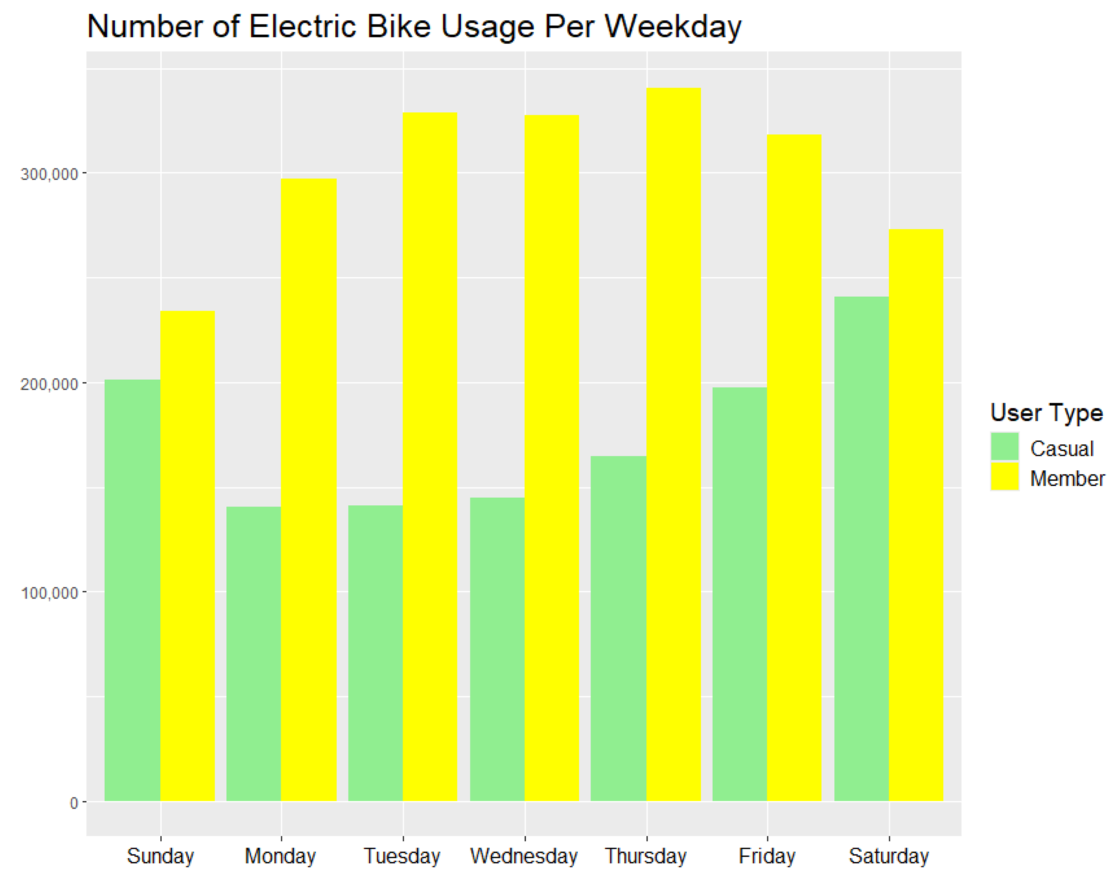
III. Plot the Count of Bike Types over the past year

```
bike_type <- combine_12_months %>%
  group_by(rideable_type) %>%
  count(member_casual)


bike_type_plot <- ggplot(data = bike_type, mapping = aes(x=rideable_type,
y=n, fill = member_casual)) +
  geom_col(position = position_dodge() ,aes(fill = member_casual))+
  scale_fill_manual(values = c("Casual" = "blue", "Member" = "green")) +
  scale_y_continuous(labels = label_comma()) +
  scale_x_discrete(labels = c("Classic Bike", "Electric Bike", "Electric
Scooter")) +
  labs(title = "Count of Bike Types Per User Type", x = "Bike Types", y =
NULL, fill ="User Type") +
  theme(axis.text.x = element_text(size = 12, color = "black"), axis.title.x
= element_text(size = 15), plot.title = element_text(size = 19), legend.title
= element_text(size = 14), legend.text = element_text(size = 12), plot.margin
= margin(l=10, r=5, t=5, b=5))

plot(bike_type_plot)
```



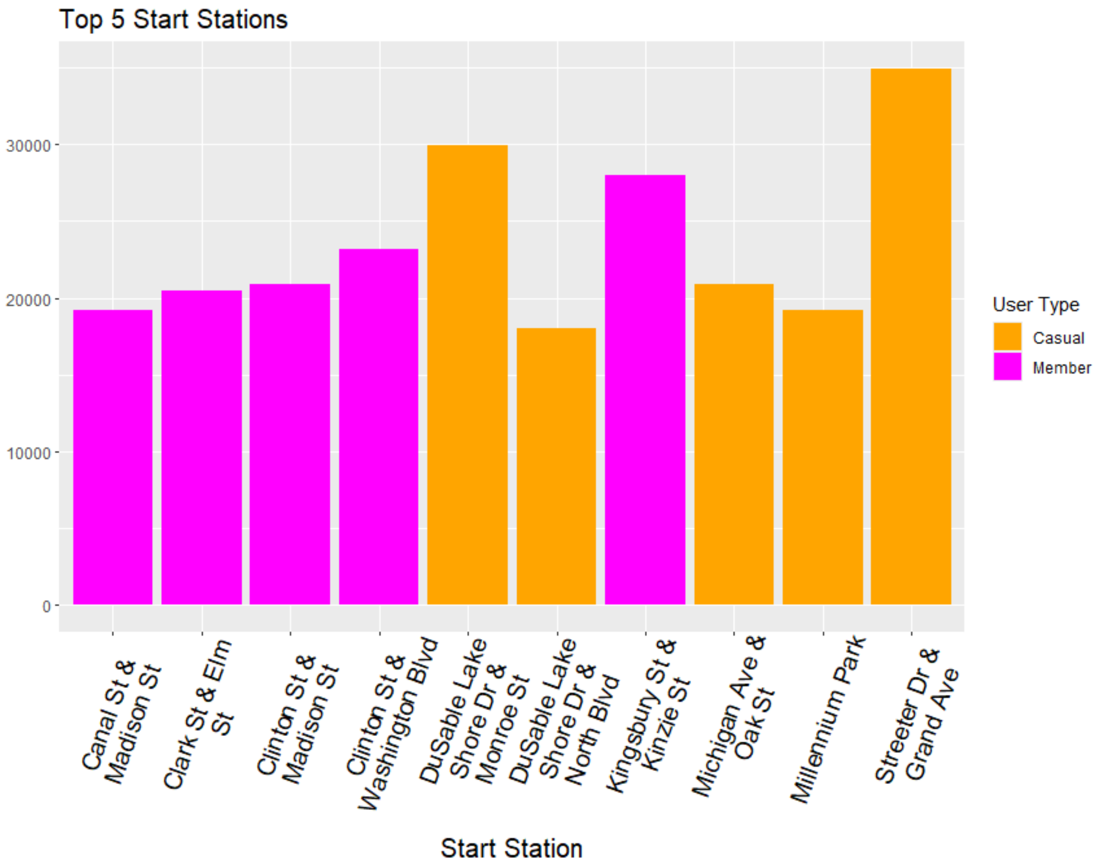
IV. Find and plot the electric bike usage over the weekday

```
elec_bike <- combine_12_months %>%
  filter(rideable_type == "electric_bike") %>%
  group_by(day_of_week, member_casual) %>%
  count(rideable_type) %>%
  ggplot(mapping = aes(x=day_of_week, y=n, fill = member_casual)) +
  geom_col(position = position_dodge())+
  scale_fill_manual(values = c("Casual" = "lightgreen", "Member" = "yellow"))
+
  labs(title = "Number of Electric Bike Usage Per Weekday", x=NULL, y=NULL,
fill= "User Type") +
  theme(axis.text.x = element_text(size = 12, color = "black"), plot.title =
element_text(size = 19), legend.title = element_text(size = 14), legend.text
= element_text(size = 12), plot.margin = margin(l=10, r=5, t=5, b=5))+
  scale_y_continuous(labels = label_comma())

plot(elec_bike)
```



VI. Counts how many people start at each start station for casual users, top 5, then plots

```
start_station <- combine_12_months %>%
  group_by(member_casual, start_station_name) %>%
  na.omit() %>%
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  top_n(5)

station_casual_plot <-  ggplot(data = start_station , mapping
=aes(x=start_station_name, y=Count, fill = member_casual)) +
  geom_col(position = position_dodge()) +
  theme(axis.text.x = element_text(angle = 70, vjust = .6, hjust = .5, size =
14, color = "black"), plot.title = element_text(size = 15), axis.title.x =
element_text(size = 15),plot.margin = margin(l=10, r=5, t=5, b=5)) +
  labs(title = "Top 5 Start Stations", x="Start Station", y=NULL, fill =
"User Type")+
  scale_fill_manual(values = c("Casual" = "orange", "Member" = "magenta")) +
  scale_x_discrete(labels = function(x) str_wrap(x, width = 15))

plot(station_casual_plot)
```


Top 5 Start Stations
30000
20000
10000
0
Canal St & Madison St
Clark St & Elm St
Clinton St & Madison St
Clinton St & Washington Blvd
DuSable Lake Shore Dr & Monroe St
DuSable Lake Shore Dr & North Blvd
Kingsbury St & Kinzie St
Michigan Ave & Oak St
Millennium Park
Streeter Dr & Grand Ave
Start Station
User Type
Casual
Member