Bike share data analysis using excel and R

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

Author: Snehith Tella
Date: 23/12/2022

Data Analysis method: following the APPASA process: Ask - Prepare - Process - Analyze - Share – Act

Data Analysis method: Ask - Prepare - Process - Analyze - Share – Act

Scenario

My role is a junior data analyst working in the marketing analyst 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. Therefore, my team wants to understand how casual riders and annual members use Cyclitic bikes differently. From these insights, my team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve my recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Ask

Business Task: Converting casual riders into membership riders

The goal of the marketing team is to design the marketing strategies at converting casual riders into annual membership members. My task is to answer one of the three questions that will guide the future marketing program – How do annual members and casual members use the bicycle bikes differently?

Prepare

Data source: The data sets of previous 12 months of trip data can download here
NOTE: As the data source files are larger, I used the data consisting of 6 months from 2020-2021. I used R studio to combine the two large datasets which consists of 1 million rows of combining 2020-2021

- 12 csv. Files
- 13 variables.
- The data has been made available by motivate International Inc. under this license
- This is a public data.

ROCCC Analysis

- Reliable – yes, not biased, the data is cleaned, given by the certificate course
- Original- yes, can locate the original public data
- Comprehensive – yes, not missing important information
- Current – yes
- Cited – yes

Process

1. Download the previous 12 months of Cyclistic trip data.

2. Unzip the files.

3. Create a folder on your desktop or Drive to house the files. Use appropriate file-naming conventions.

4. Create subfolders for the .CSV file and the .XLS or Sheets file so that we have a copy of the original data. Move the downloaded files to the appropriate subfolder.

5. Follow these instructions for Excel (a):

      a. Launch Excel, open each file, and choose to Save As an Excel Workbook file. Put it in the subfolder you created.

6. Open the spreadsheet and create a column called "ride_length." Calculate the length of each ride by subtracting the column "started_at" from the column "ended_at" (for example, =D2-C2) and format as HH:MM:SS using Format > Cells >Time > 37:30:55.

started_at	start_time	ended_at	end_time	ride_length	
04/26/20	5:45:00 PM	04/26/20	6:12:00 PM	0:26	
04/17/20	5:08:00 PM	04/17/20	5:17:00 PM	0:08	

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 that1 = Sunday and 7 = Saturday.

H
day_of_the_week
1
6

8. Now, we will Proceed to the analyze step.

Analyze

Now the data is stored appropriately and has been prepared for analysis, Now we will start putting it to work.

Open the spreadsheet application, then complete the following steps:

1. Where relevant, make columns consistent and combine them into a single worksheet.

2. Clean and transform the data to prepare for analysis.

3. Run a few calculations in one file to get a better sense of the data layout.
 ● Calculating the mean of ride_length
 ● Calculating the max ride_length
 ● Calculating the mode of day_of_week

mode_of_week	average_ride_length	max_ride_length
1	12:22:00 AM	3:18:00 PM
1	1:22:00 AM	4:18:00 PM

4. Create a pivot table to quickly calculate and visualize the data.

 ● Calculating the average ride_length for members and casual riders. Try rows =member_casual; Values = Average of ride_length.

Average_ride_length
Row Labels	Count of average_ride_length
casual	23590
member	61093
Grand Total	84683

 ● Calculating the average ride_length for users by day_of_week. Try columns = day_of_week; Rows =member_casual; Values = Average of ride_length.

Count of average_ride_length   Column Labels
Row Labels	1	2	3	4	5	6	7	Grand Total
casual	6468	2677	3651	1794	2426	2507	4066	23589
member	11429	8055	9146	6917	9256	7455	8834	61092
Grand Total	17897	10732	12797	8711	11682	9962	12900	84681

 ● Calculating the number of rides for users by day_of_week by adding Count of trip_id to   Values.

	Column Labels								
	1		2		3		4		5
Row Labels	Count of average_ride_length	Count of ride_id	Count of average_ride_length	Count of ride_id	Count of average_ride_length	Count of ride_id	Count of average_ride_length	Count of ride_id	Count of average_ride_length
casual	6468	44409	2677	24876	3651	24223	1794	24137	2426
member	11429	56725	8055	50616	9146	53721	6917	55501	9256
Grand Total	17897	101134	10732	75492	12797	77944	8711	79638	11682

5. Once we have spent some time working with the individual spreadsheets, merge them into a full-year view. We can do this with the tool you have chosen to use to perform your final analysis, either a spreadsheet, a database and SQL, or R Studio. I chose to use R studio.

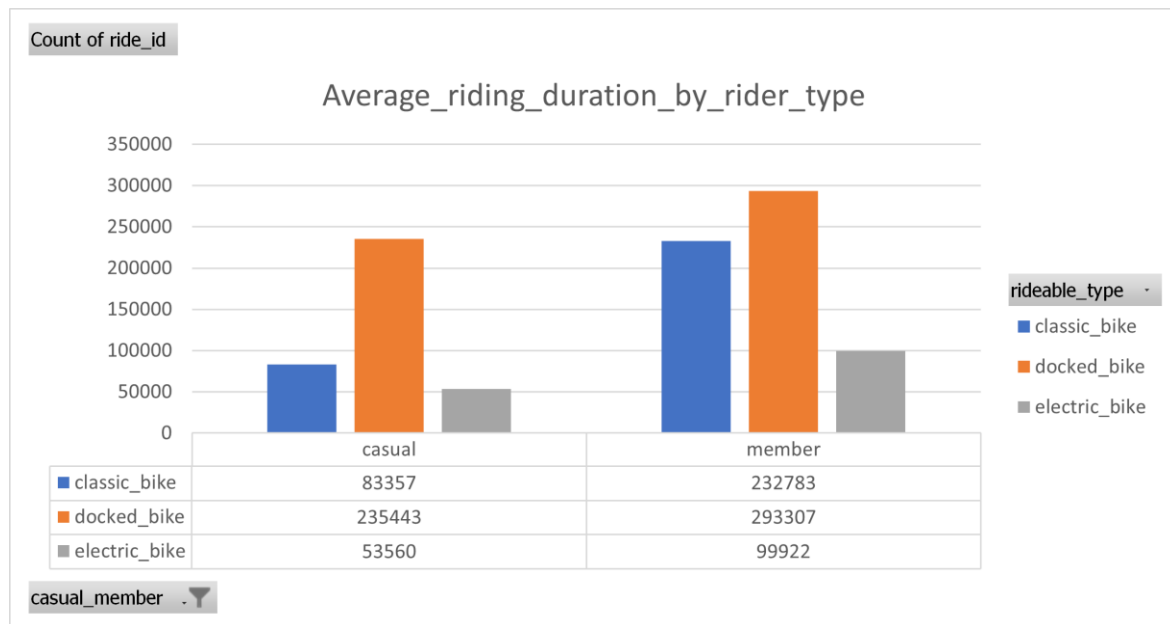6. Export the summary file for further analysis.

Share

Now that we have performed the analysis and gained some insights into the data, create visualizations to share our findings.
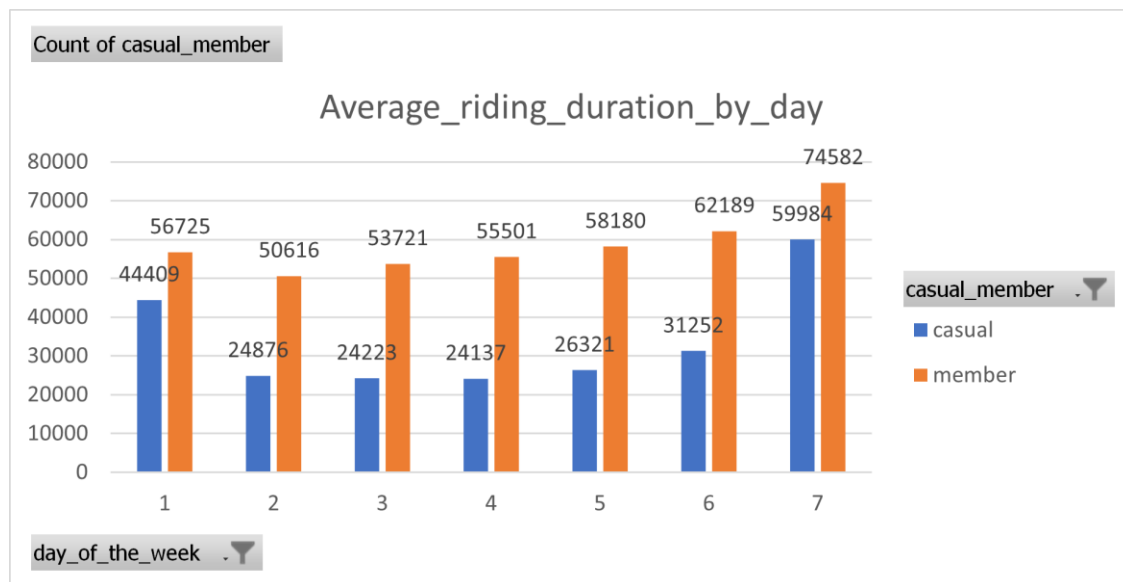
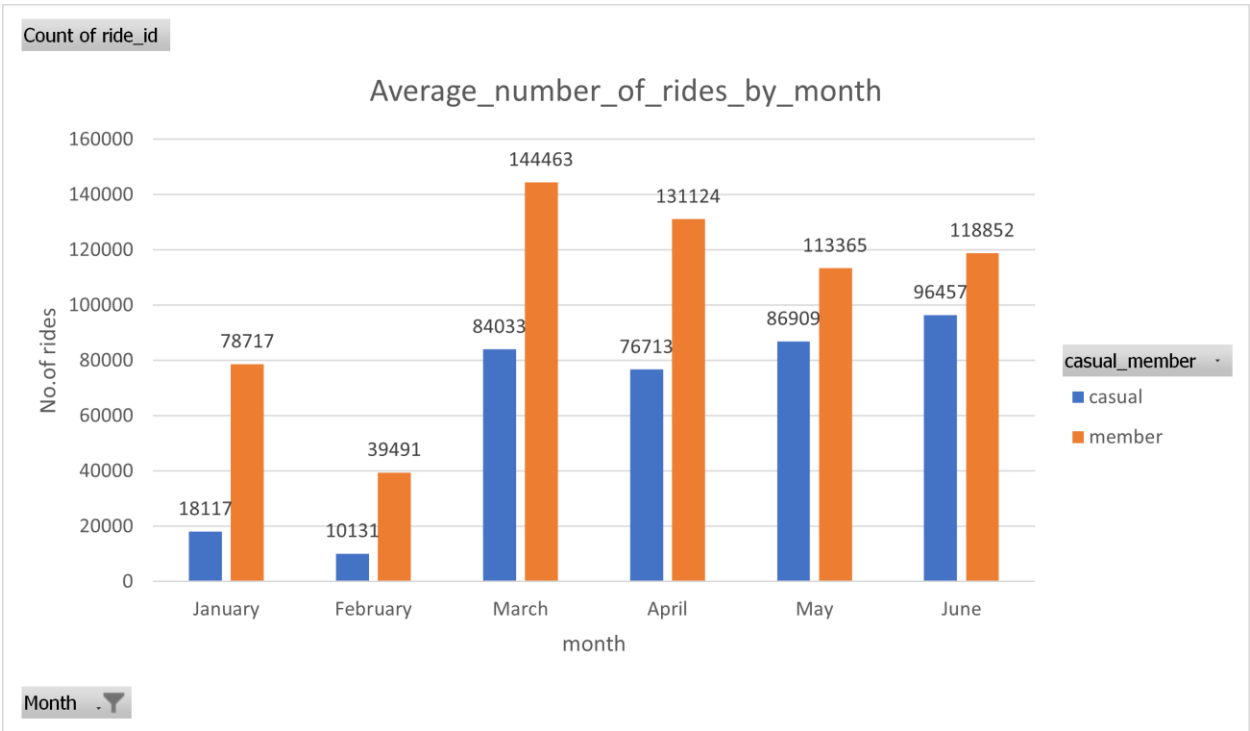Average ride length by casual and member
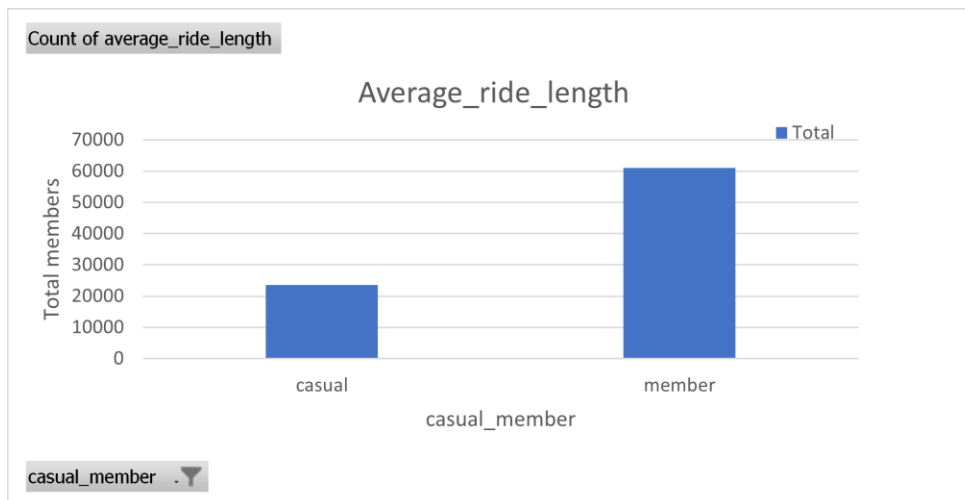
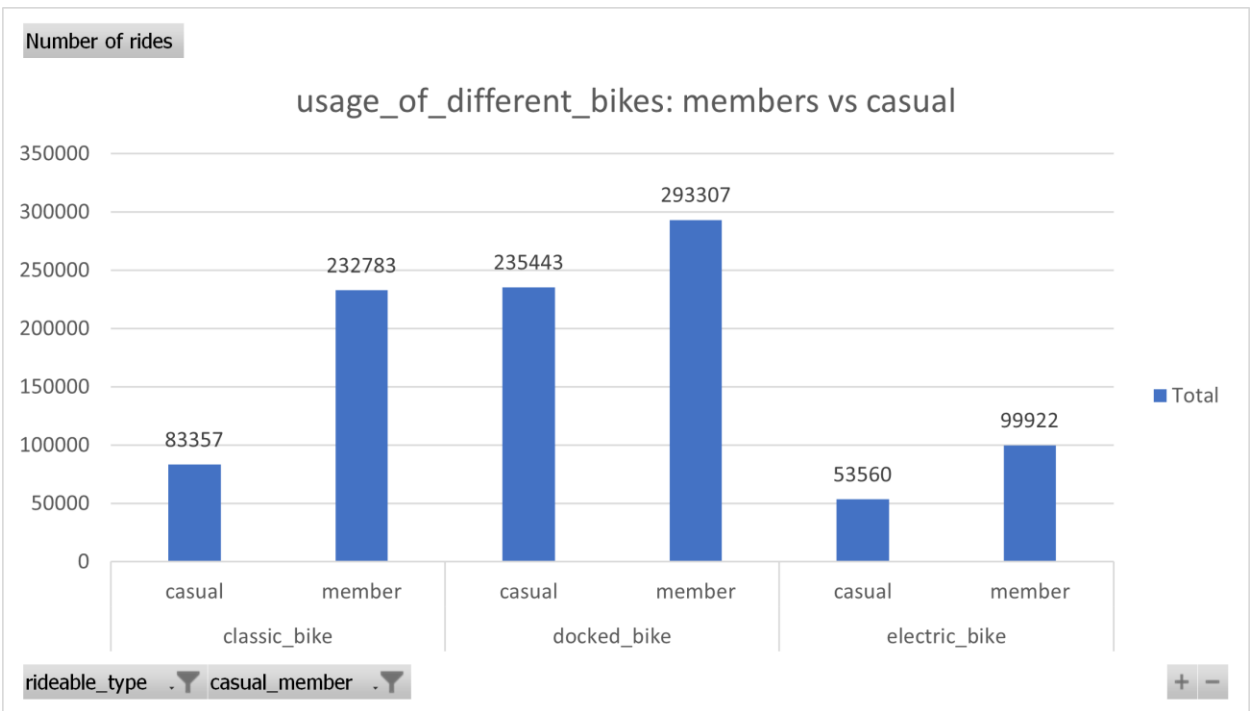Average riding duration between members and casual riders.

Average_riding_duration_by_rider_type

	casual	member
■ classic_bike	83357	232783
■ docked_bike	235443	293307
■ electric_bike	53560	99922

rideable_type
- ■ classic_bike
- ■ docked_bike
- ■ electric_bike

casual_member

Average riding duration by day: Members vs Casual riders

Count of casual_member

Average_riding_duration_by_day

casual_member
- ■ casual
- ■ member

casual values: 44409, 24876, 24223, 24137, 26321, 31252, 59984

member values: 56725, 50616, 53721, 55501, 58180, 62189, 74582

day_of_the_week

Average number of rides by month: Members vs Casual Riders

Count of ride_id

Average_number_of_rides_by_month

No.of rides

18117	10131	84033	76713	86909	96457
78717	39491	144463	131124	113365	118852

160000
140000
120000
100000
80000
60000
40000
20000
0

January   February   March   April   May   June

month

casual_member

■ casual
■ member

Month

Usage of different bikes by rider type

Number of rides

usage_of_different_bikes: members vs casual

350000
300000
250000
200000
150000
100000
50000
0

83357	232783	235443	293307	53560	99922

casual   member   casual   member   casual   member
classic_bike   docked_bike   electric_bike

■ Total

rideable_type   casual_member   + −
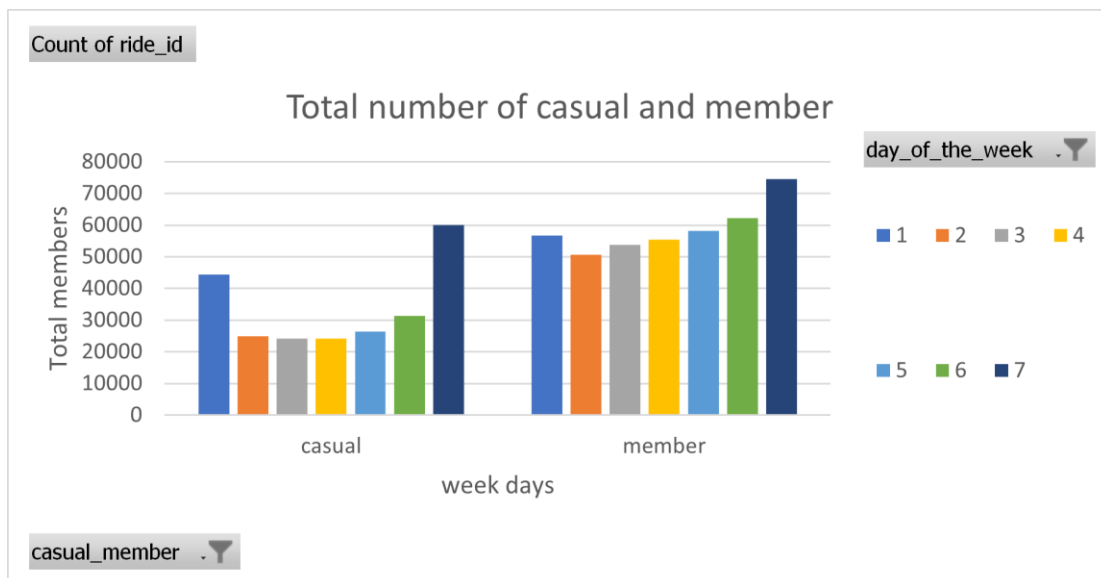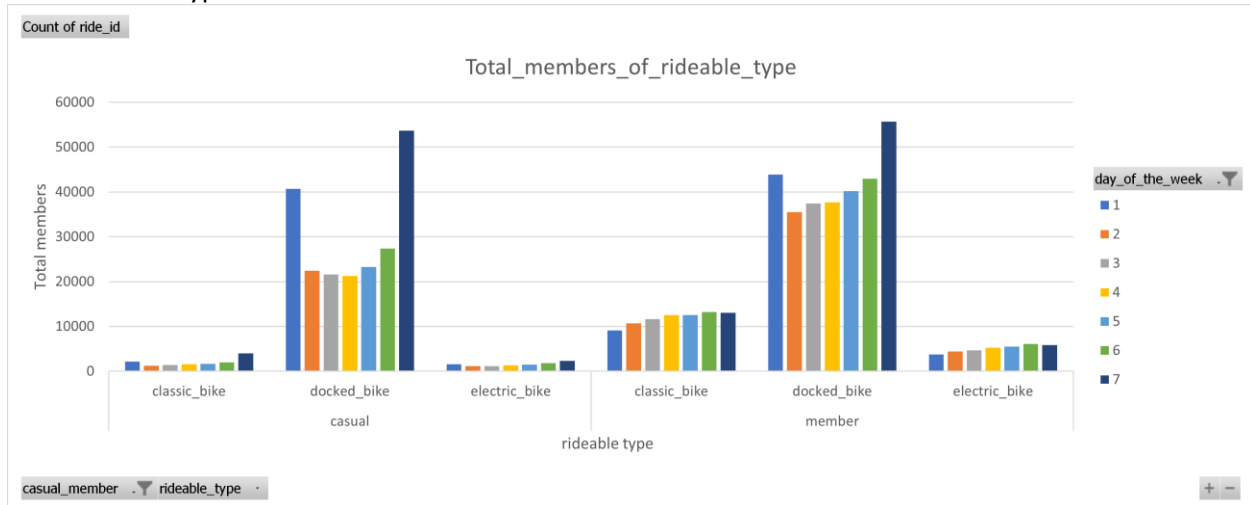
Number of rides between members and casual rides by day of week across six months
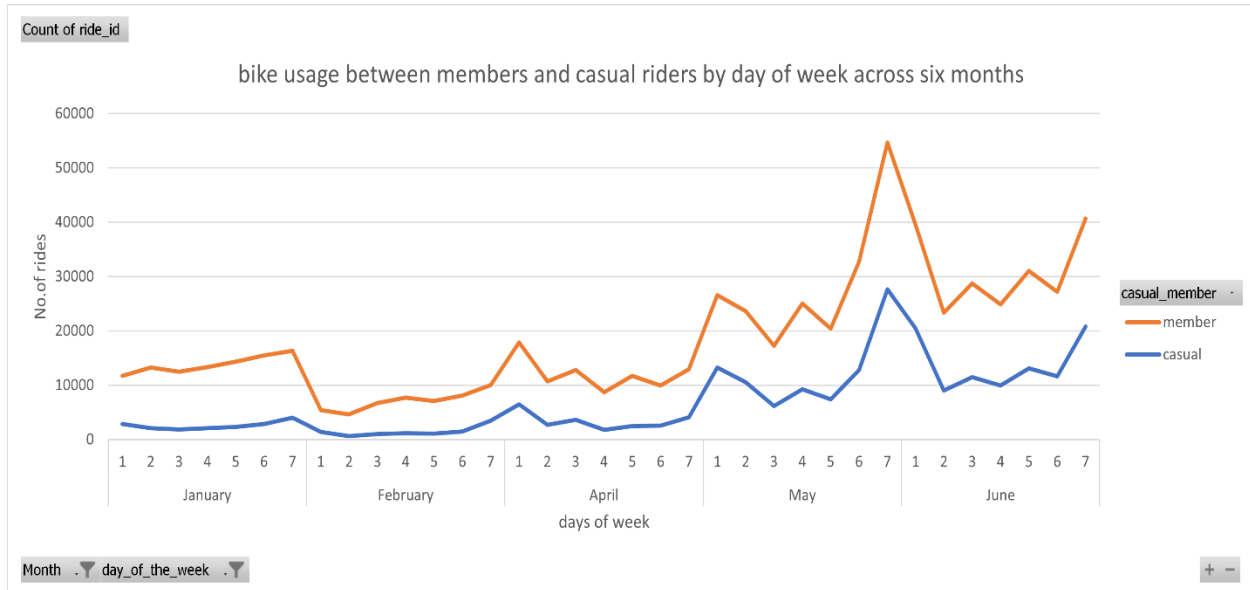
Total number of casual and members from 2020-2021

Total rideable type in 2020-2021

Key findings:

- From my analysis of 2020-2021 data set I have observed that the casual members are less when compared to membership members.
- Most of the casual ad membership people used docked bikes
- casual people in certain area choose to ride classic bikes.
- casuals prefer longer trips especially Saturday and Sunday

Act

- Thus, we could increase the ride price of the bikes for weekends to target casual members to turn into membership especially for docked bikes. Providing a good offer to membership leads casual people can attract to it.
- Providing offers for member during summer may also attract casual riders.
- Marketing more effectively in social media and advertisements.

References:

https://www.kaggle.com/alexzixinhan/bike-share-data-analysis-in-r/notebook