Cyclistic Case Study

Briefing

Founded in 2016, Cyclistic, a bikeshare company in the Chicago area featuring more than 5,800 bikes and 600 docking stations is preparing the launch of a new marketing campaign. The Director of Marketing, Lily Moreno, is convinced the success of the next phase of the company's lifecycle is in maximizing the number of annual memberships.

Previous marketing strategies have focused on brand awareness and general growth of the customer base. Due to the increased profitability of annual members as compared to casual riders, this analysis will focus on the behavioral differences between these two customer classifications relative to Cyclistic's services. These behavioral differences will provide valuable insights into marketing strategies for converting casual riders into annual members.

Data Processing Log (Page Five for Analysis)

Data manipulation and cleaning was conducted in RStudio to leverage the efficiency and accuracy of the R language. All data was collected and supplied by Cyclistic. Data has been sampled from December, 2021 to November, 2022 as directed by the case study outline. As the data was collected in house, it is believed to be accurate, complete and free of bias.

Set up environment for data manipulation/cleaning

```r

library(tidyverse)
library(data.table)


```

Import raw datasets

```r
Dec_21 <- read.csv("202112-divvy-tripdata.csv")
Jan_22 <- read.csv("202201-divvy-tripdata.csv")
Feb_22 <- read.csv("202202-divvy-tripdata.csv")
Mar_22 <- read.csv("202203-divvy-tripdata.csv")
Apr_22 <- read.csv("202204-divvy-tripdata.csv")
May_22 <- read.csv("202205-divvy-tripdata.csv")
Jun_22 <- read.csv("202206-divvy-tripdata.csv")
Jul_22 <- read.csv("202207-divvy-tripdata.csv")
Aug_22 <- read.csv("202208-divvy-tripdata.csv")
Sept_22 <- read.csv("202209-divvy-tripdata.csv")
Oct_22 <- read.csv("202210-divvy-tripdata.csv")
Nov_22 <- read.csv("202211-divvy-tripdata.csv")
```

Examine the structure of raw data for dissimilarities. All data types appear synonymous between sets.

```r
str(Dec_21)
str(Jan_22)
str(Feb_22)
str(Mar_22)
str(Apr_22)
str(May_22)
str(Jun_22)
str(Jul_22)
str(Aug_22)
str(Sept_22)
str(Oct_22)
str(Nov_22)
```

Combine individual datasets into single annual dataset.

```
annual_trips <- bind_rows(Dec_21, Jan_22, Feb_22, Mar_22, Apr_22,
                          May_22, Jun_22, Jul_22, Aug_22, Sept_22,
                          Oct_22, Nov_22)
```

Convert started_at/ended_at datatype from *CHR* to *DATE* for easier analysis.

```
annual_trips$started_at <- as.POSIXct(annual_trips$started_at, format = "%Y-%m-%d %H:%M:%S")

annual_trips$ended_at <- as.POSIXct(annual_trips$ended_at, format = "%Y-%m-%d %H:%M:%S")
```

Organize data chronologically

```
annual_trips <- annual_trips %>%
  arrange(started_at)
```

Add column for ride duration and convert to *NUM* datatype for easier analysis.

```
annual_trips$ride_duration <- difftime(annual_trips$ended_at, annual_trips$started_at, units = "secs")

annual_trips$ride_duration <- as.numeric(as.character(annual_trips$ride_duration))
```

Add columns for day, week, month, yearly analysis and isolate time and date variables.

```
# Day
annual_trips$day <- format(annual_trips$started_at, "%d")
# Week
annual_trips$week <- format(annual_trips$started_at, "%W")
# Month
annual_trips$month <- format(annual_trips$started_at, "%m")
# Year
annual_trips$year <- format(annual_trips$started_at, "%Y")
# Day of the week
annual_trips$day_of_the_week <- format(annual_trips$started_at, "%A")
# Isolate date from time
annual_trips$YMD <- format(annual_trips$started_at, "%Y-%m-%d")
# Isolate time from date
annual_trips$time_of_day <- format(annual_trips$started_at, "%H:%M:%S")
```

Cleaning Data

Instantiating clean dataframe and removing null values.

```
annual_trips_clean <- annual_trips %>%
  filter(!(is.na(start_station_name) | start_station_name == "")) %>%
  filter(!(is.na(end_station_name) | end_station_name =""))
```

Removing trips with a duration less than zero.

```
annual_trips_clean <- annual_trips_clean %>%
  filter(!(ride_duration < 0))
```

Check for trips with duplicate ride id. No duplicates found.

```
duplicate_check <- annual_trips_clean %>%
  count(ride_id) %>%
  filter(n > 1)
```

Instantiate new dataframe of all unique stations.

```
all_stations <- annual_trips_clean %>%
  group_by(start_station_name) %>%
  count(start_station_name)
```

Save cleaned data for visualization in Tableau. Two pertinent tables (annual_trips_clean and all_stations).

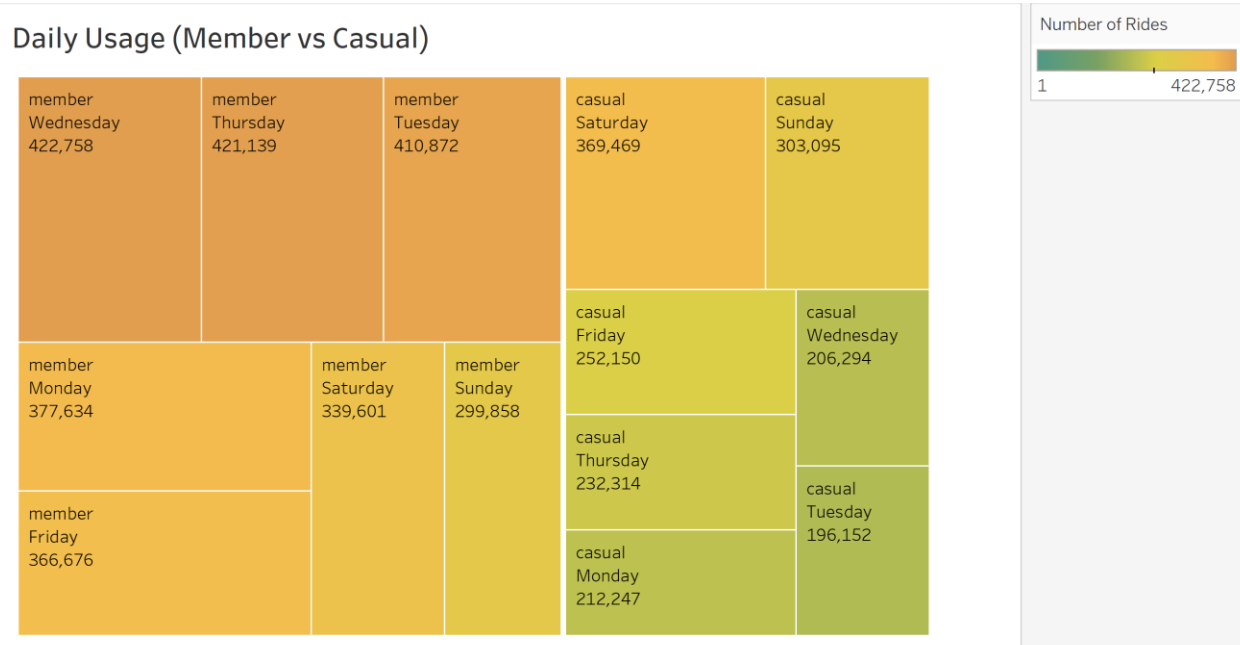
```
fwrite(annual_trips_clean, "annual_trips_clean.csv", col.names = TRUE, row.names = FALSE)

fwrite(all_stations, "all_stations.csv", col.names = TRUE, row.names = FALSE)
```

Analysis

Given the business task provided by the Director of Marketing, this analysis will focus on behavioral differences between *annual members* and *casual riders* relative to Cyclistic's services. The goal of this analysis is to identify opportunities within the current customer base for conversion from *casual riders* to *annual members*.

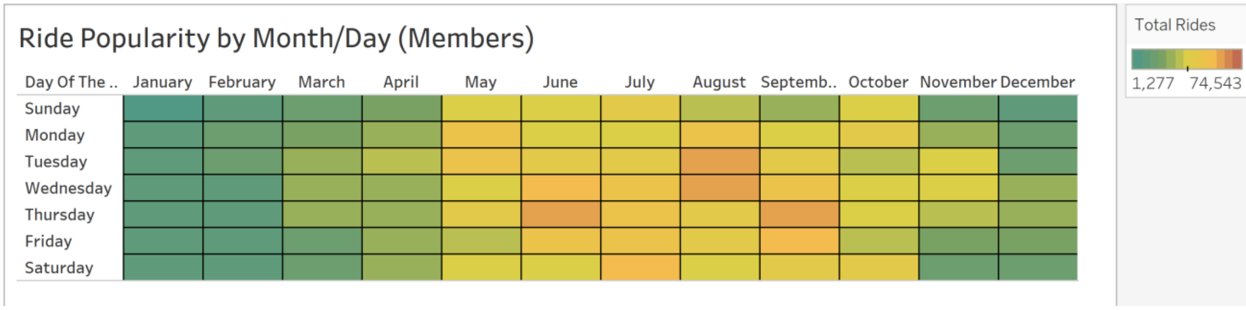
An immediate and fundamental difference between annual members and casual riders' relationship with Cyclistic's services presents itself and will be the focal point of this presentation. While annual members appear to be utilizing Cyclistic's services for leisure/recreation just as their casual counterparts, member utilization of Cyclistic's services for commuting is the defining feature of member behavior.

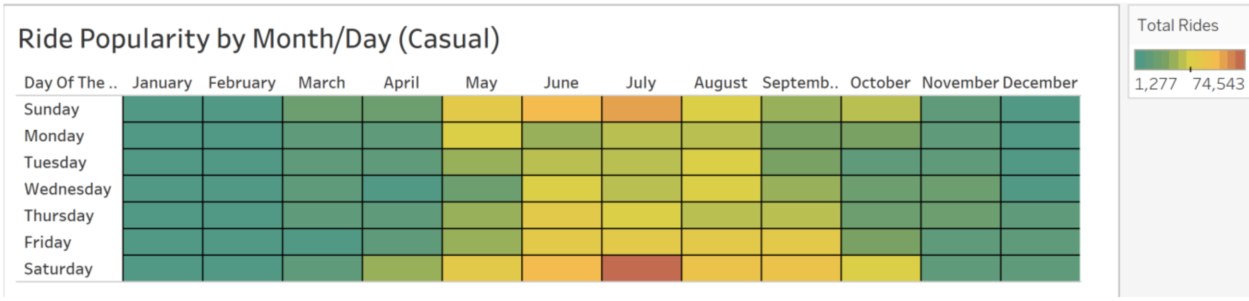
Most popular days of the week



While weekend activity is similar between both member and casual riders, member usage far exceeds that of casual riders for all weekdays with Tuesdays and Wednesdays averaging more than double the number of rides.

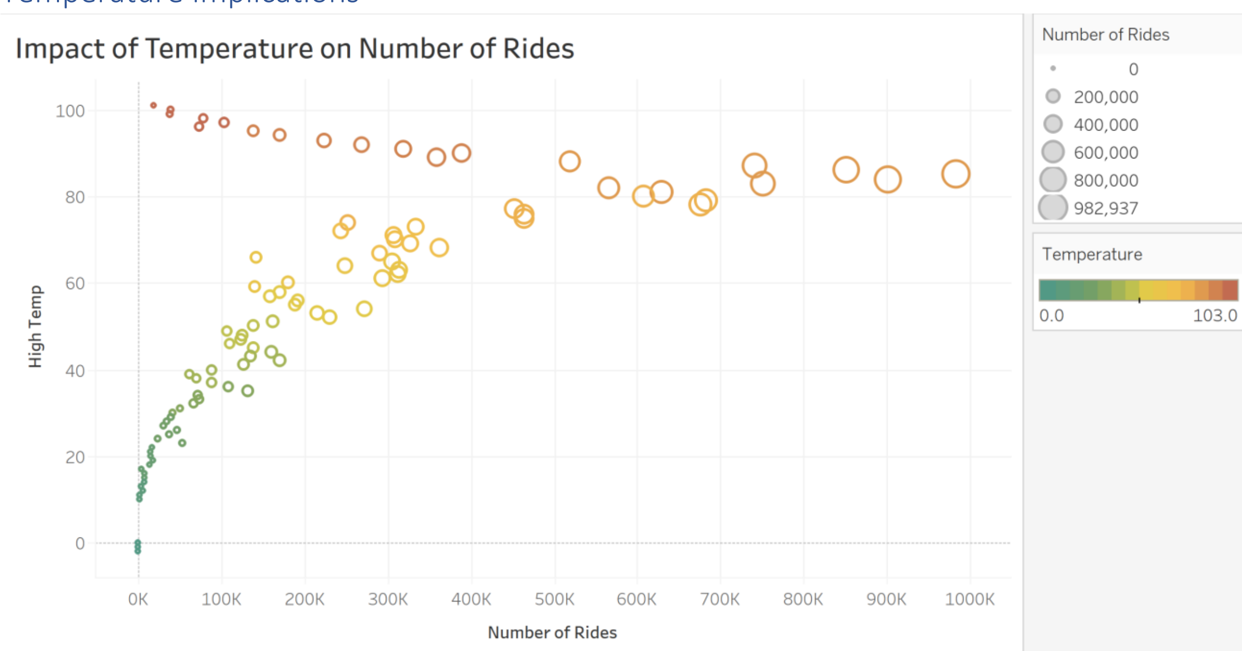
Most Popular Months





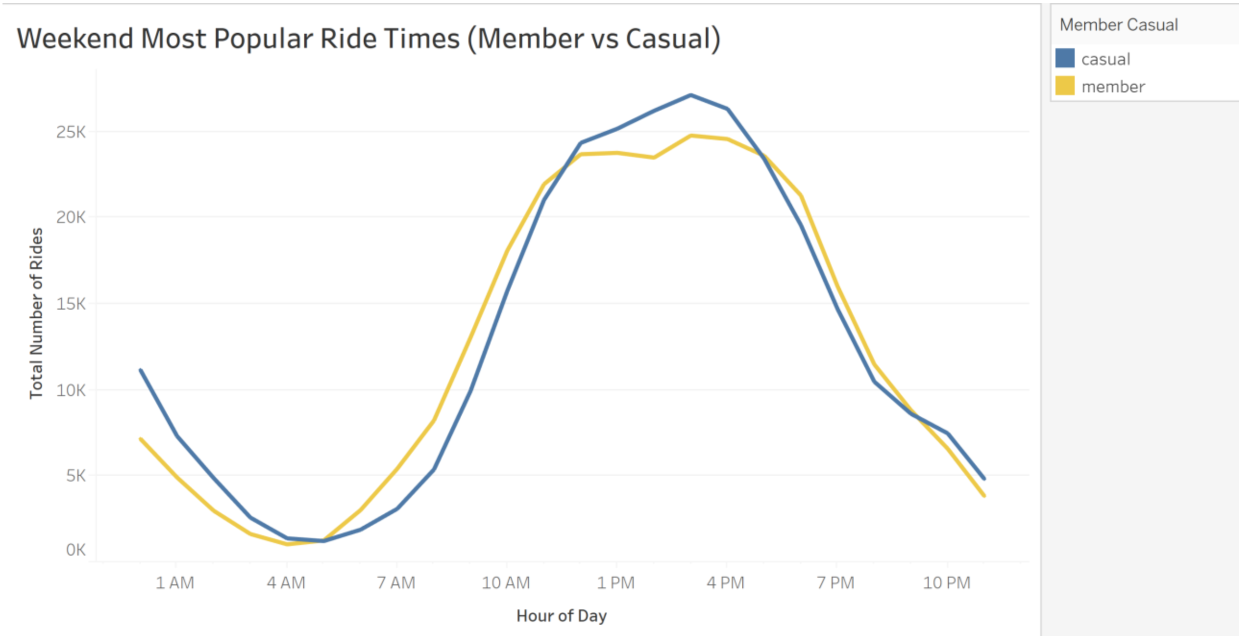
Casual riders and annual members overlap greatly in their preference for the summer months. This preference is further reinforced by the 'Impact of Temperature on Number of Rides' scatterplot (below). The heatmap most notably illustrates the relatively narrow band of usage among casual riders, primarily consisting of June-August. This is particularly important as it illustrates revenue opportunities for the company. While member revenue remains consistent throughout the year, casual revenue falls greatly from October-April (dark green rectangles).

Temperature Implications

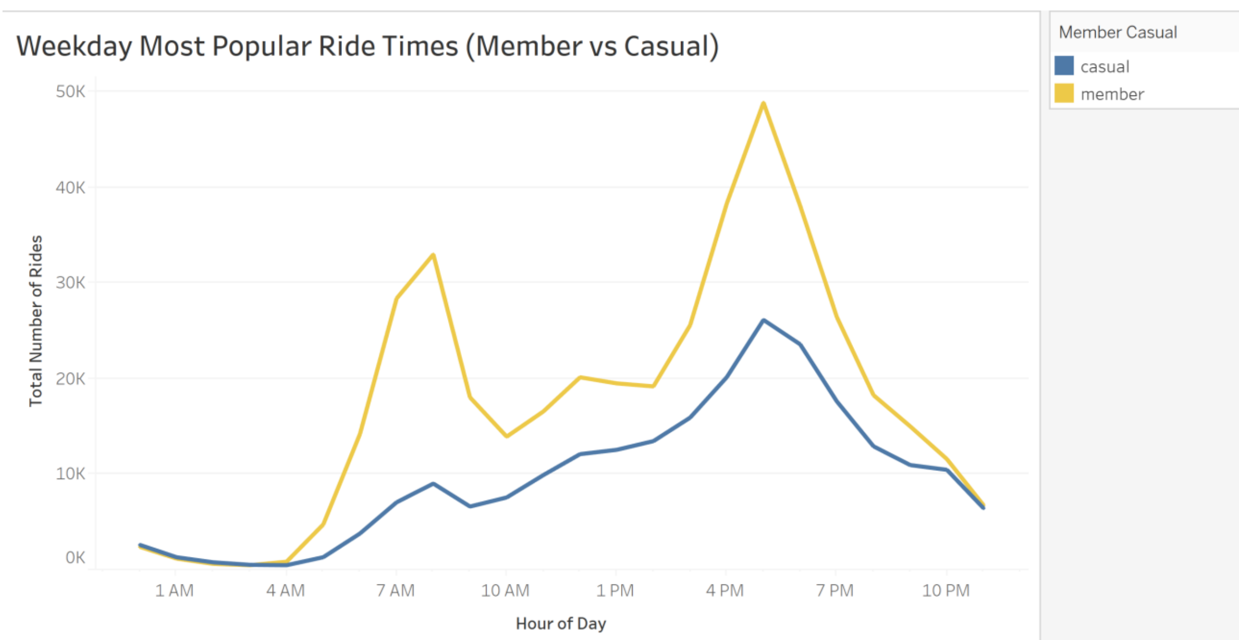


A positive correlation between temp and ride quantity. All 500k ride days exist between 78-88 degrees.

Most Popular Times

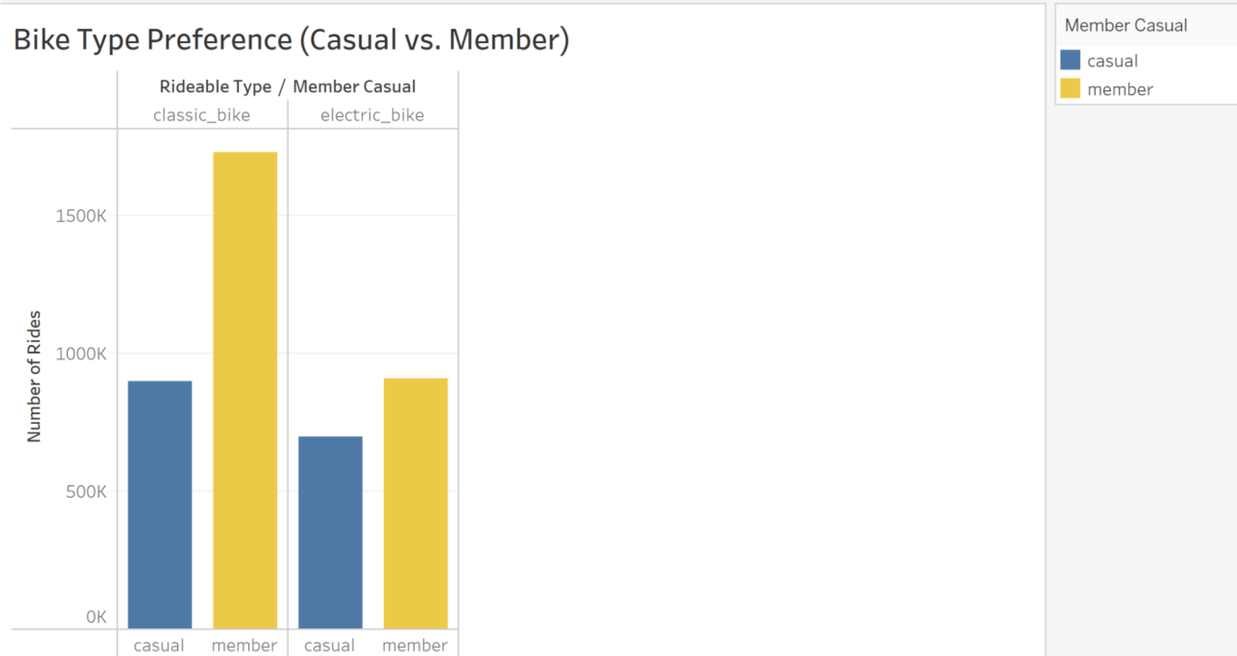


There is a synonymity between members and casual rider behavior during weekend days with usage peaking at 3:00pm.



A divide between members and casual riders reveals itself on weekday usage, with members far exceeding casual usage around the hours of 8:00am and 5:00pm. These times correlate with common work commute patterns. These same peaks, albeit greatly reduced in magnitude, are present in casual users and illustrate a *critical marketing opportunity*, as they suggest that a significant subset of casual riders utilize Cyclistic's services for commute and can benefit from membership with the company.

Most Popular Bike Type



Both members and casual riders express a preference for classic bikes over electric bikes. While this preference is slight among casual riders, there is a near 2:1 preference for classic bikes among members. (Further data will be required to examine this preference in greater detail, but this *may* highlight health/fitness as a motivating factor among members.)

Conclusions

Commuter Value

Significant opportunities exist within those casual riders who are currently using Cyclistic's services as a means of commuting to and from work. Increased value and convenience of membership utilization should be at the forefront of these strategies. These opportunities may be compounded by employer-based marketing or even employer subsidized membership for the downtown Chicago area.

Direct Digital Marketing

The increase in popularity of smartwatches/fitness trackers has created an excellent platform by which to market Cyclistic's services. With targeted ad campaigns through email and notification, direct to the wrist of perspective members, reminding fitness-conscious casual users that "Cyclistic members burned over 2.7 million calories during this morning's commute" (average Thursday morning) helps to solidify Cyclistic as a lifestyle brand rather than a casual institution.

Rideable Optics

While additional data is required to examine why a classic bike preference exists among all users, the data clearly indicates that this preference exists. Marketing should strive to feature the classic bike where applicable.

Considerations beyond the scope of this analysis

The constraints of this analysis prohibited the utilization of data pertaining to customer identity. As such, no analysis was conducted regarding age, gender, or location of the individual customers. Further marketing insights may exist through the analysis of these variables.

Dynamic Visualizations

The dynamic visualization used to support this document can be explored further here.