Cyclistic Project: Log of changes made.

Ask

What is the problem we are trying to solve?

-We want to increase the number of annual subscription-based users.

Task: Understand the difference in the way members and casual rides use the service.

Prepare

- Data was obtained from: https://divvy-tripdata.s3.amazonaws.com/index.html for 2023
- Original data was provided in .csv file. It was provided by Motivate International Inc.
- The source of the data suggests that it follows ROCCC.
- Sorted data for casual and member riders

Process

- Formatting data to make sure the formats are consistent and interpretable for plotting.
- Added a column for length of rides in "ride_length" column.
- Removed columns for station names and Id as they are not useful for this project.
- Converted latitude and longitude related columns to numbers after resolving errors with format.
- Remove rows where start time is later than end time since it is error is logging or recording data.
- Since ride_id is just a unique key for identifying different rides, we can remove duplicates based on it.
- No duplicates were found!
- day_of_week column: Numbers 1 (Monday) through 7 (Sunday).
- Added a column for length of rides in "ride_length" column using Haversine forumula:=
  6371 * ACOS(COS(RADIANS(Lat2)) * COS(RADIANS(Lat1)) * COS(RADIANS(Long2) - RADIANS(Long1))
  + SIN(RADIANS(Lat2)) * SIN(RADIANS(Lat1))).
- Removing 0 end_lng rows since it has to be an error value. Sheet 6.
- Ride time is calculated in minutes.
- Removed negative ride times.