Cyclistic Data Analysis

Jeffrey Cuevas

1/3/2022

Overview

I'm a data analyst working on the marketing team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. My team has been assigned the task of understanding how casual riders and annual members use Cyclistic bikes differently. From these insights, our team will design a new marketing strategy to convert casual riders into annual members.

Ask

Three questions will guide the future marketing program:

- How do annual members and casual riders use Cyclistic bikes differently?
- Why would casual riders buy Cyclistic annual memberships?
- How can Cyclistic use digital media to influence casual riders to become members?

This analysis will focus on how annual members and casual riders use Cyclistic bikes differently.

Prepare

I used Cyclistic's historical trip data to analyze and identify trends. Data used for this analysis consists of 12 individual .csv files representative of different months within the last year. Data was merged and prepped using Pyspark within a Jupyter Notebook. The output of the Jupyter Notebook will be used for analysis within R Studio.

Process

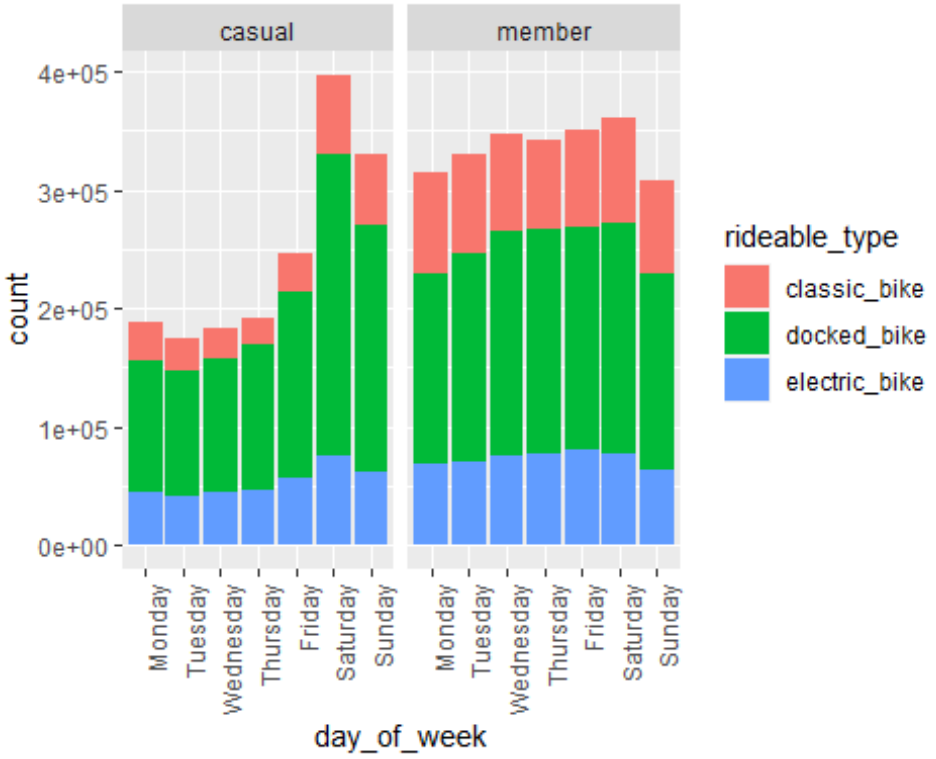
List of operations used for processing the data:

1. Read each file into a dataframe and merge all files into one dataframe.
2. Removed duplicate rows (there were none).
3. Created new column called "distance_traveled" that uses Haversine formula to calculate distance between two points based on latitude and longitude.
4. Created new column called "date_diff" that calculates the day difference between ride start time and end time.
5. Removed records where "date_diff" < 0 because those are impossible scenarios.

6. Created new column called “duration_mins” that calculates the minute difference between ride start time and end time.
7. Removed records where “duration_mins” < 0 because those are impossible scenarios.
8. Created new column called “day_of_week” using pyspark date_format function.
9. Calculated frequency distribution for day of the week.
10. Calculated frequency distribution for casual and annual members.
11. Calculated frequency distribution for bike types.
12. Removed unnecessary columns and exported to a .csv file.

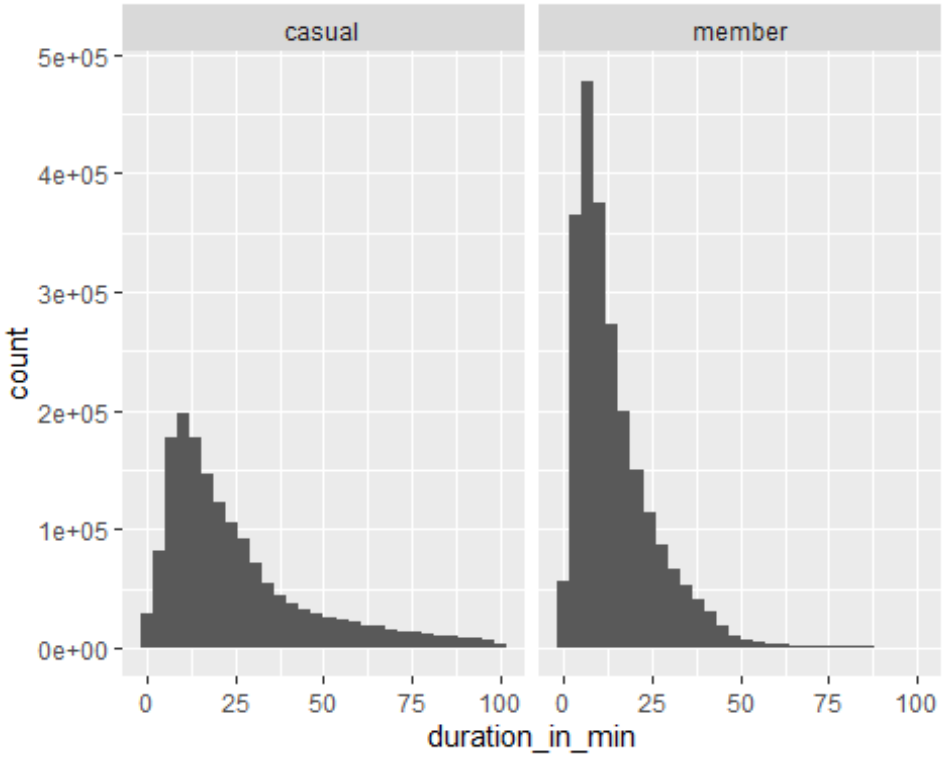
Analyze

Let’s take a look at weekly frequency distribution of the member and casual customers along with bike types:



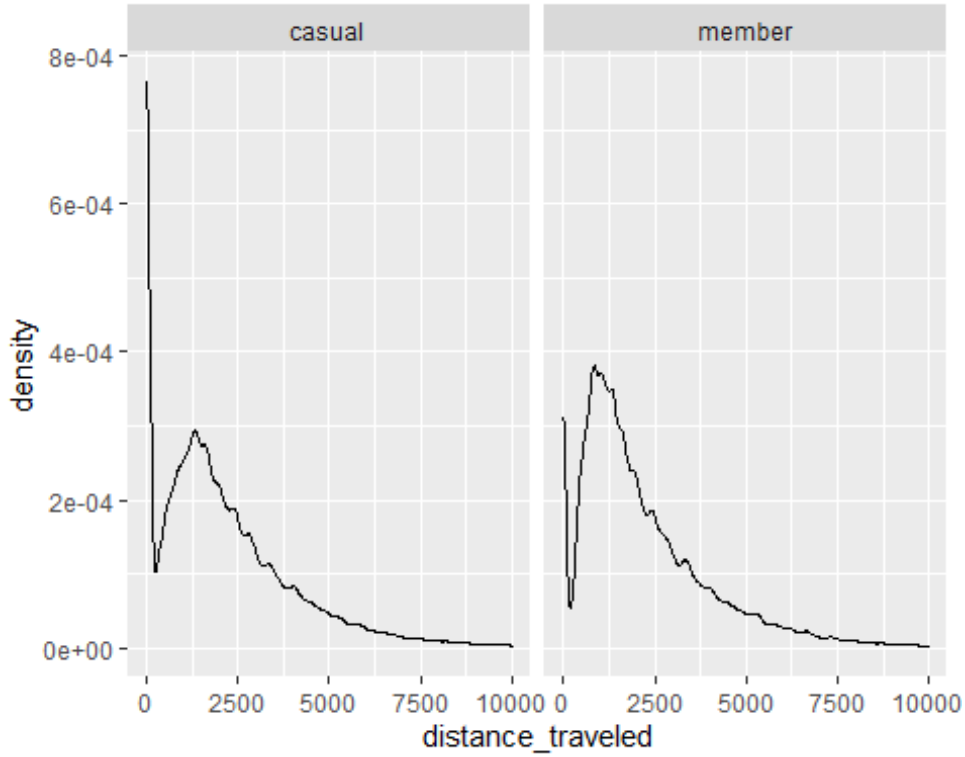
- We can infer that members are primarily made up of working people since the distribution is pretty even throughout the week and lower on Sunday.
- Distribution for casual shows a pattern of high usage on the weekends and low usage during the week.
- Docked bikes have the highest usage among both groups.

Now, let's observe trip duration behavior for member and casual:



- Casuals take longer trips than members and members take shorter trips than casuals.
- The majority of both groups take shorter trips.

Let's plot distance traveled in meters for casuals and members:



- It is difficult to make assumptions based on this plot.

We can observe a few summary statistics for both member and casual groups:

Members

```
   day_of_week     distance_traveled duration_in_min
 Monday   :315143   Min.   :    0.0   Min.   :    0.00
 Tuesday  :329035   1st Qu.:  940.5   1st Qu.:    6.32
 Wednesday:346757   Median : 1687.0   Median :   11.07
 Thursday :341550   Mean   : 2248.1   Mean   :   15.26
 Friday   :350229   3rd Qu.: 3018.5   3rd Qu.:   19.30
 Saturday :360218   Max.   :48370.8   Max.   :41271.00
 Sunday   :307507
```

Casuals

```
   day_of_week     distance_traveled duration_in_min
 Monday   :188152   Min.   :    0.0   Min.   :    0.00
 Tuesday  :174145   1st Qu.:  714.5   1st Qu.:   11.03
 Wednesday:182292   Median : 1660.8   Median :   20.20
 Thursday :191432   Mean   : 2185.4   Mean   :   41.83
 Friday   :246275   3rd Qu.: 3018.5   3rd Qu.:   38.35
```

```
 Saturday :395857   Max.   :33800.2   Max.   :54283.35
 Sunday   :329438
```

- Mean distance traveled is relatively the same for both members and casuals.
- Median distance traveled is relatively the same for both members and casuals.
- Mean duration for casuals is almost three times more than members.
- Median duration for casuals is almost twice as much as members.

Next, lets find the most popular start and end stations with their frequency for members:

Most Popular Start Stations for Members

```
         start_station_name      n
1              missing_data 119269
2          Clark St & Elm St  22633
3      Wells St & Concord Ln  17679
4        Theater on the Lake  17370
5       Broadway & Barry Ave  17309
6      Dearborn St & Erie St  17186
7  Kingsbury St & Kinzie St  17084
8    St. Clair St & Erie St  16772
9          Wells St & Elm St  16522
10       Wells St & Huron St  16113
```

Most Popular End Stations for Members

```
           end_station_name      n
1              missing_data 127946
2          Clark St & Elm St  23036
3      Wells St & Concord Ln  18037
4    St. Clair St & Erie St  17852
5      Dearborn St & Erie St  17798
6       Broadway & Barry Ave  17487
7  Kingsbury St & Kinzie St  17188
8        Theater on the Lake  16860
9          Wells St & Elm St  15860
10       Wells St & Huron St  15132
```

- Members tend to start and end trips from the same stations.

Most Popular Start Stations for Casuals

```
           start_station_name     n
1                missing_data 82656
2     Streeter Dr & Grand Ave 36559
3   Lake Shore Dr & Monroe St 28233
4              Millennium Park 24808
5          Theater on the Lake 18565
6        Michigan Ave & Oak St 18362
7  Lake Shore Dr & North Blvd 16868
8  Indiana Ave & Roosevelt Rd 15884
9      Michigan Ave & Lake St 13927
10              Shedd Aquarium 13869
```

Most Popular End Stations for Casuals

```
                end_station_name      n
1                   missing_data 101092
2        Streeter Dr & Grand Ave  39507
3      Lake Shore Dr & Monroe St  27169
4                 Millennium Park  25738
5            Theater on the Lake  20801
6          Michigan Ave & Oak St  19047
7     Lake Shore Dr & North Blvd  17991
8     Indiana Ave & Roosevelt Rd  15899
9         Michigan Ave & Lake St  13328
10 Michigan Ave & Washington St  12944
```

- Casuals tend to start and end trips from the same stations.

We can observe the most popular routes by looking at the most frequent combinations of start station and end station. #### Most Popular Member Routes

```
                                                                routes     n
1                                          missing_data - missing_data 68642
2                              Ellis Ave & 60th St - Ellis Ave & 55th St  1409
3                               MLK Jr Dr & 29th St - State St & 33rd St  1383
4                              Ellis Ave & 55th St - Ellis Ave & 60th St  1316
5                               State St & 33rd St - MLK Jr Dr & 29th St  1247
6  Lakefront Trail & Bryn Mawr Ave - Lakefront Trail & Bryn Mawr Ave  1192
7                                        Burnham Harbor - Burnham Harbor  1167
8                                     Montrose Harbor - Montrose Harbor  1131
9                              Theater on the Lake - Theater on the Lake  1123
10          Lake Shore Dr & Belmont Ave - Lake Shore Dr & Belmont Ave  1120
```

Most Popular Casual Routes

```
                                                       routes     n
1                                missing_data - missing_data 49062
2        Streeter Dr & Grand Ave - Streeter Dr & Grand Ave  8230
3    Lake Shore Dr & Monroe St - Lake Shore Dr & Monroe St  7910
4                         Millennium Park - Millennium Park  6248
5                 Buckingham Fountain - Buckingham Fountain  5726
6            Michigan Ave & Oak St - Michigan Ave & Oak St  4734
7  Indiana Ave & Roosevelt Rd - Indiana Ave & Roosevelt Rd  4272
8  Fort Dearborn Dr & 31st St - Fort Dearborn Dr & 31st St  3870
9                 Theater on the Lake - Theater on the Lake  3616
10           Michigan Ave & 8th St - Michigan Ave & 8th St  3562
```

Summary of All Findings

User_type	Amount	Avg_and_median_trip_duration	Avg_and_median_trip_distance	Busiest_day	Preferred_bike_type	Most_occured_route
Member	2,352,923 (57.9%)	15.26 min - 11.07 min	2.25 km - 1.69 km	Saturday	docked bike	Ellis Ave & 60th St - Ellis Ave & 55th St (1,409)
Casual	1,710,107 (42.1%)	41.83 min - 20.20 min	2.19 km - 1.66 km	Saturday	docked bike	Streeter Dr & Grand Ave - Streeter Dr & Grand Ave (8,230)

Share

The following are my main observations of this analysis:

- Members tend to take more rides during the week with a preference for docked bikes. This leads me to believe that members are using Cyclistic bikes for commuting.
- Casuals tend to take more rides on the weekends with a preference for docked bikes.
- The majority of trips for both members and casuals is less than 25 minutes. We can infer both user groups or using Cyclistic's bikes for short trips.
- Average distance traveled for both members and casuals is about the same, meaning there isn't much to draw from that.
- Average trip duration for casuals is almost three times higher than that of members.
- Both members and casuals appear to start and end trips from the same station.

Act

How can we convert casuals to members? There are a few ways to go about accomplishing this:

1. Offer weekend signup promotions for casual users. This could be a limited time discount offer or a $0 signup fee. Target the high traffic days (Friday - Sunday).
2. Provide an incentive program for casuals to signup for memberships. This could come in the form of partnering with local businesses that would provide a free good or service upon signup.
3. Increase bike rental fees, especially for docked bikes, on the weekends to force casual members to consider buying a membership.