Cyclistic annual members vs. casual riders

Rachel Wooliver

2025-03-27

Business Task

Cyclistic is a bikeshare company that would like to convert casual riders that pay per-ride or per-day into annual members that pay an annual fee to ride any day of the year. Annual memberships are both more cost-effective for customers and profitable for Cyclistic. The first business task is to *determine how annual members use bikes differently from casual riders.*

Data Sources

Download the last 12 months (March 2024-February 2025) of Divvy trip data found at https://divvy-tripdata.s3.amazonaws.com/index.html, and store in a subdirectory titlee "raw-files". The data have been made available by Motivate International Inc. under this license.

Note that riders' personally identifiable information is of course not publicly available. This means that we won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

This data source seems to be reliable, original, comprehensive, current, and cited.

Read the data into R:

```
# load tidyverse package
library(tidyverse, verbose = F)

# list all files
files <- as.list(dir(pattern = "divvy-tripdata", path = "raw-data", full.names = T))

# combine all of the data files into one table
trip_data <- files %>%
  # read in all of the csv's in the files list
  map_dfr(read_csv, show_col_types = FALSE)
```

Data Cleaning & Manipulation

```
library(skimr)
library(lubridate)

# Remove lat and long data which are outside the continental USA
trip_data <- trip_data %>%
  filter(end_lat < 50, end_lat > 25,
         end_lng > (-125), end_lng < (-70))
```

```r
# Calculate trip duration
trip_data <- trip_data %>%
  mutate(duration_minutes = as.numeric(ended_at - started_at)/60)
# Make sure there are no negative trip durations
trip_data <- trip_data %>%
  mutate(duration_minutes = ifelse(duration_minutes < 0, NA, duration_minutes))

# Determine if the bike is returned to the same station
trip_data <- trip_data %>%
  mutate(start_end_lng_same = (end_lng == start_lng),
         start_end_lat_same = (end_lat == start_lat),
         start_end_same = (start_end_lng_same == TRUE & start_end_lat_same==TRUE))

# Create a column for start and end stations
trip_data <- trip_data %>%
  mutate(start_end_names = paste(start_station_name, "to", end_station_name))

# Create a column for month and time of day
trip_data <- trip_data %>%
  mutate(
    weekday = wday(started_at, label = TRUE, abbr = TRUE),  # Extract weekday (abbreviated)
    month = month(started_at, label = TRUE, abbr = TRUE),    # Extract month (abbreviated)
    hour_of_day = hour(started_at)  # Extract hour of day
  )
```

Summary of Analysis

Question 1: How do total number of rides, bike type, and trip duration differ between casual riders and annual members?

Approach: Calculate the mean duration of bike ride and number of bike rides by member type and bike type.

```
## # A tibble: 2 x 4
##   member_casual mean_duration_minutes   count percent_of_rides
##   <chr>                         <dbl>   <int>            <dbl>
## 1 casual                         21   2126277             36.8
## 2 member                         12.1 3650073             63.2
```

```
## # A tibble: 6 x 5
## # Groups:   member_casual [2]
##   member_casual rideable_type    mean_duration_minutes   count percent_of_rides
##   <chr>         <chr>                            <dbl>   <int>            <dbl>
## 1 casual        classic_bike                      29.4  945084             16.4
## 2 casual        electric_bike                     14.5 1095978             19
## 3 casual        electric_scooter                  11.9   85215              1.5
## 4 member        classic_bike                      13.4 1669343             28.9
## 5 member        electric_bike                     11.1 1921608             33.3
## 6 member        electric_scooter                   8.2   59122              1
```

Results:

- Casual riders make up 36.8% of total rides, and annual members make up 63.2% of total rides.
- Electric bikes are the most-used, while electric scooters are the least-used type of ride for both casual riders and annual members.

- Ride duration differs between member types on classic bikes. Casual members on classic bikes have over two times longer rides (29.4 minutes) compared to annual members on classic bikes (13.4 minutes). The ride duration is about the same between member types for electric bikes (14.5 vs. 13.4 minutes for casual vs. annual members) and electric scooters (11.9 vs. 8.2 minutes for casual vs. annual members).

Question 2: How do start and end locations differ between casual riders and annual members?

Approach: Determine the most popular starting stations by member type.

```
## # A tibble: 880,389 x 6
## # Groups:   member_casual, start_station_name, start_lat [689,400]
##    member_casual start_station_name   start_lat start_lng count percent_of_rides
##    <chr>         <chr>                    <dbl>     <dbl> <int>            <dbl>
##  1 casual        Streeter Dr & Grand~      41.9     -87.6 44725              0.8
##  2 casual        DuSable Lake Shore ~      41.9     -87.6 29785              0.5
##  3 member        Kingsbury St & Kinz~      41.9     -87.6 24449              0.4
##  4 member        <NA>                      41.9     -87.6 22991              0.4
##  5 casual        Michigan Ave & Oak ~      41.9     -87.6 21506              0.4
##  6 member        Clinton St & Washin~      41.9     -87.6 20712              0.4
##  7 member        Clark St & Elm St         41.9     -87.6 20278              0.4
##  8 casual        DuSable Lake Shore ~      41.9     -87.6 20041              0.3
##  9 member        Clinton St & Madiso~      41.9     -87.6 19575              0.3
## 10 casual        Millennium Park           41.9     -87.6 19337              0.3
## # i 880,379 more rows
```

```
## # A tibble: 6,319 x 6
## # Groups:   member_casual, end_station_name, end_lat [4,625]
##    member_casual end_station_name         end_lat end_lng count percent_of_rides
##    <chr>         <chr>                      <dbl>   <dbl> <int>            <dbl>
##  1 casual        Streeter Dr & Grand Ave     41.9   -87.6 54297              0.9
##  2 casual        DuSable Lake Shore Dr &~    41.9   -87.6 31641              0.5
##  3 member        Kingsbury St & Kinzie St    41.9   -87.6 29993              0.5
##  4 member        Clinton St & Washington~    41.9   -87.6 27413              0.5
##  5 casual        DuSable Lake Shore Dr &~    41.9   -87.6 26385              0.5
##  6 casual        Michigan Ave & Oak St       41.9   -87.6 25561              0.4
##  7 member        Clinton St & Madison St     41.9   -87.6 25427              0.4
##  8 member        Clark St & Elm St           41.9   -87.6 24406              0.4
##  9 casual        Millennium Park             41.9   -87.6 23978              0.4
## 10 member        <NA>                        41.9   -87.6 22310              0.4
## # i 6,309 more rows
```

```
## # A tibble: 4 x 4
## # Groups:   member_casual [2]
##   member_casual start_end_same   count percent_of_rides
##   <chr>         <lgl>            <int>            <dbl>
## 1 member        FALSE          3518749             60.9
## 2 casual        FALSE          1927851             33.4
## 3 casual        TRUE            198426              3.4
## 4 member        TRUE            131324              2.3
```

Results:

- The most popular start and end stations for casual riders include Streeter Dr & Grand Ave, DuSable Lake Shore Dr & Monroe St, and Michigan Ave & Oak St.

- The most popular start and end stations for annual members include Kingsbury St & Kinzie St, Clinton St & Washington Blvd, and Clark St & Elm St.
- One of the most popular start stations for annual members does not have a name (41.89000 N, -87.63000 W).
- Annual members are much more likely to return to a different station as the start station compared to casual riders.

Question 3: How do monthly rides differ between casual riders and annual members?

Approach: Determine the most popular month, day of week, and time of day for bike rides by member type.

```
## # A tibble: 24 x 4
## # Groups:   member_casual [2]
##    member_casual month  count percent_of_rides
##    <chr>         <ord>  <int>            <dbl>
##  1 member        Sep   474284              8.2
##  2 member        Aug   437262              7.6
##  3 member        Jul   428284              7.4
##  4 member        Jun   409411              7.1
##  5 member        Oct   399758              6.9
##  6 member        May   378414              6.6
##  7 casual        Sep   345878              6
##  8 casual        Jul   319582              5.5
##  9 casual        Aug   317562              5.5
## 10 casual        Jun   300195              5.2
## # i 14 more rows
```

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday  count percent_of_rides
##    <chr>         <ord>    <int>            <dbl>
##  1 member        Wed     594272             10.3
##  2 member        Tue     557841              9.7
##  3 member        Thu     555887              9.6
##  4 member        Fri     525711              9.1
##  5 member        Mon     522958              9.1
##  6 member        Sat     480698              8.3
##  7 casual        Sat     443986              7.7
##  8 member        Sun     412706              7.1
##  9 casual        Sun     365337              6.3
## 10 casual        Fri     314299              5.4
## 11 casual        Wed     265787              4.6
## 12 casual        Thu     260748              4.5
## 13 casual        Mon     248552              4.3
## 14 casual        Tue     227568              3.9
```

```
## # A tibble: 48 x 4
## # Groups:   member_casual [2]
##    member_casual hour_of_day  count percent_of_rides
##    <chr>               <int>  <int>            <dbl>
##  1 member                 17 385271              6.7
##  2 member                 16 337034              5.8
##  3 member                 18 302539              5.2
##  4 member                  8 251194              4.3
```

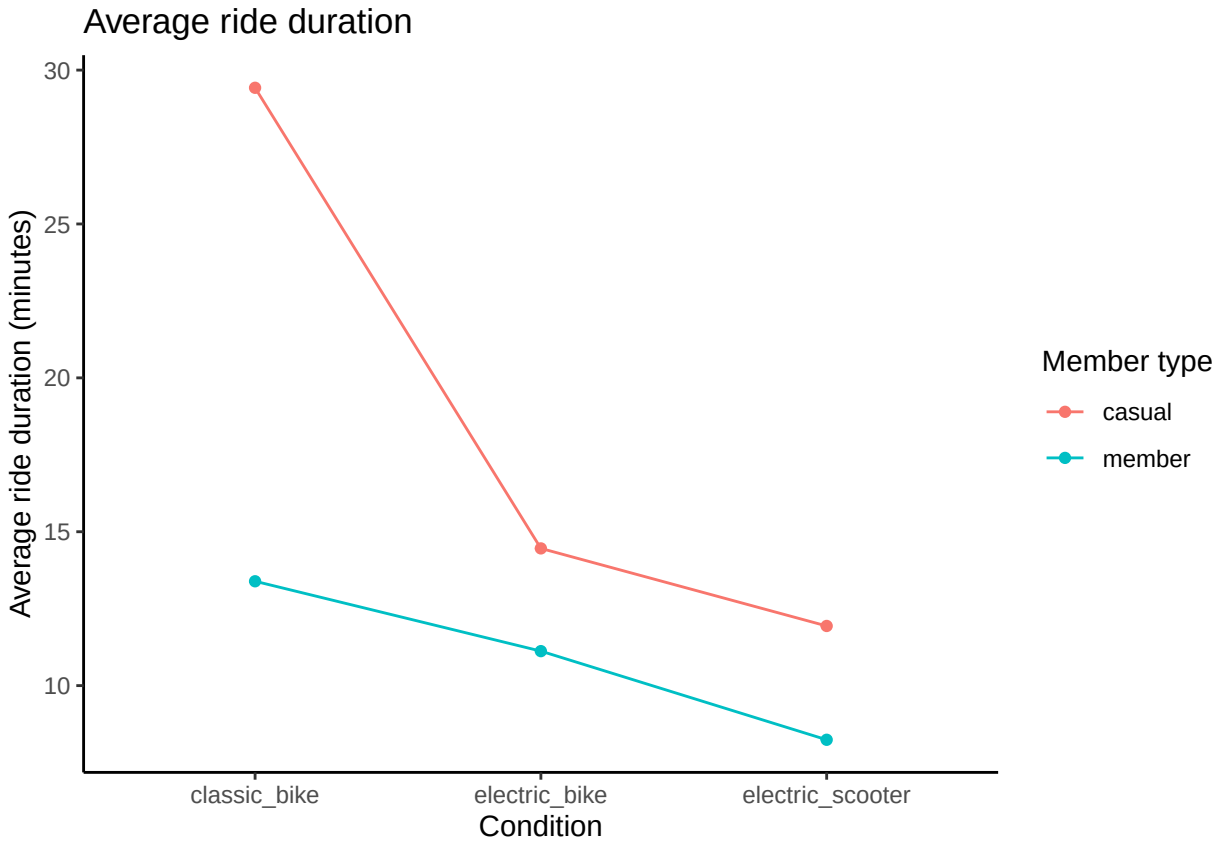
```
##  5 member                15 246339            4.3
##  6 member                19 212585            3.7
##  7 member                12 201588            3.5
##  8 casual                17 201331            3.5
##  9 member                13 201270            3.5
## 10 member                14 200829            3.5
## # i 38 more rows
```

Results:

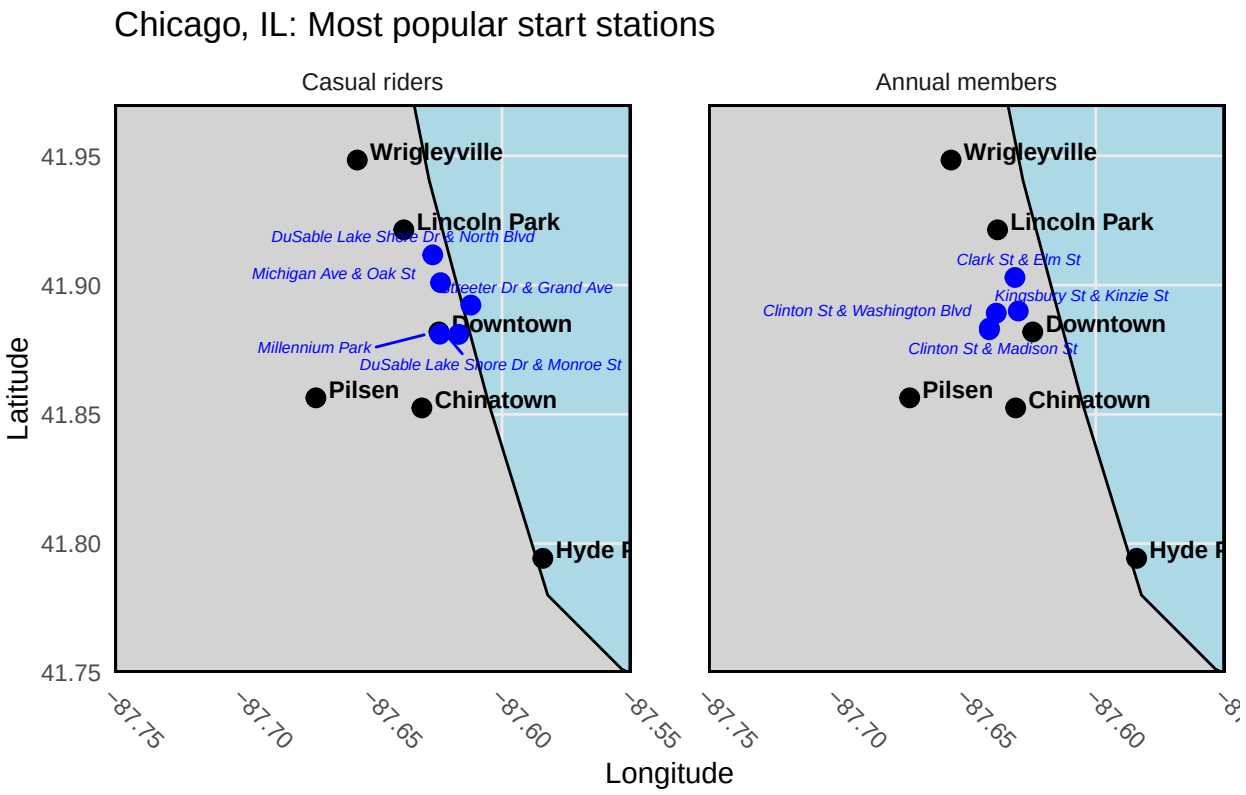
- The most popular time of year to ride is May-October (with September being the most popular month) for both member types.
- Annual members ride most often during the work week, with Tuesday and Wednesday being the most popular days. However, casual members ride most often during the weekend, with Saturday and Sunday being the most popular days.
- Annual members ride most often during rush hour, at 4-6 PM and 8 AM. Casual members also ride most often during 4-6PM, but also ride often at 2-3PM.

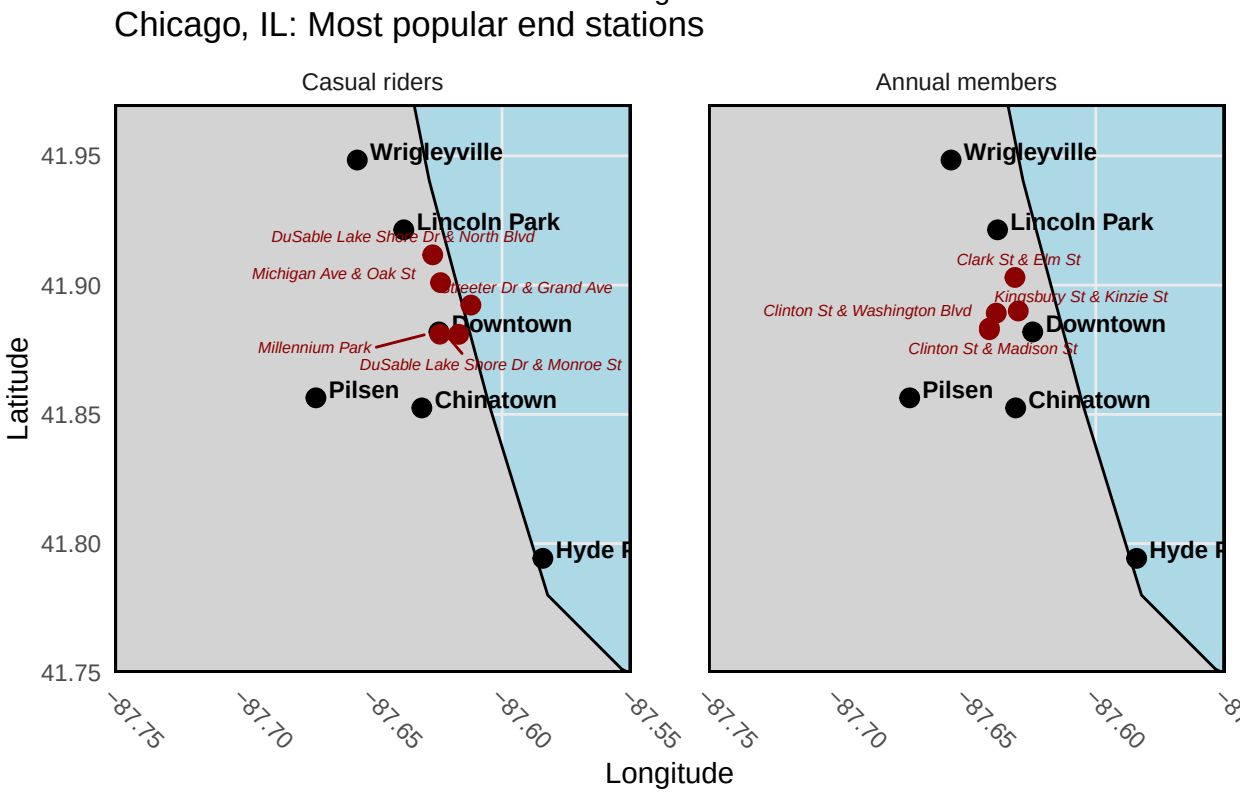
Visualizations & Key Findings

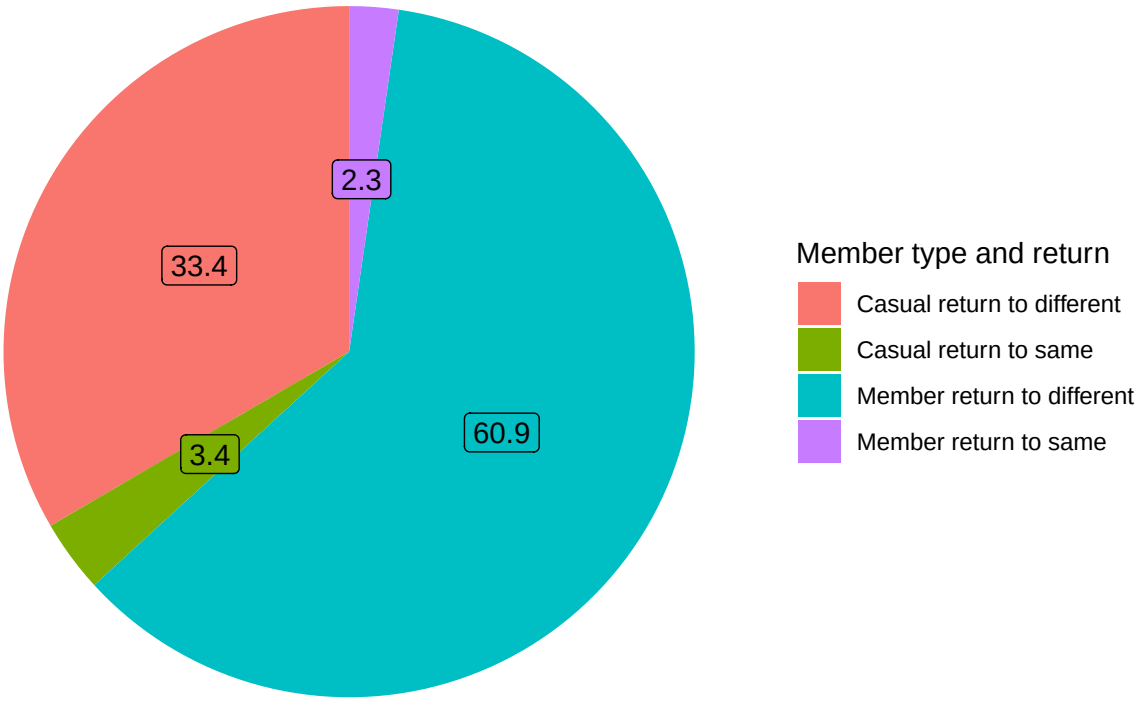
1. Casual riders take longer rides on classic bikes comapred to annual members, but have similar ride duration to annual members using electric bikes.



2. Casual riders tend to start their bike rides on the Downtown waterfront, while annual members tend to start their bike rides inland from downtown and near Hyde Park. Further, casual riders are much more likely to return their bikes to the same station they retrieved from.







3. . . . (most popular month, day of week, and time of day by member type).

Recommendations

In order to switch casual riders to annual members:

1.
2.
3.