



Cyclistic: Bike Sharing Program

Case Study



Case Study: How Does a Bike-Share Navigate Speedy Success?

Dip Patel

Scenario:

Cyclistic, a bike-share company in Chicago that features more than 5,800 bicycles and 600 docking stations. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, management wants to understand how casual riders and annual members use Cyclistic bikes differently. And from this analysis, new marketing strategy will decide how to convert casual rider into annual member.

Table Of Content:

- 1) Business Task
- 2) Data Source
- 3) Data Cleaning and manipulation
- 4) Analysis
- 5) key findings.
- 6) Best recommendation

1.ASK

Business Task

How do annual members and casual riders use Cyclistic bikes differently? What things must be done to convert casual member into annual member.

Lily Monero:

The director of marketing and responsible for development of campaigns and initiatives to promote the bike-share program.

Stakeholder perspective:

Monero believes company's future success depends on maximizing the number of annual memberships. She believes rather than creating a marketing campaign targeting all new customers, there is a very good chance to convert casual riders into members

2.Prepare

Data Source

This [data](#) is provided by Cyclistic itself. (Note: The datasets have a different name because Cyclistic is a fictional company.) Data is stored locally.

The data has been made available by Motivate International Inc. under this [license](#).

Limitation:

Data are provided for only particular period of time, it may be not have reflected whole picture. Data is available for only April month; our data will tell a story for particular time.

Another consideration made is all data provided by company are real.

data-privacy issues prohibit me from using riders' personally identifiable information. It means that I 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.

3.Process

Data Cleaning and manipulation

In this step we will clean and convert data for analysis, for that we perform certain preparations on the data.

- Microsoft Excel
1. Highlight primary key and put boundary on table.
 2. Checked for any duplicity of data. By using Conditional formatting.
 3. Some data are missing, which are replaced with NA. I have used filter to find out empty data. After that used find and replace function to fill empty place with null value. We have total 1186878 field of data available. Out of that 396 filed has no data. Which is 0.033% of total data. Which is acceptable.
 4. For checking ride_id correctly, use LEN function to count number of characters. It should be 16. If not then it is incorrect and should be corrected or deleted.

We have found following anomalies and deleted this garbage data.

ride_id	len(ride_id)
2.37907E+22	19
4.28416E+14	15
6.38012E+16	17
6.61171E+14	15
9.59599E+14	15
7.39861E+22	19
4.20238E+20	19
9.0667E+105	19
6.55391E+22	19
8.5404E+14	15
7.11981E+14	15
8.98391E+33	18
3.0177E+57	18
7.83505E+96	18

5. Put Filter for primary keys.
6. New column created name **ride_length**, which is nothing but end time – start time. It gives duration of ride length. Result converted into HHMMDD format and Extended for all data.

Sometimes data will miss guide our hypothesis if data is not properly cleaned. I have identified ride_length which have duration less than 1 minute with help of **IF function**. They can be considered as false riding. Total 975 rows data are found which has ride length less than 1 minute.

Another criterion is placed for removing garbage data is, entry is removed for those who have starting point and ending point longitudes are same. It simply indicates that user have taken the service but not availed whole service. For removing this data, I have used certain calculation to find out data.

Have created three column name Lat(latitude), Lon(longitude) and Sum(total). Lat column gives subtraction of ending latitude and starting latitude. If Both are same then answer would be zero, otherwise not. Same Lon column created for longitude data. Sum column gives summation of Lat and Lon column data.

Lat = ending_lat – Starting _lat

Lon = ending_lon – Starting_log

Sum = Lat + Lon

In short, we have deleted rows which have starting location and ending location are same and who has ride length less than 1 minute. We have 50 rows of such data.

7. Create a column called “day_of_week,” and calculate the day of the week that each ride started using the “WEEKDAY” command (for example, =WEEKDAY(C2,1)) in each file. Format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday
8. In final database consist of following characteristics:
 - Ride_id column which has ride id length of only 16 characters.
 - Unnecessary column are removed which are prepared for analysis: len(ride_id), ride_length_status, time -reference, Lat, Lon, Sum.

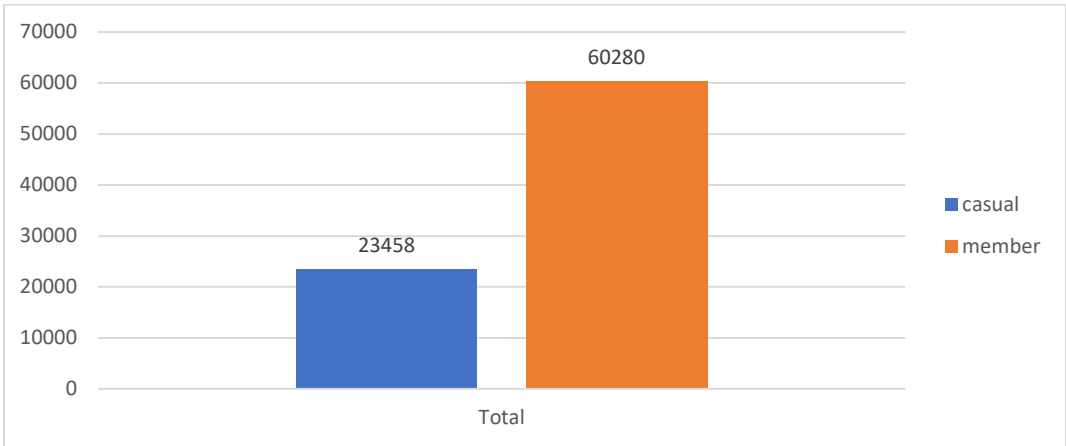
Final clean data available for analysis 83739 rows

4. Analyse

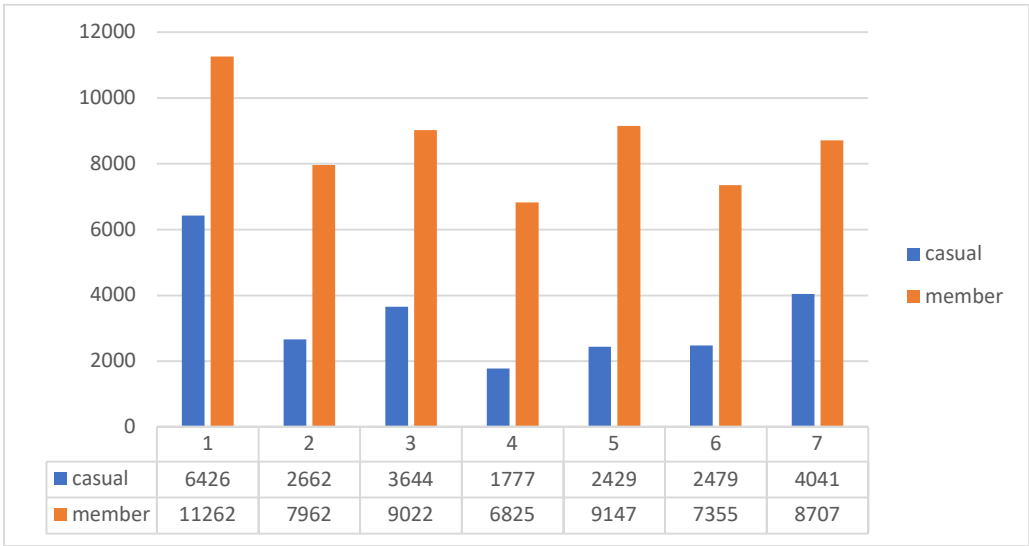
In this stage we will perform analysis on the data that cleaned in previous stage.

Number Of Users:

From the Graph we can see that casual members are very less compare to annual members. But there is still big window of opportunity to increase profit by converting this casual member into annual members.



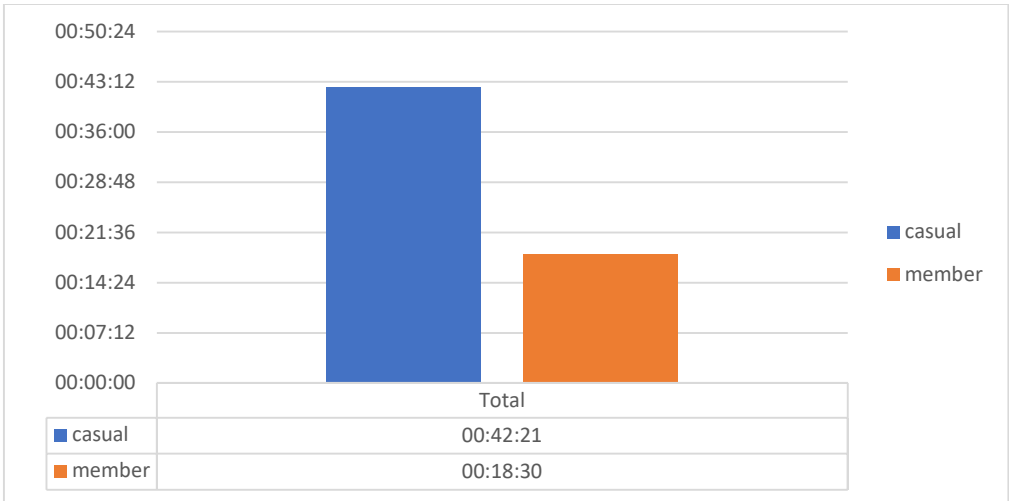
From weekday graph we can conclude that on weekend casual member use bike service often compare to other days of the week. This is vital information for using marketing strategy. (Weekday 1 = Sunday, 7 = Saturday)



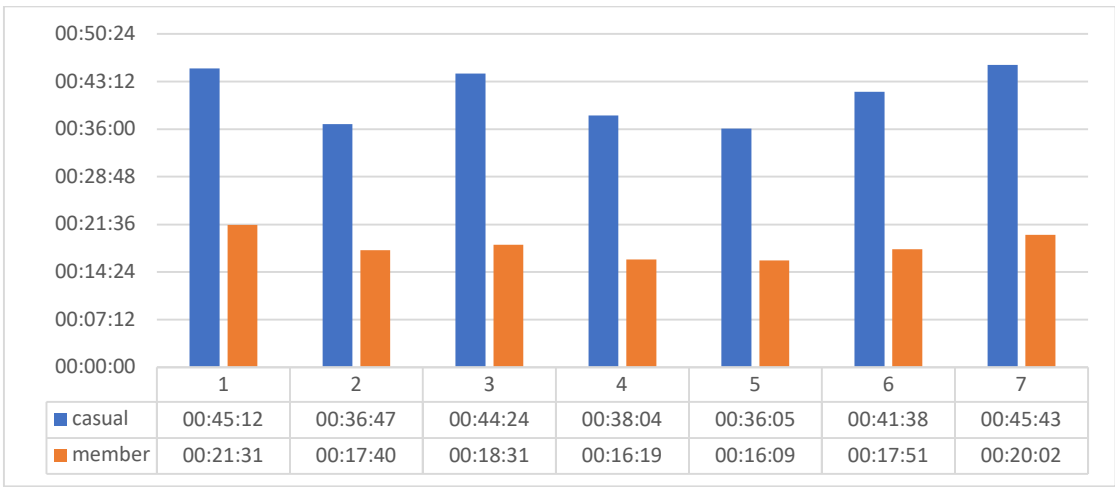
Calculated Average ride length:

Using pivot table, we have calculated average ride length time for annual and casual members. Though casual member is less than annual member in numbers, there mean time very differently. More mean time means their data are more distributed compare to annual member. They have more variation compare to members in ride length and more ride length.

	casual	member	Grand Total
Average of Ride_length	00:42:21	00:18:30	00:25:11

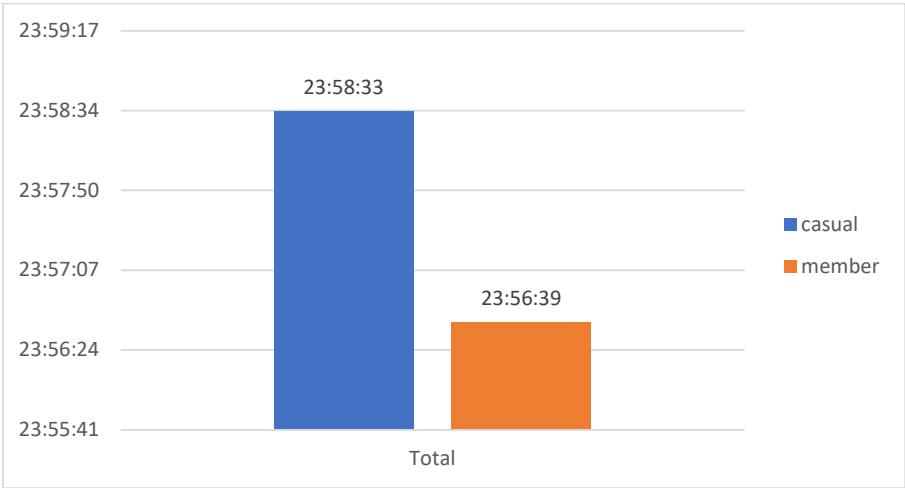


More clarity by weekday graph. Casual member average ride length is higher on every day compare to annual members. We can conclude their ride time varying drastically.

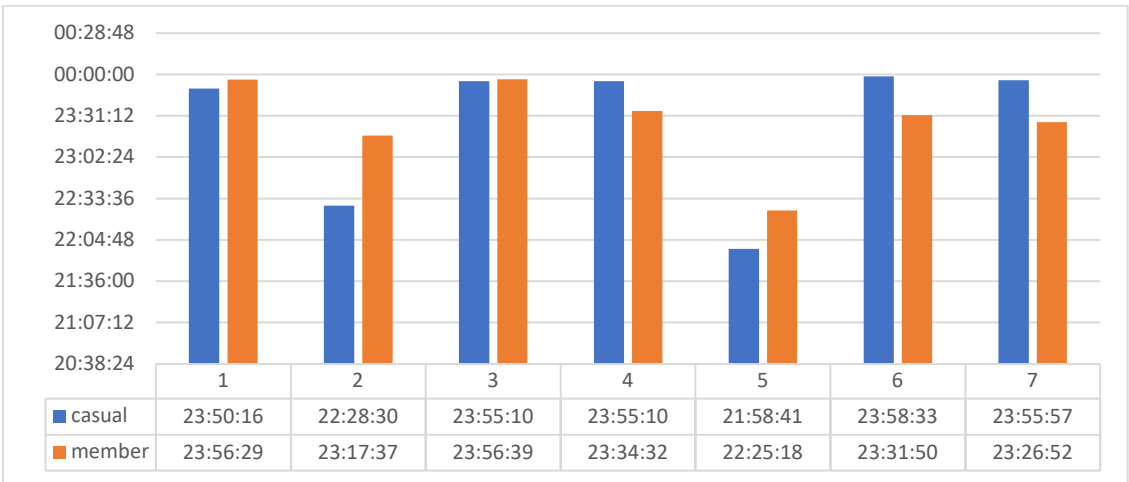


Calculated Max ride length:

We can see that annual member has the highest ride length.



But from the weekday graph we can see that in some weekday casual member has higher ride length compare to annual member.



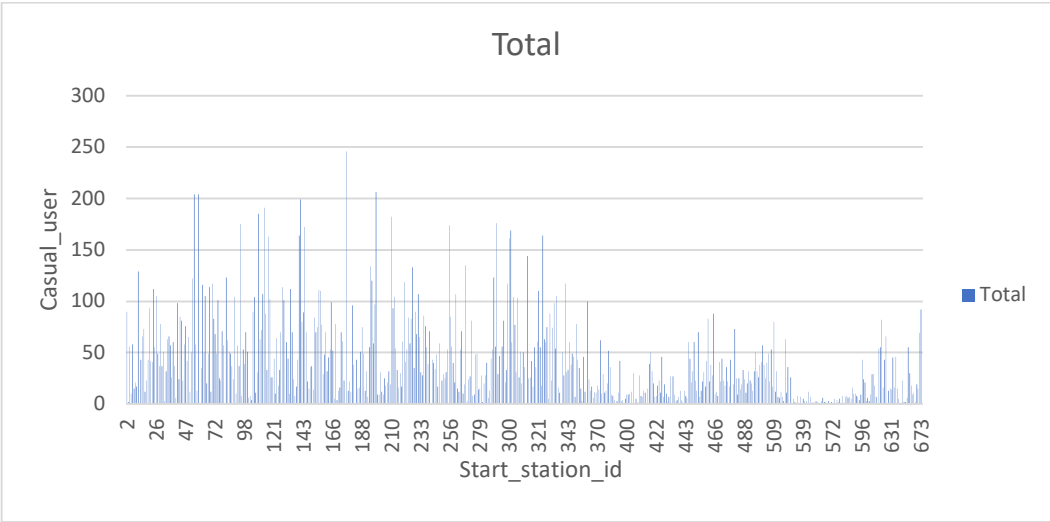
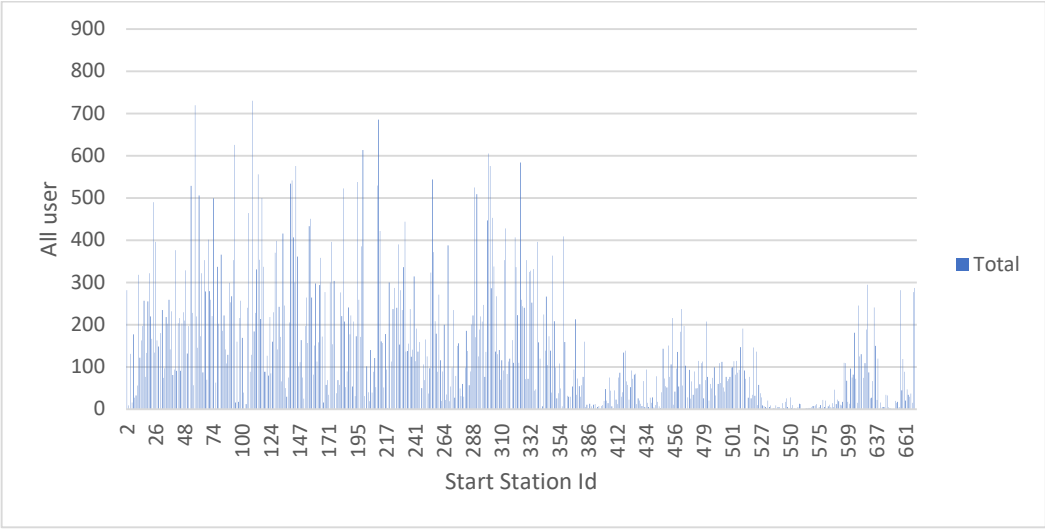
Calculated Mode for day of the week.

1 (Sunday)

Mode gives data about maximum occurrence. It's obvious that maximum ride will occur on Sunday.

Max trips starts from which started point:

From the below graph we can see that station id 176 has maximum users.



6. Share

Key findings:

- 1) Numbers of annual members are higher than casual member, still there is opportunity for company to increase profit by converting casual member into annual member.
- 2) Though annual numbers are more than casual, there average ride time is lower. Casual riders tend to have longer trips but fewer number of trips. As the need for using bike for casual member is not frequent as annual member, casual rider have no incentive to become annual member. By giving proper analysis of usage of ride, we can show user how can he save more money via choosing annual subscription.
- 3) Max ride length for annual member occurs on weekends, while for casual members it occurs on the other day. So casual user may not need bike often, that's why they are not taking annual membership.
- 4) Maximum number of user avail on weekends both annual and casual. When we are using digital platform for publicity, our most advertisement spend should occur on Friday, Saturday and Sunday. Friday night is the time when everyone planning for something.
- 5) Maximum number of starting users take bike from the station id number 176,141, 56...etc. We have information (longitude and latitude) about these stations. We should target this area more often for generate constant customers.

Recommendation:

1) Price Discount for different ride length

Since casual rider have longer rides compare to annual. Company should provide flexible rate for casual user, rather than fix rate. If ride length exceed certain distance user has some monetary benefit. This kind of benefit given to annual member as well. This way more users will attract toward platform. Eventually increase revenue for company. Monthly/Quarterly basis casual user have some statistic report when they can compare how much money they can save via using annual membership.

2) Promotion/Publicity

Social media is effective tools for advertisement. Target particular region which has more users and which area has more potential. Digital platform (Facebook, Instagram, LinkedIn...) are quite effective compare to traditional ways.

3) Marketing Strategy

Create effective add which clearly states that you have bike experience of any bike at any time without having one. Another strategy should be added for renting vehicle, where owner and user both can have benefit.

