Google capstone case study:Cyclistic Bike-Share Analysis

Gungun Aggarwal

2023-03-06

## ABOUT COMPANY

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

## Business Goal

Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. –So maximizing the number of annual members will be key to future growth. –Rather than creating a marketing campaign that targets all-new customers

**The following data analysis steps will be followed:** –Ask, Prepare, Process, Analyze, Share, Act. **loading the packages**

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(ggplot2)  
library(dplyr)  
library(janitor)

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(lubridate)

**importing CSV file in R Studio**

setwd("C:\\Users\\pc\\Desktop\\cycle\_12monthData\\MONTHS")  
Jan <- read.csv("202201-JAN.csv")  
Feb <- read.csv("202202-FEB.csv")  
Mar <- read.csv("202203-MAR.csv")  
Apr <- read.csv("202204-APR.csv")  
May <- read.csv("202205-MAY.csv")  
Jun <- read.csv("202206-JUNE.csv")  
Jul <- read.csv("202207-JULY.csv")  
Aug <- read.csv("202208-AUG.csv")  
Sep <- read.csv("202209-SEPT.csv")  
Oct <- read.csv("202210-OCT.csv")  
Nov <- read.csv("202211-NOV.csv")  
Dec <- read.csv("202212-DEC.csv")  
Jan\_23 <- read.csv("202301-JAN.csv")

**Merging into one dataframe**

rawdata <- rbind(Jan,Feb,Mar,Apr,May,Jun,Jul,Aug,Sep,Oct,Nov,Dec,Jan\_23)

**Data Cleaning Process**

rawdata<-remove\_empty(rawdata,which=c("cols"))  
rawdata<-remove\_empty(rawdata,which=c("rows"))  
dim(rawdata)  
head(rawdata) #see the first 6 rows of the data frame  
tail(rawdata) #see the last 6 rows of the data frame  
str(rawdata) #see list of columns and data types  
summary(rawdata) #statistical summary of data  
colnames(rawdata) #list of column names  
  
#changing the date formate  
rawdata$started\_at<-ymd\_hms(rawdata$started\_at)  
rawdata$ended\_at<-ymd\_hms(rawdata$ended\_at)  
  
#adding new column trip\_length  
rawdata$trip\_length<- difftime(rawdata$ended\_at,rawdata$started\_at,units = "mins")  
# removing negative values  
cleandata<-rawdata %>%   
 filter(trip\_length>0) %>%   
 drop\_na()

#### Total rows:58,58,018

–539 rows were having negative trip duration –5,985 rows had negative data

#### cleaned rows:58,51,494

## Analysing

mean(cleandata$trip\_length)#average duration of ride

## Time difference of 16.15332 mins

cat(round(max(cleandata$trip\_length)/60/24),"days")#maximum duration of ride

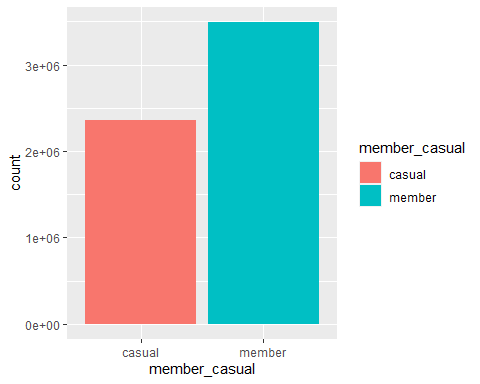
## 24 days

cleandata%>%  
 group\_by(member\_casual)%>%   
 summarise(n=n())%>%  
 mutate(percent = n\*100/sum(n))

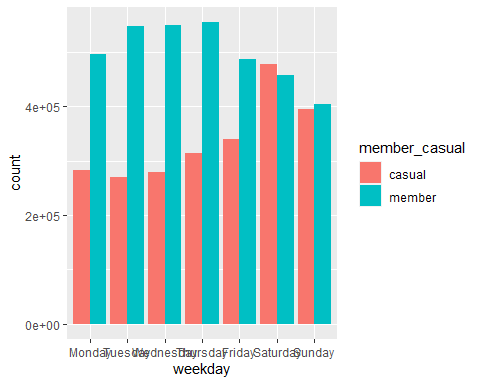
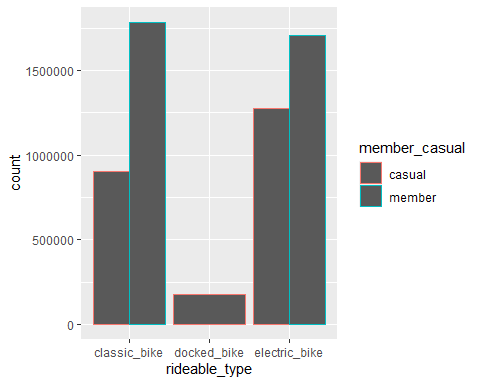
## # A tibble: 2 × 3  
## member\_casual n percent  
## <chr> <int> <dbl>  
## 1 casual 2356502 40.3  
## 2 member 3494992 59.7

## Including Plots

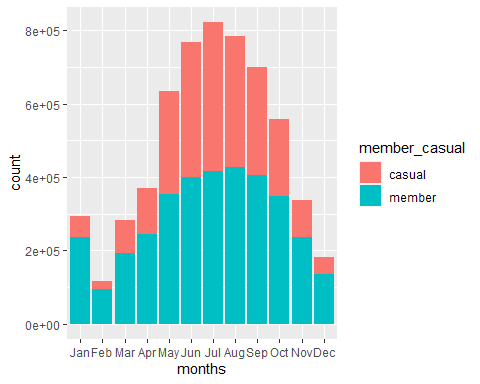
### member vs casual



## riders behaviour according to weekdays

 ## riders behaviour according to bike types 

## riders behaviour according to months



## Share

#### RECOMMENDATIONS

To convert casual riders into annual members, the following recommendations can be implemented:

**1.**We saw that casual workers uses bike more often on weekends,so if we give more discount on member riders then the casual workers are more likely to shift to be a member user of this company

**2.**We can also offer membership discount to docked type rider cause only casual rider uses them

**3.**people uses the bikes more often in summer months[MAY-OCT],SO some occasional offer can be provided limited to only member user.