**BikeShare Case Study**

**Introduction**:

This case study is part of the Coursera Google Data Analytics Capstone Project. Cyclistic, a fake BikeShare company founded in 2016 which I work for, believes that maximizing the number of future annual riders as opposed to casual riders will be key to future growth. Cyclistic financial analysts have concluded that annual riders, defined as those who purchase annual memberships, are more profitable than casual ones, defined as customers who purchase single or full day passes. My team has been tasked with identifying trends in historical Cyclistic data in order to design strategies to convert casual riders into annual riders.

**Ask**:

The question I have been tasked to investigate is, how do annual members and casual riders use Cyclistic bikes differently?

Business Task: Cyclistic wishes to increase revenue by converting casual riders into annual ones.

**Prepare**:

The data is public and has been made available by Motivate International Inc through Coursera under <https://www.divvybikes.com/data-license-agreement>.

Description: I have 12 ZIP files containing excel files for each month in 2022 for Cyclistic. The files contain information for start and end times, stations (name and id), latitudes and longitudes, and members, classified between annual and casual. The data is ROCCC as I used the most recent year and the data is provided by the company themselves.

**Process**:

I added additional columns to each excel file, titled ride\_length and day\_of\_week. Ride\_length is calculated as end time minus start time and day\_of\_week is the day of the week the ride started given by the weekday function.

I then import all excel files into R Studio for data cleaning and merging. First, I use glimpse() to see data types. Then, I use the sapply() and identical() functions to check that all data types are the same across file columns. Next, I use filter() to return any rows with ride\_length < 0 and swap the start and end times for these rows across months. I check for duplicates via sum(duplicated()) for every month and delete any duplicate rows. Finally, I merge all 12 sheets into a yearly sheet via rbind().

**Rstudio Code**:

#install packages from library

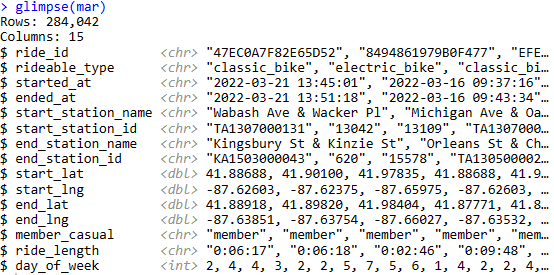
library(tidyverse)

library(lubridate)

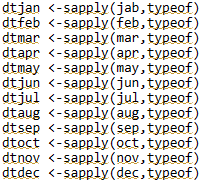
#read files



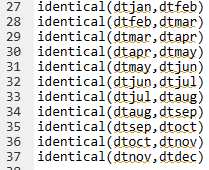
#glimpse() to see data type, example for march below



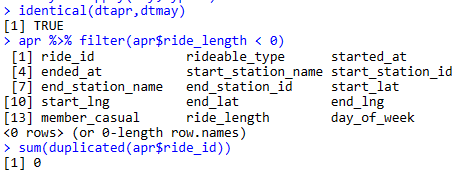
#sapply() used to apply typeof() across every column and extract a data type object.



#identical() used to compare the data types



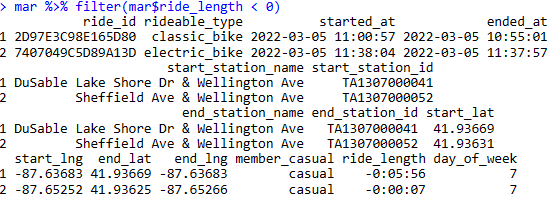




#filter() to see if ride length less than 0 for every month

mar %>% filter(mar$ride\_length < 0)

mar has 2 errors



#check for dupes

sum(duplicated(jan$ride\_id))



#merge final sheet

year <- rbind(jan,feb,mar,apr,may,jun,jul,aug,oct,nov,dec)

**Analyze**:

Most of the analysis and charts were done in Excel, with some additional R Code below.

The total number of rides per month spikes in the summer, specifically in July. There are no differences when splitting members and casual riders.

The average ride length spikes in the summer, specifically in June. Note that although both members and casual average ride lengths spike in the summer, the average ride length for members is more consistent and casual riders experience a steeper drop in average ride length going into the winter.

There is no significant difference in members’ max ride times each month with January having the shortest one at 7 hours 51 min and the rest of the month all being around 25 hours. Looking at casual riders, I notice that the max ride times for each month is significantly higher, ranging from 7 to 29 days excluding January which is the low.

Rides per day of week are lowest on Monday and highest on Saturday, increasing between Monday and Saturday then decreasing on Sunday across months.

Riders prefer electric bikes the most at 51% of total rides, with classic bikes falling 6% behind. The number of rides spikes at 4, 5, and 6 PM.

**R Studio Code**:

#Change rideable\_type to factor to get a summary count of types of bikes

year <- year %>% mutate(rideable\_type=as.factor(rideable\_type))

year <- year %>% mutate(started\_at=as.factor(started\_at))

summary(year)

#Extract hour from started\_at and convert to data frame in order to get a count of rides per hour

time <- format(as.POSIXct(year$started\_at), format = "%H")

timedf <- as.data.frame(table(time))

write.csv(timedf,"yeartime.csv")

**Share**:

I created a PowerPoint presentation to display my findings. See attached.

**Act**:

Suggestions: 1. Create a summer promotion for casual riders, targeting June through August, the three most popular months. 2. Remind casual riders (either in-app, through email, or on the bike stations) that they can convert to an annual membership if they ride regularly, such as for work. Can tie this in with a promotion. 3. Offer exclusive weekend or 4PM to 6PM perks for annual members and promote this to casual members who may want to switch.

1. Create a summer promotion program for casual riders, targeting June through August, the three most popular months.